ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
UNIVERSITY OF PIRAEUS

DEMOKRITOS

Ανίχνευση διαφορετικών αντικειμένων για αυτόνομες εφαρμογές οδήγησης

Από

Κούγκουλα Μαγδαληνή

Υποβάλλεται

για την εκπλήρωση των προϋποθέσεων λήψης

Μεταπτυχιακού Διπλώματος

στην «Τεχνητή Νοημοσύνη»

στο

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

Συγγραφέας Κούγκουλα Μαγδαληνή

ΔΠΜΣ «Τεχνητή Νοημοσύνη»

Μήνας/2022

Έγινε αποδεκτό από:

Μιχαήλ Φιλιππάκης

Αναπληρωτής Καθηγητής
Επιβλέπων

Έγινε αποδεκτό από:

Ηλίας Μαγκλογιάννης
Καθηγητής
Μέλος Εξεταστικής Επιτροπής

Έγινε αποδεκτό από:

Μαρία Χαλκίδη
Αναπληρώτρια Καθηγήτρια
Μέλος Εξεταστικής Επιτροπής

Ανίχνευση διαφορετικών αντικειμένων για αυτόνομες εφαρμογές οδήγησης

Από

Κούγκουλα Μαγδαληνή

Υποβλήθηκε στο ΔΠΜΣ «Τεχνητή Νοημοσύνη» την ΧΧ Μηνός 20ΧΧ
ως υποχρέωση για την λήψη Μεταπτυχιακού Διπλώματος Σπουδών

Abstract

The aim of this dissertation is to use real-time video to locate and classify distinct motion objects. Two ways were employed and compared to achieve this. The Berkeley DeepDrive dataset was used to train the two YOLO and Faster R-CNN models so that they could compare their performance and create a similar mAP table as well as matching diagrams of normalized total loss and average accuracy (mAP). Then, with a focus on autonomous driving and attempting to compare the models' performance, brief FPS and mAP measurement movies were generated.

Περίληψη

Στόχος της παρούσας διπλωματικής είναι ο εντοπισμός και η ταξινόμηση διάφορων αντικειμένων κίνησης μέσα από βίντεο σε πραγματικό χρόνο. Για να επιτευχθεί αυτό χρησιμοποιήθηκαν και συγκρίθηκαν δύο μοντέλα. Αρχικά πραγματοποιήθηκε η εκπαίδευση των δύο μοντέλων YOLO και Faster R-CNN στο σύνολο δεδομένων Berkeley DeepDrive έτσι ώστε να μπορέσουν να συγκριθούν οι επιδόσεις τους και να λάβουμε σαν αποτέλεσμα ένα συγκρίσιμο πίνακα mAP καθώς και αντίστοιχα διαγράμματα της ομαλοποιημένης συνολικής απώλειας και της μέσης ακρίβειας(mAP) . Έπειτα δόθηκε ιδιαίτερη έμφαση στο πλαίσιο της αυτόνομης οδήγησης και στην προσπάθεια σύγκρισης των επιδόσεων των μοντέλων δημιουργήθηκαν βίντεο μέτρησης FPS και mAP σε πραγματικό χρόνο.

Επιβλέπων: Φιλιππάκης Μιχαήλ
Ακαδημαϊκή Θέση: Αναπληρωτής Καθηγητής

ΠΕΡΙΕΧΟΜΕΝΑ

Λίστα Πινάκων

Λίστα Εικόνων

Κεφάλαιο 1

Εισαγωγή

Η υπολογιστική όραση είναι ένας καθολικός τομέας που έχει προσελκύσει την επιστημονική κοινότητα τις τελευταίες δεκαετίες με τις εφαρμογές των αυτοοδηγούμενων οχημάτων να βρίσκονται στο επίκεντρο της έρευνας. Ένας επίσης σημαντικός κλάδος της υπολογιστικής όρασης είναι η ανίχνευση αντικειμένων. Η ανίχνευση αντικειμένων διαδραματίζει σημαντικό ρόλο στην ανίχνευση άλλων οχημάτων, στην εκτίμηση του περιβάλλοντος καθώς και στην παρακολούθηση. Η διαφορά που παρουσιάζεται ανάμεσα στους αλγορίθμους ανίχνευσης αντικειμένων και στους αλγόριθμους ταξινόμησης είναι ότι στους αλγόριθμους ανίχνευσης, υπάρχει η δυνατότητα σχεδίασης ενός πλαισίου οριοθέτησης γύρω από το αντικείμενο ενδιαφέροντος για να πραγματοποιηθεί η ανίχνευση μέσα στην εικόνα. Επίσης στο ίδιο πλαίσιο ανίχνευσης αντικειμένου μπορεί να υπάρχουν πολλά πλαίσια οριοθέτησης που αντιπροσωπεύουν διαφορετικά αντικείμενα ενδιαφέροντος, όμως δεν μας δίνεται η δυνατότητα να γνωρίζουμε εκ των προτέρων των αριθμό των πλαισίων.

Μία λύση στο πρόβλημα αυτό θα ήταν η δημιουργία ενός τυπικού συνελεκτικού δικτύου ακολουθούμενο από ένα πλήρως συνδεδεμένο επίπεδο. Όμως σύμφωνα με αυτή την προσέγγιση το μήκος του επιπέδου εξόδου είναι μεταβλητό και όχι σταθερό, γεγονός που συμβαίνει επειδή ο αριθμός των εμφανίσεων των αντικειμένων ενδιαφέροντος δεν είναι καθορισμένος. Για να επιλυθεί αυτό το πρόβλημα θα μπορούσε να γίνει λήψη διαφορετικών περιοχών ενδιαφέροντος από την εικόνα και η χρήση ενός CNN για την ταξινόμηση της παρουσίας του αντικειμένου εντός αυτής της περιοχής. Το πρόβλημα με αυτήν την προσέγγιση είναι ότι τα αντικείμενα ενδιαφέροντος μπορεί να έχουν διαφορετικές χωρικές θέσεις εντός της εικόνας και διαφορετικούς λόγους διαστάσεων. Ως εκ τούτου, θα πρέπει να χρησιμοποιηθεί ένας τεράστιος αριθμός περιοχών και αυτό θα μπορούσε να μην αποδώσει υπολογιστικά. Εξαιτίας αυτού του προβλήματος αλγόριθμοι όπως R-CNN, YOLO κ.λπ. έχουν αναπτυχθεί για να εντοπίζουν αυτά τα περιστατικά γρήγορα.

Λύση στο πρόβλημα της επιλογής ενός τεράστιου αριθμού περιοχών, ο Ross Girshick [18] πρότεινε μια μέθοδο όπου χρησιμοποιεί επιλεκτική αναζήτηση για την εξαγωγή 2000 περιοχών από την εικόνα και τις ονόμασε προτάσεις περιοχής. Επομένως, η προσπάθεια ταξινόμησης ενός τεράστιου αριθμού περιοχών μπορεί να αντικατασταθεί από τις 2000 περιοχές. Αυτές οι προτάσεις περιοχών δημιουργούνται χρησιμοποιώντας τον αλγόριθμο επιλεκτικής αναζήτησης η λειτουργία του οποίου περιγράφεται παρακάτω:

Επιλεκτική αναζήτηση:

1. Πραγματοποιείται αρχική υπο-τμηματοποίηση, δημιουργία πολλών υποψήφιων περιοχών.

2. Χρήση του αλγορίθμου greedy έτσι ώστε να συνδυαστούν αναδρομικά παρόμοιες περιοχές σε μεγαλύτερες

3. Χρήση των παραγόμενων περιοχών του βήματος 2 για την δημιουργία τελικών προτάσεων υποψήφιων περιοχών.

Οι προτάσεις 2000 υποψήφιων περιοχών που αναφέρθηκαν προηγουμένως στρεβλώνονται σε ένα τετράγωνο και έπειτα τροφοδοτούνται σε ένα συνελικτικό νευρωνικό δίκτυο που ακολούθως παράγει ένα διάνυσμα χαρακτηριστικών 4096 διαστάσεων ως έξοδο. Το CNN λειτουργεί ως εξολκέας χαρακτηριστικών. Το πυκνό στρώμα εξόδου αποτελείται από τα χαρακτηριστικά που εξάγονται από την εικόνα και ακολούθως τα εξαγόμενα αυτά χαρακτηριστικά τροφοδοτούνται σε ένα SVM για να ταξινομήσει την παρουσία του αντικειμένου εντός αυτής της πρότασης υποψήφιας περιοχής. Εκτός από την πρόβλεψη της παρουσίας ενός αντικειμένου στις προτάσεις περιοχής, ο αλγόριθμος προβλέπει επίσης τέσσερις τιμές που είναι τιμές μετατόπισης για να αυξηθεί η ακρίβεια του πλαισίου οριοθέτησης. Για παράδειγμα, δεδομένης μιας πρότασης περιοχής, ο αλγόριθμος θα είχε προβλέψει την παρουσία ενός ατόμου, αλλά το πρόσωπο αυτού του ατόμου εντός αυτής της πρότασης περιοχής θα μπορούσε να είχε κοπεί στο μισό. Επομένως, οι τιμές μετατόπισης βοηθούν στην προσαρμογή του πλαισίου οριοθέτησης της πρότασης περιοχής.

Τα προβλήματα που προκύπτουν από τη χρήση των δικτύων RCNN είναι ότι χρειάζεται ακόμα πολύς χρόνος για να πραγματοποιηθεί η εκπαίδευση του δικτύου καθώς είναι αναγκαίο να ταξινομηθούν 2000 προτάσεις περιοχής ανά εικόνα. Σημαντική αδυναμία εφαρμογής των συγκριμένων δικτύων είναι η χρήση ή εφαρμογή τους σε πραγματικό χρόνο καθώς απαιτούνται περίπου 47 δευτερόλεπτα για κάθε δοκιμαστική εικόνα. Επιπρόσθετα ο αλγόριθμος επιλεκτικής αναζήτησης είναι ένας σταθερός αλγόριθμος, επομένως, δεν γίνεται μάθηση σε αυτό το στάδιο γεγονός το οποίο θα μπορούσε να οδηγήσει στη δημιουργία κακών προτάσεων για υποψήφιες περιοχές.

Στην προσπάθεια επίλυσης των μειονεκτημάτων των δικτύων R-CNN, ο Ross Girshick [19] δημιούργησε έναν ταχύτερο αλγόριθμο ανίχνευσης αντικειμένων που ονομάστηκε Fast R-CNN. Η προσέγγιση είναι παρόμοια με τον αλγόριθμο R-CNN. Η διαφορά που παρατηρείτε είναι το γεγονός ότι αντί να πραγματοποιείται είσοδος των προτάσεων περιοχών στο CNN, τροφοδοτούμε την εικόνα εισόδου στο CNN για να δημιουργήσουμε έναν συνελικτικό χάρτη χαρακτηριστικών. Από τον συνελικτικό χάρτη χαρακτηριστικών, γίνεται προσδιορισμός των περιοχών των προτάσεων και παραμόρφωση τους σε τετράγωνα και χρησιμοποιώντας ένα στρώμα συγκέντρωσης RoI αναδιαμορφώνονται σε ένα σταθερό μέγεθος ώστε να μπορεί να τροφοδοτηθεί σε ένα πλήρως συνδεδεμένο στρώμα. Από το διάνυσμα χαρακτηριστικών RoI, γίνεται χρήση ενός στρώματος softmax για να για να μπορέσει να προβλεφθεί η κλάση της προτεινόμενης περιοχής και επίσης οι τιμές μετατόπισης για το πλαίσιο οριοθέτησης.

Η αιτία που το "Fast R-CNN" είναι ταχύτερο από το R-CNN είναι επειδή δεν απαιτείται να τροφοδοτεί 2000 προτάσεις περιοχών στο συνελικτικό νευρωνικό δίκτυο κάθε φορά. Αντιθέτως, η λειτουργία συνέλιξης γίνεται μόνο

μία φορά ανά εικόνα και δημιουργείται ένας χάρτης χαρακτηριστικών από αυτήν.

Και οι δύο παραπάνω αλγόριθμοι (R-CNN & Fast R-CNN) χρησιμοποιούν επιλεκτική αναζήτηση για να ανακαλύψουν τις προτάσεις περιοχής. Η επιλεκτική αναζήτηση είναι μια αργή και χρονοβόρα διαδικασία που επηρεάζει την απόδοση του δικτύου. Επομένως, ο Shaoqing Ren et al. [20] βρήκε έναν αλγόριθμο ανίχνευσης αντικειμένων που καταργεί τον αλγόριθμο επιλεκτικής αναζήτησης και επιτρέπει στο δίκτυο να μάθει τις προτάσεις περιοχής.

Παρόμοια με τη λειτουργία του Fast R-CNN, η εικόνα παρέχεται ως είσοδος σε ένα συνελικτικό δίκτυο που παρέχει έναν συνελικτικό χάρτη χαρακτηριστικών. Αντί να χρησιμοποιείται επιλεκτικός αλγόριθμος αναζήτησης στον χάρτη χαρακτηριστικών για τον προσδιορισμό των προτάσεων περιοχής, χρησιμοποιείται ένα ξεχωριστό δίκτυο για την πρόβλεψη των προτάσεων περιοχής. Οι προβλεπόμενες προτάσεις περιοχών στη συνέχεια αναδιαμορφώνονται χρησιμοποιώντας ένα στρώμα συγκέντρωσης RoI το οποίο στη συνέχεια χρησιμοποιείται για την ταξινόμηση της εικόνας εντός της προτεινόμενης περιοχής και την πρόβλεψη των τιμών μετατόπισης για τα πλαίσια οριοθέτησης.

Όλοι οι προηγούμενοι αλγόριθμοι ανίχνευσης αντικειμένων χρησιμοποιούν περιοχές για τον εντοπισμό του αντικειμένου εντός της εικόνας. Το δίκτυο δεν εξετάζει την πλήρη εικόνα. Αντίθετα, τμήματα της εικόνας που έχουν μεγάλες πιθανότητες να περιέχουν το αντικείμενο. Το YOLO ή το You Only Look Once [8] είναι ένας αλγόριθμος ανίχνευσης αντικειμένων πολύ διαφορετικός από τους αλγόριθμους που βασίζονται στην περιοχή που φαίνονται παραπάνω. Το YOLO ένα ενιαίο συνελικτικό δίκτυο προβλέπει τα οριοθετημένα πλαίσια και τις πιθανότητες κλάσης για αυτά τα πλαίσια.

Η λειτουργία του YOLO βασίζεται στο διαχωρισμό της εικόνας σε ένα πλέγμα SxS και στο γεγονός ότι ακολούθως λαμβάνουμε m οριοθετημένα πλαίσια από αυτό τον διαχωρισμό. Για κάθε πλαίσιο οριοθέτησης, το δίκτυο εξάγει μια κλάση πιθανότητας και τις αντίστοιχες τιμές μετατόπισης. Τα οριοθετημένα πλαίσια που έχουν την πιθανότητα κλάσης πάνω από μια τιμή κατωφλίου επιλέγονται και χρησιμοποιούνται για τον εντοπισμό του αντικειμένου μέσα στην εικόνα.

Πρόκειται για ένα κατά πολύ ταχύτερο δίκτυο συγκριτικά με τους άλλους αλγόριθμους ανίχνευσης αντικειμένων. Ένα πρόβλημα που εμφανίζει ο αλγόριθμος είναι ότι δυσκολεύεται να εντοπίσει μικρά σε διαστάσεις αντικείμενα μέσα στην εικόνα γεγονός που οφείλεται στους χωρικούς περιορισμούς του αλγορίθμου.

Σημαντική έρευνα για την ανίχνευση αντικειμένων με βάση τη βαθιά μηχανική μάθηση πραγματοποιήθηκε από τους Licheng Jiao, Fan Zhang, Fang Liu, Shuyuan Yang, Lingling Li, Zhixi Feng, and Rong Qu[6]. Σύμφωνα με την έρευνα οι πιο σύγχρονοι ανιχνευτές αντικειμένων έχουν την δυνατότητα να κατηγοριοποιηθούν σε ανιχνευτές ενός σταδίου και σε ανιχνευτές δύο σταδίων [6]. Οι ανιχνευτές δύο σταδίων ανιχνεύουν αντικείμενα σε δύο στάδια:

- ένα στάδιο πρότασης περιοχής και
- μια ταξινόμηση και στάδιο εντοπισμού.

Όσο αναφορά τους ανιχνευτές ενός σταδίου, πετυχαίνουν ακριβέστερο εντοπισμό και υψηλότερη ακρίβεια αναγνώρισης αντικειμένων αλλά ταυτόχρονα χαμηλότερη ταχύτητα συμπερασμάτων. Οι ανιχνευτές ενός σταδίου προβλέπουν απευθείας και γρηγορότερα τα πλαίσια από την εικόνα εισόδου χωρίς να είναι απαραίτητο ένα στάδιο πρότασης περιοχής και επομένως εμφανίζουν υψηλότερη ταχύτητα συμπερασμάτων γεγονός που τα καθιστά καταλληλότερα για εφαρμογές που πραγματοποιούνται σε πραγματικό χρόνο [6].

Το δίκτυο R-CNN ήταν το πρώτο CNN που βασίστηκε για την λειτουργία του στην περιοχή όσο αναφορά την ανίχνευση αντικειμένων χρησιμοποιώντας επιλεκτική αναζήτηση για προτάσεις περιοχής καθώς και ένα CNN για εξαγωγή χαρακτηριστικών σε κάθε πρόταση περιοχής. Τα προβλήματα όμως που παρουσιάστηκαν από την πολύπλοκη εκπαίδευση πολλών σταδίων καθώς και ο αργός χρόνος δοκιμής, υπήρξε αφορμή για την ανάπτυξη μιας ταχύτερης έκδοσης που ονομάζεται fast R-CNN. Σύμφωνα με το Fast R-CNN πραγματοποιείτε εξάγωγή των χαρακτηριστικών μιας εικόνας μία φορά και στη συνέχεια παράγει τις προτάσεις της περιοχής, οι οποίες ακολούθως ταξινομούνται. Έπειτα υπήρξε το Faster R-CNN, το οποίο ήταν πιο αποτελεσματικό διότι ο ανιχνευτής λειτουργούσε βάσει περιοχής, και με αυτό τον τρόπο βελτίωσε ακόμη περισσότερο το CNN που βασίζεται στην περιοχή χρησιμοποιώντας μια πρόταση περιοχής δίκτυο (RPN) αντί για μια μέθοδο πρότασης εξωτερικής περιοχής για την παραγωγή προτάσεων περιοχής, που οδηγούν σε υψηλότερη ταχύτητα στο χρόνο δοκιμής [7]. Το Mask R-CNN είναι μια εκτεταμένη έκδοση του Faster R-CNN προσθέτοντας μια πλήρως συνδεδεμένη κεφαλή μάσκας στο δίκτυο. Χρησιμοποιείται κυρίως για εργασίες τμηματοποίησης [6].

Ένα άλλο δίκτυο ενός σταδίου είναι το YOLO και οι νεότερες εκδόσεις του (v2,v3,v4) και είναι ευρέως γνωστές για την ανίχνευση αντικειμένων σε πραγματικό χρόνο λόγω της υψηλής ταχύτητάς τους. Το YOLO χωρίζει την εικόνα σε ένα πλέγμα και προβλέπει οριοθετημένα πλαίσια και βαθμολογίες κλάσεων σε κάθε κελί [8]. Ο ανιχνευτής πολλαπλών κουτιών (Single-shot (SSD)) προβλέπει άμεσα βαθμολογίες κλάσεων και μετατοπίσεις κουτιών για ένα σταθερό σύνολο με προεπιλεγμένα πλαίσια οριοθέτησης διαφορετικών κλιμάκων σε κάθε τοποθεσία, σε διάφορους χάρτες χαρακτηριστικών με διαφορετικές κλίμακες [6].

Λόγω ανάγκης αντιμετώπισης προβλημάτων ανίχνευσης που αναφέρθηκαν δημιουργήθηκαν αρκετά επισημασμένα σύνολα δεδομένων με οριοθετημένα πλαίσια. Εμφανίζονται σχεδόν σε κάθε σύγκριση μεταξύ ποικίλων αρχιτεκτονικών και δικτύων. Δύο από τα πιο αναγνωρισμένα και με ευρεία χρήση σύνολα δεδομένων για ανίχνευση δεδομένων είναι το PASCAL VOC, MSCOCO (Microsoft Common Object in Context) και ILSVRC (ImageNet Large Scales Visual Recognition Challenge). Ειδικότερα, στο χώρο έρευνας για

εντοπισμό αντικειμένων σε αυτόνομη οδήγηση, μερικά από τα πιο γνωστά σύνολα δεδομένων είναι:

• KITTI: Περιλαμβάνει σκηνές κυκλοφορίας της πόλης της Καρλσρούης και οι σκηνές χωρίζονται σε 5 κατηγορίες, συμπεριλαμβανομένων των σχολιασμών (8 κλάσεις) και των πλαισίων οριοθέτησης [9].

• Waymo Open: Διαθέτει 1950 τμήματα οδήγησης και λαμβάνει υπόψη 4 κατηγορίες αντικειμένων. Τα δεδομένα λαμβάνουν επίσης υπόψη διάφορες περιβαλλοντικές και σκηνές αστικής κυκλοφορίας [10].

• nuScenes: Στο παρόν σύνολο δεδομένων υπάρχουν πολύπλοκες σκηνές οδήγησης στην πόλη με σχολιασμούς 3D αντικειμένων για αυτόνομη οδήγηση. Περιέχει 1000 σκηνές , 23 κατηγορίες αντικειμένων, περίπου 1,4 εκατομμύρια εικόνες κάμερας των πόλεων Βοστώνη και Σιγκαπούρη [11].

Κεφάλαιο 2

2.1 Μηχανική Μάθηση

Μία υπό-κατηγορία της τεχνητής νοημοσύνης είναι η μηχανική μάθηση, η οποία ορίζεται ως η ικανότητα που παρουσιάζει μια μηχανή να μιμείται την ευφυή ανθρώπινη συμπεριφορά. Τα συστήματα τεχνητής νοημοσύνης που αναπτύσσονται χρησιμοποιούνται για την εκτέλεση σύνθετων εργασιών με τρόπο πανομοιότυπο με τον τρόπο που οι άνθρωποι βρίσκουν λύση στα προβλήματα. Σκοπός της τεχνητής νοημοσύνης είναι να εφεύρει μοντέλα υπολογιστών που παρουσιάζουν «έξυπνες συμπεριφορές» όπως οι άνθρωποι. Δηλαδή οι μηχανές μπορούν να αναγνωρίσουν μια οπτική σκηνή, να καταλάβουν ένα κείμενο γραμμένο σε φυσική γλώσσα ή να εκτελέσουν μια ενέργεια στον φυσικό κόσμο. Συνεπώς θα μπορούσαμε να συμπεράνουμε ότι η μηχανική μάθηση αποτελεί έναν τρόπο χρήσης της τεχνητής νοημοσύνης. Ο ορισμός της δόθηκε τη δεκαετία του 1950 από τον πρωτοπόρο της τεχνητής νοημοσύνης, Arthur Samuel ως «το πεδίο σπουδών που δίνει στους υπολογιστές τη δυνατότητα να μαθαίνουν χωρίς να έχουν προγραμματιστεί ρητά».

Σε ορισμένες περιπτώσεις, όπως όταν επιθυμούμε την εκπαίδευση ενός υπολογιστή έτσι ώστε να αναγνωρίζει εικόνες διαφορετικών ανθρώπων, η σύνταξη ενός προγράμματος που θα ακολουθήσει το μηχάνημα αποτελεί πολλές φορές μία εξαιρετικά χρονοβόρα έως και αδύνατη διαδικασία. Σε αντίθεση με τους ανθρώπους, είναι ιδιαίτερα δύσκολο να «δείξουμε» σε έναν υπολογιστή πως να εκτελέσει την παραπάνω αναγνώριση προσώπων. Σε αυτό το πρόβλημα η μηχανική μάθηση είναι αυτή που ακολουθεί την προσέγγιση που επιτρέπει στους υπολογιστές να μάθουν να προγραμματίζουν τον εαυτό τους μέσω της εμπειρίας. Η μηχανική μάθηση αρχίζει με δεδομένα όπως αριθμούς, φωτογραφίες ή κείμενο, τραπεζικές συναλλαγές, εικόνες ανθρώπων ή ακόμα και δεδομένα χρονοσειρών από αισθητήρες ή αναφορές πωλήσεων. Τα δεδομένα συλλέγονται και ετοιμάζονται για χρήση ως δεδομένα εκπαίδευσης ή ως οι πληροφορίες στις οποίες θα εκπαιδευτεί το μοντέλο μηχανικής μάθησης. Η ύπαρξη όσο τω δυνατόν περισσότερων δεδομένων οδηγεί σε ένα καλύτερο πρόγραμμα μάθησης.

Οι προγραμματιστές αρχικά επιλέγουν ποιο μοντέλο μηχανικής μάθησης θα χρησιμοποιηθεί, φορτώνουν τα δεδομένα και στη συνέχεια το μοντέλο υπολογιστή εκπαιδεύεται για να εντοπίζει μοτίβα ή να κάνει προβλέψεις. Ο προγραμματιστής έχει τη δυνατότητα επίσης να τροποποιήσει το μοντέλο, να πραγματοποιήσει αλλαγή των παραμέτρων του, έτσι ώστε να το οδηγήσει προς πιο ακριβή αποτελέσματα. Ορισμένα δεδομένα επιλέγονται από τα δεδομένα εκπαίδευσης που θα χρησιμοποιηθούν ως δεδομένα αξιολόγησης, τα οποία δοκιμάζουν πόσο ακριβές είναι το μοντέλο μηχανικής μάθησης όταν εφαρμόζονται νέα δεδομένα. Αυτό έχει σαν αποτέλεσμα το μοντέλο να μπορεί να χρησιμοποιηθεί σε μεταγενέστερες εφαρμογές με διαφορετικά σύνολα δεδομένων.

Σύμφωνα με την βιβλιογραφία, η λειτουργία ενός συστήματος μηχανικής μάθησης ενδέχεται να είναι περιγραφική, δηλαδή ότι το σύστημα χρησιμοποιεί τα δεδομένα για να ερμηνεύσει τι συνέβη, προγνωστική, που

σημαίνει ότι το σύστημα επεξεργάζεται τα δεδομένα για να προβλέψει τι θα συμβεί ή προδιαγραφική, δηλαδή ότι το σύστημα θα μεταχειριστεί με τέτοιο τρόπο τα δεδομένα για να κάνει προτάσεις σχετικά με τη δράση που πρέπει να εκτελεστεί.

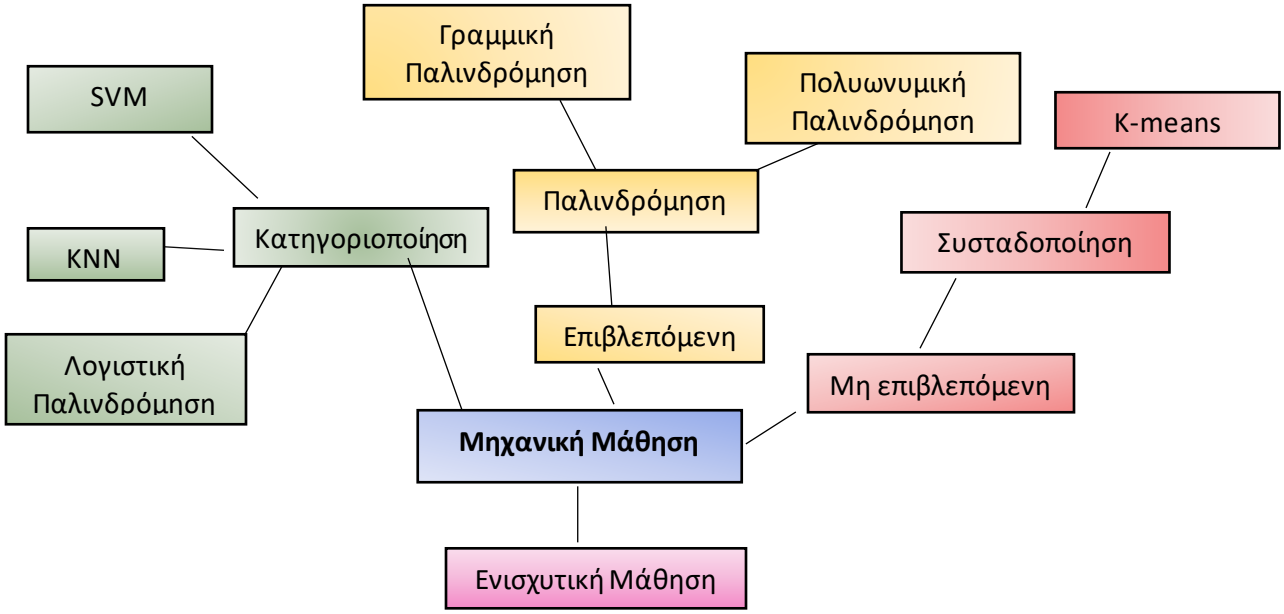
2.2 Κατηγορίες Μάθησης

Υπάρχουν τρεις υπό-κατηγορίες μηχανικής μάθησης:

- Τα εποπτευόμενα μοντέλα μηχανικής μάθησης μία κατηγορία σύμφωνα με την οποία η εκπαίδευση γίνεται με σύνολα δεδομένων με ετικέτα, τα οποία επιτρέπουν στα μοντέλα να μαθαίνουν και να αναπτύσσονται με πιο ακριβή τρόπο, με την πάροδο του χρόνου. Για παράδειγμα, ένας αλγόριθμος που επιθυμούμε να εκπαιδευτεί με εικόνες σκύλων και άλλων πραγμάτων, όλα επισημασμένα από ανθρώπους, θα επιτρέψει στο μηχάνημα να μάθει τρόπους να αναγνωρίζει από μόνο του εικόνες σκύλων. Η εποπτευόμενη μηχανική εκμάθηση αποτελεί τον πιο κοινό τύπο που χρησιμοποιείται σήμερα.

