



Εθνικό Μετσόβιο Πολυτεχνείο

Σχολή Αγρονόμων και Τοπογράφων Μηχανικών

Τομέας Τοπογραφίας

Εργαστήριο Τηλεπισκόπησης

**Σημειολογική Κατάτμηση Μέσω Συνελικτικών Νευρωνικών
Δικτύων σε Πολύ Υψηλής Ανάλυσης Εναέρια Τηλεπισκοπικά
Δεδομένα**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΚΩΣΤΑΡΕΛΟΥ ΑΙΚΑΤΕΡΙΝΗ

Αθήνα, Οκτώβριος 2020



NATIONAL TECHNICAL UNIVERSITY OF ATHENS
SCHOOL OF RURAL AND SURVEY ENGINEERING
REMOTE SENSING LABORATORY

Semantic Segmentation via Convolutional Neural Networks of Very High Resolution Aerial Datasets

DIPLOMA THESIS

KOSTARELOU AIKATERINI

Athens, October 2020



RSLab

Remote Sensing Laboratory
National Technical University of Athens

✓ Sensing ✓ Analytics ✓ Monitoring





Εθνικό Μετσόβιο Πολυτεχνείο

Σχολή Αγρονόμων και Τοπογράφων Μηχανικών

Τομέας Τοπογραφίας

Εργαστήριο Τηλεπισκόπησης

**Σημασιολογική Κατάτμηση Μέσω Συνελικτικών Νευρωνικών
Δικτύων σε Πολύ Υψηλής Ανάλυσης Εναέρια Τηλεπισκοπικά
Δεδομένα**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΚΩΣΤΑΡΕΛΟΥ ΑΙΚΑΤΕΡΙΝΗ

Τριμελής εξεταστική επιτροπή

Κωνσταντίνος Καράντζαλος

Αναπληρωτής Καθηγητής Ε.Μ.Π.

(επιβλέπων)

Μαρία Παπαδοπούλου

Καθηγήτρια Ε.Μ.Π.

Δημήτριος Αργιαλάς

Καθηγητής Ε.Μ.Π.

Αθήνα, Οκτώβριος 2020

.....
Αικατερίνη Κωσταρέλου

Διπλωματούχος Αγρονόμος και Τοπογράφος Μηχανικός Ε.Μ.Π.

© 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.

Περίληψη

Η ανάπτυξη τεχνικών βαθιάς μάθησης τα τελευταία χρόνια, μαζί με την πρόσφατη σύνταξη μεγάλων συνόλων δεδομένων εκπαίδευσης και ελέγχου, επέτρεψε την αποτελεσματική αντιμετώπιση πολλών προβλημάτων ταξινόμησης με σχετικά πολύ υψηλές ακρίβειες εντοπισμού. Συγκεκριμένα, στο πεδίο της Γεωπληροφορικής και Παρατήρησης Γης, όπου καθημερινά κάνουν την εμφάνισή τους νέα δεδομένα, εναέρια και δορυφορικά, πολύ υψηλής ανάλυσης, κρίνεται αναγκαία μια αυτοματοποιημένη ανάλυση και αξιοποίηση των δεδομένων.

Η σημασιολογική κατάτμηση είναι μία από τις βασικές τεχνολογίες για την ανάλυση, ερμηνεία και αξιοποίηση τηλεπισκοπικών δεδομένων. Σε αυτήν τη διπλωματική εργασία παρουσιάζονται δύο μοντέλα με εξαιρετικές αποδόσεις πάνω στη σημασιολογική κατάτμηση, καθώς και μία δική μας πρόταση που προκύπτει από το συνδυασμό τους.

Η διαδικασία σχεδίασης του συνελικτικού νευρωνικού μας δικτύου, βασίζεται στο δημοφιλές μοντέλο U-Net και στο μοντέλο που προτάθηκε από τους Marmanis et al., για την σημασιολογική κατάτμηση του συνόλου δεδομένων του διαγωνισμού 2-διάστατης Σημασιολογικής Ταξινόμησης της Διεθνούς Κοινότητας Φωτογραμμετρίας και Ψηφιακής Τηλεπισκόπησης.

Η εκπαίδευση και αξιολόγηση του μοντέλου μας έγινε σε σύνολο δεδομένων της βιβλιογραφίας και συγκεκριμένα του επιστημονικού οργανισμού ISPRS. Για τη σύγκριση της απόδοσής του σχετικά με τα πρώτα δύο μοντέλα, υλοποιήσαμε και τα τρία με τις ίδιες παραμέτρους, για το ίδιο σύνολο εκπαίδευσης και το ίδιο σύνολο ελέγχου. Η εκπαίδευση του νευρωνικού δικτύου έλαβε χώρα σε ένα απομακρυσμένο διαδικτυακό περιβάλλον μεγάλης υπολογιστικής ισχύος, ενώ ο απαραίτητος κώδικας γράφτηκε στη γλώσσα προγραμματισμού Python. Τέλος, για την αξιολόγηση των μοντέλων χρησιμοποιήθηκαν κάποιες μετริกές, οι οποίες ανέδειξαν την πρόταση μας ως τη βέλτιστη, για το συγκεκριμένο πρόβλημα ταξινόμησης, με ολική ορθότητα 91% και συντελεστή Kappa 0.87.

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ

Βαθιά Μάθηση, Όραση Υπολογιστή, Συνελικτικά Νευρωνικά Δίκτυα, Σημασιολογική Κατάτμηση, Ψηφιακή Τηλεπισκόπηση, U-Net

Abstract

The development of deep learning techniques in recent years, combined with the recent compilation of large sets of training and testing data, has made it possible to deal effectively with many classification problems with relatively high detection accuracy. Specifically, in the field of Geoinformatics and Earth Observation, where new data, aerial and satellite, of very high resolution, appear every day, an automated analysis and exploitation of data is considered necessary.

Semantic segmentation is one of the key technologies for analyzing, interpreting and utilizing remote sensing data. In this diploma thesis we present two models with excellent performances on the semantic segmentation task, as well as a proposed model of ours that results from their combination.

The design process of our convolutional neural network is based on the popular U-Net model and the model proposed by Marmanis et al., for the semantic segmentation of the dataset of the 2D Semantic Labeling Contest of the ISPRS (International Society for Photogrammetry and Remote Sensing).

The training and evaluation of our model was done in a set of data of the mentioned literature and specifically, of the scientific organization ISPRS. To compare its performance with the first two models, we implemented all three with the same parameters, for the same training set and the same test set. The neural network training took place in a remote internet environment of great computing power, while the necessary code was written in the Python programming language. Finally, for the evaluation of the models certain metrics were used, which highlighted our proposal as the optimal, for the specific classification problem, with a total accuracy of 91% and a Kappa coefficient of 0.87.

KEYWORDS

Deep Learning, Computer Vision, Convolutional Neural Networks, Semantic Segmentation, Remote Sensing, U-Net



RSLab

Remote Sensing Laboratory
National Technical University of Athens

✓ Sensing ✓ Analytics ✓ Monitoring



Ευχαριστίες

Η παρούσα διπλωματική εργασία εκπονήθηκε στο πλαίσιο του Προπτυχιακού Προγράμματος Σπουδών της Σχολής Αγρονόμων και Τοπογράφων Μηχανικών του Εθνικού Μετσόβιου Πολυτεχνείου.

Κατ' αρχάς θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου κ. Κωνσταντίνο Καράντζαλο για την ευκαιρία που μου έδωσε να αναπτύξω την έρευνα μου στο πεδίο της Βαθιάς Μηχανικής Μάθησης, αποκτώντας σημαντικές γνώσεις για την ακαδημαϊκή κι επαγγελματική μου πορεία, καθώς και για τη στήριξή του καθ' όλη τη διάρκειά της. Επίσης, θα ήθελα να ευχαριστήσω τον μεταδιδάκτορα κ. Μάκη Ντούσκο για τη βοήθεια και τις γνώσεις που μου παρείχε με μεγάλη χαρά και προθυμία.

Θα ήθελα να διανείμω ιδιαίτερες ευχαριστίες στους ανθρώπους μου που με στήριξαν με τον δικό τους μοναδικό τρόπο όλα αυτά τα χρόνια και που πίστεψαν σε εμένα ακόμη και στις πιο δύσκολες περιόδους που παρήλθαν.

Με κάθε σεβασμό στους προαναφερθέντες, θα ήθελα να ευχαριστήσω ξεχωριστά, εξαιρετικά κι ειλικρινά τη φίλη μου Μαρία, στην οποία οφείλω ολόκληρη την ακαδημαϊκή μου σταδιοδρομία. Χάρης στην ανιδιοτελή παραχώρηση του πολύτιμου της χρόνου, βρίσκομαι σε θέση να συγγράφω αυτό το τεύχος.

Τέλος, θα ήθελα να αφιερώσω την προκείμενη εργασία αλλά κι ολόκληρη τη ζωή κι υπόστασή μου στους γονείς μου, που δίχως την καθημερινή αυτοθυσία, αγάπη και στήριξή τους δεν θα είχα την τιμή και το προνόμιο να μορφωθώ κι εν τέλει να διαμορφωθώ τόσο ως επιστήμονας αλλά πιο σημαντικά ως άνθρωπος.

Πίνακας περιεχομένων

Περίληψη	i
Abstract	iii
Ευχαριστίες	v
Πίνακας περιεχομένων	vii
ΚΕΦΑΛΑΙΟ 1	1
Εισαγωγή	1
1.1 Τεχνητή Νοημοσύνη	1
1.2 Μηχανική Μάθηση	2
1.3 Βαθιά Μάθηση.....	3
1.3.1 Όραση Υπολογιστή.....	3
1.3.2 Βαθιά Μάθηση και Ψηφιακή Τηλεπισκόπηση	4
ΚΕΦΑΛΑΙΟ 2	5
Θεωρητικό Υπόβαθρο.....	5
2.1 Τεχνητά Νευρωνικά Δίκτυα	5
2.1.1 Αντίληπτρο.....	6
2.1.2 Πολυεπίπεδο Αντίληπτρο.....	8
2.1.2.1 Επιβλεπόμενη Μάθηση	9
2.1.2.2 Οπισθοδιάδοση.....	9
2.2 Συνελικτικά Νευρωνικά Δίκτυα.....	11
2.2.1 Αρχιτεκτονική Συνελικτικών Νευρωνικών Δικτύων	12
2.2.2 Συνελικτικά Επίπεδα.....	13
2.2.2.1 Επίπεδο Συνέλιξης.....	13
2.2.2.2 Επίπεδο Συγκέντρωσης	16
2.2.2.3 Πλήρως Συνδεδεμένο Επίπεδο	17
2.2.2.4 Εξειδικευμένα Επίπεδα	18
2.3 Εκπαίδευση	19
2.3.1 Συναρτήσεις Ενεργοποίησης.....	19

2.3.1.1	Συνάρτηση Διορθωμένης Γραμμικής Μονάδας	19
2.3.1.2	Κανονικοποιημένη Εκθετική Συνάρτηση.....	20
2.3.2	Συνάρτηση Σφάλματος.....	21
2.3.2.1	Συνάρτηση Σταυροειδούς Εντροπίας	21
2.3.3	Αλγόριθμοι Βελτιστοποίησης	22
2.3.3.1	Αλγόριθμος Στοχαστικής Κλίσης Καθόδου.....	22
2.3.4	Κανονικοποίηση.....	23
2.3.4.1	Κανονικοποίηση Παρτίδας.....	24
2.3.4.2	Κανονικοποίηση Τοπικής Απόκρισης	25
2.3.5	Υπερπροσαρμογή.....	26
2.3.5.1	Απόσυρση.....	27
2.3.5.2	Ενίσχυση Δεδομένων	28
2.3.5.3	Μεταφορά Μάθησης	28
2.3.6	Μετρικές Αξιολόγησης.....	29
2.3.6.1	Πίνακας Σύγχυσης.....	30
2.3.6.2	Ορθότητα.....	31
2.3.6.3	Ακρίβεια.....	31
2.3.6.4	Ανάκληση.....	32
2.3.6.5	Αρμονικός Μέσος F1	32
2.3.6.6	Μέσοι Όροι	33
2.3.6.7	Ο Συντελεστής Καρρα.....	34
	ΚΕΦΑΛΑΙΟ 3	36
	Μοντέλα Βαθιάς Μάθησης για Σημασιολογική Κατάτμηση.....	36
3.1	Το μοντέλο U-Net.....	36
3.2	Το μοντέλο VGG-16.....	38
3.3	Το μοντέλο FCN-Pascal	40
3.4	Το μοντέλο των Marmanis et al.	41
	ΚΕΦΑΛΑΙΟ 4	43
	Μεθοδολογία, Πειραματική Διαδικασία και Αποτελέσματα	43
4.1	Σύνολο Δεδομένων.....	43

4.2 Υλικό Υπολογιστή (Hardware)	45
4.3 Εκπαίδευση.....	46
4.4 Εργαλεία.....	46
4.5 Σχεδιασμός και Μεθοδολογία Προτεινόμενου Μοντέλου	47
4.6 Αποτελέσματα.....	50
4.6.1 Αποτελέσματα U-Net.....	50
4.6.1.1 Περιοχή 5	50
4.6.1.2 Περιοχή 7	54
4.6.1.3 Περιοχή 23	57
4.6.1.4 Περιοχή 30.....	61
4.6.1.5 Σύγκριση Μασκών Πρόβλεψης του μοντέλου U-Net με την Αληθή Βάση.....	64
4.6.2 Αποτελέσματα Marmanis et al.	66
4.6.2.1 Περιοχή 5	66
4.6.2.2 Περιοχή 7	70
4.6.2.3 Περιοχή 23	73
4.6.2.4 Περιοχή 30.....	77
4.6.2.5 Σύγκριση Μασκών Πρόβλεψης του μοντέλου Marmanis et al. με την Αληθή Βάση.....	80
4.6.3 Αποτελέσματα Προτεινόμενου Μοντέλου.....	82
4.6.3.1 Περιοχή 5	82
4.6.3.2 Περιοχή 7	86
4.6.3.3 Περιοχή 23	89
4.6.3.4 Περιοχή 30.....	93
4.6.3.5 Σύγκριση Μασκών Πρόβλεψης του Μοντέλου μας με την Αληθή Βάση	96
4.7 Αξιολόγηση Αποτελεσμάτων	97
ΚΕΦΑΛΑΙΟ 5	101
Επίλογος.....	101
5.1 Συμπεράσματα	101
5.2 Προτάσεις Βελτίωσης	102
Βιβλιογραφία	103
Κατάλογος Σχημάτων.....	106

Κατάλογος Πινάκων.....	108
Αγγλο-Ελληνικό Λεξικό.....	109

ΚΕΦΑΛΑΙΟ 1

Εισαγωγή

1.1 Τεχνητή Νοημοσύνη

Τεχνητή Νοημοσύνη (Artificial Intelligence), είναι η ικανότητα ενός ψηφιακού υπολογιστή να εκτελεί εργασίες που σχετίζονται, συνήθως, με ευφυή όντα. Ο όρος εφαρμόζεται συχνά στο έργο ανάπτυξης συστημάτων που τους έχουν αποδοθεί πνευματικές διεργασίες που χαρακτηρίζουν τον άνθρωπο, όπως η ικανότητα λογικής, ανακάλυψης νοήματος, γενίκευσης ή εκμάθησης από την εμπειρία του παρελθόντος. Από την ανάπτυξη του ψηφιακού υπολογιστή στη δεκαετία του 1940, έχει αποδειχθεί ότι οι υπολογιστές μπορούν να προγραμματιστούν για την εκτέλεση πολύπλοκων εργασιών.

Η Τεχνητή Νοημοσύνη αναφέρεται στην ικανότητα ενός υπολογιστή ή μιας μηχανής να μιμείται τις δυνατότητες του ανθρώπινου μυαλού, μαθαίνοντας από παραδείγματα και εμπειρίες, αναγνωρίζοντας αντικείμενα, κατανοώντας τη φυσική γλώσσα, λαμβάνοντας αποφάσεις κι επιλύοντας προβλήματα. Η ανάπτυξή της καθίσταται δυνατή από την ξαφνική διαθεσιμότητα μεγάλων ποσοτήτων δεδομένων και την αντίστοιχη ανάπτυξη και την ευρεία διαθεσιμότητα συστημάτων υπολογιστών που μπορούν να επεξεργαστούν όλα αυτά τα δεδομένα γρηγορότερα και με μεγαλύτερη ακρίβεια από τους ανθρώπους.

Οι στόχοι της Τεχνητής Νοημοσύνης περιλαμβάνουν τη μάθηση, τον συλλογισμό και την αντίληψη. Οι μαθησιακές διαδικασίες περιλαμβάνουν την απόκτηση νέων δηλωτικών γνώσεων, την ανάπτυξη κινητικών και γνωστικών δεξιοτήτων μέσω διδασκαλίας ή πρακτικής, την οργάνωση νέων γνώσεων σε γενικές, αποτελεσματικές αναπαραστάσεις και την ανακάλυψη νέων γεγονότων και θεωριών μέσω παρατήρησης και πειραματισμού. Από την έναρξη της εποχής των υπολογιστών, οι ερευνητές προσπαθούν να εμφυτεύσουν τέτοιες δυνατότητες σε υπολογιστές. Η επίλυση αυτού του προβλήματος ήταν, και παραμένει, ο πιο απαιτητικός στόχος μεγάλης εμβέλειας στην Τεχνητή Νοημοσύνη. Η μελέτη και η μοντελοποίηση υπολογιστών σε διαδικασίες μάθησης στις πολλαπλές εκδηλώσεις τους, αποτελούν το αντικείμενο της Μηχανικής Μάθησης.

1.2 Μηχανική Μάθηση

Η *Μηχανική Μάθηση* (Machine Learning) είναι η επιστήμη προγραμματισμού υπολογιστών, ώστε να μπορούν να μάθουν από τα δεδομένα. Τα συστήματα Μηχανικής Μάθησης μπορούν να προσαρμοστούν αποτελεσματικά κι άμεσα σε νέα δεδομένα και να αποκτήσουν πληροφορίες για μεγάλες βάσεις δεδομένων.

Έχει αποδειχθεί χρήσιμο εργαλείο για την επίλυση προβλημάτων για τα οποία οι υπάρχουσες λύσεις απαιτούν πολλές χειροκίνητες ρυθμίσεις ή μεγάλες λίστες κανόνων. Ένας αλγόριθμος Μηχανικής Μάθησης μπορεί συχνά να απλοποιήσει τον κώδικα και να αποδώσει καλύτερα. Μπορεί, επίσης, να αντιμετωπίσει πολύπλοκα προβλήματα για τα οποία δεν υπάρχει αποδοτική λύση χρησιμοποιώντας μια παραδοσιακή προσέγγιση.

Η Μηχανική Μάθηση μπορεί να χωριστεί στις παρακάτω κατηγορίες, σύμφωνα με:

- Εάν έχουν εκπαιδευτεί ή όχι με ανθρώπινη επίβλεψη. (Supervised Learning (Επιβλεπόμενη Μάθηση), Unsupervised (Μη Επιβλεπόμενη), Semi-supervised (Ημι-επιβλεπόμενη) και Reinforcement Learning (Ενισχυτική Μάθηση))
- Εάν μπορούν ή όχι να μάθουν σταδιακά κατά τη λειτουργία τους.
- Εάν λειτουργούν είτε συγκρίνοντας νέα δεδομένα με γνωστά δεδομένα, είτε αντ' αυτού ανιχνεύοντας μοτίβα στα δεδομένα εκπαίδευσης δημιουργώντας ένα μοντέλο πρόβλεψης, όπως κάνουν οι επιστήμονες.

Στην παρούσα εργασία χρησιμοποιούμε μοντέλα που έχουν εκπαιδευτεί με ανθρώπινη επίβλεψη και συγκεκριμένα, με την Επιβλεπόμενη Μάθηση. Στην Επιβλεπόμενη Μάθηση, τα δεδομένα εκπαίδευσης που τροφοδοτούνται στον αλγόριθμο περιλαμβάνουν τις επιθυμητές λύσεις, που ονομάζονται *ετικέτες* (labels). Μια τυπική εργασία είναι η πρόβλεψη μιας αριθμητικής τιμής, *στόχου* (target), δεδομένου ενός συνόλου *χαρακτηριστικών* (features), που ονομάζονται *προβλέψεις* (predictions). Αυτό το είδος εργασίας ονομάζεται *παλινδρόμηση* (regression). Μια άλλη τυπική επιβλεπόμενη μαθησιακή εργασία είναι η *ταξινόμηση* (classification), η οποία είναι η ομαδοποίηση δεδομένων σε *κλάσεις* (classes). Το μοντέλο που προτείνουμε βασίζεται στην ταξινόμηση μέσω διαδικασιών Βαθιάς Μάθησης και συγκεκριμένα στο πρόβλημα της *Σημασιολογικής Κατάτμησης* (Semantic Segmentation). [10]

1.3 Βαθιά Μάθηση

Λόγω της δυσκολίας που παρουσιάζεται κατά την εξαγωγή αφηρημένων χαρακτηριστικών από *ακατέργαστα δεδομένα* (raw data), κρίθηκε αναγκαία η εισαγωγή μιας νέας προσέγγισης, της *Βαθιάς Μάθησης* (Deep Learning). Συγκεκριμένα, είναι ένας τύπος Μηχανικής Μάθησης, και προσδιορίζεται ως μια τεχνική που επιτρέπει στα συστήματα υπολογιστών να βελτιωθούν με την εμπειρία και τα δεδομένα. Η Βαθιά Μάθηση είναι ένα ιδιαίτερο είδος Μηχανικής Μάθησης που επιτυγχάνει μεγάλη δύναμη και ευελιξία μαθαίνοντας να εκπροσωπεί τον κόσμο ως ένθετη ιεραρχία εννοιών, με κάθε έννοια να ορίζεται σε σχέση με απλούστερες έννοιες, και περισσότερες αφηρημένες παραστάσεις να υπολογίζονται με όρους λιγότερο αφηρημένων.

1.3.1 Όραση Υπολογιστή

Η *Όραση Υπολογιστή* (Computer Vision) είναι ένας από τους πιο ενεργούς ερευνητικούς τομείς για εφαρμογές Βαθιάς Μάθησης, επειδή η όραση είναι μια εργασία που είναι αβίαστη για τον άνθρωπο και πολλά ζώα, αλλά είναι δύσκολη για τους υπολογιστές. Πολλές από τις πιο δημοφιλείς εργασίες για αλγορίθμους Βαθιάς Μάθησης είναι η αναγνώριση αντικειμένων ή αναγνώριση οπτικών χαρακτηριστών.

Η *Όραση Υπολογιστή* είναι ένα πολύ ευρύ πεδίο που περιλαμβάνει μια μεγάλη ποικιλία τρόπων επεξεργασίας εικόνων και μια εκπληκτική ποικιλία εφαρμογών. Οι εφαρμογές της κυμαίνονται από την αναπαραγωγή ανθρώπινων οπτικών ικανοτήτων, όπως η αναγνώριση προσώπων, έως τη δημιουργία εντελώς νέων κατηγοριών οπτικών ικανοτήτων. Οι περισσότερες εφαρμογές της Βαθιάς Μάθησης για την *Όραση Υπολογιστή* χρησιμοποιούνται για την αναγνώριση αντικειμένων ή την ανίχνευση κάποιας μορφής, είτε αυτό σημαίνει αναφορά του αντικειμένου που υπάρχει σε μια εικόνα, σχολιασμός μιας εικόνας με πλαίσια οριοθέτησης γύρω από κάθε αντικείμενο, μεταγραφή μιας ακολουθίας συμβόλων από μια εικόνα ή επισήμανση κάθε εικονοστοιχείου σε μια εικόνα με την ταυτότητα του αντικειμένου στο οποίο ανήκει. [11]

1.3.2 Βαθιά Μάθηση και Ψηφιακή Τηλεπισκόπηση

Οι εικόνες της Ψηφιακής Τηλεπισκόπησης συχνά χαρακτηρίζονται από πολύπλοκες ιδιότητες δεδομένων με τη μορφή ετερογένειας και ανισορροπίας τάξης, καθώς και αλληλεπικαλυπτόμενες κατανομές κλάσεων. Συνδυαστικά, αυτές οι πτυχές αποτελούν σοβαρές προκλήσεις για τη δημιουργία χαρτών κάλυψης γης ή την ανίχνευση κι εντοπισμό αντικειμένων, δημιουργώντας υψηλό βαθμό αβεβαιότητας στα αποτελέσματα που λαμβάνονται, ακόμη και για τα μοντέλα με την καλύτερη απόδοση.

Τα τελευταία χρόνια, τα βαθιά *Νευρωνικά Δίκτυα* (Neural Networks) έχουν αναδειχθεί ως τα κορυφαία εργαλεία μοντελοποίησης για την ταξινόμηση και την κατάτμηση των εικονοστοιχείων της εικόνας, γενικά, και είχαν επίσης αυξανόμενο αντίκτυπο στην Τηλεπισκόπηση.

Η Τηλεπισκόπηση έχει αλλάξει ριζικά με την εισαγωγή της Βαθιάς Μάθησης. Η ταξινόμηση της κάλυψης γης, η ανίχνευση αντικειμένων κι η κατανόηση της σκηνής στις αεροφωτογραφίες βασίζονται όλο και περισσότερο σε βαθιά δίκτυα για να επιτύχουν νέα αποτελέσματα τελευταίας τεχνολογίας. Αρχιτεκτονικές, όπως τα *Συνελικτικά Νευρωνικά Δίκτυα* (Convolutional Neural Networks), μπορούν ακόμη να παράγουν σχολιασμούς σε επίπεδο εικονοστοιχείου για Σημασιολογική Κατάτμηση. Στην Τηλεπισκόπηση, η Σημασιολογική Κατάτμηση αναφέρεται συχνά ως ταξινόμηση εικόνας και η Σημασιολογική Κατάτμηση των εικόνων χωρίς τα κανάλια RGB μπορεί να εφαρμοσθεί στην ταξινόμηση κάλυψης γης, στην ταξινόμηση βλάστησης και στον πολεοδομικό σχεδιασμό.

ΚΕΦΑΛΑΙΟ 2

Θεωρητικό Υπόβαθρο

Στο παρόν κεφάλαιο γίνεται πλήρης ανάλυση της τεχνολογίας που χρησιμοποιήθηκε για την παραγωγή κι αξιολόγηση του προτεινόμενου από εμάς μοντέλου Συνελικτικού Νευρωνικού Δικτύου. Αρχικά, αναλύεται η δομή των Τεχνητών Νευρωνικών Δικτύων, ο τρόπος σύνδεσης κι εκπαίδευσής τους και το πεδίο εφαρμογών τους. Στη συνέχεια, προχωράμε στην ανάλυση των Συνελικτικών Νευρωνικών Δικτύων, όπου γίνεται εκτενής αναφορά στην αρχιτεκτονική, στην τροφοδότηση, στον τρόπο αξιολόγησής κι στις διαδικασίες βελτιστοποίησής τους.

2.1 Τεχνητά Νευρωνικά Δίκτυα

Τα *Τεχνητά Νευρωνικά Δίκτυα* (Artificial Neural Networks) αποτελούν το θεμέλιο της Τεχνητής Νοημοσύνης και λύνουν προβλήματα που έχουν αποδειχθεί αδύνατα ή δύσκολα βάσει των δεδομένων στατιστικής. Η ανάπτυξή τους βασίζεται στη λειτουργία του ανθρώπινου εγκεφάλου, απ' όπου προκύπτει κι η ονομασία τους, ωστόσο σχετίζονται αμυδρά με τους βιολογικούς ομολόγους τους. Τα Τεχνητά Νευρωνικά Δίκτυα δεν προσεγγίζουν την πολυπλοκότητα του εγκεφάλου, αλλά παρουσιάζουν δύο βασικές ομοιότητες με τα Βιολογικά Νευρωνικά Δίκτυα. Πρώτον, τα δομικά στοιχεία και των δύο δικτύων είναι απλές υπολογιστικές συσκευές που είναι βαθιά διασυνδεδεμένες. Δεύτερον, οι συνδέσεις μεταξύ των δικτύων καθορίζουν τη λειτουργικότητα του δικτύου. [15]

Η συμπεριφορά ενός Τεχνητού Νευρωνικού Δικτύου καθορίζεται πλήρως από τη δυναμική των συνδέσεων, καθώς κι από το πως το *δεδομένο εισόδου* (input) μετατρέπεται σε *δεδομένο εξόδου* (output). Σε αντίθεση με ένα λεπτομερές βιολογικό μοντέλο, χρησιμοποιεί βασικές πράξεις πινάκων και μη γραμμικούς μετασχηματισμούς ως θεμελιώδεις πράξεις. Στην πιο βασική του υλοποίηση, ένας τεχνητός *νευρώνας* (neuron) απλά μετατρέπει το δεδομένο εισόδου του x σε μία απόκριση y μέσω

μια *συνάρτησης ενεργοποίησης* (activation function) f . Η συνάρτηση ενεργοποίησης εφαρμόζεται σε μια ενεργοποίηση εισόδου, η οποία συνήθως θεωρείται ως το εσωτερικό γινόμενο μεταξύ του δεδομένου εισόδου x και των παραμέτρων (διάνυσμα βάρους) w του τεχνητού νευρώνα, που αποδίδεται με τη μορφή

$$y = f(w^T x). \quad (2.1)$$

Συνδέοντας πολλαπλούς νευρώνες δημιουργείται ένα Νευρωνικό Δίκτυο το οποίο εφαρμόζει μία μη γραμμική συνάρτηση $y = f(x; \theta)$, όπου οι f_i είναι μη γραμμικοί μετασχηματισμοί και το θ αποτελεί τις παραμέτρους του δικτύου, δηλαδή τα διανύσματα βάρους. Μετά την εκπαίδευση ενός Νευρωνικού Δικτύου, οι αναπαραστάσεις κωδικοποιούνται με κατανομημένο τρόπο σαν ένα πρότυπο που πραγματοποιείται σε όλους τους νευρώνες. [34]

2.1.1 Αντίληπτρο

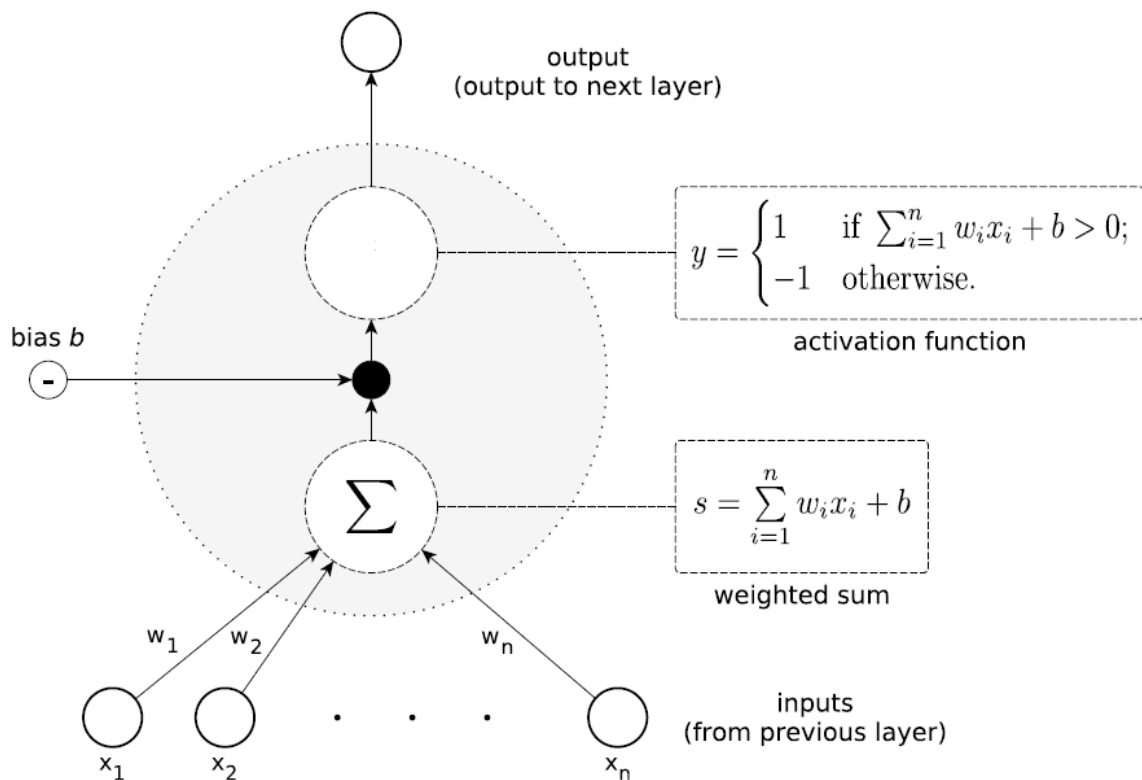
Ο πιο βασικός τύπος Τεχνητού Νευρωνικού Δικτύου είναι το *αντίληπτρο* (perceptron). Τα αντίληπτρα αποτελούνται από έναν αριθμό εξωτερικών συνδέσμων εισόδου, ένα *κατώφλι* (threshold) και ένα μονό εξωτερικό σύνδεσμο εξόδου. Επίσης, έχουν ένα εσωτερικό δεδομένο εισόδου b , το οποίο ονομάζεται *πόλωση* (bias). Το αντίληπτρο δέχεται ένα διάνυσμα πραγματικών τιμών, οι οποίες σταθμίζονται με έναν πολλαπλασιαστή. Αυτά τα βάρη τα μαθαίνει μέσω της *εκπαίδευσης* (training). Για να ενεργοποιηθεί το αντίληπτρο θα πρέπει το δεδομένο εξόδου να ισούται με 1. Η τιμή απενεργοποίησής του ισούται με -1 και στις περισσότερες περιπτώσεις, η τιμή του κατωφλιού του ισούται με 0.

