



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
Δ.Π.Μ.Σ ΕΠΙΣΤΙΜΗ ΔΕΔΟΜΕΝΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Επεξεργασία και Ταξινόμηση Δορυφορικών Δεδομένων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Δημήτριος Ν. Ζωγραφάκης

Επιβλέπων: Νεκτάριος Κοζύρης
Καθηγητής Ε.Μ.Π

Αθήνα, Ιούνιος 2021



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
Δ.Π.Μ.Σ ΕΠΙΣΤΙΜΗ ΔΕΔΟΜΕΝΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Επεξεργασία και Ταξινόμηση Δορυφορικών Δεδομένων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Δημήτριος Ν. Ζωγραφάκης

Επιβλέπων: Νεκτάριος Κοζύρης
Καθηγητής Ε.Μ.Π

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή τη 15^η Ιουνίου 2021.

.....
Νεκτάριος Κοζύρης
Καθηγητής Ε.Μ.Π

.....
Ιωάννης Κωνσταντίνου
Επ. Καθηγητής Π.Θ

.....
Ιωάννης Παπουτσής
Ερευνητής Ε.Α.Α

Αθήνα, Ιούνιος 2021

.....
Δημήτριος Ν. Ζωγραφάκης

© Δημήτριος Ν. Ζωγραφάκης, 2021.
Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Ευχαριστώ τους Καθηγητές Κοζύρη Νεκτάριο, Κωνσταντίνου Ιωάννη, τον Ερευνητή Παπουτσή Ιωάννη καθώς και τον Δρ Σοκά Δημήτρη για τη συνεργασία και την καθοδήγηση τους κατά τη διάρκεια εκπόνησης της παρούσας εργασίας.

Περιεχόμενα

Πρόλογος	6
Περίληψη	7
Abstract	7
1 Εισαγωγή	8
1.1 Κίνητρο	8
1.2 Οργάνωση Εγγράφου	9
2 Τηλεπισκόπηση	10
2.1 Ηλεκτρομαγνητικό Κύμα και Ηλεκτρομαγνητικό Φάσμα	10
2.1.1 Μέθοδοι Απεικόνισης Φωτογραφικών Δεδομένων	12
2.1.2 Θερμικές Ιδιότητες Αντικειμένων	12
2.2 Παγκόσμια Εγκάρσια Μερκατορική Προβολή	14
2.3 Αισθητήρες	15
2.3.1 Ενεργητικοί	15
2.3.2 Παθητικοί	16
2.4 Υπηρεσία Copernicus	16
2.4.1 Αποστολή Sentinel-1	16
2.4.2 Αποστολή Sentinel-2	19
3 Βαθιά Μάθηση	23
3.1 Αναδρομικό Νευρωνικό Δίκτυο (RNN)	24
3.2 Συνελκτικό Νευρωνικό Δίκτυο (CNN)	25
3.3 Παραγωγικό Αντιπαραθετικό Δίκτυο (GAN)	25
3.4 Μετασχηματιστές (Transformers)	26
3.4.1 Αρχιτεκτονική	27
3.4.2 Προσοχή (Attention)	28
3.5 Περιφερειακό Συνελκτικό Νευρωνικό Δίκτυο (R-CNN)	29
3.5.1 Fast R-CNN	30
3.5.2 Faster R-CNN	31
3.5.3 Mask R-CNN	32
3.6 DEtection TRansformer (DETR)	33
3.7 DuPLO	34
3.8 Μετρικές Αξιολόγησης Μοντέλων	35
4 Σύνολο Δεδομένων	38
4.1 Συλλογή	39
4.2 Περιγραφή	40
4.3 Επεξεργασία	41
4.3.1 Δομής Λήψης Αρχείων	41
4.3.2 Μετασχηματισμός Βάσης Δεδομένων	41
4.3.3 Ενοποίηση και Κωδικοποίηση Βάσης Δεδομένων	44
4.3.4 Οικονομία Κατά την Αποθήκευση	45
4.3.5 Τμηματοποίηση Αρχικής Εικόνας	46

4.3.6	Αφαίρεση Κενών Εικονοστοιχείων	46
4.3.7	Αφαίρεση Patches χωρίς Κατηγορίες	46
4.3.8	Αφαίρεση Χρονικών Στιγμών με μη Ακέραιο Αριθμό Καναλιών	47
4.3.9	Εντοπισμός και Εξαγωγή Μάσκας Συννεφιάς	47
4.3.10	Περαιτέρω Βελτιστοποιήσεις Βάσης Δεδομένων	49
4.3.11	Συμπλήρωση Κενών Μηνών	49
4.4	Τελική Μορφή (Σύνολο Εκπαίδευσης)	51
5	Ταξινόμηση Αγροτεμαχίων	57
5.1	Σχετική Έρευνα	57
5.2	Mask R-CNN	57
5.3	Detection Transformer (DETR)	58
5.4	DuPLO	58
5.5	Πειράματα	59
5.6	Αποτελέσματα	60
6	Συμπεράσματα	64
6.1	Παρατηρήσεις	64
6.2	Μελλοντικές Επεκτάσεις	65
6.2.1	Σύνολο Δεδομένων	65
6.2.2	Αρχιτεκτονικές Βαθιάς Μάθησης	65
	Βιβλιογραφία	68

Πρόλογος

Ο άνθρωπος είναι ένα ον με τάση να ταξινομεί την πληροφορία του περιβάλλοντος του σε κατηγορίες που αυτός καταλαβαίνει καλύτερα, προσπαθώντας με αυτόν τον τρόπο να την αποσαφηνίσει. Η παραπάνω διαδικασία αποτελεί ευεπίλυτο πρόβλημα για τον άνθρωπο αλλά είναι ιδιαίτερα απηλής όταν πραγματοποιείται από ένα μηχάνημα. Το εν λόγω κενό προσπαθεί να γεφυρώσει ένας κλάδος της Τεχνητής Νοημοσύνης, η Μηχανική Μάθηση, κατά την οποία, μαθηματικά μοντέλα εκπαιδεύονται έχοντας ως είσοδο ένα σύνολο γνωρισμάτων και ως έξοδο την επιθυμητή κατηγορία. Μετά το πέρας της εκπαίδευσης, το μοντέλο μαθαίνει να “προβλέπει” την επιθυμητή κατηγορία με βάση τα γνωρίσματα που δίνονται ως είσοδο.

Η Μηχανική Μάθηση ξεκίνησε με απλά μαθηματικά μοντέλα, τα οποία εξελίχθηκαν σε συνθετότερα με το πέρας των ετών και έφτασαν στο σημείο να συγκροτήσουν έναν νέο κλάδο, τη “Βαθιά Μάθηση”. Η Βαθιά Μάθηση αποτελεί πλέον συνώνυμο της Τεχνητής Νοημοσύνης, αφού η ενδυνάμωση της υπολογιστικής ισχύς στις νέες τεχνολογίες υπολογιστών καθώς και η παραγωγή και διάθεση δεδομένων σε παγκόσμια κλίμακα, έχουν δημιουργήσει τις δομές και έχουν θεσπίσει τα απαραίτητα πλαίσια για την εκτενή ενσωμάτωση της. Η τελευταία, δρασκελίζει τα όρια που είχαν θέσει τα μοντέλα της Μηχανικής Μάθησης επικροτώντας δαιδαλώδεις και πιο “βαθιές” αρχιτεκτονικές που οδηγούν σε μεγαλύτερη ακρίβεια κατά την πρόβλεψη.

Το πρόγραμμα Copernicus της Ε.Ε, προσφέρει ελεύθερα δεδομένα παρακολούθησης της Γης, δηλαδή δορυφορικές τηλεοπτικές απεικονίσεις. Συγκεκριμένα, θα χρησιμοποιηθούν δεδομένα από την αποστολή Sentinel 2, δηλαδή δεδομένα από παθητικούς οπτικούς αισθητήρες, για μία συγκεκριμένη περιοχή ενδιαφέροντος για το έτος 2020. Η αποστολή αποτελείται από δύο δίδυμους δορυφόρους με συχνότητα λήψης δεδομένων πέντε μέρες στον Ισημερινό και κάθε εικόνα που επιστρέφεται από τη βάση του προγράμματος, συγκροτείται από δώδεκα ζώνες δειγματοληψίας.

Ακόμα, διαθέσιμες προς το κοινό έχουν γίνει αρκετές βάσεις δεδομένων κρατών της Ε.Ε, οι οποίες περιγράφουν τις αγροτικές συνθήκες της κάθε χώρας. Η εν λόγω χαρτογράφηση των αγροτεμαχίων έχει ιδιαίτερα σημαντικό σκοπό, αφού ο έλεγχος του τύπου καλλιέργειας μπορεί πλέον, αν και δειγματοληπτικά, να γίνει σε ηλεκτρονικό επίπεδο. Βέβαια, όπως αναφέρθηκε για την περάτωση της διεργασίας αυτής απαιτείται πάλι η μεσολάβηση του ανθρώπινου παράγοντα.

Τα δορυφορικά δεδομένα, εκτός άλλων, χρησιμοποιούνται για την ανίχνευση, την παρατήρηση και εν τέλει την ταξινόμηση αγροτεμαχίων. Η διάρκεια ζωής, ο τρόπος καλλιέργειας καθώς και η εποχή αποτελούν μερικούς από τους παράγοντες που πρέπει να λαμβάνονται υπ’ όψιν κατά την τελική ταξινόμηση. Οι τελικές αποφάσεις οφείλουν να συσχετίζουν χρονικά τα δείγματα και να μην εισάγονται με τυχαία σειρά στο δίκτυο. Συνεπώς, το σύνολο δεδομένων πρέπει να θεωρείται ως χρονοσειρές εικόνων (τύπου βίντεο) και όχι σαν διακριτές εικόνες. Με αυτόν τον τρόπο, κάθε δείγμα είναι μία πολυχρωμική, πολυφασματική χρονοσειρά.

Σκοπός της παρούσας εργασίας, αποτελεί σε πρώτο στάδιο η συγκέντρωση και συνεπώς η διασταύρωση των διαθέσιμων γεωχωρικών δεδομένων της υπηρεσίας Copernicus με αυτά των βάσεων που προσφέρονται. Ένα δεύτερο στάδιο, πραγματεύεται τη διαχείριση και το μετασχηματισμό των τελικών συναθροισμένων δεδομένων σε μορφή εύπεπτη για αρχιτεκτονικές Βαθιάς Μάθησης. Τέλος, παρουσιάζονται σύνθετα μοντέλα που πραγματοποιούν γεωχωρικό διαμελισμό των αγροτεμαχίων που βρίσκονται σε μία εικόνα εισόδου.

Περίληψη

Η διαρκής ανάπτυξη συστάδων από δορυφόρους που συγκροτούν οι διάφορες διαστημικές αποστολές, παρέχουν πλέον ένα υπέρογκο πλήθος πληροφορίας σχετικά με την επιφάνεια αλλά και την ατμόσφαιρα της Γης. Πολλές υπηρεσίες κάνουν χρήση τέτοιων δεδομένων προκειμένου να προβούν σε έλεγχο καλλιέργειας, παρατήρηση του περιβάλλοντος και γενικότερα για την παροχή ασφάλειας. Οι προαναφερόμενες διεργασίες υποβοηθούνται από την παρουσία τέτοιων δορυφορικών δεδομένων, όμως απαιτούν την εμπλοκή του ανθρώπινου παράγοντα προκειμένου να επεξεργαστούν στα ενδιάμεσα βήματα. Σκοπός της παρούσας εργασίας αποτελεί η δημιουργία ενός πολυχρωρικού και πολυφασματικού κανονικοποιημένου Συνόλου Δεδομένων, που απαρτίζεται από δορυφορικά δεδομένα της Υπηρεσίας Κοπέρνικους και δεδομένα χρήσης Γης από δηλώσεις αγροτών. Ένα τελικό επίπεδο Τεχνητής Νοημοσύνης ενσωματώνεται, προκειμένου να παρέχει υποστήριξη στον ανθρώπινο παράγοντα κατά τα ενδιάμεσα βήματα αξιολόγησης των δεδομένων. Συγκεκριμένα, παρουσιάζονται αρχιτεκτονικές Βαθιών Νευρωνικών Δικτύων, δίνοντας ιδιαίτερη έμφαση σε αρχιτεκτονικές Όρασης Υπολογιστών, που είναι σε θέση να διαχωρίσουν και να ταξινομήσουν τη χρήση Γης για κάθε εικονοστοιχείο σε εικόνες εισόδου. Τα μοντέλα, εξάγουν μία δυαδική μάσκα που αποτυπώνει τη χρήση Γης για την αντίστοιχη εικόνα.

Λέξεις Κλειδιά: Τηλεπισκόπηση, Κοπέρνικους, Δορυφορικά Δεδομένα, Πολυφασματικές Εικόνες, Μηχανική Μάθηση, Βαθιά Μάθηση, Νευρωνικά Δίκτυα, ΣΝΔ, ΑΝΔ, ΜΒΔΔ, Μετασχηματιστές.

Abstract

The continuous development of satellite constellations from the various space projects, provide a tremendous amount of information regarding both the surface and the atmosphere of the Earth. A great number of services make use of such data, in order to inspect agricultural habits, to observe climate changes and, as a whole, to provide security services. The aforementioned services are supported by the presence of these available data, but the inbetween steps still require the interaction of the human factor. The sole purpose of this thesis, is to create a “normalized” multi-country, multitemporal Dataset from Copernicus Program data, as well as, farmer declarations for year 2020. A final layer of Artificial Intelligence is implemented in an attempt to assist final decision making. To be more specific, deep learning architectures are presented, with an emphasis in Computer Vision models, which are assigned to differentiate between “Crop” and “no Crop” for each pixel in an images. These models extract a segmentation mask for every image, that is used to describe the Land Use Land Cover (LULC) of the area.

Keywords: Remote Sensing, Copernicus, Satellite Data, Multispectral Imaging, Machine Learning, Deep Learning, Neural Networks, CNN, RNN, LSTM, Transformers.

Κεφάλαιο 1

Εισαγωγή

1.1 Κίνητρο

Η συλλογή και η αποθήκευση δεδομένων της Γήινης επιφάνειας γίνεται ολοένα γρηγορότερη, αποδοτικότερη, ποιοτικότερη με τη ραγδαία εξέλιξη της Τηλεπισκόπησης. Η εξέλιξη αυτή πλέον προσφέρει μία πληθώρα επιλογών ως προς τη δειγματοληψία, λ.χ τη σύλληψη δεδομένων σε πολυφασματικές ή και ακόμα υπερφασματικές απεικονίσεις. Τα εικονοστοιχεία συσχετίζονται με κάποιο μοναδικό σημείο πάνω στην επιφάνεια της Γης με απώτερο σκοπό την αναγνώριση αντικειμένων ή των επιφανειών έχοντας ως γνώρισμα την τιμή τους σε διάφορα μήκη κύματος, ή δηλαδή, τη φασματική τους υπογραφή [1].

Την τελευταία πενταετία, ένα πλήθος διαστημικών αποστολών έχει εκτοξευτεί με σκοπό την παρακολούθηση της Γης, πολλά από αυτά τα προγράμματα προσφέρουν χωρίς χρέωση τα δεδομένα τους στην κοινότητα. Για τον παραπάνω λόγο, η περίοδος αυτή χαρακτηρίζεται ως εποχή των Δωρεάν Μεγάλων Δεδομένων (Big Free Data) στην Τηλεπισκόπηση [2].

Ακόμα, τεχνικές Βαθιάς Μάθησης έχουν αρχίσει πρόσφατα να εφαρμόζονται για την επίλυση προβλημάτων της Τηλεπισκόπησης. Στο παρελθόν, αλγόριθμοι Μηχανικής Μάθησης κατακτούσαν αυτή τη θέση, ωστόσο με τον αυξανόμενο όγκο των διαθέσιμων δεδομένων από τα διάφορα δορυφορικά προγράμματα, η αντικατάσταση της με τη Βαθιά Μάθηση ήταν αναμενόμενη.

Η ανάλυση των εν λόγω δεδομένων επιφέρει θετικές αλλαγές σε πολλούς κλάδους της κοινωνίας [1]. Η ανίχνευση και διαχείριση καταστροφών αποτελεί έναν από τους τομείς, όπως για παράδειγμα η ανίχνευση ναυαγίου ενός πετρελαιοφόρου σε έναν ωκεανό ή ο προσδιορισμός ζημιών μετά από κάποια φυσική καταστροφή (π.χ πλημμύρα ή φωτιά). Περιβαντολογικοί λόγοι, λ.χ ανάλυση αιωρούμενων σωματιδίων ή ο προσδιορισμός της θερμοκρασίας ενός τόπου αποτελεί έναν ακόμα τομέα χρήσης των δορυφορικών δεδομένων.

Ίσως το βασικότερο πρόβλημα και αυτό που έχει γνωρίσει τη μεγαλύτερη προσοχή, αποτελεί αυτό της Ταξινόμησης Χρήσεων και Κάλυψης Γης (Land Use/Cover Classification; LUCL), της Ταξινόμησης Εικόνας (Scene Classification), της Αναγνώρισης Αντικειμένων (Object Detection) [3] και είναι αυτά που θα δώσει περισσότερη έμφαση η παρούσα εργασία. Η ταξινόμηση των γεωχωρικών δεδομένων αποτελεί μία εκτενής σύνθετη διαδικασία, πολλών επιπέδων [1], με αρκετά εμπόδια (υπερμεγέθη δεδομένα, γεωχωρικές και χρονικές συσχετίσεις, απουσία βάσεων για εξαγωγή κατηγοριών). Για την αντιμετώπιση τέτοιου είδους προβλημάτων, επιστρατεύονται κατά κύριο λόγο ΣΝΔ (CNN, Ενότητα 3.2) προκειμένου να εξαχθούν γεωχωρικά χαρακτηριστικά. Βέβαια, κάτι τέτοιο δε λαμβάνει υπ' όψιν τις χρονικές εξαρτήσεις των δεδομένων (χρονοσειρές), για αυτό το λόγο συνήθη είναι η χρήση δικτύων ANN (RNN, Ενότητα 3.1). Ακόμα, συχνές είναι οι αρχιτεκτονικές που συνδυάζουν τα παραπάνω σε ένα ενιαίο μοντέλο, όπως για παράδειγμα ConvLSTM.

Εξίσου σημαντικός παράγοντας είναι η έλλειψη διαθέσιμων Συνόλων Δεδομένων, προκειμένου να κατασκευαστούν αρχιτεκτονικές Βαθιάς Μάθησης, οι οποίες επιχειρούν να λύσουν τέτοια προβλήματα. Βέβαια, όταν γίνεται λόγος για Σύνολα Δεδομένων πάνω σε δορυφορικά δεδομένα, η τοποθεσία και ο χρόνος είναι αρκετά σημαντικός. Είναι απαραίτητη η φασματική υπογραφή του αντικειμένου μελέτης, τόσο σε χρόνο, τόπο αλλά και σε μήκη κύματος.

Έχοντας ως κίνητρο τα παραπάνω, θα αναπτυχθεί ένα Σύνολο Δεδομένων, το οποίο και θα χρησιμοποιηθεί για την εκπαίδευση και αξιολόγηση των προτεινόμενων αρχιτεκτονικών. Η υπηρεσία Copernicus

για το πρόγραμμα Sentinel 2, διαθέτει δύο τύπους προϊόντων, τα Level 1C και τα Level 2A. Αυτά παρουσιάζονται καλύτερα στην Ενότητα 2.4, συνοπτικά η κύρια διαφορά τους είναι η ατμοσφαιρική διόρθωση που εμπεριέχουν τα Level 2A. Συνεπώς, θα συλλεχθούν τα προϊόντα τύπου Level 1C, ώστε περαιτέρω διορθώσεις να δίνονται παραμετρικά. Όπως αναφέρθηκε, προκειμένου να παρέχεται η βέλτιστη διορατικότητα του αντικειμένου μελέτης, απαιτούνται πολυεπίπεδες δειγματοληψίες και άρα το τελικό μέγεθος αναμένεται να είναι αρκετά μεγάλο.

Εφόσον, συγκροτηθεί το προαναφερόμενο, θα κατασκευαστούν αρχιτεκτονικές Βαθιάς Μάθησης για την αντιμετώπιση προβλημάτων δυαδικής μορφής LUCL σε επίπεδο αγροτεμαχίου, δηλαδή Parcel Detection.

1.2 Οργάνωση Εγγράφου

Το παρόν έγγραφο οργανώνεται σε δύο νοητά πλαίσια το Θεωρητικό (Κεφάλαια 2 - 3) και το Πρακτικό (Κεφάλαια 4 - 6).

Κατά το πρώτο παρατίθενται το θεωρητικό υπόβαθρο που απαιτείται για τη σφαιρικότερη αντίληψη της επιστήμης της Τηλεπισκόπησης (Κεφάλαιο 2), ενώ κατά το προαναφερόμενο παρουσιάζονται συνοπτικά οι διαστημικές αποστολές Sentinel 1, Sentinel 2 της Υπηρεσίας Κοπέρνικους (Copernicus). Το Κεφάλαιο 3 απαρτίζεται από τις σύγχρονες κυρίαρχες αρχιτεκτονικές βαθιάς μάθησης, αναλύοντας τις σε σημαντικό βαθμό, παρουσιάζοντας τα προβλήματα στα οποία εφαρμόζονται καθώς και τις μετρικές που χρησιμοποιούνται προκειμένου να αξιολογηθούν τα ίδια και όλες οι παραλλαγές που προκύπτουν από διαφοροποιήσεις αυτών.

Το πρακτικό πλαίσιο, εκκινείται (Κεφάλαιο 4) περιγράφοντας τη συλλογή των δεδομένων από το πρόγραμμα Copernicus, τη συγκέντρωση των απαραίτητων Βάσεων Δεδομένων, ενδιάμεσα βήματα μετασχηματισμών και επεξεργασίας, με σκοπό την παραγωγή ενός Σύνολου Δεδομένων που θα τροφοδοτηθεί σε αρχιτεκτονικές Βαθιάς Μάθησης. Στη συνέχεια το Κεφάλαιο 5, αποτυπώνει την έρευνα που έχει υλοποιηθεί πάνω στα ανοιχτά προβλήματα που καλείται να λύσει η παρούσα εργασία. Διεξάγονται πειράματα με μία πληθώρα αρχιτεκτονικών αρχιτεκτονικές, δίνοντας ιδιαίτερη έμφαση σε μοντέλα που εφαρμόζονται σε προβλήματα Όρασης Υπολογιστών. Το κεφαλαίο αυτό, περιγράφει τυχόν ιδιαιτερότητες των μοντέλων, προτάσεις των συγγραφέων καθώς και αναλύει τη μεθοδολογία που εφαρμόστηκε για τη διεξαγωγή των πειραμάτων. Το τελευταίο κεφάλαιο (Κεφάλαιο 6) συγκεντρώνει τα διάφορα αποτελέσματα των πειραμάτων και συνοψίζει τα ευρήματα, αποτυπώνοντας παρατηρήσεις και πορίσματα που εξήχθησαν από όλο το εύρος της παρούσας εργασίας. Πιθανές εμπλουτίσεις και πορείες παρουσιάζονται λίγο πριν το πέρας της εν λόγω μελέτης.

Κεφάλαιο 2

Τηλεπισκόπηση

Η Τηλεπισκόπηση, στα αγγλικά γνωστή ως Remote Sensing, είναι η επιστήμη που ασχολείται με την απόκτηση πληροφοριών ενός αντικειμένου, μίας περιοχής ή ενός φαινομένου από απόσταση. Τα δεδομένα συλλέγονται εξ αποστάσεως από ειδικές συσκευές και στη συνέχεια αναλύονται σε χρήσιμες πληροφορίες [4].

Η πρώτη εναέρια φωτογραφία που αποτυπώθηκε ποτέ τραβήχτηκε από τον Gaspard-Félix Tournachon (γνωστό ως Nadar), το 1858 [5]. Αρκετές προσπάθειες διαδέχθηκαν αυτή του Nadar προκειμένου να χαρτογραφηθεί η επιφάνεια της Γης, πτηνά, χαρταετοί, πύραυλοι αποτελούν μερικές από αυτές.

Η ανάγκη για παρακολούθηση των αντίπαλων κινήσεων, οδήγησε στην άνθηση της τηλεπισκόπησης κατά τη διάρκεια του πρώτου Παγκοσμίου Πολέμου [5]. Αρκετά αεροσκάφη εξοπλίστηκαν με μηχανισμούς φωτογράφισης κατά τον Ψυχρό Πόλεμο, με απόγειο της εναέριας αναγνώρισης εκείνης της περιόδου, να αποτελεί η συλλογή των φωτογραφιών που τραβήχτηκαν το 1962 κατά τη κρίση των πυραύλων της Κούβας.

Η σύγχρονη εποχή της Τηλεπισκόπησης πυροδοτήθηκε από την εκτόξευση του πρώτου πολυφασματικού συστήματος επίγειας σάρωσης Landsat (Landsat Multispectral Scanner System; MSS), το 1972 [6].

Ένα απλό παράδειγμα τηλεπισκόπησης αποτελεί η λήψη μίας φωτογραφίας με τη χρήση κάμερας, εδώ η κάμερα λειτουργεί ως παθητικός αισθητήρας (βλ. Ενότητα 2.3) για να συλλέξει δεδομένα του ηλεκτρομαγνητικού φάσματος (βλ. Ενότητα 2.1) που εκπέμπει το μελετούμενο αντικείμενο, για παράδειγμα ένα δένδρο. Τα δεδομένα αυτά επεξεργάζονται στην κάμερα και στη συνέχεια παράγεται πληροφορία (φωτογραφία). Ο μελετητής με περαιτέρω νοητική επεξεργασία, μπορεί να αποσαφηνίσει το περιεχόμενο της πληροφορίας.

Σήμερα, η τηλεπισκόπηση συναντάται σε πολλούς κλάδους της καθημερινότητας όπως για παράδειγμα στην πρόγνωση του καιρού, τη μελέτη και την παρατήρηση της κλιματικής αλλαγής, την επίβλεψη ναυτικών δρομολογίων, την παγκόσμια ασφάλεια (Τσουνάμι, Ναύαγια Πετρελαίων) καθώς και σε κοινωφελείς υπηρεσίες όπως το Κτηματολόγιο.

2.1 Ηλεκτρομαγνητικό Κύμα και Ηλεκτρομαγνητικό Φάσμα

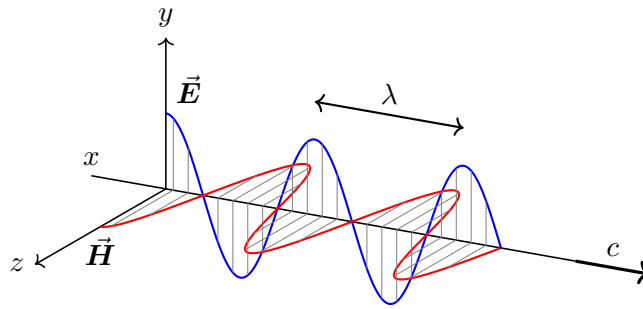
Τα κβάντα ή αλλιώς φωτόνια, όπως ονομάστηκαν από τον Άλμπερτ Αϊνστάιν, ταξιδεύουν ως ηλεκτρομαγνητικά κύματα στην ταχύτητα του φωτός. Το ηλεκτρομαγνητικό κύμα δημιουργείται από την ταλάντωση δύο πεδίων, του ηλεκτρικού και του μαγνητικού, κάθετα μεταξύ τους, ενώ η διεύθυνση διάδοσης τους είναι κάθετη με την ταλάντωση [7] (Σχήμα 2.1). Η εξίσωση 2.1 χρησιμοποιείται για να υπολογιστεί η ενέργεια που φέρει ένα φωτόνιο:

$$E = hv = h \frac{c}{\lambda} \quad (2.1)$$

όπου E η ενέργεια του φωτονίου σε Joules, h η σταθερά του Πλανκ ($6.63 \times 10^{-34} Js$), c η ταχύτητα του φωτός ($3 \times 10^8 ms^{-1}$), v η συχνότητα του κύματος και λ το μήκος του.

Έχοντας ως γνώμονα μία από τις βασικότερες εξισώσεις της Φυσικής (Εξίσωση 2.2):

$$c = v\lambda \quad (2.2)$$

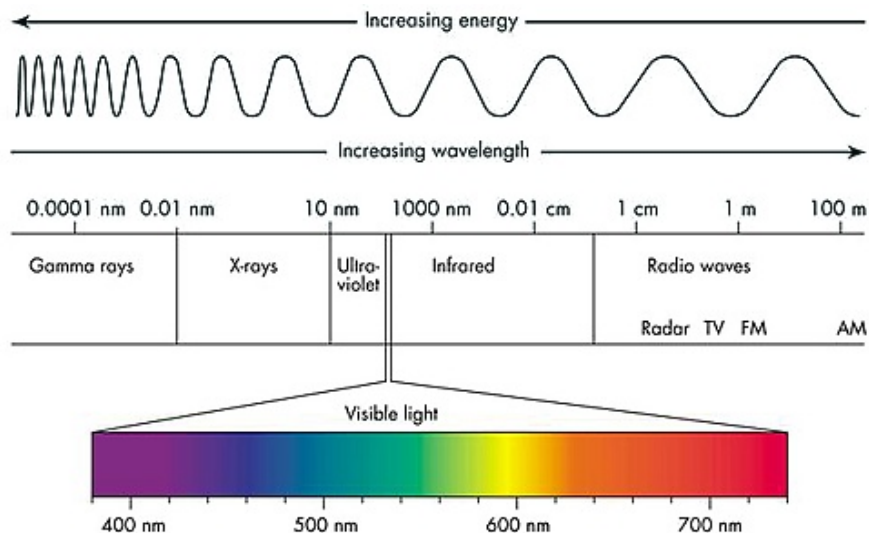


Σχήμα 2.1: Ηλεκτρομαγνητικό Κύμα.

Όπου c (σταθερά) ταχύτητα του φωτός, ν συχνότητα και λ μήκος κύματος. Παρατηρείται εύκολα ότι το μήκος κύματος και η συχνότητα είναι αντιστρόφως ανάλογα ποσά. Ως αποτέλεσμα μία μόνο μεταβλητή (συχνότητα ή μήκος κύματος) είναι αρκετή προκειμένου να προσδιοριστεί ένα σημείο μέσα στο ηλεκτρομαγνητικό φάσμα. Στην Τηλεπισκόπηση, αυτός ο προσδιορισμός γίνεται κατά κύριο λόγο με τη χρήση του μήκους κύματος (λ) και η πιο διαδεδομένη μονάδα μέτρησης μήκους ορίζεται το μικρόμετρο (δηλ. το ένα εκατομμυριοστό του μέτρου ή μm), το χιλιοστόμετρο (mm) ή σπανιότερα το μέτρο (m) [4, 8].

Ο Ήλιος, το λαμπρότερο σώμα του ουρανού, αποτελεί πηγή ενέργειας και φωτός για τον πλανήτη Γη [9]. Εκπέμπει στο διάστημα ηλιακή ακτινοβολία με τη μορφή ηλεκτρομαγνητικών κυμάτων. Η ηλιακή ακτινοβολία καταλήγει στην ατμόσφαιρα της Γης χωρίς μεγάλες αλλοιώσεις, η ατμόσφαιρα της Γης φιλτράρει ένα μέρος αυτής και καταλήγει στην επιφάνεια της, ενώ ένα άλλο μέρος ανακλάται πίσω στο διάστημα. Στην ηλιακή ακτινοβολία που ανακλάται από την επιφάνεια της Γης, βασίζεται η λήψη φωτογραφιών. Από αυτή που καταλήγει στην επιφάνεια, μερική απορροφάται και εκπέμπεται ως θερμική ενέργεια [8].

Όταν σε μία ακτίνα εμπεριέχονται φωτόνια μόνο ενός μήκους κύματος, τότε αυτή ονομάζεται Μονοχρωματική Ακτινοβολία, εν αντιθέσει ακτίνες που συμπεριλαμβάνουν φωτόνια με διαφορετικά μήκη κύματος (λ.χ Ηλιακή Ακτινοβολία) διαμορφώνουν την Πολυχρωματική Ακτινοβολία. Το ηλεκτρομαγνητικό φάσμα (Σχήμα 2.2) συγκαταλέγει όλους τους τύπους ακτινοβολίας για διαφορετικά μήκη κύματος [7].



Σχήμα 2.2: Ηλεκτρομαγνητικό Φάσμα.

Τις ιδιότητες αυτές εκμεταλλεύεται η Τηλεπισκόπηση κατά τη συλλογή δεδομένων, οι αισθητήρες σχεδιάζονται να ανιχνεύουν παθητικά διαφορετικά κανάλια ή ζώνες (οι ζώνες ταυτίζονται με την έννοια του διαστήματος, συνήθως με μεγάλο εύρος μήκους κύματος ενώ τα κανάλια αποτελούν στενότερες ζώνες) με κέντρο συγκεκριμένα μήκη κύματος (όπως για παράδειγμα η αποστολή Sentinel-2, Πίνακας 2.1) ή

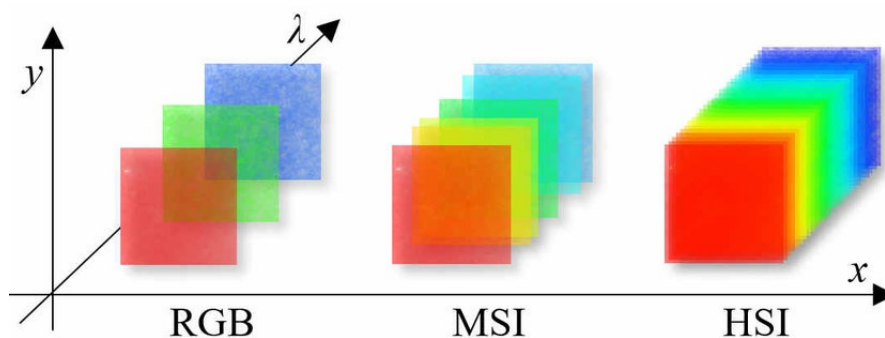
κάνουν χρήση ενεργητικών αισθητήρων τύπου RADAR (Αποστολή Sentinel-1) προκειμένου να δειγματοληψήσουν την επιφάνεια της Γης από απόσταση. Η ειδοποιός διαφορά μεταξύ των δύο τεχνολογιών αναλύεται στην Ενότητα 2.3. Συγκριτικά, τα RADAR προτιμούνται λόγω της ισχυρότερης ανοχής σε καιρικά φαινόμενα (λ.χ πυκνή συννεφιά).

Όλα τα τμήματα του Ηλεκτρομαγνητικού Φάσματος αν και παράγονται μέσω διαφορετικών διεργασιών, δε διαφέρουν σε σημαντικό βαθμό από τη γενικότερη φύση της Ηλεκτρομαγνητικής ακτινοβολίας [7].

2.1.1 Μέθοδοι Απεικόνισης Φωτογραφικών Δεδομένων

Ο κλασικός, και πλέον περισσότερο υιοθετημένος, τρόπος αναπαράστασης φωτογραφιών αποτελεί η τρικάναλη απεικόνιση ή αναλυτικότερα, η δειγματοληψία σε τρεις, αρκετά φαρδιές, ζώνες μήκους κύματος. Οι ζώνες αυτές συχνά αναφέρονται ως κόκκινο (R), πράσινο (G), μπλε (B). Η Πολυφασματική Απεικόνιση (Multispectral Imaging) επεκτείνει τα τρία κανάλια σε περισσότερα, συνήθως από τέσσερα έως δεκαπέντε δίχως περιορισμούς στο φάρδος κάθε ζώνης. Η Υπερφασματική Απεικόνιση (Hyperspectral Imaging), χρησιμοποιεί εκατοντάδες, περίπου ίδιου φάρδους, πιο στενές ζώνες [10].

Για κάθε ζώνη δειγματοληπτείται το ζητούμενο αντικείμενο και στη συνέχεια κωδικοποιείται σε μία δισδιάστατη αναπαράσταση, οι αναπαραστάσεις αυτές συνήθως ταξινομούνται ως προς το μήκος κύματος και στη συνέχεια τοποθετούνται διαδοχικά συνθέτοντας έτσι έναν φασματικό κύβο (spectral cube). Η τελική απεικόνιση λαμβάνεται συναθροίζοντας, με κάποιους συντελεστές βάρους, τις επιμέρους τιμές του κάθε εικονοστοιχείου. Το Σχήμα 2.3 αποτυπώνει τις παραπάνω περιγραφές.



Σχήμα 2.3: Διαχωρισμός μεταξύ απεικονίσεων (Πηγή: [10] Σχήμα 4.1).

Αν και οι παραπάνω τεχνολογίες αρχικά αναπτύχθηκαν για χρήσεις στην Τηλεπισκόπηση, οι πολυφασματικές και υπερφασματικές απεικονίσεις έχουν γνωρίσει ιδιαίτερη άνθηση σε πολλούς τομείς τα τελευταία χρόνια, εξαιτίας της ραγδαίας εξέλιξης της τεχνολογίας.

Η αποστολή Sentinel-2 κάνει χρήση Πολυφασματικής Απεικόνισης με 13 ζώνες (Πίνακας 2.1). Η Ultraspectral Απεικόνιση αποτελεί το μέλλον της Υπερφασματικής Απεικόνισης [7].

2.1.2 Θερμικές Ιδιότητες Αντικειμένων

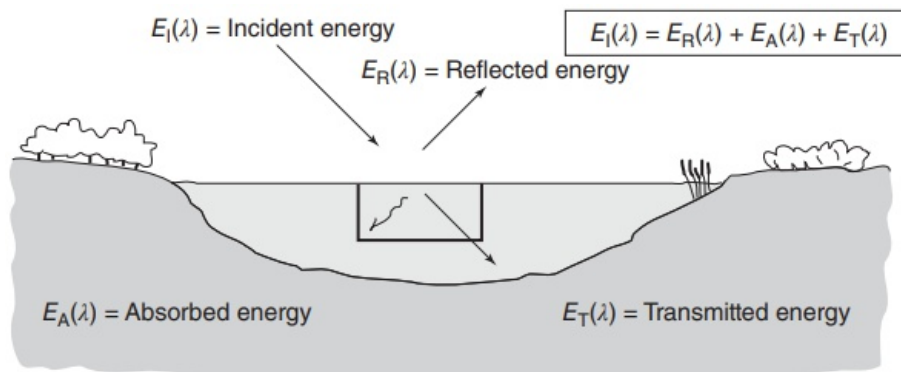
Όλα τα σώματα (στερεά, υγρά, αέρια) σε θερμοκρασίες άνω των 0 Κέλβιν (απόλυτο μηδέν) εκπέμπουν διαρκώς ηλεκτρομαγνητική ακτινοβολία, η ακτινοβολία αυτή οφείλεται στη θερμοκρασία τους και ονομάζεται Θερμική Ακτινοβολία. Η θερμική ακτινοβολία αποτελεί έναν από τους βασικότερους τρόπους μετάδοσης της θερμότητας.