- Η μηχανική μάθηση χωρίς επίβλεψη, σύμφωνα με την οποία ένα πρόγραμμα αναζητά μοτίβα σε δεδομένα χωρίς ετικέτα. Η μη εποπτευόμενη μηχανική εκμάθηση είναι σε θέση να εντοπίζει μοτίβα ή τάσεις που οι άνθρωποι δεν αναζητούν ρητά. Για παράδειγμα, ένα πρόγραμμα μηχανικής μάθησης χωρίς επίβλεψη έχει τη δυνατότητα να εξετάσει τα δεδομένα πωλήσεων στο διαδίκτυο και να διακρίνει διαφορετικούς τύπους πελατών που πραγματοποίησαν αγορές.

- Η ενισχυτική μηχανική εκμάθηση η οποία εκπαιδεύει τις μηχανές μέσω δοκιμής και λάθους ώστε να μπορούν να επιλέξουν την καλύτερη ενέργεια με τη κατασκευή ενός συστήματος ανταμοιβής. Η ενισχυτική μάθηση είναι σε θέση να εκπαιδεύσει μοντέλα να παίζουν παιχνίδια ή να εκπαιδεύει αυτόνομα οχήματα να οδηγούν υποδεικνύοντας στο μηχάνημα πότε πραγματοποίησε τις σωστές αποφάσεις, γεγονός που το βοηθά να εκπαιδευτεί με την πάροδο του χρόνου σχετικά με το ποιες ενέργειες πρέπει να εκτελεί.



Εικόνα.1: Κατηγορίες Μηχανικής Μάθησης

2.3 Τεχνικές Παλινδρόμησης

Σε αρκετές περιπτώσεις που πρέπει να διαχειριστούμε τυχαίες μεταβλητές αναγκαίο είναι να προσδιοριστεί ο τρόπος με τον οποίο μπορούν να συσχετιστούν μεταξύ τους. Παράδειγμα των συγκεκριμένων μεταβλητών αποτελούν:

- Η διάρκεια ζωής ενός οργανισμού σε μια περιοχή και το ποσοστό μόλυνσης στην περιοχή αυτή.
- Το ύψος των αποδοχών των υπαλλήλων μια εταιρείας και τα χρόνια υπηρεσίας τους.
- Οι δαπάνες κατανάλωσης σε ένα νοικοκυριό και το διαθέσιμο εισόδημα της οικογένειας.

Η ανάπτυξη ενός μαθηματικού μοντέλου αποτελεί μια στατιστική διαδικασία που συμβάλλει στην παραγωγή εξισώσεων που περιγράφουν τη σχέση μεταξύ των ανεξάρτητων μεταβλητών και της εξαρτημένης. Τα μοντέλα παλινδρόμησης μπορούν να χρησιμοποιηθούν για την πρόβλεψη μιας συνεχούς πραγματικής τιμής.

2.3.1 Απλή γραμμική παλινδρόμηση

Ο κλάδος της Στατιστικής που μελετά τη σχέση μεταξύ δύο ή περισσότερων μεταβλητών με σκοπό την πρόβλεψη μιας απ' αυτές μέσω των άλλων χαρακτηρίζεται ως ανάλυση παλινδρόμησης (regression analysis). Ιστορικά, ο όρος "regression" χρησιμοποιήθηκε για πρώτη φορά από τον Άγγλο ανθρωπολόγο Galton (1822-1911) το 1885. Με τη μελέτη του ύψους των παιδιών συγκριτικά με το ύψος των γονέων συμπέρανε ότι παιδιά με ψηλούς γονείς έχουν την τάση, κατά μέσο όρο, να είναι κοντύτερα των γονιών τους, ενώ παιδιά κοντών γονέων τείνουν, κατά μέσο όρο, να γίνονται ψηλότερα των γονιών τους.

Η πιο απλή περίπτωση παλινδρόμησης ονομάζεται απλή γραμμική παλινδρόμηση (simple linear regression), κατά την οποία υπάρχει μόνο μια ανεξάρτητη μεταβλητή *X* (independent or input variable), και η εξαρτημένη μεταβλητή *Y* (dependent or response variable), η οποία μπορεί να προσεγγιστεί σε ικανοποιητικό βαθμό από μία γραμμική συνάρτηση του X. Η περίπτωση αυτή εντοπίζεται τόσο σε πειραματικές όσο και σε μη πειραματικές μελέτες. Στις πειραματικές μελέτες ο ερευνητής καθορίζει, για παράδειγμα, από πριν τις δόσεις ενός φαρμάκου (ανεξάρτητη μεταβλητή) που παρέχει στα πειραματόζωα και απαριθμεί τις αντιδράσεις τους (εξαρτημένη μεταβλητή). Με την παλινδρόμηση μας ενδιαφέρει να προσδιορίσει μία σχέση δόσης-αντίδρασης για το συγκεκριμένο φάρμακο. Στις μη πειραματικές μελέτες ή δειγματοληψίες, πραγματοποιούνται μετρήσεις σε δύο μεταβλητές για κάθε μονάδα του δείγματος. Σε ένα δείγμα 10 μαθητών μετράμε, για παράδειγμα, το βάρος και το ύψος τους. Η διάκριση που γίνεται μεταξύ ανεξάρτητης και εξαρτημένης μεταβλητής είναι δύσκολη. Αν αυτό που μας ενδιαφέρει είναι το "τι συμβαίνει με το βάρος των παιδιών όταν αλλάζει το ύψος τους", τότε θεωρούμε ως ανεξάρτητη μεταβλητή X το ύψος και ως εξαρτημένη μεταβλητή Y το βάρος. Οπότε, ενδιαφερόμαστε για την παλινδρόμηση του βάρους (Y) πάνω στο ύψος (X). Αντίθετα, αν μας ενδιαφέρει το "τι συμβαίνει με το ύψος των παιδιών όταν αλλάζει το βάρος τους", τότε θεωρούμε ως ανεξάρτητη μεταβλητή X το βάρος και ως εξαρτημένη μεταβλητή Y το ύψος. Τότε έχουμε παλινδρόμηση του ύψους (Y) πάνω στο βάρος (X).

Όπως αναφέρθηκε και παραπάνω η απλή γραμμική παλινδρόμηση αποτελεί την πιο απλή μορφή παλινδρόμησης. Υπάρχει μόνο μία ανεξάρτητη μεταβλητή x και μία εξαρτημένη μεταβλητή y, που προσεγγίζεται ως μια γραμμική συνάρτηση του x. Η τιμή y_i της y, για κάθε τιμή x_i της x, δίνεται από τη σχέση:

$$y_i = b_0 + b_1 * x_i + e_i$$

Το b1 είναι ο συντελεστής της ανεξάρτητης μεταβλητής και bo η σταθερά. Το πρόβλημα της γραμμικής παλινδρόμησης είναι η εύρεση των παραμέτρων bo και b1 που εκφράζουν καλύτερα τη γραμμική εξάρτηση της y από τη x. Για κάθε τιμή bo και b1 ορίζεται μια διαφορετική γραμμική σχέση όπου γεωμετρικά εκφράζεται από μια ευθεία με τις εξής παραμέτρους:

• Η σταθερά bo είναι η τιμή του y για x=o

• Ο συντελεστής b1 του x είναι η κλίση (slope) της ευθείας ή αλλιώς ο συντελεστής παλινδρόμησης (regression coefficient).

Εκφράζει τη μεταβολή της μεταβλητής y όταν η μεταβλητή x μεταβληθεί κατά μία μονάδα. Ο όρος ei ονομάζεται σφάλμα παλινδρόμησης (regression error) ή

απόκλιση. Στην πράξη είναι η διαφορά της παρατηρούμενης τιμής yi δεδομένης της τιμής xi από την τιμή της πρόβλεψης που προκύπτει από το μοντέλο. Για την εύρεση της ευθείας παλινδρόμησης που ταιριάζει καλύτερα στα δεδομένα μας θα χρησιμοποιούμε μια από τις κυριότερες μεθόδους υπολογισμού εκτιμητή ευθείας, την μέθοδο ελαχίστων τετραγώνων. Ουσιαστικά πρέπει να εκτιμήσουμε τους παραμέτρους του της παλινδρόμησης, bo και b1. Ο αριθμός των ζευγών αυτών είναι άπειρος και αναζητούμε την ευθεία που περιγράφει με τον καλύτερο δυνατό τρόπο τη σχέση μεταξύ των δύο μεταβλητών. Η γραμμή παλινδρόμησης πρέπει να περνάει κοντά από τα σημεία που αντιστοιχούν τα ζεύγη παρατηρήσεων (x,y) έτσι ώστε να ελαχιστοποιούνται τα σφάλματα πρόβλεψης.

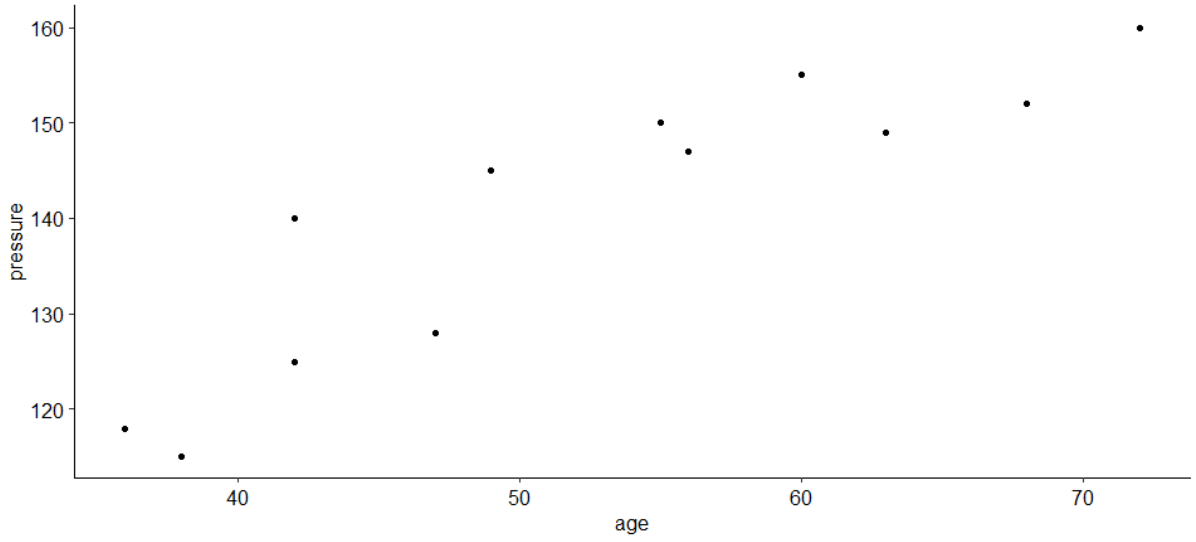
Παράδειγμα

Σε μία προσπάθεια κατανόησης της λειτουργίας της απλής γραμμικής παλινδρόμησης θεωρούμε ότι λαμβάνουμε τις ακόλουθες τιμές της πίεσης του αίματος και της αντίστοιχης ηλικίας σε έτη για n = 12 γυναίκες:

Ηλικία (*X*)	36	38	42	42	47	49	55	56	60	63	68	72
Πίεση αίματος (*Y*)	118	115	125	140	128	145	150	147	155	149	152	160

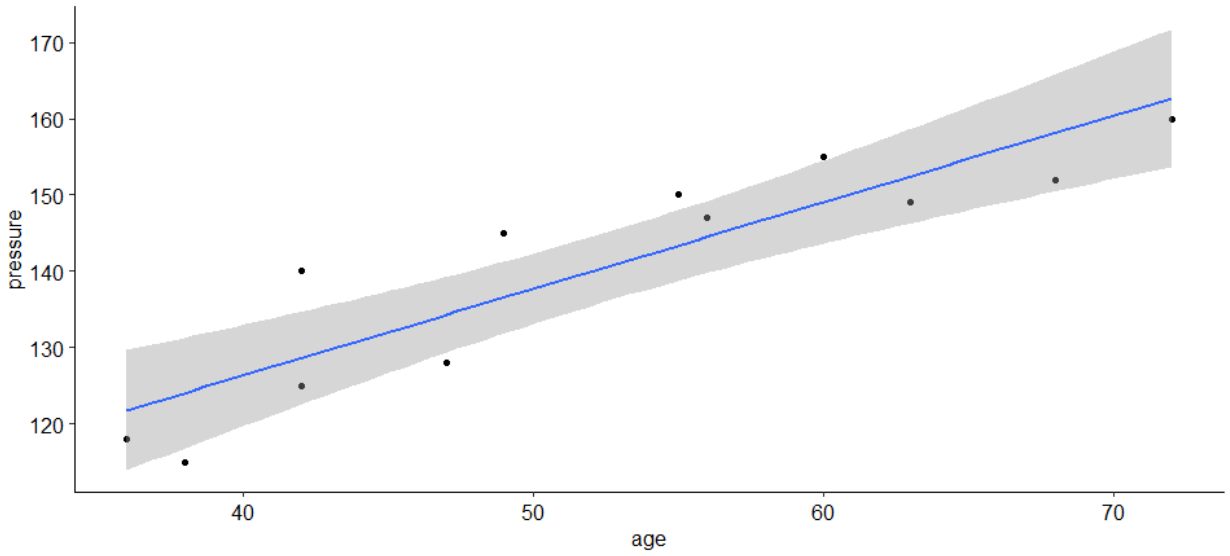
Πίνακας.1: Σύνολο Δεδομένων Απλής Γραμμικής Παλινδρόμησης

Έστω πως ο άξονας X αναπαριστά την ηλικία της κάθε γυναίκας και ο Y την πίεση του αίματος. Θέλουμε να μελετήσουμε πώς η πίεση του αίματος επηρεάζεται από την ηλικία του. Στο παρακάτω γράφημα φαίνεται πως οι τιμές της πίεσης του αίματος είναι κατανεμημένες μεταξύ των γυναικών και σύμφωνα με την αντίστοιχη ηλικία τους, για παράδειγμα το ζευγάρι τιμών (X1,Y1) = (36, 118), (X2,Y2) = (38, 115), κ.ο.κ. Εισάγουμε τα δεδομένα σε δύο μμεταβλητές – στήλες με n = 12 cases – γραμμές. Ονομάζουμε τις μεταβλητές Age (ή X) και Pressure (ή Y). Η εξίσωση της απλής γραμμικής παλινδρόμησης θα είναι: Ηλικία = b_0 + b_1 * πίεση αίματος.



Εικόνα.2: Διάγραμμα Διασποράς Συνόλου Δεδομένων Απλής Γραμμικής Παλινδρόμησης

Με δεδομένο το σύνολο δεδομένων που βλέπουμε στον πίνακα, η ευθεία της παλινδρόμησης θα είναι εξής:



Εικόνα.3: Διάγραμμα Ευθείας Γραμμικής Παλινδρόμησης

2.3.2 Πολλαπλή γραμμική παλινδρόμηση

Όταν αναφερόμαστε στην πολλαπλή παλινδρόμηση, πραγματοποιούμε μετρήσεις ταυτόχρονα για τρεις ή περισσότερες μεταβλητές από τις οποίες η μία θεωρούμε ότι είναι εξαρτημένη (Y) από τη δράση των λοιπών (Xi), π.χ. τις X1 , X2 και X3. Στην συγκεκριμένη περίπτωση ισχύουν οι εξής προϋποθέσεις για την εξαρτημένη Y μεταβλητή: οι τιμές της είναι τυχαίες, έχουν κανονική κατανομή και βρίσκονται σε αντιστοιχία με τους παρατηρούμενους συνδυασμούς των τιμών των ανεξάρτητων μεταβλητών. Επαναληπτικές μετρήσεις της Y μεταβλητής που ενδέχεται να προκύψουν, σε συνδυασμό πάντοτε με τις τιμές των ανεξάρτητων μεταβλητών, κρίνεται αναγκαίο επίσης, να έχουν κανονική κατανομή και κοινή διακύμανση. Τα μοντέλα πολλαπλής παλινδρόμησης έχουν τη δυνατότητα επίσης να διαχωριστούν επίσης σε δύο διαφορετικές προοπτικές, ως επεξηγηματικά και προβλεπτικά (Pedhazur, 1997). Τα επεξηγηματικά μοντέλα έχουν ως στόχο την εδραίωση ενός ισχυρού μοντέλου το οποίο πρέπει να επιβεβαιώνει το αποτέλεσμα των προβλέψεων εκείνων μόνο που διαθέτουν ως εφαλτήριο ικανή θεωρητική υπόσταση, απορρίπτοντας άλλες που δεν είναι σχετικές. Δηλαδή, τα μοντέλα οφείλουν να ελέγχουν αν μία σημαντική προβλεπτική μεταβλητή μπορεί, ένεκα του θεωρητικού της υπόβαθρου, να παράγει τη μέγιστη δυνατή διακύμανση οδηγώντας έτσι σε ακριβέστερες προβλέψεις. Τα προβλεπτικά ή διερευνητικά μοντέλα διαθέτουν πιο ελεύθερη θεωρητική βάση και συνεπώς είναι περισσότερο ευέλικτα αφού βασίζονται άμεσα στην απρόσκοπτη ανάλυση των στοιχείων. Τα μοντέλα αυτά επιχειρούν την ανεύρεση της ομάδας εκείνης των προσβλεπουσών μεταβλητών η οποία παρέχει το καλύτερο αποτέλεσμα πρόβλεψης, ανεξαρτήτως αν το μοντέλο προσεγγίζει ή όχι κάποιο ορθό επεξηγηματικό μηχανισμό σε θεωρητικό επίπεδο. Δεν επιδιώκουν ιδιαίτερα να ανιχνεύσουν αν οι προβλέψεις αντανακλούν κάποια πραγματική επιστημονική αιτία υπεύθυνη για την έκβαση του συγκεκριμένου αποτελέσματος.

Η πολλαπλή παλινδρόμηση έχει ευρεία επιστημονική αποδοχή διότι θεωρείται ισχυρό και ευέλικτο στατιστικό εργαλείο με πλήθος εφαρμογών σε τελείως διαφορετικά ερευνητικά πεδία (Draper & Smith, 1989, Pedhazur, 1997, Weisburg, 1985):

• Διοίκηση επιχειρήσεων και έρευνα αγοράς – εκτίμηση του βαθμού επίδοσης του προσωπικού μίας εταιρίας, διαχείριση του αριθμού και έκτασης των παραπόνων πελατών

• Προβλήματα οδικής συγκοινωνίας – διαχείριση του τύπου οδοστρώματος και είδους μεταφορικού μέσου στο χρόνο εκπλήρωσης μίας μετακίνησης

• Υπέρβαση στον αθλητισμό – τρόποι βελτίωσης των αθλητικών επιδόσεων στο στίβο, προσαρμογή ενός βέλτιστου διαιτολογίου

• Ατμοσφαιρική και υδρόβια ρύπανση με προεκτάσεις στη διαφύλαξη της δημόσιας υγείας

• Τρόποι διερεύνησης της συμπεριφοράς του δείκτη νοημοσύνης σε διαγωνιστικό επίπεδο

• Εκτίμηση της δράσης των χημικών συστατικών ενός τρόφιμου στις οργανοληπτικές ιδιότητές του. Συνοψίζοντας, η ανάλυση της παλινδρόμησης μπορεί να χρησιμοποιηθεί για την περιγραφή των ειδικών σχέσεων μεταξύ των μεταβλητών, τη διακρίβωση θεωρητικών υποθέσεων, την πρόβλεψη από λήψεις πειραματικών δεδομένων και τη δημιουργία και επαλήθευση εξισώσεων πολλαπλής παλινδρόμησης (Montgomery et al., 2012, Kleinbaum et al., 1998). Η υπολογιστική διαδικασία της ανάλυσης της πολλαπλής παλινδρόμησης και συσχέτισης είναι ιδιαίτερα κοπιαστική και στην πράξη εφικτή μόνο με τη χρήση στατιστικών προγραμμάτων Η/Υ.

Σύμφωνα με το θεσμό της απλής γραμμικής παλινδρόμησης για ένα πληθυσμό με ένα ζεύγος μεταβλητών X-Y, θα ισχύει η σχέση, $\hat{Y}=a+bX$. Όταν η εξαρτημένη μεταβλητή Y θεωρούμε ότι είναι γραμμικά εξαρτημένη, επιπλέον, και από μία δεύτερη μεταβλητή (X2) ή και από μία τρίτη (X3) ή τελικά από ένα σύνολο m μεταβλητών X, η παραπάνω σχέση διαμορφώνεται σε:

$$\hat{Y}=a + b_1 X_1 + b_2 X_2 +...+b_m X_m$$

η πιο συνεπτυγμένα σε:

$$\hat{y} = a + \sum_{i=1}^{m} b_i x_i$$

Οι συντελεστές b1 , b2 ...bm καλούνται μερικοί συντελεστές . Ο μερικός συντελεστής b1 εκφράζει το μέγεθος μεταβολής της Y, όταν μεταβάλλεται η μεταβλητή X1 κατά μία μονάδα, ενώ παράλληλα οι υπόλοιπες μεταβλητές X_i διατηρούνται σταθερές στην τιμή του μέσου όρου τους. Ή αλλιώς, ο μερικός συντελεστής b1 εκφράζει τη μέτρηση της σχέσης μεταξύ Y και X1 , θέτοντας υπό έλεγχο ταυτόχρονα τις λοιπές μεταβλητές X_i ή αλλιώς της σχέσης Y και X1 , αφού προηγούμενα απαλειφθεί (απομακρυνθεί) το αποτέλεσμα των λοιπών μεταβλητών X_i επί της Y και X1 . Παρόμοια, ο συντελεστής b2 εκφράζει το βαθμό μεταβολής της Y, όταν μεταβάλλεται μόνο η X2 κοκ. Οι συντελεστές της πολλαπλής παλινδρόμησης καλούνται μερικοί, επειδή εκφράζουν μέρος μόνο της εξαρτημένης σχέσης της Y με τις μεταβλητές Xi . Η παράμετρος a είναι η τιμή της Y, όταν όλες οι μεταβλητές X_i είναι μηδενικές.

Παράδειγμα

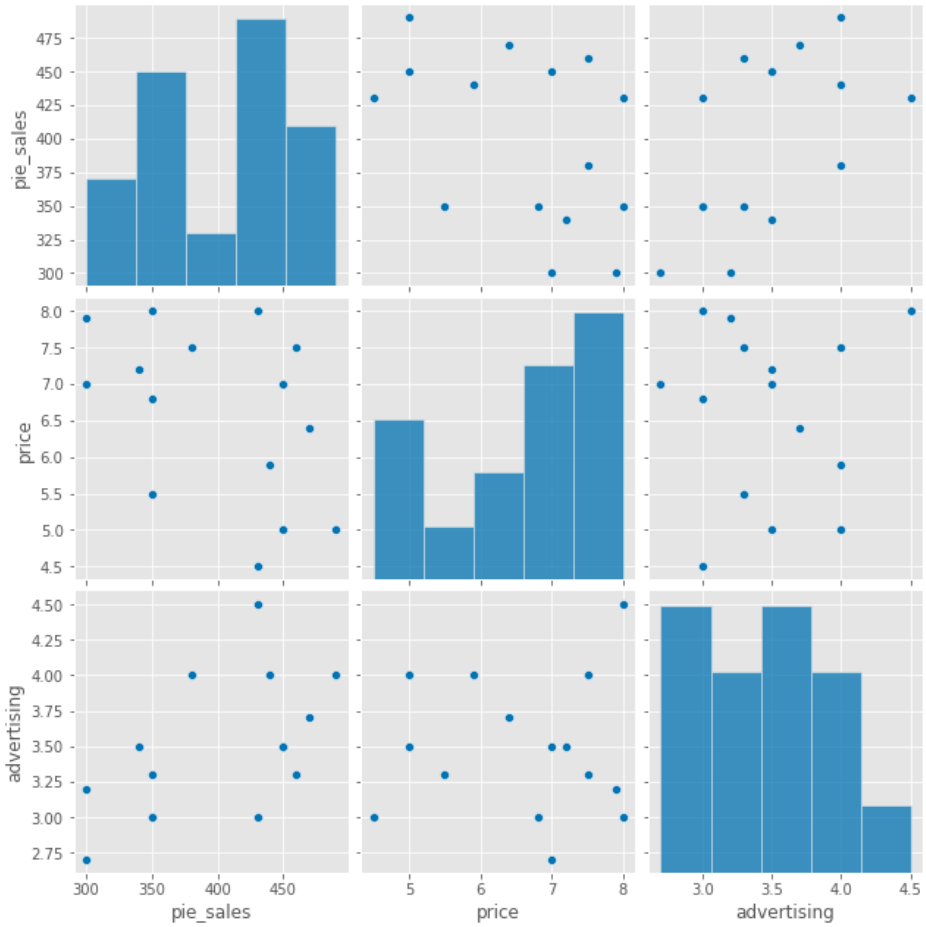
Η ανάλυση παλινδρόμησης είναι ένα εργαλείο για τη δημιουργία στατιστικών μοντέλων που χαρακτηρίζουν τις σχέσεις μεταξύ μιας εξαρτημένης μεταβλητής και μιας ή περισσότερων ανεξάρτητων μεταβλητών. Η απλή γραμμική παλινδρόμηση αναφέρεται στη μέθοδο που χρησιμοποιείται όταν υπάρχει μόνο μία ανεξάρτητη μεταβλητή, ενώ η πολλαπλή παλινδρόμηση αναφέρεται στη μέθοδο που χρησιμοποιείται όταν υπάρχουν περισσότερες από μία ανεξάρτητες μεταβλητές.

Έστω ότι έχουμε τα παρακάτω δεδομένα του πίνακα:

Εβδομάδα	Πωλήσεις Πίτας	Τιμές	Διαφημίσεις
1	350	5.5	3.3
2	460	7.5	3.3
3	350	8.0	3.0
4	430	8.0	4.5
5	350	6.8	3.0
6	380	7.5	4.0
7	430	4.5	3.0
8	470	6.4	3.7
9	450	7.0	3.5
10	490	5.0	4.0
11	340	7.2	3.5
12	300	7.9	3.2
13	440	5.9	4.0
14	450	5.0	3.5
15	300	7.0	2.7

Πίνακας.2: Σύνολο Δεδομένων Πολλαπλής Γραμμικής Παλινδρόμησης

Στο παρακάτω διάγραμμα με την οπτικοποίηση των δεδομένων για κάθε μεταβλητή μπορούμε να διακρίνουμε και κατά πόσο σχετίζεται η μία μεταβλητή με την άλλη. Εξετάζοντας την πρώτη σειρά των σχημάτων μπορούμε να δούμε ότι μπορεί να υπάρχουν σχέσεις μεταξύ τιμής, διαφήμισης και πωλήσεων. Σύμφωνα με το διάγραμμα διασποράς μεταξύ των πωλήσεων και του μοτίβου εμφάνισης της τιμής αρνητικής σχέσης, διακρίνουμε ότι όσο υψηλότερη είναι η τιμή τόσο χαμηλότερες θα είναι οι πωλήσεις. Αντιθέτως η διασπορά μεταξύ διαφήμισης και πωλήσεων εμφανίζει μια θετική σχέση, γεγονός που σημαίνει ότι όσο περισσότερα χρήματα ξοδευτούν για διαφήμιση τόσες περισσότερες πίτες θα πουληθούν.

