



ΠΑΝΕΠΙΣΤΗΜΙΟ  
ΙΩΑΝΝΙΝΩΝ

Τμήμα Πληροφορικής και Τηλεπικοινωνιών  
Πρόγραμμα Μεταπτυχιακών Σπουδών  
Πληροφορικής και Δικτύων

Διπλωματική Εργασία

Ανίχνευση αντικειμένων από δορυφορικές εικόνες με  
χρήση τεχνικών μηχανικής μάθησης

Λάμπρος Μπέλλος

Επιβλέπων καθηγητής: Πέτρος Καρβέλης

Ιωάννινα, Αύγουστος 2024

**Object Detection From Satellite Imagery  
Using Machine Learning Techniques**



ΠΑΝΕΠΙΣΤΗΜΙΟ  
ΙΩΑΝΝΙΝΩΝ

Ανίχνευση αντικειμένων από δορυφορικές εικόνες με  
χρήση τεχνικών μηχανικής μάθησης

Λάμπρος Μπέλλος

**Επιτροπή Επίβλεψης Διπλωματικής Εργασίας**

Επιβλέπων Καθηγητής: Πέτρος Καρβέλης

Επίκουρος καθηγητής

Μέλος Επιτροπής: Χρήστος Γκόγκος

Καθηγητής

Μέλος Επιτροπής:

Νικόλαος Γιαννακέας

Αναπληρωτής Καθηγητής

Ιωάννινα, Αύγουστος 2024

*«Θέλω να ευχαριστήσω τους φίλους και συναδέλφους Γιάννη, Θάλεια, Ρούλα για την παρέα σε αυτό το ταξίδι καθώς και τον συνάδελφο Δημήτρη για τις συμβουλές του στην εκπόνηση αυτής της εργασίας.»*

# Περιεχόμενα

Περιεχόμενα.....	5
Κατάλογος Εικόνων / Σχημάτων.....	8
Συνοτομογραφίες & Ακρωνύμια .....	10
Περίληψη.....	1
Abstract .....	2
1. Βασικά στοιχεία των εικόνων.....	3
1.1 Εικόνες Ασπρόμαυρες – Έγχρωμες .....	3
2. Μηχανική όραση .....	5
2.1 Εισαγωγή στη Μηχανική Όραση .....	5
2.2 Μηχανική Όραση - Λειτουργία .....	6
2.3 Συνελκτικά Νευρωνικά Δίκτυα.....	8
2.3.1 Αρχιτεκτονική Συνελκτικών Νευρωνικών Δικτύων .....	8
2.3.2 Συνέλιξη και ομαδοποίηση .....	9
3. Διαφορά μεταξύ μηχανικής όρασης και επεξεργασία εικόνας .....	11
4. Παραδείγματα μηχανικής όρασης .....	12
4.1 Ταξινόμηση εικόνας .....	12
4.2 Ανίχνευση αντικειμένων .....	13
4.3 Κατάτμηση εικόνας .....	13
4.4 Παρακολούθηση κινούμενων αντικειμένων .....	13
4.5 Ανάκτηση εικόνων.....	13
4.6 Αποκατάσταση εικόνας.....	14
5. Εφαρμογές μηχανικής όρασης.....	15
5.1 Αυτόνομα οχήματα .....	15
5.2 Google Translate App .....	15
5.3 Αναγνώριση προσώπου .....	15
5.4 Υγειονομική περίθαλψη.....	15
5.5 Γεωργία.....	16
5.6 Βιομηχανία .....	16
5.7 OCR.....	16
6. Ανίχνευση αντικειμένων (Object detection) .....	17
6.1 Ανίχνευση αντικειμένων σε εικόνα και Ταξινόμηση εικόνας .....	17
6.2 Ανίχνευση αντικειμένων και τμηματοποίηση εικόνας .....	18

6.3 Ανίχνευση αντικειμένων (Object detection), Μηχανική μάθηση (Machine learning) και Βαθιά μάθηση (Deep learning) .....	19
6.3.1 Πώς λειτουργεί η ανίχνευση αντικειμένων .....	20
6.3.2 Επιλογή περιοχής (Εντοπισμός) .....	21
6.3.3 Εξαγωγή χαρακτηριστικών .....	21
6.3.4 Ταξινόμηση.....	21
7. Αλγόριθμοι ανίχνευσης αντικειμένων .....	22
7.1 Ανίχνευση αντικειμένων ενός σταδίου .....	23
7.2 Ανιχνευτές δύο σταδίων.....	23
7.3 Σύγκριση αλγορίθμων ανίχνευσης αντικειμένων .....	25
7.3.1 Αξιολόγηση επιδόσεων μοντέλων ανίχνευσης αντικειμένων .....	25
7.3.2 Intersection over Union (IoU).....	25
7.3.3 Μέση ακρίβεια (mAP) .....	28
7.3.4 Σύνολα δεδομένων.....	29
8. YOLO (You Only Look Once) .....	32
8.1 Εισαγωγή .....	32
8.2 Ανεπίσημες εκδόσεις του YOLO .....	32
8.3 Λειτουργία.....	34
8.4 YOLOv7 .....	36
8.5 Απόδοση ανίχνευσης αντικειμένων YOLOv7 .....	37
8.6 YOLOv8 .....	39
9. Ultralytics.....	40
10. Δορυφορικές Εικόνες .....	43
10.1 Τύποι δορυφορικών εικόνων .....	43
10.2 Χρήση δορυφορικών εικόνων.....	45
10.3 Λήψη δορυφορικών εικόνων .....	46
10.4 Ανίχνευση αντικειμένων σε δορυφορικές εικόνες .....	46
10.5 Προκλήσεις στις δορυφορικές εικόνες .....	47
11. Προσανατολισμένα πλαίσια οριοθέτησης (Oriented Bounding Boxes) .....	48
11.1 Ανίχνευση αντικειμένων με προσανατολισμένα πλαίσια οριοθέτησης (Oriented Bounding Boxes).....	48
11.2 Σύνολα δεδομένων προσανατολισμένων πλαισίων οριοθέτησης (OBB).....	49
12. Σύνολο Δεδομένων DOTA .....	50
(Dataset for Object deTection in Aerial images) .....	50
12.1 Υποστηριζόμενα σύνολα δεδομένων.....	50
12.2 Δομή συνόλου δεδομένων.....	51

12.3 Εφαρμογές.....	52
12.4 Dataset YAML .....	52
13. Εκπαίδευση (Training).....	54
13.1 Μετατροπή της μορφής του συνόλου δεδομένων DOTA σε Μορφή YOLO OBB .....	54
13.2 Διαίρεση των εικόνων .....	54
13.3 Κυρίως Εκπαίδευση (Training) .....	55
14. Αποτελέσματα .....	57
14.1 Μετρήσεις ανίχνευσης αντικειμένων .....	57
14.2 Μετρήσεις ανά κλάση .....	58
14.3 Μετρήσεις ταχύτητας.....	59
14.4 Οπτικές έξοδοι.....	59
14.5 Επικύρωση (Validation) .....	61
14.6 Πρόβλεψη (Predict).....	66
15. Συμπεράσματα .....	82
Βιβλιογραφία .....	84

## Κατάλογος Εικόνων / Σχημάτων

Εικόνα 1: Τιμές φωτεινότητας εικόνας.....	3
Εικόνα 2: Σύστημα συντεταγμένων RGB.....	4
Εικόνα 3: Μοντέλο χρωμάτων σε κλίμακα του γκρι.....	5
Εικόνα 4: Εικόνα Από Την Πλευρά Του Η/Υ.....	8
Εικόνα 5: Αρχιτεκτονική Συνελκτικού Νευρωνικού Δικτύου.....	9
Εικόνα 6: Παραδείγματα Μηχανικής Όρασης.....	12
Εικόνα 7: Διαφορά ανίχνευσης και ταξινόμησης.....	17
Εικόνα 8: Τμηματοποίηση Στιγμιότυπων.....	18
Εικόνα 9: Βαθιά Μάθηση vs Μηχανική Μάθηση.....	19
Εικόνα 10: Αλγόριθμοι ανίχνευσης αντικειμένων.....	24
Εικόνα 11: IoU.....	27
Εικόνα 12: Απόδοση αλγορίθμων με βάση την ακρίβεια (AP).....	30
Εικόνα 13: Απόδοση αλγορίθμων με βάση την ταχύτητα (inference time).....	30
Εικόνα 14: Απόδοση αλγορίθμων στο σύνολο δεδομένων MS COCO.....	31
Εικόνα 15: YOLOv8.....	31
Εικόνα 16: Χρονοδιάγραμμα Ημερομηνιών Κυκλοφορίας.....	33
Εικόνα 17: Νευρωνικό Δίκτυο Συνέλιξης (CNN).....	34
Εικόνα 18: Λειτουργία Αλγορίθμου YOLO.....	35
Εικόνα 19: Σύγκριση του YOLOv7 με Άλλους Ανιχνευτές στην Μέση Ακρίβεια (AP) και στους Χρόνους Συμπερασμάτων.....	38
Εικόνα 18: Σύγκριση του YOLOv8 με Προηγούμενες Εκδόσεις YOLO στην Μέση Ακρίβεια και στην Ταχύτητα.....	39
Εικόνα 19: Ορατή Εικόνα.....	43
Εικόνα 20: Υπέρυθρη Εικόνα.....	44
Εικόνα 21: Εικόνα Υδρατμών.....	45
Εικόνα 22: Δορυφορική Εικόνα Από Το Google Earth.....	46
Εικόνα 23: Λειτουργία Μοντέλου YOLO-OBb.....	49
Εικόνα 24: Ενάεριες Εικόνες.....	51
Εικόνα 25: Περιεχόμενα Αρχείου YAMl.....	52
Εικόνα 26: Κώδικας μετατροπής δεδομένων από DOTA σε OBb.....	54
Εικόνα 27: Διαχωρισμός εικόνων σε εικόνες ανάλυσης 1024x1024.....	54
Εικόνα 28: Εκπαίδευση μοντέλου.....	55
Εικόνα 29: Πίνακας σύγχυσης.....	60
Εικόνα 30: Επικύρωση της ακρίβειας του εκπαιδευμένου μοντέλου YOLOv8x-obb.....	61
Εικόνα 31: Κανονικοποιημένος πίνακας σύγχυσης για το μοντέλο YOLOv8x-obb.....	62
Εικόνα 32: Καμπύλη ακρίβειας (Μοντέλο YOLOv8x-obb).....	63
Εικόνα 33: Καμπύλη ανάκλησης (Μοντέλο YOLOv8x-obb).....	64
Εικόνα 34: Καμπύλη ακρίβεια-ανάκλησης (Μοντέλο YOLOv8x-obb).....	64
Εικόνα 35: Καμπύλη F1 (Μοντέλο YOLOv8x-obb).....	65
Εικόνα 36: Πρόβλεψη σε εικόνα.....	66
Εικόνα 37: Πρόβλεψη σε εικόνα.....	67
Εικόνα 38: Πρόβλεψη σε εικόνα.....	68
Εικόνα 39: Πρόβλεψη σε εικόνα.....	69
Εικόνα 40: Πρόβλεψη σε εικόνα.....	70
Εικόνα 41: Πρόβλεψη σε εικόνα.....	71
Εικόνα 42: Πρόβλεψη σε εικόνα.....	72



Εικόνα 43: Πρόβλεψη σε εικόνα .....	73
Εικόνα 44: Πρόβλεψη σε εικόνα .....	74
Εικόνα 45: Πρόβλεψη σε εικόνα .....	75
Εικόνα 46: Πρόβλεψη σε εικόνα .....	76
Εικόνα 47: Πρόβλεψη σε εικόνα .....	77
Εικόνα 48: Πρόβλεψη σε εικόνα .....	78
Εικόνα 49: Πρόβλεψη σε εικόνα .....	79
Εικόνα 50: Πρόβλεψη σε εικόνα .....	80
Εικόνα 51: Πρόβλεψη σε εικόνα .....	81

## **Συντομογραφίες & Ακρωνύμια**

FPS Frames Per Second

GPU Graphics Processing Unit

AI Artificial Intelligence

ML Machine Learning

DL Deep Learning

CNN Convolutional Neural Network

SSD Single Shot Detector

R-CNN Region Convolutional Neural Network

ViT Vision Transformer

SSD Single Shot Detector

YOLO You Only Look Once

NMS Non Maximum Suppression

IoU Intesection Over Union

TP True Positive

TN True Negative

FP False Positive

FN False Negative

mAP Mean Average Precision

OBB Oriented Bounding Box

## Περίληψη

Οι δορυφορικές εικόνες διαδραματίζουν κρίσιμο ρόλο στη διαμόρφωση της κατανόησής μας για τον κόσμο, παρέχοντας πληθώρα πληροφοριών που δεν μπορούν να ληφθούν από κάμερες στο επίπεδο του εδάφους.

Η δορυφορική απεικόνιση αποκτά ολοένα και μεγαλύτερη σημασία σε πολλές εφαρμογές όπως η απομακρυσμένη επιτήρηση, η περιβαλλοντική παρακολούθηση, η εναέρια έρευνα. Όλες αυτές οι εφαρμογές περιλαμβάνουν την αναζήτηση αντικειμένων, γεγονότων ενδιαφέροντος, εγκαταστάσεων από τις δορυφορικές εικόνες.

Στις περισσότερες από αυτές, η χειροκίνητη ανίχνευση και ταξινόμηση αντικειμένων καθίσταται πολύ δύσκολη, ιδίως με τους μεγάλους όγκους δεδομένων και τον αριθμό των προς επεξεργασία δορυφορικών εικόνων.

Επιπρόσθετα, τα αντικείμενα είναι μικρά και τα οπτικά χαρακτηριστικά τους είναι εξαιρετικά δύσκολα να εντοπιστούν και να καταγραφούν, γεγονός που καθιστά την προσπάθεια ακόμη πιο δυσχερή.

Για το σκοπό αυτό διάφορες αυτοματοποιημένες τεχνικές μηχανικής μάθησης για την ανίχνευση και ταξινόμηση έχουν προταθεί και βρίσκονται σε εξέλιξη. Από αυτές τις μεθόδους, οι μέθοδοι βαθιάς μάθησης με τη χρήση νευρωνικών δικτύων είναι οι πλέον διαδεδομένες τα τελευταία χρόνια.

Περιλαμβάνουν την εξαγωγή διαφόρων χαρακτηριστικών από τις εικόνες και την ταξινόμησή τους με τη χρήση ταξινομητών μηχανικής μάθησης.

Στόχος της παρούσης διπλωματικής εργασίας είναι η αξιολόγηση του αλγορίθμου YOLO και των διαφορετικών μοντέλων του, ως μια μεθοδολογία για την αναγνώριση αντικειμένων σε δορυφορικές εικόνες.

## **Abstract**

Satellite images play a crucial role in shaping our understanding of the world, providing a wealth of information that cannot be obtained from cameras at ground level.

Satellite imagery is becoming increasingly important in many applications such as remote sensing, environmental monitoring, aerial survey. All these applications involve searching for objects, events of interest, facilities from satellite images.

In most of them, manual detection and classification of objects becomes very difficult, especially with the large volumes of data and the number of satellite images to be processed.

In addition, the objects are small, and their visual characteristics are extremely difficult to detect and record, which makes the effort even more difficult.

To this end, various automated machine learning techniques for detection and classification have been proposed and are currently under development. Of these methods, deep learning methods using neural networks are the most widespread in recent years.

They involve extracting various features from images and classifying them using machine learning classifiers.

The aim of this thesis is to evaluate the YOLO algorithm and its different models as a methodology for object recognition in satellite images.

## **Λέξεις – Κλειδιά**

Satellite Images, Machine Learning, Deep Learning, Neural Network, Computer Vision, Object Detection

# 1. Βασικά στοιχεία των εικόνων

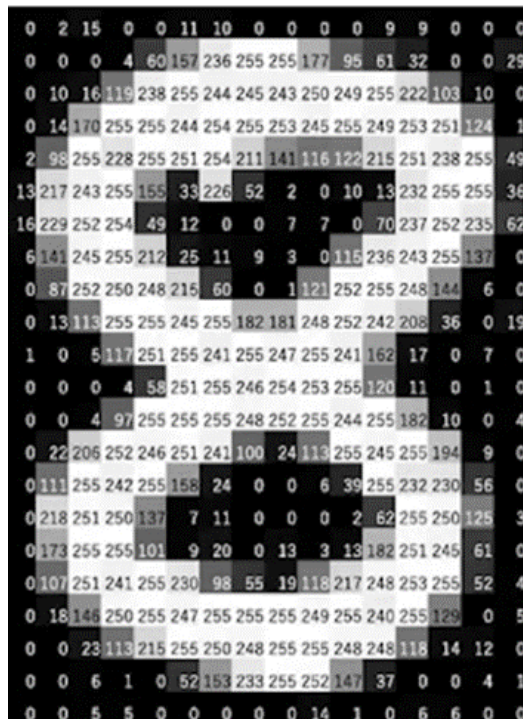
## 1.1 Εικόνες Ασπρόμαυρες – Έγχρωμες

Μια εικόνα αποτελείται από πολλά τετράγωνα στοιχεία που ονομάζονται εικονοστοιχεία. Το εικονοστοιχείο είναι η μικρότερη μονάδα μιας εικόνας, το σύνολο όλων μαζί των εικονοστοιχείων παρέχει την πληροφορία που συνιστά την πλήρη εικόνα. Ακόμα και μια μικρή εικόνα μπορεί να περιέχει εκατομμύρια τέτοιων εικονοστοιχείων διαφορετικών χρωμάτων.

Κάθε εικόνα έχει τρεις βασικές ιδιότητες:

- Μέγεθος: Αυτό είναι το ύψος και το πλάτος μιας εικόνας. Μπορεί να αναπαραχθεί σε εκατοστά, ίντσες ή ακόμα και σε εικονοστοιχεία.
- Χρωματικός χώρος: Χρωματικός χώρος είναι ένα σύνολο ή εύρος χρωμάτων διαθέσιμο για μια συσκευή και στον ψηφιακό κόσμο της απεικόνισης εκφράζει ένα πρότυπο στο προσθετικό σύστημα RGB (Red Green Blue) όπου όλα τα ορατά χρώματα προκύπτουν από τη σύνθεση των τριών βασικών παραπάνω χρωμάτων.
- Κανάλι: Αυτό είναι ένα χαρακτηριστικό του χρωματικού χώρου. Για παράδειγμα, ο χρωματικός χώρος έχει τρεις τύπους χρωμάτων ή χαρακτηριστικών γνωστών ως Κόκκινο, Πράσινο και Μπλε. Είναι σημαντικό να γίνει διάκριση μεταξύ έγχρωμων εικόνων και εικόνων κλίμακας γκρι. Μια εικόνα έχει τρία κανάλια χρώματος: κόκκινο κανάλι, πράσινο κανάλι και μπλε κανάλι. Ωστόσο, μια εικόνα σε κλίμακα γκρι έχει μόνο ένα κανάλι.

Τα χρώματα μιας εικόνας υποδηλώνονται με τις τιμές των εικονοστοιχείων της. Ένα εικονοστοιχείο μπορεί να έχει μόνο ένα χρώμα, αλλά μπορεί να συγχωνευθεί για να δημιουργήσει πολλά χρώματα.



Εικόνα 1: Τιμές φωτεινότητας εικόνας

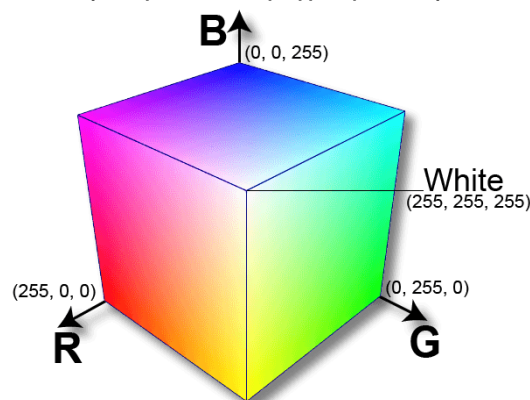
Η αναπαράσταση των χρωμάτων ως αριθμητικών τιμών είναι ένα απαραίτητο βήμα σε πολλές εφαρμογές. Για να γίνει αυτό, χρησιμοποιούμε μαθηματικά μοντέλα που περιγράφουν τρόπους αντιστοίχισης χρωμάτων σε ένα σύνολο αριθμών. Συνήθως, ένα χρωματικό μοντέλο ορίζει τρία που περιγράφονται μέσω ενός συστήματος συντεταγμένων. Κάθε χρώμα που μπορεί να αντιπροσωπεύει το μοντέλο αντιστοιχεί σε ένα σημείο αυτού του συστήματος συντεταγμένων.

Μαζί με αυτές τις τιμές, τα μοντέλα παρέχουν επίσης μια περιγραφή του τρόπου ερμηνείας αυτών των στοιχείων προκειμένου να δημιουργηθεί ένα χρώμα.

Το πιο γνωστό χρωματικό μοντέλο είναι το RGB που σημαίνει Κόκκινο-Πράσινο-Μπλε. Όπως υποδηλώνει το όνομα, αυτό το μοντέλο αντιπροσωπεύει χρώματα χρησιμοποιώντας μεμονωμένες τιμές για το κόκκινο, το πράσινο και το μπλε. Το μοντέλο RGB χρησιμοποιείται σχεδόν σε όλες τις ψηφιακές οθόνες.

Συγκεκριμένα, ένα χρώμα ορίζεται χρησιμοποιώντας τρεις ακέραιες τιμές από το 0 έως το 255 για το κόκκινο, το πράσινο και το μπλε, όπου η μηδενική τιμή σημαίνει σκούρο και η τιμή 255 σημαίνει φωτεινό. Δεδομένων των τιμών, το τελικό χρώμα ορίζεται όταν αναμειγνύουμε αυτά τα τρία βασικά χρώματα σταθμισμένα με τις τιμές τους.

Αν αναμειξουμε τα τρία χρώματα εξίσου ( $RGB = (255, 255, 255)$ ), θα έχουμε λευκό ενώ η απουσία όλων των χρωμάτων ( $RGB = (0, 0, 0)$ ) σημαίνει μαύρο. Παρακάτω υπάρχει το σύστημα συντεταγμένων RGB όπου μπορούμε να δούμε όλα τα διαφορετικά χρώματα που μπορεί να περιγράψει το μοντέλο:

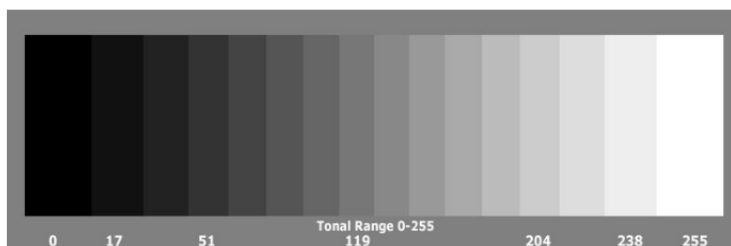


Εικόνα 2: Σύστημα συντεταγμένων RGB

Η κλίμακα του γκρι είναι το απλούστερο μοντέλο, καθώς ορίζει τα χρώματα χρησιμοποιώντας μόνο ένα στοιχείο που είναι η ένταση της φωτεινότητας. Η ποσότητα της φωτεινότητας περιγράφεται χρησιμοποιώντας μια τιμή που κυμαίνεται από 0 (μαύρο) έως 255 (λευκό) για μια εικόνα που αποθηκεύεται με χρήση 8 bit.

Από τη μία πλευρά, οι εικόνες σε κλίμακα του γκρι μεταφέρουν λιγότερες πληροφορίες από το RGB. Ωστόσο, στην επεξεργασία εικόνας η χρήση εικόνας σε κλίμακα του γκρι απαιτεί λιγότερο διαθέσιμο χώρο και είναι ταχύτερη, ειδικά όταν έχουμε να κάνουμε με πολύπλοκους υπολογισμούς.

Παρακάτω, μπορούμε να δούμε την πλήρη γκάμα χρωμάτων που μπορεί να περιγράψει το μοντέλο σε κλίμακα του γκρι:



Εικόνα 3: Μοντέλο χρωμάτων σε κλίμακα του γκρι

## 2. Μηχανική όραση

### 2.1 Εισαγωγή στη Μηχανική Όραση

Η μηχανική όραση είναι το πεδίο της τεχνητής νοημοσύνης (AI) που επιτρέπει σε υπολογιστές και συστήματα να αντλούν σημαντικές πληροφορίες από ψηφιακές εικόνες, βίντεο και να προβαίνουν σε ενέργειες ή να κάνουν συστάσεις με βάση αυτές τις πληροφορίες. Εάν η τεχνητή νοημοσύνη επιτρέπει στους υπολογιστές να σκέφτονται, η μηχανική όραση τους δίνει τη δυνατότητα να βλέπουν, να παρατηρούν και να κατανοούν. Εστιάζει στην αναπαραγωγή τμημάτων της πολυπλοκότητας του συστήματος ανθρώπινης όρασης και δίνει τη δυνατότητα στους υπολογιστές να αναγνωρίζουν και να επεξεργάζονται αντικείμενα σε εικόνες και βίντεο με τον ίδιο τρόπο που κάνουν οι άνθρωποι.

Από τη σκοπιά της μηχανικής, επιδιώκει να αυτοματοποιήσει εργασίες που μπορεί να κάνει το ανθρώπινο οπτικό σύστημα. Η μηχανική όραση ασχολείται με την αυτόματη εξαγωγή, ανάλυση και κατανόηση χρήσιμων πληροφοριών από μια μεμονωμένη εικόνα ή μια ακολουθία εικόνων. Περιλαμβάνει την ανάπτυξη μιας θεωρητικής και αλγοριθμικής βάσης για την επίτευξη αυτόματης οπτικής κατανόησης.

Χάρη στην πρόοδο στην τεχνητή νοημοσύνη και τις καινοτομίες στη βαθιά μάθηση και στα νευρωνικά δίκτυα, το πεδίο μπόρεσε να κάνει μεγάλα άλματα τα τελευταία χρόνια και κατέστη δυνατό να ξεπεράσει τους ανθρώπους σε ορισμένες εργασίες που σχετίζονται με τον εντοπισμό και την επισήμανση αντικειμένων.

Ένας από τους παράγοντες πίσω από αυτή την ανάπτυξη είναι ο μεγάλος όγκος των δεδομένων που παράγεται σήμερα και χρησιμοποιείται στη συνέχεια για την εκπαίδευση και τη βελτίωσή της. Ο κόσμος μας έχει αμέτρητες εικόνες και βίντεο μόνο από τις ενσωματωμένες κάμερες των κινητών μας συσκευών. Τα δεδομένα εικόνας μπορούν να λάβουν πολλές μορφές, όπως ακολουθίες βίντεο, προβολές από πολλές κάμερες, πολυδιάστατα δεδομένα από σαρωτή 3D ή ιατρικές συσκευές σάρωσης. Μαζί με έναν τεράστιο όγκο οπτικών δεδομένων (περισσότερες από 3 δισεκατομμύρια εικόνες μοιράζονται στο διαδίκτυο καθημερινά), η υπολογιστική ισχύς που απαιτείται για την ανάλυση των δεδομένων είναι πλέον προσβάσιμη και πιο προσιτή.

Στη Μηχανική Μάθηση (ML) και στην τεχνητή νοημοσύνη (AI), η μηχανική όραση χρησιμοποιείται για να εκπαιδεύσει το μοντέλο να αναγνωρίζει ορισμένα μοτίβα και να αποθηκεύει τα δεδομένα στην μνήμη ώστε να τα χρησιμοποιεί για την πρόβλεψη των αποτελεσμάτων στην πραγματική χρήση. Ο υπολογιστής καθοδηγείται ώστε να ερμηνεύει και να κατανοεί τις εικόνες pixel προς pixel. Ο κύριος σκοπός της χρήσης της τεχνολογίας μηχανικής όρασης σε ML και AI είναι να δημιουργηθεί ένα μοντέλο που μπορεί να λειτουργήσει μόνο του χωρίς ανθρώπινη παρέμβαση. Η όλη διαδικασία περιλαμβάνει μεθόδους απόκτησης των δεδομένων, επεξεργασίας, ανάλυσης και κατανόησης των ψηφιακών εικόνων για τη χρήση τους στο πραγματικό σενάριο. Από τεχνικής άποψης, οι υπολογιστές θα επιδιώξουν να εξάγουν οπτικά δεδομένα, να τα διαχειρίζονται και να αναλύουν τα αποτελέσματα χρησιμοποιώντας εξελιγμένα προγράμματα λογισμικού.

Οι τομείς που συνδέονται πιο στενά με την μηχανική όραση είναι η επεξεργασία και ανάλυση εικόνας. Στην επεξεργασία εικόνας, η είσοδος είναι μια εικόνα και η έξοδος είναι μια εικόνα επίσης, ενώ στην μηχανική όραση, μια εικόνα ή ένα βίντεο λαμβάνεται ως εισαγωγή και η έξοδος θα μπορούσε να είναι μια ενισχυμένη εικόνα, μια κατανόηση του περιεχομένου μιας εικόνας ή ακόμη και συμπεριφορά ενός συστήματος υπολογιστών βασισμένη σε μια τέτοια κατανόηση.

Παραδείγματα μηχανικής όρασης αποτελούν η ανίχνευση αντικειμένων, η ανίχνευση συμβάντων, η αναγνώριση αντικειμένων, η μάθηση, η κατηγοριοποίηση, η εκτίμηση κίνησης, η μοντελοποίηση 3D σκηνής, η αποκατάσταση εικόνας κ.α.

## 2.2 Μηχανική Όραση - Λειτουργία

Λειτουργεί σχεδόν το ίδιο με την ανθρώπινη όραση, εκτός από το ότι οι άνθρωποι έχουν ένα προβάδισμα. Η ανθρώπινη όραση έχει το πλεονέκτημα της διάρκειας ζωής του περιβάλλοντος για να εκπαιδεύει πώς να ξεχωρίζει τα αντικείμενα, πόσο μακριά βρίσκονται, αν κινούνται και αν υπάρχει κάτι λάθος σε μια εικόνα.

Η μηχανική όραση εκπαιδεύει τις μηχανές να εκτελούν αυτές τις λειτουργίες, αλλά πρέπει να το κάνει σε πολύ λιγότερο χρόνο με κάμερες, δεδομένα και αλγόριθμους παρά με το ανθρώπινο οπτικό σύστημα. Επειδή ένα σύστημα εκπαιδευμένο να επιθεωρεί προϊόντα ή να παρακολουθεί την παραγωγή τους μπορεί να αναλύει χιλιάδες προϊόντα ή διαδικασίες το λεπτό, παρατηρώντας ανεπαίσθητα ελαττώματα ή προβλήματα, μπορεί γρήγορα να ξεπεράσει τις ανθρώπινες δυνατότητες.

Καθώς το πεδίο της μηχανικής όρασης έχει αναπτυχθεί με νέο υλικό και αλγόριθμους, τόσο αυξάνονται τα ποσοστά ακρίβειας για την αναγνώριση αντικειμένων. Σε λιγότερο από μια δεκαετία, τα σημερινά συστήματα έχουν φτάσει στο 99 τοις εκατό ακρίβεια από 50 τοις εκατό καθιστώντας τα πιο ακριβή από τους ανθρώπους στην ταχεία αντίδραση σε οπτικές εισροές.

Τα πρώτα πειράματα στην μηχανική όραση ξεκίνησαν τη δεκαετία του 1950 και χρησιμοποιήθηκαν για πρώτη φορά εμπορικά για τη διάκριση μεταξύ



δακτυλογραφημένου και χειρόγραφου κειμένου από τη δεκαετία του 1970, σήμερα οι εφαρμογές για την μηχανική όραση έχουν αυξηθεί εκθετικά.

Η μηχανική όραση χρησιμοποιείται σε βιομηχανίες, όπως για παράδειγμα στον κλάδο της ενέργειας, τις επιχειρήσεις κοινής ωφέλειας, την κατασκευή και την αυτοκινητοβιομηχανία – ενώ η αγορά συνεχίζει να αναπτύσσεται. Αναμένεται να ξεπεράσει τα 40 δισεκατομμύρια δολάρια ΗΠΑ έως το 2030 [22].

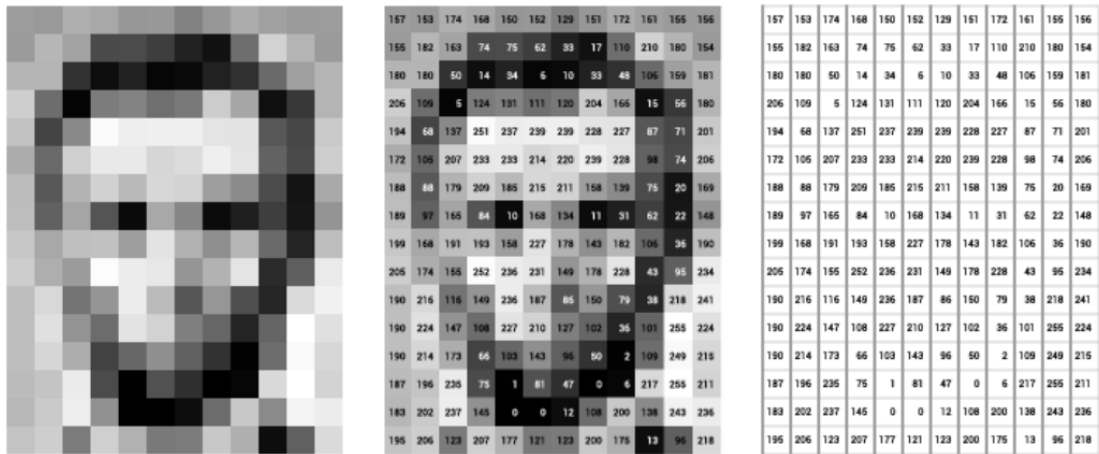
Η μηχανική όραση χρειάζεται πολλά δεδομένα. Εκτελεί αναλύσεις δεδομένων ξανά και ξανά μέχρι να διακρίνει διαφορές και τελικά να αναγνωρίσει εικόνες. Για παράδειγμα, για να εκπαιδευτεί ένας υπολογιστής ώστε να αναγνωρίζει ελαστικά αυτοκινήτου από εικόνες πρέπει να τροφοδοτηθεί με τεράστιες ποσότητες εικόνων ελαστικών και αντικειμένων που σχετίζονται με τα ελαστικά για να μάθει τις διαφορές και να αναγνωρίσει ένα ελαστικό, ειδικά ένα ελαστικό χωρίς ελαττώματα.

Η μηχανική μάθηση χρησιμοποιεί αλγοριθμικά μοντέλα που επιτρέπουν σε έναν υπολογιστή να εκπαιδεύσει τον εαυτό του σχετικά με το πλαίσιο των οπτικών δεδομένων. Εάν τροφοδοτηθούν αρκετά δεδομένα μέσω του μοντέλου, ο υπολογιστής θα αναλύσει τα δεδομένα και θα μάθει να ξεχωρίζει τη μια εικόνα από την άλλη. Τα νευρωνικά δίκτυα χρησιμοποιούν την αναγνώριση προτύπων για να διακρίνουν πολλά διαφορετικά κομμάτια μιας εικόνας. Οι αλγόριθμοι επιτρέπουν στον υπολογιστή να μαθαίνει μόνος του, αντί να τον προγραμματίζει κάποιος για να αναγνωρίσει μια εικόνα.

