



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ Μ/Υ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
ΣΧΟΛΗ ΝΑΥΤΙΛΙΑΣ ΚΑΙ ΒΙΟΜΗΧΑΝΙΑΣ
ΤΜΗΜΑΤΟΣ ΒΙΟΜΗΧΑΝΙΚΗΣ ΔΙΟΙΚΗΣΗΣ & ΤΕΧΝΟΛΟΓΙΑΣ
ΔΙΑΠΑΝΕΠΙΣΤΗΜΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«ΤΕΧΝΟ-ΟΙΚΟΝΟΜΙΚΑ ΣΥΣΤΗΜΑΤΑ»



ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

"Μηχανική όραση για ανίχνευση ανωμαλιών σε οδικές υποδομές με χρήση νευρωνικών δικτύων"

ΔΑΜΙΑΝΟΣ ΚΑΛΑΝΙΔΗΣ

Επιβλέπων καθηγητής: ΔΟΥΛΑΜΗΣ ΑΝΑΣΤΑΣΙΟΣ ΑΝΑΠΛΗΡΩΤΗΣ ΚΑΘΗΓΗΤΗΣ Ε.Μ.Π

ΑΘΗΝΑ, ΦΕΒΡΟΥΑΡΙΟΣ 2022

ΠΕΡΙΛΗΨΗ

Οι ανωμαλίες σε οδικές υποδομές είναι ένα αρκετά συχνό φαινόμενο. Με την ανίχνευση μέσω της μηχανικής όρασης μπορούμε να ανιχνεύουμε και να κατηγοριοποιούμε αυτές τις ανωμαλίες με έναν συστηματικό και αυτοματοποιημένο τρόπο. Αυτό θα βοηθούσε στην καλύτερη συντήρηση των οδικών υποδομών αλλά και στα συστήματα ADAS (Advanced Driving Assistance Systems) των οχημάτων για την αποφυγή αυτών των ανωμαλιών. Στην παρούσα εργασία γίνεται μια συνοπτική περιγραφή βασικών ορολογιών της μηχανικής όρασης και της αναγνώρισης προτύπων, των βαθιών νευρωνικών δικτύων καθώς και των ανωμαλιών οδοστρωμάτων. Στην συνέχεια μια περιγραφή του framework YOLO, της μεθοδολογίας για την διεξαγωγή πειραματικών μετρήσεων με το framework αυτό καθώς και παρουσίαση των παραγόμενων αποτελεσμάτων. Τέλος γίνεται μια συνοπτική τεchnο-οικονομική ανάλυση σχετικά με τις επίδραση των φθορών των οδοστρωμάτων καθώς και τη εφαρμογή των συστημάτων ανίχνευσης ανωμαλιών.

ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ: Μηχανική όραση

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: Μηχανική όραση, Επεξεργασία εικόνας, Νευρωνικά δίκτυα, Προηγμένα συστήματα υποβοήθησης οδηγού, Οδικές υποδομές

ABSTRACT

Irregularities in road infrastructure are a fairly common phenomenon. By detecting through mechanical vision, we can detect and categorize these abnormalities in a systematic and automated way. This would help in the better maintenance of the road infrastructure but also in the ADAS (Advanced Driving Assistance Systems) of the vehicles to avoid these anomalies. In the present work, a brief description of the basic terminology of mechanical vision and pattern recognition, deep neural networks as well as road irregularities is given. Following is a description of the YOLO framework, the methodology for conducting experimental measurements with this framework as well as a presentation of the results produced. Finally, a brief techno-economic analysis is made regarding the effect of road damage as well as the application of anomaly detection systems.

SUBJECT AREA: Computer Vision

KEYWORDS: Computer Vision, Image processing, Neural networks, ADAS, Road infrastructure

Μηχανική όραση για ανίχνευση ανωμαλιών σε οδικές υποδομές με χρήση νευρωνικών δικτύων

Στην οικογένεια μου και τους φίλους μου

ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα καθηγητή κ. Αναστάσιο Δουλάμη για την εμπιστοσύνη που μου έδειξε με την ανάθεση του θέματος και για την εξαιρετική συνεργασία.

Επίσης τον υποψήφιο διδάκτορα Ιάσονα Κατσαμένη για την πολύτιμη βοήθεια του και τις επισημάνσεις του κατά την συγγραφή της διπλωματικής.

ΠΕΡΙΕΧΟΜΕΝΑ

1.	Εισαγωγικές έννοιες και θεωρία μηχανικής όρασης.....	8
1.1.	Μηχανική όραση.....	8
1.2.	Αναγνώριση προτύπων (Pattern Recognition).....	9
1.2.1.	Αναγνώριση στιγμιότυπων (instance recognition).....	9
1.2.2.	Οπτική αναγνώριση κατηγορίας (category recognition) και κατηγοριοποίηση εικόνων (image classification).....	10
1.2.3.	Τμηματικά μοντέλα (Part-based models).....	12
1.3.	Deep networks	14
1.3.1.	Deep neural networks.....	15
1.3.2.	Συνελκτικά νευρωνικά δίκτυα (Convolutional neural networks)	18
1.3.3.	Εκπαίδευση και βελτιστοποίηση (Training and optimization)	19
1.4.	Yolo.....	21
1.4.1.	YOLO v1	21
1.4.2.	YOLO v2 (YOLO9000).....	23
1.4.3.	YOLO v3	25
1.5.	Οδοστρώματα	27
1.5.1.	Φθορές Οδοστρωμάτων	27
1.6.	ADAS (Advanced Driving Assistance Systems)	31
2.	Ανίχνευση ανωμαλιών οδοστρώματος με χρήση μηχανικής όρασης	34
2.1.	Μεθοδολογία.....	34
2.1.1.	Yolo.....	34
2.1.2.	Δομή αρχείου Yolo	35
2.1.3.	Dataset.....	35
2.1.4.	Προετοιμασία Dataset.....	40
2.1.5.	Δείγμα Εκπαίδευσης-Εκπαίδευσης-Δοκιμής	42
2.2.	Εκπαίδευση μοντέλου	43
2.3.	Δοκιμή μοντέλου και ανίχνευση	47
3.	Τεχνο-οικονομική ανάλυση	55
3.1.	Ατυχήματα Λόγω Κακής Συντήρησης Οδοστρώματος και έμμεσα/ άμεσα κόστη	55
3.2.	Κόστος αποκατάστασης φθορών οδοστρώματος.....	57
3.3.	Κόστος υλοποίησης συστημάτων ανίχνευσης	59

4. Συμπεράσματα – Μελλοντική έρευνα	60
Βιβλιογραφικές αναφορές.....	61

1. Εισαγωγικές έννοιες και θεωρία μηχανικής όρασης

1.1.Μηχανική όραση

Ως άνθρωποι, αντιλαμβανόμαστε την τρισδιάστατη δομή του κόσμου γύρω μας με εμφανή ευκολία. Ένα παράδειγμα αυτής της ικανότητας αντίληψης του τρισδιάστατου χώρου, είναι κοιτώντας ένα βάζο σε ένα τραπέζι μπορούμε να διακρίνουμε τον αριθμό των φύλλων, την απόχρωσή τους, την ανάκλαση του φωτός πάνω του και την διάκριση του ίδιου του φυτού από το φόντο πίσω του. Η μηχανική όραση, υπολογιστική όραση ή τεχνητή όραση είναι ένα επιστημονικό πεδίο της τεχνητής νοημοσύνης το οποίο επιχειρεί να αναπαράγει αλγοριθμικά την αίσθηση της όρασης, συνήθως σε ηλεκτρονικό υπολογιστή ή ρομπότ. Οι επιστήμονες στη μηχανική όραση, αναπτύσσουν παράλληλα μαθηματικά μοντέλα για την ανάκτηση των τρισδιάστατων σχημάτων και εμφάνισης των αντικειμένων σε μια εικόνα. Για παράδειγμα μπορούμε πλέον με αρκετή ακρίβεια να αναπαραστήσουμε μια τρισδιάστατη επιφάνεια από ένα σύνολο επικαλυπτόμενων εικόνων ή να αναγνωρίσουμε ένα άτομο που κινείται μπροστά από ένα περίπλοκο φόντο. Επίσης, η πρόοδος στη μηχανική όραση είναι τέτοια ώστε να έχουμε αρκετές εφαρμογές στην καθημερινότητα. Μερικές από αυτές είναι στην οπτική αναγνώριση χαρακτήρων (OCR), Machine inspection, Retail, στη δημιουργία 3D μοντέλων, στην ιατρική απεικόνιση, ασφάλεια στην αυτοκίνηση, στην αναγνώριση κινούμενων αντικειμένων, στην καταγραφή κίνησης, στα συστήματα επιτήρησης, στην αναγνώριση αποτυπωμάτων και άλλων βιομετρικών χαρακτηριστικών.

Ωστόσο, παρόλη την πρόοδο, η δυνατότητα του υπολογιστή να αναγνωρίσει μια εικόνα στο ίδιο επίπεδο όπως ένα παιδί λίγων ετών, μοιάζει για τώρα κάτι απατηλό. Γιατί είναι η όραση τόσο δύσκολη; Γιατί η όραση είναι ένα αντίστροφο πρόβλημα, στο οποίο αναζητούμε να ανακτήσουμε κάποιες άγνωστες μεταβλητές δεδομένου ανεπαρκών πληροφοριών για να προσδιοριστεί πλήρως η λύση. Ο πραγματικός κόσμος είναι φτιαγμένος από ένα μπέρδεμα αντικειμένων, που όλα περιπλέκονται μεταξύ τους και εμφανίζονται σε διαφορετικές πόζες. Επιπλέον, η εγγενής μεταβλητότητα σε μια κατηγορία αντικειμένων (π.χ., σκύλοι), λόγω σύνθετης μη άκαμπτης άρθρωσης και ακραίων παραλλαγών στο σχήμα και την εμφάνιση (π.χ. μεταξύ διαφορετικών φυλών), καθιστά όχι τόσο πιθανό να μπορούμε απλά να εφαρμόσουμε μια εξαντλητική αντιστοίχιση σε βάση δεδομένων υποδειγμάτων.

Η πιο δύσκολη εκδοχή αναγνώρισης είναι η αναγνώριση γενικής κατηγορίας (ή τάξης), η οποία μπορεί να περιλαμβάνει την αναγνώριση περιπτώσεων εξαιρετικά διαφορετικών κατηγοριών, όπως ζώα ή έπιπλα. Ορισμένες τεχνικές βασίζονται αποκλειστικά στην παρουσία χαρακτηριστικών, τις σχετικές θέσεις τους, ενώ άλλες περιλαμβάνουν την κατάτμηση της εικόνας σε σημασιολογικά σημαντικές περιοχές. Σε πολλές

περιπτώσεις, η αναγνώριση εξαρτάται σε μεγάλο βαθμό από το περιβάλλον των γύρω αντικειμένων και των στοιχείων της σκηνής.

Ενταγμένο σε όλες αυτές οι τεχνικές είναι το θέμα της μάθησης, δεδομένου ότι η χειροποίητη δημιουργία συγκεκριμένων ανιχνευτών αντικειμένων μοιάζει με μάταιη προσέγγιση δεδομένης της πολυπλοκότητας του προβλήματος.

1.2. Αναγνώριση προτύπων (Pattern Recognition)

Η επιστημονική περιοχή της αναγνώρισης προτύπων ασχολείται με την αυτόματη ανακάλυψη των υποκείμενων κανονικοτήτων στα δεδομένα με τη χρήση αλγορίθμων έτσι ώστε να πάρει αποφάσεις, όπως η ταξινόμηση των δεδομένων σε διαφορετικές κατηγορίες.

Ιστορικά υπήρχαν οι δύο πρώτες θεμελιώδεις προσεγγίσεις για την εφαρμογή ενός συστήματος αναγνώρισης προτύπων με την τρίτη να έχει ραγδαία εξέλιξη τα τελευταία χρόνια:

1. Στατιστική Αναγνώριση Προτύπων (ή θεωρητική απόφαση)

Βασίζεται σε υποκείμενα στατιστικά μοντέλα για την περιγραφή των προτύπων και των τάξεων τους.

2. Συντακτική Αναγνώριση Προτύπων (ή δομική).

Οι τάξεις αναπαριστώνται μέσω τυπικών δομών όπως γραμματική, automata, strings.

3. Νευρωνική Αναγνώριση Προτύπων

Ο ταξινομητής απεικονίζεται ως ένα δίκτυο από μικρές αυτόνομες μονάδες που επιτελούν ένα μικρό αριθμό από συγκεκριμένες πράξεις- «κύτταρα» που μιμούνται τους νευρώνες του ανθρώπινου εγκεφάλου.

Η γενική αναγνώριση αντικειμένων εμπίπτει σε δύο μεγάλες κατηγορίες, συγκεκριμένα την αναγνώριση στιγμιότυπων (instance recognition) και αναγνώριση κλάσεων (class recognition). Η πρώτη περιλαμβάνει την εκ νέου αναγνώριση ενός γνωστού 2D ή 3D άκαμπτου αντικειμένου, το οποίο ενδεχομένως παρατηρείται από μια νέα οπτική γωνία, έναντι ενός σε ένα ακατάστατου φόντου και με επιμέρους σκιάσεις. Η δεύτερη περίπτωση, η οποία είναι επίσης γνωστή ως αναγνώριση επιπέδου κατηγορίας ή αναγνώριση γενικού αντικειμένου, είναι το πολύ πιο δύσκολο πρόβλημα της αναγνώρισης οποιαδήποτε στιγμιότυπου συγκεκριμένης γενικής τάξης, όπως «γάτα», «αυτοκίνητο» ή «δέντρο».

1.2.1. Αναγνώριση στιγμιότυπων (instance recognition)

Με την πάροδο των ετών, έχουν αναπτυχθεί πολλοί διαφορετικοί αλγόριθμοι για την αναγνώριση στιγμιότυπων. Στην αρχή οι προσεγγίσεις επικεντρώνονταν στην εξαγωγή γραμμών, περιγραμμάτων ή τρισδιάστατων επιφανειών από εικόνες και αντιστοίχιση τους με γνωστά μοντέλα τρισδιάστατων αντικειμένων^[1]. Μία άλλη δημοφιλής προσέγγιση ήταν να αποκτήσουμε εικόνες από ένα μεγάλο σύνολο σημείων παρατήρησης και επίπεδων φωτισμού και να τις αναπαραστήσουμε χρησιμοποιώντας την μέθοδο της αποσύνθεσης ιδιοτιμών. Καθώς ο αριθμός των αντικειμένων ή εικόνων σε μια συλλογή αρχίζει να γίνεται πολύ μεγάλος, ο χρόνος που χρειάζεται για να συγκρίνουμε μια νέα εικόνα με κάθε εικόνα της βάσης δεδομένων μπορεί να γίνει απαγορευτικός. Αντί να συγκρίνουμε τις εικόνες μία κάθε φορά, απαιτούνται τεχνικές για τον γρήγορο περιορισμό της αναζήτησης σε μερικές πιθανές εικόνες, οι οποίες στη συνέχεια μπορούν να συγκριθούν χρησιμοποιώντας ένα πιο λεπτομερές και συντηρητικό στάδιο επαλήθευσης. Το πρόβλημα της γρήγορης εύρεσης μερικών αντιστοιχιών μεταξύ εγγράφων είναι ένα από τα κεντρικά προβλήματα στην ανάκτηση πληροφοριών (Information Retrieval)^[2]. Η βασική προσέγγιση στους αλγόριθμους γρήγορης ανάκτησης εγγράφων είναι ο εκ των προτέρων υπολογισμός ενός ανεστραμμένου ευρετηρίου μεταξύ μεμονωμένων λέξεων και των εγγράφων (ή ιστοσελίδων ή ειδήσεων) όπου εμφανίζονται. Πιο συγκεκριμένα, η συχνότητα εμφάνισης συγκεκριμένων λέξεων σε ένα έγγραφο χρησιμοποιείται για τη γρήγορη εύρεση εγγράφων που ταιριάζουν με ένα συγκεκριμένο ερώτημα. Οι Sivic και Zisserman^[3] ήταν οι πρώτοι που προσαρμόσαν τις τεχνικές IR στην οπτική αναζήτηση.

1.2.2. Οπτική αναγνώριση κατηγορίας (category recognition) και κατηγοριοποίηση εικόνων (image classification)

Η οπτική αναγνώριση κατηγορίας είναι ένα εξαιρετικά απαιτητικό πρόβλημα^[4]. Ενώ οι τεχνικές αναγνώρισης στιγμιότυπων είναι σχετικά ώριμες και χρησιμοποιούνται σε εμπορικές εφαρμογές, η αναγνώριση γενικής κατηγορίας (κλάσης) εξακολουθεί να είναι ένα σε μεγάλο βαθμό άλυτο πρόβλημα. Σκεφτείτε για παράδειγμα το σύνολο φωτογραφιών στο παρακάτω σχήμα, το οποίο δείχνει αντικείμενα που λαμβάνονται από 10 διαφορετικές οπτικές κατηγορίες. Πώς θα γράφατε ένα πρόγραμμα για να κατηγοριοποιήσετε καθεμία από αυτές τις εικόνες στην κατάλληλη κλάση, ειδικά αν σας δόθηκε επίσης η επιλογή "κανένα από τα παραπάνω»;



Εικόνα 1 : Εικόνες διαφορετικών οπτικών κατηγοριών

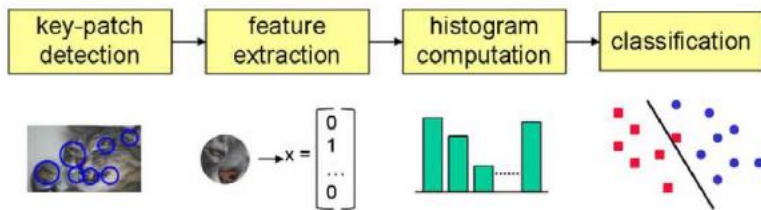
Όπως προκύπτει από αυτό το παράδειγμα, η οπτική αναγνώριση κατηγοριών είναι ένα εξαιρετικά δύσκολο πρόβλημα. Κανείς δεν έχει κατασκευάσει ακόμη ένα σύστημα που να προσεγγίζει το επίπεδο απόδοσης ενός παιδιού δύο ετών. Ωστόσο, η πρόοδος στον τομέα ήταν αρκετά δραματική, αν κριθεί το πόσο καλύτεροι είναι οι σημερινοί αλγόριθμοι σε σύγκριση με αυτούς της προηγούμενης δεκαετίας.

Υπάρχουν διάφορες προσεγγίσεις για την κατηγοριοποίηση των εικόνων (Image classification) με κυριότερες να είναι:

- Οι κλασικές προσεγγίσεις βασισμένες σε χαρακτηριστικά (part-based representations) που βασίζονται σε χειροποίητα χαρακτηριστικά και τα στατιστικά τους και προαιρετικά χρησιμοποιούν μηχανική εκμάθηση για την τελική κατηγοριοποίηση
- Τα σύγχρονα συστήματα ταξινόμησης εικόνων, τα οποία βασίζονται στα βαθιά νευρωνικά δίκτυα (deep neural networks)

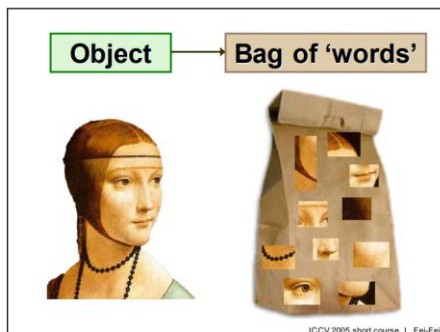
Στις προσεγγίσεις βασισμένες σε χαρακτηριστικά, αρχικά ήταν προτεινόμενη προσέγγιση οι τμηματικές αναπαραστάσεις και οι αλγόριθμοι τμηματικής αναγνώρισης, ωστόσο υπάρχουν και οι απλούστερες τύπου τσάντας των λέξεων (bag-of-features) προσεγγίσεις που αντιπροσωπεύουν αντικείμενα και εικόνες ως μη ταξινομημένες συλλογές από περιγραφείς χαρακτηριστικών.

Οι Csurka, Dance, Fan, κ.α. (2004)^[5] ήταν οι πρώτοι που χρησιμοποίησαν τον όρο τσάντα των λέξεων (bag of words) ή τσάντα με βασικά σημεία (bag of keypoints) για να περιγράψουν τέτοιες προσεγγίσεις και από τους πρώτους που απέδειξαν τη χρησιμότητα των τεχνικών που βασίζονται στη συχνότητα για την αναγνώριση κατηγοριών. Στη προσέγγιση της τσάντας των λέξεων, ο αλγόριθμος υπολογίζει απλώς την κατανομή (ιστόγραμμα) των οπτικών λέξεων που βρίσκονται στην εικόνα ερωτήματος και συγκρίνει αυτήν την κατανομή με αυτές που βρίσκονται στις εικόνες εκπαίδευσης.



Εικόνα 2 : Μια τυπική επεξεργαστική ροή για ένα σύστημα αναγνώρισης κατηγορίας τσάντας των λέξεων

Η μεγαλύτερη διαφορά από την αναγνώριση στιγμιότυπου είναι η απουσία ενός σταδίου γεωμετρικής επαλήθευσης. Το αρχικό σύστημά τους χρησιμοποίησε συγγενικές περιοχές συμμεταβλητής και περιγραφείς SIFT, κατασκευή οπτικού λεξιλογίου k-means και μαζί έναν Naïve Bayesian ταξινομητή και μηχανές υποστήριξης διανυσμάτων για ταξινόμηση.



Εικόνα 3 : Οπτική απεικόνιση αντικειμένου και τσάντας των λέξεων

1.2.3. Τμηματικά μοντέλα (Part-based models)

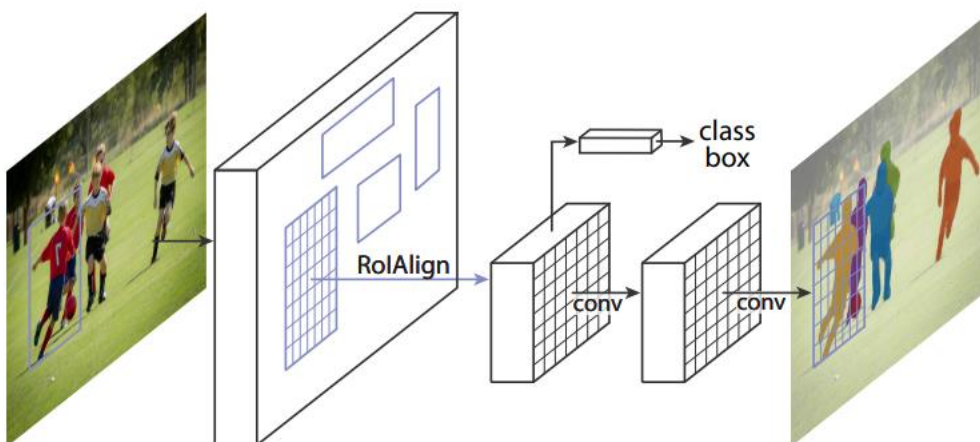
Η αναγνώριση ενός αντικειμένου με την εύρεση των συστατικών του μερών και τη μέτρηση των γεωμετρικών τους σχέσεων είναι μια από τις παλαιότερες προσεγγίσεις για την αναγνώριση

αντικειμένων^[6]. Παραδείγματα τέτοιων τμηματικών προσεγγίσεων (Part-based approaches) χρησιμοποιούνται για την αναγνώριση προσώπου και την ανίχνευση πεζών^{[7][8]}. Μερικά από τα κεντρικά ζητήματα της αναγνώρισης βάσει τμηματικών προσεγγίσεων είναι η αναπαράσταση γεωμετρικών σχέσεων, η αναπαράσταση μεμονωμένων μερών και οι αλγόριθμοι για την εκμάθηση τέτοιων περιγραφών και την αναγνώρισή τους κατά το χρόνο εκτέλεσης. Αυτά τα τμηματικά μοντέλα μπορούν να έχουν διαφορετικές τοπολογίες για τις γεωμετρικές συνδέσεις μεταξύ τους. Σε αντίθεση με τη προσέγγιση Bag of words δεν αρκεί ο εντοπισμός των επιμέρους χαρακτηριστικών μέσα στην εικόνα, αλλά αναζητείται και η επιβεβαίωση συγκεκριμένων χωρικών σχέσεων μεταξύ αυτών των χαρακτηριστικών.

