**ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ**
**ΣΧΟΛΗ ΑΓΡΟΝΟΜΩΝ ΤΟΠΟΓΡΑΦΩΝ ΜΗΧΑΝΙΚΩΝ – ΜΗΧΑΝΙΚΩΝ ΓΕΩΠΛΗΡΟΦΟΡΙΚΗΣ**
**ΔΠΜΣ ΓΕΩΠΛΗΡΟΦΟΡΙΚΗ**

**ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ**

# ΑΥΤΟΜΑΤΗ ΑΝΑΓΝΩΡΙΣΗ ΑΕΡΟΣΚΑΦΩΝ ΣΕ ΥΨΗΛΗΣ ΑΝΑΛΥΣΗΣ ΔΟΡΥΦΟΡΙΚΑ ΔΕΔΟΜΕΝΑ ΜΕ ΣΥΝΕΛΙΚΤΙΚΑ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

**Λέων Χαράλαμπος**

**Αθήνα, Δεκέμβριος 2022**

**NATIONAL TECHNICAL UNIVERSITY OF ATHENS**
**SCHOOL OF RURAL SURVEYING AND GEOINFORMATICS ENGINEERING**
**POST GRADUATE PROGRAMM GEOINFORMATICS**

**MASTER THESIS**

# AYTOMATIC AIRPLANE DETECTION FROM HIGH RESOLUTION IMAGERY SATELLITE DATA WITH CNNs

**Leon Charalampos**

**Athens, December 2022**

Τριμελής εξεταστική επιτροπή

| Κ. Καράντζαλος | Β. Καραθανάση | Π. Κολοκούσης |
|---|---|---|
| ..... | ..... | ..... |
| Αν. Καθηγητής ΕΜΠ<br>Επιβλέπων | Καθηγητής ΕΜΠ | Ε.ΔΙ.Π. ΕΜΠ |

**Χαράλαμπος Α. Λέων**
Πτυχιούχος Στρατιωτικής Σχολής Ευελπίδων, ΣΣΕ 2007

## Ευχαριστίες

Θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα καθηγητή μου κ. Κωνσταντίνο Καράντζαλο για τους νέους ορίζοντες που μου άνοιξε και την εμπιστοσύνη που μου έδειξε να φέρω σε πέρας το έργο αυτό. Θέλω επίσης να ευχαριστήσω τον συνάδελφο Κώστα που μου στάθηκε, με συμβούλευσε και με βοήθησε στη δύσκολη και πρωτόγνωρη αυτή διαδρομή. Τέλος, ένα ευχαριστώ προς την οικογένεια μου για την αμέριστη στήριξη σε όλη τη διάρκεια των σπουδών μου.

Χαράλαμπος Λέων

Δεκέμβριος 2022

# ΠΕΡΙΛΗΨΗ

Η παρούσα εργασία ασχολείται με την μελέτη, διερεύνηση, εφαρμογή και αξιολόγηση μεθόδων Μηχανικής Μάθησης για την αναγνώριση αεροσκαφών σε υψηλής ανάλυσης δορυφορικές εικόνες. Πιο συγκεκριμένα μελετήθηκαν οι τελευταίες μέθοδοι στον τομέα της όρασης υπολογιστών για σκοπούς αναγνώρισης αντικειμένων και εφαρμόστηκε μέθοδος η οποία βρίσκεται στην αιχμή της τεχνολογίας.

Αρχικά παρουσιάζονται βασικά στοιχεία θεωρίας των νευρωνικών δικτύων. Αναφέρεται ο τρόπος λειτουργίας τους, τα συστατικά τους μέρη καθώς και οι διάφοροι τύποι νευρωνικών δικτύων που συναντώνται στις μεθόδους Βαθιάς Μάθησης. Στη συνέχεια αναλύεται η μεθοδολογική προσέγγιση όπως αυτή παρουσιάζεται στο επίσημο αποθετήριο της χρησιμοποιούμενης μεθόδου.

Το dataset που χρησιμοποιήθηκε για την εκπαίδευση του δικτύου αποτελείται από 400 μη πραγματικές εικόνες οι οποίες ελήφθησαν από διαδικτυακή πηγή. Οι εικόνες απεικονίζουν το προς αναγνώριση αντικείμενο σε διάφορα στιγμιότυπα, με μεγάλη παραλακτικότητα σε συνθήκες φωτισμού, υπόβαθρα, μεγέθη και σχήμα. Σκοπός είναι να δημιουργηθεί ένας ισχυρός ανιχνευτής ο οποίος μετά από αξιολόγηση, θα χρησιμοποιηθεί σε real-case σενάριο. Μετά την εκπαίδευση του δικτύου και την μελέτη και εξαγωγή των μετρητικών, ο ανιχνευτής χρησιμοποιήθηκε για δοκιμή σε δορυφορικές εικόνες υψηλής ανάλυσης που απεικονίζουν αεροσκάφη στη βάση τους και όχι σε κάποιο κατασκευασμένο μέρος.

Τέλος παρουσιάζεται η εφαρμογή των μεθόδων αυτών πάνω στις δορυφορικές εικόνες και αξιολογούνται τα αποτελέσματα. Παρατηρούνται οι προβλέψεις σε κάθε εικόνα, και εξάγονται συμπεράσματα σχετικά με την αποτελεσματικότητα και λειτουργία των μεθόδων.

## ABSTRACT

In this thesis the study, investigation, application and evaluation of Machine Learning methods for the identification of aircraft in high resolution satellite images is attempted. More specifically, the latest methods in the field of computer science for object detection purposes were studied and applied.

Firstly, the basic elements of neural network theory are presented. Their mode of operation, their component parts as well as the various types encountered in Deep Learning methods are mentioned. Then the methodological approach is analyzed as it is presented in the official repository of the used method.

The dataset used to train the network consists of 400 computer-based images obtained from an online source. The images depict the object to be identified in various positions, with important variations in lighting conditions, backgrounds, sizes and shape. The purpose is to create a robust detector which, after evaluation, was used in a real-case scenario. After training the neural network, studying and extracting the metrics, the detector was used for testing on high-resolution satellite images depicting aircraft at their initial base.

Finally, the results of these methods on satellite images is presented and evaluated. The predictions in each image are observed, and conclusions are made regarding the operational effectiveness of the methods.

## Κατάλογος Περιεχομένων

# 1. ΕΙΣΑΓΩΓΗ

## 1.1 Image Intelligence

Η πληροφορία οποιασδήποτε μορφής και πηγής θεωρείται σημαντικός παράγοντας δύναμης και εθνικής ασφάλειας. Οι δυνατότητες ενός κρατικού οργανισμού σε ένοπλες δυνάμεις, δυνάμεις εσωτερικής και εξωτερικής ασφάλειας όσο προηγμένες και να είναι, δεν μπορούν να αναπτύξουν το μέγιστο των ικανοτήτων τους αν δεν υπάρχει αντίστοιχη υποστήριξη σε θέματα πληροφοριών. Ωστόσο η διαδικασία συλλογής, επεξεργασίας, διάθεσης και εκμετάλλευσης αποτελεί εγχείρημα ιδιαίτερα πολύπλοκο, με πολλά στάδια και πολλές διαφορετικές συνιστώσες.

Προκειμένου να επιτευχθεί αυτή η αξιοποίηση της πληροφορίας έχουν αναπτυχθεί διεθνώς προτυποποιήσεις που σκοπό έχουν να καθορίσουν διαδικασίες, καλές πρακτικές και πράξεις που θα πρέπει να λάβουν χώρα σε κάθε δυνατή περίπτωση με γνώμονα την παροχή επεξεργασμένων πληροφοριών. Οι προτυποποιήσεις αυτές έχουν οδηγήσει στην διακριτοποίηση των διαφόρων πηγών πληροφοριών αναλόγως της φύσης της πληροφορίας. Διαφορετικές κατηγορίες αποτελούν:

- Opensource intelligence (OSINT)
- Signal intelligence (SIGINT)
- Human intelligence (HUMINT)
- Acoustic intelligence (ACINT)
- Image Intelligence (IMINT)

Το Image Intelligence είναι από τις πρώτες χρονικά παραπάνω κατηγορίες ανάλυσης πληροφοριών που χρησιμοποιήθηκε, διότι το αποτέλεσμα τους είναι άμεσα αντιληπτό από τις φυσικές ανθρώπινες αισθήσεις, σε σύγκριση με τις υπόλοιπες κατηγορίες που απαιτούν κάποια ψηφιακή επεξεργασία. Περιλαμβάνει όλες τις πληροφορίες που μπορούν να συλλεχθούν μέσω φωτογραφιών, αεροφωτογραφιών και δορυφορικών εικόνων.

Τα τελευταία χρόνια η εξέλιξη της τεχνολογίας έχει οδηγήσει στην αύξηση της χωρικής ανάλυσης των Δορυφορικών Εικόνων, με ταυτόχρονη βελτίωση των τεχνικών για την αυτόματη εξαγωγή χαρακτηριστικών. Η διαδικασία IMINT ως κατεξοχήν δραστηριότητα κατά την οποία επιδιώκεται να αντληθούν πληροφορίες από δορυφορικές εικόνες, είναι ένα παράδειγμα κατά το οποίο η αυτόματη εξαγωγή και ο εντοπισμός των αντικειμένων έχει επιχειρησιακή αξία.

## 1.2 Κίνητρο

Η ανάπτυξη μεθόδων Μηχανικής Μάθησης για πλήθος εφαρμογών είναι πλέον στην αιχμή της τεχνολογίας. Πλήθος ερευνητικές ομάδες, πανεπιστήμια, δημόσιοι αλλά και ιδιωτικοί φορείς ανά τον κόσμο έχουν στρέψει τις προσπάθειες τους στην εξέλιξη και βελτίωση των μεθόδων. Απόδειξη αποτελεί ο μεγάλος αριθμός δημοσιεύσεων που επιτυγχάνεται κάθε χρόνο πάνω στον ευρύτερο τομέα. Η προοπτική και η επίδραση που φαίνεται ότι θα διαδραματίσει ο τομέας έχει οδηγήσει μεγάλους οργανισμούς (πολιτικούς και στρατιωτικούς) όπως το ΝΑΤΟ, η ΕΕ και η EDA (European Defence Agency) να χρηματοδοτούν προγράμματα διερεύνησης για την αποτελεσματικότερη χρήση της τεχνολογίας σε θέματα του ενδιαφέροντος τους.

Αποτέλεσμα της συνεχούς αυτής αναζήτησης αποτελεί η διείσδυση της τεχνολογίας και εκμετάλλευσης της για εμπορικούς σκοπούς. Πέρα από την επιστημονική διάσταση και την εξέλιξη των τεχνικών και των μεθόδων για σκοπούς παραγωγής νέας γνώσης, ιδιωτικοί φορείς ευελπιστούν να επωφεληθούν από την νέα αυτή κατάσταση. Εκμεταλλευόμενοι την τεράστια υποστήριξη που υπάρχει, την διάθεση των μεθόδων στο ευρύ κοινό, τα κεφάλαια που έχουν διατεθεί και κυρίως την τεράστια ζήτηση που υπάρχει για την αυτοματοποίηση των μεθόδων και την χρήση τους με πιο φιλικό τρόπο στον τελικό χρήστη, έχουν αναπτύξει λογισμικά τα οποία και διαθέτουν έναντι τεράστιας αμοιβής.

Έχοντας υπόψη τα ανωτέρω, κίνητρο για την ενασχόληση με την παρούσα εργασία αποτέλεσε η προσπάθεια προώθησης από ιδιωτική εταιρεία λογισμικού Τεχνητής Νοημοσύνης, η οποία έχει ενσωματώσει αλγορίθμους ελεύθερα διαθέσιμους για την αυτόματη αναγνώριση αντικειμένων μέσα από ΔΕ, έναντι τεράστιας αμοιβής. Επιδιώκεται μέσα από την εργασία αυτή αρχικά να μελετηθούν οι μέθοδοι και οι τεχνικές από την οικεία επιστημονική βιβλιογραφία, να εφαρμοστεί μια μέθοδος και να αξιολογηθεί ως προς την αποδοτικότητάς της και τέλος να αποδειχθεί ότι ακόμα και ένας μη εξοικειωμένος χρήστης μπορεί πλέον να χρησιμοποιήσει τις τεχνικές αυτές με τη βοήθεια της υποστήριξης που υπάρχει παγκοσμίως.

## 1.3 Αντικείμενο Διπλωματικής

Η παρούσα μεταπτυχιακή εργασία ασχολείται με την μελέτη και την εφαρμογή μεθόδων Μηχανικής Μάθησης για την αναγνώριση σταθμευμένων αεροσκαφών εντός δορυφορικών εικόνων. Ερευνάται η ικανότητα των εν λόγω μεθόδων, να αναγνωρίζουν το υπόψη αντικείμενο σε οποιαδήποτε διάσταση, μέγεθος, προσανατολισμό, σχήμα και είδος εμφανίζονται, καθώς και σε ποικιλία από διαφορετικά υπόβαθρα. Για να επιτευχθεί ο στόχος μελετήθηκαν οι τελευταίες τεχνικές του τομέα της Μηχανικής Μάθησης με χρήση συνελικτικών νευρωνικών δικτύων για σκοπούς αναγνώρισης αντικειμένων, και εφαρμόστηκε η τελευταία χρονολογικά μέθοδος που έχει απελευθερωθεί σε αποθετήριο στο GitHub.

Δόθηκε έμφαση στο συγκεκριμένο αντικείμενο καθώς είναι το συχνότερο που επιδιώκεται να αναγνωριστεί κατά την διαδικασία IMINT. Η παρουσία και ορθή αναγνώριση του συγκεκριμένου αντικειμένου, ο αριθμός τους και ο τύπος τους είναι πληροφορίες που έχουν ιδιαίτερη επιχειρησιακή αξία στους υπεύθυνους που λαμβάνουν αποφάσεις και στην εξαγωγή συμπερασμάτων-δράσεων. Επιπλέον η επιτυχής αναγνώριση του αντικειμένου σημαίνει ότι ακολουθώντας την ίδια μεθοδολογία θα μπορεί να αναγνωριστεί οποιοδήποτε άλλο αντικείμενο από το σύνολο των απαιτούμενων στην διαδικασία IMINT.

Η μεθοδολογία που ακολουθήθηκε ήταν σύμφωνη με τις πρακτικές που ακολουθούνται κατά την εκπαίδευση νευρωνικών δικτύων. Το dataset αντλήθηκε από διαδικτυακή πηγή παροχής δεδομένων ενώ τα συνοδευτικά αρχεία που περιγράφουν το αντικείμενο δημιουργήθηκαν από την αρχή. Το σύνολο των δεδομένων χρησιμοποιήθηκε για την εκπαίδευση αλγορίθμου αιχμής και σε τελικό στάδιο δοκιμάστηκε η αποδοτικότητα του σε σύνολο δεδομένων αξιολόγησης, καθώς και σε υψηλής ανάλυσης δορυφορικές εικόνες που απεικονίζουν αεροσκάφη επί αεροδρομίων.

Τέλος αναφέρεται ότι η ανάπτυξη ενός αλγορίθμου αιχμής είναι πέρα από τους σκοπούς της εργασίας, καθώς εκτιμάται ότι οι ήδη διαθέσιμες μέθοδοι είναι επαρκείς για την επίλυση πολλών προβλημάτων.

# 2. ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ – ΑΝΑΣΚΟΠΗΣΗ ΒΙΒΛΙΟΓΡΑΦΙΑΣ

## 2.1 Τεχνητή Νοημοσύνη - Μηχανική Μάθηση – Βαθιά Μάθηση

Στον τομέα της επιστήμης υπολογιστών ο όρος τεχνητή νοημοσύνη (Artificial Intelligence) είναι μια ευρεία έννοια η οποία αναφέρεται στην ικανότητα οποιασδήποτε ανθρωπογενούς κατασκευής (πχ υπολογιστής) να αναπτύσσει κάποιας μορφής νοημοσύνης. Η ικανότητα αυτή αποσκοπεί στην μίμηση της ανθρώπινης νοημοσύνης ως βασικός λειτουργός μάθησης νέας γνώσης και επίλυσης προβλημάτων. Μέσα από την προσέγγιση αυτή επιδιώκεται ο υπολογιστής να είναι σε θέση να αντιλαμβάνεται το περιβάλλον, να εξάγει περαιτέρω γνώση ή να μαθαίνει να εξάγει νέα γνώση. Η ικανότητα αυτή ενώ στους ανθρώπους γίνεται μέσα από εμπειρίες, για τον υπολογιστή γίνεται μέσα από υπολογιστικές διαδικασίες που επεξεργάζονται δεδομένα. Σημαντική παράμετρο αποτελεί επίσης το γεγονός ότι οι υπολογιστικές αυτές διαδικασίες δεν βασίζονται σε ντετερμινιστικές μεθόδους αλλά περιλαμβάνουν ως επί το πλείστο στοχαστικές διαδικασίες. Εφαρμογές οι οποίες βασίζονται στην τεχνητή νοημοσύνη αποτελούν η επεξεργασία φυσικής γλώσσας, η ρομποτική, η όραση υπολογιστών, η μηχανική μάθηση κα.

Ο όρος Μηχανική Μάθηση (Artificial Intelligence) αποτελεί υποκατηγορία της Τεχνητής Νοημοσύνης ο οποίος χρησιμοποιήθηκε για πρώτη φορά από τον Arthur Samuel το 1959[1]. Μέσω της Μηχανικής Μάθησης μελετώνται οι τεχνικές εκείνες που θα επιτρέπουν στον υπολογιστή να εκτελεί Τεχνητή Νοημοσύνη πάνω στις εφαρμογές που αναφέρθηκαν παραπάνω. Επίκεντρο των τεχνικών αυτών αποτελεί η όσο το δυνατόν μικρότερη παρέμβαση του ανθρώπου στις διαδικασίες μάθησης. Οι τεχνικές Μηχανικής Μάθησης χωρίζονται σε δύο κατηγορίες στην επιβλεπόμενη και μη μάθηση. Στην πρώτη κατηγορία το σύστημα μαθαίνει να αναγνωρίζει δεδομένα μέσα από αρχικά δεδομένα που το αποτέλεσμα τους είναι γνωστό. Στην δεύτερη κατηγορία το σύστημα δεν γνωρίζει εξαρχής το αποτέλεσμα αλλά αντιλαμβάνεται κοινά στοιχεία στα δεδομένα με αποτέλεσμα να προβλέπει το αποτέλεσμα μέσω από υπολογισμούς.

Ο όρος βαθιά μάθηση (Deep learning) αποτελεί υποκατηγορία της Μηχανικής Μάθησης και αναφέρεται σε αλγορίθμους με μεγάλη πολυπλοκότητα σε σχέση με τους συμβατικούς αλγορίθμους Μηχανικής Μάθησης. Η αύξηση της πολυπλοκότητας των αλγορίθμων ήταν συνάρτηση της έξαρσης στην υπολογιστική ισχύ των υπολογιστών που επέτρεψε να πραγματοποιούνται σύνθετοι υπολογισμοί σε πολύ σύντομο χρόνο.

Κοινό χαρακτηριστικό των παραπάνω αποτελεί η χρήση των Νευρωνικών Δικτύων (Neural Networks) που χρησιμοποιούνται στις εφαρμογές Μηχανικής

[1] Αμερικανός επιστήμονας στον τομέα των υπολογιστών ο οποίος έκανε γνωστό τον όρο Τεχνητή Νοημοσύνη το 1959. Επινόησε ένα είδος αυτό-εκπαιδευόμενου παιχνιδιού το οποίο αποτέλεσε το πρώτο παράδειγμα της έννοιας Τεχνητή Νοημοσύνη.

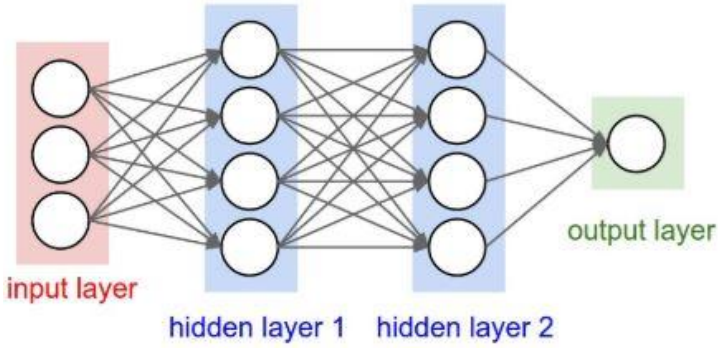
Μάθησης και Βαθιάς Μάθησης. Τα Νευρωνικά δίκτυα έχουν και αυτά κατηγορίες όπως τεχνητά Νευρωνικά Δίκτυα, Συνελικτικά Νευρωνικά Δίκτυα κα.

## 2.2 Νευρωνικά Δίκτυα

### 2.2.1 Βασικά στοιχεία

Τα νευρωνικά δίκτυα αποτελούν προσπάθεια μοντελοποίησης του τρόπου που ο ανθρώπινος εγκέφαλος εκτελεί την διαδικασία της δικής του απόκτησης νέας γνώσης. Μοντελοποιούνται μαθηματικά ως μια συλλογή από νευρώνες είτε πλήρως είτε μερικώς συνδεδεμένοι μεταξύ τους όπως στην παρακάτω εικόνα 2.1. Η παρακάτω εικόνα απεικονίζει τρία επίπεδα νευρώνων (input, hidden layer 1, hidden layer 2) τα οποία οδηγούν σε ένα επίπεδο τελικού αποτελέσματος (output layer). Η ονομασία νευρώνας εμπνεύστηκε από τον βιολογικό νευρώνα του ανθρώπινου νευρικού συστήματος, ο οποίος λειτουργεί με παρόμοιο τρόπο μεταδίδοντας πληροφορία.

Η απεικόνιση αυτή μπορεί να θεωρηθεί ως ένα απλό νευρωνικό δίκτυο το οποίο δέχεται ως είσοδο κάποιο δεδομένο το διοχετεύει εκτελώντας υπολογιστικές διαδικασίες στα κρυμμένα επίπεδα και παράγει μια έξοδο.



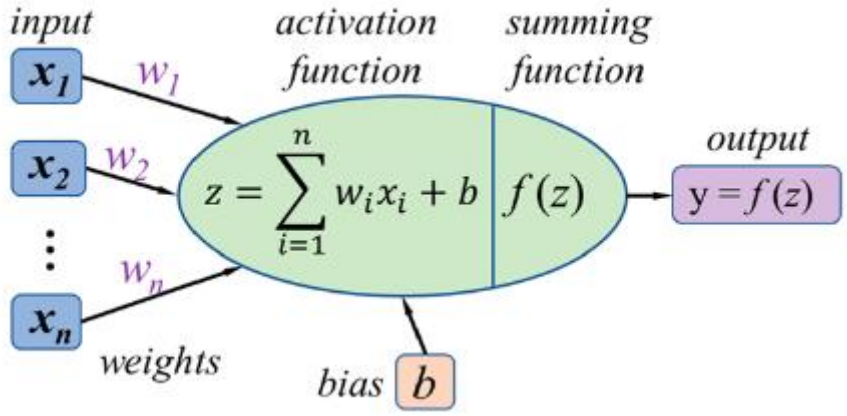

**Εικόνα 2.1:** Πλήρως συνδεδεμένοι νευρώνες
Πηγή: https://www.semanticscholar.org/paper/Multi-Class-Object-Detection-from-Aerial-Images-Schweitzer-Agrawal/

Το παραπάνω απλό δίκτυο αποτελείται από 9 νευρώνες (τέσσερις στο πρώτο κρυμμένο επίπεδο, τέσσερις στο δεύτερο και ένα στο τελικό επίπεδο). Ο κάθε νευρώνας μπορεί να θεωρηθεί ως ένας χώρος ο οποίος περιέχει μια μαθηματική τιμή κατωφλίου. Οι συνδέσεις μεταξύ των νευρώνων, οι οποίες απεικονίζονται με τα προσανατολισμένα βέλη, μπορούν να θεωρηθούν ως χώροι στους οποίους τοποθετούνται αναπροσαρμόσιμες μαθηματικές τιμές που καλούνται βάρη. Στην παραπάνω απλή απεικόνιση υπάρχουν 32 τιμές βαρών (3x4 + 4x4 + 4x1), μία σε κάθε σύνδεση.

Οι τιμές που περιγράφησαν παραπάνω αποτελούν τις παραμέτρους του δικτύου για τις οποίες απαιτείται η εύρεση της βέλτιστης τιμής τους. Αναφέρεται για

λόγους σύγκρισης ότι τα σύγχρονα συνελικτικά νευρωνικά δίκτυα μπορούν να διαθέτουν αριθμό νευρώνων με τάξη μεγέθους εκατομμυρίου και παραπάνω από 20 κρυμμένα συνελικτικά επίπεδα.

Η διαδικασία μετάδοσης της πληροφορίας διαμέσου των νευρώνων, από το αρχικό επίπεδο έως το τελικό εξαγόμενο αποτέλεσμα στο τελευταίο layer, καλείται forward pass. Η παρακάτω εικόνα 2.2 απεικονίζει τις υπολογιστικές διαδικασίες που εκτελούνται σε κάθε έναν από τους νευρώνες ενός πλήρους νευρωνικού δικτύου.

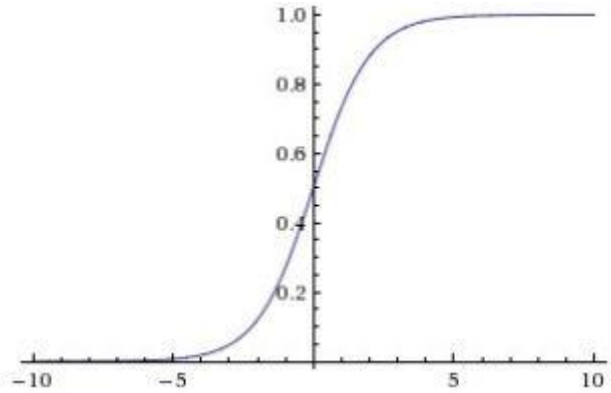



**Εικόνα 2.2 :** Διαδικασία forward pass σε έναν νευρώνα του δικτύου.

Κατά τη διαδικασία forward pass, ο κάθε νευρώνας του κάθε επιπέδου συμμετέχει σε μια μαθηματική πράξη που περιγράφεται από μια συνάρτηση. Η συνάρτηση αυτή καλείται συνάρτηση ενεργοποίησης και είναι σταθερή, δηλαδή χαρακτηρίζει ολόκληρο το δίκτυο. Στην παραπάνω εικόνα 2.3, απεικονίζεται γραφικά η εφαρμογή της συνάρτησης σε έναν τυχαίο νευρώνα. Ο κάθε νευρώνας δέχεται τιμές από νευρώνες του προηγούμενου επιπέδου (*Xi*) και εκτελεί το άθροισμα των πολλαπλασιασμών τιμών και βαρών σε κάθε σύνδεση, και προσθέτει το αποτέλεσμα στην τιμή κατωφλίου του κάθε νευρώνα. Ακολούθως η τιμή αυτή διοχετεύεται σε μια συνάρτηση ενεργοποίησης και το αποτέλεσμα αποτελεί είσοδο στο επόμενο επίπεδο. Οι πιο συχνά χρησιμοποιούμενες συναρτήσεις ενεργοποίησης έχουν ως ακολούθως:

- **Σιγμοειδής συνάρτηση**

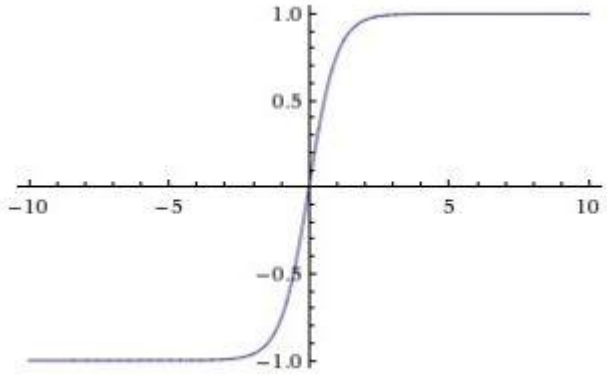
Η σιγμοειδής καμπύλη ονομάζεται από το γράμμα S που θυμίζει η γραφική της παράσταση. Χρησιμοποιείται για να μετατρέψει τις εισαγόμενες τιμές σε κανονικοποιημένο διάστημα (0, 1). Χρησιμοποιείται αρκετά συχνά σε προβλήματα ταξινόμησης όπου οι πιθανές εξαγόμενες τιμές είναι μόνο δύο. Είναι ιστορικά η πρώτη που χρησιμοποιήθηκε. Ο μαθηματικός της τύπος είναι $f(x) = \frac{1}{1+e^{-x}}$ και η γραφική της παράσταση παρουσιάζεται στην παρακάτω εικόνα 2.3.




**Εικόνα 2.3 :** Σιγμοειδής καμπύλη.

- **Tanh**

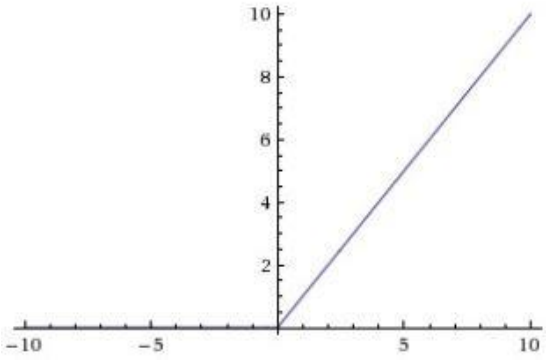
Η συνάρτηση υπερβολικής εφαπτομένης έχει ομοιότητες με την σιγμοειδή, με την διαφορά ότι οι εξαγόμενες τιμές βρίσκονται στο διάστημα [-1,1]. Προτιμάται από τη σιγμοειδή καμπύλη διότι το σύνολο των τιμών της είναι κεντροβαρικό ως προς το μηδέν, γεγονός που διευκολύνει την εκπαίδευση του δικτύου καθώς αποτρέπει την ανανέωση των παραμέτρων σε μια μόνο κατεύθυνση. Το διάγραμμα της παρουσιάζεται στην παρακάτω Εικόνα 2.4.




**Εικόνα 2.4 :** Καμπύλη υπερβολικής εφαπτομένης.

- **ReLu - Rectified Linear Unit**

Η συνάρτηση ReLu κατωφλιώνει τις τιμές μέσω της σχέσης $f(x) = \max(0, x)$ εξάγοντας μόνο θετικές τιμές. Είναι η πιο διαδεδομένη από τις προγενέστερες καθώς έχει αποδειχθεί ότι η χρήση της γενικά επιταχύνει τη διαδικασία της εκπαίδευσης του δικτύου. Συγκριτικά με τις προηγούμενες, κατά τη χρήση της το υπολογιστικό κόστος είναι μικρότερο διότι υπολογίζεται εύκολα σαν γραμμική στις θετικές τιμές και μηδενική στις αρνητικές. Το διάγραμμα της συνάρτησης παρουσιάζεται στην παρακάτω εικόνα 2.5

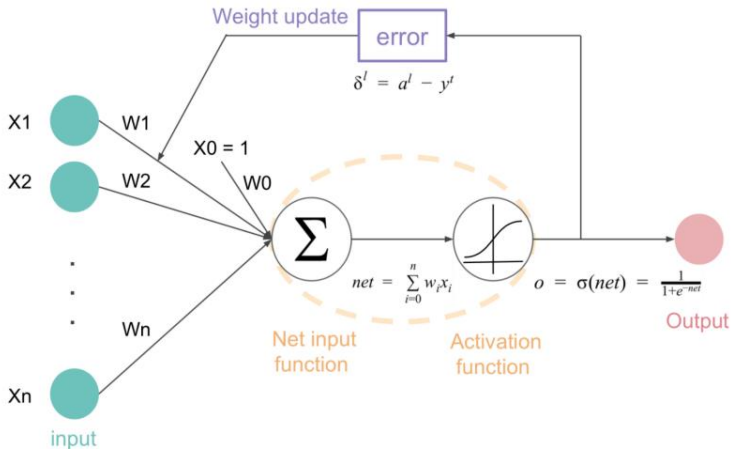



**Εικόνα 2.5 :** Καμπύλη συνάρτησης ReLu.

### 2.2.2 Εκπαίδευση δικτύου

Εκπαίδευση του δικτύου είναι η διαδικασία κατά την οποία επιδιώκεται να βρεθεί η βέλτιστη τιμή των παραμέτρων ώστε να ελαχιστοποιείται μια χρησιμοποιούμενη συνάρτηση κόστους. Αρχικώς οι τιμές του δικτύου αρχικοποιούνται με τυχαίες τιμές και ακολούθως μέσω μιας επαναληπτικής διαδικασίας, οι τιμές αναπροσαρμόζονται έως να επιτευχθεί το χαμηλότερο δυνατό κόστος.

Η συνάρτηση κόστους χρησιμοποιείται για να υποδηλώσει την αβεβαιότητα του εξαγόμενου αποτελέσματος. Στην ουσία η συνάρτηση συγκρίνει το εξαγόμενο αποτέλεσμα με την αληθή τιμή (η οποία ορίζεται από τον χρήστη πριν την εκπαίδευση) και εξάγει μια τιμή αβεβαιότητας. Μεγάλη τιμή κόστους υποδηλώνει μεγάλη αβεβαιότητα, ενώ μικρή τιμή υποδηλώνει μικρή αβεβαιότητα. Οι πιο γνωστές μέθοδοι υπολογισμού του κόστους αποτελούν οι Support Vector Machine και Softmax.




**Εικόνα 2.6:** Διαδικασία αναπροσαρμογής των παραμέτρων.
Πηγή: https://medium.com/coinmonks/implement-back-propagation-in-neural-networks

Στην παραπάνω εικόνα 2.6 απεικονίζεται η λειτουργία της επανάληψης που πραγματοποιείται σε κάθε νευρώνα του δικτύου. Ως γενικός κανόνας επιδιώκεται να βρεθεί η μεταβολή δW, ώστε τα νέα βάρη W+δW να δίνουν μικρότερη τιμή κόστους.

Η ελαχιστοποίηση του κόστους σε κάθε βήμα επιτυγχάνεται υπολογίζοντας τις μερικές παραγώγους της συνάρτησης κόστους ως προς τις παραμέτρους σε κάθε νευρώνα. Η διαδικασία υπολογισμού των μερικών παραγώγων της συνάρτησης κόστους ως προς τα βάρη του δικτύου λέγεται Gradient Descent. Η διαδικασία της συνεχούς αναπροσαρμογής των παραμέτρων υπολογίζοντας τις μερικές παραγώγους είναι γνωστή και ως οπισθοδιάδοση (backpropagation) και αποτελεί την καρδιά της εκπαίδευσης του δικτύου. Αναλυτικά τα γεγονότα της εκπαίδευσης ενός ταξινομητή για να αναγνωρίσει ένα αντικείμενο έχουν ως ακολούθως.

Αρχικά οι τυχαίες παράμετροι έχουν εξάγει ένα τυχαίο αποτέλεσμα της ταξινόμησης. Επειδή όμως είναι γνωστό το διάνυσμα της αληθούς κλάσης του αντικειμένου, μπορεί να ποσοτικοποιηθεί η διαφορά μεταξύ προβλεπόμενης και αληθούς κλάσης μέσω της συνάρτησης κόστους. Υπολογίζοντας την μερική παράγωγο της συνάρτησης κόστους, υποδεικνύεται η ποσότητα που θα πρέπει να μεταβληθούν οι τιμές των βαρών και κατωφλίων, ώστε το κόστος να μειωθεί κατά ποσότητα ίση με την διαφορά που υπολογίστηκε στο πρώτο πέρασμα της διαδικασίας.