Οι υπολογιστές επεξεργάζονται εικόνες με τη μορφή μιας συστοιχίας pixel, όπου κάθε pixel έχει ένα σύνολο τιμών, που αντιπροσωπεύουν την παρουσία και την ένταση των τριών βασικών χρωμάτων: κόκκινο, πράσινο και μπλε.

Όλα τα εικονοστοιχεία ενώνονται για να σχηματίσουν μια ψηφιακή εικόνα. Η ψηφιακή εικόνα, επομένως, γίνεται ένας πίνακας και η μηχανική όραση γίνεται μελέτη πινάκων. Ενώ οι απλούστεροι αλγόριθμοι μηχανικής όρασης χρησιμοποιούν γραμμική άλγεβρα για να χειριστούν αυτούς τους πίνακες, οι πολύπλοκες εφαρμογές περιλαμβάνουν λειτουργίες όπως συνελίξεις με εκπαιδευόμενους πυρήνες και μείωση δειγματοληψίας μέσω ομαδοποίησης με τη χρήση ενός συνελκτικού νευρωνικού δικτύου (Convolutional Neural Network) [13].

Παρακάτω είναι ένα παράδειγμα του πώς ένας υπολογιστής «βλέπει» μια μικρή εικόνα.



Εικόνα 4: Εικόνα Από Την Πλευρά Του Η/Υ [20]

Ο υπολογιστής πρέπει να εκτελέσει σύνθετους υπολογισμούς σε αυτούς τους πίνακες και να διατυπώσει σχέσεις με γειτονικά στοιχεία pixel απλώς και μόνο για να διαπιστώσει ότι αυτή η εικόνα αντιπροσωπεύει το πρόσωπο ενός ατόμου.

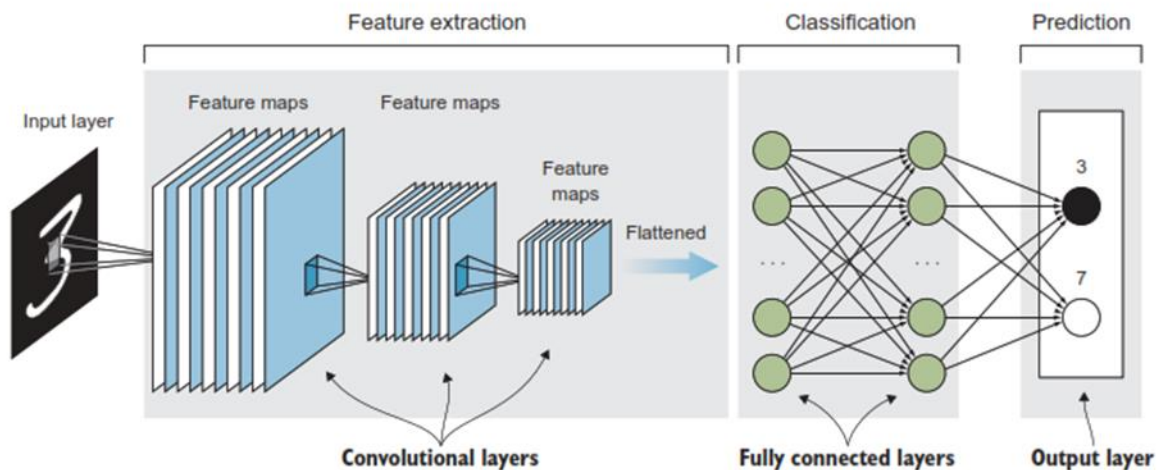
## 2.3 Συνελικτικά Νευρωνικά Δίκτυα

Τα Συνελικτικά Νευρωνικά Δίκτυα, (Convolutional Neural Networks – CNNs) αποτελούν υποκατηγορία των νευρωνικών δικτύων βαθιάς μάθησης και θεωρούνται η προτιμότερη αρχιτεκτονική νευρωνικών δικτύων για υλοποίηση εφαρμογών μηχανικής όρασης, όπως η ταξινόμηση εικόνων και η αναγνώριση αντικειμένων.

### 2.3.1 Αρχιτεκτονική Συνελικτικών Νευρωνικών Δικτύων

Η αρχιτεκτονική των CNNs βασίζεται στην αρχιτεκτονική των νευρωνικών δικτύων βαθιάς μάθησης, με τη διαφορά ότι σε αυτά δε χρησιμοποιούμε διασυνδεδεμένα κρυφά στρώματα, αλλά συνελικτικά στρώματα, τα οποία χρησιμεύουν στην εξαγωγή χαρακτηριστικών [31]. Η αρχιτεκτονική των CNN αποτελείται από τα εξής επίπεδα :

- Στρώμα Εισόδου (input layer)
- Συνελικτικά Στρώματα (convolutional layers)
- Διασυνδεδεμένα Στρώματα (connected layers)
- Στρώμα Εξόδου (output layer)



Εικόνα 5: Αρχιτεκτονική Συνελκτικού Νευρωνικού Δικτύου

Η εικόνα εισέρχεται στο στρώμα εισόδου και στη συνέχεια περνά στα συνελκτικά στρώματα για να υποστεί μείωση στις διαστάσεις της. Ακολούθως εισέρχεται στα διασυνδεδεμένα στρώματα για να ταξινομηθούν τα χαρακτηριστικά της εικόνας και τέλος, το στρώμα εξόδου του δικτύου δίνει την πρόβλεψη.

### 2.3.2 Συνέλιξη και ομαδοποίηση

Η συνέλιξη (convolution) είναι ένας πολλαπλασιασμός μεταξύ δύο πινάκων ή δύο συναρτήσεων, ο οποίος παράγει ένα τρίτο πίνακα ή μία τρίτη συνάρτηση [31]. Στα νευρωνικά δίκτυα, ο πρώτος πίνακας είναι η εικόνα εισόδου και ο δεύτερος πίνακας είναι ένα φίλτρο συνέλιξης, ο οποίος ονομάζεται φίλτρο ή πυρήνας. Ο δεύτερος (φίλτρο) είναι ένας πίνακας, με στοιχεία τις τιμές των βαρών, τα οποία αρχικοποιούνται τυχαία. Με χρήση μεθόδων βελτιστοποίησης, το CNN εκπαιδεύεται και αν χρειαστεί τις διορθώνει μέχρι να επιτευχθεί η μέγιστη βελτιστοποίηση, άρα και οι βέλτιστες τιμές των βαρών [31].

Η λογική των φίλτρων και της συνέλιξης γενικότερα, είναι η εκπαίδευση του νευρωνικού δικτύου και η ακριβέστερη εξαγωγή χαρακτηριστικών από τις εικόνες εισόδου. Με τη χρήση των φίλτρων εκπαιδεύουμε το δίκτυο να αφαιρεί τα αχρείαστα δεδομένα των εικόνων και να εστιάζει στα χρήσιμα δεδομένα – χαρακτηριστικά.

Τα στρώματα ομαδοποίησης (pooling layers) προστίθενται στα συνελκτικά στρώματα ενός CNN για να μειώσουν τις διαστάσεις τους. Η χρήση των στρωμάτων ομαδοποίησης είναι απαραίτητη σε πολύπλοκα σετ δεδομένων, στα οποία έχουν εφαρμοσθεί πολλά συνελκτικά φίλτρα, με αποτέλεσμα τα ενδιάμεσα συνελκτικά στρώματα να είναι πολύ μεγάλα.

Η πρόσθεση των στρωμάτων ομαδοποίησης ενισχύει την εξαγωγή χαρακτηριστικών από το CNN, αφού μειώνει δραστικά τις διαστάσεις των εικόνων. Η πράξη της ομαδοποίησης εφαρμόζει μία συνάρτηση μεγιστοποίησης ή μέσης τιμής στα συνελκτικά στρώματα, έτσι ώστε να μειώσει το σύνολο των

παραμέτρων που έχουν προκύψει, μειώνοντας ταυτόχρονα την πολυπλοκότητα των υπολογισμών.

Ένα συνελκτικό νευρωνικό (CNN) δίκτυο βοηθά ένα μοντέλο μηχανικής μάθησης ή βαθιάς μάθησης να «κοιτάξει» σπάζοντας τις εικόνες σε ρίχει στα οποία δίνονται ετικέτες ή επιγραφές. Χρησιμοποιεί τις ετικέτες για να εκτελέσει και κάνει προβλέψεις σχετικά με το τι «βλέπει». Το νευρωνικό δίκτυο εκτελεί συνελίξεις και ελέγχει την ακρίβεια των προβλέψεών του σε μια σειρά επαναλήψεων μέχρι να αρχίσουν να πραγματοποιούνται οι προβλέψεις. Στη συνέχεια αναγνωρίζει ή βλέπει εικόνες με τρόπο παρόμοιο με τον άνθρωπο.

Όπως ένας άνθρωπος που φτιάχνει μια εικόνα σε απόσταση, ένα CNN διακρίνει πρώτα άκρες και απλά σχήματα και στη συνέχεια συμπληρώνει πληροφορίες καθώς εκτελεί επαναλήψεις των προβλέψεών του. Ένα CNN χρησιμοποιείται για την κατανόηση μεμονωμένων εικόνων. Ένα επαναλαμβανόμενο νευρωνικό δίκτυο (RNN) χρησιμοποιείται με παρόμοιο τρόπο για εφαρμογές βίντεο για να βοηθήσει τους υπολογιστές να κατανοήσουν πώς σχετίζονται μεταξύ τους οι εικόνες σε μια σειρά καρτέ.

### **3. Διαφορά μεταξύ μηχανικής όρασης και επεξεργασία εικόνας**

Και οι δύο αποτελούν μέρος της τεχνολογίας μηχανικής μάθησης που χρησιμοποιείται κατά την επεξεργασία των δεδομένων και τη δημιουργία ενός μοντέλου.

Για παράδειγμα, η αναγνώριση αντικειμένων, η οποία είναι η διαδικασία αναγνώρισης του τύπου των αντικειμένων σε μια εικόνα, είναι ένα πρόβλημα μηχανικής όρασης. Στην μηχανική όραση, λαμβάνεται μια εικόνα ως είσοδο και μπορεί να δημιουργηθεί μια εικόνα ως έξοδος ή κάποιο άλλο είδος πληροφοριών ενώ η επεξεργασία εικόνας δεν χρειάζεται τόσο υψηλό επίπεδο κατανόησης της εικόνας.

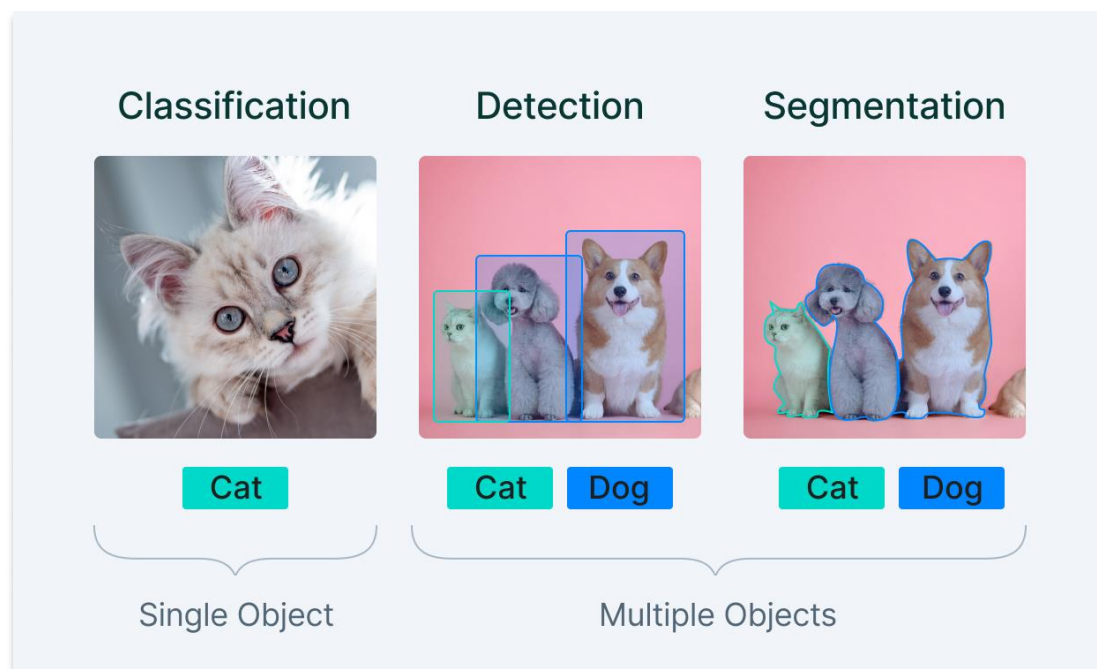
Εάν έχουμε θορυβώδεις ή θολές εικόνες, τότε στο πλαίσιο της επεξεργασίας εικόνας, η αφαίρεση του θορύβου πραγματοποιείται ώστε το αντικείμενο της εικόνας να γίνει πιο ορατό. Η επεξεργασία εικόνας περιλαμβάνει φιλτράρισμα, αφαίρεση θορύβου, ανίχνευση άκρων και επεξεργασία χρώματος. Λαμβάνεται μια εικόνα ως είσοδος και παράγεται μια άλλη εικόνα στην έξοδο που μπορεί να χρησιμοποιηθεί για την εκπαίδευση της μηχανής μέσω της μηχανικής όρασης.

Η κύρια διαφορά μεταξύ της μηχανικής όρασης και της επεξεργασίας εικόνας είναι οι στόχοι (όχι οι μέθοδοι που χρησιμοποιούνται). Για παράδειγμα, εάν ο στόχος είναι η βελτίωση της ποιότητας της εικόνας για μελλοντική χρήση, τότε ονομάζεται επεξεργασία εικόνας. Εάν ο στόχος είναι η αναγνώριση αντικειμένων, η ανίχνευση ελαττωμάτων ή η αυτόνομη οδήγηση, τότε ονομάζεται μηχανική όραση.

## 4. Παραδείγματα μηχανικής όρασης

Ουσιαστικά, οι εργασίες μηχανικής όρασης αφορούν στο να κάνουν τους υπολογιστές να κατανοούν τις ψηφιακές εικόνες καθώς και τα οπτικά δεδομένα από τον πραγματικό κόσμο. Αυτό μπορεί να περιλαμβάνει εξαγωγή, επεξεργασία και ανάλυση πληροφοριών από τέτοιες εισόδους για τη λήψη αποφάσεων.

Τα πιο δημοφιλή παραδείγματα μηχανικής όρασης στην τεχνητή νοημοσύνη περιλαμβάνουν:



Εικόνα 6: Παραδείγματα Μηχανικής Όρασης [21]

### 4.1 Ταξινόμηση εικόνας

Δεδομένης μιας ομάδας εικόνων, ο στόχος είναι να ταξινομηθούν σε ένα σύνολο προκαθορισμένων κλάσεων χρησιμοποιώντας αποκλειστικά ένα σύνολο δειγμάτων εικόνων που έχουν ήδη ταξινομηθεί. Σε αντίθεση με πολύπλοκα θέματα όπως η ανίχνευση αντικειμένων και η κατάτμηση εικόνας, τα οποία πρέπει να εντοπίζουν (ή να δίνουν θέσεις) για τα χαρακτηριστικά που ανιχνεύουν, η ταξινόμηση εικόνων ασχολείται με την επεξεργασία ολόκληρης της εικόνας στο σύνολό της και την ανάθεση μιας συγκεκριμένης ετικέτας σε αυτήν.

Η ταξινόμηση εικόνας βλέπει μια εικόνα και επιδιώκει να την ταξινομήσει (ένας σκύλος, ένα μήλο, το πρόσωπο ενός ατόμου). Πιο συγκεκριμένα, είναι σε θέση να προβλέψει με ακρίβεια ότι μια δεδομένη εικόνα ανήκει σε μια συγκεκριμένη κατηγορία. Για παράδειγμα, μια εταιρεία μέσων κοινωνικής δικτύωσης μπορεί να θέλει να το χρησιμοποιήσει για να αναγνωρίζει αυτόματα και να διαχωρίζει μη κατάλληλες εικόνες που πιθανόν θα μπορούσε να αναρτήσει κάποιος χρήστης.

## 4.2 Ανίχνευση αντικειμένων

Η ανίχνευση αντικειμένων, όπως υποδηλώνει το όνομα, αναφέρεται στον εντοπισμό αντικειμένων με τη χρήση οριοθετημένων πλαισίων. Η ανίχνευση αντικειμένων αναζητά λεπτομέρειες για μια συγκεκριμένη κατηγορία σε μια εικόνα ή ένα βίντεο και τις προσδιορίζει όποτε εμφανίζονται. Αυτές οι κατηγορίες μπορεί να είναι αυτοκίνητα, ζώα, άνθρωποι ή οτιδήποτε στο οποίο έχει εκπαιδευτεί το μοντέλο ανίχνευσης.

Προηγουμένως, οι μέθοδοι ανίχνευσης αντικειμένων χρησιμοποιούσαν διάφορες λειτουργίες για την ανίχνευση χαρακτηριστικών σε μια εικόνα και την ταξινόμηση τους με βάση τις κλασικές προσεγγίσεις μηχανικής μάθησης. Αυτή η διαδικασία, εκτός από χρονοβόρα και σε μεγάλο βαθμό ανακριβή, έχει σοβαρούς περιορισμούς στον αριθμό των αντικειμένων που μπορούν να ανιχνευθούν. Ως εκ τούτου, μοντέλα μηχανικής μάθησης όπως το YOLO, το RCNN, το SSD που χρησιμοποιούν εκατομμύρια παραμέτρους για να ξεπεράσουν αυτούς τους περιορισμούς χρησιμοποιούνται ευρέως για αυτήν την εργασία.

Η ανίχνευση αντικειμένων συχνά συνοδεύεται από την αναγνώριση αντικειμένων, γνωστή και ως ταξινόμηση αντικειμένων.

Η ανίχνευση αντικειμένων μπορεί να χρησιμοποιήσει την ταξινόμηση εικόνων για να αναγνωρίσει μια συγκεκριμένη κατηγορία εικόνας και στη συνέχεια να εντοπίσει και να καταγράψει την εμφάνισή τους σε μια εικόνα ή ένα βίντεο. Τα παραδείγματα περιλαμβάνουν τον εντοπισμό ζημιών σε μια γραμμή συναρμολόγησης ή τον εντοπισμό μηχανημάτων που απαιτούν συντήρηση.

## 4.3 Κατάτμηση εικόνας

Κατάτμηση εικόνας είναι η διαίρεση μιας εικόνας σε υποτμήματα ή υποαντικείμενα έτσι ώστε η μηχανή να μπορεί να διακρίνει ένα αντικείμενο από το φόντο ή/και άλλο αντικείμενο στην ίδια εικόνα. Ένα «τμήμα» μιας εικόνας ανήκει σε μια συγκεκριμένη κατηγορία αντικειμένου που το νευρωνικό δίκτυο έχει αναγνωρίσει, που αντιπροσωπεύεται από μια μάσκα pixel που μπορεί να χρησιμοποιηθεί για την εξαγωγή του.

## 4.4 Παρακολούθηση κινούμενων αντικειμένων

Η παρακολούθηση αντικειμένου ακολουθεί ή παρακολουθεί ένα αντικείμενο μόλις εντοπιστεί. Αυτή η εργασία εκτελείται συχνά με εικόνες που τραβήχτηκαν με τη σειρά ή με ροές βίντεο σε πραγματικό χρόνο. Τα αυτόνομα οχήματα, για παράδειγμα, δεν χρειάζεται μόνο να ταξινομήσουν και να ανιχνεύουν αντικείμενα όπως πεζούς, αλλά αυτοκίνητα και οδικές υποδομές, αλλά πρέπει να τα παρακολουθούν την κίνηση για να αποφύγουν συγκρούσεις και να τηρούν τους νόμους οδικής κυκλοφορίας.

## 4.5 Ανάκτηση εικόνων

Η ανάκτηση εικόνων με βάση το περιεχόμενο χρησιμοποιεί μηχανική όραση για περιήγηση, αναζήτηση και ανάκτηση εικόνων από μεγάλες βάσεις δεδομένων, με βάση το περιεχόμενο των εικόνων και όχι τις ετικέτες μεταδεδομένων που

σχετίζονται με αυτές. Αυτή η εργασία μπορεί να ενσωματώσει αυτόματο σχολιασμό εικόνας που αντικαθιστά τη μη αυτόματη προσθήκη ετικετών εικόνας.

#### **4.6 Αποκατάσταση εικόνας**

Η αποκατάσταση εικόνας αναφέρεται στην αποκατάσταση ή ανακατασκευή ξεθωριασμένων και παλιών αντιγράφων εικόνας που έχουν ληφθεί και αποθηκευτεί με ακατάλληλο τρόπο, οδηγώντας σε απώλεια ποιότητας της εικόνας. Οι τυπικές διαδικασίες αποκατάστασης εικόνας περιλαμβάνουν τη μείωση του πρόσθετου θορύβου μέσω μαθηματικών εργαλείων, ενώ η ανακατασκευή απαιτεί σημαντικές αλλαγές. Τα κατεστραμμένα μέρη μιας εικόνας γεμίζονται με τη βοήθεια μοντέλων παραγωγής που κάνουν μια εκτίμηση του τι προσπαθεί να μεταφέρει η εικόνα. Συχνά η διαδικασία αποκατάστασης ακολουθείται από μια διαδικασία χρωματισμού που χρωματίζει το θέμα της εικόνας (αν είναι ασπρόμαυρο) με τον πιο ρεαλιστικό δυνατό τρόπο.



## **5. Εφαρμογές μηχανικής όρασης**

Οι εφαρμογές του πραγματικού κόσμου καταδεικνύουν πόσο σημαντική είναι η μηχανική όραση στις επιχειρήσεις, την ψυχαγωγία, τις μεταφορές, την υγειονομική περίθαλψη και την καθημερινή ζωή [5]. Βασικός μοχλός για την ανάπτυξη αυτών των εφαρμογών είναι η πληθώρα οπτικών πληροφοριών από τα κινητά τηλέφωνα (smartphone), τα συστήματα ασφαλείας, τις κάμερες κυκλοφορίας και άλλες συσκευές με οπτικά όργανα. Αυτά τα δεδομένα θα μπορούσαν να διαδραματίσουν σημαντικό ρόλο στις λειτουργίες σε όλες τις βιομηχανίες, αλλά σήμερα παραμένουν αχρησιμοποίητα.

### **5.1 Αυτόνομα οχήματα**

Η μηχανική όραση είναι απαραίτητη για να ενεργοποιηθούν τα αυτόνομα αυτοκίνητα. Κατασκευαστές όπως η Tesla και άλλες χρησιμοποιούν πολλαπλές κάμερες, ραντάρ και αισθητήρες υπερήχων για τη λήψη εικόνων από το περιβάλλον, ώστε τα αυτοοδηγούμενα αυτοκίνητά τους να μπορούν να ανιχνεύουν αντικείμενα, σήμανση λωρίδας, πινακίδες και σήματα κυκλοφορίας για ασφαλή οδήγηση.

### **5.2 Google Translate App**

Για να διαβάσουμε πινακίδες σε μια ξένη γλώσσα μπορούμε να στρέψουμε την κάμερα του τηλεφώνου στις λέξεις και να χρησιμοποιήσουμε την εφαρμογή Μετάφραση Google να μας πει τι σημαίνει στη γλώσσα που επιθυμούμε. Χρησιμοποιώντας την οπτική αναγνώριση χαρακτήρων για την εικόνα και την επαυξημένη πραγματικότητα για ακριβής μετάφραση, η εφαρμογή της Google χρησιμοποιεί μηχανική όραση.

### **5.3 Αναγνώριση προσώπου**

Η αναγνώριση προσώπου είναι μια εφαρμογή της ανίχνευσης αντικειμένων όπου το κύριο αντικείμενο που ανιχνεύεται είναι το ανθρώπινο πρόσωπο. Αν και είναι παρόμοια με την ανίχνευση αντικειμένων ως εργασία, όπου τα χαρακτηριστικά εντοπίζονται, η αναγνώριση προσώπου εκτελεί όχι μόνο ανίχνευση, αλλά και αναγνώριση του προσώπου που έχει εντοπιστεί. Η αναγνώριση προσώπου χρησιμοποιείται στην καθημερινότητά μας στα κινητά τηλέφωνα αλλά και από την αστυνομία, σε συστήματα πληρωμών, σε σημεία ελέγχου ασφαλείας (π.χ. στα αεροδρόμια)

### **5.4 Υγειονομική περίθαλψη**

Δεδομένου ότι το 90 τοις εκατό όλων των ιατρικών δεδομένων βασίζονται σε εικόνες, υπάρχει πληθώρα χρήσεων για την μηχανική όραση στην ιατρική. Από την ενεργοποίηση νέων ιατρικών διαγνωστικών μεθόδων για την ανάλυση ακτινογραφιών, μαστογραφίας και άλλων σαρώσεων έως την παρακολούθηση ασθενών για την πρώιμη πρόγνωση παθήσεων και τη βοήθεια σε χειρουργικές

επεμβάσεις, αναμένεται ότι τα ιατρικά ιδρύματα και οι ασθενείς θα επωφελοούνται από την μηχανική όραση όλο και περισσότερο στο μέλλον.

## **5.5 Γεωργία**

Το 2019 παρουσιάστηκε ένα ημιαυτόνομο θεριστικό μηχάνημα που χρησιμοποιεί τεχνητή νοημοσύνη και μηχανική όραση για να αναλύσει την ποιότητα των σιτηρών κατά τη συγκομιδή και να βρει τη βέλτιστη διαδρομή μέσα από τις καλλιέργειες. Υπάρχει επίσης μεγάλη δυνατότητα για την μηχανική όραση να εντοπίσει παράσιτα έτσι ώστε τα ζιζανιοκτόνα να μπορούν να ψεκαστούν απευθείας πάνω τους αντί στις καλλιέργειες. Αυτό αναμένεται να μειώσει την ποσότητα των ζιζανιοκτόνων που απαιτούνται κατά 90 τοις εκατό.

## **5.6 Βιομηχανία**

Η μηχανική όραση βοηθά τους κατασκευαστές να λειτουργούν πιο αποτελεσματικά με διάφορους τρόπους ενώ τους δίνεται η δυνατότητα να αυτοματοποιήσουν τις διαδικασίες τους. Η προληπτική συντήρηση είναι μόνο ένα παράδειγμα όπου ο εξοπλισμός παρακολουθείται από έναν υπολογιστή ώστε να παρέμβει πριν μια βλάβη προκαλέσει χρόνο διακοπής. Η συσκευασία και η ποιότητα των προϊόντων παρακολουθούνται και τα ελαττωματικά προϊόντα μειώνονται επίσης με την μηχανική όραση. Με την ανάπτυξη τέτοιων συστημάτων, όπως η ανίχνευση ελαττωμάτων και οι επιθεωρήσεις ασφαλείας, η ποιότητα των παραγόμενων προϊόντων αυξάνεται. Επιπλέον, τα συστήματα τρισδιάστατης όρασης επιτρέπουν τη διενέργεια αποτελεσματικών επιθεωρήσεων σε μια γραμμή παραγωγής που δεν θα ήταν δυνατή από τον άνθρωπο.

## **5.7 OCR**

Μία από τις παλαιότερες εφαρμογές της όρασης υπολογιστή είναι η οπτική αναγνώριση χαρακτήρων. Με απλούς αλγόριθμους οπτικής αναγνώρισης χαρακτήρων που πειραματίζονται ήδη από το 1974, σήμερα, το OCR σήμερα με συστήματα μηχανικής μάθησης που αναπτύσσονται μπορούν να ανιχνεύουν και να μεταφράζουν κείμενο χωρίς ανθρώπινη επίβλεψη.

## 6. Ανίχνευση αντικειμένων (Object detection)

Η ανίχνευση αντικειμένων είναι το παράδειγμα της μηχανικής όρασης που ασχολείται με τον εντοπισμό και την ταξινόμηση αντικειμένων που περιέχονται σε μια εικόνα ή βίντεο. Αφορά την ανίχνευση περιπτώσεων σημασιολογικών αντικειμένων μιας συγκεκριμένης κατηγορίας (όπως άνθρωποι, κτίρια ή αυτοκίνητα) σε ψηφιακές εικόνες και βίντεο. Οι τομείς της ανίχνευσης αντικειμένων που έχουν αναπτυχθεί περιλαμβάνουν την ανίχνευση προσώπου και την ανίχνευση πεζών.

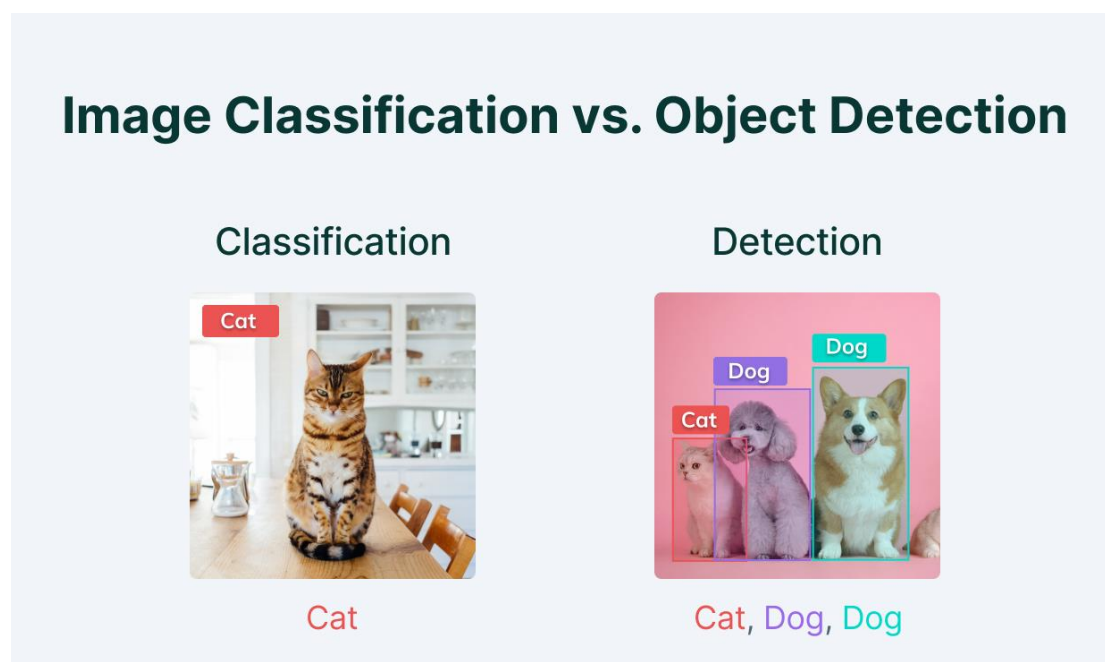
Η ανίχνευση αντικειμένων έχει εφαρμογές σε πολλούς τομείς της μηχανικής όρασης, συμπεριλαμβανομένης της ανάκτησης εικόνας και της παρακολούθησης βίντεο.

Βασίζεται στο σχέδιο οριοθέτησης πλαισίων γύρω από ανιχνευμένα αντικείμενα που μας επιτρέπουν να τα εντοπίσουμε σε μια δεδομένη σκηνή (ή πως κινούνται μέσα από αυτό).

### 6.1 Ανίχνευση αντικειμένων σε εικόνα και Ταξινόμηση εικόνας

Πρέπει εδώ να γίνει αναφορά στη διάκριση μεταξύ ταξινόμηση εικόνας και ανίχνευσης αντικειμένων σε εικόνα. Η ταξινόμηση εικόνας στέλνει μια ολόκληρη εικόνα μέσω ενός ταξινομητή (όπως ένα βαθύ νευρωνικό δίκτυο) για να παραγάγει μια ετικέτα. Οι ταξινομητές λαμβάνουν υπόψη ολόκληρη την εικόνα, αλλά δεν διαπιστώνουν πού εμφανίζεται η ετικέτα στην εικόνα.

Η ανίχνευση αντικειμένων είναι ελαφρώς πιο προηγμένη, καθώς δημιουργεί ένα πλαίσιο οριοθέτησης από το διαβαθμισμένο αντικείμενο.



Εικόνα 7: Διαφορά ανίχνευσης και ταξινόμησης [25]

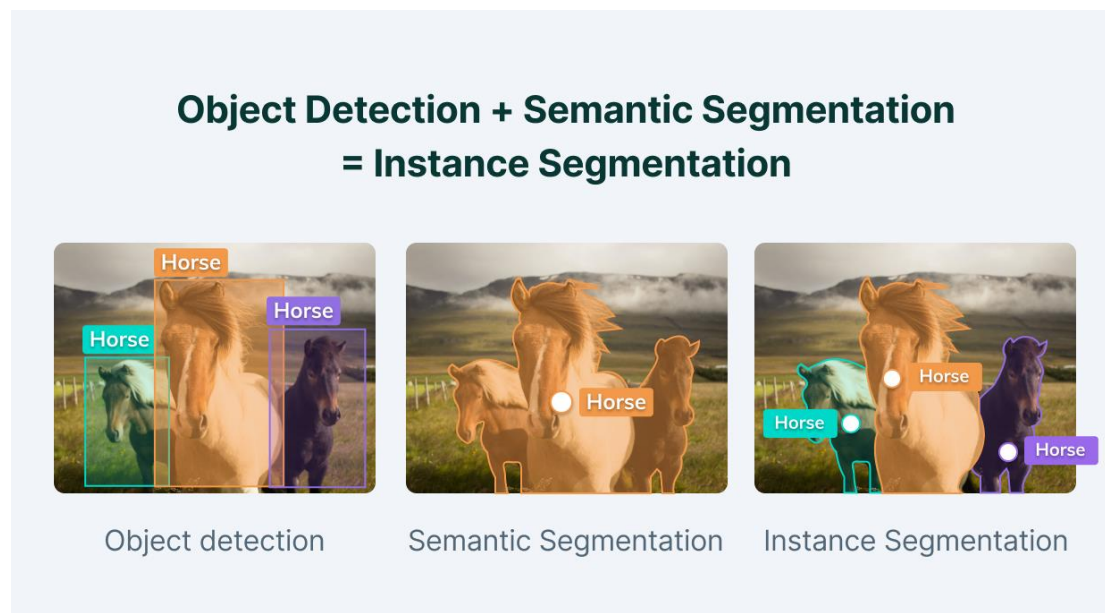
Η ταξινόμηση έχει τα πλεονεκτήματά της – είναι μια καλύτερη επιλογή για ετικέτες που δεν έχουν πραγματικά φυσικά όρια, όπως "θολή" ή "ηλιόλουστη". Ωστόσο, τα

συστήματα ανίχνευσης αντικειμένων σχεδόν πάντα ξεπερνούν τα αντίστοιχα ταξινομήσης στην ανίχνευση αντικειμένων που έχουν υλική παρουσία, όπως ένα αυτοκίνητο.