Αναγνώριση με τμηματοποίηση (Recognition with segmentation)

Η πιο απαιτητική μορφή της αναγνώρισης γενικών αντικειμένων είναι η ταυτόχρονη εκτέλεση αναγνώρισης με ακριβή τμηματοποίηση ορίων^[9], κάτι που μερικές φορές μπορεί να επιτευχθεί με την εκ των υστέρων προβολή του μοντέλου αντικειμένου στη σκηνή ή την αντιστοίχιση τμημάτων της νέας σκηνής με μοντέλα αντικειμένων που έχουν προηγηθεί (τμηματοποιηθεί).

Η κατάτμηση εικόνας είναι το έργο της εύρεσης ομάδων εικονοστοιχείων που «πάνε μαζί». Στη στατιστική, αυτό το πρόβλημα είναι γνωστό ως ανάλυση συστάδων και είναι μια ευρέως μελετημένη περιοχή με εκατοντάδες διαφορετικούς αλγόριθμους.



Εικόνα 4 : Αρχιτεκτονική του framework Mask R-CNN για την τμηματοποίηση στιγμιότυπου (instance segmentation) ^[10]

Κατανόηση πλαισίου και σκηνής (Context and scene understanding)

Το έργο της αναγνώρισης και εντοπισμού αντικειμένων μεμονωμένα από αυτό της κατανόησης της σκηνής/πλαισίου (context) στην οποία εμφανίζεται το αντικείμενο, είναι ένας σοβαρός περιορισμός, καθώς το πλαίσιο παίζει πολύ σημαντικό ρόλο στην αναγνώριση του ανθρώπινου αντικειμένου^[11].

Η μεγαλύτερη διαφορά μεταξύ των τύπου τσάντας των λέξεων μοντέλων και των μοντέλων με βάση το πλαίσιο είναι ότι τα τελευταία συνδυάζουν αντικείμενα σε σκηνές και ο αριθμός των συστατικών αντικειμένων από κάθε κλάση δεν είναι γνωστός εκ των προτέρων. Όταν ο αριθμός των αντικειμένων σε μια σκηνή και των τμημάτων σε κάθε αντικείμενο είναι άγνωστος, χρησιμοποιείται μια λανθάνουσα διεργασία Dirichlet (LDP) για τη μοντελοποίηση της δημιουργίας αντικειμένων και τμημάτων σε ένα πλαίσιο παραγωγής. Οι κατανομές για όλα τα αντικείμενα και τα μέρη μαθαίνονται από μια μεγάλη βάση δεδομένων με ετικέτες και στη συνέχεια χρησιμοποιούνται κατά την εξαγωγή συμπερασμάτων (αναγνώριση) για την επισήμανση των στοιχείων μιας σκηνής.

Η εκμάθηση του μοντέλου και οι μεγάλες συλλογές εικόνων που προσπαθούν να δημιουργήσουν ένα σύστημα αναγνώρισης χωρίς πολλά δεδομένα εκπαίδευσης για οτιδήποτε άλλο εκτός από ένα βασικό μοτίβο, όπως ένας κωδικός UPC, έχουν αποδειχθεί μια θλιβερή αποτυχία. Όταν η βάση δεδομένων των εικόνων γίνει αρκετά μεγάλη, είναι ακόμη δυνατό να αντιστοιχιστούν απευθείας πλήρεις εικόνες με την προσδοκία να βρεθεί μια καλή αντιστοίχιση^[12].

1.3. Deep networks

Από την προηγούμενη δεκαετία, τα βαθιά δίκτυα (deep networks) άρχισαν να ξεπερνούν τις προσεγγίσεις βασισμένες στη «ρηγή» μάθηση, με γνωστό παράδειγμα το διαγωνισμό ImageNet Large Scale Visual Recognition Challenge (ILSVRC) το 2012 όπου κέρδισε το βαθύ νευρωνικό δίκτυο (deep neural network) SuperVision/AlexNet των Krizhevsky, Sutskever, and Hinton^[13].

Από τότε, η ακρίβεια αναγνώρισης συνέχισε να βελτιώνεται δραματικά κυρίως λόγω των deep networks και των καλύτερων αλγορίθμων εκπαίδευσης

1.3.1. Deep neural networks

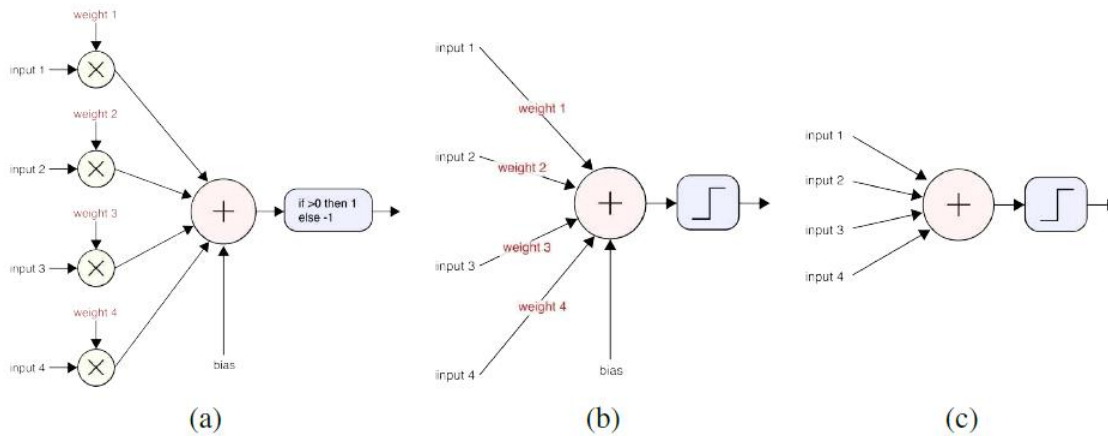
Τα βαθιά νευρωνικά δίκτυα - Deep neural networks (DNNs) παρέχουν μια ομοιόμορφη, διαφοροποιήσιμη αρχιτεκτονική υπολογισμού, ενώ ταυτόχρονα ανακαλύπτουν αυτόματα χρήσιμες εσωτερικές αναπαραστάσεις. Ενδιαφέρον για την κατασκευή υπολογιστικών συστημάτων που μιμούνται τους νευρικούς (βιολογικούς) υπολογισμούς υπήρχε από τα τέλη της δεκαετίας του 1950^[14]. Τα σημερινά πιο δημοφιλή βαθιά νευρωνικά δίκτυα είναι ντετερμινιστικά διακριτικά ανατροφοδοτούμενα δίκτυα με ενεργοποιήσεις πραγματικής αξίας, εκπαιδευμένα με χρήση μεθόδου gradient descent^[15].

Σε σύγκριση με άλλες τεχνικές μηχανικής εκμάθησης, οι οποίες συνήθως βασίζονται σε αρκετά στάδια προεπεξεργασίας για την εξαγωγή χαρακτηριστικών πάνω στα οποία μπορούν να χτιστούν ταξινομητές, οι προσεγγίσεις βαθιάς μάθησης συνήθως εκπαιδεύονται από άκρο σε άκρο, πηγαίνοντας απευθείας από τα πρωτογενή pixel στις τελικές επιθυμητές εξόδους

Τα DNNs είναι ανατροφοδοτούμενοι γράφοι υπολογισμού αποτελούμενοι από χιλιάδες απλούς διασυνδεδεμένους «νευρώνες» (μονάδες), οι οποίοι, όπως στη λογιστική παλινδρόμηση, εκτελούν

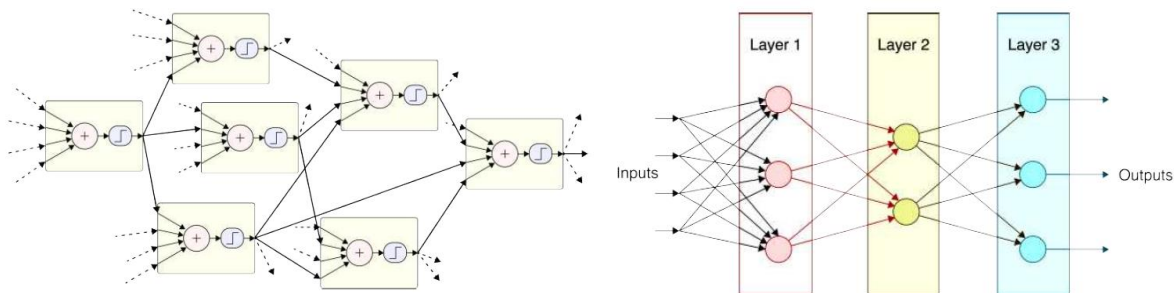
σταθμισμένα αθροίσματα των εισροών τους $s_i = \mathbf{w}_i^T \mathbf{x}_i + b_i$ και στη συνέχεια αντιστοίχιση σε μια μη γραμμική συνάρτησης ενεργοποίησης $y_i = h(s_i)$,

Στην παρακάτω εικόνα βλέπουμε την παραδείγματα τέτοιων μονάδων των DNNs. Οι πρώτες τέτοιες μονάδες ονομάζονταν perceptrons. Η διαγραμματίηση μιας τέτοιας μονάδας φαίνεται στο διάγραμμα a, στην οποία τα βάρη, που έχουν βελτιστοποιηθεί στη διαδικασία της μάθησης, αντιστοιχίζονται ανά είσοδο και γίνονται οι πολλαπλασιασμοί ανά στοιχείο εισόδου, δηλαδή έχουμε το εσωτερικό γινόμενο των διανυσμάτων εισόδου και βαρών. Στο διάγραμμα b, τα βάρη αντιστοιχίζονται ανά μονάδα πάνω στις συνδέσεις. Και στο διάγραμμα c, τα βάρη παραλείπονται και θεωρούμε ότι είναι ενσωματωμένα στις τιμές εισόδου.



Εικόνα 5 : Παραδείγματα μονάδων των DNNs

Αυτές οι μονάδες δεν συνδέονται μεταξύ τους με τη μορφή ακανόνιστου γράφου όπως στην αριστερή εικόνα αλλά συνήθως οργανώνονται σε διαδοχικά στρώματα (layers) όπως στη δεξιά εικόνα.



Εικόνα 6 : Τρόποι διασύνδεσης μονάδων DNN

Κάθε σύνολο από όλες τις μονάδες ενός layer μπορούμε να τις ορίσουμε ως τα στοιχεία ενός διανύσματος και να έχουμε τον ακόλουθο γραμμικό συνδυασμό $s_1 = W_1 x_1$; όπου W_1 ένας πίνακας βαρών με αριθμό γραμμών το πλήθος εισόδων του υπολογιζόμενου layer και αριθμό στηλών το πλήθος μονάδων του υπολογιζόμενου layer και x_1 οι τιμές των εισόδων του υπολογιζόμενου layer. Στο σταθμισμένο αυτό άθροισμα s_1 , εφαρμόζεται ένα σενάριο από συναρτήσεις ενεργοποίησης και έχουμε τον τύπο της συνάρτησης ενεργοποίησης που μας δίνει τις τιμές που δέχονται ως είσοδο οι μονάδες του κάθε επόμενου layer $x_{i+1} = y_i = h(s_i)$. Οι έξοδοι κάθε μονάδας ονομάζονται ενεργοποιήσεις (activations). Κάθε layer στο οποίο εφαρμόζεται ένας πλήρης (dense) πίνακας βαρών ονομάζεται πλήρως συνεκτικό (fully connected - FC).

Κάθε δίκτυο που αποτελείται μόνο από FC layers ονομάζεται multi-layer perceptron (MLP). Κάθε δίκτυο που αποτελείται από μη πλήρη layers ονομάζεται fully convolutional network (FCN). Κατά την επεξεργασία εικονοστοιχείων από εικόνες (ή άλλων σημάτων), τα πρώιμα στάδια επεξεργασίας χρησιμοποιούν συνελίξεις αντί για πλήρεις συνδέσεις τόσο για χωρικής ανεξαρτησία όσο και για καλύτερη αποτελεσματικότητα.

Η κανονικοποίηση (Regularization and normalization) μπορεί να χρησιμοποιηθεί για να αποτρέψει την υπερπροσαρμογή των νευρωνικών δικτύων, ώστε να μπορούν γενικεύουν καλύτερα σε νέα δεδομένα. Το Dropout είναι μια τεχνική κανονικοποίησης που παρουσίασαν οι Srivastava, Hinton το 2014 όπου σε κάθε mini-batch κατά τη διάρκεια της εκπαίδευσης κάποιο ποσοστό των μονάδων σε κάθε στρώμα μηδενίζονται. Τυχαία ρύθμιση των μονάδων στο μηδέν διοχετεύει θόρυβο στη διαδικασία εκπαίδευσης και επίσης εμποδίζει το δίκτυο να εξειδικεύει υπερβολικά τις μονάδες σε συγκεκριμένα δείγματα ή εργασίες, που και τα δύο μπορούν να βοηθήσουν στη μείωση της υπερπροσαρμογής και στη βελτίωση γενίκευση.

Η τεχνική κανονικοποίησης της κανονικοποίησης παρτίδων (Batch Normalization) βασίζεται στην ιδέα της επανακλιμάκωσης (και της ανανέωσης) των ενεργοποιήσεων σε μια δεδομένη μονάδα έτσι ώστε να έχουν μοναδιαία διακύμανση και μηδενικό μέσο όρο^[16].

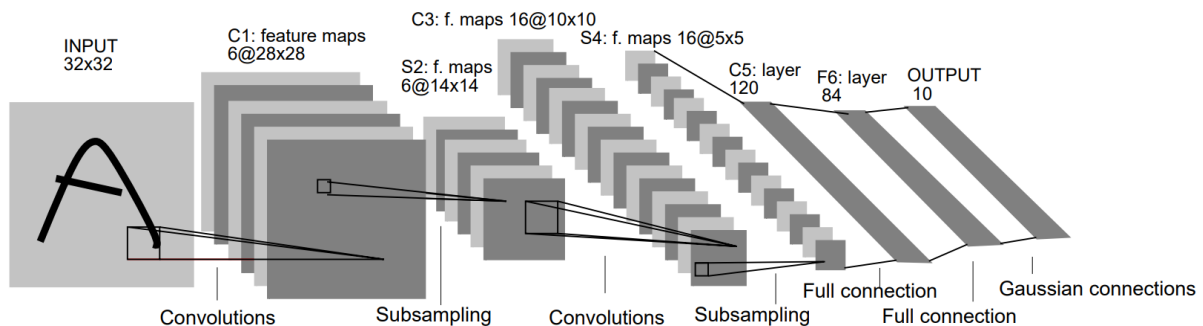
Input: Values of x over a mini-batch: $\mathcal{B} = \{x_{1\dots m}\}$; Parameters to be learned: γ, β
Output: $\{y_i = \text{BN}_{\gamma,\beta}(x_i)\}$
$\mu_{\mathcal{B}} \leftarrow \frac{1}{m} \sum_{i=1}^m x_i \quad // \text{ mini-batch mean}$
$\sigma_{\mathcal{B}}^2 \leftarrow \frac{1}{m} \sum_{i=1}^m (x_i - \mu_{\mathcal{B}})^2 \quad // \text{ mini-batch variance}$
$\hat{x}_i \leftarrow \frac{x_i - \mu_{\mathcal{B}}}{\sqrt{\sigma_{\mathcal{B}}^2 + \epsilon}} \quad // \text{ normalize}$
$y_i \leftarrow \gamma \hat{x}_i + \beta \equiv \text{BN}_{\gamma,\beta}(x_i) \quad // \text{ scale and shift}$

Εικόνα 7 : Αλγόριθμος κανονικοποίησης παρτίδων

Οι παράμετροι προς εκμάθηση είναι οι γ, β . Στην 1^η εξίσωση (mini-batch mean) υπολογίζεται η μέση τιμή του mini-batch, στην 2^η (mini-batch variance) η διακύμανση, στην 3^η (normalize) γίνεται κανονικοποίηση με βάση τις προηγούμενες δυο τιμές και στην 4^η (scale and shift) γίνεται κλιμακοποίηση και μετατόπιση της κατανομής που έχει προκύψει.

1.3.2. Συνελκτικά νευρωνικά δίκτυα (Convolutional neural networks)

Το πιο κρίσιμο συστατικό των deep networks για την επεξεργασία εικόνας και την μηχανική όραση, είναι η χρήση συνελίξεων πολλαπλών επιπέδων που μπορούν να εκπαιδευτούν. Η ιδέα των συνελκτικών νευρωνικών δικτύων έγινε δημοφιλής από τους LeCun, Bottou et al. (1998), όπου εισήγαγαν το δίκτυο LeNet-5 για την αναγνώριση ψηφίων^[17].

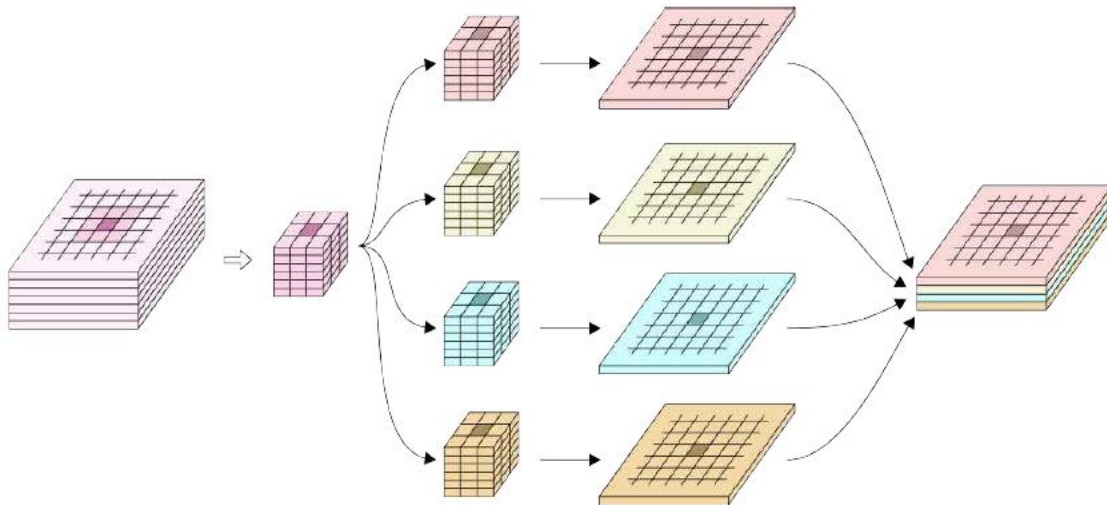


Εικόνα 8 : Αρχιτεκτονική του νευρωνικού δικτύου LeNet-5

Αντί να συνδέουν όλες τις μονάδες ενός επίπεδου με όλες τις μονάδες ενός προηγούμενου επίπεδου, τα συνελκτικά δίκτυα οργανώνουν κάθε επίπεδο σε χάρτες χαρακτηριστικών, τους οποίους μπορείτε να σκεφτείτε ως παράλληλα επίπεδα ή στρώμα καναλιών. Τα σταθμισμένα αθροίσματα εκτελούνται μόνο μέσα σε ένα μικρό τοπικό παράθυρο και τα βάρη είναι πανομοιότυπα για όλα τα εικονοστοιχεία, ακριβώς όπως στην κανονική συνέλιξη και συσχέτιση εικόνας αμετάβλητης μετατόπισης. just as in regular shift-invariant image convolution and correlation.

Σε αντίθεση με τη συνέλιξη εικόνας, ωστόσο, όπου εφαρμόζεται το ίδιο φίλτρο σε κάθε κανάλι χρώματος, οι συνελίξεις νευρωνικών δικτύων συνήθως συνδυάζουν γραμμικά τις ενεργοποιήσεις από καθένα από τα

κανάλια εισόδου C1 σε ένα προηγούμενο επίπεδο και χρησιμοποιούν διαφορετικούς πυρήνες συνέλιξης για καθένα από τα κανάλια εξόδου C2.