Για να αναπροσαρμοστούν οι παράμετροι υπολογίζεται το γινόμενο της μεταβολής των παραμέτρων με μια τιμή βήματος. Η διαδικασία ακολούθως επαναλαμβάνεται έως ότου επιτευχθεί η όσο το δυνατόν μικρότερη τιμή της συνάρτησης κόστους ή έως ότου η κάθε νέα αναπροσαρμογή να μην δίνει ουσιαστικό αποτέλεσμα στα νέα βάρη. Το μέγεθος του βήματος είναι γνωστό ως learning rate, και καθορίζει το ρυθμό με τον οποίο γίνεται η αναπροσαρμογή των παραμέτρων. Η τιμή του βήματος είναι γενικά αντικείμενο διερεύνησης και πειραματισμού διότι στην περίπτωση που ο αριθμός είναι μικρός τότε το δίκτυο θα απαιτεί αρκετό χρόνο να εκπαιδευθεί, ενώ εάν είναι μεγάλος το δίκτυο ενδεχομένως να οδηγήσει σε λάθος αποτελέσματα.

Σε πραγματικές εφαρμογές όπου τα δεδομένα εκπαίδευσης αποτελούνται από πολύ μεγάλο αριθμό εικόνων, οι παράμετροι του δικτύου μπορούν να φτάσουν πολύ μεγάλους αριθμούς. Ο μεγάλος αριθμός των παραμέτρων καθιστά τη διαδικασία υπολογισμού των μερικών παραγώγων για όλα τα δεδομένα στο ίδιο πέρασμα να απαιτεί πολύ μεγάλο υπολογιστικό κόστος. Η αντιμετώπιση αυτού του προβλήματος γίνεται με τον διαχωρισμό των δεδομένων εκπαίδευσης σε μικρά πακέτα δεδομένων, και η εφαρμογή της διαδικασίας σε κάθε πακέτο ξεχωριστά. Το πακέτο αυτό ονομάζεται αριθμός mini batch και αποτελεί παράμετρο προς διερεύνηση του δικτύου.

Η διαδικασία εκπαίδευσης γίνεται κάθε φορά με αριθμό εικόνων ίσο με τον αριθμό mini batch, και όταν έχουν ολοκληρωθεί όλες οι εικόνες τότε λέγεται ότι έχει ολοκληρωθεί μία εποχή εκπαίδευσης. Με τον τρόπο αυτό η διαδικασία της

σταδιακής αναπροσαρμογής των παραμέτρων επιτυγχάνεται σε πολύ μικρότερο χρόνο. Η τιμή του βέλτιστου αριθμού mini batch συνήθως εκλέγεται διαδοχικά δοκιμάζοντας δυνάμεις του 2 (πχ 32, 64, 128). Ωστόσο η διαδικασία που περιγράφηκε αφορά την εκπαίδευση ενός απλού νευρωνικού δικτύου. Για σκοπούς ταξινόμησης εικόνας χρησιμοποιούνται τα συνελικτικά νευρωνικά δίκτυα.

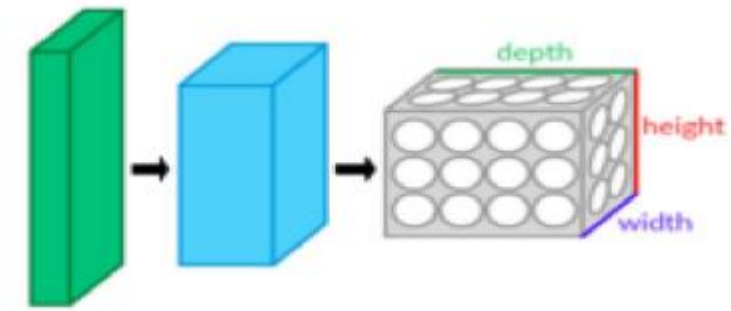
## 2.3 Συνελικτικά νευρωνικά δίκτυα

### 2.3.1 Βασικά στοιχεία

Τα Συνελικτικά νευρωνικά δίκτυα (Convolutional Neural Networks - CNN) διαδόθηκαν αρκετά και έγιναν ευρέως γνωστά για σκοπούς ταξινόμησης εικόνας κατά το διαγωνισμό ImageNet (Krizhevsky et al., 2012). Η χρήση τους είναι ήδη γνωστή από τη δεκαετία του 90, ωστόσο οι περιορισμοί σε υπολογιστική δύναμη δεν επέτρεπε την περαιτέρω χρήση και αξιοποίηση τους.

Τα δομικά στοιχεία ενός συνελικτικού νευρωνικού δικτύου είναι όμοια με τα απλά. Αποτελούνται από ένα σύνολο νευρώνων στους οποίους αντιστοιχεί μια τιμή βάρους. Το δίκτυο χρησιμοποιεί μια διαφορίσιμη συνάρτηση, η οποία λαμβάνει ως είσοδο τα pixel μιας εικόνας και εξάγει βαθμολογίες για κάθε κλάση. Επιπλέον σε αντιστοιχία με τα απλά νευρωνικά δίκτυα χρησιμοποιείται μια συνάρτηση κόστους η οποία ποσοτικοποιεί τη διαφορά της εξαγόμενης από την αληθή κατηγορία. Η κύρια διαφορά τους σε σχέση με τα απλά δίκτυα είναι ότι δέχονται ως είσοδο ολόκληρη την εικόνα, αντί για ένα διάνυσμα μιας κατεύθυνσης.

Στην παρακάτω εικόνα 2.7 απεικονίζεται σχηματικά η είσοδος μιας τριδιάστατης εικόνας σε κάθε επίπεδο. Κάθε επίπεδο μεταμορφώνει τον όγκο εισόδου σε έναν τριδιάστατο όγκο εξόδου από ενεργοποιημένους νευρώνες μέσω μιας παραγωγίσιμης συνάρτησης ενεργοποίησης.




**Εικόνα 2.7 :** Αρχιτεκτονική συνελικτικού νευρωνικού δικτύου.
Πηγή: https://ijarsct.co.in/Paper1033.pdf

Το τελευταίο επίπεδο εξόδου θα έχει επίσης τρεις διαστάσεις (1x1xN), η αρχική εικόνα θα έχει μειωθεί σε ένα τριδιάστατο διάνυσμα με τις N βαθμολογίες της κάθε κλάσης.

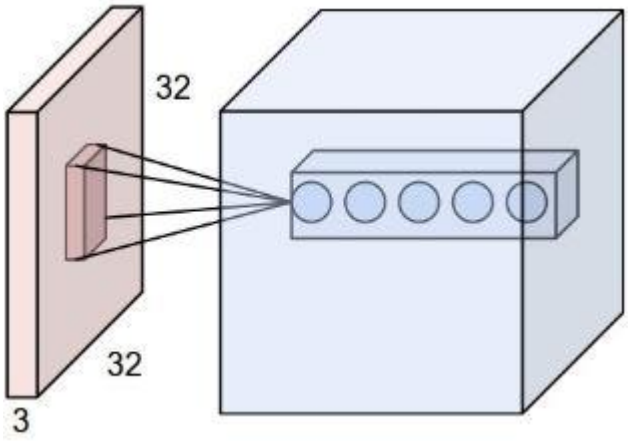
Η αρχιτεκτονική ενός συνελικτικού νευρωνικού δικτύου αποτελείται από τρία διαφορετικά επίπεδα. Οι τύποι αυτοί είναι τα συνελικτικά επίπεδα (Convolutional layers), τα Pooling επίπεδα και τα πλήρως συνδεδεμένα επίπεδα (Fully-connected layers). Η αλληλουχία των τριών αυτών τύπων στο δίκτυο μετασχηματίζει τον αρχικό όγκο της εικόνας σε έναν τελικό αποτέλεσμα με τις πιθανότητες των κλάσεων. Παρακάτω παρουσιάζονται συνοπτικά τα τρία αυτά κύρια επίπεδα συνελικτικών νευρωνικών δικτύων.

### 2.3.2 Συνελικτικό επίπεδο

Το συνελικτικό επίπεδο αποτελεί το βασικότερο στοιχείο ενός συνελικτικού νευρωνικού δικτύου, καθώς εκτελεί το μεγαλύτερο όγκο της υπολογιστικής διαδικασίας κατά την εκπαίδευση του δικτύου. Κατά την διαδικασία εκπαίδευσης επιδιώκεται η βέλτιστη τιμή των βαρών του δικτύου. Σε ένα συνελικτικό επίπεδο τα βάρη είναι οι τιμές διάφορων συνελικτικών χωρικών φίλτρων που χρησιμοποιούνται από το δίκτυο. Το μέγεθος του φίλτρου ποικίλει, αλλά συνήθως είναι μικρό (πχ 5x5x3). Καθώς το φίλτρο διαδοχικά διαπερνά όλη την εικόνα εκτελώντας τις συνελίξεις, δημιουργείται ένας χάρτης χαρακτηριστικών με τα αποτελέσματα των συνελίξεων.

Κατά τη διαδικασία της εκπαίδευσης το δίκτυο θα εκπαιδεύσει τα φίλτρα με τέτοιο τρόπο ώστε αυτά θα ενεργοποιούνται όταν θα υπάρχει στην εικόνα κάποιο είδος οπτικού χαρακτηριστικού (πχ μια ακμή). Με τον τρόπο αυτό θα δημιουργηθούν οι τιμές των φίλτρων σε κάθε επίπεδο, όπου το καθένα θα δημιουργεί έναν δισδιάστατο πίνακα χαρακτηριστικών, όμοιο με τις διαστάσεις της αρχικής εικόνας. Η διαδοχή αυτών των εξαγόμενων πινάκων κατά τη διάσταση του βάθους δημιουργεί τον όγκο του κάθε επόμενου συνελικτικού επιπέδου.

Ο κάθε νευρώνας του συνελικτικού επιπέδου συνδέεται μόνο με μια μικρή περιοχή της εικόνας εισόδου η οποία ισούται με τη διάσταση του φίλτρου, και όχι με ολόκληρη την περιοχή της εικόνας. Η χωρική έκταση αυτής της συνδεσιμότητας είναι μια από τις υπερπαραμέτρους του δικτύου (receptive field). Για παράδειγμα έστω μια εικόνα εισόδου με διαστάσεις 32x32x3 (Εικόνα 2.8), εάν η διάσταση του φίλτρου είναι 5x5, τότε κάθε νευρώνας στο συνελικτικό επίπεδο θα έχει συνδέσεις για μια περιοχή 5x5x3 αντί για ολόκληρη την εικόνα.




**Εικόνα 2.8** *:* Παράδειγμα όγκου εικόνας 1ου συνελικτικού επιπέδου.
Πηγή: https://cs217.stanford.edu/weightlayers

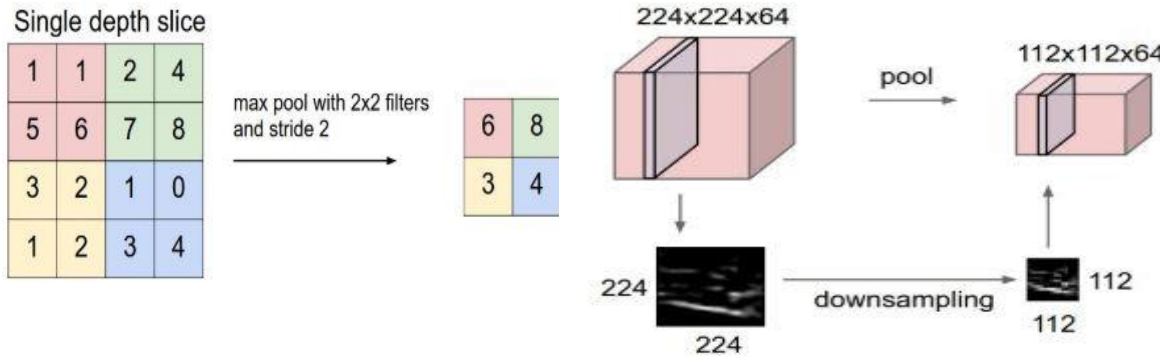
Το μέγεθος του κάθε επόμενου εξαγόμενου όγκου εξαρτάται από τρεις παραμέτρους του δικτύου οι οποίες καθορίζονται από τον χρήστη. Αυτές είναι:

- Το βάθος (depth)
- Το βήμα μετακίνησης του φίλτρου (stride)
- Η επέκταση της αρχικής εικόνας με μηδενικά (zero-padding).

Το βάθος του εξαγόμενου όγκου εξαρτάται από τον αριθμό των φίλτρων που χρησιμοποιούνται. Το βήμα μετακίνησης του φίλτρου καθορίζει τις διαστάσεις του εξαγόμενου επιπέδου κατά το επίπεδο xy. Όταν το βήμα είναι 1 το φίλτρο μετακινείται κατά μία θέση ενώ όταν το βήμα είναι 2 το φίλτρο μετακινείται κατά 2 θέσεις δημιουργώντας εξαγόμενους όγκους μικρότερων διαστάσεων. Η διάσταση της αρχικής εικόνας διατηρείται σε συνδυασμό με την επέκταση της αρχικής εικόνας με μηδενικά κατά πλάτος και ύψος (zero-padding = 1). Εάν δεν χρησιμοποιούταν η επέκταση αυτή τότε σταδιακά οι διαστάσεις των εξαγόμενων όγκων θα μειώνονταν κατά μήκος και πλάτος.

### 2.3.3 Pooling layer

Τα pooling επίπεδα δεν εισάγουν καμία επιπλέον παράμετρο στο δίκτυο διότι εκτελούν μια σταθερή συνάρτηση στο επίπεδο εισόδου. Τοποθετούνται συνήθως ανάμεσα στα συνελικτικά επίπεδα και κύρια εργασία τους είναι να μειώσουν σταδιακά τις διαστάσεις της αρχικής εικόνας και κατά συνέπεια να μειωθεί ο αριθμός των εκπαιδευόμενων παραμέτρων. Λειτουργούν ξεχωριστά σε κάθε επίπεδο ενώ ο πιο συνηθισμένος τύπος αποτελείται από φίλτρα διαστάσεων 2x2 τα οποία εφαρμόζονται σε βήμα 2 θέσεων. Στην εικόνα 2.9 η συνάρτηση δέχεται ως είσοδο 4 αριθμούς και επιλέγεται ο μέγιστος από αυτούς. Η διάσταση του βάθους παραμένει αμετάβλητη ενώ οι υπόλοιπες διαστάσεις μειώνονται κατά το ήμισυ απορρίπτοντας με αυτόν τον τρόπο το 75% των παραμέτρων.




**Εικόνα 2.9 :** Λειτουργία max Pooling.

### 2.3.4 Πλήρως συνδεδεμένα επίπεδα

Τα πλήρως συνδεδεμένα επίπεδα έχουν συνδέσεις με όλους τους νευρώνες του προηγούμενου επιπέδου σε αντιστοιχία με τα απλά νευρωνικά δίκτυα. Ωστόσο η συνδεσιμότητα αυτή δεν αφορά ολόκληρη την εικόνα, αλλά μόνο με μια μικρή περιοχή του προηγούμενου επιπέδου.

Η πιο συνηθισμένη αρχιτεκτονική νευρωνικών δικτύων αποτελείται από διαδοχικά συνελικτικά επίπεδα ακολουθούμενα από Max Pooling επίπεδα. Το πρότυπο αυτό επαναλαμβάνεται μέχρι ο τελικός όγκος να έχει μικρύνει αρκετά σε διαστάσεις. Στα τελικά επίπεδα είναι σύνηθες ο τελικός όγκος να διοχετεύεται σε πλήρη συνδεδεμένα επίπεδα, τα οποία υπολογίζουν τις βαθμολογίες για κάθε κλάση.

## 2.4 Αναγνώριση Αντικειμένων – Object Detection

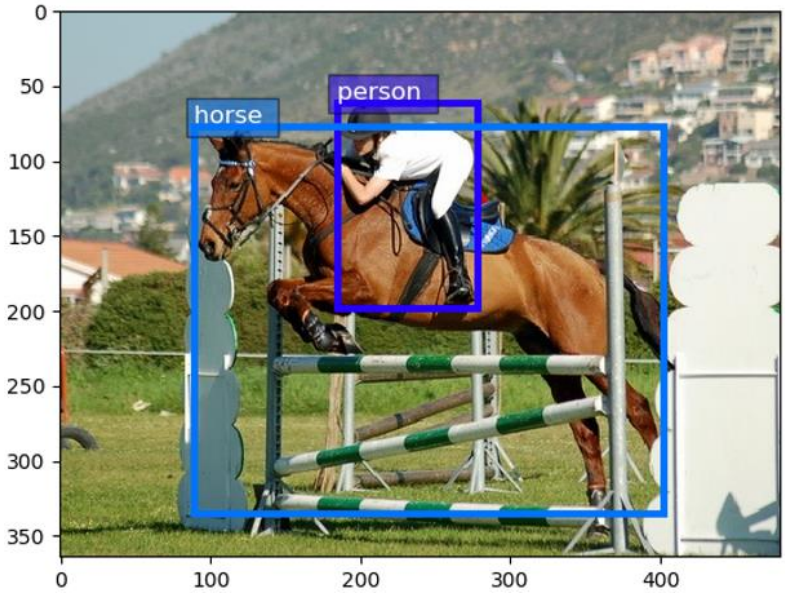
### 2.4.1 Εισαγωγή

Στην ταξινόμηση εικόνας σκοπός είναι σε ολόκληρη την εικόνα να αποδοθεί ένας χαρακτηρισμός αναλόγως του κυρίαρχου αντικειμένου που αποτυπώνεται. Το μοντέλο επιδιώκει να συνοψίσει στην εικόνα ένα ποιοτικό χαρακτηρισμό γνωστό ως labeling της εικόνας. Γενικά στο πρόβλημα της ταξινόμησης εικόνας, ένα μόνο αντικείμενο εμφανίζεται το οποίο αρκεί για να χαρακτηριστεί η εικόνα με βάση αυτό το μοναδικό αντικείμενο, όπως χαρακτηριστικά απεικονίζονται στην παρακάτω εικόνα 2.10. Στην παρακάτω εικόνα παρουσιάζονται τρία παραδείγματα και οι χαρακτηρισμοί τους από το κυρίαρχο αντικείμενο που αποτυπώνεται. Ο χαρακτηρισμός συνήθως ακολουθείται από μια βαθμολογία ή οποία αντιπροσωπεύει την πιθανότητα ορθής πρόβλεψης του αλγορίθμου.


**Εικόνα 2.10:** Προβλήματα ταξινόμησης εικόνων
Πηγή: https://vitalflux.com/classification-problems-real-world-examples/

Ωστόσο σε προβλήματα πραγματικού κόσμου, υπάρχουν πολλά σενάρια κατά τα οποία δεν αρκεί να αποδοθεί ένας χαρακτηρισμός σε ολόκληρη την εικόνα, οπότε η ταξινόμηση από μόνη της δεν αρκεί. Χαρακτηριστικό παράδειγμα απεικονίζεται στην παρακάτω εικόνα 2.11 όπου ένας χαρακτηρισμός για όλη την εικόνα δεν μπορεί να επιτευχθεί καθώς εμφανίζονται δύο κύρια αντικείμενα. Η πολυπλοκότητα του προβλήματος ανεβαίνει όταν στην εικόνα αποτυπώνονται πολλά αντικείμενα ή όταν πρόκειται για Δορυφορικές εικόνες ή αεροφωτογραφίες, όπου η σκηνή αποτύπωσης αποτελείται συνήθως από φυσική γήινη επιφάνεια στην οποία ενδεχομένως να υπάρχουν ανθρωπογενή στοιχεία.

Στην παρακάτω εικόνα το μοντέλο δεν ταξινόμησε απλώς την εικόνα, αλλά έχει εντοπίσει και τα έχει περικλείσει εντός περιγεγραμμένου τετραγώνου δύο αντικείμενα στα οποία έχει αποδώσει έναν χαρακτηρισμό (*person, horse*). Το πρόβλημα αυτό είναι γνωστό ως object detection (αναγνώριση αντικειμένου) και αποτελεί κατά έναν τρόπο εξέλιξη της ταξινόμησης εικόνας.




**Εικόνα 2.11**: Πρόβλημα αναγνώρισης αντικειμένου.
Πηγή: https://towardsdatascience.com/build-your-own-deep-learning-classification-model-in-keras-511f647980d6

Το πρόβλημα της αναγνώρισης αντικειμένων έχει εφαρμογή σε πολλά προβλήματα του πραγματικού κόσμου. Πέρα από την επιστήμη της όρασης υπολογιστών (όπως στην παραπάνω εικόνα 2.11) η επίλυση του προβλήματος έχει σημαντικό ρόλο στην επιστήμη της τηλεπισκόπησης επίσης. Ενδεικτικές εφαρμογές μπορεί να είναι στην γεωργία, στις περιβαλλοντικές αλλαγές, χρήσης/κάλυψη γης, επιτήρηση, πολεοδομικός σχεδιασμός (urban planning), χαρτογραφία κ.α. [20]. Στην παρούσα εργασία χρησιμοποιήθηκε για να υποβοηθήσει την διαδικασία του IMINT, με την οποία επιδιώκεται να αναγνωρίζονται τα αντικείμενα που αποτυπώνονται σε πάσης είδους οπτικών απεικονίσεων.

Η διαδικασία της αναγνώρισης αντικειμένων επεκτείνει την ταξινόμηση εικόνας και περιλαμβάνει δύο επιμέρους διακριτές διεργασίες. Κατά την πρώτη επιδιώκεται να βρεθεί η θέση του αντικειμένου στην εικόνα, διαδικασία γνωστή ως localization. Ο εντοπισμός μπορεί να γίνει εσωκλείοντας το αντικείμενο εντός περιγεγραμμένου τετραγώνου (Εικόνα 2.12), ενώ ακόμη έχουν αναπτυχθεί μέθοδοι κατά την οποία η διαδικασία εντοπίζει το πλήρες σχήμα του αντικειμένου [21], [26]. Στην παρακάτω εικόνα 2.13 απεικονίζεται παράδειγμα κατά το οποίο τα εξαγόμενα αντικείμενα έχουν εντοπιστεί και χρωματιστεί σε όλο το εύρος του σχήματος τους. Το εξαγόμενο αποτέλεσμα χρησιμοποιείται για την εξαγωγή διανυσματικών δεδομένων από δορυφορικές εικόνες για σκοπούς εμπλουτισμού βάσης χωρικών δεδομένων για περαιτέρω χαρτογραφική επεξεργασία.

Κατά την δεύτερη διεργασία, αφού έχει βρεθεί το αντικείμενο, επιδιώκεται να ταξινομηθεί σε μία από τις κλάσεις εκπαίδευσης. Επομένως όλα τα προβλήματα που συναντώνται κατά την ταξινόμηση, εμφανίζονται αντίστοιχα και στο στάδιο αυτό. Επιπλέον προστίθεται το πρόβλημα του εντοπισμού και της ταχύτητας

εκτέλεσης, καθώς ενδεχομένως σε μια εικόνα να περιλαμβάνονται αρκετός αριθμός αντικειμένων.


**Εικόνα 2.12:** Εφαρμογή Mask R-CNN σε δορυφορική απεικόνιση.
Πηγή: https://github.com/matterport/Mask_RCNN

Από τα παραπάνω γίνεται αντιληπτό ότι το πρόβλημα της αναγνώρισης αντικειμένων είναι ιδιαιτέρως απαιτητικό και περιλαμβάνει αρκετές προκλήσεις. Κάποιες από τις δυσκολίες που θα πρέπει να επιλυθούν ώστε να δημιουργηθεί ένας ισχυρός αλγόριθμος αναγνώρισης έχουν ως ακολούθως.

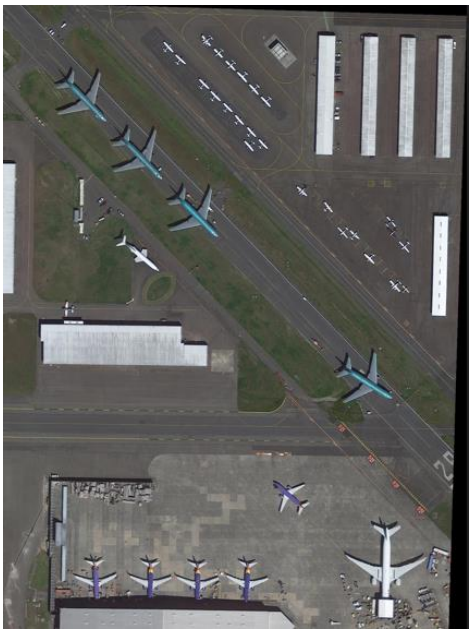
- Πολλά αντικείμενα αναγνώρισης

Τα πάρα πολλά αντικείμενα στην εικόνα την κάνουν εξαιρετικά γεμάτη από άποψη παρεχόμενου σήματος στον αλγόριθμο εκπαίδευσης. Αυτό δημιουργεί αρκετές δυσκολίες στο μοντέλο, όπως μεγάλες αποκρύψεις ή τα αντικείμενα μπορεί να είναι μικρά και η κλίμακα να είναι ασυνεπής.

- Έντονη διακύμανση αντικειμένων εντός της ίδιας κλάσης

Μια άλλη σημαντική πρόκληση για την ανίχνευση αντικειμένων είναι η σωστή ανίχνευση αντικειμένων της ίδιας κατηγορίας, τα οποία μπορεί να έχουν υψηλή

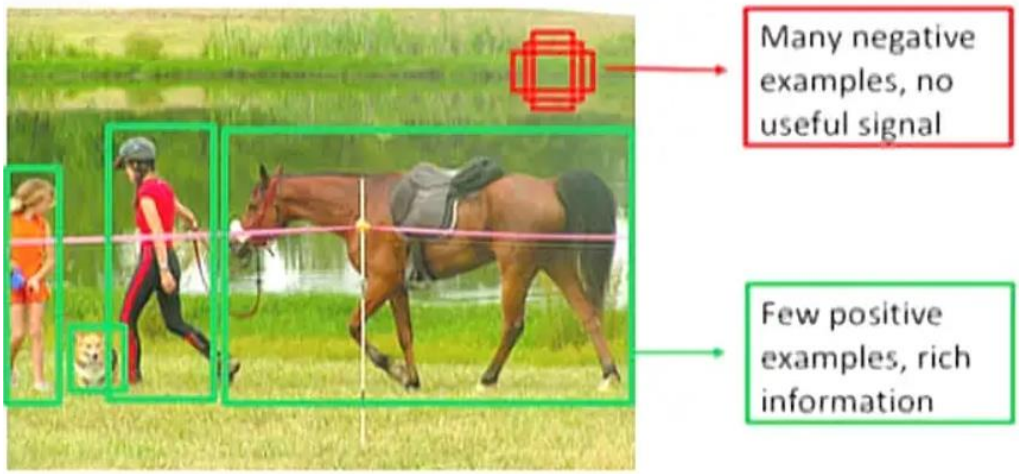
διακύμανση. Για παράδειγμα, στην παρακάτω δορυφορική εικόνα (Εικόνα 2.13) εντοπίζονται διαφορετικά είδη αεροσκαφών με διαφορετικά χαρακτηριστικά όπως μέγεθος, χρώμα, άνοιγμα φτερών, κλπ. Ο εντοπισμός αυτών των αντικειμένων της ίδιας κατηγορίας μπορεί να είναι δύσκολος.



**Εικόνα 2.13**: Δορυφορική Εικόνα με διάφορους τύπους αεροσκαφών.

- Ανισορροπία αντικειμένου-παρασκηνίου

Είναι μια πρόκληση που επηρεάζει σχεδόν όλες τις διαδικασίες αναγνώρισης, είτε πρόκειται για εικόνα, είτε για κείμενα, χρονοσειρές κτλ. Το πρόβλημα καλείται ανισορροπία κλάσης προσκηνίου-παρασκηνίου στην ανίχνευση αντικειμένων. Η ανισορροπία κλάσης μπορεί να δημιουργήσει πρόβλημα στην ταξινόμηση της εικόνας και κατ' επέκταση στην επακόλουθη ανίχνευση των αντικειμένων, όταν περιέχει πολύ λίγα κύρια αντικείμενα, ενώ το υπόλοιπο της εικόνας είναι γεμάτο με φόντο (Εικόνα 2.14). Ως αποτέλεσμα, το μοντέλο θα εξέταζε πολλές περιοχές στην εικόνα όπου οι περισσότερες θα θεωρούνταν αρνητικές. Εξαιτίας αυτών των αρνητικών, το μοντέλο δεν μαθαίνει χρήσιμες πληροφορίες με κίνδυνο να επηρεαστεί ολόκληρη η εκπαίδευση του μοντέλου.

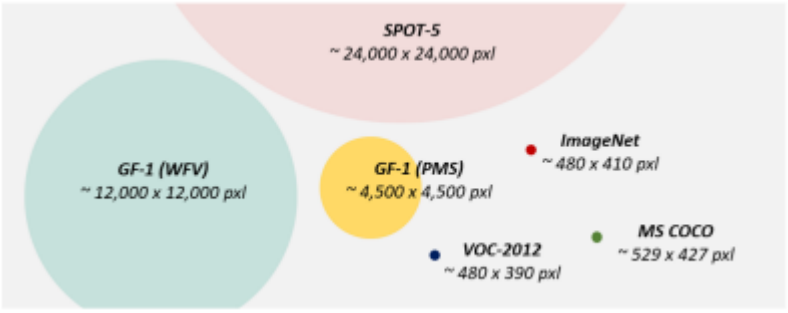



**Εικόνα 2.14:** Πρόβλημα ανισορροπίας αντικειμένου – παρασκηνίου.
Πηγή: https://towardsdatascience.com/review-retinanet-focal-loss-object-detection-38fba6afabe4

Επιπλέον των παραπάνω τα οποία εντοπίζονται γενικά στον τομέα της όρασης υπολογιστών, δυσκολίες-προκλήσεις υπάρχουν και στον εντοπισμό αντικειμένων στον τομέα της τηλεπισκόπησης [22]. Οι προκλήσεις σε αυτό το πεδίο συνοψίζονται ως εξής:

- Ανίχνευση σε "μεγάλα δεδομένα"

Λόγω του τεράστιου όγκου δεδομένων που παρέχουν οι εικόνες τηλεπισκόπησης, η αναγνώριση αντικειμένων με ταχύ και ακριβή τρόπο αποτελεί ιδιαίτερα απαιτητικό πρόβλημα. Στην παρακάτω εικόνα 2.15 παρουσιάζεται σχηματικά οι διαφορές στον παρεχόμενο όγκο δεδομένων μεταξύ μιας δορυφορικής εικόνας εμπορικού παρόχου (SPOT-5) και της μέσης εικόνας που παρέχεται από τις συλλογές δεδομένων ImageNet, VOC και COCO.




**Εικόνα 2.15**: Ανίχνευση σε μεγάλα δεδομένα
Πήγη: https://arxiv.org/pdf/1905.05055.pdf

- Επικαλύψεις

Πάνω από το 50% της επιφάνειας της γης καλύπτεται από σύννεφα κάθε μέρα. Στην παρακάτω εικόνα 2.16 παρουσιάζεται η δυσκολία αναγνώρισης αντικειμένων σε τέτοιες περιπτώσεις.


**Εικόνα 2.16:** Αποκρύψεις αντικειμένων σε ΔΕ.
Source: https://arxiv.org/pdf/1905.05055.pdf

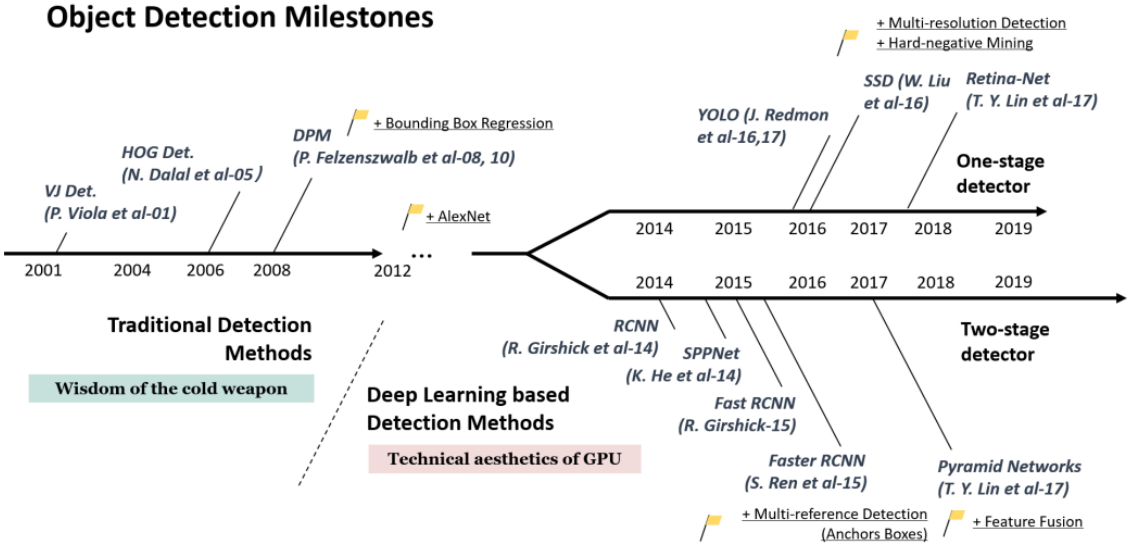
- Ανομοιογένεια δεδομένων

Η πληθώρα ύπαρξης δεκτών για λήψη απεικονίσεων, έχουν δημιουργήσει ένα πλήθος δεδομένων με διαφορετικά χαρακτηριστικά (ραδιομετρική ανάλυση, χωρική ανάλυση, χρόνος λήψης) από διαφορετικούς αισθητήρες. Αποτελεί πρόκληση η ύπαρξη ενός δικτύου που θα επεξεργάζεται εικόνες διαφορετικών χαρακτηριστικών.

### 2.4.2 Ιστορική Αναδρομή

Το πρόβλημα της ανίχνευσης αντικειμένων έχει απασχολήσει την επιστημονική κοινότητα από τη δεκαετία του 90 και θεωρείται αντικείμενο με ιδιαίτερη πρόκληση στον τομέα της όρασης υπολογιστών. Τα τελευταία χρόνια με την ανάπτυξη των τεχνικών Βαθιά Μάθησης αλλά και την υπολογιστική βοήθεια που προσέφερε η ανάπτυξη των GPU της NVIDIA, η πρόοδος στον τομέα είναι εντυπωσιακή από άποψη ακρίβειας, ταχύτητας και αποδοτικότητας. Με την εκμετάλλευση των τεχνικών και τεχνολογιών αυτών σχεδόν κάθε έτος επιτυγχάνεται ένα νέο πρότυπο αναφοράς τελευταίας τεχνολογίας.

Επιχειρώντας μια σύντομη ιστορική αναδρομή, μπορούν να διακριθούν δύο διαφορετικές εποχές στην επιστήμη της όρασης υπολογιστών. Η πρώτη εντοπίζεται πριν το 2010 που χαρακτηρίζεται από παραδοσιακές τεχνικές και προσεγγίσεις. Η δεύτερη έχει την απαρχή της το 2012, χρονιά ορόσημο για την επιστήμη, όπου πρωτοπαρουσιάστηκε το AlexNet στον διαγωνισμό ImageNet Visual Recognition.