## 6.2 Ανίχνευση αντικειμένων και τμηματοποίηση εικόνας

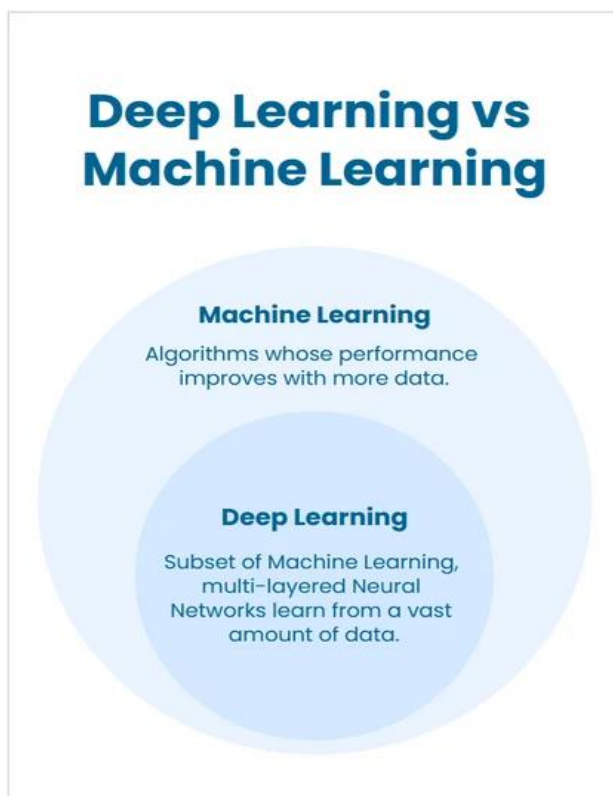
Η τμηματοποίηση εικόνας είναι η διαδικασία καθορισμού των εικονοστοιχείων μιας κατηγορίας αντικειμένων που βρίσκονται σε μια εικόνα. Η σημειακή τμηματοποίηση της εικόνας θα επισημαίνει όλα τα εικονοστοιχεία που ανήκουν σε αυτήν την ετικέτα, αλλά δεν θα καθορίσει τα όρια κάθε αντικειμένου. Αντίθετα, η ανίχνευση αντικειμένου δεν θα τμηματοποιήσει το αντικείμενο, αλλά θα καθορίσει σαφώς τη θέση κάθε μεμονωμένης παρουσίας αντικειμένου με ένα πλαίσιο.

Ο συνδυασμός της σημασιολογικής τμηματοποίησης με την ανίχνευση αντικειμένων οδηγεί σε τμηματοποίηση στιγμιότυπων, η οποία πρώτα ανιχνεύει τα στιγμιότυπα αντικειμένων και στη συνέχεια τμηματοποιεί καθένα μέσα στα ανιχνευμένα πλαίσια (γνωστά σε αυτή την περίπτωση ως περιοχές ενδιαφέροντος).



Εικόνα 8: Τμηματοποίηση Στιγμιότυπων [25]

### 6.3 Ανίχνευση αντικειμένων (Object detection), Μηχανική μάθηση (Machine learning) και Βαθιά μάθηση (Deep learning)



Εικόνα 9: Βαθιά Μάθηση vs Μηχανική Μάθηση

Μια εξαιρετικά δημοφιλής μέθοδος μηχανικής μάθησης (Machine learning) στις μέρες μας είναι η βαθιά μάθηση (Deep learning).

Η βαθιά μάθηση είναι ένα υποσύνολο της μηχανικής μάθησης που βασίζεται σε βαθιά νευρωνικά δίκτυα (Deep neural networks). Έχει συγκεκριμένα πλεονεκτήματα έναντι άλλων παραδοσιακών μορφών μηχανικής μάθησης, καθιστώντας την πιο δημοφιλή αλγοριθμική τεχνολογία της τρέχουσας εποχής. Η μηχανική μάθηση χρησιμοποιεί αλγόριθμους των οποίων η απόδοση βελτιώνεται με αυξανόμενο αριθμό δεδομένων. Από την άλλη πλευρά, η βαθιά μάθηση εξαρτάται από επίπεδα, ενώ η μηχανική μάθηση εξαρτάται από τις εισόδους δεδομένων για την εκμάθηση.

Αν και τόσο η μηχανική όσο και η βαθιά μάθηση εκπαιδεύουν αλγόριθμους να μαθαίνουν από δεδομένα, οι διαδικασίες μάθησης ή εκπαίδευσης των δύο τεχνολογιών είναι διαφορετικές. Ενώ τόσο η μηχανική μάθηση όσο και η βαθιά μάθηση εκπαιδεύουν τον υπολογιστή να μαθαίνει από τα διαθέσιμα δεδομένα, οι διαφορετικές διαδικασίες εκπαίδευσης σε καθεμία παράγουν πολύ διαφορετικά αποτελέσματα. Επιπρόσθετα η βαθιά μάθηση υποστηρίζει την επεκτασιμότητα, την εποπτευόμενη και χωρίς επίβλεψη μάθηση και τη διαστρωμάτωση της γνώσης, καθιστώντας την μια από τις πιο ισχυρές «επιστήμες μοντελοποίησης» για μηχανές εκπαίδευσης.

Χρησιμοποιεί τεχνικές για να λύσει προβλήματα πραγματικού κόσμου αξιοποιώντας νευρωνικά δίκτυα που προσομοιώνουν την ανθρώπινη λήψη αποφάσεων. Ως εκ τούτου, εκπαιδεύει τη μηχανή να κάνει αυτό που κάνει ο ανθρώπινος εγκέφαλος.

Χαρακτηρίζεται από την πολυεπίπεδη δομή της, η οποία αποτελεί το θεμέλιο των τεχνητών νευρωνικών δικτύων. Κάθε επίπεδο προσθέτει στη γνώση του προηγούμενου επιπέδου. Οι εργασίες αυτές είναι δαπανηρές, λόγω της ανάγκης για σημαντικούς υπολογιστικούς πόρους, και απαιτούν μαζικά σύνολα δεδομένων για την εκπαίδευση μοντέλων. Για τη βαθιά μάθηση, ένας τεράστιος αριθμός παραμέτρων πρέπει να κατανοηθεί από τον αλγόριθμο μάθησης, ο οποίος μπορεί αρχικά να παράγει πολλά ψευδώς θετικά αποτελέσματα.

### **6.3.1 Πώς λειτουργεί η ανίχνευση αντικειμένων**

Ο ορισμός του προβλήματος της ανίχνευσης αντικειμένου είναι να προσδιορίσει πού βρίσκονται τα αντικείμενα σε μια δεδομένη εικόνα (εντοπισμός αντικειμένου) και σε ποια κατηγορία ανήκει κάθε αντικείμενο (ταξινόμηση αντικειμένου).

Η ανίχνευση αντικειμένων μπορεί να πραγματοποιηθεί χρησιμοποιώντας είτε παραδοσιακές (1) τεχνικές επεξεργασίας εικόνας είτε σύγχρονα (2) δίκτυα βαθιάς μάθησης.

Οι τεχνικές επεξεργασίας εικόνας γενικά δεν απαιτούν ιστορικά δεδομένα για την εκπαίδευση και είναι στη φύση τους χωρίς επίβλεψη.

Οι τεχνικές αυτές δεν απαιτούν επισημασμένες εικόνες (annotated images), όπου οι άνθρωποι επισήμαναν δεδομένα χειροκίνητα (για εποπτευόμενη εκπαίδευση). Αυτές οι τεχνικές περιορίζονται από πολλούς παράγοντες, όπως πολύπλοκα σενάρια (χωρίς μονόχρωμο φόντο), μερικώς κρυμμένα αντικείμενα, φωτισμό και σκιές.

Από την άλλη πλευρά οι μέθοδοι μηχανικής μάθησης εξαρτώνται γενικά από την επιβλεπόμενη ή χωρίς επίβλεψη μάθηση, με τις εποπτευόμενες μεθόδους να αποτελούν το πρότυπο στις εργασίες μηχανικής όρασης [26]. Η απόδοση περιορίζεται από την υπολογιστική ισχύ των Graphics Processing Unit, που αυξάνεται ραγδαία χρόνο με το χρόνο.

Η ανίχνευση αντικειμένων βαθιάς μάθησης είναι σημαντικά πιο αποδοτική στα μερικώς κρυμμένα αντικείμενα, στις περίπλοκες σκηνές και στον ελλιπή φωτισμό. Απαιτείται βέβαια τεράστιος όγκος δεδομένων εκπαίδευσης. Για παράδειγμα, η επισήμανση 500.000 εικόνων για την εκπαίδευση ενός προσαρμοσμένου αλγόριθμου ανίχνευσης αντικειμένων μηχανικής μάθησης θεωρείται ένα μικρό σύνολο δεδομένων. Ωστόσο, πολλά σύνολα δεδομένων αναφοράς [10]:

- MS COCO,
- Caltech,
- KITTI,
- PASCAL
- VOC,
- V5

παρέχουν τη διαθεσιμότητα επισημασμένων δεδομένων.

Έτσι, τα παραδοσιακά μοντέλα ανίχνευσης αντικειμένων μπορούν να επιμεριστούν κυρίως σε τρία στάδια: επιλογή περιοχής, εξαγωγή χαρακτηριστικών και ταξινόμηση.

### **6.3.2 Επιλογή περιοχής (Εντοπισμός)**

Καθώς διαφορετικά αντικείμενα μπορεί να εμφανίζονται σε οποιαδήποτε θέση της εικόνας και να έχουν διαφορετικές αναλογίες διαστάσεων ή μεγέθη, είναι απαραίτητο να σαρώνεται ολόκληρη η εικόνα με ένα συρόμενο παράθυρο πολλαπλής κλίμακας. Αν και αυτή η στρατηγική μπορεί να ανακαλύψει όλες τις πιθανές θέσεις των αντικειμένων, οι αδυναμίες της είναι επίσης προφανής. Λόγω μεγάλου αριθμού υποψηφίων παραθύρων, είναι υπολογιστικά ακριβή και παράγει πάρα πολλά περιττά παράθυρα. Εάν εφαρμοστεί μόνο ένας σταθερός αριθμός, μπορεί να παραχθούν μη ικανοποιητικές περιοχές.

### **6.3.3 Εξαγωγή χαρακτηριστικών**

Για να αναγνωρίσουμε διαφορετικά αντικείμενα, πρέπει να εξαγάγουμε οπτικά χαρακτηριστικά που μπορούν να παρέχουν μια σημασιολογική και ισχυρή αναπαράσταση. Τα χαρακτηριστικά SIFT, HOG και Haar-like [4], [15] είναι αντιπροσωπευτικά.

Αυτό οφείλεται στο γεγονός ότι αυτά τα χαρακτηριστικά μπορούν να παράγουν αναπαραστάσεις σχετιζόμενες με σύνθετα κύτταρα στον ανθρώπινο εγκέφαλο. Ωστόσο, λόγω της ποικιλομορφίας των εμφανίσεων, των συνθηκών φωτισμού και του φόντου, είναι δύσκολο να σχεδιαστεί με μη αυτόματο τρόπο ένας ισχυρός περιγραφέας χαρακτηριστικών που να περιγράφει τέλεια όλα τα είδη αντικειμένων.

### **6.3.4 Ταξινόμηση**

Επιπλέον, χρειάζεται ένας ταξινομητής για να διακρίνει και να δηλώσει μια ετικέτα στην υποψήφια περιοχή.

## 7. Αλγόριθμοι ανίχνευσης αντικειμένων

Πριν η βαθιά μάθηση αρχίσει να γνωρίζει άνθηση το 2006 [2], η ανίχνευση αντικειμένων γινόταν μέσω κλασικών τεχνικών μηχανικής μάθησης. Οι τεχνικές αυτές περιλάμβαναν την τεχνική ανίχνευσης αντικειμένων viola-jones, τον HOG Detector, δημοφιλή αναλυτή χαρακτηριστικών για την ανίχνευση αντικειμένων με μηχανική όραση και επεξεργασία εικόνας καθώς και το DPM με την πρώτη παλινδρόμηση οριοθέτησης πλαισίου. Η ανίχνευση πραγματοποιείται μέσω μιας σειράς κοινών χαρακτηριστικών σε όλη την εικόνα και ταξινόμησης χρησιμοποιώντας παλινδρόμηση, ιστογραφήματα χρώματος ή δέντρα απόφασης. Οι σημερινές τεχνικές βαθιάς μάθησης ξεπερνούν σημαντικά αυτές.

Μετά το 2014, η ταχεία πρόοδος των τεχνικών βαθιάς μάθησης έχει επιταχύνει σημαντικά την βελτίωση της ανίχνευσης αντικειμένων. Με τα δίκτυα βαθιάς μάθησης και την υπολογιστική ισχύ των GPU, η απόδοση των αλγορίθμων ανίχνευσης αντικειμένων έχει βελτιωθεί σημαντικά, επιτυγχάνοντας σημαντική πρόοδο στην ανίχνευση αντικειμένων.

Οι προσεγγίσεις που βασίζονται στη βαθιά μάθηση χρησιμοποιούν αρχιτεκτονικές νευρωνικών δικτύων RetinaNet, YOLO (You Only Look Once), CenterNet, SSD (Single Shot Multibox detector), Region proposals (R-CNN, Fast-RCNN, Faster RCNN, Cascade R-CNN) [2], [10] για δυνατότητα ανίχνευσης του αντικειμένου και, στη συνέχεια, αναγνώριση σε ετικέτες.

Οι ανιχνευτές αντικειμένων των δικτύων βαθιάς μάθησης εξαγουν χαρακτηριστικά από την εικόνα εισόδου ή το πλαίσιο βίντεο.

Ένας ανιχνευτής αντικειμένων επιλύει δύο επόμενες εργασίες:

- Βρίσκει έναν αυθαίρετο αριθμό αντικειμένων (ενδεχομένως ακόμη και μηδενικό),
- Ταξινομεί κάθε μεμονωμένο αντικείμενο και κάνει εκτίμηση του μεγέθους του με ένα οριοθετημένο πλαίσιο.

Για να απλοποιηθεί η διαδικασία, μπορεί να διαχωρισθούν αυτές οι εργασίες σε δύο στάδια. Άλλες μέθοδοι συνδυάζουν και τις δύο εργασίες σε ένα βήμα (ανιχνευτές ενός σταδίου) για να επιτύχουν υψηλότερη απόδοση με κόστος όμως σε ακρίβεια στην ανίχνευση ενώ άλλοι σε δύο βήματα.

Οι αλγόριθμοι ανίχνευσης αντικειμένων ταξινομούνται ευρέως σε δύο κατηγορίες:

- Ανίχνευση αντικειμένων ενός σταδίου.
- Ανίχνευση αντικειμένων δύο σταδίων.



## 7.1 Ανίχνευση αντικειμένων ενός σταδίου

Η ανίχνευση αντικειμένων ενός σταδίου χρησιμοποιεί ένα μόνο πέρασμα της εικόνας εισόδου για να κάνει προβλέψεις σχετικά με την παρουσία και τη θέση των αντικειμένων στην εικόνα. Επεξεργάζεται μια ολόκληρη εικόνα σε ένα μόνο πέρασμα, καθιστώντας την υπολογιστικά αποδοτική.

Οι ανιχνευτές ενός σταδίου προβλέπουν πλαίσια οριοθέτησης στις εικόνες χωρίς το βήμα πρότασης περιοχής αντικειμένων. Αυτή η διαδικασία καταναλώνει λιγότερο χρόνο και επομένως μπορεί να χρησιμοποιηθεί σε εφαρμογές σε πραγματικό χρόνο.

Οι ανιχνευτές αντικειμένων ενός σταδίου δίνουν προτεραιότητα στην ταχύτητα συμπερασμάτων και είναι εξαιρετικά γρήγοροι, αλλά όχι τόσο καλοί στην αναγνώριση αντικειμένων ακανόνιστου σχήματος ή μιας ομάδας μικρών αντικειμένων.

Συμπερασματικά, η ανίχνευση αντικειμένων ενός σταδίου είναι γενικά λιγότερο ακριβής και είναι λιγότερο αποτελεσματική στην ανίχνευση μικρών αντικειμένων. Τέτοιοι αλγόριθμοι μπορούν να χρησιμοποιηθούν για την ανίχνευση αντικειμένων σε πραγματικό χρόνο σε περιβάλλοντα περιορισμένων πόρων.

Το κύριο πλεονέκτημα ενός σταδίου είναι ότι αυτοί οι αλγόριθμοι είναι γενικά ταχύτεροι από τους ανιχνευτές πολλαπλών σταδίων και δομικά πιο απλοί.

Οι πιο σημαντικοί αλγόριθμοι ανίχνευσης αντικειμένων ενός σταδίου είναι οι ακόλουθοι [3]:

- YOLO (2016)
- SSD (2016)
- RetinaNet (2017)
- YOLOv3 (2018)
- YOLOv4 (2020)
- YOLOR (2021)
- YOLOv7 (2022)
- YOLOv8 (2023)

Οι πιο δημοφιλείς ανιχνευτές ενός σταδίου περιλαμβάνουν τους YOLO, SSD και RetinaNet.

## 7.2 Ανιχνευτές δύο σταδίων

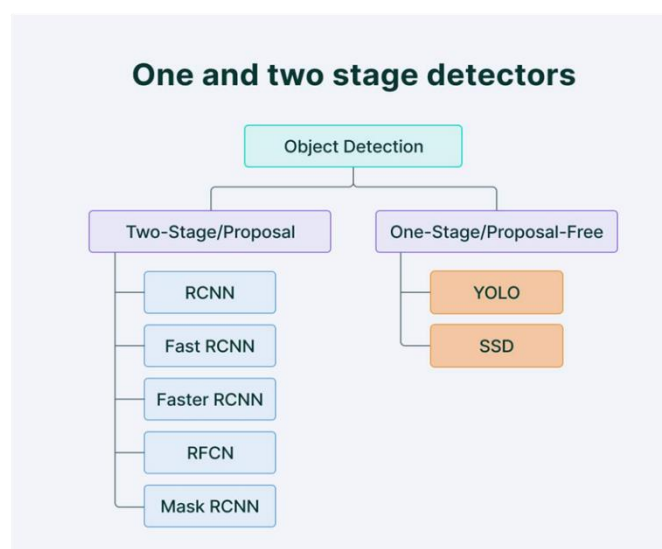
Η ανίχνευση αντικειμένων δύο σταδίων χρησιμοποιεί δύο περάσματα της εικόνας εισόδου για να κάνει προβλέψεις σχετικά με την παρουσία και τη θέση των αντικειμένων. Το πρώτο πέρασμα χρησιμοποιείται για τη δημιουργία ενός συνόλου προτάσεων ή πιθανών θέσεων αντικειμένων, και το δεύτερο πέρασμα χρησιμοποιείται για να βελτιώσει αυτές τις προτάσεις και να κάνει τελικές προβλέψεις. Αυτή η προσέγγιση είναι πιο ακριβής από την ανίχνευση αντικειμένων ενός σταδίου αλλά είναι επίσης πιο δαπανηρή υπολογιστικά.

Συνολικά, η επιλογή μεταξύ της ανίχνευσης αντικειμένων ενός σταδίου και δύο σταδίων εξαρτάται από τις συγκεκριμένες απαιτήσεις και τους περιορισμούς της εφαρμογής.

- Η αρχιτεκτονική δύο σταδίων περιλαμβάνει (1) πρόταση περιοχής αντικειμένων με συμβατικές μεθόδους μηχανικής όρασης ή δίκτυα βαθιάς μάθησης, ακολουθούμενη από (2) ταξινόμηση αντικειμένων με βάση χαρακτηριστικά που εξάγονται από την προτεινόμενη περιοχή με παλινδρόμηση οριακού πλαισίου.
- Οι μέθοδοι δύο σταδίων επιτυγχάνουν την υψηλότερη ακρίβεια ανίχνευσης αλλά είναι συνήθως πιο αργές. Λόγω των πολλών βημάτων συμπερασμάτων ανά εικόνα, η απόδοση (καρέ ανά δευτερόλεπτο) δεν είναι τόσο καλή όσο στην ανίχνευση ενός σταδίου.
- Διάφοροι ανιχνευτές δύο σταδίων περιλαμβάνουν το νευρωνικό δίκτυο περιοχής RCNN (έχει εξελιχθεί στο ταχύτερο R-CNN ή Mask R-CNN). Η τελευταία εξέλιξη είναι το RCNN (G-RCNN).
- Οι ανιχνευτές αντικειμένων δύο σταδίων βρίσκουν πρώτα μια περιοχή ενδιαφέροντος και χρησιμοποιούν αυτήν την περικυκλωμένη περιοχή για ταξινόμηση.

Οι πιο σημαντικοί αλγόριθμοι ανίχνευσης αντικειμένων δύο σταδίων [6]:

- RCNN and SPPNet (2014)
- Fast RCNN and Faster RCNN (2015)
- Mask R-CNN (2017)
- Pyramid Networks/FPN (2017)
- G-RCNN (2021)



Εικόνα 10: Αλγόριθμοι ανίχνευσης αντικειμένων

Γενικά, η ανίχνευση αντικειμένων ενός σταδίου είναι καταλληλότερη για εφαρμογές σε πραγματικό χρόνο, ενώ η ανίχνευση αντικειμένων δύο σταδίων είναι καλύτερη για εφαρμογές όπου η ακρίβεια είναι πιο σημαντική.

### 7.3 Σύγκριση αλγορίθμων ανίχνευσης αντικειμένων

#### 7.3.1 Αξιολόγηση επιδόσεων μοντέλων ανίχνευσης αντικειμένων

Για να προσδιορίσουμε και να συγκρίνουμε την προγνωστική απόδοση διαφορετικών μοντέλων ανίχνευσης αντικειμένων, χρειαζόμαστε τυποποιημένες ποσοτικές μετρήσεις.

Οι δύο πιο κοινές μετρήσεις αξιολόγησης είναι οι μετρήσεις Τομής σε Ένωση (Intersection over Union, IoU) και οι μετρήσεις Μέσης Ακρίβειας (Average Precision, AP) [6].

#### 7.3.2 Intersection over Union (IoU)

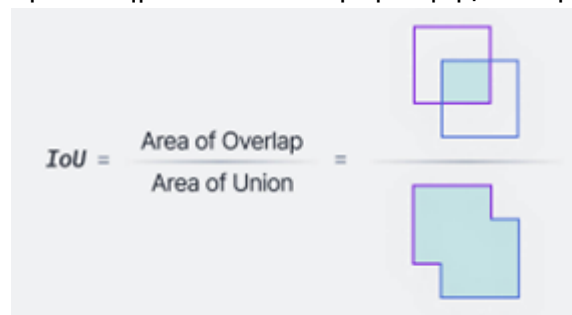
Η Τομή σε Ένωση είναι μια δημοφιλής μετρική για τη μέτρηση της ακρίβειας και τον υπολογισμό των σφαλμάτων εντοπισμού σε μοντέλα ανίχνευσης αντικειμένων.

Το πλαίσιο βασικής αλήθειας (ground truth bounding box), αντιπροσωπεύει την επιθυμητή έξοδο ενός αλγορίθμου σε μια είσοδο. Η βασική αλήθεια είναι η ιδανική έξοδος που ελπίζουμε ότι μπορεί να παράγει ο αλγόριθμός μας. Είναι επίσης το πρότυπο που ορίζεται, με βάση το οποίο αξιολογείται ένας αλγόριθμος. Όσο πιο κοντά είναι ο αλγόριθμός σας στη βασική αλήθεια τόσο το καλύτερο.

Στο πλαίσιο της παρακολούθησης αντικειμένων, η βασική αλήθεια θα αντιπροσωπεύει την "πραγματική" κατάσταση του αντικειμένου σε κάθε καρέ. Συνήθως η κατάσταση ενός αντικειμένου αναπαρίσταται από ένα οριοθετημένο ορθογώνιο που ορίζεται από ένα πλάτος, ύψος και το κέντρο του.

Για να υπολογίσουμε το IoU μεταξύ του προβλεπόμενου και του οριοθετημένου πλαισίου αλήθειας εδάφους, παίρνουμε πρώτα την περιοχή διασταύρωσης μεταξύ των δύο αντίστοιχων οριοθετημένων πλαισίων για το ίδιο αντικείμενο. Μετά από αυτό, υπολογίζουμε τη συνολική επιφάνεια που καλύπτεται από τα δύο οριοθετημένα πλαίσια- γνωστά επίσης ως "Ένωση" και την περιοχή επικάλυψης μεταξύ τους που ονομάζεται "Τομή" [24]

Η Τομή διαιρούμενη από την Ένωση μας δίνει την αναλογία της επικάλυψης προς τη συνολική επιφάνεια, παρέχοντας μια καλή εκτίμηση του πόσο κοντά είναι το οριοθετημένο πλαίσιο πρόβλεψης στο αρχικό οριοθετημένο πλαίσιο.



Για την δυαδική ταξινόμηση ο παραπάνω τύπος μπορεί να γραφεί ως εξής:

$$(IoU) = \frac{TP}{TP + FN + FP}$$

Όπου:

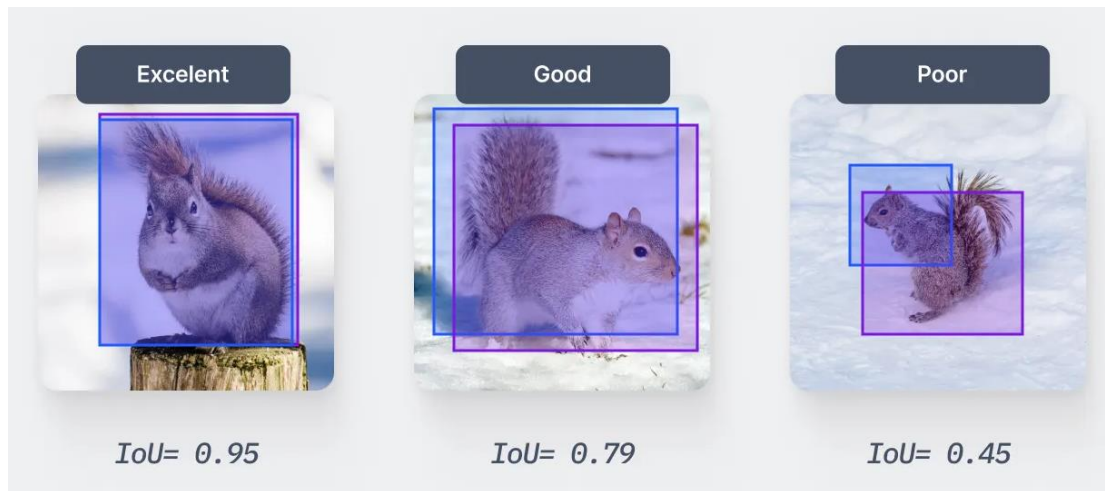
TP (Αληθώς θετικό): Το πλήθος των εικονοστοιχείων που αποτελούν το αντικείμενο σε μια εικόνα ταυτίζεται με το πλήθος των εικονοστοιχείων της βασικής αλήθειας. Το μοντέλο προέβλεψε σωστά μια ετικέτα και ταίριαξε με την βασική αλήθεια. Το μοντέλο εντόπισε τον σωστό τύπο αντικειμένου και τη σωστή θέση του.  $IoU > \text{όριο}$

TN (Αληθώς αρνητικό): Το πλήθος των εικονοστοιχείων που αποτελούν το αντικείμενο σε μια εικόνα δεν ταυτίζεται με το πλήθος των εικονοστοιχείων της βασικής αλήθειας. Το μοντέλο δεν αποτελεί μέρος της βασικής αλήθειας και ούτε προβλέπει την ετικέτα. Τα αληθώς αρνητικά αντιπροσωπεύουν το φόντο στην ανίχνευση αντικειμένων.

FP (Ψευδώς θετικό): Το πλήθος των εικονοστοιχείων που αποτελούν το αντικείμενο σε μια εικόνα ταυτίζεται με το πλήθος των εικονοστοιχείων της βασικής αλήθειας δηλ. μια ετικέτα προβλέφθηκε από το μοντέλο, ωστόσο δεν αποτελεί μέρος της βασικής αλήθειας. Το μοντέλο έχει ανιχνεύσει ένα αντικείμενο που δεν υπάρχει ή έχει δώσει μια λανθασμένη ετικέτα (Σφάλμα τύπου I).  $IoU < \text{όριο}$

FN (Ψευδώς αρνητικό): Το πλήθος των εικονοστοιχείων που αποτελούν το αντικείμενο σε μια εικόνα δεν ταυτίζεται με το πλήθος των εικονοστοιχείων της βασικής αλήθειας, δηλ. μια ετικέτα δεν έχει προβλεφθεί από το μοντέλο, αλλά αποτελεί συστατικό στοιχείο της βασικής αλήθειας. Το μοντέλο δεν εντόπισε το αντικείμενο (Σφάλμα τύπου II).

Η βαθμολογία IoU θα είναι υψηλή εάν υπάρχει μεγάλη επικάλυψη μεταξύ των προβλεπόμενων και των πεδίων της βασικής αλήθειας. Αντίθετα, μια μικρή επικάλυψη θα έχει ως αποτέλεσμα χαμηλή βαθμολογία IoU. Μια βαθμολογία IoU 1 υποδηλώνει τέλεια ταύτιση μεταξύ του προβλεπόμενου πλαισίου και του πλαισίου βασικής αλήθειας, ενώ μια βαθμολογία 0 σημαίνει ότι δεν υπάρχει επικάλυψη μεταξύ των δύο πλαισίων.



Εικόνα 11: IoU

Στο παραπάνω σχήμα, μετά τον υπολογισμό του IoU, εμφανίζονται τρεις περιπτώσεις. Στην πρώτη περίπτωση, το μοντέλο λειτουργεί αρκετά καλά, υποδεικνύοντας μεγαλύτερη ακρίβεια. Η δεύτερη περίπτωση, με IoU 0,79, το μοντέλο μας δίνει μια ικανοποιητική ανίχνευση. Τέλος, στην τρίτη περίπτωση, αποδίδει φτωχά με IoU 0,45, γεγονός που δείχνει ότι το αντικείμενο δεν ανιχνεύεται σωστά.

Η μετρική IoU είναι χρήσιμη, καθώς προσφέρει μια αριθμητική αξιολόγηση του πόσο καλά ένα μοντέλο αναγνωρίζει αντικείμενα σε μια εικόνα.

Επιπλέον, κατά την εκπαίδευση του μοντέλου μας, μπορεί να επιλεγεί μια ελάχιστη βαθμολογία IoU που απαιτείται για να θεωρηθεί ένα προβλεπόμενο πλαίσιο ως ακριβής θετική ανίχνευση, γεγονός που επιτρέπει τη χρήση της IoU για τον καθορισμό ενός ορίου για την ανίχνευση αντικειμένων. Μπορούμε να διαχειριστούμε τον συμβιβασμό μεταξύ της ακρίβειας ανίχνευσης και των ψευδώς θετικών αποτελεσμάτων επιλέγοντας το κατάλληλο όριο.

Δεν υπάρχει ένα ενιαίο συνιστώμενο όριο για την IoU, καθώς εξαρτάται σε μεγάλο βαθμό από τη συγκεκριμένη εργασία ανίχνευσης αντικειμένων και το σύνολο δεδομένων. Ωστόσο, ένα συνηθισμένο όριο που χρησιμοποιείται στην πράξη είναι το 0,5, που σημαίνει ότι ένα προβλεπόμενο πλαίσιο πρέπει να έχει IoU τουλάχιστον 0,5 με ένα πλαίσιο βασικής αλήθειας για να θεωρηθεί αληθώς θετική ανίχνευση.

Αυτό το όριο μπορεί να προσαρμοστεί με βάση τον επιθυμητό συμβιβασμό μεταξύ ακρίβειας και ανάκλησης. Για παράδειγμα, η αύξηση του ορίου θα είχε ως αποτέλεσμα λιγότερα ψευδώς θετικά αποτελέσματα, αλλά μπορεί επίσης να χάσει ορισμένα πραγματικά θετικά αποτελέσματα. Είναι σημαντικό να αξιολογήσουμε την απόδοση του μοντέλου μας σε ένα σύνολο επικύρωσης χρησιμοποιώντας διαφορετικά όρια IoU για να επιλεγεί το καταλληλότερο για την εργασία μας.

Στην ανίχνευση αντικειμένων, η IoU δεν υπολογίζει άμεσα την ακρίβεια ενός μοντέλου. Αντίθετα, είναι μια βοηθητική μετρική που αξιολογεί το βαθμό επικάλυψης μεταξύ της βασικής αλήθειας και της πρόβλεψης.

### 7.3.3 Μέση ακρίβεια (mAP)

Η ακρίβεια ορίζεται ως το ποσοστό των θετικών προβλέψεων κλάσης που ήταν πραγματικά σωστές [23]. Με άλλα λόγια, εάν ένα μοντέλο ταξινομούσε συνολικά 100 δείγματα και 70 από αυτά ανήκαν πραγματικά στη θετική κατηγορία του συνόλου δεδομένων (και 30 ήταν αρνητικά δείγματα κλάσης που προβλέπονταν λανθασμένα ως "θετικά" από τον ταξινομητή), τότε η ακρίβεια είναι 70%. Η ανάκληση ορίζεται ως το ποσοστό των πραγματικών θετικών δειγμάτων κλάσης που εντοπίστηκαν από το μοντέλο [23]. Δηλαδή αν το σύνολο δοκιμών ενός συνόλου δεδομένων αποτελείται από 100 δείγματα, πόσα από αυτά εντοπίστηκαν; Εάν τα 60 από τα θετικά δείγματα εντοπίστηκαν σωστά, τότε η ανάκληση είναι 60%.

Η μέση ακρίβεια (mAP) υπολογίζεται ως η περιοχή κάτω από μια καμπύλη ακρίβειας έναντι ανάκλησης για ένα σύνολο προβλέψεων.

Η ανάκληση υπολογίζεται ως ο λόγος των συνολικών προβλέψεων που γίνονται από το μοντέλο κάτω από μια κατηγορία με το σύνολο των υπάρχουσών ετικετών για την κατηγορία. Η ακρίβεια αναφέρεται στην αναλογία των πραγματικών θετικών σε σχέση με τις συνολικές προβλέψεις που γίνονται από το μοντέλο.

Η ανάκληση και η ακρίβεια προσφέρουν ένα συμβιβασμό που απεικονίζεται γραφικά σε μια καμπύλη με τη διαφοροποίηση του κατώτατου ορίου ταξινόμησης [6]. Η περιοχή κάτω από αυτή την καμπύλη ακρίβειας έναντι ανάκλησης μας δίνει τη μέση ακρίβεια ανά κατηγορία για το μοντέλο. Ο μέσος όρος αυτής της τιμής, που λαμβάνεται σε όλες τις κατηγορίες, ονομάζεται μέση ακρίβεια (mAP).

Στην ανίχνευση αντικειμένων, η ακρίβεια και η ανάκληση δεν χρησιμοποιούνται για προβλέψεις κλάσεων. Αντίθετα, χρησιμεύουν ως προβλέψεις οριοθετημένων πλαισίων για τη μέτρηση της απόδοσης απόφασης. Μια τιμή IoU > 0,5 λαμβάνεται ως θετική πρόβλεψη, ενώ μια τιμή IoU < 0,5 είναι αρνητική πρόβλεψη.

### 7.3.4 Σύνολα δεδομένων

Το πιο δημοφιλές σημείο αναφοράς είναι το σύνολο δεδομένων COCO της Microsoft.

Το σύνολο δεδομένων MS COCO (Common Objects in Context) είναι ένα σύνολο δεδομένων ανίχνευσης αντικειμένων μεγάλης κλίμακας, τμηματοποίησης εικόνας και υποτίτλων της Microsoft.

