



ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ

ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΤΟΠΟΓΡΑΦΙΑΣ & ΓΕΩΠΛΗΡΟΦΟΡΙΚΗΣ

ΠΜΣ: ΓΕΩΧΩΡΙΚΕΣ ΤΕΧΝΟΛΟΓΙΕΣ

ΜΕΘΟΔΟΙ ΚΑΙ ΤΕΧΝΙΚΕΣ ΑΝΙΧΝΕΥΣΗΣ & ΑΝΑΓΝΩΡΙΣΗΣ
ΣΗΜΑΤΩΝ ΟΔΙΚΗΣ ΚΥΚΛΟΦΟΡΙΑΣ

Μεταπτυχιακή Διπλωματική Εργασία

Νικολάου Αθηνά

ΑΜ:1611

Τριμελής Επιτροπή

Επιβλέποντες:

Γραμματικόπουλος Λ.

Πέτσα Ε.

Καλησπεράκης Η.

Αθήνα, Σεπτέμβριος 2020

Ευχαριστίες

Η διπλωματική αυτή εργασία σηματοδοτεί την ολοκλήρωση των μεταπτυχιακών μου σπουδών στη Γεωπληροφορική. Θα ήθελα να ευχαριστήσω θερμά τον καθηγητή μου, Λάζαρο Γραμματικόπουλο, και όλη την ομάδα του Εργαστηρίου Φωτογραμμετρίας και Όρασης Υπολογιστών, Γιώργο Καρρά, Έλλη Πέτσα και Ηλία Καλησπεράκη, για τη βοήθεια, την ενθάρρυνση και τη συνεργασία τους, όχι μόνο κατά την εκπόνηση της διπλωματικής εργασίας, αλλά καθ' όλη τη διάρκεια των προπτυχιακών και μεταπτυχιακών μου σπουδών. Ένα ακόμα μεγάλο ευχαριστώ στον Παναγιώτη Φιλντίση και στη Βάλια Κυπαρίσση για την αμέριστη βοήθεια και συμπαράστασή τους κυρίως τους τελευταίους αυτούς μήνες, καθώς και στους πολύ κοντινούς μου ανθρώπους που με κάθε τρόπο ήταν δίπλα μου σε όλη αυτή την προσπάθεια. Τέλος, ευχαριστώ πάνω από όλους τους γονείς μου και τον αδερφό μου για όλη τη στήριξη, εμπιστοσύνη και αρωγή, καθώς χωρίς αυτούς δεν θα είχα φτάσει εδώ που βρίσκομαι σήμερα.

Στους γονείς και τον αδερφό μου..

ΠΕΡΙΛΗΨΗ

Η μεταπτυχιακή διπλωματική εργασία με τίτλο «Μέθοδοι και Τεχνικές Ανίχνευσης & Αναγνώρισης σημάτων οδικής κυκλοφορίας» εκπονήθηκε στο πλαίσιο του Μεταπτυχιακού Προγράμματος «Γεωχωρικές Τεχνολογίες» του Τμήματος Μηχανικών Τοπογραφίας και Γεωπληροφορικής του Πανεπιστημίου Δυτικής Αττικής. Η εργασία είναι αποτέλεσμα εκτενούς βιβλιογραφικής έρευνας σε μεθόδους Όρασης Υπολογιστών (Computer Vision) και Μηχανικής Μάθησης (Machine Learning), με στόχο τη χρήση τους σε πλήθος εφαρμογών Προηγμένων Συστημάτων Υποβοήθησης της Οδήγησης (Advanced Driving Assistance Systems) για την ενίσχυση της ασφαλούς οδήγησης. Για τη δημιουργία τέτοιων συστημάτων απαιτούνται διάφορες επιμέρους εφαρμογές εντοπισμού και αναγνώρισης σημάτων οδικής κυκλοφορίας, φωτεινών σηματοδοτών, εμποδίων και πεζών, ακόμα και άλλων οχημάτων. Στην παρούσα εργασία παρουσιάζονται οι μέθοδοι και τεχνικές ανίχνευσης και αναγνώρισης σημάτων οδικής κυκλοφορίας, διαχωρισμένες σε δύο βασικές κατηγορίες, την προσέγγιση με κλασικές μεθόδους και την προσέγγιση με μεθόδους State – Of – The – Art. Οι κλασικές μέθοδοι αντιμετωπίζουν την ανίχνευση και την αναγνώριση του αντικειμένου ως δύο επιμέρους ανεξάρτητες διαδικασίες, οι οποίες μπορούν να αντιμετωπιστούν με μεθόδους Επεξεργασίας Εικόνας (Image Processing) ή / και Μηχανικής Μάθησης. Τα τελευταία χρόνια όμως λαμβάνουν χώρα οι σύγχρονες τεχνικές της Βαθιάς Μηχανικής Μάθησης (Deep Learning). Οι αλγόριθμοι Βαθιάς Μηχανικής Μάθησης επιτρέπουν η ανίχνευση και η αναγνώριση να γίνουν σε ένα στάδιο. Επιπλέον, τα τελευταία χρόνια αρκετές μέθοδοι ανίχνευσης και αναγνώρισης σημάτων οδικής κυκλοφορίας βασίζονται και στην τεχνολογία σαρωτών laser LiDAR. Τέλος, στα πλαίσια της μεταπτυχιακής διπλωματικής εργασίας, ακολουθεί η εφαρμογή ενός από τους State-Of-The-Art αλγορίθμους, του αλγορίθμου YOLOv3, ο οποίος εκπαιδεύτηκε σε ένα Γερμανικό σύνολο δεδομένων και τελικά εφαρμόστηκε σε πραγματικά δεδομένα που ελήφθησαν για το σκοπό αυτό.

Λέξεις Κλειδιά: Όραση Υπολογιστών, Μηχανική Μάθηση, Ανίχνευση Αντικειμένου, Εντοπισμός, Αναγνώριση, Ταξινόμηση, Βαθιά Μάθηση, Συνελκτικά Νευρωνικά Δίκτυα, Αλγόριθμος YOLO, σαρωτές LiDAR, Σήματα Οδικής Κυκλοφορίας

School of Engineering
Department of Surveying & Geoinformatics Engineering
University of West Attica, Greece
Master of Science in Geospatial Technologies

**Traffic sign detection & recognition
methods and techniques**

Nikolaou Athina

Master Thesis
September 2020

ABSTRACT

The Master Thesis "*Methods and Techniques of Detection & Recognition of Road Traffic Signs*" was conducted in the framework of the Postgraduate Program "Geospatial Technologies" of the Department of Surveying and Geoinformatics Engineering of the University of West Attica. The thesis is the result of extensive literature research on Computer Vision and Machine Learning methods, with the aim of using them in a number of applications of Advanced Driving Assistance Systems to enhance safe driving. The creation of such systems requires several separate applications for the detection and recognition of traffic road signs, traffic lights, obstacles and pedestrians, and even other vehicles. In this work, methods and techniques of detection and recognition of road signs are presented, divided into two main categories, the approach with classical methods and the approach with State - Of - The - Art methods. Classical methods treat object detection and recognition as two separate independent processes, which can be addressed by Image Processing and / or Machine Learning methods. In recent years, however, modern techniques of Deep Learning are being applied. Deep Machine Learning algorithms allow detection and recognition in one step. In addition, several methods of detecting and recognizing traffic signs are lately based also on LiDAR laser scanner technology. Finally, in the framework of the thesis, the application of one of the State-Of-The-Art algorithms, YOLOv3, is presented, which was trained using a German dataset and eventually applied to real dataset obtained for this purpose.

Keywords: Computer Vision, Machine Learning, Object Detection, Localization, Recognition, Classification, Deep Learning, Convolutional Neural Networks, YOLO algorithm, LiDAR, Traffic Signs

Περιεχόμενα

ΕΙΣΑΓΩΓΗ	1
ΓΕΝΙΚΟ ΠΛΑΙΣΙΟ	1
ΠΡΟΒΛΗΜΑΤΑ ΣΤΗ ΔΙΑΔΙΚΑΣΙΑ	2
ΔΙΑΡΘΡΩΣΗ ΕΡΓΑΣΙΑΣ	3
Η ΚΛΑΣΙΚΗ ΠΡΟΣΕΓΓΙΣΗ	4
1.1 ΕΝΤΟΠΙΣΜΟΣ // IMAGE PROCESSING METHODS	5
1.1.1 <i>Εντοπισμός με βάση το χρώμα</i>	6
1.1.1.1 Μέθοδοι κατάτμησης εικόνας	6
1.1.1.2 Ο χρωματικός χώρος RGB	8
1.1.1.3 Οι Χρωματικοί χώροι HSV & HIS	10
1.1.1.4 Άλλοι χρωματικοί χώροι CIEL*a*b*, CIEL*u*v* & YCrCb	13
1.1.2 <i>Εντοπισμός με βάση το σχήμα</i>	16
1.1.2.1 Μετασχηματισμός Hough	17
1.1.2.2 Γρήγορος Μετασχηματισμό Ακτινικής Συμμετρίας	20
1.1.2.3 Απόσταση από Όρια	20
1.1.2.4 Συνταύτιση Προτύπων	22
1.1.2.5 Αλγόριθμος RANSAC	24
1.1.2.6 Αλγόριθμοι SIFT και SURF	26
1.2 Η ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ	29
1.2.1 <i>Είδη Μηχανικής Μάθησης</i>	29
1.2.1.1 Υπερπροσαρμογή και Υποπροσαρμογή μοντέλου	32
1.2.2 <i>Τεχνικές Μηχανικής Μάθησης</i>	32
1.2.2.1 Παλινδρόμηση (Regression)	32
1.2.2.2 Ταξινόμηση (Classification)	36
1.2.2.3 Ομαδοποίηση (Clustering)	37
1.2.2.4 Ανίχνευση «Ανωμαλιών» (Anomaly detection)	38
1.2.2.5 Ανάλυση Κύριων Συνιστωσών (PCA)	39
1.2.2.6 Γενετικοί αλγόριθμοι (Genetic algorithms)	40
1.2.3 <i>Μηχανές Διανυσματικής Στήριξης</i>	42
1.2.3.1 Πραγματικός, εμπειρικός και διαρθρωτικός κίνδυνος	43
1.2.3.2 Διάσταση VC	44
1.2.3.3 Γραμμικά και μη γραμμικά διαχωρίσιμα δεδομένα	45
1.2.3.4 Συναρτήσεις Πυρήνα	46
1.2.4 <i>Νευρωνικά Δίκτυα</i>	48
1.2.4.1 Βασικές συναρτήσεις ενεργοποίησης	51
1.2.4.2 Αρχιτεκτονική νευρωνικών δικτύων	53
1.2.4.3 Perceptron	54
1.2.4.4 Πολυεπίπεδα Νευρωνικά Δίκτυα	54
1.2.4.5 Εκπαίδευση & Αξιολόγηση Νευρωνικών δικτύων	56
1.2.4.6 Αλγόριθμος οπισθοδιάδοσης σφάλματος (Back-propagation)	58
1.2.4.7 Κατάβαση Κλίσης (Gradient Descent)	60
1.3 ΕΝΤΟΠΙΣΜΟΣ // MACHINE LEARNING METHODS	61
1.3.1 <i>Η προσέγγιση Viola-Jones</i>	62
1.3.1.1 Χαρακτηριστικά τύπου Haar & Εικόνα integral	63
1.3.1.2 Ο αλγόριθμος AdaBoost	66
1.3.1.3 Ανιχνευτής Cascade	68
1.3.1.4 Η μέθοδος στον εντοπισμό σημάτων	69
1.3.2 <i>Τεχνικές εκμάθησης σχήματος</i>	72
1.3.3 <i>Τεχνικές εκμάθησης χρώματος</i>	75
1.4 ΑΝΑΓΝΩΡΙΣΗ ΜΕ ΜΕΘΟΔΟΥΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ	77

1.4.1 Μηχανές Διανυσματικής Στήριξης	77
1.4.2 Νευρωνικά Δίκτυα	82
Η ΣΥΓΧΡΟΝΗ ΠΡΟΣΕΓΓΙΣΗ (STATE OF THE ART)	88
2.1 ΕΝΤΟΠΙΣΜΟΣ ΚΑΙ ΑΝΑΓΝΩΡΙΣΗ ΣΕ ΕΝΑ ΣΤΑΔΙΟ // DEEP LEARNING METHODS	89
2.1.1 Συνελικτικά Νευρωνικά Δίκτυα (CNN)	90
2.1.1.1 Αρχιτεκτονική	91
2.1.1.2 Συνέλιξη	92
2.1.1.3 Συγκέντρωση (Pooling)	93
2.1.1.4 Μη γραμμικότητα	94
2.1.1.5 Αλγόριθμος οπισθοδιάδοσης σφάλματος	96
2.1.1.6 Ανίχνευση και Αναγνώριση με δίκτυα CNN	96
2.1.2 Αλγόριθμος R-CNN	102
2.1.3.1 Ανίχνευση και Αναγνώριση με R-CNN	104
2.1.3 Αλγόριθμος YOLO	106
2.1.3.1 Μέση αντιπροσωπευτική ακρίβεια mAP	110
2.1.3.2 Ανίχνευση και Αναγνώριση με YOLO	111
2.2 ΕΝΤΟΠΙΣΜΟΣ ΚΑΙ ΑΝΑΓΝΩΡΙΣΗ ΜΕ ΧΡΗΣΗ LIDAR	114
ΕΦΑΡΜΟΓΗ	118
3.1 Ο ΑΛΓΟΡΙΘΜΟΣ YOLOv3	118
3.2 ΤΟ ΣΥΝΟΛΟ ΔΕΔΟΜΕΝΩΝ GTSDb	121
3.3 ΕΚΠΑΙΔΕΥΣΗ ΑΛΓΟΡΙΘΜΟΥ	122
3.4 ΑΠΟΤΕΛΕΣΜΑΤΑ	123
3.4.1 Αποτελέσματα YOLOv3	124
3.4.2 Αποτελέσματα YOLOv3-tiny	125
3.4.3 Σύγκριση και αξιολόγηση αποτελεσμάτων	126
ΣΥΜΠΕΡΑΣΜΑΤΑ	138
ΒΙΒΛΙΟΓΡΑΦΙΑ	140

Εισαγωγή

Γενικό πλαίσιο

Η αυτόματη ανίχνευση και αναγνώριση αντικειμένων έχει μετατραπεί σε πραγματική πρόκληση για την υψηλή απόδοση των τεχνικών της Όρασης Υπολογιστών (Computer Vision) και της Μηχανικής Μάθησης (Machine Learning) κυρίως την τελευταία δεκαετία, ενώ απασχολεί τις κοινότητες αυτές για πάνω από 30 χρόνια. Χαρακτηριστική εφαρμογή αποτελεί η δημιουργία προηγμένων συστημάτων για την υποβοήθηση της οδήγησης (Advanced Driver Assistance Systems - ADAS) ή για την πλήρως αυτοματοποιημένη οδήγηση, δηλαδή τα επονομαζόμενα ευφυή οχήματα (Intelligent Vehicles - IVS).

Στα πλαίσια ενός τέτοιου είδους συστήματος εμπίπτει ο εντοπισμός και η αναγνώριση των σημάτων οδικής κυκλοφορίας (Road Sign Recognition – RSR). Ένα σύστημα RSR θεωρείται πολύτιμο συμπλήρωμα του συστήματος πλοήγησης GPS. Ο δυναμικός χάρτης του GPS μπορεί να εμπλουτιστεί από τους τύπους και τις θέσεις των οδικών σημάτων (που αποκτήθηκαν από το RSR) και έτσι να υποβοηθηθεί η ακρίβεια της τρέχουσας θέσης του οχήματος.



Σήματα οδικής κυκλοφορίας
ΠΗΓΗ: http://www.aigaionmetal.gr/signs_4.html

Η ανάλυση σημάτων κυκλοφορίας μπορεί να χωριστεί σε δύο προβλήματα: αυτόματος εντοπισμός (ανίχνευση) και αναγνώριση (κατηγοριοποίηση) των οδικών σημάτων κυκλοφορίας. Προκειμένου να αποφευχθεί η σύγχυση μεταξύ των εννοιών εντοπισμός, ανίχνευση, αναγνώριση και κατηγοριοποίηση, αναφέρονται τα εξής:

- Ανίχνευση (detection): Νοείται η αναζήτηση ύπαρξης ενός σήματος οδικής κυκλοφορίας στην εικόνα, δηλαδή εάν υπάρχει ή όχι.

- Εντοπισμός (localisation): Αφού έχει εξακριβωθεί ότι υπάρχει το σήμα οδικής κυκλοφορίας στην εικόνα, εντοπίζεται η θέση του και εγγράφεται σε ένα πλαίσιο (bounding box).
- Κατηγοριοποίηση ή Ταξινόμηση (classification): Η διαδικασία της κατηγοριοποίησης ή ταξινόμησης των σημάτων σε δύο ή περισσότερες κλάσεις με βάση τα χαρακτηριστικά τους.
- Αναγνώριση (recognition): Το αποτέλεσμα της διαδικασίας της κατηγοριοποίησης ή ταξινόμησης, δηλαδή ποιο είναι τελικά αυτό το σήμα οδικής κυκλοφορίας (πχ. είναι το στοπ) ή σε ποια κλάση ανήκει, αν έχει οριστεί κάποια γενικότερη κλάση (πχ. σήμα προειδοποίησης κινδύνου).

Στην παρούσα διπλωματική εργασία, από εδώ και στο εξής οι δύο έννοιες ανίχνευση και εντοπισμός θα θεωρηθούν ταυτόσημες, αφού σχεδόν πάντα όταν γίνεται η ανίχνευση ακολουθεί και ο εντοπισμός και πάντα για να γίνει ο εντοπισμός έχει προηγηθεί το στάδιο της ανίχνευσης. Αντίστοιχα, η έννοια της αναγνώρισης περιέχει την κατηγοριοποίηση ή ταξινόμηση, επομένως μπορεί να θεωρηθεί επίσης πως έχουν την ίδια σημασία.

Οι περισσότερες προσεγγίσεις στον εντοπισμό σημάτων κυκλοφορίας βασίζονται σε μεθόδους επεξεργασίας εικόνας (image processing) και πιο συγκριμένα στην εξαγωγή πληροφορίας χρώματος (color-based), συνήθως με κατωφλίωση και σχήματος (shape-based), όπως με τη χρήση μετασχηματισμού HOUGH ή ανιχνευτών (detectors) όπως ο HOG και ο Canny edge detector, με τελικό σκοπό την κατάτμηση της εικόνας.

Το πρόβλημα της αναγνώρισης και κατηγοριοποίησης των σημάτων κυκλοφορίας έχει αντιμετωπιστεί σε ένα μεγάλο βαθμό, είτε με τεχνικές ταξινόμησης όπως την απλή συνταύτιση προτύπων (template matching) είτε με πιο εκλεπτυσμένες τεχνικές, οι οποίες προέρχονται από το πεδίο της Μηχανικής Μάθησης (Machine Learning), όπως μηχανές διανυσματικής στήριξης (SVM), νευρωνικά δίκτυα (neural networks) κ.ά.

Πλέον, με τη ραγδαία ανάπτυξη κυρίως της υπολογιστικής ισχύος, έχει εμφανιστεί η πρόκληση της ταυτόχρονης επίλυσης των δύο προβλημάτων (εντοπισμός και αναγνώριση) σε ένα ενιαίο μοντέλο. Δεν είναι είδηση ότι οι τεχνικές βαθιάς μηχανικής μάθησης (deep learning) υπήρξαν καθοριστικές και ανατρεπτικές στη μηχανική μάθηση και ειδικότερα στην όραση υπολογιστών. Με παρόμοιο τρόπο που τα μοντέλα βαθιάς μάθησης έχουν παραγκωνήσει άλλα κλασσικά μοντέλα στο έργο της ταξινόμησης των εικόνων, τα μοντέλα βαθιάς μάθησης είναι πλέον state-of-the-art στην ανίχνευση και αναγνώριση αντικειμένων. Οι τεχνικές της βαθιάς μάθησης περιλαμβάνουν κυρίως συνελκτικτά νευρωνικά δίκτυα (convolutional neural networks – CNN's) και σύνολα φίλτρων. Οι μέθοδοι αυτές χρησιμοποιούνται για τον αυτόματο εντοπισμό σημάτων οδικής κυκλοφορίας.

Επιπλέον, λόγω της σημαντικής ανάπτυξης της τεχνολογίας σάρωσης με laser / LiDAR σε φορητές συσκευές τα τελευταία χρόνια, τρισδιάστατα αστικά αντικείμενα, μεταξύ άλλων και οι πινακίδες οδικής κυκλοφορίας, μπορούν να ανιχνευθούν και να ταξινομηθούν για διαφορετικούς σκοπούς, ιδιαίτερα για τη δημιουργία προηγμένων συστημάτων υποβοήθησης της οδήγησης.

Προβλήματα στη διαδικασία

Η διαδικασία του αυτόματου εντοπισμού και της κατηγοριοποίησης των σημάτων οδικής κυκλοφορίας είναι, εκ πρώτης όψεως, ως στόχος σαφής και φαίνεται να είναι αρκετά απλός. Οι οδικές πινακίδες εμφανίζονται σε προκαθορισμένες θέσεις στις κυκλοφοριακές σκηνές, και τα σχήματα, τα χρώματα και τα εικονογράμματα είναι γνωστά (λόγω των διεθνών προδιαγραφών). Ο εντοπισμός, όμως, και η αναγνώριση σημάτων με μοναδικό δεδομένο τις εικόνες που προέρχονται από κάμερα, τοποθετημένη στο αυτοκίνητο για τη δημιουργία ενός τέτοιου αυτόνομου συστήματος, είναι ένα αρκετά περίπλοκο εγχείρημα, όπου ανακύπτουν διάφορα προβλήματα. Ανάμεσά τους συγκαταλέγονται οι έντονες εναλλαγές στις συνθήκες φωτεινότητας, λόγω δυσμενών εξωτερικών παραγόντων, όπως η εναλλαγή ημέρας-νύχτας,

και καιρικών συνθηκών, όπως η βροχή, η ομίχλη ή ακόμα και η ηλιοφάνεια η οποία δημιουργεί έντονες ανακλάσεις.

Οι αποκρύψεις, οι αλλοιώσεις των πινακίδων, καθώς και η ύπαρξη στην εικόνα παρόμοιων οπτικά αντικειμένων, αποτελούν επίσης δυσκολίες οι οποίες πρέπει να αντιμετωπιστούν. Επιπλέον, δεδομένου ότι το μέγεθός αλλά και το σχήμα των εικονιζόμενων σημάτων οδικής κυκλοφορίας εξαρτάται από την απόσταση και την οπτική γωνία της κάμερας, εμφανίζονται αφινικές παραμορφώσεις. Η ανάπτυξη αλγορίθμων, οι οποίοι θα ανταποκρίνονται σε όλους τους διαφορετικούς φωτισμούς και θα χειρίζονται τις πιθανές στροφές και μεταθέσεις των σημάτων, είναι αναγκαία για την επίτευξη ικανοποιητικών αποτελεσμάτων, προκειμένου τα έξυπνα οχήματα να λειτουργούν σε πραγματικές συνθήκες κυκλοφορίας στο δρόμο.



Παραδείγματα μη ιδανικών εικόνων σημάτων οδικής κυκλοφορίας
ΠΗΓΗ: Random Google search (πάνω), Shao F. et al. (2019) [122] (κάτω)

Διάρθρωση εργασίας

Τα προαναφερθέντα περιγράφονται στα επόμενα, διαρθρωμένα σε 3 επιμέρους κεφάλαια. Στο **Κεφάλαιο 1** πραγματοποιείται μια εκτενής βιβλιογραφική ανασκόπηση των μεθόδων και τεχνικών που αποτελούν την κλασική προσέγγιση ανίχνευσης και αναγνώρισης σημάτων οδικής κυκλοφορίας. Σε αυτές συγκαταλέγεται ο εντοπισμός με μεθόδους επεξεργασίας εικόνας, ο εντοπισμός με μεθόδους μηχανικής μάθησης και η αναγνώριση ως ένα πρόβλημα ταξινόμησης της μηχανικής μάθησης. Το **Κεφάλαιο 2** αφορά επίσης μια εκτενή περιγραφή και βιβλιογραφική αναφορά των state-of-the-art τεχνικών που χρησιμοποιούνται αυτή τη στιγμή παγκοσμίως για την επιτυχή ανίχνευση και αναγνώριση πινακίδων, καθώς και του πώς αυτές λειτουργούν. Στο **Κεφάλαιο 3**, με το οποίο και ολοκληρώνεται η παρούσα μεταπτυχιακή διπλωματική εργασία, παρουσιάζεται η εφαρμογή των αλγορίθμων YOLOv3 και YOLOv3-tiny, αναλύονται δε και συγκρίνονται τα αποτελέσματά τους. Τέλος, στο **Κεφάλαιο 4** συνοψίζονται τα συμπεράσματα που προέκυψαν. Η εργασία ολοκληρώνεται με τη χρησιμοποιηθείσα βιβλιογραφία.

1

Η κλασική προσέγγιση

Αν και τα τελευταία χρόνια έχουν αναπτυχθεί διάφορες μέθοδοι για την αυτόματη ανίχνευση και την αναγνώριση οδικών σημάτων στα πεδία της όρασης υπολογιστών, επεξεργασίας εικόνας, μηχανικής μάθησης και των βαθιών νευρωνικών δικτύων, η πρώτη αναφορά μελέτης αυτοματοποιημένης αναγνώρισης οδικών πινακίδων σημειώνεται στην Ιαπωνία το 1984 (βλ. Paclik [1]).

Η πρώτη γνωστή έρευνα σχετικά με την αναγνώριση πινακίδων σε πραγματικό χρόνο ανάγεται στο 1987, όταν οι Akatsuka & Imai [2] πρότειναν, έναν πίνακα αναζήτησης στο N-RGB, προκειμένου να γίνει η κατάτμηση της εικόνας για πινακίδες οι οποίες υποδεικνύουν το όριο ταχύτητας που δεν πρέπει να υπερβαίνουν τα οχήματα. Η προσέγγιση αυτή έχει ανοχή απέναντι στις αλλαγές φωτεινότητας και στις σκιές. Από τότε και μέχρι σήμερα έχει αναπτυχθεί πληθώρα άλλων τεχνικών και μεθόδων.

Δύο ευρέως διαδεδομένες και συχνά χρησιμοποιούμενες μέθοδοι για τον εντοπισμό και την αναγνώριση αντικειμένων είναι εκείνες που βασίζονται στη χρήση των χαρακτηριστικών του ιστογράμματος με προσανατολισμούς ακμών (Histogram of Orientation Gradients - HOG) και εκείνες που βασίζονται στις μηχανές διανυσματικής στήριξης (SVM) για την ταξινόμηση, όπως έπραξαν οι Dalal & Triggs (2005) [8]. Η μέθοδος HOG εξάγει από την εικόνα πληροφορίες σχετικά με το σχήμα των αντικειμένων που περιέχει και το αποτέλεσμα της είναι ένα διάνυσμα περιγραφής σχήματος (shape descriptor) για την εικόνα. Δεδομένου ότι το σχήμα ενός αντικειμένου είναι η συνισταμένη όλων των ακμών που το αποτελούν, ο αλγόριθμος υπολογίζει ένα ιστόγραμμα με τους προσανατολισμούς των ακμών. Η ιδέα στην οποία στηρίζεται ο αλγόριθμος είναι ότι εικόνες που απεικονίζουν παρόμοιο σχήμα θα έχουν παρόμοιο ιστόγραμμα ακμών. Κατά συνέπεια η απόσταση μεταξύ των διανυσμάτων περιγραφής σχήματος αποτελεί και το μέτρο ομοιότητας των εικόνων (Μπράχου, 2010 [9]).

Παρόμοια μέθοδο υλοποίησε χρησιμοποίησε και η Αδάμ (2014) [10] στη μεταπτυχιακή διπλωματική της εργασία για τον αυτόματο εντοπισμό και την αναγνώριση σημάτων οδικής κυκλοφορίας. Υλοποιήθηκε εκεί εντοπισμός βάσει χρώματος, προκειμένου να εντοπιστούν οι περιοχές ενδιαφέροντος που πιθανώς περιέχουν πινακίδες, ακολούθησε ο κυκλικός μετασχηματισμός Hough, ο οποίος εκμεταλλεύεται τις γεωμετρικές ιδιότητες των σημάτων, και οι τελικές περιοχές ενδιαφέροντος αναπαρίστανται με την παραπάνω μέθοδο κάνοντας

χρήση των περιγραφέων HOG, ενώ στη συνέχεια για την αναγνώριση εισήχθησαν σε ταξινομητές SVM, οι οποίοι έχουν προηγουμένως εκπαιδευθεί.

Οι Escalera et al. (2011) [11] αναφέρουν πως τα οδικά σήματα παρουσιάζουν συγκεκριμένες ιδιότητες, όπως ο συνδυασμός χρώματος και σχήματος, που τα διαφοροποιούν από τα υπόλοιπα αντικείμενα. Τα συστήματα για την αυτόματη αναγνώριση των οδικών σημάτων σχεδιάζονται αποκλειστικά για τον εντοπισμό αυτών των ιδιοτήτων. Γενικά, τα συστήματα αναγνώρισης σημάτων κυκλοφορίας αποτελούνται από τη θέση της περιοχής ενδιαφέροντος και την κατάτμηση κατά την οποία δημιουργούνται πολλές δυαδικές (binary) μάσκες (φίλτρα) για τον διαχωρισμό των αντικειμένων από το φόντο. Συνήθως, εφαρμόζονται πληροφορίες χρώματος (color based detection), δεδομένου ότι τα σήματα οδήγησης χαρακτηρίζονται από έναν προκαθορισμένο αριθμό σχετικά σταθερών χρωμάτων (κόκκινο, μπλε, κίτρινο, λευκό και μαύρο), με αποτέλεσμα οι περιοχές ενδιαφέροντος να προσδιορίζονται ως συνδεδεμένα στοιχεία (connected components), τα οποία συνήθως αποτελούν σήματα κυκλοφορίας. Ακολουθεί η ανίχνευση με επαλήθευση της υπόθεσης για την παρουσία της πινακίδας, όπου συνήθως χρησιμοποιείται η πληροφορία του σχήματος (πχ. ισόπλευρα τρίγωνα, κύκλοι κ.λπ.), και τέλος η ταξινόμηση του τύπου της πινακίδας (αναγνώριση), όπου η αναγνώριση του σήματος πραγματοποιείται με χρήση κάποιας βάσης δεδομένων με όλα τα πιθανά μοντέλα των κυκλοφοριακών σημάτων και μεθόδων μηχανικής μάθησης, προκειμένου να επιτευχθεί ένα ισχυρό αποτέλεσμα αναγνώρισης.

Στα επόμενα κατηγοριοποιούνται και αναλύονται εν συντομία οι πιο διαδεδομένες μέθοδοι για τα στάδια της ανίχνευσης (εντοπισμού) και της αναγνώρισης (ταξινόμησης) με την «κλασσική» προσέγγιση που έχουν χρησιμοποιηθεί από πλήθος ερευνητών όλα αυτά τα χρόνια. Οι μέθοδοι αυτές για τον εντοπισμό των σημάτων οδικής κυκλοφορίας χρησιμοποιούν κυρίως τεχνικές από το πεδίο της επεξεργασίας εικόνας και της μηχανικής μάθησης, ενώ για την αναγνώριση των σημάτων χρησιμοποιούν αλγόριθμους μηχανικής μάθησης.

1.1 Εντοπισμός // Image Processing Methods

Το πρόβλημα της ανίχνευσης αντικειμένων ορίζεται ως η διαδικασία κατά την οποία ένα υπολογιστικό σύστημα δέχεται ως είσοδο μια εικόνα και παράγει ως έξοδο τις περιοχές της εικόνας στις οποίες εμφανίζεται κάποιο συγκεκριμένο αντικείμενο. Το αποτέλεσμα, δηλαδή, της διαδικασίας αποτελείται από την κατηγορία κάθε αντικειμένου που ανιχνεύθηκε καθώς και τη θέση και τη διάσταση του. Η ανίχνευση αντικειμένων διαφοροποιείται από την ταξινόμηση εικόνων (image classification), αφού στην ταξινόμηση εικόνων η έξοδος του υπολογιστικού συστήματος περιλαμβάνει μόνο την πληροφορία αν η εικόνα απεικονίζει ή όχι ένα συγκεκριμένο αντικείμενο, όχι τη θέση και τις διαστάσεις του (Barutimidis, 2008 [3]).

Η διαδικασία της ανίχνευσης υλοποιείται χρησιμοποιώντας τεχνικές που έχουν αναπτυχθεί στο πλαίσιο της θεωρίας αναγνώρισης προτύπων και βασίζεται στις χρωματικές και γεωμετρικές ιδιότητες των αντικειμένων. Αντίστοιχα, στην περίπτωση της ανίχνευσης σημάτων οδικής κυκλοφορίας, οι χρησιμοποιούμενες τεχνικές αξιοποιούν το χρώμα και το σχήμα των πινακίδων, αν και οι περισσότερες βασίζονται στην εξαγωγή χρωματικής πληροφορίας. Ένα φυσικό ερώτημα που τίθεται είναι ποιος είναι ο καταλληλότερος «χρωματικός χώρος» προκειμένου να διασφαλιστεί η καίρια εξαγωγή της χρωματικής πληροφορίας, απαλλαγμένη από τις επιδράσεις παραγόντων του εξωτερικού περιβάλλοντος, οι οποίοι δυσχεραίνουν το πρόβλημα της αυτόματης ανίχνευσης. Τέτοιοι παράγοντες είναι μεταβαλλόμενες συνθήκες φωτεινότητας είτε λόγω ώρας είτε λόγω καιρικών συνθηκών αλλά και η ύπαρξη σκιών και αντανακλάσεων. Δεδομένου ότι οι εικόνες που χρησιμοποιούνται για την ανίχνευση σημάτων οδικής κυκλοφορίας είναι εικόνες εξωτερικού χώρου και άρα έχουν έντονες επιδράσεις από τους παραπάνω παράγοντες, μερικοί συγγραφείς επικεντρώνονται αρκετά και στην ανίχνευση των σημάτων βάσει σχήματος (πχ. μετασχηματισμός Hough κ.ά).

Πρακτικά, ζητούμενο του πρώτου σταδίου της διαδικασίας, δηλαδή του σταδίου της ανίχνευσης, αποτελεί η εύρεση της περιοχής ενδιαφέροντος (Region of Interest - ROI), δηλαδή εκείνου του τμήματος της εικόνας, που περιέχει το υπό αναζήτηση αντικείμενο. Το στάδιο του εντοπισμού αποτελεί ιδιαίτερα κρίσιμη εργασία, καθώς όπως χαρακτηριστικά

αναφέρει η Αδάμ (2014 [10]), «η πληροφορία που απομακρύνεται κατά την εφαρμογή του δεν μπορεί να ανακτηθεί αργότερα». Ομοίως, και ο Fridtjof Stein [12] επισήμανε πως «η αξιόπιστη οπτική ανίχνευση αποτελεί το μεγαλύτερο εμπόδιο που πρέπει να ξεπεραστεί».

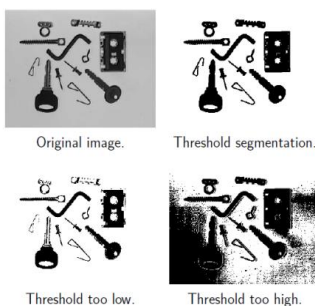
1.1.1 Εντοπισμός με βάση το χρώμα

Η ανίχνευση των σημάτων με βάση το χρώμα στηρίζεται στη φωτεινότητα και το χρώμα των εικονοψηφίδων της εκάστοτε εικόνας. Τα πέντε τυπικά χρώματα που ορίζονται για τα σήματα οδικής κυκλοφορίας είναι τα κόκκινο, μπλε, κίτρινο, λευκό και μαύρο. Οι περισσότεροι ερευνητές επιδιώκουν μια ισχυρή κατάτμηση εικόνας (image segmentation) βάση αυτών των χρωμάτων, δηλαδή μια ισχυρή ομαδοποίηση χρωματικά ομοιογενών περιοχών που έχουν το ίδιο χρώμα και οι οποίες αποτελούν τις περιοχές ενδιαφέροντος. Κατά την διαδικασία της κατάτμησης δίνεται ιδιαίτερη προσοχή στις χρησιμοποιούμενες τεχνικές ώστε να λαμβάνεται υπόψη ο μη ομοιογενής φωτισμός, δεδομένου ότι τα σφάλματα της κατάτμησης μπορούν να διαδοθούν και στα επόμενα στάδια της διαδικασίας.

1.1.1.1 Μέθοδοι κατάτμησης εικόνας

Όπως αναφέρεται στη βιβλιογραφία, η κατάτμηση (segmentation) συνοπτικά διαχωρίζεται με βάση δύο ιδιότητες της εικόνας, την ασυνέχεια (detecting discontinuities), δηλαδή την κατάτμηση εικόνας με γνώμονα τις απότομες αλλαγές που τη διέπουν, και την ομοιότητα (detecting similarities), τη διαίρεση δηλαδή της εικόνας σε παρόμοιες περιοχές σύμφωνα πάντα με προκαθορισμένα κριτήρια. Αναλυτικότερα, η κατάτμηση εικόνων κατηγοριοποιείται σε επτά βασικές μεθόδους σύμφωνα με τους Sharma et al. (2012) [13], οι οποίοι αναφέρονται στη μέθοδο των ομοιογενών περιοχών (region based techniques), τη μέθοδο συστάδων (clustering techniques), τη μέθοδο διαίρεσης-ένωσης (split-merge techniques), τη μέθοδο με γραφήματα (normalized cuts), τη μέθοδο επαύξησης της περιοχής ενδιαφέροντος (region growing), την κατωφλίωση (thresholding) και τη μέθοδο εντοπισμού ακμών (edge based techniques).

Οι Dass et al. (2012) [14] και οι Gurjinder et al. (2016) [15] πραγματοποιούν μια ανάλογη κατηγοριοποίηση των τεχνικών της κατάτμησης, με τους πρώτους να προσθέτουν τις τεχνικές που βασίζονται σε διαφορικές εξισώσεις και τη μέθοδο κατάτμησης με τη χρήση νευρωνικών δικτύων. Η τελευταία μέθοδος είναι εντελώς διαφορετική από τους συμβατικούς αλγορίθμους κατάτμησης, αφού η εικόνα χαρτογραφείται σε ένα νευρωνικό δίκτυο, όπου κάθε νευρώνας αντιπροσωπεύει ένα εικονοστοιχείο, έτσι το πρόβλημα κατάτμησης της εικόνας μετατρέπεται σε πρόβλημα ελαχιστοποίησης συγκεκριμένων συναρτήσεων. Οι Gurjinder et al. (2016) [15] ομαδοποιούν κάποιες από τις μεθόδους και να παραθέτουν εκτός της μεθόδου των νευρωνικών δικτύων και τη μέθοδο ομαδοποίησης με ασάφεια (Fuzzy c-means), ενώ παράλληλα πραγματοποιούν σύγκριση μεταξύ όλων των τεχνικών.



Κατωφλίωση

ΠΗΓΗ: Gurjinder et al. (2016) [15]

Ένεκα της μη απαίτησης για a priori πληροφορία της εικόνας και της χαμηλής υπολογιστικής πολυπλοκότητας, η τεχνική της κατωφλίωσης αποτελεί την πιο διαδεδομένη για το εν λόγω θέμα. Πρόκειται για διαδικασία της οποίας η εφαρμογή είναι εφικτό να διαχωρίσει τις εικονοψηφίδες που ανήκουν στην περιοχή ενδιαφέροντος (foreground object) και εκείνες

που ανήκουν στο υπόβαθρο (background object), αφού τα αντικείμενα μιας εικόνας συνήθως χαρακτηρίζονται από την τιμή φωτεινότητάς τους σε σχέση με το υπόβαθρο. Με τη χρήση κατωφλίων, την επιλογή δηλαδή τιμών φωτεινότητας ανάμεσα σε προκαθορισμένα όρια, πραγματοποιείται ο διαχωρισμός αυτός αποδίδοντας στην περιοχή που καλύπτει το αντικείμενο την τιμή 1, ενώ στο υπόβαθρο την τιμή 0. Έτσι, αποτέλεσμα της κατωφλίωσης είναι μια δυαδική (binary) εικόνα.

Όπως αναφέρει η Αδάμ (2014) [10], τα κατώφλια είναι είτε προκαθορισμένα, σταθερά και καθολικά, είτε προσαρμοστικά (adaptive) και καθορίζονται με βάση τις χρωματικές ιδιαιτερότητες του προς εντοπισμό αντικειμένου. Το σταθερό, καθολικό κατώφλι ορίζεται ενιαία για όλο το εύρος της εικόνας σύμφωνα με την παρακάτω σχέση:

$$G(i,j) = \begin{cases} 1 & \text{για } f(i,j) > \text{κατώφλι} \\ 0 & \text{για } f(i,j) < \text{κατώφλι} \end{cases}$$

όπου, $f(i,j)$ η εικόνα εισόδου και $G(i,j)$ η εικόνα εξόδου.

Η τιμή ενός δυναμικού κατωφλίου προσαρμόζεται στη φωτεινότητα της περιοχής γύρω από κάθε εικονοψηφίδα, επιλέγεται δηλαδή διαφορετική τιμή για κάθε pixel της εικόνας ή αλλάζει ανάλογα με την περιοχή, και ο καθορισμός των ορίων μπορεί να πραγματοποιηθεί είτε αυτόματα είτε από τον χρήστη. Σύμφωνα με τον Sergin (2004) [16] η ποιότητα των αποτελεσμάτων της κατωφλίωσης εξαρτάται από πλήθος παραγόντων, όπως είναι ο θόρυβος της εκάστοτε εικόνας, η φωτεινότητα και η αντίθεσή της αλλά και η ορθή επιλογή των κατωφλίων. Ο ίδιος κάνει λόγο για έξι διαφορετικές μεθόδους αυτόματης επιλογής κατωφλίων ανάλογα με τις πληροφορίες που εκμεταλλεύεται ο αλγόριθμος κάθε μεθόδου. Πιο συγκεκριμένα, αναφέρει τις μεθόδους οι οποίες είναι βασισμένες στο σχήμα του ιστογράμματος (histogram shape based methods) ανιχνεύοντας τις κορυφές, τις κοιλάδες και τις καμπυλώσεις του ιστογράμματος, αφού πρώτα αυτό έχει υποστεί εξομάλυνση, τις μεθόδους που βασίζονται σε αλγορίθμους συστάδας (clustering based methods), τις μεθόδους με τη χρήση εντροπίας (entropy based methods), τις μεθόδους με βάση τις ιδιότητες του αντικειμένου (object attribute based methods), τις χωρικές μεθόδους (spatial based methods) και τις τοπικές μεθόδους (local methods).

Για την προσπάθεια αυτοματοποίησης της επιλογής των κατωφλίων, αρκετές ερευνητικές εργασίες κάνουν αναφορά κυρίως στον αλγόριθμο Otsu, ενώ η πλειονότητα αυτών κάνουν λόγο για χειροκίνητη επιλογή τιμών των κατωφλίων. Ο αλγόριθμος του Otsu αποτελεί μια μη παραμετρική και μη επιβλεπόμενη μέθοδο αυτόματης επιλογής κατωφλίων για την κατάτμηση εικόνων με βάση την ομαδοποίηση. Ουσιαστικά δημιουργεί δύο κλάσεις εικονοστοιχείων (φόντο και προσκήνιο) σε μία εικόνα και βασίζεται είτε στην μεγιστοποίηση της μεταβλητότητας των διαφορετικών κλάσεων είτε στην ελαχιστοποίηση της μεταβλητότητας μεταξύ των ίδιων κλάσεων. Πιο συγκεκριμένα, για κάθε πιθανό κατώφλι (τιμές από 0 μέχρι 255) υπολογίζει την πιθανότητα ω και τη μεταβλητότητα σ^2_i για κάθε μία από τις δύο κλάσεις και επιλέγεται τελικά η τιμή που ικανοποιεί ένα από τα δύο κριτήρια (Otsu, 1975 [17]).

Όπως έχει ήδη αναφερθεί, βασικό στοιχείο για τον εντοπισμό των σημάτων οδικής κυκλοφορίας είναι τα προκαθορισμένα χρώματα που τα συνθέτουν. Κατά καιρούς έχουν προταθεί αρκετές διαφορετικές μέθοδοι για τον εντοπισμό του λευκού και του μαύρου, παρ' όλο που τα έντονα χρώματα κίτρινο, κόκκινο και μπλε καθιστούν πολύ πιο εύκολο και διαχειρίσιμο τον εντοπισμό. Υπάρχουν βέβαια και ερευνητικές εργασίες που σχετίζονται με την ανίχνευση σημάτων σε εικόνες σε κλίμακα του γκρι, όπως πράττουν οι Baro & Vitria (2012) [18], κάνοντας όμως χρήση του αλγορίθμου μηχανικής μάθησης Adaboost, όπως θα αναλυθεί στα επόμενα.

Η κατάτμηση της εικόνας με βάση το χρώμα μπορεί να πραγματοποιηθεί σε διάφορους χρωματικούς χώρους, όπως είναι οι RGB, HSI, HSV, CMYK, ο χώρος CIE Lab κ.ά., ανάλογα με τα χαρακτηριστικά του κάθε χρωματικού χώρου που είναι επιθυμητά για την επιλεγμένη μέθοδο (παράδειγμα ενός τέτοιου χαρακτηριστικού αποτελεί η ευαισθησία στο φωτισμό). Σύμφωνα με την Βrκίς (2010) [19], προκειμένου να κατανοηθεί σε βάθος ο λόγος που ορισμένοι χρωματικοί χώροι θεωρούνται πιο ευαίσθητοι στο φωτισμό και κάποιοι όχι,

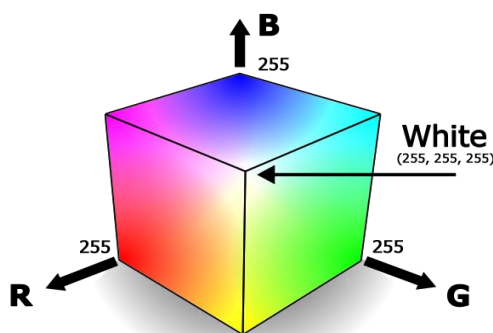
σημειώνεται πως ο κάθε χρωματικός χώρος προκύπτει από ένα χρωματικό μοντέλο, δηλαδή ένα αφηρημένο μαθηματικό μοντέλο που αναπαριστά τα χρώματα ως πλειάδες αριθμών.

Ουσιαστικά τα διάφορα χρωματικά μοντέλα έχουν αναπτυχθεί για να γίνει δυνατή η περιγραφή των χρωμάτων με μαθηματική μορφή, κατάλληλη για την επεξεργασία τους από ψηφιακά μέσα. Έτσι κατ' αρχήν η εικόνα αναλύεται σε εικονοστοιχεία, καθένα από τα οποία είναι χρωματικά ομοιογενές. Οι πιο συνηθισμένες διαστάσεις των πλειάδων, δηλαδή του χρωματικού χώρου, είναι οι τρεις ή οι τέσσερις. Σε κάποιες περιπτώσεις μπορεί να πραγματοποιηθεί εντοπισμός σε περισσότερους του ενός χρωματικούς χώρους. Οι Shadeed et al. (2003) [20] πραγματοποίησαν τον εντοπισμό των πινακίδων οδικής κυκλοφορίας σε δύο χρωματικούς χώρους. Στην τεχνική τους ενσωμάτωσαν τόσο τον χρωματικό χώρο HSV όσο και τον YUV, προκειμένου να επιτευχθούν πιο αξιόπιστα αποτελέσματα στην κατάτμηση των εικόνων.

1.1.1.2 Ο χρωματικός χώρος RGB

Η ονομασία του χώρου RGB αποτελεί ακρωνύμιο των λέξεων Red Green Blue (Κόκκινο Πράσινο Μπλε). Με τον συνδυασμό (υπέρθυση με συγκεκριμένη αναλογία) των τριών βασικών χρωμάτων προκύπτουν όλα τα υπόλοιπα χρώματα που μπορούν να εμφανιστούν σε μια οθόνη (συνήθως υπολογιστή). Στην μορφή του χρωματικού αυτού μοντέλου με βάθος χρώματος 8 δυαδικά ψηφία (bit), κάθε χρώμα μπορεί να παρασταθεί με μία τριάδα αριθμών και τιμές από 0 έως 255. Το μοντέλο βασίζεται στο γεγονός ότι όταν μία οθόνη δεν εκπέμπει φως εμφανίζεται μαύρη.

Το μοντέλο αυτό ορίζεται με έναν κύβο χρωμάτων σε ένα καρτεσιανό σύστημα συντεταγμένων, όπου οι άξονες x, y, z αναπαριστούν τα βασικά χρώματα. Στην αρχή των αξόνων είναι η κορυφή του κύβου που αντιστοιχεί στο μαύρο χρώμα, ενώ στις κορυφές του κύβου που βρίσκονται πάνω στους άξονες βρίσκονται τα βασικά χρώματα (Κόκκινο, Πράσινο, Μπλε). Βασικό πλεονέκτημα του χώρου RGB αποτελεί η υπολογιστική ταχύτητα, αφού δεν απαιτείται κάποιος μετασχηματισμός σε άλλο σύστημα.



Χρωματικός χώρος RGB

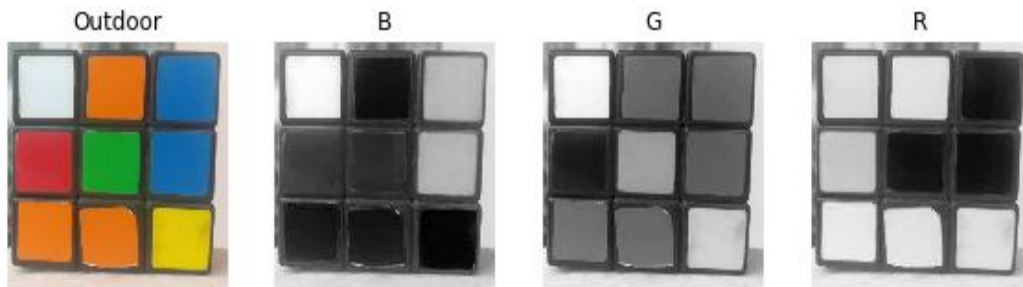
ΠΗΓΗ: <https://old.medialooks.com/mplatform/docs/CK%20Advanced.html>

Οι Benallal & Meunier (2003) [21] μελέτησαν τις αλλαγές στα χρώματα των σημάτων κυκλοφορίας ανάλογα με την ώρα της ημέρας (ανατολή, μεσημέρι, ηλιοβασίλεμα). Κατέληξαν στο γεγονός ότι οι αλλαγές του εξωτερικού φωτισμού δεν επηρεάζουν σημαντικά τις διαφοροποιήσεις των συνδεδεμένων χρωματικά περιοχών των σημάτων και πρότειναν ως πρώτο βήμα έναν απλό αλγόριθμο για την κατάτμηση των εικόνων, με σκοπό στη συνέχεια να πραγματοποιηθεί ο αυτόματος εντοπισμός.

Οι Broggi et al. (2007) [22] πραγματοποίησαν κατωφλίωση στον χώρο RGB, δεσμεύοντας τον λόγο μεταξύ των καναλιών και όχι μόνο το ίδιο το κανάλι, όπως απεικονίζεται στη συνέχεια, ενώ οι μέγιστες και ελάχιστες τιμές των παραμέτρων α , β και γ καθορίστηκαν έπειτα από δοκιμές σε πλήθος εικόνων:

$$\{\alpha_{\min} * G < R < \alpha_{\max} * G \quad \beta_{\min} * B < R < \beta_{\max} * B \quad \gamma_{\min} * B < G < \gamma_{\max} * B$$

Ένα από τα κύρια προβλήματα που αντιμετωπίστηκαν στη δημοσίευση αυτή ήταν η εξάρτηση του αποτελέσματος της κατωφλίωσης από την μεταβολή του χρώματος της πηγής φωτός κατά τη διάρκεια της αυγής ή του ηλιοβασιλέματος, οπότε κυριαρχεί το κόκκινο χρώμα με αποτέλεσμα η κατωφλίωση να μην είναι αποτελεσματική. Για την επίλυση αυτού το πρόβληματος αναπτύχθηκε μια χρωματική εξίσωση η οποία βασίζεται στην αναγνώριση του χρώματος της πηγής φωτός και την χρωματική εξισορρόπηση με τον μετασχηματισμό γάμμα (gamma correction). Για την αξιολόγηση της φωτεινότητας εντοπίζεται, συνήθως, μια επιφάνεια λευκού χρώματος στην εικόνα και συγκρίνονται οι τιμές φωτεινότητας της επιφάνειας αυτής με τις τιμές του απόλυτου λευκού (255, 255, 255). Επειδή, όμως, σε ένα δυναμικό περιβάλλον, όπως αυτό των δρόμων, είναι δύσκολο να βρεθεί μία λευκή επιφάνεια αναφοράς, οι ερευνητές χρησιμοποίησαν ως επιφάνεια αναφοράς το οδόστρωμα και το συνέκριναν με τις τιμές φωτεινότητας του οδοστρώματος σε ιδανικές συνθήκες, όπως προτείνεται στη διδακτορική διατριβή του Buluswar (2002) [23].



Διαφορετικά κανάλια Μπλε (B), Πράσινο (G), Κόκκινο (R) του χρωματικού χώρου RGB που εμφανίζεται ξεχωριστά

ΠΗΓΗ: <https://www.learnopencv.com/color-spaces-in-opencv-cpp-python/>

Οι Andrey & Jo (2006) [24] για τον υπολογισμό των κατωφλίων έκαναν χρήση μιας μεταβλητής k , η οποία για κάθε εικονοψηφίδα ισούται με:

$$k = \frac{255}{\max(R, G, B)}$$

Στην προσέγγιση των Wali et al. (2015) [25], η κάθε εικόνα εισόδου διαχωρίζεται στα κανάλια R, G και B, και για κάθε κατώφλι του εκάστοτε καναλιού εφαρμόζονται φίλτρα προκειμένου να επιλεγθούν εκείνες οι περιοχές της εικόνας όπου οι τιμές των εικονοστοιχείων αντιστοιχούν στην περιοχή ενδιαφέροντος (ROI). Για παράδειγμα, για τα σήματα κυκλοφορίας με κόκκινο φόντο, το εύρος για το κανάλι R είναι τιμές από 90 – 255, ενώ για τα κανάλια G και B το εύρος είναι 0 – 70. Τελικά, η περιοχή ενδιαφέροντος εντοπίζεται από το λογικό άθροισμα (logical sum) των τριών φιλτραρισμένων καναλιών.

Διαφορετική προσέγγιση για τον εντοπισμό των κόκκινων πινακίδων ακολουθούν οι Yuille et al. (1998) [26], υλοποιώντας τη μέθοδο επέκτασης της περιοχής ενδιαφέροντος (region growing). Σε αυτή τη μέθοδο οι περιοχές ορίζονται ξεκινώντας από κάποια αρχικά σημεία (seeds) και κατόπιν ελέγχεται η ομοιότητα όλων των γειτονικών εικονοψηφίδων με το αρχικό σημείο, με κάποιο μέτρο ομοιότητας, όπως η διαφορά φωτεινότητας, το χρώμα, η υφή ή το σχήμα. Εάν το μέτρο ομοιότητας έχει τιμή εντός συγκεκριμένων ορίων, η αρχική περιοχή ενδιαφέροντος επεκτείνεται συμπεριλαμβάνοντας τις νέες εικονοψηφίδες που έχουν παρόμοιες ιδιότητες. Η επιλογή του κριτηρίου ομοιότητας εξαρτάται από το περιεχόμενο της εικόνας. Σε εικόνες με θόρυβο η ανάπτυξη περιοχών πλεονεκτεί σε σχέση με τις τεχνικές βασισμένες στις ακμές. Η επιτυχία της μεθόδου εξαρτάται άμεσα από την επιλογή των αρχικών εικονοψηφίδων, το μέτρο ομοιότητας και τα όρια. Ειδικότερα, στη συγκεκριμένη δημοσίευση οι συγγραφείς αρχικά εντοπίζουν όσες εικονοψηφίδες εμφανίζουν τιμές:

$$R > 128, \quad B > 0.8R, \quad G > 0.8R$$

και τις ορίζουν ως αρχικά σημεία (πυρήνες). Τα παραπάνω όρια, τα οποία υπολογίστηκαν κατόπιν δοκιμών, προκύπτουν επειδή στις κόκκινες πινακίδες η κόκκινη συνιστώσα εμφανίζει τιμή αρκετά υψηλή και μεγαλύτερη από τις άλλες δύο συνιστώσες. Στη συνέχεια, εξετάζονται

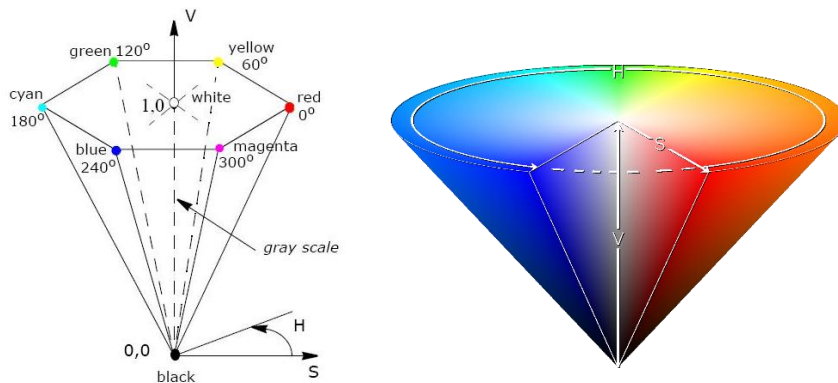
παράθυρα μεγέθους 15x15 εικονοψηφίδων γύρω από κάθε αρχικό σημείο για τον αριθμό των λευκών και κόκκινων εικονοψηφίδων που περιέχουν. Ανάλογα με τον αριθμό των λευκών και κόκκινων εικονοψηφίδων που υπάρχουν στο παράθυρο, αποφασίζεται εάν η περιοχή πρέπει να εξεταστεί περαιτέρω. Στον τελικό έλεγχο, σε κάθε ρίxel της εξεταζόμενης περιοχής αποδίδεται η τιμή 1, εάν περιλαμβάνεται στο διάστημα από $-2\sigma+m$ μέχρι $2\sigma+m$, όπου m η μέση τιμή του κόκκινου ή του λευκού χρώματος. Διαφορετικά του αποδίδεται η τιμή 0. Εάν το σύνολο των εικονοψηφίδων με τιμή διαφορετική του 0 ξεπερνά το 95%, το παράθυρο προσαρτάται στην περιοχή ενδιαφέροντος.

1.1.1.3 Οι Χρωματικοί χώροι HSV & HIS

Η πλειονότητα των ερευνητικών εργασιών για το στάδιο του εντοπισμού με βάση το χρώμα κάνει χρήση του χρωματικού μοντέλου HSV (Hue, Saturation, Value – Απόχρωση, Κορεσμός, Τόνος) ή του HIS (Hue, Saturation, Intensity – Απόχρωση, Κορεσμός, Ένταση).

Το χρωματικό μοντέλο HSV καταφέρνει να προσομοιώσει αρκετά τον τρόπο που οι άνθρωποι αντιλαμβάνονται το χρώμα και αναλύεται σε τρεις παραμέτρους. Η πρώτη αναφέρεται στη χροιά του χρώματος (Hue), η οποία γίνεται αντιληπτή λόγω του μήκους κύματος. Η δεύτερη συνιστώσα, ο κορεσμός (Saturation), αντιπροσωπεύει τον βαθμό καθαρότητάς του, δηλαδή το κατά πόσο το χρώμα έχει πρόσμιξη λευκού μέσα. Η τρίτη συνιστώσα αποτελεί την ένταση λαμπρότητας του χρώματος, κατά πόσο φωτεινό ή σκοτεινό είναι, δηλαδή αναφέρεται στο βαθμό μίξης ενός καθαρού χρώματος με το μαύρο.

Το σύνολο των τριών αυτών ιδιοτήτων μπορεί να παράγει οποιοδήποτε χρώμα βρίσκεται στη φύση. Επιτρέπει συνολικά 3.564.000 διαφορετικά χρώματα συν 101 διαβαθμίσεις του γκρι. Η τρισδιάστατη αναπαράσταση του HSV ως εξαγώνου προκύπτει από τον κύβο RGB, όπου κατά μήκος της διαγωνίου βρίσκονται οι τιμές του γκρι. Η απόχρωση δίνεται από τη γωνία με τον οριζόντιο άξονα, με το κόκκινο να είναι στις 0° , το κίτρινο στις 60° , το πράσινο στις 120° , το κυανό στις 180° , το μπλε στις 240° και το ματζέντα στις 300° . Να σημειωθεί πως τα συμπληρωματικά χρώματα έχουν διαφορά 180° . Η χρωματική καθαρότητα κυμαίνεται μεταξύ $0.0 \leq S \leq 1.0$ και είναι ο λόγος της καθαρότητας μιας συγκεκριμένης απόχρωσης προς τη μέγιστη καθαρότητα ($S=1$). Όταν $S=0$ ισχύει η κλίμακα του γκρι, δηλαδή η διαγώνιος του κύβου RGB. Η ένταση έχει και εκείνη πεδίο ορισμού μεταξύ 0 και 1.



Χρωματικός χώρος HSV

ΠΗΓΗ: https://commons.wikimedia.org/wiki/File:HSV_cone.png

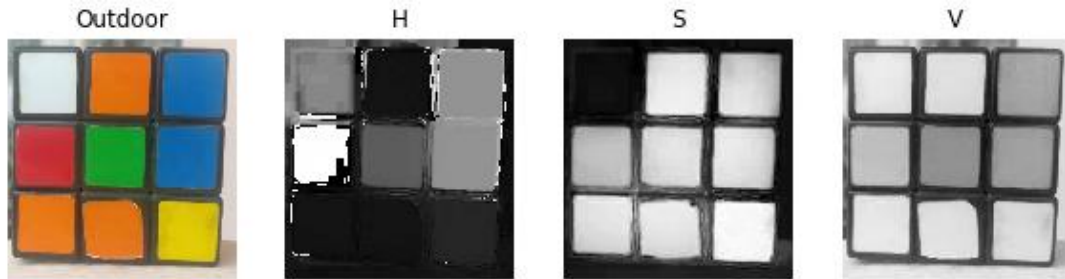
Ο μετασχηματισμός από RGB σε HSV είναι μη γραμμικός αλλά αντιστρεπτός. Οι σχέσεις που χρησιμοποιούνται για τον μετασχηματισμό από τον χρωματικό χώρο RGB σε HSV και αντίστροφα φαίνονται παρακάτω:

$$H = \begin{cases} \theta, & \text{εάν } B \leq G \\ 360 - \theta, & \text{εάν } B \geq G \end{cases}$$

$$S = 1 - \frac{3}{(R + G + B)} [\min(R, G, B)]$$

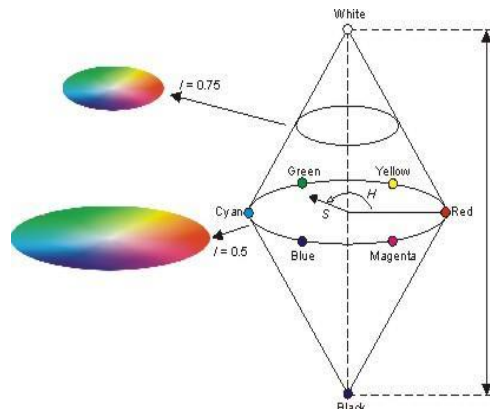
$$V = \frac{\max(R, G, B)}{255}$$

$$\text{όπου, } \theta = \cos^{-1} \left\{ \frac{\frac{1}{2}[(R-G)+(R-B)]}{[(R-G)^2+(R-G)(G-B)]^{1/2}} \right\}$$



Απόχρωση (H), Κορεσμός (S) και Τόνος (V) του χρωματικού χώρου HSV.
 ΠΗΓΗ: <https://www.learnopencv.com/color-spaces-in-opencv-cpp-python/>

Το χρωματικό μοντέλο HSI παρουσιάζει πολλές ομοιότητες με τον χρωματικό χώρο HSV. Μοναδική διαφορά των δύο χώρων είναι η τρίτη συνιστώσα και το πώς αυτή υπολογίζεται. Το μοντέλο HSI αποτελείται από τις παραμέτρους απόχρωση (Hue), κορεσμός (Saturation) και φωτεινότητα (Intensity). Οι παράμετροι του χρώματος στο μοντέλο HSI ορίζονται στην επιφάνεια του τριγώνου. Η απόχρωση H ενός χρώματος O είναι η γωνία που σχηματίζεται μεταξύ του άξονα του κόκκινου με το ευθύγραμμο τμήμα που ενώνει την αρχή (άσπρο) με το σημείο O. Όταν δηλαδή $H=0^\circ$, το χρώμα είναι κόκκινο, όταν $H=60^\circ$ το χρώμα είναι κίτρινο κ.λπ. Η χρωματική καθαρότητα S του χρώματος στο σημείο O είναι η απόσταση της ευθείας που ενώνει την αρχή με το σημείο O. Δείχνει δηλαδή το βαθμό στον οποίο το χρώμα αυτό είναι αναμιγμένο με το άσπρο. Όσο μεγαλύτερη είναι η απόσταση από την αρχή τόσο μεγαλύτερη είναι και η χρωματική καθαρότητα του χρώματος. Η ένταση στο μοντέλο HSI ορίζεται σε σχέση με τη θέση που κατέχει πάνω στη γραμμή που διαπερνά το κέντρο (άσπρο). Εντάσεις που τοποθετούνται κάτω από την επιφάνεια του τριγώνου τείνουν στο μαύρο όσο απομακρύνονται, ενώ τείνουν στο άσπρο εντάσεις που είναι πάνω από την επιφάνεια.



Χρωματικός χώρος HSI

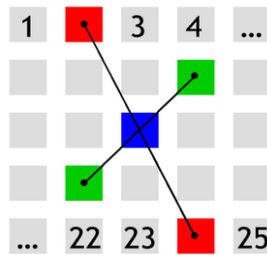
ΠΗΓΗ: <https://www.blackice.com/colorspaceHSI.htm>

Συνδυάζοντας κανείς την απόχρωση, την χρωματική καθαρότητα και την ένταση σε έναν τρισδιάστατο χώρο, καταλήγει σε μια δομή με μορφή κώνου. Οποιοδήποτε σημείο βρίσκεται πάνω στην εξωτερική επιφάνεια της δομής του μοντέλου, αντιστοιχεί σε χρώμα με πλήρη χρωματική καθαρότητα (εξαιρούνται το άσπρο όπου $S=0$ και το μαύρο όπου η παράμετρος S δεν ορίζεται). Η ένταση I εξαρτάται από τη θέση της κάθετης προβολής του σημείου πάνω στον κάθετο άξονα.

Οι χρωματικοί χώροι HSV και HIS παρουσιάζουν δυο βασικά πλεονεκτήματα. Πρώτον, οι τιμές V και I είναι ανεξάρτητες από το χρώμα και, δεύτερον, η απόχρωση H και η χρωματική καθαρότητα S είναι στενά συσχετισμένες με τον τρόπο αντίληψης του χρώματος από το ανθρώπινο μάτι. Αυτά τα χαρακτηριστικά καθιστούν τα δύο μοντέλα ιδανικά εργαλεία για την

ανάπτυξη αλγορίθμων επεξεργασίας εικόνας βασισμένων στην αίσθηση χρώματος από το ανθρώπινο οπτικό σύστημα, με ιδιαίτερα ικανοποιητικά αποτελέσματα κατάτμησης, αφού οι συνιστώσες H και S δεν επηρεάζονται από την ένταση του φωτός.

Παράδειγμα ερευνητικής εργασίας όπου χρησιμοποιήθηκε ο χρωματικός χώρος HSV στο στάδιο της κατάτμησης αποτελεί εκείνη των Vitabile et al. (2002) [27], όπου εφαρμόστηκε αρχικά ένα φίλτρο SNN (Symmetric Nearest Neighbor) για την απομάκρυνση του θορύβου. Το φίλτρο SNN συγκρίνει τις τιμές των συμμετρικών εικονοψηφίδων, επιλέγει την τιμή που βρίσκεται κοντινότερα την τιμή της κεντρικής εικονοψηφίδας και αποδίδει στην κεντρική εικονοψηφίδα τον μέσο όρο των τιμών φωτεινότητας των επιλεγμένων εικονοψηφίδων. Σημαντικό πλεονέκτημα του φίλτρου, αποτελεί το γεγονός ότι απομακρύνει τον θόρυβο, διατηρώντας όμως τις ακμές.



Φίλτρο SNN

ΠΗΓΗ: http://www.subsurfwiki.org/wiki/Symmetric_nearest_neighbour_filter

Στο επόμενο στάδιο γίνεται η μετατροπή στον χώρο HSV και επιλέγονται διάφοροι πυρήνες ανάλογα με τις τιμές H και S. Για κάθε οριζόμενη περιοχή γύρω από τον πυρήνα υπολογίζεται μια μέση τιμή και το κριτήριο ομοιότητας, που καθορίζει αν η εικονοψηφίδα θα προσαρτηθεί στον πυρήνα ή όχι. Το κριτήριο ομοιότητας είναι η ευκλείδεια απόσταση μεταξύ της μέσης τιμής και της τιμής της εικονοψηφίδας στον χώρο HSV. Η τιμή αυτή πρέπει να είναι εντός ενός κατωφλίου το οποίο είναι δυναμικό και υπολογίζεται κάθε φορά με βάση την τιμή του κορεσμού στην περιοχή του πυρήνα, όπως ορίζει η σχέση:

$$a = k - \sin(s_{seed})$$

όπου, a η παράμετρος κανονικοποίησης και s_{seed} η τιμή κορεσμού στον πυρήνα.

Στην εργασία των Wang et al. (2009) [28] επιλέχθηκαν μόνο τα εικονοστοιχεία με υψηλή τιμή για τη συνιστώσα V, καθώς μικρές τιμές της οδηγούν σε ασταθή αποτελέσματα. Τα καθολικά κατώφλια που χρησιμοποιήθηκαν για κάθε χρώμα είναι:

$H < 20$ ή $H > 300$ και $S > 0.3$ για κόκκινο χρώμα

$20 < H < 300$ και $S > 0.27$ για κίτρινο χρώμα

$180 < H < 70$ και $S > 0.25$ για μπλε χρώμα

Οι Chen et al. (2013) [29] χρησιμοποίησαν για τη μετατροπή από RGB σε HSV τις τιμές $max = \max(R, G, B)$ και $min = \min(R, G, B)$, ενώ αποδόθηκαν τιμές κατωφλίων και στις τρεις συνιστώσες και για τα τρία βασικά χρώματα, αφού πρώτα οι τιμές H, S, V ομαλοποιήθηκαν στο 255.

TABLE I.
THE H, S, V COMPONENT THRESHOLD RANGE OF THE MAIN COLOR OF TRAFFIC SIGN

Color	H	S	V
Red	$H \geq 240$ or $H \leq 10$	$S \geq 40$	$V \geq 30$
Yellow	$18 \leq H \leq 45$	$S \geq 148$	$V \geq 66$
Blue	$120 < H \leq 175$	$S \geq 127.5$	$V \geq 20$

Αντίστοιχη τακτική ακολουθήθηκε και από τους Rashmi et al. (2014) [30], όπου πραγματοποιήθηκε ανίχνευση και αναγνώριση των οδικών σημάτων κυκλοφορίας από εικόνες που πάρθηκαν με το όχημα εν κινήσει. Πρώτα εφαρμόστηκαν διάφορες τεχνικές φιλτραρίσματος για να μειωθεί η θόλωση λόγω κίνησης στις εικόνες και στη συνέχεια πραγματοποιήθηκε κατωφλίωση με τα εξής κατώφλια:

Color	H	S	V
Red	$H \geq 240$ or $H \leq 10$	$S \geq 40$	$V \geq 30$
Yellow	$18 \leq H \leq 45$	$S \geq 148$	$V \geq 66$
Blue	$120 < H \leq 175$	$S \geq 127.5$	$V \geq 20$

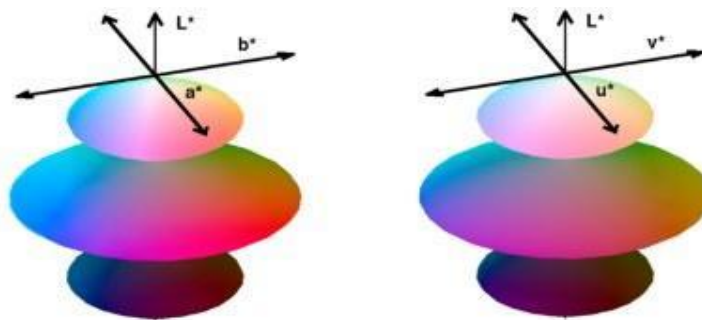
Στον χρωματικό χώρο του HSI οι Maldonado-Bascon et al. (2007) [31] μελέτησαν τα ιστογράμματα πινακίδων ειλημμένων υπό διαφορετικές καιρικές συνθήκες και συνθήκες φωτεινότητας καθώς και από διαφορετικές φωτογραφικές μηχανές, προκειμένου να υπολογίσουν σταθερά κατώφλια μόνο για τα μπλε και κόκκινα σήματα οδικής κυκλοφορίας. Τα κατώφλια αφορούν τις δύο πρώτες συνιστώσες του χώρου HSI, ενώ για το λευκό χρώμα, που δεν είναι δυνατόν να εντοπιστεί με τον συγκεκριμένο τρόπο, χρησιμοποιούν την ακόλουθη συνάρτηση.

$$f = \frac{(|R - G| + |G - B| + |B - R|)}{3D}$$

όπου D ο βαθμός εξαγωγής ενός ουδέτερου χρώματος (στη δεδομένη περίπτωση τέθηκε ίσος με 20). Εάν η παραπάνω συνάρτηση έχει τιμή μεγαλύτερη της μονάδας, καταδεικνύεται η ύπαρξη λευκού χρώματος.

1.1.1.4 Άλλοι χρωματικοί χώροι CIEL*a*b*, CIEL*u*v* & YCrCb

Ένα από τα επιθυμητά χαρακτηριστικά ενός χρωματικού συστήματος είναι η οπτική ευαισθησία για τις μικρές χρωματικά μεταβολές. Δηλαδή, μια μικρή χρωματικά αλλαγή πρέπει να γίνεται άμεσα αντιληπτή. Οι χρωματικοί χώροι που ικανοποιούν την ιδιότητα αυτή καλούνται ομοιόμορφοι χρωματικοί χώροι (perceptually uniform color spaces). Χρωματικοί χώροι όπως ο RGB απέχουν πολύ από το να ικανοποιούν την ιδιότητα αυτή. Για το λόγο αυτό η Διεθνής Ένωση Φωτεινότητας CIE (International Commission on Illumination) καθιέρωσε ως ομοιόμορφους τους χώρους CIELUV ($L^*u^*v^*$) και CIELAB ($L^*a^*b^*$). Τα δύο αυτά χρωματικά συστήματα έχουν μικρή διαφορά μεταξύ τους, ενώ παρουσιάζουν εξίσου καλή χρωματική ομοιομορφία και παρέχουν πολύ καλή δυνατότητα για τη μέτρηση της χρωματικής διαφοράς μεταξύ δύο σημείων. Βασίζονται στην αντιληπτή φωτεινότητα L^* , καθώς και σε δύο άλλες συνιστώσες χρώματος που καταγράφουν την πληροφορία χρώματος και καθορίζουν την ανάμειξη κόκκινου - πράσινου, κανάλια a^* και u^* , και κίτρινου - μπλε, κανάλια b^* και v^* , αντίστοιχα. Όπως αντίστοιχα και ο HSV, τα δύο αυτά χρωματικά μοντέλα προσομοιάζουν την ανθρώπινη όραση.



Χρωματικοί χώροι CIEL*a*b* & CIEL*u*v*

ΠΗΓΗ: https://ece.uwaterloo.ca/~nkvand/Coderep/colorspace_pascal/colorspace/colorspace.html

Οι CIEL*a*b* και CIEL*u*v* αποτελούν μη γραμμικούς, αλλά αντιστρεπτούς μετασχηματισμούς χρώματος. Δεν υπάρχουν όμως σχέσεις μετατροπής από τον RGB στον CIE Lab, καθώς ο πρώτος χώρος εξαρτάται από την συσκευή αναπαραγωγής (πχ. από την οθόνη που απεικονίζει τα χρώματα) ενώ ο δεύτερος όχι. Για τη μετατροπή του RGB στους δύο χώρους πρέπει ο πρώτος να μετασχηματιστεί σε ένα χρωματικά απόλυτο χώρο (δηλαδή ένα χώρο που δεν εξαρτάται από τη συσκευή αναπαραγωγής). Η πιο συνηθισμένη μετάβαση γίνεται από τον χρωματικό χώρο CIE XYZ με τον εξής τρόπο:

Η φωτεινότητα L^* κυμαίνεται μεταξύ 0 και 100 και ορίζεται απ' τη σχέση:

$$L^* = 116 \times f\left(\frac{Y}{Y_n}\right) - 16$$

όπου

$$f(r) = \begin{cases} r^{1/3} & \text{αν } r > 0.008856 \\ 7.7867 \times r + 16/116 & \text{αν } r \leq 0.008856 \end{cases}$$

και Y_n η φωτεινότητα του καθαρού άσπρου στο διάγραμμα CIE. Με γραμμικό μετασχηματισμό από τον χώρο RGB στον XYZ προκύπτει:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.412453 & 0.35758 & 0.180423 \\ 0.212671 & 0.71516 & 0.072169 \\ 0.019334 & 0.11919 & 0.950227 \end{bmatrix} \times \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

Οι τιμές των X_n , Y_n και Z_n αντιστοιχούν στις συντεταγμένες του λεγόμενου λευκού σημείου αναφοράς (white point reference), όπου στο πρότυπο D65 (CIE Standard Illuminants) έχουν τις τιμές: $X_n = 0.95$, $Y_n = 1$ και $Z_n = 1.089$.

Οι χρωματικές συνιστώσες u^* και v^* προσδιορίζονται από τις σχέσεις

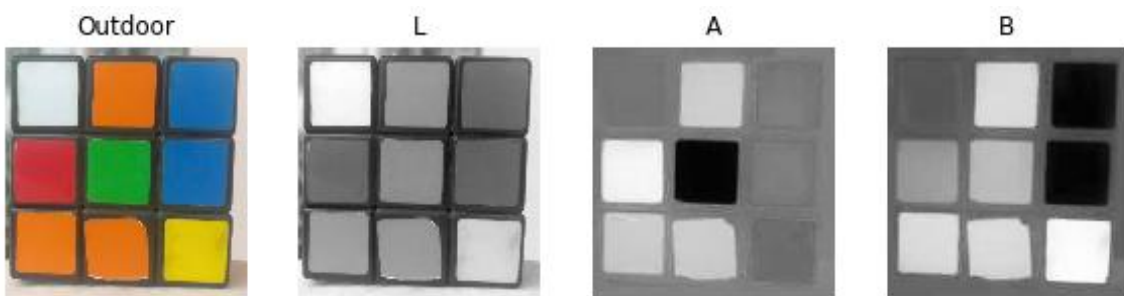
$$u^* = 13L \times \left(\frac{4X}{X+15Y+3Z} - \frac{4X_n}{X_n+15Y_n+3Z_n} \right)$$

$$v^* = 13L \times \left(\frac{9Y}{X+15Y+3Z} - \frac{9Y_n}{X_n+15Y_n+3Z_n} \right)$$

Αντίστοιχα, για το σύστημα $L^*a^*b^*$ (το οποίο είναι και πολύ περισσότερο καθιερωμένο σε σχέση με το $L^*u^*v^*$) ισχύει:

$$a^* = 500 \times \left[f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right) \right]$$

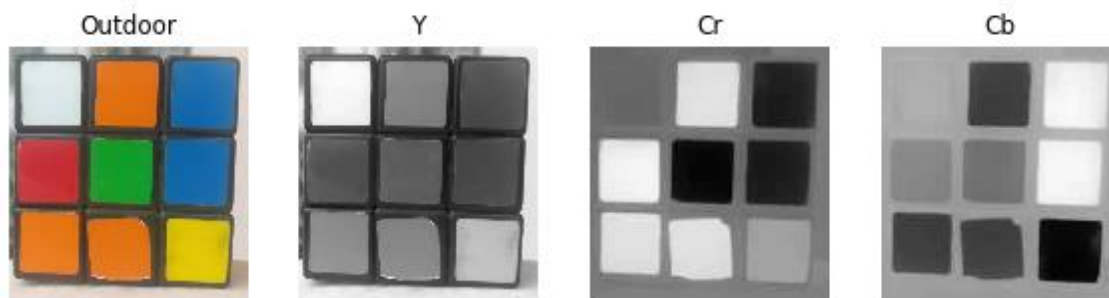
$$b^* = 200 \times \left[f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right) \right]$$



Η Φωτεινότητα (L) και τα χρωματικά στοιχεία (A, B) στο LAB χρωματικό μοντέλο
 ΠΗΓΗ: <https://www.learnopencv.com/color-spaces-in-opencv-cpp-python/>

Σε αυτό τον λιγότερο συνηθισμένο χρωματικό χώρο L^*a^*b μετασχηματίζουν οι Siogas & Dermatas (2006) [32].

Για τον εντοπισμό των πινακίδων με κόκκινο χρώμα χρησιμοποιείται κυρίως ο χώρος YCrCb (Eichner & Breckon., 2008 [33]). Ο YCrCb προέρχεται από τον χρωματικό χώρο RGB και αποτελείται από τρεις συνιστώσες, το κανάλι Y (Luminance), που λαμβάνεται από τον RGB μετά από τη χρωματική διόρθωση γάμμα (gamma correction) και καταγράφει τη φωτεινότητα, και από τα χρωματικά κανάλια Cr και Cb που αναφέρονται στην απόσταση που έχει το κόκκινο και το μπλε κανάλι του χρωματικού χώρου RGB από το κανάλι Y, αντίστοιχα. Ο χρωματικός αυτός χώρος χρησιμοποιείται στο ψηφιακό βίντεο και στα πρότυπα JPEG και MPEG.



Κανάλια του χρωματικού χώρου YCrCb

ΠΗΓΗ: <https://www.learnopencv.com/color-spaces-in-opencv-cpp-python/>

Τέλος, πέρα από τους χρωματικούς χώρους που προαναφέρθηκαν υπάρχουν και άλλοι χώροι, όπως ο CMYK, μοντέλο χρωμάτων που χρησιμοποιείται στην τυπογραφία. Η ονομασία του προέρχεται από τις λέξεις Cyan (κυανό), Magenta (ματζέντα), Yellow (κίτρινο), και Black (μαύρο), τα οποία είναι και τα βασικά χρώματα μέσω των οποίων δημιουργούνται τα υπόλοιπα χρώματα και αποχρώσεις. Το μαύρο αναφέρεται με το γράμμα K (Key), που σημαίνει χρώμα κλειδί στην παραγωγή της τετραχρωμίας. Άλλος χρωματικός χώρος είναι ο YUV, ο οποίος χρησιμοποιείται για τη μετάδοση του σήματος τηλεόρασης στο σύστημα Pal, ενώ ο YIQ (Luminance, Inphase, Quadrature) στο σύστημα NTSC αποτελεί πρότυπο για την τηλεοπτική μετάδοση. Ο HSL (Hue, Saturation, Lightness) είναι αντίστοιχος του HSV, αλλά με διαφορετικό ορισμό του κορεσμού και της έντασης. Εδώ χρησιμοποιείται η φωτεινότητα (Luminosity), αντί της λαμπρότητας και οι τιμές αναφοράς για μαύρο και γκρι είναι διαφορετικές. Το L υπολογίζεται από τον εξής τύπο:

$$L = \frac{\min(R, G, B) + \max(R, G, B)}{2}$$

Ο Zakir (2011) [34] στη διδακτορική του διατριβή παρουσιάζει μια συγκριτική μελέτη ως προς τα αποτελέσματα κατωφλίωσης στους χώρους RGB, HSV, L^*a^*b , YCbCr, CYMK, YIQ για την κατάτμηση των εικόνων και παρέχει περισσότερες πληροφορίες σχετικά με τον εντοπισμό στους διάφορους χρωματικούς χώρους. Διεξήγαγε λεπτομερή έρευνα σχετικά με την καταλληλότητα χρήσης των έξι αυτών διαφορετικών χρωματικών χώρων για την ανίχνευση σημάτων οδικής κυκλοφορίας πάνω σε τρία βασικά χρώματα που χρησιμοποιούνται ευρέως (κόκκινο, μπλε και πράσινο). Οι εικόνες τραβήχτηκαν σε τέσσερις διαφορετικές συνθήκες φωτισμού, φωτεινή / ημέρα, με βροχή, απόγευμα / σούρουπο και νύχτα. Ύστερα από την προεπεξεργασία για τη ρύθμιση του επιπέδου φωτεινότητας των εικόνων (gamma correction) και τη μετατροπή της κάθε εικόνας στον επιθυμητό χρωματικό χώρο, τίθενται κατώφλια σύμφωνα με τους διάφορους πειραματικούς περιορισμούς.

Τα πειράματά αποκάλυψαν ότι στο *CIElab*, οι χρωματικές συνιστώσες l , και b παραμένουν σε μεγάλο βαθμό σταθερές, ενώ παρατηρούνται μικρές διακυμάνσεις στην a . Στους χρωματικούς χώρους *RGB* και *YCbCr* παρατηρήθηκε ότι όλες οι παράμετροι ήταν πολύ ευαίσθητες στις αλλαγές φωτισμού καθώς και στην παρουσία περιοχών με πράσινο χρώμα, όπως τα δέντρα και το γρασίδι. Στο χώρο *YIQ*, το Q στοιχείο παραμένει σταθερό ενώ οι συνιστώσες I και Y έχουν μικρές διακυμάνσεις.

