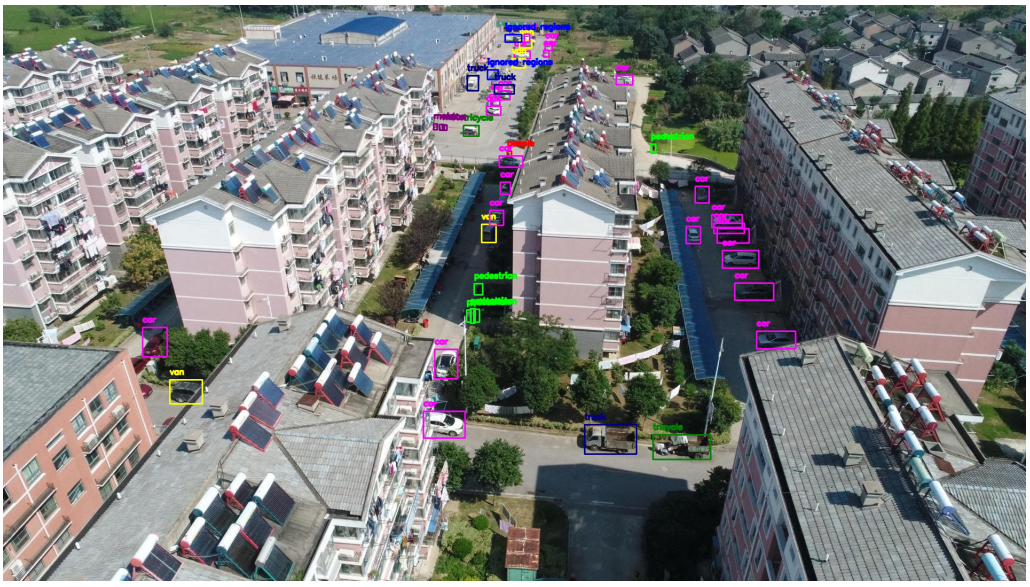
Εθνικο Μετσοβιο Πολυτεχνειο
Σχολη Ηλεκτρολογων Μηχανικων και Μηχανικων Υπολογιστων
Τομεας Τεχνολογιας Πληροφορικης και Υπολογιστων

Εντοπισμός Μικρών Αντικειμένων από Μη Επανδρωμένα Αεροσκάφη

Διπλωματικη Εργασια

του

ΝΤΟΝΤΟΡΟΥ Ε. ΗΛΙΑ



Επιβλέπων: Αθανάσιος Βουλόδημος
Επίκουρος Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2024

Εθνικο Μετσοβιο Πολυτεχνειο
Σχολη Ηλεκτρολογων Μηχανικων και Μηχανικων Υπολογιστων
Τομεας Τεχνολογιας Πληροφορικης και Υπολογιστων

Εντοπισμός Μικρών Αντικειμένων από Μη Επανδρωμένα Αεροσκάφη

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

ΝΤΟΝΤΟΡΟΥ Ε. ΗΛΙΑ

Επιβλέπων: Αθανάσιος Βουλόδημος
Επίκουρος Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 16η Ιουλίου 2024.

(Υπογραφή)	*(Υπογραφή)*	*(Υπογραφή)*
.....	.....	.....
Αθανάσιος Βουλόδημος	Γεώργιος Στάμου	Ανδρέας-Γεώργιος Σταφυλοπάτης
Επίκουρος Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2024

Εθνικο Μετσοβιο Πολυτεχνειο
Σχολη Ηλεκτρολογων Μηχανικων και Μηχανικων Υπολογιστων
Τομεας Τεχνολογιας Πληροφορικης και Υπολογιστων

ΔΗΛΩΣΗ ΜΗ ΛΟΓΟΚΛΟΠΗΣ ΚΑΙ ΑΝΑΛΗΨΗΣ ΠΡΟΣΩΠΙΚΗΣ ΕΥΘΥΝΗΣ

Με πλήρη επίγνωση των συνεπειών του νόμου περί πνευματικών δικαιωμάτων, δηλώνω ενυπογράφως ότι είμαι αποκλειστικός συγγραφέας της παρούσας Πτυχιακής Εργασίας, για την ολοκλήρωση της οποίας κάθε βοήθεια είναι πλήρως αναγνωρισμένη και αναφέρεται λεπτομερώς στην εργασία αυτή. Έχω αναφέρει πλήρως και με σαφείς αναφορές, όλες τις πηγές χρήσης δεδομένων, απόψεων, θέσεων και προτάσεων, ιδεών και λεκτικών αναφορών, είτε κατά κυριολεξία είτε βάσει επιστημονικής παράφρασης. Αναλαμβάνω την προσωπική και ατομική ευθύνη ότι σε περίπτωση αποτυχίας στην υλοποίηση των ανωτέρω δηλωθέντων στοιχείων, είμαι υπόλογος έναντι λογοκλοπής, γεγονός που σημαίνει αποτυχία στην Πτυχιακή μου Εργασία και κατά συνέπεια αποτυχία απόκτησης του Τίτλου Σπουδών, πέραν των λοιπών συνεπειών του νόμου περί πνευματικών δικαιωμάτων. Δηλώνω, συνεπώς, ότι αυτή η Πτυχιακή Εργασία προετοιμάστηκε και ολοκληρώθηκε από εμένα προσωπικά και αποκλειστικά και ότι, αναλαμβάνω πλήρως όλες τις συνέπειες του νόμου στην περίπτωση κατά την οποία αποδειχθεί, διαχρονικά, ότι η εργασία αυτή ή τμήμα της δεν μου ανήκει διότι είναι προϊόν λογοκλοπής άλλης πνευματικής ιδιοκτησίας.

(Υπογραφή)

.....
Ηλίας Ντόντορος
Διπλωματούχος
Ηλεκτρολόγος Μηχανικός
και Μηχανικός
Υπολογιστών ΕΜΠ

16 Ιουλίου 2024

Περίληψη

Η ανίχνευση αντικειμένων είναι μια βασική εργασία στον χώρο της όρασης υπολογιστών, που εμπλέκει τον εντοπισμό και την αναγνώριση αντικειμένων σε μια εικόνα ή ένα καρέ βίντεο. Οι πρόοδοι σε αυτό τον τομέα έχουν προσελκύσει σημαντικό ενδιαφέρον. Με την ανάπτυξη της Τεχνητής Νοημοσύνης και κυρίως των Νευρωνικών Δικτύων έχουν δημιουργηθεί καινούριοι τρόποι για τον εντοπισμό αντικειμένων οι οποίοι είναι πιο αποτελεσματικοί και πιο αποδοτικοί. Οι εφαρμογές που υπάρχουν είναι πολλές και η παρούσα εργασία ασχολείται με τον εντοπισμό μικρών αντικειμένων σε φωτογραφίες τραβηγμένες από μη επανδρωμένα αεροσκάφη.

Στόχος της παρούσας εργασίας είναι η ανάλυση των δυσκολιών που υπάρχουν στην συγκεκριμένη εφαρμογή καθώς και η ανάλυση και η επαλήθευση των διαφόρων συστημάτων που έχουν προταθεί τα τελευταία χρόνια για τον εντοπισμό αντικειμένων. Τα νευρωνικά μοντέλα τα οποία χρησιμοποιήθηκαν, χρησιμοποιούν διαφορετικές τεχνικές για τον εντοπισμό μικρών αντικειμένων το οποίο μας βοηθάει να καταλήξουμε σε ποια τεχνική είναι πιο αποδοτική και ποια έχει μεγάλο ερευνητικό ενδιαφέρον.

Πιο συγκεκριμένα, τα μοντέλα εκπαιδεύτηκαν σε δύο διαφορετικές εφαρμογές οι οποίες έχουν μεγάλο ενδιαφέρον, στον απλό εντοπισμό αντικειμένων και στον προσανατολισμένο εντοπισμό αντικειμένων. Για αυτό τον σκοπό χρησιμοποιήθηκαν δύο ευρέως γνωστά σύνολα δεδομένων, το VisDrone και το DOTAv1.5. Τα μοντέλα που εκπαιδεύσαμε χρησιμοποιούν το καθένα διαφορετικές αρχιτεκτονικές για τον εντοπισμό αντικειμένων όπως για παράδειγμα το Faster-RCNN [1] που έχει αρχιτεκτονική δύο σταδίων (Two Stage Detector), το YOLOv8 [2] που έχει αρχιτεκτονική ενός σταδίου (Single Stage Detector) και αναλύονται και οι μετασχηματιστές που αν και νέα τεχνολογία έχουν πολύ καλά αποτελέσματα. Οι μετρικές πάνω στις οποίες αξιολογήθηκαν τα μοντέλα είναι η μέση ακρίβεια (mean Average Precision). Τέλος, παρατίθενται τα συμπεράσματα της πειραματικής διαδικασίας και αναλύονται πιθανές επεκτάσεις με μεγάλο ερευνητικό ενδιαφέρον.

Λέξεις Κλειδιά

Τεχνητή Νοημοσύνη, Βαθιά Μηχανική Μάθηση, Συνελικτικά Νευρωνικά Δίκτυα, Ανίχνευση Μικρών Αντικειμένων, YOLOv8, Faster-RCNN, ReDet, DETR, VisDrone, DOTAv1.5

Abstract

Object detection is a key task in computer vision, which involves detecting and recognizing objects in an image or video frame. Advances in this area have attracted considerable interest. With the development of Artificial Intelligence and especially Neural Networks new ways have been created to locate objects that are more efficient and faster. The applications that exist are many and the present paper deals with them by detecting small objects in photographs taken by unmanned aerial vehicles.

The purpose of the thesis at hand is to analyze the difficulties involved in this specific application as well as the analysis and verification of the various systems that have been proposed in recent years to locate objects. The neural models use different techniques for detecting small objects which helps us to conclude in which technique is more efficient and which is of great research interest.

More specifically, the models were trained on two different applications which have large interest, in simple object detection and oriented object detection. Two well-known datasets, VisDrone and DOTAv1.5, were used for this purpose. The models that were trained each use different architectures for object detection such as for example Faster-RCNN [1] which has a two-stage architecture (Two Stage Detector), YOLOv8 [2] which has a single stage architecture (Single Stage Detector) and also transformers are explained, which are relatively new, but they have very promising results. The metrics on which the models were evaluated are the mean Average Precision (mAP). Finally, the conclusions of the experiment are listed and possible extensions of great research interest are analyzed.

Keywords

Artificial Intelligence, Deep Learning, Convolutional Neural Networks, Small Object Detection, YOLOv8, Faster-RCNN, ReDet, DETR, VisDrone, DOTAv1.5

στους γονείς μου

Ευχαριστίες

Θα ήθελα καταρχήν να ευχαριστήσω τον καθηγητή κ. Αθανάσιο Βουλόδημο και την κα. Παρασκευή Τζούβελη για την επίβλεψη αυτής της διπλωματικής εργασίας και για την ευκαιρία που μου έδωσαν να την εκπονήσω στο Εργαστήριο Συστημάτων Τεχνητής Νοημοσύνης και Μάθησης. Επίσης ευχαριστώ ιδιαίτερα την Παρασκευή Θεοφίλου και τον Νικόλαο Σπανό για την καθοδήγησή τους και την εξαιρετική συνεργασία που είχαμε. Τέλος θα ήθελα να ευχαριστήσω τους κοντινούς μου ανθρώπους για την καθοδήγηση και την ηθική συμπαράσταση που μου προσέφεραν όλα αυτά τα χρόνια.

Αθήνα, Ιούλιος 2024

Ηλίας Ντόντορος

Περιεχόμενα

Κατάλογος Σχημάτων

Κατάλογος Εικόνων

Κατάλογος Πινάκων

Πρόλογος

Η παρούσα διπλωματική εκπονήθηκε στην Αθήνα, το έτος 2024, στο Εργαστήριο Συστημάτων Τεχνητής Νοημοσύνης και Μηχανικής Μάθησης που ανήκει στον Τομέα Τεχνολογίας Πληροφορικής και Υπολογιστών του Εθνικού Μετσόβιου Πολυτεχνείου, με επιβλέποντες τον Επίκουρο Καθηγητή κ. Αθανάσιο Βουλόδημο και την κ. Παρασκευή Τζούβελη

Κεφάλαιο 1

Εισαγωγή

Τα μη επανδρωμένα αεροσκάφη με την εμφάνιση τους έχουν χρησιμοποιηθεί για ποικίλες εφαρμογές όπως για παράδειγμα στην παρακολούθηση του περιβάλλοντος στην διαχείριση καταστροφών, στην γεωργία και στις καλλιέργειες και γενικότερα στην παρακολούθηση. Η καταγραφή εικόνων από ψηλά έχει δώσει νέες δυνατότητες για την ανάλυση καταστάσεων και την λήψη αποφάσεων. Με την άνοδο της Μηχανικής Μάθησης και του κλάδου της Υπολογιστικής Όρασης έχουν βρεθεί τρόποι να γίνεται αυτόματα εντοπισμός αντικειμένων σε εικόνες το οποίο έχει προκαλέσει μεγάλο ερευνητικό ενδιαφέρον στον κλάδο και συνέχεια υπάρχουν νέες βελτιώσεις και καινούριοι αλγόριθμοι ώστε η διαδικασία να γίνεται πιο αποδοτικά με χρήση λιγότερων υπολογιστικών πόρων και με μεγαλύτερη ακρίβεια.

Ωστόσο, ένα σημαντικό πρόβλημα στην χρήση εικόνων UAV είναι ο αυτόματος εντοπισμός μικρών αντικειμένων όπως οχήματα, πεζοί και άλλοι στόχοι ενδιαφέροντος με την χρήση υπολογιστή, χωρίς να χρειάζεται ανθρώπινη παρέμβαση, καθώς συχνά αυτά τα αντικείμενα εμφανίζονται ως μικρές ομάδες εικονοστοιχείων.

Διάφοροι παράγοντες όπως η χαμηλή ανάλυση των εικόνων, η μερική απόκρυψη των αντικειμένων και οι μεταβαλλόμενες κλίμακες κάνουν τις παραδοσιακές μεθόδους Μηχανικής Μάθησης για την ανίχνευση αντικειμένων σε αεροπορικές εικόνες μη αποτελεσματικές. Παρόλες τις δυσκολίες όμως είναι πολύ σημαντικό να βρεθούν αποδοτικές λύσεις καθώς σε φωτογραφίες οι οποίες είναι τραβηγμένες από ψηλά τα περισσότερα αντικείμενα εμφανίζονται ως πολύ μικρα ασχέτως με τις πραγματικές τους διστάσεις. Τα τελευταία χρόνια η ανάπτυξη της βαθιάς μάθησης και ιδιαίτερα των συνελικτικών νευρωνικών δικτύων έχουν αποφέρει πολύ καλά αποτελέσματα και έχουν προσφέρει καινούριες λύσεις για εργασίες ανίχνευσης αντικειμένων. Τα μοντέλα νευρωνικών δικτύων εκπαιδευμένα σε σύνολα δεδομένων μεγάλης κλίμακας έχουν επιδείξει εξαιρετικές δυνατότητες στην ανίχνευση και τον εντοπισμό αντικειμένων σε διαφορετικά περιβάλλοντα και εφαρμογές.

1.1 Αντικείμενο της διπλωματικής

Στόχος της παρούσας διπλωματικής είναι η ανάλυση και η σύγκριση νευρωνικών μοντέλων που έχουν δημιουργηθεί πάνω στον εντοπισμό μικρών αντικειμένων από εικόνες τραβηγμένες από μη επανδρωμένα αεροσκάφη. Τα μοντέλα τα οποία χρησιμοποιήθηκαν στην εργασία έχουν διαφορετικές αρχιτεκτονικές και έχουν εμφανίσει πολύ καλά αποτελέσματα σε άλλες εφαρμογές. Επιπλέον τα μοντέλα συγκρίθηκαν πάνω σε δύο διαφορετικα σύνο-

λα δεδομένων από τα οποία στο ένα γίνεται προσανατολισμένος εντοπισμός αντικειμένων, δηλαδή με πλαίσιο οριοθέτησης το οποίο δεν είναι απαραίτητα παράλληλο στους άξονες.

1.2 Οργάνωση του τόμου

Η εργασία αυτή είναι οργανωμένη σε 6 κεφάλαια. Στο Κεφάλαιο 1 γίνεται η εισαγωγή της παρούσας εργασίας. Στο Κεφάλαιο 2 γίνεται μια σύντομη ανασκόπηση των Νευρωνικών Δικτύων, της Όρασης Υπολογιστών και του Εντοπισμού Αντικειμένων. Στο Κεφάλαιο 3 παρουσιάζονται λεπτομερώς τα μοντέλα που χρησιμοποιήθηκαν, η αρχιτεκτονική τους και οι βασικές βελτιώσεις και αλλαγές σε σχέση με τα υπόλοιπα. Στο Κεφάλαιο 4 αναλύεται η δομή του πειράματος αναλύονται τα δύο Σύνολα Δεδομένων που χρησιμοποιήθηκαν, το περιβάλλον που εκτελέστηκε το πείραμα και οι μετρικές αξιολόγησης των μοντέλων. Στο Κεφάλαιο 5 παρουσιάζονται τα αποτελέσματα του πειράματος. Τέλος στο Κεφάλαιο 6 δίνονται τα συμπεράσματα από το πείραμα και κάποιες πιθανές μελλοντικές επεκτάσεις.

Μέρος I

Θεωρητικό Μέρος

Κεφάλαιο 2

Νευρωνικά Δίκτυα και Όραση Υπολογιστών

Στο κεφάλαιο αυτό γίνεται μία σύντομη θεωρητική ανασκόπηση των Νευρωνικών Δικτύων, της χρήσης τους στην Όραση Υπολογιστών και αναλύονται κάποιες βασικές έννοιες που χρησιμοποιούνται στην συνέχεια της εργασίας.

2.1 Εισαγωγή

Η Όραση Υπολογιστών είναι ο επιστημονικός κλάδος που έχει ως στόχο την δημιουργία 'έξυπνων' συστημάτων που 'βλέπουν' και αντιλαμβάνονται εικόνες και βίντεο όπως ένας άνθρωπος και μπορούν να εξάγουν συμπεράσματα. Οι εφαρμογές που μπορεί να χρησιμοποιηθεί η Όραση Υπολογιστών ανήκουν σε ένα ευρύ φάσμα και μπορεί να είναι από ιατρική απεικόνιση και ρομποτική όραση μέχρι την αλληλεπίδραση ανθρώπου υπολογιστή και εναέρια παρακολούθηση. Για παράδειγμα όπως παρουσιάζεται και από αυτές τις δύο εργασίες [11] [12] με διάφορες τεχνικές της όρασης υπολογιστών μπορούν να αναγνωριστούν περιοχές ενδιαφέροντος σε ιατρικές απεικονίσεις ή ακόμα και να αναγνωριστούν ασθένειες από ακτινογραφίες χωρίς την βοήθεια κάποιου ανθρώπου. Είναι λογικό, λοιπόν, να έχει συγκεντρώσει μεγάλο ερευνητικό ενδιαφέρον.

Το μεγαλύτερο μέρος των σύγχρονων μεθόδων που χρησιμοποιούνται στην Όραση Υπολογιστών προέρχονται από την μηχανική μάθηση δηλαδή δεν χρησιμοποιούν ρητό προγραμματισμό αλλά γίνεται εκπαίδευση πάνω σε υπάρχοντα δεδομένα και εξάγεται πληροφορία από αυτά. Τα βασικά προβλήματα με τα οποία ασχολείται η μηχανική όραση είναι:

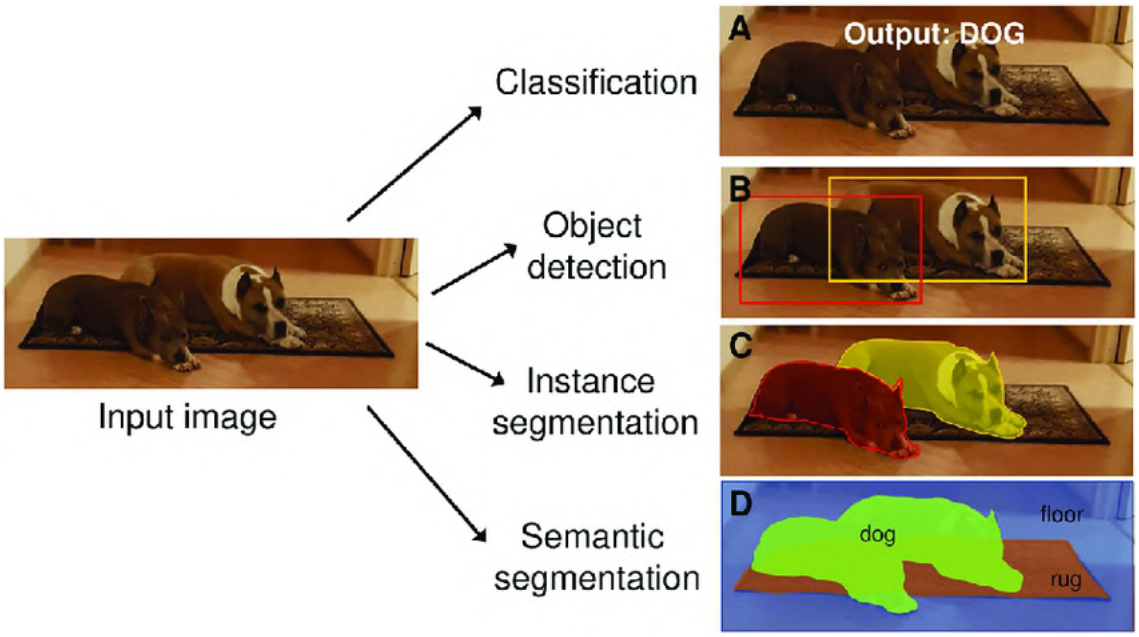
- Ταξινόμηση Εικόνων (Image Classification): δέχεται ως είσοδο μία εικόνα και προβλέπει σε ποια κλάση ανήκει η εικόνα με βάση το περιεχόμενό της. Αφορά ολόκληρη την εικόνα και όχι το κάθε εικονοστοιχείο ξεχωριστά.