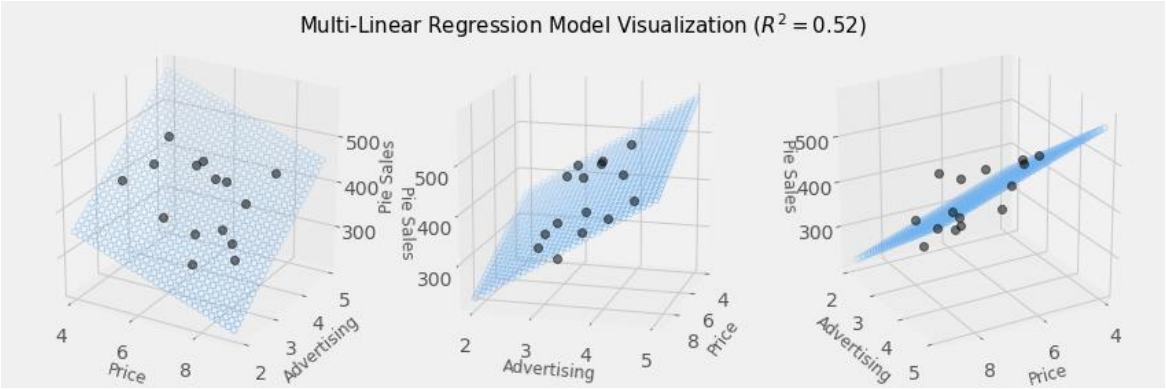


Εικόνα.4: Διάγραμμα Δεδομένων Πολλαπλής Γραμμικής Παλινδρόμησης

Εφόσον γνωρίζουμε ήδη ότι μπορεί να υπάρχουν σχέσεις μεταξύ των ανεξάρτητων και των εξαρτημένων μεταβλητών μας, χρησιμοποιώντας τη συνάρτηση της γραμμικής παλινδρόμησης καταλήγουμε στην παρακάτω εξίσωση:

$$\hat{y} = 306.5261 - 24.975 X_1 + 74.1309 X_2$$

Οι τιμές που προκύπτουν από την εξίσωση είναι οι τιμές τομής και συντελεστών των μοντέλων .Η τιμή τομής είναι η εκτιμώμενη μέση τιμή της εξαρτημένης μεταβλητής, όταν όλες οι τιμές των ανεξάρτητων μεταβλητών μας είναι 0. Συνεπώς για το παρόν πρόβλημα σημάνει ότι στην περίπτωση που πουληθούν πίτες στην τιμή 0 και ξοδευτούν διαφημιστικά έξοδα 0, θα πουληθούν περίπου 306 πίτες. Οι συντελεστές λαμβάνουν δύο τιμές για τις μεταβλητές τιμής και διαφήμισης αντίστοιχα. Αυτή η τιμή αντιπροσωπεύει τη σχέση της ανεξάρτητης μεταβλητής με την εξαρτημένη μεταβλητή, όπου μια αλλαγή ακριβώς μίας μονάδας στην ανεξάρτητη μεταβλητή θα αλλάξει την τιμή της εξαρτημένης μεταβλητής στο ίδιο ποσό με τον συντελεστή. Για παράδειγμα, αν αυξηθούν τα διαφημιστικά έξοδα κατά 10, θα αυξηθούν και οι πωλήσεις κατά περίπου 741 πίτες (74.1309 * 10).



Εικόνα.5: Διαγράμματα Οπτικοποίησης Πολλαπλής Γραμμικής Παλινδρόμησης

2.4 Κατηγοριοποίηση

Η κατηγοριοποίηση αποτελεί μια από τις βασικότερες τεχνικές της εξόρυξης δεδομένων, η οποία εξετάζοντας τα γνωρίσματα ενός στιγμιοτύπου το αντιστοιχεί σε μία προκαθορισμένη κλάση. Ένα μοντέλο κατηγοριοποίησης κατασκευάζεται λαμβάνοντας ως σύνολο εκπαίδευσης ένα πλήθος ταξινομημένων δεδομένων και χρησιμοποιώντας κατάλληλους αλγορίθμους καταφέρνει να αποδώσει σωστά την κατηγορία σε άγνωστα δεδομένα.

2.4.1 Λογιστική Παλινδρόμηση

Η λογιστική παλινδρόμηση (Logistic regression) 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) στοχεύεται η πρόβλεψη της έκβασης αυτής από ένα πλήθος προβλεπτικών μεταβλητών που μπορεί να είναι ονομαστικές, τακτικές ή ποσοτικές. Η σημαντικότερη διαφοροποίηση μεταξύ λογιστικής και γραμμικής παλινδρόμησης βασίζεται στη φύση της επιλεγμένης μεταβλητής απόκρισης, η οποία στην μεν πρώτη μπορεί να είναι κατηγορική, (τακτική ή ονομαστική, στη δε δεύτερη αποκλειστικά ποσοτική. Ενώ κατά την κλασική γραμμική παλινδρόμηση η εκτίμηση των παραμέτρων α και bi 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. Ως επακόλουθο, η πρώτη παραδέχεται την ύπαρξη ομοιογένειας (ομοσκεδαστικότητας) στα υπολείμματα των αποκρίσεων ενώ στη δεύτερη αναπτύσσεται πάντα ετεροσκεδαστικότητα σε κάθε προβλεπόμενη τιμή εξαιτίας του μεταβαλλόμενου ποσοστού διακύμανσης που αναλογεί σε αυτήν. Διακρίνονται τρεις τύποι λογιστικής παλινδρόμησης ανάλογα με την ιδιαίτερη φύση της εξαρτημένης κατηγορικής μεταβλητής η οποία μπορεί να είναι:

1. Δίτιμη ή δυαδική ή διχοτομική (binary) ή διμερής εξαρτημένη μεταβλητή. Συνίσταται από δύο κατηγορίες, όπως π.χ. είναι οι εκβάσεις επιτυχία/αποτυχία, ΝΑΙ/ΟΧΙ, γεγονός απόν/παρόν.

2. Τακτική (ordinal) μεταβλητή. Η εξαρτημένη μεταβλητή συνίσταται από τρεις ή περισσότερες κατηγορίες μεταξύ των οποίων ισχύει η έννοια της ανισότητας, όπως π.χ. σε μια ερώτηση της κλίμακας διαφωνώ καθόλου, λίγο, μέτρια, αρκετά, πολύ, στην κατάταξη ενός στρώματος υλικού ως λεπτού, μεσαίου, παχέος.

3. Ονομαστική (Nominal) ή πολυωνυμική (polynomial) ή πολυχοτομική (polychotomus) ή κατηγορική αδιαβάθμητη (non-ordered categorical) ή πολυμερής μεταβλητή απόκρισης. Περιέχει τρεις ή περισσότερες κατηγορίες χωρίς κάποια φυσική διαβάθμιση, όπως π.χ. ο χαρακτηρισμός ενός τρόφιμου ως τραγανού, μαλακού, εύθρυπτου ή του χρώματος αντικειμένων ως ερυθρού, πράσινου, κίτρινου κτλ.

Η λογιστική παλινδρόμηση επινοήθηκε ως εναλλακτική επιλογή της γραμμικής διακριτικής ανάλυσης για την ταξινόμηση των στοιχείων (ονομαστικών ή τακτικών) της εξαρτημένης, με ευρεία απήχηση σε πολλά διαφορετικά

επιστημονικά πεδία και κυρίως στην ιατρική και τις κοινωνικές επιστήμες. Χαρακτηριστικά, χρησιμοποιείται στην πρόβλεψη της:

• εμφάνισης ή μη μιας νόσου (π.χ. διαβήτης) από ένα σύνολο διαφορετικών χαρακτηριστικών του πάσχοντος ατόμου (ηλικία, φύλο, αιματολογικά, ηλεκτροκαρδιογράφημα κτλ.)

• επιλογής ενός πολιτικού κόμματος με βάση την καταγραφή των δημογραφικών στοιχείων των πολιτών, όπως είναι η ηλικία, φύλο, φυλή, τόπος διαμονής, εισόδημα, προηγούμενη ψηφοφορία

• πιθανότητας αποτυχίας μιας διεργασίας παραγωγής προϊόντος σε ένα εργοστάσιο τροφίμων

• πρόβλεψη της πρόθεσης αγοράς ενός αγαθού από έναν καταναλωτή (έρευνα αγοράς)

• πιθανότητας αθέτησης από δανειολήπτη της αποπληρωμής του δανείου του.

Λεπτομερής περιγραφή των μεθόδων της λογιστικής παλινδρόμησης παρέχεται από τα συγγράμματα των Cox & Snell (1989), των Hosmer & Lemeshow (2000), των Long & Freese (2014) και συνδυαστικά με τη χρήση των πινάκων ενδεχομένων από τους Everitt (1992) και Agresti (1996). Η κατανόηση των όρων και μαθηματικών τύπων που συνοδεύουν τη μελέτη της λογιστικής παλινδρόμησης αποτελεί κυριολεκτικά πρόκληση για τον απλό επιστήμονα.

Η εξίσωση της λογιστικής παλινδρόμησης ακολουθεί την εξίσωση της γραμμικής, δηλαδή:

$$y = b_0 + b_1 * x$$

αν πάνω σε αυτή την εξίσωση εφαρμόσουμε τη σιγμοειδή συνάρτηση έχουμε:

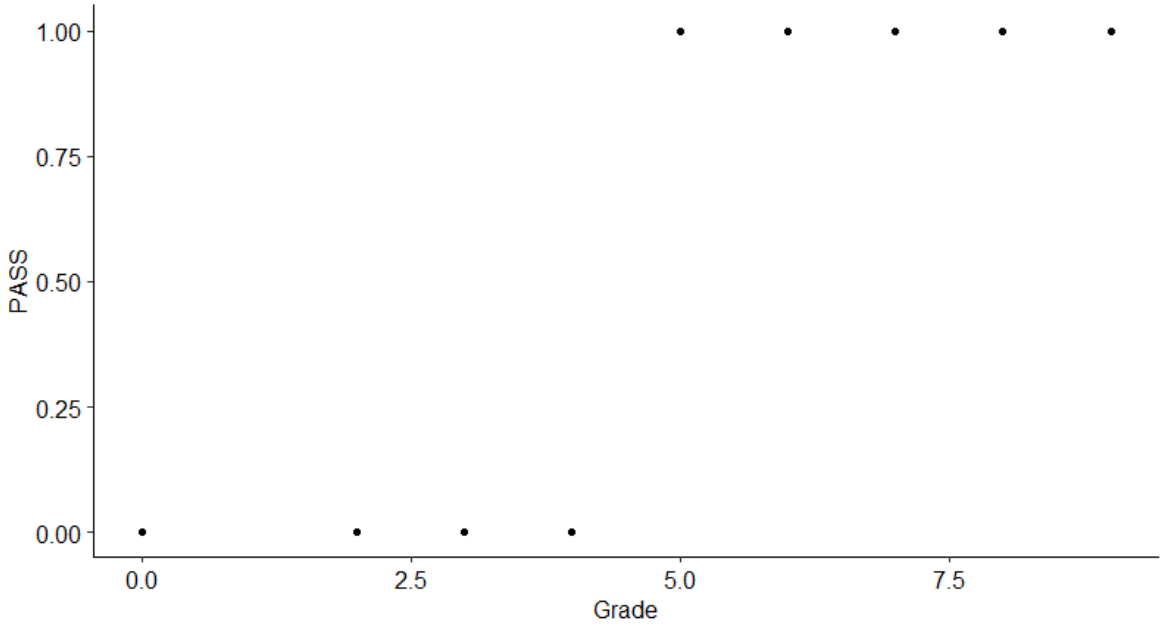
$$\sigma = 1\,1+e-y$$

λύνοντας ως προς y προκύπτει η σχέση:

$$ln(\,\sigma 1-\sigma\,) = b_0 + b_1 * x.$$

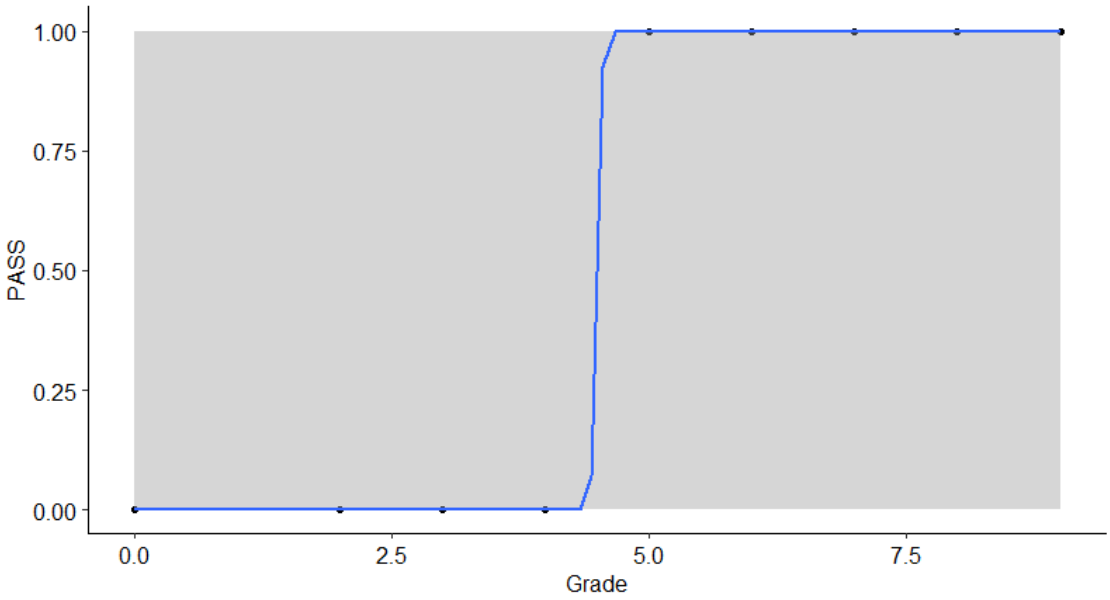
Παράδειγμα

Θεωρούμε ότι έχουμε 12 μαθητές που εξετάστηκαν σε διάφορα μαθήματα για την εισαγωγή τους με κατατακτήριες εξετάσεις. Από το σύνολο των βαθμών τους σε όλα τα μαθήματα που εξετάστηκαν υπολογίζεται ένας μέσος όρος και βάση αυτόν κρίνεται αν θα γίνει αποδεκτός ή όχι από το πανεπιστήμιο. Στο παρακάτω γράφημα βλέπουμε μία αναπαράσταση των δεδομένων μας. Μέσω της λογιστικής παλινδρόμησης αναζητούμε πιθανότητες, οι οποίες δεν υπερβαίνουν τα όρια του 0 και 1.



Εικόνα.6: Διάγραμμα Διασποράς Συνόλου Δεδομένων Λογιστικής Παλινδρόμησης

Αποτέλεσμα της εφαρμογής λογιστικής παλινδρόμησης:



Εικόνα.7: Διάγραμμα Λογιστικής Παλινδρόμησης

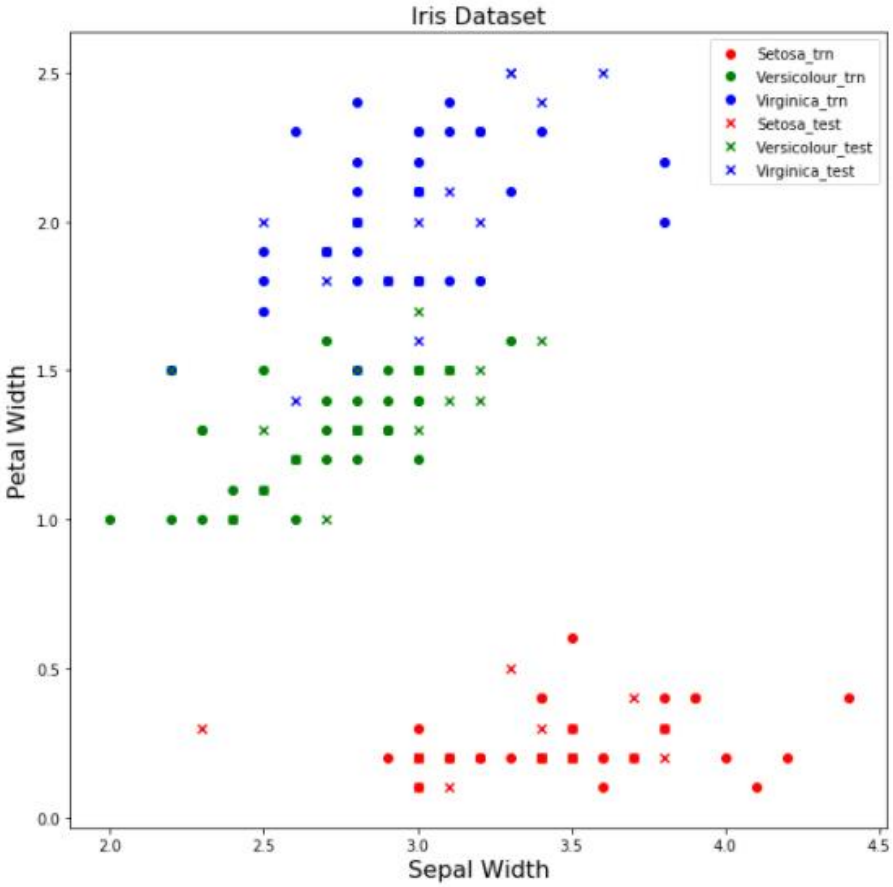
2.4.2 K-nearest neighbors

Ο αλγόριθμος k-πλησιέστερων γειτόνων (KNN) είναι ένας απλός, εύκολος στην εφαρμογή αλγόριθμος εποπτευόμενης μηχανικής μάθησης που μπορεί να χρησιμοποιηθεί για την επίλυση προβλημάτων τόσο ταξινόμησης όσο και παλινδρόμησης. Είναι εύκολο να εφαρμοστεί και να κατανοηθεί, αλλά έχει ένα σημαντικό μειονέκτημα ότι επιβραδύνεται σημαντικά καθώς αυξάνεται το μέγεθος αυτών των δεδομένων που χρησιμοποιούνται.

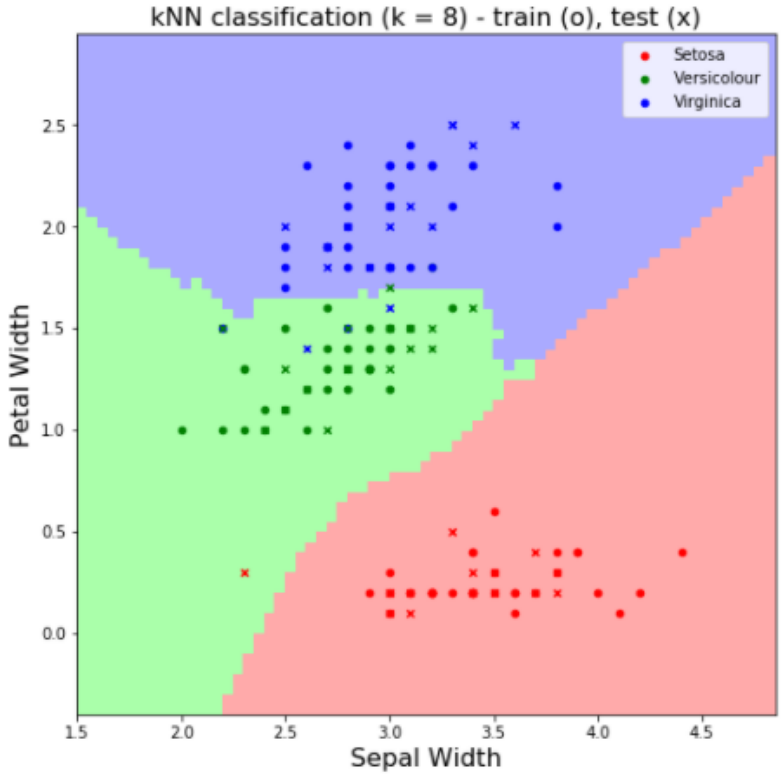
Το KNN εργάζεται βρίσκοντας τις αποστάσεις μεταξύ ενός ερωτήματος και όλων των παραδειγμάτων στα δεδομένα, επιλέγοντας τα καθορισμένα παραδείγματα αριθμών (K) που είναι πιο κοντά στο ερώτημα και, στη συνέχεια, ψηφίζει για την πιο συχνή ετικέτα (στην περίπτωση ταξινόμησης) ή υπολογίζει τον μέσο όρο των ετικετών (σε η περίπτωση της παλινδρόμησης). Στην περίπτωση της ταξινόμησης και της παλινδρόμησης, η επιλογή του σωστού K για τα δεδομένα γίνεται δοκιμάζοντας πολλά K και επιλέγοντας αυτό που λειτουργεί καλύτερα.

Παράδειγμα

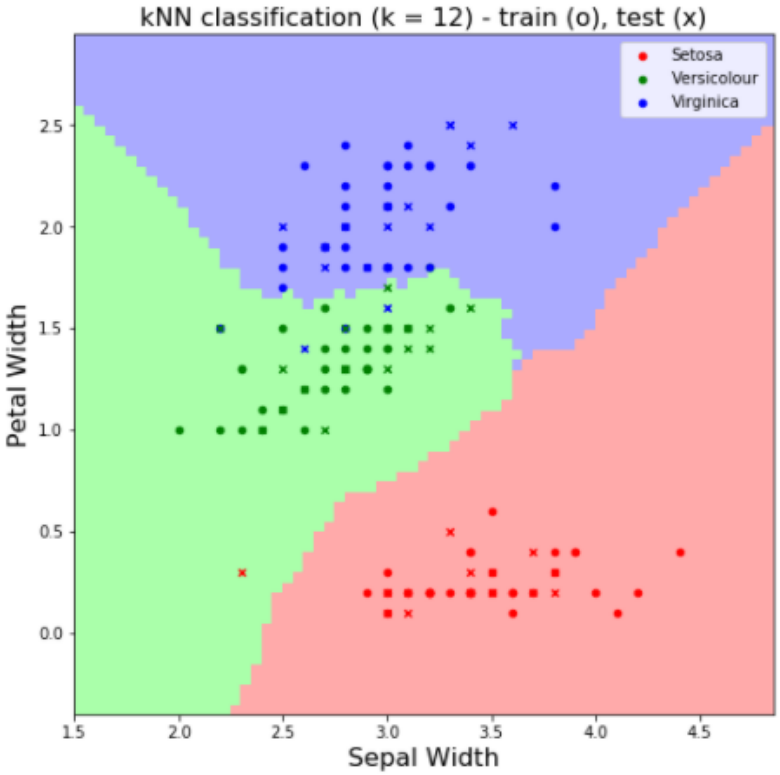
Έστω ότι εφαρμόζουμε τον αλγόριθμο KNN στο σύνολο δεδομένων IRIS το οποίο περιέχει 50 μετρήσεις για καθένα από τα τρία διαφορετικά είδη λουλουδιών: setosa, versicolor και virginica (συνολικά 150 δείγματα). Οι μετρήσεις αφορούν το μήκος και το πλάτος (σε cm) των πετάλων και των σέπαλων των λουλουδιών κάθε είδους.



Εικόνα.8:Διάγραμμα διασποράς που εμφανίζει κάθε είδος με βάση το μήκος σέπαλου (άξονας x) και το μήκος σέπαλου (άξονας y)



Εικόνα.9: Διάγραμμα διασποράς εκπαίδευσης και αξιολόγησης του ταξινομητή με κ=8 , ακρίβ εια εκπαίδευσης 0.97 και ακρίβεια διοκιμής 0.92.



Εικόνα.10: Διάγραμμα διασποράς εκπαίδευσης και αξιολόγησης του ταξινομητή με κ=8 , ακίβ εια εκπαίδευσης 0.97 και ακρίβεια διοκιμής 0.92.

2.4.3 Support Vector Machine

Ο αλγόριθμος Support Vector Machines, (SVM) είναι μία οικογένεια αλγορίθμων επιβλεπόμενης μάθησης που αναπτύχθηκαν από τον Vladimir Vapnik και χρησιμοποιούνται ευρύτατα σε προβλήματα κατάταξης. Ο αλγόριθμος SVM έχει το πλεονέκτημα να χειρίζεται πολύ καλά μεγάλο πλήθος χαρακτηριστικών και παρουσιάζει υψηλή απόδοση κατά την κατηγοριοποίηση αντικειμένων(αντικείμενο ορίζουμε μία γραμμή πίνακα (ΔΙΑΝΥΣΜΑ) που έχει ένα συγκεκριμένο πλήθος χαρακτηριστικών, χαρακτηριστικά ενός αντικειμένου είναι π.χ. πλάτος, ύψος και το βάρος ενός τραπεζιού) μεταξύ δύο (2) κατηγοριών. Ο αλγόριθμος SVM είναι ικανός στο να κατασκευάζει μοντέλα αρκετά πολύπλοκα για να επιλύει δύσκολα προβλήματα του πραγματικού κόσμου. Τα SVM μοντέλα που δημιουργούνται κατόπιν εκπαιδεύσεως του 80% περίπου των δεδομένων έχουν μια απλή λειτουργική μορφή και είναι λογικευμένα σε θεωρητικές αναλύσεις. Ο αλγόριθμος SVM περιλαμβάνει, ως ειδικές περιπτώσεις, ένα μεγάλο εύρος από νευρωνικά δίκτυα, ακτινικών συναρτήσεων (radial basis functions) και πολυωνυμικούς ταξινομητές. Σε προβλήματα κατηγοριοποίησης ο SVM μπορεί να χειριστεί στόχους με συνεχόμενο εύρος. Το SVM μοντέλο μαθαίνει την γραμμική αναδρομική συνάρτηση από τα εισαχθέντα δεδομένα προς εκπαίδευση και ακολούθως γίνεται η αυτόματη κατηγοριοποίηση των νέων «άγνωστων» δεδομένων. Ο SVM αλγόριθμος είναι ένα σύστημα που αναγνωρίζει κατηγορίες από πρότυπα και λέγεται ταξινομητής (classifier), ο ταξινομητής πρώτα εκπαιδεύεται (train) από το 70% - 80% των δεδομένων και στην συνέχεια ταξινομεί αυτόματα τα εναπομείναντα δεδομένα (test). Ο διαχωρισμός των θετικών παραδειγμάτων και των αρνητικών παραδειγμάτων επιτυγχάνεται μέσω του καταλληλότερου διαχωριστικού υπερεπιπέδου που επιλέγεται από τον αλγόριθμο SVM καθώς υπάρχουν πάρα πολλά διαχωριστικά υπερεπίπεδα μη κατάλληλα στο εκάστοτε πείραμα που πραγματοποιείται.

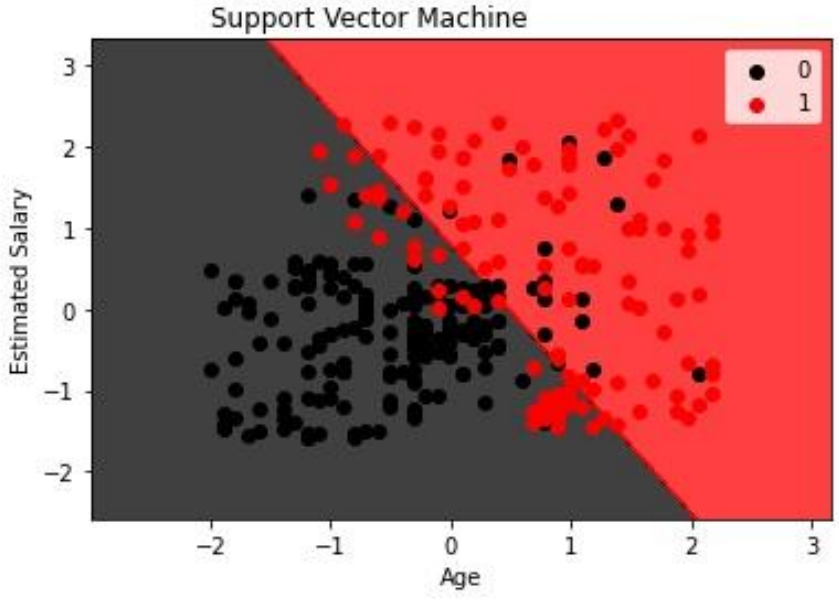
Παράδειγμα

Έστω πως έχουμε ένα σύνολο δεδομένων που περιλαμβάνει πληροφορίες για χρήστες μέσων κοινωνικής δικτύωσης και αν αγόρασαν ένα διαφημιζόμενο προϊόν ή όχι.

ID Χρήστη	Φύλο	Ηλικία	Εκτιμώμενος Μισθός	Αγορά
15624510	Άνδρας	19	19000	0
15810944	Άνδρας	35	20000	0
15668575	Γυναίκα	26	43000	0
15603246	Γυναίκα	27	57000	0
15804002	Άνδρας	19	76000	1
15728773	Άνδρας	27	58000	0
15598044	Γυναίκα	27	84000	0

Πίνακας.3: Σύνολο Δεδομένων Support Vector Machine

Εφαρμόζοντας τον αλγόριθμο SVM με γραμμικό τύπο πυρήνα πραγματοποιείτε ταξινόμηση στα δεδομένα του πίνακα με τα παρακάτω αποτελέσματα:



Εικόνα.11: Διάγραμμα διασποράς εφαρμογής αλγορίθμου SVM στα δεδομένα του παραδείγματος.

2.5 Συσταδοποίηση

Η συσταδοποίηση εμφανίζει πολλές ομοιότητες με την κατηγοριοποίηση και διαφέρει η βάση τους. Στην διαδικασία της συσταδοποίησης δεν γνωρίζουμε τι επιθυμούμε να προβλέψουμε, αλλά θέλουμε να δημιουργήσουμε και να αναγνωρίσουμε κάποια πρότυπα. Συνεπώς δημιουργούνται ομάδες δεδομένων από το σύνολο που ακολουθούν παρόμοιες συμπεριφορές.