Ωστόσο, ο χρωματικός χώρος HSV αποδείχθηκε πιο σταθερός, καθώς όλες οι συνιστώσες του παραμένουν αρκετά ομοιόμορφες υπό διαφορετικές συνθήκες φωτισμού. Ο λόγος για την παραπάνω συμπεριφορά βρέθηκε να είναι η παράμετρος H, οποία δεν μπορεί να συλλάβει καμία απόχρωση του πράσινου, που συνήθως απαντάται σε περιοχές που καλύπτονται από βλάστηση. Κάτω από το φως της ημέρας, το γρασίδι και τα δέντρα παρουσιάζουν μια ελαφριά πράσινη σκιά λόγω της αντανάκλασης του ηλιακού φωτός, ενώ οι πράσινες πινακίδες κυκλοφορίας έχουν μία χαρακτηριστική πιο σκούρα πράσινη απόχρωση. Οι συνιστώσες S και V όμως έχουν την ιδιότητα να μπορούν να διαφέρουν όταν αναζητείται περιοχή με πράσινο χρώμα χαμηλής έντασης.

Η μελέτη αυτή συμπέρανε πως ο χρωματικός χώρος HSV για τον παραπάνω λόγο αποτελεί το πιο εύρωστο χρωματικό μοντέλο που μπορεί να χρησιμοποιηθεί αποτελεσματικά για την κατάτμηση εικόνων που περιέχουν σήματα οδικής κυκλοφορίας με βάση το χρώμα υπό διαφορετικό φωτισμό και περιβαλλοντικές συνθήκες.

Παρόλο που οι αλλαγές στις συνθήκες φωτισμού επηρεάζουν τις πληροφορίες χρώματος, το χρώμα παραμένει μια χρήσιμη ένδειξη για την ανίχνευση και αναγνώριση των σημάτων οδικής κυκλοφορίας. Επιπλέον, τα σήματα του Κ.Ο.Κ. μπορεί να έχουν πολύ παρόμοιες εμφανίσεις και το χρώμα μπορεί να είναι ένα πολύ σημαντικό χαρακτηριστικό για τη διάκριση μεταξύ τους. Όταν υπάρχει ένας μεγάλος αριθμός κλάσεων στη βάση δεδομένων σημάτων κυκλοφορίας, το χρώμα φέρει πολύτιμες διακριτικές πληροφορίες που πρέπει να χρησιμοποιούνται, όποτε αυτό είναι δυνατόν. Τέλος, δεν πρέπει να παραλειφθεί το γεγονός ότι σε ορισμένες χώρες, όπως στην Ιαπωνία, υπάρχουν ζεύγη πινακίδων τα οποία διαφέρουν μόνο ως προς το χρώμα. Επομένως, η ανάπτυξη ισχυρών χρωματικών αλγορίθμων, λαμβάνοντας υπόψη την πληροφορία του χρώματος, παρουσιάζει μεγάλο ενδιαφέρον και σημαντικότητα για τα τελικά αυτόματα συστήματα υποβοήθησης της οδήγησης.

1.1.2 Εντοπισμός με βάση το σχήμα

Οι πρόσφατες εξελίξεις στην αναγνώριση αντικειμένων ανοίγουν μεγάλες ευκαιρίες για αξιόπιστη ανίχνευση σημάτων κυκλοφοριακής σήμανσης σε μη ελεγχόμενα περιβάλλοντα. Παρόλα αυτά, εξακολουθούν να υπάρχουν προκλήσεις, όπως η ανίχνευση σημάτων κυκλοφορίας σε "γεμάτες" σκηνές (cluttered scenes), ποικίλες συνθήκες φωτεινότητας και φωτισμού, αφινικές και προοπτικές παραμορφώσεις ανάλογα με την οπτική γωνία της κάμερας είτε και λόγω στροφών της ίδιας της πινακίδας, αλλαγές της κλίμακας, μερικές αποκρύψεις κ.λ.π..

Προκειμένου να επιτευχθεί ο εντοπισμός του αντικειμένου με επιτυχία, εκτός από τις χρωματικές ιδιότητες στις οποίες βασίζονται οι μέθοδοι εντοπισμού, έχουν επίσης αναπτυχθεί μέθοδοι για τον εντοπισμό με βάση το σχήμα. Αυτές οι μέθοδοι μπορούν να χρησιμοποιηθούν είτε αυτόνομα είτε σε συνδυασμό με τις μεθόδους εντοπισμού με βάση το χρώμα για ακριβέστερα αποτελέσματα.

Επί της ουσίας οι πινακίδες αποτελούνται από τα βασικά δισδιάστατα γεωμετρικά σχήματα, κύκλος, τρίγωνο, ορθογώνιο και οκτάγωνο. Η αυστηρή και συμμετρική γεωμετρία τους επιτρέπει την εξαγωγή χαρακτηριστικών σημείων για τον εντοπισμό τους.



Βασικά σχήματα σημάτων
ΠΗΓΗ: Zakir (2011) [34]

Σε ορισμένες περιπτώσεις, μάλιστα, οι μέθοδοι αυτές υπερτερούν συγκριτικά με τις μεθόδους με βάση το χρώμα, αφού δεν εξαρτώνται από τις συνθήκες φωτεινότητας της λήψης και τις καιρικές συνθήκες. Η απλότητα του σχήματος των πινακίδων, όμως, προσδίδει μια επιπλέον δυσκολία, αφού μπορεί εύκολα να συγχέονται με άλλα αντικείμενα ή μέρη αντικειμένων που

έχουν παρόμοια γεωμετρία. Επίσης, πολλές φορές οι πινακίδες πέραν των αποκρύψεων μπορεί να έχουν αλλοιωμένο σχήμα λόγω καταστροφών, κατάσταση η οποία δυσχεραίνει την ανίχνευση του αντικειμένου στην εικόνα. Αν και το υπολογιστικό κόστος σε σχέση με τον εντοπισμό με βάση το χρώμα είναι μεγαλύτερο, η ραγδαία πλέον ανάπτυξη της υπολογιστικής ισχύος δεν καθιστά το γεγονός αυτό σημαντικό μειονέκτημα.

Οι αλγόριθμοι που πλέον αναπτύσσονται και εφαρμόζονται προσπαθούν σε κάθε περίπτωση να αντιμετωπίσουν τα προβλήματα αυτά με τον καλύτερο δυνατό τρόπο.

1.1.2.1 Μετασχηματισμός Hough

Αρκετές προσεγγίσεις για την ανίχνευση σημάτων οδικής κυκλοφορίας με βάση το σχήμα εμφανίζονται συχνά στη βιβλιογραφία. Πιθανώς η πιο ευρέως διαδεδομένη προσέγγιση που χρησιμοποιείται είναι ο μετασχηματισμός Hough (Hough transformation), ο οποίος μπορεί να έχει επιτυχή αποτελέσματα ακόμα και σε επικαλυπτόμενα αντικείμενα. Ο μετασχηματισμός αυτός χρησιμοποιείται για τον εντοπισμό κυρίως γραμμών στις εικόνες, αλλά μπορεί να τροποποιηθεί με εύκολο τρόπο ώστε μέσω αυτού να εντοπίζονται και άλλα σχήματα, όπως οι κύκλοι.

Ο μετασχηματισμός Hough εισήχθη το 1962 (Hough, 1962 [35]) και αναπτύχθηκε στα πλαίσια της κατασκευής μιας ηλεκτρονικής συσκευής για την ανίχνευση της κίνησης σωματιδίων υψηλής ενέργειας. Ήταν μία από τις πρώτες προσπάθειες για αυτοματοποίηση της οπτικής διαδικασίας που μέχρι τότε απαιτούσε πολλές ανθρωποώρες εργαστηριακών μετρήσεων. Χρησιμοποιήθηκε για πρώτη φορά για ανίχνευση ακμών μία δεκαετία αργότερα από τους Duda & Hart (1972) [36] (Hough lines). Στη συνέχεια εξελίχθηκε για την ανίχνευση κύκλων ή ελλείψεων (κυκλικός μετασχηματισμός Hough – Circular Hough Transformation) και τελικά επεκτάθηκε πλήρως ώστε να εντοπίζει σχήματα τα οποία δεν παραμετροποιούνται μαθηματικά με την μορφή του γενικευμένου μετασχηματισμού Hough Generalised Hough Transformation), ο οποίος παρουσιάζει ομοιότητες με την τεχνική συνταύτισης προτύπων (Template matching).

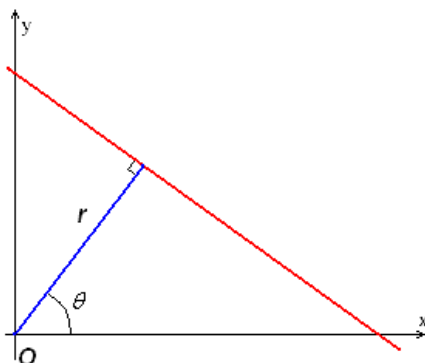
Ο μετασχηματισμός εφαρμόζεται σε μια εικόνα κλίσεων (η οποία υπόκειται και σε περαιτέρω επεξεργασία, συνήθως κατωφλίωση), αφού πρώτα εντοπιστούν οι ακμές που αποτελούν τα σύνορα μεταξύ δύο ομοιογενών περιοχών με διαφορετική φωτεινότητα. Ο πιο συχνός τελεστής εντοπισμού ακμών είναι ο Canny edge detector. Η κλίση της εικόνας (gradient) υποδηλώνει την αλλαγή της φωτεινότητας μεταξύ γειτονικών εικονοστοιχείων. Η εικόνα των κλίσεων μπορεί να προκύψει με εφαρμογή διάφορων τελεστών (μάσκες ακμών), όπως Roberts, Prewitt και Sobel.

Το βασικό χαρακτηριστικό του προσδιορισμού ευθειών με τον μετασχηματισμό Hough είναι η σχέση σημείου-ευθείας. Η λογική λοιπόν του μετασχηματισμού είναι να μετατραπούν όλα τα σημεία της εικόνας στο πεδίο (m, c) των παραμέτρων μιας ευθείας και μετά τα σημεία με την μεγαλύτερη συγκέντρωση στο πεδίο του μετασχηματισμού θα αντιστοιχούν στις «εντονότερες» ευθείες της αρχικής εικόνας. Έτσι αρκεί να βρεθούν τα τοπικά μέγιστα στο χώρο (m, c) και με αυτό τον τρόπο εξάγονται οι κυρίαρχες ευθείες στην εικόνα.

Όμως επειδή ο μετασχηματισμός Hough μετασχηματίζει τα εικονοστοιχεία σε ευθείες είναι αδύνατο να εντοπίσει κατακόρυφες ευθείες μιας και αυτές θα είχαν άπειρη κλίση. Δηλαδή, στην ευθεία $y = mx + c$, το m που αντιπροσωπεύει την κλίση θα έπρεπε να είναι άπειρο. Όποτε πρέπει να βρεθεί ένας πιο κατάλληλος μετασχηματισμός και αυτός γίνεται εάν εισαχθεί η ελάχιστη απόσταση r της ευθείας από την αρχή των αξόνων, και η γωνία που σχηματίζεται ανάμεσα στο r και τον άξονα x .

Ο προσδιορισμός των συννευθειακών σημείων του επιπέδου $[x - y]$ πραγματοποιείται με ανίχνευση των σημείων του επιπέδου $[m - c]$ από τα οποία διέρχονται περισσότερες από μια ευθείες. Στη σχέση $y = mx + c$, τα όρια των τιμών (m, c) είναι μη πεπερασμένα. Έτσι, οι ευθείες του επιπέδου $[x - y]$ που είναι παράλληλες στον άξονα y απαιτούν άπειρα ζεύγη τιμών (m, c) ώστε να προσδιοριστούν στο επίπεδο $[m - c]$. Το πρόβλημα αυτό λύθηκε με την προσέγγιση των Duda & Hart [36] οι οποίοι αντικατέστησαν το επίπεδο $[m - c]$ με το επίπεδο $[\theta - \rho]$ των

πολικών παραμέτρων μιας ευθείας. Στην περίπτωση αυτή η σχέση γίνεται (Κεσιδης, 2001 [37]): $p = x \cos \theta + y \sin \theta$



$$r = x \cos \theta + y \sin \theta$$

ΠΗΓΗ: https://en.wikipedia.org/wiki/Hough_transform

Μια σημαντική ιδιότητα του μετασχηματισμού Hough είναι η ανοχή του στο θόρυβο, αλλά και στην περίπτωση απόκρυψης του αντικειμένου και κατ'επέκταση της ευθείας. Αυτό γίνεται λόγω της ανθεκτικότητας του μετασχηματισμού από το χώρο ακμών της εικόνας στο χώρο συσσώρευσης (accumulator space), ένα κομμάτι της ευθείας που χάθηκε θα προκαλέσει απλώς ένα μικρότερο τοπικό μέγιστο λόγω της μικρότερης συσσώρευσης των σημείων της ευθείας στο αντίστοιχο σημείο του χώρου $[m - c]$.

Οι Garcia Garrido et al. (2005) [38] εφάρμοσαν τον κλασικό Μετασχηματισμό Hough για τον εντοπισμό τριγωνικών πινακίδων οδικής κυκλοφορίας στην εικόνα ακμών, η οποία προέκυψε από τη χρήση του τελεστή εντοπισμού ακμών Canny Edge Detector.

Ο ανιχνευτής ακμών Canny αναπτύχθηκε από τον John F. Canny το 1986 και χρησιμοποιεί έναν αλγόριθμο πολλαπλών σταδίων για να ανιχνεύσει ένα ευρύ φάσμα ακμών στις εικόνες. Ο αλγόριθμος διαδικασίας του τελεστή εντοπισμού ακμών Canny Edge Detector μπορεί να αναλυθεί σε 5 διαφορετικά βήματα:

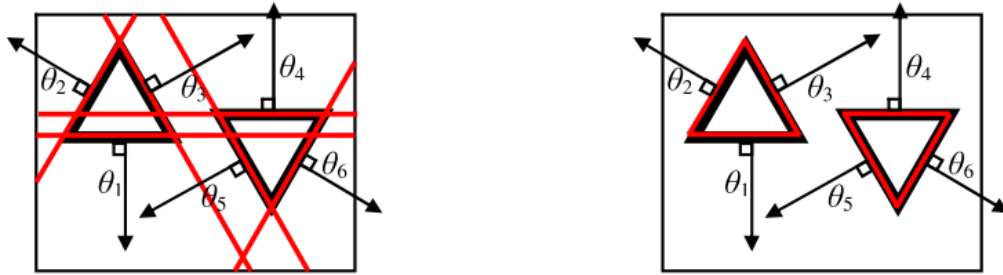
1. Εφαρμογή φίλτρου Gauss για εξομάλυνση της εικόνας και αφαίρεση θορύβου
2. Εντοπισμός των κλίσεων έντασης της εικόνας
3. Εφαρμογή φίλτρου συστολής για την απαλλαγή από ψευδή εντοπισμένα στοιχεία
4. Εφαρμογή διπλού κατωφλίου για να τον προσδιορισμό πιθανών ακμών
5. Εντοπισμός ακμών με υστέρηση (hysteresis): Ολοκλήρωση την ανίχνευσης ακμών με αποκλεισμό όλων των άλλων ακμών, οι οποίες είναι αδύναμες και δεν συνδέονται με ισχυρές ακμές.

Εφόσον ο αριθμός των ευθειών που τέμνονται μεταξύ τους μπορεί να είναι πολύ μεγάλος εάν εφαρμοστεί ο μετασχηματισμός Hough σε ολόκληρη την εικόνα, θα μπορούσαν να ανιχνευθούν περισσότερα από τα πραγματικά τρίγωνα που υπάρχουν στην εικόνα.

Με τη χρήση του μετασχηματισμού Hough ούτε η αρχή ούτε το τέλος μιας ευθείας γραμμής είναι γνωστά, καθώς μια ευθεία γραμμή που ορίζεται από αυτόν τον μετασχηματισμό εκφράζεται σε όρους των πολικών παραμέτρων θ και ρ :

$$x \cos(\theta) + y \sin(\theta) = \rho$$

Προκειμένου να ξεπεραστεί αυτό το πρόβλημα εφαρμόστηκε ο μετασχηματισμός σε κάθε ένα περίγραμμα διαδοχικά. Με αυτόν τον τρόπο, μόνο εκείνα τα τρίγωνα που υπάρχουν στην εικόνα ανιχνεύονται, όπως φαίνεται στο παρακάτω σχήμα, μειώνοντας επίσης τον υπολογιστικό χρόνο.



Ανίχνευση ακμών με μετασχηματισμό Hough, σε ολόκληρη την εικόνα (αριστερά), σε κάθε περίγραμμα, ένα προς ένα (δεξιά) ΠΗΓΗ: Garcia Garrido et al. (2005) [38]

Από τις εντοπισμένες ακμές γίνονται δεκτές μόνο όσες είναι σχεδόν κλειστές και αν εξαχθούν 3 ευθείες με γωνίες τομής 60°, τότε έχει εντοπιστεί μία τριγωνική πινακίδα, όπου ο εντοπισμός αφορά είτε το εσωτερικό είτε το εξωτερικό τρίγωνο.

Αφού εφαρμοστεί ο Hough για κάθε κλειστή ακμή, εφαρμόζεται και ο κυκλικός μετασχηματισμός για την ανίχνευση των κυκλικών σημάτων οδικής κυκλοφορίας, καθώς και του σήματος στοπ. Ένας κύκλος στο επίπεδο xy με κέντρο (χ, ψ) και ακτίνα ρ μπορεί να εκφραστεί ως:

$$f(x, p) = (x - \chi)^2 + (y - \psi)^2 - \rho^2 = 0$$

όπου ο χώρος παραμέτρων, $p = (\chi, \psi, \rho)$, πρέπει να ποσοτικοποιηθεί. Ο πίνακας συσσώρευσης «a» είναι η αναπαράσταση του κβαντισμένου χώρου των παραμέτρων. Για την ανίχνευση κύκλων, πίνακας Hough «a» θα είναι ένας τρισδιάστατος πίνακας με αρχικά όλες τις καταχωρήσεις να ισούνται με 0. Η καταχώριση $a(\chi_r, \psi_s, \rho_t)$ αυξάνεται κατά 1 για κάθε χαρακτηριστικό σημείο (x_i, y_i) στο πεδίο εικόνας, που περιέχεται στην περιφέρεια με κέντρο (χ_r, ψ_s) και ακτίνα ρ_t όπως εκφράζεται παρακάτω, όπου εισάγεται ένα περιθώριο ακρίβειας ϵ για την ακτίνα ρ_t για την αντιστάθμιση του σφάλματος κβαντισμού κατά την ψηφιοποίηση της εικόνας:

$$|(\chi_r - x_i)^2 + (\psi_s - y_i)^2 - \rho_t^2| < \epsilon$$

Για την ανίχνευση κυκλικών αντικειμένων ακολουθούνται τα ίδια κριτήρια όπως στην περίπτωση των ευθειών. Ο μετασχηματισμός εφαρμόζεται επίσης σε κάθε περίγραμμα, ένα προς ένα.

Σημαντικό χαρακτηριστικό ενός κύκλου αποτελεί το ότι το κέντρο και το κεντροειδές του είναι στην πραγματικότητα το ίδιο σημείο. Χρησιμοποιώντας αυτήν την ιδιότητα το κέντρο αναζητείται σε ένα πεδίο κοντά στο κέντρο. Όλες αυτές οι εκτιμήσεις καθιστούν το χρόνο ανίχνευσης πολύ μικρό και το σύστημα ικανό να λειτουργεί με ταχύτητα επεξεργασίας μεταξύ 5 και 50 καρέ ανά δευτερόλεπτο, ανάλογα με τον αριθμό των σημείων που εντοπίστηκαν.



Ανίχνευση κύκλων με μετασχηματισμό Hough, σε ολόκληρη την εικόνα (αριστερά), σε κάθε περίγραμμα, ένα προς ένα (δεξιά) ΠΗΓΗ: Garcia Garrido et al. (2005) [38]

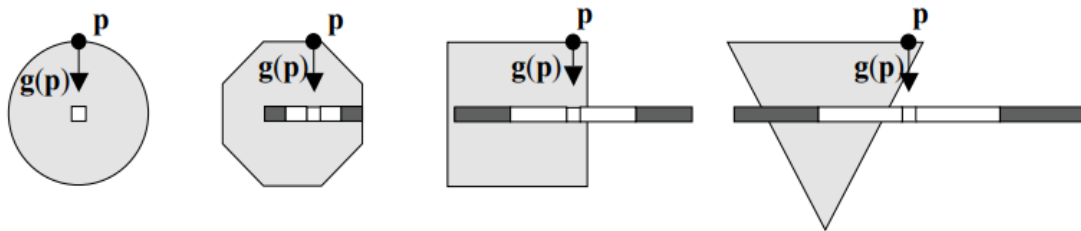
Για να επιταχυνθεί η διαδικασία, οι Piccioli et al. (1996) [39] χρησιμοποίησαν πληροφορία χρώματος για να περιοριστεί η περιοχή ενδιαφέροντος, ακολουθούμενη από γεωμετρική ανάλυση στον χάρτη ακμών για την εξαγωγή και ανίχνευση κυκλικών και τριγωνικών σχημάτων. Μετά την εξαγωγή ευθειών επίσης με τον τελεστή Canny, αυτές οι οποίες έχουν

κατάλληλο μήκος και κλίση μπορεί να είναι κατάλληλες ώστε να σημανθούν ως πινακίδες οδικής κυκλοφορίας.

1.1.2.2 Γρήγορος Μετασχηματισμό Ακτινικής Συμμετρίας

Ο γενικευμένος μετασχηματισμός Hough αποτελεί τεχνική καταγραφής αυθαίρετων σχημάτων σε μια εικόνα. Ακολουθεί το σχήμα Single Pixel Voting (SPV), όπου η βασική ιδέα είναι ότι σε μια εικόνα ακμών κάθε εικονοστοιχείο της εικόνας ψηφίζει για το πού θα ήταν το κέντρο των αντικειμένων εάν το εικονοστοιχείο ήταν στο όριο του αντικειμένου (boundary). Οι εργασίες των Loy και Barnes προορίζονταν ειδικά για ανίχνευση σημάτων οδικής κυκλοφορίας και χρησιμοποιήθηκαν ανεξάρτητα σε διάφορα συστήματα ανίχνευσης. Οι Loy & Barnes (2004) [40] προτείνουν ένα κανονικό πολύγωνο ως ανιχνευτή και το χρησιμοποιούν για την ανίχνευση των πινακίδων κυκλοφορίας. Ο ανιχνευτής βασίζεται στον Γρήγορο Μετασχηματισμό Ακτινικής Συμμετρίας (Fast Radial Symmetry Transform) και η συνολική προσέγγιση είναι παρόμοια με τον μετασχηματισμό Hough. Πρώτα, δημιουργείται από την αρχική εικόνα η εικόνα κλίσεων, η οποία στη συνέχεια κατωφλιώνεται ώστε να εξαλειφθούν τα σημεία με χαμηλά μεγέθη, τα οποία είναι απίθανο να αντιστοιχούν στις ακμές. Κάθε εναπομείναν εικονοστοιχείο ψηφίζει τότε για τις πιθανές θέσεις του κέντρου ενός κανονικού πολυγώνου.

Ένα εικονοστοιχείο εκπέμπει την ψήφο του σε πολλαπλές θέσεις που κατανέμονται κατά μήκος μιας γραμμής κάθετης προς την κλίση του εικονοστοιχείου και της οποίας η απόσταση από το εικονοστοιχείο είναι ίση με την αναμενόμενη ακτίνα του κανονικού πολυγώνου. Στην πραγματικότητα, όπως φαίνεται και στο παρακάτω σχήμα, υπάρχουν δύο γραμμές που ικανοποιούν αυτές τις απαιτήσεις, η μία στην κατεύθυνση της κλίσης και η άλλη στην αντίθετη κατεύθυνση. Και οι δύο μπορούν να χρησιμοποιηθούν εάν δεν είναι γνωστό εκ των προτέρων εάν οι πινακίδες θα είναι φωτεινότερες ή πιο σκούρες από το φόντο.



Περιοχές που σχετίζονται με τις αρνητικές και θετικές ψήφους ενός εικονοστοιχείου p κατά την αναζήτηση σε διαφορετικά σχήματα.
ΠΗΓΗ: Loy & Barnes (2004) [38]

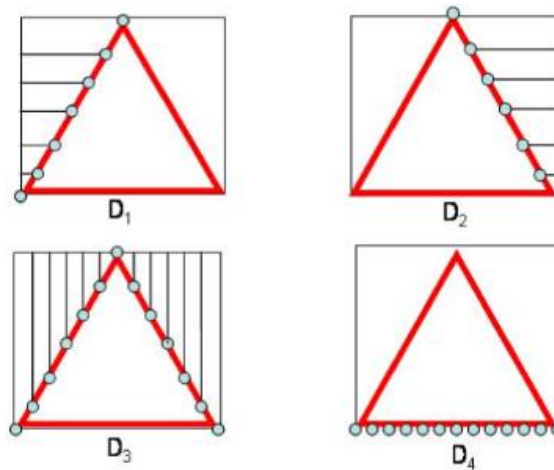
Το μήκος της γραμμής ψηφοφορίας περιορίζεται από την ακτίνα του πολυγώνου. Οι ψήφοι προς τα άκρα της γραμμής έχουν αρνητικά βάρη, προκειμένου να ελαχιστοποιηθεί η επίδραση ευθειών που είναι πολύ μεγάλες για να αποτελούν πλευρές πολυγώνου. Εκτός από την εικόνα ψηφοφορίας που προκύπτει, δημιουργείται μια ακόμα εικόνα που ονομάζεται ισοδύναμη εικόνα. Η προτεινόμενη διαδικασία ευνοεί τα ισογώνια πολύγωνα χρησιμοποιώντας την ακόλουθη ιδιότητα: εάν οι γωνίες κλίσης των εικονοστοιχείων ακμής ενός κανονικού πολυγώνου πολλαπλής όψης πολλαπλασιάζονται με n , οι γωνίες που προκύπτουν θα είναι ίσες. Η κύρια αδυναμία της προσέγγισης είναι ότι πρέπει να είναι γνωστή η ακτίνα των πολυγώνων που αναζητούνται, κάτι που δεν είναι πάντα εύκολο να επιτευχθεί. Αυτό μπορεί να επιλυθεί με δοκιμή πολλών ακτίνων, αλλά μπορεί να είναι πολύ ακριβό από την άποψη του χρόνου επεξεργασίας.

1.1.2.3 Απόσταση από Όρια

Μια άλλη πολύ γνωστή μέθοδος για ανίχνευση αντικειμένων με βάση το σχήμα είναι η τεχνική της Απόστασης από Όρια (Distance to Borders), η οποία μάλιστα παρουσιάζει ανοχή σε πιθανές μεταθέσεις των αντικειμένων. Ουσιαστικά με τη μέθοδο αυτή δημιουργείται ένα εφαπτόμενο τετράπλευρο (bounding box) γύρω από την περιοχή ενδιαφέροντος. Η απόσταση

του εφαπτόμενου τετραπλεύρου από την εξωτερική ακμή της περιοχής ενδιαφέροντος μπορεί να αποδώσει το σχήμα της περιοχής. Η μέθοδος αυτή μπορεί να αποκτήσει ανοχή και σε ενδεχόμενες στροφές, στην περίπτωση που το εφαπτόμενο τετράπλευρο ταυτιστεί με τον οριζόντιο άξονα, καθώς και ανοχή ως προς την κλίμακα στις περιπτώσεις που ακολουθήσει κανονικοποίηση των αποστάσεων με τέτοιον τρόπο ώστε να λαμβάνονται υπόψη οι διαστάσεις του τετράπλευρου.

Οι Maldonado - Bascon et al. (2007) [31] χρησιμοποιούν τα αποτελέσματα της μεθόδου Απόσταση από Όρια σε διαφορετικά Support Vector Machines. Όπως έχει ήδη αναφερθεί τα χρώματα και τα σχήματα των πινακίδων οδικής κυκλοφορίας είναι προκαθορισμένα, επομένως γίνονται οι απαραίτητοι έλεγχοι με βάση το χρώμα και στη συνέχεια ελέγχεται και το αντίστοιχο σχήμα. Αν, για παράδειγμα, η περιοχή ενδιαφέροντος έχει μπλε χρώμα τότε γίνεται έλεγχος για κυκλική πινακίδα, εάν είναι κίτρινη ακολουθεί έλεγχος για τριγωνικό σχήμα, ενώ αν εντοπιστεί κόκκινη περιοχή ενδιαφέροντος ελέγχεται αν το σχήμα είναι κυκλικό ή τριγωνικό. Στη συνέχεια υπολογίζονται οι αποστάσεις και εισάγονται στις διαφορετικές Μηχανές Διανυσματικής Στήριξης (Support Vector Machines - SVM) με βάση το χρώμα, και κατατάσσονται σε δύο κλάσεις. Εάν παραπάνω από μια πλευρές ανήκουν στην ίδια κλάση τότε κατατάσσεται η περιοχή ως σήμα, διαφορετικά απορρίπτεται ως θόρυβος.



Μέθοδος απόσταση από όρια για τριγωνική πινακίδα
 ΠΗΓΗ: Maldonado - Bascon et al. (2007) [31]



(a)



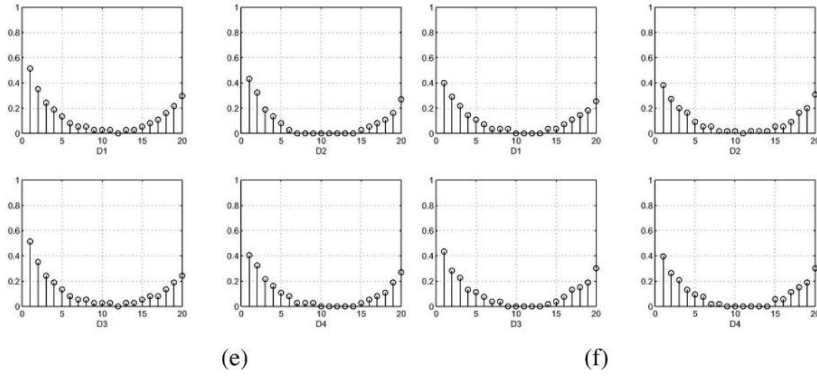
(b)



(c)



(d)



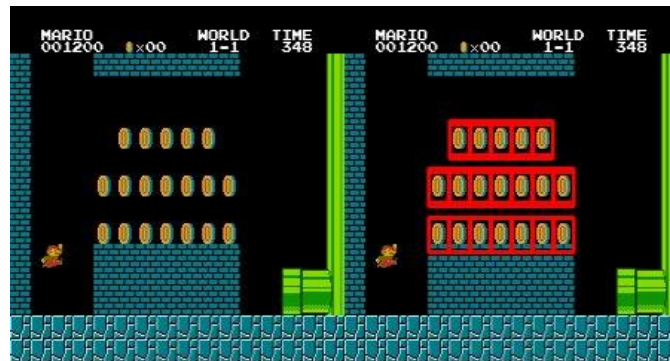
Μέθοδος απόσταση από όρια για κυκλική πινακίδα

(a) - (b) Αρχικές εικόνες. (c) - (d) Αποτελέσματα κατάτμησης με βάση το χρώμα (e) - (f) Διανύσματα μεθόδου Απόστασης από Όρια

ΠΗΓΗ: Maldonado - Bascon et al. (2007) [31]

1.1.2.4 Συνταύτιση Προτύπων

Σε αρκετές δημοσιεύσεις που αφορούν τον εντοπισμό με βάση το σχήμα χρησιμοποιείται αρκετά συχνά η τεχνική συνταύτισης προτύπων. Πρόκειται για μια τεχνική που αφορά την αναζήτηση και εύρεση της θέσης μικρών τμημάτων (templates) μιας εικόνας σε μια μεγαλύτερη εικόνα και τα οποία είναι γνωστά a priori. Συνταύτιση προτύπων μπορεί να επιτύχει κανείς χρησιμοποιώντας διάφορα μέτρα ομοιότητας, όπως η συσχέτιση (Cross correlation) ή το άθροισμα των απόλυτων διαφορών (Sum of absolute differences). Οι κύριες προκλήσεις στην υλοποίηση της συνταύτισης προτύπων είναι: η απόκρυψη, η ανίχνευση μηάκαμπτων μετασχηματισμών (non-rigid transformations), ο φωτισμός, οι αλλαγές στην κλίμακα καθώς και οι αλλαγές στο φόντο [41].



Συνταύτιση προτύπων με πολλαπλά αντικείμενα

ΠΗΓΗ: https://docs.opencv.org/master/d4/dc6/tutorial_py_template_matching.html

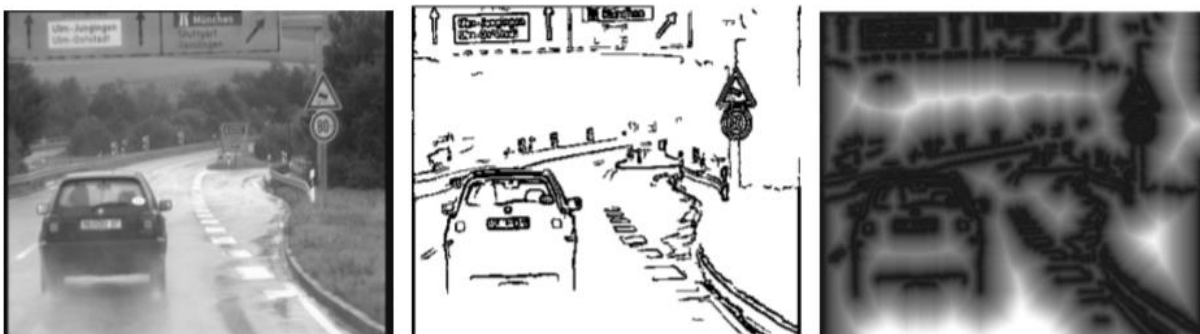
Η αντιμετώπιση του προβλήματος της συνταύτισης προτύπων προσεγγίζεται με δύο μεθόδους, την επιφανειακή συνταύτιση (Area-based approach) και τη συνταύτιση χαρακτηριστικών (Feature-based approach). Στην πρώτη περίπτωση συγκρίνονται οι τιμές φωτεινότητας κάθε εικονοψηφίδας του προτύπου με τις αντίστοιχες τιμές κάθε εικονοψηφίδας της περιοχής που εξετάζεται στην εικόνα. Στην προσέγγιση που βασίζεται στην εξαγωγή χαρακτηριστικών σημείων (με τη χρήση περιγραφών, πχ. HOG, SIFT, SURF), χρησιμοποιούνται χαρακτηριστικά σημεία της εικόνας με βάση διάφορες ιδιότητες, όπως σχήματα, υφές, χρώματα που έχουν εξαχθεί τόσο από το πρότυπο όσο και από την περιοχή ενδιαφέροντος, τα οποία συγκρίνονται μεταξύ τους, Αδάμ (2014) [10].

Η διαδικασία μπορεί να επιταχυνθεί με τη χρήση πυραμίδων εικόνων ή με εικόνες ακμών. Οι πυραμίδες εικόνων ουσιαστικά αποτελούν μια σειρά εικόνων σε διαφορετικές κλίμακες, οι οποίες σχηματίζονται με εφαρμογή αλληπαλλήλων φίλτρων εξομάλυνσης στην αρχική εικόνα,

προκειμένου να δημιουργηθεί μια ακολουθία εικόνων μειωμένης ανάλυσης (Adelson et al. [42]). Επιπλέον, μπορούν να γίνουν βελτιώσεις στη μέθοδο συνταύτισης προτύπων με περισσότερα από ένα πρότυπα, τα οποία μπορούν να έχουν διαφορετικές κλίμακες και στροφές.

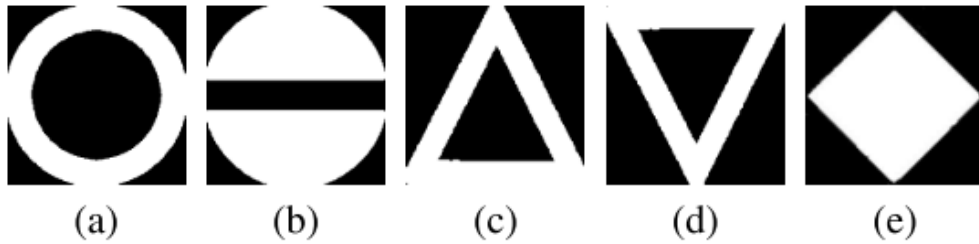
Ο Gavrilu (1999) [43] χρησιμοποιεί τη συνταύτιση προτύπων υλοποιώντας έναν μετασχηματισμό απόστασης κάθε εικονοστοιχείου από τις ακμές τις εικόνας για την ανίχνευση του σχήματος της πινακίδας. Αναλυτικότερα, πρώτα εξάγονται οι ακμές από την αρχική εικόνα και στη συνέχεια δημιουργείται η εικόνα μετασχηματισμού απόστασης, στην οποία η τιμή που δίνεται σε κάθε εικονοστοιχείο αντιπροσωπεύει την απόστασή του από την πλησιέστερη ακμή. Για να εντοπιστεί το σχήμα ενδιαφέροντος, η βασική ιδέα είναι να ταιριάξει ένα πρότυπο (για παράδειγμα, ένα κανονικό τρίγωνο) με την εικόνα του μετασχηματισμού απόστασης. Προκειμένου να βρεθεί η βέλτιστη αντιστοίχιση, το πρότυπο στρέφεται, αλλάζει κλίμακα και υπόκειται σε μετάθεση.

Στην επέκταση της βασικής ιδέας του Gavrilu (1999) [43], οι ακμές διαφοροποιούνται ανάλογα με τον προσανατολισμό τους, έτσι ώστε να υπολογιστούν ξεχωριστές εικόνες μετασχηματισμού απόστασης βάσει διακριτών διαστημάτων του προσανατολισμού των ακμών και τα πρότυπα να διαχωρίζονται σε μέρη με βάση τους προσανατολισμούς των ακμών τους. Το χρησιμοποιούμενο μέτρο ομοιότητας είναι ένα άθροισμα μετρήσεων της αντιστοίχισης μεταξύ των εικόνων μετασχηματισμού αποστάσεων και των προτύπων συγκεκριμένων προσανατολισμών. Επίσης, χρησιμοποιείται μια ιεραρχία προτύπων, δηλαδή παρόμοια πρότυπα ομαδοποιούνται σε πρωτότυπα και, μόλις βρεθεί το πρωτότυπο, η διαδικασία βρίσκει το καλύτερο πρότυπο μέσα στο πρωτότυπο. Η διαδικασία αυτή εξοικονομεί κυρίως υπολογιστικό κόστος.



Δημιουργία εικόνας μετασχηματισμού απόστασης. Από αριστερά προς τα δεξιά: η αρχική εικόνα, η εικόνα ακμών και η εικόνα μετασχηματισμού απόστασης.
ΠΗΓΗ: Gavrilu (1999) [43]

Στην εργασία των Broggi et al. (2007) [22] δημιουργούνται ως πρότυπα πέντε δυαδικές εικόνες, μία για κάθε σχήμα που πρέπει να ανιχνευθεί. Κάθε περιοχή ενδιαφέροντος επαναπροσδιορίζεται, ώστε να αποκτήσει διαστάσεις 50x50 εικονοψηφίδες, προκειμένου να ταυτίζεται με τις διαστάσεις του προτύπου. Με αυτόν τον τρόπο μειώνεται η υπολογιστική πολυπλοκότητα και ο υπολογιστικός χρόνος, ενώ παράλληλα ο αλγόριθμος μπορεί να είναι ανεκτικός σε αλλαγές κλίμακας, χωρίς να χρειάζεται να χρησιμοποιηθούν πρότυπα διαφορετικών μεγεθών. Το μέτρο ομοιότητας για τη συνταύτιση είναι ο λόγος των σωστά συνταυτισμένων εικονοψηφίδων μεταξύ προτύπου και περιοχής ενδιαφέροντος ως προς τον αριθμό των λευκών εικονοψηφίδων σε κάθε δυαδική εικόνα. Η συνταύτιση θεωρείται επιτυχής εάν ο λόγος υπερβαίνει ένα προκαθορισμένο όριο. Η περιοχή ενδιαφέροντος ταυτίζεται με το πρότυπο, το οποίο εμφανίζει τον μέγιστο λόγο κατά τη διαδικασία συνταύτισης.



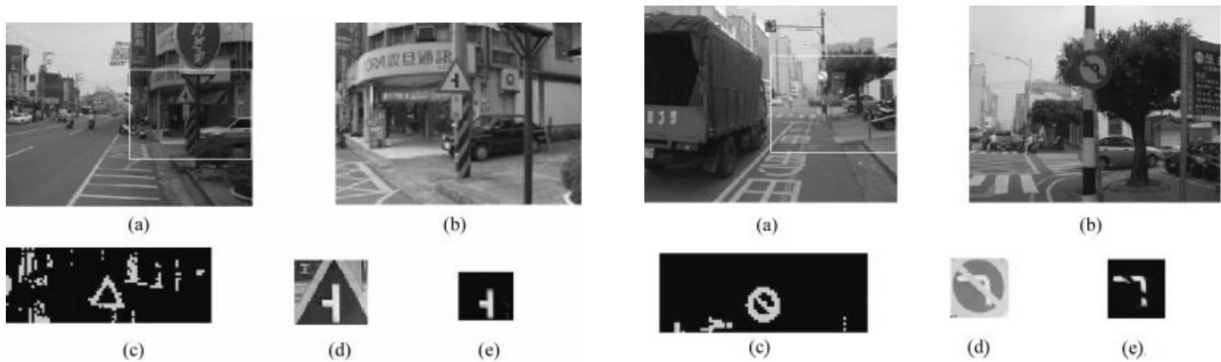
Μοτίβα αναζητούμενων σχημάτων
 ΠΗΓΗ: Broggi et al. (2007) [22]

Το σύστημα ανίχνευσης οδικών σημάτων των Hsu et al. (2001) [44] αποτελείται από τρία στάδια. Αρχικά, εντοπίζονται οι περιοχές ενδιαφέροντος (ROI) για τις οποίες εξάγεται η κόκκινη συνιστώσα στον χρωματικό χώρο RGB. Έπειτα, δύο πρότυπα, ένα για τα τριγωνικά και ένα για τα κυκλικά σήματα, μετακινούνται μέσα στην περιοχή ενδιαφέροντος και σε κάθε θέση υπολογίζεται ο αριθμός των κόκκινων εικονοψηφίδων που εντοπίζονται εντός του δακτυλίου (διπλό περίγραμμα κύκλου και τριγώνου). Τα πρότυπα που εφαρμόζονται έχουν μεταβλητό μέγεθος, ώστε ο αλγόριθμος να μην εξαρτάται από την κλίμακα της λήψης.



Το τριγωνικό και κυκλικό πρότυπο μεταβλητού μεγέθους
 ΠΗΓΗ: Hsu et al. (2001) [44]

Εάν ο αριθμός των κόκκινων εικονοψηφίδων υπερβεί ένα προκαθορισμένο κατώφλι σε μία συγκεκριμένη θέση, επιλέγεται ως τελική περιοχή ενδιαφέροντος η περιοχή με κέντρο το κέντρο του προτύπου στη θέση αυτή και διαστάσεις ίσες με τις διαστάσεις του προτύπου.



Αποτελέσματα ανίχνευσης των τριγωνικών και κυκλικών οδικών σημάτων κυκλοφορίας.
 ΠΗΓΗ: Hsu et al. (2001) [44]

1.1.2.5 Αλγόριθμος RANSAC

Στην πραγματικότητα, οι περισσότεροι από τους ανιχνευτές οδικών πινακίδων βασίζονται στον εντοπισμό με βάση το σχήμα στην εικόνα κλίσης της αρχικής εικόνας. Μία άλλη κατηγορία που συναντάται στη βιβλιογραφία αποτελεί η εφαρμογή περιγράμματος Contour Fitting (CF), (Belaroussi et al. (2010) [45]), στην οποία συγκαταλέγεται η ανίχνευση ακμών με τη βοήθεια του αλγορίθμου RANSAC (RANdom SAMple Consensus). Ο αλγόριθμος RANSAC δημοσιεύθηκε για πρώτη φορά από τους Fischler & Bolles το 1981 [46].



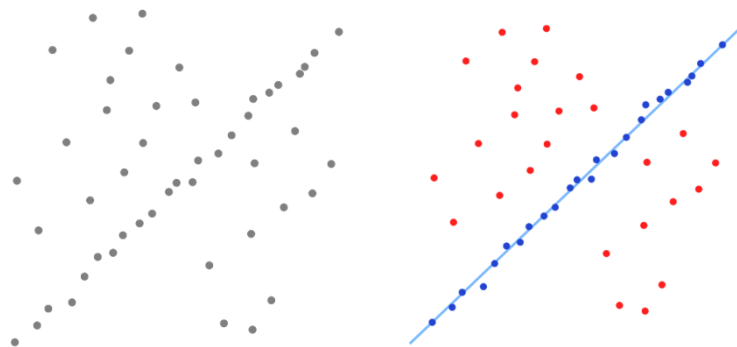
Παραδείγματα ανιχνεύσεων CF
 ΠΗΓΗ: Belaroussi et al. (2010)[45]

Ο RANSAC ουσιαστικά αποτελεί έναν αλγόριθμο εντοπισμού χονδροειδών σφαλμάτων. Έρευνες μεταξύ διαφόρων αλγορίθμων εντοπισμού χονδροειδών σφαλμάτων διαπίστωσαν ότι αυτός συγκαταλέγεται ανάμεσα στους αποτελεσματικότερους, ακόμα και σε περιπτώσεις όπου οι εσφαλμένες ομολογίες είναι περισσότερες από τις ορθές (Καλησπεράκης, 2010 [78]). Εν συντομία, επιλέγονται τυχαία ζευγάρια εικονοψηφίδων που έχουν αναγνωριστεί ότι βρίσκονται πάνω σε ακμή για να σχηματίσει μια υπόθεση γραμμής και στη συνέχεια ελέγχεται πόσα άλλα τέτοια εικονοστοιχεία εμπίπτουν σε αυτήν τη γραμμή. Γραμμές με επαρκώς μεγάλο αριθμό inliers (συμβατές εικονοψηφίδες) επιλέγονται στη συνέχεια ως τα επιθυμητά τμήματα γραμμής. Τα υπόλοιπα δεδομένα θεωρούνται χονδροειδή σφάλματα ή ακραίες τιμές (outliers). Οι ακραίες τιμές μπορεί να προέρχονται, για παράδειγμα, από ακραίες τιμές θορύβου ή από εσφαλμένες μετρήσεις ή εσφαλμένες υποθέσεις σχετικά με την ερμηνεία των δεδομένων Szeliski (2010) [47].

Πρόκειται για έναν μη ντετερμινιστικό αλγόριθμο με την έννοια ότι παράγει ένα λογικό αποτέλεσμα μόνο με μια συγκεκριμένη πιθανότητα, με αυτήν την πιθανότητα να αυξάνεται καθώς επιτρέπονται πολλές επαναλήψεις. Ο αριθμός N των επαναλήψεων που απαιτούνται ώστε κάποιο από τα δείγματα να περιλαμβάνει μόνο inliers με πιθανότητα p δίνεται από την εξής σχέση:

$$N = \frac{\log(1 - p)}{\log(1 - (1 - e)^s)}$$

όπου s ο αριθμός των ελάχιστων σημείων που απαιτούνται σε κάθε δείγμα και e η εκτίμηση του ποσοστού των εσφαλμένων υποθέσεων.



Αποτέλεσμα RANSAC, οι ακραίες τιμές δεν επηρεάζουν το αποτέλεσμα.
 ΠΗΓΗ: https://en.wikipedia.org/wiki/Random_sample_consensus

Οι Arlicot et al. (2009) [48] προχωρούν σε κατωφλίωση ώστε να προκύψουν διάφορες ακμές με την εφαρμογή του τελεστή Canny edge detector, οι οποίες τροφοδοτούνται στον RANSAC. Επιπλέον λαμβάνουν υπόψη τους και τις προοπτικές παραμορφώσεις της λήψης, που έχουν ως αποτέλεσμα οι κυκλικές πινακίδες να εμφανίζονται ως ελλείψεις. Στη συνέχεια, δημιουργούν διάφορα δείγματα δεδομένων αποτελούμενα από τρία σημεία. Αν και η έλλειψη χρειάζεται πέντε σημεία για τον πλήρη ορισμό της, οι ερευνητές εκμεταλλεύονται εδώ το θεώρημα του Pascal, σύμφωνα με το οποίο μία έλλειψη μπορεί να οριστεί πλήρως από τρία μόνο σημεία. Με βάση αυτό, ο αλγόριθμος δημιουργεί διάφορες υπο-ομάδες αποτελούμενες από τρία σημεία. Από κάθε τριάδα σημείων ορίζεται μία έλλειψη και στη συνέχεια ελέγχεται πόσα από τα σημεία που απαρτίζουν το σύνολο δεδομένων και δεν συμμετείχαν στον υπολογισμό των παραμέτρων της έλλειψης ανήκουν σε αυτήν. Πρακτικά υπολογίζεται η απόσταση των υπόλοιπων σημείων από την έλλειψη και τα σημεία με απόσταση μικρότερη ενός προκαθορισμένου ορίου αποτελούν τα inliers. Τελικά, επιλέγεται η λύση με τα περισσότερα συμβατά δεδομένα και βάσει αυτών υπολογίζεται η τελική έλλειψη. Εάν ολοκληρωθεί ο αριθμός των ελέγχων χωρίς να εντοπιστεί λύση με τον επιθυμητό αριθμό inliers, η περιοχή ενδιαφέροντος απορρίπτεται.

Οι Kim & Lee (2016) [49] προκειμένου να έχουν έναν πιο ισχυρό αλγόριθμο απέναντι σε διαφορετικές συνθήκες λήψης, όπως αλλαγές στον καιρό και τον φωτισμό, προτείνουν την ανίχνευση με RANSAC - βασισμένη στην εκτίμηση κίνησης. Χρησιμοποιώντας εκτίμηση κίνησης δύο σημείων, επιλέγουν τα εσωτερικά σημεία χαρακτηριστικών εντός της πινακίδας κυκλοφορίας και στη συνέχεια η βέλτιστη κίνηση υπολογίζεται με τα inliers μέσω συνόρθωσης δέσμης (bundle adjustment). Αυτή η προσέγγιση βελτιώνει σημαντικά την απόδοση του αλγορίθμου. Ανίχνευση χρώματος και σχήματος που βασίζεται σε RANSAC για κυκλικές πινακίδες χρησιμοποιεί και το σύστημα των Eichner & Breckon (2008) [33].

1.1.2.6 Αλγόριθμοι SIFT και SURF

Ο αλγόριθμος SIFT είναι ίσως ο πιο δημοφιλής αλγόριθμος εξαγωγής χαρακτηριστικών σημείων που χρησιμοποιείται στην Όραση Υπολογιστών. Το όνομα του προκύπτει από τα αρχικά Scale Invariant Feature Transform και έχει προταθεί από τον David Lowe το 1999. Μια δεύτερη πιο εκτεταμένη εργασία έγινε από τον ίδιο το 2004 [79]. Ο αλγόριθμος SIFT αποτελείται από έναν ανιχνευτή χαρακτηριστικών και από έναν περιγραφέα (descriptor). Για κάθε αντικείμενο σε μια εικόνα μπορούν να εντοπιστούν σημεία ενδιαφέροντος, τα οποία μπορούν να χρησιμοποιηθούν για να περιγράψουν το αντικείμενο με μαθηματικό τρόπο. Στην περίπτωση του αλγορίθμου SIFT αυτά τα σημεία εντοπίζονται κοντά σε σαφή σημεία (corners) των αντικειμένων, καθώς αυτά φέρουν διακριτή πληροφορία ανεξάρτητη από μεταβολές όπως η στροφή και οι αλλαγές φωτισμού.

Με τη χρήση μιας γκαουσιανής κατανομής, η εικόνα μετασχηματίζεται από το χρωματικό χώρο στο χώρο κλίμακας (scale space) που προτάθηκε από τον Lindeberg (1994). Πρακτικά, ο αλγόριθμος αναζητεί τα σημεία ενδιαφέροντος που διατηρούνται αναλλοίωτα στις μεταβολές της κλίμακας σε πυραμίδες εικόνων. Για την δημιουργία τους πραγματοποιείται συνέλιξη με φίλτρο εξομάλυνσης Gauss με διαφορετικές τυπικές αποκλίσεις, με αποτέλεσμα να παράγονται εικόνες με διαφορετική κλίμακα:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)$$

$$\text{όπου } G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}$$

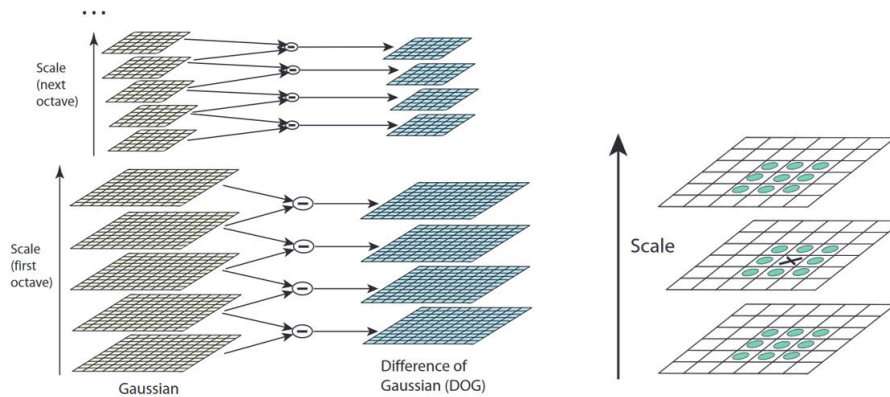
$I(x, y)$ η αρχική εικόνα

$L(x, y, \sigma)$ η εικόνα συνέλιξης

Έπειτα, οι διαδοχικές εικόνες διαφορετικής κλίμακας κάθε επιπέδου αφαιρούνται μεταξύ τους με αποτέλεσμα να δημιουργείται η συνάρτηση διαφοράς (Difference of Gaussian - DoG) ή αλλιώς η πυραμίδα DoG, η οποία προσεγγίζει το φίλτρο LoG (Laplacian of Gaussian) και

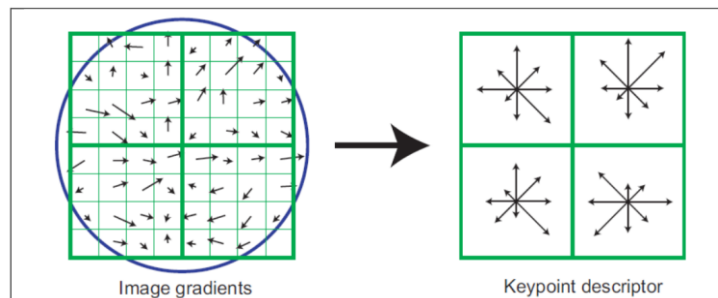
πρακτικά μπορεί να υπολογιστεί από την διαφορά δύο διαδοχικών κλιμάκων που διαφέρουν κατά την σταθερά k :

$$D(x, y, \sigma) = [G(x, y, k\sigma) - G(x, y, \sigma)] * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma)$$



Πυραμίδες Gaussian - DoG και σύγκριση εικονοψηφίδων με γειτονικά για εξαγωγή σημείου ενδιαφέροντος. ΠΗΓΗ: Lowe (2004) [79]

Στη συνέχεια, η συνάρτηση διαφοράς των Γκαουσιανών (DoG), συνδυαζόμενη με μια παρεμβολή στο χώρο κλίμακας, παράγει τα σημεία ενδιαφέροντος. Ο περιγραφέας είναι άμεσα συνδεδεμένος με τη διαδικασία ανάθεσης ενός προσανατολισμού σε κάθε σημείο ενδιαφέροντος, όπως αυτό έχει προκύψει από τον ανιχνευτή. Συγκεκριμένα, ο περιγραφέας είναι ένα διάνυσμα, το μέτρο και η διεύθυνση του οποίου υπολογίζονται από πληροφορίες της γειτονιάς του εκάστοτε σημείου ενδιαφέροντος. Τα καλύτερα αποτελέσματα προκύπτουν όταν λαμβάνεται πληροφορία από μια γειτονιά 4x4 εικονοστοιχείων. Η διεύθυνση του διανύσματος τυποποιείται έτσι ώστε να λαμβάνει προσανατολισμούς ανά 45°. Με αυτό το τρόπο, τοπικές πληροφορίες ενός αντικειμένου προσδιορίζονται με έναν και μοναδικό τρόπο, γεγονός που εξασφαλίζει τη μοναδικότητα κάθε χαρακτηριστικού σημείου ενός αντικειμένου, Κουσκουρίδας κ.ά. (2009) [80].



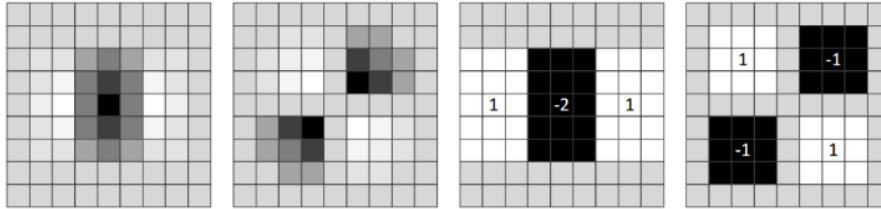
Υπολογισμός περιγραφικών χαρακτηριστικών. Εδώ φαίνεται ενδεικτικά περιγραφέας 2x2 υπολογισμένους από 8x8 δείγματα. ΠΗΓΗ: Lowe (2004) [79]

Άλλος ένας γρήγορος και εύρωστος μορφολογικός αλγόριθμος για την αναπαράσταση εικόνων και την εξαγωγή χαρακτηριστικών σημείων είναι ο αλγόριθμος SURF (Speeded Up Robust Features). Η προσέγγιση του SURF διαφέρει από αυτή του SIFT σε δύο σημεία που τον καθιστούν αρκετά ταχύτερο. Επιπλέον, τα αποτελέσματα του SURF όσον αφορά το ποσοστό επιτυχίας ταύτισης σε σχέση με αυτά του SIFT είναι παρόμοια, αν και οι συγγραφείς ισχυρίζονται ότι ο SURF πετυχαίνει καλύτερα αποτελέσματα από τον SIFT (Bay et al., 2006 [81]).

Ο SURF αποτελείται από τρία βασικά βήματα:

- Ανίχνευση των σημείων ενδιαφέροντος
- Περιγραφή των σημείων ενδιαφέροντος
- Ταύτιση χαρακτηριστικών

Όπως στον SIFT, έτσι και στον SURF τα δύο πρώτα βήματα στηρίζονται σε μια scale-space αναπαράσταση της εικόνας και σε διαφορικές παραγώγους πρώτης και δεύτερης τάξης. Αυτό που κάνει τον SURF να διαφέρει είναι ότι αυτές οι διαδικασίες επιταχύνονται κάνοντας χρήση της εικόνας ολοκλήρωσης (integral image - αναλύεται παρακάτω στην ενότητα 1.3.1.1) και τεχνικών φιλτραρίσματος κουτιού (box filter techniques) για την προσομοίωση του scale-space και την επιλογή των σημείων ενδιαφέροντος, αντίστοιχα.



Προσέγγιση LoG με φίλτρο κουτιών (box filter techniques)
ΠΗΓΗ: Zhao et al. (2013) [82]

Στο στάδιο της ανίχνευσης, υπολογίζονται τα τοπικά μέγιστα του τελεστή Hessian (Hessian determinant operator), ο οποίος βασίζεται στους υπολογισμούς του Εσσιανού πίνακα για κάθε ένα σημείο στην εικόνα και εφαρμόζονται στην αναπαράσταση scale-space για την επιλογή των υποψήφιων σημείων ενδιαφέροντος. Στην κλασική περίπτωση το τρίτο βήμα, ταυτίζει τους περιγραφείς δύο εικόνων. Γίνονται όλες οι συγκρίσεις υπολογίζοντας την ευκλείδεια απόσταση ανάμεσα σε όλα τα πιθανά ζευγάρια περιγραφέων. Με βάση την εξαγωγή διανυσμάτων των περιγραφέων, στο τέλος γίνεται η σύγκρισή τους και προκύπτει αν αυτά είναι ίδια σε κάθε εικόνα

Ο SURF είναι ένας αλγόριθμος ο οποίος παρουσιάζει μεγάλη αντοχή στις αλλαγές κλίμακας και στροφών. Αυτές οι δυνατότητες τον καθιστούν ιδανικό για αντιστοίχιση και ανίχνευση αντικειμένων. Δυστυχώς όμως, λόγω της υπολογιστικής πολυπλοκότητάς του, η εφαρμογή του αλγορίθμου παραμένει περιορισμένη σε συστήματα πραγματικού χρόνου. Οι Zhao et al. (2013) [82] παρουσιάζουν ένα σύστημα εντοπισμού σημάτων οδικής κυκλοφορίας βασισμένο σε SURF σε πραγματικό χρόνο, εκμεταλλευόμενοι τον παραλληλισμό και τους πόρους της Επιτόπιας Συστοιχίας Προγραμματιζόμενων Πυλών (FPGA - Field Programmable Gate Array).



Ταύτιση χαρακτηριστικών πινακίδας STOP με φυσική εικόνα
ΠΗΓΗ : Zhao et al. (2013) [82]

Το FPGA είναι ένα ολοκληρωμένο κύκλωμα σχεδιασμένο να διαμορφώνεται και μετά την κατασκευή του - εξ ου και ο όρος " Field Programmable", το οποίο βοηθάει στο να εκτελεστούν γρήγορα σύνθετες και συνδυαστικές λειτουργίες, παρέχοντας αρκετή μνήμη. Με αυτόν τον τρόπο ο SURF καθίσταται καλύτερη επιλογή για πραγματικού χρόνου εφαρμογές ανίχνευσης και αναγνώρισης, όπως η δημιουργία συστημάτων υποβοήθησης της οδήγησης (ADAS).

1.2 Η Μηχανική Μάθηση

Η Μηχανική Μάθηση (Machine Learning - ML) είναι η μελέτη και κατασκευή αλγορίθμων υπολογιστών που μπορούν να μαθαίνουν από τα δεδομένα, να κάνουν προβλέψεις σχετικά με αυτά και οι οποίοι βελτιώνονται αυτόματα μέσω της εμπειρίας τους. Θεωρείται ως ένα υποσύνολο της επιστήμης των υπολογιστών και κλάδος της Τεχνητής Νοημοσύνης (Artificial Intelligence), που αναπτύχθηκε από τη μελέτη της αναγνώρισης προτύπων και της υπολογιστικής θεωρίας μάθησης στην τεχνητή νοημοσύνη (Artificial Intelligence - AI). Το 1959, ο Arthur Lee Samuel ορίζει τη Μηχανική Μάθηση ως:

«Πεδίο μελέτης που δίνει στους υπολογιστές την ικανότητα να μαθαίνουν, χωρίς να έχουν ρητά προγραμματιστεί.»

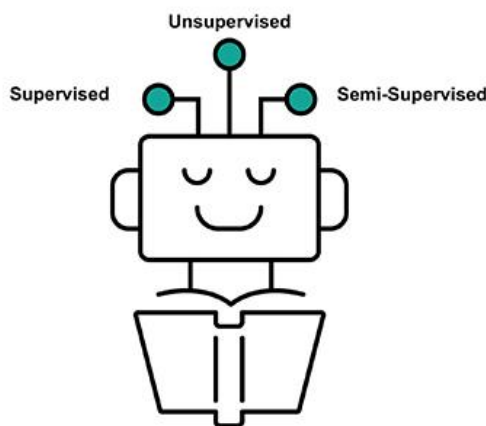
Οι αλγόριθμοι μηχανικής εκμάθησης δημιουργούν ένα μαθηματικό μοντέλο βασισμένο σε δείγματα δεδομένων, γνωστά ως δεδομένα εκπαίδευσης (training data), προκειμένου να εξάγουν τις προβλέψεις ή τις αποφάσεις χωρίς να έχουν προγραμματιστεί ρητά να το κάνουν. Οι αλγόριθμοι μηχανικής μάθησης χρησιμοποιούνται σε μια ευρεία ποικιλία εφαρμογών, όπως η Όραση Υπολογιστών (Computer Vision), όπου είναι δύσκολο ή ανέφικτο να αναπτυχθούν συμβατικοί αλγόριθμοι για την εκτέλεση των απαιτούμενων εργασιών.

Με βάση τα παραπάνω, μπορεί να δοθεί ο ακόλουθος εναλλακτικός ορισμός για τη Μηχανική Μάθηση:

«Μηχανική Μάθηση ονομάζεται η ικανότητα ενός υπολογιστικού συστήματος να δημιουργεί μοντέλα ή πρότυπα από ένα σύνολο δεδομένων.»

1.2.1 Είδη Μηχανικής Μάθησης

Οι τύποι αλγορίθμων μηχανικής μάθησης διαφέρουν ως προς την προσέγγισή τους, τον τύπο δεδομένων που εισάγουν και εξάγουν, και τον τύπο εργασίας ή προβλήματος που σκοπεύουν να λύσουν, συνήθως δε ταξινομούνται σε τρεις μεγάλες κατηγορίες:



ΠΗΓΗ: <https://abhitronix.github.io/2018/12/14/humanoid-Brain-2/>

- **Επιβλεπόμενη μάθηση - Supervised learning (αλλιώς επιτηρούμενη μάθηση ή μάθηση με επίβλεψη)**

Είναι η διαδικασία, όπου ο αλγόριθμος κατασκευάζει μια συνάρτηση που απεικονίζει δεδομένες εισόδους (σύνολο εκπαίδευσης) σε γνωστές επιθυμητές εξόδους, με απώτερο στόχο τη γενίκευση της συνάρτησης αυτής και για εισόδους με άγνωστη έξοδο. Χρησιμοποιείται σε προβλήματα:

- Ταξινόμησης (Classification)
- Παλινδρόμησης (Regression)
- Πρόβλεψης (Prediction)

- **Μη επιβλεπόμενη μάθηση - Unsupervised learning (αλλιώς μη επιτηρούμενη μάθηση ή μάθηση χωρίς επίβλεψη)**

Είναι η διαδικασία όπου ο αλγόριθμος δημιουργεί ένα μοντέλο για κάποιο σύνολο εισόδων υπό μορφή παρατηρήσεων χωρίς να γνωρίζει τις επιθυμητές εξόδους. Η μη επιτηρούμενη μάθηση μπορεί να είναι αυτοσκοπός (στοχεύοντας δηλαδή στον εντοπισμό κρυμμένων μοτίβων σε δεδομένα) ή μέσο για ένα τέλος (χαρακτηριστικό της μάθησης). Χρησιμοποιείται σε προβλήματα:

- Ομαδοποίησης (Clustering)
- Ανάλυσης Συσχετισμών (Association Analysis)

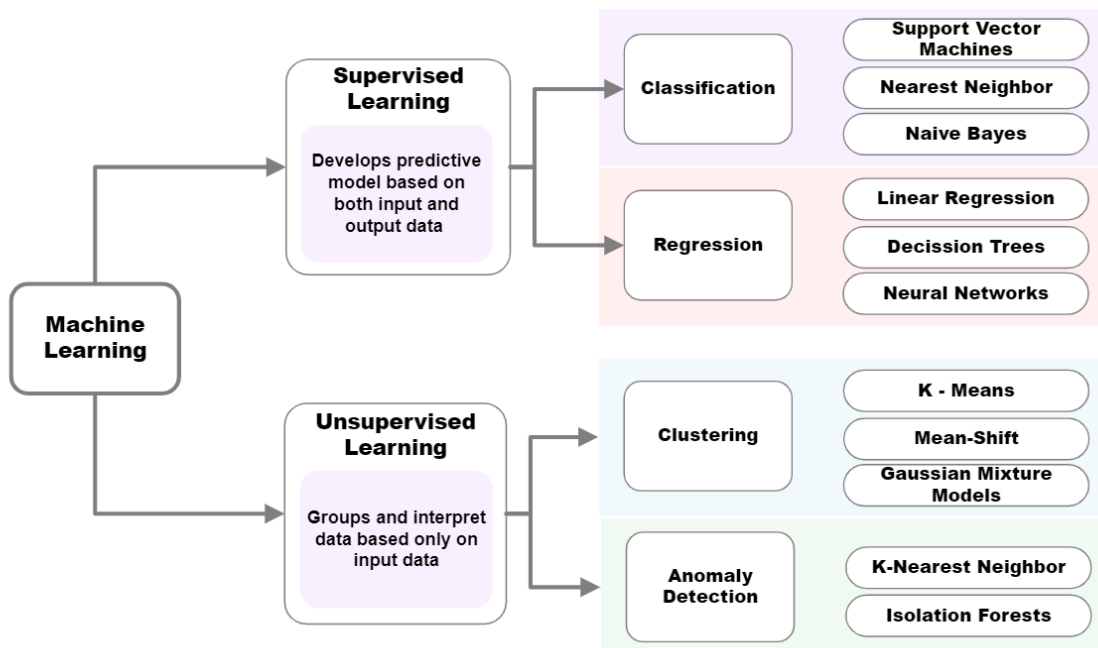
- **Ενισχυτική μάθηση (Reinforcement learning)**

Είναι η διαδικασία κατά την οποία ο αλγόριθμος μαθαίνει μια στρατηγική ενεργειών αλληλεπιδρώντας άμεσα με ένα δυναμικό περιβάλλον στο οποίο πρέπει να επιτευχθεί ένας συγκεκριμένος στόχος. Χρησιμοποιείται κυρίως σε προβλήματα σχεδιασμού (planning), όπως για παράδειγμα ο έλεγχος κίνησης ρομπότ και η βελτιστοποίηση εργασιών σε εργοστασιακούς χώρους, ή για να μάθει να παίζει ένα παιχνίδι εναντίον κάποιου αντιπάλου. Επιπλέον, η ενισχυτική μάθηση χρησιμοποιείται και για την υποβοήθηση της οδήγησης (Driving Assistance Systems) ή για την πλήρως αυτοματοποιημένη οδήγηση σε ευφυή οχήματα (IVS).

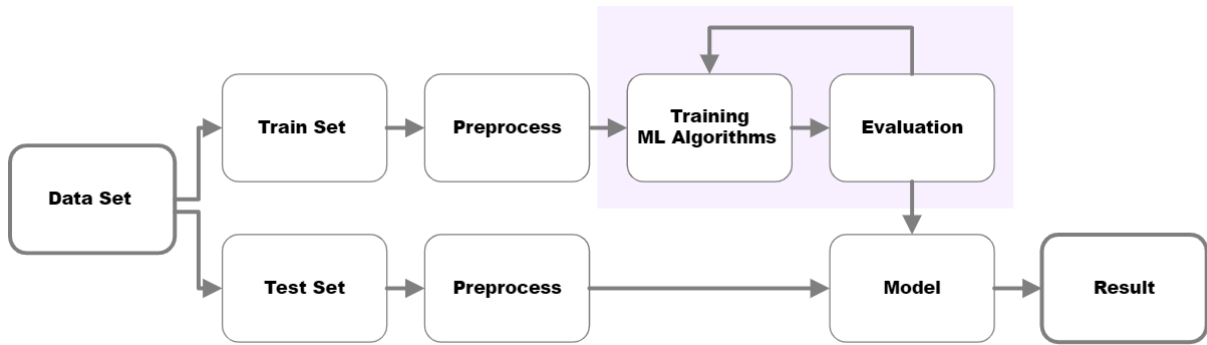
- **Ημι-επιτηρούμενη μάθηση (Semi-supervised learning)**

Μεταξύ της επιτηρούμενης και της μη επιτηρούμενης μάθησης είναι η ημι-επιτηρούμενη μάθηση, όπου στον αλγόριθμο δίνεται ένα σύνολο δεδομένων εκπαίδευσης με κάποια (συνχά πολλά) από τα αποτελέσματα – επιθυμητές εξόδους να λείπουν.

Για κάθε πρόβλημα προς επίλυση στο χώρο της Μηχανικής Μάθησης υπάρχει ένας κατάλληλος τρόπος μάθησης και για κάθε τρόπο μάθησης υπάρχει τουλάχιστον ένας κατάλληλος αλγόριθμος που μπορεί να χρησιμοποιηθεί. Η βασικότερη φάση κάθε αλγόριθμου είναι η εκπαίδευση, όπου ο αλγόριθμος χρησιμοποιεί ως είσοδο ένα σύνολο δεδομένων εκπαίδευσης (training set) με στόχο τη δημιουργία νέας γνώσης. Επιπλέον, μπορεί είτε να αξιοποιεί λιγότερο ή περισσότερο την υπάρχουσα γνώση είτε να μην τη χρησιμοποιήσει καθόλου. Ορισμένοι αλγόριθμοι δέχονται ως είσοδο μόνο παρατηρήσεις και άλλοι λαμβάνουν υπόψη τους λίγο ή περισσότερο την προ-υπάρχουσα γνώση (Γεωργούλη, 2015 [50]).



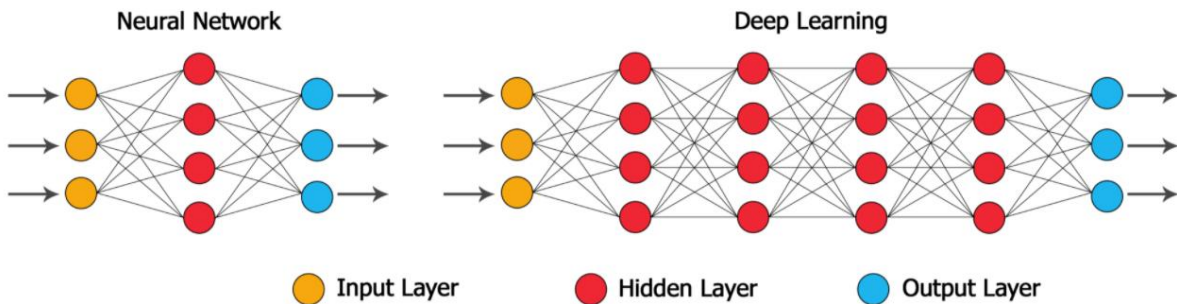
Διαφορές Επιβλεπόμενης και Μη επιβλεπόμενης Μάθησης με παραδείγματα αλγορίθμων



Γενικός τρόπος λειτουργίας των αλγορίθμων Μηχανικής Μάθησης

• **Βαθιά Μηχανική Μάθηση (Deep learning)**

Οι μέθοδοι Βαθιάς Μάθησης (Deep Learning) είναι μέθοδοι αναπαραστατικής μάθησης με πολλαπλά επίπεδα αναπαράστασης. Κάθε νέο επίπεδο δημιουργείται έπειτα από μετασχηματισμό του προηγούμενου επιπέδου μέσω απλών μη γραμμικών μεθόδων, και αποτελεί έναν καινούριο (πιο αφηρημένο) τρόπο αναπαράστασης των δεδομένων. Ουσιαστικά η βαθιά μάθηση χρησιμοποιεί τεχνητά νευρωνικά δίκτυα πολλαπλών επιπέδων (δηλαδή δίκτυα στα οποία υπάρχουν περισσότερα από ένα κρυφά επίπεδα), προκειμένου να παρέχει αποτελέσματα ακριβείας. Αποτελεί την βασική τεχνολογία αιχμής πίσω από τη δημιουργία συστημάτων για την υποβοήθηση της οδήγησης (Driving Assistance Systems), ή για την πλήρως αυτοματοποιημένη οδήγηση, επιτρέποντας την αναγνώριση των σημάτων οδικής κυκλοφορίας, των πεζών, των φαναριών κλπ.



Διαφορά μεταξύ ενός νευρικού δικτύου και ενός βαθιού νευρωνικού δικτύου
 ΠΗΓΗ: <https://alphabold.com/neural-networks-and-deep-learning-an-overview/>

• **Μεταφορά Μάθησης (Transfer Learning)**

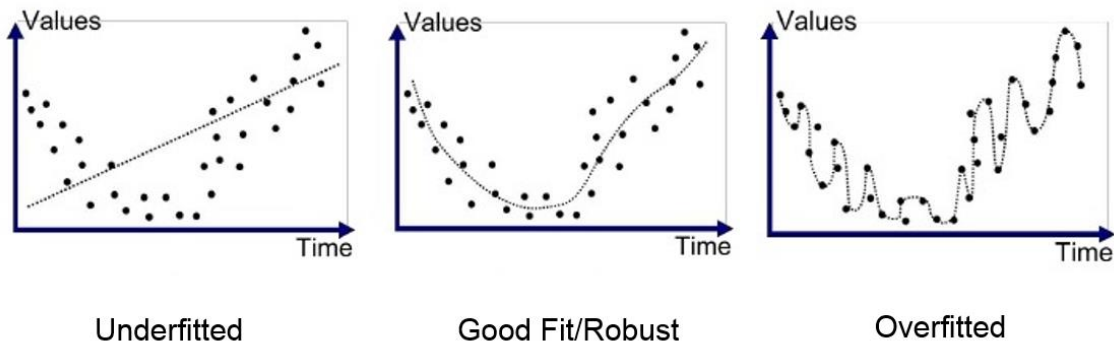
Η μεταφορά μάθησης είναι μια μέθοδος που επιτρέπει να χρησιμοποιηθούν οι γνώσεις που αποκτήθηκαν από άλλες εργασίες προκειμένου να αντιμετωπιστεί ένα νέο αλλά παρόμοιο πρόβλημα γρήγορα και αποτελεσματικά. Για παράδειγμα, στην περίπτωση των ανθρώπων η χρήση γνώσεων για ποδηλασία θα διευκολύνει την εκμάθηση ποδηλασίας. Αυτό μειώνει την ανάγκη για δεδομένα που σχετίζονται με το πρόβλημα το οποίο χρήζει αντιμετώπισης.

Μεταξύ άλλων κατηγοριών μηχανικής μάθησης, υπάρχουν ακόμα οι διαδικασίες εκμάθησης Self-learning, Feature learning, Sparse dictionary learning καθώς και η τεχνική Meta learning που μαθαίνει στην μηχανή να αναπτύσσει τις δικές της επαγωγικές μεθόδους, βασιζόμενη σε προηγούμενη εμπειρία.

1.2.1.1 Υπερπροσαρμογή και Υποπροσαρμογή μοντέλου

Αφού ολοκληρωθεί η εκπαίδευση του μοντέλου από τα δεδομένα, στόχος είναι η γενίκευση (generalization) του μοντέλου ώστε να μπορεί να κάνει όσο το δυνατόν ακριβέστερες προβλέψεις για καινούρια δεδομένα. Ένα μοντέλο που δεν οδηγεί στη γενίκευση της εμπειρίας του συστήματος σημαίνει ότι δεν μπορεί να κάνει ακριβείς προβλέψεις για νέα δεδομένα που συναντά. Στην περίπτωση λοιπόν που το μοντέλο είναι προσαρμοσμένο σε ένα συγκεκριμένο σύνολο δεδομένων σε τέτοιο βαθμό ώστε να αποδίδει πολύ ακριβή αποτελέσματα στα δεδομένα εκπαίδευσης, αλλά δεν προσαρμόζεται καλά σε νέα δεδομένα ή δεν προβλέπει νέες παρατηρήσεις αξιόπιστα, παρατηρείται το φαινόμενο της υπερπροσαρμογής - υπερμοντελοποίησης (overfitting). Το φαινόμενο αυτό προκαλείται συνήθως όταν το στατιστικό μοντέλο έχει περισσότερες παραμέτρους από αυτές που είναι αναγκαίες για να εκφράσει τα χαρακτηριστικά των δεδομένων.

Από την άλλη πλευρά, ένα μοντέλο πρόβλεψης το οποίο δεν είναι αρκετά σύνθετο μπορεί να αποτύχει να μοντελοποιήσει επιτυχώς τα δεδομένα εκπαίδευσης και να μη μπορεί να προσαρμοστεί σε νέα δεδομένα, οδηγώντας σε ατελή μάθηση. Η ατελής μάθηση, ή αλλιώς υποπροσαρμογή (underfitting), αποτελεί ακριβώς το αντίθετο φαινόμενο από την υπερπροσαρμογή. Ένα τέτοιο μοντέλο μηχανικής μάθησης αναγνωρίζεται εύκολα λόγω των χαμηλών τιμών απόδοσης, τόσο στα δεδομένα εκπαίδευσης όσο και στα δεδομένα ελέγχου. Συχνά, το φαινόμενο αυτό συμβαίνει όταν τα δεδομένα για την εκπαίδευση ενός μοντέλου ακριβείας είναι πάρα πολύ λίγα, καθώς επίσης και όταν γίνεται προσπάθεια προσαρμογής γραμμικού μοντέλου σε μη γραμμικά δεδομένα.



Ατελής Μάθηση και Υπερμοντελοποίηση
 ΠΗΓΗ: <https://medium.com/greyatom/what-is-underfitting-and-overfitting-in-machine-learning-and-how-to-deal-with-it-6803a989c76>

1.2.2 Τεχνικές Μηχανικής Μάθησης

Η εκτέλεση μίας διαδικασίας Μηχανικής Μάθησης περιλαμβάνει τη δημιουργία ενός μοντέλου, το οποίο εκπαιδεύεται σε ορισμένα δεδομένα εκπαίδευσης και στη συνέχεια μπορεί να κάνει προβλέψεις για νέα δεδομένα. Στην υπάρχουσα βιβλιογραφία έχουν χρησιμοποιηθεί και διερευνηθεί διάφοροι τύποι μοντέλων για συστήματα μηχανικής μάθησης προκειμένου να αντιμετωπιστούν προκλήσεις όπως τα ανεπαρκή δεδομένα, τα δεδομένα κακής ποιότητας και τα δεδομένα με διαφορετικά μεταξύ τους χαρακτηριστικά. Στα επόμενα παρουσιάζονται τα προβλήματα που καλείται να αντιμετωπίσει η Μηχανική Μάθηση καθώς και τα διάφορα μοντέλα που χρησιμοποιούνται.

1.2.2.1 Παλινδρόμηση (Regression)

Τα μοντέλα παλινδρόμησης χρησιμοποιούνται στα προβλήματα επιτηρούμενης μάθησης όπου η μεταβλητή εξόδου λαμβάνει συνεχόμενες τιμές. Στόχος των μοντέλων είναι η εύρεση μιας συνάρτησης f , η οποία αντιστοιχίζει τα δεδομένα εισόδου x σε μία συνεχόμενη μεταβλητή

εξόδου y . Για παράδειγμα, με τη χρήση ενός μοντέλου παλινδρόμησης μπορεί να αναλυθεί η σχέση μεταξύ εισοδήματος (ή και άλλων παραγόντων) και χρόνου εργασίας.

1. Γραμμική Παλινδρόμηση (Linear Regression)

Η γραμμική παλινδρόμηση είναι μια από τις πιο γνωστές τεχνικές μοντελοποίησης, η οποία καθιερώνει μια σχέση μεταξύ εξαρτημένης μεταβλητής (Y) και μιας ή περισσότερων ανεξάρτητων μεταβλητών (X) μέσω μιας ευθείας γραμμής βέλτιστης προσαρμογής στα δεδομένα (επίσης γνωστής ως ευθεία παλινδρόμησης). Η συγκεκριμένη μέθοδος εστιάζει μόνο στη μέση τιμή της εξαρτημένης μεταβλητής και περιορίζεται σε γραμμικά προβλήματα.

Το μοντέλο της γραμμικής παλινδρόμησης εκφράζεται μαθηματικά από την παρακάτω εξίσωση:

$$Y_i = b_0 + b_1 X_i + \varepsilon_i, \quad i = 1, 2, \dots, n$$

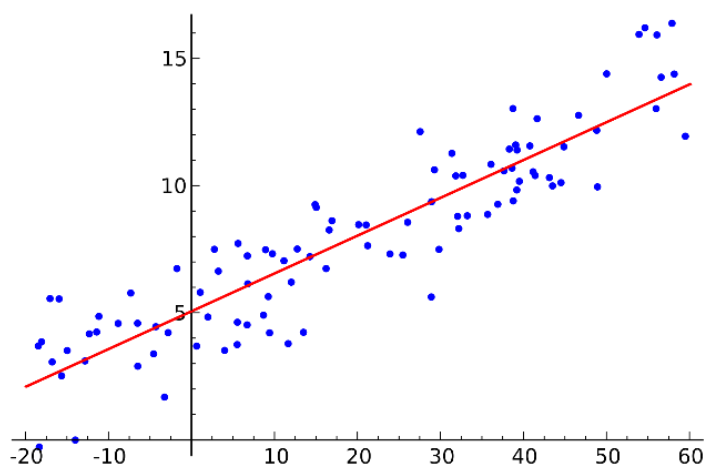
όπου b_0 , b_1 είναι δυο άγνωστες σταθερές (καλούνται και τεταγμένη ή intercept και κλίση ή slope, αντίστοιχα), ενώ οι $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ είναι ανεξάρτητες τυχαίες μεταβλητές που ακολουθούν την κανονική κατανομή και καλούνται σφάλματα των μετρήσεων.

Αυτή η εξίσωση μπορεί να χρησιμοποιηθεί για την πρόβλεψη της τιμής Y με βάση δεδομένες μεταβλητές πρόβλεψης. Η γραμμική παλινδρόμηση επηρεάζεται από ακραίες τιμές παρατηρήσεων και τα δεδομένα πρέπει να είναι ανεξάρτητα μεταξύ τους.

