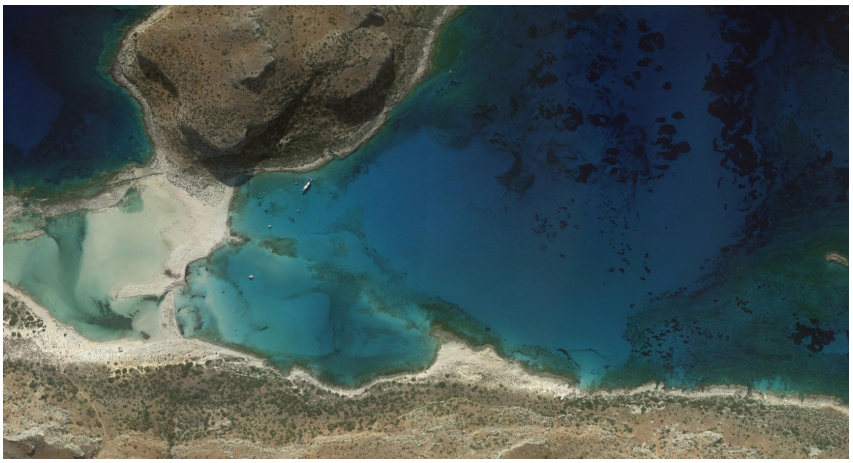
Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Τομέας Τεχνολογίας Πληροφορικής και
Υπολογιστών



Ανίχνευση Παραλιών σε Δορυφορικές Εικόνες με Χρήση Συνελικτικών Νευρωνικών Δικτύων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΕΛΕΝΗ ΜΑΘΙΟΥΛΑΚΗ

Επιβλέπων: Στέφανος Κόλλιας
Καθηγητής Ε.Μ.Π.

Συμμετοχή στην Επίβλεψη: Τζούβελη Παρασκευή
μέλος ΕΔΙΠ

Αθήνα, Νοέμβριος 2020

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Τομέας Τεχνολογίας Πληροφορικής και
Υπολογιστών

Ανίχνευση Παραλιών σε Δορυφορικές Εικόνες με Χρήση Συνελικτικών Νευρωνικών Δικτύων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΕΛΕΝΗ ΜΑΘΙΟΥΛΑΚΗ

Επιβλέπων: Στέφανος Κόλλιας
Καθηγητής Ε.Μ.Π.

Συμμετοχή στην Επίβλεψη: Τζούβελη Παρασκευή
μέλος ΕΔΙΠ

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 20η Νοεμβρίου 2020.

Στέφανος Κόλλιας	Γιώργος Στάμου	Ανδρέας-Γεώργιος Σταφυλοπάτης
Καθηγητής Ε.Μ.Π.	Αν. Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.

Αθήνα, Νοέμβριος 2020

.....

Ελένη Μαθιουλάκη

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

Περίληψη

Ο εντοπισμός αντικειμένων και η σημασιολογική κατάτμηση εικόνας αποτελούσαν ανέκαθεν δύο από τα δυσκολότερα προβλήματα της όρασης υπολογιστών. Με τη ραγδαία ανάπτυξη του τομέα της Τεχνητής Νοημοσύνης που έχει συντελεστεί τα τελευταία χρόνια, κυρίως λόγω της εμφάνισης των Βαθιών Νευρωνικών Δικτύων, έχουν προταθεί συστήματα που φαίνεται να δίνουν αποτελεσματικές και αποδοτικές λύσεις για προβλήματα αυτής της κατηγορίας με χρήση Συνελικτικών Νευρωνικών Δικτύων.

Στόχο της παρούσας εργασίας αποτελεί η υλοποίηση και η εκπαίδευση ενός τέτοιου συστήματος, το οποίο θα αναλύει δορυφορικές εικόνες με σκοπό να εντοπίσει τις περιοχές τους που αντιστοιχούν σε παραλίες. Συγκεκριμένα, το ζητούμενο του μοντέλου είναι ο εντοπισμός των ορίων της κάθε παραλίας σε επίπεδο εικονοστοιχείου, αποτελεί δηλαδή συνδυασμό των προβλημάτων του εντοπισμού αντικειμένων σε μία εικόνα και της εύρεσης των ορίων του με ακρίβεια. Στη συνέχεια, το σύστημα αυτό μπορεί να χρησιμοποιηθεί ως εργαλείο καταγραφής και παρακολούθησης των ελληνικών (και όχι μόνο) παραλιών ώστε να δημιουργηθεί μία πλήρης βάση δεδομένων και να διευκολυνθεί η μελέτη και η προστασία τους από καταστροφικά φυσικά φαινόμενα και υπέρμετρη τουριστική εκμετάλλευση.

Για την εκπαίδευση του μοντέλου απαραίτητη ήταν η δημιουργία ενός συνόλου δεδομένων αποτελούμενου από δορυφορικές εικόνες και τις αντίστοιχες ετικέτες. Οι εικόνες που χρησιμοποιούνται προέρχονται από τη δορυφορική αποστολή Sentinel-2 και συγκεντρώθηκαν μέσω του Google Earth Engine, ενώ οι τοποθεσίες των παραλιών αντλήθηκαν από το OpenStreetMap. Συνολικά συγκεντρώθηκαν πάνω από 3000 εικόνες της αποστολής Sentinel-2 που καλύπτουν όλη την ακτογραμμή της Ελλάδας, ενώ οι αντίστοιχες ετικέτες αφορούν πάνω από 5000 παραλίες.

Αφού δημιουργήθηκε το σύνολο δεδομένων, χρησιμοποιήθηκε για να εκπαιδεύσει δύο διαφορετικά μοντέλα, το Mask R-CNN, state-of-the-art δίκτυο κατάτμησης στιγμιοτύπων εικόνων, και το Rotated Mask R-CNN, μία τροποποίησή του που λαμβάνει υπόψη το χαρακτηριστικό σχήμα των παραλιών ώστε να βελτιώσει την απόδοση. Πράγματι, το Mask R-CNN πετυχαίνει Μέση Ακρίβεια (mean Average Precision) ίση με 43.5%, ενώ το Rotated Mask R-CNN 45.0%, τιμές αρκετά ικανοποιητικές δεδομένης της ποιότητας των εικόνων και της δυσκολίας του προβλήματος.

Λέξεις κλειδιά

Δορυφορικές εικόνες, Sentinel-2, Google Earth Engine, OpenStreetMap, Εντοπισμός Παραλίας, Τεχνητή Νοημοσύνη, Βαθιά Μηχανική Μάθηση, Συνελικτικά Νευρωνικά Δίκτυα, Ανίχνευση Αντικειμένων, Κατάτμηση Εικόνων, Mask R-CNN, Rotated Mask R-CNN

Abstract

Object detection and semantic segmentation in images have always been two of the most difficult problems in the field of computer vision. Given the rapid evolution of Artificial Intelligence (AI) in recent years, mostly due to the emergence of Deep Neural Networks, systems that seem to provide effective and efficient solutions for these kind of problems using Convolutional Neural Networks (CNNs) have been proposed.

The purpose of the thesis at hand is the implementation and training of such a system, that will analyse satellite imagery data in order to detect the areas that correspond to beaches. More specifically, the objective of this particular model is to identify the boundaries of the objects at the detailed pixel level, thus combining the object detection and the image segmentation problems. Afterwards, this system could be used as a monitoring and registering tool for greek beaches, so as to create a complete database and facilitate their study and protection from destructive physical phenomena and excessive touristic exploitation.

For the purpose of training the model, it was necessary to create a dataset consisting of satellite images and the corresponding labels. The imagery used is from the Sentinel-2 satellite mission and was collected using Google Earth Engine, while the locations of the beaches have been extracted from the OpenStreetMap database. In total, the dataset comprises more than 3000 Sentinel-2 images, covering the entirety of the greek coastline, while more than 5000 beaches are labeled.

After its creation, the dataset was used to train two different models, the Mask R-CNN, state-of-the-art network for instance segmentation and the Rotated Mask R-CNN, an altered version that takes the distinct shape of the beaches into consideration in order to improve accuracy. Indeed, the Mask R-CNN achieves a mean Average Precision equal to 43.5%, while the Rotated Mask R-CNN 45.0%. The obtained results were quite satisfactory, given the image quality, resolution and the problem difficulty.

Key words

Satellite Imagery, Sentinel-2, Google Earth Engine, OpenStreetMap, Beach Detection, Artificial Intelligence, Deep Learning, Convolutional Neural Networks, Object Detection, Image Segmentation, Mask R-CNN, Rotated Mask R-CNN

Ευχαριστίες

Θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα καθηγητή αυτής της διπλωματικής εργασίας, κ. Στέφανο Κόλλια, για τη δυνατότητα που μου έδωσε να ασχοληθώ με το συγκεκριμένο θέμα και την εμπιστοσύνη που μου έδειξε. Επίσης, ευχαριστώ ιδιαίτερα την Δρ. Παρασκευή Τζούβελη για την πολύτιμη καθοδήγηση και τις συμβουλές της, καθώς και την υποδειγματική συνεργασία που είχαμε καθ' όλη τη διάρκεια της εκπόνησης της διπλωματικής.

Τέλος, δε θα μπορούσα να μην ευχαριστήσω την οικογένειά μου και τους φίλους μου, που σε όλο το διάστημα των σπουδών μου με ανέχονται, με στηρίζουν και με καθοδηγούν, και χωρίς τους οποίους η προσπάθεια όλων αυτών των χρόνων δε θα ήταν εφικτή.

Ελένη Μαθιουλάκη,

Αθήνα, 20η Νοεμβρίου 2020

Περιεχόμενα

Κατάλογος πινάκων

Κατάλογος σχημάτων

Κεφάλαιο 1

Εισαγωγή

1.1 Κίνητρο

Σε ολόκληρο τον κόσμο, οι παράκτιες ζώνες και ειδικά οι παραλίες αποτελούν περιοχές ζωτικής σημασίας, όχι μόνο ως πολύτιμα οικοσυστήματα αλλά και ως παράγοντες οικονομικής και τουριστικής ανάπτυξης. Ανέκαθεν οι άνθρωποι φαίνεται να προτιμούσαν για την εγκατάστασή τους παράκτιες περιοχές, με πάνω από το 50% του παγκόσμιου πληθυσμού να ζει το 1996 σε απόσταση μικρότερη των 60 χιλιομέτρων από τη θάλασσα, και τον πληθυσμό στις περιοχές αυτές να αυξάνεται με γρήγορους ρυθμούς [1].

Οι κίνδυνοι που απειλούν τη διατήρηση και την ευημερία των περιοχών αυτών, ωστόσο, είναι πολυάριθμοι. Η επέμβαση του ανθρώπου, είτε μέσω έργων υποδομής (όπως για παράδειγμα εργασίες αποχέτευσης ή προσάμμωσης), είτε στο πλαίσιο της αστικοποίησης και την υπέρμετρης τουριστικοποίησης των γύρω περιοχών, έχει πολλές φορές οδηγήσει σε αλόγιστη εκμετάλλευση των φυσικών πόρων, αφήνοντας τα οικοσυστήματα των παραλιών και των παραθαλάσσιων περιοχών εκτεθειμένα. Ταυτόχρονα, η κλιματική αλλαγή και η συνεπακόλουθη άνοδος της στάθμης της θάλασσας [2] απειλεί σταδιακά τις παράκτιες περιοχές σε πολλές χώρες του κόσμου με εξαφάνιση.

Είναι καθοριστικής σημασίας, λοιπόν, η ανάπτυξη εργαλείων συστηματικής καταγραφής και ανάλυσης δεδομένων σχετικά με την ακριβή τοποθεσία, το μέγεθος και την κατάσταση της κάθε παραλίας, ώστε να είναι εφικτή η παρατήρηση των μεταβολών που λαμβάνουν χώρα και ο έγκαιρος σχεδιασμός των κατάλληλων μέτρων προστασίας.

Εξαιρετικά χρήσιμο εργαλείο σε αυτή την προσπάθεια αποτελεί η δορυφορική τηλεπισκόπηση, χάρη στην ανάπτυξη της οποίας υπάρχουν δημόσια διαθέσιμες πολυφασματικές απεικονίσεις υψηλής χρονικής και χωρικής ανάλυσης, με παγκόσμια κάλυψη. Η δωρεάν, δημόσια διάθεση των δορυφορικών εικόνων αυτών επέτρεψε την ανάπτυξη πλήθους εφαρμογών εντοπισμού αντικειμένων και κατάτμησης, αντίστοιχων με την παρούσα εργασία. Ενδεικτικά αναφέρονται εφαρμογές εντοπισμού δρόμων [3], κτηρίων [4], πλοίων [5], υδάτινων σωμάτων [6], πετρελαιοκηλίδων [7] ή και θερμοκηπίων [8].

1.2 Σκοπός της εργασίας

Ο στόχος της παρούσας εργασίας είναι η υλοποίηση και η εκπαίδευση ενός μοντέλου κατάτμησης εικόνων, το οποίο θα δέχεται σαν είσοδο μία δορυφορική εικόνα και θα εντοπίζει σε

αυτήν όλες τις περιοχές που αντιστοιχούν σε παραλίες, με ακρίβεια εικονοστοιχείου. Δευτερεύον στόχο της εργασίας αποτελεί η δημιουργία του συνόλου δεδομένων (dataset) που θα χρησιμοποιηθεί για την εκπαίδευση του μοντέλου.

1.3 Δομή της εργασίας

Η εργασία διαιρείται σε τρία μέρη.

Στο πρώτο μέρος καλύπτεται το Θεωρητικό Υπόβαθρο που είναι απαραίτητο για την κατανόηση της εργασίας. Συγκεκριμένα, στο Κεφάλαιο 2 πραγματοποιείται μία εισαγωγή στην Δορυφορική Τηλεπισκόπηση, η οποία περιλαμβάνει σύντομη αναφορά στις ιδιότητες του ηλεκτρομαγνητικού φάσματος, και αναλυτικές πληροφορίες σχετικά με τη δορυφορική αποστολή του Sentinel-2 και τη δομή των δορυφορικών εικόνων που θα χρησιμοποιηθούν στην εργασία. Στη συνέχεια, στο Κεφάλαιο 3 πραγματοποιείται μια εισαγωγή στον τρόπο με τον οποίο τα Νευρωνικά Δίκτυα χρησιμοποιούνται σε προβλήματα όπως αυτό που θα μας απασχολήσει. Συγκεκριμένα, ορίζονται έννοιες που θα είναι απαραίτητες στη συνέχεια της εργασίας, περιγράφονται τα βασικά προβλήματα της μηχανικής όρασης που θα μας απασχολήσουν και γίνεται συνοπτική αναφορά στις διάφορες αρχιτεκτονικές που προτείνονται.

Το δεύτερο μέρος αφορά την προετοιμασία των δεδομένων για την εκπαίδευση των μοντέλων. Στο Κεφάλαιο 4 περιγράφεται αναλυτικά η διαδικασία συγκέντρωσης, οργάνωσης και προεπεξεργασίας των δεδομένων που θα χρησιμοποιηθούν.

Το τρίτο μέρος της εργασίας αφορά την πειραματική διαδικασία. Στο Κεφάλαιο 5 περιγράφεται η μεθοδολογία που θα ακολουθηθεί, και συγκεκριμένα αναλύεται η δομή και η λειτουργία των δύο μοντέλων που θα χρησιμοποιηθούν, καθώς και των τμημάτων τους. Στο Κεφάλαιο 6 περιγράφεται η πειραματική διάταξη και παρουσιάζονται τα αποτελέσματα της εκπαίδευσης και της αξιολόγησης κάθε μοντέλου. Τέλος, στο Κεφάλαιο 7 πραγματοποιείται η σύνοψη της εργασίας και η εξαγωγή των τελικών συμπερασμάτων.

Μέρος Ι

Θεωρητικό Υπόβαθρο

Κεφάλαιο 2

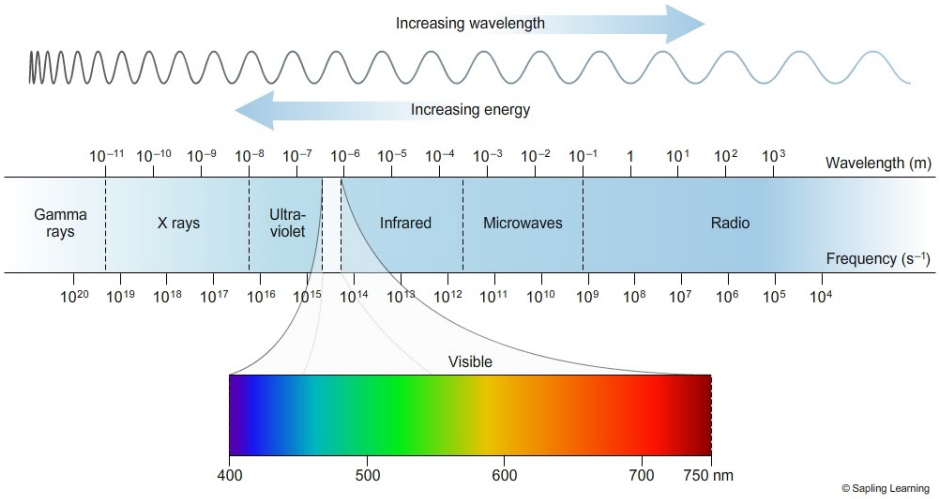
Δορυφορική Τηλεπισκόπηση

2.1 Ηλεκτρομαγνητικό Φάσμα

Πριν να προχωρήσουμε στη μελέτη της δομής και των ιδιοτήτων των δορυφορικών εικόνων, είναι απαραίτητη η αναφορά στο ηλεκτρομαγνητικό φάσμα και τις περιοχές στις οποίες διαιρείται.

Ηλεκτρομαγνητικά ονομάζονται τα κύματα τα οποία διαδίδονται σε κάποιο μέσο (ατμόσφαιρα, νερό, υλικά σώματα) λόγω της συγχρονισμένης ταλάντωσης ενός ηλεκτρικού και ενός μαγνητικού πεδίου, σε επίπεδα κάθετα μεταξύ τους και κάθετα προς την διεύθυνση διάδοσης. Τα ηλεκτρομαγνητικά κύματα διαδίδονται στο κενό (και κατά προσέγγιση στην ατμόσφαιρα) με την ταχύτητα του φωτός ($c = 299.792.458$ m/s), και χαρακτηρίζονται από την συχνότητα και το μήκος κύματός τους, το γινόμενο των οποίων ισούται με την ταχύτητα c.

$$c = \lambda f$$



Σχήμα 2.1: Το ηλεκτρομαγνητικό φάσμα [1]

Ως ηλεκτρομαγνητικό φάσμα ορίζεται η ταξινόμηση της ηλεκτρομαγνητικής ακτινοβολίας σύμφωνα με τη συχνότητα, το μήκος κύματος ή την ενέργεια της. Το ηλεκτρομαγνητικό φάσμα περιέχει μεγάλο εύρος διαφορετικών μηκών κύματος, και έχει διαιρεθεί στις ζώνες που φαίνονται παρακάτω. Στο πλαίσιο της συγκεκριμένης εργασίας, σημαντικότερο θεωρείται το τμήμα του

[1] https://sites.google.com/site/chempendix/em-spectrum

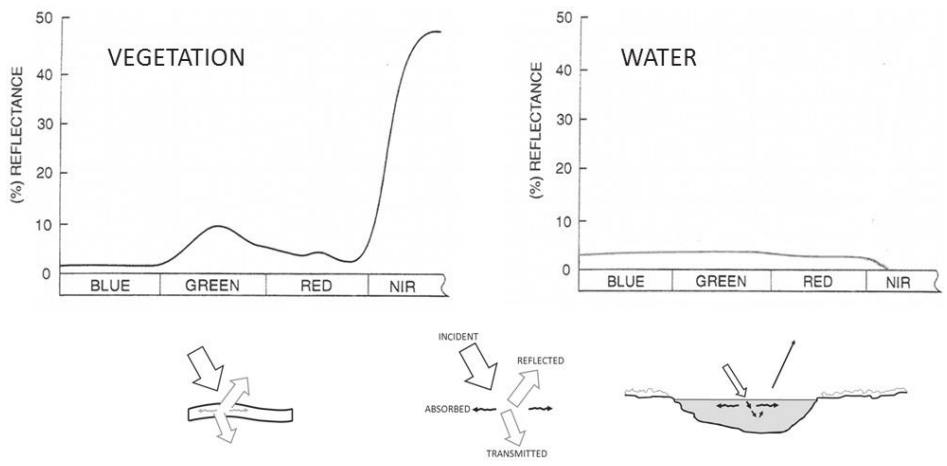
φάσματος που δημιουργείται από την ηλιακή ακτινοβολία, και συγκεκριμένα οι περιοχές του υπεριώδους, του ορατού και του υπέρυθρου.

Περιοχή Φάσματος	**Μήκος Κύματος**
Ραδιοκύματα	> 0.3 m
Μικροκύματα	1 mm - 0.3 m
Υπέρυθρη Ακτινοβολία	700 nm - 1 mm
Ορατή Ακτινοβολία	400 nm - 700 nm
Υπεριώδης Ακτινοβολία	10 nm - 400 nm
Ακτίνες X	0.01 nm - 10 nm
Ακτίνες γ	< 0.01 nm

Πίνακας 2.1: Περιοχές Ηλεκτρομαγνητικού Φάσματος [2]

2.2 Φασματική Υπογραφή

Μία από τις χαρακτηριστικές ιδιότητες της ύλης είναι η απορρόφηση και η ανάκλαση της προσπίπτουσας ηλεκτρομαγνητικής ακτινοβολίας. Η σύσταση καθώς και τα φυσικά χαρακτηριστικά του κάθε υλικού επηρεάζουν το ποσοστό της ηλεκτρομαγνητικής ακτινοβολίας που απορροφάται σε κάθε μήκος κύματος. Κατά συνέπεια, η κατανομή της ανακλώμενης ακτινοβολίας από ένα αντικείμενο σε συνάρτηση του μήκους κύματος, που ονομάζεται φασματική απόκριση, μπορεί να θεωρηθεί ως "φασματική υπογραφή" του υλικού, η οποία μπορεί να χρησιμοποιηθεί για την αναγνώρισή του και τη μελέτη των χαρακτηριστικών του. [9]



Σχήμα 2.2: Διαφορά μεταξύ φασματικών υπογραφών βλάστησης και νερού [9]

Ενδεικτικά, στο σχήμα φαίνεται η διαφορά μεταξύ της φασματικής υπογραφής κάποιου είδους βλάστησης και του νερού. Αμέσως γίνεται αντιληπτό ότι οι δύο καμπύλες είναι εντελώς διαφορετικές, τόσο ως προς το πλάτος όσο και ως προς τη μορφή τους. Το ίδιο ισχύει, σε μικρότερο

[2] `https://www.esa.int/Science_Exploration/Space_Science/Integral/The_electromagnetic_spectrum`

βαθμό, και για υλικά πιο παρόμοια μεταξύ τους, όπως για παράδειγμα διαφορετικά είδη βλάστησης ή πετρωμάτων. Συμπεραίνουμε, λοιπόν, ότι η φασματική υπογραφή των υλικών μπορεί να χρησιμοποιηθεί για την ταξινόμηση και τον χαρακτηρισμό τους, και κατά συνέπεια και για τον εντοπισμό συγκεκριμένων περιοχών, που αποτελεί το αντικείμενο αυτής της διπλωματικής.

2.3 Δορυφορική Τηλεπισκόπηση (Remote Sensing)

Η τηλεπισκόπηση (remote sensing), σύμφωνα με έναν από τους ευρύτερους ορισμούς της, είναι η συγκέντρωση φυσικών δεδομένων για ένα αντικείμενο χωρίς την άμεση επαφή με αυτό [10]. Ετυμολογικά, άλλωστε, προέρχεται από το αρχαίο επίρρημα "τήλε" (από απόσταση) και το ρήμα "επισκοπώ" (εξετάζω οπτικώς).Στο πλαίσιο των περισσότερων σύγχρονων εφαρμογών όμως ορίζεται ως η εξαγωγή πληροφοριών σχετικά με τις χερσαίες και υδάτινες επιφάνειες της Γης βάσει της αλληλεπίδρασης των υλικών που βρίσκονται επάνω σε αυτή με ηλεκτρομαγνητική ακτινοβολία από μία ή περισσότερες περιοχές του ηλεκτρομαγνητικού φάσματος [9].

Ιστορικά, οι πρώτες εφαρμογές της τηλεπισκόπησης ήταν κυρίως συνδεδεμένες με τη χαρτογραφία και τη στρατιωτική αναγνώριση. Γρήγορα όμως έγιναν αντιληπτές οι δυνατότητες εφαρμογής της σε τομείς όπως η μετεωρολογία, η μελέτη του περιβάλλοντος και των φυσικών φαινομένων, αλλά και η παρατήρηση των ανθρωπογενών δραστηριοτήτων και του τρόπου με τον οποίο επιδρούν στο περιβάλλον [9]. Η επιτήρηση και η πρόβλεψη φυσικών καταστροφών, η πρόβλεψη του καιρού, η παρακολούθηση των μεταβολών στον αστικό ιστό και την αγροτική δραστηριότητα του ανθρώπου με στόχο τη βιώσιμη ανάπτυξη αποτελούν λίγα μόνο παραδείγματα εφαρμογών στις οποίες η συμβολή της τηλεπισκόπησης είναι καίρια. Είναι σαφής λοιπόν η ανάγκη για συνολική και συστηματική τηλεπισκοπική παρατήρηση της Γης.

Καθοριστικό ρόλο στην ανάπτυξη της τηλεπισκόπησης διαδραμάτισε προφανώς η ραγδαία βελτίωση των δυνατοτήτων των δορυφορικών συστημάτων καταγραφής και ανάλυσης δεδομένων, η οποία οδήγησε στην πλήρη επικράτηση της δορυφορικής τηλεπισκόπησης για την παρατήρηση της Γης. Την εκτόξευση του πρώτου δορυφόρου, Sputnik-1 το 1957 ακολούθησαν εκατοντάδες δορυφορικές αποστολές, με αποτέλεσμα σήμερα να υπάρχουν σε τροχιά γύρω από τη Γη πάνω από 200 δορυφόροι γεωπαρατήρησης [11]. Ενδεικτικά αναφέρονται οι δορυφορικές αποστολές SPOT και Pleiades του Γαλλικού Εθνικού Κέντρου Διαστημικών Ερευνών (Centre National d'Etudes Spatiales), το πρόγραμμα Landsat της Αμερικανικής Εθνικής Υπηρεσίας Αεροναυπηγικής και Διαστήματος (National Aeronautics and Space Administration - NASA) και το πρόγραμμα Copernicus του Ευρωπαϊκού Οργανισμού Διαστήματος (European Space Agency - ESA).

2.4 Πρόγραμμα Copernicus

Το πρόγραμμα Copernicus αποτελεί μία πρωτοβουλία της Ευρωπαϊκής Επιτροπής (European Commission) σε συνεργασία με τον Ευρωπαϊκό Οργανισμό Διαστήματος (European Space Agency - ESA). Ζητούμενο του προγράμματος είναι να δημιουργηθεί ένα ενοποιημένο σύστημα μέσω του οποίου θα επιτυγχάνεται η συγκέντρωση μεγάλου όγκου δεδομένων γεωπαρατήρησης, και η διοχέτευσή τους σε ένα ευρύ πεδίο εφαρμογών, όπως η παρακολούθηση των κλιματικών

μεταβολών, η αειφόρος ανάπτυξη, η διαχείριση των αστικών περιοχών και των φυσικών πόρων, ο τοπικός και περιφερειακός σχεδιασμός και η εξυπηρέτηση ανθρωπιστικών αναγκών. Στα πλαίσια του προγράμματος, δεδομένα γεωπαρατήρησης συγκεντρώνονται με παγκόσμια κάλυψη και συνεχή ενημέρωση, και στη συνέχεια επεξεργάζονται και αναλύονται από τις υπηρεσίες του προγράμματος οι οποίες αφορούν 6 βασικές θεματικές ενότητες: τα ατμοσφαιρικά δεδομένα, το θαλάσσιο περιβάλλον, την επιφάνεια της Γης, την κλιματική αλλαγή, την ασφάλεια και τη διαχείριση έκτακτων καταστάσεων. [12]

Για τις ανάγκες του Copernicus, ο Ευρωπαϊκός Διαστημικός Οργανισμός δημιούργησε μία νέα, εξειδικευμένη οικογένεια δορυφόρων που ονομάζονται Sentinel. [13]

- Η αποστολή **Sentinel-1** χρησιμοποιεί προηγμένα όργανα ραντάρ για να παρέχει, ανεξαρτήτως καιρικών συνθηκών, ημερήσιες και νυκτερινές εικόνες τόσο των χερσαίων όσο και των θαλάσσιων περιοχών. Αποτελείται από δύο δορυφόρους, τον Sentinel-1A και τον Sentinel-1B, οι οποίοι εκτοξεύθηκαν στις 3 Απριλίου 2014 και στις 25 Απριλίου 2016 αντίστοιχα.
- Η αποστολή **Sentinel-2** παρέχει υψηλής χωρικής ανάλυσης πολυφασματικές απεικονίσεις για την παρακολούθηση και τη μελέτη της κάλυψης του εδάφους, των υδάτινων δικτύων και της βλάστησης. Αποτελείται από δύο δορυφόρους, τον Sentinel-2A και τον Sentinel-2B, οι οποίοι εκτοξεύθηκαν στις 22 Ιουνίου 2015 και στις 7 Μαρτίου 2017 αντίστοιχα.
- Η αποστολή **Sentinel-3** παρέχει υψηλής ακρίβειας και αξιοπιστίας δεδομένα σχετικά με την τοπογραφία της επιφάνειας της θάλασσας, τις επιφανειακές θερμοκρασίες της γης, το χρωματισμό των ωκεανών και της ξηράς. Αποτελείται από δύο δορυφόρους, τον Sentinel-3A και τον Sentinel-3B, οι οποίοι εκτοξεύθηκαν στις 16 Φεβρουαρίου 2016 και στις 25 Απριλίου 2018 αντίστοιχα.
- Οι αποστολές **Sentinel-4**, **Sentinel-5**, **Sentinel-5 Precursor** είναι σχεδιασμένες για την μελέτη της σύνθεσης της ατμόσφαιρας, και συγκεκριμένα την παρακολούθηση σε παγκόσμιο επίπεδο της ποιότητας του αέρα, της παρουσίας του όζοντος και της ηλιακής ακτινοβολίας, με υψηλή χωρική και χρονική ανάλυση. Η εκτόξευση των Sentinel-4 και Sentinel-5 είναι προγραμματισμένη για το 2023 και το 2021 αντίστοιχα, ενώ ο Sentinel-5P εκτοξεύθηκε στις 13 Οκτωβρίου 2017, ως πρόδρομος του Sentinel-5.
- Η αποστολή **Sentinel-6**, τέλος, παρέχει υψηλής ακρίβειας υψομετρικά δεδομένα για τη μέτρηση της παγκόσμιας στάθμης της θάλασσας, με στόχο κυρίως τη συλλογή και ανάλυση πληροφοριών σχετικά με τις κλιματικές μεταβολές, τα θαλάσσια ρεύματα και το ύψος των κυμάτων. Η αποστολή θα αποτελείται από δύο δορυφόρους, τον Sentinel-6 Michael Freilich, ο οποίος εκτοξεύθηκε στις 21 Νοεμβρίου 2020 και τον Sentinel-6B, ο οποίος θα εκτοξευθεί το 2025.

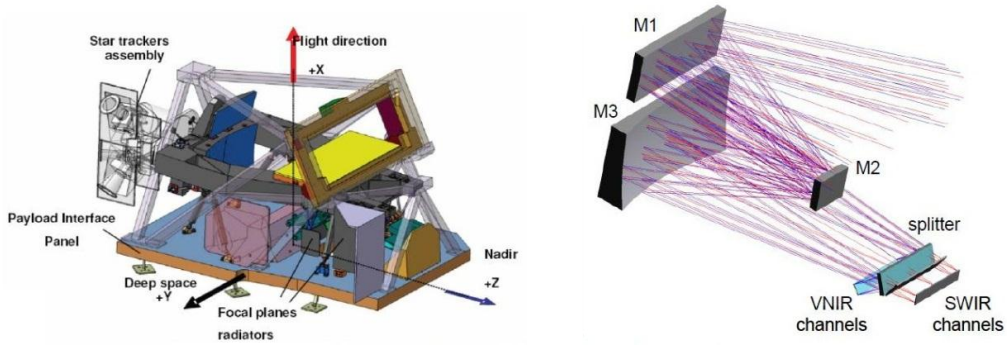
2.5 Sentinel-2

Καθώς η παρούσα εργασία σχετίζεται με την κατάτμηση των δορυφορικών εικόνων με βάση τη μορφολογία του εδάφους, επιλέχθηκαν ως πλέον κατάλληλα τα δεδομένα της αποστολής Sentinel-2 [14].

