



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΕΠΙΚΟΙΝΩΝΙΩΝ, ΗΛΕΚΤΡΟΝΙΚΗΣ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ

Ανίχνευση Οχημάτων με Χρήση Υπολογιστικής Όρασης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ
της
Αικατερίνης Γρετσίστα

Επιβλέπουσα: Θεοδώρα Βαρβαρίγου
Καθηγήτρια Ε.Μ.Π

Αθήνα, Νοέμβριος 2019



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΕΠΙΚΟΙΝΩΝΙΩΝ, ΗΛΕΚΤΡΟΝΙΚΗΣ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ

Ανίχνευση Οχημάτων με Χρήση Υπολογιστικής Όρασης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ
της
Αικατερίνης Γρετσίστα

Επιβλέπουσα: Θεοδώρα Βαρβαρίγου
Καθηγήτρια Ε.Μ.Π

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 1η Απριλίου 2019.

(Υπογραφή)
.....
Θεοδώρα Βαρβαρίγου
Καθηγήτρια Ε.Μ.Π.

(Υπογραφή)
.....
Εμμανουήλ Βαρβαρίγος
Καθηγητής Ε.Μ.Π

(Υπογραφή)
.....
Συμεών Παπαβασιλείου
Καθηγητής Ε.Μ.Π

Αθήνα, Νοέμβριος 2019

.....
Γρετσίστα Αικατερίνη
Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright © Γρετσίστα Αικατερίνη, 2019.
Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς την συγγραφέα. Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν την συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Περίληψη

Οι σημερινές κοινωνίες είναι άρρηκτα συνδεδεμένες με τεχνολογικά επιτεύγματα. Οι περισσότεροι άνθρωποι που ζουν σε χώρες του αναπτυσσόμενου κόσμου δεν θα μπορούσαν να φανταστούν την ζωή τους χωρίς την έξυπνη κινητή τους συσκευή. Είναι γεγονός πως η τεχνολογία έχει διευκολύνει σε σημαντικό βαθμό τις ζωές των ανθρώπων. Αυτό που είναι ιδιαίτερα ενδιαφέρον ωστόσο, είναι το γεγονός πως πλέον απλοί χρήστες είναι σε θέση να έχουν πρόσβαση σε πιο εξειδικευμένες τεχνολογίες, σε αντίθεση με παλιότερες δεκαετίες που αυτό ήταν προνόμιο μόνο των κυβερνητικών οργανώσεων ή μεγάλων οργανισμών και εταιριών. Μία από αυτές τις τεχνολογίες αποτελεί και το θέμα της συγκεκριμένης Διπλωματικής Εργασίας, τα μη επανδρωμένα αεροσκάφη (Unmanned Aerial Vehicle – UAV). Τα Drone, όπως ευρέως αποκαλούνται τα μη επανδρωμένα εναέρια οχήματα, διανέμονται πλέον σε αρκετά προσιτές τιμές από πολλές εταιρίες παγκοσμίως.

Οι περισσότεροι από εμάς έχουμε ονειρευτεί, μία συσκευή, όπως για παράδειγμα το κινητό μας τηλέφωνο να μπορούσε να εντοπίσει για εμάς το αυτοκίνητό μας ή ίσως να μας βρει μία ελεύθερη θέση στάθμευσης, γλιτώνοντάς μας από πολύ κόπο και σίγουρα ενέργεια. Η εποχή που αυτό θα είναι εύκολα υλοποιήσιμο και προσβάσιμο από όλους τις χρήστες έξυπνων κινητών συσκευών (smartphones, tablets κ.τ.λ.) είναι πολύ πιο κοντά από όσο θα φανταζόταν κανείς. Στην προσέγγιση αυτού του στόχου μπορούν να συμβάλλουν καθοριστικά τα μη επανδρωμένα εναέρια οχήματα.

Για την αντιμετώπιση του προβλήματος εντοπισμού αντικειμένων, όπως οχήματα, σπίτια αλλά ακόμα και ζώα και άνθρωποι, έχουν προταθεί πολλοί αλγόριθμοι οι οποίοι βασίζονται σε θεμελιώδεις αρχές της μηχανικής μάθησης (machine learning), της βαθιάς μάθησης (deep learning) καθώς και της τεχνητής νοημοσύνης (Artificial Intelligence-AI). Ωστόσο υπάρχουν ακόμα μερικές προκλήσεις αναφορικά με την καθολικότητα εφαρμογής αυτών των πρωτοκόλλων, οι οποίες παρουσιάζονται εκτενώς στην πορεία της εργασίας. Προκειμένου να μπορέσει μία τέτοια εφαρμογή να χρησιμοποιείται ευρέως, θα πρέπει να ξεπεραστούν τουλάχιστον μερικές από αυτές τις δυσκολίες.

Στην συγκεκριμένη εργασία προτείνεται η χρήση υπολογιστικής όρασης, μέσω μη επανδρωμένου αεροσκάφους ως μία πιθανή λύση για την άρση μερικών εκ των περιορισμών που συναντώνται κατά την προσπάθεια ανίχνευσης αντικειμένων. Πιο συγκεκριμένα, στο πλαίσιο της εργασίας, αναπτύσσεται μια εφαρμογή για έξυπνες συσκευές με λειτουργικό iOS, η οποία επιτρέπει στον χρήστη τον χειρισμό μη επανδρωμένου αεροσκάφους της εταιρείας DJI. Η DJI προσφέρει το απαραίτητο Kit Ανάπτυξης Λογισμικού (Software Development Kit - SDK) για την δημιουργία εφαρμογών. Σκοπός της εφαρμογής είναι η σάρωση περιοχών που επιλέγονται μέσω των χαρτών της Apple και η αναγνώριση οχημάτων στις περιοχές αυτές, χρησιμοποιώντας το λογισμικό OpenCV.

Λέξεις Κλειδιά: μη επανδρωμένα αεροσκάφη, αναγνώριση αντικειμένων, μηχανική μάθηση, έξυπνες κινητές συσκευές, τεχνητή νοημοσύνη, xCode, γλώσσα προγραμματισμού swift.

Abstract

Today's societies are inextricably linked to technological achievements. Most people living in the developed world could not imagine their lives without their smart mobile device. It is a fact that technology has greatly facilitated people's lives. What is particularly interesting, however, is the fact that ordinary users are now able to access more sophisticated technologies, as opposed to previous decades, when that was only a privilege for government organizations or large organizations and companies. One of these technologies is also the subject of this Diploma Thesis, Unmanned Aerial Vehicle (UAV). Drones, as unmanned aerial vehicles are commonly called, are now widely available at affordable prices from many companies worldwide.

Most of us have dreamed that a device, such as our cell phone, could locate our car for us or perhaps find us a free parking space, saving us a lot of trouble and definitely energy. The era when this is easily accessible to all users of smart devices (smartphones, tablets, etc.) is much closer than one might imagine. Unmanned aerial vehicles can make a decisive contribution towards this goal.

To address the problem of locating objects, such as vehicles, homes but even animals and humans, many algorithms have been proposed based on fundamental principles of machine learning, deep learning and Artificial Intelligence-AI. However, there are still some challenges regarding the universality of application of these protocols, which are widely presented in the course of the work. In order for such an application to be widely used, at least some of these difficulties must be overcome.

This Thesis proposes the use of computational vision via unmanned aircraft as a possible solution to remove some of the constraints encountered when attempting to detect objects. In particular, as part of this paper, an application is developed for smart devices running iOS, which allows the user to operate the unmanned aircraft of the DJI company. DJI offers the necessary Software Development Kit (SDK) for creating applications. The purpose of the application is to scan areas selected via Apple maps and identify vehicles in those areas using OpenCV software.

Keywords: unmanned aircrafts, object detection, machine learning, smart mobile devices, Artificial Intelligence, xCode, swift programming.

Ευχαριστίες

Θα ήθελα αρχικά να ευχαριστήσω την καθηγήτρια Θεοδώρα Βαρβαρίγου που με εμπιστεύτηκε με την ανάθεση της συγκεκριμένης διπλωματικής εργασίας. Παρόλο που οι μέχρι πρωτίνως γνώσεις μου στον τομέα του προγραμματισμού και του deep learning ήταν αρκετά περιορισμένες, σε αυτή την πρώτη επαφή με την παραγωγή ακαδημαϊκού έργου, ήθελα να προκαλέσω τον εαυτό μου να μπει σε άγνωστα μονοπάτια. Είμαι ευγνώμων που καθ' όλη την διάρκεια εκπόνησης της εργασίας, είχα την υποστήριξη και καθοδήγηση όχι μόνο της κυρίας Βαρβαρίγου, αλλά και του υποψηφίου μεταπτυχιακού, Αναστασίου Νικολακόπουλου. Επίσης θα ήθελα θερμά να ευχαριστήσω τον υπεύθυνο του εργαστηρίου, Βρεττό Μούλο, καθώς και ολόκληρη την ομάδα του εργαστηρίου των Ψηφιακών Συστημάτων, για την καθοριστική στήριξή τους αλλά κυρίως την υπομονή τους.

Τους τελευταίους μήνες έχω μάθει πολύ περισσότερα πράγματα από ότι πίστευα πως θα μπορούσα σε τόσο μικρό χρονικό διάστημα. Ένα από τα κυριότερα μαθήματα που πήρα είναι να μην φοβάμαι να πειραματίζομαι και να ασχολούμαι με τομείς που δεν ανήκουν στην ζώνη ασφαλείας μου. Με την ολοκλήρωση της διπλωματικής μου, κλείνει για μένα το κεφάλαιο της φοίτησής μου στην σχολή των Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών με τον πλέον ευχάριστο τρόπο. Ανυπομονώ να προχωρήσω στο επόμενο κεφάλαιο της ακαδημαϊκής και επαγγελματικής μου πορείας, κουβαλώντας μαζί μου πολλές εμπειρίες, γνωριμίες και γνώσεις τις οποίες ανυπομονώ να εφαρμόσω στην πράξη. Ελπίζω κάθε κεφάλαιο της ζωής μου να με αποζημιώνει τόσο όσο αυτό της φοίτησής μου στο Εθνικό Μετσόβιο Πολυτεχνείο. Τέλος θα ήθελα ευχαριστήσω από καρδιάς την οικογένειά μου και τους ανθρώπους που ήταν δίπλα μου όλα αυτά τα δημιουργικά χρόνια, τις στιγμές που ήμουν πεπεισμένη πως είμαι μέλος μίας εκ των πλέον δυναμικών ομάδων στην Ελλάδα, ως φοιτήτρια των ΗΜΜΥ, αλλά κυρίως και όταν αμφέβαλλα για τις επιλογές μου και ανησυχούσα για το μέλλον μου. Η στήριξη και η αγάπη σας είναι παραπάνω από πολύτιμη για εμένα.

Πίνακας Περιεχομένων

Περίληψη	5
Abstract	7
Ευχαριστίες	9
Πίνακας Περιεχομένων	10
Κατάλογος Σχημάτων	13
1. Εισαγωγή	15
1.1 Μη επανδρωμένα αεροσκάφη - drones	15
1.2 Drones και φωτογραμμετρία	16
1.3 Drones for good	17
1.4 Αντικείμενο και σκοπός της διπλωματικής	20
2. Σχετικές Εφαρμογές	21
2.1 Ανίχνευση ατόμων και οχημάτων σε πραγματικό χρόνο από εικόνες UAV	21
2.2 Γρήγορη ανίχνευση αντικειμένων βασισμένη σε βαθμίδες υψηλών τάξεων και παλινδρόμηση της διαδικασίας Gauss για εικόνες UAV	23
2.3 DroneSAR Search & Rescue	24
3. Μηχανική Μάθηση	25
3.1 Η λειτουργία της μηχανικής μάθησης	25
3.2 Παραδείγματα μηχανικής μάθησης	26
3.3 Τύποι μοντέλων μηχανικής μάθησης	26
3.3.1 Δέντρα αποφάσεων	27
3.3.2 K-means ομαδοποίηση (clustering)	27
3.3.3 Νευρωνικά δίκτυα (Neural Networks-NN)	27
3.3.4 Μάθηση ενδυνάμωσης (Reinforcement Learning)	28
3.4 Το μέλλον της μηχανικής μάθησης	28
3.5 Αναγνώριση προσώπου	29
4. Ανίχνευση Αντικειμένων	30
4.1 Αλγόριθμοι μηχανικής μάθησης για ανίχνευση αντικειμένων	31
4.1.1 Ιστόγραμμα προσανατολισμένων διαβαθμίσεων (Histogram of Oriented Gradients-HOG)	32
4.1.2 Πλαίσιο ανίχνευσης αντικειμένων Viola-Jones	32
4.2 Τεχνικές ταξινόμησης	32
4.2.1 Αλγόριθμος KNN (K Κοντινότεροι Γείτονες)	33

4.2.2 Μηχανή Υποστήριξης Διανυσμάτων (SUPPORT VECTOR MACHINE-SVM)	33
4.3 CNN (Convolutional Neural Network/Συνελικτικό Νευρωνικό Δίκτυο)	34
4.4 Ανίχνευση αντικειμένων με χρήση τεχνικών βαθιάς μάθησης	34
4.4.1 Σύγκριση τεχνικών μηχανικής μάθησης έναντι αυτών βαθιάς μάθησης	35
4.4.2 Αλγόριθμοι ανίχνευσης αντικειμένων βαθιάς μάθησης	35
4.4.2.1 R-CNN	36
4.4.2.2 Fast R-CNN	37
4.4.2.3 Faster R-CNN	40
4.4.2.4 Δίκτυο πρότασης περιοχής (Region Proposal Network-RPN)	41
4.4.2.4.1 Mask R-CNN	42
4.4.2.5 Shot MultiBox Detector (SSD)	43
4.4.2.6 YOLO	44
4.4.2.6.1 Περιορισμοί του YOLO	45
4.4.2.6.2 Πλεονεκτήματα του YOLO	46
5. Η εφαρμογή <i>DJI GoTo</i>	47
5.1 Γενικά για την εφαρμογή	47
5.2 Περιγραφή της λειτουργίας της εφαρμογής	48
5.2.1 Γιατί επιλέχθηκε η χρήση Drone	49
5.2.2 Mavic Pro	50
5.2.3 Χρήση αλγορίθμου ανίχνευσης αντικειμένων	51
5.2.3.1 Πρόβλεψη κουτιού οριοθέτησης	51
5.2.3.2 Πρόβλεψη κλάσης	52
5.2.3.3 Προβλέψεις σε κλίμακες	52
5.2.3.4 Απόσπαση στοιχείων: Darknet-53	53
5.2.3.5 Εκπαίδευση	53
5.2.3.6 Συμπεράσματα	53
5.3 Γραφικό περιβάλλον και παρουσίαση του κώδικα της εφαρμογής	54
5.3.1 Υλοποίηση της εφαρμογής	54
5.3.2 Εφαρμογή του αλγορίθμου YOLO για την ανίχνευση των οχημάτων	66
5.4 Γραφικό περιβάλλον	68
6. Μετρήσεις και αποτελέσματα	72
6.1 Αποτελέσματα αλγορίθμου στον εντοπισμό αυτοκινήτων	72
6.1.1 Αποδοτικότητα αλγορίθμου	72
6.1.2 Χρόνος εκτέλεσης αλγορίθμου	75
6.1.3 Σύγκριση λειτουργικότητας των διαφορετικών αναλύσεων εικόνων	75

7. Προτάσεις για περαιτέρω έρευνα	76
7.1 Επίλυση προβλημάτων και βελτιώσεις	77
7.2 Επεκτάσεις	78
7.2.1 Εύρεση θέσης στάθμευσης	78
7.2.1.1 Ανίχνευση θέσεων στάθμευσης	78
7.2.1.2 Προσδιορισμός κατειλημμένων θέσεων στάθμευσης	80
7.2.2 Ανάλυση κίνησης πελατών	82
7.2.3 Ανάλυση κυκλοφορίας στους αυτοκινητόδρομους	82
7.2.4 Εντοπισμός επιζώντων	82
Βιβλιογραφία	84

Κατάλογος Σχημάτων

1.3.1: Drone σκανάρει απομακρυσμένη περιοχή	18
1.3.2: Drones για τον καθαρισμό των θαλασσών από διαρροές πετρελαίου	19
1.3.3: Όταν ανιχνεύεται διαρροή πετρελαίου τα βακτήρια τοποθετούνται στον Bio-Cleaner	20
4.4.2.1.1: Ο αλγόριθμος επιλεκτικής αναζήτησης λαμβάνει κατ' επανάληψη προτάσεις περιοχής	36
4.4.2.1.2: Το pipeline εκπαίδευσης του R-CNN, λαμβάνει μια σειρά προτάσεων περιοχής και παρέχει μια ετικέτα ταξινόμησης	37
4.4.2.2.1: Σύγκριση των αλγορίθμων εντοπισμού αντικειμένων	40
4.4.2.3.1: Αρχιτεκτονική μοντέλου Faster R-CNN. Αποτέλεσμα της ανίχνευσης αντικειμένων με Faster R-CNN	41
4.4.2.3.2: Σύγκριση των αλγορίθμων εντοπισμού αντικειμένων	41
4.4.2.4.1.1: Αποτέλεσμα της ανίχνευσης αντικειμένων με χρήση του αλγορίθμου Mask R-CNN	42
4.4.2.4.1.2: Παράδειγμα από το σύνολο δεδομένων COCO	43
4.4.2.5.1: Αρχιτεκτονική μοντέλου SSD	44
4.4.2.6.1: Τα στάδια χρήσης του YOLO για ανίχνευση αντικειμένων	45
4.4.2.6.2.1: Αποτέλεσμα χρήσης του YOLO για ανίχνευση αντικειμένων	46
5.2.1: Διάγραμμα ροής για την περιγραφή της λειτουργίας της εφαρμογής	48
5.2.2: Γενική Λειτουργία του Gimbal	49
5.2.3: Gimbal του Μοντέλου Mavic Pro	49
5.2.2.1: Mavic Pro	50
5.2.3.2.1: Διαφορά μεταξύ προβλεπόμενων και πραγματικών κουτιών	52
5.2.3.6.1: Σύγκριση αλγορίθμων με χρήση του COCO dataset σε 0.5 mAP	54
5.3.1.1: Κώδικας εγγραφής της εφαρμογής και σύνδεσης του αεροσκάφους	55
5.3.1.2: Σύνδεση εφαρμογής με το μη επανδρωμένο αεροσκάφος	56
5.3.1.3: Εστίαση στην θέση του αεροσκάφους στον χάρτη	57
5.3.1.4: Εμφάνιση του αεροσκάφους στον χάρτη	57
5.3.1.5: Αναγνώριση σημείων ως waypoints	58
5.3.1.6: Προσθήκη σημείων στον χάρτη	58

5.3.1.7: Έλεγχος επαρκούς αριθμού σημείων και διαγραφή σημείων	59
5.3.1.8: Ενημέρωση των συντεταγμένων και της κατεύθυνσης κίνησης του drone	60
5.3.1.9: Εφαρμογή του flight controller για τον έλεγχο της κίνησης του drone	60
5.3.1.10: Εισαγωγή default τιμών για μεγέθη που αφορούν την πτήση	61
5.3.1.11: Μετατροπή των χαρακτήρων σε πραγματικούς αριθμούς και finish, cancel buttons	61
5.3.1.12: Κλήση μεθόδων setupVideoPreviewer() και resetVideoPreview() για την λειτουργία της κάμερας	63
5.3.1.13: Οι delegate μέθοδοι της DJISDKManagerDelegate, η viewWillDisappear	63
5.3.1.14: Η συνάρτηση που αφορά την λήψη φωτογραφιών	65
5.3.1.15: Η συνάρτηση που αφορά την καταγραφή βίντεο	65
5.3.1.16: Συνάρτηση για την αλλαγή κατάστασης από λήψη φωτογραφιών σε καταγραφή βίντεο και το αντίστροφο	66
5.3.2.1: Εισαγωγή της εικόνας, των labels του COCO και των αντίστοιχων weights	67
5.3.2.2: Εισαγωγή των κλάσεων του dataset και δημιουργία blob για τα bounding boxes	67
5.3.2.3: Δημιουργία του bounding box για κάθε όχημα που εντοπίζεται	68
5.4.1: Η κύρια οθόνη της εφαρμογής <i>DJI GoTo</i>	69
5.4.2: Εμφάνιση του παραθύρου <i>Waypoint Configuration</i> μετά το πάτημα του κουμπιού <i>Config</i>	70
5.4.3: Στιγμιότυπο οθόνης για χρήση της ενσωματωμένης κάμερας του drone	71
5.4.4: Ο DJI Assistant 2 Simulator προσομοιώνει την πτήση του αεροσκάφους	72
6.1.1.1: Ποσοστά επιτυχίας του YOLO στην εύρεση οχημάτων για τα διαφορετικά μεγέθη εικόνας	73
6.1.1.2: Διάγραμμα για τα ποσοστά επιτυχίας εντοπισμού οχημάτων	74
6.1.2.1: Γράφημα για τον χρόνο εκτέλεσης του YOLO στις διαφορετικές αναλύσεις εικόνας	75
6.1.3.1: Η εικόνα που λαμβάνουμε σαν έξοδο του αλγορίθμου YOLO	76
7.2.1.1.1: Η έξοδος μετά την εφαρμογή του αλγορίθμου ανίχνευσης Canny edge	79
7.2.1.1.2: Η έξοδος μετά την εφαρμογή του αλγορίθμου hough line transform	79
7.2.1.1.3: Το αποτέλεσμα μετά την ομαδοποίηση των συντεταγμένων του άξονα x	80
7.2.1.1.4: Το τελικό αποτέλεσμα της ανίχνευσης των θέσεων στάθμευσης	80
7.2.1.2.1: Εύρεση κατειλημμένων θέσεων με χρήση CNN	81

Κεφάλαιο 1

Εισαγωγή

1.1 Μη επανδρωμένα αεροσκάφη

Τα μη επανδρωμένα εναέρια οχήματα (Unnamed Aerial Vehicles-UAV) ή drones είναι μια κατηγορία αεροσκαφών, τα οποία έχουν την δυνατότητα να πραγματοποιούν πτήσεις χωρίς την παρουσία πιλότων, είτε αυτόνομα, είτε μέσω τηλεκατεύθυνσης. Τα μη επανδρωμένα ιπτάμενα οχήματα συνήθως έχουν τη μορφή μικρού αεροπλάνου ή ελικοπτέρου με έναν ή περισσότερους κινητήρες και έλικες συντονισμένους για πλήρως ελεγχόμενη πτήση από ειδικό πρόγραμμα ή χειριστήριο εδάφους αεροσκαφών. Αποτελούνται από τα συστατικά μέρη του αεροσκάφους, τα ωφέλιμα φορτία των αισθητήρων και τον σταθμό ελέγχου εδάφους. [1]

Για την πραγματοποίηση αυτόνομης πτήσης προϋποτίθεται η χρήση εξειδικευμένου λογισμικού, το οποίο ελέγχει το σχέδιο πτήσης, που σε συνεργασία με τους onboard αισθητήρες και το GPS, καθοδηγεί το αεροσκάφος. Ο έλεγχος της πτήσης, πραγματοποιείται είτε με ηλεκτρονικό εξοπλισμό επί του αεροσκάφους, είτε μέσω εξοπλισμού ελέγχου από το έδαφος. Στην περίπτωση όπου ο έλεγχος πραγματοποιείται εξ αποστάσεως, από το έδαφος, το εναέριο όχημα ονομάζεται RPV (Remote Piloted Vehicle) και απαιτεί αξιόπιστη ασύρματη επικοινωνία για έλεγχο. Τα UAV, τα οποία πιθανώς αντιμετωπίζουν δυσκολίες αναφορικά με την εμβέλεια ή την επικοινωνία, μπορούν να ενισχυθούν με την τοποθέτηση ειδικών συστημάτων ελέγχου πάνω σε οχήματα ή σε ρυμουλκούμενα σε μικρή απόσταση από αυτά.

Τα μη επανδρωμένα αεροσκάφη, τα οποία χρησιμοποιούνται πλέον κατά κόρον για παρατήρηση και τακτικό σχεδιασμό, ταξινομούνται με βάση το εύρος του ύψους, την αντοχή και το βάρος, και υποστηρίζουν μια ευρεία σειρά εφαρμογών, συμπεριλαμβανομένων στρατιωτικών και εμπορικών εφαρμογών. Οι μικρότερες κατηγορίες αεροσκαφών, συνοδεύονται συχνά από σταθμούς ελέγχου στο έδαφος, οι οποίοι μπορεί να αποτελούνται από φορητούς υπολογιστές και άλλα εύκολα μεταφέρσιμα στοιχεία.

Τα drone, τα οποία είναι εξοπλισμένα με κάμερες υψηλής ακρίβειας, επιτρέπουν στους χειριστές να εκτελούν λεπτομερή ανάλυση εικόνας και δομής. Παρόλο που η αρχική χρήση των μη επανδρωμένων εναέριων οχημάτων, περιοριζόταν κυρίως σε στρατιωτικές εφαρμογές, μιας και η πρώτη τους εμφάνιση έγινε στον Β Παγκόσμιο Πόλεμο, σήμερα οι σχετικές εφαρμογές έχουν ξεπεράσει κάθε προσδοκία. Οι ερευνητές ανακαλύπτουν συνεχώς νέους τρόπους με τους οποίους η χρήση μη επανδρωμένων αεροσκαφών, μπορεί να συντελέσει στην βελτίωση της ζωής των ανθρώπων και την ανάπτυξη των κοινωνιών. [2]

1.2 Drones και φωτογραμμετρία

Η ιδέα της ανίχνευσης αντικειμένων, όπως αυτοκινήτων, μέσα από εικόνες ή βίντεο δεν είναι ιδιαίτερα πρωτοποριακή, μιας και συναντάται από την απαρχή των μηχανών λήψης βίντεο και των ενσωματωμένων αισθητήρων. Οι μέθοδοι ανίχνευσης που απαιτούν χρήση αισθητήρων συνοδεύονται από υψηλό κόστος δημιουργίας και διατήρησης, καθώς και από ιδιαίτερη πολυπλοκότητα εξαιτίας της απαίτησης για ενσωμάτωση πολλαπλών διαφορετικών πηγών δεδομένων, καθεμία εκ των οποίων μπορεί να δέχεται μία περιορισμένη ζώνη δεδομένων ως είσοδο. Ωστόσο, χάρη στην ταχύρρυθμη αύξηση του όγκου των πληροφοριών με την οποία ερχόμαστε αντιμέτωποι στην καθημερινότητά μας, σε συνδυασμό με την επανάσταση στον τομέα της μηχανικής μάθησης, βρισκόμαστε πλέον σε θέση να προσεγγίζουμε θέματα τα οποία απασχολούν τις ανθρώπινες κοινωνίες εδώ και δεκαετίες, από νέες διορατικές σκοπιές.

Η ανάπτυξη της βιομηχανίας των μη επανδρωμένων αεροσκαφών (drones) και η επικράτησή τους στην αγορά, έχουν προσφέρει απίστευτες δυνατότητες καινοτομίας. Η χρήση των μη επανδρωμένων αεροσκαφών έχει επιφέρει σημαντική μείωση στο κόστος λήψης εναέριων δεδομένων και έχει ανοίξει δρόμους για νέες τεχνολογίες αιχμής σε τομείς όπως η φωτογραμμετρία. Η ραγδαία αύξηση των δεδομένων που υπάρχουν στην διάθεσή μας, σε συνδυασμό με την μηχανική μάθηση και τα Συστήματα Γεωγραφικών Πληροφοριών (GIS) μας δίνει την δυνατότητα να εντρυφήσουμε με πιο παραγωγικό τρόπο σε νέους τομείς ενδιαφέροντας όπως αυτός της αναγνώρισης αντικειμένων.

Η φωτογραμμετρία είναι η επιστήμη, η οποία ασχολείται με την απόκτηση αξιόπιστων μετρήσεων από φωτογραφίες και ψηφιακές εικόνες. Η έξοδος της φωτογραμμετρικής διαδικασίας μπορεί να είναι συμβολικοί χάρτες, στρώματα συστημάτων γεωγραφικών πληροφοριών, τρισδιάστατα μοντέλα πραγματικών αντικειμένων και χάρτες ορθοφωτογραφίας. Στην εναέρια φωτογραμμετρία, η οποία αποτελεί έναν από τους γενικούς τύπους φωτογραμμετρίας, ο αισθητήρας βρίσκεται πάνω σε δορυφόρο, επανδρωμένο αεροσκάφος ή όπως στην προκειμένη περίπτωση, σε μη επανδρωμένο αεροσκάφος και έχει κάθετη κατεύθυνση σε σχέση με το έδαφος. Κατά την διάρκεια της πτήσης, συγκεντρώνονται πολλαπλές επικαλυπτόμενες εικόνες, οι οποίες επεξεργάζονται προκειμένου να παραχθούν ψηφιακά δεδομένα ανύψωσης και ορθοφωτογραφικά μωσαϊκά.

Οι εικόνες χαρακτηρίζονται από προοπτική γεωμετρία, γεγονός το οποίο έχει ως πιθανή συνέπεια την παραμόρφωση της εικόνας, ενδεχομένως με διαφορετικό τρόπο κάθε φορά. Οι ορθοεικόνες είναι γεωμετρικά διορθωμένες προκειμένου η τελική εικόνα να προσομοιάζει την αξιοπιστία ενός χάρτη. Η διαδικασία με την οποία επιτυγχάνεται η διόρθωση της γεωμετρικής παραμόρφωσης, η οποία κληροδοτείται στις εικόνες που έχουν προκύψει από τηλεανίχνευση, προκειμένου να παραχθεί ορθοεικόνα που θα προσεγγίζει την ακρίβεια ενός χάρτη, ονομάζεται ορθοαναγωγή. Κατόπιν, μία ομάδα από επεξεργασμένες ορθοεικόνες συνθέτουν το ορθομωσαϊκό. Προκειμένου να είναι δυνατή η δημιουργία ορθομωσαϊκού, είναι απαραίτητη η γνώση των συντεταγμένων των αισθητήρων, καθώς και ένα βαθμονομημένο γεωμετρικό μοντέλο για τον αισθητήρα.

Μερικές φορές η λήψη των εικόνων συνοδεύεται από γνώση της θέσης, όταν παραδίδονται στον χρήστη. Σε αντίθετη περίπτωση, οι εικόνες θα πρέπει να προσαρμοστούν στο σύστημα ελέγχου εδάφους. Οι διαδικασίες ρύθμισης χρησιμοποιούν τη βαθμονόμηση του αισθητήρα, τις πληροφορίες προσανατολισμού αισθητήρα, τα σημεία ελέγχου εδάφους, τα σημεία σύνδεσης και ένα DTM (digital terrain model) προκειμένου να παράγουν τις ακριβείς θέσεις. Αυτό με τη σειρά του επιτρέπει τη δημιουργία χαρτών ορθών εικόνων. Τα μεμονωμένα ορθογραφικά σήματα αντισταθμίζονται στη συνέχεια και εξισορροπούν το χρώμα για να δημιουργήσουν έναν απλό χάρτη ορθοφωτογραφιών. Αυτό το ψηφιδωτό ορθοφωτογραφίας είναι ακριβές σε μια συγκεκριμένη κλίμακα ακρίβειας κλίμακας και μπορεί να χρησιμοποιηθεί για την πραγματοποίηση μετρήσεων καθώς και για τη δημιουργία και την ενημέρωση επιπέδων κλάσεων χαρακτηριστικών GIS. [3]

1.3 Drones for good

Στις απαρχές του Διαδικτύου, κανείς ενδεχομένως δεν θα μπορούσε να φανταστεί πως η τεχνολογία θα έφτανε στο σημείο που είναι σήμερα. Η ραγδαία, σχεδόν τρομακτική, εξάπλωση των μέσων κοινωνικής δικτύωσης, η επικοινωνία μεταξύ διαφορετικών μηχανών, η αστείρευτη πρόσβαση στην γνώση, η εδραίωση του Bitcoin και των Blockchain είναι μερικά από τα παραδείγματα που επιβεβαιώνουν την εξέλιξη της τεχνολογίας, επιτεύγματα της οποίας αποτελούν πλέον κομμάτια της καθημερινότητας ενός μεγάλου μέρους του παγκοσμίου πληθυσμού. Παρόλα αυτά, δεν είναι λίγοι οι άνθρωποι οι οποίοι αδυνατούν να έχουν πρόσβαση στο Διαδίκτυο με κάποιο μέσο και έτσι παραμένουν αποκομμένοι από αυτή την ατέρμονη πηγή πληροφόρησης.

Τα μη επανδρωμένα αεροσκάφη, θα μπορούσαν να αποτελέσουν μία καίρια παρέμβαση στο πρόβλημα της πρόσβασης στις πληροφορίες και όχι μόνο, ακόμα και στα πλέον απομακρυσμένα μέρη. Χάρη στην χρήση μη επανδρωμένων αεροσκαφών, έχει καταστεί δυνατή η παράδοση φαρμακευτικού εξοπλισμού σε μέρη που υπάρχει ανάγκη. Συγκεκριμένα, μία από τις πρώτες δοκιμές για την προκειμένη εφαρμογή των μη επανδρωμένων αεροσκαφών, πραγματοποιήθηκε μετά τον σεισμό του 2010 στην Αϊτή. Τρία ημι-αυτόνομα μη επανδρωμένα αεροσκάφη, τα οποία πλοηγούνταν από τους χειριστές εξ αποστάσεως, κατάφεραν το 2012 να παραδώσουν φάρμακα σε αστικές, αλλά και αγροτικές περιοχές. Η χρήση των μη επανδρωμένων αεροσκαφών για την πραγμάτωση αυτού του έργου έχει αρκετά πλεονεκτήματα, μιας και δίνεται η δυνατότητα στοχευμένου σχεδιασμού των προμηθειών προς παράδοση, ακριβώς όταν χρειάζονται σε αντίθεση με την ενίοτε λανθασμένη πρόβλεψη των αναγκών κατά την εφαρμογή των παραδοσιακών τρόπων διανομής. [4]



Σχήμα 1.3.1: Drone σκανάρει απομακρυσμένη περιοχή

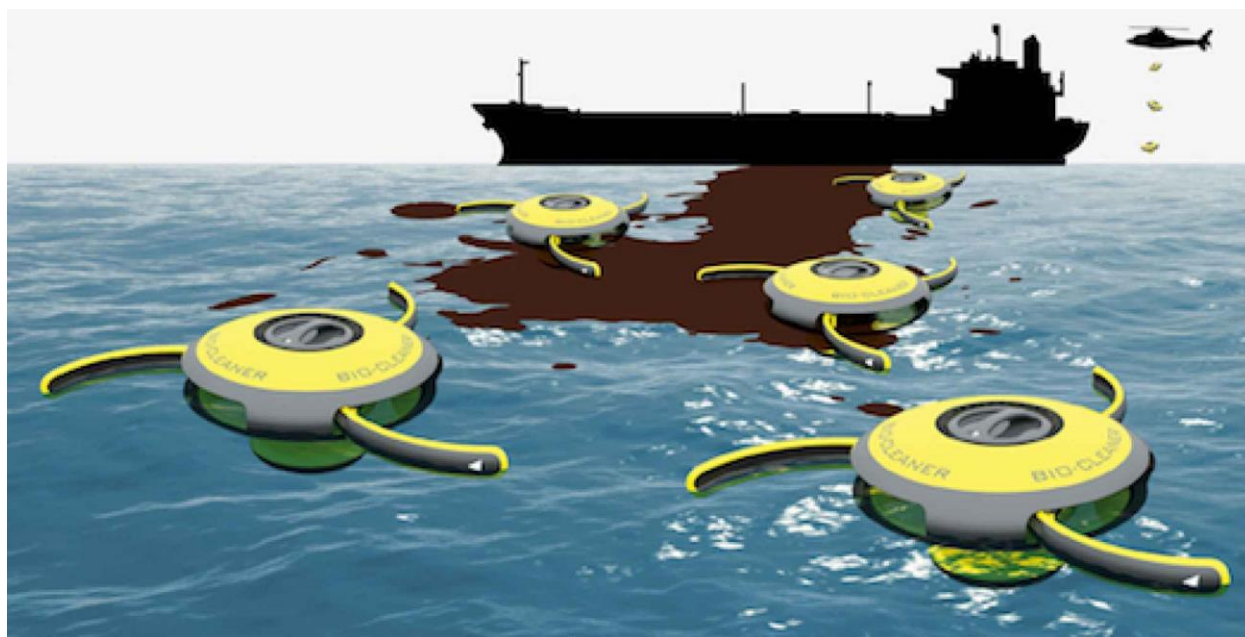
Ένα ακόμα από την πληθώρα των επιτευγμάτων των μη επανδρωμένων αεροσκαφών είναι η διανομή προμηθειών, πληροφοριών και διαγνωστικών εργαλείων μεταξύ ιατρικών κέντρων. Την ίδια σχεδόν περίοδο με την επιτυχημένη προσπάθεια στην Αϊτή, διεξάγει αντίστοιχο πείραμα στην Δομινικανή Δημοκρατία που αφορούσε την μεταφορά προμηθειών και διαγνωστικού εξοπλισμού από μεγάλα προς μικρότερα κέντρα υγείας σε απομακρυσμένες περιοχές, καθώς και την μεταφορά δειγμάτων προς τα μεγαλύτερα κέντρα για ανάλυση. Η εφαρμογή αυτή θα επωφελούσε ιδιαίτερα τους ανθρώπους, οι οποίοι ζουν σε αγροτικές περιοχές και δεν έχουν την δυνατότητα να μεταφερθούν σε κάποιο μεγαλύτερο ιατρικό κέντρο. Είναι αναμφίβολο πως οι εφαρμογές που προκύπτουν από την χρήση μη επανδρωμένων αεροσκαφών είναι πολλές και ιδιαίτερα σημαντικές.