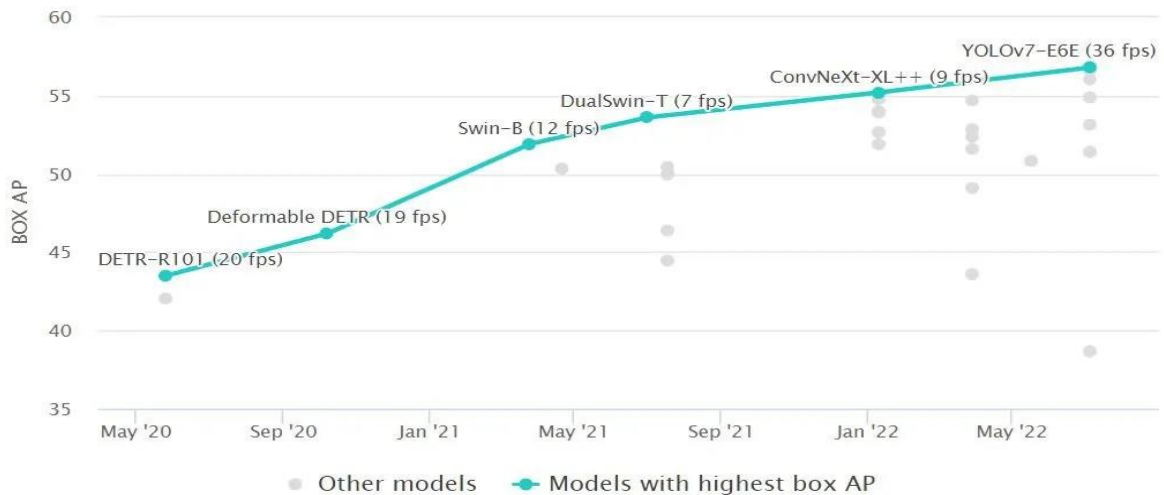
Αποτελείται από πάνω από 330.000 εικόνες, 90 διαφορετικές κατηγορίες αντικειμένων (αυτοκίνητα, άτομα, αθλητικές μπάλες, ποδήλατα, σκύλοι, γάτες, άλογα κ.λπ.). Το σύνολο δεδομένων COCO χρησιμοποιείται ευρέως στην έρευνα μηχανικής όρασης για την εκπαίδευση και την αξιολόγηση πολλών μοντέλων ανίχνευσης και τμηματοποίησης αντικειμένων. Το σύνολο δεδομένων έχει δύο κύρια μέρη: τις εικόνες και τους σχολιασμούς τους. Οι εικόνες οργανώνονται σε μια ιεραρχία καταλόγων, με τον κατάλογο ανώτατου επιπέδου να περιέχει υποκαταλόγους για την εκπαίδευση, την επικύρωση και τα σύνολα δοκιμών.

Στις μέρες μας νεότερα σύνολα δεδομένων, όπως το OpenImages, επικρατούν ως προεκπαιδευτικά σύνολα δεδομένων.

Τα σύνολα δεδομένων χρησιμοποιούνται συχνά για να συγκριθεί η απόδοση της ανίχνευσης αντικειμένων σε πραγματικό χρόνο. Η μορφή τους ερμηνεύεται αυτόματα από προηγμένες βιβλιοθήκες νευρωνικών δικτύων.

Η κατανόηση οπτικών σκηνών είναι ένας πρωταρχικός στόχος της μηχανικής όρασης: Περιλαμβάνει την αναγνώριση των αντικειμένων που υπάρχουν, τον εντοπισμό των αντικειμένων σε 2D και 3D, τον προσδιορισμό των χαρακτηριστικών του αντικειμένου και τον χαρακτηρισμό της σχέσης μεταξύ των αντικειμένων. Επομένως, οι αλγόριθμοι για την ανίχνευση και την ταξινόμηση αντικειμένων μπορούν να εκπαιδευτούν χρησιμοποιώντας το σύνολο δεδομένων. Τα διαφορετικά μοντέλα συνήθως αξιολογούνται σύμφωνα με μια μέτρηση Μέσης Ακρίβειας (AP). Στη συνέχεια, θα συγκρίνουμε τους καλύτερους αλγόριθμους ανίχνευσης αντικειμένων σε πραγματικό χρόνο. Είναι σημαντικό να σημειωθεί ότι η επιλογή του αλγορίθμου εξαρτάται από την χρήση και την εφαρμογή. Διαφορετικοί αλγόριθμοι υπερέχουν σε διαφορετικές εργασίες (π.χ., ο R-CNN δείχνει τα καλύτερα αποτελέσματα για την ανίχνευση πεζών στον δρόμο).

Στο σύνολο δεδομένων MS COCO και με βάση τη Μέση Ακρίβεια (AP), ο καλύτερος αλγόριθμος ανίχνευσης αντικειμένων σε πραγματικό χρόνο είναι ο YOLOv7, ακολουθούμενος από το Vision Transformer (ViT) όπως το Swin και το DualSwin. PP-YOLOE, YOLOR, YOLOv4 και EfficientDet.



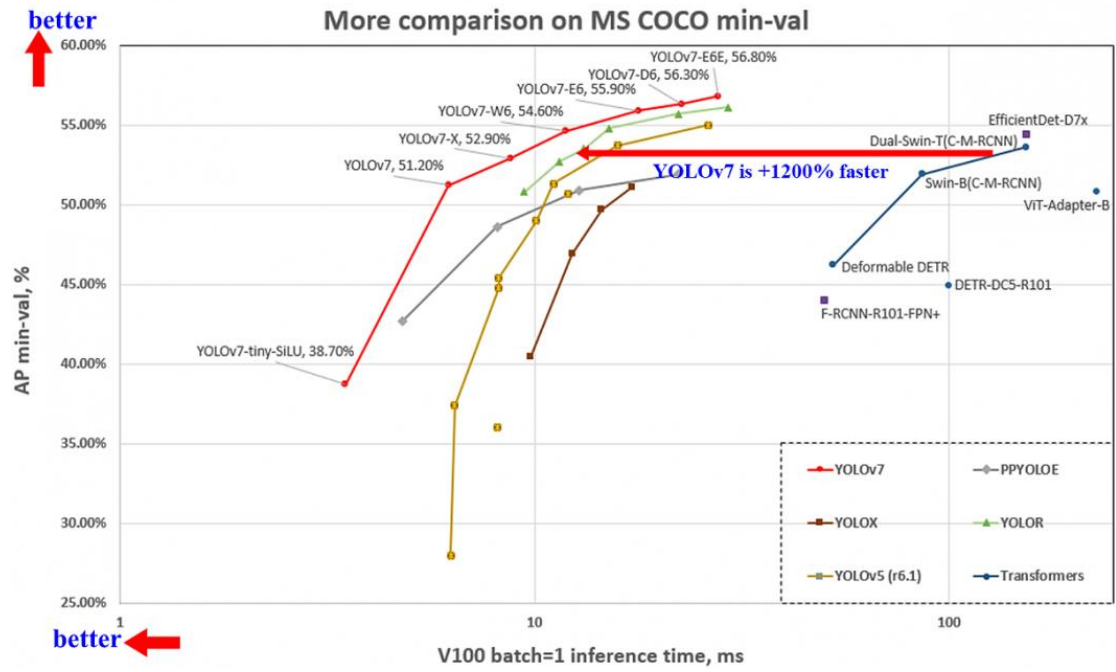
Εικόνα 12: Απόδοση αλγορίθμων με βάση την ακρίβεια (AP)

Στο σύνολο δεδομένων MS COCO, μια σημαντική μέτρηση αναφοράς είναι ο χρόνος συμπερασμάτων (ms/Frame, χαμηλότερο είναι καλύτερο) ή Frames per second (FPS, υψηλότερο είναι καλύτερο). Οι ραγδαίες πρόοδοι στην τεχνολογία μηχανικής όρασης είναι πολύ ορατές όταν εξετάζουμε συγκρίσεις χρόνου συμπερασμάτων. Με βάση τους τρέχοντες χρόνους συμπερασμάτων (ο χαμηλότερος είναι καλύτερος), το YOLOv7 επιτυγχάνει 3,5ms ανά καρέ, σε σύγκριση με το YOLOv4 12ms ή το δημοφιλές YOLOv3 29ms. Να σημειωθεί πως η εισαγωγή του YOLO (ανιχνευτής ενός σταδίου) οδήγησε σε δραματικά ταχύτερους χρόνους συμπερασμάτων σε σύγκριση με οποιεσδήποτε προηγούμενες καθιερωμένες μεθόδους, όπως η μέθοδος δύο σταδίων Mask R-CNN (333ms).

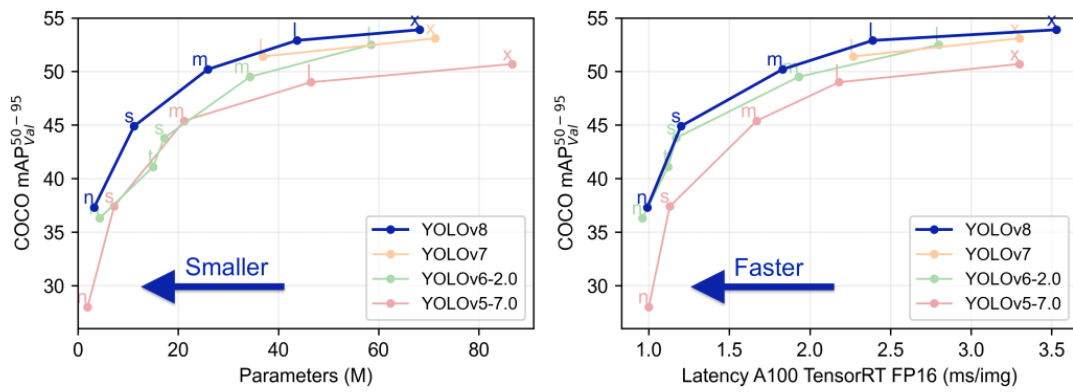


Εικόνα 13: Απόδοση αλγορίθμων με βάση την ταχύτητα (inference time)





Εικόνα 14: Απόδοση αλγορίθμων στο σύνολο δεδομένων MS COCO [6]



Εικόνα 15: YOLOv8 [6]

## 8. YOLO (You Only Look Once)

### 8.1 Εισαγωγή

Ο YOLO "You Only Look Once", είναι μια δημοφιλής οικογένεια αλγορίθμων ανίχνευσης αντικειμένων σε πραγματικό χρόνο που έχει φέρει επανάσταση στο πεδίο της μηχανικής όρασης. Είναι γρήγορος και αποτελεσματικός, καθιστώντας τον μια εξαιρετική επιλογή για εργασίες ανίχνευσης αντικειμένων σε πραγματικό χρόνο. Έχει επιτύχει επιδόσεις αιχμής σε διάφορα σημεία αναφοράς και έχει υιοθετηθεί ευρέως σε διάφορες εφαρμογές πραγματικού κόσμου.

Ένα από τα κύρια πλεονεκτήματά του είναι η γρήγορη ταχύτητα του, η οποία του επιτρέπει να επεξεργάζεται εικόνες σε πραγματικό χρόνο. Είναι κατάλληλος για εφαρμογές όπως η παρακολούθηση βίντεο, αυτόνομα οχήματα και η επαυξημένη πραγματικότητα. Επιπλέον, έχει απλή αρχιτεκτονική και απαιτεί ελάχιστα δεδομένα κατάρτισης, καθιστώντας την εφαρμογή εύκολη σε νέες εργασίες.

Παρά τους περιορισμούς όπως η δυσκολία ανίχνευσης μικρών αντικειμένων και η αδυναμία εκτέλεσης ταξινόμησης αντικειμένων σε ορισμένες περιπτώσεις, ο αλγόριθμος έχει αποδειχθεί ένα πολύτιμο εργαλείο για την ανίχνευση αντικειμένων και έχει ανοίξει πολλές νέες δυνατότητες για ερευνητές και επαγγελματίες.

Ο αρχικός αλγόριθμος YOLO κυκλοφόρησε για πρώτη φορά το 2015, δημιουργήθηκε από τους Joseph Redmon, Ali Farhadi και Santosh Divvala [6].

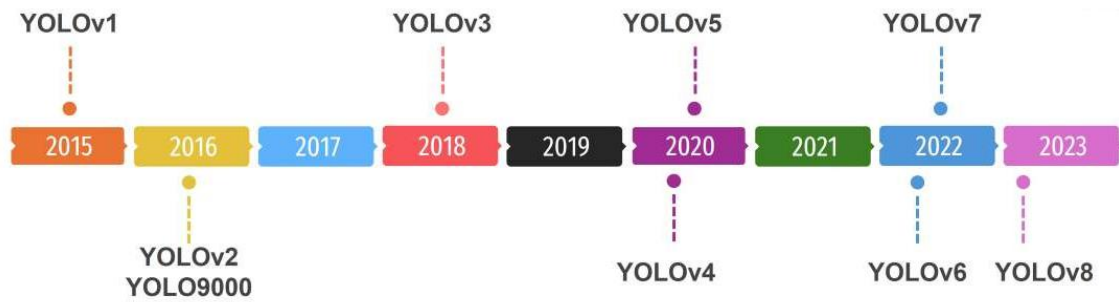
Η αρχιτεκτονική του ήταν πολύ πιο γρήγορη από άλλους ανιχνευτές αντικειμένων και έγινε τελευταίας τεχνολογίας για εφαρμογές μηχανικής όρασης σε πραγματικό χρόνο.

Αρκετές νέες εκδόσεις του ίδιου μοντέλου έχουν προταθεί από την αρχική κυκλοφορία του YOLO, καθεμία από τις οποίες βασίζεται και βελτιώνει τον προκάτοχό της παρέχοντας σημαντική αύξηση στην απόδοση και την αποδοτικότητα του αλγορίθμου. Οι εκδόσεις από το YOLOv1 στο δημοφιλές YOLOv3 δημιουργήθηκαν από τον τότε πτυχιούχο φοιτητή Joseph Redmon και τον σύμβουλο Ali Farhadi. Ο YOLOv4 εισήχθη από τον Alexey Bochkovskiy, ο οποίος συνέχισε την κληρονομιά αφού ο Redmon είχε σταματήσει την έρευνά του για την μηχανική όραση λόγω ηθικών ανησυχιών. Ο YOLOv7 είναι η τελευταία επίσημη έκδοση YOLO που δημιουργήθηκε από τους αρχικούς δημιουργούς της αρχιτεκτονικής YOLO. Αναμένεται ότι πολλά εμπορικά δίκτυα θα μετακινηθούν απευθείας από το YOLOv4 στο v7, παρακάμπτοντας όλους τους άλλες εκδόσεις.

### 8.2 Ανεπίσημες εκδόσεις του YOLO

Υπήρχαν κάποιες διαφωνίες στην κοινότητα της μηχανικής όρασης όταν άλλοι ερευνητές και εταιρείες δημοσίευσαν τα μοντέλα τους ως εκδόσεις YOLO. Ένα δημοφιλές παράδειγμα είναι ο YOLOv5 που δημιουργήθηκε από την εταιρεία Ultralytics. Ο αλγόριθμος είναι παρόμοιος με ο YOLOv4 αλλά χρησιμοποιεί ένα διαφορετικό framework, το PyTorch αντί για το DarkNet. Ωστόσο, ο δημιουργός του YOLOv4, Alexey Bochkovskiy παρείχε σημεία αναφοράς συγκρίνοντας τον YOLOv4

με τον YOLOv5, δείχνοντας ότι η έκδοση 4 είναι ίση ή καλύτερη. Ένα άλλο παράδειγμα είναι ο YOLOv6 ο οποίο δημοσιεύθηκε από την κινεζική εταιρεία Meituan (εξ ου και το πρόθεμα MT του YOLOv6). Και υπάρχει επίσης μια ανεπίσημη έκδοση που κυκλοφόρησε το χρόνο πριν από την επίσημη έκδοση YOLOv7 (υπάρχουν δύο YOLOv7). Τόσο το YOLOv5 όσο και το YOLOv6 δεν θεωρούνται μέρος της επίσημης σειράς YOLO αλλά εμπνεύστηκαν από την αρχική αρχιτεκτονική ενός σταδίου του Yolo. Οι επικριτές υποστηρίζουν ότι οι εταιρείες προσπαθούν να επωφεληθούν από την φήμη του YOLO και ότι οι αλγόριθμοι δεν εξετάστηκαν επαρκώς ή δεν δοκιμάστηκαν υπό τις ίδιες συνθήκες. Ως εκ τούτου, θεωρείται ότι το επίσημο YOLOv7 πρέπει να είναι το πραγματικό YOLOv5.

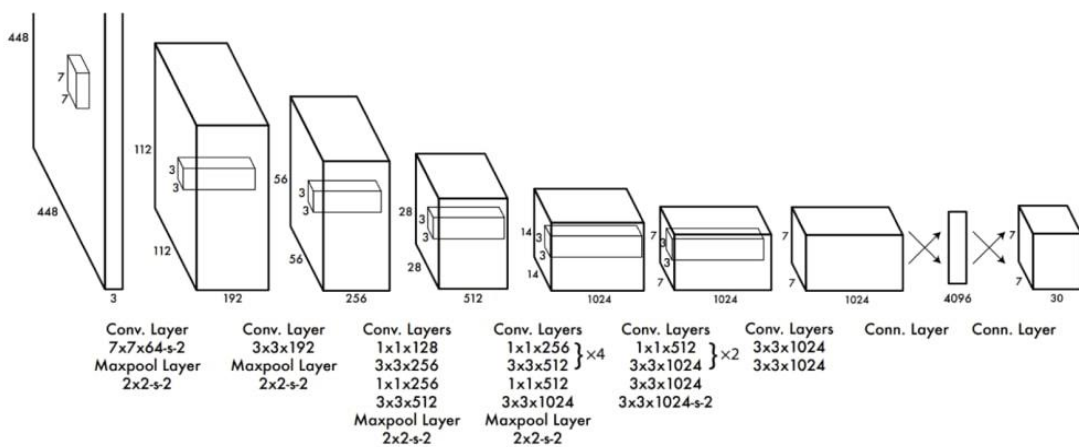


Εικόνα 16: Χρονοδιάγραμμα Ημερομηνιών Κυκλοφορίας [6]

### 8.3 Λειτουργία

Ο YOLO προτείνει τη χρήση ενός νευρωνικού δικτύου που κάνει προβλέψεις οριοθετημένων πλαισίων και πιθανοτήτων κλάσεων ταυτόχρονα. Διαφέρει από την προσέγγιση που ακολουθούσαν οι προηγούμενοι αλγόριθμοι ανίχνευσης αντικειμένων, οι οποίοι επαναχρησιμοποίησαν ταξινομητές για την εκτέλεση ανίχνευσης. Ακολουθώντας μια θεμελιωδώς διαφορετική προσέγγιση στον εντοπισμό αντικειμένων, ο YOLO πέτυχε καλύτερα αποτελέσματα, ξεπερνώντας άλλους αλγόριθμους ανίχνευσης αντικειμένων σε πραγματικό χρόνο με μεγάλη διαφορά. Ενώ αλγόριθμοι όπως το Faster RCNN λειτουργούν ανιχνεύοντας πιθανές περιοχές ενδιαφέροντος χρησιμοποιώντας ένα δίκτυο πρότασης περιοχής και στη συνέχεια πραγματοποιούν αναγνώριση σε αυτές τις περιοχές ξεχωριστά, ο YOLO εκτελεί όλες τις προβλέψεις του με τη βοήθεια ενός μόνο πλήρως συνδεδεμένου επιπέδου. Οι μέθοδοι που χρησιμοποιούν δίκτυα πρότασης περιοχής εκτελούν πολλαπλές επαναλήψεις για την ίδια εικόνα, ενώ ο YOLO χρειάζεται μία μόνο επανάληψη.

Ο αλγόριθμος YOLO δέχεται μια εικόνα ως είσοδο και στη συνέχεια χρησιμοποιεί ένα συνελικτικό νευρωνικό δίκτυο για να ανιχνεύσει αντικείμενα στην εικόνα. Η αρχιτεκτονική του μοντέλου του CNN που αποτελεί τη ραχοκοκαλιά του YOLO φαίνεται στην εικόνα 17 [11].

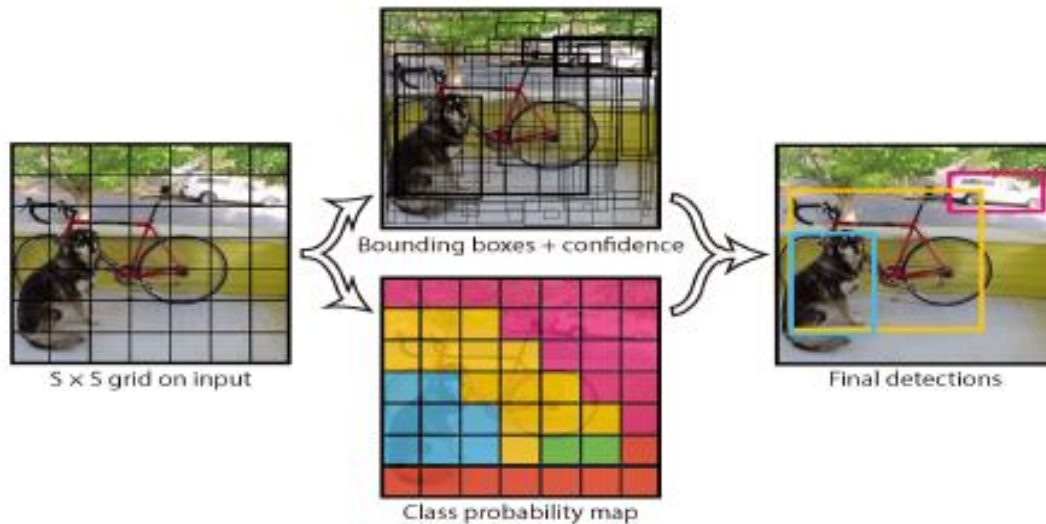


**The Architecture.** Our detection network has 24 convolutional layers followed by 2 fully connected layers. Alternating  $1 \times 1$  convolutional layers reduce the features space from preceding layers. We pretrain the convolutional layers on the ImageNet classification task at half the resolution ( $224 \times 224$  input image) and then double the resolution for detection.

Εικόνα 17: Νευρωνικό Δίκτυο Συνέλιξης (CNN)

Τα πρώτα 20 επίπεδα συνέλιξης του μοντέλου είναι προεκπαιδευμένα χρησιμοποιώντας το ImageNet (Βάση δεδομένων εικόνων) συνδέοντας ένα προσωρινό μέσο όρο συγκέντρωσης και πλήρως συνδεδεμένο επίπεδο. Στη συνέχεια, αυτό το προεκπαιδευμένο μοντέλο μετατρέπεται για ανίχνευση, καθώς προηγούμενες έρευνες έδειξαν ότι η προσθήκη συνέλιξης και συνδεδεμένων επιπέδων σε ένα προεκπαιδευμένο δίκτυο βελτιώνει την απόδοση. Το τελικό πλήρως συνδεδεμένο επίπεδο του YOLO προβλέπει και τις πιθανότητες κλάσης και τις συντεταγμένες οριοθέτησης πλαισίου.

Ο YOLO διαιρεί μια εικόνα εισόδου σε ένα πλέγμα  $S \times S$  [4], [11]. Εάν το κέντρο ενός αντικειμένου βρίσκεται σε ένα κελί πλέγματος, αυτό το κελί πλέγματος είναι υπεύθυνο για την ανίχνευση αυτού του αντικειμένου.



Εικόνα 18: Λειτουργία Αλγορίθμου YOLO

Στον αλγόριθμο YOLO, η πιθανότητα κλάσης για κάθε κελί πλέγματος και η εμπιστοσύνη για κάθε οριοθετημένο πλαίσιο υπολογίζονται ως εξής:

Πιθανότητα κλάσης για κάθε κελί πλέγματος:

- Για κάθε κελί πλέγματος, το νευρωνικό δίκτυο προβλέπει πολλαπλά οριοθετημένα πλαίσια μαζί με τις βαθμολογίες εμπιστοσύνης για αυτά τα πλαίσια. Αυτές οι βαθμολογίες εμπιστοσύνης αντικατοπτρίζουν πόσο σίγουρο είναι το μοντέλο ότι το πλαίσιο περιέχει ένα αντικείμενο και πόσο ακριβές πιστεύει ότι είναι το προβλεπόμενο πλαίσιο
- Κάθε πρόβλεψη οριοθετημένου πλαισίου περιλαμβάνει μια βαθμολογία εμπιστοσύνης για το πλαίσιο και ένα σύνολο πιθανοτήτων κλάσης.
- Η πιθανότητα κλάσης για κάθε κελί πλέγματος υπολογίζεται πολλαπλασιάζοντας τη βαθμολογία εμπιστοσύνης του πλαισίου οριοθέτησης με τις πιθανότητες κλάσης που προβλέπονται για το εν λόγω πλαίσιο.
- Η τελική πιθανότητα κλάσης για κάθε κελί πλέγματος προκύπτει από τη μέγιστη πιθανότητα κλάσης μεταξύ όλων των οριοθετημένων πλαισίων που προβλέπονται για το συγκεκριμένο κελί.

Βαθμολογία εμπιστοσύνης για κάθε οριοθετημένο πλαίσιο (όχι εμπιστοσύνη για κάθε κλάση):

- Η βαθμολογία εμπιστοσύνης για κάθε οριοθετημένο πλαίσιο στο YOLO αντιπροσωπεύει πόσο σίγουρο είναι το μοντέλο ότι το οριοθετημένο πλαίσιο περιέχει ένα αντικείμενο και πόσο ακριβές είναι το πλαίσιο όσον αφορά τη θέση και το μέγεθός του.

- Η βαθμολογία εμπιστοσύνης υπολογίζεται ως το γινόμενο της πιθανότητας να υπάρχει αντικείμενο στο πλαίσιο (βαθμολογία αντικειμενικότητας) και της τομής επί της ένωσης (IoU) μεταξύ του προβλεπόμενου πλαισίου και του πλαισίου της βασικής αλήθειας ( $C = Pr(object) * IoU$ ).

- Κατά τη διάρκεια της εκπαίδευσης, η βαθμολογία εμπιστοσύνης συγκρίνεται με ένα όριο (συνήθως 0,5) για να καθοριστεί αν το προβλεπόμενο πλαίσιο περιέχει αντικείμενο ή όχι.

- Κατά τη διάρκεια της δοκιμής, η βαθμολογία εμπιστοσύνης χρησιμοποιείται για να φιλτράρει τις αδύναμες ανιχνεύσεις και να κρατήσει μόνο τα πλαίσια οριοθέτησης με υψηλή βαθμολογία εμπιστοσύνης.

Τέλος, μια βασική τεχνική που χρησιμοποιείται στα μοντέλα YOLO είναι η μη μέγιστη καταστολή (Non Maximum Suppression) [6]. Το NMS είναι ένα βήμα μετα-επεξεργασίας που χρησιμοποιείται για τη βελτίωση της ακρίβειας και της αποτελεσματικότητας της ανίχνευσης αντικειμένων. Στην ανίχνευση αντικειμένων, είναι σύνηθες να δημιουργούνται πολλαπλά πλαίσια οριοθέτησης για ένα μεμονωμένο αντικείμενο σε μια εικόνα. Αυτά τα πλαίσια οριοθέτησης μπορεί να επικαλύπτονται ή να βρίσκονται σε διαφορετικές θέσεις, αλλά όλα αντιπροσωπεύουν το ίδιο αντικείμενο. Η τεχνική NMS χρησιμοποιείται για τον εντοπισμό και την αφαίρεση περιττών ή εσφαλμένων πλαισίων οριοθέτησης και για την έξοδο ενός μόνο πλαισίου οριοθέτησης για κάθε αντικείμενο στην εικόνα.

## 8.4 YOLOv7

Η έκδοση του YOLO, YOLOv7, παρέχει πολύ βελτιωμένη ακρίβεια ανίχνευσης αντικειμένων σε πραγματικό χρόνο χωρίς να αυξάνει τον απαιτούμενο χρόνο συμπερασμάτων [8]. Όπως δείχνουν τα σημεία αναφοράς, σε σύγκριση με άλλους γνωστούς ανιχνευτές αντικειμένων, ο YOLOv7 μπορεί να μειώσει αποτελεσματικά περίπου κατά 50% τον χρόνο υπολογισμού και να επιτύχει μεγαλύτερη ταχύτητα συμπερασμάτων και μεγαλύτερη ακρίβεια ανίχνευσης.

Υλοποιεί μια ταχύτερη και ισχυρότερη αρχιτεκτονική δικτύου που παρέχει μια πιο αποτελεσματική μέθοδο ολοκλήρωσης χαρακτηριστικών, πιο ακριβή απόδοση ανίχνευσης αντικειμένων και αυξημένη απόδοση ετικετών και εκπαίδευσης μοντέλων. Ως αποτέλεσμα, ο YOLOv7 απαιτεί πολλές φορές φθηνότερο υπολογιστικό υλικό από άλλα μοντέλα βαθιάς εκμάθησης. Μπορεί να εκπαιδευτεί πολύ πιο γρήγορα σε μικρά σύνολα δεδομένων χωρίς προ-εκπαιδευμένα βάρη.

Μία από τις κύριες βελτιώσεις, σε σχέση με τις προηγούμενες εκδόσεις, είναι η χρήση anchor boxes. Τα anchor boxes είναι ένα σύνολο προκαθορισμένων πλαισίων με διαφορετικούς λόγους διαστάσεων που χρησιμοποιούνται για την ανίχνευση αντικειμένων διαφορετικών σχημάτων.

Ο YOLOv7 χρησιμοποιεί εννέα anchor boxes, τα οποία του επιτρέπουν να ανιχνεύει ένα ευρύτερο φάσμα σχημάτων και μεγεθών αντικειμένων σε σύγκριση με προηγούμενες εκδόσεις, συμβάλλοντας έτσι στη μείωση του αριθμού των ψευδώς θετικών ανιχνεύσεων.

Μια σημαντική βελτίωση στο YOLOv7 είναι η χρήση μιας νέας συνάρτησης απώλειας που ονομάζεται "εστιακή απώλεια". Οι προηγούμενες εκδόσεις του χρησιμοποιούσαν μια τυπική λειτουργία απώλειας διασταυρούμενης εντροπίας, η οποία είναι γνωστό ότι είναι λιγότερο αποτελεσματική στην ανίχνευση μικρών αντικειμένων.

Η εστιακή απώλεια καταπολεμά αυτό το ζήτημα μειώνοντας την απώλεια για καλά ταξινομημένα παραδείγματα και εστιάζοντας στα αντικείμενα που είναι δύσκολο να ανιχνευθούν.

Ο YOLOv7 έχει επίσης υψηλότερη ανάλυση από τις προηγούμενες εκδόσεις. Επεξεργάζεται εικόνες σε ανάλυση 608 με 608 εικονοστοιχεία, η οποία είναι υψηλότερη από την ανάλυση 416 με 416 που χρησιμοποιείται στο YOLOv3. Αυτή η υψηλότερη ανάλυση επιτρέπει στο να ανιχνεύει μικρότερα αντικείμενα και να έχει υψηλότερη ακρίβεια συνολικά.

## **8.5 Απόδοση ανίχνευσης αντικειμένων YOLOv7**

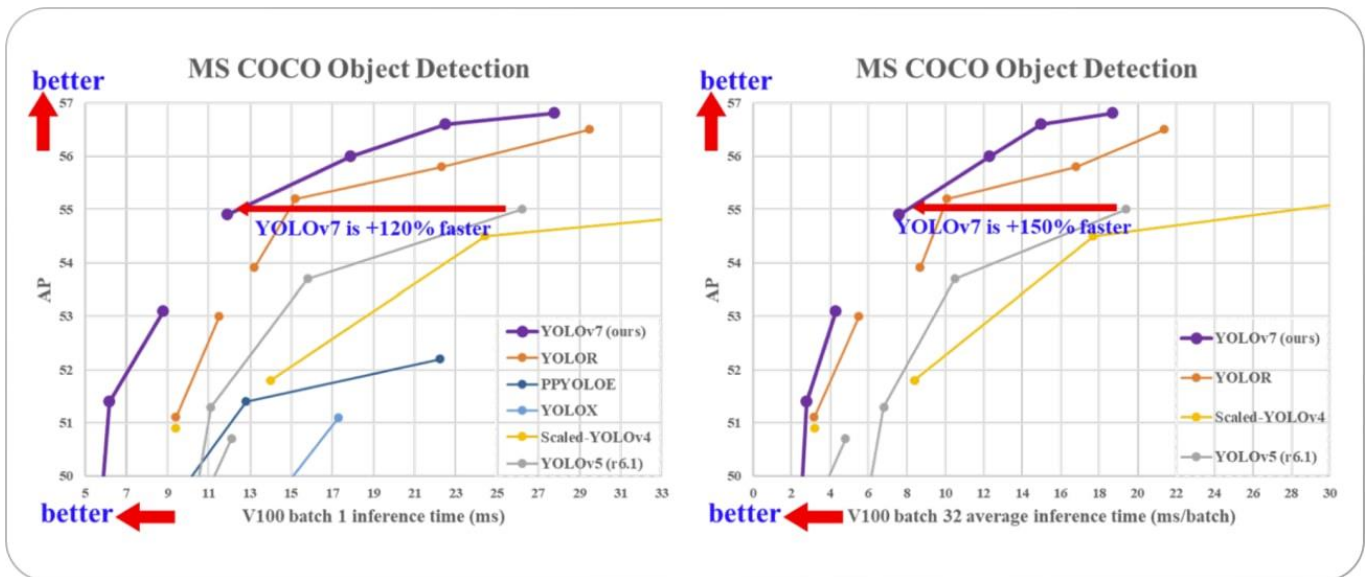
Η απόδοση αξιολογήθηκε με βάση προηγούμενες εκδόσεις YOLO (YOLOv4 και YOLOv5) και YOLOR ως μέτρο σύγκρισης. Ο YOLOv7 παρουσιάζει την καλύτερη ισορροπία ταχύτητας-ακρίβειας σε σύγκριση με υπόλοιπους ανιχνευτές αντικειμένων [6].

Σε γενικές γραμμές, ο YOLOv7 ξεπερνά όλους τους προηγούμενους ανιχνευτές αντικειμένων τόσο από την άποψη ταχύτητας όσο και της ακρίβειας, επιτυγχάνοντας την υψηλότερη ακρίβεια μεταξύ όλων των άλλων μοντέλων ανίχνευσης αντικειμένων σε πραγματικό χρόνο.

Μπορεί να επεξεργαστεί εικόνες με ρυθμό έως 160 FPS, πολύ γρηγορότερα από άλλους αλγόριθμους ανίχνευσης αντικειμένων. Αυτό τον καθιστά κατάλληλο για ευαίσθητες εφαρμογές σε πραγματικό χρόνο, όπως επιτήρηση και αυτόνομη οδήγηση, όπου οι υψηλότερες ταχύτητες επεξεργασίας είναι κρίσιμες.

Όσον αφορά την ακρίβεια, ο YOLOv7 έχει καλή απόδοση σε σύγκριση με άλλους αλγόριθμους ανίχνευσης αντικειμένων. Επιτυγχάνει μέση ακρίβεια 37,2% σε όριο IoU (τομή πάνω από ένωση) 0,5 στο δημοφιλές σύνολο δεδομένων COCO, το οποίο είναι συγκρίσιμο με άλλους προηγμένους αλγόριθμους ανίχνευσης αντικειμένων.