- Ανίχνευση Αντικειμένου (Object Detection): δέχεται ως είσοδο μία εικόνα και εντοπίζει την ακριβή θέση διάφορων αντικειμένων που απεικονίζονται σε αυτή καθώς και σε ποια κατηγορία ανήκουν. Όπως και πριν δεν γίνεται εντοπισμός σε επίπεδο εικονοστοιχείων.

- Σημασιολογική Κατάτμηση Εικόνας (Semantic Segmentation): δέχεται ως είσοδο μία εικόνα και προβλέπει την κλάση κάθε εικονοστοιχείου.

- Κατάτμηση Στιγμιοτύπων Εικόνας (Instance Segmentation): δέχεται ως είσοδο μία εικόνα και εντοπίζει τα αντικείμενα που υπάρχουν μέσα σε αυτή (και την κλάση τους)

σε επίπεδο εικονοστοιχείου. Πρόκειται δηλαδή για έναν συνδυασμό της ανίχνευσης αντικειμένου και της σημασιολογικής κατάτμησης.



Σχήμα 2.1: *Αναπαράσταση βασικών προβλημάτων Υπολογιστικής Όρασης [3]*

Στα πλαίσια της παρούσας εργασίας θα εξεταστεί το δεύτερο πρόβλημα. Συγκεκριμένα η ανίχνευση μικρών αντικειμένων σε εικόνες τραβηγμένες από μεγάλο υψόμετρο. Παρακάτω αναλύονται κάποιες γενικές έννοιες σχετικά με τον εντοπισμό αντικειμένων.

2.2 Εισαγωγικές Έννοιες Ανίχνευσης Αντικειμένων

2.2.1 Πλαίσιο Οριοθέτησης (Bounding Box)

Ως πλαίσιο οριοθέτησης ενός αντικειμένου σε μία εικόνα ορίζεται το μικρότερο δυνατό ορθογώνιο τμήμα της εικόνας στο εσωτερικό του οποίου βρίσκεται ολόκληρο το αντικείμενο. Το Πλαίσιο Οριοθέτησης μπορεί να είναι παράλληλο στους άξονες της εικόνας ή προσανατολισμένο με βάση τον προσανατολισμό του αντικειμένου. Στην πρώτη περίπτωση η οποία είναι και πιο συνηθισμένη χρειάζονται 4 τιμές για την πληρη αναπαράσταση του πλαισίου, ενώ στην δεύτερη περίπτωση χρειάζονται τουλάχιστον 4 αλλά συνήθως χρησιμοποιούνται 8.

2.2.2 Είδη Ανιχνευτών

Στην ανίχνευση αντικειμένων, οι αλγόριθμοι διακρίνονται σε ανιχνευτές ενός σταδίου, δύο σταδίων και μετασχηματιστές. Αυτή η διάκριση επηρεάζει την ταχύτητα, την ακρίβεια και την πολυπλοκότητα του αλγορίθμου.

Οι ανιχνευτές ενός σταδίου είναι πιο γρήγοροι, καθώς εκτελούν τη διαδικασία ανίχνευσης απευθείας στην εικόνα χωρίς προηγούμενη διαδικασία πρότασης περιοχών. Είναι απλούστεροι στην υλοποίηση και συνήθως χρησιμοποιούνται για εφαρμογές πραγματικού χρόνου. Ωστόσο, συνήθως παρουσιάζουν χαμηλότερη ακρίβεια.

Οι ανιχνευτές δύο σταδίων παρουσιάζουν υψηλότερη ακρίβεια, καθώς έχουν περισσότερο χρόνο για να εξετάσουν και να προσαρμόσουν τις προτάσεις περιοχών. Είναι πιο πολύπλοκοι στην υλοποίηση και περισσότερο κατάλληλοι για σκηνές με πολλαπλά και περίπλοκα αντικείμενα. Παρουσιάζουν υψηλή απόδοση σε μεγάλες βάσεις δεδομένων, αλλά απαιτούν περισσότερο χρόνο εκτέλεσης.

Οι μετασχηματιστές (Transformers) έχουν εισάγει μια νέα προσέγγιση στην ανίχνευση αντικειμένων, βασισμένοι στην αρχιτεκτονική κωδικοποιητή-αποκωδικοποιητή και προσοχής (attention) [13]. Οι μετασχηματιστές χρησιμοποιούν μηχανισμούς αυτοπροσοχής για να επεξεργαστούν την εικόνα και να εντοπίσουν αντικείμενα, επιτρέποντας την παράλληλη επεξεργασία των δεδομένων και την ανίχνευση σχέσεων μεταξύ τους. Αυτή η προσέγγιση επιτρέπει στους μετασχηματιστές να επιτυγχάνουν υψηλή ακρίβεια και να αντιμετωπίζουν καλύτερα τα μικρά και περίπλοκα αντικείμενα. Παρόλο που οι μετασχηματιστές χρησιμοποιήθηκαν για προβλήματα φυσικής γλώσσας στην αρχή, παρουσιάζουν υψηλές επιδόσεις και στα προβλήματα υπολογιστικής όρασης. Απαιτούν σημαντικά μεγαλύτερη υπολογιστική ισχύ και πόρους για την εκπαίδευση και την εκτέλεση σε σύγκριση με τους παραδοσιακούς ανιχνευτές ενός ή δύο σταδίων και η πολυπλοκότητά τους καθιστά την υλοποίησή τους πιο απαιτητική, αλλά οι δυνατότητές τους για βελτιωμένη ανίχνευση αντικειμένων καθιστούν αυτή την επένδυση πόρων επωφελή σε πολλές εφαρμογές.

2.3 Σύντομη Ανασκόπηση

Καθώς η ανίχνευση αντικειμένων έγινε πολύ δημοφιλής τομέας με την άνοδο των υπολογιστών αναπτύχθηκαν διάφορες τεχνικές. Κάποιες από αυτές που αποτέλεσαν σταθμούς καταγράφονται παρακάτω.

- Viola-Jones [14]: Παρουσιάστηκε το 2001 από τους Viola και Jones. Χρησιμοποιούσε χαρακτηριστικά Haar και μια συνεκτική αρχιτεκτονική για τον εντοπισμό προσώπων. Αν και σχεδιάστηκε αρχικά για τον εντοπισμό προσώπων, αποτέλεσε τη βάση για μελλοντικές μεθόδους ανίχνευσης αντικειμένων.

- Histogram of Oriented Gradients (HOG) [15]: Προτάθηκε από τους Dalal και Triggs το 2005. Χρησιμοποιούσε ιστογράμματα κατανομής κλίσεων σε περιοχές της εικόνας για τον εντοπισμό αντικειμένων. Ήταν δημοφιλής για την αντοχή του σε αλλαγές στο φωτισμό και την εμφάνιση.

- Template Matching [16]: Η συνταγή αυτή αναζητούσε προκαθορισμένα πρότυπα στην εικόνα για τον εντοπισμό αντικειμένων. Είναι μια απλή μέθοδος, αλλά ευαίσθητη σε αλλαγές στη θέση, το μέγεθος και τον προσανατολισμό των αντικειμένων.

Με την ραγδαία ανάπτυξη της μηχανικής μάθησης και κυρίως της βαθιάς μάθησης ξεκίνησαν να αναπτύσσονται και μοντέλα νευρωνικών δικτύων με σκοπό τον εντοπισμό αντικειμένων τα οποία έχουν δώσει πολύ καλά αποτελέσματα. Κάποια από αυτά καταγράφονται παρακάτω:

- R-CNN (Region-Based Convolutional Neural Network) [17]: Η μέθοδος R-CNN αποτέλεσε μια σημαντική επανάσταση στον τομέα της ανίχνευσης αντικειμένων. Η βασική ιδέα ήταν να εφαρμοστεί ένα νευρωνικό δίκτυο σε προτάσεις περιοχών που εξήχθησαν από την εικόνα. Αυτό βελτίωσε την ακρίβεια σε σχέση με παλαιότερες μεθόδους, όπως οι χειροκίνητες χαρακτηριστικές και τα παραδοσιακά νευρωνικά δίκτυα.

- Fast R-CNN [18]: Η εξέλιξη της μεθόδου R-CNN οδήγησε στο Fast R-CNN , το οποίο βελτίωσε την ταχύτητα και την αποτελεσματικότητα. Εισήγαγε τον έννοια της κοινής χρήσης των χαρακτηριστικών συνέλιξης σε ολόκληρη την εικόνα, αντί να εξάγει χαρακτηριστικά για κάθε πρόταση ξεχωριστά.

- Faster R-CNN [1]: Το Faster R-CNN πήγε ένα βήμα παραπέρα εισάγοντας ένα δίκτυο πρότασης περιοχών (Region Proposal Network - RPN), επιταχύνοντας σημαντικά τη διαδικασία της ανίχνευσης αντικειμένων. Αυτή η προσέγγιση έκανε το Faster R-CNN ακόμα πιο αποτελεσματικό. Στο επόμενο κεφάλαιο θα αναλυθεί αναλυτικά η αρχιτεκτονική του Faster-RCNN καθώς είναι ένα από τα μοντέλα που χρησιμοποιήθηκαν στην πειραματική διαδικασία.

- You Only Look Once (YOLO) [19]: Το YOLO προσέφερε μια πλήρως διαφορετική προσέγγιση. Διαιρώντας την εικόνα σε πλέγμα, το YOLO προβλέπει απευθείας τα πλαίσια περιοχών και τις πιθανότητες κλάσης, κάνοντας το ιδιαίτερα αποδοτικό και κατάλληλο για πραγματικό χρόνο.

2.4 Εντοπισμός Μικρών Αντικειμένων

Η ανίχνευση μικρών αντικειμένων είναι μία υποκατηγορία του γενικότερου προβλήματος του εντοπισμού αντικειμένων της υπολογιστικής όρασης η οποία παρουσιάζει επιπρόσθετες δυσκολίες λόγω του μικρού μεγέθους και της χαμηλής ανάλυσης των αντικειμένων. Ιδιαίτερα όταν ο εντοπισμός αντικειμένων γίνεται από φωτογραφίες τραβηγμένες από μεγάλο υψόμετρο, που είναι και το αντικείμενο της παρούσα εργασίας, υπάρχουν επιπλέον δυσκολίες. Παραδοσιακά συστήματα ανίχνευσης αντικειμένων συχνά αποτυγχάνουν να εντοπίσουν μικρά αντικείμενα καθιστώντας αναγκαία την ανάπτυξη εξειδικευμένων τεχνικών για αυτό το πρόβλημα.

Μέχρι στιγμής δεν υπάρχει κάποιος οριστικός ορισμός για το ποια αντικείμενα θεωρούνται μικρά και κάθε σύνολο δεδομένων χρησιμοποιεί ένα δικό του ορισμό. Από αυτούς τους ορισμούς που έχουν αποδοθεί οι πιο δημοφιλείς είναι ότι τα μικρά αντικείμενα καταλαμβάνουν λιγότερα από 1024 pixels ή 32×32 pixels (σε εικόνα μεγέθους 480×480). Πέρα από το μικρό μέγεθος των αντικειμένων και την έλλειψη λεπτομερειών, στον εντοπισμό αντικειμένων σε αεροφωτογραφίες υπάρχουν επιπλέον παράγοντες όπως ο φωτισμός, η ταχύτητα που κινείται το μέσο που τραβάει την φωτογραφία και η ορατότητα λόγω καιρού που προσθέτουν επιπλέον δυσκολία σε αυτή την εφαρμογή και πρέπει να ληφθούν υπόψιν.

Για να αντιμετωπιστούν όλες αυτές οι δυσκολίες έχουν προταθεί διάφορες τεχνικές όπως η υπερανάλυση (super-resolution), η χρήση συμφραζομένων πληροφοριών (context-based information) και η μάθηση πολλαπλών κλιμάκων (multi-scale representation learning). Η

υπερανάλυση στοχεύει στην αύξηση της ανάλυσης της εικόνας, χρησιμοποιώντας τεχνικές όπως τα GANs (Generative Adversarial Networks) [20]. Η χρήση GAN για την αύξηση της ανάλυση των εικόνων έχει χρησιμοποιηθεί και σε άλλες εφαρμογές όπως για παράδειγμα σε αυτή [21] που χρησιμοποιήθηκε για να αυξηθεί η ανάλυση εικόνων από δορυφόρους ώστε να γίνει πιο εύκολος ο εντοπισμός δασών. Στην συγκεκριμένη εργασία το αντικείμενο ήταν η τμηματοποίηση των εικόνων αλλά μπορεί με ακριβώς τον ίδιο τρόπο να βοηθήσει και στον εντοπισμό αντικειμένων. Η χρήση συμφραζομένων πληροφοριών εκμεταλλεύεται το γεγονός ότι τα περισσότερα αντικείμενα συνήθως εμφανίζονται σε συγκεκριμένα περιβάλλοντα και συνυπάρχουν με άλλα σχετικά αντικείμενα. Η μάθηση πολλαπλών κλιμάκων συνεπάγεται την καταγραφή πληροφοριών σε διάφορες κλίμακες, δίνοντας τη δυνατότητα για μια σφαιρική κατανόηση του οπτικού πλαισίου και βελτιώνοντας την ικανότητα του μοντέλου να εντοπίζει μικρά αντικείμενα.

Ένα αντιπροσωπευτικό παράδειγμα της εφαρμογής εντοπισμού μικρών αντικειμένων παρουσιάζεται στην εργασία [22]. Στην εργασία αυτή αναπτύσσεται ένα πλαίσιο βαθιάς μάθησης για την ανίχνευση, παρακολούθηση και εκτίμηση αποστάσεων μη συνεργατικών εναέριων οχημάτων χρησιμοποιώντας οπτικούς αισθητήρες. Η συγκεκριμένη έρευνα εστιάζει στην πλήρως αυτόνομη πτήση και στην αποφυγή εναέριων συγκρούσεων μέσω του εντοπισμού αντικειμένων και εκτίμησης της απόστασης τους. Η εργασία αυτή αποτελεί ένα εξαιρετικό παράδειγμα του πώς οι τεχνικές ανίχνευσης μικρών αντικειμένων μπορούν να εφαρμοστούν σε πρακτικά προβλήματα, αντιμετωπίζοντας τις προκλήσεις του μικρού μεγέθους, της χαμηλής ανάλυσης και των περίπλοκων περιβαλλοντικών συνθηκών

Συνολικά, η ανίχνευση μικρών αντικειμένων αποτελεί ένα πεδίο με πολλές δυσκολίες αλλά και με μεγάλο ερευνητικό ενδιαφέρον. Οι εφαρμογές που μπορεί να σκεφτεί κανείς είναι πολλές και για αυτό υπάρχουν πολλές βελτιώσεις στον τομέα αλλά ακόμα υπάρχει μεγάλο περιθώριο ανάπτυξης.

Κεφάλαιο 3

Θεωρητική Ανάλυση Νευρωνικών Μοντέλων

Σε αυτό το κεφάλαιο παρουσιάζονται οι αρχιτεκτονικές των νευρωνικών δικτύων που χρησιμοποιήθηκαν σε αυτή την εργασία.

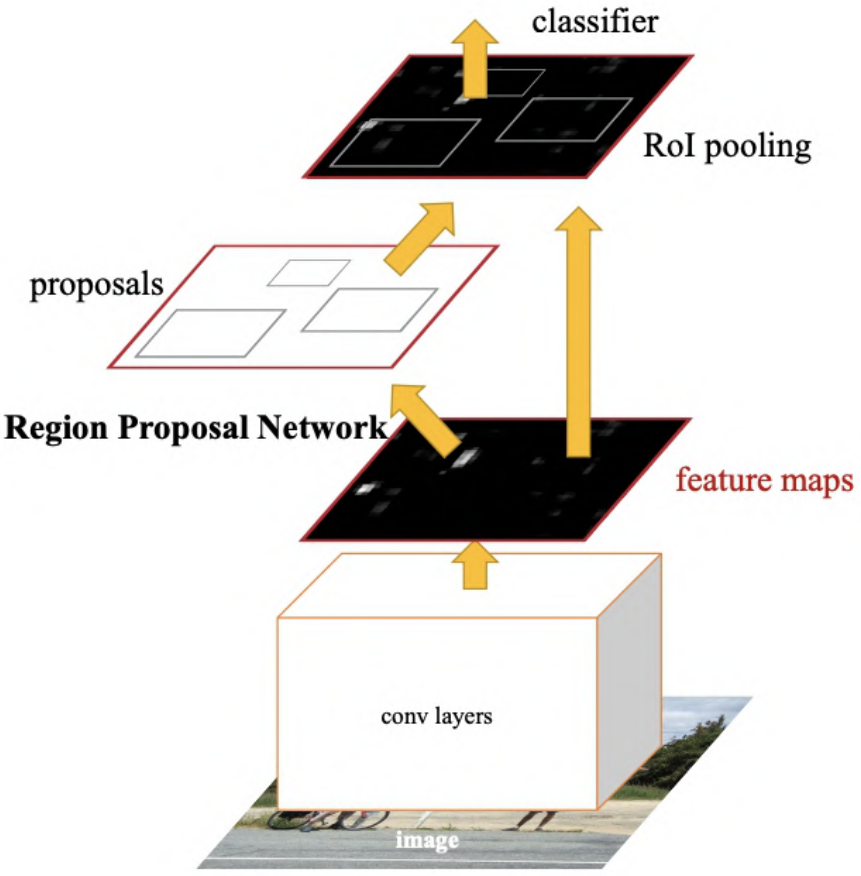
3.1 Faster-RCNN

Το Faster R-CNN (Region Convolutional Neural Network) [1] είναι ένας από τους πιο δημοφιλείς και αποδοτικούς αλγορίθμους ανίχνευσης αντικειμένων δύο σταδίων. Η κύρια ιδέα πίσω από το Faster R-CNN είναι η ενσωμάτωση ενός δικτύου πρότασης περιοχών (Region Proposal Network, RPN) απευθείας στο CNN, επιτρέποντας την ταχύτερη και ακριβέστερη ανίχνευση αντικειμένων.

Η αρχιτεκτονική του Faster R-CNN 3.1 αποτελείται από τα ακόλουθα κύρια μέρη: το πρώτο μέρος του μοντέλου στο οποίο εισάγεται η εικόνα είναι ένα βαθύ συνελικτικό νευρωνικό δίκτυο (CNN), όπως το VGG16 [23] ή το ResNet [24], το οποίο χρησιμοποιείται για την εξαγωγή χαρακτηριστικών της εικόνας. Τα χαρακτηριστικά αυτά είναι οι ενεργοποιήσεις από τα τελευταία στρώματα του πρώτου σταδίου. Το RPN είναι ένα μικρό δίκτυο που διαχειρίζεται την πρόταση περιοχών ενδιαφέροντος (RoIs). Το RPN γλιστράει ένα μικρό παράθυρο πάνω στον χάρτη χαρακτηριστικών από την έξοδο του CNN backbone και για κάθε θέση, προβλέπει εάν υπάρχει ένα αντικείμενο και τις συντεταγμένες των προτεινόμενων περιοχών. Το RPN χρησιμοποιεί άγκυρες (anchors) σε διαφορετικές κλίμακες και αναλογίες για να προβλέψει περιοχές ενδιαφέροντος.

Μετά την πρόταση των περιοχών από το RPN, οι προτεινόμενες περιοχές (RoIs) περνούν από μία διαδικασία που ονομάζεται RoI Pooling [17]. Αυτή η διαδικασία χαρτογραφεί τις περιοχές στον χάρτη χαρακτηριστικών και εξάγει σταθερού μεγέθους χαρακτηριστικά για κάθε RoI. Το RoI Pooling εξασφαλίζει ότι οι προτεινόμενες περιοχές έχουν σταθερό μέγεθος για να μπορούν να εισαχθούν στα επόμενα στρώματα του δικτύου. Τα χαρακτηριστικά που εξάγονται από το RoI Pooling εισάγονται σε δύο ξεχωριστά δίκτυα: ένα δίκτυο για την ταξινόμηση, το οποίο προβλέπει την κατηγορία του αντικειμένου για κάθε περιοχή ενδιαφέροντος και ένα δίκτυο για την εκτίμηση των συντεταγμένων του πλαισίου οριοθέτησης του αντικειμένου (bounding box regression).