Όταν η ηλεκτρομαγνητική ακτινοβολία συναντά οποιαδήποτε επιφάνεια, ένα μέρος από αυτή ανακλάται, ένα άλλο μέρος απορροφάται και το υπόλοιπο μεταδίδεται. Σύμφωνα με την Αρχή Διατήρησης της Ενέργειας, για E_I συνολική αρχική ενέργεια, E_R ενέργεια που ανακλάται, E_A ενέργεια που απορροφάται και E_T ενέργεια που μεταδίδεται [4] ισχύει η παρακάτω σχέση (Εξίσωση 2.3):

$$E_I(\lambda) = E_R(\lambda) + E_A(\lambda) + E_T(\lambda) \quad (2.3)$$

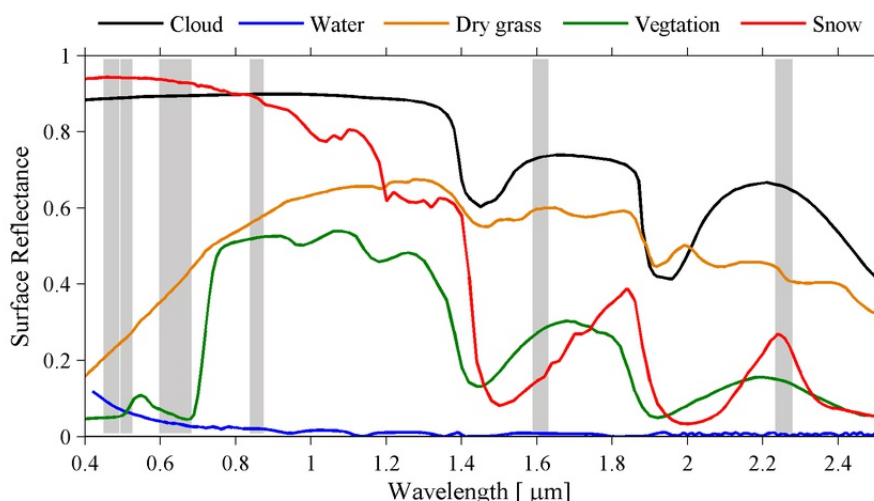
Στο Σχήμα 2.4 παρουσιάζεται μία τέτοια αλληλεπίδραση για επιφάνεια νερού, η Εξίσωση 2.3 ισχύει για διαφορετικές επιφάνειες αλλά οι συντελεστές των επιμέρους ενεργειών αλλάζουν. Επιπλέον, για ίδιες

επιφάνειες, οι επιμέρους αναλογίες μπορεί να αυξάνονται ή να μειώνονται για διαφορετικά μήκη κύματος. Η αναγνώριση των αντικειμένων με βάση της ενέργειας πραγματοποιείται κάνοντας εκμετάλλευση των παραπάνω γνωρισμάτων.



Σχήμα 2.4: Αλληλεπιδράσεις ηλεκτρομαγνητικής ακτινοβολίας με την επιφάνεια της Γης (Πηγή: [4] Σχήμα 1.6).

Οι κοινές κάμερες και φακοί λειτουργούν σε μήκη κύματος του ορατού φως. Εκεί που τελειώνει το ορατό φως, ξεκινάει η υπέρυθη ακτινοβολία, η οποία παρουσιάζει αρκετές ομοιότητες με το ορατό φως και για το λόγο, με ελάχιστες αλλαγές μπορούν να χρησιμοποιηθούν οι ίδιες συσκευές. Ωστόσο, από τη μέση του φάσματος των υπέρυθρων μέχρι την αρχή του φάσματος των μικροκυμάτων η δειγματοληψία δεν μπορεί πραγματοποιηθεί με τον ίδιο τρόπο. Αυτό οφείλεται κυρίως στον τρόπο αλληλεπίδρασης των εν λόγω μηκών κύματος με την ατμόσφαιρα καθώς και τον τύπο πληροφορίας που αποτυπώνουν. Συγκεκριμένα, μετρήσεις σε αυτά τα μήκη κύματος παρέχουν γνώση αναφορικά με τη Γήινη επιφάνεια και τη θερμοκρασία διαφόρων σωμάτων πάνω σε αυτή. Κύρια χρήση αποτελεί η αναγνώριση αντικειμένων με βάση τη θερμική τους ενέργεια [8]. Οι πιο κοινές συσκευές που παρέχουν τέτοιες μετρήσεις ονομάζονται θερμικοί σαρωτές (thermal scanners). Στο Σχήμα 2.5 παρουσιάζονται μετρήσεις σε διάφορα μήκη κύματος για πέντε επιφάνειες, οι μετρήσεις αυτές ονομάζονται φασματικές υπογραφές (spectral signatures) και χρησιμοποιούνται συχνά για την κατηγοριοποίηση των επιφανειών.

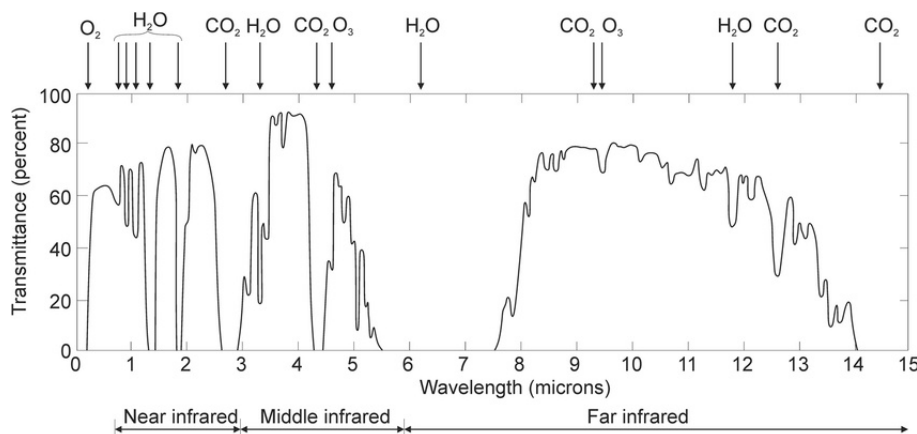


Σχήμα 2.5: Φασματικές υπογραφές ως συνάρτηση του μήκους κύματος (Πηγή: [11] Σχήμα 1).

Η κοινή πηγή εκπεμπόμενης θερμικής ακτινοβολίας είναι η ηλιακή ακτινοβολία. Τα σώματα στην επιφάνεια της Γης λαμβάνουν και απορροφούν την ακτινοβολία και στη συνέχεια την εκπέμπουν πίσω. Η ποσότητα και η φασματική κατανομή αυτής της εκπομπής εξαρτάται από τις θερμικές τους ιδιότητες. Για ανάλυση, χρησιμοποιείται ένα θερμογράφημα (thermogram) το οποίο περιέχει διάφορες τιμές επιπέδων του γκρι. Συνήθως, οι χαμηλότερες θερμοκρασίες απεικονίζονται με σκούρους τόνους ενώ οι υψηλότερες με πιο ανοιχτούς. Οι αντιθέσεις της θερμικής φωτεινότητας, αποτελούν τη βάση για τον προσδιορισμό

των χαρακτηριστικών [8].

Τα διάφορα αερολύματα και σωματίδια (νερού ή/και πάγου) που συνθέτουν την ατμόσφαιρα ποικίλουν. Για το λόγο αυτό, η ατμόσφαιρα παρουσιάζει διαφορετική συμπεριφορά ως προς τη διαπερατότητα της ακτινοβολίας στα διάφορα μήκη κύματος της. Όλα αυτά, δηλαδή οι διαδικασίες απορρόφησης, σκέδασης και σπινθηρισμού προσθέτουν θόρυβο ή και εξασθενούν το σήμα κατά τη διαδικασία δειγματοληψίας. Συνεπώς, η δειγματοληψία γίνεται σε παράθυρα που καθορίζονται από τις προαναφερόμενες διεργασίες [12] (Σχήμα 2.6).



Σχήμα 2.6: Ποσοστό διαπερατότητας ως συνάρτηση του μήκους κύματος (Πηγή: [12] Σχήμα 4).

2.2 Παγκόσμια Εγκάρσια Μερκατορική Προβολή

Οι χαρτογραφικές προβολές (map projections) αποσκοπούν στην απεικόνιση της Γήινης επιφάνειας σε δισδιάστατο χώρο, όπως για παράδειγμα σε ένα κομμάτι χαρτί [13]. Η διατήρηση των γεωμετρικών χαρακτηριστικών του προβαλλόμενου σχήματος αποτελεί μία αρκετά δύσκολη πρόκληση κατά τη δημιουργία της απεικόνισης, για το λόγο αυτό υπάρχει πληθώρα χαρτογραφικών προβολών, η κάθε μία με τη δική της μαθηματική ανάλυση.

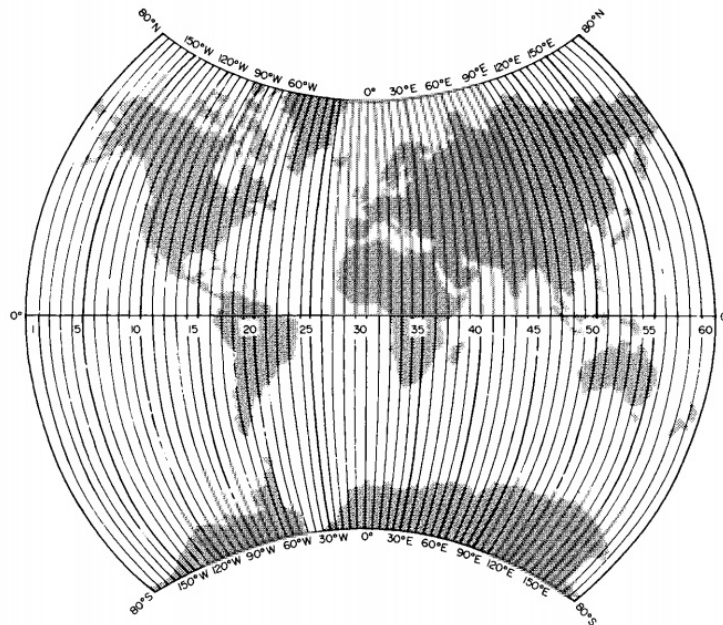
Οι προβολές συνοδεύονται από ένα σύστημα αναφοράς (Coordinate Reference System) που αντιστοιχίζει σημεία της προβολής με πραγματικά μέρη πάνω στην επιφάνεια της Γης [13].

Η Παγκόσμια Εγκάρσια Μερκατορική Προβολή, στα αγγλικά Universal Transverse Mercator (UTM), αποτελεί ένα από τα πιο διάσημα συστήματα γεωγραφικών συντεταγμένων. Ιστορικά, το σύστημα UTM υιοθετήθηκε από τον Αμερικάνικο στρατό στα τέλη της δεκαετίας του 1940 [14], χρησιμοποιείται ακόμα και σήμερα στις αποστολές του Copernicus [15]. Βασίζεται στη Μερκατορική Προβολή, μία κυλινδρική παρουσίαση χάρτη που προτάθηκε τον δέκατο έκτο αιώνα από τον Γκεράντους Μερκάτορ.

Το εν λόγω σύστημα διαχωρίζει την επιφάνεια της Γης σε εξήντα ζώνες (Σχήμα 2.7), με κάθε ζώνη να χωρίζεται ανά έξι μοίρες γεωγραφικού μήκους. Η πρώτη ζώνη έχει τη δυτικότερη της άκρη πάνω στον Αντιμεσημβρινό (180° δυτικός Μεσημβρινός) και την ανατολικότερη πάνω στον 164° δυτικό Μεσημβρινό. Συνεπώς, το κέντρο της ζώνης αυτής διέρχεται από τον 177° δυτικό Μεσημβρινό. Οι ζώνες απαριθμούνται στο σημείο τομής του κεντρικού μεσημβρινού κάθε ζώνης με τον Ισημερινό (π.χ για τη ζώνη 1, σημείο τομής 177° δυτικού Μεσημβρινού με Ισημερινό), με διεύθυνση από τα δυτικά προς τα ανατολικά και βήμα έξι μοίρες γεωγραφικού μήκους [14]. Μετά τον αριθμό της ζώνης, συμπεριλαμβάνεται ένας από τους χαρακτήρες “S” ή “N”, με σκοπό να δηλωθεί αν το σημείο βρίσκεται Βόρεια (North) ή Νότια (South) του Ισημερινού.

Αρχικά τα όρια του γεωγραφικού πλάτους ήταν από 80° B έως 80° N, όπως φαίνεται στο Σχήμα 2.7, ωστόσο τα όρια αυτά στη συνέχεια μετακινήθηκαν από 84° B έως 80° N [14].

Προκειμένου να υπολογιστεί ένα σημείο ενδιαφέροντος (Point of Interest), αρκούν τρεις μεταβλητές. Η ζώνη στην οποία βρίσκεται, η ανατολική (easting) και δυτική (northing) τιμή. Η ανατολική τιμή ορίζεται ως η απόσταση από τον κεντρικό Μεσημβρινό της ζώνης στην οποία βρίσκεται το σημείο, η δυτική ως η απόσταση από τον Ισημερινό και οι δύο τιμές μετρούνται σε μέτρα [13].



Σχήμα 2.7: Διαχωρισμός ζωνών υπό την UTM προβολή (Πηγή: [14] Σχήμα 16.06).

Ως απόρροια, αν θεωρηθεί ως σημείο ενδιαφέροντος η Αθήνα (ζώνη 34), υπολογίζοντας την ανατολική (739.545,42) και τη δυτική τιμή (4.207.529,27) μπορεί να εκφραστεί σε αναπαράσταση UTM ως 34S 739545.42 4207529.27. Προσδιορίζοντας έτσι το σημείο μοναδικά, ενώ παράλληλα μεταβιβάζοντας την ελάχιστη δυνατή πληροφορία.

2.3 Αισθητήρες

Τα όργανα που χρησιμοποιεί η Τηλεπισκόπηση για τη μέτρηση της ηλεκτρομαγνητικής ακτινοβολίας ονομάζονται Αισθητήρες. Οι αισθητήρες μοιάζουν με τον όρο κάμερα, όμως διαφοροποιούνται από αυτόν αφού μπορούν να πραγματοποιήσουν δειγματοληψία σε όλο το φάσμα της ηλεκτρομαγνητικής ακτινοβολίας. Εν αντιθέσει, στην πλειοψηφία οι κάμερες “δειγματοληπτούν” κατά κανόνα στα μήκη κύματος της ορατής ακτινοβολίας [16]. Συνήθης πρακτική, αποτελεί η κατηγοριοποίηση συγκεκριμένων ευρών του μήκους κύματος σε ζώνες (bands) συχνότητας.

Οι αισθητήρες αυτοί τοποθετούνται εναέρια σε αεροσκάφη, στο διάστημα σε δορυφόρους, είτε σε επίγειους σταθμούς [7]. Ανάλογα, τον τρόπο δειγματοληψίας οι αισθητήρες διακρίνονται σε δύο ευρείες κατηγορίες, τους παθητικούς και τους ενεργητικούς.

Ένα χαρακτηριστικό παράδειγμα αποτελεί μία φωτογραφική μηχανή, όταν ο χρήστης τραβάει μία φωτογραφία τότε η κάμερα λειτουργεί ως παθητικός αισθητήρας αφού δημιουργεί πληροφορία (φωτογραφία) με βάση τη ποσότητα της ακτινοβολίας που φτάνει στο φακό της. Η ακτινοβολία αυτή εκπέμπεται ήδη από τα αντικείμενα και **δεν την παράγει η κάμερα**. Όταν όμως οι περιστάσεις απαιτούν τη χρήση φλας, πιθανόν για κάποια νυχτερινή λήψη, τότε η κάμερα αρχίζει να λειτουργεί ως ενεργητικός αισθητήρας λόγω του ότι παράγει μόνη της την ακτινοβολία που χρειάζεται για να δημιουργήσει πληροφορία. Η κάμερα μετράει τη ποσότητα της ακτινοβολίας που επιστρέφεται από τα αντικείμενα και είναι σε θέση να δημιουργήσει την τελική πληροφορία.

2.3.1 Ενεργητικοί

Όπως αναφέρθηκε, ένας ενεργητικός αισθητήρας είναι ένα όργανο (ραντάρ ή λέιζερ) που εκπέμπει ακτινοβολία με σκοπό να μετρήσει την ακτινοβολία που θα ανακλάται από το μελετούμενο αντικείμενο (π.χ. Επιφάνεια της Γης). Η πληθώρα των ενεργητικών αισθητήρων λειτουργεί στην περιοχή των μικροκυμάτων του ηλεκτρομαγνητικού φάσματος (μήκος κύματος 1mm - 1m) [7].

Τα βασικά δομικά συστατικά ενός συστήματος ραντάρ συγκαταλέγουν έναν πομπό (transmitter), έναν δέκτη (receiver), ένα σύστημα κεραιών (antenna array) και μία συσκευή εγγραφής (recorder). Ο πομπός

μεταδίδει παλμούς μικροκυμάτων σε μία συγκεκριμένη συχνότητα, ο δέκτης δέχεται το ανακλώμενο σήμα από την κεραία και στη συνέχεια το ενισχύει ή το φιλτράρει. Το σύστημα των κεραιών μεταδίδει μία στενή δέσμη μικροκυμάτων και αποτελείται από οδηγούς (guides), που καθορίζουν τα πρότυπα για τη μετάδοση του ηλεκτρομαγνητικού κύματος. Σε πολλές περιπτώσεις, το σύστημα κεραιών χρησιμοποιείται για τη μετάδοση αλλά και τη λήψη του σήματος. Η συσκευή εγγραφής μεταφράζει το σήμα σε εικόνα και το αποθηκεύει [4].

Μία θετική πτυχή των ενεργητικών αισθητήρων αποτελεί η χρήση ενέργειας που έχει ήδη μελετηθεί και είναι γνωστή σε βάθος η συμπεριφορά της σε ποικίλες συνθήκες. Ως αποτέλεσμα, η σύγκριση της ενέργειας που μεταδίδεται και αυτής που λαμβάνεται μπορεί να πραγματοποιηθεί με μεγαλύτερη ακρίβεια [8]. Ακόμα, η αναγνώριση αεροζόλ, συννεφιάς και βροχόπτωσης είναι εφικτή με τη χρήση της ηχώς [7]. Τέλος, αφού τα όργανα παράγουν την ακτινοβολία που χρειάζονται, η δειγματοληψία είναι εφικτή κατά τη διάρκεια της μέρας αλλά και της νύχτας [4].

2.3.2 Παθητικοί

Εν αντιθέσει με τους ενεργητικούς, οι παθητικοί αισθητήρες χρησιμοποιούν ενέργεια που είναι ήδη διαθέσιμη προκειμένου να δημιουργήσουν πληροφορία. Συγκεκριμένα, αυτός ο τύπος αισθητήρα μετράει την ακτινοβολία που φτάνει από το μελετούμενο αντικείμενο στο φακό [7]. Η ισχύς της εν λόγω ακτινοβολίας συσχετίζεται σε μεγάλο βαθμό με την ποσότητα εκπομπής θερμικής ακτινοβολίας του αντικείμενου που μελετάται. Για τον λόγο αυτό, αν και λειτουργούν στο φάσμα των μικροκυμάτων, παρουσιάζουν περισσότερες ομοιότητες με τους θερμικούς αισθητήρες [8].

Κάνοντας χρήση διαφορετικών τύπων ραδιόμετρων και φασματομέτρων, οι παθητικοί αισθητήρες είναι σε θέση να ανιχνεύσουν τμήματα του φάσματος των μικροκυμάτων, εκπομπές (υπέρυθρη ακτινοβολία) και ανάκλαση (ορατό φως). Κύριο προτέρημα των οποίων αποτελεί η δυνατότητα δειγματοληψίας ολόκληρου του Πλανήτη [7].

2.4 Υπηρεσία Copernicus

Η ονομασία Copernicus προέρχεται από τον Ευρωπαϊκό επιστήμονα Νικόλαο Κοπέρνικο [17], το πρόγραμμα ήταν παλαιότερα γνωστό ως Global Monitoring for Environment and Security (GMES) [18, 19]. Η εν λόγω υπηρεσία είναι ένα πρόγραμμα που διαχειρίζεται η Ευρωπαϊκή Επιτροπή, σε συνεργασία με τα κράτη μέλη της Ε.Ε, τον Ευρωπαϊκό Οργανισμό Διαστήματος (ESA) [20], κ.α [21], με κύριο σκοπό του προγράμματος να την Παρατήρηση της Γης. Τα δεδομένα αυτά ενσωματώνονται και χρησιμοποιούνται σε υπηρεσίες που σχετίζονται με θέματα ασφάλειας και περιβάλλοντος.

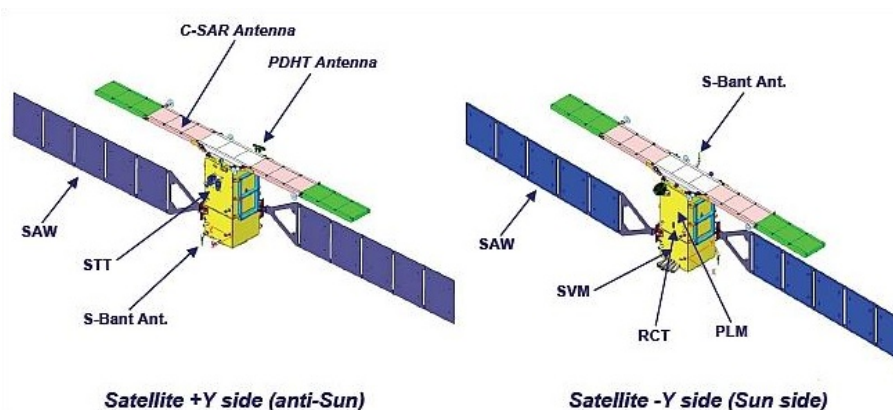
Συνολικά, έξι διαστημικές αποστολές απαρτίζουν την υπηρεσία Copernicus. Ο κάθε δορυφόρος, ανάλογα την κατηγορία στην οποία εντάσσεται, προσφέρει συγκεκριμένους τύπους δεδομένων. Συγκεκριμένα, τρεις αποστολές ασχολούνται με την παρατήρηση της επιφάνειας της Γης (Sentinel-1, Sentinel-2, Sentinel-3), δύο με την μελέτη της ατμόσφαιρας (Sentinel-4, Sentinel-5), ενώ η έκτη και τελευταία αποστολή (Sentinel-6) θα αφιερωθεί στην μελέτη της παγκόσμιας στάθμης της θάλασσας, την ωκεανογραφία και σε μελέτες για την κλιματική αλλαγή [22].

Η παρούσα μελέτη επικεντρώνεται σε δεδομένα εδάφους, συνεπώς οι επόμενες Ενότητες (2.4.1 - 2.4.2) θα αφιερωθούν στην περιγραφή των αποστολών (Sentinel-1, Sentinel-2) που προσφέρουν τέτοια δεδομένα. Παρατίθεται μία σύντομη ιστορική αναδρομή για κάθε αποστολή, περιγραφή των δορυφόρων και των διαθέσιμων λειτουργιών, ενώ η θεματική ενότητα φτάνει στο πέρας της με την αποτύπωση των τύπων δεδομένων που είναι προσβάσιμα από το κοινό.

2.4.1 Αποστολή Sentinel-1

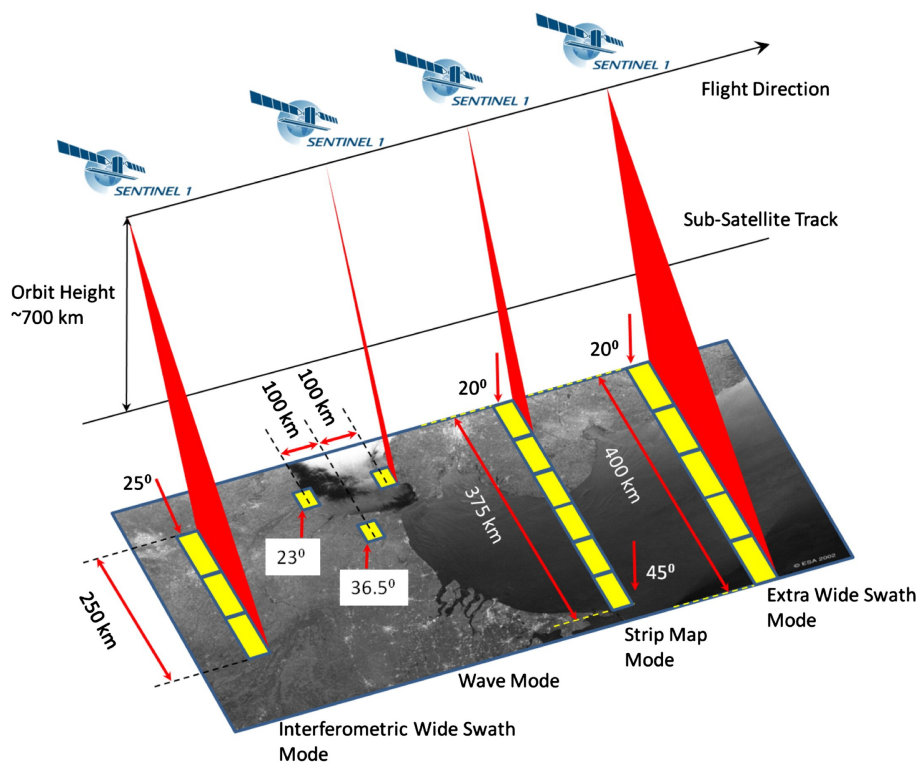
Η αποστολή Sentinel-1 αποτελείται από δύο δίδυμους δορυφόρους, τον δορυφόρο Sentinel-1A και τον δορυφόρο Sentinel-1B, οι οποίοι μπήκαν σε τροχιά στις 3 Απριλίου 2014 και στις 25 Απριλίου 2016, αντίστοιχα [22]. Οι δορυφόροι μοιράζονται το ίδιο τροχιακό επίπεδο και η αρχιτεκτονική τους μπορεί να παρατηρηθεί καλύτερα στο Σχήμα 2.8. Βασικός εξοπλισμός των δορυφόρων αποτελεί ένα Ραντάρ Συνθετικού Ανοίγματος (SAR), κύρια εφαρμογή αυτού του τύπου ραντάρ αποτελεί η ανακατασκευή αντικειμένων σε δισδιάστατες ή τρισδιάστατες εικόνες [23], προκειμένου να δειγματοληπτηθούν δεδομένα

στεριάς ή θάλασσας οποιαδήποτε στιγμή της ημέρας ή της νύχτας, χωρίς να επηρεάζονται σε σημαντικό βαθμό από τα καιρικά φαινόμενα [19].



Σχήμα 2.8: Περιγραφή Sentinel-1 [24].

Οι δορυφόροι κάνουν χρήση C-band SAR με κεντρική συχνότητα τα 5.405 Ghz και δειγματοληπτούν στις ακόλουθες λειτουργίες: **Stripmap (SM)**, **Interferometric Wide swath (IW)**, **Extra-Wide swath (EW)**, **Wave mode (WV)** [25]. Οι προαναφερόμενες λειτουργίες απεικονίζονται στο Σχήμα 2.9.



Σχήμα 2.9: Λειτουργίες Sentinel-1 (Πηγή [25] Σχήμα 1).

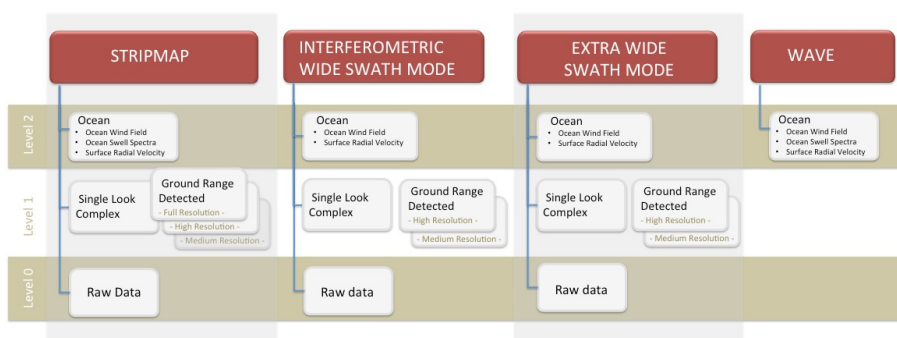
Αναλυτικότερα, η λειτουργία **Stripmap (SM)** [26] προσφέρει μετρήσεις μήκους 80 χιλιομέτρων σε ανάλυση 5 μέτρων επί 5 μέτρων. Οι μετρήσεις αυτές χωρίζονται σε 6 τμήματα, κάθε ένα από τα οποία μπορούν να επιλεγθούν αλλάζοντας την γωνία πρόσπτωσης (beam incidence angle) ή το εύρος της δέσμης (elevation beamwidth) [25].

Η λειτουργία **Interferometric Wide Swath Mode (IW)** αποτελεί την προεπιλεγμένη μέθοδο απόκτησης δεδομένων στεριάς. Κατά τη λειτουργία αυτή, ο δορυφόρος κάνει χρήση της τεχνικής Terrain Observation with Progressive Scans SAR (TOPSAR) που βασίζεται στο φαινόμενο Ντόπλερ (Doppler) [27], αντί της παραδοσιακής ScanSAR. Οι μετρήσεις πραγματοποιούνται σε ανάλυση 5 μέτρων επί 20 μέτρων, για αρκετά μεγάλο μήκος (250 χιλιόμετρα) [28].

Η λειτουργία **Extra Wide Swath Mode (EW)** παρέχει μετρήσεις μήκους έως 400 χιλιόμετρα με ανάλυση 20 μέτρων επί 40 μέτρων. Στη συγκεκριμένη περίπτωση ενσωματώνεται η τεχνική TOPSAR, όπως στην λειτουργία IW και ως απόρροια η επιλογή EW μπορεί κάλλιστα να λειτουργήσει για συμβολομετρία (interferometry) [25]. Κύριες χρήσεις της EW περιλαμβάνουν επίβλεψη θαλάσσιων εκτάσεων, ναυτικών δρομολογίων καθώς και των πολικών περιοχών με κύριους σκοπούς την εξαγωγή στατιστικών για το περιβάλλον (λ.χ λιώσιμο των πάγων), παγκόσμια ασφάλεια (π.χ έγκαιρη διάγνωση ναυαγίων πετρελαίου) [29].

Τέλος, η **Wave Mode (WV)** αποτελείται από εικόνες 20 χιλιομέτρων επί 20 χιλιομέτρων, κάθε 100 χιλιόμετρα που λαμβάνονται εναλλασσόμενα από δύο διαφορετικές γωνίες πρόσπτωσης. Οι κυματομορφές που λαμβάνονται από την ίδια γωνία πρόσπτωσης απέχουν 200 χιλιόμετρα [30]. Χρησιμοποιείται ευρέως για την παρατήρηση των κυμάτων του ωκεανού [25].

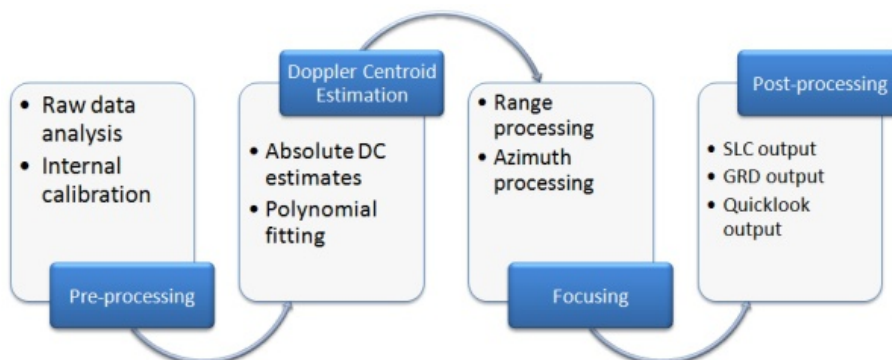
Κάθε μία από τις λειτουργίες που περιγράφηκαν παραπάνω (SM, IW, EW, WV) προσφέρουν 3 Τύπους Δεδομένων (Level0, Level1, Level2), όπως απεικονίζονται στο Σχήμα 2.10. Συγκεκριμένα, για τις λειτουργίες SM, IW, EW είναι διαθέσιμα δεδομένα μονής πόλωσης (HH/VV) ή διπλής πόλωσης (VV+VH/HH+HV) αλλά για την λειτουργία WV διατίθενται μόνο δεδομένα μονής πόλωσης (HH/VV) [31].



Σχήμα 2.10: Προϊόντα Sentinel-1 (Πηγή [31] Σχήμα 1).

Τα δεδομένα **Επιπέδου-0 (Level0)** είναι διαθέσιμα για τις λειτουργίες SM, IW, EW και αποτελούνται από ωμά SAR δεδομένα (SAR raw data). Θεωρούνται η βάση για την εξαγωγή δεδομένων των υψηλότερων επιπέδων (Level1, Level2) και συγκαταλέγουν πληροφορία σχετικά με την τροχιά και τις εσωτερικές μετρήσεις του δορυφόρου, ενώ δεν απουσιάζει ο θόρυβος. Η συμπίεση πραγματοποιείται με τη χρήση του αλγορίθμου Dynamic Block Adaptive Quantization (FDBAQ) [32], προκειμένου να χρησιμοποιηθούν τα ωμά δεδομένα αρχικά πρέπει να αποσυμπίεστούν και στην συνέχεια να διοχετευθούν σε έναν επεξεργαστή SAR [31, 33].

Εφαρμόζοντας μία σειρά αλγορίθμων στα δεδομένα Επιπέδου-0 (Σχήμα 2.11), λαμβάνονται τα δεδομένα **Επιπέδου-1**. Τα προαναφερόμενα, είναι διαθέσιμα για τις λειτουργίες SM, IW, EW και χωρίζονται σε δύο ευρείες κατηγορίες: **Single Look Complex (SLC)** και **Ground Range Detected (GRD)** [34].



Σχήμα 2.11: Επεξεργασία επιπέδου-0 σε επίπεδο-1 [35].

Η κατηγορία **Single Look Complex (SLC)** περιλαμβάνει εστιασμένα SAR δεδομένα, τα οποία μπορούν να συσχετιστούν σε ένα επίγειο σύστημα γεωγραφικών συντεταγμένων χρησιμοποιώντας πληρο-

φορίας τροχιάς και στάσης του δορυφόρου. Προκειμένου να ληφθεί το τελικό προϊόν, δειγματοληπτείται κάθε διάσταση μία φορά (Single Look). Κατά την δειγματοληψία λαμβάνονται υπ' όψιν πραγματικά, φανταστικά δείγματα καθώς και όλη το εύρος του σήματος, διατηρώντας τις πληροφορίες φάσης [31, 34].

Η κατηγορία **Ground Range Detected (GRD)** περιλαμβάνει επίσης εστιασμένα SAR δεδομένα, τα οποία έχουν δειγματοληπτηθεί περισσότερο από μία φορές (Multi Look) και έχουν αντιστοιχιστεί σε επίγειες συντεταγμένες χρησιμοποιώντας ένα μοντέλο ελλειψοειδούς Γης, χωρίς να διατηρηθούν πληροφορίες φάσης. Τα τελικά δεδομένα, εμφανίζουν λιγότερα στίγματα (speckles), ωστόσο παρέχουν χειρότερη ανάλυση. Προσφέρονται τρεις υποκατηγορίες σχετικά με την ανάλυση: Πλήρης Ανάλυση (FR), Υψηλή Ανάλυση (HR) και Μεσαία Ανάλυση (MR). Για τις επιλογές IW, EW διατίθενται οι αναλύσεις MR, HR, για την WV μόνο ανάλυση MR και τέλος για την επιλογή SM διατίθενται όλες οι αναλύσεις (FR, HR, MR) [31, 34].

Στο τελευταίο επίπεδο, βρίσκονται τα δεδομένα **Επιπέδου-2 (Level-2)** που προκύπτουν από περαιτέρω επεξεργασία του Επιπέδου-1. Χωρίζονται σε 3 κατηγορίες συνιστώσων, ονομαστικά **Ocean Swell spectra (OSW)**, **Surface Radial Velocity (RVL)** και **Ocean Wind field (OWI)** [36]. Η συνιστώσα **OSW** αποτελεί μία δισδιάστατη απεικόνιση του φάσματος της επιφάνειας του ωκειανού, προσφέρει εκτιμήσεις για την κατεύθυνση και την ταχύτητα του ανέμου ανά φάσμα (swell spectrum) και προκύπτει από την επεξεργασία των λειτουργιών SM, WV. Η **RVL** υπολογίζεται ως η διαφορά μεταξύ των πλεγμάτων Doppler του Επιπέδου-2 και Επιπέδου-1. Η συνιστώσα **OWI** υπολογίζεται σε απόσταση 10 μέτρα από την επιφάνεια της Γης για τις λειτουργίες SM, IW, EW και αποτελεί μία εκτίμηση της ταχύτητας και της κατεύθυνσης του ανέμου [31, 36].

Ο κάθε δορυφόρος παρέχει μέση περίοδο επισκεψιμότητας δώδεκα μέρες, ενώ ο συνδυασμός των δύο δορυφόρων προσφέρει περίοδο επανεξέτασης έξι μέρες στον Ισημερινό. Ωστόσο είναι σημαντικό να ειπωθεί ότι η απόσταση μεταξύ των τροχιών των δύο δορυφόρων ποικίλει ανάλογα το γεωγραφικό πλάτος στο οποίο βρίσκονται, για αυτό το λόγο η περίοδος των έξι ημερών δεν μπορεί να είναι παντού σταθερή [37].

2.4.2 Αποστολή Sentinel-2

Η αποστολή Sentinel-2, όπως και η Sentinel-1, αποτελείται από δύο δίδυμους δορυφόρους. Ο δορυφόρος Sentinel-2A αποτελεί την πρώτη εκτόξευση του προγράμματος στις 22 Ιουνίου 2015, ενώ ο δίδυμος του, Sentinel-2B μπήκε σε τροχιά στις 7 Μαρτίου του 2017. Αμφότεροι, προσφέρουν δεδομένα υψηλής ανάλυσης σχετικά με την μορφολογία του εδάφους, επιπρόσθετα τους δορυφόρους αυτούς χρησιμοποιούν υπηρεσίες έκτακτης ανάγκης [22].



Σχήμα 2.12: Δορυφόρος Sentinel-2 (Πηγή [38] Σχήμα 1).