Ωστόσο, θα πρέπει να σημειωθεί ότι ο YOLOv7 είναι λιγότερο ακριβής από τους ανιχνευτές δύο σταδίων, όπως το Faster R-CNN και το Mask R-CNN, που τείνουν να επιτυγχάνουν υψηλότερη μέση ακρίβεια στο σύνολο δεδομένων COCO, αλλά απαιτούν και μεγαλύτερους χρόνους συμπερασμάτων.



Εικόνα 19: Σύγκριση του YOLOv7 με Άλλους Ανιχνευτές στην Μέση Ακρίβεια (AP) και στους Χρόνους Συμπερασμάτων [27]

Ο YOLOv7 είναι ένας ισχυρός και αποτελεσματικός αλγόριθμος ανίχνευσης αντικειμένων, αλλά έχει μερικούς περιορισμούς:

Όπως πολλοί αλγόριθμοι ανίχνευσης αντικειμένων, δυσκολεύεται να ανιχνεύσει μικρά αντικείμενα. Μπορεί να αποτύχει να ανιχνεύσει με ακρίβεια αντικείμενα σε σκηνές με πολλά αντικείμενα ή αντικείμενα που βρίσκονται μακριά από την κάμερα.

Δυσκολεύεται επίσης στην ανίχνευση αντικειμένων σε διαφορετικές κλίμακες. Δηλαδή αντικειμένων που είναι είτε πολύ μεγάλα είτε πολύ μικρά σε σύγκριση με τα άλλα αντικείμενα στη σκηνή.

Είναι ευαίσθητος σε αλλαγές στον φωτισμό ή άλλες περιβαλλοντικές συνθήκες, οπότε δεν μπορεί να χρησιμοποιηθεί σε πραγματικές εφαρμογές όπου οι συνθήκες φωτισμού μπορεί να διαφέρουν.

Τέλος είναι υπολογιστικά απαιτητικός, γεγονός που μπορεί να καταστήσει δύσκολη την εκτέλεση σε πραγματικό χρόνο σε συσκευές περιορισμένων πόρων, όπως έξυπνα τηλέφωνα ή άλλες συσκευές.



## 8.6 YOLOv8

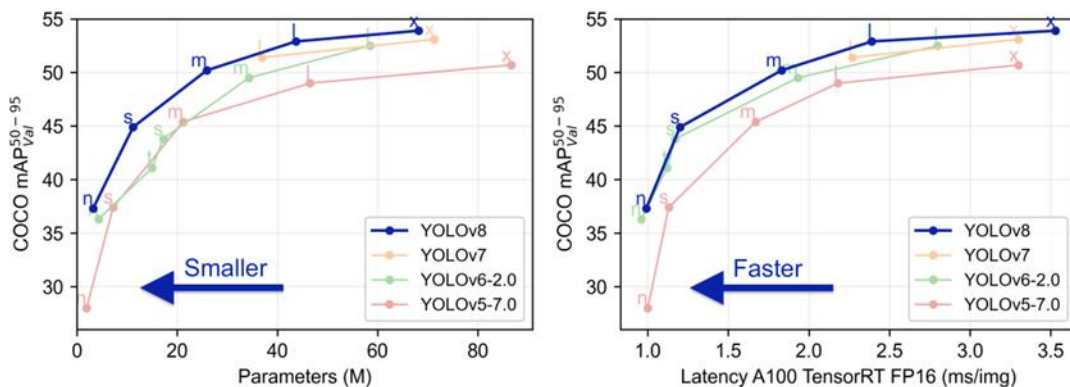
Ο YOLOv8 είναι η τελευταία έκδοση των αλγορίθμων ανίχνευσης αντικειμένων YOLO σε πραγματικό χρόνο, προσφέροντας κορυφαίες επιδόσεις όσον αφορά την ακρίβεια και την ταχύτητα. Συνδυάζοντας τις τελευταίες εξελίξεις από προηγούμενες εκδόσεις, ο YOLOv8 εισάγει νέα χαρακτηριστικά και βελτιστοποιήσεις που τον καθιστούν ιδανική επιλογή για διάφορες εργασίες ανίχνευσης αντικειμένων σε ένα ευρύ φάσμα εφαρμογών.

Ένα από τα βασικά πλεονεκτήματα του YOLOv8 είναι η ευελιξία του [18]. Ως ένα μοντέλο που υποστηρίζει όλα τα προηγούμενα μοντέλα YOLO, είναι εύκολο για τους χρήστες να εναλλάσσονται μεταξύ διαφορετικών εκδόσεων και να αξιολογούν τις επιδόσεις τους, καθιστώντας το μια εξαιρετική επιλογή για όσους θέλουν να αξιοποιήσουν την τελευταία τεχνολογία YOLO και ταυτόχρονα να μπορούν να χρησιμοποιούν τα παλαιότερα μοντέλα τους.

Μαζί με την ευελιξία του, ο YOLOv8 διαθέτει αρκετές άλλες καινοτομίες που το καθιστούν ισχυρό εργαλείο για ένα ευρύ φάσμα εργασιών ανίχνευσης αντικειμένων και τμηματοποίησης εικόνας. Επιπλέον, ο YOLOv8 είναι εξαιρετικά αποδοτικός και μπορεί να εκτελεστεί σε μια ποικιλία υλικού, από CPU έως GPU.

Μειονέκτημα του αλγορίθμου είναι η δυσκολία στην ανίχνευση αντικειμένων σε ακατάστατες σκηνές. Επιπλέον, το μοντέλο μπορεί να δυσκολεύεται να ανιχνεύσει μικρά αντικείμενα ή αντικείμενα με χαμηλή αντίθεση.

Συμπερασματικά, ο YOLOv8 είναι ένα στιβαρό και προσαρμόσιμο εργαλείο για την ανίχνευση αντικειμένων που προσφέρει τεχνολογία αιχμής καθώς και τη δυνατότητα χρήσης και σύγκρισης όλων των προηγούμενων εκδόσεων του YOLO.



Εικόνα 18: Σύγκριση του YOLOv8 με Προηγούμενες Εκδόσεις YOLO στην Μέση Ακρίβεια και στην Ταχύτητα [6]

## 9. Ultralytics

Το ultralytics YOLOv8 [17] είναι ένα ευέλικτο framework σχεδιασμένο για να καλύπτει ολόκληρο τον κύκλο ζωής των μοντέλων μηχανικής μάθησης – από την πρόσληψη δεδομένων και την εκπαίδευση μοντέλων έως την επικύρωση των αποτελεσμάτων και την ανάπτυξη. Κάθε λειτουργία εξυπηρετεί έναν συγκεκριμένο σκοπό και έχει σχεδιαστεί για να προσφέρει την ευελιξία και την αποδοτικότητα που απαιτούνται για διαφορετικές εργασίες και περιπτώσεις χρήσης.

Η σειρά προσφέρει ένα ευρύ φάσμα μοντέλων, το καθένα ειδικευμένο για συγκεκριμένες εργασίες στην μηχανική όραση. Τα μοντέλα αυτά ικανοποιούν διάφορες απαιτήσεις, από την ανίχνευση αντικειμένων σε πιο σύνθετες εργασίες όπως η τμηματοποίηση περιπτώσεων, η ανίχνευση θέσεων/σημείων-κλειδιά, η προσανατολισμένη ανίχνευση αντικειμένων και η ταξινόμηση.

Κάθε παραλλαγή της σειράς YOLOv8 βελτιστοποιείται για την αντίστοιχη εργασία της, εξασφαλίζοντας υψηλή απόδοση και ακρίβεια. Επιπλέον, αυτά τα μοντέλα είναι συμβατά με διάφορους τρόπους λειτουργίας, συμπεριλαμβανομένων των συμπερασμάτων, της επικύρωσης, της εκπαίδευσης και της εξαγωγής, διευκολύνοντας τη χρήση τους σε διάφορα στάδια ανάπτυξης.

Ενδεικτικά οι λειτουργίες αυτές είναι:

### Λειτουργία εκπαίδευσης (Training)

Η λειτουργία εκπαίδευσης χρησιμοποιείται για την εκπαίδευση ενός μοντέλου YOLOv8 σε ένα προσαρμοσμένο σύνολο δεδομένων. Σε αυτή τη λειτουργία, το μοντέλο εκπαιδεύεται χρησιμοποιώντας ένα καθορισμένο σύνολο δεδομένων και παραμέτρους. Η διαδικασία εκπαίδευσης περιλαμβάνει τη βελτιστοποίηση των παραμέτρων του μοντέλου, ώστε να μπορεί να προβλέπει με ακρίβεια τις κατηγορίες και τις θέσεις των αντικειμένων σε μια εικόνα.

### Λειτουργία επικύρωσης (Validation)

Η λειτουργία επικύρωσης χρησιμοποιείται για την επικύρωση ενός μοντέλου YOLOv8 μετά την εκπαίδευσή του. Σε αυτή τη λειτουργία, το μοντέλο αξιολογείται σε ένα σύνολο επικύρωσης για τη μέτρηση της ακρίβειας και της απόδοσης γενίκευσής του. Αυτή η λειτουργία μπορεί να χρησιμοποιηθεί για τη ρύθμιση των παραμέτρων του μοντέλου ώστε να βελτιωθεί η απόδοσή του.

### Λειτουργία πρόβλεψης (Predict)

Η λειτουργία Πρόβλεψη χρησιμοποιείται για την πραγματοποίηση προβλέψεων με τη χρήση ενός εκπαιδευμένου μοντέλου YOLOv8 σε νέες εικόνες ή βίντεο. Σε αυτή τη λειτουργία, το μοντέλο φορτώνεται από ένα αρχείο ελέγχου και ο χρήστης μπορεί να παρέχει εικόνες ή βίντεο για την εκτέλεση συμπερασμάτων. Το μοντέλο προβλέπει τις κλάσεις και τις θέσεις των αντικειμένων στις εικόνες ή το βίντεο εισόδου.

### Λειτουργία εξαγωγής (Export)

Η λειτουργία εξαγωγής χρησιμοποιείται για την εξαγωγή ενός μοντέλου YOLOv8 σε μορφή που μπορεί να χρησιμοποιηθεί για ανάπτυξη. Σε αυτή τη λειτουργία, το μοντέλο μετατρέπεται σε μορφή που μπορεί να χρησιμοποιηθεί από άλλες εφαρμογές λογισμικού ή συσκευές υλικού. Αυτή η λειτουργία είναι χρήσιμη κατά την ανάπτυξη του μοντέλου σε περιβάλλοντα παραγωγής.

Το Ultralytics YOLOv8 είναι ένα framework τεχνητής νοημοσύνης που υποστηρίζει πολλαπλές εργασίες μηχανικής όρασης. Μπορεί να χρησιμοποιηθεί για την ανίχνευση αντικειμένων, τμηματοποίηση, ανίχνευση προσανατολισμένων αντικειμένων, ταξινόμηση και ανίχνευση θέσης. Κάθε μία από αυτές τις εργασίες έχει διαφορετικό στόχο και περίπτωση χρήσης:

### Ανίχνευση αντικειμένων

Η ανίχνευση είναι η κύρια εργασία που υποστηρίζεται από το YOLOv8. Περιλαμβάνει την ανίχνευση αντικειμένων σε ένα καρέ εικόνας ή βίντεο και τη σχεδίαση πλαισίων οριοθέτησης γύρω από αυτά. Τα αντικείμενα που ανιχνεύονται ταξινομούνται σε διάφορες κατηγορίες με βάση τα χαρακτηριστικά τους. Το YOLOv8 μπορεί να ανιχνεύσει πολλαπλά αντικείμενα σε ένα μόνο καρέ εικόνας ή βίντεο με μεγάλη ακρίβεια και ταχύτητα.

### Τμηματοποίηση

Η τμηματοποίηση είναι μια εργασία που περιλαμβάνει την κατάτμηση μιας εικόνας σε διαφορετικές περιοχές με βάση το περιεχόμενο της εικόνας. Σε κάθε περιοχή αποδίδεται μια ετικέτα με βάση το περιεχόμενό της. Η εργασία αυτή είναι χρήσιμη σε εφαρμογές όπως η τμηματοποίηση εικόνων και η ιατρική απεικόνιση. Το YOLOv8 χρησιμοποιεί μια παραλλαγή της αρχιτεκτονικής U-Net για την εκτέλεση της τμηματοποίησης.

### Ταξινόμηση

Η ταξινόμηση είναι μια εργασία που περιλαμβάνει την ταξινόμηση μιας εικόνας σε διάφορες κατηγορίες. Το YOLOv8 μπορεί να χρησιμοποιηθεί για την ταξινόμηση εικόνων με βάση το περιεχόμενό τους. Χρησιμοποιεί μια παραλλαγή της αρχιτεκτονικής EfficientNet για την ταξινόμηση.

### Ανίχνευση θέσης

Η ανίχνευση θέσης/σημείου κλειδιού είναι μια εργασία που περιλαμβάνει την ανίχνευση συγκεκριμένων σημείων σε ένα καρέ εικόνας ή βίντεο. Τα σημεία αυτά αναφέρονται ως σημεία κλειδιά και χρησιμοποιούνται για την παρακολούθηση της κίνησης ή την εκτίμηση της πόζας. Το YOLOv8 μπορεί να ανιχνεύσει σημεία-κλειδιά σε ένα καρέ εικόνας ή βίντεο με μεγάλη ακρίβεια και ταχύτητα.

### Ανίχνευση προσανατολισμένων αντικειμένων

Η ανίχνευση προσανατολισμένων αντικειμένων πηγαίνει ένα βήμα παραπέρα από την κανονική ανίχνευση αντικειμένων με την εισαγωγή μιας επιπλέον γωνίας για

τον ακριβέστερο εντοπισμό αντικειμένων σε μια εικόνα. Το YOLOv8 μπορεί να ανιχνεύσει περιστρεφόμενα αντικείμενα σε μια εικόνα ή ένα καρέ βίντεο με μεγάλη ακρίβεια και ταχύτητα.

## 10. Δορυφορικές Εικόνες

Οι δορυφορικές εικόνες αναφέρονται σε εικόνες της Γης που λαμβάνονται από δορυφόρους που βρίσκονται σε τροχιά γύρω από τον πλανήτη. Αυτοί οι δορυφόροι είναι εξοπλισμένοι με διάφορους αισθητήρες για την ανίχνευση ορατού και υπέρυθρου φωτός, καθώς και ακτινοβολίας μικροκυμάτων για τη δημιουργία εικόνων υψηλής ανάλυσης. Αυτές οι εικόνες ενώνονται για να δημιουργήσουν οπτικές αναπαραστάσεις της Γης που παρέχουν δεδομένα για το κλίμα, τη γεωγραφία και τις ανθρωπογενείς δομές.

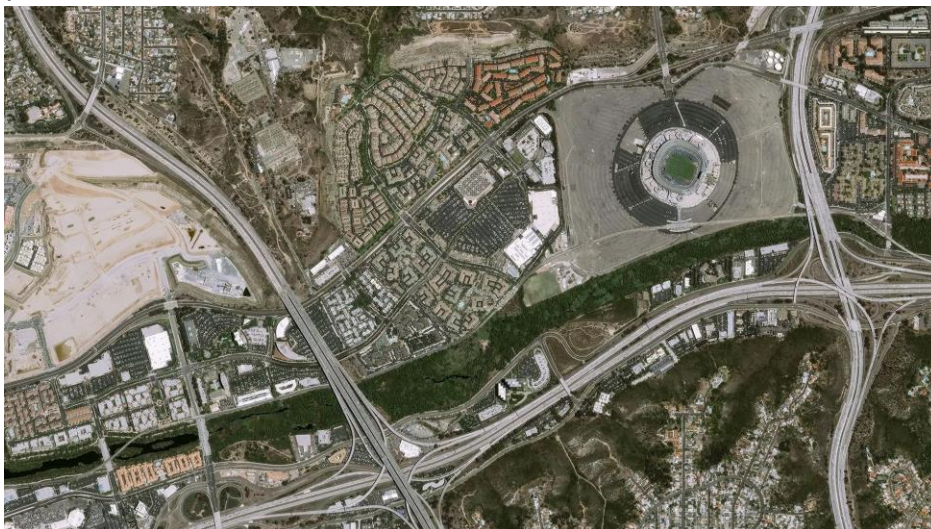
Οι δορυφορικές εικόνες διαδραματίζουν κρίσιμο ρόλο στη διαμόρφωση της κατανόησής μας για τον κόσμο, παρέχοντας πληθώρα πληροφοριών που δεν μπορούν να ληφθούν από κάμερες στο επίπεδο του εδάφους. Για παράδειγμα, οι δορυφορικές εικόνες μπορούν να παρακολουθήσουν την εξάπλωση της αποψίλωσης των δασών, τις αλλαγές στις θερμοκρασίες των ωκεανών ή να εντοπίσουν αλλαγές σε περιοχές της γης. Συνολικά, τα δεδομένα που συγκεντρώνονται από δορυφορικές εικόνες μπορούν να μας βοηθήσουν να λάβουμε αποφάσεις για τον πλανήτη μας.

### 10.1 Τύποι δορυφορικών εικόνων

Υπάρχουν μερικοί κύριοι τύποι δορυφορικών εικόνων, καθένας από τους οποίους συλλαμβάνει ένα διαφορετικό φάσμα πληροφοριών. Οι τρεις πιο συνηθισμένοι τύποι δορυφορικών εικόνων είναι [16]:

- Ορατές εικόνες

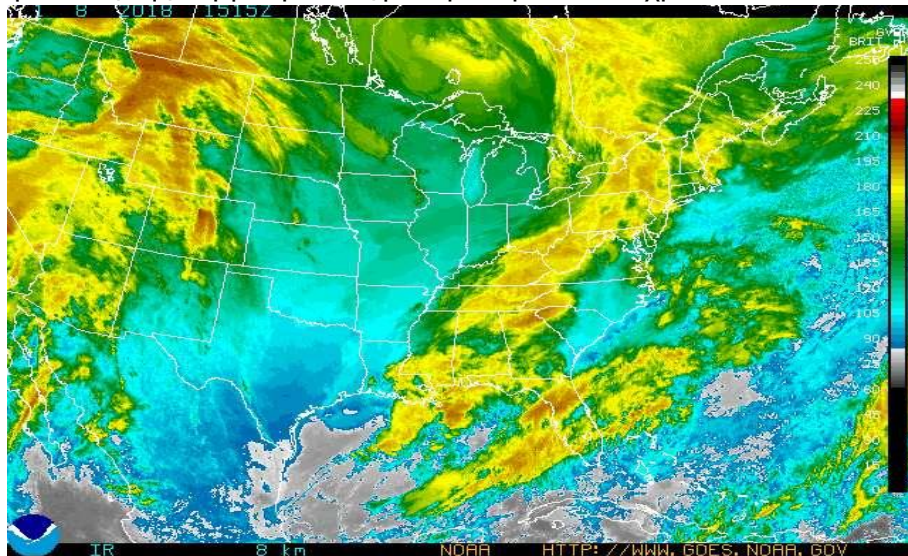
Αυτός ο τύπος εικόνων χρησιμοποιεί το φάσμα του ορατού φωτός (τα χρώματα του ουράνιου τόξου) για να συλλάβει μια αναπαράσταση της επιφάνειας της Γης. Αυτές οι εικόνες μοιάζουν πολύ με τις συνηθισμένες φωτογραφίες που μπορεί να τραβήξει κάποιος από μια ψηφιακή φωτογραφική μηχανή. Η βασική διαφορά είναι ότι οι δορυφορικές εικόνες λαμβάνονται από εξαιρετικά μεγάλο υψόμετρο από έναν δορυφόρο σε τροχιά.



Εικόνα 19: Ορατή Εικόνα

- Υπέρυθρες εικόνες

Οι υπέρυθρες εικόνες είναι ένας τύπος δορυφορικών εικόνων που καταγράφει εικόνες της επιφάνειας της Γης χρησιμοποιώντας υπέρυθρο φως. Οι υπέρυθρες εικόνες προσφέρουν μια διαφορετική άποψη της επιφάνειας της Γης από τις ορατές εικόνες. Ενώ οι ορατές εικόνες καταγράφουν εικόνες μόνο εντός του ορατού φάσματος, οι υπέρυθρες εικόνες μπορούν να ανιχνεύσουν ενέργεια στο υπέρυθρο φάσμα, ειδικά τη θερμοκρασία που εκπέμπεται από αντικείμενα. Αυτή η ικανότητα παρέχει μια ευδιάκριτη προοπτική της επιφάνειας της Γης για να παρακολουθείται η εξάπλωση των πυρκαγιών, η ποσότητα του νερού στο έδαφος των καλλιεργειών ή οι μεγάλες διακυμάνσεις της θερμοκρασίας με την πάροδο του χρόνου.

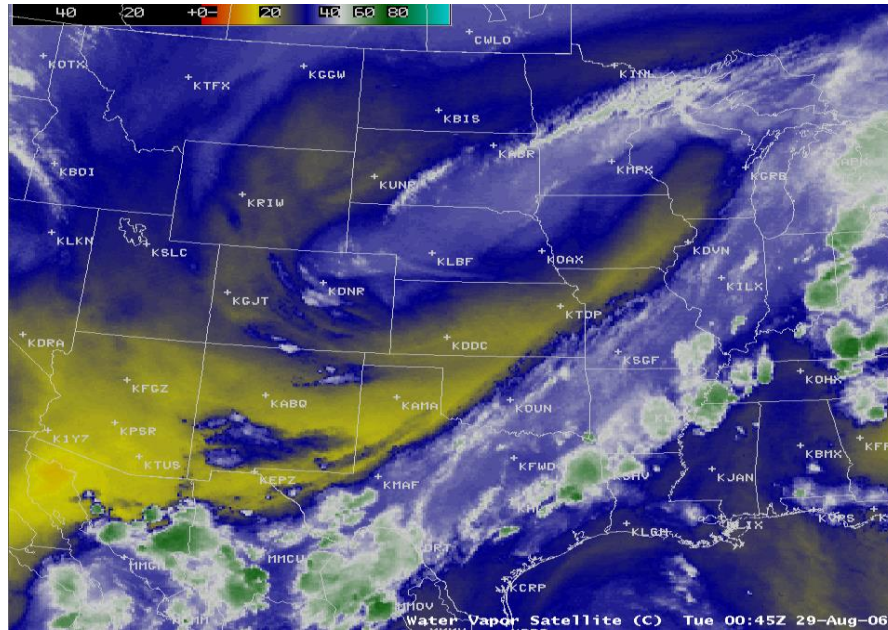


Εικόνα 20: Υπέρυθρη Εικόνα

- Εικόνες υδρατμών

Οι εικόνες υδρατμών είναι ένας τύπος δορυφορικών εικόνων που χρησιμοποιεί αισθητήρες για την ανίχνευση και τη μέτρηση της ποσότητας υδρατμών στην ατμόσφαιρα της Γης. Σε αντίθεση με τις ορατές εικόνες, που καταγράφουν εικόνες στο φάσμα του ορατού φωτός και τις υπέρυθρες εικόνες, οι οποίες ανιχνεύουν τη θερμοκρασία που εκπέμπεται από αντικείμενα, οι εικόνες υδρατμών ανιχνεύουν την ενέργεια μικροκυμάτων που εκπέμπεται από τα μόρια του νερού. Αυτό επιτρέπει μια πιο συγκεκριμένη κατανόηση της κατανομής και της κίνησης των υδρατμών στην ατμόσφαιρα, χρήσιμη για την πρόγνωση του καιρού και τις κλιματικές μελέτες.





Εικόνα 21: Εικόνα Υδρατμών

## 10.2 Χρήση δορυφορικών εικόνων

Οι δορυφορικές εικόνες χρησιμοποιούνται για διάφορους σκοπούς, συμπεριλαμβανομένης της δημιουργίας χαρτών και εικόνων, της ανίχνευσης αλλαγών και της πρόβλεψης καιρού [14].

- **Εικόνες και χάρτες**  
Δορυφορικές εικόνες μπορούν να χρησιμοποιηθούν για τη δημιουργία λεπτομερών χαρτών της επιφάνειας της Γης. Αυτοί οι χάρτες μπορούν να χρησιμοποιηθούν για διάφορους σκοπούς, συμπεριλαμβανομένης της ναυσιπλοΐας, του αστικού σχεδιασμού και της περιβαλλοντικής διαχείρισης.
- **Ανίχνευση αλλαγών**  
Οι δορυφορικές εικόνες μπορούν να χρησιμοποιηθούν για την παρακολούθηση αλλαγών στην επιφάνεια της Γης με την πάροδο του χρόνου. Για παράδειγμα, οι δορυφορικές εικόνες μπορούν να χρησιμοποιηθούν για την παρακολούθηση της αποψίλωσης των δασών, της ανάπτυξης των αστικών περιοχών ή της έκτασης των ζημιών μετά από φυσική καταστροφή.
- **Πρόγνωση καιρού**  
Οι δορυφορικές εικόνες μπορούν να χρησιμοποιηθούν για την παρακολούθηση των καιρικών προτύπων, συμπεριλαμβανομένης της κίνησης των σύννεφων, της θέσης των καταιγίδων, και των αλλαγών στη θερμοκρασία.
- **Εφαρμογές Άμυνας**  
Οι δορυφορικές εικόνες μπορούν να χρησιμοποιηθούν για στρατιωτικούς σκοπούς

### 10.3 Λήψη δορυφορικών εικόνων

Οι δορυφορικές εικόνες λαμβάνονται από εμπορικούς ή κυβερνητικούς δορυφόρους. Οι ερευνητές βασίζονται σε μεγάλο βαθμό σε βάσεις δεδομένων δορυφορικών εικόνων ανοιχτού κώδικα που ενημερώνονται συχνά. Μπορούμε να λάβουμε δορυφορικές εικόνες σε εύρος ανάλυσης και ιστορικές εικόνες από παρόχους δορυφορικών εικόνων ανοιχτού κώδικα, όπως το Google Earth Pro και το Bing Maps. Σχετικά μικρά αντικείμενα όπως αεροσκάφη, οχήματα και πλοία μπορούν να ανιχνευτούν από εικόνες υψηλότερης ανάλυσης (π.χ. περίπου 0,5 m/pixel) ενώ αντικείμενα μεγάλου μεγέθους, π.χ. αεροδρόμια, δρόμοι και μεγάλα κτίρια με μέση ανάλυση (π.χ. περίπου 1 m/pixel)



Εικόνα 22: Δορυφορική Εικόνα Από Το Google Earth

### 10.4 Ανίχνευση αντικειμένων σε δορυφορικές εικόνες

Η ανίχνευση αντικειμένων σε δορυφορικές εικόνες είναι η εργασία για την εύρεση συγκεκριμένης παρουσίας μιας ή περισσότερων κατηγοριών αντικειμένων, όπως ένα αεροπλάνο, και τον εντοπισμό της θέσης τους.

Στην ανάλυση δορυφορικών εικόνων, η ανίχνευση αντικειμένων έπαιξε πάντα ζωτικό ρόλο σε πολλές εφαρμογές, όπως συστήματα γεωγραφικών πληροφοριών, περιβαλλοντική παρακολούθηση, γεωργία ακριβείας και πολεοδομικό σχεδιασμό [12]. Για το σκοπό αυτό, χρειαζόμαστε ισχυρούς αλγόριθμους για αυτόματη ανίχνευση αντικειμένων σε δορυφορικές εικόνες. Η βαθιά μάθηση παρέχει μια αποτελεσματική λύση για την ανίχνευση αντικειμένων στις εικόνες αυτές.

Η ανίχνευση αντικειμένων επηρεάζεται από δύο είδη προκλήσεων: η πρώτη σχετίζεται με την ακρίβεια και η δεύτερη σχετίζεται με την αποτελεσματικότητα. Η πρώτη αφορά το ευρύ φάσμα των αλλαγών εντός μιας κατηγορίας ενώ η δεύτερη αφορά τον τεράστιο αριθμό κατηγοριών αντικειμένων και την ταξινόμηση στην σωστή κλάση αντικειμένου[12]. Η ακρίβεια των μοντέλων ανίχνευσης αντικειμένων βασίζεται σε μεγάλο βαθμό σε αυτούς τους δύο



παράγοντες. Για κάθε κατηγορία αντικειμένων, τα αντικείμενα εμφανίζονται με πολλές παραλλαγές, π.χ. μέγεθος, θέση, χρώμα, υφή και ιδιότητα ανάκλασης ενός υλικού. Η εμφάνιση τους επηρεάζεται επίσης από το περιβάλλον και τις συνθήκες φωτισμού. Η παραμόρφωση, ο φωτισμός, η θέση, η κλίμακα, το θάμπωμα, η ακαταστασία του φόντου και η σκίαση είναι όλοι παράγοντες που επηρεάζουν την εμφάνιση των αντικειμένων [6] [10], [12].

Ειδική αναφορά πρέπει να γίνει και στην υπολογιστική ισχύς όταν πρόκειται για την ανίχνευση αντικειμένων. Για τον βέλτιστο ανιχνευτή, οι συσκευές έχουν λιγότερη ισχύ επεξεργασίας και αποθηκευτικό χώρο. Όταν υπάρχει μεγάλος αριθμός αντικειμένων σε μια μεμονωμένη εικόνα με πάρα πολλές κατηγορίες, το αποτέλεσμα υποβαθμίζεται.

Τέλος, τα μοντέλα ανίχνευσης αντικειμένων απαιτούν μεγάλο όγκο επισημασμένων δεδομένων. Η επισήμανση αυτή των δεδομένων είναι μια από τις πιο χρονοβόρες και δαπανηρές εργασίες.

## 10.5 Προκλήσεις στις δορυφορικές εικόνες

Οι ταξινομητές εμφανίζουν συχνά φτωχά αποτελέσματα σε δορυφορικές εικόνες λόγω διαφορετικών συνθηκών. Οι χρονικές εποχιακές διαφορές, η ώρα της ημέρας καθώς και οι καιρικές μεταβολές επηρεάζουν τις δορυφορικές εικόνες.

Σε χωρική ανάλυση, τα αντικείμενα είναι πολύ μικρού μεγέθους και συγκεντρώνονται πυκνά σε δορυφορικές εικόνες, π.χ. τα αυτοκίνητα, είναι μόνο ~15 pixel σε εικόνες υψηλής ανάλυσης [1], [12], [32].

Επίσης, τα αντικείμενα έχουν πολλούς προσανατολισμούς (για παράδειγμα τα πλοία έχουν οποιονδήποτε προσανατολισμό που κυμαίνεται από 0 έως 360 μοίρες).

Πρέπει ακόμη να σημειωθεί ότι υπάρχει σχετική έλλειψη δεδομένων στις δορυφορικές εικόνες και τα αντικείμενα δεν είναι πάντα ορατά σε σχήμα. Οι εικόνες είναι πολύ υψηλής ανάλυσης (εκατοντάδες megapixel), ενώ οι περισσότεροι αλγόριθμοι λαμβάνουν εικόνες εισόδου συνήθως σε μερικές εκατοντάδες pixel.

Επομένως, είναι δύσκολο για έναν ταξινομητή να ανιχνεύσει αντικείμενα από συμβατικά σύνολα δεδομένων για λόγους που αναφέρθηκαν.

Για αυτό, χρειαζόμαστε ένα εξειδικευμένο είδος δεδομένων για δορυφορικές εικόνες ώστε η ανίχνευση των αντικειμένων σε αυτές να είναι υπολογιστικά λιγότερο δαπανηρή και χρονικά αποδοτική.

## **11. Προσανατολισμένα πλαίσια οριοθέτησης (Oriented Bounding Boxes)**

### **11.1 Ανίχνευση αντικειμένων με προσανατολισμένα πλαίσια οριοθέτησης (Oriented Bounding Boxes)**

Για να αντιμετωπιστούν οι προκλήσεις στην ανίχνευση αντικειμένων από δορυφορικές εικόνες επιλέχθηκε στο πλαίσιο της διπλωματικής η χρήση του μοντέλου της ανίχνευσης αντικειμένων σε προσανατολισμένα πλαίσια οριοθέτησης.

Τα Oriented Bounding Boxes (OBBs) είναι ένας τύπος πλαισίων οριοθέτησης που χρησιμοποιείται στην μηχανική όραση για να περιβάλλει ένα αντικείμενο με μεγαλύτερη ακρίβεια από ό,τι ένα οριοθετημένο πλαίσιο ευθυγραμμισμένο με τους άξονες συντεταγμένων.

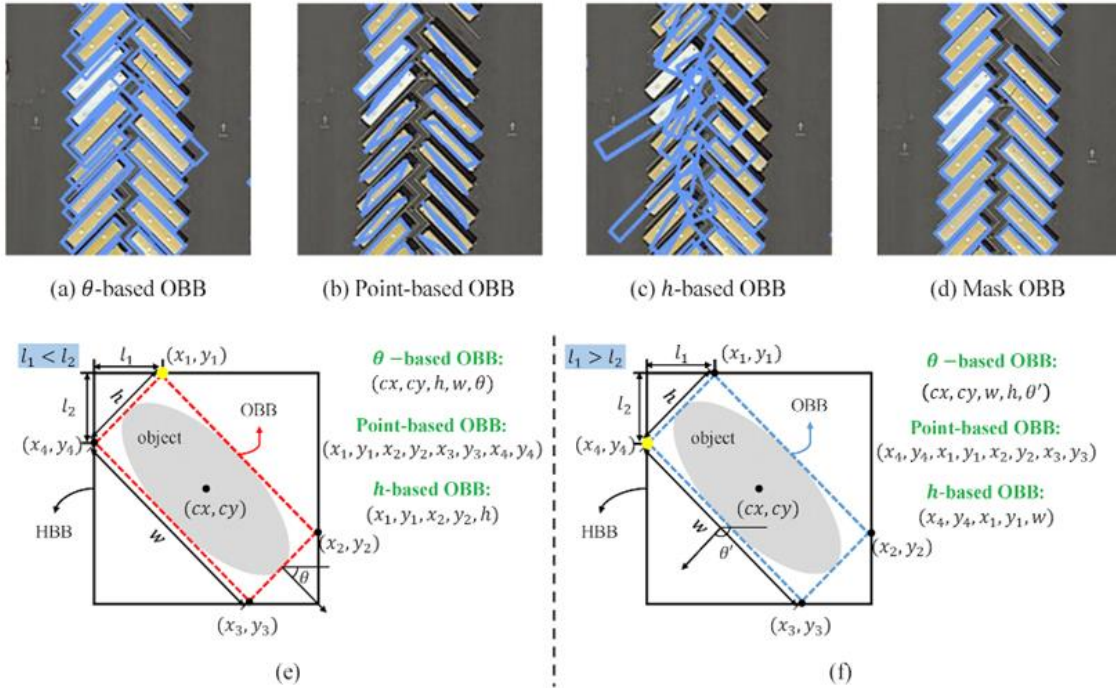
Το μοντέλο ανίχνευσης προσανατολισμένων αντικειμένων περιλαμβάνει μια πρόσθετη γωνία για τη βελτίωση της ακρίβειας εντοπισμού αντικειμένων στις εικόνες. Σε αντίθεση με τα κανονικά πλαίσια οριοθέτησης, τα οποία είναι ορθογώνια ευθυγραμμισμένα με τους άξονες  $x$  και  $y$ , τα OBB μπορούν να περιστρέφονται για να ταιριάζουν καλύτερα στον προσανατολισμό του αντικειμένου, επιτρέποντάς του να ευθυγραμμιστεί περισσότερο με το πραγματικό σχήμα και τον προσανατολισμό του αντικειμένου που περικλείει.