Μελλοντικά, χάρη στην χρήση των μη επανδρωμένων αεροσκαφών θα ήταν ακόμα δυνατή η άμεση βοήθεια μετά από μία καταστροφή, ή και η πρόβλεψη της εξέλιξής της, όπως σε περίπτωση πυρκαγιάς. Στο εργαστήριο των Ψηφιακών Συστημάτων του Εθνικού Μετσόβιου Πολυτεχνείου, το οποίο διευθύνει η υπεύθυνη της παρούσας διπλωματικής εργασίας, κυρία Βαρβαρίγου, αναπτύσσεται μία εφαρμογή η οποία δέχεται δεδομένα που αφορούν μία πυρκαγιά σε εξέλιξη μέσω μη επανδρωμένου αεροσκάφους και σύμφωνα με συνιστώσες όπως η ταχύτητα και η φορά του ανέμου, το υψόμετρο και η θερμοκρασία, προβλέπεται η εξέλιξη της πυρκαγιάς, η οποία αναπαρίσταται και γραφικά. Η επιτυχία μίας τέτοιας πρωτοβουλίας, θα μπορούσε να έχει εκπληκτικά αποτελέσματα, ειδικά σε μία χώρα όπως η Ελλάδα, όπου κατά του καλοκαιρινούς κυρίως μήνες ο αριθμός των πυρόπληκτων περιοχών είναι πολύ μεγάλος.

Επιπρόσθετα, είναι πολύ πιθανό στο άμεσο μέλλον να παρέχεται η δυνατότητα μεταφοράς φαγητού και νερού σε περιοχές που δεν είναι εύκολα προσβάσιμες. Το εγχείρημα αυτό είναι ζωτικής σημασίας σε καιρούς καταστροφών, πολέμων ή κρίσεων. Προκειμένου να μπορεί να επιτευχθεί κάτι τέτοιο, είναι απαραίτητη η αύξηση του φορτίου, το οποίο είναι ικανό να μετακινήσει ένα μη επανδρωμένο αεροσκάφος. Ενδεικτικά, κάποια μη επανδρωμένα αεροσκάφη είναι ικανά να μεταφέρουν φορτίο βάρους δύο κιλών, καλύπτοντας μία απόσταση δέκα χιλιομέτρων σε δεκαπέντε λεπτά. Καθώς η τεχνογνωσία που αφορά τα μη επανδρωμένα αεροσκάφη εξελίσσεται, η δυνατότητα μεταφοράς φορτίου αυξάνεται προσεγγίζοντας έτσι την διεκπεραίωση της εν λόγω εφαρμογής.

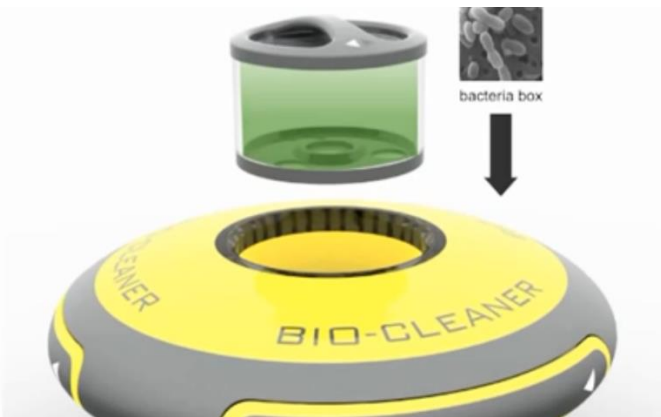
Οι παραπάνω τεχνολογικές εξελίξεις θα μπορούσαν να συντελέσουν στην προσπάθεια αποαστικοποίησης μεγάλων αστικών κέντρων, μιας και οι κάτοικοι αγροτικών περιοχών θα μπορούν να εξασφαλίζουν ίσες ευκαιρίες στην κάλυψη των βασικών αναγκών.

Μία ακόμη χρήση των μη επανδρωμένων αεροσκαφών προς το κοινωφελές καλό θα μπορούσε να είναι ο καθαρισμός των θαλασσών από ρυπογόνες ουσίες. Η ρύπανση των ωκεανών αποτελεί ένα εξαιρετικά φλέγον ζήτημα. Όσο και αν προετοιμαζόμαστε για μία πιθανή διαρροή πετρελαίου στην θάλασσα, ίσως να μην είναι ποτέ αρκετό, με αποτέλεσμα μεγάλες πετρελαιοκηλίδες να απειλούν την υποθαλάσσια ζωή. Ο Hsu Sean σχεδίασε ένα ιδιαίτερου τύπου μη επανδρωμένο αεροσκάφος το οποίο πιθανώς θα μπορούσε να συνεισφέρει καθοριστικά στο πρόβλημα του καθαρισμού των πετρελαιοκηλίδων. Το σχεδιαστικό αυτό έργο όχι μόνο συμβάλλει στον καθαρισμό της ίδιας της διαρροής, αλλά και περιλαμβάνει έναν πομπό που χρησιμοποιεί ακουστικά κύματα υψηλής συχνότητας για να αποθαρρύνει τους υποθαλάσσιους οργανισμούς να εισέλθουν στην καθαρισμένη ζώνη.



Σχήμα 1.3.2: Drones για τον καθαρισμό των θαλασσών από διαρροές πετρελαίου

Το σύστημα Bio-Cleaner είναι μια αυτόνομη λύση που ενσωματώνει μια ποικιλία εργαλείων καταπολέμησης της ρύπανσης, όπως τα βακτήρια βιοαποικοδόμησης και τους ακουστικούς συναγερμούς που αποσκοπούν στη απομάκρυνση των θαλάσσιων ζώων κατά τον καθαρισμό της διαρροής. Χρησιμοποιείται τεχνολογία βιοαισθητήρων για την παρακολούθηση της ροής του πετρελαίου καθώς ακολουθείται η διαρροή ενώ παράλληλα απελευθερώνονται τα βακτήρια με σκοπό να υποβαθμιστεί ο ρύπος, βελτιώνοντας έτσι τη διαδικασία καθαρισμού και εμποδίζοντας τη περαιτέρω εξάπλωση της ρύπανσης. Στις παραδοσιακές μεθόδους καθαρισμού των ωκεανών, χρησιμοποιούνται χημικά τα οποία δημιουργούν δευτερογενή ρύπανση των θαλασσών, καθώς επίσης σπαταλούνται πολλοί ανθρώπινοι πόροι και χρόνος. [5]



Σχήμα 1.3.3: Όταν ανιχνεύεται διαρροή πετρελαίου τα βακτήρια τοποθετούνται στον Bio-Cleaner

1.4 Αντικείμενο και σκοπός της διπλωματικής

Ο σκοπός της συγκεκριμένης εργασίας είναι καταρχάς η ανάπτυξη μιας προσαρμοσμένης εφαρμογής σε λειτουργικό περιβάλλον iOS, η οποία θα μπορεί να εντοπίζει την θέση του μη επανδρωμένου αεροσκάφους στον χάρτη και να επιστρέφει τις συντεταγμένες του. Κατόπιν θα δέχεται ως είσοδο από τον χρήστη τα επιθυμητά σημεία στο χάρτη σύμφωνα με τα οποία θα διαγράψει την πορεία του. Κατά την διάρκεια της πτήσης, ο χρήστης μπορεί να παρακολουθεί στον χάρτη την χάραξη της πορείας του αεροσκάφους καθώς μετακινείται από το ένα σημείο στο επόμενο. Διαθέτοντας ως εξοπλισμό ένα μικρό μη επανδρωμένο αεροσκάφος της εταιρίας DJI, εφοδιασμένο με ενσωματωμένη κάμερα, είναι εφικτό ο χρήστης να τραβά φωτογραφίες και βίντεο την επιθυμητή χρονική στιγμή. Συγκεκριμένα, παρέχεται στον χρήστη η δυνατότητα όρασης πρώτου προσώπου, χάρη στην οποία εμφανίζεται στην οθόνη του iPhone ό,τι καταγράφεται από την κάμερα του αεροσκάφους. Η εφαρμογή που ελέγχει το μη επανδρωμένο αεροσκάφος μέσω έξυπνης κινητής συσκευής της εταιρίας Apple, γράφηκε σε γλώσσα προγραμματισμού swift στο γραφικό περιβάλλον Xcode.

Το δεύτερο κομμάτι της εργασίας αφορά την επεξεργασία των εικόνων και των βίντεο που τραβήχτηκαν από την κάμερα στο μη επανδρωμένο αεροσκάφος. Επιθυμούμε να ανιχνεύουμε την ύπαρξη οχημάτων στην περιοχή των εγκαταστάσεων του Εθνικού Μετσόβιου Πολυτεχνείου. Για τον σκοπό αυτό, βασιζόμαστε στις θεμελιώδεις αρχές της υπολογιστικής όρασης και της μηχανικής μάθησης. Οι πιο δημοφιλείς αλγόριθμοι μηχανικής μάθησης καθώς και ο αλγόριθμος που χρησιμοποιήθηκε στην παρούσα Διπλωματική Εργασία με σκοπό την ανίχνευση οχημάτων, αλλά και όχι μόνο, παρουσιάζονται στην συνέχεια.

Παρόλο που εφαρμογές που επιτελούν παρόμοιες με τις προαναφερθείσες λειτουργίες μπορούν να βρεθούν στην αγορά, αυτές απευθύνονται κυρίως σε επαγγελματικούς σκοπούς. Στην προκειμένη περίπτωση, η εφαρμογή απευθύνεται κυρίως σε απλούς χρήστες, οι οποίοι διαθέτουν μη επανδρωμένο αεροσκάφος της DJI και έχουν πρόσβαση σε έξυπνη κινητή συσκευή της Apple. Φυσικά, η εφαρμογή μπορεί σχετικά εύκολα να τροποποιηθεί, ώστε να μπορεί να χρησιμοποιηθεί και από χρήστες έξυπνων κινητών συσκευών με λογισμικό Android. Η εφαρμογή

παρέχει την δυνατότητα σε καθημερινούς χρήστες να εκμεταλλευτούν τις δυνατότητες αναγνώρισης που μπορεί να τους παρέχει η κάμερα του μη επανδρωμένου εναέριου οχήματος.

Εκτός από την προφανή χρήση της εφαρμογής που αφορά τον εντοπισμό οχημάτων υπάρχουν και άλλοι τρόποι εκμετάλλευσης της ανίχνευσης αντικειμένων. Ενδεικτικά, είναι εφικτή η αναγνώριση ύπαρξης ανθρώπων ή ζώων στις φωτογραφίες και τα βίντεο, γεγονός που μπορεί να έχει τεράστια οφέλη όταν πρόκειται για παράδειγμα για φυσικές καταστροφές ή καταστάσεις αναγκαστικής εκκένωσης χώρων. Διανύουμε μία περίοδο συνειδητοποίησης των δυνατοτήτων που εν δυνάμει παρέχονται από την χρήση εναέριας καταγραφής εικόνων με την βοήθεια των μη επανδρωμένων αεροσκαφών. Οι εφαρμογές σταματούν μόνο εκεί που σταματά η φαντασία των ανθρώπων.

Κεφάλαιο 2

Σχετικές εφαρμογές

Όπως προαναφέρθηκε, την παρούσα στιγμή στην αγορά συναντώνται ποικίλες εφαρμογές που αφορούν την ανίχνευση αντικειμένων με χρήση υπολογιστικής όρασης από μη επανδρωμένο εναέριο όχημα. Στο κεφάλαιο αυτό, παρουσιάζονται μερικές από αυτές. Κάθε εφαρμογή έχει διαφορετικό σκοπό και καλύπτει διαφορετικές ανάγκες. Έτσι, υπάρχουν εφαρμογές, οι οποίες είναι ανοιχτού κώδικα, γεγονός που σημαίνει πως είναι προσβάσιμες από όλους χωρίς την προϋπόθεση καταβολής κάποιου χρηματικού ποσού. Συνήθως αυτές οι εφαρμογές απευθύνονται σε πιο απλούς χρήστες και οι δυνατότητες που παρέχουν είναι σχετικά πιο περιορισμένες σε σχέση με τις περισσότερο επαγγελματικές εφαρμογές, οι οποίες συχνά απαιτούν την καταβολή χρημάτων. Συγκεκριμένα, παρουσιάζονται τα αποτελέσματα από τέσσερις ερευνητικές δουλειές. Οι δύο πρώτες υλοποιήσεις παρέχονται δωρεάν, σε αντίθεση με την τελευταία.

2.1 Ανίχνευση ατόμων και οχημάτων σε πραγματικό χρόνο από εικόνες UAV

Μια γενική και ισχυρή προσέγγιση για τον εντοπισμό ανθρώπων και οχημάτων σε πραγματικό χρόνο από ένα μη επανδρωμένο αεροσκάφος (UAV) είναι ένας σημαντικός στόχος στο πλαίσιο της πλήρως αυτόνομης ανάπτυξης UAV για την εναέρια αναγνώριση και επιτήρηση. Παρουσιάζετε μια προσέγγιση για την αυτόματη ανίχνευση οχημάτων βασισμένη στη χρήση πολλαπλών εκπαιδευμένων κλιμακωτών ταξινομητών Haar (cascaded Haar classifiers), η οποία επιβεβαιώνεται με την χρήση θερμικών εικόνων. [6]

Η αλληλουχία Haar (Haar cascade), είναι ένας αλγόριθμος μηχανικής μάθησης, ο οποίος χρησιμοποιείται για αναγνώριση αντικειμένων. Αναγνωρίζει αντικείμενα σε μία εικόνα ή ένα βίντεο

και βασίζεται στην έννοια των χαρακτηριστικών, η οποία προτάθηκε το 2001 από τους Paul Viola και Michael Jones στην εργασία τους με τίτλο “Ταχεία αναγνώριση αντικειμένων με χρήση ενισχυμένης αλληλουχίας απλών χαρακτηριστικών”. Αποτελεί μία προσέγγιση με βάση τις αρχές της μηχανικής μάθησης όπου η cascade συνάρτηση εκπαιδεύεται από πλήθος θετικών και αρνητικών εικόνων και κατόπιν χρησιμοποιώντας τα αποτελέσματα της εκπαίδευσης, ανιχνεύονται αντικείμενα σε νέες εικόνες. [7]

Τα αποτελέσματα που παρουσιάζονται δείχνουν την επιτυχή ανίχνευση οχημάτων και ανθρώπων υπό διαφορετικές συνθήκες τόσο σε αστικά, όσο και σε απομονωμένα αγροτικά περιβάλλοντα με ελάχιστη ψευδή ανίχνευση. Επί του παρόντος, ο αλγόριθμος σημειώνει ποσοστό ανίχνευσης ατόμων που κυμαίνεται περίπου στο 70% και αυτοκινήτων στο 80%. Παρόλα αυτά, ο συνολικός ρυθμός ανίχνευσης αντικειμενικών για κάθε πρότυπο πτήσης υπερβαίνει το 90%.

Προκειμένου να επιτευχθεί η μέγιστη ποιότητα εικόνας καθώς και να αποφευχθεί η υπέρβαση του εύρους ζώνης μετάδοσης, η δειγματοληψία στην υλοποίηση αυτή γίνεται με 1 fps (frame per second – πλαίσιο ανά δευτερόλεπτο). Όπως είναι λογικό, ο εντοπισμός ενός οχήματος είναι ευκολότερος σε σχέση με τον εντοπισμό ενός ανθρώπου, λόγω του όγκου που καταλαμβάνουν. Επιπλέον, ένας ακόμα παράγοντας, πέρα από το μέγεθος, που δυσχεραίνει την πιθανότητα έγκυρης ανίχνευσης ανθρώπων, είναι η χαμηλή αντίθεση που συχνά συναντάται ανάμεσα στον άνθρωπο και το περιβάλλον του.

Μία πολύ αποτελεσματική λύση στο συγκεκριμένο πρόβλημα αποτελούν οι θερμικές εικόνες. Ωστόσο, παρόλο που η αρχική κατάσταση βελτιώνεται χάρη στις θερμικές εικόνες, η ποιότητα της θερμικής εικόνας ενδέχεται να επηρεαστεί από τις εκάστοτε καιρικές συνθήκες, γεγονός που αποτελεί σημαντική πρόκληση. Με την ολοκλήρωση της διεργασίας της επεξεργασίας της εικόνας, πέρα από τα αποτελέσματα της ανίχνευσης, παρέχεται επίσης για κάθε αντικείμενο, ένας δείκτης εμπιστοσύνης. Ο δείκτης αυτός, θα μπορούσε να χρησιμοποιηθεί ως είσοδος σε κάποιο δευτερεύον σύστημα ελέγχου με σκοπό την ακριβέστερη ταξινόμηση των αποτελεσμάτων. Η αποτελεσματικότητα της ανίχνευσης μπορεί να αυξηθεί επιπλέον με τη χρήση πολλαπλών ταξινομητών και αισθητήρων.

Όσον αφορά τα τεχνικά μέρη που αφορούν την ανίχνευση αντικειμένων σε πραγματικό χρόνο με χρήση εικόνων από μη επανδρωμένο εναέριο όχημα, στο σύστημα λήψης φωτογραφιών περιλαμβάνεται ένα μικρό μη επανδρωμένο αεροσκάφος με σταθερά πτερύγια, εφοδιασμένο με οπτική και θερμική κάμερα, το οποίο είναι ρυθμισμένο να πετάει σε υψόμετρο περίπου 60 μέτρων, με την κάμερα στραμμένη προς το έδαφος υπό γωνία 45 μοιρών, δημιουργώντας έτσι απόσταση από το στόχο 180 μέτρα, σε ευθεία γραμμή. Οι εικόνες αποστέλλονται στο έδαφος μέσω ραδιοζεύκτη, καθώς η ανίχνευση πραγματοποιείται σε τετραπύρρηνο επεξεργαστή, συχνότητας 2.33 GHz, κατασκευασμένο από την Intel.

Για τον εντοπισμό των αντικειμένων, χρησιμοποιείται ένα ζεύγος εικόνων, μια απλή οπτική εικόνα και μια θερμική. Αρχικά, οι δύο εικόνες επεξεργάζονται ανεξάρτητα η μία από την άλλη. Εν συνεχεία, οι εξαγόμενες από την επεξεργασία πληροφορίες, συνδυάζονται για το τελικό αποτέλεσμα. Προκειμένου να επιτευχθεί ο εντοπισμός των οχημάτων, η οπτική εικόνα ελέγχεται με πολλά διαφορετικά cascaded Haar classifiers που έχουν εκπαιδευτεί για εντοπισμό οχημάτων σε διαφορετικούς προσανατολισμούς και φωτισμό. Στο επόμενο στάδιο, αφού δηλαδή έχει με επιτυχία εντοπιστεί κάποιο όχημα με τα παραπάνω μοντέλα, ελέγχεται η αντίστοιχη θερμική εικόνα στο σημείο εντοπισμού του εκάστοτε οχήματος. Αντίστοιχα πραγματοποιείται και η ανίχνευση ατόμων, η οποία όπως προαναφέρθηκε σημειώνει χαμηλότερα ποσοστά επιτυχίας

λόγω μεγέθους του αντικειμένου προς εντοπισμό καθώς και έλλειψη αντίθεσης με τα περιβάλλοντα στοιχεία.

2.2 Γρήγορη ανίχνευση αντικειμένων βασισμένη σε βαθμίδες υψηλών τάξεων και παλινδρόμηση της διαδικασίας Gauss για εικόνες UAV

Τα μη επανδρωμένα εναέρια οχήματα (UAV) αποκτούν εικόνες που χαρακτηρίζονται από εξαιρετικό επίπεδο λεπτομέρειας, το οποίο απαιτεί μεθόδους επεξεργασίας και ανάλυσης ικανές να εκμεταλλεύονται αποτελεσματικά το πλούσιο πληροφοριακό τους περιεχόμενο. Ενδεικτικά, η ανίχνευση συγκεκριμένων κατηγοριών αντικειμένων (π.χ. αυτοκίνητα, στέγες) αντιπροσωπεύει ένα σημαντικό αλλά δύσκολο έργο για αυτές τις εικόνες. Το μεγαλύτερο μέρος της σχετικής βιβλιογραφίας έχει στόχο να προτείνει μεθόδους ικανές να παρέχουν ικανοποιητική ακρίβεια ανίχνευσης. Ωστόσο, συνήθως δίνουν περισσότερη βαρύτητα σε μια συγκεκριμένη κατηγορία αντικειμένων, χωρίς να στοχεύουν τόσο στην ελαχιστοποίηση του χρόνου επεξεργασίας.

Το προκείμενο έργο, παρουσιάζει μια νέα και γρήγορη μεθοδολογική εναλλακτική λύση. Πέραν του ότι η προτεινόμενη μέθοδος χαρακτηρίζεται από ιδιαίτερη ταχύτητα, είναι επίσης μια γενική μέθοδος ανίχνευσης που χρήζει σχετικά καθολικής εφαρμογής που μπορεί να προσαρμοστεί σε οποιαδήποτε κατηγορία αντικειμένων μετά από κατάλληλη φάση εκπαίδευσης. Αποτελείται από το σχεδιασμό ενός μη γραμμικού φίλτρου που συνδυάζει χαρακτηριστικά μεταβαλλόμενης εικόνας με διαφορετική αλληλουχία και Gaussian process (GP) modelling. Οι βαθμίδες υψηλής ευκρίνειας της εικόνας επιτρέπουν την λεπτομερή καταγραφή πληροφοριών σχετικά με τη δομή της εξεταζόμενης κατηγορίας αντικειμένων. Το μοντέλο GP, που τροφοδοτείται με κλίσεις υψηλής τάξης, αποδίδει μια εκτίμηση της παρουσίας του αντικειμένου ενδιαφέροντος για μια δεδομένη θέση μέσα στην εικόνα. Στο πλαίσιο του παρόντος paper, πραγματοποιήθηκαν δύο ξεχωριστά σύνολα πειραμάτων, με στόχο την εκτίμηση της προτεινόμενης μεθόδου ανίχνευσης μιας δεδομένης κατηγορίας αντικειμένων κάθε φορά, συγκεκριμένα οχημάτων και ηλιακών συλλεκτών, αντίστοιχα. Τα αποτελέσματα σε πραγματικές εικόνες από μη επανδρωμένο εναέριο όχημα (UAV), ανάλυσης 2 cm και από τρία κανάλια (κόκκινο, πράσινο και μπλε-RGB) παρέχουν ιδιαίτερα ενδιαφέροντα στοιχεία στον ανιχνευτή.

Η ιδιαιτερότητα της εφαρμογής αυτής, έγκειται στο γεγονός πως το ύψος, στο οποίο πετά το μη επανδρωμένο αεροσκάφος είναι ψηλότερο από το σύνηθες. Στην λύση που προτείνεται εδώ, το υψόμετρο του UAV ανέρχεται περίπου στα δύο χιλιόμετρα, κάτι που δυσχεραίνει τον εντοπισμό των αντικειμένων. Όταν δε, πρόκειται για ανίχνευση αντικειμένων μικρότερης κλίμακας, όπως για παράδειγμα ανίχνευση ανθρώπων, τότε ο εντοπισμός μέσω της ανθρώπινης φιγούρας καθίσταται ιδιαίτερα απαιτητικός έως σχεδόν αδύνατος, αφού σε κάθε εικόνα τα pixels που αντιστοιχούν ανά άνθρωπο δεν είναι επαρκή. Ωστόσο, ο εντοπισμός των ατόμων μπορεί να διευκολυνθεί, χάρη στην παρατήρηση των σκιών που δημιουργούν οι φιγούρες.

Για την υλοποίηση, αρχικά χρησιμοποιείται ένας αλγόριθμος που εντοπίζει πιθανούς στόχους κάνοντας χρήση Maximally Stable External Regions (MSER). Κατόπιν οι ανιχνευθέντες στόχοι παρέχονται ως είσοδοι σε διάφορα φίλτρα προκειμένου να εντοπιστούν οι άνθρωποι και οι σκιές τους. Ο εντοπισμός ενός ανθρώπου προϋποθέτει την ύπαρξη ταιριαστού ζεύγους ανθρώπου-σκιάς. Για την εξαγωγή των επιθυμητών αποτελεσμάτων, απορρίπτονται περιπτώσεις όπου ο άνθρωπος είναι πολύ μεγαλύτερος ή πολύ μικρότερος σε σχέση με μια μέση περίπτωση. Για την εύρεση των σκιών χρησιμοποιούνται στοιχεία όπως η γωνία λήψης και η ημέρα και ώρα

που λήφθηκε η εικόνα. Μία περαιτέρω βελτίωση του ρυθμού εντοπισμού επιτυγχάνεται, λαμβάνοντας επίσης υπόψη τα μοτίβα αναγνώρισης, τα οποία προέρχονται από ιστορικά δεδομένα. Για περιπτώσεις όπως ο εντοπισμός πυρκαγιών, γίνεται χρήση θερμικής κάμερας, αφού οι υψηλές θερμοκρασίες επιτρέπουν την ανίχνευση με αρκετά μεγάλο ποσοστό ευστοχίας. Σε αυτή την περίπτωση, για να εντοπιστούν πιθανές εστίες πυρκαγιάς, χρησιμοποιείται ο MSER, όπως και με τους ανθρώπους. Επιπλέον, για την επιβεβαίωση πιθανής πυρκαγιάς χρησιμοποιείται καταρχάς η αντίστοιχη εικόνα που λήφθηκε από την έγχρωμη κάμερα, και ακολουθεί χρωματικός έλεγχος. [8]

2.3 DroneSAR Search & Rescue

Το DroneSAR–Search and Rescue είναι ένα λογισμικό, συμβατό με drone της εταιρίας DJI, όπως και η εφαρμογή που αναπτύχθηκε στα πλαίσια της παρούσας Διπλωματικής Εργασίας, το οποίο σχεδιάστηκε με σκοπό να μειώνει τους κινδύνους που εγκυμονεί μία πιθανή διάσωση προσωπικού και να αυξάνει τις πιθανότητες διάσωσης αγνοουμένων προσώπων. Η σκέψη πίσω από την σχεδίαση του προκειμένου λογισμικού είναι η προσφορά ενός εργαλείου του οποίου θα έχει την ικανότητα να ενισχύει το προσωπικό διάσωσης. Ωστόσο, η συγκεκριμένη καινοτομία αναμένεται να αλλάξει καθοριστικά τον τρόπο διεξαγωγής διασώσεων.

Το DroneSAR είναι σχεδιασμένο να παρέχει τις απαραίτητες πληροφορίες σε περιπτώσεις όπου ο χρόνος και οι πόροι είναι περιορισμένοι. Το λογισμικό λειτουργεί σύμφωνα με τις ανάγκες των πιλότων. Συγκεκριμένα, παρέχεται μια σειρά από επιλογές αποστολής, όπως το *customgrid function*, βάσει του οποίου ο πιλότος μπορεί να επιλέξει το μέγεθος και το σχήμα της επιθυμητής περιοχής. Επιπλέον, με το *Waypoint Mission* μπορούν να τοποθετηθούν σημεία πάνω στον χάρτη που να οριοθετούν την περιοχή στην οποία θα κινηθεί το drone και να του δίνουν ξεκάθαρη διαδρομή. Όπως και στην εφαρμογή *DJI GoTo* που δημιουργήθηκε για τον σκοπό της περάτωσης της προκείμενης Διπλωματικής Εργασίας, η τοποθέτηση σημείων στον χάρτη πραγματοποιείται μέσω αγγίγματος της οθόνης. Ο αριθμός σημείων που θα τοποθετηθούν στον χάρτη, εξαρτάται από τον πιλότο ή από τον εκάστοτε χρήστη της εφαρμογής. Μία ακόμα αποστολή που παρέχει η εφαρμογή *Search & Rescue* είναι η *Lat/Long Mission*, η οποία είναι χρήσιμη όταν ο χρήστης γνωρίζει το γεωγραφικό μήκος και πλάτος της περιοχής ενδιαφέροντος.

Κατά την διάρκεια εκτέλεσης κάθε αποστολής, παρέχονται πληροφορίες στον χρήστη μέσω της κινητής συσκευής που έχει εγκατασταθεί η εφαρμογή. Γίνεται ζωντανή μετάδοση βίντεο από το μέρος που πετάει το μη επανδρωμένο εναέριο όχημα και παρέχεται η δυνατότητα κοινοποίησης τοποθεσίας. [9]

Κεφάλαιο 3

Μηχανική Μάθηση (Machine Learning-ML)

Η μηχανική μάθηση είναι ο προγραμματισμός των υπολογιστών που επιτρέπει στις εφαρμογές λογισμικού να γίνουν ακριβέστερες στην πρόβλεψη αποτελεσμάτων χωρίς να είναι αναγκαία η ρητή τους καθοδήγηση. Η βασική ιδέα της μηχανικής μάθησης είναι η δημιουργία αλγορίθμων που μπορούν να λάβουν δεδομένα εισόδου και να χρησιμοποιήσουν στατιστική ανάλυση για να προβλέψουν μια έξοδο, ενώ φροντίζουν να ενημερώνουν τις εξόδους καθώς θα διατίθενται νέα δεδομένα. Ένα μεγάλο μειονέκτημα αρκετών προγραμματιστικών εργαλείων είναι ότι μένουν αμετάβλητα, ενώ τα δεδομένα που επεξεργάζονται μεταβάλλονται στο χρόνο. Αντίθετα, τα μοντέλα μηχανικής μάθησης προσαρμόζονται συνεχώς σε όλες τις αλλαγές μέσω των δεδομένων εισόδου που επεξεργάζονται, γεγονός που τα καθιστά εξαιρετικά χρήσιμα, ιδιαίτερα όταν πρόκειται για επεξεργασία μεγάλου όγκου και υψηλής πολυπλοκότητας δεδομένων.

Οι διαδικασίες που σχετίζονται με την μηχανική μάθηση είναι παρόμοιες με εκείνες της εξόρυξης δεδομένων και της πρότυπης μοντελοποίησης. Συγκεκριμένα, το κοινό χαρακτηριστικό είναι η αναγκαιότητα σάρωσης των δεδομένων για την εύρεση μοτίβων, έτσι ώστε να προσαρμοστούν αναλόγως οι ενέργειες του προγράμματος. Πολλοί άνθρωποι έρχονται σε επαφή με την έννοια της μηχανικής μάθησης, όταν μία αγορά τους από το διαδίκτυο, έχει ως αποτέλεσμα την προβολή διαφημίσεων που σχετίζονται με την αγορά αυτή σε διάφορες σελίδες στο διαδίκτυο. Αυτό συμβαίνει επειδή οι μηχανισμοί συστάσεων (Recommender Systems-RE) χρησιμοποιούν την μηχανική μάθηση για να προσαρμόσουν την εμφάνιση διαφημίσεων στο διαδίκτυο σχεδόν σε πραγματικό χρόνο. Πέρα από το εξατομικευμένο μάρκετινγκ, άλλες συνήθεις περιπτώσεις χρήσης μηχανικής μάθησης περιλαμβάνουν την ανίχνευση απάτης, το φιλτράρισμα ανεπιθύμητων μηνυμάτων, την ανίχνευση απειλών για την ασφάλεια δικτύων, την επεξεργασία αστρονομικών δεδομένων, τη πρόβλεψη του καιρού και τις μηχανές αναζήτησης στο διαδίκτυο. Οι παραπάνω αποτελούν μερικές μόνο από τις υπεράριθμες εφαρμογές που συμβάλλει η μηχανική μάθηση.

3.1 Η λειτουργία της μηχανικής μάθησης

Οι αλγόριθμοι μηχανικής μάθησης συχνά κατηγοριοποιούνται ως επιτηρούμενοι και μη επιτηρούμενοι. Οι εποπτευόμενοι αλγόριθμοι απαιτούν από έναν επιστήμονα δεδομένων ή από έναν αναλυτή δεδομένων με δεξιότητες μηχανικής μάθησης να παρέχουν τόσο την είσοδο, όσο και την επιθυμητή έξοδο, καθώς και τα απαραίτητα σχόλια σχετικά με την ακρίβεια των προβλέψεων κατά τη διάρκεια της εκπαίδευσης του αλγορίθμου. Οι αναλυτές δεδομένων προσδιορίζουν ποιες μεταβλητές ή χαρακτηριστικά, πρέπει να αναλυθούν και να χρησιμοποιηθούν για την ανάπτυξη των προβλέψεων. Μόλις ολοκληρωθεί η εκπαίδευση, ο

αλγόριθμος εφαρμόζει τις γνώσεις που έχει αποκτήσει κατά την εκπαίδευση, σε νεοεισελθέντα δεδομένα.

Οι μη επιτηρούμενοι αλγόριθμοι δεν χρειάζεται να εκπαιδευτούν με τα επιθυμητά αποτελέσματα εξόδου. Αντ' αυτού, χρησιμοποιείται μια επαναληπτική προσέγγιση για την ανασκόπηση των δεδομένων και την εξαγωγή συμπερασμάτων. Οι μη επιτηρούμενοι αλγόριθμοι μηχανικής μάθησης, οι οποίοι ονομάζονται επίσης νευρωνικά δίκτυα, παρέχουν λύσεις σε πιο σύνθετα θέματα επεξεργασίας από τα εποπτευόμενα συστήματα, συμπεριλαμβανομένης της αναγνώρισης εικόνων, της μετατροπής της ομιλίας σε κείμενο και της δημιουργίας φυσικών γλωσσών. Τα νευρωνικά δίκτυα λειτουργούν συνδυάζοντας εκατομμύρια παραδείγματα δεδομένων εκπαίδευσης, με σκοπό τον αυτόματο εντοπισμό συσχετίσεων μεταξύ πολλών μεταβλητών, γεγονός που αποτελεί ιδιαίτερος απαιτητικό έργο. Με την ολοκλήρωση της διαδικασίας της εκπαίδευσης, ο αλγόριθμος μπορεί να χρησιμοποιήσει την τράπεζα των γνώσεων που έχει προσκομίσει κατά την εκπαίδευση, για την ερμηνεία νέων δεδομένων. Αυτοί οι αλγόριθμοι είναι εφικτοί μόνο στην εποχή των μεγάλων δεδομένων (big data), καθώς απαιτούν τεράστιες ποσότητες δεδομένων εκπαίδευσης.

3.2 Παραδείγματα μηχανικής μάθησης

Οι υπολογιστές παρά την τεράστια υπολογιστή τους ισχύ, πολλές φορές υστερούν στην λήψη αποφάσεων σε σχέση με έναν άνθρωπο. Αυτό είναι το πρόβλημα αντιμετωπίζεται χάρη στην μηχανική μάθηση, η οποία χρησιμοποιείται σήμερα σε ένα ευρύ φάσμα εφαρμογών. Ένα από τα πιο γνωστά παραδείγματα είναι η σελίδα News Feed του Facebook. Η σελίδα αυτή χρησιμοποιεί τη μηχανική μάθηση για την εξατομίκευση κάθε μέλους. Εάν ένα μέλος σταματάει συχνά για να διαβάσει ή να αντιδράσει σε αναρτήσεις ενός συγκεκριμένου φίλου, το News Feed θα αρχίσει να δείχνει περισσότερα από τη δραστηριότητα του φίλου του υψηλότερα στη σελίδα. Στην πραγματικότητα, το λογισμικό χρησιμοποιεί απλώς στατιστική ανάλυση και αναλυτική πρόβλεψη για τον εντοπισμό μοτίβων και αξιοποίησή της ανάμεσα στα δεδομένα του χρήστη.