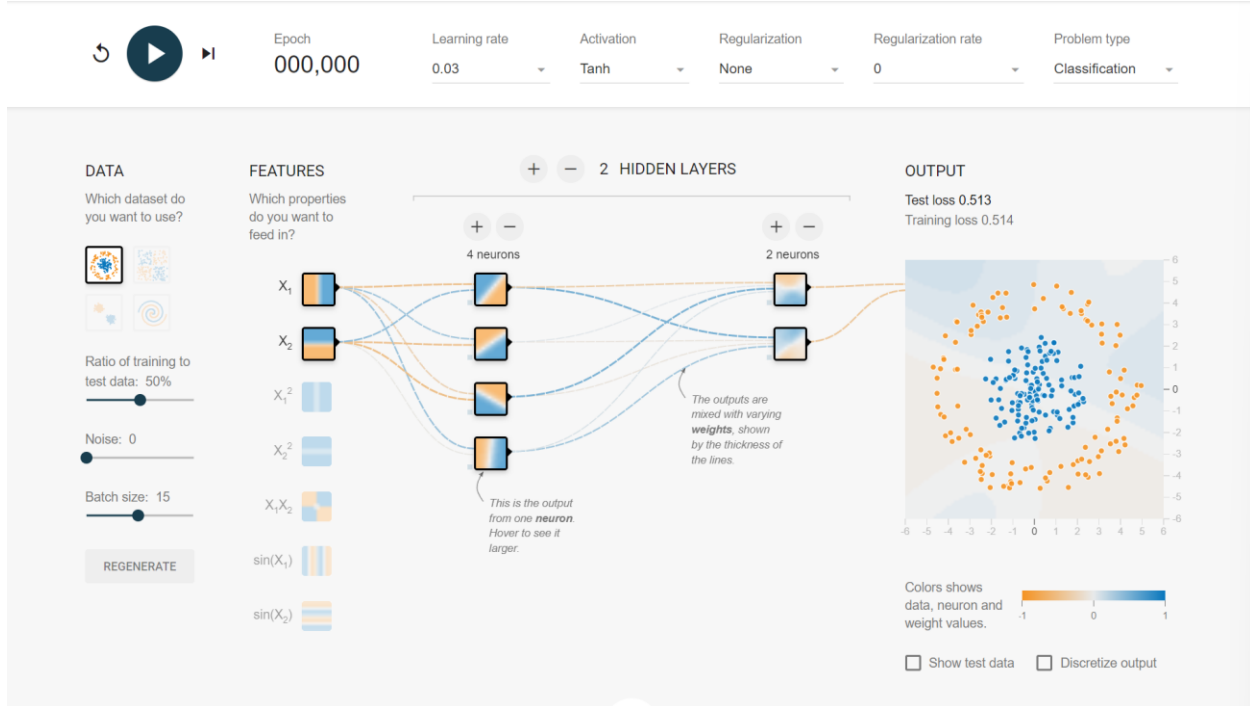
Εικόνα 9 : 2D συνέλιξη με πολλαπλά κανάλια εισόδου και εξόδου

Αυτό είναι λογικό, καθώς το κύριο καθήκον στα επίπεδα των συνελκτικών νευρωνικών δικτύων είναι η κατασκευή τοπικών χαρακτηριστικών και στη συνέχεια ο συνδυασμός τους με διαφορετικούς τρόπους για την παραγωγή πιο διακριτικών και σημασιολογικά σημαντικών χαρακτηριστικών

1.3.3. Εκπαίδευση και βελτιστοποίηση (Training and optimization)

Έστω ότι έχουμε όλα τα στοιχεία που απαιτούνται για την εκπαίδευση ενός νευρωνικού δικτύου. Έχουμε ορίσει την τοπολογία του δικτύου ως προς τα μεγέθη και τα βάθη κάθε επιπέδου, καθορίσαμε τις λειτουργίες ενεργοποίησης, προσθέσαμε όρους τακτοποίησης, καθορίσαμε τη συνάρτηση απώλειας και αρχικοποιήσαμε τα βάρη. Έχουμε περιγράψει ακόμη και τον τρόπο υπολογισμού των διαβαθμίσεων, δηλαδή των παραγών της κανονικοποιημένης απώλειας σε σχέση με όλα τα βάρη μας. Αυτό που χρειαζόμαστε σε αυτό το σημείο είναι κάποιος αλγόριθμος για να μετατρέψουμε αυτές τις κλίσεις σε ενημερώσεις βάρους που θα βελτιστοποιήσουν τη συνάρτηση απώλειας και θα δημιουργήσουν ένα δίκτυο που γενικεύεται καλά σε νέα, αγρησιμοποιήτα δεδομένα.

Το παρακάτω σχήμα δείχνει πως τα βαθιά δίκτυα ενημερώνουν τα βάρη και χαράσσουν ένα χώρο λύσης κατά τη διάρκεια της εκπαίδευσης^[18].



Εικόνα 10 : Εφαρμογή προσομοίωσης νευρωνικών δικτύων

Βάσεις δεδομένων αναγνώρισης εικόνων και σύνολα δοκιμών

Εκτός από την ταχεία πρόοδο στη μηχανική μάθηση και τις τεχνικές στατιστικής μοντελοποίησης, ένα από τα βασικά συστατικά για τη συνεχή βελτίωση των αλγορίθμων αναγνώρισης ήταν η αυξημένη διαθεσιμότητα και ποιότητα των βάσεων δεδομένων αναγνώρισης εικόνων.

Ορισμένες από αυτές τις βάσεις δεδομένων, όπως αυτές για την αναγνώριση προσώπου και την τοπικοποίηση, χρονολογούνται πάνω από δύο δεκαετίες. Οι πιο πρόσφατες και δημοφιλείς, όπως η βάση δεδομένων PASCAL, ανανεώνονται κάθε χρόνο με όλο και πιο απαιτητικά προβλήματα



Εικόνα 11 : Δείγματα εικόνων από τον διαγωνισμό PASCAL Visual Object Classes Challenge 2008

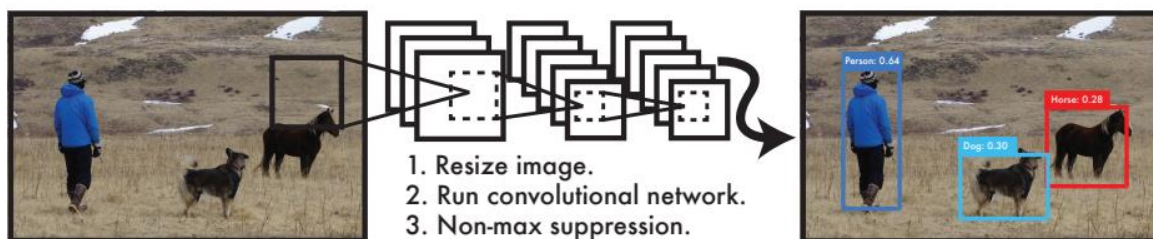
1.4. Yolo

Η κατηγοριοποίηση εικόνων είναι μία από τις πολλές συναρπαστικές εφαρμογές των συνελκτικών νευρωνικών δικτύων. Εκτός από την απλή κατηγοριοποίηση εικόνων, υπάρχουν πολλά συναρπαστικά προβλήματα στην μηχανική όραση, με την ανίχνευση αντικειμένων να είναι ένα από τα πιο ενδιαφέροντα. Το YOLO ("You Only Look Once") είναι ένας αποτελεσματικός αλγόριθμος αναγνώρισης αντικειμένων σε πραγματικό χρόνο, ο οποίος περιγράφηκε για πρώτη φορά στη θεμελιώδη εργασία του 2015 από τους Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi.

Ο YOLO είναι ένας αλγόριθμος που χρησιμοποιεί νευρωνικά δίκτυα για την παροχή ανίχνευσης αντικειμένων σε πραγματικό χρόνο. Αυτός ο αλγόριθμος είναι δημοφιλής λόγω της ταχύτητας και της ακρίβειάς του. Έχει χρησιμοποιηθεί σε διάφορες εφαρμογές για την ανίχνευση σημάτων κυκλοφορίας, ανθρώπων, παρκόμετρων και ζώων.

1.4.1. YOLO v1

Όπως αναφέρθηκε η πρώτη έκδοση του YOLO (YOLO v1) περιεγράφηκε το 2015 από τους Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi. Όπως φαίνεται και στην παρακάτω εικόνα, αυτοί περιγέαφουν τον αλγόριθμο ως ένα ενιαίο συνελκτικό δίκτυο προβλέπει ταυτόχρονα πολλαπλά πλαίσια οριοθέτησης και πιθανότητες κλάσης για αυτά τα πλαίσια^[19].



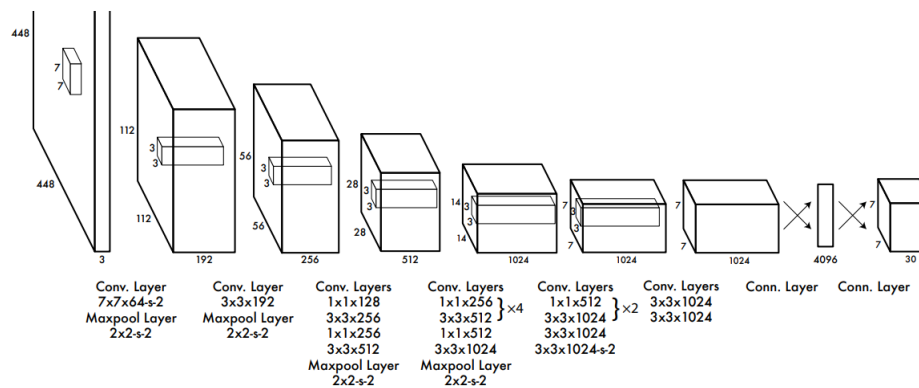
Εικόνα 12 : Η επεξεργασία εικόνων με το YOLO είναι απλή και απλή. Το σύστημά (1) αλλάζει το μέγεθος της εικόνας εισόδου σε 448×448 , (2) εκτελεί ένα ενιαίο συνελκτικό δίκτυο στην εικόνα και (3) ορίζει τα κατώφλια των ανιχνεύσεων που προκύπτουν με βάση την εμπιστοσύνη του μοντέλου

Το YOLO εκπαιδεύεται σε πλήρεις εικόνες και βελτιστοποιεί άμεσα την απόδοση ανίχνευσης. Αυτό το ενοποιημένο μοντέλο έχει πολλά πλεονεκτήματα σε σχέση με τις παραδοσιακές μεθόδους ανίχνευσης αντικειμένων.

1. Το YOLO είναι εξαιρετικά γρήγορο. Δεδομένου ότι πλαισιώνουμε την ανίχνευση ως πρόβλημα παλινδρόμησης, δεν χρειαζόμαστε μια περίπλοκη διοχέτευση. Απλώς τρέχουμε το νευρωνικό μας δίκτυο σε μια νέα εικόνα τη στιγμή της δοκιμής για να προβλέψουμε τις ανιχνεύσεις.

2. Το YOLO λαμβάνει υπόψιν του συνολικά την εικόνα όταν κάνει προβλέψεις. Σε αντίθεση με τις τεχνικές συρόμενου παραθύρου και περιοχής, το YOLO βλέπει ολόκληρη την εικόνα κατά τη διάρκεια της εκπαίδευσης και του χρόνου δοκιμής, επομένως κωδικοποιεί σιωπηρά πληροφορίες σχετικά με τα συμφραζόμενα σχετικά με τις κλάσεις καθώς και την εμφάνισή τους.

3. Το YOLO μαθαίνει γενικεύσιμες αναπαραστάσεις αντικειμένων. Όταν εκπαιδεύεται σε φυσικές εικόνες και δοκιμάζεται σε έργα τέχνης, το YOLO ξεπερνά τις κορυφαίες μεθόδους ανίχνευσης όπως το DPM και το R-CNN με μεγάλη διαφορά. Δεδομένου ότι το YOLO είναι εξαιρετικά γενικεύσιμο, είναι λιγότερο πιθανό να καταρρεύσει όταν εφαρμόζεται σε νέους τομείς ή απρόσμενες εισόδους.



Εικόνα 13 : Αρχιτεκτονική του δικτύου ανίχνευσης του YOLO

Μειονεκτήματα: Το YOLO εξακολουθεί να υστερεί σε σχέση με τα υπερσύγχρονα συστήματα ανίχνευσης στην ακρίβεια. Ενώ μπορεί να αναγνωρίσει γρήγορα αντικείμενα στις εικόνες, δυσκολεύεται να εντοπίσει με ακρίβεια ορισμένα αντικείμενα, ειδικά μικρά.

Περιορισμοί: Το YOLO επιβάλλει ισχυρούς χωρικούς περιορισμούς στις προβλέψεις πλαισίων οριοθέτησης αφού κάθε κελί πλέγματος προβλέπει μόνο δύο πλαίσια και μπορεί να έχει μόνο μία κλάση. Αυτός ο χωρικός περιορισμός περιορίζει τον αριθμό των κοντινών αντικειμένων που μπορεί να προβλέψει

το μοντέλο. Δυσκολεύεται με μικρά αντικείμενα που εμφανίζονται σε ομάδες, όπως κοπάδια πουλιών. Εφόσον μαθαίνει να προβλέπει οριοθέτηση πλαισίων από δεδομένα, δυσκολεύεται να γενικεύσει σε αντικείμενα σε νέους ή ασυνήθιστους λόγους διαστάσεων ή διαμορφώσεις. Επίσης χρησιμοποιεί σχετικά χονδροειδή χαρακτηριστικά για την πρόβλεψη οριοθετημένων πλαισίων, καθώς η αρχιτεκτονική του YOLO έχει πολλαπλά επίπεδα μείωσης δειγματοληψίας από την εικόνα εισόδου.

1.4.2. YOLO v2 (YOLO9000)

Η επόμενη έκδοση του YOLO παρουσιάστηκε το 2016. Οι βελτιώσεις σε σχέση με την πρώτη έκδοση εμφανίζονται στους ακόλουθους τομείς^[20]:

- **Κανονικοποίηση παρτίδας (Batch Normalization).** Η κανονικοποίηση παρτίδας οδηγεί σε σημαντικές βελτιώσεις στη σύγκλιση, ενώ εξαλείφει την ανάγκη για άλλες μορφές τακτοποίησης. Προσθέτοντας κανονικοποίηση παρτίδας σε όλα τα συνελκτικά επίπεδα στο YOLO, έχουμε πάνω από 2% βελτίωση στο mAP (mean Average Precision), δηλαδή στην ακρίβεια των ανιχνευτών αντικειμένων. Η κανονικοποίηση παρτίδας βοηθά επίσης στην τακτοποίηση του μοντέλου. Με την κανονικοποίηση παρτίδας μπορούμε να αφαιρέσουμε το dropout χωρίς υπερβολική προσαρμογή
- **Ταξινομητές υψηλής ανάλυσης (High Resolution Classifier).** Όλες οι μέθοδοι ανίχνευσης τελευταίας τεχνολογίας χρησιμοποιούν ταξινομητή προεκπαιδευμένο στο ImageNet. Το αρχικό YOLO εκπαιδεύει το δίκτυο ταξινομητή στα 224×224 και αυξάνει την ανάλυση σε 448 για ανίχνευση. Για το YOLOv2 πρώτα ρυθμίζουμε το δίκτυο ταξινόμησης στην πλήρη ανάλυση 448×448 για 10 epochs, δηλαδή 10 φορές διαπέραση όλου του dataset) στο ImageNet. Αυτό δίνει στο δίκτυο χρόνο να προσαρμόσει τα φίλτρα του ώστε να λειτουργούν καλύτερα σε είσοδο υψηλότερης ανάλυσης
- **Πρόβλεψη μετατοπίσεων (Convolutional With Anchor Boxes).** Το YOLO προβλέπει τις συντεταγμένες των πλαισίων οριοθέτησης (bounding boxes) απευθείας χρησιμοποιώντας πλήρως συνδεδεμένα στρώματα πάνω από τον συνελκτικό εξαγωγέα χαρακτηριστικών. Η πρόβλεψη μετατοπίσεων αντί για συντεταγμένες απλοποιεί το πρόβλημα και διευκολύνει την εκμάθηση του δικτύου. Αφαιρούμε τα πλήρως συνδεδεμένα επίπεδα από το YOLO και χρησιμοποιούμε πλαίσια αγκύρωσης για να προβλέψουμε τα πλαίσια οριοθέτησης.
- **Dimension Clusters.** Στο Yolo οι διαστάσεις του πλαισίου επιλέγονται με το χέρι. Το δίκτυο μπορεί να μάθει να προσαρμόζει κατάλληλα τα πλαίσια, αλλά αν επιλέξουμε καλύτερες προτεραιότητες για να ξεκινήσει το δίκτυο, μπορούμε να διευκολύνουμε το δίκτυο να μάθει να

προβλέπει καλές ανιχνεύσεις. Αντί να επιλέγουμε προηγούμενα με το χέρι, εκτελούμε την ομαδοποίηση k-means στα πλαίσια οριοθέτησης του σετ εκπαίδευσης για αυτόματη εύρεση καλών priors

- **Direct location prediction.** Όταν χρησιμοποιούμε κουτιά αγκύρωσης με το YOLO, αντιμετωπίζουμε ένα δεύτερο ζήτημα: την αστάθεια του μοντέλου, ειδικά κατά τις πρώτες επαναλήψεις. Το μεγαλύτερο μέρος της αστάθειας προέρχεται από την πρόβλεψη των θέσεων (x, y) για το πλαίσιο. Αυτή η διατύπωση είναι απεριόριστη, επομένως κάθε κουτί αγκύρωσης μπορεί να καταλήξει σε οποιοδήποτε σημείο της εικόνας, ανεξάρτητα από τη θέση που προέβλεψε το πλαίσιο. Με την τυχαία εκκίνηση το μοντέλο χρειάζεται πολύ χρόνο για να σταθεροποιηθεί για να προβλέψει λογικές μετατοπίσεις. Αντί να προβλέψουμε μετατοπίσεις, ακολουθούμε την προσέγγιση του YOLO και προβλέπουμε συντεταγμένες τοποθεσίας σε σχέση με τη θέση του κελιού πλέγματος
- **Fine-Grained Features.** Αυτό το τροποποιημένο YOLO προβλέπει ανιχνεύσεις σε έναν χάρτη χαρακτηριστικών 13x13. Αν και αυτό είναι αρκετό για μεγάλα αντικείμενα, μπορεί να επωφεληθεί από τα πιο λεπτά χαρακτηριστικά για τον εντοπισμό μικρότερων αντικειμένων. Τα πιο γρήγορα R-CNN και SSD εκτελούν τα δίκτυα προτάσεων τους σε διάφορους χάρτες χαρακτηριστικών του δικτύου για να λάβουν μια σειρά αναλύσεων. Ακολουθούμε μια διαφορετική προσέγγιση, προσθέτοντας απλώς ένα επίπεδο διέλευσης που φέρνει χαρακτηριστικά από ένα προηγούμενο επίπεδο σε ανάλυση 26x26.

	YOLO									YOLOv2
batch norm?		✓	✓	✓	✓	✓	✓	✓	✓	✓
hi-res classifier?			✓	✓	✓	✓	✓	✓	✓	✓
convolutional?				✓	✓	✓	✓	✓	✓	✓
anchor boxes?				✓	✓					✓
new network?					✓	✓	✓	✓		✓
dimension priors?						✓	✓	✓		✓
location prediction?						✓	✓	✓		✓
passthrough?							✓	✓		✓
multi-scale?								✓		✓
hi-res detector?										✓
VOC2007 mAP	63.4	65.8	69.5	69.2	69.6	74.4	75.4	76.8		78.6

Εικόνα 14 : Συγκριση απόδοσης του YOLO v1 , της εφαρμογής των επιμέρους βελτιστοποιήσεων και του YOLOv2 με κριτήριο το metric mAP (Mean Average Precision)

Επίσης, έγινε αλλαγή από το YOLO v1 που χρησιμοποιεί ένα προσαρμοσμένο δίκτυο που βασίζεται στην αρχιτεκτονική του Googlenet, σε ένα νέο μοντέλο ταξινόμησης που ονομάζεται Darknet-19, το οποίο

Μηχανική όραση για ανίχνευση ανωμαλιών σε οδικές υποδομές με χρήση νευρωνικών δικτύων

χρησιμοποιεί το στρώμα καθολικής μέσης χωρικής συγκέντρωσης για να κάνει προβλέψεις καθώς και φίλτρα 1 1 για τη συμπίεση της αναπαράστασης χαρακτηριστικών μεταξύ 3 x 3 συνελκτικές στρώσεις. Επίσης έχει 19 συνελκτικά στρώματα και 5 στρώματα maxpooling.

1.4.3. YOLO v3

Η αρχιτεκτονική του YOLO v2 εξακολουθούσε να στερείται μερικά από τα πιο σημαντικά στοιχεία που αποτελούν πλέον βασικό στοιχείο στους περισσότερους αλγόριθμους τελευταίας τεχνολογίας^[21]. Ήταν χωρίς υπολειμματικά μπλοκ, χωρίς συνδέσεις παράλειψης και χωρίς δειγματοληψία.

Η έκδοση v3 παρουσιάστηκε το 2018 έχει βελτιώσεις σε σχέση με την v2 στα ακόλουθα:

- **Bounding Box Prediction.**

YOLOv3 προβλέπει μια βαθμολογία αντικειμενικότητας για κάθε πλαίσιο οριοθέτησης χρησιμοποιώντας λογιστική παλινδρόμηση

- **Class Prediction**

Για την πρόβλεψη, χρησιμοποιούνται ανεξάρτητοι λογιστικοί ταξινομητές. Για την εκπαίδευση, χρησιμοποιείται δυαδική διασταυρούμενη εντροπία για τις προβλέψεις των κλάσεων.

- **Predictions Across Scales**

YOLOv3 predicts προβλέπει πλαίσια σε 3 διαφορετικές κλίμακες

- **Feature Extractor**

Χρησιμοποιείται ένα νέο δίκτυο για την εκτέλεση εξαγωγής χαρακτηριστικών που έχει 53 συνελκτικά επίπεδα και ονομάζεται Darknet-53. Το Darknet-53 είναι ένα συνελκτικό νευρωνικό δίκτυο που λειτουργεί ως ραχοκοκαλιά για την προσέγγιση ανίχνευσης αντικειμένων YOLOv3

	Type	Filters	Size	Output
	Convolutional	32	3 × 3	256 × 256
	Convolutional	64	3 × 3 / 2	128 × 128
1x	Convolutional	32	1 × 1	
	Convolutional	64	3 × 3	
	Residual			128 × 128
	Convolutional	128	3 × 3 / 2	64 × 64
2x	Convolutional	64	1 × 1	
	Convolutional	128	3 × 3	
	Residual			64 × 64
	Convolutional	256	3 × 3 / 2	32 × 32
8x	Convolutional	128	1 × 1	
	Convolutional	256	3 × 3	
	Residual			32 × 32
	Convolutional	512	3 × 3 / 2	16 × 16
8x	Convolutional	256	1 × 1	
	Convolutional	512	3 × 3	
	Residual			16 × 16
	Convolutional	1024	3 × 3 / 2	8 × 8
4x	Convolutional	512	1 × 1	
	Convolutional	1024	3 × 3	
	Residual			8 × 8
	Avgpool		Global	
	Connected		1000	
	Softmax			

Εικόνα 15 : Δομή του Darknet-53

1.5. Οδοστρώματα

Ως οδοστρώματα ορίζουμε την η επιφάνεια του δρόμου που στρώνεται με άσφαλτο ή τσιμέντο και πάνω της κινούνται τα οχήματα

1.5.1. Φθορές Οδοστρωμάτων

Γενικά, κάθε φθορά του οδοστρώματος κατατάσσεται και εκτιμάται ανάλογα με τον τύπο, τη σοβαρότητα και την έκταση στην οποία παρατηρείται. Στις περισσότερες περιπτώσεις οι φθορές ταξινομούνται σε τρία επίπεδα ανάλογα με τον βαθμό σοβαρότητας που εμφανίζουν ενώ η έκταση τους μετριέται ή εκτιμάται ως ποσοστό του μήκους που εμφανίζεται το πρόβλημα επί του συνολικού μήκους του απογραφόμενου τμήματος.

Σοβαρότητα (severity) : με τον όρο σοβαρότητα φθοράς νοείται η εκτίμηση της σφοδρότητας της φθοράς ή ο βαθμός επικινδυνότητας που ενέχει. Οι φθορές ανάλογα με τη σφοδρότητα που εμφανίζουν κατατάσσονται σε χαμηλού, μεσαίου και υψηλού βαθμού σοβαρότητας.

Έκταση (extent) : με τον όρο έκταση φθοράς νοείται η ποσοτική εκτίμηση του μεγέθους της φθοράς. Ανάλογα με τον τύπο της φθοράς η έκταση μπορεί να μετριέται σε μονάδες μήκους, εμβαδόν κάλυψης ή πλήθος εμφάνισης φθορών συγκεκριμένου τύπου.