Η αποστολή Sentinel-2, όπως προαναφέρθηκε, αποτελείται από δύο δορυφόρους, οι οποίοι

κινούνται σε μέσο ύψος 786 km από την επιφάνεια της Γης με γωνιακή απόκλιση 180° μεταξύ τους. Καλύπτει την περιοχή μεταξύ −56° και 84° γεωγραφικού πλάτους. Με τη χρήση των δύο δορυφόρων επιτυγχάνεται η μείωση του χρόνου επαναδιέλευσης στο μισό, από 10 σε 5 ημέρες (στον Ισημερινό, υπό συνθήκες χωρίς νέφη). [15]

Ο κάθε ένας από τους δορυφόρους Sentinel-2 είναι εξοπλισμένος με υψηλής χωρικής ικανότητας πολυφασματικό σαρωτή MSI (MultiSpectral Instrument) με εύρος πεδίου (FOV) 290 km. Πρόκειται για παθητικού τύπου σύστημα το οποίο λειτουργεί συλλέγοντας την ανακλώμενη από τη Γη ηλιακή ακτινοβολία. Η εισερχόμενη ακτίνα φωτός διαχωρίζεται σε κατάλληλο φίλτρο και εστιάζεται σε δύο ξεχωριστά συγκροτήματα εστιακού επιπέδου (focal plane assemblies), ένα για το ορατό και εγγύς-υπέρυθρο τμήμα του φάσματος (Visible Near Infrared - VNIR) και ένα για το υπέρυθρο βραχέων κυμάτων (Short Wave Infrared - SWIR). Στη συνέχεια πραγματοποιείται φασματικός διαχωρισμός σε ζώνες με χρήση αντίστοιχων φίλτρων. [15]

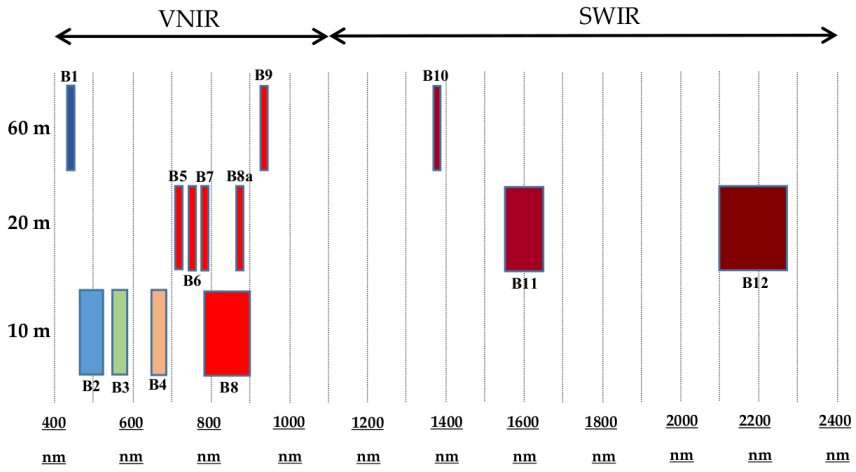


(a) Ολοκληρωμένη όψη οργάνου

(b) Διάταξη διαχωρισμού οπτικής δέσμης σε SWIR και VNIR

Σχήμα 2.3: Εσωτερική διαμόρφωση πολυφασματικού απεικονιστή (MSI) [16]

Ο πολυφασματικός απεικονιστής πραγματοποιεί μετρήσεις σε 13 φασματικές ζώνες (443-2.190nm) με χωρικές αναλύσεις μεταξύ 10 και 60m.



Σχήμα 2.4: Οι 13 φασματικές ζώνες του MSI ανά χωρική ανάλυση [16]

Συγκεκριμένα:

- 4 ζώνες έχουν χωρική ανάλυση 10m: το μπλε (490 nm), το πράσινο (560 nm), το ερυθρό (665 nm) και το εγγύς υπέρυθρο - NIR (842 nm)
- 6 ζώνες έχουν χωρική ανάλυση 20m: 4 στενές ζώνες, που χρησιμοποιούνται κυρίως για χαρακτηρισμό της βλάστησης στο όριο του ερυθρού (705 nm, 740 nm, 783 nm, 865 nm) και 2 ευρύτερες ζώνες υπέρυθρου βραχέων κυμάτων - SWIR (1610 nm και 2190 nm) για εφαρμογές όπως ο εντοπισμός χιονιού, πάγου ή νεφών ή η αξιολόγηση των επιπέδων υγρασίας της βλάστησης.
- 3 ζώνες έχουν χωρική ανάλυση 60m: των αερολυμάτων (443 nm), των υδρατμών (945 nm) και των νεφών (1375 nm). Οι ζώνες αυτές χρησιμοποιούνται σε εφαρμογές όπως ο εντοπισμός νεφών και οι ατμοσφαιρικές διορθώσεις. [15]

Ζώνη	**Ανάλυση (m)**	**Μήκος Κύματος (nm)**	**Εύρος Ζώνης (nm)**	**Περιγραφή**
B1	60	443	20	Aerosols
B2	10	490	65	Blue
B3	10	560	35	Green
B4	10	665	30	Red
B5	20	705	15	Red Edge 1
B6	20	740	15	Red Edge 2
B7	20	783	20	Red Edge 3
B8	10	842	115	NIR
B8B	20	865	20	Red Edge 4
B9	60	945	20	Water Vapor
B10	60	1375	30	Cirrus
B11	20	1610	90	SWIR 1
B12	20	2190	180	SWIR 2

Πίνακας 2.2: Φασματικές ζώνες πολυφασματικού απεικονιστή (MSI) [15]

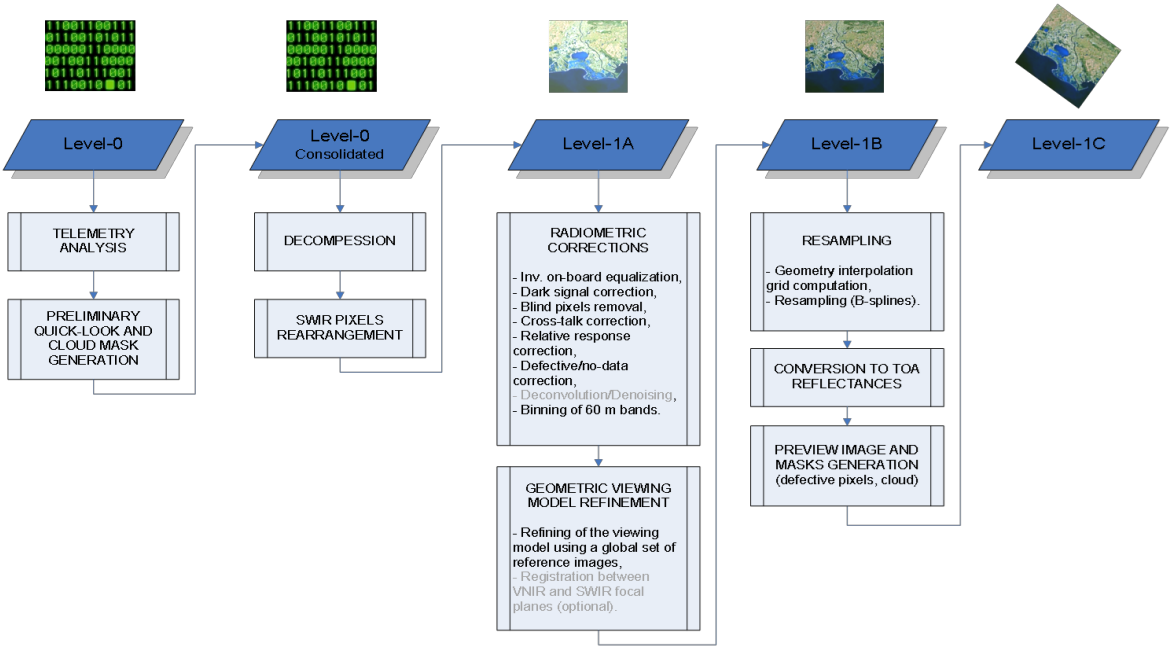
Τα δεδομένα που συγκεντρώνονται από το MSI περνάνε στη συνέχεια από διαδοχικά στάδια επεξεργασίας πριν παραχθούν τα τελικά προϊόντα στα οποία θα έχουν πρόσβαση οι χρήστες.

Τα στάδια αυτά, τα οποία απεικονίζονται στο Σχήμα 2.5, είναι 5, αλλά δημόσια διαθέσιμα είναι μόνο τα δύο τελευταία:

- Το επίπεδο Level-1C (L1C) παρέχει εικόνες Top-Of-Atmosphere (TOA) μετά από πλήρη ραδιομετρική διόρθωση και ευθυγράμμιση, σε χαρτογραφική γεωμετρία.
- Το επίπεδο Level-2A (L2A) παρέχει εικόνες Bottom-Of-Atmposphere (BOA), που προκύπτουν μετά από εντοπισμό των νεφών και κατάλληλη επεξεργασία των δεικτών Aerosol Optical Thickness και Water Vapour. Επιπλέον, πραγματοποιείται ταξινόμηση εδάφους (scene classification), και τα αποτελέσματα προστίθενται στα δεδομένα. Η επεξεργασία του επιπέδου Level-1C ώστε να προκύψει το επίπεδο Level-2A πραγματοποιείται από το πακέτο λογισμικού Sen2Cor της ESA. [17]

Τόσο τα δεδομένα του επιπέδου Level-1C όσο και αυτά του επιπέδου Level-2A είναι οργανωμένα ως μωσαϊκό τετράγωνων εικόνων (tiles) η κάθε μία εκ των οποίων αντιστοιχεί σε έκταση $100km^2$. Οι συντεταγμένες των εικόνων αναφέρονται στο σύστημα γεωγραφικών συντεταγμένων WGS84 (Universal Transverse Mercator/World Geodetic System 1984).

Για την παρούσα εργασία επιλέχθηκε η χρήση δεδομένων από το επίπεδο Level-2A, καθώς τα Bottom-Of-Atmposphere δεδομένα είναι σαφώς πιο κατάλληλα για τη συγκεκριμένη εφαρμογή.



Σχήμα 2.5: Η αλυσίδα επεξεργασίας των δεδομένων του Sentinel-2 [16]

Κεφάλαιο 3

Νευρωνικά Δίκτυα στην Υπολογιστική Όραση

3.1 Εισαγωγή

Η όραση υπολογιστών ή μηχανική όραση είναι ο επιστημονικός κλάδος που έχει ως στόχο τη δημιουργία συστημάτων που "βλέπουν", αντιλαμβάνονται δηλαδή μέσω της εικόνας ή του βίντεο τον κόσμο και εξάγουν συμπεράσματα. Οι εφαρμογές που σχετίζονται με την όραση υπολογιστών μπορεί να κυμαίνονται από τη ρομποτική όραση και την αλληλεπίδραση ανθρώπου-υπολογιστή μέχρι και την ιατρική απεικόνιση και διάγνωση, και η σημασία τους είναι καθοριστική στο σύγχρονο κόσμο. Είναι επόμενο, λοιπόν, τα προβλήματα που απασχολούν την όραση υπολογιστών να αποτελούν αντικείμενο συνεχούς έρευνας και η προσπάθεια βελτίωσης των μεθόδων που χρησιμοποιούνται να είναι αδιάκοπη.

Σχεδόν το σύνολο των σύγχρονων μεθόδων που χρησιμοποιούνται στα πλαίσια της όρασης υπολογιστών εντάσσονται στον τομέα της Μηχανικής Μάθησης, προσανατολίζονται δηλαδή στην επίλυση του κάθε προβλήματος όχι μέσω ρητού προγραμματισμού, αλλά μέσω της εκπαίδευσης πάνω σε διαθέσιμα δεδομένα και της εξαγωγής πληροφορίας από αυτά. Ανάλογα με τη φύση της κάθε εφαρμογής μπορεί να χρησιμοποιηθεί μεγάλη ποικιλία αρχιτεκτονικών και αλγορίθμων μηχανικής μάθησης. Για να γίνει κατανοητό το ζητούμενο της συγκεκριμένης εργασίας και ο τρόπος με τον οποίο συνδέεται με την αντίστοιχη επιλογή αρχιτεκτονικής, είναι απαραίτητο να αναφερθούν οι βασικές κατηγορίες προβλημάτων με τα οποία ασχολείται η μηχανική όραση, οι οποίες είναι οι ακόλουθες:

- Ταξινόμηση Εικόνων (Image Classification): δέχεται ως είσοδο μία εικόνα και προβλέπει τι απεικονίζεται, επιστρέφοντας ως έξοδο την πιθανότητα η εικόνα να ανήκει σε κάποια κλάση. Αφορά την εικόνα ως σύνολο, και δεν πραγματοποιείται σε επίπεδο εικονοστοιχείου.
- Ανίχνευση Αντικειμένου (Object Detection): δέχεται ως είσοδο μία εικόνα και εντοπίζει τη θέση των διαφόρων αντικειμένων σε αυτήν, καθώς και την κλάση τους. Όπως και η ταξινόμηση εικόνων, δεν πραγματοποιείται σε επίπεδο εικονοστοιχείου.
- Σημασιολογική Κατάτμηση Εικόνας (Semantic Segmentation): δέχεται ως είσοδο μία εικόνα και προβλέπει την κλάση κάθε εικονοστοιχείου.
- Κατάτμηση Στιγμιοτύπων Εικόνας (Instance Segmentation): αποτελεί συνδυασμό της ανίχνευσης αντικειμένου με την σημασιολογική κατάτμηση. Δέχεται ως είσοδο μία εικόνα και εντοπίζει τα διάφορα αντικείμενα που υπάρχουν σε αυτή (και την κλάση τους) σε επίπεδο εικονοστοιχείου.

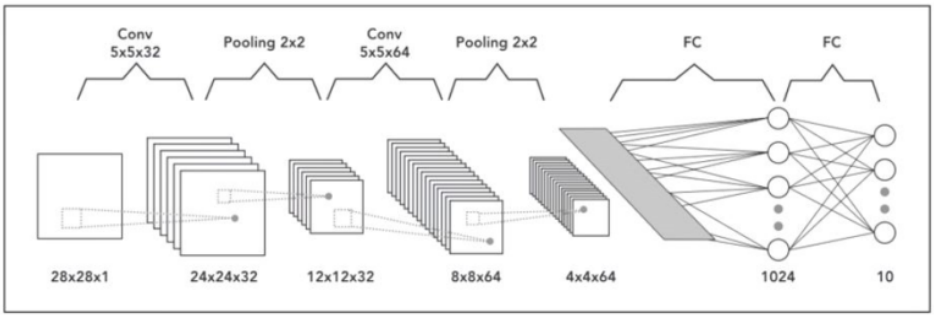
Το πρόβλημα που εξετάζεται στα πλαίσια της παρούσας εργασίας ανήκει στην τελευταία κατηγορία. Συγκεκριμένα, στόχος είναι όχι μόνο να αποφασίσουμε εάν το κάθε εικονοστοιχείο

ανήκει σε κάποια παραλία ή όχι, αλλά να αναγνωρίσουμε την κάθε παραλία (το κάθε στιγμιότυπο που ανήκει στην κλάση) ως ξεχωριστή οντότητα, και να επιλέξουμε όλα τα εικονοστοιχεία που την αποτελούν.

Παρακάτω θα αναλυθούν αρχικά κάποιες εισαγωγικές έννοιες σχετικά με τη δομή των Συνελικτικών Νευρωνικών Δικτύων και το πρόβλημα της ανίχνευσης αντικειμένων. Στη συνέχεια θα γίνει μία συνοπτική περιγραφή των αρχιτεκτονικών που χρησιμοποιούνται για κάθε ένα από τα προβλήματα της υπολογιστικής όρασης.

3.2 Συνελικτικά Νευρωνικά Δίκτυα (CNN)

Τα Συνελικτικά Νευρωνικά Δίκτυα είναι ίσως τα πιο διαδεδομένα μοντέλα όσον αφορά την ανάλυση εικόνας και την όραση υπολογιστών, καθώς υπερισχύουν των παραδοσιακών, Πλήρως Συνδεδεμένων Νευρωνικών Δικτύων σε δύο βασικά σημεία.



Σχήμα 3.1: Απλό Συνελικτικό Νευρωνικό Δίκτυο [1]

Αρχικά, στην περίπτωση ενός Πλήρως Συνδεδεμένου Δικτύου (Fully Connected Network), όλοι οι νευρώνες του κάθε επιπέδου είναι συνδεδεμένοι με όλους τους νευρώνες του επόμενου. Κατά συνέπεια, όταν η είσοδος του δικτύου είναι μία εικόνα, και ακόμη περισσότερο όταν αποτελείται από πάνω από ένα κανάλια (πχ RGB, πολυφασματικές εικόνες), ο αριθμός των συνδέσεων και κατά συνέπεια ο όγκος των υπερπαραμέτρων αυξάνεται δραματικά, με σαφείς δυσμενείς επιπτώσεις τόσο στην απαιτούμενη υπολογιστική ισχύ όσο και στο πλήθος των δεδομένων που απαιτούνται για μία επιτυχημένη εκπαίδευση. Αντιθέτως, στην περίπτωση ενός συνελικτικού δικτύου τα πλήρως συνδεδεμένα επίπεδα αντικαθίστανται από συνελικτικά επίπεδα, στα οποία ο κάθε νευρώνας συνδέεται με συγκεκριμένους νευρώνες του επόμενου επιπέδου, μέσω συνέλιξης με τα κατάλληλα φίλτρα. Τα φίλτρα αυτά έχουν τα ίδια βάρη για όλους τους νευρώνες του ίδιου επιπέδου και είναι οργανωμένα σε πλέγμα ώστε να εφαρμόζονται σε διαφορετικές περιοχές τις εικόνας, με αποτέλεσμα ο όγκος των παραμέτρων να είναι σημαντικά μικρότερος.

Επιπλέον, εάν χρησιμοποιήσουμε μία εικόνα ως είσοδο σε ένα Πλήρως Συνδεδεμένο Δίκτυο, αυτή θα λειτουργεί σαν μονοδιάστατο διάνυσμα, με αποτέλεσμα να μην μπορούμε να εκμεταλλευτούμε τις χωρικές συσχετίσεις μεταξύ των τιμών γειτονικών εικονοστοιχείων. Αντιθέτως, τα συνελικτικά επίπεδα εφαρμόζουν τεχνικές κυλιόμενου παραθύρου, με αποτέλεσμα στα Συνελικτικά Νευρωνικά Δίκτυα η χωρική συσχέτιση μεταξύ των δεδομένων να διατηρείται.

Παρακάτω επεξηγούνται συνοπτικά τα διαφορετικά στρώματα ενός συνελικτικού δικτύου.

[1] `https://engmrk.com/module-22-implementation-of-cnn-using-keras/`

Συνελικτικό Επίπεδο (Convolutional Layer)

Για την περιγραφή του τρόπου λειτουργίας του συνελικτικού επιπέδου πρέπει αρχικά να ορίσουμε την έννοια της συνέλιξης εικόνων.

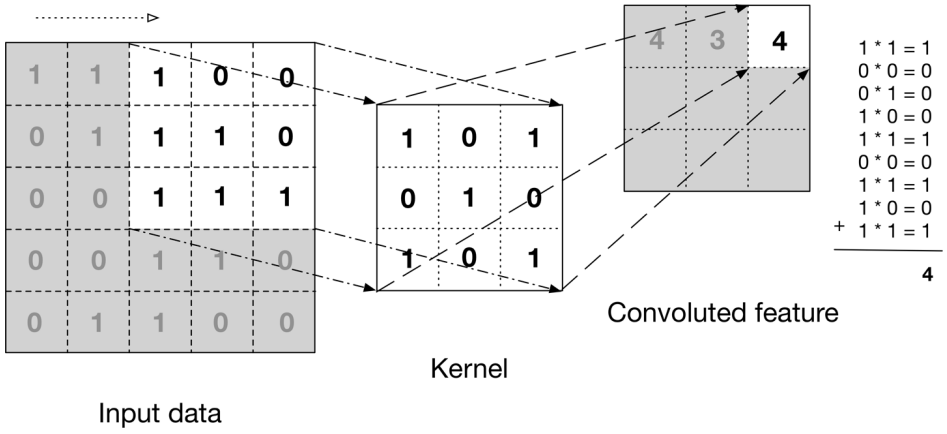
Στο πεδίο της επεξεργασίας εικόνας, πολύ συχνά χρησιμοποιούμε την έννοια του πυρήνα (kernel) ή φίλτρου (filter), ο οποίος μπορεί να οριστεί ως ένα παράθυρο συγκεκριμένης, μικρής διάστασης n x n ίδιου βάθους με την εικόνα (για παράδειγμα 3 σε περίπτωση RBG εικόνας). Σε αναλογία με τη διακριτή συνέλιξη μονοδιάστατων σημάτων, που ορίζεται ως

$$(f * g)[n] = f[n] * g[n] = \sum_{m=-\infty}^{\infty} f[m]g[n-m]$$

όπου f, g μονοδιάστατα διακριτά σήματα, η συνέλιξη των διδιάστατων διακριτών σημάτων της εικόνας f με τον πυρήνα g ορίζεται ως

$$(f * g)[x, y] = f[x, y] * g[x, y] = \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} f[n_1, n_2]g[x-n_1, y-n_2]$$

Σε κάθε σημείο της εικόνας, δηλαδή, τα στοιχεία του πυρήνα πολλαπλασιάζονται με τα εικονοστοιχεία της αντίστοιχης περιοχής, και το αποτέλεσμα τοποθετείται στην κατάλληλη θέση του πίνακα εξόδου. Η έξοδος κάθε τέτοιας πράξης ονομάζεται χάρτης ενεργοποίησης (activation map) ή χάρτης χαρακτηριστικών (feature map), καθώς η τιμή του χάρτη σε κάθε θέση εκφράζει την πιθανότητα με την οποία το επιθυμητό χαρακτηριστικό βρίσκεται σε αυτή την περιοχή της αρχικής εικόνας.



Σχήμα 3.2: Συνελικτικό Επίπεδο [2]

Τα βάρη του πυρήνα αποτελούν εκπαιδεύσιμες παραμέτρους του δικτύου. Αν και σύμφωνα με τον τυπικό ορισμό της διακριτής συνέλιξης εικόνων, ο πυρήνας ολισθαίνει μόνο ένα χωρικό βήμα σε κάθε μετακίνησή του, στην πράξη κάποιες φορές το βήμα ολίσθησης που χρησιμοποιείται (stride) είναι μεγαλύτερο. Επίσης, συνήθως χρησιμοποιούνται πάνω από ένας πυρήνας ανά συνελικτικό επίπεδο, δίνοντας ως έξοδο πολλαπλούς χάρτες ενεργοποίησης που αντιστοιχούν σε διαφορετικά χαρακτηριστικά (ένα για κάθε πυρήνα), και κατά συνέπεια η έξοδος του κάθε συνελικτικού επιπέδου είναι μία τρισδιάστατη "εικόνα" μεγάλου βάθους, που αποτελείται από διαφορετικούς χάρτες ενεργοποίησης.

[2] `https://www.researchgate.net/figure/The-convolution-operation-Source-14_fig2_342692021`

Επίπεδο Ενεργοποίησης (Activation Layer)

Τα περισσότερα συστήματα τα οποία καλείται να προσεγγίσει ένα Συνελικτικό Νευρωνικό Δίκτυο είναι πραγματικά συστήματα, και για το λόγο αυτό η συμπεριφορά τους δεν είναι απόλυτα γραμμική. Προκειμένου να εισαχθεί η απαραίτητη μη-γραμμικότητα στο δίκτυο, το κάθε συνελικτικό επίπεδο ακολουθείται από ένα επίπεδο ενεργοποίησης το οποίο εφαρμόζει στη έξοδό του μία (μη γραμμική) συνάρτηση ενεργοποίησης φ. Η πιο ευρέως χρησιμοποιούμενη συνάρτηση ενεργοποίησης είναι η Rectified Linear Unit (ReLU),

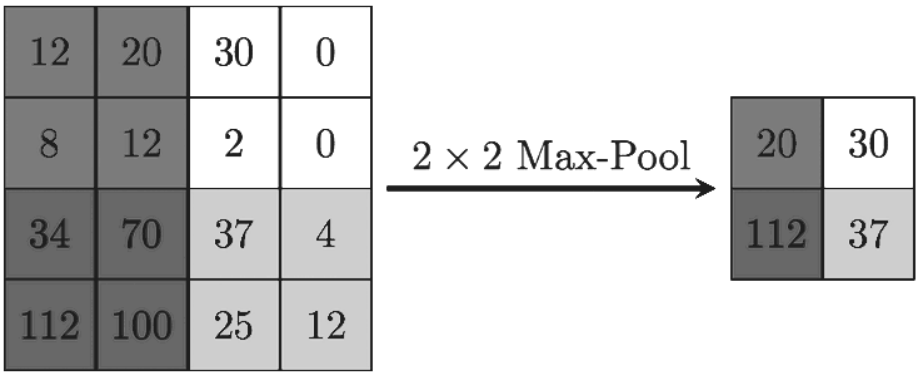
$$\phi(x) = max(0, x)$$

καθώς έχει αποδειχθεί ότι επιταχύνει την εκπαίδευση του δικτύου, απλοποιώντας το back-propagation.

Επίπεδο Υποδειγματοληψίας (Pooling Layer)

Ο σκοπός των επιπέδων υποδειγματοληψίας είναι να μειώνουν τις διαστάσεις των χαρτών ενεργοποίησης που προκύπτουν από τα συνελικτικά επίπεδα. Λειτουργούν χωρίζοντας τον χάρτη σε μη επικαλυπτόμενα τμήματα, για κάθε ένα από τα οποία επιλέγουν μία αντιπροσωπευτική τιμή. Στην πιο συνηθισμένη περίπτωση η τιμή αυτή υπολογίζεται ως το μέγιστο (υποδειγματοληψία μεγίστου - max pooling), αλλά μπορεί να χρησιμοποιηθεί και ο μέσος όρος (υποδειγματοληψία μέσου όρου) ή και τυχαία επιλογή (στοχαστική υποδειγματοληψία).

Επιπλέον της μείωσης των παραμέτρων και άρα της βελτίωσης της ταχύτητας εκπαίδευσης, η ύπαρξη του επιπέδου υποδειγματοληψίας μειώνει και την πιθανότητα υπερεκπαίδευσης του δικτύου.



Σχήμα 3.3: Επίπεδο Υποδειγματοληψίας [3]

Επίπεδο Κανονικοποίησης Παρτίδας (Batch Normalization Layer)

Όπως είδαμε μέχρι τώρα, τα Συνελικτικά Νευρωνικά Δίκτυα αποτελούνται από διαδοχικά επίπεδα συνελίξεων, με την έξοδο του κάθε επιπέδου να αποτελεί την είσοδο του επόμενου. Λόγω της δομής αυτής, στις περιπτώσεις Συνελικτικών Νευρωνικών Δικτύων με μεγάλο αριθμό επιπέδων, συχνά παρατηρείται ένα φαινόμενο που ονομάζεται "internal covariate shift", κατά το

[3] https://computersciencewiki.org/index.php/Max-pooling_/_Pooling

οποίο η προσαρμογή των παραμέτρων του δικτύου κατά την εκπαίδευση προκαλεί αλλαγή στην κατανομή των ενεργοποιήσεων των διαφόρων επιπέδων (κυρίως των τελευταίων).

Για την αποφυγή του συγκεκριμένου φαινομένου, συνηθίζεται η προσθήκη επιπέδων κανονικοποίησης παρτίδας, τα οποία εξασφαλίζουν την κανονικοποίηση των δεδομένων κάθε παρτίδας (batch) σε κάθε επίπεδο. Η κανονικοποίηση γίνεται με χρήση των στατιστικών χαρακτηριστικών του υποσυνόλου, ώστε ο μέσος όρος να ισούται με 0 και η διακύμανση να είναι μοναδιαία.

Πλήρως Συνδεδεμένο Επίπεδο (Fully Connected Layer)

Το πρώτο μέρος κάθε Συνελικτικού Νευρωνικού Δικτύου αποτελείται από διαδοχικές επαναλήψεις των επιπέδων που περιγράψαμε παραπάνω (Συνελικτικά - Ενεργοποίησης - Υποδειγματοληψίας - Κανονικοποίησης), ονομάζεται Δίκτυο Εξαγωγής Χαρακτηριστικών και έχει ως έξοδο έναν τελικό τρισδιάστατο χάρτη ενεργοποίησης μεγάλου βάθους.

Στην περίπτωση των δικτύων ταξινόμησης, η έξοδος αυτή αναδιατάσσεται σε διάνυσμα (flatten) και δίνεται ως είσοδος σε ένα σύνολο πλήρως συνδεδεμένων επιπέδων, το οποίο αναλαμβάνει την τελική ταξινόμηση της εικόνας. Τα επίπεδα αυτά συνήθως χρησιμοποιούν σαν συνάρτηση ενεργοποίησης τη ReLU, με εξαίρεση το τελευταίο το οποίο χρησιμοποιεί την SoftMax:

$$\sigma(x)_i = \frac{exp(x_i)}{\sum_j exp(x_j)}$$

3.3 Εισαγωγικές Έννοιες Ανίχνευσης Αντικειμένων

3.3.1 Πλαίσιο Οριοθέτησης (Bounding Box)

Ως πλαίσιο οριοθέτησης ενός αντικειμένου σε μία εικόνα ορίζεται το μικρότερο δυνατό ορθογώνιο τμήμα της εικόνας στο εσωτερικό του οποίου βρίσκεται ολόκληρο το αντικείμενο. Για την περιγραφή ενός πλαισίου οριοθέτησης είναι απαραίτητες 4 τιμές, οι οποίες μπορούν να είναι ενδεικτικά:

- οι συντεταγμένες (i_0, j_0) της κάτω αριστερής και (i_1, j_1) της πάνω δεξιάς γωνίας του
- οι συντεταγμένες (i_0, j_0) της κάτω αριστερής γωνίας, το πλάτος w και το ύψος h
- οι συντεταγμένες (i_c, j_c) του κέντρου, το πλάτος w και το ύψος h

3.3.2 Περιοχή Ενδιαφέροντος (Region of Interest - ROI)

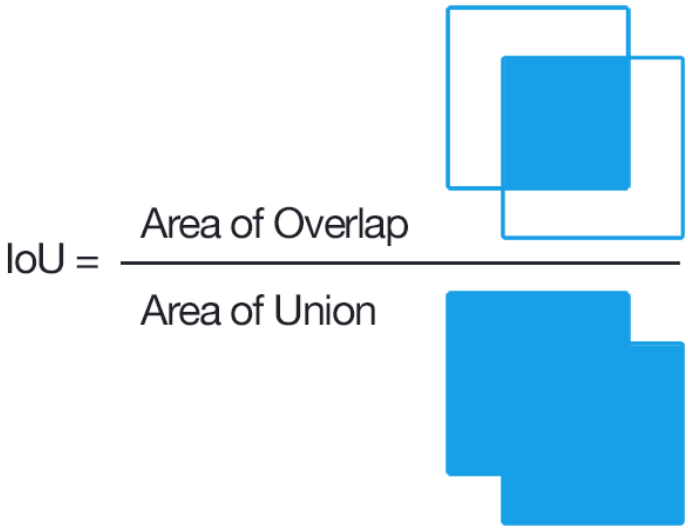
Ως περιοχή ενδιαφέροντος (Region of Interest) ή πρόταση περιοχής (region proposal) ορίζεται μία ορθογώνια περιοχή της εικόνας εισόδου η οποία θεωρητικά είναι πιθανό να περιέχει ένα αντικείμενο. Οι περιοχές αυτές μπορούν να υπολογιστούν με χρήση είτε κάποιου εξωτερικού αλγορίθμου όπως ο Selective Search [18] ή ο Edge Box detection [19], είτε ενός Δικτύου Προτάσεων Περιοχών (Region Proposal Network - RPN) [20].

3.3.3 Λόγος τομής προς Ένωση (Intersection over Union)

Για την υλοποίηση οποιουδήποτε συστήματος ανίχνευσης αντικειμένων είναι απαραίτητος ο ορισμός μίας μετρικής της ομοιότητας μεταξύ δύο αντικειμένων. Η μετρική αυτή χρησιμοποιείται τόσο για τη σύγκριση της πρόβλεψης με το αληθινό αντικείμενο (ground truth) για την

αξιολόγησή της, όσο και για τη σύγκριση διαφορετικών προβλέψεων μεταξύ τους με σκοπό την απαλοιφή υπερβολικά όμοιων προβλέψεων. Το πιο ευρέως χρησιμοποιούμενο μέγεθος για το σκοπό αυτό είναι ο λόγος της τιμής προς την ένωση (Intersection over Union - IoU).

Στην περίπτωση της σύγκρισης αντικειμένων που περιγράφονται από πλαίσια οριοθέτησης, το IoU ορίζεται ως το εμβαδόν της τομής των δύο πλαισίων, δηλαδή της περιοχής που ανήκει τόσο στο ένα πλαίσιο όσο και στο άλλο, προς το εμβαδόν της ένωσής τους, δηλαδή της συνολικής περιοχής που καλύπτουν και τα δύο πλαίσια, όπως φαίνεται στο Σχήμα 3.4.



Σχήμα 3.4: Ορισμός IoU [4]

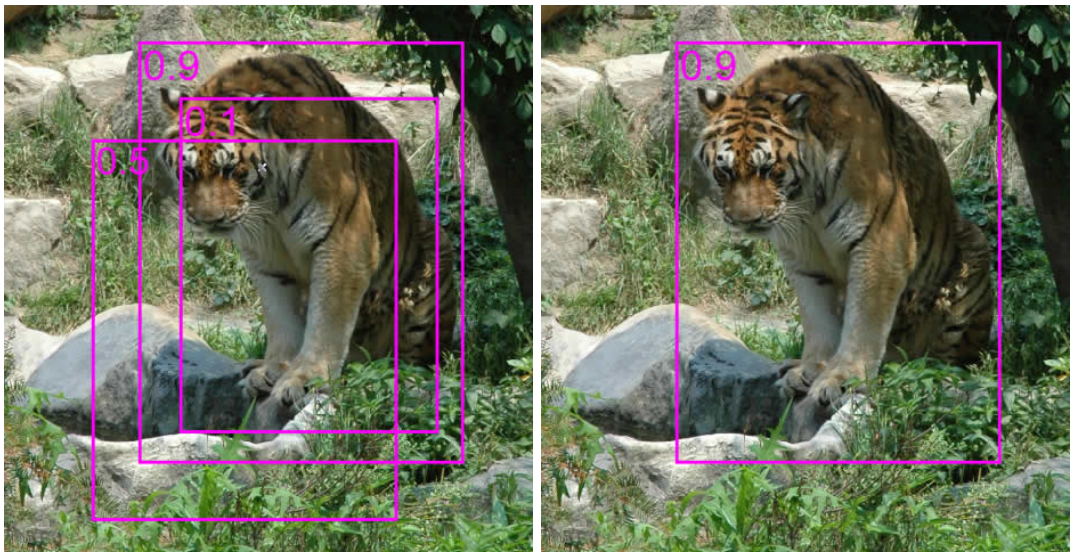
Όπως αναφέρθηκε, η IoU ως συνεχές μέγεθος χρησιμοποιείται για την ποσοτικοποίηση της ποιότητας των προβλέψεων. Για τη μετατροπή της από συνεχές μέγεθος σε διακριτό (0: μη ταύτιση αντικειμένων, 1: ταύτιση αντικειμένων) χρησιμοποιείται ένα κατώφλι (IoU threshold), το οποίο στις περισσότερες εφαρμογές ανίχνευσης αντικειμένων τίθεται ίσο με 0.5. Στο σημείο αυτό είναι σημαντικά να σημειωθεί ότι μετά από εξέταση των αποτελεσμάτων των πειραμάτων, στο πλαίσιο της παρούσας εργασίας επιλέχθηκε να αποκλίνουμε από τη συνήθη πρακτική και να θέσουμε μία χαμηλότερη τιμή, λόγω της ιδιαιτερότητας των δεδομένων. Συγκεκριμένα, τα αντικείμενα που πρέπει να εντοπίσουμε έχουν μακρόστενο σχήμα με πολύ μικρό πλάτος, γεγονός που σε συνδυασμό με τη σχετικά χαμηλή ανάλυση των δεδομένων οδηγεί συχνά σε χαμηλές τιμές IoU ακόμη και για καλές προβλέψεις.