**Εικόνα 2.17**: Χάρτης εξέλιξης τεχνικών Αναγνώρισης Αντικειμένων
Πηγή: https://arxiv.org/pdf/1905.05055.pdf

Στην παραπάνω εικόνα 2.17 απεικονίζεται ο χάρτης εξέλιξης των τεχνικών από το 2001 εώς το σήμερα. Αν και αναφορές και βιβλιογραφία για το αντικείμενο μπορούν να βρεθούν και προγενέστερα του παραπάνω χρονικού χάρτη, η πρώτη σοβαρή προσπάθεια έγινε το 2001 όταν οι P. Viola και M. Jones [23, 24] ανέπτυξαν αλγόριθμο ο οποίος αναγνώριζε σε πραγματικό χρόνο ανθρώπινα πρόσωπα. Ο αλγόριθμος, ο οποίος προς τιμήν τους ονομάστηκε «Viola-Jones detector», πέτυχε απόδοση 10 φορές ταχύτερα από τους συνήθεις της εποχής. Η υλοποιούμενη μεθοδολογία έφερε στο προσκήνιο την έννοια του κινούμενου παραθύρου (sliding window), διαδικασία κατά την οποία ένα κινούμενο παράθυρο διατρέχει όλες τις πιθανές θέσεις και πιθανές κλίμακες εντός των διαστάσεων μιας εικόνας, ώστε να εντοπίσει πιθανές θέσεις στις οποίες αποτυπώνονται ανθρώπινα πρόσωπα. Παρόλο που η σύλληψη με τα σημερινά δεδομένα φαίνεται απλή, ξεκάθαρη και υλοποιήσιμη, για τις τότε υπολογιστικές δυνατότητες ήταν εκτός εποχής.

Οι επόμενες μέθοδοι που ακολούθησαν όπως το ιστόγραμμα των προσανατολισμένων κλίσεων (Histogram of Oriented Gradients - HOG) [25] και Deformable Parts Model (DPM) [26] βασίζονται στην εξαγωγή χαρακτηριστικών από την εικόνα όπως ακμές, γωνίες και κλίσεις από τις τιμές ραδιομετρίας των pixel της εικόνας. Βασική λογική της πρώτης μεθόδου είναι ότι το σχήμα ενός

αντικειμένου μπορεί να περιγραφεί από την κατανομή των εντάσεων των κλίσεων που εμφανίζονται στο αντικείμενο. Η δεύτερη μέθοδος αρχικώς αποτέλεσε εξέλιξη της πρώτης και ακολούθως αναπτύχθηκαν διάφορες εξελίξεις και παραλλαγές. Η μέθοδος ακολουθεί την λογική «διαίρει και βασίλευε» όπου η εκπαίδευση μπορεί να θεωρηθεί ως εκμάθηση ενός σωστού τρόπου αποσύνθεσης ενός αντικειμένου, και ο εντοπισμός μπορεί να θεωρηθεί ως ένα σύνολο ανιχνεύσεων σε διαφορετικά μέρη του αντικειμένου. Για παράδειγμα, το πρόβλημα της ανίχνευσης ενός «αυτοκίνητου» μπορεί να θεωρηθεί ως η ανίχνευση του παραθύρου, του σώματος και των τροχών του.

Αν και οι σημερινοί ανιχνευτές έχουν ξεπεράσει κατά πολύ τις δύο προηγούμενες μεθόδους όσον αφορά την αποδοτικότητα και ακρίβεια ανίχνευσης, πολλές από τις τεχνικές που εισήγαγαν εξακολουθούν να επηρεάζουν τις σύγχρονες μεθόδους.

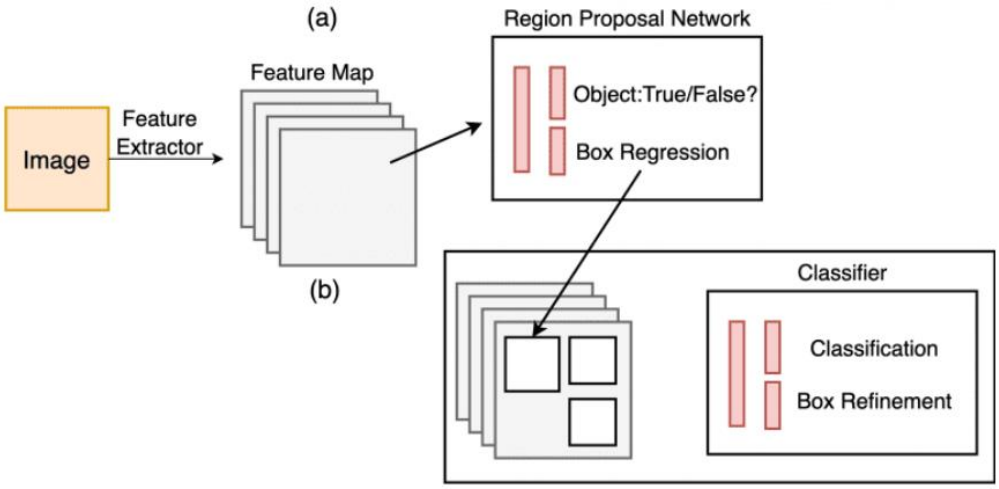
Η παρουσίαση του AlexNet το 2012 δικαίως θεωρείται ως το τέλος των παραδοσιακών τεχνικών και η αρχή της εποχής των τεχνικών βαθιάς μάθησης στην επιστήμη της όρασης υπολογιστών. Η μέθοδος χρησιμοποίησε συνελικτικά νευρωνικά δίκτυα για να ταξινομήσει της 1000 εικόνες του ImageNet και πέτυχε σημαντική απόδοση στον διαγωνισμό ImageNet LSVRC-2012 με ακρίβεια 84.7% ποσοστό πολύ καλύτερο από οτιδήποτε είχε δοκιμαστεί μέχρι στιγμής. Από την καινοτομία αυτή ήταν θέμα χρόνου η τεχνική να αξιοποιηθεί περαιτέρω και για σκοπούς αναγνώρισης αντικειμένων.

Το 2014 οι Girshick et al. [27] πρότειναν έναν τρόπο να χρησιμοποιηθούν συνελικτικά χαρακτηριστικά για αναγνώριση αντικειμένων, εισάγοντας την μέθοδο R-CNN. Από τότε και μεταγενέστερα η εξέλιξη στον τομέα ήταν ραγδαία. Όπως απεικονίζεται και στην παραπάνω εικόνα 2.17 οι τεχνικές βαθιάς μάθησης εξελίχθηκαν σε δύο κύριες κατηγορίες. Η πρώτη κατηγορία λέγεται *αλγόριθμοι αναγνώρισης δύο σταδίων* (two-stage detection algorithms) και περιλαμβάνει ταξινομητές όπως RCNN, Fast-RCNN, Faster-RCNN, Mask-RCNN. Η δεύτερη λέγεται *αλγόριθμοι ενός σταδίου (one-stage detectors)* και περιλαμβάνει ταξινομητές όπως SSD, Yolo, EfficientDet κτλ.

### 2.4.3 Two-stage detectors - RCNN

Σε αντίθεση με την μεθοδολογία που ακολουθείται στις προσεγγίσεις single Stage, στις two-stage η διαδικασία εφαρμόζει γενικά δύο βήματα. Όπως παρατηρείται στην παρακάτω εικόνα 2.18, από την αρχική εικόνα εξάγεται ένας χάρτης χαρακτηριστικών, ο οποίος τροφοδοτείται σε ένα δίκτυο που λέγεται Region Proposal Network. Δουλεία του δικτύου αυτού είναι να προτείνει πιθανές περιοχές στις οποίες εντοπίζονται αντικείμενα. Ακολούθως, σε δεύτερο διακριτό βήμα οι εξαγόμενες πιθανές περιοχές διοχετεύονται σε έναν ταξινομητή για να βρεθούν τα πιθανά αντικείμενα. Γνωστοί αλγόριθμοι που ακολουθούν την λογική αυτή είναι η οικογένεια αλγορίθμων RCNN (Region Based CNN). Οι two-stage μέθοδοι γενικά θεωρούνται ότι παρέχουν καλύτερα αποτελέσματα από τις single-

stage, θυσιάζοντας όμως τον παράγοντα χρόνο στην εκπαίδευση και στην αναγνώριση των αντικειμένων.




**Εικόνα 2.18:** Two-Stage detectors
Πηγή: pyimagesearch.com/2022/04/04/introduction-to-the-yolo-family/

Η μέθοδος RCNN (Region Based Neural Network) [27] αποτέλεσε την πρώτη προσπάθεια αξιοποίησης των νευρωνικών δικτύων για αναγνώριση αντικειμένων. Η υλοποίηση της μεθόδου γίνεται σε δύο διαφορετικά βήματα ως εξής:
Αρχικά ένας αλγόριθμος επιλεκτικής αναζήτησης (selective research) προτείνει περιοχές της εικόνας που δυνατόν να υπάρχει κάποιο αντικείμενο, ορίζοντας με αυτό τον τρόπο περιοχές ενδιαφέροντος. Στη συνέχεια κάθε περιοχή ενδιαφέροντος τροφοδοτείται σε ένα συνελικτικό νευρωνικό δίκτυο που έχει εκπαιδευθεί στο ImageNet για την εξαγωγή χαρακτηριστικών. Τέλος, ταξινομητές SVM χρησιμοποιούνται για την πρόβλεψη της παρουσίας ενός αντικειμένου σε κάθε περιοχή και την αναγνώριση του αντικειμένου. Παρόλο που η μέθοδος είχε επιτυχία, το κύριο της μειονέκτημα είναι το γεγονός της εύρεσης πολλαπλών επικαλυπτόμενων προτεινόμενων περιοχών οι οποίες διοχετεύονται στο δίκτυο. Ως αποτέλεσμα δαπανάται υπολογιστική ισχύ για πλεονάζουσα πληροφορία και ο χρόνος ανίχνευσης να είναι αρκετά μεγάλος. Ενδεικτικά σε κάθε εικόνα προτείνονται περίπου 2000 περιοχές.

Το 2015 προτάθηκε η μέθοδος Fast RCNN [28] η οποία αποτελεί μια βελτίωση της RCNN. Για να αντιμετωπιστεί το πρόβλημα των πολλαπλών επικαλυπτόμενων τετραγώνων, βελτιώθηκε η διαδικασία αντικαθιστώντας τις προτεινόμενες περιοχές με έναν χάρτη χαρακτηριστικών ο οποίος χρησιμοποιείται για την διαδικασία της αναγνώρισης. Η μέθοδος αύξησε την ακρίβεια από το 58.8% (RCNN) σε 70% ενώ η ταχύτητα εντοπισμού αυξήθηκε κατά περίπου 200%.

Το 2015 προτάθηκε επίσης η μέθοδος Faster RCNN [29] λίγο μετά την δημοσίευση της προηγούμενης μεθόδου. Καινοτομία αποτελεί η εισαγωγή της έννοιας Regional Proposed Network (RPN) η οποία χρησιμοποιείται για την

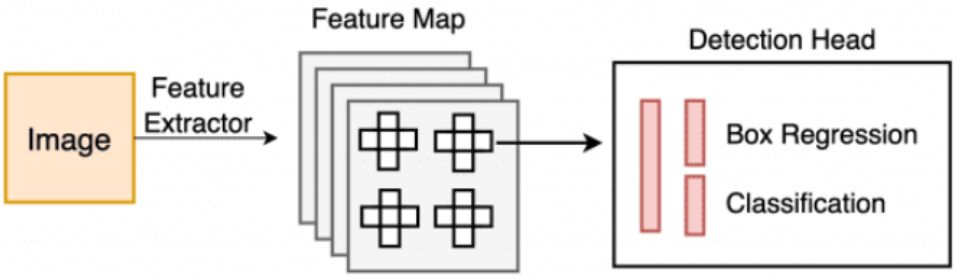
εξαγωγή των προτεινόμενων περιοχών σχεδόν με μηδενικό κόστος. Οι πιθανές αυτές περιοχές δεν δημιουργούνται ξεχωριστά από ένα διαφορετικό μοντέλο, αλλά έχουν ενσωματωθεί στην διαδικασία του νευρωνικού δικτύου.

Η τέταρτη μέθοδος στην εξέλιξη της οικογένειας των ταξινομητών RCNN μεθόδων αποτελεί η μέθοδος Mask R-CNN [30]. Οι προγενέστερες τεχνικές αναγνωρίζουν τα διαφορετικά αντικείμενα στις διάφορες θέσεις της εικόνας και καταδεικνύουν με ένα χρωματισμένο περιγεγραμμένο τετράγωνο. Στη νέα αυτή μέθοδο ο ταξινομητής καταδεικνύει το αντικείμενο και τα συγκεκριμένα pixel που απαρτίζουν το αντικείμενο. Η νέα τεχνική ονομάστηκε Mask R-CNN διότι τα αντικείμενα αναδεικνύονται με τη μορφή μάσκας σχηματίζοντας το σχήμα του αντικειμένου με ένα διαφορετικό διάφανο χρωματισμό.

### 2.4.4 Single-stage detectors – You Only Look Once

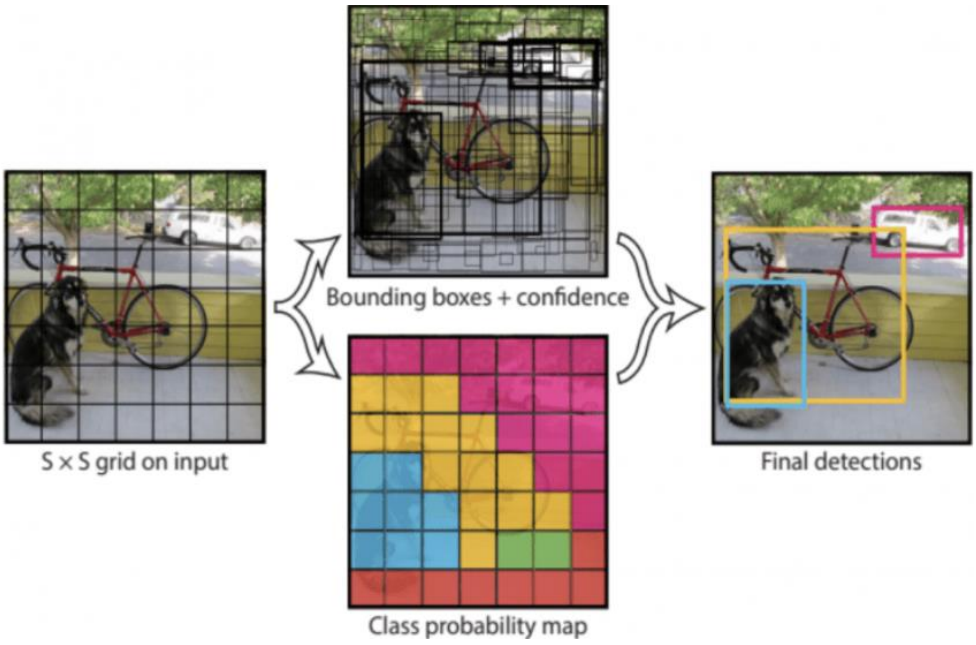
Η μέθοδος Yolo (You only Look Once) προτάθηκε από τον R. Joseph et al. το 2015 [31] και αποτέλεσε σημείο αναφοράς στην εξέλιξη των τεχνικών καθώς ήταν η πρώτη single-stage μέθοδος που υλοποιήθηκε. Με τη νέα αυτή τεχνική εγκαταλείφθηκε η παλαιότερη φιλοσοφία των two-stage τεχνικών, οι οποίες ακολουθούσαν τον γενικό κανόνα εξαγωγής προτεινόμενων περιοχών και ακολούθως ανίχνευση. Η νέα αυτή μέθοδος αντιμετώπισε το πρόβλημα ως πρόβλημα παλινδρόμησης.

Η μέθοδος είναι εξαιρετικά γρήγορη με μεγάλο ποσοστό ακρίβειας, διότι δεν απαιτεί την ύπαρξη δύο ξεχωριστών δικτύων στην αρχιτεκτονική της σε αντίθεση με τις two-stage μεθόδους. Όπως παρατηρείται στην παρακάτω εικόνα 2.19 χρησιμοποιείται ένα δίκτυο σε ολόκληρη την εικόνα και προβλέπει αντικείμενα και πιθανότητες για κάθε αντικείμενο ταυτόχρονα. Το δίκτυο είναι κατασκευασμένο να εκπαιδεύεται με αντίστοιχο τρόπο που γίνεται η ταξινόμηση μιας εικόνας, με αποτέλεσμα η ταχύτητα απόδοσης να είναι εξαιρετικά μεγάλη. Το πρώτο μοντέλο που δημιουργήθηκε πέτυχε ταχύτητα 45 fps (frames per seconds) σε ένα μηχάνημα Titan X GPU.



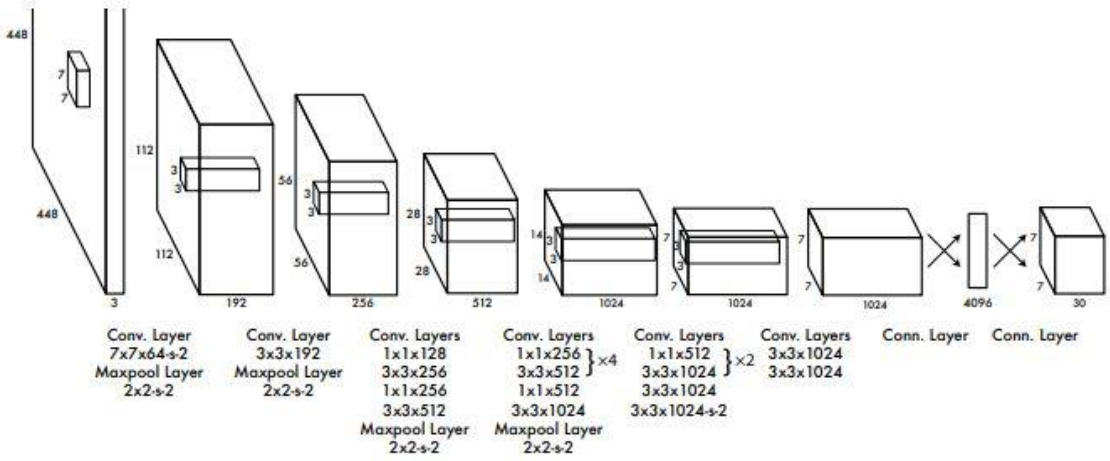

***Εικόνα 2.19:*** Single Stage Detectors
Πηγή: pyimagesearch.com/2022/04/04/introduction-to-the-yolo-family/

Όπως παρατηρείται στην παρακάτω εικόνα 2.20 η μέθοδος αρχικά διαιρεί την εικόνα σε έναν κάνναβο 7x7 κελιών. Το κελί που περιέχει το κεντροειδές του αντικειμένου είναι υπεύθυνο να αναγνωρίσει το αντικείμενο. Κάθε κελί ανιχνεύει συνολικά Β πιθανά τετράγωνα αντικειμένων καθώς και C πιθανότητες κλάσεων. Μια βαθμολογία τοποθετείται σε κάθε τετράγωνο που δείχνει τη βεβαιό του μοντέλου ότι το περιγεγραμμένο τετράγωνο περιέχει αντικείμενο.




**Εικόνα 2.20:** Διαδικασία εντοπισμού Yolo.
https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/Redmon_You_Only_Look_CVPR_2016_paper.pdf

Η αρχιτεκτονική του δικτύου (Εικόνα 2.21) αποτελείται από 24 διαδοχικά συνελικτικά νευρωνικά δίκτυα με ενδιάμεσα Max pooling επίπεδα τα οποία καταλήγουν σε δύο Fully-connected επίπεδα. Η μείωση των διαστάσεων των βαθύτερων συνελικτικών επιπέδων γίνεται χρησιμοποιώντας φίλτρα 1x1.



**Εικόνα 2.21 :** Αρχιτεκτονική Yolo.
Πηγή: You Only Look Once: Unified, Real-Time Object Detection [31]

Η μέθοδος πέτυχε μετρητική mAP (mean Average Precision [2]) 63.4% στο dataset VOC07 και 57.9% στο dataset VOC12. Μετά την πρώτη αυτή προσέγγιση προτάθηκαν νέες βελτιώσεις οι οποίες περαιτέρω βελτιώνουν την ακρίβεια ανίχνευσης διατηρώντας παράλληλα πολύ υψηλή ταχύτητα. Παρά την εντυπωσιακή της απόδοση σε ταχύτητα και ακρίβεια πρόβλεψης, η YoLo είναι λιγότερο ακριβής από τις προγενέστερες μεθόδους two-stage, ειδικά στην αναγνώριση αντικειμένων με ακανόνιστο σχήμα ή πολύ μικρών αντικειμένων. Οι επόμενες βελτιώσεις της μεθόδου δίνουν μεγαλύτερη έμφαση στην επίλυση αυτού του προβλήματος.

### 2.4.5 YoLov2 – YoLo9000

Το 2017 οι Joseph Redmon και Ali Farhadi πρότειναν μια εξελιγμένη μορφή της μεθόδου την οποία ονόμασαν *Yolo9000: Better, Faster, Stronger* [32]. Οι συγγραφείς πρότειναν δύο εκδοχές της μεθόδου, της οποίες ονόμασαν YoLov2 και YoLo9000, με μικρές διαφορές μεταξύ τους κυρίως στη διαδικασία εκπαίδευσης.

Το βελτιωμένο μοντέλο YOLOv2 χρησιμοποίησε διάφορες νέες τεχνικές για να ξεπεράσει τις sigle-stage μεθόδους, τόσο σε ταχύτητα όσο και σε ακρίβεια, αλλά και να διορθώσει τις αστοχίες που παρατηρήθηκαν με την προηγούμενη έκδοση YoLov1. Νέες τεχνικές οι οποίες, μεταξύ άλλων, εισήχθησαν με την μέθοδο έχουν ως ακολούθως:

- Batch κανονικοποίηση
- Ταξινόμηση υψηλής ανάλυσης
- Χρήση αγκύρων
- Εκπαίδευση σε πολλαπλές κλίμακες
- Χρήση δικτύου DarkNet
- κα.

Η batch κανονικοποίηση βοήθησε στη βελτίωση της σύγκλισης της διαδικασίας εκπαίδευσης του δικτύου και εξάλειψε την ανάγκη για άλλες τεχνικές τακτοποίησης (πχ dropout) χωρίς το δίκτυο να υπερπροσαρμόζεται (overfitting) στα δεδομένα εκπαίδευσης. Προσθέτοντας ένα επίπεδο κανονικοποίησης σε όλα τα συνελικτικά επίπεδα βελτιώθηκε ο δείκτης mAP κατά 2%.

Στην Yolov1 πριν από την εκπαίδευση του δικτύου λάμβανε χώρα μια διαδικασία ταξινόμησης των εικόνων στο σετ δεδομένων Imagenet, με μέγεθος εισόδου 224x224 και ακολούθως οι εικόνες γινόντουσαν upscale για την διαδικασία του

[2] Βλ. παρ. 3.3.3

object detection. Στην YoLov2 οι συγγραφείς διατήρησαν την αρχική ταξινόμηση των 224x224, ακολούθως επαναλαμβάνουν την ταξινόμηση με μέγεθος εισόδου εικόνων 448x448 για 10 εποχές εκπαίδευσης και τέλος γίνεται η διαδικασία του object detection. Με αυτό τον τρόπο το δίκτυο κέρδισε χρόνο, προσάρμοσε τα βάρη του στις upscale εικόνες οι οποίες χρησιμοποιούνται για την τελική διαδικασία αναγνώρισης αντικειμένων. Με τον τρόπο αυτό η μετρητική mAP αυξήθηκε κατά 4%.

Εμπνευσμένοι από τον FastRCNN, η μέθοδος εισάγει την έννοια του anchor (άγκυρα). Τα τελευταία πλήρως συνδεδεμένα επίπεδα καταργούνται και στη θέση τους τοποθετούνται τα anchor boxes για την πρόβλεψη των τετραγώνων. Τα anchor boxes είναι πολλαπλά τετράγωνα με διαφορετικά μεγέθη τοποθετημένα σε κάθε pixel, σε όλα τα δυνατά επίπεδα κλίμακας. Κατά τη διαδικασία της εκπαίδευσης η σύνδεση μεταξύ του κάθε Anchor και των Ground Truth τετραγώνων γίνεται με το δείκτη IoU με τιμή επικάλυψης μεγαλύτερη από 0.5.

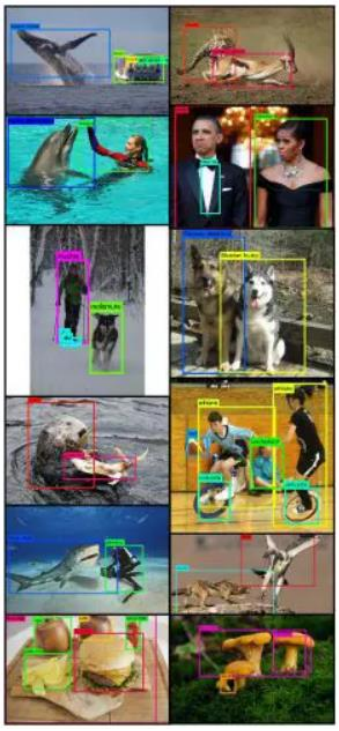
Μια άλλη τεχνική που χρησιμοποιήθηκε ήταν η εκπαίδευση πολλαπλών κλιμάκων που επέτρεπε στο δίκτυο να προβλέψει σε διάφορα μεγέθη της εικόνας, επιτρέποντας έτσι μια ισορροπία μεταξύ ταχύτητας και ακρίβειας.

Η μέθοδος χρησιμοποίησε μια αρχιτεκτονική ταξινόμησης των εικόνων που ονομάστηκε Darknet-19 ως βάση για τον εντοπισμό των αντικειμένων. Η αρχιτεκτονική αυτή εμπνεύστηκε από προγενέστερη δουλειά και ομοιάζει με την VGG-16. Όλο το δίκτυο αποτελείται από 19 συνελικτικά επίπεδα και 5 max-pooling επίπεδα.

Η YoLov2 εκπαιδεύτηκε σε σύνολα δεδομένων ανίχνευσης όπως το Pascal VOC και το MS COCO. Η YoLo9000 σχεδιάστηκε για να προβλέπει περισσότερες από 9000 διαφορετικές κατηγορίες αντικειμένων εκπαιδεύοντάς το δίκτυο του στα σύνολα δεδομένων MS COCO και ImageNet.

Σε ανάλυση εισόδου 416×416 pixels, η YOLOv2 πέτυχε 76,8 mAP στο σύνολο δεδομένων VOC 2007 και 67 FPS σε μηχάνημα Titan X GPU. Στο ίδιο σύνολο δεδομένων με είσοδο 544×544 pixels, η YOLOv2 πέτυχε 78,6 mAP και 40 FPS.

Στην παρακάτω εικόνα 2.22 απεικονίζονται κάποιες από τις 9000 κλάσεις που είναι εκπαιδευμένο το δίκτυο να αναγνωρίζει, σε παραδείγματα πραγματικού κόσμου.



Εικόνα 2.22 : Παραδείγματα ανιχνεύσεων Yolo9000 σε πραγματικό χρόνο.
Πηγή:
https://openaccess.thecvf.com/content_cvpr_2017/papers/Redmon_YOLO9000_Better_Faster_CVPR_2017_paper.pdf

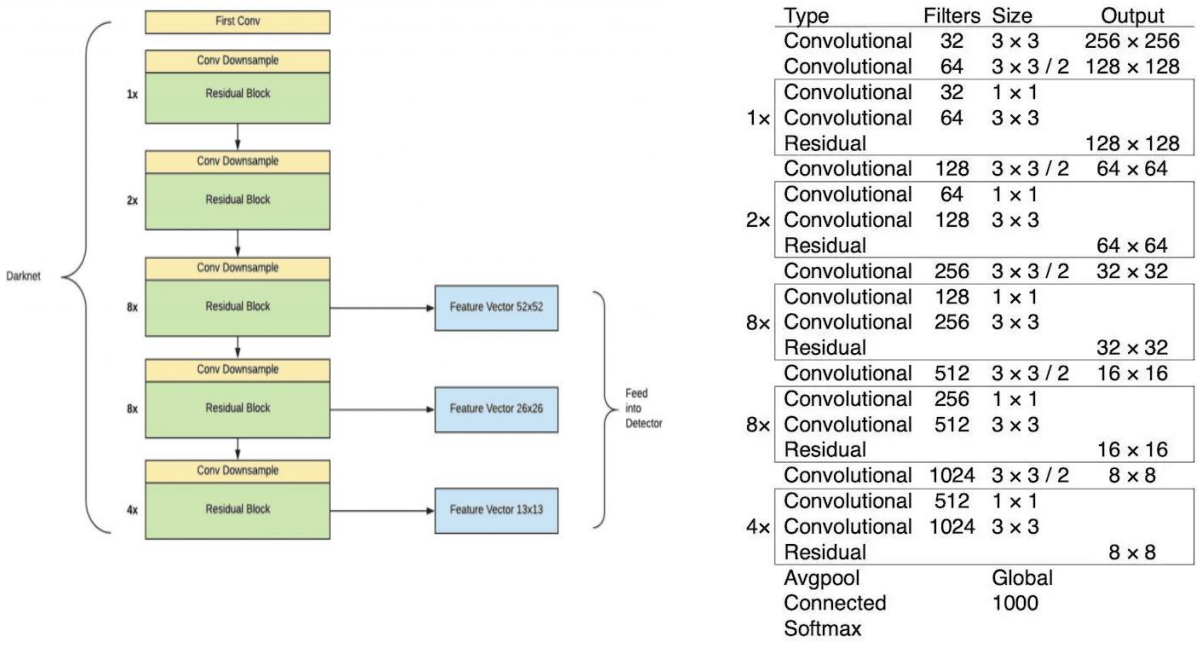
### 2.4.6 YoLov3

Το 2018 οι Joseph Redmon και Ali Farhadi δημοσίευσαν τη μέθοδο *Yolov3: An Incremental Improvement* [33]. Οι συγγραφείς πραγματοποίησαν αλλαγές στην αρχιτεκτονική του δικτύου βασιζόμενοι στις τεχνικές που δημιουργήθηκαν κατά τις προηγούμενες μεθόδους. Η νέα μέθοδος εισήγαγε μια νέα αρχιτεκτονική την οποία ονόμασαν Darknet-53, ως βάση για την αναγνώριση αντικειμένων και την εξαγωγή χαρακτηριστικών. Το δίκτυο Darknet-53 είναι πιο βαθύ από το προηγούμενο Darknet-19, πιο γρήγορο και με καλύτερη ακρίβεια.

Το δίκτυο αποτελείται από 53 συνελικτικά επίπεδα, το οποία έχουν προεκπαιδευθεί στην ταξινόμηση εικόνας χρησιμοποιώντας το σετ δεδομένων ImageNet. Για την αναγνώριση αντικειμένων προστίθενται επιπλέον 53 επίπεδα πριν το βασικό δίκτυο, ανεβάζοντας τον συνολικό αριθμό των επιπέδων σε 106, αριθμό συγκριτικά πολύ μεγαλύτερο από το προηγούμενο Darknet-19 που είχε 19 επίπεδα.

Η αρχιτεκτονική του δικτύου παρουσιάζεται στην παρακάτω εικόνα 2.23. Το κάθε συνελικτικό επίπεδο ακολουθείται από ένα Residual Block επίπεδο με διάφορα μεγέθη residuals blocks (1x, 2x, 4x, 8x). Για να μειωθεί η χωρική διάσταση των χαρτών χαρακτηριστικών (δηλαδή των επιπέδων που προκύπτουν μετά από κάθε συνελικτικό επίπεδο), χρησιμοποιείται πίνακας συνέλιξης με βήμα μετακίνησης του φίλτρου (stride) 2 πριν από κάθε Residual Block. Ο αριθμός των φίλτρων

ξεκινά με 32 και διπλασιάζεται σε κάθε επίπεδο και έως τα 1024 φίλτρα. Το κάθε residual block ξεκινάει από ένα φίλτρο 1x1, ακολουθούμενο από φίλτρο 3x3 και τέλος από μια σύνδεση παράκαμψης. Τέλος, για να πραγματοποιηθεί η ταξινόμηση των αντικειμένων στα τελικά επίπεδα, έχουν τοποθετηθεί πλήρως συνδεδεμένα επίπεδα και ένας ταξινομητής softmax για την εξαγωγή των πιθανοτήτων σε μια από τις 100 κλάσεις του δικτύου.




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

**Εικόνα 2.23:** Αρχιτεκτονική δικτύου YoLov3.
Πηγή: towardsdatascience.com/dive-really-deep-into-yolo-v3-a-beginners-guide

Κατά την αναγνώριση των αντικειμένων, τα στρώματα μετά την τελευταία ομάδα αφαιρούνται, και έτσι προκύπτει ο σκελετός του ανιχνευτή. Το δίκτυο ανιχνεύει αντικείμενα σε πολλαπλές κλίμακες διότι σε καθεμία από τις τρεις τελευταίες υπολειπόμενες ομάδες, προσαρτάται ένα στρώμα ανίχνευσης για να κάνει προβλέψεις αντίστοιχων αντικειμένων, όπως φαίνεται στο παραπάνω σχήμα 2.23 (αριστερά). Ενώ στις προηγούμενες εκδόσεις η αναγνώριση γινόταν μόνο στο τελευταίο επίπεδο, στην παρούσα έκδοση τα αντικείμενα ανιχνεύονται σε τρία διαφορετικά στάδια του δικτύου. Το δίκτυο εξάγει χαρακτηριστικά και στα τρία αυτά στάδια χρησιμοποιώντας μια λειτουργεία αντίστοιχη με το FPN (Feature pyramid network).

| | Backbone | AP | AP50 | AP75 | APs | APm | APl |
|---|---|---|---|---|---|---|---|
| *Two-Stage Methods* | | | | | | | |
| Faster-RCNN +++ | ResNet-101-C4 | 34.9 | 55.7 | 37.4 | 15.6 | 38.7 | 50.9 |
| Faster-RCNN with FPN | ResNet-101-FPN | 36.2 | 59.1 | 39 | 18.2 | 39 | 48.2 |
| Faster-RCNN by G-RMI | Inception-ResNet-v2 | 34.7 | 55.5 | 36.7 | 13.5 | 38.1 | 52 |
| Faster-RCNN with TDM | Inception-ResNet-v2-TDM | 36.8 | 57.7 | 39.2 | 16.2 | 39.8 | **52.1** |
| *Single-Stage Methods* | | | | | | | |
| YOLOv2 | DarkNet-19 | 21.6 | 44 | 19.2 | 5 | 22.4 | 35.5 |
| SSD513 | ResNet-101-SSD | 31.2 | 50.4 | 33.3 | 10.2 | 34.5 | 49.8 |
| DSSD513 | ResNet-101-DSSD | 33.2 | 53.3 | 35.2 | 13 | 35.4 | 51.1 |
| RetinaNet | ResNet-101-FPN | 39.1 | 59.1 | 42.3 | 21.8 | 42.7 | 50.2 |
| RetinaNet | ResNeXt-101-FPN | **40.8** | **61.1** | **44.1** | **24.1** | **44.2** | 51.2 |
| YOLOv3 - 608 x 608 | Darknet-53 | 33 | 57.9 | 34.4 | 18.3 | 35.4 | 41.9 |

**Εικόνα 2.24:** Συγκριτικός πίνακας μεταξύ μεθόδων Yolov3 και λοιπών ανιχνευτών.
Πηγή: https://arxiv.org/pdf/1804.02767.pdf

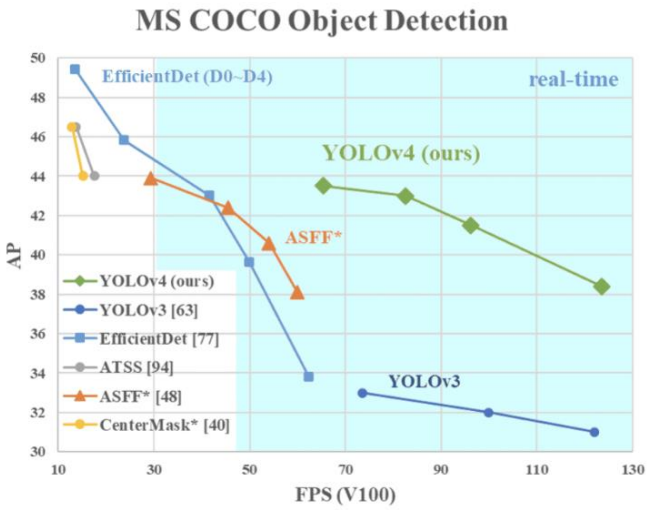
Από τον παραπάνω πίνακα 2.24 παρατηρείται ότι η μέθοδος είναι αρκετά μακριά από άλλα μοντέλα όπως το RetinaNet. Ωστόσο στα αποτελέσματα AP50 (Average Precision με δείκτη IoU 0.5), το YoLov3 έχει αρκετά ικανοποιητικά αποτελέσματα και αποδίδει στο ίδιο επίπεδο με όλους σχεδόν τους ανιχνευτές. Καθώς αυξάνεται το όριο IoU στο 0,75, η απόδοση πέφτει σημαντικά, υποδεικνύοντας ότι το YOLOv3 δυσκολεύεται να ευθυγραμμίσει ικανοποιητικά το περιγεγραμμένο τετράγωνο με το αντικείμενο. Οι προηγούμενες εκδόσεις του YoLo δυσκολεύονταν να ανιχνεύσουν μικρά αντικείμενα, αλλά το YOLOv3 με την προσέγγιση πολλαπλής κλίμακας αποδίδει σχετικά καλά. Ωστόσο, έχει συγκριτικά χειρότερη απόδοση σε αντικείμενα μεσαίου και μεγαλύτερου μεγέθους. Τέλος συμπεραίνεται ότι η μέθοδος αποδίδει καλύτερα στην μετρητική AP50 και είναι ο ταχύτερος ανιχνευτής μεταξύ όλων των προγενέστερων.