Για την εκτίμηση των συντελεστών ενός μοντέλου γραμμικής παλινδρόμησης χρησιμοποιείται είτε η μέθοδος της Σταδιακής Κατάβασης της Κλίσης (Gradient Descent), η οποία επιχειρεί να βρει τους συντελεστές του μοντέλου που ελαχιστοποιούν μία συγκεκριμένη συνάρτηση κόστους, είτε η μέθοδος των ελαχίστων τετραγώνων (OLS). Η μέθοδος αυτή είναι ίσως η παλαιότερη διαδικασία εκτίμησης παραμέτρων. Διατυπώθηκε για πρώτη φορά από τον Gauss, στις αρχές του 19^{ου} αιώνα. Η μέθοδος της Σταδιακής Κατάβασης είναι εντελώς διαφορετική: δεν παρέχει κλειστούς τύπους υπολογισμού των συντελεστών (όπως η κανονική εξίσωση στην μέθοδο των ελαχίστων τετραγώνων), αλλά βασίζεται στη διενέργεια επαναλήψεων για την εύρεση των κατάλληλων τιμών τους.

Όταν υπάρχουν περισσότερες από μία ανεξάρτητες μεταβλητές, η διαδικασία ονομάζεται πολλαπλή γραμμική παλινδρόμηση. Το μαθηματικό μοντέλο αποτελεί που χρησιμοποιείται αποτελεί γενίκευση του παραπάνω μοντέλου και έχει την εξής μορφή για παραμέτρους b_0, b_1, \dots, b_{p-1} .

$$Y = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_{p-1} X_{p-1} + \varepsilon$$



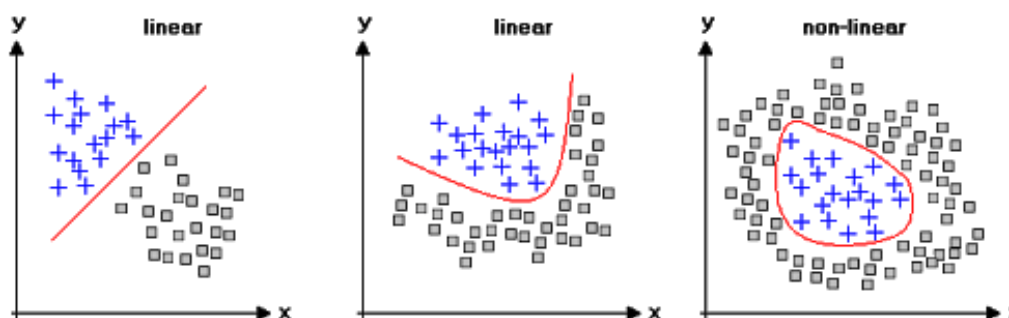
Παράδειγμα απλής γραμμικής παλινδρόμησης, η οποία έχει μια ανεξάρτητη μεταβλητή
 ΠΗΓΗ: https://el.wikipedia.org/wiki/Γραμμική_παλινδρόμηση

2. Μη Γραμμική Παλινδρόμηση (Non-Linear Regression)

Υπάρχουν όμως πολλά φαινόμενα που δεν μπορούν να περιγραφούν ικανοποιητικά από γραμμικά μοντέλα. Η μη γραμμική παλινδρόμηση αποτελεί έναν τύπο ανάλυσης παλινδρόμησης, στον οποίο οι παρατηρήσεις μοντελοποιούνται από μια συνάρτηση που είναι μη γραμμικός συνδυασμός των παραμέτρων του μοντέλου και εξαρτάται από μία ή περισσότερες ανεξάρτητες μεταβλητές.

Κάποια συνηθισμένα παραδείγματα συναρτήσεων που δεν είναι γραμμικές και χρησιμοποιούνται εκτεταμένα για την εκτίμηση των παραμέτρων είναι οι εκθετικές και οι λογαριθμικές συναρτήσεις. Στη μη γραμμική παλινδρόμηση μπορεί να συμπεριληφθεί οποιαδήποτε περίπτωση μη γραμμικής συνάρτησης, με την θεώρηση πως το μη γραμμικό μοντέλο εκφράζεται ως εξής:

$$Y_i = f\left(X_i, \gamma\right) + \varepsilon_i$$



Γραμμική και Μη Γραμμική Παλινδρόμηση

ΠΗΓΗ: http://www.statistics4u.com/fundstat_eng/cc_linvsnonlin.html

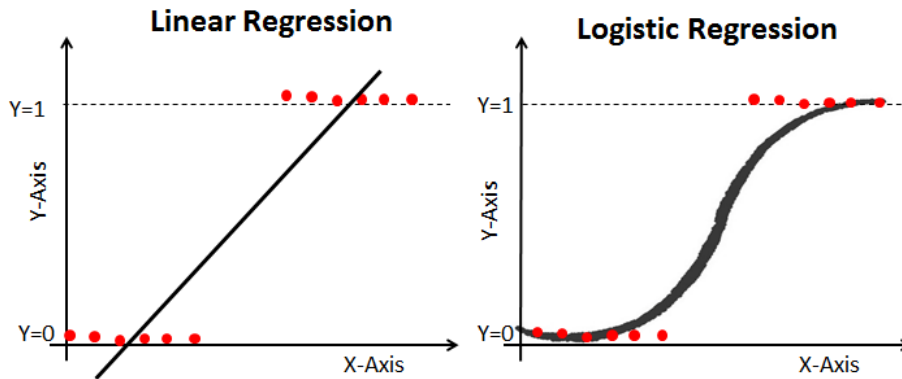
3. Λογιστική Παλινδρόμηση (Logistic regression)

Η λογιστική παλινδρόμηση (logistic regression) επεκτείνει τις τεχνικές της παλινδρόμησης σε περιπτώσεις όπου η έξοδος έχει διακριτή μορφή (χωρίζεται σε κατηγορίες). Αποτελεί στην ουσία ένα μοντέλο ταξινόμησης των τιμών μιας μεταβλητής απόκρισης Y με βάση τη θεωρία των πιθανοτήτων. Στο μοντέλο αυτό όπου η μεταβλητή Y συνήθως έχει δυαδικό χαρακτήρα (λαμβάνει δύο τιμές) επιδιώκεται η πρόβλεψη της έκβασης αυτής από ένα πλήθος προβλεπτικών μεταβλητών που μπορεί να είναι ονομαστικές, τακτικές ή ποσοτικές. Επομένως, χρησιμοποιείται για τη λύση του προβλήματος της Ταξινόμησης.

Η συνάρτηση που χρησιμοποιείται για τη λογιστική παλινδρόμηση είναι η σιγμοειδής συνάρτηση (λογιστική συνάρτηση), μια μαθηματική συνάρτηση η οποία έχει μορφή S και ονομάζεται επίσης και σιγμοειδής καμπύλη:

$$f(x) = \frac{1}{1 + e^{-x}}$$

Ενώ κατά την κλασική γραμμική παλινδρόμηση η εκτίμηση των παραμέτρων a και b_i γίνεται με τη μέθοδο των ελάχιστων τετραγώνων, κατά τη λογιστική παλινδρόμηση η εκτίμηση των παραμέτρων γίνεται με τη μέθοδο του λόγου πιθανοφάνειας (μέθοδος συνήθως εφαρμοζόμενη στα γενικευμένα γραμμικά υποδείγματα), δηλαδή επιλέγονται οι πιο πιθανοφανείς τιμές των παραμέτρων, προκειμένου να οδηγήσουν στα παρατηρούμενα αποτελέσματα. Αξίζει να σημειωθεί ότι πρώτος ο Ohlson, το 1980, χρησιμοποίησε το λογιστικό υπόδειγμα πιθανότητας (probit/logit model) σε έρευνά του σχετικά με την πρόβλεψη της πτώχευσης επιχειρήσεων.



Γραμμική και Λογιστική Παλινδρόμηση

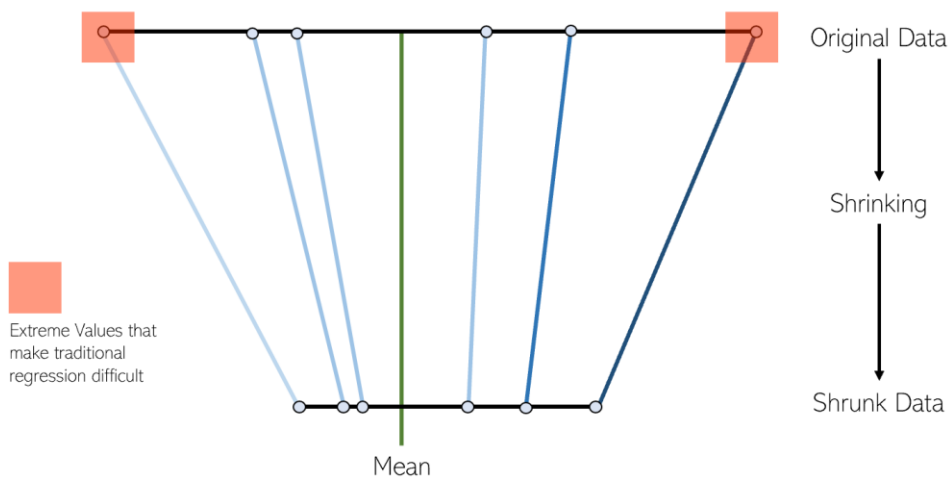
ΠΗΓΗ: <https://www.datacamp.com/community/tutorials/understanding-logistic-regression-python>

4. Παλινδρόμηση LASSO (Least Absolute Selection Shrinkage Operator Regression)

Ο στόχος της παλινδρόμησης LASSO είναι η απόκτηση του υποσυνόλου των προβλέψεων που ελαχιστοποιούν το σφάλμα πρόβλεψης για μια μεταβλητή ποσοτικής απόκρισης. Ο αλγόριθμος λειτουργεί επιβάλλοντας έναν περιορισμό στις παραμέτρους του μοντέλου που προκαλεί συστολή των συντελεστών παλινδρόμησης για ορισμένες μεταβλητές στο μηδέν.

Η μέθοδος, με τον τρόπο που ελαχιστοποιεί τη συνάρτηση κόστους, εκτός του ότι μειώνει τους συντελεστές είναι και ικανή να κάνει επιλογή χαρακτηριστικών (feature selection). Το παραπάνω καθίσταται εφικτό μέσω της κανονικοποιημένης παραμέτρου $\lambda \geq 0$, της οποίας η τιμή όσο πιο μεγάλη είναι τόσο περισσότερες εκτιμήσεις συντελεστών θα οδηγηθούν στο να έχουν τιμή ίση με το μηδέν. Αυτό πολύ απλά σημαίνει ότι συντελεστές που θα λάβουν τιμή ίση με το μηδέν καταδεικνύουν ανεξάρτητες μεταβλητές που δεν είναι σημαντικές για την πρόβλεψη της τιμής της εξαρτημένης μεταβλητής. Κατά συνέπεια, η παλινδρόμηση LASSO επιτρέπει την προσδιορισμό των σημαντικών χαρακτηριστικών (μεταβλητών) στο μοντέλο παλινδρόμησης.

Άλλη μία δημοφιλής μέθοδος είναι η παλινδρόμηση Ridge, η οποία καταφέρνει να μειώσει το πλήθος των συντελεστών, όσο πιο μεγάλη είναι η τιμή της παραμέτρου λ , δίχως ωστόσο να οδηγεί τους συντελεστές να λάβουν την τιμή μηδέν. Έτσι, αν και κατορθώνει να μειώσει τη διακύμανση (με τη μείωση των συντελεστών) δεν μπορεί να υποδείξει ποιες μεταβλητές είναι σημαντικές και ποιες όχι όπως η παλινδρόμηση LASSO.

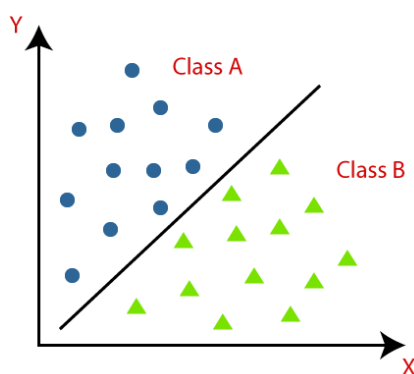


Απλοποιημένη απεικόνιση της διαδικασίας συρρίκνωσης / ελαχιστοποίησης
ΠΗΓΗ: <https://towardsdatascience.com/5-regression-algorithms-you-need-to-know-theory-implementation-37993382122d>

1.2.2.2 Ταξινόμηση (Classification)

Στην Μηχανική Μάθηση, η ταξινόμηση είναι μια επιβλεπόμενη μαθησιακή διαδικασία, η οποία μπορεί να θεωρηθεί ως μέσο κατηγοριοποίησης ορισμένων άγνωστων αντικειμένων σε ένα διακριτό σύνολο κλάσεων (σε δύο ή περισσότερες). Τα φίλτρα Spam είναι ένα παράδειγμα ταξινόμησης, όπου οι εισοδοί είναι τα emails ή άλλα μηνύματα και οι κλάσεις είναι "spam" και "όχι spam".

Η ταξινόμηση επιχειρεί να μάθει τη σχέση μεταξύ ενός συνόλου μεταβλητών και των χαρακτηριστικών τους από τα δεδομένα που εισάγονται στο μοντέλο και τα χρησιμοποιεί ώστε να μάθει να ταξινομή τις νέες παρατηρήσεις που θα δοθούν στο μοντέλο. Σε αντίθεση με την παλινδρόμηση, η μεταβλητή εξόδου της ταξινόμησης είναι μια κατηγορία, όχι μια τιμή, όπως "Πράσινο ή Μπλε", "φρούτα ή ζώα", κ.λπ. Δεδομένου ότι ο αλγόριθμος ταξινόμησης είναι μια τεχνική εποπτευόμενης μάθησης, λαμβάνει δεδομένα εισόδου σημασμένα (labeled), το οποίο σημαίνει πως περιέχει δεδομένα εισόδου με την αντίστοιχη έξοδο τους.



Ταξινόμηση Δεδομένων

ΠΗΓΗ: <https://www.javatpoint.com/classification-algorithm-in-machine-learning>

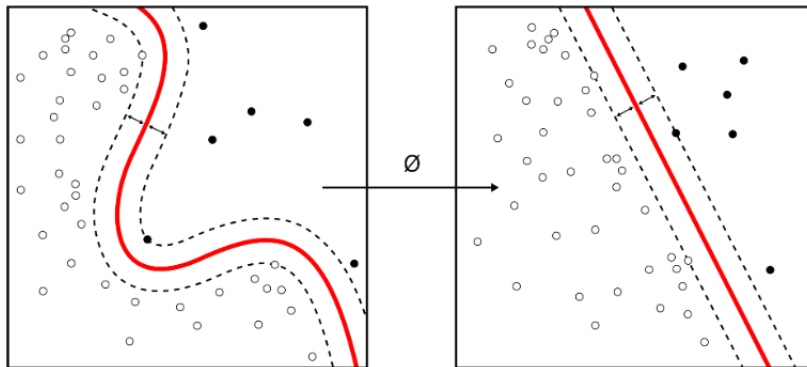
Στο πρόβλημα της ταξινόμησης υπάρχουν δύο τύποι μάθησης (learners):

1. **Μάθηση Lazy:** Αποθηκεύεται πρώτα το σύνολο των δεδομένων εκπαίδευσης (training dataset) χωρίς το σύστημα να μαθαίνει από αυτά, μέχρι να λάβει το σύνολο δεδομένων δοκιμής (test dataset). Στην περίπτωση αυτή, η ταξινόμηση γίνεται με βάση τα πιο σχετικά μεταξύ τους δεδομένα που αποθηκεύονται στο σύνολο των δεδομένων εκπαίδευσης, αφού ληφθούν και τα δεδομένα test. Η φάση της εκπαίδευσης διαρκεί λιγότερο χρόνο, αλλά το μοντέλο έχει μεγαλύτερη καθυστέρηση στο να εξάγει τις προβλέψεις. Παράδειγμα αποτελεί ο αλγόριθμος K-Nearest Neighbours (K-NN).
2. **Μάθηση Eager:** Αναπτύσσεται ένα μοντέλο ταξινόμησης το οποίο μαθαίνει μόλις λαμβάνει τα δεδομένα εκπαίδευσης, προτού λάβει το σύνολο δεδομένων δοκιμής. Επομένως, σε αντίθεση με τη μάθηση Lazy, η Eager χρειάζεται περισσότερο χρόνο για τη διαδικασία της εκπαίδευσης και λιγότερο χρόνο για την εξαγωγή της πρόβλεψης. Παραδείγματα τέτοιων αλγορίθμων είναι τα Δέντρα Απόφασης (Decision Trees), τα Naïve Bayes και τα Τεχνητά Νευρωνικά Δίκτυα.

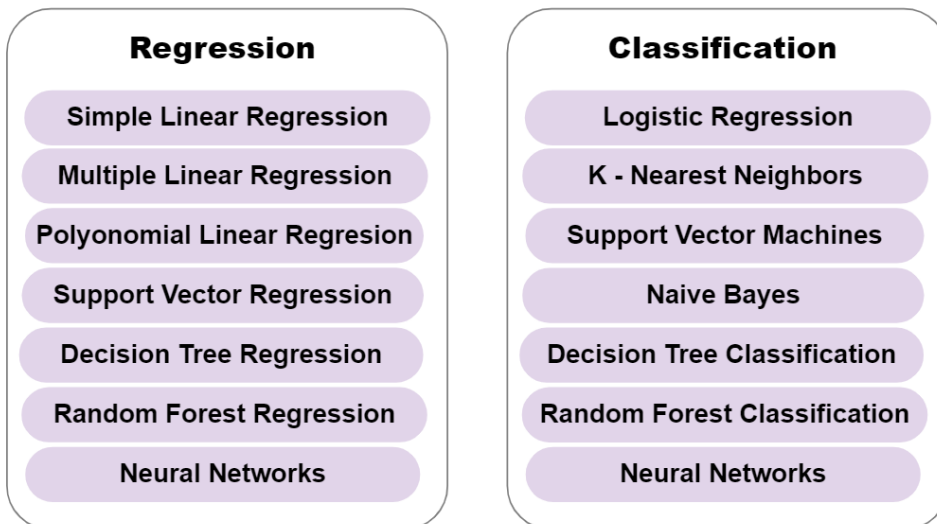
Υπάρχουν αρκετοί αλγόριθμοι στη μηχανική μάθηση που χρησιμοποιούνται προκειμένου να λυθεί το πρόβλημα της ταξινόμησης. Ανάλογα με το εάν τα χρησιμοποιούμενα μοντέλα είναι γραμμικά ή μη γραμμικά, αυτά διακρίνονται σε:

- **Γραμμικά Μοντέλα**
 - Λογιστική Παλινδρόμηση (Logistic Regression)
 - Μηχανές διανυσματικής στήριξης (Support Vector Machines)
- **Μη - Γραμμικά Μοντέλα**
 - K-Nearest Neighbours
 - Μηχανές Kernel (SVM Kernel machines)
 - Ταξινομητές Naïve Bayes

- Δέντρα Απόφασης (Decision Trees)
- Τεχνητά Νευρωνικά Δίκτυα (Artificial Neural Networks)



Μηχανές Kernel για τον υπολογισμό των μη - γραμμικά διαχωρίσιμων συναρτήσεων σε μια υψηλότερη διάσταση, όπου η συνάρτηση είναι γραμμικά διαχωρίσιμη.
ΠΗΓΗ: https://el.wikipedia.org/wiki/Στατιστική_ταξινόμηση



Παλινδρόμηση και Ταξινόμηση – Αλγόριθμοι

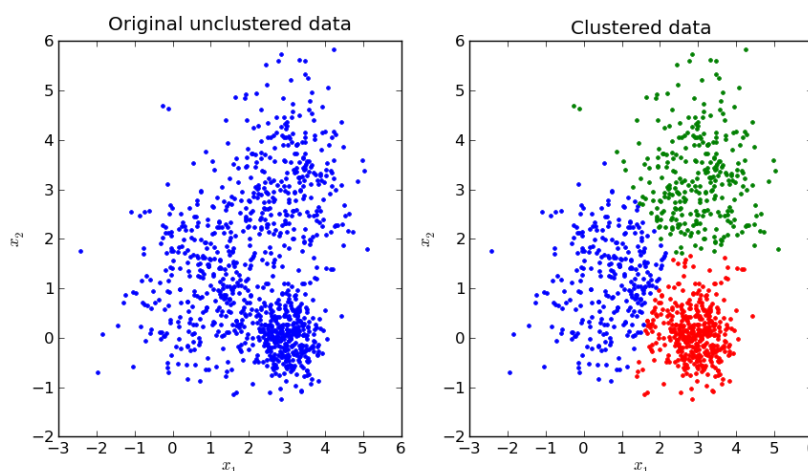
1.2.2.3 Ομαδοποίηση (Clustering)

Στην ομαδοποίηση ή αλλιώς συσταδοποίηση, ένα σύνολο δεδομένων εισόδου πρόκειται να χωριστεί σε ομάδες. Σε αντίθεση με την ταξινόμηση, οι ομάδες δεν είναι γνωστές εκ των προτέρων, καθιστώντας αυτόν τον διαχωρισμό τυπική εργασία μη επιτηρούμενης μάθησης. Η καταχώρηση αντικειμένων (δεδομένα, πρότυπα, διανύσματα) σε ίδια ομάδα μεταφράζεται ως ομοιότητα των αντικειμένων αυτών και αντίστροφα (αντικείμενα που ανήκουν σε διαφορετικές ομάδες είναι ανόμοια). Η ομοιότητα ή μη μεταξύ των αντικειμένων ουσιαστικά εξαρτάται από το συγκεκριμένο πρόβλημα και τη μορφή των δεδομένων.

Υπάρχει ένα μεγάλο πλήθος από αλγόριθμους συσταδοποίησης που έχουν προταθεί και ο καθένας τους βασίζεται σε διαφορετική φιλοσοφία. Σχεδόν όλοι τους δέχονται ένα σύνολο παραμέτρων που μπορεί να είναι το πλήθος των ομάδων, διανύσματα αρχικοποίησης που απαιτούν από τον αλγόριθμο κάποιες υποθέσεις για την πυκνότητα των διανυσμάτων στο χώρο και άλλες διάφορες παραμέτρους. Με διαφοροποίηση αυτών των παραμέτρων προκύπτει ένα σύνολο από αλγόριθμους σε κάθε βασική κατηγορία, όπως παρακάτω:

- Αλγόριθμοι k - Means

- Αλγόριθμοι k – Median
- Ιεραρχικοί Αλγόριθμοι Συσταδοποίησης (Hierarchical clustering algorithms)
- Συγκεντρωτικοί Αλγόριθμοι (Agglomerative algorithms)
- Διαχωριστικοί Αλγόριθμοι (Divisive algorithms)
- Αλγόριθμος DBSCAN (Density-Based Spatial Clustering of Applications with Noise)



Συσταδοποίηση Δεδομένων με k – Means αλγόριθμο

ΠΗΓΗ: <https://towardsdatascience.com/k-means-data-clustering-bce3335d2203>

1.2.2.4 Ανίχνευση «Ανωμαλιών» (Anomaly detection)

Η ανίχνευση ανωμαλιών χρησιμοποιείται για την ανίχνευση και αναγνώριση ανομοιογενών δεδομένων σε ένα σύνολο παρατηρήσεων, για παράδειγμα χρησιμοποιείται για τον εντοπισμό απάτης με πιστωτικές κάρτες. Ως ανωμαλίες ορίζονται οι παρατηρήσεις οι οποίες αποκλίνουν από κάποιο πρότυπο και δεν ακολουθούν το υπόλοιπο μοτίβο. Στην όραση υπολογιστών η ανίχνευση ανωμαλιών εντοπίζει ακραίες τιμές και ανωμαλίες σε ένα σετ εικόνων.

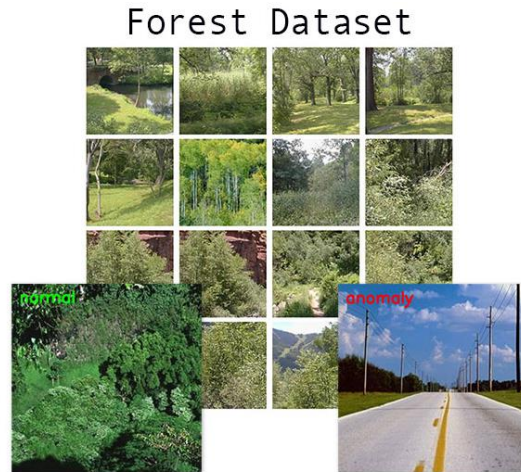
Οι αλγόριθμοι ανίχνευσης ανωμαλιών μπορούν να χωριστούν σε δύο υποκατηγορίες:

- **Ανίχνευση ακραίων τιμών (Outlier detection):**

Σε αυτή την κατηγορία το σύνολο δεδομένων εισόδου περιέχει παραδείγματα τόσο κανονικών δεδομένων όσο και ανομοιογενών δεδομένων. Οι αλγόριθμοι είναι μη εποπτευόμενης μάθησης και επιδιώκουν να ταιριάζουν τις περιοχές από τα δεδομένα εκπαίδευσης όπου τα κανονικά δεδομένα είναι πιο συγκεντρωμένα, απομονώνοντας τις περιοχές ανομοιογένειας. Στην ουσία ο αλγόριθμος ομαδοποιεί τα δεδομένα σε inliers (κανονικά δεδομένα) και outliers (ακραίες τιμές / θόρυβος). Συνήθως, γίνεται χρήση αυτών των μεθόδων για τον καθαρισμό και την προεπεξεργασία συνόλων δεδομένων πριν εφαρμοστούν πρόσθετες τεχνικές μηχανικής εκμάθησης.

- **Ανίχνευση Καινοτομίας (Novelty detection):**

Σε αντίθεση με την ανίχνευση ακραίων τιμών, η οποία περιλαμβάνει παραδείγματα κανονικών και παράταιρων δεδομένων, οι αλγόριθμοι ανίχνευσης καινοτομίας περιέχουν μόνο κανονικά δεδομένα κατά τη διάρκεια της εκπαίδευσης. Πρόκειται για μια μορφή εποπτευόμενης μάθησης, όπου εξαρτάται από τον αλγόριθμο ανίχνευσης καινοτομίας να προβλέψει εάν ένα δεδομένο είναι inlier ή outlier κατά τη διάρκεια της δοκιμής.

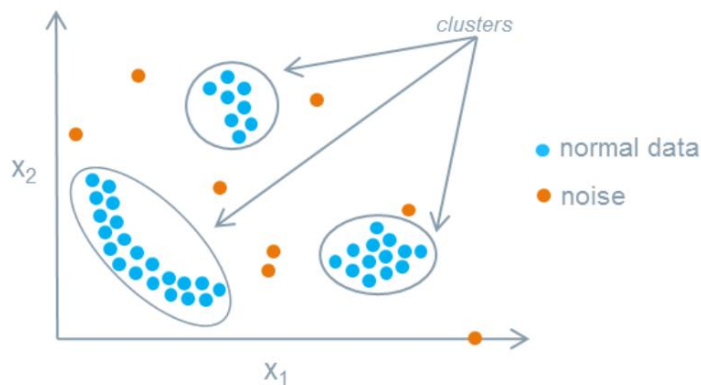


Δεδομένα εκπαίδευσης για την Ανίχνευση Ανωμαλιών

ΠΗΓΗ: <https://www.pyimagesearch.com/2020/01/20/intro-to-anomaly-detection-with-opencv-computer-vision-and-scikit-learn/>

Υπάρχουν αρκετοί αλγόριθμοι στη μηχανική μάθηση που χρησιμοποιούνται προκειμένου να επιτευχθεί η ανίχνευση των «ανωμαλιών». Παραδείγματα εποπτευόμενων μεθόδων περιλαμβάνουν την ανίχνευση ανομοιογενών δεδομένων, όπου χρησιμοποιούνται Νευρωνικά δίκτυα, δίκτυα Bayes, και τη μέθοδο K-πλησιέστερων γειτόνων (k-NN). Οι εποπτευόμενες μέθοδοι παρέχουν ένα καλύτερο ποσοστό ανίχνευσης «ανωμαλιών» χάρη στην ικανότητά τους να κωδικοποιούν οποιαδήποτε αλληλεξάρτηση μεταξύ μεταβλητών και να συμπεριλαμβάνουν προηγούμενα δεδομένα σε οποιοδήποτε προγνωστικό μοντέλο.

Οι μη εποπτευόμενες μέθοδοι ανίχνευσης «ανωμαλιών» δεν εξαρτώνται από δεδομένα εκπαίδευσης με χειροκίνητη επισήμανση. Αυτές οι μέθοδοι βασίζονται στη στατιστική παραδοχή ότι τα περισσότερα από τα δεδομένα εισροής είναι φυσιολογικά και μόνο ένα μικρό ποσοστό αποτελεί ακραία δεδομένα. Ορισμένες από τις μη επιβλεπόμενες μεθόδους είναι η μέθοδος K-Means, η προσέγγιση ομαδοποίησης με βάση την πυκνότητα (Density-based Clustering Approach - DBSCAN) και η Ανάλυση βάσει υποθέσεων (Hypothesis-based analysis).



Συσταδοποίηση για την Ανίχνευση Ανωμαλιών

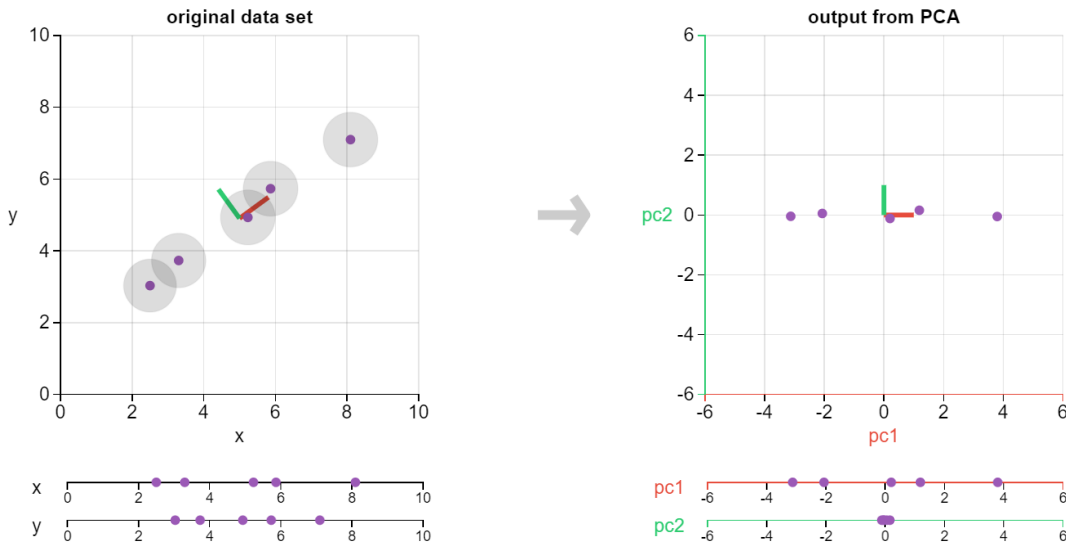
ΠΗΓΗ: <https://developer.mindsphere.io/apis/analytics-anomalydetection/api-anomalydetection-overview.html>

1.2.2.5 Ανάλυση Κύριων Συνιστωσών (PCA)

Η Ανάλυση Κύριων Συνιστωσών (Principle Component Analysis – PCA) είναι μια στατιστική διαδικασία, η οποία χρησιμοποιεί έναν ορθογώνιο μετασχηματισμό για να μετατρέψει ένα σύνολο παρατηρήσεων, πιθανώς συσχετισμένων μεταβλητών, σε ένα σύνολο από γραμμικά ανεξάρτητες μεταβλητές, οι οποίες ονομάζονται κύριες συνιστώσες. Ο αριθμός των κύριων συνιστωσών είναι μικρότερος ή ίσος από τον αρχικό αριθμό των μεταβλητών των

παρατηρήσεων. Ο μετασχηματισμός είναι ορισμένος με τέτοιο τρόπο, ώστε η πρώτη συνιστώσα να περιγράφει τη μέγιστη δυνατή διασπορά των αρχικών μεταβλητών και κάθε επόμενη συνιστώσα να εξηγεί επίσης με τη σειρά της τη μέγιστη δυνατή διασπορά υπό τον περιορισμό ότι είναι ορθογώνια με την προηγούμενή της. Τα παραγόμενα διανύσματα αποτελούν ένα ασυσχέτιστο ορθογώνιο σύνολο βάσης. Η ανάλυση κυρίων συνιστωσών είναι ευαίσθητη στην σχετική κλίμακα των αρχικών μεταβλητών.

Η παραπάνω μέθοδος αποτελεί κύρια τεχνική για την εξαγωγή χαρακτηριστικών και εγγυάται την εύρεση του καλύτερου γραμμικού μετασχηματισμού που μειώνει τον αριθμό των διαστάσεων με την ελάχιστη απώλεια πληροφοριών.



Παράδειγμα 2D PCA

ΠΗΓΗ: <https://towardsdatascience.com/a-one-stop-shop-for-principal-component-analysis-5582fb7e0a9c>

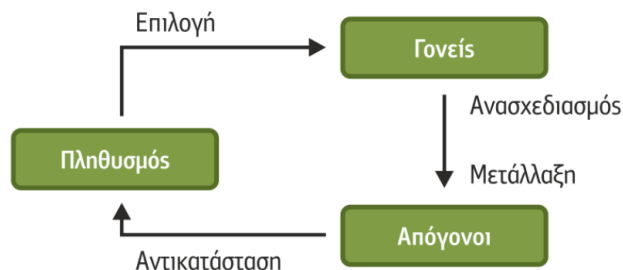
1.2.2.6 Γενετικοί αλγόριθμοι (Genetic algorithms)

Ένας γενετικός αλγόριθμος (GA) ακολουθεί μια στρατηγική αναζήτησης (heuristic search), η οποία αναπτύσσεται με Εξελικτική Μάθηση (Evolutionary Learning) και μιμείται τη διαδικασία της φυσικής επιλογής χρησιμοποιώντας μεθόδους όπως αυτή της μετάλλαξης και της διασταύρωσης, προκειμένου να δημιουργήσει καινούρια γονότυπα με την ελπίδα εύρεσης αποτελεσματικών λύσεων σε ένα συγκεκριμένο πρόβλημα. Στη Μηχανική Μάθηση, γενετικοί αλγόριθμοι χρησιμοποιήθηκαν τη δεκαετία του 1980 και του 1990. Αντίστροφα, τεχνικές μηχανικής μάθησης έχουν χρησιμοποιηθεί για την βελτίωση της απόδοσης γενετικών και εξελικτικών αλγορίθμων.

Οι αλγόριθμοι αυτοί αποτελούν ένα νέο τρόπο προγραμματισμού, κατάλληλο για επίλυση προβλημάτων σε χώρους αναζήτησης ιδιαίτερα πολύπλοκους και προσαρμόσιμο σε ποικίλα και ευμετάβλητα περιβάλλοντα. Περιλαμβάνουν ένα σύνολο μεθόδων εμπνευσμένων από τη θεωρία της εξέλιξης. Αρχίζουν από έναν συνήθως επιλεγμένο πληθυσμό και επιδιώκουν την παραγωγή ατόμων με υψηλή απόδοση. Η λειτουργία των γενετικών αλγορίθμων βασίζεται στη δυνατότητά τους να αναπαράγουν τον πληθυσμό σε γενεές. Κάθε γενεά προκύπτει από την προηγούμενη, αφού επιλεγούν οι γονείς και εφαρμοστούν οι γενετικοί τελεστές της διασταύρωσης και της μετάλλαξης. Οι είσοδοι και οι έξοδοι του προβλήματος αναπαρίστανται ως συμβολοσειρές (strings) σταθερού μήκους, ενός συγκεκριμένου αλφαβήτου (συνήθως δυαδικού), με ρόλο αντίστοιχο του χρωμοσώματος στη Γενετική. Κάθε σημείο του χώρου επίλυσης του προβλήματος αντιπροσωπεύεται από μια συμβολοσειρά ή δομή. Απαιτείται η δυνατότητα πληροφόρησης για την ικανότητα των διάφορων συμβολοσειρών του πληθυσμού, σύμφωνα με τη συνάρτηση καταλληλότητας (fitness function), που καλείται επίσης και συνάρτηση απόδοσης, συνάρτηση προσαρμογής ή αντικειμενική συνάρτηση. Ο

χρόνος στην εξέλιξη της διαδικασίας υπολογίζεται σε διακριτά διαστήματα τα οποία καλούνται γενεές, παράδειγμα $N(t)$: ο πληθυσμός κατά τη γενεά t .

Τυπικά, ένας εξελικτικός αλγόριθμος αρχικοποιεί τον πληθυσμό του σε τυχαίες τιμές, αν και μπορεί να χρησιμοποιηθεί προηγούμενη γνώση του πεδίου εφαρμογής (εάν υπάρχει), για να επιδράσει στην αρχικοποίηση του πληθυσμού. Ακολουθεί η αξιολόγηση του πληθυσμού με απόδοση αντίστοιχων τιμών ποιότητας σε κάθε άτομο του πληθυσμού στο συγκεκριμένο περιβάλλον. Η αξιολόγηση γίνεται μέσω της συνάρτησης ποιότητας (αντιπροσωπευτικής του συγκεκριμένου περιβάλλοντος), η οποία μπορεί να είναι πολύ απλή, όπως ο υπολογισμός μιας απλής συνάρτησης, ή ιδιαίτερα πολύπλοκη, όπως η εκτέλεση μιας πολύπλοκης προσομοίωσης. Η επιλογή συνήθως υλοποιείται σε δύο βήματα: επιλογή γονέων και επιβίωση γονέων. Κατά την επιλογή των γονέων καθορίζεται ποια άτομα θα γίνουν γονείς και πόσους απογόνους/παιδιά (offsprings/children) θα αποκτήσουν. Οι απόγονοι δημιουργούνται μέσω ανασυνδυασμού των γονέων, δηλαδή με την ανταλλαγή πληροφορίας μεταξύ των γονέων και μέσω μετάλλαξης η οποία διαταράσσει περαιτέρω τους απογόνους. Ακολουθεί η χρήση της συνάρτησης ποιότητας για την αξιολόγηση των απογόνων και τελικά η επιλογή των ατόμων του πληθυσμού που θα επιβιώσουν στην επόμενη γενιά. Η διαδικασία αυτή ονομάζεται εξελικτικός κύκλος (evolutionary cycle).



Ο εξελικτικός κύκλος των Γενετικών Αλγορίθμων
ΠΗΓΗ: Γεωργούλη (2015) [50]

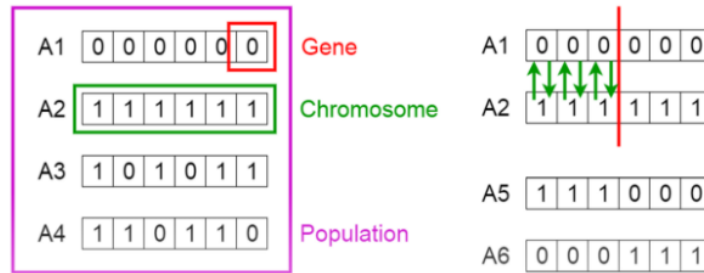
Επομένως, τρία είναι τα βασικά λειτουργικά τμήματα ενός γενετικού αλγορίθμου:

1. Τμήμα πληθυσμού, με τις δομές και τεχνικές για τη δημιουργία και τη διαχείρισή του.
2. Τμήμα εκτίμησης της απόδοσης, με τη συνάρτηση καταλληλότητας.
3. Τμήμα αναπαραγωγής, με τεχνικές για τη δημιουργία νέων δομών με την εφαρμογή των γενετικών τελεστών.

Στα παραπάνω, οι βασικοί μηχανισμοί που συνδέουν τον ΓΑ με το πρόβλημα που επιλύει είναι ο τρόπος κωδικοποίησης της δομής που αναπαριστά μια λύση του προβλήματος, καθώς και η συνάρτηση καταλληλότητάς που επιστρέφει τη μέτρηση της απόδοσης της δομής. Επιπλέον, ο ΓΑ απαιτεί να οριστεί ένα όριο καταλληλότητας, το οποίο θα αντιστοιχεί στην επιθυμητή τιμή απόδοσης που πρέπει να επιτευχθεί από μια δομή, ώστε να τερματιστεί ο ΓΑ, και 3 επιπλέον παράμετροι:

1. το πλήθος N των δομών του αρχικού πληθυσμού,
2. το ποσοστό r του πληθυσμού που θα αναπαραχθεί σε μια γενεά,
3. το ποσοστό m των απογόνων που θα υποστούν μετάλλαξη.

Οι μηχανισμοί λειτουργίας του αλγορίθμου ποικίλλουν ως προς την εφαρμογή και τα αποτελέσματα. Στο βιβλίο του «Genetic Algorithms in Search, Optimization and Machine Learning» (1989), ο Goldberg περιγράφει τι κάνει τους γενετικούς αλγορίθμους να λειτουργούν και εισάγει τον απλό γενετικό αλγόριθμο ή κανονικό γενετικό αλγόριθμο (canonical genetic algorithm): έναν εξελικτικό αλγόριθμο που εκτελεί ενέργειες σε μια συμβολοσειρά δυαδικών ψηφίων με τους ακόλουθους τρεις γενετικούς τελεστές: αναπαραγωγή, διασταύρωση, μετάλλαξη (Γεωργούλη (2015) [50]).



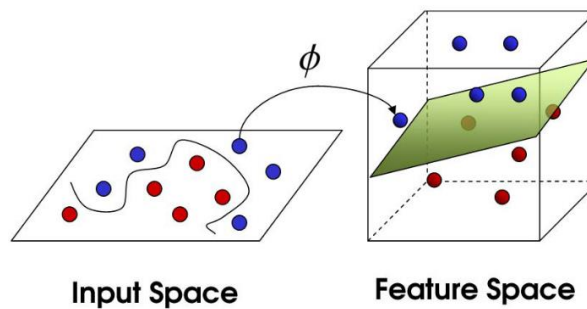
Γενετικοί Αλγόριθμοι ΠΗΓΗ: <https://towardsdatascience.com/introduction-to-genetic-algorithms-including-example-code-e396e98d8bf3>

1.2.3 Μηχανές Διανυσματικής Στήριξης

Η πρώτη αναφορά στις Μηχανές Διανυσματικής Στήριξης (Support Vector Machines – SVMs) πραγματοποιήθηκε το 1979 από τον Vapnik και τους συνεργάτες του, η ευρεία διάδοση και χρήση τους ξεκίνησε όμως μετά το 1995 [83]. Πρόκειται για μία ομάδα αλγορίθμων επιτηρούμενης μάθησης, οι οποίοι αρχικά χρησιμοποιήθηκαν για τη λύση του προβλήματος της κατηγοριοποίησης δεδομένων, ενώ αργότερα εφαρμόστηκαν και σε προβλήματα παλινδρόμησης. Συγκριτικά με άλλες παραδοσιακές μεθόδους ταξινόμησης, η μέθοδος SVM ταξινομητών παρουσίασε μεγάλη ικανότητα καθολικής εφαρμογής.

Η βασική ιδέα της κατασκευής τους στηρίζεται στην αρχή ελαχιστοποίησης του Δομικού Διαρθρωτικού Κινδύνου (Structural Risk Minimization - SRM) και στη διάσταση VC κατά Vapnik Chervonenkis (VC) [83].

Η κατηγοριοποίηση των δεδομένων με Μηχανές Διανυσματικής Στήριξης, εάν δεν μπορεί να λυθεί γραμμικά, στηρίζεται στη μεταφορά των δεδομένων σε έναν διανυσματικό χώρο περισσότερων διαστάσεων, στο «χώρο χαρακτηριστικών», στον οποίο τα δεδομένα είναι γραμμικά διαχωριζόμενα και άρα το πρόβλημα μπορεί να λυθεί γραμμικά. Στον «χώρο χαρακτηριστικών» γίνεται ο υπολογισμός για την εύρεση ενός βέλτιστου υπερεπιπέδου (hyperplane), το οποίο παίζει τον ρόλο επιφάνειας λήψης απόφασης, διαχωρίζοντας τα δεδομένα και δημιουργώντας το μέγιστο περιθώριο διαχωρισμού (margins) των δύο κλάσεων.



Απεικόνιση στον χώρο των χαρακτηριστικών

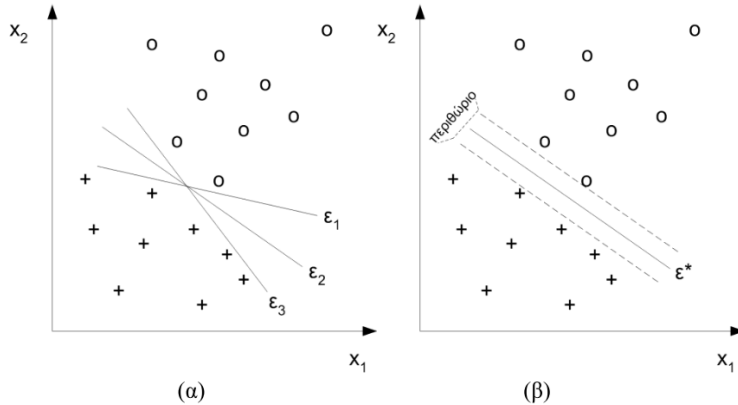
ΠΗΓΗ: <https://towardsdatascience.com/the-kernel-trick-c98cdbcaeb3f>

Ουσιαστικά, οι ταξινομητές SVM χρησιμοποιούν τις πιο ακραίες περιπτώσεις ανάμεσα στις δύο κλάσεις έτσι ώστε να δημιουργηθεί το μοντέλο. Έστω, για παράδειγμα, ότι σε μία εικόνα υπάρχει ένα σήμα οδικής κυκλοφορίας και στόχος είναι να αναγνωριστεί αν πρόκειται για το σήμα STOP ή για το σήμα ορίου ταχύτητας. Κατά την προσαρμογή των δεδομένων, ο αλγόριθμος SVM χρησιμοποιεί τα σήματα STOP που μοιάζουν πολύ με τα σήματα ορίου ταχύτητας και τα σήματα ορίου ταχύτητας που μοιάζουν πολύ με STOP για να ορίσει τελικά το μέγιστο περιθώριο διαχωρισμού μεταξύ των δύο κλάσεων.

Τα σημεία αυτά που στο παράδειγμα βρίσκονται πάνω στα όρια του περιθωρίου, ονομάζονται διανύσματα υποστήριξης (Support Vectors, SV) και είναι εκείνα που ταξινομούνται δυσκολότερα. Αποτελούν τα κρίσιμότερα για την ταξινόμηση στοιχεία του δείγματος

εκπαίδευσης και εάν τα SV μετακινηθούν θα αλλάξουν την θέση του διαχωριστικού υπερεπιπέδου.

Η μεταφορά από τον ένα χώρο στον άλλο επιτυγχάνεται μέσω συναρτήσεων πυρήνων (kernels), συναρτήσεις που δίνουν το εσωτερικό γινόμενο στο «χώρο χαρακτηριστικών» εκτελώντας υπολογισμούς στο «χώρο δεδομένων». Κάθε SVM είναι ένας δυαδικός ταξινομητής, έχει δηλαδή τη δυνατότητα κατηγοριοποίησης σε δύο κλάσεις. Η θεωρία της μεθόδου ταξινομητών SVM επεκτάθηκε στην πορεία, ώστε να μπορεί να κατηγοριοποιεί δεδομένα σε πολλαπλές κλάσεις (multiclass classification). Η μέθοδος των SVM στηρίζεται στο γεγονός ότι στο πρόβλημα βελτιστοποίησης τα δεδομένα εκπαίδευσης «συμμετέχουν» μόνο με τη μορφή εσωτερικών γινομένων.



Πρόβλημα δύο γραμμικώς διαχωρίσιμων κατηγοριών: (α) πολλαπλές ενδεχόμενες επιφάνειες απόφασης $\epsilon_1, \epsilon_2, \epsilon_3$ κ.λπ. και (β) η βέλτιστη επιφάνεια απόφασης ϵ^*
 ΠΗΓΗ: Καμπουρλάζος (2015) [84]

1.2.3.1 Πραγματικός, εμπειρικός και διαρθρωτικός κίνδυνος

Τα δεδομένα εκπαίδευσης αποτελούνται από ένα σύνολο ιδιοτήτων (attributes) και από αποκρίσεις (labels), που καθορίζουν την τάξη στην οποία ανήκουν. Σε ένα πρόβλημα δυαδικής ταξινόμησης όπου οι δύο πιθανές αποκρίσεις λαμβάνουν τιμές $\{-1, +1\}$ ενός συνόλου συναρτήσεων:

$$\{f_a(x) \in \Lambda\} \text{ με } f_a: \mathbb{R}^N \rightarrow \{-1, +1\}$$

όπου Λ ένα σύνολο αυθαίρετων παραμέτρων, $x \in \mathbb{R}^N$ και y η απόκριση του κάθε διανύσματος ιδιοτήτων και σύνολο δειγμάτων άγνωστης κατανομής πιθανότητας $P(x, y)$, η οποία καθορίζει την πιθανότητα σωστής ταξινόμησης του παραδείγματος. Οι συναρτήσεις f_a καλούνται υπόθεση και το σύνολο $\{f_a(x) \in \Lambda\}$ καλείται χώρος της υπόθεσης H .

Ο αναμενόμενος κίνδυνος αναφέρεται στις εσφαλμένες ταξινομήσεις των νέων δειγμάτων, ουσιαστικά ελέγχει πόσο καλή είναι η υπόθεση για την πρόβλεψη του σωστού y για το σημείο x , και ορίζεται ως:

$$R(a) = \int \frac{1}{2} |f_a(x) - y| P(x, y) dx dy$$

Προκειμένου να ελαχιστοποιηθεί ο αναμενόμενος κίνδυνος, εφόσον δίνονται τα δείγματα της άγνωστης κατανομής πιθανότητας $P(x, y)$, μπορεί να υπολογιστεί μια στοχαστική προσέγγιση η οποία καλείται εμπειρικός κίνδυνος. Ο αναμενόμενος κίνδυνος αναφέρεται στο σφάλμα της ήδη εκπαιδευμένης μηχανής σε νέα δείγματα, ενώ ο εμπειρικός κίνδυνος μόνο στα δείγματα του συνόλου εκπαίδευσης. Ουσιαστικά, ο εμπειρικός κίνδυνος ταυτίζεται με τον μέσο αριθμό λαθών που προκύπτουν κατά τη διαδικασία εκπαίδευσης και προκύπτει πάντα ο ίδιος για συγκεκριμένο σύνολο δεδομένων εκπαίδευσης και παραμέτρων a .

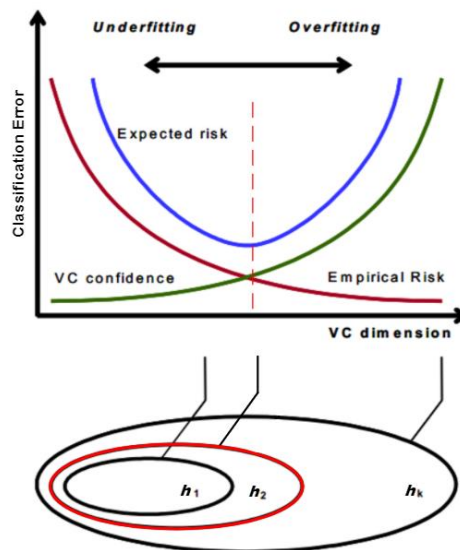
$$R_{emp}(a) = \frac{1}{2l} \sum_{i=1}^l |f_a(x_i) - y_i|$$

Η σχέση που συνδέει τον αναμενόμενο με τον εμπειρικό κίνδυνο είναι η παρακάτω ανισοσύτητα. Όπως έχει αναφερθεί, στόχος είναι ο αναμενόμενος κίνδυνος να είναι όσο το δυνατόν μικρότερος και για να επιτευχθεί αυτό πρέπει ο εμπειρικός κίνδυνος, όσο και ο λόγος της διάστασης VC (h) να είναι όσο το δυνατόν μικρότερα.

$$R(a) \leq R_{emp}(a) + \sqrt{\frac{h \left(\log \frac{2l}{h} + 1 \right) - \log \frac{\eta}{4}}{l}}$$

Η ανισότητα είναι ένας τρόπος να γίνει η εκτίμηση του σφάλματος της μηχανής σε νέα δείγματα με χρήση μόνο του σφάλματος εκπαίδευσης που προκύπτει από τα υπάρχοντα δείγματα εκπαίδευσης και τη διάσταση VC. Η σχέση αυτή ορίζει τον διαρθρωτικό κίνδυνο, που οφείλει να ελαχιστοποιείται σε μία μηχανή διανυσματικής στήριξης.

Προκειμένου να ελαχιστοποιηθεί ο διαρθρωτικός κίνδυνος είναι απαραίτητο να βρεθεί μία ισορροπία μεταξύ του εμπειρικού κινδύνου και της διάστασης VC. Για τον σκοπό αυτό δημιουργούνται διάφορες υποομάδες δεδομένων εκπαίδευσης, με αυξανόμενη διαδοχικά τη διάσταση VC. Για κάθε υποομάδα εκπαιδεύεται ένα σύστημα ώστε να ελαχιστοποιηθεί το άθροισμα του ρυθμού λάθους εκπαίδευσης.



Ελαχιστοποίηση διαρθρωτικού κινδύνου
 ΠΗΓΗ: <https://slideplayer.com/slide/5206673/>

1.2.3.2 Διάσταση VC

Μια άλλη πολύ σημαντική ιδιότητα ενός ταξινομητή είναι η πολυπλοκότητά του η οποία εκφράζεται με την έννοια της ικανότητας (capacity) ως εξής: πολύπλοκοι ταξινομητές, δηλαδή ταξινομητές με πολλές ανεξάρτητες παραμέτρους, θεωρείται ότι έχουν μεγάλη ικανότητα. Ένα διαδεδομένο μέτρο της ικανότητας ενός ταξινομητή είναι η διάσταση Vapnik Chervonenkis (VC) [83].

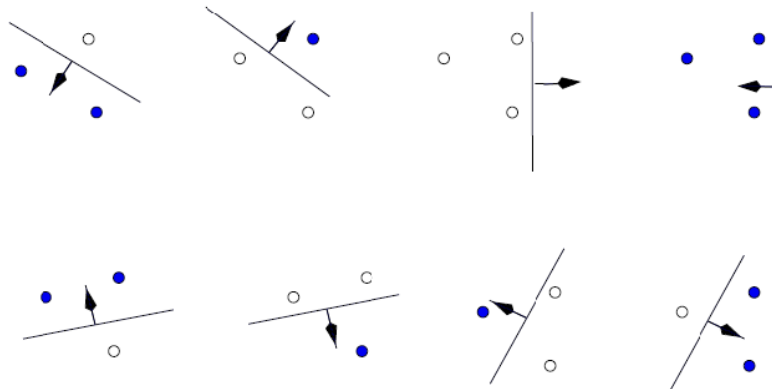
Η διάσταση VC αποτελεί χαρακτηριστική ιδιότητα μιας οικογένειας συναρτήσεων $\{f_a(x)\}$, όπου x είναι ένα σύνολο παραμέτρων. Έστω συναρτήσεις οι οποίες αναφέρονται στην περίπτωση ταξινόμησης σε δύο κλάσεις, ισχύει:

$$f_a(x) \in \{-1, +1\} \forall x, a$$

Ένα δεδομένο σύνολο n σημείων μπορεί να χαρακτηριστεί με 2^n τρόπους. Εάν για κάθε χαρακτηρισμό μπορεί να βρεθεί ένα μέλος της οικογένειας $\{f_a(x)\}$ το οποίο να καθορίζει τον χαρακτηρισμό αυτό, δηλαδή η $\{f_a(x)\}$ να διαχωρίζει το δεδομένο σύνολο με όλους τους δυνατούς τρόπους, τότε: η διάσταση VC για μια οικογένεια συναρτήσεων $\{f_a(x)\}$ ορίζεται ως ο μέγιστος αριθμός των δεδομένων εκπαίδευσης ο οποίος μπορεί να διαχωριστεί από την $\{f_a(x)\}$

Έρευνες έχουν αποδείξει ότι η δυνατότητα γενίκευσης ενός ταξινομητή είναι αντιστρόφως ανάλογη της πολυπλοκότητας και ανάλογη της προσαρμογής του ταξινομητή στα δεδομένα εκπαίδευσης. Συνεπώς, κατά την ανάπτυξη ενός μοντέλου ταξινόμησης είναι απαραίτητο να εξεταστούν τα δύο αυτά στοιχεία (πολυπλοκότητα - προσαρμογή) σε ένα ενιαίο πλαίσιο. Στη φιλοσοφία αυτή βασίζεται η μεθοδολογία των μηχανών διανυσμάτων υποστήριξης (SVM), στην οποία η έννοια της πολυπλοκότητας περιγράφεται από τη διάσταση VC και την άμεση σχέση που έχει με το περιθώριο διάκρισης των κατηγοριών.

Η διάσταση VC είναι ενδεικτική της ικανότητας μάθησης του συστήματος. Η ικανότητα μάθησης πρέπει να επιλεγεί με τρόπον ώστε να μην είναι πολύ μεγάλη αλλά ούτε πολύ μικρή, καθώς και στις δύο περιπτώσεις δημιουργούνται προβλήματα στη γενίκευση.



Δυνατοί διαχωρισμοί 3 σημείων στο 2-στατο χώρο από προσανατολισμένες ευθείες γραμμές.
ΠΗΓΗ: Burges (1998) [85]

1.2.3.3 Γραμμικά και μη γραμμικά διαχωρίσιμα δεδομένα

Αρχικά, για την κατανόηση της λειτουργίας των ταξινομητών SVM γίνεται η υπόθεση πως το σύνολο των δεδομένων είναι γραμμικά διαχωρίσιμο (Hard Margin). Ανάλογα με το χώρο στον οποίο βρίσκονται οι παρατηρήσεις, τα δεδομένα εκπαίδευσης μπορούν να διαχωρίζονται:

- από ένα σημείο στον μονοδιάστατο χώρο
- από μία ευθεία γραμμή στο δισδιάστατο χώρο
- από ένα επίπεδο στον τρισδιάστατο χώρο
- από ένα υπερεπίπεδο (hyperplane) σε μεγαλύτερες διαστάσεις

Στο δισδιάστατο χώρο δεδομένων είναι φανερό πως το δείγμα εκπαίδευσης είναι δυνατόν να διαχωριστεί από διάφορες ευθείες. Η ευθεία ή το υπερεπίπεδο που πετυχαίνει την ασφαλέστερη κατηγοριοποίηση είναι αυτή που απέχει περισσότερο από τα στοιχεία των δύο κλάσεων. Το περιθώριο ενός υπερεπιπέδου σε ένα οποιαδήποτε πρόβλημα κατηγοριοποίησης ορίζεται να είναι το διπλάσιο της ελάχιστης απόστασης της κοντινότερης παρατήρησης από την ευθεία ή το υπερεπίπεδο. Εξ ορισμού, στο χώρο που καταλαμβάνεται από το περιθώριο δεν υπάρχουν παρατηρήσεις από καμία κατηγορία. Λαμβάνοντας υπόψη τα παραπάνω το βέλτιστο υπερεπίπεδο ορίζεται ως το υπερεπίπεδο εκείνο που ορίζει το μέγιστο περιθώριο. Άρα, το πρόβλημα που πρέπει να αντιμετωπιστεί είναι η εύρεση της ευθείας ή του υπερεπιπέδου που πετυχαίνει τον βέλτιστο διαχωρισμό.

Ένα τέτοιο υπερεπίπεδο περιγράφεται από την εξίσωση:

$$w^T x + b = 0$$

Όπου, $w = (w_1, w_2, \dots, w_n)$ το διάνυσμα βαρών που είναι κάθετο στο υπερεπίπεδο, $x = (x_1, x_2, \dots, x_n)$ το διάνυσμα εισόδου και b το κατώφλι.

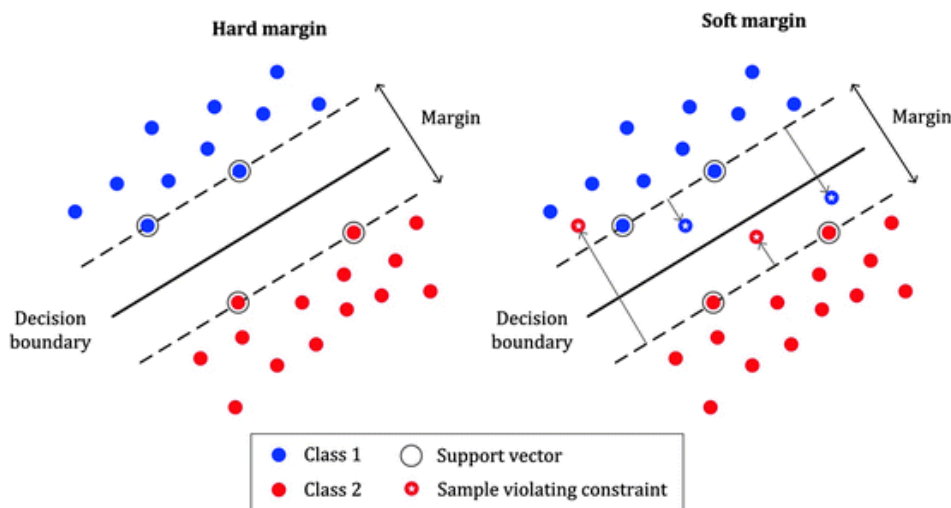
Συχνά τα δεδομένα δεν είναι γραμμικά διαχωρίσιμα στον «χώρο δεδομένων» που έχουν τις ιδιότητες τους. Οι μηχανές διανυσμάτων στήριξης λειτουργούν όμως και σε αυτές τις περιπτώσεις μέσω προβολής των διανυσμάτων των ιδιοτήτων στον «χώρο χαρακτηριστικών». Το 1965, ο Cover επισημαίνει ότι:

«Ένα μη-γραμμικώς διαχωρίσιμο πρόβλημα αναγνώρισης προτύπων μπορεί να μετασχηματιστεί σε γραμμικώς διαχωρίσιμο σε ένα χώρο περισσότερων διαστάσεων.»

Εάν $\Phi(x)$ η συνάρτηση που προβάλλει τα διανύσματα και K μία συνάρτηση πυρήνα, το υπερεπίπεδο που διαχωρίζει τα δεδομένα στον «χώρο χαρακτηριστικών» εκφράζεται ως:

$$w = \sum_{i=1}^{Ns} a_i y_i \Phi(s_i) \Phi(x) + b = \sum_{i=1}^{Ns} a_i y_i K(s_i, x) + b$$

Στην προηγούμενη περίπτωση (Hard Margin) του «τέλειου» γραμμικού διαχωρισμού, με τη χρήση αυστηρότερων παραμέτρων στο μοντέλο, εξασφαλιζόταν η σωστή κατηγοριοποίηση όλων των παρατηρήσεων. Σε πολλές εφαρμογές τα αντικείμενα δεν είναι γραμμικώς διαχωρίσιμα ή ακόμη και αν είναι, συχνά προτιμάται η επιλογή ενός μεγαλύτερου περιθωρίου επιτρέποντας σε κάποιες παρατηρήσεις να βρίσκονται μέσα σε αυτό ή ακόμη και να ταξινομηθούν λανθασμένα, προκειμένου να μπορέσει το μοντέλο να γενικεύσει σωστά αποφεύγοντας την υπερμοντελοποίηση (overfitting). Παρακάτω, στην αριστερή εικόνα, στο ίδιο σύνολο δεδομένων που ανήκουν σε δύο κλάσεις εφαρμόζεται ένας ταξινομητής μικρού περιθωρίου που δεν κάνει λάθη, ενώ στη δεξιά εικόνα ο κατηγοριοποιητής ταξινομεί λάθος ένα στοιχείο αλλά δημιουργεί μεγαλύτερο περιθώριο. Τέτοια μοντέλα ταξινομητών SVM που επιτρέπουν λάθη προκειμένου να μεγιστοποιηθεί το περιθώριο ονομάζονται ταξινομητές SVM χαλαρού περιθωρίου (Soft Margin).



Διαχωρισμένα δεδομένα χαλαρού και μη περιθωρίου

ΠΗΓΗ: <https://medium.com/@ankitnitjsr13/math-behind-svm-support-vector-machine-864e58977fdb>

1.2.3.4 Συναρτήσεις Πυρήνα

Όπως έχει ήδη αναφερθεί, οι συναρτήσεις πυρήνα (kernel functions) είναι απεικονίσεις των διανυσμάτων εισόδου x στο σύνολο \mathbf{R} , οι οποίες έχουν συγκεκριμένη μορφή και ιδιότητες και γενικεύουν σε μεγάλο βαθμό τις εφαρμογές των αλγορίθμων ταξινόμησης. Πρόκειται δηλαδή

για συναρτήσεις που δίνουν το εσωτερικό γινόμενο στον «χώρο χαρακτηριστικών» εκτελώντας υπολογισμούς στο «χώρο δεδομένων». Σύμφωνα με τη μέθοδο των SVMs, μετασχηματίζονται κατάλληλα τα διανύσματα εισόδου ώστε να επιτευχθεί η πιο καθολική λύση του προβλήματος της ταξινόμησης. Οι πιο συχνές συναρτήσεις πυρήνα που απαντώνται στην πράξη είναι:

1. Γραμμική συνάρτηση πυρήνα (linear kernel)

Η γραμμική συνάρτηση πυρήνα αποτελεί την απλούστερη μορφή γραμμικού διαχωρισμού των δεδομένων.

$$K(x, y) = x \cdot y$$

2. Πολυωνυμική συνάρτηση πυρήνα (polynomial kernel)

Η πολυωνυμική συνάρτηση πυρήνα αποτελεί μία από τις ευρύτερα χρησιμοποιούμενες συναρτήσεις, όταν τα δεδομένα δεν είναι γραμμικά διαχωρίσιμα:

$$K(x, y) = (x \cdot y)^p$$

ή

$$K(x, y) = (\gamma (x \cdot y) + \delta)^p$$

όπου, γ , δ , p οι παράμετροι της συνάρτησης και υποχρεωτικά $\gamma > 0$.

3. Γκαουσιανή συνάρτηση ακτινικής βάσης (Gaussian radial basis function)

Άλλη μία συνάρτηση πυρήνα που προβάλλει με μη γραμμικό τρόπο το εσωτερικό γινόμενο των διανυσμάτων σε έναν χώρο μεγαλύτερων διαστάσεων είναι η Γκαουσιανή συνάρτηση ακτινικής βάσης.

$$K(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right)$$

Το εύρος τιμών της Γκαουσιανής συνάρτησης ακτινικής βάσης περιορίζεται στο διάστημα $[0, 1]$, σε αντίθεση με την πολυωνυμική συνάρτηση που το εύρος τιμών της είναι το διάστημα

$[0, \infty]$.

4. Εκθετική συνάρτηση ακτινικής βάσης (Exponential radial basis function)

Μία μικρή παραλλαγή της Γκαουσιανής συνάρτησης ακτινικής βάσης είναι η εκθετική συνάρτηση ακτινικής βάσης:

$$K(x, y) = \exp\left(-\frac{\|x - y\|}{2\sigma^2}\right)$$

5. Σιγμοειδής συνάρτηση (Sigmoid function)

Μια συνήθης συνάρτηση πυρήνα είναι η σιγμοειδής συνάρτηση. Η χρήση της στους ταξινομητές SVM ισοδυναμεί με τη χρήση ενός νευρωνικού δικτύου πρόσθιας τροφοδότησης, που αποτελείται από δύο στρώματα:

$$K(x, y) = \tanh(\gamma (x \cdot y) + \delta)$$

Η επιλογή της καταλληλότερης συνάρτησης πυρήνα απαιτεί σε βάθος γνώση και κατανόηση του προβλήματος, καθώς και μελέτη των συναρτήσεων πυρήνα που χρησιμοποιούνται πιο συχνά. Αρκετές φορές χρησιμοποιούνται περισσότερες της μίας. Όταν έχει πλέον επιλεγεί η βέλτιστη συνάρτηση πυρήνα, για τον υπολογισμό των κατάλληλων παραμέτρων είναι χρήσιμο να πραγματοποιείται η διαδικασία της «εσωτερικής επαλήθευσης» (cross validation). Η διαδικασία αυτή εφαρμόζεται συχνά στις στατιστικές μεθόδους εκμάθησης (statistical learning methods) και στοχεύει στην πρόβλεψη της ακρίβειας της μεθόδου σε νέα δεδομένα

Στην περίπτωση που απαιτείται να χρησιμοποιηθούν πιο σύνθετες συναρτήσεις είναι δυνατόν να κατασκευαστούν νέες που συνδυάζουν ήδη γνωστές συναρτήσεις πυρήνα μέσω του παρακάτω θεωρήματος:

Έστω K_1 και K_2 γνωστές συναρτήσεις πυρήνα, τότε και οι ακόλουθες συναρτήσεις είναι επίσης συναρτήσεις πυρήνα:

$$K(x, z) = K_1(x, z) + K_2(x, z)$$

$$K(x, z) = a \cdot K_1(x, z)$$

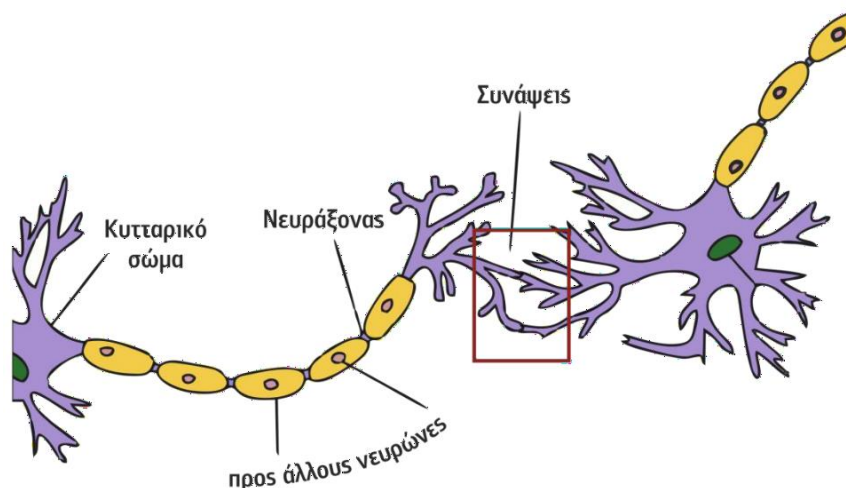
$$K(x, z) = K_1(x, z) \cdot K_2(x, z)$$

1.2.4 Νευρωνικά Δίκτυα

Ο όρος «Τεχνητά Νευρωνικά Δίκτυα» (ANN - Artificial Neural Networks) χρησιμοποιήθηκε αρχικώς ως μία έννοια για να περιγράψει ένα μαθηματικό μοντέλο προσομοίωσης της πολύπλοκης λειτουργίας του ανθρώπινου εγκεφάλου. Από την στιγμή που διαπιστώθηκε ότι ο ανθρώπινος εγκέφαλος λειτουργεί εντελώς διαφορετικά από τον τρόπο που λειτουργούν οι Ηλεκτρονικοί Υπολογιστές (Η/Υ) για να κάνουν υπολογισμούς, ξεκίνησε η προσπάθεια προσομοίωσης των μεθόδων μάθησης και λειτουργίας του ανθρώπινου εγκεφάλου από τους Η/Υ.

Ο εγκέφαλος είναι δομημένος με τέτοιο τρόπο ώστε να επιτρέπει την παράλληλη επεξεργασία δεδομένων και τη δυνατότητα συνεχούς μάθησης μέσω της αλληλεπίδρασης με το περιβάλλον. Τα βασικά αυτά χαρακτηριστικά συμβάλλουν στην ικανότητα, αφενός να εκτελεί δύσκολα καθήκοντα όπως ταχύτατη αναγνώριση μορφών, ταξινόμηση κ.ά., αφετέρου να εξελίσσεται συνεχώς μαθαίνοντας από το περιβάλλον του κατά την αλληλεπίδρασή του με αυτό. Ο εγκέφαλος αποτελείται από διακριτά στοιχεία, τους νευρώνες (neurons), που επικοινωνούν το ένα με το άλλο. Οι νευρώνες συνιστούν το βασικό δομικό κομμάτι του ανθρώπινου εγκεφάλου. Ένας νευρώνας διαχωρίζεται από τα υπόλοιπα κύτταρα με μια μεμβράνη και έχει την ικανότητα να μεταφέρει ηλεκτρικά σήματα από το νευρώνα αυτόν προς τους υπόλοιπους νευρώνες με τους οποίους επικοινωνεί. Κάθε νευρώνας αποτελείται από 3 κύρια τμήματα:

1. Τις διακλαδώσεις, οι οποίες ονομάζονται δενδρίτες (dendrites), που λειτουργούν ως κανάλια εισόδου για το νευρώνα και σκοπός τους είναι να λαμβάνουν το σήμα
2. Το κυρίως κυτταρικό σώμα (cell body)
3. Τον άξονα του κυττάρου ή νευροάξονα (αχον), που συνδέει ένα νευρώνα με άλλους νευρώνες.



Φυσικοί διασυνδεδεμένοι νευρώνες
ΠΗΓΗ: Γεωργούλη (2015) [50]

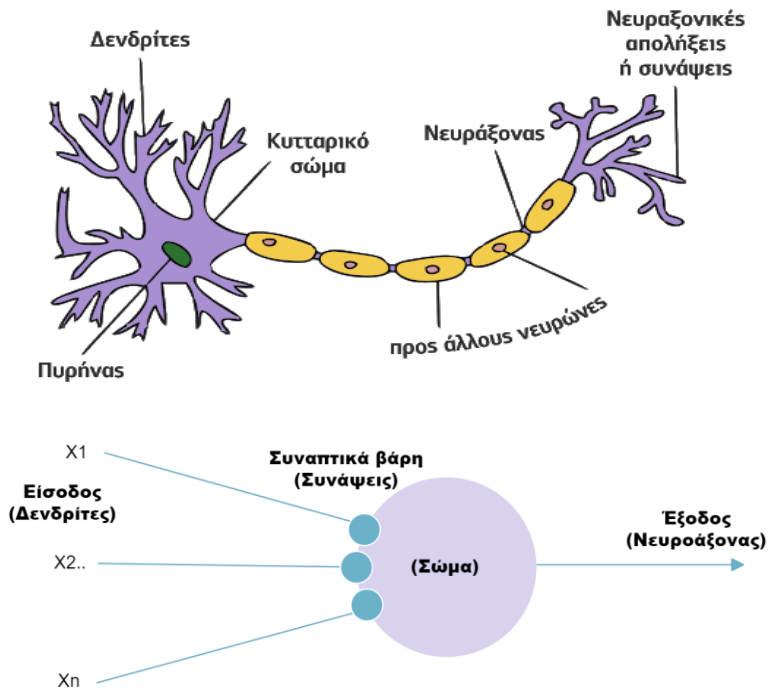
Η δομή του τεχνητού νευρωνικού δικτύου μιμείται κατά το δυνατό εκείνη του βιολογικού νευρωνικού δικτύου, ώστε να εμφανίζει παρόμοιες ιδιότητες. Κατ' αναλογία επομένως με ένα δίκτυο νευρώνων εγκεφάλου, ένα τεχνητό δίκτυο αποτελείται από ένα σύνολο τεχνητών νευρώνων που αλληλεπιδρούν, συνδεδεμένοι μεταξύ τους με τις λεγόμενες συνάψεις (synapses). Ο βαθμός αλληλεπίδρασης ποικίλει για κάθε ζεύγος νευρώνων και καθορίζεται από τα λεγόμενα συναπτικά βάρη.

Αντίστοιχα, σε ένα τεχνητό νευρωνικό δίκτυο οι νευρώνες αποτελούν τα δομικά στοιχεία του δικτύου. Κάθε τέτοιος κόμβος δέχεται ένα σύνολο αριθμητικών εισόδων από διαφορετικές πηγές (είτε από άλλους νευρώνες είτε από το περιβάλλον), εκτελεί έναν υπολογισμό, αξιοποιώντας αυτές τις πληροφορίες, και παράγει μία έξοδο. Η εν λόγω έξοδος είτε κατευθύνεται στο περιβάλλον είτε τροφοδοτείται ως είσοδος σε άλλους νευρώνες του δικτύου. Υπάρχουν τρεις τύποι νευρώνων:

1. Οι νευρώνες εισόδου
2. Οι νευρώνες εξόδου
3. Οι υπολογιστικοί ή κρυφοί νευρώνες.

Συνοπτικά για τη κάθε κατηγορία, οι νευρώνες εισόδου δεν επιτελούν κανέναν υπολογισμό, μεσολαβούν απλώς ανάμεσα στις περιβαλλοντικές εισόδους του δικτύου και τους υπολογιστικούς νευρώνες, ενώ οι νευρώνες εξόδου είναι αυτοί που διοχετεύουν στο περιβάλλον τις τελικές αριθμητικές εξόδους του δικτύου. Παράλληλα, οι υπολογιστικοί νευρώνες πολλαπλασιάζουν κάθε είσοδό τους με το αντίστοιχο συναπτικό βάρος και υπολογίζουν το ολικό άθροισμα των γινομένων. Το άθροισμα αυτό τροφοδοτείται ως όρισμα στη συνάρτηση ενεργοποίησης, την οποία υλοποιεί εσωτερικά κάθε κόμβος. Η τιμή που λαμβάνει η συνάρτηση για το εν λόγω όρισμα είναι και η έξοδος του νευρώνα για τις τρέχουσες εισόδους και βάρη.

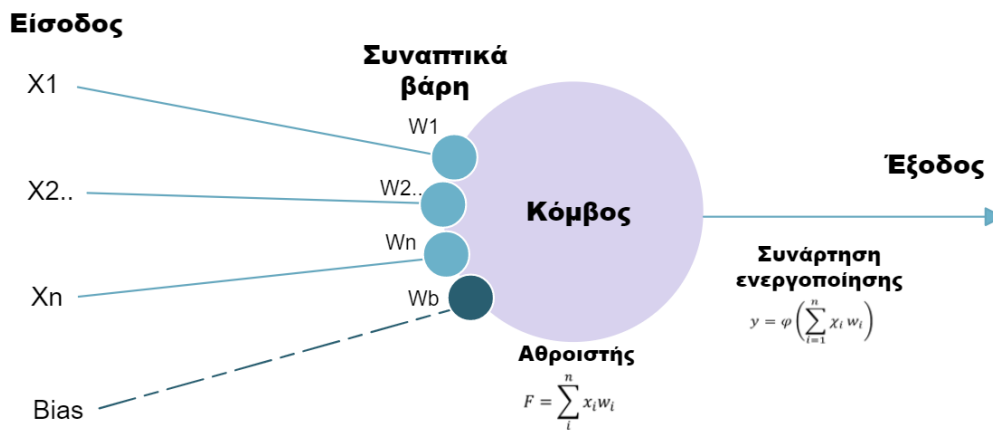
Οι νευρώνες με κοινά χαρακτηριστικά ονομάζονται επίπεδο νευρώνων και μεταξύ της εισόδου και της εξόδου ενός νευρωνικού δικτύου μπορεί να υπάρχουν περισσότερα του ενός επίπεδα, τα λεγόμενα υπολογιστικά ή κρυφά επίπεδα νευρώνων (hidden layers). Τα επίπεδα αυτά ονομάζονται έτσι διότι οι τιμές τους δεν εμφανίζονται στα δεδομένα εκπαίδευσης.



Ο φυσικός νευρώνας σε σχέση με τον στοιχειώδη τεχνητό νευρώνα
ΠΗΓΗ: Γεωργούλη (2015) [50] (πάνω εικόνα)

Συνοπτική περιγραφή ενός νευρωνικού δικτύου:

- Τα νευρωνικά δίκτυα συνήθως οργανώνονται σε επίπεδα (layers), τα οποία καλούνται και στρώματα. Τα ενδιάμεσα επίπεδα καλούνται κρυφά επίπεδα (hidden layers) και δεν είναι απαραίτητο να υπάρχουν.
- Οι εισοδοί παρουσιάζονται στο δίκτυο μέσω του επιπέδου εισόδου (input layer) το οποίο επικοινωνεί με έναν ή περισσότερα κρυμμένα επίπεδα. Τα κρυμμένα επίπεδα συνδέονται με το επίπεδο εξόδου (output layer) από το οποίο εξάγεται το αποτέλεσμα.
- Τα επίπεδα αποτελούνται από έναν αριθμό κόμβων (nodes) που είναι συνδεδεμένοι με τέτοιο τρόπο μεταξύ τους, ώστε ένας κόμβος να έχει συνδέσμους με πολλούς άλλους κόμβους του ίδιου ή άλλου επιπέδου.
- Οι κόμβοι επιδρούν σε άλλους κόμβους με το να διεγείρουν ή να αναστέλλουν την ενεργοποίησή τους. Για να επιτευχθεί αυτό, κάθε κόμβος περιέχει έναν αθροιστή που αθροίζει τα εισερχόμενα σήματα στον νευρώνα, τα οποία πολλαπλασιάζονται με το αντίστοιχο συναπτικό βάρος της αντίστοιχης σύναψης από την οποία εισήλθαν, και μια συνάρτηση ενεργοποίησης μέσω της οποίας παράγεται μία μοναδική έξοδος, εάν το άθροισμα υπερβαίνει μία τιμή κατωφλίου (threshold value). Συνήθως το κανονικοποιημένο εύρος της εξόδου ενός νευρώνα είναι το κλειστό σύνολο $[0,1]$ ή $[-1,1]$.
- Περιλαμβάνει επίσης και ένα εξωτερικό βάρος bias. Το βάρος bias έχει ως αποτέλεσμα την αύξηση ή την μείωση της τιμής που δίνει ως είσοδο το δίκτυο στην συνάρτηση ενεργοποίησης ανάλογα με το αν είναι αρνητικό ή θετικό. Το βάρος αυτό ονομάζεται μεροληψία ή πόλωση (bias). Ουσιαστικά, διασφαλίζει ότι ακόμη και αν το σταθμισμένο άθροισμα των εισόδων σε έναν νευρώνα είναι μηδέν, υπάρχει κάποια τιμή για την ενεργοποίηση του νευρώνα. Παράδειγμα, αν το bias είναι αρνητικό τότε ο βεβαρημένος συνδυασμός των εισόδων πρέπει να παραγάγει μία θετική τιμή μεγαλύτερη του $-bias$, έτσι ώστε να αναγκάσει τον νευρώνα που ταξινομεί να έχει τιμή άνω του κατωφλίου 0. Χωρικά, το bias μεταβάλλει την θέση (αλλά όχι τον προσανατολισμό) του συνόρου απόφασης.



Αρχιτεκτονική τεχνητού νευρώνα

Υπάρχουν πολλές κατηγορίες νευρωνικών δικτύων, ανάλογα με την αρχιτεκτονική τους και τον τρόπο εκπαίδευσής τους. Βασικά στοιχεία που πρέπει να καθοριστούν κατά τη δημιουργία τους είναι:

- Ο αριθμός των ενδιάμεσων κρυφών επιπέδων
- Ο αριθμός των κόμβων ανά επίπεδο
- Ο τρόπος σύνδεσης των κόμβων μεταξύ τους
- Η συνάρτηση ενεργοποίησης

- Οι τιμές των αρχικών βαρών μεταξύ των κόμβων
- Οι αλγόριθμοι (κανόνες εκπαίδευσης) που χρησιμοποιούνται, για να ενισχυθούν οι σύνδεσμοι μεταξύ των κόμβων κατά τη διαδικασία της εκπαίδευσης.

1.2.4.1 Βασικές συναρτήσεις ενεργοποίησης

Όπως έχει ήδη αναφερθεί, ο τεχνητός νευρώνας δίνει έξοδο μέσω της συνάρτησης ενεργοποίησης μόνο όταν το σταθμισμένο άθροισμα των εισόδων είναι μεγαλύτερο μιας ορισμένης τιμής κατώφλιου (threshold value) θ . Μια συνάρτηση ενεργοποίησης ενός νευρωνικού δικτύου για την έξοδο του νευρώνα, είναι της μορφής:

$$y_k = \varphi \left(\sum_{i=1}^n x_{ki} w_{ki} \right)$$

όπου y_k η έξοδος του k νευρώνα, φ η συνάρτηση ενεργοποίησης, x_{ki} είναι η i -οστή είσοδος και w_{ki} το i -οστό συναπτικό βάρος του k νευρώνα.

Η συνάρτηση ενεργοποίησης μπορεί να είναι βηματική (step function), γραμμική (linear function), μη γραμμική (non-linear function), στοχαστική (stochastic function). Παρακάτω παρουσιάζονται οι ευρύτερα χρησιμοποιούμενες συναρτήσεις ενεργοποίησης:

- **Βηματική συνάρτηση ενεργοποίησης ή αλλιώς συνάρτηση κατώφλι (Threshold function):**

Ένας νευρώνας που χρησιμοποιεί τέτοιου είδους συνάρτηση ενεργοποίησης ακολουθεί το μοντέλο των McCulloch–Pitts. Σε αυτό το μοντέλο η έξοδος του νευρώνα παίρνει την τιμή 1 αν η ολική δυνατότητα ενεργοποίησης του νευρώνα είναι μη αρνητική, διαφορετικά παίρνει την τιμή 0:

$$\varphi(U_k) = \begin{cases} 1, & \text{αν } U_k \geq 0 \\ 0, & \text{αν } U_k < 0 \end{cases}$$

Αντίστοιχα, η έξοδος του νευρώνα k με την χρήση μιας τέτοιας συνάρτησης εκφράζεται ως:

$$y_k = \begin{cases} 1, & \text{αν } U_k \geq 0 \\ 0, & \text{αν } U_k < 0 \end{cases}$$

όπου U_k είναι η δυνατότητα ενεργοποίησης του νευρώνα:

$$U_k = \sum_{j=1}^m w_{kj} x_j - b_k$$

Η βηματική συνάρτηση έχει το μειονέκτημα ότι η παράγωγός της είναι μηδέν. Δεδομένου ότι μάθηση στα νευρωνικά δίκτυα είναι η μεταβολή των τιμών των βαρών και μεταβολή σχετίζεται με την παράγωγο, η βηματική συνάρτηση δεν θεωρείται βολική ως συνάρτηση ενεργοποίησης των νευρώνων στα νευρωνικά δίκτυα.