2.5.1 K-means

Η ομαδοποίηση K-means είναι ένας τύπος μάθησης χωρίς επίβλεψη, που χρησιμοποιείται όταν υπάρχουν δεδομένα χωρίς ετικέτα (δηλαδή δεδομένα χωρίς καθορισμένες κατηγορίες ή ομάδες). Ο στόχος αυτού του αλγορίθμου είναι να βρει ομάδες στα δεδομένα, με τον αριθμό των ομάδων που αντιπροσωπεύονται από τη μεταβλητή K. Ο αλγόριθμος λειτουργεί επαναληπτικά για να εκχωρήσει κάθε σημείο δεδομένων σε μία από τις ομάδες K με βάση τα χαρακτηριστικά που παρέχονται. Τα σημεία δεδομένων ομαδοποιούνται με βάση την ομοιότητα χαρακτηριστικών. Τα αποτελέσματα του αλγόριθμου ομαδοποίησης K-means είναι:

- Τα κεντροειδή των συμπλεγμάτων K, τα οποία μπορούν να χρησιμοποιηθούν για την επισήμανση νέων δεδομένων
- Ετικέτες για τα δεδομένα εκπαίδευσης (κάθε σημείο δεδομένων εκχωρείται σε ένα μόνο σύμπλεγμα)

Κάθε κέντρο ενός συμπλέγματος είναι μια συλλογή από τιμές χαρακτηριστικών που καθορίζουν τις ομάδες που προκύπτουν. Η εξέταση των βαρών των κεντροειδών χαρακτηριστικών μπορεί να χρησιμοποιηθεί για την ποιοτική ερμηνεία του είδους της ομάδας που αντιπροσωπεύει κάθε σύμπλεγμα.

Ο αλγόριθμος ομαδοποίησης K-means χρησιμοποιείται για την εύρεση ομάδων που δεν έχουν επισημανθεί ρητά στα δεδομένα. Αυτό μπορεί να χρησιμοποιηθεί για την επιβεβαίωση επιχειρηματικών υποθέσεων σχετικά με τους τύπους ομάδων που υπάρχουν ή για τον εντοπισμό άγνωστων ομάδων σε πολύπλοκα σύνολα δεδομένων. Μόλις εκτελεστεί ο αλγόριθμος και οριστούν οι ομάδες, όλα τα νέα δεδομένα μπορούν εύκολα να αντιστοιχιστούν στη σωστή ομάδα. Αυτός είναι ένας ευέλικτος αλγόριθμος που μπορεί να χρησιμοποιηθεί για κάθε τύπο ομαδοποίησης. Μερικά παραδείγματα περιπτώσεων χρήσης είναι:

Ταξινόμηση μετρήσεων αισθητήρων:

- Ανίχνευση τύπων δραστηριότητας σε αισθητήρες κίνησης
- Ομαδικές εικόνες
- Ξεχωριστός ήχος
- Προσδιορίστε ομάδες στην παρακολούθηση της υγείας

Ανίχνευση ρομπότ ή ανωμαλιών:

- Διαχωρίστε έγκυρες ομάδες δραστηριοτήτων από bots
- Ομαδοποιήστε έγκυρη δραστηριότητα για να καθαρίσετε τον εντοπισμό ακραίων τιμών

Επιπλέον, η παρακολούθηση εάν ένα σημείο δεδομένων παρακολουθείται αλλάζει μεταξύ ομάδων με την πάροδο του χρόνου μπορεί να χρησιμοποιηθεί για τον εντοπισμό σημαντικών αλλαγών στα δεδομένα. Ο αλγόριθμος ομαδοποίησης K-means χρησιμοποιεί επαναληπτική βελτίωση για να παράγει ένα τελικό αποτέλεσμα. Οι είσοδοι του αλγορίθμου είναι ο αριθμός των συστάδων K και το σύνολο δεδομένων. Το σύνολο δεδομένων είναι μια συλλογή χαρακτηριστικών για κάθε σημείο δεδομένων. Οι αλγόριθμοι ξεκινούν με αρχικές εκτιμήσεις για τα K κεντροειδή, τα οποία μπορούν είτε να

δημιουργηθούν τυχαία είτε να επιλεγούν τυχαία από το σύνολο δεδομένων. Στη συνέχεια, ο αλγόριθμος επαναλαμβάνεται μεταξύ δύο βημάτων:

1. Βήμα ανάθεσης δεδομένων:

Κάθε κέντρο ορίζει ένα από τα συμπλέγματα. Σε αυτό το βήμα, κάθε σημείο δεδομένων εκχωρείται στο πλησιέστερο κέντρο του, με βάση το τετράγωνο της Ευκλείδειας απόστασης. Πιο τυπικά, εάν c_i είναι η συλλογή των κεντροειδών στο σύνολο C, τότε κάθε σημείο δεδομένων x εκχωρείται σε ένα σύμπλεγμα με βάση:

$$\arg\min_{c_i \in C} dist(C_i, x)^2$$

όπου dist(·) είναι η τυπική (L2) Ευκλείδεια απόσταση. Έστω το σύνολο των εκχωρήσεων σημείων δεδομένων για κάθε κέντρο συμπλέγματος i^{th} να είναι S_i.

2. Βήμα ενημέρωσης Centroid:

Σε αυτό το βήμα, τα κεντροειδή υπολογίζονται εκ νέου. Αυτό γίνεται λαμβάνοντας τον μέσο όρο όλων των σημείων δεδομένων που έχουν εκχωρηθεί στο σύμπλεγμα αυτού του κέντρου.

$$C_i = \frac{1}{|S_i|} \sum x_i \in S_i^{x_i}$$

Ο αλγόριθμος επαναλαμβάνεται μεταξύ των βημάτων ένα και δύο μέχρι να ικανοποιηθεί ένα κριτήριο διακοπής (δηλαδή, κανένα σημείο δεδομένων δεν αλλάζει συστάδες, το άθροισμα των αποστάσεων ελαχιστοποιείται ή επιτυγχάνεται κάποιος μέγιστος αριθμός επαναλήψεων). Αυτός ο αλγόριθμος είναι εγγυημένο ότι συγκλίνει σε ένα αποτέλεσμα. Το αποτέλεσμα μπορεί να είναι ένα τοπικό βέλτιστο (δηλαδή όχι απαραίτητα το καλύτερο δυνατό αποτέλεσμα), που σημαίνει ότι η αξιολόγηση περισσότερων από μία εκτέλεσης του αλγορίθμου με τυχαιοποιημένα αρχικά κεντροειδή μπορεί να δώσει καλύτερο αποτέλεσμα.

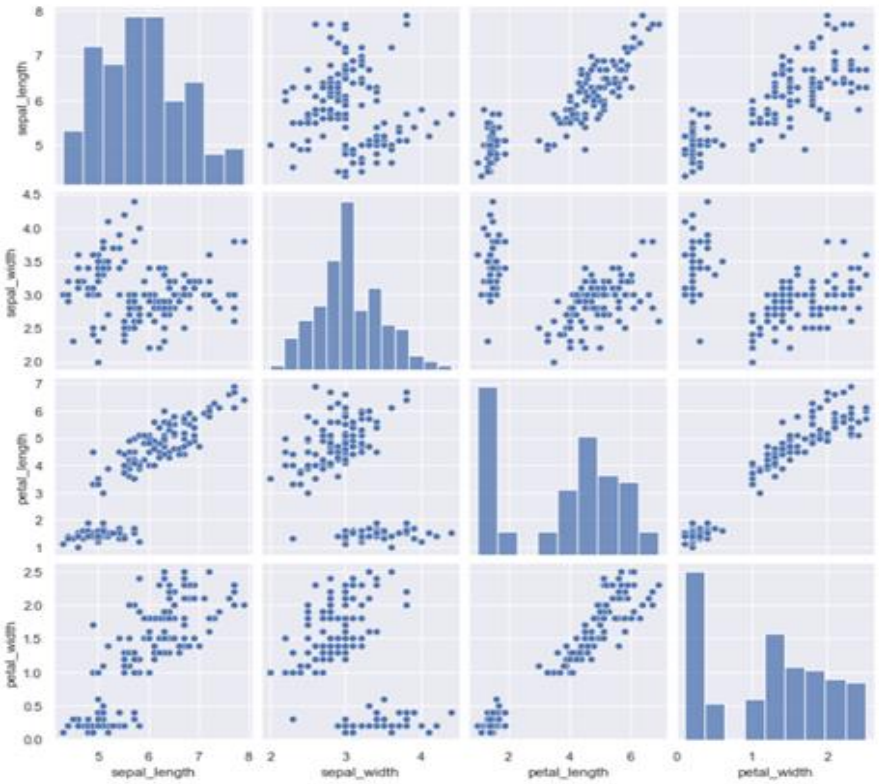
Ο αλγόριθμος που περιγράφεται παραπάνω βρίσκει τις ετικέτες συστάδων και συνόλων δεδομένων για ένα συγκεκριμένο προεπιλεγμένο K. Για να βρει τον αριθμό των συστάδων στα δεδομένα, ο χρήστης πρέπει να εκτελέσει τον αλγόριθμο ομαδοποίησης K-means για μια περιοχή τιμών K και να συγκρίνει τα αποτελέσματα . Γενικά, δεν υπάρχει μέθοδος για τον προσδιορισμό της ακριβούς τιμής του K, αλλά μια ακριβής εκτίμηση μπορεί να ληφθεί χρησιμοποιώντας τις ακόλουθες τεχνικές. Μία από τις μετρήσεις που χρησιμοποιείται συνήθως για τη σύγκριση αποτελεσμάτων σε διαφορετικές

τιμές του K είναι η μέση απόσταση μεταξύ των σημείων δεδομένων και του κέντρου συστάδας τους. Δεδομένου ότι η αύξηση του αριθμού των συστάδων θα μειώνει πάντα την απόσταση από τα σημεία δεδομένων, η αύξηση του K θα μειώνει πάντα αυτή τη μέτρηση, στο άκρο να φτάσει στο μηδέν όταν το K είναι το ίδιο με τον αριθμό των σημείων δεδομένων. Επομένως, αυτή η μέτρηση δεν μπορεί να χρησιμοποιηθεί ως ο μοναδικός στόχος. Αντίθετα, η μέση απόσταση από το κέντρο ως συνάρτηση του K απεικονίζεται γραφικά και το "σημείο αγκώνα", όπου ο ρυθμός μείωσης μετατοπίζεται απότομα, μπορεί να χρησιμοποιηθεί για τον χονδρικό προσδιορισμό του K.

Υπάρχουν πολλές άλλες τεχνικές για την επικύρωση του K, συμπεριλαμβανομένης της διασταυρούμενης επικύρωσης(cross-validation), των κριτηρίων πληροφοριών(information criteria), της μεθόδου θεωρητικού άλματος πληροφοριών(the information theoretic jump method), της μεθόδου σιλουέτας(the silhouette method) και του αλγόριθμου G-means. Επιπλέον, η παρακολούθηση της κατανομής των σημείων δεδομένων μεταξύ των ομάδων παρέχει μια εικόνα για το πώς ο αλγόριθμος διαχωρίζει τα δεδομένα για κάθε

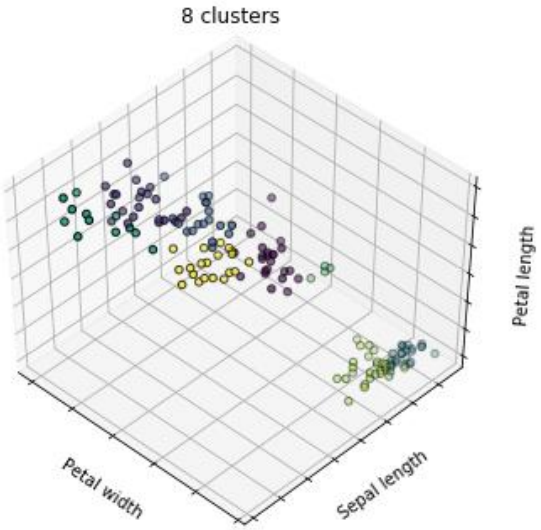
Παράδειγμα

Έστω ότι έχουμε το σύνολο δεδομένων IRIS και πραγματοποιούμε συσταδοποίηση με τον αλγόριθμο K-means με μεγέθη συστάδας 8, 3 και 4:

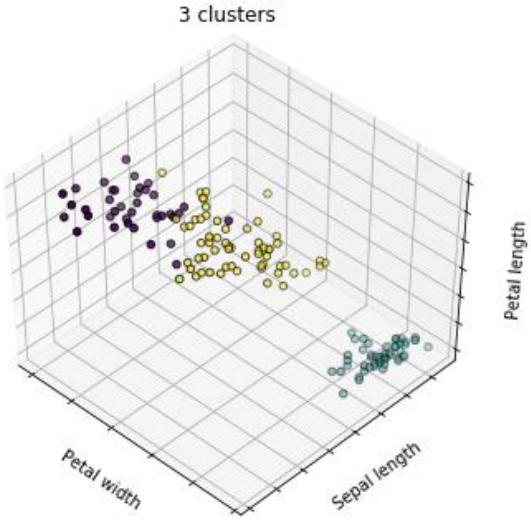


Εικόνα.12: Διάγραμμα Διασποράς του συνόλου δεδομένων

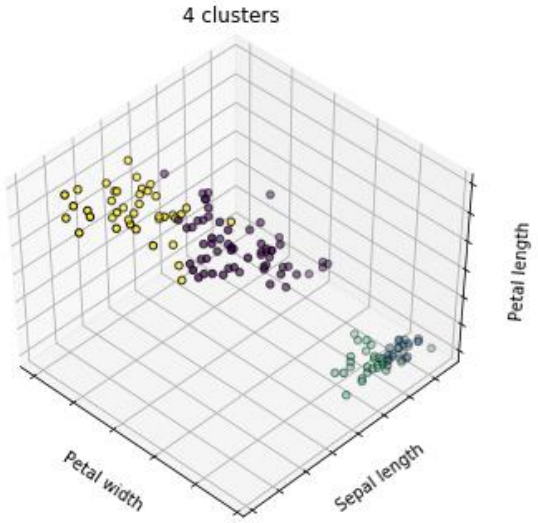
Ένα διάγραμμα διασποράς του συνόλου δεδομένων IRIS που υποδεικνύει συσχετίσεις μεταξύ των μεταβλητών. Τα χρώματα υποδεικνύουν τις τρεις κατηγορίες στα δεδομένα και επίσης υπάρχει ένα ιστόγραμμα της μεταβλητής που περιλαμβάνεται στη διαγώνιο.



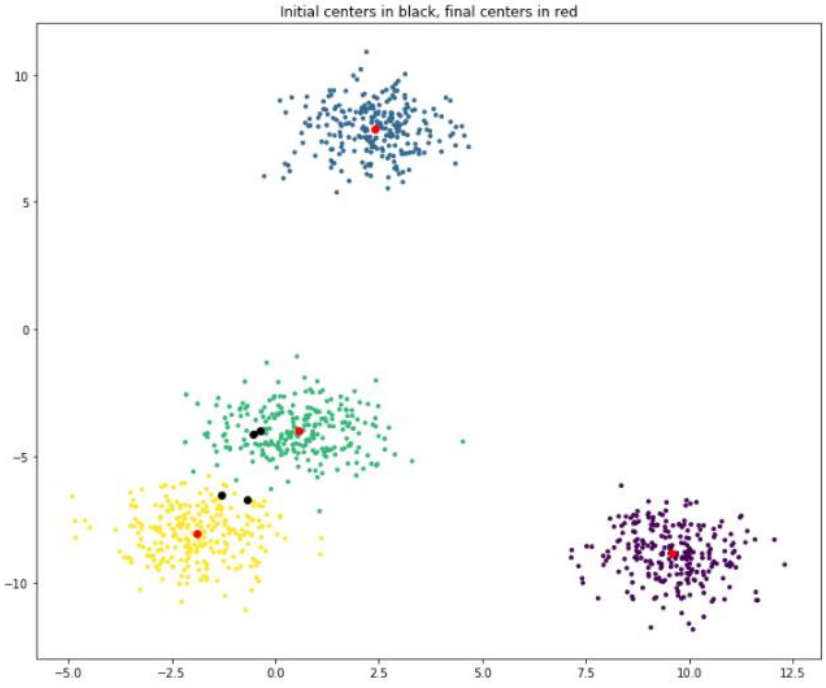
Εικόνα.13:Εφαρμογή αλγορίθμου K-means στο σύνολο δεδομένων IRIS με μέγεθος συστάδας 8.



Εικόνα.14:Εφαρμογή αλγορίθμου K-means στο σύνολο δεδομένων IRIS με μέγεθος συστάδας 3.



Εικόνα.15:Εφαρμογή αλγορίθμου K-means στο σύνολο δεδομένων IRIS με μέγεθος συστάδας 4.



Εικόνα.16:Διάγραμμα διασποράς με εφαρμογή αλγορίθμου K-means και παρουσίαση των νέων κεντροειδών.

Κεφάλαιο 3

3.1 Βαθιά Μηχανική Μάθηση

Η βαθιά μηχανική μάθηση αποτελεί έναν τύπο μοντέλου εκπαίδευσης μηχανικής μάθησης που λειτουργεί πλησιέστερα στον τρόπο με τον οποίο λαμβάνει αποφάσεις ο ανθρώπινος εγκέφαλος, δηλαδή ότι οι αλγόριθμοι που δημιουργούνται είναι περισσότερο πολύπλοκοι και περιλαμβάνουν περισσότερα στρώματα. Συνεπώς αντί για ένα μόνο στρώμα, χρησιμοποιούνται πολλαπλά περίπλοκα στρώματα για επεξεργασία των δεδομένων. Η ενσωμάτωση των στρωμάτων απαιτεί την εφαρμογή ενός νευρωνικού δικτύου, το οποίο αποτελεί ένα σύστημα που επιτρέπει την επικοινωνία μεταξύ των επιπέδων. Αυτή η διαδικασία θα λέγαμε ότι παρουσιάζει περισσότερα κοινά με τη μάθηση χωρίς επίβλεψη, καθώς αποτελεί επίσης μία αυτοματοποιημένη διαδικασία.

Θα μπορούσαμε να παρομοιάσουμε τη λειτουργία του συγκεκριμένου τύπου μάθησης ως έναν ισχυρό εγκέφαλο, για τον οποίο απαιτείται μεγάλος όγκος δεδομένων για εκπαίδευση. Απαιτούνται τόσα πολλά δεδομένα που πριν από την επιστήμη του big data και toy cloud computing, οι ποσότητες δεδομένων και η επεξεργασία τους δεν αποτελούσε εύκολη υπόθεση. Σε αυτό το σημείο ας επισημάνουμε ότι το γεγονός της ύπαρξης πολλών δεδομένων, δεν σημαίνει απαραίτητα ότι τα δεδομένα πρέπει να είναι δομημένα. Η βαθιά μηχανική μάθηση έχει τη δυνατότητα να επεξεργάζεται τόσο μη επισημασμένα όσο και μη δομημένα δεδομένα. Επίσης η συγκεκριμένη μέθοδος μάθησης δημιουργεί αρκετά πιο σύνθετα στατιστικά μοντέλα. Με κάθε νέο στοιχείο δεδομένων, το μοντέλο γίνεται πιο σύνθετο, αλλά και πιο ακριβές.

Η βαθιά μηχανική μάθηση χωρίζεται σε τρείς κατηγορίες με την πρώτη να αναφέρεται ως Transfer Learning. Αν συγκριθεί με άλλα μοντέλα εκπαίδευσης, χρειάζεται λιγότερο χρόνο για υπολογισμό καθώς και ότι δεν απαιτεί σχεδόν τόσα δεδομένα. Το Transfer Learning επεξεργάζεται τα δεδομένα με ένα προϋπάρχον δίκτυο το οποίο απαιτείται να έχει μια διεπαφή για να μπορέσει να χρησιμοποιηθεί . Στη συνέχεια, το δίκτυο τροφοδοτείται με νέα δεδομένα, δηλαδή με σενάρια που δεν έχει ακόμη αντιμετωπίσει για να λάβει την αντίδραση και να μάθει από τα δεδομένα. Με την πάροδο του χρόνου, μπορούν να γίνουν προσαρμογές για να ταιριάζουν στα νέα δεδομένα καθώς και να ενσωματωθούν τα παλιά δεδομένα. Στη συνέχεια, οι νέες εργασίες μπορούν να εκτελεστούν σύμφωνα με όσα έχει μάθει το δίκτυο και μπορούν ακόμη και να αποκτήσουν πιο συγκεκριμένες ικανότητες κατηγοριοποίησης με τη νέα προσαρμοστικότητά του. Ουσιαστικά, απαιτείται μια υπάρχουσα διεπαφή, η οποία πρέπει να τροφοδοτείται με νέα σενάρια και αυτές οι νέες συνθήκες βελτιώνουν το δίκτυο ώστε να είναι πιο διευρυμένο όσον αφορά την αποδοχή δεδομένων, αλλά πιο επιλεκτικό και συγκεκριμένο όσον αφορά την κατηγοριοποίηση.

Μια άλλη μέθοδος βαθιάς μηχανικής μάθησης είναι η μέθοδος Dropout. Για παράδειγμα στην εποπτευόμενη μάθηση, μπορούν να δημιουργηθούν κατηγορίες έτσι ώστε να περιλαμβάνουν δεδομένα που δεν χρειάζεται να

επεξεργαστούν. Με απλά λόγια, η υπερπροσαρμογή(overfitting) είναι ένα φαινόμενο το οποίο λαμβάνει χώρα όταν οι κατηγορίες είναι πολύ συγκεκριμένες στο σημείο που έχουν υποχωρήσει πάρα πολύ. Για την καταπολέμηση της υπερπροσαρμογής, η μέθοδος εγκατάλειψης (dropout) επιλέγει τυχαίους κόμβους/μονάδες στο νευρωνικό δίκτυο για «αποχώρηση». Με την απαλλαγή από αυτούς τους τυχαίους κόμβους, παλεύει να εξαλείψει ταξινομήσεις που ξεπερνούν υπερβολικά τα όρια. Η διαδικασία αυτή λαμβάνει χώρα κατά τη διάρκεια της προπόνησης.

Η επόμενη μέθοδος βαθιάς μηχανικής μάθησης είναι η εκπαίδευση από την αρχή. Σύμφωνα με τη συγκεκριμένη μέθοδο ο χρήστης δημιουργεί και εκπαιδεύει το δίκτυο από το τίποτα. Η αρχιτεκτονική δικτύου δημιουργείται σταδιακά σύμφωνα με τις ανάγκες και απαιτείται ένα σύνολο δεδομένων με ετικέτα για τη διαμόρφωση αυτού του δικτύου. Τουλάχιστον στα αρχικά στάδια της προπόνησης, απαιτείται τα δεδομένα να έχουν ετικέτα, διότι δεν μπορούν να συμπεριληφθούν στο σετ εκπαίδευσης,. Για να συλληφθούν όλα τα σενάρια ή τουλάχιστον η πλειοψηφία, απαιτείται μεγάλος όγκος δεδομένων για τη δοκιμή και επίσης περισσότερος χρόνος συγκριτικά με άλλες μεθόδους, επομένως δεν αποτελεί τόσο συνηθισμένη μέθοδο.

Η τελευταία μέθοδος βαθιάς μηχανικής μάθησης είναι η μέθοδος μείωσης του ρυθμού μάθησης, γνωστή και ως μέθοδος προσαρμοστικών ρυθμών μάθησης. Ο ρυθμός εκμάθησης καθορίζει πόση αλλαγή θα συμβεί μόλις μετρηθεί το ποσοστό σφάλματος. Παρακολουθώντας τον ρυθμό εκμάθησης, ο χρήστης μπορεί να προσαρμόσει το μοντέλο για να αυξηθεί η απόδοση και να μειωθεί ο χρόνος εκπαίδευσης. Με την πάροδο του χρόνου, ο στόχος είναι να μειωθεί ο ρυθμός εκμάθησης, έτσι ώστε να μην χρειάζεται να αλλάξουν πολλά όταν αποστέλλονται δεδομένα μέσω του μοντέλου. Εάν τα ποσοστά είναι πολύ υψηλά, έχουμε πάρα πολλές αλλαγές στο δίκτυο γεγονός που το κάνει ασταθές κατά τη διάρκεια της προπόνησης. Αλλά αν το ποσοστό είναι πολύ μικρό, τότε η απόδοση μπορεί να φτάσει σε σημείο που τα δεδομένα θα μπορούσαν ακόμη και να κολλήσουν κατά τη διάρκεια της προπόνησης. Με την εξισορρόπηση του ρυθμού εκμάθησης, η ακρίβεια μπορεί να βελτιωθεί και η απόδοση της εκπαίδευσης μπορεί να αυξηθεί σημαντικά.

3.2 Τεχνητά Νευρωνικά Δίκτυα

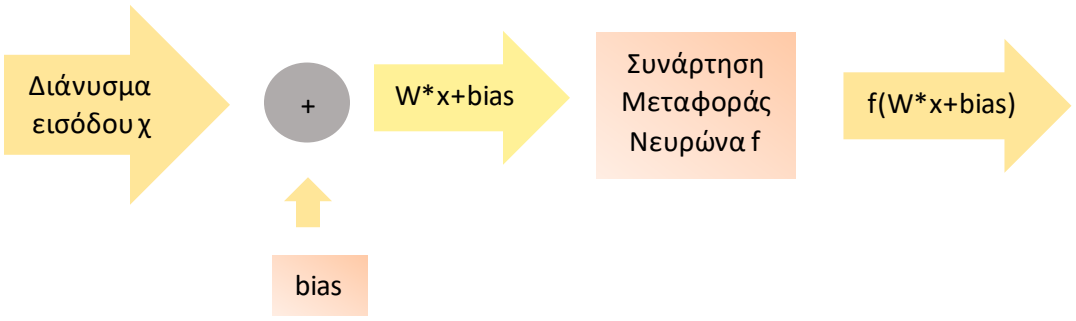
Τα νευρωνικά δίκτυα είναι σημαντικά ως πηγή επικοινωνίας διότι οι κόμβοι, οι οποίοι περιέχουν τα επίπεδα των αλγορίθμων, μπορούν να αξιολογήσουν ακατέργαστα δεδομένα και να βρουν κρυφά μοτίβα. Όπως η μάθηση χωρίς επίβλεψη, η βαθιά μάθηση θα αλλάξει και θα προσαρμοστεί στα δεδομένα με την πάροδο του χρόνου, γεγονός που της επιτρέπει να βελτιώνεται με την πάροδο του χρόνου. Τα νευρωνικά δίκτυα κατασκευάζονται επίσης για πιο σύνθετα ζητήματα και λύνουν ακόμη και σενάρια πραγματικής ζωής.

Αλλά τι ακριβώς υποτίθεται ότι είναι ένας κόμβος; Ένας κόμβος υποτίθεται ότι αντιπροσωπεύει την είσοδο στην οποία θα πρέπει να ληφθεί μια απόφαση, η οποία στη συνέχεια συνδέεται με κρυφά επίπεδα. Αυτά τα κρυφά επίπεδα

αποτελούν μέρος του δικτύου λήψης αποφάσεων. Μόλις ληφθεί μια απόφαση, συνδέεται με το επίπεδο εξόδου, το οποίο εμφανίζει το αποτέλεσμα αυτής της απόφασης. Ακριβώς όπως οι νευρώνες στον εγκέφαλο, η είσοδος μπορεί να αναπηδήσει γύρω από το κρυφό στρώμα δικτύου, πυροδοτώντας κόμβους όπως οι νευρώνες πυροδοτούν αντιδράσεις στον εγκέφαλο. Αυτό ενθαρρύνει τη λήψη πιο ακριβών και περιγραφικών αποφάσεων.

