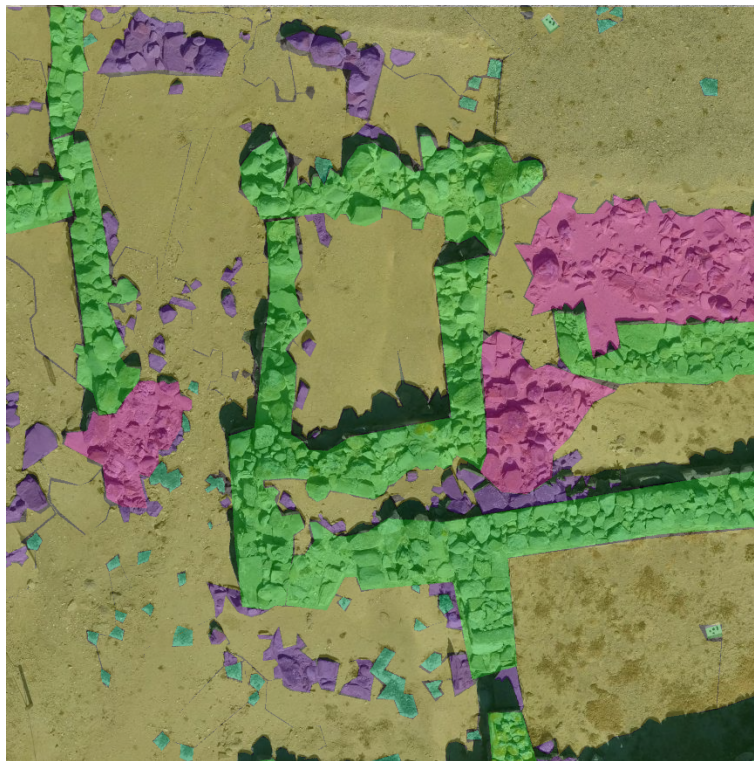




ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΑΓΡΟΝΟΜΩΝ ΚΑΙ ΤΟΠΟΓΡΑΦΩΝ ΜΗΧΑΝΙΚΩΝ
ΤΟΜΕΑΣ ΤΟΠΟΓΡΑΦΙΑΣ
ΕΡΓΑΣΤΗΡΙΟ ΦΩΤΟΓΡΑΜΜΕΤΡΙΑΣ

ΣΗΜΑΣΙΟΛΟΓΙΚΗ ΚΑΤΑΤΜΗΣΗ RGB ΚΑΙ RGB-D ΕΙΚΟΝΩΝ ΜΕ ΧΡΗΣΗ ΜΕΘΟΔΩΝ ΒΑΘΙΑΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ



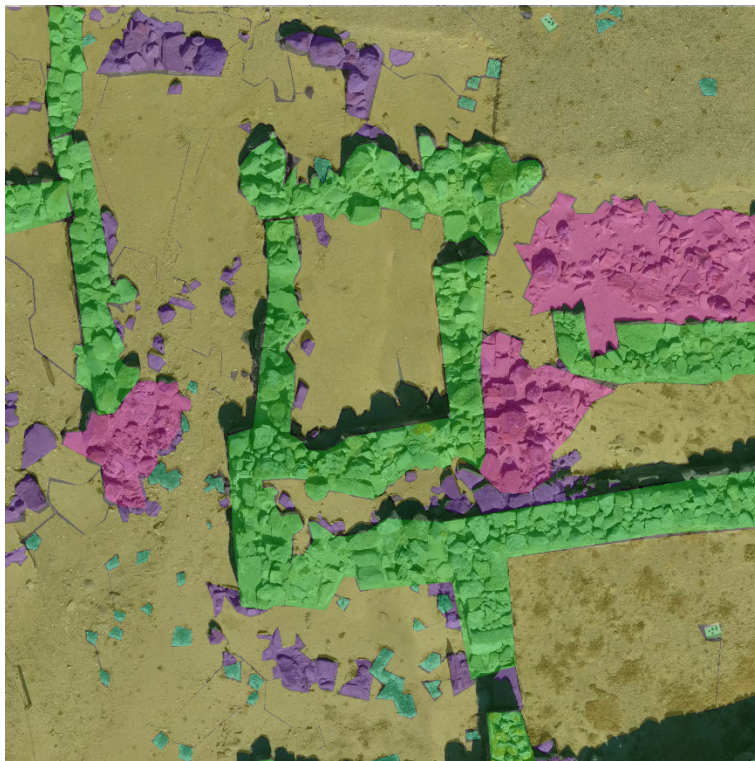
ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ: ΤΣΙΡΩΝΗΣ ΠΙΕΤΡΟΣ
ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ: ΑΝΔΡΕΑΣ ΓΕΩΡΓΟΠΟΥΛΟΣ