Θεωρούμε διάνυσμα εισόδου $x = \langle x_1, \dots, x_n \rangle$, τα εκπαιδευμένα βάρη W_1, \dots, W_n και τα δεδομένα εξόδου y , που υπολογίζονται από τον τύπο

$$y = \begin{cases} 1, & \text{αν } \sum_{i=1}^n W_i x_i + b > 0 \\ -1, & \text{για τις υπόλοιπες περιπτώσεις.} \end{cases} \quad (2.2)$$

Ονομάζουμε το $z = \sum_{i=1}^n W_i x_i + b$, ως σταθμισμένο δεδομένο εισόδου και ως κατάσταση (state) του αντιλήπτρου το $s = z + b$. Για την ενεργοποίηση του αντιλήπτρου, πρέπει το s να ξεπερνά την τιμή του κατωφλιού.

Τα μεμονωμένα αντιλήπτρα περιορίζονται στην εκμάθηση γραμμικά διαχωρίσιμων συναρτήσεων. Γενικά, ένα πρόβλημα θεωρείται γραμμικό κι οι κλάσεις του διαχωρίζονται γραμμικά σε ένα n -διάστατο χώρο, εάν το όριο αποφάσεων είναι ένα $(n-1)$ -διάστατο υπερεπίπεδο (hyperplane). [31]

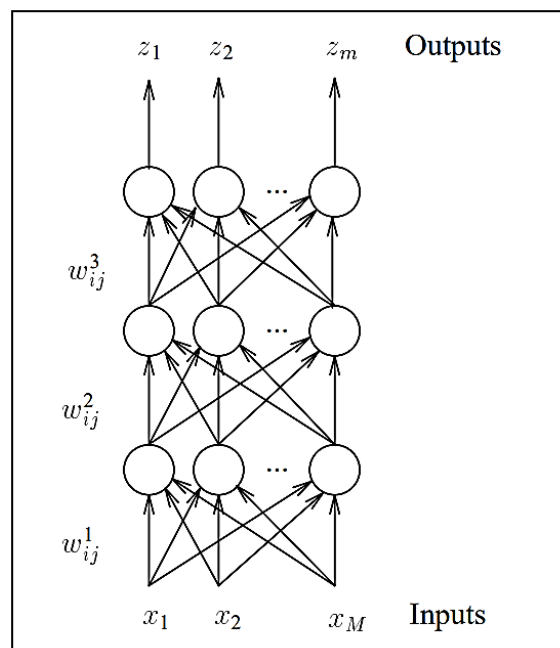


Σχήμα 2. 1: Η δομή του αντιλήπτρου. [31]

2.1.2 Πολυεπίπεδο Αντίληπτρο

Ένα *πολυεπίπεδο αντίληπτρο* (multilayer perceptron) είναι μια παραλλαγή του αρχικού αντίληπτρου. Έχει ένα ή περισσότερα *κρυφά επίπεδα* (hidden layers) ανάμεσα στα δεδομένα εισόδου κι εξόδου του, οι νευρώνες είναι οργανωμένοι σε *επίπεδα* (layers) κι οι συνδέσεις έχουν πάντα κατεύθυνση από τα χαμηλότερα προς τα υψηλότερα επίπεδα.

Το πολυεπίπεδο αντίληπτρο μπορεί να χρησιμοποιηθεί για την διαδικασία της ταξινόμησης. Κάθε νευρώνας στο δίκτυο εκτελεί μια απλή λειτουργία. Το δεδομένο εισόδου του δικτύου αποτελεί ένα διάνυσμα χαρακτηριστικών που έχει εξαχθεί από το αντικείμενο που πρόκειται να ταξινομηθεί και το δεδομένο εξόδου είναι ένα κωδικοποιημένο τμήμα του οποίου ένα στοιχείο εξάγεται με υψηλή τιμή υποδηλώνοντας μία κλάση κι όλα τα υπόλοιπα με χαμηλές τιμές. Τα βάρη που συνδέουν τους νευρώνες καθορίζονται από ένα σύνολο διανυσματικών χαρακτηριστικών που αποτελούν το *σύνολο εκπαίδευσης* (training set). [25]



Σχήμα 2. 2: Πολυεπίπεδο Αντίληπτρο. Οι εκθέτες χρησιμοποιούνται για να δηλώσουν τα επίπεδα. Άρα το x_i^j είναι το δεδομένο εξόδου του νευρώνα i στο επίπεδο j και το w_{ij}^k είναι το βάρος που συνδέει τον νευρώνα i στο επίπεδο $k - 1$ με τον νευρώνα j στο επίπεδο k . Το επίπεδο 1 αποτελεί το πρώτο κρυφό επίπεδο και τα δεδομένα εισόδου μπορούν να θεωρηθούν ως το επίπεδο 0. [25]

Για την εκπαίδευσή τους τα πολυεπίπεδα αντίληπτρα χρησιμοποιούν έναν αλγόριθμο Επιβλεπόμενης Μάθησης που ονομάζεται *Οπισθοδιάδοση* (Backpropagation).

2.1.2.1 Επιβλεπόμενη Μάθηση

Σε ένα πολυεπίπεδο δίκτυο που χρησιμοποιείται η Επιβλεπόμενη Μάθηση, τα δεδομένα εισόδου μπορούν να καταγραφούν σε μια εσωτερική αναπαράσταση και τα δεδομένα εξόδου που παράγονται αποτελούν την αντιπροσώπευση αυτής της εσωτερικής αναπαράστασης αντί του αρχικού μοτίβου. Στο δίκτυο παρέχεται ένα σύνολο παραδειγμάτων αποτελούμενο από ζεύγη δεδομένων εισόδου-εξόδου (δηλαδή, το σύνολο εκπαίδευσης) και έναν τροποποιημένο πίνακα βαρών, ώστε να προσεγγιστεί η συνάρτηση από την οποία το σύνολο εκπαίδευσης γενίκευσε ή παρήγαγε εύλογα αποτελέσματα για το σύνολο εισόδου με το οποίο τροφοδοτήθηκε. Επί του παρόντος, η πιο διαδεδομένη τεχνική για την επίτευξη αυτού του τύπου μάθησης σε ένα Τεχνητό Νευρωνικό Δίκτυο είναι το πολυεπίπεδο αντίληπτρο που χρησιμοποιεί τον αλγόριθμο Οπισθοδιάδοσης. [5]

2.1.2.2 Οπισθοδιάδοση

Ο αλγόριθμος Οπισθοδιάδοσης είναι μια μέθοδος παρακολούθησης της μάθησης. Χρησιμοποιεί τις μεθόδους του μέσου τετραγωνικού σφάλματος και της *Κλίσης Καθόδου* (Gradient Descent) για να πραγματοποιήσει την τροποποίηση στο βάρος σύνδεσης του δικτύου. Η τροποποίηση στο βάρος σύνδεσης του δικτύου στοχεύει στην επίτευξη του ελάχιστου αθροίσματος σφαλμάτων των τετραγώνων. Σε αυτόν τον αλγόριθμο, πρώτα δίνεται μια μικρή τιμή στην τιμή σύνδεσης του δικτύου και στη συνέχεια, επιλέγεται ένα εκπαιδευτικό δείγμα για τον υπολογισμό της διαβάθμισης του σφάλματος σε σχέση με αυτό το δείγμα.

Η διαδικασία του περιγράφεται με τα εξής δύο βήματα: **Την Εμπρόσθια Διάδοση (Forward Propagation) του λειτουργικό σήματος**: το σήμα εισόδου διαδίδεται από το επίπεδο εισόδου, μέσω

του κρυφού επιπέδου, στο επίπεδο εξόδου. Κατά τη διάρκεια της Εμπρόσθιας Διάδοσης του λειτουργικού σήματος, η τιμή του βάρους και η τιμή της αντιστάθμισης διατηρούνται σταθερές και η κατάσταση κάθε επιπέδου του νευρώνα θα ασκήσει επίδραση μόνο στο επόμενο επίπεδο του νευρώνα. Στην περίπτωση που το αναμενόμενο δεδομένο εξόδου δεν επιτευχθεί στο επίπεδο εξόδου, μπορεί να αλλάξει στην Οπισθοδιάδοση του σήματος σφάλματος.

Την Οπισθοδιάδοση του σήματος σφάλματος: η διαφορά μεταξύ του πραγματικού δεδομένου εξόδου με του αναμενόμενου δεδομένου εξόδου του δικτύου ορίζεται ως το σήμα σφάλματος. Στην Οπισθοδιάδοση του σήματος σφάλματος, αυτό διαδίδεται από την άκρη του δεδομένου εξόδου στο επίπεδο εισόδου μέσω του τρόπου επίπεδο-προς-επίπεδο. Η τιμή του βάρους του δικτύου ρυθμίζεται από το σφάλμα ανατροφοδότησης. Η συνεχής τροποποίηση της τιμής βάρους και της τιμής αντιστάθμισης εφαρμόζεται ώστε να πλησιάσει όσο το δυνατόν πιο κοντά το δεδομένου εξόδου του δικτύου το αναμενόμενο.

Η ιδεολογία που καθοδηγεί τους κανόνες μάθησης του αλγορίθμου της Οπισθοδιάδοσης είναι ότι η τροποποίηση της τιμής βάρους και της τιμής κατωφλίου του δικτύου πρέπει να γίνεται κατά την αρνητική κατεύθυνση κλίσης που αντικατοπτρίζει την ταχύτερη πτώση της λειτουργίας.

$$x_{k+1} = x_k - \eta_k g_k, \quad (2.3)$$

στον παραπάνω τύπο το x_k αντιπροσωπεύει τον πίνακα της ισχύουσας τιμής βάρους και τιμής κατωφλίου, το g_k αντιπροσωπεύει την κλίση της ισχύουσας συνάρτησης και το η_k αντιπροσωπεύει τον ρυθμό μάθησης.

Για ένα δίκτυο Οπισθοδιάδοσης 3 επιπέδων, υποθέτουμε πως ο νευρώνας του επιπέδου εισόδου είναι x_i , ο νευρώνας του κρυφού επιπέδου είναι y_j και ο νευρώνας του επιπέδου εξόδου είναι z_l . Η τιμή του βάρους του δικτύου μεταξύ του νευρώνα του επιπέδου εισόδου και του νευρώνα του κρυφού δικτύου είναι w_{ji} και η τιμή του βάρους του δικτύου μεταξύ των νευρώνων του κρυφού επιπέδου και του επιπέδου εξόδου είναι v_{lj} . Όταν η αναμενόμενη τιμή του νευρώνα εξόδου είναι t_l , η συνάρτηση ενεργοποίησης είναι $f(\cdot)$. Ο υπολογισμός του τύπου του μοντέλου εκφράζεται ακολούθως:

Εμπρόσθια Διάδοση: δεδομένο εξόδου του υπολογιστικού δικτύου.

Δεδομένο εξόδου του νευρώνα του κρυφού δικτύου

$$y_j = f\left(\sum_i w_{ji}x_i - \theta_j\right) = f(\text{net}_j) \quad (2.4)$$

$$\text{όπου, } \text{net}_j = \sum_i w_{ji}x_i - \theta_j.$$

Υπολογισμός δεδομένου εξόδου του νευρώνα εξόδου

$$z_l = f\left(\sum_j v_{lj}y_j - \theta_l\right) = f(\text{net}_l) \quad (2.5)$$

$$\text{όπου, } \text{net}_l = \sum_j v_{lj}y_j - \theta_l.$$

Σφάλμα του νευρώνα εξόδου

$$E = \frac{1}{2} \sum_l (t_l - z_l)^2 = \frac{1}{2} \sum_l (t_l - f(\sum_j v_{lj}y_j - \theta_l))^2 = \quad (2.6)$$

$$\frac{1}{2} \sum_l (t_l - f(\sum_j v_{lj}f(\sum_i w_{ji}x_i - \theta_j) - \theta_l))^2.$$

Στην Οπισθοδιάδοση υιοθετείται η μέθοδος της κλίσης καθόδου για τη ρύθμιση της τιμής βάρους όλων των επιπέδων και του αλγορίθμου εκμάθησης της τιμής βάρους. [18]

2.2 Συνελικτικά Νευρωνικά Δίκτυα

Ο όρος *Βαθύ Νευρωνικό Δίκτυο* (Deep Neural Network) αναφέρεται σε Τεχνητά Νευρωνικά Δίκτυα με πολλαπλά επίπεδα. Το ενδιαφέρον για βαθύτερα κρυφά επίπεδα άρχισε πρόσφατα να ξεπερνά την απόδοση των κλασικών μεθόδων σε διαφορετικά πεδία, δείχνοντας ιδιαίτερα εξαιρετική απόδοση σε πολλά προβλήματα Όρασης Υπολογιστή και Μηχανικής Μάθησης. Ένα από τα πιο δημοφιλή Βαθιά Νευρωνικά Δίκτυα είναι το Συνελικτικό Νευρωνικό Δίκτυο.

2.2.1 Αρχιτεκτονική Συνελικτικών Νευρωνικών Δικτύων

Το Συνελικτικό Νευρωνικό Δίκτυο δέχεται ως δεδομένου εισόδου έναν τανυστή 3 τάσεων, παραδείγματος χάριν μια εικόνα με H γραμμές, W στήλες και 3 κανάλια (RGB). Ωστόσο και τανυστές εισόδου υψηλότερων τάσεων, μπορούν να αντιμετωπιστούν με παρόμοιο τρόπο. Στη συνέχεια, το δεδομένο εισόδου περνά από μια σειρά βημάτων επεξεργασίας, τα οποία ονομάζονται επίπεδα.

Μια περιληπτική περιγραφή της δομής των Συνελικτικών Νευρωνικών Δικτύων, αποτελεί η κάτωθι.

$$x^1 \rightarrow \boxed{w^1} \rightarrow x^2 \rightarrow \dots \rightarrow x^{L-1} \rightarrow \boxed{w^{L-1}} \rightarrow x^L \rightarrow \boxed{w^L} \rightarrow z. \quad (2.7)$$

Η παραπάνω εξίσωση απεικονίζει πως το Συνελικτικό Νευρωνικό Δίκτυο τρέχει από επίπεδο σε επίπεδο μια Εμπρόσθια Διάδοση. Το δεδομένο εισόδου x^1 αποτελεί συνήθως μια εικόνα (τανυστής 3 τάσεων). Περνάει μέσω μιας επεξεργασίας στο πρώτο επίπεδο, που είναι το πρώτο κουτί. Συμβολίζουμε τις παραμέτρους που εμπλέκονται στη διαδικασία του πρώτου επιπέδου ως τανυστή w^1 . Το δεδομένο εξόδου του πρώτου επιπέδου είναι το x^2 , το οποίο, επίσης, λειτουργεί κι ως δεδομένο εισόδου στο δεύτερο επίπεδο επεξεργασίας. Η επεξεργασία συνεχίζεται μέχρι όλα τα επίπεδα του Συνελικτικού Νευρωνικού Δικτύου να έχουν ολοκληρωθεί, δίνοντάς μας το δεδομένο εξόδου x^L . Ωστόσο, προστίθεται ένα επιπλέον επίπεδο Οπισθοδιάδοσης σφάλματος.

Ας υποθέσουμε πως το πρόβλημα που έχουμε είναι η ταξινόμηση μιας εικόνας με C κλάσεις. Μια συνηθισμένη στρατηγική είναι να εξάγουμε το x^L ως διάνυσμα διαστάσεων C , όπου η i -οστή είσοδος κωδικοποιεί την πρόβλεψη. Για να κάνουμε το x^L μια συνάρτηση μάζας πιθανότητας, θέτουμε την επεξεργασία στο $(L - 1)$ -οστό επίπεδο ως μια κανονικοποιημένη εκθετική συνάρτηση του x^{L-1} . Σε άλλες εφαρμογές, το x^L μπορεί να έχει άλλες μορφές κι ερμηνείες.

Το τελευταίο επίπεδο είναι επίπεδο σφάλματος. Ας υποθέσουμε ότι t είναι η αντίστοιχη τιμή στόχου για το δεδομένο εισόδου x^1 , τότε μια συνάρτηση κόστους ή σφάλματος μπορεί να χρησιμοποιηθεί για να μετρηθεί η απόκλιση μεταξύ της πρόβλεψης x^L του Συνελικτικού Νευρωνικού Δικτύου και του στόχου t . Για παράδειγμα, μια απλή συνάρτηση σφάλματος που θα μπορούσε να χρησιμοποιηθεί είναι

$$z = \frac{1}{2} \|t - x^L\|^2, \quad (2.8)$$

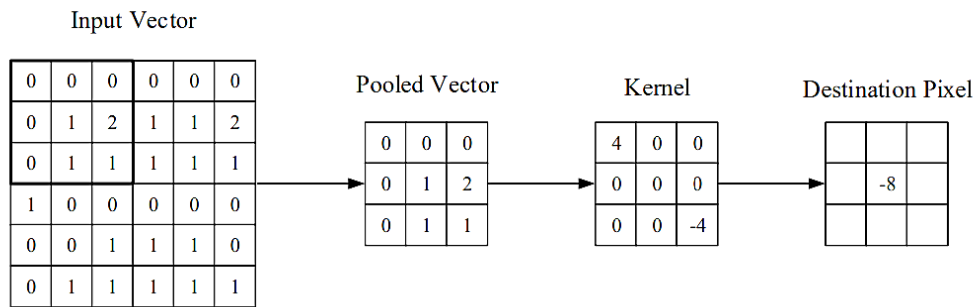
αν και συνήθως χρησιμοποιούνται πιο περίπλοκες συναρτήσεις. Το τετραγωνισμένο σφάλμα $L2$ μπορεί να χρησιμοποιηθεί για προβλήματα παλινδρόμησης. Σε προβλήματα ταξινόμησης χρησιμοποιείται η Σταυροειδής Εντροπία (Cross Entropy). Η αληθής βάση (ground truth) σε ένα πρόβλημα ταξινόμησης αποτελεί μια μεταβλητή κατηγοριοποίησης t . Αρχικά την μετατρέπουμε σε διάνυσμα C , διαστάσεων t . Αμφότερες t και x^L είναι συναρτήσεις μάζας πιθανότητας, και το σφάλμα της Σταυροειδούς Εντροπίας μετράει τη μεταξύ τους απόσταση. Ως εκ τούτου, μπορούμε να ελαχιστοποιήσουμε τη Σταυροειδή Εντροπία. Η εξίσωση (2.8) μοντελοποιεί τη συνάρτηση σφάλματος σε επίπεδο σφάλματος, του οποίου η επεξεργασία μοντελοποιείται ως κουτί με παραμέτρους w^L . Σημειώνεται πως κάποια επίπεδα μπορεί να μην έχουν παραμέτρους, γι' αυτό το w^i μπορεί να είναι κενό για κάποια i . Το επίπεδο της κανονικοποιημένης εκθετικής συνάρτησης αποτελεί ένα τέτοιο παράδειγμα.[35]

2.2.2 Συνελικτικά Επίπεδα

2.2.2.1 Επίπεδο Συνέλιξης

Τα επίπεδα Συνέλιξης έχουν κομβικό ρόλο στη λειτουργία των Συνελικτικών Νευρωνικών Δικτύων. Οι παράμετροι των επιπέδων εστιάζουν στη χρήση *μαθησιακών φίλτρων* (kernels). Αυτά είναι συνήθως μικρά σε χωρική διάσταση, αλλά απλώνονται σε όλο το βάθος του δεδομένου εισόδου. Όταν τα δεδομένα φτάσουν σε ένα επίπεδο Συνέλιξης, το επίπεδο συνελίσσει κάθε φίλτρο σε όλη τη χωρική διάσταση του δεδομένου εισόδου, ώστε να παραχθεί ένας δισδιάστατος χάρτης ενεργοποίησης.

Καθώς περνάμε μέσα από το δεδομένο εισόδου, υπολογίζεται το βαθμωτό γινόμενο για κάθε τιμή στο μαθησιακό φίλτρο. Έτσι, το δίκτυο εκπαιδεύεται ώστε όταν συνάντα ένα συγκεκριμένο χαρακτηριστικό σε μια δοσμένη θέση του δεδομένου εισόδου, να χρησιμοποιεί τα κατάλληλα μαθησιακά φίλτρα. Αυτό ονομάζεται *ενεργοποίηση* (activation).



Σχήμα 2. 3: Μια οπτική αναπαράσταση ενός συνελκτικού επιπέδου. Το κεντρικό στοιχείο του μαθησιακού φίλτρου τοποθετείται πάνω από το διάνυσμα εισόδου, το οποίο στη συνέχεια υπολογίζεται και αντικαθίσταται με ένα σταθμισμένο άθροισμα του εαυτού του και τυχόν γειτονικών εικονοστοιχείων.[22]

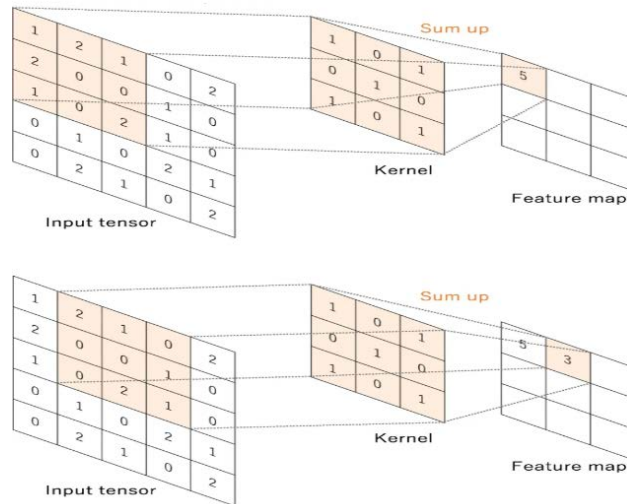
Κάθε μαθησιακό φίλτρο θα έχει έναν αντίστοιχο χάρτη ενεργοποίησης, ο οποίος θα στοιβάζεται κατά μήκος της διάστασης βάθους για να σχηματίσει τον πλήρη όγκο του δεδομένου εξόδου από το συνελκτικό επίπεδο.

Η εκπαίδευση των δικτύων με δεδομένα εισόδου όπως οι εικόνες, οδηγεί σε μοντέλα που είναι πολύ μεγάλα για αποτελεσματική εκπαίδευση. Αυτό ανέρχεται στον πλήρως συνδεδεμένο τρόπο των τυπικών νευρώνων, οπότε για να μετριαστεί αυτό, κάθε νευρώνας σε ένα συνελκτικό επίπεδο συνδέεται μόνο με μικρή περιοχή του όγκου του δεδομένου εισόδου. Η διάσταση αυτής της περιοχής αναφέρεται συνήθως ως το μέγεθος του *δεκτικού πεδίου* (receptive field) του νευρώνα. Το μέγεθος της συνδεσιμότητας μέσω του βάθους είναι σχεδόν πάντα ίσο με το βάθος του δεδομένου εισόδου.

Τα συνελκτικά επίπεδα μπορούν, επίσης, να μειώσουν σημαντικά την πολυπλοκότητα του μοντέλου μέσω της βελτιστοποίησης του δεδομένου εξόδου του, που πραγματοποιείται μέσω τριών υπερπαραμέτρων (hyperparameters), του *βάθους* (depth), του *βήματος* (stride) και της *ρυθμικής επένδυσης* (zero-padding).

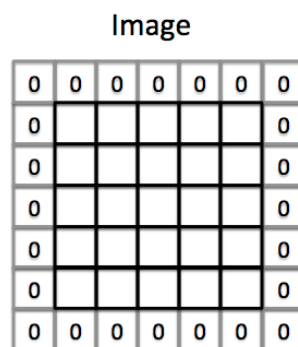
- Το **βάθος** του όγκου του δεδομένου εξόδου που παράγεται από τα συνελκτικά επίπεδα μπορεί να ρυθμιστεί χειροκίνητα μέσω του αριθμού των νευρώνων εντός του επιπέδου στην ίδια περιοχή του δεδομένου εισόδου. Η μείωση αυτής της υπερπαραμέτρου μπορεί να ελαχιστοποιήσει σημαντικά τον συνολικό αριθμό νευρώνων του δικτύου, αλλά μπορεί επίσης να μειώσει σημαντικά τις δυνατότητες αναγνώρισης προτύπων του μοντέλου.

- Το **βήμα** στο οποίο ρυθμίζουμε το βάθος γύρω από τη χωρική διάσταση του δεδομένου εισόδου για να τοποθετήσουμε το δεικτικό πεδίο. Για παράδειγμα, αν θέταμε το βήμα ως 1, τότε θα είχαμε ένα πολύ επικαλυπτόμενο δεικτικό πεδίο που παράγει εξαιρετικά μεγάλες ενεργοποιήσεις. Εναλλακτικά, θέτοντας το βήμα ως μεγαλύτερο αριθμό θα μειώσει την επικάλυψη και θα παράγει ένα δεδομένο εξόδου χαμηλότερων χωρικών διαστάσεων.



Σχήμα 2. 4: Παράδειγμα Συνέλιξης με μέγεθος μαθησιακού φίλτρου 3×3 , χωρίς επένδυση και βήμα 1.[36]

- Η **μηδενική επένδυση** είναι η απλή διαδικασία πλήρωσης του περιγράμματος του δεδομένου εισόδου και είναι μια αποτελεσματική μέθοδος για να δοθεί περαιτέρω έλεγχος ως προς τη διάσταση των όγκων του δεδομένου εξόδου.



Σχήμα 2. 5: Εικόνα με μηδενική επένδυση. [6]

Είναι σημαντικό να κατανοήσουμε ότι με τη χρήση αυτών των τεχνικών, αλλάζει η χωρική διάσταση του δεδομένου εξόδου των συνελικτικών επιπέδων. Γι' αυτόν τον υπολογισμό, χρησιμοποιείται ο ακόλουθος τύπος,

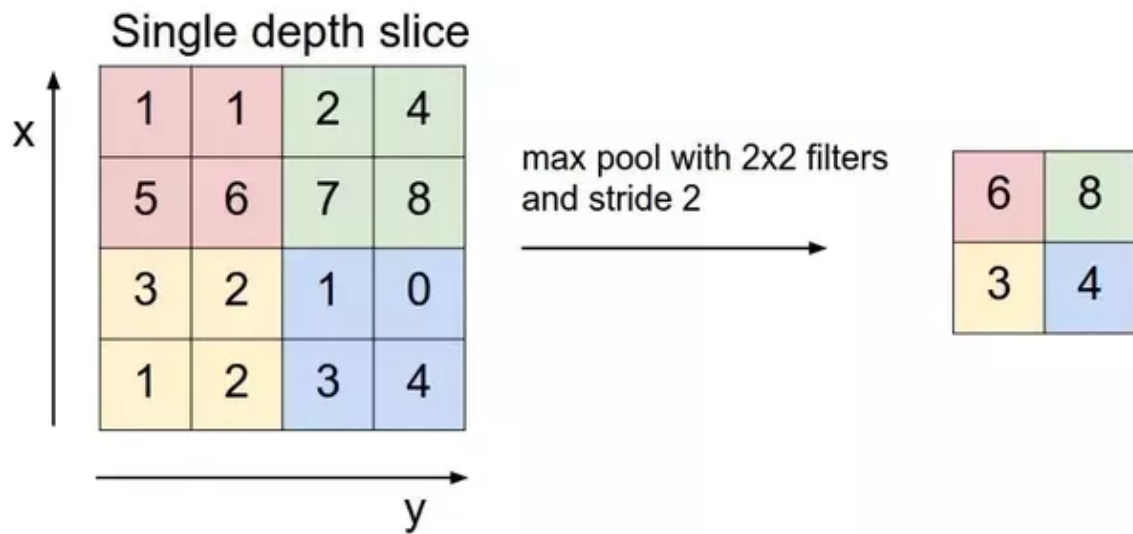
$$\frac{(V-R)+2Z}{S+1}, \quad (2.9)$$

όπου το V αντιπροσωπεύει το μέγεθος του όγκου του δεδομένου εισόδου (ύψος \times πλάτος \times βάθος), το R το δεκτικό πεδίο, το Z είναι η ποσότητα της μηδενικής επένδυσης και το S το βήμα. Εάν το υπολογισμένο αποτέλεσμα από αυτήν την εξίσωση δεν είναι ίσο με θετικό ακέραιο, τότε το βήμα έχει ρυθμιστεί λανθασμένα, καθώς οι νευρώνες δεν θα μπορούν να ταιριάζουν σωστά σε όλο το δεδομένο εισόδου. [22]

2.2.2.2 Επίπεδο Συγκέντρωσης

Τα *επίπεδα συγκέντρωσης* (pooling layers) στοχεύουν στη σταδιακή μείωση της διαστατικότητας της αναπαράστασης, και συνεπώς, στη μείωση του αριθμού των παραμέτρων και της υπολογιστικής πολυπλοκότητας του μοντέλου. Το επίπεδο συγκέντρωσης εφαρμόζεται σε κάθε χάρτη ενεργοποίησης στο δεδομένο εισόδου και κλιμακώνει τη διάστασή του χρησιμοποιώντας τη λειτουργία «MAX». Στα περισσότερα Συνελικτικά Νευρωνικά Δίκτυα, αυτό προκύπτει με τη μορφή Επιπέδων *Μέγιστης Συγκέντρωσης* (Max Pooling) με μαθησιακά φίλτρα διαστάσεων 2×2 που εφαρμόζονται με βήμα 2 κατά μήκος των χωρικών διαστάσεων του δεδομένου εισόδου. Αυτό κλιμακώνει τον χάρτη ενεργοποίησης στο 25% του αρχικού μεγέθους, διατηρώντας παράλληλα τον όγκο του βάθους στο τυπικό του μέγεθος.

Λόγω της καταστροφικής φύσης του επιπέδου συγκέντρωσης, υπάρχουν μόνο δύο γενικά παρατηρούμενες μέθοδοι μέγιστης συγκέντρωσης. Συνήθως, το βήμα και τα φίλτρα των επιπέδων συγκέντρωσης ρυθμίζονται και τα δύο σε 2×2 , τα οποία θα επιτρέψουν στο επίπεδο να επεκταθεί στο σύνολο της χωρικής διάστασης του δεδομένου εισόδου. Επιπλέον, μπορεί να χρησιμοποιηθεί επικάλυψη συγκέντρωσης, όπου το βήμα είναι ρυθμισμένο σε 2 με μέγεθος μαθησιακού φίλτρου 3. Μέγεθος μαθησιακού φίλτρου μεγαλύτερο του 3, μπορεί να μειώσει σημαντικά την απόδοση του μοντέλου. [22]

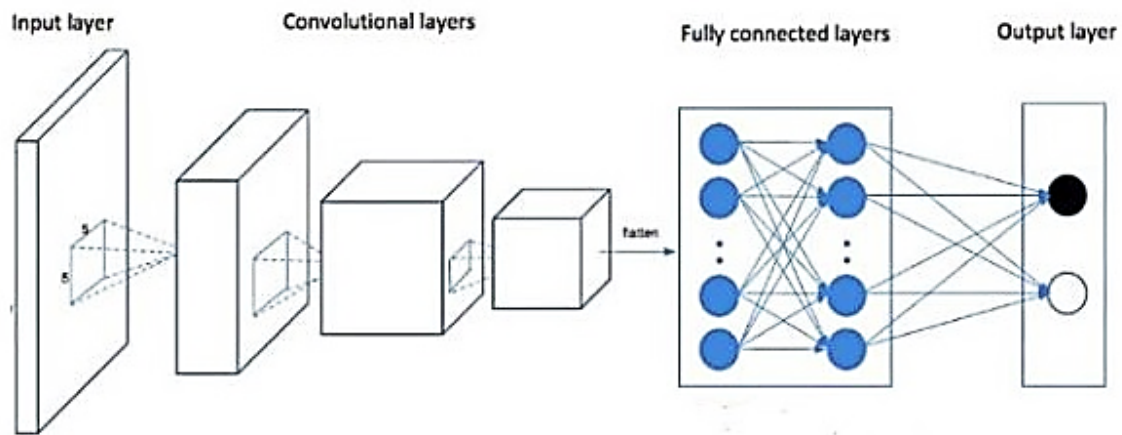


Σχήμα 2. 6: Διαδικασία μέγιστης συγκέντρωσης. [2]

2.2.2.3 Πλήρως Συνδεδεμένο Επίπεδο

Το Πλήρως Συνδεδεμένο επίπεδο (Fully Connected layer) είναι παρόμοιο με τον τρόπο που οι νευρώνες διατάσσονται σε ένα παραδοσιακό νευρωνικό δίκτυο. Επομένως, κάθε νευρώνας σε ένα Πλήρως Συνδεδεμένο επίπεδο συνδέεται άμεσα με κάθε νευρώνα τόσο στο προηγούμενο όσο και στο επόμενο επίπεδο.