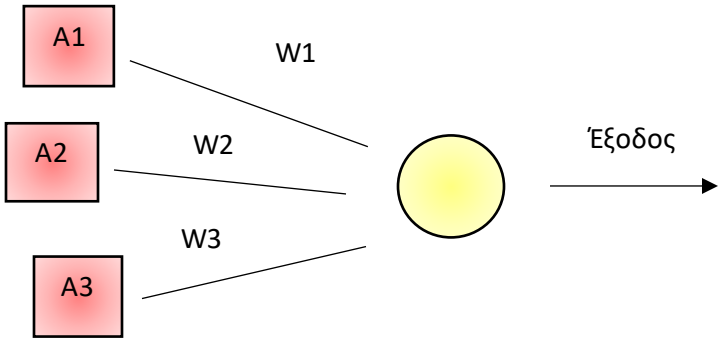
3.2.1 Ο νευρώνας

Τα τεχνικά νευρωνικά δίκτυα αποτελούν μια προσπάθεια μαθηματικής προσομοίωσης της λειτουργίας του ανθρώπινου εγκεφάλου σε σχέση με τον τρόπο λήψης αποφάσεων. Αποτελούνται λοιπόν από ένα σύνολο νευρώνων (στοιχείων) οι οποίοι συνδέονται μεταξύ τους. Όπως και στα βιολογικά νευρωνικά δίκτυα ολόκληρη η λειτουργία του δικτύου καθορίζεται από τις συνδέσεις μεταξύ των νευρώνων (στοιχείων). Το κάθε στοιχείο – νευρώνας υλοποιεί μια συνάρτηση μεταφοράς f λαμβάνοντας μια είσοδο και αποδίδοντας μια έξοδο, η οποία ταυτόχρονα μπορεί να είναι είσοδος για κάποιον άλλον νευρώνα. Σχηματικά η λειτουργία του νευρώνα είναι:



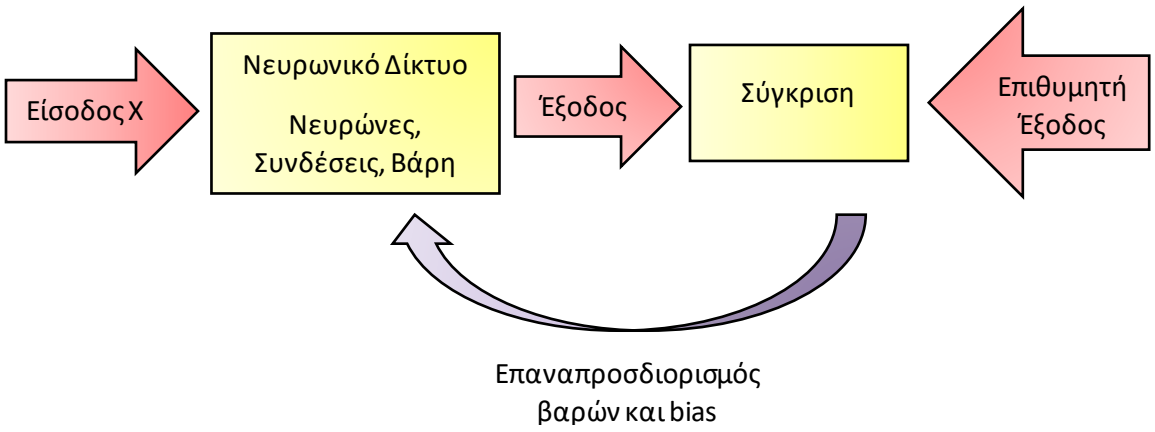
Εικόνα.17: Λειτουργία του νευρώνα

Ένας νευρώνας λαμβάνει στην γενική περίπτωση ως είσοδο ένα διάνυσμα x. Το κάθε στοιχείο του διανύσματος εισέρχεται σταθμισμένο πολλαπλασιάζεται με έναν συντελεστή w το οποίο το ονομάζουμε βάρος και αθροίζεται με τα υπόλοιπα στοιχεία του διανύσματος καθώς και με έναν συντελεστή ο οποίος ονομάζεται bias. Σε ένα νευρωνικό δίκτυο η κάθε σύνδεση μεταξύ νευρώνων καθορίζεται από το βάρος της σύνδεσης w. Ως είσοδος στον νευρώνα λοιπόν εισέρχεται η παράσταση ($x_1*w_1+x_2*w_2+"...+x_n*w_n+b$).Το bias μπορούμε να το θεωρήσουμε ως μια μοναδιαία είσοδο πολλαπλασιαζόμενη με τον συντελεστή b. Στον νευρώνα υλοποιείται η συνάρτηση μεταφοράς f και λαμβάνουμε την τελική έξοδο του.



Εικόνα.18: Συνάρτηση Μεταφοράς και Τελική Έξοδος Νευρώνα

Ένα νευρωνικό δίκτυο αποτελείται από ένα σύνολο νευρώνων οι οποίοι συνδέονται μεταξύ τους και η ταυτότητα της σύνδεσης καθορίζεται από το βάρος w. Στα πολυεπίπεδα νευρωνικά δίκτυα έχουμε διαφορετικά επίπεδα όπου το καθένα αποτελείται από ένα σύνολο νευρώνων. Η εκπαίδευση ενός νευρωνικού δικτύου συνίσταται στο να ρυθμίσουμε τις συνδέσεις του δικτύου προκειμένου να υλοποιήσουμε μια συγκεκριμένη συνάρτηση και το δίκτυο να μας δώσει ή να πλησιάσει μια επιθυμητή έξοδο. Η ρύθμιση των συνδέσεων γίνεται με την ρύθμιση των βαρών. Στοιχείο της εκπαίδευσης του δικτύου είναι και η ρύθμιση των bias. Θα πρέπει να αναφέρουμε ότι σε ένα νευρωνικό δίκτυο δεν είναι απαραίτητο ο αριθμός των νευρώνων που αποτελούν το κάθε επίπεδο να είναι ο ίδιος. Επίσης δεν είναι απαραίτητο όλοι οι νευρώνες να υλοποιούν την ίδια συνάρτηση μεταφοράς. Στο σχήμα παρακάτω φαίνεται η διαδικασία εκπαίδευσης ενός δικτύου με την αναπροσαρμογή των συνδέσεων μεταξύ των νευρώνων:



Εικόνα. 19: Διαδικασία Εκπαίδευσης ενός Δικτύου με την Αναπροσαρμογή των Συνδέσεων μεταξύ των Νευρώνων

3.3 Συνελικτικά Νευρωνικά Δίκτυα

Ένας τομέας στον οποίο η βαθιά μηχανική μάθηση έχει σημειώσει θεαματική επιτυχία είναι η επεξεργασία εικόνας. Αυτό επιτυγχάνεται με τη χρήση των συνελικτικών δικτύων, που επιτρέπουν τη προσθήκη περισσότερων επιπέδων στο δίκτυο και χρήση της ανάστροφης διάδοσης για την εκμάθηση των βαρών. Όμως ο αριθμός των βαρών αυξάνεται υπερβολικά και, κατά συνέπεια, ο όγκος των δεδομένων εκπαίδευσης που απαιτείται για την επίτευξη ικανοποιητικής ακρίβειας μπορεί να καταστεί υπερβολικά μεγάλος και, ως εκ τούτου, μη ρεαλιστικός.

Τα δίκτυα που περιλαμβάνουν συνελικτικά επίπεδα ονομάζονται συνελικτικά νευρωνικά δίκτυα (ΣΝΔ). Η βασική ιδιότητά τους είναι ότι μπορούν να εντοπίζουν χαρακτηριστικά εικόνων όπως φωτεινά ή σκοτεινά σημεία (ή συγκεκριμένο χρώμα), τις ακμές σε διάφορους προσανατολισμούς, πρότυπα, και ούτω καθεξής. Τα εν λόγω χαρακτηριστικά αποτελούν τη βάση για τον εντοπισμό πιο αφηρημένων χαρακτηριστικών όπως τα αυτιά μιας γάτας, το ρύγχος του σκύλου, το μάτι ενός ανθρώπου, ή το οκταγωνικό σχήμα ενός σήματος στοπ. Κανονικά, η εκπαίδευση ενός νευρωνικού δικτύου με σκοπό να εντοπίζει αυτού του είδους τα χαρακτηριστικά με βάση τα εικονοστοιχεία της εικόνας εισόδου είναι δύσκολη, διότι τα χαρακτηριστικά μπορεί να εμφανίζονται σε διαφορετικές θέσεις, με διαφορετικούς προσανατολισμούς και σε διαφορετικά μεγέθη στην εικόνα: η μετακίνηση του αντικειμένου ή της γωνίας λήψης της κάμερας μεταβάλλει δραστικά τις τιμές των εικονοστοιχείων, ακόμη και αν το ίδιο το αντικείμενο φαίνεται να παραμένει το ίδιο στα μάτια μας. Για την εκμάθηση του εντοπισμού ενός σήματος στοπ σε όλες αυτές τις διαφορετικές συνθήκες απαιτούνται τεράστιες ποσότητες δεδομένων εκπαίδευσης, διότι το δίκτυο θα εντοπίζει μόνο το σήμα στις συνθήκες στις οποίες εμφανίζεται στα δεδομένα εκπαίδευσης. Ως εκ τούτου, για παράδειγμα, ένα σήμα στοπ στην άνω δεξιά γωνία μιας εικόνας θα εντοπιστεί μόνο εφόσον στα δεδομένα εκπαίδευσης περιλαμβάνεται εικόνα με το σήμα στην άνω δεξιά γωνία. Τα ΣΝΔ μπορούν να αναγνωρίζουν το αντικείμενο οπουδήποτε, ανεξάρτητα από το πού έχει παρατηρηθεί στις εικόνες εκπαίδευσης.

Τα ΣΝΔ χρησιμοποιούν ένα έξυπνο τέχνασμα για να μειώνουν τον όγκο των δεδομένων εκπαίδευσης που απαιτείται για τον εντοπισμό αντικειμένων σε διαφορετικές συνθήκες. Βασικά, το τέχνασμα συνίσταται στη χρήση των ίδιων βαρών εισόδου για πολλούς νευρώνες – ούτως ώστε όλοι οι εν λόγω νευρώνες να ενεργοποιούνται από το ίδιο πρότυπο – αλλά με διαφορετικά εικονοστοιχεία εισόδου. Για παράδειγμα, μπορούμε να έχουμε ένα σύνολο νευρώνων που ενεργοποιούνται από το μυτερό αυτί μιας γάτας. Όταν η είσοδος είναι η φωτογραφία μιας γάτας, ενεργοποιούνται δύο νευρώνες, ένας για το αριστερό αυτί και ένας άλλος για το δεξί. Επίσης, μπορούμε να επιτρέψουμε τη λήψη των εικονοστοιχείων εισόδου του νευρώνα από μικρότερη ή μεγαλύτερη περιοχή, ούτως ώστε διαφορετικοί νευρώνες να ενεργοποιούνται από το αυτί που θα εμφανίζεται σε διαφορετικές κλίμακες (μεγέθη) και, ως εκ τούτου, να μπορούμε να εντοπίσουμε τα αυτιά μιας μικρής γάτας, ακόμη και αν στα δεδομένα εκπαίδευσης περιλαμβάνονταν εικόνες μόνο μεγάλων γατών.

Οι συνελικτικοί νευρώνες τοποθετούνται συνήθως στα κάτω επίπεδα του δικτύου, όπου πραγματοποιείται η επεξεργασία των ανεπεξέργαστων εικονοστοιχείων εισόδου. Οι βασικοί νευρώνες όπως ο νευρώνας perceptron, τοποθετούνται στα ανώτερα επίπεδα, όπου πραγματοποιείται η επεξεργασία της εξόδου των κάτω επιπέδων. Τα κάτω επίπεδα μπορούν συνήθως να εκπαιδευτούν μέσω μη επιβλεπόμενης μάθησης, χωρίς να αποσκοπούν σε κάποια συγκεκριμένη εργασία πρόβλεψης. Τα βάρη τους θα εκπαιδευτούν να εντοπίζουν χαρακτηριστικά που εμφανίζονται συχνά στα δεδομένα εισόδου. Ως εκ τούτου, στην περίπτωση φωτογραφιών ζώων, τα τυπικά χαρακτηριστικά θα είναι αυτιά και ρύγχη, ενώ στην περίπτωση φωτογραφιών κτιρίων, τα χαρακτηριστικά είναι αρχιτεκτονικά στοιχεία όπως τοίχοι, στέγες, παράθυρα, και ούτω καθεξής. Σε περίπτωση που ως δεδομένα εισόδου χρησιμοποιείται συνδυασμός διαφόρων αντικειμένων και σκηνών, τότε, τα χαρακτηριστικά που θα μαθαίνονται στα κάτω επίπεδα θα είναι λίγο έως πολύ γενικού χαρακτήρα. Αυτό σημαίνει ότι τα προ-εκπαιδευμένα συνελικτικά επίπεδα μπορούν να χρησιμοποιηθούν σε πολλές διαφορετικές εργασίες επεξεργασίας εικόνας. Αυτό είναι ιδιαίτερα σημαντικό, δεδομένου ότι είναι εύκολη η συγκέντρωση ουσιαστικά απεριόριστων ποσοτήτων μη επισημασμένων δεδομένων εκπαίδευσης – εικόνων χωρίς ετικέτες – τα οποία μπορούν να χρησιμοποιηθούν για την εκπαίδευση των κάτω επιπέδων. Τα ανώτερα επίπεδα εκπαιδεύονται πάντοτε μέσω τεχνικών επιβλεπόμενης μηχανικής μάθησης όπως η ανάστροφη διάδοση.

Η λειτουργία των CNN συνοψίζεται στα εξής τέσσερα βήματα:

1. Συνέλιξη και εφαρμογή μη γραμμικότητας (Convolution operation – ReLU layer)

2. Συγκέντρωση (Pooling)

3. Κανονικοποίηση (Flattening)

4. Πλήρης Σύνδεση (Full Connection)

3.3.1 Επίπεδο Συνέλιξης

Ο πρωταρχικός σκοπός της συνέλιξης(Convolution) είναι η εξαγωγή χαρακτηριστικών από την εικόνα εισόδου. Το Convolution διατηρεί τη χωρική σχέση μεταξύ των pixel μαθαίνοντας χαρακτηριστικά εικόνας χρησιμοποιώντας μικρά τετράγωνα δεδομένων εισόδου. Κάθε εικόνα μπορεί να θεωρηθεί ως μια μήτρα τιμών pixel. Έστω μια εικόνα 5 x 5 της οποίας οι τιμές pixel είναι μόνο 0 και 1 . Σε αυτό το σημείο ας σημειωθεί ότι για μια εικόνα σε κλίμακα του γκρι, οι τιμές pixel κυμαίνονται από 0 έως 255, ο πράσινος πίνακας παρακάτω είναι μια ειδική περίπτωση όπου οι τιμές pixel είναι μόνο 0 και 1:

1	1	1	0	0
0	1	1	1	0
0	0	1	1	1
0	0	1	1	0
0	1	1	0	0

Εικόνα.20: Εικόνα 5χ5 με τιμές pixel 0 και 1

Ας θεωρήσουμε έναν άλλο πίνακα 3 x 3 όπως φαίνεται παρακάτω:

1	0	1
0	1	0
1	0	1

Εικόνα.21: Εικόνα 3χ3 με τιμές pixel 0 και 1

Στη συνέχεια, η συνέλιξη της εικόνας 5 x 5 και του πίνακα 3 x 3 μπορεί να υπολογιστεί όπως φαίνεται στην εικόνα παρακάτω:

1 *1	1 *0	1 *1	0	0
0 *0	1 *1	1 *0	1	0
0 *1	0 *0	1 *1	1	1
0	0	1	1	0
0	1	1	0	0

Εικόνα

4		

Πίνακας Εξόδου
(Convoled Feature)

1	1 *1	1 *0	0 *1	0
0	1 *0	1 *1	1 *0	0
0	0 *1	1 *0	1 *1	1
0	0	1	1	0
0	1	1	0	0

Εικόνα

4	3	

Πίνακας Εξόδου
(Convoled Feature)

1	1	1 *1	0 *0	0 *1
0	1	1 *0	1 *1	0 *0
0	0	1 *1	1 *0	1 *1
0	0	1	1	0
0	1	1	0	0

Εικόνα

4	3	4

Πίνακας Εξόδου
(Convoled Feature)

1	1	1	0	0
0 *1	1 *0	1 *1	1	0
0 *0	0 *1	1 *0	1	1
0 *1	0 *0	1 *1	1	0
0	1	1	0	0

Εικόνα

4	3	4
2		

Πίνακας Εξόδου
(Convoled Feature)

1	1	1	0	0
0	1 *1	1 *0	1 *1	0
0	0 *0	1 *1	1 *0	1
0	0 *1	1 *0	1 *1	0
0	1	1	0	0

Εικόνα

4	3	4
2	4	

Πίνακας Εξόδου
(Convoled Feature)

1	1	1	0	0
0	1	1 *1	1 *0	0 *1
0	0	1 *0	1 *1	1 *0
0	0	1 *1	1 *0	0 *1
0	1	1	0	0

Εικόνα

4	3	4
2	4	3

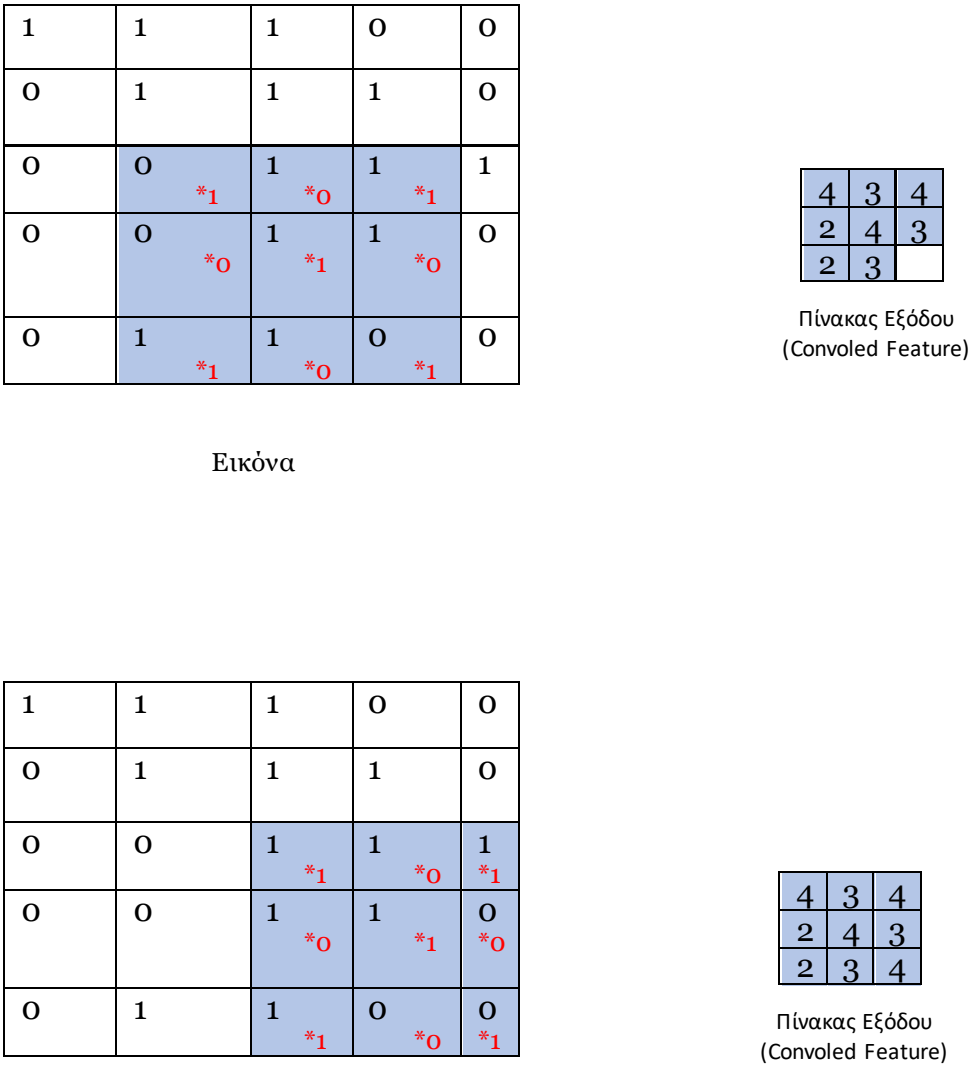
Πίνακας Εξόδου
(Convoled Feature)

1	1	1	0	0
0	1	1	1	0
0 *1	0 *0	1 *1	1	1
0 *0	0 *1	1 *0	1	0
0 *1	1 *0	1 *1	0	0

Εικόνα

4	3	4
2	4	3
2		

Πίνακας Εξόδου
(Convoled Feature)



Εικόνα.22: Η λειτουργία Convolution. Ο πίνακας εξόδου ονομάζεται Convolved Feature ή Feature Map

Όπως φαίνεται και στην εικόνα ο γαλάζιος πίνακας σύρεται πάνω από την αρχική μας εικόνα κατά 1 pixel και για κάθε θέση, υπολογίζουμε τον πολλαπλασιασμό βάσει στοιχείων (μεταξύ των δύο πινάκων) και προσθέτουμε τις εξόδους πολλαπλασιασμού για να πάρουμε τον τελικό ακέραιο που σχηματίζεται ένα μόνο στοιχείο του πίνακα εξόδου . Ας σημειωθεί ότι ο πίνακας 3×3 «βλέπει» μόνο ένα μέρος της εικόνας εισόδου σε κάθε διασκελισμό. Στην ορολογία του CNN, ο πίνακας 3×3 ονομάζεται «φίλτρο» ή «πυρήνας» ή «ανιχνευτής χαρακτηριστικών» και ο αντίστοιχος πίνακας που σχηματίζεται με την ολίσθηση του φίλτρου πάνω από την εικόνα και τον υπολογισμό του προϊόντος κουκκίδας ονομάζεται «Συζευγμένο χαρακτηριστικό» ή

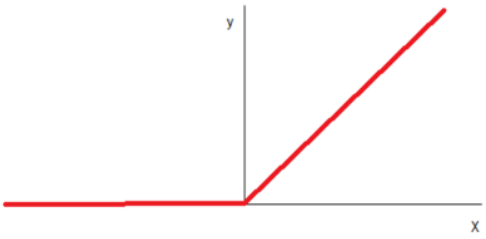
«Ενεργοποίηση», Χάρτης» ή «Χάρτης Χαρακτηριστικών». Είναι σημαντικό να σημειωθεί ότι τα φίλτρα λειτουργούν ως ανιχνευτές χαρακτηριστικών από την αρχική εικόνα εισόδου. Είναι προφανές από την παραπάνω κινούμενη εικόνα ότι διαφορετικές τιμές του πίνακα φίλτρου θα παράγουν διαφορετικούς χάρτες Χαρακτηριστικών για την ίδια εικόνα εισόδου.

Αυτό επιτυγχάνεται με ένα φίλτρο το οποίο ολισθαίνει πάνω από την εικόνα εισόδου (λειτουργία συνέλιξης) για να δημιουργήσει έναν χάρτη χαρακτηριστικών. Η περιέλιξη ενός άλλου φίλτρου, πάνω από την ίδια εικόνα δίνει έναν διαφορετικό χάρτη χαρακτηριστικών αντίστοιχα. Είναι σημαντικό να σημειωθεί ότι η λειτουργία Convolution καταγράφει τις τοπικές εξαρτήσεις στην αρχική εικόνα. Συνεπώς δύο διαφορετικά φίλτρα δημιουργούν διαφορετικούς χάρτες χαρακτηριστικών από την ίδια αρχική εικόνα. Στην πράξη, ένα CNN μαθαίνει μόνο του τις τιμές αυτών των φίλτρων κατά τη διάρκεια της εκπαιδευτικής διαδικασίας (αν και πρέπει ακόμα να καθορίσουμε παραμέτρους όπως ο αριθμός των φίλτρων, το μέγεθος του φίλτρου, η αρχιτεκτονική του δικτύου κ.λπ. πριν από τη διαδικασία εκπαίδευσης). Όσο περισσότερος αριθμός φίλτρων έχουμε, τόσο περισσότερα χαρακτηριστικά εικόνας εξάγονται και τόσο καλύτερο γίνεται το δίκτυό μας στην αναγνώριση μοτίβων σε εικόνες που δεν εμφανίζονται.

Το μέγεθος του χάρτη χαρακτηριστικών (Convolved Feature) ελέγχεται από τρεις παραμέτρους που πρέπει να αποφασίσουμε πριν από την εκτέλεση του βήματος συνέλιξης:

- Βάθος: Το βάθος αντιστοιχεί στον αριθμό των φίλτρων που χρησιμοποιούμε για τη λειτουργία συνέλιξης. Για παράδειγμα έστω ότι πραγματοποιείτε συνέλιξη μίας αρχικής εικόνας χρησιμοποιώντας τρία διαφορετικά φίλτρα, δημιουργώντας έτσι τρεις διαφορετικούς χάρτες χαρακτηριστικών αντίστοιχα. Οι τρεις αυτοί χάρτες χαρακτηριστικών μπορούν να θεωρηθούν και ως stacked πίνακες δύο διαστάσεων, επομένως, το «βάθος» του χάρτη χαρακτηριστικών θα είναι τρία.
- Διασκελισμός(Stride): Διασκελισμός είναι ο αριθμός των εικονοστοιχείων με τα οποία σύρουμε τον πίνακα φίλτρου πάνω από τον πίνακα εισόδου. Όταν ο διασκελισμός είναι 1 τότε μετακινούμε τα φίλτρα ένα pixel τη φορά. Όταν ο διασκελισμός είναι 2, τότε τα φίλτρα πηδούν κατά 2 pixel κάθε φορά καθώς τα σύρουμε. Έχοντας ένα μεγαλύτερο βήμα θα δημιουργήσει μικρότερους χάρτες χαρακτηριστικών.
- Μηδενική συμπλήρωση(zero padding): Μερικές φορές, είναι βολικό να συμπληρώνουμε τον πίνακα εισόδου με μηδενικά γύρω από το περίγραμμα, έτσι ώστε να μπορούμε να εφαρμόσουμε το φίλτρο σε συνοριακά στοιχεία του πίνακα εικόνας εισόδου. Ένα ωραίο χαρακτηριστικό του zero padding είναι ότι μας επιτρέπει να ελέγχουμε το μέγεθος των χαρτών χαρακτηριστικών. Η προσθήκη μηδενικής συμπλήρωσης ονομάζεται επίσης ευρεία συνέλιξη και η μη χρήση μηδενικής συμπλήρωσης θα ήταν μια στενή συνέλιξη.

Μια πρόσθετη λειτουργία που ονομάζεται ReLU χρησιμοποιείτε ορισμένες φορές μετά από τη λειτουργία συνέλιξης. Το ReLU σημαίνει Rectified Linear Unit και είναι μια μη γραμμική λειτουργία. Η έξοδος του δίνεται από:



Output = Max (zero, Input)

Εικόνα.23: Συνάρτηση εξόδου ReLu και αναπαράσταση με διάγραμμα

Το ReLU είναι μια λειτουργία βάσει στοιχείων (εφαρμόζεται ανά pixel) και αντικαθιστά όλες τις αρνητικές τιμές εικονοστοιχείων στον χάρτη χαρακτηριστικών με μηδέν. Ο σκοπός του ReLU είναι να εισαγάγει τη μη γραμμικότητα, καθώς τα περισσότερα από τα δεδομένα του πραγματικού κόσμου που θα θέλαμε να μάθει το δίκτυό μας θα ήταν μη γραμμικά (Η συνέλιξη είναι μια γραμμική πράξη – πολλαπλασιασμός και πρόσθεση πινάκων βάσει στοιχείων, επομένως εφαρμόζουμε τη μη γραμμικότητα εισάγοντας μια μη γραμμική συνάρτηση όπως η ReLU). Άλλες μη γραμμικές συναρτήσεις όπως το tanh ή το σιγμοειδές μπορούν επίσης να χρησιμοποιηθούν αντί του ReLU, αλλά έχει βρεθεί ότι το ReLU αποδίδει καλύτερα στις περισσότερες περιπτώσεις.

3.3.2. Επίπεδο Συγκέντρωσης (Pooling)

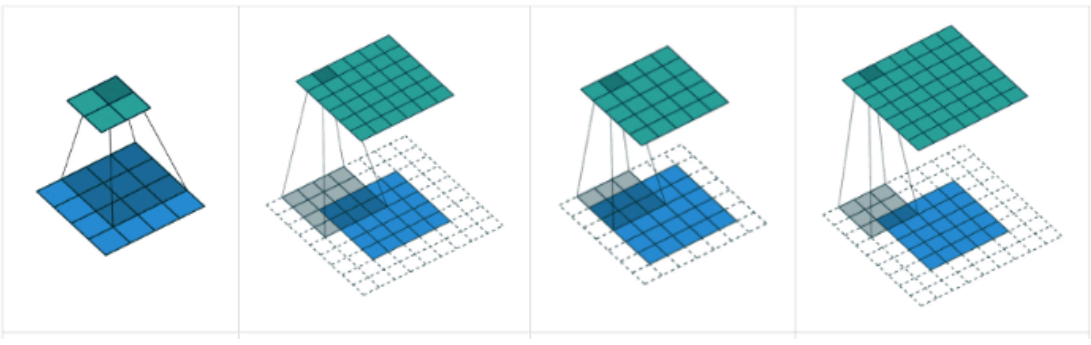
Η χωρική συγκέντρωση (ονομάζεται επίσης υποδειγματοληψία ή μείωση δειγματοληψίας) μειώνει τη διάσταση κάθε χάρτη χαρακτηριστικών, αλλά διατηρεί τις πιο σημαντικές πληροφορίες. Η χωρική συγκέντρωση μπορεί να είναι διαφορετικών τύπων: Μέγιστο, Μέσο, Άθροισμα κ.λπ. Στην περίπτωση του Max Pooling, ορίζουμε μια χωρική γειτονιά (για παράδειγμα, ένα παράθυρο 2×2) και παίρνουμε το μεγαλύτερο στοιχείο από τον διορθωμένο χάρτη χαρακτηριστικών εντός αυτού του παραθύρου. Αντί να πάρουμε το μεγαλύτερο στοιχείο, θα μπορούσαμε επίσης να πάρουμε τον μέσο όρο (Μέση συγκέντρωση) ή το άθροισμα όλων των στοιχείων σε αυτό το παράθυρο. Στην πράξη, το Max Pooling έχει αποδειχθεί ότι λειτουργεί καλύτερα.