Οι δίδυμοι δορυφόροι, όπως απεικονίζονται στο Σχήμα 2.12, έχουν διαφορά φάσης 180° μεταξύ τους και έχουν σχεδιαστεί για να προσφέρουν υψηλής ανάλυσης εικόνες σε συνδυασμό με υψηλή συχνότητα δειγματοληψίας (περίοδος πέντε μέρες στον Ισημερινό). Η δειγματοληψία πραγματοποιείται σε δεκατρία κανάλια, τέσσερα κανάλια στα 10 μέτρα, έξι στα 20 μέτρα και τρία στα 60 μέτρα ανάλυσης [39]. Κύριες χρήσεις των δεδομένων από την αποστολή Sentinel-2 περιλαμβάνουν παρακολούθηση νερού, δασών,

φυσικών πόρων, καλλιεργειών και βλάστησης, ενώ διεργασίες για χωροταξικό σχεδιασμό μπορούν ακόμα να επωφεληθούν [40].

Η αρχιτεκτονική προσφέρει οπτικό πεδίο (Field of View) 290 χιλιομέτρων, ενώ στη δειγματοληψία προσφέρει μετρήσεις το εξειδικευμένο πολυφασματικό όργανο (MultiSpectral Instrument) [38].

Το MSI συλλέγει την ανακλώμενη ακτινοβολία της Γης σε δεκατρία κανάλια από **Visible και Near Infrared (VNIR)** έως **Short Wavelength Infrared (SWIR)**, καθώς ο δορυφόρος κινείται. Στον Πίνακα 2.1 παρουσιάζονται αναλυτικά τα κεντρικά μήκη κύματος καθώς και το εύρος ζώνης για κάθε ένα από τα δεκατρία κανάλια του κάθε δορυφόρου. Λόγω της γεωμετρικής του διάταξης, οι παρατηρήσεις της γήινης επιφάνειας διαφέρουν μεταξύ των φασματικών ζωνών [41].

Πίνακας 2.1: Φασματικές Ζώνες για Δορυφόρους Sentinel-2A (S2A) και Sentinel-2B (S2B) [41].

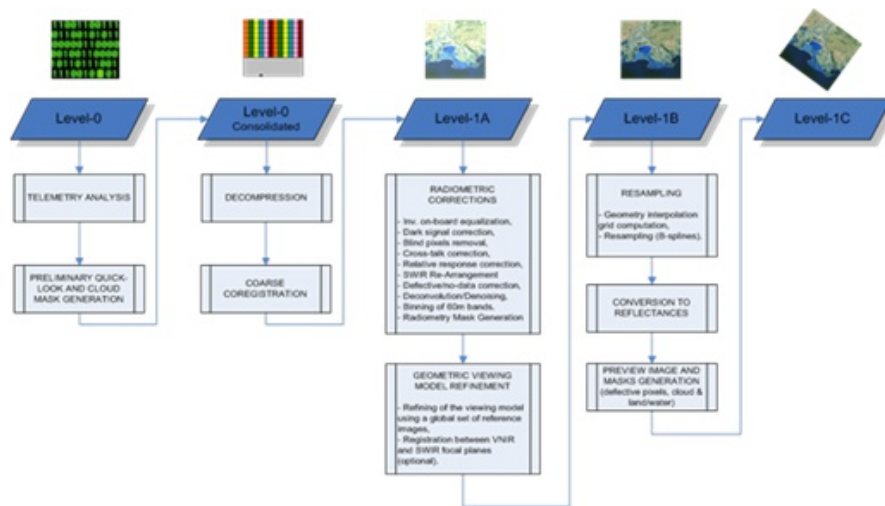
Κανάλι	Κέντρο Μήκους Κύματος (nm)		Εύρος Ζώνης (nm)		Ανάλυση (m)
	S2A	S2B	S2A	S2B	
1	442.7	442.3	21	21	60
2	492.4	492.1	66	66	10
3	559.8	559.0	36	36	10
4	664.6	665.0	31	31	10
5	704.1	703.8	15	16	20
6	740.5	739.1	15	15	20
7	782.8	779.7	20	20	20
8	832.8	833.0	106	106	10
8a	864.7	864.0	21	22	20
9	945.1	943.2	20	21	60
10	1373.5	1376.9	31	30	60
11	1613.7	1610.4	91	94	20
12	2202.4	2185.7	175	185	20

Όπως η Αποστολή Sentinel-1, έτσι και η αποστολή Sentinel-2 προσφέρει τρεις τύπους δεδομένων. Διαθέσιμα προς το χρήστη είναι μόνο τα δεδομένα Level-1C και Level-2A, τα Level-2A μπορούν είτε να παραχθούν από το χρήστη με το λογισμικό Sentinel-2 Toolbox [42], είτε κατευθείαν από τον δορυφόρο αξιοποιώντας το Payload Data Ground Segment (PDGS) [43]. Τα στάδια επεξεργασίας κάθε επιπέδου προς το επόμενο, απεικονίζονται καλύτερα στο Σχήμα 2.13.

Συγκεκριμένα, κατά το πρώτο **Επίπεδο (Level-0)** εμπεριέχονται συμπίεσμένα χωρίς επεξεργασία (raw) δεδομένα σε μορφή ISP [45]. Το επίπεδο αυτό αποτελεί τη βάση παραγωγής των επομένων, συγκαταλέγει πληροφορίες σχετικά με πακέτα βοηθητικών πηγών, μεταδεδομένα καθώς και σχολιασμούς των ISP που αντιστοιχούν στις συμπίεσμένες εικόνες [46].

Το **Επίπεδο 1 (Level-1)**, χωρίζεται σε τρεις υποκατηγορίες **Level-1A, Level-1B, Level-1C**. Στο κοινό γίνονται διαθέσιμα μόνο τα δεδομένα **Level-1C**. Το **Level-1A** λαμβάνεται εφαρμόζοντας αποσυμπίεση στα “ωμά” δεδομένα του Level-0 [47]. Κατά το επίπεδο **Level-1B** πραγματοποιούνται ραδιομετρικές διορθώσεις, όπως διόρθωση διασταύρωσης συνδιαλέξεων (crosstalk), ελάττωση παρεμβολών των εικονοστοιχείων, κ.α [48]. Τα δεδομένα **Level-1C** επαναδειγματοληπτούνται με σταθερή απόσταση δειγματοληψίας (Στήλη Ανάλυση Πίνακας 2.1) αναλόγως το κανάλι, συγκαταλέγονται μετεωρολογικές μετρήσεις/προβλέψεις καθώς και μάσκες συννεφιάς [49]. Οι μάσκες αυτές προσφέρουν πληροφόρηση σχετικά με το ποσοστό συννεφιάς του κάθε εικονοστοιχείου και τον τύπο του [50].

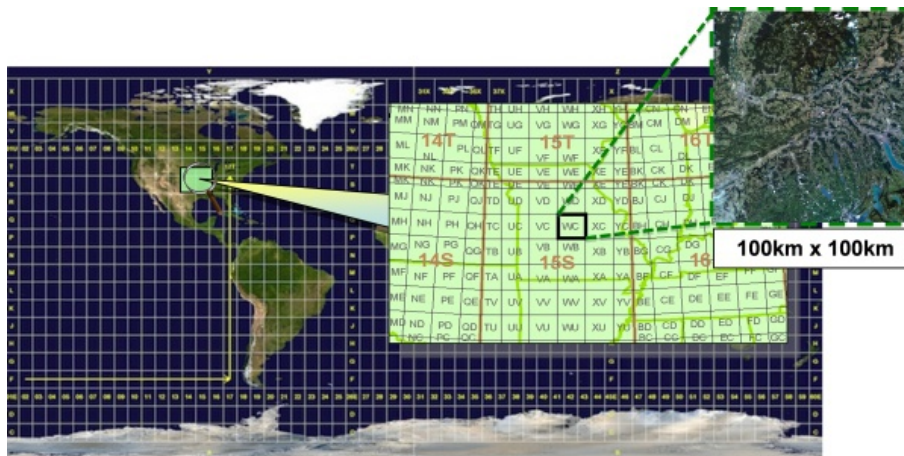
Επεξεργάζοντας περαιτέρω τα Level-1C, προκύπτουν τα Level-2A. Αμφότερα, έχουν ανάλυση 100



Σχήμα 2.13: Προϊόντα Sentinel-2 [44].

επί 100 τετραγωνικά χιλιόμετρα σε προβολή UTM [49]. Τα παραπροϊόντα των Level-2A χρησιμοποιούνται ως χάρτες οπτικού πάχους των αιωρούμενων σωματιδίων (Aerosol Optical Thickness), χάρτες υδρατμών και για την μέτρηση της επιφανειακής ανάκλασης στα χαμηλότερα επίπεδα της ατμόσφαιρας (Bottom of the Atmosphere) [51].

Τα δεδομένα διατίθενται σε κομμάτια (granules) σταθερού μεγέθους, μέσα σε μία τροχιά. Τα κομμάτια αυτά, είναι το ελάχιστο αδιαίρετο υποπροϊόν. Για τα Level-1C, Level-2A τα κομμάτια ταυτίζονται με την έννοια ενός τετράγωνου (100 επί 100 τετραγωνικών χιλιομέτρων) προβολή UTM (Ενότητα 2.2). Κατά την επιλογή περιοχής ενδιαφέροντος, επιστρέφονται όλα τα κομμάτια που εμπεριέχονται μέσα σε αυτήν (Σχήμα 2.14) [15].



Σχήμα 2.14: Διαχωρισμός προϊόντων Level-1C, Level-2A [52].

Η αποστολή Sentinel-2 παρέχει κάλυψη σε χερσαίες επιφάνειες μεταξύ γεωγραφικού πλάτους 56°N και 84°B, αυτό συμπεριλαμβάνει τα ύδατα σε έκταση έως είκοσι χιλιομέτρων από την ακτή, σε όλα τα νησιά της Ευρωπαϊκής Ένωσης, σε κλειστές θάλασσες και τέλος σε νησιά των οποίων η έκταση είναι μεγαλύτερη των εκατό τετραγωνικών χιλιομέτρων [53].

Μέση περίοδος επισκεψιμότητας των περιοχών ορίζονται οι πέντε μέρες, ενώ λόγω της αλληλεπικάλυψης των τροχών των δύο δορυφόρων, ορισμένες περιοχές θα καταγράφονται συχνότερα υπό διαφορετικές οπτικές γωνίες [53].

Κεφάλαιο 3

Βαθιά Μάθηση

Όπως αναφέρει ο Ian Goodfellow και οι συνεργάτες του, η Βαθιά Μάθηση αποτελεί ένα είδος Μηχανικής Μάθησης που επιτυγχάνει με αρκετή δύναμη και ευελιξία, μαθαίνοντας να αποτυπώνει τον κόσμο ως μία εμφωλευμένη ιεραρχία από σενάρια. Κάθε πολύπλοκο σενάριο περιγράφεται ως συνδυασμός απλούστερων, ενώ σύνθετες αναπαραστάσεις υπολογίζονται με βάση απλοποιημένων όρων [54]. Η εκρηκτική αύξηση της επίδοσης της Βαθιάς Μάθησης υποβοηθήθηκε τα τελευταία χρόνια από τον τεράστιο όγκο των διαθέσιμων δεδομένων καθώς και τις βελτιώσεις των υπολογιστικών συστημάτων.

Ο Geoffrey E. Hinton στο [55] αναλύει το παράδοξο λόγο που τα οπτικά πεδία αυξάνονται στα υψηλότερα επίπεδα του οπτικού φλοιού. Εξηγεί ότι στα υψηλότερα αυτά επίπεδα, μπορούν να αναπαρασταθούν περισσότερες και συνθετότερες οντότητες, ωστόσο το πλήθος αυτών ελαττώνεται με αποτέλεσμα η αναγνώριση τους να γίνεται με μεγαλύτερη ακρίβεια.

Επιπρόσθετα, με βάση το παραπάνω βιβλίο [54], η ιστορική αναδρομή της Βαθιάς Μάθησης χωρίζεται σε τρία επιμέρους κύματα: τη Κυβερνητική (cybernetics), το Συνδεδετισμός (connectionism) και τελικά τη Βαθιά Μάθηση (Deep Learning). Ακόμα, συχνός είναι ο όρος Τεχνητά Νευρωνικά Δίκτυα (Artificial Neural Networks; ANNs) στο πλαίσιο της Βαθιάς Μάθησης, αφού οι πρώτοι αλγόριθμοι σχεδιάστηκαν με βάση τη διαδικασία μάθησης του εγκεφάλου [54].

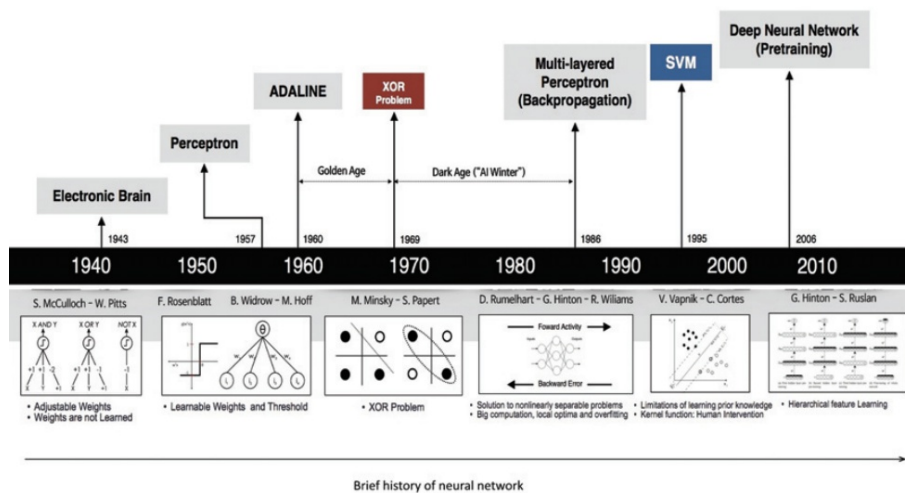
Κατά το πρώτο κύμα της Κυβερνητικής (1940-1960), η αρχή της μοντελοποίησης ξεκίνησε έχοντας ως πρότυπο τη βιολογική μάθηση ή αλλιώς προσπαθώντας να γίνει κάποιος μορφής αναπαράστασης των νευρώνων σε μαθηματικά μοντέλα. Το 1943, εισάγεται από τους Mc Culloch και Pitts, το πρώτο τέτοιο μοντέλο [56]. Ο Frank Rosenblatt (1957), δημοσιεύει την πρώτη έκδοση του αλγορίθμου Perceptron [57], ο αλγόριθμος αυτός είναι ένας δυαδικός ταξινομητής δηλαδή έχοντας ως είσοδο χαρακτηριστικά μπορεί να προβλέψει αν η είσοδος ανήκει ή όχι σε μία κλάση. Ο Perceptron πάσχει από αδυναμία ταξινόμησης αντικειμένων που δε χωρίζονται γραμμικά (λ.χ συνάρτηση XOR). Επίσης, αξιοσημείωτο αποτελεί το δίκτυο ADALINE, το οποίο βασίζεται στην προηγούμενη δουλειά των McCulloch και Pitts και το MADALINE, που δημιουργείται χρησιμοποιώντας πολλαπλά ADALINE.

Το δεύτερο κύμα του Συνδεδετισμού (1980-2000) συμβάλει ίσως με τη σημαντικότερη πρόταση, που χρησιμοποιείται με λίγες παραλλαγές ακόμα και σήμερα σε διάφορες αρχιτεκτονικές. Λόγος γίνεται για τη μέθοδο εκπαίδευση με οπίσθια διάδοση (backpropagation) [54], η οποία χρονολογείται στο 1960. Γνωστή έγινε όμως κυρίως στο [58] (1986) από τους D. Rumelhart, Geoffrey E. Hinton, R. J. Williams και από τον LeCun [59] όταν εκπαίδευσε το πρώτο ΣΝΔ (Ενότητα 3.2).

Τέλος, σημαντικότερα ευρήματα κατά την περίοδο της Βαθιάς Μάθησης (2006-) αποτελούν οι Βαθιές Μηχανές Boltzmann [60] (2006), το AlexNet [61] (2012), το R-CNN [62] (2013), το VGG [63] (2014), τα GAN [64] (2014), η τεχνική dropout [65] (2014), το ResNet [66] (2015), το U-Net [67] (2015) και Transformers [68] (2017).

Η παραπάνω ιστορική αναδρομή απεικονίζεται καλύτερα στο Σχήμα 3.1, το οποίο δεν περιλαμβάνει σημαντικές αρχιτεκτονικές που προτάθηκαν μετά το 2010.

Στις επόμενες ενότητες, παρουσιάζονται τα πιο δημοφιλή μοντέλα των τελευταίων ετών καθώς και συνθετότερες, πιο βαθιές αρχιτεκτονικές που ενσωματώνουν τα προαναφερόμενα. Στην Ενότητα 3.5, γίνεται μία περιγραφή της αρχιτεκτονικής R-CNN και των προτεινόμενων επεκτάσεων του, που έχουν γνωρίσει ιδιαίτερη επιτυχία σε προβλήματα Όρασης Υπολογιστών (Computer Vision).

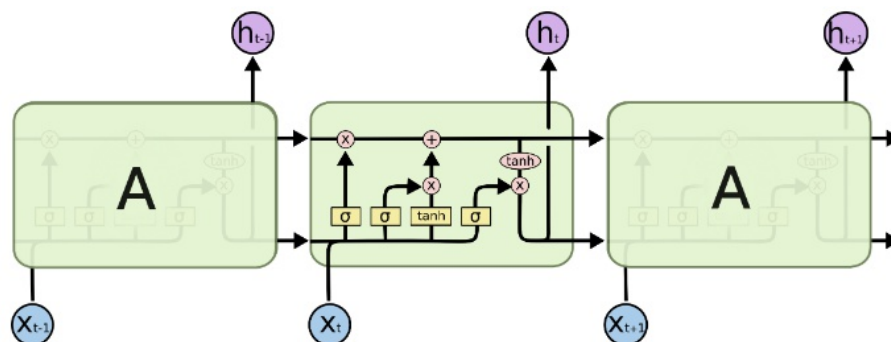


Σχήμα 3.1: Ιστορία Βαθιάς Μάθησης [69].

3.1 Αναδρομικό Νευρωνικό Δίκτυο (RNN)

Το πρώτο είδος Αναδρομικών Νευρωνικών Δικτύων, μία αρχική έκδοση των δικτύων Hopfield προτάθηκαν το 1970 από τους Wilhelm Lenz, Ernst Ising [70], απέκτησε δημοσιότητα όμως από τον John Hopfield το 1982 [71]. Πρώτη αναφορά σε Αναδρομικά Νευρωνικά Δίκτυα (RNN) έγινε το 1986 από τον D. Rumelhart και τους συνεργάτες του [58]. Σήμερα, ευρέως γνωστά στον κλάδο της Βαθιάς Μάθησης αποτελούν τα Μακράς και Βραχείας Διάρκειας Δίκτυα (Long Short Term Memory; LSTMs) που εισήχθησαν το 1997 [72].

Τα τελευταία ακόμα προσφέρουν λύσεις σε αρκετά προβλήματα αν και στην πληθώρα των λύσεων αυτών συνεργάζονται με άλλες δημοφιλείς αρχιτεκτονικές. Έχουν προταθεί αρκετές τροποποιήσεις και εναλλακτικές προσεγγίσεις, όμως στον πυρήνα τους τα δίκτυα αυτά αποτελούνται από τα ίδια δομικά υλικά (Σχήμα 3.2). Συναντώνται κυρίως σε μορφή αλυσίδας, δηλαδή η έξοδος του ενός τροφοδοτεί την είσοδο του επομένου. Καλούνται να λύσουν προβλήματα όπου τα δεδομένα βρίσκονται σε χρονοσειρές, αφού έχουν σχεδιαστεί να “διατηρούν” μνήμη σε πληροφορία που έχει παρέλθει, δημιουργώντας με αυτόν τον τρόπο πρότυπα στο χώρο του χρόνου.



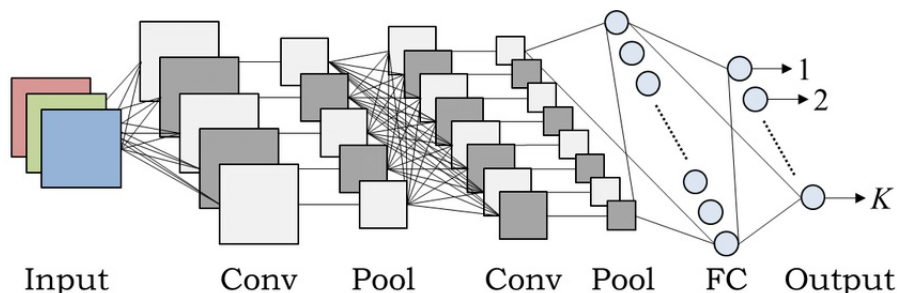
Σχήμα 3.2: Αρχιτεκτονική LSTM [73].

Τα πρώτα RNN δίκτυα μοιάζουν με αυτά του Σχήματος 3.2, με την ειδοποιός διαφορά ότι ήταν αρκετά πιο απλοϊκά διότι απαρτιζόνταν μόνο από τη συνάρτηση tanh. Ο κύριος τρόπος δράσης των LSTMs μπορεί να αντιληφθεί θεωρώντας μία νοητή ευθεία γραμμή, η οποία διέρχεται από την αρχή προς το τέλος της αλυσίδας μεταβιβάζοντας γνώση. Η γραμμή αυτή ενημερώνεται σε κάθε επιμέρους κελί, το κελί είναι υπεύθυνο να διαγράψει ή να προσθέσει δεδομένα στη γραμμή [73].

Αξιοσημείωτες εναλλακτικές των LSTMs αποτελούν τα LSTM με “peerhole connections” [74] και τα Επαναλαμβανόμενα Δίκτυα με Πύλες (Gated Recurrent Unit; GRUs) [75].

3.2 Συνελκτικό Νευρωνικό Δίκτυο (CNN)

Η πρώτη μορφή Συνελκτικού Νευρωνικού Δικτύου, το neocognitron [76] προτάθηκε το 1980 από τον Kunihiko Fukushima και βασιζόταν σε προηγούμενες δουλειές των Hubel και Wiesel. Εδραιώθηκε όταν ο LeCun πραγματοποίησε εκπαίδευση με οπίσθια διάδοση (backpropagation) το 1989 [59]. Όπως τα LSTMs, έτσι και τα CNNs μπορούν να διαφέρουν αρκετά μεταξύ των υλοποιήσεων, όμως τα βασικά δομικά τους μέλη είναι τα ίδια.



Σχήμα 3.3: Αρχιτεκτονική CNN (Πηγή [77] Σχήμα 1).

Λαμβάνοντας ως γνώμονα το Σχήμα 3.3, τα συνελκτικά νευρωνικά δίκτυα μπορούν να χωριστούν σε τέσσερα επίπεδα, το επίπεδο των συνελιξεων (convolutions), το επίπεδο της υποδειγματοληψίας (pooling), το επίπεδο της συνάρτησης ενεργοποίησης (activation function) και αυτό της ταξινόμησης (classification).

Αρχικά, ως γίνεται η υπόθεση ότι στο δίκτυο εισάγεται μία εικόνα με ένα μόνο κανάλι (δηλαδή ασπρόμαυρη), ένα πλήθος φίλτρων (filters) ή πυρήνων (kernels) εφαρμόζεται κατά μήκος της εικόνας, υπολογίζοντας το εσωτερικό γινόμενο μεταξύ της εικόνας και του φίλτρου. Η εφαρμογή των φίλτρων μπορεί να ερμηνευτεί ως ένα μικρό παράθυρο, το οποίο διαπερνά την εικόνα. Τα παράθυρα μπορούν να έχουν επικάλυψη ή και όχι. Η παραπάνω διαδικασία αποτελεί το επίπεδο των συνελιξεων.

Όπως γίνεται αντιληπτό αν τα φίλτρα έχουν μεγάλο βάθος ή χρησιμοποιούνται διαδοχικά πολλαπλά τέτοια επίπεδα, η διάσταση του παραγόμενου τανυστή ή χάρτη χαρακτηριστικών (feature map), του είναι αρκετά μεγάλη. Για να λυθεί αυτό το πρόβλημα, συνήθη είναι η εφαρμογή ενός επιπέδου υποδειγματοληψίας, υπάρχουν πολλές μέθοδοι υποδειγματοληψίας στις περισσότερες περιπτώσεις χρησιμοποιείται αυτή της μέγιστης υποδειγματοληψίας (max pooling). Η τελευταία, εφαρμόζεται όπως το φίλτρο στις συνελιξεις με τη διαφορά ότι δεν υπολογίζει το εσωτερικό γινόμενο αλλά διατηρεί το μεγαλύτερο στοιχείο που βρίσκεται στο αντίστοιχο παράθυρο κάθε χρονική στιγμή, πραγματοποιώντας έτσι ελάττωση των διαστάσεων του αρχικού τανυστή.

Το επίπεδο της συνάρτησης ενεργοποίησης στις περισσότερες περιπτώσεις αποτελείται από ReLU ($\max(0, x)$), αφού εκπαιδεύει τα δίκτυα γρηγορότερα λόγω της απλότητας της και οδηγεί σε λιγότερες ποινές κατά την κανονικοποίηση.

Σε βαθιές αρχιτεκτονικές όπως το AlexNet (2012) [61], το VGG (2014) [63], το ResNet (2015) [66] εναλλάσσονται τα παραπάνω επίπεδα πολλαπλές φορές.

Εφόσον παρέλθουν τα προαναφερόμενα επίπεδα και έχουν προκύψει τα αντίστοιχα διανύσματα τότε ενσωματώνεται ένα τελικό επίπεδο συνδεδεμένων νευρώνων προκειμένου να ληφθεί η τελική πρόβλεψη.

3.3 Παραγωγικό Αντιπαραθετικό Δίκτυο (GAN)

Η μεθοδολογία εκμάθησης ενός Παραγωγικό Αντιπαραθετικό Δικτύου (Generative Adversarial Network; GAN) [64] προτάθηκε από τον Ian Goodfellow και τους συνεργάτες του το 2014. Στην εν λόγω δημοσίευση, προτείνεται η εκπαίδευση παραγωγικών μοντέλων με χρήση της Αντιδικητικής διαδικασίας (Adversarial Process). Η μέθοδος αυτή περιλαμβάνει την εκπαίδευση δύο μοντέλων, ενός παραγωγικού (Generative) μοντέλου G και ενός διαχωριστικού (discriminative ή conditional) μοντέλου D .

Η εκπαίδευση των δύο μοντέλων εξηγείται από την Εξίσωση 3.1 και μπορεί να περιγραφεί ως ένα παιχνίδι μεταξύ δύο αντιπάλων, κατά το οποίο ο παίχτης G προσπαθεί να παράξει δείγματα τόσο ρεαλιστικά ώστε να ξεγελάσει τον παίχτη D . Ο G εκπαιδεύεται με βάση πόσα από τα δείγματα του ξεγελάνε

τον D , όταν ο τελευταίος δεν μπορεί να απαντήσει με βεβαιότητα για την αυθεντικότητα των δειγμάτων που λαμβάνει, τότε θεωρείται ότι ο G προσεγγίζει την κατανομή των δειγμάτων για την οποία έχει εκπαιδευτεί.

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))] \quad (3.1)$$

Στην πράξη, η εκπαίδευση τέτοιων δικτύων είναι αρκετά δυσκολότερη και συνήθως γίνεται σε δύο στάδια. Αρχικά, εκπαιδεύεται ο D πάνω σε ένα σύνολο δεδομένων, προκειμένου να μάθει τα χαρακτηριστικά των αντικειμένων που θα διαχωρίζει σε ψεύτικα (0) ή αληθινά (1). Συνεπώς, ο D είναι ένας δυαδικός ταξινομητής που εκτός της τελικής απόφασης (0/1), παράγει επίσης απώλειες (losses) για την επίδοση του G αλλά και για τον εαυτό του D . Οι απώλειες του D χρησιμοποιούνται από τον ίδιο για τη ρύθμιση των βαρών του, ενώ οι απώλειες του D για τον G αγνοούνται και χρησιμοποιούνται μόνο κατά την εκπαίδευση του τελευταίου.

Από την άλλη, η εκπαίδευση του G είναι συνθετότερη διαδικασία. Αρχικά, εφόσον τα ΠΑΔ παράγουν “νέα” δεδομένα, λογικό είναι να μην απαιτείται κάποια είσοδος δεδομένων στον G . Για το λόγο αυτόν, δειγματοληπτείται από κατανομή (συνήθως Γκαουσιανή) τυχαίο δείγμα z , το δείγμα αυτό χαρακτηρίζεται ως “θόρυβος”. Ο G χρησιμοποιεί αυτό το τυχαίο δείγμα για να παράξει κάτι ρεαλιστικό ή δηλαδή να προσεγγίσει την κατανομή των δεδομένων που δέχεται το μοντέλο για εκπαίδευση. Εφόσον παραχθούν τα δείγματα, συμψηφίζονται με τα αληθινά δεδομένα και τροφοδοτούνται στον D . Ανάλογα τις ταξινομήσεις του D προκύπτουν οι απώλειες του G , οι οποίες και διοχετεύονται σε αυτόν προκειμένου να ενημερώσει τα βάρη του. Θεωρείται ότι το μοντέλο συγκλίνει ή μαθαίνει, όταν τα δεδομένα που παράγει ο G ταξινομούνται λάθος από τον D .

Η διαδικασία της εκπαίδευσης πραγματοποιείται εναλλάξ, δηλαδή όταν βελτιστοποιείται ο D , ο G δεν ενημερώνει τα βάρη του και το αντίστροφο. Στο Σχήμα 3.4i παρουσιάζονται τα αποτελέσματα του πρώτου GAN όπως παρουσιάστηκε από τον Ian Goodfellow το 2014, ενώ στο Σχήμα 3.4ii ένα σύγχρονο ΠΑΔ StyleGan2-ADA [78] (2020), το οποίο μάλιστα και εκπαιδεύεται πάνω σε σύνολα δεδομένων με περιορισμένο αριθμό δειγμάτων.



(i) Πρώτο GAN (Πηγή [64] Σχήμα 2).

(ii) StyleGan2-ADA (Πηγή [78] Σχήμα 10).

Σχήμα 3.4: Παραδείγματα ΠΑΔ.

3.4 Μετασηματιστές (Transformers)

Ίσως από τις σημαντικότερες αρχιτεκτονικές των τελευταίων ετών αποτελούν οι Μετασηματιστές (Transformers) [68] (2017). Αποτελούν τον αντικαταστάτη των ANN (RNN), συγκεκριμένα των ΜΒΔΔ (LSTM), αφού βασικό προτέρημά τους είναι η παραλληλοποίηση της εκπαίδευσης των επιπέδων τους, δεν απαιτούν τη σειριακή επεξεργασία της εισόδου με άμεσο αποτέλεσμα τη ταχύτερη εκπαίδευση. Το παραπάνω σε συνδυασμό με τα εξαιρετικά αποτελέσματα πάνω σε προβλήματα Επεξεργασίας Φυσικής Γλώσσας (Natural Language Processing; NLP), τα τοποθετούν στις κορυφαίες επιλογές.

Γνωστότερα μοντέλα, που παρέχονται προ εκπαιδευμένα (pretrained) για άμεση υλοποίηση και χρήση είναι το BERT (Bidirectional Encoder Representations from Transformers) [79] (2018) και τη σειρά GPT (Generative Pre-trained Transformer), με το τελευταίο μοντέλο GPT-3 [80] (2020). Στην αρχική πρόταση του το BERT χωρίζεται σε δύο υπο μοντέλα το BERT-base με δώδεκα κωδικοποιητές (encoders) και δώδεκα bidirectional self-attention heads, το BERT-large που είναι διπλάσιο σε μέγεθος.

Βέβαια, αρκετές νέες αρχιτεκτονικές έχουν προκύψει από την οικογένεια του BERT, όπως για παράδειγμα το ALBERT [81] (2019), DistilBERT [82] (2019) και RoBERTa [83] (2019) μερικές από αυτές.

Επιπρόσθετα, όπως παρουσιάζεται το Vision Transformer (ViT) [84] υπάρχουν προοπτικές αντικατάστασης των ΣΝΔ με δίκτυα Μετασχηματιστών, οι συγγραφείς ισχυρίζονται ότι επιτυγχάνουν αποτελέσματα αντάξια με εκείνα των σύγχρονων και κατά πολύ πιο περίπλοκων αρχιτεκτονικών ΣΝΔ, ενώ παράλληλα απαιτούν λιγότερο κόστος εκπαίδευσης.

Τέλος, η Google πρόσφατα εκπαίδευσε έναν μετασχηματιστή ενός τρισεκατομμυρίου παραμέτρων μεγέθους περίπου έξι φορές όσο το μέγεθος του GPT-3, το οποίο είχε κινήσει ήδη αρκετές δυσανασχετίσεις για το αποτύπωμα άνθρακα που δημιουργήσε η εκπαίδευση του. Η μεθοδολογίες και οι τεχνικές που χρησιμοποιήθηκαν για να αυξήσουν το μέγεθος του μοντέλου, δίχως να αυξηθεί γραμμικά ή και εκθετικά το κόστος εκπαίδευσης περιγράφονται στο [85].

3.4.1 Αρχιτεκτονική

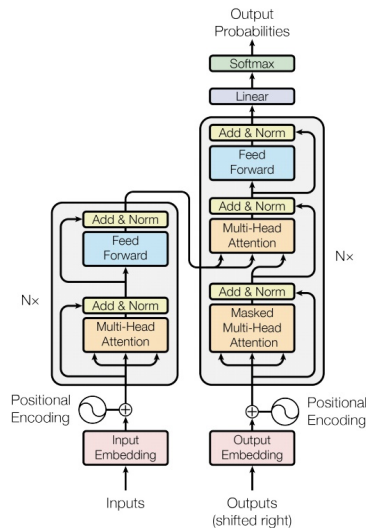
Όπως αναφέρθηκε έχουν δημιουργηθεί και θα δημιουργηθούν αρκετές νέες αρχιτεκτονικές βασισμένες στους Μετασχηματιστές. Παρακάτω θα περιγραφεί αυτή, όπως παρουσιάστηκε στη δημοσίευση [68] που τους εισήγαγε.

Το σημαντικότερο κομμάτι του συγκεκριμένου τύπου δικτύου αποτελεί το επίπεδο Προσοχής, το οποίο και παρουσιάζεται αναλυτικά στην Ενότητα 3.4.2. Ακόμα, ιδιαίτερη αξία προσδίδει ο τρόπος κωδικοποίησης της εισόδου που παρουσιάζεται από τους συγγραφείς, αφού κωδικοποιηθούν οι λέξεις με learned embeddings στη συνέχεια ενσωματώνεται η θέση τους με βάση τις Εξισώσεις 3.2.

$$PE_{(pos,2i)} = \sin(pos/10000^{2i/d_{model}})$$

$$PE_{(pos,2i+1)} = \cos(pos/10000^{2i/d_{model}})$$
(3.2)

Το προαναφερόμενο αποτελεί εξαιρετική προσθήκη διότι καθιστά το μοντέλο ικανό να επεξεργάζεται προτάσεις διαφορετικού μήκους από αυτές που εκπαιδεύτηκε. Οι εν λόγω κωδικοποιήσεις ενσωματώνονται στα embeddings των λέξεων της εισόδου και στη συνέχεια τροφοδοτούνται στον Κωδικοποιητή και στον Αποκωδικοποιητή, όπως φαίνεται στο Σχήμα 3.5.



Σχήμα 3.5: Αρχιτεκτονική Transformer (Πηγή [68] Σχήμα 1).

Προχωρώντας ψηλότερα στο Σχήμα 3.5, το αριστερό γκρι κουτί ονομάζεται Κωδικοποιητής (Encoder) και το δεξί Αποκωδικοποιητής (Decoder), τα εσωτερικά στοιχεία των οποίων παρουσιάζουν αρκετές ομοιότητες.

Αναλυτικότερα, το πρώτο υπό επίπεδο του Κωδικοποιητή ονομάζεται μηχανισμός αυτόπροσοχής με πολλαπλές κεφαλές (multi-head self-attention mechanism) και το δεύτερο ένα απλό πλήρως διασυνδεδεμένο πρόσθιας μετάδοσης νευρωνικό δίκτυο.

Ο αποκωδικοποιητής ενσωματώνει ακριβώς τα ίδια υπό επίπεδα με τον Κωδικοποιητή, με τη μόνη διαφορά να είναι ένα επιπλέον υπό επίπεδο στη βάση του, το οποίο υλοποιεί μηχανισμό αυτόπροσοχής με πολλαπλές κεφαλές, ενσωματώνοντας όμως μίας μορφής μάσκας στην έξοδο του Κωδικοποιητή.

Προστίθεται κανονικοποίηση σε κάθε έξοδο των αναφερόμενων υπό επιπέδων καθώς και συνδέσεις υπολοίπων (residual connections). Η τελική αρχιτεκτονική δημιουργεί $N = 6$ Κωδικοποιητές και $N = 6$ Αποκωδικοποιητές, βέβαια το N αποτελεί υπερπαράμετρο του δικτύου.

3.4.2 Προσοχή (Attention)

Η διαδικασία της προσοχής είναι ίσως το κυριότερο κομμάτι των Μετασχηματιστών. Το κείμενο της εισόδου σε τέτοια μοντέλα μεταφράζεται σε σύμβολα (tokens) και στη συνέχεια κωδικοποιείται όπως παρουσιάστηκε στην Ενότητα 3.4.1, τα κωδικοποιημένα διανύσματα δίνονται ως είσοδος στο δίκτυο.

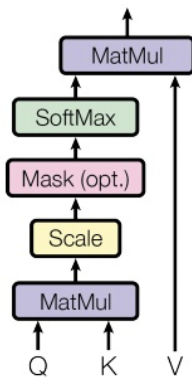
Έστω τρεις μήτρες W^K, W^Q, W^V , οι οποίες μαθαίνονται κατά τη διάρκεια της εκπαίδευσης. Τότε προκύπτουν τα διανύσματα k, q, v του κάθε συμβόλου ως το εσωτερικό γινόμενο της διανυσματικής παράστασης τους x με τις μήτρες W^K, W^Q, W^V , αντίστοιχα. Το μέγεθος των παραγόμενων διανυσμάτων, όπως προτείνεται [68] είναι 64 διαστάσεις έκαστο.

$$Attention(A, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (3.3)$$

Αναλύοντας την Εξίσωση 3.3, αρχικά υπολογίζεται το εσωτερικό γινόμενο του q με όλα τα διαφορετικά k της πρότασης εισόδου και στη συνέχεια διαιρείται με την τετραγωνική ρίζα του μήκους του k (δηλαδή $\sqrt{d_k}$). Έστω για παράδειγμα, το πρώτο σύμβολο t_1 μίας πρότασης x , στο οποίο αντιστοιχεί διάνυσμα q_1 , τότε υπολογίζονται τα εσωτερικά γινόμενα $q_1 \cdot k_1, \dots, q_1 \cdot k_n$, με n το πλήθος των διαφορετικών συμβόλων της πρότασης εισόδου, στη συνέχεια το κάθε ένα διαιρείται με $\sqrt{64} = 8$.