Η μηχανική μάθηση εισέρχεται επίσης σε μια σειρά επιχειρησιακών εφαρμογών. Τα συστήματα διαχείρισης σχέσεων πελατών (Customer Relationship Management-CRM) χρησιμοποιούν μαθησιακά μοντέλα για να αναλύσουν τα μηνύματα ηλεκτρονικού ταχυδρομείου και να προτρέψουν τα μέλη της ομάδας πωλήσεων να ανταποκριθούν ιεραρχικά στα πιο σημαντικά μηνύματα. Τα πιο προηγμένα συστήματα μπορούν ακόμη και να προτείνουν πιθανές αποτελεσματικές απαντήσεις. Επίσης, τα συστήματα ανθρώπινου δυναμικού (Human Resources-HR) χρησιμοποιούν αντίστοιχα μοντέλα για τον προσδιορισμό των χαρακτηριστικά των αποτελεσματικών υπαλλήλων, βάση των οποίων θα αναζητούν τους κατάλληλους ανθρώπους για τις ανοικτές θέσεις. Επιπλέον, η μηχανική μάθηση διαδραματίζει σημαντικό ρόλο στην αυτόματη οδήγηση οχημάτων.

3.3 Τύποι μοντέλων μηχανικής μάθησης

Η ύπαρξη σχεδόν απεριόριστων χρήσεων της μηχανικής μάθησης, έχει οδηγήσει στην ανάγκη δημιουργίας πληθώρας αλγορίθμων και μοντέλων μηχανικής μάθησης. Οι αλγόριθμοι είναι ποικίλων πολυπλοκοτήτων, με μερικούς από αυτούς να είναι αρκετά απλοί, αναφορικά με το

σκεπτικό και την χρήση και άλλους να είναι ιδιαίτερα περίπλοκοι. Το ίδιο συμβαίνει και με τα μοντέλα που χρησιμοποιούνται κατά κόρον σε εφαρμογές μηχανικής μάθησης. Μερικά από τα πιο συχνά χρησιμοποιούμενα μοντέλα παρουσιάζονται ακολούθως:

3.3.1 Δέντρα αποφάσεων

Αυτά τα μοντέλα χρησιμοποιούν παρατηρήσεις σχετικά με συγκεκριμένες ενέργειες και εντοπίζουν μια βέλτιστη διαδρομή για την επίτευξη ενός επιθυμητού αποτελέσματος. Ένα δέντρο απόφασης αποτελεί ένα εργαλείο υποστήριξης αποφάσεων που χρησιμοποιεί ένα μοντέλο τύπου δέντρου για την λήψη αποφάσεων και τις πιθανές συνέπειές τους, συμπεριλαμβανομένων των αποτελεσμάτων συμβάντων τυχαίων γεγονότων, του κόστους πόρων και της χρησιμότητας. Είναι ένας τρόπος προβολής ενός αλγορίθμου που περιέχει μόνο δηλώσεις υπό όρους ελέγχου.

3.3.2 K-means ομαδοποίηση (clustering)

Ανήκει στους μη εποπτευόμενους αλγορίθμους, το οποίο σημαίνει πως δεν απαιτείται από τον προγραμματιστή να παρέχει δεδομένα εκπαίδευσης, μιας και το μοντέλο καταλαβαίνει ποια είναι τα χαρακτηριστικά που διαφοροποιούν κάθε κλάση. Έτσι, το μοντέλο αυτό ομαδοποιεί έναν καθορισμένο αριθμό σημείων δεδομένων σε έναν συγκεκριμένο αριθμό ομάδων που βασίζονται σε παρόμοια χαρακτηριστικά. Το K αποτελεί τον αριθμό των διαφορετικών κλάσεων που προκύπτουν μετά την ομαδοποίηση. Επίσης, στο μοντέλο συναντώνται K διαφορετικά κέντρα (centroids), ο αριθμός των οποίων ισούται με το πλήθος των διαστάσεων του μοντέλου. Το κάθε σημείο ομαδοποιείται με την κλάση στην οποία το κέντρο βρίσκεται πλησιέστερα. Στη συνέχεια χρειάζεται να βρεθεί το κέντρο των δεδομένων κάθε κλάσης μετά την εισαγωγή των νέων και να μεταφερθεί το νέο κέντρο εκεί. Στην αρχή της διαδικασίας, το κέντρο τοποθετείται τυχαία στον χώρο. Έπειτα η διαδικασία επαναλαμβάνεται με το νέο κέντρο και πραγματοποιούνται οι απαιτούμενες παρεμβάσεις. Η ομαδοποίηση ολοκληρώνεται όταν δεν γίνονται πλέον άλλες παρεμβάσεις αναφορικά με την μεταφορά δεδομένων από μία κλάση σε άλλη. Παρόλο που είναι πιο αργός από τους αλγορίθμους επιτηρούμενης μάθησης, είναι ο γρηγορότερος από τους υπόλοιπους αλγορίθμους ομαδοποίησης.

3.3.3 Νευρωνικά δίκτυα (Neural Networks-NN)

Αποτελούν μοντέλα βαθιάς μάθησης, τα οποία χρησιμοποιούν μεγάλες ποσότητες δεδομένων εκπαίδευσης για τον εντοπισμό συσχετισμών μεταξύ πολλών μεταβλητών για να μάθουν να επεξεργάζονται εισερχόμενα δεδομένα στο μέλλον. Οι νευρώνες μπορούν να βρίσκονται σε μία από τις εξής δύο καταστάσεις, να πυροδοτούν ή όχι, γεγονός που μεταφράζεται σε 1 και 0 αντίστοιχα. Το νευρικό δίκτυο αποτελεί ένα συνδεδεμένο στρώμα, ή και στρώματα, νευρώνων, ενώ συναντώνται και κρυμμένα στρώματα. Το δίκτυο αυτό μπορεί να είναι πλήρως συνδεδεμένο, όταν κάθε νευρώνας κάθε στρώματος συνδέεται με όλους τους νευρώνες του επόμενου στρώματος ακριβώς μία φορά. Καθεμία από αυτές τις συνδέσεις έχει βάρος και ένα ακόμα μέγεθος που ονομάζεται προκατάληψη. Το πρώτο επίπεδο του δικτύου αποτελεί την είσοδό του και το τελευταίο την έξοδο. Κατά την διάρκεια εκπαίδευσης του δικτύου, ρυθμίζονται τα βάρη και οι προκαταλήψεις κάθε σύνδεσης έτσι ώστε να πάρουμε το σωστό αποτέλεσμα στην έξοδο. Τυπικά, ένα νευρωνικό δίκτυο χρειάζεται πολύ μεγάλο όγκο πληροφοριών προκειμένου να ολοκληρωθεί με επιτυχία η διαδικασία της εκπαίδευσης.

3.3.4 Μάθηση Ενδυνάμωσης (Reinforcement Learning)

Αυτός ο τομέας της μηχανικής μάθησης περιλαμβάνει μοντέλα που χρειάζονται πολλές προσπάθειες ολοκλήρωσης μιας διαδικασίας. Ασχολείται με τον τρόπο με τον οποίο οι πράκτορες λογισμικού πρέπει να αναλάβουν δράση σε ένα περιβάλλον, ώστε να μεγιστοποιήσουν κάποια έννοια σωρευτικής ανταμοιβής. Η ενίσχυση της μάθησης είναι ένα από τα τρία βασικά παραδείγματα εκμάθησης μηχανών, παράλληλα με την εποπτευόμενη και την μη επιτηρούμενη μάθηση. Διαφέρει από την εποπτευόμενη μάθηση στο ότι τα επισημασμένα ζεύγη εισόδου-εξόδου δεν χρειάζεται να παρουσιαστούν και οι υπο-βέλτιστες ενέργειες δεν χρειάζεται να διορθωθούν ρητά. Αντ' αυτού, η εστίαση είναι στην εύρεση ισορροπίας μεταξύ της εξερεύνησης (της άγνωστης επικράτειας) και της εκμετάλλευσης (της σημερινής γνώσης). Λειτουργεί ως εξής: τα βήματα που παράγουν ευνοϊκά αποτελέσματα επιβραβεύονται και τα βήματα που παράγουν ανεπιθύμητα αποτελέσματα τιμωρούνται μέχρι ο αλγόριθμος να μάθει τη βέλτιστη διαδικασία.

3.4 Το μέλλον της μηχανικής μάθησης

Παρόλο που οι αλγόριθμοι μηχανικής μάθησης έχουν κάνει την εμφάνισή τους εδώ και δεκαετίες, πλέον έχουν αποκτήσει νέα δημοτικότητα καθώς η τεχνητή νοημοσύνη (Artificial Intelligence-AI) έχει προαχθεί σε εξέχουσα θέση. Οι πλέον προηγμένες εφαρμογές AI βασίζονται σε μοντέλα βαθιάς μάθησης.

Οι πλατφόρμες μηχανικής μάθησης συγκαταλέγονται στις πιο ανταγωνιστικές τεχνολογίες, με τους περισσότερους σημαντικούς πωλητές, όπως οι Amazon, Google, Microsoft, IBM και άλλοι, να αγωνίζονται να αποκτήσουν πελάτες για υπηρεσίες πλατφόρμας που καλύπτουν το φάσμα δραστηριοτήτων μηχανικής μάθησης, συμπεριλαμβανομένων της συλλογής δεδομένων, της δημιουργία μοντέλων καθώς και της εκπαίδευση και ανάπτυξης εφαρμογών. Όσο η μηχανική μάθηση συνεχίζει να εξελίσσεται στις επιχειρησιακές λειτουργίες και η τεχνητή νοημοσύνη καθίσταται όλο και πιο πρακτική στις επιχειρηματικές ρυθμίσεις, ο ανταγωνισμός μεταξύ των παρόχων πλατφόρμας μηχανικής μάθησης αυξάνεται κατακόρυφα.

Ένας από τους πλέον αναπτυσσόμενους τομείς της μηχανικής μάθησης είναι η αναγνώριση αντικειμένων. Εκτός από την ανίχνευση αντικειμένων όπως τα αυτοκίνητα, το οποίο αποτελεί κύριο θέμα της παρούσας Διπλωματικής Εργασίας, ένας ακόμα κλάδος που αναπτύσσεται ραγδαία είναι αυτός της αναγνώρισης προσώπων, για την οποία κρίνεται σημαντικό να γίνει ειδικά αναφορά. [10]

3.5 Αναγνώριση προσώπου

Η αναγνώριση προσώπου είναι μια κατηγορία βιομετρικού λογισμικού που χαρτογραφεί μαθηματικά τα χαρακτηριστικά ενός προσώπου και αποθηκεύει τα δεδομένα ως faceprint. Το λογισμικό χρησιμοποιεί αλγόριθμους βαθιάς μάθησης για να συγκρίνει μια ζωντανή λήψη ή μια ψηφιακή εικόνα με την αποθηκευμένη αποτύπωση του προσώπου (faceprint) προκειμένου να επαληθεύσει την ταυτότητα ενός ατόμου.

Συγκεκριμένα, το λογισμικό αναγνωρίζει 80 κόμβους σε ένα ανθρώπινο πρόσωπο. Σε αυτό το πλαίσιο, τα σημεία κόμβων είναι τελικά σημεία που χρησιμοποιούνται για τη μέτρηση των μεταβλητών του προσώπου, όπως το μήκος ή το πλάτος της μύτης, το βάθος των οπών και το σχήμα των ζυγωματικών. Το σύστημα λειτουργεί καταγράφοντας δεδομένα για κομβικά σημεία από μια ψηφιακή εικόνα του προσώπου ενός ατόμου και αποθηκεύοντας τα δεδομένα που προκύπτουν ώστε να τα χρησιμοποιήσει ως βάση για τις μετέπειτα συγκρίσεις.

Ωστόσο, η αναγνώριση προσώπου ενδέχεται να εγείρει προβλήματα αναφορικά με την ιδιωτικότητα των ανθρώπων που εμπλέκονται. Αντίστοιχα θέματα ασφάλειας και προστασίας της ιδιωτικής ζωής μπορεί να προκύψουν όχι μόνο από την αναγνώριση ενός ανθρώπου, αλλά ακόμα και τον εντοπισμό κάποιου προσωπικού αντικειμένου όπως το αυτοκίνητο ή το σπίτι. Επί του παρόντος, στην Ευρώπη γίνονται ενέργειες για θέσπιση νόμων που θα άρουν πιθανές απειλές ως προς το δικαίωμα ιδιωτικότητας των ανθρώπων. Παρόλα αυτά, στις Ηνωμένες Πολιτείες δεν υπάρχουν νόμοι που να προστατεύουν ειδικά τα βιομετρικά δεδομένα ενός ατόμου. Τα συστήματα αναγνώρισης προσώπου μελετώνται για την ασφάλεια των αεροδρομίων και εκτιμάται ότι τα πρόσωπα της πλειονότητας των κατοίκων των Ηνωμένων Πολιτειών της Αμερικής έχουν ήδη αποτυπωθεί. Τα δεδομένα από ένα σύστημα αναγνώρισης προσώπου μπορεί να ληφθούν και να αποθηκευτούν εν αγνοία του ίδιου του ανθρώπου, γεγονός που μπορεί να αποβεί εξαιρετικά επικίνδυνο, μιας και τα δεδομένα αυτά θα μπορούσαν για παράδειγμα να χρησιμοποιηθούν από κυβερνητικές υπηρεσίες ή διαφημιστές για την παρακολούθηση των ατόμων.

Η αναγνώριση προσώπου αποτελεί μία υποκατηγορία της ανίχνευσης αντικειμένων, η οποία αναλύεται στην συνέχεια. [11]

Κεφάλαιο 4

Ανίχνευση Αντικειμένων

Η ανίχνευση αντικειμένων αποτελεί μια τεχνολογία που σχετίζεται με την υπολογιστική όραση και την επεξεργασία εικόνων, η οποία ασχολείται με την ανίχνευση περιπτώσεων σημασιολογικών αντικειμένων μιας συγκεκριμένης κλάσης, όπως για παράδειγμα άνθρωποι, αυτοκίνητα ή κτίρια, σε ψηφιακές εικόνες και βίντεο. Οι τομείς ανίχνευσης αντικειμένων, οι οποίοι χαίρουν περισσότερης δημοτικότητας και είναι οι πλέον ερευνημένοι, περιλαμβάνουν ανίχνευση προσώπου και ανίχνευση πεζών. Η ανίχνευση αντικειμένων έχει εφαρμογές σε πολλούς τομείς της υπολογιστικής όρασης, συμπεριλαμβανομένης της ανάκτησης εικόνων και της βιντεοεπιτήρησης, όπως η παρακολούθηση μιας μπάλας κατά τη διάρκεια ενός ποδοσφαιρικού αγώνα, η παρακολούθηση της κίνησης ενός μπαστουνιού χόκεϊ ή ακόμα και η παρακολούθηση ενός ατόμου σε ένα βίντεο. [12]

Η ανίχνευση ενός αντικειμένου είναι άρρηκτα συνδεδεμένη με την ταξινόμησή των αντικειμένων σε κλάσεις έτσι ώστε να παρέχεται η πληροφορία σχετικά με το είδος αντικειμένου που αναγνωρίστηκε. Κάθε τάξη αντικειμένων έχει τα δικά της ειδικά χαρακτηριστικά που βοηθούν στην ταξινόμηση της κλάσης. Ένα κλασικό παράδειγμα είναι πως όλοι οι κύκλοι είναι στρογγυλοί. Έτσι, όταν αναζητούνται κύκλοι, ψάχνουμε για αντικείμενα που βρίσκονται σε συγκεκριμένη απόσταση από ένα σημείο, το οποίο αποτελεί το κέντρο του κύκλου. Ομοίως, όταν αναζητούνται τετράγωνα, χρειαζόμαστε αντικείμενα που είναι κάθετα στις γωνίες και έχουν ίσα πλάτη πλευρών. Μια παρόμοια προσέγγιση χρησιμοποιείται για την αναγνώριση προσώπου όπου πρέπει να βρεθούν χαρακτηριστικά που κάνουν τους ανθρώπους ξεχωριστούς όπως τα μάτια, η μύτη, τα χείλη και χαρακτηριστικά όπως το χρώμα του δέρματος και η απόσταση μεταξύ των ματιών.

Οι μέθοδοι ανίχνευσης αντικειμένων εν γένει εμπίπτουν είτε σε προσεγγίσεις που βασίζονται στη μηχανική μάθηση είτε σε προσεγγίσεις βασιζόμενες στη βαθιά μάθηση. Για τις προσεγγίσεις μηχανικής μάθησης, καθίσταται απαραίτητο να οριστούν πρώτα τα χαρακτηριστικά και στη συνέχεια να χρησιμοποιηθεί μια τεχνική όπως η μηχανή φορέα υποστήριξης (SVM) για την ταξινόμηση. Από την άλλη πλευρά, οι τεχνικές βαθιάς εκμάθησης είναι σε θέση να κάνουν ανίχνευση αντικειμένων από άκρο σε άκρο χωρίς να καθορίζουν συγκεκριμένα χαρακτηριστικά και βασίζονται συνήθως σε Συνελικτικά Νευρωνικά Δίκτυα (CNN). Στην συνέχεια αναλύονται τεχνικές τόσο της μηχανικής όσο και της βαθιάς μάθησης, παρουσιάζονται οι δημοφιλέστεροι αλγόριθμοι ανίχνευσης αντικειμένων για καθεμία από τις δύο κατηγορίες και συγκρίνονται μεταξύ τους.

Ο ορισμός του προβλήματος της ανίχνευσης αντικειμένων είναι να προσδιοριστεί που βρίσκονται τα αντικείμενα σε μια δεδομένη εικόνα (εντοπισμός αντικειμένων) και σε ποια κατηγορία ανήκει το κάθε αντικείμενο (ταξινόμηση αντικειμένων). Έτσι, τα παραδοσιακά μοντέλα ανίχνευσης αντικειμένων, δηλαδή αυτά που βασίζονται στις αρχές της μηχανικής μάθησης, μπορούν να διαιρεθούν σε τρεις φάσεις: επιλογή περιοχής, εξαγωγή χαρακτηριστικών και ταξινόμηση.

Επιλογή περιοχής πληροφορίας:

Καθώς διαφορετικά αντικείμενα, τα οποία έχουν διαφορετικές αναλογίες και διαστάσεις, μπορεί να εμφανιστούν σε οποιαδήποτε θέση σε μία εικόνα, μία εύλογη επιλογή είναι η σάρωση ολόκληρης της εικόνας με ένα συρόμενο παράθυρο πολλαπλών κλιμάκων (multi-scale sliding window). Αν και αυτή η εξαντλητική στρατηγική μπορεί να βρει όλες τις πιθανές θέσεις των αντικειμένων, τα μειονεκτήματά της είναι προφανείς. Λόγω του μεγάλου αριθμού υποψηφίων παραθύρων, είναι υπολογιστικά εξαιρετικά ακριβό και παράγει επίσης πολλά περιττά παράθυρα. Από την άλλη πλευρά, αν εφαρμοστεί μόνο ένας σταθερός αριθμός πρότυπων συρόμενων παραθύρων, υπάρχει ο κίνδυνος να υπάρξουν περιοχές που να μην έχουν σαρωθεί.

Εξαγωγή χαρακτηριστικών:

Για την αναγνώριση διαφορετικών αντικειμένων, είναι απαραίτητη η εξαγωγή οπτικών χαρακτηριστικών, τα οποία μπορούν να παρέχουν μια σημασιολογική και πληροφοριακή αναγνώριση. Τα χαρακτηριστικά SIFT, HOG και Haar είναι τα πλέον αντιπροσωπευτικά και αναλύονται στην συνέχεια. Αυτό οφείλεται στο γεγονός ότι αυτά τα χαρακτηριστικά μπορούν να παράγουν αναπαραστάσεις που σχετίζονται με σύνθετα κύτταρα στον ανθρώπινο εγκέφαλο. Ωστόσο, λόγω της ποικιλομορφίας των εμφανίσεων, των συνθηκών φωτισμού και του περιβάλλοντος, η μη αυτόματη σχεδίαση ενός ισχυρού περιγραφέα στοιχείων που να περιγράφει τέλεια όλα τα είδη αντικειμένων, αποτελεί μία πρόκληση.

Ταξινόμηση:

Ένας ταξινομητής κρίνεται απαραίτητος για να διακρίνει ένα αντικείμενο-στόχο από τις άλλες κατηγορίες αντικειμένων και να κάνει τις αναπαραστάσεις πιο ιεραρχικές για οπτική αναγνώριση. Συνήθως η Μηχανή Υποστήριξης Διανυσμάτων (SUPPORT VECTOR MACHINE-SVM), το AdaBoost και το παραμορφώσιμο βάσει τεμαχίων μοντέλο (Deformable Part-based Model-DPM) είναι αξιόπιστες επιλογές. Μεταξύ αυτών των ταξινομητών, το DPM αποτελεί ένα ευέλικτο μοντέλο, μιας και συνδυάζει τα μέρη αντικειμένων με το κόστος παραμόρφωσης για την αντιμετώπιση σοβαρών παραμορφώσεων.

4.1 Αλγόριθμοι μηχανικής μάθησης για ανίχνευση αντικειμένων

Υπάρχουν πολλές προσεγγίσεις μηχανικής μάθησης, τις οποίες μπορούμε να χρησιμοποιήσουμε για να ανιχνεύσουμε ένα αντικείμενο σε μια εικόνα. Ακολούθως παρουσιάζεται μία λίστα από τους πλέον διαδεδομένους αλγορίθμους, οι οποίοι χρησιμοποιούνται στο πρώτο στάδιο της ανίχνευσης αντικειμένων, δηλαδή στον ορισμό των χαρακτηριστικών των αντικειμένων όπως αναφέρθηκε παραπάνω :

4.1.1 Ιστόγραμμα προσανατολισμένων διαβαθμίσεων (Histogram of Oriented Gradients-HOG)

Το ιστόγραμμα προσανατολισμένων διαβαθμίσεων (HOG) είναι ένας περιγραφέας χαρακτηριστικών που χρησιμοποιείται στην υπολογιστική όραση και την επεξεργασία εικόνας για τον σκοπό της ανίχνευσης αντικειμένων. Η τεχνική αυτή μετρά περιστατικά προσανατολισμού κλίσης σε εντοπισμένα τμήματα μιας εικόνας. Αυτή η μέθοδος είναι παρόμοια με εκείνη των ιστογραμμάτων προσανατολισμού άκρου, των περιγραφέν μετασχηματισμού κλίμακας αμετάβλητων χαρακτηριστικών και των πλαισίων σχήματος, αλλά διαφέρει στο ότι υπολογίζεται σε ένα πυκνό πλέγμα ομοιόμορφα τοποθετημένων κυψελών και χρησιμοποιεί επικαλυπτόμενη τοπική κανονικοποίηση αντίθεσης για την επίτευξη της μέγιστης δυνατής ακρίβειας.

Προκειμένου να μπορέσουμε να εντοπίσουμε το σύνολο των αυτοκινήτων σε μία εικόνα χρησιμοποιώντας τον συγκεκριμένο αλγόριθμο, εκπαιδεύουμε έναν ανιχνευτή αντικειμένων HOG και τον σέρνουμε πάνω από την επιθυμητή εικόνα κάθε φορά. Η προσέγγιση αυτή είναι παλαιότερη και αδυνατεί να χειριστεί με επιτυχία κάποια πιο πολύπλοκα σενάρια, όμως παρόλα αυτά, είναι σχετικά γρήγορη. Στο πρόβλημα της ανίχνευσης αυτοκινήτων από μία εικόνα για παράδειγμα, ο αλγόριθμος αντιμετωπίζει δυσκολία στον εντοπισμό αυτοκινήτων που περιστρέφονται σε διαφορετικούς προσανατολισμούς.

4.1.2 Πλαίσιο ανίχνευσης αντικειμένων Viola-Jones

Αποτελεί το πρώτο πλαίσιο ανίχνευσης αντικειμένων που παρέχει ανταγωνιστικά ποσοστά ανίχνευσης αντικειμένων σε πραγματικό χρόνο. Παρόλο που μπορεί να εκπαιδευτεί για να ανιχνεύσει μια ποικιλία κατηγοριών αντικειμένων, είχε ως κύριο σκοπό την ανίχνευση προσώπων. Η διαδικασία αναγνώρισης ενός προσώπου είναι εύκολη όταν ένας άνθρωπος καλείται να το κάνει χρησιμοποιώντας τις αισθήσεις του. Ωστόσο, ένας υπολογιστής χρειάζεται ακριβείς οδηγίες και περιορισμούς. Για να καταστεί η εργασία πιο διαχειρίσιμη, το συγκεκριμένο πλαίσιο απαιτεί πλήρη μετωπική όψη των προσώπων. Έτσι, για να ανιχνευθεί, ολόκληρο το πρόσωπο πρέπει να δείχνει προς την κάμερα και δεν πρέπει να είναι κεκλιμένο σε καμία πλευρά. Οι περιορισμοί αυτοί φαινομενικά θα μπορούσαν να μειώσουν σε κάποιο βαθμό την χρησιμότητα του αλγορίθμου, επειδή το βήμα ανίχνευσης ακολουθείται συχνότερα από ένα βήμα αναγνώρισης, στην πράξη όμως αυτά τα όρια είναι αποδεκτά, καθιστώντας το πλαίσιο αρκετά αποτελεσματικό.

Παραπάνω παρουσιάστηκαν δύο εκ των σημαντικότερων προσεγγίσεων της μηχανικής μάθησης. Μία ακόμα δημοφιλή προσέγγιση αποτελεί ο **Μετασχηματισμός Χαρακτηριστικών Αμετάβλητης Κλίμακας (Scale-invariant feature transform-SIFT)**.

4.2 Τεχνικές ταξινόμησης

Αφού πρώτα έχουμε εξαγάγει τα χαρακτηριστικά που διαφοροποιούν κάθε κλάση με χρήση κάποιου από τους αλγορίθμους που μόλις παρουσιάστηκαν, στις προσεγγίσεις μηχανικής μάθησης είναι απαραίτητο να χρησιμοποιήσουμε εκ νέου κάποιες τεχνικές με σκοπό την

ταξινόμηση και ομαδοποίηση των αντικειμένων στις επιθυμητές κλάσεις. Αυτό επιτυγχάνεται με χρήση τεχνικών όπως είναι ο KNN ή ο SVM.

4.2.1 Αλγόριθμος KNN (Κ Κοντινότεροι Γείτονες)

Αποτελεί μία τεχνική ταξινόμησης, όπου το k είναι ένας ακέραιος αριθμός, ο οποίος αναπαριστά τον αριθμό των πλησιέστερων γειτόνων, τους οποίους καλείται να ψάξει ο αλγόριθμος. Ο ακέραιος αυτός θα πρέπει να είναι μονός αριθμός, έτσι ώστε να είμαστε σε θέση να διαλέγουμε πάντα την κλάση που περιλαμβάνει τα περισσότερα στοιχεία. Η επιλογή του k πρέπει να γίνεται με ιδιαίτερη προσοχή. Αν επιλεγεί πολύ μεγάλη τιμή για τον ακέραιο, τα αποτελέσματα μπορεί να είναι παραπλανητικά, διότι η κλάση που θα ταίριαζε περισσότερο στο στοιχείο που επιθυμούμε να ταξινομήσουμε, ενδέχεται να μην διαθέτει αρκετά μεγάλο αριθμό στοιχείων. Κάθε φορά χρειάζεται να υπολογίσουμε την απόσταση του στοιχείου προς ταξινόμηση με όλα τα υπόλοιπα στοιχεία.

4.2.2 Μηχανή Υποστήριξης Διανυσμάτων (SUPPORT VECTOR MACHINE-SVM)

Ο αλγόριθμος αυτός χρησιμοποιείται για ταξινόμηση δεδομένων, καθώς προσπαθεί να δημιουργήσει hyperplane: ένας γραμμικός τρόπος για να διαιρέσουμε τα δεδομένα μας. Το hyperplane, το οποίο ουσιαστικά αποτελεί μια ευθεία γραμμή που χωρίζει τον χώρο σε επίπεδα, πρέπει να απέχει την ίδια απόσταση από τα δύο κοντινότερα σημεία κάθε κατηγορίας. Μπορούμε να δημιουργήσουμε απεριόριστα hyperplanes, το ιδανικό ωστόσο είναι να δημιουργηθεί ένα που έχει τη μεγαλύτερη δυνατή απόσταση από ένα από τα σημεία κάθε αντίθετης τάξης. Όσο μεγαλύτερη είναι αυτή η απόσταση, και κατ' επέκταση όσο μεγαλύτερο είναι το περιθώριο, μπορούμε να διαχωρίσουμε με μεγαλύτερη ακρίβεια κάθε τάξη, έτσι ώστε η τελική ταξινόμηση να είναι αποδοτικότερη.

Τα σημεία που είναι τα πλησιέστερα σε κάθε πλευρά είναι τα διανύσματα υποστήριξης. Στην πραγματικότητα τα δεδομένα δεν είναι ιδανικά για το σχεδιασμό ενός hyperplane, για αυτό πρέπει να μετατραπούν σε μια μορφή, η οποία να επιτρέπει την ακριβέστερη δυνατή σχεδίασή του. Για το σκοπό αυτό, χρησιμοποιούμε μια λειτουργία πυρήνα. $f(x, y) = z$, προσθέτοντας μια επιπλέον διάσταση. Τότε κάθε σημείο θα έχει συντεταγμένες (x, y, z) . Χάρη στην προσθήκη διάστασης, μπορούμε να διαιρέσουμε με επιτυχία τα δεδομένα κάθε τάξης και να δημιουργήσουμε ένα hyperplane, το οποίο θα είναι επιφάνεια (δύο διαστάσεων).

Πέρα από τους αλγορίθμους που βασίζονται στις αρχές της μηχανικής μάθησης, υπάρχουν και εκείνοι που αναπτύσσονται σύμφωνα με τις αρχές της βαθιάς μάθησης. Στην συνέχεια επεξηγείται ακροθιγώς τι είναι τα Συνελικτικά Νευρωνικά Δίκτυα (CNN) και παρουσιάζονται μερικά από τα πλεονεκτήματα των τεχνικών βαθιάς μάθησης έναντι αυτών μηχανικής μάθησης. Επίσης αναλύονται μερικοί από τους σημαντικότερους αλγορίθμους βαθιάς μάθησης για ανίχνευση αντικειμένων.

4.3 CNN (Convolutional Neural Network / Συνελικτικό Νευρωνικό Δίκτυο)

Στη βαθιά μάθηση, ένα Συνελικτικό Νευρωνικό Δίκτυο (CNN, ή ConvNet) είναι μια κλάση βαθιά νευρωνικών δικτύων, που συνήθως εφαρμόζεται στην ανάλυση οπτικών εικόνων. Τα CNN είναι κανονικοποιημένες εκδόσεις πολλαπλών στρώσεων, γεγονός το οποίο συνήθως σημαίνει πλήρως συνδεδεμένα δίκτυα, δηλαδή κάθε νευρώνας σε ένα στρώμα συνδέεται με όλους τους νευρώνες στο επόμενο στρώμα. Η πλήρης σύνδεση αυτών των δικτύων τα καθιστά επιρρεπή στην υπερφόρτωση δεδομένων. Ένας τυπικός τρόπος κανονικοποίησης θα μπορούσε να περιλαμβάνει την προσθήκη κάποιας μορφής μέτρησης βαρών για τη αποφυγή απωλειών. Ωστόσο, το CNN υιοθετεί μια διαφορετική προσέγγιση όσον αφορά τη κανονικοποίηση, μιας και εκμεταλλεύεται το ιεραρχικό πρότυπο των δεδομένων και συγκεντρώνει τα πιο περίπλοκα μοτίβα χρησιμοποιώντας μικρότερα και απλούστερα πρότυπα. Ως εκ τούτου, στην κλίμακα της συνδεσιμότητας και της πολυπλοκότητας, τα CNN βρίσκονται στο κατώτερο άκρο.

Τα Συνελικτικά Νευρωνικά Δίκτυα είναι επίσης γνωστά ως αμετάβλητα ως προς την μετατόπιση ή τον χώρο τεχνητά νευρωνικά δίκτυα (Shift Invariant or Space Invariant Artificial Neural Networks SIANN), βασισμένα στην αρχιτεκτονική των κοινών βαρών τους και στα χαρακτηριστικά αμετάβλητης μετάφρασης. Τα συνελικτικά δίκτυα εμπνεύστηκαν από βιολογικές διεργασίες, αφού το πρότυπο σύνδεσης μεταξύ νευρώνων μοιάζει με την οργάνωση του ζωτικού οπτικού φλοιού. Οι μεμονωμένοι φλοιώδεις νευρώνες ανταποκρίνονται σε ερεθίσματα μόνο σε μια περιορισμένη περιοχή του οπτικού πεδίου που είναι γνωστή ως το δεκτικό πεδίο. Τα δεκτικά πεδία διαφορετικών νευρώνων επικαλύπτονται εν μέρει έτσι ώστε να καλύπτουν ολόκληρο το οπτικό πεδίο.

Τα CNN χρησιμοποιούν σχετικά μικρή προεπεξεργασία σε σύγκριση με άλλους αλγόριθμους ταξινόμησης εικόνων. Αυτό σημαίνει ότι το δίκτυο μαθαίνει τα φίλτρα που στους παραδοσιακούς αλγόριθμους κατασκευάστηκαν χειρωνακτικά. Η ανεξαρτησία από προηγούμενες γνώσεις και ανθρώπινη προσπάθεια στον σχεδιασμό χαρακτηριστικών αποτελεί σημαντικό πλεονέκτημα. [13]

4.4 Ανίχνευση αντικειμένων με χρήση τεχνικών βαθιάς μάθησης

Η ανίχνευση αντικειμένων, η οποία αποτελεί ένα από τα θεμελιώδη προβλήματα υπολογιστικής όρασης, είναι σε θέση να παρέχει πολύτιμες πληροφορίες για τη σημασιολογική κατανόηση των εικόνων και βίντεο, καθώς επίσης σχετίζεται με πολλές εφαρμογές, συμπεριλαμβανομένων της ταξινόμησης εικόνων, της ανάλυσης ανθρώπινης συμπεριφοράς, της αναγνώριση προσώπων και της αυτόνομης οδήγησης. Η πρόοδος σε αυτούς τους τομείς, κληρονομώντας χαρακτηριστικά από τα νευρωνικά δίκτυα και συναφή συστήματα μάθησης, αναπτύσσει αλγόριθμους νευρωνικού δικτύου, και έχει επίσης μεγάλες επιπτώσεις σε τεχνικές ανίχνευσης αντικειμένων που μπορούν να θεωρηθούν ως συστήματα μάθησης.

Οι παραδοσιακές μέθοδοι ανίχνευσης αντικειμένων βασίζονται σε επιφανειακά εκπαιδευμένες αρχιτεκτονικές. Οι επιδόσεις τους μειώνονται σημαντικά, όταν σαν είσοδος

παρέχονται σύνθετα σύνολα που συνδυάζουν πολλαπλά χαμηλού επιπέδου χαρακτηριστικά εικόνων, με υψηλού επιπέδου περιεχόμενο από ανιχνευτές αντικειμένων. Η ταχεία ανάπτυξη της βαθιάς μάθησης, είχε σαν αποτέλεσμα την δημιουργία ισχυρότερων εργαλείων για την αντιμετώπιση των υφιστάμενων προβλημάτων σε παραδοσιακές αρχιτεκτονικές, τα οποία είναι σε θέση να διαχειρίζονται χαρακτηριστικά υψηλού επιπέδου πολύ αποτελεσματικότερα.

4.4.1 Σύγκριση τεχνικών μηχανικής μάθησης έναντι αυτών βαθιάς μάθησης

Ενδεικτικά, κάποια από τα πλεονεκτήματα της χρήσης Συνελικτικών Νευρωνικών Δικτύων (CNN) έναντι παραδοσιακών μεθόδων συνοψίζονται στα ακόλουθα:

- Η ιεραρχική αναπαράσταση χαρακτηριστικών, η οποία είναι οι πολυεπίπεδες αναπαραστάσεις από τα *pixel* σε σημασιολογικά χαρακτηριστικά υψηλού επιπέδου που αντλούνται από μια ιεραρχική δομή πολλαπλών σταδίων. Μπορεί να μάθει από τα δεδομένα αυτόματα και οι κρυφοί παράγοντες των δεδομένων εισόδου μπορούν να απελευθερωθούν μέσω μη γραμμικών αντιστοιχίσεων πολλαπλών επιπέδων
- Σε σύγκριση με τα παραδοσιακά επιφανειακά μοντέλα, μια βαθύτερη αρχιτεκτονική παρέχει εκθετικά αυξημένες ικανότητες.
- Η αρχιτεκτονική του CNN παρέχει την δυνατότητα από κοινού βελτιστοποίησης διαφόρων εργασιών, οι οποίες χαρακτηρίζονται από συνάφεια ως προς την υλοποίηση (π.χ., το γρήγορο RCNN-Fast RCNN συνδυάζει την ταξινόμηση και την κλίση του πλαισίου οριοθέτησης).
- Κάνοντας χρήση της μεγάλης ικανότητας εκμάθησης των *deep CNNs*, μερικές κλασσικές προκλήσεις υπολογιστικής όρασης μπορούν να αναδιατυπωθούν ως προβλήματα μετασχηματισμού δεδομένων μεγάλων διαστάσεων και να λυθούν από διαφορετική οπτική γωνία.