Αθήνα, Οκτώβριος 2021



**NATIONAL AND TECHNICAL UNIVERSITY
RURAL AND SURVEING ENGINEERING
SECTION OF TOPOGRAPHY
LAB OF PHOTOGRAMMETRY**

SEMANTIC SEGMENTATION OF RGB AND RGB-D IMAGES WITH THE USE OF METHODS OF DEEP MACHINE LEARNING



**DIPLOMA THESIS: TSIRONIS PETROS
THESIS SUPERVISOR: ANDREAS GEORGOPOULOS**

Athens, October 2021

Ευχαριστίες

Με την ολοκλήρωση της διπλωματικής εργασίας θα ήθελα να ευχαριστήσω τα άτομα τα οποία συνέβαλαν με τον δικό τους ξεχωριστό τρόπο κατά την εκπόνηση αυτής.

Αρχικά θα ήθελα να ευχαριστήσω τον υπεύθυνο καθηγητή κύριο Ανδρέα Γεωργόπουλο, Καθηγητή ΣΑΤΜ για την εμπιστοσύνη, την αφοσίωση και τις διδαχές που μου χάρισε κατά την διάρκεια των σπουδών μου και ειδικά κατά την διάρκεια εκπόνησης της διπλωματικής εργασίας στο τμήμα Αγρονόμων και Τοπογράφων Μηχανικών.

Τον Παναγιώτη Αγραφιώτη, Μεταδιδακτορικό Ερευνητή, ο οποίος ήταν και ο εμπνευστής του θέματος της διπλωματικής, καθώς και ο άνθρωπος που βρισκόταν δίπλα μου καθ' όλη την διάρκεια εκπόνησης της παρούσας διπλωματικής με τις συμβουλές και τις γνώσεις του.

Την κυρία Σεβαστή Ταπεινάκη, μέλος ΕΔΙΠ, για την διάθεση δεδομένων για την ορθότερη τεκμηρίωση της συγκεκριμένης διπλωματικής εργασίας.

Επίσης να ευχαριστήσω τους συμφοιτητές μου, οι οποίοι με στήριξαν, όντας δίπλα μου, και με τον διάλογο με βοήθησαν να γίνομαι καλύτερος σε κάθε βήμα των σπουδών μου.

Τέλος να ευχαριστήσω την μητέρα μου και τον αδερφό μου που χωρίς την ανιδιοτελή τους βοήθεια δεν θα μπορούσα να φτάσω στο σημείο που βρίσκομαι σήμερα σαν άνθρωπος, αλλά και στο επίπεδο γνώσεων, καθώς μου χάρισαν απλόχερα τα στοιχεία της εργατικότητας και της περιέργειας να μαθαίνω και να προπορεύομαι.

Περιεχόμενα

Ευχαριστίες.....	1
Περίληψη.....	6
Abstract.....	7
1. Εισαγωγή.....	8
1.1 Στόχος της Διπλωματικής.....	8
2. Βιβλιογραφική ανασκόπηση.....	9
3. Θεωρητικό Υπόβαθρο.....	11
3.1 Εισαγωγή στη Μηχανική Μάθηση.....	11
3.2 Νευρωνικά Δίκτυα.....	15
3.2.1 Αλγόριθμος Back-propagation.....	16
3.2.2 Συναρτήσεις Ενεργοποίησης.....	17
3.3 Συνελκτικά Νευρωνικά Δίκτυα.....	20
3.3.1 Επίπεδο Συνέλιξης (Convolutional Layer).....	21
3.3.2 Επίπεδο Υπό-Δειγματοληψίας (Pooling Layer).....	22
3.3.3 Πλήρως Συνδεδεμένο Επίπεδο (Fully Connected Layer).....	23
3.4 Αρχιτεκτονικές Βαθιάς Μηχανικής Μάθησης και η Επιστήμη της Μηχανικής Όρασης.....	23
4. Υλοποίηση.....	25
4.1 Δεδομένα και Λογισμικά προεπεξεργασίας.....	25
4.1.1 Ο Οικισμός της Πολιόχνης.....	25
4.1.2 Ορθοφωτογραφίες.....	26
4.1.3 Λογισμικά και κώδικες Προεπεξεργασίας.....	26
4.2 Αρχιτεκτονικές CNN.....	29
4.2.1 SegNet.....	29
4.2.2 U-Net.....	30
4.2.3 ResNet.....	31
4.3 Κώδικας εκπαίδευσης.....	32
4.3.1 Περαιτέρω Επεξήγηση βασικών στοιχείων.....	33
4.3.2 Εμβάθυνση στη Λειτουργία του Κώδικα.....	34
5. Εκπαίδευση και αποτελέσματα.....	36
5.1 Στάδιο πρώτο.....	36
5.1.1 SegNet RGB.....	37
5.1.2 SegNet RGB-D.....	39