Το κύριο μειονέκτημα ενός Πλήρως Συνδεδεμένου επιπέδου, είναι ότι περιλαμβάνει πολλές παραμέτρους που χρειάζονται πολύπλοκους υπολογισμούς κατά τη διαδικασία εκπαίδευσης. [2]



Σχήμα 2. 7: Πλήρως συνδεδεμένα επίπεδα σε ένα Συνελικτικό Νευρωνικό Δίκτυο. [23]

2.2.2.4 Εξειδικευμένα Επίπεδα

Στην παρούσα διπλωματική γίνεται χρήση των κάτωθι εξειδικευμένων επιπέδων:

- **Επίπεδο Αποσυγκέντρωσης (Unpooling Layer):** Κατά τη διάρκεια της λειτουργίας του επιπέδου συγκέντρωσης, δημιουργείται ένας πίνακας που καταγράφει τη θέση της μέγιστης τιμής. Η λειτουργία Αποσυγκέντρωσης θα εισαγάγει την ομαδοποιημένη τιμή στην αρχική θέση, με τα υπόλοιπα στοιχεία να μηδενίζονται. Η Αποσυγκέντρωση καταγράφει παραδείγματα συγκεκριμένων δομών εντοπίζοντας τις αρχικές τοποθεσίες με ισχυρές ενεργοποιήσεις πίσω στο χώρο της εικόνας. Ως αποτέλεσμα, ανακατασκευάζει αποτελεσματικά τη λεπτομερή δομή.
- **Επίπεδο Μετασχηματισμού (Transposed Layer):** Οι συνελίξεις Μετασχηματισμού λειτουργούν με εναλλαγή των εμπρός και πίσω διαδόσεων μιας Συνέλιξης. Το μαθησιακό φίλτρο καθορίζει μια Συνέλιξη, αλλά αν είναι μια άμεση Συνέλιξη ή μια Συνέλιξη Μετασχηματισμού καθορίζεται από τον τρόπο υπολογισμού των εμπρός και προς τα πίσω διαδόσεων. Είναι πάντοτε δυνατό να μιμηθούμε μια Συνέλιξη Μετασχηματισμού με μια άμεση Συνέλιξη. Το μειονέκτημα είναι ότι συνήθως περιλαμβάνει την προσθήκη πολλών στηλών και γραμμών μηδενικών στο δεδομένο εισόδου, με αποτέλεσμα μια πολύ λιγότερο

αποτελεσματική εφαρμογή. [7]

- **Επίπεδο Ανάστροφης Συνέλιξης (Deconvolutional Layer):** Η Ανάστροφη Συνέλιξη αναστρέφει τη διαδικασία μια Συνέλιξης. Τα στρώματα της Ανάστροφης Συνέλιξης πυκνοποιούν τις αραιές ενεργοποιήσεις. Το δεδομένο εξόδου του επιπέδου Ανάστροφης Συνέλιξης είναι ένας διευρυμένος και πυκνός χάρτης ενεργοποίησης. Η ενεργοποίηση σχετίζεται στενά με τις στοχευμένες κλάσεις που ενισχύονται, ενώ οι θορυβώδεις ενεργοποιήσεις από άλλες περιοχές καταστέλλονται αποτελεσματικά. Μια σημαντική λειτουργία του επιπέδου είναι η συνδεσιμότητα μεταξύ των τιμών των δεδομένων εισόδου και εξόδου.

2.3 Εκπαίδευση

2.3.1 Συναρτήσεις Ενεργοποίησης

Οι συναρτήσεις ενεργοποίησης είναι συναρτήσεις που χρησιμοποιούνται σε νευρωνικά δίκτυα για τον υπολογισμό του σταθμισμένου αθροίσματος του δεδομένου εισόδου και των πολώσεων, από τη χρήση των οποίων αποφασίζεται εάν ένας νευρώνας μπορεί να ενεργοποιηθεί ή όχι. Χειρίζεται τα δεδομένα που παρουσιάζονται μέσω κάποιας επεξεργασίας καθόδου, συνήθως της κλίσης καθόδου και στη συνέχεια παράγει ένα δεδομένο εξόδου για το νευρικό δίκτυο, το οποίο περιέχει τις παραμέτρους στα δεδομένα.

2.3.1.1 Συνάρτηση Διορθωμένης Γραμμικής Μονάδας

Η συνάρτηση ενεργοποίησης της *Διορθωμένης Γραμμικής Μονάδας* (Rectified Linear Unit) προτάθηκε από τους Nair και Hinton το 2010, και έκτοτε, είναι η πιο ευρέως χρησιμοποιούμενη συνάρτηση ενεργοποίησης για εφαρμογές Βαθιάς Μάθησης με σύγχρονα αποτελέσματα μέχρι και σήμερα. Είναι

μια συνάρτηση ενεργοποίησης που προσφέρει ταχύτερη μάθηση, καλύτερη απόδοση και γενίκευση στη Βαθιά Μάθηση σε σύγκριση με άλλες. Αντιπροσωπεύει μια σχεδόν γραμμική συνάρτηση και επομένως, διατηρεί τις ιδιότητες των γραμμικών μοντέλων που διευκολύνουν τη βελτιστοποίηση, με μεθόδους διαβάθμισης.

Η συνάρτηση ενεργοποίησης της διορθωμένης γραμμικής μονάδας, εκτελεί μια λειτουργία κατωφλίου σε κάθε δεδομένο εισόδου, όπου οι τιμές μικρότερες από το μηδέν ρυθμίζονται στο μηδέν,

$$f(x) = \max(0, x) = \begin{cases} x_i, & \text{για } x_i \geq 0 \\ 0, & \text{για } x_i < 0. \end{cases} \quad (2.10)$$

Αυτή η συνάρτηση διορθώνει τις τιμές των δεδομένων εισόδου που είναι μικρότερες από το μηδέν, αναγκάζοντάς τις να μηδενιστούν και εξαλείφοντας το πρόβλημα της κλίσης καθόδου. Η συνάρτηση έχει χρησιμοποιηθεί στα κρυφά επίπεδα των Βαθιών Νευρωνικών Δικτύων με άλλες συναρτήσεις ενεργοποίησης και χρησιμοποιείται στα επίπεδα εξόδου των δικτύων. Τυπικά παραδείγματα βρίσκονται σε εφαρμογές ταξινόμησης και αναγνώρισης ομιλίας.

Το κύριο πλεονέκτημα της χρήσης της Διορθωμένης Γραμμικής Μονάδας στον υπολογισμό είναι ότι εγγυάται ταχύτερο υπολογισμό, καθώς δεν υπολογίζει εκθετικά ή με διαιρέσεις, με συνολικά βελτιωμένη ταχύτητα υπολογισμού. Μια άλλη ιδιότητά της είναι ότι εισάγει την αδράνεια στα κρυφά επίπεδα, καθώς συμπιέζει τις τιμές μεταξύ μηδέν και μέγιστου. Ωστόσο, έχει δύο περιορισμούς. Πρώτον, υπερπροσαρμόζεται εύκολα σε σύγκριση με τη λειτουργία της σιγμοειδούς, αν και η τεχνική *Απόσυρσης* (Dropout) έχει υιοθετηθεί για τη μείωση της επίδρασης της *Υπερπροσαρμογής* (Overfitting). Δεύτερον, είναι εύθραυστη κατά τη διάρκεια της εκπαίδευσης, εξαλείφοντας κάποιες κλίσεις. Αυτό οδηγεί και στην εξάλειψη ορισμένων νευρώνων, με αποτέλεσμα να μην ενημερώνονται τα βάρη. [21]

2.3.1.2 Κανονικοποιημένη Εκθετική Συνάρτηση

Η *Κανονικοποιημένη Εκθετική Συνάρτηση* (Softmax) είναι ένας άλλος τύπος συνάρτησης ενεργοποίησης που χρησιμοποιείται στον νευρωνικό υπολογισμό. Χρησιμοποιείται για τον υπολογισμό της κατανομής πιθανότητας από ένα διάνυσμα πραγματικών αριθμών. Παράγει ένα δεδομένο εξόδου, το οποίο είναι μια περιοχή τιμών μεταξύ 0 και 1, με το άθροισμα των πιθανοτήτων να είναι ίσο με 1. Η Κανονικοποιημένη Εκθετική Συνάρτηση υπολογίζεται χρησιμοποιώντας τη σχέση,

$$f(x_i) = \frac{\exp(x_i)}{\sum_j \exp(x_j)}. \quad (2.11)$$

Χρησιμοποιείται σε μοντέλα πολλών κλάσεων, όπου επιστρέφει τις πιθανότητες κάθε κλάσης, με την κλάση στόχου να έχει την υψηλότερη πιθανότητα. [21]

2.3.2 Συνάρτηση Σφάλματος

Τα εποπτευόμενα Βαθιά Νευρωνικά Δίκτυα συχνά απαιτούν μεγάλο αριθμό εκπαιδευτικών δειγμάτων για να επιτύχουν υψηλό επίπεδο απόδοσης. Επίσης, παρόλο που είναι δυνατός ο σχολιασμός μεγάλης κλίμακας, κάποιο σφάλμα κατά τη διαδικασία επισημάνσης είναι συχνά αναπόφευκτο και τα λανθασμένα δείγματα μπορούν να επηρεάσουν την απόδοση των μοντέλων που έχουν εκπαιδευτεί σε αυτά τα δεδομένα. Για να δούμε αν πράγματι βελτιώνεται η εκπαίδευση κι οι πραγματικές τιμές βρίσκονται κοντά με τις τιμές εξόδου, χρησιμοποιούμε τις *συναρτήσεις σφάλματος* (loss functions).

2.3.2.1 Συνάρτηση Σταυροειδούς Εντροπίας

Η *συνάρτηση σφάλματος Σταυροειδούς Εντροπίας* (Cross Entropy loss) χρησιμοποιείται για προβλήματα ταξινόμησης και είναι δημοφιλής λόγω των εξαιρετικών ταχυτήτων σύγκλισης. Είναι ένα βαθμονομημένο σφάλμα, και παρέχει καλές εκτιμήσεις πιθανότητας. Χρησιμοποιείται, συνήθως, μετά από ένα επίπεδο Κανονικοποιημένης Εκθετικής Συνάρτησης που ομαλοποιεί την έξοδο του δικτύου. Η Σταυροειδής Εντροπία μεταξύ δύο διακριτών κατανομών $\mathbf{p} \in [0,1]^N$ και $\mathbf{q} \in (0,1]^N$ και μεγέθους N είναι:

$$H(\mathbf{p}, \mathbf{q}) = - \sum_{i=1}^N p_i \log q_i, \quad (2.12)$$

όπου, το \mathbf{p} αντιστοιχεί στον στόχο της ταξινόμησης και το \mathbf{q} στο δεδομένο εξόδο του επιπέδου της Κανονικοποιημένης Εκθετικής Συνάρτησης, δηλαδή είναι η προβλεπόμενη πιθανότητα ανά κλάση.

Είναι σημαντικό να σημειωθεί πως η πραγματική τιμή της συνάρτησης σφάλματος δεν επηρεάζει τη διαδικασία εκπαίδευσης, καθώς μόνο η κλίση της χρησιμοποιείται κατά την Οπισθοδιάδοση. [20]

2.3.3 Αλγόριθμοι Βελτιστοποίησης

Για ένα πρόβλημα Βαθιάς Μάθησης, πρώτα ορίζουμε μια συνάρτηση σφάλματος. Μόλις έχουμε τη συνάρτηση, μπορούμε να χρησιμοποιήσουμε έναν *αλγόριθμο βελτιστοποίησης* (optimization algorithm) για να ελαχιστοποιήσουμε το σφάλμα. Οι *βελτιστοποιητές* (optimizers) είναι αλγόριθμοι ή μέθοδοι που χρησιμοποιούνται για την αλλαγή των χαρακτηριστικών του νευρικού δικτύου, όπως βάρη και ρυθμός μάθησης, προκειμένου να μειωθούν τα σφάλματα.

2.3.3.1 Αλγόριθμος Στοχαστικής Κλίσης Καθόδου

Ο αλγόριθμος *Στοχαστικής Κλίσης Καθόδου* (Stochastic Gradient Descent) αντί να υπολογίζει ακριβώς την κλίση του $En(fw)$, κάθε επανάληψη εκτιμά αυτήν την κλίση σύμφωνα με ένα μόνο τυχαία επιλεγμένο παράδειγμα z_t :

$$w_{t+1} = w_t - \gamma_t \nabla_w Q(z_t, w_t). \quad (2.13)$$

Η στοχαστική διαδικασία $\{w_t, t = 1, \dots\}$ εξαρτάται από τα παραδείγματα που επιλέγονται τυχαία σε κάθε επανάληψη.

Δεδομένου ότι ο στοχαστικός αλγόριθμος δεν χρειάζεται να θυμάται ποια παραδείγματα χρησιμοποιήθηκαν κατά τις προηγούμενες επαναλήψεις, μπορεί να επεξεργαστεί παραδείγματα εν κινήσει σε ένα αναπτυσσόμενο σύστημα. Σε μια τέτοια κατάσταση, η Στοχαστική Κλίση Καθόδου βελτιστοποιεί άμεσα τον αναμενόμενο κίνδυνο, καθώς τα παραδείγματα αντλούνται τυχαία από την αληθή βάση του εδάφους.

Η ταχύτητα σύγκλισης της Στοχαστικής Κλίσης Καθόδου περιορίζεται στην πραγματικότητα από τη

θορυβώδη προσέγγιση της πραγματικής κλίσης. Όταν οι απολαβές μειώνονται πολύ αργά, η διακύμανση της παραμέτρου εκτίμησης w_t μειώνεται εξίσου αργά. Όταν οι απολαβές μειώνονται πολύ γρήγορα, η προσδοκία της παραμέτρου w_t εκτιμάται με πολύ χρόνο για να προσεγγίσει το βέλτιστο. Υπό επαρκείς συνθήκες τακτοποίησης, η καλύτερη ταχύτητα σύγκλισης επιτυγχάνεται χρησιμοποιώντας απολαβές $\gamma_t \sim t^{-1}$. Η προσδοκία του υπολειπόμενου σφάλματος στη συνέχεια μειώνεται με παρόμοια ταχύτητα, $E \rho \sim t^{-1}$. [4]

Μια πετυχημένη επέκταση της Στοχαστικής Κλίσης Καθόδου γίνεται με την προσθήκη της ορμής (momentum). Η ορμή είναι ένα ισχυρό ευρετικό για την επιτάχυνση της σύγκλισης των μεθόδων βελτιστοποίησης. Η Στοχαστική Κλίση Καθόδου μαζί με την ορμή, λειτουργεί ως εξής:

$$m_t = \beta m_{t-1} + (1 - \beta) \nabla F_i(\theta t), \theta_{t+1} = \theta_t - \alpha_t m_t. \quad (2.14)$$

Υπάρχουν δύο τρόποι για να επιτευχθεί καλύτερο ποσοστό σύγκλισης από τη Στοχαστική Κλίση Καθόδου. Πρώτον, αξιοποιώντας κόλπα, όπως τη μείωση διακύμανσης, πιο προηγμένες μέθοδοι βελτιστοποίησης μπορούν να επιτύχουν βελτιωμένο ρυθμό σύγκλισης που συνδυάζει τη θεωρητική βελτίωση τόσο της ορμής όσο και της Στοχαστικής Κλίσης Καθόδου. Δεύτερον, εξετάζοντας περισσότερο τη δομή του προβλήματος, απλούστερες παραλλαγές της Στοχαστικής Κλίσης Καθόδου μπορούν να επιτύχουν επιτάχυνση. [32]

2.3.4 Κανονικοποίηση

Η κανονικοποίηση (normalization) είναι μια τεχνική που εφαρμόζεται συχνά ως μέρος της προετοιμασίας δεδομένων. Ο στόχος της κανονικοποίησης είναι να αλλαχθούν οι τιμές των αριθμητικών στηλών στο σύνολο δεδομένων σε μια κοινή κλίμακα, χωρίς να παραμορφώνονται οι διαφορές στα εύρη τιμών.

2.3.4.1 Κανονικοποίηση Παρτίδας

Ένα από τα προβλήματα που αντιμετωπίζουν τα Βαθιά Νευρωνικά Δίκτυα είναι η αλλαγή της κατανομής των ενεργοποιήσεων των επιπέδων τους κατά την εκπαίδευσή, που οφείλεται στην αλλαγή των παραμέτρων τους, γνωστή ως *εσωτερική μετατόπιση μεταβλητών* (internal covariate shift). Στην εκκίνηση του αλγορίθμου, οι παράμετροι αρχικοποιούνται τυχαία και κατά την εκπαίδευση του δικτύου οι τιμές τους μεταβάλλονται. Μεταβολές προκύπτουν και στις εισακτές τιμές της στατιστικής κατανομής κατά τις επαναλήψεις κι οι αλλαγές στην κατανομή εξόδου του προηγούμενου επιπέδου μπορούν να μετατοπίσουν την κατανομή των εισόδων του επόμενου επιπέδου, με αποτέλεσμα να επιβραδύνεται η εκπαιδευτική διαδικασία, καθώς τα επίπεδα προσπαθούν να προσαρμοστούν στην καινούρια κατανομή.

Για τη σταθεροποίηση του δικτύου, χρησιμοποιούμε την *Κανονικοποίηση Παρτίδας* (Batch Normalization). Αντί να αφαιρέσουμε τη συσχέτιση των χαρακτηριστικών κάθε δεδομένου εισόδου κι εξόδου ανά επίπεδο, την κανονικοποιούμε ξεχωριστά για κάθε μονόμετρο χαρακτηριστικό, θέτοντας τη μέση τιμή 0 και τη διακύμανση 1. Για ένα επίπεδο με d -διάστατο δεδομένο εισόδου $x = (x^{(1)} \dots x^{(d)})$, κανονικοποιούμε κάθε διάσταση,

$$\hat{x}^{(k)} = \frac{x^{(k)} - E[x^{(k)}]}{\sqrt{\text{Var}[x^{(k)}]}}, \quad (2.15)$$

όπου η μέση τιμή κι η διακύμανση υπολογίζονται για όλο το σύνολο εκπαίδευσης. Βέβαια, με αυτήν τη διαδικασία δημιουργείται ο κίνδυνος να αλλάξει η αντιπροσώπευση του επιπέδου, οπότε γίνεται η εισαγωγή ενός ζεύγους παραμέτρων, $\gamma^{(k)}$ και $\beta^{(k)}$, το οποίο σταθμίζει και μετατοπίζει κάθε ενεργοποίηση,

$$y^{(k)} = \gamma^{(k)} \hat{x}^{(k)} + \beta^{(k)}. \quad (2.16)$$

Στην περίπτωση των παρτίδων, όπου κάθε στάδιο εκπαίδευσης βασίζεται στην πληρότητα του συνόλου εκπαίδευσης, δεν είναι πρακτικό να γίνει χρήση ολόκληρου για την κανονικοποίηση των ενεργοποιήσεων, γι' αυτό χρησιμοποιούμε *μίνι-παρτίδες* (mini-batches), όπου κάθε μία παράγει εκτιμήσεις της μέσης τιμής και της διακύμανσης για κάθε ενεργοποίηση. Παρακάτω παρατίθεται ο αλγόριθμος της Κανονικοποίησης Παρτίδων:

Input: Values of x over a mini-batch: $\mathcal{B} = \{x_1 \dots x_m\}$;
Parameters to be learned: γ, β
Output: $\{y_i = \text{BN}_{\gamma, \beta}(x_i)\}$

$$\begin{aligned} \mu_{\mathcal{B}} &\leftarrow \frac{1}{m} \sum_{i=1}^m x_i && // \text{ mini-batch mean} \\ \sigma_{\mathcal{B}}^2 &\leftarrow \frac{1}{m} \sum_{i=1}^m (x_i - \mu_{\mathcal{B}})^2 && // \text{ mini-batch variance} \\ \hat{x}_i &\leftarrow \frac{x_i - \mu_{\mathcal{B}}}{\sqrt{\sigma_{\mathcal{B}}^2 + \epsilon}} && // \text{ normalize} \\ y_i &\leftarrow \gamma \hat{x}_i + \beta \equiv \text{BN}_{\gamma, \beta}(x_i) && // \text{ scale and shift} \end{aligned}$$

Σχήμα 2. 8: Μετατροπή Κανονικοποίησης Παρτίδας, εφαρμοσμένη σε μια ενεργοποίηση x σε μια μίνι-παρτίδα.

Η Κανονικοποίηση Παρτίδας είναι μια παραγωγική μετατροπή που εισάγει κανονικοποιημένες ενεργοποιήσεις στο δίκτυο σταθεροποιώντας τη διαδικασία εκπαίδευσης, επιταχύνοντάς τη μέσω της χρήσης μεγαλύτερων ρυθμών μάθησης και μειώνοντας την Υπερπροσαρμογή, οδηγώντας, έτσι, στην ελάττωση χρήσης επιπέδων Απόσυρσης, βοηθώντας το δίκτυο να διατηρήσει μεγαλύτερη πληροφορία. [14]

2.3.4.2 Κανονικοποίηση Τοπικής Απόκρισης

Πρόκειται για μία κανονικοποίηση που εφαρμόζει την τεχνική της πλάγιας αναστολής (lateral inhibition), όπως συμβαίνει με τους πραγματικούς νευρώνες. Η πλάγια αναστολή αναφέρεται στο φαινόμενο όπου ένας διεγερμένος νευρώνας καταφέρει να περιορίσει τους γειτονικούς του, φτάνοντας, έτσι, σε μια κορύφωση υπό μορφή τοπικού μεγίστου, με αποτέλεσμα να δημιουργείται αντίθεση κι αυτή με τη σειρά της να προκαλεί αύξηση της αισθητηριακής αντίληψης.

Η κανονικοποίηση μπορεί να συμβεί μέσα στο ίδιο το κανάλι ή σε όλα τα κανάλια. Και οι δύο αυτές μέθοδοι τείνουν να ενισχύουν τον διεγερμένο νευρώνα, ενώ μειώνουν τους περιβάλλοντες νευρώνες. Η Κανονικοποίηση Τοπικής Απόκρισης (Local Response Normalization) δίνεται από την παρακάτω συνάρτηση

$$b^i_{x,y} = \alpha_{x,y}^i \left(k + a \sum_{j=\max(0,i-\frac{n}{2})}^{\min(N-1,i+\frac{n}{2})} (\alpha_{x,y}^j)^2 \right)^\beta, \quad (2.17)$$

όπου $\alpha_{x,y}^i$ η δραστηριότητα του νευρώνα που υπολογίζεται εφαρμόζοντας φίλτρο i στη θέση (x, y) . Οι σταθερές k , a και β είναι υπερπαραμέτροι των οποίων οι τιμές αποφασίζονται μέσω του συνόλου επαλήθευσης. Η Κανονικοποίηση Τοπικής Απόκρισης, αφορά μια κανονικοποίηση στη φωτεινότητα και συμβάλλει στη μείωση των ποσοστών σφάλματος πρόβλεψης. [16]

2.3.5 Υπερπροσαρμογή

Η αρχή της οικονομίας απαιτεί τη χρήση μοντέλων και διαδικασιών που περιέχουν όλα όσα είναι απαραίτητα για τη μοντελοποίηση. Για παράδειγμα, εάν ένα μοντέλο παλινδρόμησης με 2 προβλέψεις είναι αρκετό για να εξηγήσει το πρόβλημα που τίθεται, τότε δεν πρέπει να χρησιμοποιούνται περισσότερες από αυτές τις 2 προβλέψεις. Η Υπερπροσαρμογή είναι η χρήση μοντέλων ή διαδικασιών που παραβιάζουν αυτήν την οικονομία, δηλαδή, περιλαμβάνουν περισσότερους όρους από ό,τι είναι απαραίτητοι ή χρησιμοποιούν πιο περίπλοκες προσεγγίσεις από ό,τι είναι απαραίτητες.

Είναι χρήσιμο να διακρίνουμε δύο τύπους Υπερπροσαρμογής, πρώτον χρησιμοποιώντας ένα μοντέλο που είναι πιο ευέλικτο από ό,τι πρέπει και δεύτερον, η χρήση ενός μοντέλου που περιλαμβάνει άσχετα στοιχεία.

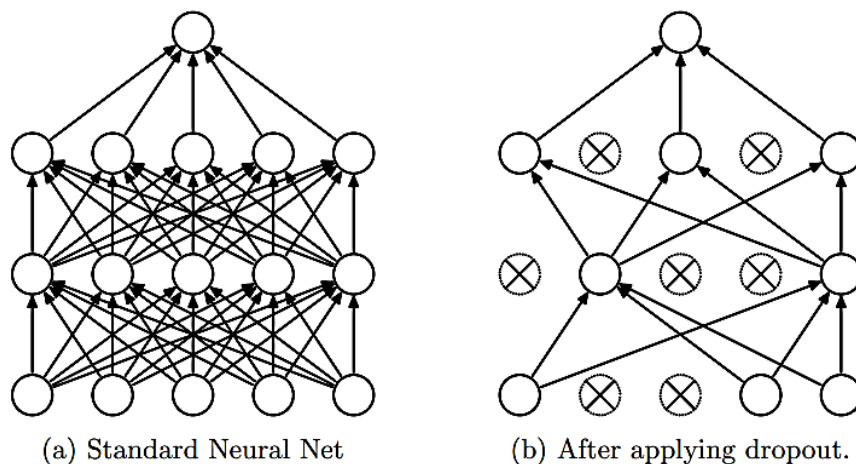
Η Υπερπροσαρμογή είναι ανεπιθύμητη για διάφορους λόγους. Η προσθήκη προβλέψεων που δεν εκτελούν καμία χρήσιμη λειτουργία σημαίνει ότι στη μελλοντική χρήση της παλινδρόμησης για να γίνουν προβλέψεις θα πρέπει να μετρηθούν και να καταγραφούν αυτοί οι προγνωστικοί παράγοντες, ώστε να μπορούν να αντικατασταθούν οι τιμές τους στο μοντέλο. Αυτό όχι μόνο σπαταλά πόρους, αλλά, επίσης, επεκτείνει τις πιθανότητες μη εντοπισμένων σφαλμάτων σε βάσεις δεδομένων για να οδηγήσει σε λάθη πρόβλεψης. Ακόμη, σε ένα πρόβλημα επιλογής χαρακτηριστικών, μοντέλα που περιλαμβάνουν περιττές προβλέψεις οδηγούν σε χειρότερες αποφάσεις. Η προσθήκη άσχετων προβλέψεων μπορεί να επιδεινώσει τις προβλέψεις, επειδή οι συντελεστές που τους προστίθενται προσθέτουν τυχαία παραλλαγή στις επόμενες προβλέψεις. [13]

Για να αποφύγουμε την Υπερπροσαρμογή, χρησιμοποιούμε διάφορες μεθόδους που προσεγγίζουν το

πρόβλημα με διαφορετικό τρόπο, προσφέροντας οφέλη από την άποψη ενός συνδυασμού απόδοσης γενίκευσης, διαμόρφωσης και υπολογιστικής πολυπλοκότητας.

2.3.5.1 Απόσυρση

Τα Νευρωνικά Δίκτυα της Βαθιάς Μάθησης, αντιμετωπίζουν τον κίνδυνο της Υπερπροσαρμογής και για την εξάλειψή του, παρουσιάστηκε η μέθοδος της Απόσυρσης. Αποτελεί μια μέθοδο κατά την οποία τυχαία δεδομένα εξόδου αγνοούνται ή «αποσύρονται», δημιουργώντας την αίσθηση ενός δικτύου όπου κάθε επίπεδο με το επόμενο, έχουν διαφορετικούς νευρώνες και διαφορετική συνδεσιμότητα. Αγνοώντας ή «αποσύροντας» τυχαία δεδομένα εξόδου, εννοούμε την προσωρινή αφαίρεσή τους από το δίκτυο.



Σχήμα 2. 9: Μοντέλο Απόσυρσης. **Αριστερά (a):** Ένα κανονικό νευρωνικό δίκτυο. **Δεξιά (b):** Ένα παράδειγμα αποτελέσματος της χρήσης της Απόσυρσης στο αριστερό μοντέλο. Οι διαγραμμένοι νευρώνες έχουν υποστεί απόσυρση.

[30]

Εισάγεται μια υπερπαράμετρος p που ορίζει είτε την πιθανότητα απόσυρσης των δεδομένων εξόδου των επιπέδων είτε την πιθανότητα διατήρησής τους. Αυτή η πιθανότητα είναι ανεξάρτητη από τα

υπόλοιπα επίπεδα και μπορεί να επιλεγθεί χρησιμοποιώντας ένα σύνολο δεδομένων ή θέτοντας την στο 0.5, όταν πρόκειται για διατήρηση δεδομένων εξόδου σε κρυφά επίπεδα και κοντά στο 1.0, όταν πρόκειται για διατήρηση δεδομένων εισόδου σε ορατά επίπεδα. [30]

2.3.5.2 Ενίσχυση Δεδομένων

Η *Ενίσχυση Δεδομένων* (Data Augmentation) προσεγγίζει το πρόβλημα της Υπερπροσαρμογής από το σύνολο δεδομένων εκπαίδευσης. Αυτό γίνεται υπό την προϋπόθεση ότι περισσότερες πληροφορίες μπορούν να εξαχθούν από το αρχικό σύνολο δεδομένων μέσω ενισχύσεων. Αυτές διογκώνουν τεχνητά το μέγεθος του συνόλου δεδομένων εκπαίδευσης είτε μέσω αναδίπλωσης δεδομένων είτε μέσω *υπερδειγματοληψίας* (upsampling). Οι ενισχύσεις αναδίπλωσης δεδομένων μεταμορφώνουν τις υπάρχουσες εικόνες έτσι ώστε να διατηρείται η ετικέτα τους. Αυτό περιλαμβάνει ενισχύσεις όπως, γεωμετρικούς και χρωματικούς μετασχηματισμούς, τυχαία διαγραφή, ανταγωνιστική εκπαίδευση και μεταφορά νευρικού τρόπου. Οι ενισχύσεις υπερδειγματοληψίας δημιουργούν συνθετικά στοιχεία και τις προσθέτουν στο εκπαιδευτικό σύνολο. Αυτές περιλαμβάνουν τη μίξη εικόνων, τις ενισχύσεις των χαρακτηριστικών του χώρου και τα *Παραγωγικά Ανταγωνιστικά Δίκτυα* (Generative Adversarial Networks - GANS). [27]

2.3.5.3 Μεταφορά Μάθησης

Η *μεταφορά μάθησης* (transfer learning) είναι ένα άλλο ενδιαφέρον παράδειγμα για την αποφυγή της Υπερπροσαρμογής. Λειτουργεί εκπαιδευοντας ένα δίκτυο σε ένα μεγάλο σύνολο δεδομένων και στη συνέχεια χρησιμοποιώντας αυτά τα βάρη ως τα αρχικά βάρη σε ένα νέο πρόβλημα ταξινόμησης. Συνήθως, αντιγράφονται μόνο τα βάρη των συνελκτικών επιπέδων, αντί για ολόκληρο το δίκτυο, συμπεριλαμβανομένων των Πλήρως Συνδεδεμένων επιπέδων. Αυτό είναι πολύ αποτελεσματικό, καθώς πολλά σύνολα δεδομένων εικόνων μοιράζονται χωρικά χαρακτηριστικά χαμηλού επιπέδου που μαθαίνονται καλύτερα με μεγάλα δεδομένα. [27]

2.3.6 Μετρικές Αξιολόγησης

Η προετοιμασία των δεδομένων καθώς και η εκπαίδευση των μοντέλων αποτελούν τις πιο βασικές διαδικασίες της Μηχανικής Μάθησης, ωστόσο, είναι εξίσου σημαντικό να μετριέται κι η απόδοση των εκπαιδευσιμων μοντέλων. Γι' αυτό χρησιμοποιούνται *μετρικές αξιολόγησης* (metrics). Χρησιμοποιώντας διαφορετικές μετρικές για την αξιολόγηση της απόδοσης, χωρίς να εξαρτόμαστε μόνο από την ακρίβεια, είμαστε σε θέση να βελτιώσουμε τη συνολική προβλέψιμη ισχύ του μοντέλου μας προτού το διαθέσουμε για παραγωγή σε δεδομένα που δεν εμφανίζονται.