Χάρη σε αυτά τα πλεονεκτήματα, το CNN εφαρμόστηκε ευρέως σε πολλούς τομείς έρευνας, όπως η υπέρ-ανάλυση εικόνας, η ταξινόμηση ή η ανάκτηση εικόνων, η αναγνώριση προσώπων, η ανίχνευση πεζών και η ανάλυση βίντεο.

4.4.2 Αλγόριθμοι ανίχνευσης αντικειμένων βαθιάς μάθησης

Οι σημαντικότεροι αλγόριθμοι βαθιάς μάθησης που χρησιμοποιούνται κατά κόρων στην ανίχνευση αντικειμένων είναι οι εξής:

- Αλγόριθμοι Προτάσεων Περιοχής (R-CNN, Fast R-CNN, Faster R-CNN, Mask R-CNN)
- Ανιχνευτής Πολλαπλών Θυρίδων μιας Λήψης (Single Shot MultiBox Detector -SSD)
- You Only Look Once (YOLO)

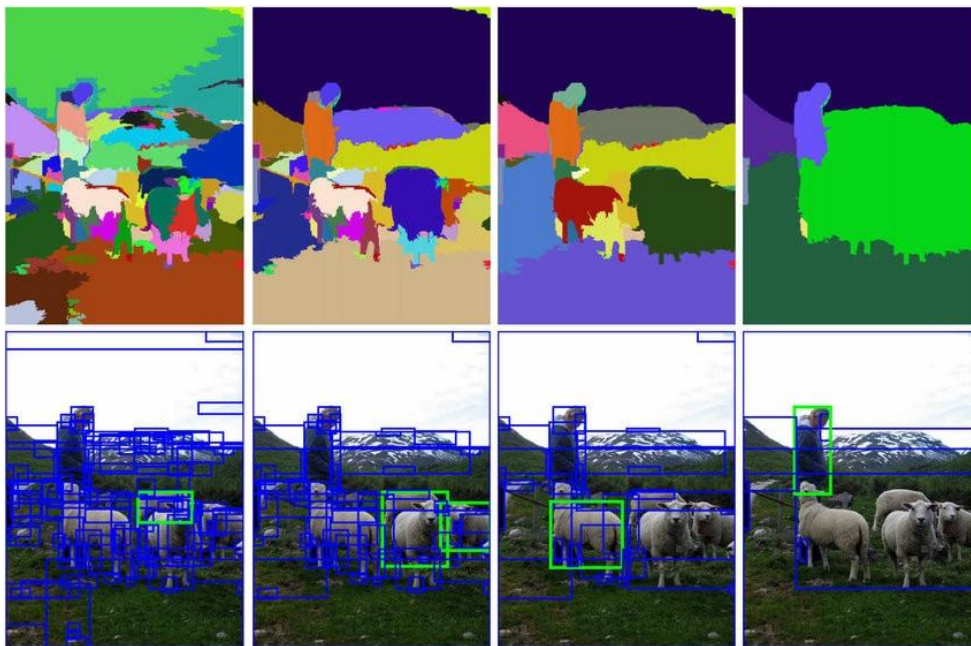
Αλγόριθμοι όπως Faster R-CNN, Mask R-CNN ή YOLO αποτελούν μια νεότερη, *deep-learning* προσέγγιση που συνδυάζει την ακρίβεια των CNN με έξυπνη σχεδίαση και αποτελεσματικότητα. Το γεγονός αυτό επιταχύνει σημαντικά τη διαδικασία ανίχνευσης, η οποία είναι σχετικά γρήγορη,

όταν πρόκειται για μία κάρτα γραφικών (Graphics Processing Unit-GPU), εφόσον έχουμε αρκετά δεδομένα για την εκπαίδευση του μοντέλου. [14]

4.4.2.1 R-CNN

Μία από τις πρώτες καινοτομίες της χρήσης των Συνελικτικών Νευρωνικών Δικτύων-CNN σε ένα σύστημα ανίχνευσης αντικειμένων είναι το R-CNN ή "Regions with CNN", το οποίο κατάφερε να πετύχει πολύ υψηλότερη απόδοση ανίχνευσης αντικειμένων από άλλες δημοφιλείς μεθόδους κατά την εποχή δημιουργίας του.

Το πρώτο στάδιο του R-CNN pipeline είναι η δημιουργία προτάσεων περιοχής (region proposals), δηλαδή η υπόθεση πως κάποια περιοχή σε μια εικόνα θα μπορούσε να ανήκει σε ένα συγκεκριμένο αντικείμενο. Προκειμένου να προταθούν οι διάφορες περιοχές, συχνά γίνεται χρήση του αλγορίθμου επιλεκτικής αναζήτησης. Ο αλγόριθμος επιλεκτικής αναζήτησης δημιουργεί διατμήσεις της εικόνας που θα μπορούσαν να ανήκουν σε ένα αντικείμενο, βασισμένες στο χρώμα, την υφή, το μέγεθος και το σχήμα, και συνδυάζει παρόμοιες περιοχές προκειμένου να σχηματίσουν αντικείμενα. Με αυτόν τον τρόπο δημιουργούνται προτάσεις αντικειμένων με διαφορετικές κλίμακες. [15]



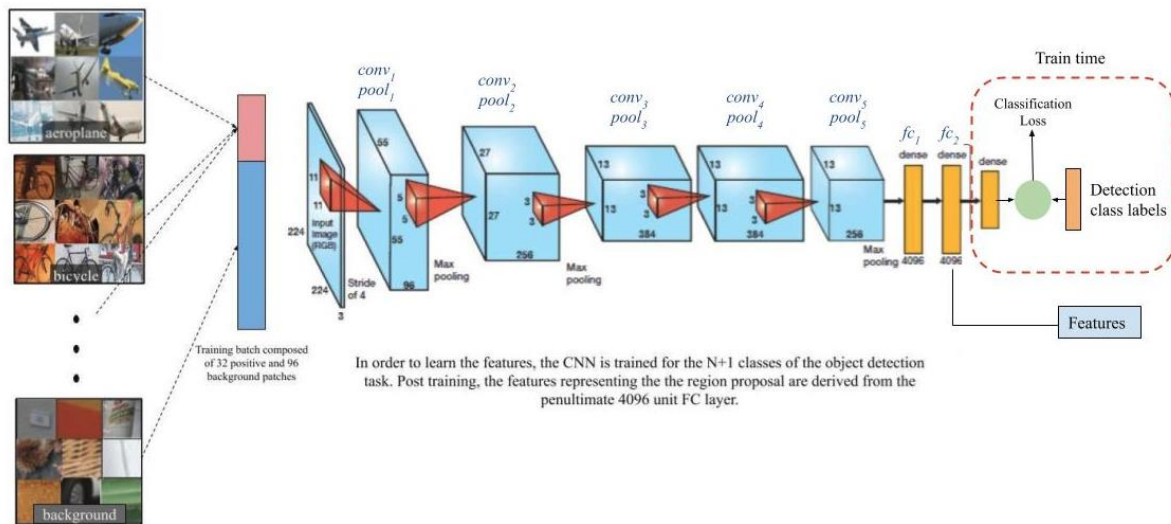
Σχήμα 4.4.2.1.1: Ο αλγόριθμος επιλεκτικής αναζήτησης λαμβάνει κατ' επανάληψη προτάσεις περιοχής

Το πρώτο στάδιο στο pipeline του αλγορίθμου R-CNN αφορά την εξαγωγή χαρακτηριστικών από τις προτάσεις περιοχής. Το δίκτυο πρέπει να ρυθμιστεί με ακρίβεια για να εκπαιδευτεί αφενός ως προς τα οπτικά χαρακτηριστικά των νέων τύπων εικόνων - παραμορφωμένες προτάσεις περιοχής και αφετέρου ως προς ειδικές κατηγορίες στόχων του μικρότερου συνόλου δεδομένων για την διαδικασία ανίχνευσης.

Όλες οι προτάσεις της περιοχής που δημιουργούνται από κάθε εικόνα, ως αποτέλεσμα χρήσης του αλγορίθμου επιλεκτικής αναζήτησης, μετατρέπονται σε εισόδους σταθερού μεγέθους, ανεξάρτητα από το πραγματικό μέγεθος ή την αναλογία διαστάσεων, προκειμένου να

τελειοποιηθεί το CNN. Μια πρόσθετη παράμετρος p χρησιμοποιείται για να υποδείξει την ποσότητα της πιθανής διαστολής του αρχικού πλαισίου οριοθέτησης ώστε να συμπεριλάβει κάποιο πλαίσιο από την περιοχή γύρω από αυτήν. Το κάθε πλαίσιο λαμβάνει μία ετικέτα ταξινόμησης, ανάλογα με το ποσοστό ταύτισης με κάποιο αντικείμενο. Σε περίπτωση που το ποσοστό αυτό είναι χαμηλότερο από κάποιο κατώφλι, τότε το πλαίσιο αντιμετωπίζεται ως φόντο.

Στο τέλος του πρώτου σταδίου, το δίκτυο εκπαιδεύεται με ρυθμό εκμάθησης αρκετά μικρότερο από εκείνον στην αρχή της διαδικασίας εκπαίδευσης. Σε κάθε επανάληψη δοκιμάζεται ένας αριθμός παραθύρων τα οποία είναι θετικά σε όλες τις κατηγορίες και μερικά που ανήκουν στην τάξη του φόντου, για να διασφαλιστεί ότι υπάρχει επαρκής αναπαράσταση από τις θετικές κατηγορίες κατά τη διάρκεια της εκπαίδευσης. [16], [17]



Σχήμα 4.4.2.1.2: Το pipeline εκπαίδευσης του R-CNN, λαμβάνει μια σειρά προτάσεων περιοχής και παρέχει μια ετικέτα ταξινόμησης

4.4.2.2 Fast R-CNN

Το Fast R-CNN αποτελεί έναν αλγόριθμο ανίχνευσης αντικειμένων, ο οποίος βασίζεται στις αρχές της βαθιάς μάθησης και ουσιαστικά αποτελεί διάδοχο του R-CNN. Οι τεχνολογικές εξελίξεις και η σημαντική αύξηση της υπολογιστικής δύναμης από την δημοσίευση της πρώτης εργασίας αναφορικά με το R-CNN, μας βοηθούν να μπορούμε εύκολα να ανιχνεύσουμε τα προφανή μειονεκτήματα των R-CNNs.

Καταρχάς για την εκπαίδευση των R-CNN αλγορίθμων, απαιτείται δαπανηρή εκπαίδευση πολλαπλών σταδίων. Οι ξεχωριστές διαδικασίες εκπαίδευσης που απαιτούνται για όλα τα στάδια του δικτύου, η τελειοποίηση ενός CNN σε προτάσεις περιοχής αντικειμένων, η εκμάθηση ενός SVM για την ταξινόμηση του διανύσματος χαρακτηριστικών κάθε πρότασης από το CNN

αποδεικνύεται επιβάρυνση χρόνου και πόρων. Για παράδειγμα, για να εκπαιδεύσουμε το SVM, θα χρειαζόμασταν τα χαρακτηριστικά χιλιάδων πιθανών προτάσεων περιοχής να έχουν γραφτεί στο δίσκο από το προηγούμενο στάδιο. Ένα επιπλέον αρνητικό χαρακτηριστικό του αλγορίθμου R-CNN είναι ο αργός χρόνος δοκιμής. Δεδομένου του αγωγού πολλαπλών σταδίων, η ανίχνευση χρησιμοποιώντας ένα απλό δίκτυο VGG (Ομάδα Οπτικής Γεωμετρίας-Visual Geometry Group) ως τη CNN βάση, διαρκεί περί τα 47 δευτερόλεπτα ανά εικόνα. [18]

Μία μελέτη που δημοσιεύθηκε το 2015, το Spatial Pyramid Pooling Networks (SPP-Net), επιλύει ένα μέρος του προβλήματος. Η κύρια συμβολή της εργασίας ήταν καταρχάς η χρήση πολυδιάστατης συγκέντρωσης χαρακτηριστικών εικόνας για τη βελτίωση της απόδοσης του δικτύου στην ταξινόμηση και την ανίχνευση καθώς επίσης και η χρήση εικόνων εισόδου αυθαίρετου μεγέθους για την κατάρτιση του CNN παρά τη χρήση πλήρως συνδεδεμένων (Fully Connected-FC) στρωμάτων. Στον αγωγό ανίχνευσης, εισήρθε μια σημαντική έννοια, η εξαγωγή ενός ενιαίου χάρτη χαρακτηριστικών από ολόκληρη την εικόνα μόνο μία φορά και συγκέντρωση των χαρακτηριστικών των προτάσεων περιοχής αυθαίρετου μεγέθους για αυτήν την εικόνα, από αυτόν τον ενιαίο χάρτη χαρακτηριστικών. Αυτό επέτρεψε στο δίκτυο να εκτελεί εντολές ταχύτερα από το R-CNN. Αντί να εκτελούνται CNN προτάσεις περιοχής 2000 από μια εικόνα ξεχωριστά, όπως το R-CNN, υπάρχει μόνο μία προώθηση, από την οποία μπορούν να εξαχθούν τα χαρακτηριστικά όλων των υποψηφίων προτάσεων. [19]

Fast R-CNN pipeline

Λεπτομέρειες Αρχιτεκτονικής

Παρακάτω παρουσιάζονται μερικά στοιχεία της αρχιτεκτονικής του αλγορίθμου Fast R-CNN, προκείμενου να γίνουν αντιληπτοί οι λόγοι βελτίωσης των επιδόσεων σε σχέση με αυτές των δικτύων R-CNN.

Το Fast R-CNN αποτελείται από ένα CNN (συνήθως προ-εκπαιδευμένο στην εργασία ταξινόμησης ImageNet) με το τελευταίο στρώμα συγκέντρωσης να αντικαθίσταται από ένα στρώμα συγκέντρωσης ROI (ROI pooling) και το τελικό του στρώμα FC από δύο κλάδους, $a(K+1)$ κατηγορίας softmax και έναν κλάδο παλινδρόμησης συγκεκριμένης κατηγορίας.

Ολόκληρη η εικόνα τροφοδοτείται στη βάση (backbone) του CNN και εξάγονται τα χαρακτηριστικά από την τελευταία στρώση συνέλιξης. Ανάλογα με το CNN που χρησιμοποιείται, οι χάρτες χαρακτηριστικών εξόδου είναι πολύ μικρότεροι από το αρχικό μέγεθος εικόνας. Αυτό εξαρτάται από το βήμα της ραχοκοκαλιάς του CNN, το οποίο είναι συνήθως 16 στην περίπτωση μιας βάσης VGG. Εν τω μεταξύ, λαμβάνονται τα παράθυρα προτάσεων αντικειμένων από έναν αλγόριθμο πρότασης περιοχής όπως ο αλγόριθμος επιλεκτικής αναζήτησης. Στη συνέχεια, το τμήμα του χάρτη χαρακτηριστικών του κορμού που ανήκει σε ένα παράθυρο εισάγεται στο στρώμα συγκέντρωσης ROI.

Το στρώμα συγκέντρωσης ROI είναι μια ειδική περίπτωση του στρώματος συγκέντρωσης πυραμίδων (Spatial Pyramid Pooling-SPP) με ένα μόνο επίπεδο πυραμίδας. Το στρώμα διαιρεί τα χαρακτηριστικά από τα επιλεγμένα παράθυρα πρότασης, τα οποία προέρχονται από τον αλγόριθμο πρότασης περιοχής, σε υποπαράθυρα μεγέθους u / Y και π / Π και εκτελεί μια διαδικασία συγκέντρωσης σε κάθε ένα από αυτά τα υποπαράθυρα. Αυτό οδηγεί σε χαρακτηριστικά εξόδου σταθερού μεγέθους ($Y \times \Pi$) ανεξάρτητα από το μέγεθος της εισόδου. Τα Y και Π επιλέγονται έτσι ώστε η έξοδος να είναι συμβατή με το πρώτο πλήρως συνδεδεμένο

επίπεδο του δικτύου. Όπως η κανονική συγκέντρωση, η συγκέντρωση ROI πραγματοποιείται σε κάθε κανάλι ξεχωριστά.

Τα χαρακτηριστικά εξόδου από το στρώμα συγκέντρωσης ROI τροφοδοτούνται στη συνέχεια στα διαδοχικά στρώματα FC και στους κλάδους softmax και BB-regression. Ο κλάδος ταξινόμησης softmax παράγει τιμές πιθανότητας για κάθε ROI που ανήκει στις K κατηγορίες και στην μία κατηγορία φόντου. Η έξοδος διακλάδωσης παλινδρόμησης BB χρησιμοποιείται για να κάνει πιο ακριβείς τα κιβώτια οριοθέτησης από τον αλγόριθμο της πρότασης περιοχής.

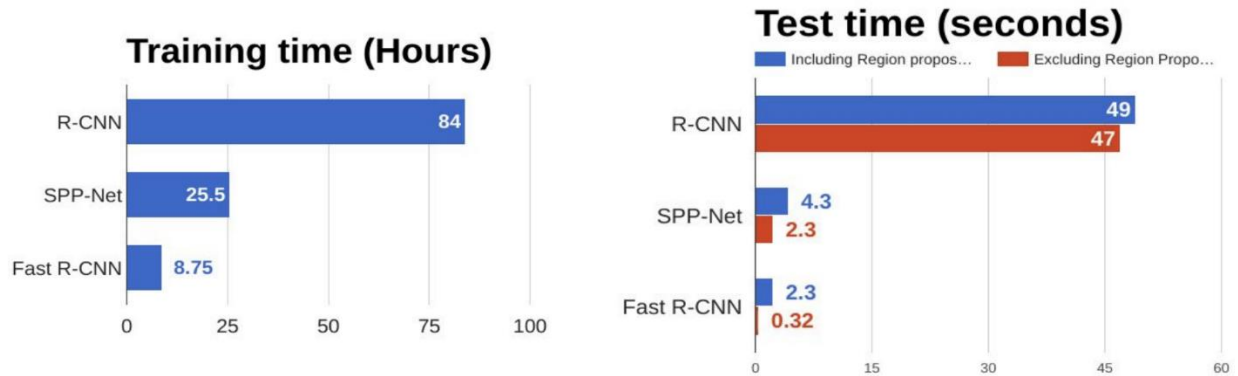
Εκπαίδευση

Η στρατηγική δειγματοληψίας, δηλαδή η δειγματοληψία όλων των ROI σε μια παρτίδα από πολύ λίγες εικόνες, επιτρέπει την εκπαίδευση ολόκληρου του συνόλου, συμπεριλαμβανομένων της δημιουργίας χαρακτηριστικών, της ταξινόμησης και της παλινδρόμησής τους, σε σχέση με τα SPP δίκτυα. Στα SPP δίκτυα τα περισσότερα από τα ROI εισόδου προέρχονται από διαφορετικές εικόνες, οπότε η διαδικασία εκπαίδευσης με σκοπό την ανίχνευση το μόνο που καταφέρνει είναι να τελειοποιεί τις πλήρως συνδεδεμένες στρώσεις μετά την παραγωγή χαρακτηριστικών επειδή δεν είναι εφικτή η ενημέρωση των διαφορετικών βαρών πριν από το στρώμα SPP. Τα πραγματικά χαρακτηριστικά εξακολουθούν να προέρχονται από ένα προ-εκπαιδευμένο δίκτυο που εκπαιδεύτηκε για ταξινόμηση, γεγονός που περιορίζει την ακρίβεια των δικτύων SPP.

Στον αγωγό πολλαπλών κλιμάκων του Fast R-CNN, οι εικόνες εισόδου μετατρέπονται σε ένα τυχαίο μέγεθος δειγματοληψίας στο χρόνο εκπαίδευσης για να εισαχθεί η αμετάβλητη κλίμακα. Κατά την ώρα δοκιμής (test time), κάθε εικόνα τροφοδοτείται στο δίκτυο σε πολλαπλές σταθερές κλίμακες. Για κάθε ROI, τα χαρακτηριστικά συγκεντρώνονται από μία μόνο από αυτές τις κλίμακες, που επιλέγονται έτσι ώστε το υποψήφιο παράθυρο να έχει αριθμό pixel πλησιέστερο στα 224 x 224. Ωστόσο, ο αγωγός ενιαίας κλίμακας μπορεί να λειτουργεί σχεδόν εξίσου καλά με σημαντικά χαμηλότερο κόστος χρόνου υπολογισμού.

Το Fast R-CNN πέτυχε τις καλύτερες επιδόσεις κατά τη στιγμή της δημοσίευσής του σε VOC12, VOC10 και VOC07 (Virtual Object Classes).

Το Fast R-CNN επεξεργάζεται εικόνες 45 φορές γρηγορότερα από το R-CNN κατά το χρόνο δοκιμής και 9 φορές ταχύτερα κατά τον χρόνο εκπαίδευσης. Επίσης εκπαιδεύει 2,7 φορές ταχύτερα και εκτελεί δοκιμαστικές εικόνες 7 φορές ταχύτερα από το SPP-Net. Το σύστημα έχει απόδοση 66% mAP. Για τα μεγαλύτερα δίκτυα η βελτιστοποίηση των στρώσεων μετατροπής (και όχι μόνο των στρωμάτων FC όπως στην περίπτωση των SPPnets) είναι πολύ σημαντική για την επίτευξη σημαντικά καλύτερου mAP. Ωστόσο, για τα μικρότερα δίκτυα, η βελτίωση αυτή μπορεί να μην είναι τόσο μεγάλη. [20]

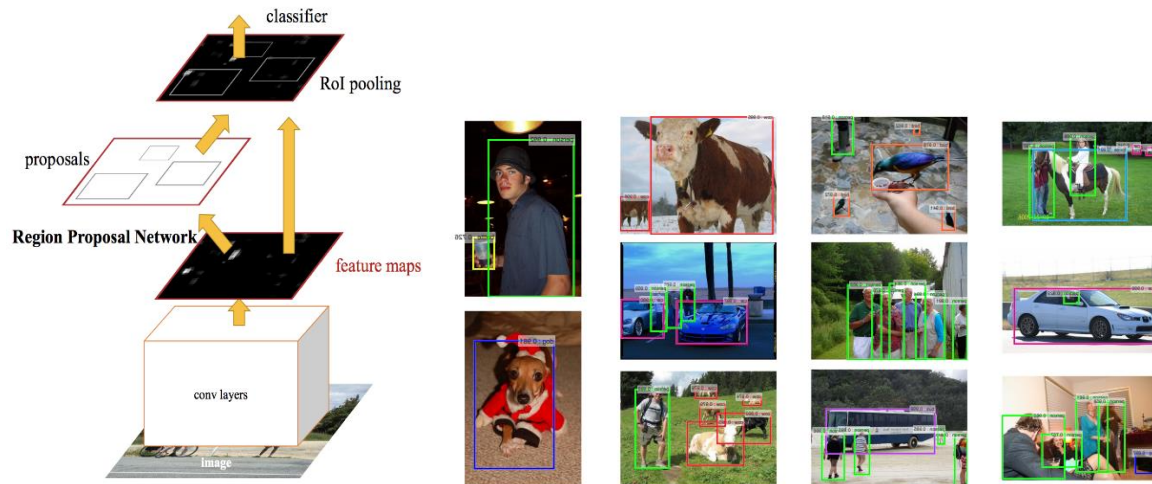


Σχήμα 4.4.2.2.1: Σύγκριση των αλγορίθμων εντοπισμού αντικειμένων

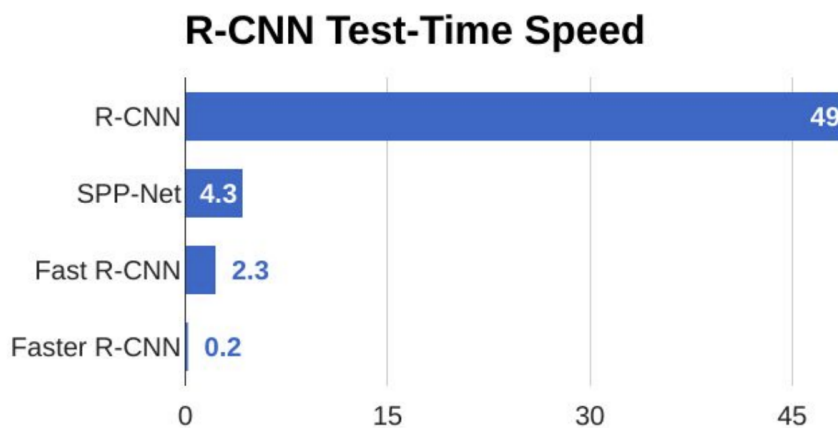
4.4.2.3 Faster R-CNN

Η πιο ευρέως διαδεδομένη έκδοση της οικογένειας R-CNN, είναι το Faster R-CNN, το οποίο δημοσιεύθηκε για πρώτη φορά το 2015. Η εξέλιξη μεταξύ των διαφορετικών CNN εκδόσεων αφορούσαν συνήθως την υπολογιστική αποτελεσματικότητα, καθώς αλλάζει η ενσωμάτωση των διαφορετικών σταδίων κατάρτισης, την μείωση του χρόνου δοκιμής και την βελτίωση της απόδοσης, η οποία μετράτε σε mAP. Αυτά τα δίκτυα αποτελούνται συνήθως καταρχάς από έναν αλγόριθμο πρότασης περιοχής για τη δημιουργία πλαισίων οριοθέτησης ή τοποθεσιών πιθανών αντικειμένων στην εικόνα. Επίσης, απαραίτητη είναι η ύπαρξη ενός σταδίου δημιουργίας χαρακτηριστικών για την εξαγωγή των χαρακτηριστικών αυτών των αντικειμένων, συνήθως χρησιμοποιώντας CNN. Ακόμα, στα δίκτυα αυτά συναντάται το στρώμα ταξινόμησης προκειμένου να γίνει η πρόβλεψη αναφορικά με την κλάση που θα μπορούσε να ανήκει το αντικείμενο. Τέλος, αναπόσπαστο κομμάτι των δικτύων είναι μια στρώση παλινδρόμησης, έτσι ώστε οι συντεταγμένες του πλαισίου οριοθέτησης αντικειμένου να προσδιορίζονται με περισσότερη ακρίβεια.

Το μόνο αυτόνομο τμήμα του δικτύου που απέμεινε στο Fast R-CNN ήταν ο αλγόριθμος της πρότασης περιοχής. Τόσο ο αλγόριθμος R-CNN όσο και ο Fast R-CNN χρησιμοποιούν αλγόριθμους βασισμένους σε CPU (Κεντρική Μονάδα Επεξεργασίας-Central Processing Unit) με βάση την περιοχή, π.χ. τον αλγόριθμο επιλεκτικής αναζήτησης που διαρκεί περίπου 2 δευτερόλεπτα ανά εικόνα και εκτελείται με υπολογισμό της CPU. Η μελέτη σχετικά με τον αλγόριθμο Faster R-CNN επιδιορθώνει τον περιορισμό χρησιμοποιώντας ένα άλλο συνελικτικό δίκτυο, το RPN (Δίκτυο Πρότασης Περιοχής-Region Proposal Network) για να δημιουργήσει τις προτάσεις περιοχής. Αυτό, όχι μόνο μειώνει της διάρκειας της πρότασης περιοχής κατά μέσο όρο από 2 δευτερόλεπτα σε 10 μιλιδευτερόλεπτα (ms) ανά εικόνα, αλλά και επιτρέπει στο στάδιο της πρότασης περιοχής να μοιράσει στρώματα με τα στάδια ανίχνευσης που αναλύονται στην συνέχεια, προκαλώντας μια συνολική βελτίωση στην αναπαράσταση χαρακτηριστικών. [21]



Σχήμα 4.4.2.3.1: Αρχιτεκτονική μοντέλου Faster R-CNN. Αποτέλεσμα της ανίχνευσης αντικειμένων με Faster R-CNN



Σχήμα 4.4.2.3.2: Σύγκριση των αλγορίθμων εντοπισμού αντικειμένων

4.4.2.4 Δίκτυο πρότασης περιοχής (Region Proposal Network-RPN)

Αρχιτεκτονική

Το δίκτυο πρότασης περιοχής ξεκινάει με την εικόνα-είσοδο που τροφοδοτείται στο κεντρικό νευρωνικό δίκτυο περιστροφής. Τα χαρακτηριστικά εξόδου του δικτύου κορμού είναι συνήθως πολύ μικρότερα από την εικόνα εισόδου ανάλογα με το βήμα του δικτύου κορμού. Για κάθε σημείο του χάρτη εξόδων, το δίκτυο πρέπει να καταλάβει αν υπάρχει ένα αντικείμενο στην εικόνα εισόδου στην αντίστοιχη θέση και να υπολογίσει το μέγεθός του. Αυτό γίνεται με την τοποθέτηση ενός συνόλου πινεζών στην εικόνα εισόδου για κάθε θέση στον χάρτη χαρακτηριστικών εξόδου από το δίκτυο κορμού. Αυτά τα άγκιστρα δείχνουν πιθανά αντικείμενα σε διάφορα μεγέθη και αναλογίες σε αυτή τη θέση.

Καθώς το δίκτυο μετακινείται μεταξύ των εικονοστοιχείων στον χάρτη χαρακτηριστικών εξόδου, πρέπει να ελέγξει εάν αυτά τα αντίστοιχα άγκιστρα που καλύπτουν την εικόνα εισόδου περιέχουν πραγματικά αντικείμενα και να τελειοποιήσουν τις συντεταγμένες αυτών των αγκυρών για να δώσουν πλαίσια οριοθέτησης ως προτάσεις αντικειμένων ή περιοχές ενδιαφέροντος.

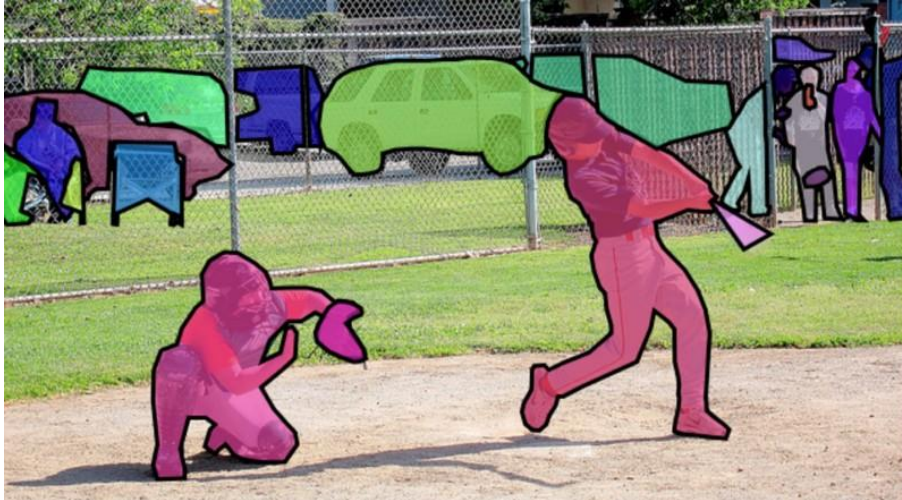
4.4.2.4.1 Mask R-CNN

Παρόλο που το Mask R-CNN αποτελεί μία σχετικά νέα προσθήκη στα μοντέλα βαθιάς μάθησης, αποτελεί ίσως την ιδανική λύση όταν επιθυμούμε να φτάσουμε στο επιθυμητό αποτέλεσμα, διαθέτοντας τον μικρότερο δυνατό πλήθος δεδομένων προς εκπαίδευση. Η αρχιτεκτονική αυτή είναι σχεδιασμένη με τέτοιο τρόπο ώστε να καθίσταται δυνατή η ανίχνευση αντικειμένων σε ολόκληρη την εικόνα με υπολογιστικά αποτελεσματικό τρόπο, χωρίς να προϋποτίθεται η χρήση κυλιόμενου παραθύρου. Έτσι, ο αλγόριθμος τρέχει με σχετικά γρήγορες ταχύτητες. Σε περίπτωση που υπάρχει διαθέσιμη μία κάρτα γραφικών νέας τεχνολογίας, ο αλγόριθμος είναι σε θέση να εντοπίζει αντικείμενα από βίντεο υψηλής ευκρίνειας με ρυθμό αρκετών καρέ ανά δευτερόλεπτο. Επιπλέον το Mask R-CNN παρέχει στους χρήστες επιπρόσθετες πληροφορίες αναφορικά με το κάθε αντικείμενο που εντοπίζει. Σε αντίθεση με τους περισσότερους αλγόριθμους ανίχνευσης αντικειμένων, οι οποίοι επιστρέφουν ως έξοδο τα αντικείμενα που έχουν εντοπιστεί περιφραγμένα από ένα πλαίσιο, το Mask R-CNN εκτός από την τοποθεσία των αντικειμένων παρέχει το ακριβές περίγραμμά τους (mask).



Σχήμα 4.4.2.4.1.1: Αποτέλεσμα της ανίχνευσης αντικειμένων με χρήση του αλγορίθμου Mask R-CNN

Στο διαδίκτυο υπάρχει διαθέσιμο ένα δημοφιλές σύνολο δεδομένων (dataset), το COCO (Κοινά Αντικείμενα Στο Πλαίσιο-Common Objects In Context), το οποίο διαθέτει εικόνες με περιεγραμμένα αντικείμενα. Σε αυτό το σύνολο δεδομένων, συναντώνται περισσότερες από 12.000 εικόνες με ήδη περιεγραμμένα αυτοκίνητα.



Σχήμα 4.4.2.4.1.2: Παράδειγμα από το σύνολο δεδομένων COCO

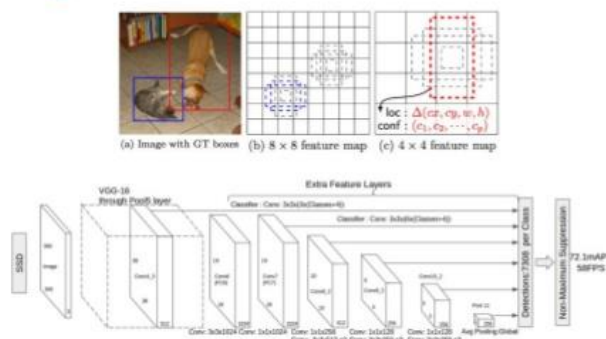
4.4.2.5 Shot MultiBox Detector (SSD)

Το SSD (Single Shot MultiBox Detector) αποτελεί έναν δημοφιλή αλγόριθμο βαθιάς μάθησης για ανίχνευση αντικειμένων. Είναι σε γενικές γραμμές ταχύτερο από το Faster R-CNN. Είναι το πρώτο μοντέλο που δεν επαναπροσδιορίζει τα pixels ή τα χαρακτηριστικά για να τοποθετήσει τα πλαίσια, ενώ παράλληλα πετυχαίνει πάρα πολύ μεγάλα ποσοστά ακρίβειας. Ένα τυπικό δίκτυο CNN συρρικνώνει σταδιακά το μέγεθος του χάρτη χαρακτηριστικών και αυξάνει το βάθος καθώς προχωρά στα βαθύτερα στρώματα. Τα βαθιά στρώματα καλύπτουν μεγαλύτερα δεκτικά πεδία και κατασκευάζουν πιο αφηρημένη αναπαράσταση, ενώ τα πιο πάνω στρώματα καλύπτουν μικρότερες δεκτικές περιοχές. Χρησιμοποιώντας αυτές τις ιδιότητες των στρωμάτων, μπορούμε να χρησιμοποιήσουμε τα ρηχά στρώματα για να προβλέψουμε μικρά αντικείμενα και τα βαθύτερα στρώματα για να προβλέψουμε μεγάλα αντικείμενα, καθώς τα μικρά αντικείμενα δεν χρειάζονται μεγαλύτερα δεκτικά πεδία, μιας και αυτά ενδέχεται να προκαλέσουν σύγχυση όταν πρόκειται για μικρά αντικείμενα.