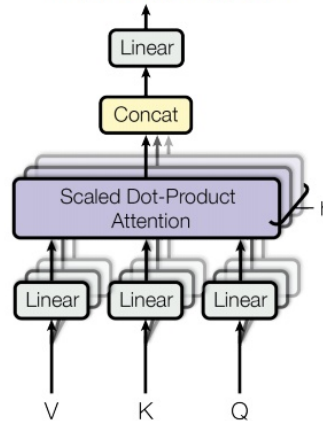
Η διαίρεση λειτουργεί σαν κανονικοποίηση, αφού επιτρέπει την καλύτερη διαχείριση των βαρών. Εφόσον, υπολογιστούν οι παραπάνω τιμές, χρησιμοποιείται η συνάρτηση softmax προκειμένου να υπολογιστεί η πιθανοφάνεια των τιμών. Η πιθανότητα κάθε συμβόλου, λειτουργώντας σαν βάρος, πολλαπλασιάζεται με το διάνυσμα v . Τελικώς, τα προαναφερόμενα διανύσματα αθροίζονται σε ένα τελικό διάνυσμα z .

Scaled Dot-Product Attention



(i) Scaled Dot-Product Attention (Πηγή [68] Σχήμα 2 (αριστερά)).

Multi-Head Attention



(ii) Multi-Head Attention (Πηγή [68] Σχήμα 2 (δεξιά)).

Σχήμα 3.6: Επίπεδο Προσοχής (Attention).

Η παραπάνω διαδικασία περιγράφει τη διαδικασία προσοχής (attention) Σχήμα 3.6i, επιστρατεύονται πολλαπλές κεφαλές προσοχής (multi-head attention) Σχήμα 3.6ii, οι οποίες πραγματοποιούν παράλληλα την ίδια διαδικασία h (προτείνεται $h = 8$ [68]) φορές. Δηλαδή εκπαιδεύονται W_h^K, W_h^Q, W_h^V μήτρες, οι οποίες σε τελική έξοδο παράγουν z_1, \dots, z_h . Τα τελευταία, συναθροίζονται και υπολογίζεται το εσωτερικό γινόμενο με μήτρα W^O , με σκοπό να ληφθεί το τελικό z . Η μήτρα W^O προστίθεται στις

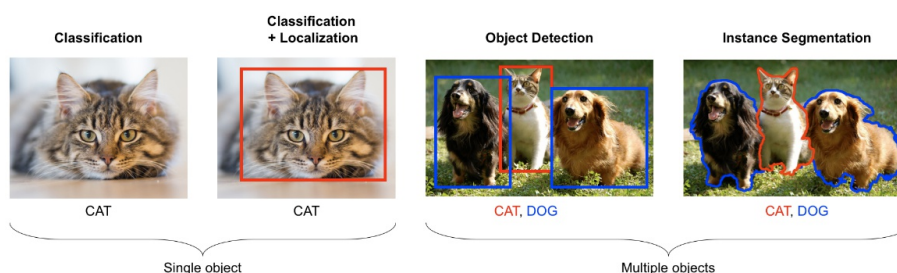
μήτρες που μαθαίνονται κατά την εκπαίδευση. Οι συγγραφείς παραθέτουν ότι κάθε μία από τις h κεφαλές εκπαιδεύεται να δίνει προσοχή σε διαφορετικά σημεία και άρα η χρήση πολλαπλών, ενισχύει την αποτελεσματικότητα του μοντέλου.

Τελευταίο δομικό συστατικό των Μετασχηματιστών, αποτελεί η μόνη διαφορά μεταξύ Κωδικοποιητών και Αποκωδικοποιητών, δηλαδή το αρχικό επιπλέον επίπεδο που ενσωματώνουν οι Αποκωδικοποιητές. Το επίπεδο πολλαπλών κεφαλών προσοχής με μάσκα (masked multi-head attention), υλοποιεί την παραπάνω ρουτίνα με ειδοποιό διαφορά την παρουσία μάσκας, δηλαδή τα σύμβολα που βρίσκονται σε επόμενες θέσεις από τη θέση της λέξης που επεξεργάζεται ο Κωδικοποιητής, φιλτράρονται έξω από το μοντέλο. Η τεχνική αυτή είναι αναγκαία, αφού ενεργοποιεί την εκπαίδευση στο μοντέλο, αν απουσίαζε το μοντέλο δε θα μάθαινε να προβλέπει διότι θα είχε στη διάθεση του τις επόμενες λέξεις και θα χρησιμοποιούσε αυτές ως έξοδο.

Με βάση το [68] δημιουργούνται έξι παράλληλοι Κωδικοποιητές και έξι παράλληλοι Αποκωδικοποιητές, ωστόσο δεν αποτελεί αναγκαία συνθήκη όπως και το πλήθος των κεφαλών.

3.5 Περιφερειακό Συνελκτικό Νευρωνικό Δίκτυο (R-CNN)

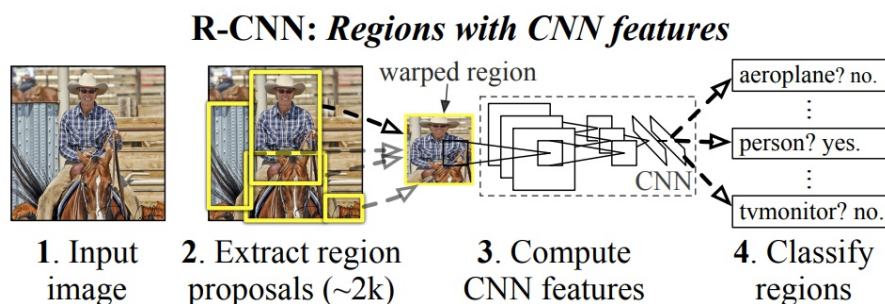
Μέχρι τώρα έχει γίνει αναφορά ταξινόμησης (classification), η οποία αποτελεί υποσύνολο της ταξινόμησης αντικειμένων (object detection). Κατά την τελευταία, απαιτείται τόσο η ταξινόμηση αλλά και ο εντοπισμός αυτών, εξάγοντας ορθογώνια που περικλείουν τα αντικείμενα στα δεδομένα εισόδου. Για παράδειγμα, έχοντας ως είσοδο μία εικόνα με διάφορα φρούτα σε μία πιατέλα, το μοντέλο οφείλει να παράξει σαν έξοδο τις συντεταγμένες των ορθογώνιων που εμπεριέχουν τα διαφορετικά είδη φρούτων αλλά και να τα ταξινομήσει σε κατηγορίες.



Σχήμα 3.7: Προβλήματα στην Όραση Υπολογιστών [86].

Τα παραπάνω αποτυπώνονται καλύτερα στο Σχήμα 3.7, ακόμα παρουσιάζονται οι έννοιες του Classification + Localization, δηλαδή ο εντοπισμός ενός αντικειμένου μέσα σε μία εικόνα καθώς και του Instance Segmentation, ή αναλυτικότερα η εξαγωγή μίας μάσκας (συνήθως δυαδικής απεικόνισης) για το που βρίσκεται το αντικείμενο μέσα στην εικόνα. Το Instance Segmentation διαφέρει από το Classification + Localization και το Object Detection, αφού δεν παράγει ορθογώνια που μπορεί να εμπεριέχουν το αντικείμενο αλλά προσπαθεί να εντοπίσει τη γεωμετρία που το ερμηνεύει.

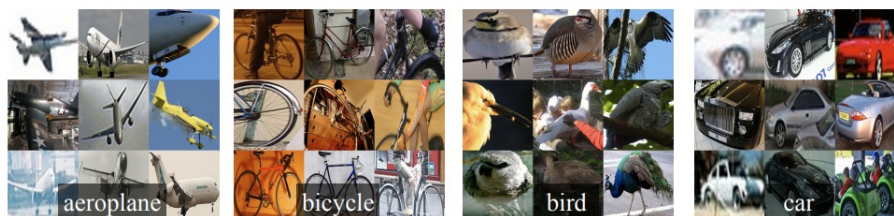
Η αρχιτεκτονική R-CNN [62] εισήχθη το 2013 από τους R. Girshick, J. Donahue, T. Darrell, J. Malik, ως μία λύση για ανίχνευση αντικειμένων σε προβλήματα Όρασης Υπολογιστών. Η παραπάνω πρόταση παρουσιάζεται στο Σχήμα 3.8.



Σχήμα 3.8: Αρχιτεκτονική R-CNN (Πηγή [62] Σχήμα 1).

Το προτεινόμενο μοντέλο στο [62] κάνει χρήση επιλεκτικής αναζήτησης (Selective Search) [87], για να εξάγει περίπου δύο χιλιάδες περιοχές ενδιαφέροντος (Regions of Interest; RoIs). Βέβαια όπως θα αναφερθεί και στην Ενότητα 3.5.2, η επιλογή αυτού του αλγορίθμου αναζήτησης έχει δραματική επιβράδυνση στην εκπαίδευση του μοντέλου.

Εφόσον παρέλθει η Επιλεκτική Αναζήτηση, τα εξαγόμενα RoIs τροφοδοτούνται σε ΣΝΔ με απώτερο σκοπό την εξαγωγή χαρακτηριστικών. Εδώ δημιουργείται το εξής πρόβλημα, οι επιμέρους περιοχές (ορθογώνια) που έχουν ληφθεί από την αρχική εικόνα δεν έχουν όλες τις ίδιες διαστάσεις. Προκειμένου να λύσουν το πρόβλημα αυτό, οι συγγραφείς ελαττώνουν όλες αυτές τις περιοχές σε τετράγωνο συγκεκριμένων διαστάσεων (227×227). Ένα παράδειγμα δίνεται στο Σχήμα 3.9.



Σχήμα 3.9: Τετραγωνισμός Ορθογωνίων RoIs (Πηγή [62] Σχήμα 2).

Το επίπεδο των ΣΝΔ παράγει ένα διάνυσμα 4096 για κάθε περιοχή, δηλαδή περίπου 2000×4096 , τα διανύσματα αυτά εισάγονται σε Μηχανή Διανυσμάτων Υποστήριξης (SVM) προκειμένου να ληφθεί η τελική κατηγορία της κάθε περιοχής.

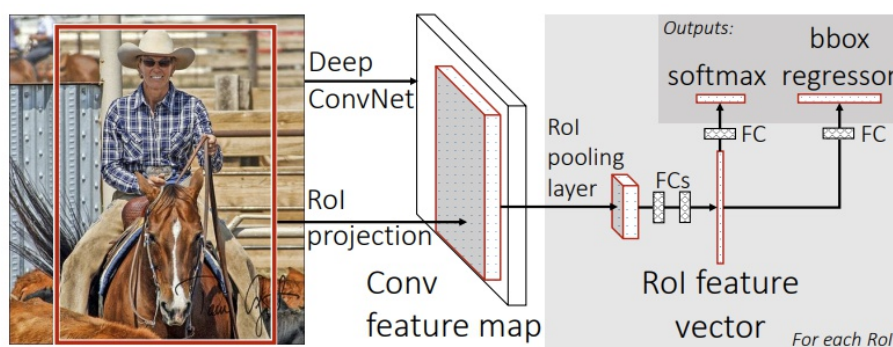
Δύο βασικά προβλήματα στην παραπάνω μεθοδολογία είναι πρώτον η αδυναμία εκπαίδευσης όλης της αρχιτεκτονικής σαν ένα ενιαίο μοντέλο και δεύτερον ο χρόνος που απαιτείται για την πραγματοποίηση πρόβλεψης. Το τελευταίο λύνεται σε επόμενες προτάσεις (Fast R-CNN, Faster R-CNN).

Η παρούσα δημοσίευση πρότείνει επίσης το σενάριο της εφαρμογής παλινδρόμησης (bounding-box regression) για την πρόβλεψη του ορθογωνίου που περικλείει την κλάση, η παλινδρόμηση μαθαίνει από τα χαρακτηριστικά που εξάγονται από το ΣΝΔ και προτείνει βελτιώσεις στις ήδη υπολογισμένες περιοχές ενδιαφέροντος.

Η αρχιτεκτονική του R-CNN παρουσιάζεται στο Σχήμα 3.8. Οι μετρικές αξιολόγησης τέτοιων μοντέλων παρουσιάζονται κατά την Ενότητα 3.8.

3.5.1 Fast R-CNN

Σε επόμενη δημοσίευση του ο Ross Girshick (2015) προτείνει τη χρήση ΣΝΔ στην αρχική εικόνα και στην συνέχεια προβολή των ROIs, όπως αυτά υπολογίζονταν στο [62]. Η νέα αρχιτεκτονική ονομάζεται Fast R-CNN [88], βασίζεται στο τότε υπερσύγχρονο VGG-16, το οποίο και εκπαιδεύει $\times 9$ φορές πιο γρήγορα, ενώ επιτυγχάνει $\times 213$ γρηγορότερες προβλέψεις και καλύτερα αποτελέσματα από το R-CNN. Η αρχιτεκτονική του Fast R-CNN παρουσιάζεται στο Σχήμα 3.10.



Σχήμα 3.10: Αρχιτεκτονική Fast R-CNN (Πηγή [88] Σχήμα 1).

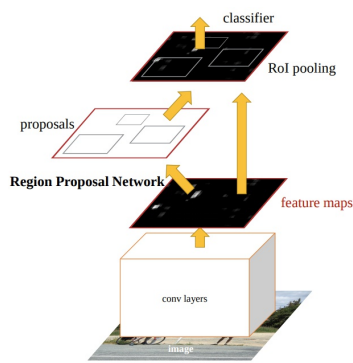
Αξιοσημείωτη αποτελεί η προσθήκη του επιπέδου υποδειγματοληψίας των περιοχών ενδιαφέροντος (RoI pooling layer), η συγκεκριμένη τεχνική χρησιμοποιείται τόσο στο Fast R-CNN, αλλά και στο διά-

δοχο του Faster R-CNN. Έχοντας τις προβλέψεις για τις πιθανές προτάσεις, αυτές διαιρούνται με τον συντελεστή κλίμακας (scale factor) και κβαντίζονται (συνήθως με στρογγυλοποίηση προς τα κάτω) σε φυσικούς αριθμούς. Ύστερα με ένα επίπεδο υποδειγματοληψίας με μέγιστο ή μέσο, δημιουργείται ένας τελικός τανυστής που τροφοδοτείται στα επίπεδα ταξινόμησης.

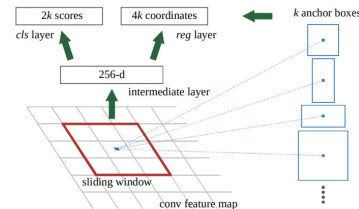
Όπως γίνεται αντιληπτό, η παραπάνω μεθοδολογία δε λαμβάνει υπ' όψιν αρκετή πληροφορία αφού κατά τον κβαντισμό αμελεί σημεία. Το RoIAlign (Ενότητα 3.5.3) λύνει το πρόβλημα αυτό κάνοντας χρήση διγραμμικής παρεμβολής (bilinear interpolation).

3.5.2 Faster R-CNN

Λίγο καιρό μετά το Fast R-CNN, προτείνεται το Faster R-CNN [89] (2015) από τους S. Ren, K. He, R. Girshick, J. Sun. Το νέο δίκτυο αντικαθιστά τον αλγόριθμο επιλεκτικής αναζήτησης με ένα Δίκτυο Πρότασης Περιφερειών (Region Proposal Network; RPN), εμπνευσμένο από το (Fully Convolutional Network; FCN) [90], καθώς και συνδυάζει τις δύο αρχιτεκτονικές RPN και Fast R-CNN για να λύσει το πρόβλημα της αναγνώρισης αντικειμένων (Σχήμα 3.11i).



(i) Faster R-CNN (Πηγή [89] Σχήμα 2).



(ii) RPN (Πηγή [89] Σχήμα 3 (αριστερά))

Σχήμα 3.11: Αρχιτεκτονική Faster R-CNN.

Το δίκτυο RPN (Σχήμα 3.11ii) είναι πλέον υπεύθυνο να παράξει τις πιθανές περιοχές ενδιαφέροντος (proposals), ένα μικρό παράθυρο μεγέθους $n \times n$ (οι συγγραφείς πρότειναν $n = 3$) εφαρμόζεται κατά μήκος των χαρακτηριστικών (feature map) που παράγεται από το τελευταίο επίπεδο των ΣΝΔ (ZF/VGG). Τα αποτελέσματα δίνονται ως είσοδοι σε δύο πλήρως διασυνδεδεμένα επίπεδα (fully connected layers), το ένα επιτυγχάνει παλινδρόμηση (box-regression; reg) και το άλλο ταξινόμηση (box-classification; cls) [89]. Ο reg προβλέπει τις συντεταγμένες των προτάσεων, ενώ ο cls την πιθανότητα της πρότασης να εμπεριέχει το αντικείμενο.

Αρκετό ενδιαφέρον, παρουσιάζεται στην προσθήκη του επιπέδου με κουτιών με στηρίγματα (anchor ή prior boxes). Συγκεκριμένα, οι προτάσεις που εξάγονται από το RPN προβάλλονται πάνω σε προκαθορισμένου σχήματος κουτιά με σκοπό να βρεθούν εκείνα που τα πλαισιώνουν καλύτερα. Για κάθε τέτοιο κουτί υπολογίζεται το κλάσμα της τομής προς την ένωση (Intersection Over Union; IoU), βλ. Εξίσωση 3.4, ή αλλιώς γνωστό ως δείκτης Jaccard. Εφόσον ο δείκτης αυτός βρίσκεται πάνω από κάποιο κατώφλι (συνήθως $\theta \geq 0.5$), το αντικείμενο που δίνει το μεγαλύτερο IoU πρέπει να προβλεφθεί.

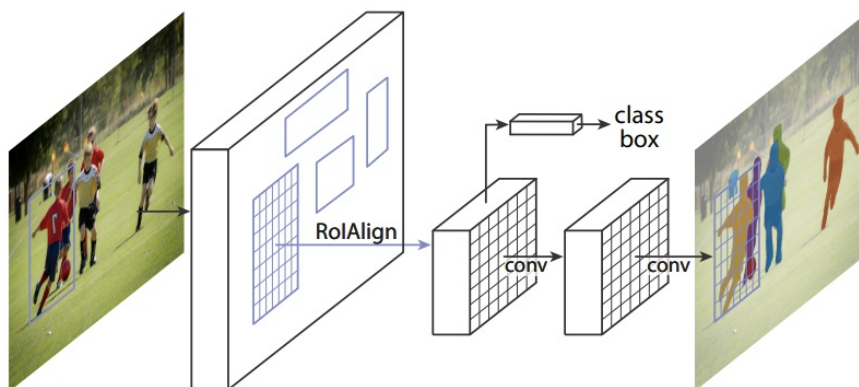
$$IoU = \frac{|A \cap B|}{|A \cup B|} \quad (3.4)$$

Πλεονέκτημα της προαναφερόμενης μεθόδου είναι αμεταβλητότητα κατά τη μετατόπιση που προσφέρει, τόσο σε επίπεδο στηρίγματος αλλά και κατά τις συναρτήσεις που υπολογίζουν τις προτάσεις με βάση τα στηρίγματα [89].

Η νέα αρχιτεκτονική είναι ταχύτερη, πιο ακριβής και πλέον μπορεί να χρησιμοποιηθεί σε συστήματα πραγματικού χρόνου, π.χ σε κάμερες, ενώ παρέχει το θετικό ότι πλέον το κομμάτι του RPN μπορεί να μάθει τις γεωμετρίες των κατηγοριών που επρόκειτο να ταξινομήσει και να μην τις εξάγει στατικά με βάση κάποιον αλγόριθμο.

3.5.3 Mask R-CNN

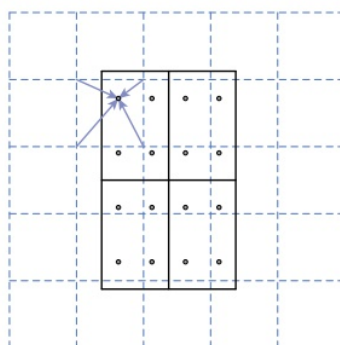
Το 2017 προστίθεται στην οικογένεια των R-CNN, το Mask R-CNN [91], η αρχιτεκτονική του οποίου απεικονίζεται στο Σχήμα 3.12. Το Mask R-CNN ενσωματώνεται ένα επιπλέον δίκτυο FPN [92] στο τελικό επίπεδο του Faster R-CNN με απώτερο σκοπό να προσδώσει μία εκτίμηση της γεωμετρίας του αντικειμένου προς ταξινόμηση. Η γεωμετρία δίνεται με μορφή δυαδικής μάσκας με διαστάσεις ίδιες εκείνης της εικόνας εισόδου.



Σχήμα 3.12: Αρχιτεκτονική Mask R-CNN (Πηγή [91] Σχήμα 1).

Το Mask R-CNN ενσωματώνει την τεχνική του **RoIAlign** (Σχήμα 3.13), προκειμένου να δώσει λύση στο σκληρό κβαντισμό του RoI Pooling. Αποτελεί πλέον τη κυρίαρχη μεθοδολογία σε αρχιτεκτονικές Όρασης Υπολογιστών, αφού διατηρεί και άρα λαμβάνει αποφάσεις με περισσότερη πληροφορία από τη RoI Pooling.

Το τέχνασμα που χρησιμοποιήθηκε αποτελεί την ανοχή δεκαδικών ψηφίων κατά τη διαίρεση με το συντελεστή κλίμακας (δηλαδή αφαιρεί το στοιχείο του κβαντισμού). Στη συνέχεια, το “ορθογώνιο” που προκύπτει διαιρείται νοητά σε n ίσα κομμάτια, το n προκύπτει ως το εμβαδόν του επιπέδου υποδειγματοληψίας (pooling layer). Εφόσον παρέλθει η παραπάνω διαίρεση, με βάση τη διγραμμική παρεμβολή πραγματοποιείται δειγματοληψία των χαρακτηριστικών (feature map) που έχουν επικάλυψη με το κάθε κομμάτι.



Σχήμα 3.13: RoIAlign (Πηγή [91] Σχήμα 3).

Για παράδειγμα με βάση το Σχήμα 3.13, ας γίνει η υπόθεση ότι το επίπεδο υποδειγματοληψίας έχει διαστάσεις 2×2 και ότι ο χάρτης χαρακτηριστικών (feature map) είναι το μπλε διακεκομμένο πλέγμα στο παρασκήνιο. Εφόσον παρέλθει η διαδικασία εξαγωγής πιθανών προτάσεων, επιλέγεται μία πρόταση και διαιρείται με το συντελεστή κλίμακας (scale factor) το αποτέλεσμα παρουσιάζεται ως μαύρο ορθογώνιο στο εν λόγω σχήμα. Το ορθογώνιο αυτό, χωρίζεται σε 4 ίσα κομμάτια (RoI bins). Το κάθε κομμάτι, επικαλύπτει συγκεκριμένες τιμές του χάρτη χαρακτηριστικών, χρησιμοποιείται η τεχνική της διγραμμικής παρεμβολής προκειμένου να γίνει δειγματοληψία σε τέσσερα σημεία από για κάθε ένα RoI bin. Οι τελικές τιμές συναθροίζονται είτε με βάση τη μέγιστη τιμή ή το μέσο και στη συνέχεια τροφοδοτούνται στο επίπεδο ταξινόμησης [91].

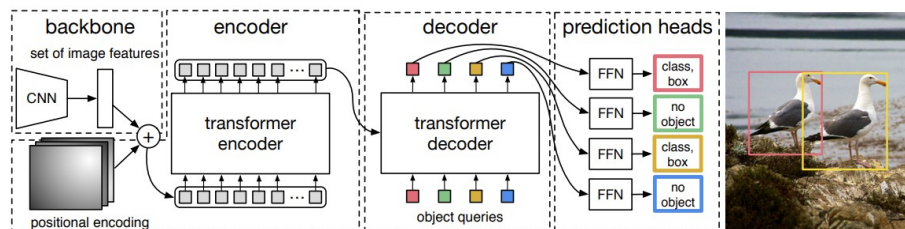
Μία ακόμα μέθοδος που έχει προταθεί αποτελεί η **RoI Warping** [93] που βασίζεται και αυτή στη διγραμμική παρεμβολή, μοιάζει αρκετά στη RoIAlign, αλλά χρησιμοποιεί κβαντισμό κατά την προβολή στο χάρτη χαρακτηριστικών.

Οι συγγραφείς του [91] συγκρίνουν τις μεθοδολογίες RoIAlign και RoIWarp, τονίζοντας τη σημασία της ευθυγράμμισης που δεν ενσωματώνουν οι RoIPool, RoIAlign.

Τελευταία επέκταση της οικογένειας των R-CNN αποτελεί το Mesh R-CNN [94] (2019), το οποίο προσπαθεί να δημιουργήσει ένα τρισδιάστατο πλέγμα των αντικειμένων που αναγνωρίζει. Βασίζεται εξολοκλήρου στο Mask R-CNN, το επεκτείνει ωστόσο προκειμένου να εξάγει τη προαναφερόμενη πληροφορία.

3.6 DETection TRansformer (DETR)

Η αρχιτεκτονική DETection TRansformer [95] που προτάθηκε το 2020, συνδυάζει μοντέλα Μετασχηματιστών (Ενότητα 3.4.1) μαζί με δίκτυο FPN [92] για την εξαγωγή προβλέψεων αντικειμένων (object detection) με τη μορφή κουτιών (bounding boxes, Σχήμα 3.14). Οι συγγραφείς παραθέτουν, ότι η συγκεκριμένη αρχιτεκτονική μπορεί αρκετά εύκολα, με λίγες τροποποιήσεις, να χρησιμοποιηθεί για εξαγωγή μασκών (segmentation) των αντικειμένων προς αναγνώριση (3.15).



Σχήμα 3.14: Αρχιτεκτονική Detection Transformers (Πηγή [95] Σχήμα 2).

Το τελικό δίκτυο, αποτελείται από τρία απλά δομικά συστατικά. Ένα αρχικό δίκτυο ΣΝΔ (CNN backbone) για την εξαγωγή χαρακτηριστικών (π.χ ResNet50), ένα δίκτυο Μετασχηματιστών και ένα δίκτυο εμπρόσθιας τροφοδότησης (feed forward network; FFN).

Συγκεκριμένα, προτείνονται, για το ΣΝΔ μέγεθος $2048 \times \frac{H_o}{32} \times \frac{W_o}{32}$, όπου W_o, H_o διαστάσεις τις εικόνας εισόδου. Κατά την είσοδο στον κωδικοποιητή, εφαρμόζεται ένα στάδιο συνελίξεων 1×1 , μετασχηματίζοντας το χάρτη χαρακτηριστικών από $2048 \times \frac{H_o}{32} \times \frac{W_o}{32}$ σε μικρότερο $d \times \frac{H_o}{32} \times \frac{W_o}{32}$, και στη συνέχεια σε $d \times \frac{H_o}{32} \cdot \frac{W_o}{32}$.

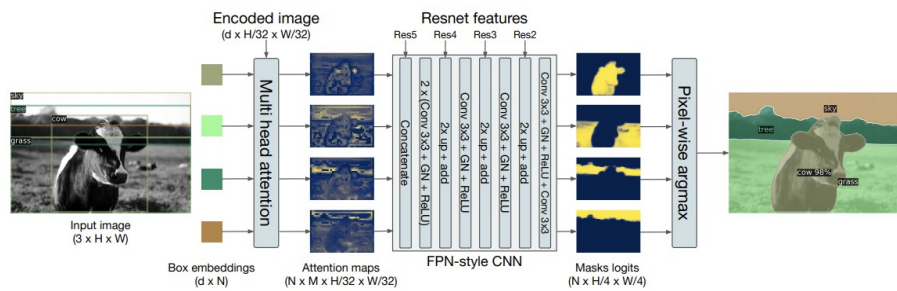
Κάθε κωδικοποιητής αποτελείται από ένα στάδιο πολλαπλών κεφαλών προσοχής (multihead attention Σχήμα 3.6ii) και ένα δίκτυο εμπρόσθιας τροφοδότησης, τονίζεται η σημαντικότητα του FFN από τους συγγραφείς κατά τη διεξαγωγή των πειραμάτων. Χρησιμοποιούνται έξι κωδικοποιητές και έξι αποκωδικοποιητές με οχτώ κεφαλές προσοχής. Η συνάρτηση λάθους (loss function) βασίζεται στον Ουγγρικό Αλγόριθμο (Hungarian algorithm) [96], με σκοπό να βρεθεί μία διμερής αντιστοίχιση [95].

Όπως αποτυπώνεται στο Σχήμα 3.14, η εικόνα διοχετεύεται πρώτα στο ΣΝΔ, τα χαρακτηριστικά που προκύπτουν μετασχηματίζονται από πίνακα σε διάνυσμα. Οι αρχικές θέσεις των εικονοστοιχείων διατηρούνται και συναθροίζονται ως positional encoding στο διάνυσμα που προκύπτει. Το διάνυσμα με τη σειρά του εισάγεται στον κωδικοποιητή του δικτύου, η έξοδος του κωδικοποιητή μαζί με τα object queries, τα οποία μαθαίνονται κατά την εκπαίδευση, συνεχίζουν στον αποκωδικοποιητή. Τέλος, η έξοδος του αποκωδικοποιητή δίνεται ως είσοδος στο δίκτυο εμπρόσθιας τροφοδότησης, με σκοπό να εξαχθούν οι κατηγορίες και τα “κουτιά” [95].

Όσον αφορά το κομμάτι που πραγματεύεται την εξαγωγή της μάσκας (segmentation, Σχήμα 3.15), όπως και στο Mask-RCNN [91], προστίθεται ένα επιπλέον επίπεδο ταξινόμησης της εξόδου του αποκωδικοποιητή. Η ταξινόμηση εφαρμόζεται πάνω στα “κουτιά” (predicted bounding box) που εξάγει το μοντέλο, επομένως για κάθε πρόβλεψη προσφέρεται το κουτί που την πλαισιώνει, η ονομασία της κατηγορίας στην οποία ανήκει το αντικείμενο καθώς και μία δυαδική μάσκα που αντιστοιχεί σε αυτό.

Τέλος, παρατίθεται ότι η εκπαίδευση του τελικού αυτού επιπέδου μπορεί να γίνει είτε σε μία ολοκληρωμένη αρχιτεκτονική ή σε δύο στάδια. Κατά την τελευταία επιλογή, εκπαιδεύεται το DETR πάνω

σε πρόβλημα object detection, εφόσον φτάσει κάποια ικανοποιητικά επίπεδα “κλειδώνουν” τα βάρη και πραγματοποιείται εκπαίδευση του επιπέδου ταξινόμησης.

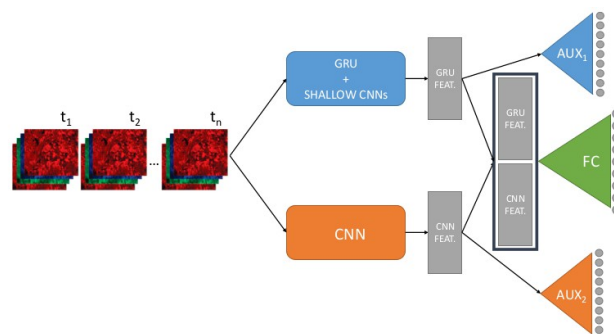


Σχήμα 3.15: Προσθήκη Κεφαλής για Object Segmentation (Πηγή [95] Σχήμα 8).

3.7 DuPLO

Η αρχιτεκτονική “DuPLO: A DUal view Point deep Learning architecture for time series classificatiOn” [97] προτάθηκε το 2018 και αποτελεί από τα πρώτα μοντέλα που λαμβάνουν υπ’ όψιν τη χρονική υπογραφή της εκάστοτε καλλιέργειας, προτού προβούν σε τελική ταξινόμηση.

Η εσωτερική αρχιτεκτονική του μοντέλου απεικονίζεται στο Σχήμα 3.16 και όπως φαίνεται προκύπτει από την συνένωση δύο μικρότερων αρχιτεκτονικών, κάθε μία γνωστή για την κυριαρχία της στα προβλήματα που αντιμετωπίζει.



Σχήμα 3.16: Αρχιτεκτονική DuPLO (Πηγή [97] Σχήμα 1).

Συγκεκριμένα, το πρώτο τμήμα (GRU + Shallow CNN branch), για τη χρονική συσχέτιση χρησιμοποιεί μοντέλα τύπου GRU (RNN, Ενότητα 3.1) πάνω σε χάρτη χαρακτηριστικών που εξάγεται από “ρηχά” ΣΝΔ. Το δεύτερο σκέλος (CNN branch), εξάγει χαρακτηριστικά με αρκετά βαθύτερα ΣΝΔ. Τα επιμέρους χαρακτηριστικά ενώνονται και εισάγονται σε τελικό δίκτυο εμπρόσθιας τροφοδότησης για την τελική ταξινόμηση.

Κατά την εκπαίδευση, κάθε σκέλος παράγει τη δικιά του συνάρτηση σφάλματος, ενώ το σταθμισμένο άθροισμα αποτελεί τη συνάρτηση σφάλματος όλης της αρχιτεκτονικής.

Τα δύο στάδια, δέχονται διαφορετικές εισόδους. Αναλυτικότερα, το CNN branch λαμβάνει είσοδο μεγέθους $C \times 5 \times 5$, όπου $C = 5 \cdot T$, δηλαδή 5 κανάλια (B02, B03, B04, B08 και το Normalized Difference Vegetation Index; NDVI) για T χρονικές στιγμές. Από το CNN branch προκύπτει ένα διάνυσμα 1024 διαστάσεων.

Όπως προαναφέρθηκε, το στάδιο GRU + Shallow CNN, συνδυάζει ένα “ρηχό” ΣΝΔ που εφαρμόζεται στα δεδομένα εισόδου, μεγέθους $5 \times 5 \times 5$ για κάθε χρονική στιγμή και παράγει έξοδο διάνυσμα μεγέθους 64 διαστάσεων. Τα αποτελέσματα διοχετεύεται σε δίκτυο GRU, το οποίο παράγει τελικό διάνυσμα 1024 διαστάσεων, με σκοπό να υπάρχει ομοιόμορφη εκπροσώπηση των χαρακτηριστικών της εισόδου και από τα δύο τμήματα.

Το 5×5 και στις δύο περιπτώσεις καθορίζει το μέγεθος της εικόνας, ενώ ως κατηγορία του 5×5 θεωρείται το κεντρικό εικονοστοιχείο, δηλαδή η θέση (2, 2) [97].

Τελική έξοδος του μοντέλου, αποτελεί ο συνδυασμός των επιμέρους, δηλαδή ένα διάνυσμα 2048 διαστάσεων, το οποίο με τη σειρά του εισάγεται σε ένα δίκτυο εμπρόσθιας τροφοδοσίας, με σκοπό τη τελική ταξινόμηση του αγροτεμαχίου.

3.8 Μετρικές Αξιολόγησης Μοντέλων

Έστω οι όροι

- **TP**: Πλήθος δειγμάτων που ταξινομήθηκαν στην κλάση που άνηκαν (Κλάση Positive).
- **FP**: Πλήθος δειγμάτων που ταξινομήθηκαν στην κλάση που δεν άνηκαν (Κλάση Positive).
- **TN**: Πλήθος δειγμάτων που ταξινομήθηκαν στην κλάση που άνηκαν (Κλάση Negative).
- **FN**: Πλήθος δειγμάτων που ταξινομήθηκαν στην κλάση που άνηκαν (Κλάση Negative).

Θεωρώντας τα παραπάνω, οι κύριες μετρικές αξιολόγησης των μοντέλων είναι

- **Accuracy (Εξίσωση 3.5)**: Ποσοστό των ορθών προβλέψεων που έκανε το μοντέλο.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.5)$$

- **Precision (Εξίσωση 3.6)**: Ποσοστό ακρίβειας των προβλέψεων του μοντέλου, δηλαδή πόσες από τις προβλέψεις που έκανε άνηκαν στην ίδια κλάση.

$$Precision = \frac{TP}{TP + FP} \quad (3.6)$$

- **Recall (Εξίσωση 3.7)**: Ποσοστό πληρότητας των προβλέψεων του μοντέλου, δηλαδή από τις προβλέψεις που ανήκουν στην κλάση, πόσες όντως έγιναν σωστά.

$$Recall = \frac{TP}{TP + FN} \quad (3.7)$$

- **F1 Score (Εξίσωση 3.8)**: Αρμονικός μέσος μεταξύ Precision και Recall.

$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (3.8)$$

Βέβαια, αυτές οι μετρικές μπορούν να γενικευτούν και να χρησιμοποιηθούν σε προβλήματα με περισσότερες από μία κλάσεις. Λειτουργούν ως τους κύριους τρόπους αξιολόγησης μοντέλων ταξινόμησης, όμως όπως περιγράφηκε στην Ενότητα 3.5. Τα προβλήματα Όρασης Υπολογιστών, αποτελούν άλλη κατηγορία προβλημάτων και άρα πρέπει να χρησιμοποιηθούν άλλες μετρικές. Γνωστότερη αποτελεί η mAP (mean Average Precision), η οποία συνδέεται με τα παραπάνω αλλά κάνει κάποιες παραδοχές προκειμένου να συμπεριλάβει το γεγονός ότι οι επιμέρους κλάσεις μπορούν να έχουν κάποια (ποσοστιαία) επικάλυψη.

Υπάρχουν αρκετές εκδοχές υπολογισμού του mAP, ωστόσο θα παρουσιαστεί μόνο αυτή του COCO mAP ή COCO AP (η εκδοχή αυτή θεωρεί ισάξιες τις δύο ορολογίες [98]). Συγκεκριμένα, υπολογίζεται για κάθε δείγμα το ποσοστό επικάλυψης με βάση το δείκτη Jaccard (Εξίσωση 3.4). Θεωρούνται διαφορετικά επίπεδα με κατώφλια επικάλυψης (IoU thresholds) από το 0.5 ως και το 0.95 με βήμα 0.05 (.50 : .05 : .95), έτσι για παράδειγμα το επίπεδο με κατώφλι 0.5 συμβολίζεται AP_{50} . Για κάθε ένα από τα επίπεδα υπολογίζονται πόσα δείγματα ικανοποιούν το κατώφλι με βάση το δείκτη Jaccard, με αυτόν τον τρόπο προκύπτουν τα προαναφερόμενα TP, FP, TN, FN.

Υπολογίζονται τα Precision και Recall για κάθε IoU, δημιουργείται η γραφική παράσταση (PR Curve) και υπολογίζεται το εμβαδόν. Για να γίνει το παραπάνω, διαλέγονται έντεκα σημεία με ίση απόσταση στο εύρος 0 – 1 (δηλαδή 0, 0.1, ..., 1), για τα οποία εφαρμόζεται η Εξίσωση 3.9.

$$AP = \frac{1}{11} \sum_{i=0}^{10} AP_{i \cdot 10^{-1}} \quad (3.9)$$

Ο μέσος όρος των AP που προκύπτουν για κάθε κατηγορία θεωρείται ως το τελικό AP ή mAP του συγκεκριμένου IoU, με βάση το COCO AP, τελικό σκορ αποτελεί ο μέσος όρος AP όλων των IoU.