1.5.1.1. Τύποι Φθορών

1.5.1.1.1. Ρηγματώσεις (Cracking).

Ρηγμάτωση είναι η διακοπή της συνέχειας σε μια επιφάνεια. Οι ρηγματώσεις στα ασφαλτικά οδοστρώματα παρουσιάζουν μια μεγάλη ποικιλία μορφών, αφού ποικίλα είναι και τα αίτια που τις προκαλούν.

Ρηγμάτωσης τύπου αλιγάτορα (alligator cracks)

Οι ρωγμές τύπου αλιγάτορα είναι διακλαδιζόμενες και αλληλοσυνδεόμενες ρωγμές που σχηματίζουν πολυγωνικά κομμάτια (μπλοκ) όμοια με αυτά του δέρματος του αλιγάτορα^[22]. Σε ορισμένες περιπτώσεις τα κομμάτια αυτά δίνουν τη εντύπωση ότι είναι σχεδόν έτοιμα να αποκολληθούν. Η φθορά αρχικά εμφανίζεται σαν



Εικόνα 16 : Ρηγμάτωσης τύπου αλιγάτορα

διαμήκης ρωγμή στην τροχιά των τροχών και επιπλέον καταπόνηση με την πάροδο του χρόνου οδηγεί στη προαναφερθείσα μορφή.

Επίπεδα σοβαρότητας:

Χαμηλή: Διακλαδιζόμενες, διαμήκεις, ασυνεχείς και περιορισμένου εύρους ρωγμές που αρχίζουν να συνδέονται μεταξύ τους και να σχηματίζουν το μοτίβο ενός αλιγάτορα χωρίς όμως να παρατηρείται αποφλοίωση του οδοστρώματος.



Εικόνα 17 : Χαμηλό επίπεδο φθοράς σε ρηγμάτωσεις τύπου αλιγάτορα

Μέση: Οι ρωγμές σχηματίζουν ένα πλέγμα το οποίο ακολουθεί ξεκάθαρα το μοτίβο του αλιγάτορα. Στις παρυφές της ρηγματωμένης περιοχής το οδόστρωμα πιθανόν να παρουσιάζει σημεία αποφλοίωσης. Το εύρος των ρωγμών μπορεί να είναι μεγαλύτερο από 0,6 cm χωρίς όμως ακόμη να αποκολλώνται κομμάτια από το οδόστρωμα.



Εικόνα 18 : Μέσο επίπεδο φθοράς σε ρηγμάτωσεις τύπου αλιγάτορα

Υψηλή: Το οδόστρωμα παρουσιάζει πλέον ένα καλοσχηματισμένο μοτίβο ρωγμών με ιδιαίτερα εμφανή σημεία αποφλοίωσης. Κομμάτια του οδοστρώματος έχουν πλήρως αποκολληθεί και ενδεχομένως φύγει εντελώς από τη θέση τους. Λεπτόκοκκο υλικό από το υπόστρωμα βγαίνει στην επιφάνεια.



Εικόνα 19 : Υψηλό επίπεδο φθοράς σε ρηγμάτωσεις τύπου αλιγάτορα

1.5.1.1.2. Ρηγμάτωση στα άκρα του οδοστρώματος (Edge cracking)

Οι ρωγμές αυτές είναι συνήθως επιμήκεις και παρατηρούνται περίπου 30-50 cm από τα άκρα του οδοστρώματος με ή χωρίς εγκάρσιες διακλαδώσεις προς το έρεισμα. Ανάλογα με την σοβαρότητα της φθοράς, μπορεί να μοιάζουν στην μορφή με τις ρωγμές τύπου αλιγάτορα, διαφοροποιούνται όμως ως προς τα αίτια, που στην προκειμένη περίπτωση είναι κυρίως η έλλειψη πλευρικής υποστήριξης του οδοστρώματος και η υποχώρηση των υποκείμενων στρώσεων λόγω ξηρασίας εδάφους, δράσης παγετού, ανεπαρκούς αποστράγγισης, κακής συμπύκνωσης, συρρίκνωσης λόγω υγρασίας του εδάφους της περιοχής ή μειωμένου πάχους των στρώσεων στα σημεία αυτά.



Εικόνα 20 : Ρηγμάτωση στα άκρα του οδοστρώματος

Η έκταση της φθοράς στο κάθε τμήμα του οδοστρώματος χαρακτηρίζεται ως:

Μικρή: Όταν το 1/2 ή περισσότερο του τμήματος παρουσιάζει μικρές ρωγμές ή όταν το 1/3 ή περισσότερο του τμήματος παρουσιάζει μικρές ή μεσαίες ρωγμές.

Μέση: Όταν τουλάχιστον το 1/2 του τμήματος παρουσιάζει μεσαίες ρωγμές ή όταν τουλάχιστον το 1/3 του τμήματος παρουσιάζει μικρές, μεσαίες και μεγάλες ρωγμές.

Μεγάλη: Όταν τουλάχιστον το 1/3 του τμήματος παρουσιάζει μεγάλες ρωγμές.

Τα επίπεδα σοβαρότητας φθοράς ταξινομούνται ως:

Μικρές: Φθορές (αποκόλληση αδρανούς υλικού) παρατηρείται κατά μήκος των ακμών του οδοστρώματος.

Μεσαίες: Οι φθορές κατά μήκος των ακμών του οδοστρώματος είναι τέτοιες που απαιτούνται τοπικές επισκευές για την αποκατάστασή τους.

Μεγάλες: Οι φθορές κατά μήκος των ακμών του οδοστρώματος είναι τέτοιες που η απόσταση από τον άξονα της οδού μέχρι την ακμή του οδοστρώματος είναι μικρότερη από 3 m.

1.5.1.1.3. Λακκούβες (Potholes).

Οι λακκούβες είναι οπές διάφορων μεγεθών και μορφής λεκανών μέσα στο οδόστρωμα. Δημιουργούνται από αδυναμίες του ασφαλτομίγματος ή της δομής του οδοστρώματος όπως έλλειψη συνδετικού υλικού στο ασφαλτόμιγμα, μειωμένο πάχος ασφαλτικής στρώσης, τοπική αστοχία των στρώσεων – κυρίως της βάσης - κατά την κατασκευή και ανεπαρκή αποστράγγιση της οδού. Επίσης, μια βλάβη του δικτύου υδροδότησης ή λιμνάζοντα νερά είναι δυνατόν να προκαλέσουν την υποχώρηση της ασφάλτου. Επιρρεπές σε μετατοπίσεις είναι κυρίως το αργιλώδες έδαφος που εντοπίζεται πολύ συχνά κοντά σε χωράφια.



Εικόνα 21 : Λακούβα (Pothole)

Η έντονη βροχόπτωση δεν επηρεάζει την κατάσταση του οδοστρώματος. Όμως, αν η ασφαλτος έχει ρωγμές και επιβαρύνεται από βαριά οχήματα, το νερό μπορεί να επιφέρει αλλαγές στο μέγεθος και το βάθος της λακκούβας.

Επίπεδα σοβαρότητας:

Χαμηλή: άνοιγμα λακκούβας μικρότερο των 8 cm και βάθος όχι μεγαλύτερο από 5 cm.

Μέση: άνοιγμα λακκούβας 8-30 cm και βάθος από 5 cm μέχρι 10 cm.

Υψηλή: άνοιγμα λακκούβας μεγαλύτερο από 30 cm και βάθος μεγαλύτερο 10 cm.

1.5.1.1.4. Φθορές οριζόντιας σήμανσης (Markings wear).

Η οριζόντια σήμανση των οδών είναι το τμήμα της σήμανσης που αποτελείται από ενδείξεις επάνω στο οδόστρωμα. και ο σκοπός της είναι είτε από μόνη της, είτε σε συνδυασμό με άλλες διατάξεις ελέγχου της κυκλοφορίας, να καθοδηγήσει τον οδηγό, να ρυθμίσει την κυκλοφορία, να καθορίσει ειδικές περιοχές του οδοστρώματος και να καταστήσει εμφανή ορισμένα επικίνδυνα σημεία της οδού. Επειδή η οριζόντια σήμανση εφαρμόζεται στην επιφάνεια του οδοστρώματος επηρεάζονται άμεσα από τις προηγούμενα αναφερόμενες φθορές που δημιουργούνται στην κοινή τους επιφάνεια και επιπρόσθετα ξεθωριάζουν λόγω χρήσης με την τριβή των ελαστικών αλλά και λόγω περιβαλλοντικών συνθηκών (υγρασία, ηλιακή ακτινοβολία κλπ.)



Εικόνα 22 : Παράδειγμα φθοράς οριζόντιας σήμανσης

1.6. ADAS (Advanced Driving Assistance Systems)

Με τον όρο Προηγμένα συστήματα υποβοήθησης οδηγού (Advanced Driving Assistance Systems - ADAS) περιγράφουμε ένα σύνολο συστημάτων που είναι απαραίτητα για να παρακολουθούν διάφορες παραμέτρους που σχετίζονται με το όχημα και το περιβάλλον του οχήματος για τον εντοπισμό δυνητικά επικίνδυνων καταστάσεων σε ένα πρώιμο στάδιο.

Αναφορικά με την ταξινόμηση της Εταιρείας Διεθνών Μηχανικών Αυτοκινήτων (SAE), το ADAS φτάνει στα εμπορικά οχήματα μόνο μέχρι το επίπεδο 2, το οποίο αναλαμβάνει καθήκοντα όπως η ανίχνευση του περιβάλλοντος οχημάτων, η προειδοποίηση των οδηγών για έκτακτη ανάγκη και η εκτέλεση μιας ή περισσότερων απλών λειτουργιών ελέγχου, όπως για παράδειγμα έλεγχος ταχύτητας και προσαρμογής, φρένο έκτακτης ανάγκης εκτέλεση και ούτω



SAE J3016™ LEVELS OF DRIVING AUTOMATION™

Learn more here: [sae.org/standards/content/j3016_202104](https://www.sae.org/standards/content/j3016_202104)

Copyright © 2021 SAE International. The summary table may be freely copied and distributed AS-IS provided that SAE International is acknowledged as the source of the content.

	SAE LEVEL 0™	SAE LEVEL 1™	SAE LEVEL 2™	SAE LEVEL 3™	SAE LEVEL 4™	SAE LEVEL 5™
What does the human in the driver's seat have to do?	You are driving whenever these driver support features are engaged – even if your feet are off the pedals and you are not steering			You are not driving when these automated driving features are engaged – even if you are seated in "the driver's seat"		
	You must constantly supervise these support features; you must steer, brake or accelerate as needed to maintain safety			When the feature requests, you must drive	These automated driving features will not require you to take over driving	

Copyright © 2021 SAE International.

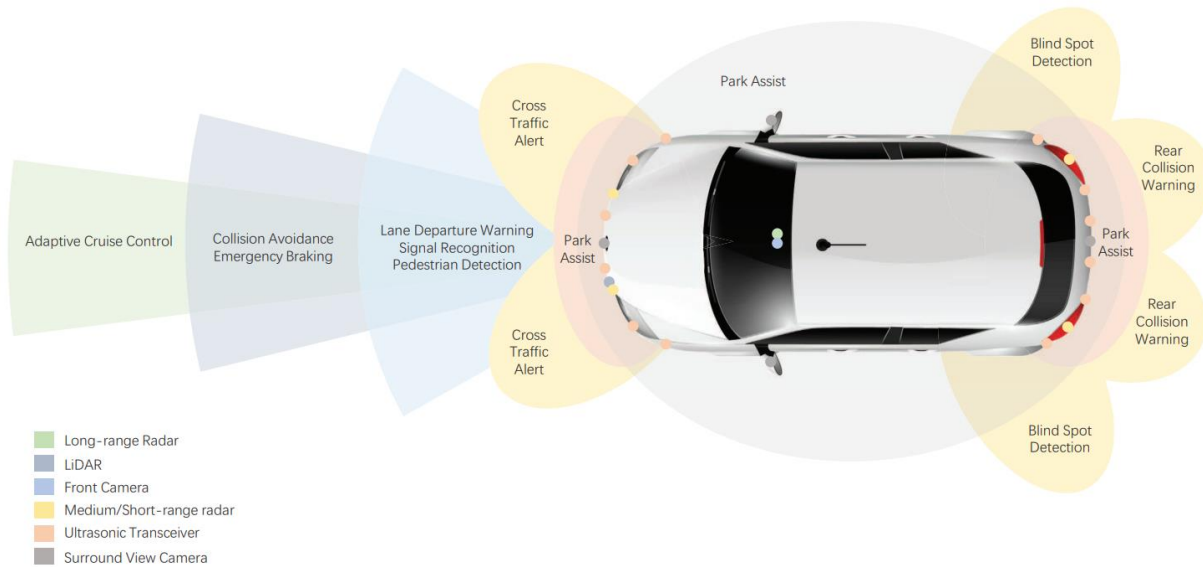
	These are driver support features			These are automated driving features		
What do these features do?	These features are limited to providing warnings and momentary assistance	These features provide steering OR brake/acceleration support to the driver	These features provide steering AND brake/acceleration support to the driver	These features can drive the vehicle under limited conditions and will not operate unless all required conditions are met	This feature can drive the vehicle under all conditions	
Example Features	<ul style="list-style-type: none"> • automatic emergency braking • blind spot warning • lane departure warning 	<ul style="list-style-type: none"> • lane centering OR • adaptive cruise control 	<ul style="list-style-type: none"> • lane centering AND • adaptive cruise control at the same time 	<ul style="list-style-type: none"> • traffic jam chauffeur 	<ul style="list-style-type: none"> • local driverless taxi • pedals/steering wheel may or may not be installed 	<ul style="list-style-type: none"> • same as level 4, but feature can drive everywhere in all conditions

Εικόνα 23 : Επίπεδα αυτοματοποιημένης οδήγησης σύμφωνα με το ADAS

καθεξής^[23].

Το ADAS εισήχθη για πρώτη φορά στην αγορά εμπορικών οχημάτων με το χαρακτηριστικό ABS (Antilock Braking System) στα τέλη της δεκαετίας του 1980^[24]. Αυτού του είδους οι ADAS λειτουργίες παρέχονται μόνο ως πολυτελή χαρακτηριστικά για πολυτελείς μάρκες τον 20ό αιώνα και στις αρχές του 2000. Ωστόσο, τα συστήματα ADAS αναπτύχθηκε γρήγορα από το 2000, χάρη στους σχετικούς διαγωνισμούς και προκλήσεις αυτόνομης οδήγησης. Είναι πλέον ένα πολύπλοκα και καλά ανεπτυγμένα συστήματα αποτελούμενα τόσο από αρχιτεκτονική υλικού για την αντίληψη όσο και σχεδιασμό λογισμικού για μεταεπεξεργασία. Βελτιώνουν την οδηγική άνεση βοηθώντας ή αναδιαμορφώνοντας τις οδηγικές συμπεριφορές ώστε να αποφεύγονται τα τροχαία ατυχήματα για να διασφαλιστεί η οδηγική ασφάλεια.

Οι συμβατικοί αισθητήρες αντίληψης του ADAS, όπως το LiDAR (αισθητήρας ανίχνευσης φωτός και εμβέλεια), κάμερα, ραντάρ, πομποδέκτης υπερήχων και κάποια άλλοι αισθητήρες εντοπισμού φαίνονται στην παρακάτω εικόνα



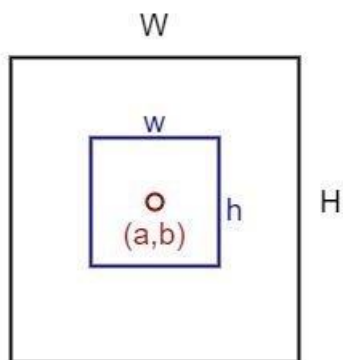
Εικόνα 24 : Είδη και θέση αισθητήρων ADAS σε όχημα

Για την ανίχνευση ανωμαλιών οδοστρώματος με μηχανική όρασης ενδιαφερόμαστε κυρίως για τις κάμερες μπροστινής θέασης και ενδεχομένως της περιφερειακής θέασης. Οι κάμερες είναι από τους πρωτοπόρους αισθητήρες στα συστήματα ADAS λόγω της παροχής τους ευρείας αντίληψης για το περιβάλλον και της αποδεκτής τιμής τους. Επίσης όταν χρησιμοποιούνται σε ζεύγη παρέχουν μέσω της στερεοσκοπικής της εικόνας δυνατότητα μέτρησης αποστάσεων.

Το μέγεθος της παρτίδας διαιρείται σε υποδιαιρέσεις (subdivisions) κατά τη φόρτωση του δικτύου, δηλαδή πόσες εικόνες μπορούν να χωρέσουν στη GPU που εκτελεί τον αλγόριθμο ταυτόχρονα. Με τις υποδιαιρέσεις προκύπτει το πλήθος των mini-batches (Batch/subdivisions = mini-batches). Μια επανάληψη (iteration) αποτελεί τον αριθμό των batches που χρειαζόμαστε για να ολοκληρωθεί ένα Epoch. Ένα Epoch έχουμε όταν κάνουμε μια πλήρη διαπέραση του training dataset.

2.1.2. Δομή αρχείου Yolo

Το Yolo Text file format είναι ένας τύπος αρχείου κειμένου, που χρησιμοποιείται από τον αλγόριθμο του YOLO. Αυτός ο τύπος αρχείου καθορίζει σε κάθε γραμμή του αρχείου την απόλυτη θέση του αντικειμένου στην εικόνα και τον αριθμό της κλάση στην οποία ανήκει, και αυτή την περιγραφή θα χρησιμοποιήσουμε για να κάνουμε την εκπαίδευση στο σύνολο δεδομένων εκπαίδευσης (training dataset). Ένα παράδειγμα μιας τέτοιας εγγραφής “2 0.553419 0.653416 0.448718 0.693168”



W = Image Width
H = Image Height
w = Absolute Width
h = Absolute Height
(a,b) = Absolute x,y coordinates

YOLO FORMAT = <Class Number><a/W><b/H><w/W><h/H>

Εικόνα 26 : Δομή εγγραφών (γραμμών) του Yolo Text file format

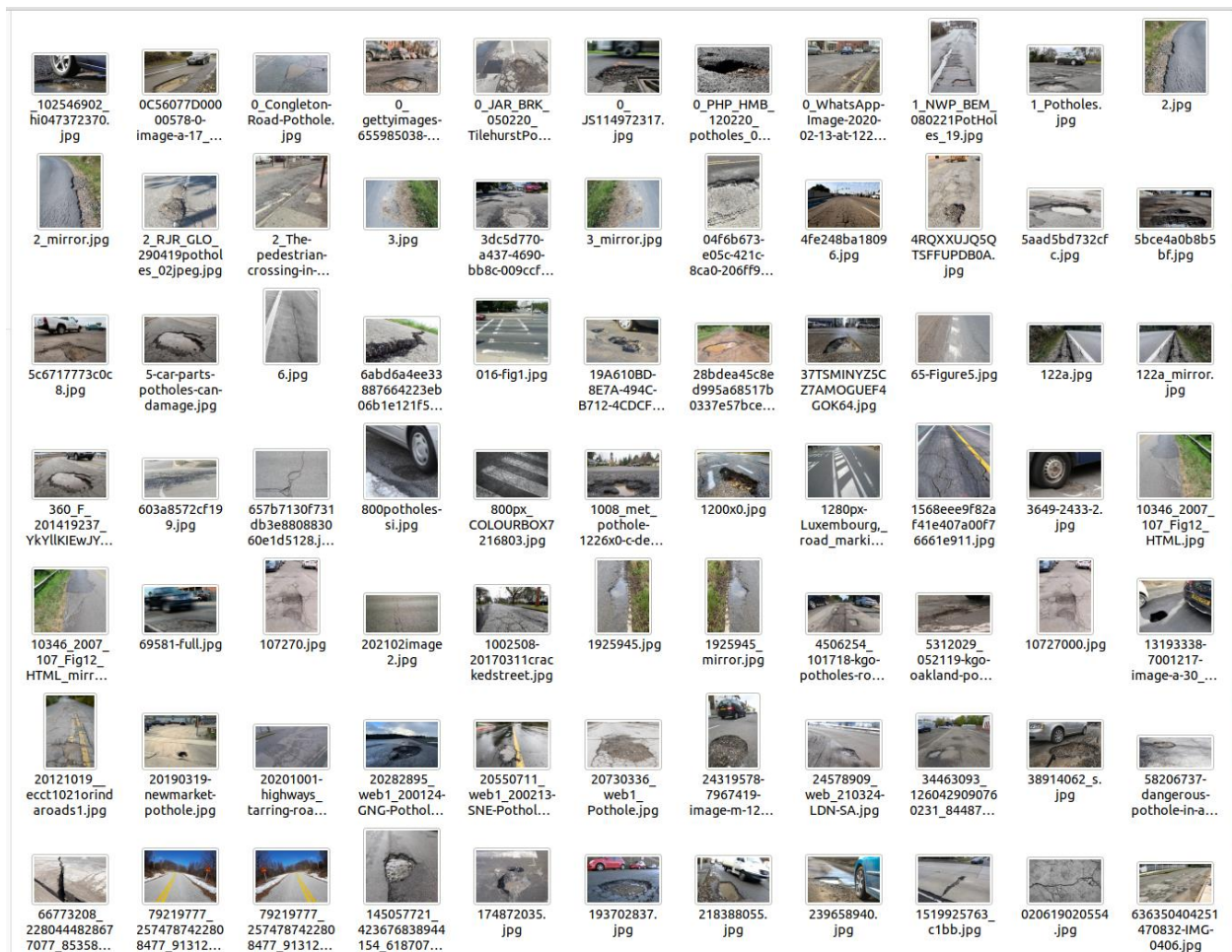
2.1.3. Dataset

Μηχανική όραση για ανίχνευση ανωμαλιών σε οδικές υποδομές με χρήση νευρωνικών δικτύων