Προτού παραθέσουμε τις διάφορες μετρικές που χρησιμοποιούνται, κάνουμε μια αναφορά στις βασικές έννοιες για τη μέτρηση της απόδοσης. Οι έννοιες ορίζονται ως προς μία κλάση αναφοράς.

- **Αληθώς Θετική περίπτωση (True Positive – TP):** Ως αληθώς θετική ορίζεται η περίπτωση όπου η πρόβλεψή μας ήταν θετική και το αποτέλεσμα βγήκε θετικό, δηλαδή ένα εικονοστοιχείο ανήκει σε μία κλάση κι αυτό είναι ορθό.
- **Ψευδώς Θετική περίπτωση (False Positive – FP):** Ως ψευδώς θετική ορίζεται η περίπτωση όπου η πρόβλεψή μας ήταν αρνητική και το αποτέλεσμα βγήκε θετικό, δηλαδή ένα εικονοστοιχείο δεν ανήκει σε μία κλάση κι αυτό είναι ορθό.
- **Αληθώς Αρνητική περίπτωση (True Negative – TN):** Ως αληθώς αρνητική ορίζεται η περίπτωση όπου η πρόβλεψή μας ήταν θετική και το αποτέλεσμα βγήκε αρνητικό, δηλαδή ένα εικονοστοιχείο ανήκει σε μία κλάση κι αυτό είναι λανθασμένο.
- **Ψευδώς Αρνητική περίπτωση (False Negative – FN):** Ως ψευδώς αρνητική ορίζεται η περίπτωση όπου η πρόβλεψή μας ήταν αρνητική και το αποτέλεσμα βγήκε αρνητικό, δηλαδή ένα εικονοστοιχείο δεν ανήκει σε μία κλάση κι αυτό είναι λανθασμένο.

Αυτές οι έννοιες μπορούν να συνοψισθούν σε έναν *πίνακα σύγχυσης* (confusion matrix) ώστε να οπτικοποιηθεί η απόδοση του μοντέλου.

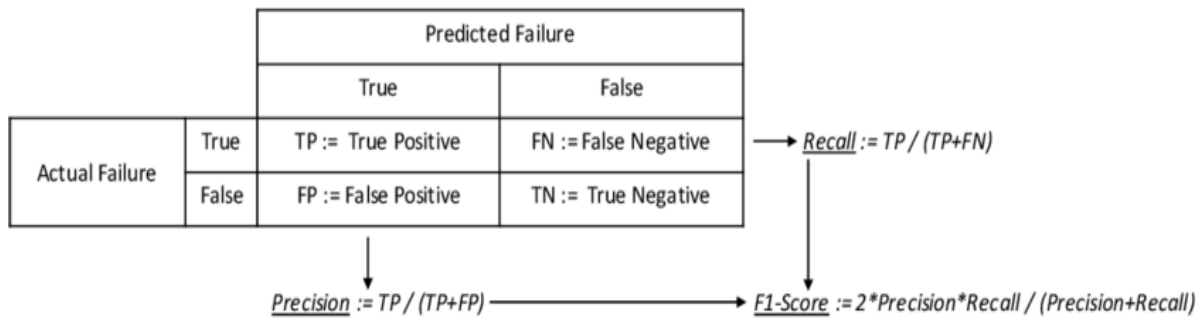
	Actual = Yes	Actual = No
Predicted = Yes	TP	FP
Predicted = No	FN	TN

Σχήμα 2. 10: Πίνακας σύγκρισης δυαδικής ταξινόμησης. [33]

2.3.6.1 Πίνακας Σύγκρισης

Ο πίνακας σύγκρισης αποτελεί μία μέτρηση απόδοσης για το πρόβλημα της ταξινόμησης, όπου τα δεδομένα εξόδου μπορεί να είναι 2 ή περισσότερες κλάσεις. Είναι ένας πίνακας με 4 διαφορετικούς συνδυασμούς μεταξύ των τιμών πρόβλεψης και των πραγματικών τιμών, που επιτρέπει την εύκολη αναγνώριση σύγκρισης μεταξύ κλάσεων, όπως για παράδειγμα το πότε μία κλάση ταξινομήθηκε λανθασμένα ως άλλη.

Ο αριθμός των σωστών και λανθασμένων προβλέψεων συνοψίζεται με τιμές μέτρησης και κατανέμεται ανά κλάση. Ο πίνακας σύγκρισης δείχνει τους τρόπους με τους οποίους το μοντέλο ταξινόμησης μπερδεύεται όταν κάνει προβλέψεις. Μας δίνει μια εικόνα όχι μόνο για τα λάθη που γίνονται από έναν ταξινομητή (classifier), αλλά πιο σημαντικά και για τα είδη των σφαλμάτων. Τα περισσότερα μέτρα απόδοσης υπολογίζονται από τον πίνακα σύγκρισης. [29]



Σχήμα 2. 11: Πίνακας σύγκρισης με οπτικοποίηση εξαγωγής βασικών μέτρων απόδοσης. [29]

2.3.6.2 Ορθότητα

Η *ορθότητα* (accuracy) ταξινόμησης είναι μια μετρική που συνοψίζει την απόδοση ενός μοντέλου ταξινόμησης ως τον αριθμό των σωστών προβλέψεων διαιρούμενο με τον συνολικό αριθμό προβλέψεων. Είναι εύκολο να υπολογιστεί και μπορεί να γίνει κατανοητή διαισθητικά, καθιστώντας την πιο κοινή μετρική που χρησιμοποιείται για την αξιολόγηση μοντέλων ταξινόμησης.

Δίνεται από τον τύπο:

$$Accuracy = \frac{True\ Positives + True\ Negatives}{True\ Positives + True\ Negatives + False\ Positive + False\ Negatives}. \quad (2.18)$$

2.3.6.3 Ακρίβεια

Η *ακρίβεια* (precision) είναι μια μετρική που ποσοτικοποιεί τον αριθμό των σωστών θετικών προβλέψεων. Είναι ο λόγος των σωστά προβλεπόμενων θετικών παραδειγμάτων δια του συνολικού αριθμού των θετικών παραδειγμάτων που είχαν προβλεφθεί.

Υπολογίζεται ως ο αριθμός των αληθώς θετικών διαιρούμενος με τον συνολικό αριθμό των αληθώς θετικών και των ψευδών θετικών. Η ακρίβεια των προβλέψεων είναι της τάξης των μειονοτήτων.

Το αποτέλεσμα της κυμαίνεται μεταξύ της τιμής 0.0 και της τιμής 1.0.

Δίνεται από τον τύπο:

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \quad (2.19)$$

2.3.6.4 Ανάκληση

Η *ανάκληση* (recall) είναι μια μετρική που ποσοτικοποιεί τον αριθμό των σωστών θετικών προβλέψεων που έγιναν από όλες τις θετικές προβλέψεις που θα μπορούσαν να είχαν γίνει. Υπολογίζεται ως ο αριθμός των αληθώς θετικών διαιρούμενος με τον συνολικό αριθμό των αληθώς θετικών και των ψευδώς αρνητικών.

Σε αντίθεση με την ακρίβεια που σχολιάζει μόνο τις σωστές θετικές προβλέψεις από όλες τις θετικές προβλέψεις, η ανάκληση παρέχει μια ένδειξη χαμένων θετικών προβλέψεων. Με αυτόν τον τρόπο, παρέχει κάποια έννοια της κάλυψης της θετικής περίπτωσης.

Το αποτέλεσμα της κυμαίνεται μεταξύ της τιμής 0.0 και της τιμής 1.0.

Δίνεται από τον τύπο:

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \quad (2.20)$$

2.3.6.5 Αρμονικός Μέσος F1

Ο *αρμονικός μέσος F1* (F1-Score) είναι μια μετρική που παρέχει έναν συνδυασμό τόσο της ακρίβειας όσο και της ανάκλησης σε ένα μόνο μέτρο που περιλαμβάνει και τις δύο ποσότητες. Είναι ο αρμονικός

μέσος όρος των δύο κλασμάτων και είναι η πιο κοινή μετρική που χρησιμοποιείται σε μη ισορροπημένες ταξινομήσεις.

Το αποτέλεσμά του κυμαίνεται μεταξύ της τιμής 0.0 και της τιμής 1.0.

Δίνεται από τον τύπο:

$$F1 - Score = \frac{2 * Precision * Recall}{Precision + Recall} \quad (2.21)$$

2.3.6.6 Μέσοι Όροι

Στη μέθοδο *micro μέσου όρου* (micro-average), αθροίζονται τα μεμονωμένα αληθώς θετικά, ψευδώς θετικά και ψευδώς αρνητικά του συστήματος για διαφορετικά σύνολα και τα εφαρμόζουμε συνδυαστικά αναλόγως τη μετρική που θέλουμε να υπολογίσουμε, για ένα σύνολο δεδομένων.

Στη μέθοδο *macro μέσου όρου* (macro-average), παίρνουμε τη μέση τιμή της ακρίβειας και της ανάκλησης των κλάσεων του δικτύου.

Για τον υπολογισμό ενός *σταθμισμένου μέσου όρου* (weighted average), κάθε συνεισφορά τάξης στον μέσο όρο σταθμίζεται από τον σχετικό αριθμό παραδειγμάτων που είναι διαθέσιμα για αυτόν.

Για παράδειγμα, για ένα σύνολο δεδομένων με δύο κλάσεις ο μικρός μέσος όρος υπολογίζεται ως εξής:

$$\text{Micro - average precision} = \frac{\text{True Positives}_1 + \text{True Positives}_2}{\text{True Positives}_1 + \text{True Positives}_2 + \text{False Positives}_1 + \text{False Positives}_2} \quad (2.22)$$

$$\text{Micro – average recall} = \frac{\text{True Positives}_1 + \text{True Positives}_2}{\text{True Positives}_1 + \text{True Positives}_2 + \text{False Negatives}_1 + \text{False Negatives}_2} \quad (2.23)$$

$$\text{Micro – average F1 – Score} = \frac{2 * \text{Micro – average Precision} * \text{Micro – average Recall}}{\text{Micro – average Precision} + \text{Micro – average Recall}} \quad (2.24)$$

2.3.6.7 Ο Συντελεστής Kappa

Ο συντελεστής Kappa είναι μια μετρική που συγκρίνει μια *παρατηρούμενη ορθότητα* (observed accuracy) με μια *αναμενόμενη ορθότητα* (expected accuracy). Χρησιμοποιείται όχι μόνο για την αξιολόγηση ενός μεμονωμένου ταξινομητή, αλλά και για την αξιολόγηση των ταξινομητών μεταξύ τους. Επιπλέον, λαμβάνει υπόψη την τυχαία πιθανότητα, πράγμα που σημαίνει ότι είναι λιγότερο παραπλανητική μετρική από την ορθότητα.

Παρατηρούμενη ορθότητα είναι ο αριθμός των στοιχείων που ταξινομήθηκαν σωστά σε ολόκληρο τον πίνακα σύγχυσης. Για τον υπολογισμό της προστίθεται ο αριθμός των περιπτώσεων που ο ταξινομητής Μηχανικής Μάθησης συμφώνησε με την ετικέτα της αληθούς βάσης και διαιρείται με τον συνολικό αριθμό στοιχείων.

Αναμενόμενη ακρίβεια είναι η τιμή που ορίζεται ως η ακρίβεια που αναμένεται να επιτύχει κάθε τυχαίος ταξινομητής βάσει του πίνακα σύγχυσης. Σχετίζεται άμεσα με τον αριθμό παρουσιών κάθε τάξης, μαζί με τον αριθμό των περιπτώσεων που ο ταξινομητής Μηχανικής Μάθησης συμφώνησε με την ετικέτα της αληθούς βάσης.

Ο συντελεστής kappa μπορεί στη συνέχεια να υπολογιστεί χρησιμοποιώντας τόσο την παρατηρούμενη ακρίβεια, όσο και την αναμενόμενη ακρίβεια από τον τύπο,

$$Kappa = \frac{Observed\ Accuracy - Expected\ Accuracy}{1 - Expected\ Accuracy}. \quad (2.25)$$

Στην ουσία, είναι ένα μέτρο του πόσο κοντινά τα στοιχεία που ταξινομούνται από τον ταξινομητή Μηχανικής Μάθησης ταιριάζουν με τα δεδομένα που επισημαίνονται ως αληθής βάση, ελέγχοντας την ακρίβεια ενός τυχαίου ταξινομητή, όπως μετριέται από την αναμενόμενη ακρίβεια. Ο συντελεστής kappa όχι μόνο μπορεί να ρίξει φως στην απόδοση του ίδιου του ταξινομητή, αλλά για ένα μοντέλο είναι άμεσα συγκρίσιμος με τον συντελεστή kappa οποιουδήποτε άλλου μοντέλου που χρησιμοποιείται για την ίδια εργασία ταξινόμησης.

Δεν υπάρχει τυποποιημένη ερμηνεία του. Ο Landis και ο Koch [17] θεωρούν το < 0.00 ως ανεπαρκές, 0 - 0.20 ως ισχνό, το 0.21 - 0.40 ως δίκαιο, το 0.41 - 0.60 ως μέτριο, το 0.61 - 0.80 ως ουσιαστικό και το 0.81 - 1 ως σχεδόν τέλειο. Ο Fleiss [8] θεωρεί τους συντελεστές > 0.75 εξαιρετικούς, 0.40 - 0.75 ως δίκαιους προς καλούς και < 0.40 ως ανεπαρκείς. Είναι σημαντικό να σημειωθεί ότι και οι δύο κλίμακες είναι κάπως αυθαίρετες. Τουλάχιστον δύο περαιτέρω σκέψεις πρέπει να ληφθούν υπόψη κατά την ερμηνεία του συντελεστή kappa. Πρώτον, θα πρέπει πάντοτε να συγκρίνεται με έναν συνοδευτικό πίνακα σύγχυσης, εάν είναι δυνατόν, για να ληφθεί η πιο ακριβής ερμηνεία. Δεύτερον, οι αποδεκτές στατιστικές τιμές ποικίλλουν ανάλογα με το θέμα της ταξινόμησης.

ΚΕΦΑΛΑΙΟ 3

Μοντέλα Βαθιάς Μάθησης για Σημασιολογική Κατάτμηση

Σε αυτό το κεφάλαιο παρουσιάζονται κάποια από τα πιο διαδεδομένα μοντέλα για τη Σημασιολογική Κατάτμηση οπτικών δεδομένων πολύ υψηλής ανάλυσης στα οποία βασιστήκαμε για τη σύνθεση του προτεινόμενου Συνελικτικού Νευρωνικού μας μοντέλου, αλλά και για τη σύγκρισή του, ώστε να εξάγουμε τα καλύτερα αποτελέσματα. Αρχικά, γίνεται μια λεπτομερής περιγραφή των μοντέλων και της αρχιτεκτονικής τους κι έπειτα μια οπτική αναπαράστασή τους. Τα μοντέλα που αναπτύσσονται είναι το **U-Net**, που αποτελεί τη βάση του δικού μας μοντέλου, το **VGG-16** και το **FCN-Pascal** για την εξήγηση του **μοντέλου των Marmanis et al.**, από το οποίο προκύπτει η τελική μας ταξινόμηση.

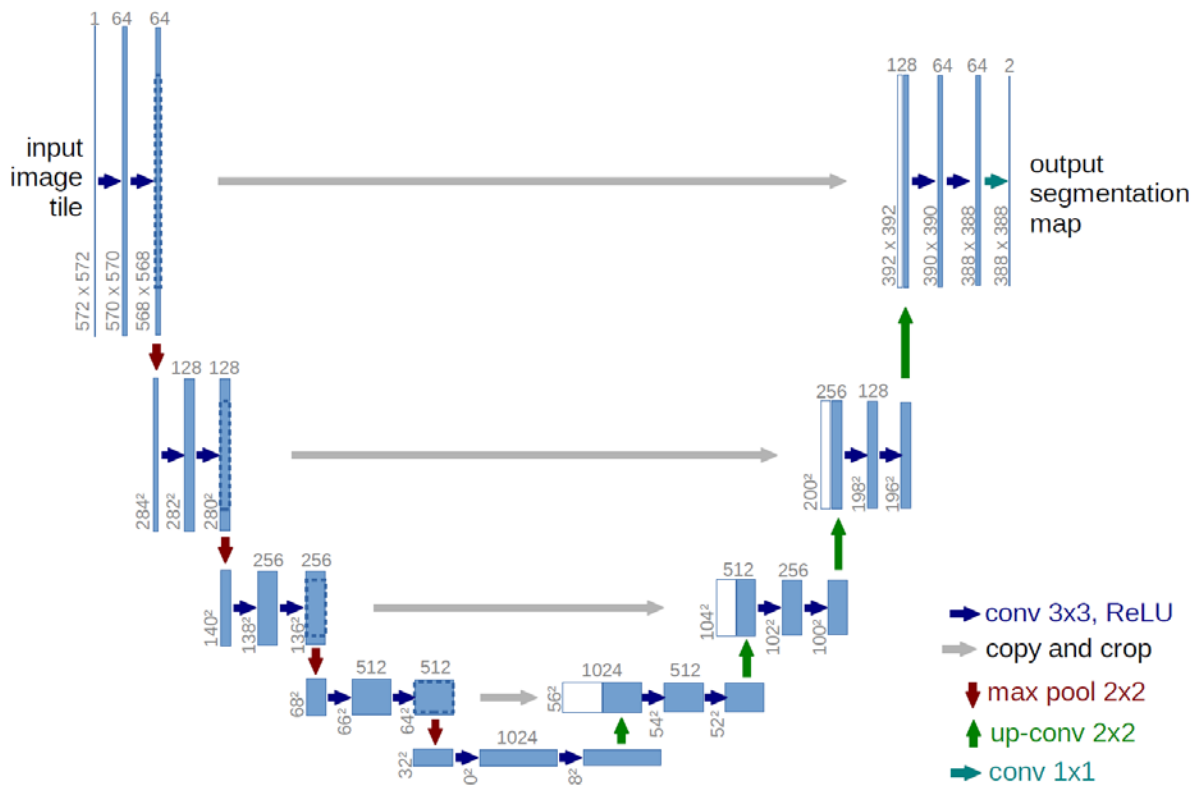
3.1 Το μοντέλο U-Net

Το μοντέλο U-Net [24] έχει χτιστεί σύμφωνα με τη λογική των Πλήρως Συνδεδεμένων δικτύων. Η κεντρική ιδέα του συγκεκριμένου δικτύου παραμένει ίδια, η εικόνα εισόδου ακολουθεί ένα *μονοπάτι συστολής* (contracting path), όπου η εικόνα συμπιέζεται σε μικρότερες διαστάσεις κι ένα *μονοπάτι διαστολής* (expansive path), όπου η εικόνα ανοικοδομείται στις αρχικές της. Η διαφορά που εισάγεται στο συγκεκριμένο δίκτυο είναι οι *συνδέσεις παράκαμψης* (skip connections). Αυτές εφαρμόζονται στα σημεία όπου έχουμε ίδιες διαστάσεις κι επιτρέπουν στο κομμάτι της υπερδειγματοληψίας τη διατήρηση του μεγάλου αριθμού πληροφοριών περιβάλλοντος, ώστε το δίκτυο να μπορεί να διαδώσει της πληροφορίες περιεχομένου σε επίπεδα υψηλότερης ανάλυσης. Αυτό έχει ως απόρροια, να δημιουργείται μια συμμετρία μεταξύ των μονοπατιών συστολής και διαστολής, δημιουργώντας την αίσθηση ενός σχήματος που θυμίζει το αγγλικό γράμμα «U», εξ ου κι η ονομασία «U-Net». Το δίκτυο δεν έχει Πλήρως Συνδεδεμένα επίπεδα και χρησιμοποιεί μόνο το έγκυρο κομμάτι κάθε Συνέλιξης, δηλαδή η μάσκα κατάτμησης περιέχει μόνο τα εικονοστοιχεία που αντιστοιχούν στην εικόνα εισόδου.

Αυτή η διαδικασία μαζί με την πρόβλεψη εικονοστοιχείων στα σύνορα της εικόνας μέσω εξαγωγής πληροφοριών περιεχομένου μέσω κατοπτρισμού της εικόνας εισόδου, επιτρέπουν τη συνεχόμενη κατάτμηση αυθαίρετα μεγάλων εικόνων.

Πιο αναλυτικά, η αρχιτεκτονική του δικτύου αποτελείται από ένα μονοπάτι συστολής στα αριστερά κι ένα κομμάτι διαστολής στα δεξιά, όπως φαίνεται και στο Σχήμα 3. 1. Το μονοπάτι της συστολής ακολουθεί την τυπική αρχιτεκτονική των συνελικτικών δικτύων. Απαρτίζεται από επαναλαμβανόμενες εφαρμογές διπλών Συνελίξεων 3×3 (χωρίς επένδυση ακριανών τιμών), τις οποίες ακολουθεί η συνάρτηση της Διορθωμένης Γραμμικής Μονάδας κι η εφαρμογή Μέγιστης Συγκέντρωσης με βήμα 2 για την επίτευξη της υποδειγματοληψίας (downsampling). Σε κάθε βήμα υποδειγματοληψίας διπλασιάζουμε τον αριθμό των συμμετεχόντων καναλιών. Κάθε βήμα στο μονοπάτι της διαστολής αποτελείται από την υπερδειγματοληψία των χαρτογραφημένων χαρακτηριστικών, η οποία ακολουθείται από μία Συνέλιξη 2×2 που μειώνει στο μισό τον αριθμό των συμμετεχόντων καναλιών, μια συνένωση με τα αντίστοιχα χαρτογραφημένα χαρακτηριστικά που έχουν αποκοπεί από το μονοπάτι της συστολής και δύο Συνελίξεις 3×3 , που τις ακολουθεί μία Συνάρτηση Διορθωμένης Γραμμικής Μονάδας. Στο τελικό επίπεδο, μία Συνέλιξη 1×1 χρησιμοποιείται για να αντιστοιχήσει κάθε διάνυσμα στο επιθυμητό αριθμό κλάσεων. Το δίκτυο αποτελείται από 23 συνελικτικά επίπεδα στο σύνολό του.

Τα μειονεκτήματα του δικτύου είναι ότι χάνεται μεγάλος όγκος χωρικής πληροφορίας λόγω των επιπέδων μέγιστης συγκέντρωσης και δεν έχει την ικανότητα να συσχετίσει το περιεχόμενο με τις συντεταγμένες του. [24]



Σχήμα 3. 1: Αρχιτεκτονική δικτύου U-Net. Παράδειγμα για εικονοστοιχεία 32×32 στη χαμηλότερη ανάλυση. Κάθε μπλε κουτί αντιστοιχεί σε έναν πολυκάναλο χάρτη. Ο αριθμός των καναλιών αναγράφεται πάνω από τα κουτιά. Τα λευκά κουτιά αποτελούν τα αντιγραμμένα χαρακτηριστικά. Τα βέλη προσδιορίζουν τις διάφορες λειτουργίες. [24]

3.2 Το μοντέλο VGG-16

Το μοντέλο VGG-16 σχεδιάστηκε για την Πρόκληση Οπτικής Αναγνώρισης Μεγάλης Κλίμακας του ImageNet 2012 και ξεχώρισε για τα άριστα συνολικά του αποτελέσματα [28]. Σημαντικά χαρακτηριστικά της αρχιτεκτονικής VGG είναι οι σχετικά λίγες εκπαιδευσιμες παράμετροι ανά επίπεδο, λόγω της χρήσης μικρών Συνελικτικών φίλτρων μεγέθους 3×3 . Αυτό καθιστά δυνατή την εκπαίδευση πολύ βαθιών δικτύων με 16 (ή ακόμη και 19) επίπεδα σε εύλογο χρόνο. [19]

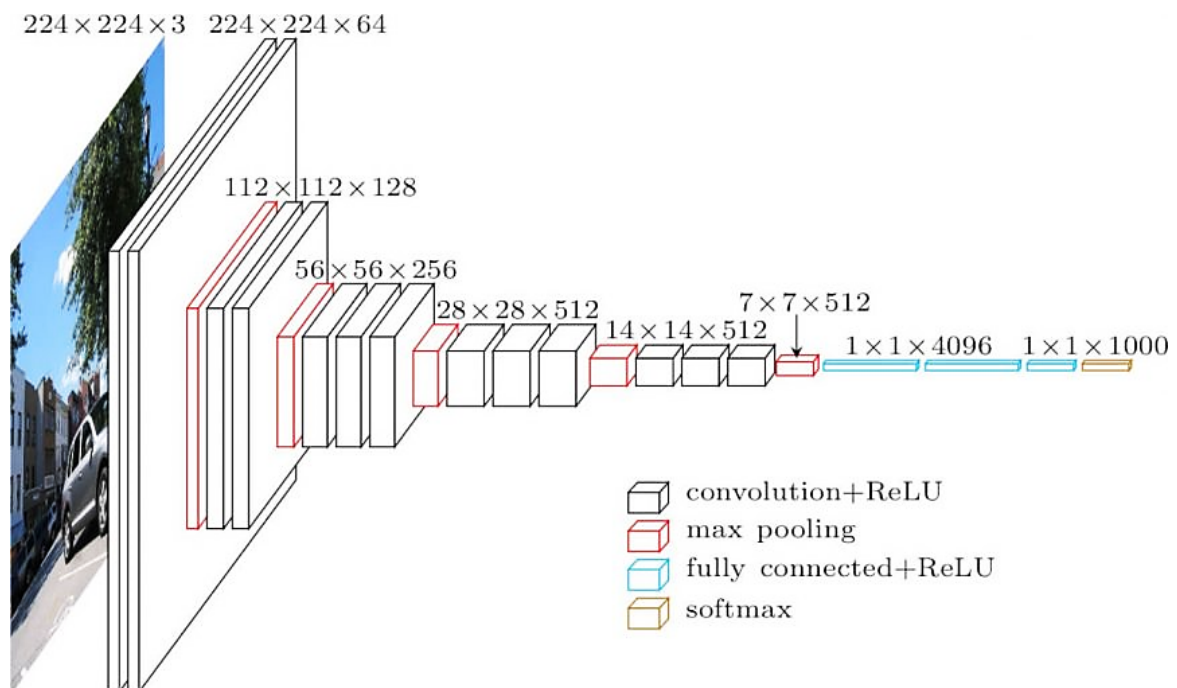
Συγκεκριμένα, κατά την εκπαιδευτική διαδικασία, το δεδομένο εισόδου είναι μια RGB εικόνα που παραμένει σε σταθερό μέγεθος 224×224 . Η μόνη προεπεξεργασία που συμβαίνει, είναι η αφαίρεση

της μέσης τιμής RGB, η οποία υπολογίζεται στο σύνολο εκπαίδευσης από κάθε εικονοστοιχείο. Η εικόνα περνά μέσα από μια στοίβα Συνελικτικών επιπέδων, όπου χρησιμοποιούνται φίλτρα 3×3 . Σε μία εκ των διατάξεων χρησιμοποιούνται, επίσης, Συνελικτικά φίλτρα 1×1 , τα οποία μπορούν να θεωρηθούν ως γραμμική μεταμόρφωση των καναλιών εισόδου (ακολουθούμενη από μη-γραμμικότητα). Το βήμα συνέλιξης είναι σταθερό στο 1 εικονοστοιχείο και η επένδυση των ακριανών τιμών για το επίπεδο εισόδου κρατά σταθερή τη χωρική ανάλυση μετά τη συνέλιξη. Υλοποιούνται 5 επίπεδα Μέγιστης Συγκέντρωσης 2×2 , τα οποία ακολουθούν κάποια από τα Συνελικτικά επίπεδα, και με βήμα 2.

Μια στοίβα Συνελικτικών επιπέδων διαδέχονται 3 Πλήρως Συνδεδεμένα επίπεδα, που καταλήγουν σε ένα επίπεδο Κανονικοποιημένης Εκθετικής Συνάρτησης.

Όλα τα κρυφά επίπεδα είναι εξοπλισμένα με τη συνάρτηση ενεργοποίησης, Διορθωμένη Γραμμική Μονάδα, και μόνο σε ένα επίπεδο εφαρμόζεται η Κανονικοποίηση Τοπικής Απόκρισης.