Τέλος, για τα προβλήματα NLP έχουν δημοσιευθεί αρκετά σύνολα δεδομένων που χρησιμοποιούνται για αξιολόγηση των τελικών μοντέλων, ωστόσο οι πιο δημοφιλείς μετρικές που παρατίθενται σχεδόν σε κάθε δημοσίευση είναι η BLEU (Bilingual Evaluation Understudy) [99] και η ROUGE (Recall Oriented Understudy for Gisting Evaluation) [100].

Κεφάλαιο 4

Σύνολο Δεδομένων

Η παρούσα εργασία θα αξιοποιήσει δεδομένα που προέρχονται από την υπηρεσία Copernicus [101] και συγκεκριμένα της διαστημικής αποστολής Sentinel-2 [102]. Όπως αναφέρθηκε στην Υποενότητα 2.4.2, το πρόγραμμα Sentinel-2 παρέχει πρόσβαση σε πολυφασματικά δεδομένα υπό τη μορφή δύο ειδών προϊόντων (Level-1C, Level-2A). Θα πραγματοποιηθεί συλλογή των δεδομένων τύπου Level-1C αφού αυτά εμπεριέχουν πληροφορία σχετικά με τη μορφολογία της Γήινης επιφάνειας.

Ως περιοχή ενδιαφέροντος (AOI) ορίζεται η Καταλονία, μία σχετικά μικρή περιφέρεια της Ισπανίας. Η Καταλονία έχει πρωτεύουσα τη Βαρκελώνη και τρεις επίσημες γλώσσες τα Καταλονικά, τα Αρανικά και τα Ισπανικά. Στο Σχήμα 4.1 παρουσιάζεται το πολύγωνο που περικλείει τη Καταλονία και με βάση του οποίου θα ζητηθούν τα δεδομένα από τη διαστημική αποστολή Sentinel-2.



Σχήμα 4.1: Καταλονία, περιοχή ενδιαφέροντος.

Όλα τα επίπεδα/στάδια της εργασίας θα γραφτούν σε Python 3.8.5 [103], ενώ για την οπτικοποίηση των δεδομένων θα χρησιμοποιηθεί η βιβλιοθήκη matplotlib [104] και το λογισμικό ελεύθερου κώδικα QGIS [105]. Τα επίπεδα (layers) για το QGIS αποκτήθηκαν από Natural Earth [106] και συγκεκριμένα χρησιμοποιήθηκαν τα populated places [107], boundary lines [108] και Natural Earth II with Shaded Relief, Water, and Drainages [109]. Τέλος, η ιστοσελίδα του Copernicus προσφέρει αρχείο σε μορφή kml [110] που εμπεριέχει τις συνταγμένες κάθε tile, ωστόσο επειδή το αρχείο αυτό δεν ήταν εφικτό να φορτωθεί χρησιμοποιήθηκαν τα αρχεία που έχει μετατρέψει ο χρήστης “justinelliotmeyers” στο εν λόγω repository [111].

Η βάση δεδομένων από την οποία θα εξαχθούν οι κατηγορίες με βάση των γεωμετριών τους, προσφέρεται από το Departament d’Agricultura, Ramaderia, Pesca i Alimentació, Generalitat de Catalunya [112]. Συγκεκριμένα, θα χρησιμοποιηθούν δεδομένα της χρονολογίας 2020 [113], όπως παρουσιάζονται στο Σχήμα 4.2 πάνω στην περιοχή ενδιαφέροντος.

Η μικρότερη διαίρεση του τελικού Συνόλου Δεδομένων ονομάζεται patch, κάθε tile αποτελείται από 900 patches. Το patch εσωτερικά για τα κανάλια με 10 μέτρα περιέχει 366×366 εικονοστοιχεία, 183×183



Σχήμα 4.2: Περιοχή ενδιαφέροντος και βάση δεδομένων (DUN 2020).

για τα 20 μέτρα και 61×61 για αυτά με 60 μέτρα ανάλυση. Στο patch, περιλαμβάνονται κατηγορίες ταξινόμησης (labels και parcels) σε ανάλυση 10 μέτρων.

4.1 Συλλογή

Έχοντας κωδικοποιήσει τα σημεία του Σχήματος 4.1 σε ζεύγη γεωγραφικού πλάτους και μήκους, δημιουργείται ένα πολύγωνο όπου το πρώτο και το τελευταίο σημείο του πολυγώνου πρέπει να έχουν την ίδια γεωγραφική υπογραφή. Το εν λόγω πολύγωνο στη συνέχεια κωδικοποιείται και αποθηκεύεται υπό τη μορφή αρχείου geojson [114].

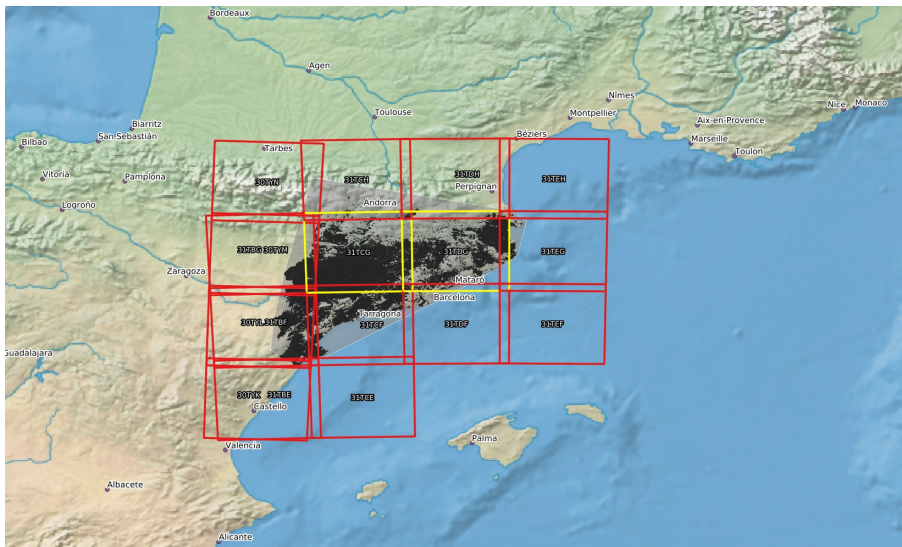
Μέσω της βιβλιοθήκης sentinelsat [115] και ενός λογαριασμού στο Copernicus Open Access Hub [116] μπορεί να ξεκινήσει η λήψη των αρχείων. Η βιβλιοθήκη λαμβάνει τα στοιχεία χρήστη του λογαριασμού, την περιοχή ενδιαφέροντος κωδικοποιημένη σε Well Known Text (WKT) από το geojson που βρίσκεται και παραμέτρους όπως το όνομα του δορυφόρου που θα χρησιμοποιηθεί, το είδος των προϊόντων που θα επιστραφούν, ενώ προσφέρονται παράμετροι για συνθήκες καιρού (π.χ ποσοστό συννεφιάς). Όπως αναφέρθηκε στην ενότητα 2.3, οι παθητικοί αισθητήρες δεν είναι τόσο αποδοτικοί κατά περιόδους πυκνής συννεφιάς. Αν τα δεδομένα έχουν δείκτη (ποσοστό) συννεφιάς μεγαλύτερο από το ζητούμενο, τότε αμελούνται. Προτείνεται, να χρησιμοποιηθεί το ποσοστό αποδοχής 0 – 10% συννεφιάς.

Μία από τις μεγαλύτερες δυσκολίες κατά τη διαδικασία λήψης των δεδομένων εμπεριέχεται στα δεδομένα που βρίσκονται στο αρχείο. Η υπηρεσία copernicus προσφέρει πρόσβαση σε απέραντο πλήθος, τεράστιων σε μέγεθος, δεδομένων γεγονός που προκαλεί προβλήματα στην αποθήκευση και διανομή τους. Προκειμένου να αντιμετωπιστούν τέτοιες δυσκολίες δημιουργήθηκε το αρχείο (Long Term Archive [117]). Κάθε προϊόν του δορυφόρου έχει διάρκεια ζωής δώδεκα μήνες, κατά αυτή την περίοδο τα δεδομένα βρίσκονται σε κατάσταση Online και μπορούν να γίνουν άμεσα διαθέσιμα στο χρήστη. Εφόσον παρέλθει το χρονικό διάστημα των δώδεκα μηνών, το προϊόν μπαίνει σε κατάσταση Offline και μετακινείται στο αρχείο. Προκειμένου ο χρήστης να λάβει πρόσβαση σε Offline δεδομένα, πρέπει να το ζητήσει και στη συνέχεια να περιμένει μέχρι το σύστημα να το φέρει σε κατάσταση Online. Μόλις γίνει Online είναι διαθέσιμο προς λήψη. Η αναμονή του χρήστη μεταξύ του αιτήματος και τη μετατροπή από Offline σε Online, διαρκεί συνήθως έως μία μέρα, ενώ προϊόντα που έχουν ανακτηθεί από το αρχείο μένουν σε κατάσταση Online για τουλάχιστον τρεις μέρες (ενδεχομένως να αλλάζει ανάλογα τη ζήτηση).

Σκοπός του γράφοντα αποτελεί η συλλογή δεδομένων τουλάχιστον ενός χρόνου, αυτό ποικίλει να αλλάξει αναλόγως τις απαιτήσεις σε υλικό (μέγεθος δίσκου, υπολογιστική ισχύ) καθώς και το χρόνο λήψης των δεδομένων από το αρχείο.

4.2 Περιγραφή

Όπως αναφέρθηκε κατά την Παγκόσμια Εγκάρσια Μερκατορική Προβολή (Ενότητα 2.2), η Γη χωρίζεται σε ζώνες. Η περιοχή ενδιαφέροντος που μελετάται εμπεριέχεται μεταξύ δύο ζωνών, της 30 και της 31. Οι εν λόγω ζώνες χωρίζονται σε επιμέρους τετράγωνα των εκατό τετραγωνικών χιλιομέτρων (Σχήμα 2.14), τα μοναδικά τετράγωνα που περικλείουν την Καταλονία είναι όπως ακολουθούν: **30TYK, 30TYL, 30TYM, 30TYN, 31TBE, 31TBF, 31TBG, 31TCE, 31TCF, 31TCG, 31TCH, 31TDF, 31TDG, 31TDH, 31TEG, 31TEH**. Στο Σχήμα 4.3 παρουσιάζονται τα εν λόγω tiles, μαζί με τα δεδομένα της βάσης και την περιοχή ενδιαφέροντος. Μπορεί εύκολα να παρατηρηθεί ότι υπάρχει μία στοιχειώδης επικάλυψη στα άκρα των τετραγώνων. Τα υπόλοιπα τετράγωνα έχουν αφαιρεθεί για λόγους διορατικότητας.



Σχήμα 4.3: Περιοχή ενδιαφέροντος, βάση δεδομένων και Tiles.

Από την έκτη Δεκεμβρίου του 2016, υιοθετήθηκε νέος τρόπος κωδικοποίησης των προϊόντων Sentinel-2 Level-1C (Naming Convention [118]), η κωδικοποίηση αυτή περιγράφεται παρακάτω.

Κάθε τέτοιο τετράγωνο που βρίσκεται μέσα στο χρονικό διάστημα που ζητήθηκε καθώς και ικανοποιηθεί διάφορες συνθήκες (λ.χ συννεφιάς), επιστρέφεται από το API του copernicus μέσω της βιβλιοθήκης sentinelsat. Τα δεδομένα βρίσκονται στη συμπιεσμένη μορφή (Ψευδοκώδικας 4.1):

Ψευδοκώδικας 4.1: Πρότυπο κωδικοποίησης ονόματος αρχείων.

```
MMM_MSIXXX_YYYYMMDDHHMMSS_Nxxyy_ROOO_Txxxxx_<Product  
Discriminator>
```

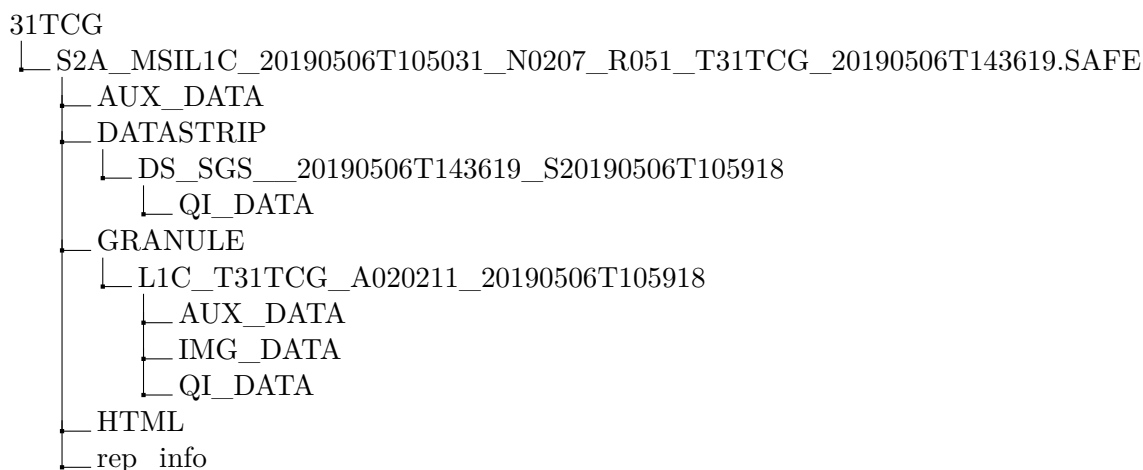
όπου **MMM** ο κωδικός αποστολής π.χ “S2B”, **MSIXXX** το είδος των προϊόντων που ζητήθηκαν π.χ “MSIL1C”, **Nxxyy** αριθμός Payload Data Ground Segment [43], **ROOO** σχετική τροχία του δορυφόρου, **Txxxxx** αριθμός Tile.

Το όνομα του αρχείου περιέχει δύο ημερομηνίες, η πρώτη “**YYYYMMDDHHMMSS**” αποτελεί τη χρονική στιγμή της δειγματοληψίας, ενώ η δεύτερη “**<Product Discriminator>**” χρησιμοποιείται για την διαφοροποίηση τελικών προϊόντων από το ίδιες δειγματοληψίες.

Στο Σχήμα 4.4 παρουσιάζονται τα περιεχόμενα όταν επιχειρείται αποσυμπίεση ενός τέτοιου αρχείου. Ο φάκελος βρίσκεται σε μορφή “.SAFE” (καθιερωμένη μορφή αρχείου της Ευρώπης [43]).

Τα δεδομένα των δεκατριών καναλιών (Πίνακας 2.1) που θα χρησιμοποιηθούν βρίσκονται μέσα στο φάκελο “IMG_DATA” σε μορφή “.jp2”.

Σχετικά με τη Βάση Δεδομένων, όπως αναφέρεται στο εγχειρίδιο Crop Map Data Source [119], αποτελείται από εννέα στήλες ονομαστικά “Campanya”, “Provincia”, “Comarca”, “Municipi”, “ID_MUN”, “Grup”, “Cultiu”, “Seca_Regad”, “HA” και έρχεται σε μορφή shapefile. Φορτώνοντας το εν λόγω shapefile



Σχήμα 4.4: Δομή αποσυμπιεσμένου αρχείου Sentinel-2 Level-1C.

με τη βοήθεια της βιβλιοθήκης `georand` [120], λαμβάνεται μία επιπλέον στήλη (`geometry`) που αναφέρει τη γεωμετρία στην οποία ανήκει κάθε γραμμή.

Συλλέγοντας δορυφορικά δεδομένα δώδεκα μηνών για την προεπιλεγμένη περιοχή, γρήγορα συγκεντρώθηκαν αρχεία μεγέθους μισού TB. Γεγονός που καθιστά αδύνατη, τουλάχιστον σε προσωπικό ηλεκτρονικό υπολογιστή, την εκπαίδευση νευρωνικών δικτύων. Συνεπώς, επιλέχθηκαν τα πιο κεντρικά `tiles`, δηλαδή τα “31TCG, 31TDG” για μελέτη, αφού περιγράφουν το μεγαλύτερο μέρος της Βάσης Δεδομένων.

4.3 Επεξεργασία

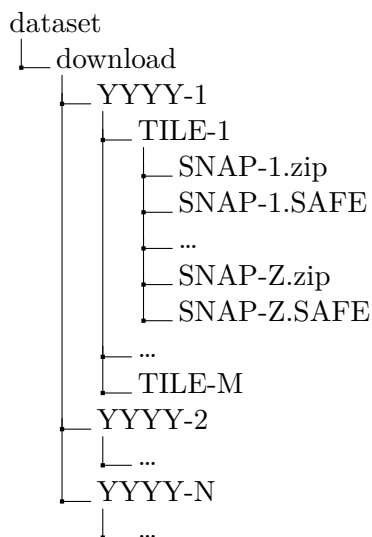
Η επεξεργασία αποτελείται από τρία βήματα. Αρχικά, η λήψη και αποσυμπίεση των αρχείων. Στη συνέχεια περιγραφεί ο τρόπος αντιμετώπισης της εξαγωγής των κατηγοριών από τις βάσεις, κάποιες βελτιστοποιήσεις που πραγματοποιήθηκαν μαζί με ανάλογες αναφορές στη διαφορά χρόνων και τέλος θα παρουσιαστεί η διαδικασία που ακολουθήθηκε προκειμένου να πραγματοποιηθούν η τμηματοποίηση των αρχικών εικόνων. Προστίθενται τεχνικές απόρριψης κενών ή θορυβωδών δεδομένων, προκειμένου να ελαττωθεί η ανάγκη για αποθηκευτικό χώρο αλλά και η υπολογιστική ισχύ που απαιτείται για την εκπαίδευση του τελικού μοντέλου. Σε τελικό στάδιο, παρουσιάζεται η μορφή του τελειωτικού, επεξεργασμένου συνόδου δεδομένων που θα χρησιμοποιηθεί ως είσοδο στην αρχιτεκτονική Βαθιάς Μάθησης.

4.3.1 Δομής Λήψης Αρχείων

Μετά το πέρας της λήψης των αρχείων, έχει κατασκευαστεί η δομή του Σχήματός 4.5. Όπου “YYYY” χρονολογία, “TILE” ο κωδικός του `tile` και “SNAP” τα δεδομένα για μία δεδομένη χρονική στιγμή. Να τονιστεί ότι υποστηρίζονται δυναμικά από το σύστημα λήψεις σε οποιοδήποτε διάστημα χρονολογιών, για οποιαδήποτε αριθμό `tile`, είτε τα δεδομένα είναι Online, είτε Offline. Ωστόσο λόγω περιορισμού διαθέσιμων υπολογιστικών και αποθηκευτικών πόρων, λήψεις θα γίνουν για τα `tiles` “31TCG, 31TDG” (κίτρινα πλαίσια στο Σχήμα 4.3) σε χρονικό διάστημα από 1 Ιανουαρίου 2020 έως 31 Δεκεμβρίου 2020. Το συνολικό μέγεθος των κατεβασμένων δορυφορικών δεδομένων υπολογίζεται προσεγγιστικά στα 90GB (40GB συμπίεσμένα και 50GB μετά την αποσυμπίεση). Η δομή των αποσυμπιεσμένων αρχείων παρουσιάζεται στο Σχήμα 4.4.

4.3.2 Μετασχηματισμός Βάσης Δεδομένων

Επιχειρώντας αποσυμπίεση στη βάση δεδομένων, αυτή βρίσκεται σε μορφή ESRI Shapefile, η οποία αποτελεί τη συνήθης μορφή αποθήκευσης γεωχωρικών δεδομένων. Τα `shapefiles` υποστηρίζονται από τα `georand` χωρίς περαιτέρω προεργασία, ωστόσο μετά από πειράματα και όπως φαίνεται στον Ψευδοκώδικα 4.2 είναι αρκετά αργά κατά το διάβασμα και το γράψιμο (I/O). Όντας παρακινούμενος από το



Σχήμα 4.5: Δομή κατεβασμένων αρχείων.

προαναφερόμενο, καθώς η εκπόνηση της εργασίας απαιτούσε αρκετό I/O, υιοθετήθηκε μία νέα μορφή αποθήκευσης γεωχωρικών δεδομένων που έρχεται με τη νέα έκδοση του `pyarrow` [121], η `feather` [122]. Η νέα έκδοση του `pyarrow` 2.0.0 προσφέρει ακόμα τη μορφή `parquet`, ωστόσο και οι δύο εκπροσωπίσεις βρίσκονται σε πρώιμο στάδιο, αφού δεν έχουν καθιερωθεί πρότυπα σχετικά με τα δεδομένα που θα επισυνάπτονται πάνω στα αρχεία. Σίγουρα, το μεταβατικό στάδιο συγκροτεί σχεδόν πάντα τη δυσκολότερη περίοδο, παράυτα οι νέες μορφές κωδικοποιούν όλα τα δεδομένα της βάσης χρησιμοποιώντας μόνο ένα τελικό αρχείο, ενώ η ESRI Shapefile χρησιμοποιεί πολλαπλά.

Ψευδοκώδικας 4.2: Χρόνοι επεξεργασίας αρχείου Shapefile.

```

# Read Shapefile
%%timeit
gdf = geopandas.read_file('path/to/old.shp')
37.4 s ± 293 ms per loop (mean ± std. dev. of 7 runs, 1 loop each)

# Write to Shapefile
%%timeit
gdf.to_file('path/to/new.shp')
3min 38s ± 2.05 s per loop (mean ± std. dev. of 7 runs, 1 loop each)

# Write to Feather
%%timeit
gdf.to_feather('path/to/new.feather')
41.6 s ± 119 ms per loop (mean ± std. dev. of 7 runs, 1 loop each)

```

Τα shapefiles διαβάζονται μέσω `geopandas` και στη συνέχεια αποθηκεύονται ως `feather`. Από εκεί και μετά, χρησιμοποιούνται μόνο τα αντίστοιχα `feather`. Προφανώς, η πρώτη ανάγνωση των shapefiles θα είναι η πιο αργή, αλλά η απόσβεση του κόστους γίνεται με κάθε επόμενη ανάγνωση αφού το κόστος πληρώνεται μία μόνο φορά. Ο απαιτούμενος χρόνος ανάγνωσης `feather`, που είναι στην ουσία ίδιο με την αρχική βάση (shapefile), αποτυπώνεται στον Ψευδοκώδικα 4.3.

Τα πειράματα διεξήχθησαν σε περιβάλλον Python 3.8.5, Jupyter Notebook 6.1.5 και `geopandas` (0.8.3) [123] με χρήση της εντολής `%%timeit`, η οποία επαναλαμβάνει τις εντολές του κάθε κελιού για ένα πλήθος επαναλήψεων και προσφέρει μέσο όρο και τυπική απόκλιση χρόνου. Συγκρίνοντας τους Ψευδοκώδικες 4.2 & 4.3, παρατηρείται ότι η χρήση `feather` έχει νόημα, αφού υπερτερεί σε κάθε δοκιμή. Χαρακτηριστικότερο παράδειγμα και αυτό που οδήγησε στην υιοθέτηση του `feather`, αποτελεί ο χρόνος εγγραφής. Όπως φαίνεται στον Ψευδοκώδικα 4.2 η εγγραφή σε shapefile απαιτεί κατά μέσο όρο 3 λεπτά

Ψευδοκώδικας 4.3: Χρόνοι επεξεργασίας αρχείου Feather.

```
# Read Feather
%%timeit
gdf.read_feather('path/to/old.feather')
11.3 s ± 38.6 ms per loop (mean ± std. dev. of 7 runs, 1 loop each)

# Write to Feather
%%timeit
gdf.to_feather('path/to/new.feather')
41.8 s ± 641 ms per loop (mean ± std. dev. of 7 runs, 1 loop each)

# Write to Shapefile
%%timeit
gdf.to_file('path/to/new.shp')
3min 35s ± 1.67 s per loop (mean ± std. dev. of 7 runs, 1 loop each)
```

και 38 δευτερόλεπτα με τυπική απόκλιση 2 δεύτερα (αρκετά χαμηλή, άρα ο μέσος όρος αρκεί για να περιγραφεί), από την άλλη πλευρά η εγγραφή δεδομένων σε μορφή feather χρειάζεται κατά μέσο όρο μόνο 42 δεύτερα με τυπική απόκλιση 119 μιλιδευτερόλεπτα (επίσης αδιάφορη).

Επιπρόσθετα, να τονιστεί ότι οι δοκιμές πραγματοποιήθηκαν για σχετικά μικρές βάσεις δεδομένων (1.5GB) και παρόλα αυτά η διαφορά είναι αξιοσημείωτη. Οι διαφορές αποτελούν αποφασιστικό παράγοντα στη διαχείριση μεγαλύτερων βάσεων δεδομένων. Ωστόσο οι μορφές feather, parquet όσον αφορά τα geodataframes βρίσκονται ακόμα σε πρώιμο στάδιο (τουλάχιστον κατά την έκδοση 0.8.3), αφού δεν έχουν καθιερωθεί τα μεταδεδομένα των εν λόγω formats και για το λόγο αυτό πρέπει να ενσωματώνονται με μεγάλη προσοχή στα προτζεκτς, τουλάχιστον μέχρι να καθιερωθούν οι συμβάσεις των μεταδεδομένων.

Ακόμα, επειδή η κωδικοποίηση της βάσης δεδομένων καθώς και αυτή των αρχείων που έχουν ληφθεί, δε βρίσκονται στο ίδιο σύστημα συντεταγμένων (Ενότητα 2.2), απαιτείται η επαναπροβολή της βάσης στο σύστημα των δεδομένων. Αυτή είναι μία σχετικά εύκολη διαδικασία αλλά απαιτεί αρκετό χρόνο, αφού κάθε γεωμετρία πρέπει να υπολογιστεί εκ νέου για το καινούργιο σύστημα. Να σημειωθεί ότι το σύστημα αναφοράς μπορεί να αλλάζει από κάποια δορυφορική εικόνα σε κάποια άλλη, αυτό συμβαίνει διότι η περιοχή ενδιαφέροντος βρίσκεται μεταξύ δύο διαφορετικών ζωνών (Σχήμα 4.3). Ένας εύκολος τρόπος να ελαττωθεί το πλήθος μετατροπών μεταξύ των συστημάτων αποτελεί η ταξινόμηση των δεδομένων με βάση τη ζώνη στην οποία βρίσκονται, αρκετά εύκολο με τη δομή που χρησιμοποιείται κατά την αποθήκευση (Σχήμα 4.5). Στον Ψευδοκώδικα 4.4 αποτυπώνεται ο χρόνος αλλαγής προβολής της βάσης δεδομένων από EPSG:25831 σε EPSG:32631.

Ψευδοκώδικας 4.4: Αλλαγή προβολής βάσης δεδομένων.

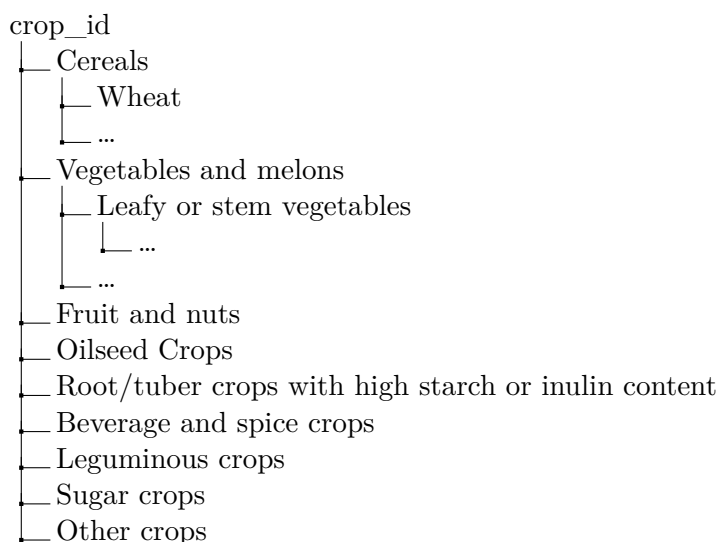
```
# Default crs
gdf.to_feather('path/to/db.feather')
band_data = rasterio.open('path/to/data.jp2', driver='JP2OpenJPEG')
print(gdf.crs) # EPSG:25831
print(band_data.crs.data['init']) # EPSG:32631

# Change crs
%%timeit
gdf.to_crs(band_data.crs.data['init'])
2min 7s ± 641 ms per loop (mean ± std. dev. of 7 runs, 1 loop each)
```

4.3.3 Ενοποίηση και Κωδικοποίηση Βάσης Δεδομένων

Προτού προχωρήσει η διαδικασία τμηματοποίησης των εικόνων που έχουν συλλεχθεί, απαιτείται μία πολύ σημαντική ενέργεια πάνω στη Βάση των Δεδομένων. Η ενέργεια αυτή περιλαμβάνει την κωδικοποίηση των εγγραφών που περικλείει στο σύνολο της η Βάση Δεδομένων. Συγκεκριμένα, κάθε εγγραφή, δηλαδή κάθε διαφορετικό αγροτεμάχιο, θα αποκτήσει ένα συγκεκριμένο αναγνωριστικό κλειδί που θα ονομαστεί `parcel_id`. Επιπρόσθετα, αγροτεμάχια που απευθύνονται στην ίδια κατηγορία π.χ δύο διαφορετικά αγροτεμάχια στη Βάση Δεδομένων, που με βάση το χώρο βρίσκονται σε διαφορετικές τοποθεσίες, αλλά καλλιεργούνται και στις δύο λ.χ ρύζι, θα κωδικοποιηθούν με το ίδιο αναγνωριστικό κλειδί ή όπως θα αναφέρεται `crop_id`.

Η κωδικοποίηση των διάφορων καλλιεργειών πραγματοποιείται με βάση ενός μοντέλου που προτείνεται από τον Food and Agriculture Organization of the United Nations (FAO) [124]. Το μοντέλο αποτελείται από 9 βασικές κατηγορίες, ονομαστικά Cereals, Vegetables and melons, Fruit and nuts, Oilseed Crops, Root/tuber crops with high starch or inulin content, Beverage and spice crops, Leguminous crops, Sugar crops και Other crops. Κάθε μία χωρίζεται έως και 9 υποκατηγορίες, οι οποίες πάλι χωρίζονται έως 9 μικρότερες κατηγορίες. Στο προτεινόμενο μοντέλο, υπάρχει και μία τέταρτη διαίρεση, αλλά δε θα ληφθεί υπ' όψιν για την παρούσα υλοποίηση. Το τελικό μοντέλο κωδικοποίησης, μπορεί να αποτυπώσει ένα τεράστιο αριθμό διαφορετικών καλλιεργειών με μόνο τρία ακέραια ψηφία. Η όλη διαδικασία μπορεί να αντιληφθεί ευκολότερα, όταν αποτυπωθεί με τη μορφή δένδρου (Σχήμα 4.6).



Σχήμα 4.6: Εξαγωγή `crop_id`.

Με βάση την παραπάνω λογική, κατά τη δημιουργία του τελικού Συνόλου Δεδομένων (Ενότητα 4.4), εξάγονται δύο “labels” για το κάθε patch, κάθε ένα από τα οποία έχουν τις ίδιες διαστάσεις με αυτές του patch. Το πρώτο και αυτό που αναφέρεται ως labels, αποτελεί έναν τελικό πίνακα που σε κάθε κελί περιέχεται η κατηγορία της καλλιέργειας ή το `crop_id`. Ενώ το δεύτερο, το οποίο αναφέρεται ως parcels, είναι ένας πίνακας όπου το κάθε κελί είναι ο κωδικός του αγροτεμαχίου ή το `parcel_id`. Το παραπάνω απεικονίζεται στα Σχήματα 4.15, 4.17, όμως λόγω του τεράστιου εύρους τόσο των διαφορετικών αγροτεμαχίων, όπως και των κατηγοριών, οι αποχρώσεις των colormaps του matplotlib δεν επαρκούν για την επαρκή απεικόνισή τους.

Δημιουργείται το Σχήμα 4.7, που αποτελεί ένα απλοϊκό παράδειγμα επεξήγησης των όσων αναφέρθηκαν. Ο λόγος που διατηρούνται τα διαφορετικά αγροτεμάχια, είναι διότι η συμπεριφορά της ανάπτυξης των ίδιων καλλιεργειών μπορεί να διαφέρει από αγροτεμάχιο σε αγροτεμάχιο. Διατηρώντας περισσότερη πληροφορία, παρέχεται μεγαλύτερη δυναμικότητα κατά την ανάπτυξη αλγορίθμων ή/και μοντέλων.

341	341	0	611
0	341	0	611
0	0	0	611
611	611	0	611

(i) Labels.

1	1	0	2
0	1	0	2
0	0	0	2
3	3	0	2

(ii) Parcels.

Σχήμα 4.7: Παράδειγμα Labels και Patches.

Όπως φαίνεται στο Σχήμα 4.7, κοιτώντας των πίνακα των αγροτεμαχίων (Parcels), τα αγροτεμάχια 1, 2 βρίσκονται σε διαφορετικές τοποθεσίες, έχουν διαφορετικό σχήμα, αλλά εμπεριέχουν την ίδια καλλιέργεια (611). Το παραπάνω μοτίβο συναντάται συχνά σε όλη την έκταση του συνόλου δεδομένων, αφού ίδια αγροτεμάχια μπορούν να χωρίζονται από δρόμους, λίμνες, βουνά, κ.α. Τα διάφορα αγροτεμάχια εμφανίζονται σε πολλαπλά είδη σχημάτων και δεν θα έπρεπε να θεωρούνται ένα τεράστιο αγροτεμάχιο με βάση την κατηγορία.

Το παραπάνω πρόβλημα, μπορεί να γίνει καλύτερα αντιληπτό όταν για παράδειγμα παρατηρείται μία εικόνα που περιλαμβάνει διαφορετικά ένα σύνολο από γάτες. Κάθε γάτα πρέπει να θεωρείται ως μία ξεχωριστή οντότητα, η οποία μοιράζεται κοινά χαρακτηριστικά με τη μεγαλύτερη κατηγορία “γάτα”, όμως διαφοροποιείται από αυτή ανάλογα τη ράτσα της εκάστοτε γάτας. Το ίδιο ακριβώς πρέπει να ληφθεί υπ’ όψιν για όλο το πλήθος των αγροτεμαχίων.

4.3.4 Οικονομία Κατά την Αποθήκευση

Ένα αρκετά ενδιαφέρον στοιχείο που συχνά παραβλέπεται είναι οι τύποι αποθήκευσης των δεδομένων στο δίσκο. Ο τρόπος αναπαράστασης των δεδομένων σε δυαδική μορφή για την αποθήκευσή τους είναι τετριμμένος, με λίγα λόγια, όσο μεγαλύτερο εύρος για αναπαράσταση χρειάζεται τόσο περισσότερα δυαδικά ψηφία πρέπει να χρησιμοποιηθούν για την αποθήκευσή τους.

Αν και η Python είναι δυναμική γλώσσα προγραμματισμού, με την έννοια ότι οι μεταβλητές της μπορούν να αλλάξουν τύπους καθ’ όλη τη διάρκεια του προγράμματος, οι βιβλιοθήκες που χρησιμοποιούνται για την επιτάχυνση των υπολογισμών απαιτούν στατικούς τύπους δεδομένων. Αυτό μπορεί εύκολα να ξεπεραστεί χρησιμοποιώντας τους μεγαλύτερους, σε εύρος, τύπους μεταβλητών προκειμένου να εκτελεστούν οι απαραίτητοι υπολογισμοί, ωστόσο αυτό αποτελεί κακό σχεδιασμό διότι δεσμεύεται περισσότερη μνήμη κατά τις ρουτίνες αλλά και περισσότερη θέση στο δίσκο κατά την αποθήκευση. Αυτό που πρέπει να πραγματοποιηθεί είναι μία έρευνα για τις μέγιστες και ελάχιστες τιμές που μπορεί να λάβει κάθε εικονοστοιχείο του τελικού συνόλου δεδομένων και να χρησιμοποιηθούν οι κατάλληλοι τύποι.

Η κωδικοποίηση των κατηγοριών των αγροτεμαχίων της Βάσης Δεδομένων (parcel_id) γίνεται με ακέραιους αριθμούς με εύρος $[0, 999]$, άρα απαιτούνται ακέραιες μεταβλητές μεγαλύτερες ή ίσες του μηδενός που λαμβάνουν τιμές μέχρι το χίλια. Επομένως, η μεταβλητή uint16 (unsigned integer 16 bit) με εύρος $[0, 2^{16} - 1]$ είναι αρκετή.

Για τον ίδιο λόγο, επειδή όμως τα αγροτεμάχια ταυτίζονται με τις εγγραφές της Βάσης Δεδομένων και μπορούν να αυξηθούν δραματικά για μεγάλες Βάσεις Δεδομένων, χρησιμοποιείται ο uint32 (unsigned integer 32 bit) με εύρος $[0, 2^{32} - 1]$.

Για την αποθήκευση των δεδομένων από τους δορυφόρους χρησιμοποιείται ο προεπιλεγμένος τύπος uint16, αφού η ελάχιστη τιμή κάθε εικονοστοιχείου μπορεί να είναι 0 (όχι δειγματοληψία) και η μέγιστη 10000.

Συνοπτικά, χρησιμοποιώντας τύπους δεδομένων συγκεκριμένους για το πρόβλημα μπορεί να μειώσει αισθητά τους απαιτούμενους πόρους επεξεργασίας και αποθήκευσης. Στη συγκεκριμένη περίπτωση, αν θεωρηθεί ότι σε κάθε περίπτωση ήταν να χρησιμοποιηθεί ο μεγαλύτερος τύπος δεδομένων (π.χ uint64), κάθε εικονοστοιχείο του patch θα απαιτούσε $64 + 64 + 64 \cdot (64 * 13) = 960$ bits.

Η αναλυτικά για όλο το patch, το patch περιέχει περίπου 50 χρονικές στιγμές, άρα τα δεδομένα για τα κανάλια είναι επί 50. Για τα κανάλια με τα 10 μέτρα ανάλυση (4 στο σύνολο) και τα αρχεία labels, parcels (2) έχουν $366 \cdot 366 = 133956$ εικονοστοιχεία ή $(4 \cdot 50 + 2) \cdot 133956 = 27059112$, τα κανάλια με 20 μέτρα ανάλυση (6) είναι $6 \cdot 50 \cdot 183 \cdot 183 = 10046700$ και τελικά αυτά των 60 μέτρων (3 συνολικά) $3 \cdot 50 \cdot 61 \cdot 61 = 558150$. Αθροίζοντας όλα τα παραπάνω, προκύπτουν $37.663.962$ εικονοστοιχεία των 64 bit άρα $241.0493.568$ bits ή 287.35 MB. Αυτό για ένα patch, στη συγκεκριμένη περίπτωση δημιουργούνται 1.800 patches ή 505.11 GB.