5.1.3	U-Net RGB	41
5.1.4	U-Net RGB-D	44
5.1.5	ResNet50_fcn RGB	46
5.1.6	ResNet50_dlv3 RGB	48
5.1.7	ResNet101_fcn RGB	51
5.1.8	ResNet101_dlv3 RGB	53
5.2	Στάδιο Δεύτερο	55
5.2.1	SegNet RGB	56
5.2.2	SegNet RGB-D.....	58
5.2.3	U-Net RGB	59
5.2.4	U-Net RGB-D	61
5.2.5	ResNet50_fcn RGB	63
5.2.6	ResNet50_dlv3 RGB	65
5.2.7	ResNet101_fcn RGB	67
5.2.8	ResNet101_dlv3 RGB	69
5.3	Τελική Επιλογή.....	71
6.	Συμπεράσματα.....	72
6.1	Αξιολόγηση	72
6.1.1	Αξιολόγηση Σημασιολογική Κατάτμησης σε RGB εικόνες.....	72
6.1.2	Αξιολόγηση Σημασιολογικής Κατάτμησης σε RGB-D εικόνες	75
6.1.3	Σύγκριση Επιλεγμένων Μεθόδων Σημασιολογικής Κατάτμησης για RGB και RGB-D εικόνες	77
6.2	Προβλήματα.....	79
6.3	Μελλοντικές Χρήσεις	79
	Βιβλιογραφία	80

Κατάλογος Εικόνων

<i>Εικόνα 3.1: Βασικές Εφαρμογές της Τεχνητής Νοημοσύνης</i>	<i>12</i>
<i>Εικόνα 3.2: Δομή Νευρικού Κυττάρου.....</i>	<i>15</i>
<i>Εικόνα 3.3: Γραφική Παράσταση Βαθμιδωτής Συνάρτησης.....</i>	<i>17</i>
<i>Εικόνα 3.4: Γραφική παράσταση Σιγμοειδής Συνάρτησης</i>	<i>18</i>
<i>Εικόνα 3.5: Γραφική παράσταση Συνάρτησης Εφαπτομένης Υπερβολής</i>	<i>19</i>
<i>Εικόνα 3.6: Γραφική παράσταση Συνάρτησης Διορθωμένης Μονάδας.....</i>	<i>19</i>
<i>Εικόνα 3.7: Συνελκτικό Νευρωνικό Δίκτυο[18]</i>	<i>20</i>