Για την εκπαίδευσή του, ο αλγόριθμος δημιουργεί για κάθε σημείο χαρακτηριστικών μια σειρά από προτεραιότητες (priors), τις οποίες στη συνέχεια χρησιμοποιεί για να ταιριάζει με τα κουτιά αληθείας, προκειμένου να επιτευχθεί ο προσδιορισμός των ετικετών και των πλαισίων οριοθέτησης. Σε αυτή την φάση της αντιστοίχισης, ενισχύουμε τους θετικούς στόχους, καθώς τα πλαίσια έχουν το αντικείμενο που τους έχει ανατεθεί, προσαρμόζοντας τα κουτιά αληθείας σε πολλαπλές προτεραιότητες. Ωστόσο, εξακολουθούν να υπάρχουν πολύ περισσότερες ασύγκριτες προτεραιότητες. Αυτό συμβαίνει διότι ο τεράστιος αριθμός προτεραιοτήτων που έχουν επισημανθεί ως φόντο καθιστά το σύνολο δεδομένων μη ισορροπημένο. Για την άρση της ανισορροπίας στο σύνολο δεδομένων, συχνά χρησιμοποιείται το Hard Negative Mining. Η ιδέα είναι να μετρούνται μόνο τα αρχικά priors με την υψηλότερη εμπιστοσύνη στον υπολογισμό της συνάρτησης συνολικών ζημιών, ενώ τα υπόλοιπα αγνοούνται. Η αναλογία μεταξύ των priors του φόντου και αυτών που αντιστοιχίζονται μειώνεται αισθητά. Αφού ο αλγόριθμος έχει παραγάγει με επιτυχία τα δεδομένα εκπαίδευσης, υπάρχει μόνο ένα βήμα που λείπει ακόμα, η αύξηση των

δεδομένων, η οποία βοηθά τον αλγόριθμο να καταλάβει πως τα δεδομένα είναι αμετάβλητα. Στην πραγματικότητα, σε αντίθεση με το Faster R-CNN, η αύξηση δεδομένων διαδραματίζει ουσιαστικό ρόλο στον SSD. [22]

SSD: Single Shot MultiBox Detector



Liu et al. [SSD: Single Shot MultiBox Detector](#), arXiv 2015

29

Σχήμα 4.4.2.5.1: Αρχιτεκτονική μοντέλου SSD

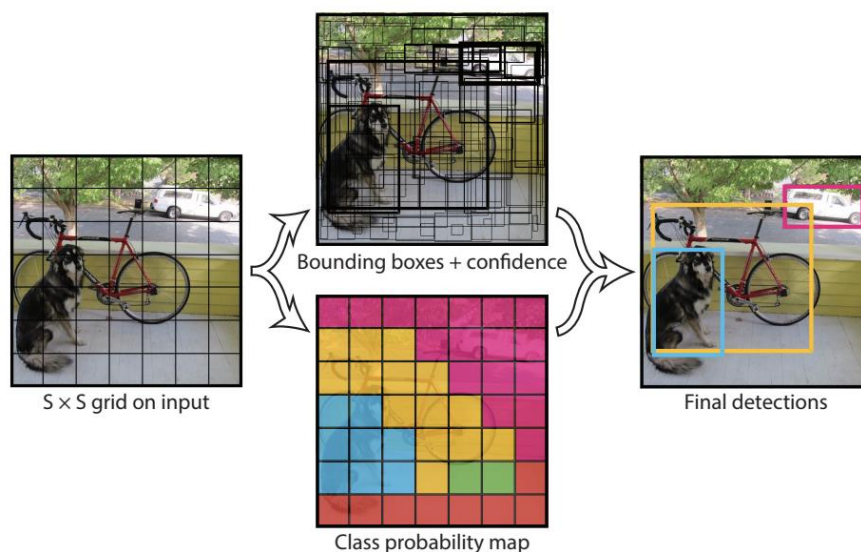
Ο συγκεκριμένος αλγόριθμος, όπως όλοι άλλωστε, παρουσιάζει κάποια μειονεκτήματα. Τα ρηχά στρώματα σε ένα νευρικό δίκτυο ενδέχεται να μην παράγουν αρκετά χαρακτηριστικά υψηλού επιπέδου για να κάνουν πρόβλεψη για μικρά αντικείμενα. Επομένως, το SSD έχει χαμηλότερες επιδόσεις στα μικρότερα αντικείμενα σε σύγκριση με τα μεγαλύτερα αντικείμενα. Η ανάγκη σύνθετης αύξησης των δεδομένων υποδεικνύει επίσης ότι χρειάζεται μεγάλος αριθμός δεδομένων για εκπαίδευση, γεγονός που δυσχεραίνει την διαδικασία. Για παράδειγμα, το SSD βελτιώνεται εάν το μοντέλο έχει πρώτα εκπαιδευτεί σε σύνολο δεδομένων COCO.

4.4.2.6 YOLO

Σε αντίθεση με άλλους αλγόριθμους ανίχνευσης αντικειμένων, οι οποίοι επαναχρησιμοποιούν τους ταξινομητές προκειμένου να εντοπίσουν τα αντικείμενα σε μία εικόνα ή ένα βίντεο, ο αλγόριθμος YOLO, αντιμετωπίζει την ανίχνευση σαν μια διαδικασία παλινδρόμησης σε χωρικά διαχωρισμένα πλαίσια οριοθέτησης και σχετικές πιθανότητες κλάσης. Ένα απλό νευρωνικό δίκτυο έχει την δυνατότητα να προβλέπει τα κουτιά που οριοθετούν τα αντικείμενα (bounding boxes) και τις πιθανότητες κλάσης απευθείας από πλήρεις εικόνες σε μία αξιολόγηση.

Δεδομένου ότι ολόκληρος ο αγωγός ανίχνευσης (detection pipeline) αποτελεί ένα ενιαίο δίκτυο, οι επιδόσεις ανίχνευσης μπορούν να βελτιστοποιηθούν άμεσα από άκρο σε άκρο. Σε σύγκριση με τα πλέον σύγχρονα συστήματα ανίχνευσης, ο YOLO είναι περισσότερο επιρρεπής σε σφάλματα εντοπισμού, όμως έχει λιγότερες πιθανότητες να προβλέψει ψευδώς θετικά το παρασκήνιο (background). Ο YOLO μαθαίνει πολύ γενικές αναπαραστάσεις αντικειμένων, γεγονός που τον κάνει να υπερτερεί έναντι άλλων μεθόδων ανίχνευσης, συμπεριλαμβανομένων των DPM και R-CNN, όταν παρατηρείται γενίκευση από φυσικές εικόνες σε άλλους τομείς όπως έργα τέχνης.

Οι άνθρωποι βλέπουν μια εικόνα και αμέσως γνωρίζουν ποια αντικείμενα βρίσκονται στην εικόνα, πού βρίσκονται και πώς αλληλοεπιδρούν. Το ανθρώπινο οπτικό σύστημα είναι γρήγορο και ακριβές, επιτρέποντάς μας να εκτελούμε σύνθετα καθήκοντα όπως οδήγηση με μικρή συνειδητή σκέψη. Οι αλγόριθμοι για ανίχνευση αντικειμένων, οι οποίοι χαρακτηρίζονται από ακρίβεια σε υψηλές ταχύτητες, θα μπορούσαν να βοηθήσουν τους υπολογιστές να οδηγούν αυτοκίνητα χωρίς εξειδικευμένους αισθητήρες, κάποιες βοηθητικές συσκευές να μεταδίδουν σε πραγματικό χρόνο χρήσιμες πληροφορίες σε χρήστες και τα ανταποκρινόμενα ρομποτικά συστήματα να ξεκλειδώσουν τις δυνατότητές τους. Τα υπάρχοντα συστήματα ανίχνευσης επανατοποθετούν τους ταξινομητές για να εκτελέσουν ανίχνευση. Για την ανίχνευση ενός αντικειμένου, τα συστήματα αυτά λαμβάνουν έναν ταξινομητή για αυτό το αντικείμενο και το αξιολογούν σε διάφορες τοποθεσίες και κλίμακες σε μια δοκιμαστική εικόνα. Συστήματα όπως τα μοντέλα παραμορφώσιμων μερών (Deformable Parts Models-DPM) χρησιμοποιούν μια προσέγγιση συρόμενων παραθύρων, όπου ο ταξινομητής εκτελείται σε ομοιόμορφα τοποθετημένες θέσεις σε ολόκληρη την εικόνα. [23]



Σχήμα 4.4.2.6.1: Τα στάδια χρήσης του YOLO για ανίχνευση αντικειμένων

4.4.2.6.1 Περιορισμοί του YOLO

Ο YOLO επιβάλλει ισχυρούς χωρικούς περιορισμούς στις προβλέψεις οριοθέτησης των κουτιών, δεδομένου ότι κάθε κυψέλη πλέγματος προβλέπει μόνο δύο πλαίσια και μπορεί να έχει μόνο μία τάξη. Αυτός ο χωροταξικός περιορισμός περιορίζει τον αριθμό των κοντινών αντικειμένων που μπορεί να προβλέψει το μοντέλο. Ο YOLO αντιμετωπίζει δυσκολίες σχετικά με μικρά αντικείμενα που εμφανίζονται κατά ομάδες, όπως για παράδειγμα σμήνη πουλιών. Δεδομένου ότι το μοντέλο μαθαίνει να προβλέπει τα κουτιά οριοθέτησης από δεδομένα, δυσκολεύεται να γενικεύσει σε αντικείμενα που είναι νέα ή ασυνήθιστα αναφορικά με τις αναλογίες ή τις διαστάσεις. Επίσης, ο αλγόριθμος χρησιμοποιεί σχετικά τραχιά χαρακτηριστικά για την πρόβλεψη των πλαισίων οριοθέτησης μιας και η αρχιτεκτονική του έχει πολλαπλά στρώματα δειγματοληψίας από την εικόνα εισόδου. Τέλος, ενώ εκπαιδεύουμε το μοντέλο σε μια λειτουργία απώλειας που προσεγγίζει την απόδοση ανίχνευσης, η λειτουργία απώλειας αντιμετωπίζει τα

λάθη με τον ίδιο τρόπο, είτε πρόκειται για μικρά, είτε για μεγάλα κιβώτια οριοθέτησης. Ένα μικρό σφάλμα σε ένα μεγάλο κιβώτιο είναι γενικά καλοήθες όμως ένα μικρό σφάλμα σε ένα μικρό κουτί, έχει πολύ μεγαλύτερη επιρροή στην αξιοπιστία του τελικού αποτελέσματος.

4.4.2.6.2 Πλεονεκτήματα του YOLO

Σε σύγκριση με άλλα δίκτυα ταξινόμησης προτάσεων περιοχής, όπως το Fast R-CNN, τα οποία εκτελούν ανίχνευση σε διάφορες προτάσεις περιοχής και έτσι συχνά καταλήγουν να προβλέπουν διάφορες περιοχές σε μια εικόνα παραπάνω από μία φορές, η αρχιτεκτονική YOLO περνά την εικόνα μια φορά μέσω του FCNN (πλήρως συνελκτικό νευρωνικό δίκτυο) και η έξοδος που προκύπτει, είναι η επιθυμητή πρόβλεψη. Το ενιαίο συνελκτικό δίκτυο προβλέπει συγχρόνως πολλαπλά πλαίσια οριοθέτησης και πιθανότητες τάξης για αυτά τα πλαίσια. Ο YOLO εκπαιδεύεται σε πλήρεις εικόνες και βελτιστοποιεί άμεσα την απόδοση ανίχνευσης. Αυτό το ενοποιημένο μοντέλο έχει πολλά πλεονεκτήματα σε σχέση με τις παραδοσιακές μεθόδους ανίχνευσης αντικειμένων, τα οποία αναλύονται εν συνεχεία. Παρακάτω φαίνεται ένα παράδειγμα εξόδου με χρήση του αλγορίθμου YOLO.



Σχήμα 4.4.2.6.2.1: Αποτέλεσμα χρήσης του YOLO για ανίχνευση αντικειμένων

Πρώτον, ο YOLO είναι εξαιρετικά γρήγορος αλγόριθμος ανίχνευσης αντικειμένων. Από τη στιγμή που η ανίχνευση αντιμετωπίζεται ως πρόβλημα παλινδρόμησης, η χρήση ενός πολύπλοκου αγωγού (pipeline), δεν καθίσταται πια απαραίτητη. Αρκεί απλά το νευρωνικό δίκτυο να εκτελεστεί σε μια νέα εικόνα στο χρόνο δοκιμής, προκειμένου να μπορούν να προβλεφθούν οι ανιχνεύσεις. Χάρη στην γρήγορη ταχύτητα πρόβλεψης του αλγορίθμου, δύναται η επεξεργασία βίντεο συνεχούς ροής σε πραγματικό χρόνο σε λιγότερο από μερικά χιλιοστά του δευτερολέπτου.

Επιπλέον σε αντίθεση με τις τεχνικές που βασίζονται σε προτάσεις συρόμενων παραθύρων και προτάσεων περιοχής, ο YOLO βλέπει ολόκληρη την εικόνα κατά τη διάρκεια της εκπαίδευσης και του χρόνου δοκιμών. Έτσι κωδικοποιεί έμμεσα πληροφορίες σχετικά με τις τάξεις καθώς και την εμφάνισή τους. Το Fast R-CNN, η οποία αποτελεί μια από τις πλέον αξιόπιστες μεθόδους ανίχνευσης, συγχέει τις ιδιομορφίες του φόντου σε μια εικόνα με τα αντικείμενα επειδή αδυνατεί να διακρίνει το ευρύτερο περιεχόμενο. Ο YOLO κάνει λιγότερο από τον μισό αριθμό σφαλμάτων φόντου σε σύγκριση με το Fast R-CNN.

Ένα ακόμα πλεονέκτημα του YOLO σε σχέση με άλλους αλγορίθμους ανίχνευσης αντικειμένων είναι η δυνατότητά του να μαθαίνει γενικευμένες αναπαραστάσεις αντικειμένων. Για παράδειγμα, όπως αναφέρθηκε και παραπάνω όταν εκπαιδεύεται σε φυσικές εικόνες και δοκιμάζεται σε έργα τέχνης, ο YOLO ξεπερνάει κατά πολύ τις επιδόσεις των κορυφαίων μεθόδων ανίχνευσης όπως το DPM και το R-CNN. Δεδομένου ότι ο YOLO είναι εξαιρετικά γενικεύσιμος, είναι λιγότερο πιθανό να καταρρεύσει όταν εφαρμόζεται σε νέους τομείς ή απρόσμενες εισόδους.

Το δίκτυο χρησιμοποιεί χαρακτηριστικά από ολόκληρη την εικόνα για να προβλέψει κάθε πλαίσιο οριοθέτησης. Προβλέπει επίσης ταυτόχρονα όλα τα πλαίσια οριοθέτησης από όλες τις κατηγορίες για μια εικόνα. Ο σχεδιασμός της αρχιτεκτονικής του YOLO επιτρέπει εκπαίδευση από άκρο σε άκρο και ταχύτητες σε πραγματικό χρόνο διατηρώντας παράλληλα υψηλή ακρίβεια. Το σύστημα χωρίζει την εικόνα εισόδου σε ένα δίκτυο $S \times S$. Αν το κέντρο ενός αντικειμένου πέσει σε ένα κελί δικτύου, αυτό το κελί δικτύου είναι υπεύθυνο για την ανίχνευση αυτού του αντικειμένου.

Κάθε κυψελίδα πλέγματος προβλέπει καμπύλες οριοθέτησης και βαθμούς εμπιστοσύνης για αυτά τα πλαίσια. Αυτές οι βαθμολογίες εμπιστοσύνης αντανακλούν πόσο βέβαιο είναι το μοντέλο ότι το κουτί περιέχει ένα αντικείμενο και επίσης πόσο ακριβές πιστεύει πως είναι αυτό που προβλέπει. Εάν κανένα αντικείμενο δεν υπάρχει σε αυτό το κελί, οι βαθμοί εμπιστοσύνης θα πρέπει να είναι μηδέν. Διαφορετικά, θέλουμε η βαθμολογία εμπιστοσύνης να ισούται με τη διασταύρωση κατά την ένωση (intersection over union-IOU) μεταξύ του προβλεπόμενου κιβωτίου και της αλήθειας. Κάθε πλαίσιο οριοθέτησης αποτελείται από 5 προβλέψεις: x , y , w , h , και εμπιστοσύνη. Οι συντεταγμένες (x, y) αντιπροσωπεύουν το κέντρο του κουτιού σε σχέση με τα όρια του κελιού του δικτύου. Το πλάτος και το ύψος προβλέπονται σε σχέση με ολόκληρη την εικόνα. Τέλος, η πρόβλεψη εμπιστοσύνης αντιπροσωπεύει το IOU μεταξύ του προβλεπόμενου κουτιού και οποιουδήποτε βασικού κουτιού αλήθειας. [24]

Κεφάλαιο 5

Η εφαρμογή *DJI GoTo*

5.1 Γενικά για την εφαρμογή

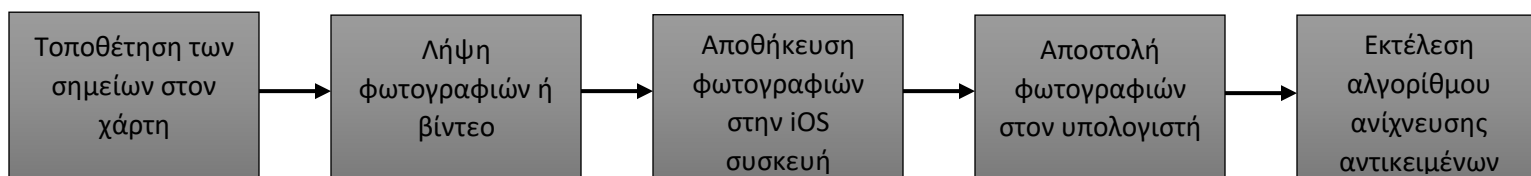
Ο εντοπισμός αντικειμένων με χρήση μη επανδρωμένων αεροσκαφών έχει πολλές διαφορετικές προσεγγίσεις και μπορεί να επεκταθεί με ποικίλους τρόπους. Η συγκεκριμένη εργασία εστιάζει κυρίως στον εντοπισμό οχημάτων. Η εφαρμογή που δημιουργήθηκε μπορεί να χρησιμοποιηθεί από τους χρήστες για πολλές διαφορετικές λειτουργίες. Υπάρχει αρκετή ελευθερία όσον αφορά τον χειρισμό της. Ενδεικτικά, ο χρήστης έχει την δυνατότητα να εισάγει ο ίδιος τιμές όπως η ταχύτητα και το ύψος του μη επανδρωμένου αεροσκάφους και να τις μεταβάλλει κατά βούληση. Η εφαρμογή δεν απαιτεί από τους χρήστες την γνώση προγραμματισμού ή άλλων εξειδικευμένων γνώσεων, μιας και το γραφικό περιβάλλον τους καθοδηγεί με σαφήνεια κάνοντας χρήση βοηθητικών μηνυμάτων και pop up παραθύρων.

Η εφαρμογή, η οποία στο πλαίσιο της Διπλωματικής εργασίας ονομάστηκε *DJI GoTo* μπορεί να χρησιμοποιηθεί είτε για εμπορικούς, είτε για ερευνητικούς σκοπούς, αλλά ακόμα και για προσωπική χρήση. Μερικοί από τους τρόπους αξιοποίησης της εφαρμογής αναφέρονται στην συνέχεια. Ωστόσο, στο κομμάτι της εισαγωγής της συγκεκριμένης Διπλωματικής Εργασίας έχει αφιερωθεί ένα κεφάλαιο σχετικά με το θέμα *Drones for Good*, όπου παρουσιάζονται πιθανές χρήσεις των μη επανδρωμένων αεροσκαφών και της ανίχνευσης αντικειμένων προς όφελος του κοινωνικού συνόλου, όπως για παράδειγμα η παράδοση τροφίμων ή φαρμάκων σε απομακρυσμένες περιοχές, όταν υπάρχει ανάγκη. Οι περιπτώσεις αυτές προφανώς απαιτούν ταχύτητα και ακρίβεια, διαφορετικά τα αποτελέσματα ενδέχεται να είναι καταστροφικά.

5.2 Περιγραφή της λειτουργίας της εφαρμογής

Στο πλαίσιο της συγκεκριμένης Διπλωματικής Εργασίας, δημιουργήθηκε μία εφαρμογή η οποία δίνει την δυνατότητα σε χρήστες έξυπνων συσκευών της εταιρίας Apple (iPhone ή iPad) να χειρίζονται ένα μη επανδρωμένο αεροσκάφος DJI. Συγκεκριμένα, η εφαρμογή γράφτηκε στην γλώσσα προγραμματισμού Swift προκειμένου να είναι συμβατή με λογισμικό iOS, το οποίο αναπτύχθηκε και διανέμεται αποκλειστικά από την Apple. Η έκδοση της Swift στην οποία έγινε η ανάπτυξη της εφαρμογής είναι η Swift 5.0, η οποία είναι η νεότερη διαθέσιμη. Ο κώδικας γράφτηκε στο ολοκληρωμένο περιβάλλον ανάπτυξης Xcode, το οποίο υποστηρίζεται από συσκευές με λογισμικό iOS .

Το γραφικό περιβάλλον της εφαρμογής σχεδιάστηκε για το iPhone 5S, το οποίο και είχαμε διαθέσιμο στο εργαστήριο. Η διαφοροποίηση στην σχεδίαση αφορά τις διαφορετικές προδιαγραφές σχεδιασμού της κάθε συσκευής: σχήμα και μέγεθος οθόνης, μέγεθος ασφαλούς περιοχής (το εκμεταλλεύσιμο μέρος της οθόνης) και προσανατολισμός οθόνης αν πρόκειται για iPad. Η εφαρμογή μπορεί εύκολα να τροποποιηθεί προκειμένου να ταιριάζει με τις προδιαγραφές και άλλων συσκευών της Apple. Το μη επανδρωμένο αεροσκάφος που χρησιμοποιήθηκε κατασκευάστηκε από την εταιρία DJI, όπως προαναφέρθηκε, και το μοντέλο του είναι το Mavic Pro. Τα χαρακτηριστικά του εναέριου οχήματος καθώς και φωτογραφίες του, υπάρχουν σε επόμενο κεφάλαιο της εργασίας. Ακολουθώντας παρουσιάζονται τα διαδοχικά στάδια της λειτουργίας της εφαρμογής σε μορφή διαγράμματος ροής.

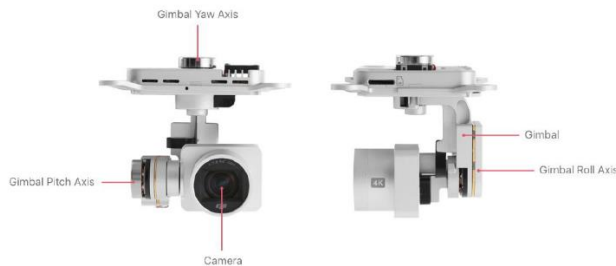


Σχήμα 5.2.1: Διάγραμμα ροής για την περιγραφή της λειτουργίας της εφαρμογής

Όπως φαίνεται και από το παραπάνω σχήμα, το πρώτο στάδιο της εφαρμογής είναι ο ορισμός των σημείων στον χάρτη, τα οποία ο χρήστης επιθυμεί να επισκεφτεί το μη επανδρωμένο εναέριο όχημα. Αφού έχει ολοκληρωθεί η διαδικασία της απογείωσης και καθώς το μη επανδρωμένο διαγράφει την πορεία που του έχει υποβάλλει ο χρήστης, είναι στην δική του ευχέρεια το πότε επιθυμεί να τραβήξει φωτογραφίες ή βίντεο μιας και του παρέχεται οπτική επαφή με ό,τι καταγράφεται από την κάμερα του αεροσκάφους κάθε στιγμή. Οι φωτογραφίες και τα βίντεο αποθηκεύονται στην συσκευή που εκτελείται η εφαρμογή. Στην συνέχεια αποστέλλονται στον υπολογιστή, ο οποίος είναι συμβατός με λογισμικό iOS. Τα στάδια αυτά αναλύονται περαιτέρω στην πορεία.

Ο βέλτιστος εντοπισμός αντικειμένων με χρήση του μη επανδρωμένου αεροσκάφους προϋποθέτει την σωστή χρήση του gimbal της κάμερας. Συγκεκριμένα, πολύ σημαντικό ρόλο στην λήψη αξιόπιστων φωτογραφιών και βίντεο παίζει η κλίση του gimbal. Σε περιπτώσεις όπου ο χρήστης επιλέγει η πτήση του αεροσκάφους να κρατηθεί σε χαμηλό υψόμετρο, τότε η αντίστοιχη κλίση πρέπει να είναι εξίσου μικρή. Σε αντίθετη περίπτωση ο εντοπισμός συνήθως δεν θα είναι εφικτός, αφού στις φωτογραφίες οι οντότητες των αντικειμένων δεν θα φαίνονται ολοκληρωμένες. Αυτό συμβαίνει διότι, ο εντοπισμός αυτοκινήτων καθίσταται αδύνατος σε περίπτωση που η λήψη προς ανάλυση περιέχει μόνο κομμάτι της οροφής τους. Σε αυτή την υλοποίηση οι 2 από τους 3 άξονες περιστροφής του gimbal διατηρούνται σταθεροί και μεταβάλλεται μόνο ο gimbal pitch axis.

Οι προτεινόμενες τιμές από την DJI κυμαίνονται μεταξύ -30 και -60 μοιρών, ανάλογα με το υψόμετρο. Η μαθηματική σχέση που περιγράφει την απόσταση σε ευθεία γραμμή μεταξύ κάμερας και εδάφους είναι ο λόγος του ύψους της πτήσης προς το ημίτονο της κλίσης του gimbal. Από την σχέση αυτή μπορεί να εξαχθεί το συμπέρασμα πως όσο η απόσταση πλησιάζει την τιμή του ύψους τόσο δυσκολότερος είναι και ο εντοπισμός αφού η εικόνες θα περιλαμβάνουν μόνο την πάνω όψη των αντικειμένων.



Σχήμα 5.2.2: Γενική Λειτουργία του Gimbal



Σχήμα 5.2.3: Gimbal του Μοντέλου Mavic Pro

5.2.1 Γιατί επιλέχθηκε η χρήση Drone

Η πρόοδος της τεχνολογίας UAV(Unnamed Aerial Vehicle) έχει φτάσει σε ένα σημείο, όπου προσφέρει την δυνατότητα για εικόνες τηλεπισκόπησης εξαιρετικά υψηλής ανάλυσης που περιλαμβάνουν άφθονες χωρικές, αλλά και όχι μόνο, πληροφορίες. Η ταχύρρυθμη εξέλιξη της τεχνολογίας που περιβάλλει τα μη επανδρωμένα εναέρια αεροσκάφη, αποτέλεσε το κατάλληλο υπόβαθρο για την διεξαγωγή πολυάριθμων μελετών. Ως αποτέλεσμα, αναπτύχθηκαν πολλές νέες εφαρμογές για την ανάλυση εικόνας UAV, συμπεριλαμβανομένης της παρακολούθησης της βλάστησης, της ανάλυσης αστικών περιοχών, της διαχείρισης καταστροφών, της παρακολούθησης των αγωγών πετρελαίου και φυσικού αερίου, της ανίχνευσης και χαρτογράφησης αρχαιολογικών χώρων και της ανίχνευσης αντικειμένων, που φυσικά αποτελεί και το αντικείμενο μελέτης της παρούσας Διπλωματικής Εργασίας.

Ο εντοπισμός και η μέτρηση αυτοκινήτων αποτελεί έναν τομέα ενεργητικής έρευνας στον κλάδο ανίχνευσης αντικειμένων κιάλας από το 2001. Για τον εντοπισμό των αυτοκινήτων, χρειάζεται ο υπολογισμός ενός μέτρου ομοιότητας χρησιμοποιώντας έναν κατάλογο αυτοκινήτων ως αναφορά. Τα UAV μπορούν να αποκτήσουν εικόνες με εξαιρετικά υψηλό επίπεδο χωρικών λεπτομερειών σε σύγκριση με τα τυπικά συστήματα απομακρυσμένης απόκτησης όπως οι δορυφόροι ή τα αερομεταφερόμενα. Τα μη επανδρωμένα αεροσκάφη είναι ιδανικά για την ανίχνευση συγκεκριμένης κατηγορίας αντικειμένων, γεγονός το οποίο δημιουργεί πληθώρα δυνατοτήτων στην ανάπτυξη διαφόρων εφαρμογών ενδιαφέροντος αναγνώρισης.

Ένα επιπλέον πολύ σημαντικό πλεονέκτημα των μη επανδρωμένων αεροσκαφών είναι το χαμηλό τους κόστος σε συνδυασμό με την υψηλή ευελιξία που προσφέρουν, γεγονός το οποίο καθιστά ιδιαίτερως δημοφιλείς τις πλατφόρμες που βασίζονται σε UAV και σε εφαρμογές ανάλυσης εικόνας UAV.

5.2.2 Mavic Pro

Το μοντέλο του μη επανδρωμένου αεροσκάφους πάνω το οποίο χρησιμοποιήσαμε στο εργαστήριο, προκειμένου να ελέγξουμε την λειτουργία της εφαρμογής, είναι το Mavic Pro Platinum της εταιρίας DJI. Το συγκεκριμένο μοντέλο μπορεί να είναι μικρό, με βάρος μόλις 743 γραμμάρια και διαστάσεις 83mm, 83mm και 198mm σε ύψος, πλάτος και μήκος αντίστοιχα όταν είναι σε κατάσταση δίπλωσης, ωστόσο έχει πολλές δυνατότητες. Το συμπαγές μέγεθός του κρύβει μία υψηλή πολυπλοκότητα που το καθιστά μία από τις πλέον εκλεπτυσμένες ιπτάμενες κάμερες που έχουν κατασκευαστεί από την DJI. Διαθέτει 24 υπολογιστικούς πυρήνες υψηλών επιδόσεων, ένα καινοτόμο σύστημα μετάδοσης με εμβέλεια 7 χιλιομέτρων, 5 αισθητήρες ορατότητας και μια κάμερα 4K σταθεροποιημένη από μηχανισμό τριών αξόνων. Μπορεί να υποστηρίξει μέχρι και 30 λεπτά πτήσης και διαθέτει σύστημα αποφυγής εμποδίων, το οποίο του επιτρέπει πτήσεις με μεγάλη ακρίβεια. Ο θόρυβος που προκαλεί το συγκεκριμένο μη επανδρωμένο είναι κατά 4dB χαμηλότερος σε σχέση με τα προηγούμενα μοντέλα, με αποτέλεσμα την βελτίωση της εμπειρίας της πτήσης. [25]



Το Μοντέλο σε Αναδίπλωση



Άνωθεν Όψη του Μοντέλου



Κάτωθεν Όψη του Μοντέλου



Η Κάμερα του Μοντέλου

Σχήμα 5.2.2.1: Mavic Pro

5.2.3 Χρήση αλγορίθμου ανίχνευσης αντικειμένων

Αφού ο χρήστης έχει τραβήξει τις φωτογραφίες και τα βίντεο με την βοήθεια της κάμερας του μη επανδρωμένου αεροσκάφους, το επόμενο στάδιο είναι η επεξεργασία των εικόνων και πιο συγκεκριμένα, η σάρωσή τους προκειμένου να ανιχνευτούν αυτοκίνητα. Ο αλγόριθμος εντοπισμού αντικειμένων που χρησιμοποιήθηκε είναι ο YOLOv3 και ο κώδικας που αφορά την ανίχνευση είναι γραμμένος σε γλώσσα προγραμματισμού Python. Ο αλγόριθμος YOLOv3 αποτελεί την τρίτη κατά σειρά έκδοση του YOLO. Περιέχει κάποιες μικρές διαφοροποιήσεις από τις προηγούμενες εκδόσεις με σκοπό να έχει καλύτερη αποτελεσματικότητα και μεγαλύτερη ταχύτητα στην εκτέλεση. Ακολούθως παρουσιάζεται ο τρόπος με τον οποίο ο YOLOv3 πραγματοποιεί τις εξής λειτουργίες:

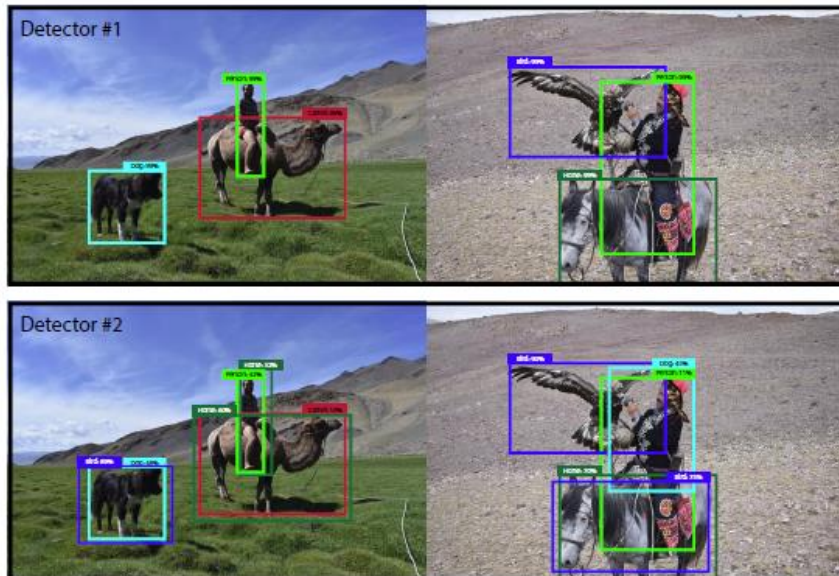
1. Πρόβλεψη κουτιού οριοθέτησης
2. Πρόβλεψη κλάσης
3. Προβλέψεις σε κλίμακες
4. Απόσπαση στοιχείων: Darknet-53
5. Εκπαίδευση

5.2.3.1 Πρόβλεψη κουτιού οριοθέτησης

Αυτό το στάδιο παραμένει ίδιο σε σύγκριση με την προηγούμενη έκδοση του αλγορίθμου, YOLOv2. Κατά την πρόβλεψη του κουτιού οριοθέτησης, προβλέπονται τέσσερις μεταβλητές, οι t_x , t_y , t_w και t_h , οι οποίες αντιστοιχούν στις συντεταγμένες του κέντρου του κουτιού (δισδιάστατος χώρος), το πλάτος και το ύψος του κουτιού αντίστοιχα. Επιπλέον, υπάρχει η βαθμολογία αντικειμενικότητας, η οποία προβλέπεται με τη χρήση λογιστικής παλινδρόμησης. Λαμβάνει την τιμή 1 αν το πλαίσιο οριοθέτησης προηγουμένως επικαλύπτει ένα γειτονικό αντικείμενο αλήθειας περισσότερο από οποιοδήποτε άλλο πλαίσιο οριοθέτησης. Για κάθε γειτονικό αντικείμενο ορίζεται μόνο ένα προηγούμενο πλαίσιο οριοθέτησης. Σε περίπτωση που το προηγούμενο κουτί οριοθέτησης δεν είναι το πλέον ταιριαστό, αλλά αλληλεπικαλύπτεται με ένα αντικείμενο αλήθειας πάνω από κάποιο κατώφλι αγνοούμε την πρόβλεψη. Ο YOLOv3 χρησιμοποιεί το κατώτατο όριο του 0.5.

5.2.3.2 Πρόβλεψη κλάσης

Κάθε παράθυρο προβλέπει τις κατηγορίες τις οποίες μπορεί να περιέχει το πλαίσιο οριοθέτησης χρησιμοποιώντας ταξινόμηση πολλαπλών ετικετών. Δεν χρησιμοποιείται softmax μιας και όπως διαπιστώθηκε, δεν είναι απαραίτητο προκειμένου να επιτευχθούν υψηλές επιδόσεις. Αντ' αυτού χρησιμοποιούνται ανεξάρτητοι υλικοτεχνικοί ταξινομητές. Κατά τη διάρκεια της εκπαίδευσης γίνεται χρήση δυαδικής cross-entropy απώλειας για την τάξη πρόβλεψης. Αυτή η διατύπωση βοηθάει όταν μεταβαίνουμε σε πιο περίπλοκους τομείς όπως το σύνολο δεδομένων εικόνων ανοιχτού κώδικα (Open Images Dataset). Σε αυτό το σύνολο δεδομένων υπάρχουν πολλές επικαλυπτόμενες ετικέτες, όπως για παράδειγμα η ετικέτα Γυναίκα με την ετικέτα Άνθρωπος. Σε περίπτωση που γινόταν χρήση softmax, θα υπήρχε επιβολή της υπόθεσης πως κάθε κουτί έχει ακριβώς μια κατηγορία, γεγονός που συχνά δεν συμβαίνει. Μια πολυεπίπεδη προσέγγιση μπορεί να προβλέψει με περισσότερη επιτυχία τα δεδομένα. Όπως φαίνεται στην ακόλουθη εικόνα, οι διαφορές μεταξύ των προβλεπόμενων και των πραγματικών πλαισίων οριοθέτησης είναι σχεδόν μηδαμινές.