Εκτελώντας την παραπάνω απλή προσέγγιση, δηλαδή χρήση (uint16, uint32) όπου είναι εφικτό, το τελικό μέγεθος του patch υπολογίζεται σε 72.0938 MB. Δηλαδή 126.73 GB για 1.800 patches. Πετυχαίνοντας ελάττωση 75% στον απαιτούμενο χώρο αποθήκευσης. Σημειώνεται, ότι πέρα από την παραπάνω προσέγγιση, εφαρμόζεται και εσωτερική συμπίεση χωρίς απώλειες, επομένως το τελικό μέγεθος είναι ακόμα χαμηλότερο.

4.3.5 Τμηματοποίηση Αρχικής Εικόνας

Η αρχική εικόνα από το Sentinel 2 μεγέθους 10980×10980 , θα χωριστεί σε $\frac{n}{k} \times \frac{n}{k}$ τετράγωνα, όπου k ο λόγος $\frac{10}{band\ resolution}$ με $band\ resolution$ την ανάλυση της εκάστοτε ζώνης. Τα τετράγωνα αυτά θα ονομαστούν και θα αναφέρονται ως patches.

Οι παραγόμενες εικόνες είναι ευκολότερες προς τη διαχείριση, καθιστούν ικανή τη χρήση αλγορίθμων μηχανικής μάθησης. Επίσης, όπως παρουσιάζεται στις επόμενες ενότητες, μπορούν να επεξεργαστούν περαιτέρω και να απορριφθούν επιμέρους patches και όχι ολόκληρη η αρχική εικόνα, γεγονός που συμβάλει στη διατήρηση πληροφορίας ενώ παράλληλα εξοικονομεί χώρο και υπολογιστική ισχύ.

4.3.6 Αφαίρεση Κενών Εικονοστοιχείων

Το Σχήμα 4.8 αποτυπώνει μία χρονική στιγμή κατά την οποία, ο δορυφόρος δεν είναι ικανός να πραγματοποιήσει δειγματοληψία στην περιοχή και άρα τα δεδομένα απουσιάζουν (μαύρα εικονοστοιχεία). Αυτό μπορεί να οφείλεται σε τυχόν επικάλυψη που έχουν οι δίδυμοι δορυφόροι. Τα δεδομένα αυτά δεν προσφέρουν διορατικότητα κατά την ταξινόμηση αλλά και επιβαρύνουν πρώτον υπολογιστικά το δίκτυο και δεύτερον σε επίπεδο αποθηκευτικού χώρου. Αποφασίζεται λοιπόν, να απορριφθούν οι εικόνες που παρουσιάζουν πληρότητα κάτω από ένα κατώφλι θ (εδώ χρησιμοποιείται $\theta = 1$ (100%), δηλαδή αν έστω και ένα εικονοστοιχείο έχει τιμή μηδέν εκείνη τη χρονική στιγμή.

Η παραπάνω απόφαση αφαιρεί έναν αξιοσημείωτο αριθμό patches, τα οποία στη πληθώρα τους είναι εξ ολοκλήρου μαύρα (μηδέν). Τονίζεται ότι η παραπάνω κατάσταση δε εντοπίζεται συνεχώς, παρά μόνο σε ελάχιστες χρονικές στιγμές.

4.3.7 Αφαίρεση Patches χωρίς Κατηγορίες

Κατά τη δημιουργία του Συνόλου Δεδομένων, παρατηρήθηκαν patches στα οποία δεν περιέχονται αγροτεμάχια. Τέτοιες καταστάσεις συμβαίνουν όταν η Βάση Δεδομένων που χρησιμοποιείται, δεν συγκροτεί εγγραφές με γεωμετρίες που να εντάσσονται στο εν λόγω patch. Οι περιπτώσεις αυτές δεν είναι πολλές, τουλάχιστον για τα tiles 31TCG, 31TDG.

Σίγουρα όμως, αν γινόταν εκμετάλλευση και των δεκαέξι tiles (Σχήμα 4.3), τότε αρκετά patches θα είχα δεδομένα χωρίς ουσιαστική κατηγοριοποίηση. Θα έδειχναν δηλαδή σε τεράστιες εκτάσεις που δεν καλλιεργούνται επειδή δε γίνεται να εκμεταλλευτούν (π.χ θάλασσες, βουνά, κ.α) ή/και να έδειχναν σε άλλες χώρες για τις οποίες δεν έχουν φορτωθεί οι ανάλογες βάσεις δεδομένων.

Αν θεωρηθούν ως background κατηγορία όλα τα εικονοστοιχεία που δεν αντιστοιχούν σε κάποιο αγροτεμάχιο, τότε διατηρώντας όλα τα παραπάνω δεδομένα, η κατηγορία background υπερέχει αριθμητικά από όλες τις άλλες, οδηγώντας με αυτόν τον τρόπο σε ανισόρροπες κατανομές μεταξύ των κατηγοριών.

Ακόμα, τέτοια δεδομένα ανεβάζουν σημαντικά το τελικό μέγεθος του Συνόλου Δεδομένων, χωρίς να προσφέρουν περισσότερη πληροφορία για το πρόβλημα που προσπαθεί να λυθεί.

Για τους παραπάνω λόγους, αποφασίστηκε να απορριφθούν τέτοιες περιπτώσεις, δηλαδή οποιοδήποτε patch δεν περιέχει αγροτεμάχια να μη λαμβάνεται υπ' όψιν.



Σχήμα 4.8: Εικόνα Δορυφόρου Sentinel 2B στις 22/05/2020.

4.3.8 Αφαίρεση Χρονικών Στιγμών με μη Ακέραιο Αριθμό Καναλιών

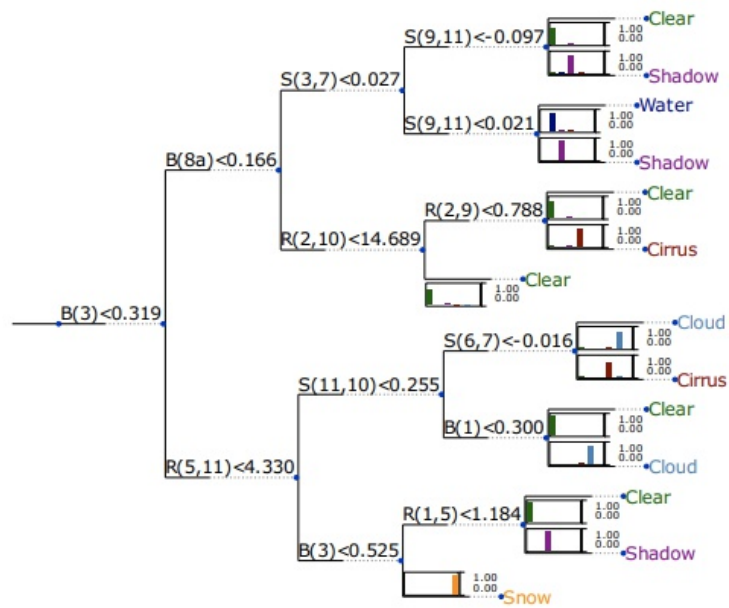
Μετά το πέρας των παραπάνω διαδικασιών, που αφαιρούν σταδιακά δεδομένα, υπάρχουν αρκετές περιπτώσεις όπου έχουν μείνει δεδομένα μόνο για συγκεκριμένα κανάλια και όχι για όλα. Τέτοιες περιπτώσεις προφανώς και δεν μπορούν να αξιοποιηθούν, διότι δεν περιέχουν σφαιρική εικόνα για την εκάστοτε χρονική στιγμή καθώς και δε συνεισφέρουν στην κανονικοποίηση του τελικού Συνόλου Δεδομένων, αφού δεν πληρούν τα κριτήρια της πληρότητας (completeness) και της συνέπειας (consistency).

Δεδομένα που δεν περιέχουν τη συγκεκριμένη χρονική στιγμή σε όλα τα κανάλια (13) δειγματοληψίας, απορρίπτονται.

4.3.9 Εντοπισμός και Εξαγωγή Μάσκας Συννεφιάς

Αν και όχι όλα, αρκετά στιγμιότυπα από τα δεδομένα εμπεριέχουν σύννεφα. Ζητώντας εξαρχής εικόνες με έως και 10% συνολική συννεφιά από το δορυφόρο λύνει σε έναν βαθμό το πρόβλημα, ωστόσο το ποσοστό αυτό ανάγεται σε όλη την αρχική εικόνα. Όταν λοιπόν, εξάγονται τα επιμέρους patches συναντάται η κατάσταση κατά την οποία ολόκληρα patches περιέχουν μόνο εικόνα από σύννεφο. Μία από τις μεθοδολογίες που χρησιμοποιείται για την εξάλειψη της συννεφιάς, είναι η συγκέντρωση των patches σε χρονικές περιόδους και στη συνέχεια χρήση κάποιας μετρικής π.χ μέσου όρου ή διαμέσου.

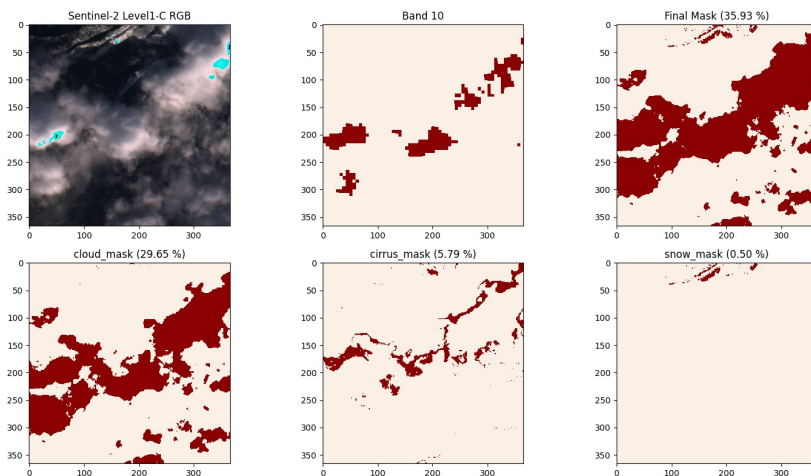
Το παραπάνω, αν και αποτελεσματικό όταν υπάρχει μόνο μία χρονική στιγμή με σύννεφα, αποτυγχάνει εάν αρκετές ή/και όλες οι χρονικές στιγμές περιέχουν μόνο εικόνες με σύννεφα. Συνεπώς, ενσωματώνεται μία ακόμα τεχνική εφόσον παρέλθει η χρονική συνάθροιση όπως προτείνεται στο [125]. Συγκεκριμένα, προτείνεται ένα δένδρο απόφασης με βάση τιμές που προκύπτουν από τις επιμέρους ζώνες δειγματοληψίας ή/και συναθροίζοντας τες. Για να λυθεί το πρόβλημα, κύριος σκοπός αποτελεί ο εντοπισμός των Σύννεφων (Cloud) και της σκιάς αυτών (Shadow) στα φύλλα του δένδρου. Το εν λόγω Δένδρο Απόφασης απεικονίζεται στο Σχήμα 4.9 και με βάση τους συγγραφείς πετυχαίνει 91% ακρίβεια στις ταξινομήσεις του.



Σχήμα 4.9: Δένδρο Απόφασης για Αφαίρεση Συννεφιάς (Πηγή: [125] Σχήμα 8).

Στην περίπτωση της παρούσας εργασίας, υπολογίζεται μία τελική δυαδική μάσκα των θέσεων που ικανοποιούν τις επιμέρους συνθήκες Cloud, Cirrus, Snow του Σχήματος 4.9. Η παρουσία των σύννεφων ή/και του χιονιού, μπορεί οπτικά σε μερικές περιπτώσεις να μην αλλοιώνουν το περιεχόμενο της φωτογραφίας, αλλά σε επίπεδο εικονοστοιχείου αλλάζουν τελείως την ταυτότητα του.

Τα αποτελέσματα της χρήσης του αλγόριθμου Hollstein [125], αποτυπώνονται καλύτερα στο Σχήμα 4.10, όπου παρουσιάζεται το προϊόν του δορυφόρου σε RGB (Ζώνες 4, 3, 2). Η Ζώνη 10 ή αλλιώς cirrus band [126], σε διαστάσεις ίδιες με του patch, κύρια εφαρμογή της οποία είναι η ανίχνευση της συννεφιάς. Τέλος, δίνονται οι επιμέρους μάσκες, τα ποσοστά τους καθώς και η τελική μάσκα που χρησιμοποιείται για την απόρριψη.

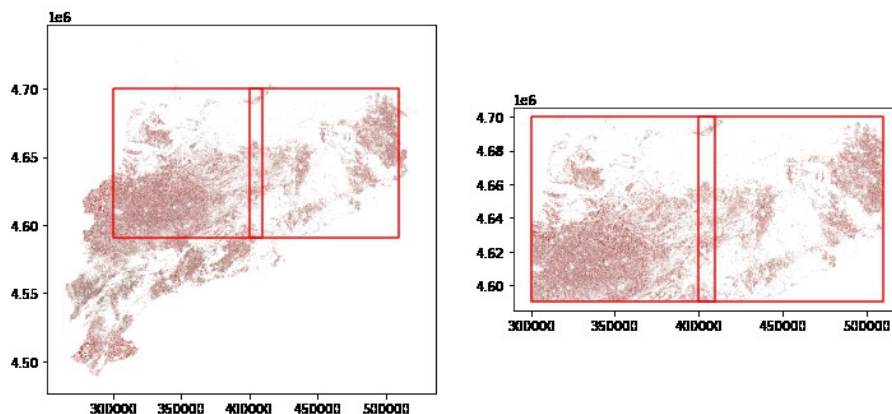


Σχήμα 4.10: Εξαγωγή Μασκών με τη χρήση του Αλγόριθμου Hollstein [125] (Ημ. Προϊόντος 2020/01/31).

Προκειμένου, να δημιουργηθεί το Σχήμα 4.10 είναι απαραίτητη η “υποβάθμιση” των δορυφορικών ζωνών από το εύρος των ακεραίων [0, 10000] σε [0, 255]. Για το λόγο αυτόν, η παρουσία “αλλοιωμένων” (corrupted) εικονοστοιχείων (λ.χ μπλε) είναι αναμενόμενη. Τονίζεται, ότι το παραπάνω πραγματοποιείται μόνο για τη δημιουργία των σχημάτων και ότι τα προϊόντα του δορυφόρου παραμένουν αμετάβλητα.

4.3.10 Περαιτέρω Βελτιστοποιήσεις Βάσης Δεδομένων

Όπως με τα προβλήματα που προκαλεί η δημιουργία πολλαπλών αρχείων (Ενότητα 4.4), για τους ίδιους λόγους η διαχείριση Βάσεων Δεδομένων δυσχεραίνει όσο αυτές μεγαλώνουν σε μέγεθος. Στην παρούσα περίπτωση δεν αποτελεί πρόβλημα, αφού η Βάση αναφοράς περιέχει πλήθος αγροτεμαχίων μικρότερο του ενός εκατομμυρίου καθώς και το εμβαδόν τους είναι περιορισμένο.



Σχήμα 4.11: Βάση Δεδομένων πριν (αριστερά) και μετά (δεξιά) το Overlay.

Βέβαια από τις διαδικασίες που εκτελούνται πάνω στη Β.Δ, όπως για παράδειγμα ρουτίνες για δημιουργία ράστερ, μπορεί να παρατηρηθεί ότι ο απαιτούμενος χρόνος θα αποτελέσει πρόβλημα σε μεγαλύτερες εκτάσεις Γης ή σε χώρες με εκατομμύρια αγροτεμάχια.

Μία λύση που χρησιμοποιήθηκε σε αυτή την εργασία και προτείνεται είναι η χρήση του Ελάχιστου Περιβάλλον Ορθογωνίου (Minimum Bounding Rectagular; MBR) που πλαισιώνει τα tiles που χρησιμοποιούνται. Εφόσον, δηλαδή, παρέλθει οποιαδήποτε μορφή προ-επεξεργασίας της Βάσης τότε δημιουργείται ένα GeoDataFrame με μία μόνο εγγραφή, τη γεωμετρία του MBR όλων των tiles. Στη συνέχεια, πραγματοποιείται overlay διαδικασία για να “κοπεί” η αρχική βάση πάνω στο MBR. Το τελικό αποτέλεσμα περιλαμβάνει την τομή της Βάσης και του MBR.

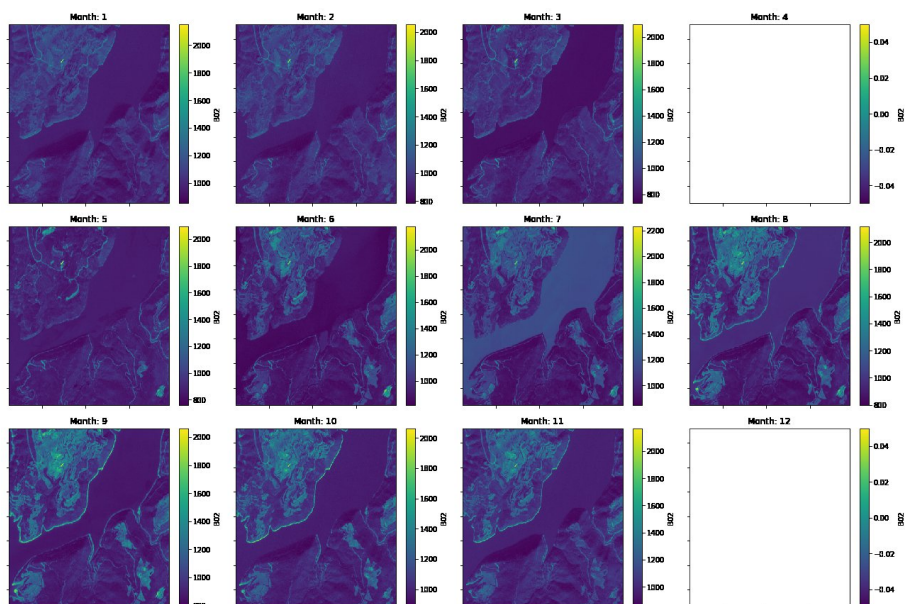
Το παραπάνω απεικονίζεται καλύτερα στο Σχήμα 4.11, όπου με κόκκινο παρουσιάζονται τα δύο tiles (31TCG, 31TDG) που έχουν επιλεγθεί. Όπως φαίνεται, τα tiles παρουσιάζουν μία σχετική επικάλυψη, αυτό είναι συχνό φαινόμενο, ωστόσο δε δημιουργεί κάποιο πρόβλημα κατά τη διαδικασία του overlay. Στην αριστερή εικόνα, παρουσιάζεται η Βάση στην αρχική της κατάσταση και δεξιά, μετά την επεξεργασία της. Με βάση το Σχήμα, τα περιεχόμενα της Βάσης έχουν ελαττωθεί αρκετά, διατηρώντας πάντα άθικτη την περιοχή μελέτης.

4.3.11 Συμπλήρωση Κενών Μηνών

Για τους διάφορους προαναφερόμενους λόγους, ενδέχεται πολλές χρονικές στιγμές να έχουν φιλτραριστεί και άρα απορριφθεί. Αυτό δημιουργεί πρόβλημα, διότι αν για παράδειγμα θεωρηθεί ως συχνότητα συνάθροισης των δεδομένων ο μήνας, δεν υπάρχει σταθερός αριθμός παρατηρήσεων ανά μήνα. Βέβαια, η συχνότητα ομαδοποίησης είναι ένα νοητό πλαίσιο που χρησιμοποιείται για την καλύτερη ανασκόπηση και παρατήρησης των δεδομένων.

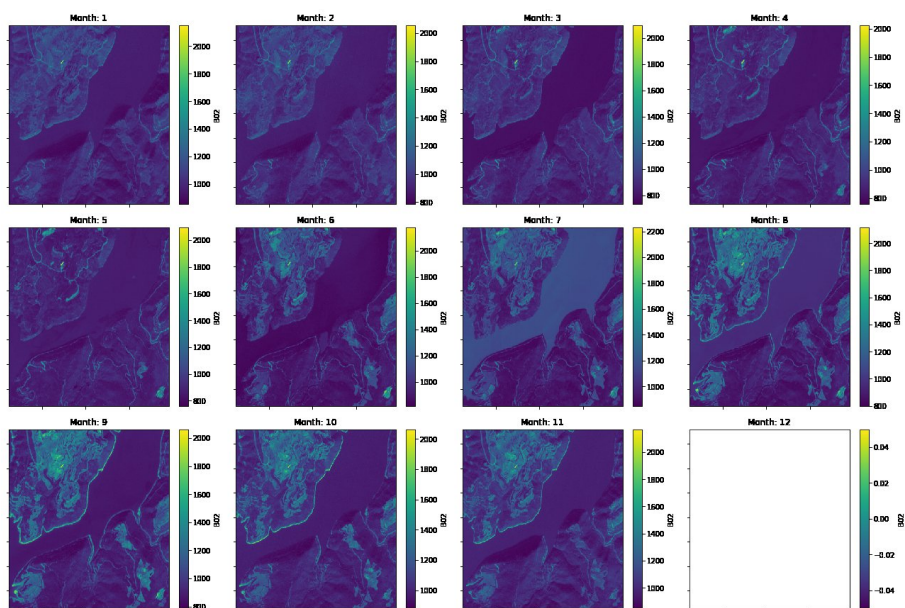
Στην πραγματικότητα, η ομαδοποίηση είναι κάτι που συμβαίνει γιατί προσδίδει σταχαστικότητα στον τρόπο φόρτωσης και εισαγωγής των δεδομένων στο μοντέλο. Για τα επόμενα παραδείγματα, θεωρείται ως συχνότητα ομαδοποίησης ο ένας μήνας και σαν συνάρτηση συνάθροισης χρησιμοποιείται η διάμεσος σε επίπεδο εικονοστοιχείου, έτσι δημιουργείται το Σχήμα 4.12. Το εν λόγω σχήμα, αποτυπώνει το κανάλι B02 για το συγκεκριμένο patch, το συγκεκριμένο tile και τη συγκεκριμένο έτος. Το ίδιο συμβαίνει και για τα υπόλοιπα κανάλια του συγκεκριμένου patch, αφού απαραίτητη προϋπόθεση του συνόλου δεδομένων είναι η παρουσία όλων των καναλιών μετά το στάδιο της προ-επεξεργασίας (Ενότητα 4.3.8).

Το Σχήμα 4.12 απεικονίζει αρκετά καλά το πρόβλημα που εξηγήθηκε παραπάνω. Όπως φαίνεται, οι μήνες 4 και 12 είναι κενοί. Η συμπλήρωση δεδομένων που δε βρίσκονται σε πρώτες ή τελευταίες θέσεις ή καλύτερα σε θέσεις, όπου υπάρχει δειγματοληψία για προηγούμενες και μελλοντικές χρονικές στιγμές,



Σχήμα 4.12: Συνάθροιση Καναλιού Β02 ανά Μήνα με Διάμεσο (i).

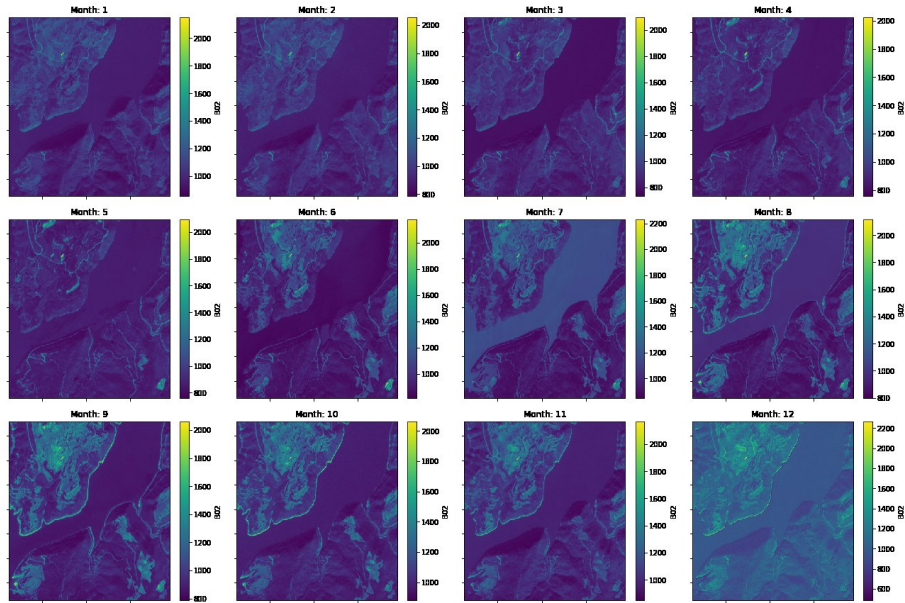
είναι αρκετά εύκολη και όπως θα φανεί οδηγεί σε ικανοποιητικά αποτελέσματα.



Σχήμα 4.13: Συνάθροιση Καναλιού Β02 ανά Μήνα με Διάμεσο (ii).

Για τη συμπλήρωση τέτοιων χρονικών στιγμών, χρησιμοποιείται η μεθοδολογία linear interpolation και τα αποτελέσματα παρουσιάζονται στο Σχήμα 4.13.

Όπως παρατηρείται από το Σχήμα 4.13 επιτυγχάνεται μία ικανοποιητική συμπλήρωση του μήνα που έλειπε (Μήνας 4). Όμως, η διαδικασία έχει αποτύχει να συμπληρώσει τον τελευταίο μήνα (Μήνα 12). Σε αυτή την περίπτωση, η συμπλήρωση είναι αρκετά δυσκολότερη αφού πρέπει να γίνει μία πρόβλεψη με βάση την κατανομή που ακολουθούν τα δεδομένα σε προηγούμενες χρονικές στιγμές. Η ρουτίνα που εφαρμόζεται ονομάζεται extrapolation και προφανώς το αποτέλεσμα αναμένεται να είναι αμφιλεγόμενο εφόσον βασίζεται σε πρόβλεψη. Η παραπάνω πρόταση υποστηρίζεται από το Σχήμα 4.14, με βάση το οποίο, η χρονική στιγμή που συμπληρώθηκε, και μεν ακολουθεί τη γενικότερη δομή της περιοχής μελέτης, ωστόσο η αλλαγή των τιμών σε όλο το εύρος των εικονοστοιχείων παρουσιάζει αισθητή διαφορά.



Σχήμα 4.14: Συνάθροιση Καναλιού B02 ανά Μήνα με Διάμεσο (iii).

4.4 Τελική Μορφή (Σύνολο Εκπαίδευσης)

Η τελική μορφή του συνόλου δεδομένων είναι εμπνευσμένη από το BigEarthNet [127], με την ειδοποιός διαφορά ότι η παρούσα προσέγγιση χωρίζει το κάθε tile σε $x \times x$ patches και πραγματοποιείται τη διαχείριση δεδομένων τύπου Level-1C εν αντιθέσει των ατμοσφαιρικά διορθωμένων που παρέχει το BigEarthNet.

Κάθε patch εμπεριέχει n στιγμιότυπα για κάθε ένα από τα διαθέσιμα κανάλια της εν λόγω περιοχής, ενώ συμπεριλαμβάνονται αρχεία labels.tiff, parcels.tiff και annotations.csv. Το αρχείο labels.tiff εμπεριέχει τις κατηγορίες των αγροτεμαχίων, το parcels.tiff αποτυπώνει τα διαφορετικά αγροτεμάχια και το annotations.csv αποτελείται από τις γεωμετρίες, τα πλαίσια (bounding boxes) και την κατηγορία του κάθε αγροτεμαχίου, αντίστοιχα. Το προαναφερόμενο σενάριο απεικονίζει καλύτερα στο Σχήμα 4.17.

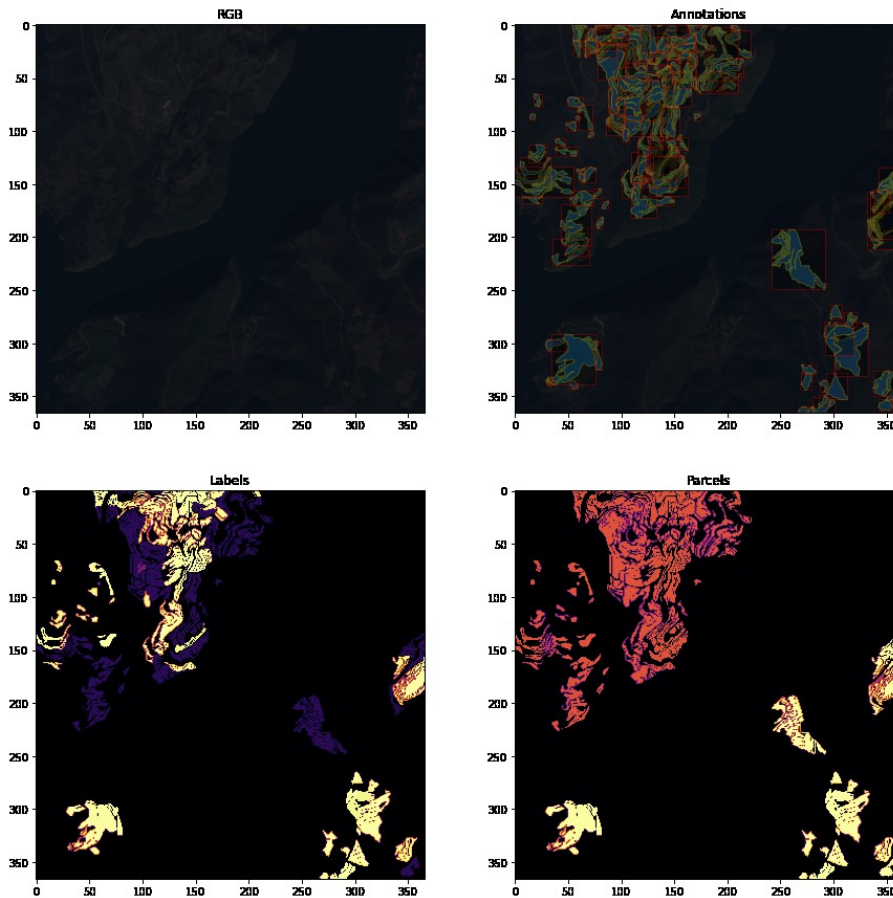
Ο Πίνακας 4.1 περιγράφει την εσωτερική δομή του αρχείου annotations.csv. Συγκεκριμένα, κάθε εγγραφή του αρχείου αυτού περιέχει πέντε στήλες. Η Στήλη **parcel_id** αναφέρεται στο μοναδικό αναγνωριστικό του αγροτεμαχίου μέσα στη Βάση Δεδομένων, το πεδίο **crop_type** στην ευρύτερη κλάση που ανήκει το εν λόγω αγροτεμάχιο. Η Στήλη **segmentation** αποτελεί έναν μετασχηματισμό του πολυγώνου που περικλείει το εκάστοτε αγροτεμάχιο, από επίπεδο γεωγραφικών συντεταγμένων, σε επίπεδο εικονοστοιχείων μέσα στο συγκεκριμένο patch. Η στήλη **area** είναι το εμβαδόν του αγροτεμαχίου, ενώ τέλος το πεδίο **bbox** είναι το Ελάχιστο Περιβάλλον Ορθογώνιο (Minimum bounding rectangle; MBR) που περικλείει την περίμετρο.

Σημείωση, τα πεδία **segmentation**, **bbox** αναφέρονται σε θέσεις εικονοστοιχείων στο τωρινό patch και άρα λαμβάνουν τιμές στο διάστημα $[0, width)$ όπου $width$ το μήκος της εικόνας (θεωρείται ότι η αρχική εικόνα “κόβεται” σε ακέραιο αριθμό τετράγωνων patch). Ακόμα, το πεδίο **bbox** ακολουθεί τη μορφή XYWH_ABS και όχι τη XYXY.

Πίνακας 4.1: Αρχείο annotations.csv.

parcel_id	crop_type	segmentation	area	bbox
0	150	[[1, 239, 1, 240], [...]]	8.5	(0.0, 234.0, 1.0, 12.0)
1	150	[[...], ..., [...]]	11.5	(0.0, 221.0, 5.0, 10.0)
2	150	[[...]]	11	(0.0, 222.0, 1.0, 14.0)
3	150	[[...]]	4	(26.0, 198.0, 2.0, 4.0)

Ο πίνακας 4.1 αποτυπώνεται καλύτερα στο Σχήμα 4.15, όπου φαίνεται ξεκάθαρα η εσωτερική δομή των αρχείων. Σημαντικό πλεονέκτημα αποτελεί ότι δε γίνεται κάποια υπόθεση ότι όλη η εικόνα αποτελείται από μία μόνο κλάση. Ακόμα, η εξαγωγή των annotations επιτρέπει την επιστράτευση αρχιτεκτονικών Όρασης Υπολογιστών. Η χρήση μεγάλων σε εμβადόν εικόνων, εξασφαλίζει την εξαγωγή περισσότερων χωρικών συσχετίσεων, αφού το μοντέλο μπορεί να μάθει χαρακτηριστικά με βάση τους γείτονες.



Σχήμα 4.15: Annotations για το 2020_31TCG_patch_14_14 (Ημ. Προϊόντος 2020/01/31).

Η παραπάνω λογική λειτουργεί αρκετά καλά για τη συγκεκριμένη περίπτωση, δηλαδή όταν η περιοχή μελέτης είναι αρκετά μικρή. Βέβαια, το σύνολο δεδομένων δημιουργείται για δύο tiles ($num_tiles = 2$), μίας μόνο χρονιάς ($num_years = 1$). Συνεπώς, αν θεωρηθεί ότι δεν έχουν εξαχθεί χρονικές στιγμές (λόγω π.χ μαύρων εικονοστοιχείων), για μέγεθος patch 366×366 , δημιουργούνται 900 $num_patches$ ανά tile, κάθε ένα από τα οποία περιέχει περίπου 50 χρονικές στιγμές ($num_timestamps = 50$) για 13 κανάλια ($num_channels$). Άρα προκύπτει ότι αναμένονται 1.175.400 αρχεία (Εξίσωση 4.1).

$$num_files = num_years \cdot num_tiles \cdot num_patches \cdot (num_timestamps \cdot num_channels + 3) \quad (4.1)$$

Το πλήθος είναι αρκετά μεγάλο ακόμα και για ένα tile, αν η παραπάνω λογική επεκταθεί σε μεγαλύτερο χωρικό εύρος, θα αποτελέσει ανασταλτικό παράγοντα αξιοποίησης του αποτελέσματος. Αφού, οι διεργασίες I/O για τη διαχείριση μεγάλου αριθμού αρχείων, ακόμα και αν αυτά είναι ασήμαντου μεγέθους, απαιτούν σημαντικό χρόνο περάτωσης.

Επιπρόσθετα, σε αυτό το στάδιο έχει επιτυγχανθεί σε μεγάλο βαθμό η ευκολότερη διαχείριση των δορυφορικών εικόνων, διότι από το αρχικό μέγεθος 10980×10980 , βρίσκονται πλέον σε 366×366 , ένα μέγεθος δηλαδή που μπορεί εύκολα πλέον να διοχετευθεί σε αρχιτεκτονικές δικτύων βαθιάς μάθησης. Ωστόσο, όπως αναφέρθηκε παραπάνω το πλήθος των αρχείων αποτελεί αξιοσημείωτο εμπόδιο που δεν μπορεί απλά να αγνοηθεί και πέρα από το όνομα του εκάστοτε αρχείου, δεν υπάρχει ουσιαστική χρονική συσχέτιση μεταξύ των αρχείων μέσα στο ίδιο patch.

Για την αντιμετώπιση των παραπάνω, θα επιστρατευθούν τρεις επιπλέον τεχνολογίες, η βιβλιοθήκη της Python με όνομα xarray [128, 129], τα εργαλεία NCO NetCDF Operators [130] και εν τέλει το σαν container η κωδικοποίηση netCDF4 (ή αλλιώς γνωστή ως HDF5). Η διαδικασία που θα ακολουθηθεί περιγράφεται στη συνέχεια, η βασική ιδέα είναι ότι πλέον για κάθε patch δημιουργείται ένα τελικό netcdf4 (.nc) αρχείο με μοναδικό χαρακτηριστικό όνομα, το έτος, το tile και τον αριθμό του patch και εσωτερικά του περιέχονται όλες οι χρονοστιγμές για κάθε ένα από τα κανάλια, πλέον σε μορφή DataArray του xarray, συσχετισμένα στο χρόνο.

Με τον τρόπο αυτό, κάθε κανάλι διαβάζεται ως μία χρονοσειρά και είναι έτοιμο για περαιτέρω επεξεργασία (π.χ συνάθροιση ανά μήνα με δειγματικό μέσο) ή οποιαδήποτε άλλη ρουτίνα. Επιπλέον, ενσωματώνονται τα labels.tiff και parcels.tiff του συγκεκριμένου patch και πλέον, τα αρχεία αυτά μπορούν κάλλιστα να είναι χρονοσειρές, λ.χ σε περιπτώσεις εναλλασσόμενων καλλιεργειών μέσα στο ίδιο έτος. Τέλος, δίνονται επιλογές εσωτερικής συμπίεσης των netcdf αρχείων, ελαττώνοντας περαιτέρω το χώρο που καταλαμβάνουν.

Αρχικά, η διαδικασία διαδραματίζεται σε επίπεδο patch, δηλαδή ότι περιγράφεται λαμβάνει χώρα για κάθε patch και εφόσον τελειώσει η διαδικασία, επεξεργάζεται το επόμενο στη σειρά. Με τη βοήθεια του xarray και έχοντας ως engine το rasterio, διαβάζονται τα .tiff που έχουν παραχθεί σε προηγούμενο στάδιο. Επομένως, για κάθε ένα από τα δέκατρία κανάλια αναζητούνται τα ανάλογα .tiff, φορτώνονται ως xarray DataArray και δημιουργείται νέα συντεταγμένη χρονο (το όνομα του καναλιού και ο χρόνος μπορούν να βρεθούν από το όνομα του αρχείου). Όταν βρεθούν και φορτωθούν όλα τα .tiff για το συγκεκριμένο κανάλι αποθηκεύονται στο δίσκο με τη μορφή .nc (netcdf).

Εφόσον, δημιουργηθούν τα απαραίτητα netcdf για όλα τα κανάλια του συγκεκριμένου patch, χρησιμοποιείται η συνάρτηση “netgag” του NCO για να συγκεντρωθούν ως groups σε ένα τελικό netcdf Dataset (Ψευδοκώδικας 4.5).

Ψευδοκώδικας 4.5: Τελική ονομασία patch σε μορφή netcdf4.

```
YYYY_TTTTT_patch_XX_YY.nc
```

Όπου YYYY έτος (π.χ 2020), TTTTT κωδικός tile (π.χ 31TCG), XX γραμμή του patch (π.χ 15) και YY στήλη του tile (π.χ 15).