Εικόνα 3.8: Διαδικασία Συνέλιξης[19]	21
Εικόνα 3.9: Μέθοδοι Υπό-Δειγματοληψίας[20]	22
Εικόνα 3.10: Πλήρως Συνδεδεμένο Επίπεδο	23
Εικόνα 4.1: Ορθοφωτογραφία και χάρτης βάθους (DEM) τμήματος του Αρχαιολογικού Χώρου	26
Εικόνα 4.2: Κομμάτι δεδομένων αλήθειας πεδίου (Ground Truth) στα αριστερά και πραγματικής εικόνας στα δεξιά.....	28
Εικόνα 4.3: Προσθήκη βάθους σε RGB εικόνα και δημιουργία RGB-D εικόνας	29
Εικόνα 4.4: Παράσταση Αρχιτεκτονικής SegNet με τροποποίηση ως προς τα δεδομένα εισαγωγής και εξαγωγής[14]	30
Εικόνα 4.5: Παράσταση Αρχιτεκτονικής U-Net με τροποποίηση ως προς τα δεδομένα εισαγωγής και εξαγωγής [12]	31
Εικόνα 4.6: Λογική των Δομών Συντόμευσης της ResNet[31]	32
Εικόνα 4.7: παράσταση αρχιτεκτονικής DeepLabV3 με Atrous Spatial Pyramid Pooling και Atrous Convolution με τροποποίηση ως προς τα δεδομένα εισαγωγής και εξαγωγής [32].....	32
Εικόνα 5.1: Γραφική παράσταση Μέσης απώλειας(Mean loss) και διασταυρούμενης ακρίβειας(Cross-validation accuracy) για το μοντέλο αρχιτεκτονικής SegNet σε εικόνες RGB	56
Εικόνα 5.2: Γραφική παράσταση Μέσης απώλειας(Mean loss) και διασταυρούμενης ακρίβειας(Cross-validation accuracy) για το μοντέλο αρχιτεκτονικής SegNet σε εικόνες RGB-D	58
Εικόνα 5.3: Γραφική παράσταση Μέσης απώλειας(Mean loss) και διασταυρούμενης ακρίβειας(Cross-validation accuracy) για το μοντέλο αρχιτεκτονικής U-Net σε εικόνες RGB	59
Εικόνα 5.4: Γραφική παράσταση Μέσης απώλειας(Mean loss) και διασταυρούμενης ακρίβειας(Cross-validation accuracy) για το μοντέλο αρχιτεκτονικής U-Net σε εικόνες RGB-D	61
Εικόνα 5.5: Γραφική παράσταση Μέσης απώλειας(Mean loss) και διασταυρούμενης ακρίβειας(Cross-validation accuracy) για το μοντέλο αρχιτεκτονικής ResNet50_fcn με βήμα 0.01.....	63
Εικόνα 5.6: Γραφική παράσταση Μέσης απώλειας(Mean loss) και διασταυρούμενης ακρίβειας(Cross-validation accuracy) για το μοντέλο αρχιτεκτονικής ResNet50_fcn με βήμα 0.001.....	64
Εικόνα 5.7: Γραφική παράσταση Μέσης απώλειας(Mean loss) και διασταυρούμενης ακρίβειας(Cross-validation accuracy) για το μοντέλο αρχιτεκτονικής ResNet50_dln3 με βήμα 0.001.....	65
Εικόνα 5.8: Γραφική παράσταση Μέσης απώλειας(Mean loss) και διασταυρούμενης ακρίβειας(Cross-validation accuracy) για το μοντέλο αρχιτεκτονικής ResNet101_fcn με βήμα 0.01.....	67
Εικόνα 5.9: Γραφική παράσταση Μέσης απώλειας(Mean loss) και διασταυρούμενης ακρίβειας(Cross-validation accuracy) για το μοντέλο αρχιτεκτονικής ResNet101_fcn με βήμα 0.001.....	68
Εικόνα 5.10: Γραφική παράσταση Μέσης απώλειας(Mean loss) και διασταυρούμενης ακρίβειας(Cross-validation accuracy) για το μοντέλο αρχιτεκτονικής ResNet101_dln3 με βήμα 0.001.....	69
Εικόνα 6.1: Εικόνα προβολής διαφορών όσον αφορά την κατηγορίας γκρεμισμάτων	72
Εικόνα 6.2: Εικόνα σύγκρισης πρόβλεψης (Prediction) και αλήθειας πεδίου (Ground Truth)	75

Κατάλογος Πινάκων

Πίνακας 1: Πίνακας στοιχείων κατάτμησης των εικόνων.....	28
Πίνακας 2: Αποτελέσματα εκπαίδευσης και δοκιμών SegNet σε εικόνες RGB με διάρκεια 5 εποχών	37
Πίνακας 3: Γραφικές παραστάσεις διασταυρούμενων ακριβειών και απωλειών για αρχιτεκτονική SegNet σε εικόνες RGB με διάρκεια 5 εποχών.....	38