Το Faster R-CNN είναι γνωστό για την υψηλή του ακρίβεια στην ανίχνευση αντικειμένων, καθώς εκμεταλλεύεται τις χωρικές πληροφορίες που εξάγονται από το backbone και την αποτελεσματική πρόταση περιοχών. Παρόλο που είναι πιο αργό από τους ανιχνευτές ενός



Σχήμα 3.1: *Αρχιτεκτονική Faster-RCNN [1]*

σταδίου όπως το YOLO [19], είναι πολύ πιο γρήγορο από τους παραδοσιακούς αλγόριθμους δύο σταδίων, λόγω της ενσωμάτωσης του RPN στο CNN. Μπορεί να ανιχνεύσει αντικείμενα σε διάφορες κλίμακες και αναλογίες, καθιστώντας το κατάλληλο για σκηνές με πολλαπλά και μικρά αντικείμενα. Ωστόσο, η εκπαίδευση και η εκτέλεση του Faster R-CNN απαιτεί σημαντικούς υπολογιστικούς πόρους, ιδιαίτερα όταν χρησιμοποιούνται βαθιά δίκτυα για το backbone. Παρότι το Faster R-CNN είναι πιο γρήγορο από άλλους αλγόριθμους δύο σταδίων, μπορεί να μην είναι κατάλληλο για εφαρμογές πραγματικού χρόνου σε σύγκριση με τους ανιχνευτές ενός σταδίου. Παρά τα μειονεκτήματα αυτά, το Faster R-CNN αποτελεί ένα από τα πιο ισχυρά εργαλεία για την ανίχνευση αντικειμένων.

3.2 Μοντέλα YOLO

Τα μοντέλα YOLO (You Only Look Once) είναι μια σειρά από εξαιρετικά αποτελεσματικά μοντέλα ανίχνευσης αντικειμένων που έχουν μεταμορφώσει το πεδίο της υπολογιστικής όρασης. Η κύρια φιλοσοφία πίσω από τα μοντέλα YOLO είναι η ταυτόχρονη ανίχνευση και ταξινόμηση αντικειμένων σε μία εικόνα με μία μόνο διαδικασία (One Stage Detectors), καθιστώντας τα εξαιρετικά γρήγορα σε σύγκριση με τα παραδοσιακά μοντέλα ανίχνευσης δύο σταδίων. Η βασική διαδικασία που ακολουθούν τα μοντέλα YOLO είναι η διαίρεση της ει-

κόνας σε ένα πλέγμα κελιών και για το κάθε κελί γίνονται προβλέψεις για το αν υπάρχει κάποιο αντικείμενο μέσα σε αυτό. Έχουν δημιουργηθεί 10 βασικές εκδόσεις YOLO (v1 - v10) αλλά απο αυτές τις 10 βασικές εκδόσεις έχουν βγει πολλές παραλλαγές, όπως για παράδειγμα το HIC-YOLOv5 [4] που είναι μία παραλλαγή του YOLOv5 [25]. Η παρουσα εργασία θα ασχοληθεί με τα YOLOv8 [2], YOLOv9 [2], YOLOv10 [6] και HIC-YOLOv5, αλλά για να γίνουν πιο κατανοητές οι αρχιτεκτονικές αυτών των μοντέλων θα αναλυθεί και η αρχιτεκτονική του απλού YOLOv5 πάνω στο οποίο είναι βασισμένα τα υπόλοιπα μοντέλα.

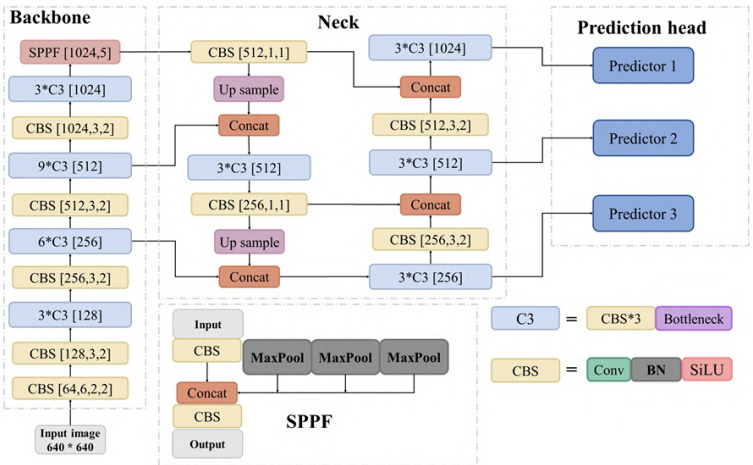
3.2.1 YOLOv5

Το YOLOv5 δημιουργήθηκε από την Ultralytics και είναι σχεδιασμένα για να προσφέρει υψηλή ακρίβεια και ταχύτητα. Αποτελείται από τρία μέρη: το Backbone, το Neck και το Head.

Το Backbone είναι υπεύθυνο για την εξαγωγή των βασικών χαρακτηριστικών από την εικόνα εισόδου. Χρησιμοποιεί το CSPDarknet53 μία παραλλαγή του Darknet53 [26]που ενσωματώνει το Cross Stage Partial Network (CSPNet) [27] για να βελτιώσει την ροή της πληροφορίας και να μειώσει την πολυπλοκότητα των υπολογισμών διαιρώντας τον χάρτη χαρακτηριστικών σε δύο μέρη και συγχωνεύοντας τα δύο αυτά μέρη στο τέλος κάθε μπλοκ.

Το Neck είναι υπεύθυνο να συγκεντρώνει τα χαρακτηριστικά από διαφορετικά επίπεδα του Backbone και να τα συνδυάζει για την τελική πρόβλεψη. Χρησιμοποιεί το PANet (Path Aggregation Network) [28] το οποίο εξάγει πλούσια χαρακτηριστικά από διάφορα επίπεδα και βοηθάει την ροή της πληροφορίας επιτρέποντας έτσι την καλύτερη ανίχνευση αντικειμένων σε διάφορες κλίμακες.

Το Head είναι υπεύθυνο για τις τελικές προβλέψεις και περιλαμβάνει τρεις διαστάσεις ανίχνευσης που η κάθε μία είναι υπευθυνη για ένα διαφορετικό μεγεθος αντικειμένων (μικρά, μεσαία, μεγάλα). Κάθε μια από αυτές είναι υπεύθυνη για την πρόβλεψη των πλαισίων οριοθέτησης, των κλάσεων και των επιπέδων εμπιστοσύνης και χρησιμοποιεί ένα σύνολο από συνελικτικά επίπεδα για αυτό τον σκοπό.



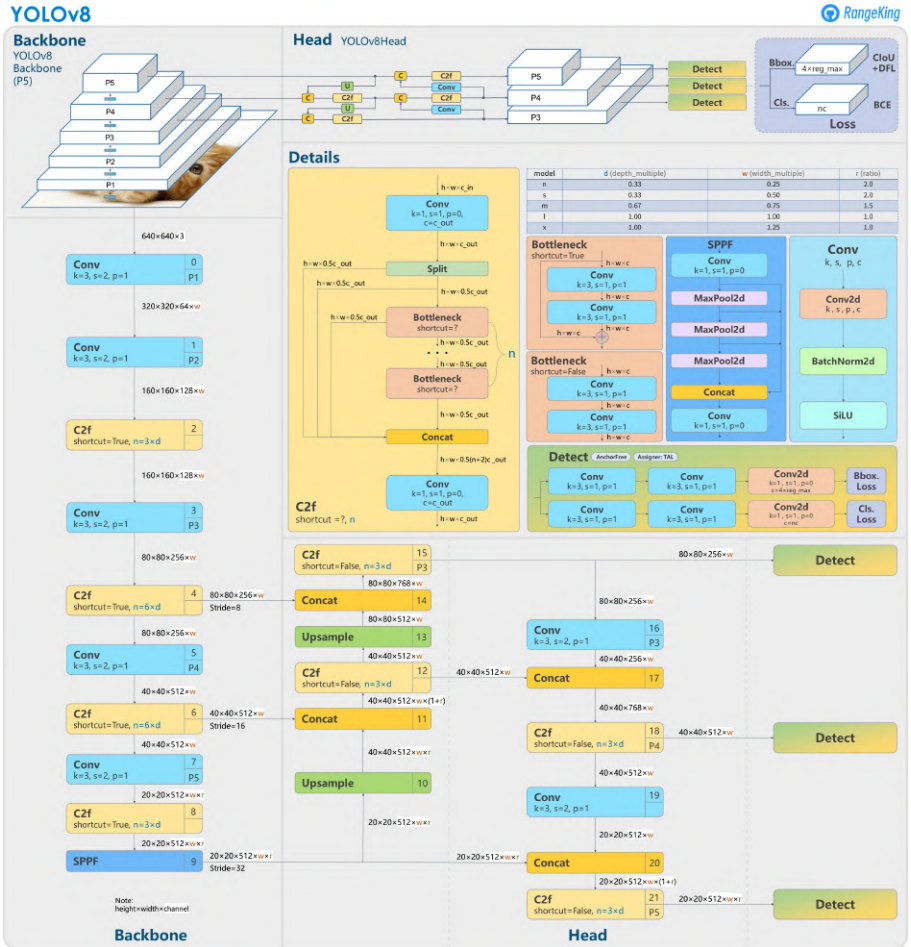
Σχήμα 3.2: *Αρχιτεκτονική YOLOv5 [4]*

Το YOLOv5 χρησιμοποιεί άγκυρες (anchors) οι οποίες είναι προσαρμοσμένες να καλύπτουν ένα μεγάλο εύρος διαστάσεων αντικειμένων και ακόμη χρησιμοποιεί και την τεχνική AutoAnchor, η οποία προσαρμόζει τις άγκυρες στο εκάστοτε σύνολο δεδομένων βελτιώνοντας έτσι την επίδοση του. Γενικότερα το YOLOv5 αποτέλεσε σταθμός στους One Stage Detectors λόγω της επίδοσής του και αποτέλεσε βάση για πολλές βελτιωμένες εκδόσεις, όπως για παράδειγμα το HIC-YOLOv5.

3.2.2 YOLOv8

Το YOLOv8 [2] είναι βασισμένο στο YOLOv5 αλλά έχει καποιες αλλαγές που φέρνουν πολύ καλά αποτελέσματα στην επίδοσή του χωρίς να αυξάνουν το υπολογιστικό κόστος.

Για Backbone χρησιμοποιεί παρόμοια αρχιτεκτονική με το YOLOv5 με κάποιες αλλαγές στο CSP που πλέον ονομάζεται C2f module και συνδυάζει χαρακτηριστικά υψηλού επιπέδου με πληροφορίες περιβάλλοντος. Επίσης το YOLOv8 χρησιμοποιεί ένα μοντέλο χωρίς άγκυρες με ένα αποσυνδεδεμένο head για την ανεξάρτητη επεξεργασία της ύπαρξης αντικειμένων, της ταξινόμησης και της παλινδρόμησης. Αυτός ο σχεδιασμός επιτρέπει σε κάθε branch να εστιάσει στο δικό του task, βελτιώνοντας τη συνολική ακρίβεια του μοντέλου.



Σχήμα 3.3: *Αρχιτεκτονική YOLOv8 (πηγή)*

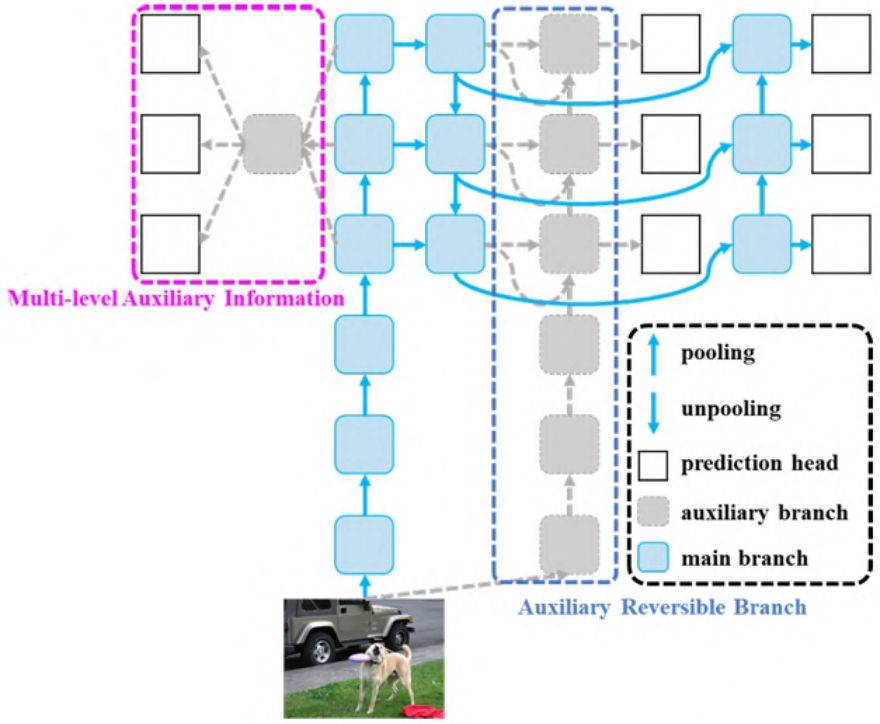
Τέλος η ομάδα που δημιούργησε το YOLOv8 (Ultralytics) έχει φτιάξει τις απαραίτητες τροποποιήσεις ώστε να μπορεί να χρησιμοποιηθεί το μοντέλο και για αλλά προβλήματα

υπολογιστικής όρασης, όπως για παράδειγμα η σημασιολογική κατάτμηση και ο προσανατολισμένος εντοπισμός αντικειμένων κάτι το οποίο θα χρησιμοποιηθεί στην παρούσα εργασία στην σύγκριση που έγινε με το σύνολο δεδομένων DOTAv1.5 [10]

3.2.3 YOLOv9

Το YOLOv9 [5] είναι πολύ πρόσφατο μοντέλο της σειράς YOLO το οποίο φέρνει νέες ιδέες για να αντιμετωπιστούν κάποια προβλήματα που υπάρχουν στις τεχνικές εντοπισμού αντικειμένου. Είναι βασισμένο στην αρχιτεκτονική των προκάτοχών του και το κύριο πρόβλημα που προσπαθεί να αντιμετωπίσει το YOLOv9 είναι η απώλεια πληροφορίας στα βαθιά νευρωνικά δίκτυα και για αυτό προτείνονται δυό νέα στοιχεία: το Programmable Gradient Information (PGI) και το Generalized Efficient Layer Aggregation Network (GELAN).

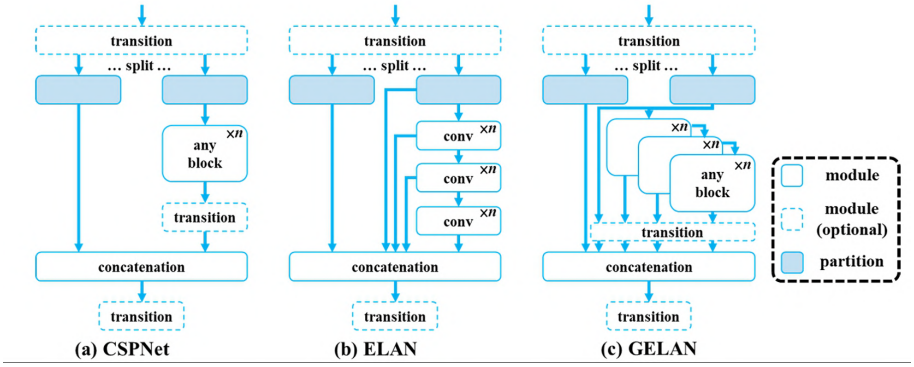
Το PGI είναι μία ιδέα για την βελτιστοποίηση της διαδικασίας εκπαίδευσης των βαθιών νευρωνικών μοντέλων και για την μείωση της απώλειας πληροφορίας. Δημιουργεί αξιόπιστες κλίσεις μέσω ενός βοηθητικού αντιστρέψιμου κλάδου που διατηρεί τις βασικές πληροφορίες των χαρακτηριστικών σε όλα τα στρώματα του δικτύου και έτσι η εκπαίδευση του μοντέλου γίνεται πιο αποδοτική καθώς τα χαρακτηριστικά των αντικειμένων δεν χάνονται στα πιο βαθιά επίπεδα. Η αρχιτεκτονική του φαίνεται στην εικόνα 3.4



Σχήμα 3.4: *Αρχιτεκτονική Programmable Gradient Information (PGI) [5]*

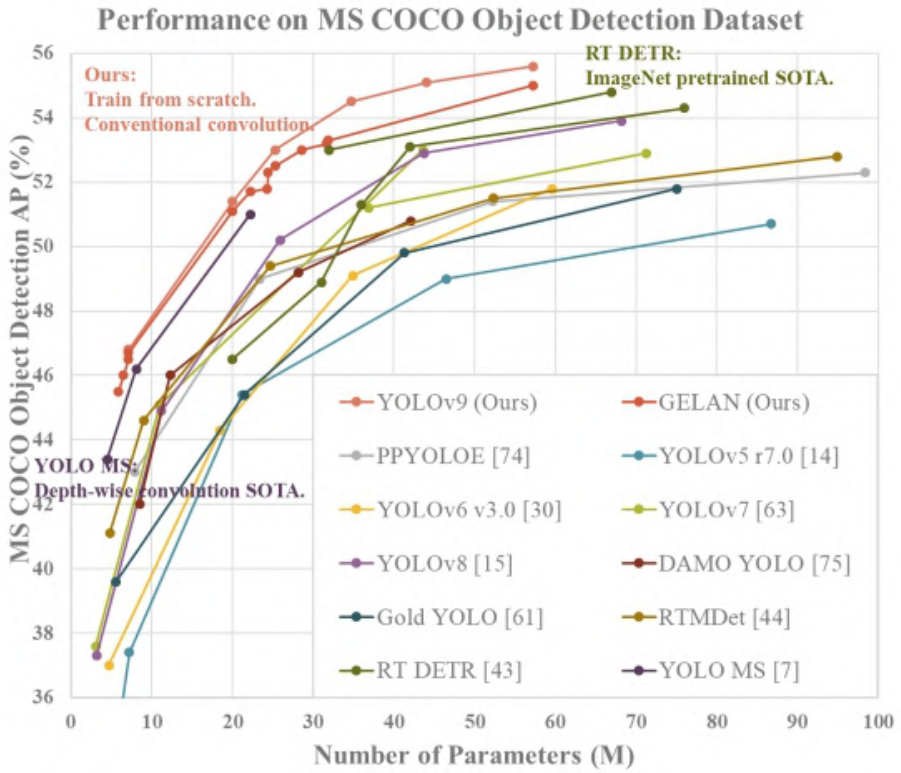
Το GELAN είναι βασισμένο στο ELAN [29] και στο CSPNet [27] και στην ουσία έχει την αρχική αρχιτεκτονική του ELAN αλλά εκεί που υπάρχουν συνελίξεις στο ELAN, το GELAN μπορεί να έχει οποιοδήποτε άλλο block και στην προκειμένη περίπτωση έχουν επιλέξει να βάλουν το CSPNet, όπως φαίνεται και στην εικόνα 3.5.

Με αυτά τα δύο νέα χαρακτηριστικά το YOLOv9 έχει καταφέρει να έχει πολύ καλά αποτελέσματα. Οι παράμετροι του έχουν μειωθεί σε σχέση με το YOLOv8 και μπορεί να ανιχνεύει



Σχήμα 3.5: *Αρχιτεκτονική Generalized Efficient Layer Aggregation Network (GELAN) [5]*

αντικείμενα σε πραγματικό χρόνο, παρόλο όμως που έχει μειωθεί το μέγεθος του μοντέλου έχει βελτιωθεί η ακρίβεια του ξεπερνώντας τον προκάτοχό του.



Σχήμα 3.6: *Αριθμός παραμέτρων και απόδοση στο MS COCO [5]*

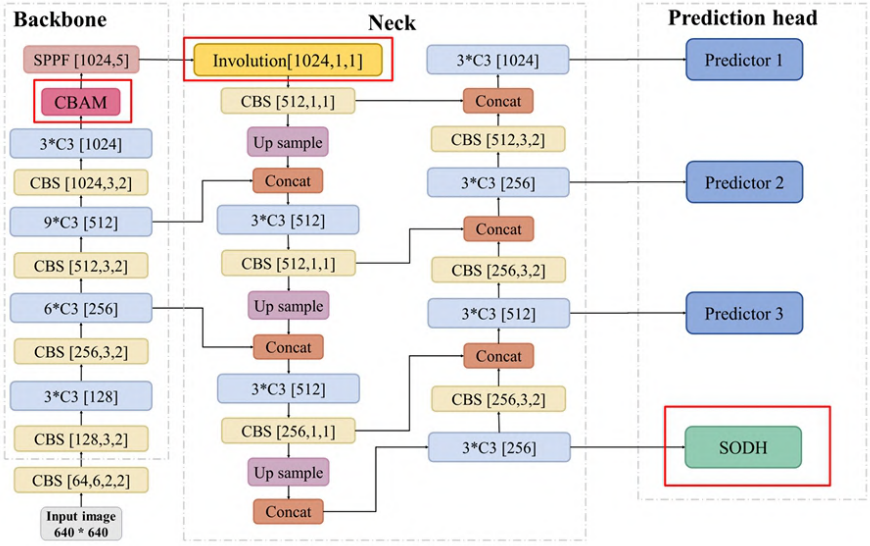
3.2.4 HIC-YOLOv5

Το HIC-YOLOv5 [4] είναι ένα μοντέλο βασισμένο στο πολύ δημοφιλές YOLOv5 (η αρχιτεκτονική του οποίου φαίνεται στο Σχήμα 3.2) και προτείνει κάποιες βελτιώσεις για την αύξηση της ακρίβειας και την μείωση του υπολογιστικού κόστους στην ανίχνευση μικρών αντικειμένων. Για αυτό τον σκοπό περιλαμβάνει τρεις βελτιώσεις στην αρχιτεκτονική του οι οποίες αναλύονται στην συνέχεια.