Σχήμα 5.2.3.2.1: Διαφορά μεταξύ προβλεπόμενων και πραγματικών κουτιών

5.2.3.3 Προβλέψεις σε κλίμακες

Ο YOLOv3 προβλέπει τα κουτιά οριοθέτησης, χρησιμοποιώντας 3 διαφορετικές κλίμακες. Το σύστημα εξάγει χαρακτηριστικά από αυτές τις κλίμακες όπως το FPN (Feature Pyramid Network). Επιπλέον προστίθενται αρκετά στρώματα συνέλιξης στον βασικό αποσπαστή χαρακτηριστικών, το Darknet-53, το οποίο αναλύεται στην συνέχεια. Η τελευταία από αυτές τις στρώσεις προβλέπει το πλαίσιο οριοθέτησης, την αντικειμενικότητα και τις προβλέψεις κλάσης. Στα πειράματα που διεξάχθηκαν από τους δημιουργούς του αλγορίθμου με το σύνολο δεδομένων COCO, προβλέφθηκαν 3 κουτιά σε κάθε κλίμακα. Επομένως, ο τανυστής εξόδου είναι $N \times N \times [3$

$\times (4 + 1 + 80)$], δηλ. 4 μετατοπίσεις πλαισίου, 1 πρόβλεψη αντικειμενικότητας και 80 προβλέψεις κλάσης.

Στη συνέχεια, λαμβάνεται ο χάρτης χαρακτηριστικών από τα δύο προηγούμενα επίπεδα και δειγματοληπτούμε ξανά με $\times 2$. Επίσης παίρνουμε έναν προηγούμενο χάρτη χαρακτηριστικών από το δίκτυο και τον συγχωνεύουμε με τα χαρακτηριστικά που προκύπτουν από αυτόν που έχουμε εκτελέσει αναδειγματοληψία, χρησιμοποιώντας συγκόλληση. Αυτή είναι η τυπική αρχιτεκτονική κωδικοποιητή-αποκωδικοποιητή. Αυτή η μέθοδος μας επιτρέπει να έχουμε πιο σημαντικές σημασιολογικές πληροφορίες από τις βελτιωμένες λειτουργίες και τις πιο λεπτομερείς πληροφορίες από τον παλαιότερο χάρτη χαρακτηριστικών.

Κατόπιν, προστίθενται μερικά επιπλέον συνθετικά στρώματα για την επεξεργασία αυτού του νέου χάρτη χαρακτηριστικών και τελικά να επέλθει η πρόβλεψη ενός παρόμοιου τανυστή με το διπλάσιο μέγεθος. Ο ίδιος σχεδιασμός εκτελείται για μια ακόμη φορά με σκοπό την πρόβλεψη των κουτιών για την τελική κλίμακα. Έτσι οι προβλέψεις για την τρίτη κλίμακα επωφελούνται από το σύνολο των προηγούμενων υπολογισμών, καθώς και από τα λεπτά χαρακτηριστικά από το αρχικό δίκτυο. Τέλος, και αυτή η έκδοση του αλγορίθμου χρησιμοποιεί την ομαδοποίηση k-means.

5.2.3.4 Απόσπαση στοιχείων: Darknet-53

Σε αντίθεση με το YOLOv2, το οποίο χρησιμοποιεί το δίκτυο ταξινόμησης Darknet-19 για την εξαγωγή των χαρακτηριστικών, το YOLOv3 χρησιμοποιείται ένα πολύ βαθύτερο δίκτυο, το Darknet-53, το οποίο περιέχει 53 στρώματα συνέλιξης. Το Darknet-53 αποτελεί ένα υβρίδιο μεταξύ των Darknet-19 και ενός νέου ταξινομητή δικτύου. Αυτό το νέο δίκτυο είναι πολύ πιο ισχυρό από το Darknet-19 και πιο αποδοτικό από το ResNet-101 ή το ResNet-152. Το Darknet-53 επιτυγχάνει επίσης τον υψηλότερο αριθμό λειτουργιών κινητής υποδιαστολής ανά δευτερόλεπτο. Αυτό σημαίνει ότι η δομή του δικτύου χρησιμοποιεί καλύτερα τη GPU, καθιστώντας την πιο αποτελεσματική για την αξιολόγηση και επομένως ταχύτερη. Αυτό οφείλεται κυρίως στο γεγονός ότι τα ResNets έχουν πάρα πολλά στρώματα και δεν είναι πολύ αποτελεσματικά.

Κάθε δίκτυο εκπαιδεύεται με τις ίδιες ρυθμίσεις και δοκιμάζεται σε ακρίβεια ενιαίας αποκοπής 256×256 . Οι χρόνοι εκτέλεσης μετριούνται σε ένα Titan X στα 256×256 . Το Darknet-53 είναι καλύτερο από ResNet-101 από άποψη ακρίβειας και σύμφωνα με μετρήσεις είναι 1,5 φορές γρηγορότερο. Το Darknet-53 έχει παρόμοια απόδοση με το ResNet-152 και είναι 2 φορές ταχύτερο.

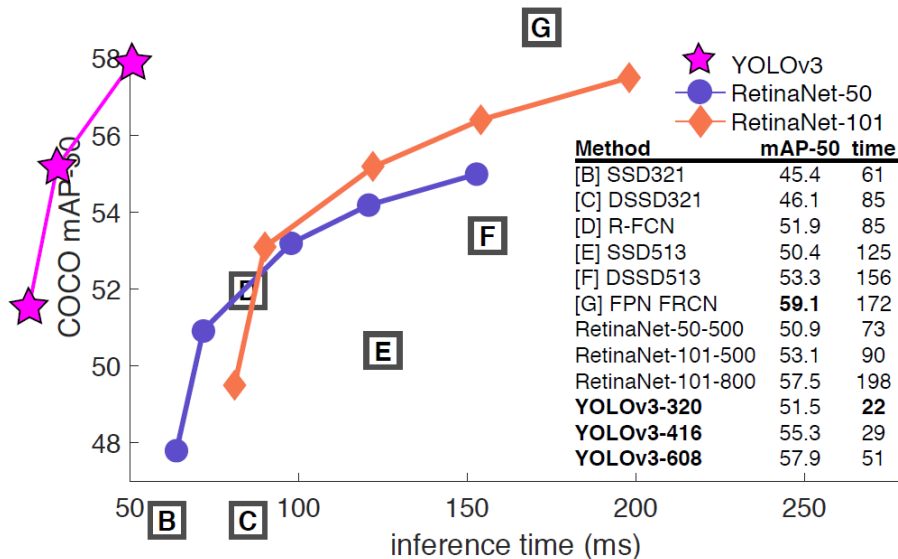
5.2.3.5 Εκπαίδευση

Ο αλγόριθμος εκπαιδεύεται σε πλήρεις εικόνες χωρίς σκληρή αρνητική εξόρυξη με το πλαίσιο νευρωνικού δικτύου Darknet, το οποίο είναι κατάλληλο για εκπαίδευση και για δοκιμές. Χρησιμοποιείται εκπαίδευση πολλαπλών επιπέδων και εξομάλυνση παρτίδας.

5.2.3.6 Συμπεράσματα

Συμπερασματικά, ο YOLOv3 είναι ένας εξαιρετικά αξιόπιστος αλγόριθμος, ο οποίος χαρακτηρίζεται από την ταχύτητά του. Οι ανάγκες της εφαρμογής που αναπτύχθηκε στο πλαίσιο

της προκειμένης Διπλωματικής Εργασίας απαιτούν έναν αλγόριθμο που να μπορεί να επεξεργαστεί τον όγκο των δεδομένων, ο οποίος ομολογουμένως σε αυτό το στάδιο δεν είναι αρκετά μεγάλος, σε μικρό χρονικό διάστημα, ώστε να παρέχεται το αποτέλεσμα στον χρήστη σε πραγματικό χρόνο. Για αυτό επιλέχθηκε ο συγκεκριμένος αλγόριθμος. Όπως φαίνεται και από το παρακάτω διάγραμμα, ο YOLOv3 έχει πολύ καλά αποτελέσματα με χρήση του COCO dataset. Είναι σχεδόν χίλιες φορές γρηγορότερος συγκριτικά με το Fast R-CNN. [26]



Σχήμα 5.2.3.6.1: Σύγκριση αλγορίθμων με χρήση του COCO dataset σε 0.5 mAP

5.3 Γραφικό περιβάλλον και παρουσίαση του κώδικα της εφαρμογής

5.3.1 Υλοποίηση της εφαρμογής

Ο κώδικας της εφαρμογής γράφτηκε στη γλώσσα προγραμματισμού Swift, όπως έχει ήδη αναφερθεί και για την υλοποίηση του εργαστήκαμε στο περιβάλλον ανάπτυξης Xcode. Για τις περισσότερες λειτουργίες της εφαρμογής χρησιμοποιείται το Mobile SDK που παρέχει η εταιρεία DJI. Παρακάτω παρουσιάζονται τα διάφορα στάδια της εφαρμογής, καθώς και τα σημαντικότερα μέρη του κώδικα κάθε φορά. [27]

Ενεργοποίηση εφαρμογής και σύνδεση αεροσκάφους

Καταρχάς, είναι απαραίτητη η ενεργοποίηση της εφαρμογής από τον προγραμματιστή. Αυτό συμβαίνει μόνο την πρώτη φορά που ενεργοποιείται η εφαρμογή. Συγκεκριμένα, για την ενεργοποίηση της εφαρμογής, ο προγραμματιστής θα πρέπει να έχει δημιουργήσει ένα μοναδικό κλειδί στην ιστοσελίδα της DJI, προκειμένου να αποκτήσει το δικαίωμα ανάπτυξης μίας εφαρμογής κάνοντας χρήση του SDK της DJI. Σε περίπτωση που υπάρχει κάποιο λάθος κατά την διαδικασία ενεργοποίησης της εφαρμογής, εμφανίζεται σχετικό μήνυμα. Ο κώδικας που αφορά το κομμάτι αυτό είναι ο εξής:

```

func appRegisteredWithError(_ error: Error?) {
    var message = "Register App Succeeded!"
    if (error != nil) {
        message = "Register App Failed! Please enter App Key and check the
network."
    }
    else {
        DJISDKManager.startConnectionToProduct()
        DJISDKManager.appActivationManager().delegate = self
        activationState =
DJISDKManager.appActivationManager().appActivationState
        aircraftBindingState =
DJISDKManager.appActivationManager().aircraftBindingState
    }
    showAlertView(withTitle: "App Registration", withMessage: message)
}

func manager(_ manager: DJIAppActivationManager?, didUpdate
appActivationState: DJIAppActivationState) {
    activationState = appActivationState
    updateUI()
}

```

Σχήμα 5.3.1.1: Κώδικας εγγραφής της εφαρμογής και σύνδεσης του αεροσκάφους

Σύμφωνα με τον παραπάνω κώδικα, σε περίπτωση που δεν συναντάται κάποιο σφάλμα κατά την εγγραφή της εφαρμογής, δηλαδή το κλειδί που ελήφθει είναι πράγματι σωστό και μοναδικό, εκκινεί η σύνδεση με το προϊόν, το οποίο στην προκειμένη περίπτωση είναι το μοντέλο Mavic Pro Platinum της DJI. Αυτό επιτυγχάνεται με την εντολή `DJISDKManager.startConnectionToProduct()`.

Η εφαρμογή που δημιουργήθηκε δεν απαιτεί για την εκκίνησή της, την εισαγωγή στοιχείων από τον χρήστη. Αρχικά, όταν το μη επανδρωμένο αεροσκάφος συνδεθεί με την κινητή συσκευή, μέσω του τηλεχειριστηρίου, εμφανίζεται σχετικό μήνυμα σύνδεσης στην πρώτη οθόνη της εφαρμογής. Επιπλέον, κατά την επιτυχή σύνδεση, αναγνωρίζεται και εμφανίζεται στην οθόνη το μοντέλο του μη επανδρωμένου αεροσκάφους. Αν δεν έχει πραγματοποιηθεί σύνδεση με κάποιο μη επανδρωμένο αεροσκάφος ή υπάρχει κάποιο πρόβλημα κατά την σύνδεση, εμφανίζεται σχετικό μήνυμα στον χρήστη και στην θέση όπου αναμένεται να εμφανίζεται το μοντέλο του αεροσκάφους, υπάρχει η ένδειξη *Unknown*. Παρακάτω παρουσιάζεται ο κώδικας ελέγχου σύνδεσης της εφαρμογής με κάποιο μη επανδρωμένο. Επίσης παρατίθενται τα κομμάτια κώδικα που αντιστοιχούν στις δύο καταστάσεις που παρουσιάστηκαν (επιτυχής και ανεπιτυχής σύνδεση):

```

func updateStatusBased(on newConnectedProduct: DJIBaseProduct?) {
    if newConnectedProduct != nil {
        connectStatusLabel.text = "Status: Product Connected"
        let model = newConnectedProduct?.model
        modelNameLabel.text = "Model: \ \(model!)"
    }
    else {
        connectStatusLabel.text = "Status: Product Not Connected"
        modelNameLabel.text = "Model: Unknown"
    }
}

func productConnected(_ product: DJIBaseProduct?) {
    NSLog("Product Connected")
    updateStatusBased(on: product)
}

func productDisconnected() {
    NSLog("Product Disconnected")
    let product = DJIBaseProduct()
    updateStatusBased(on: product)
}

func manager(_ manager: DJIAppActivationManager?, didUpdate
aircraftBindingState: DJIAppActivationAircraftBindingState) {
    self.aircraftBindingState = aircraftBindingState
    updateUI()
}

```

Σχήμα 5.3.1.2: Σύνδεση εφαρμογής με το μη επανδρωμένο αεροσκάφος

Σε περίπτωση που το newConnectedProduct είναι διάφορο του nil, σημαίνει πως έχει ανιχνευτεί μη επανδρωμένο αεροσκάφος. Σε αυτή την περίπτωση, ο χρήστης ενημερώνεται με σχετικό μήνυμα καθώς επίσης ανανεώνεται η ετικέτα που αφορά το μοντέλο του αεροσκάφους. Επίσης ενημερώνεται η κατάσταση σύνδεσης του αεροσκάφους, έτσι ώστε αν υπάρχει συνδεδεμένο αεροσκάφος να μην είναι δυνατή νέα σύνδεση. Σε περίπτωση που το αεροσκάφος για οποιοδήποτε λόγο σταματήσει την σύνδεση με την συσκευή, η εφαρμογή ενημερώνεται και πάλι, προκειμένου να πραγματοποιηθεί εκ νέου η σύνδεση ή να δημιουργηθεί νέα σύνδεση.

Δημιουργία του map view και προσθήκη των waypoints

Στο δεύτερο στάδιο της εφαρμογής, αφού έχει ανιχνευτεί επιτυχής σύνδεση, εμφανίζεται η θέση του αεροσκάφους στον χάρτη. Η εφαρμογή κάνει αυτόματη εστίαση στην συγκεκριμένη θέση έτσι ώστε, αφενός να καθίσταται δυνατός ο έλεγχος της ορθότητά της και αφετέρου να διευκολύνεται η τοποθέτηση σημείων στον χάρτη που να σηματοδοτούν την διαδρομή που θα ακολουθηθεί. Ο χρήστης είναι αυτός που επιλέγει τα σημεία σύμφωνα με τα οποία θα κινηθεί το μη επανδρωμένο. Αυτά εμφανίζονται πάνω στον χάρτη με την μορφή πινεζών σε μωβ χρώμα.


```

//shows the map screenshot
@IBOutlet weak var mapView: MKMapView!

//focus mapView to the drone's current location.
func focusMap() {
    debugview.text += "drone location\(\droneLocation)\n"
    if CLLocationCoordinate2DIsValid(droneLocation) {
        debugview.text += "inside focus button for drone location\n"
        var region: MKCoordinateRegion
        region = MKCoordinateRegion()
        region.center = droneLocation
        region.span.latitudeDelta = CLLocationDegrees(0.001)
        region.span.longitudeDelta = CLLocationDegrees(0.001)
        mapView?.setRegion(region, animated: true)
    }
}

override func viewWillAppear(_ animated: Bool) {
    super.viewWillAppear(animated)
    startUpdateLocation()
}

//startUpdateLocation is stopped
override func viewWillDisappear(_ animated: Bool) {
    super.viewWillDisappear(animated)
    locationManager.stopUpdatingLocation()
}

```

Σχήμα 5.3.1.3: Εστίαση στην θέση του αεροσκάφους στον χάρτη

```

public class DJIAircraftAnnotationView: MKAnnotationView {

    override init(annotation: MKAnnotation?, reuseIdentifier: String?) {
        super.init(annotation: annotation, reuseIdentifier: reuseIdentifier)
        isEnabled = false
        isDraggable = false
        image = imageLiteral(resourceName: "aircraft.png")
    }

    func updateHeading(_ heading: Float) { //changes aircraft's rotation
        transform = CGAffineTransform.identity
        transform = CGAffineTransform(rotationAngle: CGFloat(heading))
    }
}

```

Σχήμα 5.3.1.4: Εμφάνιση του αεροσκάφους στον χάρτη

Αφού έχει πραγματοποιηθεί η εστίαση στην θέση του αεροσκάφους, η οποία λογικά θα συμπίπτει με την θέση του χρήστη, γεγονός που διευκολύνει τον έλεγχο της αξιοπιστίας της εστίασης, ο χρήστης πρέπει να τοποθετήσει τα σημεία στον χάρτη. Η ενέργεια αυτή πραγματοποιείται όταν ο χρήστης αγγίζει την οθόνη του κινητής συσκευής στο επιθυμητό σημείο (tap action). Αυτό ανιχνεύεται από την εφαρμογή, εμφανίζεται σχετικό μήνυμα στην περιοχή debugging και ο συντεταγμένες του σημείου αποθηκεύονται σε έναν πίνακα, έτσι ώστε να τροφοδοτηθούν στο αεροσκάφος με την εκκίνηση της πτήσης του. Ο χρήστης έχει την δυνατότητα να προσθέτει και να αφαιρεί waypoints κατά βούληση, αρκεί αυτό να συμβαίνει πριν το αεροσκάφος ξεκινήσει την διαδρομή του και ο αριθμός τους να μην είναι λιγότερος από δύο.

```
func mapView(_ mapView: MKMapView, viewFor annotation: MKAnnotation) ->
MKAnnotationView? {
    debugview.text += "inside map view\n"
    if (annotation is MKPointAnnotation) {
        let pinView = MKPinAnnotationView(annotation: annotation,
reuseIdentifier: "Pin_Annotation")
        pinView.pinTintColor = .purple
        return pinView
    }
    else if (annotation is DJIAircraftAnnotation) {
        let annoView = DJIAircraftAnnotationView(annotation: annotation,
reuseIdentifier: "Aircraft_Annotation")
        (annotation as? DJIAircraftAnnotation)?.annotationView = annoView
        return annoView
    }
}
return nil
}
```

Σχήμα 5.3.1.5: Αναγνώριση σημείων ως waypoints

```
//adds waypoints by tapping
@objc func addWaypoints(_ gesture: UITapGestureRecognizer?) {
    let point: CGPoint? = gesture?.location(in: mapView)
    if gesture?.state == .ended {
        if isEditingPoints {
            mapController.add(point!, with: mapView)
        }
    }
}

func addBtn(_ button: UIButton?, withActionInGSButtonVC GSBtnVC:
DJIGSButtonController?) {
    if isEditingPoints {
        isEditingPoints = false
        button?.setTitle("Add", for: .normal)
    }
    else {
        isEditingPoints = true
        button?.setTitle("Finished", for: .normal)
    }
}
```

Σχήμα 5.3.1.6: Προσθήκη σημείων στον χάρτη

```

func clearBtnAction(inGSButtonVC GSBtnVC: DJIGSButtonController?) {
    mapController.cleanAllPoints(with: mapView)
}

func configBtnAction(inGSButtonVC GSBtnVC: DJIGSButtonController?) {
    let editPoints = mapController.editPoints
    if (editPoints == [] || editPoints.count < 2) {
        //DJIWaypointMissionMinimumWaypointCount is 2.
        debugview.text += "No or not enough waypoints for mission\n"
        return
    }

    UIView.animate(withDuration: 0.25, animations: {
        self.waypointConfigVC.view.alpha = 1.0
    })
    waypointMission.removeAllWaypoints()

    for i in 0..editPoints.count {
        let location = editPoints[i] as? CLLocation
        if CLLocationCoordinate2DIsValid((location?.coordinate)!) {
            var waypoint: DJIWaypoint? = nil
            if let coordinate = location?.coordinate {
                waypoint = DJIWaypoint(coordinate: coordinate)
            }
            waypointMission.add(waypoint!)
        }
    }
}

```

Σχήμα 5.3.1.7: Έλεγχος επαρκούς αριθμού σημείων και διαγραφή σημείων

Μετά την προσθήκη των επιθυμητών σημείων που καθορίζουν την διαδρομή που έχει επιλέξει ο χρήστης για το μη επανδρωμένο αεροσκάφος, πρέπει οι συντεταγμένες των σημείων να παρουσιαστούν στο αεροσκάφος ως σημεία στόχος. Το αεροσκάφος θα κινηθεί ανάλογα με την σειρά προσθήκης των σημείων, δηλαδή έτσι όπως έχουν αποθηκευτεί στον πίνακα των συντεταγμένων-στόχων. Προκειμένου το αεροσκάφος να μπορέσει να κινηθεί προς το επόμενο σημείο, θα πρέπει να γίνει ενημέρωση των συντεταγμένων, καθώς και μίας μεταβλητής η οποία καθορίζει την κατεύθυνση προς την οποία κινείται το μπροστινό μέρος του αεροσκάφους. Η μεταβλητή αυτή έχει ονομαστεί *heading*. Η διαδικασία που μόλις περιγράφεται υλοποιείται με κώδικα όπως φαίνεται παρακάτω:

```

public class DJIAircraftAnnotation: NSObject {
    var coordinate = CLLocationCoordinate2D()
    var annotationView = DJIAircraftAnnotationView()

    init(coordinate: CLLocationCoordinate2D) {
        super.init()
        self.coordinate = coordinate
    }

    func setCoordinate(_ newCoordinate: CLLocationCoordinate2D) {
        coordinate = newCoordinate
    }

    //updates DJIAircraftAnnotationView's heading
    func updateHeading(_ heading: Float) {
        annotationView.updateHeading(heading)
    }
}

```

Σχήμα 5.3.1.8: Ενημέρωση των συντεταγμένων και της κατεύθυνσης κίνησης του drone

Επιπλέον, προκειμένου η κίνηση του μη επανδρωμένου να μπορεί να ελέγχεται καθ' όλη την διάρκεια της πτήσης, θα πρέπει να εφαρμοστεί ένας flight controller ως εξής:

```

func flightController(_ fc: DJIFlightController, didUpdate state:
DJIFlightControllerState) {
    if(state.aircraftLocation != nil) {
        drone.position.latitude = (state.aircraftLocation?.coordinate.latitude)!
        drone.position.longitude =
(state.aircraftLocation?.coordinate.longitude)!
        let heading = state.attitude.yaw
        heading >= 0 ? heading : heading + 360.0
        drone.rotation = heading
        let dronelat = drone.position.latitude
        let dronelong = drone.position.longitude
        let camera = GMSCameraPosition.camera(withLatitude: dronelat,
longitude: dronelong, zoom: 16.0)
        mapView.camera = camera
    }
    else {
        showAlertView(withTitle: "Drone not in GPS range", withMessage:
"Error")
    }
}
}

```

Σχήμα 5.3.1.9: Εφαρμογή του flight controller για τον έλεγχο της κίνησης του drone

Καθορισμός μεταβλητών που αφορούν την πτήση

Ο χρήστης έχει την δυνατότητα να καθορίζει κάποιες μεταβλητές που αφορούν την ποιότητα πτήσης του αεροσκάφους. Για παράδειγμα, είναι σε θέση να ρυθμίζει το υψόμετρο στο οποίο επιθυμεί να πετά το drone, το οποίο θα παραμένει σταθερό, μέχρι να του δοθεί η οδηγία να επιστρέψει στην θέση που έχει οριστεί ως *Home*, είτε μετά το πέρας της αποστολής, είτε λόγω κάποιας απρόσμενης εξέλιξης, όπως οι καιρικές συνθήκες ή μη επαρκής εναπομείνουσα μπαταρία. Εκτός από το υψόμετρο, άλλα μεγέθη τα οποία επιδέχονται ρύθμιση από τον χρήστη είναι η ταχύτητα αυτόματης πτήσης, η μέγιστη ταχύτητα πτήσης, η θέση επιστροφής και η κατεύθυνση του αεροσκάφους. Η επεξεργασία των παραπάνω μεταβλητών επιτυγχάνεται με το πάτημα του κουμπιού config, για το οποίο εκτενέστερος λόγος θα γίνει στην παρουσίαση του γραφικού περιβάλλοντος της εφαρμογής. Σε περίπτωση που ο χρήστης δεν επιθυμεί να επεξεργαστεί τα εν δυνάμει μεταβαλλόμενα μεγέθη που αναφέρθηκαν, έχουν οριστεί κάποιες default τιμές, όπως φαίνεται στην συνέχεια:

```
func initUI() {
    altitudeTextField.text = "100" //Set the altitude to 100
    autoFlightSpeedTextField.text = "8" //Set the autoFlightSpeed to 8
    maxFlightSpeedTextField.text = "10" //Set the maxFlightSpeed to 10
    actionSegmentControl.selectedSegmentIndex = 1 //Set the finishAction to
    DJIWaypointMissionFinishedGoHome
    headingSegmentControl.selectedSegmentIndex = 0 //Set the headingMode to
    DJIWaypointMissionHeadingAuto
}
}
```

Σχήμα 5.3.1.10: Εισαγωγή default τιμών για μεγέθη που αφορούν την πτήση

```
func cancelBtnAction(in waypointConfigVC: DJIWaypointConfigViewController?) {
    UIView.animate(withDuration: 0.25, animations: {
        self.waypointConfigVC.view.alpha = 0
    })
}

func finishBtnAction(in waypointConfigVC: DJIWaypointConfigViewController?) {
    UIView.animate(withDuration: 0.25, animations: {
        self.waypointConfigVC.view.alpha = 0
    })
    for i in 0..
```

Σχήμα 5.3.1.11: Μετατροπή των χαρακτήρων σε πραγματικούς αριθμούς και finish, cancel buttons

Σε αυτό το σημείο, ο χρήστης είναι σε θέση να εκκινήσει την εφαρμογή συνδέοντας το μη επανδρωμένο με το gimbal και την συσκευή της Apple που φέρει την εφαρμογή και να διαπιστώσει αν η σύνδεση είναι επιτυχής. Σε περίπτωση επιτυχούς σύνδεσης, ο χρήστης παρατηρεί στον χάρτη το σημείο όπου βρίσκεται το drone. Πριν την έναρξη της πτήσης, τοποθετούνται τα waypoints στον χάρτη σύμφωνα με τα οποία θα διαγράψει την πορεία του το μη επανδρωμένο. Μπορούν να προστεθούν ή να αφαιρεθούν σημεία. Επίσης ο χρήστης μπορεί να επέμβει σε μεγέθη όπως η μέγιστη ταχύτητα πτήσης. Όταν έχουν ολοκληρωθεί τα προαναφερθέντα βήματα, το αεροσκάφος είναι έτοιμο για πτήση.

Λειτουργίας της ενσωματωμένης κάμερας

Στο πλαίσιο της εφαρμογής που δημιουργήθηκε για τον σκοπό της Διπλωματικής εργασίας, η κάμερα η οποία είναι ενσωματωμένη στο μη επανδρωμένο μπορεί να επιτελέσει τις εξής δύο λειτουργίες: λήψη φωτογραφιών και βιντεοσκόπηση. Στην πραγματικότητα η ανίχνευση αντικειμένων είναι πολύ ευκολότερη από φωτογραφίες σε αντίθεση με τα βίντεο, μιας και τα αντικείμενα είναι στατικά και μία σάρωση ολόκληρη της εικόνας αρκεί προκειμένου να αποφανθούμε για την ύπαρξη αντικειμένων. Έτσι, η αναγνώριση αντικειμένων μέσω στατικών εικόνων αποδεικνύεται πολύ πιο αποτελεσματική, για αυτό και αποτελεί προτεραιότητα για την εργασία.

Παρακάτω παρατίθενται οι μέθοδοι `setupVideoPreviewer` και `resetVideoPreview`. Στην πρώτη εκ των δύο μεθόδων ορίζεται η μεταβλητή `frnPreviewView` ως `Superview` του `MovieGLView` της τάξης (class) `DJIVideoPreviewer`, προκειμένου να εμφανίζεται το video stream και στη συνέχεια να δημιουργείται ένα αντικείμενο της κλάσης `DJIBaseProduct`, του οποίου ελέγχουμε την ιδιότητα μέσω μίας εντολής συνθήκης `if`.

Σε περίπτωση που το μοντέλο του μη επανδρωμένου είναι ένα από τα: A3, N3, Matrice 600 ή Matrice 600 Pro, καλείται η μέθοδος `add(self, with: nil)` της `DJIVideoFeeder` για να προστεθεί το `DJICameraViewControlller` ως listener του `secondaryVideoFeed` για τροφοδοσία βίντεο. Διαφορετικά, προσθέτουμε το `DJICameraViewControlller` ως ακροατή του `primaryVideoFeed` για βίντεο feed. Τέλος, γίνεται κλήση της μεθόδου `start` του `DJIVideoPreviewer` για να ξεκινήσει η αποκωδικοποίηση του βίντεο.

Επιπλέον, στη μέθοδο `resetVideoPreview`, καλούμε την μέθοδο `unSetView` του `DJIVideoPreviewer`, προκειμένου καταρχάς να αφαιρεθεί πρώτα η `MovieGLView` της κλάσης `DJIVideoPreviewer` από την εφαρμογή `frnPreviewView`. Στη συνέχεια, ελέγχεται επίσης το μοντέλο του τρέχοντος προϊόντος με μια εντολή `if`. Εάν το μοντέλο του drone είναι A3, N3, Matrice 600 ή Matrice 600 Pro, καλείται η μέθοδος `remove(self)` της κλάσης `DJIVideoFeeder` για την αφαίρεση του ακροατή `DJICameraViewControlller` από το `secondaryVideoFeed` για την τροφοδοσία βίντεο, διαφορετικά αφαιρούμε τον ακροατή `DJICameraViewControlller` από το `primaryVideoFeed`.

```

func setupVideoPreviewer() {
    DJIVideoPreviewer.instance()?.setView(fpvPreviewView) //to show the Video
    Stream first
    let product: DJIBaseProduct? = DJISDKManager.product()
    if product?.model == DJIAircraftModelNameA3 || product?.model ==
    DJIAircraftModelNameN3 || product?.model == DJIAircraftModelNameMatrice600 ||
    product?.model == DJIAircraftModelNameMatrice600Pro {
        DJISDKManager.videoFeeder()?.secondaryVideoFeed.add(self, with: nil)
    }
    else {
        DJISDKManager.videoFeeder()?.primaryVideoFeed.add(self, with: nil)
    }
    DJIVideoPreviewer.instance().start()
}

func resetVideoPreview() {
    DJIVideoPreviewer.instance().unSetView()
    let product: DJIBaseProduct? = DJISDKManager.product()
    if product?.model == DJIAircraftModelNameA3 || product?.model ==
    DJIAircraftModelNameN3 || product?.model == DJIAircraftModelNameMatrice600 ||
    product?.model == DJIAircraftModelNameMatrice600Pro {
        DJISDKManager.videoFeeder()?.secondaryVideoFeed.remove(self)
    }
    else {
        DJISDKManager.videoFeeder()?.primaryVideoFeed.remove(self)
    }
}

```

Σχήμα 5.3.1.12: Κλήση μεθόδων `setupVideoPreviewer()` και `resetVideoPreview()` για την λειτουργία της κάμερας

Κατόπιν υλοποιούνται οι delegate μέθοδοι της `DJISDKManagerDelegate` οι: `fetchCamera`, `productConnected` και `productDisconnected`, καθώς και η μέθοδος `viewWillDisappear`, όπως φαίνεται παρακάτω.

```

//DJISDKManagerDelegate delegate method
func fetchCamera() -> DJICamera? {
    if !(DJISDKManager.product() != nil) {
        return nil
    }
    if (DJISDKManager.product() is DJIAircraft) {
        return (DJISDKManager.product() as? DJIAircraft)?.camera
    }
    else if (DJISDKManager.product() is DJIHandheld) {
        return (DJISDKManager.product() as? DJIHandheld)?.camera
    }
    return nil
}

```

```

var prodCon = false
//DJISDKManagerDelegate delegate method
func productConnected(_ product: DJIBaseProduct?) {
    if product != nil {
        debugView.text += "product connected\n"
        product?.delegate = self as? DJIBaseProductDelegate
        let camera: DJICamera? = fetchCamera()
        if camera != nil {
            camera?.delegate = self
            prodCon = true
        }
        setupVideoPreviewer()
    }
    else {
        debugView.text += "product connection not successful\n"
    }
}

//DJISDKManagerDelegate delegate method
func productDisconnected() {
    let camera: DJICamera? = fetchCamera()
    if prodCon {
        camera?.delegate = nil
    }
    resetVideoPreview()
}

override func viewWillDisappear(_ animated: Bool) {
    super.viewWillDisappear(animated)
    let camera: DJICamera? = fetchCamera()
    if prodCon {
        camera?.delegate = nil
    }
    resetVideoPreview()
}

```

Σχήμα 5.3.1.13: Οι delegate μέθοδοι της DJISDKManagerDelegate και η viewWillDisappear

Πρώτον, δημιουργείται η μέθοδος - (DJICamera *) fetchCamera για τη λήψη του ενημερωμένου αντικείμενου τύπου DJICamera. Πριν επιστραφεί το αντικείμενο DJICamera, πρέπει να ελεγχθεί αν το αντικείμενο του προϊόντος DJISDKManager ανήκει στην κλάση DJIAircraft ή στην κλάση DJIHandheld. Δεδομένου ότι το στοιχείο της κάμερας του αεροσκάφους ή της συσκευής χειρός μπορεί να αλλάξει ή να αποσυνδεθεί, το αντικείμενο της κάμερας πρέπει να ελέγχεται κάθε φορά που καλείται προς χρήση προκειμένου να διασφαλίζεται πως εκτελούνται οι σωστές μέθοδοι για το κάθε αντικείμενο κάμερας αντίστοιχα.

Όταν ανιχνεύεται συνδεδεμένο αντικείμενο, καλείται η μέθοδος productConnected, η οποία με την σειρά της ενεργοποιεί την setupVideoPreviewer, ενώ όταν το αντικείμενο αποσυνδέεται η productDisconnected καλεί την resetVideoPreview. Μέχρι στιγμής ο παραπάνω κώδικας είναι αρκετός ώστε να δώσει την δυνατότητα στον χρήστη να παρακολουθήσει ό,τι καταγράφεται από την κάμερα του αεροσκάφους σε πραγματικό χρόνο(first person view).