Η λειτουργία του Max Pooling είναι να μειώνει προοδευτικά το χωρικό μέγεθος της αναπαράστασης εισόδου. Ειδικότερα, η συγκέντρωση (pooling):

1. κάνει τις αναπαραστάσεις εισόδου (διάσταση χαρακτηριστικών) μικρότερες και πιο διαχειρίσιμες
2. μειώνει τον αριθμό των παραμέτρων και των υπολογισμών στο δίκτυο, επομένως ελέγχεται η υπερπροσαρμογή(overfitting)
3. κάνει το δίκτυο αμετάβλητο σε μικρούς μετασχηματισμούς, παραμορφώσεις και μεταφράσεις στην εικόνα εισόδου.
4. μας βοηθά να φτάσουμε σε μια σχεδόν αμετάβλητη αναπαράσταση της εικόνας μας (ο ακριβής όρος είναι «ισοδύναμος»). Αυτό είναι πολύ σημαντικό αφού μπορούμε να ανιχνεύσουμε αντικείμενα σε μια εικόνα ανεξάρτητα από το πού βρίσκονται.

3.3.3 Κανονοικοποίηση (Flattening)

Υπάρχουν και άλλοι μέθοδοι που μπορούμε να εφαρμόσουμε με τη συνέλιξη, όπως είναι το padding. Η κανονικοποίηση μετατρέπει τα δεδομένα σε έναν μονοδιάστατο πίνακα για την εισαγωγή τους στο επόμενο επίπεδο. Κανονικοποιούμε την έξοδο των συνελικτικών στρωμάτων για να δημιουργήσουμε ένα ενιαίο μακρό διάνυσμα χαρακτηριστικών. Και συνδέεται με το τελικό μοντέλο ταξινόμησης, το οποίο ονομάζεται πλήρως συνδεδεμένο στρώμα. Με άλλα λόγια, βάζουμε όλα τα δεδομένα pixel σε μια γραμμή και κάνουμε συνδέσεις με το τελικό επίπεδο. Όμως τα pixel της εικόνας δεν υποβάλλονται σε επεξεργασία με τον ίδιο αριθμό. Τα pixel στις γωνίες είναι λιγότερο μετρημένα από εκείνα στη μέση. Αυτό σημαίνει ότι τα pixel δεν παίρνουν το ίδιο βάρος. Επιπλέον, εάν συνεχίσουμε να εφαρμόζουμε απλώς τη συνέλιξη, ενδέχεται να χάσουμε τα δεδομένα πολύ γρήγορα. Το padding είναι το τέχνασμα που μπορούμε να χρησιμοποιήσουμε για να διορθωθεί αυτό το πρόβλημα. Το padding σημαίνει την παροχή πρόσθετων pixel στο όριο των δεδομένων.



Εικόνα.24: Παράδειγμα συνέλιξης με padding

Πηγή: *The Most Intuitive and Easiest Guide for Convolutional Neural Network*

https://towardsdatascience.com/the-most-intuitive-and-easiest-guide-for-convolutional-neural-network-3607be47480

Η εικόνα εισόδου έχει 4x4 pixel και το φίλτρο έχει 3x3. Δεν υπάρχει padding, το οποίο ονομάζεται "έγκυρο"(valid). Το αποτέλεσμα γίνεται δεδομένα 2x2 pixel (4–3+1 = 2). Μπορούμε να δούμε ότι τα δεδομένα εξόδου είναι μειωμένα.

Στο τρίτο παράδειγμα υπάρχει ένα στρώμα γεμίσματος με τα κενά pixel. Η εικόνα εισόδου έχει 5x5 pixel και το φίλτρο έχει 3x3. Έτσι το αποτέλεσμα παίρνει 5x5 pixel (5 + 1*2–3 + 1= 5), το οποίο είναι το ίδιο μέγεθος με την εικόνα εισόδου. Το ονομάζουμε "ίδιο"(same). Μπορούμε ακόμη και να κάνουμε το αποτέλεσμα μεγαλύτερο από τα δεδομένα εισόδου, αλλά οι δύο περιπτώσεις χρησιμοποιούνται περισσότερο.

Επίσης ας σημειωθεί ότι ένα φίλτρο δεν είναι απαραίτητο πάντα να μετακινείται ένα pixel τη φορά. Μπορεί να κινείται δύο βήματα ή τρία βήματα τη φορά τόσο με οριζόντιο όσο και με κατακόρυφο τρόπο, το οποίο ονομάζεται «βήμα»(stride).

3.3.4 Πλήρως Συνδεδεμένο Επίπεδο

Το Fully Connected layer είναι ένα παραδοσιακό Perceptron πολλαπλών επιπέδων που χρησιμοποιεί μια συνάρτηση ενεργοποίησης softmax στο επίπεδο εξόδου (μπορούν επίσης να χρησιμοποιηθούν άλλοι ταξινομητές όπως το SVM).Ο όρος "Πλήρως συνδεδεμένος" υπονοεί ότι κάθε νευρώνας στο προηγούμενο στρώμα συνδέεται με κάθε νευρώνα στο επόμενο στρώμα. Η έξοδος από τα επίπεδα συνέλιξης και συγκέντρωσης αντιπροσωπεύει χαρακτηριστικά υψηλού επιπέδου της εικόνας εισόδου. Ο σκοπός του επιπέδου Fully Connected είναι να χρησιμοποιήσει αυτές τις δυνατότητες για την ταξινόμηση της εικόνας εισόδου σε διάφορες κλάσεις με βάση το σύνολο δεδομένων εκπαίδευσης.

Εκτός από την ταξινόμηση, η προσθήκη ενός πλήρως συνδεδεμένου στρώματος είναι επίσης ένας (συνήθως) φθηνός τρόπος εκμάθησης μη γραμμικών συνδυασμών αυτών των χαρακτηριστικών. Τα περισσότερα από τα χαρακτηριστικά από τα επίπεδα συνέλιξης και συγκέντρωσης μπορεί να είναι καλά για την εργασία ταξινόμησης, αλλά οι συνδυασμοί αυτών των χαρακτηριστικών μπορεί να είναι ακόμη καλύτεροι. Το άθροισμα των πιθανοτήτων εξόδου από το πλήρως συνδεδεμένο επίπεδο είναι 1. Αυτό διασφαλίζεται χρησιμοποιώντας το Softmax ως συνάρτηση ενεργοποίησης στο επίπεδο εξόδου του πλήρως συνδεδεμένου επιπέδου. Η συνάρτηση Softmax παίρνει ένα διάνυσμα αυθαίρετων βαθμολογιών πραγματικών τιμών και το συμπυκνώνει σε ένα διάνυσμα τιμών μεταξύ μηδέν και ενός που αθροίζονται σε ένα.

Κεφάλαιο 4

4.1 Αλγόριθμοι ανίχνευσης αντικειμένων σε αυτόνομη οδήγηση

Η ικανότητα ενός υπολογιστή για εξαγωγή πληροφοριών από ψηφιακές εικόνες ή βίντεο εμπίπτει στο πεδίο της όρασης του υπολογιστή (CV). Αυτό το πεδίο αναπτύσσεται ραγδαία παράλληλα με την άνοδο της βαθιάς μηχανικής μάθησης και χρησιμοποιείται σε συνδυασμό με πολλά προβλήματα στην τεχνητή νοημοσύνη (AI). Ένα κοινό πρόβλημα στο πεδίο του Computer Vision(CV), το οποίο για μεγάλο χρονικό διάστημα θεωρήθηκε ότι είναι δύσκολο να επιλυθεί, είναι η ταξινόμηση εικόνας. Σήμερα, η συνεχώς αυξημένη δημοτικότητα και η ουσιαστική έρευνα στη βαθιά μηχανική μάθηση έχουν επιτρέψει σε ένα ειδικό είδος βαθύ νευρωνικού δικτύου, που ονομάζεται Convolutional Neural Network (CNN), να βελτιώσει την απόδοση ταξινόμησης της ανθρώπινης εικόνας.

Ωστόσο, για ένα αυτοκινούμενο όχημα, η απλή ταξινόμηση εικόνων από το περιβάλλον δεν είναι αρκετή. Αντ 'αυτού, το σύστημα πρέπει να μπορεί να εντοπίζει και να ταξινομεί πολλά αντικείμενα χρησιμοποιώντας εικόνες από κάμερα. Αυτή η διαδικασία ονομάζεται συνήθως ανίχνευση αντικειμένων και είναι πολύ πιο δύσκολη από την απλή ταξινόμηση εικόνων σε διακριτές τάξεις. Επιπλέον, η ασφαλής οδήγηση απαιτεί κάτι περισσότερο από την απλή ανίχνευση του δρόμου στον οποίο οδηγούμε, μαζί με άλλα οχήματα, άτομα και ζώα. Πρέπει επίσης το σύστημα να είναι σε θέση να γνωρίζει πώς είναι πιο πιθανό να ενεργήσουν και πώς να ανταποκριθούμε. Για να μπορέσουμε να εξαγάγουμε αυτά τα συμπεράσματα και να χρησιμοποιήσουμε τις πληροφορίες για να ενεργήσουμε αναλόγως, είναι επιθυμητό να υπάρχει ένα σύστημα ικανό να επεξεργάζεται εικόνες από ενσωματωμένες κάμερες σε πραγματικό χρόνο.

Η βελτίωση αυτοκινούμενων οχημάτων συνεχίζεται από τη δεκαετία του '60, όμως η αλληλεπίδραση και η κυκλοφορία σε πραγματικούς δρόμους με σενάρια πραγματικής κυκλοφορίας, είναι κάτι που δεν ήταν δυνατό μέχρι τα τελευταία δέκα χρόνια. Κατά τη διάρκεια αυτών των ετών, τα αυτοκίνητα χωρίς οδηγό έχουν χρησιμοποιήσει τεχνολογίες όπως Radar, LiDAR, GPS και άλλους αισθητήρες για χαρτογράφηση του περιβάλλοντος του αυτοκινήτου. Ωστόσο, τα τελευταία χρόνια, έχουν εμφανιστεί κάποιες αρχιτεκτονικές βαθύ νευρωνικού δικτύου (DNN) ικανές για ανίχνευση αντικειμένων ζωντανής ροής βίντεο, όπως το YOLO (You Only Look Once) με δυνατότητα χρήσης ως μέρος αυτόνομων συστημάτων οχημάτων. Ένα μειονέκτημα με αυτές τις αρχιτεκτονικές ανίχνευσης αντικειμένων, μαζί με άλλα βαθιά νευρωνικά δίκτυα, είναι ότι απαιτούν τεράστιες ποσότητες δεδομένων με ετικέτες για την εκπαίδευση - στην περίπτωση αυτόνομων οχημάτων, εικόνες με αντικείμενα, όπως αυτοκίνητα, οδικές πινακίδες, δρόμους , και οι άνθρωποι με ακριβείς σχολιασμούς. Η δημιουργία αυτών των συνόλων δεδομένων είναι τόσο χρονοβόρα όσο και δαπανηρή λόγω των απαιτήσεων ώστε οι εικόνες να είναι αντιπροσωπευτικές και με ποικιλία συνθηκών οδήγησης και σεναρίων. Μία πιθανή εναλλακτική λύση για τη συλλογή δεδομένων σε πραγματικές σκηνές οδήγησης θα ήταν η χρήση εικονικών περιβαλλόντων για τη συλλογή συνθετικών δεδομένων. Οι ανιχνευτές αντικειμένων βαθιών νευρωνικών

δικτύων θα μπορούσαν τότε να εκπαιδευτούν χρησιμοποιώντας αυτά τα εικονικά δεδομένα για να αναγνωρίσουν αντικείμενα πραγματικής ζωής σε σκηνές πραγματικής ζωής.

Υπάρχουν δύο κύριοι αρχιτεκτονικοί τύποι DNN για την ανίχνευση αντικειμένων: δίκτυα με βάση την περιοχή (δύο στάδια) και δίκτυα με παλινδρόμηση (onestage). Στα πλαίσια δύο σταδίων, το πρώτο βήμα αποτελείται από προτάσεις περιοχής ανεξάρτητες από την κατηγορία, ακολουθούμενες από την εξαγωγή χαρακτηριστικών CNN από αυτές τις περιοχές. Οι κατηγοριοποιητές ειδικών κατηγοριών δεύτερου βήματος χρησιμοποιούνται για τον προσδιορισμό των ετικετών κατηγορίας των προτάσεων. Τα περισσότερα δίκτυα δύο σταδίων παράγουν χιλιάδες προτάσεις περιοχής κατά τη διάρκεια της δοκιμής, η οποία συνοδεύεται από υψηλό υπολογιστικό κόστος. Τα ταχύτερα δίκτυα ανιχνευτών αντικειμένων δύο σταδίων σήμερα είναι ταχύτερα R-CNN και R-FCN, τα οποία είναι σε θέση να επεξεργάζονται εικόνες σε περίπου 5-6 FPS.

Σε αντίθεση με τους ανιχνευτές αντικειμένων δύο σταδίων, τα δίκτυα ενός σταδίου προβλέπουν άμεσα τις πιθανότητες κλάσης και τις αντισταθμίσεις πλαισίου οριοθέτησης από πλήρεις εικόνες με ένα μόνο δίκτυο CNN feedforward. Αυτή η απλούστερη και πιο κομψή προσέγγιση εξαλείφει τη δημιουργία προτάσεων περιοχής και τα επόμενα στάδια δειγματοληψίας χαρακτηριστικών και επιτρέπει στο δίκτυο να βελτιστοποιείται από άκρο σε άκρο απευθείας στην απόδοση ανίχνευσης. Αν και αυτή η βελτιστοποίηση απόδοσης ανίχνευσης έρχεται με μια μικρή μείωση στην ακρίβεια σε σύγκριση με δίκτυα δύο σταδίων, τα δίκτυα ενός σταδίου συχνά ισχυρίζονται ότι έχουν δυνατότητες ανίχνευσης αντικειμένων σε πραγματικό χρόνο.

Η ανίχνευση αντικειμένων είναι μία από τις πιο σημαντικές απαιτήσεις για αυτόνομη πλοήγηση και αποτελείται από εντοπισμό και ταξινόμηση αντικειμένων. Επομένως, χρειάζονται ακριβείς αλγόριθμοι ανίχνευσης αντικειμένων. Μια πρόκληση για παράδειγμα είναι η επεξεργασία πολλών υποψηφίων θέσεων αντικειμένων (συχνά αποκαλούνται «προτάσεις»(proposal). Αυτοί οι υποψήφιοι παρέχουν μόνο ακατέργαστο εντοπισμό που πρέπει να βελτιωθεί για να επιτευχθεί ακριβής εντοπισμός . Ωστόσο, οι λύσεις σε αυτά τα προβλήματα συχνά θέτουν σε κίνδυνο την ταχύτητα, την ακρίβεια ή την απλότητα . Τα πρόσφατα υπερσύγχρονα μοντέλα βαθιάς μάθησης που αντιμετωπίζουν το πρόβλημα της ανίχνευσης αντικειμένων περιλαμβάνουν Περιφερειακά Βασικά Νευρωνικά Δίκτυα (R-CNN) και τις βελτιωμένες εκδόσεις τους Fast R-CNN και Faster R-CNN, σχεδιασμένα για απόδοση μοντέλου και πρώτα εισήχθη το 2013. Ένα δεύτερο μοντέλο ανίχνευσης αντικειμένων που παρουσιάστηκε το 2015 είναι το YOLO, σχεδιασμένο για ταχύτητα και χρήση σε πραγματικό χρόνο.

4.2 Σύνολο Δεδομένων

Το BDD100K περιλαμβάνει εκατοντάδες χιλιάδες εικόνες σχολιασμένες με ετικέτες επιπέδου εικόνας, πλαίσια οριοθέτησης αντικειμένων, κινητές περιοχές, σήματα λωρίδας και τμηματοποίηση παρουσιών πλήρους καρέ. Ωστόσο, στην παρούσα διπλωματική θα χρησιμοποιηθούν μόνο οι οριοθετημένες εικόνες με σχόλια.

Εικόνα.25: Στιγμιότυπα από το σύνολο δεδομένων BDD100K

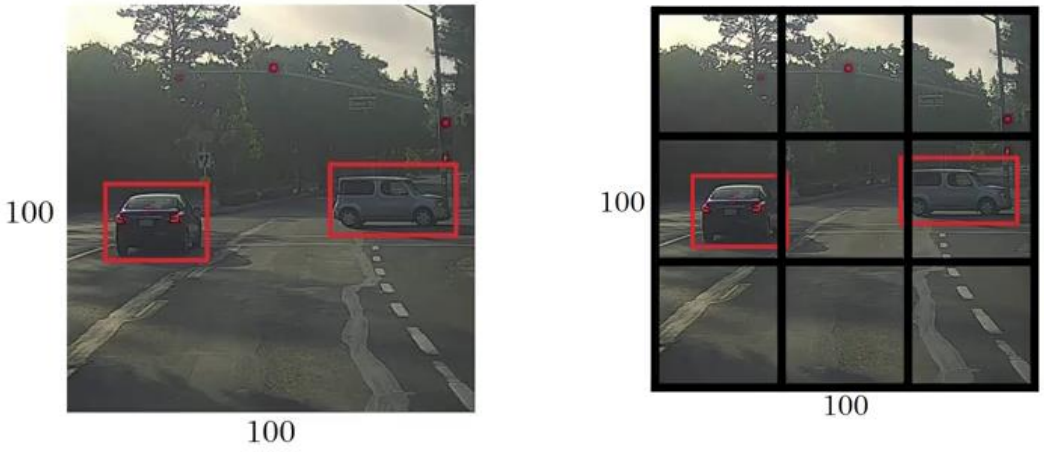
Πηγή: : https://bdd-data.berkeley.edu/

Το σύνολο δεδομένων BDD100K, δημιουργήθηκε για να αξιολογήσει το όριο των ανιχνευτών αντικειμένων, παρουσιάζοντας μια ποικιλία από διάφορες συνθήκες οδήγησης που μπορεί να αντιμετωπίσει ένα αυτο-οδηγημένο όχημα στην καθημερινή ζωή. Οι εικόνες συλλέχθηκαν στην πόλη της Νέας Υόρκης, στο Μπέρκλεϋ και στο Σαν Φρανσίσκο, σε διάφορες καταστάσεις οδήγησης και καλύπτει διαφορετικές καιρικές συνθήκες, συμπεριλαμβανομένων των συνθηκών ηλιοφάνειας, συννεφιά και βροχή, καθώς και εικόνες που καταγράφονται τόσο κατά τη διάρκεια της ημέρας όσο και κατά τη διάρκεια της νύχτας, όπως φαίνεται στο σχήμα παραπάνω. Η ποικιλομορφία αυτή δίνει τη δυνατότητα στο BDD100K να αποτελεί αρκετά αληθοφανή και αξιόπιστη αναπαράσταση του φυσικού κόσμου που θα συναντούσε ένα σύστημα αυτό-οδηγούμενου οχήματος. Το BDD100K περιλαμβάνει πάνω από 1,8 εκατομμύρια οριοθετημένα κουτιά (bounding box) με διαφορετικά αντικείμενα με εμφανίσεις και περιβάλλοντα από δέκα διαφορετικές κατηγορίες (λεωφορείο, φανάρι, πινακίδα, άτομο, ποδήλατο, φορτηγό, μοτοσικλέτα, αυτοκίνητο, τρένο, αναβάτης).

4.3 YOLO (You only look once)

Η οικογένεια των τεχνικών R-CNN διαχειρίζεται κυρίως περιοχές για τον εντοπισμό των αντικειμένων που υπάρχουν μέσα σε μία εικόνα. Η συγκεκριμένη κατηγορία δικτύων δεν σκανάρει ολόκληρη την εικόνα, παρά μόνο εκείνα τα μέρη των εικόνων που παρουσιάζουν υψηλότερη πιθανότητα να περιέχουν ένα αντικείμενο. Αντιθέτως το σύστημα YOLO, επεξεργάζεται την ανίχνευση αντικειμένων με διαφορετικό τρόπο. Λαμβάνει ολόκληρη την εικόνα σε ένα μόνο παράδειγμα και προβλέπει τις συντεταγμένες του πλαισίου οριοθέτησης και τις πιθανότητες κλάσεων για αυτά τα πλαίσια. Το μεγαλύτερο πλεονέκτημα της χρήσης του YOLO είναι η εξαιρετική του ταχύτητα – είναι απίστευτα γρήγορο και μπορεί να επεξεργαστεί έως και 45 καρέ ανά δευτερόλεπτο. Το YOLO αντικατοπτρίζει επίσης τη γενικευμένη αναπαράσταση αντικειμένων. Πρόκειται για έναν από τους καλύτερους αλγόριθμους για την ανίχνευση αντικειμένων και έχει επιτύχει συγκριτικά παρόμοια απόδοση με τους αλγόριθμους R-CNN. Για την καλύτερη κατανόηση του αλγορίθμου που χρησιμοποιήθηκε στην παρούσα διπλωματική, σκόπιμο θα ήταν να αναλύσουμε τα βήματα που ακολουθούνται από το YOLO για την ανίχνευση αντικειμένων σε μια δεδομένη εικόνα.

Ο αλγόριθμος YOLO υποδέχεται αρχικά μια εικόνα εισαγωγής, και στη συνέχεια το επιλεγμένο πλαίσιο χωρίζει την εικόνα εισόδου σε πλέγματα (έστω ένα πλέγμα 3 X 3). Η ταξινόμηση καθώς και ο εντοπισμός εικόνων εφαρμόζονται σε κάθε πλέγμα. Έπειτα το YOLO προβλέπει τα οριοθετημένα πλαίσια και τις αντίστοιχες πιθανότητες κλάσης για αντικείμενα που θα εντοπιστούν.



Εικόνα.26: Στιγμιότυπα Ανίχνευσης Αντικειμένων με τον αλγόριθμο YOLO

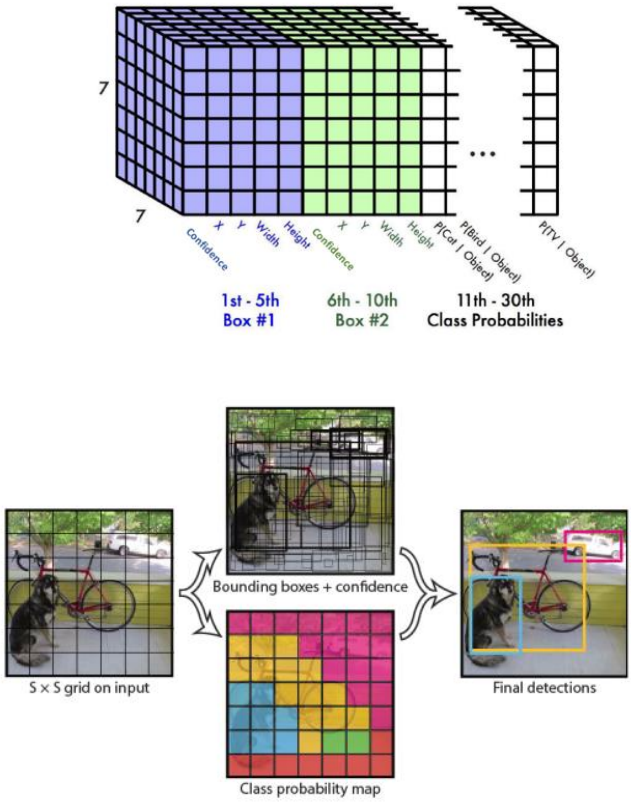
Πηγή: A Practical Guide to Object Detection using the Popular YOLO Framework – Part III (with Python codes), *https://www.analyticsvidhya.com/blog/2018/12/practical-guide-object-detection-yolo-framewor-python/*

Ο αλγόριθμος YOLO αποτελεί έναν σύγχρονος αλγόριθμο ανίχνευσης αντικειμένων που αναπτύχθηκε και δημοσιεύθηκε το 2015 από τους Redmon et al.. Το όνομα του αλγορίθμου βασίζεται στο γεγονός ότι ο αλγόριθμος

σκανάρει μόνο μια φορά την εικόνα και χρειάζεται μόνο μία μετάβαση προς τα εμπρός μέσω του νευρωνικού δικτύου για να πραγματοποιήσει προβλέψεις, σε αντίθεση με άλλους αλγόριθμους ανίχνευσης αντικειμένων τελευταίας τεχνολογίας που λειτουργούν με προτάσεις περιοχής και εξετάζουν την εικόνα πολλές φορές. Το σύστημα YOLO αποτελείται από ένα μονό άκρο σε άκρο συνελικτικό νευρωνιικό δίκτυο που επεξεργάζεται εικόνες RGB μεγέθους 448 x 448 και πραγματοποιεί τις προβλέψεις του πλαισίου οριοθέτησης για τη δεδομένη εικόνα. Δηλαδή τροποποιεί εκ νέου την ανίχνευση αντικειμένων ως ένα πρόβλημα παλινδρόμησης, κατευθείαν από εικονοστοιχεία εικόνας έως συντεταγμένες οριοθετημένου πλαισίου και πιθανότητες κλάσης. Ο αλγόριθμος διαιρεί την εικόνα εισόδου σε πλέγμα S x S (έστω S = 7). Για κάθε κελί πλέγματος προβλέπει B κουτιά οριοθέτησης (B = 2), όπου κάθε κουτί οριοθέτησης διαμορφώνεται από 4 συντεταγμένες και βαθμολογία εμπιστοσύνης για την πρόβλεψη και πιθανότητες κλάσης C ανά κελί πλέγματος λαμβάνοντας την υψηλότερη ως την τελική τάξη. Όλες αυτές οι προβλέψεις κωδικοποιούνται ως αισθητήρας S x S x (B * 5 + C) που εξάγεται από το νευρωνικό δίκτυο. Συνεπώς η λειτουργία που πραγματοποιεί ο αλγόριθμος, είναι η αναγνώριση αντικειμένων στην εικόνα και η χαρτογράφηση τους στο κελί πλέγματος που περιέχει το κέντρο του αντικειμένου. Το συγκεκριμένο κελί πλέγματος θα αποτελεί το καθοριστικό στοιχείο για την πρόβλεψη του τελικού πλαισίου οριοθέτησης του αντικειμένου και θα περιλαμβάνει την υψηλότερη βαθμολογία εμπιστοσύνης.

Στο παράδειγμα που παρατίθεται στην παρακάτω εικόνα κάθε κελί με πλέγμα 7 x 7 αντιπροσωπεύεται από ένα διάνυσμα μεγέθους 30 που αντιπροσωπεύει μια συγκεκριμένη περιοχή της εικόνας. Κάθε διάνυσμα περιλαμβάνει 2 προβλέψεις οριοθέτησης (5 τιμές καθεμία) και 20 πιθανότητες κλάσης υπό όρους P(κλάση|αντικείμενο). Αρχικά για να μπορέσει να πραγματοποιηθεί η εξαγωγή μιας έγκυρης πρόβλεψης αναγκαίο είναι να επιλεχθεί το πλαίσιο οριοθέτησης με την υψηλότερη βαθμολογία εμπιστοσύνης και να διασταυρωθεί εάν η βαθμολογία εμπιστοσύνης βρίσκεται πάνω από ένα προκαθορισμένο όριο (όριο = 0,25) και στη συνέχεια να το εξάγει ως έγκυρη πρόβλεψη. Η συγκεκριμένη βαθμολογία εμπιστοσύνης αντικατοπτρίζει το προηγούμενο στην υπό όρους πιθανότητα για την πρόβλεψη κλάσης που δηλώνει την πιθανότητα το δεδομένο κελί πλέγματος να είναι το κέντρο ενός αντικειμένου με ένα σωστό πλαίσιο οριοθέτησης. Προκειμένου να γίνει η εξαγωγή της πρόβλεψης κλάσης, ο αλγόριθμος YOLO βγάζει την υπό όρους πιθανότητα με την υψηλότερη βαθμολογία. Έπειτα οριοθετεί χωρικά κάθε πλαίσιο οριοθέτησης με τέσσερις συντεταγμένες (X, Y, Πλάτος, Ύψος), όπου (X, Y) αντιπροσωπεύουν το κέντρο του πλαισίου οριοθέτησης σε σχέση με το κελί, ενώ (Πλάτος, Ύψος) αντιπροσωπεύουν το πλάτος και το ύψος του οριοθετημένου πλαισίου σε σύγκριση με ολόκληρη την εικόνα. Το γεγονός αυτό μπορεί να οδηγήσει στο φαινόμενο κατά το οποίο ένα πλαίσιο οριοθέτησης ενδέχεται να είναι μεγαλύτερο από το κελί όπου είχε προβλεφθεί. Το κελί χρησιμοποιείται μόνο ως σημείο αγκύρωσης για την πρόβλεψη. Τέλος, μειονέκτημα αυτού του αλγορίθμου αποτελεί το γεγονός ότι κάθε κελί είναι σε θέση να προβλέψει μόνο ένα αντικείμενο. Στην περίπτωση κατά την οποία

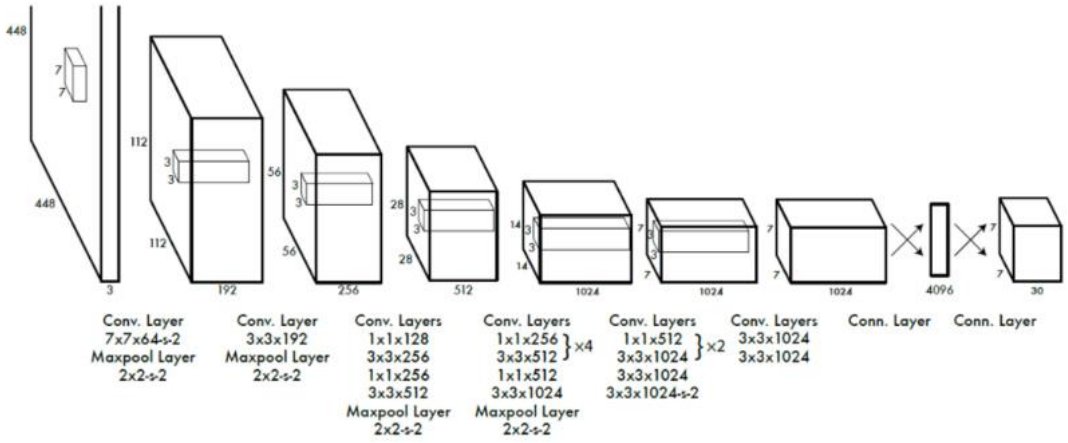
μεγάλο πλήθος αντικειμένων έχουν τα κεντρικά τους σημεία στο ίδιο κελί, μόνο ένα θα μπορέσει να προβλεφθεί.