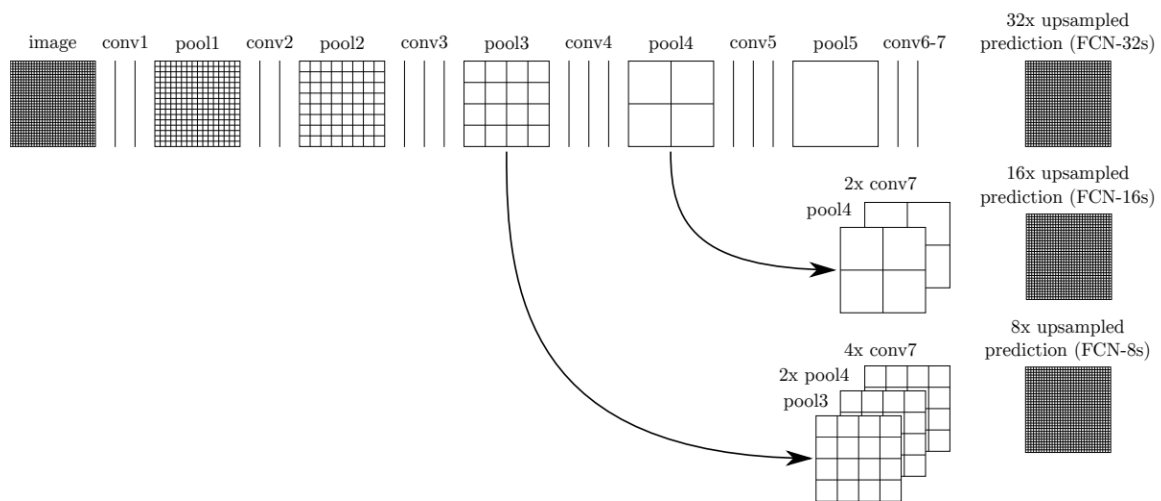
Αυτό το μοντέλο έχει επιτύχει 92.7% ορθότητα στο σύνολο του ImageNet, το οποίο περιέχει 14 εκατομμύρια εικόνες που ανήκουν σε 1000 κλάσεις. Αυτό καθιστά την επιλογή των βαρών του πρώτη σε προτίμηση για τη διαδικασία της Σημασιολογικής Κατάτμησης. [27]



Σχήμα 3. 2: Αρχιτεκτονική του μοντέλου VGG-16. [12]

3.3 Το μοντέλο FCN-Pascal

Το μοντέλο FCN-Pascal είναι ένα ισχυρό δίκτυο προ-εκπαιδευμένο στη βάση δεδομένων του Pascal VOC Context με στόχο τη Σημασιολογική Κατάτμηση. Αποσκοπεί σε υπερδευματοληπτικές προβλέψεις εντός δικτύου και σφάλματος ως προς τα εικονοστοιχεία. Χρησιμοποιεί συνδέσεις παράκαμψης, οι οποίες μαθαίνουν *διατεμαχικά* (end-to-end) με σκοπό τη σημασιολογική βελτίωση και τη χωρική ακρίβεια του αποτελέσματος. Η εκπαίδευση γίνεται με την Κανονικοποιημένη Εκθετική Συνάρτηση σφάλματος ανά εικονοστοιχείο και επαληθεύεται με τον μέσο λόγο *τομής προς ένωση* (Intersection over Union – IoU), όπου η μέση τιμή λαμβάνεται σε όλες τις κλάσεις και αγνοεί εικονοστοιχεία που αποκρύπτονται από την αληθή βάση. Τα κατώτερα επίπεδα του έχουν την ίδια διάταξη με το δίκτυο VGG-16. Πρόκειται για ένα δίκτυο με διαδοχική διαδικασία εκπαίδευσης η οποία ξεκινά από ένα πιο «ρηχό» μοντέλο και σταδιακά προσθέτει επίπεδα με σκοπό την εκμάθηση σύνδεσης *κατευθυνόμενου άκωνου γράφου* (DAG-connection) από κατώτερα συνελικτικά επίπεδα σε υψηλά αναστροφικά. [26]

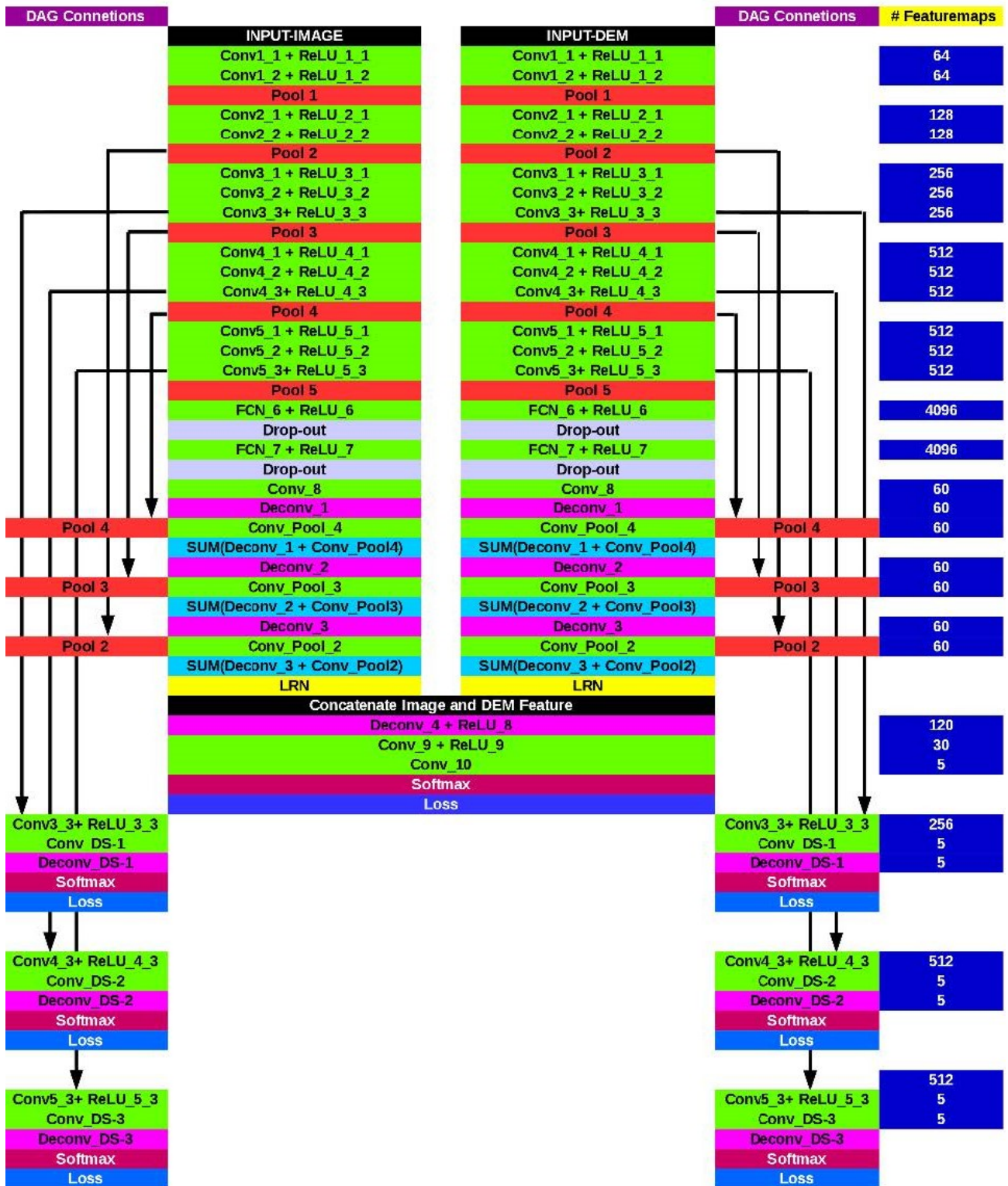


Σχήμα 3. 3: Αρχιτεκτονική FCN-Pascal. Τα επίπεδα συγκέντρωσης και πρόβλεψης εμφανίζονται ως πλέγματα, ενώ τα ενδιάμεσα επίπεδα με οριζόντιες γραμμές. **Πρώτη γραμμή (FCN-32s):** Μονής ροής δίκτυο, υπερδευματοληπία με βήμα 32, προβλέψεις πίσω στα εικονοστοιχεία σε ένα βήμα. **Δεύτερη γραμμή (FCN-16s):** Συνδυασμός προβλέψεων από το τελικό επίπεδο και το επίπεδο «pool4», με βήμα 16, επιτρέπει την πρόβλεψη λεπτομέρειας υψηλότερης ανάλυσης, ενώ διατηρεί σημασιολογική πληροφορία υψηλού επιπέδου. **Τρίτη γραμμή (FCN-8s):** Πρόσθετες προβλέψεις από το επίπεδο «pool3», με βήμα 8, παρέχουν μεγαλύτερη ακρίβεια. [26]

3.4 Το μοντέλο των Marmanis et al.

Το μοντέλο που προτείνεται από τους Marmanis et al., είναι μια επέκταση του δικτύου FCN-Pascal. Χρησιμοποιεί παντού φίλτρα Συνέλιξης 3×3 . Συγκριτικά με την αρχική διάταξη, έχει προστεθεί ακόμη μία σύνδεση παράκαμψης με σκοπό την έγχυση χαρακτηριστικών υψηλής ανάλυσης από ένα αρκετά πρώιμο επίπεδο, ώστε να αντιπροσωπεύεται καλύτερα η υψηλότερη λεπτομέρεια των ορίων των κλάσεων.

Ως δεδομένα εισόδου χρησιμοποιούνται τόσο η ένταση της εικόνας, όσο και το ψηφιακό υψομετρικό μοντέλο (digital elevation model – DEM). Επειδή τα υψομετρικά δεδομένα και τα δεδομένα της έντασης απαιτούν διαφορετικές αναπαραστάσεις χαρακτηριστικών, δημιουργούνται δύο ξεχωριστά μονοπάτια με την ίδια αρχιτεκτονική τα οποία συγχωνεύονται λίγο πριν το τελικό επίπεδο ώστε να εξαχθεί η πιθανότητα κάθε κλάσης. Αυτό επιτρέπει στο δίκτυο να μάθει ανεξάρτητα σύνολα σημαντικών χαρακτηριστικών για τις δύο δεδομένα εισόδου, τα οποία καθοδηγούνται από την ίδια συνάρτηση σφάλματος. Μια τελευταία τροποποίηση γίνεται με σκοπό να αποφευχθεί η υπερβολική αύξηση ενεργοποιήσεων σε έναν μικρό αριθμό νευρώνων και να προληφθεί ο σχηματισμός αιχμών, των οποίων η υπερβολική επιρροή μπορεί να διακόψει την εκπαίδευση του δικτύου κι αυτή είναι η πρόσθεση της Κανονικοποίησης Τοπικής Απόκρισης, ως τελευταίου επιπέδου των δύο ξεχωριστών μονοπατιών, ακριβώς πριν συγχωνευθούν. [19]



Σχήμα 3. 4: Αρχιτεκτονική του δικτύου Marmanis et al. Στα αριστερά τα επίπεδα κι οι συνδέσεις και στα δεξιά ο αριθμός των φίλτρων ανά επίπεδο. Όλα τα φίλτρα Συνέλιξης είναι 3×3 , όλα τα επίπεδα Μέγιστης Συγκέντρωσης είναι 2×2 . [19]

ΚΕΦΑΛΑΙΟ 4

Μεθοδολογία, Πειραματική Διαδικασία και Αποτελέσματα

Σε αυτό το κεφάλαιο παρουσιάζεται η αρχιτεκτονική που προτείνουμε για τη σημασιολογική κατάτμηση εναέριων τηλεπισκοπικών δεδομένων πολύ υψηλής ανάλυσης. Πρωταρχικά, αναλύεται το σύνολο δεδομένων που χρησιμοποιήθηκε για την πειραματική μας διαδικασία κι η επεξεργασία που υπόκειται κατά την υλοποίηση του αλγορίθμου. Ύστερα, περιγράφεται λεπτομερώς η μεθοδολογία που ακολουθεί το μοντέλο μας και τα εργαλεία που θέσαμε σε εφαρμογή για την προγραμματιστική του υλοποίηση. Τέλος, μέσω της πειραματικής διαδικασίας εξετάζουμε το μοντέλο μας σε σύγκριση με τα δύο μοντέλα που συνέβαλαν στη σχεδίαση της αρχιτεκτονικής του Συνελικτικού Νευρωνικού μας μοντέλου και παραθέτουμε αναλυτικά τα αποτελέσματα που προκύπτουν.

4.1 Σύνολο Δεδομένων

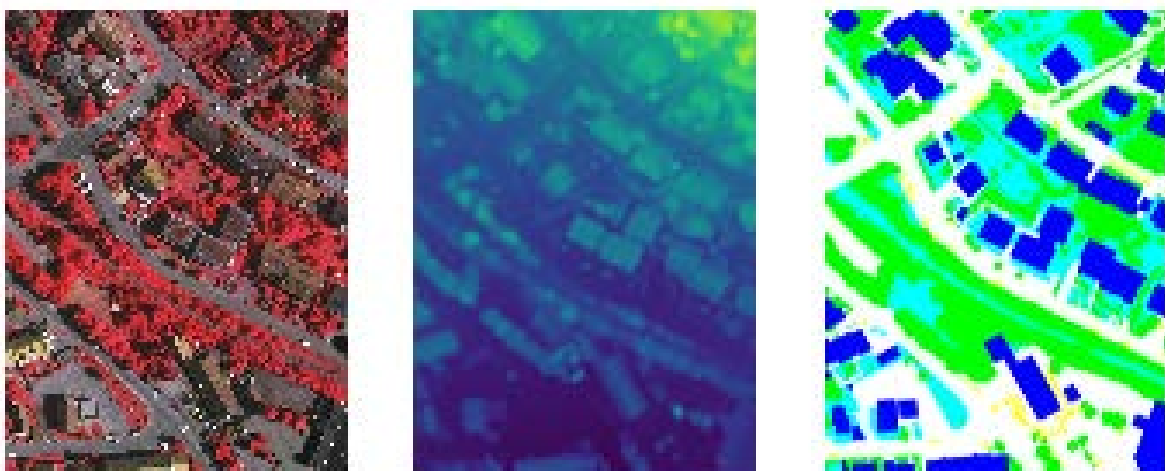
Για την ανάπτυξη της αρχιτεκτονικής του Συνελικτικού Νευρωνικού μας μοντέλου, βασιστήκαμε στο σύνολο δεδομένων του διαγωνισμού *2-διάστατης Σημασιολογικής Ταξινόμησης της Διεθνούς Κοινότητας Φωτογραμμετρίας και Ψηφιακής Τηλεπισκόπησης* (2D Semantic Labeling Contest - International Society for Photogrammetry and Remote Sensing). Παρέχεται ένα σύνολο δεδομένων που αντιπροσωπεύει τις περιοχές του Πότσνταμ και του Βάιχινγκεν. Περιλαμβάνει δύο σύνολα εναέριων δεδομένων που έχουν εξαχθεί με υπερσύγχρονα μέσα. Το πρώτο αφορά ορθοφωτογραφίες πολύ υψηλής ανάλυσης (TOP) και το δεύτερο τα αντίστοιχα ψηφιακά μοντέλα επιφάνειάς τους (DSM), που προκύπτουν από πυκνές τεχνικές αντιστοίχισης εικόνων. Και οι δύο περιοχές αντιπροσωπεύουν αστικά τοπία. Παρόλο που το Βάιχινγκεν είναι ένα σχετικά μικρό χωριό με πολλά αποκομμένα και μικρά πολυώροφα κτίρια, ενώ το Πότσνταμ είναι μια ιστορική πόλη που καλύπτεται από οικοδομικά τετράγωνα, στενούς δρόμους και πυκνοκατοικημένη δομή.

Κάθε σύνολο δεδομένων έχει κατηγοριοποιηθεί χειρωνακτικά στις 6 πιο διαδεδομένες κατηγορίες χρήσεων γης. Παρέχονται τα δεδομένα ταξινόμησης (κατηγοριοποιημένες εικόνες) προσεγγιστικά των μισών εικόνων. Οι κατηγορίες/κλάσεις που έχουν προσδιοριστεί είναι οι εξής:

- Διαπεράστες Επιφάνειες (Impervious surfaces) (RGB: 255, 255, 255)
- Κτίριο (Building) (RGB: 0, 0, 255)
- Χαμηλή Βλάστηση (Low vegetation) (RGB: 0, 255, 255)
- Δέντρο (Tree) (RGB: 0, 255, 0)
- Αυτοκίνητο (Car) (RGB: 255, 255, 0)
- Συσσώρευση/Φόντο (Clutter/background) (RGB: 255, 0, 0)

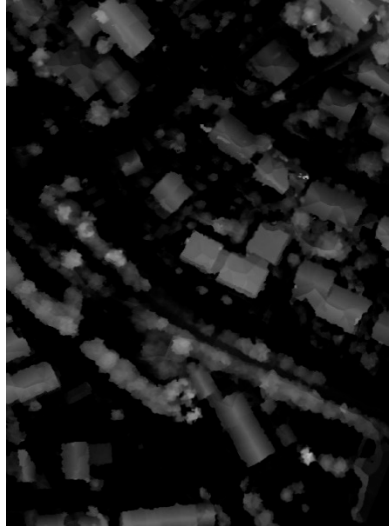
Η κλάση συσσώρευση/φόντο εμπεριέχει σώματα νερού κι άλλα αντικείμενα τα οποία δείχνουν πολύ διαφορετικά από οποιοδήποτε άλλο (όπως γήπεδα τένις, πισίνες κ.λπ.) και συνήθως δεν αποτελούν κομμάτι ενδιαφέροντος της σημασιολογικής κατάτμησης αστικών τοπίων. [1]

Εφόσον βασιζόμαστε στην έρευνα των Marmanis et al., επιλέξαμε το σύνολο δεδομένων του Βάιχιγγκεν για την πειραματική διαδικασία του συνελικτικού μας δικτύου. Παρακάτω δίνεται παράδειγμα απεικόνισης του συνόλου.



Σχήμα 4. 1: Παράδειγμα απεικόνισης του συνόλου δεδομένων του Βάιχιγγκεν για την περιοχή 11. **Αριστερά:** Ορθοφωτογραφία. **Μέση:** Ψηφιακό Μοντέλο Επιφάνειας (DSM). **Δεξιά:** Αληθής Βάση (Ground Truth).

Εδώ χρειάζεται να αναφέρουμε, πως η έρευνα χρησιμοποιεί ένα σύνολο δεδομένων ψηφιακών υψομετρικών μοντέλων (DEM), το οποίο δεν μας παρέχεται. Αντ' αυτού, για την εξαγωγή των δικών μας αποτελεσμάτων, αποφασίσαμε να χρησιμοποιήσουμε το σύνολο δεδομένων των *κανονικοποιημένων ψηφιακών μοντέλων εδάφους* [9] (normalized digital surface model – nDSM).



Σχήμα 4. 2: Παράδειγμα απεικόνισης Κανονικοποιημένου Ψηφιακού Μοντέλου Επιφάνειας (nDSM) για την περιοχή 11. [9]

4.2 Υλικό Υπολογιστή (Hardware)

Για την υλοποίηση του αλγορίθμου δουλέψαμε στα δωρεάν διαδικτυακά περιβάλλοντα ανάπτυξης ανοιχτού κώδικα Google Colab και Kaggle. Και τα δύο προσφέρονται για διεργασίες στους τομείς της Μηχανικής Μάθησης και της Επιστήμης Δεδομένων, καθώς παρέχουν υπολογιστικά ισχυρές κάρτες γραφικώς (Graphics Processing Unit – GPU) και δυνατές κεντρικές μονάδες επεξεργασίας (Central Processing Unit– CPU).

Συγκεκριμένα, το Google Colab δίνει τη δυνατότητα εκπαίδευσης σε κάρτα γραφικών NVIDIA Tesla K80 και μνήμη τυχαίας προσπέλασης (Random Access Memory - RAM) 12GB ενώ το Kaggle δίνει τη

δυνατότητα εκπαίδευσης σε κάρτα γραφικών Nvidia Telsa P100 σε και μνήμη τυχαίας προσπέλασης 13GB.

4.3 Εκπαίδευση

Βασιστήκαμε στον αλγόριθμο των Audebert et al. [3] για την εισαγωγή, την ενίσχυση και την κατηγοριοποίηση των περιοχών για τη δημιουργία των συνόλων εκπαίδευσης και ελέγχου, καθώς και για τις συναρτήσεις της εκπαίδευσης και του ελέγχου και προσθέσαμε τις απαραίτητες εντολές, ώστε να γίνει εισαγωγή και του συνόλου του ψηφιακού μοντέλου επιφάνειας.

Οι τιμές που επιλέχθηκαν για την εκπαίδευση του μοντέλου, προκύπτουν μέσω της *ευρετικής τεχνικής* (heuristic method). Προσπαθούμε για το καλύτερο αποτέλεσμα βάσει του υλικού υπολογιστή και των εργαλείων που διαθέτουμε.

Καταλήξαμε στις τιμές, 5 για το μέγεθος παρτίδας και 256×256 για το μέγεθος παραθύρου. Χρησιμοποιούμε τη Στοχαστική Κλίση Καθόδου, με ορμή 0.9, ρυθμό μάθησης 0.01 και μείωση βαρών 0.0005 και για συνάρτηση απώλειας τη Σταυροειδή Εντροπία για 5 εποχές. Για βάρη επιλέξαμε τα προ-εκπαιδευμένα βάρη του δικτύου VGG-16.

Επιπλέον, εφαρμόζουμε Ενίσχυση Δεδομένων μέσω καθρεπτισμού εικόνας και περιστροφής και για αντιστοιχία αποτελεσμάτων με την έρευνα των Marmanis et al., ορίζονται ως σύνολα εκπαίδευσης και ελέγχου οι περιοχές [1, 3, 11, 13, 15, 17, 21, 26, 34, 32, 28, 37] και [5, 7, 23, 30], αντίστοιχα.

4.4 Εργαλεία

Η υλοποίηση του αλγορίθμου του μοντέλου έγινε σε γλώσσα Python 3.0. Χρησιμοποιήθηκε το PyTorch, ένα λογισμικό Μηχανικής Μάθησης και ανοιχτού κώδικα το οποίο παρέχεται δωρεάν και βασίζεται στο Torch. Έχει δύο κύρια χαρακτηριστικά, πρώτον υπολογίζει ταυνοστές με ισχυρή

επιτάχυνση της κάρτας γραφικών και δεύτερον, έχει τη δυνατότητα αυτόματης παραγωγής για χτίσιμο κι εκπαίδευση Νευρωνικών δικτύων. Ένας ταχυστής PyTorch είναι σχετικά ίδιος με μια συστοιχία του NumPy, δηλαδή δε γνωρίζει τίποτα όσον αφορά τη Βαθιά Μάθηση και πρόκειται για μία n -διάστατη συστοιχία που χρησιμοποιείται για αριθμητικούς υπολογισμούς, με τη σημαντική διαφορά ότι μπορεί να υλοποιηθεί μέσω της κάρτας γραφικών ή της μητρικής. Το προτιμήσαμε γιατί συγκριτικά με άλλα λογισμικά και βιβλιοθήκες, παρέχει μέγιστη ταχύτητα και προσαρμοστικότητα.

Όπως αναφέρθηκε και πιο πάνω, χρησιμοποιήθηκε κι η, επίσης, ανοιχτού κώδικα βιβλιοθήκη NumPy η οποία χρησιμοποιεί συστοιχίες. Έχει συναρτήσεις που μπορούν να υλοποιήσουν πεδία όπως, η γραμμική άλγεβρα, ο μετασχηματισμός Φουριέρ και πίνακες. Οι γραφικές παραστάσεις οπτικοποιήθηκαν με τη βοήθεια της βιβλιοθήκης Matplotlib κι η πρόοδος της κάθε εποχής με το πακέτο (package) tqdm. Τα δομοστοιχεία (modules) skimage, sklearn.metrics και random χρησιμοποιήθηκαν για ανάλυση εικόνων, υλοποίηση μετρικών και παραγωγή τυχαίων μεταβλητών, αντίστοιχα.

4.5 Σχεδιασμός και Μεθοδολογία Προτεινόμενου Μοντέλου

Η αρχιτεκτονική που προτείνουμε βασίζεται στο μοντέλο U-Net και στο μοντέλο που προτάθηκε από τους Marmanis et al., για το σύνολο δεδομένων του διαγωνισμού 2-διάστατης Σηματολογικής Ταξινόμησης της Διεθνούς Κοινότητας Φωτογραμμετρίας και Ψηφιακής Τηλεπισκόπησης. Πρόκειται για ένα διατερματικό, πλήρως αυτοματοποιημένο δίκτυο με σκοπό τη Σηματολογική Κατάτμηση εναέριων τηλεπισκοπικών δεδομένων υψηλής ανάλυσης. Όπως και τα αρχικά μοντέλα, βασίζεται στη λογική των Πλήρως Συνδεδεμένων δικτύων. Προκειμένου να χρησιμοποιήσουμε τόσο την ένταση της εικόνας όσο και τα χαρακτηριστικά του κανονικοποιημένου ψηφιακού μοντέλου επιφάνειας ως δεδομένα εισόδου, πραγματοποιείται παράλληλη επεξεργασία μέσα στο δίκτυο δύο δομικά πανομοιότυπων σκελών. Η τελική κατηγοριοποίηση εκτελείται στη συνένωση τους.

Ειδικότερα, ακολουθείται ο συνελικτικός τρόπος των Marmanis et al. προσαρμοσμένος στη λογική του U-Net. Αρχικά, γίνεται παράλληλη εισαγωγή των δεδομένων εισόδου, δηλαδή της ορθοφωτογραφίας και του κανονικοποιημένου ψηφιακού μοντέλου επιφάνειας. Ακολουθούν 2 πλάκες αποτελούμενες από 2 επίπεδα Συνελίξεων 3×3 κι ένα επίπεδο Μέγιστης Συγκέντρωσης 2×2 κι

έπειτα, 3 πλάγιες αποτελούμενες από 3 επίπεδα Συνελίξεων 3×3 και κι ένα επίπεδο Μέγιστης Συγκέντρωσης 2×2 . Σε κάθε Συνέλιξη έχουμε μηδενική επένδυση αρχικών τιμών η οποία ισούται με 1, καθώς και Κανονικοποίηση Παρτίδας στην οποία εφαρμόζεται η συνάρτηση ενεργοποίησης της Διορθωμένης Γραμμικής Μονάδας. Αυτή η διαδικασία συγκροτεί το μονοπάτι συστολής, όπου όμοια με το U-Net, σε κάθε βήμα υποδειγματοληψίας διπλασιάζεται ο αριθμός των συμμετεχόντων καναλιών.

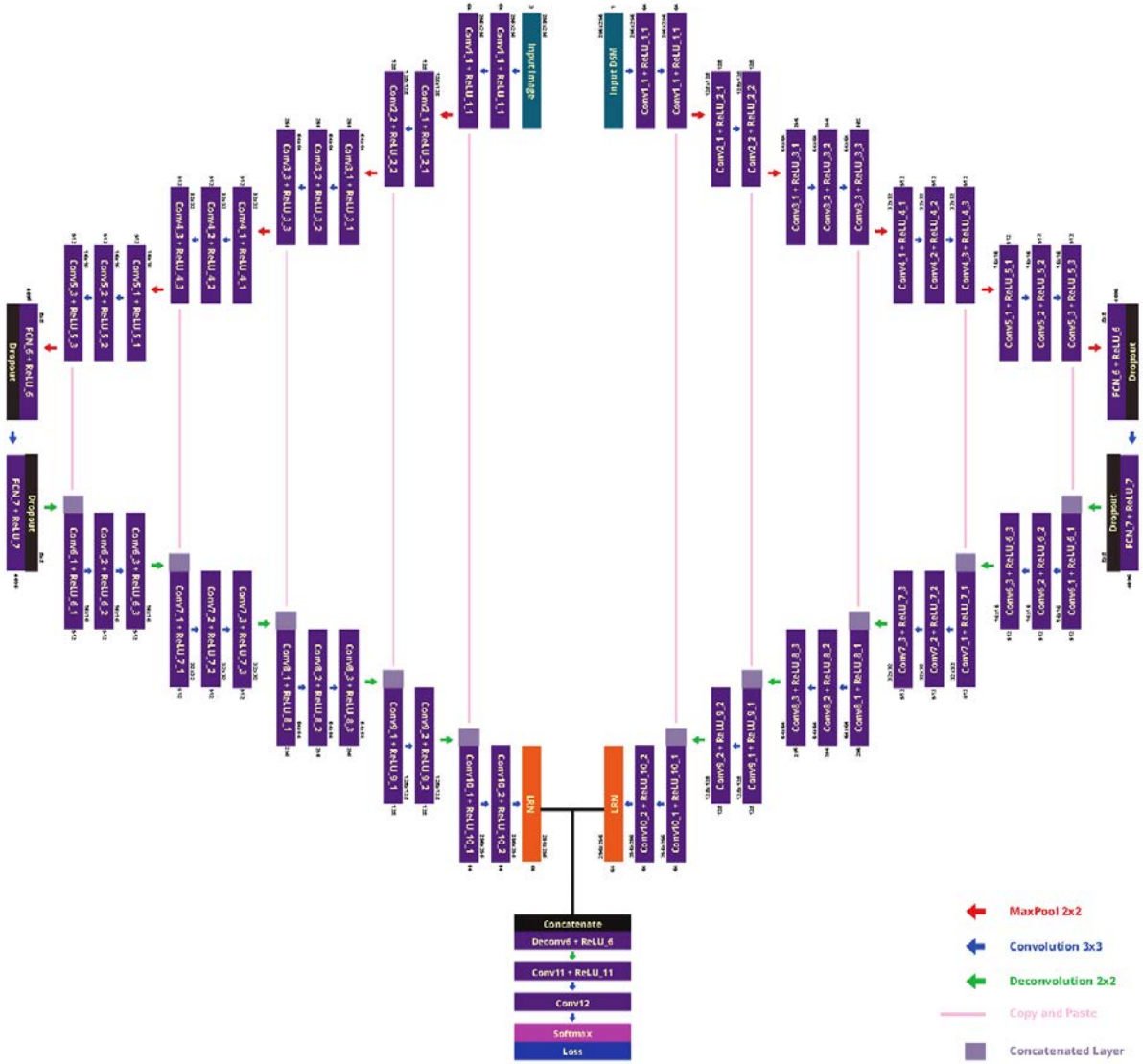
Σε αντίθεση με το U-Net, πριν το μονοπάτι της διαστολής εισάγονται 2 Πλήρως Συνδεδεμένα επίπεδα, όπου το καθένα συνοδεύεται από 1 επίπεδο Απόσυρσης. Εφόσον αργότερα γίνεται χρήση της Κανονικοποίησης Τοπικής Απόκρισης, θέτουμε το επίπεδο Απόσυρσης στο 0.05. Η ένταξη αυτών των επιπέδων έγινε ως προσπάθεια αποσυμφόρησης του δικτύου. Φυσικά, όπως και πριν, σε κάθε Πλήρως Συνδεδεμένο επίπεδο εφαρμόζεται η συνάρτηση ενεργοποίησης της Διορθωμένης Γραμμικής Μονάδας.

Μέχρι στιγμής η πρώτη διαφοροποίηση με το μοντέλο των Marmanis et al., ήταν η παράβλεψη των συνδέσεων των κατευθυνόμενων άκυκλων γράφων. Η δεύτερη εφαρμόζεται στο μονοπάτι της διαστολής, όπου ακολουθείται η αρχιτεκτονική του U-Net. Μετά το τελευταίο Πλήρως Συνδεδεμένο επίπεδο, εφαρμόζεται μία Ανάστροφη Συνέλιξη 2×2 . Το αποτέλεσμα αυτού του επιπέδου συνενώνεται με τα χαρτογραφημένα χαρακτηριστικά που έχουν αποκοπεί από το αντίστοιχο επίπεδο κι έχουν υποστεί μια Μέγιστη Αποσυγκέντρωση 2×2 . Ακολουθείται η αρχιτεκτονική του μονοπατιού της συστολής, με τη διαφορά ότι σε κάθε βήμα υπερδειγματοληψίας, στο τελευταίο επίπεδο αντί της Μέγιστης Συγκέντρωσης, εφαρμόζεται μία Ανάστροφη Συνέλιξη 2×2 , το οποίο συνενώνεται με τα αποκομμένα χαρτογραφημένα χαρακτηριστικά του αντίστοιχου επιπέδου.

Η διαφορά με το μοντέλο του U-Net, είναι πως στο τελικό επίπεδο πριν τη συνένωση των δύο αποτελεσμάτων, επιμένουμε σε Συνέλιξη 3×3 κι εφαρμόζουμε μια Κανονικοποίηση Τοπικής Απόκρισης, την οποία μέσω της ευρετικής τεχνικής θέτουμε 5.

Τέλος, γίνεται συνένωση των αποτελεσμάτων των δύο σκελών και προχωράμε όπως στην καταληκτική αρχιτεκτονική των Marmanis et al., εκτελώντας μια Ανάστροφη Συνέλιξη 3×3 με μια Κανονικοποίηση Παρτίδας, στην οποία εφαρμόζεται η συνάρτηση ενεργοποίησης της Διορθωμένης Γραμμικής Μονάδας, ένα επίπεδο Συνελίξης 3×3 με μια Κανονικοποίηση Παρτίδας, στην οποία εφαρμόζεται η συνάρτηση ενεργοποίησης της Διορθωμένης Γραμμικής Μονάδας κι ένα επίπεδο Συνελίξης 3×3 . Για την κατηγοριοποίηση του αποτελέσματος εξόδου τελείται η Κανονικοποιημένη Εχθετική

Συνάρτηση.



Σχήμα 4. 3: Οπτικοποίηση του προτεινόμενου μοντέλου μας.

4.6 Αποτελέσματα

Παρακάτω παρατίθενται τα αποτελέσματα των περιοχών 5, 7, 23 και 30, για το δίκτυο U-Net, το δίκτυο Marmanis et al. και το προτεινόμενο δίκτυο από εμάς. Σε αντίθεση με την έρευνα των Marmanis et al., γίνεται προσπάθεια κατάτμησης και των 6 ορισμένων κλάσεων.

4.6.1 Αποτελέσματα U-Net

4.6.1.1 Περιοχή 5

Confusion Matrix

Classes	Roads	Buildings	Low veg.	Trees	Cars	Clutter
Roads	1398204	94991	18847	3755	0	0
Buildings	37972	2258939	4255	1384	0	0
Low veg.	42391	4608	155871	60939	0	0
Trees	7241	702	6868	277825	0	0
Cars	39729	4704	259	254	0	0
Clutter	0	0	0	0	0	0

Classification report

Classes	n(truth)	n(classified)	Accuracy	Precision	Recall	F1-Score
Roads	1525537	1515797	94.46%	0.92	0.92	0.92
Buildings	2363944	2302550	96.64%	0.98	0.96	0.97
Low veg.	186100	263809	96.87%	0.59	0.84	0.69
Trees	344157	292636	98.16%	0.95	0.81	0.87
Cars	0	44946	98.98%	0.0	0.0	0.0
Clutter	0	0	100%	0.0	0.0	0.0

	Precision	Recall	F1-Score
Micro Average	0.93	0.93	0.93
Macro Average	0.57	0.59	0.58
Weighted Average	0.92	0.92	0.92

Kappa: 0.87

Total Accuracy: 92.56%

Πίνακας 4. 1: Πρώτος Έλεγχος Περιοχής 5.

Confusion Matrix

Classes	Roads	Buildings	Low veg.	Trees	Cars	Clutter
Roads	1398204	94991	18847	3755	0	0
Buildings	37972	2258939	4255	1384	0	0
Low veg.	42391	4608	155871	60939	0	0
Trees	7241	702	6868	277825	0	0
Cars	39729	4704	259	254	0	0
Clutter	0	0	0	0	0	0

Classification report

Classes	n(truth)	n(classified)	Accuracy	Precision	Recall	F1-Score
Roads	1525537	1515797	94.46%	0.92	0.92	0.92
Buildings	2363944	2302550	96.64%	0.98	0.96	0.97
Low veg.	186100	263809	96.87%	0.59	0.84	0.69
Trees	344157	292636	98.16%	0.95	0.81	0.87
Cars	0	44946	98.98%	0.0	0.0	0.0
Clutter	0	0	100%	0.0	0.0	0.0

	Precision	Recall	F1-Score
Micro Average	0.93	0.93	0.93
Macro Average	0.57	0.59	0.58
Weighted Average	0.92	0.92	0.92

Kappa: 0.87

Total Accuracy: 92.56%

Πίνακας 4. 2: Δεύτερος Έλεγχος Περιοχής 5.