- **Γραμμική συνάρτηση:**

Η κλασική γραμμική συνάρτηση ενεργοποίησης μπορεί να είναι της μορφής

$$\varphi(U_k) = U_k$$

Για τη γραμμική συνάρτηση ενεργοποίησής Piecewise:

$$\varphi(\cdot) = \begin{cases} 1, & \text{αν } U \geq \frac{1}{2} \\ U, & \text{αν } \frac{1}{2} > U > -\frac{1}{2} \\ 0, & \text{αν } U \leq -\frac{1}{2} \end{cases}$$

όπου, ο παράγοντας ενίσχυσης (amplification factor) μέσα στην γραμμική περιοχή της συνάρτησης θεωρείται μονάδα. Το είδος αυτών των συναρτήσεων μπορεί να θεωρηθεί μια προσέγγιση σε μη γραμμικό ενισχυτή. Οι δυο περιπτώσεις παρακάτω είναι ειδικές μορφές της Piecewise-Linear Function:

- Ένας γραμμικός συνδυαστής χρησιμοποιείται όταν η γραμμικότητα της περιοχής ενεργοποίησης διατηρείται χωρίς να υπάρχουν κορεσμοί
- Η Piecewise-Linear Function μετατρέπεται σε συνάρτηση κατώφλι αν ο παράγοντας ενίσχυσης της γραμμικής περιοχής γίνει απείρωσ μεγάλος.

• **Μη γραμμική συνάρτηση:**

Συνηθέστερα, χρησιμοποιούνται μη γραμμικές συναρτήσεις, όπως οι σιγμοειδείς συναρτήσεις (sigmoid functions) και οι Γκαουσιανές συναρτήσεις (Gaussian functions). Η λογιστική σιγμοειδής συνάρτηση είναι η ευρύτερα χρησιμοποιούμενη συνάρτηση ενεργοποίησης για την κατασκευή νευρωνικών δικτύων. Ορίζεται ως μια γνησίως αύξουσα συνάρτηση που είναι ομαλή και ασυμπτωτική. Ένα παράδειγμα σιγμοειδούς συνάρτησης είναι η παρακάτω:

$$\varphi(U_k) = \frac{1}{1 + e^{(-au)}}$$

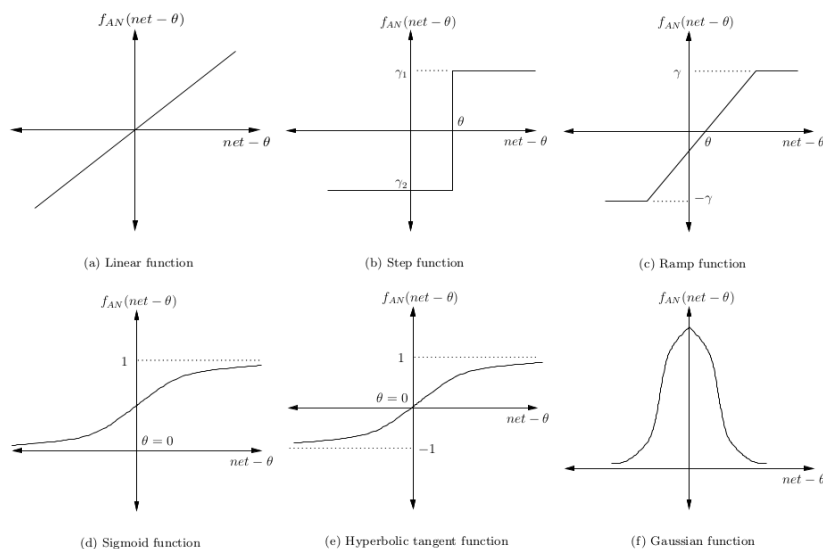
όπου α είναι η παράμετρος κλίσης της σιγμοειδούς συνάρτησης. Με μεταβολή του α μπορεί να μεταβληθεί η κλίση της σιγμοειδούς. Στο όριο, καθώς το α τείνει προς το άπειρο, η σιγμοειδής συνάρτηση μετατρέπεται σε συνάρτηση κατώφλι. Αντίθετα με την συνάρτηση κατώφλι, που παίρνει τιμές 0 ή 1, η σιγμοειδής συνάρτηση μπορεί να πάρει όλες τις τιμές στο διάστημα $[0,1]$. Επίσης, η σιγμοειδής συνάρτηση είναι διαφορίσιμη ενώ η συνάρτηση κατώφλι δεν είναι.

Όλες οι παραπάνω συναρτήσεις ενεργοποίησης παίρνουν τιμές από το 0 ως το +1, κάποιες φορές όμως είναι επιθυμητό η συνάρτηση ενεργοποίησης να παίρνει τιμές από το -1 ως το +1. Σε αυτή την περίπτωση η συνάρτηση ενεργοποίησης παίρνει μια αντισυμμετρική μορφή ως προς την αρχή των αξόνων. Συγκεκριμένα η συνάρτηση κατώφλι γίνεται:

$$\varphi(\cdot) = \begin{cases} 1, & \text{αν } U \geq 0 \\ U, & \text{αν } U = 0 \\ 0, & \text{αν } U < 0 \end{cases}$$

Η σιγμοειδής συνάρτηση παίρνει τη μορφή της υπερβολικής εφαπτομένης ως εξής:

$$\varphi(U_k) = \tanh(U_k)$$



Συναρτήσεις ενεργοποίησης (θ : κατώφλι)

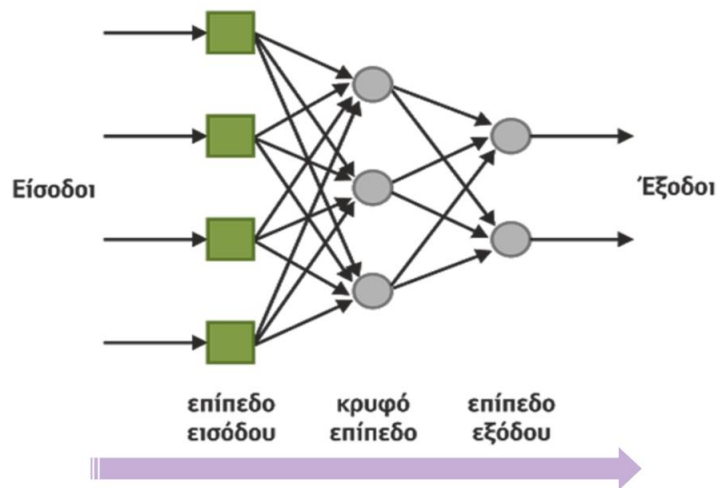
ΠΗΓΗ: <http://www.turingfinance.com/misconceptions-about-neural-networks/>

1.2.4.2 Αρχιτεκτονική νευρωνικών δικτύων

Ο τρόπος με τον οποίο οι νευρώνες ενός νευρωνικού δικτύου δομούνται είναι στενά συνδεδεμένος με τον αλγόριθμο εκμάθησης που χρησιμοποιείται για την εκπαίδευση του δικτύου. Υπάρχουν δυο βασικές κατηγορίες νευρωνικών δικτύων:

- **Δίκτυα πρόσθιας τροφοδότησης (Feed-Forward networks)**

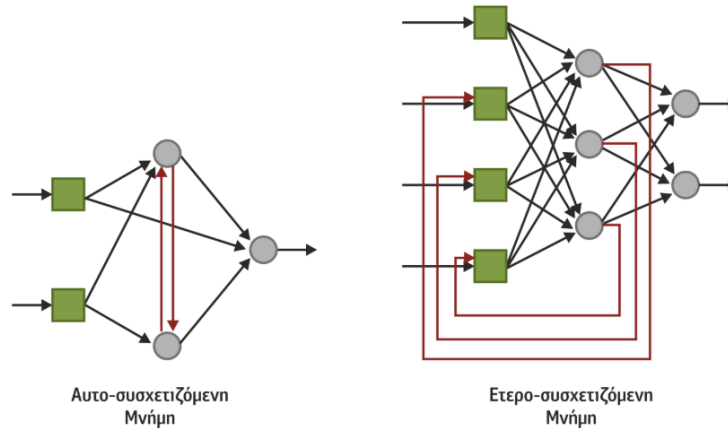
Ως πρόσθιας τροφοδότησης νευρωνικά δίκτυα νοούνται τα δίκτυα στα οποία τα σήματα κατευθύνονται από την είσοδο προς την έξοδο. Στην περίπτωση που ένα νευρωνικό δίκτυο αποτελείται από ένα μόνο επίπεδο νευρώνων, το δίκτυο είναι ένα αυστηρά πρόσθιας τροφοδότησης (feed-forward) δίκτυο και καλείται Δίκτυο Ενός Επιπέδου. Με τον όρο 'ένα επίπεδο' νοείται το επίπεδο εξόδου που περιέχει και τους νευρώνες όπου γίνονται οι υπολογισμοί. Στην περίπτωση νευρωνικού δικτύου πολλαπλών επιπέδων, δηλαδή όταν υφίστανται κρυφά επίπεδα, οι κόμβοι εισόδου στο πρώτο επίπεδο (επίπεδο εισόδου) του δικτύου παρέχουν τα στοιχεία των προτύπων εισόδου με την μορφή διανυσμάτων που τροφοδοτούν τους νευρώνες του δεύτερου επιπέδου, δηλαδή του πρώτου κρυφού επιπέδου. Τα σήματα εξόδου των νευρώνων του δεύτερου επιπέδου (πρώτου κρυφού επιπέδου) γίνονται σήματα εισόδου στο τρίτο επίπεδο και συνεχίζεται κατά αυτό τον τρόπο η ροή των σημάτων μεταξύ των επιπέδων του δικτύου μέχρι να φτάσουν στο επίπεδο εξόδου.



Νευρωνικό δίκτυο πρόσθιας τροφοδότησης (Feed-Forward network)
 ΠΗΓΗ: : Γεωργούλη (2015) [50]

- **Δίκτυα οπίσθιας τροφοδότησης ή Αναδρομικά Νευρωνικά Δίκτυα (Feed-Backward networks ή Recurrent Neural Networks -RNN)**

Τα δίκτυα αυτά αφορούν την περίπτωση όπου οι έξοδοι κάποιων νευρώνων γίνονται εισοδοί σε νευρώνες προηγούμενων επιπέδων και τότε θεωρείται πως υπάρχει ανάδραση, δηλαδή περιέχεται τουλάχιστον ένας βρόχος ανατροφοδότησης στο δίκτυο. Αυτό σημαίνει ότι σε έναν τουλάχιστον νευρώνα, το σήμα εξόδου επηρεάζει το σήμα που έρχεται στην είσοδο του νευρώνα. Αν η ανατροφοδότηση αφορά κόμβους στο ίδιο επίπεδο, τότε τα δίκτυα καλούνται αυτοσυσχετιζόμενες μνήμες (autoassociated memories), διαφορετικά καλούνται ετεροσυσχετιζόμενες μνήμες (heteroassociated memories). Στην περίπτωση που ένα νευρωνικό δίκτυο αποτελείται από ένα μόνο επίπεδο νευρώνων, όπου κάθε νευρώνας επιστρέφει το σήμα εξόδου του σαν σήμα εισόδου σε όλους τους άλλους νευρώνες του επιπέδου, τότε στο σχήμα αυτό δεν υπάρχουν κρυφοί νευρώνες ούτε βρόχοι, όπου ένας νευρώνας να επιστρέφει το σήμα εξόδου σαν είσοδο στον εαυτό του.



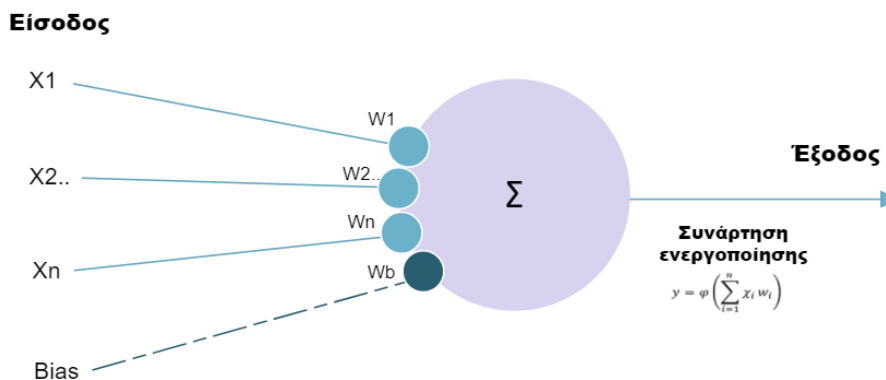
Νευρωνικό δίκτυο οπίσθιας τροφοδότησης (Feed-Backward network)
 ΠΗΓΗ: : Γεωργούλη (2015) [50]

1.2.4.3 Perceptron

Ένα από τα απλούστερα νευρωνικά δίκτυα που προσομοιώνουν τον φυσικό νευρώνα είναι ο στοιχειώδης Perceptron (basic Perceptron), δηλαδή ένα τεχνητό νευρωνικό δίκτυο που αποτελείται από έναν μόνο νευρώνα πρόσθιας τροφοδότησης (feed forward). Αν και ο όρος Perceptron (Rosenblatt, 1962) χρησιμοποιήθηκε αρχικά για το στοιχειώδες νευρωνικό δίκτυο πολλών εισόδων και μόνο μιας εξόδου, έχει επικρατήσει να χαρακτηρίζονται ως Perceptrons όλα τα τεχνητά νευρωνικά δίκτυα πρόσθιας τροφοδότησης που δεν περιέχουν στην αρχιτεκτονική τους κρυφά επίπεδα. Ο Perceptron λοιπόν είναι ένας δυαδικός ταξινομητής, δηλαδή μία συνάρτηση η οποία απεικονίζει την είσοδο x (ένα διάνυσμα με πραγματικές τιμές) σε μία τιμή εξόδου y του νευρώνα k (μία και μοναδική δυαδική τιμή). Συνεπώς η τιμή της y_k (0 ή 1) χρησιμοποιείται για να ταξινομήσει το x ως θετικό ή αρνητικό στοιχείο στην περίπτωση ενός δυαδικού προβλήματος ταξινόμησης.

Αν διαθέτουν την κατάλληλη δομή, οι Perceptrons είναι ικανοί να επιλύουν οποιαδήποτε γραμμική συνάρτηση. Η έξοδος y_k του Perceptron για ένα διάνυσμα εισόδου $x = (x_1, x_2, \dots, x_n)$ δίνεται μέσω της συνάρτησης ενεργοποίησης φ ως ακολούθως:

$$y_k = \varphi \left(\sum_{i=1}^n x_{ki} w_{ki} \right)$$



Ο στοιχειώδης Perceptron (basic Perceptron)

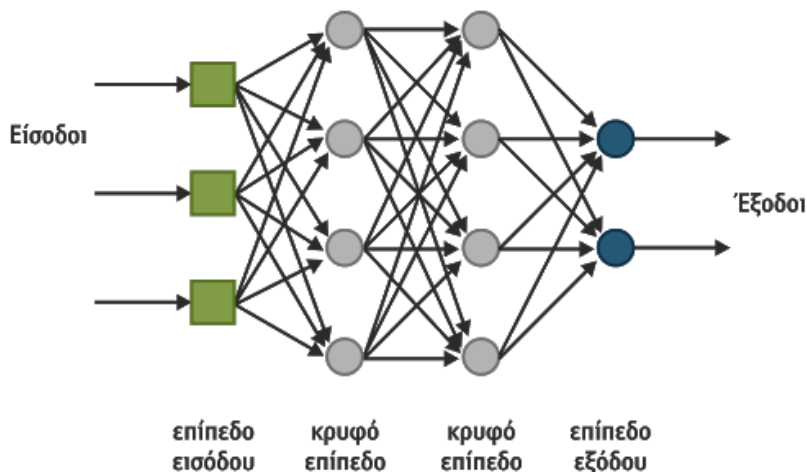
1.2.4.4 Πολυεπίπεδα Νευρωνικά Δίκτυα

Το γεγονός ότι τα νευρωνικά δίκτυα που αποτελούνται από ένα μόνο επίπεδο νευρώνων, όπως ο Perceptron, δεν είναι σε θέση να επιλύσουν προβλήματα με μη γραμμικά διαχωρίσιμες

τιμές εξόδου οδήγησε στην ανάγκη προσθήκης κρυφών επιπέδων στη δομή των Perceptrons και στη δημιουργία πολυεπίπεδων νευρωνικών δικτύων. Τα πολυεπίπεδα νευρωνικά δίκτυα διαφέρουν από τα προηγούμενα λόγω της ύπαρξης ενός ή περισσότερων κρυφών επιπέδων, των οποίων οι νευρώνες καλούνται κρυφοί νευρώνες. Η λειτουργία των κρυφών νευρώνων είναι να παρεμβάλλονται μεταξύ των κόμβων εισόδου και των νευρώνων εξόδου του δικτύου. Με την προσθήκη ενός ή περισσότερων κρυφών επιπέδων, το δίκτυο έχει την δυνατότητα να προσεγγίζει συναρτήσεις μεγαλύτερης πολυπλοκότητας.

Οι κόμβοι εισόδου του δικτύου παρέχουν τα στοιχεία των προτύπων εισόδου με την μορφή διανυσμάτων, που γίνονται εισοδοί στους νευρώνες του δεύτερου επιπέδου, δηλαδή του πρώτου κρυφού επιπέδου. Η διαδικασία συνεχίζεται μέχρι να φτάσουν στο επίπεδο εξόδου. Όπως έχει ήδη αναφερθεί, συνήθως οι νευρώνες σε κάθε επίπεδο του δικτύου παίρνουν ως εισοδο μόνο τα σήματα εξόδου από τους νευρώνες του προηγούμενου επιπέδου, και άρα αποτελούν δίκτυα πρόσθιας τροφοδότησης. Στην πλειοψηφία των εφαρμογών χρησιμοποιούνται τέτοια δίκτυα ενός κρυφού επιπέδου με πλήρως συνδεδεμένους κόμβους.

Το νευρωνικό δίκτυο του παρακάτω σχήματος με τα δύο κρυφά επίπεδα καλείται πλήρως διασυνδεδεμένο (fully connected), με την έννοια ότι κάθε κόμβος σε κάθε επίπεδο του δικτύου είναι συνδεδεμένος με κάθε κόμβο του επόμενου επιπέδου του δικτύου. Στην περίπτωση που κάποιες από τις συνδέσεις δεν υπάρχουν τότε το δίκτυο είναι μερικώς διασυνδεδεμένο (partially connected).



Πολυεπίπεδο νευρωνικό δίκτυο, πρόσθιας τροφοδότησης, δύο κρυφών επιπέδων, πλήρως διασυνδεδεμένο. ΠΗΓΗ: : Γεωργούλη (2015) [50]

Ένα ευρέως χρησιμοποιούμενο νευρωνικό δίκτυο για την ταξινόμηση και αναγνώριση αντικειμένων είναι ο Perceptron πολλαπλών επιπέδων (Multilayer perceptron - MLP). Ο Perceptron πολλαπλών επιπέδων είναι ένα νευρωνικό δίκτυο πρόσθιας τροφοδοσίας με ένα ή περισσότερα κρυφά επίπεδα μεταξύ των επιπέδων εισόδου και εξόδου. Στην περίπτωση που υπάρχουν περισσότερα από δύο κρυφά επίπεδα, τότε η χρήση του εν λόγω νευρωνικού δικτύου αποτελεί τεχνική βαθιάς μάθησης.

Γενικότερα, δεν υπάρχει καθολικά συμφωνημένο κατώτατο όριο για το πότε ένα πολυεπίπεδο νευρωνικό δίκτυο θεωρείται βαθιά ή όχι μάθησης. Για την διάκριση αυτή, χρησιμοποιείται συχνά ο όρος CAP (Credit Assignment Path) που αναφέρεται στην αλυσίδα μετασχηματισμών από την είσοδο στην έξοδο του δικτύου. Για ένα νευρωνικό δίκτυο πρόσθιας τροφοδοσίας, το βάθος των CAP, και συνεπώς το βάθος του δικτύου, είναι ο αριθμός των κρυφών επιπέδων συν ένα. Οι περισσότεροι ερευνητές στον τομέα συμφωνούν ότι η βαθιά μάθηση έχει πολλαπλά μη γραμμικά επίπεδα περισσότερα των $CAP > 2$ και ο Schmidhuber (2015) [103] θεωρεί ότι στην περίπτωση που εντοπίζονται περισσότερα των $CAP > 10$ επιπέδων το δίκτυο αποτελεί τεχνικής πολύ βαθιάς μάθησης.

1.2.4.5 Εκπαίδευση & Αξιολόγηση Νευρωνικών δικτύων

Η πιο σημαντική ιδιότητα των νευρωνικών δικτύων είναι η ικανότητά τους να μαθαίνουν από το περιβάλλον τους και να βελτιώνουν την απόδοσή τους μέσω της διαδικασίας της μάθησης. Ακριβώς για αυτήν τους τη δυνατότητα που έχουν να μαθαίνουν μετά από εκπαίδευση, είναι τόσο δημοφιλή στο χώρο των συστημάτων Μηχανικής Μάθησης. Η εκπαίδευσή τους γίνεται βάσει αλγορίθμων που μπορεί να υιοθετούν μοντέλα μάθησης με επίβλεψη, χωρίς επίβλεψη ή ενισχυτικής μάθησης.

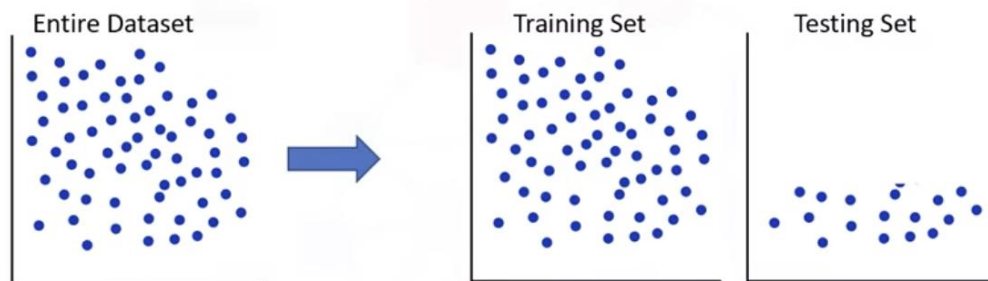
- **Εκπαίδευση με επίβλεψη**, όπου το νευρωνικό δίκτυο μαθαίνει να απεικονίζει δεδομένες εισόδους σε εξόδους εκ των προτέρων γνωστές (σύνολο εκπαίδευσης), με απώτερο στόχο τη γενίκευση της αναγνώρισης αυτής και για παρεμφερείς εισόδους στο μέλλον.
- **Εκπαίδευση χωρίς επίβλεψη**, όπου το νευρωνικό δίκτυο κατασκευάζει απεικονίσεις από μια αναπαράσταση σε μια άλλη.
- **Ενισχυτική Εκπαίδευση**, όπου ο αλγόριθμος μαθαίνει μια στρατηγική ενεργειών για δεδομένου τύπου παρατηρήσεις.

Ένα νευρωνικό δίκτυο μαθαίνει από το περιβάλλον του μέσω μιας επαναληπτικής διαδικασίας προσαρμογών των συναπτικών βαρών στις συνάψεις του δικτύου και των τιμών των κατωφλίων του. Στην ιδανική περίπτωση, μετά από κάθε επανάληψη της διαδικασίας μάθησης το δίκτυο αποκτά περισσότερη γνώση για το περιβάλλον του (Γεωργούλη, 2015 [50]).

Θα πρέπει να σημειωθεί ότι δεν υπάρχει ένας μοναδικός αλγόριθμος εκπαίδευσης κατάλληλος για όλες τις περιπτώσεις. Αντίθετα, υπάρχει ένα σύνολο αλγορίθμων που χρησιμοποιούνται κατά περίπτωση ανάλογα με τα πλεονεκτήματα και τα μειονεκτήματά τους καθώς και ανάλογα με το είδος του προβλήματος. Το βασικό σημείο στο οποίο διαφέρουν οι αλγόριθμοι εκπαίδευσης είναι ο τρόπος με τον οποίο προσαρμόζουν τα βάρη Wkj των συνάψεων τους.

Μετά την εκπαίδευση του δικτύου ακολουθεί η φάση της αξιολόγησης (evaluation), δηλαδή του ελέγχου του δικτύου με τη βοήθεια ενός νέου συνόλου δειγμάτων τα οποία δεν είχαν πάρει μέρος στη διαδικασία εκπαίδευσης. Υπάρχουν δύο τύποι διαδικασίας αξιολόγησης που μπορούν να χρησιμοποιηθούν, η μέθοδος όπου το σύνολο των δεδομένων ελέγχου είναι υποσύνολο του συνόλου δεδομένων εκπαίδευσης (Train & Test in the same dataset) και η μέθοδος όπου τα δεδομένα χωρίζονται σε δεδομένα εκπαίδευσης και δεδομένα ελέγχου (Train/Test Split).

Η πρώτη προσέγγιση αξιολόγησης είναι η πιο απλή. Γίνεται χρήση ολόκληρου του συνόλου δεδομένων για την εκπαίδευση και ενός μόνο υποσυνόλου αυτών για τον έλεγχο. Τα δεδομένα ελέγχου συνοδεύονται με τις αντίστοιχες ετικέτες (εξόδους), αλλά εκείνες δεν χρησιμοποιούνται στην πρόβλεψη παρά μόνο για έλεγχο, ως πραγματικές τιμές (ground truth).

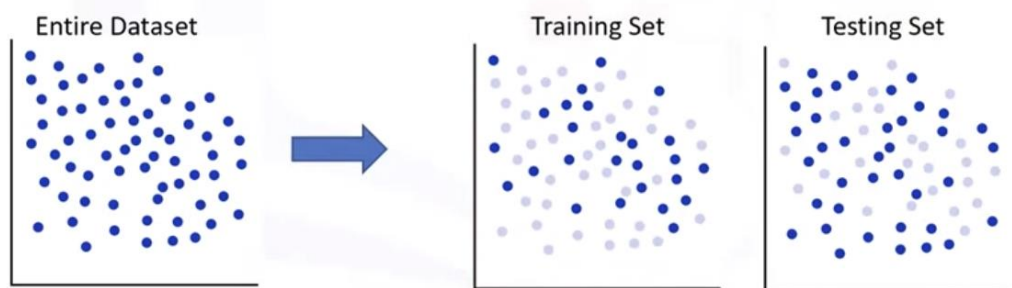


Δεδομένα εκπαίδευσης και ελέγχου στο ίδιο σύνολο

ΠΗΓΗ: <https://medium.com/@toprak.mhmt/in-this-blog-ill-be-covering-model-evaluation-cc91557a04e2>

Στη δεύτερη προσέγγιση αξιολόγησης (Train/Test Split), όπου επιλέγεται ένα τμήμα του συνόλου δεδομένων για εκπαίδευση και το υπόλοιπο χρησιμοποιείται για δοκιμές, οι προβλεπόμενες τιμές που προκύπτουν για τα δεδομένα δοκιμών συγκρίνονται και πάλι με τις πραγματικές τιμές τους (ground truth). Η προσέγγιση αυτή παρέχει πιο έγκυρη αξιολόγηση

της ακρίβειας του μοντέλου, αφού το μοντέλο δεν έχει εκπαιδευτεί καθόλου πάνω στα δεδομένα ελέγχου.



Δεδομένα εκπαίδευσης και ελέγχου διαχωρισμένα (Split)

ΠΗΓΗ: <https://medium.com/@toprak.mhmt/in-this-blog-ill-be-covering-model-evaluation-cc91557a04e2>

Το στάδιο της αξιολόγησης ενός δικτύου στοχεύει στην μέτρηση της απόδοσης του χρησιμοποιούμενου μοντέλου, ενώ παράλληλα παίζει βασικό ρόλο στην ανάπτυξη του δικτύου, καθώς κατά την διενέργειά του προκύπτουν πληροφορίες για τομείς που απαιτούν βελτίωση. Η ακρίβεια (*Precision*) και η ανάκληση (*Recall*) είναι δύο απλά αλλά χρήσιμα ποσοτικά μέτρα που εκφράζουν την ποιότητα των προβλέψεων του δικτύου. Η αποθήκευση του αριθμού των πραγματικών θετικών, ψευδών θετικών, πραγματικών αρνητικών και ψευδών αρνητικών προβλέψεων του μοντέλου επιτρέπει να υπολογιστούν οι τιμές ακρίβειας και ανάκλησης.

Η ακρίβεια μπορεί να οριστεί ως το ποσοστό των σχετικών αποτελεσμάτων, ενώ η ανάκληση χαρακτηρίζεται ως το ποσοστό των σχετικών αποτελεσμάτων που ταξινομούνται σωστά από το μοντέλο. Η ακρίβεια προσπαθεί να απαντήσει στην ερώτηση ποιο είναι το ποσοστό θετικών ταυτοποιήσεων που ήταν πραγματικά σωστά. Βοηθάει όταν το κόστος των ψευδών θετικών είναι υψηλό και υπολογίζεται ως εξής:

$$Precision = \frac{\text{Θετικά Σωστά ταξινομημένα δεδομένα}}{\text{Πραγματικά Αποτελέσματα (σύνολο θετικών)}} = \frac{\text{Θετικά Σωστά}}{\text{Θετικά Σωστά} + \text{Θετικά Λάθος}}$$

Η ανάκληση προσπαθεί να απαντήσει στην ερώτηση ποιο ποσοστό των πραγματικά θετικών ταυτοποιήθηκε σωστά. Βοηθάει όταν το πλήθος των ψευδών αρνητικών είναι υψηλό και υπολογίζεται ως εξής:

$$Recall = \frac{\text{Θετικά Σωστά ταξινομημένα δεδομένα}}{\text{Αποτελέσματα πρόβλεψης}} = \frac{\text{Θετικά Σωστά}}{\text{Θετικά Σωστά} + \text{Αρνητικά Λάθος}}$$

Κατά την αξιολόγηση ενός δικτύου, ένα ακόμη μέτρο που χρησιμοποιείται είναι εκείνο της ορθότητας (*Accuracy*) που εκφράζει το πόσο κοντά είναι τα αποτελέσματα του μοντέλου σε μία συγκεκριμένη τιμή. Συγκεκριμένα, η ορθότητα εκφράζει το πόσο συχνά είναι σωστό το μοντέλο συνολικά, ενώ η ακρίβεια εκφράζει το πόσο συχνά το μοντέλο πετυχαίνει σωστή θετική πρόβλεψη. Η συνολική ορθότητα ενός μοντέλου ορίζεται ως:

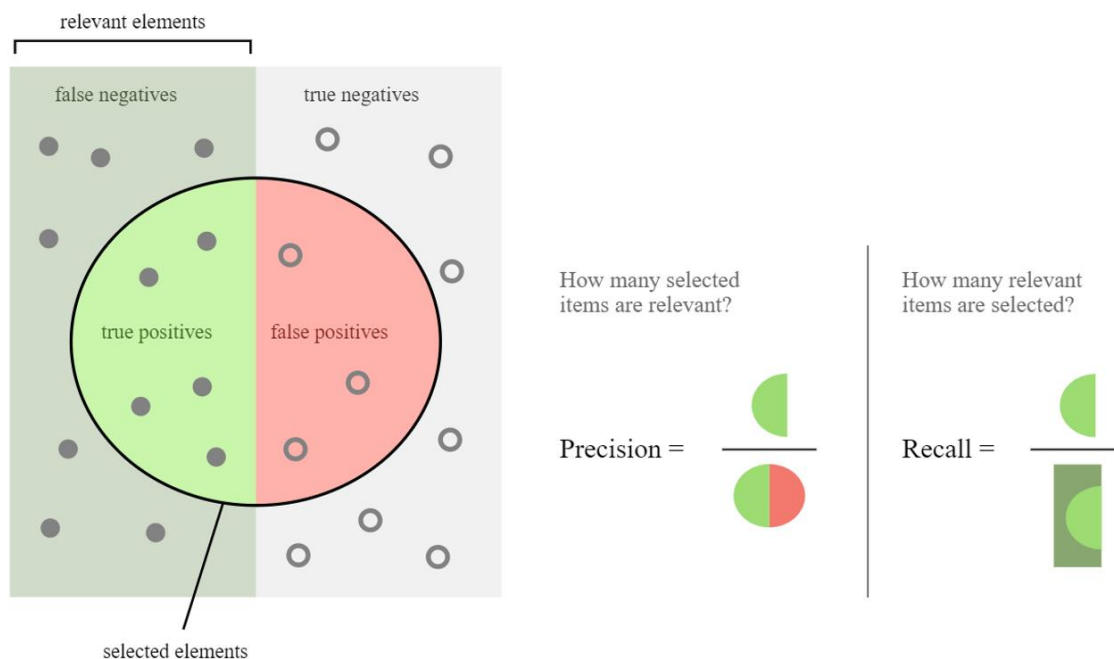
$$Accuracy = \frac{\text{Σύνολο σωστών προβλέψεων}}{\text{Σύνολο προβλέψεων}}$$

$$Accuracy = \frac{\text{Θετικά Σωστά} + \text{Θετικά Λάθος}}{\text{Θετικά Σωστά} + \text{Θετικά Λάθος} + \text{Αρνητικά Σωστά} + \text{Αρνητικά Λάθος}}$$

Τέλος, το *F1-Score* λαμβάνει υπόψη τόσο την ακρίβεια όσο και την ανάκληση για να αξιολογήσει τελικά την συνολική απόδοση του μοντέλου. Μια καλή βαθμολογία F1 σημαίνει ότι το μοντέλο έχει λίγα ψευδή θετικά και λίγα ψευδή αρνητικά, επομένως εξαγει σωστά

αποτελέσματα. Σκοπός είναι η τιμή F1 να είναι όσο πιο κοντά στο 1, ενώ αν είναι 0 σημαίνει πως το μοντέλο έχει αποτύχει.

$$F1Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$



Σχηματική αναπαράσταση Precision και Recall
 ΠΗΓΗ: https://en.wikipedia.org/wiki/Precision_and_recall

1.2.4.6 Αλγόριθμός οπισθοδιάδοσης σφάλματος (Back-propagation)

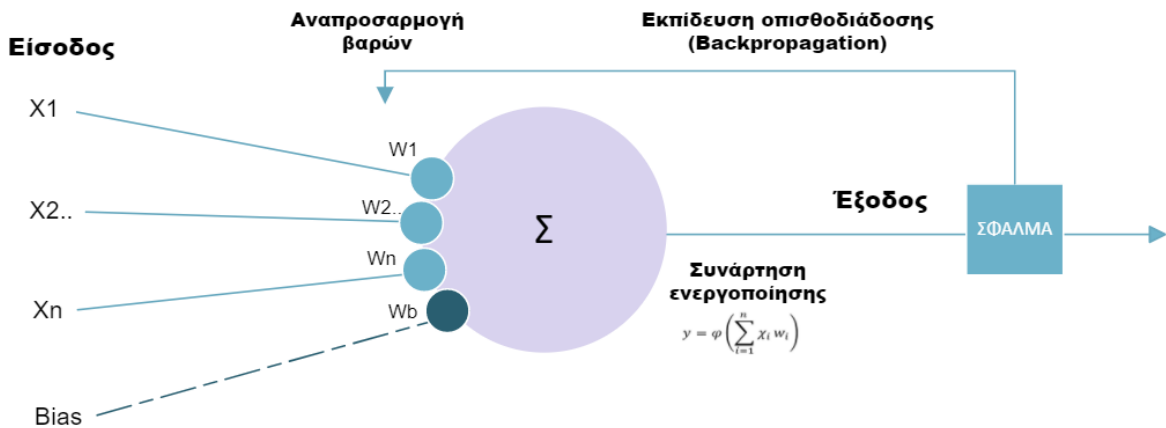
Ο αλγόριθμος οπισθοδιάδοσης εντάσσεται σε μια οικογένεια μεθόδων που χρησιμοποιούνται για την αποτελεσματική εκπαίδευση των τεχνητών νευρωνικών δικτύων, ακολουθώντας έναν αλγόριθμο βελτιστοποίησης βασισμένο στον υπολογισμό και ελαχιστοποίηση της κλίσης της συνάρτησης ενεργοποίησης. Το κύριο χαρακτηριστικό της οπισθοδιάδοσης είναι η επαναληπτική, αναδρομική και αποδοτική διαδικασία για τον υπολογισμό της μεταβολής των βαρών για τη βελτίωση του δικτύου έως ότου αυτό είναι σε θέση να εκτελέσει το έργο για το οποίο εκπαιδεύεται.

Η οπισθοδιάδοση απαιτεί οι παράγωγοι των συναρτήσεων ενεργοποίησης κατά το σχεδιασμό του δικτύου να είναι γνωστοί. Η αυτόματη διαφόριση είναι μια τεχνική που παρέχει αυτόματα και αναλυτικά τις παραγώγους των συναρτήσεων στον αλγόριθμο εκπαίδευσης. Στο πλαίσιο της μάθησης, η οπισθοδιάδοση χρησιμοποιείται συνήθως από τον αλγόριθμο βελτιστοποίησης μέσω κατάβασης της κλίσης (gradient descent), ώστε να ρυθμίσει το βάρος των νευρώνων υπολογίζοντας την κλίση της συνάρτησης κόστους. Η οπισθοδιάδοση υπολογίζει τις κλίσεις, ενώ η στοχαστική κατάβαση της κλίσης (stochastic gradient descent) χρησιμοποιεί τις κλίσεις για την εκπαίδευση του μοντέλου (μέσω βελτιστοποίησης).

Η μέθοδος αυτή αναπτύχθηκε ανεξάρτητα σε διάφορες παραλλαγές από τους Bryson & Ho (1969), Werbos (1974) και Parker (1982), αλλά έγινε ευρέως γνωστή και προωθήθηκε πολύ από το έργο "Parallel distributed processing" των Rumelhart & McClelland (1986), το οποίο άνοιξε πολλές εφαρμογές και νέα πεδία, και ανακίνησε το ενδιαφέρον σε όλη την περιοχή των νευρωνικών δικτύων.

Στα πολυεπίπεδα νευρωνικά δίκτυα, ο συνηθέστερος τρόπος επιβλεπόμενης μάθησης είναι η μάθηση με οπισθοδιάδοση σφάλματος (back-propagation). Η κεντρική ιδέα είναι αυτή που προαναφέρθηκε και είναι αρκετά απλή. Το δίκτυο ξεκινά τη διαδικασία εκπαίδευσης από τυχαίες τιμές βαρών. Εάν δώσει λάθος απάντηση στην έξοδο (που είναι και το πιο πιθανό) τότε τα συναπτικά βάρη διορθώνονται, έτσι ώστε το λάθος να γίνει μικρότερο. Ο τελικός

στόχος της εκπαίδευσης είναι η ελαχιστοποίηση μιας συνάρτησης κόστους που βασίζεται στο σφάλμα, έτσι ώστε το πραγματικό αποτέλεσμα κάθε νευρώνα στο δίκτυο να πλησιάζει το επιθυμητό αποτέλεσμα για τον νευρώνα κατά μια στατιστική έννοια. Η διαδικασία αυτή αποτελεί μια εποχή (epoch), δηλαδή ένα πέρασμα, είσοδος-έξοδος-είσοδος, και επαναλαμβάνεται πολλές φορές προκειμένου σταδιακά το λάθος να ελαττώνεται μέχρις ότου γίνει πολύ μικρό και ανεκτό. Στο σημείο αυτό το δίκτυο έχει εκπαιδευτεί με την ακρίβεια που είχε απαιτηθεί. Η σειρά που τροφοδοτούνται τα δεδομένα στο δίκτυο σε κάθε εποχή θα πρέπει να μεταβάλλεται τυχαία. Η εκπαίδευση του δικτύου με τον αλγόριθμο οπισθοδιάδοσης σφάλματος περιλαμβάνει τόσο την πρόσθια τροφοδότηση της εκπαίδευσης όσο και την οπίσθια τροφοδότηση.



Αλγόριθμός οπισθοδιάδοσης σφάλματος (Back-propagation)

Συνοψίζοντας, ο αλγόριθμος οπισθοδιάδοσης σφάλματος (Back-propagation) έχει την ακόλουθη μορφή:

1. **Είσοδος x παρατηρήσεων**
Ορίζεται η αντίστοιχη ενεργοποίηση a για το επίπεδο εισόδου και z_j^l η τιμή ενεργοποίησης του νευρώνα j του l -οστού επιπέδου
2. **Εμπρόσθια τροφοδότηση**
Για κάθε επίπεδο $l = 2, 3, \dots, L$ υπολογίζονται τα $z^l = w^l a^{l-1} + b^l$ και τα $a^l = \sigma(z^l)$
3. **Σφάλμα εισόδου**
Υπολογίζεται το διάνυσμα $\delta^L = \nabla_{\alpha} C \odot \sigma'(z^L)$, σφάλμα ενός επιπέδου όπου η συνάρτηση κόστους $C = \frac{1}{2} \|y(x) - a^L(x)\|^2$
4. **Οπισθοδιάδοση σφάλματος**
Για κάθε $l = L - 1, L - 2, \dots, 2$ υπολογίζεται το σφάλμα δ^l με το σφάλμα δ^{l+1} του επόμενου στρώματος $\delta^l = ((w^{l+1})^T \delta^{l+1}) \odot \sigma'(z^l)$
5. **Έξοδος y**
Η κλίση της συνάρτησης κόστους δίνεται από τον υπολογισμό των μερικών παραγώγων:

Εξίσωση για τον ρυθμό μεταβολής του κόστους σε σχέση με οποιοδήποτε βάρος στο δίκτυο $\frac{\partial C}{\partial w_{jk}^l} = \alpha_k^{l-1} \delta_j^l$

Εξίσωση για τον ρυθμό μεταβολής του κόστους σε σχέση με οποιαδήποτε πόλωση στο δίκτυο $\frac{\partial C}{\partial b_j^l} = \delta_j^l$

1.2.4.7 Κατάβαση Κλίσης (Gradient Descent)

Όπως έχει ήδη αναφερθεί, προκειμένου να επιτευχθεί η βελτιστοποίηση της μάθησης με οπισθοδιάδοση σφάλματος (back-propagation) είναι απαραίτητη η ελαχιστοποίηση μίας συνάρτησης κόστους $C(x)$ με μεταβολή του x . Η βασική ιδέα του αλγορίθμου οπισθοδιάδοσης είναι να προσαρμόζονται τα συναπτικά βάρη σε κάθε βήμα της επανάληψης, έτσι ώστε η συνάρτηση κόστους (cost function) να ελαχιστοποιείται.

Η συνάρτηση C ονομάζεται συνάρτηση ελαχίστων τετραγώνων (Mean Squared Error – MSE) που αντιστοιχεί στα σφάλματα, γνωστή και ως τετραγωνική (quadratic) συνάρτηση κόστους. Σκοπός του αλγορίθμου εκπαίδευσης είναι να βρει όλα τα βάρη και τις πολώσεις ώστε αυτή η συνάρτηση κόστους να τείνει στο μηδέν. Μία ευρέως διαδεδομένη διαδικασία μέσω της οποίας ικανοποιείται η παραπάνω συνθήκη ονομάζεται Μέθοδος Κατάβασης Κλίσης (Gradient Descent). Η κλίση της επιφάνειας σφάλματος σε οποιοδήποτε σημείο είναι ένα διάνυσμα που έχει κατεύθυνση προς την πιο απότομη κάθοδο (Steepest descent).

Στόχος της διαδικασίας αποτελεί η εύρεση του σημείου στο οποίο η συνάρτηση κόστους C επιτυγχάνει το ολικό ελάχιστο (global minimum). Αυτό προσεγγίζεται εύκολα με το να υπολογιστούν οι μερικές παραγώγους και να βρεθεί το σημείο στο οποίο η συνάρτηση κόστους έχει ακρότατο.

Ως συνάρτηση κόστους $C(w, b)$ ορίζεται η:

$$C(w, b) = \frac{1}{2n} \sum_x \|y(x) - a\|^2$$

όπου w είναι η συλλογή όλων των βαρών στο δίκτυο και b όλων των πολώσεων. Επίσης, n είναι ο συνολικός αριθμός των εισόδων εκπαίδευσης και a είναι το διάνυσμα των αληθών τιμών εξόδων του δικτύου στην τιμή εισόδου x .

Η μέθοδος κατάβασης δυναμικού ορίζει ως αναγκαία συνθήκη για το βέλτιστο v^* την:

$$\nabla C(v^*) = 0, \text{ όπου } v = v_1, v_2, \dots, v_n.$$

Υπάρχουν περιπτώσεις στις οποίες υπάρχουν πολλά τοπικά ελάχιστα στη συνάρτηση κόστους. Τέτοιες περιπτώσεις συναντώνται πολύ συχνά στους αλγορίθμους νευρωνικών δικτύων βαθιάς μάθησης. Σε αυτές τις περιπτώσεις η συνάρτηση κόστους μπορεί πολλές φορές να πάρει αρκετά μικρή τιμή σε ένα σημείο, χωρίς το σημείο αυτό να αποτελεί κατ' ανάγκη ολικό ελάχιστο.

Ισχύει ότι: $\Delta C(v) \approx \nabla C(v) * \Delta v$

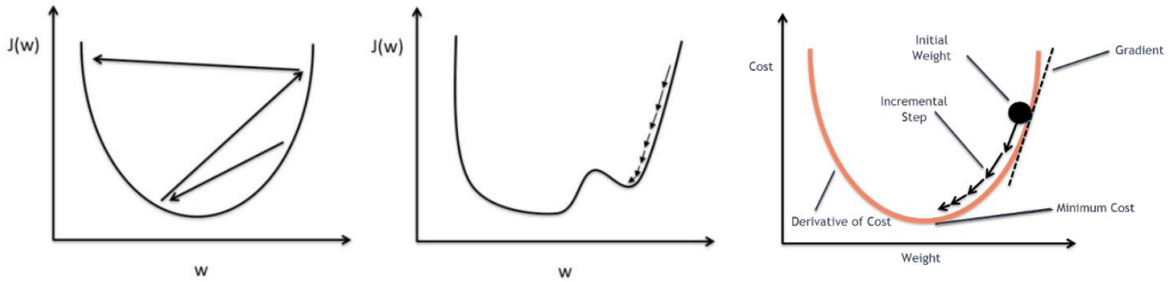
Η κατάβαση κλίσης (Gradient Descent), ξεκινάει από ένα τυχαίο σημείο και προχωράει διαδοχικά σε άλλα σημεία, τέτοια ώστε η συνάρτηση κόστους $C(v)$, να μειώνεται σε κάθε επανάληψη, σύμφωνα με τη σχέση:

$$C(v(k+1)) < C(v(k))$$

όπου $v(k)$ η τιμή του διανύσματος κατά το βήμα k και αντίστοιχα $v(k+1)$ η τιμή του διανύσματος κατά το βήμα $k+1$.

Η μέθοδος βασίζεται στην παρατήρηση ότι εάν μια πραγματική συνάρτηση είναι διαφορίσιμη σε ένα σημείο, τότε η τιμή της $\Delta C(v)$ μειώνεται γρήγορα όσο πλησιάζει στο ακρότατο. Ο κανόνας κατάβασης δυναμικού είναι $v \rightarrow v' = v - \eta \nabla C$, όπου η είναι ο ρυθμός μάθησης (learning rate). Ο ρυθμός μάθησης, ο οποίος στην πραγματικότητα δεν είναι άλλο από έναν συντελεστή που καθορίζει το βήμα που χρησιμοποιείται για την εύρεση του (τοπικού) ελάχιστου, θα πρέπει να είναι τέτοιος ώστε να μην είναι πολύ μικρός, αλλά ούτε και πολύ μεγάλος, προκειμένου να λειτουργήσει σωστά ο αλγόριθμος. Στην περίπτωση που ο ρυθμός είναι πολύ μικρός, ο αλγόριθμος εκπαιδεύεται πάρα πολύ αργά, ενώ στην περίπτωση που ο ρυθμός είναι αρκετά μεγάλος μπορεί να υπάρξει αδυναμία εύρεσης (τοπικού) ελάχιστου και άρα αδυναμία σύγκλησης (overshooting).

Η διαδικασία επαναλαμβάνεται μέχρι τη σύγκλιση σε ολικό ελάχιστο.



Αριστερά: μεγάλο learning rate (overshooting), μέση: μικρό learning rate, πολλές επαναλήψεις μέχρι τη σύγκλιση και εγκλωβισμός σε τοπικό ελάχιστο, δεξιά: κανονικό learning rate

Μια μέθοδος επιτάχυνσης της εκπαίδευσης είναι η μέθοδος της στοχαστικής κατάβασης κλίσης (stochastic gradient descent). Η μέθοδος αυτή χρησιμοποιεί ένα υποσύνολο των δεδομένων εκπαίδευσης προκειμένου να υπολογίσει την κλίση της συνάρτησης κόστους VC . Ο αριθμός των δεδομένων εκπαίδευσης που χρησιμοποιούνται σε μία επανάληψη του αλγορίθμου ονομάζεται μέγεθος παρτίδας (batch size), η επιλογή της οποίας είναι τυχαία. Εδώ, ο όρος "στοχαστικός" προέρχεται από το γεγονός ότι η κλίση (gradient) που βασίζεται σε ένα δείγμα εκπαίδευσης είναι μια "στοχαστική προσέγγιση" της "πραγματικής" κλίσης της συνάρτησης κόστους. Ο στοχαστικός αυτός χαρακτήρας της μεθόδου έχει το αποτέλεσμα της μείωσης της πιθανότητας η διαδικασία μάθησης να παγιδευτεί σε ένα τοπικό ελάχιστο. Παράλληλα όμως παρουσιάζει και μεγαλύτερη αστάθεια, λόγω του ότι οι ενημερώσεις γίνονται στην κατεύθυνση πολλών μικρών συνιστωσών της πραγματικής κλίσης (Παπαδόπουλος, 2018 [104]).

1.3 Εντοπισμός // Machine Learning Methods

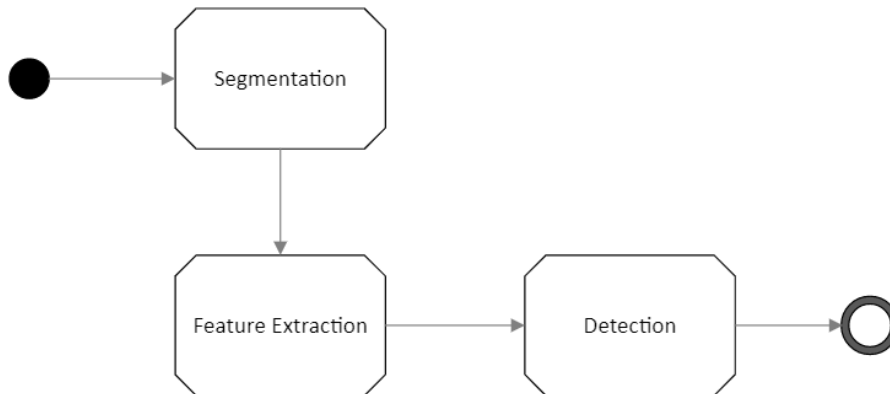
Στις προσεγγίσεις που περιγράφηκαν για τον εντοπισμό των σημάτων οδικής κυκλοφορίας, η προηγούμενη γνώση του προβλήματος (το αναμενόμενο χρώμα και το σχήμα μίας περιοχής ενδιαφέροντος ROI) «κωδικοποιείται» και τροφοδοτεί το μοντέλο με μεθόδους Επεξεργασίας Εικόνας σε ξεχωριστό στάδιο της λύσης. Ωστόσο, αυτή η γνώση θα μπορούσε επίσης να ανακαλυφθεί στο πλαίσιο ενός αλγορίθμου Μηχανικής Μάθησης.

Επιπλέον, οι προηγούμενες μέθοδοι μοιράζονται μια κοινή αδυναμία απέναντι στη διαχείριση δυσμενών συνθηκών, όπως είναι οι αλλαγές φωτισμού, οι αποκρύψεις, η αλλαγή κλίμακας, η στροφή και οι μετακινήσεις. Εντούτοις, αυτά τα προβλήματα μπορούν να αντιμετωπιστούν με μεγαλύτερη ευκολία με αλγορίθμους της Μηχανικής Μάθησης, αλλά απαιτείται μια ευρεία βάση δεδομένων προκειμένου να γίνει η εκπαίδευση των αλγορίθμων.

Στη βιβλιογραφία έχουν προταθεί διάφορες μέθοδοι όπως οι Μηχανές Διανυσματικής Στήριξης (SVM), η μέθοδος των Viola & Jones [51] με τον αλγόριθμο Adaboost, αλγόριθμοι γραμμικής ταξινόμησης για πάνω από δύο κλάσεις, Linear Discriminant Analysis κ.λπ. Ευρέως διαδεδομένα είναι επίσης τα Νευρωνικά Δίκτυα ως μέρος της οικογένειας αλγορίθμων Μηχανικής Μάθησης, πχ. Berger et al. [52]. Η State-of-state προσέγγιση αφορά αλγορίθμους βαθιάς μάθησης προκειμένου να επιτευχθεί το βέλτιστο αποτέλεσμα, όπως για παράδειγμα η μέθοδος των Xiong et al. (2016) [53], η οποία όμως θα αναλυθεί σε επόμενο κεφάλαιο.

Βιβλιογραφικά, σύμφωνα με τους Mukhometzianov & Wang (2017) [55], είναι δυνατόν να διακριθούν οι προσεγγίσεις με στόχο την ανίχνευση σε δύο ομάδες. Η πρώτη ομάδα περιλαμβάνει έρευνες που χρησιμοποιούν μεθόδους όπως αυτή των Viola & Jones, Μηχανές Διανυσματικής Στήριξης (SVM) κ.λπ. απευθείας σε εικόνες. Η δεύτερη ομάδα περιλαμβάνει μεθόδους που χρησιμοποιούν διαφορετικές τεχνικές προεπεξεργασίας για να εξαγάγουν χαρακτηριστικά και στη συνέχεια να τα χρησιμοποιήσουν ως δεδομένα εισαγωγής για μεθόδους ανίχνευσης με μηχανική μάθηση.

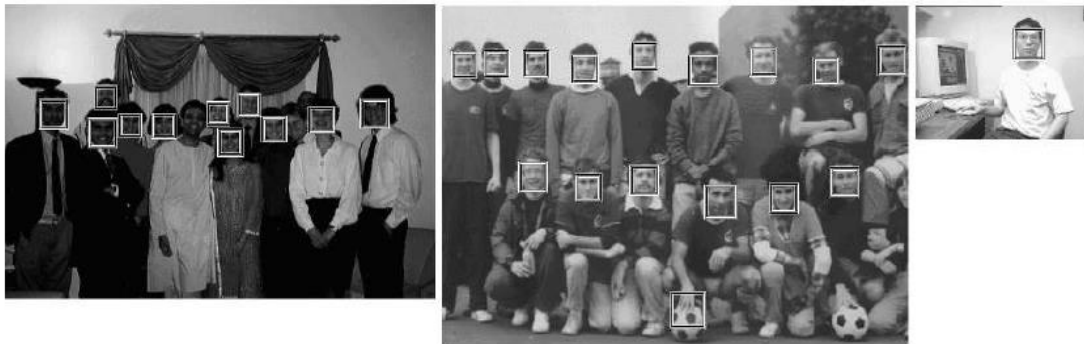
Σύμφωνα με τους Dewan & Vig et al. (2017) [54], ο διαχωρισμός των βιβλιογραφικών προσεγγίσεων μπορεί να γίνει με βάση το εάν εφαρμόζονται τεχνικές που είναι βασισμένες σε μεθόδους εκμάθησης χρώματος ή τεχνικές εκμάθησης σχήματος.



Η γενική ροή που ακολουθείται από τους τυπικούς αλγόριθμους ανίχνευσης σημάτων οδικής κυκλοφορίας ΠΗΓΗ: Møgelmoose et al (2012) [125]

1.3.1 Η προσέγγιση Viola-Jones

Μία από τις πιο δημοφιλείς προσεγγίσεις εκμάθησης σχήματος, που εξακολουθεί να χρησιμοποιείται ευρέως, είναι η προαναφερθείσα μέθοδος των Viola & Jones (2001) [51], που προτάθηκε για αξιόπιστη και σε πραγματικό χρόνο ανίχνευση αντικειμένων. Ειδικότερα, η μέθοδος αναπτύχθηκε και αφορά την ανίχνευση προσώπων (face detection).

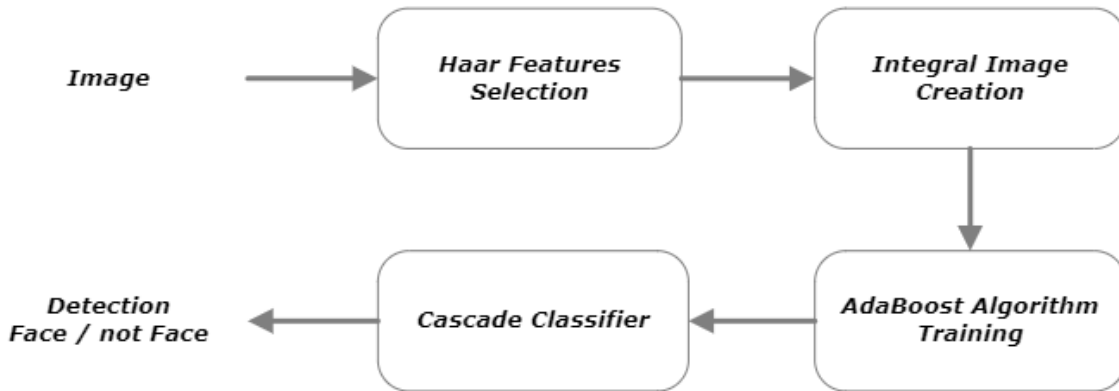


Αποτελέσματα ανίχνευσης προσώπου μας σε μια σειρά δοκιμαστικών εικόνων από το σετ δοκιμών MIT + CMU. ΠΗΓΗ: Viola & Jones (2001) [51]

Η προσέγγιση τους είναι γρήγορη και σχετικά απλή, τόσο που επιτρέπει την ανίχνευση προσώπου σε πραγματικό χρόνο και με χαμηλή υπολογιστική ισχύ. Για την περιγραφή των εικόνων χρησιμοποιούνται τα χαρακτηριστικά τύπου Haar (Haar features), τα οποία είναι παρόμοια με τα συνελκτικά φίλτρα Kernel. Η ταξινόμηση των υποψήφιων περιοχών της εικόνας γίνεται με τον αλγόριθμο AdaBoost. Την ολοκληρωμένη αυτή μέθοδο, παρουσιάζει και στη διπλωματική του εργασία ο Βαρυτιμίδης (2008) [3] για την ανίχνευση αντικειμένων και τον ημιαυτόματο χαρακτηρισμό εικόνων.

Λόγω του ότι, αφενός μεν, τα χαρακτηριστικά τύπου Haar παρουσιάζουν πολύ καλά αποτελέσματα ανίχνευσης, αφού η διαδικασία της ανίχνευσης αντικειμένων δεν επηρεάζεται από τις διαφορές στη φωτεινότητα από εικόνα σε εικόνα και, αφετέρου δε, απαιτούν πολύ λίγο χρόνο για την ολοκλήρωση της διαδικασίας, αναπτύχθηκαν και χρησιμοποιήθηκαν για πρώτη φορά με στόχο την ανίχνευση ανθρώπων από τους Oren et al. (1997) [4]. Στη συνέχεια, παρουσιάστηκε ένα γενικότερο πλαίσιο για ανίχνευση αντικειμένων από τους Parageorgiou et al. (1998) [5], με το οποίο ανιχνεύθηκαν άνθρωποι και πρόσωπα. Το 2001

οι Viola & Jones επέκτειναν τη μέθοδο όπου, πέρα από τα επιπλέον χαρακτηριστικά τύπου Haar που πρόσθεσαν, εισήγαγαν τη χρήση πινάκων προστιθέμενου εμβαδού (integral image) για το γρήγορο υπολογισμό των χαρακτηριστικών και τη χρήση του αλγορίθμου AdaBoost, όπως προαναφέρθηκε. Αργότερα, οι Lienhart & Maydt (2002) [6] επέκτειναν το σύστημα ανίχνευσης των Viola & Jones προσθέτοντας το σύνολο των χαρακτηριστικών τύπου Haar που έχουν στραφεί. Η υλοποίηση του συστήματος αυτού ενσωματώθηκε σε βιβλιοθήκη της OpenCV [7] και είναι διαθέσιμη ελεύθερα ως λογισμικό ανοιχτού κώδικα (Βαρυτιμίδης, 2008 [3]).



Διάγραμμα ροής αλγορίθμου Viola-Jones

Αναλυτικότερα, ο αλγόριθμος που προτάσσουν οι Viola & Jones (2001) [51] προκειμένου να επιτραπεί η γρήγορη και ακριβής ανίχνευση αποτελείται από 3 συστατικά μέρη (Yi-Qing Wang, 2014 [56]). Αρχικά, για κάθε εικόνα δημιουργείται το ολοκλήρωμα της εικόνας (integral image), πάνω στην οποία υπολογίζονται τα χαρακτηριστικά τύπου Haar που έχουν επιλεγεί, και στη συνέχεια επιλέγονται τα πιο ισχυρά χαρακτηριστικά με τη χρήση του αλγορίθμου AdaBoost. Τέλος, δημιουργούνται πολλοί και διαφορετικοί διαδοχικά συνδεδεμένοι απλοί ταξινομητές (Cascade Detectors) για την αύξηση της ταχύτητας της διαδικασίας, καταλήγοντας σε ένα ολοκληρωμένο σύστημα ανίχνευσης το οποίο λειτουργεί με μεγάλη απόδοση και σε πραγματικό χρόνο. Ακολουθεί εκτενέστερη περιγραφή των επιμέρους τεχνικών του αλγορίθμου.

1.3.1.1 Χαρακτηριστικά τύπου Haar & Εικόνα integral

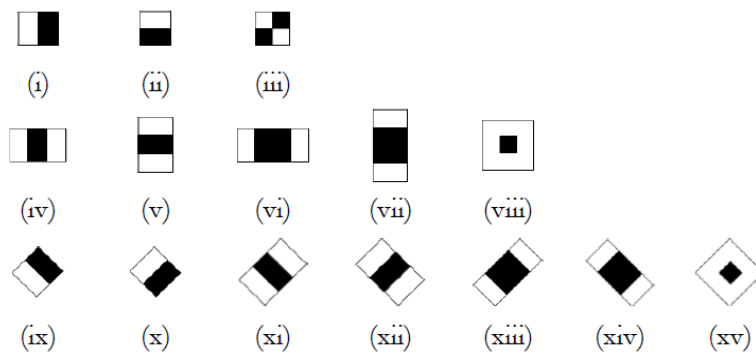
Για την αναπαράσταση του περιεχομένου μιας εικόνας και την άντληση πληροφορίας από αυτήν, με στόχο την ανίχνευση, απαιτείται πρώτα να γίνει η κατάλληλη επεξεργασία της εικόνας. Στη μέθοδο των Viola-Jones εξετάζονται τα χαρακτηριστικά τύπου Haar. Το πλήρες όνομα των χαρακτηριστικών είναι Haar - Wavelet αφού προκύπτουν από την εφαρμογή του μετασχηματισμού κυματιδίων (Discrete Wavelet Decomposition - DWT) σε μια εικόνα με χρήση των συναρτήσεων τύπου Haar. Παραδείγματα χαρακτηριστικών Haar φαίνονται στην επόμενη εικόνα.

Τα κλασικά χαρακτηριστικά Haar (edge features) είναι σχετικά απλά και μπορούν να εντοπίσουν οριζόντιες και κατακόρυφες ακμές, καθώς και διαγώνιες γραμμές, στην εικόνα (i, ii, iii). Για την αναπαράσταση γραμμών, ράβδων και τετραγώνων δίνουν καλύτερα αποτελέσματα τα χαρακτηριστικά τα iv, v, vi, vii και viii (line features), τα οποία υπολογίζονται χωρίς να αυξάνεται ιδιαίτερα η πολυπλοκότητα. Η αναπαράσταση των διαγώνιων σχημάτων βελτιώνεται αισθητά με τη χρήση των στραμμένων κατά 45° χαρακτηριστικών.

Η χρήση των συναρτήσεων Haar στο μετασχηματισμό Wavelet ξεκινά από την παρατήρηση ότι η τιμή της φωτεινότητας κάθε εικονοστοιχείου επηρεάζεται έντονα από τις αλλαγές στο φωτισμό της σκηνής. Αυτή η αλλαγή, όμως, επηρεάζει ομοιόμορφα όλα τα εικονοστοιχεία της

εικόνας. Έτσι, η τιμή μιας συνάρτησης που εξετάζει τη μέση διαφορά ανάμεσα σε δύο ή τρεις περιοχές της ίδιας εικόνας θα παραμένει σε μεγάλο βαθμό ανεπηρέαστη. Χρησιμοποιώντας, λοιπόν, τις συναρτήσεις Haar, η διαδικασία της ανίχνευσης αντικειμένων δεν επηρεάζεται από ομοιόμορφες μεταβολές φωτεινότητας από εικόνα σε εικόνα.

Οι συναρτήσεις Haar υπολογίζουν τη διαφορά ανάμεσα στους μέσους όρους των τιμών των εικονοστοιχείων δύο (ή τριών) περιοχών. Αν μία συνάρτηση Haar υπολογίζεται για ένα ορθογώνιο μισό άσπρο και μισό μαύρο (binary), τότε υπολογίζεται ο μέσος όρος των εικονοστοιχείων που βρίσκονται μέσα στο άσπρο ορθογώνιο καθώς και αυτών που βρίσκονται μέσα στο μαύρο ορθογώνιο. Έπειτα, ο μέσος όρος του μαύρου ορθογωνίου αφαιρείται από τον μέσο όρο του άσπρου. Η τιμή που προκύπτει αποτελεί την τιμή του χαρακτηριστικού Haar (Μαρινέλλης, 2018) [58].



Χαρακτηριστικά τύπου Haar (κλασικά χαρακτηριστικά, χαρακτηριστικά για γραμμές και χαρακτηριστικά στραμμένα κατά 45°)
 ΠΗΓΗ ΕΙΚΟΝΑΣ: Βαρυτιμίδης, 2008 [3]

Τα χαρακτηριστικά Haar αποτελούνται από δύο ή περισσότερες ορθογώνιες περιοχές που περικλείονται σε ένα πρότυπο. Η τιμή χαρακτηριστικού f ενός χαρακτηριστικού τύπου Haar που έχει k ορθογώνια λαμβάνεται όπως στην ακόλουθη εξίσωση:

$$f = \sum_{i=1}^k w^{(i)} \mu^{(i)}$$

όπου $\mu^{(i)}$ είναι η μέση ένταση των εικονοστοιχείων στην εικόνα x που περικλείεται από το ορθογώνιο και $w^{(i)}$ είναι το βάρος που αντιστοιχεί στο ορθογώνιο.

Παραδοσιακά, τα βάρη που αποδίδονται στα ορθογώνια ενός προτύπου Haar ορίζονται ως προεπιλεγμένοι ακέραιοι αριθμοί έτσι ώστε να ικανοποιείται η ακόλουθη εξίσωση:

$$\sum_{i=1}^k w^{(i)} = 0$$

Όσο πληθαίνουν τα χαρακτηριστικά τύπου Haar σε μια εικόνα τόσο αυξάνεται και η ποσότητα της πληροφορίας που αντλείται από αυτήν. Για να επιτευχθεί μια πολύ πιο πυκνή αναπαράσταση, οι συναρτήσεις Haar εφαρμόζονται κάθε φορά με μεταξύ τους απόσταση ενός εικονοστοιχείου. Επιπλέον, σε κάθε επανάληψη στο μετασχηματισμό Wavelet το μέγεθος των συναρτήσεων Haar κανονικά διπλασιάζεται. Εάν οριστεί ότι το μέγεθος των συναρτήσεων Haar θα αυξάνει κάθε φορά κατά ένα μόνο εικονοστοιχείο, τότε το σύνολο των χαρακτηριστικών Haar σε μία εικόνα γίνεται υπερπολλαπλάσιο του αρχικού.

Ο αριθμός των χαρακτηριστικών που προκύπτουν από την εφαρμογή κάθε Haar διαφέρει. Για τον υπολογισμό του πλήθους των χαρακτηριστικών Haar σε κάποιο παράθυρο εικόνας πλάτους W και ύψους H , ισχύουν τα παρακάτω:

όπου $X = \frac{W}{w}$ και $Y = \frac{H}{h}$ οι μέγιστοι συντελεστές κλίμακας σε πλάτος και ύψος, αντίστοιχα, που αυξάνεται σε κάθε επανάληψη. Το πλήθος των χαρακτηριστικών που προκύπτουν από την εφαρμογή ενός κατακόρυφου Haar $w \times h$ στο παράθυρο εικόνας είναι:

$$XY \left(W + 1 - w \frac{X + 1}{2} \right) \left(H + 1 - h \frac{Y + 1}{2} \right)$$

ενώ για τα στρεφόμενα χαρακτηριστικά Haar ισχύει:

$$XY \left(W + 1 - z \frac{X+1}{2} \right) \left(H + 1 - z \frac{Y+1}{2} \right), \text{ όπου } z = w + h$$

Τα χαρακτηριστικά μπορούν πολύ γρήγορα να υπολογιστούν σε γραμμικό χρόνο, με τη βοήθεια μιας νέας αναπαράστασης εικόνας που ονομάζεται ολοκληρωμένη εικόνας (Integral Image). Συχνά στη βιβλιογραφία αυτή αναφέρεται και ως Πίνακας Προστιθέμενου Εμβαδού (Summed Area Table).

Η ολοκληρωμένη εικόνα είναι μία εικόνα της οποίας κάθε εικονοστοιχείο περιέχει το άθροισμα όλων των προηγούμενων εικονοστοιχείων που βρίσκονται πάνω και αριστερά του τρέχοντος. Έτσι, ένα εικονοστοιχείο στη θέση $(0,0)$ σε μια ολοκληρωμένη εικόνα διατηρεί την τιμή του $(0,0)$, αλλά ένα εικονοστοιχείο στη θέση $(1,1)$ σε μια ολοκληρωμένη εικόνα περιέχει το άθροισμα των εικονοστοιχείων $(0,0)$, $(1,0)$, $(0,1)$ και $(1,1)$ της αρχικής εικόνας. Μαθηματικά αυτό σε ένα συγκεκριμένο σημείο (x,y) μπορεί να εκφραστεί ως:

$$ii(x,y) = \sum_{x'=0}^{x \leq x} \sum_{y'=0}^{y' \leq y} i(x',y')$$

όπου $ii(x,y)$ είναι η τιμή του ολοκληρώματος εικόνας και $i(x,y)$ είναι η τιμή της πραγματικής εικόνας. Χρησιμοποιώντας τις ακόλουθες σχέσεις:

$$s(x,y) = s(x,y - 1) + i(x,y)$$

$$ii(x,y) = ii(x - 1,y) + s(x,y)$$

όπου $s(x,y)$ είναι το άθροισμα των εικονοστοιχείων μιας γραμμής και

$$s(x,-1) = 0, ii(-1,y) = 0$$

παίρνει κανείς την έκφραση:

$$ii(x,y) = i(x,y) + ii(x,y - 1) + ii(x - 1,y) - ii(x - 1,y - 1)$$

Με τον τρόπο αυτό υπολογίζεται το ολοκλήρωμα εικόνας για όλα τα σημεία σε φυσικό χρόνο. Το άθροισμα των εικονοστοιχείων σε ένα ορθογώνιο μπορεί να υπολογιστεί με χρήση 4 αναφορών στον πίνακα. Όταν ένα ορθογώνιο μπορεί να υπολογιστεί με 4 αναφορές στον πίνακα, η διαφορά μεταξύ δύο ορθογωνίων μπορεί να υπολογιστεί με 8 αναφορές στον πίνακα. Επειδή όμως τα πιο απλά χαρακτηριστικά τύπου Haar μοιράζονται δύο γωνίες μεταξύ των ορθογωνίων, μπορούν να υπολογιστούν με μόνο 6 αναφορές. Τα χαρακτηριστικά μεσαίας λωρίδας μπορούν να υπολογιστούν σε 8 αναφορές. Έτσι, μπορεί να υπολογιστεί εξαιρετικά γρήγορα ένας μεγάλος αριθμός χαρακτηριστικών τύπου Haar δημιουργώντας μια ολοκληρωμένη εικόνα μία φορά και αναφερόμενη σε αυτήν για κάθε υπολογισμό χαρακτηριστικών.

Original	Integral	Original	Integral
5 2 3 4 1	5 7 10 14 15	5 2 3 4 1	5 7 10 14 15
1 5 4 2 3	6 13 20 26 30	1 5 4 2 3	6 13 20 26 30
2 2 1 3 4	8 17 25 34 42	2 2 1 3 4	8 17 25 34 42
3 5 6 4 5	11 25 39 52 65	3 5 6 4 5	11 25 39 52 65
4 1 3 2 6	15 30 47 62 81	4 1 3 2 6	15 30 47 62 81

$5 + 2 + 3 + 1 + 5 + 4 = 20$
 $5 + 4 + 2 + 2 + 1 + 3 = 17$
 $34 - 14 - 8 + 5 = 17$

Η τιμές της ολοκληρωμένης εικόνας για σημείο και ορθογώνιο με 6 αναφορές
 ΠΗΓΗ: <https://vinsol.com/blog/2016/06/28/computer-vision-face-detection/>

Η ανίχνευση του εκάστοτε αντικειμένου θα πρέπει να γίνει σε κάθε δυνατή θέση της εικόνας, καθώς επίσης και σε κάθε δυνατή κλίμακα. Με τον παραπάνω τρόπο, τα χαρακτηριστικά μπορούν όχι μόνο να υπολογιστούν γρήγορα αλλά και να αλλάξουν το μέγεθος τους εύκολα κατά την αναζήτηση αντικειμένων μεγαλύτερων από το αρχικό μέγεθος των χαρακτηριστικών. Για τον έλεγχο σε κάθε κλίμακα, άλλες μέθοδοι δημιουργούν μια πυραμίδα από μικρύνσεις της εικόνας και εφαρμόζουν σε κάθε κλίμακα της πυραμίδας τον ανιχνευτή, διατηρώντας σταθερό το μέγεθός του. Αντίθετα εδώ, αντί για την εφαρμογή σταθερού μεγέθους ανιχνευτή σε διαφορετικά επίπεδα πυραμίδας εικόνας, τα χαρακτηριστικά αλλάζουν το μέγεθος και αξιολογούνται για άλλη μια φορά στην ίδια ολοκληρωμένη εικόνα. Αυτό σημαίνει πως για έναν αμετάβλητο ανιχνευτή κλίμακας πρέπει να υπολογιστεί μόνο μία ολοκληρωμένη εικόνα και δεν απαιτείται αλλαγή μεγέθους. Αυτό είναι εφικτό, καθώς τα χαρακτηριστικά τύπου Haar μπορούν να μεταβληθούν σε μέγεθος.

Μετά τη συμβολή των Viola & Jones (2001) [51] έχουν προταθεί πολλά επιτυχημένα συστήματα που βασίζονται σε χαρακτηριστικά τύπου Haar. Οι Pavanì et al. (2010) [59] προτείνουν τροποποιημένα χαρακτηριστικά Haar, των οποίων τα ορθογώνια έχουν βέλτιστα βάρη, χωρίς τη χρήση του Πίνακα Προστιθέμενου Εμβαδού. Η τιμή των προτεινόμενων χαρακτηριστικών υπολογίζεται ακριβώς όπως τα κλασικά, εκτός από το ότι τα προεπιλεγμένα βάρη των ορθογώνιων του, $w^{(i)}$, αντικαθίστανται με βέλτιστες τιμές, $\hat{w}^{(i)}$. Επομένως η σχέση μετατρέπεται ως εξής:

$$f = \sum_{i=1}^k \hat{w}^{(i)} \mu^{(i)}$$

Τα αποτελέσματα δείχνουν ότι οι ανιχνευτές αντικειμένων που βασίζονται στα προτεινόμενα χαρακτηριστικά είναι πιο ακριβείς και ταχύτεροι από τους ανιχνευτές αντικειμένων που έχουν κατασκευαστεί με παραδοσιακά χαρακτηριστικά τύπου Haar, ενώ παράλληλα διατηρούν την απλότητα των παραδοσιακών χαρακτηριστικών.

1.3.1.2 Ο αλγόριθμος AdaBoost

Στη μέθοδο Viola & Jones (2001) [51] τα χαρακτηριστικά μπορούν να υπολογιστούν πάρα πολύ γρήγορα. Η ελάχιστη ανάλυση του ανιχνευτή τους έχει οριστεί στα 24x24. Επομένως, το σύνολο των χαρακτηριστικών ορθογώνιων είναι αρκετά μεγάλο, πάνω από 180.000, όπως αναφέρουν ενδεικτικά οι συγγραφείς. Συνεπώς, η διαδικασία υπολογισμού του συνόλου των χαρακτηριστικών για όλα τα παράθυρα μέσα σε μια εικόνα παραμένει χρονοβόρα. Για το λόγο αυτό χρησιμοποιήθηκε ο αλγόριθμος AdaBoost, τόσο για την επιλογή των χαρακτηριστικών όσο και για την εκπαίδευση του ταξινομητή.

Ο αλγόριθμος εκμάθησης AdaBoost είναι ένας αλγόριθμος της Μηχανικής Μάθησης. Ανήκει στην κατηγορία των boosting algorithms, χρησιμοποιείται δηλαδή για να αυξήσει την απόδοση ενός οποιουδήποτε απλού αλγορίθμου ταξινόμησης (weak classifier). Γενικά, ως

απλοί ή αδύναμοι ταξινομητές αναφέρονται αυτοί που από μόνοι τους δεν καταφέρνουν να ταξινομήσουν ορθά ένα σύνολο δειγμάτων, δηλαδή η ακρίβειά τους είναι λίγο μεγαλύτερη από εκείνη της τυχαίας ταξινόμησης. Ο AdaBoost ουσιαστικά είναι μια τεχνική ταξινόμησης για την κατασκευή ενός ισχυρού ταξινομητή, συνδυάζοντας πολλούς απλούς ταξινομητές. Με αυτόν τον τρόπο βελτιώνεται η ακρίβεια ταξινόμησης αυξάνοντας τα βάρη των εσφαλμένων ταξινομημένων δεδομένων.

Η βελτίωση του ασθενούς αλγόριθμου ταξινόμησης πραγματοποιείται με το να καλείται ο αλγόριθμος να επιλύσει μια αλληλουχία προβλημάτων ταξινόμησης. Αρχικά, όλα τα δεδομένα (θετικά και αρνητικά) παίρνουν μια τιμή βάρους, η οποία είναι ίδια για όλα. Αυτές οι εισοδοί δίνονται στον αλγόριθμο και πραγματοποιείται ο πρώτος κύκλος εκμάθησης, όπου ο αλγόριθμος ταξινομεί όλα τα δεδομένα με κάθε διαθέσιμη συνάρτηση ταξινόμησης. Έπειτα, οι συναρτήσεις ταξινόμησης διατάσσονται σύμφωνα με τα αποτελέσματά τους λαμβάνοντας υπόψη το βάρος κάθε δεδομένου. Επιλέγεται ένας μικρός αριθμός συναρτήσεων ταξινόμησης από αυτές με τα καλύτερα αποτελέσματα, που αποτελούν τον πρώτο αδύναμο ταξινομητή. Ο πρώτος κύκλος εκμάθησης ολοκληρώνεται και τα βάρη των δεδομένων ισοσταθμίζονται, δίνοντας μεγαλύτερο βάρος στα δεδομένα που ταξινομήθηκαν λανθασμένα από τον πρώτο ασθενή ταξινομητή. Έτσι, στον δεύτερο κύκλο εκμάθησης ο αλγόριθμος ταξινόμησης θα θεωρήσει πιο σημαντικά τα δεδομένα που ταξινομήθηκαν λανθασμένα από τον προηγούμενο ταξινομητή. Τα βήματα επαναλαμβάνονται διαδοχικά, μέχρι το επίπεδο του συνολικού λόγου λανθασμένης ταξινόμησης να ελαχιστοποιηθεί. Κάθε γύρος ενίσχυσης επιλέγει ένα χαρακτηριστικό από τα 180.000 πιθανά χαρακτηριστικά. Τελικά, ο ισχυρός ταξινομητής προκύπτει από τον συνδυασμό των αδύναμων ταξινομητών που επιλέχθηκαν και ένα κατώφλι. Κατά την διαδικασία της ταξινόμησης ενός υποπαραθύρου εικόνας από τον ισχυρό ταξινομητή, εφαρμόζονται στο υποπάρθυρο όλοι οι αδύναμοι ταξινομητές. Τα αποτελέσματα των αδύναμων ταξινομητών αθροίζονται, και αν το άθροισμα ξεπερνά το κατώφλι του ταξινομητή, το υπό εξέταση αντικείμενο ταξινομείται ως θετικό, αλλιώς ως αρνητικό (Μαρινέλλης, 2018 [58]). Ο αλγόριθμος των Viola & Jones για την υλοποίηση του AdaBoost για την εκπαίδευση των ταξινομητών παρατίθεται παρακάτω:

- Έστω σύστημα από παραδείγματα / δεδομένα εικόνων $(x_1, y_1) \dots (x_n, y_n)$, όπου $y_i = 0, 1$ για αρνητικά (negative) και θετικά (positive) παραδείγματα, αντίστοιχα.
- Αρχικοποίηση των βαρών $w_{1,i} = \frac{1}{2m}, \frac{1}{2l}$ για $y = 0, 1$ αντίστοιχα, όπου m και l ο αριθμός των αρνητικών και των θετικών, αντίστοιχα.
- Για $t = 1, \dots, T$:
 1. Κανονικοποίηση των βαρών,

$$w_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{j=1}^n w_{t,j}}$$

έτσι ώστε η παράμετρος w_t να αναπαριστά μια κατανομή πιθανότητας

2. Για κάθε χαρακτηριστικό j γίνεται εκπαίδευση ενός ταξινομητή h_j , ο οποίος είναι περιορισμένος να χρησιμοποιεί μόνο ένα χαρακτηριστικό. Το ποσοστό σφάλματος του ταξινομητή (ϵ_j) εκτιμάται μέσω της σχέσης:

$$\epsilon_j = \sum_i |h_j(x_i) - y_i|$$

3. Επιλογή του ταξινομητή h_t με το μικρότερο ποσοστό σφάλματος ϵ_t
4. Ενημέρωση των βαρών:

$$w_{t+1,i} = w_{t,i} \beta_t^{1-e_i}$$

όπου $e_i = 0$ εάν το παράδειγμα x_i είναι ορθώς ταξινομημένο,

$$\text{διαφορετικά } e_i = 1 \text{ και } \beta_t = \frac{e_t}{1-e_t}$$

- Ο τελικός ισχυρός ταξινομητής είναι:

$$h(x) = \begin{cases} 1 & \sum_{t=1}^T a_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^T a_t \\ 0 & \text{σε άλλη περίπτωση} \end{cases} \quad \text{όπου, } a_i = \log\left(\frac{1}{\beta_t}\right)$$

Συχνό πρόβλημα στη Μηχανική μάθηση αποτελεί η «κατάρα» της διαστατικότητας (curse of dimensionality). Ο όρος εισήχθη το 1961 από τον Bellman και αναφέρεται στο πρόβλημα της ανάλυσης δεδομένων πολλών μεταβλητών καθώς αυξάνει η διάσταση. Αυτό πρακτικά σημαίνει ότι για δεδομένο αριθμό δειγμάτων υπάρχει μια μέγιστη διάσταση των διανυσμάτων χαρακτηριστικών πάνω από την οποία η απόδοση του ταξινομητή θα μειώνεται. Επιπλέον, μπορεί να μειώνεται όχι μόνο η απόδοση του ταξινομητή από άποψη ταχύτητας εκπαίδευσης και εκτέλεσης, αλλά στην πραγματικότητα να μειώνεται και η προγνωστική ισχύς του.

Σε αντίθεση με τα νευρωνικά δίκτυα και τις μηχανές διανυσματικής μάθησης, η διαδικασία εκπαίδευσης AdaBoost επιλέγει μόνο εκείνες τις λειτουργίες που είναι γνωστό ότι βελτιώνουν την προγνωστική ισχύ του μοντέλου, μειώνοντας τη διάσταση και δυνητικά βελτιώνοντας τον χρόνο εκτέλεσης, καθώς δεν χρειάζεται να υπολογιστούν άσχετα χαρακτηριστικά. Έτσι, η διαδικασία της εκμάθησης πραγματοποιείται σχεδόν σε πραγματικό χρόνο.