### 2.4.7 YoLov4

Το 2020 οι Bochkovskiy et al. δημοσίευσαν την μέθοδο *YOLOv4: Optimal Speed and Accuracy of Object Detection* [34]. Οι συγγραφείς χρησιμοποίησαν με τον καλύτερο δυνατό τρόπο υπάρχουσες νέες τεχνικές και χαρακτηριστικά συνδυάζοντάς με τέτοιο τρόπο που οδήγησε σε έναν ανιχνευτή που επιτυγχάνει βέλτιστη ταχύτητα και ακρίβεια, ξεπερνώντας πολλούς άλλους προηγμένους ανιχνευτές.

Είναι γενικά γνωστό ότι τα πιο ακριβή μοντέλα νευρωνικών δικτύων δεν λειτουργούν ικανοποιητικά σε πραγματικό χρόνο και απαιτούν πολλές GPU για εκπαίδευση. Με το YOLOv4, οι συγγραφείς προσπαθούν να αντιμετωπίσουν αυτό το ζήτημα δημιουργώντας ένα CNN που λειτουργεί σε πραγματικό χρόνο, σε μια συμβατική GPU και για το οποίο η εκπαίδευση απαιτεί μόνο μία παραδοσιακή GPU. Έτσι, οποιοσδήποτε διαθέτει μία συμβατική GPU μπορεί να εκπαιδεύσει αυτόν τον εξαιρετικά γρήγορο και ακριβή ανιχνευτή. Για να επιβεβαιώσουν το

μοντέλο τους, οι συγγραφείς δοκίμασαν το YOLOv4 σε διάφορες αρχιτεκτονικές GPU. Παρατηρείται από το παρακάτω διάγραμμα της εικόνας 2.25 ότι η μέθοδος έχει διπλάσια ταχύτητα από το δίκτυο EfficientDet, ενώ βελτιώνει ακόμα την ακρίβεια της προγενέστερης μεθόδου YoLov3 στο σετ δεδομένων COCO.




**Εικόνα 2.25:** Σύγκριση του προτεινόμενου YOLOv4 με άλλων σύγχρονων ανιχνευτών.
Πηγή: https://arxiv.org/pdf/2004.10934v1.pdf

Ο κύριος στόχος των δημιουργών της μεθόδου είναι η κατασκευή ενός νευρωνικού δικτύου που θα λειτουργεί γρήγορα και θα μπορεί να χρησιμοποιηθεί σε συστήματα παραγωγής. Για να αναπτυχθεί ένα τέτοιο μοντέλο χαμηλού κόστους, θα πρέπει να βρεθεί η βέλτιστη ισορροπία μεταξύ των κάτωθι παραμέτρων:

- Ανάλυση δεδομένων εισόδου
- Αριθμός συνελικτικών επιπέδων του δικτύου
- Συνολικός αριθμός παραμέτρων στο δίκτυο
- Αριθμός επιπέδων εξόδου (αριθμός χρησιμοποιούμενων φίλτρων)

Λαμβάνοντας υπόψη τα παραπάνω οι δημιουργοί ανακάλυψαν ότι ένα μοντέλο που είναι βέλτιστο για ταξινόμηση δεν είναι πάντα βέλτιστο για έναν ταξινομητή. Σε αντίθεση με τον ταξινομητή, ο ανιχνευτής χρειάζεται τα κάτωθι:

- Μεγαλύτερη ανάλυση δεδομένων εισόδου – για ανίχνευση αντικειμένων μικρού μεγέθους.
- Περισσότερα επίπεδα
- Περισσότερες παραμέτρους – για μεγαλύτερη δυνατότητα του μοντέλου να αναγνωρίζει πολλαπλά αντικείμενα διαφορετικών μεγεθών σε μια εικόνα.

Μελετώντας τις συνιστώσες που χρησιμοποιούνται γενικά στον τομέα, οι δημιουργοί χρησιμοποίησαν την κάτωθι αρχιτεκτονική για την ανάπτυξη της μεθόδου YoLov4.

Ως backbone: CSPDarknet53

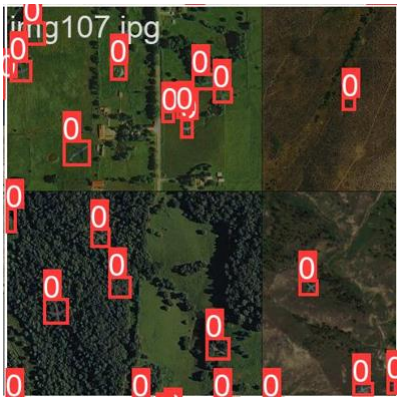
Neck: SPP, PAN

Head: YoLoV3

Οι τρεις αυτές συνιστώσες αποτελούν γενικά όλα τα μέρη της αρχιτεκτονικής ενός ανιχνευτή. Το backbone (πχ VGG, ResNet, DenseNet, and Darknet-19/53) χρησιμοποιείται για την εξαγωγή χαρακτηριστικών από την προεκπαίδευση του δικτύου σε κάποιο σετ δεδομένων (συνήθως στο ImageNet). Head χρησιμοποιείται για την εξαγωγή περιεγραμμένων τετραγώνων και των labels των κλάσεων. Συνήθως διακρίνεται σε δύο κατηγορίες (Single Stage – Two Stages) όπως έχει περιγραφεί παραπάνω. Τα τελευταία χρόνια σύγχρονοι ανιχνευτές χρησιμοποιούν επίσης Neck, που είναι η πρόσθεση επιπλέον επιπέδων μεταξύ backbone και Head, με σκοπό την συλλογή χαρακτηριστικών εικόνας από διαφορετικά στάδια του backbone.

Εκτός από τη μελέτη των υφιστάμενων λειτουργιών που ήδη χρησιμοποιούνται, η μέθοδος έφερε και κάποιες καινοτομίες επίσης. Μια από αυτές είναι η χρήση διαφορετικού τρόπου δεδομένων εκπαίδευσης που λέγεται Data Augmentation (Εικόνα 2.26). Καλείται η τεχνική κατά την οποία τέσσερις εικόνες εκπαίδευσης συνενώνονται σε μία εικόνα με συγκεκριμένες αναλογίες. Το όφελος από τη χρήση της τεχνικής είναι:

- Το δίκτυο βλέπει περισσότερες πληροφορίες του περιβάλλοντος της εικόνας ακόμη και έξω από το κανονικό τους πλαίσιο.
- Επιτρέπει στο μοντέλο να μάθει να αναγνωρίζει αντικείμενα σε μικρότερη κλίμακα από το συνηθισμένο.
- Η κανονικοποίηση batch θα έχει 4 φορές μείωση επειδή θα υπολογίζει στατιστικά στοιχεία για τέσσερις διαφορετικές εικόνες σε κάθε επίπεδο. Αυτό θα μείωνε την ανάγκη για μεγάλο μέγεθος mini batch κατά τη διάρκεια της εκπαίδευσης.




**Εικόνα 2.26:** Παράδειγμα Data Augmentation.

### 2.4.8 YoLov5

Δύο μήνες μετά την κυκλοφορία του YoLov4 ο Glenn Jocher, ιδρυτής και διευθύνων σύμβουλος της Ultralytics, κυκλοφόρησε την εφαρμογή ανοιχτού κώδικα του YOLOv5 στο GitHub . Το μοντέλο YoLov5 προσφέρει μια οικογένεια ανιχνευτών προεκπαιδευμένων στο σύνολο δεδομένων MS COCO. Η συγκεκριμένη μέθοδος αποτελεί την μοναδική της οικογένειας YoLo που δεν έχει δημοσιευτεί με επίσημη ερευνητική τεκμηρίωση γεγονός που αρχικά οδήγησε σε κάποιες διαμάχες.

Την παρούσα χρονική περίοδο το YoLov5 είναι ένα μοντέλο τελευταίας τεχνολογίας με τεράστια υποστήριξη και εύκολο στη χρήση και στην παραγωγή. Όλο το project υλοποιείται στο PyTorch framework, εξαλείφοντας τους περιορισμούς του Darknet (το οποίο έχει αναπτυχθεί στη γλώσσα προγραμματισμού C χωρίς προοπτική περιβάλλοντος παραγωγής). Το Darknet έχει εξελιχθεί με την πάροδο του χρόνου και είναι ένα εξαιρετικό ερευνητικό πλαίσιο για εργασία, εκπαίδευση, περαιτέρω ανάπτυξη και εξαγωγή συμπερασμάτων. Ωστόσο, έχει μικρότερη κοινότητα και επομένως λιγότερη υποστήριξη.

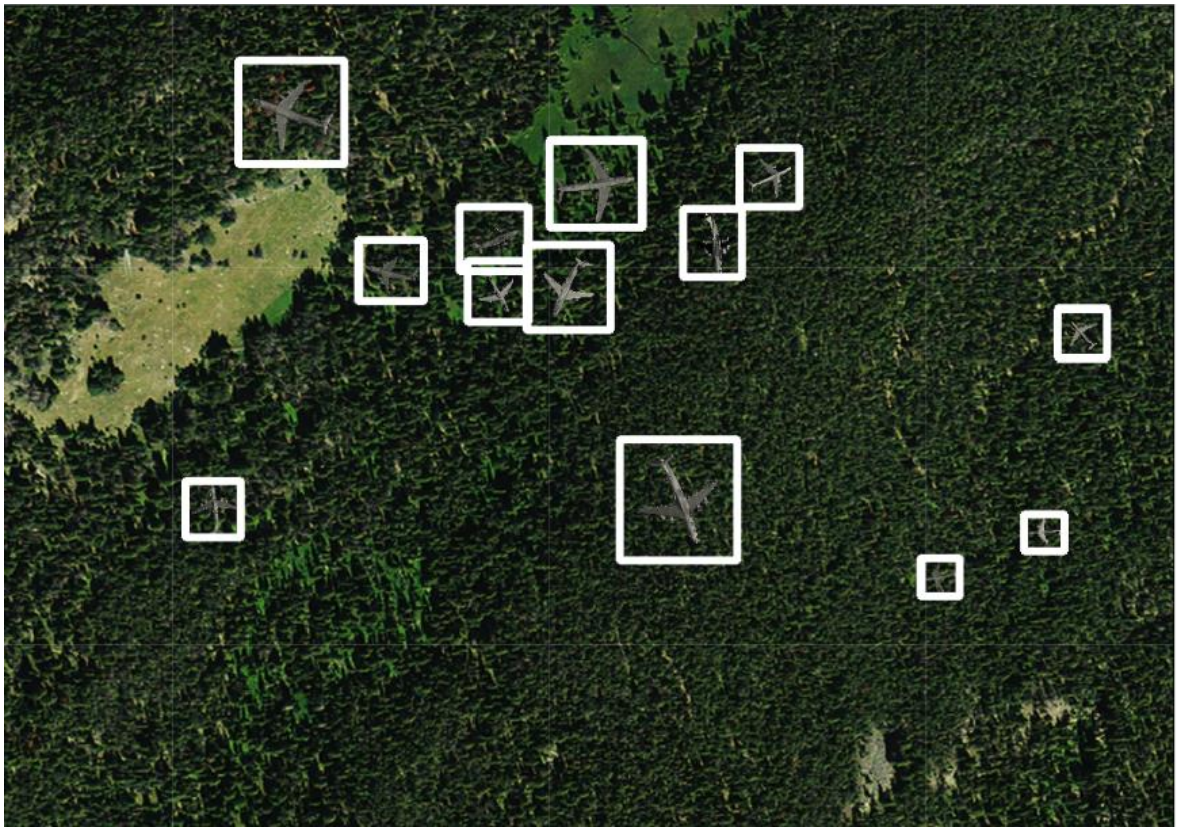
Αυτή η τεράστια αλλαγή του YOLO προς το PyTorch διευκόλυνε τους προγραμματιστές να τροποποιήσουν την αρχιτεκτονική και να την εξάγουν απευθείας σε πολλά περιβάλλοντα ανάπτυξης. Με τον τρόπο αυτό η μέθοδος είναι εύκολα προσβάσιμη στο ευρύ κοινό, επιτρέποντας διαδικασίες εκπαίδευσης, οπτικοποίησης μετρητικών και ανίχνευσης αντικειμένων.

Στην παρούσα εργασία για τους παραπάνω λόγους χρησιμοποιήθηκε αυτή η εκδοχή. Ακολουθήθηκε η διαδικασία που περιγράφεται στο επίσημο αποθετήριο και εξάγονται αποτελέσματα, μετρητικές και συμπεράσματα.

# 3. ΜΕΘΟΔΟΛΟΓΙΑ

## 3.1 Χρησιμοποιούμενα Δεδομένα

Τα διατιθέμενα δεδομένα αφορούν το σύνολο των εικόνων που θα χρησιμοποιηθούν από την διαδικασία εκπαίδευσης για να δημιουργηθεί ο ταξινομητής ενδιαφέροντος. Σε γενικότερο πλαίσιο η δημιουργία ενός ισχυρού ταξινομητή προϋποθέτει τη χρήση ενός μεγάλου αριθμού εικόνων στις οποίες θα απεικονίζονται σε τυχαίες θέσεις τα προς αναγνώριση αντικείμενα. Τα υπόψη αντικείμενα θα πρέπει να έχουν υψηλή παραλακτικότητα σε υπόβαθρα, συνθήκες φωτισμού, ποσοστό απεικόνισης του αντικειμένου, κλίμακα αντικειμένου και σχήμα, ώστε τα βάρη του δικτύου κατά την διαδικασία μάθησης να προσαρμοστούν σε όσο το δυνατόν περισσότερες διαφοροποιήσεις.



**Εικόνα 3.1 :** Ενδεικτικό παράδειγμα χρησιμοποιούμενης εικόνας εκπαίδευσης.
Πηγή: https://www.kaggle.com/datasets/aceofspades914/cgi-planes-in-satellite-imagery-w-bboxes

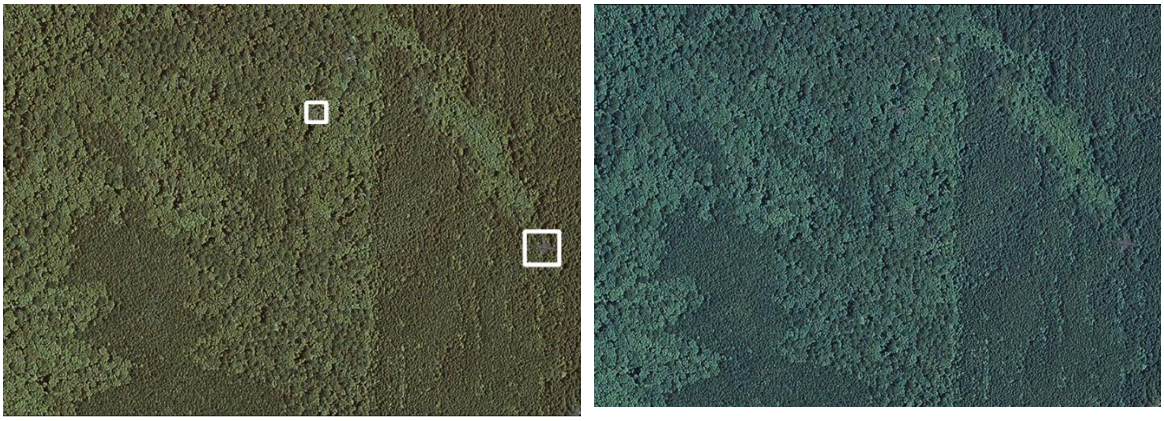
Τα δεδομένα που χρησιμοποιήθηκαν στην παρούσα εργασία αφορούν 400 εικόνες εκπαίδευσης και 100 εικόνες ελέγχου ακολουθούμενα από τα αντίστοιχα αρχεία που περιγράφουν τα προς απεικόνιση αντικείμενα σε μορφή xml και csv. Ενδεικτικό παράδειγμα εικόνας παρουσιάζεται στο παραπάνω σχήμα 3.1. Αναφέρεται επίσης ότι το μέγεθος των εικόνων είναι 1000x700 pixels (το μέγεθος θα μεταβληθεί κατά την διαδικασία εκπαίδευσης ώστε να ταιριάζει με την αρχιτεκτονική του χρησιμοποιούμενου δικτύου) ενώ ένα μέσο μέγεθος περιγεγραμμένου τετραγώνου είναι περίπου 60x60 pixel (Εικόνα 3.2).

```
</source>
<size>
    <width>1000</width>
    <height>700</height>
    <depth>3</depth>
</size>
<segmented>0</segmented>

<object>
    <name>plane</name>
    <pose>Unspecified</pose>
    <truncated>0</truncated>
    <difficult>0</difficult>
    <bndbox>
        <xmin>410</xmin>
        <ymin>356</ymin>
        <xmax>469</xmax>
        <ymax>419</ymax>
    </bndbox>
</object>
<object>
    <name>plane</name>
    <pose>Unspecified</pose>
    <truncated>0</truncated>
    <difficult>0</difficult>
    <bndbox>
        <xmin>635</xmin>
        <ymin>276</ymin>
        <xmax>738</xmax>
```

**Εικόνα 3.2:** xml αρχείο με τα περιγεγραμμένα τετράγωνα των αντικειμένων.

Το σύνολο δεδομένων αποτελείται από δορυφορικές εικόνες που έχουν δημιουργηθεί από υπολογιστή και απεικονίζουν αεροπλάνα-στόχους. Τα δεδομένα λήφθηκαν από γνωστή διαδικτυακή πηγή [34] αποθήκευσης, διάθεσης και διαμοίρασης δεδομένων για σκοπούς εκπαίδευσης δικτύων με τεχνικές μηχανικής μάθησης. Τα αντικείμενα απεικονίζονται με διαφορετικούς προσανατολισμούς, με διακυμάνσεις στην κλίμακα απεικόνισης, με παραμορφώσεις, αποκρύψεις, διαφορετικές συνθήκες φωτισμού και δυσκολία διάκρισής τους με το υπόβαθρο.

Σε ορισμένες περιπτώσεις η διαφοροποίηση μεταξύ αντικειμένου και υποβάθρου είναι δυσδιάκριτη στο ανθρώπινο μάτι όπως απεικονίζεται στο παρακάτω σχήμα (Εικόνα 3.3).



**Εικόνα 3.3:** Εικόνα εκπαίδευσης
Πηγή: https://www.kaggle.com/datasets/aceofspades914/cgi-planes-in-satellite-imagery-w-bboxes

Στις αρχικές προσεγγίσεις αναγνώρισης αντικειμένων η διαδικασία θα μετέτρεπε την εικόνα αρχικά σε grayscale και στην συνέχεια θα γινόταν αναζήτηση των ακμών ή της μεταβολής της ραδιομετρίας του προς αναγνώριση αντικειμένου. Εάν το αντικείμενο και το υπόβαθρο δεν είχαν σημαντικές διαφοροποιήσεις μεταξύ τους, τότε τέτοιες προσεγγίσεις συνήθως αποτυγχάνουν. Στην προσέγγιση της βαθιάς μηχανικής μάθησης επιδιώκεται, μετά την εκπαίδευση, να αναγνωριστεί πρώτα το αντικείμενο (recognition) με κάποια πιθανότητα πρόβλεψης και ακολούθως να εσωκλειστεί εντός περιγεγραμμένου τετραγώνου (localization). Ομοίως, εάν δεν υπάρχει μεγάλη διαφοροποίηση μεταξύ αντικειμένου και υποβάθρου, το αντικείμενο είναι δύσκολο να αναγνωριστεί.

Επιπλέον ότι το σύνολο των 400 εικόνων γενικά θεωρείται μικρό για να εκτελεστεί εκπαίδευση αλγορίθμου βαθιάς μηχανικής μάθησης. Στις περιπτώσεις που δεν υπάρχουν επιπλέον δεδομένα, μπορεί να χρησιμοποιηθεί διαδικασία γνωστή ως data augmentation για εμπλουτισμό του dataset με επιπλέον δημιουργημένες εικόνες από τις αρχικές εικόνες. Διάφορες τεχνικές data augmentation περιλαμβάνουν:

- Οριζόντια-κάθετη μετακίνηση του αντικειμένου
- Αλλαγή κλίμακας
- Περιστροφή του αντικειμένου
- Αποκοπή μέρους του αντικειμένου
- Συνδυασμός των παραπάνω

Στην παρούσα διαδικασία εκπαίδευσης δεν χρησιμοποιήθηκε data augmentation πέρα από το αρχικό σύνολο των 400 εικόνων, καθώς εκτιμήθηκε a priori ότι μια μόνο κλάση εκπαίδευσης δεν απαιτεί την δημιουργία επιπλέον διαφοροποιήσεων από τις ήδη υπάρχουσες στο αρχικό σύνολο δεδομένων. Επομένως, το χρησιμοποιηθέν dataset αποτελεί ένα ιδανικό παράδειγμα για να διευκρινιστεί η αποδοτικότητα των μεθόδων βαθιάς μηχανικής μάθησης σε ακραίες συνθήκες απεικόνισης αντικειμένων, με μικρό αριθμό δεδομένων εκπαίδευσης, χωρίς να χρησιμοποιηθεί data augmentation.

Η ροή εργασίας που ακολουθήθηκε συνοπτικά περιλαμβάνει τα κάτωθι βήματα:

1. Συλλογή δεδομένων
2. Διαμόρφωση συνοδευτικών αρχείων
3. Καθορισμός παραμέτρων εκπαίδευσης
4. Εκπαίδευση δικτύου
5. Αξιολόγηση
6. Δοκιμή σε παρεμφερής εικόνες
7. Δοκιμή σε πραγματικά σενάρια

## 3.2 Υλοποίηση Διαδικασίας Εκπαίδευσης Δικτύου

Από τις διατιθέμενες μεθόδους που μελετήθηκαν στο προηγούμενο κεφάλαιο επιλέχθηκε να χρησιμοποιηθεί για την εκπαίδευση του δικτύου η πιο πρόσφατη έκδοση της μεθόδου YoLo, ήτοι YoLo v5. Θεωρήθηκε σκόπιμο να χρησιμοποιηθεί η συγκεκριμένη μέθοδος πρωτίστως διότι βρίσκεται στην αιχμή της τεχνολογίας και επιπλέον, σύμφωνα με τους δημιουργούς, διότι παράγει γρήγορα και αξιόπιστα αποτελέσματα. Παράλληλα η διαδικασία εκπαίδευσης περιγράφεται με αρκετά σαφή και εύκολα χρησιμοποιήσιμο και κατανοητό για τον μέσο χρήστη τρόπο. Η διαδικασία εφαρμόστηκε σύμφωνα με τις οδηγίες που παρέχονται στο επίσημο αποθετήριο github των δημιουργών [29] και βασίζεται στην βιβλιοθήκη ανοιχτού κώδικα Pytorch[3].

Σε γενικές γραμμές η υλοποιούμενη μεθοδολογία περιλαμβάνει τη χρησιμοποίηση ενός όσο το δυνατόν μεγαλύτερου αριθμού εικόνων με τα αντίστοιχα για την κάθε εικόνα αρχεία που περιέχουν τα ground truth δεδομένα, τα οποία τροφοδοτούν την αρχιτεκτονική του δικτύου, ώστε αυτό να εκπαιδευθεί να αναγνωρίζει τα επιθυμητά αντικείμενα. Μετά το πέρας της εκπαίδευσης, τα αποτελέσματα ελέγχονται τόσο ποιοτικά όσο και ποσοτικά, με τον υπολογισμό κατάλληλων μετρητικών στοιχείων. Τα βήματα της διαδικασίας, όπως αυτά αναφέρονται στις οδηγίες του επίσημου αποθετηρίου, είναι πλήρως αυτοματοποιημένα με ελάχιστη παρέμβαση από τον χρήστη και παρουσιάζονται συνοπτικά παρακάτω.

### 3.2.1 Εγκατάσταση Απαιτήσεων

Η εκπαίδευση ενός νευρωνικού δικτύου για σκοπούς αναγνώρισης αντικειμένων σε οπτικές απεικονίσεις, αποτελεί μια διαδικασία αρκετά χρονοβόρα και κοστοβόρα από άποψη υπολογιστικού κόστους. Οι δημιουργοί της μεθόδου θέλοντας να ελαχιστοποιήσουν την αντικειμενική δυσκολία των χρηστών στη σύνθεση του κατάλληλου υπολογιστικού περιβάλλοντος για την εκπαίδευση των δικτύων, ανέπτυξαν την μεθοδολογία τους με τρόπο ώστε να μπορεί να αξιοποιεί

[3] https://pytorch.org/ Πρόκειται για βιβλιοθήκη η αποία αναπτύχθηκε από την Facebook έχοντας ως αφετηρία την Torch αλλά γραμμένη σε python. Μαζί με την Tensorflow αποτελούν τις πιο ευρύτατα χρησιμοποιούμενες βιβλιοθήκες στον τομέα της μηχανικής μάθησης.

υφιστάμενες υποδομές που προσφέρουν μεγάλες εταιρείες στον χώρο (google colab). Με τον τρόπο αυτό η εστίαση επικεντρώνεται περισσότερο στην ωφέλιμη διαδικασία (δημιουργία δεδομένων εκπαίδευσης, συνοδευτικών αρχείων, εκπαίδευση δικτύου, αξιολόγηση, χρήση δικτύου κα) και λιγότερο στην υποδομή (σε λογισμικό και υλισμικό) που απαιτείται για την εκπαίδευση.

Ως πρώτο βήμα της διαδικασίας απαιτείται να εγκατασταθούν όλες οι συνιστώσες οι οποίες είναι απαραίτητες στο δίκτυο για να εκτελέσει την διαδικασία εκπαίδευσης. Λίστα με τις απαραίτητες απαιτήσεις σε λογισμικό της διαδικασίας παρουσιάζεται στο παρακάτω σχήμα Εικόνα 3.4.

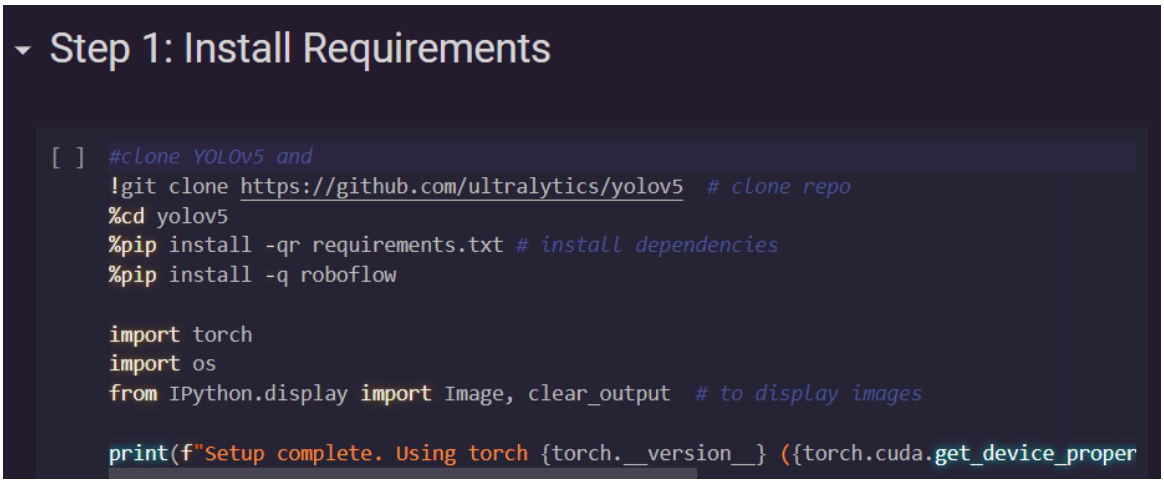
```
# YOLOv5 requirements
# Usage: pip install -r requirements.txt

# Base ----------------------------------------
matplotlib>=3.2.2
numpy>=1.18.5
opencv-python>=4.1.1
Pillow>=7.1.2
PyYAML>=5.3.1
requests>=2.23.0
scipy>=1.4.1
torch>=1.7.0
torchvision>=0.8.1
tqdm>=4.64.0
protobuf<=3.20.1  # https://github.com/ultralytics/yolov5/issues/8012

# Logging -------------------------------------
tensorboard>=2.4.1
# wandb
# clearml

# Plotting ------------------------------------
pandas>=1.1.4
seaborn>=0.11.0

# Export --------------------------------------
# coremltools>=5.2  # CoreML export
# onnx>=1.9.0  # ONNX export
# onnx-simplifier>=0.4.1  # ONNX simplifier
# nvidia-pyindex  # TensorRT export
# nvidia-tensorrt  # TensorRT export
# scikit-learn==0.19.2  # CoreML quantization
# tensorflow>=2.4.1  # TFLite export (or tensorflow-cpu, tensorflow-aarch64)
# tensorflowjs>=3.9.0  # TF.js export
# openvino-dev  # OpenVINO export

# Extras --------------------------------------
ipython  # interactive notebook
psutil  # system utilization
thop>=0.1.1  # FLOPs computation
# albumentations>=1.0.3
# pycocotools>=2.0  # COCO mAP
# roboflow
```

**Εικόνα 3.4 :** Απαιτήσεις συστήματος
Πηγή: https://github.com/ultralytics/yolov5/blob/master/requirements.txt

Ενδεικτικά αναφέρονται ότι απαιτούνται πέρα από την γλώσσα προγραμματισμού python, βιβλιοθήκες όπως numpy, matplotlib, opencv γνωστές στο πεδίο της επεξεργασίας εικόνας αλλά και οι βιβλιοθήκες torch, tensorflow γνωστές στο πεδίο της μηχανικής μάθησης. Γίνεται κατανοητό ότι η χειροκίνητη μία προς μία εγκατάσταση τους, καθώς και η διερεύνηση τυχόν ασυμβατοτήτων μεταξύ των διαφορετικών εκδόσεων απαιτεί χρόνο, κόπο και ενδεχομένως εξοικείωση.

Το google Colab έχει αναπτυχθεί από την Google και αποτελεί ένα διαδικτυακό χώρο ο οποίος επιτρέπει την εγγραφή και εκτέλεση της γλώσσας python σε διαδικτυακό περιβάλλον. Χρησιμοποιείται ευρέως σε εφαρμογές μηχανικές μάθησης και ανάλυσης δεδομένων. Πρόκειται για μια διαδικτυακή υπηρεσία που ενσωματώνει το γνωστό project ανοιχτού κώδικα Jupyter Notebook, χωρίς να απαιτεί καμιά εγκατάσταση τοπικά από τον χρήστη, ενώ παράλληλα παρέχει πρόσβαση δωρεάν σε πλήθος υπολογιστικών πηγών [32].

Η χρήση του google colab καθιστά την διαδικασία της εγκατάστασης των απαιτήσεων εφαρμόσιμη με μια απλή εντολή, με την μαζική εκτέλεση εντολών εντός του ίδιου κελιού όπως χαρακτηριστικά απεικονίζεται στο παρακάτω σχήμα 3.5.




**Εικόνα 3.5:** Εγκατάσταση απαιτήσεων μέσα από το περιβάλλον του google colab

Το παραπάνω σετ εντολών δημιουργεί έναν κλώνο του αρχικού project από το github αποθετήριο, στον χώρο που διατίθεται δωρεάν από την εταιρεία στον κάθε χρήστη. Για να περιοριστεί η ασύμμετρη και μαζική δωρεάν εκμετάλλευση της παρεχόμενης υποδομής, ο διατιθέμενος χώρος είναι πεπερασμένος σε ψηφιακό χώρο και χρόνο (διαπιστώθηκε ότι πέρα από τις 5 ώρες συνεχούς χρήσης γίνεται αυτόματη αποσύνδεση από τον προσωπικό λογαριασμό του χρήστη και διαγραφή όλων των δεδομένων που είχαν μεταφορτωθεί στον παρεχόμενο χώρο). Ακολούθως όλες οι απαιτήσεις εγκαθίστανται ταυτοχρόνως μέσω της γνωστής διαδικασίας εγκατάστασης που προσφέρει η γλώσσα python (pip install).

Πέρα από τις απαιτήσεις σε λογισμικό, οι απαιτήσεις σε υλισμικό έχουν ομοίως σημαντική συνεισφορά στην διαδικασία εκπαίδευσης. Η συνιστώσα παίζει επιπλέον σημαντικό ρόλο, διότι αφορά περισσότερο στην επένδυση της υλικής υποδομής που πρέπει να δαπανηθεί για την εξασφάλιση ταχύτητας στην διαδικασία. Σε γενικό πλαίσιο οι αλγόριθμοι μηχανικής μάθησης έχουν δημιουργηθεί ώστε να καταναλώνουν πόρους του φιλοξενούμενου συστήματος μέσω των υπολογιστικών μονάδων (processing unit) CPU και GPU. Η κάθε μονάδα έχει τα δικά της χαρακτηριστικά και η επιλογή του κατάλληλου συστήματος γίνεται με βάση το πρόβλημα που διερευνάται και τις ανάγκες σε ταχύτητα, σε υπολογιστικό κόστος και δαπανόμενη ενέργεια.

Οι μονάδες CPU[4] καλούνται επεξεργαστές γενικής-χρήσης διότι χρησιμεύουν για την υποστήριξη σχεδόν οποιασδήποτε υπολογιστικής διαδικασίας. Η χρήση της προτιμάται στον τομέα της μηχανικής μάθησης όταν οι αλγόριθμοι επεξεργάζονται δεδομένα που αφορούν χρονοσειρές ή άλλου τύπου διαδοχικών δεδομένων, για πιο κλασσικές μορφές μηχανικής μάθησης (support vector machine, recurrent neural networks) και γενικά σε διαδικασίες που δεν είναι data-intensive.