Εικόνα.27: Λειτουργία Αλγορίθμου YOLO

Πηγή:https://www.pyimagesearch.com/2018/11/12/yolo-object-detection-with-opencv/

Όσο αναφορά την αρχιτεκτονική του μοντέλου YOLO, περιλαμβάνει συνολικά 24 συνελικτικά στρώματα που ακολουθούνται από 4 επίπεδα συγκέντρωσης και 2 πλήρως συνδεδεμένα επίπεδα, όπως φαίνεται και στην εικόνα παρακάτω. Επεξεργάζεται και πραγματοποιεί 1 x 1 συνελίξεις για να μειώσει τον αριθμό των χαρτών χαρακτηριστικών το οποίο υποκινείται από τα Inception Modules του GoogLeNet. Επιπλέον, εφαρμόζει τη λειτουργία ενεργοποίησης Leaky ReLu μετά από όλα τα επίπεδα εκτός από το τελευταίο και χρησιμοποιεί διαρροή μεταξύ των δύο πλήρως συνδεδεμένων στρωμάτων προκειμένου να αντιμετωπιστεί η υπερβολική τοποθέτηση.



Εικόνα.28: Αρχιτεκτονική Αλγορίθμου YOLO

Πηγή: Sik-Ho Tsang. Review: Yolov1 — you only look once (object detection).https://towardsdatascience.com/yolov1-you-only-look-once-object-detection-e1f3ffec8a89

Ο αλγόριθμος YOLO χρησιμοποιεί μια προσαρμοσμένη συνάρτηση απώλειας προκειμένου να ελέγχει τη διαφορετική έξοδο τομείς και η επιρροή τους στην τελική απώλεια με τη χρήση ειδικών υπερπαραμέτρων :

1. πρώτος όρος: τιμωρεί τις κακές θέσεις για τις κεντρικές συντεταγμένες εάν το κελί περιέχει ένα αντικείμενο.
2. δεύτερος όρος: τιμωρεί τις κακές τιμές πλάτους και ύψους του πλαισίου οριοθέτησης. Η τετραγωνική ρίζα παρουσιάζεται έτσι ώστε τα σφάλματα σε μικρά οριοθετημένα κουτιά να είναι πιο επιζήμια από τα σφάλματα σε μεγάλα οριοθετημένα κουτιά.
3. τρίτος όρος: τιμωρεί τις μικρές βαθμολογίες εμπιστοσύνης για κελιά που περιέχουν ένα αντικείμενο.
4. τέταρτος όρος: τιμωρεί μεγάλες βαθμολογίες εμπιστοσύνης για κελιά που δεν περιέχουν αντικείμενο.
5. πέμπτος όρος: απλή τετραγωνισμένη απώλεια ταξινόμησης.

$$\lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{obj} [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2]$$

$$+\lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{obj} [(\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2]$$

$$+\sum_{i=0}^{S^2}\sum_{j=0}^{B} 1_{ij}^{obj}(c_i - \hat{c}_i)^2$$

$$+\lambda_{noobj}\sum_{i=0}^{S^2}\sum_{j=0}^{B} 1_{ij}^{noobj}(c_i - \hat{c}_i)^2$$

$$+\sum_{i=0}^{S^2} 1_i^{obj}\sum_{c \in classes}(p_i(c) - \hat{p}_i(c))^2$$

Εικόνα.29: Συνάρτηση Απώλειας

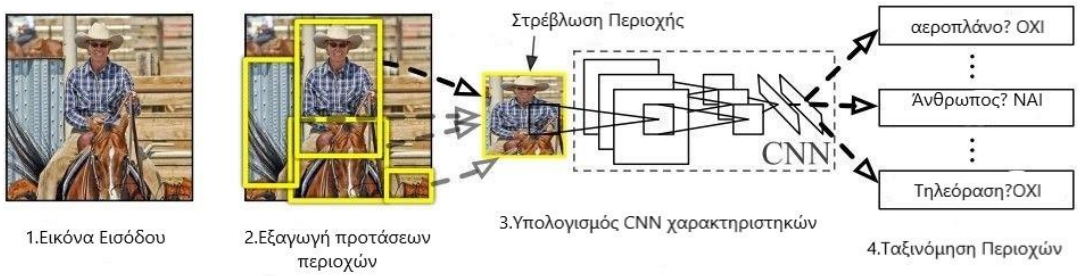
Στην παρούσα διπλωματική εφαρμόστηκε και εκπαιδεύτηκε το YOLO εξ ολοκλήρου από την αρχή χρησιμοποιώντας μόνο το BDD100K σύνολο δεδομένων. Δεδομένου ότι περιλαμβάνονται πολλά αντικείμενα ανάμεσα στις εικόνες του συνόλου δεδομένων BDD100K, αυξήθηκε το μέγεθος διαχωρισμού για τον αλγόριθμο YOLO από 7 σε 14 έτσι ώστε να μειώσει το πρόβλημα πολλαπλών κέντρων αντικειμένων που εμπίπτουν σε ένα κελί. Για να επιτευχθεί η αύξηση στον αριθμό των παραμέτρων εξόδου από 1127 σε 4508 αλλά και το ποσό των παραμέτρων μέσα στα τελευταία 5 επίπεδα, χρησιμοποιήθηκε και αναπτύχθηκε ένα πολύ λεπτότερο πλέγμα. Η πραγματοποίηση αυτού του πλέγματος έγινε με την προσαρμογή στον διασκελισμό του 23ου συνελικτικού στρώματος από 2 έως 1 για διατήρηση του μεγέθους εξόδου 14 x 14.

Στη συνέχεια πραγματοποιήθηκε προσθήκη ομαλοποίηση παρτίδας μεταξύ όλων των στρωμάτων για να μπορέσει να αυξηθεί η ταχύτητα και να διατηρηθούν οι αρχικές υπερπαράμετροι απώλειας κατά τη διάρκεια της εκπαίδευσης (coord=5 και noobj=0,5). Η εκπαίδευση του YOLO πραγματοποιήθηκε για 100 epochs με ρυθμό μάθησης 1e-5 και παρτίδα μέγεθος 10 και συνεπώς το σύστημα YOLO παράγει 2 προβλέψεις οριοθέτησης ανά κελί πλέγματος σε 14 x 14 πλέγμα. Η εικόνα που δίνεται ως είσοδος έχει διάσταση (3, 448, 448) και ο αλγόριθμος παράγει ως έξοδο έναν τανυστή μεγέθους (14, 14, 23).

4.4 R-CNN

Το 2014, μια ομάδα ερευνητών στο UC Berkely ανέπτυξε ένα βαθύ συνελικτικό δίκτυο που ονομάζεται R-CNN (συντομογραφία για το συνελικτικό νευρωνικό δίκτυο που βασίζεται σε περιοχή) που μπορεί να ανιχνεύσει 80 διαφορετικούς τύπους αντικειμένων σε εικόνες. Σε σύγκριση με τον γενικό αγωγό των τεχνικών ανίχνευσης αντικειμένων, η κύρια συμβολή του R-CNN απλώς εξάγει τα χαρακτηριστικά που βασίζονται σε ένα συνελικτικό νευρωνικό δίκτυο

(CNN). Εκτός από αυτό, όλα είναι παρόμοια με τη γενική γραμμή ανίχνευσης αντικειμένων. Το επόμενο σχήμα δείχνει τη λειτουργία του μοντέλου R-CNN.



Εικόνα.30: Λειτουργία του μοντέλου R-CNN

Πηγή: Faster R-CNN Explained for Object Detection Tasks, https://blog.paperspace.com/faster-r-cnn-explained-object-detection/

Το R-CNN αποτελείται από 3 κύριες ενότητες:

1. Η πρώτη ενότητα δημιουργεί 2.000 προτάσεις περιοχών χρησιμοποιώντας τον αλγόριθμο Επιλεκτικής Αναζήτησης.
2. Αφού αλλάξει το μέγεθος σε ένα σταθερό προκαθορισμένο μέγεθος, η δεύτερη ενότητα εξάγει ένα διάνυσμα χαρακτηριστικών μήκους 4.096 από κάθε πρόταση περιοχής.
3. Η τρίτη ενότητα χρησιμοποιεί έναν προ εκπαιδευμένο αλγόριθμο SVM για να ταξινομήσει την πρόταση περιοχής είτε στο φόντο είτε σε μία από τις κατηγορίες αντικειμένων.

Το μοντέλο R-CNN έχει ορισμένα μειονεκτήματα:

1. Είναι ένα μοντέλο πολλαπλών σταδίων, όπου κάθε στάδιο είναι ένα ανεξάρτητο στοιχείο. Έτσι, δεν μπορεί να εκπαιδευτεί από άκρο σε άκρο.
2. Αποθηκεύει τις εξαγόμενες δυνατότητες από το προ εκπαιδευμένο CNN στο δίσκο για να εκπαιδεύσει αργότερα τα SVM. Αυτό απαιτεί εκατοντάδες gigabyte αποθήκευσης.
3. Το R-CNN εξαρτάται από τον αλγόριθμο επιλεκτικής αναζήτησης για τη δημιουργία προτάσεων περιοχής, κάτι που απαιτεί πολύ χρόνο. Επιπλέον, αυτός ο αλγόριθμος δεν μπορεί να προσαρμοστεί στο πρόβλημα ανίχνευσης.
4. Κάθε πρόταση περιοχής τροφοδοτείται ανεξάρτητα στο CNN για εξαγωγή χαρακτηριστικών. Αυτό καθιστά αδύνατη την εκτέλεση του R-CNN σε πραγματικό χρόνο.

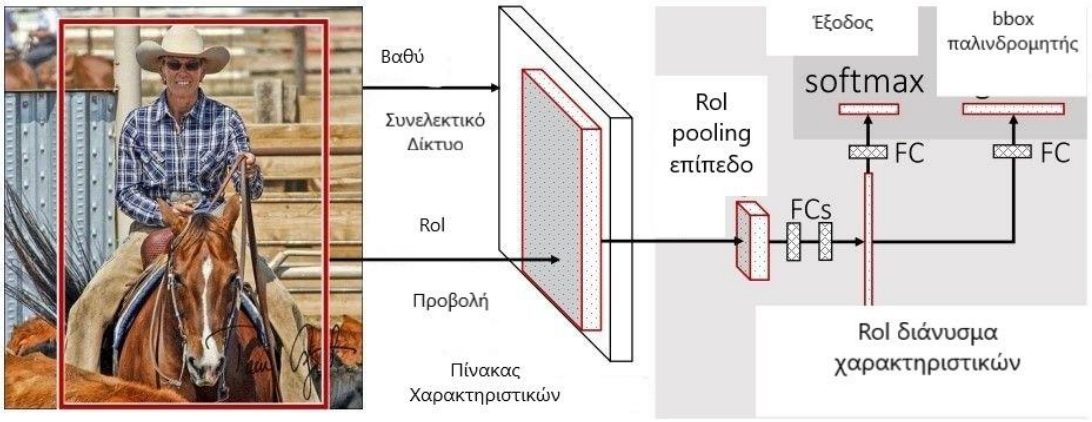
Ως επέκταση του μοντέλου R-CNN, προτείνεται το μοντέλο Fast R-CNN για να ξεπεράσουμε κάποιους περιορισμούς. Μια γρήγορη επισκόπηση του Fast R-CNN δίνεται στην επόμενη ενότητα.

4.5 Fast R-CNN

Fast R-CNN είναι ένας ανιχνευτής αντικειμένων που αναπτύχθηκε αποκλειστικά από τον Ross Girshick, ερευνητή AI του Facebook και πρώην ερευνητή της Microsoft. Το Fast R-CNN ξεπερνά αρκετά προβλήματα στο R-CNN. Όπως υποδηλώνει το όνομά του, ένα πλεονέκτημα του Fast R-CNN έναντι του R-CNN είναι η ταχύτητά του.

1. Πρότεινε ένα νέο επίπεδο που ονομάζεται ROI Pooling που εξάγει διανύσματα χαρακτηριστικών ίσου μήκους από όλες τις προτάσεις (δηλαδή ROI) στην ίδια εικόνα.
2. Σε σύγκριση με το R-CNN, το οποίο έχει πολλαπλά στάδια (δημιουργία πρότασης περιοχής, εξαγωγή χαρακτηριστικών και ταξινόμηση με χρήση SVM), το Faster R-CNN δημιουργεί ένα δίκτυο που έχει μόνο ένα στάδιο.
3. Το ταχύτερο R-CNN μοιράζεται υπολογισμούς (δηλαδή υπολογισμούς συνελικτικού επιπέδου) σε όλες τις προτάσεις (δηλαδή ROI) αντί να κάνει τους υπολογισμούς για κάθε πρόταση ανεξάρτητα. Αυτό γίνεται χρησιμοποιώντας το νέο επίπεδο συγκέντρωσης ROI, το οποίο κάνει το Fast R-CNN πιο γρήγορο από το R-CNN.
4. Το Fast R-CNN δεν αποθηκεύει προσωρινά τις εξαγόμενες δυνατότητες και επομένως δεν χρειάζεται τόσο πολύ χώρο αποθήκευσης δίσκου σε σύγκριση με το R-CNN, το οποίο χρειάζεται εκατοντάδες gigabyte.
5. Το γρήγορο R-CNN είναι πιο ακριβές από το R-CNN.

Η γενική αρχιτεκτονική του Fast R-CNN φαίνεται παρακάτω. Το μοντέλο αποτελείται από ένα μονοστάδιο, σε σύγκριση με τα 3 στάδια στο R-CNN. Απλώς δέχεται μια εικόνα ως είσοδο και επιστρέφει τις πιθανότητες κλάσης και τα πλαίσια οριοθέτησης των ανιχνευόμενων αντικειμένων.



Εικόνα.31: Αρχιτεκτονική του μοντέλου Fast R-CNN

Πηγή: Faster R-CNN Explained for Object Detection Tasks, https://blog.paperspace.com/faster-r-cnn-explained-object-detection/

Ο χάρτης χαρακτηριστικών από το τελευταίο συνελικτικό επίπεδο τροφοδοτείται σε ένα επίπεδο συγκέντρωσης ROI. Ο λόγος είναι να εξαχθεί ένα διάνυσμα χαρακτηριστικών σταθερού μήκους από κάθε πρόταση περιοχής. Με απλά λόγια, το επίπεδο συγκέντρωσης ROI λειτουργεί διαιρώντας κάθε πρόταση περιοχής σε ένα πλέγμα κελιών. Η λειτουργία max pooling εφαρμόζεται σε κάθε κελί του πλέγματος για να επιστρέψει μια μεμονωμένη τιμή. Όλες οι τιμές από όλα τα κελιά αντιπροσωπεύουν το διάνυσμα χαρακτηριστικών. Εάν το μέγεθος του πλέγματος είναι 2×2, τότε το μήκος του διανύσματος χαρακτηριστικών είναι 4.

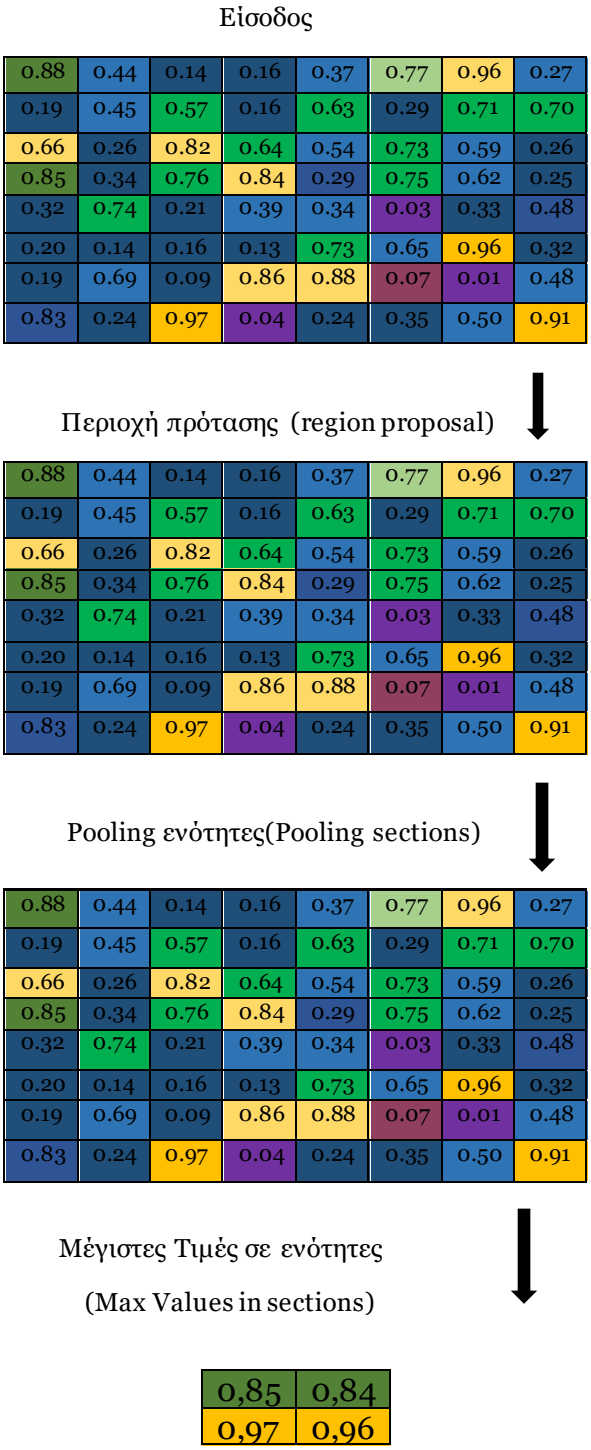
Fast R-CNN, το τελευταίο τμήμα της αρχιτεκτονικής λαμβάνει τους χάρτες χαρακτηριστικών υψηλής ανάλυσης από το ραχοκοκαλιά δικτύου και οι προτάσεις περιοχών από το δίκτυο προτάσεων περιοχής ως είσοδο. Να πάρω ένα σταθερό μέγεθος από τις προτάσεις περιοχών διαφορετικού μεγέθους, χρησιμοποιείται μια μέθοδος που ονομάζεται συγκέντρωση ROI , που σημαίνει συγκέντρωση περιοχών ενδιαφέροντος.

Για κάθε περιοχή ενδιαφέροντος από τη λίστα εισόδου, χρειάζεται ένα τμήμα του χάρτη χαρακτηριστικών εισόδου

που αντιστοιχεί σε αυτό και το κλιμακώνει σε κάποιο προκαθορισμένο μέγεθος, στην περίπτωσή μας 7x7. Η κλιμάκωση γίνεται με:

1. Διαίρεση της πρότασης περιοχής σε τμήματα ίσου μεγέθους, όπου είναι ο αριθμός των ενοτήτων ίδια με τη διάσταση της εξόδου
2. Εύρεση της μεγαλύτερης τιμής σε κάθε ενότητα
3. Αντιγραφή αυτών των μέγιστων τιμών στο buffer εξόδου

Μετά από αυτό υπάρχουν μόνο 2 πλήρως συνδεδεμένα στρώματα που ακολουθούνται από δύο ξεχωριστά στρώματα εξόδου τα οποία προβλέψτε τη βαθμολογία softmax για κάθε τάξη και τα διορθωμένα πλαίσια οριοθέτησης για κάθε περιοχή πρότασης.



Εικόνα.32: Παράδειγμα συγκέντρωσης της περιοχής ενδιαφέροντος

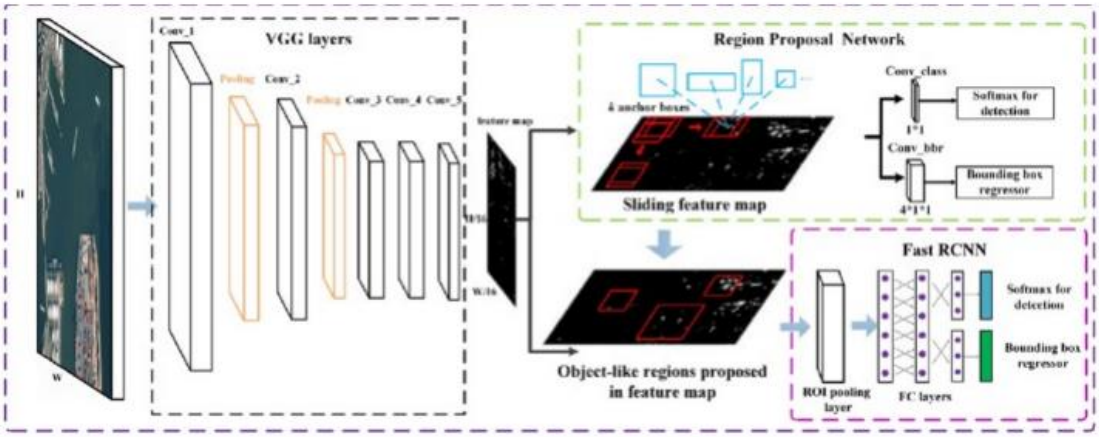
Ως συναρτήσεις απώλειας για την παλινδρόμηση χρησιμοποιούν την απώλεια καταγραφής και για τα πλαίσια οριοθέτησης χρησιμοποιήστε την ομαλή απώλεια L1 και αναδιαδώστε τις απώλειες μαζί με παράγοντες που καθορίζουν τον τρόπο πολύ το πλαίσιο οριοθέτησης και η απώλεια ταξινόμησης θα πρέπει να επηρεάσουν ολόκληρη την απώλεια.

$$L_{loc}(t^{u}, u) = \sum_{i \in \{x,y,w,h\}} smooth_{L1}(t_i^{u} - u_i)$$

$$smooth_{L1}(x) = \begin{cases} 0{,}5x^2 \text{εάν} |x| < 1 \\ |x| - 0{,}5 \text{διαφορετικά} \end{cases}$$

4.6 Faster R-CNN

Η αρχιτεκτονική του Faster R-CNN αποτελείται από τρία μέρη: τον κορμό του δικτύου, την περιοχή δίκτυο πρότασης (RPN) και η παλαιότερη έκδοση αυτού του αλγορίθμου που ονομάζεται γρήγορο R-CNN. Η ραχοκοκαλιά του δικτύου είναι γενικά ένα δίκτυο ταξινόμησης όπως το VGG-Net ή το ResNet προ εκπαιδευμένο ένα σύνολο δεδομένων ταξινόμησης εικόνων. Χρησιμοποιείται για τη δημιουργία χαρτών χαρακτηριστικών υψηλής ανάλυσης και απαιτεί μέγεθος εικόνας 640 x 640 pixel. Στην παρούσα διπλωματική χρησιμοποιήθηκε το ResNet50 ως βασικό δίκτυο προ εκπαιδευμένο στο σύνολο δεδομένων ImageNet.



Εικόνα.33: Αρχιτεκτονική μοντέλου Faster R-CNN

Πηγή: Shilin Zhou Juanping Zhao Zhipeng Deng, Hao Sun. Multi-scale object detection in remote sensing imagery with convolutional neural networks. In ISPRS Journal of Photogrammetry and Remote Sensing 145, 2018.

Το δίκτυο πρότασης περιοχής αποτελείται από ένα ενιαίο συνελικτικό στρώμα το οποίο στη συνέχεια διαιρείται σε 2 ξεχωριστά συνελικτικά επίπεδα για την πρόβλεψη βαθμολογίας ταξινόμησης και οριοθέτησης πλαισίων για το προτάσεις της περιοχής. Το δίκτυο χρησιμοποιεί προκαθορισμένα κουτιά αγκύρωσης για να δημιουργήσει περίπου 2.000 προτάσεις της περιοχής. Μια

άγκυρα είναι ένα μικρότερο μέρος μιας εικόνας. Στην παρούσα διπλωματική χρησιμοποιήθηκαν άγκυρες με τα μεγέθη {128, 256, 512} με αναλογίες διαστάσεων {0,5, 1, 2} που οδηγούν σε 9 διαφορετικά κουτιά αγκύρωσης. Κάθε μία από αυτές τις άγκυρες θα σύρεται πάνω από το παράθυρο με ένα βήμα 16 και θα αφαιρεί τα επικαλυπτόμενα μέρη της εικόνας. Η βαθμολογία κατάταξης αποφασίζει για όλες τις άγκυρες εάν αυτές περιλαμβάνουν ένα αντικείμενο ή όχι. Οι παλινδρομήσεις οριοθέτησης είναι για την καλύτερη αναγνώριση των αντικειμένων στις άγκυρες. Μετά την πρόβλεψη ο αριθμός των προτάσεων περιφέρειας θα μειωθεί, με μη μέγιστη καταστολή.

Κεφάλαιο 5

5.1 Αποτελέσματα

Η εκπαίδευση του YOLO έγινε για 80 epochs με μειωμένο ρυθμό μάθησης 1e-5 και μέγεθος παρτίδας 10. Το Faster R-CNN εκπαιδεύτηκε για 60 epochs με μειωμένο ρυθμό μάθησης 1e-4 και μέγεθος παρτίδας 16. Η κανονικοποιημένη συνολική απώλεια και το mAP της προόδου της προπόνησης φαίνονται στον πίνακα παρακάτω.

	mAP	FPS
YOLO	18,4	210
Faster R-CNN	41,6	17,1

Πίνακας.4: Πίνακας σύγκρισης μοντέλων

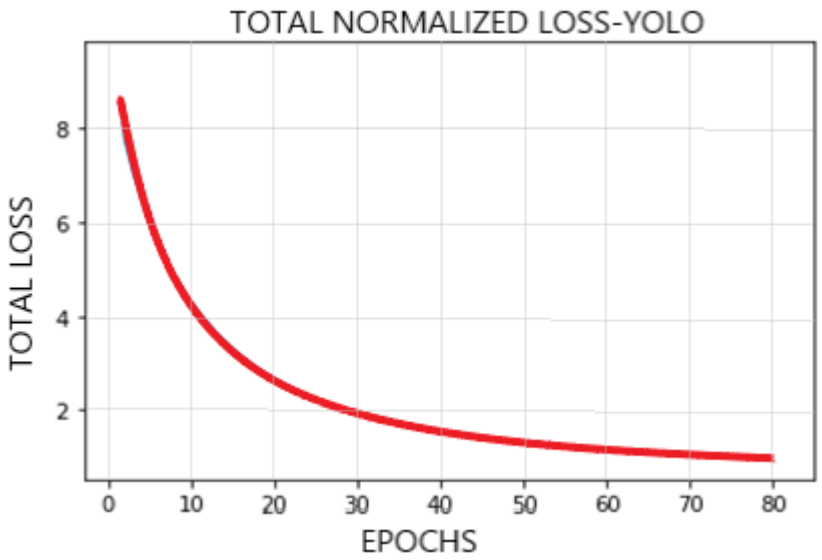
Μετά την εκπαίδευση, οι δύο αλγόριθμοι χρησιμοποιήθηκαν για την αναγνώριση αντικειμένων στο σετ δοκιμής. Τα αποτελέσματα φαίνονται στον πίνακα 4. Το Faster R-CNN έχει μεγαλύτερη ακρίβεια αλλά χαμηλότερο FPS. Συγκριτικά το YOLO έχει πολύ υψηλότερο FPS, αλλά και πολύ χαμηλότερη ακρίβεια γεγονός που οφείλεται στην αρχιτεκτονική του. Το FPS είναι ένας κοινός δείκτης για την αξιολόγηση της ταχύτητας ανίχνευσης μοντέλου. Αναφέρεται στον αριθμό των εικόνων που μπορούν να υποστούν επεξεργασία ανά δευτερόλεπτο.

Γενικά, FPS πάνω από 30 θεωρείται ότι έχει επιτύχει ανίχνευση σε πραγματικό χρόνο. Το Faster R-CNN έχει το υψηλότερο mAP μεταξύ των αλγορίθμων. Σε σύγκριση με το YOLO το mAP του Faster R-CNN είναι 41,6% δηλαδή σχεδόν το διπλάσιο από την τιμή του YOLO που είναι στο 18,4. Αυτό δείχνει ότι ο αλγόριθμος δύο σταδίων έχει πλεονεκτήματα όσον αφορά την ακρίβεια ανίχνευσης σε σύγκριση με τους αλγόριθμους που ολοκληρώνουν την επεξεργασία τους σε ένα στάδιο. Το YOLO μπορεί να προβλέψει πολλαπλούς περιορισμούς κουτιά και τις κατηγορίες τους ταυτόχρονα, και η ταχύτητα ανίχνευσης είναι μεγαλύτερη από εκείνη του Faster R-CNN.