Υπάρχουν τέσσερις διαφορετικές εκδοχές του αλγορίθμου AdaBoost. Η αρχική εκδοχή ονομάζεται Διακριτός AdaBoost (Discrete AdaBoost – DAB), καθώς η συνάρτηση ταξινόμησης κάθε αδύναμου ταξινομητή παίρνει μόνο δύο διακριτές τιμές, τις $-1, 1$ ανάλογα με το αν ένα δείγμα ταξινομείται ως θετικό ή αρνητικό. Η δεύτερη εκδοχή ονομάζεται Πραγματικός AdaBoost (Real AdaBoost – RAB), καθώς η συνάρτηση ταξινόμησης έχει ως πεδίο τιμών ολόκληρο το διάστημα $[0,1]$. Άλλη εκδοχή είναι ο Ήπιος AdaBoost (Gentle AdaBoost – GAB), ο οποίος βασίζεται στον Πραγματικό AdaBoost, παράγει όμως τα επιμέρους βήματα χρησιμοποιώντας τη μέθοδο Newton αντί να χρησιμοποιεί ακριβή υπολογισμό σε κάθε βήμα. Τέλος, υπάρχει και ο LogitBoost, ο οποίος έχει δύο παραλλαγές, αυτή που χρησιμοποιεί δύο κλάσεις και αυτή που χρησιμοποιεί J κλάσεις.

1.3.1.3 Ανιχνευτής Cascade

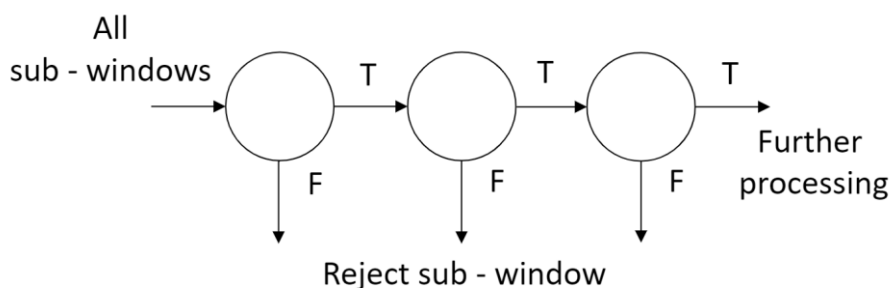
Οι ταξινομητές που δημιουργούνται με τον αλγόριθμο AdaBoost παρέχουν ιδιαίτερα υψηλή ακρίβεια για την ανίχνευση του αντικειμένου, όμως στην περίπτωση του ενός ανιχνευτή είναι αρκετά αργοί για χρήση σε εφαρμογή πραγματικού χρόνου. Η βασική ιδέα είναι ότι μπορούν να κατασκευαστούν μικρότεροι και κατά συνέπεια πιο αποδοτικοί και γρήγοροι ανιχνευτές.

Η μέθοδος του ανιχνευτή Cascade Detector αντί για έναν μεγάλο και χρονοβόρο ανιχνευτή αποτελείται από επίπεδα (stages), όπου το καθένα είναι ένας ισχυρός ανιχνευτής. Επομένως, χρησιμοποιεί διαδοχικά πολλούς μικρότερους, ενισχυμένους (boosted) και άρα πιο γρήγορους. Στην πραγματικότητα πρόκειται για πολλούς μικρούς ταξινομητές, οι οποίοι τελικά απορρίπτουν τις εικόνες στις οποίες δεν υπάρχει το αντικείμενο ενδιαφέροντος.

Ο στόχος του κάθε επιπέδου είναι να καθορίσει κατά πόσο ένα δοθέν παράθυρο εισόδου είναι σίγουρα πρόσωπο ή όχι. Αν το παράθυρο ταξινομηθεί ότι δεν περιέχει πρόσωπο από κάποιο επίπεδο, τότε απορρίπτεται. Αν όμως ένα παράθυρο ταξινομηθεί ως πιθανό πρόσωπο, τότε συνεχίζει στο επόμενο επίπεδο του ταξινομητή Cascade. Είναι προφανές ότι όσα περισσότερα επίπεδα καταφέρει να περάσει ένα παράθυρο χωρίς να απορριφθεί τόσο αυξάνει η πιθανότητα να περιλαμβάνει τελικά πρόσωπο. Αρχικά, χρησιμοποιούνται απλούστεροι ταξινομητές και πολύ ταχύτεροι, οι οποίοι απορρίπτουν γρήγορα την πλειονότητα των αρνητικών εισόδων με πολύ λίγη επεξεργασία. Αυτό συμβαίνει καθώς ελέγχουν λιγότερα χαρακτηριστικά Haar. Αργότερα, γίνεται χρήση πιο σύνθετων και χρονοβόρων που ελέγχουν περισσότερα χαρακτηριστικά Haar, ώστε να μειωθεί ο λόγος των λανθασμένων ανιχνεύσεων. Ένας Cascade ταξινόμησης αποτελεί στην πραγματικότητα ένα εκφυλισμένο δέντρο αποφάσεων (Degenerated Decision Tree).

Πιο αναλυτικά ο τρόπος λειτουργίας του ανιχνευτή είναι ως εξής:

- Η εικόνα εισόδου χωρίζεται σε πολλαπλά παράθυρα.
- Κάθε παράθυρο αποτελεί είσοδο του ανιχνευτή Cascade.
- Σε κάθε επίπεδο, το παράθυρο ελέγχεται εάν περιέχει το αντικείμενο ενδιαφέροντος ή όχι και το ταξινομεί.
- Εάν το παράθυρο είναι αρνητικό, τότε απορρίπτεται και τα βήματα επαναλαμβάνονται για άλλο παράθυρο.
- Εάν το παράθυρο είναι θετικό, τότε είναι πολύ πιθανόν να περιέχεται το αντικείμενο ενδιαφέροντος, επομένως θα περάσει στο επόμενο στάδιο με έναν πιο ισχυρό/χρονοβόρο ταξινομητή, ο οποίος ελέγχει περισσότερα χαρακτηριστικά Haar.
- Εάν το παράθυρο περάσει όλα τα στάδια του ταξινομητή σημαίνει πως περιέχει το αντικείμενο.



Σχηματική απεικόνιση ενός ανιχνευτή cascade
ΠΗΓΗ: Viola & Jones (2001) [51]

1.3.1.4 Η μέθοδος στον εντοπισμό σημάτων

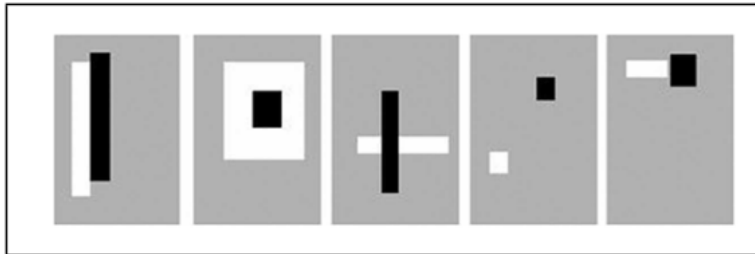
Η έρευνα των Viola & Jones αποτελεί ορόσημο στο επιστημονικό πεδίο της Όρασης Υπολογιστών. Διάφοροι άλλοι ερευνητές έχουν εφαρμόσει επιτυχώς τον συγκεκριμένο αλγόριθμο ανίχνευσης και παραλλαγές του, σε πολλές άλλες κατηγορίες αντικειμένων. Μεταξύ άλλων, εντοπίστηκαν επιτυχώς και οι πινακίδες οδικής κυκλοφορίας.



Αποσύνθεση δεύτερου επιπέδου εικόνας Stop Stop με χρήση Haar Wavelet
ΠΗΓΗ: Escalera et al. (2011) [11]

Οι Baro et al. (2009) [63] παρουσιάζουν μία νέα προσέγγιση ενισχυμένου (boosting) αλγορίθμου, ο οποίος όχι μόνο μειώνει δραστικά τον χρόνο εκπαίδευσης αλλά επιτρέπει επίσης τη χρήση πολύ μεγάλου όγκου χαρακτηριστικών. Προτείνουν έναν ανιχνευτή attentional cascade, ο οποίος αποτελείται από ένα σύνολο ταξινομητών. Κάθε είσοδος του ανιχνευτή είναι η περιοχή ενδιαφέροντος που έχει ανιχνευθεί από τον προηγούμενο ταξινομητή και ο αλγόριθμος AdaBoost χρησιμοποιείται για την εκπαίδευση των ταξινομητών αυτών.

Πιο συγκεκριμένα, οι ερευνητές κάνουν χρήση των διαχωρισμένων δίπολων (dissociated dipoles), ενός πιο γενικού τύπου χαρακτηριστικών από τα χαρακτηριστικά τύπου Haar. Τα διαχωρισμένα δίπολα έχουν παρουσιαστεί από τους Balas & Sinha (2003) [64] και είναι ένα είδος χαρακτηριστικών που αποτελείται από ένα ζεύγος ορθογώνιων περιοχών, που ονομάζονται διεγερτικό δίπολο (excitatory) και το ανασταλτικό δίπολο (inhibitory), αντίστοιχα. Η μαύρη περιοχή αντιστοιχεί στο ανασταλτικό δίπολο και η λευκή στο διεγερτικό δίπολο. Η μέση τιμή όλων των εικονοστοιχείων στο ανασταλτικό δίπολο αφαιρείται από τη μέση τιμή των εικονοστοιχείων των διεγερτικών δίπολων. Όπως στην περίπτωση των χαρακτηριστικών Haar, χρησιμοποιείται η ολοκληρωμένη εικόνα (integral image) για τον υπολογισμό του αθροίσματος των εικονοστοιχείων εντός των ορθογώνιων περιοχών.



Διαχωρισμένα δίπολα.
ΠΗΓΗ: Balas & Sinha (2003) [64]

Τα διαχωρισμένα δίπολα καθιστούν την κλασική προσέγγιση υπολογιστικά ανέφικτη, καθώς αυξάνουν κατά πάρα πολύ τα χαρακτηριστικά. Προκειμένου να αντιμετωπιστεί αυτός ο περιορισμός ως προς τη διάσταση του χώρου των χαρακτηριστικών, ορίστηκε μια εξελικτική προσέγγιση του AdaBoost.

Ο εξελικτικός (evolutionary) αλγόριθμος AdaBoost εκπαιδεύεται πάνω στα στάδια του ανιχνευτή cascade, χρησιμοποιώντας ένα σύνολο $\langle (x_1, y_1), \dots, (x_m, y_m) \rangle$ δειγμάτων που ταξινομήθηκαν ως θετικά δείγματα στα προηγούμενα στάδια του ανιχνευτή. Ουσιαστικά ο αλγόριθμος evolutionary AdaBoost χρησιμοποιεί επαναληπτικά έναν γενετικό αλγόριθμο για την ελαχιστοποίηση του σταθμισμένου σφάλματος και για τη θεσμοθέτηση των παραμέτρων ενός νέου αδύναμου ταξινομητή που προστίθεται στο τελικό σύνολο. Ο εξελικτικός αδύναμος ταξινομητής ελαχιστοποιεί τη συνάρτηση σταθμισμένου σφάλματος ε του AdaBoost ως:

$$\varepsilon = \sum_{i:h(x_i) \neq y_i} w_i$$

όπου $X = \{(x_i, y_i) | i = 1 : m\}$ είναι τα ζεύγη δείγμα-ετικέτα που συνθέτουν τα δεδομένα εκπαίδευσης, $W = \{w_1, w_2, \dots, w_m\}$ είναι η κατανομή βαρών AdaBoost πάνω στα δεδομένα εκπαίδευσης και το $h(x_i)$ αντιστοιχεί στην ετικέτα που προβλέπεται από την υπόθεση h για το αντικείμενο εκπαίδευσης x_i .

Αυτή η αρχιτεκτονική επιτρέπει την εύκολη απόρριψη περιοχών χωρίς αντικείμενο με χαμηλό υπολογιστικό κόστος, ενώ οι πιο πολύπλοκες περιοχές αναλύονται σε βάθος. Παρέχει υψηλή απόδοση και ακρίβεια και παρέχει δυναμικά καλύτερα αποτελέσματα σε περιπτώσεις θορύβου, παραμορφώσεων, μερικών αποκρύψεων και μειωμένου φωτισμού.

Οι Brkić et al. (2009) [74] εφαρμόζουν τον ανιχνευτή αντικειμένων Viola-Jones για την ανίχνευση τριγωνικών προειδοποιητικών σημάτων. Το πρώτο στάδιο του ανιχνευτή cascade

χρησιμοποιεί χαρακτηριστικά εύκολα ερμηνεύσιμα από άνθρωπο: ένα χαρακτηριστικό ευαίσθητο στο κάτω άκρο της πινακίδας και τρία χαρακτηριστικά ευαίσθητα στις αλλαγές ακμών κοντά στην κορυφή της πινακίδας.

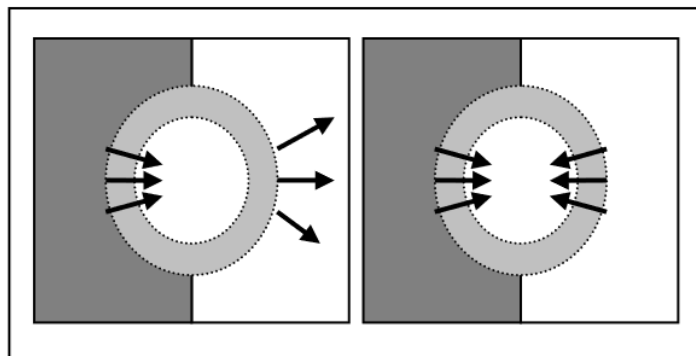


Τα πρώτα τέσσερα χαρακτηριστικά στο πρώτο επίπεδο του cascade ανιχνευτή
ΠΗΓΗ: Brkić et al. (2009) [74]

Η ανίχνευση των σημάτων οδικής κυκλοφορίας στους Priscariu et al. (2010) [62] πραγματοποιείται επίσης με ταξινομητές cascade, οι οποίοι ακολουθούνται από μια μηχανή διανυσματικής στήριξης που λειτουργεί σε κανονικοποιημένα κανάλια RGB, πυραμίδες HOG και χαρακτηριστικά σημεία επιλεγμένα από τον αλγόριθμο AdaBoost.

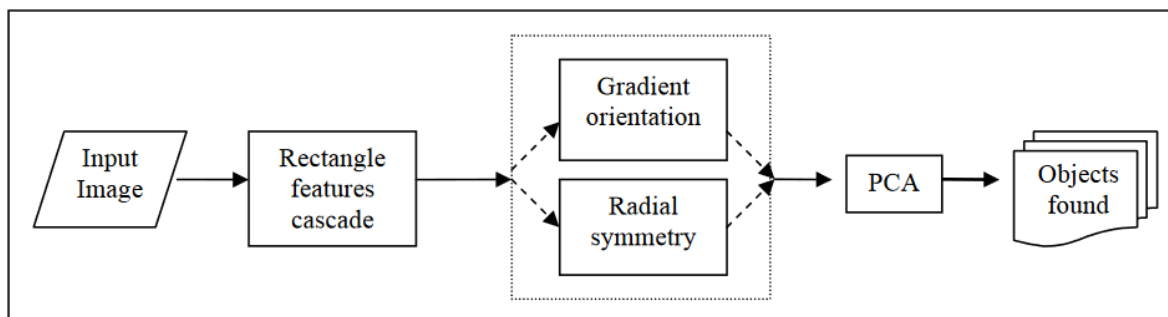
Στην εργασία των Baro & Vitria (2012) [18], προκειμένου να εντοπιστούν πινακίδες κυκλοφορίας σε εικόνες Geonap, η πρώτη προσέγγιση ήταν να προσαρμοστεί το σύστημα ανίχνευσης αντικειμένων που πρότειναν οι Viola & Jones και να γενικευτεί ώστε να ανιχνεύονται τα απλά σχήματα των πινακίδων αντί για πρόσωπα. Το πρόβλημα της μεθόδου ήταν πως έπρεπε να εκπαιδευτούν πολλά στάδια προκειμένου να υπάρχει μια λογική αναλογία ψευδών σφαλμάτων, αλλά από ένα συγκεκριμένο στάδιο αρχίζει να ειδικεύεται στο σετ δεδομένων της εκπαίδευσης με αποτέλεσμα να αποτυγχάνει στα δεδομένα της δοκιμής (over fitting).

Η λύση που υιοθετήθηκε ήταν ο συνδυασμός ενός ανιχνευτή cascade με λιγότερα στάδια με άλλες μεθόδους που εξαλείφουν τα ψευδή σφάλματα. Το αποτέλεσμα είναι ένας ετερογενής ταξινομητής cascade. Η εικόνα εισόδου πηγαίνει κατευθείαν στον ορθογώνιο ταξινομητή cascade. Σε κάθε ένα στάδιο αφαιρούνται μερικά αρνητικά υποπαράθυρα, ενώ τα θετικά περνούν όλα τα στάδια. Μετά τον ανιχνευτή, χρησιμοποιούνται διαφορετικά είδη χαρακτηριστικών για κάθε τύπο πινακίδας. Πιο συγκεκριμένα, χρησιμοποιούνται οι αλλαγές προσανατολισμού κλίσης (changes in gradient orientation) για εξαγωγή χαρακτηριστικών σημείων από τις τριγωνικές πινακίδες και επιπλέον η ακτινική συμμετρία (radial symmetry) για τα κυκλικά σήματα οδικής κυκλοφορίας, τα οποία είναι δύσκολο να περιγραφούν με τα ορθογώνια χαρακτηριστικά τύπου Haar.



Αλλαγές προσανατολισμού κλίσης.
ΠΗΓΗ: Baro & Vitria (2012) [18]

Στο τέλος του ταξινομητή εξαλείφονται όλα τα ψευδή σφάλματα με την ανάλυση κύριων συνιστωσών (PCA). Αυτή η μέθοδος χρησιμοποιεί καθολικές πληροφορίες, αντί για τις τοπικές πληροφορίες που χρησιμοποιήθηκαν στα προηγούμενα στάδια, ώστε να εντοπιστούν τελικά οι πινακίδες από κάθε τύπο.



Η δομή του ετερογενή ταξινομητή cascade
 ΠΗΓΗ: Baro & Vitria (2012) [18]

Οι Jeon et. al. (2014) [61] έκαναν χρήση των χαρακτηριστικών Haar για ανίχνευση πινακίδων ορίου ταχύτητας σε πραγματικό χρόνο. Εστίασαν κυρίως στην ακρίβεια και την ταχύτητα της μεθόδου που προτείνουν. Εκπαιδεύουν έναν ταξινομητή χρησιμοποιώντας τον Ήπιο AdaBoost (Gentle AdaBoost). Σε αντίθεση με τον Πραγματικό AdaBoost (Real AdaBoost) που θα μπορούσε να προβλέψει την αξιοπιστία, ο Gentle AdaBoost εμφανίζει καλή απόδοση στην ταξινόμηση μεταξύ του αντικειμένου και του φόντου, αφού μειώνει και διορθώνει το βάρος των λανθασμένων ταξινομημένων δειγμάτων. Για τη βελτίωση του ποσοστού ανίχνευσης εφαρμόζονται φίλτρα Kernel και η μέθοδος ισοστάθμισης ιστογράμματος (histogram equalization) στα δείγματα εκπαίδευσης.

Σε μια συγκριτική μελέτη που έκαναν οι Das & Mohammadi (2015) [62] για να αξιολογήσουν τη σκοπιμότητα αλγορίθμων της ανίχνευσης και της αναγνώρισης των σημάτων οδικής κυκλοφορίας, μελέτησαν μεταξύ άλλων και την ανίχνευση με χαρακτηριστικά Haar και ανιχνευτή cascade. Συγκεκριμένα, μελέτησαν την ταξινόμηση Haar - Cascade, την ταξινόμηση με βάση το χρώμα, την εξαγωγή χαρακτηριστικών με τη μέθοδο HoG, την ταξινόμηση με μηχανή διανυσματικής στήριξης (SVM) και τα πολυεπίπεδα νευρωνικά δίκτυα Perceptron (Multi-layer Perceptron classifier - MLP), καθώς και συνδυασμούς των μεθόδων αυτών. Τα καλύτερα αποτελέσματα επιτεύχθηκαν όταν συνδυάστηκαν μαζί οι ταξινομητές Haar και οι ταξινομητές Perceptron πολλαπλών επιπέδων. Ο συνδυασμός των δύο κατηγοριών επέτρεψε ακρίβεια ανίχνευσης περίπου 80% και ακρίβεια αναγνώρισης 99.9%.

Οι Chen et al. (2016) [67] εξέτασαν τόσο τον αλγόριθμο AdaBoost όσο και τα Μοντέλα Παλινδρόμησης Διανυσμάτων Υποστήριξης (SVR) για την ανίχνευση των πινακίδων κυκλοφορίας. Η προτεινόμενη τεχνική αξιολογείται σε τρία σύνολα δεδομένων (German - GTSD, Belgium - BTSD, Swedish - STSD), αλλά αυτή η προσέγγιση δεν είναι σε πραγματικό χρόνο.

1.3.2 Τεχνικές εκμάθησης σχήματος

Οι αλγόριθμοι μηχανικής μάθησης όπως οι Μηχανές Διανυσματικής Στήριξης (Support Vector Machines - SVMs) και τα Νευρωνικά Δίκτυα (Neural Networks - NNs) μπορούν να χρησιμοποιηθούν για την περιγραφή και εξαγωγή σχήματος. Οι Μηχανές Διανυσματικής Στήριξης και τα Νευρωνικά Δίκτυα χρησιμοποιούνται για την ανίχνευση σημάτων οδικής κυκλοφορίας λόγω της ικανότητάς τους να εντοπίζουν με ακρίβεια τα σχήματα. Τα Νευρωνικά Δίκτυα εκπαιδεύονται για κάθε ένα από τα διαφορετικά σύνολα πινακίδων. Ωστόσο, η προσθήκη περισσότερων πινακίδων συνεπάγεται και πάλι εκπαίδευση του δικτύου και ως εκ τούτου χειροκίνητη επιλογή δειγμάτων εκπαίδευσης. Οι Μηχανές Διανυσματικής Στήριξης παρουσιάζουν ανοχή στην στροφή, τη μετάθεση και στις μερικές αποκρύψεις των σημάτων οδικής κυκλοφορίας (Dewan et al., 2017 [54]).

Μια αρχιτεκτονική για τον εντοπισμό των κυκλικών πινακίδων οδικής σημασίας παρουσιάζουν οι Hoferlin et al. (2009) [68]. Αποτελείται από δύο πολυεπίπεδα Perceptrons (MLP). Το πρώτο επίπεδο χρησιμοποιεί ως εισαγωγή χαρακτηριστικά εξαγόμενα με SIFT και το δεύτερο χρησιμοποιεί χαρακτηριστικά εξαγόμενα με SURF. Η απόδοση του συστήματος που περιέχει 133 πινακίδες κυκλοφορίας ανήλθε σε ποσοστό επιτυχίας 96.4%.

Από τις πιο δημοφιλείς κατηγορίες αλγορίθμων που χρησιμοποιούνται ευρέως σε συστήματα ανίχνευσης σημάτων κυκλοφορίας είναι οι προαναφερθείσες Μηχανές Διανυσματικής Στήριξης (SVM). Οι Zaklouta et al. (2011) [69] χρησιμοποιούν τη μέθοδο HOG που πρότειναν οι Dalal & Triggs (2005) [8] για ανίχνευση πεζών λόγω της ανοχής σε αλλαγές κλίμακας. Οι περιγραφείς HOG υπολογίζονται και χρησιμοποιούνται ως χαρακτηριστικά για την εκπαίδευση ενός γραμμικού ταξινομητή SVM. Τα διανύσματα υποστήριξης που προκύπτουν συνδυάζονται σε ένα μόνο με πολλαπλασιασμό κάθε διανύσματος με το βάρος του και πρόσθεσή του σε ένα καθολικό διάνυσμα. Το βάρος α είναι η εμπιστοσύνη του διανύσματος στήριξης v_i και υπολογίζεται ως συνάρτηση του σφάλματος ταξινόμησης:

$$v = \sum_i^n a_i v_i$$

Αυτή η μείωση σε ένα διάνυσμα υποστήριξης επιταχύνει την ανίχνευση καθώς οι περιγραφείς HOG των υποψηφίων περιοχών συγκρίνονται με ένα μόνο διάνυσμα αντί για περισσότερα. Προκειμένου να βελτιωθεί η ακρίβεια του ανιχνευτή SVM, χρησιμοποιείται ο μορφολογικός τελεστής Blackhat για να φιλτραριστούν οι υποψήφιες περιοχές που έχουν εντοπιστεί, και στη συνέχεια η εικόνα κατωφλιώνεται. Ο μετασχηματισμός Blackhat ορίζεται ως η διαφορά μεταξύ του κλεισίματος και της εικόνας εισόδου. Με αυτόν τον τρόπο δεν γίνεται αναζήτηση σε ένα μεγάλο μέρος της εικόνας και μειώνεται ο αριθμός των σφαλμάτων. Αυτό το μορφολογικό φίλτρο δίνει έμφαση στα σκοτεινά εικονοστοιχεία με υψηλή αντίθεση προς το κοντινό τους περιβάλλον, κάτι που βοηθάει ιδιαίτερα στην περίπτωση των σημάτων οδικής κυκλοφορίας.



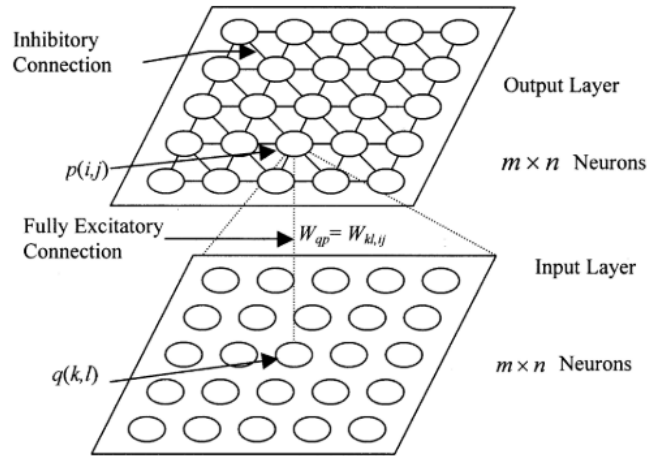
(a) Initial detection using SVM detector with 576 value HOG descriptor



(b) Blackhat filtering removes some false alarms

Αποτελέσματα εντοπισμού χωρίς και με μετασχηματισμό Blackhat
ΠΗΓΗ: Zaklouta et al. (2011) [69]

Νευρωνικά δίκτυα χρησιμοποιούνται από τους Fang et al. (2003) [71] για να αναγνωρίσουν το σχήμα των πλαισίων των σημάτων, αλλά αυτή η διαδικασία δεν είναι γραμμική. Στο σύστημά τους αναπτύσσονται δύο νευρωνικά δίκτυα για την εξαγωγή χαρακτηριστικών χρώματος και σχήματος. Και τα δύο αυτά νευρωνικά δίκτυα κατασκευάζονται από δύο στρώματα νευρώνων, ένα για είσοδο και ένα για έξοδο. Ο αριθμός των νευρώνων σε κάθε στρώμα είναι ίδιος με τον αριθμό των εικονοστοιχείων στην εικόνα εισόδου. Οι συνάψεις μεταξύ των επιπέδων εισόδου και εξόδου είναι πλήρως συνδεδεμένες, όπου (i, j) είναι η θέση του νευρώνα p στο layer εξόδου, (k, l) είναι η θέση του νευρώνα q στο layer εισόδου και w το βάρος μεταξύ νευρώνων q και p .

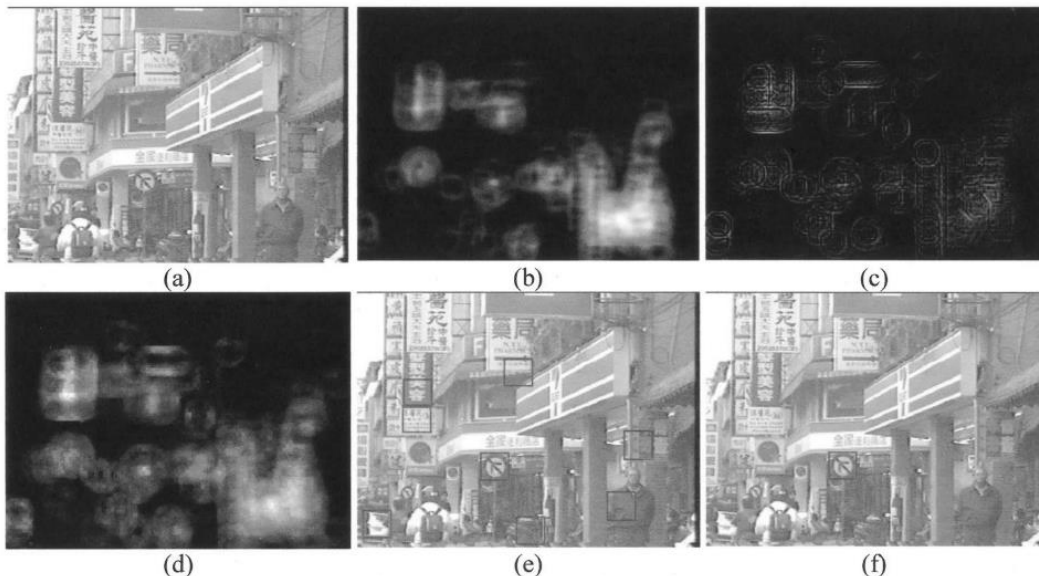


Νευρωνικό δίκτυο για ανίχνευση χαρακτηριστικών χρώματος ή σχήματος
 ΠΗΓΗ: Fang et al. (2003) [71]

Στην προσέγγιση αυτή επιλέχθηκε το μοντέλο HSI ως κατάλληλο για τα χαρακτηριστικά χρώματος, δεδομένου ότι βασίζεται στην αντίληψη του ανθρώπινου ματιού για το χρώμα. Συγκεκριμένα, μετατράπηκαν τις τιμές R,G,B κάθε εικονοστοιχείου σε τιμές Απόχρωσης (Hue). Οι πληροφορίες σχήματος βοηθούν στην εξάλειψη ακατάλληλων σημείων χρώματος, όπως τα εικονοστοιχεία ενός κόκκινου κτιρίου ή ενός κόκκινου αυτοκινήτου (δηλαδή μια μεγάλη κόκκινη περιοχή που δεν είναι πινακίδα). Η εξαγωγή χαρακτηριστικών σχήματος βασίστηκε στην ανίχνευση ακμών. Έτσι, οι νευρώνες στο στρώμα εισόδου του νευρωνικού δικτύου δρουν ως ανιχνευτές ακμών. Έστω (r,g,b) οι τιμές χρώματος RGB του εικονοστοιχείου (k,l) στην έγχρωμη εικόνα και η είσοδος στον νευρώνα q και x η έξοδος του νευρώνα, τότε ισχύει:

$$x = \sqrt{(I_{kl} - I_{k-1l})^3 + (I_{kl} - I_{kl-1})^2} \quad \text{και} \quad I = \frac{(r+g+b)}{3}$$

Η εξίσωση θα μπορούσε να αντικατασταθεί από οποιονδήποτε άλλο ανιχνευτή ακμών, συμπεριλαμβανομένων των Roberts, Sobel ή Laplace. Στην πραγματικότητα, ο εντοπισμός όλων των ακμών στην εικόνα δεν είναι απαραίτητος, καθώς απαιτούνται μόνο ακμές με συγκεκριμένες χρωματικές μεταβολές όπως συμβαίνει στα σήματα.



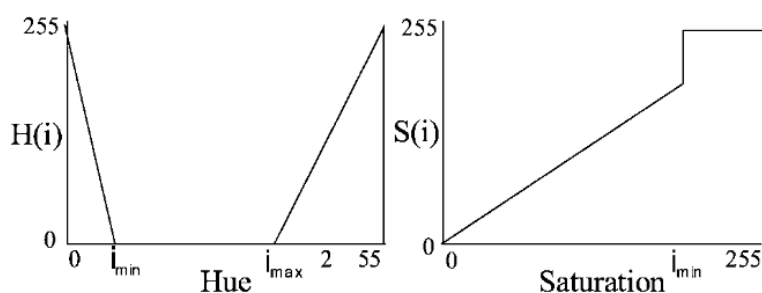
Παράδειγμα της διαδικασίας ανίχνευσης πινακίδων. (a) Αρχική εικόνα εισαγωγής. (b) Χρωματικός χάρτης χαρακτηριστικών εικόνας εισόδου. (c) Χάρτης χαρακτηριστικών σχήματος της εικόνας εισόδου. (d) Χάρτης ενσωμάτωσης χαρακτηριστικών χρώματος και σχήματος (e) Αποτέλεσμα ενσωμάτωσης. (f) Αποτέλεσμα μετά το βήμα επαλήθευσης.

ΠΗΓΗ: Fang et al. (2003) [71]

1.3.3 Τεχνικές εκμάθησης χρώματος

Εκτός από τις τεχνικές εκμάθησης σχήματος αλλά και συνδυασμού σχήματος – χρώματος, υπάρχουν και οι τεχνικές βασισμένες εξ ολοκλήρου στην εκπαίδευση χρώματος. Και σε αυτήν την περίπτωση πρωταρχικό ρόλο παίζουν οι γενετικοί αλγόριθμοι, οι μηχανές διανυσματικής στήριξης και τα νευρωνικά δίκτυα, αλλά και η τεχνική του αλγορίθμου AdaBoost των Viola – Jones.

Σε εργασία τους οι Escalera et al. (2003) [65] χρησιμοποίησαν γενετικούς αλγόριθμους στο στάδιο της ανίχνευσης. Έδειξαν ότι η χρήση έγχρωμων πινάκων αναζήτησης (LUT) στο χώρο HSI για τον εντοπισμό των περιοχών ενδιαφέροντος, καθώς και η χρήση γενετικών αλγορίθμων εντός των περιοχών ενδιαφέροντος, παρήγαγε πολύ καλύτερα τελικά αποτελέσματα από άλλες μεθόδους εντοπισμού. Έλαβαν μέρος μόνο τα κανάλια της Απόχρωσης και του Κορεσμού (Hue και Saturation) προκειμένου να αποφευχθούν οι αλλαγές στην κατάσταση του φωτισμού και κατασκευάζονται δύο πίνακες αναζήτησης (LUT), ένας για την Απόχρωση και ένας για τον Κορεσμό. Η γενική ιδέα είναι ότι εάν ένα κανάλι έχει λάθος τιμή θα μπορούσε να βοηθηθεί από το άλλο.



Colour classification LUTs
ΠΗΓΗ: Escalera et al. (2003) [65]

Επιπλέον, εφάρμοσαν τον γενετικό αλγόριθμο, μια παράλληλη αναζήτηση σε διαφορετικές κατευθύνσεις και στη συνέχεια μία διαδικασία βελτιστοποίησης που μιμείται τη φυσική εξέλιξη και επιλογή. Βασικά βήματα που ακολουθήθηκαν:

1. Αρχικοποίηση πληθυσμού.
2. Αξιολόγηση φυσικής κατάστασης για κάθε άτομο (individual).
3. Καλή επιλογή λύσεων για την παραγωγή νέου πληθυσμού.
4. Παραγωγή του νέου πληθυσμού.
5. Αξιολόγηση των νέων αποτελεσμάτων.
6. Ανταλλαγή του παλαιού πληθυσμού με το νέο.

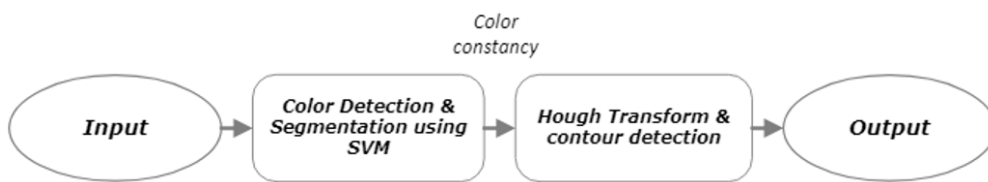
Τα βήματα 3 έως 6 γίνονται έναν καθορισμένο αριθμό φορών (γενιές) ή έως ότου η λύση που παρέχεται από το καλύτερο άτομο φτάσει μια συγκεκριμένη τιμή. Το βασικό πλεονέκτημα της προσέγγισής τους με τη χρήση των γενετικών αλγορίθμων είναι ότι επιτρέπει την αποτελεσματική ανίχνευση σημάτων οδικής κυκλοφορίας ανεξάρτητα από τη θέση, την κλίμακα, την στροφή, τις μερικές αποκρύψεις, την παρουσία άλλων αντικειμένων στην εικόνα που έχουν ίδιο χρώμα και τις διακυμάνσεις στις καιρικές συνθήκες.

Οι Kaplan et al. (2012) [66] εφάρμοσαν επίσης γενετικούς αλγόριθμους μαζί με προσαρμοστικές διορθώσεις φωτεινότητας (Adaptive Brightness Correction). Συγκεκριμένα, για τις παραμέτρους του γενετικού αλγορίθμου χρησιμοποιούν συντελεστές του αφινικού μετασχηματισμού (Affine Transformation Coefficients). Με μέσο όρο 9 ms ανά καρτέ επεξεργασίας πέτυχαν το 91% των αληθών ανιχνεύσεων, το 4,5% των ψευδών ανιχνεύσεων κατά μέσο όρο, τα οποία εκπαιδεύτηκαν και δοκιμάστηκαν σε βίντεο 512 x 288 που είχαν προγενέστερα συλλεχθεί χειροκίνητα. Αυτή η απόδοση καθιστά αυτή τη μέθοδο εν δυνάμει κατάλληλη για εφαρμογές πραγματικού χρόνου.

Ένα μοντέλο βασισμένο σε χρώμα, στο οποίο δεν πραγματοποιείται κατωφλίωση, προτάθηκε από τους Deguchi et al. (2011) [72]. Γίνεται χρήση ενός διαβαθμισμένου ταξινομητή που

εκπαιδεύτηκε με αλγόριθμο AdaBoost, παρόμοια τεχνική με εκείνη που πρότειναν οι Viola και Jones (2001), αλλά για την εξαγωγή χαρακτηριστικών σημείων χρησιμοποιήθηκαν τα Μοτίβα Τοπικής Κατάταξης (Local Rank Patterns), αντί για τα χαρακτηριστικά τύπου Haar, τα οποία είναι κατάλληλα για την ανίχνευση 2D μοτίβων σε πραγματικό χρόνο.

Στους Tam et al. (2010) [76] παρουσιάζεται μια προσέγγιση για την ανίχνευση κυκλοφοριακών πινακίδων με την τεχνική εκμάθησης χρώματος που βασίζεται σε Μηχανές Διανυσματικής Στήριξης (SVM). Αντί να γίνεται επεξεργασία σε κάθε εικονοστοιχείο, αυτή η προσέγγιση χρησιμοποιεί ένα μπλοκ εικονοστοιχείων ως διάνυσμα εισόδου στη μέθοδο SVM με στόχο την ταξινόμηση χρώματος, όπου η διάσταση κάθε διανύσματος μπορεί να επεκταθεί από μια ομάδα γειτονικών εικονοστοιχείων. Αυτό βοηθά στον χειρισμό της διαφοροποίησης τόσο στα δεδομένα εκπαίδευσης όσο και στα δεδομένα δοκιμών. Για πιο έγκυρα αποτελέσματα ανίχνευσης, εξάγονται και χαρακτηριστικά σχήματος μέσω του μετασχηματισμού Hough, όπου ανιχνεύεται το περίγραμμα των κυκλικών και τριγωνικών αντικειμένων για την επαλήθευση των υποψήφιων περιοχών.



Μοντέλο ανίχνευσης ΠΗΓΗ: Tam et al. (2010) [76]

Άλλος ένας αλγόριθμος βασισμένος σε Μηχανές Διανυσματικής Στήριξης (SVM) παρουσιάστηκε από τους Moreno et al. (2010) [70] για ανίχνευση πινακίδων, ο οποίος ταξινομεί τα εικονοστοιχεία χρησιμοποιώντας πληροφορίες χρώματος.

Μια διαφορετική προσέγγιση υιοθετήθηκε από τους Bahlmann et al. (2005) [73], όπου εκπαιδεύτηκαν συγκεκριμένης κλίμακας ανιχνευτές σημάτων οδικής κυκλοφορίας χρησιμοποιώντας χαρακτηριστικά τύπου Haar που παραμετροποιούνται με βάση το χρώμα. Τα ισχυρότερα χαρακτηριστικά επιλέχθηκαν μέσω ενός cascade αλγόριθμου AdaBoost, από ένα σύνολο πολλών και διαφορετικών χαρακτηριστικών τα οποία είχαν καθοριστεί σε πολλαπλές αναπαραστάσεις χρώματος: κανάλια απλού R, G και B, κανονικοποιημένα κανάλια των R, G και B (r, g, b) και ένα κανάλι γκριζας κλίμακας:

$$r = R/S, \quad g = G/S \text{ και } b = B/S \text{ με } S = R + G + B$$

$$\text{κανάλι grayscale} = S/3$$

Με αυτόν τον τρόπο, συνήχθη αυτόματα από τα δεδομένα μια καταλληλότερη αναπαράσταση χρώματος αντί να επιλεγθεί αυθαίρετα από τους συγγραφείς. Παρά τα υψηλής ακρίβειας τελικά αποτελέσματα, ένα από τα κύρια μειονεκτήματα αυτής της προσέγγισης είναι το υψηλό υπολογιστικό κόστος, πράγμα που καθιστά δύσκολη τη ανάπτυξη ενός συστήματος σε πραγματικό χρόνο. Όπως σημειώνουν οι συγγραφείς, για ανάλυση βίντεο 384×288 pixel επιτυγχάνεται ταχύτητα επεξεργασίας μόνο 10 καρτέ ανά δευτερόλεπτο.



Τα top 6 χαρακτηριστικά του ανιχνευτή. Τα εικονοστοιχεία κάτω από τις λευκές περιοχές σταθμίζονται κατά +1, η μαύρη περιοχή κατά -1. Το φόντο υποδεικνύει το κανάλι χρώματος στο οποίο υπολογίζονται μεμονωμένα τα χαρακτηριστικά, εδώ αντιστοιχεί σε $r, R, G, r, S/3, g$.
ΠΗΓΗ: Bahlmann et al. (2005) [73]

Μια από τις προσεγγίσεις που εφαρμόζεται στην εργασία των Lopez-Montiel et al. (2019) [75] αφορά την κατάτμηση εικόνας με τον αλγόριθμο συσταδοποίησης K-means. Αυτή η μη

επιβλεπόμενη τεχνική της Μηχανικής Μάθησης χρησιμοποιεί τα κανάλια από τον χρωματικό χώρο LAB για τη δημιουργία διαφορετικών συστάδων με τις ίδιες χρωματικές ιδιότητες. Με αυτόν τον τρόπο είναι εφικτό να εντοπιστούν περιοχές στην εικόνα με αντικείμενα που πληρούν αυτές τις προϋποθέσεις.

Ο αλγόριθμος K-means ελαχιστοποιεί: $W(C)$ όταν $D = ||x_i - x_{i'}||^2$

$$W(C) = \frac{1}{2} \sum_{k=1}^K \sum_{c(i)=k} \sum_{c(i')=k} ||x_i - x_{i'}||^2 = \sum_{k=1}^K N_k \sum_{c(i)=k} ||x_i - \bar{x}_k||^2$$

1.4 Αναγνώριση με μεθόδους Μηχανικής Μάθησης

Στα πλαίσια δημιουργίας συστημάτων για την υποβοήθηση της οδήγησης (Advanced Driver Assistance Systems - ADAS) ή για την πλήρως αυτοματοποιημένη οδήγηση, η αναγνώριση (recognition) αποτελεί το δεύτερο σημαντικό βήμα της κλασικής προσέγγισης που έχει αναφερθεί ως τώρα.

Η αναγνώριση στην γενική περίπτωση έχει ως στόχο την ταξινόμηση (classification) των περιοχών ενδιαφέροντος στις κλάσεις που έχουν οριστεί. Αν η ταξινόμηση δεν καταστεί δυνατή, η περιοχή ενδιαφέροντος θεωρείται εσφαλμένη και απομακρύνεται ως θόρυβος. Το μεγαλύτερο πρόβλημα που καλείται να λύσει η αναγνώριση είναι η κατηγοριοποίηση ενός ή περισσότερων αντικειμένων σε γενικές κλάσεις. Οι αλγόριθμοι ταξινόμησης προσεγγίζουν το πρόβλημα είτε με τη χρήση προτύπων (template-based approach) είτε με τη χρήση ταξινομητών (classifier-based approach).

Στην πρώτη περίπτωση πρόκειται για μία από τις πιο απλές και πρωτοεμφανισθείσες προσεγγίσεις της αναγνώρισης αντικειμένων, τη συνταύτιση προτύπων (template matching). Ο όρος "matching", γενικότερα, χρησιμοποιείται στην αναγνώριση προτύπων για να δηλώσει την ομοιότητα ανάμεσα σε δύο οντότητες του ίδιου είδους. Απαραίτητη προϋπόθεση είναι η ύπαρξη προκαθορισμένων προτύπων (templates), δηλαδή εν προκειμένω εικόνων με τα σήματα οδικής κυκλοφορίας. Κάθε περιοχή ενδιαφέροντος που έχει προκύψει από το στάδιο του εντοπισμού συγκρίνεται με τα διάφορα πρότυπα. Τελικά, η πινακίδα που απεικονίζεται στην περιοχή ενδιαφέροντος κατατάσσεται στην αντίστοιχη ομάδα του προτύπου με το οποίο εμφανίζει τη μέγιστη ομοιότητα (ή την ελάχιστη διαφορά). Το μέτρο της σύγκρισης, συνήθως προϊόν μιας συσχέτισης, στηρίζεται στο σύνολο των δεδομένων εκπαίδευσης. Η προσέγγιση αυτή απαιτεί μεγάλο υπολογιστικό κόστος, πρόβλημα που τα τελευταία χρόνια έχει αντιμετωπιστεί με την χρήση γρήγορων επεξεργασιών.

Στη δεύτερη περίπτωση, για την οποία ακολούθως γίνεται ενδεικτική αναφορά σε σχετικές ερευνητικές εργασίες, οι αλγόριθμοι προέρχονται από το πεδίο της μηχανικής μάθησης (Machine Learning). Εφαρμόζονται αλγόριθμοι όπου, με τη βοήθεια εικόνων σημάτων οδικής κυκλοφορίας, ο κανόνας ταξινόμησης που τους διέπει εκπαιδεύεται ώστε να κατηγοριοποιεί νέα δεδομένα. Υπάρχουν πολλές τεχνικές μηχανικής μάθησης που μπορούν να χρησιμοποιηθούν για την αναγνώριση πινακίδων, τα νευρωνικά δίκτυα και οι μηχανές διανυσματικής στήριξης αποτελούν τις πιο δημοφιλείς.

1.4.1 Μηχανές Διανυσματικής Στήριξης

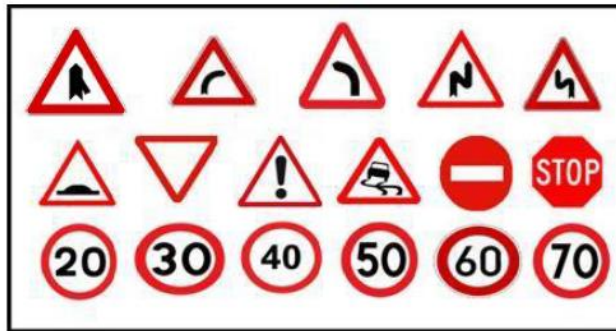
Στην περίπτωση των υποβοηθούμενων συστημάτων οδήγησης για την αναγνώριση των σημάτων οδικής κυκλοφορίας χρησιμοποιούνται ευρέως οι Μηχανές Διανυσματικής Μάθησης (SVM). Οι Dalal & Triggs (2005) [8] προέβησαν σε αναγνώριση των σημάτων οδικής κυκλοφορίας με τη χρήση μηχανών διανυσματικής στήριξης (SVM). Ως δεδομένα εισόδου δόθηκαν οι περιγραφείς ιστογράμματος με προσανατολισμούς ακμών (Histogram of Orientation Gradients – HOG) που είχαν εξαχθεί από την εικόνα. Αντίστοιχα έπραξαν και οι Chang & Swinson (2011) [86], με τη διαφορά ότι, ενώ οι περιγραφείς HOG υπολογίζουν σε

σύνολα εικονοψηφίδων τα μέτρα των κλίσεων σε προκαθορισμένο αριθμό διευθύνσεων, εδώ εφαρμόστηκαν σε έξι κατηγορίες για τον εντοπισμό μόνο δύο σημάτων. Το σύνολο των δεδομένων αυτών αποτελεί τα δεδομένα εκπαίδευσης του συστήματος αναγνώρισης.

Άλλη μία ερευνητική εργασία που χρησιμοποιεί τη μέθοδο SVM είναι των Hua et al. (2010) [87]. Το θέμα προσεγγίζεται με εξαγωγή χαρακτηριστικών από την εικόνα και τροφοδότηση τους στη μηχανή διανυσματικής στήριξης. Τα χαρακτηριστικά που χρησιμοποιούνται είναι οι περιγραφείς SIFT, οι οποίοι είναι ανθεκτικοί σε μετασχηματισμούς κλίμακας και στροφής, αλλαγές φωτεινότητας και ύπαρξη θορύβου. Το σύνολο δεδομένων που χρησιμοποιήθηκε για την αξιολόγηση αποτελείται από 130 εικόνες (50 × 50 εικονοστοιχεία) και έλαβαν ποσοστό ακρίβειας 93% με την απαίτηση χρόνου 0.098 ms ανά εικόνα.

Οι Greenhalgh & Mehdi (2012) [97] χρησιμοποίησαν τη διαδοχική δομή (cascaded structure) των ταξινομητών SVM και των χαρακτηριστικών HOG για να αναγνωρίσουν τα σήματα οδικής κυκλοφορίας. Ο αλγόριθμος SURF χρησιμοποιήθηκε για να συνταιριάζει δύο πινακίδες κυκλοφορίας ακόμα και αν μεταξύ τους έχουν μικρές διαφορές στον προσανατολισμό τους.

Μια διαφορετική προσέγγιση για ανίχνευση και αναγνώριση πινακίδων είναι αυτή των Kumaraswamy et al. (2011) [88], όπου οι αλγόριθμοι που εφαρμόζονται δοκιμάζονται σε ενσωματωμένο περιβάλλον σε πραγματικό χρόνο. Οι αλγόριθμοι εκπαιδεύονται για τον εντοπισμό και την αναγνώριση σημαντικών απαγορευτικών και προειδοποιητικών σημάτων από βίντεο που καταγράφονται σε πραγματικό χρόνο. Το σύστημα ανίχνευσης και αναγνώρισης πινακίδων κυκλοφορίας αποτελείται από τρία στάδια. Αρχικά, επιλέγονται οι περιοχές ενδιαφέροντος με κατωφλίωση στον χρωματικό χώρο RGB και εξάγονται σε μέγεθος 64x64 μετά από διόρθωση των στροφών και αφού περικοπούν. Στη συνέχεια οι περιοχές αυτές ταξινομούνται βάσει σχήματος με τη βοήθεια ενός μη γραμμικού ταξινομητή SVM. Ο μη γραμμικός ταξινομητής SVM εκπαιδεύεται με βάση τα χαρακτηριστικά της απόστασης από τα όρια (distance to borders), και με αυτόν τον τρόπο επιτρέπεται η ταξινόμηση των σημάτων σε τρεις κλάσεις, κύκλος, τρίγωνο και ανεστραμμένο τρίγωνο. Τέλος, όσες πινακίδες ταξινομήθηκαν ως κύκλος ή τρίγωνο αποστέλλονται στο στάδιο αναγνώρισης προτύπου, όπου επίσης εκπαιδεύονται μη γραμμικοί ταξινομητές SVM πολλαπλών κλάσεων.



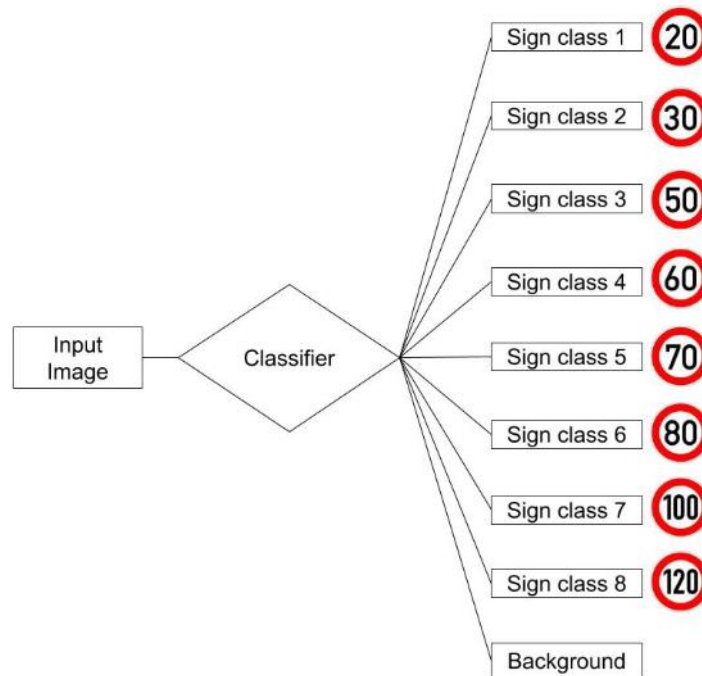
Σήματα που αναγνωρίζονται από την ενσωματωμένη εφαρμογή
ΠΗΓΗ: Kumaraswamy et al. (2011) [87]

Στην μελέτη επισκόπησης σχετικά με τις μεθόδους αναγνώρισης των Saadna & Behloul (2017) [91] αναφέρεται πως οι Yakimov et al (2016) [90] πέτυχαν αναγνώριση σηματοδότησης κυκλοφορίας σε πραγματικό χρόνο χρησιμοποιώντας τεχνολογία CUDA σε φορητή GPU Nvidia Tegra K1, η οποία περιέχει 192 πυρήνες γραφικών και 4 πυρήνες CPU αρχιτεκτονικής ARM. Οι συγγραφείς πρότειναν έναν τροποποιημένο γενικευμένο αλγόριθμο μετασχηματισμού Hough (GHT) για την ταξινόμηση και δείχνουν ότι ο αλγόριθμος πέτυχε έναν καλό συμβιβασμό μεταξύ του χρόνου εκτέλεσης και της ακρίβειας με τις προεπεξεργασμένες εικόνες σε σύγκριση με τα αποτελέσματα των Houben et al. (2013) [89], οι οποίοι μεταξύ άλλων μεθόδων δοκίμασαν και τις μηχανές διανυσματικής υποστήριξης σε συνδυασμό με την τεχνική κυλιόμενου παραθύρου (sliding window).

Algorithm	Accuracy (%)	time (FPS)
Sliding window + SVM [89]	100	1
GHT with preprocessing [90]	97.3	43
GHT without preprocessing [90]	89.3	25
Viola–Jones [89]	90.81	15
HOG [89]	70.33	20

Σύγκριση αλγορίθμων με βάση χρόνο εκτέλεσης και ακρίβεια
 ΠΗΓΗ: Saadna & Behloul (2017) [91]

Οι Berkaya et al. (2016) [92], στο πλαίσιο της αναγνώρισης κυκλικών σημάτων οδικής κυκλοφορίας ορίου ταχύτητας, κάνουν χρήση τριών διαφορετικών περιγραφέων HOG, τοπικών δυαδικών προτύπων (LBP) και φίλτρου Gabor, για την εξαγωγή χαρακτηριστικών που τροφοδοτούν στη συνέχεια σε Μηχανές Διανυσματικής Στήριξης. Οι υποψήφιες περιοχές ταξινομούνται με χρήση της μεθόδου SVM πολλαπλών κλάσεων. Γενικά, οι ταξινομητές SVM χρησιμοποιούνται ως δυαδικοί ταξινομητές. Ωστόσο, μια ταξινόμηση πολλαπλών κλάσεων SVM μπορεί να πραγματοποιηθεί αν συνδυαστούν πολλοί δυαδικοί ταξινομητές SVM ένας-εναντίον-ενός (one-versus-one). Οι διαφορετικές κλάσεις ορίζονται με βάση τις διαφορετικές τιμές ορίου ταχύτητας, όπως για παράδειγμα 30, 40, 50 km κ.λπ. Επιπλέον, ορίζεται μια κλάση για το φόντο και τα αντικείμενα που τελικά δεν αποτελούν κυκλικά σήματα οδικής κυκλοφορίας ορίου ταχύτητας και θα πρέπει να απορρίπτονται. Τελικά, δημιουργούνται στο σύνολο εννέα κατηγορίες, οκτώ για τις πινακίδες και μία για το φόντο. Για κάθε μία κατηγορία εκπαιδεύεται ένα SVM, και τα SVM όλων των κατηγοριών συνδυάζονται για να σχηματίσουν ένα SVM πολλαπλών κλάσεων.



Ταξινομητής SVM πολλαπλών κλάσεων για αναγνώριση σήματος ορίου ταχύτητας
 ΠΗΓΗ: Berkaya et al. (2016) [92]

Στην εργασία των Mandeljc et al. (2015) [93] παρουσιάζεται συγκριτική μελέτη για διάφορους τύπους χαρακτηριστικών αλλά και διαφορετικών χρωματικών χώρων. Σχετικά με την ταξινόμηση των πινακίδων κυκλοφορίας για την αναγνώριση, επικεντρώνονται σε

γραμμικούς ταξινομητές SVM δοκιμάζοντας διάφορες συνθέσεις πολλαπλών κλάσεων. Πιο συγκεκριμένα, γίνεται χρήση τριών διαφορετικών τύπων ταξινομητών του SVM και μιας διαδικτυακής μεθόδου. Οι μηχανές διανυσματικής στήριξης έχουν γίνει δημοφιλείς για την επίλυση προβλημάτων ταξινόμησης λόγω της γρήγορης και απλής διαδικτυακής εφαρμογής τους σε σύνολα δεδομένων μεγάλης κλίμακας.

1. Ταξινομητής ένας εναντίων όλων (one-vs-rest)

Περιλαμβάνει μια σειρά δυαδικών ταξινομητών, μια για κάθε κατηγορία πινακίδας. Κάθε ταξινομητής εκπαιδεύεται με τον τρόπο ένας εναντίων όλων (one-vs-rest), με τα δεδομένα της αντίστοιχης κλάσης να αποτελούν θετικά και τις περιπτώσεις όλων των άλλων κατηγοριών αρνητικά δείγματα. Κατά τη διάρκεια της πρόβλεψης το δείγμα δηλώνεται ότι ανήκει στην κλάση με την υψηλότερη εκτιμώμενη πιθανότητα. Χρησιμοποιήθηκε το πακέτο LIBLINEAR, το οποίο χρησιμοποιεί γραμμικό ταξινομητή SVM με την προεπιλεγμένη επιλογή επίλυσης και χωρίς προκατάληψη ($\text{bias} = 0$), Mandeljc et al. (2015) [93].

Το LIBLINEAR είναι μια βιβλιοθήκη ανοιχτού κώδικα για μεγάλης κλίμακας γραμμική ταξινόμηση. Υποστηρίζει την ταξινόμηση με λογιστική παλινδρόμηση και γραμμικές μηχανές διανυσματικής υποστήριξης. Υλοποιήθηκε από τους Fan et al. (2008) [94] και είναι διαθέσιμο στον ιστότοπο: www.csie.ntu.edu.tw/~cjlin/liblinear.

2. Ταξινομητής πολλαπλών κλάσεων με δομή δέντρου (Multiclass tree)

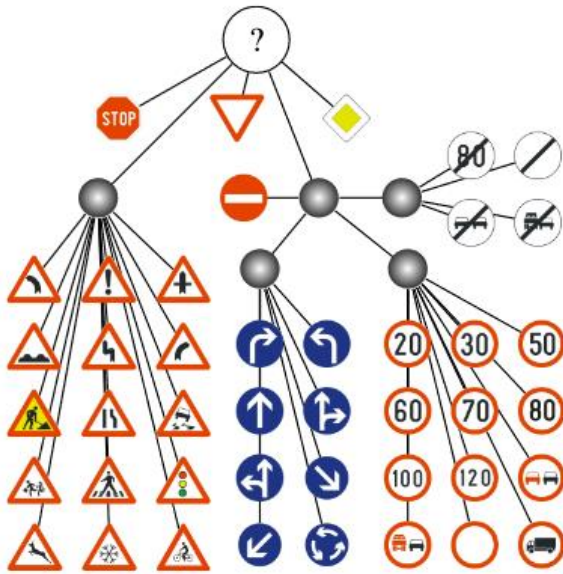
Η δομή του ταξινομητή που μοιάζει με δέντρο αποτελεί μια επέκταση του κανονικού «επίπεδου» ταξινομητή ένας εναντίων όλων (one-vs-rest). Οι κλάσεις των πινακίδων διαχωρίζονται χειροκίνητα σε υπερκλάσεις με τους δικούς τους κόμβους στο δέντρο, με βάση τις κοινές τους ιδιότητες (πχ. μπλε στρογγυλές πινακίδες, τριγωνικές πινακίδες, κ.λπ.). Κάθε κλαδί του δέντρου τελειώνει με έναν κόμβο φύλλων, που αντιστοιχεί σε έναν ταξινομητή που αντιπροσωπεύει μία τελική κλάση. Όλοι οι ταξινομητές εκπαιδεύονται και πάλι με λογική ένας εναντίων όλων (one-vs-rest). Ωστόσο, κάθε ταξινομητής εκπαιδεύεται ως ένας εναντίων όλων σε σχέση με τους αδελφικούς κόμβους του (κόμβοι στο ίδιο επίπεδο) και μεταδίδει στους θυγατρικούς κόμβους του μόνο τα θετικά του δείγματα. Κατά τη διάρκεια της πρόβλεψης, όλοι οι κόμβοι στο ίδιο επίπεδο χρησιμοποιούνται για την ταξινόμηση των δεδομένων, και ο κλάδος με την υψηλότερη εκτιμώμενη πιθανότητα χρησιμοποιείται για περαιτέρω ταξινόμηση, έως ότου επιτευχθεί ο κόμβος των φύλλων.

3. Ταξινομητής πολλαπλών κλάσεων Crammer & Singer.

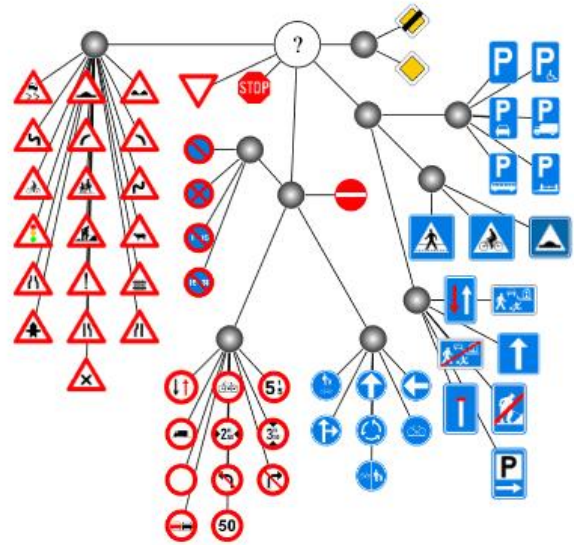
Στη συνέχεια, οι Mandeljc et al. (2015) [93], αξιολόγησαν τη μέθοδο SVM πολλαπλών κλάσεων των Crammer & Singer (2002) [95], το οποίο παρέχεται επίσης από το πακέτο LIBLINEAR. Σε αντίθεση με τις προηγούμενες δύο υλοποιήσεις, η συγκεκριμένη βελτιστοποιεί άμεσα το πρόβλημα των πολλών κλάσεων και δεν απαιτεί αποσύνθεση σε δυαδικά SVM τα οποία θα συνενωθούν μετά. Και πάλι, χρησιμοποιείται η προεπιλεγμένη επιλογή επίλυσης και η προκατάληψη τίθεται στο μηδέν.

4. Γραμμικός ταξινομητής LaRank (online)

Στους Bordes et al. (2007) [96] βελτιστοποιείται με τη χρήση του αλγορίθμου LaRank το ίδιο πρόβλημα με εκείνο της υλοποίησης των Crammer & Singer (2002) [95], αλλά αυτό γίνεται με διαδικτυακό τρόπο, πράγμα που σημαίνει ότι η εκπαίδευση μπορεί να πραγματοποιηθεί για ένα δείγμα κάθε φορά. Στην πράξη, πραγματοποιούνται πολλαπλές επαναλήψεις στο σετ δεδομένων εκπαίδευσης και με μεταβολή της σειράς των δεδομένων στην αρχή κάθε περιόδου. Με αυτόν τον τρόπο επιτυγχάνεται μια εκπαίδευση ισοδύναμη μιας μαζικής εκπαίδευσης. Ο ταξινομητής αξιολογείται μετά την πρώτη και μετά τη δέκατη περίοδο εκπαίδευσης. Καθώς πρόκειται για διαδικτυακή μέθοδο, αξιολογείται επίσης σε ένα σύνολο δεδομένων δοκιμής στο διαδίκτυο, όπου μετά την πρόβλεψη της ετικέτας ενός δοκιμαστικού δείγματος χρησιμοποιείται η πραγματική της ετικέτα για να ενημερωθεί ο ταξινομητής πριν ταξινομηθεί το επόμενο δείγμα.



(a) Tree structure for GTSRB dataset.



(b) Tree structure for BTSC dataset.

Οπτικοποίηση των δομών που μοιάζουν με δέντρα, με όλες τις κλάσεις πινακίδων που υπάρχουν σε κάθε σύνολο δεδομένων.

ΠΗΓΗ: Mandeljc et al. (2015) [93]

Συνολικά, συμπεραίνεται ότι η χρήση πολυκαναλικής αναπαράστασης εικόνας, αντί εικόνας σε κλίμακα του γκρι, τείνει να βελτιώσει την απόδοση της ταξινόμησης. Το καλύτερο αποτέλεσμα (εκτός σύνδεσης) επιτυγχάνεται από τον ταξινομητή πολλαπλών κλάσεων με δομή δέντρου, με ακρίβεια ταξινόμησης 98.26%. Ο γραμμικός ταξινομητής LaRank SVM φτάνει σχεδόν τη βέλτιστη ακρίβεια ταξινόμησης μετά από μία περίοδο εκπαίδευσης και δεν κερδίζει πολλά με την επανάληψη της εκπαίδευσης για δέκα εποχές, ενώ τα αποτελέσματα του διαδικτυακού σεναρίου δοκιμών δείχνουν σταθερά περαιτέρω βελτίωση στα αποτελέσματα της ταξινόμησης.

Οι Chaiyakhon et al. (2015) [98] επίσης εφάρμοσαν μία συγκριτική μελέτη χρησιμοποιώντας διαφορετικές τεχνικές προεπεξεργασίας των εικόνων και εφαρμογή τεσσάρων διαφορετικών συναρτήσεων πυρήνα για την ταξινόμηση με SVM. Τρία διαφορετικά είναι τα μοντέλα για την εκπαίδευση των εικόνων, συμπεριλαμβανομένων των αρχικών εικόνων που μετατρέπονται σε εικόνες κλίμακας του γκρι, των περιοχών ενδιαφέροντος που εξήχθησαν με τη βοήθεια του ανιχνευτών ακμών Canny σε συνδυασμό με τον μετασχηματισμό Hough και των περιοχών ενδιαφέροντος με χαρακτηριστικά κανονικής κατεύθυνσης (Normal direction feature). Τα μοντέλα εκπαιδεύτηκαν με 270 εικόνες. Στη διαδικασία αξιολόγησης της ταξινόμησης χρησιμοποιήθηκαν για δοκιμή 90 εικόνες. Τέλος, χρησιμοποιούνται τέσσερις διαφορετικές συναρτήσεις πυρήνα για τη διαδικασία ταξινόμησης με SVM (γραμμική, πολυωνυμική, σιγμοειδής και RBF).

Ο αλγόριθμος τμηματοποίησης της εικόνας που χρησιμοποιεί τη λειτουργία γραμμικού πυρήνα SVM είναι αρκετά αποτελεσματικός για την ταξινόμηση. Μπορεί να βελτιώσει την ακρίβεια της ταξινόμησης από 67.80% σε 73% και 88.60%. Η αύξηση είναι της τάξεως 5% και 20.8%. Αυτή η αύξηση οφείλεται στο γεγονός ότι οι περιοχές ενδιαφέροντος περικλύονται, τροφοδοτώντας το SVM μόνο με την περιοχή των πινακίδων κυκλοφορίας, γεγονός που εξαλείφει το περιττό υπόβαθρο και θόρυβο της εικόνας. Τα αποτελέσματα από κάθε μέθοδο παρατίθενται στην παρακάτω εικόνα.

SVM	Original		ROI		Normal direction	
	Accuracy		Accuracy		Accuracy	
	Train	Test	Train	Test	Train	Test
Linear	1.000	0.678	1.000	0.730	1.000	0.886
poly	1.000	0.467	1.000	0.511	0.958	0.545
RBF	1.000	0.067	1.000	0.067	1.000	0.365
Sigmoid	0.960	0.044	0.096	0.044	0.0975	0.287

Αποτελέσματα ταξινόμησης για κάθε μέθοδο
 ΠΗΓΗ: Chaiyakhan et al. (2015) [98]

1.4.2 Νευρωνικά Δίκτυα

Υπάρχουν πολλές τεχνικές μηχανικής μάθησης που μπορούν να χρησιμοποιηθούν για την αναγνώριση πινακίδων, η χρήση των νευρωνικών δικτύων είναι μία από τις πιο δημοφιλείς. Τα νευρωνικά δίκτυα χρησιμοποιούνται ευρέως στην ανάλυση εικόνας για πληθώρα εργασιών, όπως κατάτμηση εικόνας (image segmentation) και αναγνώριση αντικειμένων (object detection and recognition). Στο πρόβλημα της αναγνώρισης σημάτων οδικής κυκλοφορίας χρησιμοποιούνται για ταξινόμηση των περιοχών ενδιαφέροντος που έχουν προκύψει από το στάδιο της ανίχνευσης.

Η ίδια λογική ισχύει εκτός από τους ταξινομητές SVM και στα νευρωνικά δίκτυα. Τα προς ταξινόμηση δεδομένα εισόδου μπορούν να είναι και εδώ οι τιμές του χρώματος των εικονοσηφίδων (template based approach) ή χαρακτηριστικά που έχουν εξαχθεί από την εικόνα (feature based approach). Η επιλογή του τρόπου περιγραφής των δεδομένων που εισάγονται στο σύστημα ταξινόμησης θα πρέπει να γίνει έτσι ώστε να αποφευχθεί η χρήση πολλών παραμέτρων (overfitting), καθώς στην περίπτωση αυτή μειώνεται η ακρίβεια της ταξινόμησης. Αντιστοίχως, αν τα δεδομένα χαρακτηρίζονται από πολύ λίγες παραμέτρους, πιθανόν να μην είναι εφικτός ο επιτυχής διαχωρισμός τους. Για τον λόγο αυτό θα πρέπει να μελετάται και να ελέγχεται με λεπτομέρεια ο τρόπος με τον οποίο θα περιγράφονται τα δεδομένα. Για τις περιπτώσεις όπου τα προς εντοπισμό αντικείμενα ενδέχεται να παρουσιάζουν στροφές, μεταθέσεις και αλλαγές κλίμακας, η δημιουργία ενός συστήματος που βασίζεται στα χαρακτηριστικά είναι η βέλτιστη λύση. Η επιτυχία και η ακρίβεια της μεθόδου εξαρτώνται από τη σωστή επιλογή των χαρακτηριστικών που θα εξαχθούν από τα δεδομένα.

Τα χαρακτηριστικά που συνήθως χρησιμοποιούνται για ένα σύστημα νευρωνικών δικτύων επιλέγονται κατάλληλα, ώστε να εμφανίζουν ανοχή σε μετασχηματισμούς ομοιότητας. Αυτό αποτελεί σημαντικό πλεονέκτημα έναντι της προσέγγισης με συνταύτιση προτύπων, που για να εξασφαλίσει αυτή την ανεξαρτησία θα πρέπει να συμπεριλάβει στο σύνολο των δεδομένων πολλαπλές εικόνες του ίδιου αντικειμένου υπό διαφορετικούς πιθανούς μετασχηματισμούς.

Πολλοί ερευνητές προτείνουν για την επίτευξη της αναγνώρισης τη χρήση νευρωνικών δικτύων επειδή έχουν την ικανότητα να γενικεύουν από τα δεδομένα εκπαίδευσης. Ορισμένα μάλιστα μπορούν να κάνουν σωστά την ταξινόμηση παρά την παρουσία αποκρύψεων. Μια προϋπόθεση είναι ότι το νευρωνικό δίκτυο δεν πρέπει να επανεκπαιδεύεται όταν περιλαμβάνονται νέα δεδομένα, αλλά θα πρέπει να έχει τη δυνατότητα να αυξήσει τον όγκο των πληροφοριών χωρίς να χάσει αυτό που έχει ήδη μάθει. Η διαδικασία εκτελείται σε δύο βήματα. Πρώτον, η εξαγωγή των πληροφοριών που περιέχονται στην πινακίδα και, το δεύτερο, η αναγνώρισή τους από το νευρωνικό δίκτυο.

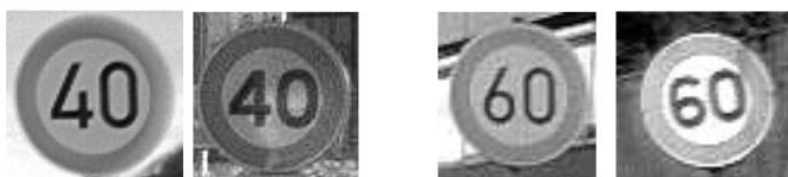
Για τους παραπάνω λόγους, στην εργασία των Escalera et al. (2003) [65] μετά το στάδιο της ανίχνευσης με γενετικό αλγόριθμο ακολουθεί η κατάτμηση των εικόνων στον χώρο RGB για την εξαγωγή πληροφοριών που περιέχονται στις πινακίδες, και επιλέγεται ένα νευρωνικό δίκτυο ως μέθοδος για τη λύση του προβλήματος της ταξινόμησης. Το νευρωνικό δίκτυο που χρησιμοποιείται για την ταξινόμηση ακολουθεί το πρότυπο Adaptive Resonance Theory για

δυναμικές εισόδους, ART1. Ένα τέτοιο δίκτυο είναι ικανό να αναπτύξει σταθερή ομαδοποίηση αυθαίρετων ακολουθιών προτύπων εισόδου από μόνο του.

Αποτελείται από δύο στρώματα με bottom up and top down μεταξύ τους μονοπάτια. Το επίπεδο εξόδου έχει σχεδιαστεί ως «ανταγωνιστικό» δίκτυο, ενεργοποιείται δηλαδή μόνο ο κόμβος που λαμβάνει τη μεγαλύτερη συνολική είσοδο. Αυτός ο κόμβος δημιουργεί ένα μοτίβο top down που συγκρίνεται με την είσοδο. Εάν η διαφορά είναι αρκετά μεγάλη, ο κόμβος αναστέλλεται και ο κύκλος ξεκινά ξανά. Όταν η ομοιότητα είναι αρκετά υψηλή τότε αυτό μπορεί να σημαίνει πως το πρωτότυπο διάνυσμα της κλάσης ταιριάζει αρκετά με το τρέχον διάνυσμα εισόδου. Εάν όχι, σημαίνει πως έχει παρουσιαστεί ένα διαφορετικό μοτίβο και έτσι δημιουργούνται νέοι νευρώνες. Με αυτό τον τρόπο, δεν είναι απαραίτητο να εκπαιδευτεί ξανά το δίκτυο εάν πρέπει να προστεθούν νέα μοτίβα. Το δίκτυο βελτιώνει και εμπλουτίζει τις γνώσεις του μόνο του με τη χρήση.

Ουσιαστικά τα νευρωνικά δίκτυα ARN1 μιμούνται την ανθρώπινη ικανότητα συνδυασμού της μάθησης και της μνήμης. Κατά τη διάρκεια της λειτουργίας ενός ART1, το δίκτυο ομαδοποιεί τα δεδομένα σε συστάδες, όπου κάθε νευρώνας εξόδου είναι το κέντρο ενός συμπλέγματος. Στην περίπτωση εδώ, ο τύπος των σημάτων είναι γνωστός εκ των προτέρων, επομένως οι νευρώνες εξόδου επισημαίνονται με την κλάση του κάθε προτύπου εκπαίδευσης (διαφορετικές πινακίδες).

Οι Mourtarde et al. (2007) [99] παρουσιάζουν έναν εύρωστο αλγόριθμο ανίχνευσης και αναγνώρισης σημάτων ορίου ταχύτητας για αμερικανικές και ευρωπαϊκές πινακίδες. Αρκούνταν μόνο στον εντοπισμό με βάση το σχήμα (ορθογώνια ή κύκλοι) πράγμα που καθιστά τον εντοπισμό ανθεκτικό απέναντι στις μεταβολές των χρωμάτων και του φωτισμού. Αντί για καθολική αναγνώριση, το σύστημα ταξινομεί (ή απορρίπτει) τις υποψήφιες περιοχές με το όριο ταχύτητας τμηματοποιώντας πιθανά ψηφία μέσα τους και στη συνέχεια εφαρμόζεται οπτική αναγνώριση των ψηφίων με νευρωνικά δίκτυα (Optical Digit Recognition - ODR). Αυτό βοηθά στον χειρισμό των διαφορετικών σημάτων οδικής κυκλοφορίας ορίου ταχύτητας από χώρα σε χώρα, αρκεί μόνο τα ψηφία να αναγνωρίζονται σωστά. Το νευρωνικό δίκτυο ODR είναι ένα πολυεπίπεδο perceptron με 10 εξόδους (1 για κάθε τιμή ψηφίου) εκπαιδευμένο σε ειδικά κατασκευασμένες βάσεις δεδομένων ψηφίων που εξάγονται από σχετικές ενδείξεις ορίου ταχύτητας σε βίντεο που έχουν εγγραφεί από κάμερα οχήματος. Για παράδειγμα, η βάση δεδομένων ψηφίων της ΕΕ που εμπλουτίζεται συνεχώς με νέα παραδείγματα, που εξάγονται από βίντεο που έχουν εγγραφεί σε διαφορετικές ευρωπαϊκές χώρες, περιείχε εν έτει 2007 συνολικά 2762 παραδείγματα ψηφίων και 2789 αρνητικά παραδείγματα (μη ψηφία). Το συνολικό ποσοστό ανίχνευσης που επιτεύχθηκε είναι περίπου 90%, με εσφαλμένη ταξινόμησης κάτω από 1%. Το σύστημα επεξεργάζεται σε πραγματικό χρόνο εικόνες βίντεο 640x480, σε περίπου 20 καρέ/δευτερόλεπτο, σε έναν τυπικό φορητό υπολογιστή διπλού πυρήνα 2,13GHz.



Απεικόνιση της διαφορετικότητας των ψηφίων σε χώρες της ΕΕ: Γερμανικά σήματα (αριστερά κάθε ζεύγους) σε σύγκριση με τα αντίστοιχα της Γαλλίας
ΠΗΓΗ: Mourtarde et al. (2007) [99]

Στην εργασία των Medici et al. (2008) [100] οι εικόνες εισόδου εισάγονται στο σύστημα ανίχνευσης, ώστε να γίνει εξαγωγή των περιοχών ενδιαφέροντος. Οι υποψήφιες έγχρωμες εικόνες περικόπτονται, αφού όλες οι εικόνες που τροφοδοτούνται στο στάδιο της αναγνώρισης πρέπει να έχουν ενιαίες διαστάσεις, επομένως πραγματοποιείται επαναπροσδιορισμός των διαστάσεων σε 50x50. Στη συνέχεια μετατρέπονται σε κλίμακα του γκρι και ενισχύονται με μια τεχνική αντίθεσης (contrast stretching technique), προκειμένου να αντισταθμιστεί η φωτεινότητά τους.

Κάθε υποψηφία περιοχή ενδιαφέροντος που έχει υποστεί επεξεργασία προηγουμένως και εν δυνάμει περιέχει πινακίδα, εισάγεται σε ένα νευρωνικό δίκτυο πολλαπλών επιπέδων perceptron πρόσθιας τροφοδότησης. Δημιουργούνται συνολικά έξι νευρωνικά δίκτυα, ένα για κάθε κλάση πινακίδας που έχει δηλωθεί πως πρέπει να αναγνωρισθεί (πληροφοριακές, απαγορευτικές, κίνδυνος, στοπ, απαγόρευση στάθμευσης και υποχρεωτικές). Νευρωνικό δίκτυο για περιοχές ενδιαφέροντος που δεν απεικονίζουν σήματα δεν δημιουργήθηκε, αφού οι συγκεκριμένες περιοχές θα απομακρύνονται όταν παρουσιάζονται πολύ χαμηλές τιμές σε όλα τα στρώματα εξόδου. Η εκπαίδευση των νευρωνικών δικτύων πραγματοποιείται με αλγόριθμο ανάστροφης διάδοσης σφάλματος, με χρήση προσομοιωμένων και πραγματικών δεδομένων που απεικόνιζαν δεδομένα καλής ποιότητας, κακής ποιότητας και δείγματα που δεν περιείχαν σήματα οδικής κυκλοφορίας.



Δείγμα εικόνων με πινακίδες: (a) προσομοίωσης, (b) καλής ποιότητας, (c) κακής ποιότητας
 ΠΗΓΗ: Medici et al. (2008) [100]

Κάθε επίπεδο εισόδου αποτελείται από 2.500 νευρώνες, διαστάσεις ίσες με τον αριθμό εικονοψηφίδων της εικόνας με την οποία τροφοδοτείται στο δίκτυο. Τα επίπεδα εξόδου σε κάθε νευρωνικό δίκτυο απαρτίζονται από τόσους νευρώνες όσα τα ξεχωριστά σήματα που ανήκουν σε κάθε κατηγορία. Πραγματοποιήθηκαν διάφορες δοκιμές για να αποφασιστεί εάν τα «κρυμμένα» επίπεδα πρέπει να είναι ένα ή δύο, καθώς και για το πόσους νευρώνες πρέπει να έχει το κάθε κρυφό επίπεδο. Η διαδικασία της εκπαίδευσης πραγματοποιήθηκε για 50.000, 100.000 και 200.000 επαναλήψεις.

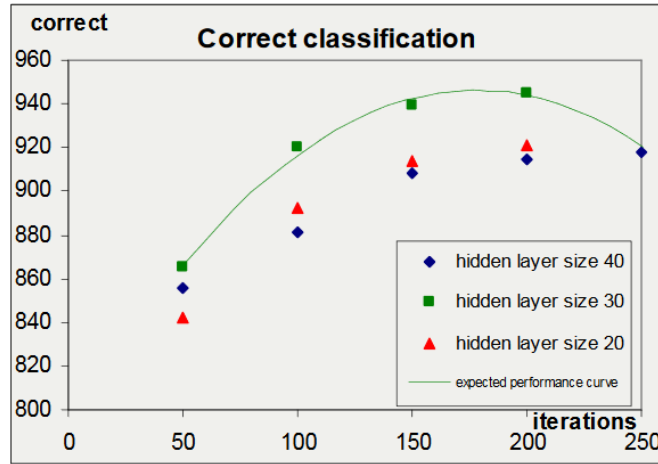
Net Name	# Hidden Layer	Geometry	# Output
Prohibitions	2	115+65	37
Information	1	50	10
Obligation	2	90+40	26
Danger	1	175	42
No-Parking	1	40	3
Stop	1	80	3

Γεωμετρία του νευρωνικού δικτύου
 ΠΗΓΗ: Medici et al. (2008) [100]

Το δίκτυο με ένα πρώτο επίπεδο 115 νευρώνων και ένα δεύτερο επίπεδο 65 νευρώνων αποδείχθηκε πως έχει τα καλύτερα αποτελέσματα και χρησιμοποιείται στην τελική εφαρμογή, όπου μπορεί να ταξινομήσει περισσότερες από 200 διαφορετικές ιταλικές πινακίδες σε πραγματικό χρόνο, με ποσοστό αναγνώρισης 80% έως 90%.

Άλλη μια εργασία με ταξινόμηση πινακίδων διαφορετικών ορίων ταχύτητας με τη χρήση νευρωνικών δικτύων είναι αυτή των Eichner & Breckon (2008) [33]. Μετά τη διαδικασία του εντοπισμού απομακρύνονται οι δακτύλιοι των κυκλικών σημάτων, ώστε στα δεδομένα εισόδου του νευρωνικού δικτύου να περιλαμβάνεται μόνο το εσωτερικό τμήμα της πινακίδας και πραγματοποιείται επανασύσταση εικόνας με διαστάσεις 20x20. Το πολυεπίπεδο νευρωνικό δίκτυο πρόσθιας τροφοδότησης που χρησιμοποιήθηκε αποτελείται από 400 νευρώνες εισόδου, 30 νευρώνες κρυφού επιπέδου και 12 νευρώνες εξόδου που αντιστοιχούν στις πινακίδες ορίου ταχύτητας για Βρετανία και Πολωνία: 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 km, όριο εθνικής ταχύτητας και ψευδώς-θετικό (δηλαδή όχι πινακίδα). Το μέγεθος του κρυφού επιπέδου ορίστηκε με εμπειρικό πειραματισμό, ενώ πρώτα πραγματοποιήθηκαν

διάφορες δοκιμές σχετικά με τον αριθμό των νευρώνων που πρέπει να έχει (20, 30 και 40 νευρώνες). Δοκιμές επίσης πραγματοποιήθηκαν και για τον αριθμό επαναλήψεων που τελικά ορίστηκε να είναι 150 με ανάλυση της αναμενόμενης καμπύλης απόδοσης. Αυτή η τιμή εξασφαλίζει καλή απόδοση συστήματος και ταυτόχρονα αποτρέπει την υπερβολική προσαρμογή του ταξινομητή στα δεδομένα εκπαίδευσης.