3.3.4 Καταστολή μη μεγίστων (Non-Maximum Suppression)

Ένα από τα βασικότερα προβλήματα που πρέπει να αντιμετωπιστούν στο πλαίσιο της ανίχνευσης αντικειμένων είναι η ύπαρξη πολλών προβλέψεων με μικρές διαφορές οι οποίες αντιστοιχούν στο ίδιο αντικείμενο. Η καταστολή μη μεγίστων (Non-Maximum Supression - NMS) είναι ένας άπληστος (greedy) αλγόριθμος ο οποίος χρησιμοποιείται στις περισσότερες σύγχρονες αρχιτεκτονικές ανίχνευσης αντικειμένων ώστε να συγχωνεύσει αυτά τα αλληλοεπικαλυπτόμενα πλαίσια οριοθέτησης.

[4] https://www.pyimagesearch.com/2016/11/07/intersection-over-union-iou-for-object-detection/

Αρχικά, ο αλγόριθμος ταξινομεί όλα τα πλαίσια σε αύξουσα σειρά ως προς την πιθανότητά τους να αντιστοιχούν σε κάποιο αντικείμενο. Στη συνέχεια, επιλέγει το πλαίσιο με τη μεγαλύτερη πιθανότητα και, συγκρίνοντάς το με κάθε ένα από τα πλαίσια με μικρότερη πιθανότητα, απορρίπτει όσα έχουν επικάλυψη IoU μικρότερη από μία προκαθορισμένη τιμή. Η τιμή αυτή αποτελεί μία από τις υπερπαραμέτρους του συστήματός μας. Η διαδικασία αυτή επαναλαμβάνεται όσες φορές είναι απαραίτητο.



(a) πριν την NMS (b) μετά την NMS

Σχήμα 3.5: Πλαίσια οριοθέτησης ανιχνευμένων αντικειμένων πριν και μετά την NMS [5]

3.4 Πρόβλημα Ανίχνευσης Αντικειμένων

3.4.1 Περιγραφή

(a) Ταξινόμηση Εικόνας (b) Ανίχνευση Αντικειμένου

Σχήμα 3.6: Ταξινόμηση και Ανίχνευση Αντικειμένου [6]

Ένα από τα πιο σημαντικά προβλήματα που απασχολούν την όραση υπολογιστών είναι η ανίχνευση αντικειμένων σε εικόνες (Object Detection). Το πρόβλημα αυτό περιγράφεται ως εξής: Με δεδομένη μία εικόνα εισόδου, πρέπει να προβλεφθεί η τοποθεσία και η έκταση των αντικειμένων που ανήκουν σε ένα σύνολο προκαθορισμένων κλάσεων, και να αποδοθεί η σωστή

[5] http://www.interstellarengine.com/ai/Non-maximum-suppression.html
[6] https://towardsdatascience.com/detection-and-segmentation-through-convnets-47aa42de27ea

κλάση στο κάθε ένα. Η τοποθεσία των αντικειμένων συνήθως εκφράζεται ως το ελάχιστο πλαίσιο οριοθέτησης που περικλείει εξ ολοκλήρου το αντικείμενο.

Προφανώς, για τη λύση αυτού του προβλήματος δεν αρκεί ένα Συνελικτικό Νευρωνικό Δίκτυο, καθώς η ίδια η φύση του προβλήματος είναι εντελώς διαφορετική από αυτή της απλής ταξινόμησης. Πέρα από την αναγνώριση των αντικειμένων και την πρόβλεψη των κλάσεων τους, υπεισέρχεται και το πρόβλημα του εντοπισμού (localization). Θα μπορούσαμε να περιγράψουμε την ανίχνευση αντικειμένων ως τη σύνθεση δύο διαφορετικών προβλημάτων: ένα πρόβλημα ταξινόμησης και ένα πρόβλημα παλινδρόμησης, γνωστό και ως bounding box regression [21].

Τα τελευταία χρόνια έχει προταθεί μεγάλος αριθμός εξειδικευμένων στην ανίχνευση αντικειμένων μοντέλων, τα οποία χωρίζονται σε δύο κατηγορίες ως προς τη δομή τους [22]:

- Τα μοντέλα ενός σταδίου (one-step models), όπως τα YOLO, Multibox, AttentionNet, G-CNN, χρησιμοποιούν ένα feed forward CNN για να προσδιορίσουν την τοποθεσία των αντικειμένων ενδιαφέροντος. Το γεγονός ότι σε αυτή την κατηγορία μοντέλων δεν πραγματοποιούνται region proposals τα καθιστά απλούστερα και ταχύτερα, αλλά η απόδοσή τους είναι μειωμένη, κυρίως όταν απαιτείται και κατάτμηση της εικόνας, όπως στην περίπτωσή μας [22]. Κατά συνέπεια, δε θα μας απασχολήσουν στο πλαίσιο της εργασίας.
- Η άλλη κατηγορία (two-step models ή region-based models) περιλαμβάνει μοντέλα όπως το R-CNN [23], το Fast R-CNN [24], το FPN, το Faster R-CNN [20] και το R-FPN. Τα μοντέλα αυτά ενώ έχουν αρκετές διαφορές, έχουν περίπου κοινή δομή. Αρχικά χρησιμοποιούν, ως πρώτο βήμα, έναν αλγόριθμο (πχ Selective Search [18]) ή ένα μοντέλο (συνήθως ένα region-based CNN) που δέχεται σαν είσοδο την εικόνα και προτείνει διαφορετικές πιθανές περιοχές ενδιαφέροντος. Στη συνέχεια (δεύτερο βήμα), χρησιμοποιείται ένα Συνελικτικό Νευρωνικό Δίκτυο ως feature extractor ώστε να υπολογιστεί ο χάρτης χαρακτηριστικών κάθε περιοχής ενδιαφέροντος, ο οποίος τελικά δίνεται ως είσοδος σε ένα πλήρως συνδεδεμένο επίπεδο (ή εναλλακτικά μία Μηχανή Διανυσμάτων Υποστήριξης (Support Vector Machine - SVM) [25], το οποίο επιστρέφει το αποτέλεσμα της ταξινόμησης, συνοδευόμενο από ένα confidence score που δηλώνει την πιθανότητα της πρόβλεψης [22].

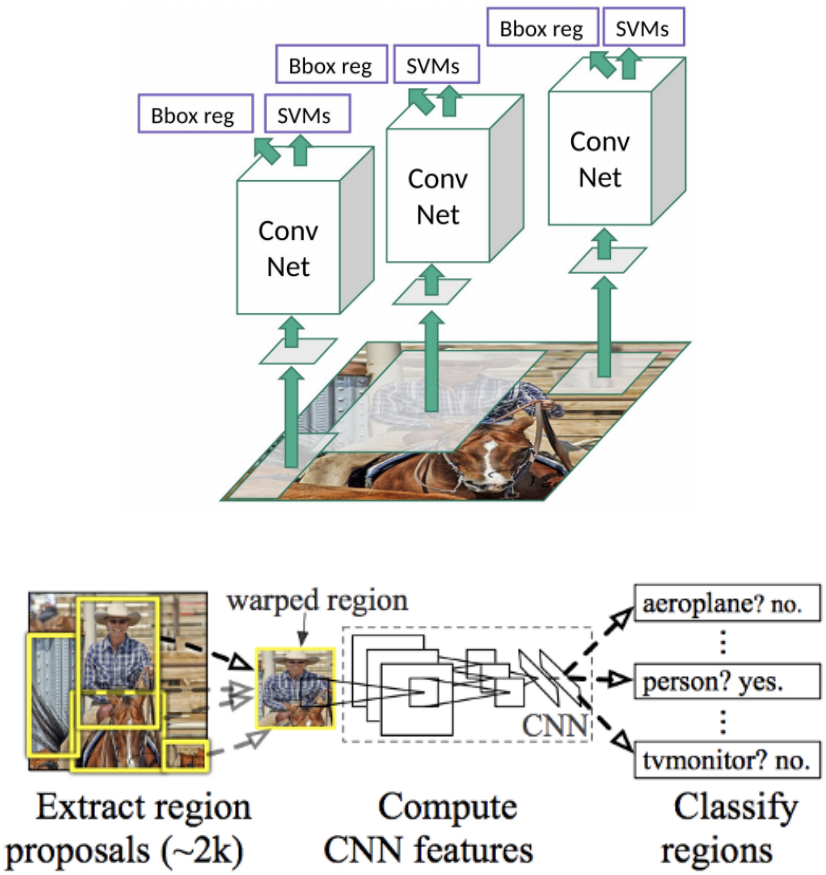
3.4.2 R-CNN

Η λειτουργία του Mask R-CNN, του μοντέλου που θα μας απασχολήσει στο πλαίσιο της εργασίας, βασίζεται σε πολύ μεγάλο βαθμό στα R-CNN, Fast R-CNN και Faster R-CNN, αφού στην πραγματικότητα προέκυψε από αυτά μέσω συνεχών βελτιώσεων και προσθηκών. Λόγω της εξαιρετικά παρόμοιας δομής και λειτουργίας τους, κρίθηκε απαραίτητο να γίνει μία συνοπτική αναφορά στα χαρακτηριστικά και στις διαφορές τους, ώστε να γίνει καλύτερα κατανοητή η περιγραφή του στη συνέχεια.

Το 2014, προτάθηκε το μοντέλο R-CNN από τον R. Girshick et al., με σκοπό να συνδυάσει προτάσεις περιοχών (region proposals) με Συνελικτικά Νευρωνικά Δίκτυα (CNN) ώστε να εντοπίζει αντικείμενα σε εικόνες με χρήση πλαισίων οριοθέτησης [23].

Για να το πετύχει αυτό, το R-CNN χρησιμοποιεί τον αλγόριθμο Selective Search [18] ώστε να παράγει 2000 προτάσεις περιοχών ανά εικόνα. Στη συνέχεια, η κάθε περιοχή, μετά από προσαρμογή του μεγέθους της, δίνεται ως είσοδος σε ένα CNN, του οποίου η έξοδος είναι ένα διάνυσμα 4096 χαρακτηριστικών (feature vector). Τέλος, μία Μηχανή Διανυσμάτων Υποστήρι-

ξης (SVM) δέχεται ως είσοδο το διάνυσμα και αποφασίζει για την ύπαρξη ή όχι αντικειμένων σε αυτό.



Σχήμα 3.7: Δομή R-CNN [26, 27]

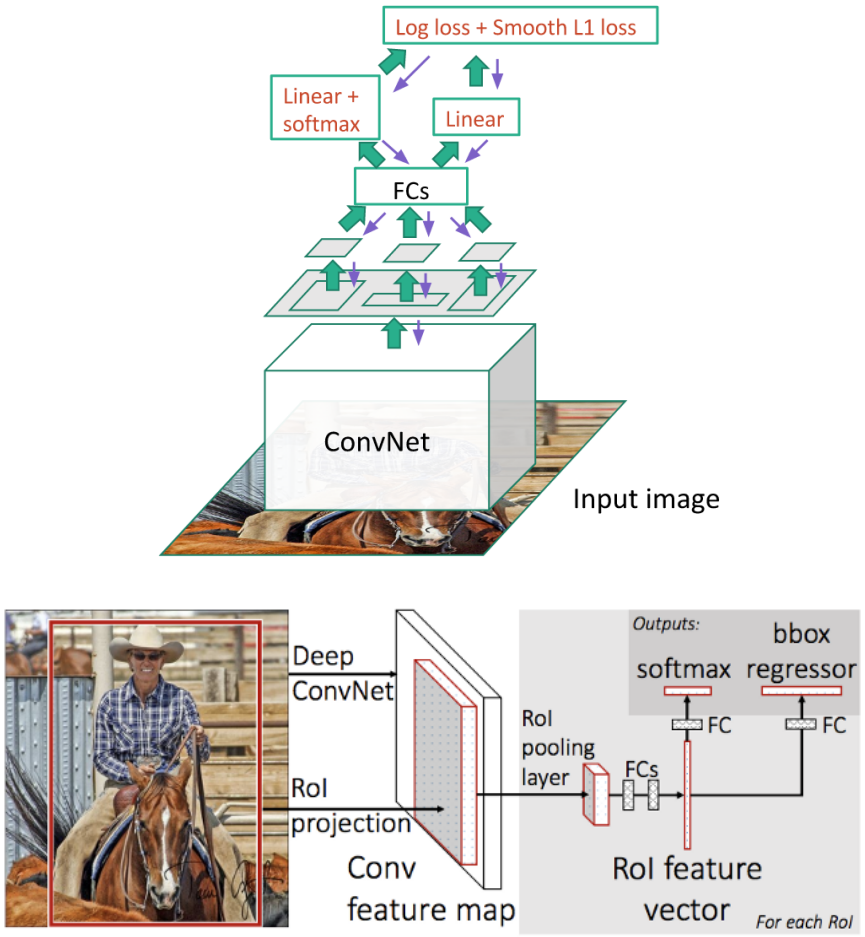
Αν και το R-CNN πέτυχε σημαντικά καλύτερη απόδοση σε σχέση με τις υπάρχουσες αρχιτεκτονικές, είχε κάποια σαφή μειονεκτήματα, κυρίως σχετικά με την ταχύτητα [26]. Τα βασικότερα από αυτά είναι:

- Η μελέτη 2000 προτάσεων ανά εικόνα οδηγεί αναπόφευκτα σε πολύ μεγάλο χρόνο εκπαίδευσης, καθώς πρέπει να υπολογιστεί ξεχωριστά το διάνυσμα χαρακτηριστικών κάθε περιοχής.
- Πέρα από το χρόνο εκπαίδευσης, και ο χρόνος του testing είναι απαγορευτικά μεγάλος (πάνω από 40 δευτερόλεπτα ανά εικόνα), με αποτέλεσμα να μην είναι δυνατή η χρήση του μοντέλου για εφαρμογές πραγματικού χρόνου.
- Επίσης, η συμπεριφορά του αλγορίθμου Selective Search 2000 περιοχών είναι προκαθορισμένη, με αποτέλεσμα το πρώτο τμήμα της αναγνώρισης να μην βελτιώνεται μέσω εκπαίδευσης.

3.4.3 Fast R-CNN

Οι συγγραφείς του R-CNN, αντιλαμβανόμενοι τα παραπάνω προβλήματα, πρότειναν το 2015 το βελτιωμένο Fast R-CNN [24].

Στην πραγματικότητα, κατόρθωσαν να βελτιώσουν σημαντικά την ταχύτητα του μοντέλου αλλάζοντας απλώς τη σειρά των επιπέδων του. Αντί να δίνεται ως είσοδος στο CNN κάθε μία από τις περιοχές ενδιαφέροντος, το CNN εξάγει τα χαρακτηριστικά ολόκληρης της εικόνας σε μορφή χάρτη χαρακτηριστικών, και στη συνέχεια για κάθε περιοχή απομονώνεται το αντίστοιχο τμήμα. Με αυτό τον τρόπο, η εξαγωγή των χαρακτηριστικών γίνεται μόνο μία φορά αντί για 2000, γεγονός που είναι προφανές ότι βελτιώνει κατά πολύ την ταχύτητα της εκπαίδευσης και της πρόβλεψης.



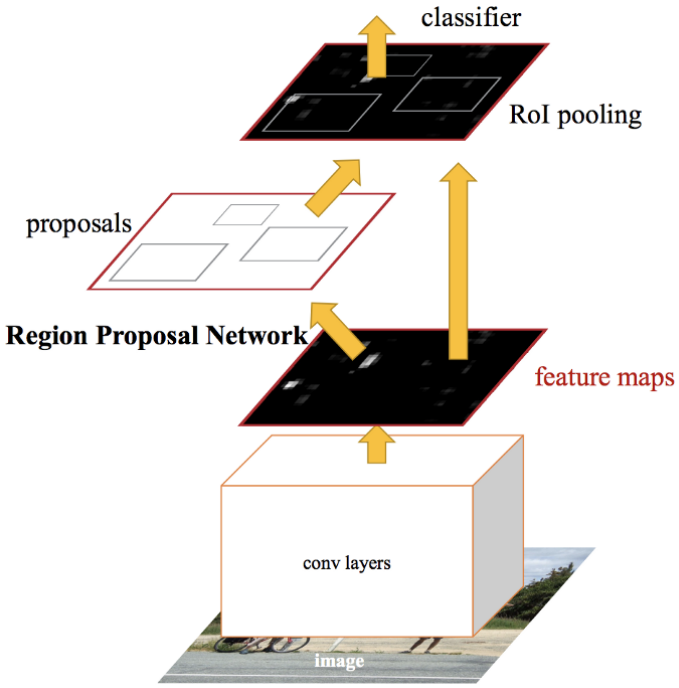
Σχήμα 3.8: Δομή Fast R-CNN [26, 27]

3.4.4 Faster R-CNN

Παρά τη βελτίωση της επίδοσης που παρατηρήθηκε με τη μετάβαση από το R-CNN στο Fast R-CNN, υπήρχε ακόμα κάποιο περιθώριο για εξέλιξη, αφού μόνο τα δύο πρώτα από τα τρία προβλήματα του R-CNN που αναφέρθηκαν στην προηγούμενη υποενότητα είχαν αντιμετωπιστεί.

Οι Shaoqing Ren et al. μελετώντας τις αδυναμίες του Selective Search, πρότειναν την αντικατάστασή του από τον δικό τους αλγόριθμο πρότασης περιοχών. Συγκεκριμένα, αντιλήφθηκαν ότι ο χάρτης χαρακτηριστικών που παράγεται από το συνελικτικό τμήμα του Fast R-CNN μπορεί να χρησιμοποιηθεί αποτελεσματικά και για το πρόβλημα της πρότασης περιοχών, αντικαθιστώντας τις αργές μεθόδους όπως η Selective Search με ένα εκπαιδεύσιμο Νευρωνικό Δίκτυο, γνωστό

ως Δίκτυο Πρότασης Περιοχών [20]. Το μοντέλο που προέκυψε ευφάνταστα ονομάστηκε Faster R-CNN και η δομή του φαίνεται στην εικόνα.

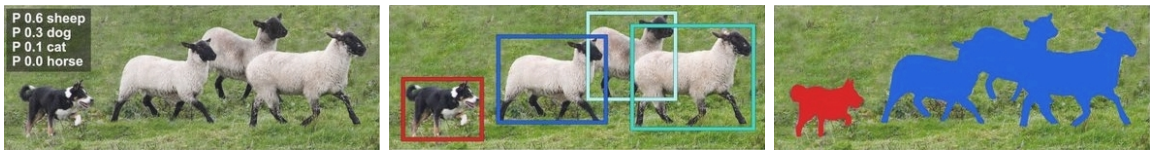


Σχήμα 3.9: Δομή Faster R-CNN [20]

3.5 Πρόβλημα Σημασιολογικής Κατάτμησης

3.5.1 Περιγραφή

Ως σημασιολογική κατάτμηση ορίζεται το πρόβλημα της ταξινόμησης σε επίπεδο εικονοστοιχείου, δηλαδή της ανάθεσης σε κάθε εικονοστοιχείο της εικόνας της κλάσης στην οποία ανήκει. Σε αντίθεση με το πρόβλημα της ανίχνευσης αντικειμένων, δεν υπάρχει η έννοια της οντότητας που ανήκει σε κάποια κλάση, αλλά η πρόβλεψη πρέπει να πραγματοποιείται με χωρική ανάλυση ενός εικονοστοιχείου.



(a) Ταξινόμηση Εικόνας (b) Ανίχνευση Αντικειμένου (c) Σημασιολογική Κατάτμηση

Σχήμα 3.10: Ταξινόμηση, Ανίχνευση Αντικειμένου και Σημασιολογική Κατάτμηση [7]

Μία προφανής, αφελής πρώτη προσέγγιση του ζητούμενου θα ήταν η υλοποίηση ενός μοντέλου με διαδοχικά συνελικτικά επίπεδα, του οποίου η έξοδος θα είχε την ίδια διάσταση με

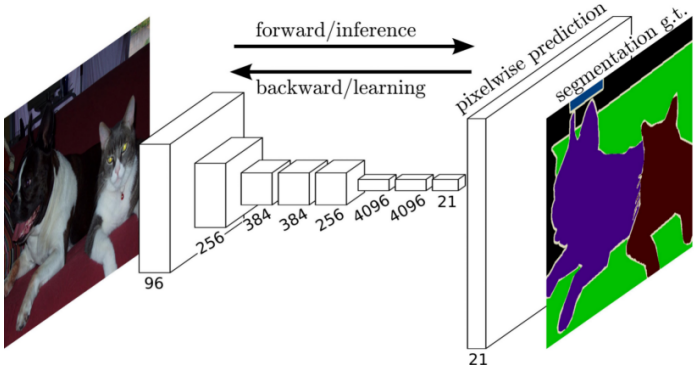
[7] https://towardsdatascience.com/detection-and-segmentation-through-convnets-47aa42de27ea

την είσοδο. Με τον τρόπο αυτό, μετά την εκπαίδευση του μοντέλου το κάθε εικονοστοιχείο της εξόδου θα αποτελούσε την πρόβλεψη για την κλάση του αντίστοιχου εικονοστοιχείου της αρχικής εικόνας. Παρόλα αυτά, εύκολα γίνεται αντιληπτό ότι ένα τέτοιου τύπου μοντέλο, στο οποίο η διάσταση της εικόνας διατηρείται αμείωτη κατά μήκος του δικτύου, θα διέθετε απαγορευτική υπολογιστική πολυπλοκότητα.

Για το λόγο αυτό, οι πιο συνηθισμένες προσεγγίσεις για το πρόβλημα της σημασιολογικής κατάτμησης αφορούν δίκτυα με δομή κωδικοποιητή - αποκωδικοποιητή (encoder-decoder). Η διάσταση δηλαδή της εικόνας μειώνεται αρχικά (encoder), παράγοντας χαμηλότερης ανάλυσης χάρτες χαρακτηριστικών οι οποίοι έχουν πολύ καλά αποτελέσματα για την ταξινόμηση μεταξύ των κλάσεων, και στη συνέχεια αυξάνεται και πάλι (decoder), μέχρι να προκύψει ο τελικός χάρτης κατάτμησης.

3.5.2 Πλήρως Συνελικτικά Δίκτυα (FCN)

Το Πλήρως Συνελικτικό Δίκτυο (Fully Convolutional Network - FCN) για Σημασιολογική Κατάτμηση εικόνας προτάθηκε το 2015 από τον Jonathan Long και την ομάδα του UC Berkeley [28]. Συγκεκριμένα, με αφετηρία κάποια αρχιτεκτονική συνελικτικού δικτύου, αντικαθιστώντας τα πλήρως συνδεδεμένα επίπεδα με πλήρως συνελικτικά επίπεδα, κατόρθωσαν να δημιουργήσουν ένα μοντέλο που θα έχει ως έξοδο εικόνα και όχι απλώς μία ταξινόμηση.

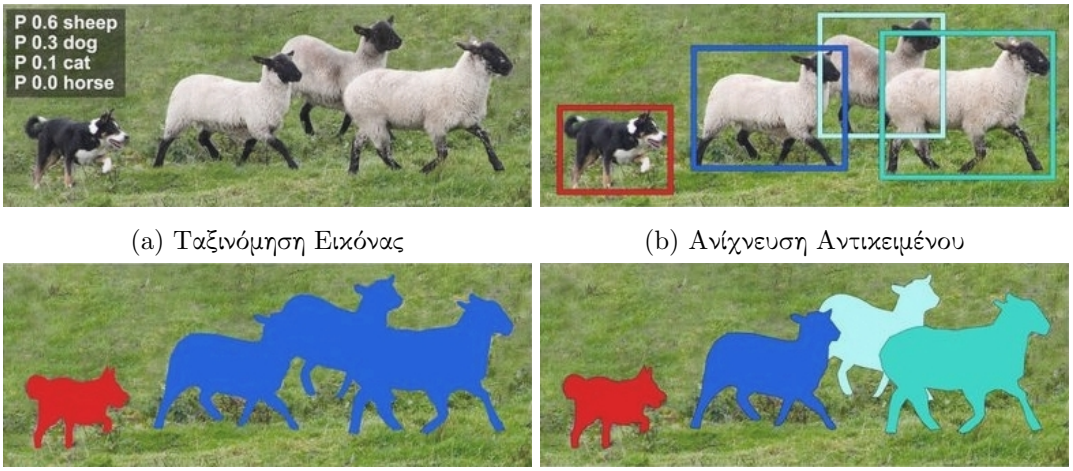


Σχήμα 3.11: Δομή FCN [28]

Η αρχιτεκτονική ακολουθεί το μοντέλο της κωδικοποίησης-αποκωδικοποίησης. Το τμήμα του κωδικοποιητή είναι υπεύθυνο για την εξαγωγή των χαρακτηριστικών και είναι εκπαιδευμένο με βάση το πρόβλημα της ταξινόμησης. Χρησιμοποιώντας συνελικτικά επίπεδα, προκαλεί μείωση της διάστασης της εικόνας (downsampling). Στη συνέχεια, το τμήμα του αποκωδικοποιητή αναλαμβάνει την επαναφορά της αρχικής ανάλυσης, δηλαδή την προβολή του χαμηλής ανάλυσης χάρτη χαρακτηριστικών που προέκυψε από τον κωδικοποιητή στην αρχική εικόνα. Ο αποκωδικοποιητής αποτελείται από μία σειρά αντίστροφες συνελίξεις (backwards convolutions) ή αποσυνελίξεις (deconvolutions), οι οποίες πραγματοποιούν αύξηση της χωρικής ανάλυσης με χρήση διγραμμικής παρεμβολής (bilinear interpolation). Ένα ακόμη σημαντικό χαρακτηριστικό του FCN είναι η χρήση παρακαμπτήριων συνδέσεων (skip connections), που εκμεταλλεύονται τις παρόμοιες διαστάσεις των εκατέρωθεν επιπέδων του FCN και συνδέουν σειριακά τους χάρτες ενεργοποίησης του κωδικοποιητή με την αντίστοιχή δομή που προκύπτει μετά από κάθε αποσυνέλιξη.

3.6 Πρόβλημα Κατάτμησης Στιγμιοτύπων

3.6.1 Περιγραφή



(a) Ταξινόμηση Εικόνας (b) Ανίχνευση Αντικειμένου

(c) Σημασιολογική Κατάτμηση (d) Κατάτμηση Στιγμιοτύπων

Σχήμα 3.12: Ταξινόμηση, Ανίχνευση Αντικειμένου και Κατάτμηση Εικόνας [8]

Το πρόβλημα της Κατάτμησης Στιγμιοτύπων αποτελεί τον συνδυασμό των δύο προηγούμενων προβλημάτων, της Ανίχνευσης Αντικειμένων και της Σημασιολογικής Κατάτμησης. Η Κατάτμηση Στιγμιοτύπων (Instance Segmentation) στοχεύει στον εντοπισμό των διαφορετικών αντικειμένων σε μία εικόνα όχι με χρήση πλαισίων οριοθέτησης, όπως στην περίπτωση της Ανίχνευσης Αντικειμένων, αλλά με ακρίβεια εικονοστοιχείου. Το κάθε εικονοστοιχείο, δηλαδή, θα ταξινομείται σε μία κλάση, όπως στη Σημασιολογική Κατάτμηση, αλλά τα διαφορετικά αντικείμενα θα έχουν άλλη μάσκα, ακόμα κι αν ανήκουν στην ίδια κλάση.

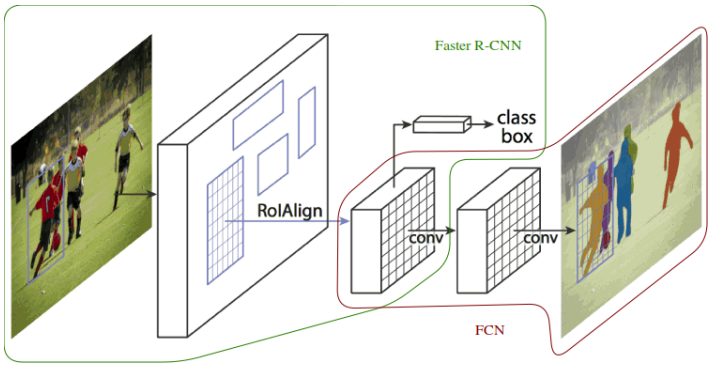
Όπως εύκολα γίνεται αντιληπτό, πρόκειται για ένα ιδιαίτερα απαιτητικό ζητούμενο, αφού πρέπει να λυθούν ταυτόχρονα και τα δύο προαναφερθέντα προβλήματα. Από τη δημοσίευση της πρώτης εργασίας που ασχολήθηκε εκτενώς με το θέμα το 2014 [29], η Κατάτμηση Στιγμιοτύπων έχει αποτελέσει αντικείμενο πολλών διαφορετικών προσεγγίσεων, οι περισσότερες εκ των οποίων βασίζονται σε Συνελικτικά Νευρωνικά Δίκτυα. [25] Μία από τις πιο πρόσφατες και υποσχόμενες, με βάση τα έως τώρα αποτελέσματα, είναι αυτή του Mask R-CNN.

3.6.2 Mask R-CNN

Όπως αναφέρθηκε στην προηγούμενη υποενότητα, από την πρόταση του μοντέλου R-CNN για την ανίχνευση αντικειμένων μέχρι σήμερα η βελτιστοποίησή του έχει απασχολήσει ένα σημαντικό αριθμό ερευνητικών εργασιών, με πιο σημαντικές αυτές των Fast-RCNN και Faster R-CNN. Αντί, όμως, η επόμενη σημαντική τροποποίηση του μοντέλου να αφορά για ακόμα μία φορά την ταχύτητά του, οδηγώντας πιθανότατα στη σύλληψη του Fastest ή του EvenFaster R-CNN, αυτή προέκυψε το 2017 από την ανάγκη προσαρμογής του στο πρόβλημα της Κατάτμησης Εικόνων. Συγκεκριμένα, η ερευνητική ομάδα της Facebook, Facebook AI Research (FAIR), δημιούργησε το Mask R-CNN επιχειρώντας να εκμεταλλευτεί την καλή απόδοση του

[8] https://towardsdatascience.com/detection-and-segmentation-through-convnets-47aa42de27ea

Faster R-CNN για τη δημιουργία ενός μοντέλου Κατάτμησης Στιγμιοτύπων [30].



Σχήμα 3.13: Δομή Mask R-CNN [30]

Για την προσαρμογή του Faster R-CNN στο νέο πρόβλημα, προφανώς χρειάστηκαν ορισμένες τροποποιήσεις στη δομή του. Αρχικά, προστέθηκε ένα επιπλέον τμήμα υπεύθυνο για την πρόβλεψη της μάσκας του κάθε αντικειμένου, παράλληλα με τα υπάρχοντα τμήματα της ταξινόμησης και της παλινδρόμησης του πλαισίου οριοθέτησης. Όπως και στην περίπτωση του Faster R-CNN, το πρώτο στάδιο του δικτύου αποτελείται από ένα δίκτυο RPN, για την πρόταση των υποψηφίων περιοχών. Στο Mask R-CNN, όμως, χρησιμοποιείται επιπλέον και ένα τμήμα για τον υπολογισμό των μασκών, αντίστοιχο με ένα Πλήρως Συνελικτικό Δίκτυο (FCN).

Μέρος ΙΙ

Προετοιμασία

Κεφάλαιο 4

Δημιουργία Συνόλου Δεδομένων

Για την επίτευξη του στόχου της παρούσας εργασίας απαραίτητη ήταν αρχικά η δημιουργία ενός συνόλου δεδομένων (dataset) που θα περιλαμβάνει επαρκή αριθμό δορυφορικών εικόνων, η κάθε μία εκ των οποίων θα συνοδεύεται από τις επιθυμητές εξόδους (ground truth) των σημείων που αντιστοιχούν σε κάποια γνωστή παραλία. Για τους λόγους που αναλύθηκαν στο Κεφάλαιο 2, καταλήξαμε ότι καταλληλότερες για τη συγκεκριμένη εφαρμογή είναι οι λήψεις της δορυφορικής αποστολής Sentinel-2 του προγράμματος Copernicus της ESA.

Πέραν από την επιλογή της πηγής των δορυφορικών εικόνων, ωστόσο, απαραίτητη ήταν η λήψη αποφάσεων και σχεδιαστικών επιλογών για μία σειρά ζητημάτων, τα οποία κινούνται σε τρεις βασικούς άξονες:

- Επιλογή κατάλληλου API για λήψη εικόνων: Οι εικόνες που προέρχονται από τον Sentinel-2 είναι δημόσια διαθέσιμες, και για το λόγο αυτό έχουν αναπτυχθεί διάφορες πλατφόρμες για τη λήψη τους, οι οποίες διαφοροποιούνται όσον αφορά την οργάνωση των δεδομένων, τις δυνατότητες που προσφέρουν αλλά και τη διεπαφή με το χρήστη.
- Οργάνωση των δεδομένων: Ο τρόπος με τον οποίο θα επιλέξουμε να οργανώσουμε τα δορυφορικά δεδομένα σε συγκεκριμένων διαστάσεων εικόνες θα διαδραματίσει καθοριστικό ρόλο στην επιτυχία της διαδικασίας της εκπαίδευσης του Νευρωνικού Δικτύου.
- Εύρεση ετικετών (ground truth): Απαραίτητη για τη δημιουργία του dataset είναι και η εύρεση όσο το δυνατόν ακριβέστερων δεδομένων σχετικά με την τοποθεσία των παραλιών. Επισήμως, δεν υπάρχει κάποια διαθέσιμη βάση που να διαθέτει τέτοιας φύσης πληροφορίες. Θα πρέπει, λοιπόν, να βρούμε κάποια εναλλακτική πηγή.