Αρχικά, έχει προστεθεί μία επιπλέον κεφαλή πρόβλεψης η οποία είναι ειδικά σχεδια-

σμένη να εντοπίζει μικρά αντικείμενα μικρών διαστάσεων, χρησιμοποιώντας χαρακτηριστικά υψηλής ανάλυσης τα οποία κάνουν τον εντοπισμό των μικρών αντικειμένων πιο εύκολο καθώς είναι πιο ευδιάκριτα. Δεύτερον, χρησιμοποιείται ένα involution block ανάμεσα στο backbone και στο neck αυξάνοντας την πληροφορία που διατηρείται στα κανάλια βελτιώνοντας την συνολική απόδοση του μοντέλου. Τέλος, εφαρμόζεται το Convolutional Block Attention Module (CBAM) στο τέλος του backbone το οποίο δίνει έμφαση στις πληροφορίες του καναλιού και στις χωρικές πληροφορίες.

Το CBAM αποτελείται από δύο blocks: το Channel Attention Module και το Spatial Attention Module. Αυτά τα δύο blocks παράγουν αντίστοιχα ένα χάρτη προσοχής καναλιού και ένα χωρικό χάρτη προσοχής οι οποίοι πολλαπλασιάζονται με τον αρχικό χάρτη χαρακτηριστικών για να διευκολύνουν την βελτίωση των χαρακτηριστικών. Η εισαγωγή του CBAM στο backbone και όχι στο neck όπως σε προηγούμενες εργασίες, μειώνει σημαντικά το υπολογιστικό κόστος καθώς οι χάρτες χαρακτηριστικών έχουν μικρότερες διαστάσεις.



Σχήμα 3.7: *Αρχιτεκτονική HIC-YOLOv5 [4]*

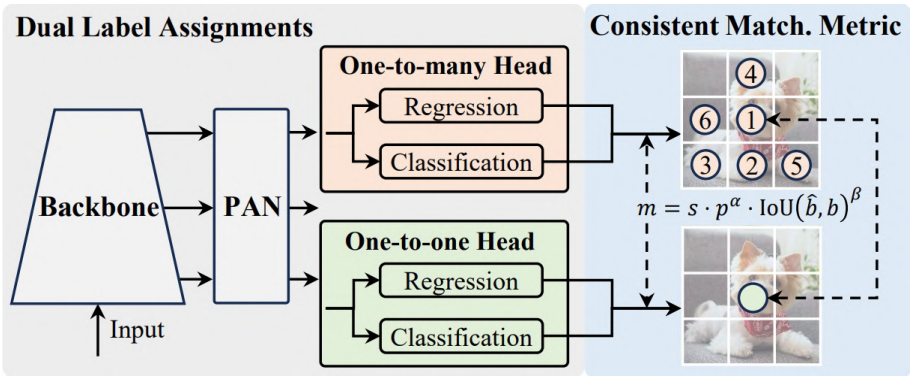
Με αυτές τις βελτιώσεις το HIC-YOLOv5 καταφέρνει να έχει πολύ καλα αποτελέσματα σε εφαρμογές ανίχνευσης μικρών αντικειμένων χωρίς να αυξάνεται το υπολογιστικό κόστος.

3.2.5 YOLOv10

Το YOLOv10 [6] είναι η πιο πρόσφατη έκδοση της οικογένειας YOLO και εισάγει πολλές καινοτομίες και βελτιώσεις στην αρχιτεκτονική. Έχει δημιουργηθεί με σκοπό τον εντοπισμό αντικειμένων σε πραγματικό χρόνο αλλά χωρίς να υπάρχει μείωση στην ακρίβεια των προβλέψεων. Επίσης με τις αλλαγές που έχουν γίνει στην αρχιτεκτονική έχει επιτευχθεί σημαντική μείωση στον αριθμό των παραμέτρων κάνοντας το μοντέλο πιο ελαφρύ και γρήγορο.

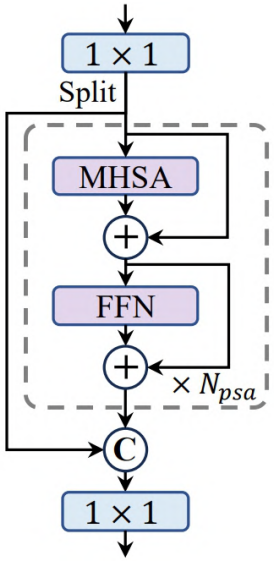
Η πιο σημαντική καινοτομία που έχει φέρει το YOLOv10 είναι η κατάργηση της καταστολής μη μεγίστου/Non-Maximum Suppression (NMS) που στα υπόλοιπα μοντέλα χρησιμοποιείται γιατί κάθε κεφαλή εντοπισμού παράγει πολλαπλές προβλέψεις για ένα αντικείμενο και πρέπει να όλες αυτές οι προβλέψεις να συνδυαστούν ώστε να υπάρξει μία και μονα-

δική πρόβλεψή για κάθε αντικείμενο με όσο μεγαλύτερη ακρίβεια γίνεται. Το YOLOv10 χρησιμοποιεί δύο διαδικασίες για την πρόβλεψη αντικειμένων, μία κατά την διάρκεια της εκπαίδευσης που είναι ίδια με τα υπόλοιπα μοντέλα και παράγει πολλαπλές προβλέψεις για ένα αντικείμενο και μία κατά της διάρκεια της ανάλυσης (inference) που παράγει μόνο μία πρόβλεψη για κάθε αντικείμενο. Η πρώτη επιτρέπει στο μοντέλο να συλλέγει περισσότερες πληροφορίες και να μαθαίνει πιο αποτελεσματικά το εκάστοτε σύνολο δεδομένων και η δεύτερη μειώνει την πολυπλοκότητα του μοντέλου αυξάνοντας την ταχύτητα του. Για να γεφυρωθεί το χάσμα ανάμεσα στις δύο διαδικασίες το YOLOv10 χρησιμοποιεί μια μετρική αντιστοίχισης που εξασφαλίζει οτι η διαδικασία της πρόβλεψης κατα την εκπαίδευση και κατα την ανάλυση αποφέρουν όσο το δυνατόν ίδια αποτελέσματα.



Σχήμα 3.8: *Αρχιτεκτονική του YOLOv10 [6]*

Επιπλέον ολόκληρος ο σχεδιασμός του YOLOv10 έχει γίνει με βάση την ακρίβεια και την απόδοση και ενσωματώνει αρκετές καινοτομίες, όπως μία πιο ελαφριά κεφαλή ταξινόμησης, χρήση μεγαλύτερων πυρήνων συνέλιξης και μία μερική αυτοπροσοχή εμπνευσμένη από τους μετασχηματιστές που έχει υλοποιηθεί από το block που φαίνεται στο σχήμα 3.9.



Σχήμα 3.9: *Αρχιτεκτονική του block Αυτοπροσοχής στο YOLOv10[6]*

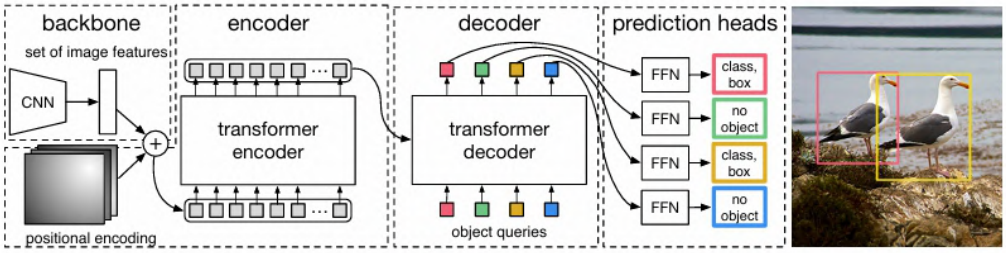
Συνολικά το YOLOv10 έφερε σημαντικές βελτιώσεις στην αρχιτεκτονική των YOLO και έχει καταφέρει να μειώσει τις παραμέτρους και την πολυπλοκότητα του μοντέλου κρατώντας

την ακρίβεια και την απόδοση σε πολύ υψηλά επίπεδα. Βέβαια όπως αναφέρει και η ίδια η ομάδα που το δημιούργησε έχει κάποιους περιορισμούς, όπως για παράδειγμα ότι παραμένει ένα μικρό χάσμα στην απόδοση της προσέγγισης χωρίς NMS σε σχέση με την παραδοσιακή προσέγγιση.

3.3 DETR

Ο DETR (Detection Transformers) [30] είναι από τις πρώτες προσπάθειες να μπει η αρχιτεκτονική των μετασχηματιστών (transformers) [13] στην όραση υπολογιστών μετά την τεράστια επιτυχία που είχε στα Large Language Models. Αυτή η προσέγγιση απομακρύνεται από την παραδοσιακή χρήση των CNNs και χρησιμοποιεί transformers για τον εντοπισμό της θέσης και την κατηγορία των αντικειμένων μεσα στην εικόνα.

Ο DETR αποτελείται από δύο κύρια μέρη όπως φαίνεται και στο σχήμα 3.10, ένα CNN Backbone και ένα μετασχηματιστή. Το πρώτο μέρος χρησιμοποιείται για την εξαγωγή χαρακτηριστικών από την εικόνα και αυτά τα χαρακτηριστικά εισέρχονται στο δεύτερο μέρος, τον μετασχηματιστή, ο οποίος αποτελείται απο μια ακολουθία επιπέδων encoder και decoder. Ο encoder αναλαμβάνει να μετασχηματίσει ταν χαρακτηριστικά απο το backbone σε μία αναπαράσταση υψηλότερων διαστάσεων και ο decoder χρησιμοποιεί αυτές τις αναπαραστάσεις για να παράγει προβλέψεις για την θέση και την κατηγορία των αντικειμένων.



Σχήμα 3.10: *Αρχιτεκτονική DETR*

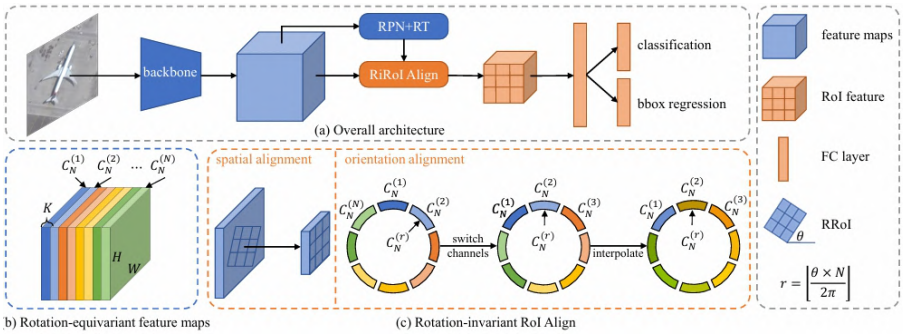
Ο DETR χρησιμοποιεί δύο καινοτομίες: τα object queries και την bipartite matching loss. Τα object queries είναι ενσωματώσεις που επιτρέπουν στον μετασχηματιστή να εστιάζει σε συγκεκριμένες περιοχές για την ανίχνευση αντικειμένων και κατά την διαδικασία του decoding κάθε object querie προσπαθεί να ανιχνεύσει ένα αντικείμενο στην εικόνα. Η bipartite matching loss χρησιμοποιεί ένα συνδυασμό των Hungarian loss και L1 loss για να συγκρίνει τις προβλέψεις με το ground truth αντί για παραδοσιακούς τρόπους που βασίζονται στην απόσταση, όπως η IoU loss.

Η κύρια συνεισφορά του DETR είναι η απλότητα της αρχιτεκτονικής του και η ικανότητα του να ανιχνεύει αντικείμενα χωρίς την χρήση προκαθορισμένων anchors όπως είναι συνηθισμένο. Παρόλες τις καινοτομίες που έφερε ο DETR, οι επιδόσεις του δεν ήταν πολύ καλές και υπήρξαν πολλές προσπάθειες να γίνουν μικρές αλλαγές για να βελτιωθεί.

3.4 ReDet

Ο ReDet (Rotation-equivariant Detector) [7] σχεδιάστηκε ειδικά για την ανίχνευση αντικειμένων από αεροφωτογραφίες, αυτό όμως που τον κάνει να ξεχωρίζει από τους υπόλοιπους ανιχνευτές είναι η χρήση περιστρεφόμενων πλαισίων οριοθέτησης. Για να το καταφέρει αυτό αποτελεσματικά χρησιμοποιεί δίκτυα τα οποία είναι ισοδύναμα στις περιστροφές των αντικειμένων.

Στις φωτογραφίες οι οποίες είναι τραβηγμένες από μεγάλο υψόμετρο τα αντικείμενα έχουν μικρές διαστάσεις και αυθαίρετο προσανατολισμό, για αυτό τον λόγο είναι πολύ αποτελεσματικό τα μοντέλα να μαθαίνουν τα χαρακτηριστικά των αντικειμένων χωρίς να εξαρτιούνται απο τον προσανατολισμό που έχει το αντικείμενο στην εικόνα. Ο ReDet αποτελείται από δύο κύρια μέρη, το πρώτο αφορά την εξαγωγή χαρακτηριστικών ισοδύναμων στις περιστροφές των αντικειμένων και το δεύτερο αφορά την ευθυγράμμιση των περιοχών ενδιαφέροντος.



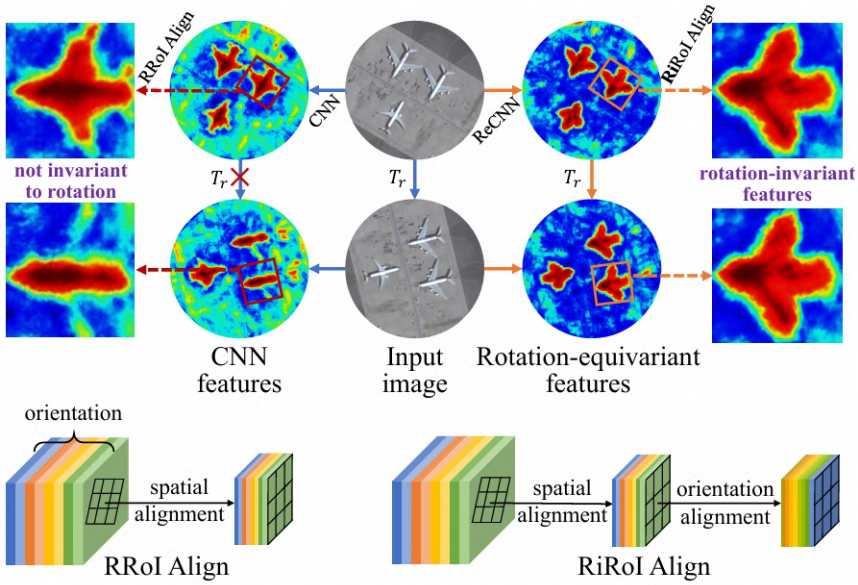
Σχήμα 3.11: *Αρχιτεκτονική ReDet [7]*

Το πρώτο μέρος (backbone) χρησιμοποιεί το ReResNet με ReFPN τα οποία είναι βασισμένα στα πολύ διαδεδομένα ResNet [24] και FPN [31] αλλά τα επίπεδα τους έχουν αλλάξει ώστε να γίνουν ανθεκτικά στις περιστροφές των αντικειμένων. Το δεύτερο μέρος είναι το RiRoI align (Rotation-invariant RoI Align) το οποίο ευθυγραμμίζει τα χαρακτηριστικά στον χωρικό άξονα αλλά και στον άξονα της περιστροφής, προσφέροντας έτσι χαρακτηριστικά πλήρως ανθεκτικά στην περιστροφή. Τέλος όπως φαίνεται και στο σχήμα 3.11 υπάρχει ένα πλήρως συνδεδεμένο επίπεδο από το οποίο βγαίνουν τελικά οι προβλέψεις για την θέση και την κατηγορία των αντικειμένων.

Ο ReDet έχει καταφέρει να έχει μικρό μέγεθος σε σχέση με όμοια μοντέλα που έχουν δημιουργηθεί αλλά ταυτόχρονα έχει κορυφαίες επιδόσεις σε όλα τα σύνολα δεδομένων που έχει δοκιμαστεί.

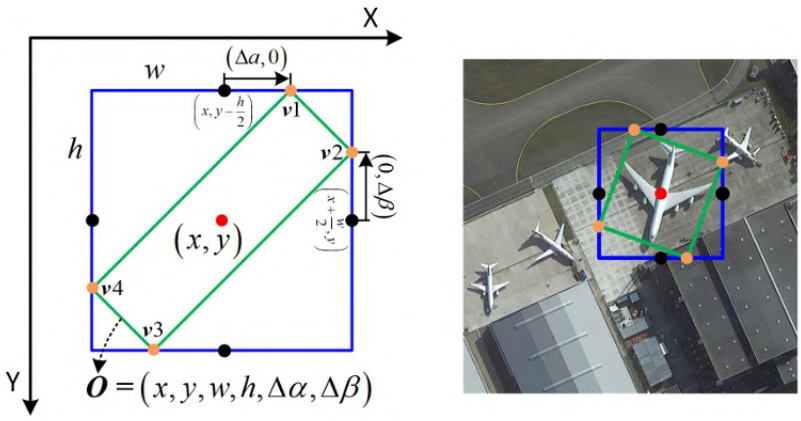
3.5 Oriented R-CNN

Το Oriented-RCNN [8] είναι ένα ισχυρό και αποδοτικό μοντέλο δύο σταδίων, το οποίο είναι και αυτό για την ανίχνευση προσανατολισμένων αντικειμένων. Οι παραδοσιακοί αλγόριθμοι δύο σταδίων έχουν μεγάλες υπολογιστικές απαιτήσεις για να ανιχνεύσουν προσανατολισμένα αντικείμενα και για αυτό είναι πολύ πιο αργοί και δεν μπορούν να χρησιμοποιηθούν σε εφαρμογές πραγματικού χρόνου. Το Oriented-RCNN στοχεύει να βελτιώσει



Σχήμα 3.12: *Αναπαράσταση χαρακτηριστικών αναλόγως την περιστροφή [7]*

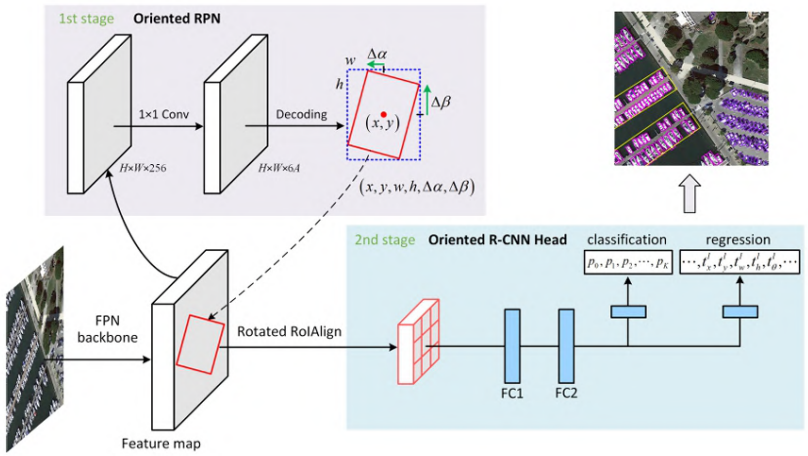
αυτή την διαδικασία χρησιμοποιώντας δύο καινούρια στοιχεία.



Σχήμα 3.13: *Αναπαράσταση μετατόπισης μεσαίου σημείου [8]*

Το Oriented-RCNN αποτελείται από δύο κύρια μέρη, το πρώτο είναι ένα προσανατολισμένο Region Proposal Network (RPN) και το δεύτερο είναι μία προσανατολισμένη κεφαλή R-CNN. Το προσανατολισμένο RPN παράγει ένα αραιό σύνολο υψηλής ποιότητας προσανατολισμένων προτάσεων απευθείας απο τους χάρτες χαρακτηριστικών που παράγονται από το FPN [31]. Σε αντίθεση με τα παραδοσιακά RPN που χρησιμοποιούν οριζόντιες άγκυρες, το προσανατολισμένο RPN χρησιμοποιεί μια νέα αναπαράσταση μετατόπισης μεσαίου σημείου για την πρόβλεψη προσανατολισμένων πλαισίων οριοθέτησης.