Η χρήση της GPU[5] προτιμάται για διαδικασίες βαθιάς μηχανικής μάθησης η οποία απαιτεί παράλληλες υπολογιστικές διαδικασίες με τη χρήση μεγάλου όγκου δεδομένων (εικόνες, βίντεο). Η χρήση του χαρακτηρίζεται από υψηλή κατανάλωση ενέργειας και δεν ενδείκνυται για διαδικασίες που δεν μπορούν να εκμεταλλευτούν τον παράλληλο προγραμματισμό (πχ διαδικασίες διαδοχικών υπολογισμών σε χρονοσειρές δεδομένων). Οι διαδικασίες αναγνώρισης αντικειμένων σε οπτικές απεικονίσεις, εξαιτίας της μαζικής χρήσης συνελίξεων σε όλο το εύρος της καταγεγραμμένης ραδιομετρίας των εικόνων (κατά πλάτος, μήκος και βάθος της εικόνας) εντάσσονται στην διαδικασία του παράλληλου προγραμματισμού το οποίο συνεπάγεται ότι η χρήση της GPU ενδείκνυται για τέτοιου είδους εφαρμογές.

Το περιβάλλον google colab επιτρέπει στον χρήστη την εκμετάλλευση εκτός από λογισμικό και υπολογιστική υποδομή μεταξύ τριών διαφορετικών επιλογών:

1. CPU
2. GPU

[4] CPU (Central Processing Unit): Αποτελεί κύριο μέρος οποιουδήποτε ψηφιακού συστήματος, αποτελείται από την κύρια μνήμη, την μονάδα ελέγχου και τη μονάδα αριθμητικής λογικής. Η μονάδα ελέγχου ρυθμίζει και εκτελεί όλες τις βασικές λειτουργίες του υπολογιστή [30].

[5] GPU (Graphic Processing Unit): Ηλεκρονικό κύκλωμα ικανό να αποδιδεί γραφικά για εμφάνιση εικόνας σε ηλεκτρονικό υπολογιστή. Αρχικά αναπτύχθηκαν για να επιταχύνουν εργασίες 2D-3D rendering, και εξελίχθηκαν για να υποστηρίζουν παράλληλη επεξεργασία γενικής χρήσης σε μεγάλο εύρος εφαρμογών (πχ mining crypto), πλέον είναι αναπόσπαστο κομμάτι στις εφαρμογές μηχανικής μάθησης που κάνουν χρήση μεγάλου όγκου δεδομένων.

3. TPU[6]

Στην παρούσα διαδικασία έγινε δοκιμή και εκτέλεση της διαδικασίας εκπαίδευσης του δικτύου επιλέγοντας τις επιλογές για CPU και GPU. Διαπιστώθηκε ότι ο χρόνος εκτέλεσης διαφέρει αρκετά μεταξύ των δύο επιλογών αποδεικνύοντας την καταλληλότητα του επεξεργαστή GPU για εφαρμογές μηχανικής μάθησης που εμπλέκουν εικόνες.

### 3.2.2 Συγκέντρωση Δεδομένων Εκπαίδευσης

Η εκπαίδευση του μοντέλου συνίσταται στον προσδιορισμό της κατάλληλης τιμής των βαρών μεταξύ των νευρώνων του χρησιμοποιούμενου δικτύου, ώστε αυτά να μπορούν, μέσω μιας εσωτερικής διαδικασίας υπολογισμού συνελίξεων, να αντιστοιχούν τα αντικείμενα στην εικόνα με μια από τις κλάσεις εκπαίδευσης που έχουν χρησιμοποιηθεί για την βελτιστοποίηση τους. Για να πραγματοποιηθεί η εκπαίδευση απαιτείται ένας μεγάλος αριθμός από εικόνες με επισημασμένους τους στόχους που επιθυμεί το δίκτυο να αναγνωρίζει. Η επισήμανση των στόχων γίνεται με την κατάδειξη των διαστάσεων και της τοποθεσίας του καθενός εντός της εικόνας όπου εμφανίζεται.

Σκοπός της διαδικασίας σε αυτό το στάδιο είναι να διαμορφωθεί το dataset που έχει επιλεγεί, με κατάλληλο τρόπο ώστε αυτό να είναι σε θέση να τροφοδοτήσει το δίκτυο. Η ορθή εκτέλεση της διαδικασίας απαιτεί κάθε εικόνα να ακολουθείται από μοναδικό .txt αρχείο στο οποίο θα περιγράφεται το περιγεγραμμένο τετράγωνο που εσωκλείει τον στόχο εντός της εικόνας. Η διαδικασία αυτή λέγεται labeling και γενικά είναι χρονοβόρα, καθώς απαιτεί ο χρήστης να ελέγξει μία προς μία όλες τις διαθέσιμες εικόνες και να παράξει τα αντίστοιχα συνοδευτικά αρχεία.

Η διαδικασία του labeling της μεθόδου YoLo v5 απαιτεί να παραχθεί για κάθε εικόνα ένα συνοδευτικό αρχείο το οποίο θα περιέχει τις περιγραφές των στόχων με την κάτωθι μορφή:

- Μία εγγραφή για κάθε εμφανιζόμενο αντικείμενο στην εικόνα.
- Κάθε εγγραφή θα περιέχει 5 τιμές της μορφής: class x_center y_center width height με κενά μεταξύ τους (όχι κόμμα).
- Οι παραπάνω συντεταγμένες θα είναι σε κανονικοποιημένη μορφή μεταξύ των τιμών 0-1.
- Οι κλάσεις θα κωδικοποιούνται με ακέραιους θετικούς αριθμούς ξεκινώντας από το 0.

Από τις παραπάνω προδιαγραφές ενδιαφέρον παρουσιάζει το γεγονός ότι οι συντεταγμένες των παραπάνω στοιχείων δεν απαιτούνται σε τιμές σύμφωνα με

[6] TPU (Tensor Processing Unit): Ενσωματωμένα ηλεκτρονικά κυκλώματα που είναι βελτιστοποιημένα συγκεκριμένα για την επεξεργασία πινάκων.

το σύστημα αναφοράς της ψηφιακής εικόνας (συνήθως πάνω αριστερή γωνία), αλλά σε κανονικοποιημένη μορφή μεταξύ 0-1. Η κανονικοποίηση επιτυγχάνεται με την διαίρεση της x_συντεταγμένης με το πλάτος της εικόνας και της y_συντεταγμένης με το ύψος της εικόνας (εξισώσεις 1 και 2).

Χαρακτηριστικό παράδειγμα απεικονίζεται στο σχήμα 3.6 της παρακάτω εικόνας.


**Εικόνα 3.6:** Παράδειγμα δημιουργίας labels
Πηγή: https://github.com/ultralytics/yolov5/wiki/Train-Custom-Data

$$Xnorm = \frac{Xim}{width} \qquad (1)$$

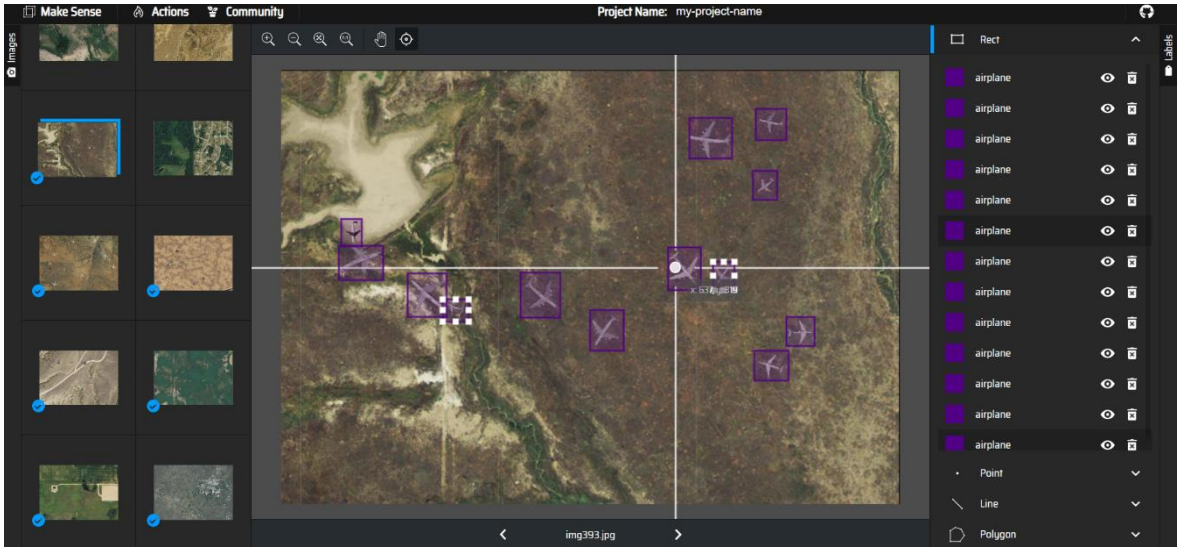
$$Ynorm = \frac{Yim}{height} \qquad (2)$$

$$width\ norm = \frac{X2 - X1}{width} \qquad (3)$$

$$height\ norm = \frac{Y2 - Y1}{height} \qquad (4)$$

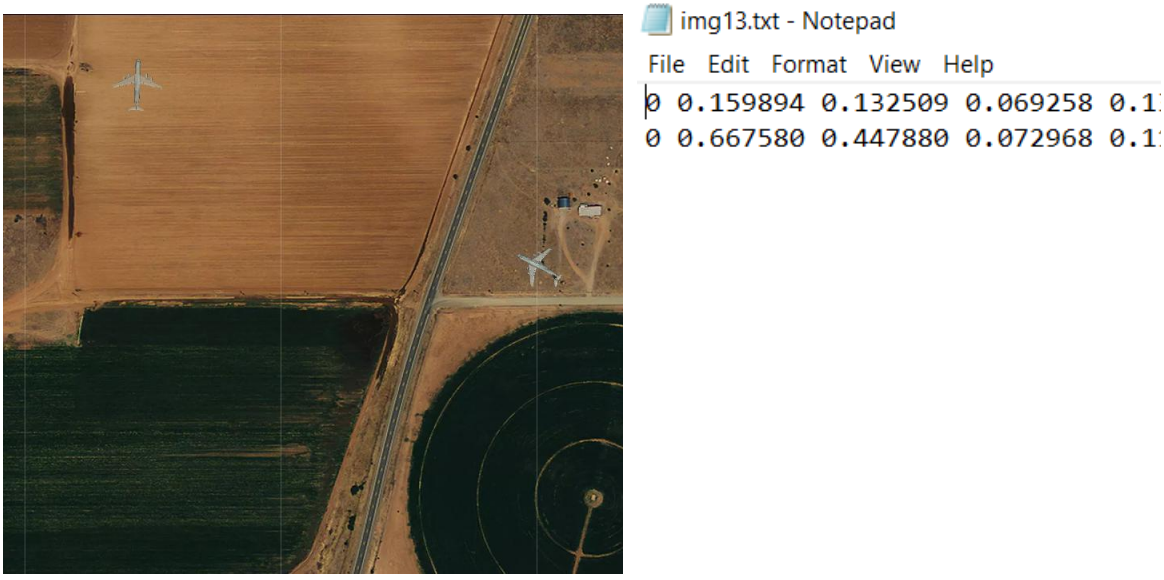
Η παραπάνω διαδικασία θα πρέπει να εφαρμοστεί σε όλες τις εικόνες που θα χρησιμοποιηθούν για την εκπαίδευση. Μπορεί να εκτελεστεί είτε με τον χειροκίνητο υπολογισμό χρησιμοποιώντας τις εξισώσεις (1) - (4), είτε με τη χρήση εργαλείων που υπάρχουν διαθέσιμα στο διαδίκτυο, όπως το Roboflow

Annotate[7] και το makesense.ai[8]. Επιλέχθηκε να χρησιμοποιηθεί το makesense.ai το περιβάλλον εργασίας του οποίου απεικονίζεται στην εικόνα 3.7.




**Εικόνα 3.7:** Δημιουργία labeling - makesense.ai

Για κάθε μια εικόνα από το διαθέσιμο dataset, εισήχθηκε μέσα στο εργαλείο makesense.ai, σχεδιάστηκαν τα περιγεγραμένα τετράγωνα που εσωκλείουν όλα τα εμφανιζόμενα αντικείμενα στην εικόνα, και αυτομάτως εξήχθησαν τα .txt αρχεία με την μορφή που περιγράφηκε παραπάνω. Στο τέλος της διαδικασίας προκύπτουν 400 εικόνες εκπαίδευσης με 400 συνοδευτικά .txt. Ενδεικτικό παράδειγμα παρουσιάζεται στην εικόνα 3.8.




**Εικόνα 3.8 :** Παράδειγμα εικόνας με το αντίστοιχο συνοδευτικό αρχείο

[7] https://roboflow.com/annotate?ref=ultralytics
[8] https://www.makesense.ai/

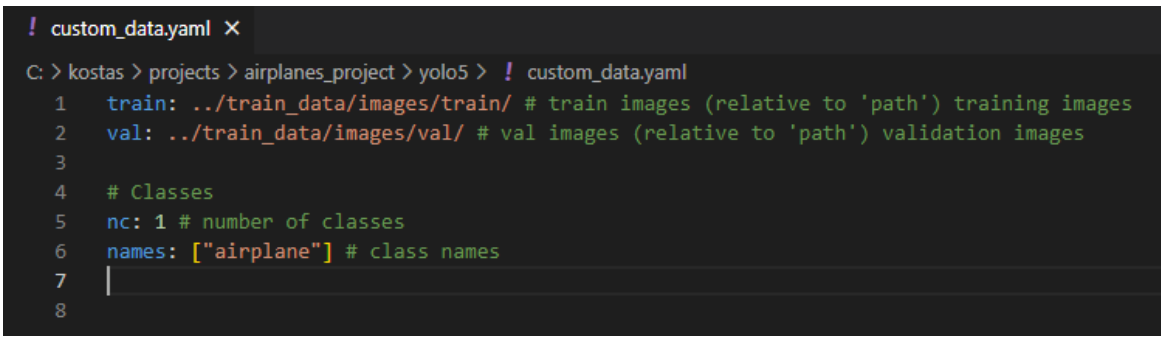
### 3.2.3 Εκπαίδευση Δικτύου

Η παραπάνω διαδικασία της δημιουργίας των συνοδευτικών αρχείων αποσκοπούσε στην δημιουργία των δεδομένων σε κατάλληλη μορφή ώστε το δίκτυο να εκπαιδευθεί να αναγνωρίζει τα επιθυμητά αντικείμενα. Η διαδικασία της εκπαίδευσης πραγματοποιείται μέσα από το περιβάλλον του google colab αξιοποιώντας τον αντίστοιχο αλγόριθμο *train.py* που έχει αναπτυχθεί από τους δημιουργούς του project.

Πρώτο βήμα της διαδικασίας είναι να δημιουργηθεί η κατάλληλη δομή εντός φακέλων με τις συνολικές εικόνες. Το dataset μοιράζεται σε ένα ποσοστό 80%-20% σε ένα υποσύνολο που θα χρησιμοποιηθεί για την εκπαίδευση (*images/train*) και ένα που θα χρησιμοποιηθεί για την αξιολόγηση (*images/val*). Αντίστοιχα δημιουργείται και φάκελος με τα labels των εικόνων. Η δομή των φακέλων απεικονίζεται σε δενδριτική μορφή ως ακολούθως:

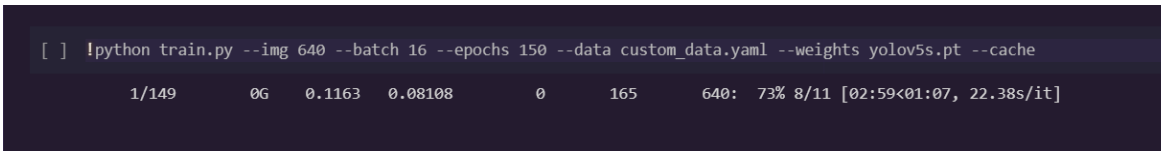
```
/content/train_data
├── README.dataset.txt
├── README.roboflow.txt
├── images
│   ├── train
│   └── val
├── labels
│   ├── train
│   └── val
└── custom_data.yaml
6 directories, 3 files
```

Η παραπάνω δομή θα πρέπει να αποτυπώνεται και στο αρχείο custom_data.yaml όπου το σύστημα θα διαβάσει τη διαδρομή των εικόνων που θα χρησιμοποιηθούν για την εκπαίδευση του δικτύου. Στο ίδιο αρχείο ορίζεται ο αριθμός των κλάσεων καθώς και το όνομα της κλάσης. Στην συγκεκριμένη περίπτωση ο αριθμός των κλάσεων είναι ένας (nc=1) και το όνομα της μοναδικής κλάσης είναι airplane (Εικόνα 3.9).




**Εικόνα 3.9:** path εικόνων εκπαίδευσης

Με την ολοκλήρωση των παραπάνω η διαδικασία της εκπαίδευσης έγκειται στην εκτέλεση του αντίστοιχου αλγορίθμου (Εικόνα 3.10). Η εκτέλεση και η επακόλουθη διαδικασία υπολογισμών εκτελούνται στον χώρο του colab αξιοποιώντας την παρεχόμενη λογισμική και υλισμική (cpu ή gpu) υποδομή.




**Εικόνα 3.10:** Εκτέλεση αλγορίθμου εκπαίδευσης του δικτύου**.**

Ο χρήστης καλείται να ορίσει ένα σύνολο παραμέτρων (ορίζονται σαν --flags κατά την εκτέλεση του κώδικα) ανάλογα με τις ανάγκες και τις απαιτήσεις του project. Κάποιες από τις standard παραμέτρους που πρέπει να οριστούν αναφέρονται ακολούθως:

- *SIZE:* Αφορά το μέγεθος αναπροσαρμογής των εικόνων προτού διοχετευθούν στο δίκτυο. Η αρχιτεκτονική του δικτύου απαιτεί μέγεθος εικόνων μεγέθους 640 pixels.

- *BATCH_SIZE: Αφορά τον* αριθμό των εικόνων που τροφοδοτούνται ως μία παρτίδα στο δίκτυο για ένα forward πέρασμα. Τροποποιείται ανάλογα με τη διαθέσιμη μνήμη GPU. Συνήθως ορίζεται σαν δυνάμεις του 32

- *EPOCHS:* Αφορά τον αριθμό των φορών που θα εκπαιδευθεί ο αλγόριθμος σε όλα τα δεδομένα εκπαίδευσης.

- *WEIGHS:* Αφορά το αρχείο με τα προεκπαιδευμένα που θα χρησιμοποιηθούν για την υφιστάμενη διαδικασία εκπαίδευσης.

Οι τιμές που επιλέχθηκαν παρουσιάζονται στον παρακάτω πίνακα:

| παράμετρος | τιμή |
|---|---|
| **--img** | 640 |
| **--batch** | 16 |
| **--epochs** | 150 |
| **--weighs** | Yolov5s |
| **--cache** | yes |

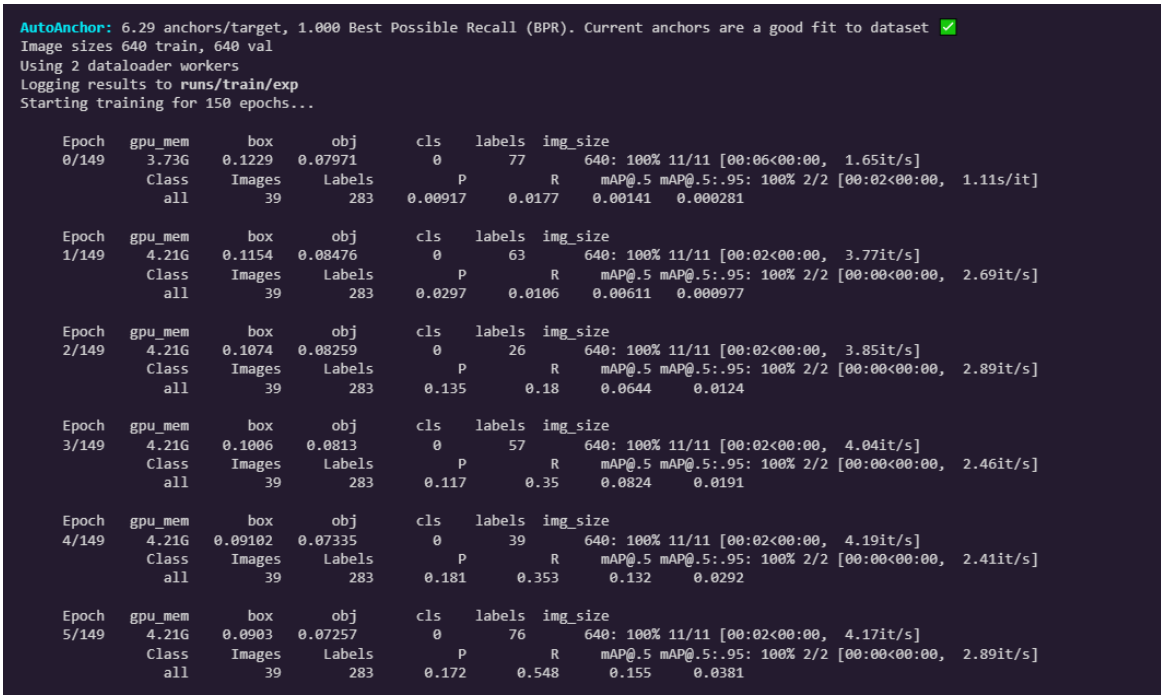
**Πίνακας 3-1:** Τιμές παραμέτρων

Στη συγκεκριμένη περίπτωση η εκπαίδευση πραγματοποιήθηκε με χρήση gpu αριθμός batch ίσος με 16, αριθμός εποχών εκπαίδευσης ίσος με 150 και ρυθμός μάθησης 0,00005 (Εικόνα 3.10 – Πίνακας 1). Οι τιμές των παραμέτρων καθορίζονται συνήθως μετά από πειραματισμό και εξαρτώνται από το είδος των δεδομένων εκπαίδευσης. Ένα απαιτητικό dataset απαιτεί χρήση gpu, μικρό αριθμό batch, μεγάλο αριθμό εποχών και μικρό ρυθμό μάθησης. Το συγκεκριμένο dataset κρίνεται δύσκολο εξαιτίας των περιορισμών των αντικειμένων που περιγράφηκαν παραπάνω και γι' αυτό τέθηκαν αυστηρές τιμές παραμέτρων.

Εκτελώντας τον παραπάνω αλγόριθμο το σύστημα αυτομάτως μεταφορτώνει στο project εξωτερικό αρχείο με ήδη εκπαιδευμένα βάρη σε κάποιο διαφορετικό dataset. Ο χρήστης έχει τη δυνατότητα να επιλέξει ανάμεσα σε διαφορετικές επιλογές προεκπαιδευμένων μοντέλων. Στην προκειμένη περίπτωση επιλέχθηκε να χρησιμοποιηθεί το μοντέλο YOLOv5s. Το ήδη εκπαιδευμένο μοντέλο έχει εκπαιδευθεί στο dataset COCO val2017 [33]. Το συγκεκριμένο dataset αποτελείται από ένα πολύ μεγάλο αριθμό εικόνων που περιλαμβάνουν 80 κατηγορίες εκπαίδευσης.

Η χρήση ενός ήδη εκπαιδευμένου μοντέλου αποσκοπεί στην σημαντική μείωση της διαδικασίας εκπαίδευσης εφαρμόζοντας την διαδικασία transfer learning. Η υπόψη διαδικασία είναι συνήθης στον τομέα της μηχανικής μάθησης, ειδικά στα πεδία της όρασης υπολογιστών (computer vision) και στην επεξεργασία φυσικής γλώσσας (natural language processing), κατά την οποία μια υφιστάμενη διαδικασία μάθησης χρησιμοποιείται για γενίκευση σε μια διαφορετική αλλά παρεμφερή κατάσταση. Στην προκειμένη περίπτωση τα βάρη του δικτύου έχουν ήδη προσαρμοστεί να αναγνωρίζουν συγκεκριμένες κατηγορίες μέσω του dataset COCO, ενώ με την παρούσα διαδικασία βελτιστοποιούνται με τρόπο ώστε να προσαρμόζονται στο συγκεκριμένο πρόβλημα. Πέρα από τη σημαντική μείωση του χρόνου εκπαίδευσης η διαδικασία transfer learning βελτιώνει και την αποδοτικότητα του δικτύου.

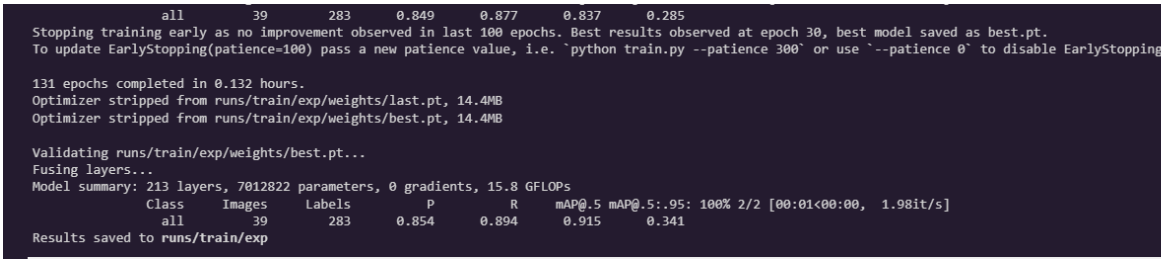
Μετά από τα παραπάνω, η διαδικασία της εκπαίδευσης του δικτύου είναι πλήρως αυτοματοποιημένη με μηδενική παρέμβαση από τον χρήστη. Με την έναρξη της εκπαίδευσης (Εικόνα 3.11) κάθε εικόνα μαζί με τα συνοδευτικά αρχεία που περιγράφουν τους στόχους, διοχετεύεται μέσα στο δίκτυο και η εκπαίδευση ακολουθεί τη διαδικασία που περιγράφηκε σε προγενέστερο κεφάλαιο. Η κάθε εποχή εκπαίδευσης ολοκληρώνεται μετά την διοχέτευση όλου του αριθμού των εικόνων στο δίκτυο. Οι εικόνες διοχετεύονται ανα αριθμό αντίστοιχο με την παράμετρο batch. Με το πέρας της κάθε εποχής εξάγονται μετρητικά στοιχεία ελέγχου της διαδικασίας.




**Εικόνα 3.11:** Διαδικασία εκπαίδευσης του δικτύου

Η διαδικασία εκπαίδευσης συνεχίζεται εως ότου επέλθει ο αριθμός των εποχών που έχει ορίσει ο χρήστης ή εως το σύστημα αντιληφθεί ότι η κάθε νέα εποχή δεν προσφέρει τίποτα στη βελτιστοποίηση. Η σύγκλιση της τιμής σε μια σταθερή και ελάχιστη τιμή σημαίνει και τον τερματισμό της εκπαίδευης του δικτύου. Υπάρχει επίσης η δυνατότητα εξαγωγής διαγράμματος το οποίο δείχνει την πρόοδο της βελτιστοποίησης.

Στην πρώτη δοκιμή που εκτελέστηκε, το σύστημα αυτό-τερματίστηκε μετά από 100 εποχές εκπαίδευσης, εξάγοντας μήνυμα ότι η κάθε νέα εποχή δεν προσφέρει επιπλέον βελτίωση των βαρών (Εικόνα 3.12). Ο ορισμός ενός ποσοστού των εικόνων πριν την διαδικασία εκπαίδευσης ως validation data, αποσκοπεί στον έλεγχο της διαδικασίας με την εξαγωγή κατάλληλων μετρητικών.




**Εικόνα 3.12 :** Τερματισμός διαδικασίας εκπαίδευσης

Περιοδικά το σύστημα αποθηκεύει αρχείο με τα μέχρι εκείνη τη στιγμή βάρη εκπαίδευσης, ώστε αφενός να υπάρχει backup σε περίπτωση που για κάποιο λόγο

τερματιστεί η διαδικασία και αφετέρου να μπορεί να συνεχιστεί η εκπαίδευση του δικτύου μελλοντικά από τη δεδομένη χρονική στιγμή και όχι από την αρχή. Μετά το πέρας της εκπαίδευσης εξάγεται το τελικό αρχείο το οποίο περιέχει την αρχιτεκτονική και τα βελτιστοποιημένα βάρη του δικτύου. Το αρχείο μπορεί πλέον να χρησιμοποιηθεί σε νέα ανεξάρτητη διαδικασία αναγνώρισης των αντικειμένων που έχει εκπαιδευθεί να αναγνωρίζει.

Η δοκιμή του δικτύου γίνεται σε νέα άγνωστα δεδομένα τα οποία περιέχουν τις κλάσεις εκπαίδευσης-βελτιστοποίησης. Η επιτυχία της διαδικασίας να αναγνωρίσει τα αντικείμενα διαπιστώνεται αρχικά με οπτική παρατήρηση ποιοτικά και κατά κύριο λόγο ποσοτικά με κατάλληλους δείκτες.

## 3.3 Δείκτες Αξιολόγησης

Η επίλυση ενός προβλήματος με τεχνικές μηχανικής και βαθιάς μηχανικής μάθησης, μπορεί να επιτευχθεί πλέον με χρήση πληθώρας επιλογών από διαφορετικές αρχιτεκτονικές. Για παράδειγμα σε ένα πρόβλημα ταξινόμησης εικόνας μπορεί να χρησιμοποιηθεί το δίκτυο VGG16 ή το ResNet 50 ανάμεσα σε πλήθος ακόμα επιλογών με διαφορετικά χαρακτηριστικά και δυνατότητες. Η επιλογή ανάμεσα στο καταλληλότερο μοντέλο υποβοηθάται από τη χρήση κατάλληλων μετρητικών στοιχείων, μέσω των οποίων επιδιώκεται η αντικειμενική αξιολόγηση των αποτελεσμάτων. Επιπλέον για σκοπούς αναγνώρισης αντικειμένων (object detection) η αξιολόγηση θα πρέπει να λάβει υπόψη ότι δεν αρκεί μόνο να ταξινομηθεί σωστά το αντικείμενο, αλλά και να βρεθεί σωστά η θέση του πάνω στην εικόνα (localization).

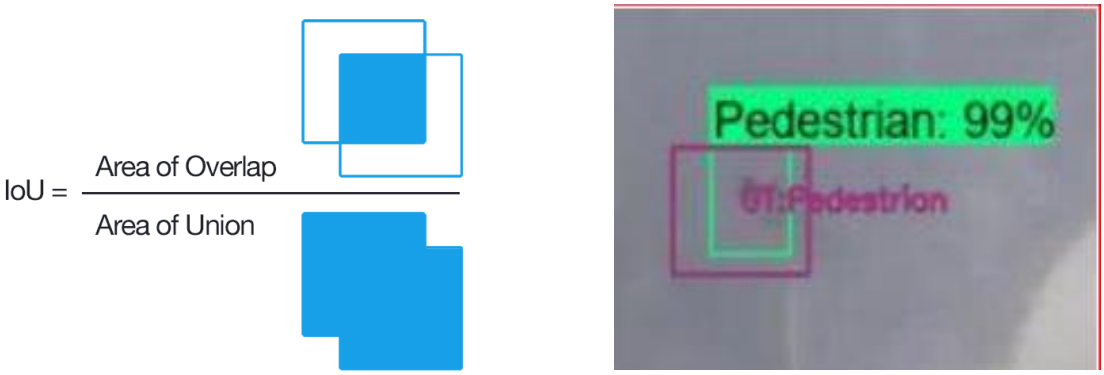
Αφού το μοντέλο που επιλέχθηκε έχει εκπαιδευθεί, ακολούθως αξιολογείται η αποδοτικότητα του στην ικανότητα να προβλέψει τις κλάσεις εκπαίδευσης πάνω στο σύνολο των δεδομένων για αξιολόγηση. Τα πιο ευρέως χρησιμοποιούμενα μετρητικά στοιχεία κατάλληλα για αξιολόγηση της αποδοτικότητας του δικτύου, που χρησιμοποιείται για σκοπούς αναγνώρισης αντικειμένων, είναι οι δείκτες:

- Precision
- Recall
- Mean Average Precision (mAP)

Για να κατανοηθούν οι παραπάνω δείκτες απαιτείται όπως περιγραφεί πρώτα ο δείκτης Intersection Over Union (IoU) που αποτελεί βασικό στοιχείο της διαδικασίας υπολογισμού των υπολοίπων μετρητικών. Παρακάτω ακολουθεί η συνοπτική παρουσίαση των δεικτών που θα χρησιμοποιηθούν στην παρούσα εργασία για την αξιολόγηση των αποτελεσμάτων της εκπαίδευσης του δικτύου YoLov5.

### 3.3.1 Δείκτης Intersection Over Union

Ο υπόψη δείκτης, γνωστός και ως Jaccard Index[9], ποσοτικοποιεί την επικάλυψη μεταξύ των ground-truth και των αναγνωρισμένων από το μοντέλο περιγεγραμμένων τετραγώνων των αντικειμένων που απεικονίζονται στην εικόνα. Ο δείκτης υπολογίζει τον λόγο της τομής των επιφανειών των δύο τετραγώνων με την ένωση τους. Στο παρακάτω σχήμα 3.13 (δεξιά) απεικονίζεται ένα παράδειγμα δύο τετραγώνων από ένα μοντέλο αναγνώρισης αντικειμένων, και σχηματικά (αριστερά) οι περιοχές των τετραγώνων που υπολογίζονται από τον δείκτη.




**Εικόνα 3.13:** Δείκτης Intersection Over Union
Πηγή: Διπλωματική Βασίλη Κωνσταντίνου, 2018

Ο δείκτης IoU συγκρίνει δύο περιοχές της εικόνας υπολογίζοντας τιμές pixel ως προς την πάνω αριστερή γωνία της ψηφιακής εικόνας, οπότε είναι καθαρός αριθμός. Οι τιμές που λαμβάνει είναι κανονικοποιημένες και κυμαίνονται από 0 εώς 1. Η τιμή 0 υποδηλώνει ότι δεν υπάρχει καμία απολύτως επικάλυψη μεταξύ του προβλεφθέντος και του αληθούς αντικειμένου, ενώ η τιμή 1 υποδηλώνει πλήρης επικάλυψη. Στην πραγματικότητα είναι σχεδόν απίθανο ότι το προβλεφθέν αντικείμενο θα συμπέσει ακριβώς με το αληθές, οπότε αυθαίρετα ορίζεται ότι μια τιμή δείκτη ίση ή μεγαλύτερη από 0.5 τυπικά θεωρείται ως καλή πρόβλεψη του αντικειμένου από το μοντέλο.

Η χρήση του παραπάνω δείκτη είναι συχνή σε προβλήματα ταξινόμησης εικόνας όπου το αποτέλεσμα είναι δυαδικό (σωστό, λάθος). Ωστόσο η χρήση του επεκτείνεται και για σκοπούς αναγνώρισης αντικειμένων με την επέκταση του δείκτη και τον καθορισμό νέων μετρητικών όπως το Precision, Recall και το Mean Average Precision (mAP).

[9] Στατιστικός δείκτης, επίσης γνωστός ως Jaccard similarity coefficient, ο οποίος χρησιμοποιείται για την μέτρηση της ομοιότητας μεταξύ πεπερασμένων συνόλων δειγμάτων. Αναπτύχθηκε από τον Grove Karl Gilbert το 1884.