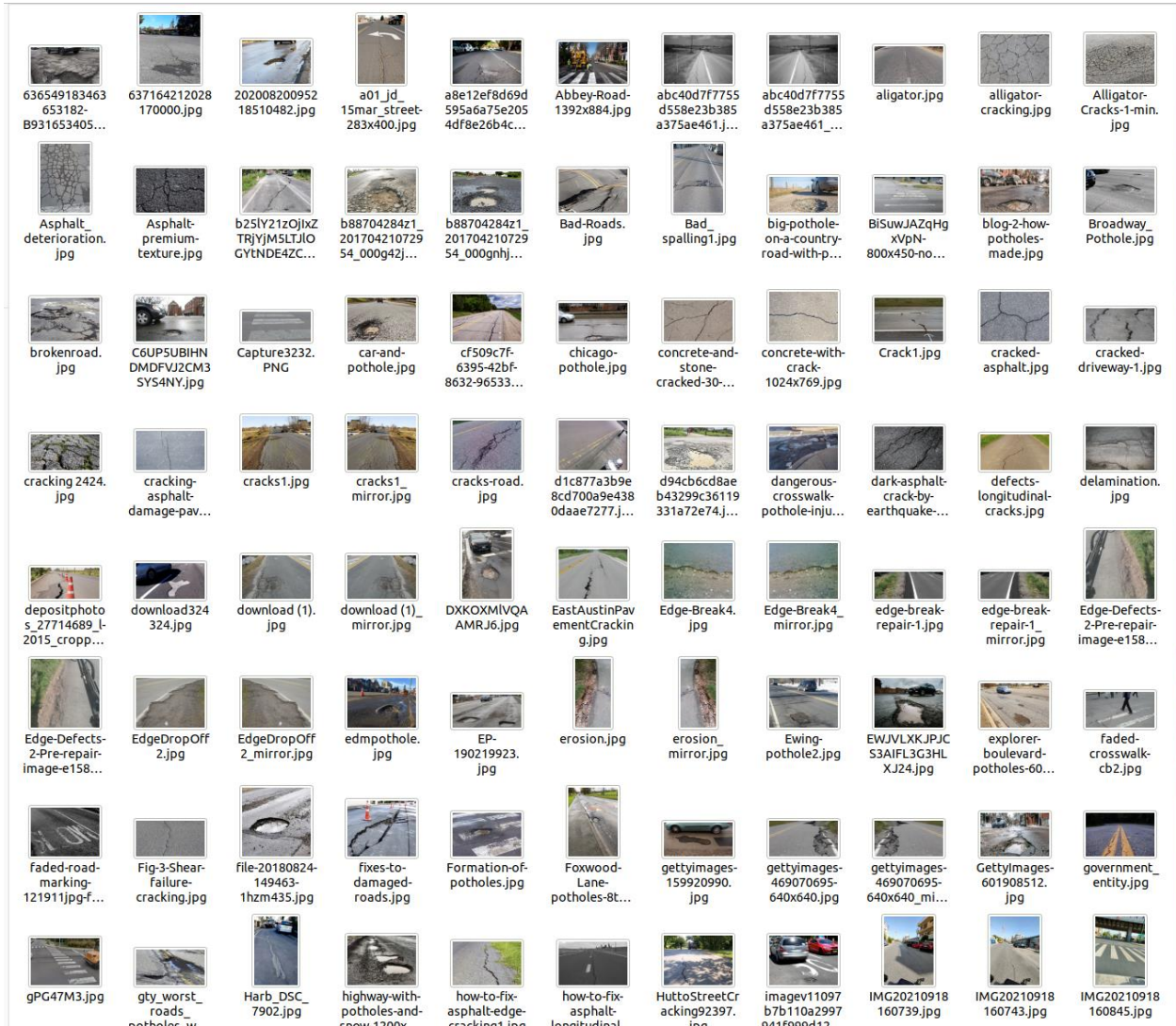
Το dataset περιείχε ένα σύνολο από εικόνες που περιείχαν 4 κατηγορίες φθορών οδοστρώματος:

1. Ρωγμές (cracking)
2. αστοχίες στην άκρη(edge failure)
3. φθορά σημάτων (markings wear)
4. λακκούβες (potholes)

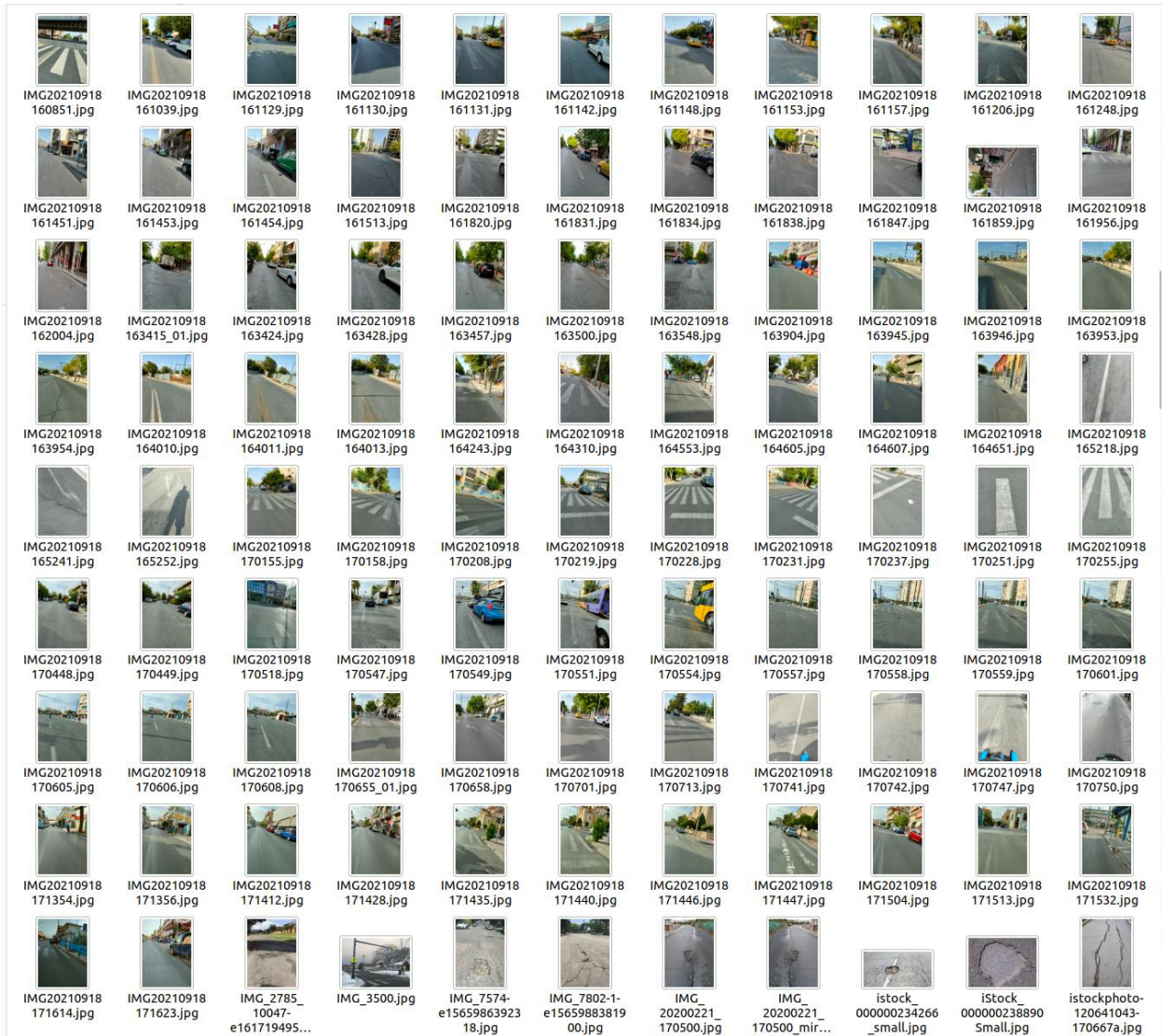
Οι εικόνες αυτές είναι οι ακόλουθες :



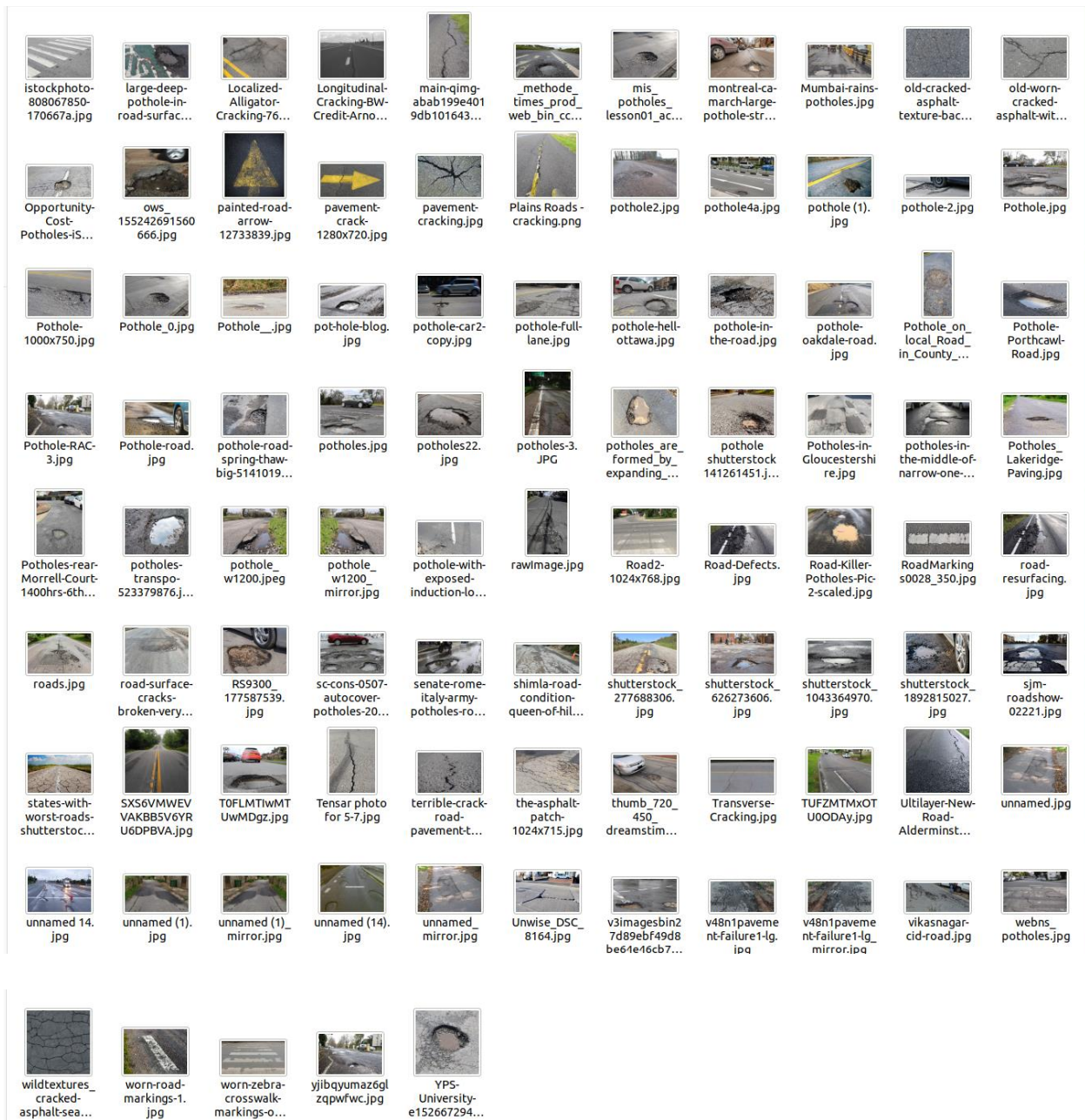
Μηχανική όραση για ανίχνευση ανωμαλιών σε οδικές υποδομές με χρήση νευρωνικών δικτύων



Μηχανική όραση για ανίχνευση ανωμαλιών σε οδικές υποδομές με χρήση νευρωνικών δικτύων




Μηχανική όραση για ανίχνευση ανωμαλιών σε οδικές υποδομές με χρήση νευρωνικών δικτύων



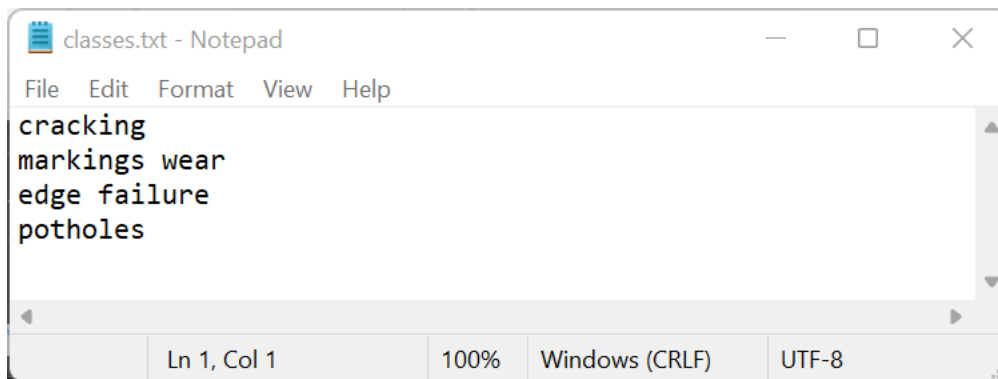
Εικόνα 27 : Σύνολο εικόνων dataset

2.1.4. Προετοιμασία Dataset

Για την προετοιμασία του dataset ώστε να το επεξεργαστούμε με το YOLO, χρειάζεται να κάνουμε μια προεργασία με το δείγμα εκπαίδευσης μας. Αυτό είναι το labeling, δηλαδή η επισήμανση σε κάθε εικόνα περιοχών όπου εμφανίζονται αντικείμενα κάποιας κλάσης. Για την εργασία αυτή χρησιμοποίησα το εργαλείο LabelImg ([GitHub - tzutalin/labelImg](https://github.com/tzutalin/labelImg):  LabelImg is a graphical image annotation tool and label object bounding boxes in images)

Το εργαλείο αυτό αποθηκεύει αυτά τα μεταδεδομένα για κάθε εικόνα ως αρχείο με δομή περιγραφής Yolo.

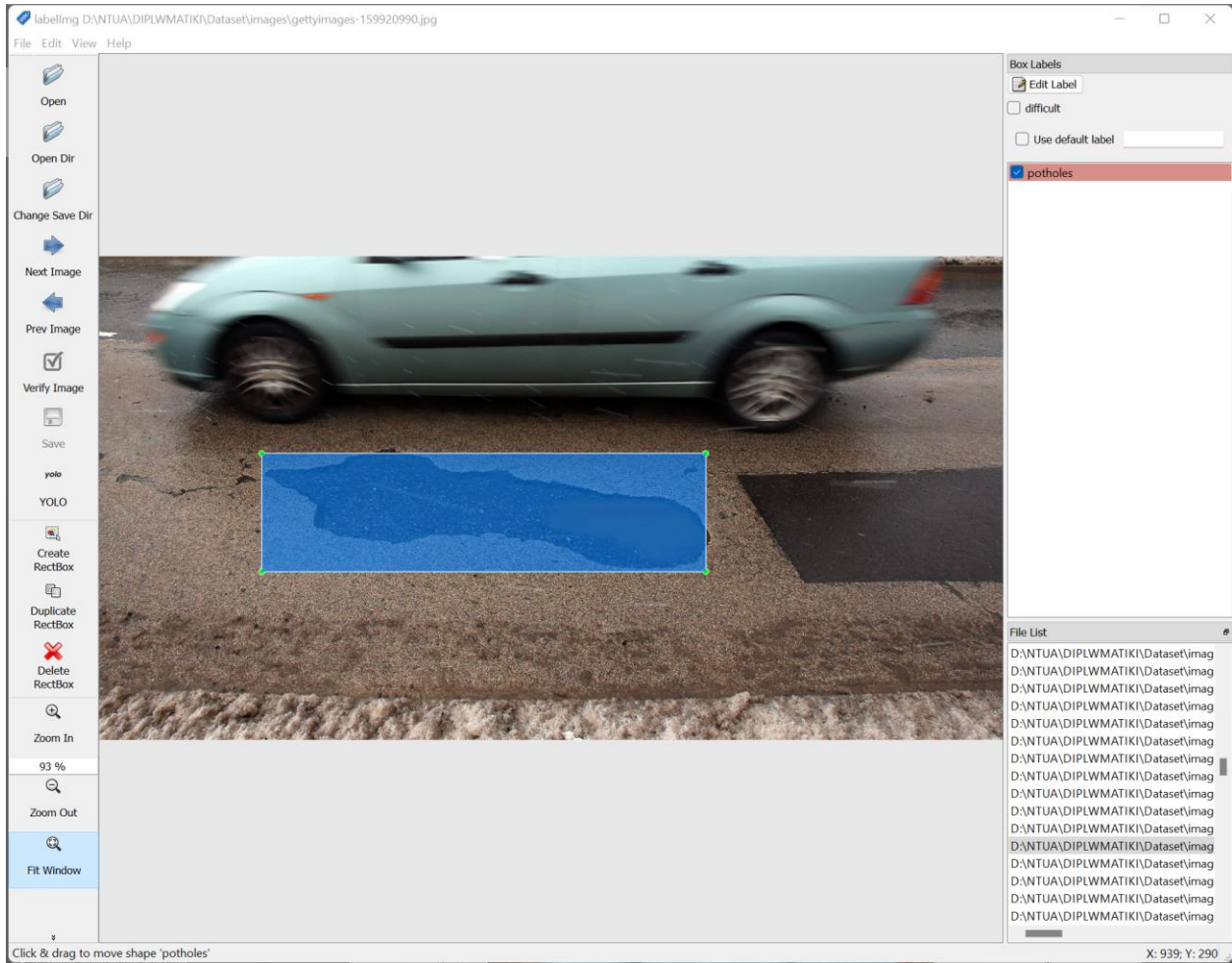
Για το dataset έχουμε δημιουργήσει ένα αρχείο κλάσεων που περιγράφει λεκτικά τις 4 κλάσεις μας. Σε αυτό κάθε γραμμή περιγράφει κάθε κλάση.



```
classes.txt - Notepad
File Edit Format View Help
cracking
markings wear
edge failure
potholes
Ln 1, Col 1 100% Windows (CRLF) UTF-8
```

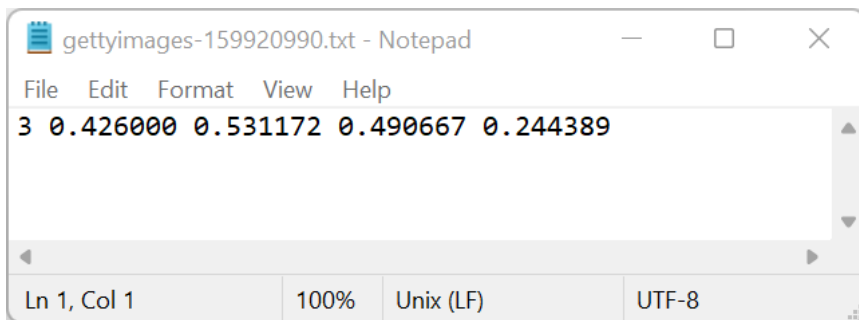
Εικόνα 28 : Αρχείο περιγραφής των κλάσεων στο YOLO

Στην παρακάτω εικόνα φαίνεται το εργαλείο labeling και μια εικόνα από το dataset που έχουμε ανοίξει σε αυτό. Παρατηρούμε την οριοθέτηση ενός τμήματος της εικόνας με ένα ορθογώνιο στο οποίο έχουμε αντιστοιχήσει τη κλάση potholes



Εικόνα 29 : Μαρκάρισμα περιοχής που ανήκει στην κλάση potholes με το εργαλείο LabelImg

Αφού τελειώσουμε με το μαρκάρισμα των περιοχών της εικόνας, δημιουργείται το αντίστοιχο YOLO text file format αρχείο όπως φαίνεται παρακάτω, για την παραπάνω εικόνα. Εμφανίζεται μια γραμμή-εγγραφή για το ορθογώνιο που επέλεξα με πρώτο τον αριθμό που αντιστοιχεί στην γραμμή που περιγράφει την κλάση potholes στο αρχείο κλάσεων.



Εικόνα 30 : Αρχείο τύπου Yolo Text file format με τα στοιχεία των μαρκαρισμένων κλάσεων που ανήκουν σε μια εικόνα

Μηχανική όραση για ανίχνευση ανωμαλιών σε οδικές υποδομές με χρήση νευρωνικών δικτύων

Για το σύνολο των 382 εικόνων του dataset έγιναν τα ακόλουθα μαρκαρίσματα

Κατηγορία φθορών οδοστρώματος	Σύνολο μαρκαρισμάτων
cracking	447
markings wear	406
edge failure	59
potholes	239

2.1.5. Δείγμα Εκπαίδευσης-Εκπαίδευσης-Δοκιμής

Χρησιμοποιήθηκαν όλα τα δεδομένα(εικόνες) για δεδομένα εκπαίδευσης και δεδομένα επαλήθευσης (τα δεδομένα εκπαίδευσης χρησιμοποιούνται για την εκπαίδευση του μοντέλου και τα δεδομένα επαλήθευσης χρησιμοποιούνται για την προσαρμογή του μοντέλου).