Πίνακας 4:Αποτελέσματα εκπαίδευσης και δοκιμών SegNet σε εικόνες RGB-D με διάρκεια 5 εποχών...	40
Πίνακας 5: Γραφικές παραστάσεις διασταυρούμενων ακριβειών και απωλειών για αρχιτεκτονική SegNet σε εικόνες RGB-D με διάρκεια 5 εποχών.....	41
Πίνακας 6: Αποτελέσματα εκπαίδευσης και δοκιμών U-Net σε εικόνες RGB με διάρκεια 5 εποχών	42
Πίνακας 7:Γραφικές παραστάσεις διασταυρούμενων ακριβειών και απωλειών για αρχιτεκτονική U-Net σε εικόνες RGB με διάρκεια 5 εποχών.....	43
Πίνακας 8:Αποτελέσματα εκπαίδευσης και δοκιμών U-Net σε εικόνες RGB-D με διάρκεια 5 εποχών.....	44
Πίνακας 9: Γραφικές παραστάσεις διασταυρούμενων ακριβειών και απωλειών για αρχιτεκτονική U-Net σε εικόνες RGB-D με διάρκεια 5 εποχών.....	45
Πίνακας 10: Αποτελέσματα εκπαίδευσης και δοκιμών ResNet50_fcn σε εικόνες RGB με διάρκεια 5 εποχών	47
Πίνακας 11:Γραφικές παραστάσεις διασταυρούμενων ακριβειών και απωλειών για αρχιτεκτονική ResNet50_fcn σε εικόνες RGB με διάρκεια 5 εποχών	48
Πίνακας 12: Αποτελέσματα εκπαίδευσης και δοκιμών ResNet50_dlv3 σε εικόνες RGB με διάρκεια 5 εποχών	49
Πίνακας 13:Γραφικές παραστάσεις διασταυρούμενων ακριβειών και απωλειών για αρχιτεκτονική ResNet50_dlv3 σε εικόνες RGB με διάρκεια 5 εποχών	50
Πίνακας 14: Αποτελέσματα εκπαίδευσης και δοκιμών ResNet101_fcn σε εικόνες RGB με διάρκεια 5 εποχών	51
Πίνακας 15: Γραφικές παραστάσεις διασταυρούμενων ακριβειών και απωλειών για αρχιτεκτονική ResNet101_fcn σε εικόνες RGB με διάρκεια 5 εποχών	52
Πίνακας 16: Αποτελέσματα εκπαίδευσης και δοκιμών ResNet101_dlv3 σε εικόνες RGB με διάρκεια 5 εποχών	54
Πίνακας 17: Γραφικές παραστάσεις διασταυρούμενων ακριβειών και απωλειών για αρχιτεκτονική ResNet101_dlv3 σε εικόνες RGB με διάρκεια 5 εποχών	55
Πίνακας 18: Αποτελέσματα δοκιμών αρχιτεκτονικής SegNet με βήμα 0.001 για εικόνες RGB.....	57
Πίνακας 19: Αποτελέσματα δοκιμών αρχιτεκτονικής SegNet με βήμα 0.001 για εικόνες RGB-D	58
Πίνακας 20: Αποτελέσματα δοκιμών αρχιτεκτονικής U-Net με βήμα 0.001 για εικόνες RGB.....	60
Πίνακας 21: Αποτελέσματα δοκιμών αρχιτεκτονικής U-Net με βήμα 0.001 για εικόνες RGB-D	62
Πίνακας 22: Αποτελέσματα δοκιμών αρχιτεκτονικής ResNet50_fcn με βήμα 0.01.....	63
Πίνακας 23: Αποτελέσματα δοκιμών αρχιτεκτονικής ResNet50_fcn με βήμα 0.001.....	64
Πίνακας 24: Αποτελέσματα δοκιμών αρχιτεκτονικής ResNet50_dlv3 με βήμα 0.001.....	66
Πίνακας 25: Αποτελέσματα δοκιμών αρχιτεκτονικής ResNet101_fcn με βήμα 0.01.....	67
Πίνακας 26: Αποτελέσματα δοκιμών αρχιτεκτονικής ResNet101_fcn με βήμα 0.001.....	68
Πίνακας 27: Αποτελέσματα δοκιμών αρχιτεκτονικής ResNet101_dlv3 με βήμα 0.001.....	70
Πίνακας 28: Εικόνες δοκιμών χωρίς αληθινή βάση για αρχιτεκτονική ResNet101_fcn με βήμα 0.001	74
Πίνακας 29:Εικόνες δοκιμών χωρίς αληθινή βάση για αρχιτεκτονική SegNet με βήμα 0.001	76
Πίνακας 30: Πίνακας σύγκρισης εικόνων από μοντέλα ResNet101_fcn και SegNet.....	77