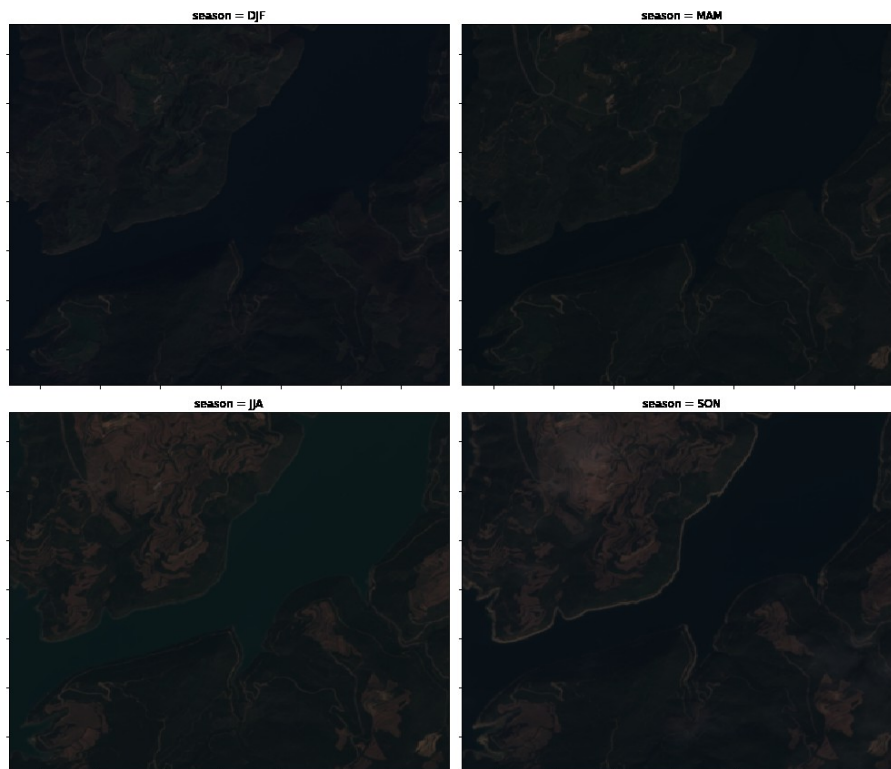
Μετά το πέρας της προαναφερόμενης ρουτίνας, η Εξίσωση 4.1 μπορεί να ξαναγραφεί ως Εξίσωση 4.2, όπου 2 το αρχείο netcdf και το αρχείο annotations.csv που δεν έχει ενσωματωθεί με κάποιο τρόπο ακόμα. ο νέος αριθμός των αρχείων είναι 3.600, άρα επιτυγχάνεται μείωση **99.7%** του πλήθους των απαιτούμενων αρχείων.

$$num_files = num_years \cdot num_tiles \cdot num_patches \cdot 2 \quad (4.2)$$

Βέβαια, μπορεί να γίνει μία επιπλέον επέκταση, η υιοθέτηση ενός αρχείου Common Objects in Context (COCO) [98]. Τα αρχεία σε μορφή COCO αποτελούν μία συμβατική χαρτογράφηση της διαδρομής των αρχείων δεδομένων (netcdf) μέσα στο δίσκο καθώς και επιπλέον μεταδεδομένων που μπορούν να φέρουν μαζί τους. Εξ ορισμού, ένα τέτοιο αρχείο πρέπει να περιέχει έξι καταλόγους, ανάλογα το πρόβλημα που καλείται να λύσει. Ονομαστικά, περιέχει, έναν κατάλογο με μεταδεδομένα του αρχείου (info), έναν κατάλογο με τις άδειες χρήσεις (licenses), έναν με τις κατηγορίες (categories), έναν με τις διαδρομές των αρχείων (images) και έναν annotations, που εσωτερικά του περιέχει τις στήλες του Πίνακα 4.1. Τέλος, ο κατάλογος segment_info χρησιμοποιείται σε προβλήματα rapoptic και άρα δε θα συμπεριληφθεί στο τελικό COCO. Παράγονται συνολικά 3 COCO αρχεία, ένα για εκπαίδευση (train), ένα για αξιολόγηση (val) κατά την εκπαίδευση και ένα τελικό (test) που θα αξιολογήσει τη τελική κατάσταση του μοντέλου.

Τα πλεονεκτήματα της υιοθέτησης του λεγόμενου COCO formatting, είναι αρκετά και φαίνονται αφού χρησιμοποιούνται σαν είσοδο σε πολλές σύγχρονες αρχιτεκτονικές βαθιάς μάθησης για προβλήματα Όρασης Υπολογιστών, διότι λειτουργούν σαν γεννήτριες και δεν απαιτούν το εξ ολοκλήρου φόρτωμα των δεδομένων στη μνήμη.

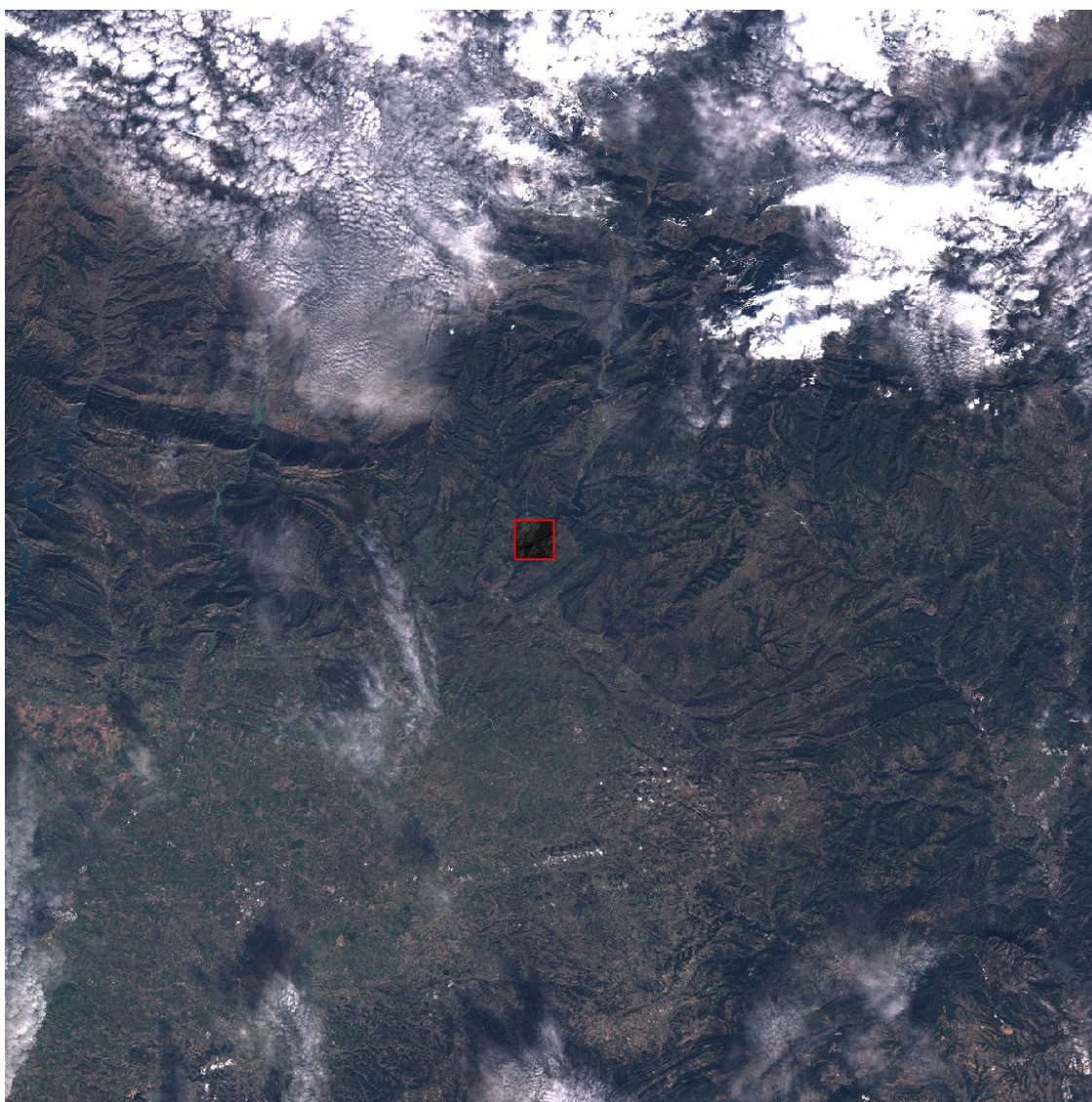
Η δημιουργία ενός τέτοιου αρχείου, καθιστά ικανό τον πειραματισμό με μία πληθώρα αρχιτεκτονικών με μικρές αλλαγές καθώς και τη χρήση τεχνικών μεταβίβασης της μάθησης (transfer learning), η οποία



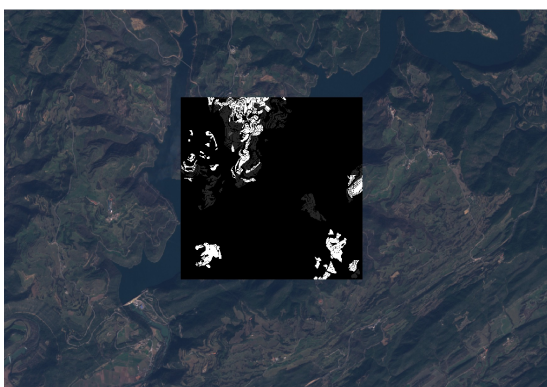
Σχήμα 4.16: 2020_31TCG_14_14 σε κάθε εποχή (Ομαδοποίηση με διάμεσο).

με βάση τη βιβλιογραφία [131, 132] δε φαίνεται να διευκολύνει τη σύγκλιση μοντέλων για γεωχωρικά δεδομένα. Αντιθέτως, παρατηρείται καλύτερη ακρίβεια στα μοντέλα τα οποία εκπαιδεύονται εξαρχής.

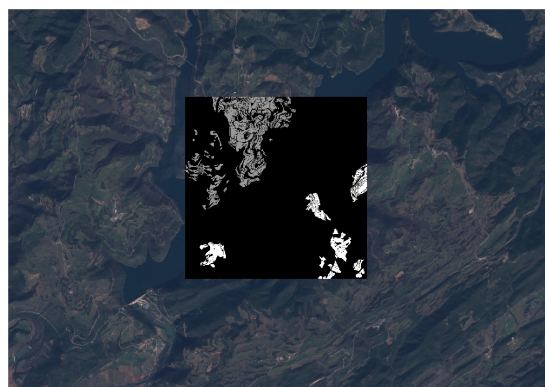
Το Σχήμα 4.16, συναθροίζει όλα τα στιγμιότυπα του 2020_31TCG_patch_14_14 στις τέσσερις εποχές. Χρησιμοποιούνται τα κανάλια κόκκινο, πράσινο, μπλε για οπτικοποίηση και ομαδοποιούνται με βάση τη διάμεση τους τιμή. Το παρόν, σε συνδυασμό με τα υπόλοιπα σχήματα του κεφαλαίου, προσφέρει μία ξεκάθαρη εικόνα για τη σημαντικότητα του χρόνου στη λήψη αποφάσεων κατά την ταξινόμηση αγροτεμαχίων.



(i) 2020_31TCG_patch_14_14 μέσα στην αρχική εικόνα.



(ii) 2020_31TCG_patch_14_14 Labels.



(iii) 2020_31TCG_patch_14_14 Parcels.

Σχήμα 4.17: Τμηματοποίηση Αρχικής Εικόνας σε Patches

Κεφάλαιο 5

Ταξινόμηση Αγροτεμαχίων

5.1 Σχετική Έρευνα

Οι κύριες κατηγορίες που ταξινομούνται οι διάφορες γνωστές αρχιτεκτονικές για προβλήματα ταξινόμησης σε δορυφορικά δεδομένα, απαριθμούνται σε δύο. Ονομαστικά, οι “N-to-1” και οι “N-to-N”, όπου N το πλήθος των παρατηρήσεων [133].

Συγκριτικά με τις κατηγορίες των μοντέλων, υπάρχουν τρεις μεθοδολογίες συγχώνευσης δεδομένων, σε επίπεδο εικονοστοιχείου (pixel-level), σε επίπεδο χαρακτηριστικών (feature-level) και σε επίπεδο απόφασης (decision-level). Οι πρώτες δύο κυριαρχούν στην πληθώρα των υλοποιήσεων [134].

Η πρώτη κατηγορία μοντέλων, δηλαδή η “N-to-1”, περικλείει όλες εκείνες τις αρχιτεκτονικές που λαμβάνουν ως είσοδο N αριθμό παρατηρήσεων και παράγουν μόνο μία έξοδο [133], λ.χ το όνομα της κλάσης. Εδώ υπάρχει η περισσότερη μελέτη, αφού η ταξινόμηση μίας εικόνας εισόδου σε επίπεδο εικονοστοιχείου (pixel-based classification) αποτελεί ένα διάσημο κλάδο της Τηλεπισκόπησης που έχει γνωρίσει μεγάλο ενδιαφέρον από νωρίς. Γνωστοί αλγόριθμοι Μηχανικής Μάθησης, όπως ο Random Forest και τα SVM (Support Vector Machines), χρησιμοποιήθηκαν [133, 134], ενώ με ενσωμάτωση τεχνικών όπως η επιλογή χαρακτηριστικών (Feature Selection), οι επιδόσεις τους μπορούν να βελτιωθούν περαιτέρω [134].

Η δεύτερη, “N-to-N” προσπαθεί να αντιμετωπίσει το πρόβλημα των εναλλασσόμενων καλλιιεργειών, τέτοια μοντέλα δίνουν ως έξοδο πολλαπλές τιμές N , για τις διάφορες χρονικές στιγμές της εκάστοτε καλλιέργειας. Μοντέλα RNN, συγκεκριμένα LSTM, συγκροτούν αυτή την κατηγορία λόγω της φύσης τους, που τα επιτρέπει να συσχετίζουν χαρακτηριστικά στο χρόνο [133].

Η διεξαγωγή των πειραμάτων θα εστιάσει περισσότερο σε αρχιτεκτονικές Όρασης Υπολογιστών όπως το Mask R-CNN [91] και το DETR [95]. Ενώ ακόμα, πειράματα θα γίνουν με ένα μοντέλο από τη βιβλιογραφία, που λαμβάνει υπ’ όψιν τόσο το χρονικό αλλά και το χωρικό κομμάτι, αποτελεί το DuPLO [97], για το οποίο θα πραγματοποιηθούν πειράματα. Όλα τα πειράματα θα αξιολογηθούν σε επίπεδο εικονοστοιχείου.

5.2 Mask R-CNN

Μία πρώτη υλοποίηση Βαθιάς Αρχιτεκτονικής αποτελεί το Mask R-CNN (Ενότητα 3.5.3) που προσφέρεται μέσω της βιβλιοθήκης Detectron2 [135]. Η βιβλιοθήκη προσφέρει πολλές λειτουργίες, όμως θα χρησιμοποιηθεί μόνο αυτή της ανίχνευσης αντικειμένων (Ενότητα 3.5).

Βασικό πρόβλημα αποτελεί, πρώτον, το γεγονός ότι το δίκτυο δέχεται εικόνες και όχι βίντεο. Επομένως, δεν υφίσταται η χρονική συσχέτιση των εικόνων στο εσωτερικό του δικτύου. Βέβαια, μπορούν να δοθούν οι επιμέρους εικόνες με τυχαία σειρά αλλά πάλι αυτό δεν αρκεί αφού αναιρεί τη χρονική υπογραφή της ανάπτυξης της καλλιέργειας.

Επιπρόσθετα, η παρουσία σύννεφων δημιουργεί θόρυβο στις εικόνες και γιαυτό χρονικές στιγμές που υπερβαίνουν ένα κατώφλι συννεφιάς, όπως παρουσιάστηκε ο αλγόριθμος Hollstein (Ενότητα 4.3.9), εντοπίζει σε σχετικά καλό βαθμό την παρουσία συννεφιάς. Για τον λόγο αυτόν, στα πειράματα που θα διεξαχθούν θα προστεθούν “κανάλια” που αποτυπώνουν την παρουσία συννεφιάς με τη μορφή δυαδικής μάσκας ανά εικόνα.

Λαμβάνοντας υπ' όψιν τα παραπάνω, μία μέθοδος προσέγγισης και αυτή που χρησιμοποιήθηκε για τη συγκεκριμένη αρχιτεκτονική, περιλαμβάνει τη συνάθροιση των καναλιών “B01, B02, B03, B04, B05, B06, B07, B08, B8A, B09, B11, B12”, σε διαφορετικούς συνδυασμούς και σε ορισμένες περιπτώσεις αυτή του καναλιού της συννεφιάς. Αυτό υποβοηθείται από τη δυνατότητα εισόδου πολλαπλών καναλιών που προσφέρει το Detectron2. Συνεπώς, κάθε δείγμα που εισάγεται στο δίκτυο είναι μία εικόνα με $num_bands \cdot num_intervals$ κανάλια, όπου num_bands πλήθος των καναλιών των επιμέρους εικόνων και $num_intervals$ πλήθος των χρονικών διαστημάτων που πραγματοποιήθηκε η συνάθροιση (λ.χ για συνάθροιση ενός μήνα τότε $num_intervals = 12$).

Σαν επιπλέον πείραμα, που θα δώσει και ένα ικανοποιητικό μέτρο σύγκρισης με το DETR, θα θεωρηθεί $num_intervals = 1$, δηλαδή μία εικόνα τη φορά. Αμελώντας με αυτόν τον τρόπο, τη σημαντικότητα της συσχέτισης στο χρόνο. Το παραπάνω, θα διεξαχθεί για διαφορετικούς συνδυασμούς καναλιών και πάντα με τα ίδια κανάλια που θα χρησιμοποιηθούν ως είσοδο στο DETR.

5.3 Detection Transformer (DETR)

Το DETR [84], όπως παρουσιάστηκε (Ενότητα 3.6), επιστρατεύει Μετασχηματιστές (Ενότητα 3.4) και προτείνεται ως εναλλακτική του Mask R-CNN και συγκρίνεται με αυτό, σε προβλήματα που αφορούν Object Detection και Segmentation. Συνεπώς, ισχύουν τα προβλήματα που αναφέρθηκαν και για το Mask R-CNN (Ενότητα 5.2), δηλαδή δεν υπάρχει υποστήριξη για βίντεο και η συννεφιά αποτελεί ακόμα ένα σοβαρό εμπόδιο.

Η αρχιτεκτονική αυτή παρουσιάζει, όμως δύο μεγάλες αδυναμίες. Αρχικά, δεν προτείνεται για εικόνες, οι οποίες ενδέχεται να περιέχουν περισσότερα από από 100 διαφορετικά αντικείμενα. Αναφορικά, στο τελικό Σύνολο Δεδομένων υπάρχουν περιπτώσεις με patches που υπερβαίνουν τα 1000 αγροτεμάχια.

Το δεύτερο πρόβλημα εμφανίζεται κατά τη είσοδο των δεδομένων στο δίκτυο. Το DETR, στην υλοποιημένη του μορφή, δέχεται μόνο 3 κανάλια εισόδου, δημιουργώντας με αυτόν τον τρόπο, ένα μέγιστο όριο στην πληροφορία που μπορεί να εισαχθεί στο δίκτυο. Το πλήθος των καναλιών, μπορεί να αυξηθεί, τροποποιώντας το backbone δίκτυο της όλης αρχιτεκτονικής. Ωστόσο, κάτι τέτοιο απαιτεί την ολική εκπαίδευση του δικτύου (μαζί με το backbone), αυξάνοντας δραματικά την πολυπλοκότητα και το κόστος εκπαίδευσης, γεγονός που δυσκολεύει τη διεξαγωγή πολλαπλών πειραμάτων (όπως με το Mask R-CNN).

Η εκπαίδευση του DETR γίνεται σε δύο στάδια, κατά το πρώτο, το μοντέλο εκπαιδεύεται να λύσει το πρόβλημα σε Object Detection. Εφόσον, φτάσει σε επιθυμητά επίπεδα, τα βάρη του δικτύου “κλειδώνουν” και η κεφαλή ταξινόμησης αρχίζει να εκπαιδεύεται πάνω στην τελική ταξινόμηση του κάθε εικονοστοιχείου, δηλαδή σαν τελική έξοδος δίνεται μία μάσκα με τη γεωμετρία του αντικειμένου προς αναγνώριση.

Τέλος, οι Μετασχηματιστές είναι διαβόητοι για το πλήθος των εποχών που απαιτούν κατά την εκπαίδευση προκειμένου να συγκλίνουν. Χαρακτηριστικά στη δημοσίευση του DETR [84], αναφέρεται πως αυτό έχει εκπαιδευτεί για 300 εποχές. Αυτό απαιτεί αρκετά μεγάλο εμπόδιο κατά τον πειραματισμό με περισσότερες διαφοροποιήσεις του DETR.

5.4 DuPLO

Μία αρχιτεκτονική που συνδυάζει το χωροχρόνο αποτελεί το DuPLO [97], το οποίο όπως αναφέρθηκε στην Ενότητα 3.7, αντιμετωπίζει το πρόβλημα ταξινόμησης ενσωματώνοντας δύο διαφορετικά τμήματα. Το ένα τμήμα μαθαίνει καλύτερα το χώρο και το άλλο το χρόνο. Η τελική ταξινόμηση προκύπτει από τη συνένωση των χαρακτηριστικών που έχουν εξαχθεί από το τελικό δίκτυο.

Στην εν λόγω δημοσίευση, παρατίθεται ότι η εκπαίδευση του δικτύου διαρκεί 300 εποχές, με batch size 128. Βέβαια, σαν ένα batch θεωρείται ένα μικρό κομμάτι (5×5) της εικόνας εισόδου. Επομένως, κάθε patch που έχει διαστάσεις (366×366) πρέπει να γίνει pad στον κοντινότερο διαιρέτη του 5, δηλαδή 370×370 , με μηδενικά (0). Χρησιμοποιείται το μηδέν, διότι με βάση τον Sentinel 2, εικονοστοιχεία με τιμή μηδέν, μεταφράζονται ως κενά. Εφόσον παρέλθει αυτή η διαδικασία, το κάθε patch μεταφράζεται ως $370^2/5^2 = 5476$ δείγματα.

Ωστόσο, δεν έχει αντιμετωπιστεί ακόμα το χρονικό κομμάτι. Αν γίνει η υπόθεση, ότι το patch συνθροίζεται με συχνότητα ανά μήνα, τότε κάθε patch περιέχει 12 εικόνες των 366×366 , που γίνονται

370 × 370. Το DuPLO, στο χρονικό κομμάτι απαιτεί $5 \times T$, όπου T χρονικό παράθυρο. Εδώ μπορούν να γίνουν αρκετοί πειραματισμοί για το μέγεθος του παραθύρου, ωστόσο η τιμή που θα επιλεγεί, θα διατηρηθεί για το τελικό μοντέλο των προβλέψεων. Με βάση αυτό, η επιλογή μεγάλων τιμών (λ.χ 12) δεν είναι ρεαλιστική επιλογή, αφού για μελλοντικές προβλέψεις το μοντέλο θα απαιτεί παρατηρήσεις του ίδιου σημείου για 12 μήνες. Έτσι, η τελική τιμή T που επιλέχθηκε είναι οι 6 μήνες, αφού μικρότερες τιμές ανεβάζουν σημαντικά το πλήθος των δειγμάτων.

Συνεπώς, για $T = 6$, το κάθε patch αντιστοιχεί σε $5476 \cdot (12/6) = 10952$ δείγματα και για 1800 patches (2 tiles), ο τελικός αριθμός των δειγμάτων είναι 19.713.600 ή για batch size 128, προκύπτουν 154.012,5 batches.

Η τελική είσοδος του μοντέλου αποτελείται από εικόνα μεγέθους 5×5 στα κανάλια “B02, B03, B04, B08” και το NVDI (5 κανάλια) για $T = 6$ συνεχόμενες χρονικές στιγμές, άρα $5T \times 5 \times 5$.

5.5 Πειράματα

Κατά τη διάρκεια των πειραμάτων, παρατηρήθηκε ότι υπάρχουν κενοί μήνες, δηλαδή μήνες χωρίς δεδομένα από τους δορυφόρους. Η δυσκολία αυτή αντιμετωπίστηκε με γραμμική παρεμβολή, όπως περιγράφεται στην Ενότητα 4.3.11.

Ακόμα, στην πραγματικότητα, λόγω των ελλিপών δεδομένων είναι αρκετά δύσκολο να εκπαιδευτεί αρχιτεκτονική που να λύνει το πρόβλημα ταξινόμησης σε όλες τις κατηγορίες που εμπεριέχει η βάση. Έχοντας ως γνώμονα το παραπάνω, το πρόβλημα ανάγεται σε πρόβλημα δυαδικής ταξινόμησης, δηλαδή με βάση εικόνας εισόδου $H \times W$ απαιτείται πρόβλεψη ίσης διάστασης ($H \times W$) με αυτή της εισόδου, που να περιγράφει ποια εικονοστοιχεία περιέχουν αγροτεμάχια και ποια όχι.

Τα μοντέλα δέχονται σαν είσοδο τα ίδια αρχεία coco (train, val) και αξιολογούνται με τις ίδιες μετρικές πάνω στο test set, συγκεκριμένα ανακατασκευάζοντας μία τελική μάσκα από τις επιμέρους, μπορεί να γίνει σύγκριση της με τη μάσκα των πραγματικών τιμών σε επίπεδο εικονοστοιχείου και να εξαχθούν οι εν λόγω μετρικές. Για κάθε μοντέλο, θα υπολογιστεί accuracy και macro average των precision και recall.

Τέλος, προκειμένου να λειτουργήσουν οι κανονικοποιήσεις των εισόδων και άρα τα μοντέλα, πρέπει να προσκομισθούν οι μέσοι όροι και οι τυπικές αποκλίσεις των καναλιών που θα χρησιμοποιηθούν για την εκπαίδευση του δικτύου. Αυτό είναι αναγκαίο, διότι ο τομέας των δορυφορικών εικόνων διαφέρει αρκετά από τις τυπικές εικόνες της καθημερινότητας. Αφενός βρίσκονται στην πλειοψηφία τους στο εύρος $[0 - 255]$, αφετέρου αντικατοπτρίζουν διαφορετικά αντικείμενα (λ.χ γάτες, αυτοκίνητα, κ.α).

Σχετικά με το Mask-RCNN, διεξήχθησαν αρκετά πειράματα με τις αρχιτεκτονικές. Αναλυτικά, χρησιμοποιήθηκαν δύο διαφορετικά μοντέλα, το **R50-FPN**, το οποίο χρησιμοποιεί ResNet50 για την εξαγωγή του χάρτη χαρακτηριστικών και στη συνέχεια δίκτυο FPN για την τελική εξαγωγή των μασκών που μπορεί να εμπεριέχουν τα αντικείμενα. Το δεύτερο δίκτυο, **R101-FPN**, διαφέρει από το πρώτο αφού ενσωματώνει το ResNet101 για την εξαγωγή του χάρτη χαρακτηριστικών.

Προσφέρεται ένα πλήθος εκπαιδευμένων αρχιτεκτονικών ως προς τους συνδυασμούς, π.χ R50-C4 (Faster R-CNN) [89] και R50-DC5 (Dilated-C5) [136]. Ωστόσο, τέτοια μοντέλα απαιτούν αρκετούς υπολογιστικούς πόρους και η επίδοση δε δικαιολογεί την υιοθέτησή τους. Σε ευρύτερο πλαίσιο, τα μοντέλα που ενσωματώνουν δίκτυο FPN ισορροπούν τη σχέση κόστους απόδοσης.

Όσον αφορά το DETR, όπως αναλύθηκε τα κανάλια που υποστηρίζει είναι μέχρι τρία. Προκειμένου να εισαχθεί περισσότερη πληροφορία μέσα στο δίκτυο, πρέπει να γίνει τροποποίηση του δικτύου εξαγωγής χάρτη χαρακτηριστικών (δηλαδή το backbone), το οποίο πρέπει να “ξε-παγώσει” τα βάρη του και να εκπαιδευτεί από την αρχή. Ανεβάζοντας αρκετά την πολυπλοκότητα του μοντέλου, αφού στην απλούστερη του μορφή απαιτούνται δύο στάδια εκπαίδευσης για να χρησιμοποιηθεί ως segmentation αρχιτεκτονική.

Το DETR παρέχεται από τους συγγραφείς και σε Detectron2 engine, επομένως θα χρησιμοποιηθεί αυτή για να υπάρχει κοινός κώδικας που φορτώνει τα δεδομένα στα μοντέλα. Προς στιγμή, υποστηρίζεται μόνο σαν backbone το δίκτυο R50, άρα θα συγκριθεί με Mask RCNN που εφαρμόζει το ίδιο backbone στα ίδια κανάλια.

Σε όλες τις περιπτώσεις, τα μοντέλα αξιολογούνται στο ίδιο σύνολο δεδομένων, το οποίο διαφέρει από αυτό της εκπαίδευσης ως προς την τοποθεσία. Συγκεκριμένα, τα μοντέλα αξιολογούνται σε περιοχές

που δεν έχουν ξαναδεί (διαφορετικά patches), έχοντας ως στόχο την καλύτερη γενίκευση του μοντέλου. Με βάση το παραπάνω, το μοντέλο αναγκάζεται να μάθει τις συμπεριφορές των διαφορετικών καλλιεργειών, έτσι ώστε να χρησιμοποιηθεί σε γειτονικές περιοχές. Τονίζεται, ότι αυτό ίσως δεν είναι εφικτό για τοποθεσίες που διαφέρουν σημαντικά σε κλίμα, τύπους καλλιεργειών, κ.α. Το τελικό μοντέλο, χρησιμοποιείται σε περιοχές που δεν έχει δει αλλά βρίσκονται χωρικά κοντά σε αυτές που εκπαιδεύτηκε.

Για τα μοντέλα που δέχονται σαν είσοδο μία εικόνα, π.χ έναν μήνα και όχι τα συναθροισμένα κανάλια πολλών μηνών, ως μήνας αξιολόγησης επιλέχθηκε ο Ιούνιος. Ο Ιούνιος αποτελεί έναν μήνα που τα περισσότερα φυτά δείχνουν το τελικό στάδιο ανάπτυξης τους και άρα μπορούν, τουλάχιστον από άνθρωπο, να ταξινομηθούν με μεγαλύτερη ακρίβεια.

5.6 Αποτελέσματα

Πίνακας 5.1: Αποτελέσματα Mask R-CNN για LULC (§: macro average).

model	interval	channels	precision§	recall§	accuracy
R50-FPN + Cloud Masks	5	8	0.61	0.62	0.69
R50-FPN No Masks	5	7	0.63	0.64	0.72
R50-FPN No Masks	10	13	0.60	0.61	0.69
R101-FPN + Cloud Masks	10	8	0.63	0.65	0.71
R101-FPN + Backbone Retrain	5	8	0.57	0.60	0.70
R101-FPN No Masks	3	7	0.62	0.63	0.72
R101-FPN No Masks	5	7	0.65	0.65	0.75
R101-FPN No Masks	10	7	0.66	0.67	0.75
R101-FPN No Masks	5	13	0.64	0.64	0.75
R101-FPN No Masks	10	13	0.62	0.62	0.73

Ο Πίνακας 5.1 παρουσιάζει τα αποτελέσματα για διάφορες μορφές του MASK R-CNN. Όπως αναφέρεται και στη βιβλιογραφία του μοντέλου [135], το R101 όντως παράγει καλύτερες μετρικές σε σχέση με το R50. Ακόμα, η χρήση των μασκών συννεφιάς ως επιπλέον κανάλι στα δεδομένα δε φαίνεται να βοηθάει το μοντέλο, αλλά να το μπερδεύει περισσότερο. Τονίζεται επίσης, ότι η παραγωγή μασκών απαιτεί επιπλέον χρόνο κατά την εκπαίδευση, συνεισφέροντας στην μεγάλη πολυπλοκότητα.

Η χρήση μικρών διαστημάτων παρατήρησης, δηλαδή πόσους συνεχόμενους μήνες βλέπει το μοντέλο σε κάθε δείγμα, φαίνεται να βελτιώνει την ικανότητα τους στις προβλέψεις. Με ιδανικό μέγεθος παραθύρου να αποτελούν οι δέκα μήνες. Η αύξηση των καναλιών, δε φαίνεται να προσφέρει διορατικότητα στο μοντέλο. Τέλος το ξεκλείδωμα και άρα η επανεκπαίδευση των βαρών του Backbone δικτύου, ενώ ανεβάζει σημαντικά το χρόνο εκπαίδευσης, δε συνεπάγεται με καλύτερες προβλέψεις.

Όσον αφορά τις αρχιτεκτονικές DuPLO και DETR, ο χρόνος που απαιτείται για την περάτωση της εκπαίδευσης είναι αρκετά μεγάλος, τουλάχιστον με βάση τους διαθέσιμους υπολογιστικούς πόρους. Ακόμα, όπως αναφέρθηκε, στην υλοποίηση του DETR δεν παρέχεται backbone δίκτυο R101, όπως στις υλοποιήσεις του Mask R-CNN. Επομένως, ο μόνος τρόπος σύγκρισης είναι η εκπαίδευση του Mask R-CNN στα ίδια κανάλια με το ίδιο backbone δίκτυο.

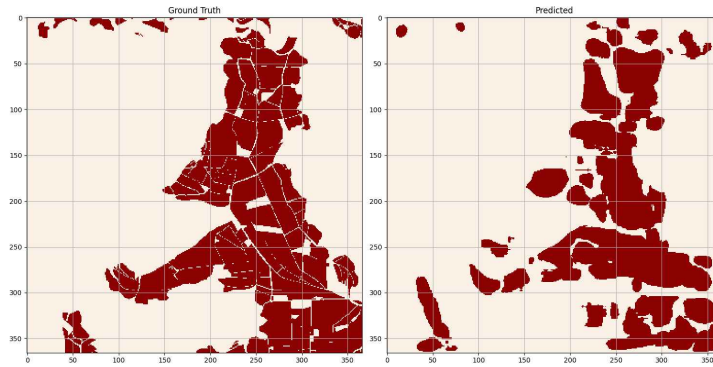
Το DuPLO δεν κάνει χρήση κάποιου backbone δικτύου και άρα τα αποτελέσματα του συγκρίνονται απευθείας με αυτά των άλλων δύο αρχιτεκτονικών.

Στο Σχήμα 5.1 παρουσιάζονται οι προβλέψεις του Mask R-CNN για διάφορα patches. Συγκεκριμένα για τα patches (2020 31TDG patch 14 08), (2020 31TCG patch 19 12), (2020 31TCG patch 19 29), (2020 31TDG patch 20 17), δίνεται αριστερά η πραγματική εικόνα αναφοράς (labels) και δεξιά η πρόβλεψη του μοντέλου. Το μοντέλο, προβλέπει πολλαπλές δυαδικές μάσκες, στο μέγεθος της εικόνας εισόδου δηλαδή

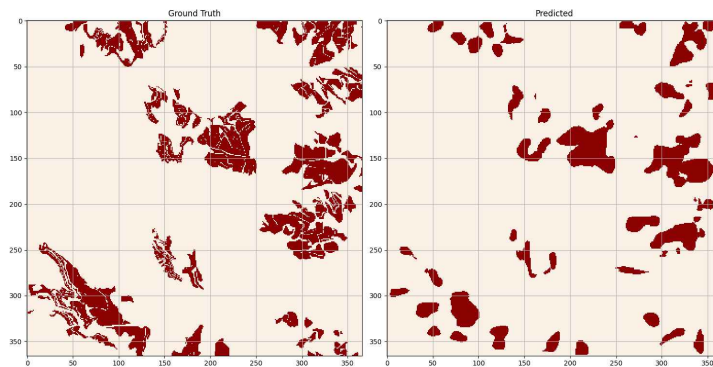
Πίνακας 5.2: Αποτελέσματα DETR και Mask-RCNN για LULC (§: macro average).

model	bands	precision§	recall§	accuracy
DETR R50	B02, B03, B04	0.50	0.53	0.63
Mask-RCNN R50	B02, B03, B04	0.52	0.56	0.65
DuPLO	B02, B03, B04, B08, NVDI	0.74	0.76	0.81

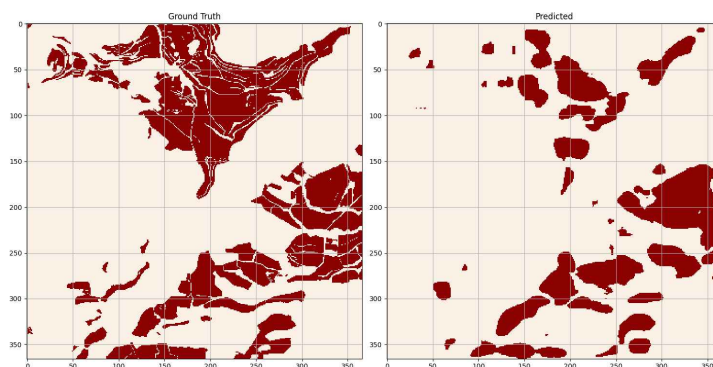
366 × 366 μία για κάθε πρόβλεψη. Η τελική μάσκα που παρουσιάζεται (δεξιά) αποτελείται από το λογικό η (οr) όλων των μασκών.



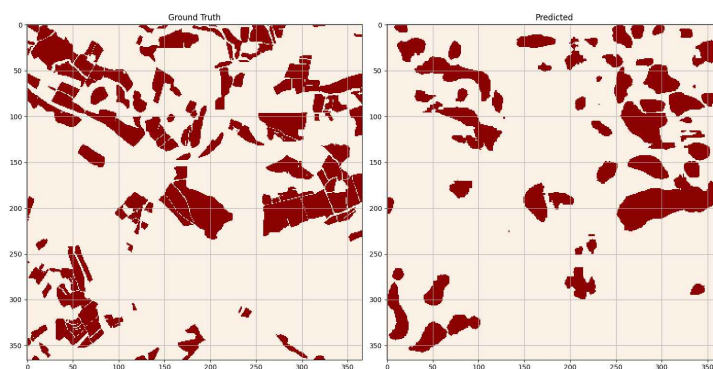
(i) 2020_31TDG_patch_14_08.



(ii) 2020_31TCG_patch_19_12.



(iii) 2020_31TCG_patch_19_29.



(iv) 2020_31TDG_patch_20_17.

Σχήμα 5.1: Προβλέψεις του Mask R-CNN με backbone R101.