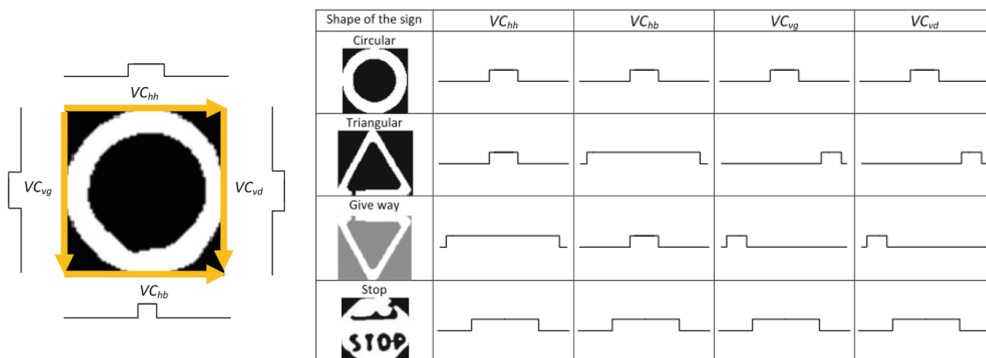


Αποτελέσματα εκπαίδευσης νευρικού δικτύου – καμπύλη απόδοσης
 ΠΗΓΗ: Eichner & Breckon (2008) [33]

Αποτέλεσμα κάθε νευρώνα εξόδου είναι μία πιθανότητα, η οποία λαμβάνει τιμή μεταξύ 0 και 1, έτσι ο νευρώνας για τον οποίο εντοπίζεται η μέγιστη πιθανότητα είναι εκείνος που καθορίζει το αποτέλεσμα. Για να θεωρηθεί έγκυρη η ταξινόμηση πρέπει η μέγιστη πιθανότητα να απέχει από την αμέσως επόμενη περισσότερο από 0.5.

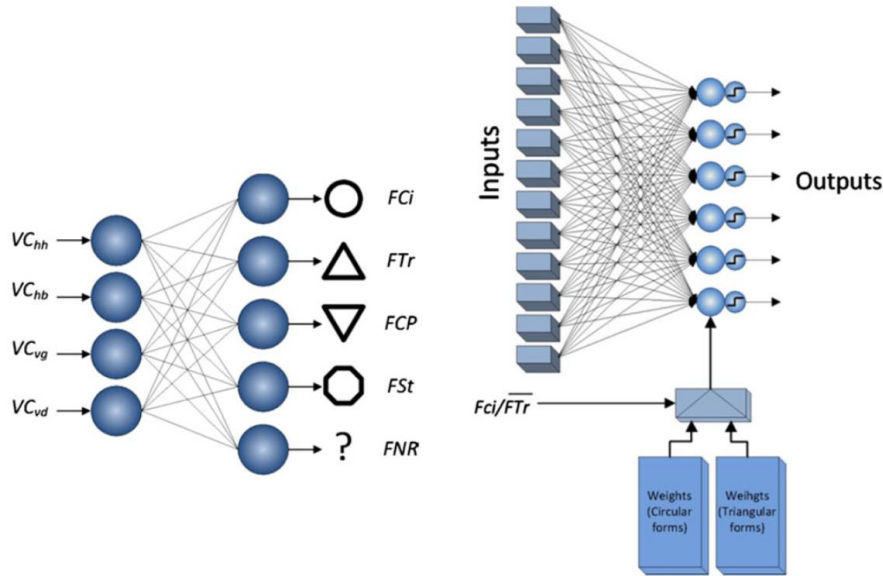
Στους Chokri et al. (2013) [101] η αναγνώριση των πινακίδων ολοκληρώνεται σε τρία κύρια βήματα. Πρώτον, γίνεται κατάτμηση εικόνας με βάση το χρώμα, στον χρωματικό χώρο HSV. Δεύτερον, εξάγεται το σχήμα του σήματος κυκλοφορίας με τη χρήση νευρωνικού δικτύου και, τέλος, το σήμα κυκλοφορίας αναγνωρίζεται με τη χρήση ενός δευτέρου νευρωνικού δικτύου.

Το πρώτο νευρωνικό δίκτυο επιτρέπει την αναγνώριση του τύπου σχήματος, εάν το σχήμα είναι κυκλικό, τριγωνικό, αντεστραμμένο τρίγωνο ή οκτάγωνο STOP ή οποιαδήποτε άλλη μορφή. Οι εισοδοί του νευρωνικού δικτύου είναι τέσσερα διανύσματα, επομένως τέσσερις νευρώνες, που καθορίζουν τις τιμές των εικονοστοιχείων στις τέσσερις άκρες της πινακίδας που έχει μετατραπεί σε δυαδική εικόνα: VC_{hh} , VC_{hb} , VC_{vg} and VC_{vd} . Αποτελείται από 5 νευρώνες εξόδου, 4 για κάθε αναγνωρισμένο σχήμα και 1 για οποιαδήποτε άλλη μορφή που δεν αντιστοιχεί στις 4 γνωστές. Η έξοδος του πρώτου νευρωνικού δικτύου δείχνει το σχήμα της πινακίδας, όπου το περιεχόμενο θα αναγνωριστεί από το δεύτερο νευρωνικό δίκτυο.



Τα τέσσερα διανύσματα που καθορίζουν τις τιμές των εικονοστοιχείων στις άκρες του εικονογράμματος (αριστερά) & οι μορφές των τεσσάρων διανυσμάτων για τα διαφορετικά σχήματα των πινακίδων (δεξιά). ΠΗΓΗ: Chokri et al. (2013) [101]

Το δεύτερο νευρωνικό δίκτυο αναλύει μόνο το περιεχόμενο των κυκλικών και τριγωνικών πινακίδων, επομένως ορίζονται δύο πίνακες για να περιέχουν τα βάρη του δικτύου για τους δύο τύπους εικόνων. Κατά τη διάρκεια της εκπαίδευσης μπορούν να εναλλάσσονται αυτοί οι πίνακες για την εκμετάλλευση των αντίστοιχων βαρών που αφορούν κάθε σχήμα. Το επίπεδο εξόδου αποτελείται από 6 νευρώνες. Ο συνολικός αριθμός των αναγνωρίσιμων πινακίδων ισούται με 130, ενώ το ποσοστό αναγνώρισης είναι περίπου 95% για εικόνες με καλό φωτισμό και περίπου 84% για εικόνες με χαμηλό φωτισμό.



Δομή του πρώτου νευρωνικού δικτύου για ταξινόμηση των σχημάτων (αριστερά) & δομή του δεύτερου νευρωνικού δικτύου για ταξινόμηση περιεχομένου (δεξιά).
 ΠΗΓΗ: Chokri et al. (2013) [101]

Οι Abedin et al. (2016) [102] δημιουργούν για τις πινακίδες οδικής κυκλοφορίας του Μπαγκλαντές ένα νέο σύνολο δεδομένων που περιέχει 200 εικόνες αυτοκινητοδρόμων σε διάφορες καιρικές συνθήκες και συνθήκες φωτισμού. Χρησιμοποιούν τον αλγόριθμο SURF, ο οποίος παρουσιάζει αντοχή σε περιστροφές, κλίση και αποκρύψεις για το στάδιο της ανίχνευσης, ενώ στο στάδιο της αναγνώρισης επιλέγεται η επίλυση με τη βοήθεια νευρωνικού δικτύου ανάστροφης διάδοσης (Back-propagation algorithm) πολλαπλών επιπέδων. Για την εκπαίδευση του νευρωνικού δικτύου εισάγονται τα χαρακτηριστικά που εξήχθησαν με τον αλγόριθμο SURF στο προηγούμενο στάδιο για κάθε περιοχή ενδιαφέροντος. Το δίκτυο αποτελείται από 64 νευρώνες στο επίπεδο εισόδου, 10 νευρώνες στο κρυφό επίπεδο και 4 νευρώνες στο επίπεδο εξόδου. Η απόδοση του ταξινομητή απεικονίζεται μεταξύ άλλων και με την καμπύλη ROC (Receiver Operating Characteristic). Το πραγματικό θετικό ποσοστό που επιτυγχάνεται είναι 97% με ψευδές θετικό ποσοστό 3%.

Η καμπύλη ROC είναι μια γραφική παράσταση που απεικονίζει την απόδοση ενός ταξινομητή. Δημιουργείται με σχεδίαση των αληθώς θετικών αποτελεσμάτων (TPR) έναντι των ψευδών θετικών (FPR) σε διάφορες ρυθμίσεις κατωφλίου. Ουσιαστικά, δείχνει τη σχετική μεταβολή των ποσοστών «ορθής λήψης σήματος» από τις πραγματικές μετρήσεις σήματος και «εσφαλμένου συναγερμού» από τις πραγματικές μετρήσεις θορύβου, συναρτήσεως του σημείου απόφασης, δηλαδή της μέτρησης με βάση την οποία ο δέκτης θεωρεί ότι κάθε άλλη μέτρηση μεγαλύτερη αυτής αποτελεί σήμα. Το ποσοστό των αληθώς θετικών αποτελεσμάτων TPR (True Positive Rate) ενός ταξινομητή εκτιμάται ως εξής:

$$TPR = \frac{TP}{P} = \frac{\text{θετικά σωστά ταξινομημένα δεδομένα}}{\text{σύνολο θετικών}}$$

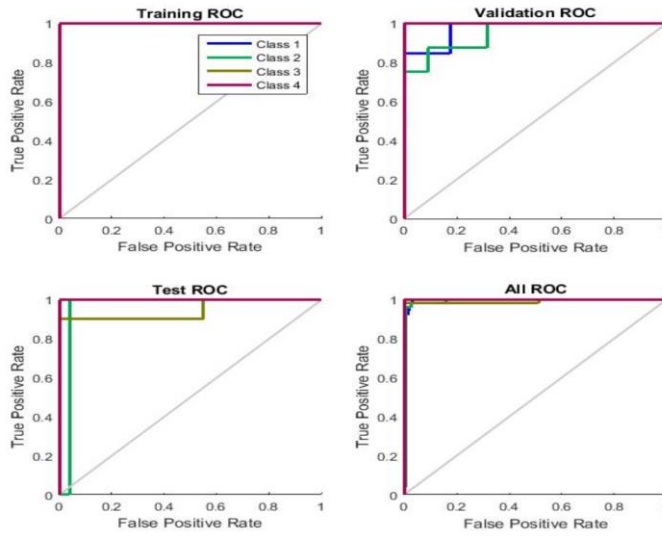
Το ποσοστό των ψευδών θετικών αποτελεσμάτων FPR (False Positive Rate) ενός ταξινομητή εκτιμάται ως εξής:

$$FPR = \frac{FP}{P} = \frac{\text{αρνητικά λανθασμένα ταξινομημένα δεδομένα}}{\text{σύνολο αρνητικών}}$$

Επιπρόσθετοι όροι συνδεδεμένοι με τις ROC καμπύλες είναι η ευαισθησία (sensitivity) και η “ειδικότητα” (specificity) οι οποίες ορίζονται ως εξής:

$$\text{Sensitivity} = SE = TPR$$

$$\text{Specificity} = \frac{TN}{FP + TN} = \frac{\text{αληθώς αρνητικά}}{\text{ψευδώς θετικά} + \text{αληθώς αρνητικά}} = 1 - FPR$$

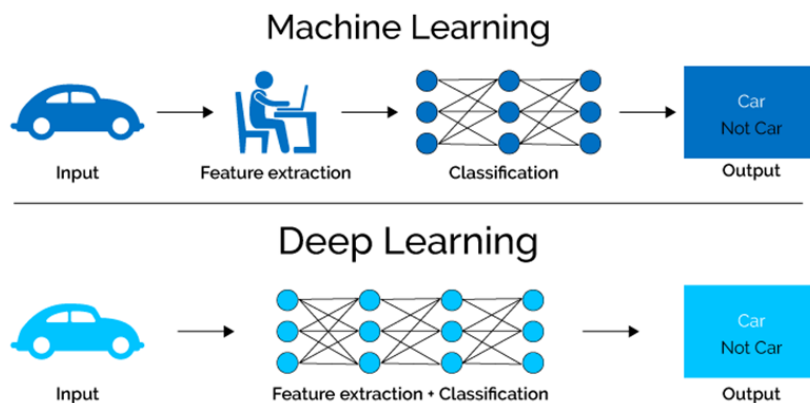


ROC καμπύλες ταξινομητή
 ΠΗΓΗ: Abedin et al. (2016) [102]

2

Η σύγχρονη προσέγγιση (State of the Art)

Υπενθυμίζεται εδώ ότι η βαθιά μάθηση (Deep Learning), υποσύνολο της τεχνητής νοημοσύνης και εξέλιξη της μηχανικής μάθησης, χρησιμοποιεί τεχνητά νευρωνικά δίκτυα πολλαπλών επιπέδων για να παρέχει προηγμένη ακρίβεια σε εργασίες όπως η ταξινόμηση και η ανίχνευση αντικειμένων. Ο όρος «βαθύ» (Deep) είναι ένας τεχνικός όρος που αναφέρεται στον αριθμό των κρυφών επιπέδων (hidden layers) που περιέχει ένα δίκτυο. Εάν υπάρχουν ένα ή δύο κρυμμένα επίπεδα στο δίκτυο, τότε αυτό ταξινομείται ως απλό ή ρηχό (shallow network), εάν όμως υπάρχουν περισσότερα από 3 κρυμμένα επίπεδα, τότε ταξινομείται ως δίκτυο βαθιάς μάθησης και ονομάζεται βαθύ νευρωνικό δίκτυο (Deep Neural Network).



Διαφορά μηχανικής μάθησης και βαθιάς μηχανικής μάθησης
ΠΗΓΗ: <https://abhitronix.github.io/2018/12/14/humanoid-Brain-2/>

Οι αλγόριθμοι βαθιάς μάθησης μαθαίνουν μέσω του τεχνητού νευρικού δικτύου που λειτουργεί κατ' εικόνα και καθ' ομοίωση του ανθρώπινου εγκεφάλου, επιτρέποντας στην μηχανή να αναλύει δεδομένα με μια συγκεκριμένη διαδικασία, παρόμοια όπως κάνουν οι άνθρωποι. Δηλαδή, οι μηχανές βαθιάς μάθησης δεν απαιτούν από έναν άνθρωπο - προγραμματιστή να τους πει τι να κάνουν με τα δεδομένα ώστε να αρχίσουν να μαθαίνουν από αυτά. Αυτό πραγματοποιείται από την εξαιρετική ποσότητα δεδομένων που συλλέγουν και καταναλώνουν οι ίδιοι. Τα δεδομένα δηλαδή είναι το καύσιμο για την λειτουργία των

μοντέλων βαθιάς μάθησης. Τα βαθιά νευρωνικά δίκτυα, πολλαπλών κρυφών επιπέδων, επεξεργάζονται τα δεδομένα εισόδου με περισσότερες μαθηματικές λειτουργίες και για το λόγο αυτό έχουν μεγαλύτερη υπολογιστική πολυπλοκότητα. Το υπολογιστικό κόστος είναι ένα από τα χαρακτηριστικά γνωρίσματα της βαθιάς μάθησης και είναι ένας λόγος για τον οποίο οι GPU έχουν ζήτηση για την εκπαίδευση τέτοιων μοντέλων.

Δεν είναι υπερβολή να λεχθεί ότι η βαθιά μάθηση χρησιμοποιείται με ποικίλους τρόπους σε όλες τις βιομηχανίες. Για παράδειγμα, η βαθιά μάθηση προσπαθεί να βοηθήσει τη βιομηχανία υγείας και περίθαλψης σε εργασίες όπως η ανίχνευση καρκίνου. Οι Hu et al. (2018) [133], έτσι, παρέχουν μια επισκόπηση για τη βαθιά μάθηση και τις δημοφιλείς αρχιτεκτονικές που χρησιμοποιούνται για την ανίχνευση και τη διάγνωση του καρκίνου. Στον κλάδο των υπηρεσιών διαδικτύου και κινητής τηλεφωνίας υπάρχει πληθώρα εφαρμογών που χρησιμοποιούν μεθόδους βαθιάς μάθησης είτε για ταξινόμηση εικόνας είτε για αναγνώριση ομιλίας. Εφαρμογές όπως το Skype της Microsoft, το Google Voice, το Apple Siri, αποτελούν τέτοιες εφαρμογές που μετατρέπουν την ομιλία σε κείμενο. Σε αυτό το πλαίσιο, οι Nassif et al. (2019) [134] παρουσιάζουν μια διεξοδική ανασκόπηση των διάφορων μελετών που έχουν διεξαχθεί για την αναγνώριση ομιλίας. Στα μέσα ενημέρωσης, την ψυχαγωγία και τις ειδήσεις, πολλές εφαρμογές όπως η εξατομίκευση σε πραγματικό χρόνο ή συστήματα προτάσεων όπως το Netflix χρησιμοποιούν τη βαθιά μάθηση. Παρότι το μοντέλο της αυτόματης μετάφρασης δεν είναι καινούργιο, η βαθιά μάθηση συμβάλλει στη βελτίωση της εμπειρίας αυτόματης μετάφρασης κειμένου.

Μια άλλη εντυπωσιακή ικανότητα της βαθιάς μάθησης είναι να αναγνωρίσει μια εικόνα και να δημιουργήσει μια συνεκτική λεζάντα με τη σωστή δομή πρότασης που ανταποκρίνεται ή και περιγράφει την εικόνα. Άλλος ένας τομέας όπου η βαθιά μάθηση επικρατεί είναι αυτός της ρομποτικής, ο οποίος πλέον όλο και πιο συχνά περιλαμβάνει την αλληλεπίδραση ανθρώπου με ρομπότ (Human Robot Interaction - HRI) σε πολλές διαφορετικές πτυχές της ανθρώπινης ζωής. Τα ρομπότ πρέπει να έχουν την ικανότητα να αντιλαμβάνονται και να κατανοούν τους διαφορετικούς τρόπους που χρησιμοποιούν οι άνθρωποι για επικοινωνία, όπως ομιλία ή κινήσεις σώματος. Προκειμένου να ενισχυθεί η εμπειρία μεταξύ ανθρώπου και ρομπότ, οι Filntisis et al. (2018) [135] διερευνούν μεθόδους που μπορούν να επεκτείνουν τις δυνατότητες ενός συστήματος αναγνώρισης της κίνησης με χρήστες, όπως τα μικρά παιδιά, όπου υπάρχουν περιορισμένα δεδομένα για εκπαίδευση των αλγορίθμων και πολλές τεχνικές αιχμής αντιμετωπίζουν δυσκολίες. Η βαθιά μάθηση είναι ακόμα και μέρος του γνωστού αλγόριθμου AlphaGo της DeepMind, ο οποίος νίκησε τον πρώην παγκόσμιο πρωταθλητή Lee Sedol στο παιχνίδι Go στις αρχές του 2016.

Φυσικά, η βαθιά μάθηση αποτελεί βασική τεχνολογία αιχμής και πίσω από τη δημιουργία συστημάτων για την υποβοήθηση της οδήγησης (Driving Assistance Systems) ή για την πλήρως αυτοματοποιημένη οδήγηση, επιτρέποντας την αναγνώριση σημάτων οδικής κυκλοφορίας, των πεζών, των φωτεινών σηματοδοτών κλ.π.

2.1 Εντοπισμός και Αναγνώριση σε ένα στάδιο // Deep Learning Methods

Όπως έχει ήδη αναφερθεί, για την ανίχνευση και αναγνώριση σημάτων οδικής κυκλοφορίας στις περισσότερες περιπτώσεις χρησιμοποιείται η εξαγωγή χαρακτηριστικών σημείων - feature extraction (feature based τεχνικές), είτε χρώματος (κατωφλιώσεις σε διάφορους χρωματικούς χώρους) είτε σχήματος (Hough Transformation, HOG, SHIFT/ SURF, Distance to Borders κ.ά.) τα οποία τροφοδοτούν έπειτα για την αναγνώριση έναν αλγόριθμο ταξινόμησης Μηχανικής Μάθησης, όπως είναι για παράδειγμα οι ταξινομητές SVM. Αντίστοιχα, αναφέρθηκαν και οι περιπτώσεις όπου ο εντοπισμός γίνεται με βάση κάποιον αλγόριθμο Μηχανικής Μάθησης, όπως η χρήση του αλγορίθμου AdaBoost, και στη συνέχεια η αναγνώριση επίσης με κάποια μέθοδο μηχανικής μάθησης (SVM, Νευρωνικών Δικτύων κ.ά.).

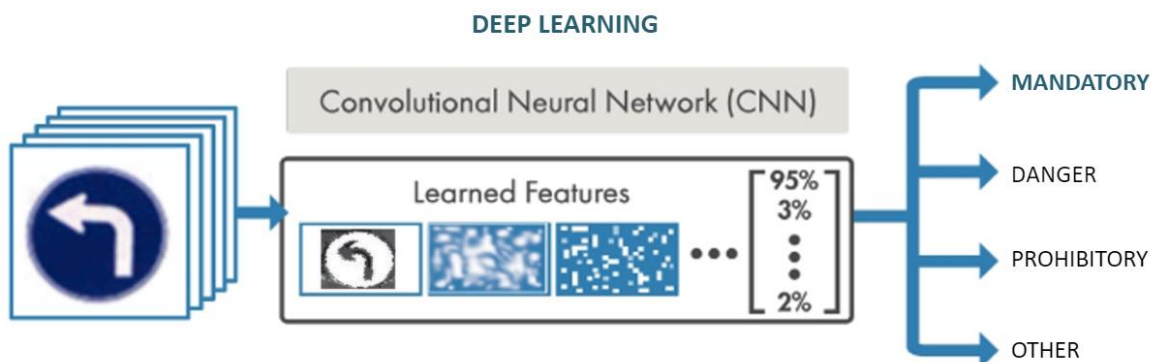
Κοινό χαρακτηριστικό των εν λόγω μεθόδων είναι πως ο ανίχνευση και η αναγνώριση γίνονται σε δύο στάδια, καθότι πρέπει να γίνει η εξαγωγή των χαρακτηριστικών σημείων για τον εντοπισμό. Στο κεφάλαιο αυτό, παρουσιάζονται οι τεχνικές αιχμής (state-of-the-art), που

λαμβάνουν χώρα τα τελευταία χρόνια με τη χρήση της Βαθιάς Μηχανικής Μάθησης (Deep Learning). Οι αλγόριθμοι βαθιάς μηχανικής μάθησης επιτρέπουν να γίνεται η ανίχνευση και η αναγνώριση σε ένα στάδιο.

Οι αλγόριθμοι που έχουν αναπτυχθεί για την αναγνώριση σημάτων οδικής κυκλοφορίας με βαθιά μάθηση, και οι οποίες είναι πολύ ελπιδοφόρες για την ανάπτυξη προηγμένων συστημάτων υποστήριξης οδηγών (ADAS) και αυτόνομων οχημάτων, ποικίλουν. Μεγαλύτερη απήχηση στον ευρύτερο επιστημονικό κλάδο έχουν τα Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks - CNN). Η ιδιαιτερότητά τους οφείλεται στην ικανότητά τους να ανιχνεύουν και να αναγνωρίζουν συγκεκριμένα (πολλά και διαφορετικά) στοιχεία εντός των εικόνων, επιτυγχάνοντας με αυτό τον τρόπο μεγάλες ακρίβειες στην κατηγοριοποίηση.

Η οικογένεια αλγορίθμων Περιφερειακών Συνελικτικών Νευρωνικών Δικτύων (Region based CNN - R-CNN) είναι μία προσέγγιση βαθιάς μάθησης που χρησιμοποιείται συστηματικά στην αναγνώριση πινακίδων. Η μέθοδος συνδυάζει προτάσεις ορθογώνιων περιοχών με τα συνελικτικά νευρωνικά δίκτυα. Αν και η ακρίβεια των αποτελεσμάτων που παρέχουν είναι αρκετά υψηλή, δεν καθίστανται ιδανικοί για την χρήση σε εφαρμογές πραγματικού χρόνου λόγω υπολογιστικής πολυπλοκότητας.

Τα τελευταία χρόνια έχει αναπτυχθεί πολύ η οικογένεια των αλγορίθμων YOLO, η οποία επίσης βασίζεται σε συνελικτικά δίκτυα. Οι συγκεκριμένοι αποτελούν αλγορίθμους ανίχνευσης αντικειμένων ενός σταδίου και επιτρέπουν για πρώτη φορά την ανίχνευση αντικειμένων σε πραγματικό χρόνο. Οι μέθοδοι ενός σταδίου είναι χρήσιμες για την παραγωγή ταχύτερων αποτελεσμάτων με μειωμένη όμως ακρίβεια. Εξαιρεση αποτελεί η τελευταία βελτίωση του YOLO (YOLOv4), όπου τα αποτελέσματα τόσο της ακρίβειας όσο και της ταχύτητας είναι αξιοσημείωτα.



Προσέγγιση βαθιάς μάθησης κατηγοριοποίησης σημάτων οδικής κυκλοφορίας

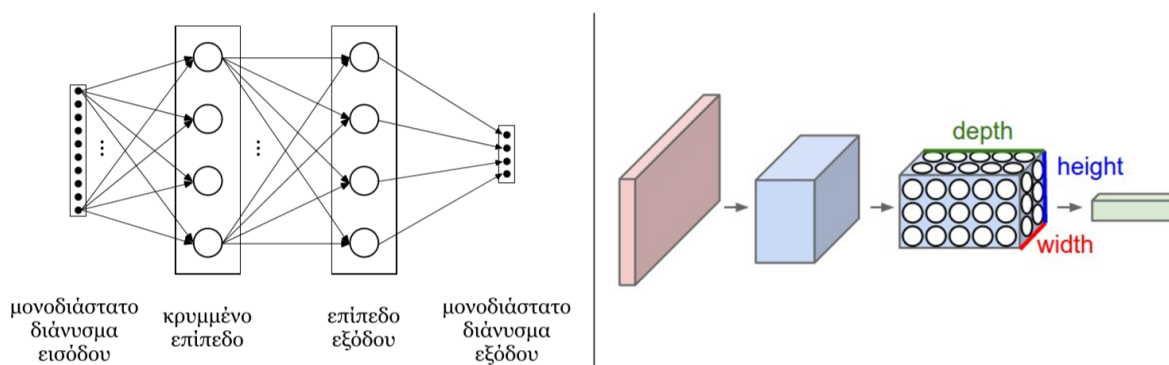
2.1.1 Συνελικτικά Νευρωνικά Δίκτυα (CNN)

Τα Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks - CNN) είναι μια παραλλαγή πολυεπίπεδων νευρωνικών δικτύων και αποτελούν αλγορίθμους βαθιάς μηχανικής μάθησης. Είναι σχεδιασμένα ώστε να αναγνωρίζουν οπτικά μοτίβα απευθείας από εικόνες με ελάχιστη προεπεξεργασία. Ονομάζονται «συνελικτικά» λόγω της διαδικασίας της συνέλιξης που εφαρμόζεται στα επίπεδα του δικτύου.

Τα συνελικτικά νευρωνικά δίκτυα άρχισαν να εμφανίζονται στο πεδίο της ανίχνευσης και αναγνώρισης αντικειμένων κυρίως από το 2012. Τη χρονιά αυτή, το νευρωνικό δίκτυο AlexNet (με 5 συνελικτικά επίπεδα) κέρδισε τον διαγωνισμό ImageNet Large Scale Visual Recognition Challenge (ILSVRC). Στον συγκεκριμένο διαγωνισμό, διάφορες ερευνητικές ομάδες αξιολογούν τους αλγορίθμους τους πάνω σε δεδομένο σετ εικόνων (ImageNet2) και προσπαθούν να επιτύχουν τη μεγαλύτερη δυνατή ακρίβεια. Τη χρονιά εκείνη το υψηλότερο ποσοστό επέτυχε η ερευνητική ομάδα των Alex Krizhevsky, Ilya Sutskever και Geoffrey E. Hinton χρησιμοποιώντας το συνελικτικό νευρωνικό δίκτυο (Convolutional Neural Network) AlexNet, επιτυγχάνοντας ένα top-5 ποσοστό σφάλματος 15.3%, 10.8 ποσοστιαίες μονάδες

πάνω από το δεύτερο. Το 2014, νικητής του διαγωνισμού ILSVRC ήταν το GoogleNet (Inception_V1, 22 συνελκτικά επίπεδα) από την Google, το οποίο πέτυχε το top-5 ποσοστό σφάλματος 6.67%.

Τα δίκτυα CNN αποτελούνται γενικά από δύο κύρια μέρη: το πρώτο, που είναι υπεύθυνο για την εξαγωγή των συνελκτικών χαρακτηριστικών (convolutional features) για την διαδικασία της ανίχνευσης, και το κυρίως νευρωνικό, το οποίο εκπαιδεύεται χρησιμοποιώντας αυτά τα χαρακτηριστικά. Το δεύτερο μέρος συνήθως είναι ένας γραμμικός ή μη ταξινομητής. Μία διαφορά από τα κλασικά νευρωνικά δίκτυα είναι η υπόθεση ότι οι εισοδοί τους είναι εικόνες ή έχουν μορφή και χαρακτηριστικά εικόνων. Στα κλασικά νευρωνικά δίκτυα η έννοια της εικόνας και του χώρου δεν έχουν καθοριστική σημασία όπως έχουν στα δίκτυα CNN. Επιπλέον, βασίζονται στη πράξη της συνέλιξης, ενώ τα κλασικά νευρωνικά βασίζονται στην πράξη του εσωτερικού γινομένου. Τέλος, η πληθώρα επιπέδων που τα διέπει τους επιτρέπει να κάνουν αρκετές πρόσθετες διαδικασίες, όπως είναι η μείωση της διάστασης των χαρακτηριστικών. Άλλο ένα προσόν των δικτύων CNN είναι ότι είναι ευκολότερα στην εκπαίδευση και έχουν πολύ λιγότερες παραμέτρους από τα κλασικά νευρωνικά δίκτυα με τον ίδιο αριθμό κρυφών επιπέδων.



Ένα κλασικό νευρωνικό δίκτυο (αριστερά), ένα συνελκτικό νευρωνικό δίκτυο (δεξιά) του οποίου τα επίπεδα νευρώνων οργανώνονται σε 3 διαστάσεις.
 ΠΗΓΗ: Μελέτης (2015) [136]

Ένα τυπικό βαθύ δίκτυο CNN, προκειμένου να στηρίξει τη διαδικασία της ανίχνευσης και της ταξινόμησης, συνήθως έχει τρία σύνολα επιπέδων, ένα σύνολο συνελκτικών επιπέδων με τις συναρτήσεις ενεργοποίησης και τα συγκεντρωτικά επίπεδα (pooling layers), τα οποία επαναλαμβάνονται πολλές φορές. Αυτά τα επίπεδα ακολουθούνται από μερικά πλήρως συνδεδεμένα επίπεδα (fully connected layers).

2.1.1.1 Αρχιτεκτονική

Η δομή και λειτουργία των δικτύων CNN συνοψίζεται στα ακόλουθα βήματα:

1. Συνελκτικά επίπεδα (Convolutional layers)

Σε ένα συνελκτικό επίπεδο περιλαμβάνονται νευρώνες οργανωμένοι σε παράλληλα επίπεδα, που εφαρμόζουν σε όλα τα κανάλια της εισόδου φίλτρα τα οποία έχουν προκύψει από τη διαδικασία μάθησης μέσω του αλγορίθμου οπισθοδιάδοσης (back-propagation). Ο κύριος στόχος του επιπέδου είναι να ανιχνεύσει διαφορετικά μοτίβα ή χαρακτηριστικά από μια εικόνα εισόδου. Η "βελτιστότητα" (principle of optimality) των εξαγόμενων χαρακτηριστικών αποτελεί την κύρια αιτία που οδήγησαν τα συνελκτικά νευρωνικά δίκτυα στην μεγάλη ανάπτυξη και επιτυχία. Στον δυναμικό προγραμματισμό, ο οποίος χρησιμοποιείται για την επίλυση προβλημάτων βελτιστοποίησης, των οποίων η λύση μπορεί να θεωρηθεί σαν μια ακολουθία αποφάσεων, ως αρχή της βελτιστότητας ορίζεται:

«Οποιαδήποτε υπακολουθία της βέλτιστης ακολουθίας αποφάσεων είναι επίσης βέλτιστη για το υποπρόβλημα που αντιστοιχεί στη συγκεκριμένη ακολουθία».

2. Συναρτήσεις ενεργοποίησης (Activation functions)

Για την εφαρμογή της μη γραμμικότητας στα συνελκτικά νευρωνικά δίκτυα η πιο συχνά χρησιμοποιούμενη είναι η συνάρτηση ReLU (Rectified Linear Unit), η οποία ενεργοποιεί όλους τους νευρώνες ταυτόχρονα.

3. Επίπεδα κανονικοποίησης (Normalization layers)

Τα επίπεδα αυτά δεν είναι πάντοτε απαραίτητα, ωστόσο χρησιμοποιούνται διότι κανονικοποιούν τις σχετικές διαφορές γειτονικών τιμών χαρακτηριστικών, ώστε να έχουν χρησιμοποιήσιμες τιμές από τα επόμενα επίπεδα. Μπορεί να εφαρμοσθεί εντός ενός χαρακτηριστικού ή και μεταξύ ενός συνόλου χαρακτηριστικών.

4. Επίπεδα χωρικής υποδειγματοληψίας (Pooling layers)

Εκτός από τα επίπεδα του φιλτραρίσματος, αναγκαία είναι η περιοδική ύπαρξη επιπέδων τα οποία μειώνουν τις χωρικές διαστάσεις της αναπαράστασης, ελαττώνοντας το πλήθος προς μάθηση παραμέτρων και τους υπολογισμούς σε ένα δίκτυο, ελαχιστοποιώντας ταυτόχρονα τις πιθανότητες για υπερ-εκπαίδευση.

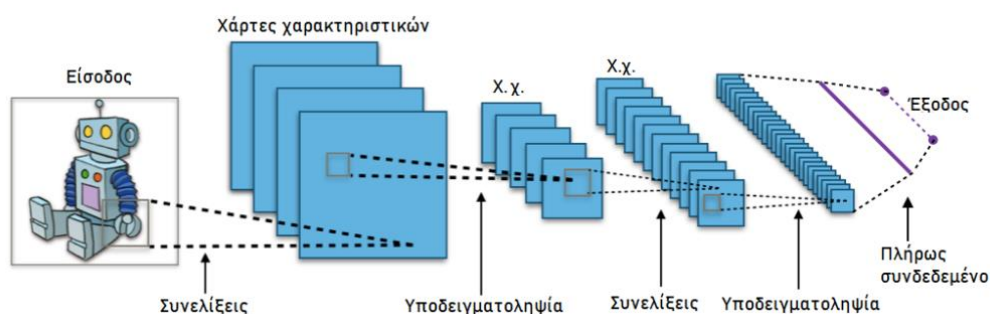
5. Πλήρως συνδεδεμένα επίπεδα (Fully connected layers)

Αυτά τα επίπεδα είναι ίδια με τα επίπεδα που χρησιμοποιούνται στα κλασικά νευρωνικά δίκτυα. Περιλαμβάνουν νευρώνες που δέχονται ένα διάνυσμα χαρακτηριστικών εισόδου και εξάγουν μία απόκριση.

6. Συναρτήσεις σφάλματος (Loss functions)

Σημαντικό ρόλο στην απόδοση του δικτύου διαδραματίζει η δυναμική σχέση των σφαλμάτων των αποτελεσμάτων με τις τιμές των συναπτικών βαρών. Και στα συνελκτικά νευρωνικά δίκτυα χρησιμοποιείται ο αλγόριθμος οπισθοδιάδοσης σφάλματος (back propagation, optimization error) και η επαναληπτική μέθοδος gradient descent για να ελαχιστοποιηθεί η συνάρτηση κόστους.

Από τα παραπάνω επίπεδα μόνο τα συνελκτικά και τα πλήρως συνδεδεμένα έχουν νευρώνες με τη συνήθη έννοια, επομένως μόνο αυτά εισάγουν βάρη προς εκπαίδευση στο δίκτυο. Τα υπόλοιπα επίπεδα επιτελούν μία προκαθορισμένη λειτουργία. Επομένως, η σειρά τοποθέτησης και οργάνωσής τους πρέπει να ακολουθεί συγκεκριμένα μοτίβα (layer patterns). Τα συνελκτικά τοποθετούνται στην αρχή, ενώ τα πλήρως συνδεδεμένα τοποθετούνται συνήθως στο τέλος. Και δύο αυτά επίπεδα μπορεί να ακολουθούνται από επίπεδα συγκέντρωσης, τα οποία μειώνουν τις χωρικές διαστάσεις των χαρακτηριστικών. Τα επίπεδα κανονικοποίησης μπορούν να προστεθούν προαιρετικά πριν ή μετά από τα επίπεδα συγκέντρωσης.



Αρχιτεκτονική συνελκτικού νευρωνικού δικτύου

ΠΗΓΗ: <https://www.jessicayung.com/explaining-tensorflow-code-for-a-convolutional-neural-network/>

2.1.1.2 Συνέλιξη

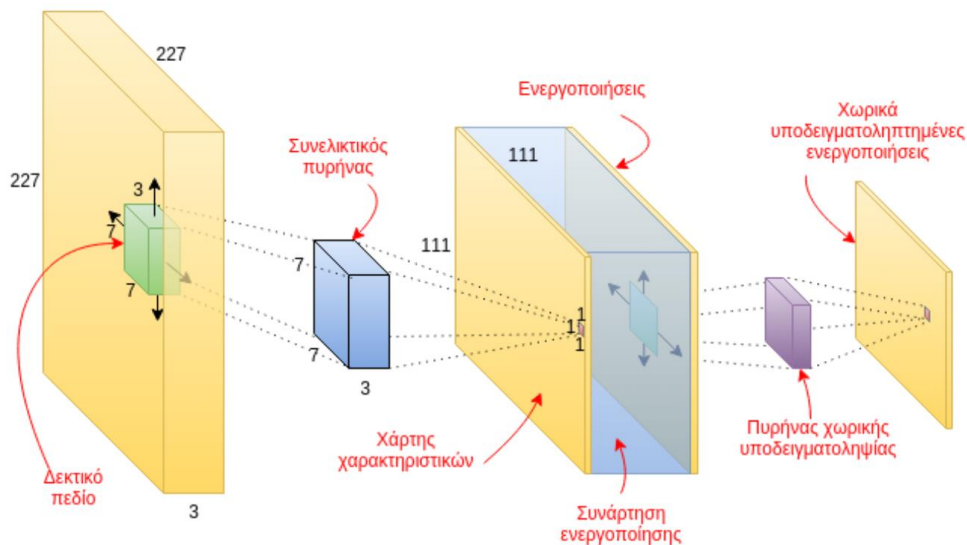
Τα επίπεδα συνέλιξης είναι ο πυρήνας των μοντέλων CNN. Οι παράμετροι ενός επιπέδου συνέλιξης είναι μία σειρά από δισδιάστατα φίλτρα, τα οποία όμως εκτείνονται σε όλο το βάθος του όγκου εισόδου. Επομένως, ένα επίπεδο συνέλιξης χρησιμοποιεί ένα σύνολο από φίλτρα τα οποία εντοπίζουν την παρουσία συγκεκριμένων χαρακτηριστικών ή μοτίβων που παρουσιάζονται στην αρχική εικόνα εισόδου. Πιο συγκεκριμένα, η είσοδος σε ένα επίπεδο

συνέλιξης είναι μια εικόνα $m \times m \times r$, όπου m το είναι το ύψος και το πλάτος της εικόνας, ενώ r ο αριθμός των καναλιών, π.χ. $r = 3$ για RGB. Το επίπεδο συνέλιξης έχει k φίλτρα (kernels) μεγέθους $n \times n \times q$, όπου n είναι μικρότερο από τη διάσταση της εικόνας και q μπορεί να είναι ίδιου μεγέθους με τον αριθμό των καναλιών ή μικρότερου και μπορεί να ποικίλει για κάθε kernel, συνήθως όμως διατηρούν τη διάσταση του βάρους της εικόνας.

Διαφορετικά φίλτρα τα οποία εντοπίζουν διαφορετικά χαρακτηριστικά περιστρέφονται στην εικόνα εισόδου, όπου το καθένα εφαρμόζεται σε κάθε εικόνα για να παράγουν k χάρτες χαρακτηριστικών (feature maps) μεγέθους $(m - n) + 1$, οι οποίοι προκύπτουν ως έξοδος και περνούν στο επόμενο επίπεδο. Για παράδειγμα, όταν η συνέλιξη εφαρμοστεί σε έγχρωμη εικόνα εισόδου $32 \times 32 \times 3$ (RGB) με φίλτρο $5 \times 5 \times 3$, το αποτέλεσμα της εξόδου θα είναι $28 \times 28 \times 3$ με τις χωρικές διαστάσεις (ύψος και πλάτος) να μειώνονται. Ωστόσο, το βήμα μετατόπισης (stride) του φίλτρου πάνω στην εισοδο είναι και αυτό μία παράμετρος των επιπέδων συνέλιξης. Καθορίζει τον αριθμό των εικονοστοιχείων κατά το οποίο θα μετακινείται το δεκτικό πεδίο του συνελικτικού επιπέδου οριζοντίως και κατακορυφώς. Με βήμα μετατόπισης S διάφορο της μονάδας η έξοδος είναι $(m - n)/S + 1$.

Ένα πρόβλημα που εμφανίζεται στην περίπτωση των μοντέλων CNN με μεγάλο αριθμό κρυφών επιπέδων είναι η γρήγορη μείωση των διαστάσεων μήκους και πλάτους του όγκου, το οποίο είναι αποτέλεσμα της διαδοχικής εφαρμογής πράξεων συνέλιξης. Προκειμένου να αντιμετωπιστεί το πρόβλημα αυτό χρησιμοποιείται η παράμετρος zero-padding P , η οποία γεμίζει με μηδενικά το περίγραμμα της εισόδου, βοηθώντας σημαντικά στη διατήρηση των χωρικών διαστάσεων της εισόδου. Με την εισαγωγή της υπέρ-παραμέτρου zero-padding η εξίσωση υπολογισμού εξόδου ορίζεται ως $(m - n + 2P)/S + 1$.

Όλες οι παραπάνω παράμετροι (αριθμός φίλτρων kernel, μέγεθος φίλτρων, το βήμα μετατόπισης και η ποσότητα zero-padding) ονομάζονται υπερ-παραμέτροι (Hyperparameters) των επιπέδων συνέλιξης.



Επίπεδο συνέλιξης μονού φίλτρου, με βήμα 2 και padding 0

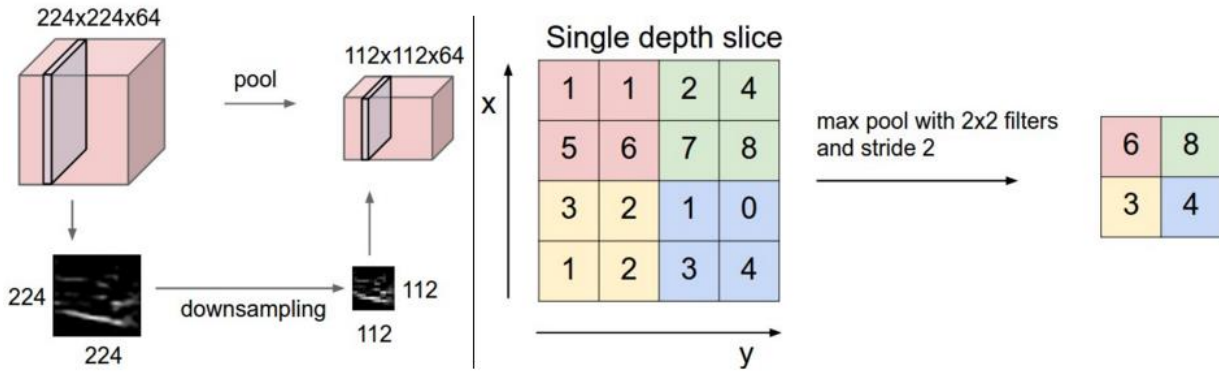
ΠΗΓΗ: <https://docplayer.gr/47274240-Aristoteleio-panepistimio-thessalonikis.html>

2.1.1.3 Συγκέντρωση (Pooling)

Τα επίπεδα συγκέντρωσης ή αλλιώς υποδειγματοληψίας (pooling layers) στα μοντέλα CNN μειώνουν τις χωρικές διαστάσεις της εικόνας (ύψος και πλάτος), εφαρμόζοντας μια συνάρτηση χωρικής υποδειγματοληψίας. Η διαφορά του συνελικτικού επιπέδου με το επίπεδο υποδειγματοληψίας είναι ότι το συνελικτικό επίπεδο ανιχνεύει μοτίβα σε υποπεριοχές της εικόνας εισόδου, ενώ το επίπεδο υποδειγματοληψίας χρησιμεύει στη σταδιακή μείωση της εικόνας εισόδου, τη μείωση του χρόνου εκπαίδευσης του δικτύου και την αποφυγή της υπερμοντελοποίησης (overfitting). Αντίστοιχα με το συνελικτικό επίπεδο, το επίπεδο

συγκέντρωση ή δειγματοληψίας δέχεται επίσης παραμέτρους για τη δήλωση του μεγέθους του φίλτρου και του βήματος σάρωσης.

Οι επικρατέστερες κατηγορίες του pooling είναι τα max, sum και average pooling, ενώ μπορεί τα παράθυρα που χρησιμοποιούνται να επικαλύπτονται ή και όχι ανάλογα με τις ανάγκες του προβλήματος. Στο επόμενο διάγραμμα παρουσιάζεται ο πιο συνηθισμένος τύπος συγκέντρωσης του μέγιστου επιπέδου συγκέντρωσης (max pooling), όπου το παράθυρο μετακινείται με βάση το βήμα και λαμβάνει τη μέγιστη τιμή στο παράθυρο ως έξοδο.



Επίπεδο υποδειγματοληψίας / συγκέντρωσης - Pooling layer
 ΠΗΓΗ: https://leonardoaraujosantos.gitbook.io/artificial_intelligence/machine_learning/deep_learning/pooling_layer

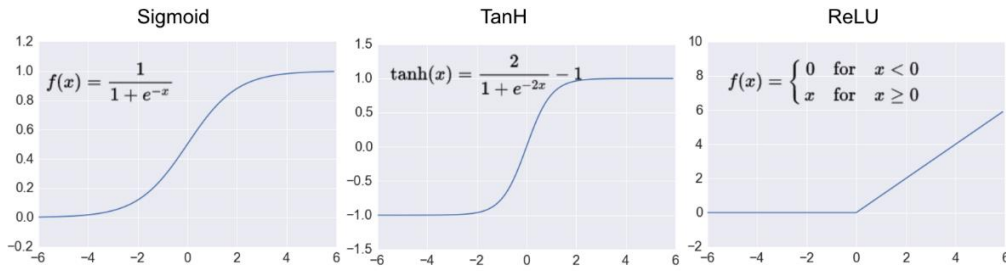
Όταν έχουν πλέον ληφθεί τα χαρακτηριστικά διανύσματα που περιγράφηκαν νωρίτερα, αποφασίζεται το μέγεθος των παραθύρων που θα γίνει pool στα δεδομένα. Τότε διαιρούνται τα χαρακτηριστικά σε περιοχές και λαμβάνονται το μέγιστο (max) ή ο μέσος όρος (mean) του παραθύρου, το οποίο αποτελεί το νέο χαρακτηριστικό. Αυτές οι pooled περιοχές μπορούν πλέον να χρησιμοποιηθούν για ταξινόμηση. Επομένως, μετά τη διαδικασία της συνέλιξης και την δημιουργία των χαρτών χαρακτηριστικών, κάθε χαρακτηριστικό υποδειγματοληπτείται/ελαττώνεται τυπικά με mean ή max pooling σε $p \times p$ συνεχείς περιοχές, όπου το p παίρνει τυπικά τιμές μεταξύ 2 και κοντά στο 5 για μεγάλες εικόνες εισόδου. Πριν ή μετά το pooling layer ακολουθεί μια προσθήκη bias και συνάρτησης μη γραμμικότητας σε κάθε χάρτη χαρακτηριστικών (Παπαδόπουλος, 2016 [137]).

2.1.1.4 Μη γραμμικότητα

Στα δίκτυα CNN η εισαγωγή της μη γραμμικότητας (non-linearity) δίνει σημαντικά πλεονεκτήματα για τη λύση πολύπλοκων προβλημάτων. Μερικά παραδείγματα μη γραμμικών συναρτήσεων ενεργοποίησης αποτελούν η υπερβολική εφραπτομένη και η σιγμοειδής, που έχουν ήδη αναφερθεί σε προηγούμενο κεφάλαιο, και η ευρέως χρησιμοποιούμενη συνάρτηση ReLU.

Η ανορθωμένη Γραμμική Μονάδα (Rectified Linear Unit - ReLU), γνωστή αλλιώς και ως συνάρτηση ράμπας, χρησιμοποιείται όλο και περισσότερο στα κρυφά επίπεδα των βαθιών συνελκτικών νευρωνικών δικτύων, μιας και επιτυγχάνουν καλύτερη εκπαίδευση. Συγκεκριμένα, αυξάνει τις μη γραμμικές ιδιότητες της συνάρτησης ενεργοποίησης και του συνολικού δικτύου χωρίς να επηρεάζει τα δεκτικά πεδία του συνελκτικού επιπέδου. Το μειονέκτημά της συνάρτησης ReLU είναι ότι κατά την διάρκεια της εκπαίδευσης του δικτύου τα βάρη ανανεώνονται με τέτοιο τρόπο που ο νευρώνας μπορεί να μην ενεργοποιηθεί ποτέ. Αυτό έχει σαν αποτέλεσμα να μην λειτουργήσει ποτέ ο συγκεκριμένος νευρώνας παραμένοντας ανενεργός.

$$\varphi(x) = \begin{cases} 0, & \text{για } x < 0 \\ x, & \text{για } x \geq x \end{cases}$$



Μη γραμμικές συναρτήσεις ενεργοποίησης CNN δικτύων

ΠΗΓΗ: https://www.researchgate.net/figure/Activation-function-plots-for-Sigmoid-Tanh-and-ReLU_fig4_337173375

Μια παραλλαγή της συνάρτησης ReLU που επίσης χρησιμοποιείται αρκετά είναι η συνάρτηση Leaky ReLU και προσπαθεί να λύσει το προαναφερθέν πρόβλημα. Αντί να μηδενίζεται για $x < 0$ έχει μία πάρα πολύ μικρή κλίση:

$$\varphi(x) = \begin{cases} x, & \text{για } x > 0 \\ 0.1x, & \text{για } x \leq 0 \end{cases}$$

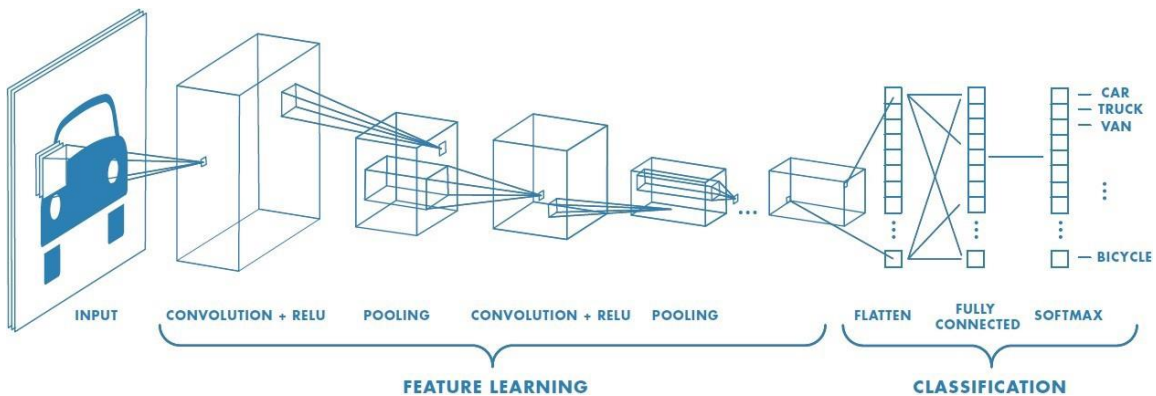
Επιπλέον, υπάρχει και η παραμετρική συνάρτηση ReLU, γνωστή ως PReLU, η οποία πάει ένα βήμα περαιτέρω αυτήν την ιδέα παραμετροποιώντας τον συντελεστή, ώστε να μαθαίνει μαζί με τις άλλες παραμέτρους του δικτύου και να πάρει τη βέλτιστη τιμή:

$$\varphi(x) = \begin{cases} x, & \text{για } x > 0 \\ \alpha x, & \text{για } x \leq 0 \end{cases}$$

Οι συναρτήσεις ενεργοποίησης ReLU θεωρούνται οι καταλληλότερες για τα δίκτυα CNN, λόγω του ότι οι απαιτούμενες πράξεις είναι πολύ λιγότερες, αφού ενεργοποιούν όλους τους νευρώνες ταυτόχρονα, απλοποιούν και επιταχύνουν αρκετά τον αλγόριθμο back propagation.

Όταν χρειάζεται να αντιμετωπιστούν προβλήματα κατηγοριοποίησης, οι προηγούμενες συναρτήσεις δεν μπορούν να βοηθήσουν πολύ. Για παράδειγμα, η σιγμοειδής συνάρτηση μπορεί να χειριστεί μέχρι δύο κλάσεις, κάτι το οποίο πολύ συχνά δεν είναι αρκετό. Η συνάρτηση Softmax, η οποία είναι μία γενίκευση της λογιστικής συνάρτησης, μπορεί να βοηθήσει σε αυτό το πρόβλημα. Χρησιμοποιείται συχνά στο τελευταίο στρώμα ενός δικτύου και λειτουργεί συμπυκνώνοντας τις τιμές αυτές, έτσι ώστε να είναι μεταξύ 0 και 1, αλλά και το άθροισμα τους να ισούται με τη μονάδα. Έτσι, κάθε τιμή εξόδου που προκύπτει από μία συνάρτηση Softmax είναι ισοδύναμη με μία κατηγορική συνάρτηση πιθανότητας:

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}}, \text{ για } j = 1, \dots, K$$



Παράδειγμα χρήσης συναρτήσεων σε ένα δίκτυο CNN

ΠΗΓΗ: <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>

2.1.1.5 Αλγόριθμος οπισθοδιάδοσης σφάλματος

Η εκπαίδευση ενός συνελκτικού νευρωνικού δικτύου γίνεται επίσης με την χρήση του αλγορίθμου οπισθοδιάδοσης σφάλματος (backpropagation) για την ενημέρωση των βαρών από το επίπεδο εξόδου προς το επίπεδο εισόδου. Το σφάλμα κατηγοριοποίησης διαδίδεται από τα πλήρως συνδεδεμένα επίπεδα στα αραιότερα συνδεδεμένα επίπεδα φτάνοντας στο επίπεδο εισόδου. Όπως έχει ήδη αναφερθεί, μόνο τα συνελκτικά και τα πλήρως συνδεδεμένα επίπεδα έχουν νευρώνες με τη συνήθη έννοια, επομένως μόνο αυτά εισάγουν βάρη προς εκπαίδευση στο δίκτυο. Τα υπόλοιπα επίπεδα απλώς επιτελούν μία προκαθορισμένη λειτουργία. Το σφάλμα $\delta_{x,y}^{m,\lambda-1}$ που αντιστοιχεί σε μια εικονοψηφίδα m στη θέση x,y για το επίπεδο $\lambda - 1$ δίνεται ως εξής:

$$\delta_{x,y}^{m,\lambda-1} = \sum_{n=1}^N \sum_{i=d}^x \sum_{j=g}^y \delta_{i,j}^{n,\lambda} k_{x-i,y-j}^{nm} f'(u_{i,j}^{n,\lambda})$$

$$d = \max(0, x - k_o + 1), \quad g = \max(0, y - k_o + 1)$$

Η συνάρτηση ενημέρωσης των βαρών ορίζεται ως εξής:

$$k_{x,y}^n = k_{i,j}^0 - e \left(\sum_{i=0}^{M_0} \sum_{j=0}^{M_0} \delta_{i,j}^\lambda a_{i+x,j+y}^{\lambda-1} \right)$$

όπου το $k_{x,y}^n$ είναι η εικονοψηφίδα που βρίσκεται στη θέση i,j και ενώνει τα εικονοστοιχεία m,n , ενώ f η συνάρτηση ενεργοποίησης και k_o το πλήθος των εικονοστοιχείων, M_0 το μέγεθος των εικόνων και N το πλήθος των εικόνων του επιπέδου λ .

Παρατηρείται ότι το σφάλμα μειώνεται όσο γίνεται η προώθησή του στα αρχικά επίπεδα του δικτύου. Υπάρχουν, επίσης, διάφορες μέθοδοι για να μειωθούν τα σφάλματα εκπαίδευσης ενός δικτύου. Για δίκτυα σαν τα CNN, των οποίων η εκπαίδευση μπορεί να κρατήσει αρκετές ημέρες, έχει αναπτυχθεί μια μέθοδος σύγκρισης μοντέλων αρκετά αποτελεσματική με πολύ χαμηλό υπολογιστικό κόστος. Η τεχνική αυτή, η οποία ονομάζεται Dropout, είναι μία τεχνική όπου τυχαία επιλεγμένοι νευρώνες αγνοούνται κατά την διάρκεια της εκπαίδευσης. Εγκαταλείπονται εντελώς τυχαία. Αυτό σημαίνει πως η συμβολή τους στη διαδικασία της οπισθοδιάδοσης του σφάλματος δεν είναι εφικτή και τα βάρη τους δεν συνεισφέρουν στην εκπαίδευση τη δεδομένη στιγμή. Έτσι, κάθε φορά που μια είσοδος παρουσιάζεται στο δίκτυο, εκείνο χρησιμοποιεί διαφορετική αρχιτεκτονική με όλες τις αρχιτεκτονικές να μοιράζονται τα ίδια βάρη.

Εάν οι νευρώνες εγκαταλειφθούν τυχαία από το δίκτυο κατά τη διάρκεια της εκπαίδευσης, οι άλλοι νευρώνες θα πρέπει να μπουν και να χειριστούν την αναπαράσταση που απαιτείται για να κάνουν προβλέψεις για τους νευρώνες που λείπουν. Αυτό πιστεύεται ότι οδηγεί σε πολλές ανεξάρτητες εσωτερικές αναπαραστάσεις που μαθαίνονται από το δίκτυο. Το αποτέλεσμα είναι ότι το δίκτυο γίνεται λιγότερο ευαίσθητο στα συγκεκριμένα βάρη των νευρώνων. Αυτό με τη σειρά του οδηγεί σε ένα δίκτυο ικανό για καλύτερη γενίκευση και είναι λιγότερο πιθανό να υπάρξει υπέρ-μοντελοποίηση. Η χρήση του Dropout σχεδόν διπλασιάζει τον αριθμό των επαναλήψεων που απαιτούνται για σύγκλιση.

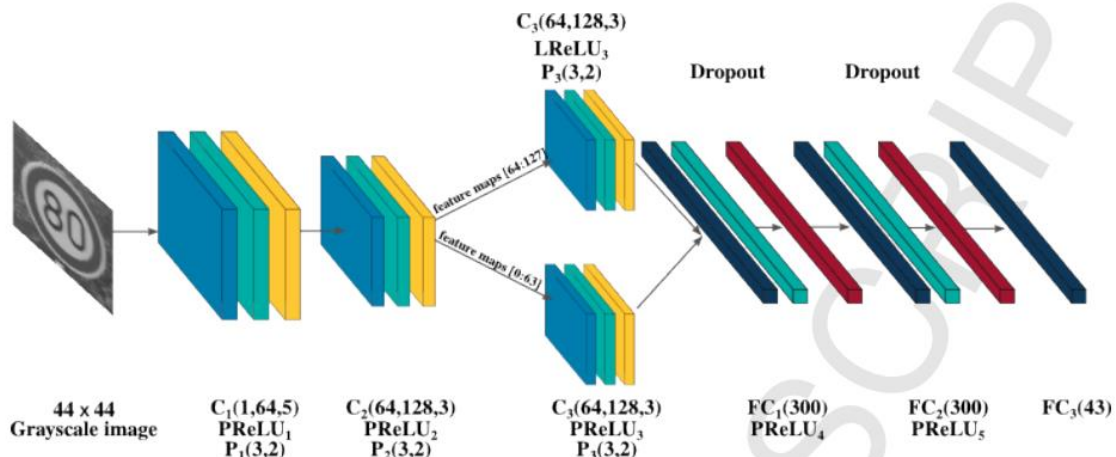
2.1.1.6 Ανίχνευση και Αναγνώριση με δίκτυα CNN

Οι περισσότεροι ερευνητές που εμπνεύστηκαν από τις εφαρμογές των νευρωνικών δικτύων τα τελευταία χρόνια άρχισαν να εφαρμόζουν τη βαθιά μάθηση με τη χρήση συνελκτικών νευρωνικών δικτύων για την αντιμετώπιση των προβλημάτων της ανίχνευσης και αναγνώρισης των σημάτων οδικής κυκλοφορίας.

Σύμφωνα με τους Liu et al. (2019) [105], η χρήση των δικτύων CNN στις πινακίδες κυκλοφορίας ξεκίνησε με τους Wu et al. (2013) [106], με στόχο την ανίχνευση των πινακίδων και την αναγνώρισή τους ως μίας από τις τρεις μεγάλες κατηγορίες σημάτων:

«Απαγορευτικές», «Υποχρεωτικές» και «Κινδύνου». Οι εικόνες έχουν αρχικά υποστεί προεπεξεργασία από έναν ταξινομητή SVM, ώστε να μετατραπούν από RGB σε ασπρόμαυρες. Με αυτόν τον τρόπο μπορεί εύκολα να αποφευχθεί το πρόβλημα της ευαισθησίας στη διαφορά του χρώματος, λόγω των διαφορετικών συνθηκών φωτισμού. Στη συνέχεια, μέσω ενός δικτύου CNN εποπτευόμενης μάθησης, εντοπίζονται οι περιοχές ενδιαφέροντος και από κάθε επίπεδο νευρώνων εξάγονται τα διακριτά χαρακτηριστικά τα οποία εν τέλει θα ταξινομηθούν. Το δίκτυο 2-επιπέδων είναι πλήρως συνδεδεμένο με 100 νευρώνες στο πρώτο επίπεδο και δύο στο δεύτερο. Οι συγγραφείς εξασφάλισαν επιτυχία της καμπύλης ακρίβειας - ανάκλησης (precision - recall curve) 99.72% στα προειδοποιητικά σήματα κινδύνου και 97.62% στα υποχρεωτικά σήματα οδικής κυκλοφορίας, ενώ στο σύνολο των δεδομένων ελέγχου επιτεύχθηκε ποσοστό κατά μέσο όρο 98.68%. Παρά την υψηλή ακρίβεια, αυτό το μοντέλο δεν μπορεί να κάνει ανίχνευση σε πραγματικό χρόνο, καθώς ο χρόνος επεξεργασίας είναι αρκετά μεγάλος.

Λόγω της επιτυχίας του δικτύου CNN στην ταξινόμηση σημάτων κυκλοφορίας, οι Aghdam et al. (2016) [107] πρότειναν ένα «ελαφρύ» σε αριθμητικές πράξεις και ακριβές δίκτυο CNN με κυλιόμενο παράθυρο (sliding window) χρησιμοποιώντας διευρυνόμενες συνελίξεις (dilated convolutions) σε εικόνες υψηλής ανάλυσης. Η βασική ιδέα του αλγορίθμου είναι να χωριστούν τα ενδιάμεσα συνελικτικά επίπεδα σε δύο μέρη, με στόχο να μειωθεί η τρίτη διάσταση των συνελικτικών πυρήνων (kernels) και να χρησιμοποιηθεί ως συνάρτηση ενεργοποίησης η LReLU (Leaky Rectified Linear Units), αντί για συναρτήσεις ενεργοποίησης που έχουν μεγάλο υπολογιστικό κόστος, όπως η συνάρτηση της υπερβολής. Η συνάρτηση LReLU χρησιμοποιεί μόνο μία σύγκριση και έναν πολλαπλασιασμό για τον υπολογισμό της εξόδου, με αποτέλεσμα να μειώνεται δραστικά το υπολογιστικό κόστος. Στα υπόλοιπα επίπεδα, γίνεται χρήση της παραμετρικής συνάρτησης ReLU (PReLU). Στο παρακάτω σχήμα παρατίθεται η αρχιτεκτονική του δικτύου CNN, όπου με μπλε, πράσινο και κίτρινο χρώμα υποδηλώνεται μια συνέλιξη με συνάρτηση ReLU και ένα επίπεδο συγκέντρωσης P (k, s) (pooling), με μέγεθος συγκέντρωσης $k \times k$ και βήμα s, αντίστοιχα. Και ως C (c, n, k) υποδηλώνεται η συνέλιξη πυρήνα μεγέθους $k \times k \times c$. Τέλος, ο αριθμός στις μονάδες συνάρτησης LReLU δείχνει τον συντελεστή διαρροής της συνάρτησης ενεργοποίησης.

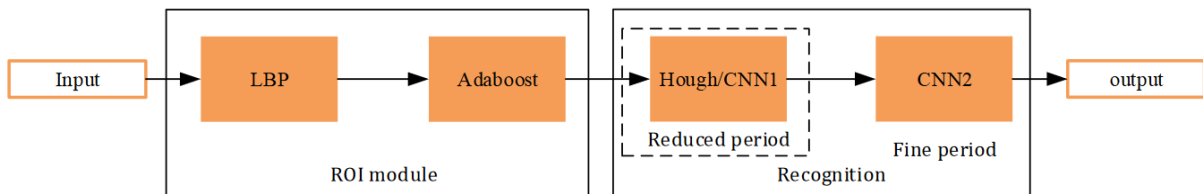


Αρχιτεκτονική του δικτύου CNN
 ΠΗΓΗ: Aghdam et al. (2016) [107]

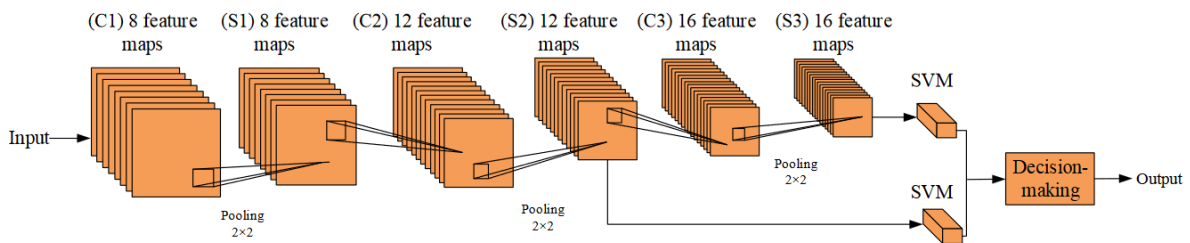
Το δίκτυο εντοπίζει τις πινακίδες κυκλοφορίας με μέση ακρίβεια ίση με 99.89%, στο σύνολο δεδομένων για την ανίχνευση γερμανικών σημάτων κυκλοφορίας, ενώ για την ταξινόμηση η οποία γίνεται με βάση το σχήμα είναι σε θέση να ταξινομήσει σωστά το 99.55% των δεδομένων ελέγχου. Επιπλέον, η εκτέλεση χρόνου στην GPU (GeForce GTX 980) είναι 26.506 ms, η οποία ισοδυναμεί με την επεξεργασία 3772 καρτέ ανά δευτερόλεπτο. Τα αποτελέσματα αυτά καθιστούν την προσέγγιση αποτελεσματική για μια εφαρμογή πραγματικού χρόνου.

Τα διαδοχικά συνελικτικά νευρωνικά δίκτυα (cascade CNN) χρησιμοποιήθηκαν από τους Zang et al. (2016) [108] για τη μείωση ψευδώς θετικών περιοχών που προέκυψαν όταν χρησιμοποιήθηκε ο τοπικός δυαδικός ανιχνευτής χαρακτηριστικών (LBP) σε συνδυασμό με

τον ταξινομητή AdaBoost. Η δομή του δικτύου CNN περιλαμβάνει ένα επίπεδο εισόδου, τρία συνελκτικά επίπεδα με διαφορετικά μεγέθη πυρήνα και τρία επίπεδα δειγματοληψίας. Το κομμάτι της ταξινόμησης αντικαθίσταται από SVM, όπου τα αποτελέσματα από το δεύτερο και το τρίτο επίπεδο δειγματοληψίας εισάγονται σε δύο διαφορετικούς ταξινομητές SVM για ταξινόμηση, ενώ εφαρμόζεται ένα μπλοκ λήψης αποφάσεων για τον προσδιορισμό του τελικού αποτελέσματος αναγνώρισης. Τα αποτελέσματα που επιτεύχθηκαν για το γερμανικό σύνολο δεδομένων (GTSDB) ήταν 99.45% για τις απαγορευτικές πινακίδες, 96.50% για τις υποχρεωτικές πινακίδες και 98.33% για τις πινακίδες κινδύνου.



Αρχιτεκτονική του συστήματος (1)
ΠΗΓΗ: Zang et al. (2016) [108]



Αρχιτεκτονική του συστήματος (2)
ΠΗΓΗ: Zang et al. (2016) [108]

Οι Xie et al. (2016) [109] παρατήρησαν πως το 80% των λανθασμένα ταξινομημένων πινακίδων έχουν το ίδιο χρώμα, σχήμα και εικονόγραμμα. Για να ξεπεραστεί αυτό το πρόβλημα, πρότειναν επίσης ένα διαδοχικό συνελκτικό νευρωνικό δίκτυο (cascade CNN), αλλά δύο σταδίων. Το πρώτο στάδιο του CNN δικτύου εκπαιδεύεται στην ετικέτα (label) της κατηγορίας, ενώ το δεύτερο στάδιο εκπαιδεύεται σε υπερ-κλάσεις (super-classes) ξεχωριστά, σύμφωνα με το σχήμα και το εικονόγραμμα των σημάτων. Η ακρίβεια της προτεινόμενης μεθόδου είναι 97.94% στο γερμανικό dataset GTSRB, ενώ ο αριθμός των σφαλμάτων μειώνεται με το διαδοχικό δίκτυο CNN κατά από 50% (από 430 σε 202). Ο χρόνος εκτέλεσης της προτεινόμενης μεθόδου δεν αναφέρεται στην εργασία.

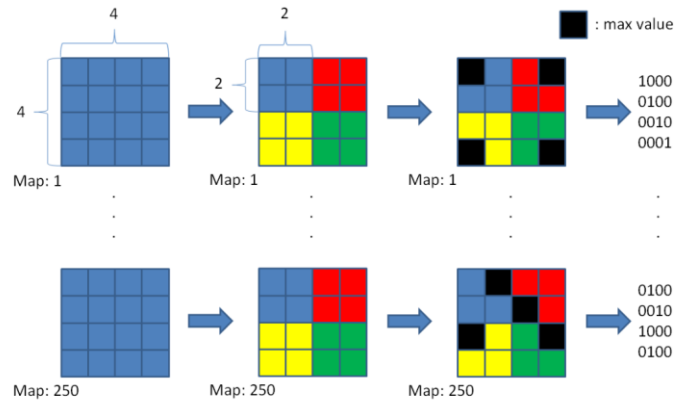
Στην εργασία των Qian et al. (2016) [110] τα δίκτυα CNN χρησιμοποιήθηκαν ως εργαλείο εξαγωγής χαρακτηριστικών και το πολυστρωματικό perceptron (MLP) ως ταξινομητής. Σε σύγκριση με τα κλασικά συνελκτικά νευρωνικά δίκτυα, στο επίπεδο μέγιστης συγκέντρωσης οι συγγραφείς δεν χρησιμοποιούν τις μέγιστες τιμές αλλά τις μέγιστες θέσεις τους. Οι μέγιστες θέσεις συγκέντρωσης (Max Pooling Positions - MPPs) συνίστανται στην κωδικοποίηση κάθε θέσης μέγιστης τιμής σε δυαδική τιμή 4 bit και στη συνέχεια σε συνδυασμό όλων των τιμών μέγιστων θέσεων για τη λήψη της δυνατότητας MPPs.



Επισκόπηση Συστήματος
ΠΗΓΗ: Qian et al. (2016) [110]

Δεδομένου ότι το τελικό στρώμα του προτεινόμενου μοντέλου CNN έχει 250 χαρακτηριστικών με νευρώνες 4×4 , τα εξαγόμενα χαρακτηριστικά είναι 4000. Για κάθε χάρτη χαρακτηριστικών 4×4 , εκτελείται μέγιστη συγκέντρωση με μέγεθος πυρήνα 2×2 και βήμα

2, και καταγράφεται η θέση κάθε μέγιστης τιμής. Στη συνέχεια, κάθε καταγεγραμμένη θέση κωδικοποιείται σε δυαδική τιμή 4 bit. Τέλος, ολόκληρη η ακολουθία MPPs μπορεί να υλοποιηθεί με συνδυασμό όλων των δυαδικών τιμών. Η διάσταση της ακολουθίας MPPs είναι επίσης 4000. Η ακρίβεια που επιτυγχάνεται με την χρήση MPPs αυξάνεται στο 98.86% στο GTSRB.



Επισκόπηση Συστήματος
 ΠΗΓΗ: Qian et al. (2016) [110]

Οι Qian et al. (2015) [111] πρότειναν την ανίχνευση και αναγνώριση των κινεζικών σημάτων οδικής κυκλοφορίας εφαρμόζοντας ένα βαθύ συνελκτικό δίκτυο, το οποίο είχε υψηλή απόδοση σε σχέση με το ποσοστό ανίχνευσης και την ακρίβεια αναγνώρισης. Σε σύγκριση με άλλες δημοσιευμένες μεθόδους που συνήθως περιορίζονται σε ένα προκαθορισμένο σύνολο πινακίδων κυκλοφορίας, το προτεινόμενο σύστημα είναι πιο ολοκληρωμένο καθώς οι περιοχές ενδιαφέροντος περιλαμβάνουν τις πινακίδες κυκλοφορίας με ψηφία, αγγλικά γράμματα και κινεζικούς χαρακτήρες. Το σύστημα βασίζεται σε ένα δίκτυο CNN πολλαπλών εργασιών (multi-task CNN) που έχει εκπαιδευτεί για τον αποτελεσματικό εντοπισμό και την απόκτηση χαρακτηριστικών για τον εντοπισμό και την ταξινόμηση διαφορετικών πινακίδων κυκλοφορίας και κειμένων.

Το δίκτυο αποτελείται από τρία συνδεδεμένα στάδια συνέλιξης και δύο τελικά επίπεδα softmax. Κάθε στάδιο συνέλιξης αποτελείται από το επίπεδο συνέλιξης, το επίπεδο μη γραμμικής ενεργοποίησης και το επίπεδο μέγιστης συγκέντρωσης (max pooling). Ως συνάρτηση ενεργοποίησης χρησιμοποιείται η συνάρτηση ReLU. Το τελικό επίπεδο μέγιστης συγκέντρωσης μοιράζεται από δύο πλήρως συνδεδεμένα επίπεδα, δηλαδή το πολλαπλό επίπεδο perceptron (MLP) που αντιστοιχεί στην πρόβλεψη και την ταξινόμηση, αντίστοιχα. Το πρώτο MLP έχει σχεδιαστεί για ανίχνευση, με 64 κρυφά επίπεδα και 2 επίπεδα εξόδου για θετική/αρνητική απόφαση. Το δεύτερο MLP έχει σχεδιαστεί για την ταξινόμηση συγκεκριμένων κατηγοριών ενός αντικειμένου που εντοπίστηκε. Υπάρχουν 96 κλάσεις αντικειμένων, συμπεριλαμβανομένων πινακίδων κυκλοφορίας, ψηφίων, αγγλικών γραμμάτων και κινεζικών χαρακτήρων.

Layer	Type	Feature maps & Size		kernal
1	Input	1 map with 48×48 neurons		
2	Convolution	100 maps with 44×44 neurons		5×5
3	Max pooling	100 maps with 22×22 neurons		2×2
4	Convolution	150 maps with 20×20 neurons		3×3
5	Max pooling	150 maps with 10×10 neurons		2×2
6	Convolution	250 maps with 8×8 neurons		3×3
7	Max pooling	250 maps with 4×4 neurons		2×2
8	Fully connection	512 neurons		
		Task 1	Task 2	
9	Fully connection	64 neurons	256 neurons	
10	Softmax	2 neurons	96 neurons	

Λεπτομέρειες των παραμέτρων του CNN δικτύου βαθιάς μάθησης
 ΠΗΓΗ: Qian et al. (2015) [111]

Πρόσφατα σε δημοσίευση του Islam (2019) [112] χρησιμοποιήθηκαν δύο ξεχωριστά συνελκτικά νευρωνικά δίκτυα για σκοπούς ανίχνευσης και αναγνώρισης 28 διαφορετικών

σημάτων οδικής κυκλοφορίας, τα οποία χρησιμοποιούνται σε όλο τον κόσμο. Το πρώτο συνελκτικό δίκτυο ταξινομεί την πινακίδα για τον εντοπισμό, με ένα μόνο στάδιο συνέλιξης και συνάρτηση ενεργοποίησης ReLU, και δυο πλήρως συνδεδεμένα επίπεδα με συνάρτηση ενεργοποίησης ReLU και Softmax. Αντίστοιχη δομή έχει και το δεύτερο δίκτυο CNN με τη διαφορά πως αποτελείται από δύο στάδια συνέλιξης για την ταξινόμηση των πινακίδων με βάση το σχήμα.

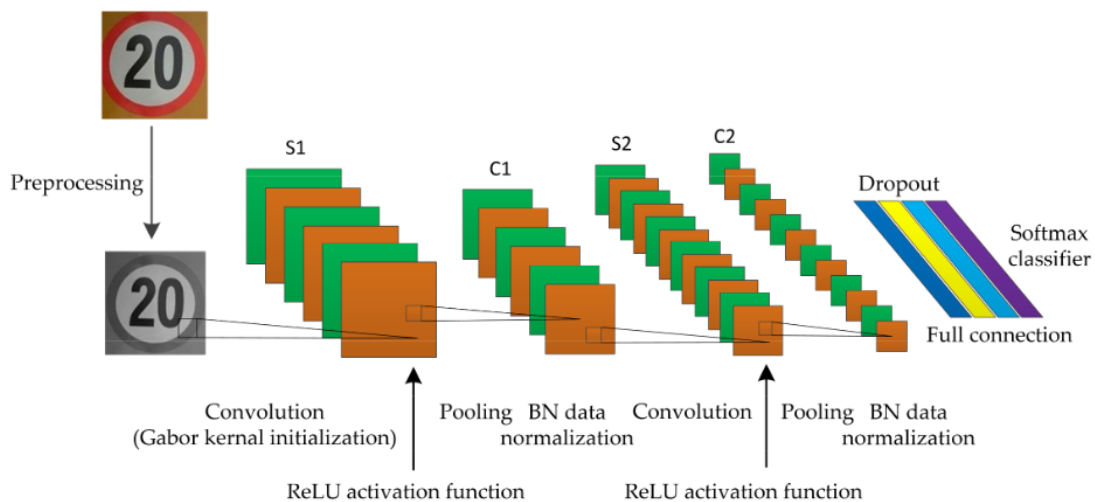


Διαφορές μεταξύ αντίστοιχων σημάτων από διαφορετικές χώρες
 ΠΗΓΗ: Islam (2019) [112]

Δημιουργήθηκε ένα σύνολο δεδομένων εκπαίδευσης και ελέγχου, αποτελούμενο από 40.000 εικόνες για την εκπαίδευση του πρώτου δικτύου CNN με 28000 θετικές εικόνες (εικόνες που περιέχουν σήματα κυκλοφορίας) και 12000 αρνητικές εικόνες (εικόνες που δεν περιέχουν σήματα κυκλοφορίας). Για το δεύτερο δίκτυο CNN χρησιμοποιήθηκαν 3600 εικόνες για την εκπαίδευση, εκ των οποίων 2400 θετικές και 1200 αρνητικές. Οι εικόνες αρχικά υφίστανται επεξεργασία για τον εντοπισμό της περιοχής ενδιαφέροντος, η οποία στη συνέχεια τροφοδοτεί δύο ταξινομητές CNN για ταξινόμηση. Τα αποτελέσματα της εφαρμογής είναι αρκετά υψηλά αφού επιτυγχάνεται συνολική ακρίβεια 0.90.

Οι Alghamghama et al. (2019) [113] χρησιμοποίησαν βαθιά συνελκτικά νευρωνικά δίκτυα για την ανάπτυξη ενός συστήματος αυτόματης ανίχνευσης και αναγνώρισης οδικής σήμανσης (ATRS). Το προτεινόμενο σύστημα λειτουργεί σε πραγματικό χρόνο. Η αρχιτεκτονική του δικτύου deep-CNN που προτείνεται αποτελείται από δύο συνελκτικά επίπεδα, δύο επίπεδα μέγιστης συγκέντρωσης (max pooling), ένα επίπεδο εγκατάλειψης (dropout) και τρία πυκνά επίπεδα (dense layers). Επιτεύχθηκε ακρίβεια 100% για την εποχή 150 για κάθε σύνολο διαφορετικού μεγέθους.

Ένας βελτιωμένος αλγόριθμος ανίχνευσης και αναγνώρισης πινακίδων κυκλοφορίας για έξυπνα οχήματα προτείνεται από τους Cao et al. (2019) [114], ώστε να αντιμετωπίσει προβλήματα όπως το πόσο εύκολα επηρεάζεται η παραδοσιακή ανίχνευση των σημάτων κυκλοφορίας από το περιβάλλον και την ισχνή απόδοση που έχουν οι μέθοδοι βαθιάς μάθησης για αναγνώριση σε πραγματικό χρόνο. Οι συγγραφείς χρησιμοποιούν το κλασικό μοντέλο συνελκτικού νευρωνικού δικτύου LeNet-5, το οποίο βελτίωσαν προσθέτοντας τον πυρήνα Gabor ως αρχικό συνελκτικό πυρήνα του κόμβου και προσθέτοντας τη διαδικασία ομαλοποίησης (batch normalization) μετά το επίπεδο συγκέντρωσης (pooling), ενώ τέλος επιλέγεται η μέθοδος Adam ως αλγόριθμος βελτιστοποίησης.



Το μοντέλο της βελτιωμένης δομής του δικτύου LeNet-5.
 ΠΗΓΗ: Cao et al. (2019) [114]

Συνοπτικά, η βελτίωση του μοντέλου για το δίκτυο LeNet-5 περιλαμβάνει τα ακόλουθα πέντε βήματα:

1. Ο πυρήνας Gabor χρησιμοποιείται ως ο αρχικός συνελκτικός πυρήνας μεταξύ του επιπέδου εισόδου και του πρώτου συνελκτικού επιπέδου. Το δισδιάστατο φίλτρο Gabor, το οποίο έχει καλή προσαρμοστικότητα στο φως, είναι ένα φίλτρο διέλευσης ζώνης του οποίου η συνάρτηση έχει ως εξής:

$$g(x, y, f, \theta) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left(-\frac{k_1^2}{2\sigma_x^2} - \frac{k_2^2}{2\sigma_y^2}\right) \exp(i(f_x x + f_y y))$$

$$k_1 = x\cos\theta + y\sin\theta, k_2 = -x\sin\theta + y\cos\theta$$

όπου f είναι η κεντρική συχνότητα του εύρους ζώνης, θ η χωρική κατεύθυνση της οποίας η τιμή κυμαίνεται $[0, \pi]$, σ_x και σ_y είναι οι τυπικές αποκλίσεις στις κατευθύνσεις x και y , αντίστοιχα, $f_x = f \cdot \cos\theta$ και $f_y = f \cdot \sin\theta$ είναι και οι δύο συχνότητες του διαστήματος.

2. Μετά από κάθε επίπεδο συγκέντρωσης, η διαδικασία ομαλοποίησης (batch normalization - BN) προστίθεται για ομαλοποίηση δεδομένων.

Input: Mini-batch input $x: B = \{x_1, \dots, m\}$

Output: Normalized network response $\{y_i = BN_{\gamma, \beta}(x_i)\}$

Ομαλοποίηση: $\hat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}}$

όπου ϵ είναι ο ελάχιστος θετικός αριθμός που χρησιμοποιείται για την αποφυγή διαίρεσης με το 0.

3. Η συνάρτηση ReLU επιλέγεται ως συνάρτηση ενεργοποίησης, λόγω του ότι είναι απλή στον υπολογισμό και λειτουργεί αποτελεσματικά, γεγονός που βοηθά στη μείωση της υπολογιστικής πολυπλοκότητας και την επιτάχυνση της σύγκλισης του αλγορίθμου.
4. Η μέθοδος Adam επιλέγεται ως αλγόριθμος βελτιστοποίησης. Αυτή η μέθοδος είναι ένας διευρυμένος αλγόριθμος βελτιστοποίησης πρώτης τάξης που βασίζεται στη μέθοδο stochastic gradient descent, ο οποίος μπορεί δυναμικά να προσαρμόσει τον ρυθμό εκμάθησης των σχετικών παραμέτρων χρησιμοποιώντας την εκτίμηση ροής της κλίσης. Η μέθοδος Adam μπορεί να ελέγξει κάθε επαναληπτικό ρυθμό εκπαίδευσης εντός ενός συγκεκριμένου εύρους, διασφαλίζοντας έτσι την ομαλή ενημέρωση των παραμέτρων του δικτύου:

$$\theta_t = \theta_{t-1} - \frac{\eta}{\sqrt{\theta_t + \epsilon}} \hat{m}_t$$

5. Η εγκατάλειψη (dropout) προστίθεται στα πλήρως συνδεδεμένα επίπεδα.