Στο κεφάλαιο αυτό θα εξεταστούν τα παραπάνω προβλήματα και θα αναλυθούν οι σχεδιαστικές επιλογές που πραγματοποιήθηκαν καθώς και οι λόγοι που οδήγησαν σε αυτές.

4.1 Συγκέντρωση Δορυφορικών Εικόνων

Η ελεύθερη διάθεση μεγάλου όγκου δορυφορικών εικόνων, σε συνδυασμό με το αυξημένο ενδιαφέρον του σύγχρονου επιστημονικού κόσμου για εφαρμογές σχετικές με την παρακολούθηση των περιβαλλοντικών φαινομένων και της ανθρώπινης δραστηριότητας στον πλανήτη, έχουν οδηγήσει στην ανάπτυξη πληθώρας εργαλείων που διευκολύνουν την επεξεργασία και τη λήψη δορυφορικών εικόνων.

Πέρα από την απευθείας λήψη από την σελίδα του Copernicus Programme, για την οποία απαιτείται μία δωρεάν εγγραφή (Copernicus Open Access Hub - SciHub Copernicus: `https:`

//scihub.copernicus.eu/), πλέον έχουν κάνει την εμφάνισή τους και πολλές άλλες υπηρεσίες οι οποίες προσφέρουν επιπλέον δυνατότητες. Για παράδειγμα, ιδιαίτερα χρήσιμες είναι οι διάφορες διαδικτυακές ιστοσελίδες που επιτρέπουν την αναζήτηση, προεπισκόπηση και στη συνέχεια λήψη δορυφορικών εικόνων και παραγώγων τους, με παγκόσμια κάλυψη και υψηλή χρονική ανάλυση, με πιο χαρακτηριστικά παραδείγματα τον Earth Explorer (`https://earthexplorer.usgs.gov/`) της Αρχής Γεωλογικών Ερευνών των ΗΠΑ (US Geological Survey), τον EO Browser (`https://apps.sentinel-hub.com/eo-browser/`) και τον Land Viewer (`https://eos.com/landviewer/`) του EOS (Earth Observing System).

Επίσης, αξίζει να αναφερθεί ότι εταιρίες που παρέχουν υποδομές νέφους (cloud computing), όπως η Amazon (Amazon Web Services - Open Data on AWS) και η Google (Google Cloud Storage) φιλοξενούν τα δεδομένα του Sentinel-2 και άλλων δορυφόρων και επιτρέπουν την επεξεργασία τους στο πλαίσιο των υπηρεσιών τους χωρίς να απαιτείται λήψη από το χρήστη. Το γεγονός αυτό διευκολύνει σε μεγάλο βαθμό την ανάπτυξη γεωχωρικών εφαρμογών, της οποίας η βασικότερη ίσως δυσκολία (ειδικά όταν αφορά δορυφορικές εικόνες υψηλής ανάλυσης) είναι η επεξεργασία του τεράστιου όγκου των δεδομένων που απαιτεί μεγάλη υπολογιστική ισχύ.

Για το λόγο αυτό, τα πιο διαδεδομένα πλέον εργαλεία λήψης δορυφορικών εικόνων είναι αυτά που επιτρέπουν, μέσω της κατάλληλης Διασύνδεσης Προγραμματισμού Εφαρμογών (Application Programming Interface - API), την επεξεργασία των εικόνων στους server κάποιου παρόχου, και τη συνέχεια τη λήψη του τελικού προϊόντος. Σε αυτή την κατηγορία ανήκει και το Google Earth Engine, που αποφασίσαμε να χρησιμοποιήσουμε στην συγκεκριμένη εφαρμογή.

4.1.1 Google Earth Engine

Το Earth Engine της Google είναι μία πλατφόρμα επιστημονικής ανάλυσης και οπτικοποίησης γεωχωρικών συνόλων δεδομένων μεγάλης κλίμακας, η οποία προορίζεται κυρίως για χρήση στο πλαίσιο ακαδημαϊκών, μη κερδοσκοπικών ή και κυβερνητικών προγραμμάτων. Στο δημόσιο αρχείο που διατηρεί φιλοξενούνται δορυφορικά δεδομένα από διαφορετικές πηγές και σε μεγάλο βάθος χρόνου, το οποίο συχνά φτάνει τα 40 έτη, ενώ ανανεώνεται με μεγάλη συχνότητα. Ταυτόχρονα, παρέχει ένα ιδιαίτερα ευέλικτο και εύχρηστο API (Διασύνδεση Προγραμματισμού Εφαρμογών) το οποίο μπορεί να ενσωματωθεί σε οποιαδήποτε προϋπάρχουσα εφαρμογή και επιτρέπει την άντληση δεδομένων από ένα μεγάλο εύρος πηγών, με απλό και εύκολο τρόπο.

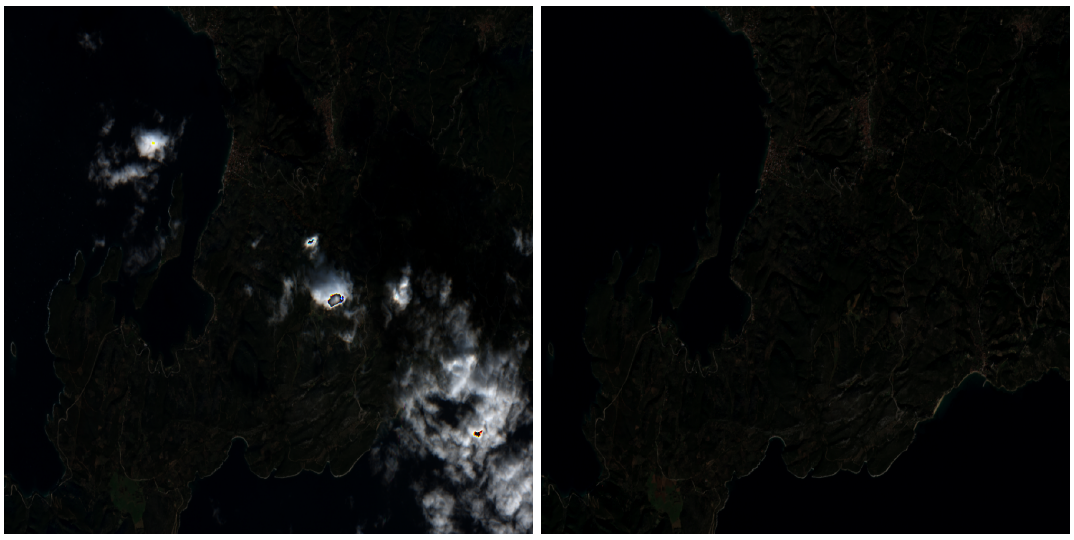
Το σημαντικότερο πλεονέκτημά του όμως, είναι ότι ταυτόχρονα παρέχει στο χρήστη τη δυνατότητα να ολοκληρώσει την αναζήτηση και την προεπεξεργασία των εικόνων πριν ξεκινήσει τη λήψη τους, απευθείας στους server του Google Cloud, με αποτέλεσμα να μειώνεται τόσο η υπολογιστική ισχύς που απαιτείται όσο και ο όγκος των δεδομένων που θα χρειαστεί να ληφθούν. Για παράδειγμα, στην περίπτωση της εφαρμογής που αναπτύσσουμε στα πλαίσια της παρούσας εργασίας, όπως θα δούμε στη συνέχεια, δεν είναι απαραίτητο να χρησιμοποιηθούν όλες οι ζώνες που παρέχουν οι μετρήσεις του Sentinel-2. Με τη χρήση του Google Earth Engine, η επιλογή των επιθυμητών ζωνών πραγματοποιείται πριν τη λήψη, με αποτέλεσμα ο όγκος των απαραίτητων δεδομένων να μειώνεται περίπου στο μισό.

Σημειώνεται ότι για να επιτραπεί η πρόσβαση στο Google Earth Engine χρειάζεται να γίνει ειδικό αίτημα πρόσβασης στο `https://signup.earthengine.google.com/`, το οποίο περιέχει τα στοιχεία του χρήστη και το γενικό σκοπό του project. Η επεξεργασία του αιτήματος απαιτεί περίπου μία ημέρα και στη συνέχεια η χρήση της υπηρεσίας είναι ελεύθερη, μόνο με ορισμένους

περιορισμούς που αφορούν κυρίως το μέγιστο μέγεθος της κάθε εικόνας.

4.1.2 Προεπεξεργασία των εικόνων

Το βασικότερο ίσως ζήτημα που οφείλει να διαχειριστεί η προεπεξεργασία των εικόνων είναι η εξασφάλιση της καταλληλότητάς τους. Για τις ανάγκες του συνόλου δεδομένων που επιχειρούμε να δημιουργήσουμε, είναι απαραίτητη μόνο μία δορυφορική εικόνα ανά περιοχή. Θα μπορούσαμε, λοιπόν, να ορίσουμε μία συγκεκριμένη περιοχή και ημερομηνία και να χρησιμοποιήσουμε την αντίστοιχη εικόνα. Η στρατηγική αυτή, όμως, έχει ένα πολύ βασικό μειονέκτημα: η ποιότητα ενός σχετικά μεγάλου ποσοστού των εικόνων του αρχείου του Sentinel-2 δεν ανταποκρίνεται στις ανάγκες της εφαρμογής μας, για δύο λόγους. Πρώτον, είναι εξαιρετικά συνηθισμένο τη στιγμή της λήψης της δορυφορικής εικόνας να παρεμβάλλονται σύννεφα, τα οποία προφανώς προκαλούν απώλεια πληροφορίας στα σημεία τα οποία καλύπτουν. Δεύτερον, για λόγους που αφορούν τη λειτουργία του αισθητήρα MSI, την τροχιά του δορυφόρου αλλά και τη ραδιομετρική διόρθωση και ευθυγράμμιση που υφίστανται τα δεδομένα για το σχηματισμό των τελικών εικόνων του επιπέδου L1C, κάποιες εικόνες δεν είναι ολοκληρωμένες, αλλά υπάρχουν περιοχές τους που λείπουν. Δεν είναι, λοιπόν, όλες οι εικόνες κατάλληλες για να συμπεριληφθούν στα δεδομένα.



(a) Δορυφορική εικόνα με σύννεφα (b) Δορυφορική εικόνα χωρίς σύννεφα

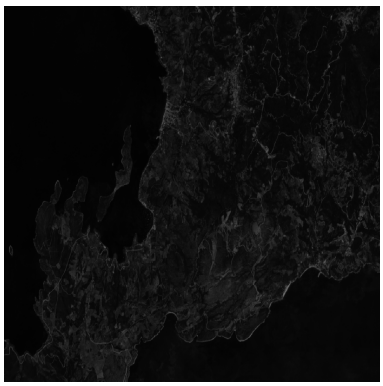
Σχήμα 4.1: Δορυφορικές εικόνες της ίδιας περιοχής με και χωρίς σύννεφα

Για την αντιμετώπιση των παραπάνω προβλημάτων, επιλέξαμε να συγκεντρώνουμε, για κάθε περιοχή, ένα σύνολο εικόνων που αφορούν επαρκές χρονικό διάστημα, και στη συνέχεια να κατεβάζουμε ως τελική εικόνα το διάμεσο (median) όλων των εικόνων του συνόλου αυτού. Εφόσον οι ατέλειες στις οποίες αναφερθήκαμε είναι τελείως τυχαίες, η επιλογή αρκετά μεγάλου χρονικού διαστήματος εγγυάται ότι η τελική εικόνα θα είναι αντιπροσωπευτική της περιοχής, χωρίς απώλεια πληροφορίας. Από την άλλη, βέβαια, αξίζει να σημειωθεί ότι το διάστημα αυτό δεν πρέπει να είναι υπερβολικά μεγάλο, ώστε να αποφευχθεί το ενδεχόμενο να παρατηρηθούν γεωλογικές αλλαγές στην περιοχή οι οποίες θα επηρεάσουν την τελική εικόνα.

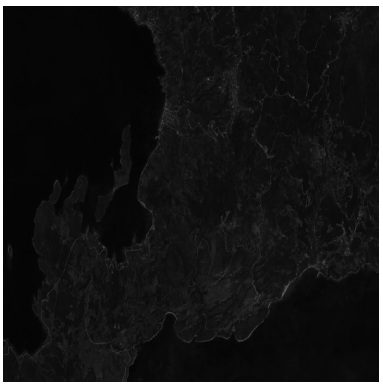
Το δεύτερο ζήτημα που αφορά την προεπεξεργασία των εικόνων είναι η επιλογή των κατάλληλων ζωνών (bands) τις οποίες θα χρησιμοποιήσουμε. Η επιλογή αυτή, όπως εύκολα γίνεται

αντιληπτό, υπαγορεύεται από το κατά πόσο η κάθε ζώνη περιλαμβάνει πληροφορία η οποία είναι χρήσιμη για την ανίχνευση της παραλίας και το διαχωρισμό της από την υπόλοιπη εικόνα.

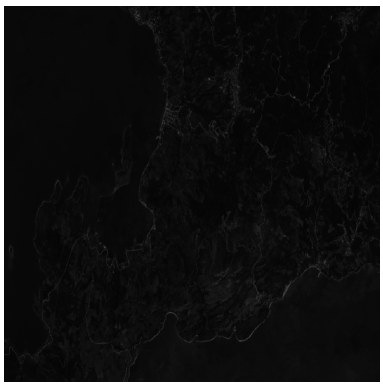
(a) Ζώνες R,G,B



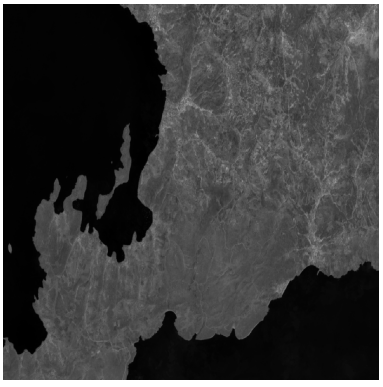
(b) Ζώνη R



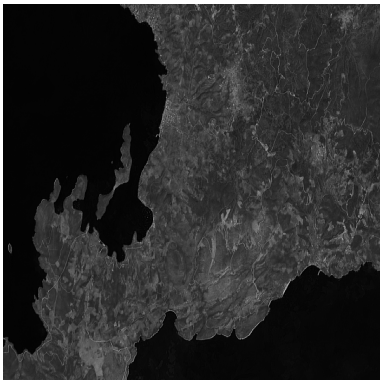
(c) Ζώνη G



(d) Ζώνη B



(e) Ζώνη NIR



(f) Ζώνη SWIR1

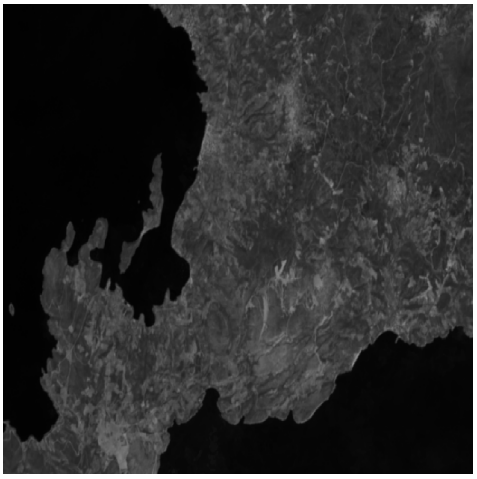
Σχήμα 4.2: Δορυφορικές ζώνες της ίδιας εικόνας

Προφανώς, οι γνωστές ζώνες του οπτικού φάσματος (R, G, B) είναι εξαιρετικά χρήσιμες για το στόχο αυτό. Από τις υπόλοιπες ζώνες, επιλέχθηκε να χρησιμοποιηθούν οι εξής:

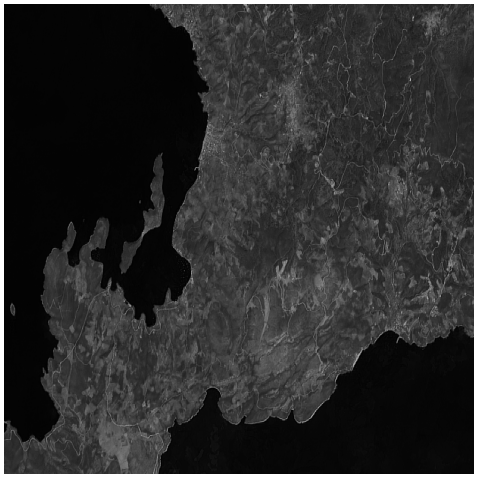
- NIR: Η ζώνη του εγγύς υπέρυθρου χρησιμοποιείται γενικά για την ανίχνευση σωμάτων

νερού και υδάτινων επιφανειών, καθώς η ακτινοβολία στη συγκεκριμένη ζώνη απορροφάται έντονα από το νερό, ενώ αντιθέτως ανακλάται έντονα από τη βλάστηση και τα γήινα πετρώματα [31]. Για το λόγο αυτό άλλωστε, μαζί με το πράσινο, χρησιμοποιείται για τον υπολογισμό του κατά McFeeters [31] κανονικοποιημένου δείκτη διαφοράς νερού (NDWI), ενός από τους πιο κοινούς τρόπους εύρεσης της ακτογραμμής και των ορίων των υδάτινων σωμάτων γενικότερα.

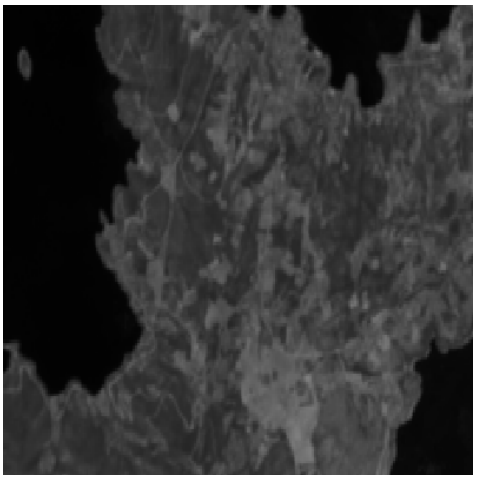
- SWIR1: Η ζώνη του υπέρυθρου βραχέων κυμάτων, από την άλλη, χρησιμοποιείται συχνά για τον υπολογισμό της υγρασίας του εδάφους, λόγω της ακρίβειας με την οποία μπορεί να προσδιορίσει το περιεχόμενο των διαφόρων υλικών σε νερό [32]. Κατά συνέπεια, προφανώς μπορεί να χρησιμεύσει για τη διάκριση του εδάφους της παραλίας από τις γύρω περιοχές, καθώς εκ φύσεως το περιεχόμενό τους σε υγρασία θα διαφέρει.



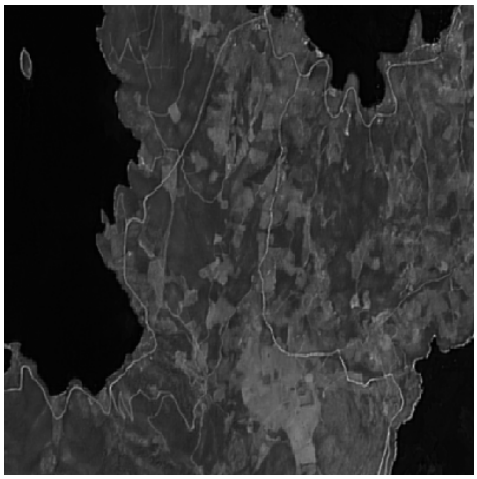
(a) Ζώνη SWIR1 πριν το pansharpening



(b) Ζώνη SWIR1 μετά το pansharpening



(c) Ζώνη SWIR1 πριν το pansharpening (σε μεγέθυνση)



(d) Ζώνη SWIR1 μετά το pansharpening (σε μεγέθυνση)

Σχήμα 4.3: Αποτελέσματα pansharpening

Τέλος, πριν την αποθήκευση των δορυφορικών εικόνων, πρέπει να ενοποιήσουμε τις 5 ζώνες (R, G, B, NIR, SWIR-1) που επιλέξαμε στο ίδιο αρχείο '.tiff'. Όπως είδαμε στην προηγούμενη ενότητα, όμως, οι ζώνες R, G, B, NIR έχουν χωρική ανάλυση 10m, ενώ η SWIR-1 έχει 20m. Κατά συνέπεια, πρέπει να προσαρμόσουμε την τελευταία ώστε να μπορεί να ενσωματωθεί στο

ίδιο αρχείο με τις υπόλοιπες.

Όπως προτείνεται στη σύγχρονη βιβλιογραφία [33, 34], προκειμένου να επιτευχθεί όσο το δυνατόν μεγαλύτερη ακρίβεια στη βελτίωση της ανάλυσης της εικόνας, χρησιμοποιήθηκε παγχρωματική όξυνση (pansharpening). Η διαδικασία της παγχρωματικής όξυνσης είναι μία μέθοδος που αποσκοπεί στην αύξηση της χωρικής ανάλυσης πολυφασματικών εικόνων εκμεταλλευόμενη τη χωρική πληροφορία από την υψηλής ανάλυσης παγχρωματική εικόνα. Είναι αλλιώς γνωστή και ως συγχώνευση αναλύσεων (resolution merge) ή ολοκλήρωση εικόνων (image integration). Στη συγκεκριμένη περίπτωση, εφόσον τα δεδομένα του Sentinel-2 δε διαθέτουν κάποια έτοιμη παγχρωματική ζώνη, χρησιμοποιήθηκε η μέση τιμή των ζωνών με ανάλυση 10m (B2, B3, B4, B8). Το αποτέλεσμα φαίνεται στην παραπάνω εικόνα.

4.1.3 Λήψη των εικόνων

Η λήψη των εικόνων πραγματοποιήθηκε με χρήση του Python API του Google Earth Engine [35], το οποίο παρέχεται και μέσω του Python Package Index (pip). Το πακέτο αυτό δίνει τη δυνατότητα να επιλεχθούν όλες οι διαθέσιμες δορυφορικές εικόνες από κάποια συλλογή (στην περίπτωσή μας αυτή του Sentinel-2), για συγκεκριμένη περιοχή (η οποία ορίζεται από τις αντίστοιχες συντεταγμένες), μεταξύ συγκεκριμένων ημερομηνιών. Επιπλέον, επιτρέπει την επεξεργασία του συνόλου αυτών των εικόνων πριν τη λήψη τους, όπως για παράδειγμα φιλτράρισμα ανάλογα με τις ιδιότητες των εικόνων, επιλογή των επιθυμητών μόνο ζωνών και εκτέλεση υπολογισμών στο σύνολό τους.

Για τη προεπεξεργασία που περιγράφηκε στην προηγούμενη ενότητα καθώς και τη λήψη των δορυφορικών εικόνων του Sentinel δημιουργήθηκε κατάλληλο εργαλείο γραμμένο σε python (το οποίο στο εξής θα ονομάζουμε tile_downloader) που δέχεται σαν ορίσματα:

- το πλαίσιο οριοθέτησης (bounding box) σε μορφή (ελάχιστο γεωγραφικό μήκος, ελάχιστο γεωγραφικό πλάτος, μέγιστο γεωγραφικό μήκος, μέγιστο γεωγραφικό πλάτος) της περιοχής στην οποία θέλουμε να αντιστοιχεί η εικόνα
- το χρονικό διάστημα στο οποίο επιθυμούμε να έχει πραγματοποιηθεί η λήψη της εικόνας, σε μορφή αρχικής ημερομηνίας - τελικής ημερομηνίας
- το όνομα που θα δοθεί στην εικόνα

και στη συνέχεια επεξεργάζεται τα δεδομένα ώστε να καταλήξει σε μία μόνο εικόνα ανά περιοχή, η οποία αποθηκεύεται σε μορφή .tiff.

Για τη δημιουργία των εικόνων του συνόλου δεδομένων, χρησιμοποιήθηκε το χρονικό διάστημα από την 01-06-2018 έως την 01-09-2018. Συνειδητά επιλέχθηκαν θερινοί μήνες, κατά τους οποίους είναι λιγότερο πιθανό έντονα καιρικά φαινόμενα να επηρεάσουν την ποιότητα των εικόνων. Ο τρόπος που επιλέχθηκαν οι διαστάσεις κάθε εικόνας και οι ακριβείς συντεταγμένες του περιβάλλοντος κουτιού αναλύονται στην Ενότητα 4.3.

Ο αλγόριθμος που εκτελείται από τον tile_downloader είναι ο εξής:

- Αρχικά, με βάση τα δεδομένα που δίνονται ως ορίσματα, επιλέγεται από το Google Earth Engine API μία συλλογή εικόνων (ImageCollection), η οποία περιέχει όλες τις διαθέσιμες εικόνες για τη συγκεκριμένη περιοχή, εντός των συγκεκριμένων ημερομηνιών.
- Οι εικόνες του Sentinel L1C περιλαμβάνουν μεταξύ άλλων στα μεταδεδομένα τους, την ιδιότητα cloudy_pixel_percentage, η οποία περιγράφει το ποσοστό των εικονοστοιχείων της εικόνας που αντιστοιχούν σε σύννεφα. Η ιδιότητα αυτή χρησιμοποιείται για να φιλ-

τράρει τις εικόνες, αφαιρώντας από τη συλλογή αυτές με ποσοστό μεγαλύτερο του 20%, ώστε η ποιότητα της τελικής εικόνας να είναι όσο το δυνατόν καλύτερη.

- Στη συνέχεια, επιλέγονται οι ζώνες της δορυφορικής εικόνας τις οποίες επιθυμούμε να χρησιμοποιήσουμε. Συγκεκριμένα, επιλέγονται οι ζώνες B2 (B), B3 (G), B4 (R), B8 (NIR), B11 (SWIR1).
- Πραγματοποιείται υπολογισμός του median της συλλογής.
- Πραγματοποιείται παγχρωματική όξυνση, και το σύνολο των επιλεγμένων ζωνών αποθηκεύεται σε ένα αρχείο '.tiff'.

Σε αυτό το σημείο είναι σημαντικό να γίνει κατανοητό ότι οι εικόνες που κατεβαίνουν και αποθηκεύονται με την παραπάνω διαδικασία δεν είναι κοινές δισδιάστατες εικόνες, αλλά πρόκειται για ειδικό τύπο αποθήκευσης και επεξεργασίας χωρικών δεδομένων που ονομάζονται δεδομένα κανάβου (raster data). Το βασικό χαρακτηριστικό του συγκεκριμένου τύπου δεδομένων που μας ενδιαφέρει στα πλαίσια της παρούσας εργασίας είναι ότι η γεωγραφική θέση του κάθε εικονοστοιχείου (γεωγραφικό μήκος/πλάτος) υποδηλώνεται από τη θέση του στον πίνακα (γραμμή/στήλη). Έτσι, η διαδικασία εντοπισμού μίας εικόνας στην επιφάνεια της Γης, αλλά και η ευθυγράμμιση με άλλες εικόνες του ίδιου τύπου απλοποιούνται σημαντικά.

4.2 Συγκέντρωση Ετικετών

Μετά την ολοκλήρωση της λήψης των δορυφορικών εικόνων, απαραίτητη είναι και η συγκέντρωση των αντίστοιχων ετικετών, οι οποίες να προσδιορίζουν τη θέση των παραλιών πάνω στην κάθε εικόνα. Μετά από διεξοδική αναζήτηση, καταλήξαμε ότι δεν υπάρχει κάποια επίσημη ευρωπαϊκή ή ελληνική πηγή, ινστιτούτο ή περιβαλλοντική οργάνωση που να παρέχει οργανωμένες και πλήρεις αυτές τις πληροφορίες. Για το λόγο αυτό, επιλέξαμε να αντλήσουμε δεδομένα από το OpenStreetMap, τα οποία αν και δεν είναι πλήρη και περιέχουν σε κάποιο βαθμό ανακρίβειες, είναι σίγουρα επαρκή ώστε να εκπαιδευτεί το Νευρωνικό Δίκτυο, οδηγώντας αργότερα σε μία πληρέστερη, πιο ακριβή και ολοκληρωμένη βάση δεδομένων.

4.2.1 OpenStreetMap (OSM)

Το OpenStreetMap ξεκίνησε το 2004 από το Ηνωμένο Βασίλειο και αποσκοπεί στη δημιουργία ενός δωρεάν, ελεύθερα προσβάσιμου και επεξεργάσιμου χάρτη ολόκληρου του κόσμου που χτίζεται σε μεγάλο βαθμό μέσω της συνεισφοράς εθελοντών και κυκλοφορεί με άδεια ανοικτού περιεχομένου. Ο χάρτη αυτός από τη στιγμή της δημιουργίας του δεν έχει σταματήσει να αναπτύσσεται και πλέον καλύπτει περιοχές σε όλον τον κόσμο, περιλαμβάνοντας τόσο γεωχωρικά δεδομένα και εικόνες όσο και ένα εξαιρετικά λεπτομερές σύστημα ετικετών που περιγράφουν τα χαρακτηριστικά του κάθε σημείου του χάρτη, είτε αφορούν το φυσικό περιβάλλον είτε ανθρώπινη δραστηριότητα. Τα χαρακτηριστικά αυτά συνοδεύουν το χάρτη ως metadata, και περιγράφονται με χρήση ζευγών κλειδιού-τιμής, συνδεδεμένων με κάποιο συγκεκριμένο σημείο του χάρτη. Για παράδειγμα, μία περιοχή στο χάρτη που αντιστοιχεί σε ένα δάσος θα συνοδεύεται, μεταξύ άλλων, από την ετικέτα (tag) "landuse=forest", ένα μουσείο από το "tourism=museum", ένας αυτοκινητόδρομος από το "highway=motorway" κοκ. Η ετικέτα που χρησιμοποιείται για να περιγράψει τις παραλίες είναι η "natural=beach".

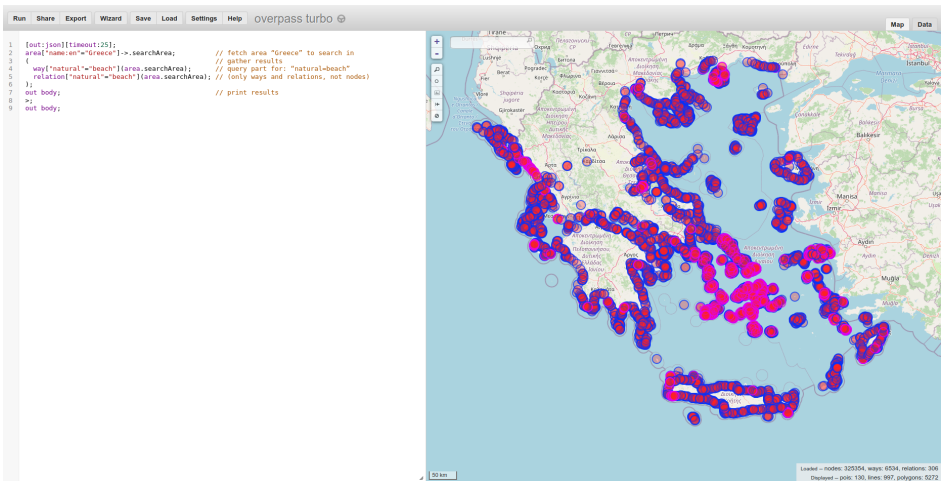
Για την άντληση δεδομένων και την αναζήτηση συγκεκριμένων tag στη βάση δεδομένων του OSM έχει αναπτυχθεί η διεπαφή Overpass API, η οποία φιλτράρει τα δεδομένα σύμφωνα με τα κριτήρια που δίνει ο χρήστης με μορφή αιτήματος (request) και επιστρέφει όλα τα αποτελέσματα που ταιριάζουν σε αυτά ως λίστα.