Αυτό είναι ιδιαίτερα χρήσιμο για εφαρμογές που απαιτούν ακριβή τοποθέτηση αντικειμένων, όπως εναέριες ή δορυφορικές εικόνες.

Η έξοδος ενός προσανατολισμένου ανιχνευτή αντικειμένων είναι ένα σύνολο προσανατολισμένων οριοθετημένων πλαισίων που περικλείουν ακριβώς τα αντικείμενα στην εικόνα, μαζί με ετικέτες κλάσης και βαθμολογίες εμπιστοσύνης για κάθε πλαίσιο. Η ανίχνευση αντικειμένων είναι μια καλή επιλογή όταν πρέπει να εντοπιστούν αντικείμενα ενδιαφέροντος σε μια σκηνή, αλλά δεν χρειάζεται να γνωρίζουμε ακριβώς πού βρίσκεται το αντικείμενο ή το ακριβές σχήμα του.

## 11.2 Σύνολα δεδομένων προσανατολισμένων πλαισίων οριοθέτησης (OBB)

Η εκπαίδευση ενός μοντέλου ανίχνευσης αντικειμένων με προσανατολισμένα πλαίσια οριοθέτησης (OBB) απαιτεί ένα λεπτομερές σύνολο δεδομένων. Η μορφή YOLO OBB ορίζει τα οριοθετημένα πλαίσια από τα τέσσερα γωνιακά σημεία τους με συντεταγμένες κανονικοποιημένες μεταξύ 0 και 1. Η μορφή είναι:



Εικόνα 23: Λειτουργία Μοντέλου YOLO-OBB

`class_index, x1, y1, x2, y2, x3, y3, x4, y4`

Εσωτερικά, ο αλγόριθμος YOLO επεξεργάζεται τις απώλειες και τις εξόδους σε μορφή `xywhr`, η οποία αναπαριστά το κεντρικό σημείο του οριοθετημένου πλαισίου ( $xy$ ), το πλάτος  $w$ , το ύψος  $h$  και την περιστροφή  $\theta$ .

Το μοντέλο μπορεί να απεικονιστεί, σύμφωνα με την εικόνα 23 με τρεις τρόπους:

- Point - based OBB ( $x_1, y_1, x_2, y_2, x_3, y_3, x_4, y_4$ ) – Απεικόνιση με βάση τις συντεταγμένες κορυφών OBB
- $\theta$  - based OBB ( $cx, cy, h, w, \theta$ ) – Απεικόνιση με βάση τη γωνία περιστροφής  $\theta$  του OBB
- $h$  - based OBB ( $x_1, y_1, x_2, y_2, h$ ) – Απεικόνιση με βάση το ύψος  $h$  του OBB

## 12. Σύνολο Δεδομένων DOTA (Dataset for Object deTection in Aerial images)

### 12.1 Υποστηριζόμενα σύνολα δεδομένων

Στο framework ultralytics, υποστηρίζονται τα ακόλουθα σύνολα δεδομένων με προσανατολισμένα πλαίσια οριοθέτησης (Oriented Bounding Boxes):

- DOTA-v1 και v2: Σύνολα δεδομένων μεγάλης κλίμακας για την ανίχνευση αντικειμένων σε αεροφωτογραφίες με έμφαση στην ανίχνευση από εναέρια οπτική γωνία, ενώ υποστηρίζουν προσανατολισμένα πλαίσια οριοθέτησης (Oriented Bounding Boxes).  
Η συλλογή τους έγινε από το Google Earth, το δορυφόρο GF-2 και άλλες αεροφωτογραφίες.  
Περιλαμβάνουν έως 18 κατηγορίες αντικείμενων, με 11.268 εικόνες με 1.793.658 στιγμιότυπα. Οι εικόνες του συνόλου δεδομένων χωρίζονται σε σύνολα εκπαίδευσης, επικύρωσης, δοκιμής. Για να αποφευχθεί το πρόβλημα της υπερπροσαρμογής (overfitting), το ποσοστό του συνόλου εκπαίδευσης και επικύρωσης είναι μικρότερο από το σύνολο δοκιμής.
- DOTA8: Ένα μικρό υποσύνολο 8 εικόνων του πλήρους συνόλου δεδομένων DOTA, κατάλληλο για δοκιμαστικές ροές εργασίας και ελέγχους συνεχούς ολοκλήρωσης (CI) της εκπαίδευσης OBB στο αποθετήριο ultralytics.

Το DOTA αποτελεί ένα εξειδικευμένο σύνολο δεδομένων μεγάλης κλίμακας, με έμφαση στην ανίχνευση αντικειμένων σε αεροφωτογραφίες. Προσφέρει επισημασμένες εικόνες που αποτυπώνουν ένα ευρύ φάσμα εναέριων εικόνων με προσανατολισμένα πλαίσια οριοθέτησης (Oriented Bounding Boxes) και μπορεί να χρησιμοποιηθεί για την ανάπτυξη και την αξιολόγηση ανιχνευτών αντικειμένων στις εναέρια εικόνες.

Οι εικόνες συλλέγονται από διαφορετικούς αισθητήρες και πλατφόρμες. Κάθε εικόνα έχει μέγεθος από 800 × 800 έως 20.000 × 20.000 και περιέχει αντικείμενα που παρουσιάζουν μεγάλη ποικιλία κλίμακας, προσανατολισμών και σχημάτων. Οι εικόνες στη συνέχεια «σχολιάζονται» από ειδικούς στην ερμηνεία αεροφωτογραφιών καταγράφοντας αντικείμενα διαφορετικής κλίμακας, προσανατολισμού και σχήματος

Ο σχολιασμός εικόνας είναι η διαδικασία επισημάνσης ή ταξινόμησης μιας εικόνας χρησιμοποιώντας κείμενο, εργαλεία σχολιασμού ή και τα δύο, για να αποτυπωθούν τα χαρακτηριστικά δεδομένων που θέλουμε να αναγνωρίσει το μοντέλο από μόνο του [28], Όταν επισημανθεί μια εικόνα, προστίθενται μεταδεδομένα σε ένα σύνολο δεδομένων.

Η επεξεργασία αυτή σηματοδοτεί τις λειτουργίες που θέλουμε να αναγνωρίζει το σύστημα μηχανικής εκμάθησης ενώ μπορούμε να χρησιμοποιήσουμε τις εικόνες για να εκπαιδύσουμε το μοντέλο μας χρησιμοποιώντας εποπτευόμενη εκμάθηση. Μόλις αναπτυχθεί το μοντέλο, είναι σε θέση να αναγνωρίζει εκείνες τις δυνατότητες στις εικόνες που δεν έχουν σχολιαστεί και, ως εκ τούτου, να λάβει μια απόφαση ή να προβεί σε κάποια ενέργεια. Ο σχολιασμός εικόνας

χρησιμοποιείται συνήθως για την αναγνώριση αντικειμένων και ορίων και για την τμηματοποίηση εικόνων. Για καθεμία από αυτές τις χρήσεις, απαιτείται σημαντικός όγκος δεδομένων για την εκπαίδευση, την επικύρωση και τη δοκιμή ενός μοντέλου μηχανικής μάθησης για να επιτευχθεί το επιθυμητό αποτέλεσμα.



Εικόνα 24: Ενάερεις Εικόνες

Το παραπάνω στιγμιότυπο υπογραμμίζει την πολυπλοκότητα των εναέριων σκηνών και τη σημασία των επισημασμένων προσανατολισμένων πλαισίων οριοθέτησης, που αποτυπώνουν τα αντικείμενα στο φυσικό τους προσανατολισμό.

## 12.2 Δομή συνόλου δεδομένων

Το DOTA παρουσιάζει μια δομημένη διάταξη προσαρμοσμένη στις προκλήσεις ανίχνευσης αντικειμένων OBB:

- Εικόνες: Μια τεράστια συλλογή αεροφωτογραφιών υψηλής ανάλυσης που αποτυπώνουν ποικίλα εδάφη και δομές.
- Προσανατολισμένα πλαίσια οριοθέτησης: Σημειώσεις με τη μορφή περιστρεφόμενων ορθογώνιων που περικλείουν αντικείμενα ανεξάρτητα



από τον προσανατολισμό τους, ιδανικά για τη σύλληψη αντικειμένων όπως αεροπλάνα, πλοία και κτίρια.

### 12.3 Εφαρμογές

Το DOTA σύνολο δεδομένων χρησιμεύει ως σημείο αναφοράς για την εκπαίδευση και την αξιολόγηση μοντέλων ειδικά προσαρμοσμένων για την ανάλυση αεροφωτογραφιών. Με τη συμπερίληψη των επισημάνσεων σε προσανατολισμένα πλαίσια οριοθέτησης (OBB), παρέχεται η δυνατότητα της ανάπτυξης εξειδικευμένων μοντέλων ανίχνευσης αντικειμένων που ανταποκρίνονται στις ιδιαιτερότητες των εναέριων εικόνων.

### 12.4 Dataset YAML

Τα σύνολα δεδομένων ενσωματώνουν ένα αρχείο YAML (Yet Another Markup Language) που περιγράφει λεπτομερώς τη διαμόρφωση του συνόλου δεδομένων. Για το DOTA v1, το Ultralytics παρέχει το αρχείο DOTAv1.yaml.



```
# Ultralytics YOLO 🚀, AGPL-3.0 license
# DOTA 1.0 dataset https://captain-whu.github.io/DOTA/index.html for object detection in aerial images by Wuhan
# Documentation: https://docs.ultralytics.com/datasets/obb/dota-v2/
# Example usage: yolo train model=yolov8n-obb.pt data=DOTAv1.yaml
# parent
# ├── ultralytics
# └── datasets
#     └── dota1 ← downloads here (2GB)

# Train/val/test sets as 1) dir: path/to/imgs, 2) file: path/to/imgs.txt, or 3) list: [path/to/imgs1, path/to/imgs2, ...]
path: ../datasets/DOTAv1 # dataset root dir
train: images/train # train images (relative to 'path') 1411 images
val: images/val # val images (relative to 'path') 458 images
test: images/test # test images (optional) 937 images

# Classes for DOTA 1.0
names:
  0: plane
  1: ship
  2: storage tank
  3: baseball diamond
  4: tennis court
  5: basketball court
  6: ground track field
  7: harbor
```

Εικόνα 25: Περιεχόμενα Αρχείου YAML

Το σύνολο δεδομένων αποτελείται από εικόνες χωρισμένες σε τρεις κατηγορίες υπό μορφή καταλόγων:

- Train

Εικόνες του συνόλου δεδομένων που θα χρησιμοποιηθούν για την εκπαίδευση

- Val

Εικόνες του συνόλου δεδομένων που θα χρησιμοποιηθούν για την αξιολόγηση και την επικύρωση των αποτελεσμάτων

- Test

Εικόνες του συνόλου δεδομένων που θα χρησιμοποιηθούν για την ανίχνευση των αντικειμένων

Οι εικόνες κατηγοριοποιούνται στις ακόλουθες κλάσεις:

- Plane (Αεροπλάνο)
- Ship (Πλοίο)
- Storage tank (Δεξαμενές αποθήκευσης)
- Baseball diamond (Γήπεδο Baseball)
- Tennis court (Γήπεδο Tennis)
- Basketball court (Γήπεδο Basketball)
- Ground track field (Στάδιο)
- Harbor (Λιμάνι)
- Bridge (Γέφυρα)
- Large vehicle (Μικρό Όχημα)
- Small vehicle (Μεγάλο Όχημα)
- Helicopter (Ελικόπτερο)
- Roundabout (Κυκλικός Κόμβος)
- Soccer ball field (Γήπεδο Ποδοσφαίρου)
- Swimming pool (Πισίνα)

## 13. Εκπαίδευση (Training)

Για την υλοποίηση της εργασίας επιλέχθηκε το σύνολο δεδομένων DOTAv1 ενώ από τις λειτουργίες του μοντέλου ultralytics YOLOv8 είναι απαραίτητες οι λειτουργίες Εκπαίδευση (Train), Επικύρωση (Validation), Πρόβλεψη (Predict). Η γλώσσα προγραμματισμού που χρησιμοποιήθηκε για την ανάπτυξη του κώδικα είναι η Python.

Τα βήματα που ακολουθήθηκαν για την εκπαίδευση του μοντέλου είναι:

### 13.1 Μετατροπή της μορφής του συνόλου δεδομένων DOTA σε Μορφή YOLO OBB

Η μετάβαση των ετικετών από τη μορφή συνόλου δεδομένων DOTA στη μορφή YOLO OBB μπορεί να επιτευχθεί ως εξής:

```
from ultralytics.data.converter import convert_dota_to_yolo_obb
convert_dota_to_yolo_obb("path/to/DOTA")
```

Εικόνα 26: Κώδικας μετατροπής δεδομένων από DOTA σε OBB

Αυτός ο μηχανισμός μετατροπής είναι σημαντικός για σύνολα δεδομένων σε μορφή DOTA, διασφαλίζοντας την ευθυγράμμιση με τη μορφή Ultralytics YOLO OBB.

### 13.2 Διαίρεση των εικόνων

Για να εκπαιδευτεί το σύνολο δεδομένων DOTA, διαχωρίζονται οι αρχικές εικόνες του συνόλου με υψηλή ανάλυση σε εικόνες με ανάλυση 1024x1024 πολλαπλής κλίμακας.

```
from ultralytics.data.split_dota import split_trainval, split_test

# split train and val set, with labels.
split_trainval(
    data_root='DOTAv1',
    save_dir='DOTAv1-split',
    rates=[0.5, 1.0, 1.5], # multiscale
    gap=500
)

# split test set, without labels.
split_test(
    data_root='DOTAv1',
    save_dir='DOTAv1-split',
    rates=[0.5, 1.0, 1.5], # multiscale
    gap=500
)
```

Εικόνα 27: Διαχωρισμός εικόνων σε εικόνες ανάλυσης 1024x1024



### 13.3 Κυρίως Εκπαίδευση (Training)

Το μοντέλο ultralytics YOLOv8 προσφέρει μεγέθη βαρών σύμφωνα με τον πίνακα:

Model	Weights
YOLOv8n-obb	Nano model
YOLOv8s-obb	Small model
YOLOv8m-obb	Medium model
YOLOv8l-obb	Large model
YOLOv8x-obb	Extra Large model

Πίνακας 1: Μεγέθη βαρών μοντέλου

Τα βάρη είναι αριθμητικές τιμές που σχετίζονται με τις συνδέσεις μεταξύ των νευρώνων στο νευρωνικό δίκτυο. Καθορίζουν την ισχύ αυτών των συνδέσεων και, με τη σειρά τους, την επιρροή που έχει η έξοδος ενός νευρώνα στην είσοδο ενός άλλου νευρώνα. Ουσιαστικά τα βάρη είναι συντελεστές που ρυθμίζουν την επίδραση των εισερχόμενων δεδομένων.

Πραγματοποιήθηκε εκτεταμένη εκπαίδευση κάνοντας χρήση όλων των μεγεθών βαρών του μοντέλου.

```
from ultralytics import YOLO
from ultralytics.data.converter import convert_dota_to_yolo_obb

convert_dota_to_yolo_obb('DOTAv1-split2')
# Load a model
# model = YOLO('yolov8n-obb.yaml') # build a new model from YAML
# model = YOLO('yolov8n-obb.pt') # load a pretrained model (recommended for training)
model = YOLO('yolov8x-obb.yaml').load('yolov8x.pt') # build from YAML and transfer weights

# Train the model
results = model.train(data='DOTAv1.yaml', epochs=300, imgsz=736, batch=8, name="NEW_split")
```

Εικόνα 28: Εκπαίδευση μοντέλου

Οι ρυθμίσεις εκπαίδευσης για τα μοντέλα YOLO περιλαμβάνουν διάφορες παραμέτρους και διαμορφώσεις που χρησιμοποιούνται κατά τη διαδικασία εκπαίδευσης. Αυτές οι ρυθμίσεις επηρεάζουν την απόδοση, την ταχύτητα και την ακρίβεια του μοντέλου.

Οι ρυθμίσεις αυτές κυρίως περιλαμβάνουν το μέγεθος των βαρών του μοντέλου, τις εποχές (κάθε εποχή αντιπροσωπεύει ένα πλήρες πέρασμα σε ολόκληρο το σύνολο δεδομένων - η τιμή μπορεί να επηρεάσει τη διάρκεια της εκπαίδευσης και την απόδοση του μοντέλου), το μέγεθος batch (δηλ. πόσες εικόνες μπορεί να επεξεργάζεται το μοντέλο ταυτόχρονα, πριν ενημερωθούν οι εσωτερικές παράμετροι του μοντέλου), το μέγεθος των εικόνων που θα εισαχθούν για εκπαίδευση (όλες οι εικόνες αλλάζουν μέγεθος σε αυτή τη διάσταση πριν εισαχθούν στο μοντέλο - επηρεάζει την ακρίβεια του μοντέλου και την υπολογιστική πολυπλοκότητα).

Η εκπαίδευση του μοντέλου έγινε σε 300 εποχές, στο μέγεθος batch δίνεται η τιμή 8, ενώ στο μέγεθος των εικόνων δίνεται η τιμή 736. Το μοντέλο θα αλλάξει το

μέγεθος των εικόνων εισόδου ώστε η μεγαλύτερη διάστασή τους να είναι 736, διατηρώντας την αρχική αναλογία διαστάσεων.

Η χρήση μεγαλύτερων μεγεθών εικόνας και μεγέθους batch, ενώ όπως διαπιστώθηκε επηρεάζουν θετικά την ακρίβεια και την απόδοση του μοντέλου, αυξάνουν σημαντικά τις απαιτήσεις μνήμης και το χρόνο εκπαίδευσης. Για αυτό το λόγο συνιστάται ιδιαίτερα η χρήση μιας ισχυρής GPU με αρκετή μνήμη.

Σχετικά με τον αριθμό των εποχών ενώ η προεπιλογή είναι 100, επιλέχθηκε η τιμή 300 μια και διαπιστώθηκε κατά τη διάρκεια της εκπαίδευσης βελτίωση της απόδοσης μέχρι τις 300 εποχές. Μετά τις 300 εποχές δεν παρουσιάστηκε καμία βελτίωση για αυτό συνιστάται να μην υπερβαίνουν οι εποχές αυτό το όριο γιατί είναι πιθανό να παρουσιαστεί το πρόβλημα της υπερπροσαρμογής (overfitting). Η υπερπροσαρμογή είναι μια ανεπιθύμητη συμπεριφορά μηχανικής μάθησης που εμφανίζεται όταν το μοντέλο δίνει ακριβείς προβλέψεις για δεδομένα εκπαίδευσης αλλά όχι για νέα δεδομένα.

Ο υπολογιστής που είχα στη διάθεσή μου για την εκπαίδευση δεν είχε δυνατότητα για μέγεθος εικόνας μεγαλύτερο του 736 και μέγεθος batch μεγαλύτερο του 8 λόγω περιορισμών στο hardware.

## 14. Αποτελέσματα

Οι μετρήσεις επιδόσεων είναι βασικά εργαλεία για την αξιολόγηση της ακρίβειας και της αποτελεσματικότητας των μοντέλων ανίχνευσης αντικειμένων. Υποδεικνύουν το πόσο αποτελεσματικά ένα μοντέλο μπορεί να αναγνωρίσει και να εντοπίσει αντικείμενα μέσα σε εικόνες. Επιπλέον, βοηθούν στην κατανόηση του χειρισμού των ψευδώς θετικών και ψευδώς αρνητικών αποτελεσμάτων από το μοντέλο. Αυτές οι γνώσεις είναι ζωτικής σημασίας για την αξιολόγηση και τη βελτίωση της απόδοσης του μοντέλου.

### 14.1 Μετρήσεις ανίχνευσης αντικειμένων

Μετρήσεις που δεν είναι σημαντικές μόνο για το YOLOv8 αλλά είναι ευρέως εφαρμόσιμες σε διάφορα μοντέλα ανίχνευσης αντικειμένων είναι οι εξής:

**Τομή σε Ένωση (IoU):** Η IoU είναι ένα μέτρο που ποσοτικοποιεί την επικάλυψη μεταξύ ενός προβλεπόμενου πλαισίου οριοθέτησης και ενός πλαισίου οριοθέτησης βασικής αλήθειας. Παίζει θεμελιώδη ρόλο στην αξιολόγηση της ακρίβειας του εντοπισμού αντικειμένων.

**Μέση ακρίβεια (AP):** Η AP υπολογίζει την περιοχή κάτω από την καμπύλη ακρίβειας-ανάκλησης, παρέχοντας μια ενιαία τιμή που περιλαμβάνει την απόδοση ακρίβειας και ανάκλησης του μοντέλου.

**Μέσος όρος Μέσης Ακρίβειας (mAP):** Η mAP επεκτείνει την έννοια της AP υπολογίζοντας τις μέσες τιμές AP σε πολλαπλές κλάσεις αντικειμένων. Αυτό είναι χρήσιμο σε σενάρια ανίχνευσης αντικειμένων πολλαπλών κλάσεων για την παροχή μιας ολοκληρωμένης αξιολόγησης της απόδοσης του μοντέλου.

**Ακρίβεια (Precision) και ανάκληση (Recall):** Η ακρίβεια ποσοτικοποιεί το ποσοστό των αληθώς θετικών προβλέψεων μεταξύ όλων των θετικών προβλέψεων, αξιολογώντας την ικανότητα του μοντέλου να αποφεύγει τα ψευδώς θετικά αποτελέσματα. Από την άλλη πλευρά, η ανάκληση υπολογίζει το ποσοστό των αληθώς θετικών μεταξύ όλων των θετικών προβλέψεων, μετρώντας την ικανότητα του μοντέλου να ανιχνεύει όλες τις περιπτώσεις μιας κλάσης.

**Βαθμολογία F1:** Η βαθμολογία F1 είναι ο μέσος όρος της ακρίβειας και της ανάκλησης, παρέχοντας μια ισορροπημένη αξιολόγηση της απόδοσης ενός μοντέλου, λαμβάνοντας υπόψη τόσο τα ψευδώς θετικά όσο και τα ψευδώς αρνητικά αποτελέσματα.

Οι διαθέσιμες μετρήσεις στο μοντέλο Ultralytics YOLOv8 είναι οι ακόλουθες:

## 14.2 Μετρήσεις ανά κλάση

Ένα από τα τμήματα της εξόδου είναι η ανάλυση των μετρικών επιδόσεων ανά κλάση. Αυτές οι λεπτομερείς πληροφορίες είναι χρήσιμες για να καταλάβουμε πόσο καλά τα πάει το μοντέλο για κάθε συγκεκριμένη κλάση, ειδικά σε σύνολα δεδομένων με ποικίλο φάσμα κατηγοριών αντικειμένων. Για κάθε κλάση στο σύνολο δεδομένων παρέχονται τα εξής:

**Κλάση:** Αυτή δηλώνει το όνομα της κατηγορίας αντικειμένου, όπως «πρόσωπο», «αυτοκίνητο» ή «σκύλος».

**Εικόνες:** Αυτή η μέτρηση μας δηλώνει τον αριθμό των εικόνων στο σύνολο επικύρωσης που περιέχουν την κατηγορία αντικειμένου.

**Περιπτώσεις (Instances):** Αυτή παρέχει τον αριθμό των φορών που εμφανίζεται η κλάση σε όλες τις εικόνες του συνόλου επικύρωσης.

**Box (P, R, mAP50, mAP50-95):** Αυτή η μέτρηση παρέχει πληροφορίες σχετικά με την απόδοση του μοντέλου στον εντοπισμό αντικειμένων:

- **P (Ακρίβεια - Precision):** Η ακρίβεια των εντοπισμένων αντικειμένων, υποδεικνύοντας πόσες ανιχνεύσεις ήταν σωστές. Η ακρίβεια είναι ένα μέτρο που δείχνει από όλες τις προβλέψεις που γίνονται από το μοντέλο, ποιο ποσοστό των προβλέψεων είναι πραγματικά σωστές.

$$Precision = \frac{TP}{TP + FP} = \frac{TP}{all\ Detections} \quad \text{ή}$$

$$Precision = \frac{CorrectPredictions}{CorrectPredictions+IncorrectPrediction} \quad [9]$$

- **R (Recall - Ανάκληση):** Η ικανότητα του μοντέλου να εντοπίζει όλες τις περιπτώσεις αντικειμένων στις εικόνες. Η ανάκληση είναι ένα μέτρο που υποδεικνύει από όλες τις βασικές αλήθειες ποιο ποσοστό προβλέφθηκε σωστά.

$$Recall = \frac{TP}{TP + FN} = \frac{TP}{all\ Ground\ Truths} \quad \text{ή}$$

$$Recall = \frac{CorrectPredictions}{CorrectPredictions+MissingDetections} \quad [9]$$

- **mAP50:** Μέσος όρος μέσης ακρίβειας που υπολογίζεται σε ένα όριο τομής σε ένωση (IoU) 0,50 για πολλαπλές κλάσεις αντικειμένων. Είναι ένα μέτρο της ακρίβειας του μοντέλου λαμβάνοντας υπόψη μόνο τις «εύκολες» ανιχνεύσεις.
- **mAP75:** Μέσος όρος μέσης ακρίβειας που υπολογίζεται σε ένα όριο τομής σε ένωση (IoU) 0,75. Είναι ένα μέτρο της ακρίβειας του μοντέλου

λαμβάνοντας υπόψη μόνο τις ανιχνεύσεις που θεωρούνται σωστές, με όριο  $IoU > 0,70$ .

- mAP50-95: Ο μέσος όρος της μέσης ακρίβειας που υπολογίζεται σε διαφορετικά όρια  $IoU$ , που κυμαίνονται από 0,50 έως 0,95.

Παρέχει μια ολοκληρωμένη εικόνα της απόδοσης του μοντέλου σε διάφορα επίπεδα δυσκολίας ανίχνευσης.

### 14.3 Μετρήσεις ταχύτητας

Η ταχύτητα εξαγωγής συμπερασμάτων μπορεί να είναι εξίσου κρίσιμη με την ακρίβεια, ειδικά σε σενάρια ανίχνευσης αντικειμένων σε πραγματικό χρόνο. Αυτή η μέτρηση αναλύει το χρόνο που απαιτείται για διάφορα στάδια της διαδικασίας επικύρωσης, από την προεπεξεργασία έως τη μετα-επεξεργασία.

### 14.4 Οπτικές έξοδοι

Η συνάρτηση `model.val()`, εκτός από την παραγωγή αριθμητικών μετρήσεων, παράγει επίσης οπτικές εξόδους που μπορούν να παρέχουν μια πιο διαισθητική κατανόηση της απόδοσης του μοντέλου. Ακολουθεί μια ανάλυση των οπτικών αποτελεσμάτων:

Καμπύλη βαθμολογίας F1 (`F1_curve.png`): Αυτή η καμπύλη αντιπροσωπεύει τη βαθμολογία F1 σε διάφορα κατώτατα όρια. Η ερμηνεία αυτής της καμπύλης μπορεί να προσφέρει πληροφορίες σχετικά με την ισορροπία του μοντέλου μεταξύ ψευδώς θετικών και ψευδώς αρνητικών αποτελεσμάτων σε διάφορα κατώτατα όρια.

Καμπύλη ακρίβειας-ανάκλησης (`PR_curve.png`): Μια οπτικοποίηση για κάθε πρόβλημα ταξινόμησης, αυτή η καμπύλη παρουσιάζει τις ανταλλαγές μεταξύ ακρίβειας και ανάκλησης σε ποικίλα όρια.

Καμπύλη ακρίβειας (`P_curve.png`): Μια γραφική αναπαράσταση των τιμών ακρίβειας σε διαφορετικά κατώτατα όρια. Αυτή η καμπύλη βοηθά στην κατανόηση του τρόπου με τον οποίο μεταβάλλεται η ακρίβεια καθώς αλλάζει το όριο.

Καμπύλη ανάκλησης (`R_curve.png`): Αντίστοιχα, αυτή η γραφική παράσταση απεικονίζει πώς μεταβάλλονται οι τιμές ανάκλησης σε διαφορετικά όρια.

Πίνακας σύγχυσης (`confusion_matrix.png`): Ο πίνακας σύγχυσης παρέχει μια λεπτομερή εικόνα των αποτελεσμάτων, παρουσιάζοντας τις μετρήσεις των αληθώς θετικών, των αληθώς αρνητικών, των ψευδώς θετικών και των ψευδώς αρνητικών για κάθε κλάση ή κατηγορία.

		Actual	
		Positive	Negative
Predicted	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

Εικόνα 29: Πίνακας σύγχυσης

Κανονικοποιημένος πίνακας σύγχυσης (confusion\_matrix\_normalized.png): Αυτή η απεικόνιση είναι μια κανονικοποιημένη έκδοση του πίνακα σύγχυσης. Αναπαριστά τα δεδομένα σε αναλογίες και όχι σε ακατέργαστες μετρήσεις. Αυτή η μορφή καθιστά απλούστερη τη σύγκριση των επιδόσεων μεταξύ των κλάσεων. Ετικέτες παρτίδας επικύρωσης (val\_batchX\_labels.jpg): Αυτές οι εικόνες απεικονίζουν τις ετικέτες βασικής αλήθειας για ξεχωριστές παρτίδες από το σύνολο δεδομένων επικύρωσης. Παρέχουν μια σαφή εικόνα για το ποια είναι τα αντικείμενα και τις αντίστοιχες θέσεις τους σύμφωνα με το σύνολο δεδομένων. Προβλέψεις παρτίδων επικύρωσης (val\_batchX\_pred.jpg): Σε αντίθεση με τις εικόνες των ετικετών, αυτές οι εικόνες εμφανίζουν τις προβλέψεις που έκανε το μοντέλο YOLOv8 για τις αντίστοιχες παρτίδες. Συγκρίνοντάς τες με τις εικόνες ετικέτας, μπορείτε εύκολα να αξιολογήσετε πόσο καλά το μοντέλο ανιχνεύει και ταξινομεί τα αντικείμενα οπτικά.

## 14.5 Επικύρωση (Validation)

Η λειτουργία επικύρωσης του YOLOv8 μπορεί να χρησιμοποιηθεί για τον υπολογισμό των προαναφερθέντων μετρήσεων αξιολόγησης.

Η επικύρωση είναι ένα κρίσιμο βήμα στην μηχανική μάθηση, που επιτρέπει να αξιολογηθεί η ποιότητα των εκπαιδευμένων μοντέλων.

Η λειτουργία αυτή στο Ultralytics YOLOv8 παρέχει μια ισχυρή σουίτα εργαλείων και μετρήσεων για την αξιολόγηση της απόδοσης των μοντέλων ανίχνευσης αντικειμένων. Η χρήση της είναι απλή. Μόλις έχουμε ένα εκπαιδευμένο μοντέλο, μπορεί να κληθεί η συνάρτηση `model.val()`. Αυτή η συνάρτηση θα επεξεργαστεί στη συνέχεια το σύνολο δεδομένων επικύρωσης και θα επιστρέψει μια ποικιλία μετρήσεων απόδοσης.

```
OOB_Val.py > ...
1 from ultralytics import YOLO
2 from ultralytics.data.converter import convert_dota_to_yolo_obb
3
4 model = YOLO('runs/obb/NEW_split5/weights/best.pt') # load a custom model
5 # Validate the model
6 metrics = model.val(data='DOTAv1.yaml') # no arguments needed, dataset and settings remembered
7 metrics.box.map # map50-95(B)
8 metrics.box.map50 # map50(B)
9 metrics.box.map75 # map75(B)
10 metrics.box.maps # a list contains map50-95(B) of each category
11 print ('')
12 print (['model weights: Yolo8x'])
13 print ('map50-95')
14 print (metrics.box.map)
15 print ('')
16 print ('map50') (function) map50: Any
17 print (metrics.box.map50)
18 print ('')
19 print ('map75')
```

Εικόνα 30: Επικύρωση της ακρίβειας του εκπαιδευμένου μοντέλου YOLOv8x-obb

Class	Images	Instances	P	R	mAP50	mAP50-95
all	458	28853	0.781	0.547	0.608	0.461
Plane	458	2531	0.926	0.727	0.815	0.681
Ship	458	8960	0.882	0.569	0.694	0.502
Tennis court	458	760	0.959	0.908	0.943	0.907
Basketball court	458	132	0.74	0.496	0.543	0.48
Ground track field	458	144	0.538	0.556	0.554	0.429
Harbor	458	2090	0.829	0.775	0.815	0.514
Bridge	458	463	0.597	0.224	0.271	0.129
Large vehicle	458	4387	0.833	0.803	0.82	0.646
Small vehicle	458	5438	0.584	0.619	0.609	0.447
Soccer ball field	458	153	0.733	0.379	0.433	0.382
Swimming pool	458	440	0.741	0.55	0.587	0.355

Πίνακας 2: Αποτελέσματα επικύρωσης (mAP50, mAP70, mAP50-95) του μοντέλου YOLOv8x-obb

Στον παρακάτω πίνακα αναγράφεται η μέση ακρίβεια mAP50, mAP70, mAP50-95, σε μοντέλα με όλα τα μεγέθη βαρών μετά την εκπαίδευση που πραγματοποιήθηκε, αξιοποιώντας τη λειτουργία επικύρωσης:

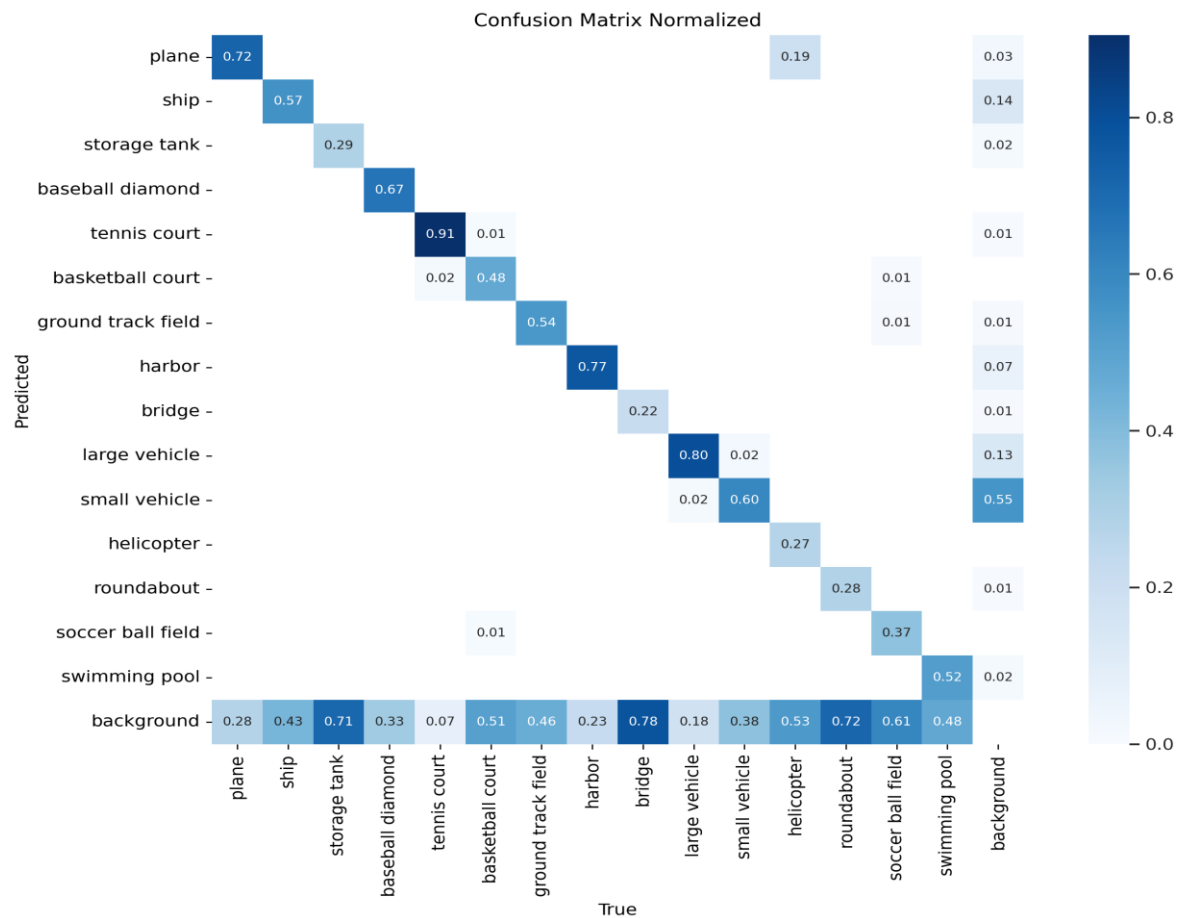
Model	mAP50	mAP75	mAP50-95	Inference Time (ms)
YOLOv8n-obb	0.47	0.37	0.33	3.1
YOLOv8s-obb	0.50	0.40	0.36	3.3
YOLOv8m-obb	0.58	0.49	0.43	4.7
YOLOv8l-obb	0.60	0.51	0.45	5.9
YOLOv8x-obb	0.61	0.52	0.46	8.6

Πίνακας 3: Μέση ακρίβεια (mAP50, mAP70, mAP50-95) και χρόνος συμπερασμάτων

Οι τιμές αυτές μας δείχνουν το ποσοστό της πρόβλεψης του κάθε μοντέλου που καλύπτει τη βασική αλήθεια (ground truth) στο σύνολο δεδομένων, ουσιαστικά το ποσοστό των θετικών προβλέψεων σε κάθε περίπτωση που υπερβαίνουν το  $IoU > 0.50$ , το  $IoU > 0.75$  και το μέσο όρο των 0,50 έως 0.95.