Παρατηρείται σταθερότητα στις τιμές των δεικτών κατά τους δύο ελέγχους. Η υψηλή τιμή του δείκτη kappa αποδίδει αξιοπιστία στην υψηλή τιμή του total accuracy. Παρά τις μέτριες τιμές στη γραμμή του macro, αντιλαμβανόμαστε από τους δείκτες micro και weighted πως κάθε κλάση συμμετέχει στην έκβαση των αποτελεσμάτων σε σχεδόν άριστο βαθμό ενισχύοντας την εμπιστοσύνη των αποτελεσμάτων.

4.6.1.2 Περιοχή 7

Confusion Matrix

Classes	Roads	Buildings	Low veg.	Trees	Cars	Clutter
Roads	1211649	113688	14538	6905	0	0
Buildings	19900	1184705	6839	8112	0	0
Low veg.	24153	10360	437542	84844	0	0
Trees	10920	808	105866	1076789	0	0
Cars	69714	4460	936	142	0	0
Clutter	0	0	0	0	0	0

Classification report

Classes	n(truth)	n(classified)	Accuracy	Precision	Recall	F1-Score
Roads	1336336	1346780	94.09%	0.90	0.91	0.90
Buildings	1314021	1219556	96.26%	0.97	0.90	0.94
Low veg.	565721	556899	94.37%	0.79	0.77	0.78
Trees	1176792	1194383	95.05%	0.90	0.92	0.91
Cars	0	75252	98.29%	0.0	0.0	0.0
Clutter	0	0	100%	0.0	0.0	0.0

	Precision	Recall	F1-Score
Micro Average	0.89	0.89	0.89
Macro Average	0.59	0.58	0.59
Weighted Average	0.89	0.88	0.88

Kappa: 0.85

Total Accuracy: 89.02%

Πίνακας 4. 3: Πρώτος Έλεγχος Περιοχής 7.

Confusion Matrix

Classes	Roads	Buildings	Low veg.	Trees	Cars	Clutter
Roads	2609853	208679	33385	10660	0	0
Buildings	57872	3443644	11094	9496	0	0
Low veg.	66544	14968	593413	145783	0	0
Trees	18161	1510	112734	1354614	0	0
Cars	109443	9164	1195	396	0	0
Clutter	0	0	0	0	0	0

Classification report

Classes	n(truth)	n(classified)	Accuracy	Precision	Recall	F1-Score
Roads	2861873	2862577	94.27%	0.91	0.91	0.91
Buildings	3677965	3522106	96.45%	0.98	0.94	0.96
Low veg.	751821	820708	95.62%	0.72	0.79	0.75
Trees	1520949	1487019	96.61%	0.91	0.89	0.90
Cars	0	120198	98.64%	0.0	0.0	0.0
Clutter	0	0	100%	0.0	0.0	0.0

	Precision	Recall	F1-Score
Micro Average	0.91	0.91	0.91
Macro Average	0.59	0.59	0.59
Weighted Average	0.91	0.90	0.90

Kappa: 0.87

Total Accuracy: 90.8%

Πίνακας 4. 4: Δεύτερος Έλεγχος Περιοχής 7.

Παρατηρείται βελτίωση στο σύνολο των αποτελεσμάτων κατά τον δεύτερο έλεγχο. Η υψηλή τιμή του δείκτη kappa αποδίδει αξιοπιστία στην υψηλή τιμή του total accuracy. Παρά τις μέτριες τιμές στη γραμμή του macro, αντιλαμβανόμαστε από τους δείκτες micro και weighted πως κάθε κλάση συμμετέχει στην έμβαση των αποτελεσμάτων σε σχεδόν άριστο βαθμό ενισχύοντας την εμπιστοσύνη των αποτελεσμάτων.

4.6.1.3 Περιοχή 23

Confusion Matrix

Classes	Roads	Buildings	Low veg.	Trees	Cars	Clutter
Roads	604940	15317	65073	16896	0	0
Buildings	10652	787348	12523	3429	0	0
Low veg.	76976	18679	972250	123993	0	0
Trees	12864	8702	169749	1422581	0	0
Cars	10057	7	179	45	0	0
Clutter	3679	16	2697	0	0	0

Classification report

Classes	n(truth)	n(classified)	Accuracy	Precision	Recall	F1-Score
Roads	719168	702226	95.12%	0.86	0.84	0.85
Buildings	830069	813952	98.4%	0.97	0.95	0.96
Low veg.	1222471	1191898	89.17%	0.82	0.80	0.81
Trees	1566944	1613896	92.26%	0.88	0.91	0.89
Cars	0	10288	99.76%	0.0	0.0	0.0
Clutter	0	6392	99.85%	0.0	0.0	0.0

	Precision	Recall	F1-Score
Micro Average	0.87	0.87	0.87
Macro Average	0.59	0.58	0.59
Weighted Average	0.87	0.87	0.87

Kappa: 0.82

Total Accuracy: 87.29%

Πίνακας 4. 5: Πρώτος Έλεγχος Περιοχής 23.

Confusion Matrix

Classes	Roads	Buildings	Low veg.	Trees	Cars	Clutter
Roads	3214793	223996	98458	27556	0	0
Buildings	68524	4230992	23617	12925	0	0
Low veg.	143520	33647	1565663	269776	0	0
Trees	31025	10212	282483	2777195	0	0
Cars	119500	9171	1374	441	0	0
Clutter	3679	16	2697	0	0	0

Classification report

Classes	n(truth)	n(classified)	Accuracy	Precision	Recall	F1-Score
Roads	3581041	3564803	94.55%	0.90	0.89	0.90
Buildings	4508034	4336058	97.09%	0.98	0.94	0.96
Low veg.	1974292	2012606	93.49%	0.78	0.79	0.79
Trees	3087893	3100915	95.18%	0.90	0.90	0.90
Cars	0	130486	99.01%	0.0	0.0	0.0
Clutter	0	6392	99.95%	0.0	0.0	0.0

	Precision	Recall	F1-Score
Micro Average	0.90	0.90	0.90
Macro Average	0.59	0.59	0.59
Weighted Average	0.90	0.89	0.89

Kappa: 0.86

Total Accuracy: 89.64%

Πίνακας 4. 6: Δεύτερος Έλεγχος Περιοχής 23.

Παρατηρείται βελτίωση στο σύνολο των αποτελεσμάτων κατά τον δεύτερο έλεγχο. Η υψηλή τιμή του δείκτη kappa αποδίδει αξιοπιστία στην υψηλή τιμή του total accuracy. Παρά τις μέτριες τιμές στη γραμμή του macro, αντιλαμβανόμαστε από τους δείκτες micro και weighted πως κάθε κλάση συμμετέχει στην έμβαση των αποτελεσμάτων σε σχεδόν άριστο βαθμό ενισχύοντας την εμπιστοσύνη των αποτελεσμάτων.

4.6.1.4 Περιοχή 30

Confusion Matrix

Classes	Roads	Buildings	Low veg.	Trees	Cars	Clutter
Roads	1181344	30708	40037	12075	0	0
Buildings	25340	1170540	3506	2147	0	0
Low veg.	52138	42689	720979	143981	0	0
Trees	4394	608	43648	926084	0	0
Cars	49413	4887	1549	196	0	0
Clutter	0	0	0	0	0	0

Classification report

Classes	n(truth)	n(classified)	Accuracy	Precision	Recall	F1-Score
Roads	1312629	1264164	95.2%	0.93	0.90	0.92
Buildings	1249432	1201533	97.53%	0.97	0.94	0.96
Low veg.	809719	959787	92.65%	0.75	0.89	0.81
Trees	1084483	974734	95.35%	0.95	0.85	0.90
Cars	0	56045	98.74%	0.0	0.0	0.0
Clutter	0	0	100%	0.0	0.0	0.0

	Precision	Recall	F1-Score
Micro Average	0.90	0.90	0.90
Macro Average	0.60	0.60	0.60
Weighted Average	0.90	0.89	0.89

Kappa: 0.86

Total Accuracy: 89.74%

Πίνακας 4. 7: Πρώτος Έλεγχος Περιοχής 30.

Confusion Matrix

Classes	Roads	Buildings	Low veg.	Trees	Cars	Clutter
Roads	4396137	254704	138495	39631	0	0
Buildings	93864	5401532	27123	15072	0	0
Low veg.	195658	76336	2286642	413757	0	0
Trees	35419	10820	326131	3703279	0	0
Cars	168913	14058	2923	637	0	0
Clutter	3679	16	2697	0	0	0

Classification report

Classes	n(truth)	n(classified)	Accuracy	Precision	Recall	F1-Score
Roads	4893670	4828967	94.72%	0.91	0.90	0.90
Buildings	5757466	5537591	97.21%	0.98	0.94	0.96
Low veg.	2784011	2972393	93.28%	0.77	0.82	0.79
Trees	4172376	4075649	95.22%	0.91	0.89	0.90
Cars	0	186531	98.94%	0.0	0.0	0.0
Clutter	0	6392	99.96%	0.0	0.0	0.0

	Precision	Recall	F1-Score
Micro Average	0.90	0.90	0.90
Macro Average	0.60	0.59	0.59
Weighted Average	0.90	0.89	0.89

Kappa: 0.86

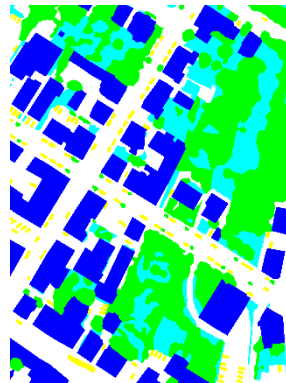
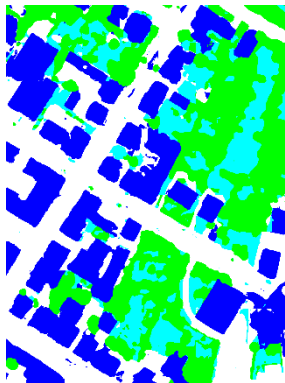
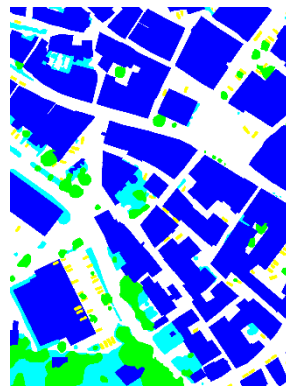
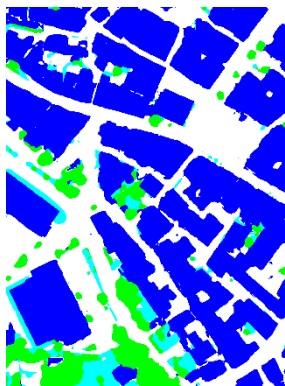
Total Accuracy: 89.66%

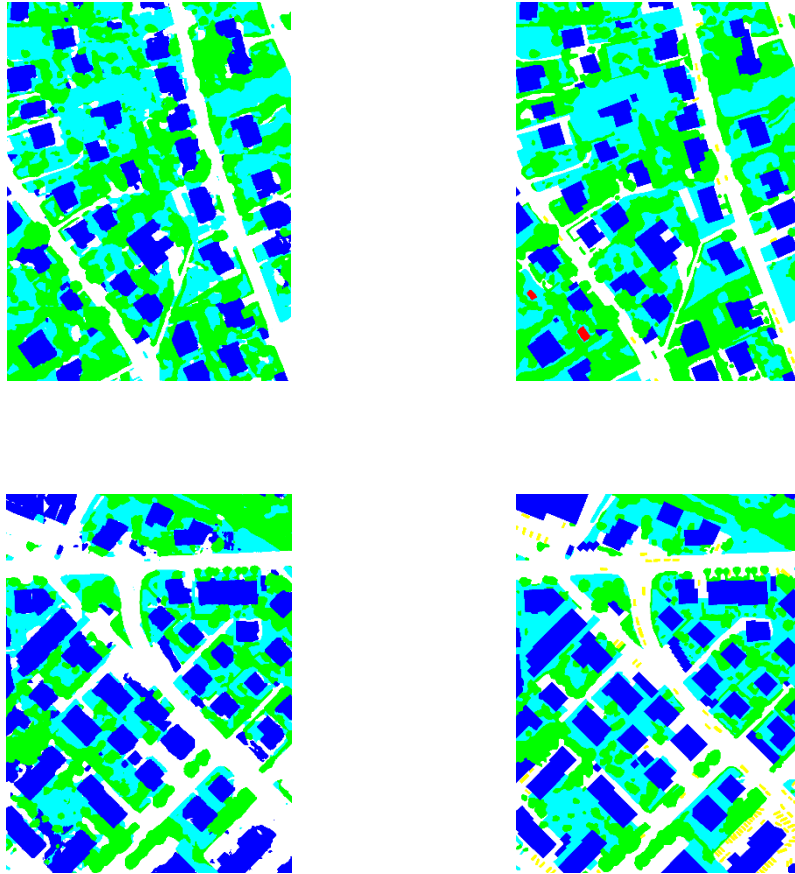
Πίνακας 4. 8: Δεύτερος Έλεγχος Περιοχής 30.

Παρατηρείται ελάχιστη μείωση σε κάποιες τιμές κατά τον δεύτερο έλεγχο, συγκεκριμένα στο accuracy κάθε κλάσης, επηρεάζοντας αναπόφευκτα τον δείκτη macro. Η υψηλή τιμή του δείκτη kappa

αποδίδει αξιοπιστία στην υψηλή τιμή του total accuracy. Παρά τις μέτριες τιμές στη γραμμή του macro, αντιλαμβανόμαστε από τους δείκτες micro και weighted πως κάθε κλάση συμμετέχει στην έκβαση των αποτελεσμάτων σε σχεδόν άριστο βαθμό ενισχύοντας την εμπιστοσύνη των αποτελεσμάτων.

4.6.1.5 Σύγκριση Μασκών Πρόβλεψης του μοντέλου U-Net με την Αληθή Βάση





Σχήμα 4. 4: Αριστερά: Πρόβλεψη. Δεξιά: Αληθής Βάση. Από πάνω προς τα κάτω: Περιοχή 5, Περιοχή 7, Περιοχή 23, Περιοχή 30.

Το εύρος των τιμών kappa και total accuracy κυμαίνεται μεταξύ 0.82 – 0.87 και 87.29% - 92.56%, αντίστοιχα, και κάθε κλάση συμμετέχει σε σχεδόν άριστο βαθμό στα αποτελέσματα αποδίδοντας αξιοπιστία στα αποτελέσματα του μοντέλου.

4.6.2 Αποτελέσματα Marmanis et al.

4.6.2.1 Περιοχή 5

Confusion Matrix

Classes	Roads	Buildings	Low veg.	Trees	Cars	Clutter
Roads	1442481	46224	22208	2069	2815	0
Buildings	200259	2085226	14410	2126	529	0
Low veg.	35845	2362	165716	59562	324	0
Trees	8402	320	15937	267977	0	0
Cars	24467	4809	689	73	14908	0
Clutter	0	0	0	0	0	0

Classification report

Classes	n(truth)	n(classified)	Accuracy	Precision	Recall	F1-Score
Roads	1711454	1515797	92.26%	0.95	0.84	0.89
Buildings	2138941	2302550	92.26%	0.91	0.97	0.94
Low veg.	218960	263809	96.58%	0.63	0.76	0.69
Trees	331807	292636	98%	0.92	0.81	0.86
Cars	18576	44946	99.24%	0.33	0.80	0.47
Clutter	0	0	100%	0.0	0.0	0.0

	Precision	Recall	F1-Score
Micro Average	0.90	0.90	0.90
Macro Average	0.62	0.70	0.66
Weighted Average	0.90	0.90	0.90

Kappa: 0.83

Total Accuracy: 89.97%

Πίνακας 4. 9: Πρώτος Έλεγχος Περιοχής 5.

Confusion Matrix

Classes	Roads	Buildings	Low veg.	Trees	Cars	Clutter
Roads	1442481	46224	22208	2069	2815	0
Buildings	200259	2085226	14410	2126	529	0
Low veg.	35845	2362	165716	59562	324	0
Trees	8402	320	15937	267977	0	0
Cars	24467	4809	689	73	14908	0
Clutter	0	0	0	0	0	0

Classification report

Classes	n(truth)	n(classified)	Accuracy	Precision	Recall	F1-Score
Roads	1711454	1515797	92.26%	0.95	0.84	0.89
Buildings	2138941	2302550	92.26%	0.91	0.97	0.94
Low veg.	218960	263809	96.58%	0.63	0.76	0.69
Trees	331807	292636	98%	0.92	0.81	0.86
Cars	18576	44946	99.24%	0.33	0.80	0.47
Clutter	0	0	100%	0.0	0.0	0.0

	Precision	Recall	F1-Score
Micro Average	0.90	0.90	0.90
Macro Average	0.62	0.70	0.66
Weighted Average	0.90	0.90	0.90

Kappa: 0.83

Total Accuracy: 89.97%

Πίνακας 4. 10: Δεύτερος Έλεγχος Περιοχής 5.

Παρατηρείται σταθερότητα στις τιμές των δεικτών κατά τους δύο ελέγχους. Η υψηλή τιμή του δείκτη kappa αποδίδει αξιοπιστία στην υψηλή τιμή του total accuracy. Παρά τις μέτριες τιμές στη γραμμή του macro, αντιλαμβανόμαστε από τους δείκτες micro και weighted πως κάθε κλάση συμμετέχει στην έκβαση των αποτελεσμάτων σε σχεδόν άριστο βαθμό ενισχύοντας την εμπιστοσύνη των αποτελεσμάτων.

4.6.2.2 Περιοχή 7

Confusion Matrix

Classes	Roads	Buildings	Low veg.	Trees	Cars	Clutter
Roads	1241004	79717	18942	3233	3884	0
Buildings	105318	1096677	8953	8506	102	0
Low veg.	29075	6908	451334	69104	478	0
Trees	7318	2087	146640	1038332	6	0
Cars	42338	3145	939	137	σ28693	0
Clutter	0	0	0	0	0	0

Classification report

Classes	n(truth)	n(classified)	Accuracy	Precision	Recall	F1-Score
Roads	1425053	1346780	93.4%	0.92	0.87	0.90
Buildings	1188534	1219556	95.11%	0.90	0.92	0.91
Low veg.	626808	556899	93.6%	0.81	0.72	0.76
Trees	1119312	1194383	94.6%	0.87	0.93	0.90
Cars	33163	75252	98.84%	0.38	0.87	0.53
Clutter	0	0	100%	0.0	0.0	0.0

	Precision	Recall	F1-Score
Micro Average	0.88	0.88	0.88
Macro Average	0.65	0.72	0.68
Weighted Average	0.88	0.88	0.88

Kappa: 0.83

Total Accuracy: 87.78%

Πίνακας 4. 11: Πρώτος Έλεγχος Περιοχής 7.

Confusion Matrix

Classes	Roads	Buildings	Low veg.	Trees	Cars	Clutter
Roads	2683485	125941	41150	5302	6699	0
Buildings	305577	3181903	23363	10632	631	0
Low veg.	64920	9270	617050	128666	802	0
Trees	15720	2407	162577	1306309	6	0
Cars	66805	7954	1628	210	43601	0
Clutter	0	0	0	0	0	0

Classification report

Classes	n(truth)	n(classified)	Accuracy	Precision	Recall	F1-Score
Roads	3136507	2862577	92.83%	0.94	0.86	0.89
Buildings	3327475	3522106	94.49%	0.90	0.96	0.93
Low veg.	845768	820708	95.09%	0.75	0.73	0.74
Trees	1451119	1487019	96.31%	0.88	0.90	0.89
Cars	51739	120198	99.04%	0.36	0.84	0.51
Clutter	0	0	100%	0.0	0.0	0.0

	Precision	Recall	F1-Score
Micro Average	0.89	0.89	0.89
Macro Average	0.64	0.71	0.67
Weighted Average	0.89	0.89	0.89

Kappa: 0.84

Total Accuracy: 88.88%

Πίνακας 4. 12: Δεύτερος Έλεγχος Περιοχής 7.

Παρατηρείται ελάχιστη μείωση σε κάποιες τιμές κατά τον δεύτερο έλεγχο, συγκεκριμένα στο accuracy κάθε κλάσης, επηρεάζοντας αναπόφευκτα τον δείκτη macro. Η υψηλή τιμή του δείκτη kappa

αποδίδει αξιοπιστία στην υψηλή τιμή του total accuracy. Παρά τις μέτριες τιμές στη γραμμή του macro, αντιλαμβανόμαστε από τους δείκτες micro και weighted πως κάθε κλάση συμμετέχει στην έκβαση των αποτελεσμάτων σε σχεδόν άριστο βαθμό ενισχύοντας την εμπιστοσύνη των αποτελεσμάτων.

4.6.2.3 Περιοχή 23

Confusion Matrix

Classes	Roads	Buildings	Low veg.	Trees	Cars	Clutter
Roads	581822	9967	97338	12579	520	0
Buildings	38087	747218	23620	4949	78	0
Low veg.	100922	7115	990366	87994	5501	0
Trees	12060	3807	252626	1344125	1278	0
Cars	4703	4	386	117	5078	0
Clutter	4516	2	260	0	1614	0

Classification report

Classes	n(truth)	n(classified)	Accuracy	Precision	Recall	F1-Score
Roads	742110	702226	93.53%	0.83	0.78	0.81
Buildings	768113	813952	97.98%	0.92	0.97	0.94
Low veg.	1364596	1191898	86.73%	0.83	0.73	0.77
Trees	1449764	1613896	91.35%	0.83	0.93	0.88
Cars	14069	10288	99.67%	0.49	0.36	0.42
Clutter	0	6392	99.85%	0.0	0.0	0.0

	Precision	Recall	F1-Score
Micro Average	0.85	0.85	0.85
Macro Average	0.65	0.63	0.64
Weighted Average	0.85	0.86	0.85

Kappa: 0.78

Total Accuracy: 84.56%

Πίνακας 4. 13: Πρώτος Έλεγχος Περιοχής 23.

Confusion Matrix

Classes	Roads	Buildings	Low veg.	Trees	Cars	Clutter
Roads	3265307	135908	138488	17881	7219	0
Buildings	343664	3929121	46983	15581	709	0
Low veg.	165842	16385	1607416	216660	6303	0
Trees	27780	6214	415203	2650434	1284	0
Cars	71508	7958	2014	327	48679	0
Clutter	4516	2	260	0	1614	0

Classification report

Classes	n(truth)	n(classified)	Accuracy	Precision	Recall	F1-Score
Roads	3878617	3564803	93.06%	0.92	0.84	0.88
Buildings	4095588	4336058	95.64%	0.91	0.96	0.93
Low veg.	2210364	2012606	92.33%	0.80	0.73	0.76
Trees	2900883	3100915	94.67%	0.85	0.91	0.88
Cars	65808	130486	99.25%	0.37	0.74	0.50
Clutter	0	6392	99.95%	0.0	0.0	0.0

	Precision	Recall	F1-Score
Micro Average	0.87	0.87	0.87
Macro Average	0.64	0.70	0.67
Weighted Average	0.88	0.88	0.88

Kappa: 0.83

Total Accuracy: 87.45%

Πίνακας 4. 14: Δεύτερος Έλεγχος Περιοχής 23.

Παρατηρείται σημαντική βελτίωση στο σύνολο των αποτελεσμάτων κατά τον δεύτερο έλεγχο. Η διακύμανση στον δείκτη kappa αποδίδει μέτρια αξιοπιστία στην υψηλή τιμή του total accuracy. Παρά τις μέτριες τιμές στη γραμμή του macro, αντιλαμβανόμαστε από τους δείκτες micro και weighted πως κάθε κλάση συμμετέχει στην έμβαση των αποτελεσμάτων σε σχεδόν άριστο βαθμό ενισχύοντας την εμπιστοσύνη των αποτελεσμάτων.

4.6.2.4 Περιοχή 30

Confusion Matrix

Classes	Roads	Buildings	Low veg.	Trees	Cars	Clutter
Roads	1152000	27919	71436	11486	1323	0
Buildings	66385	1125852	5487	3151	658	0
Low veg.	98765	32864	698955	127789	1414	0
Trees	4152	1006	72949	896605	22	0
Cars	26617	4984	4487	611	19346	0
Clutter	0	0	0	0	0	0

Classification report

Classes	n(truth)	n(classified)	Accuracy	Precision	Recall	F1-Score
Roads	1347919	1264164	93.09%	0.91	0.85	0.88
Buildings	1192625	1201533	96.8%	0.94	0.94	0.94
Low veg.	853314	959787	90.68%	0.73	0.82	0.77
Trees	1039642	974734	95.04%	0.92	0.86	0.89
Cars	22763	56045	99.1%	0.35	0.85	0.49
Clutter	0	0	100%	0.0	0.0	0.0

	Precision	Recall	F1-Score
Micro Average	0.87	0.87	0.87
Macro Average	0.64	0.72	0.68
Weighted Average	0.87	0.87	0.87

Kappa: 0.83

Total Accuracy: 87.35%

Πίνακας 4. 15: Πρώτος Έλεγχος Περιοχής 30.

Confusion Matrix

Classes	Roads	Buildings	Low veg.	Trees	Cars	Clutter
Roads	4417307	163827	209924	29367	8542	0
Buildings	410049	5054973	52470	18732	1367	0
Low veg.	264607	49249	2306371	344449	7717	0
Trees	31932	7220	488152	3547039	1306	0
Cars	98125	12942	6501	938	68025	0
Clutter	4516	2	260	0	1614	0

Classification report

Classes	n(truth)	n(classified)	Accuracy	Precision	Recall	F1-Score
Roads	5226536	4828967	93.07%	0.91	0.85	0.88
Buildings	5288213	5537591	95.93%	0.91	0.96	0.93
Low veg.	3063678	2972393	91.92%	0.78	0.75	0.76
Trees	3940525	4075649	94.76%	0.87	0.90	0.88
Cars	88571	186531	99.21%	0.36	0.77	0.49
Clutter	0	6392	99.96%	0.0	0.0	0.0

	Precision	Recall	F1-Score
Micro Average	0.87	0.87	0.87
Macro Average	0.64	0.71	0.67
Weighted Average	0.87	0.88	0.88

Kappa: 0.83

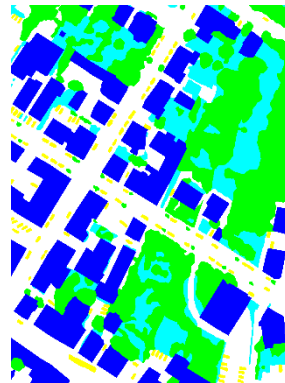
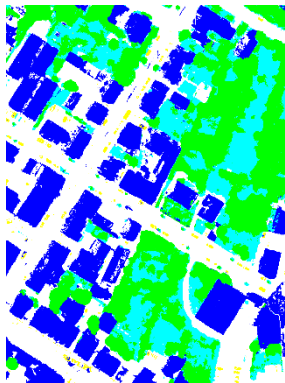
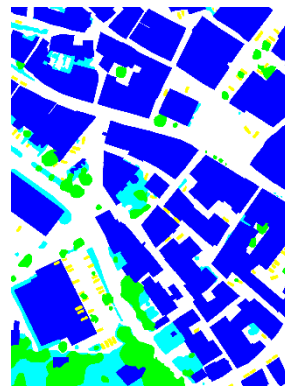
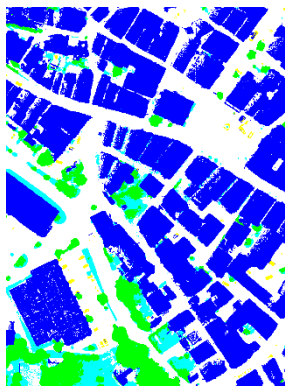
Total Accuracy: 87.43%

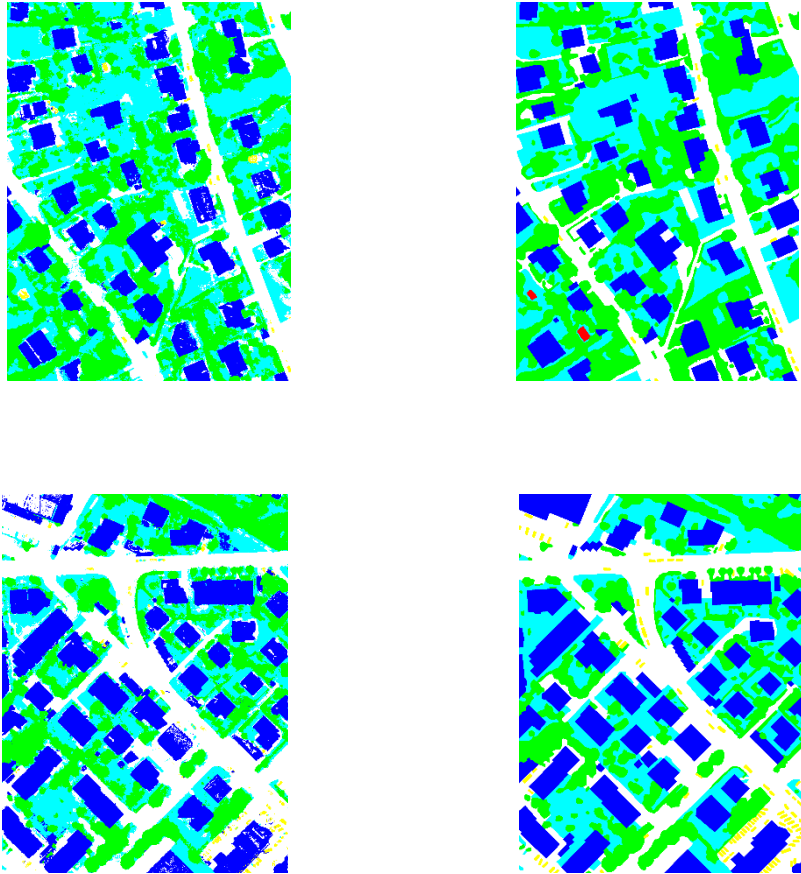
Πίνακας 4. 16: Δεύτερος Έλεγχος Περιοχής 30.

Παρατηρούνται αμελητέες αυξομειώσεις σε κάποιες τιμές και στους δύο ελέγχους, οπότε εξετάζοντας σε βαθμό εικονοστοιχείου, αξιολογούμε πως με την πάροδο του δεύτερου ελέγχου, επιτιμάται

βελτίωση στα αποτελέσματα. Η υψηλή τιμή του δείκτη kappa αποδίδει αξιοπιστία στην υψηλή τιμή του total accuracy. Παρά τις μέτριες τιμές στη γραμμή του macro, αντιλαμβανόμαστε από τους δείκτες micro και weighted πως κάθε κλάση συμμετέχει στην έκβαση των αποτελεσμάτων σε σχεδόν άριστο βαθμό ενισχύοντας την εμπιστοσύνη των αποτελεσμάτων.

4.6.2.5 Σύγκριση Μασκών Πρόβλεψης του μοντέλου Marmanis et al. με την Αληθή Βάση





Σχήμα 4. 5: Αριστερά: Πρόβλεψη. Δεξιά: Αληθής Βάση. Από πάνω προς τα κάτω: Περιοχή 5, Περιοχή 7, Περιοχή 23, Περιοχή 30.

Το εύρος των τιμών kappa και total accuracy κυμαίνεται μεταξύ 0.78 – 0.84 και 84.56% - 89.97%, αντίστοιχα, και κάθε κλάση συμμετέχει σε σχεδόν άριστο βαθμό στα αποτελέσματα αποδίδοντας αξιοπιστία στα αποτελέσματα του μοντέλου.

4.6.3 Αποτελέσματα Προτεινόμενου Μοντέλου

4.6.3.1 Περιοχή 5

Confusion Matrix

Classes	Roads	Buildings	Low veg.	Trees	Cars	Clutter
Roads	1399709	90065	15517	3309	7197	0
Buildings	27506	2271080	3047	860	57	0
Low veg.	42443	6726	153528	61002	110	0
Trees	6599	975	8227	276811	24	0
Cars	6852	5812	39	95	32148	0
Clutter	0	0	0	0	0	0

Classification report

Classes	n(truth)	n(classified)	Accuracy	Precision	Recall	F1-Score
Roads	1483109	1515797	95.49%	0.92	0.94	0.93
Buildings	2374658	2302550	96.94%	0.99	0.96	0.97
Low veg.	180358	263809	96.9%	0.58	0.85	0.69
Trees	342077	292636	98.17%	0.95	0.81	0.87
Cars	39536	44946	99.54%	0.72	0.81	0.76
Clutter	0	0	100%	0.0	0.0	0.0

	Precision	Recall	F1-Score
Micro Average	0.94	0.94	0.94
Macro Average	0.69	0.73	0.71
Weighted Average	0.94	0.94	0.94

Kappa: 0.89

Total Accuracy: 93.52%

Πίνακας 4. 17: Πρώτος Έλεγχος Περιοχής 5.