2.2. Εκπαίδευση μοντέλου

Για την εκπαίδευση του μοντέλου χρησιμοποίησα τη πλατφόρμα Google Colaboratory. Σε αυτή έτρεξα το κατάλληλο `ipynb` script το οποίο καλεί το Darknet για να κάνει την εκπαίδευση στο δείγμα εκπαίδευσης που του παρέixa. Το `.ipynb` είναι ένα `python` notebook το οποίο περιέχει τον κώδικα για την παραμετροποίηση και εκτέλεση της εφαρμογής στη πλατφόρμα Google Colaboratory. Το `python` script στο οποίο βασίζεται το `ipynb` είναι το ακόλουθο :

```
1  # -*- coding: utf-8 -*-
2  """Train_YoloV3_.ipynb
3
4  Automatically generated by Colaboratory.
5
6  Original file is located at
7  https://colab.research.google.com/drive/1FH08zeYyx2Xw0op1XdIn_AViEMlvubAt
8
9  **Connect google drive**
10 """
11
12 # Check if NVIDIA GPU is enabled
13 !nvidia-smi
14
15 from google.colab import drive
16 drive.mount('/content/gdrive')
17 !ln -s /content/gdrive/My\ Drive/ /mydrive
18 !ls /mydrive
19
20 """**1) Clone the Darknet**
21
22 """
23
24 !git clone https://github.com/AlexeyAB/darknet
25
26 """**2) Compile Darknet using Nvidia GPU**
27
28 """
29
30 # Commented out IPython magic to ensure Python compatibility.
31 # change makefile to have GPU and OPENCV enabled
32 # %cd darknet
33 !sed -i 's/OPENCV=0/OPENCV=1/' Makefile
34 !sed -i 's/GPU=0/GPU=1/' Makefile
35 !sed -i 's/CUDNN=0/CUDNN=1/' Makefile
36 !make
37
38 """**3) Configure Darknet network for training YOLO V3**""
39
40 !cp cfg/yolov3.cfg cfg/yolov3_training.cfg
41
42 !sed -i 's/batch=1/batch=8/' cfg/yolov3_training.cfg
43 !sed -i 's/subdivisions=1/subdivisions=4/' cfg/yolov3_training.cfg
44 !sed -i 's/max_batches = 500200/max_batches = 8000/' cfg/yolov3_training.cfg
45 !sed -i '610 s@classes=80@classes=4@' cfg/yolov3_training.cfg
46 !sed -i '696 s@classes=80@classes=4@' cfg/yolov3_training.cfg
47 !sed -i '783 s@classes=80@classes=4@' cfg/yolov3_training.cfg
48 !sed -i '603 s@filters=255@filters=27@' cfg/yolov3_training.cfg
49 !sed -i '689 s@filters=255@filters=27@' cfg/yolov3_training.cfg
50 !sed -i '776 s@filters=255@filters=27@' cfg/yolov3_training.cfg
51
52 # Create folder on google drive so that we can save there the weights
53 !mkdir "/mydrive/yolov3"
54
55 !echo -e 'cracking\nmarkings wear\nedge failure\npotholes\n' > data/obj.names
56 !echo -e 'classes=4\ntrain = data/train.txt\nvalid = data/test.txt\nnames = data/obj.names\nbackup = /mydrive/yolov3' > data/obj.data
57 !mkdir data/obj
58
59 # Download weights darknet model 53
60 !wget https://pjreddie.com/media/files/darknet53.conv.74
61
62 """**4) Extract Images**
63
64 The images need to be inside a zip archive called "images.zip" and they need to be inside the folder "yolov3" on Google Drive
65 """
66
67 !unzip /mydrive/yolov3/images.zip -d data/obj
68
69 import glob
70 images_list = glob.glob("data/obj/*.jpg")
71 print(images_list)
72
73 #Create training.txt file
74 file = open("data/train.txt", "w")
75 file.write("\n".join(images_list))
76 file.close()
77
78 """**6) Start the training**""
79
80 # Start the training
81 !./darknet detector train data/obj.data cfg/yolov3_training.cfg darknet53.conv.74 -dont_show
82
```

Εικόνα 31 : Δομή αρχείου `ipynb`

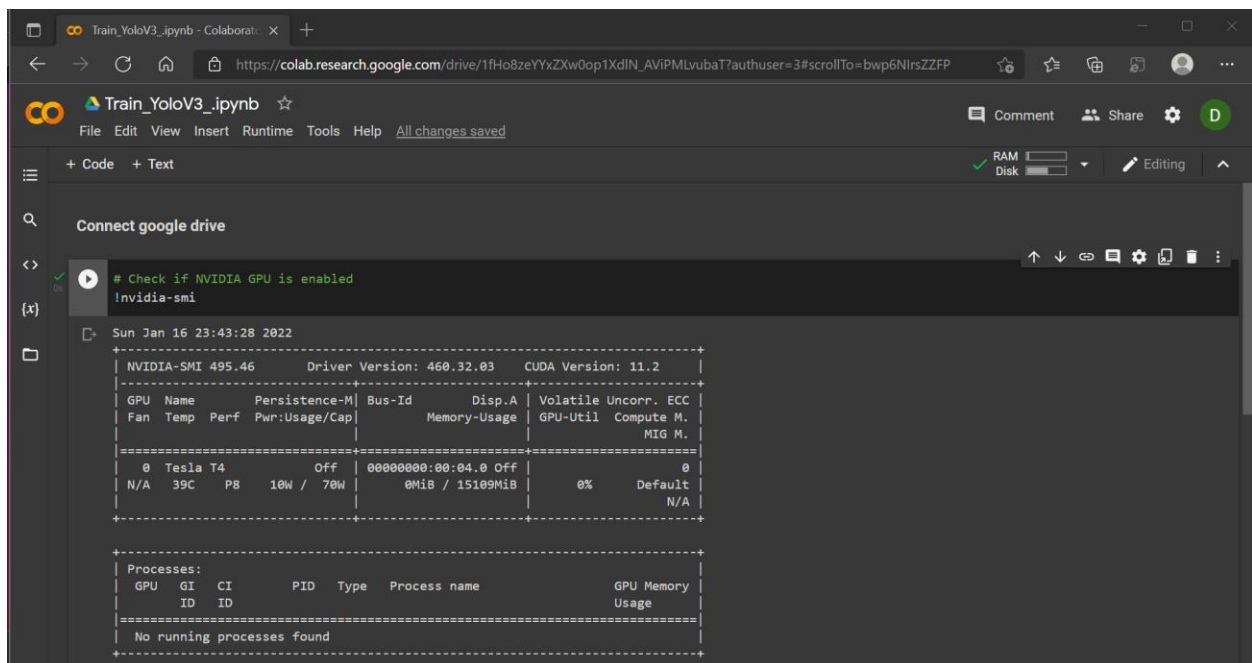
Μηχανική όραση για ανίχνευση ανωμαλιών σε οδικές υποδομές με χρήση νευρωνικών δικτύων

Κατά τη διάρκεια της εκπαίδευσης του μοντέλου ρυθμίζονται κάποιες παράμετροι του Darknet ανάλογα πως θέλουμε να εκπαιδύσουμε το μοντέλο μας. Οι σημαντικότερες που αλλάζουμε κάθε φορά που εκτελούμε την εκπαίδευση του μοντέλου μας και που έχουμε περιγράψει νωρίτερα είναι :

batch: Το μέγεθος της παρτίδας (Batch size)

subdivisions: Οι υποδιαιρέσεις, δηλαδή το σύνολο των εικόνων που επεξεργάζεται μαζί η GPU

max_batches: Το μέγιστο πλήθος των iterations

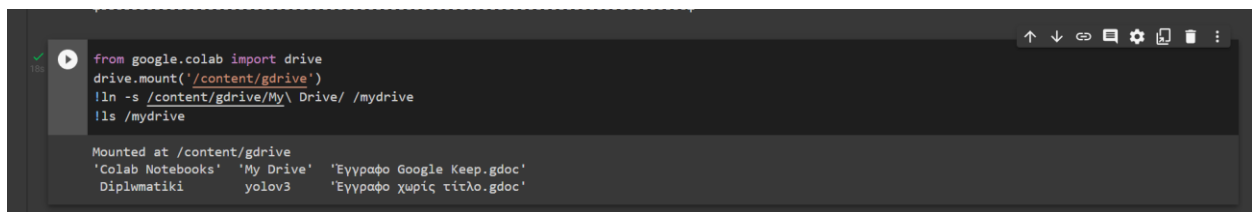


```
# Check if NVIDIA GPU is enabled
nvidia-smi

Sun Jan 16 23:43:28 2022

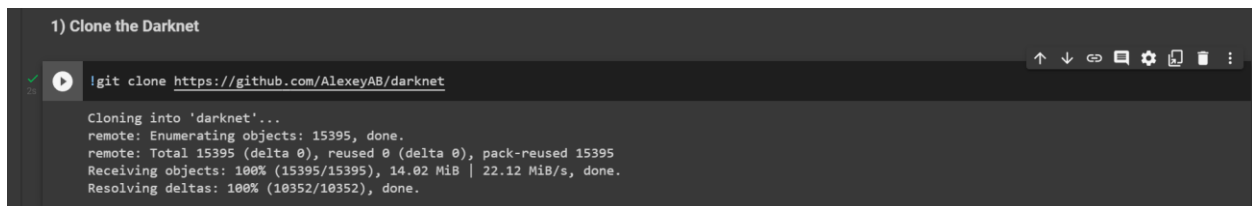
+-----+
| NVIDIA-SMI 495.46      | Driver Version: 460.32.03   | CUDA Version: 11.2   |
+-----+-----+
| GPU  Name            Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC  |
| Fan  Temp  Perf    Pwr:Usage/Cap|  Memory-Usage | GPU-Util  Compute M. |
|-----+-----+-----+
| 0   Tesla T4              Off          | 00000000:00:04:0  Off  |    0          0      |
| N/A   39C   P8      10W /  70W   |  0MiB / 15109MiB |    0%      Default  |
+-----+-----+-----+

Processes:
GPU  GI  CI          PID  Type  Process name          GPU Memory
ID   ID   ID                                     Usage
-----+-----+
No running processes found
```



```
from google.colab import drive
drive.mount('/content/gdrive')
!ln -s /content/gdrive/My\ Drive/ /mydrive
!ls /mydrive

Mounted at /content/gdrive
'Colab Notebooks' 'My Drive' 'Έγγραφο Google Keep.gdoc'
'Diplmatiki'      'yolov3' 'Έγγραφο χωρίς τίτλο.gdoc'
```



```
1) Clone the Darknet

!git clone https://github.com/AlexeyAB/darknet

Cloning into 'darknet'...
remote: Enumerating objects: 15395, done.
remote: Total 15395 (delta 0), reused 0 (delta 0), pack-reused 15395
Receiving objects: 100% (15395/15395), 14.02 MiB | 22.12 MiB/s, done.
Resolving deltas: 100% (10352/10352), done.
```

```
2) Compile Darknet using Nvidia GPU

[4] # change makefile to have GPU and OPENCV enabled
%cd darknet
!sed -i 's/OPENCV=0/OPENCV=1/' Makefile
!sed -i 's/GPU=0/GPU=1/' Makefile
!sed -i 's/CUDNN=0/CUDNN=1/' Makefile
!make

./src/im2col_kernels.cu(1389): warning: unrecognized #pragma in device code

nvcc -gencode arch=compute_35,code=sm_35 -gencode arch=compute_50,code=[sm_50,compute_50] -gencode arch=compute_52,code=[sm_52,compute_52] -gencode
nvcc warning : The 'compute_35', 'compute_37', 'compute_50', 'sm_35', 'sm_37' and 'sm_50' architectures are deprecated, and may be removed in a futu
nvcc -gencode arch=compute_35,code=sm_35 -gencode arch=compute_50,code=[sm_50,compute_50] -gencode arch=compute_52,code=[sm_52,compute_52] -gencode
```

```
3) Configure Darknet network for training YOLO V3

[5] !cp cfg/yolov3.cfg cfg/yolov3_training.cfg
```

Στο αρχείο yolov3_training.cfg που περιέχει τις ρυθμίσεις για το training, κάνουμε τις ακόλουθες αλλαγές :

batch=8, subdivisions=4, max_batches = 8000, classes=4, filters=27

Η συνάρτηση για τον υπολογισμό της μεταβλητής filters είναι : $(\text{number of classes} + 5) * 3$

```
[6] !sed -i 's/batch=1/batch=8/' cfg/yolov3_training.cfg
!sed -i 's/subdivisions=1/subdivisions=4/' cfg/yolov3_training.cfg
!sed -i 's/max_batches = 500200/max_batches = 8000/' cfg/yolov3_training.cfg
!sed -i '610 s@classes=80@classes=4@' cfg/yolov3_training.cfg
!sed -i '696 s@classes=80@classes=4@' cfg/yolov3_training.cfg
!sed -i '783 s@classes=80@classes=4@' cfg/yolov3_training.cfg
!sed -i '603 s@filters=255@filters=27@' cfg/yolov3_training.cfg
!sed -i '689 s@filters=255@filters=27@' cfg/yolov3_training.cfg
!sed -i '776 s@filters=255@filters=27@' cfg/yolov3_training.cfg
```

Ορίζουμε ένα φάκελο στο Google Drive για να αποθηκευτούν τα αποτελέσματα των υπολογισμών του training

```
[7] # Create folder on google drive so that we can save there the weights
!mkdir "/mydrive/yolov3"
```

Ορίζουμε τα ονόματα και το πλήθος των κλάσεων της εκπαίδευσης

```
[8] !echo -e 'cracking\nmarkings wear\nedge failure\npotholes\n' > data/obj.names
!echo -e 'classes = 4\ntrain = data/train.txt\nvalid = data/test.txt\nnames = data/obj.names\nbackup = /mydrive/yolov3' > data/obj.data
!mkdir data/obj
```

Για την εκπαίδευση χρησιμοποιούμε την βάση δεδομένων εικόνων Imagenet που περιέχει συνολικά βάρη που είναι προ-εκπαιδευμένα από διάφορα μοντέλα. Για την εκπαίδευση μας χρησιμοποιούμε τα βάρη από το μοντέλο darknet53.

```
[9] # Download weights darknet model 53
!wget https://pjreddie.com/media/files/darknet53.conv.74

--2022-01-16 23:57:46-- https://pjreddie.com/media/files/darknet53.conv.74
Resolving pjreddie.com (pjreddie.com)... 128.208.4.108
Connecting to pjreddie.com (pjreddie.com)|128.208.4.108|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 162482580 (155M) [application/octet-stream]
Saving to: 'darknet53.conv.74'

darknet53.conv.74 100%[=====] 154.96M 40.0MB/s in 4.3s

2022-01-16 23:57:50 (36.4 MB/s) - 'darknet53.conv.74' saved [162482580/162482580]
```

Μηχανική όραση για ανίχνευση ανωμαλιών σε οδικές υποδομές με χρήση νευρωνικών δικτύων

Αποσυμπιέζουμε το αρχείο που περιέχει τις εικόνες με τα annotation τους, που δημιουργήσαμε κατά την προετοιμασία του dataset

```
4) Extract Images
The images need to be inside a zip archive called "images.zip" and they need to be inside the folder "yolov3" on Google Drive

[10] !unzip /mydrive/yolov3/images.zip -d data/obj
Archive: /mydrive/yolov3/images.zip
  inflating: data/obj/0_gettyimages-655985038-170667a.jpg
  inflating: data/obj/0_gettyimages-655985038-170667a.txt
  inflating: data/obj/0_JAR_BRK_050220_TilehurstPotHole_02jfif (1).jpg

[11] import glob
images_list = glob.glob("data/obj/*.jpg")
print(images_list)
['data/obj/thumb_720_450_dreamstime_xl_30104007-custom (1).jpg', 'data/obj/IMG20210918170158.jpg', 'data/obj/EdgeDropOff2_mirror.jpg', 'data/obj/3_mir

[12] #Create training.txt file
file = open("data/train.txt", "w")
file.write("\n".join(images_list))
file.close()
```

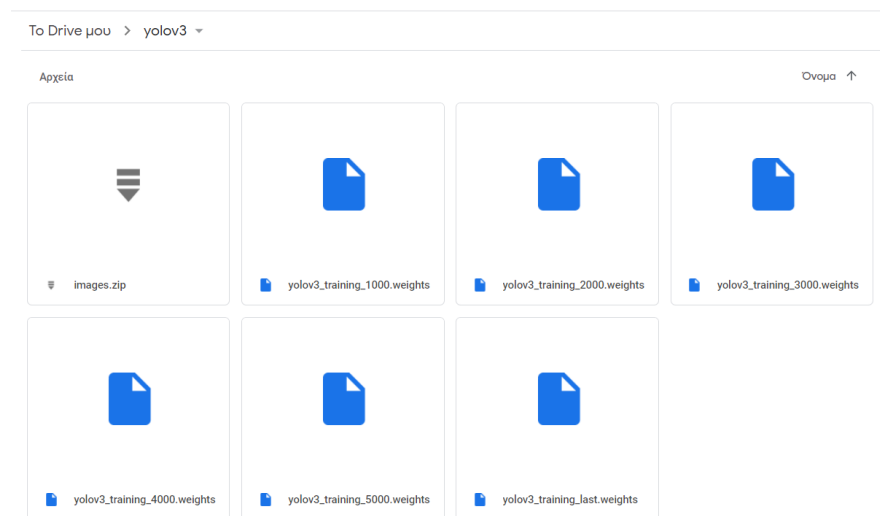
Πλέον έχουμε ρυθμίσει το σύστημα εκπαίδευσης και είμαστε έτοιμοι να ξεκινήσουμε την εκπαίδευση

```
6) Start the training

# Start the training
!./darknet detector train data/obj.data cfg/yolov3_training.cfg darknet53.conv.74 -dont_show

CUDA-version: 11010 (11020), cuDNN: 7.6.5, GPU count: 1
OpenCV version: 3.2.0
yolov3_training
 0 : compute_capability = 750, cudnn_half = 0, GPU: Tesla T4
net.optimized_memory = 0
mini_batch = 2, batch = 8, time_steps = 1, train = 1
  layer  filters  size/strd(dil)  input           output
 0 Create CUDA-stream - 0
Create cudnn-handle 0
conv      32      3 x 3/ 1      416 x 416 x  3 -> 416 x 416 x  32 0.299 BF
1 conv    64      3 x 3/ 2      416 x 416 x 32 -> 208 x 208 x  64 1.595 BF
2 conv    32      1 x 1/ 1      208 x 208 x 64 -> 208 x 208 x  32 0.177 BF
```

Με την ολοκλήρωση της διαδικασίας, τα βάρη που προκύπτουν από την εκπαίδευση αποθηκεύονται στο google drive



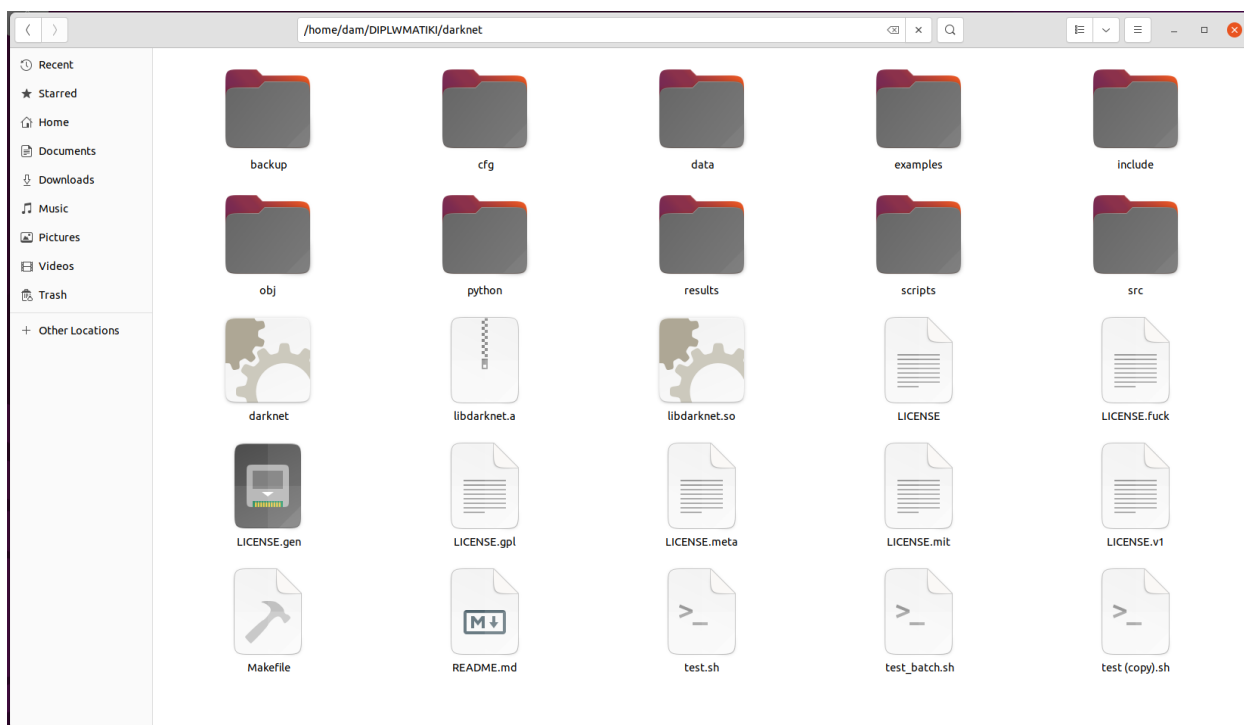
2.3. Δοκιμή μοντέλου και ανίχνευση

Η εκπαίδευση έγινε με διάφορες παραμέτρους εκπαίδευσης (αριθμός κλάσεων , αριθμός batch, αριθμός subdivisions) και προέκυπταν τα αντίστοιχα αρχεία βαρών κάθε φορά.

Για την δοκιμή του κάθε μοντέλου που προέκυψε από την εκπαίδευση , εκτέλεσα την ανίχνευση των 4 κλάσεων στις εικόνες του dataset με βάση αυτό το μοντέλο.

Κατά την διαδικασία της ανίχνευσης κάθε ανιχνευόμενη κλάση ανιχνευόταν με ένα ποσοστό βεβαιότητας. Για την ανίχνευση μπορούσε να οριστεί ένα κατώφλι ανίχνευσης, ώστε όσες περιοχές της εικόνας που ανιχνεύονταν με ποσοστό βεβαιότητας χαμηλότερο του κατωφλίου να αγνοούνται.

Η εκτέλεση των δοκιμών έγινε τοπικά σε υπολογιστή με λειτουργικό Linux όπου εγκαταστάθηκε το Darknet.



Εικόνα 32 : Φάκελος αρχείων του Darknet

Για τα μοντέλα που προέκυψαν από την εκπαίδευση που έτρεξε στο Colaboratory για το dataset μου, εκτέλεσα δοκιμές ανίχνευσης.