Περίληψη

Η Μηχανική Μάθηση είναι ένας από τους σημαντικότερους τομείς των υπολογιστικών συστημάτων τα τελευταία χρόνια λόγω της ραγδαίας εξέλιξής της. Αυτό συνέβη καθώς πολλοί επιστημονικοί τομείς, όπως και η φωτογραμμετρία, βασίζονται πλέον στις μεθόδους αυτής, και κυρίως στην χρήση νευρωνικών δικτύων συνέλιξης (Convolutional Neural Networks, CNN), για την ανάπτυξη μοντέλων που διευκολύνουν και επιταχύνουν σημαντικά την έρευνα. Ειδικότερα, όπως και εδώ, τα δίκτυα αυτά χρησιμοποιούνται κατά κύριο λόγο για επίλυση προβλημάτων αναγνώρισης προτύπων στην Μηχανική Όραση (CV). Τα CNNs ανήκουν στις αλγοριθμικές αρχιτεκτονικές εκμάθησης αναπαραστάσεων, συνεπώς δεν χρειάζονται εξαγωγή χαρακτηριστικών από ανθρώπινο παράγοντα, αλλά εξάγουν αυτόματα τα χαρακτηριστικά που χρειάζονται βασισμένα στα δεδομένα που τους έχουν ήδη δοθεί.

Η παρούσα διπλωματική εξετάζει πιο από τα CNN που χρησιμοποιούνται δίνει το καλύτερο δυνατό αποτέλεσμα για το πρόβλημα σημασιολογικής κατάτμησης ορθοφωτογραφιών μνημείων RGB και RGB-D, οι οποίες είναι εικόνες υπό συγκεκριμένη κλίμακα, οι οποίες μπορούν να αντιμετωπιστούν ως χάρτες και επιτρέπουν να πραγματοποιούνται μετρήσεις ακριβείας επάνω σε αυτές. Τα CNN που χρησιμοποιήθηκαν για την μελέτη είναι τα SegNet, U-Net, ResNet50 και 101 με βάσεις FCN και DeepLabV3. Για κάθε μια από τις παραπάνω αρχιτεκτονικές πραγματοποιήθηκαν στατιστικές μετρήσεις σε εικόνες, στις οποίες είχε πραγματοποιηθεί από πριν ταξινόμηση των διάφορων κατηγοριών που επιλέχθηκαν για την μελέτη. Βάσει των μετρήσεων αυτών επιλέχθηκε η αρχιτεκτονική με το καλύτερο δυνατό αποτέλεσμα για τον αυτοματισμό του κατάτμηση των ορθοφωτογραφιών που χρησιμοποιήθηκαν, καθώς και άλλων στο μέλλον.

Επιπλέον παρουσιάζονται οι ρυθμίσεις που πραγματοποιήθηκαν για την βέλτιστη επιλογή του κάθε ενός CNN και η σημαντικότητα αυτών στον κώδικα που χρησιμοποιήθηκε για την βέλτιστη επίλυση του προβλήματος.

Με βάση τα αποτελέσματα παρατηρείται η σημαντικότητα της επιλογής της αρχιτεκτονικής CNN και πόσο σημαντικός είναι ο ανθρώπινος παράγοντας και οι επιλογές ρυθμίσεων για την τελική εμφάνιση της σημασιολογικής κατάτμησης της κάθε εικόνας.

Abstract

Machine learning is one of the most important areas of computational systems of the recent years, because of its fast evolution. This happened because many scientific areas, like photogrammetry, are based on machine learning methods, and especially on the use of Convolutional Neural Networks (CNN), for the development of models that can reduce the amount of time and optimize surveys. Specifically, as in this study, these networks are used primarily for solving problems of pattern recognition in Computer Vision (CV). CNNs are categorized as algorithmic architectures of learning patterns, so they do not have the need of a human aspect for learning, but they extract automatically the characteristics based on the data that they were given.

The current diploma thesis examines, which of the CNN architectures that is used gives the best output for the problem of semantic segmentation of RGB and RGB-D orthoimages of monument sites. Orthoimages are images of specific scale, which can be seen as maps and accurate measurements can be made on them. The architectures that were used for this research are SegNet, U-Net and ResNet50 and 101 with FCN and DeepLabV3 backbone. For each one of these architectures statistical measurements on images were made, on which were already drawn the chromatic categories that are used on this research. Based on those measurements the best architectures for solving the problem of automating the categorization of the cultural orthoimages used, and many other in the feature were selected.