Το δεύτερο μέρος του Oriented R-CNN, η προσανατολισμένη κεφαλή R-CNN, βελτιώνει τις προσανατολισμένες προτάσεις που παράγονται από το RPN. Αυτό το στάδιο περιλαμβάνει ευθυγράμμιση της περιοχής ενδιαφέροντος (RoI) με περιστροφή, όπου τα χαρακτηριστικά κάθε πρότασης εξάγονται και χρησιμοποιούνται για την εκτέλεση ταξινόμησης και παλινδρόμησης.



Σχήμα 3.14: *Αρχιτεκτονική Oriented R-CNN [8]*

Το Oriented R-CNN παρουσίασε σημαντική πρόοδο στον τομέα της ανίχνευση προσανατολισμένων αντικειμένων καθώς απλοποίησε την διαδικασία χωρίς να επιβαρύνει την ακρίβεια. Οι καινοτομίες που έφερε αξίζουν να ερευνηθούν περαιτέρω και να αξιοποιηθούν σαν βάση για μελλοντικές βελτιώσεις στην ανίχνευση αντικειμένων.

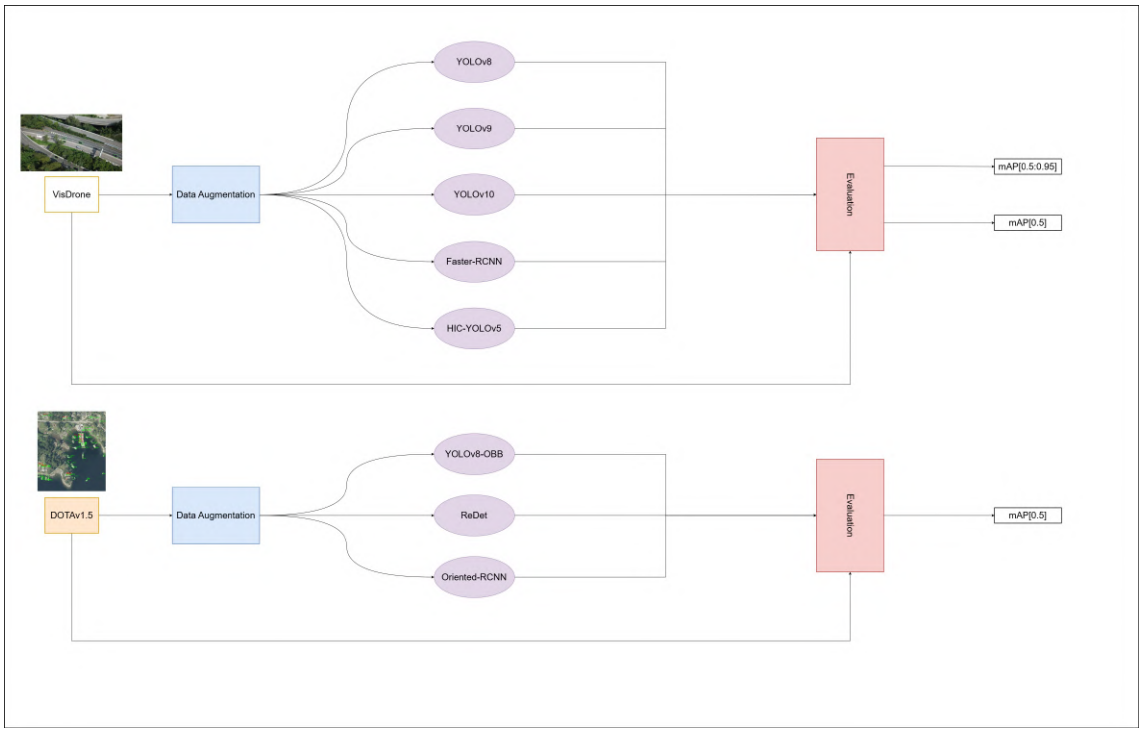
Μέρος II

Πειραματικό Μέρος

Κεφάλαιο 4

Υλοποίηση

Σε αυτό το κεφάλαιο αναλύεται η δομή του πειράματος και αναλύονται τα σύνολα δεδομένων που χρησιμοποιήθηκαν στο πείραμα, οι μετρικές για την αξιολόγηση των μοντέλων και το περιβάλλον που έγινε η εκπαίδευση των μοντέλων. Η συνολική αρχιτεκτονική του πειράματος φαίνεται στην παρακάτω εικόνα και αμέσως μετά αναλύονται με την σειρά τα τμήματα του πειράματος.



Σχήμα 4.1: *Αρχιτεκτονική του πειράματος*

4.1 Σύνολα Δεδομένων

Στο πείραμα χρησιμοποιούνται δύο σύνολα δεδομένων τα οποία αναλύονται λεπτομερώς παρακάτω.

4.1.1 VisDrone

Το Visdrone [9] είναι ένα εξειδικευμένο σύνολο δεδομένων σχεδιασμένο για την ανίχνευση αντικειμένων σε εικόνες που συλλέγονται από μη επανδρωμένα αεροσκάφη (UAV). Το Visdrone δημιουργήθηκε με στόχο να καλύψει τις απαιτήσεις της ανίχνευσης μικρών αντικειμένων σε εναέριες λήψεις και προσφέρει ένα ευρύ φάσμα περιστάσεων που ανταποκρίνονται στον πραγματικό κόσμο.

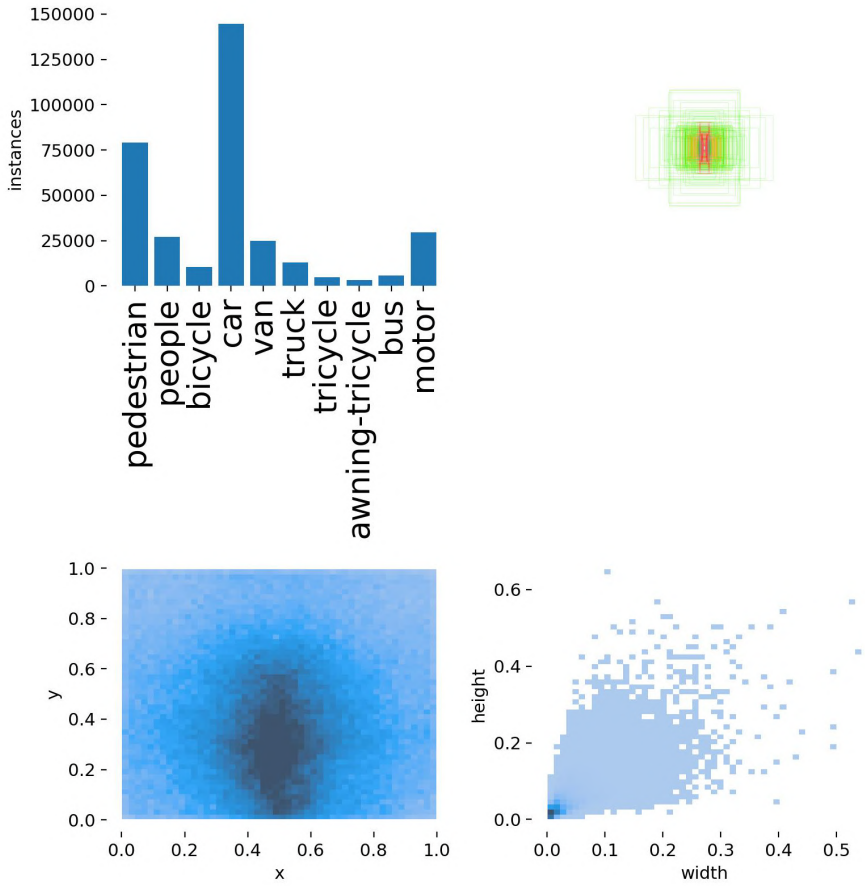
Οι εικόνες που περιέχονται στο VisDrone συλλέγονται από διάφορα είδη UAV, που πετούν σε διαφορετικά υψόμετρα, σε διαφορετικές ώρες της ημέρες και έτσι υπάρχουν ποικίλες συνθήκες φωτισμού και καιρού. Επίσης, οι εικόνες είναι τραβηγμένες και σε αστικές περιοχές αλλά και σε αγροτικές ζώνες με διαφορετικούς τύπους εδάφους και περιβάλλοντος. Στο σύνολο δεδομένων υπάρχουν συνολικά 10,209 εικόνες και χωρίζονται σε 6,471 εικόνες για εκπαίδευση, 548 εικόνες για επαλήθευση και 3,190 εικόνες για testing (από αυτές μόνο οι 1,610 είναι διαθέσιμες στο κοινό και οι υπόλοιπες έχουν δοθεί μόνο για διαγωνισμούς). Η ανάλυση των εικόνων ποικίλλει καλύπτοντας από 640×480 έως 1920×1080 και τα αντικείμενα που έχουν επισημανθεί ανήκουν σε δέκα κλάσεις: πεζοί, άνθρωποι, ποδήλατα, αυτοκίνητα, βανάκια, φορτηγά, τρίκυκλα, καροτσάκια, λεωφορεία και μηχανάκια (pedestrian, people, bicycle, car, van, truck, tricycle, awning-tricycle, bus, motor). Όπως φαίνεται και από τα παραδείγματα 4.1 τα περισσότερα αντικείμενα που εμφανίζονται στις εικόνες είναι μικρά και φαίνονται ξεκάθαρα οι δυσκολίες που παρουσιάζονται στον εντοπισμό των αντικειμένων. Αυτό κάνει το συγκεκριμένο σύνολο δεδομένων ιδανικό για την παρούσα εργασία.

Όπως φαίνεται και από το σχήμα 4.2 τα αντικείμενα στις εικόνες δεν είναι ισοκατανεμημένα στις κλάσεις αλλά αυτό είναι αναμενόμενο, καθώς αυτό ισχύει και στην πραγματικότητα. Επίσης φαίνεται ότι τα περισσότερα αντικείμενα που υπάρχουν είναι μικρά και είναι διασκορπισμένα σε όλη την περιοχή των εικόνων.

4.1.2 DOTAv1.5

Το DOTA (Dataset for Object Detection in Aerial Images) [10] είναι ένα από τα πιο εκτενή και ευρέως χρησιμοποιούμενα σύνολα δεδομένων για ανίχνευση αντικειμένων σε εναέριες εικόνες. Όπως και το VisDrone, έτσι και αυτό δημιουργήθηκε για την εξέλιξη των τεχνικών εντοπισμού μικρών αντικειμένων σε εικόνες τραβηγμένες από μεγάλο υψόμετρο.

Το DOTA περιέχει 2,806 εικόνες οι οποίες έχουν διαστάσεις από 800×800 έως 20000×20000 και έχουν συλλεχθεί από διαφορετικές πηγές όπως από δορυφόρους, από το Google Earth και από drones. Οι εικόνες καλύπτουν ένα μεγάλο εύρος διαφορετικών φόντων καθώς είναι παρμένες από περιοχές με διαφορετική γεωγραφία και τύπους εδάφους. Οι 2,806 εικόνες είναι μοιρασμένες σε 1,411 για την εκπαίδευση των μοντέλων, 458 για την επαλήθευση και 937 για το testing, για τις οποίες όμως δεν έχουν δοθεί στο κοινό τα σωστά πλαίσια οριοθέτησης στο κοινό, οπότε στην παρούσα εργασία θα χρησιμοποιηθούν οι εικόνες για την επαλήθευση και για τους σκοπούς του testing. Στο DOTAv1.5 τα αντικείμενα χωρίζονται σε 16 κλάσεις: αεροπλάνο, πλοίο, δεξαμενή αποθήκευσης, γήπεδο μπέιζμπολ, γήπεδο τένις, γήπεδο μπάσκετ, γήπεδο στίβου, λιμάνι, γέφυρα, μεγάλο όχημα, μικρό όχημα, ελικόπτερο, κυκλικός κόμβος, γήπεδο ποδοσφαίρου, πισίνα και γερανός κοντέινερ (plane, ship, storage tank, baseball diamond, tennis court, basketball court, ground track field, harbor,

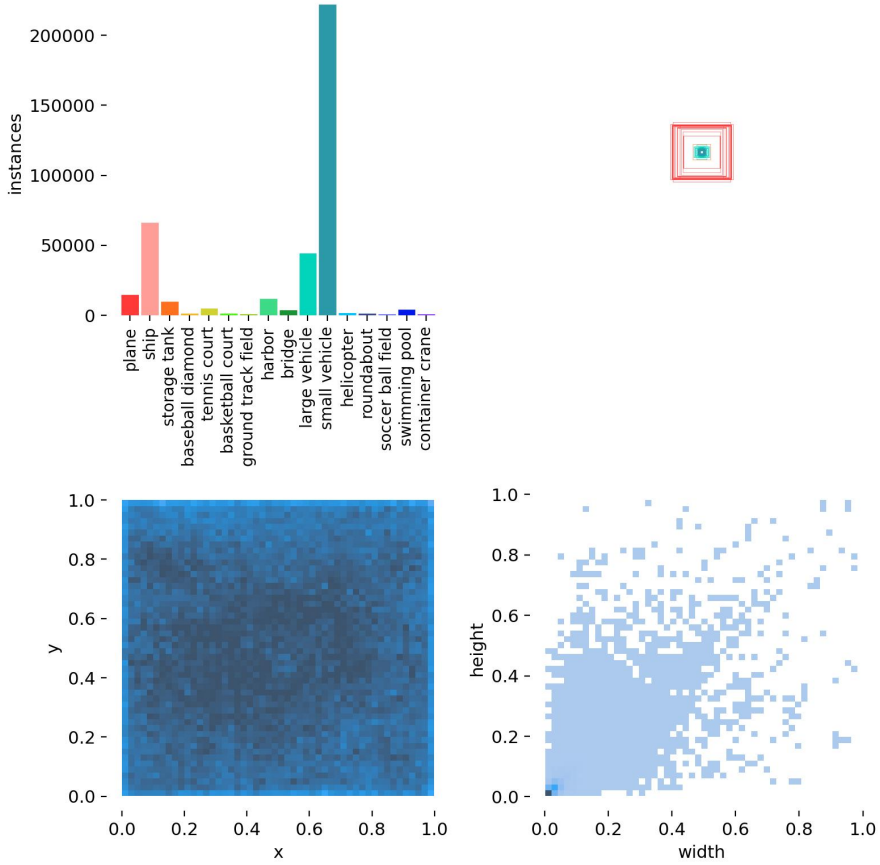


Σχήμα 4.2: *Annotations του VisDrone Dataset [9]*

bridge, large vehicle, small vehicle, helicopter, roundabout, soccer ball field, swimming pool container crane).

Αυτό που κάνει το DOTA να ξεχωρίζει από το VisDrone και τα υπόλοιπα σύνολα δεδομένων είναι τα πλαίσια οριοθέτησης του, τα οποία είναι προσανατολισμένα σύμφωνα με τον προσανατολισμό των αντικειμένων μέσα στην εικόνα και δεν είναι απαραίτητα παράλληλα στους άξονες της φωτογραφίας. Μερικά παραδείγματα φαίνονται στις εικόνες παρακάτω 4.2.

Επειδή οι εικόνες έχουν πολύ μεγάλο εύρος ανάλυσης για να γίνει πιο εύκολη η εκπαίδευση των μοντέλων και να επιτύχουν καλύτερες επιδόσεις ακολουθείτε μια διαδικασία χωρισμού των εικόνων σε επιμέρους εικόνες μεγέθους 1024×1024 με επικάλυψη 200 pixels (padding). Με αυτή την διαδικασία το τελικό σύνολο δεδομένων περιέχει 15,749 εικόνες για εκπαίδευση και 5,297 εικόνες για επαλήθευση όλες μεγέθους 1024×1024 (παραδείγματα φαίνονται παρακάτω 4.3). Παρόλο που γίνεται μεγέθυνση των εικόνων τα περισσότερα αντικείμενα παραμένουν πολύ μικρά και αυτό φαίνεται και στο σχήμα 4.3, στο οποίο φαίνονται η κατανομή των αντικειμένων στις κλάσεις, οι τοποθεσίες των αντικειμένων μέσα στις εικόνες και τα μεγέθη των αντικειμένων σε σχέση με το μέγεθος της εικόνας.



Σχήμα 4.3: *Annotations του DOTAv1.5 Dataset*

4.2 Επαύξηση Δεδομένων

Για την αύξηση των συνόλων δεδομένων που χρησιμοποιούνται και για την αύξηση της απόδοσης της διαδικασίας εκπαίδευσης των μοντέλων χρησιμοποιήθηκαν κάποιες τεχνικές επαύξησης δεδομένων. Οι τεχνικές αυτές στην ουσία αλλάζουν κάποιες εικόνες τυχαία ώστε να γίνουν τα μοντέλα πιο ανθεκτικά στις αλλαγές και να μπορούν να γενικεύουν καλύτερα σε νέα δεδομένα. Παρακάτω περιγράφονται συνοπτικά οι τεχνικές που χρησιμοποιήθηκαν και ποιες τεχνικές χρησιμοποιήθηκαν σε ποια μοντέλα:

- Θάμπωμα (Blur):Το θάμπωμα εφαρμόζεται στις εικόνες για να μειωθεί ο θόρυβος και οι λεπτομέρειες. Αυτή η τεχνική βοηθά στη δημιουργία πιο γενικευμένων χαρακτηριστικών.

- Μέσο Θάμπωμα (MedianBlur): Η τεχνική αυτή χρησιμοποιεί τον μεσαίο όρο των εικονοστοιχείων σε ένα παράθυρο συγκεκριμένου μεγέθους για να θαμπώσει την εικόνα. Είναι ιδιαίτερα αποτελεσματική στην απομάκρυνση των salt-and-pepper θορύβων, βελτιώνοντας την ποιότητα της εικόνας.

- Μετατροπή σε Γκρι (ToGray): Η μετατροπή των έγχρωμων εικόνων σε κλίμακα του γκρι μειώνει την πολυπλοκότητα των δεδομένων και επιτρέπει στα μοντέλα να εστιάσουν στα χωρικά χαρακτηριστικά των αντικειμένων.

Εικόνα 4.1: *Παραδείγματα εικόνων από το VisDrone [9]*

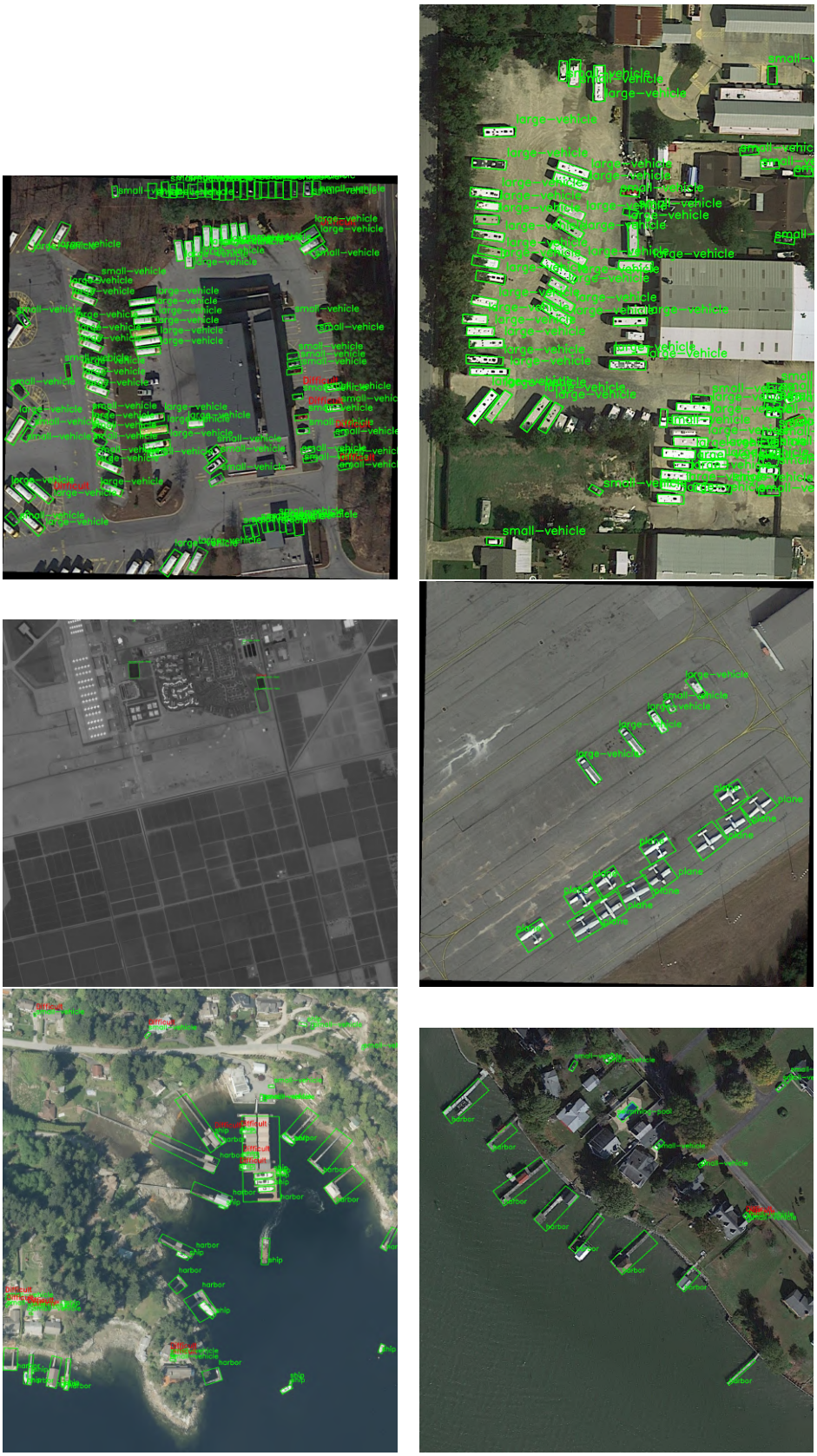
- Ιστογραμμική Ισοστάθμιση Προσαρμοστικής Αντίθεσης (CLAHE): Η CLAHE βελτιώνει την αντίθεση της εικόνας μέσω της τοπικής ισοστάθμισης του ιστογράμματος, κάνοντας τις λεπτομέρειες των μικρών αντικειμένων πιο ευδιάκριτες και βελτιώνοντας την ακρίβεια της ανίχνευσης.