Η ανίχνευση πραγματοποιείται με την εκτέλεση της εντολής :

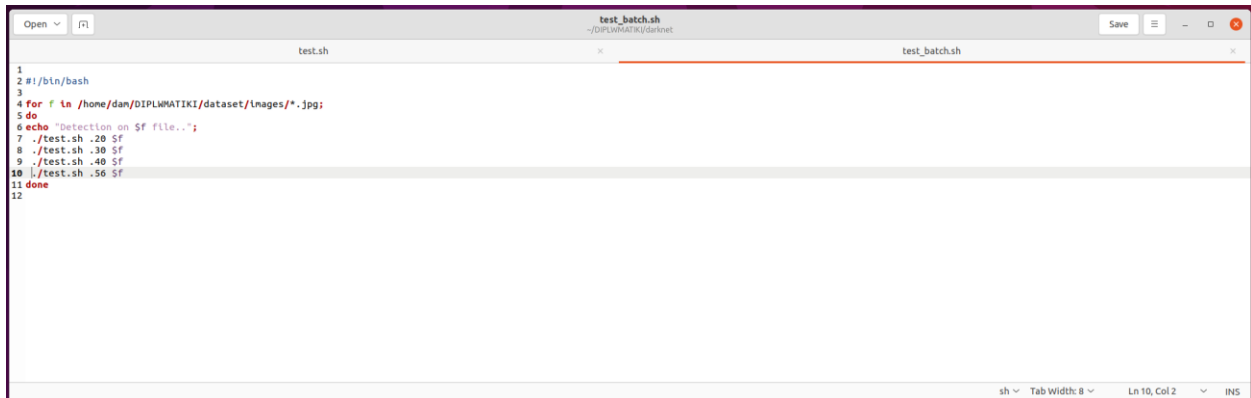

```
./darknet detector test <αρχείο πληροφοριών κλάσεων> <αρχείο ρυθμίσεων παραμέτρων εκτέλεσης>
-thresh <επίπεδο threshold με τιμές από 0.0 έως 1.0> <μοντέλο εκπαίδευσης> <αρχείο που θα γίνει η
δοκιμή ανίχνευσης των κλάσεων>
```

Για την επιλογή του μοντέλου εκπαίδευσης όπως και των παραμέτρων threshold, που θα χρησιμοποιούταν κάθε φορά για την ανίχνευση δημιουργήσα τα κατάλληλα bash script .



```
1 #!/bin/bash
2 # arg1 threshold of detection 0 to 1 , 0 detects all
3 # arg2 file to test
4 batch="64"
5 subdivisions="4"
6 model="/home/dan/DIPLWMATIKI/models/22.9.2021_dataset/training_4_categories__batch_"$batch"_subdivisions_"$subdivisions"/yolov3_training_last.weights
7 dataset_version="22.9.2021_dataset"
8
9 #4 classes
10 ./darknet detector test cfg/coco_cracking_markings_wear_edge_fallure_potholes.data cfg/yolov3_classes_4.cfg -thresh $1 $model $2
11
12 s=$2
13 filename="${s##*/}"
14 echo $filename
15
16
17 if [ ! -d /home/dan/DIPLWMATIKI/detections/detects_for_$dataset_verion/batch_"$batch"_subdivisions_"$subdivisions"/with_threshold_$1 ]
18 then
19     mkdir -p /home/dan/DIPLWMATIKI/detections/detects_for_$dataset_verion/batch_"$batch"_subdivisions_"$subdivisions"/with_threshold_$1
20 fi
21 mv /home/dan/DIPLWMATIKI/darknet/predictions.jpg /home/dan/DIPLWMATIKI/detections/detects_for_$dataset_verion/batch_"$batch"_subdivisions_"$subdivisions"/with_threshold_$1/"$filename"_detection_thresh_"$1".jpg
22
23
```

Εικόνα 33 : Bash script (test.sh) παραμετροποιημένης εκτέλεσης ανίχνευσης εικόνας και αποθήκευσης εικόνας αποτελέσματος



```
1
2 #!/bin/bash
3
4 for f in /home/dan/DIPLWMATIKI/dataset/images/*.jpg;
5 do
6 echo "Detection on $f file..";
7 ./test.sh .20 $f
8 ./test.sh .30 $f
9 ./test.sh .40 $f
10 ./test.sh .50 $f
11 done
12
```

Εικόνα 34 : Bash script (test_batch.sh) εκτέλεσης του test.sh για το σύνολο των εικόνων εντός φακέλου με τα επιλεγμένα threshold

Κατά την εκτέλεση της ανίχνευσης για κάθε εικόνα εμφανίζονται στο τερματικό πρώτα τα βάρη του μοντέλου που φορτώθηκε (Εικόνα 35) και στην συνέχεια το σύνολο των περιοχών που ανιχνεύθηκαν με πληροφορίες για την κλάση που ανήκουν και ο βαθμός βεβαιότητας με τον οποίο ανιχνεύθηκαν (Εικόνα 36)

```
0_gettyimages-655985038-170667a.jpg
```

layer	filters	size	input	output	
0 conv	32	3 x 3 / 1	608 x 608 x 3	608 x 608 x 32	0.639 BFLOPs
1 conv	64	3 x 3 / 2	608 x 608 x 32	304 x 304 x 64	3.407 BFLOPs
2 conv	32	1 x 1 / 1	304 x 304 x 64	304 x 304 x 32	0.379 BFLOPs
3 conv	64	3 x 3 / 1	304 x 304 x 32	304 x 304 x 64	3.407 BFLOPs
4 res	1		304 x 304 x 64	304 x 304 x 64	
5 conv	128	3 x 3 / 2	304 x 304 x 64	152 x 152 x 128	3.407 BFLOPs
6 conv	64	1 x 1 / 1	152 x 152 x 128	152 x 152 x 64	0.379 BFLOPs
7 conv	128	3 x 3 / 1	152 x 152 x 64	152 x 152 x 128	3.407 BFLOPs
8 res	5		152 x 152 x 128	152 x 152 x 128	
9 conv	64	1 x 1 / 1	152 x 152 x 128	152 x 152 x 64	0.379 BFLOPs
10 conv	128	3 x 3 / 1	152 x 152 x 64	152 x 152 x 128	3.407 BFLOPs
11 res	8		152 x 152 x 128	152 x 152 x 128	
12 conv	256	3 x 3 / 2	152 x 152 x 128	76 x 76 x 256	3.407 BFLOPs
13 conv	128	1 x 1 / 1	76 x 76 x 256	76 x 76 x 128	0.379 BFLOPs
14 conv	256	3 x 3 / 1	76 x 76 x 128	76 x 76 x 256	3.407 BFLOPs
15 res	12		76 x 76 x 256	76 x 76 x 256	
16 conv	128	1 x 1 / 1	76 x 76 x 256	76 x 76 x 128	0.379 BFLOPs
17 conv	256	3 x 3 / 1	76 x 76 x 128	76 x 76 x 256	3.407 BFLOPs
18 res	15		76 x 76 x 256	76 x 76 x 256	

Εικόνα 35 : Στιγμιότυπο εκτέλεσης ανίχνευσης όπου φαίνεται η φόρτωση των βαρών του μοντέλου

```
Loading weights from /home/dam/DIPLWMAIKI/models/22.9.2021_dataset/training_4_categories__batch_64_subdivisions_8/yolo
t.weights...Done!
/home/dam/DIPLWMAIKI/dataset/images/0_gettyimages-655985038-170667a.jpg: Predicted in 10.163996 seconds.
potholes: 60%
potholes: 56%
markings_wear: 58%
cracking: 63%
cracking: 60%
cracking: 58%
cracking: 57%
0_gettyimages-655985038-170667a.jpg
```

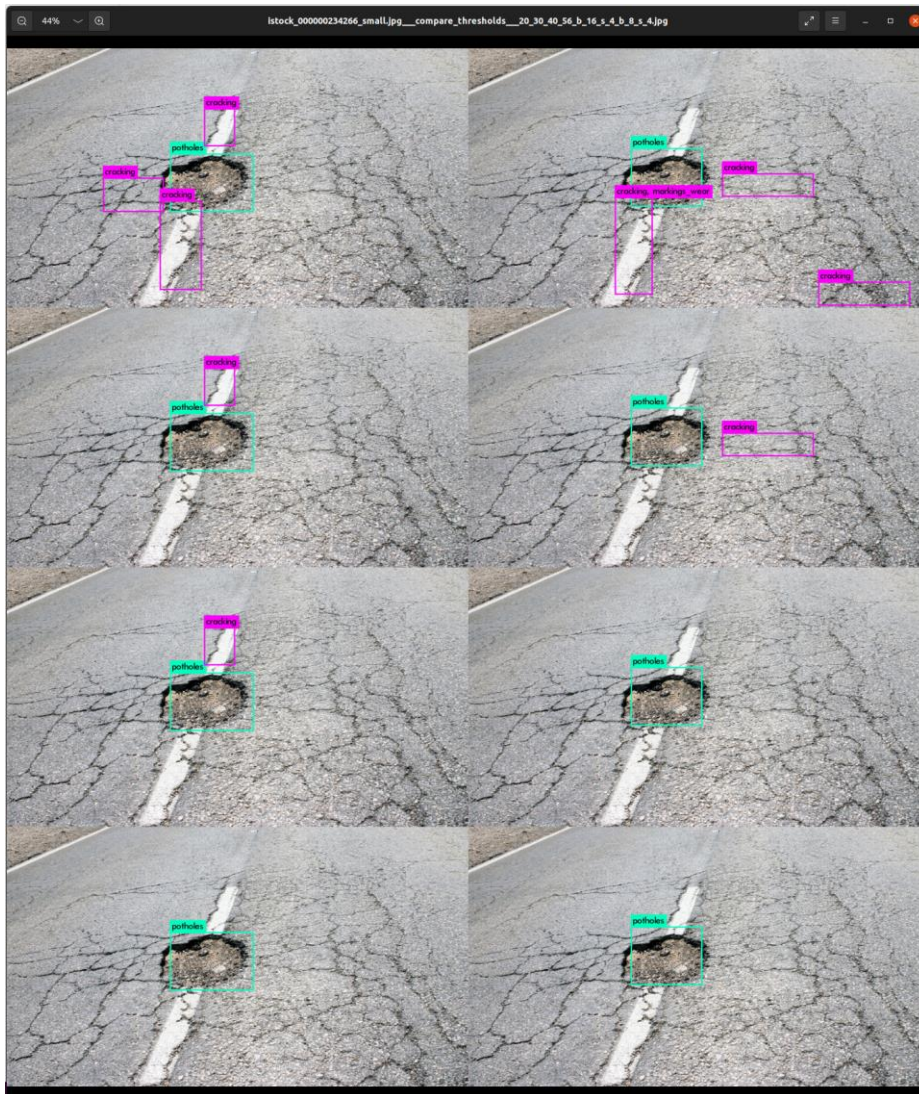
Εικόνα 36 : Στιγμιότυπο εκτέλεσης ανίχνευσης όπου φαίνονται οι ανιχνεύσεις στην εξεταζόμενη εικόνα

Στην παρακάτω εικόνα βλέπουμε ανιχνεύσεις με διαφορετικά επίπεδα βεβαιότητας, ξεκινώντας από πάνω με threshold 20% , συνεχίζοντας με 30%, 40% και καταλήγοντας σε 56%. Παρατηρούμε ότι όσο χαμηλότερο το threshold η ανίχνευση είναι λιγότερο αυστηρή παρουσιάζοντας περιοχές που ανήκουν σε κάποια κλάση με μικρότερη βεβαιότητα.

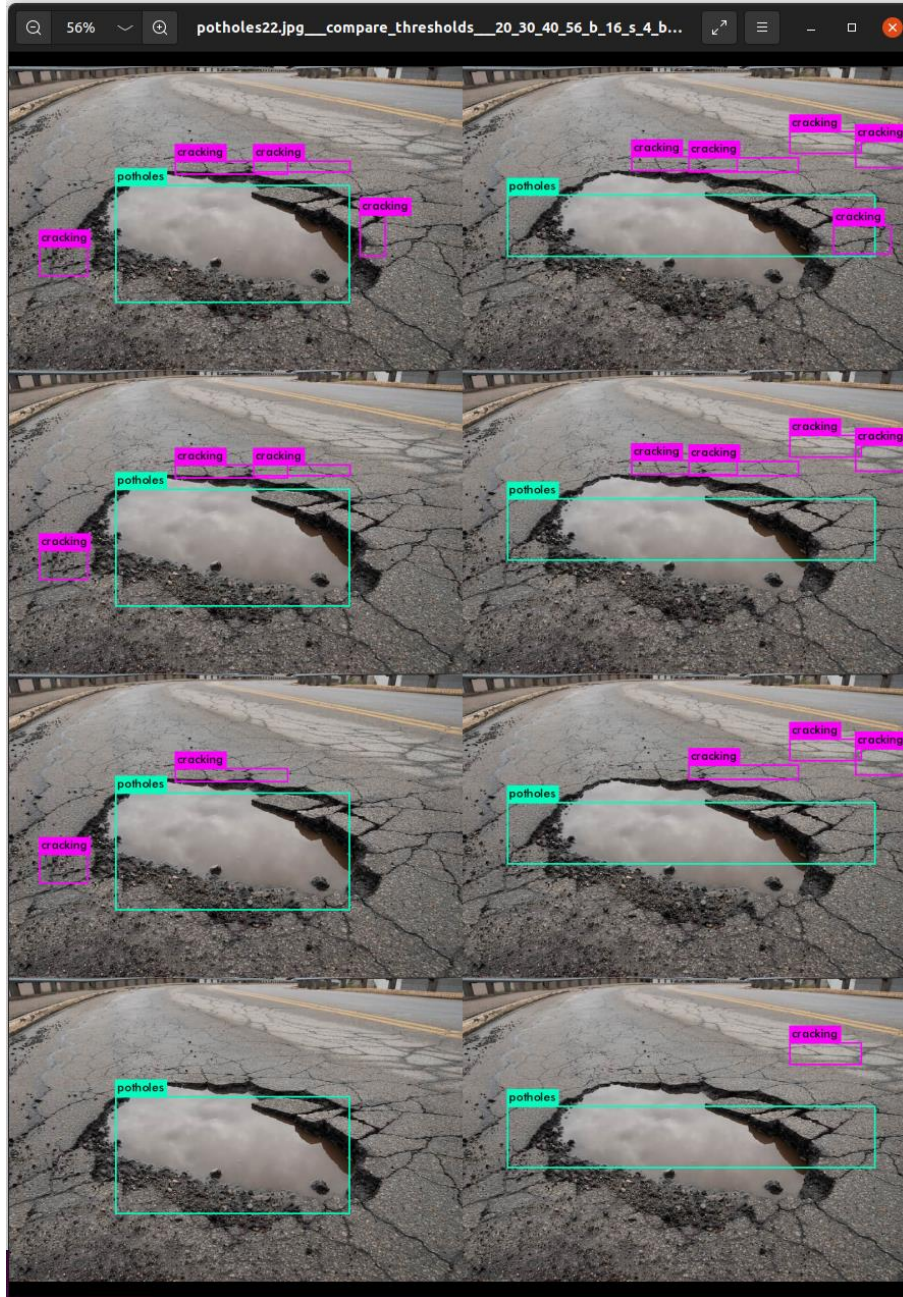


Εικόνα 37 : Παράδειγμα ανίχνευσης κλάσεων ανωμαλιών ξεκινώντας από πάνω με threshold=20 φτάνοντας προς τα κάτω με threshold=56

Επίσης ακολουθεί μια εικόνα που δείχνει πως επιδρούν οι παράμετροι batch και subdivision. Βλέπουμε ότι διαφοροποιούνται ως προς την ποιοτική ανίχνευση της κάθε κλάσης. Αυτό φαίνεται πιο καθαρά σε περιοχές μικρότερης βεβαιότητας όπου η διαφορετική ομαδοποίηση στο μοντέλο εκπαίδευσης, κατά την ανίχνευση επηρεάζει τα βάρη των μοντέλων με διαφορετικό τρόπο. Γενικότερα ισχύει ότι όσο μικραίνει η τιμή του batch τόσο αυξάνεται η ταχύτητα εκπαίδευσης και η εκπαίδευση χρειάζεται μικρότερο κύκλο iterations και όσο μεγαλώνει η τιμή του μειώνεται η ικανότητα του μοντέλου ως προς τη γενίκευση.

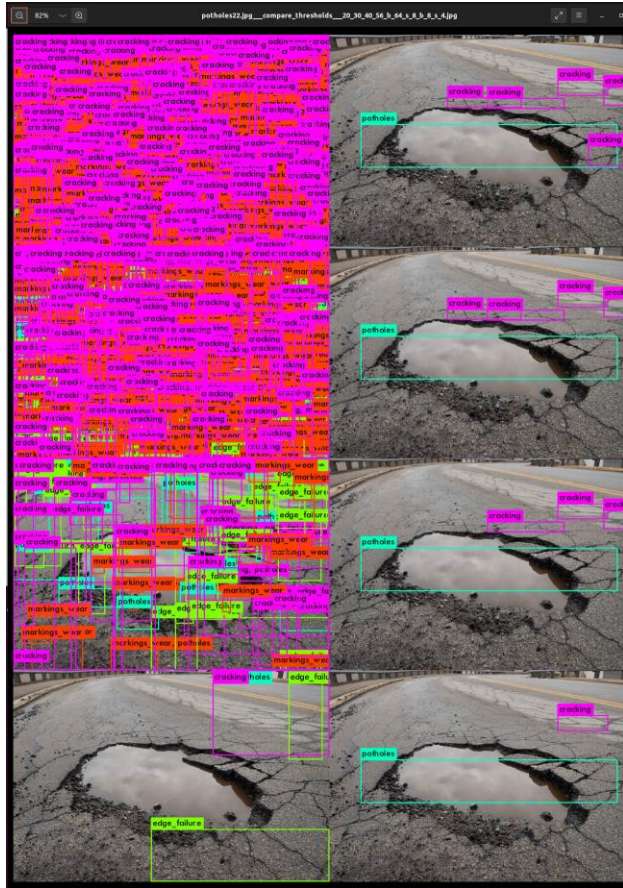


Εικόνα 38 : Πλέγμα 2x4 εικόνων όπου στις στήλες έχουμε παραμέτρους batch=16, subdivision=4 για την 1^η στήλη και batch=8, subdivision=4 για την 2^η στήλη και τιμές threshold= 20,30,40,56 από πάνω προς τα κάτω

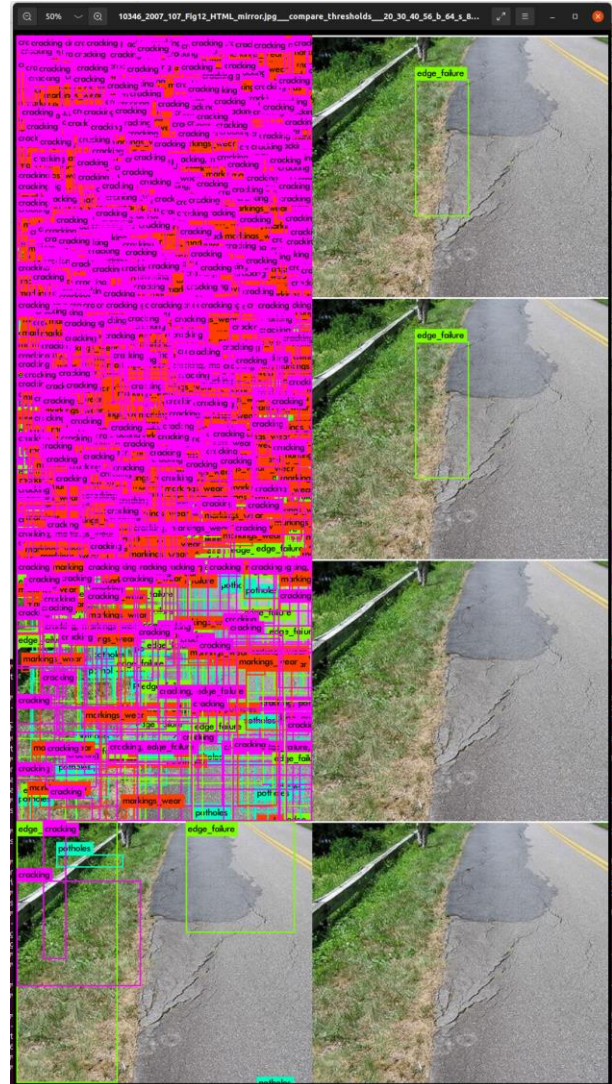


Εικόνα 39 : Πλέγμα 2x4 εικόνων όπου στις στήλες έχουμε παραμέτρους batch=16, subdivision=4 για την 1η στήλη και batch=8, subdivision=4 για την 2η στήλη και τιμές threshold= 20,30,40,56 από πάνω προς τα κάτω

Μηχανική όραση για ανίχνευση ανωμαλιών σε οδικές υποδομές με χρήση νευρωνικών δικτύων



Εικόνα 40 : Πλέγμα 2x4 εικόνων όπου στις στήλες έχουμε παραμέτρους batch=64, subdivision=8 για την 1η στήλη και batch=8, subdivision=4 για την 2η στήλη και τιμές threshold= 20,30,40,56 από πάνω προς τα κάτω



Εικόνα 41 : Πλέγμα 2x4 εικόνων όπου στις στήλες έχουμε παραμέτρους batch=64, subdivision=8 για την 1η στήλη και batch=8, subdivision=4 για την 2η στήλη και τιμές threshold= 20,30,40,56 από πάνω προς τα κάτω

Επίσης δοκιμάστηκε η διαφορά στην ανίχνευση, βλέποντας πως την επηρεάζει το dataset που χρησιμοποιείται για την εκπαίδευση. Από το αρχικό dataset των 382 εικόνων που αναφέρθηκε προηγουμένως χρησιμοποιήθηκε ένα υποσύνολο 354 εικόνων με το ακόλουθο πλήθος μαρκαρισμένων περιοχών

Κατηγορία φθορών οδοστρώματος	Σύνολο μαρκαρισμάτων
cracking	211
markings wear	89
edge failure	17
potholes	79

Μηχανική δράση για ανίχνευση ανωμαλιών σε οδικές υποδομές με χρήση νευρωνικών δικτύων

Παρατηρούμε ότι αυτό έχει σε σχέση με το αρχικό dataset λιγότερα μαρκαρίσματα (markings wear) τόσο σε πλήθος όσο και σε αναλογικά. Αυτή η ποιοτική διαφορά φαίνεται στην παρακάτω εικόνα με λιγότερα false positives ειδικά για αυτήν την κατηγορία.



Εικόνα 42 : Πλέγμα 2x4 εικόνων όπου στις στήλες έχουμε παραμέτρους batch=64, subdivision=8 για την 1η στήλη και batch=64, subdivision=8 για την 2η στήλη και τιμές threshold= 20,30,40,56 από πάνω προς τα κάτω. Η αριστερή στήλη αναφέρεται σε ανιχνεύσεις με training από το υποσύνολο του dataset και η δεύτερη από το πλήρες

3. Τεχνο-οικονομική ανάλυση

3.1. Ατυχήματα Λόγω Κακής Συντήρησης Οδοστρώματος και έμμεσα/ άμεσα κόστη

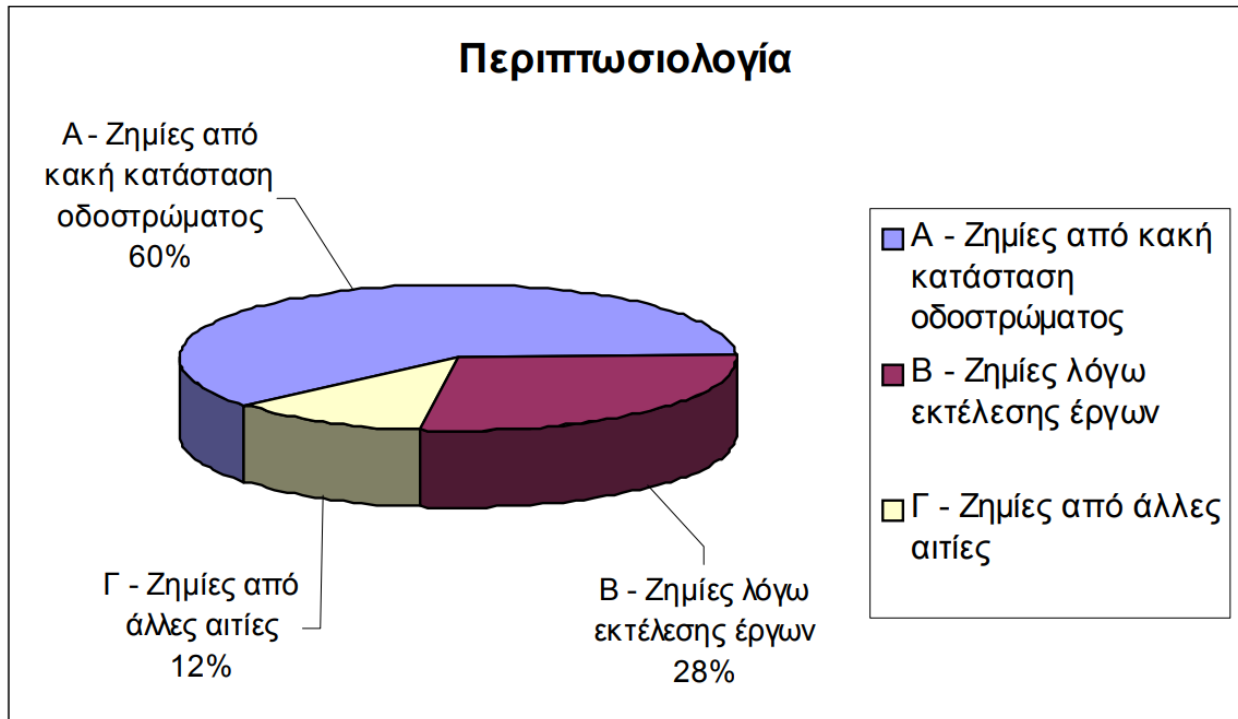
Τα κόστη που προκύπτουν λόγω της κακής κατάστασης των οδοστρώματων δεν είναι αμελητέα. Ακολουθεί μια σχετική αναφορά από το Συνήγορο του Πολίτη^[25]. Ήδη από το δεύτερο έτος της λειτουργίας του (1999), ο Συνήγορος του Πολίτη (ΣτΠ) άρχισε να δέχεται αναφορές με αίτημα τη διαμεσολάβηση της Αρχής προς τους Οργανισμούς Τοπικής Αυτοδιοίκησης Α΄ και Β΄ βαθμού (στο εξής: ΟΤΑ Α΄ και Β΄ βαθμού), προκειμένου οι τελευταίοι να προβούν στην αποκατάσταση ενός ορισμένου τύπου ζημιών : Σε 207 αναφορές που το 2004 είχαν κατατεθεί στον ΣτΠ, ζητήθηκε η άμεση, δηλαδή χωρίς προσφυγή στη δικαιοσύνη, χρηματική ικανοποίηση πολιτών για βλάβες που υπέστησαν τα οχήματά τους από κακοτεχνίες του δημοτικού και επαρχιακού οδικού δικτύου (λ.χ. λακκούβες), οφειλόμενες σε έλλειψη συντήρησης ή αδυναμία διατήρησης αυτού σε κατάσταση που να ανταποκρίνεται στον λειτουργικό του προορισμό. Τα αιτήματα αποζημίωσης των πολιτών αφορούσαν και ατυχήματα από ελλιπή ή πλημμελή σήμανση οδών, ζημίες από πτώση κλαδιών ή κορμών δένδρων, μετακινήσεις κάδων απορριμμάτων ή και από άλλες παρεμφερείς αιτίες. Στη συντριπτική πλειονότητα των περιπτώσεων οι εμπλεκόμενοι ΟΤΑ αρνήθηκαν να καταβάλουν οποιαδήποτε αποζημίωση.