4.2.2 Λήψη Ετικετών

Προκειμένου να έχουμε όσο το δυνατόν περισσότερα διαθέσιμα δεδομένα, αποφασίστηκε να αντληθούν από το OpenStreetMap όλες οι διαθέσιμες τοποθεσίες παραλιών στην Ελλάδα. Για το σκοπό αυτό, δημιουργήθηκε ένα ακόμα εργαλείο σε μορφή python script, το οποίο στο εξής θα ονομάζεται osm_data_downloader και χρησιμοποιεί το overpy, τον python wrapper του Overpass API ο οποίος είναι διαθέσιμος μέσω του Python Package Index (pip). Συγκεκριμένα, το αίτημα που χρησιμοποιήθηκε για τη συλλογή του συνόλου των παραλιών της Ελλάδας φαίνεται παρακάτω. Συνολικά εντοπίστηκαν 5209 παραλίες. Στην εικόνα φαίνεται η έξοδος της διαδραστικής διαδικτυακής διεπαφής του Overpass API, Overpass Turbo, για ακριβώς το ίδιο αίτημα.

```
[out:json][timeout:25];
area[''name:en''=''Greece'']->.searchArea;          // fetch search area ''Greece''
(                                                   // gather results
  way[''natural''=''beach''](area.searchArea);      // query for ''natural=beach''
  relation[''natural''=''beach''](area.searchArea); // (ways and relations)
);
out body;                                           // print results
>;
out body;
```

Σχήμα 4.4: Αίτημα Overpass API

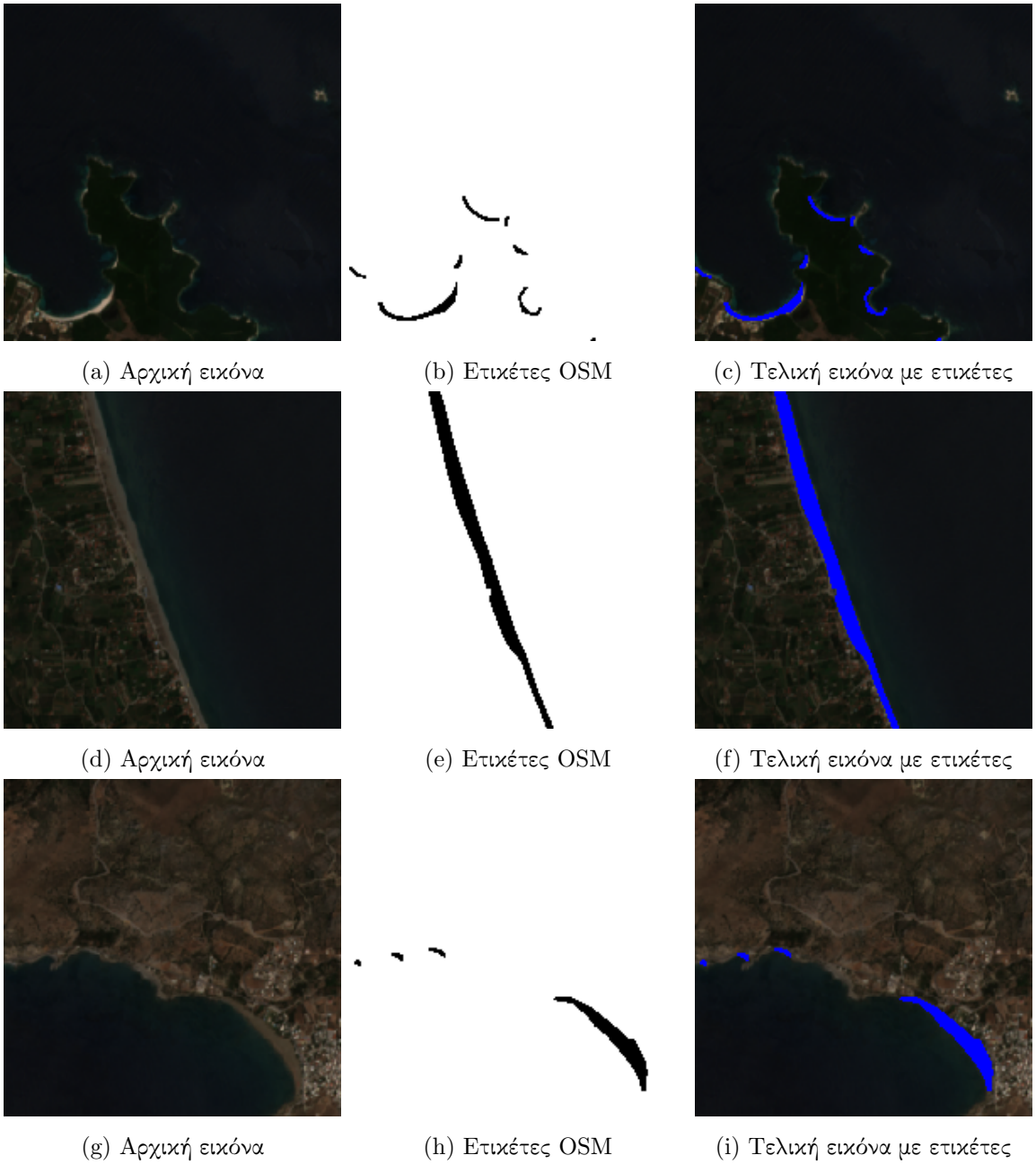


Σχήμα 4.5: Αποτελέσματα αναζήτησης παραλιών

Ο osm_data_downloader χρησιμοποιεί την παραπάνω σύνταξη ώστε να συγκεντρώσει τις

περιοχές όλων των παραλιών που βρίσκονται σε ελληνικές ακτές. Η κάθε παραλία αναπαριστάται από ένα πολύγωνο (polygon) που περιλαμβάνει τις συντεταγμένες του περιγράμματός της. Όλες οι περιοχές καταγράφονται σε ένα αρχείο '.GeoJSON', το οποίο θα αποτελέσει στο εξής τις ετικέτες του συνόλου δεδομένων μας.

Στην εικόνα φαίνονται κάποια παραδείγματα παραλιών, παράλληλα με την αντίστοιχη δορυφορική εικόνα.



(a) Αρχική εικόνα (b) Ετικέτες OSM (c) Τελική εικόνα με ετικέτες

(d) Αρχική εικόνα (e) Ετικέτες OSM (f) Τελική εικόνα με ετικέτες

(g) Αρχική εικόνα (h) Ετικέτες OSM (i) Τελική εικόνα με ετικέτες

Σχήμα 4.6: Εικόνες με ετικέτες

4.3 Οργάνωση Δεδομένων

Στις προηγούμενες υποενότητες αναλύθηκε ο τρόπος συλλογής των δορυφορικών εικόνων και των αντίστοιχων ετικετών. Στη συνέχεια, μένει να εξηγηθεί πώς τα δεδομένα αυτά οργανώθηκαν ώστε να δομήσουν με το βέλτιστο τρόπο το σύνολο δεδομένων που θα χρησιμοποιηθεί για την εκπαίδευση.

4.3.1 Μέγεθος εικόνων

Αρχικά, εξαιτίας της φύσης του προβλήματος, οι περιοχές που πρέπει να εντοπίζονται σε κάθε εικόνα είναι πολύ μικρές σε σχέση με το μέγεθός της. Ακόμα και στις περιπτώσεις που το μήκος μιας παραλίας είναι υπολογίσιμο, το πλάτος της δεν ξεπερνά ποτέ τα λίγα εικονοστοιχεία, δεδομένου ότι η χωρική ανάλυση των δορυφορικών εικόνων του Sentinel-2 είναι ίση με 10m. Το μικρό μέγεθος των αντικειμένων που θέλουμε να εντοπίσουμε αποτελεί γενικά παράγοντα που δυσχεραίνει την απόδοση του Νευρωνικού Δικτύου. Αυτό αποτελεί γνωστό πρόβλημα, το οποίο απασχολεί σημαντικά τη σύγχρονη βιβλιογραφία [36, 37], αφού σχετίζεται με την ίδια τη δομή των συνελικτικών δικτύων. Ένα τυπικό Συνελικτικό Νευρωνικό Δίκτυο δομείται από διαδοχικά συνελικτικά επίπεδα και επίπεδα pooling, τα οποία σταδιακά μειώνουν τις διαστάσεις των εικόνων εισόδου. Αυτό έχει σαν αποτέλεσμα σε κάθε κρυφό επίπεδο να χάνεται χωρική πληροφορία, και τα αντικείμενα που έχουν πολύ μικρές διαστάσεις συχνά να εξαφανίζονται και να μη φτάνουν μέχρι τα τελευταία επίπεδα.

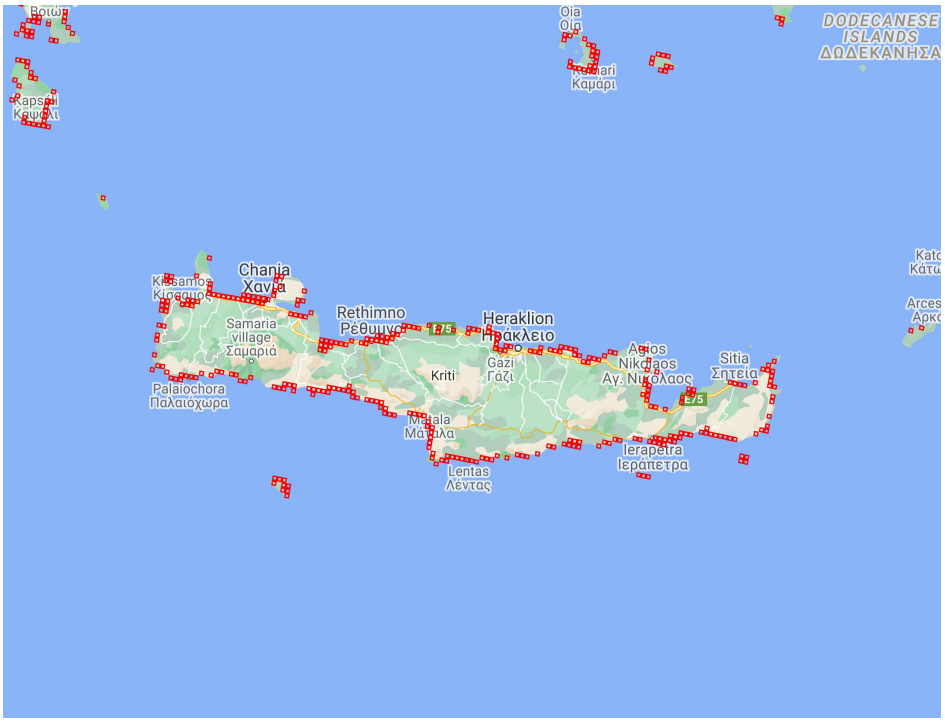
Ο βασικός τρόπος που προτείνεται προκειμένου να καταπολεμηθεί αυτή η τάση είναι η μείωση του μεγέθους των εικόνων. Όταν οι συνολικές διαστάσεις της εικόνας είναι μικρότερες, τα μικρά ή λεπτά αντικείμενα καταλαμβάνουν μεγαλύτερο μέρος της εικόνας, με αποτέλεσμα η απόδοση να βελτιώνεται. Για το λόγο αυτό, το μέγεθος των εικόνων που χρησιμοποιήθηκε στην εργασία επιλέχθηκε να είναι ίσο με 200*200 εικονοστοιχεία, ή 2km*2km σε πραγματικές διαστάσεις.

4.3.2 Δημιουργία Πλέγματος

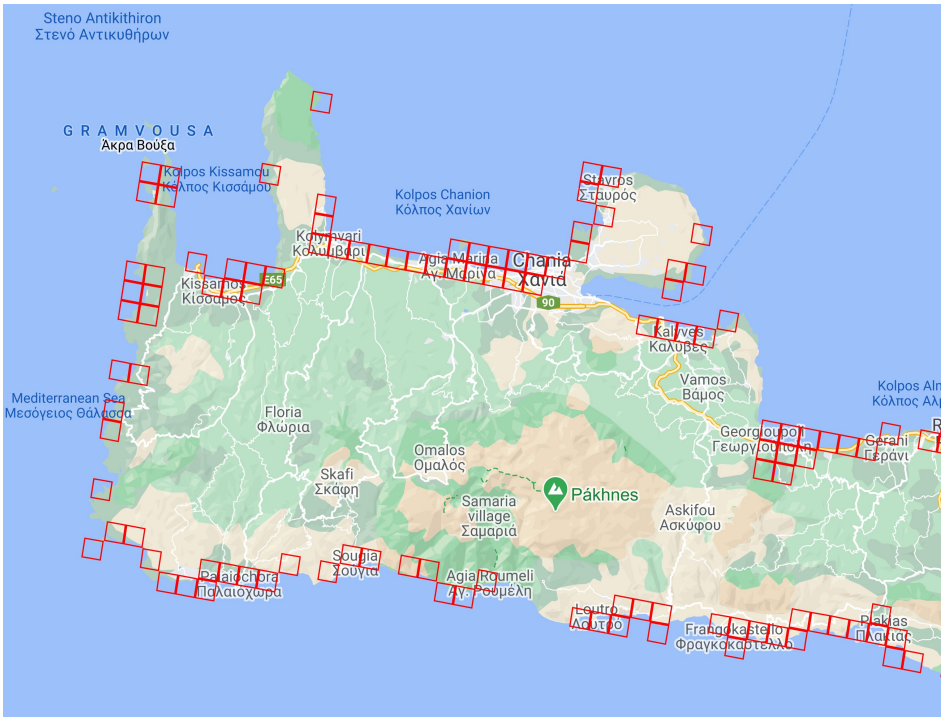
Με βάση τα παραπάνω, πρέπει ολόκληρη η επιφάνεια της Ελλάδας να χωριστεί σε δορυφορικές εικόνες μεγέθους 200*200 εικονοστοιχείων και να συμπεριληφθούν στα δεδομένα μας όσες από αυτές περιέχουν κάποια παραλία. Για να γίνει αυτή η κατάτμηση, χρειαζόμαστε ένα πλέγμα (grid) το οποίο να χωρίζει με κατάλληλο τρόπο την επιφάνεια της Ελλάδας ώστε να υπολογίσουμε τις ακριβείς συντεταγμένες κάθε εικόνας. Τέτοιου τύπου πλέγματα, γνωστά και ως Ευρωπαϊκά Πλέγματα Αναφοράς (European Reference Grids), διαθέτει στην ιστοσελίδα του (`https://esdac.jrc.ec.europa.eu/content/european-reference-grids`) το Ευρωπαϊκό Κέντρο Δεδομένων Εδάφους (European Soil Data Centre - ESDAC). Καθώς όμως τα μόνα διαθέσιμα πλέγματα έχουν διάσταση εικόνας 1km και 10km, χρειάστηκε να κατασκευάσουμε το νέο πλέγμα 2km με βάση αυτό του 1km, συνενώνοντας τα μικρότερα τμήματα ανά 4.

Στη συνέχεια, εκμεταλλευόμενοι το αρχείο '.GeoJSON' το οποίο περιλαμβάνει τις περιοχές όλων των ελληνικών παραλιών που θα χρησιμοποιήσουμε, επεξεργαζόμαστε το πλέγμα που κατασκευάσαμε κρατώντας μόνο τις περιοχές αυτές που έχουν κοινό έδαφος με κάποια παραλία.

Στο χάρτη που ακολουθεί φαίνονται οι περιοχές που παρέμειναν στο πλέγμα, δηλαδή αυτές που περιέχουν έστω και τμήμα μίας παραλίας, για το νησί της Κρήτης, καθώς το αντίστοιχο σχήμα για ολόκληρη την Ελλάδα ήταν αρκετά δυσδιάκριτο.



(a) Πλέγμα περιοχών



(b) Πλέγμα περιοχών σε μεγέθυνση

Σχήμα 4.7: Πλέγμα περιοχών με διαστάσεις 2km*2km

Μέρος ΙΙΙ

Πειραματική Διαδικασία

Κεφάλαιο 5

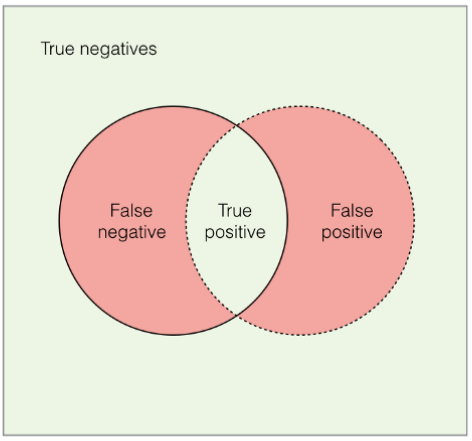
Μεθοδολογία

5.1 Μετρικές Αξιολόγησης Ανίχνευσης Αντικειμένων

Σε αυτή την υποενότητα θα αναλυθούν οι μετρικές που χρησιμοποιούνται στις εφαρμογές ανίχνευσης αντικειμένων και κατάτμησης στιγμιοτύπων, συμπεριλαμβανομένης της παρούσας εργασίας.

5.1.1 Precision και Recall

Ίσως η πιο χρήσιμη μετρική για το συγκεκριμένο πρόβλημα είναι η Μέση Ακρίβεια ή mean Average Precision (mAP), η οποία χρησιμοποιείται στην πλειοψηφία των σύγχρονων εφαρμογών ανίχνευσης αντικειμένων.



Σχήμα 5.1: Αναπαράσταση TP, TN, FP, FN [1]

Για να την ορίσουμε, πρέπει αρχικά να οριστούν οι έννοιες των True Positive, False Positive και False Negative, και στη συνέχεια του Precision και του Recall στα πλαίσια αυτού του τύπου προβλημάτων.

- Ως True Positive θεωρούμε την πρόβλεψη για την οποία το IoU μεταξύ του πλαισίου οριοθέτησης της πρόβλεψης και του πλαισίου οριοθέτησης του ground truth είναι μεγαλύτερο ή ίσο του προκαθορισμένου κατωφλίου iou_thres, ενώ ταυτόχρονα προβλέπεται σωστά η

[1] https://medium.com/@klintcho/explaining-precision-and-recall-c770eb9c69e9

κλάση του. Στο εξής ως TruePositive θα συμβολίζουμε τον αριθμό αυτών των περιπτώσεων σε κάποιο σύνολο δεδομένων.

- Ως False Positive θεωρούμε την πρόβλεψη για την οποία μία από τις δύο παραπάνω συνθήκες δεν ισχύει, δηλαδή είτε το πλαίσιο οριοθέτησης της πρόβλεψης έχει IoU με το πλαίσιο οριοθέτησης του ground truth μικρότερο του προκαθορισμένου κατωφλίου iou_thres, είτε η πρόβλεψη για την κλάση του αντικειμένου είναι λανθασμένη. Στο εξής ως FalsePositive θα συμβολίζουμε τον αριθμό αυτών των περιπτώσεων σε κάποιο σύνολο δεδομένων.
- Ως False Negative ορίζουμε τις περιπτώσεις στις οποίες δεν ανιχνεύτηκε ένα υπάρχον αντικείμενο. Στο εξής ως FalseNegative θα συμβολίζουμε τον αριθμό αυτών των περιπτώσεων σε κάποιο σύνολο δεδομένων.

Η Ακρίβεια (Precision) εκφράζει την ικανότητα του μοντέλου να ανιχνεύει μόνο τα σωστά αντικείμενα, και όχι τα λάθος [38]. Σύμφωνα με τα παραπάνω, ισούται με το ποσοστό των προβλέψεων που είναι σωστές, δηλαδή με το λόγο των σωστών προβλέψεων προς τις συνολικές προβλέψεις [39].

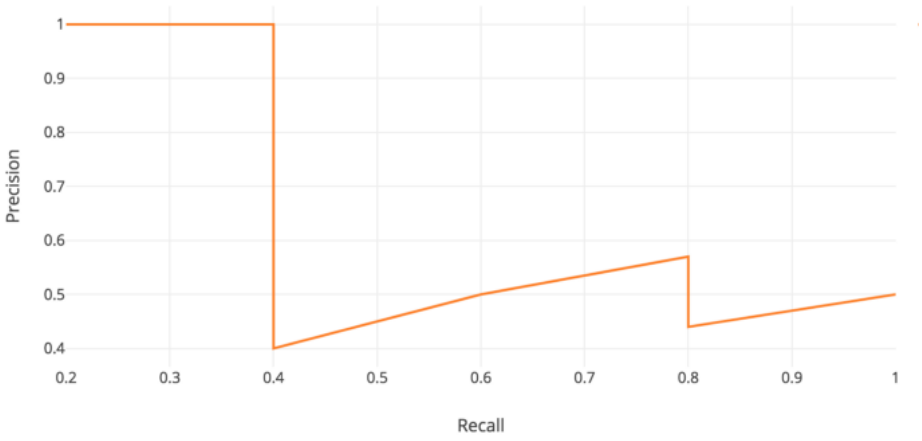
$$Precision = \frac{TruePositive}{TruePositive + FalsePositive}$$

Η Ανάκληση (Recall), από την άλλη, εκφράζει την ικανότητα του μοντέλου να ανιχνεύει όλα τα αντικείμενα [38]. Σύμφωνα με τα παραπάνω, ισούται με το ποσοστό των αντικειμένων που ανιχνεύθηκαν σωστά, δηλαδή με το λόγο των σωστών προβλέψεων προς το συνολικό αριθμό αντικειμένων [39].

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative}$$

5.1.2 Precision-Recall Curve

Η καμπύλη του Precision ως συνάρτηση του Recall χρησιμοποιείται ευρέως στην αξιολόγηση της ανίχνευσης αντικειμένων.



Σχήμα 5.2: Παράδειγμα καμπύλης Precision-Recall [39]

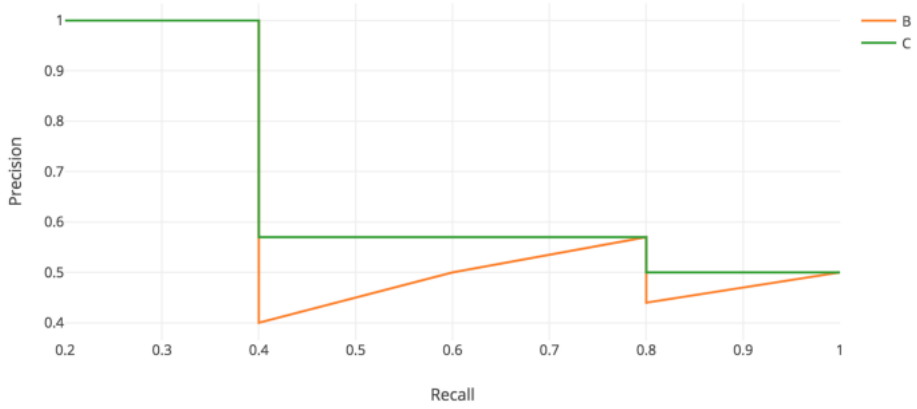
Κατά τη διάρκεια της εκπαίδευσης ενός μοντέλου ανίχνευσης αντικειμένων, ο αριθμός των True Positive προφανώς αυξάνεται, όσο εντοπίζονται νέα αντικείμενα, με αποτέλεσμα το Recall

να αυξάνεται (αύξουσα συνάρτηση). Ταυτόχρονα όμως, αυξάνεται και ο αριθμός των False Positive, γεγονός που μπορεί να οδηγήσει σε μείωση του Precision. Εάν λοιπόν τοποθετήσουμε τα ζεύγη (Precision, Recall) που παρατηρούνται κατά τη διάρκεια της εκπαίδευσης σε ένα διάγραμμα, προκύπτει μία καμπύλη όπως αυτή στο Σχήμα 5.2.

5.1.3 Average Precision (AP)

Η Μέση Ακρίβεια (Average Precision - AP) υπολογίζεται ως το εμβαδόν της επιφάνειας που βρίσκεται κάτω από την καμπύλη Precision-Recall, αφού πρώτα εξομαλυνθούν οι κάθετες διακυμάνσεις της (μετατραπεί δηλαδή σε φθίνουσα συνάρτηση χωρίς αυξομειώσεις). Η νέα καμπύλη φαίνεται στο ακόλουθο σχήμα και δίνεται από τον τύπο

$$P_{interp}(r) = max_{r' \leq r} p(r')$$



Σχήμα 5.3: Νέα καμπύλη Precision-Recall [39]

Ένας αρκετά συνηθισμένος τρόπος υπολογισμού της AP είναι αυτός που προτάθηκε στο PASCAL Visual Object Classes (VOC) [38], σύμφωνα με τον οποίο η AP ισούται με το μέσο όρο των τιμών του Precision για 11 τιμές του Recall: 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, δηλαδή

$$AP = \frac{1}{11} \sum_{r \in (0, 0.1, \ldots, 1)} AP_r$$

Ως mean Average Precision (mAP) ορίζεται ο μέσος όρος των τιμών του AP για όλες τις κλάσεις. Σε περίπτωση που υπάρχει μόνο μία κλάση, όπως στην παρούσα εργασία, ταυτίζεται με το AP.

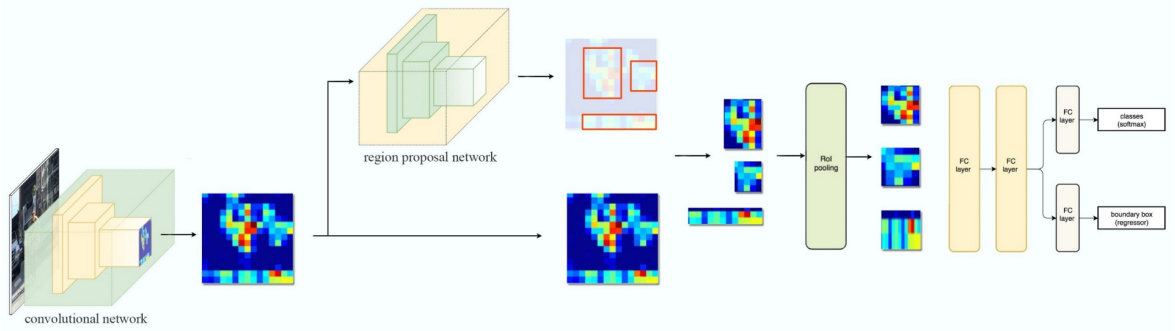
$$mAP = \frac{1}{N_{classes}} \sum_{1 \leq i \leq N_{classes}} AP(i)$$

Η χρήση της μετρικής mAP είναι ευρέως διαδεδομένη. Ενδεικτικά, ήταν η μετρική που χρησιμοποιήθηκε σε εξαιρετικά σημαντικούς διαγωνισμούς όπως οι PASCAL VOC [38], ImageNet [40] και Microsoft COCO (Common Objects in Context) [41].

5.2 Αρχιτεκτονική Mask R-CNN

Όπως αναφέρθηκε στην προηγούμενη ενότητα (Ενότητα 3.6.2), το Mask R-CNN προτάθηκε ως επέκταση του Faster R-CNN, και όπως και αυτό αποτελείται από δύο βασικά συστατικά:

- Αρχικά, το μοντέλο περιλαμβάνει ένα Δίκτυο Προτάσεων Περιοχής (Region Proposal Network - RPN). Το δίκτυο αυτό λαμβάνει ως είσοδο μία εικόνα και προτείνει υποψήφια πλαίσια οριοθέτησης όπου είναι πιθανό να βρίσκονται τα αντικείμενα (Σχήμα 5.6).
- Στη συνέχεια, τα προτεινόμενα πλαίσια οριοθέτησης δίνονται ως είσοδοι στο επόμενο στάδιο, το οποίο χρησιμοποιεί τα χαρακτηριστικά κάθε πλαισίου για την τελική πρόβλεψη.

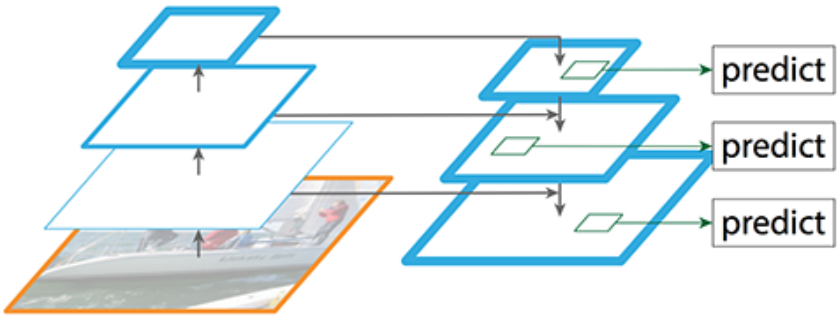


Σχήμα 5.4: Πλήρης αρχιτεκτονική Mask R-CNN [42]

5.2.1 Backbone

Ως είσοδος του Δικτύου Προτάσεων Περιοχής δεν δίνεται απευθείας η αρχική εικόνα. Όπως αναφέρθηκε (Ενότητα 3.4.3) το Fast R-CNN εισήγαγε πρώτο την ιδέα ενός αρχικού CNN το οποίο χρησιμοποιείται για την εξαγωγή του χάρτη χαρακτηριστικών ολόκληρης της εικόνας, σε αντίθεση με την εξαγωγή για κάθε πλαίσιο οριοθέτησης, ώστε να επιταχυνθούν οι υπολογισμοί του μοντέλου. Το μοντέλο αυτό, το οποίο αποτελεί το πρώτο συστατικό του Mask R-CNN, ονομάζεται backbone δίκτυο και ακολουθείται από το RPN και την κεφαλή του δικτύου.

Στην αρχική δημοσίευση [30], οι συγγραφείς πρότειναν τη χρήση ως backbone ενός ResNet [43] μαζί με ένα Δίκτυο Πυραμίδας Χαρακτηριστικών (Feature Pyramid Network - FPN), το οποίο επιτρέπει την εξαγωγή των χαρακτηριστικών σε διαφορετικές κλίμακες [44].



Σχήμα 5.5: Δίκτυο Πυραμίδας Χαρακτηριστικών [44]

Το FPN αποτελείται από ένα bottom-up και ένα top-down τμήμα, τα οποία συνδέονται μέσω

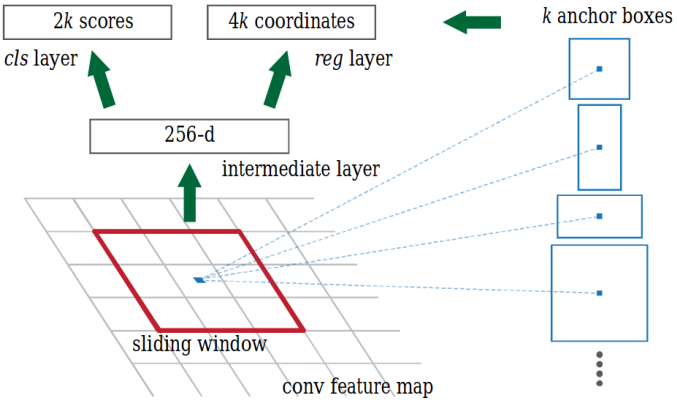
παρακαμπτηρίων συνδέσεων, όπως φαίνεται στο Σχήμα 5.5. Κατά μήκος της bottom-up διαδρομής, τα εξαγόμενα χαρακτηριστικά είναι υψηλού επιπέδου, λέμε δηλαδή ότι είναι σημασιολογικά σημαντικά, αλλά η χωρική ανάλυση είναι χαμηλή. Λόγω της συνθήκης αυτής, το μοντέλο αντιμετωπίζει δυσκολία στον εντοπισμό αντικειμένων το μέγεθος των οποίων είναι αρκετά μικρό ώστε να χάνονται στους χαμηλής ανάλυσης χάρτες [44].

Για το λόγο αυτό υλοποιείται το top-down τμήμα, με υπερδειγματοληψία (upsamling) των χαρτών χαρακτηριστικών υψηλού επιπέδου με χρήση της μεθόδου κοντινότερου γείτονα και παρακαμπτηρίων συνδέσεων [44]. Έτσι, τα χαρακτηριστικά που εξάγονται από το top-down τμήμα έχουν υψηλή χωρική ανάλυση, αν και σημασιολογικά είναι πιο ασήμαντα (χαμηλού επιπέδου).

Στη συνέχεια, το RPN παράγει τις περιοχές ενδιαφέροντος. Αναλόγως του μεγέθους μίας περιοχής ενδιαφέροντος, επιλέγουμε το επίπεδο του χάρτη χαρακτηριστικών που έχει την κατάλληλη κλίμακα.

5.2.2 Δίκτυο Προτάσεων Περιοχής

Μετά τον υπολογισμό του από το backbone δίκτυο, ο χάρτης χαρακτηριστικών δίνεται ως είσοδος στο Δίκτυο Προτάσεων Περιοχής (Region Proposal Network - RPN), ένα Πλήρως Συνελικτικό Δίκτυο, το οποίο χρησιμοποιεί κοινά συνελικτικά επίπεδα για τα δύο ζητούμενα, αυτό της παραγωγής προτάσεων περιοχών και αυτό της ανίχνευσης.



Σχήμα 5.6: Δίκτυο Προτάσεων Περιοχής [20]

Αρχικά, το RPN χρησιμοποιεί ένα συρόμενο παράθυρο για να σαρώσει τον χάρτη. Το κέντρο του παραθύρου αυτού σε κάθε θέση ονομάζεται άγκυρα, και για κάθε άγκυρα ορίζεται ένας προκαθορισμένος αριθμός (k) πλαισίων οριοθέτησης αναφοράς [30]. Τα πλαίσια αυτά είναι συνήθως 9, αφού προκύπτουν από όλους τους πιθανούς συνδυασμούς 3 προεπιλεγμένων μεγεθών και 3 προεπιλεγμένων αναλογιών (1:2, 1:1, 2:1), όπως φαίνεται στο Σχήμα 5.6.

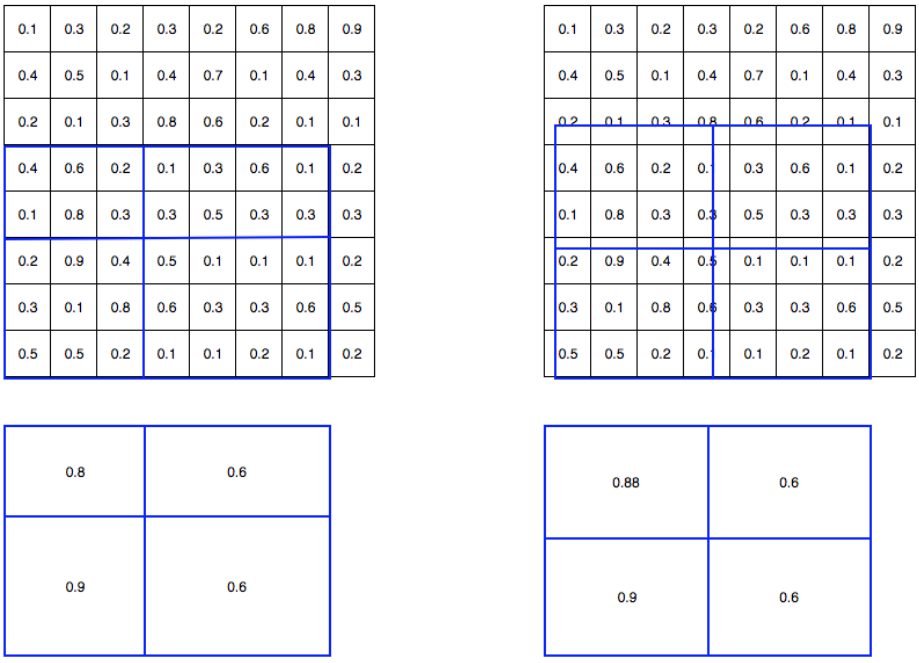
Όταν παραχθούν τα πλαίσια οριοθέτησης για όλες τις άγκυρες, το RPN υπολογίζει για κάθε ένα από αυτά δύο τιμές: την πιθανότητα να περιέχουν κάποιο αντικείμενο (ή, σύμφωνα με μία άλλη διατύπωση, να βρίσκονται στο προσκήνιο της εικόνας) και το ακριβές πλαίσιο οριοθέτησης, όπως προκύπτει από τη βελτίωση του πλαισίου αναφοράς μέσω παλινδρόμησης. Το RPN, δηλαδή, χρησιμοποιεί κοινά συνελικτικά επίπεδα από τα οποία προκύπτουν οι προτάσεις των περιοχών, αλλά δύο διαφορετικές κεφαλές (διαφορετικά τελικά πλήρως συνδεδεμένα επίπεδα) για τα δύο διαφορετικά προβλήματα της ταξινόμησης και της παλινδρόμησης πλαισίου.