Τα πειράματα ταξινόμησης και αναγνώρισης πραγματοποιούνται με βάση το γερμανικό σύνολο δεδομένων. Τα πειραματικά αποτελέσματα δείχνουν ότι ο ακριβής ρυθμός αναγνώρισης των πινακίδων κυκλοφορίας φτάνει το 99.75% και ο μέσος χρόνος επεξεργασίας ανά καρτέ είναι 5.4 msec. Σε σύγκριση με άλλους αλγορίθμους state-of-the-art, ο προτεινόμενος έχει αξιοσημείωτη ακρίβεια και απόδοση σε πραγματικό χρόνο, ισχυρή ικανότητα γενίκευσης και υψηλή αποτελεσματικότητα εκπαίδευσης. Αυτή η σημαντική βελτίωση του μοντέλου έχει μεγάλη σημασία για τη μείωση του ποσοστού ατυχημάτων και την ενίσχυση της κατάστασης της οδικής ασφάλειας, παρέχοντας μια ισχυρή τεχνική εγγύηση για τη σταθερή ανάπτυξη έξυπνων συστημάτων υποβοήθησης της οδήγησης (ADAS) ή για την πλήρως αυτοματοποιημένη οδήγηση.

Σε πρόσφατη εργασία των Liang et al. (2019) [115] για την εφαρμογή ενός συστήματος εντοπισμού και αναγνώρισης, αναπτύχθηκε ένα δίκτυο δύο σταδίων. Στο πρώτο στάδιο για τον εντοπισμό υιοθετείται μια βαθιά αρχιτεκτονική πυραμίδας χαρακτηριστικών με πλευρικές συνδέσεις, η οποία καθιστά πιο ευαίσθητα τα σημασιολογικά χαρακτηριστικά μικρών αντικειμένων. Στο στάδιο της ταξινόμησης, χρησιμοποιείται ένα πυκνά συνδεδεμένο συνελκτικό δίκτυο για την ενίσχυση της μετάδοσης των χαρακτηριστικών και του

πολυπλέγματος, γεγονός που οδηγεί σε ακριβέστερη ταξινόμηση με μικρότερο αριθμό παραμέτρων. Τα πειράματα απέδειξαν ότι η προτεινόμενη μέθοδος επιτυγχάνει μια πολύ καλή απόδοση και ξεπερνά τις άλλες state-of-the-art μεθόδους. Η αρχιτεκτονική πυραμίδας κάνει πλήρη χρήση λειτουργιών χαμηλότερου επιπέδου για την ανίχνευση πολύ μικρών σημάτων κυκλοφορίας και η κάθετη σύνδεση βοηθά στην ενίσχυση της σχέσης μεταξύ λειτουργιών χαμηλού επιπέδου και υψηλού επιπέδου.

2.1.2 Αλγόριθμος R-CNN

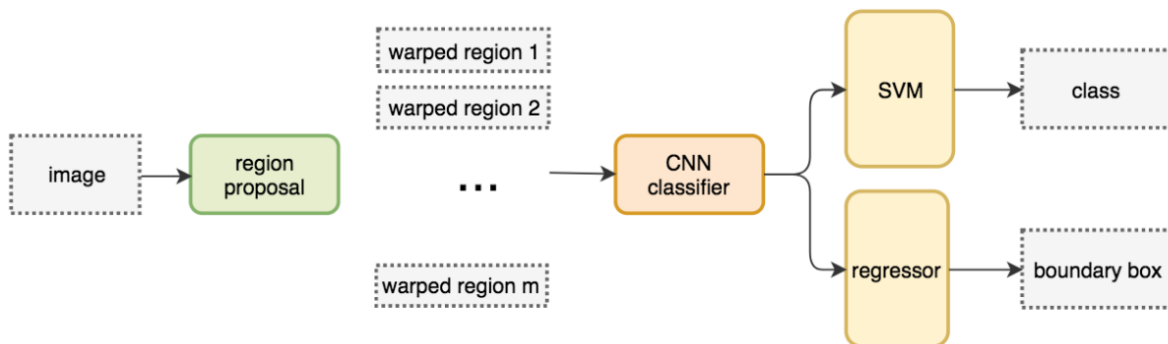
Μία άλλη αρκετά διαδεδομένη μέθοδος που χρησιμοποιείται για τον εντοπισμό και την αναγνώριση σημάτων οδικής κυκλοφορίας είναι τα Συνελικτικά Νευρωνικά Δίκτυα ανά περιοχές (R-CNN, Fast R-CNN, Faster-RCNN)

Η λογική των Συνελικτικών Νευρωνικών Δικτύων ανά περιοχές (Regions with CNN features - R-CNN) βασίζεται σε μία προσέγγιση των Girshick et al. (2014) [116] τριών σταδίων:

1. Εξάγονται οι πιθανές περιοχές ενδιαφέροντος (ROIs) με βάση μια μέθοδο πρότασης περιοχής (Region Proposal Method)
2. Ένα δίκτυο CNN εξαγεί χαρακτηριστικά από αυτές.
3. Κάθε περιοχή ενδιαφέροντος ταξινομείται με τη μέθοδο SVM

Για να παρακάμψουν το πρόβλημα της επιλογής ενός τεράστιου αριθμού περιοχών, οι Girshick et al. (2014) πρότειναν μια μέθοδο όπου χρησιμοποιεί τη επιλεκτική αναζήτηση (Selective Search) για να εξαγάγει μόλις 2000 περιοχές από την εικόνα. Αυτή η μέθοδος πρότασης των πιθανών περιοχών ενδιαφέροντος (ROIs) ονομάστηκε μέθοδος πρότασης περιοχής (Region Proposal Method). Επομένως, αντί να ταξινομηθεί ένας τεράστιος αριθμός περιοχών ταξινομούνται μόνο 2000 περιοχές. Οι περιοχές αυτές, αφού αρχικά ανασχηματιστούν ώστε να έχουν όλες το ίδιο μέγεθος, δίνονται ως είσοδος σε ένα δίκτυο CNN για την εξαγωγή χαρακτηριστικών, ενώ παράγεται ένα διάνυσμα χαρακτηριστικών 4096 διαστάσεων ως έξοδος.

Στη συνέχεια, τα εξαγόμενα χαρακτηριστικά τροφοδοτούν ένα ταξινομητή SVM για να ταξινομήσουν την παρουσία του αντικειμένου μέσα στην πρόταση της υποψήφιας περιοχής. Εκτός από την πρόβλεψη της παρουσίας ενός αντικειμένου εντός των προτάσεων περιοχής, ο αλγόριθμος προβλέπει επίσης τέσσερις τιμές που. Οι τιμές αυτές αφορούν τη μετατόπιση του πλαισίου οριοθέτησης (bounding box), ώστε να έχει μεγαλύτερη ακρίβεια. Για παράδειγμα, δεδομένης μιας πρότασης περιοχής, ο αλγόριθμος θα είχε προβλέψει την παρουσία ενός σήματος οδικής κυκλοφορίας, αλλά το σήμα μέσα σε αυτήν την πρόταση περιοχής θα μπορούσε να έχει κοπεί στα μισά. Επομένως, οι τιμές μετατόπισης βοηθούν στην προσαρμογή του πλαισίου οριοθέτησης της πρότασης περιοχής.

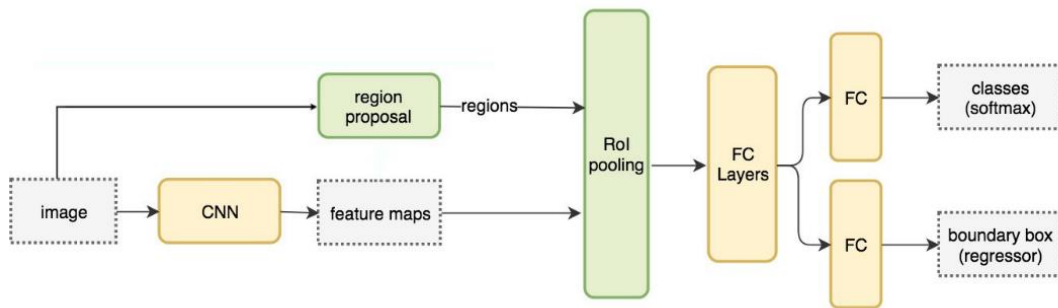


Δίκτυο R-CNN
ΠΗΓΗ: Μαρινέλλης (2018) [57]

Το δίκτυο R-CNN, παρότι παρέχει αρκετή ακρίβεια στα αποτελέσματά του, χρειάζεται πάρα πολύ χρόνο για να εκπαιδευτεί, καθώς θα πρέπει να ταξινομήσει 2000 προτάσεις περιοχής

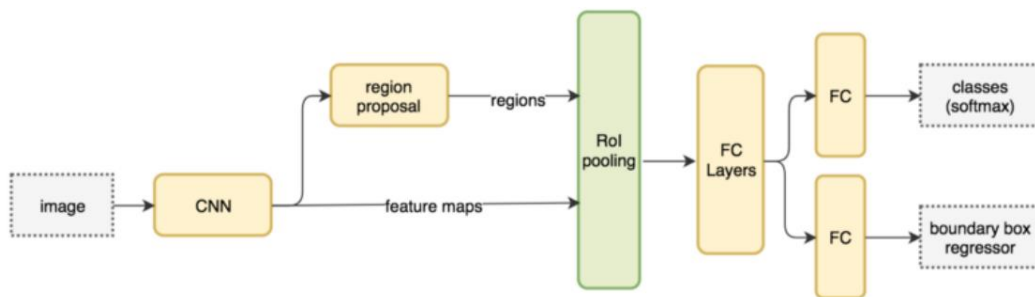
ανά εικόνα. Για κάθε εικόνα ελέγχου χρειάζεται περίπου 47 δευτερόλεπτα, επομένως δεν μπορεί να εφαρμοστεί σε εφαρμογές πραγματικού χρόνου.

Οι ίδιοι συγγραφείς (Girshick, 2015 [117]) έλυσαν μερικά από τα μειονεκτήματα του δικτύου R-CNN προκειμένου να δημιουργηθεί ένας ταχύτερος αλγόριθμος ο οποίος ονομάστηκε Fast R-CNN. Το δίκτυο επιλύει το παραπάνω πρόβλημα εφαρμόζοντας τη μέθοδο εξαγωγής χαρακτηριστικών κατευθείαν πάνω στην αρχική εικόνα. Ουσιαστικά, η προσέγγιση είναι παρόμοια με τον αλγόριθμο R-CNN, όμως, αντί να γίνεται τροφοδότηση των 2000 προτάσεων περιοχής στο δίκτυο CNN κάθε φορά, η λειτουργία της συνέλιξης γίνεται μία φορά για κάθε εικόνα, μέσω της εισαγωγής της απευθείας στο δίκτυο CNN, ώστε να δημιουργηθεί ένας χάρτης χαρακτηριστικών από αυτήν. Το αποτέλεσμα είναι η παραγωγή των χαρτών χαρακτηριστικών όλων των ROIs σε μόλις δύο βήματα. Οι χάρτες αυτοί υπόκεινται πρώτα σε ανασχηματισμό ώστε να αποκτήσουν ίσες διαστάσεις και στη συνέχεια δίνονται ως είσοδος στα πλήρως συνδεδεμένα επίπεδα του δικτύου για την πρόβλεψη της κλάσης και την εύρεση του πλαισίου οριοθέτησης. Με αυτή την προσέγγιση ο συνολικός χρόνος εκτέλεσης της μεθόδου για την ανίχνευση του αντικειμένου ελαττώνεται σημαντικά.



Δίκτυο Fast R-CNN
 ΠΗΓΗ: Μαρινέλλης (2018) [57]

Και οι δύο παραπάνω αλγόριθμοι (R-CNN & Fast R-CNN) χρησιμοποιούν επιλεκτική αναζήτηση για να μάθουν τις προτάσεις της περιοχής. Η επιλεκτική αναζήτηση είναι μια αργή και χρονοβόρα διαδικασία που επηρεάζει την απόδοση του δικτύου. Έτσι, οι Ren et al. (2015) [118] βρήκαν έναν αλγόριθμο ο οποίος εξαλείφει τον αλγόριθμο επιλεκτικής αναζήτησης και επιτρέπει στο δίκτυο να μάθει τις προτάσεις περιοχής. Ο αλγόριθμος αυτός ονομάστηκε Faster R-CNN.



Δίκτυο Faster R-CNN
 ΠΗΓΗ: Μαρινέλλης (2018) [57]

Παρόμοια με το δίκτυο Fast R-CNN, η εικόνα παρέχεται ως είσοδος σε ένα συνελκτικό δίκτυο το οποίο παρέχει έναν χάρτη συνελκτικών χαρακτηριστικών. Αντί να χρησιμοποιεί επιλεκτικό αλγόριθμο αναζήτησης στο χάρτη χαρακτηριστικών για τον προσδιορισμό των προτάσεων περιοχής, χρησιμοποιείται εσωτερικά ένα ξεχωριστό δίκτυο για την πρόβλεψη των προτάσεων περιοχής. Το δίκτυο εξαγωγής περιοχών ενδιαφέροντος (Regional Proposal Network - RPN) είναι πιο αποτελεσματικό και χρειάζεται μόλις 10 χιλιοστά του δευτερολέπτου για να δημιουργήσει τα ROIs. Από το χάρτη χαρακτηριστικών της αρχικής εικόνας και τις

προτεινόμενες περιοχές ενδιαφέροντος σχηματίζονται οι χάρτες χαρακτηριστικών των περιοχών ενδιαφέροντος. Οι χάρτες αυτοί, αφού πρώτα υποβληθούν σε ανασχηματισμό ώστε να αποκτήσουν ίσες διαστάσεις, εισάγονται στα πλήρως συνδεδεμένα επίπεδα του δικτύου για την ταξινόμηση του αντικειμένου και τον προσδιορισμό του πλαισίου οριοθέτησης.

2.1.3.1 Ανίχνευση και Αναγνώριση με R-CNN

Οι Xiong et al. (2016) [119] εκπαίδευσαν ένα μοντέλο ανίχνευσης πινακίδων κυκλοφορίας που βασίζεται σε βαθιά δίκτυα CNN χρησιμοποιώντας το Region Proposal Network (RPN) στο δίκτυο Fast R-CNN. Λειτουργώντας σε περιβάλλον GPU NVIDIA GTX980Ti 6GB, ο μέσος χρόνος ανίχνευσης είναι περίπου 51.5 msec ανά εικόνα με αποτέλεσμα ανίχνευσης πάνω από 99% σε συνεχή ακολουθία εικόνων. Η βάση δεδομένων που χρησιμοποίησαν αφορά κινεζικές πινακίδες κυκλοφορίας με 7 κύριες κατηγορίες. Παρ' όλο που δεν δόθηκαν πληροφορίες σχετικά με την ακρίβεια και την ανάκληση, φαίνεται ότι αυτή η μέθοδος είναι κατάλληλη για ανίχνευση σε πραγματικό χρόνο. Ένα μοντέλο Faster R-CNN με χρήση της επιλεκτικής αναζήτησης σε δύο μέρη για τον εντοπισμό των υποψήφιων περιοχών προτάθηκε από τους Zuo et al. (2017) [120]. Στη συνέχεια γίνεται χρήση ενός δικτύου CNN για την εξαγωγή χαρακτηριστικών, την ταξινόμηση και την τροποποίηση παραμέτρων. Η απόδοση του δικτύου έφτασε έως 0.3449 στην τιμή mAP (mean Average Precision), εξαιτίας ορισμένων άλυτων προβλημάτων, όπως η αγνόηση των πολύ μικρών πινακίδων ή λόγω επικαλύψεων. Η έννοια και το πως δημιουργείται η τιμή mAP θα αναλυθεί σε επόμενη ενότητα.

Το άρθρο των Rehlaender et al. (2019) [121] παρουσιάζει έναν αλγόριθμο ανίχνευσης σημάτων οδικής κυκλοφορίας που βασίζεται σε μοντέλο R-CNN. Σε αυτήν την προσέγγιση ένας αλγόριθμος τμηματοποίησης προτείνει τις περιοχές που ενδέχεται να περιέχουν πινακίδες, ο αριθμός των οποίων μειώνεται μέσω μιας γραμμικής λογικής κατωφλίου (Linear Threshold Logic - LTL). Οι περισσότερες περιοχές που προτείνονται από τον αλγόριθμο τμηματοποίησης, αφενός, δεν ταιριάζουν με την αναλογία των πλευρών των πινακίδων κυκλοφορίας και, αφετέρου, δεν μπορούν να είναι πινακίδες κυκλοφορίας επειδή υπάρχει αναντιστοιχία μεταξύ μεγέθους και θέσης σχετικά με την προοπτική του αναλυόμενου πλαισίου. Για το λόγο αυτό οι συγγραφείς βρίσκουν το σημείο φυγής στην εικόνα. Το μέγεθος όλων των αντικειμένων γίνεται γραμμικά μικρότερο όσο πιο κοντά φτάνουν στο σημείο φυγής. Επιπλέον, τα αντικείμενα στην αριστερή και τη δεξιά πλευρά του δρόμου έχουν διαφορετικά μεγέθη.

Η εικόνα μπορεί να χωριστεί σε τεταρτημόρια. Για κάθε τεταρτημόριο, το ελάχιστο και το μέγιστο όριο πρέπει να καθορίζεται ξεχωριστά. Μια προτεινόμενη περιοχή αξιολογείται από το διακριτικό σημείο (x_2, y_2) που είναι στο πρώτο τεταρτημόριο, το σημείο (x_1, y_2) στο δεύτερο τεταρτημόριο, το (x_1, y_1) στο τρίτο και στο (x_2, y_1) τέταρτο. Αυτό σημαίνει ότι μια περιοχή μπορεί να βρίσκεται και στα τέσσερα τεταρτημόρια ταυτόχρονα. Εάν τουλάχιστον μία λογική αποδέχεται την περιοχή, η περιοχή διατηρείται έγκυρη και διαβιβάζεται στο δίκτυο CNN. Έτσι, οι περιοχές μπορούν να αξιολογηθούν με βάση τη θέση τους. Λόγω των περιορισμών μεγέθους, δεν μπορούν όλες οι περιοχές να είναι πινακίδες κυκλοφορίας και μπορούν να διαγραφούν. Το κατώτατο και το ανώτατο όριο του κατωφλίου καθορίζονται από μια γραμμική συνάρτηση σε σχέση με τα x και y . Η αρχή του συστήματος συντεταγμένων μετακινείται στο κέντρο. Ως εκ τούτου, ορίζεται η παρακάτω συνάρτηση κατωφλίου για κάθε τεταρτημόριο q :

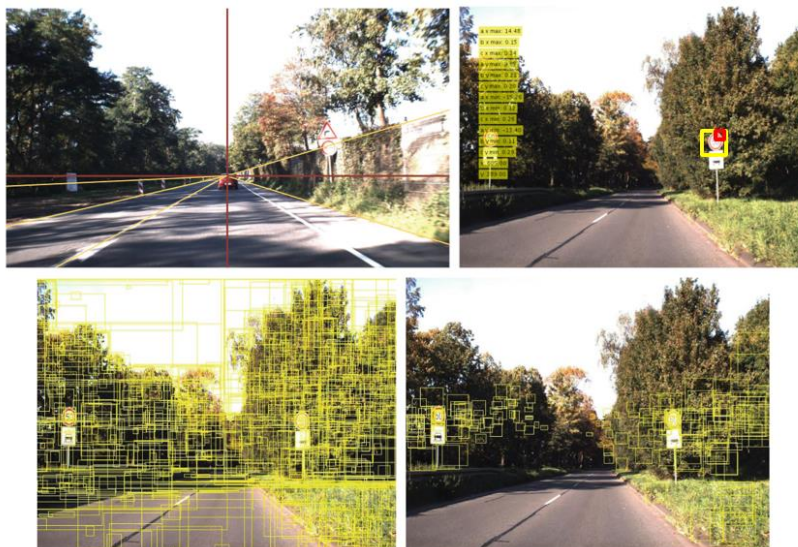
$$\Delta x_{min,x}^q = a_{min,x}^q + b_{min,x}^q \cdot |x| + c_{min,x}^q \cdot |y|$$

$$\Delta x_{min,y}^q = a_{min,y}^q + b_{min,y}^q \cdot |x| + c_{min,y}^q \cdot |y|$$

$$\Delta x_{max,x}^q = a_{max,x}^q + b_{max,x}^q \cdot |x| + c_{max,x}^q \cdot |y|$$

$$\Delta x_{max,y}^q = a_{max,y}^q + b_{max,y}^q \cdot |x| + c_{max,y}^q \cdot |y|$$

Η διαδικασία και το αποτέλεσμα μείωσης των περιοχών μιας εικόνας απεικονίζεται στην παρακάτω εικόνα. Ο αρχικός αριθμός των 6402 προτεινόμενων περιοχών μειώθηκε σε τελικό αριθμό 386. Αυτό αντιστοιχεί σε μείωση 94%.



Η διαδικασία και το αποτέλεσμα μείωσης των περιοχών μιας εικόνας
ΠΗΓΗ: Rehlaender et al. (2019) [121]

Οι περιοχές, έπειτα, αναλύονται από ένα δίκτυο CNN το οποίο στη συνέχεια τροφοδοτεί με τα διανύσματα των 4096-χαρακτηριστικών μια μηχανή διανυσματικής υποστήριξης (SVM) δύο επιπέδων, για ταχύτερη εκπαίδευση. Το πρώτο επίπεδο χρησιμοποιείται για την κατηγοριοποίηση της οικογένειας της εικόνας, ενώ το δεύτερο για την ταξινόμηση της πινακίδας κυκλοφορίας. Η χρήση του διπλού ταξινομητή SVM είχε επίσης ως αποτέλεσμα μια ελαφρά βελτίωση του ποσοστού ταξινόμησης. Αφού, λόγω του γεγονότος ότι το ποσοστό εσφαλμένης ταξινόμησης σε διαφορετικές κατηγορίες πινακίδων ήταν υψηλότερο από το αναμενόμενο, προστέθηκε το δεύτερο επίπεδο SVM για κλάσεις με ομοιότητες όπως των πινακίδων ταχύτητας, των απαγορευτικών πινακίδων, των προειδοποιητικών πινακίδων, των λευκών πινακίδων και των υποχρεωτικών πινακίδων. Για τη βελτίωση της εκπαίδευσης, τα χαρακτηριστικά μειώθηκαν και τροποποιήθηκαν με τον αλγόριθμο LASSO, ο οποίος μπόρεσε να μειώσει τον αριθμό των χαρακτηριστικών κατά περισσότερο από 50%. Η συνολική ακρίβεια ταξινόμησης αυτής της τεχνικής ανέρχεται στο 86%.

Μία σημαντική βελτίωση του δικτύου Fast R-CNN υλοποιείται από τους Shao et al. (2019) [122]. Αρχικά, προτείνεται ένας αλγόριθμος πρότασης περιοχής, που βασίζεται σε απλοποιημένα κυματίδια Gabor (SGWs) και τον τελεστή εξαγωγής περιοχών MSER (maximally stable extremal regions). Με αυτόν τον τρόπο, λαμβάνονται a priori πληροφορίες για την περιοχή ενδιαφέροντος, οι οποίες θα χρησιμοποιηθούν για τη βελτίωση του δικτύου Faster R-CNN. Η μέθοδος αυτή ονομάζεται από τους συγγραφείς ως δίκτυο πιθανών προτάσεων περιοχών (HP-RPN). Δεύτερον, για να λυθεί το πρόβλημα ότι ο αλγόριθμος Faster R-CNN δεν μπορεί να ανιχνεύσει αποτελεσματικά μικρές πινακίδες, προτείνεται μια μέθοδος που συνδυάζει τα χαρακτηριστικά του τρίτου, τέταρτου και πέμπτου επιπέδου του VGG16 για τον εμπλουτισμό των χαρακτηριστικών των μικρών πινακίδων. Τρίτον, προτείνεται η μέθοδος της δευτερεύουσας περιοχής ενδιαφέροντος (secondary region of interest method - SROI) σε σχήμα σταυρού, για την ενίσχυση των χαρακτηριστικών προς ανίχνευση των αντικειμένων και τη βελτίωση της ικανότητας ταξινόμησης του Faster R-CNN. Τα πειραματικά αποτελέσματα δείχνουν ότι η μέθοδος βελτιώνει την απόδοση ανίχνευσης, ειδικά για μικρούς στόχους, ενώ επίσης δείχνουν ότι η ακρίβεια ανίχνευσης των απαγορευτικών, υποχρεωτικών και σημάτων κινδύνου ήταν 99.53%, 98.40% και 98.44%, αντίστοιχα. Η συνολική ακρίβεια ανίχνευσης των τριών τύπων πινακίδων κυκλοφορίας είναι 99.01%. Η ταχύτητα επεξεργασίας ολόκληρου του αλγορίθμου είναι κοντά στα 9.3 καρέ ανά δευτερόλεπτο σε έναν κανονικό φορητό υπολογιστή. Τα παραπάνω ικανοποιούν τις ανάγκες εφαρμογών πραγματικού χρόνου.



Πινακίδες κυκλοφορίας με δευτερεύουσες περιοχές ενδιαφέροντος (SROI) σε σχήμα σταυρού.
ΠΗΓΗ: Shao et al. (2019) [122]

Μια άλλη αξιοσημείωτη προσπάθεια, όχι τόσο για τα αποτελέσματα όσο για τον σκοπό που αντιπροσωπεύει, είναι αυτή των Nassar et al. (2019) [123]. Η μέθοδος τους συνίσταται στην επεξεργασία εικόνων του Google Street View από γαλλικές πόλεις, προκειμένου να τοποθετηθούν γεωγραφικά οι πινακίδες στάθμευσης ανθρώπων με αναπηρικά προβλήματα, όπου το σημείο στάθμευσης δεν είναι διαθέσιμο σε συστήματα GIS. Για να το πετύχουν αυτό, βασίζονται στη μέθοδο ανίχνευσης και αναγνώρισης αντικειμένων Faster R-CNN με δίκτυο RPN (Region Proposal Networ). Η προσέγγιση αυτή βοηθά στη χαρτογράφηση της ακριβούς τοποθεσίας όπου υπάρχουν χώροι στάθμευσης. Οι εν λόγω τοποθεσίες μπορούν να χρησιμοποιηθούν για την κατασκευή υπηρεσιών ή την ενημέρωση υπηρεσιών χαρτογράφησης στο διαδίκτυο, όπως το Open Street Map. Διαθέσιμο σύνολο δεδομένων για πινακίδες στάθμευσης αναπηρικών αμαξιδίων δεν υπάρχει. Η πειραματική τους εκστρατεία πραγματοποιήθηκε στη Γαλλία, αλλά η προτεινόμενη μέθοδος μπορεί να εφαρμοστεί άμεσα σε άλλες χώρες. Όμως τους παρείχαν κάποιοι φορείς τις θέσεις GPS των χώρων στάθμευσης στο Παρίσι. Οι συντεταγμένες GPS συνδυάστηκαν, και συγκεντρώθηκαν οι σχετικές πανοραμικές εικόνες από το Google Street View. Απέκτησαν ένα ικανό σύνολο πανοραμικών εικόνων με σήματα από διάφορες οπτικές γωνίες. Το συνολικό σύνολο των δεδομένων χωρίστηκε σε ένα υποσύνολο δεδομένων εκπαίδευσης της τάξεως του 70%, 15% δεδομένα ελέγχου και 15% δεδομένα επικύρωσης. Η συνολική ακρίβεια ανίχνευσης της μεθόδου ήταν 85%. Η ακρίβεια ανίχνευσης της επιδαπέδιας σήμανσης ήταν 83.2% και για τα κατακόρυφα σήματα 86.8%. Είναι επίσης σημαντικό να σημειωθεί ότι οι ακρίβειες είναι πολύ κοντά μεταξύ τους, δεδομένων των πολλών διαφορετικών μεγεθών που έχουν τα σήματα. Η προτεινόμενη μέθοδος είναι έτοιμη για χρήση σε διάφορες εφαρμογές για την υποστήριξη πρωτοβουλιών Open Data, για βοήθεια σε πολεοδόμους και για συμμετοχή σε διαδικτυακές υπηρεσίες χαρτογράφησης.



Πινακίδες στάθμευσης αναπηρικών αμαξιδίων που βρέθηκαν στο σύνολο δεδομένων της Γαλλίας.
ΠΗΓΗ: Nassar et al. (2019) [123].

2.1.3 Αλγόριθμος YOLO

Τα μοντέλα στην οικογένεια των R-CNN βασίζονται όλα στις προτεινόμενες περιοχές, από τις οποίες στη συνέχεια εξαγονται χαρακτηριστικά σημεία. Επομένως η ανίχνευση πραγματοποιείται σε δύο στάδια. Πρώτον, το μοντέλο προτείνει ένα σύνολο περιοχών ενδιαφέροντος με επιλεγμένη αναζήτηση (Selective Search) ή δίκτυο περιφερειακών προτάσεων (Regional Proposal Network - RPN). Οι προτεινόμενες περιοχές είναι «αραιές» και τα χαρακτηριστικά σημεία μπορεί να είναι άπειρα. Τότε, στο δεύτερο στάδιο ένας ταξινομητής επεξεργάζεται μόνο τις υποψήφιες περιοχές.

Η άλλη διαφορετική προσέγγιση παραλείπει το στάδιο πρότασης της περιοχής και εκτελεί τον εντοπισμό και την ταξινόμηση απευθείας σε ένα ενιαίο συνελκτικό νευρωνικό δίκτυο

βασισμένο στην παλινδρόμηση. Οι αλγόριθμοι αυτοί ονομάζονται αλγόριθμοι ανίχνευσης αντικειμένων ενός σταδίου. Οι μέθοδοι δύο σταδίων έχουν καλύτερη απόδοση όσον αφορά την ακρίβεια εντοπισμού και ταξινόμησης, ωστόσο με αυξημένο υπολογιστικό κόστος και αρκετές καθυστερήσεις. Οι μέθοδοι ενός σταδίου είναι χρήσιμες για την παραγωγή ταχύτερων αποτελεσμάτων με μειωμένη όμως ακρίβεια.

Μία απλή προσέγγιση ανίχνευσης ενός σταδίου με πολύ καλά αποτελέσματα, αλλά και υψηλή ταχύτητα, αποτελεί ο αλγόριθμος βαθιάς μάθησης YOLO (You Only Look Once). Ο YOLO δημοσιεύτηκε από τους Redmon et al. (2016) [124], ενώ υπήρξαν και βελτιώσεις του όπως ο YOLOv2 (επίσης γνωστός ως YOLO9000) που δημοσιεύτηκε αργότερα το 2016 [127] και ο YOLOv3 που δημοσιεύτηκε το 2018 [126], καθώς και ο YOLOv4 που δημοσιεύτηκε τον Απρίλιο του 2020 από μια διαφορετική ομάδα επιστημόνων [128], επιτρέποντας για πρώτη φορά την ανίχνευση αντικειμένων σε πραγματικό χρόνο. Ο αλγόριθμος ενσωματώνει τα καθήκοντα εντοπισμού και ταξινόμησης σε ένα ενιαίο συνελκτικό νευρωνικό δίκτυο, το οποίο προβλέπει ταυτόχρονα πολλαπλά πλαίσια οριοθέτησης και πιθανότητες κλάσης στην οποία μπορεί να ανήκει ένα αντικείμενο που εγγράφεται σε ένα πλαίσιο, απευθείας από πλήρεις εικόνες σε μία αξιολόγηση. Ο αλγόριθμος, «βλέπει» μόνο μία φορά μια εικόνα για να προβλέψει ποια αντικείμενα υπάρχουν και πού βρίσκονται.

Δεδομένου ότι ολόκληρη η διαδικασία της ανίχνευσης και αναγνώρισης είναι ένα μόνο δίκτυο, μπορεί να βελτιστοποιηθεί απευθείας end-to-end στην απόδοση ανίχνευσης. Η ενοποιημένη αρχιτεκτονική του YOLO είναι εξαιρετικά γρήγορη. Το βασικό μοντέλο YOLO επεξεργάζεται εικόνες σε πραγματικό χρόνο με 45 καρέ ανά δευτερόλεπτο. Σε σύγκριση με τα σύγχρονα συστήματα ανίχνευσης, ο αλγόριθμος YOLO κάνει περισσότερα σφάλματα εντοπισμού, αλλά είναι λιγότερο πιθανό να προβλέψει ψευδώς θετικά στο παρασκήνιο, λόγω του ότι στην εκπαίδευση βλέπει ολόκληρη την εικόνα με αποτέλεσμα να κωδικοποιεί πληροφορίες συναφείς με τις κλάσεις και την εμφάνισή τους. Επιπλέον, το μοντέλο μαθαίνει πολύ γενικές αναπαραστάσεις αντικειμένων και, δεδομένου αυτού, είναι λιγότερο πιθανό να «καταρρεύσει» όταν εφαρμόζεται σε νέα πεδία ή μη αναμενόμενα δεδομένα εισαγωγής. Παρ' όλο που η ακρίβεια του μοντέλου YOLO είναι μικρότερη σε σύγκριση με άλλα μοντέλα, η ικανότητά του να ανιχνεύει και να ταξινομεί αντικείμενα σε πραγματικό χρόνο υπερτερεί αυτού του μειονεκτήματος.



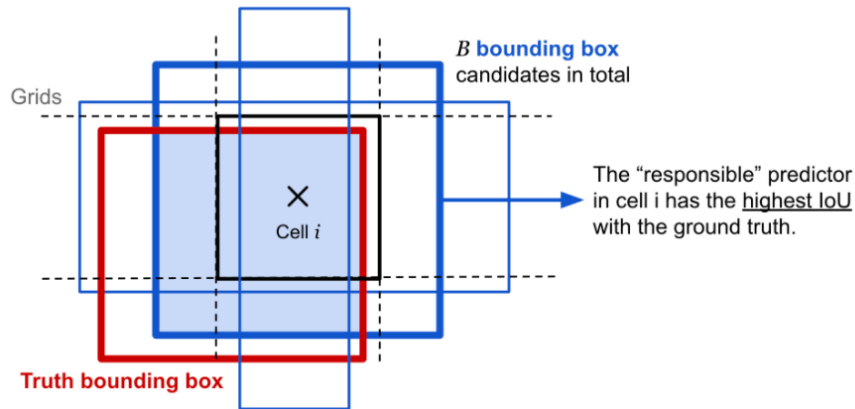
Παραδείγματα ανίχνευσης και αναγνώρισης με YOLO

ΠΗΓΗ: 1. <https://www.kaggle.com/valentynsichkar/traffic-signs-detection-by-yolo-v3-opencv-keras>,
2. <https://www.youtube.com/watch?v=ynGpj458PVc>

Το δίκτυο χρησιμοποιεί χαρακτηριστικά σημεία από ολόκληρη την εικόνα για να προβλέψει κάθε πλαίσιο οριοθέτησης. Προβλέπει, επίσης, όλα τα πλαίσια οριοθέτησης για μια εικόνα ταυτόχρονα. Η εικόνα εισόδου διαιρείται σε πλέγμα κελιών $S \times S$. Εάν το κέντρο ενός αντικειμένου πέσει σε ένα κελί, αυτό το κελί είναι «υπεύθυνο» για την ανίχνευση αυτού του αντικειμένου. Κάθε κελί προβλέπει πλαίσια οριοθέτησης B και βαθμολογίες εμπιστοσύνης για αυτά τα πλαίσια. Αυτές οι βαθμολογίες εμπιστοσύνης αντικατοπτρίζουν το βαθμό εμπιστοσύνης του μοντέλου για το ότι το πλαίσιο οριοθέτησης περιέχει ένα αντικείμενο ενδιαφέροντος και, επίσης, με ποια ακρίβεια εκτιμά ότι είναι όντως αυτό που προβλέπει. Τελικά ορίζεται η εμπιστοσύνη (confidence) ως:

$$P_r(Object) * IOU_{prediction}^{truth}$$

Εάν δεν υπάρχει αντικείμενο σε αυτό το κελί, οι βαθμολογίες εμπιστοσύνης θα πρέπει να είναι μηδέν. Διαφορετικά, το σκορ εμπιστοσύνης πρέπει να ισούται με την τομή προς ένωση (intersection over union - IOU) μεταξύ του πλαισίου οριοθέτησης που έχει προβλεφθεί και της πραγματικότητας.



Πλαίσια οριοθέτησης

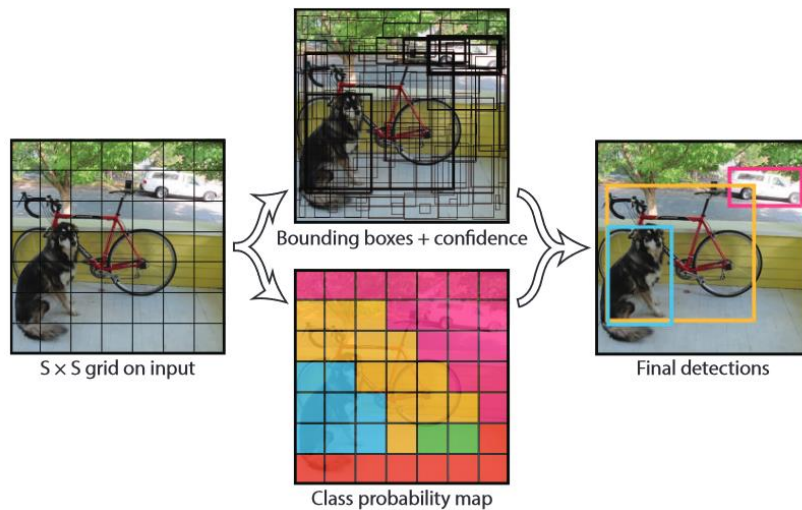
ΠΗΓΗ: <https://lilianweng.github.io/lil-log/2018/12/27/object-detection-part-4.html>

Κάθε πλαίσιο οριοθέτησης αποτελείται από 5 προβλέψεις x, y, w, h και την εμπιστοσύνη. Οι συντεταγμένες (x, y) αναπαριστούν το κέντρο του πλαισίου σε σχέση με τα όρια του κελιού πλέγματος. Το πλάτος w και το ύψος h προβλέπονται σε σχέση με ολόκληρη την εικόνα. Τέλος, η πρόβλεψη εμπιστοσύνης αντιπροσωπεύει το IOU του πλαισίου οριοθέτησης που έχει προβλεφθεί ($b. b_{pred}$) και του πραγματικού πλαισίου οριοθέτησης ($b. b_{truth}$):

$$IOU_{prediction}^{truth} = \frac{area(b. b_{pred} \cap b. b_{truth})}{area(b. b_{pred} \cup b. b_{truth})}$$

Κάθε κελί πλέγματος προβλέπει επίσης C εξαρτώμενες πιθανότητες κλάσης, $P_r(Class_i | Object)$, όπου C ο αριθμός των κλάσεων. Αυτές οι πιθανότητες εξαρτώνται από το κελί πλέγματος που περιέχει ένα αντικείμενο. Προβλέπεται μόνο ένα σύνολο πιθανοτήτων κατηγορίας ανά κελί πλέγματος, ανεξάρτητα από τον αριθμό των κουτιών B . Κατά τη διάρκεια των ελέγχων πολλαπλασιάζονται οι εξαρτώμενες πιθανότητες κλάσης και οι μεμονωμένες προβλέψεις εμπιστοσύνης του πλαισίου, το οποίο δίνει τους βαθμούς εμπιστοσύνης για κάθε κατηγορία. Αυτές οι βαθμολογίες αντιπροσωπεύουν τόσο την πιθανότητα εμφάνισης αυτής της κλάσης στο πλαίσιο οριοθέτησης όσο και το πόσο καλά ταιριάζει στο αντικείμενο το πλαίσιο οριοθέτησης που προβλέφθηκε:

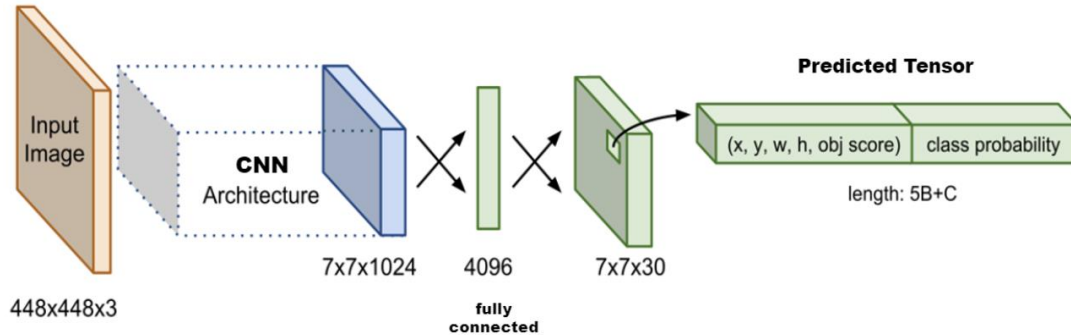
$$P_r(Class_i | Object) * P_r(Object) * IOU_{prediction}^{truth} = Pr(Class_i) * IOU_{prediction}^{truth}$$



Μοντέλο YOLO

ΠΗΓΗ: Redmon et al. (2016) [124]

Το δίκτυο αποτελείται από 24 συνελκτικά επίπεδα και δύο πλήρως συνδεδεμένα επίπεδα. Το τελικό επίπεδο προβλέπει τόσο τις πιθανότητες κλάσης όσο και τις συντεταγμένες του πλαισίου οριοθέτησης. Η έξοδος του μοντέλου (πρόβλεψη) είναι ένα διάνυσμα $S \times S \times (5B + C)$ για κάθε εικόνα. Ως συνάρτηση ενεργοποίησης χρησιμοποιείται σε όλα τα επίπεδα η συνάρτηση Leaky RELU.



Αρχιτεκτονική δικτύου YOLO

ΠΗΓΗ: <https://www.mdpi.com/2076-3417/10/2/612/htm>

Στην έξοδο του μοντέλου βελτιστοποιείται το άθροισμα τετραγώνων των αποκλίσεων. Ωστόσο, με αυτόν τον τρόπο σταθμίζεται το σφάλμα εντοπισμού εξίσου με το σφάλμα ταξινόμησης, γεγονός που μπορεί να μην είναι ιδανικό. Επιπλέον, σε κάθε εικόνα πολλά κελιά πλέγματος δεν περιέχουν κανένα αντικείμενο. Αυτό ωθεί τους βαθμούς εμπιστοσύνης αυτών των κελιών προς το μηδέν και συχνά υπερνικά την κλίση των κελιών που περιέχουν αντικείμενα. Αυτό μπορεί να οδηγήσει σε αστάθεια του μοντέλου, με αποτέλεσμα η εκπαίδευση να αποκλίνει νωρίς. Προκειμένου να διορθωθεί αυτό, χρησιμοποιούνται δύο παράμετροι, λ_{coord} και λ_{noobj} , ώστε να αυξηθεί το κόστος των προβλέψεων συντεταγμένων του πλαισίου οριοθέτησης και να μειωθεί το κόστος των προβλέψεων εμπιστοσύνης για τα κουτιά που δεν περιέχουν τα αντικείμενα ενδιαφέροντος. Η βελτιστοποίηση του αθροίσματος τετραγώνων των αποκλίσεων εξισώνει επίσης το σφάλμα των μεγάλων πλαισίων οριοθέτησης με εκείνο των μικρών. Για να αντιμετωπιστεί αυτό, εν μέρει, προβλέπεται η τετραγωνική ρίζα του πλάτους και του ύψους του πλαισίου οριοθέτησης αντί απλώς του πλάτους και του ύψους.

Βάσει των παραπάνω, ο YOLO χρησιμοποιεί μία συνάρτηση κόστους για τον καθορισμό των πλαισίων οριοθέτησης καθώς και για την ταξινόμηση των αντικειμένων. Η συνάρτηση κόστους «πειθαρχεί» μόνο το σφάλμα ταξινόμησης εάν υπάρχει αντικείμενο στο αντίστοιχο κελί πλέγματος, εξ ου και η εξαρτώμενη πιθανότητα κλάσης. Επιπλέον, «πειθαρχεί» το σφάλμα των συντεταγμένων του πλαισίου οριοθέτησης εάν αυτός ο προγνωστικός δείκτης είναι «υπεύθυνος» για το πραγματικό πλαίσιο οριοθέτησης (ground truth box), δηλαδή αν έχει το υψηλότερο IOU από οποιονδήποτε άλλο προγνωστικό παράγοντα σε αυτό το κελί πλέγματος:

$$\begin{aligned}
 Cost = & \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{obj} (x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 + \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{obj} (\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2 \\
 & + \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{obj} (C_i - \hat{C}_i)^2 + \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{noobj} (C_i - \hat{C}_i)^2 \\
 & + \sum_{i=0}^{S^2} 1_i^{obj} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2
 \end{aligned}$$

όπου το 1_{ij}^{obj} δηλώνει εάν το αντικείμενο εμφανίζεται στα κελιά i, j και το 1_i^{obj} δηλώνει ότι ο προγνωστικός παράγοντας j του πλαισίου οριοθέτησης στο κελί i είναι «υπεύθυνος» για αυτήν την πρόβλεψη. Το πρώτο μέρος της παραπάνω συνάρτησης σχετίζεται με το κόστος των προβλέψεων πλαισίου οριοθέτησης, το δεύτερο μέρος αφορά το κόστος της ταξινόμησης, ενώ το τρίτο μέρος έχει να κάνει με το κόστος πρόβλεψης εμπιστοσύνης. Το $p_i(c)$ δίνει την

εμπιστοσύνη του πραγματικού πλαισίου οριοθέτησης, το οποίο δεν είναι άλλο από τον λόγο IOU του πραγματικού πλαισίου οριοθέτησης με το πλαίσιο οριοθέτησης που έχει προβλεφθεί. Το $\hat{p}_i(c)$ δίνει την προβλεπόμενη εμπιστοσύνη του δικτύου.

Αυτή τη στιγμή, όπως προαναφέρθηκε, η οικογένεια YOLO έχει τέσσερις κύριες παραλλαγές που δημοσιεύτηκαν προκειμένου να βελτιώσουν κάθε φορά την προηγούμενη έκδοση.

1. **YOLOv1**: YOLO Unified, Real-Time Object Detection (Redmon et al., 2016 [124]). Ο πρωτοποριακός αλγόριθμος ενοποίησε σε ένα ενιαίο συνελικτικό νευρωνικό δίκτυο τα καθήκοντα εντοπισμού και ταξινόμησης, επιτρέποντας για πρώτη φορά την ανίχνευση αντικειμένων σε πραγματικό χρόνο.
2. **YOLOv2**: YOLO9000 - Better, Faster, Stronger (Redmon et al., 2016 [126]) μπορεί να ανιχνεύσει πάνω από 9.000 αντικείμενα, εξ ου και το όνομα YOLO9000. Έχει εκπαιδευτεί σε σύνολα δεδομένων ImageNet και MS COCO, και έχει επιτύχει mAP 16%, που δεν αποτελεί πολύ καλό ποσοστό.
 - ο Χρησιμοποιώντας εκπαίδευση πολλαπλών κλιμάκων, το μοντέλο έμαθε να γενικεύει και να εντοπίζει αντικείμενα με διαφορετικούς λόγους διαστάσεων ή σχηματισμών.
 - ο Η ομαλοποίηση των χαρακτηριστικών βοήθησε στη βελτίωση της μέσης ακρίβειας για μικρά αντικείμενα, αλλά εξακολουθούν να υστερούν σε σύγκριση με τα ταχύτερα μοντέλα RCNN.
 - ο Καθώς η συνολική μέση ακρίβεια mAP αυξήθηκε (από 63.4% σε 78.6%), τα σφάλματα εντοπισμού μειώθηκαν σε κάποιο ποσό. Το mAP στο 0.5 IOU εξακολουθεί να υστερεί συγκριτικά με τον αλγόριθμο Faster - RCNN
3. **YOLOv3**: An Incremental Improvement (Redmon et al., 2018 [127]). Αυτό το μοντέλο είναι σημαντικά μεγαλύτερο από τα προηγούμενα δύο και έχει επιτύχει mAP 57.9%, όπου μέχρι πρότινος ήταν το καλύτερο αποτέλεσμα από την οικογένεια αλγορίθμων YOLO.
 - ο Η μέση ακρίβεια για τα μικρά αντικείμενα βελτιώθηκε, δίνοντας καλύτερα αποτελέσματα από τον αλγόριθμο Faster RCNN.
 - ο Η μέση ακρίβεια mAP αυξήθηκε σημαντικά.
 - ο Αφού αυξήθηκε η mAP τα σφάλματα εντοπισμού μειώθηκαν.
 - ο Οι προβλέψεις σε διαφορετικές κλίμακες ή αναλογίες διαστάσεων για το ίδιο αντικείμενο βελτιώθηκαν λόγω της προσθήκης της μεθόδου τύπου πυραμίδας.
4. **YOLOv4**: Optimal Speed and Accuracy of Object Detection (Bochkovskiy et al., 2020 [128]). Ο νέος εξελιγμένος αλγόριθμος YOLOv4 επιτυγχάνει εξαιρετική ταχύτητα και ακρίβεια.
 - ο Είναι ένα αποτελεσματικό και ισχυρό μοντέλο ανίχνευσης αντικειμένων που επιτρέπει σε οποιονδήποτε με GPU 1080 Ti ή 2080 Ti να εκπαιδεύσει έναν εξαιρετικά γρήγορο και με μεγάλη ακρίβεια ανιχνευτή.
 - ο Επαληθεύεται η επίδραση των προηγμένων μεθόδων ανίχνευσης αντικειμένων "Bag-of-Freebies" και "Bag-of-Specials" κατά την εκπαίδευση του ανιχνευτή με τα αποτελέσματα που παρέχει.
 - ο Οι τροποποιημένες μέθοδοι αιχμής, όπως CBN (Cross-iteration batch normalization), PAN (Path aggregation network) κ.λπ., καθιστούν τον YOLOv4 έναν σύγχρονο ανιχνευτή ο οποίος είναι ταχύτερος και πιο ακριβής από όλους τους διαθέσιμους εναλλακτικούς ανιχνευτές

2.1.3.1 Μέση αντιπροσωπευτική ακρίβεια mAP

Η απόδοση της ανίχνευσης μετρείται με τη μέση αντιπροσωπευτική ακρίβεια mAP (mean Average Precision), η οποία είναι το πιο συχνά χρησιμοποιούμενο μέτρο αξιολόγησης για την

ανίχνευση αντικειμένων. Ουσιαστικά, πρόκειται για τη μέση ακρίβεια πολλαπλών κλάσεων. Η μέση ακρίβεια υπολογίζει κατά πόσον τα ανιχνευόμενα πλαίσια οριοθέτησης αντιστοιχούν στα αντίστοιχα πραγματικά. Η μέση ακρίβεια είναι η περιοχή κάτω από την καμπύλη ανάκτησης ακριβείας, η οποία ενσωματώνει την ακρίβεια και την ανάκληση, ενώ κυμαίνεται από 0 έως 1. Για να υπολογιστεί η mAP πρέπει πρώτα να υπολογιστεί η AP (average precision), η οποία θεωρείται η πλέον κατάλληλη για ανίχνευση αντικειμένων. Για κάθε πρόβλεψη του μοντέλου υπολογίζονται η ακρίβεια (precision) και η ανάκληση (recall), και ταξινομούνται με αύξουσα ανάκληση. Η AP υπολογίζει την τιμή της precision (p) για κάθε τιμή της ανάκλησης (r) από 0 έως 1 και εκφράζεται ως εξής:

$$AP = \int_0^1 p(r)dr$$

Για να θεωρηθεί μια πρόβλεψη σωστή και να συμβάλει στην AP θα πρέπει να επιτυγχάνει μετρική απόδοσης IoU μεγαλύτερη από 0.5, η οποία υπολογίζεται, όπως έχει αναφερθεί παραπάνω, διαφορετικά πιο απλά ως:

$$IoU = \frac{\text{Θετικά Σωστά}}{\text{Θετικά Σωστά} + \text{Θετικά Λάθος} + \text{Αρνητικά Λάθος}}$$



Οπτικοποίηση IoU

ΠΗΓΗ: https://commons.wikimedia.org/wiki/File:Intersection_over_Union_-_poor,_good_and_excellent_score.png

Στην πράξη το ολοκλήρωμα ισοδυναμεί με το έξης άθροισμα της συνάρτησης:

$$AP = \sum_k^N P(k)\Delta r(k)$$

όπου N ο αριθμός των προβλέψεων και k οι προβλέψεις μέχρι μια δεδομένη ανάκληση. Σε προβλήματα με πολλαπλές κλάσεις υπολογίζεται για κάθε κλάση η average precision και μετά η mean average precision (mAP) ως μέσος όρος των ξεχωριστών AP.

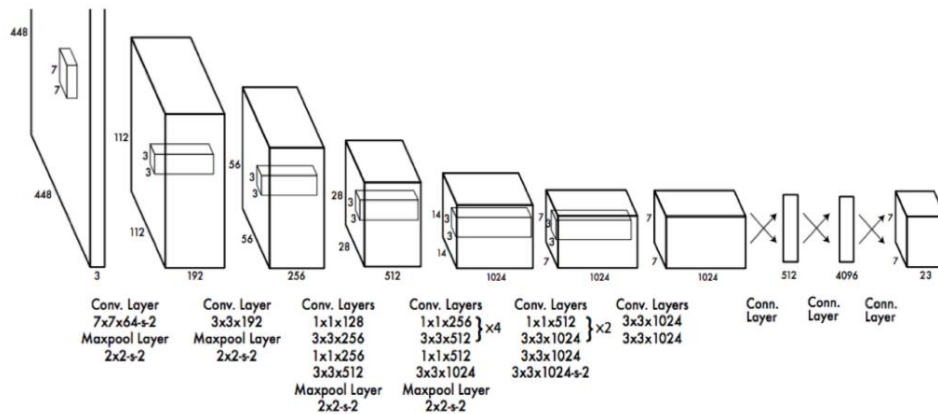
2.1.3.2 Ανίχνευση και Αναγνώριση με YOLO

Δεδομένου ότι οι αλγόριθμοι της οικογένειας YOLO επιτρέπουν την ανίχνευση και αναγνώριση αντικειμένων σε πραγματικό χρόνο, αποτελούν κατάλληλη μέθοδο για το πρόβλημα ανίχνευσης και αναγνώρισης των σημάτων οδικής κυκλοφορίας, το οποίο απαιτεί υψηλή ταχύτητα και ακρίβεια. Ένα μοντέλο βασισμένο στην αρχιτεκτονική του YOLO χρησιμοποιεί ο Vikram (2018) [129], τροποποιώντας όμως τη συνάρτηση κόστους καθώς οι πινακίδες από το βέλγικο σύνολο δεδομένων που χρησιμοποιεί είναι πολύ μικρές. Επομένως, χρειάζεται το μοντέλο να δώσει μεγαλύτερη έμφαση στην ανίχνευση και αναγνώριση μικρού μεγέθους πινακίδων. Η τροποποίηση της συνάρτησης κόστους «πειθαρχεί» το μοντέλο περισσότερο στην περίπτωση που δεν εντοπίζει με ακρίβεια ένα μικρό σήμα οδικής κυκλοφορίας σε σχέση με ένα μεγαλύτερο. Διαισθητικά, το μοντέλο μαθαίνει να εντοπίζει μικρότερα αντικείμενα στις εικόνες.

$$Cost = \left(\lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{obj} (x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 + \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{obj} (\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2 \right) \times \frac{1}{\sqrt{h_i \times w_i}}$$

Το αποτέλεσμα του μοντέλου με την μετρική mAP έδωσε τιμή 0.332 με χρήση της κανονικής συνάρτησης κόστους που χρησιμοποιούν οι Redmon et al. (2016) [124] στο YOLO και 0.525 με χρήση της τροποποιημένης συνάρτησης κόστους. Το αποτέλεσμα είναι εμφανώς καλύτερο, αυτό όμως που αξίζει επίσης να αναφερθεί είναι πως η τροποποιημένη συνάρτηση κόστους μειώνεται απότομα κατά τη διάρκεια της εκπαίδευσης, γεγονός που δείχνει ότι το μοντέλο μαθαίνει με πολύ γρήγορο ρυθμό.

Οι Wang et al. (2016) [130] επίσης έκαναν χρήση του αλγόριθμου YOLO για την ανίχνευση και αναγνώριση πινακίδων. Οι βασικές αρχιτεκτονικές τροποποιήσεις που προτείνουν αφορούν τα τελευταία επίπεδα μετά τα συνελκτικά επίπεδα. Δεδομένου ότι υπάρχουν μόνο 13 κατηγορίες πινακίδων κυκλοφορίας, το τελευταίο επίπεδο απαιτεί C = 13. Λόγω περιορισμών στη μνήμη, εισήχθη ένα συνδεδεμένο επίπεδο υπόδειγματοληψίας πριν από τα δύο συνδεδεμένα επίπεδα, μειώνοντας ουσιαστικά τον αριθμό των παραμέτρων κατά 84%.



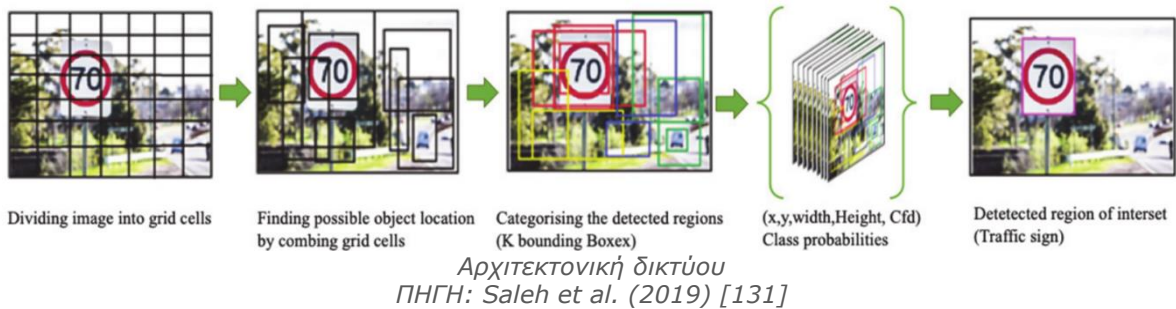
Αρχιτεκτονική δικτύου
 ΠΗΓΗ: Wang et al. (2016) [130]

Επιπλέον, αυτοί ενσωμάτωσαν τη δημιουργία χαρτών απεικόνισης των προβολών ή χάρτη σημαντικότητας (Saliency Maps) στο C. Ένας χάρτης Saliency είναι στην πραγματικότητα μια εικόνα που δείχνει τη μοναδική ποιότητα κάθε εικονοστοιχείου. Ο στόχος ενός τέτοιου χάρτη είναι να απλοποιήσει ή και να αλλάξει την αναπαράσταση μιας εικόνας σε κάτι που έχει περισσότερο νόημα και είναι πιο εύκολο να αναλυθεί. Για παράδειγμα, εάν ένα εικονοστοιχείο έχει υψηλό επίπεδο γκρι τότε αυτό το εικονοστοιχείο θα εμφανίζεται στον χάρτη εντονότερα. Πρόκειται για ένα είδος κατάτμησης εικόνας. Το τελικό αποτέλεσμα mAP που επιτυγχάνεται είναι 32%.

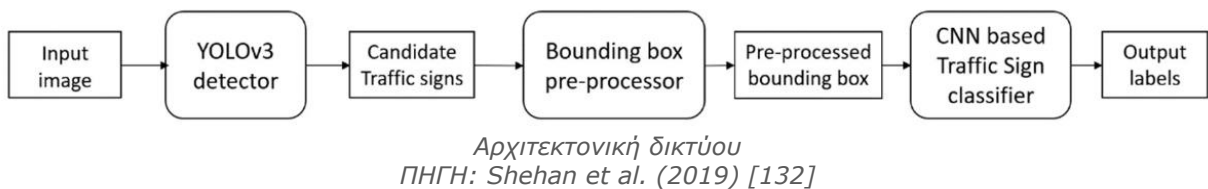
Εφαρμογή του YOLO πραγματοποιήσαν οι Saleh et al. (2019) [131] για μια αποτελεσματική προσέγγιση με τη χρήση μίας μονοσκοπικής (monocular) κάμερας για την αναγνώριση σημάτων οδικής κυκλοφορίας σε πραγματικό χρόνο και την εκτίμηση απόστασης μεταξύ σήματος οδικής κυκλοφορίας και θέσης της κάμερας. Στόχος είναι η ειδοποίηση των οδηγών χωρίς να αποσπάται η προσοχή τους από την οδήγηση. Η αρχιτεκτονική τους βασίστηκε στον αλγόριθμο YOLOv2. Συμπεριέλαβαν 19 συνελκτικά επίπεδα και 3 συνδεδεμένα επίπεδα. Το μοντέλο παρέχει περίπου μέση ακρίβεια mAP 97% στο σύνολο δεδομένων δοκιμής με υψηλή ταχύτητα περίπου 60 ms/καρέ, το οποίο πληροί τις απαιτήσεις για την εφαρμογή σε πραγματικό χρόνο

Προκειμένου να εκτιμηθεί η απόσταση μεταξύ των πινακίδων κυκλοφορίας και της θέσης της κάμερας χρειάζεται, πρώτα, εντοπισμός του αντικειμένου ενδιαφέροντος. Επομένως, αφού εντοπιστεί με επιτυχία η πινακίδα, μπορεί να εκτιμηθεί το περίγραμμα που αντιστοιχεί στο αντικείμενο, το οποίο επιστρέφει το πλαίσιο οριοθέτησης μαζί με το πλάτος, το ύψος και τις

συντεταγμένες (x, y) του πλαισίου σε εικονοστοιχεία. Τέλος, χρησιμοποιούνται τα ορθογώνια μεγέθη του πλαισίου οριοθέτησης που παρέχονται από τον ταξινομητή για να εκτιμηθεί η απόσταση της πινακίδας κυκλοφορίας.



Ο αλγόριθμος YOLOv3 χρησιμοποιεί μια εξειδικευμένη αρχιτεκτονική δικτύου βασισμένη σε πυραμίδα και έχει αρκετές διαφορές σε σύγκριση με τις προηγούμενες εκδόσεις, προκειμένου να αντιμετωπίσει προβλήματα όπως η ακρίβεια και η ανίχνευση αντικειμένων μικρού μεγέθους στην εικόνα. Την προσέγγιση αυτή ακολουθούν και οι Shehan et al. (2019) [132] για την ανίχνευση, ενώ για την αναγνώριση αξιοποιούν έναν ταξινομητή που βασίζεται σε δίκτυο CNN, χρησιμοποιώντας επίσης το γερμανικό σύνολο δεδομένων.



Η μέθοδος αποτελείται από τον ανιχνευτή YOLO εκπαιδευμένο να εντοπίζει τις υποψήφιες πινακίδες, ένα πλαίσιο οριοθέτησης το οποίο επεξεργάζεται μεγεθύνοντας κατά 25% το εντοπισμένο πλαίσιο οριοθέτησης του YOLO, ενώ περικόπτεται και μεταβάλλεται το μέγεθος των πλαισίων όπου περιέχονται σήματα σε 48x48, και στη συνέχεια ένας ταξινομητής CNN αναγνωρίζει σε ποια από τις 43 κλάσεις ανήκει η πινακίδα.



Αποτελέσματα ανίχνευσης και αναγνώρισης της μεθόδου
ΠΗΓΗ: Shehan et al. (2019) [132]

Η υψηλότερη απόδοση ακρίβειας mAP του ανιχνευτή που επιτεύχθηκε αντιστοιχεί στο 92.2% στο σετ δεδομένων ελέγχου, με περίπου 10 καρτέ ανά δευτερόλεπτο. Το αντίστοιχο ποσοστό ακρίβειας για την αναγνώριση ανέρχεται σε 96.46% σε 12630 εικόνες χωρίς επιπλέον δεδομένα, όταν προστέθηκαν όμως επιπλέον δεδομένα με την τεχνική της αύξησης δεδομένων (data augmentation) η ακρίβεια βελτιώθηκε στο 99.6%.

Η αύξηση των δεδομένων (data augmentation) αποτελεί μια τεχνική για την αύξηση της ποικιλομορφίας του σετ εκπαίδευσης, με εφαρμογή τυχαίων (αλλά ρεαλιστικών) μετασχηματισμών, όπως η στροφή της εικόνας.

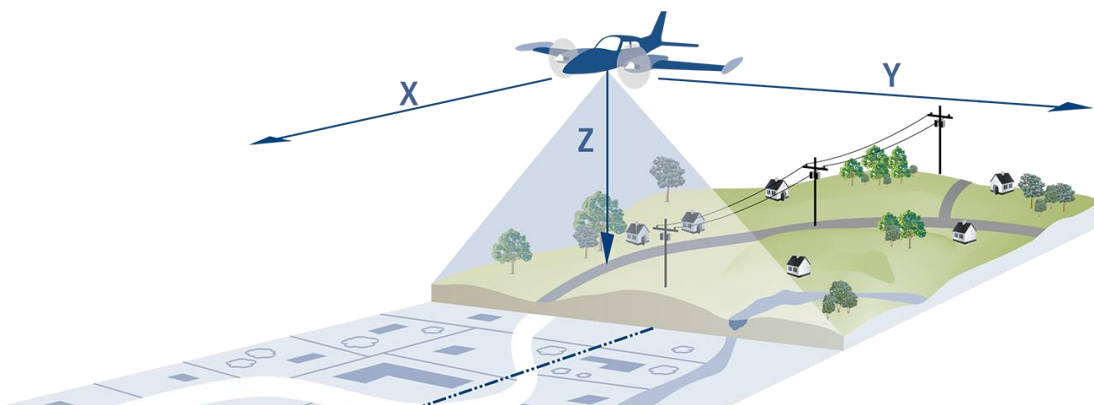
2.2 Εντοπισμός και Αναγνώριση με χρήση LiDAR

Λόγω της μεγάλης βελτίωσης της τεχνολογίας των εναέριων σαρωτών laser LiDAR (Light Detection And Ranging), αλλά και της σημαντικής ανάπτυξης των φορητών συσκευών laser LiDAR, εμφανίστηκαν τα τελευταία χρόνια αρκετές μέθοδοι ανίχνευσης και αναγνώρισης σημάτων οδικής κυκλοφορίας που βασίζονται στην τεχνολογία LiDAR. Σημειώνεται, πως η τεχνολογία LiDAR αποτελεί μια μέθοδο ενεργητικής τηλεμέτρησης, η οποία χρησιμοποιεί τεχνικές τρισδιάστατης σάρωσης με δέσμη laser για τη συλλογή 3D δεδομένων υψηλής πυκνότητας και ακρίβειας. Σήμερα, χρησιμοποιείται ευρέως για την ανάκτηση της τρισδιάστατης δομής της επιφάνειας του εδάφους και της βλάστησης υπό οποιοσδήποτε συνθήκες φωτισμού, παρουσιάζοντας σημαντική μείωση του κόστους και του χρόνου έναντι των παραδοσιακών μεθόδων με μετρήσεις πεδίου.

Τα συστήματα LiDAR ακτινοβολούν προς το στόχο μια ανεξάρτητα δημιουργούμενη ενεργειακή δέσμη ακτινοβολίας, ενώ καταγράφουν παράλληλα την ανακλώμενη από το στόχο ποσότητά της μαζί με τις μεταβολές που υφίσταται το αρχικό εκπεμπόμενο σήμα από την ανάκλαση. Οι διαφορές των χρόνων επιστροφής των παλμών καθώς και των μηκών κύματος χρησιμοποιούνται για να παραχθεί η τρισδιάστατη αναπαράσταση του στόχου και για τον υπολογισμό της ακριβούς απόστασης από το αντικείμενο. Γνωρίζοντας την ακριβή θέση και προσανατολισμό του αεροσκάφους, και γενικότερα της φέρουσας πλατφόρμας, την ίδια στιγμή είναι δυνατό να καθοριστούν οι χωρικές συντεταγμένες κάθε επιστρεφόμενου παλμού, παράγοντας ένα συνολικό νέφος σημείων με γεωαναφορά. Τα παραγόμενα με αυτό τον τρόπο δεδομένα μπορούν να προβληθούν ως ένα τρισδιάστατο (3D) μοντέλο, παρέχοντας λεπτομερή πληροφορία για την περιοχή ενδιαφέροντος ακόμα και σε περιπτώσεις πυκνής βλάστησης.

Η χρήση της τεχνολογίας LiDAR προσφέρει υψηλής ποιότητας λύσεις σε ένα ευρύ πεδίο εφαρμογών, όπως ενδεικτικά:

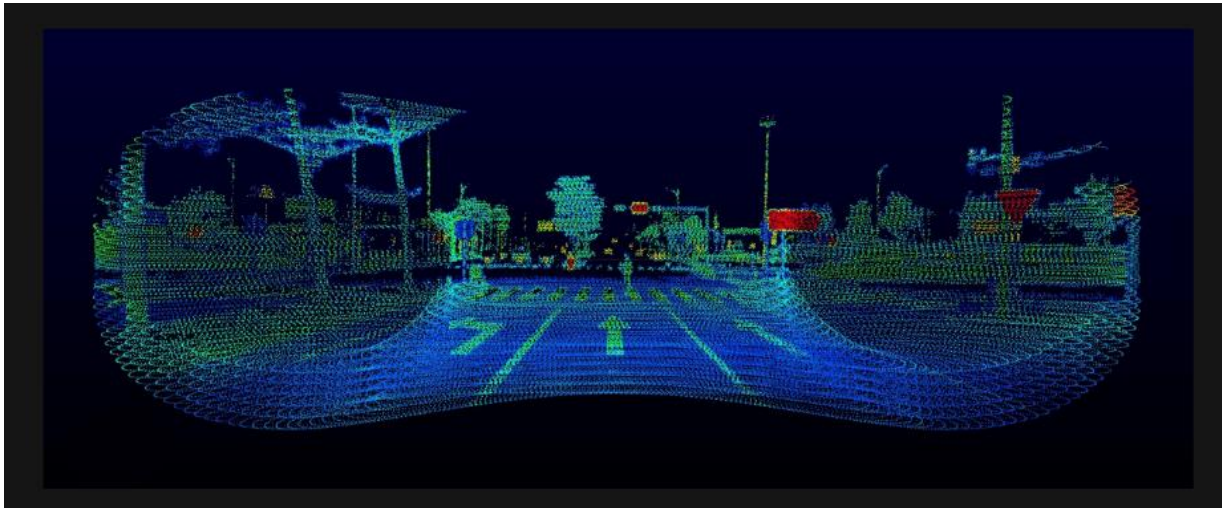
- Παραγωγή ψηφιακών μοντέλων εδάφους / επιφανείας και ισοϋψών γραμμών.
- Απομόνωση και μελέτη εδάφους, με δυνατότητα εφαρμογής στη μελέτη και το σχεδιασμό υδρογραφικών δικτύων, οδικών ή σιδηροδρομικών έργων, ογκομέτρηση στοιχείων εδάφους (εκσκαφών, επιχωμάτων), χωροταξικό σχεδιασμό κ.ά.
- Χαρτογράφηση υδρογραφικού δικτύου και λεκανών απορροής, έλεγχος διάβρωσης εδαφών και αποτύπωση μεταβολών.
- Χαρτογράφηση ακτογραμμών.
- Τρισδιάστατη απεικόνιση κτιρίων και εξαγωγή τους σε ψηφιακή διανυσματική μορφή. Δυνατότητα προσομοίωσης πλημμυρών, θορύβου, ρύπανσης και εφαρμογές τηλεπικοινωνιών.
- Αποτύπωση γραμμών μεταφοράς ενέργειας, όπως καλώδια υψηλής τάσης, σωλήνες και αγωγοί.



Τεχνολογία αερομεταφερόμενου LiDAR

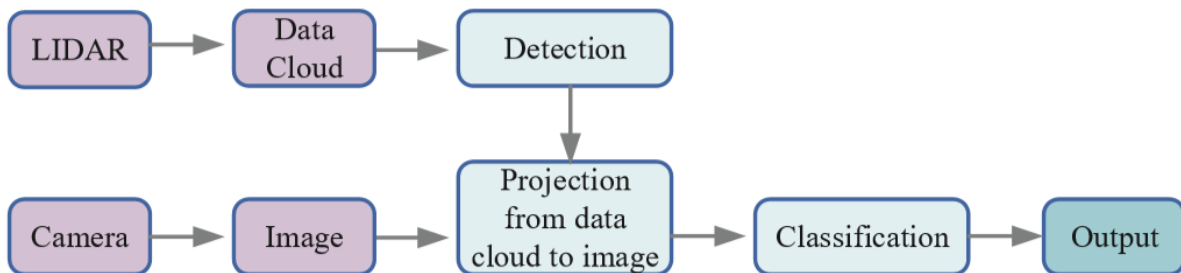
ΠΗΓΗ: <https://www.anditi.com/not-all-lidar-is-equal-part-2-the-role-of-the-imu/>

Όλες οι τεχνικές που έχουν περιγραφεί έως τώρα έχουν ως δεδομένα εισόδου εικόνες RGB (2D) από φωτογραφική μηχανή. Η τεχνολογία LiDAR παρέχει ένα 3D νέφος σημείων. Στην προκειμένη περίπτωση του προβλήματος της ανίχνευσης και αναγνώρισης των πινακίδων οδικής κυκλοφορίας χρησιμοποιείται συνήθως ο συνδυασμός των εικόνων και του νέφους σημείων, ενώ υπάρχουν και έρευνες που χρησιμοποιούν μόνο την τεχνολογία LiDAR. Οι μέθοδοι που βασίζονται και σε LiDAR έχουν μεγάλο δυναμικό για τη διαχείριση συνθηκών, όπως οι ακραίες καιρικές συνθήκες, η ισχυρή ομίχλη, η δυνατή βροχή και η ισχυρή χιονόπτωση, ή όπως βασικά προβλήματα της ανίχνευσης των πινακίδων τη νύχτα, αντανάκλαση προβολών και ελλιπής φωτισμός. Ωστόσο, μικρό μέρος μόνο ερευνητών έχουν το δικό τους ενσωματωμένο/φορητό LiDAR για τη συλλογή δεδομένων και την έρευνα.

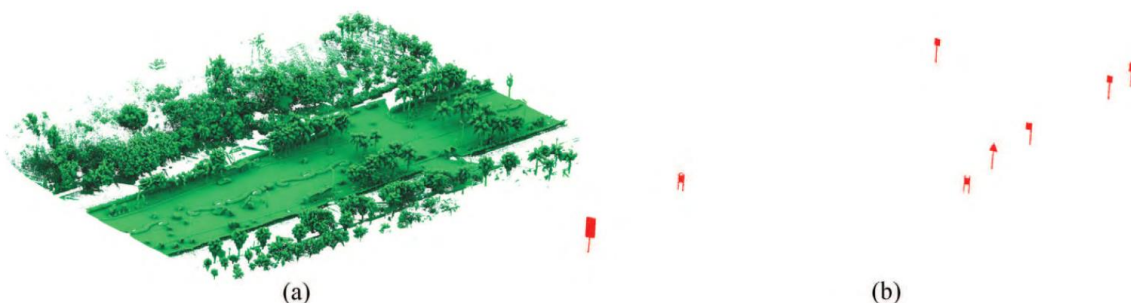


Τρισδιάστατο νέφος σημείων από φορητό αισθητήρα LiDAR ενσωματωμένο σε όχημα
ΠΗΓΗ: <https://www.spar3d.com/news/lidar/djis-livox-introduces-500-meter-lidar-with-unique-scanning-pattern/>

Οι Liu & Wang (2019) [141], στην κατηγοριοποίηση των μεθόδων ανίχνευσης και αναγνώρισης των πινακίδων οδικής κυκλοφορίας που πραγματοποιούν για την βιβλιογραφική ανασκόπηση στην εργασία τους, συγκαταλέγουν και τις μεθόδους βασισμένες σε LiDAR. Η διαδικασία της ανίχνευσης και αναγνώρισης σημάτων με συσχέτιση δεδομένων LiDAR και κάμερας ακολουθεί την διαδρομή που φαίνεται στην παρακάτω εικόνα. Παράγεται από το LiDAR το τρισδιάστατο νέφος σημείων, το οποίο υποβάλλεται σε επεξεργασία για την ανίχνευση πινακίδων. Στις μεθόδους ανίχνευσης και αναγνώρισης σημάτων οδικής κυκλοφορίας χρησιμοποιούνται συνήθως διαφορετικές διαδικασίες επεξεργασίας του νέφους σημείων (κατωφλίωση, φιλτράρισμα, ομαδοποίηση) . Στη συνέχεια, τα εντοπισμένα στο νέφος σημείων σήματα σχετίζονται με τις αντίστοιχες θέσεις τους σε εικόνες RGB. Τέλος, εκτελείται μια διαδικασία ταξινόμησης της κατηγορίας των πινακίδων κυκλοφορίας από το σύνολο εικόνων RGB. Αυτή τη διαδικασία ακολούθησαν και οι Yu et al. (2016) [142].



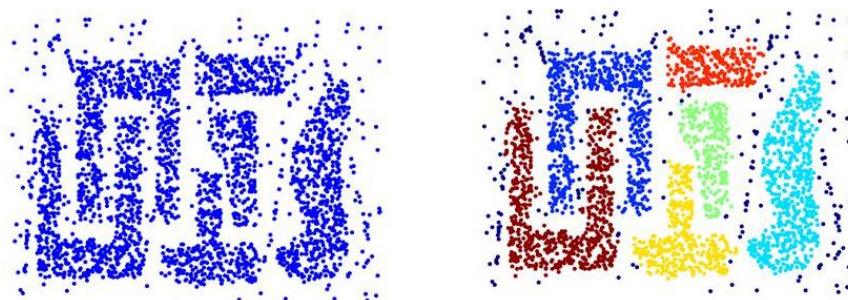
Διαδικασία ανίχνευσης και αναγνώρισης σημάτων με συσχέτιση δεδομένων LiDAR και κάμερας
ΠΗΓΗ: Liu & Wang (2019) [141]



Απεικόνιση αποτελεσμάτων ανίχνευσης πινακίδων από τρισδιάστατο νέφος σημείων: a) νέφος σημείων πριν από επεξεργασία, b) πινακίδες που ανιχνεύθηκαν
ΠΗΓΗ: Yu et al. (2016) [142]

Οι Pei et al. (2013) [143] παρουσιάζουν μια μέθοδο ανίχνευσης με βάση τη σάρωση με laser για πινακίδες κυκλοφορίας, δέντρα, τοίχους κτιρίων και φράγματα. Σε αυτή τη μελέτη, οι πινακίδες αναγνωρίστηκαν έως και κατά 86%, χωρίς τη χρήση επιπλέον εικόνων για την αναγνώριση. Αντίστοιχα, μόνο με τη χρήση τεχνολογίας LiDAR και λαμβάνοντας υπόψη τις ραδιομετρικές και γεωμετρικές πληροφορίες που δημιουργούνται με σάρωση με laser, οι Riveiro et al. (2017) [144] ασχολήθηκαν με τις αντανάκλαστικές πινακίδες οδικής κυκλοφορίας. Πρώτα δημιούργησαν έναν χάρτη έντασης του νέφους σημείων και έπειτα επιλέχθηκαν με κατωφλίωση τα εικονοστοιχεία των έντονα αντανάκλαστικών επιφανειών. Στη συνέχεια, και αφού το νέφος σημείων φιλτράρεται, εφαρμόζεται ο αλγόριθμος DBSCAN (Density-Based Spatial Clustering of Applications with Noise) για την ομαδοποίηση και την αναγνώριση χαρακτηριστικών.

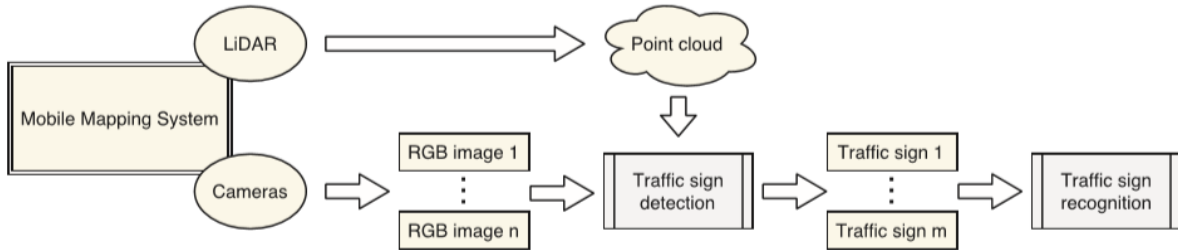
Ο DBSCAN είναι ένας αλγόριθμος βασισμένος στην πυκνότητα, όπου η πυκνότητα είναι ένας αριθμός σημείων (MinPts) μέσα σε προκαθορισμένη ακτίνα (Eps). Ο αλγόριθμος αυτός διαχωρίζει τα σημεία σε: βασικά σημεία που βρίσκονται στο εσωτερικό μιας συστάδας, οριακά που βρίσκονται στη γειτονιά ενός βασικού σημείου και τα σημεία που αποτελούν θόρυβο, τα οποία και αφαιρούνται στη συνέχεια. Κάθε ομάδα συνδεδεμένων βασικών σημείων δημιουργείται ως διαφορετική συστάδα στις οποίες εντάσσονται και τα οριακά σημεία.



DBSCAN αρχικά σημεία και συστάδες
ΠΗΓΗ: <https://slideplayer.com/slide/3900625/>

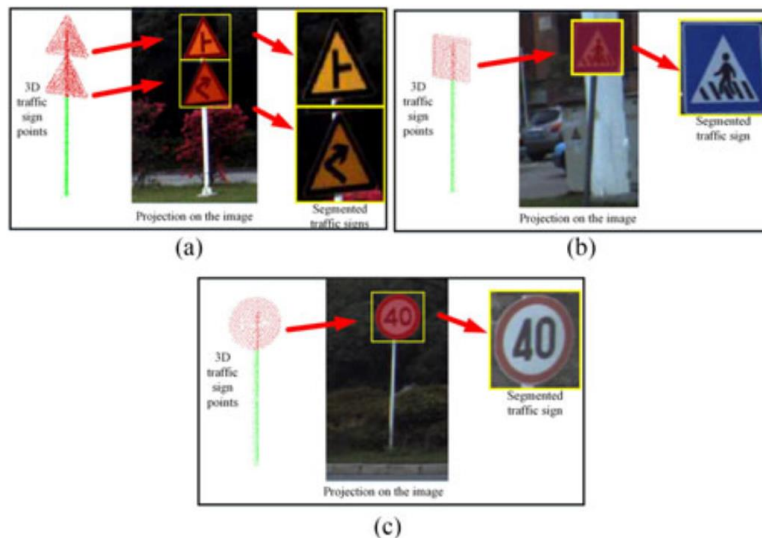
Για τη συσχέτιση δεδομένων ενός σημείου laser και του αντίστοιχου εικονοστοιχείου εικόνας, οι παράμετροι της κάμερας και του LiDAR πρέπει να υπολογίζονται εκ των προτέρων, όπως αναφέρουν οι Deng et al. (2018) [145]. Στη μελέτη τους, συσχέτισαν δεδομένα LiDAR και κάμερας από την συλλογή έγχρωμων σαρώσεων laser, οι οποίες ταυτόχρονα περιέχουν χρωματικές και γεωμετρικές πληροφορίες, ενώ συνδυάζουν τρεις διαφορετικούς χρωματικούς χώρους RGB, HSV, και CIE-L*a*b με την ανακλαστικότητα του laser. Η ανίχνευση επιτυγχάνεται με τη χρήση Μηχανών Διανυσματικής Στήριξης (SVM), ενώ η αναγνώριση με τον συνδυασμό HOG και ταξινομητή SVM. Συνολικά, το ποσοστό ανίχνευσης της μεθόδου έχει επιτυχία έως και 95.87% και ο ρυθμός αναγνώρισης φτάνει ακόμη και το 95.07% για πινακίδες κυκλοφορίας εντός απόστασης μέχρι 100 m, με μέσο χρόνο λειτουργίας περίπου 33.25 msec ανά καρτέ.

Ένα αποτελεσματικό σύστημα ανίχνευσης και αναγνώρισης σημάτων οδικής κυκλοφορίας σε δύο στάδια παρουσιάστηκε από τους Arcos-García et al. (2017) [146]. Το ανακλαστικό υλικό των πινακίδων κυκλοφορίας χρησιμοποιήθηκε για το σχεδιασμό της διαδικασίας ανίχνευσης βάσει τρισδιάστατων νεφών σημείων από LiDAR. Στη συνέχεια, σχεδιάστηκε ένα βαθύ νευρωνικό δίκτυο για την ταξινόμηση των εικόνων RGB στις οποίες έχουν πρώτα προβληθεί τα δεδομένα από το νέφος σημείων.



Σύστημα ανίχνευσης και αναγνώρισης
ΠΗΓΗ: Arcos-García et al. (2017) [146]

Οι Guan et al. (2018) [147] συγκεντρώνουν αρχικά τα δεδομένα (3D νέφος σημείων και 2D εικόνες) με όχημα εξοπλισμένο με LiDAR και κάμερες RGB. Στη συνέχεια, επεξεργάζονται το νέφος σημείων ώστε να εντοπίζονται αυτόματα τα σήματα κυκλοφορίας με βάση τις ανακλαστικές ιδιότητές τους. Χρησιμοποιείται η a priori γνώση του ύψους του στύλου των πινακίδων και του πλάτους του δρόμου για την αφαίρεση της πλειονότητας των σημείων από τα δεδομένα. Αφού οι πινακίδες από το νέφος σημείων προβληθούν στις αντίστοιχες ψηφιακές εικόνες, ακολουθεί η αναγνώριση με τη χρήση αλγορίθμου βαθιάς μηχανικής μάθησης. Επιτεύχθηκε η ακρίβεια ανίχνευσης 86.8% και η συνολική απόδοση της ταξινόμησης των σημάτων κυκλοφορίας πέτυχε ποσοστό αναγνώρισης 93.3%.