Confusion Matrix

Classes	Roads	Buildings	Low veg.	Trees	Cars	Clutter
Roads	1399709	90065	15517	3309	7197	0
Buildings	27506	2271080	3047	860	57	0
Low veg.	42443	6726	153528	61002	110	0
Trees	6599	975	8227	276811	24	0
Cars	6852	5812	39	95	32148	0
Clutter	0	0	0	0	0	0

Classification report

Classes	n(truth)	n(classified)	Accuracy	Precision	Recall	F1-Score
Roads	1483109	1515797	95.49%	0.92	0.94	0.93
Buildings	2374658	2302550	96.94%	0.99	0.96	0.97
Low veg.	180358	263809	96.9%	0.58	0.85	0.69
Trees	342077	292636	98.17%	0.95	0.81	0.87
Cars	39536	44946	99.54%	0.72	0.81	0.76
Clutter	0	0	100%	0.0	0.0	0.0

	Precision	Recall	F1-Score
Micro Average	0.94	0.94	0.94
Macro Average	0.69	0.73	0.71
Weighted Average	0.94	0.94	0.94

Kappa: 0.89

Total Accuracy: 93.52%

Πίνακας 4. 18: Δεύτερος Έλεγχος Περιοχής 5.

Παρατηρείται σταθερότητα στις τιμές των δεικτών κατά τους δύο ελέγχους. Η υψηλή τιμή του δείκτη kappa αποδίδει αξιοπιστία στην υψηλή τιμή του total accuracy. Οι υψηλές τιμές στη γραμμή του macro κι οι πιο υψηλές τιμές στους δείκτες micro και weighted μας δείχνουν πως κάθε κλάση συμμετέχει στην έκβαση των αποτελεσμάτων σε σχεδόν άριστο βαθμό ενισχύοντας την εμπιστοσύνη των αποτελεσμάτων.

4.6.3.2 Περιοχή 7

Confusion Matrix

Classes	Roads	Buildings	Low veg.	Trees	Cars	Clutter
Roads	1202631	117657	15983	4580	5929	0
Buildings	14572	1192537	3889	8558	0	0
Low veg.	25819	8971	442108	79992	9	0
Trees	7506	1761	109175	1075940	1	0
Cars	15040	5716	259	153	54084	0
Clutter	0	0	0	0	0	0

Classification report

Classes	n(truth)	n(classified)	Accuracy	Precision	Recall	F1-Score
Roads	1265568	1346780	95.29%	0.89	0.95	0.92
Buildings	1326642	1219556	96.33%	0.98	0.90	0.94
Low veg.	571414	556899	94.44%	0.79	0.77	0.78
Trees	1169223	1194383	95.18%	0.90	0.92	0.91
Cars	60023	75252	99.38%	0.72	0.90	0.80
Clutter	0	0	100%	0.0	0.0	0.0

	Precision	Recall	F1-Score
Micro Average	0.90	0.90	0.90
Macro Average	0.71	0.74	0.73
Weighted Average	0.90	0.90	0.90

Kappa: 0.87

Total Accuracy: 90.31%

Πίνακας 4. 19: Πρώτος Έλεγχος Περιοχής 7.

Confusion Matrix

Classes	Roads	Buildings	Low veg.	Trees	Cars	Clutter
Roads	2602340	207722	31500	7889	13126	0
Buildings	42078	3463617	6936	9418	57	0
Low veg.	68262	15697	595636	140994	119	0
Trees	14105	2736	117402	1352751	25	0
Cars	21892	11528	298	248	86232	0
Clutter	0	0	0	0	0	0

Classification report

Classes	n(truth)	n(classified)	Accuracy	Precision	Recall	F1-Score
Roads	2748677	2862577	95.39%	0.91	0.95	0.93
Buildings	3701300	3522106	96.64%	0.98	0.94	0.96
Low veg.	751772	820708	95.67%	0.73	0.79	0.76
Trees	1511300	1487019	96.68%	0.91	0.90	0.90
Cars	99559	120198	99.46%	0.72	0.87	0.78
Clutter	0	0	100%	0.0	0.0	0.0

	Precision	Recall	F1-Score
Micro Average	0.92	0.92	0.92
Macro Average	0.71	0.74	0.72
Weighted Average	0.92	0.92	0.92

Kappa: 0.88

Total Accuracy: 91.92%

Πίνακας 4. 20: Δεύτερος Έλεγχος Περιοχής 7.

Παρατηρείται βελτίωση στο σύνολο των αποτελεσμάτων κατά τον δεύτερο έλεγχο. Η υψηλή τιμή του δείκτη kappa αποδίδει αξιοπιστία στην υψηλή τιμή του total accuracy. Οι υψηλές τιμές στη γραμμή

του macro κι οι πιο υψηλές τιμές στους δείκτες micro και weighted μας δείχνουν πως κάθε κλάση συμμετέχει στην έμβαση των αποτελεσμάτων σε σχεδόν άριστο βαθμό ενισχύοντας την εμπιστοσύνη των αποτελεσμάτων.

4.6.3.3 Περιοχή 23

Confusion Matrix

Classes	Roads	Buildings	Low veg.	Trees	Cars	Clutter
Roads	590430	17305	78655	15485	351	0
Buildings	8407	787065	15774	2706	0	0
Low veg.	56807	28780	994691	111313	307	0
Trees	7440	7512	178330	1420614	0	0
Cars	1179	0	0	67	9042	0
Clutter	4194	0	2198	0	0	0

Classification report

Classes	n(truth)	n(classified)	Accuracy	Precision	Recall	F1-Score
Roads	668457	702226	95.62%	0.84	0.88	0.86
Buildings	840662	813952	98.14%	0.97	0.94	0.95
Low veg.	1269648	1191898	89.12%	0.83	0.78	0.81
Trees	1550185	1613896	92.56%	0.88	0.92	0.90
Cars	9700	10288	99.96%	0.88	0.93	0.90
Clutter	0	6392	99.85%	0.0	0.0	0.0

	Precision	Recall	F1-Score
Micro Average	0.88	0.88	0.88
Macro Average	0.73	0.74	0.74
Weighted Average	0.88	0.88	0.88

Kappa: 0.83

Total Accuracy: 87.63%

Πίνακας 4. 21: Πρώτος Έλεγχος Περιοχής 23.

Confusion Matrix

Classes	Roads	Buildings	Low veg.	Trees	Cars	Clutter
Roads	3192770	225027	110155	23374	13477	0
Buildings	50485	4250682	22710	12124	57	0
Low veg.	125069	44477	1590327	252307	426	0
Trees	21545	10248	295732	2773365	25	0
Cars	23071	11528	298	315	95274	0
Clutter	4194	0	2198	0	0	0

Classification report

Classes	n(truth)	n(classified)	Accuracy	Precision	Recall	F1-Score
Roads	3417134	3564803	95.47%	0.90	0.93	0.91
Buildings	4541962	4336058	97.14%	0.98	0.94	0.96
Low veg.	2021420	2012606	93.51%	0.79	0.79	0.79
Trees	3061485	3100915	95.32%	0.89	0.91	0.90
Cars	109259	130486	99.63%	0.73	0.87	0.79
Clutter	0	6392	99.95%	0.0	0.0	0.0

	Precision	Recall	F1-Score
Micro Average	0.91	0.91	0.91
Macro Average	0.72	0.74	0.73
Weighted Average	0.91	0.91	0.91

Kappa: 0.87

Total Accuracy: 90.5%

Πίνακας 4. 22: Δεύτερος Έλεγχος Περιοχής 23.

Παρατηρείται σημαντική βελτίωση στο σύνολο των αποτελεσμάτων κατά τον δεύτερο έλεγχο. Η υψηλή τιμή του δείκτη kappa αποδίδει αξιοπιστία στην υψηλή τιμή του total accuracy. Οι υψηλές τιμές στη γραμμή του macro κι οι πιο υψηλές τιμές στους δείκτες micro και weighted μας δείχνουν πως κάθε κλάση συμμετέχει στην έμβαση των αποτελεσμάτων σε σχεδόν άριστο βαθμό ενισχύοντας την εμπιστοσύνη των αποτελεσμάτων.

4.6.3.4 Περιοχή 30

Confusion Matrix

Classes	Roads	Buildings	Low veg.	Trees	Cars	Clutter
Roads	1154007	39089	52320	13988	4760	0
Buildings	8291	1186829	5062	1350	1	0
Low veg.	49284	62081	707059	140641	722	0
Trees	2941	1545	47538	922708	2	0
Cars	7056	3637	423	238	44691	0
Clutter	0	0	0	0	0	0

Classification report

Classes	n(truth)	n(classified)	Accuracy	Precision	Recall	F1-Score
Roads	1221579	1264164	96.01%	0.91	0.94	0.93
Buildings	1293181	1201533	97.28%	0.99	0.92	0.95
Low veg.	812402	959787	91.96%	0.74	0.87	0.80
Trees	1078925	974734	95.33%	0.95	0.86	0.90
Cars	50176	56045	99.62%	0.80	0.89	0.84
Clutter	0	0	100%	0.0	0.0	0.0

	Precision	Recall	F1-Score
Micro Average	0.90	0.90	0.90
Macro Average	0.73	0.75	0.74
Weighted Average	0.90	0.90	0.90

Kappa: 0.87

Total Accuracy: 90.1%

Πίνακας 4. 23: Πρώτος Έλεγχος Περιοχής 30.

Confusion Matrix

Classes	Roads	Buildings	Low veg.	Trees	Cars	Clutter
Roads	4346777	264116	162475	37362	18237	0
Buildings	58776	5437511	27772	13474	58	0
Low veg.	174353	106558	2297386	392948	1148	0
Trees	24486	11793	343270	3696073	27	0
Cars	30127	15165	721	553	139965	0
Clutter	4194	0	2198	0	0	0

Classification report

Classes	n(truth)	n(classified)	Accuracy	Precision	Recall	F1-Score
Roads	4638713	4828967	95.6%	0.90	0.94	0.92
Buildings	5835143	5537591	97.17%	0.98	0.93	0.96
Low veg.	2833822	2972393	93.12%	0.77	0.81	0.79
Trees	4140410	4075649	95.32%	0.91	0.89	0.90
Cars	159435	186531	99.62%	0.75	0.88	0.81
Clutter	0	6392	99.96%	0.0	0.0	0.0

	Precision	Recall	F1-Score
Micro Average	0.90	0.90	0.90
Macro Average	0.72	0.74	0.73
Weighted Average	0.90	0.90	0.90

Kappa: 0.87

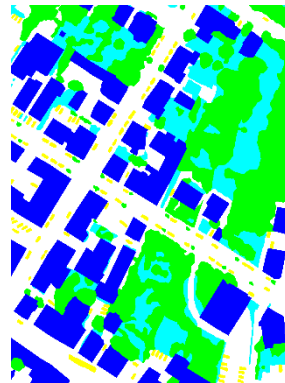
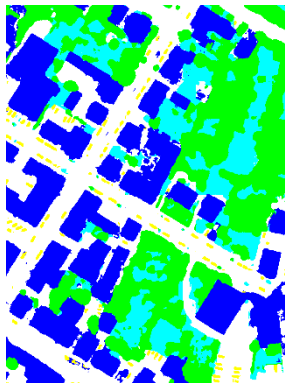
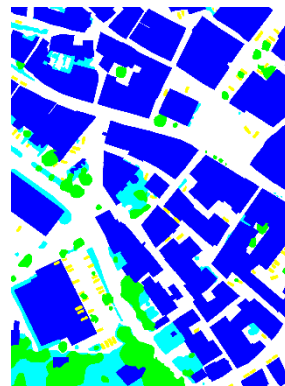
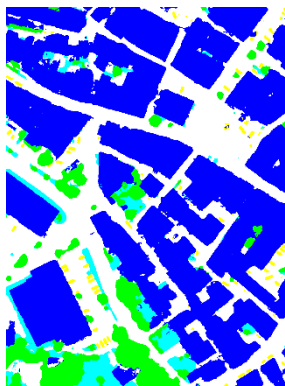
Total Accuracy: 90.4%

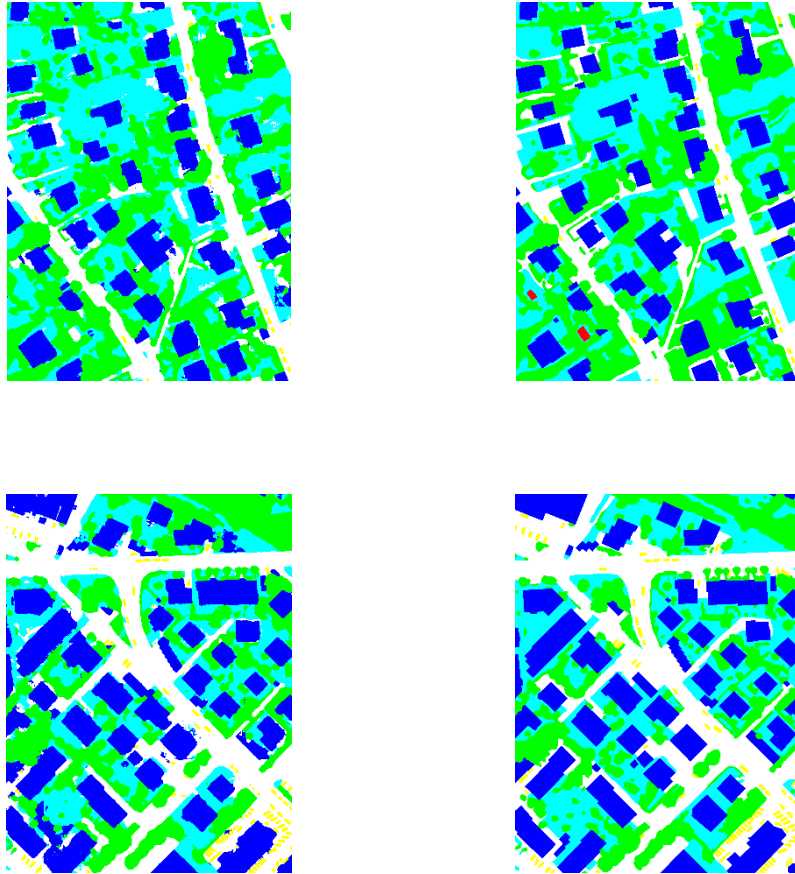
Πίνακας 4. 24: Δεύτερος Έλεγχος Περιοχής 30.

Παρατηρείται βελτίωση στο σύνολο των αποτελεσμάτων κατά τον δεύτερο έλεγχο, πλην του δείκτη macro, καθότι προκύπτουν απειροελάχιστες αυξομειώσεις στο εύρος του συνόλου του δείκτη

accuracy. Η υψηλή τιμή του δείκτη kappa αποδίδει αξιοπιστία στην υψηλή τιμή του total accuracy. Οι υψηλές τιμές στη γραμμή του macro κι οι πιο υψηλές τιμές στους δείκτες micro και weighted μας δείχνουν πως κάθε κλάση συμμετέχει στην έκβαση των αποτελεσμάτων σε σχεδόν άριστο βαθμό ενισχύοντας την εμπιστοσύνη των αποτελεσμάτων.

4.6.3.5 Σύγκριση Μασκών Πρόβλεψης του Μοντέλου μας με την Αληθή Βάση





Σχήμα 4. 6: Αριστερά: Πρόβλεψη. Δεξιά: Αληθής Βάση. Από πάνω προς τα κάτω: Περιοχή 5, Περιοχή 7, Περιοχή 23, Περιοχή 30.

Το εύρος των τιμών kappa και total accuracy κυμαίνεται μεταξύ 0.83 – 0.89 και 87.63% - 93.52%, αντίστοιχα, και κάθε κλάση συμμετέχει σε άριστο βαθμό στα αποτελέσματα αποδίδοντας αξιοπιστία στα αποτελέσματα του μοντέλου.

4.7 Αξιολόγηση Αποτελεσμάτων

Παρατηρούμε ότι κατά τη διάρκεια των ελέγχων, τα αποτελέσματα μας βελτιώνονται σημαντικά στην πλειοψηφία τους. Στο μοντέλο που έχει προταθεί από τους Marmanis et al. και στο μοντέλο U-Net, αποκαλύπτονται κάποιες αδυναμίες στην προσπάθεια ταξινόμησης και των 6 κλάσεων, η οποία γίνεται

εμφανής κατά την προσπάθεια οπτικοποίησης της κλάσης «Συσσωρευση/Φόντο», όσον αφορά το πρώτο μοντέλο και στο ποσοστό συμμετοχής και κατ' επέκταση της οπτικοποίησης τους, στις κλάσεις «Αυτοκίνητο» και «Συσσωρευση/Φόντο», όσον αφορά το δεύτερο.

Συγκεκριμένα, προκύπτουν οι μέσοι όροι του δείκτη kappa 0.86, 0.83 και 0.87 για το μοντέλο U-Net, το μοντέλο των Marmanis et al. και το δικό μας, αντίστοιχα, και οι μέσοι όροι του δείκτη total accuracy 90.16%, 88% και 91%, για τα αντίστοιχα δίκτυα. Οι μέσοι όροι των δύο αυτών δεικτών, με κύριο το kappa, μας επιβεβαιώνουν την υπόθεσή μας, πως ο συνδυασμός των δύο προηγούμενων μοντέλων, μπορεί να αποφέρει ένα ισχυρό διατεματικό δίκτυο Σηματολογικής Κατάτμησης ακατέργαστων δεδομένων.

Πιο αναλυτικά, παρατηρούμε ότι κατά τη διάρκεια εκπαίδευσης του μοντέλου U-Net, δεν συμμετέχει ούτε η κλάση «Αυτοκίνητο», αλλά ούτε κι η κλάση «Συσσωρευση/Φόντο». Αυτό προκύπτει από την εξέταση των πινάκων σύγχυσης, σε κάθε έλεγχο των περιοχών, όπου τα εικονοστοιχεία που συμμετέχουν είναι μηδενικά. Αυτό συμβάλλει στους μέτριους μακρινούς μέσους όρους κι εν τέλει, στην ίδια την εκπαίδευση που οδηγεί στην αξιολόγηση του μοντέλου, αφού δεν μπορούμε να αρκестούμε μόνο στους δείκτες της ολικής ορθότητας και του συντελεστή kappa. Παρά τις υψηλές τιμές των δύο τελευταίων, το ποσοστό συμμετοχής κάθε κλάσης στο μοντέλο παραμένει σε μέτρια επίπεδα, παρουσιάζοντας μια ελάχιστη βελτίωση με το πέρας των εποχών. Λεπτομερώς, η κλάση που συμμετέχει περισσότερο είναι η κλάση «Κτίριο» με τιμή που αγγίζει το 0.98 σε κάθε έλεγχο. Ακολουθούν, με μικρές μεταξύ τους διαφορές, οι κλάσεις «Δρόμοι» και «Δέντρο» με τιμές συμμετοχής κοντά στο 0.91 και τέλος, η κλάση «Χαμηλή Βλάστηση», με τιμή κοντά στο 0.73. Οι μέσες τιμές του αρμονικού μέσου F1, αγγίζουν την τιμή 0.81. Αυτό μας δείχνει ότι πρόκειται για ένα μοντέλο, του οποίου τα αποτελέσματα είναι ικανοποιητικά, αλλά θα χρειαζόταν αρκετές βελτιώσεις για να συμμετάσχει έμπιστα στις εργασίες ταξινόμησης.

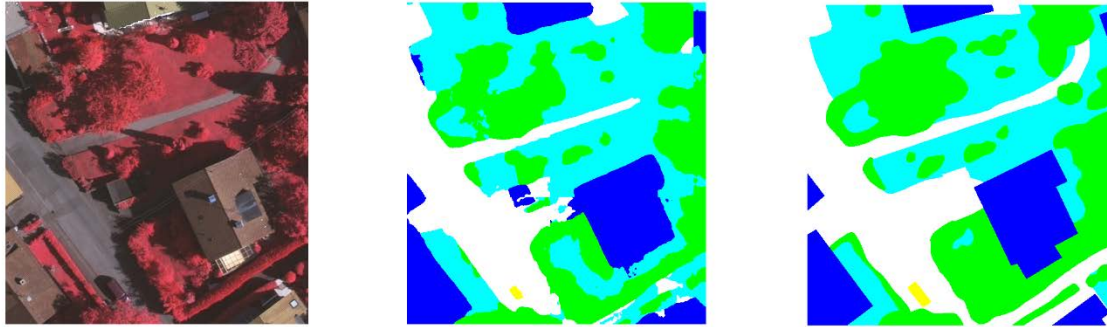
Ακολουθεί το μοντέλο των Marmanis et al. Στο συγκεκριμένο μοντέλο παρατηρείται η έναρξη συμμετοχής των κλάσεων «Αυτοκίνητο» και «Συσσωρευση/Φόντο», που αδυνατούσε να ταξινομήσει το μοντέλο U-Net. Αυτό συμβάλλει άμεσα στην αύξηση των δεικτών των μακρινών μέσων όρων, δίνοντας πιο ενθαρρυντικά αποτελέσματα για την αξιοπιστία του μοντέλου. Συγκριτικά με το προηγούμενο μοντέλο, έχουμε αύξηση συμμετοχής της κάθε κλάσης, θέτοντας το μοντέλο ήδη αρκετά έμπιστο για διαδικασίες ταξινόμησης εικονοστοιχείων. Συγκεκριμένα, για το μοντέλο των Marmanis et al. με μια πρώτη εξέταση παρατηρείται πτώση τόσο της ολικής ορθότητας, όσο και του συντελεστή kappa. Αν, όμως, εξετάσουμε προσεκτικά τους πίνακές μας, παρατηρούμε σημαντικά υψηλές τιμές στη συμμετοχή των κλάσεων, υπενθυμίζοντας για ποιο λόγο δεν πρέπει να αρκούμε μόνο στους δύο προαναφερθέντες δείκτες. Με τιμές συμμετοχής στην κλάση «Κτίριο» που φτάνουν

το 0.92, στις κλάσεις «Δρόμοι» και «Δέντρο» με τιμές συμμετοχής κοντά στο 0.91, με μικρές μεταξύ τους διαφορές, στην κλάση «Χαμηλή Βλάστηση» με τιμή συμμετοχής κοντά στο 0.75 και στην κλάση «Αυτοκίνητο» κοντά στο 0.37, αλλά και με τη σταθερότητα σε σχετικά αρκετά υψηλές τιμές στους δείκτες ακρίβειας κι ανάκλησης, που προκύπτει από τις μέσες τιμές του αρμονικού μέσου F1, που αγγίζουν την τιμή 0.81, παρατηρούμε πως πρόκειται για ένα μοντέλο εμπιστοσύνης στην διαδικασία ταξινόμησης.

Τελευταίο είναι το μοντέλο που προτείνουμε. Έχοντας κρατήσει τα χαρακτηριστικά του μοντέλου Marmanis et al και του μοντέλου U-Net, παρατηρείται κι εδώ η συμμετοχή των κλάσεων «Αυτοκίνητο» και «Συσσώρευση/Φόντο», αλλά κι η σταθερή βελτίωση των τιμών των μετρικών που έχουν χρησιμοποιηθεί. Η αδυναμία οπτικοποίησης της τελευταίας, παραμένει κι εδώ ένα πρόβλημα, το οποίο, όμως, μπορεί να επιλυθεί, όπως θα αναλύσουμε παρακάτω. Συγκριτικά με τα δύο προηγούμενα έχουμε, αρχικά, αύξηση της ολικής ορθότητας και του συντελεστή kappa. Έπειτα, εξετάζοντας, πιο προσεκτικά τους πίνακες, παρατηρούμε μικρή αύξηση συμμετοχής των κλάσεων. Η κλάση «Κτίριο» συμμετέχει με τιμή κοντά στο 0.99, η κλάση «Δρόμοι» και «Δέντρο», συμμετέχουν με τιμές μεταξύ 0.90 και 0.92. κι οι κλάσεις «Χαμηλή Βλάστηση» και «Αυτοκίνητο», συμμετέχουν με τιμές μεταξύ 0.74 και 0.76. Οι μέσες τιμές του αρμονικού μέσου F1, αγγίζουν την τιμή 0.85, οπότε μπορούμε εύλογα να αποφανθούμε, βάσει όλων των αποτελεσμάτων, ότι πρόκειται για ένα πολύ έμπιστο μοντέλο για τη διαδικασία της ταξινόμησης των εικονοστοιχείων στις αντίστοιχες κλάσεις τους και με τις κατάλληλες ρυθμίσεις στις υπερπαραμέτρους και το αντίστοιχο υλικό υπολογιστή, είναι σε θέση να εξάγει άριστα αποτελέσματα. Η κλάση «Συσσώρευση/Φόντο», μπορεί στους δείκτες ακρίβειας κι ανάκλησης να μην έχει τιμές, αλλά παρατηρούμε στον πίνακα σύγχυσης, πως σε όποια εικόνα αληθούς βάσης συμμετέχει, τα εικονοστοιχεία της προσπαθούν να ταξινομηθούν και στο δικό μας μοντέλο.

Στις μάσκες πρόβλεψης, μετά τα στάδια της Συγκέντρωσης και την Ανάστροφης Συνέλιξης, μπορεί να δημιουργηθούν ελλείψεις ευκρίνειας κι αιχμηρότητας στις άκρες σε κάποια σημεία, κάτι αναμενόμενο γιατί αρκετά χαρακτηριστικά των εικόνων χάνονται κατά τη διαδικασία συστολής και δεν μπορούν ανακτηθούν στη διαδικασία διαστολής. Αυτό γίνεται έντονα εμφανές στο μοντέλο των Marmanis et al., όπου η οπτική απόδοση των μασκών έχει μια τραχύτητα και δεν εμφανίζεται συνέχεια στον τρόπο αποτύπωσης του χρώματος κάθε κλάσης. Το μοντέλο U-Net, παρά της μέτριας συμμετοχής των κλάσεων, έχει μια πιο αιχμηρή απεικόνιση. Ο συνδυασμός και των δύο, δίνει πράγματι τις πιο αντιπροσωπευτικές απεικονίσεις, εφόσον, όπως παρατηρούμε στις μάσκες πρόβλεψης της πρότασής μας, διατηρούμε τα χαρακτηριστικά του μοντέλου των Marmanis et al. με την χρωματική απόδοση της κλάσης «Αυτοκίνητο» και τα χαρακτηριστικά του μοντέλου U-Net, όπου διατηρείται η συνέχεια κι η αιχμηρότητα της οπτικής απόδοσης των κλάσεων.

Τέλος, επιθεωρώντας τα δεδομένα που μας παρέχονται, αποκαλύπτονται γρήγορα ορισμένα λάθη στην αληθή βάση. Σε αρκετές περιπτώσεις το μοντέλο μας καταφέρει να κάνει καλύτερη ταξινόμηση, όπως φαίνεται στην Σχήμα 4. 7. Αυτές οι αστοχίες στην αληθή βάση, μπορεί να προκύψουν από την ανακρίβεια των αναλυτών σε συνδυασμό με την αβεβαιότητα της τοποθεσίας των χαρακτηριστικών.



Σχήμα 4. 7: Από αριστερά στα δεξιά: Αρχική Εικόνα, Πρόβλεψη, Αληθής Βάση

Επίσης, όσον αφορά την κλάση «Συσσώρευση/Φόντο», παρατηρείται μια δυσκολία απόδοσης και στην αληθή βάση, καθώς υπάρχουν ίδια ή παρόμοια τμήματα που έχουν αποδοθεί με την κλάση «Κτίριο» σε κάποια σημεία, ενώ σε άλλα έχουν αποδοθεί με την κλάση «Συσσώρευση/Φόντο». Αυτό συμβάλλει σε κάποια αρνητικά αποτελέσματα όλων των μοντέλων, όπου η ταξινόμηση των εικονοστοιχείων έχει γίνει ορθά, αλλά ο στόχος μας είναι λάθος, αξιολογώντας τα, έτσι, ως μη ισχυρά, ερμηνεία που δεν ισχύει σε ορισμένες περιπτώσεις.

ΚΕΦΑΛΑΙΟ 5

Επίλογος

5.1 Συμπεράσματα

Η παρούσα εργασία είχε ως στόχο την πρόταση ενός νέου διατετακτικού Συνελικτικού Νευρωνικού Δικτύου για τη Σηματολογική Κατάτμηση εναερίων τηλεπισκοπικών δεδομένων υψηλής ανάλυσης.

Παρουσιάστηκαν τα μοντέλα στα οποία βασιστήκαμε, τόσο για την αρχιτεκτονική, όσο και για την εκπαίδευση του δικού μας. Έχοντας θέσει τις απαραίτητες μεταβλητές για την υλοποίηση του αλγορίθμου μας, μέσω ευρετικής διαδικασίας, ακολούθησαν τα πειράματα κι η σύγκριση των αποτελεσμάτων. Παρατηρήθηκε ότι το μοντέλο U-Net έδινε καλύτερα αποτελέσματα έναντι των Marmanis et al., κάτι που αναμέναμε εφόσον το τελευταίο είχε δημιουργηθεί για λιγότερες κλάσεις και χαμηλότερα δεδομένα εισόδου.

Βάσει αυτών των αποτελεσμάτων αποφασίστηκε να γίνει μια προσπάθεια ένωσης των δύο μοντέλων με σκοπό την βέλτιστη απόδοση και την εξαγωγή λεπτομερούς ταξινόμησης. Πράγματι, ο συνδυασμός τους επέφερε ισχυρά αποτελέσματα και καλύτερη ανάδειξη όλων των κλάσεων, πράγμα αρκετά ενθαρρυντικό αν αναλογιστούμε πως προέκυψαν μέσω μεταβλητών που απαρτίζουν μια απλή εκπαίδευση. Το πρόβλημα της οπτικής απόδοσης της κλάσης «Συσώρευση/Φόντο» μπορεί να επιμένει, αλλά τα αποτελέσματα του μοντέλου μας είναι ενθαρρυντικά κι εφόσον υλοποιηθεί με αποδοτικότερες υπερπαραμέτρους, θα δοθεί η δυνατότητα κι οπτικής της αντιπροσώπευσης.

Επιπρόσθετα, η εργασία μας έχει ενδιαφέρον για μελλοντικές έρευνες, όπου λόγω χαμηλού προϋπολογισμού δεν καθίσταται δυνατή η πρόσβαση σε υπερσύγχρονα μέσα υλοποίησης, απόκτησης μεγάλου όγκου δεδομένων υψηλής ανάλυσης και φυσικά εφαρμογών που έχουν αναπτυχθεί για την ανάλυση κι επεξεργασία εικόνων. Αξιόπιστα κι αξιόλογα αποτελέσματα, μπορούν πλέον να εξαχθούν και με λιγότερες εποχές, μικρότερα μεγέθη παρτίδας και σύνολα δεδομένων, παρέχοντας ευκαιρίες βελτιστοποίησης των μοντέλων και κατ' επέκταση ανάπτυξη του προκείμενου επιστημονικού πεδίου.

5.2 Προτάσεις Βελτίωσης

Για τη βελτίωση της πρότασης μας, μια σημαντική παρέμβαση, θα μπορούσε να γίνει στην ίδια την αρχιτεκτονική του μοντέλου μας. Κάποιες συσσωρεύσεις κρυφών επιπέδων, στο πέρας του μονοπατιού συστολής, έναντι των Πλήρως Συνδεδεμένων επιπέδων, θα μπορούσαν να διατηρήσουν κάποια από τα υψηλά χαρακτηριστικά των εικόνων, ώστε η αναγκαία χωρική πληροφορία για την ταξινόμηση των εικονοστοιχείων να μη χαθεί.

Φυσικά, μια άλλη πρόταση είναι η υλοποίηση του δικτύου με διαφορετικά δεδομένα. Αν πρόκειται για το ίδιο σύνολο που χρησιμοποιήθηκε, θα μπορούσε να αντικατασταθεί το nDSM, με το DEM, που προτείνεται από τους Marmanis et al. Επίσης, θα μπορούσαμε να αλλάξουμε εντελώς το σύνολο δεδομένων και να χρησιμοποιήσουμε ένα ακόμη πιο σύγχρονο, στο οποίο θα έχει επιβεβαιωθεί πως οι ανθρώπινες αστοχίες κατάτμησης είναι μηδαμινές έως μηδενικές.

Με τα υπάρχοντα δεδομένα, μια βελτίωση θα μπορούσε να παρουσιαστεί παρεμβαίνοντας στις κλάσεις, ακολουθώντας μια διαφορετική προσέγγιση, όπου αντί των 6 υπαρχόντων, να παραλειπόταν εκείνη που συμμετείχε λιγότερο κι αδυνατούσε να οπτικοποιηθεί, όπως η κλάση «Συσώρευση/Φόντο».