Με βάση τις προβλέψεις αυτές του RPN, επιλέγουμε για κάθε αντικείμενο το πιο πιθανό πλαίσιο οριοθέτησης. Αφού επιλεχθούν όλα τα τελικά πλαίσια, εφαρμόζεται καταστολή μη μεγίστων (NMS) (Ενότητα 3.3.4) ώστε να αφαιρεθούν τα αλληλοεπικαλυπτόμενα αντικείμενα.

5.2.3 ROI align

Οι περιοχές ενδιαφέροντος που προκύπτουν από το RPN στη συνέχεια θα χρησιμοποιηθούν για την ταξινόμηση του αντικειμένου και τον υπολογισμό της μάσκας. Πριν συμβεί αυτό όμως, πρέπει να επιλεγεί για κάθε μία από αυτές το αντίστοιχο τμήμα του χάρτη χαρακτηριστικών. Όπως είναι αναμενόμενο, τα αντικείμενα σε μια εικόνα μπορεί να έχουν πολύ διαφορετικά μεγέθη, και κατά συνέπεια οι περιοχές ενδιαφέροντος που προκύπτουν από τα πλαίσια οριοθέτησης του RPN έχουν πολύ διαφορετικές διαστάσεις. Για να χρησιμοποιηθούν ως είσοδοι του τελικού τμήματος του Mask R-CNN, όμως, πρέπει να μετατραπούν σε περιοχές ίδιου μεγέθους [22].

Ο προκάτοχος του Mask R-CNN, το Faster R-CNN, χρησιμοποιεί για το λόγο αυτό ένα επίπεδο υποδειγματοληψίας περιοχών ενδιαφέροντος (ROI pooling), το οποίο μειώνει τη χωρική ανάλυση κάθε περιοχής ώστε να έχουν όλες το ίδιο μέγεθος [20]. Εάν όμως η αρχική διάσταση της περιοχής δεν είναι τέλεια διαιρέσιμη με την τελική, η μέθοδος αυτή δημιουργεί ένα πρόβλημα ευθυγράμμισης, όπως φαίνεται στο Σχήμα 5.7.



(a) ROI Pooling (b) ROI Align

Σχήμα 5.7: Σύγκριση ROI Pooling και ROI Align [42]

Όπως εύκολα γίνεται αντιληπτό, αν και η ανακρίβεια αυτή δεν ήταν καθοριστική για το πρόβλημα της ανίχνευσης αντικειμένων, προκαλεί σημαντική μείωση της ακρίβειας των μασκών που προβλέπονται κατά την κατάτμηση, η οποία πρέπει να γίνεται σε επίπεδο εικονοστοιχείου [30]. Για να αντιμετωπιστεί το πρόβλημα, οι συγγραφείς του Mask R-CNN πρότειναν τη χρήση ενός επιπέδου ευθυγράμμισης περιοχών ενδιαφέροντος (ROI Align) αντί της υποδειγματοληψίας (ROI Pooling) [30]. Συγκεκριμένα, το κάθε στοιχείο του χάρτη χαρακτηριστικών κβαντίζεται

με χρήση διγραμμικής παρεμβολής (συνήθως σε 4 υπο-στοιχεία), και λαμβάνει την τιμή που προκύπτει από τα υπο-στοιχεία αυτά, είτε ως μέσος όρος είτε ως μέγιστο.

5.2.4 Κεφαλή Δικτύου

Αφού τα πλαίσια οριοθέτησης του RPN χρησιμοποιηθούν για τη επιλογή των αντιστοίχων περιοχών του χάρτη χαρακτηριστικών, δίνονται ως είσοδος στο τελευταίο τμήμα του Mask R-CNN. Το τμήμα αυτό ονομάζεται και κεφαλή του δικτύου, και είναι αυτό που πραγματοποιεί την πρόβλεψη της κλάσης του αντικειμένου, του πλαισίου οριοθέτησης και της μάσκας του. Η πρόβλεψη της κλάσης (πρόβλημα ταξινόμησης) και του πλαισίου οριοθέτησης (πρόβλημα ανίχνευσης) πραγματοποιούνται με τον ίδιο τρόπο που πραγματοποιούνται και στα δίκτυα ανίχνευσης αντικειμένων (Ενότητα 3.4). Η πρόβλεψη της μάσκας πραγματοποιείται από ένα Πλήρως Συνελικτικό Δίκτυο (Ενότητα 3.5.2).

5.2.5 Συνάρτηση Κόστους

Η κατάτμηση στιγμιοτύπων (Ενότητα 3.6), για την οποία χρησιμοποιείται το Mask R-CNN, αποτελεί συνδυασμό τριών διαφορετικών προβλημάτων:

- ανίχνευση αντικειμένων στην εικόνα (Ενότητα 3.4)
- ταξινόμηση των αντικειμένων αυτών στη σωστή κλάση
- σημασιολογική κατάτμηση της εικόνας ώστε να προβλεφθεί με ακρίβεια εικονοστοιχείου η θέση των αντικειμένων (Ενότητα 3.5)

Κατά συνέπεια, η συνάρτηση κόστους που θα χρησιμοποιηθεί για την εκπαίδευση του δικτύου πρέπει να συνδυάζει τις συναρτήσεις κόστους των τριών διαφορετικών αυτών υποπροβλημάτων:

$$L = L_{cls} + L_{box} + L_{mask}$$

Πρόβλημα ταξινόμησης Η πρόβλεψη του δικτύου για την κλάση μίας περιοχής ενδιαφέροντος δίνεται ως μία διακριτή κατανομή πιθανότητας στο σύνολο των K κλάσεων [24],

$$p = (p_0, p_1, ..., p_K)$$

Για τον υπολογισμό της απώλειας του προβλήματος ταξινόμησης χρησιμοποιείται το λογαριθμικό σφάλμα για πραγματική κλάση (ground truth) u [24]:

$$L_{cls}(p, u) = -log(p_u)$$

Πρόβλημα παλινδρόμησης πλαισίου οριοθέτησης Η πρόβλεψη του δικτύου για το πλαίσιο οριοθέτησης μίας περιοχής ενδιαφέροντος δίνεται ως K διανύσματα αποκλίσεων, το κάθε ένα εκ των οποίων αντιστοιχεί στο πλαίσιο οριοθέτησης της αντίστοιχης κλάσης k [24],

$$t^k = (t^k_x, t^k_y, t^k_w, t^k_h)$$

Για τον υπολογισμό της απώλειας της παλινδρόμησης πλαισίου χρησιμοποιείται σφάλμα smooth-L1 για πραγματική κλάση (ground truth) u [24]:

$$L_{box}(t^u, v) = \sum_{i \in \{x,y,w,h\}} smooth_{L_1}(t^u_i - u_i)$$

όπου

$$smooth_{L_1}(x) = \begin{cases} 0.5x^2 & if|x| < 1 \\ |x| - 0.5 & otherwise \end{cases}$$

Πρόβλημα κατάτμησης Τέλος, η πρόβλεψη του δικτύου για τη μάσκα ενός αντικειμένου δίνεται ως K δυαδικές μάσκες $\hat{y}^k$ με διάσταση $m * m$, μία για κάθε κλάση k [30].

Για τον υπολογισμό της απώλειας της κατάτμησης χρησιμοποιείται το μέσο cross-entropy σφάλμα [30]:

$$L_{mask}(y, \hat{y}^k) = -\frac{1}{m^2} \sum_{1 \leq i,j \leq m} [y_{ij} log \hat{y}_{ij}^k + (1 - y_{ij}) log(1 - \hat{y}_{ij}^k)$$

5.3 Αρχιτεκτονική Rotated Mask R-CNN

5.3.1 Εισαγωγή

Ζητούμενο της συγκεκριμένης εργασίας αποτελεί ο ακριβής εντοπισμός των παραλιών σε δορυφορικές εικόνες προερχόμενες από τον Sentinel 2, οι οποίες έχουν χωρική ανάλυση 10 μέτρων. Αυτό έχει σαν αποτέλεσμα στην πλειοψηφία των περιπτώσεων οι περιοχές που πρέπει να επιλεχθούν να έχουν σχήμα μακρόστενο και πάχος ελάχιστων εικονοστοιχείων (λίγες παραλίες στην Ελλάδα έχουν πλάτος άνω των 30 μέτρων, που αντιστοιχούν σε 3 εικονοστοιχεία).

Ο εντοπισμός αντικειμένων με τέτοιο σχήμα αποτελεί ένα από τα δυσκολότερα προβλήματα στα πλαίσια της κατάτμησης εικόνων σε πραγματικό χρόνο, το οποίο ακόμα δεν έχει επιλυθεί. Χαρακτηριστικές εφαρμογές στις οποίες εμφανίζεται αυτό το πρόβλημα είναι η αποφυγή των ηλεκτροφόρων καλωδίων κατά την πτήση των μη επανδρωμένων αεροσκαφών (drones) [45], η ανίχνευση λεπτών δομών σε ιατρικές εικόνες [46] καθώς και ο εντοπισμός ρωγμών σε δομικές επιφάνειες [47].

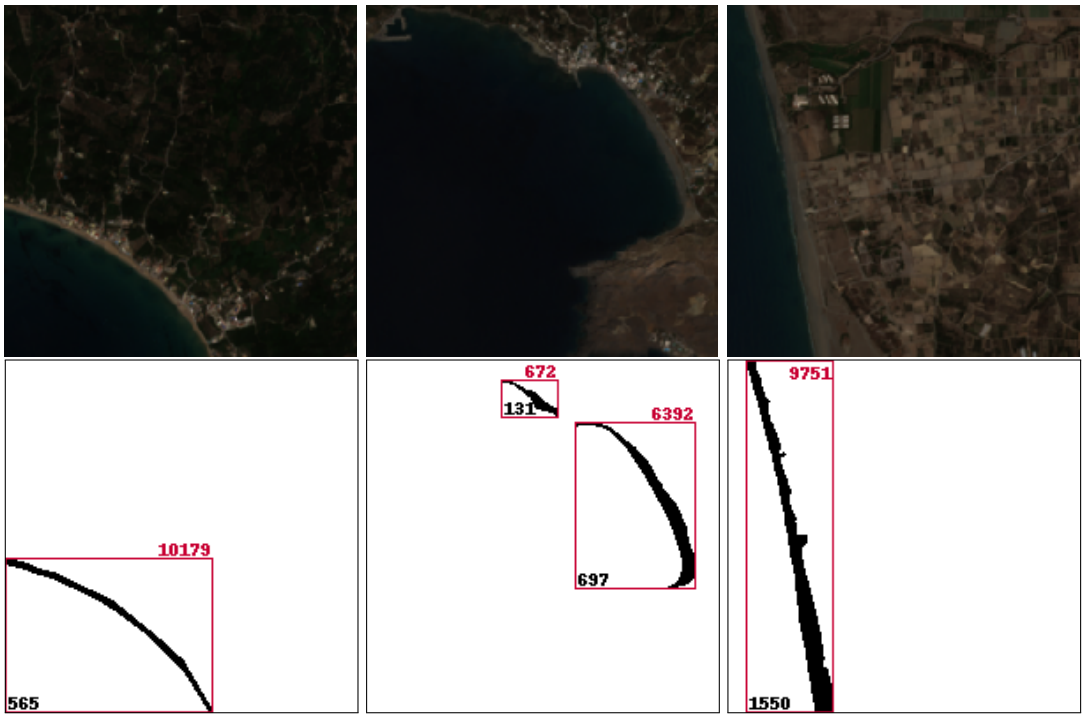
Στο [48] οι Feng et al. ορίζουν το Λόγο Αντικειμένου προς Περιοχή (Object-Region Percentage ORP) ως

$$ORP = \frac{\text{Εμβαδόν Αντικειμένου}}{\text{Εμβαδόν Ελάχιστου Πλαισίου Οριοθέτησης}}$$

Στην περίπτωση που ένα αντικείμενο έχει μακρόστενο σχήμα, μικρό πλάτος και κλίση σε σχέση με τον οριζόντιο και τον κάθετο άξονα, το πλαίσιο οριοθέτησής του θα καλύπτει δυσανάλογα μεγάλη επιφάνεια, και κατά συνέπεια οι αντίστοιχες τιμές του ORP θα είναι εξαιρετικά μικρές, όπως φαίνεται στο Σχήμα 5.8. Ακόμα και στην τρίτη περίπτωση, όπου η κλίση δεν είναι μεγάλη, το ποσοστό ελάχιστα ξεπερνά το 15%.

Όπως έδειξαν οι Feng et al. στο [48], τα αντικείμενα με χαμηλές τιμές ORP αποτελούν πρόκληση για μοντέλα ανίχνευσης αντικειμένων όπως το Faster R-CNN. Λόγω του σχήματός τους, έχουν ιδιαίτερα χαμηλό σηματοθορυβικό λόγο (SNR) με αποτέλεσμα η περιττή πληροφορία στο παρασκήνιο (background) του πλαισίου οριοθέτησης να κάνει την πρόβλεψη εξαιρετικά δύσκολη. Για να επαληθεύσουν την υπόθεσή τους, χρησιμοποιώντας το σύνολο δεδομένων COCO (Common Objects in Context) [41] υπολόγισαν για κάθε αντικείμενο την αντίστοιχη τιμή του ORP και αξιολόγησαν την επίδοση του Faster R-CNN για διάφορα εύρη των τιμών αυτών. Με τον τρόπο αυτό έδειξαν ότι όταν η τιμή του ORP είναι χαμηλότερη του 30% η απόδοση του

δικτύου μειώνεται κατά περίπου 30%, ενώ για χαμηλότερες τιμές η μείωση είναι δραματική (πάνω από 60% για ORP μικρότερο του 10%).



(a) ORP = 5.5% (b) ORP = 19.5%, 10.9% (c) ORP = 15.89%

Σχήμα 5.8: Λόγος Αντικειμένου/Περιοχής (ORP)

Τα παραπάνω είναι εμφανές ότι θα επηρεάσουν την απόδοση του Mask R-CNN που θα χρησιμοποιήσουμε, αφού εξ ορισμού τα αντικείμενα στο σύνολο δεδομένων μας έχουν πολύ χαμηλό ORP. Για το λόγο αυτό, αποφασίσαμε, εκτός από την εκπαίδευση του απλού Mask R-CNN, να συμπεριλάβουμε στην εργασία και μία τροποποίηση του μοντέλου, που θα επιχειρεί να ελαττώσει την επίδραση του σχήματος των αντικειμένων στα αποτελέσματα. Συγκεκριμένα, επιλέξαμε να συμπεριλάβουμε και το Rotated Mask R-CNN, όπως αυτό υλοποιήθηκε από τον S. Looi [49], το οποίο αντικαθιστά τα παραδοσιακά πλαίσια οριοθέτησης με πλαίσια οριοθέτησης με περιστροφή (Rotated Bounding Box), όπως περιγράφονται στο [50]. Η βελτίωση της ποιότητας της πρόβλεψης είναι εμφανής στο Σχήμα 5.9.

Σχήμα 5.9: Παράδειγμα Rotated Mask R-CNN [49]

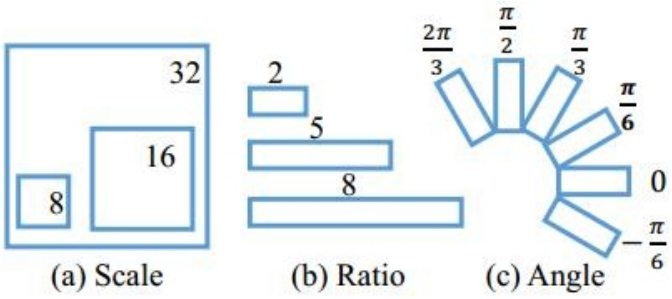
5.3.2 Δίκτυο Προτάσεων Περιοχής με Περιστροφή

Η δομή του Rotated Mask R-CNN είναι γενικά ίδια με αυτήν του Mask R-CNN, με εξαίρεση προφανώς το τμήμα του Δικτύου Προτάσεων Περιοχής, το οποίο πρέπει να τροποποιηθεί ώστε να παράγει περιοχές ενδιαφέροντος με περιστροφή (Rotated Regions of Interest - RROI). Το Δίκτυο Προτάσεων Περιοχής με Περιστροφή (Rotated Region Proposal Network - RRPN) προτάθηκε από τους J. Ma et al. στο [50] το 2018 για τη βελτίωση της ανίχνευσης συμβολοσειρών (που επίσης έχουν χαμηλό ORP) σε φωτογραφίες.

Η κάθε περιοχή ενδιαφέροντος με περιστροφή (Rotated Region of Interest) ορίζεται από ένα πλαίσιο ενδιαφέροντος με περιστροφή (Rotated Bounding Box). Το πλαίσιο ενδιαφέροντος με περιστροφή ορίζεται από 5 τιμές, οι οποίες μπορούν να είναι ενδεικτικά:

- οι συντεταγμένες (i_0, j_0) της κάτω αριστερής και (i_1, j_1) της πάνω δεξιάς γωνίας του και η γωνία προσανατολισμού ϕ
- οι συντεταγμένες (i_0, j_0) της κάτω αριστερής γωνίας, το πλάτος w, το ύψος h και η γωνία προσανατολισμού ϕ
- οι συντεταγμένες (i_c, j_c) του κέντρου, το πλάτος w, το ύψος h και η γωνία προσανατολισμού ϕ

Όπως και στην περίπτωση του RPN, το RRPN χρησιμοποιεί άγκυρες σε κάθε θέση του χάρτη χαρακτηριστικών ώστε να παράγει τα πιθανά πλαίσια οριοθέτησης. Η διαφορά μεταξύ τους έγκειται στο ότι στην περίπτωση του RRPN οι άγκυρες δε διαφέρουν μεταξύ τους μόνο ως προς το μέγεθος και την αναλογία, αλλά και ως προς τη γωνία προσανατολισμού τους, όπως φαίνεται στο σχήμα Σχήμα 5.10



Σχήμα 5.10: Άγκυρες RRPN [50]

Όπως και στο RPN, μετά την παραγωγή των πλαισίων οριοθέτησης με περιστροφή για όλες τις άγκυρες, το RRPN υπολογίζει για κάθε ένα από αυτά την πιθανότητα να βρίσκεται στο προσκήνιο της εικόνας (να περιέχει κάποιο αντικείμενο) και μέσω παλινδρόμησης (regression) υπολογίζει το βέλτιστο πλαίσιο οριοθέτησης. Η γωνία προσανατολισμού συμμετέχει κανονικά στην παλινδρόμηση του πλαισίου, ως μία ακόμα παράμετρος περιγραφής του (Ενότητα 5.3.3).

5.3.3 Συνάρτηση Κόστους

Η συνάρτηση κόστους που θα χρησιμοποιηθεί για την εκπαίδευση του Rotated Mask R-CNN, όπως και του απλού Mask R-CNN (Ενότητα 5.2.5), συνδυάζει τις συναρτήσεις κόστους των τριών υποπροβλημάτων που απαρτίζουν την κατάτμηση στιγμιοτύπων: της ταξινόμησης, της

επιλογής πλαισίου οριοθέτησης και της κατάτμησης της εικόνας:

$$L = L_{cls} + L_{box} + L_{mask}$$

Το τμήμα του Rotated Mask R-CNN που αφορά την ταξινόμηση και τον υπολογισμό της μάσκας ταυτίζεται με αυτό του Mask R-CNN, οπότε τα L_{cls}, L_{mask} ορίζονται ακριβώς όπως στην Ενότητα 5.2.5.

Από την άλλη, η συνάρτηση σφάλματος της παλινδρόμησης του πλαισίου οριοθέτησης πρέπει να προσαρμοστεί ώστε να συμπεριλάβει και τη γωνία προσανατολισμού του πλαισίου. Η πρόβλεψη του πλαισίου οριοθέτησης μίας περιοχής ενδιαφέροντος δίνεται ως K διανύσματα αποκλίσεων, το κάθε ένα εκ των οποίων αντιστοιχεί στο πλαίσιο οριοθέτησης με περιστροφή της αντίστοιχης κλάσης k [50],

$$t^k = (t^k_x, t^k_y, t^k_w, t^k_h, t^k_r)$$

Για τον υπολογισμό της απώλειας της παλινδρόμησης πλαισίου χρησιμοποιείται σφάλμα smooth-L1 για πραγματική κλάση (ground truth) u [50]:

$$L_{box}(t^u, v) = \sum_{i \in \{x,y,w,h,r\}} smooth_{L_1}(t^u_i - u_i)$$

όπου

$$smooth_{L_1}(x) = \begin{cases} 0.5x^2 & if|x| < 1 \\ |x| - 0.5 & otherwise \end{cases}$$

Κεφάλαιο 6

Εκπαίδευση και Αξιολόγηση των Μοντέλων

6.1 Πειραματική Διάταξη

6.1.1 Hardware

Η εκτέλεση των πειραμάτων πραγματοποιήθηκε στους servers του Εργαστηρίου Συστημάτων Τεχνητής Νοημοσύνης και Μάθησης της Σχολής Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών, στο Εθνικό Μετσόβιο Πολυτεχνείο. Συγκεκριμένα, ο server που χρησιμοποιήθηκε διαθέτει 2 Μονάδες Επεξεργασίας Γραφικών (GPU) Nvidia GeForce GTX 1080, η κάθε μία εκ των οποίων διαθέτει μνήμη 8 GB.

6.1.2 Λογισμικό

Mask R-CNN Η υλοποίηση που χρησιμοποιήθηκε για το Mask R-CNN προέρχεται από το δημόσιο GitHub repository της Facebook AI Research (FAIR) [51], και αποτελεί μέρος του detectron [52], του λογισμικού της που ενσωματώνει state-of-the-art αλγορίθμους ανίχνευσης αντικειμένων. Όλος ο κώδικας είναι γραμμένος σε python3, ενώ για την υλοποίηση του μοντέλου Mask R-CNN χρησιμοποιήθηκε η βιβλιοθήκη PyTorch 1.0.

Το μοντέλο της FAIR προορίζεται για την ανίχνευση αντικειμένων σε φωτογραφικές εικόνες, οι οποίες είναι γενικά είτε Grayscale είτε RGB, διαθέτουν δηλαδή μία ή τρεις ζώνες. Για τη χρήση του μοντέλου με τις δορυφορικές εικόνες του συνόλου δεδομένων που κατασκευάσαμε (Κεφάλαιο 4), οι οποίες διαθέτουν 5 ζώνες, χρειάστηκε να γίνουν αρκετές προσαρμογές στον κώδικα.

Rotated Mask R-CNN Η υλοποίηση του Rotated Mask R-CNN που χρησιμοποιήθηκε είναι γραμμένη από τον S. Looi [49]. Είναι βασισμένη στο Mask R-CNN της Facebook AI Research (FAIR) [51], και μεγάλο μέρος του κώδικα είναι κοινό. Οι τροποποιήσεις που έγιναν σχετικά με τη διάσταση των εικόνων προφανώς ισχύουν και σε αυτή την εκδοχή του μοντέλου.

6.1.3 Παραμετροποίηση μοντέλου

Τα πλήρη αρχεία παραμετροποίησης των δύο μοντέλων, τα οποία περιέχουν αναλυτικές πληροφορίες για τη δομή του κάθε τμήματός τους, βρίσκονται στο παράρτημα (Appendix A). Τα σημεία που αξίζει να σημειωθούν είναι:

Δεδομένα

Οι δορυφορικές εικόνες του συνόλου δεδομένων μας (Κεφάλαιο 4) έχουν διάσταση $200 * 200 * 5$. Πριν δοθούν σαν είσοδος στο μοντέλο, κανονικοποιούνται ανά batch και η διάστασή τους μετατρέπεται σε $400 * 400 * 5$. Επίσης, πραγματοποιείται αύξηση των δεδομένων (Data Augmentation), με οριζόντια και κάθετα flip της εικόνας με πιθανότητα 50%.

Backbone δίκτυο

Ως backbone δίκτυο χρησιμοποιήθηκε το ResNet-50-FPN. Αν και υπήρχε διαθέσιμο το προεκπαιδευμένο δίκτυο, οι δυνατότητες εκμετάλλευσής του για Μεταφορά Μάθησης ήταν περιορισμένες, καθώς τα δεδομένα διαθέτουν 5 ζώνες, αντίθετα με τις RGB εικόνες στις οποίες έχει εκπαιδευτεί. Παρόλα αυτά, για τις ζώνες RGB χρησιμοποιήθηκαν τα προεκπαιδευμένα βάρη.

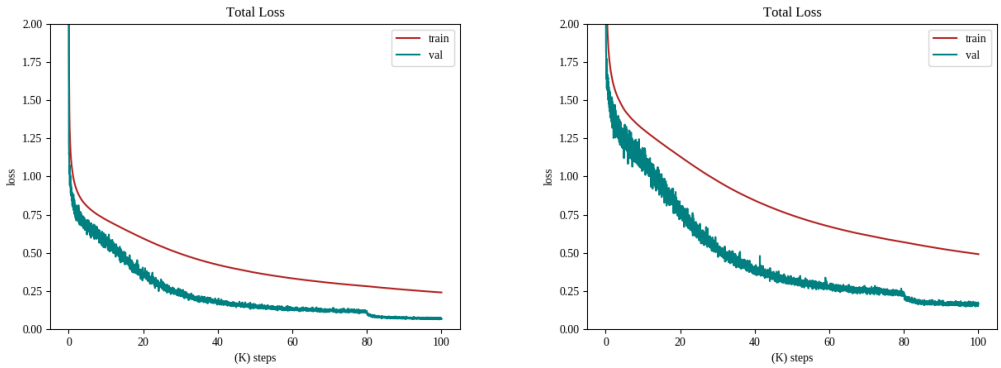
Άγκυρες

Για την πρόταση περιοχών από το RPN χρησιμοποιήθηκαν 15 διαφορετικές άγκυρες, με μεγέθη 32, 64, 128, 256 και 512 εικονοστοιχείων και αναλογίες 1:2, 1:1 και 2:1. Για τις προτάσεις του RRPN χρησιμοποιήθηκαν 45 άγκυρες, λόγω των 3 γωνιών προσανατολισμού $-30°$, $-60°$, $-90°$ που χρησιμοποιήθηκαν.

6.2 Εκπαίδευση μοντέλου

Η εκπαίδευση του μοντέλου πραγματοποιήθηκε με batch_size 12 εικόνων. Σύμφωνα με τους κανόνες δρομολόγησης του detectron [52], είχε διάρκεια 100000 βημάτων, με learning rate ίσο με 0.005 για τα πρώτα 80000 βήματα και 0.0005 για τα υπόλοιπα. Χρησιμοποιήθηκε SGD optimizer και weight decay ίσο με 0.0001.

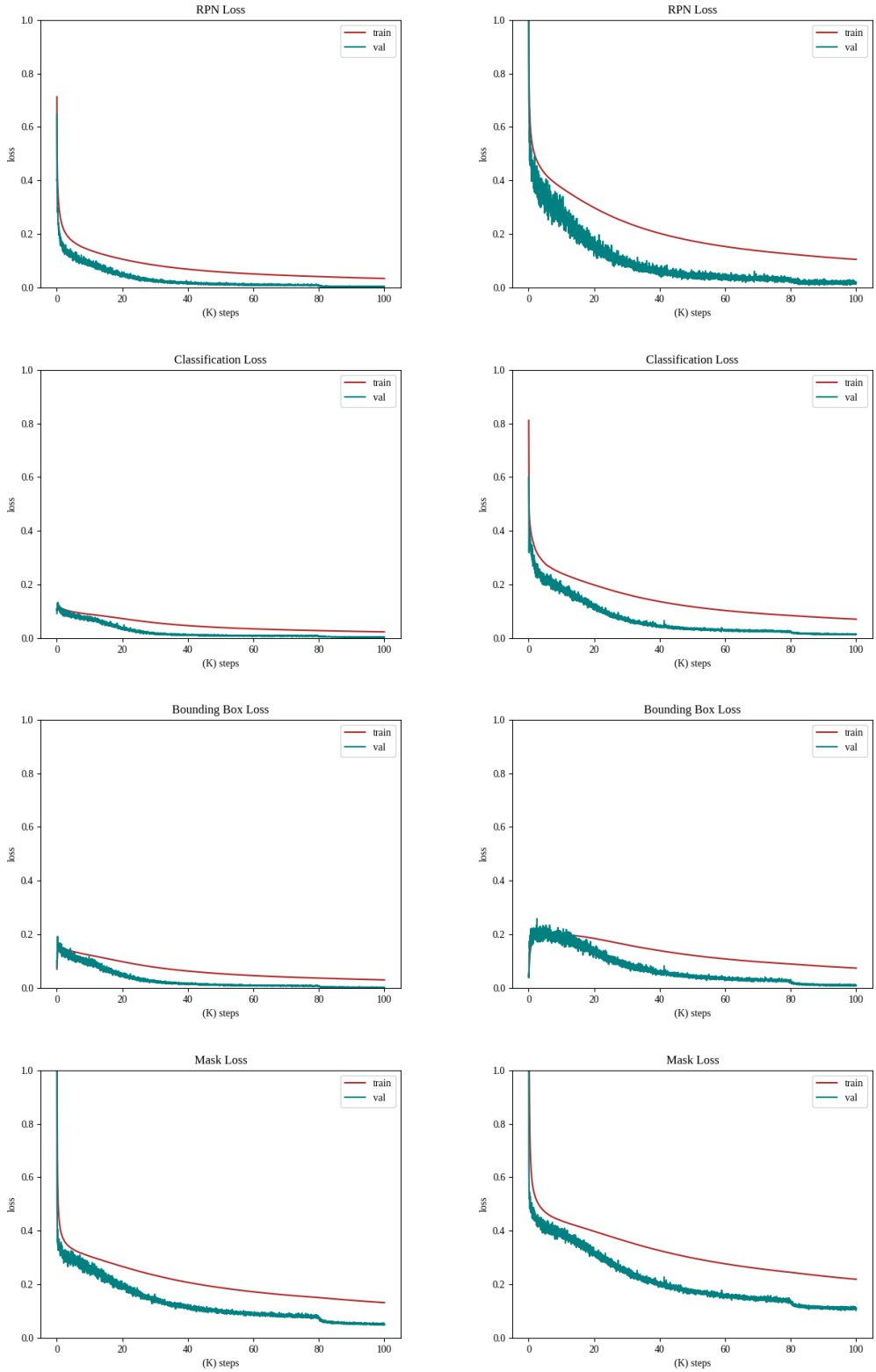
Παρακάτω φαίνεται η εξέλιξη της τιμής της συνάρτησης κόστους κατά την εκπαίδευση των δύο δικτύων. Για να διευκολυνθεί η σύγκριση των αποτελεσμάτων, το αριστερό σχήμα αφορά το Mask R-CNN, ενώ το δεξί το Rotated Mask R-CNN. Στο Σχήμα 6.1 φαίνεται η εξέλιξη της τιμής της συνάρτησης κόστους (Ενότητα 5.2.5) κατά την εκπαίδευση των δύο μοντέλων.



(a) Mask R-CNN (b) Rotated Mask R-CNN

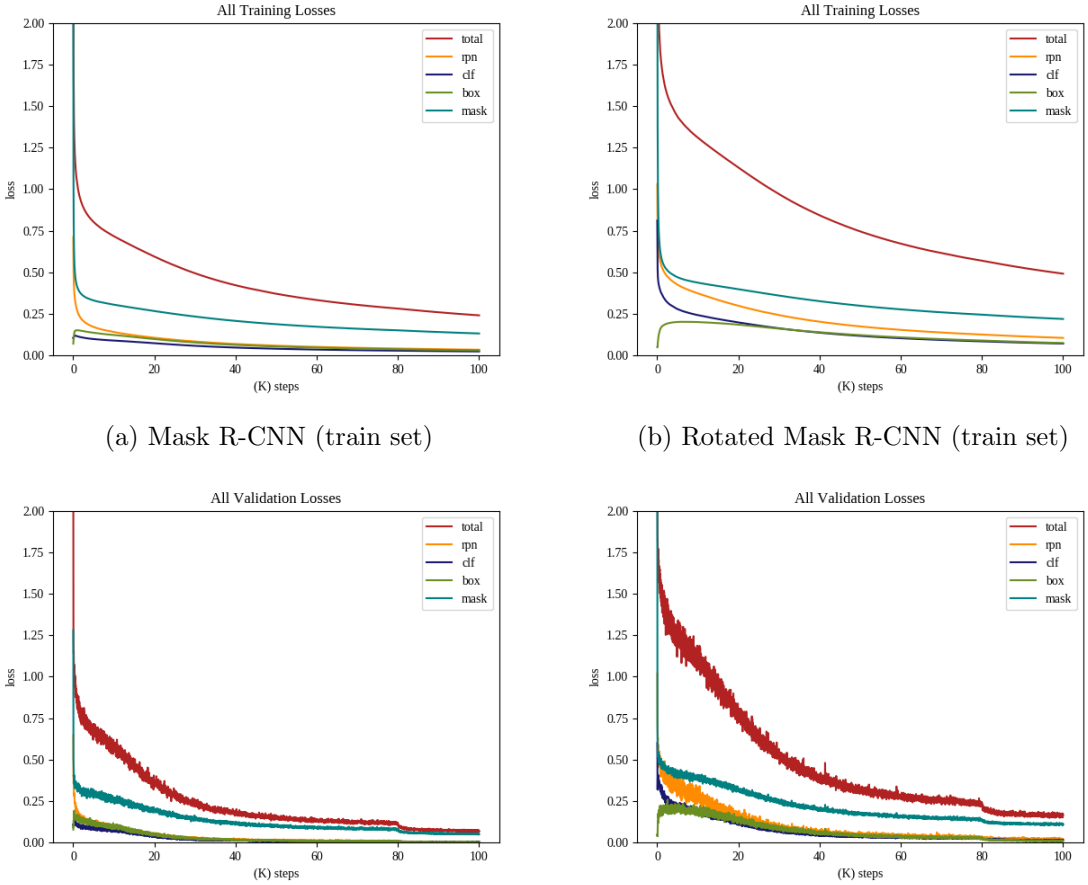
Σχήμα 6.1: Συνολικό Loss

Όπως είδαμε στην Ενότητα 5.2.5 η συνάρτηση κόστους υπολογίζεται ως το άθροισμα των επιμέρους συναρτήσεων κόστους των υποπροβλημάτων που καλείται να λύσει. Η εξέλιξη των επιμέρους αυτών συναρτήσεων κόστους κατά τη διάρκεια της εκπαίδευσης των δύο μοντέλων φαίνεται στο Σχήμα 6.2 (αριστερά για το Mask R-CNN, δεξιά για το Rotated Mask R-CNN).