- Αναστροφή (RandomFlip): Η τυχαία αναστροφή κάποιων εικόνων, οριζόντια ή κάθετα, αυξάνει την ποικιλία των δεδομένων εκπαίδευσης και βοηθά το μοντέλο να αναγνωρίζει αντικείμενα ανεξαρτήτως προσανατολισμού.

- Αλλαγή Απόχρωσης, Κορεσμού και Φωτεινότητας (HSV-Hue, HSV-Saturation, HSV-Value): Οι μετασχηματισμοί ΗΣ" επιτρέπουν την τροποποίηση της απόχρωσης, του κορεσμού και της φωτεινότητας της εικόνας, αυξάνοντας την ποικιλία των χρωματικών συνθηκών που το μοντέλο μπορεί να αντιμετωπίσει.

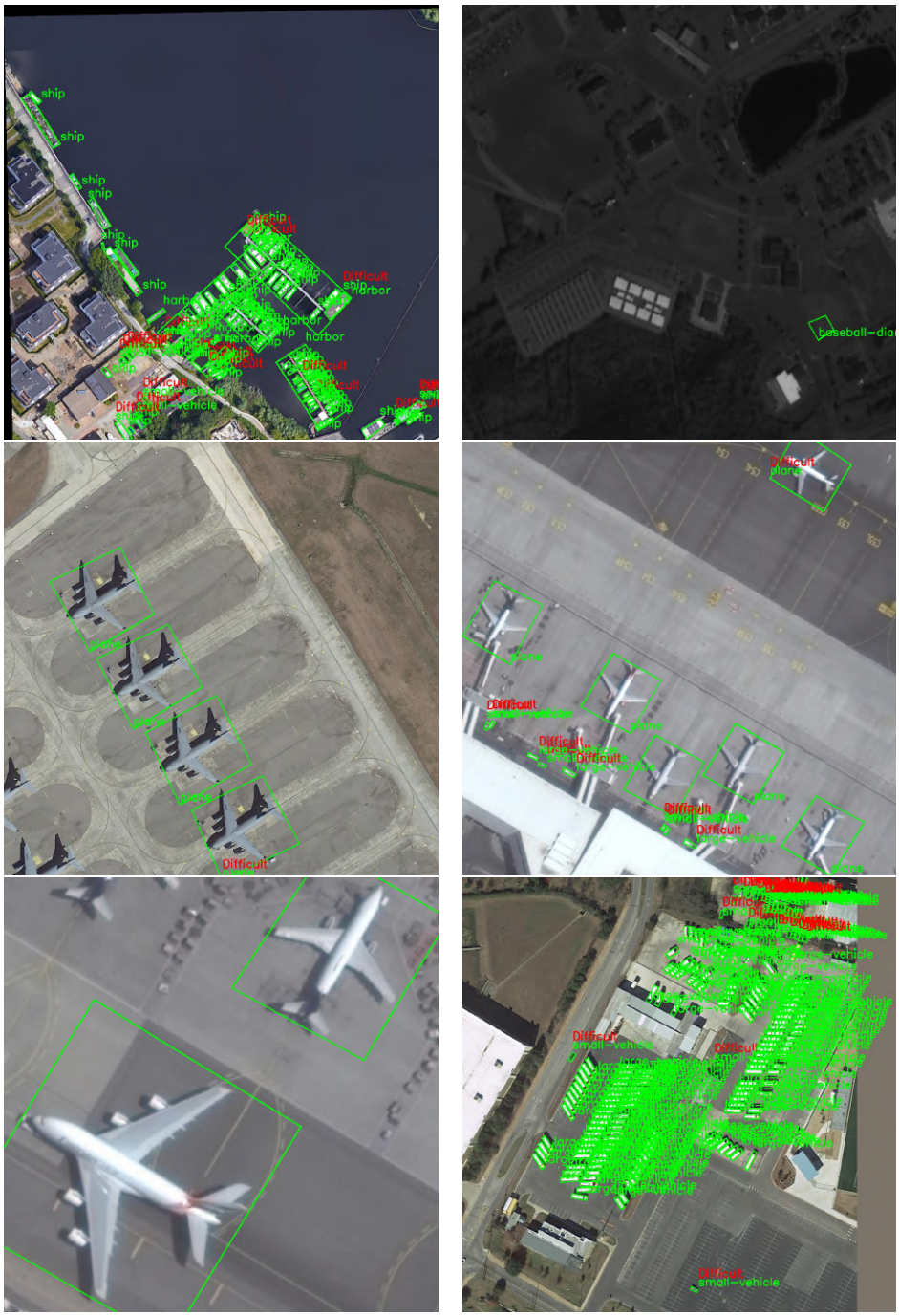
- Περιστροφή (Rotation): Η περιστροφή των εικόνων σε τυχαίες γωνίες βοηθά το μοντέλο να αναγνωρίζει αντικείμενα ανεξαρτήτως προσανατολισμού, βελτιώνοντας την ικανότητά του να γενικεύει σε αδημοσίευτα δεδομένα.

- Μωσαϊκό (Mosaic): Η τεχνική μωσαϊκού συνδυάζει τμήματα από τέσσερις διαφορετικές εικόνες σε μία νέα εικόνα. Αυτή η μέθοδος αυξάνει σημαντικά την ποικιλία των



Εικόνα 4.2: *Παραδείγματα εικόνων από το DOTAv1.5 [10]*

δεδομένων και βοηθά το μοντέλο να αναγνωρίζει αντικείμενα σε ποικίλα περιβάλλοντα.



Εικόνα 4.3: *Παραδείγματα εικόνων από το DOTAv1.5 μετά τον διαχωρισμό [10]*

- Κανονικοποίηση (Normalize): Η κανονικοποίηση των εικονοστοιχείων των εικόνων σε ένα συγκεκριμένο εύρος τιμών συμβάλλει στην ομογενοποίηση των δεδομένων εισόδου, βελτιώνοντας την απόδοση του μοντέλου και επιταχύνοντας την εκπαίδευση.
- Ανάμειξη (Mixup): Η τεχνική ανάμειξης συνδυάζει δύο εικόνες και τις αντίστοιχες ετικέτες τους, δημιουργώντας νέα παραδείγματα εκπαίδευσης. Αυτό βοηθά στη μείωση της υπερβολικής προσαρμογής και στην αύξηση της γενίκευσης του μοντέλου.

Από τις παραπάνω τεχνικές τα YOLOv8, YOLOv9 και YOLOv10 χρησιμοποιούν: Θάμπωμα, Μέσο Θάμπωμα, Μετατροπή σε Γκρι και Ιστογραμμική Ισοστάθμιση Προσαρμοστικής

Αντίθεσης. Το HIC-YOLOv5 χρησιμοποιεί: Αλλαγή Απόχρωσης, Κορεσμού και Φωτεινότητας, Αναστροφή, Περιστροφή, Μωσαϊκό και Ανάμειξη. Το Faster-RCNN χρησιμοποιεί μόνο Αναστροφή και τα Oriented-RCNN και ReDet χρησιμοποιούν: Αναστροφή και Κανονικοποίηση

4.3 Περιβάλλον Εκτέλεσης Πειράματος

Τα πειράματα εκτελέστηκαν στην δωρεάν πλατφόρμα της Google που ονομάζεται Kaggle η οποία προσφέρει δωρεάν πρόσβαση σε Virtual Machines εξοπλισμένα με 8 επεξεργαστές vCPUs, 29GB μνήμη (RAM) και 2 κάρτες γραφικών NVIDIA T4 GPU με 15 ΓΒ μνήμη η κάθε μία. Αυτά τα χαρακτηριστικά ήταν αρκετά για την υλοποίηση των πειραμάτων σε λογικό χρόνο.

Πέρα από το Kaggle, χρησιμοποιήθηκαν και κάποια open-source εργαλεία με έτοιμα νευρωνικά μοντέλα. Το ένα από αυτά έχει δημιουργηθεί από την ομάδα Ultralytics [32], η οποία έχει δημιουργήσει και πολλά από τα μοντέλα YOLO. Ο κώδικας που έχουν δημιουργήσει παρέχει εύκολη πρόσβαση σε όλα τα μοντέλα YOLO καθώς και μεγάλη ευελιξία στην δημιουργία νέων μοντέλων. Με την χρήση του πακέτου Ultralytics μέσα σε λίγες γραμμές κώδικα μπορεί κανείς να εκπαιδεύσει μοντέλα σε ένα μεγάλο πλήθος συνόλων δεδομένων, να επαληθεύσει ήδη εκπαιδευμένα μοντέλα, να κάνει fine-tuning σε προεκπαιδευμένα μοντέλα και να ενσωματώσει μοντέλα σε άλλες εφαρμογές.

Τα άλλα δύο εργαλεία ονομάζονται mmdetection [33] και mmrotate [34] και έχουν αναπτυχθεί από την ομάδα OpenMMLab και παρέχουν έτοιμο κώδικας για την χρήση νευρωνικών μοντέλων για τον εντοπισμό αντικειμένων με πλαίσια οριοθέτησης παράλληλα στους άξονες της εικόνας ή όχι αντίστοιχα. Είναι φτιαγμένα με την ίδια λογική με τον προηγούμενο αλλά προσφέρουν πολύ μεγαλύτερη ευχέρεια σε αλλαγές στα μοντέλα και τα επιμέρους στοιχεία τους.

4.4 Μετρικές

Για την αξιολόγηση της απόδοσης ενός αλγορίθμου ανίχνευσης αντικειμένων έχουν δημιουργηθεί κάποιες μετρικές οι οποίες είναι αποδεκτές από όλους. Οι πιο συνηθισμένες από αυτές είναι οι IoU, Precision, Recall και mAP από τις οποίες στην παρούσα εργασία χρησιμοποιείται η mAP (Μέση Ακρίβεια Πρόβλεψης) καθώς αυτή θεωρείται η κύρια μετρική για την αξιολόγηση της ανίχνευσης αντικειμένων. Οι λεπτομερείς ορισμοί τους δίνονται παρακάτω:

1. IoU (Συντελεστής Διασταύρωσης Ένωσης): Ο υπολογισμός του IoU γίνεται με τον υπολογισμό της περιοχής επικάλυψης μεταξύ της προβλεπόμενης περιοχής του αντικειμένου (A) και της πραγματικής περιοχής του αντικειμένου (B) διά το συνολικό εμβαδόν των δύο. Η φόρμουλα εκφράζεται ως:

$$IoU = \frac{A \cap B}{A \cup B}$$

Η τιμή του IoU κυμαίνεται από 0 έως 1 και όσο μεγαλύτερη η τιμή, τόσο πιο ακριβές το μοντέλο. Μια χαμηλότερη τιμή του αριθμητή υποδεικνύει ότι η πρόβλεψη απέτυχε να προβλέψει ακριβώς την πραγματική περιοχή του αντικειμένου. Αντιθέτως, μια υψηλότερη τιμή του παρονομαστή υποδεικνύει μια μεγαλύτερη προβλεπόμενη περιοχή, με αποτέλεσμα μια χαμηλότερη τιμή IoU.

2. Precision (Ακρίβεια): Η ακρίβεια αντιπροσωπεύει το ποσοστό των δειγμάτων που προβλέφθηκαν σωστά στο σύνολο των θετικών προβλέψεων. Μπορεί να εκφραστεί ως:

$$\text{Precision} = \frac{\text{Πραγματικά Θετικά}}{\text{Πραγματικά Θετ ικά + Ψευδή Θετικά}}$$

3. Recall (Ανάκληση): Η ανάκληση αντιπροσωπεύει το ποσοστό των δειγμάτων που είναι πραγματικά θετικά και προβλέπονται σωστά. Μπορεί να εκφραστεί ως:

$$\text{Recall} = \frac{\text{Πραγματικά Θετικά}}{\text{Πραγματικά Θετικά + Ψευδή Αρνητικά}}$$

4. mAP (Μέση Ακρίβεια Πρόβλεψης): Η Μέση Ακρίβεια (AP) είναι μια μέτρηση των βαθμολογιών Ακρίβειας σε διάφορα κατώφλια IoU κατά μήκος της καμπύλης Ακρίβειας-Ανάκλησης (PR), και υπολογίζεται ως ένας ζυγισμένος μέσος όρος. Η mAP είναι ο μέσος όρος της μέσης ακρίβειας για όλες τις κατηγορίες αντικειμένων. Συγκεκριμένα, το mAP@0.5 αντιπροσωπεύει το mAP όταν το όριο IoU είναι 0.5, ενώ το mAP@[0.5:0.95] είναι το μέσο mAP όταν το IoU κυμαίνεται από 0.5 έως 0.95 με βημα 0.05.

Κεφάλαιο 5

Πειραματικά Αποτελέσματα

Σε αυτό το κεφάλαιο παρουσιάζονται τα πειραματικά αποτελέσματα. Τα αποτελέσματα παρατίθενται σε ποσοτική και σε ποιοτική μορφή στους παρακάτω πίνακες και σε κάποια τυχαία παραδείγματα εικόνων που επιλέχθηκαν για να συγκριθούν τα μοντέλα.

5.1 Αποτελέσματα με βάση το Σύνολο Δεδομένων

Όπως αναλύθηκε και στο προηγούμενο κεφάλαιο χρησιμοποιήθηκαν δύο σύνολα δεδομένων που έχουν δημιουργηθεί για την εφαρμογή που πραγματεύεται η παρούσα εργασία τα αποτελέσματα των μοντέλων πάνω στα δύο σύνολα δεδομένων παρουσιάζονται παρακάτω.

5.1.1 VisDrone

Παρακάτω στον πίνακα 5.1 παρουσιάζονται τα αποτελέσματα των μοντέλων στο σύνολο δεδομένων VisDrone. Το μοντέλο που αποδίδει καλύτερα είναι το YOLOv8-Pretrained δηλαδή το YOLOv8 που είναι πρωτα εκπαιδευμένο σε ένα πολύ μεγάλο σύνολο δεδομένων και στην παρούσα εργασία έγινε επανεκπαίδευση (fine-tuning) στο VisDrone. Παρόλο που το YOLOv8 φαίνεται να αποδίδει πολύ καλύτερα πρέπει να σημειωθεί και η πολύ καλή απόδοση του YOLOv10 ιδιαίτερα η προεκπαιδευμένη έκδοση αλλα και το HIC-YOLOv5 το οποίο έχει κάτω από τις μισές παραμέτρους του YOLOv10 και σχεδόν 5 φορές λιγότερες παραμέτρους από το YOLOv8 αποδίδει σχεδόν ισάξια με τα υπόλοιπα και μία λίγο μεγαλύτερη έκδοσή του μπορεί να απέδιδε καλύτερα από όλα.

Models	Parameters	Epochs	Precision	Recall	mAP0.5	mAP[0.5:0.95]
YOLOv9	25.536M	50	0.473	0.38	0.366	0.217
YOLOv8	43.637M	50	0.533	0.419	0.419	0.252
YOLOv8-Pretrained	43.637M	10	**0.541**	**0.441**	**0.442**	**0.271**
Faster-RCNN	41.753M	24	-	-	0.399	0.248
Faster-RCNN-Pretrained	41.753M	10	-	-	0.409	0.251
HIC-YOLOv5	**9.335M**	50	0.487	0.417	0.402	0.226
YOLOv10	25.780M	50	0.485	0.404	0.392	0.236
YOLOv10-pretrained	25.780M	50	0.52	0.423	0.421	0.256

Πίνακας 5.1: *Αποτελέσματα μοντέλων στο VisDrone*

5.1.2 DOTAv1.5

Παρακάτω στον πίνακα 5.2 παρουσιάζονται τα αποτελέσματα των μοντέλων στο σύνολο δεδομένων DOTAv1.5, στο οποίο τα αντικείμενα σημειώνονται με πλαίσια οριοθέτησης τα οποία είναι προσανατολισμένα σύμφωνα με τον προσανατολισμό του αντικειμένου στην εικόνα. Το μοντέλο που αποδίδει καλύτερα είναι το YOLOv8-OBB δηλαδή η ειδική έκδοση του YOLOv8 που έχει δημιουργηθεί για τον εντοπισμό προσανατολισμένων αντικειμένων. Παρόλο που το YOLOv8 φαίνεται να αποδίδει καλύτερα πρέπει να σημειωθεί και η πολύ καλή απόδοση του ReDet που έχει απόδοση σχεδόν ισάξια με το YOLOv8 και μπορεί να έχει περισσότερες παραμέτρους αλλά χρειάζεται μόλις 12 εποχές κατά την εκπαίδευση του.

Models	Parameters	Epochs	mAP0.5
YOLOv8-OBB	26.409M	50	0.679
ReDeT	31.650M	12	0.671
Oriented-RCNN	41.140M	24	0.603

Πίνακας 5.2: *Αποτελέσματα μοντέλων στο DOTAv1.5*

5.2 Αποτελέσματα με βάση την Αρχιτεκτονική των Μοντέλων

Models	Architecture	Parameters	Epochs	Dataset	mAP0,5	mAP[0,5:0,95]	Release Date
YOLOv9	One-Stage	25.536M	50	Visdrone	0.366	0.217	Feb 2024
YOLOv8	One-Stage	43.637M	50	Visdrone	0.419	0.252	Jan 2023
YOLOv8-Pretrained	One-Stage	43.637M	10	Visdrone	0.442	0.271	Jan 2023
Faster-RCNN	Two-Stage	41.753M	24	Visdrone	0.399	0.248	Apr 2015
Faster-RCNN-Pretrained	Two-Stage	41.753M	10	Visdrone	0.409	0.251	Apr 2015
HIC-YOLOv5	One-Stage	9.335M	50	Visdrone	0.402	0.226	Sep 2023
YOLOv10	One-Stage	25.780M	50	Visdrone	0.392	0.236	May 2024
YOLOv10-pretrained	One-Stage	25.780M	50	Visdrone	0.421	0.256	May 2024
YOLOv8-OBB	One-Stage	26,409M	50	DOTAv1.5	0.679	0.506	Jan 2023
ReDeT	Two-Stage	31.650M	12	DOTAv1.5	0.671	-	Mar 2021
Oriented-RCNN	Two-Stage	41.140M	24	DOTAv1.5	0.603	-	Aug 2021

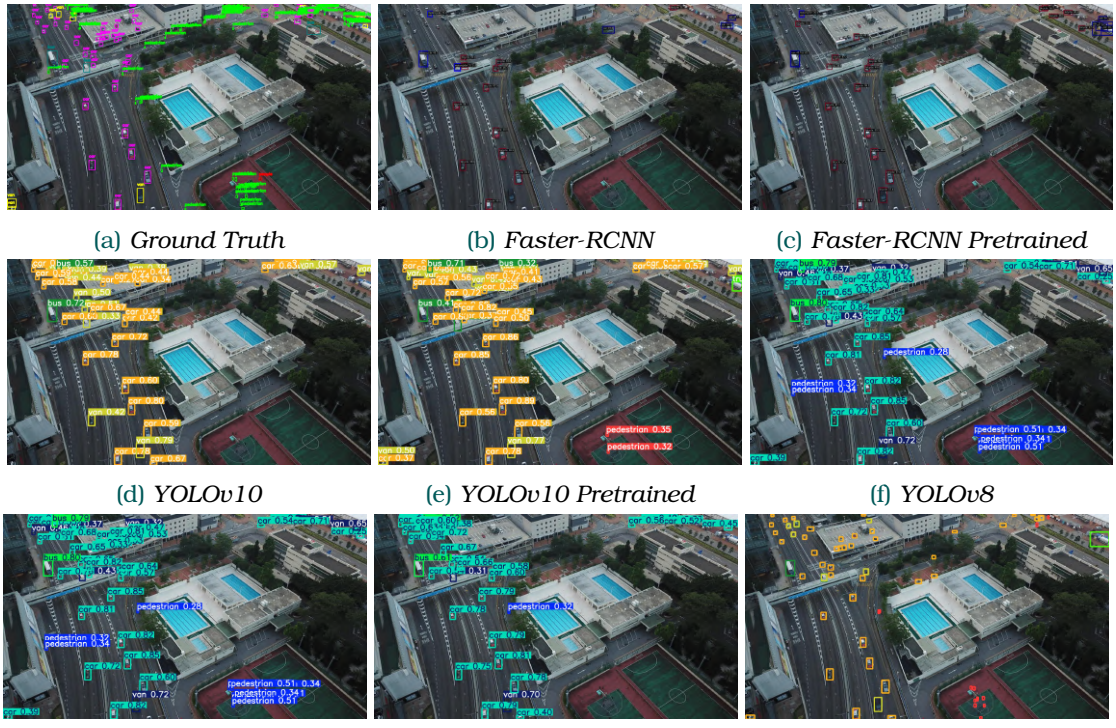
Πίνακας 5.3: *Αποτελέσματα μοντέλων με βάση την Αρχιτεκτονική*

Στον πίνακα 5.3 συνοψίζονται τα αποτελέσματα και των δύο συνόλων δεδομένων και σημειώνεται η κατηγορία που ανήκει το κάθε μοντέλο-ανιχνευτής (ενός σταδίου/ δύο σταδίων). Και οι δύο κατηγορίες ανιχνευτών φαίνεται να τα πηγαίνουν πολύ καλά αλλά οι ανιχνευτές ενός σταδίου έχουν λίγο καλύτερες επιδόσεις. Βέβαια σε κάθε εφαρμογή που χρησιμοποιούνται μοντέλα έχει και διαφορετικές απαιτήσεις, όπως την ταχύτητα των προβλέψεων και την ακρίβεια που χρειάζεται. Για παράδειγμα σε θέματα ιατρικής φύσης η ακρίβεια είναι πολύ σημαντική ενώ η ταχύτητα με την οποία γίνονται οι προβλέψεις δεν παίζει σημαντικό ρόλο οπότε σε μια τέτοια εφαρμογή θα ήταν προτιμότερο ένα μοντέλο με μεγάλη ακρίβεια χωρίς να ενδιαφέρει τόσο η ταχύτητα του. Από την άλλη σε μία εφαρμογή όπως αυτή που εξετάζει η παρούσα εργασία είναι πολύ σημαντική η ταχύτητα που γίνονται οι προβλέψεις καθώς οι ταχύτητες των μη επανδρωμένων αεροσκαφών είναι πολύ μεγάλη και για αυτό θα ήταν προτιμότερο ένα μοντέλο με μεγάλη ταχύτητα προβλέψεων ακόμα και αν η ακρίβεια δεν ήταν η βέλτιστη. Στην γενική περίπτωση οι ανιχνευτές ενός σταδίου είναι πιο γρήγοροι

χρειάζονται περισσότερες εποχές εκπαίδευσης, αλλά πετυχαίνουν πολύ μεγαλύτερες ταχύτητες προβλέψεων αφού έχουν εκπαιδευτεί, ενώ οι ανιχνευτές δύο σταδίων έχουν μεγαλύτερη ακρίβεια αλλά η ταχύτητα τους είναι πολύ μικρότερη.