### 3.3.2 Δείκτες Precision και Recall

Οι δείκτες Precision και Recall εμφανίζονται συχνά σε προβλήματα μηχανικής μάθησης που αφορούν ταξινόμηση εικόνας. Πριν γίνει η επεξήγηση τους απαιτείται όπως ορισθούν οι όροι True Positive, False Positive, False Negative και True Negative οι οποίοι χρησιμοποιούνται για να υπολογισθούν οι δείκτες Precision και Recall. Έστω ότι σε κάποιο πρόβλημα δυαδικής ταξινόμησης οι πιθανές τιμές πρόβλεψης του μοντέλου είναι 1 (positive) ή 0 (negative). Οι τιμές 0 και 1 αποτελούν την αριθμητική αναπαράσταση οποιαδήποτε ποιοτικής ταξινόμησης με φυσική σημασία για τον άνθρωπο (πχ γάτα-σκύλος, άντρας-γυναίκα, αεροπλάνο-ελικόπτερο κτλ.) Ορίζονται οι παρακάτω έννοιες:

**True positive (TP):** Είναι ο αριθμός των αντικειμένων που έχουν αναγνωριστεί από το μοντέλο ορθώς, συγκεκριμένα στην δυαδική ταξινόμηση τα δεδομένα με τον χαρακτηρισμό 1 έχουν ταξινομηθείς ως 1.

**False Positive (FP):** Είναι ο αριθμός των αντικειμένων που έχουν αναγνωριστεί και ταξινομηθεί στη μια κλάση ενώ ανήκουν στην άλλη, συγκεκριμένα τα αντικείμενα με χαρακτηρισμό 0 έχουν ταξινομηθεί ως κλάση 1.

**False Negative (FN):** Είναι ο αριθμός των αντικειμένων που έχουν αναγνωριστεί και ταξινομηθεί στη μια κλάση ενώ ανήκουν στην άλλη, συγκεκριμένα τα αντικείμενα με χαρακτηρισμό 1 έχουν ταξινομηθεί ως κλάση 0.

**True Negative (TN):** Είναι ο αριθμός των αντικειμένων που έχουν αναγνωριστεί από το μοντέλο ορθώς, δηλαδή τα δεδομένα με τον χαρακτηρισμό 0 έχουν ταξινομηθείς ως κλάση 0.

Οι παραπάνω όροι χρησιμοποιούνται για την βαθύτερη κατανόηση και την εξαγωγή περαιτέρω δεικτών για την αποδοτικότητα του μοντέλου. Με την βοήθεια των παραπάνω ορίζονται οι μετρητικές Precision και Recall ως ακολούθως.

Precision είναι ο λόγος μεταξύ των ορθών ταξινομημένων αντικειμένων της κλάσης 1 (Positive) προς των συνολικών αριθμό των αντικειμένων που έχουν ταξινομηθεί ως Positive. Ο συγκεκριμένος δείκτης μετράει την ακρίβεια του μοντέλου στην ταξινόμηση των δειγμάτων ως Positive. Πέρα των ορθών positive αντικειμένων, ο δείκτης λαμβάνει υπόψη του τα αντικείμενα που λανθασμένα έχουν ταξινομηθεί ως Positive (πχ ένα κομμάτι εικόνας που έχει αναγνωρισθεί ως αεροπλάνο ενώ δεν υπάρχει αεροπλάνο στο συγκεκριμένο τμήμα της εικόνας). Από τον ορισμό του λόγου γίνεται αντιληπτό ότι όσο μεγαλύτερος είναι ο αριθμητής τόσο μεγαλύτερη είναι η τιμή του δείκτη, όπου μεγάλος αριθμητής σημαίνει μεγάλο πλήθος αντικειμένων που ανιχνεύθηκαν ορθώς στην κλάση Positive.

$$Precision = \frac{TP}{TP + FP}$$

Recall είναι ο λόγος μεταξύ των ορθών ταξινομημένων αντικειμένων της κλάσης 1 (Positive) προς των συνολικών αριθμό των Positive αντικειμένων. Όσο μεγαλύτερος ο δείκτης τόσο μεγαλύτερος ο αριθμός των Positive αντικειμένων που έχουν ανιχνευθεί. Στο παρακάτω κλάσμα ο παρονομαστής (TP+FN) υποδηλώνει τον συνολικό αριθμό των Ground Truth αντικειμένων (καθώς ένα FN αντικείμενο ανήκει στην ουσία στην κλάση Positive).

$$Recall = \frac{TP}{TP + FN}$$

Ενώ ο δείκτης Precision λαμβάνει υπόψη του το πως έχουν ταξινομηθεί τα Negative αντικείμενα, ο δείκτης Recall είναι ανεξάρτητος στον τρόπο που έχουν ταξινομηθεί τα Negative αντικείμενα και λαμβάνει υπόψη μόνο στο πως έχουν ταξινομηθεί τα Positive αντικείμενα. Ο δείκτης λαμβάνει την τιμή 1 (ή 100%) όταν όλα τα positive αντικείμενα έχουν ταξινομηθεί ως positive. Αντιθέτως λαμβάνει μικρή τιμή όταν Positive δείγματα έχουν ταξινομηθεί ως Negative.

### 3.3.3 Δείκτες στην Αναγνώριση Αντικειμένων

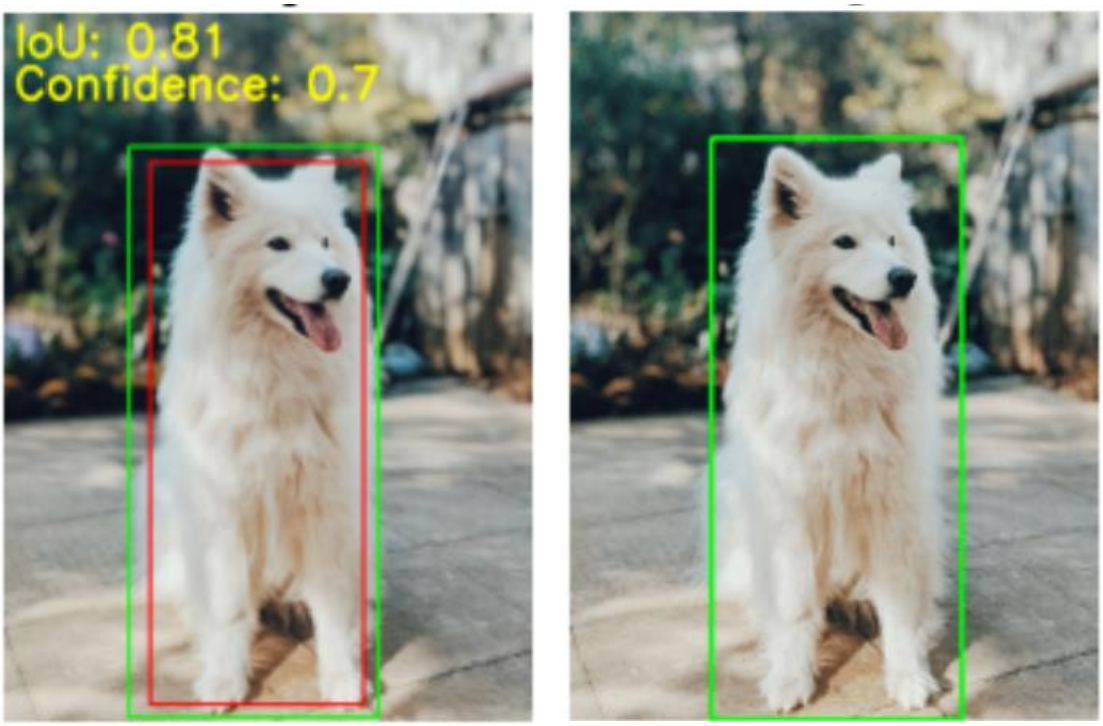
Όλα τα παραπάνω αφορούν μετρητικές υπό την οπτική της ταξινόμησης ενός αντικειμένου σε μια εικόνα εντός δύο κλάσεων (0 ή 1) που μπορεί να υποδηλώνει οποιαδήποτε ποιοτική κατηγοριοποίηση. Στην παρούσα παράγραφο θα επεκταθούν οι έννοιες ώστε να κατανοηθούν υπό την οπτική της αναγνώρισης αντικειμένων (object detection). Για σκοπούς αναγνώρισης αντικειμένων ουσιαστικό ρόλο παίζουν οι δείκτες *IoU* και το *επίπεδο εμπιστοσύνης* (confidence score) με τη βοήθεια των οποίων υπολογίζονται τα TP, FP, FN και ακολούθως οι δείκτες Precision και Recall.

Ως επίπεδο εμπιστοσύνης ορίζεται ως η πιθανότητα του αντικειμένου που περικλείεται από το περιγεγραμμένο τετράγωνο να είναι όντως το σωστό αντικείμενο. Κάθε ταξινομητής εξάγει αποτελέσματα προγνώσεων επί της εικόνας, παρέχοντας για κάθε τετράγωνο ένα επίπεδο εμπιστοσύνης που υποδηλώνει την ακρίβεια της πρόγνωσης του αντικειμένου. Υπο την παραπάνω οπτική ορίζονται οι έννοιες ως ακολούθως.

Η πρόβλεψη από το μοντέλο θεωρείται True Positive όταν ικανοποιεί δύο όρους:

- Το επίπεδο εμπιστοσύνης του τετραγώνου είναι μεγαλύτερο από ένα κατώφλι το οποίο έχει ορίσει ο χρήστης (είναι υπερπαράμετρος του συστήματος).
- Ο δείκτης IoU μεταξύ των δύο τετραγώνων πρέπει να είναι μεγαλύτερος από το κατώφλι που έχει ορίσει ο χρήστης.

Οι δυο παραπάνω προϋποθέσεις σημαίνουν ότι το αντικείμενο προκειμένου να θεωρείται ότι έχει αναγνωριστεί σωστά από το σύστημα, θα πρέπει όχι μόνο να αναγνωριστεί σωστά αλλά και στη σωστή θέση στην εικόνα (localization). Στην παρακάτω εικόνα 3.14 απεικονίζεται μια True Positive αναγνώριση αντικειμένου με δείκτη IoU 81% και επίπεδο εμπιστοσύνης 70%, που σημαίνει ότι τα δύο τετράγωνα (ground truth και πρόγνωσης) έχουν επικάλυψη ικανή ώστε να θεωρείται ότι αναγνωρίστηκε αντικείμενο, και η πιθανότητα το αντικείμενο να ανήκει σε συγκεκριμένη κλάση είναι 70%.

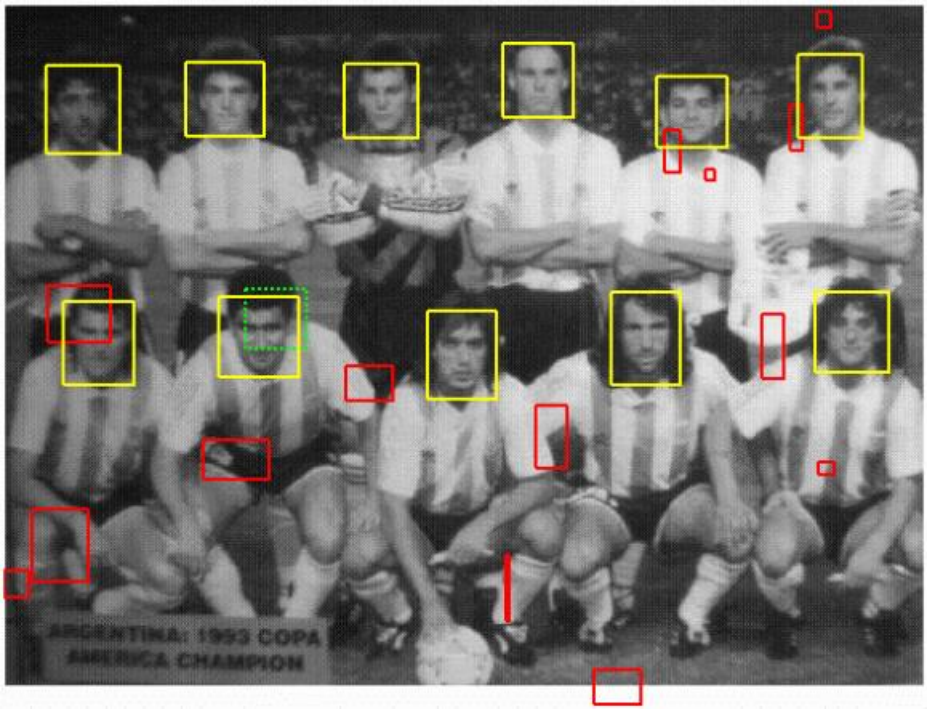



**Εικόνα 3.14:** True Positive (αριστερά) – False Negative (δεξιά)
Πηγή: https://pyimagesearch.com/

Η πρόβλεψη από το μοντέλο θεωρείται False Positive όταν συμβαίνουν τα κάτωθι:

- Το μοντέλο έχει αναγνωρίσει ένα αντικείμενο με υψηλό επίπεδο εμπιστοσύνης, αλλά το αντικείμενο δεν είναι παρόν εντός του τετραγώνου, άρα ο δείκτης IoU είναι μηδέν.
- Ο δείκτης IoU είναι μικρότερος από την ορισθείσα τιμή κατωφλίου.
- Το εξαγόμενο περιγεγραμμένο τετράγωνο ευθυγραμμίζεται σε μεγάλο βαθμό με το αληθές τετράγωνο, αλλά η πρόβλεψη της κλάσης του αντικειμένου είναι λανθασμένη.

Η παρακάτω εικόνα 3.15 απεικονίζει με πράσινο χρώμα το True Positive, με κίτρινο τα Ground Truth και με κόκκινο τα False Negative που αναγνωρίστηκαν από το μοντέλο.




**Εικόνα 3.15:** Αναγνώριση FN αντικειμένων
Πηγή:*https://www.cc.gatech.edu/classes/AY2016/cs4476_fall/results/proj5/html/pkundra3/index.html*

Το αποτέλεσμα θεωρείται False Negative όταν ενώ υπάρχει αντικείμενο το μοντέλο δεν μπόρεσε να το αναγνωρίσει όπως απεικονίζεται στην εικόνα 3.14 (δεξιά). Στο συγκεκριμένο παράδειγμα ενώ υπάρχει το ground truth τετράγωνο που περικλείει το αντικείμενο, ο αλγόριθμος πρόβλεψης δεν εξήγαγε κάποιο αποτέλεσμα επομένως ο δείκτης IoU είναι μηδέν.

Τέλος, True Negative γενικά δεν χρησιμοποιείται διότι δεν εμφανίζεται πουθενά στον υπολογισμό των δεικτών Recall και Precision. Ένα True Negative αντικείμενο θα ήταν εάν οι δείκτες IoU και το επίπεδο εμπιστοσύνης θα ήταν λιγότερα από αυτά που είχε ορίσει ο χρήστης. Μια τέτοια κατηγοριοποίηση εμφανίζεται συχνά σε προβλήματα αναγνώρισης αντικειμένων, διότι το background καλύπτει πολύ μεγαλύτερο μέρος της εικόνας από τα προς αναγνώριση αντικείμενα.

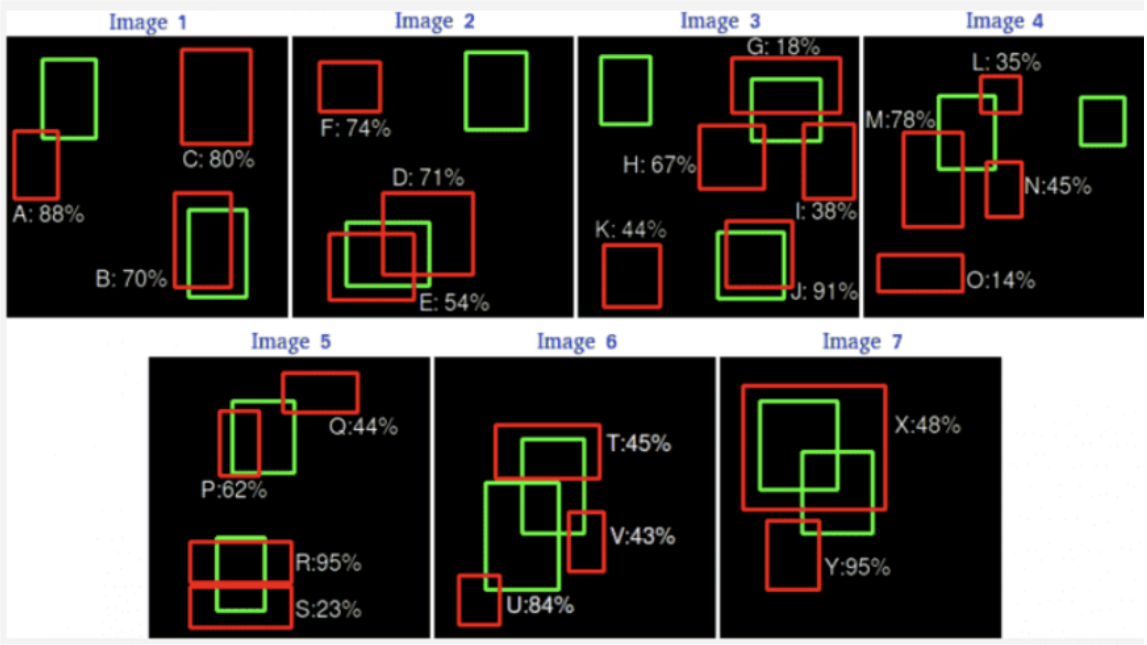
Από τα παραπάνω συμπεραίνεται ότι εάν το αποτέλεσμα έχει υψηλούς δείκτες Precision και Recall, σημαίνει ότι το μοντέλο προβλέπει σωστά τα Positive δείγματα και επιπλέον ότι προβλέπει τα περισσότερα Positive δείγματα δηλαδή δεν αγνοεί ή βλέπει τα δείγματα ως Negative. Εάν το μοντέλο έχει υψηλό Precision και χαμηλό Recall, τότε προβλέπει με ακρίβεια τα δείγματα ως Positive αλλά μόνο λίγα από το συνολικό αριθμό τους (τα περισσότερα αναγνωρίζονται

ως False Negative). Σε κάθε περίπτωση επιδιώκεται το μοντέλο να έχει υψηλούς δείκτες Precision και Recall.

### 3.3.4 Καμπύλη Precision-Recall

Οι δείκτες αυτοί μπορούν να αναπαρασταθούν σε μια γραφική παράσταση η οποία καλείται καμπύλη Precision-Recall. Αποτελεί έναν τρόπο οπτικοποίησης των τιμών των δύο δεικτών σε διαφορετικές τιμές κατωφλίων που έχουν οριστεί από τον χρήστη. Η καμπύλη βοηθά στην επιλογή της καλύτερης δυνατής τιμής κατωφλίου που μεγιστοποιεί τις τιμές των δύο δεικτών, επομένως μεγιστοποιεί και την αποδοτικότητα του μοντέλου.

Η καμπύλη Precision-Recall είναι ένας καλός τρόπος αξιολόγησης της απόδοσης ενός ανιχνευτή, σχεδιάζοντας μια καμπύλη για κάθε κλάση σε διαφορετικά επίπεδα τιμών εμπιστοσύνης. Με τη βοήθεια της συγκεκριμένης καμπύλης μπορεί να υπολογιστεί ο δείκτης Average Precision για την κάθε κλάση του μοντέλου, και ακολούθως ο δείκτης mean Average Precision (mAP) που μετρά την αποδοτικότητα του μοντέλου στο σύνολο των κλάσεων πρόβλεψης. Στην παρακάτω εικόνα 3.16 απεικονίζεται ένα παράδειγμα 7 εικόνων με 15 ground truth αντικείμενα (απεικονιζόμενα με πράσινο χρώμα) και τα 24 πολύγωνα των αντίστοιχων προβλέψεων που έχουν εξαχθεί από το μοντέλο με τα επίπεδα εμπιστοσύνης για το καθένα (με κόκκινο χρώμα).



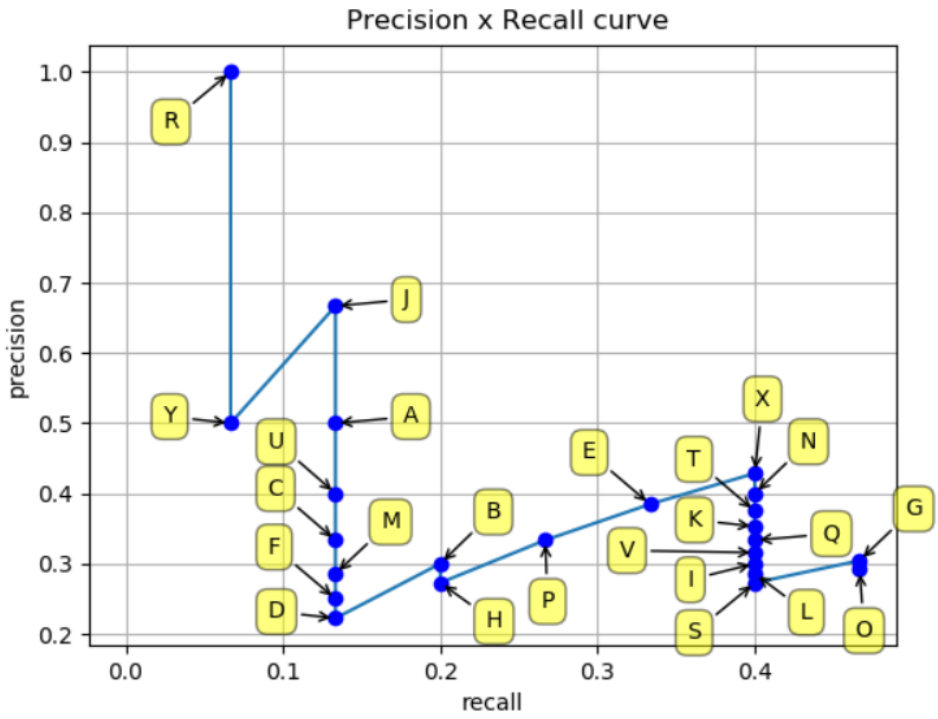

**Εικόνα 3.16**: Ground truth αντικείμενα (πράσινο χρώμα) και 24 πολύγωνα εξαγόμενα από το μοντέλο με τις αντίστοιχες τιμές εμπιστοσύνης (κόκκινο χρώμα).
Πηγή: https://github.com/rafaelpadilla/Object-Detection-Metrics

Τα αποτελέσματα των παραπάνω προβλέψεων τοποθετούνται σε πίνακα σε φθίνουσα σειρά επιπέδου εμπιστοσύνης, όπως στην παρακάτω εικόνα 3.17. Κάθε εγγραφή του πίνακα αντιστοιχεί σε μια διαφορετική πρόβλεψη για την οποία υπολογίζονται τα TP και FP και αθροιστικά υπολογίζονται τα συνολικά TP (Acc TP) και FP (accFP) λαμβάνοντας υπόψη την τρέχουσα και τις προγενέστερες εγγραφές. Ακολούθως για κάθε εγγραφή υπολογίζονται οι δείκτες Precision και Recall χρησιμοποιώντας τις εξισώσεις όπως αυτές παρουσιάστηκαν παραπάνω.

| Images | Detections | Confidences | TP | FP | Acc TP | Acc FP | Precision | Recall |
|---|---|---|---|---|---|---|---|---|
| Image 5 | R | 95% | 1 | 0 | 1 | 0 | 1 | 0.0666 |
| Image 7 | Y | 95% | 0 | 1 | 1 | 1 | 0.5 | 0.0666 |
| Image 3 | J | 91% | 1 | 0 | 2 | 1 | 0.6666 | 0.1333 |
| Image 1 | A | 88% | 0 | 1 | 2 | 2 | 0.5 | 0.1333 |
| Image 6 | U | 84% | 0 | 1 | 2 | 3 | 0.4 | 0.1333 |
| Image 1 | C | 80% | 0 | 1 | 2 | 4 | 0.3333 | 0.1333 |
| Image 4 | M | 78% | 0 | 1 | 2 | 5 | 0.2857 | 0.1333 |
| Image 2 | F | 74% | 0 | 1 | 2 | 6 | 0.25 | 0.1333 |
| Image 2 | D | 71% | 0 | 1 | 2 | 7 | 0.2222 | 0.1333 |
| Image 1 | B | 70% | 1 | 0 | 3 | 7 | 0.3 | 0.2 |
| Image 3 | H | 67% | 0 | 1 | 3 | 8 | 0.2727 | 0.2 |
| Image 5 | P | 62% | 1 | 0 | 4 | 8 | 0.3333 | 0.2666 |
| Image 2 | E | 54% | 1 | 0 | 5 | 8 | 0.3846 | 0.3333 |
| Image 7 | X | 48% | 1 | 0 | 6 | 8 | 0.4285 | 0.4 |
| Image 4 | N | 45% | 0 | 1 | 6 | 9 | 0.4 | 0.4 |
| Image 6 | T | 45% | 0 | 1 | 6 | 10 | 0.375 | 0.4 |
| Image 3 | K | 44% | 0 | 1 | 6 | 11 | 0.3529 | 0.4 |
| Image 5 | Q | 44% | 0 | 1 | 6 | 12 | 0.3333 | 0.4 |
| Image 6 | V | 43% | 0 | 1 | 6 | 13 | 0.3157 | 0.4 |
| Image 3 | I | 38% | 0 | 1 | 6 | 14 | 0.3 | 0.4 |
| Image 4 | L | 35% | 0 | 1 | 6 | 15 | 0.2857 | 0.4 |
| Image 5 | S | 23% | 0 | 1 | 6 | 16 | 0.2727 | 0.4 |
| Image 3 | G | 18% | 1 | 0 | 7 | 16 | 0.3043 | 0.4666 |
| Image 4 | O | 14% | 0 | 1 | 7 | 17 | 0.2916 | 0.4666 |

**Εικόνα 3.17**: Πίνακας αποτελεσμάτων Εικόνας 3.16
Πηγή: https://github.com/rafaelpadilla/Object-Detection-Metrics

Αφού έχουν υπολογιστεί οι τιμές Precision και Recall, οι τιμές μπορούν να τοποθετηθούν σε διάγραμμα όπου στον άξονα των xx τοποθετούνται οι τιμές Recall και στον άξονα yy οι τιμές Precision όπως στην παρακάτω εικόνα 3.18.




**Εικόνα 3.18**: Καμπύλη Precision-Recall
Πηγή: https://github.com/rafaelpadilla/Object-Detection-Metrics

Στο παραπάνω σχήμα απεικονίζονται οι προτεινόμενες περιοχές με το υψηλότερο επίπεδο εμπιστοσύνης R επάνω αριστερά με πράσινο χρώμα και την ανίχνευση με το χαμηλότερο επίπεδο εμπιστοσύνη κάτω δεξιά σε κόκκινο χρώμα. Χρησιμοποιώντας την καμπύλη μπορεί να υπολογιστεί το εμβαδόν κάτω από την γραφική παράσταση με τον άξονα xx, το οποίο στην ουσία υποδηλώνει τον δείκτη Average Precision. Πριν πραγματοποιηθεί ο υπολογισμός του εμβαδού προηγείται η εξομάλυνση της γραφικής παράστασης.

Τέλος, για να υπολογιστεί η μετρητική mean Average Precision (mAP), υπολογίζονται πρώτα το Average Precision της κάθε κλάσης ξεχωριστά με τον τρόπο που υποδείχθηκε παραπάνω και στη συνέχεια υπολογίζεται ο μέσος όρος όλων των AP της κάθε κλάσης. Για παράδειγμα, στην εικόνα 3.19 εμφανίζονται τα AP στο σύνολο δεδομένων PASCAL VOC για διάφορους ταξινομητές (π.χ. SSD300, SSD512 κτλ), καθώς και ο μέσος όρος όλων των AP ώστε τελικώς να ληφθεί η μέση ακρίβεια του ταξινομητή.

| Method | data | mAP | aero | bike | bird | boat | bottle | bus | car | cat | chair | cow | table | dog | horse | mbike | person | plant | sheep | sofa | train | tv |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Fast [6] | 07 | 66.9 | 74.5 | 78.3 | 69.2 | 53.2 | 36.6 | 77.3 | 78.2 | 82.0 | 40.7 | 72.7 | 67.9 | 79.6 | 79.2 | 73.0 | 69.0 | 30.1 | 65.4 | 70.2 | 75.8 | 65.8 |
| Fast [6] | 07+12 | 70.0 | 77.0 | 78.1 | 69.3 | 59.4 | 38.3 | 81.6 | 78.6 | 86.7 | 42.8 | 78.8 | 68.9 | 84.7 | 82.0 | 76.6 | 69.9 | 31.8 | 70.1 | 74.8 | 80.4 | 70.4 |
| Faster [2] | 07 | 69.9 | 70.0 | 80.6 | 70.1 | 57.3 | 49.9 | 78.2 | 80.4 | 82.0 | 52.2 | 75.3 | 67.2 | 80.3 | 79.8 | 75.0 | 76.3 | 39.1 | 68.3 | 67.3 | 81.1 | 67.6 |
| Faster [2] | 07+12 | 73.2 | 76.5 | 79.0 | 70.9 | 65.5 | 52.1 | 83.1 | 84.7 | 86.4 | 52.0 | 81.9 | 65.7 | 84.8 | 84.6 | 77.5 | 76.7 | 38.8 | 73.6 | 73.9 | 83.0 | 72.6 |
| Faster [2] | 07+12+COCO | 78.8 | 84.3 | 82.0 | 77.7 | 68.9 | 65.7 | 88.1 | 88.4 | 88.9 | 63.6 | 86.3 | 70.8 | 85.9 | 87.6 | 80.1 | 82.3 | 53.6 | 80.4 | 75.8 | 86.6 | 78.9 |
| SSD300 | 07 | 68.0 | 73.4 | 77.5 | 64.1 | 59.0 | 38.9 | 75.2 | 80.8 | 78.5 | 46.0 | 67.8 | 69.2 | 76.6 | 82.1 | 77.0 | 72.5 | 41.2 | 64.2 | 69.1 | 78.0 | 68.5 |
| SSD300 | 07+12 | 74.3 | 75.5 | 80.2 | 72.3 | 66.3 | 47.6 | 83.0 | 84.2 | 86.1 | 54.7 | 78.3 | 73.9 | 84.5 | 85.3 | 82.6 | 76.2 | 48.6 | 73.9 | 76.0 | 83.4 | 74.0 |
| SSD300 | 07+12+COCO | 79.6 | 80.9 | 86.3 | 79.0 | **76.2** | 57.6 | 87.3 | 88.2 | 88.6 | 60.5 | 85.4 | **76.7** | **87.5** | **89.2** | 84.5 | 81.4 | 55.0 | 81.9 | **81.5** | 85.9 | 78.9 |
| SSD512 | 07 | 71.6 | 75.1 | 81.4 | 69.8 | 60.8 | 46.3 | 82.6 | 84.7 | 84.1 | 48.5 | 75.0 | 67.4 | 82.3 | 83.9 | 79.4 | 76.6 | 44.9 | 69.9 | 69.1 | 78.1 | 71.8 |
| SSD512 | 07+12 | 76.8 | 82.4 | 84.7 | 78.4 | 73.8 | 53.2 | 86.2 | 87.5 | 86.0 | 57.8 | 83.1 | 70.2 | 84.9 | 85.2 | 83.9 | 79.7 | 50.3 | 77.9 | 73.9 | 82.5 | 75.3 |
| SSD512 | 07+12+COCO | **81.6** | **86.6** | **88.3** | **82.4** | 76.0 | **66.3** | **88.6** | **88.9** | **89.1** | **65.1** | **88.4** | 73.6 | 86.5 | 88.9 | **85.3** | **84.6** | **59.1** | **85.0** | 80.4 | **87.4** | **81.2** |

**Εικόνα 3.19:** mAP και AP 20 κλάσεων στο σύνολο δεδομένων PASCAL VOC
Πηγή: https://github.com/rafaelpadilla/Object-Detection-Metrics

Το AP σε κάθε κατηγορία δίνει μια πιο λεπτομερή αξιολόγηση του ανιχνευτή, καθώς υποδεικνύει τις κατηγορίες που ο ανιχνευτής είχε καλή απόδοση και τις κατηγορίες στις οποίες είχε κακή απόδοση.

`

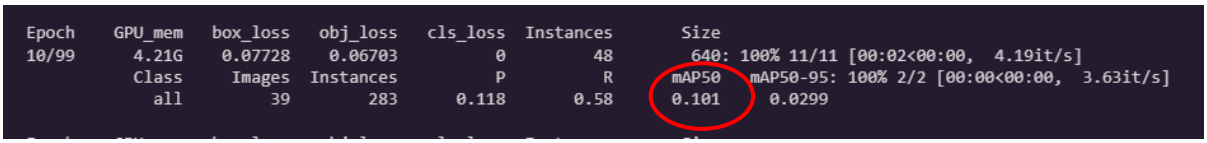
# 4. ΠΕΙΡΑΜΑΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΚΑΙ ΑΞΙΟΛΟΓΗΣΗ

## 4.1 Εισαγωγή

Ακολουθώντας την παραπάνω διαδικασία εκπαιδεύτηκε το δίκτυο που αναφέρθηκε σε ανωτέρω παραγράφους και ακολούθως παρουσιάζονται τα αποτελέσματα. Παρουσιάζονται επίσης οι μετρητικές που εξηγήθηκαν παραπάνω οι οποίες αυτόματα εξήχθησαν από την διαδικασία εκπαίδευσης. Αναφέρεται επίσης ότι για χάρη πειραματισμού εκτελέστηκαν δυο διαφορετικές διαδικασίες εκπαίδευσης του δικτύου. Η πρώτη δοκιμή αφορούσε την εκπαίδευση όπως αυτή παρουσιάστηκε παραπάνω, ενώ η δεύτερη εκπαίδευση πραγματοποιήθηκε χρησιμοποιώντας μόνο τα τελευταία επίπεδα του δικτύου παγώνοντας τα αρχικά επίπεδα. Στις παρακάτω παραγράφους αναφέρονται τα αποτελέσματα της εκπαίδευσης στις δύο διαφορετικές διαδικασίες και οι μετρητικές που εξήχθησαν πάνω στα δεδομένα αξιολόγησης του δικτύου όπως αυτά ορίστηκαν κατά την έναρξης της εκπαίδευσης (80%-20%). Ακολούθως, περιγράφονται τα αποτελέσματα των μετρητικών πάνω σε νέο dataset παρεμφερές με το αρχικό που χρησιμοποιήθηκε για την εκπαίδευση, που αποτελείται από επιπλέον 100 κατασκευασμένες εικόνες. Τέλος, η αποδοτικότητα του εκπαιδευμένου δικτύου δοκιμάστηκε σε τρείς δορυφορικές εικόνες υπερύψηλής ανάλυσης (30 εκ.) που απεικονίζουν τρεις πραγματικές σκηνές αεροδρομίων (όχι κατασκευασμένες όπως το χρησιμοποιηθέν dataset), με σκοπό να ελεγχθεί η ικανότητα του να ανιχνεύσει το αεροπλάνο σε τρία διαφορετικά σενάρια του πραγματικού κόσμου.