Σχήμα 6.2: Συναρτήσεις Κόστους επιμέρους υποπροβλημάτων

Τέλος, στο Σχήμα 6.3 έχουν σχεδιαστεί όλα τα παραπάνω losses, ξεχωριστά για την εκπαίδευση και την αξιολόγηση.



(a) Mask R-CNN (train set)

(b) Rotated Mask R-CNN (train set)

(c) Mask R-CNN (test set)

(d) Rotated Mask R-CNN (test set)

Σχήμα 6.3: Ανάλυση Συνάρτησης Κόστους στα Δεδομένα Εκπαίδευσης και Αξιολόγησης

Με βάση τα παραπάνω, η εκπαίδευση του Rotated Mask R-CNN φαίνεται να χαρακτηρίζεται από μεγαλύτερες τιμές της συνάρτησης κόστους σε σχέση με το Mask R-CNN. Παρατηρώντας πιο προσεκτικά, ωστόσο, γίνεται αντιληπτό ότι για τη διαφορά αυτή ευθύνονται κυρίως οι μεγάλες τιμές της συνάρτησης κόστους του Δικτύου Προτάσεων Περιοχής (RPN) του Rotated Mask R-CNN.

Αναλυτική σύγκριση της απόδοσης των δύο μοντέλων γίνεται στην Ενότητα 6.5.

6.3 Ποιοτική Αξιολόγηση Αποτελεσμάτων

Καθώς ο συνολικός αριθμός των εικόνων του test set είναι υπερβολικά μεγάλος ώστε να συμπεριληφθούν όλες στην παρούσα εργασία, στο Σχήμα 6.3 παρατίθενται ενδεικτικά 10 από αυτές, τυχαία επιλεγμένες.

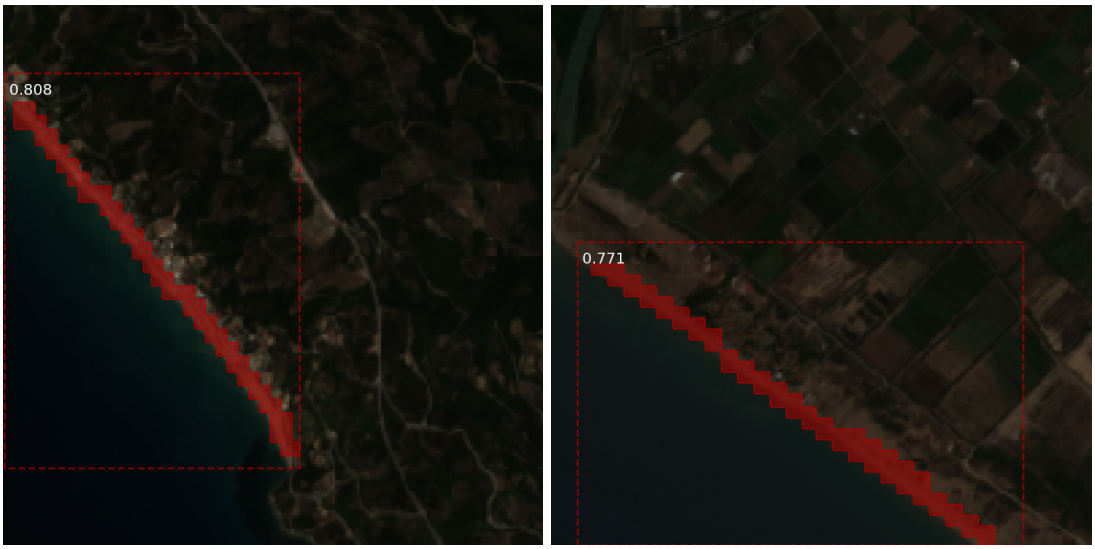
Ground Truth

Mask R-CNN

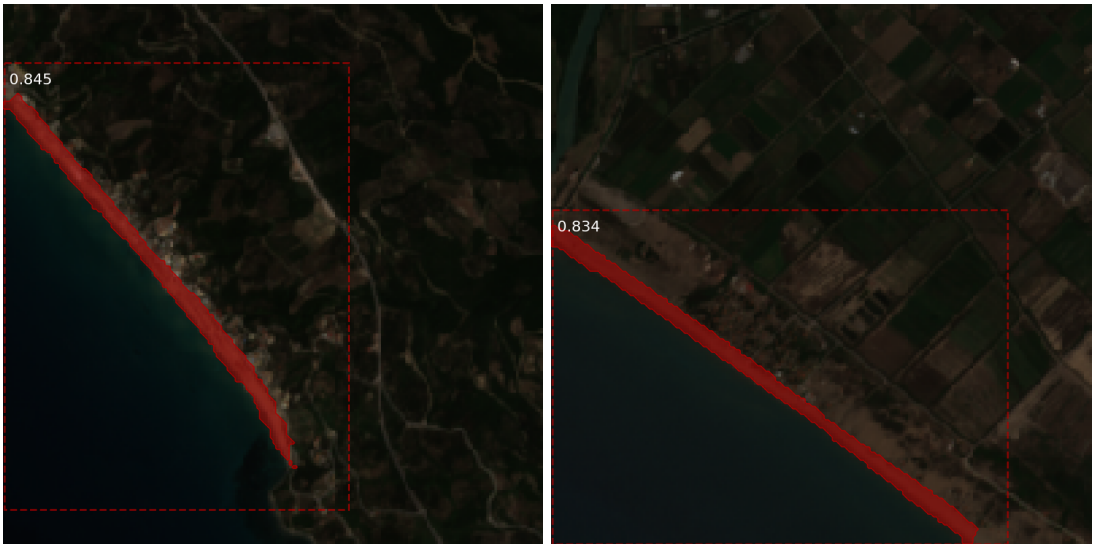
Rotated Mask R-CNN

Σχήμα 6.3: Αποτελέσματα Εντοπισμού

Ground Truth



Mask R-CNN



Rotated Mask R-CNN

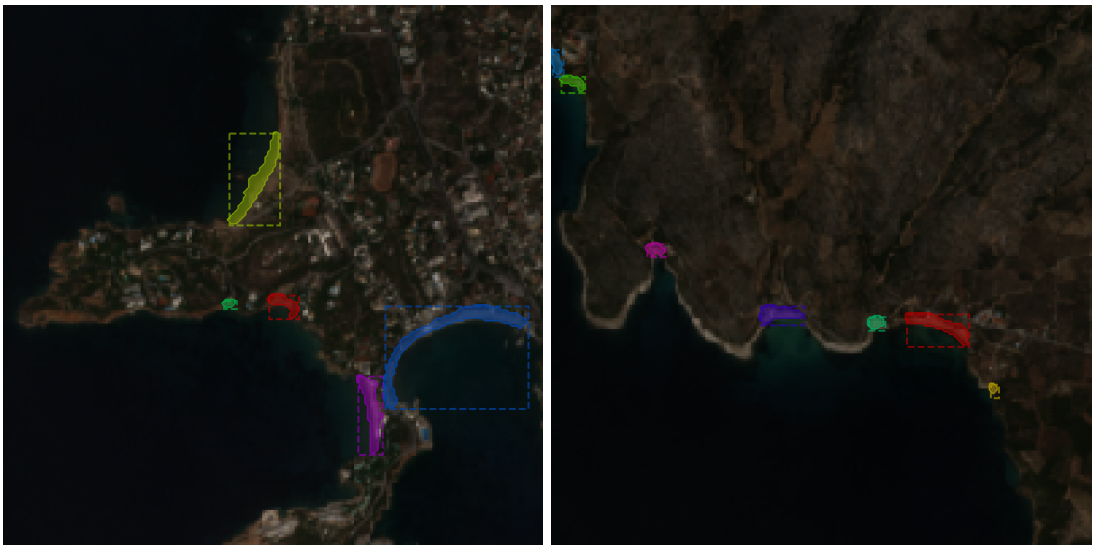
Σχήμα 6.3: Αποτελέσματα Εντοπισμού (συνέχεια)

Ground Truth

Mask R-CNN

Rotated Mask R-CNN

Σχήμα 6.3: Αποτελέσματα Εντοπισμού (συνέχεια)



Ground Truth

Mask R-CNN

Rotated Mask R-CNN

Σχήμα 6.3: Αποτελέσματα Εντοπισμού (συνέχεια)

Ground Truth

Mask R-CNN

Rotated Mask R-CNN

Σχήμα 6.3: Αποτελέσματα Εντοπισμού (συνέχεια)

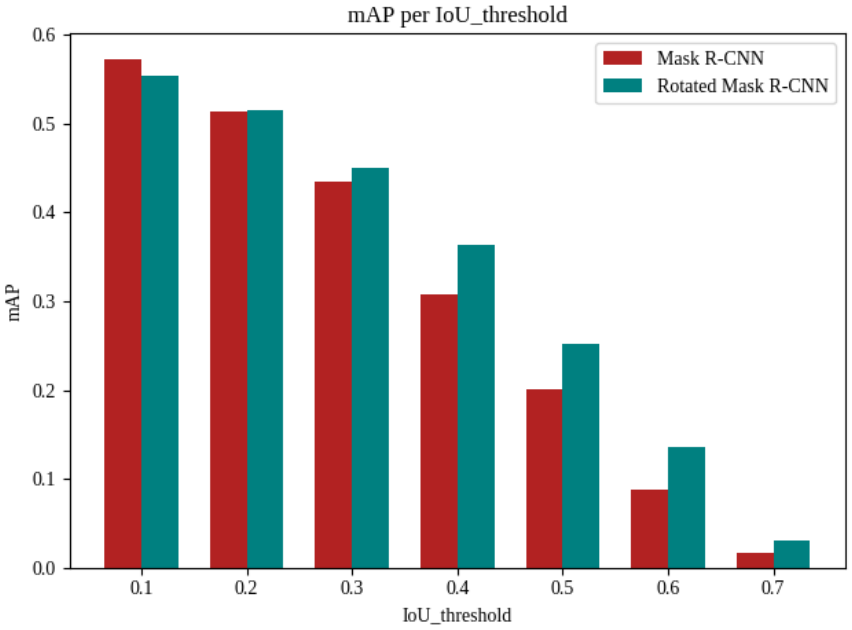
6.4 Ποσοτική Αξιολόγηση Αποτελεσμάτων

6.4.1 Average Precision (mAP) ανά κατώφλι IoU

Η βασική μετρική που χρησιμοποιήθηκε για την αξιολόγηση των δύο μοντέλων είναι η Μέση Ακρίβεια (mean Average Precision - mAP) (Ενότητα 5.1). Για τον υπολογισμό της mAP απαιτείται η επιλογή ενός κατωφλίου IoU για τον υπολογισμό της επιτυχημένης πρόβλεψης. Η τιμή αυτή ουσιαστικά περιγράφει πόσο "αυστηρή" είναι η αξιολόγησή μας ως προς την ακρίβεια των αποτελεσμάτων. Μία μικρή τιμή κατωφλίου δέχεται ως σωστές και προβλέψεις που ταυτίζονται σε μικρό βαθμό με το αντίστοιχο αντικείμενο του ground truth, με αποτέλεσμα να οδηγεί σε μεγαλύτερες τιμές της mAP, και αντίστροφα. Όπως συνηθίζεται σε παρόμοια προβλήματα της βιβλιογραφίας [41, 38], η απόδοση των μοντέλων αξιολογήθηκε για διαφορετικές τιμές του κατωφλίου, στο εύρος 0.1 - 0.7. Τα αποτελέσματα φαίνονται στον Πίνακα 6.1 και στο Σχήμα 6.4.

Κατώφλι IoU	**mAP**	
	Mask R-CNN	**Rotated Mask R-CNN**
0.1	0.573	0.554
0.2	0.513	0.515
0.3	0.435	0.450
0.4	0.307	0.363
0.5	0.201	0.252
0.6	0.088	0.136
0.7	0.016	0.031

Πίνακας 6.1: mean Average Precision (mAP) ανά κατώφλι IoU



Σχήμα 6.4: mean Average Precision (mAP) ανά κατώφλι IoU

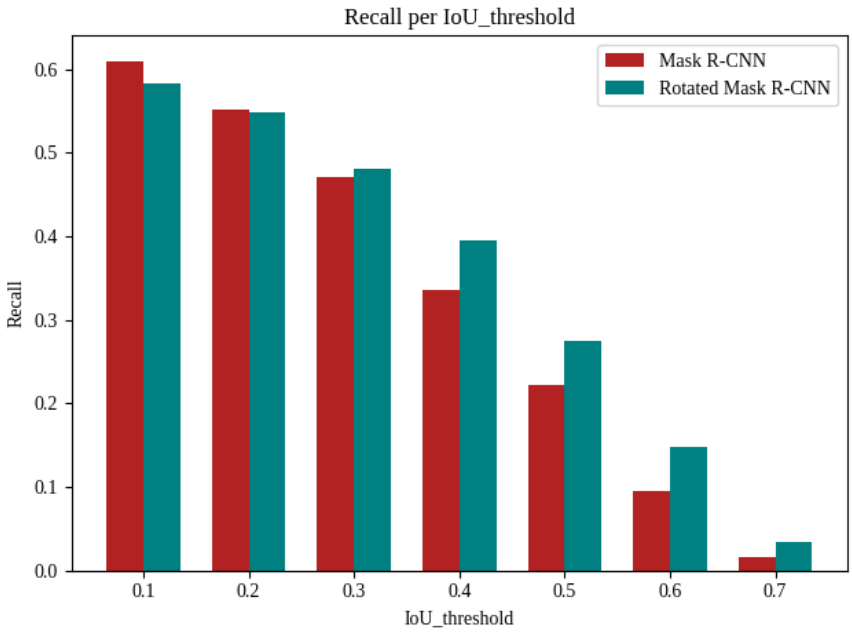
6.4.2 Recall ανά κατώφλι IoU

Καθώς το σύνολο δεδομένων μας δεν είναι πλήρες, δηλαδή δεν περιέχει όλες ανεξαιρέτως τις παραλίες της Ελλάδας, είναι λογικό και επιθυμητό το μοντέλο μας να εντοπίζει και παραλίες που δεν υπάρχουν στην βάση δεδομένων. Έχει νόημα, λοιπόν, να υπολογίσουμε και την ανάκληση (Recall) του μοντέλου (Ενότητα 5.1), η οποία περιγράφει πόσα από τα υπάρχοντα αντικείμενα εντοπίστηκαν (TruePositive), αγνοώντας αυτά που εντοπίστηκαν χωρίς να περιλαμβάνονται στο ground truth (FalsePositive).

Τα αποτελέσματα φαίνονται παρακάτω, στον Πίνακα 6.2 και στο Σχήμα 6.5.

Κατώφλι IoU	**Recall**	
	Mask R-CNN	**Rotated Mask R-CNN**
0.1	0.610	0.584
0.2	0.551	0.548
0.3	0.471	0.481
0.4	0.335	0.396
0.5	0.223	0.275
0.6	0.095	0.148
0.7	0.017	0.035

Πίνακας 6.2: Recall ανά κατώφλι IoU



Σχήμα 6.5: Recall ανά κατηγορία μεγέθους

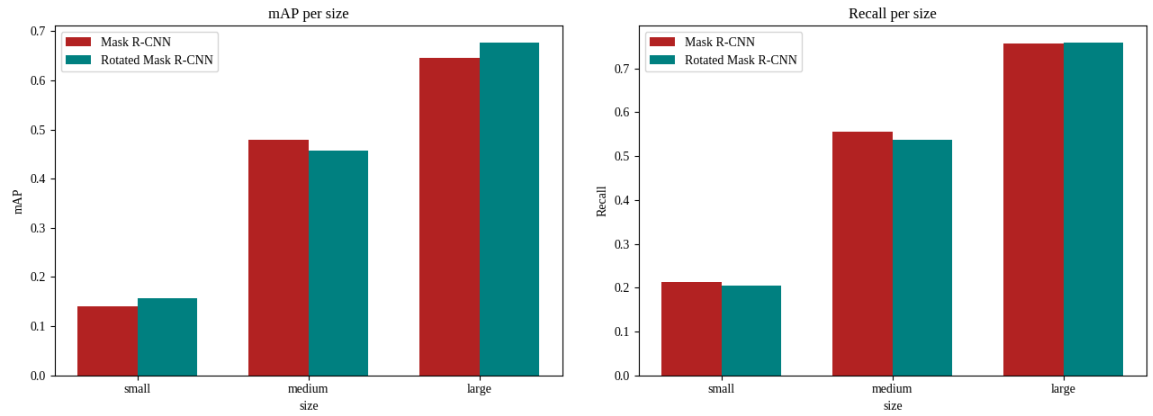
6.4.3 Average Precision και Recall ανά μέγεθος Περιοχής

Πέραν από την αξιολόγηση για διαφορετικές τιμές του κατωφλίου IoU, σύμφωνα με αντίστοιχες εργασίες [41] πραγματοποιήθηκε και αξιολόγηση για διαφορετικά μεγέθη αντικειμένων. Συγκεκριμένα, τα συνολικά 1154 αντικείμενα του test set χωρίστηκαν σε 3 κατηγορίες (small, medium, large) ανάλογα με το εμβαδόν τους και η αξιολόγηση έγινε για κάθε μία από αυτές τις κατηγορίες ξεχωριστά.

Στον Πίνακα 6.3 και το Σχήμα 6.6 παρακάτω παρουσιάζονται τα αποτελέσματα για τα δύο μοντέλα, για κατώφλι IoU ίσο με 0.2. Η τιμή αυτή επιλέχθηκε καθώς για αυτήν τα δύο μοντέλα παρουσιάζουν την ίδια απόδοση στο σύνολο των δεδομένων (Ενότητα 6.4.1).

Μέγεθος			**mAP**		**Recall**	
	Όρια (τμ)	#	**M R-CNN**	**RM R-CNN**	**M R-CNN**	**RM R-CNN**
small	< 2500	323	0.140	0.156	0.212	0.206
medium	2500 - 10000	536	0.480	0.457	0.555	0.538
large	> 10000	295	0.645	0.677	0.757	0.759

Πίνακας 6.3: mAP ανά κατηγορία μεγέθους



(a) mAP

(b) Recall

Σχήμα 6.6: mAP και Recall ανά κατηγορία μεγέθους

6.4.4 Average Precision και Recall ανά τιμή ORP

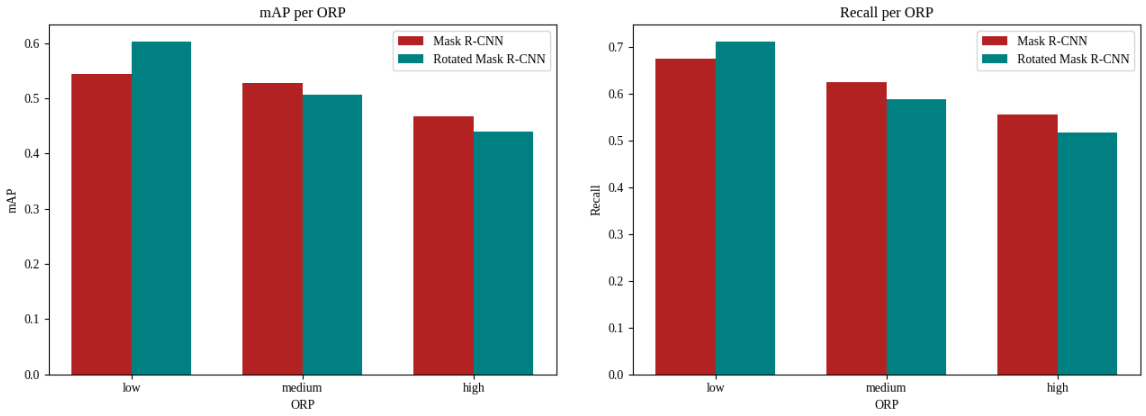
Όπως αναφέρθηκε στην Ενότητα 5.3.1, ένας από τους λόγους για τους οποίους επιλέξαμε να υλοποιήσουμε και το Rotated Mask R-CNN εκτός του απλού Mask R-CNN είναι η αδυναμία ορθής κατάτμησης των περιοχών με χαμηλό Λόγο Αντικειμένου προς Περιοχή (ORP) από το δεύτερο. Πραγματοποιήσαμε, λοιπόν, αξιολόγηση των δύο μοντέλων για αντικείμενα με διαφορετικές τιμές ORP. Συγκεκριμένα, τα αντικείμενα του test set χωρίστηκαν σε 3 κατηγορίες (low, medium, high) ανάλογα με την τιμή του ORP τους και η αξιολόγηση έγινε για κάθε μία από αυτές τις κατηγορίες ξεχωριστά.

Στον Πίνακα 6.4 και το Σχήμα 6.7 παρουσιάζονται τα αποτελέσματα για τα δύο μοντέλα,

για κατώφλι IoU ίσο με 0.2 και μέγεθος μεγαλύτερο από 25 εικονοστοιχεία (για μικρότερα αντικείμενα ο υπολογισμός του ORP δεν έχει τόσο νόημα). Η τιμή του κατωφλίου επιλέχθηκε καθώς για αυτήν τα δύο μοντέλα παρουσιάζουν την ίδια απόδοση στο σύνολο των δεδομένων (Ενότητα 6.4.1).

Μέγεθος			**mAP**		**Recall**	
	Όρια	#	**M R-CNN**	**RM R-CNN**	**M R-CNN**	**RM R-CNN**
low	0.0 - 0.3	247	0.545	0.603	0.675	0.712
medium	0.3 - 0.6	426	0.528	0.506	0.625	0.588
high	0.6 - 1.0	158	0.468	0.439	0.555	0.517

Πίνακας 6.4: mAP ανά τιμή ORP



(a) mAP

(b) Recall

Σχήμα 6.7: mAP και Recall ανά τιμή ORP

6.5 Ανάλυση Αποτελεσμάτων

6.5.1 Ιδιαιτερότητες Δεδομένων

Πριν προχωρήσουμε στην ανάλυση των αποτελεσμάτων, είναι σημαντικό να προσδιοριστεί η αναμενόμενη απόδοση σε σχέση με το σύνολο δεδομένων που διαθέτουμε.

Δορυφορικές Εικόνες Οι δορυφορικές εικόνες του Sentinel που χρησιμοποιήθηκαν έχουν χωρική ανάλυση 10 μέτρων. Όπως εύκολα γίνεται αντιληπτό, η ανάλυση αυτή δεν επαρκεί για έναν εξαιρετικά ακριβή εντοπισμό των παραλιών, οι οποίες σε πολλές περιπτώσεις έχουν πλάτος ελάχιστων εικονοστοιχείων. Το γεγονός αυτό είναι βέβαιο ότι θα οδηγήσει σε λιγότερο ακριβή αποτελέσματα, επομένως είναι αναμενόμενη μία μειωμένη τιμή της Μέσης Ακρίβειας, ειδικά για μεγάλες τιμές του κατωφλίου IoU (Ενότητα 6.4.1).

Ετικέτες Οι ετικέτες που χρησιμοποιήθηκαν για τη δημιουργία του συνόλου δεδομένων προέρχονται από τη βάση δεδομένων του OpenStreetMap, και είναι στην πλειοψηφία τους καταγεγραμμένες από τους χρήστες. Κατά συνέπεια:

- Το σύνολο δεδομένων δεν είναι πλήρες, αντίθετα σίγουρα υπάρχουν παραλίες που απεικονίζονται στις δορυφορικές εικόνες που χρησιμοποιούνται αλλά δεν αντιστοιχούν σε κάποια ετικέτα. Είναι, λοιπόν, όχι μόνο λογικό αλλά και επιθυμητό το μοντέλο να εντοπίζει και παραλίες που δεν υπάρχουν στο σύνολο δεδομένων, οπότε η πραγματική του απόδοση μπορούμε να υποθέσουμε ότι είναι μεγαλύτερη από αυτήν που υπολογίζουμε στο test set.
- Τα όρια των παραλιών εξ ορισμού δεν χαρακτηρίζονται από μεγάλη ακρίβεια, αλλά σε πολλές περιπτώσεις είναι πολύ απλοποιημένα. Υπάρχει, δηλαδή, το ενδεχόμενο το μοντέλο, το οποίο εκμεταλλεύεται την πληροφορία που περιέχεται στην εικόνα, να καταλήξει σε ένα πιο ακριβές περίγραμμα της παραλίας. Κατά συνέπεια, δεν έχει νόημα να χρησιμοποιούμε ιδιαίτερα αυστηρά κριτήρια (κατώφλι IoU) σχετικά με το ποιες προβλέψεις είναι σωστές, αφού μία μικρή τιμή IoU πιθανότατα σημαίνει ότι το αντικείμενο της πρόβλεψης διαφέρει από αυτό του grund truth γιατί είναι σωστότερο.

6.5.2 Σύγκριση Εκπαίδευσης

Όπως είδαμε στην Ενότητα 6.2, η συνάρτηση κόστους της εκπαίδευσης, και πιο συγκεκριμένα η συνάρτηση κόστους του Δικτύου Προτάσεων Περιοχής (RPN), παίρνει μεγαλύτερες τιμές για το Rotated Mask R-CNN σε σχέση με το Mask R-CNN. Αν και η διαφορά αυτή μας προδιαθέτει να υποθέσουμε ότι και η απόδοση του Mask R-CNN θα είναι καλύτερη, το γεγονός ότι οφείλεται κυρίως στο RPN υπονοεί ότι η μειωμένη απόδοση θα αφορά μόνο την πρόταση περιοχών, και όχι απαραίτητα την ποιότητα των αποτελεσμάτων.

Συγκεκριμένα, η διαφορά αυτή το πιθανότερο είναι ότι εξηγείται από το γεγονός ότι για τα δύο δίκτυα χρησιμοποιήθηκε ακριβώς η ίδια παραμετροποίηση, ώστε να μην επηρεαστεί η σύγκρισή τους από τις διαφορές των παραμέτρων. Τόσο στο Mask R-CNN όσο και στο Rotated Mask R-CNN το Δίκτυο Πρότασης Περιοχών παράγει τον ίδιο αριθμό προτάσεων (Appendix A), παρόλο που το RRPN (το RPN του Rotated Mask R-CNN) επιλέγει για κάθε άγκυρα και πλαίσια οριοθέτησης με περιστροφή, τα οποία τριπλασιάζουν τον αριθμό των προτάσεων ανά άγκυρα. Το Rotated RPN, δηλαδή, εξαναγκάζεται να απορρίψει μεγαλύτερο ποσοστό των αρχικών διαθέσιμων προτάσεών του σε σχέση με το RPN του Mask R-CNN, γεγονός που πιθανότατα εξηγεί τη διαφορά στην απόδοση των δύο Δικτύων Προτάσεων Περιοχής. Η αύξηση του συνολικού αριθμού προτάσεων είναι πολύ πιθανό ότι θα έλυνε το πρόβλημα.

6.5.3 Τελική Απόδοση Μοντέλων

Εξετάζοντας τις τιμές του Average Precision των δύο μοντέλων για διαφορετικές τιμές κατωφλίου IoU, παρατηρούμε ότι καθώς αυξάνουμε την τιμή του κατωφλίου μειώνεται η απόδοση του μοντέλου. Η μείωση αυτή είναι, προφανώς, αναμενόμενη, καθώς είναι λογικό όσο γίνονται πιο απαιτητικά τα κριτήριά μας σχετικά με το ποιες προβλέψεις θεωρούνται σωστές, ο αριθμός των σωστών προβλέψεων να γίνεται μικρότερος. Το ίδιο ισχύει προφανώς και για την ανάκληση (Recall). Στις γνωστότερες αντίστοιχες εργασίες [41, 38], ως βασική τιμή κατωφλίου επιλέγεται το 0.5. Παρόλα αυτά, με δεδομένες τις ιδιαιτερότητες του συνόλου δεδομένων που αναφέρθηκαν

στην Ενότητα 6.5.1 καθώς και την ανάλυση των ποιοτικών αποτελεσμάτων του μοντέλου (Ενότητα 6.3), αυτή η τιμή είναι υπερβολικά αυστηρή για την συγκεκριμένη εργασία, και επιλέχθηκε να μετρηθεί η τελική απόδοση για κατώφλι IoU ίσο με 0.3.

Mask R-CNN Με βάση τα παραπάνω, η Μέση Ακρίβεια (mean Average Precision - mAP) του Mask R-CNN στο test set του συνόλου δεδομένων, με κατώφλι IoU ίσο με 0.3 είναι ίση με (Ενότητα 6.4.1)

$$mAP_{0.3} = 43.5\%$$

ενώ η Ανάκληση (Recall) είναι ίση με (Ενότητα 6.4.2)

$$Recall_{0.3} = 47.1\%$$

Rotated Mask R-CNN Όσον αφορά το Rotated Mask R-CNN, η Μέση Ακρίβεια (mean Average Precision - mAP) του στο test set του συνόλου δεδομένων, με κατώφλι IoU ίσο με 0.3 είναι ίση με (Ενότητα 6.4.1)

$$mAP_{0.3} = 45.0\%$$

ενώ η Ανάκληση (Recall) είναι ίση με (Ενότητα 6.4.2)

$$Recall_{0.3} = 48.1\%$$

6.5.4 Σύγκριση μεταξύ Μοντέλων

Παρατηρώντας τη σύγκριση μεταξύ της απόδοσης του Mask R-CNN και του Rotated Mask R-CNN για διαφορετικές τιμές κατωφλίου (Σχήμα 6.4), βγάζουμε το συμπέρασμα ότι αν και για χαμηλές τιμές κατωφλίου (μικρότερες του 0.2) το Mask R-CNN έχει καλύτερη απόδοση, η αύξηση του κατωφλίου συνεπάγεται ξεκάθαρη υπεροχή του Rotated Mask R-CNN. Συγκεκριμένα, η απόδοση του Rotated Mask R-CNN είναι κατά 3.5% καλύτερη για IoU κατώφλι ίσο με 0.3, κατά 18.2% καλύτερη για IoU κατώφλι ίσο με 0.4 και κατά 25.4% καλύτερη για IoU κατώφλι ίσο με 0.5.

Με βάση τα παραπάνω, συμπεραίνουμε ότι ενώ το Mask R-CNN εμφανίζει πιο καλά αποτελέσματα στον εντοπισμό της περιοχής των αντικειμένων (όπως αναλύθηκε στην Ενότητα 6.5.2), αντιμετωπίζει μεγαλύτερη δυσκολία στην κατάτμηση, και δεν προσδιορίζει το περίγραμμα των αντικειμένων με την ακρίβεια του Rotated Mask R-CNN. Όσο αυστηροποιούνται τα κριτήρια επιλογής μίας πρόβλεψης ως σωστής, λοιπόν, το Mask R-CNN αποδίδει χειρότερα από το Rotated Mask R-CNN. Πέρα από την μελέτη των τιμών των μετρικών, αυτό επιβεβαιώνεται και από την επισκόπηση των εικόνων στην Ενότητα 6.3. Ενδεικτικά στο Σχήμα 6.3, στην δεύτερη σελίδα, όπου οι διαστάσεις των εικόνων το επιτρέπουν, βλέπουμε ότι το σχήμα της παραλίας που προβλέπει το Mask R-CNN είναι τελείως ανακριβές ενώ το αντίστοιχο του Rotated Mask R-CNN είναι πολύ κοντά σε αυτό του ground truth. Παρόλα αυτά, σε πολλές από τις εικόνες το Rotated Mask R-CNN δεν έχει καταφέρει να εντοπίσει όλα τα αντικείμενα, σε αντίθεση με το Mask R-CNN.

Όσον αφορά την απόδοση ως προς το μέγεθος περιοχής (Ενότητα 6.4.3), είναι σαφές ότι η απόδοση είναι καλύτερη για ευμεγέθη αντικείμενα. Ενδιαφέρον παρουσιάζει το γεγονός ότι

για τα μικρά και τα μεγάλα αντικείμενα το Rotated Mask R-CNN έχει ελαφρώς βελτιωμένη απόδοση, ακόμα και με την τιμή κατωφλίου 0.2 με την οποία έγιναν οι μετρήσεις.

Ενδιαφέροντα είναι επίσης τα στοιχεία που προκύπτουν από τη σύγκριση της απόδοσης των δύο μοντέλων σε διαφορετικά εύρη Λόγου Αντικειμένου Περιοχής (Ενότητα 6.4.4), όπου αποδεικνύεται και πειραματικά η υπόθεση (Ενότητα 5.3.1) ότι το Rotated Mask R-CNN είναι καταλληλότερο για αντικείμενα με μικρές τιμές ORP.