5.3 Ποιοτικά Αποτελέσματα

Παρακάτω παρουσιάζονται 5 τυχαία επιλεγμένα εικόνες από κάθε σύνολο δεδομένων με τα σωστά πλαίσια οριοθέτησης όπως παρουσιάζονται στο σύνολο δεδομένων και μετά οι προβλέψεις του κάθε μοντέλου για κάθε εικόνα.

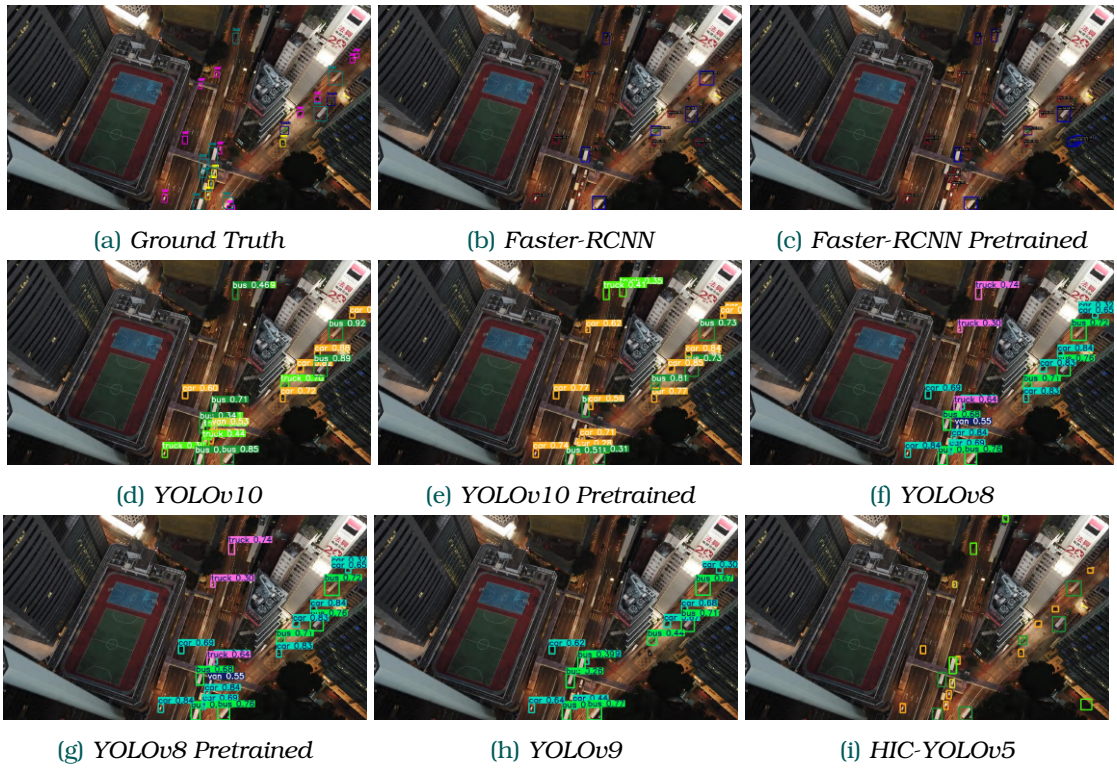


(a) *Ground Truth* (b) *Faster-RCNN* (c) *Faster-RCNN Pretrained*

(d) *YOLOv10* (e) *YOLOv10 Pretrained* (f) *YOLOv8*

(g) *YOLOv8 Pretrained* (h) *YOLOv9* (i) *HIC-YOLOv5*

Εικόνα 5.1: *Παραδείγματα προβλέψεων από το VisDrone (1)*

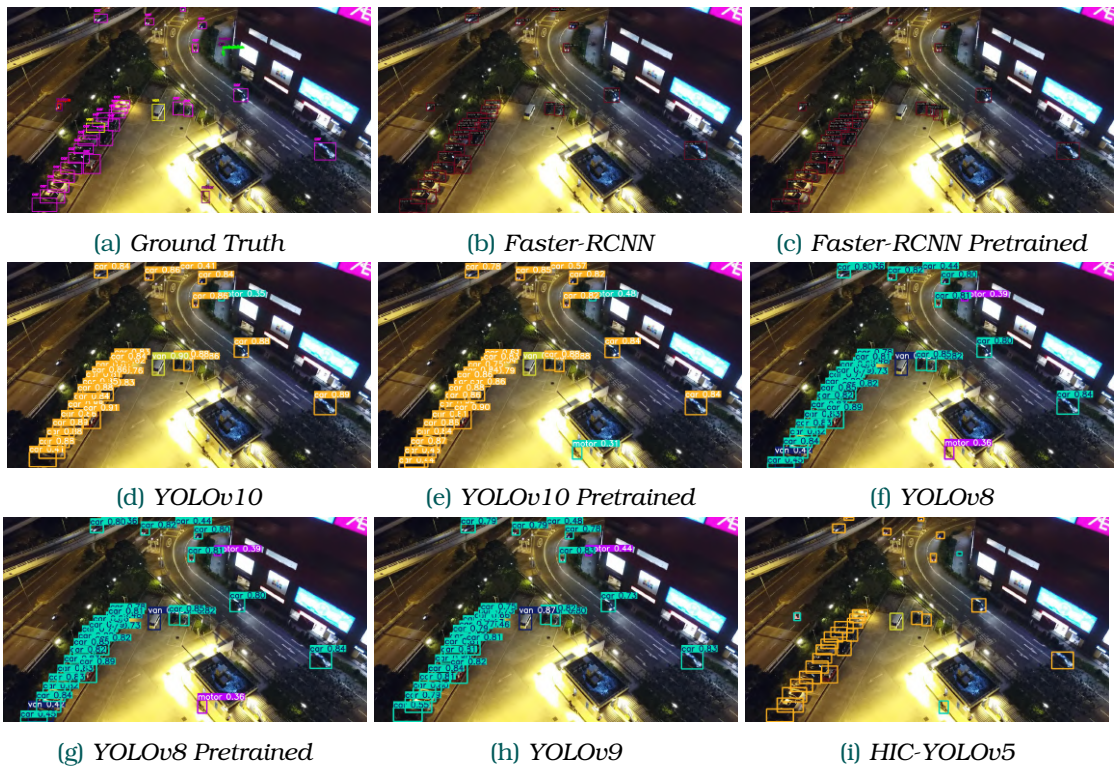


(a) *Ground Truth* (b) *Faster-RCNN* (c) *Faster-RCNN Pretrained*

(d) *YOLOv10* (e) *YOLOv10 Pretrained* (f) *YOLOv8*

(g) *YOLOv8 Pretrained* (h) *YOLOv9* (i) *HIC-YOLOv5*

Εικόνα 5.2: *Παραδείγματα προβλέψεων από το VisDrone (2)*

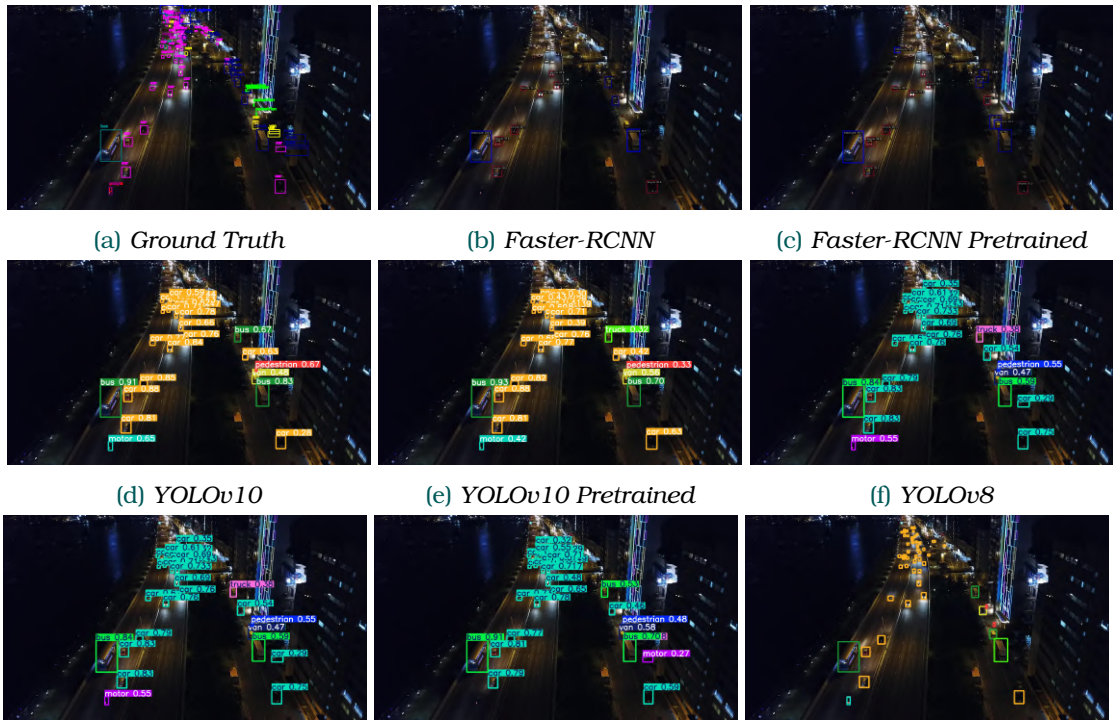


(a) *Ground Truth* (b) *Faster-RCNN* (c) *Faster-RCNN Pretrained*

(d) *YOLOv10* (e) *YOLOv10 Pretrained* (f) *YOLOv8*

(g) *YOLOv8 Pretrained* (h) *YOLOv9* (i) *HIC-YOLOv5*

Εικόνα 5.3: *Παραδείγματα προβλέψεων από το VisDrone (3)*



(a) *Ground Truth* (b) *Faster-RCNN* (c) *Faster-RCNN Pretrained*

(d) *YOLOv10* (e) *YOLOv10 Pretrained* (f) *YOLOv8*

(g) *YOLOv8 Pretrained* (h) *YOLOv9* (i) *HIC-YOLOv5*

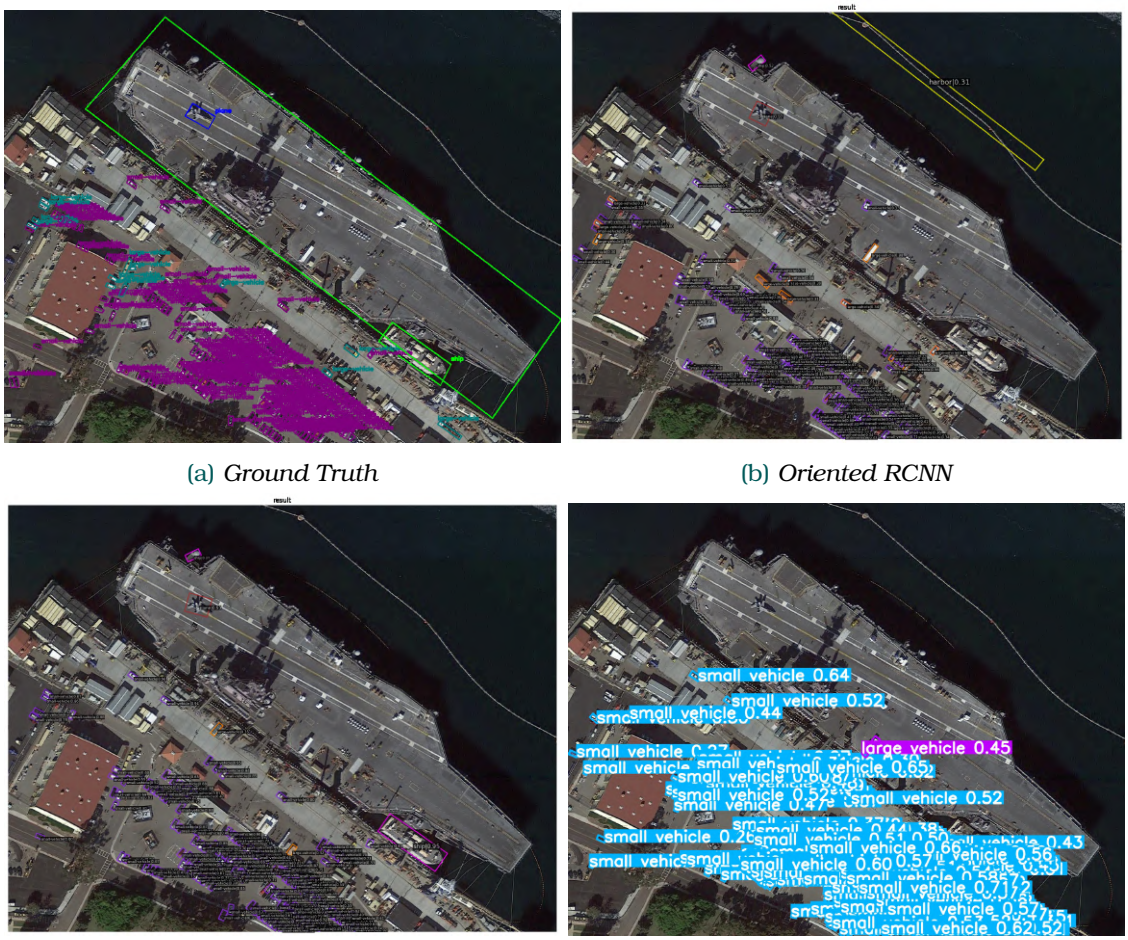
Εικόνα 5.4: *Παραδείγματα προβλέψεων από το VisDrone (4)*

(a) *Ground Truth* (b) *Faster-RCNN* (c) *Faster-RCNN Pretrained*

(d) *YOLOv10* (e) *YOLOv10 Pretrained* (f) *YOLOv8*

(g) *YOLOv8 Pretrained* (h) *YOLOv9* (i) *HIC-YOLOv5*

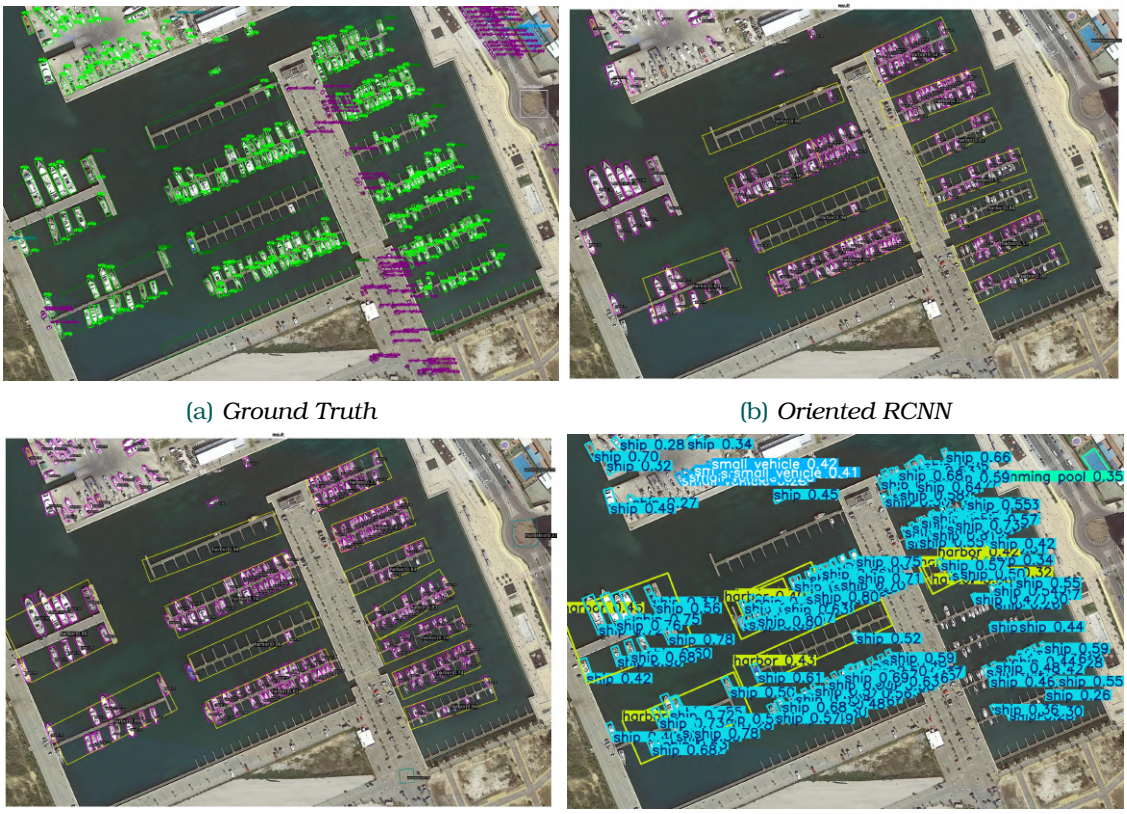
Εικόνα 5.5: *Παραδείγματα προβλέψεων από το VisDrone (5)*