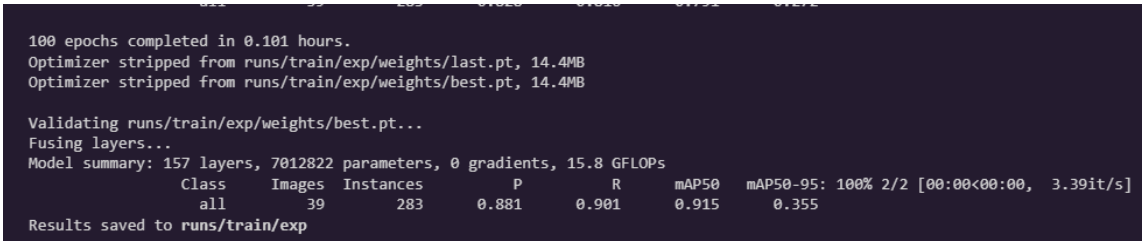
## 4.2 Αποτελέσματα εκπαίδευσης

Στο πρώτο πείραμα που εκτελέστηκε επιλέχθηκαν τα προεκπαιδευμένα βάρη του δικτύου yolov5s.pt τα οποία αυτομάτως μεταφορτώνονται από το επίσημο αποθετήριο. Το δίκτυο έχει εκπαιδευθεί από το dataset MS COCO και χρησιμοποιείται στην παρούσα διαδικασία για την προσαρμογή των βαρών στην εκπαίδευση των νέων δεδομένων. Εφόσον όλες οι παράμετροι έχουν οριστεί σωστά η διαδικασία εκπαίδευσης είναι πλήρως αυτοματοποιημένη προσαρμόζοντας τις 80 κλάσεις εκπαίδευσης στη μοναδική κλάση της παρούσας περίπτωσης. Στην παρακάτω εικόνα φαίνεται χαρακτηριστικά πως μετά από 10 εποχές εκπαίδευσης, ο αλγόριθμός έχει επιτύχει μετρητική mAP50, δηλαδή mean Average Precision με δείκτη IoU 0.5, της τάξης του 10%.




**Εικόνα 4.1:** Map για 10 εποχές εκπαίδευσης.

Όταν η εκπαίδευση του δικτύου ολοκληρωθεί εμφανίζονται τα συνολικά αποτελέσματα όπως στην παρακάτω εικόνα 4.2.




**Εικόνα 4.2:** Τελικά εξαγόμενα αποτελέσματα.

Τα παραπάνω αποτελέσματα εξήχθησαν με βάση την απόδοση του αλγορίθμου στο 20% των δεδομένων που χρησιμοποιήθηκαν για αξιολόγηση, ποσοστό το οποίο ορίστηκε πριν την έναρξη της εκπαίδευσης. Σύμφωνα με τα παραπάνω ο αλγόριθμος πέτυχε στις εικόνες αξιολόγησης τα κάτωθι αποτελέσματα:

| | |
|---|---|
| ***Precision*** | *0.881* |
| ***Recall*** | *0.901* |
| ***mAP50*** | *0.915* |
| ***mAP50-95*** | *0.355* |

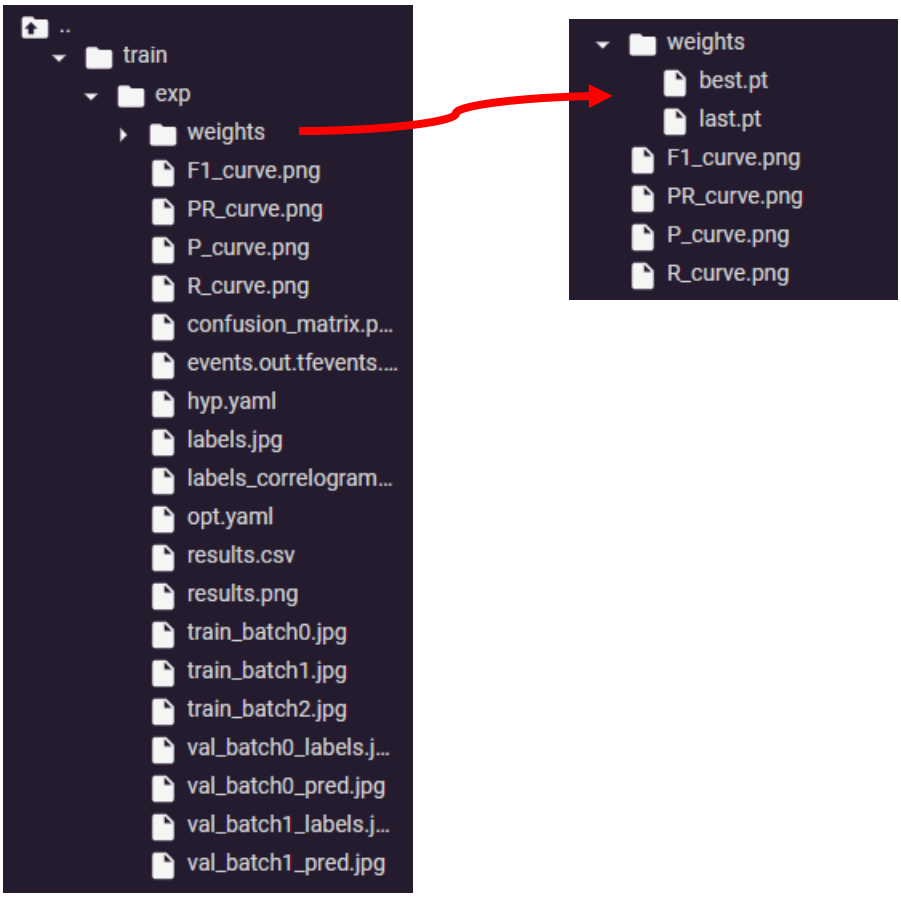
Τα εξαγόμενα αποτελέσματα δείχνουν ότι η χρησιμοποιούμενη μέθοδος πέτυχε δείκτη mean Average Precision για IoU 0.5 της τάξης 0.915. Ο τελευταίος δείκτης (*mAP50-95*) σημαίνει ότι ο αλγόριθμος εξετάζει διαδοχικά τον δείκτη mAP από τιμές IoU 0.5 εώς 0.95 αυξάνοντας κάθε φορά την τιμή κατά 0.05. Δηλαδή ο δείκτης IoU γίνεται σταδιακά αυστηρότερος καθώς προσεγγίζει επικάλυψη 100%. Στην περίπτωση αυτή η τελική τιμή είναι της τάξης του 0.355. Τα παραπάνω αποτελέσματα κρίνονται άκρως ικανοποιητικά, που σημαίνει ότι το δίκτυο πέτυχε τον σκοπό του και μπορεί να αναγνωρίζει με μεγάλο ποσοστό επιτυχίας τα αεροσκάφη.

Με το πέρας της εκπαίδευσης του δικτύου, η μέθοδος εξάγει όλα τα αποτελέσματα αξιολόγησης του δικτύου εντός συγκεκριμένου προκαθορισμένου φακέλου[10]. Συγκεκριμένα, όπως φαίνεται στην εικόνα 4.3 μεταξύ διαφόρων μετρητικών εξάγονται, οι καμπύλες των γραφικών παραστάσεων precision, recall, precision-recall και τα αρχεία που αποτυπώνουν τα βάρη του δικτύου. Τα βάρη του δικτύου αποτελούνται από δύο αρχεία:

[10] *yolov5/runs/train/exp*

- best.pt
- last.pt

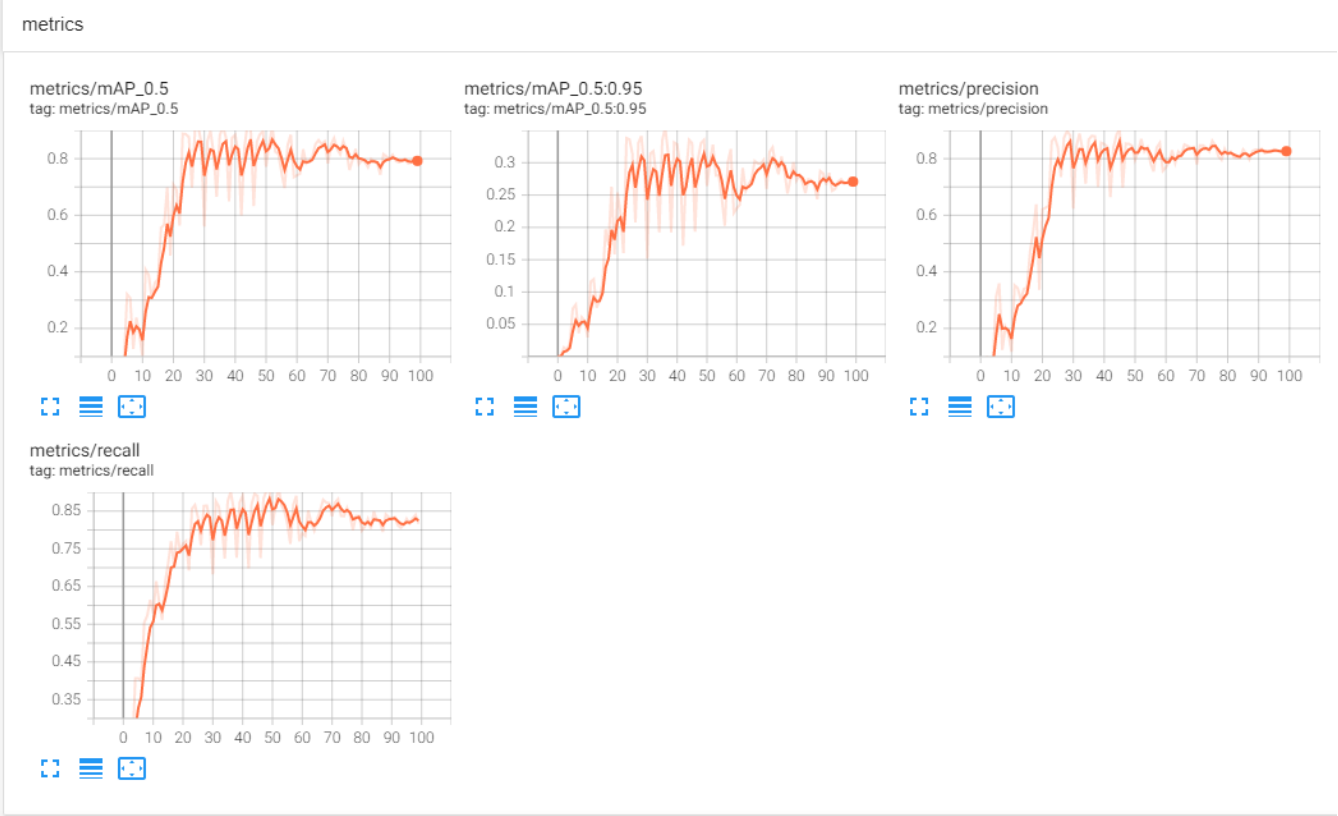
Τα υπόψη αρχεία υποδηλώνουν τιμές βαρών που περιοδικά αποθηκεύονται από τη διαδικασία (best.pt), για να μπορούν να χρησιμοποιηθούν για την συνέχιση της εκπαίδευσης σε μεταγενέστερο χρόνο χωρίς να απαιτείται να ξεκινήσει η διαδικασία από την αρχή. Το αρχείο last.pt υποδηλώνει την τελική μορφή του προσαρμοσμένου πλέον δικτύου στα νέα δεδομένα εκπαίδευσης.




**Εικόνα 4.3:** Αποτελέσματα εκπαίδευσης δικτύου – Βελτιστοποιημένα βάρη του δικτύου

Οι μετρητικές Precision, Recall και mAP όπως παρουσιάστηκαν παραπάνω απεικονίζονται και σε αντίστοιχα διαγράμματα όπως φαίνεται στην παρακάτω εικόνα 4.4. Η εικόνα απεικονίζει τις τιμές των μετρητικών ανά εποχή εκπαίδευσης μέχρι την τελική τιμή των 100 εποχών εκπαίδευσης που είχε τεθεί ως κριτήριο τερματισμού της διαδικασίας. Ως γενική παρατήρηση οι γραφικές παραστάσεις και των τεσσάρων μετρητικών παρουσιάζουν παρεμφερή μορφή, είναι δηλαδή γνησίως αύξουσες. Από 0 εως 30 εποχές η κλίση των παραστάσεων είναι μεγάλη, που σημαίνει ότι η διαδικασία συνεχώς μεταβάλει τις αντίστοιχες τιμές λόγω συνεχούς αναπροσαρμογής των βαρών και επακόλουθη μεταβολή στα αποτελέσματα των προβλέψεων. Οι καμπύλες τείνουν να σταθεροποιηθούν μετά την παρέλευση 20-30 εποχών εκπαίδευσης. Η σταθεροποίηση των τιμών μετά

από σχετικά σύντομες εποχές, μπορεί να ερμηνευτεί εξαιτίας της μίας και μοναδικής κλάσης εκπαίδευσης που είχε ο αλγόριθμος να μάθει. Η σύγκλιση της τιμής μετά τις 30 εποχές σημαίνει ότι οι 100 εποχές εκπαίδευσης ήταν υπεραρκετές για το σύστημα και επομένως περισσότερες εποχές δεν θα προσέφεραν αισθητή βελτίωση στην διαδικασία με τα υπάρχοντα δεδομένα εκπαίδευσης.




**Εικόνα 4.4:** Διαγράμματα Map 0.5, mAP 0.5-0.95, Precision, Recall.

## 4.3 Αποτελέσματα εκπαίδευσης με χρήση παγωμένων επιπέδων

Μετά τον πρώτο πειραματισμό εκτελέστηκε εκ νέου η εκπαίδευση με διαφοροποίηση στα συμμετέχοντα συνελικτικά επίπεδα του δικτύου. Κατά την πρώτη δοκιμή η διαδικασία εκπαίδευσης μετέβαλε ολόκληρη την υποδομή των βαρών ανάμεσα στους νευρώνες του δικτύου. Ωστόσο τίθεται το ερώτημα εάν πράγματι απαιτείται να γίνει κάτι τέτοιο, από τη στιγμή που το αρχικώς χρησιμοποιημένο δίκτυο έχει προεκπαιδευθεί σε ένα πολύ μεγαλύτερο σετ δεδομένων. Στην περίπτωση όπου απαιτείται να προσαρμοστούν τα βάρη σε ένα πολύ μικρό dataset από το αρχικώς χρησιμοποιηθέν για την εκπαίδευση του, η αφαίρεση από την διαδικασία μάθησης των αρχικών επιπέδων θα μειώσει αρκετά το χρονικό διάστημα εκπαίδευσης. Εάν τα νέα δεδομένα εκπαίδευσης δεν προσθέτουν μεγάλη πολυπλοκότητα, όπως στην παρούσα περίπτωση, τότε δεν αναμένονται σημαντικές διαφοροποιήσεις στα τελικά αποτελέσματα και στον χρόνο εκπαίδευσης.

Στη συνέχεια πραγματοποιήθηκε εκ νέου εκπαίδευση παγώνοντας τα 11 πρώτα συνελικτικά επίπεδα. Η τεχνική διαδικασία για να πραγματοποιηθεί αυτό είναι εξαιρετικά απλή καθώς απαιτείται μόνο να τοποθετηθεί στην εκτέλεση του αλγορίθμου ένα flag με τις αντίστοιχες παραμέτρους. Η εξαίρεση των πρώτων επιπέδων σημαίνει ότι τα αντίστοιχα βάρη θα μείνουν ανεπηρέαστα από την διαδικασία, και η αναπροσαρμογή θα γίνει στα εναπομείναντα βάρη των 15 υπόλοιπων επιπέδων του δικτύου. Τα τελικά αποτελέσματα της διαδικασίας παρουσιάζονται στην παρακάτω εικόνα 4.5.

```
100 epochs completed in 0.094 hours.
Optimizer stripped from runs/train/exp2/weights/last.pt, 14.4MB
Optimizer stripped from runs/train/exp2/weights/best.pt, 14.4MB

Validating runs/train/exp2/weights/best.pt...
Fusing layers...
Model summary: 157 layers, 7012822 parameters, 0 gradients, 15.8 GFLOPs
               Class     Images  Instances          P          R      mAP50   mAP50-95: 100% 2/2 [00:00<00:00,  3.36it/s]
                 all         39        283      0.859      0.843      0.867      0.331
Results saved to runs/train/exp2
```

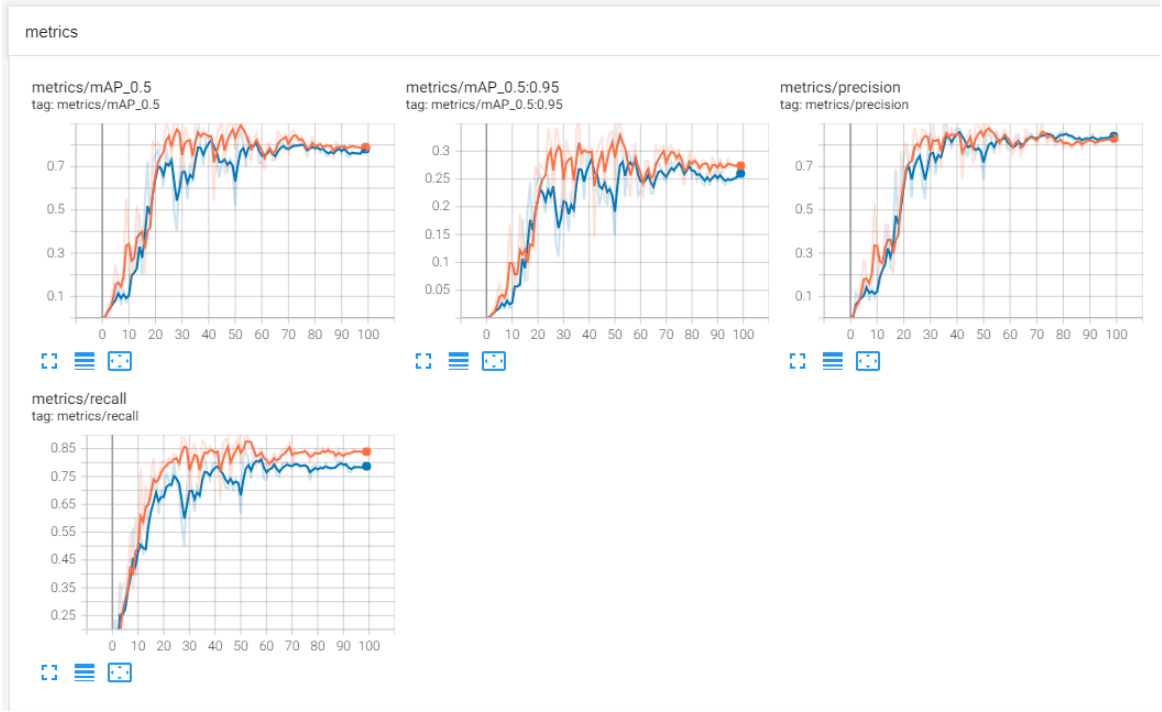
**Εικόνα 4.5:** Τελικά εξαγόμενα αποτελέσματα εκπαίδευσης με παγωμένα τα αρχικά επίπεδα.

Η διαδικασία εκτελέστηκε εντός της δικτυακής υποδομής με εικονικό μηχάνημα Tesla T4 GPU και torch 1.12. Για λόγους σύγκρισης μεταξύ των δύο διαφορετικών διαδικασιών εκπαιδεύσεων στον παρακάτω πίνακα παρουσιάζονται τα αποτελέσματα.

| | **1η εκπαίδευση** | **2η εκπαίδευση** |
|---|---|---|
| **Time (h)** | 0.101 | 0.094 |
| ***Precision*** | *0.881* | 0.859 |
| ***Recall*** | *0.901* | 0.843 |
| ***mAP50*** | *0.915* | 0.867 |
| ***mAP50-95*** | *0.355* | 0.331 |
| ***Εποχές*** | *100* | 100 |

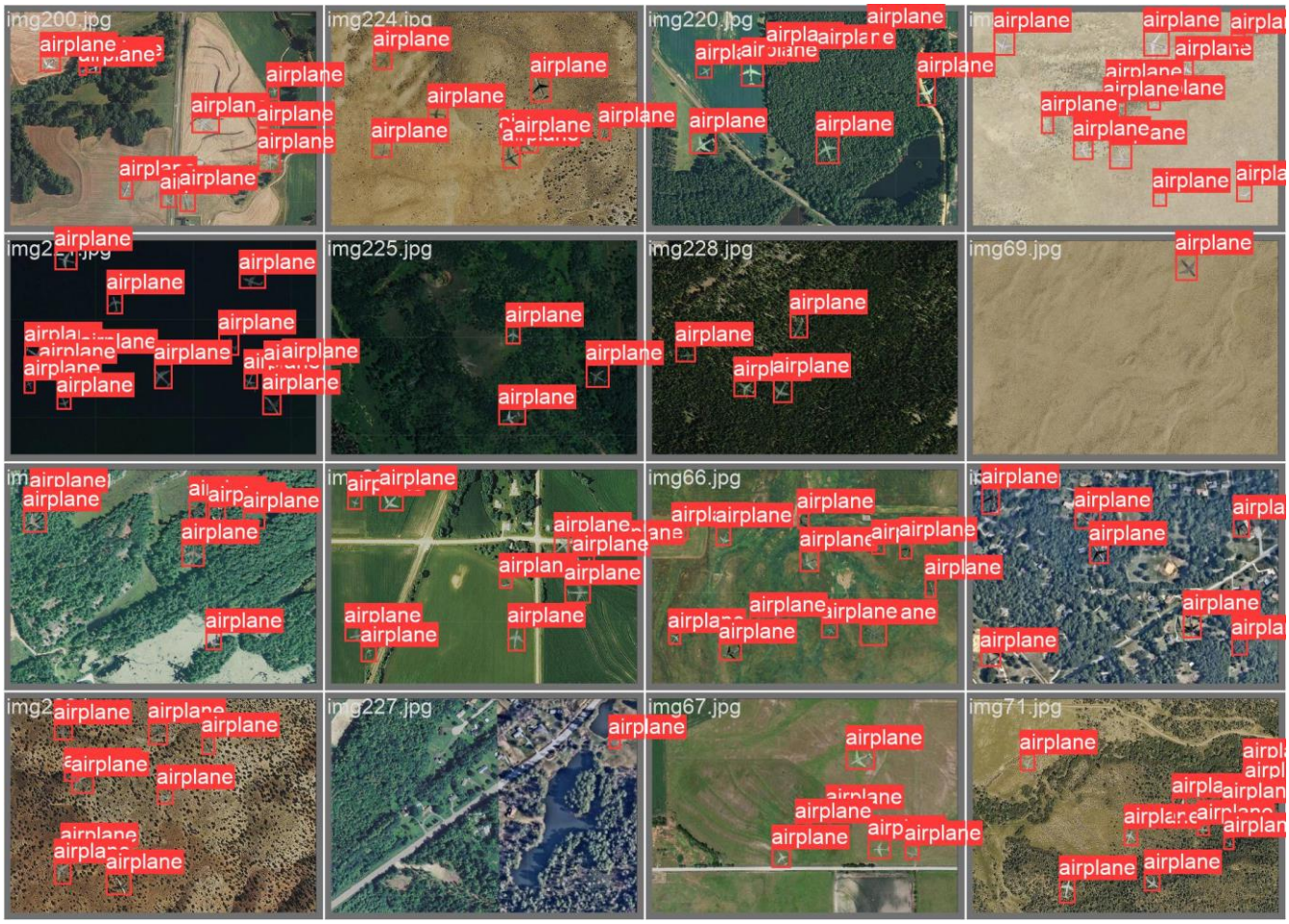
Ο χρόνος εκτέλεσης του αλγορίθμου κατά την δεύτερη δοκιμή μειώθηκε ελαφρώς καθώς από τα 0.101 έπεσε στο 0.094 h. Όπως ήταν αναμενόμενο οι υπόλοιποι δείκτες επίσης μειώθηκαν ελαφρώς, χωρίς ωστόσο να παρατηρούνται σημαντικές διαφοροποιήσεις.

Στα παρακάτω διαγράμματα (Εικόνα 4.16) παρουσιάζονται επίσης στο ίδιο σύστημα αξόνων οι μετρητικές precision, recall και mAP και των δύο παραπάνω δοκιμών. Παρατηρείται πως η εικόνα που παρουσιάστηκε κατά την πρώτη δοκιμή επαναλαμβάνεται και στην δεύτερη. Οι τιμές στην δεύτερη περίπτωση παρουσιάζουν μια μικρή διακύμανση αλλά η γενική τάση των γραφικών παραστάσεων επαναλαμβάνεται. Οι παραστάσεις είναι αύξουσες, παρουσιάζουν μεγάλη κλίση στις πρώτες 30 εποχές ενώ ακολούθως συγκλίνουν προς μια σχετικά υψηλή τιμή καθώς τείνουν στις 100 εποχές. Παρατηρείται επίσης ότι πέραν των 100 εποχών κάθε νέα εποχή δε συνεισφέρει ουσιαστικά στην περαιτέρω βελτιστοποίηση των βαρών.




**Εικόνα 4.6:** Σύγκριση δεικτών μεταξύ των δυο εκπαιδεύσεων.

Ενδεχομένως τα παραπάνω αποτελέσματα να παρουσίαζαν διαφορετική εικόνα εάν το dataset περιελάμβανε περισσότερες κλάσεις εκπαίδευσης. Ως γενική παρατήρηση οι δυο δοκιμές έχουν παρεμφερή εικόνα με ελάχιστα καλύτερες τιμές να παρουσιάζονται στην αναπροσαρμογή όλων των βαρών του δικτύου σε σύγκριση με την αναπροσαρμογή στα τελευταία μόνο επίπεδα.




**Εικόνα 4.7:** Ενδεικτικά αποτελέσματα

Τέλος, στην παραπάνω εικόνα 4.7 παρουσιάζονται ενδεικτικά αποτελέσματα από τις εικόνες που χρησιμοποιήθηκαν για την αξιολόγηση του δικτύου. Με βάση τα παραπάνω αποτελέσματα εξήχθησαν οι δείκτες που παρουσιάστηκαν παραπάνω.

## 4.4 Αποτελέσματα σε ανεξάρτητες εικόνες του ίδιου dataset

Στα μέχρι τώρα πειράματα έχει φανεί ότι η συμπεριφορά του αλγορίθμου είναι παρεμφερής στην χρήση παγωμένων ή όχι επιπέδων του δικτύου. Στην παρούσα παράγραφο δοκιμάστηκε η αποδοτικότητα του δικτύου σε νέες εικόνες από το ίδιο αρχικό dataset των κατασκευασμένων εικόνων. Παρόλο που οι δύο προηγούμενες εκπαιδεύσεις εξήγαγαν παρεμφερή αποτελέσματα, προτιμήθηκε να χρησιμοποιηθεί το δίκτυο που είχε εκπαιδευθεί με το σύνολο των επιπέδων καθώς οι δείκτες αξιολόγησης ήταν ελαφρώς βελτιωμένοι.

Στη παρούσα παράγραφο τον πειραματισμό διατηρήθηκε το ίδιο σετ δεδομένων εκπαίδευσης και ενισχύθηκε το σετ των δεδομένων αξιολόγησης με 100 επιπλέον διαφορετικές εικόνες. Τα αποτελέσματα όπως εξήχθησαν από τον αλγόριθμο παρουσιάζονται στην παρακάτω εικόνα 4.8.

```
100 epochs completed in 0.120 hours.
Optimizer stripped from runs/train/exp/weights/last.pt, 14.4MB
Optimizer stripped from runs/train/exp/weights/best.pt, 14.4MB

Validating runs/train/exp/weights/best.pt...
Fusing layers...
Model summary: 157 layers, 7012822 parameters, 0 gradients, 15.8 GFLOPs
               Class     Images  Instances          P          R      mAP50   mAP50-95: 100% 2/2 [00:00<00:00,  2.68it/s]
                 all        139       2831      0.884      0.919      0.929      0.367
Results saved to runs/train/exp
```

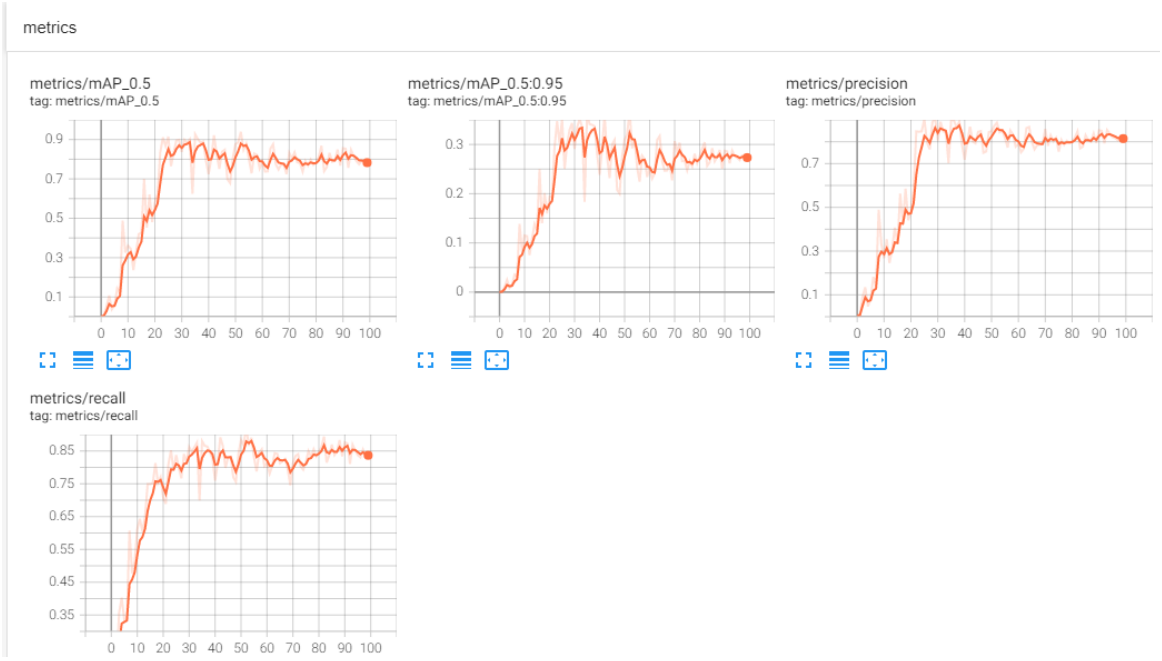
**Εικόνα 4.8:** Αποτελέσματα επιπλέον εικόνων αξιολόγησης.

Για λόγους σύγκρισης με τα αποτελέσματα των υπολοίπων πειραματισμών, συγκεντρωτικά παρουσιάζονται στον παρακάτω πίνακα.

| | **1η εκπαίδευση** | **2η εκπαίδευση** | **Αποτελέσματα Αξιολόγησης** |
|---|---|---|---|
| **Time (h)** | 0.101 | 0.094 | |
| ***Precision*** | *0.881* | 0.859 | **0.884** |
| ***Recall*** | *0.901* | 0.843 | **0.919** |
| ***mAP50*** | *0.915* | 0.867 | **0.929** |
| ***mAP50-95*** | *0.355* | 0.331 | **0.367** |
| ***Εποχές*** | *100* | 100 | |

**Πίνακας 4-1:** Πίνακας συγκεντρωτικών αποτελεσμάτων.α

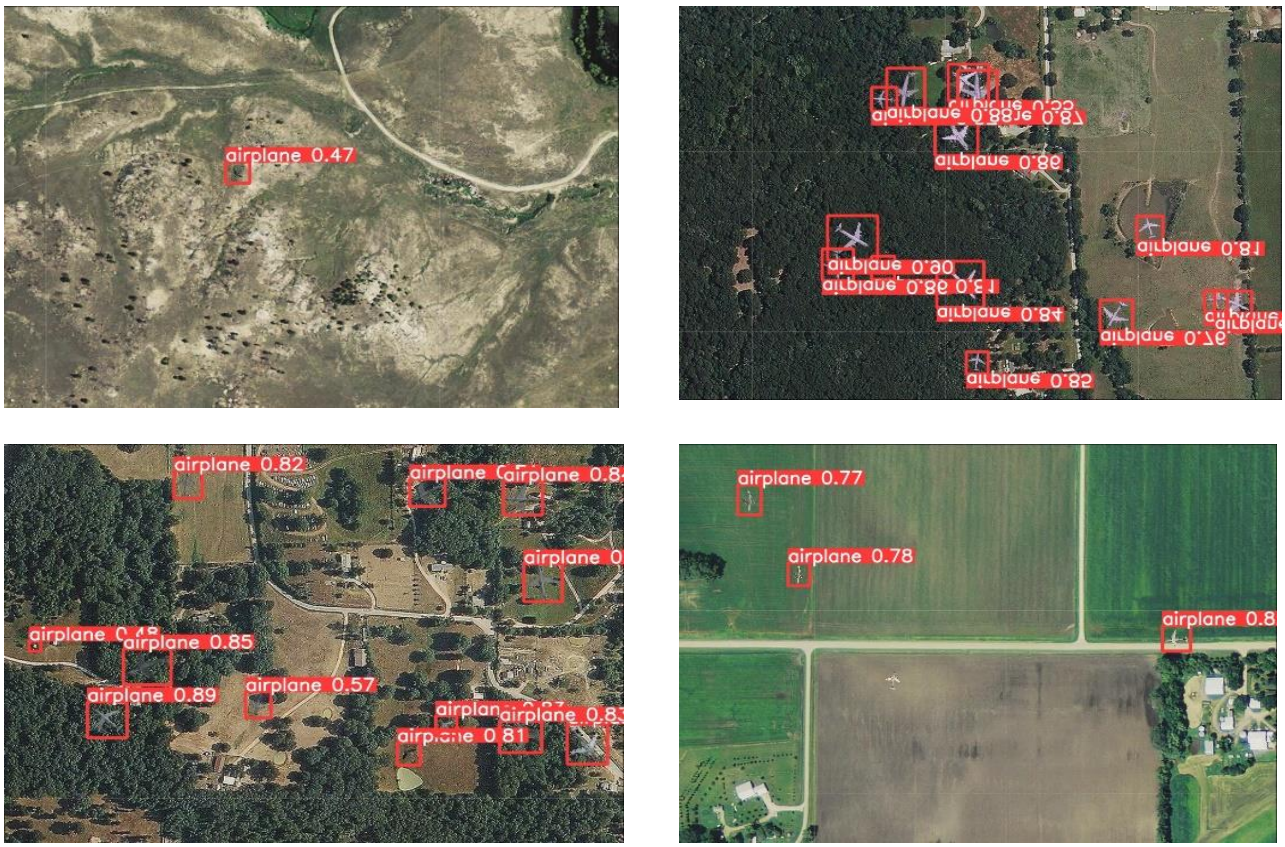
Όπως αναμενόταν οι τιμές έρχονται σε πλήρη αρμονία με τα αποτελέσματα των προηγούμενων δοκιμών. Οι τιμές τείνουν να προσεγγίσουν, με ελάχιστη διαφοροποίηση, τις αντίστοιχες τιμές του πρώτου πειραματισμού που εκτελέστηκε. Επιπλέον στην παρακάτω εικόνα 4.9 παρουσιάζονται οι γραφικές παραστάσεις των μετρητικών που εξάγονται με το πέρας της διαδικασίας.




**Εικόνα 4.9:** Γραφικές παραστάσεις μετρητικών νέου πειραματισμού.

Όπως ήταν αναμενόμενο οι παραπάνω γραφικές παραστάσεις έχουν την ίδια συμπεριφορά με τα αποτελέσματα των προηγούμενων δοκιμών. Η φύση των καμπυλών βρίσκεται σε συμφωνία με τις καμπύλες των προγενέστερων δοκιμών. Τα παραπάνω αποτελέσματα είναι αναμενόμενα διότι οι νέες εικόνες αξιολόγησης δεν παρουσιάζουν μεγάλη παραλακτικότητα σε κλίμακα, φωτεινότητα, χωρική ανάλυση και δυσκολία εντοπισμού. Η μεγάλη ομοιότητα των δεδομένων οδηγεί σε ίδια συμπεριφορά στους εξαγόμενους δείκτες.