Δείγματα πινακίδων από νέφος σημείων τα οποία προβάλλονται στις αντίστοιχες ψηφιακές εικόνες τους: (a) τριγωνικές, (b) τετράγωνες, και (c) κυκλικές πινακίδες.

ΠΗΓΗ: Οι Guan et al. (2018) [147]

Η συσχέτιση δεδομένων LiDAR και κάμερας είναι ένας πολλά υποσχόμενος τρόπος για την ανίχνευση και αναγνώριση των πινακίδων κυκλοφορίας, όμως λόγω της ιδιαιτερότητας του 3D νέφους σημείων οι περισσότερες από τις παραδοσιακές μεθόδους βαθιάς μάθησης είναι δύσκολο να εφαρμοστούν άμεσα στο πρόβλημα ανίχνευσης από νέφος σημείων. Όπως έχει ήδη αναφερθεί, χρησιμοποιούνται συνήθως διαφορετικές διαδικασίες προεπεξεργασίας του νέφους σημείων, όπως η κατωφλίωση, το φιλτράρισμα και η ομαδοποίηση. Σε αντίθεση με τις μεθόδους βαθιάς μάθησης, αυτές οι διαδικασίες συνήθως εξακολουθούν να βασίζονται σε χειροκίνητα σχεδιασμένα μοντέλα και κατώφλια, τα οποία μπορεί να οδηγήσουν σε περιορισμένη δυνατότητα γενίκευσης.

3

Εφαρμογή

Στα πλαίσια της μεταπτυχιακής διπλωματικής εργασίας χρησιμοποιήθηκε η μέθοδος του YOLOv3, αφού μέχρι την εμφάνιση της νέας του εκδοχής (το 2020, και ενώ η εκπόνηση της παρούσας εργασίας είχε προχωρήσει) αποτελούσε και ακόμα βέβαια αποτελεί, έναν από τους αλγόριθμους αιχμής (state-of-the-art) για ανίχνευση και αναγνώριση σημάτων οδικής κυκλοφορίας σε πραγματικό χρόνο, ως συστατικό μέρος ενός συστήματος υποβοήθησης οδήγησης.

Συγκεκριμένα, υλοποιήθηκε η εκπαίδευση και η αξιολόγηση του YOLOv3 και YOLOv3-tiny σε PyTorch της Ultralytics που υπάρχει στο GitHub [139], πάνω στο γερμανικό σύνολο δεδομένων GTSDb πινακίδων οδικής κυκλοφορίας, με σκοπό τη σύγκριση τους. Ο YOLOv3-tiny είναι ένας μικρότερος αλλά ταχύτερος αλγόριθμος YOLOv3. Χρησιμοποιεί πιο ελαφρύ δίκτυο ως βασικό δίκτυο και, όπως ο YOLOv3, ακολουθεί τη λογική του δικτύου πυραμίδας για τις πολλαπλές κλίμακες.

Τέλος, ακολούθησε η εφαρμογή των μοντέλων σε ένα σύνολο 30 περίπου εικόνων που πήρε η γράφουσα σε διαφορετικές ώρες της ημέρας από το εσωτερικό εν κινήσει οχήματος, υπό διαφορετικές συνθήκες φωτισμού και με διαφορετικά κινητά τηλέφωνα. Θα πρέπει να υπενθυμιστεί πως τα μοντέλα εκπαιδεύτηκαν σε σύνολο γερμανικών πινακίδων και εφαρμόστηκαν πειραματικά στο τέλος σε ένα σύνολο ελληνικών πινακίδων.

Στα επόμενα αναλύονται τα βασικά βήματα που ακολουθεί ο αλγόριθμος YOLOv3, καθώς και οι διαφοροποιήσεις και βελτιώσεις από τις προηγούμενες εκδόσεις.

3.1 Ο αλγόριθμος YOLOv3

Το 2018 οι Redmon et al. [127] δημοσίευσαν την 3^η έκδοση του αλγορίθμου YOLO. Στον YOLOv3, όπως έχει ήδη αναφερθεί, βελτιώθηκε η μέση ακρίβεια για τα μικρά αντικείμενα, η μέση αντιπροσωπευτική ακρίβεια mAP αυξήθηκε με αποτέλεσμα να μειωθούν και τα σφάλματα εντοπισμού, ενώ οι προβλέψεις για διαφορετικές κλίμακες ή αναλογίες διαστάσεων για το ίδιο αντικείμενο βελτιώθηκαν επίσης σημαντικά. Ο YOLOv3 βασίστηκε στα προηγούμενα μοντέλα, ενώ οι τροποποιήσεις που πέτυχαν τις παραπάνω βελτιώσεις αφορούν τα εξής τέσσερα συστατικά μέρη του αλγορίθμου:

1. Πρόβλεψη του πλαισίου οριοθέτησης.
2. Πρόβλεψη της κλάσης.
3. Προβλέψεις σε διαφορετικές κλίμακες.
4. Εξαγωγή χαρακτηριστικών.

Πρόβλεψη του πλαισίου οριοθέτησης

Το YOLOv3, όπως και το YOLOv2, χρησιμοποιεί συστάδες διαστάσεων για τη δημιουργία των πλαισίων οριοθέτησης. Το δίκτυο προβλέπει 4 συντεταγμένες για κάθε πλαίσιο οριοθέτησης, t_x, t_y, t_w, t_h . Εάν το κελί είναι μετατοπισμένο από την επάνω αριστερή γωνία της εικόνας κατά (c_x, c_y) και το πλαίσιο οριοθέτησης έχει προηγουμένως πλάτος και ύψος p_w, p_h , τότε οι προβλέψεις για το αντίστοιχο προβλεπόμενο πλαίσιο b με κέντρο (b_x, b_y) και μέγεθος (b_w, b_h) αντιστοιχούν σε:

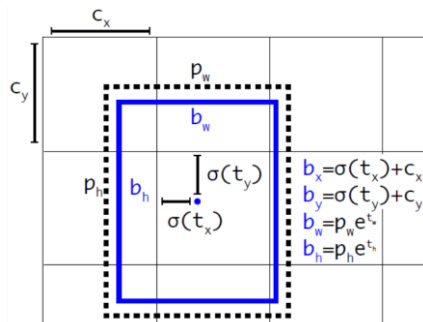
$$b_x = \sigma(t_x) + c_x$$

$$b_y = \sigma(t_y) + c_y$$

$$b_w = p_w e^{t_w}$$

$$b_h = p_h e^{t_h}$$

Οι κεντρικές συντεταγμένες του πλαισίου σε σχέση με τη θέση της εφαρμογής του φίλτρου προβλέπονται με χρήση μιας σιγμοειδούς συνάρτησης.



Πρόβλεψη πλαισίου οριοθέτησης, πλαίσιο που προβλέφθηκε (μπλε), προηγούμενο πλαίσιο (μαύρες διακεκομμένες γραμμές) ΠΗΓΗ: Redmon et al. [127]

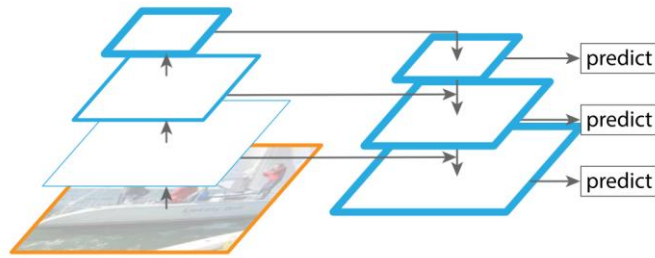
Κατά τη διάρκεια της εκπαίδευσης, χρησιμοποιείται το άθροισμα της απώλειας του τετραγωνικού σφάλματος. Προβλέπεται η τιμή "αντικειμενικότητας" (objectness score) για κάθε πλαίσιο οριοθέτησης με βάση τη μέθοδο της λογιστικής παλινδρόμησης. Η τιμή ισούται με 1 στην περίπτωση που το πλαίσιο οριοθέτησης προηγουμένως επικαλύπτει το πραγματικό πλαίσιο περισσότερο από οποιοδήποτε άλλο προηγούμενο πλαίσιο οριοθέτησης. Εάν το πλαίσιο οριοθέτησης δεν είναι το καλύτερο αλλά επικαλύπτει το πραγματικό πλαίσιο περισσότερο από οποιοδήποτε άλλο, αγνοείται η πρόβλεψη που ακολουθεί.

Πρόβλεψη κλάσης

Μια προσέγγιση με πολλές ετικέτες μοντελοποιεί καλύτερα τα δεδομένα. Για το λόγο αυτό ο YOLOv3 χρησιμοποιεί ανεξάρτητους ταξινομητές (independent logistic classifiers) για κάθε κατηγορία αντί για ένα κανονικό επίπεδο softmax. Η χρήση ενός softmax επιβάλλει την υπόθεση ότι κάθε πλαίσιο περιέχει ακριβώς μια κατηγορία που συχνά δεν ισχύει. Επομένως, κάθε πλαίσιο προβλέπει τις πολλαπλές κατηγορίες που μπορεί να περιέχει το πλαίσιο οριοθέτησης. Κατά τη διάρκεια της εκπαίδευσης χρησιμοποιείται δυαδική απώλεια εντροπίας (binary cross-entropy) για τις προβλέψεις της τάξης.

Προβλέψεις σε διαφορετικές κλίμακες

Για να υποστηρίξει την ανίχνευση σε ποικίλες κλίμακες ο YOLOv3 προβλέπει πλαίσια σε 3 διαφορετικές κλίμακες. Στη συνέχεια, τα χαρακτηριστικά εξάγονται από κάθε κλίμακα με μια μέθοδο παρόμοια με εκείνη των δικτύων πυραμίδας (Feature Pyramid Network - FPN). Τα δίκτυα πυραμίδας αποτελούν μια από πάνω προς τα κάτω αρχιτεκτονική με πλευρικές συνδέσεις για την κατασκευή χαρτών υψηλού επιπέδου σημασιολογικών χαρακτηριστικών (high-level semantic feature maps) σε όλες τις κλίμακες.



Δίκτυο πυραμίδας FPN
ΠΗΓΗ: Lin et al. (2017) [140]

Χρησιμοποιώντας την παραπάνω μέθοδο, ο YOLOv3 αποκτά την ικανότητα να προβλέπει καλύτερα σε διαφορετικές κλίμακες. Τα προγενέστερα πλαίσια οριοθέτησης που δημιουργούνται χρησιμοποιώντας συστάδες διαστάσεων χωρίζονται σε 3 κλίμακες, ώστε να υπάρχουν 3 προγενέστερα πλαίσια οριοθέτησης ανά κλίμακα και έτσι συνολικά 9 προγενέστερα πλαίσια οριοθέτησης. Επομένως:

- Χρησιμοποιούνται 3 διαφορετικές κλίμακες.
- Τα χαρακτηριστικά εξάγονται από αυτές τις κλίμακες με τη μέθοδο δικτύου πυραμίδας FPN.
- Πολλά συνελικτικά στρώματα προστίθενται στο βασικό δίκτυο Darknet-53 (το οποίο αναφέρεται στην επόμενη ενότητα για την εξαγωγή των χαρακτηριστικών).
- Το τελευταίο από αυτά τα επίπεδα προβλέπει το πλαίσιο οριοθέτησης, την αντικειμενικότητα και τις προβλέψεις κλάσης.
- Το διάνυσμα εξόδου είναι:

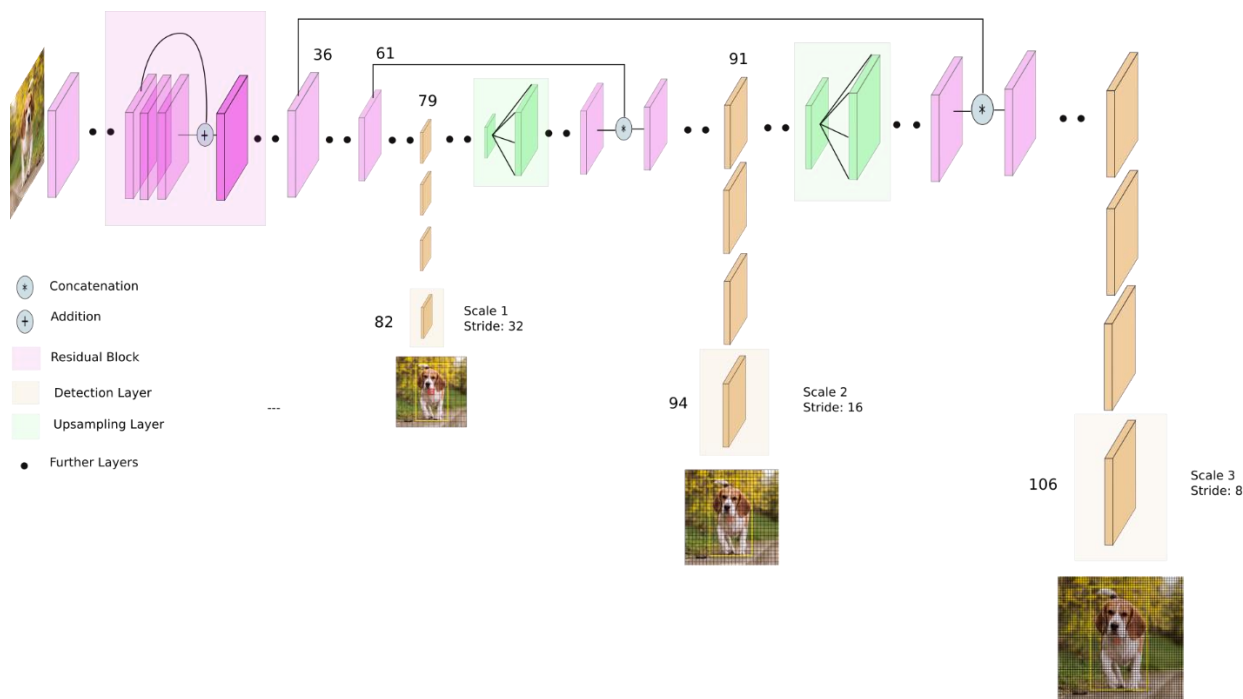
$$N \times N \times [\text{number of scales} \times (\text{bounding box offsets} + \text{objectness} + \text{class predictions})]$$

Εξαγωγή χαρακτηριστικών: Darknet-53

Για την εξαγωγή των χαρακτηριστικών χρησιμοποιείται ένα δίκτυο κορμού πολύ βαθύτερο από του YOLOv2. Συγκεκριμένα, αποτελείται από 53 συνελικτικά επίπεδα, για το λόγο αυτό ονομάζεται Darknet-53. Το δίκτυο χρησιμοποιεί διαδοχικές συνελίξεις 3×3 και 1×1 , αλλά τώρα έχει και κάποιες συνδέσεις συντόμευσης όπως φαίνονται παρακάτω.

	Type	Filters	Size	Output
	Convolutional	32	3×3	256×256
	Convolutional	64	$3 \times 3 / 2$	128×128
1x	Convolutional	32	1×1	128×128
	Convolutional	64	3×3	
	Residual			128×128
	Convolutional	128	$3 \times 3 / 2$	64×64
2x	Convolutional	64	1×1	64×64
	Convolutional	128	3×3	
	Residual			64×64
	Convolutional	256	$3 \times 3 / 2$	32×32
8x	Convolutional	128	1×1	32×32
	Convolutional	256	3×3	
	Residual			32×32
	Convolutional	512	$3 \times 3 / 2$	16×16
8x	Convolutional	256	1×1	16×16
	Convolutional	512	3×3	
	Residual			16×16
	Convolutional	1024	$3 \times 3 / 2$	8×8
4x	Convolutional	512	1×1	8×8
	Convolutional	1024	3×3	
	Residual			8×8
	Avgpool		Global	
	Connected		1000	
	Softmax			

Darknet-53
ΠΗΓΗ: Redmon et al. [127]



Αρχιτεκτονική δικτύου YOLOv3

ΠΗΓΗ: <https://towardsdatascience.com/yolo-v3-object-detection-53fb7d3bfe6b>

3.2 Το σύνολο δεδομένων GTSDDB

Υπάρχουν αρκετές βάσεις με σύνολα δεδομένων για πινακίδες οδικής κυκλοφορίας που μπορούν να χρησιμοποιηθούν ελεύθερα για ερευνητικούς σκοπούς. Ανάμεσα τους είναι το Γερμανικό GTSDDB, το Βελγικό BTSD, το Σουηδικό STSD, το Ιταλικό DITS και άλλα. Το Γερμανικό σύνολο δεδομένων είναι το πιο διαδεδομένο. Στην παρούσα εφαρμογή επιλέχθηκε αυτό καθότι οι γερμανικές πινακίδες έχουν αρκετά κοινά με τις ελληνικές. Η βασική διαφορά είναι μόνο πως οι τριγωνικές ελληνικές πινακίδες (κινδύνου) είναι συνήθως κίτρινες στο εσωτερικό τους, ενώ οι γερμανικές είναι λευκές.



Διαφορά ελληνικών (αριστερά) και γερμανικών (δεξιά) τριγωνικών πινακίδων

Οι εικόνες του γερμανικού συνόλου δεδομένων GTSDDB συλλέχθηκαν κοντά στο Μπόχουμ της Γερμανίας, σε διάφορες διαδρομές με αυτοκίνητο την άνοιξη και το φθινόπωρο του 2010. Καταγράφονται διαφορετικά σενάρια τοπίου, αστικά, αγροτικά, αυτοκινητόδρομοι, κατά τη διάρκεια της ημέρας και το σούρουπο, με διαφορετικές καιρικές συνθήκες. Τα καταγεγραμμένα σήματα κυκλοφορίας ορίζονται από τη Σύμβαση της Βιέννης για τα οδικά σήματα κυκλοφορίας και τα σήματα που εναρμονίζουν την εμφάνισή τους σε 62 χώρες.

Το σύνολο δημιουργήθηκε προκειμένου να χρησιμοποιηθεί για την αξιολόγηση του προβλήματος της ανίχνευσης με μία μόνο εικόνα. Αποτελείται από 900 εικόνες 1360 × 800 εικονοστοιχείων. Οι εικόνες διαχωρίζονται σε 600 εικόνες εκπαίδευσης και 300 εικόνες αξιολόγησης. Κατηγοριοποιούνται σε 3 κλάσεις: υποχρεωτικές, προειδοποιητικές και απαγορευτικές (mandatory, warning, prohibitory).



Ενδεικτικές εικόνες από το Γερμανικό σύνολο δεδομένων GTSDDB
ΠΗΓΗ: [138]

3.3 Εκπαίδευση Αλγορίθμου

Η εκπαίδευση και αξιολόγηση των αλγορίθμων YOLOv3 και YOLOv3-tiny υλοποιήθηκε με βάση τα βήματα που παρουσιάζονται στο [139], με τη χρήση της Python3 και της βιβλιοθήκης PyTorch .

Η PyTorch είναι μια βιβλιοθήκη μηχανικής εκμάθησης ανοιχτού κώδικα σε Python. Χρησιμοποιείται ευρέως για εφαρμογές όπως η όραση υπολογιστών (Computer Vision - CV) και η επεξεργασία φυσικής γλώσσας (Natural Language Processing - NLP), και αναπτύχθηκε κυρίως από το εργαστήριο AI Research της Facebook. Πολλά τμήματα λογισμικού Βαθιάς Μάθησης έχουν δημιουργηθεί σε PyTorch, συμπεριλαμβανομένων του Tesla Autopilot, ενός προηγμένου συστήματος υποστήριξης οδηγού της Tesla, και της γλώσσας προγραμματισμού Pyro της Uber η οποία συνδυάζει τη Βαθιά Μάθηση με μοντέλα Bayesian με σκοπό την επιτάχυνση της έρευνας για την αξιόπιστη μεταφορά όλων.

Εκτός από την εγκατάσταση της PyTorch, βασικό υπολογιστικό εξάρτημα για την ανίχνευση και αναγνώριση με αλγόριθμους βαθιάς μηχανικής μάθησης, οι οποίοι χρειάζονται ενδεχομένως και ολόκληρες ημέρες για την εκπαίδευση τους μέχρι να συγκλίνουν, αποτελεί η χρήση μιας «δυνατής» κάρτας γραφικών (Graphic Processing Unit – GPU). Η κάρτα γραφικών που χρησιμοποιήθηκε εδώ σε λειτουργικό σύστημα Ubuntu είναι η GeForce GTX 1080 Ti της NVIDIA, η οποία μπορεί να σηκώσει παιχνίδια τελευταίας τεχνολογίας, με εξαιρετική απόδοση σε ταχύτητα.

Όπως έχει ήδη αναφερθεί, το σύνολο δεδομένων που χρησιμοποιήθηκε για την εκπαίδευση του αλγορίθμου είναι το γερμανικό σύνολο δεδομένων GTSDb [138], το οποίο αποτελείται από το σύνολο των 900 εικόνων, οι οποίες χωρίζονται σε 600 εικόνες εκπαίδευσης και 300 εικόνες αξιολόγησης. Στην παρούσα εφαρμογή το σύνολο των δεδομένων χωρίστηκε τυχαία σε 66% εικόνες εκπαίδευσης και 33% εικόνες αξιολόγησης. Συγκεκριμένα, πρόκειται για 491 εικόνες εκπαίδευσης και 245 εικόνες αξιολόγησης, σύνολο 736 εικόνες. Η διαφορά που προκύπτει από τις 900 φωτογραφίες είναι λόγω του ότι οι 159 εικόνες δεν είχαν το πραγματικό πλαίσιο οριοθέτησης (ground-truth).

Οι εικόνες για την ανίχνευση και αναγνώριση ταξινομούνται σε τέσσερις διαφορετικές κλάσεις: απαγορευτικές, προειδοποιητικές/κινδύνου, υποχρεωτικές και άλλες (prohibitory, danger, mandatory, other). Για την εκπαίδευση του αλγορίθμου οι εικόνες δεν υπέστησαν καμία επιπλέον επεξεργασία. Ο ρυθμός μάθησης ορίστηκε ως 0.001 με βελτιστοποίηση στοχαστικής κατάβασης κλίσης (Stochastic Gradient Descent - SGD) και αριθμό παρτίδας που διαδίδονται μέσω του αλγορίθμου (batch size) ίσο με 8, οι δε εποχές ορίστηκαν ως 200. Το κατώφλι IoU διατήρηθηκε σταθερό στο 0.5. Ακριβώς οι ίδιοι παράμετροι ορίστηκαν και για τα δύο δίκτυα YOLOv3 και YOLOv3-tiny.



Οι 43 διαφορετικές πινακίδες του συνόλου GTSDb. Από πάνω προς τα κάτω, χωρίζονται στις τέσσερις κλάσεις (απαγορευτικές, προειδοποιητικές/κινδύνου, υποχρεωτικές και άλλες)

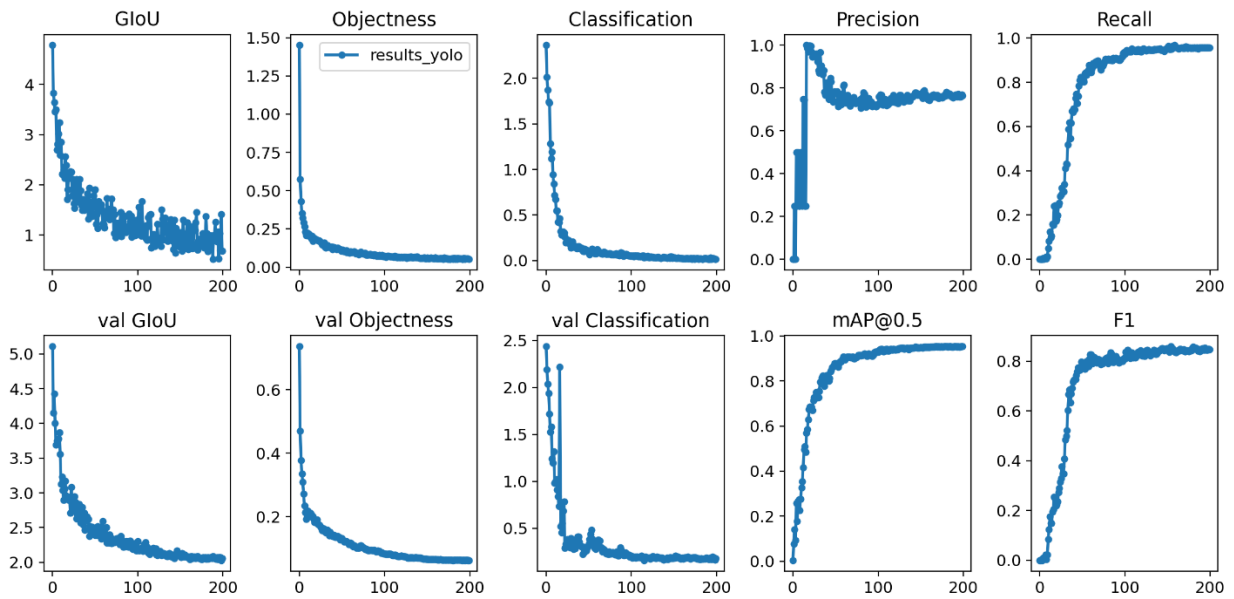
3.4 Αποτελέσματα

Παρουσιάζονται εδώ τα αποτελέσματα εκπαίδευσης και αξιολόγησης των δύο αλγορίθμων YOLOv3 και YOLOv3-tiny, καθώς και η σύγκρισή τους. Επιπλέον, ακολουθούν τα αποτελέσματα εφαρμογής των μοντέλων σε ένα σύνολο 30 περίπου εικόνων τραβηγμένων από διαφορετικά κινητά τηλέφωνα από το εσωτερικό εν κινήσει οχήματος.

3.4.1 Αποτελέσματα YOLOv3

Ο αλγόριθμος στο σύνολο του πετυχαίνει ακρίβεια/precision 76.5% και ανάκληση/recall 95.7%, ενώ το F1score ανέρχεται στο 84.7%. Όπως έχει ήδη αναφερθεί, καλή βαθμολογία F1 σημαίνει ότι το μοντέλο έχει χαμηλά ψευδώς θετικά και χαμηλά ψευδώς αρνητικά, επομένως εξάγει σωστά αποτελέσματα στην ταξινόμηση. Η μέση αντιπροσωπευτική ακρίβεια ανίχνευσης mAP είναι 95.3%.

Στα επόμενα διαγράμματα φαίνεται η εξέλιξη των βασικών μετρητικών σε κάθε εποχή. Αντίστοιχα, οπτικοποιείται και η ελαχιστοποίηση των σφαλμάτων με το σφάλμα αντικειμενικότητας να ελαχιστοποιείται στην εκπαίδευση στο 0.0533, ενώ το σφάλμα της ταξινόμησης στο 0.0147, όπου ο GIoU η ελαχιστοποίηση του γενικευμένου σφάλματος IoU εκπαίδευσης και val IoU η ελαχιστοποίηση σφάλματος IoU αξιολόγησης. Αντίστοιχα ισχύουν για την ελαχιστοποίηση του σφάλματος της αντικειμενικότητας και της ταξινόμησης.



YOLOv3: Διαγράμματα εξέλιξης των βασικών μετρητικών σε κάθε εποχή.

Ακολουθούν παραδείγματα εικόνων από τα αποτελέσματα της αξιολόγησης. Όπως φαίνεται και στις εικόνες, οι πινακίδες συνήθως εντοπίζονται και ταξινομούνται σωστά σε μία από τις τέσσερις κλάσεις (απαγορευτικές, προειδοποιητικές/κινδύνου, υποχρεωτικές και άλλες). Το γνωστό πρόβλημα της ανίχνευσης των πολύ μικρών αντικειμένων, παρ' ότι έχει αυξηθεί η mAP, συνεχίζει να υφίσταται σε ένα ποσοστό. Αυτό γίνεται εύκολα αντιληπτό από την πρόβλεψη της κλάσης που υπάρχει δίπλα από το πλαίσιο οριοθέτησης: όσο μικρότερο είναι το σήμα οδικής κυκλοφορίας στην εικόνα τόσο μικρότερη και η πιθανότητα της πρόβλεψης. Αντίστοιχα, υπάρχουν περιπτώσεις όπου ο αλγόριθμος λανθασμένα εντοπίζει πινακίδα στο φόντο ενώ υπάρχουν και περιπτώσεις που λόγω αποκρύψεων αποτυγχάνει πλήρως να εντοπίσει την πινακίδα. Στο γενικότερο πλαίσιο η ποιότητα των αποτελεσμάτων του είναι ιδιαίτερα υψηλή.





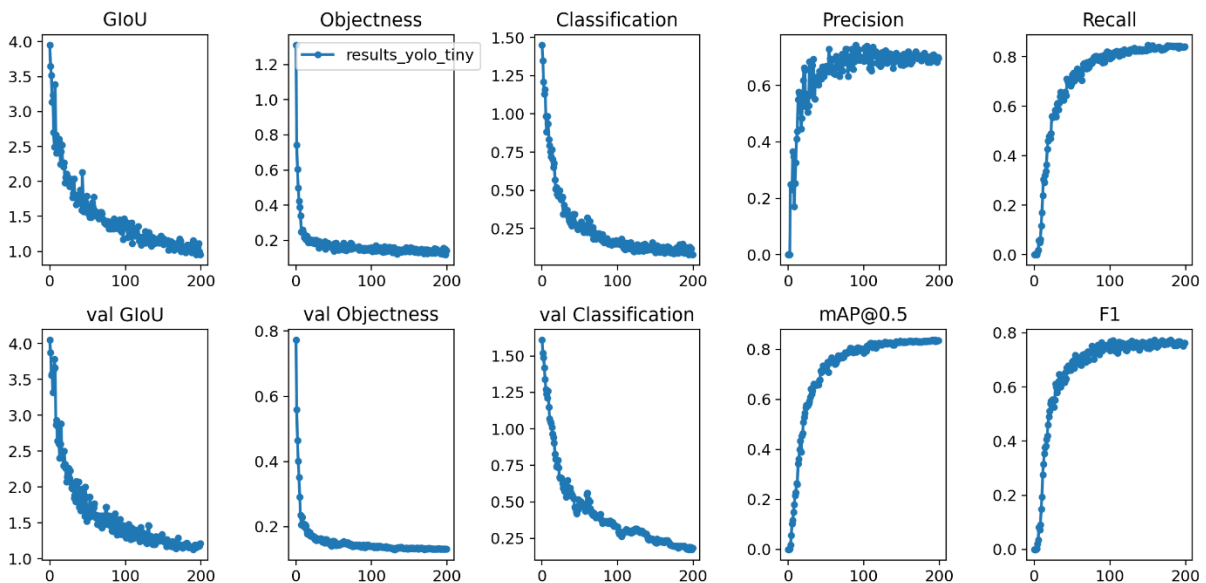
Παράδειγματα επιτυχημένης ανίχνευσης και ταξινόμησης από το σύνολο των δεδομένων αξιολόγησης του YOLOv3



Παράδειγμα αποτυχημένης ανίχνευσης και αναγνώρισης φόντου ως πινακίδα (αριστερά) και παράδειγμα πλήρους αποτυχίας ανίχνευσης σήματος αναγόρευσης κινδύνου λόγω αποκρύψεων (δεξιά) από το σύνολο των δεδομένων αξιολόγησης του YOLOv3

3.4.2 Αποτελέσματα YOLOv3-tiny

Ο tiny αλγόριθμος, ο οποίος αποτελεί ελαφρότερη έκδοση με μικρότερο δίκτυο στον βασικό σκελετό του, πετυχαίνει συνολική ακρίβεια/precision 69.8%, και ανάκληση/recall 84.2%, ενώ το F1score ανέρχεται στο 76.3%. Η μέση αντιπροσωπευτική ακρίβεια ανίχνευσης mAP είναι 83.5%. Το σφάλμα αντικειμενικότητας ελαχιστοποιείται στην στο 0.141, ενώ το σφάλμα της ταξινόμησης στο 0.0798.



YOLOv3 - tiny: Διαγράμματα εξέλιξης των βασικών μετρητικών σε κάθε εποχή.



Παραδείγματα επιτυχημένης ανίχνευσης και ταξινόμησης από το σύνολο των δεδομένων αξιολόγησης του YOLOv3 - tiny

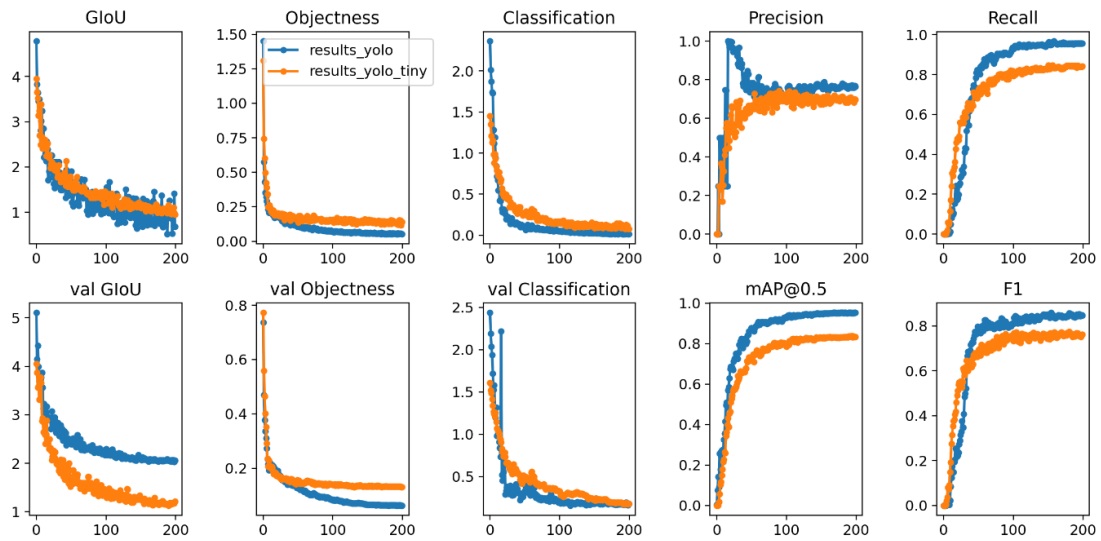


Παράδειγμα πλήρους αποτυχίας ανίχνευσης σημάτων του YOLOv3 - tiny

3.4.3 Σύγκριση και αξιολόγηση αποτελεσμάτων

Οι δύο αλγόριθμοι έχουν πολύ καλά αποτελέσματα, με τον YOLOv3 πάντως να προηγείται τουλάχιστον πάνω από 10 μονάδες στη μέση αντιπροσωπευτική ακρίβεια ανίχνευσης mAP. Η υψηλότερη απόδοση που επιτεύχθηκε αντιστοιχεί στο 95.3% του YOLOv3, ενώ του tiny

ανέρχεται σε 83.5%. Το αντίστοιχο ποσοστό ακρίβειας για την αναγνώριση ανέρχεται σε 84.7% από τον αλγόριθμο YOLOv3 και 76.3% από τον tiny. Η διαφορές των μετρητικών των δύο αλγορίθμων και της εξέλιξής τους σε κάθε εποχή οπτικοποιούνται στα παρακάτω διαγράμματα. Εμφανώς, ο YOLOv3 (μπλε) πετυχαίνει καλύτερες ακρίβειες.



YOLOv3 (μπλε) και YOLOv3-tiny (πορτοκαλί): Συγκριτικά διαγράμματα εξέλιξης των βασικών μετρητικών σε κάθε εποχή

Στην συνέχεια παρατίθενται παραδείγματα εικόνων και από τους δύο αλγορίθμους (YOLOv3 αριστερά και YOLOv3-tiny δεξιά). Στις περισσότερες περιπτώσεις ο YOLOv3 πετυχαίνει καλύτερη εκτίμηση από τον YOLOv3-tiny.

Το πρόβλημα της ανίχνευσης των πολύ μικρών αντικειμένων, υφίσταται και στους δύο αλγορίθμους, με τον YOLOv3 να μπορεί να ανιχνεύσει σε ορισμένες περιπτώσεις όπου ο tiny αποτυγχάνει (πχ. έβδομο ζευγάρι εικόνων). Στις μεγάλες αποκρύψεις φαίνεται να αποτυγχάνουν και οι δύο αλγόριθμοι, ενώ ο tiny σε περιπτώσεις όπου ο YOLOv3 λανθασμένα εντοπίζει πινακίδα στο φόντο φαίνεται να μην επηρεάζεται έντονα από αντίστοιχα φαινόμενα.







Σύγκριση αποτελεσμάτων YOLOv3 (αριστερά) και YOLOv3 – tiny (δεξιά)

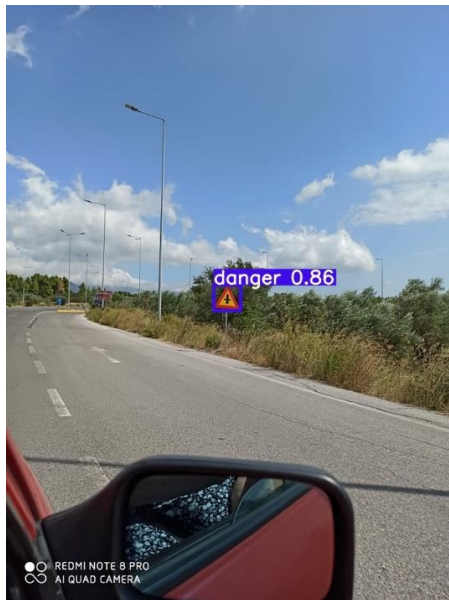
Στη συνέχεια, ακολούθησε η εφαρμογή των δύο μοντέλων σε ένα σύνολο 30 περίπου εικόνων που ελήφθησαν για της ανάγκες αυτής της εργασίας από διαφορετικά κινητά τηλέφωνα από το εσωτερικό εν κινήσει οχήματος. Στις εικόνες δεν προηγήθηκε κάποια προεπεξεργασία ούτε πρόκειται, επιπλέον, για εικόνες συγκεκριμένων διαστάσεων και αναλύσεων, αφού είναι τραβηγμένες από διαφορετικές συσκευές τηλεφώνου. Καταγράφονται επίσης διαφορετικά σενάρια τοπίου, αστικά, αγροτικά, αυτοκινητόδρομοι, κατά τη διάρκεια της ημέρας και της νύχτας με διαφορετικές καιρικές συνθήκες.

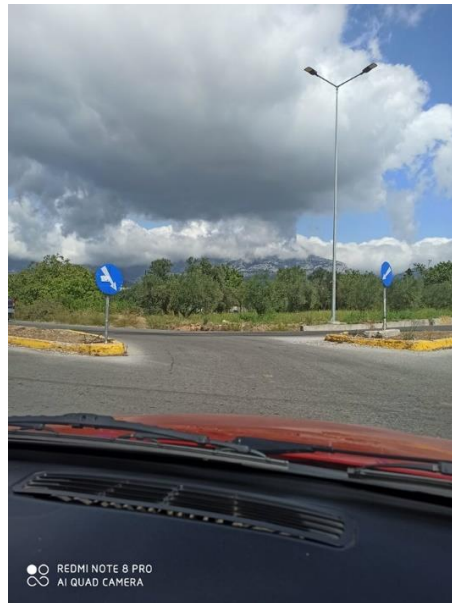
Στόχος ήταν να αναδειχθεί ότι δεν υπάρχει ανάγκη για ιδιαίτερο εξοπλισμό, όπως για παράδειγμα συγκεκριμένη φωτογραφική μηχανή. Έτσι θα μπορούσε ο αλγόριθμος να ενσωματωθεί σε εφαρμογή συσκευής τηλεφώνου ώστε η ανίχνευση και αναγνώριση να γίνεται από τη συσκευή του εκάστοτε οδηγού, η οποία θα τοποθετείται σε συγκεκριμένη θέση μέσα στο όχημα. Θα μπορούσε ακόμα να εφαρμοστεί για παλαιότερα μοντέλα οχημάτων που δεν έχουν ενσωματωμένο σύστημα υποβοήθησης οδήγησης, εξασφαλίζοντας με αυτόν τον τρόπο ασφαλέστερη οδήγηση.

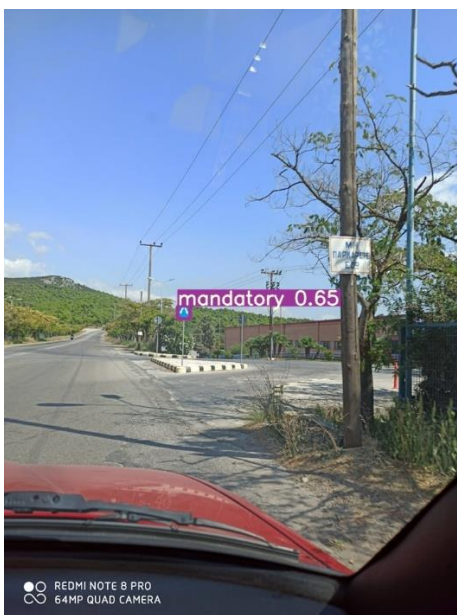
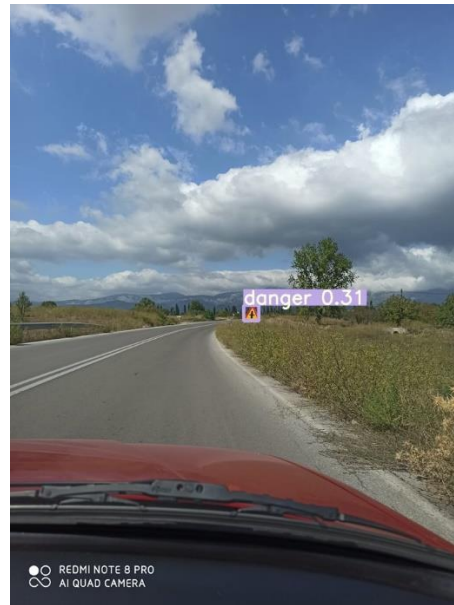
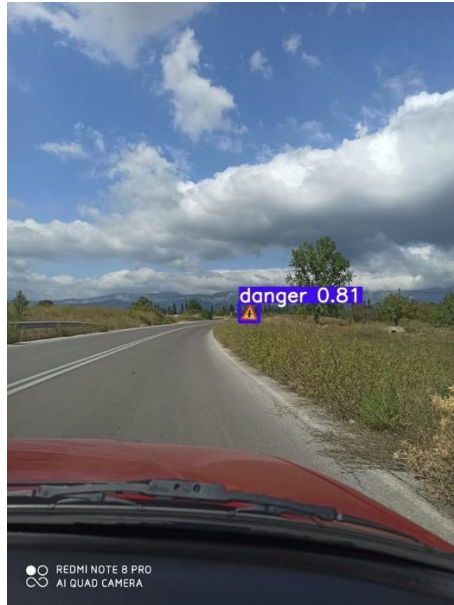
Αριστερά παρατίθενται οι εικόνες με αποτελέσματα από YOLOv3 και δεξιά από YOLOv3-tiny.



Μέθοδοι και τεχνικές ανίχνευσης & αναγνώρισης σημάτων οδικής κυκλοφορίας

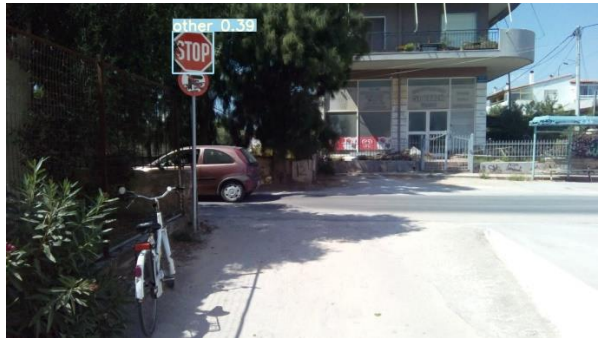
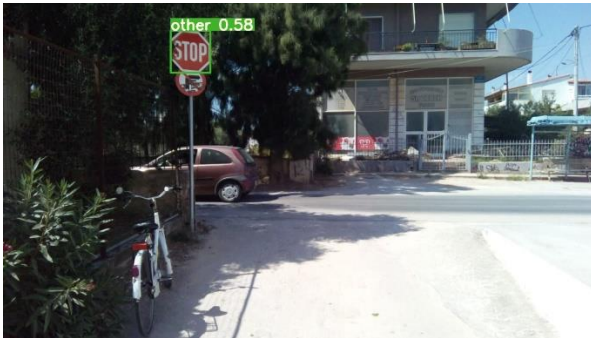




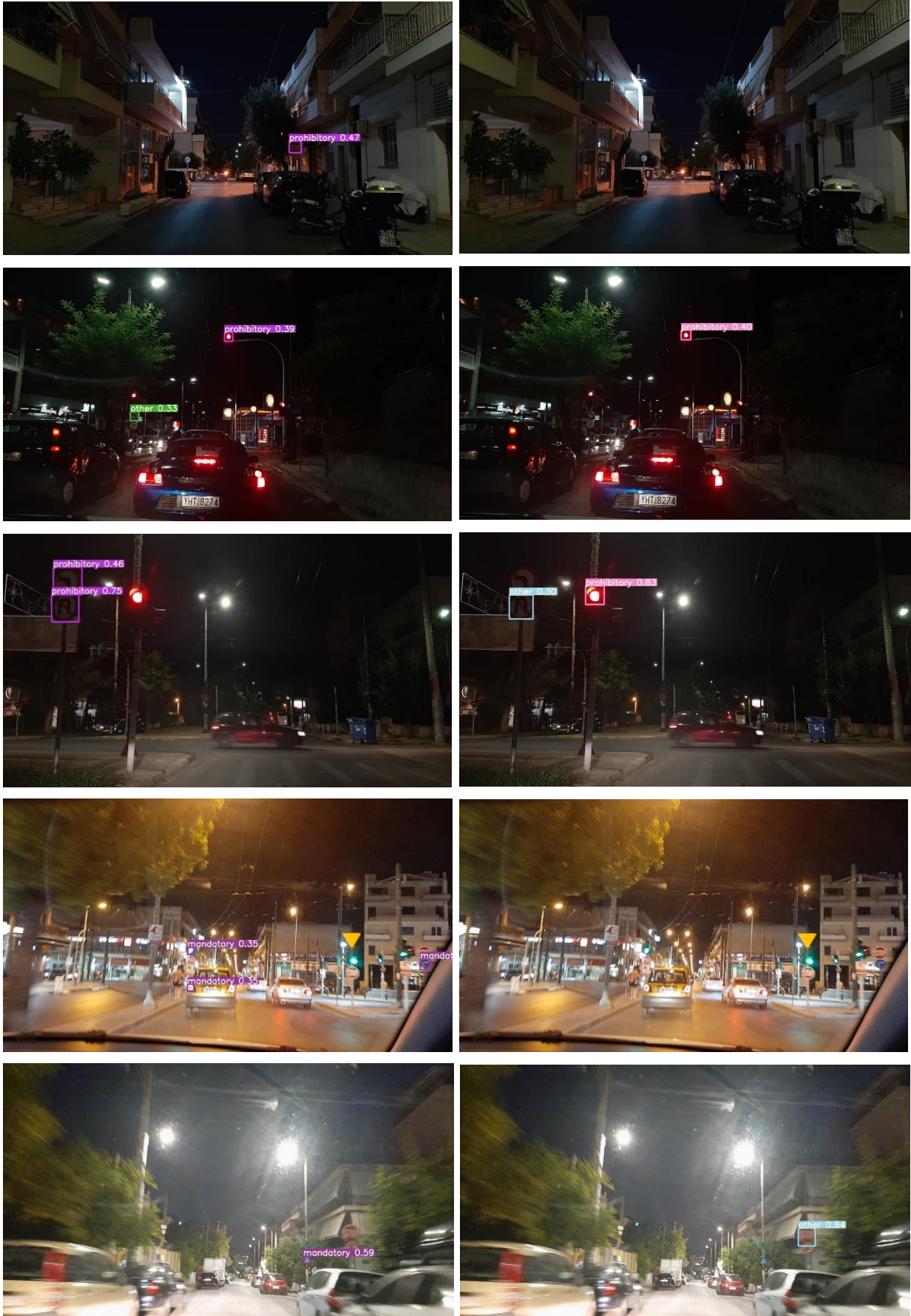


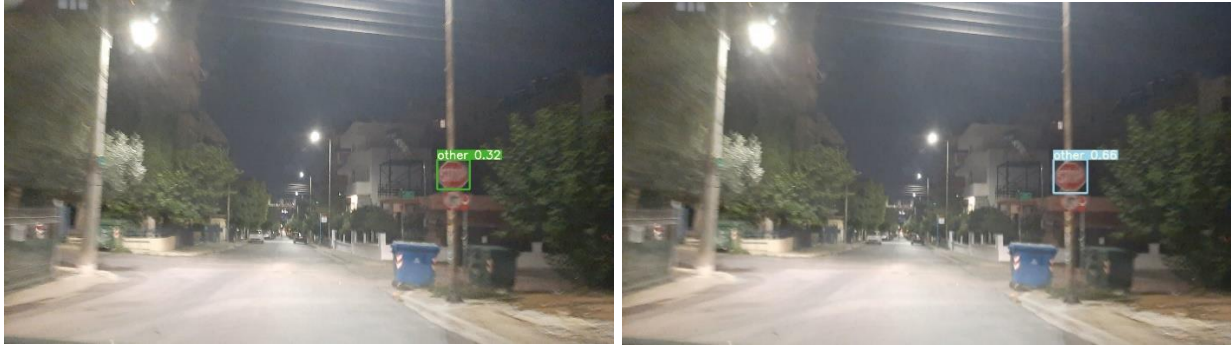


Μέθοδοι και τεχνικές ανίχνευσης & αναγνώρισης σημάτων οδικής κυκλοφορίας









Σύγκριση αποτελεσμάτων YOLOv3(αριστερά) και YOLOv3-tiny(δεξιά)

Όπως γίνεται άμεσα αντιληπτό, ο YOLOv3 επιτυγχάνει γενικώς καλύτερα αποτελέσματα στην ανίχνευση και στην αναγνώριση της κλάσης των πινακίδων. Επιπλέον, σε πολλές περιπτώσεις φαίνεται πως ο tiny δεν είναι ιδιαίτερα ανθεκτικός απέναντι σε στροφές. Το πρόβλημα με τις πολύ μικρές πινακίδες στις εικόνες όπου αυτές δεν ανιχνεύονται παρουσιάζεται στις παρακάτω δύο εικόνες. Στις νυχτερινές λήψεις, σαφώς, ανεβαίνει το επίπεδο δυσκολίας της ανίχνευσης και αναγνώρισης, με τον αλγόριθμο του YOLOv3 να πετυχαίνει υψηλότερο ποσοστό.



Περιπτώσεις μικρών απεικονιζόμενων πινακίδων που δεν ανιχνεύτηκαν από τους αλγόριθμους.

4

Συμπεράσματα

Στην παρούσα εργασία παρουσιάστηκαν κατ' αρχάς, κατά το δυνατόν πληρέστερα και αναλυτικότερα, και σχολιάστηκαν οι σύγχρονες μέθοδοι και τεχνικές ανίχνευσης και αναγνώρισης σημάτων οδικής κυκλοφορίας. Οι κλασικές μέθοδοι αντιμετωπίζουν την ανίχνευση και την αναγνώριση του εν λόγω αντικειμένου ως δύο επιμέρους ανεξάρτητες διαδικασίες, οι οποίες μπορούν να αντιμετωπιστούν με μεθόδους Επεξεργασίας Εικόνας (Image Processing) ή/και Μηχανικής Μάθησης. Οι περισσότερες προσεγγίσεις στον εντοπισμό σημάτων κυκλοφορίας βασίζονται σε μεθόδους εξαγωγής πληροφορίας χρώματος σε συνδυασμό με εξαγωγή πληροφορίας σχήματος για πιο εύρωστους αλγορίθμους.

Οι μέθοδοι αιχμής (State-of-the-Art) χρησιμοποιούν αλγόριθμους Βαθιάς Μηχανικής Μάθησης, οι οποίοι επιτρέπουν την ανίχνευση και αναγνώριση σε ένα στάδιο. Με αυτή την ανίχνευση και αναγνώριση ενός σταδίου δεν υπάρχει η ανάγκη για εξαγωγή χαρακτηριστικών σημείων από έναν άνθρωπο-προγραμματιστή, αλλά οι αλγόριθμοι μαθαίνουν μέσα από τα δεδομένα. Σε αντίθεση βέβαια με τις μεθόδους της κλασικής προσέγγισης, είναι απαραίτητος εδώ ο μεγάλος όγκος των δεδομένων αλλά και η αντίστοιχη υπολογιστική ισχύς. Παράλληλα, στους αλγόριθμους Βαθιάς Μάθησης τίθεται το θέμα της επίλυσης πολύ πιο σύνθετων προβλημάτων.

Επιπλέον, αναφέρθηκε πως αρκετές μέθοδοι ανίχνευσης και αναγνώρισης σημάτων οδικής κυκλοφορίας βασίζονται και στην τεχνολογία σαρωτών laser LiDAR. Η συσχέτιση δεδομένων LiDAR και κάμερας είναι ένας πολλά υποσχόμενος τρόπος για την ανίχνευση και αναγνώριση των πινακίδων κυκλοφορίας, όμως λόγω της ιδιαιτερότητας του 3D νέφους σημείων οι περισσότερες από τις παραδοσιακές μεθόδους Βαθιάς Μάθησης είναι δύσκολο να εφαρμοστούν άμεσα στο πρόβλημα ανίχνευσης με νέφος σημείων. Σε αντίθεση με τις μεθόδους βαθιάς μάθησης, οι διαδικασίες προεπεξεργασίας του νέφους σημείων συνήθως εξακολουθούν να βασίζονται σε χειροκίνητα σχεδιασμένα μοντέλα και κατώφλια, τα οποία μπορεί να οδηγήσουν σε χαμηλή ικανότητα γενίκευσης.

Στα πλαίσια της μεταπτυχιακής διπλωματικής εργασίας εφαρμόστηκε ένας από τους αλγορίθμους State-of-the-Art, ο αλγόριθμος YOLO. Συγκεκριμένα, υλοποιήθηκε η εκπαίδευση και αξιολόγηση των αλγορίθμων YOLOv3 και YOLOv3-tiny, οι οποίοι εκπαιδεύτηκαν σε Γερμανικό σύνολο δεδομένων και τελικά εφαρμόστηκαν και σε πραγματικά πειραματικά δεδομένα που ελήφθησαν στο πλαίσιο της εργασίας για το σκοπό αυτό. Ο αλγόριθμος YOLOv3-tiny, αν και ταχύτερος συγκριτικά με τον YOLOv3, πετυχαίνει

χαμηλότερα ποσοστά ανίχνευσης και αναγνώρισης. Θα μπορούσε να χρησιμοποιηθεί σε εφαρμογές που επιτακτική ανάγκη είναι η ταχύτητα και όχι τόσο η ακρίβεια του αλγορίθμου.

Αδιαμφισβήτητα ο αλγόριθμος YOLOv3, αλλά και όλοι οι αλγόριθμοι της οικογένειας YOLO με τελευταίο τον YOLOv4, είναι οι ταχύτεροι ανιχνευτές αντικειμένων που υπάρχουν αυτή τη στιγμή, και κυρίως λύσεις σε πραγματικό χρόνο, επιτυγχάνοντας μάλιστα υψηλά ποσοστά ανίχνευσης και αναγνώρισης. Εξίσου αδιαμφισβήτητη είναι η ολοένα αυξανόμενη τάση δημιουργίας εφαρμογών και συστημάτων για την υποβοήθηση της οδήγησης (Advanced Driver Assistance Systems - ADAS) ή για την πλήρως αυτοματοποιημένη οδήγηση. Τα δύο παραπάνω συντελούν στο συμπέρασμα που καθιστά τους αλγορίθμους YOLO ιδανικούς για τη χρήση τους σε τέτοιου είδους εφαρμογές και συστήματα.

Δεν είναι πολύ μακριά στο μέλλον η στιγμή που τα πλήρως αυτόνομα οχήματα θα διατεθούν στο ευρύ κοινό. Τα αυτόνομα αυτοκίνητα είναι η επόμενη σελίδα της σύγχρονης αυτοκίνησης, και οι εταιρείες κινούνται ήδη προς αυτή την κατεύθυνση μαζί με πλήθος ερευνητών, οι οποίοι καλούνται να βελτιώσουν τους αλγορίθμους Βαθιάς Μηχανικής Μάθησης, ώστε να μπορούν να χρησιμοποιηθούν σε εφαρμογές πραγματικού χρόνου.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] Paclik P. (1999). <http://euler.fd.cvut.cz/research/rs2/files/skoda-rs-survey.html>
- [2] Akatsuka H. and Imai S. (1987), Road signposts recognition system. *Proc. SAE Vehicle Highway Infrastructure: Safety Compatibility*, pp. 189-196.
- [3] Βαρυτιμιδης, Χ.Ι., (2008), *Ανίχνευση Αντικειμένων και Ημιαυτόματος Χαρακτηρισμός Εικόνων*. Διπλωματική Εργασία, Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών, Εθνικό Μετσόβιο Πολυτεχνείο.
- [4] Oren M., Papageorgiou C., Sinha P., Osuna E. and Poggio T. (1997). Pedestrian detection using wavelet templates. *IEEE Proc. Computer Vision and Pattern Recognition (CVPR)*.
- [5] Papageorgiou C., Oren M. and Poggio T. (1998), A general framework for object detection. *IEEE Proc. International Conference on Computer Vision (ICCV)*
- [6] Lienhart R. and Maydt J. (2002), An extended set of Haar-like features for rapid object detection. *IEEE Proc. International Conference on Image Processing (ICIP)*, pp. 900-903.
- [7] OpenCV. <https://opencv.org/>
- [8] Dalal N. Triggs B. (2005), Histograms of oriented gradients for human detection. *IEEE Proc. Computer Vision and Pattern Recognition (CVPR)*, vol. 1, pp 886-893.
- [9] Μπράχου Ε. (2010), *Μέθοδοι Ταξινόμησης Εικόνας*. Μεταπτυχιακή εργασία, Πανεπιστήμιο Ιωαννίνων.
- [10] Αδάμ Κ. (2014), *Αυτόματος Εντοπισμός και Αναγνώριση Σημάτων Οδικής Κυκλοφορίας*. Μεταπτυχιακή εργασία ΣΑΤΜ, Εθνικό Μετσόβιο Πολυτεχνείο.
- [11] Escalera S., Baró, X., Pujol, O., Vitrià, J., Radeva, P. (2011), Background on traffic sign detection and recognition. *Traffic-Sign Recognition Systems, Springer*, pp. 5-13.
- [12] Stein Fridtjof, Cleopatra project at PAC (Parallel Application Center) <http://www.pac.soton.ac.uk/>
- [13] Sharma N., Mishra M., Shrivastava M. (2012). Colour image segmentation techniques and issues: an approach. *International Journal of Scientific & Technology Research*, 1(4), pp. 9-12.
- [14] Dass R., Priyanka, Devi S., (2012), Image segmentation techniques. *International Journal of Electronics & Communication Technology*, 3(1), pp. 66-70.
- [15] Gurjinder K., Randeep K., Mandeep S., Rishamjot K.D. (2016), Image segmentation methods and techniques: a review. *International Journal of Engineering Science and Computing*, 6(5), pp. 6096-6099.
- [16] Sezgin M. (2004), Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging*, 13(1), pp. 146-168.
- [17] Otsu N. (1975), A threshold selection method from gray-level histograms. *Automatica*, 11, pp. 23-27.
- [18] Baró X., Vitrià J., (2012), *Traffic Sign Detection on Greyscale Image*, Report, Centre de Visió per Computador & Dept. Informàtica UAB, Barcelona, Spain
- [19] Brkić K. (2010), *An Overview of Traffic Sign Detection Methods*. Dept. of Electronics, Microelectronics, Computer and Intelligent Systems, Faculty of Electrical Engineering and Computing, Zagreb, Croatia.

- [20] Shadeed W.G., Abu-Al-Nadi D.I., Mismar M.J, (2003), Traffic road sign detection in colored images, *Proc. 10th IEEE International Conference on Electronics, Circuits, and Systems (ICECS)*, pp. 890-893
- [21] Benallal M., Meunier J. (2003), Real-time color segmentation of road signs. *Proc. Canadian Conference on Electrical and Computer Engineering*, vol. 3, pp. 1823-1826.
- [22] Broggi A., Cerri P., Medici P., Porta P.P., & Ghisio G. (2007). Real time road signs recognition. *Proc. IEEE Intelligent Vehicles Symposium*, pp.981-986.
- [23] Buluswar S.D., (2002), *Color-Based Models for Outdoor Machine Vision*. Ph.D. Dissertation, University of Massachusetts, Amherst.
- [24] Andrey V., and Kang H.J., (2006), Automatic detection and recognition of traffic signs using geometric structure analysis. *SICE-ICASE International Joint Conference*, pp. 1451-1456.
- [25] Wali S.B., Hannan M.A., Hussain A., Samad A.S., (2015), An automatic traffic sign detection and recognition system based on colour segmentation, shape matching, and SVM. *Mathematical Problems in Engineering*, vol. 2015, no. 250461.
- [26] Yuille A., Snow D., and Nitzberg M., (1998), Sign finder, using colour to detect, localize and identify informational. *Proc. 6th International Conference on Computer Vision*.
- [27] Vitabile S., Gentile A., and Sorbello F. (2002), A neural network based automatic road sign recognizer. *International Joint Conference on Neural Networks, Honolulu, USA*.
- [28] Wang Y.Q., Liu L.M., Yang Z. (2009), Traffic sign detection based on fixed color combination and intensity restraint. *Proc. IEEE International Symposium on Computer Network and Multimedia Technology*, pp. 1-5.
- [29] Chen Y., Xie Y., Wang Y., (2013), Detection and recognition of traffic signs based on HSV vision model and shape features. *Journal of Computers*, 8(5), pp. 1366-1370
- [30] Rashmi H, Prashanth Kumar G, Shashidhar M, (2014), Automatic Tracking of Traffic Signs Based on HSV, *INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT)* Volume 03
- [31] Maldonado-Bascon S., Lafuente-Arroyo S., Gil-Jimenez P., Gomez-Moreno H., Lopez-Ferreras F (2007), Road-sign detection and recognition based on support vector machines. *IEEE Transactions on Intelligent Transportation Systems*, 8(2), pp.264-278
- [32] Siogkas, G.K., Dermatas, E.S. (2006), Detection, tracking and classification of road signs in adverse conditions. *Proc. IEEE Mediterranean Electrotechnical Conference (MELECON)*, pp. 537-540.
- [33] Eichner M.L., & Breckon T.P. (2008), Integrated speed limit detection and recognition from real-time video. *Proc. IEEE Intelligent Vehicles Symposium*, pp. 626-631.
- [34] Zakir U. (2011), *Automatic Road Sign Detection and Recognition*. Doctoral Dissertation, Loughborough University, United Kingdom.
- [35] Hough P.V.C. (1962), Methods and means for recognizing complex patterns, *U. S. Patent 3069654*.
- [36] Duda R.O., Hart P.E. (1972), Use of the Hough transformation to detect lines and curves in pictures. *Communications of the ACM*, 15(1), pp. 11-15.
- [37] Κεσιδης. Α. (2001), *Αντίστροφος Μετασχηματισμός Hough και οι Εφαρμογές του στην Ψηφιακή Επεξεργασία Εικόνας*, Διδακτορική Διατριβή, Δημοκρίτειο Πανεπιστήμιο Θράκης (ΔΠΘ).

- [38] Garcia Garrido M.Á., Sotelo M.Á. & Martín-Gorostiza E. (2005), Fast road sign detection using Hough transform for assisted driving of road vehicles. *Computer Aided Systems Theory - EUROCAST 2005*, LNCS 3643, Springer, pp. 543-548.
- [39] Piccioli G., De Micheli E., Parodi P., Campani M. (1996), Robust method for road sign detection and recognition. *Image and Vision Computing*, 14(3), pp. 209-223.
- [40] Loy G., Barnes N. (2004), Fast shape-based road sign detection for a driver assistance system, *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 70-75.
- [41] https://en.wikipedia.org/wiki/Template_matching
- [42] Adelson E.H., Anderson C.H., Bergen J.R., Burt P.J. and Ogden J.M. (1983), Pyramid methods in image processing, http://web.mit.edu/persci/people/adelson/pub_pdfs/RCA84.pdf
- [43] Gavrilu D. (1999), Traffic sign recognition revisited, *DAGM-Symposium*, pp. 86-93.
- [44] Hsu S.H., Huang C. L. (2001), Road sign detection and recognition using matching pursuit method. *Image and Vision Computing*, 19(3), pp. 119-129.
- [45] Belaroussi R., Foucher P., Tarel J.-P., Soheilian B., Charbonnier P., Paparoditis N. (2010), Road sign detection in images: a case study, *20th International Conference on Pattern Recognition*.
- [46] Fischler M.A., Bolles R. C. (1981), Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6), pp. 381-395.
- [47] Szeliski R. (2010), *Computer Vision: Algorithms and Applications*, draft c, Springer.
- [48] Arlicot A., Soheilian B. & Paparoditis N. (2009), Circular road sign extraction from street level images using colour, shape and texture databases maps. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXVIII, Part 3/W4, pp. 205-210.
- [49] Kim S.U., Lee J.W. (2016), Traffic sign recognition, and tracking using RANSAC-based motion estimation for autonomous vehicles. *Journal of Institute of Control*, 22(2), pp. 110-116
- [50] Γεωργούλη Κ. (2015), *Τεχνητή Νοημοσύνη*, © ΣΕΑΒ – Αποθετήριο «ΚΑΛΛΙΠΟΣ»
- [51] Viola P., Jones M. (2001), Rapid object detection using a boosted cascade of simple features. *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*
- [52] Berger M., Forechi A., De Souza A.F., de Oliveira Neto J., Veronese L., Badue C. (2012), Traffic sign recognition with VG-RAM Weightless Neural Networks, *12th International Conference on Intelligent Systems Design and Applications (ISDA)*.
- [53] Xiong C., Wang C., Ma W., Shan Y. (2016), A traffic sign detection algorithm based on deep convolutional neural network, *Proc. IEEE International Conference on Signal and Image Processing (ICSIP)*.
- [54] Dewan P., Vig R., Shukla N., Das B.K. (2017), An overview of traffic signs recognition methods. *International Journal of Computer Applications*, 168(11), pp. 7-11.
- [55] Mukhometzianov R. and Wang Y. (2017), Review. Machine learning techniques for traffic sign detection, *Computer Science, ArXiv*.
- [56] Wang Y.-Q. (2014), *An Analysis of the Viola-Jones Face Detection Algorithm*, Image Processing On Line (IPOL), <http://dx.doi.org/10.5201/ipol.2014.104>
- [57] Μαρινέλλης Γ. (2018), *Αυτόματη Προσθήκη Επισημάνσεων για Πρόσωπα και Αντικείμενα σε Πολυμεσικό Περιεχόμενο*, Διπλωματική Εργασία, Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών, Εθνικό Μετσόβιο Πολυτεχνείο.

- [58] Pavani S.-K., Delgado D., Frangi A.F. (2010), Haar-like features with optimally weighted rectangles for rapid object detection. *Pattern Recognition*, 43, pp.160-172.
- [59] Escalera S., Radeva P. (2012), Fast greyscale road sign model matching and recognition, *ResearchGate*
- [60] Jeon W.J., Ruiz Sanchez G.A., Lee T., Choi Y., Woo B., Lim K., Byun H. (2014), Real-time detection of speed limit traffic signs on the real road using Haar-like features and boosted cascade, *Proc 8th International Conference on Ubiquitous Information Management and Communication*
- [61] Das S., Mohammadi M. (2015), *Mobile Traffic Sign Recognition*. Stanford University.
- [62] Prisacariu V.A., Timofte R., Zimmermann K., Reid I., Van Gool L. (2010), Integrating object detection with 3D tracking towards a better driver assistance system. *Proc. IEEE 20th International Conference on Pattern Recognition (ICPR)*, pp. 3344–3347.
- [63] Baro X., Escalera S., Vitri J., Pujol O., Radeva P. (2009), Traffic sign recognition using evolutionary adaboost detection and forest-ECOC classification. *IEEE Transactions on Intelligent Transportation Systems*, 10(1), pp. 113–126.
- [64] Balas B.J. and Sinha P. (2003), STICKS: Image-representation via non-local comparisons, *Journal of Vision*, vol. 3, no. 9.
- [65] De la Escalera A., Armingol J.M., Mata M. (2003), Traffic sign recognition and analysis for intelligent vehicles. *Image and Vision Computing*, 21(3), pp. 247-258.
- [66] Kaplan K., Kurtul C. and Akin H.L., (2012), Real-time traffic sign detection and classification method for intelligent vehicles, *Proc. IEEE International Conference on Vehicular Electronics and Safety*, Istanbul, Turkey.
- [67] Chen T., Lu S. (2016), Accurate and efficient traffic sign detection using discriminative Adaboost and support vector regression. *IEEE Transactions on Vehicular Technology*, 65(6), pp. 4006-4015.
- [68] Hoferlin B., Zimmermann K. (2009), Towards reliable traffic sign recognition. *Proc. IEEE Intelligent Vehicles Symposium*, pp. 324-329.
- [69] Zaklouta F., Stanculescu B. (2011), Warning traffic sign recognition using a HOG-based Kd tree. *Proc. IEEE intelligent Vehicles Symposium*, pp. 1019-1024.
- [70] Moreno H.G., Bascón S.M., Jiménez P.G. and Arroyo S.F. (2010), Goal evaluation of segmentation algorithms for traffic sign recognition, *IEEE Transaction on Intelligent Transportation Systems*, 11(4), pp. 917-930.
- [71] Fang C.Y., Chen S.W., Fuh C.S. (2003), Road-sign detection and tracking. *IEEE Transactions on Vehicular Technology*, 52(5), pp. 1329-1341.
- [72] Deguchi et al. (2011), Intelligent traffic sign detector: Adaptive learning based on online gathering of training samples. *IEEE IV Symp., 2011*, pp. 72–77.
- [73] Bahlmann C., Zhu Y., Ramesh V., Pellkofer M., Koehler T. (2005), A system for traffic sign detection, tracking and recognition using color, shape, and motion information. *Proc. IEEE Intelligent Vehicles Symposium*, pp. 255-260.
- [74] Brkić K., Pinz A. and Segvić S., (2009) Traffic sign detection as a component of an automated traffic infrastructure inventory system, *Visual Learning*, OCG Schriftenreihe, vol. 254, pp. 129-140.
- [75] Lopez-Montiel M., Rubio Y., Sánchez-Adame M., Orozco-Rosas U. (2019), Evaluation of algorithms for traffic sign detection, *Optics and Photonics for Information Processing XIII, Proc. SPIE*, vol. 11136

- [76] Tam T. Le, Son T. Tran, Seichii Mita, and Thuc D. Nguyen (2010), Real time traffic sign detection using color and shape-based features. *Asian Conference on Intelligent Information and Database Systems*, Springer, pp. 268-278
- [77] L. Chen, Q. Li, M. Li, and Q. Mao (2011), Traffic sign detection and recognition for intelligent vehicle. *IEEE Intelligent Vehicles Symposium (IV)*, pp. 908–913
- [78] Καλησπεράκης Η. (2010), *Η Επιπολική Γεωμετρία στον Προβολικό και τον Ευκλείδειο Χώρο*. Διδακτορική Διατριβή, ΣΑΤΜ, ΕΜΠ
- [79] Lowe D. (2004), Distinctive image features from scale invariant key points. *International Journal of Computer Vision*, 2(60), pp. 91-110.
- [80] Κουσκουρίδας Ρ., Μπελαγιάννης Β., Χρυσοστόμου Δ., Γαστεράτος Α, (2009), Συγκριτική μελέτη αλγορίθμων εξαγωγής χαρακτηριστικών, 1^ο Πανελλήνιο Συνέδριο Ρομποτικής, ΤΕΕ, Αθήνα.
- [81] Bay H., Tuytelaars T., Van Gool L. (2006), "Surf: Speeded up robust features", *Proc. European Conference on Computer Vision*, Springer, pp. 404-417.
- [82] Zhao J., Zhu S., and Huang X. (2013), Real-time traffic sign detection using SURF features on FPGA, *IEEE High Performance Extreme Computing Conference (HPEC)*.
- [83] Vapnik V., Kotz S. (2006). *Estimation of Dependences Based on Empirical Data*. Springer.
- [84] Καμπουράζος Β. (2015), *Εφαρμογές της Υπολογιστικής Νοημοσύνης*. ΤΕΙ Ανατολικής Μακεδονίας και Θράκης, © Αποθετήριο «ΚΑΛΛΙΠΟΣ»
- [85] Burges C. J. (1998), A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2), pp. 121-167.
- [86] Chang K., Swinson T. (2011), *Using Support Vector Machines to Classify Traffic Signs for an Autonomous Vehicle*, www.academia.edu.
- [87] Hua X., Zhua X., Lia D., Li H., (2010), *Traffic sign recognition using scale invariant feature transform and SVM*. *Special Joint Symposium of ISPRS Technical Commission IV and AutoCarto in Conjunction with ASPRS/CaGIS Fall specialty conference*, pp. 15–19.
- [88] Kumaraswamy R., Prabhu L.V., Suchithra K., Sreejith Pai P.S., (2011), SVM based classification of traffic signs for realtime embedded platform, *International Conference on Advances in Computing and Communications*, Spinger, pp. 339-348.
- [89] Houben S., Stallkamp J., Salmen J., Schlipsing M., Igel C., (2013), Detection of traffic signs in real-world images: The German Traffic Sign Detection Benchmark. *Proc. IEEE International Joint Conference on Neural Networks (IJCNN)*.
- [90] Yakimov P.Y., (2016), Real-time road signs recognition using mobileGPU. In: *CEUR Workshop Proceedings*, vol. 1638, pp. 477-483.
- [91] Saadna Y., Behloul A. (2017), An overview of traffic sign detection and classification methods. *International Journal of Multimedia Information Retrieval*, 6, pp. 193–210.
- [92] Berkaya S.K., Gunduz H., Ozsen O., Akinlar C., Gunal S. (2016), On circular traffic sign detection and recognition. *Expert Systems with Applications*, 48, pp. 67–75.
- [93] Mandeljc R., Tabernik D., Kristan M., Skocaj D., (2015), *Traffic Sign Classification with Batch and On-Line Linear Support Vector Machines*. Visual Cognitive Systems Laboratory, Faculty of Computer and Information Science, University of Ljubljana.
- [94] Fan R.-E., Chang K.-W., Hsieh C.-J., Wang X.-R., Lin C.-J. (2008), Liblinear: A library for large linear classification. *The Journal of Machine Learning Research*, 9, pp. 1871–1874.

- [95] Crammer K. and Singer Y., (2002), On the algorithmic implementation of multiclass kernel-based vector machines. *The Journal of Machine Learning Research.*, 2, pp. 265–292.
- [96] Bordes A., Bottou L., Gallinari P., Weston J. (2007), Solving multiclass support vector machines with LaRank. *Proc. 24th International Conference on Machine Learning (ICML)*, pp. 89–96.
- [97] Greenhalgh J. and Mehdi M., (2012), Real-time detection and recognition of road traffic signs, *IEEE Transactions on Intelligent Transportation Systems*, 13(4), pp. 1498-1506.
- [98] Chaiyakhan K., Hirunyanakul A., Chanklan R., Kerdprasop K., Kerdprasop N. (2015), Traffic sign classification using support vector machine and image segmentation, *International Conference on Industrial Application Engineering*.
- [99] Moutarde F., Bargeton A., Herbin A. and Chanussot L., (2007), Robust on vehicle real-time visual detection of American and European speed limit signs with a modular Traffic Signs Recognition system, *Proc. IEEE Intelligent Vehicles Symposium*, pp. 1122-1126
- [100] Medici P., Caraffi C., Cardarelli E., Porta P.P., Ghisio G. (2008), Real time road signs classification. *Proc. IEEE International Conference on Vehicular Electronics and Safety*, pp. 253-258.
- [101] Souani C., Faieth H. and Besbes K., (2013), Efficient algorithm for automatic road sign recognition and its hardware implementation, *Journal of Real Time Image Processing*, 9, pp. 79–93.
- [102] Abedin M.Z., Dhar P., Deb K. (2016), Traffic sign recognition using SURF: speeded up robust feature descriptor and artificial neural network classifier. *Proc. IEEE 9th International Conference on Electrical and Computer Engineering (ICECE)*, pp. 198-201.
- [103] Schmidhuber J. (2015), Deep learning in neural networks: an overview. *Neural Networks*. 61, pp. 85-117.
- [104] Παπαδόπουλος Α. (2018), *Πρόβλεψη Τροχιών σε Δεδομένα Κίνησης με Βαθιά Νευρωνικά Δίκτυα*, Μεταπτυχιακή Διπλωματική Εργασία, ΠΜΣ Εφαρμοσμένη Στατιστική, Πανεπιστήμιο Πειραιώς,
- [105] Liu C., Li S., Chang F., Wang Y. (2019), Machine vision based traffic sign detection methods: review, analyses and perspectives, *IEEE Access*, 99: DOI: 10.1109/ ACCESS. 2019.292494
- [106] Wu Y., Liu Y., Li J., Liu H., and Hu X., (2013), Traffic sign detection based on convolutional neural networks, *Proc. International Joint Conference on Neural Networks*, Dallas, USA, pp. 1-7.
- [107] Aghdam H.H., Heravi E.J., Puig D., (2016), A practical approach for detection and classification of traffic signs using Convolutional Neural Networks. *Robotics and Autonomous Systems*, 84, pp. 97–112.
- [108] Zang D., Zhang J., Zhang D., Bao M., Cheng J., Tang K., (2016), Traffic sign detection based on cascaded convolutional neural networks. *Proc. 17th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, pp. 201–206.
- [109] Xie K., Ge S., Ye Q., Luo Z., (2016), Traffic sign recognition based on attribute-refinement cascaded convolutional neural networks. *Proc. Pacific Rim Conference on Multimedia*, Springer, pp. 201–210
- [110] Qian R., Yue Y., Coenen F., Zhang B. (2016), Traffic sign recognition with convolutional neural network based on max pooling positions. *Proc. IEEE 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, pp. 578–582.
- [111] Qian R., Zhang B., Yue Y., Wang Z., Coenen F. (2015), Robust Chinese traffic sign detection and recognition with deep convolutional neural network, *Proc. 11th International Conference on Natural Computation (ICNC)*.
- [112] Islam T. (2019), Traffic sign detection and recognition based on convolutional neural networks, *Proc. International Conference on Advances in Computing, Communication and Control (ICAC3)*.

- [113] Alghmghama D.A., Latif G., Alghazo J., Alzubaidi L., (2019), Autonomous traffic sign (ATSR) detection and recognition using deep CNN, *Proc. 16th International Learning & Technology Conference*.
- [114] Cao J., Song C., Peng S., Xiao F., Song S., (2019), Improved traffic sign detection and recognition algorithm for intelligent vehicles, *Sensors*, 19, 4021.
- [115] Liang Z., Shao J., Zhang D., Gao L., (2019), Traffic sign detection and recognition based on pyramidal convolutional networks, *Neural Computing and Applications*, 32, pp. 6533–6543
- [116] Girshick R., Donahue J., Darrell T., Malik J., (2014), Rich feature hierarchies for accurate object detection and semantic segmentation. *Proc. IEEE International Conference on Computer Vision (ICCV)*.
- [117] Girshick R. (2015), Fast R-CNN, *IEEE International Conference on Computer Vision (ICCV)*
- [118] Ren S., He K., Girshick R., Sun J., (2015), Faster R-CNN: towards real-time object detection with region proposal networks, *Proc. 28th International Conference on Neural Information Processing Systems*, vol. 1, pp. 91–99
- [119] Xiong C., Wang C., Ma W. and Shan Y., (2016), A traffic sign detection algorithm based on deep convolutional neural network, *IEEE International Conference on Signal and Image Processing (ICSIP)*.
- [120] Zuo Z., Yu K., Zhou Q., Wang X. and Li T., (2017), Traffic Signs Detection Based on Faster RCNN, *Proc. IEEE 37th International Conference on Distributed Computing Systems Workshops*.
- [121] Rehlaender P., Schroeder M., Chadha G., Schwung A., (2019), Traffic sign detection using R-CNN, *Proc. INNS Big Data and Deep Learning Conference*.
- [122] Shao F., Wang X., Meng F., Zhu J., Wang D. Dai J., (2019), Improved Faster R-CNN traffic sign detection based on a second region of interest and highly possible regions proposal network, *Sensors*, 19(10):2288.
- [123] Nassar A.S., Lefèvre S., (2019), Automated mapping of accessibility signs with deep learning from ground-level imagery and open data, *Proc. Joint Urban Remote Sensing Event (JURSE)*, pp. 1-4.
- [124] Redmon J., Divvala S., Girshick R., and Farhadi A., (2016), You only look once - unified, real-time object detection, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779–788.
- [125] Møgelmoose A., Trivedi M.M., Moeslund T.B. (2012), Vision-based traffic sign detection and analysis for intelligent driver assistance systems: perspectives and survey, *IEEE Transactions on Intelligent Transportation Systems*, 13(4), pp. 1484 – 1497.
- [126] Redmon J. and Farhadi A., (2016), YOLO9000: better, faster, stronger, *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*
- [127] Redmon J., Farhadi A., (2018), Yolov3: An incremental improvement. *arXiv 2018, arXiv:1804.02767v1*.
- [128] Bochkovskiy A., Wang C.-Y., Liao H.-Y. M., (2020), YOLOv4: optimal speed and accuracy of object detection, *arXiv:2004.10934v1 [cs.CV]*
- [129] Vikram M., (2018), *A YOLO Based Approach for Traffic Sign Detection*, Project Report, Dept. of Information Technology National Institute of Technology Karnataka, Surathkal
- [130] Wang C.Y., Royce C.-Y., (2016), *Traffic Sign Detection using You Only Look Once Framework*, Report, Stanford University.

- [131] Saleh S., Khwandah S.A., Mumtaz A., Heller A., Hardt W., (2019), Traffic signs recognition and distance estimation using a monocular camera, *Proc. 6th International Conference Actual Problems of System and Software Engineering*, pp. 407-418.
- [132] Rajendran S.P., Shine L., Pradeep R., Vijayaraghavan S., (2019), Real-time traffic sign recognition using YOLOv3 based detector, *Proc. 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*.
- [133] Hu Z., Tang J., Wang Z., Zhang K., Zhang L., Sun Q., (2018), Deep learning for image-based cancer detection and diagnosis – A survey, *Pattern Recognition*, 83, pp. 134-149.
- [134] Nassif A.B., Shahin I., Attili I., Azzeh M., Shaalan K. (2019), *Speech recognition using deep neural networks: a systematic review*, *IEEE Access*, 7, pp. 19143- 19165.
- [135] Efthymiou N., Koutras P., Filntisis P.P., Potamianos G., Maragos P., (2018), Multi-view fusion for action recognition in child-robot interaction, *Proc. 25th IEEE International Conference on Image Processing (ICIP)*, pp. 455-459.
- [136] Μελέτης Π., (2015), *Τεχνικές Βαθιάς Μάθησης και Εφαρμογές στην Ανίχνευση Προσώπων σε Εικόνες*, Διπλωματική Εργασία, Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών, Εθνικό Μετσόβιο Πολυτεχνείο.
- [137] Παπαδόπουλος Α. (2016), *Συνελικτικά Νευρωνικά Δίκτυα στην Όραση Υπολογιστών*, Μεταπτυχιακή Εργασία, Δ.Π.Μ.Σ. «Ηλεκτρονική και Επεξεργασία της Πληροφορίας», Πανεπιστήμιο Πατρών,
- [138] *The Traffic German Sign Detection Benchmark – TGSDB*
<http://benchmark.ini.rub.de/?section=gtsdb&subsection=news>
- [139] *Ultralytics: YOLOv3 in PyTorch free code on GitHub*
https://github.com/ultralytics/yolov3?fbclid=IwAR0JtmcspvGFbJXho_o6KoV0ukw_1eoyYwLSrWxMcP5P_thyOog5ALA151U#readme
- [140] Lin T.-Y., Dollar P., Girshick R., He K., Hariharan B., and Belongie S., (2017), Feature pyramid networks for object detection, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2117–2125.
- [141] Liu C., Wang Y. (2019), Machine vision based traffic sign detection methods: review, analyses and perspectives, *IEEE Access*, vol.7, pp. 86578-86596.
- [142] Yu Y., Li J., Wen C., Guan H., Luo H., and Wang C., (2016), Bag-of-visual-phrases and hierarchical deep models for traffic sign detection and recognition in mobile laser scanning data, *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 113, pp. 106–123.
- [143] Pei D., Sun F., and Liu H., (2013), Supervised low-rank matrix recovery for traffic sign recognition in image sequences, *IEEE Signal Processing Letters*, 20(3), pp. 241–244.
- [144] Riveiro B., Díaz-Vilariño L., Conde-Carnero B., Soilán, M. and Arias P., (2017), Automatic segmentation and shape-based classification of retro-reflective traffic signs from mobile LiDAR data, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(1), pp. 295–303.
- [145] Deng Z. and Zhou L., (2018). *Detection and recognition of traffic planar objects using colorized laser scan and perspective distortion rectification*, *IEEE Transactions on Intelligent Transportation Systems*, 19(5), pp. 1485–1495.
- [146] Arcos-García A., Soilán M., Álvarez-García J.A., and Riveiro B., (2017), *Exploiting synergies of mobile mapping sensors and deep learning for traffic sign recognition systems*”, *Expert SystEms and Applications*, 89(15), pp. 286–295.
- [147] Guan H., Yan W., Yu Y., Zhong L., and Li D., (2018), *Robust traffic-sign detection and classification using mobile LiDAR data with digital images*”, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(5), pp. 1715–172.

- How do you feel about your master degree?
 - Like looking for dinosaurs.