Κεφάλαιο 7

Επίλογος και Μελλοντικές Επεκτάσεις

Σύμφωνα με την ανάλυση των πειραματικών αποτελεσμάτων (Ενότητα 6.5.4), καταλήξαμε στο συμπέρασμα ότι για τη συγκεκριμένη παραμετροποίηση και τα δύο μοντέλα διαθέτουν τόσο πλεονεκτήματα όσο και μειονεκτήματα. Συγκεκριμένα, το Mask R-CNN χαρακτηρίζεται από μεγαλύτερη ικανότητα εντοπισμού της περιοχής των αντικειμένων, αλλά δεν προβλέπει με ακρίβεια τα όρια της περιοχής, ενώ το Rotated Mask R-CNN εντοπίζει τα όρια των περιοχών με πολύ ικανοποιητική ακρίβεια, αλλά συχνά δεν καταφέρνει να εντοπίσει κάποιες από τις περιοχές. Εξ αιτίας αυτού του tradeoff, η επιλογή του καλύτερου μεταξύ των δύο δεν είναι αυτονόητη, αλλά εξαρτάται από την εφαρμογή στην οποία θα χρησιμοποιηθεί.

Η διαφορά αυτή, όπως αναφέρθηκε στην Ενότητα 6.5.2, το πιθανότερο είναι ότι εξηγείται λόγω της ίδιας ακριβώς παραμετροποίησης η οποία επιλέχθηκε ώστε να είναι η σύγκριση των μοντέλων όσο το δυνατόν αντικειμενικότερη. Αυτό είχε σαν αποτέλεσμα το Δίκτυο Πρότασης Περιοχών να παράγει και στις δύο περιπτώσεις τον ίδιο αριθμό προτάσεων (Appendix A), παρόλο που το RRPN (το RPN του Rotated Mask R-CNN) επιλέγει για κάθε άγκυρα επιπλέον πλαίσια οριοθέτησης με περιστροφή, τα οποία τριπλασιάζουν τον αριθμό των προτάσεων ανά άγκυρα. Κατά συνέπεια, μία μελλοντική βελτίωση της παρούσας εργασίας θα μπορούσε να αφορά την επανάληψη των πειραμάτων με αυξημένο (πιθανώς τριπλάσιο) αριθμό προτάσεων για το RRPN.

Παράλληλα, όπως τονίστηκε στην Ενότητα 5.3.1, βασική ευθύνη για τη μειωμένη απόδοση των μοντέλων φέρει η χαμηλή χωρική ανάλυση των δεδομένων, η οποία σε συνδυασμό με το λεπτό σχήμα των περισσότερων παραλιών καθιστά πολύ δύσκολο τον εντοπισμό τους. Επανάληψη της εκπαίδευσης με χρήση δορυφορικών εικόνων καλύτερης ανάλυσης είναι απόλυτα βέβαιο ότι θα οδηγούσε σε κατά πολύ βελτιωμένα αποτελέσματα.

Παρά τις πιθανές χρήσιμες βελτιώσεις που αναφέρονται παραπάνω, η απόδοση του δικτύου είναι σίγουρα επαρκής, ιδιαίτερα δεδομένης της δυσκολίας του προβλήματος. Πρόκειται, άλλωστε, για ένα πρόβλημα που δύσκολα μπορεί να λυθεί με τέλεια αποτελέσματα ακόμα και από την ανθρώπινη αντίληψη, όπως καταλαβαίνει κανείς παρατηρώντας τις δορυφορικές εικόνες.

Διαθέτοντας το τελικό, εκπαιδευμένο μοντέλο, μπορούμε να το χρησιμοποιήσουμε, όπως αναφέρθηκε εισαγωγικά, για τη δημιουργία μίας βάσης δεδομένων. Επιπλέον, θα μπορούσε να χρησιμοποιηθεί ως μέρος ενός pipeline, όπου ο ρόλος του θα ήταν να απομονώσει τα εικονοστοιχεία που ανήκουν στην παραλία ώστε στη συνέχεια να πραγματοποιηθεί ταξινόμησή της ως προς τον τύπο εδάφους με ένα Συνελικτικό Νευρωνικό Δίκτυο.

Τέλος, αξίζει να σημειωθεί ότι βασικό πλεονέκτημα της διαδικασίας που ακολουθήθηκε στο πλαίσιο της εργασίας είναι ότι είναι εξ ολοκλήρου ανεξάρτητη του είδους του αντικειμένου που αφορούσε ο εντοπισμός, συμπεριλαμβανομένης της δημιουργίας του συνόλου δεδομένων. Με

ελάχιστες τροποποιήσεις, δηλαδή, η διαδικασία θα μπορούσε να επαναληφθεί για οποιοδήποτε από τα χιλιάδες χαρακτηριστικά που περιλαμβάνονται στο OpenStreetMap, ώστε για παράδειγμα να παραχθεί ένα μοντέλο που αντί παραλιών θα αναγνωρίζει πλατείες, καταρράκτες ή καλλιεργήσιμες εκτάσεις εσπεριδοειδών. Ίσως η πιο σημαντική λοιπόν βελτίωση που προτείνεται ως συνέχεια αυτής της εργασίας, είναι η μετατροπή της σε ένα παραμετροποιήσιμο εργαλείο, το οποίο απαιτώντας ως είσοδο μόνο το όνομα του αντικειμένου που θα πρέπει να εντοπίζει το μοντέλο που θα προκύψει, θα συλλέγει τις δορυφορικές εικόνες μέσω του Google Earth Engine, θα συγκεντρώνει τις ετικέτες μέσω του OSM, θα οργανώνει το σύνολο δεδομένων και θα εκπαιδεύει το δίκτυο, επιστρέφοντας το στην τελική μορφή του. Σε δεύτερη φάση, θα ήταν αρκετά απλό το εργαλείο να προσαρμοστεί και για την ταυτόχρονη αναγνώριση πολλών διαφορετικών αντικειμένων.

Βιβλιογραφία

[1] R. K. Turner, S. Subak, and W. N. Adger. Pressures, trends, and impacts in coastal zones: Interactions between socioeconomic and natural systems. *Environmental Management*, 20(2):159–173, March 1996.

[2] F. Bosello and E. De Cian. Climate change, sea level rise, and coastal disasters. A review of modeling practices. *Energy Economics*, 46:593–605, 2014.

[3] M. Mokhtarzade and M.J. Valadan Zoej. Road detection from high-resolution satellite images using artificial neural networks. *International Journal of Applied Earth Observation and Geoinformation*, 9(1):32–40, 2007.

[4] K. Zhao, J. Kang, J. Jung, and G. Sohn. Building Extraction from Satellite Images Using Mask R-CNN with Building Boundary Regularization. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 242–2424, 2018.

[5] X. Nie, M. Duan, H. Ding, B. Hu, and E. K. Wong. Attention Mask R-CNN for Ship Detection and Segmentation From Remote Sensing Images. *IEEE Access*, 8:9325–9334, 2020.

[6] S. Dhyakesh, A. Ashwini, S. Supraja, C. M. R. Aasikaa, M. Nithesh, J. Akshaya, and S. Vivitha. Mask R-CNN for Instance Segmentation of Water Bodies from Satellite Image. In A. Haldorai, A. Ramu, S. Mohanram, and M. Chen, editors, *2nd EAI International Conference on Big Data Innovation for Sustainable Cognitive Computing*, pages 301–307, Cham, 2021. Springer International Publishing.

[7] A. Emna, B. Alexandre, P. Bolon, M. Veronique, C. Bruno, and O. Georges. Offshore Oil Slicks Detection From SAR Images Through The Mask-RCNN Deep Learning Model. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, 2020.

[8] M. Li, Z. Zhang, L. Lei, X. Wang, and X. Guo. Agricultural Greenhouses Detection in High-Resolution Satellite Images Based on Convolutional Neural Networks: Comparison of Faster R-CNN, YOLO v3 and SSD. *Sensors*, 20, August 2020.

[9] J. Campbell and R. Wynne. *Introduction to Remote Sensing*. The Guilford Press, 2011.

[10] J. Lintz and D. S. Simonett. *Remote Sensing of Environment*. Addison-Wesley, 1976.

[11] World Meteorological Organization - Observing Systems Capability Analysis and Review Tool. `https://www.wmo-sat.info/oscar/satellites`. Accessed: 15/11/2020.

[12] Copernicus Programme. https://www.copernicus.eu/en/. Accessed: 15/11/2020.

[13] Sentinel Mission. https://sentinel.esa.int/web/sentinel/missions. Accessed: 15/11/2020.

[14] Sentinel-2 Mission. http://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-2. Accessed: 15/11/2020.

[15] Sentinel-2 User Handbook. https://sentinel.esa.int/documents/247904/685211/Sentinel-2_User_Handbook, 2015. ESA Standard Document, Issue 1, Revision 2.

[16] F. Gascon, C. Bouzinac, O. Thépaut, M. Jung, B. Francesconi, J. Louis, V. Lonjou, B. Lafrance, S. Massera, A. Gaudel-Vacaresse, F. Languille, B. Alhammoud, F. Viallefont, B. Pflug, J. Bieniarz, S. Clerc, L. Pessiot, T. Trémas, E. Cadau, R. De Bonis, C. Isola, P. Martimort, and V. Fernandez. Copernicus Sentinel-2A Calibration and Products Validation Status. *Remote Sens*, 9(584), 2017.

[17] Google Earth Engine Sentinel-2 L2A Dataset. https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_SR. Accessed: 15/11/2020.

[18] J. R. R. Uijlings, K. E. A. van de Sande, and T. Gevers et al. Selective Search for Object Recognition. *International Journal of Computer Vision*, 104:154—171, 2013.

[19] C. L. Zitnick and P. Dollar. Edge Boxes: Locating Object Proposals from Edges. In D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, editors, *Computer Vision – ECCV 2014*, volume 8693 of *Lecture Notes in Computer Science*. Springer, Cham, 2014.

[20] S. Ren, K. He, R. B. Girshick, and J. Sun. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada*, pages 91–99, 2015.

[21] M. Waleed Zafar. Object Detection and Segmentation using Region-based Deep Learning Architectures. Master's thesis, TU Dortmund University, December 2018.

[22] C. Lóopez Góomez. Deep Active Learning for Instance Segmentation. Master's thesis, Eindhoven University of Technology, August 2019.

[23] R. B. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 580–587, 2014.

[24] R. B. Girshick. Fast r-cnn. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 1440–1448, 2015.

[25] J. Pinto. Masknet: An Instance Segmentation Algorithm. Master's thesis, Chalmers University of Technology, 2017.

[26] R. Gandhi. R-CNN, Fast R-CNN, Faster R-CNN, YOLO — Object Detection Algorithms. https://towardsdatascience.com/r-cnn-fast-r-cnn-faster-r-cnn-yolo-object-detection-algorithms-36d53571365e. Accessed: 15/11/2020.

[27] F. Li, J. Johnson, and S. Yeung. Stanford cs231 lecture 11: Detection and segmentation. http://cs231n.stanford.edu/slides/2017/cs231n_2017_lecture11.pdf, 2018. Accessed: 15/11/2020.

[28] J. Long, E. Shelhamer, and T. Darrell. Fully Convolutional Networks for Semantic Segmentation. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3431–3440, 2015.

[29] B. Hariharan, P. Arbelaez, R. B. Girshick, and J. Malik. Simultaneous Detection and Segmentation. In D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, editors, *Computer Vision – ECCV 2014*, volume 8695 of *Lecture Notes in Computer Science*. Springer, Cham, 2014.

[30] K. He, G. Gkioxari, P. Dollár, and R. B. Girshick. Mask r-cnn. *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2980–2988, 2017.

[31] S. K. McFeeters. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, 17(7):1425–1432, 1996.

[32] J. Yue, J. Tian, Q. Tian, K. Xu, and N. Xu. Development of soil moisture indices from differences in water absorption between shortwave-infrared bands. *ISPRS Journal of Photogrammetry and Remote Sensing*, 154(8):216–230, 2019.

[33] M. Selva, B. Aiazzi, F. Butera, L. Chiarantini, and S. Baronti. Hyper-sharpening of hyperspectral data: A first approach. In *2014 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pages 1–4, 2014.

[34] Q. Wang, W. Shi, Z. Li, and P. M. Atkinson. Fusion of Sentinel-2 images. *Remote Sensing of Environment*, 187:241–252, 2016.

[35] Google Earth Engine Python API. https://developers.google.com/earth-engine/guides/python_install. Accessed: 15/11/2020.

[36] K. Tong, Y. Wu, and F. Zhou. Recent advances in small object detection based on deep learning: A review. *Image and Vision Computing*, 97:103910, 2020.

[37] G. Chen, H. Wang, K. Chen, Z. Li, Z. Song, Y. Liu, W. Chen, and A. Knoll. A Survey of the Four Pillars for Small Object Detection: Multiscale Representation, Contextual Information, Super-Resolution, and Region Proposal. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, pages 1–18, 2020.

[38] M. Everingham, L. Van Gool, and C. K. I. Williams et al. The Pascal Visual Object Classes (VOC) Challenge. *International Journal of Computer Vision*, 88:303–338, 2010.

[39] J. Hui. mAP (mean Average Precision) for Object Detection. https://jonathan-hui.medium.com/map-mean-average-precision-for-object-detection-45c121a31173. Accessed: 15/11/2020.

[40] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-Fei. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009.

[41] Microsoft Common Objects in Context. https://cocodataset.org/#home. Accessed: 15/11/2020.

[42] Image segmentation with Mask R-CNN. https://jonathan-hui.medium.com/image-segmentation-with-mask-r-cnn-ebe6d793272. Accessed: 15/11/2020.

[43] K. He, X. Zhang, S. Ren, and J. Sun. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 770–778. IEEE Computer Society, 2016.

[44] T. Lin, P. Dollár, R. B. Girshick, K. He, B. Hariharan, and S. J. Belongie. Feature Pyramid Networks for Object Detection. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 936–944. IEEE Computer Society, 2017.

[45] R. Madaan, D. Maturana, and S. Scherer. Wire Detection using Synthetic Data and Dilated Convolutional Networks for Unmanned Aerial Vehicles. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3487–3494, 2017.

[46] M. Holtzman-Gazit, R. Kimmel, N. Peled, and D. Goldsher. Segmentation of thin structures in volumetric medical images. *IEEE Transactions on Image Processing*, 15(2):354–363, 2006.

[47] Z. Liu, Y. Cao, Y. Wang, and W. Wang. Computer vision-based concrete crack detection using U-net fully convolutional networks. *Automation in Construction*, 104:129–139, 2019.

[48] F. Fang, L. Li, H. Zhu, and J. H. Lim. Combining Faster R-CNN and Model-Driven Clustering for Elongated Object Detection. *IEEE Transactions on Image Processing*, 29:2052–2065, 2020.

[49] S. Looi. Rotated Mask R-CNN: From Bounding Boxes To Rotated Bounding Boxes. https://github.com/mrlooi/rotated_maskrcnn, 2019. Accessed: 15/11/2020.

[50] J. Ma, W. Shao, H. Ye, L. Wang, H. Wang, Y. Zheng, and X. Xue. Arbitrary-Oriented Scene Text Detection via Rotation Proposals. *IEEE Transactions on Multimedia*, 20(11):3111–3122, 2018.

[51] F. Massa and R. Girshick. maskrcnn-benchmark: Fast, modular reference implementation of Instance Segmentation and Object Detection algorithms in PyTorch.

https://github.com/facebookresearch/maskrcnn-benchmark, 2018. Accessed: 15/11/2020.

[52] R. Girshick, I. Radosavovic, G. Gkioxari, P. Dollár, and K. He. Detectron. https://github.com/facebookresearch/detectron, 2018. Accessed: 15/11/2020.

Παράρτημα Α

Αρχεία παραμετροποίησης μοντέλων

Α.1 Mask R-CNN

```
AMP_VERBOSE: False
DATALOADER:
  ASPECT_RATIO_GROUPING: True
  NUM_WORKERS: 4
  SIZE_DIVISIBILITY: 32
DATASETS:
  TEST: ('beaches_2k_test',)
  TRAIN: ('beaches_2k_train', 'beaches_2k_val')
DTYPE: float32
INPUT:
  BRIGHTNESS: 0.0
  CONTRAST: 0.0
  HORIZONTAL_FLIP_PROB_TRAIN: 0.5
  HUE: 0.0
  MAX_SIZE_TEST: 1333
  MAX_SIZE_TRAIN: 1333
  MIN_SIZE_TEST: 400
  MIN_SIZE_TRAIN: (400,)
  PIXEL_MEAN: [17.58, 20.57, 21.54, 39.64, 41.4]
  PIXEL_STD: [1.0, 1.0, 1.0, 1.0, 1.0]
  ROTATE_DEGREES_TRAIN: (-90.0, 90.0)
  ROTATE_PROB_TRAIN: 0.0
  SATURATION: 0.0
  TO_BGR255: True
  VERTICAL_FLIP_PROB_TRAIN: 0.5
MODEL:
  BACKBONE:
    CONV_BODY: R-50-FPN
    FREEZE_CONV_BODY_AT: -1
  CLS_AGNOSTIC_BBOX_REG: False
  DEVICE: cuda
  FBNET:
    ARCH: default
    ARCH_DEF:
    BN_TYPE: bn
    DET_HEAD_BLOCKS: []
    DET_HEAD_LAST_SCALE: 1.0
    DET_HEAD_STRIDE: 0
    DW_CONV_SKIP_BN: True
    DW_CONV_SKIP_RELU: True
    KPTS_HEAD_BLOCKS: []
    KPTS_HEAD_LAST_SCALE: 0.0
    KPTS_HEAD_STRIDE: 0
    MASK_HEAD_BLOCKS: []
    MASK_HEAD_LAST_SCALE: 0.0
    MASK_HEAD_STRIDE: 0
    RPN_BN_TYPE:
    RPN_HEAD_BLOCKS: 0
    SCALE_FACTOR: 1.0
    WIDTH_DIVISOR: 1
  FPN:
    USE_GN: False
    USE_RELU: False
  GROUP_NORM:
    DIM_PER_GP: -1
    EPSILON: 1e-05
    NUM_GROUPS: 32
```

```
  IM_CHANNELS: 5
  KEYPOINT_ON: False
  MASKIOU_ON: True
  MASK_ON: True
  META_ARCHITECTURE: GeneralizedRCNN
  RESNETS:
    BACKBONE_OUT_CHANNELS: 256
    DEFORMABLE_GROUPS: 1
    NUM_GROUPS: 1
    RES2_OUT_CHANNELS: 256
    RES5_DILATION: 1
    STAGE_WITH_DCN: (False, False, False, False)
    STEM_FUNC: StemWithFixedBatchNorm
    STEM_OUT_CHANNELS: 64
    STRIDE_IN_1X1: True
    TRANS_FUNC: BottleneckWithFixedBatchNorm
    WIDTH_PER_GROUP: 64
    WITH_MODULATED_DCN: False
  RETINANET:
    ANCHOR_SIZES: (32, 64, 128, 256, 512)
    ANCHOR_STRIDES: (8, 16, 32, 64, 128)
    ASPECT_RATIOS: (0.5, 1.0, 2.0)
    BBOX_REG_BETA: 0.11
    BBOX_REG_WEIGHT: 4.0
    BG_IOU_THRESHOLD: 0.4
    FG_IOU_THRESHOLD: 0.5
    INFERENCE_TH: 0.05
    LOSS_ALPHA: 0.25
    LOSS_GAMMA: 2.0
    NMS_TH: 0.4
    NUM_CLASSES: 81
    NUM_CONVS: 4
    OCTAVE: 2.0
    PRE_NMS_TOP_N: 1000
    PRIOR_PROB: 0.01
    SCALES_PER_OCTAVE: 3
    STRADDLE_THRESH: 0
    USE_C5: True
  RETINANET_ON: False
  ROI_BOX_HEAD:
    CONV_HEAD_DIM: 256
    DILATION: 1
    FEATURE_EXTRACTOR: FPN2MLPFeatureExtractor
    MLP_HEAD_DIM: 1024
    NUM_CLASSES: 2
    NUM_STACKED_CONVS: 4
    POOLER_RESOLUTION: 7
    POOLER_SAMPLING_RATIO: 2
    POOLER_SCALES: (0.25, 0.125, 0.0625, 0.03125)
    PREDICTOR: FPNPredictor
    USE_GN: False
  ROI_HEADS:
    BATCH_SIZE_PER_IMAGE: 512
    BBOX_REG_ANGLE_RELATIVE: True
    BBOX_REG_WEIGHTS: (10.0, 10.0, 5.0, 5.0)
    BG_IOU_THRESHOLD: 0.3
    DETECTIONS_PER_IMG: 10
    FG_IOU_THRESHOLD: 0.3
    NMS: 0.01
    POSITIVE_FRACTION: 0.25
    SCORE_THRESH: 0.4
    SOFT_NMS:
      METHOD: 1
      SCORE_THRESH: 0.01
      SIGMA: 0.5
    USE_FPN: True
    USE_SOFT_NMS: True
  ROI_KEYPOINT_HEAD:
    CONV_LAYERS: (512, 512, 512, 512, 512, 512, 512, 512)
    FEATURE_EXTRACTOR: KeypointRCNNFeatureExtractor
    MLP_HEAD_DIM: 1024
    NUM_CLASSES: 17
    POOLER_RESOLUTION: 14
    POOLER_SAMPLING_RATIO: 0
    POOLER_SCALES: (0.0625,)
    PREDICTOR: KeypointRCNNPredictor
    RESOLUTION: 14
    SHARE_BOX_FEATURE_EXTRACTOR: True
```

```
    ROI_MASKIOU_HEAD:
      CONV_LAYERS: (256, 256, 256, 256)
      LOSS_WEIGHT: 1.0
      MLP_HEAD_DIM: 1024
      USE_GN: False
      USE_NMS: True
    ROI_MASK_HEAD:
      CONV_LAYERS: (256, 256, 256, 256)
      DILATION: 1
      FEATURE_EXTRACTOR: MaskRCNNFPNFeatureExtractor
      MLP_HEAD_DIM: 1024
      POOLER_RESOLUTION: 14
      POOLER_SAMPLING_RATIO: 2
      POOLER_SCALES: (0.25, 0.125, 0.0625, 0.03125)
      POSTPROCESS_MASKS: False
      POSTPROCESS_MASKS_THRESHOLD: 0.5
      PREDICTOR: MaskRCNNC4Predictor
      RESOLUTION: 28
      SHARE_BOX_FEATURE_EXTRACTOR: False
      USE_GN: False
      WITH_CLASSIFIER: False
    ROTATED: False
    RPN:
      ANCHOR_ANGLES: (-90, -60, -30)
      ANCHOR_SIZES: (32, 64, 128, 256, 512)
      ANCHOR_STRIDE: (4, 8, 16, 32, 64)
      ASPECT_RATIOS: (0.5, 1.0, 2.0)
      BATCH_SIZE_PER_IMAGE: 256
      BBOX_REG_ANGLE_RELATIVE: True
      BBOX_REG_WEIGHTS: (1.0, 1.0, 1.0, 1.0)
      BG_IOU_THRESHOLD: 0.1
      FG_IOU_THRESHOLD: 0.5
      FPN_POST_NMS_PER_BATCH: True
      FPN_POST_NMS_TOP_N_TEST: 1000
      FPN_POST_NMS_TOP_N_TRAIN: 6000
      MIN_SIZE: 0
      NMS_THRESH: 0.5
      POSITIVE_FRACTION: 0.5
      POST_NMS_TOP_N_TEST: 1000
      POST_NMS_TOP_N_TRAIN: 2000
      PRE_NMS_TOP_N_TEST: 1000
      PRE_NMS_TOP_N_TRAIN: 2000
      RPN_HEAD: SingleConvRPNHead
      STRADDLE_THRESH: -1
      USE_FPN: True
    RPN_ONLY: False
    WEIGHT: catalog://ImageNetPretrained/MSRA/R-50
    WEIGHT_LOAD_OPTIMIZER: True
    WEIGHT_LOAD_SCHEDULER: True
  OUTPUT_DIR: checkpoints/not_rotated/mscoco_msrcnn
  PATHS_CATALOG: rotated_maskrcnn-master/maskrcnn_benchmark/config/paths_catalog.py
  SOLVER:
    BASE_LR: 0.005
    BIAS_LR_FACTOR: 2
    CHECKPOINT_PERIOD: 2500
    GAMMA: 0.1
    IMS_PER_BATCH: 12
    MAX_ITER: 100000
    MOMENTUM: 0.9
    OPTIMIZER: SGD
    STEPS: (80000, 105000)
    WARMUP_FACTOR: 0.3333333333333333
    WARMUP_ITERS: 500
    WARMUP_METHOD: linear
    WEIGHT_DECAY: 0.0001
    WEIGHT_DECAY_BIAS: 0
  TEST:
    BBOX_AUG:
      ENABLED: False
      H_FLIP: False
      MAX_SIZE: 4000
      SCALES: ()
      SCALE_H_FLIP: False
    DETECTIONS_PER_IMG: 100
    EXPECTED_RESULTS: []
    EXPECTED_RESULTS_SIGMA_TOL: 4
    IMS_PER_BATCH: 2
```

A.2 Rotated Mask R-CNN

```
AMP_VERBOSE: False
DATALOADER:
  ASPECT_RATIO_GROUPING: True
  NUM_WORKERS: 4
  SIZE_DIVISIBILITY: 32
DATASETS:
  TEST: ('beaches_2k_test',)
  TRAIN: ('beaches_2k_train', 'beaches_2k_val')
DTYPE: float32
INPUT:
  BRIGHTNESS: 0.0
  CONTRAST: 0.0
  HORIZONTAL_FLIP_PROB_TRAIN: 0.5
  HUE: 0.0
  MAX_SIZE_TEST: 1333
  MAX_SIZE_TRAIN: 1333
  MIN_SIZE_TEST: 400
  MIN_SIZE_TRAIN: (400,)
  PIXEL_MEAN: [17.58, 20.57, 21.54, 39.64, 41.4]
  PIXEL_STD: [1.0, 1.0, 1.0, 1.0, 1.0]
  ROTATE_DEGREES_TRAIN: (-90.0, 90.0)
  ROTATE_PROB_TRAIN: 0.0
  SATURATION: 0.0
  TO_BGR255: True
  VERTICAL_FLIP_PROB_TRAIN: 0.5
MODEL:
  BACKBONE:
    CONV_BODY: R-50-FPN
    FREEZE_CONV_BODY_AT: -1
  CLS_AGNOSTIC_BBOX_REG: False
  DEVICE: cuda
  FBNET:
    ARCH: default
    ARCH_DEF:
    BN_TYPE: bn
    DET_HEAD_BLOCKS: []
    DET_HEAD_LAST_SCALE: 1.0
    DET_HEAD_STRIDE: 0
    DW_CONV_SKIP_BN: True
    DW_CONV_SKIP_RELU: True
    KPTS_HEAD_BLOCKS: []
    KPTS_HEAD_LAST_SCALE: 0.0
    KPTS_HEAD_STRIDE: 0
    MASK_HEAD_BLOCKS: []
    MASK_HEAD_LAST_SCALE: 0.0
    MASK_HEAD_STRIDE: 0
    RPN_BN_TYPE:
    RPN_HEAD_BLOCKS: 0
    SCALE_FACTOR: 1.0
    WIDTH_DIVISOR: 1
  FPN:
    USE_GN: False
    USE_RELU: False
  GROUP_NORM:
    DIM_PER_GP: -1
    EPSILON: 1e-05
    NUM_GROUPS: 32
  IM_CHANNELS: 5
  KEYPOINT_ON: False
  MASKIOU_ON: True
  MASK_ON: True
  META_ARCHITECTURE: GeneralizedRCNN
  RESNETS:
    BACKBONE_OUT_CHANNELS: 256
    DEFORMABLE_GROUPS: 1
    NUM_GROUPS: 1
    RES2_OUT_CHANNELS: 256
    RES5_DILATION: 1
    STAGE_WITH_DCN: (False, False, False, False)
    STEM_FUNC: StemWithFixedBatchNorm
    STEM_OUT_CHANNELS: 64
    STRIDE_IN_1X1: True
    TRANS_FUNC: BottleneckWithFixedBatchNorm
    WIDTH_PER_GROUP: 64
    WITH_MODULATED_DCN: False
```

```
  RETINANET:
    ANCHOR_SIZES: (32, 64, 128, 256, 512)
    ANCHOR_STRIDES: (8, 16, 32, 64, 128)
    ASPECT_RATIOS: (0.5, 1.0, 2.0)
    BBOX_REG_BETA: 0.11
    BBOX_REG_WEIGHT: 4.0
    BG_IOU_THRESHOLD: 0.4
    FG_IOU_THRESHOLD: 0.5
    INFERENCE_TH: 0.05
    LOSS_ALPHA: 0.25
    LOSS_GAMMA: 2.0
    NMS_TH: 0.4
    NUM_CLASSES: 81
    NUM_CONVS: 4
    OCTAVE: 2.0
    PRE_NMS_TOP_N: 1000
    PRIOR_PROB: 0.01
    SCALES_PER_OCTAVE: 3
    STRADDLE_THRESH: 0
    USE_C5: True
  RETINANET_ON: False
  ROI_BOX_HEAD:
    CONV_HEAD_DIM: 256
    DILATION: 1
    FEATURE_EXTRACTOR: FPN2MLPFeatureExtractor
    MLP_HEAD_DIM: 1024
    NUM_CLASSES: 2
    NUM_STACKED_CONVS: 4
    POOLER_RESOLUTION: 7
    POOLER_SAMPLING_RATIO: 2
    POOLER_SCALES: (0.25, 0.125, 0.0625, 0.03125)
    PREDICTOR: FPNPredictor
    USE_GN: False
  ROI_HEADS:
    BATCH_SIZE_PER_IMAGE: 512
    BBOX_REG_ANGLE_RELATIVE: True
    BBOX_REG_WEIGHTS: (10.0, 10.0, 5.0, 5.0, 1.0)
    BG_IOU_THRESHOLD: 0.3
    DETECTIONS_PER_IMG: 10
    FG_IOU_THRESHOLD: 0.3
    NMS: 0.01
    POSITIVE_FRACTION: 0.25
    SCORE_THRESH: 0.4
    SOFT_NMS:
      METHOD: 1
      SCORE_THRESH: 0.01
      SIGMA: 0.5
    USE_FPN: True
    USE_SOFT_NMS: True
  ROI_KEYPOINT_HEAD:
    CONV_LAYERS: (512, 512, 512, 512, 512, 512, 512, 512)
    FEATURE_EXTRACTOR: KeypointRCNNFeatureExtractor
    MLP_HEAD_DIM: 1024
    NUM_CLASSES: 17
    POOLER_RESOLUTION: 14
    POOLER_SAMPLING_RATIO: 0
    POOLER_SCALES: (0.0625,)
    PREDICTOR: KeypointRCNNPredictor
    RESOLUTION: 14
    SHARE_BOX_FEATURE_EXTRACTOR: True
  ROI_MASKIOU_HEAD:
    CONV_LAYERS: (256, 256, 256, 256)
    LOSS_WEIGHT: 1.0
    MLP_HEAD_DIM: 1024
    USE_GN: False
    USE_NMS: True
  ROI_MASK_HEAD:
    CONV_LAYERS: (256, 256, 256, 256)
    DILATION: 1
    FEATURE_EXTRACTOR: MaskRCNNFPNFeatureExtractor
    MLP_HEAD_DIM: 1024
    POOLER_RESOLUTION: 14
    POOLER_SAMPLING_RATIO: 2
    POOLER_SCALES: (0.25, 0.125, 0.0625, 0.03125)
    POSTPROCESS_MASKS: False
    POSTPROCESS_MASKS_THRESHOLD: 0.5
    PREDICTOR: MaskRCNNC4Predictor
    RESOLUTION: 28
```

```
    SHARE_BOX_FEATURE_EXTRACTOR: False
    USE_GN: False
    WITH_CLASSIFIER: False
  ROTATED: True
  RPN:
    ANCHOR_ANGLES: (-90, -60, -30)
    ANCHOR_SIZES: (32, 64, 128, 256, 512)
    ANCHOR_STRIDE: (4, 8, 16, 32, 64)
    ASPECT_RATIOS: (0.5, 1.0, 2.0)
    BATCH_SIZE_PER_IMAGE: 256
    BBOX_REG_ANGLE_RELATIVE: True
    BBOX_REG_WEIGHTS: (1.0, 1.0, 1.0, 1.0, 1.0)
    BG_IOU_THRESHOLD: 0.1
    FG_IOU_THRESHOLD: 0.5
    FPN_POST_NMS_PER_BATCH: True
    FPN_POST_NMS_TOP_N_TEST: 1000
    FPN_POST_NMS_TOP_N_TRAIN: 6000
    MIN_SIZE: 0
    NMS_THRESH: 0.5
    POSITIVE_FRACTION: 0.5
    POST_NMS_TOP_N_TEST: 1000
    POST_NMS_TOP_N_TRAIN: 2000
    PRE_NMS_TOP_N_TEST: 1000
    PRE_NMS_TOP_N_TRAIN: 2000
    RPN_HEAD: SingleConvRPNHead
    STRADDLE_THRESH: -1
    USE_FPN: True
  RPN_ONLY: False
  WEIGHT: catalog://ImageNetPretrained/MSRA/R-50
  WEIGHT_LOAD_OPTIMIZER: True
  WEIGHT_LOAD_SCHEDULER: True
OUTPUT_DIR: checkpoints/rotated/mscoco_msrcnn
PATHS_CATALOG: rotated_maskrcnn-master/maskrcnn_benchmark/config/paths_catalog.py
SOLVER:
  BASE_LR: 0.005
  BIAS_LR_FACTOR: 2
  CHECKPOINT_PERIOD: 2500
  GAMMA: 0.1
  IMS_PER_BATCH: 12
  MAX_ITER: 100000
  MOMENTUM: 0.9
  OPTIMIZER: SGD
  STEPS: (80000, 105000)
  WARMUP_FACTOR: 0.3333333333333333
  WARMUP_ITERS: 500
  WARMUP_METHOD: linear
  WEIGHT_DECAY: 0.0001
  WEIGHT_DECAY_BIAS: 0
TEST:
  BBOX_AUG:
    ENABLED: False
    H_FLIP: False
    MAX_SIZE: 4000
    SCALES: ()
    SCALE_H_FLIP: False
  DETECTIONS_PER_IMG: 100
  EXPECTED_RESULTS: []
  EXPECTED_RESULTS_SIGMA_TOL: 4
  IMS_PER_BATCH: 2
```