Η ταχύτητα ανίχνευσης για το YOLO υπερβαίνει κατά πολύ τα 30 FPS, κάτι που είναι πολύ πιο γρήγορο από ό,τι στην περίπτωση του Faster R-CNN που έχει μόλις 17,1 FPS. Αν λαμβάνεται υπόψη η αποτελεσματικότητα ανίχνευσης, το YOLO αποδίδει καλύτερα μεταξύ των δύο μοντέλων, ενώ το Faster R-CNN δεν πληροί την απαίτηση σε πραγματικό χρόνο. Αυτό περιορίζει τις δυνατότητές του και καταδεικνύει το πλεονέκτημα του αλγορίθμου ενός σταδίου σχετικά με την ταχύτητα ανίχνευσης.

Με βάση την ανάλυση των παρακάτω πειραματικών αποτελεσμάτων, το Faster R-CNN είναι πιο κατάλληλο εάν απαιτείται η υψηλότερη mAP αναγνώρισης , αλλά το YOLO μπορεί να είναι πιο κατάλληλο για χρήση όταν η προτεραιότητα είναι η απόδοση σε πραγματικό χρόνο και είναι εφικτό να αποδεχθεί ένα ελαφρώς χαμηλότερο mAP. Επομένως, πιστεύουμε ότι το YOLO μπορεί να εφαρμοστεί με μεγαλύτερη αποτελεσματικότητα σε εφαρμογές πραγματικού χρόνου.

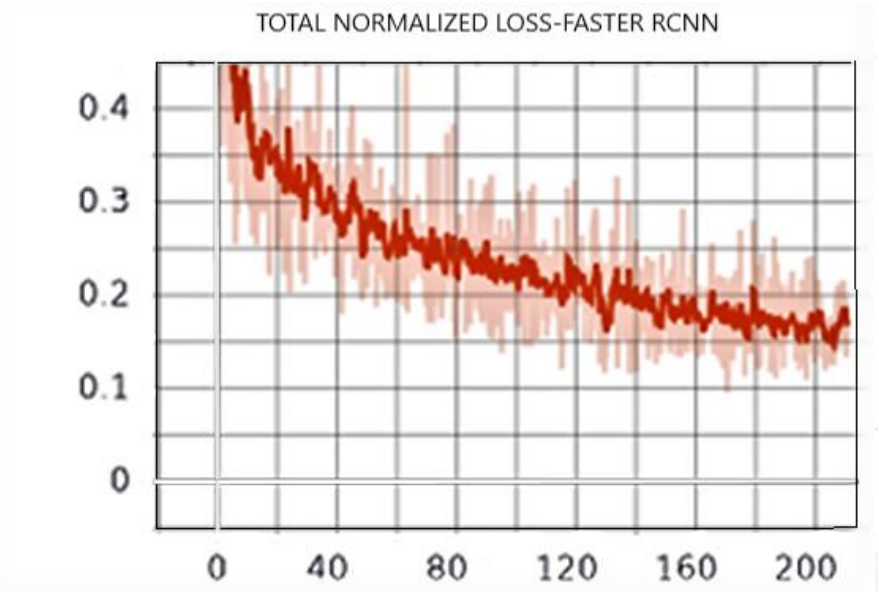
Κατά τη διάρκεια της εκπαίδευσης του συστήματος YOLO, η ομαλοποιημένη συνολική απώλεια παρουσιάζεται στο παρακάτω διάγραμμα:



Εικόνα.34: YOLO-Διάγραμμα ομαλοποιημένης συνολικής απώλειας

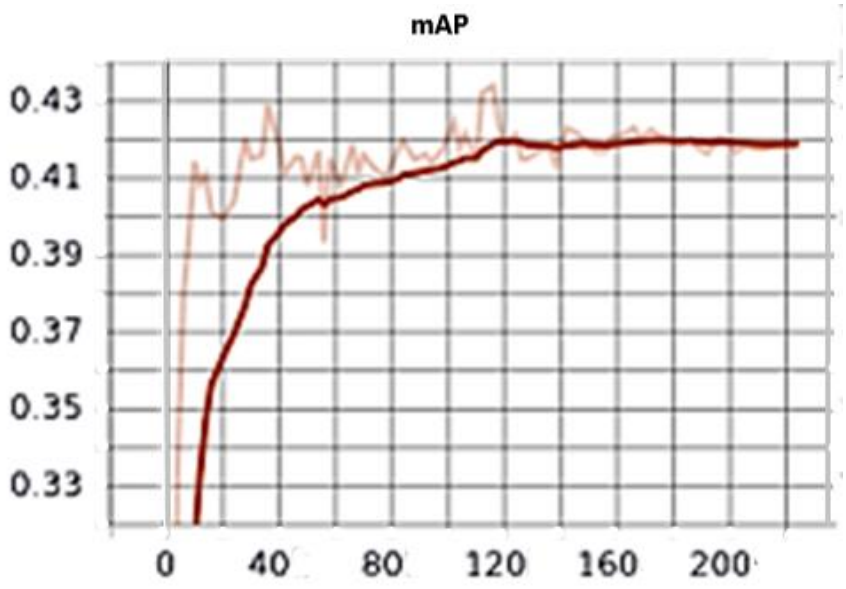
Παρατηρούμε ότι η ομαλοποιημένη συνολική απώλεια για τον αλγόριθμο YOLO μειώνεται όσο αυξάνεται ο αριθμός των epochs.

Και αντίστοιχα κατά τη διάρκεια της εκπαίδευσης του Faster RCNN, η ομαλοποιημένη συνολική απώλεια καθώς και η μέση ακρίβεια(mAP) παρουσιάζεται στα παρακάτω διαγράμματα:



Εικόνα.35: Faster R-CNN- Διάγραμμα ομαλοποιημένης συνολικής απώλειας

Παρατηρούμε ότι η ομαλοποιημένη συνολική απώλεια για τον αλγόριθμο Faster R-CNN είναι μικρότερη συγκριτικά με τον αλγόριθμο YOLO και μειώνεται καθώς αυξάνεται ο αριθμός των βημάτων.

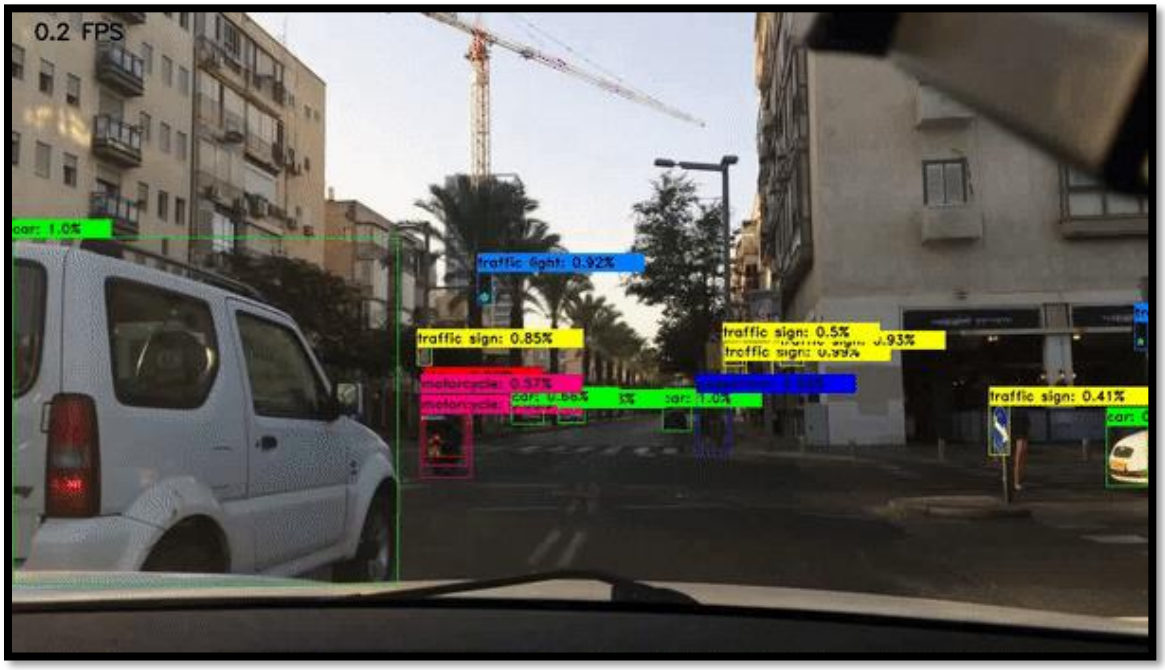


Εικόνα.36: Faster R-CNN- Διάγραμμα μέσης ακρίβειας

Σύμφωνα με το διάγραμμα της εικόνας 36, παρατηρούμε ότι μέση ακρίβεια αυξάνεται όσο αυξάνεται ο αριθμός των βημάτων, και ειδικότερα μετά το πέρας των 30 epochs παρατηρούνται πιο απότομες αλλαγές αύξησης.

Αποτελέσματα της εφαρμογής YOLO και Faster RCCN στο σύνολο δεδομένων BDD100K:

Faster R-CNN

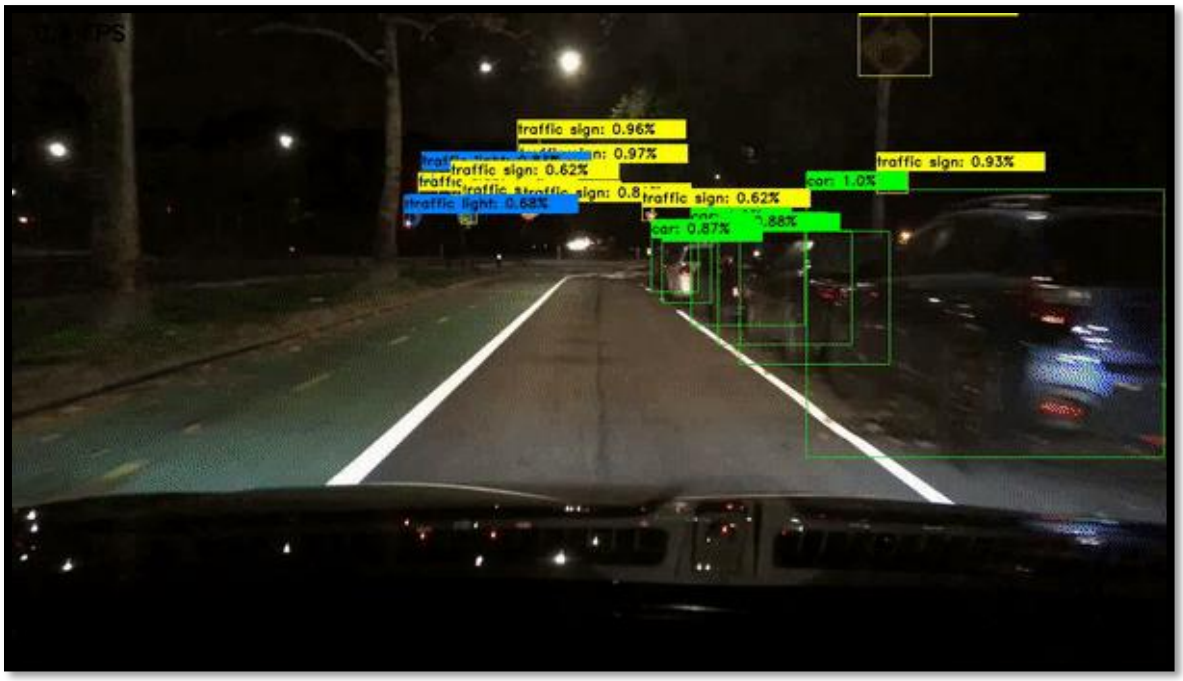


YOLO

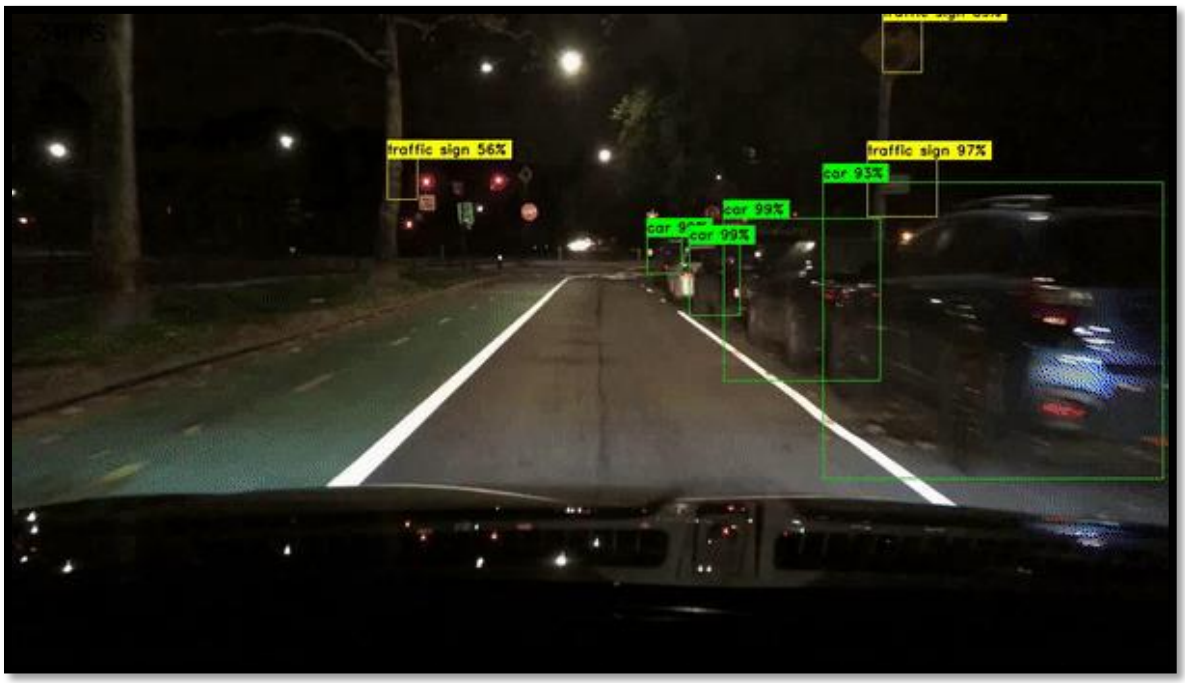
Σύμφωνα με την σύγκριση των παραπάνω εικόνων παρατηρούμε ότι υπό κανονικές συνθήκες οδήγησης και περιβάλλοντος ο αλγόριθμος YOLO δεν εντοπίζει τα μικρά σε μέγεθος αντικείμενα, αντίθετα με τον αλγόριθμο Faster R-CNN που εμφανίζεται πιο αποτελεσματικός στην ανίχνευση μακρινών και

μικρών σε μέγεθος αντικειμένων. Επίσης παρατηρούμε ότι ο αριθμός των frames ανά δευτερόλεπτο (FPS) για τον αλγόριθμο YOLO είναι στα 73 και για το Faster R-CNN στα 0,2 γεγονός που επαληθεύει τα αποτελέσματα των μετρήσεων του πίνακα 4, δηλαδή ότι ο αλγόριθμος YOLO αποδίδει καλύτερα σε πραγματικό χρόνο.

Faster R-CNN



YOLO



Σε συνθήκες νύχτας και χαμηλού φωτισμού και πάλι ο αλγόριθμος Faster R-CNN εντοπίζει τα αντικείμενα που είναι πιο απομακρυσμένα και μικρότερα σε

μέγεθος, όμως ο αλγόριθμος YOLO εμφανίζει σημαντικά υψηλότερα ποσοστά ακρίβειας στη αναγνώριση των αντικειμένων που εντοπίζει.

Faster RCNN

YOLO

Σε συνθήκες περιβάλλοντος όπως το χιόνι που παρεμβάλει σημαντικά στην εικόνα ο αλγόριθμος YOLO καταφέρνει να επιτύχει και πάλι μεγαλύτερο αριθμό frame με περίπου 14 φορές πιο μεγάλο αριθμό από το Faster R-CNN.

Επίσης ο αλγόριθμος Faster R-CNN αναγνωρίζει τα πιο μακρινά και μικρότερα αντικείμενα με μεγαλύτερη ακρίβεια από αυτή του αλγορίθμου YOLO.

Faster RCNN

YOLO

Σε συνθήκες περιβάλλοντος με βροχή ο αλγόριθμος Faster R-CNN καταφέρνει να αναγνωρίσει μακρινά και μικρότερα σε μέγεθος αντικείμενα σε σύγκριση με τον αλγόριθμο YOLO, ο οποίος καταφέρνει να επιτύχει πολύ υψηλότερη ακρίβεια στα αντικείμενα που εντοπίζει. Επίσης και πάλι στη συγκεκριμένη

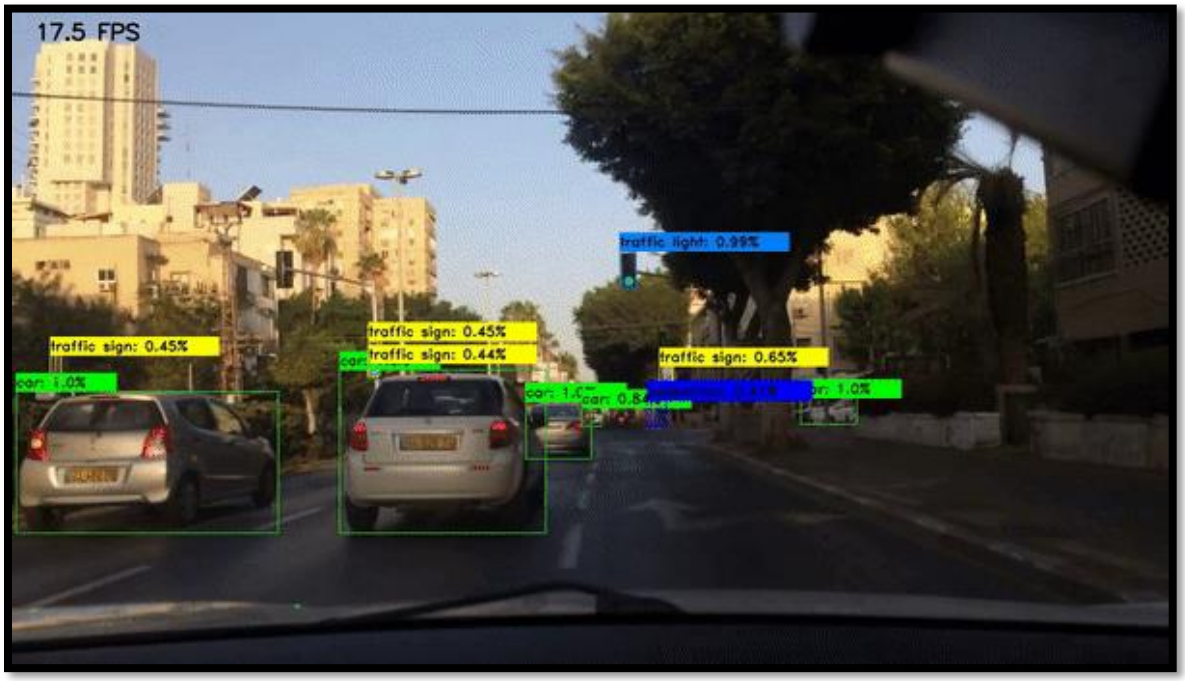
περίπτωση ο αλγόριθμος YOLO παρουσιάζει περίπου 6 φορές μεγαλύτερο αριθμό frame από το Faster R-CNN.

Faster RCNN

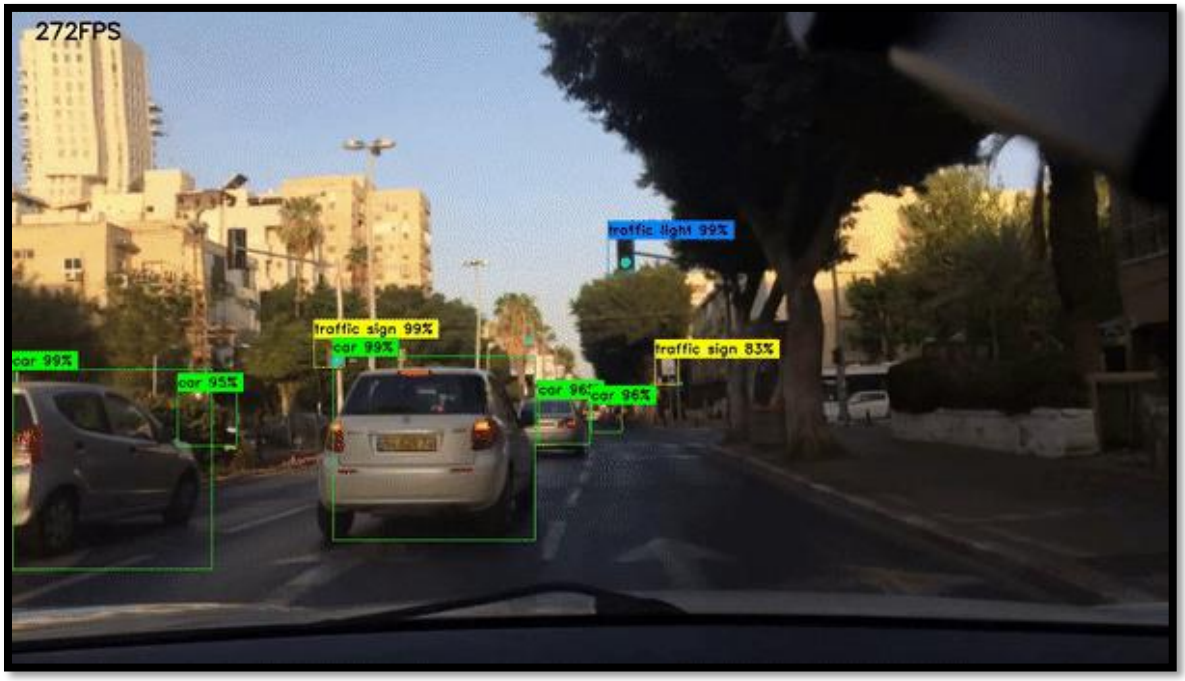
YOLO

Σε ένα ακόμη παράδειγμα με συνθήκες νύχτας και χαμηλού φωτισμού οι δύο αλγόριθμοι καταφέρνουν να εντοπίσουν τα ίδια αντικείμενα με τον αλγόριθμο YOLO να πετυχαίνει καλύτερη ακρίβεια πρόβλεψης των αντικειμένων.

Faster RCNN



YOLO



Επίσης σε αυτό το παράδειγμα με κανονικές συνθήκες περιβάλλοντος για ακόμη μία φορά ο αλγόριθμος Faster R-CNN εντοπίζει τα μακρινά αντικείμενα αλλά ο αλγόριθμος YOLO πετυχαίνει μεγαλύτερη ακρίβεια πρόβλεψης καθώς επίσης και ο αριθμός των frames είναι περίπου 16 φορές μεγαλύτερος από εκείνο του Faster R-CNN.

5.2 Συμπεράσματα

Μέσα από την σύγκριση των δύο μοντέλων προκύπτει ότι το YOLO και το Faster RCNN μοιράζονται σαφώς κάποιες ομοιότητες. Και τα δύο χρησιμοποιούν μια δομή δικτύου που βασίζεται σε κουτί αγκύρωσης, και επίσης και τα δύο χρησιμοποιούν παλινδρόμηση. Πράγματα που διαφέρουν στο YOLO από το Faster RCNN είναι ότι πραγματοποιεί ταξινόμηση και παλινδρόμηση οριοθέτησης ταυτόχρονα. Κρίνοντας από την έκδοση, είναι λογικό ότι το YOLO απαιτούσε έναν πιο αποτελεσματικό τρόπο για να κάνει παλινδρόμηση και ταξινόμηση. Ωστόσο, το YOLO έχει το μειονέκτημά του στην ανίχνευση αντικειμένων. Το YOLO δυσκολεύεται να ανιχνεύσει αντικείμενα που είναι μικρά και κοντά το ένα στο άλλο λόγω μόνο δύο κουτιών αγκύρωσης σε ένα πλέγμα που προβλέπουν μόνο μία κατηγορία αντικειμένων. Δεν γενικεύεται καλά όταν τα αντικείμενα στην εικόνα εμφανίζουν σπάνιες πτυχές αναλογίας. Το Faster R-CNN από την άλλη, ανιχνεύει καλά μικρά αντικείμενα, καθώς περιλαμβάνει άγκυρες σε ένα μόνο πλέγμα, ωστόσο αποτυγχάνει ορισμένες φορές να κάνει ανίχνευση σε πραγματικό χρόνο με την αρχιτεκτονική δύο βημάτων. Στα πλαίσια της αυτόνομης οδήγησης, με την παρούσα διπλωματική υλοποιήθηκε και εκπαιδεύτηκε ο ανιχνευτής ενός σταδίου YOLO και ο ανιχνευτής δύο σταδίων Faster R-CNN στο σύνολο δεδομένων BDD100K. Τα αποτελέσματα της αξιολόγησης και σύγκρισης των δύο μοντέλων έδειξαν ότι το Faster R-CNN έχει μεγαλύτερη ακρίβεια αλλά χαμηλότερο FPS, συγκριτικά με το YOLO το οποίο παρουσιάζει πολύ υψηλότερο FPS, αλλά και πολύ χαμηλότερη ακρίβεια λόγω της πιο απλής αρχιτεκτονικής του. Μελλοντική εργασία θα μπορούσε να περιλαμβάνει περαιτέρω πειράματα με νεότερα μοντέλα, όπως οι πιο πρόσφατα ανεπτυγμένες εκδόσεις του YOLO, καθώς για την ανάπτυξη του κώδικα στην παρούσα διπλωματική χρησιμοποιήθηκε η πρώτη έκδοση του YOLO. Επίσης μακροπρόθεσμα θα μπορούσε να αναπτυχθεί η επίτευξη αποτελεσμάτων με μεγαλύτερη ακρίβεια και υψηλότερο FPS που είναι πιο ωφέλιμα για τις εφαρμογές της αυτόνομης οδήγησης.

Κεφάλαιο 6

6.1 Βιβλιογραφία-Αναφορές

[1] Gene Lewis. Object detection for autonomous vehicles, 2014.

[2] Ross Girshick. Fast r-cnn. In Proceedings of the IEEE international conference on computer
vision, pages 1440–1448, 2015.

[3] Jason Brownlee. A gentle introduction to object recognition with deep learning. Machine
Learning Mastery, 5, 2019.

[4] Fisher Yu, Haofeng Chen, Xin Wang, Wenqi Xian, Yingying Chen, Fangchen Liu, Vashisht Madhavan, and Trevor Darrell. Bdd100k: A diverse driving dataset for heterogeneous multitask learning. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 2636–2645, 2020.

[5] Prajjwal Bhargava. On generalizing detection models for unconstrained environments. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, pages 0–0, 2019.

[6] Licheng Jiao, Fan Zhang, Fang Liu, Shuyuan Yang, Lingling Li, Zhixi Feng, and Rong Qu. A survey of deep learning-based object detection. IEEE Access, 7:128837–128868, 2019.

[7] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. arXiv preprint arXiv:1506.01497, 2015.

[8] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 779–788, 2016.

[9] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The kitti dataset. The International Journal of Robotics Research, 32(11):1231–1237, 2013.

[10] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, et al. Scalability in perception for autonomous driving: Waymo open dataset. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 2446–2454, 2020.

[11] Holger Caesar, Varun Bankiti, Alex H Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 11621–11631, 2020.

[12] Sik-Ho Tsang. Review: Yolov1 — you only look once (object detection). https://towardsdatascience.com/yolov1-you-only-look-once-object-detection-e1f3ffec8a89 zuletzt besucht:
15.03.21, 2018.

[13] Manish Chablani. Yolo – you only look once, real time object detection explained. https://towardsdatascience.com/yolo-you-only-look-once-real-time-object-detection-explained 492dc9230006
zuletzt besucht: 15.03.21, 2017.
11
[14] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1–9, 2015.

[15] Shilin Zhou Juanping Zhao Zhipeng Deng, Hao Sun. Multi-scale object detection in remote sensing imagery with convolutional neural networks. In ISPRS Journal of Photogrammetry and Remote Sensing 145, 2018.

[16] Mata. faster rcnn in rpn the anchor, sliding windows, proposals of understanding. https://www.programmersought.com/article/31012543832/ zuletzt besucht: 15.03.21.

[17] Tomasz Grel. Region of interest pooling explained. https://deepsense.ai/ region-of-interest-pooling-explained/ zuletzt besucht: 15.03.21, 2017.

[18]Girshick, Ross, et al. "Rich feature hierarchies for accurate object detection and semantic segmentation." Proceedings of the IEEE conference on computer vision and pattern recognition. 2014.

[19]Girshick, Ross. "Fast r-cnn." Proceedings of the IEEE international conference on computer vision. 2015.

[20]Ren, Shaoqing, et al. "Faster r-cnn: Towards real-time object detection with region proposal networks." Advances in neural information processing systems. 2015.

[21]He, Kaiming, et al. "Mask r-cnn." Proceedings of the IEEE international conference on computer vision. 2017.