Προκειμένου ο χρήστης να είναι σε θέση να τραβάει φωτογραφίες όποτε το επιθυμεί, θα πρέπει να αναπτυχθεί η συνάρτηση captureAction, ενώ για να είναι δυνατή η λήψη βίντεο, απαραίτητη είναι η δημιουργία της συνάρτησης recordAction. Για να μπορεί να γίνει μετάβαση μεταξύ της κατάστασης capture που αφορά τη λήψη φωτογραφιών και της record που αφορά την

καταγραφή βίντεο, αναπτύχθηκε η συνάρτηση `changeWorkModeAction` όπως φαίνεται στις παρακάτω εικόνες.

```
@IBAction func captureAction(_ sender: Any) {
    let camera: DJICamera? = fetchCamera()
    if camera != nil {
        camera?.setShootPhotoMode(DJICameraShootPhotoMode(rawValue: 0x00)!
, withCompletion:
        //invoke the following method with 1 second delay to control the camera to
shoot a photo

        { error in
            DispatchQueue.main.asyncAfter(deadline: DispatchTime.now() +
Double(Int64(1 * Double(NSEC_PER_SEC))) / Double(NSEC_PER_SEC),
            execute: {
                camera?.startShootPhoto(completion: { error in
                    if error != nil {
                        var alertMessage: String? = nil
                        alertMessage = "Take Photo Error: \(error.debugDescription)"
                        self.debuggView.text += alertMessage! + "\n"
                    }
                })
            })
        })
    }
}
```

Σχήμα 5.3.1.14: Η συνάρτηση που αφορά την λήψη φωτογραφιών

```
@IBAction func recordAction(_ sender: Any) {
    let camera: DJICamera? = fetchCamera()
    if camera != nil {
        var alertForError : String? = nil
        if isRecording {
            camera?.stopRecordVideo(completion: { error in
                if error != nil {
                    alertForError = "Stop Record Video Error:\(error.debugDescription)"
                    self.debuggView.text += alertForError! + "\n"
                }
            })
        }
        else {
            camera?.startRecordVideo(completion: { error in
                if error != nil {
                    alertForError = "Start Record Video Error:\(error.debugDescription)"
                    self.debuggView.text += alertForError! + "\n"
                }
            })
        }
    }
}
```

Σχήμα 5.3.1.15: Η συνάρτηση που αφορά την καταγραφή βίντεο

```

@IBAction func changeWorkModeAction(_ sender: Any) {
    let segmentControl = sender as? UISegmentedControl
    let camera: DJICamera? = fetchCamera()

    if camera != nil {
        if segmentControl?.selectedSegmentIndex == 0 {
            //Take photo
            camera?.setMode(DJICameraMode(rawValue: 0x00)!, withCompletion: {
                error in
                    if error != nil {
                        var displayMessage: String? = nil
                        displayMessage = "Set DJICameraModeShootPhoto Failed
\\(error.debugDescription)"
                        self.debuggView.text += displayMessage! + "\\n"
                    }
                })
        }
        else if segmentControl!.selectedSegmentIndex == 1 {
            //Record video
            camera?.setMode(DJICameraMode(rawValue: 0x01)!, withCompletion: {
                error in
                    if error != nil {
                        var showMessage: String? = nil
                        showMessage = "Set DJICameraModeShootVideo Failed
\\(error.debugDescription)"
                        self.debuggView.text += showMessage! + "\\n"
                    }
                })
        }
    }
}

```

Σχήμα 5.3.1.16: Συνάρτηση για την αλλαγή κατάστασης από λήψη φωτογραφιών σε καταγραφή βίντεο και το αντίστροφο

5.3.2 Εφαρμογή του αλγορίθμου YOLO για την ανίχνευση των οχημάτων

Καταρχάς για να μπορέσουμε να εφαρμόσουμε τον αλγόριθμο προϋποτίθεται να είναι εγκατεστημένα τα εξής πακέτα (packages): OpenCV και NumPy. Το OpenCV (Open Computer Vision-Υπολογιστή Όραση Ανοικτού Κώδικα) αποτελεί μια βιβλιοθήκη λειτουργιών προγραμματισμού που στοχεύει κυρίως στην υπολογιστική όραση σε πραγματικό χρόνο. Αρχικά αναπτύχθηκε από την Intel, στη συνέχεια υποστηρίχθηκε από το Willow Garage και στη συνέχεια το Itseez. Η βιβλιοθήκη είναι πολλαπλών πλατφορμών και είναι δωρεάν για χρήση υπό την άδεια ανοικτού κώδικα. Το OpenCV απογείωσε την έρευνα στον τομέα της θέασης παρέχοντας όχι μόνο ανοιχτό αλλά και βελτιστοποιημένο κώδικα για βασική υποδομή υπολογιστικής όρασης. Αυτό επιτεύχθηκε διότι το OpenCV παρέχει μια κοινή υποδομή την οποία μπορούν να αναπτύξουν οι προγραμματιστές, με αποτέλεσμα ο κώδικας να είναι πιο εύκολα αναγνώσιμος και μεταβιβάσιμος. Η έκδοση που χρησιμοποιήσαμε είναι η *OpenCV 3.4.2+*. Το NumPy αποτελεί μία από της βιβλιοθήκες της γλώσσας προγραμματισμού Python, το οποίο χρησιμοποιείται κυρίως για χειρισμό μεγάλων και πολύπλοκων πινάκων, αλλά και για άλλες μαθηματικές πράξεις, συμπεριλαμβανομένων πολλών συναρτήσεων.

Επιπλέον χρησιμοποιήσαμε το COCO dataset, για το οποίο έχει γίνει λόγος παραπάνω, μιας και περιέχει την κλάση των αυτοκινήτων που μας ενδιαφέρει και είναι ιδιαιτέρως εύχρηστο. Κατόπιν φαίνονται κάποια screenshot του κώδικα και τα απαραίτητα σχόλια.

```

# construct the argument parse and parse the arguments
ap = argparse.ArgumentParser()
ap.add_argument("-i", "--image", required=True,
                help="path to input image")
ap.add_argument("-y", "--yolo", required=True,
                help="base path to YOLO directory")
ap.add_argument("-c", "--confidence", type=float, default=0.5,
                help="minimum probability to filter weak detections")
ap.add_argument("-t", "--threshold", type=float, default=0.3,
                help="threshold when applying non-maxima suppression")
args = vars(ap.parse_args())

# load the COCO class labels our YOLO model was trained on
labelsPath = os.path.sep.join([args["yolo"], "coco.names"])
LABELS = open(labelsPath).read().strip().split("\n")

# initialize a list of colors to represent each possible class label
np.random.seed(42)
COLORS = np.random.randint(0, 255, size=(len(LABELS), 3),
                            dtype="uint8")

# derive the paths to the YOLO weights and model configuration
weightsPath = os.path.sep.join([args["yolo"], "yolov3.weights"])
configPath = os.path.sep.join([args["yolo"], "yolov3.cfg"])

```

Σχήμα 5.3.2.1: Εισαγωγή της εικόνας, των labels του COCO και των αντίστοιχων weights

```

# load our YOLO object detector trained on COCO dataset (80 classes)
print("[INFO] loading YOLO from disk...")
net = cv2.dnn.readNetFromDarknet(configPath, weightsPath)

# load our input image and grab its spatial dimensions
image = cv2.imread(args["image"])
(H, W) = image.shape[:2]

# determine only the *output* layer names that we need from YOLO
ln = net.getLayerNames()
ln = [ln[i] - 1 for i in net.getUnconnectedOutLayers()]

|# construct a blob from the input image and then perform a forward
|# pass of the YOLO object detector, giving us our bounding boxes and
|# associated probabilities
blob = cv2.dnn.blobFromImage(image, 1 / 255.0, (416, 416),
                              swapRB=True, crop=False)

net.setInput(blob)
start = time.time()
layerOutputs = net.forward(ln)
end = time.time()

# show timing information on YOLO
print("[INFO] YOLO took {:.6f} seconds".format(end - start))

|# initialize our lists of detected bounding boxes, confidences, and
|# class IDs, respectively
boxes = []
confidences = []
classIDs = []

```

Σχήμα 5.3.2.2: Εισαγωγή των κλάσεων του dataset και δημιουργία blob για τα bounding boxes

```

# ensure at least one detection exists
if len(idxs) > 0:
    # loop over the indexes we are keeping
    for i in idxs.flatten():
        # extract the bounding box coordinates
        (x, y) = (boxes[i][0], boxes[i][1])
        (w, h) = (boxes[i][2], boxes[i][3])

        # draw a bounding box rectangle and label on the image
        color = [int(c) for c in COLORS[classIDs[i]]]
        cv2.rectangle(image, (x, y), (x + w, y + h), color, 2)

# show the output image
cv2.imshow("Image", image)
cv2.waitKey(0)

```

Σχήμα 5.3.2.3: Δημιουργία του bounding box για κάθε όχημα που εντοπίζεται

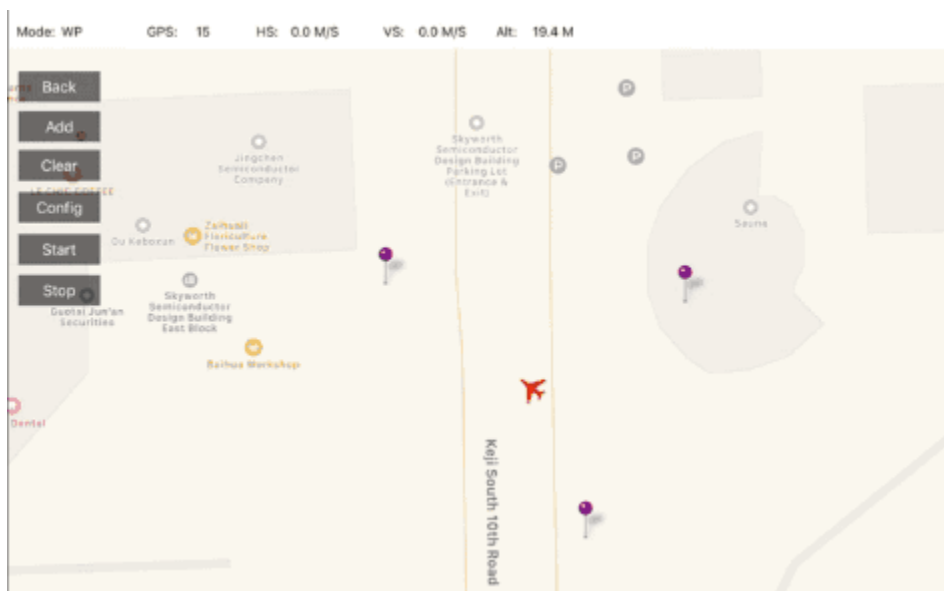
5.4 Γραφικό περιβάλλον

Σε αντίθετη περίπτωση δεν πραγματοποιείται η εκτέλεση της εφαρμογής *DJI GoTo* και παράλληλα εμφανίζεται κατάλληλο μήνυμα στον χρήστη για έλεγχο της σύνδεσης. Όταν η εγγραφή πραγματοποιηθεί εμφανίζεται αντίστοιχο μήνυμα και αναμένεται η σύνδεση του μη επανδρωμένου αεροσκάφους.

Κατά την εκκίνηση της εφαρμογής *DJI GoTo*, ο χρήστης αρχικά βρίσκεται στην πρώτη οθόνη της εφαρμογής, στην οποία εμφανίζεται σχετικό μήνυμα όταν επιτευχθεί σύνδεση με κάποιο μοντέλο μη επανδρωμένου αεροσκάφους της DJI. Σε περίπτωση επιτυχούς σύνδεσης, εμφανίζεται επίσης το μοντέλο του αεροσκάφους που ανιχνεύεται από την εφαρμογή. Πέρα από την περίπτωση να δημιουργηθεί κάποια δυσλειτουργία κατά την σύνδεση, μπορεί να αντιμετωπιστεί πρόβλημα κατά την εκκίνηση της εφαρμογής για διάφορους λόγους. Σε κάθε περίπτωση, ο χρήστης ενημερώνεται με σχετικά μηνύματα που τον καθοδηγούν προς ενέργειες με σκοπό την άρση των περιορισμών. Όταν το όνομα του μοντέλου κατά την σύνδεση είναι αυτό που αναμένεται, ο χρήστης με το πάτημα του κουμπιού *Proceed* μπορεί να μεταβεί στην δεύτερη οθόνη της εφαρμογής.

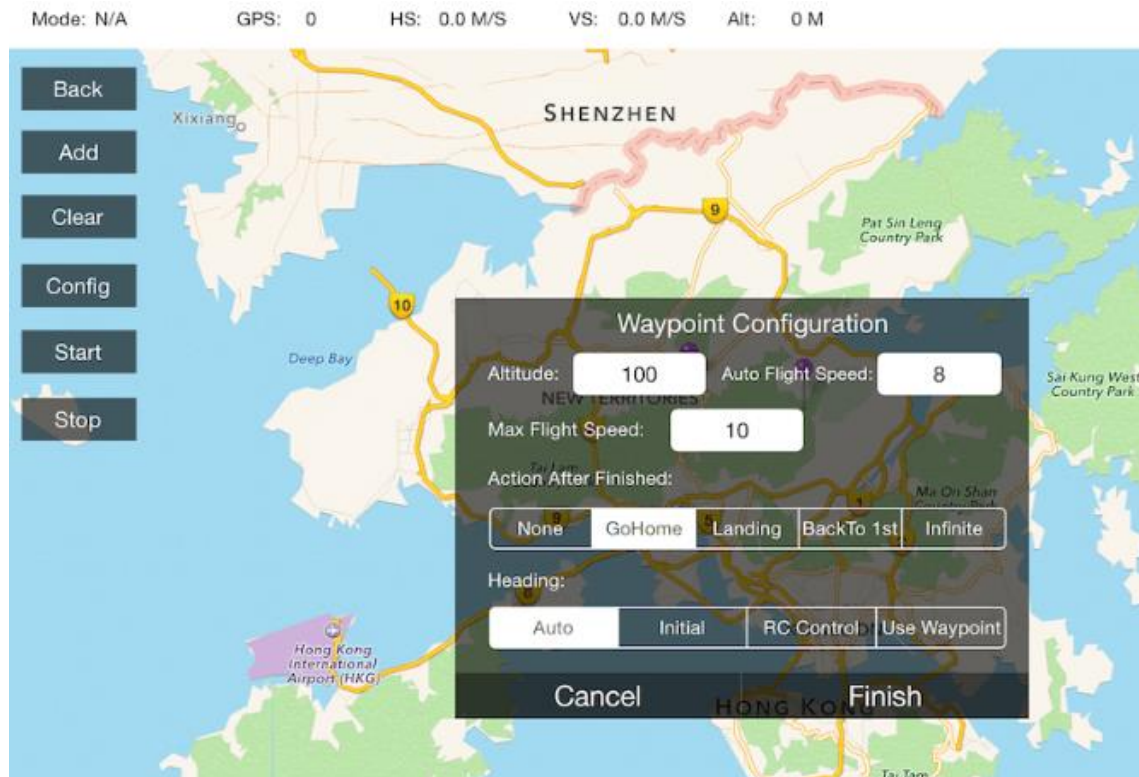
Στην δεύτερη, και κύρια, οθόνη εμφανίζεται η θέση όπου βρίσκεται το μη επανδρωμένο. Συγκεκριμένα, η εφαρμογή εστιάζει στην θέση του αεροσκάφους, η οποία απεικονίζεται με ένα μικρό κόκκινο αεροσκάφος. Επιπλέον, υπάρχουν κάποια κουμπιά τα οποία επιτελούν συγκεκριμένες λειτουργίες. Το πρώτο κουμπί είναι το *Edit*. Ο χρήστης, πατώντας το συγκεκριμένο κουμπί, μπορεί να προσθέσει σημεία στον χάρτη, τα οποία θα αποτελέσουν τα σημεία που συνθέτουν την διαδρομή που θα ακολουθήσει το μη επανδρωμένο αεροσκάφος. Τα σημεία αυτά απεικονίζονται σαν μωβ πινέζες στον χάρτη. Για να μπορεί να ξεκινήσει η πτήση για μία επιλεγμένη διαδρομή, θα πρέπει να σημεία στον χάρτη να είναι τουλάχιστον τρία. Σε διαφορετική περίπτωση εμφανίζεται κατάλληλο μήνυμα στον χρήστη στο λευκό περιθώριο στο κάτω μέρος της εφαρμογής, τον χώρο για την επισήμανση των λαθών (debugging area). Μόλις πατηθεί το

κουμπί *Edit*, μετονομάζεται σε *Back* και παράλληλα εμφανίζονται άλλα πέντε κουμπιά τα οποία είχαν οριστεί να είναι κρυμμένα από το αρχικό View. Σε περίπτωση λανθασμένης τοποθέτησης σημείου από τον χρήστη, μπορεί να ενεργοποιηθεί το κουμπί *Clear*, προκειμένου να σβηστούν τα σημεία από τον χάρτη. Αν επιθυμεί να προσθέσει σημεία, αρκεί να πατήσει το κουμπί *Add* και να κάνει tap στα επιθυμητά σημεία στον χάρτη. Στην συνέχεια παρατίθεται ένα στιγμιότυπο οθόνης αυτού του σταδίου της εφαρμογής, όπου φαίνονται η θέση του μη επανδρωμένου αεροσκάφους στον χάρτη καθώς και οι θέσεις τριών σημείων που έχουν προστεθεί.



Σχήμα 5.4.1: Η κύρια οθόνη της εφαρμογής *DJI GoTo*

Όπως φαίνεται και στο παραπάνω στιγμιότυπο, ένα από τα κουμπιά είναι το *Config*. Όταν ο χρήστης πατάει το συγκεκριμένο κουμπί, εμφανίζεται ένα παράθυρο στην κύρια οθόνη με τίτλο *Waypoint Configuration*. Με την ενεργοποίηση του κουμπιού, δίνεται η δυνατότητα στον χρήστη να παρέμβει σε μερικά από τα στοιχεία που αφορούν την πτήση του αεροσκάφους. Συγκεκριμένα, θα μπορούσε να αλλάξει το υψόμετρο στο οποίο είναι ρυθμισμένο να πετά το μη επανδρωμένο εναέριο όχημα. Η προεπιλεγμένη τιμή είναι τα εκατό μέτρα. Επιπλέον, μία μεταβλητή στην οποία θα μπορούσε να παρέμβει ο χρήστης και να την προσαρμόσει στις ανάγκες του, είναι η ταχύτητα και ειδικότερα η ταχύτητα του αεροσκάφους όταν βρίσκεται σε κατάσταση αυτόματης πλοήγησης και η μέγιστη ταχύτητα κατά την πτήση. Οι default τιμές είναι οχτώ και δέκα μέτρα το δευτερόλεπτο αντίστοιχα. Η προεπιλογή των τιμών αυτών είναι χρήσιμη γιατί με αυτόν τον τρόπο δεν προϋποτίθεται από τον χρήστη να πληκτρολογήσει τις τιμές όταν ξεκινά η λειτουργία της εφαρμογής. Εκτός των παραπάνω, περιέχονται δύο μπάρες με επιλογές στο κουτί, η πρώτη αφορά την ενέργεια του αεροσκάφους μετά την περάτωση της αποστολής του και η δεύτερη τον προσανατολισμό του αεροσκάφους. Σαν ενέργεια αποπεράτωσης έχει επιλεγεί η επιστροφή στο σημείο που έχει οριστεί ως *Home* και ο προσανατολισμός γίνεται αυτόματα. Σε οποιαδήποτε στιγμή μπορεί να ενεργοποιηθεί το κουμπί *Cancel*. Σε αυτή την περίπτωση, οποιαδήποτε αλλαγή δεν θα ληφθεί υπόψιν. Αντίθετα, με το πάτημα του κουμπιού *Finish*, ο χρήστης αποθηκεύει τις αλλαγές και εξέρχεται από το *Waypoint Configuration*.



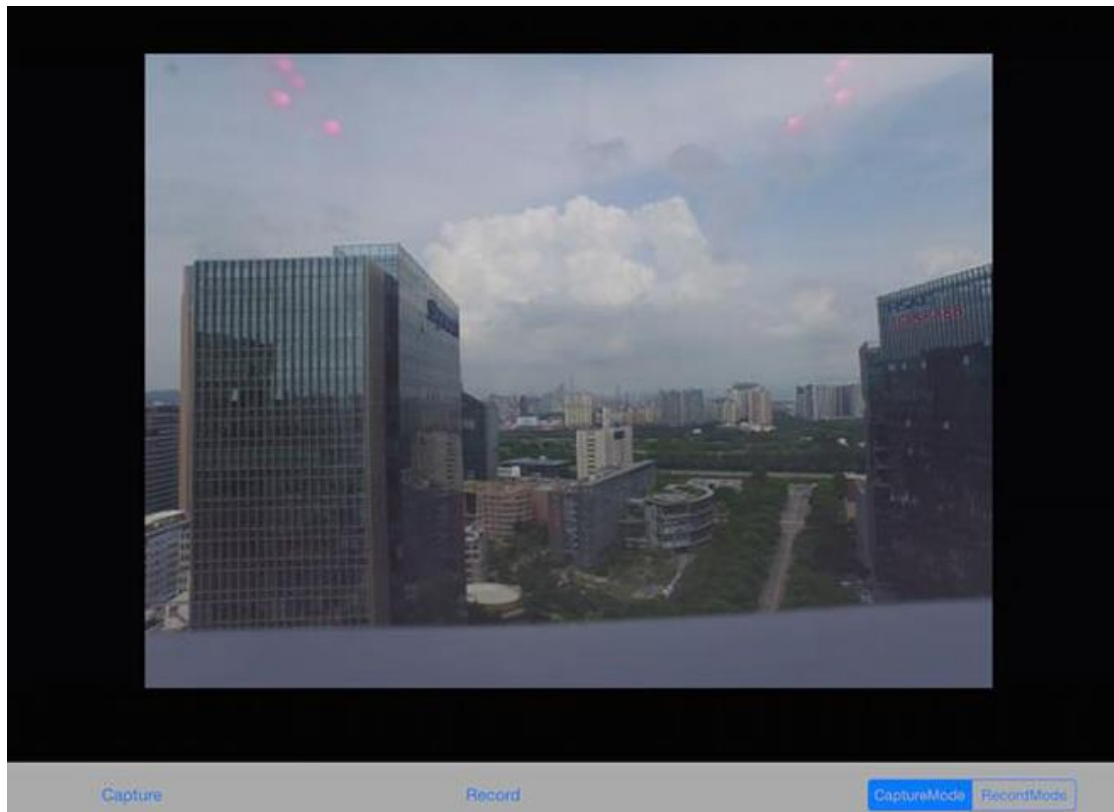
Σχήμα 5.4.2: Εμφάνιση του παραθύρου *Waypoint Configuration* μετά το πάτημα του κουμπιού *Config*

Αφού έχει ολοκληρωθεί η διαδικασία επεξεργασίας των στοιχείων που αφορούν την πτήση, ο χρήστης μπορεί να επιλέξει την έναρξη της πτήσης με το πάτημα του κουμπιού *Start*. Μετά την έναρξη της πτήσης, σε οποιαδήποτε στιγμή, ο χρήστης μπορεί να επιλέξει την διακοπή της αποστολής, πατώντας το κουμπί *Stop*. Όπως προαναφέρθηκε, με την λήξη της αποστολής, το αεροσκάφος θα επιστρέψει στην θέση που έχουμε ορίσει ως θέση *Home*, η οποία συνήθως είναι η θέση αφετηρίας της πτήσης.

Εκτός από τα κουμπιά αυτά, όπως φαίνεται και από το παραπάνω στιγμιότυπο οθόνης, υπάρχει επίσης μία μπάρα στο πάνω μέρος, όπου μετρούνται κάποια μεγέθη τα οποία μεταβάλλονται σε πραγματικό χρόνο κατά την διάρκεια της πτήσης. Αριστερότερα συναντάται το μέγεθος *Mode*, το οποίο δείχνει τον τύπο της πτήσης. Έπειτα υπάρχει το μέγεθος *GPS*, οποίο ορίζει την παρούσα δορυφορική μέτρηση μέσω GPS. Ακόμα, φαίνονται η κάθετη και η οριζόντια ταχύτητα του αεροσκάφους καθώς και το υψόμετρο στο οποίο βρίσκεται κάθε στιγμή. Οι δείκτες αυτοί προορίζονται για την βελτίωση της εμπειρίας του χρήστη, μιας και η προσφορά αντίστοιχων πληροφοριών καθ' όλη την διάρκεια της πτήσης, κρίνεται αναγκαία.

Πέρα από τις δύο πρώτες οθόνες, ο χρήστης έχει την δυνατότητα μετάβασης σε μία τρίτη οθόνη, όπου μπορεί με το πάτημα του αντίστοιχου κουμπιού να τραβάει φωτογραφίες ή βίντεο κατά βούληση. Ειδικότερα, ο χρήστης μπορεί να έχει οπτική επαφή με ό,τι καταγράφει η κάμερα που είναι ενσωματωμένη στο μη επανδρωμένο αεροσκάφος σε πρώτο πρόσωπο (*first person view*) και να παίρνει φωτογραφίες ή και βίντεο όποτε κρίνει σκόπιμο. Η λήψη φωτογραφιών πραγματοποιείται μέσω της ενεργοποίησης του κουμπιού *Capture*, ενώ η λήψη βίντεο, μέσω του κουμπιού *Record*. Πριν πατήσει κάποιο από τα δύο κουμπιά, ο χρήστης θα πρέπει να έχει διαλέξει

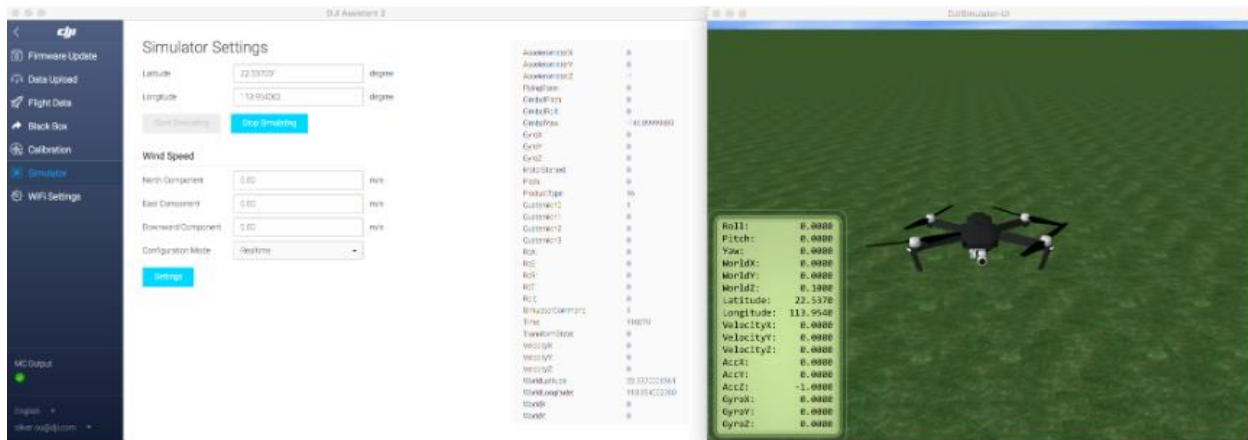
σε ποια κατάσταση επιθυμεί να βρίσκεται το drone, κατάσταση λήψης φωτογραφιών ή κατάσταση λήψης βίντεο. Αυτό μπορεί να ρυθμιστεί από την μπάρα που προσφέρει τις δύο παραπάνω επιλογές στο κάτω δεξιά μέρος της οθόνης.



Σχήμα 5.4.3: Στιγμιότυπο οθόνης για χρήση της ενσωματωμένης κάμερας του drone

Έκτος από την πολιτική που ακολουθείται στην συγκεκριμένη εφαρμογή, δηλαδή η συχνότητα λήψης φωτογραφιών και καταγραφής βίντεο να καθορίζεται από τον χρήστη, μία άλλη δυνατότητα θα ήταν η εφαρμογή να σχεδιάζόταν ώστε να παίρνει έναν συγκεκριμένο αριθμό φωτογραφιών όταν το drone βρίσκεται στο κάθε σημείο από το οποίο του έχει υποδεχθεί να περάσει, καθώς και λίγο πριν ή μετά από αυτό. Ωστόσο μιας και η εφαρμογή μας παρέχει live feed, κρίθηκε ευκολότερο οι φωτογραφίες και τα βίντεο να λαμβάνονται χειροκίνητα. Η λειτουργία του βίντεο δεν χρησιμοποιήθηκε ιδιαίτερα στην συγκεκριμένη εργασία μιας και η ανίχνευση αντικειμένων είναι ευκολότερη όταν υπάρχει σταθερή εικόνα. Ωστόσο η εφαρμογή θα μπορούσε να χρησιμοποιηθεί ανεξάρτητα για πληθώρα διαφορετικών έργων, τα οποία πιθανών να επωφελούνται από την χρήση της λειτουργίας λήψης βίντεο.

Σε περίπτωση που κάποιος δεν διαθέτει drone, η DJI παρέχει την δυνατότητα για προσομοίωση της εμπειρίας της πτήσης μέσω του DJI Assistant 2 Simulator. Χάρη στον προσομοιωτή, ο χρήστης μπορεί να παρατηρεί την κίνηση του αεροσκάφους καθώς και την μεταβολή των μεγεθών αναφορικά με την κλίση, τον προσανατολισμό, τις συντεταγμένες και άλλα.



Σχήμα 5.4.4: Ο DJI Assistant 2 Simulator προσομοιώνει την πτήση του αεροσκάφους

Κεφάλαιο 6

Μετρήσεις και αποτελέσματα

Το μη επανδρωμένο αεροσκάφος που χρησιμοποιήθηκε για την μελέτη των αποτελεσμάτων όπως έχει ήδη αναφερθεί, είναι το Mavic Pro Platinum και διαθέτει κάμερα ακρίβειας 12 megapixels, ενώ παρέχει την δυνατότητα λήψης βίντεο 4K/30fps. Δηλαδή προσφέρεται ανάλυση 4000 pixels πλάτος επί 3000 pixels ύψος. Κατά την διάρκεια των μετρήσεων, ανιχνεύτηκαν τα αυτοκίνητα τα οποία μπορούν να εντοπιστούν στην εκάστοτε εικόνα κάθε φορά. Οι εικόνες που λήφθηκαν από την ενσωματωμένη κάμερα του αεροσκάφους, είναι από τον χώρο του Εθνικού Μετσόβιου Πολυτεχνείου και πιο συγκεκριμένα από τους κατ' εξοχήν χώρους στάθμευσης, όπου μπορούν να παρατηρηθούν αρκετά αυτοκίνητα είτε σε κίνηση, είτε σε ακινησία. Κρίθηκε σκόπιμο το υψόμετρο που πετά το αεροσκάφος να διατηρηθεί πάνω από τα δώδεκα μέτρα, έτσι ώστε να αποφευχθεί ο κίνδυνος συντριβής του drone με κάποιο από τα αρκετά εμπόδια που θα συναντούσε σε χαμηλότερο υψόμετρο πτήσης.

6.1 Αποτελέσματα αλγορίθμου στον εντοπισμό αυτοκινήτων

6.1.1 Αποδοτικότητα αλγορίθμου

Εκτελέσαμε τον αλγόριθμο εντοπισμού αντικειμένων, ο οποίος είναι ο YOLOv3 τρεις φορές για κάθε εικόνα. Συγκεκριμένα έγινε έλεγχος της αποτελεσματικότητας και της αξιοπιστίας του αλγορίθμου για τα εξής μεγέθη εικόνας: 512x512, 768x768 και 1024x1024. Προκειμένου να κρίνουμε την αξιοπιστία του αλγορίθμου, πρέπει να ληφθεί υπόψιν πόσα αντικείμενα ανιχνεύτηκαν ορθά από το σύνολο των αντικειμένων που βρίσκονται στις εικόνες. Επίσης ένα ακόμα κριτήριο αποτελεσματικότητας του αλγορίθμου είναι φυσικά και ο χρόνος εκτέλεσης. Όπως έχει προαναφερθεί, η επιλογή του συγκεκριμένου αλγορίθμου ανίχνευσης αντικειμένων έγινε κυρίως με βάση το γεγονός πως ο YOLO αποτελεί έναν από τους γρηγορότερους αλγορίθμους

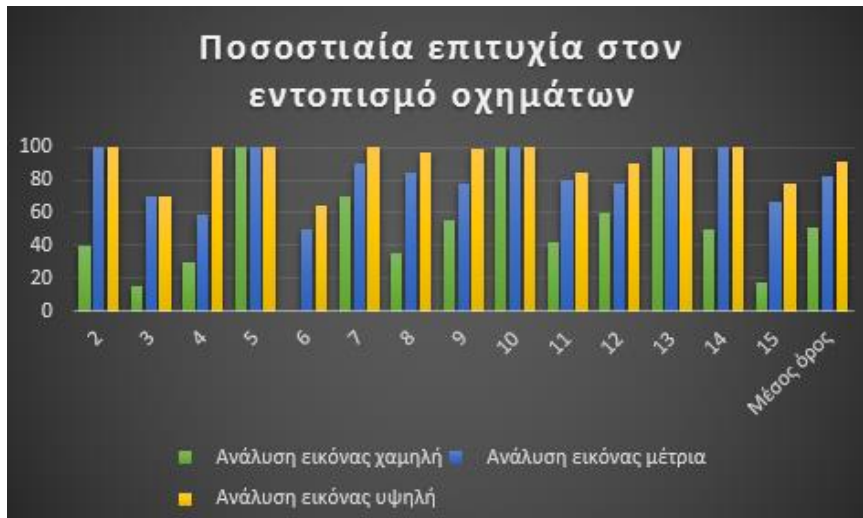
εντοπισμού. Η ταχύτητα εκτέλεσης είναι ιδιαίτερος σημαντική στο συγκεκριμένο εγχείρημα μιας και επιζητούμε ανιχνεύσεις οχημάτων σε πραγματικό χρόνο. Με αυτόν τον τρόπο η εφαρμογή θα μπορούσε να πάρει πολλές ενδιαφέρουσες επεκτάσεις, όπως η ανίχνευση της κοντινότερης θέσης στάθμευσης για τους οδηγούς αυτοκινήτων. Μερικές από τις δυνατές επεκτατικές λύσεις που θα μπορούσε να προσφέρει η εφαρμογή, αναφέρονται στο επόμενο κεφάλαιο.

Στη συνέχεια παρουσιάζονται με τη μορφή πινάκων και διαγραμμάτων τα αποτελέσματα που προέκυψαν από την χρήση του αλγορίθμου σε εικόνες που λήφθηκαν από την ενσωματωμένη κάμερα του drone στην περιοχή κοντά στην πύλη Κατεχάκη στο ΕΜΠ. Παρατηρείται πως μεταβάλλοντας το μέγεθος της εικόνας που προσφέρεται προς σάρωση με σκοπό την ανίχνευση αυτοκινήτων, αλλάζουν τα εξαγόμενα αποτελέσματα. Συγκεκριμένα, όπως είναι άλλωστε αναμενόμενο, υπάρχει μεγαλύτερη ακρίβεια στις προβλέψεις όταν το μέγεθος της εικόνας ήταν μεγαλύτερο. Ευτυχώς οι διαφορές δεν ήταν ουσιαστικά πολύ μεγάλες, γεγονός που επιβεβαιώνει την καλή λειτουργία του αλγορίθμου. Η ακρίβειά του αλγορίθμου διατηρείται σε υψηλά επίπεδα, παρόλο που σε κάποιες περιπτώσεις συναντώνται κάποια εμπόδια τα οποία ενδεχομένως ευθύνονται για τυχόν λανθασμένες ανιχνεύσεις. Μία ακόμα πιθανή αιτία λανθασμένων προβλέψεων είναι και η μεγάλη απόσταση από την οποία λήφθηκαν οι φωτογραφίες προκειμένου να αποφευχθεί η σύγκρουση του αεροσκάφους με κάποιο δέντρο ή κτήριο.

Ποσοστά ακρίβειας σχετικά με τον επιτυχή εντοπισμό οχημάτων

Εικόνα	Ανάλυση εικόνας		
	χαμηλή	μέτρια	υψηλή
1			
2	40	100	100
3	15	70	70
4	30	59	100
5	100	100	100
6	0	50	65
7	70	90	100
8	35	85	97
9	55	78	99
10	100	100	100
11	42	80	85
12	60	78	90
13	100	100	100
14	50	100	100
15	17	67	78
Μέσος όρος	51	82.64286	91.71429

Σχήμα 6.1.1.1: Ποσοστά επιτυχίας του YOLO στην εύρεση οχημάτων για τα διαφορετικά μεγέθη εικόνας



Σχήμα 6.1.1.2: Διάγραμμα για τα ποσοστά επιτυχίας εντοπισμού οχημάτων

Σε γενικές γραμμές, το μοντέλο ανταποκρίνεται με μεγάλη ακρίβεια στις περιπτώσεις όπου το μέγεθος της εικόνας είναι μεγάλο, δηλαδή για εικόνες 768x768 και 1024x1024, μιας και εντοπίζονται σχεδόν όλα τα αυτοκίνητα. Οι δύο αυτές διαφορετικές εκτελέσεις του αλγορίθμου παρουσιάζουν πολύ παρόμοια αποτελέσματα. Η εκτέλεση του αλγορίθμου στις εικόνες με την μεγαλύτερη ανάλυση υπερτερεί σε περιπτώσεις όπου τα αντικείμενα βρίσκονται είτε πολύ μακριά, είτε κοντά τους εμπόδια. Σε περίπτωση που το αντικείμενο προς ανίχνευση ήταν μικρότερου όγκου, όπως για παράδειγμα στην προσπάθεια ανίχνευσης ανθρώπων, ενδεχομένως τα ποσοστά επιτυχίας να ήταν χαμηλότερα. Αυτό συμβαίνει διότι τα εμπόδια που συναντώνται στην περιοχή ενδιαφέροντος έχουν πολύ περισσότερο αντίκτυπο σε αντικείμενα μικρότερης κλίμακας. Ακόμα και οι σκιές των δέντρων, των κτιρίων ή των αυτοκινήτων μπορούν να αλλοιώσουν τα αποτελέσματα στην περίπτωση αυτή. Μία σημαντική παρατήρηση είναι πως σε καμία εικόνα δεν υπήρξε λανθασμένος εντοπισμός. Όσες φορές εντοπίστηκε κάτι ήταν σωστό.