Ακόμη, επωφελομένη παρέμβαση θα ήταν κι η επανεξέταση του συνόλου της αληθούς βάσης, καθώς υπάρχουν αρκετές αστοχίες που συγκρίνοντας τα αποτελέσματα διαφόρων ερευνών δε θεωρούνται αμελητέες κι ελλοχεύουν το κίνδυνο να καταστήσουν το σύνολο ξεπερασμένο.

Μια άλλη πολύ σημαντική βελτίωση της παρούσας μελέτης, θα ήταν η εκπαίδευση του μοντέλου σε πιο ισχυρό υλικό υπολογιστή. Θα δινόταν η δυνατότητα να αυξηθούν σε σημαντικό βαθμό ο αριθμός των εποχών και το μέγεθος της παρτίδας και βέβαια, να χρησιμοποιηθεί μεγαλύτερου μεγέθους παράθυρο εικονοστοιχείων, παρέχοντας υψηλότερη ανάλυση. Φυσικά, θα επέρχονταν σημαντικές αλλαγές και στις μετρικές του δικτύου, όπου πάλι με ευρετικό τρόπο, θα καταλήγαμε στις επιθυμητές τιμές.

Βιβλιογραφία

- [1] 2D Semantic Labeling Contest. Retrieved from <https://www2.isprs.org/commissions/comm2/wg4/benchmark/semantic-labeling/>
- [2] Albawi, S., Mohammed, T. A., & Al-Zawi, S. Understanding of a convolutional neural network. *2017 International Conference on Engineering and Technology (ICET)*, 2017. doi: [10.1109/ICEngTechnol.2017.8308186](https://doi.org/10.1109/ICEngTechnol.2017.8308186)
- [3] Audebert, N., Le Saux, B., & Lefèvre, S. Beyond RGB: Very High Resolution Urban Remote Sensing With Multimodal Deep Networks. *arXiv preprint [arXiv:1711.08681v1](https://arxiv.org/abs/1711.08681v1) [cs.NE]*, 2017.
- [4] Bottou, L. Large-Scale Machine Learning with Stochastic Gradient Descent. *Proceedings of COMPSTAT'2010*, pp. 177–186, 2010. doi: [10.1007/978-3-7908-2604-3_16](https://doi.org/10.1007/978-3-7908-2604-3_16)
- [5] Chaturvedi, DK. Artificial Neural Network and Supervised Learning. *Soft Computing: Techniques and its Applications in Electrical Engineering*, Berlin Heidelberg: Springer, pp. 23–50, 2018. doi: [10.1007/978-3-540-77481-5_3](https://doi.org/10.1007/978-3-540-77481-5_3)
- [6] Chen, T. (2017). *What is “padding” in Convolutional Neural Network?*. Retrieved from <https://medium.com/machine-learning-algorithms/what-is-padding-in-convolutional-neural-network-c120077469cc>
- [7] Dumoulin, V., & Visin, F. A guide to convolution arithmetic for deep learning. *arXiv preprint [arXiv:1603.07285v2](https://arxiv.org/abs/1603.07285v2) [stat.ML]*, 2016.
- [8] Fleiss, J. L. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5), pp. 378–382, 1971. doi: [10.1037/h0031619](https://doi.org/10.1037/h0031619).
- [9] Gerke, M. Normalized DSM-Heights Encoded in Dm-See Report for Details. 2014. Retrieved from https://www.researchgate.net/profile/Markus_Gerke/publication
- [10] Géron, A. Hands-On Machine Learning with Scikit-Learn and TensorFlow. O'Reilly Media, Inc, pp. 4-9, 2017.
- [11] Goodfellow, I., Bengio, Y. & Courville, A. Deep Learning. *MIT Press*, pp. 452-453, 2016. Retrieved from www.deeplearningbook.org

- [12] Hassan, M. u. (2018). *VGG16 – Convolutional Network for Classification and Detection*. Retrieved from <https://neurohive.io/wp-content/uploads/2018/11/vgg16-neural-network.jpg>
- [13] Hawkins, D. M. The Problem of Overfitting. *Journal of Chemical Information and Computer Sciences*, vol. 44(1), pp. 1–12, 2004. doi: [10.1021/ci0342472](https://doi.org/10.1021/ci0342472)
- [14] Ioffe, S., & Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv preprint [arXiv:1502.03167](https://arxiv.org/abs/1502.03167) [cs.LG]*, 2015.
- [15] Jørgensen, S. E. (Ed.). Artificial Neural Networks: Multilayer Perceptron for Ecological Modeling. *Ecological Model Types*, vol. 28, pp. 123-140, 2016. doi: [10.1016/B978-0-444-63623-2.00007-4](https://doi.org/10.1016/B978-0-444-63623-2.00007-4).
- [16] Krizhevsky, A., Sutskever, I., & Hinton, G. E. ImageNet Classification with Deep Convolutional Neural Networks. *NIPS'12: Proceedings of the 25th International Conference on Neural Information Processing Systems*, vol. 1, pp. 1097–1105, 2012.
- [17] Landis, J., & Koch, G. The measurement of observer agreement for categorical data. *Biometrics*, vol. 33(1), pp. 159-74, 1977.
- [18] Li, J., Cheng, J., Shi, J., & Huang, F. Brief Introduction of Back Propagation (BP) Neural Network Algorithm and Its Improvement. *Advances in Computer Science and Information Engineering*, pp. 553–558, 2012. doi: [10.1007/978-3-642-30223-7_87](https://doi.org/10.1007/978-3-642-30223-7_87)
- [19] Marmanis, D., Wegner, J. D., Galliani, S., Schindler, K., Datcu, M., & Stilla, U. Semantic segmentation of aerial images with an ensemble of CNNs. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 3, p. 473, 2016.
- [20] Martinez, M., & Stiefelhagen, R. Taming the Cross Entropy Loss. *arXiv preprint [arXiv:1810.05075v1](https://arxiv.org/abs/1810.05075v1) [cs.LG]*, 2018.
- [21] Nwankpa, C., Ijomah, W., Gachagan, A., & Marshall, S. Activation Functions: Comparison of trends in Practice and Research for Deep Learning. *arXiv preprint [arXiv:1811.03378v1](https://arxiv.org/abs/1811.03378v1) [cs.LG]*, 2018.
- [22] O’Shea, K., & Nash, R. An Introduction to Convolutional Neural Networks. *arXiv preprint [arXiv:1511.08458v2](https://arxiv.org/abs/1511.08458v2) [cs.NE]*, 2015
- [23] Parekh, M. (2019). A Brief Guide to Convolutional Neural Network(CNN). Retrieved from <https://medium.com/nybles/a-brief-guide-to-convolutional-neural-network-cnn-642f47e88ed4>

- [24] Ronneberger, O., Fischer, P., & Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv preprint* [arXiv:1505.04597v1](https://arxiv.org/abs/1505.04597v1) [cs.CV], 2015.
- [25] Ruck, D. W., Rogers, S. K., & Kabrisky, M. Feature Selection Using a Multilayer Perceptron. *Journal of Neural Network Computing*, vol. 2, No 2, pp. 40-48, 1990.
- [26] Shelhamer, E., Long, L., & Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *arXiv preprint* [arXiv:1605.06211v1](https://arxiv.org/abs/1605.06211v1) [cs.CV], 2016.
- [27] Shorten, C., & Khoshgoftaar, T.M. A survey on Image Data Augmentation for Deep Learning. *J Big Data* **6**, 60, 2019. doi: [10.1186/s40537-019-0197-0](https://doi.org/10.1186/s40537-019-0197-0)
- [28] Simonyan, K., & Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv preprint* [arXiv:1409.1556v6](https://arxiv.org/abs/1409.1556v6) [cs.CV], 2014.
- [29] Spiegel, S., Mueller, F., Wiesmann, D., & Bird, J. Cost-Sensitive Learning for Predictive Maintenance. *arXiv preprint* [arXiv:1809.10979v1](https://arxiv.org/abs/1809.10979v1) [cs.LG], 2018.
- [30] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.
- [31] Staudemeyer, R. C., & Rothstein Morris, E. Understanding LSTM -- a tutorial into Long Short-Term Memory Recurrent Neural Networks. *arXiv preprint* [arXiv:1909.09586v1](https://arxiv.org/abs/1909.09586v1) [cs.NE], 2019.
- [32] Sun, R. Optimization for deep learning: theory and algorithms. *arXiv preprint* [arXiv:1912.08957v1](https://arxiv.org/abs/1912.08957v1) [cs.LG], 2019.
- [33] Swalin, A. (2018). *Choosing the Right Metric for Evaluating Machine Learning Models — Part 2*. Retrieved from <https://medium.com/usf-msds/choosing-the-right-metric-for-evaluating-machine-learning-models-part-2-86d5649a5428>
- [34] Van Gerven, M., & Bohte, S. Artificial neural networks as models of neural information processing. *Frontiers Media SA*, p. 12, 2018.
- [35] Wu, J. Introduction to Convolutional Neural Networks. *National Key Lab for Novel Software Technology Nanjing University*, 2017.
- [36] Yamashita, R., Nishio, M., Do, R.K.G. et al. Convolutional neural networks: an overview and application in radiology. *Insights Imaging* **9**, pp. 611–629, 2018. doi: [10.1007/s13244-018-0639-9](https://doi.org/10.1007/s13244-018-0639-9)

Κατάλογος Σχημάτων

ΣΧΗΜΑ 2. 1: Η ΔΟΜΗ ΤΟΥ ΑΝΤΙΛΗΠΤΡΟΥ. [31]	7
ΣΧΗΜΑ 2. 2: ΠΟΛΥΕΠΙΠΕΔΟ ΑΝΤΙΛΗΠΤΡΟ. ΟΙ ΕΚΘΕΤΕΣ ΧΡΗΣΙΜΟΠΟΙΟΥΝΤΑΙ ΓΙΑ ΝΑ ΔΗΛΩΣΟΥΝ ΤΑ ΕΠΙΠΕΔΑ. ΑΡΑ ΤΟ x_{ij} ΕΙΝΑΙ ΤΟ ΔΕΔΟΜΕΝΟΥ ΕΞΟΔΟΥ ΤΟΥ ΝΕΥΡΩΝΑ i ΣΤΟ ΕΠΙΠΕΔΟ j ΚΑΙ ΤΟ w_{ijk} ΕΙΝΑΙ ΤΟ ΒΑΡΟΣ ΠΟΥ ΣΥΝΔΕΕΙ ΤΟΝ ΝΕΥΡΩΝΑ i ΣΤΟ ΕΠΙΠΕΔΟ $k - 1$ ΜΕ ΤΟΝ ΝΕΥΡΩΝΑ j ΣΤΟ ΕΠΙΠΕΔΟ k . ΤΟ ΕΠΙΠΕΔΟ 1 ΑΠΟΤΕΛΕΙ ΤΟ ΠΡΩΤΟ ΚΡΥΦΟ ΕΠΙΠΕΔΟ ΚΑΙ ΤΑ ΔΕΔΟΜΕΝΑ ΕΙΣΟΔΟΥ ΜΠΟΡΟΥΝ ΝΑ ΘΕΩΡΗΘΟΥΝ ΩΣ ΤΟ ΕΠΙΠΕΔΟ 0. [25]	8
ΣΧΗΜΑ 2. 3: ΜΙΑ ΟΠΤΙΚΗ ΑΝΑΠΑΡΑΣΤΑΣΗ ΕΝΟΣ ΣΥΝΕΛΙΚΤΙΚΟΥ ΕΠΙΠΕΔΟΥ. ΤΟ ΚΕΝΤΡΙΚΟ ΣΤΟΙΧΕΙΟ ΤΟΥ ΜΑΘΗΣΙΑΚΟΥ ΦΙΛΤΡΟΥ ΤΟΠΟΘΕΤΕΙΤΑΙ ΠΑΝΩ ΑΠΟ ΤΟ ΔΙΑΝΥΣΜΑ ΕΙΣΟΔΟΥ, ΤΟ ΟΠΟΙΟ ΣΤΗ ΣΥΝΕΧΕΙΑ ΥΠΟΛΟΓΙΖΕΤΑΙ ΚΑΙ ΑΝΤΙΚΑΘΙΣΤΑΤΑΙ ΜΕ ΕΝΑ ΣΤΑΘΜΙΣΜΕΝΟ ΑΘΡΟΙΣΜΑ ΤΟΥ ΕΑΥΤΟΥ ΤΟΥ ΚΑΙ ΤΥΧΟΝ ΓΕΙΤΟΝΙΚΩΝ ΕΙΚΟΝΟΣΤΟΙΧΕΙΩΝ. [22].....	14
ΣΧΗΜΑ 2. 4: ΠΑΡΑΔΕΙΓΜΑ ΣΥΝΕΛΙΞΗΣ ΜΕ ΜΕΓΕΘΟΣ ΜΑΘΗΣΙΑΚΟΥ ΦΙΛΤΡΟΥ 3×3 , ΧΩΡΙΣ ΕΠΕΝΔΥΣΗ ΚΑΙ ΒΗΜΑ 1. [36].....	15
ΣΧΗΜΑ 2. 5: ΕΙΚΟΝΑ ΜΕ ΜΗΔΕΝΙΚΗ ΕΠΕΝΔΥΣΗ. [6]	15
ΣΧΗΜΑ 2. 6: ΔΙΑΔΙΚΑΣΙΑ ΜΕΓΙΣΤΗΣ ΣΥΓΚΕΝΤΡΩΣΗΣ. [2].....	17
ΣΧΗΜΑ 2. 7: ΠΛΗΡΩΣ ΣΥΝΔΕΔΕΜΕΝΑ ΕΠΙΠΕΔΑ ΣΕ ΕΝΑ ΣΥΝΕΛΙΚΤΙΚΟ ΝΕΥΡΩΝΙΚΟ ΔΙΚΤΥΟ. [23]	18
ΣΧΗΜΑ 2. 8: ΜΕΤΑΤΡΟΠΗ ΚΑΝΟΝΙΚΟΠΟΙΗΣΗΣ ΠΑΡΤΙΔΑΣ, ΕΦΑΡΜΟΣΜΕΝΗ ΣΕ ΜΙΑ ΕΝΕΡΓΟΠΟΙΗΣΗ x ΣΕ ΜΙΑ ΜΙΝΙ-ΠΑΡΤΙΔΑ.	25
ΣΧΗΜΑ 2. 9: ΜΟΝΤΕΛΟ ΑΠΟΣΥΡΣΗΣ. ΑΡΙΣΤΕΡΑ (Α): ΈΝΑ ΚΑΝΟΝΙΚΟ ΝΕΥΡΩΝΙΚΟ ΔΙΚΤΥΟ. ΔΕΞΙΑ (Β): ΈΝΑ ΠΑΡΑΔΕΙΓΜΑ ΑΠΟΤΕΛΕΣΜΑΤΟΣ ΤΗΣ ΧΡΗΣΗΣ ΤΗΣ ΑΠΟΣΥΡΣΗΣ ΣΤΟ ΑΡΙΣΤΕΡΟ ΜΟΝΤΕΛΟ. ΟΙ ΔΙΑΓΡΑΜΜΕΝΟΙ ΝΕΥΡΩΝΕΣ ΕΧΟΥΝ ΥΠΟΣΤΕΙ ΑΠΟΣΥΡΣΗ. [30]	27
ΣΧΗΜΑ 2. 10: ΠΙΝΑΚΑΣ ΣΥΓΧΥΣΗΣ ΔΥΑΔΙΚΗΣ ΤΑΞΙΝΟΜΗΣΗΣ. [33].....	30
ΣΧΗΜΑ 2. 11: ΠΙΝΑΚΑΣ ΣΥΓΧΥΣΗΣ ΜΕ ΟΠΤΙΚΟΠΟΙΗΣΗ ΕΞΑΓΩΓΗΣ ΒΑΣΙΚΩΝ ΜΕΤΡΩΝ ΑΠΟΔΟΣΗΣ. [29].....	31
ΣΧΗΜΑ 3. 1: ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΔΙΚΤΥΟΥ U-NET. ΠΑΡΑΔΕΙΓΜΑ ΓΙΑ ΕΙΚΟΝΟΣΤΟΙΧΕΙΑ 32×32 ΣΤΗ ΧΑΜΗΛΟΤΕΡΗ ΑΝΑΛΥΣΗ. ΚΑΘΕ ΜΠΛΕ ΚΟΥΤΙ ΑΝΤΙΣΤΟΙΧΕΙ ΣΕ ΕΝΑΝ ΠΟΛΥΚΑΝΑΛΟ ΧΑΡΤΗ. Ο ΑΡΙΘΜΟΣ ΤΩΝ ΚΑΝΑΛΙΩΝ ΑΝΑΓΡΑΦΕΤΑΙ ΠΑΝΩ ΑΠΟ ΤΑ ΚΟΥΤΙΑ. ΤΑ ΛΕΥΚΑ ΚΟΥΤΙΑ ΑΠΟΤΕΛΟΥΝ ΤΑ ΑΝΤΙΓΡΑΜΜΕΝΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ. ΤΑ ΒΛΕΦΗ ΠΡΟΣΔΙΟΡΙΖΟΥΝ ΤΙΣ ΔΙΑΦΟΡΕΣ ΛΕΙΤΟΥΡΓΙΕΣ. [24]	38
ΣΧΗΜΑ 3. 2: ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΤΟΥ ΜΟΝΤΕΛΟΥ VGG-16. [12].....	39
ΣΧΗΜΑ 3. 3: ΑΡΧΙΤΕΚΤΟΝΙΚΗ FCN-PASCAL. ΤΑ ΕΠΙΠΕΔΑ ΣΥΓΚΕΝΤΡΩΣΗΣ ΚΑΙ ΠΡΟΒΛΕΨΗΣ ΕΜΦΑΝΙΖΟΝΤΑΙ ΩΣ ΠΛΕΓΜΑΤΑ, ΕΝΩ ΤΑ ΕΝΔΙΑΜΕΣΑ ΕΠΙΠΕΔΑ ΜΕ ΟΡΙΖΟΝΤΙΕΣ ΓΡΑΜΜΕΣ. ΠΡΩΤΗ ΓΡΑΜΜΗ (FCN-32s): ΜΟΝΗΣ ΡΟΗΣ ΔΙΚΤΥΟ, ΥΠΕΡΔΕΙΓΜΑΤΟΛΗΨΙΑ ΜΕ ΒΗΜΑ 32, ΠΡΟΒΛΕΨΕΙΣ ΠΙΣΩ ΣΤΑ ΕΙΚΟΝΟΣΤΟΙΧΕΙΑ ΣΕ ΕΝΑ ΒΗΜΑ. ΔΕΥΤΕΡΗ ΓΡΑΜΜΗ (FCN-16s): ΣΥΝΔΥΑΣΜΟΣ ΠΡΟΒΛΕΨΕΩΝ ΑΠΟ ΤΟ ΤΕΛΙΚΟ ΕΠΙΠΕΔΟ ΚΑΙ ΤΟ ΕΠΙΠΕΔΟ «POOL4», ΜΕ ΒΗΜΑ 16, ΕΠΙΤΡΕΠΕΙ ΤΗΝ ΠΡΟΒΛΕΨΗ ΛΕΠΤΟΜΕΡΕΙΑΣ ΥΨΗΛΟΤΕΡΗΣ ΑΝΑΛΥΣΗΣ, ΕΝΩ ΔΙΑΤΗΡΕΙ	

ΣΗΜΑΣΙΟΛΟΓΙΚΗ ΠΛΗΡΟΦΟΡΙΑ ΥΨΗΛΟΥ ΕΠΙΠΕΔΟΥ. ΤΡΙΤΗ ΓΡΑΜΜΗ (FCN-8s): ΠΡΟΣΘΕΤΕΣ ΠΡΟΒΛΕΨΕΙΣ ΑΠΟ ΤΟ ΕΠΙΠΕΔΟ «FOOL3», ΜΕ ΒΗΜΑ 8, ΠΑΡΕΧΟΥΝ ΜΕΓΑΛΥΤΕΡΗ ΑΚΡΙΒΕΙΑ. [26].....	40
ΣΧΗΜΑ 3. 4: ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΤΟΥ ΔΙΚΤΥΟΥ ΜΑΡΜΑΝΙΣ ET AL. ΣΤΑ ΑΡΙΣΤΕΡΑ ΤΑ ΕΠΙΠΕΔΑ ΚΙ ΟΙ ΣΥΝΔΕΣΕΙΣ ΚΑΙ ΣΤΑ ΔΕΞΙΑ Ο ΑΡΙΘΜΟΣ ΤΩΝ ΦΙΛΤΡΩΝ ΑΝΑ ΕΠΙΠΕΔΟ. Όλα τα φίλτρα συνέλιξης είναι 3×3 , όλα τα επίπεδα μεγίστης συγκέντρωσης είναι 2×2 . [19]	42
ΣΧΗΜΑ 4. 1: ΠΑΡΑΔΕΙΓΜΑ ΑΠΕΙΚΟΝΙΣΗΣ ΤΟΥ ΣΥΝΟΛΟΥ ΔΕΔΟΜΕΝΩΝ ΤΟΥ ΒΑΙΧΙΝΓΚΕΝ ΓΙΑ ΤΗΝ ΠΕΡΙΟΧΗ 11. ΑΡΙΣΤΕΡΑ: ΟΡΘΟΦΩΤΟΓΡΑΦΙΑ. ΜΕΣΗ: ΨΗΦΙΑΚΟ ΜΟΝΤΕΛΟ ΕΠΙΦΑΝΕΙΑΣ (DSM). ΔΕΞΙΑ: ΑΛΗΘΗΣ ΒΑΣΗ (GROUND TRUTH)	44
ΣΧΗΜΑ 4. 2: ΠΑΡΑΔΕΙΓΜΑ ΑΠΕΙΚΟΝΙΣΗΣ ΚΑΝΟΝΙΚΟΠΟΙΗΜΕΝΟΥ ΨΗΦΙΑΚΟΥ ΜΟΝΤΕΛΟΥ ΕΠΙΦΑΝΕΙΑΣ (NDSM) ΓΙΑ ΤΗΝ ΠΕΡΙΟΧΗ 11. [9]	45
ΣΧΗΜΑ 4. 3: ΟΠΤΙΚΟΠΟΙΗΣΗ ΤΟΥ ΠΡΟΤΕΙΝΟΜΕΝΟΥ ΜΟΝΤΕΛΟΥ ΜΑΣ.	49
ΣΧΗΜΑ 4. 4: ΑΡΙΣΤΕΡΑ: ΠΡΟΒΛΕΨΗ. ΔΕΞΙΑ: ΑΛΗΘΗΣ ΒΑΣΗ. ΑΠΟ ΠΑΝΩ ΠΡΟΣ ΤΑ ΚΑΤΩ: ΠΕΡΙΟΧΗ 5, ΠΕΡΙΟΧΗ 7, ΠΕΡΙΟΧΗ 23, ΠΕΡΙΟΧΗ 30.	65
ΣΧΗΜΑ 4. 5: ΑΡΙΣΤΕΡΑ: ΠΡΟΒΛΕΨΗ. ΔΕΞΙΑ: ΑΛΗΘΗΣ ΒΑΣΗ. ΑΠΟ ΠΑΝΩ ΠΡΟΣ ΤΑ ΚΑΤΩ: ΠΕΡΙΟΧΗ 5, ΠΕΡΙΟΧΗ 7, ΠΕΡΙΟΧΗ 23, ΠΕΡΙΟΧΗ 30.	81
ΣΧΗΜΑ 4. 6: ΑΡΙΣΤΕΡΑ: ΠΡΟΒΛΕΨΗ. ΔΕΞΙΑ: ΑΛΗΘΗΣ ΒΑΣΗ. ΑΠΟ ΠΑΝΩ ΠΡΟΣ ΤΑ ΚΑΤΩ: ΠΕΡΙΟΧΗ 5, ΠΕΡΙΟΧΗ 7, ΠΕΡΙΟΧΗ 23, ΠΕΡΙΟΧΗ 30.	97
ΣΧΗΜΑ 4. 7: ΑΠΟ ΑΡΙΣΤΕΡΑ ΣΤΑ ΔΕΞΙΑ: ΑΡΧΙΚΗ ΕΙΚΟΝΑ, ΠΡΟΒΛΕΨΗ, ΑΛΗΘΗΣ ΒΑΣΗ	100

Κατάλογος Πινάκων

ΠΙΝΑΚΑΣ 4. 1: ΠΡΩΤΟΣ ΈΛΕΓΧΟΣ ΠΕΡΙΟΧΗΣ 5.....	51
ΠΙΝΑΚΑΣ 4. 2: ΔΕΥΤΕΡΟΣ ΈΛΕΓΧΟΣ ΠΕΡΙΟΧΗΣ 5.....	53
ΠΙΝΑΚΑΣ 4. 3: ΠΡΩΤΟΣ ΈΛΕΓΧΟΣ ΠΕΡΙΟΧΗΣ 7.....	55
ΠΙΝΑΚΑΣ 4. 4: ΔΕΥΤΕΡΟΣ ΈΛΕΓΧΟΣ ΠΕΡΙΟΧΗΣ 7.....	56
ΠΙΝΑΚΑΣ 4. 5: ΠΡΩΤΟΣ ΈΛΕΓΧΟΣ ΠΕΡΙΟΧΗΣ 23.....	58
ΠΙΝΑΚΑΣ 4. 6: ΔΕΥΤΕΡΟΣ ΈΛΕΓΧΟΣ ΠΕΡΙΟΧΗΣ 23.....	60
ΠΙΝΑΚΑΣ 4. 7: ΠΡΩΤΟΣ ΈΛΕΓΧΟΣ ΠΕΡΙΟΧΗΣ 30.....	62
ΠΙΝΑΚΑΣ 4. 8: ΔΕΥΤΕΡΟΣ ΈΛΕΓΧΟΣ ΠΕΡΙΟΧΗΣ 30.....	63
ΠΙΝΑΚΑΣ 4. 9: ΠΡΩΤΟΣ ΈΛΕΓΧΟΣ ΠΕΡΙΟΧΗΣ 5.....	67
ΠΙΝΑΚΑΣ 4. 10: ΔΕΥΤΕΡΟΣ ΈΛΕΓΧΟΣ ΠΕΡΙΟΧΗΣ 5.....	69
ΠΙΝΑΚΑΣ 4. 11: ΠΡΩΤΟΣ ΈΛΕΓΧΟΣ ΠΕΡΙΟΧΗΣ 7.....	71
ΠΙΝΑΚΑΣ 4. 12: ΔΕΥΤΕΡΟΣ ΈΛΕΓΧΟΣ ΠΕΡΙΟΧΗΣ 7.....	72
ΠΙΝΑΚΑΣ 4. 13: ΠΡΩΤΟΣ ΈΛΕΓΧΟΣ ΠΕΡΙΟΧΗΣ 23.....	74
ΠΙΝΑΚΑΣ 4. 14: ΔΕΥΤΕΡΟΣ ΈΛΕΓΧΟΣ ΠΕΡΙΟΧΗΣ 23.....	76
ΠΙΝΑΚΑΣ 4. 15: ΠΡΩΤΟΣ ΈΛΕΓΧΟΣ ΠΕΡΙΟΧΗΣ 30.....	78
ΠΙΝΑΚΑΣ 4. 16: ΔΕΥΤΕΡΟΣ ΈΛΕΓΧΟΣ ΠΕΡΙΟΧΗΣ 30.....	79
ΠΙΝΑΚΑΣ 4. 17: ΠΡΩΤΟΣ ΈΛΕΓΧΟΣ ΠΕΡΙΟΧΗΣ 5.....	83
ΠΙΝΑΚΑΣ 4. 18: ΔΕΥΤΕΡΟΣ ΈΛΕΓΧΟΣ ΠΕΡΙΟΧΗΣ 5.....	85
ΠΙΝΑΚΑΣ 4. 19: ΠΡΩΤΟΣ ΈΛΕΓΧΟΣ ΠΕΡΙΟΧΗΣ 7.....	87
ΠΙΝΑΚΑΣ 4. 20: ΔΕΥΤΕΡΟΣ ΈΛΕΓΧΟΣ ΠΕΡΙΟΧΗΣ 7.....	88
ΠΙΝΑΚΑΣ 4. 21: ΠΡΩΤΟΣ ΈΛΕΓΧΟΣ ΠΕΡΙΟΧΗΣ 23.....	90
ΠΙΝΑΚΑΣ 4. 22: ΔΕΥΤΕΡΟΣ ΈΛΕΓΧΟΣ ΠΕΡΙΟΧΗΣ 23.....	92
ΠΙΝΑΚΑΣ 4. 23: ΠΡΩΤΟΣ ΈΛΕΓΧΟΣ ΠΕΡΙΟΧΗΣ 30.....	94
ΠΙΝΑΚΑΣ 4. 24: ΔΕΥΤΕΡΟΣ ΈΛΕΓΧΟΣ ΠΕΡΙΟΧΗΣ 30.....	95

Αγγλο-Ελληνικό Λεξικό

Αγγλικός Όρος

Activation Function

Array

Artificial Intelligence

Back Propagation

Background

Batch

Batch Size

Bias

Class

Classification

Classification Data

Classifier

Clutter

Computer Vision

Concatenation

Contracting Path

Convolution

Cross Entropy

DAG

Dataset

Deconvolution

Ελληνικός Όρος

Συνάρτηση Ενεργοποίησης

Συστοιχία

Τεχνητή Νοημοσύνη

Εμπρόσθια Διάδοση

Φόντο

Παρτίδα

Μέγεθος Παρτίδας

Πόλωση

Κλάση

Ταξινόμηση

Σύνολο Ταξινόμησης

Ταξινομητής

Συσσωρευση

Όραση Υπολογιστή

Συνένωση

Μονοπάτι Συστολής

Συνέλιξη

Σταυροειδής Εντροπία

Κατευθυνόμενος Άκυκλος Γράφος

Σύνολο Δεδομένων

Ανάστροφη Συνέλιξη

Deep Learning	Βαθιά Μάθηση
Depth	Βάθος
Downsample	Υποδειγματοληψία
End-to-End	Διατερματικός
Epoch	Εποχή
Expansive Path	Μονοπάτι Διαστολής
Forward Propagation	Εμπρόσθια Διάδοση
Fully Connected Network	Πλήρως Συνδεδεμένο Δίκτυο
Ground Truth	Αληθής Βάση
Hardware	Υλικό Υπολογιστή
Hidden Layer	Κρυφό Επίπεδο
Hyperparameter	Υπερπαραμέτρος
Hyperplane	Υπερεπίπεδο
Input	Δεδομένο Εισόδου
Iteration	Επανάληψη
Kernel	Μαθησιακό Φίλτρο
Label	Ετικέτα
Label Image	Κατηγοριοποιημένη Εικόνα
Lateral Inhibition	Πλάγια Αναστολή
Layer	Επίπεδο
Learning Rate	Ρυθμός Μάθησης
Local Response Normalization	Κανονικοποίηση Τοπικής Απόκρισης
Loss Function	Συνάρτηση Απώλειας
Machine Learning	Μηχανική Μάθηση
Max Pooling	Μέγιστη Συγκέντρωση

Max Unpooling	Μέγιστη Αποσυγκέντρωση
Metric	Μετρική
Module	Δομοστοιχείο
Momentum	Ορμή
Multi-channel	Πολυκάναλος
Multilayer Perceptron	Πολυεπίπεδο Αντίληπτρο
Neural Network	Νευρωνικά Δίκτυα
Neuron	Νευρώνας
Normalization	Κανονικοποίηση
Open-Source	Ανοιχτός Κώδικας
Optimizer	Βελτιστοποιητής
Output	Δεδομένο Εξόδου
Package	Πακέτο
Padding	Επένδυση
Perceptron	Αντίληπτρο
Pixel	Εικονοστοιχείο
Prediction	Πρόβλεψη
Random Access Memory	Μνήμη Τυχαίας Προσπέλασης
Raw data	Ακατέργαστα Δεδομένα
Receptive Field	Δεκτικό Πεδίο
Rectified Linear Unit (ReLU)	Διορθωμένη Γραμμική Μονάδα
Semantic Segmentation	Σημασιολογική Κατάτμηση
Skip Connection	Σύνδεση Υπερπήδησης
Softmax	Κανονικοποιημένη Εκθετική Συνάρτηση
State	Κατάσταση

Stochastic Gradient Descent	Στοχαστική Κλίση Καθόδου
Stride	Βήμα
Supervised learning	Επιβλεπόμενη μάθηση
Tensor	Τανυστής
Test Set	Σύνολο Ελέγχου
Threshold	Κατώφλι
Training Set	Σύνολο Εκπαίδευσης
Unit	Μονάδα
Upsample	Υπερδειγματοληψία
Validation Set	Σύνολο Επαλήθευσης
Weight	Βάρος
Window Size	Μέγεθος Παραθύρου