Από τον συνολικό αριθμό των 207 σχετικών αναφορών, 178 αφορούσαν ΟΤΑ α΄ βαθμού και 29 ΟΤΑ β΄ βαθμού. Η περαιτέρω στατιστική ανάλυση των αναφορών αυτών αναδεικνύει τρεις μεγάλες κατηγορίες αιτημάτων αποζημίωσης:

α) Αποζημίωση για υλικές ζημίες που προκλήθηκαν σε όχημα από κακή κατάσταση του οδοστρώματος, η οποία με τη σειρά της μπορεί να οφείλεται είτε σε εξαρχής κακοτεχνία (π.χ. ανωμαλίες του ασφαλτοτάπητα, φρεάτια που προεξέχουν ή είναι υπερβολικά βυθισμένα), είτε σε μετέπειτα φθορά αυτού (καθίζηση ή βύθιση του οδοστρώματος που προκλήθηκε από την πάροδο του χρόνου ή λόγω καιρικών συνθηκών, σε συνδυασμό με ολιγωρία επισκευής αυτού από τον αρμόδιο φορέα).

β) Αποζημίωση για υλικές ζημίες που προκλήθηκαν σε οχήματα λόγω της εκτέλεσης έργων στο οδόστρωμα από συνεργεία δήμου ή νομαρχίας, από υπεργολάβο ή, τέλος, από Δημοτικές Επιχειρήσεις Ύδρευσης Αποχέτευσης (στο εξής: ΔΕΥΑ).

γ) Αποζημίωση για υλικές ζημιές που προκλήθηκαν σε οχήματα από άλλες αιτίες, όπως πτώση κλαδιών δέντρων, μετακίνηση (κύλιση) κάδου απορριμμάτων, ελλιπής σήμανση οδικής κυκλοφορίας, κολωνάκια πεζοδρομίου



Εικόνα 43 : Κατηγορίες αιτημάτων αποζημίωσης

Table 2 Comparison of road crash costs with particular reference to human costs

Costs	Cost (\$ million)	Human costs as a per cent of total human cost	Road crash costs	Comparison of cost components
Human costs				
Long-term care	1 990	23.7%	1,990	13.3%
Labour in the workplace	1 625	19.4%	1,625	10.8%
Labour in the household	1 493	17.8%	1,493	10.0%
Quality of life	1 769	21.1%	1,769	11.8%
Legal and Insurance Admin costs	813	9.7%	813	5.4%
Medical and other associated costs	361	4.3%	361	2.4%
Correctional services	17	0.2%	17	0.1%
Workplace disruption	313	3.7%	313	2.1%
Funeral	3	0.0%	3	0.0%
Coroner	1	0.0%	1	0.0%
Total human costs	8,385	100.0%		56.0%
Total vehicle costs			4,110	27.4%
Total general costs			2,485	16.6%
Total of all crash costs			14,980	100.0%

Source BTRE estimates based on BTE (2000)

Εικόνα 44 : Σχέση τροχαίων ατυχημάτων και απώλειας ανθρώπινων ζωών

3.2.Κόστος αποκατάστασης φθορών οδοστρώματος

Η συντήρηση των φθορών οδοστρώματος μπορεί να επεκτείνει τον εκτιμώμενο χρόνο ζωής των οδοστρωμάτων και να καθυστερήσει την ανάγκη για αποκατάσταση. Οι επεμβάσεις αυτές μπορούν να εκτιμηθούν αν είναι κατάλληλο να γίνουν με ανάλυση του κόστους κύκλου ζωής. Συνεπώς η έγκαιρη διάγνωση των φθορών και η κατάλληλη αποκατάσταση είναι κρίσιμος παράγοντας για την συνολική διάρκεια ζωής του οδοστρώματος, την διατήρηση της ποιότητας τους και συνεπώς των εμμέσων (αποζημιώσεις λόγω ατυχημάτων από φθορές των οδοστρωμάτων, ζημιές στο στόλο των οχημάτων των πολιτών, απώλειες ανθρώπινων ζωών και των άμεσων κοστών (γρηγορότερη καταστροφή και ανάγκη συνολικής αποκατάστασης των οδοστρωμάτων)^[26]. Στην διεθνή βιβλιογραφία υπάρχουν γενιές οδηγίες για τις επεμβάσεις αποκατάστασης και για την εκτιμώμενη διάρκεια ζωής κάθε επέμβασης. Φυσικά η πραγματική διάρκεια ζωής μπορεί να διαφέρει εξαιτίας των τοπικών συνθηκών κυκλοφορίας, περιβάλλοντος και ποιότητας κατασκευής. Ακολουθούν πίνακες με στοιχεία σχετικά με την εκτιμώμενη διάρκεια ζωής του οδοστρώματος μετά από συγκεκριμένη επέμβαση συντήρησης εξαιτίας φθοράς ορισμένης έκτασης^[27].

Ρηγμάτωση λόγω κόπωσης				
Σοβαρότητα	Έκταση (% μήκους)	Πάχος ασφαλτικής στρώσης (mm)	Τυπικές επεμβάσεις	Εκτιμώμενη διάρκεια ζωής (χρόνια)
Χαμηλή	1-10	<200	“Μπάλωμα” ή τοπική πλήρωση ρωγμών	5
Μέση	11-25	>200	Επίστρωση 50mm	12
Μέση	11-25	<200	Επίστρωση 75mm	12
Μέση	25-50	>200	Επίστρωση 75mm	12
Μέση	25-50	<200	Επίστρωση 150mm	12
Υψηλή	>50	>200	Επίστρωση 100mm	12
Υψηλή	>50	<200	Ανακατασκευή	15
Εγκάρσια ρηγμάτωση				
Σοβαρότητα	Έκταση (απόσταση σε m)	Τυπικές επεμβάσεις		Εκτιμώμενη διάρκεια ζωής (χρόνια)
Χαμηλή	>15	Καθαρισμός και σφράγιση		8
Μέση	10-15	Καθαρισμός και σφράγιση		5
Μέση	10-15	Καθαρισμός και σφράγιση με ασφαλτόμιγμα τύπου Slurry		3
Υψηλή	<10	Καθαρισμός/σφράγιση και επίστρωση 50mm		10
Παραμορφώσεις - Αυλάκωση				
Σοβαρότητα (mm)	Έκταση (% μήκους)	Τυπικές επεμβάσεις		Εκτιμώμενη διάρκεια ζωής (χρόνια)
<6	>25	Καμία επέμβαση		-
6-12	>25	Ισοπέδωση		7
12-25	>25	Ισοπέδωση		4
>25	>50	Φρεζάρισμα και επίστρωση 50mm		10

Εικόνα 45 : Σχέση τύπου φθοράς οδοστρώματος και εκτιμώμενος χρόνος ζωής με την εφαρμογή διορθωτικών επεμβάσεων

3.3.Κόστος υλοποίησης συστημάτων ανίχνευσης

Τα κόστη υλοποίησης των συστημάτων ανίχνευσης ανωμαλιών οδοστρώματος διαφοροποιούνται ανάλογα με τον σκοπό που επιτελεί το κάθε σύστημα. Επιμερίζονται κυρίως σε χρήση υλικού , ανάπτυξη λογισμικού καθώς και σε ανάπτυξη συμπληρωματικών υπηρεσιών και υποδομών. Το κόστος ενός συστήματος που χρησιμοποιείται αυτόνομα σε ένα αυτοκίνητο για την αποφυγή ανωμαλιών είναι σαφώς μικρότερο από μια συνολική λύση εποπτείας ενός ολόκληρου αυτοκινητοδρόμου. Πέρα από τα κόστη υλοποίησης πρέπει να λαμβάνουμε υπόψιν και τα οφέλη από την χρήση αυτών συστημάτων ανίχνευσης ανωμαλιών^[28] και συνεπακόλουθα τις αποσβέσεις που προκύπτουν επί αυτών των κοστών. Η εποπτεία και διατήρηση σε καλή κατάσταση του οδοστρώματος επεκτείνει και την διάρκεια ζωής του οδοστρώματος αλλά και μειώνει και τις φθορές των οχημάτων που το χρησιμοποιούν.

Σε ένα αυτοκίνητο ένα βασικό σύστημα ανίχνευσης ανωμαλιών οδοστρώματος είναι πιο εύκολο να αναλυθεί και αποτελείται κυρίως από το υλικό (Κάμερες στα απαραίτητα σημεία του οχήματος, σύστημα επεξεργασίας πληροφοριών, οθόνες διεπαφής) και το λογισμικό. Αυτά τα συστήματα στο μέλλον και με εξέλιξη των έξυπνων συστημάτων στα αυτοκίνητα θα μπορούσαν να επεκταθούν με δυνατότητες όπως η επικοινωνία μεταξύ των οχημάτων και η ενημέρωση για ύπαρξη ανωμαλιών προς οχήματα που δεν έχουν προσεγγίσει ακόμα την ανωμαλία στο οδόστρωμα, χρήση εξιδεικευμένων καμερών για δύσκολες συνθήκες οδήγησης βασιζόμενοι σε αντίστοιχες μελέτες^[29] κλπ. Για αυτά τα σενάρια χρήσης η μηχανική όραση θα βοηθούσε για αναγνώριση και αποφυγή ανωμαλιών οδοστρώματος.

Σε μια ολοκληρωμένη υλοποίηση αυτό το κόστος είναι πιο δύσκολο να υπολογιστεί καθώς είναι πιο πολύπλοκο. Μπορεί να περιλαμβάνει πολλά συστήματα που επικοινωνούν μεταξύ τους για την καταγραφή της κατάστασης του οδοστρώματος , ένα κεντρικό σύστημα για την συγκέντρωση των δεδομένων και εξαγωγή συμπερασμάτων. Για αυτά τα σενάρια χρήσης η μηχανική όραση θα βοηθούσε για αναγνώριση και διόρθωση ανωμαλιών οδοστρώματος.

4. Συμπεράσματα – Μελλοντική έρευνα

Από την θεωρητική ανάγνωση και τις διαρκείς εξελίξεις στην μηχανική όραση είδαμε πλέον ότι είναι εφικτή και εφαρμόσιμη η ανίχνευση και η κατηγοριοποίηση διαφόρων τύπων ανωμαλιών οδοστρώματος αλλά και συναφών κατηγοριών αναγνώρισης κατηγοριών ^{[31][32][33][34][35]}. Αυτό επιβεβαιώθηκε και από την εκτέλεση του YOLO framework κατά την εκπόνηση της τρέχουσας εργασίας με της επιτυχή ανίχνευση ανωμαλιών του οδοστρώματος σε πραγματικά δεδομένα. Επίσης η ανίχνευση ανωμαλιών δεν είναι χρήσιμη μόνο στο επίπεδο μια έξυπνης λειτουργίας ενός συστήματος που βελτιώνει την οδηγική εμπειρία ή απλά αποτυπώνει καλύτερα την κατάσταση του οδοστρώματος ενός οδικού δικτύου στο διαχειριστή του, αλλά έχει και τεχνοοικονομικές προεκτάσεις με πολλαπλά οφέλη.

Η ανίχνευση ανωμαλιών μπορεί να βελτιωθεί περαιτέρω με χρήση μοντέλων και μεθόδων που χρησιμοποιούν Semantic Segmentation, όπου η ανάλυση γίνεται εξαντλητικά σε επίπεδο pixel και το αντικείμενο ανιχνεύεται όχι απλά εντός ενός ορθογωνίου που το περικλείει αλλά με το πραγματικό του περίγραμμα^[30]. Βέβαια τέτοια μοντέλα παρόλο που έχουν το πλεονέκτημα ότι ανιχνεύουν αντικείμενα με μεγαλύτερη χωρική ακρίβεια, έχουν μειονέκτημα απαιτούν αρκετή επεξεργαστική ισχύ, σε αντίθεση με τα μοντέλα όπως το YOLO που απαιτούν μικρότερη επεξεργαστική ισχύ αλλά εντοπίζουν λιγότερο ακριβέστερα τις ζητούμενες κλάσεις-αντικείμενα, καθιστώντας τα δυο είδη μοντέλων συμπληρωματικά. Τα μοντέλα όπως το YOLO είναι κατάλληλα για εφαρμογές πραγματικού χρόνου και ανάγκης άμεσης λήψης αποφάσεων και τα μοντέλα Semantic Segmentation σε εφαρμογές εκτέλεσης αναλυτικότερων εργασιών και μεγάλου όγκου δεδομένων και όχι απαραίτητα πραγματικού χρόνου. Παραδείγματα εφαρμογών των πρώτων είναι η ανίχνευση ανωμαλιών σε πραγματικό χρόνο από ένα αυτοκίνητο που κινείται εντός αυτοκινητοδρόμου και των δεύτερων η αναλυτική σκιαγράφηση των ανωμαλιών ενός αυτοκινητοδρόμου από την διαχειρίστρια εταιρία του αυτοκινητοδρόμου για την τεχνοοικονομική ανάλυση και την στοχευμένη δρομολόγηση των απαιτούμενων εργασιών επισκευής χρήση αυτόνομων μηχανημάτων επιδιόρθωσης. Επίσης η ενσωμάτωση όλο και περισσότερο προηγμένων συστημάτων ενεργητικής ασφάλεια στα οχήματα και η θέσπιση νόμων που επιβάλουν την ύπαρξη στο βασικό εξοπλισμό ορισμένων λειτουργιών αυτών, θα βοηθήσει στην περεταίρω έρευνα στον τομέα της μηχανικής όρασης και της ανίχνευσης ανωμαλιών οδοστρώματος.

Βιβλιογραφικές αναφορές

- [1] Mundy, J. L. (2006). Object recognition in the geometric era: A retrospective. In Ponce, J., Hebert, M., Schmid, C., and Zisserman, A. (eds), *Toward Category-Level Object Recognition*, pp. 3–28, Springer, New York.
- [2] Baeza-Yates, R. and Ribeiro-Neto, B. (1999). *Modern Information Retrieval*. Addison Wesley.
- [3] Sivic, J. and Zisserman, A. (2009). Efficient visual search of videos cast as text retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(4):591–606.
- [4] Richard Szeliski (2022) *Computer Vision: Algorithms and Applications 2nd Edition*, Springer
- [5] Csurka, G., Dance, C. R., Fan, L., Willamowski, J., and Bray, C. (2004). Visual categorization with bags of keypoints. In *ECCV International Workshop on Statistical Learning in Computer Vision*.
- [6] Fischler, M. A. and Elschlager, R. A. (1973). The representation and matching of pictorial structures. *IEEE Transactions on Computers*, 22(1):67–92.
- [7] Moghaddam, B. and Pentland, A. (1997). Probabilistic visual learning for object representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):696–710.
- [8] Felzenszwalb, P., McAllester, D., and Ramanan, D. (2008). A discriminatively trained, multiscale, deformable part model. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*.
- [9] Fergus, R. (2007). Combined segmentation and recognition. In *CVPR 2007 Short Course on Recognizing and Learning Object Categories*. <http://people.csail.mit.edu/torralba/shortCourseRLOC/>
- [10] He, K., Gkioxari, G., Dollar, P., and Girshick, R. (2017). Mask R-CNN. In *IEEE International Conference on Computer Vision (ICCV)*.
- [11] Oliva, A. and Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Sciences*, 11(12):520–527.
- [12] Torralba, A., Freeman, W. T., and Fergus, R. (2008). 80 million tiny images: a large dataset for non-parametric object and scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(11):1958–1970.
- [13] Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 1097–1105.
- [14] Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386.
- [15] Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986b). Learning internal representations by error propagation. *Nature*, 323:533–536.
- [16] Ioffe, S. & Szegedy, C.. (2015). Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *Proceedings of the 32nd International Conference on Machine Learning in Proceedings of Machine Learning Research*

- [17] Lecun, Yann & Bottou, Leon & Bengio, Y. & Haffner, Patrick. (1998). Gradient-Based Learning Applied to Document Recognition. Proceedings of the IEEE. 86. 2278 - 2324. 10.1109/5.726791.
- [18] A Neural Network Playground – TensorFlow. Ανακτήθηκε από <http://playground.tensorflow.org/#>
- [19] Redmon, Joseph & Divvala, Santosh & Girshick, Ross & Farhadi, Ali. (2016). You Only Look Once: Unified, Real-Time Object Detection. 779-788. 10.1109/CVPR.2016.91.
- [20] Redmon, Joseph & Farhadi, Ali. (2017). YOLO9000: Better, Faster, Stronger. 6517-6525. 10.1109/CVPR.2017.690.
- [21] Redmon, Joseph & Farhadi, Ali. (2018). YOLOv3: An Incremental Improvement.
- [22] Παναγοπούλου Μαίρη. 2009. Διερεύνηση των Παραγόντων Επιρροής της Εξέλιξης της Κατάστασης του Οδοστρώματος. Πάτρα : Πανεπιστήμιο Πατρών, Τμήμα Πολιτικών Μηχανικών, Εργαστήριο Συγκοινωνιακών Έργων.
- [23] SAE updates J3016 Levels of Automated Driving graphic to reflect evolving standard. Ανακτήθηκε από <https://www.sae.org/news/2019/01/sae-updates-j3016-automated-driving-graphic>
- [24] M. Galvani, “History and future of driver assistance,” IEEE Instrumentation & Measurement Magazine, vol. 22, no. 1, pp. 11–16, 2019
- [25] Καλλιόπη Σπανού (2015). Αστική ευθύνη των Οργανισμών Τοπικής Αυτοδιοίκησης - Η περίπτωση υλικών ζημιών σε οχήματα από ανωμαλίες του οδοστρώματος ή άλλες παρεμφερείς αιτίες. Ανακτήθηκε από <https://www.synigoros.gr/resources/docs/188408.pdf>
- [26] Valuing the Benefits of Highways Maintenance, Ανακτήθηκε από https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/487937/highways-maintenance-benefits-summary.pdf
- [27] Αποστολάκη Λεμονιά (2011). Ανάλυση κόστους κύκλου ζωής οδοστρωμάτων , Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης
- [28] Eftychios Protopapadakis, Iason Katsamenis, and Anastasios Doulamis. 2020. Multi-label deep learning models for continuous monitoring of road infrastructures. In Proceedings of the 13th ACM International Conference on Pervasive Technologies Related to Assistive Environments (PETRA '20). Association for Computing Machinery, New York, NY, USA, Article 66, 1–7. DOI:<https://doi.org/10.1145/3389189.3397997>
- [29] Iason Katsamenis, Eftychios Protopapadakis, Athanasios Voulodimos, Dimitris Dres, and Dimitris Drakoulis. 2020. Man overboard event detection from RGB and thermal imagery: possibilities and limitations. In Proceedings of the 13th ACM International Conference on Pervasive Technologies Related to Assistive Environments (PETRA '20). Association for Computing Machinery, New York, NY, USA, Article 67, 1–6. DOI:<https://doi.org/10.1145/3389189.3397998>
- [30] Katsamenis, Iason & Protopapadakis, Eftychios & Doulamis, Anastasios & Doulamis, Nikolaos & Voulodimos, Athanasios. (2020). Pixel-Level Corrosion Detection on Metal Constructions by Fusion of Deep Learning Semantic and Contour Segmentation. 10.1007/978-3-030-64556-4_13.

- [31] Eftychios Protopapadakis, Athanasios Voulodimos, Anastasios Doulamis, Nikolaos Doulamis, and Tania Stathaki. 2019. Automatic crack detection for tunnel inspection using deep learning and heuristic image post-processing. *Applied Intelligence* 49, 7 (July 2019), 2793–2806. DOI:<https://doi.org/10.1007/s10489-018-01396-y>
- [32] Loupos, Konstantinos & Makantasis, Konstantinos & Protopapadakis, Eftychios & Doulamis, Anastasios & Doulamis, Nikolaos. (2015). Deep Convolutional Neural Networks for efficient vision based tunnel inspection. 10.1109/ICCP.2015.7312681.
- [33] Loupos, Konstantinos & Doulamis, Anastasios & Stentoumis, Christos & Protopapadakis, Eftychios & Makantasis, Konstantinos & Doulamis, Nikolaos & Amditis, Angelos & Chrobocinski, P. & Victores, Juan & Montero, Roberto & Menendez, Elisabeth & Balaguer, Carlos & Lopez-Tarazon, Rafael & Cantero, Miquel & Navarro, Roman & Roncaglia, Alberto & Belsito, Luca & Camarinopoulos, Stephanos & Komodakis, Nikolaos & Singh, Praveer. (2018). Autonomous robotic system for tunnel structural inspection and assessment. *International Journal of Intelligent Robotics and Applications*. 2. 10.1007/s41315-017-0031-9.
- [34] Protopapadakis, Eftychios & Makantasis, Konstantinos & Kopsiaftis, George & Doulamis, Nikolaos & Amditis, Angelos. (2016). Crack Identification Via User Feedback, Convolutional Neural Networks and Laser Scanners for Tunnel Infrastructures. 725-734. 10.5220/0005853007250734.
- [35] Protopapadakis, Eftychios & Voulodimos, Athanasios & Doulamis, Anastasios. (2017). Data sampling for semi-supervised learning in vision-based concrete defect recognition. 1-6. 10.1109/IISA.2017.8316454.