(a) *Ground Truth* (b) *Oriented RCNN*

(c) *ReDet* (d) *YOLOv8-OBB*

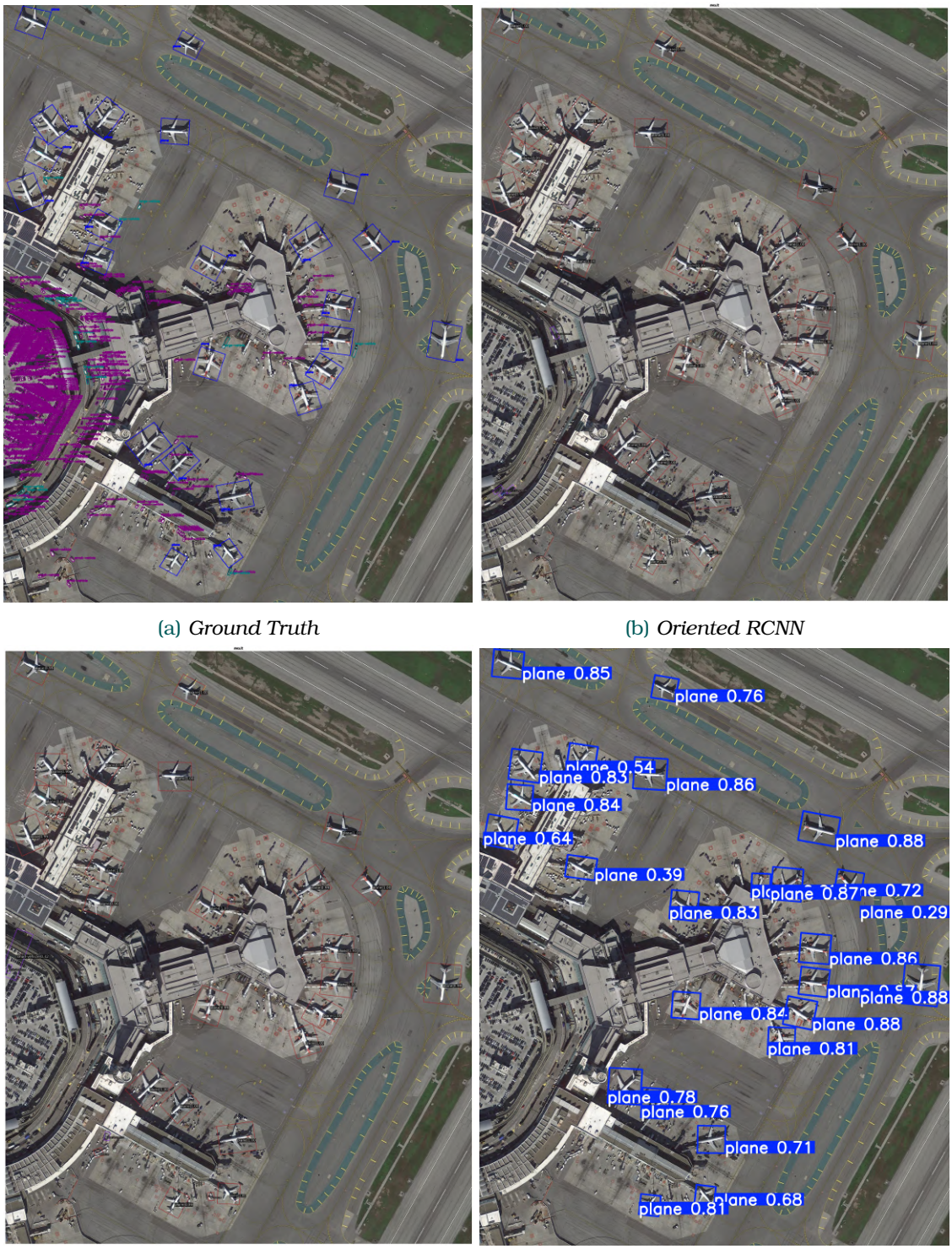
Εικόνα 5.6: *Παραδείγματα προβλέψεων από το DOTAv1.5 (1)*



(a) *Ground Truth* (b) *Oriented RCNN*

(c) *ReDet* (d) *YOLOv8-OBB*

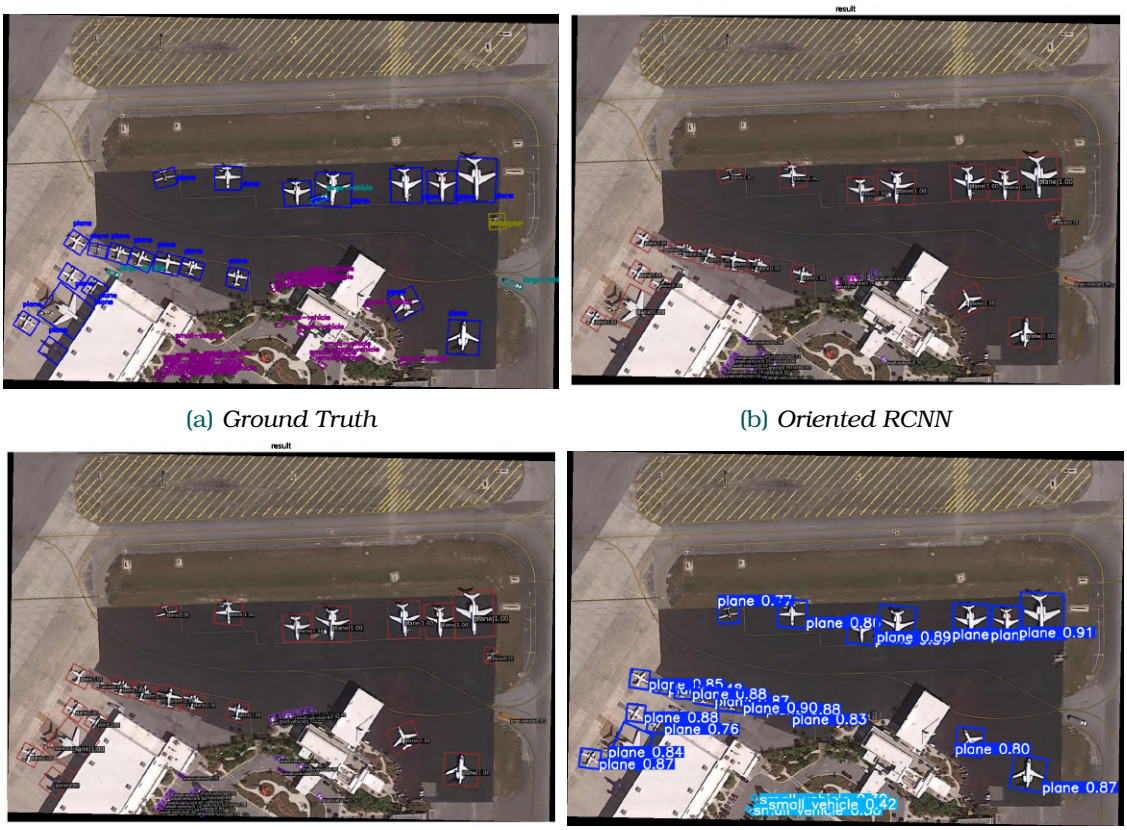
Εικόνα 5.7: *Παραδείγματα προβλέψεων από το DOTAv1.5 (2)*



(a) *Ground Truth* (b) *Oriented RCNN*

(c) *ReDet* (d) *YOLOv8-OBB*

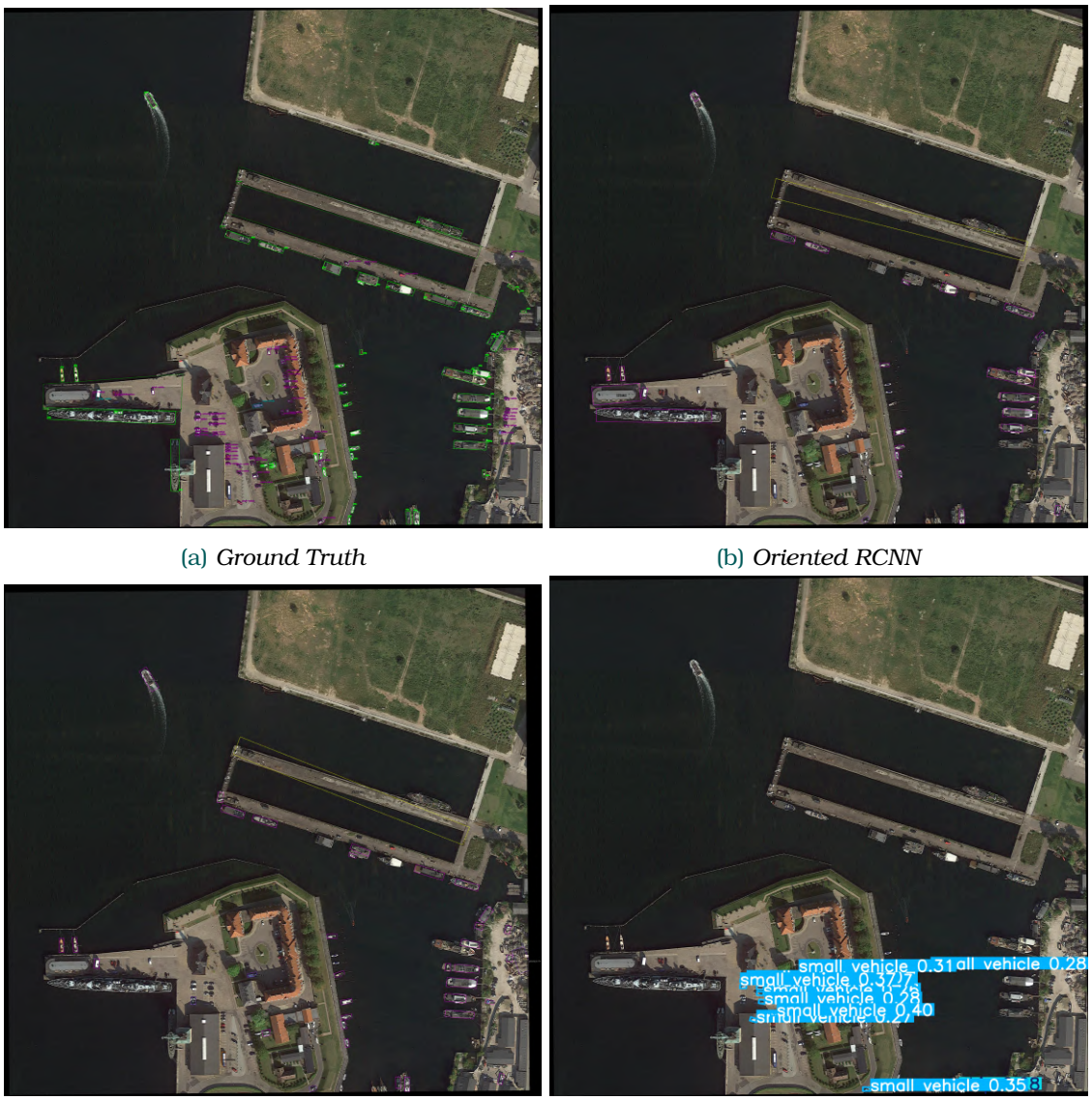
Εικόνα 5.8: *Παραδείγματα προβλέψεων από το DOTAv1.5 (3)*



(a) *Ground Truth* (b) *Oriented RCNN*

(c) *ReDet* (d) *YOLOv8-OBB*

Εικόνα 5.9: *Παραδείγματα προβλέψεων από το DOTAv1.5 (4)*



(a) *Ground Truth*

(b) *Oriented RCNN*

(c) *ReDet*

(d) *YOLOv8-OBB*

Εικόνα 5.10: *Παραδείγματα προβλέψεων από το DOTAv1.5 (5)*

Μέρος III

Επίλογος

Κεφάλαιο 6

Συμπεράσματα και Μελλοντικές Επεκτάσεις

Σ

6.1 Συμπεράσματα

Στα πλαίσια της παρούσας διπλωματικής εργασίας εξετάστηκε η απόδοση διάφορων μοντέλων ανίχνευσης μικρών αντικειμένων σε οπτικά δεδομένα από μη επανδρωμένα αεροσκάφη (UAV-vision). Τα μοντέλα που χρησιμοποιήθηκαν έχουν αρχιτεκτονική ενός ή δύο σταδίων με αρκετές διαφοροποιήσεις μεταξύ τους και τα σύνολα δεδομένων είναι ευρέως γνωστά και πολύ δημοφιλή για την συγκεκριμένη εφαρμογή.

Τα μοντέλα ενός σταδίου (YOLO) είναι γνωστά για την ταχύτητα τους στις προβλέψεις και φαίνεται οτι έχουν πολύ καλα αποτελέσματα στα σύνολα δεδομένων που χρησιμοποιήθηκαν. Ειδικότερα το YOLOv8 φαίνεται να έχει τα καλύτερα αποτελέσματα από όλα τα υπόλοιπα μοντέλα και στα δύο σύνολα δεδομένων, αλλά ταυτόχρονα είναι και το μοντέλο με τις περισσότερες παραμέτρους που σημαίνει οτι χρειάζεται περισσότερος χρόνος για να εκπαιδευτεί. Τα YOLOv9 και YOLOv10 έχουν και αυτά πολύ καλές επιδόσεις με το YOLOv10 να έχει λίγο καλύτερες, αλλά έχουν πολύ λιγότερες παραμέτρους απο το YOLOv8. Επίσης το HIC-YOLOv5 που έχει 5 φορές λιγότερες παραμέτρους από το YOLOv8 έχει αποτελέσματα καλύτερα από τα YOLOv9 και YOLOv10.

Τα μοντέλα δύο σταδίων έχουν επιτύχει και αυτά σχετικά καλά αποτελέσματα με το Faster-RCNN να έχει ίδιες επιδόσεις με το HIC-YOLOv5 αλλά με πολύ περισσότερες παραμέτρους και το ReDet να επιτυγχάνει ίσες επιδόσεις με το YOLOv8 στο DOTAv1.5 με λίγες παραπάνω παραμέτρους. Τέλος το Oriented-RCNN δεν καταφέρνει ισάξιες επιδόσεις με τα YOLOv8 και ReDet ενώ έχει αρκετές παραπάνω παραμέτρους. Δεδομένου ότι οι ανιχνευτές δύο σταδίων γενικά είναι πιο αργοί στην διαδικασία της πρόβλεψης και γενικότερα χρειάζονται περισσότερες παραμέτρους, μάλλον για την συγκεκριμένη εφαρμογή που πραγματεύεται η παρούσα εργασία δεν είναι η κατάλληλη επιλογή. Βέβαια αυτό ισχύει στην γενική περίπτωση καθώς το ReDet πετυχαίνει ισάξια αποτελέσματα με τους ανιχνευτές ενός σταδίου και αν ο αριθμός των παραμέτρων δεν ήταν τόσο σημαντικός σε κάποια εφαρμογή τότε θα ήταν εξίσου καλή επιλογή.

Τέλος όπως φαίνεται και από τις προβλέψεις που παρουσιάστηκαν στο προηγούμενο κεφάλαιο όλα τα μοντέλα έχουν κάποιες ελλείψεις και πολλές φορές αδυνατούν να εντοπίσουν όλα τα αντικείμενα, ιδιαίτερα όταν τα αντικείμενα δεν είναι ευδιάκριτα η αποκρύπτονται

μερικώς.

6.2 Μελλοντικές Επεκτάσεις

Ενώ στην παρούσα εργασία εξετάστηκαν πολλά μοντέλα με διαφορες αρχιτεκτονικές και δύο σύνολα δεδομένων που είναι αντιπροσωπευτικά της ερευνητικής περιοχής του αντικειμένου της εργασίας υπάρχουν ακόμα πολλές επεκτάσεις που θα μπορούσαν να γίνουν. Αρχικά θα μπορούσαν να ελεγχθούν περισσότερα μοντέλα ή να δημιουργηθούν και νέα συνδυάζοντας τα καλύτερα επιμέρους στοιχεία των μοντέλων όπω για παράδειγμα το backbone του ReDet (ReResNet) να αντικαταστήσει το backbone του YOLOv8. Επίσης θα μπορούσε να αναλυθεί η επίδοση των μοντέλων με την χρήση διαφορετικών συναρτήσεων απώλειας για να βρεθεί η βέλτιστη για την συγκεκριμένη εφαρμογή και γενικότερα θα μπορούσε να ερευνηθεί η χρήση διαφορετικών υπερ-παραμέτρων για τα μοντέλα ώστε να βρεθούν οι βέλτιστες για το συγκεκριμένο πρόβλημα.

Βιβλιογραφία

[1] Shaoqing Ren, Kaiming He, Ross Girshick και Jian Sun. *Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks*, 2016.

[2] Glenn Jocher, Ayush Chaurasia και Jing Qiu. *Ultralytics YOLO*, 2023.

[3] Szuzina Fazekas, Bettina Budai, Róbert Stollmayer, Pál Kaposi και Viktor Bérczi. *Artificial intelligence and neural networks in radiology - Basics that all radiology residents should know*. *Imaging*, 14:73–81, 2022.

[4] Shiyi Tang, Shu Zhang και Yini Fang. *HIC-YOLOv5: Improved YOLOv5 For Small Object Detection*, 2023.

[5] Chien Yao Wang και Hong Yuan Mark Liao. *YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information*. 2024.

[6] Ao Wang, Hui Chen, Lihao Liu και et al. *YOLOv10: Real-Time End-to-End Object Detection*. *arXiv preprint arXiv:2405.14458*, 2024.

[7] Jiaming Han, Jian Ding, Nan Xue και Gui Song Xia. *ReDet: A Rotation-equivariant Detector for Aerial Object Detection*, 2021.

[8] Xingxing Xie, Gong Cheng, Jiabao Wang, Xiwen Yao και Junwei Han. *Oriented R-CNN for Object Detection*, 2021.

[9] Pengfei Zhu, Longyin Wen, Dawei Du, Xiao Bian, Heng Fan, Qinghua Hu και Haibin Ling. *Detection and tracking meet drones challenge*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11):7380–7399, 2021.

[10] Jian Ding, Nan Xue, Gui Song Xia, Xiang Bai, Wen Yang, Michael Yang, Serge Belongie, Jiebo Luo, Mihai Datcu, Marcello Pelillo και Liangpei Zhang. *Object Detection in Aerial Images: A Large-Scale Benchmark and Challenges*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, σελίδες 1–1, 2021.

[11] Natalia Salpea, Paraskevi Tzouveli και Dimitrios Kollias. *Medical Image Segmentation: A Review of Modern Architectures*. *Computer Vision - ECCV 2022 Workshops*Leonid Karlinsky, Tomer Michaeli και Ko Nishino, επιμελητές, σελίδες 691–708, Cham, 2023. Springer Nature Switzerland.

[12] Georgios Sapountzakis, Paraskevi Antonia Theofilou και Paraskevi Tzouveli. *Covid-19 Detection From X-Rays Images Using Deep Learning Methods*. *2023 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW)*, σελίδες 1–5, 2023.

[13] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser και Illia Polosukhin. *Attention Is All You Need*, 2023.

[14] Paul Viola και Michael Jones. *Rapid Object Detection using a Boosted Cascade of Simple Features*. τόμος 1, σελίδες I–511, 2001.

[15] N. Dalal και B. Triggs. *Histograms of oriented gradients for human detection*. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, τόμος 1, σελίδες 886–893 ολ. 1, 2005.

[16] Nilamani Bhoi και Mihir Mohanty. *Template Matching based Eye Detection in Facial Image*. *International Journal of Computer Applications*, 12, 2010.

[17] Ross Girshick, Jeff Donahue, Trevor Darrell και Jitendra Malik. *Rich feature hierarchies for accurate object detection and semantic segmentation*, 2014.

[18] Ross Girshick. *Fast R-CNN*, 2015.

[19] Joseph Redmon, Santosh Divvala, Ross Girshick και Ali Farhadi. *You Only Look Once: Unified, Real-Time Object Detection*, 2016.

[20] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville και Yoshua Bengio. *Generative Adversarial Networks*, 2014.

[21] Eleftherios Lymperopoulos, Paraskevi Tzouveli και Stefanos Kollias. *Satellite image super-resolution for forest localization*. *2023 International Conference on Machine Intelligence for GeoAnalytics and Remote Sensing (MIGARS)*, τόμος 1, σελίδες 1–4, 2023.

[22] Vasileios Karampinis, Anastasios Arsenos, Orfeas Filippopoulos, Evangelos Petrongonas, Christos Skliros, Dimitrios Kollias, Stefanos Kollias και Athanasios Voulodimos. *Ensuring UAV Safety: A Vision-only and Real-time Framework for Collision Avoidance Through Object Detection, Tracking, and Distance Estimation*, 2024.

[23] Karen Simonyan και Andrew Zisserman. *Very Deep Convolutional Networks for Large-Scale Image Recognition*, 2015.

[24] Kaiming He, Xiangyu Zhang, Shaoqing Ren και Jian Sun. *Deep Residual Learning for Image Recognition*, 2015.

[25] Glenn Jocher. *Ultralytics YOLOv5*, 2020.

[26] Joseph Redmon και Ali Farhadi. *YOLOv3: An Incremental Improvement*, 2018.

[27] Chien Yao Wang, Hong Yuan Mark Liao, I Hau Yeh, Yueh Hua Wu, Ping Yang Chen και Jun Wei Hsieh. *CSPNet: A New Backbone that can Enhance Learning Capability of CNN*, 2019.

[28] Shu Liu, Lu Qi, Haifang Qin, Jianping Shi και Jiaya Jia. *Path Aggregation Network for Instance Segmentation*, 2018.

[29] Chien Yao Wang, Hong Yuan Mark Liao και I Hau Yeh. *Designing Network Design Strategies Through Gradient Path Analysis*, 2022.

[30] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov και Sergey Zagoruyko. *End-to-End Object Detection with Transformers*, 2020.

[31] Tsung Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan και Serge Belongie. *Feature Pyramid Networks for Object Detection*, 2017.

[32] Glenn Jocher, Ayush Chaurasia και Jing Qiu. *Ultralytics YOLO*, 2023.

[33] MMDetection Contributors. *OpenMMLab Detection Toolbox and Benchmark*, 2018.

[34] MMRotate Contributors. *OpenMMLab rotated object detection toolbox and benchmark*, 2022.