Ενδεικτικά αποτελέσματα διαφόρων ανιχνεύσεων παρουσιάζονται στην παρακάτω εικόνα 4.10. Παρατηρείται ότι ο αλγόριθμος έχει εκπαιδευθεί ικανοποιητικά, καθώς στις περισσότερες των περιπτώσεων έχει αναγνωρίσει το επιθυμητό αντικείμενο. Σημαντικές περιπτώσεις αστοχίας του αλγορίθμου γενικά δεν παρατηρούνται.




**Εικόνα 4.10**: Ενδεικτικά αποτελέσματα ανιχνεύσεων.


**Εικόνα 4.11:** Ενδεικτικό αποτέλεσμα ανίχνευσης

Στην παραπάνω εικόνα 4.11 απεικονίζεται ένα παράδειγμα ανίχνευσης αεροσκάφους με μεγάλο βαθμό δυσκολίας. Το ανιχνευμένο αντικείμενο έχει εξαιρετικά μικρό μέγεθος σε σχέση με τις διαστάσεις της εικόνας, το σχήμα του είναι δυσδιάκριτο με γυμνό μάτι και η φωτεινότητά του επίσης δεν παραπέμπει σε σαφώς περιγεγραμμένο αντικείμενο. Παρόλο των παραπάνω περιορισμών ο αλγόριθμος κατάφερε να εντοπίσει το συγκεκριμένο απαιτητικό αντικείμενο με έναν δείκτη εμπιστοσύνης με τιμή 0.47. Επιπλέον δεν παρατηρούνται στην εικόνα False Positives (δηλαδή λάθος αναγνωρισμένα αντικείμενα ως αεροπλάνα), παρόλο που η εικόνα διαθέτει πλήθος αντικειμένων (πχ διάσπαρτοι θάμνοι, δέντρα, βράχοι) τα οποία ομοιάζουν σε σχήμα, μέγεθος και χρώμα με το μοναδικό αεροπλάνο που αποτυπώνεται στην εικόνα. Το παραπάνω παράδειγμα αποτελεί περίπτωση κατά την οποία ο αλγόριθμος πέτυχε ικανοποιητικά την αναγνώριση του αντικειμένου.


**Εικόνα 4.12:** Ενδεικτικά αποτελέσματα ανίχνευσης – μη ανίχνευσης.

Στην παραπάνω εικόνα 4.12 απεικονίζεται ένα παράδειγμα όπου ο αλγόριθμος κατάφερε να εντοπίσει τα 3 από τα 4 απεικονιζόμενα αντικείμενα. Όλα τα αντικείμενα παρουσιάζουν σχετική ομοιότητα ως προς το σχήμα, μέγεθος και χρώμα ενώ αλλάζει μόνο ο προσανατολισμός των αντικειμένων. Το αντικείμενο που δεν αναγνωρίστηκε εμφανίζεται εντός μωβ πλαισίου. Το συγκεκριμένο παράδειγμα αποτελεί περίπτωση αστοχίας ενός σχετικά απλού αντικειμένου με σαφώς περιγεγραμμένο σχήμα και χωρίς αποκρύψεις. Πιθανή εξήγηση της

συμπεριφοράς αυτή αποτελεί το γεγονός της χρησιμοποίησης ενός σχετικά μικρού αριθμού εικόνων εκπαίδευσης (400 εικόνες) οι οποίες φαίνεται πως δεν είναι αρκετές ώστε να περιγράψουν όλες τις δυνατές περιπτώσεις.

## 4.5 Αποτελέσματα σε ΔΕ υψηλής ανάλυσης

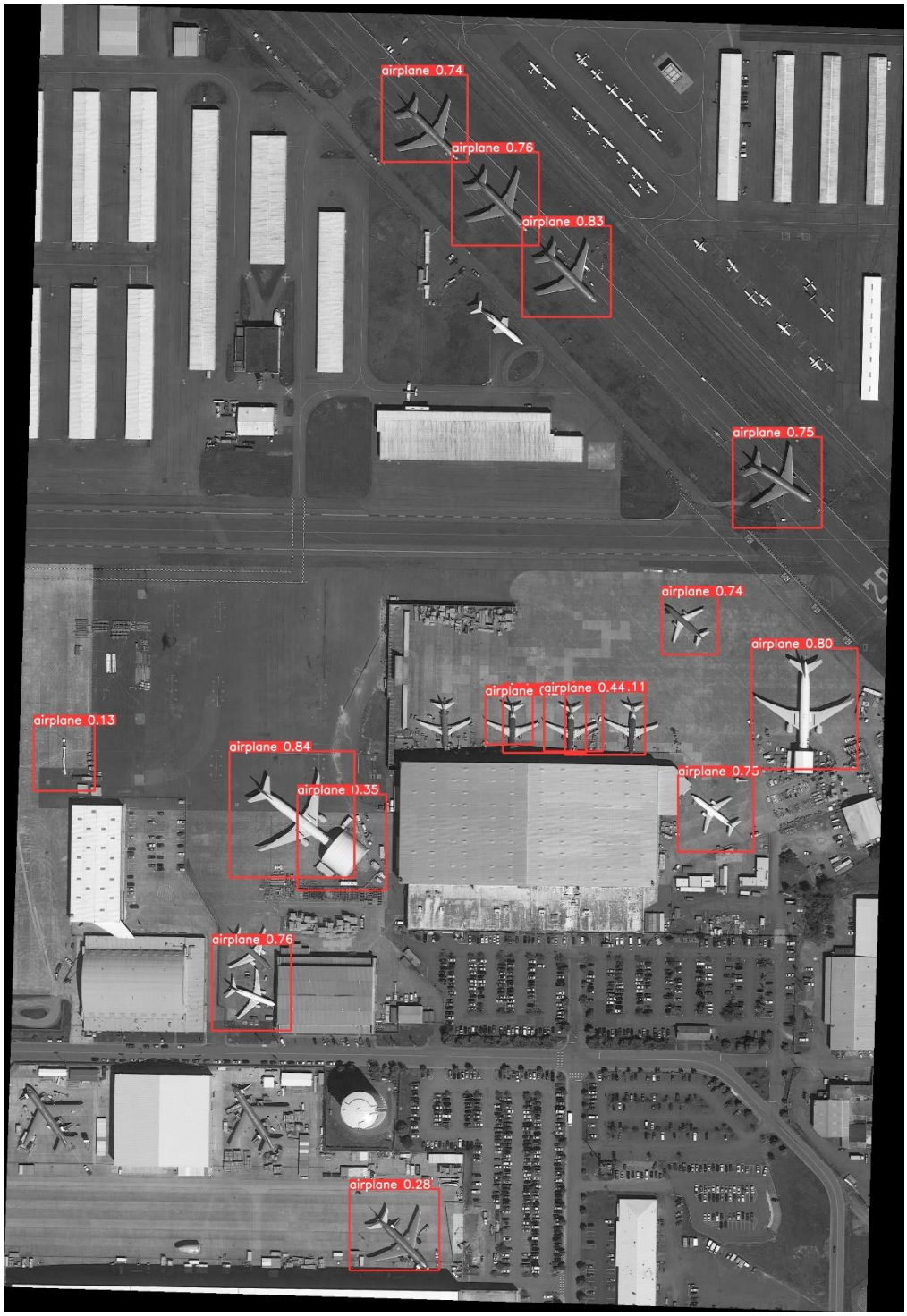
Στους μέχρι τώρα πειραματισμούς εκπαιδεύτηκε ένα δίκτυο με δεδομένα εκπαίδευσης που προέρχονται από τεχνητές εικόνες στις οποίες έχουν εμφυτευτεί ηλεκτρονικά τα προς αναγνώριση αντικείμενα (αεροσκάφη). Η αποδοτικότητα του αλγορίθμου ελέγχθηκε με δεδομένα τα οποία προέρχονται από το ίδιο dataset και σύμφωνα με τις εξαγόμενες μετρητικές, τα αποτελέσματα εκτιμήθηκαν ικανοποιητικά. Ωστόσο τίθεται το ερώτημα ποια θα ήταν η συμπεριφορά του δικτύου εάν δοκιμαζόταν σε ένα πιο αληθινό σενάριο.

Στην παρούσα παράγραφο παρουσιάζονται τα αποτελέσματα της χρήσης του εκπαιδευμένου δικτύου σε τρεις υπέρ-υψηλής ανάλυσης Δορυφορικές Εικόνες. Συγκεκριμένα δοκιμάστηκε το δίκτυο που εκπαιδεύθηκε με το παραπάνω dataset σε τρία πραγματικά σενάρια (real case scenarios) τα οποία απεικονίζουν τρεις διαφορετικές δορυφορικές σκηνές αεροδρομίων, στα οποία βρίσκονται προσγειωμένα διαφόρων τύπων αεροσκάφη. Οι εικόνες που χρησιμοποιήθηκαν αποτελούνται από:

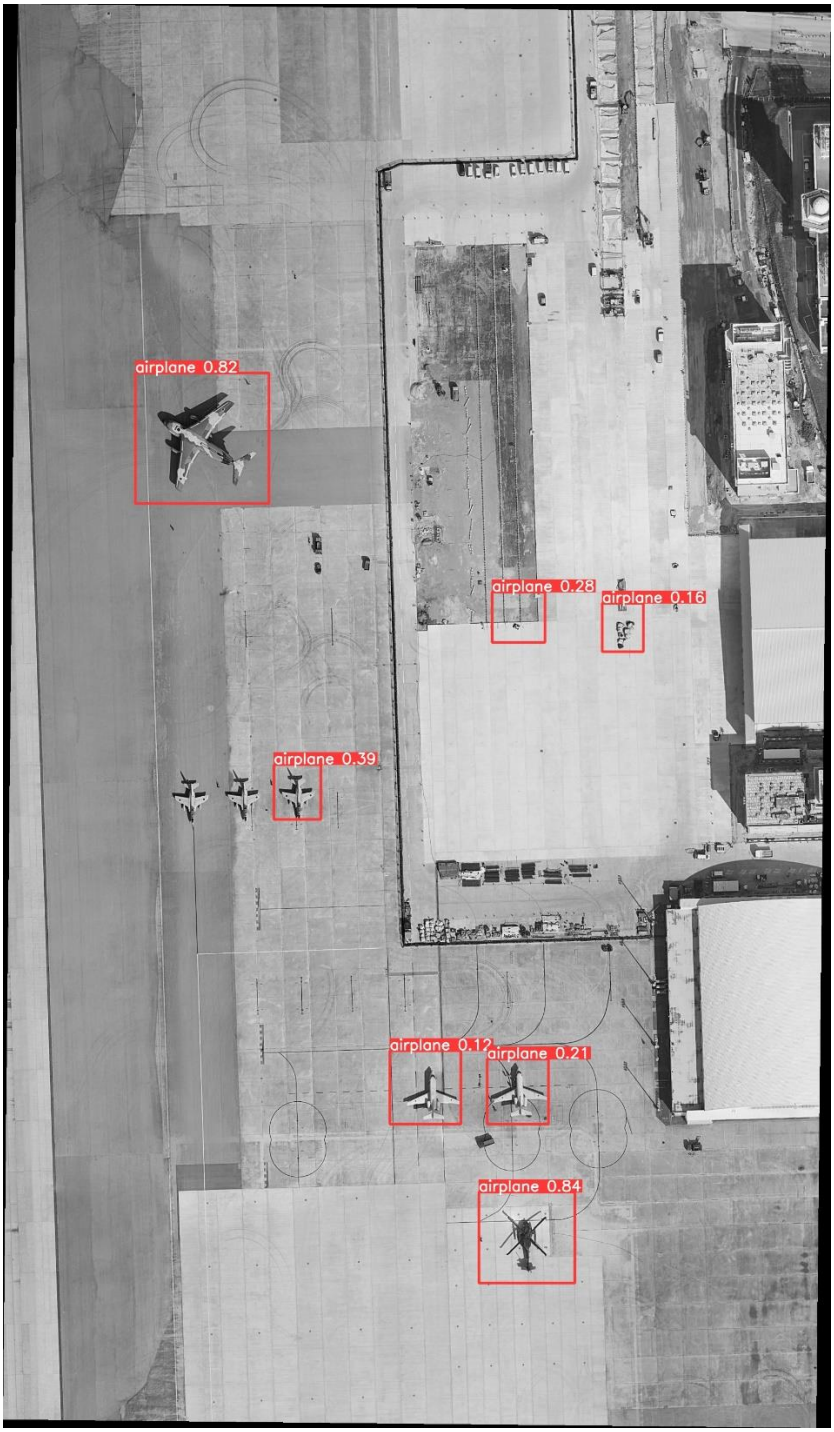
- Μία παγχρωματική εικόνα 30 cm που απεικονίζει πολιτικό αεροδρόμιο
- Μία παγχρωματική εικόνα 30cm που απεικονίζει στρατιωτικό αεροδρόμιο
- Μια pansharpened εικόνα 30 cm που απεικονίζει στρατιωτικό αεροδρόμιο

Στην παρακάτω εικόνα 4.13 απεικονίζεται το αποτέλεσμα της αναγνώρισης που προέκυψε από τη χρήση της παγχρωματικής εικόνας σε πολιτικό αεροδρόμιο. Παρατηρούνται τα κάτωθι:

- Στην εικόνα απεικονίζονται αεροσκάφη διαφόρων κλιμάκων αντικειμένου (μεγάλα, μικρά, μεσαία).
- Ο αλγόριθμος αναγνώρισε συνολικά τα δεκαέξι από τα δεκαεφτά μεγάλα και μεσαία αεροσκάφη που βρίσκονται σταθμευμένα – 16 True Positive.
- Ένα μεγάλο αεροσκάφος δεν αναγνωρίστηκε καθόλου.
- Δύο αντικείμενα λανθασμένα αναγνωρίστηκαν ως αεροσκάφη – 2 False Positive.
- Η σειρά των μικρής κλίμακας αεροσκαφών που βρίσκονται πάνω και δεξιά της εικόνας δεν έχουν αναγνωριστεί καθόλου.




**Εικόνα 4.13:** Παγχρωματική ΔΕ 30 cm πολιτικού αεροδρομίου.




**Εικόνα 4.14:** Παγχρωματική ΔΕ 30 cm στρατιωτικού αεροδρομίου.

Στην παραπάνω εικόνα 4.14 απεικονίζεται το αποτέλεσμα της αναγνώρισης που προέκυψε από τη χρήση της παγχρωματικής εικόνας σε στρατιωτικό αεροδρόμιο. Παρατηρούνται τα κάτωθι:

- Η εικόνα απεικονίζει ένα μεγάλο αεροσκάφος και 5 μικρά.
- Ο αλγόριθμος αναγνώρισε συνολικά τα 4 από τα 6 αεροσκάφη που βρίσκονται σταθμευμένα – 4 True Positive.

- Αναγνωρίστηκαν ένα αεροσκάφος μεγάλης κλίμακας και τρία αεροσκάφη μικρότερης κλίμακας.
- Δεν παρατηρούνται αντικείμενα που λανθασμένα αναγνωρίζονται ως αεροσκάφη – 0 False Positive.
- Δύο αεροσκάφη μικρότερης κλίμακας από τα πέντε δεν αναγνωρίστηκαν καθόλου.
- Η εικόνα απεικονίζει έταιρο πτητικό μέσο (ελικόπτερο) το οποίο ορθώς δεν αναγνωρίστηκε παρόλο που το σχήμα του ενδεχομένως παραπέμπει ως αεροσκάφος.


**Εικόνα 4.15:** Pansharpened ΔΕ 30 cm στρατιωτικού αεροδρομίου

Στην παραπάνω εικόνα 4.15 απεικονίζεται το αποτέλεσμα της αναγνώρισης που προέκυψε από τη χρήση της pansharpened εικόνας σε στρατιωτικό αεροδρόμιο. Η εικόνα δημιουργήθηκε χρησιμοποιώντας μια αρχική παγχρωματική χωρικής ανάλυσης 0.3 μ., με παρεμβολή χρώματος από την αντίστοιχη πολυφασματική εικόνα χωρικής ανάλυσης 1.2 μ. Η προκύπτουσα εικόνα είναι μια φυσική έγχρωμη σύνθετη χωρικής ανάλυσης 0.3 μ η οποία τροφοδοτήθηκε στο εκπαιδευμένο δίκτυο. Παρατηρούνται τα κάτωθι:

- Η εικόνα περιλαμβάνει πλήθος αεροσκαφών μικρού και μεσαίου μεγέθους.
- Αναγνωρίστηκαν ορθώς συνολικά εννιά αεροσκάφη μεσαίου μεγέθους από τα δέκα που εμφανίζονται στην εικόνα – 9 True Positive.
- Παρατηρείται ένα αντικείμενο που λανθασμένα αναγνωρίστηκε ως αεροσκάφος – 1 False Positive.
- Παρατηρείται ένα αεροσκάφος μεσαίου μεγέθους που δεν αναγνωρίστηκε καθόλου.
- Όλα τα αεροσκάφη μικρού μεγέθους που εμφανίζονται στην πάνω δεξιά μεριά της εικόνας δεν έχουν αναγνωριστεί καθόλου.

Από τις παραπάνω δοκιμές προκύπτει ότι η αποδοτικότητα του αλγορίθμου είναι ικανοποιητική καθόσον δεν παρατηρούνται σημαντικές αστοχίες. Παρόλο που το δίκτυο είναι εκπαιδευμένο σε ένα εντελώς διαφορετικό dataset, ο αλγόριθμος πέτυχε να αναγνωρίσει σε μεγάλο βαθμό τα αεροσκάφη μεσαίου και μεγάλου μεγέθους. Οι αστοχίες παρατηρήθηκαν στα μικρού μεγέθους αεροσκάφη όπου ο αλγόριθμος δεν κατάφερε να τα αναγνωρίσει καθόλου.

Ως γενική παρατήρηση εξάγεται ότι το εκπαιδευμένο δίκτυο είναι κατάλληλο να αναγνωρίζει τα αντικείμενα σε ένα εύρος υποβάθρων, φωτεινότητας, χωρικής ανάλυσης και δυσκολίας εντοπισμού, αλλά αδυνατεί να αναγνωρίσει τα αντικείμενα με έντονη διαφοροποίηση στην κλίμακα. Εκτιμάται ότι το πρόβλημα μπορεί να λυθεί με την χρήση δεδομένων εκπαίδευσης που να καλύπτουν και την παράμετρο της κλίμακας των απεικονιζόμενων αντικειμένων.

# 5. ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΠΡΟΟΠΤΙΚΕΣ

## 5.1 Γενικά Συμπεράσματα

Στο κεφάλαιο αυτό παρουσιάζονται γενικά και ειδικά συμπεράσματα που προέκυψαν από την ενασχόληση, την μελέτη και τις δοκιμές πάνω στην μεθοδολογία Yolo5 για την αναγνώριση του συγκεκριμένου αντικειμένου.

Οι τεχνικές Μηχανικής Μάθησης είναι πλέον ευρύτατα διαδομένες και διαθέσιμες στο ευρύ κοινό για πλήθος εφαρμογών. Στην παρούσα εργασία χρησιμοποιήθηκε μια μεθοδολογία από ένα open source project το οποίο βρίσκεται διαθέσιμο σε αποθετήριο στο github. Η χρήση του συγκεκριμένου project δεν περιορίζεται αμιγώς σε project όρασης υπολογιστών, αλλά μπορεί να επεκταθεί στην κάλυψη πλήθος εργασιών που να προσαρμόζονται κάθε φορά στις ανάγκες του χρήστη. Στην συγκεκριμένη περίπτωση χρησιμοποιήθηκε το υπόψη project για την υποστήριξη εξαγωγής ενός συγκεκριμένου αντικειμένου με απώτερο σκοπό στην υποβοήθηση της διαδικασίας IMINT.

Η εξέλιξη στην λογισμική υποδομή των μεθόδων Μηχανικής Μάθησης είναι επίσης εντυπωσιακή. Κατά τα αρχικά στάδια ο χρήστης έπρεπε να προγραμματίσει μόνος του τις διαδικασίες που επιτρέπουν την εκπαίδευση ενός δικτύου, γεγονός το οποίο αποτελούσε τροχοπέδη στην περαιτέρω εξέλιξη και προώθηση στο ευρύ κοινό των μεθόδων. Ακολούθως, εμφανίστηκαν βιβλιοθήκες οι οποίες αυτοματοποιούν τις διαδικασίες και επιτρέπουν στον χρήστη να ασχοληθεί περισσότερο με τον συγκεκριμένο πρόβλημα παρά με την υλοποίηση του. Βιβλιοθήκες και Framework όπως Tensorflow, Pytorch, Keras είναι πλέον αναπόσπαστο κομμάτι σε κάθε πρόβλημα που μπορεί να επιλυθεί με μεθόδους Μηχανικής Μάθησης και Νευρωνικών Δικτύων.

Εκτός από την εντυπωσιακή ανάπτυξη των βιβλιοθηκών, εντυπωσιακή ήταν επίσης και η ανάπτυξη της υποδομής που επιτρέπουν στον χρήστη να χρησιμοποιεί τις βιβλιοθήκες. Το μειονέκτημα της χρήσης των βιβλιοθηκών σε ανεξάρτητο περιβάλλον, ήταν ο πολύ αργός χρόνος που απαιτείται για την εκπαίδευση του δικτύου, χρόνος ο οποίος ενδεχομένως να φτάνει σε τάξη μεγέθους από ώρες μέχρι μέρες και που εξαρτάται από τα χαρακτηριστικά και τις επιδόσεις της εκάστοτε υλισμικής υποδομής (Hardware). Πλέον το πρόβλημα έχει σε μεγάλο βαθμό ξεπεραστεί, καθώς μεγάλες εταιρείες στο χώρο έχουν δημιουργήσει και διαθέτουν στο ευρύ κοινό, με κάποιους περιορισμούς, Cloud based πλατφόρμες με χρήση online υποδομής βέλτιστων προδιαγραφών.

Στην παρούσα εργασία, όπως περιγράφηκε σε ανωτέρω παραγράφους, έγινε εκτεταμένη χρήση μιας cloud based πλατφόρμας για την επίλυση του συγκεκριμένου προβλήματος, και δεν απαιτήθηκε η τοπική εγκατάσταση βιβλιοθηκών. Η εμπειρία από την χρήση της πλατφόρμας εκτιμάται θετική καθώς, πέρα από την ευκολία που παρέχει στην εγκατάσταση των ιδιαίτερα πολύπλοκων βιβλιοθηκών, δεν παρατηρήθηκαν δυσκολίες στην χρήση της. Το περιβάλλον εργασίας, η διαδικασία αποθήκευσης, η επιλογή των παραμέτρων (πχ χρήση

GPU, CPU, TPU) καθώς και ο κειμενογράφος συγγραφής του κώδικα είναι εύκολα κατανοητά και χρησιμοποιήσιμα από τον μέσο χρήστη. Με τον τρόπο αυτό δαπανάται περισσότερος χρόνος στην επίλυση του προβλήματος και λιγότερος στις ενέργειες που απαιτούνται για τα διαδικαστικά.

Τέλος η διαδικασία εκπαίδευσης του δικτύου βασίστηκε στο αποθετήριο του συγκεκριμένου project ακολουθώντας πιστά τα βήματα όπως περιγράφονται στο παρεχόμενο documentation. Η διαδικασία σε γενικές γραμμές είναι πλήρως αυτοματοποιημένη και εύκολα επεκτάσιμη και προσαρμόσιμη στις εκάστοτε ανάγκες του χρήστη.

## 5.2 Ειδικά Συμπεράσματα

Σε γενικές γραμμές το συγκεκριμένο dataset μπορεί να χαρακτηριστεί απαιτητικό εξαιτίας κυρίως της δυσκολίας ανεύρεσης μεγάλου όγκου εικόνων που να αποτυπώνουν το προς αναγνώριση αντικείμενο. Οι κατασκευασμένες εικόνες που χρησιμοποιήθηκαν αποσκοπούν στην κάλυψη αυτής της δυσκολίας, γι’ αυτό τα αεροσκάφη έχουν τοποθετηθεί σε διάφορους προσανατολισμούς και σε ένα ευρύ φάσμα υποβάθρων με σκοπό να καλύψουν όσο το δυνατόν περισσότερες επιλογές.

Με γνώμονα το παραπάνω η αποδοτικότητα του αλγορίθμου, όπως αποδείχθηκε από τις εξαγόμενες μετρητικές, ήταν αρκετά ικανοποιητική στα δεδομένα αξιολόγησης του ίδιου dataset. Τα εξαγόμενα διαγράμματα συμπεραίνεται ότι οι 100 εποχές εκπαίδευσης είναι υπεραρκετές για το συγκεκριμένο dataset καθώς οι μετρητικές precision και recall τείνουν να σταθεροποιούνται μετά από 30 εποχές.

Η εκπαίδευση του δικτύου με χρήση παγωμένων ή όχι επιπέδων φαίνεται ότι δεν βελτιώνει ουσιαστικά την αποδοτικότητα καθώς οι δείκτες είναι παρεμφερείς και στις δύο περιπτώσεις. Ο χρόνος εκπαίδευσης είναι ελαφρώς μικρότερος στη χρήση παγωμένων επιπέδων. Εκτιμάται ότι ο χρόνος και η αποδοτικότητα θα διέφεραν σε πιο απαιτητικό dataset με περισσότερα δεδομένα και περισσότερες κλάσεις εκπαίδευσης.

Η αποδοτικότητα του αλγορίθμου σε πραγματικές συνθήκες απέδωσε αρκετά ικανοποιητικά, με μοναδική σημαντική αστοχία την αδυναμία εντοπισμού μικρής κλίμακας αεροσκαφών. Εκτιμάται ότι η χρήση αντίστοιχων δεδομένων που να καλύπτει τη παράμετρο της κλίμακας θα αντιμετωπίσει το πρόβλημα. Ως γενική παρατήρηση η εκπαίδευση του δικτύου σε διαφορετικό dataset και η ικανότητα του να αναγνωρίζει τα αντικείμενα σε πραγματικές εικόνες αποδεικνύουν την ανθεκτικότητα της μεθόδου.

Τέλος, αναφέρεται ότι απαιτούνται αρκετές εργατοώρες για την δημιουργία των Annotation εκπαίδευσης. Στο συγκεκριμένο dataset δημιουργήθηκαν χειροκίνητα όλα τα labels που αφορούσαν τα δεδομένα εκπαίδευσης και αξιολόγησης στο σύνολο των εικόνων. Η διαδικασία θα ήταν ακόμα περισσότερο χρονοβόρα εάν σε κάθε εικόνα υπήρχαν περισσότερα από μια κλάση εκπαίδευσης

του δικτύου. Το πρόβλημα της διαμόρφωσης των labels με τρόπο που απαιτεί η διαδικασία Yolo5 επιλύθηκε με χρήση λογισμικού, διαφορετικά ο χρόνος δημιουργίας κατάλληλης μορφής annotation θα ήταν ακόμη περισσότερος.

## 5.3 Μελλοντική Επέκταση

Στην παρούσα εργασία αποπειράθηκε να στηθεί μια μεθοδολογία η οποία θα αναγνωρίζει αεροσκάφη μέσα από Δορυφορικές Εικόνες υψηλής ανάλυσης. Λαμβάνοντας υπόψη τα αποτελέσματα εκτιμάται ότι η εργασία πέτυχε το σκοπό της σε μεγάλο ποσοστό. Στην παρούσα παράγραφο θα αναφερθούν κάποιες προτάσεις για μελλοντικές επεκτάσεις.

Αρχικά αναφέρεται ότι απώτερος σκοπός της εργασίας είναι να δημιουργηθεί ένας πυρήνας ώστε μελλοντικά να μπορεί να επεκταθεί στις υπόλοιπες κατηγορίες της διαδικασίας IMINT. Έχοντας σαν απαρχή την εκπαίδευση μιας κλάσης, με αντίστοιχη διαδικασία μπορεί να επιτευχθεί η εκπαίδευση των υπολοίπων κλάσεων.

Το σύνολο της εργασίας υλοποιήθηκε εντός ψηφιακού παρόχου, ο οποίος όμως, πέρα από τα αναμφισβήτητα πλεονεκτήματα, παρέχει τις υπηρεσίες του δωρεάν για πεπερασμένο χρονικό διάστημα. Επομένως η χρήση του αλγορίθμου εκτός cloud περιβάλλοντος, θα δώσει την ευελιξία της χρησιμοποίησης χωρίς περιορισμούς.

Μετά την τοπική εγκατάσταση ο αλγόριθμος μπορεί να επεκταθεί με νέα δεδομένα εκπαίδευσης ώστε να αναγνωρίζει και επιπλέον κλάσεις της διαδικασίας IMINT. Επιπλέον, με χρήση κατάλληλων δεδομένων η κάθε κλάση μπορεί να ταξινομηθεί σε επιπλέον κατηγοριοποίηση ώστε να επιμεριστεί περαιτέρω (πχ να αναγνωρίζει και το συγκεκριμένο είδος αεροσκάφους ή κάποιου άλλου αντικειμένου).

Η διαδικασία που περιγράφηκε παραπάνω μπορεί να αποτελέσει την αρχή για την ανάπτυξη μιας ολοκληρωμένης εφαρμογής η οποία πέρα από την αναγνώριση των αντικειμένων, θα μπορεί να αποτελεί ένα σημαντικό εργαλείο ανάλυσης και λήψης αποφάσεων. Προς αυτή την κατεύθυνση οι δυνατότητες του λογισμικού θα μπορούν να περιλαμβάνουν κάποιες από τις κάτω προτάσεις:

- Επέκταση της διαδικασίας αναγνώρισης ώστε να εξάγονται μετρητικά στοιχεία από τα εξαγόμενα αντικείμενα όπως συντεταγμένες, μήκη, πλάτη, αζιμούθια.
- Εισαγωγή της χωρικής διάστασης στην διαδικασία ώστε ο αλγόριθμος να αναγνωρίζει το αντικείμενο αλλά και να το συσχετίζει με τη θέση του στον χώρο. Παράδειγμα μπορεί να αποτελεί η εξαγωγή μηνύματος στην περίπτωση που δεν ανιχνεύεται κάποιο αντικείμενο (πχ σταθμευμένο πλοίο) ενώ συνήθως στην ίδια γεωγραφική περιοχή εντοπίζεται κατά κάποια παρελθοντική στιγμή.
- Δυνατότητες μέτρησης πλήθους αντικειμένων που εντοπίζονται καθώς και περαιτέρω ανάλυση τους με ποιοτικά και ποσοτικά χαρακτηριστικά όπως

δυνατότητες του μέσου, προσωπικό που δύναται να φιλοξενηθεί, προσωπικό που απαιτείται για την συντήρηση του κα.

Τα παραπάνω αποτελούν λίγα μόνο από κάποιες προτάσεις που μπορούν να χρησιμοποιηθούν για περαιτέρω επέκταση της εργασίας. Γίνεται κατανοητό ότι σκοπός της εργασίας δεν είναι η ανάπτυξη κάποιου αλγορίθμου αιχμής, αλλά περισσότερο ή μελέτη της διαδικασίας εκπαίδευσης ενός δικτύου για σκοπούς αναγνώρισης και εντοπισμού συγκεκριμένων αντικειμένων μέσα από υψηλής ανάλυσης δορυφορικές εικόνες.

# ΒΙΒΛΙΟΓΡΑΦΙΑ


[1] David Jacobs, "Image Gradients", 2005.

[2] Navneet Dalal, Bill Triggs, "Histograms of Oriented Gradients for Human Detection", 2010.

[3] Yichuan Tang, "Deep Learning using Linear Support Vector Machines".

[4] Yoshua Bengio, "Practical Recommendations for Gradient-Based Training of Deep Architectures", 2012.

[5] Yan Le Cun, "Efficient BackProp", 1998.

[6] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks".

[7] Vincent Dumoulin, Francesco Visin, "A guide to convolution arithmetic for deep learning".

[8] Yichuan Tang, "Deep Learning using Linear Support Vector Machines", 2015.

[9] Michael Nielsen, "Neural Networks and Deep Learning", 2015

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

[11] Pedro Felzenszwalb, Daniel Huttenlocher, "Efficient Graph-Based Image Segmentation", 2004.

[12] Ross Girshick, "Fast R-CNN", 2015

[13] Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks", 2016, arXiv: 1506.01497

[14] Kaiming He, Georgia Gkioxari, Piotr Dollár, Ross Girshick, "Mask R-CNN", 2017.

[16] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Szegedy Alexander C. Berg, 2015, "SSD: Single Shot MultiBox Detector", arXiv:1512.02325

[17] A. Robicquet, A. Sadeghian, A. Alahi, S. Savarese, "Learning Social Etiquette: Human Trajectory Prediction in Crowded Scenes", European Conference on Computer Vision (ECCV), 2016.

[18] Faisal Bashir, Fatih Porikli, "Performance Evaluation of Object Detection and Tracking Systems", 2006

[19] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors", 2012.

[20] P. Deepan, L.R. Sudha, in The Cognitive Approach in Cloud Computing and Internet of Things Technologies for Surveillance Tracking Systems, 2020

[21] He, K., Gkioxari, G., Dollár, P. and Girshick, R., 2017. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 2961-2969).

[22] Object Detection in 20 Years: A Survey Zhengxia Zou, Zhenwei Shi, Member, IEEE, Yuhong Guo, and Jieping Ye, Senior Member, IEEE.

[23] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on, vol. 1. IEEE, 2001, pp. I–I.

[24] P. Viola and M. J. Jones, "Robust real-time face detection," International journal of computer vision, vol. 57, no. 2, pp. 137–154, 2004.

[25] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, vol. 1. IEEE, 2005, pp. 886–893.

[26] P. Felzenszwalb, D. McAllester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," in Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. IEEE, 2008, pp. 1–8.

[27] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 580–587.

[28] R. Girshick, "Fast r-cnn," in Proceedings of the IEEE international conference on computer vision, 2015, pp. 1440–1448.

[29] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in Advances in neural information processing systems, 2015, pp. 91–99.

[30] Kaiming He, Georgia Gkioxari, Piotr Dollár, Ross Girshick, "Mask R-CNN", 2017.

[31] [15] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi, 2015, "You Only Look Once: Unified, Real-Time Object Detection", arXiv : 1506.02640.

[32] Redmon, J., & Farhadi, A. (2017). YOLO9000: better, faster, stronger. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7263-7271).

[33] Redmon, Joseph, and Ali Farhadi. "Yolov3: An incremental improvement." arXiv preprint arXiv:1804.02767 (2018).

[34] Bochkovskiy, Alexey, Chien-Yao Wang, and Hong-Yuan Mark Liao. "Yolov4: Optimal speed and accuracy of object detection." *arXiv preprint arXiv:2004.10934* (2020).

[35] Καρκάλου Ε., Συνελικτικά Νευρωνικά Δίκτυα για την πυκνή συνταύτιση ζεύγους εικόνων. Μεταπτυχιακή εργασία, ΔΠΜΣ «Γεωπληροφορική»,2017.

[36] Φίλιππας Δ., Εντοπισμός και Παρακολούθηση Κινούμενων Οχημάτων από RGB και Υπερφασματικά Δεδομένα Βίντεο. Μεταπτυχιακή εργασία, ΠΜΣ «Τεχνολογίες Πληροφορικής και Τεχνολογιών», 2017.

[37] Βασίλη Κωνσταντίνος, Συνελικτικά νευρωνικά δίκτυα για την αναγνώριση αντικειμένων σε εναέρια υψηλής ανάλυσης δεδομένα βίντεο. Διπλωματική εργασία, ΣΑΤΜ, 2018.