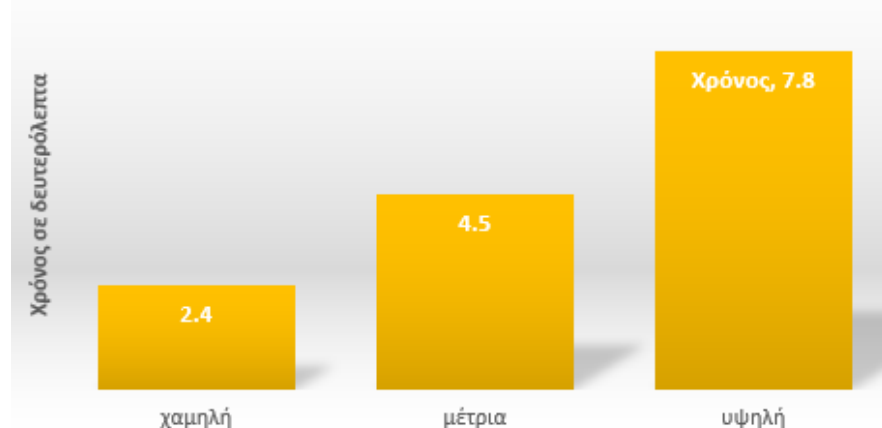
Παρατηρείται ωστόσο πως η απόδοση του αλγορίθμου στις εικόνες μικρότερου μεγέθους δεν είναι καλή, με ποσοστό επιτυχίας να είναι κατά μέσο όρο 51%, γεγονός που σημαίνει πως σε αυτό το μέγεθος εικόνας δεν ανιχνεύονται περίπου τα μισά οχήματα. Το αποτέλεσμα αυτό είναι αρκετά λογικό μιας και η εικόνα είναι πολύ μικρότερη από ότι η αρχική. Μία αρκετά ενδιαφέρουσα παρατήρηση ωστόσο είναι πως ακόμα και στην χαμηλότερη ανάλυση, αν υπάρχουν ιδανικές συνθήκες, οι προβλέψεις είναι επαρκείς. Με τον όρο ιδανικές συνθήκες εννοείται πως η απόσταση μεταξύ των οχημάτων είναι αρκετά μεγάλη και δεν παρεμβάλλονται σχεδόν καθόλου εμπόδια μεταξύ τους. Αυτό σημαίνει πως σε συνθήκες όπως αυτές που επικρατούν συνήθως στους αυτοκινητοδρόμους, εκτός από τις περιπτώσεις υπερβολικής κίνησης ή ατυχημάτων, η μικρότερη ανάλυση θα ήταν αρκετή για την αξιόπιστη λειτουργία της εφαρμογής.

Συμπερασματικά, σύμφωνα με τα αποτελέσματα αποδοτικότητας του αλγορίθμου σε εικόνες διαφορετικών αναλύσεων, οι εικόνες με την υψηλότερη ανάλυση, παρέχουν ακριβέστερα αποτελέσματα. Ωστόσο η διαφορά δεν είναι μεγάλη σε σχέση με τις εικόνες με την δεύτερη κατά σειρά υψηλότερη ανάλυση. Αρκεί να εξετάσουμε αν αξίζει χρονικά η επεξεργασία των εικόνων με καλύτερη ανάλυση. Λογικά, υποθέτουμε πως στις περιπτώσεις όπου ο χρόνος εκτέλεσης αποτελεί το σημαντικότερο κριτήριο επιλογής αλγορίθμου, τότε ενδεχομένως η ανάλυση 1024x1024 να μην αποτελεί την ιδανική λύση. Παρακάτω γίνεται παρουσίαση των αποτελεσμάτων που αφορούν στον χρόνο εκτέλεσης του αλγορίθμου και για τις τρεις περιπτώσεις.

6.1.2 Χρόνος εκτέλεσης αλγορίθμου

Όπως προβλέψαμε, όσο μεγαλύτερη είναι η ανάλυση των εικόνων, τόσο περισσότερος είναι και ο χρόνος που χρειάζεται ο αλγόριθμος για την επεξεργασία τους. Πιο συγκεκριμένα, όταν αυξάνουμε την ανάλυση από τα 512x512 στα 1024x1024, ο χρόνος εκτέλεσης σχεδόν τετραπλασιάζεται και από τα 768x768 στα 1024x1024, διπλασιάζεται, γεγονός που δηλώνει πως η σχέση ανάλυσης εικόνας και χρόνου εκτέλεσης αλγορίθμου είναι κατά προσέγγιση γραμμική. Ποσοτικά το αποτέλεσμα αυτό φαίνεται στην συνέχεια.

Χρόνος εκτέλεσης αλγορίθμου

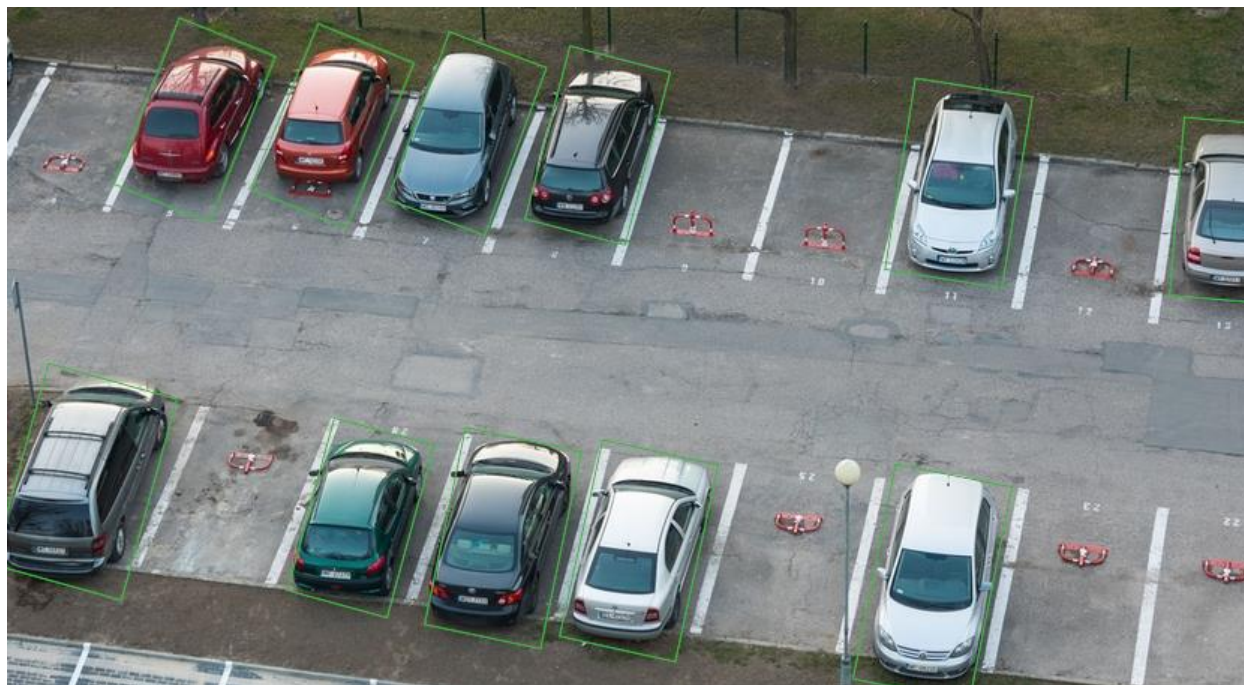


Σχήμα 6.1.2.1: Γράφημα για τον χρόνο εκτέλεσης του YOLO στις διαφορετικές αναλύσεις εικόνας

6.1.3 Σύγκριση λειτουργικότητας των διαφορετικών αναλύσεων εικόνων

Σε αυτή την υποενότητα γίνεται συγχώνευση των αποτελεσμάτων που έχουν σχολιαστεί στις δύο προηγούμενες ενότητες. Συγκεκριμένα, μελετάται η σχέση αποτελεσματικότητας και χρόνου εκτέλεσης για καθεμία από τις διαφορετικές αναλύσεις των εικόνων. Ο συνδυασμός αυτός συνθέτει την απόδοση του αλγορίθμου στις τρεις διαφορετικές περιπτώσεις. Όταν η ανάλυση αυξάνεται από 512x512 σε 768x768, παρατηρείται πως η αποτελεσματικότητα στην ανίχνευση αυξάνεται κατά περίπου 20%, ενώ ταυτόχρονα ο χρόνος εκτέλεσης του αλγορίθμου αυξάνεται σχεδόν κατά 100%. Σε αυτή την περίπτωση, παρόλο που ο χρόνος εκτέλεσης έχει μεγάλη διαφορά, κρίνεται πιο συνετό να επιλεγεί η μεγαλύτερη ανάλυση στις εικόνες, διότι διαφορετικά η αποτελεσματικότητα είναι πολύ χαμηλή σε σημείο που δεν θα μας συνέφερε να την θυσιάσουμε για να πετύχουμε γρηγορότερη ταχύτητα εκτέλεσης. Αντίθετα, περαιτέρω αύξηση σε 1024x1024 δεν θεωρείται ιδιαίτερα επικερδής, αφού ειδικά όταν πρόκειται για οχήματα το ποσοστό αποτελεσματικότητας δεν αλλάζει αισθητά, ενώ το τμήμα του χρόνου είναι πολύ μεγάλο. Επομένως καταλήγουμε στο συμπέρασμα πως η ιδανικότερη περίπτωση ανάλυσης εικόνας είναι το 768x768, μιας και προσφέρει τον καλύτερο συνδυασμό στην ακρίβεια των αποτελεσμάτων και τον χρόνο που απαιτείται για την πλήρη εκτέλεση του αλγορίθμου ανίχνευσης των αντικειμένων.

Στην συνέχεια παρουσιάζεται η έξοδος του αλγορίθμου όταν η ανάλυση ήταν 768x768. Όπως φαίνεται, ο αλγόριθμος έχει ανιχνεύσει επιτυχώς όλα τα αυτοκίνητα, ακόμα και το όχημα που δεν είναι εξ' ολοκλήρου μέσα στο επιθυμητό κάδρο.



Σχήμα 6.1.3.1: Η εικόνα που λαμβάνουμε σαν έξοδο του αλγορίθμου YOLO

Κεφάλαιο 7

Προτάσεις για περαιτέρω έρευνα

Στην παρούσα εργασία αναπτύχθηκε μία εφαρμογή για ανίχνευση αντικειμένων και πιο συγκεκριμένα, οχημάτων με χρήση υπολογιστικής όρασης με την βοήθεια της κάμερας ενός μη επανδρωμένου αεροσκάφους. Η εφαρμογή καινοτομεί στο γεγονός πως αν υπάρχει διαθέσιμος υπολογιστής στον οποίο μπορούμε να σταλούν οι εικόνες που λαμβάνονται από το μη επανδρωμένο εναέριο όχημα, μπορούμε σε real time να σκανάρουμε μια αρκετά μεγάλη περιοχή και να είμαστε σε θέση να απαντήσουμε ακριβώς πόσα αυτοκίνητα βρίσκονται στην επιθυμητή περιοχή. Μιας και οχήματα έχουν αρκετά μεγάλο όγκο, οι πιθανότητες λάθους στην ανίχνευση είναι αρκετά χαμηλές. Ωστόσο, όπως έχει αναφερθεί, ο αλγόριθμος που προτείνεται για την ανίχνευση αντικειμένων μπορεί να ανιχνεύσει και αντικείμενα πολύ μικρότερου όγκου. Παρόλα αυτά, όπως διαπιστώθηκε η σχέση όγκου αντικειμένου και αποτελεσματικότητας του αλγορίθμου είναι ανάλογη, με την δεύτερη να μειώνεται αισθητά όταν ο όγκος των αντικειμένων προς ανίχνευση μειώνεται. Αυτό αποτελεί ένα από τα προβλήματα που χρήζουν επίλυσης όσων αφορά την λύση που προτείνεται. Επομένως, υπάρχει περιθώριο βελτιώσεων της προτεινόμενης λύσης και παράλληλα θα μπορούσαν να υπάρχουν ερευνητικά περαιτέρω επεκτάσεις. Οι διαπιστώσεις αυτές παρουσιάζονται στην συνέχεια.

7.1 Επίλυση προβλημάτων και βελτιώσεις

Η παρούσα Διπλωματική Εργασία πραγματεύεται την ανίχνευση αντικειμένων και συγκεκριμένα τον εντοπισμό αυτοκινήτων μέσα από εικόνες που έχουν ληφθεί από μη επανδρωμένο εναέριο όχημα. Η ανίχνευση αντικειμένων μέσα από εικόνες και βίντεο χαίρει σημαντικής ανάπτυξης, μιας και αναμένεται να διευκολύνει τις ζωές των ανθρώπων σε πολυάριθμους τομείς. Ο εντοπισμός ανθρώπων και αυτοκινήτων είναι οι πλέον δημοφιλείς στον κλάδο του object detection. Παρόλα αυτά, οι αλγόριθμοι εντοπισμού που αναφέρθηκαν και παραπάνω δεν περιορίζονται στην ανίχνευση ατόμων και οχημάτων, καλύπτοντας έτσι ένα ευρύ φάσμα προβλημάτων.

Η λύση που προτείνεται στο πλαίσιο της προκείμενης Διπλωματικής Εργασίας, θα μπορούσε μελλοντικά να βελτιωθεί ώστε να πληροί αυστηρότερες προϋποθέσεις κυρίως αναφορικά με τον χρόνο εκτέλεσης του αλγορίθμου. Σε περίπτωση που δεν πρόκειται για εφαρμογή που προορίζεται για επαγγελματική χρήση, θα μπορούσε να γίνει εστίαση σε μικρότερο αριθμό κατηγοριών, ώστε να έχουμε πιο εξειδικευμένο στόχο, γεγονός που θα βελτιώνει σημαντικά τον χρόνο εκτέλεσης. Επιπλέον, η κάμερα του drone, θα μπορούσε να ενισχυθεί με την προσθήκη μίας θερμικής κάμερας. Αυτή η λύση χρησιμοποιείται ήδη σε πολλές εφαρμογές, μερικές από τις οποίες έχουν παρουσιαστεί πρωτίτερα, στο κεφάλαιο των σχετικών εφαρμογών, γεγονός το οποίο προσδίδει σημαντική μείωση στην ταχύτητα εκτέλεσης.

Η περαιτέρω αύξηση της ταχύτητας εκτέλεσης της εφαρμογής, θα μπορούσε να επέλθει με την πραγματοποίηση του αλγορίθμου εντοπισμού onboard, γεγονός που δεν θα προϋπέθετε την αποστολή των φωτογραφιών σε κάποιο υπολογιστή. Εκτός από την αύξηση του ρυθμού επεξεργασίας, με αυτό τον τρόπο ελαχιστοποιούνται επίσης προβλήματα απώλειας δεδομένων που ενδέχεται να προκύψουν κατά την αποστολή των φωτογραφιών από το μη επανδρωμένο αεροσκάφος προς τον υπολογιστή. Ωστόσο, το τίμημα για την εξοικονόμηση χρόνου, είναι η ανάγκη προσθήκης υλικού, γεγονός το οποίο αυξάνει και το κόστος του εγχειρήματος. Συγκεκριμένα, για να επιτευχθεί η onboard επεξεργασία των δεδομένων, χρειάζεται η τοποθέτηση κάποιου μικροϋπολογιστή όπως ένα raspberry ή για ακόμα ταχύτερα αποτελέσματα κάποια κάρτα γραφικών στο drone. Πέρα από την χρηματική επιβάρυνση, η λύση αυτή στερείται αυτονομίας αφού προκειμένου να λειτουργήσουν οι μικροεπεξεργαστές απαιτούν τροφοδοσία.

Μία ακόμη βελτίωση της εφαρμογής θα μπορούσε να είναι η μετατροπή της ώστε να είναι συμβατή με λογισμικό Android, το οποίο θα δίνει την δυνατότητα σύνδεσης με τους χάρτες της Google. Το παραπάνω εγχείρημα δεν κρίνεται ιδιαίτερος απαιτητικό, μιας και η DJI προσφέρει σαφείς και λεπτομερείς οδηγίες για την δημιουργία αντίστοιχης εφαρμογής σε γλώσσα προγραμματισμού Java, η οποία χρησιμοποιείται κατά κόρον για την ανάπτυξη εφαρμογών που προορίζονται για Android.

7.2 Επεκτάσεις

Εκτός από τις βελτιωτικές ενέργειες που αφορούν κυρίως την βελτίωση της ταχύτητας εκτέλεσης της εφαρμογής, μερικές διακριτικές αλλαγές θα μπορούσαν να κάνουν την εφαρμογή επωφελή σε πολλούς κλάδους.

7.2.1 Εύρεση θέσης στάθμευσης

Καταρχάς θα ήταν πολύ ενδιαφέρον να μπορούσε να γίνει ανίχνευση των κατειλημμένων θέσεων στάθμευσης. Υπάρχουν ερευνητικές προτάσεις για εύρεση θέσης στάθμευσης, όμως αυτές είτε χρησιμοποιούν τις κάμερες που μπορεί να βρίσκονται σε κάποιο χώρο, όπως για παράδειγμα το πάρκινγκ του αεροδρομίου ή κάποιος άλλος περιφραγμένος χώρος στάθμευσης. Ωστόσο δεν υπάρχουν προτάσεις που να χρησιμοποιούν τις εικόνες ή τα βίντεο που μπορούν να ληφθούν από την ενσωματωμένη κάμερα ενός μη επανδρωμένου αεροσκάφους.

Προκειμένου να λυθεί ένα σύνθετο πρόβλημα, είναι προτιμότερο να χωριστεί σε επιμέρους απλούστερα προβλήματα. Το ίδιο συμβαίνει και για τα προβλήματα μηχανικής μάθησης. Αφού το πρόβλημα έχει διασπαστεί σε λιγότερα περίπλοκα προβλήματα, χρησιμοποιώντας τα ως οδηγό και επιστρατεύοντας διάφορα εργαλεία της μηχανικής μάθησης, η εύρεση λύσης είναι πολύ γρηγορότερη και με λιγότερες πιθανότητες λάθους. Έτσι το πρόβλημα της ανίχνευσης θέσης στάθμευσης, αφού έχουμε λάβει τις εικόνες από την κάμερα του μη επανδρωμένου, μπορεί να χωριστεί σε δύο διασυνδεδεμένα απλούστερα, όσον αφορά την επεξεργασία των εικόνων. Στο πρώτο μέρος ανιχνεύονται όλες οι θέσεις στάθμευσης που είναι ορατές στην εκάστοτε εικόνα και στο δεύτερο προσδιορίζεται η κατάσταση της κάθε θέσης, κατειλημμένη ή διαθέσιμη.

7.2.1.1. Ανίχνευση θέσεων στάθμευσης

Όταν πρόκειται για μέρος το οποίο διαθέτει διαχωριστικές γραμμές μεταξύ των θέσεων στάθμευσης, αυτό που πρέπει να κάνουμε προκειμένου να καταχωρήσουμε κάθε θέση ξεχωριστά είναι να ανιχνεύσουμε τις γραμμές που τις διαχωρίζουν. Η βασική ιδέα είναι πως όλες οι διαχωριστικές γραμμές των θέσεων στάθμευσης είναι παράλληλες μεταξύ τους και οι θέσεις στάθμευσης που βρίσκονται στην ίδια στήλη είναι περίπου εξίσου διαχωρισμένες. Αρχικά μπορεί να χρησιμοποιηθεί η ανίχνευση Canny edge, η οποία θα έχει ως αποτέλεσμα τον εντοπισμό των γωνιών στην εικόνα. Στην συνέχεια παρουσιάζεται η έξοδος του πρώτου σταδίου της διαδικασίας, έχοντας σαν είσοδο την εικόνα ενός χώρου στάθμευσης αεροδρομίου. [28]



Σχήμα 7.2.1.1.1: Η έξοδος μετά την εφαρμογή του αλγορίθμου ανίχνευσης Canny edge

Στο επόμενο στάδιο, επιστρατεύοντας την τεχνική hough line transform στην εικόνα που έχει ληφθεί ως έξοδος από το προηγούμενο στάδιο, λαμβάνουμε την αρχική εικόνα με σχεδιασμένες όλες τις γραμμές που ανιχνεύτηκαν. Έπειτα απομονώνουμε τις γραμμές που ταιριάζουν στον προσανατολισμό των διαχωριστικών γραμμών που αναζητάμε. Για παράδειγμα μπορεί να χρειαστεί να απομονώσουμε όλες τις οριζόντιες γραμμές, όπως φαίνεται και στην εικόνα παρακάτω που αφορά αυτό το στάδιο της διαδικασίας.



Σχήμα 7.2.1.1.2: Η έξοδος μετά την εφαρμογή του αλγορίθμου hough line transform

Παρόλο που οι hough lines αναγνωρίζουν τις γραμμές στάθμευσης αρκετά καλά, η έξοδος δεν είναι καθαρή, μιας και πολλές γραμμές στάθμευσης ανιχνεύονται πολλές φορές, ενώ μερικές χάνονται. Προκειμένου το αποτέλεσμα να είναι πιο καθαρό, μπορούμε να χρησιμοποιήσουμε τις συντεταγμένες που επέστρεψε το hough line transform και ομαδοποιώντας τις συντεταγμένες για τον άξονα που μας ενδιαφέρει (για παράδειγμα τον άξονα x) προσδιορίζουμε τις κύριες λωρίδες στάθμευσης.



Σχήμα 7.2.1.1.3: Το αποτέλεσμα μετά την ομαδοποίηση των συντεταγμένων του άξονα x

Τώρα γνωρίζουμε πού βρίσκονται όλες οι λωρίδες στάθμευσης. Προκειμένου να εντοπίσουμε κάθε μεμονωμένη θέση στάθμευσης, υποθέτουμε πως είναι όλες του ίδιου μεγέθους. Έτσι, μπορούμε να εκχωρήσουμε σε κάθε θέση ένα αναγνωριστικό και να αποθηκεύσουμε τις συντεταγμένες του σε ένα λεξικό, ώστε να μπορούν να ανακτηθούν αργότερα.



Σχήμα 7.2.1.1.4: Το τελικό αποτέλεσμα της ανίχνευσης των θέσεων στάθμευσης

7.2.1.2. Προσδιορισμός κατειλημμένων θέσεων στάθμευσης

Προκειμένου να μπορούμε να αποφανθούμε αν μία θέση είναι κατειλημμένη ή όχι μπορούμε να διαλέξουμε έναν από τους παρακάτω τρόπους.

Χρήση του OpenCV ώστε να ελεγχθεί εάν το χρώμα των εικονοστοιχείων ενός σημείου ευθυγραμμίζεται με το χρώμα ενός κενού σημείου στάθμευσης. Πρόκειται για μια απλή προσέγγιση αλλά είναι επιρρεπής σε σφάλματα. Για παράδειγμα, η αλλαγή στο φωτισμό επηρεάζει το χρώμα μιας άδειας θέσης στάθμευσης, γεγονός που θα δυσκολέψει τη λειτουργία αυτής της λογικής μέσα στην ημέρα. Επίσης, είναι πιθανό να προκληθεί σύγχυση με τα αυτοκίνητα γκρι χρώματος.

Χρήση αλγορίθμων ανίχνευσης αντικειμένων για τον εντοπισμό όλων των οχημάτων και κατόπιν έλεγχος αν η θέση του αυτοκινήτου επικαλύπτεται με θέση στάθμευσης. Ο τρόπος αυτός, όπως έχει αναφερθεί και παραπάνω, μπορεί να αντιμετωπίζει δυσκολίες αν ο χώρος στάθμευσης είναι πολύ μεγάλος και κατ' επέκταση το μέγεθος των αντικειμένων προς ανίχνευση σχετικά μικρό.

Χρήση ενός CNN για να βρεθεί κάθε θέση στάθμευσης και να γίνει πρόβλεψη για το αν είναι άδεια ή όχι. Αυτή ενδεχομένως είναι και η καλύτερη λύση.



Σχήμα 7.2.1.2.1: Εύρεση κατειλημμένων θέσεων με χρήση CNN

Επιπλέον, μία ενδιαφέρουσα προέκταση της εφαρμογής θα μπορούσε να είναι ο υπολογισμός του χρόνου μετάβασης με το αυτοκίνητο προς τον επιθυμητό προορισμό σε συνδυασμό με τον υπολογισμό ενός εκτιμώμενου χρόνου για το παρκάρισμα. Αυτό θα ήταν ιδιαίτερα ωφέλιμο, αφού δεν θα χρειαζόταν να μαντέψουμε πόσος χρόνος θα χρειαστεί για το παρκάρισμα. Στο ίδιο πλαίσιο, ο χρήστης θα μπορούσε να κάνει κράτηση μίας θέσης στάθμευσης. Αυτό προϋποθέτει την αρίθμηση και τον ξεκάθαρο διαχωρισμό των θέσεων στάθμευσης, το οποίο ομολογουμένως είναι ένα απαιτητικό έργο. Τέλος, μία ιδέα θα μπορούσε να είναι η ύπαρξη μίας εφαρμογής για την ανίχνευση θέσεων στάθμευσης αναπήρων. Με μία τέτοια εφαρμογή θα διευκολυνόταν ακόμα ο εντοπισμός των οχημάτων που δεν έχουν δικαιοδοσία να είναι σταθμευμένα σε αυτές τις θέσεις.

7.2.2 Ανάλυση κίνησης πελατών

Η εύρεση χώρου στάθμευσης, όπως οι περισσότεροι γνωρίζουμε, μπορεί να αποδειχθεί ιδιαίτερα δύσκολο και χρονοβόρο εγχείρημα, ειδικά σε μία μεγάλη πόλη. Οι περισσότεροι από εμάς έχουμε ονειρευτεί, μία συσκευή, όπως για παράδειγμα το κινητό μας τηλέφωνο να μπορούσε να εντοπίσει για εμάς την πλησιέστερη θέση πάρκινγκ, γλιτώνοντάς μας από πολύ κόπο και σίγουρα ενέργεια. Ωστόσο, η ανίχνευση των κατειλημμένων θέσεων στάθμευσης, θα μπορούσε να φανεί χρήσιμη και ακόμα και επικερδής, όχι μόνο στους ενδιαφερόμενους για την εύρεση θέσης παρκαρίσματος, αλλά και για πολλές επιχειρήσεις. Τα δεδομένα μετά από δημογραφική ανάλυση ή ανάλυση κίνησης πελατών θα μπορούσαν να προσφέρουν πληροφορίες, όπως ποιες είναι οι ώρες αιχμής σε κάθε περιοχή σύμφωνα με τον αριθμό παρκαρισμένων οχημάτων και πόσοι από τους πελάτες τους προσεγγίζουν τον χώρο με αυτοκίνητο.

Θα μπορούσε να γίνει ταξινόμηση των αυτοκινήτων τα οποία έχουν ανιχνευτεί ανάλογα με την μάρκα ή το μέγεθος ή ακόμα και το χρώμα. Σύμφωνα με αυτά τα δεδομένα θα μπορούσαν να εξαχθούν πληροφορίες για τους κατόχους των αυτοκινήτων όπως το εισόδημα, η οικογενειακή κατάσταση, ακόμα και οι πολιτικές προτιμήσεις. Επειδή είναι σημαντικό η διαδικασία αυτή να είναι ασφαλής και να μην παραβιάζει την ιδιωτικότητα των χρηστών που συμμετέχουν, θα μπορούσε όποιος ιδιοκτήτης αυτοκινήτου δεν επιθυμεί να συμμετέχει σε μία τέτοια έρευνα να φέρει κάποιο διακριτικό στην οροφή του αυτοκινήτου του, το οποίο θα ανιχνεύεται και θα έχει ως αποτέλεσμα το όχημα να αποκλείεται από την διαδικασία ταξινόμησης.

7.2.3 Ανάλυση κυκλοφορίας σε αυτοκινητόδρομους

Τα αυτοκίνητα είναι εύκολα ανιχνεύσιμα λόγω του όγκου που καταλαμβάνουν. Όσοι ζουν σε μεγαλουπόλεις και κινούνται με το αυτοκίνητο γνωρίζουν πως πολλές φορές η κίνηση στους δρόμους μπορεί να γίνει ανυπόφορη. Η εφαρμογή που προτείνεται στην παρούσα Διπλωματική Εργασία θα μπορούσε επίσης να χρησιμοποιηθεί ώστε να μετριέται σε πραγματικό χρόνο ο αριθμός των αυτοκινήτων στους δρόμους. Ενδεχομένως μία τέτοια εφαρμογή να μην χρειαζόταν την χρήση GPS. Θα μπορούσε να δημιουργηθεί εφαρμογή είτε κάποιο άλλου είδους πλατφόρμα, η οποία θα ενημερώνει τους χρήστες για τις ακριβείς συνθήκες της κυκλοφορίας στους δρόμους την ζητούμενη στιγμή. Επίσης θα είναι δυνατή η παρακολούθηση των ατυχημάτων και η άμεση ενημέρωση της τροχαίας, της οδικής βοήθειας ή ασθενοφόρου μέσω της εφαρμογής. Αυτό θα μπορούσε να γλιτώσει πολύτιμο χρόνο και ακόμα και ζωές σε περιπτώσεις που δεν υπάρχει κάποιος να καλέσει το ασθενοφόρο την φλέγουσα στιγμή.

7.2.4 Εντοπισμός επιζώντων

Όπως έχει αναφερθεί, η εφαρμογή θα μπορούσε να χρησιμοποιηθεί και για εντοπισμό ανθρώπων αρκετά αποτελεσματικά, μιας και ο αλγόριθμος ανίχνευσης αντικειμένων YOLO, σημειώνει υψηλές επιδόσεις σε αντικείμενα με όγκο όσο ένας άνθρωπος. Έτσι, η εφαρμογή θα μπορούσε να χρησιμοποιηθεί για εντοπισμό ανθρώπων σε περιπτώσεις πολέμων ή και φυσικών καταστροφών. Η έκταση στην οποία πιθανόν να βρίσκονται άνθρωποι είναι συχνά αποκεντρωμένη, ενώ το μέγεθος της είναι αρκετά μεγάλο. Ενδεχομένως το προσωπικό που ασχολείται με την έρευνα να μην είναι επαρκές, με αποτέλεσμα να απαιτείται αρκετός κρίσιμος χρόνος ώστε να διερευνηθούν όλα τα σημεία. Πιο συγκεκριμένα όταν πρέπει να επισκεφθούν μεγάλες μη κατοικήσιμες περιοχές, το μη επανδρωμένο αεροσκάφος μπορεί να τις σκανάρει πολύ

ευκολότερα και γρηγορότερα, καθιστώντας την έρευνα αρκετά αποτελεσματικότερη. Δε δαπανάται άσκοπα ανθρώπινο δυναμικό, αλλά ερευνούν μόνο μέρη που το μη επανδρωμένο επισήμανε ως σημεία ενδιαφέροντος.

Βιβλιογραφία

- [1] "Unnamed Aerial Vehicles – ScienceDirect." [Ηλεκτρονικό].
Available: <https://www.sciencedirect.com/topics/engineering/unmanned-aerial-vehicles>
- [2] Wikipedia, "Unnamed Aerial Vehicle." [Ηλεκτρονικό].
Available: https://en.wikipedia.org/wiki/Unmanned_aerial_vehicle
- [3] "Parking Lot Vehicle Detection Using Deep Learning - Medium." [Ηλεκτρονικό].
Available: <https://medium.com/geoai/parking-lot-vehicle-detection-using-deep-learning-49597917bc4a>
- [4] "Drones to deliver medicine and food? Drones for disaster relief? Why not? – Ideas.Ted." [Ηλεκτρονικό].
Available: <https://ideas.ted.com/6-ways-drones-can-be-used-for-good/>
- [5] "Drones For Good: Sea Cleaning Drones - Drone Universities." [Ηλεκτρονικό].
Available: <https://www.droneuniversities.com/drones/drones-for-good-sea-cleaning-drones/>
- [6] Gaszczak, Anna, Toby P. Breckon, and Jiwan Han. "Real-time people and vehicle detection from UAV imagery." Intelligent Robots and Computer Vision XXVIII: Algorithms and Techniques (2011): 78780B.
- [7] "Deep Learning Haar Cascade Explained - Will Berger." [Ηλεκτρονικό].
Available: <http://www.willberger.org/cascade-haar-explained/>
- [8] "A fast object detector based on high-order gradients and Gaussian process regression for UAV images - ResearchGate." [Ηλεκτρονικό].
Available: https://www.researchgate.net/publication/277076267_A_fast_object_detector_based_on_high-order_gradients_and_Gaussian_process_regression_for_UAV_images
- [9] "DroneSAR | Emergency Response specific software for DJI drones." [Ηλεκτρονικό].
Available: <https://www.dronesarpilot.com>
- [10] "What is machine learning? - Search Enterprise AI, TechTarget." [Ηλεκτρονικό].
Available: <https://searchenterpriseai.techtarget.com/definition/machine-learning-ML>
- [11] "Facial recognition - Search Enterprise AI, TechTarget." [Ηλεκτρονικό].
Available: <https://searchenterpriseai.techtarget.com/definition/facial-recognition>
- [12] Wikipedia, "Object Detection." [Ηλεκτρονικό].
Available: https://en.wikipedia.org/wiki/Object_detection
- [13] "Snagging Parking Spaces with Mask R-CNN and Python – Medium, Machine Learning." [Ηλεκτρονικό].
Available: <https://medium.com/@ageitgey/snagging-parking-spaces-with-mask-r-cnn-and-python-955f2231c400>

- [14] Zhong-Qiu Zhao, Shou-tao Xu, and Xindong Wu. "Object Detection with Deep Learning: A Review." IEEE Transactions on Neural Networks and Learning Systems (2019): 1-6
Available: <https://arxiv.org/pdf/1807.05511.pdf>
- [15] Uijlings, J. R. R. et al. "Selective Search for Object Recognition." International Journal of Computer Vision 104.2 (2013): 11-12
- [16] Girshick, Ross et al. "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation." 2014 IEEE Conference on Computer Vision and Pattern Recognition (2014): 7-12
- [17] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton "ImageNet Classification with Deep Convolutional Neural Networks", Published in NIPS (2012)
- [18] Uijlings, J. R. R. et al. "Selective Search for Object Recognition." International Journal of Computer Vision 104.2 (2013): 1-4
- [19] He Kaiming et al. "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition." Lecture Notes in Computer Science (2014): 2-4
- [20] Girshick, Ross. "Fast R-CNN." 2015 IEEE International Conference on Computer Vision (ICCV) (2015):1-4
- [21] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun, Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, NIPS'15 Proceedings (2015): 1-5
- [22] "Understand Single Shot MultiBox Detector (SSD) and Implement it in Pytorch - Medium." [Ηλεκτρονικό].
Available: <https://medium.com/@smallfishbigsea/understand-ssd-and-implement-your-own-caa3232cd6ad>
- [23] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi, et al. "You only look once: Unified, real-time object detection." Proceedings of the IEEE conference on computer vision and pattern recognition. (2016)
- [24] Joseph Redmon, Ali Farhadi, et al. "YOLO9000: Better, Faster, Stronger" (2016)
Available: <https://arxiv.org/pdf/1612.08242v1.pdf>
- [25] "Mavic Pro - DJI Developer." [Ηλεκτρονικό].
Available: <https://www.dji.com/gr/mavic>
- [26] Redmon Joseph and Ali Farhadi, University of Washington. "Yolov3: An incremental improvement." [Ηλεκτρονικό].
Available: <https://pjreddie.com/media/files/papers/YOLOv3.pdf>
- [27] "DJI Mobile SDK Documentation - DJI Developer." [Ηλεκτρονικό].
Available: <https://developer.dji.com/mobile-sdk/documentation/introduction/index.html>

[28] "Find where to park in real time using OpenCV and Tensorflow - Towards Data Science." [Ηλεκτρονικό].
Available:<https://towardsdatascience.com/find-where-to-park-in-real-time-using-opencv-and-tensorflow-4307a4c3da03>