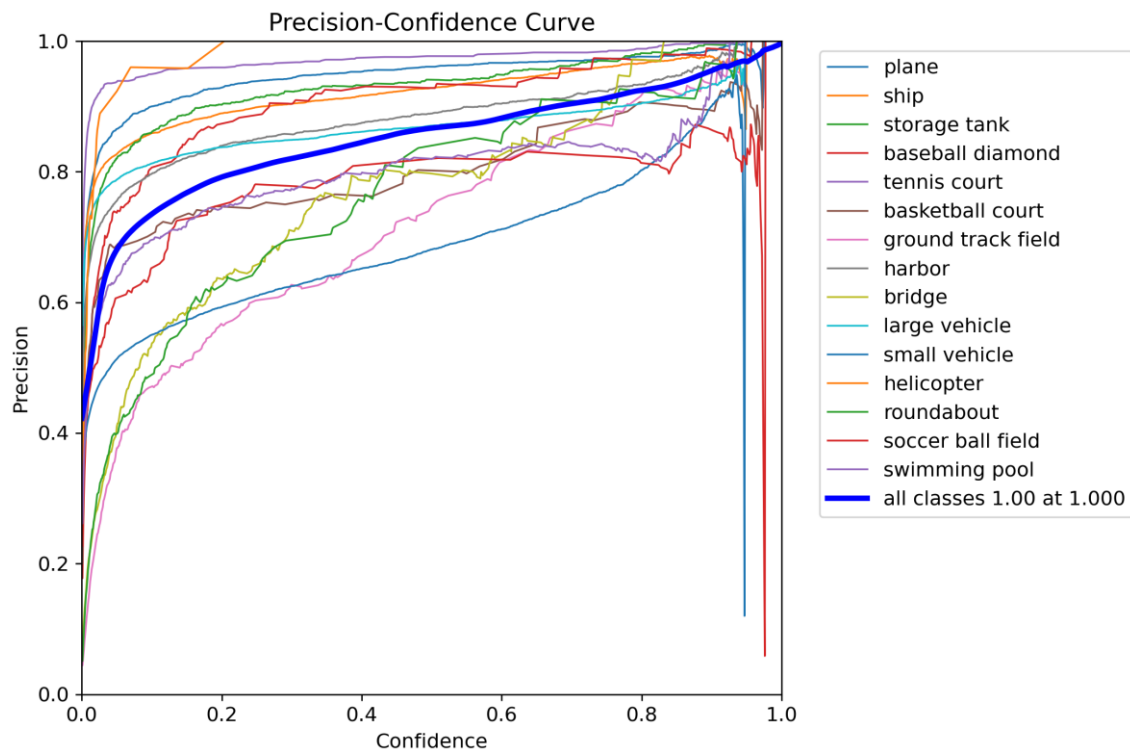
Πρέπει να σημειωθεί ότι όσο μεγαλύτερο είναι το  $IoU$  όριο (threshold) τόσο μικρότερο θα είναι και το mAP.

Στη συνέχεια παρατίθεται ο πίνακας σύγχυσης (confusion matrix) για το μοντέλο YOLOv8x-obb ο οποίος συσχετίζει τις προβλέψεις με τις ετικέτες της κλάσης:



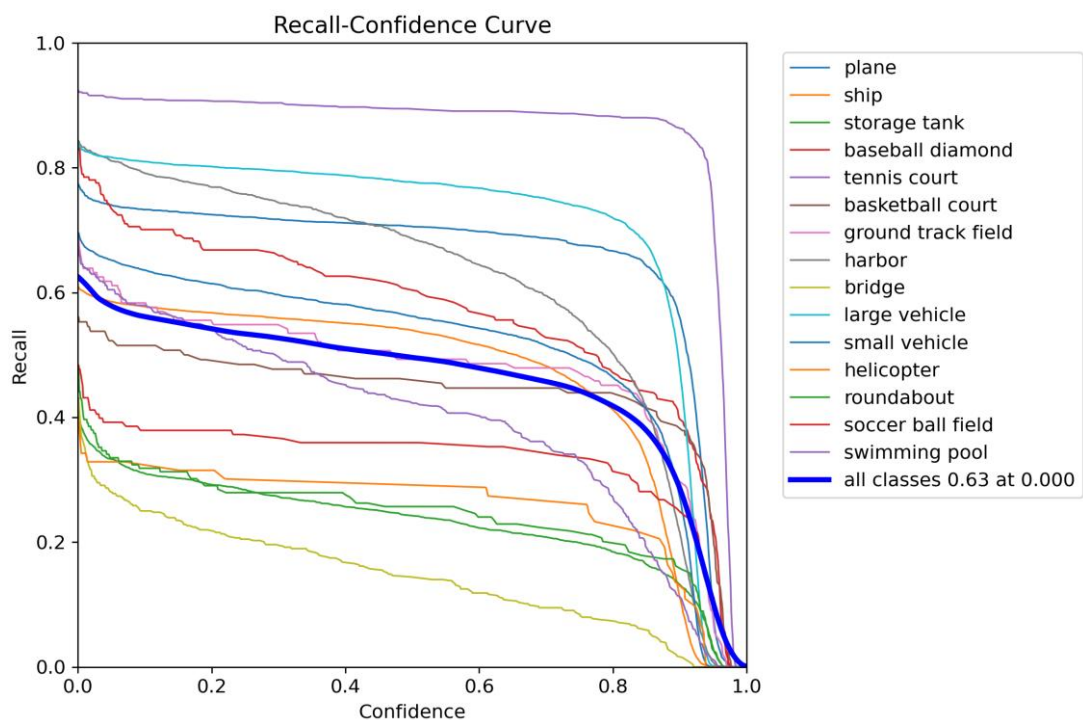
Εικόνα 31: Κανονικοποιημένος πίνακας σύγχυσης για το μοντέλο YOLOv8x-obb





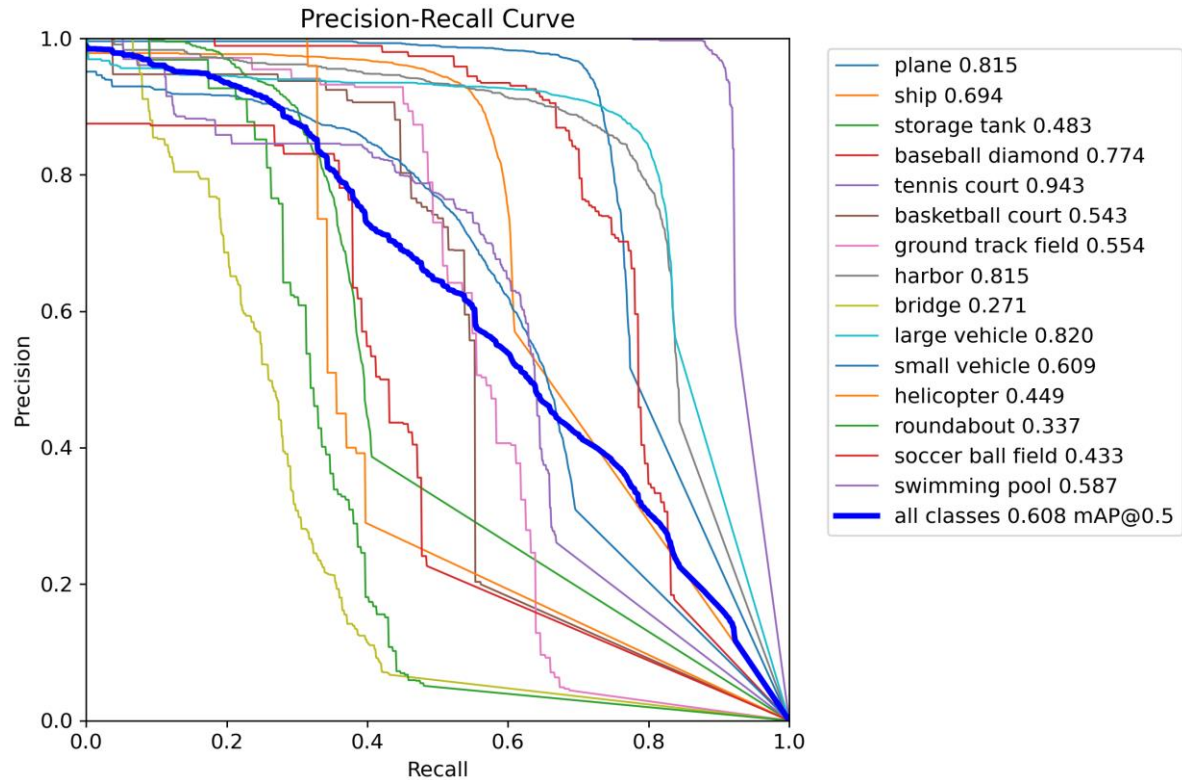
Εικόνα 32: Καμπύλη ακρίβειας (Μοντέλο YOLOv8x-obb)

Γραφική παράσταση των τιμών ακρίβειας του μοντέλου ως συνάρτηση του ορίου βαθμολογίας εμπιστοσύνης του μοντέλου. Η ακρίβεια σχετίζεται με τη σωστή πρόβλεψη και δείχνει σε ποιο βαθμό οι θετικές προβλέψεις του μοντέλου μπορούν να είναι αξιόπιστες



Εικόνα 33: Καμπύλη ανάκλησης (Μοντέλο YOLOv8x-obb)

Γραφική παράσταση των τιμών ανάκλησης του μοντέλου ως συνάρτηση του ορίου βαθμολογίας εμπιστοσύνης του μοντέλου. Η ανάκληση μετρά την ικανότητα του μοντέλου να εντοπίζει θετικά δείγματα και καταδεικνύει τυχόν προβλέψεις που δεν θα έπρεπε να έχει χάσει το μοντέλο.



Εικόνα 34: Καμπύλη ακρίβεια-ανάκλησης (Μοντέλο YOLOv8x-obb)

Η καμπύλη ανάκλησης-ακρίβειας απαιτεί τις ακόλουθες εισόδους για να δημιουργηθεί:

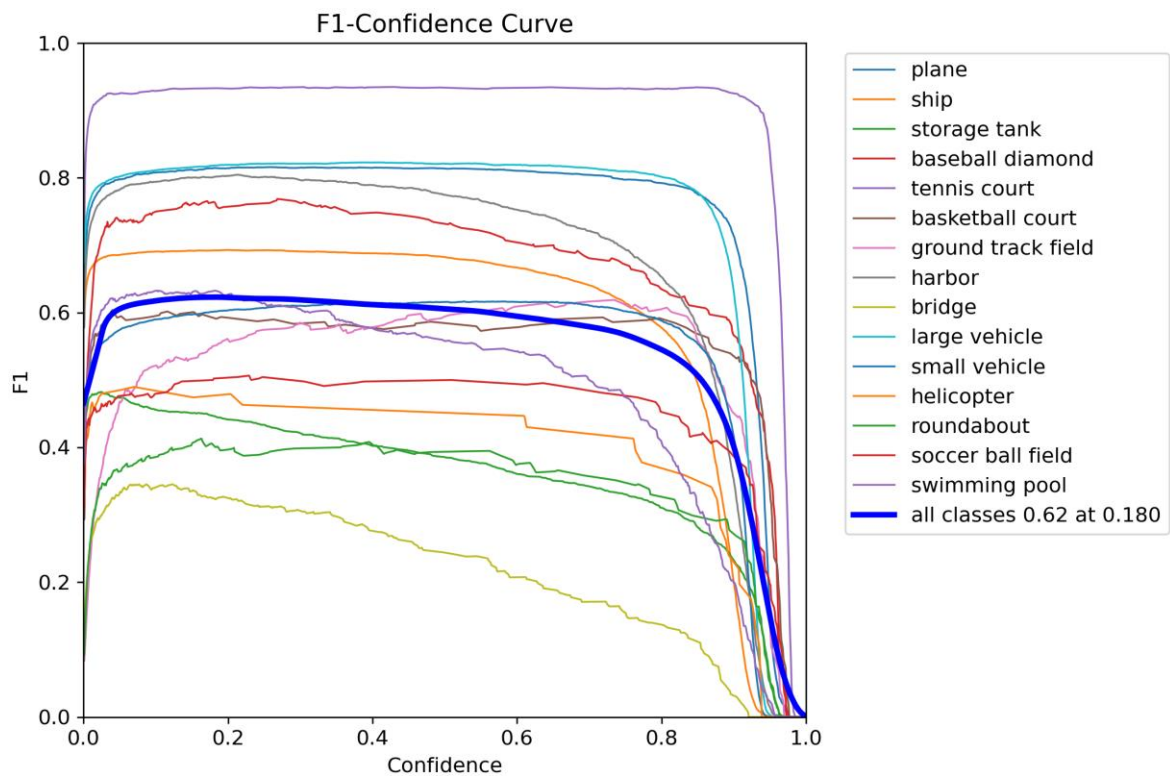
- Τις ετικέτες της βασικής αλήθειας.
- Τις βαθμολογίες πρόβλεψης των δειγμάτων.
- Μερικά όρια για τη μετατροπή των αποτελεσμάτων πρόβλεψης σε ετικέτες κλάσης.

Η ακρίβεια και η ανάκληση δεν μπορούν να χρησιμοποιηθούν ανεξάρτητα η μία από την άλλη, γι' αυτό και χρησιμοποιούμε μια καμπύλη που τις συσχετίζει.

Ο λόγος γι' αυτό είναι απλός. Όταν υπάρχει υψηλή ανάκληση αλλά χαμηλή ακρίβεια, το μοντέλο παράγει πολλά ψευδώς θετικά αποτελέσματα (τα αρνητικά δείγματα κατηγοριοποιούνται ως θετικά), παρόλο που ταξινομεί σωστά τα θετικά δείγματα. Αντίθετα, με υψηλή ακρίβεια αλλά χαμηλή ανάκληση, το μοντέλο είναι ακριβές μόνο όταν ταξινομεί τα δείγματα ως θετικά- αλλά μπορεί να το κάνει μόνο με μερικά από τα θετικά δείγματα.

Συνοψίζοντας, η καμπύλη ακρίβειας-ανάκλησης αποτυπώνει το συμβιβασμό μεταξύ των δύο κριτηρίων και μεγιστοποιεί τη συνδυασμένη επίδρασή τους.

Η εξέταση του συμβιβασμού μεταξύ των τιμών ακρίβειας και ανάκλησης σε διάφορα κατώτατα όρια είναι σημαντική λόγω του ρόλου και των δύο αυτών μετρικών. Η καμπύλη βοηθά στην επιλογή του ιδανικού ορίου για τη βελτιστοποίηση και των δύο μετρικών. Η υψηλή ανάκληση και η υψηλή ακρίβεια αντιπροσωπεύονται και οι δύο από μια υψηλή περιοχή κάτω από την καμπύλη, όπου η υψηλή ακρίβεια συσχετίζεται με χαμηλό ποσοστό ψευδώς θετικών αποτελεσμάτων και η υψηλή ανάκληση συσχετίζεται με χαμηλό ποσοστό ψευδώς αρνητικών αποτελεσμάτων. Οι υψηλές βαθμολογίες και για τις δύο δείχνουν ότι ο ταξινομητής παράγει αποτελέσματα με υψηλή ακρίβεια και υψηλό ποσοστό καλών ευρημάτων (υψηλή ανάκληση).



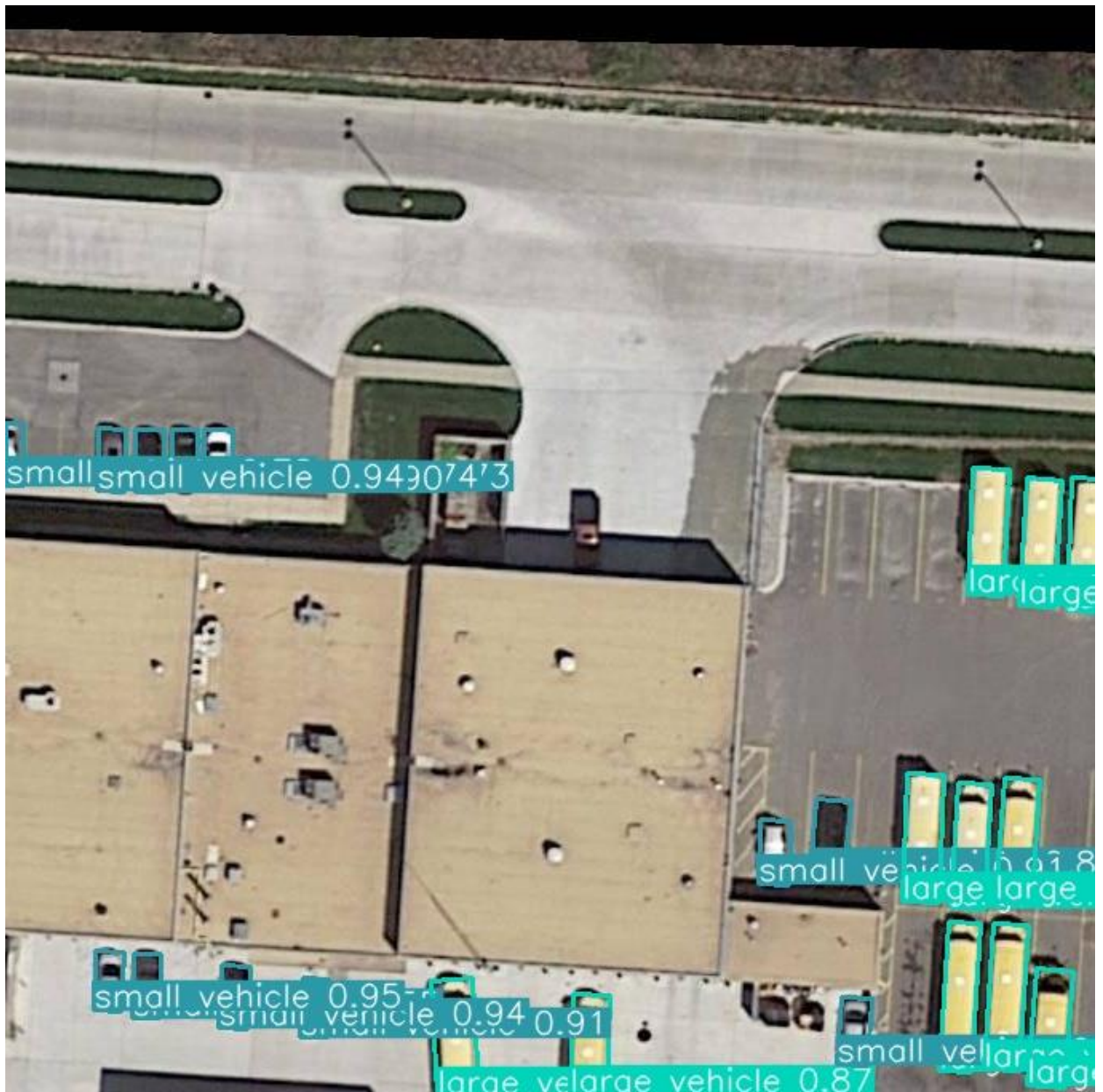
Εικόνα 35: Καμπύλη F1 (Μοντέλο YOLOv8x-obb)

Η καμπύλη F1 είναι ο μέσος όρος της ακρίβειας και της ανάκλησης, παρέχοντας μια ισορροπημένη αξιολόγηση της απόδοσης ενός μοντέλου, λαμβάνοντας υπόψη τόσο τα ψευδώς θετικά όσο και τα ψευδώς αρνητικά αποτελέσματα

## 14.6 Πρόβλεψη (Predict)

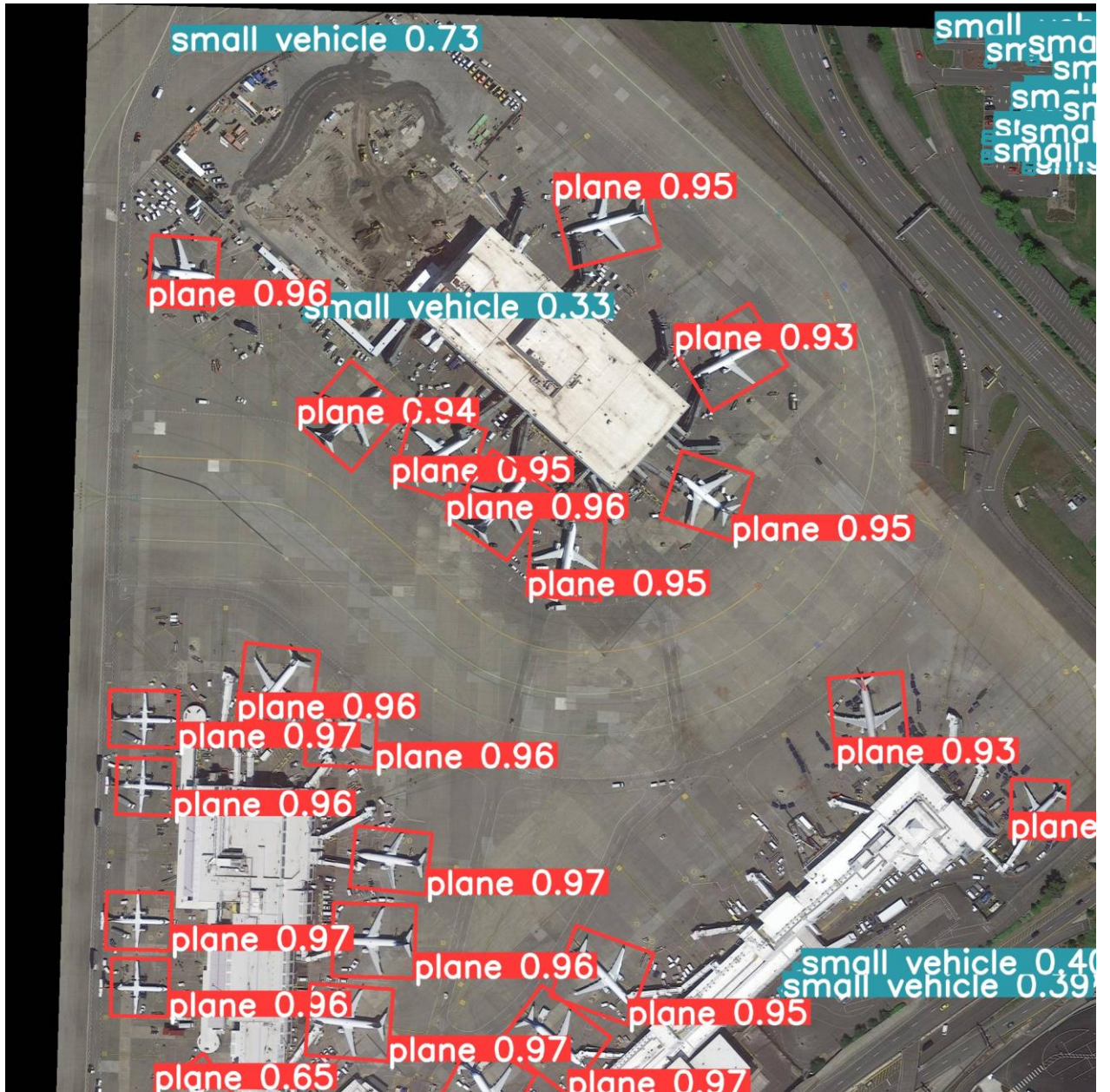
Η βαθμολογία εμπιστοσύνης (confidence score) για κάθε οριοθετημένο πλαίσιο στο YOLO αντιπροσωπεύει πόσο σίγουρο είναι το μοντέλο ότι το οριοθετημένο πλαίσιο περιέχει ένα αντικείμενο και πόσο ακριβές είναι το πλαίσιο όσον αφορά τη θέση και το μέγεθός του.

Η βαθμολογία εμπιστοσύνης υπολογίζεται ως το γινόμενο  $(Pr(\text{Object}) * IoU)$ , της πιθανότητας να υπάρχει αντικείμενο στο πλαίσιο (βαθμολογία αντικειμενικότητας) και της τομής επί της ένωσης (IoU) μεταξύ του προβλεπόμενου πλαισίου και του πλαισίου της βασικής αλήθειας.