Κεφάλαιο 6

Συμπεράσματα

6.1 Παρατηρήσεις

Σημαντικότερη παρατήρηση, αποτελεί ο χρόνος εκπαίδευσης του DuPLO και του DETR πάνω στο κοινό σύνολο δεδομένων. Όλες οι παραλλαγές του Mask R-CNN, εκπαιδεύονται εντός ωρών. Ωστόσο, το DuPLO για λιγότερες (5) από τις προτεινόμενες εποχές (300), με διπλάσιο batch size (256) από το προτεινόμενο (128), ολοκληρώνει την εκπαίδευση του σε διάστημα πέντε ημερών. Το DETR για 500 εποχές, από τις 300 που προτείνονται από τους συγγραφείς, με backbone δίκτυο R50 και 3 κανάλια, απαιτεί περίπου δύο ολόκληρες μέρες για το πρώτο σκέλος της εκπαίδευσης. Το δεύτερο σκέλος ολοκληρώνεται μετά από μισή μέρα. Το γεγονός αυτό, δε βοηθάει στον περαιτέρω πειραματισμό με τις εν λόγω αρχιτεκτονικές. Ακόμα, το DuPLO κάνει τη βασική υπόθεση ότι για κάθε grid 5×5 , το περιγράφει μόνο το κεντρικό εικονοστοιχείο, αυτό προφανώς δεν ισχύει κατά μήκος του Συνόλου Δεδομένων.

Ο Πίνακας 5.2 παρουσιάζει τα αποτελέσματα με τις διάφορες αρχιτεκτονικές. Όπως φαίνεται, το DuPLO πετυχαίνει τις καλύτερες μετρικές, ωστόσο ο χρόνος εκπαίδευσης, πρόβλεψης και η υπόθεση που κάνει για την είσοδο 5×5 , δεν το καθιστούν την ιδανική επιλογή.

Το DETR φαίνεται να συμβαδίζει με το Mask R-CNN, για ίδιο backbone. Όμως το Mask R-CNN υπερτερεί όταν εφαρμόζεται σε αυτό το backbone R101. Επιπρόσθετα, το DETR, όπως προτείνεται, έχει περιορισμό στα κανάλια εισόδου. Τα δορυφορικά δεδομένα, είναι πολυφασματικά και το μέλλον της Τηλεπισκόπησης περιλαμβάνει υπερφασματικές εικόνες (δηλαδή πολλά περισσότερα κανάλια εισόδου). Η αποτυχία εισαγωγής των πολλαπλών καναλιών στο δίκτυο, σημαίνει έλλειψη πληροφορίας στο εσωτερικό του δικτύου και ενδεχομένως να οδηγήσει σε χειρότερες συσχετίσεις. Το DETR, όπως και κάθε άλλο δίκτυο που περιλαμβάνει αρχιτεκτονικές Μετασχηματιστών, το πλήθος των εποχών που απαιτούνται για τη σύγκλιση είναι μεγάλο, το οποίο μεταφράζεται ως μεγαλύτερος χρόνος εκπαίδευσης.

Ακόμα, ο Πίνακας 5.1 μπορεί να μεταφραστεί ως ένα πείραμα που αποτυπώνει τη σημαντικότητα του χρόνου κατά την ταξινόμηση καλλιιεργειών. Παρουσιάζεται ξεκάθαρα ότι όσο αυξάνεται το παράθυρο παρατήρησης, τόσο το μοντέλο βελτιώνεται στις προβλέψεις του. Εδώ βέβαια, οι επιπλέον χρονικές στιγμές προστίθενται στο μοντέλο ως επιπλέον κανάλια, ενδεχομένως η χρήση μίας συνθετότερης αρχιτεκτονικής που συσχετίζεται τα δεδομένα σε επίπεδο χρόνου να βελτιώσει δραματικά τα αποτελέσματα του μοντέλου.

Όπως απεικονίζεται στο Σχήμα 5.1, η αρχιτεκτονική Mask R-CNN φαίνεται να μαθαίνει την κατανομή των αγροτεμαχίων. Οι προβλέψεις που δίνει βρίσκονται αρκετά κοντά στις πραγματικές, βέβαια παρατηρώντας καλύτερα το σχήμα και συγκεκριμένα τη δεξιά εικόνα κάθε φορά, οι προτάσεις εμφανίζουν μεγαλύτερη άμβλυνση στο περίβλημά τους, σε σχέση με το πιο “απότομο” και αιχμηρό της πραγματικής εικόνας.

Επιπλέον, παρατηρείται ότι οι προβλέψεις δεν παρουσιάζουν επιμέρους αγροτεμάχια, όπως αυτά διαχωρίζονται (λ.χ δρόμων). Οι γραμμές αυτές διακρίνονται εύκολα με το ανθρώπινο μάτι, σε όλο το εύρος τους. Όμως, το μοντέλο φαίνεται να μπερδεύεται αρκετά στη διατήρηση τέτοιων συνθηκών. Οφείλεται όμως, να ληφθεί υπ’ όψιν ότι το Mask R-CNN προβλέπει ολόκληρα αντικείμενα μέσα σε μία εικόνα εισόδου και εφόσον το πρόβλημα είναι δυαδικής φύσης, ενδεχομένως να δημιουργείται περαιτέρω σύγχυση στο εσωτερικό της αρχιτεκτονικής.

Ενδέχεται, σε άλλες κατηγορίες προβλημάτων, όπου τα γειτονικά αγροτεμάχια ανήκουν σε διαφορε-

τικές κλάσεις (λ.χ μήλα και αχλάδια) το μοντέλο να επιτυγχάνει καλύτερο διαχωρισμό των πλαισίων του εκάστοτε αγροτεμαχίου.

Τέλος, η εισαγωγή της μάσκας συννεφιάς ως επιπλέον κανάλι στο δίκτυο δε φαίνεται να βελτιώνει τις μετρικές, αντιθέτως ενδέχεται να προκαλεί σύγχυση στο μοντέλο.

Με βάση όλα τα παραπάνω, το καταλληλότερο μοντέλο και αυτό που προτείνεται είναι το Mask R-CNN με backbone R101 και 70 (10 · 7) κανάλια εισόδου.

6.2 Μελλοντικές Επεκτάσεις

Η παρούσα εργασία μπορεί να αποκτήσει αρκετές μελλοντικές πορείες, κυρίως λόγω της δυναμικότητας. Οι επεκτάσεις αυτές θα χωριστούν με γνώμονα δύο κατηγορίες, περαιτέρω επεκτάσεις ή και τροποποιήσεις που έχουν ως στόχο το Σύνολο Δεδομένων (Ενότητα 6.2.1) καθώς και πιθανές εφαρμογές ή και προβλήματα που μπορούν να λυθούν με Αρχιτεκτονικές Βαθιάς Μάθησης (Ενότητα 6.2.2).

6.2.1 Σύνολο Δεδομένων

Ίσως η ευκολότερη επέκταση που μπορεί να λάβει χώρα είναι η δημιουργία patches για tiles σε άλλες χώρες. Η διαδικασία γίνεται σε δύο βήματα, πρέπει να συλλεχθούν τα δεδομένα Sentinel-2 από την υπηρεσία Copernicus για την περιοχή ενδιαφέροντος. Στη συνέχεια, απαιτείται μία Βάση Δεδομένων, η οποία περιέχει στήλη με γεωμετρίες. Η κάθε γεωμετρία πρέπει να ανταποκρίνεται σε ένα μόνο αγροτεμάχιο της περιοχής μελέτης. Οι διάφορες κατηγορίες των καλλιεργειών πρέπει επίσης να μετατραπούν από συμβολοσειρές σε ακέραιους αριθμούς με βάση την κωδικοποίηση του FAO (Ενότητα 4.3.3).

Επιπρόσθετα, η συνένωση δορυφορικών δεδομένων από το πρόγραμμα Sentinel-1 (Ενότητα 2.4.1), δηλαδή δεδομένων ραντάρ, είναι κάτι εφικτό και αρκετά ενδιαφέρον. Τα ραντάρ παρέχουν ποιοτικότερα δεδομένα, αφού δεν επηρεάζονται από τις καιρικές συνθήκες όπως η συννεφιά, που δημιουργεί τεράστιο πρόβλημα στα δορυφορικά δεδομένα με εικόνες. Βέβαια, η συλλογή και η επεξεργασία εμπεριέχουν μεγαλύτερο βαθμό δυσκολίας ως προς την κανονικοποίηση στο υπάρχον Σύνολο Δεδομένων.

Ένα συχνό φαινόμενο, το οποίο δεν προβλέπεται από τη δημιουργία του Συνόλου Δεδομένων που παρουσιάστηκε, αποτελεί η ύπαρξη της συγκαλλιέργειας. Αγροτεμάχια στα οποία καλλιεργούνται δύο ή περισσότερες καλλιέργειες κατά τη διάρκεια της χρονιάς συγκροτούν την τεχνική της συγκαλλιέργειας. Η αντιμετώπιση τέτοιων περιπτώσεων, απαιτεί η εικόνα των κατηγοριών (labels) να αλλάζει την αντίστοιχη χρονική στιγμή που αλλάζει ή που υπερτερεί ο άλλος τύπος της καλλιέργειας, συνεπώς το τελικό αρχείο labels πρέπει να είναι χρονοσειρά και όχι μία στατική εικόνα για όλη τη χρονιά. Κάτι τέτοιο είναι εφικτό και υλοποιήσιμο, εφόσον υπάρχουν οι σχετικές Βάσεις Δεδομένων. Το παραπάνω λύνει και το πρόβλημα των εναλλασσόμενων, όταν αυτές πραγματοποιούνται στη μέση του χρόνου. Αν οι καλλιέργειες, έχουν αλλάξει πριν τη δημιουργία της Βάσης Αναφοράς, τότε οι εναλλασσόμενες καλλιέργειες δεν αποτελούν πρόβλημα.

6.2.2 Αρχιτεκτονικές Βαθιάς Μάθησης

Στις αρχιτεκτονικές βαθιάς μάθησης υπάρχουν αρκετές υλοποιήσεις και προτάσεις καθώς και πολλά προβλήματα που μπορούν να τεθούν προς αντιμετώπιση.

Αρχικά, η παρούσα εργασία εξετάζει μερικές από τις διαθέσιμες αρχιτεκτονικές, με κύρια έμφαση σε μοντέλα που αντιμετωπίζουν προβλήματα Όρασης Υπολογιστών. Υπάρχει ένα τεράστιο πλήθος υλοποιήσεων που δεν έχουν εφαρμοστεί πάνω στο Σύνολο Δεδομένων, όπως το Raster Vision [137]. Ωστόσο, δεν είναι απαραίτητο να χρησιμοποιηθούν έτοιμα μοντέλα, ο όλος σκοπός της δημιουργίας ενός ‘κανονικοποιημένου’ Συνόλου Δεδομένων, είναι να ελαχιστοποιήσει τη δουλειά που απαιτείται για τη συγκρότηση δεδομένων και να επιτρέψει τη σύνθεση νέων πολυπλοκότερων αρχιτεκτονικών για το εκάστοτε πρόβλημα.

Ένα ακόμα πεδίο εφαρμογής αποτελεί η ανίχνευση συννεφιάς με μοντέλα ΠΑΔ (GAN; Ενότητα 3.3), εδώ υπάρχει αρκετή μελέτη όπως το Cloud-GAN [138] και το STGAN [139] και θα μπορούσε να χρησιμοποιηθεί ως εναλλακτική στην εξαγωγή μασκών συννεφιάς που προτείνεται στην παρούσα εργασία με τον αλγόριθμο hollstein (Ενότητα 4.3.9).

Αρκετό ενδιαφέρον, με πληθώρα δημοσιεύσεων παρουσιάζει το Super Resolution (SR). Σε αυτή την κατηγορία προβλημάτων, επιστρατεύονται ΠΑΔ [140, 141] ή ΣΝΔ [142] με σκοπό να βελτιώσουν την ανάλυση της εικόνας. Τα δορυφορικά δεδομένα δημιουργούν αρκετές δυσμενείς καταστάσεις διότι στις περισσότερες περιπτώσεις, ο λόγος του μεγέθους του αντικειμένου μελέτης προς το συνολικό μέγεθος της εικόνας είναι αρκετά μικρό. Ενδεχομένως, η εφαρμογή ΠΑΔ ή ΣΝΔ να βελτιώσει τα αποτελέσματα τις προαναφερόμενες καταστάσεις, όπου δηλαδή η ανάλυση της δορυφορικής εικόνας είναι περιορισμένη.

Στην τηλεπισκόπηση, ανωμαλία θεωρείται όταν διαστρεβλώνεται η φασματική υπογραφή των αντικειμένων [143]. Ο Εντοπισμός Ανωμαλιών (Anomaly Detection), προσπαθεί να διορθώσει το παραπάνω με βάση χαρακτηριστικά που έχουν προκύψει από παρελθοντική εμπειρία. Δεν υπάρχει αρκετή έρευνα σε αυτόν τον τομέα εκτός του [144], που χρησιμοποιεί ΣΝΔ για τη διόρθωση του εικονοστοιχείου ελέγχοντας τα γειτονικά εικονοστοιχεία [143].

Τέλος, άμεση απόρροια της αύξησης των διαθέσιμων δεδομένων, αποτελεί και αύξηση του όγκου που αυτά καταλαμβάνουν, πόσο μάλλον για πολυφασματικές δορυφορικές εικόνες υψηλής ανάλυσης. Απαιτείται λοιπόν, η υλοποίηση κάποιας μεθόδου συμπίεσης της πληροφορίας, προκειμένου η τελευταία να διαχειρίζεται καλύτερα καθώς και να απαιτεί λιγότερους υπολογιστικούς πόρους κατά την επεξεργασία από τα μοντέλα. Στη βιβλιογραφία [3], η συμπίεση χρησιμοποιείται ως επί το πλείστον με τη μορφή αρχιτεκτονικών Autoencoders, για την εξαγωγή χαρακτηριστικών ή για τη μείωση θορύβου.

Βιβλιογραφία

- [1] Zhe Jiang and Shashi Shekhar. *Spatial Big Data Science*. 07 2017.
- [2] Nataliia Kussul, Mykola Lavreniuk, Sergii Skakun, and Andrey Shelestov. Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geoscience and Remote Sensing Letters*, PP:1–5, 03 2017.
- [3] Lei Ma, Yu Liu, Xueliang Zhang, Yuanxin Ye, Gaofer Yin, and Brian Alan Johnson. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 152:166–177, 2019.
- [4] Jonathan Chipman Thomas Lillesand, Ralph W. Kiefer. *Remote Sensing and Image Interpretation, 7th Edition*. John Wiley & Sons, Inc., 2015.
- [5] Caroline Johnson. Bird’s eye viewfinder: 160 years of aerial photography, Sat, July 28, 2018. <https://airandspace.si.edu/stories/editorial/birds-eye-viewfinder-160-years-aerial-photography>, Τελευταία Πρόσβαση 2021-01-16.
- [6] Robert A. Schowengerdt. Chapter 1 - the nature of remote sensing. In Robert A. Schowengerdt, editor, *Remote Sensing (Third Edition)*, pages 1 – X. Academic Press, Burlington, third edition edition, 2007.
- [7] Ni-Bin Chang and Kaixu Bai. *Multisensor Data Fusion and Machine Learning for Environmental Remote Sensing*. CRC Press, Inc., USA, 1st edition, 2018.
- [8] Shu-Kun Lin. Introduction to remote sensing. fifth edition. by james b. campbell and randolph h. wynne, the guilfordpress, 2011; 662 pages. isbn 978-1-60918-176-5. *Remote Sensing*, 5(1):282–283, Jan 2013.
- [9] Εθνικό Αστεροσκοπείο Αθηνών Ινστιτούτο Αστρονομίας & Αστροφυσικής. Ηλιακό Σύστημα - Ο Ήλιος. https://www.astro.noa.gr/journal/Basic/journal_solhelios.html, Τελευταία Πρόσβαση 2020-11-09.
- [10] Lou Gevaux. *3D-hyperspectral imaging and optical analysis of skin for the human face*. PhD thesis, 10 2019.
- [11] Xiaoyong Zhuge, X. Zou, and Yuan Wang. A fast cloud detection algorithm applicable to monitoring and nowcasting of daytime cloud systems. *IEEE Transactions on Geoscience and Remote Sensing*, 55:1–9, 11 2017.
- [12] Janusz Mikołajczyk, Zbigniew Bielecki, Maciej Bugajski, Jozef Piotrowski, Jacek Wojtas, Waldemar Gawron, Dariusz Szabra, and Artur Prokopiuk. Analysis of free-space optics development. *Metrology and Measurement Systems*, 24, 12 2017.
- [13] Documentation QGIS 2.8. A gentle introduction to gis, coordinate reference systems. https://docs.qgis.org/2.8/en/docs/gentle_gis_introduction/coordinate_reference_systems.html, Τελευταία Πρόσβαση 2020-11-14.
- [14] D.H. MALING. Chapter 16 - the transverse mercator projection. In D.H. MALING, editor, *Coordinate Systems and Map Projections (Second Edition)*, pages 336 – 363. Pergamon, Amsterdam, second edition edition, 1992.
- [15] Copernicus: Europe’s eyes on Earth. Sentinel-2 multispectral instrument (msi) product types. <https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi/product-types>, Τελευταία Πρόσβαση 2020-10-30.
- [16] ESA EduspaceGR. Αισθητήρες. https://www.esa.int/SPECIALS/Eduspace_GR/SEMBFR4PVFG_0.html, Τελευταία Πρόσβαση 2020-11-16.
- [17] Copernicus: Europe’s eyes on earth. why copernicus? <https://www.copernicus.eu/en/about-copernicus/why-copernicus>, Τελευταία Πρόσβαση 2020-10-13.
- [18] Global monitoring for environment and security (gmes). <https://www.eea.europa.eu/themes/climate/links/physical-science-on-climate/global-monitoring-for-environment-and-security-gmes>, Τελευταία Πρόσβαση 2020-10-13.
- [19] Copernicus: Europe’s eyes on Earth. Sentinel-1 overview. <https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-1/overview>, Τελευταία Πρόσβαση 2020-10-18.
- [20] The european space agency. <https://www.esa.int/>, Τελευταία Πρόσβαση 2020-10-13.

-
- [21] Copernicus: Europe's eyes on earth. what is copernicus? <https://web.archive.org/web/20181103182626/http://www.copernicus.eu/main/overview>, Τελευταία Πρόσβαση 2020-10-13.
- [22] Copernicus: Europe's eyes on Earth. Discover our satellites. <https://www.copernicus.eu/en/about-copernicus/infrastructure/discover-our-satellites>, Τελευταία Πρόσβαση 2020-10-19.
- [23] Martin Kirscht and Carsten Rinke. 3d reconstruction of buildings and vegetation from synthetic aperture radar (sar) images. In *In Proceedings of LAPR Workshop on Machine Vision Applications*, pages 17–19, 1998.
- [24] Copernicus: Europe's eyes on Earth. Sentinel-1: Description (Εικόνα). Τελευταία Πρόσβαση 2020-10-18.
- [25] Copernicus: Europe's eyes on Earth. Sentinel-1 instrument payload. <https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-1/instrument-payload>, Τελευταία Πρόσβαση 2020-10-18.
- [26] Copernicus: Europe's eyes on Earth. Sentinel-1 instrument payload: Stripmap (sm). <https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-1-sar/acquisition-modes/stripmap>, Τελευταία Πρόσβαση 2020-10-18.
- [27] F. De Zan and A. Monti Guarnieri. Topsar: Terrain observation by progressive scans. *IEEE Transactions on Geoscience and Remote Sensing*, 44(9):2352–2360, 2006.
- [28] Copernicus: Europe's eyes on Earth. Sentinel-1 instrument payload: Interferometric wide swath (iw). <https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-1-sar/acquisition-modes/interferometric-wide-swath>, Τελευταία Πρόσβαση 2020-10-19.
- [29] Copernicus: Europe's eyes on Earth. Sentinel-1 instrument payload: Extra wide swath (ew). <https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-1-sar/acquisition-modes/extra-wide-swath>, Τελευταία Πρόσβαση 2020-10-22.
- [30] Copernicus: Europe's eyes on Earth. Sentinel-1 instrument payload: Wave (wv). <https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-1-sar/acquisition-modes/wave>, Τελευταία Πρόσβαση 2020-10-22.
- [31] Copernicus: Europe's eyes on Earth. Sentinel-1 data products. <https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-1/data-products>, Τελευταία Πρόσβαση 2020-10-22.
- [32] P. Snoeij, E. Attema, A. M. Guarnieri, and F. Rocca. Fdmaq a novel encoding scheme for sentinel-1. In *2009 IEEE International Geoscience and Remote Sensing Symposium*, volume 1, pages 1–44–1–47, 2009.
- [33] Copernicus: Europe's eyes on Earth. Sentinel-1: Data products (level0). <https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-1-sar/product-types-processing-levels/level-0>, Τελευταία Πρόσβαση 2020-10-22.
- [34] Copernicus: Europe's eyes on Earth. Sentinel-1: Data products (level1). <https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-1-sar/product-types-processing-levels/level-1>, Τελευταία Πρόσβαση 2020-10-23.
- [35] Copernicus: Europe's eyes on Earth. Sentinel-1: Level-1 data products (Εικόνα). Τελευταία Πρόσβαση 2020-10-23.
- [36] Copernicus: Europe's eyes on Earth. Sentinel-1: Data products (level2). <https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-1-sar/product-types-processing-levels/level-2>, Τελευταία Πρόσβαση 2020-10-23.
- [37] Copernicus: Europe's eyes on Earth. Sentinel-1 revisit and coverage. <https://sentinel.esa.int/web/sentinel/user-guides/sentinel-1-sar/revisit-and-coverage>, Τελευταία Πρόσβαση 2020-11-17.
- [38] Copernicus: Europe's eyes on Earth. Sentinel-2 satellite description. <https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-2/satellite-description>, Τελευταία Πρόσβαση 2020-10-29.
- [39] Copernicus: Europe's eyes on Earth. Sentinel-2 overview. <https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-2/overview>, Τελευταία Πρόσβαση 2020-10-29.
- [40] Copernicus: Europe's eyes on Earth. Sentinel-2 overview (user guides). <https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2/msi/overview>, Τελευταία Πρόσβαση 2020-10-29.
- [41] Copernicus: Europe's eyes on Earth. Sentinel-2 multispectral instrument (msi) overview. <https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-2-msi/msi-instrument>, Τελευταία Πρόσβαση 2020-10-29.
- [42] Copernicus: Europe's eyes on Earth. Sentinel-2 toolbox. <https://sentinels.copernicus.eu/web/sentinel/toolboxes/sentinel-2>, Τελευταία Πρόσβαση 2020-10-30.
- [43] Copernicus: Europe's eyes on Earth. Sentinel-2 payload data ground segment (pdgs). <https://sentinel.esa.int/web/sentinel/missions/sentinel-2/ground-segment/core-ground-segment/pdgs>, Τελευταία Πρόσβαση 2020-10-30.
- [44] Copernicus: Europe's eyes on Earth. Sentinel-2 multispectral instrument (msi) processing levels (Εικόνα). Τελευταία Πρόσβαση 2020-10-30.
-

-
- [45] esa earthnet online. 2.5.2 level 0 instrument source packet description. <https://earth.esa.int/handbooks/asar/CNTR2-5-2.html>, Τελευταία Πρόσβαση 2020-10-30.
- [46] Copernicus: Europe’s eyes on Earth. Sentinel-2 multispectral instrument (msi) product types level-0. <https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi/product-types/level-0>, Τελευταία Πρόσβαση 2020-10-30.
- [47] Copernicus: Europe’s eyes on Earth. Sentinel-2 multispectral instrument (msi) product types level-1a. <https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi/product-types/level-1a>, Τελευταία Πρόσβαση 2020-10-31.
- [48] Copernicus: Europe’s eyes on Earth. Sentinel-2 multispectral instrument (msi) product types level-1b. <https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi/product-types/level-1b>, Τελευταία Πρόσβαση 2020-10-31.
- [49] Copernicus: Europe’s eyes on Earth. Sentinel-2 multispectral instrument (msi) product types level-2a. <https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi/product-types/level-2a>, Τελευταία Πρόσβαση 2020-10-30.
- [50] Copernicus: Europe’s eyes on Earth. Sentinel-2 multispectral instrument (msi) products and algorithms level1-c processing cloud masks. <https://sentinel.esa.int/web/sentinel/technical-guides/sentinel-2-msi/level-1c/cloud-masks>, Τελευταία Πρόσβαση 2020-10-30.
- [51] Copernicus: Europe’s eyes on Earth. Sentinel-2, multispectral instrument (msi), products and algorithms, level2-a processing overview, product formatting, level-2a product formatting. <https://sentinel.esa.int/web/sentinel/technical-guides/sentinel-2-msi/level-2a/product-formatting>, Τελευταία Πρόσβαση 2020-10-30.
- [52] Copernicus: Europe’s eyes on Earth. Sentinel-2 multispectral instrument (msi) product types (Εικόνα). Τελευταία Πρόσβαση 2020-10-31.
- [53] Copernicus: Europe’s eyes on Earth. Sentinel-2 revisit and coverage. <https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/revisit-coverage>, Τελευταία Πρόσβαση 2020-11-17.
- [54] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.
- [55] James A. Anderson and Edward Rosenfeld. *Talking Nets: An Oral History of Neurocomputing*. MIT Press, Cambridge, MA, USA, 1998.
- [56] Warren S McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, 1943.
- [57] F. Rosenblatt. *The Perceptron, a Perceiving and Recognizing Automaton Project Para*. Report: Cornell Aeronautical Laboratory. Cornell Aeronautical Laboratory, 1957.
- [58] D. Rumelhart, Geoffrey E. Hinton, and R. J. Williams. Learning representations by back-propagating errors. *Nature*, 323:533–536, 1986.
- [59] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1(4):541–551, 1989.
- [60] Ruslan Salakhutdinov and Geoffrey Hinton. Deep boltzmann machines. In David van Dyk and Max Welling, editors, *Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics*, volume 5 of *Proceedings of Machine Learning Research*, pages 448–455, Hilton Clearwater Beach Resort, Clearwater Beach, Florida USA, 16–18 Apr 2009. PMLR.
- [61] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 25, pages 1097–1105. Curran Associates, Inc., 2012.
- [62] Ross B. Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. *CoRR*, abs/1311.2524, 2013.
- [63] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv 1409.1556*, 09 2014.
- [64] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks, 2014.
- [65] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(56):1929–1958, 2014.
- [66] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.
-

-
- [67] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015.
- [68] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. *CoRR*, abs/1706.03762, 2017.
- [69] Wei Di, Anurag Bhardwaj, and Jianing Wei. *Deep Learning Essentials: Your Hands-on Guide to the Fundamentals of Deep Learning and Neural Network Modeling*. Packt Publishing, 2018.
- [70] Ernst Ising. Beitrag zur Theorie des Ferromagnetismus. *Zeitschrift für Physik*, 31(1):253–258, February 1925.
- [71] J J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8):2554–2558, 1982.
- [72] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9:1735–80, 12 1997.
- [73] Christopher Olah. Understanding lstm networks, Posted on August 27, 2015. <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>, Τελευταία Πρόσβαση 2021-02-03.
- [74] F. A. Gers and J. Schmidhuber. Recurrent nets that time and count. In *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium*, volume 3, pages 189–194 vol.3, 2000.
- [75] Kyunghyun Cho, Bart van Merriënboer, Çağlar Gülçehre, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *CoRR*, abs/1406.1078, 2014.
- [76] Kunihiko Fukushima. Neocognitron: A hierarchical neural network capable of visual pattern recognition. *Neural Networks*, 1(2):119–130, 1988.
- [77] Akinori Hidaka and Takio Kurita. Consecutive dimensionality reduction by canonical correlation analysis for visualization of convolutional neural networks. volume 2017, pages 160–167, 12 2017.
- [78] Tero Karras, Miika Aittala, Janne Hellsten, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Training generative adversarial networks with limited data. In *Proc. NeurIPS*, 2020.
- [79] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2019.
- [80] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners. 2020.
- [81] Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. Albert: A lite bert for self-supervised learning of language representations, 2020.
- [82] Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter, 2020.
- [83] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach, 2019.
- [84] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale, 2020.
- [85] William Fedus, Barret Zoph, and Noam Shazeer. Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity, 2021.
- [86] Rajalingappaa Shanmugamani Giuseppe Bonaccorso, Armando Fandango. *Python: Advanced Guide to Artificial Intelligence*. Packt Publishing, 2018.
- [87] Jasper Uijlings, K. Sande, T. Gevers, and A.W.M. Smeulders. Selective search for object recognition. *International Journal of Computer Vision*, 104:154–171, 09 2013.
- [88] Ross B. Girshick. Fast R-CNN. *CoRR*, abs/1504.08083, 2015.
- [89] Shaoqing Ren, Kaiming He, Ross B. Girshick, and Jian Sun. Faster R-CNN: towards real-time object detection with region proposal networks. *CoRR*, abs/1506.01497, 2015.
-

-
- [90] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. *CoRR*, abs/1411.4038, 2014.
- [91] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross B. Girshick. Mask R-CNN. *CoRR*, abs/1703.06870, 2017.
- [92] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection, 2017.
- [93] Jifeng Dai, Kaiming He, and Jian Sun. Instance-aware semantic segmentation via multi-task network cascades. *CoRR*, abs/1512.04412, 2015.
- [94] Georgia Gkioxari, Jitendra Malik, and Justin Johnson. Mesh R-CNN. *CoRR*, abs/1906.02739, 2019.
- [95] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers, 2020.
- [96] H. W. Kuhn and Bryn Yaw. The hungarian method for the assignment problem. *Naval Res. Logist. Quart.*, pages 83–97, 1955.
- [97] Roberto Interdonato, Dino Ienco, Raffaele Gaetano, and Kenji Ose. Duplo: A dual view point deep learning architecture for time series classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 149:91–104, 2019.
- [98] COCO. Common object in context. <https://cocodataset.org/>, Τελευταία Πρόσβαση 2021-01-14.
- [99] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, Pennsylvania, USA, July 2002. Association for Computational Linguistics.
- [100] Chin-Yew Lin. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pages 74–81, Barcelona, Spain, July 2004. Association for Computational Linguistics.
- [101] European Union. Copernicus: Europe’s eyes on earth. <https://www.copernicus.eu/>, Τελευταία Πρόσβαση 2020-10-13.
- [102] Copernicus: Europe’s eyes on Earth. Sentinel-2. <https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-2>, Τελευταία Πρόσβαση 2020-11-26.
- [103] Python. <https://www.python.org/>, Τελευταία Πρόσβαση 2020-11-26.
- [104] matplotlib. Matplotlib: Visualization with python. <https://matplotlib.org/>, Τελευταία Πρόσβαση 2020-11-26.
- [105] QGIS. A free and open source geographic information system. <https://qgis.org/en/site/>, Τελευταία Πρόσβαση 2020-11-26.
- [106] Natural Earth. <https://www.naturalearthdata.com/>, Τελευταία Πρόσβαση 2020-11-28.
- [107] Natural Earth. Populated places, (populated places) version 4.1.0. <https://www.naturalearthdata.com/downloads/10m-cultural-vectors/>, Τελευταία Πρόσβαση 2020-11-28.
- [108] Natural Earth. Admin 0 - boundary lines, (land boundaries) version 4.1.0. <https://www.naturalearthdata.com/downloads/10m-cultural-vectors/>, Τελευταία Πρόσβαση 2020-11-28.
- [109] Natural Earth. Natural earth ii with shaded relief, water, and drainages (large size) version 3.2.0. <https://www.naturalearthdata.com/downloads/10m-raster-data/10m-natural-earth-2/>, Τελευταία Πρόσβαση 2020-11-28.
- [110] Copernicus: Europe’s eyes on Earth. Sentinel-2, data products. <https://sentinel.esa.int/web/sentinel/missions/sentinel-2/data-products>, Τελευταία Πρόσβαση 2020-11-29.
- [111] justinelliottmeyers. Sentinel-2-shapefile-index. <https://github.com/justinelliottmeyers/Sentinel-2-Shapefile-Index>, Τελευταία Πρόσβαση 2020-11-28.
- [112] Generalitat de Catalunya. Department of agriculture, livestock, fisheries and food. <http://agricultura.gencat.cat/ca/inici/>, Τελευταία Πρόσβαση 2020-11-27.
- [113] Generalitat de Catalunya. Department of agriculture, livestock, fisheries and food, services, cartography and gis, thematic applications, sigpac, crop map. <http://agricultura.gencat.cat/ca/serveis/cartografia-sig/aplicatiu-tematic-geoinformacio/sigpac/mapa-cultius/>, Τελευταία Πρόσβαση 2020-11-27.
- [114] GEOJSON. The geojson specification (rfc 7946). <https://geojson.org/>, Τελευταία Πρόσβαση 2020-11-26.
- [115] SentinelSat. <https://sentinelat.readthedocs.io/en/stable/>, Τελευταία Πρόσβαση 2020-11-26.
- [116] Copernicus Open Access Hub. <https://scihub.copernicus.eu/dhus/#/home>, Τελευταία Πρόσβαση 2020-11-26.
-

-
- [117] Copernicus Open Access Hub. Lta. <https://scihub.copernicus.eu/userguide/LongTermArchive>, Τελευταία Πρόσβαση 2020-11-26.
- [118] Copernicus: Europe’s eyes on Earth. Sentinel-2, user guides, sentinel-2 msi, product naming convention. <https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/naming-convention>, Τελευταία Πρόσβαση 2020-11-27.
- [119] Generalitat de Catalunya. Department of agriculture, livestock, fisheries and food, services, cartography and gis, thematic applications, sigpac, crop map (source of crop map data pdf). http://agricultura.gencat.cat/web/content/09-desenvolupament-rural/comu/dades_obertes/origen-dades-mapa-cultius.pdf, Τελευταία Πρόσβαση 2020-11-27.
- [120] GeoPandas 0.8.0. <https://geopandas.org/>, Τελευταία Πρόσβαση 2020-11-27.
- [121] Apache Arrow. Apache arrow 2.0.0 release, 22 Oct 2020. <https://arrow.apache.org/blog/2020/10/22/2.0.0-release/>, Τελευταία Πρόσβαση 2021-01-16.
- [122] Apache Arrow Documentation. Feather file format. <https://arrow.apache.org/docs/python/feather.html>, Τελευταία Πρόσβαση 2021-01-16.
- [123] Kelsey Jordahl, Joris Van den Bossche, Martin Fleischmann, Jacob Wasserman, James McBride, Jeffrey Gerard, Jeff Tratner, Matthew Perry, Adrian Garcia Badaracco, Carson Farmer, Geir Arne Hjelle, Alan D. Snow, Micah Cochran, Sean Gillies, Lucas Culbertson, Matt Bartos, Nick Eubank, maxalbert, Aleksey Bilogur, Sergio Rey, Christopher Ren, Dani Arribas-Bel, Leah Wasser, Levi John Wolf, Martin Journois, Joshua Wilson, Adam Greenhall, Chris Holdgraf, Filipe, and François Leblanc. *geopandas/geopandas: v0.8.1*, July 2020.
- [124] Food and Agriculture Organization of the United Nations. *A system of integrated agricultural censuse and surveys. Volume 1, World programme for the census of agriculture 2010*. FAO statistical development series, no. 11. FAO, Rome, 2005.
- [125] André Hollstein, Karl Segl, Luis Guanter, Maximilian Brell, and Marta Enesco. Ready-to-use methods for the detection of clouds, cirrus, snow, shadow, water and clear sky pixels in sentinel-2 msi images. *Remote Sensing*, 8(8), 2016.
- [126] THE EUROPEAN SPACE AGENCY. Sentinel online: Home, user guides, sentinel-2 msi, processing levels, level-2. <https://sentinel.esa.int/web/sentinel/user-guides/Sentinel-2-msi/processing-levels/level-2>, Τελευταία Πρόσβαση 2021-01-21.
- [127] BigEarthNet. A large-scale sentinel benchmark archive v1.0. <http://bigearth.net/>, Τελευταία Πρόσβαση 2021-01-01.
- [128] S. Hoyer and J. Hamman. xarray: N-d labeled arrays and datasets in python. *Journal of Open Research Software*, 5(1), 2017.
- [129] Stephan Hoyer, Clark Fitzgerald, Joe Hamman, et al. xarray: v0.17.0, February 2021.
- [130] Charles Zender. Analysis of self-describing gridded geoscience data with netcdf operators (nco). *Environmental Modelling & Software*, 23:1338–1342, 10 2008.
- [131] Kaiming He, Ross Girshick, and Piotr Dollár. Rethinking imagenet pre-training, 2018.
- [132] Gencer Sumbul, Jian Kang, Tristan Kreuziger, Filipe Marcelino, Hugo Costa, Pedro Benevides, Mario Caetano, and Begüm Demir. Bigearthnet dataset with a new class-nomenclature for remote sensing image understanding, 2020.
- [133] Jorge Andres Chamorro Martinez, Laura Elena Cué La Rosa, Raul Queiroz Feitosa, Ieda Del’Arco Sanches, and Patrick Nigri Happ. Fully convolutional recurrent networks for multivariate crop recognition from multitemporal image sequences. *ISPRS Journal of Photogrammetry and Remote Sensing*, 171:188–201, 2021.
- [134] Aiyim Orynbaikyzy, Ursula Gessner, Benjamin Mack, and Christopher Conrad. Crop type classification using fusion of sentinel-1 and sentinel-2 data: Assessing the impact of feature selection, optical data availability, and parcel sizes on the accuracies. *Remote Sensing*, 12(17), 2020.
- [135] Yuxin Wu, Alexander Kirillov, Francisco Massa, Wan-Yen Lo, and Ross Girshick. Detectron2. <https://github.com/facebookresearch/detectron2>, 2019.
- [136] Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei. Deformable convolutional networks, 2017.
- [137] Azavea. Raster vision. <https://github.com/azavea/raster-vision>.
- [138] Praveer Singh and Nikos Komodakis. Cloud-gan: Cloud removal for sentinel-2 imagery using a cyclic consistent generative adversarial networks. In *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, pages 1772–1775, 2018.
-

-
- [139] Vishnu Sarukkai, Anirudh Jain, Burak Uzkent, and Stefano Ermon. Cloud removal in satellite images using spatiotemporal generative networks. In *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1785–1794, 2020.
- [140] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. Photo-realistic single image super-resolution using a generative adversarial network, 2017.
- [141] Zhongyuan Wang, Kui Jiang, Peng Yi, Zhen Han, and Zheng He. Ultra-dense gan for satellite imagery super-resolution. *Neurocomputing*, 398:328–337, 2020.
- [142] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks, 2015.
- [143] Xiao Xiang Zhu, Devis Tuia, Lichao Mou, Gui-Song Xia, Liangpei Zhang, Feng Xu, and Friedrich Fraundorfer. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4):8–36, Dec 2017.
- [144] W. Li, G. Wu, and Q. Du. Transferred deep learning for anomaly detection in hyperspectral imagery. *IEEE Geoscience and Remote Sensing Letters*, 14(5):597–601, 2017.