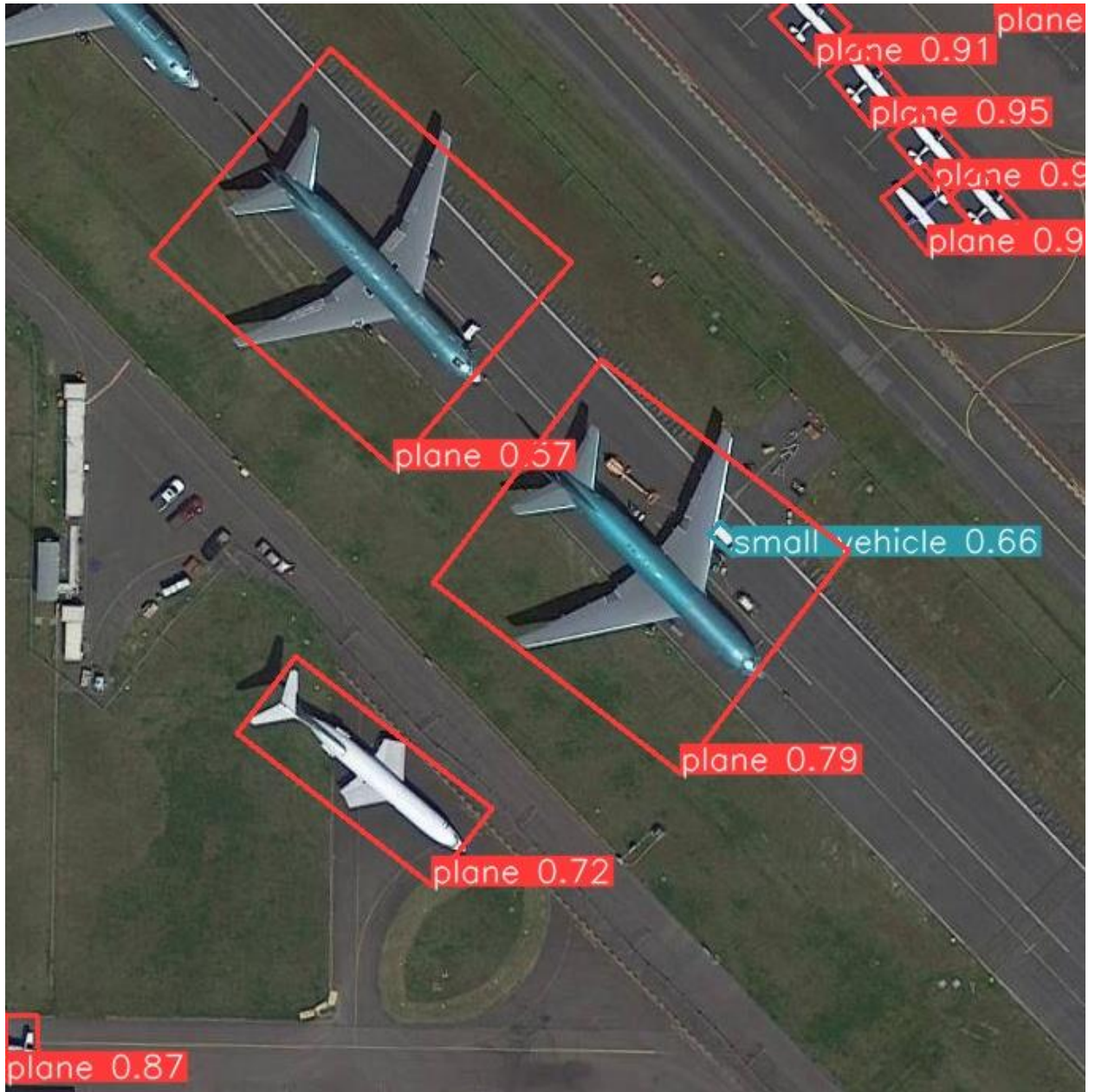


Εικόνα 36: Πρόβλεψη σε εικόνα





Εικόνα 37: Πρόβλεψη σε εικόνα

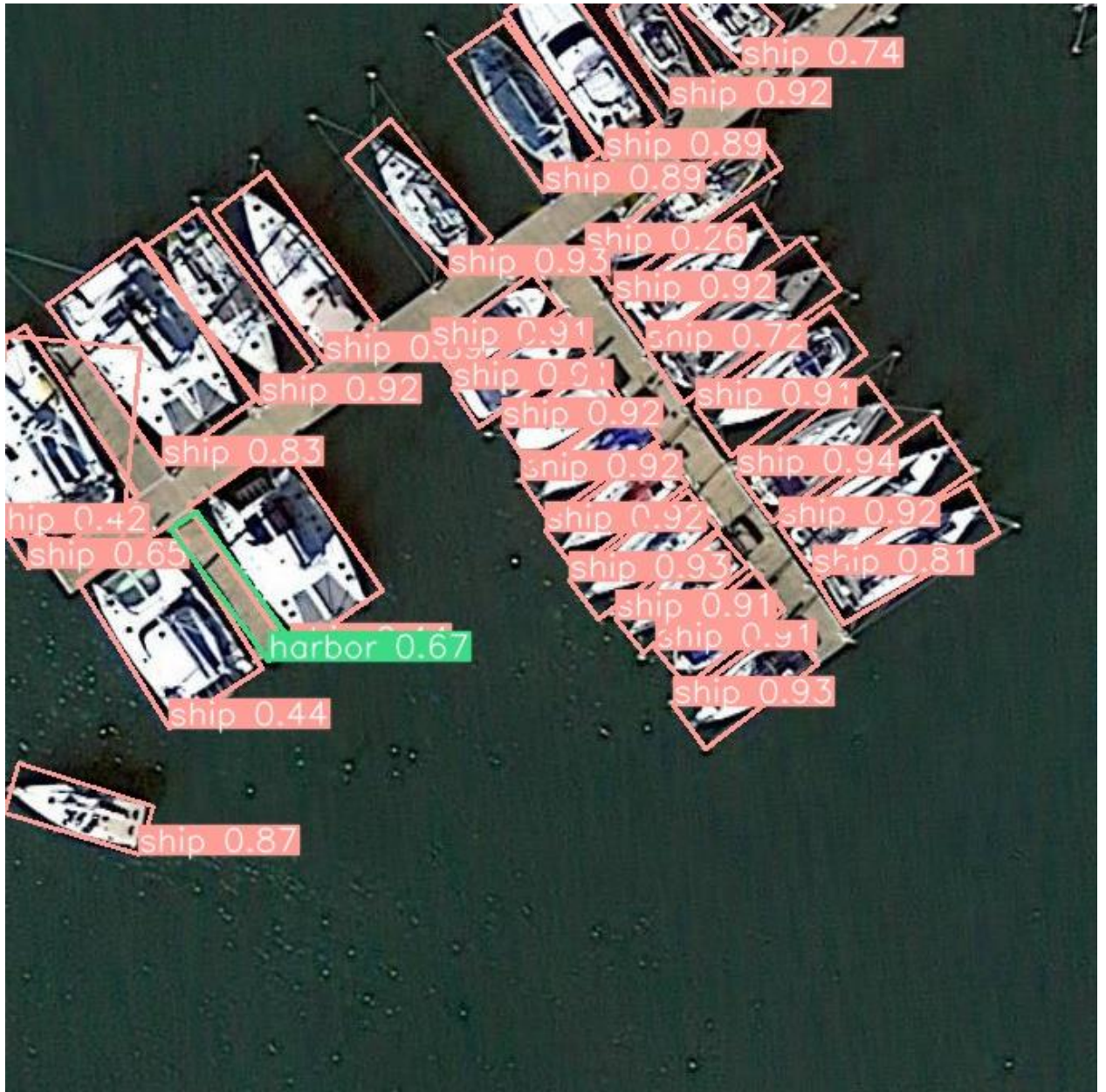


Εικόνα 38: Πρόβλεψη σε εικόνα



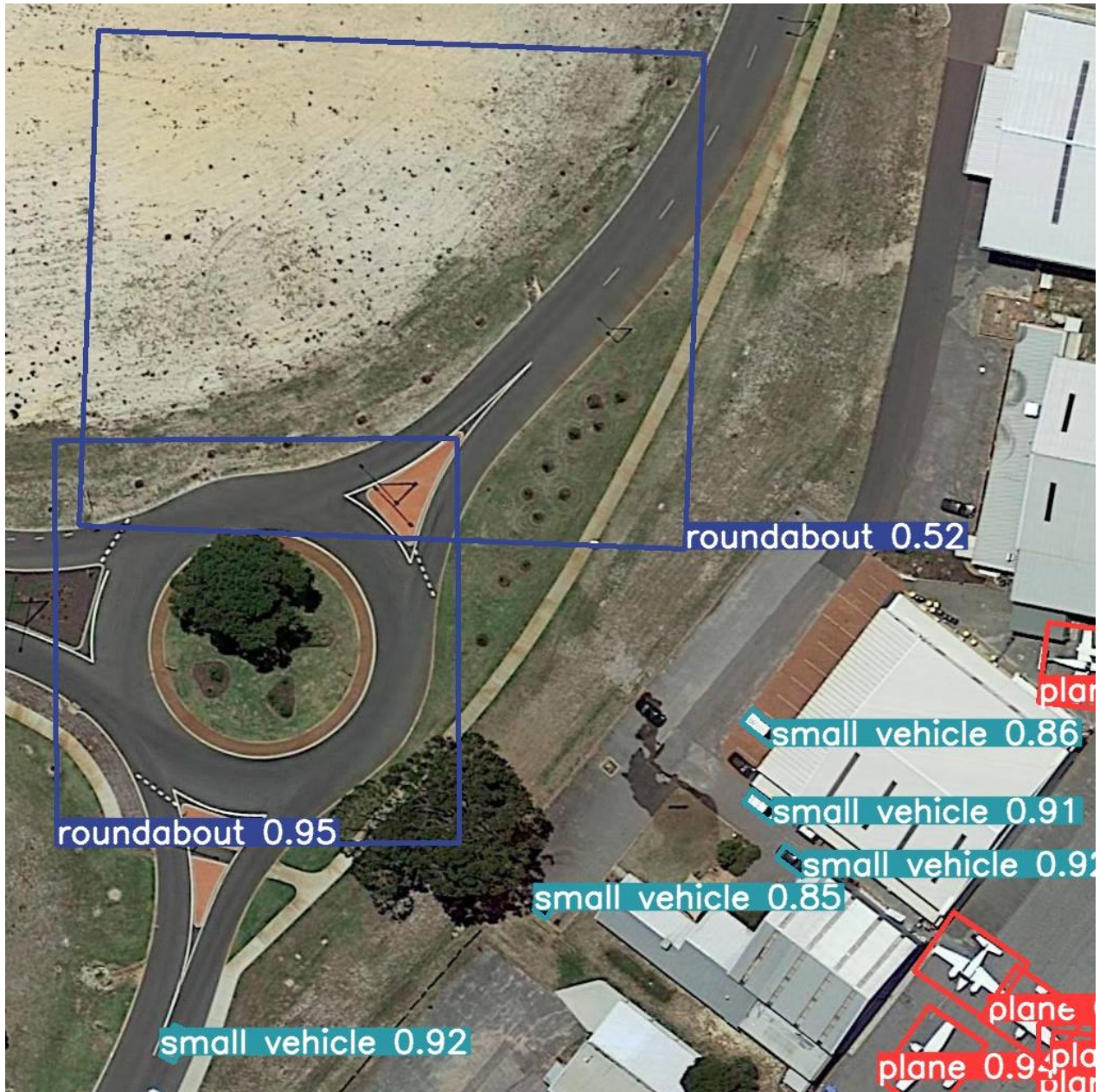


Εικόνα 39: Πρόβλεψη σε εικόνα



Εικόνα 40: Πρόβλεψη σε εικόνα





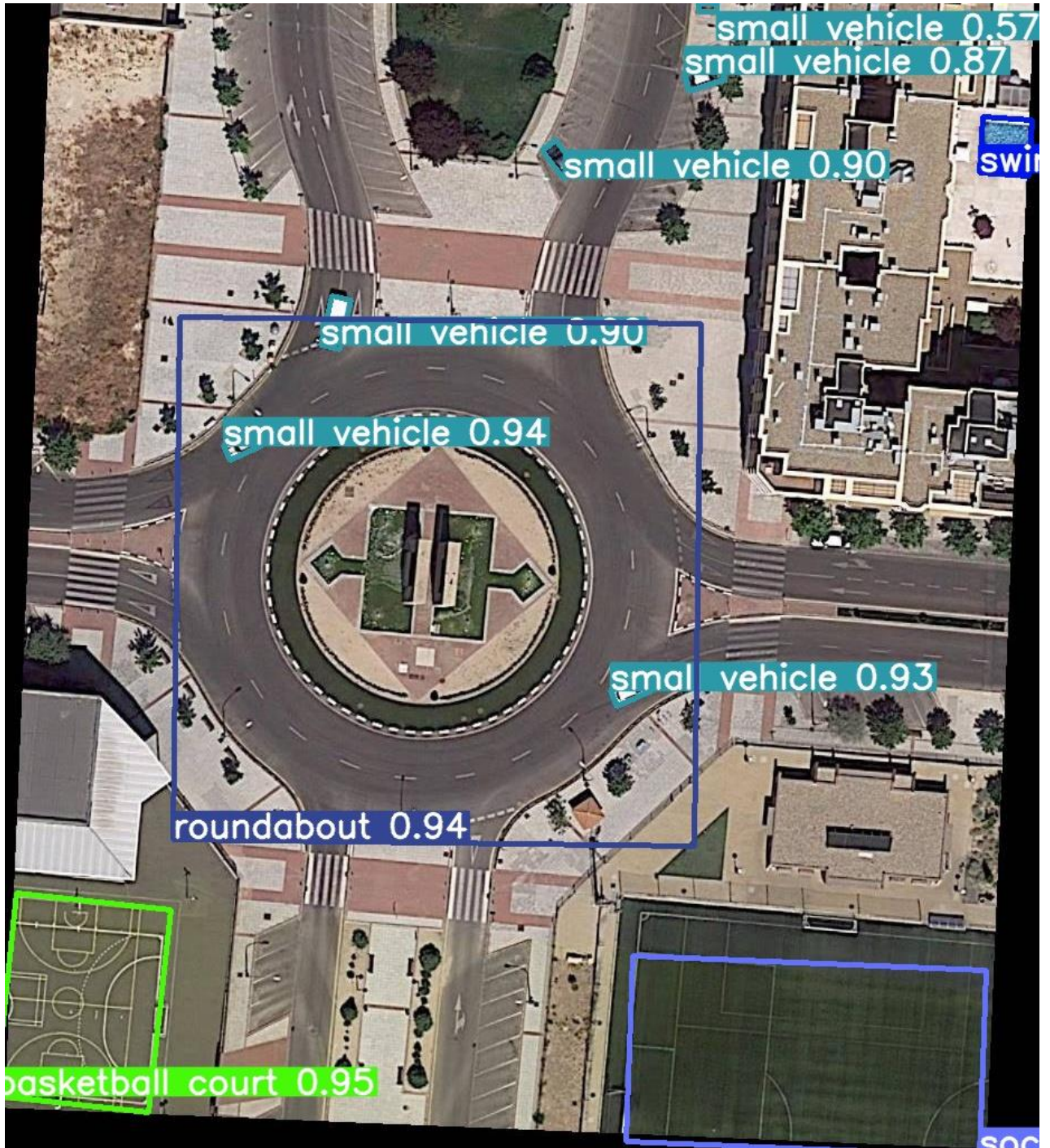
Εικόνα 41: Πρόβλεψη σε εικόνα





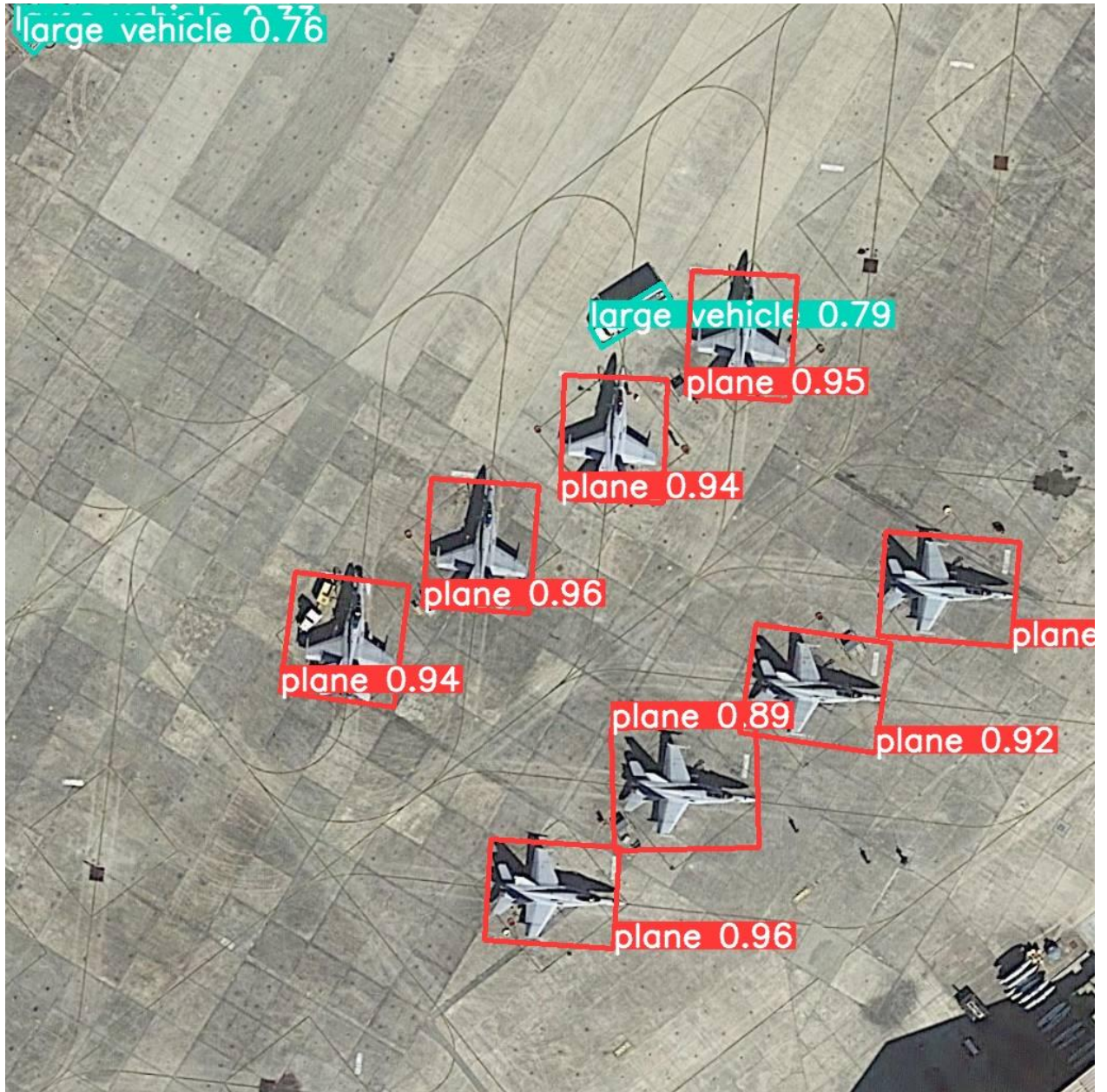
Εικόνα 42: Πρόβλεψη σε εικόνα





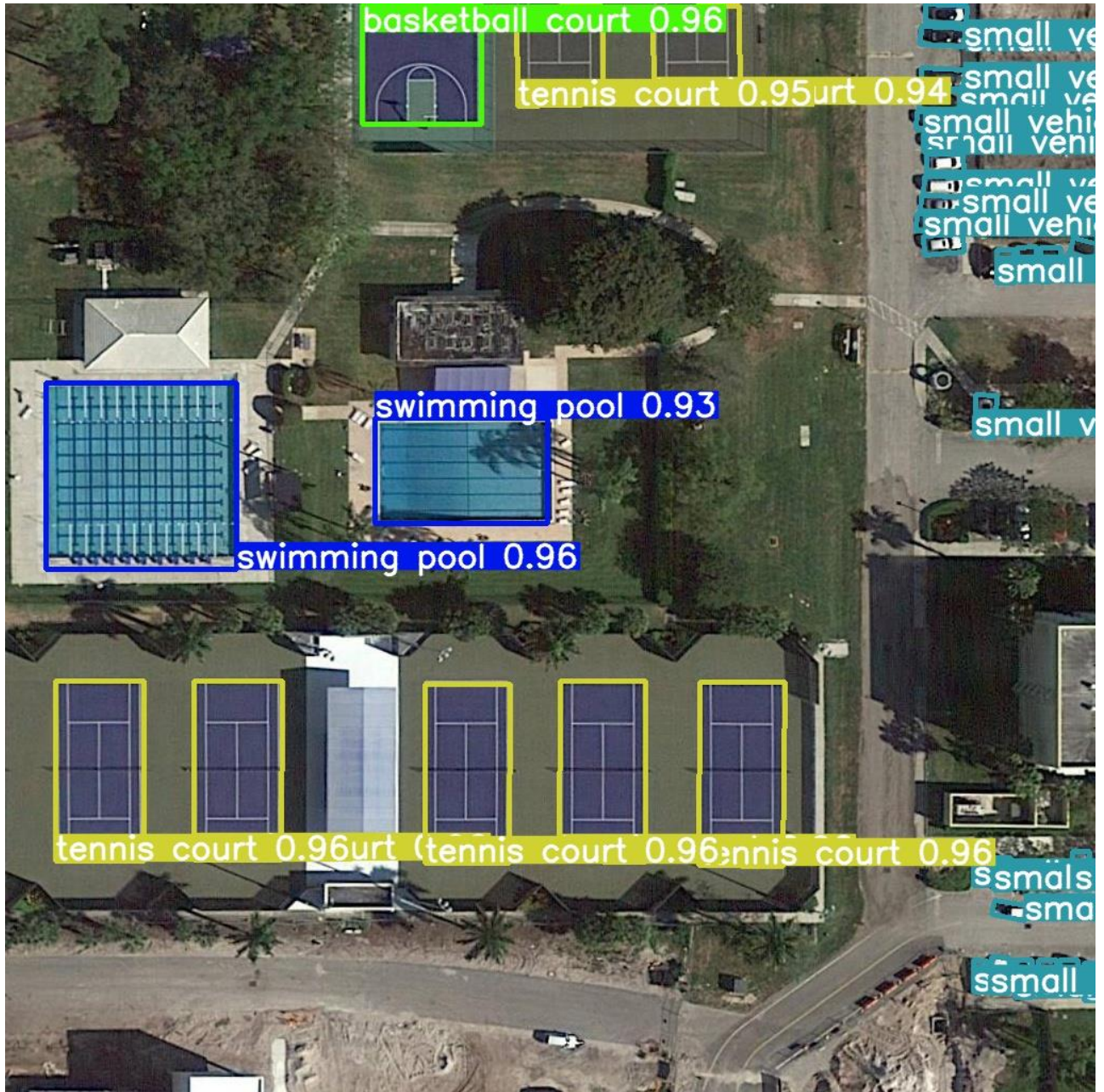
Εικόνα 43: Πρόβλεψη σε εικόνα





Εικόνα 44: Πρόβλεψη σε εικόνα





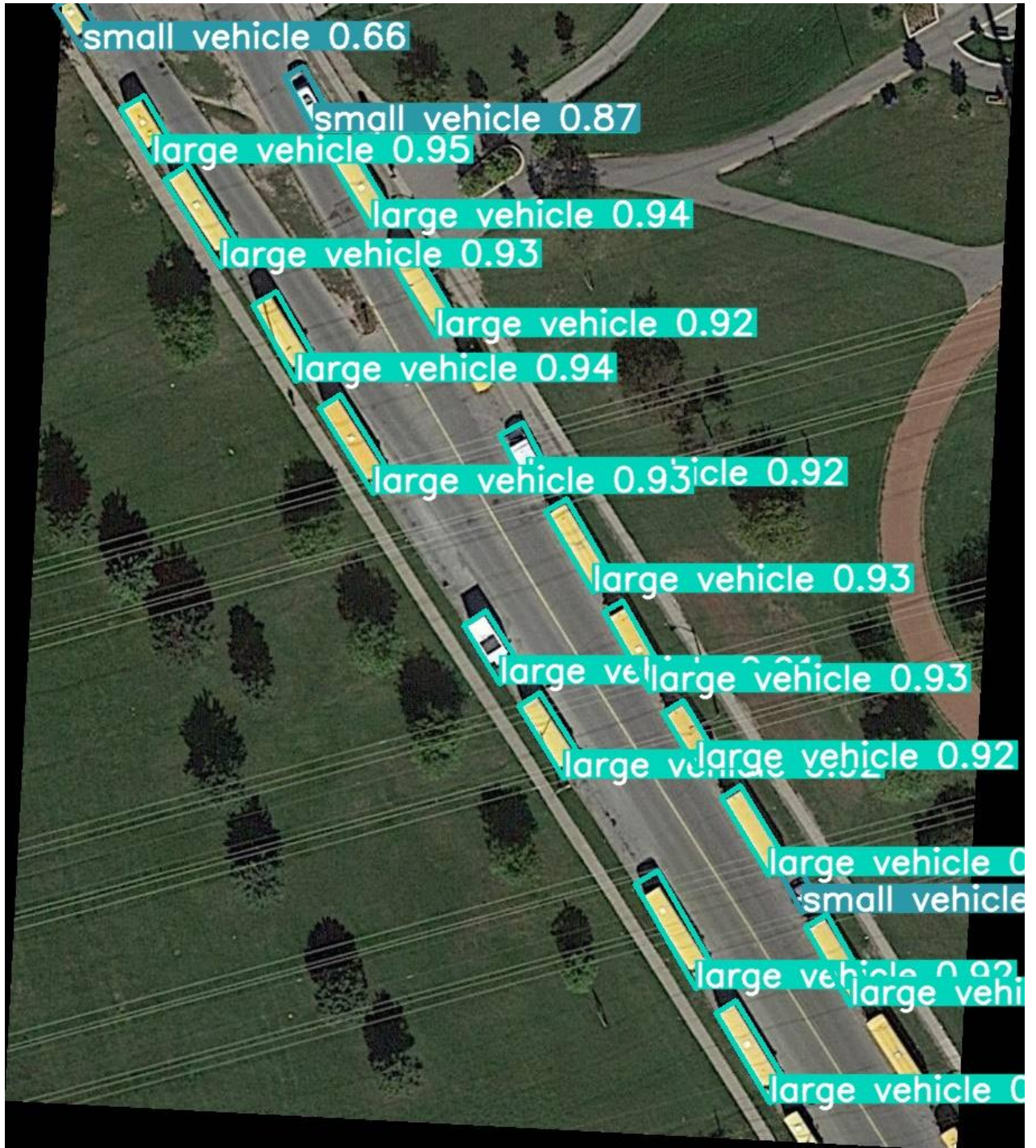
Εικόνα 45: Πρόβλεψη σε εικόνα





Εικόνα 46: Πρόβλεψη σε εικόνα





Εικόνα 47: Πρόβλεψη σε εικόνα





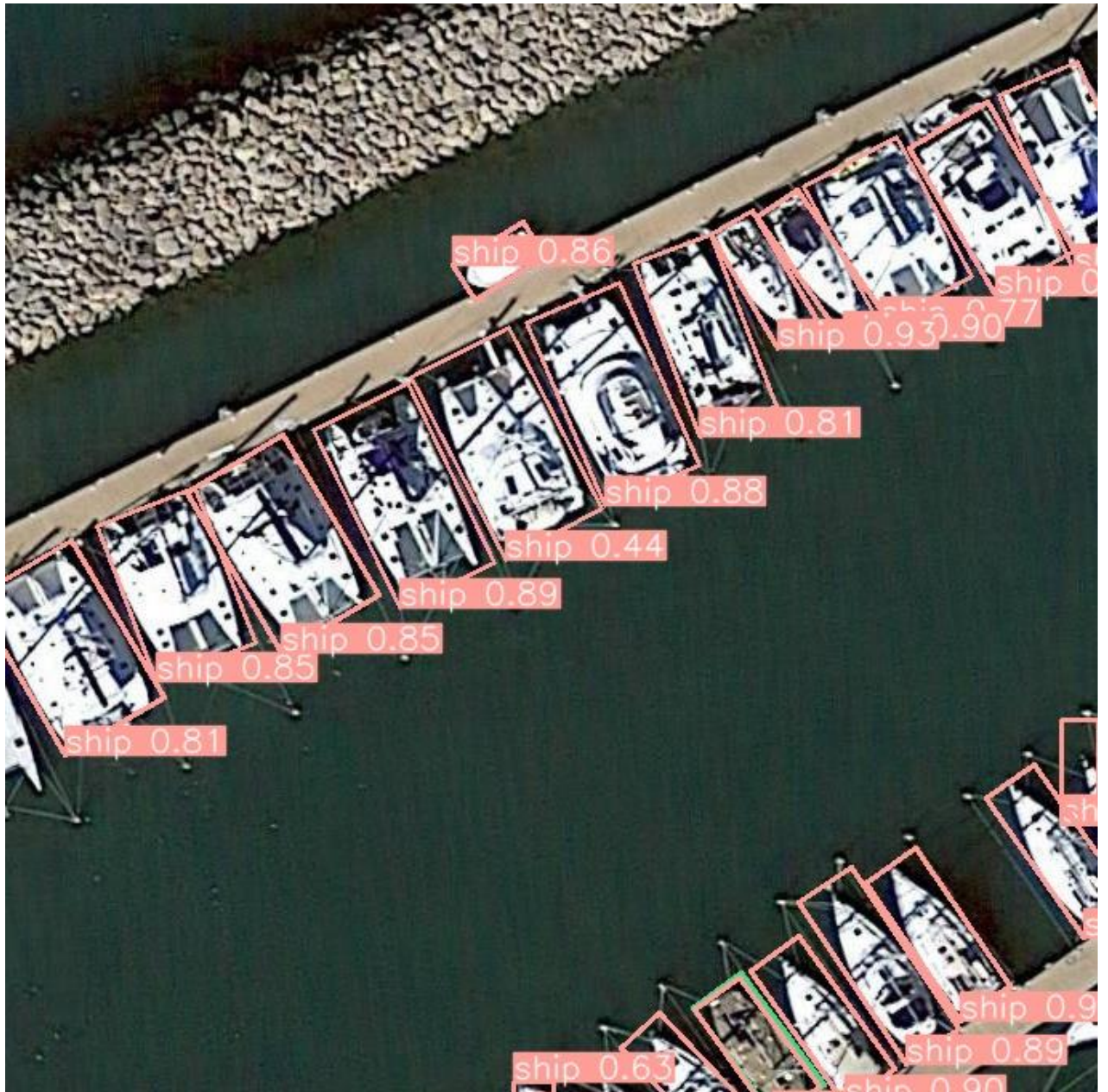
Εικόνα 48: Πρόβλεψη σε εικόνα





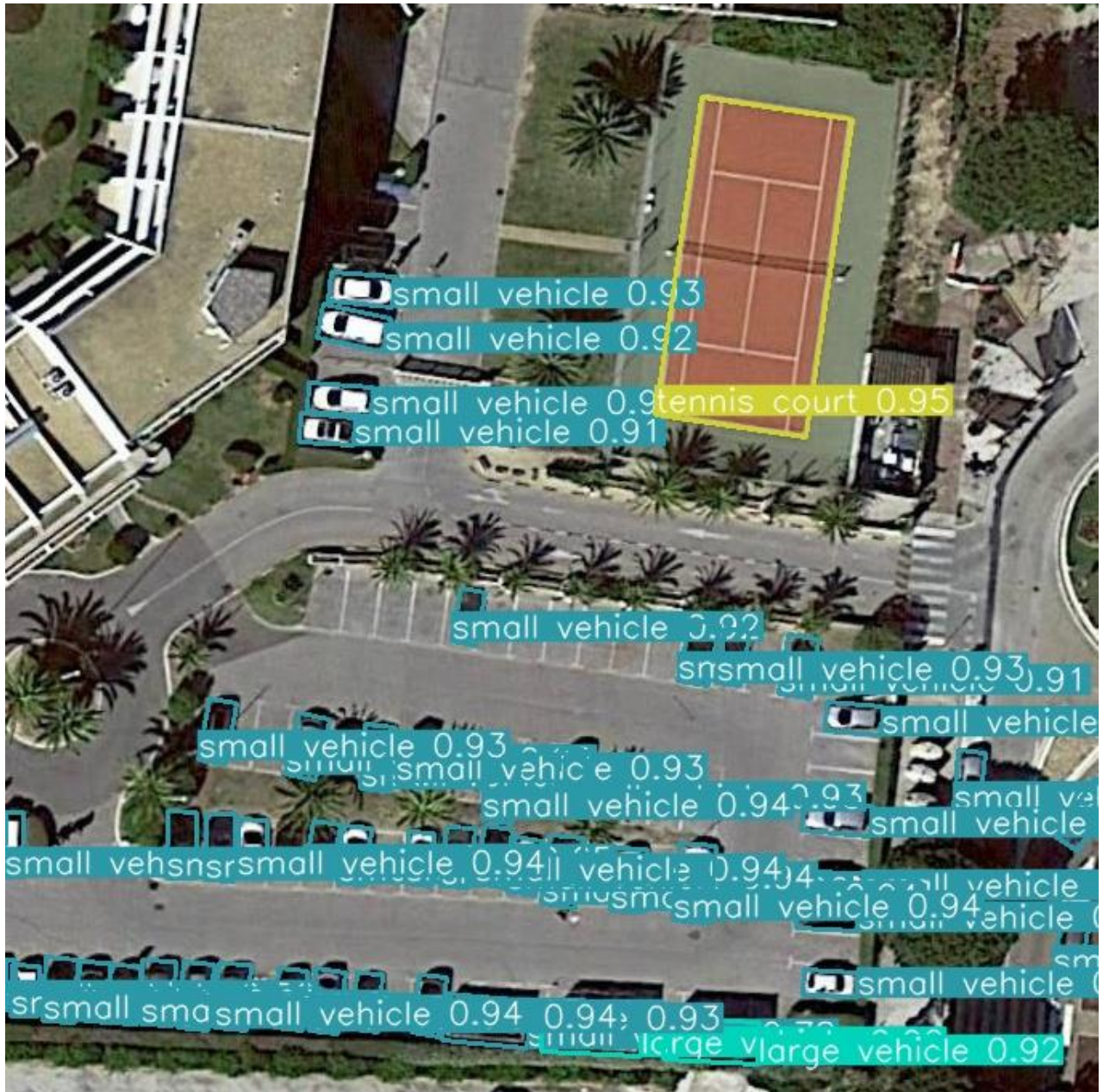
Εικόνα 49: Πρόβλεψη σε εικόνα





Εικόνα 50: Πρόβλεψη σε εικόνα





Εικόνα 51: Πρόβλεψη σε εικόνα

## 15. Συμπεράσματα

Από τα αποτελέσματα των μετρήσεων μέσω της μεθόδου επικύρωσης εξάγονται συμπεράσματα για την αξιολόγηση του μοντέλου Yolov8-OBB στην ανίχνευση αντικειμένων σε δορυφορικές εικόνες.

Το μοντέλο μπορεί να θεωρηθεί ως μια αξιόλογη μεθοδολογία για την ανίχνευση αντικειμένων σε δορυφορικές εικόνες με ικανοποιητικά αποτελέσματα σε μια ποικιλία κατηγοριών αντικειμένων. Θέτοντας το σύνηθες όριο IoU 0.5 για την εγκυρότητα των θετικών προβλέψεων παρατηρούμε ότι το μοντέλο παράγει ικανοποιητικό αριθμό θετικών προβλέψεων για την πλειάδα των κλάσεων που περιέχονται στο σύνολο δεδομένων.

Η εκτεταμένη εκπαίδευση που πραγματοποιήθηκε χρησιμοποιώντας το σύνολο των μεγεθών των βαρών απέδειξε ότι το μέγεθος επηρεάζει την απόδοση και την ακρίβεια του μοντέλου, καθώς όσο μεγαλύτερο είναι τόσο παρουσιάζεται βελτίωση στα αποτελέσματα της ανίχνευσης.

Ένας δεύτερος ιδιαίτερα σημαντικός παράγοντας που έχει σημαντική επίδραση στα αποτελέσματα επηρεάζοντας την απόδοση, την ταχύτητα και την ακρίβεια είναι οι παράμετροι εκπαίδευσης του μοντέλου.

Όπως αναφέρθηκε και στο κεφάλαιο της εκπαίδευσης, οι παράμετροι αυτοί είναι:

- Οι εποχές εκπαίδευσης

Δηλ. ο αριθμός των πλήρη περασμάτων του μοντέλου στο σύνολο δεδομένων του μοντέλου. Να σημειωθεί ότι το framework έχει ως προεπιλογή για το μοντέλο την τιμή 100. Από τις δοκιμές που έγιναν συνιστάται η τιμή αυτή να αυξηθεί έως τις 300 εποχές όχι όμως παραπάνω γιατί είναι πιθανό να παρουσιαστεί το πρόβλημα της υπερπροσαρμογής (overfitting). Η υπερπροσαρμογή είναι μια ανεπιθύμητη συμπεριφορά μηχανικής μάθησης που εμφανίζεται όταν το μοντέλο δίνει ακριβείς προβλέψεις για δεδομένα εκπαίδευσης αλλά όχι για νέα δεδομένα

- Batch

Δηλ. πόσες εικόνες μπορεί να επεξεργάζεται το μοντέλο ταυτόχρονα, πριν ενημερωθούν οι εσωτερικές παράμετροί του. Η αύξηση της συγκεκριμένης τιμής επηρεάζει θετικά την απόδοση του μοντέλου

- Το μέγεθος των εικόνων που θα εισαχθούν για εκπαίδευση

- Και σε αυτή την περίπτωση επηρεάζεται θετικά η ακρίβεια του μοντέλου με την αύξηση της τιμής του μεγέθους των εικόνων που θα τροφοδοτηθεί το μοντέλο.
- Για να εφαρμοστούν οι αυξημένες τιμές στις παραμέτρους απαιτείται, όπως αποδείχθηκε, ισχυρό hardware.

Το κύριο μειονέκτημα του μοντέλου αφορά την απόδοσή του στην ανίχνευση εξαιρετικά μικρών αντικείμενων των δορυφορικών εικόνων. Η απόδοση κρίθηκε, σε ορισμένες περιπτώσεις, προβληματική στα συγκεκριμένα αντικείμενα λόγω της εγγενούς αδυναμίας του αλγορίθμου YOLO στον εντοπισμό τους [6].

Μια πρόταση για μελλοντική έρευνα είναι να δοκιμαστούν και άλλοι αλγόριθμοι εκτός του YOLO, όπως ο RCNN (δύο σταδίων) στην ανίχνευση αντικειμένων σε δορυφορικές εικόνες. Όπως είδαμε στην αξιολόγηση των αλγορίθμων, οι αλγόριθμοι ενός σταδίου, όπως ο YOLO, υπερέχουν στην ταχύτητα και είναι καταλληλότεροι για εφαρμογές πραγματικού χρόνου (Real time). Στην περίπτωση όμως της ανίχνευσης αντικειμένων σε δορυφορικές εικόνες υπάρχουν εφαρμογές που μπορούν να θεωρηθούν ότι δεν είναι πραγματικού χρόνου, όπως η περιβαλλοντική παρακολούθηση και ο πολεοδομικός σχεδιασμός. Σε αυτές θα μπορούσαν να εφαρμοστούν αλγόριθμοι δύο σταδίων που υπερτερούν στην ακρίβεια των εντοπισμένων αντικειμένων.

## Βιβλιογραφία

- [1] Adam Van Etten (2018) You Only Look Twice: Rapid Multi-Scale Object Detection in Satellite Imagery, <https://arxiv.org/pdf/1805.09512>
- [2] Zhong-Qiu Zhao, Shou-tao Xu, Xindong Wu (2019) Object Detection with Deep Learning: A Review, <https://arxiv.org/pdf/1807.05511>
- [3] Alexey Bochkovskiy, Chien-Yao Wang, Hong-Yuan Mark Liao (2020) YOLOv4: Optimal Speed and Accuracy of Object Detection, <https://arxiv.org/pdf/2004.10934>
- [4] Zhong-Qiu Zhao, Member, IEEE, Peng Zheng, Shou-tao Xu, and Xindong Wu, Fellow, IEEE, Object Detection with Deep Learning: A Review
- [5] Asharul Islam Khan, Salim Al-Habsib (2019) Machine Learning in Computer Vision
- [6] Juan R. Terven, Diana M. Cordova-Esparza (2024) A COMPREHENSIVE REVIEW OF YOLO ARCHITECTURES IN COMPUTER VISION: FROM YOLOV1 TO YOLOV8 AND YOLO-NAS
- [7] Deepthi S., Sandeep Kumar, Dr. Suresh L. (2021) Detection and Classification of Objects in Satellite Images using Custom CNN
- [8] Chien-Yao Wang, Alexey Bochkovskiy, Hong-Yuan Mark Liao (2022) YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors
- [9] Dimitrios Mpouziotas, Eleftherios Mastrapas, Nikos Dimokas, Petros Karvelis, Evripidis Glavas (2022) Object Detection for Low Light Images
- [10] Adam Van Etten (2018) Satellite Imagery Multiscale Rapid Detection with Windowed Networks
- [11] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi (2016) You Only Look Once: Unified, Real-Time Object Detection
- [12] Arsalan Tahir, Harif Suliman Munawar, Junaid Akram, Muhammad Adil, Shehryan Ali, Abbas Z. Kouzani, M. A. Parvez Mahmud (2022), Automatic Target Detection from Satellite Imagery Using Machine Learning
- [13] Yann Lecun, Yoshua Bengio, Geoffrey Hinton (2015), Deep learning
- [14] Chaitanya Malladi (2017), Detection of Objects in Satellite using Supervised and Unsupervised Learning Methods
- [15] Gonzalez Rafael, Woods Richard (2018), Digital Image Processing
- [16] What is Satellite imagery, [mapbox.com](https://www.mapbox.com)
- [17] Ultralytics, [Ultralytics.com](https://ultralytics.com)
- [18] Dmitry Spodarets (2023), A Guide to the YOLO Family of Computer Vision Models, [dataphoenix.info](https://dataphoenix.info)

- [19] Dharmaraj (2022), Convolutional Neural Networks (CNN) — Architecture Explained, [medium.com](https://medium.com)
- [20] Ilija Mihajlovic (2019) Everything You Ever Wanted To Know About Computer Vision, [towardsdatascience.com](https://towardsdatascience.com)
- [21] Hmrishav Bandyopadhyay (2022), What Is Computer Vision? [Basic Tasks & Techniques], [v7labs.com](https://v7labs.com)
- [22] The Power of Computer Vision in AI: Unlocking the Future!, [simplilearn.com](https://simplilearn.com)
- [23] Rohit Kundu (2022), Precision vs. Recall: Differences, Use Cases & Evaluation, [v7labs.com](https://v7labs.com)
- [24] Pedro Azevedo (2022), Object Detection State of the Art 2022, [medium.com](https://medium.com)
- [25] Alberto Rizzoli (2021), The Ultimate Guide to Object Detection, [v7labs.com](https://v7labs.com)
- [26] Gaudenz Boesch (2024), Object Detection in 2024: The Definitive Guide, [viso.ai](https://viso.ai)
- [27] Gaudenz Boesch (2023), YOLOv7: A Powerful Object Detection Algorithm (2024 Guide), [viso.ai](https://viso.ai)
- [28] Machine Learning Space Blog, [machinelearningspace.com/](https://machinelearningspace.com/)
- [29] Gaudenz Boesch (2023) Image Annotation: Best Software Tools and Solutions in 2024, [viso.ai](https://viso.ai)
- [30] Hmrishav Bandyopadhyay (2024), Image Annotation: Definition, Use Cases & Types, [viso.ai](https://viso.ai)
- [31] Ελευθέριος Κ. Μαστραπάς, Διπλωματική εργασία (2022), Βελτιστοποίηση εικόνων χαμηλού φωτισμού με χρήση μεθόδων μηχανικής μάθησης
- [32] Jimwang Wang, Jian Ding, Haowen Guo, Wensheng Cheng, Ting Pan, Wen Yang (2019), Mask OBB: A Semantic Attention-Based Mask Oriented Bounding Box Representation for Multi-Category Object Detection in Aerial Images