Also the best settings used are presented that were used for the final architecture selection and the importance of those for the optimization of solving the problem with the use of a certain algorithm.

Based on the results one can comprehend the importance of selecting the architecture, of the human choices and the selection of settings used for the semantic segmentation of every image.

1. Εισαγωγή

Η Τεχνητή Νοημοσύνη (Artificial Intelligence) είναι μια ευρέως διαδεδομένη πλέον έννοια. Σε αυτήν γίνεται αναφορά συνήθως όταν γίνεται λόγος για κάποια ρομποτική οντότητα ή κάποια έξυπνα συστήματα (συνήθως υπολογιστικά).

Η μελέτη για την δημιουργία τέτοιων συστημάτων οδήγησε στην έννοια τη Μηχανικής Μάθησης (Machine Learning). Η αναφορά αυτή έγινε για τον προσδιορισμό όλων εκείνων των τεχνικών που οδήγησαν σε αυτά τα έξυπνα συστήματα στον κόσμο του προγραμματισμού. Αυτή η έννοια της Μηχανικής Μάθησης έδωσε και την δυνατότητα για επίλυση σύνθετων προβλημάτων. Το 1952, όπου ο Arthur Samuel αναφέρθηκε πρώτη φορά σε αυτόν τον όρο[1], ήταν αδύνατο να διανοηθεί κανείς την δύναμη που έδωσε έκτοτε αυτή, στην σύγχρονη εξέλιξη της μηχανικής μάθησης.

Οι παραπάνω τεχνολογίες χρησιμοποιούνται κυρίως για ανάλυση δεδομένων, αυτόματη αναγνώριση φωνής, επεξεργασία φυσικής γλώσσας, επεξεργασία οπτικού στυλ, ιατροφαρμακευτικές εφαρμογές, αναγνώριση οικονομικής απάτης, στρατιωτικές εφαρμογές και, πιο σημαντικό για την συγκεκριμένη διπλωματική, αναγνώριση στοιχείων εικόνας.

1.1 Στόχος της Διπλωματικής

Στόχος αυτής της διπλωματικής είναι η διερεύνηση των δυνατοτήτων της εφαρμογής αυτών των σύγχρονων τεχνικών στην γεωμετρική τεκμηρίωση μνημείων και συγκεκριμένα η εκπαίδευση ενός συστήματος έτσι ώστε να μπορεί να αναγνωρίζει με την καλύτερη δυνατή ακρίβεια τα στοιχεία ενός μνημείου που έχουν επιλεγθεί να εμφανίζονται σε οποιαδήποτε ορθοφωτογραφία οποιουδήποτε μνημείου. Αυτό γίνεται με την χρήση μηχανικής μάθησης και συγκεκριμένα Αρχιτεκτονικών Βαθιάς Μηχανικής Μάθησης (Deep Learning Architectures), οι οποίες προσμοιάζουν την λειτουργία των νευρώνων του ανθρώπου για την αναγνώριση χαρακτηριστικών στα πρότυπα των εικόνων που έχουν χρησιμοποιηθεί, και στα δεδομένα που έχουν δοθεί στον αλγόριθμο. Το τελικό αυτό εκπαιδευμένο σύστημα θα μπορεί να πραγματοποιήσει την κατάτμηση (semantic segmentation) της ορθοφωτογραφίας σε πολύ μικρότερο χρονικό διάστημα από αυτό που θα χρειαζόταν ένας άνθρωπος.

2. Βιβλιογραφική ανασκόπηση

Η σημασιολογική κατάτμηση μπορεί να χαρακτηριστεί ως η διαδικασία απόδοσης μιας κατηγορίας σε κάθε τμήμα της εικόνας με παρόμοια χαρακτηριστικά. Αυτό το τμήμα μπορεί να είναι ακόμα και ένα εικονοστοιχείο ή ένα κομμάτι της εικόνας που απαρτίζεται από μερικά εικονοστοιχεία[2]. Συνηθίζεται βέβαια σε εικόνες όπως αυτές που χρησιμοποιήθηκαν για την συγκεκριμένη έρευνα να κατηγοριοποιούνται ανά κομμάτια εικόνας, καθώς λόγω της φύσης αυτών δεν μπορεί ένα αντικείμενο να περιγράφεται από ένα μόνο εικονοστοιχείο. Βέβαια σε κάθε εικονοστοιχείο στο τέλος της κατάτμησης αποδίδεται μια κατηγορία. Πολλοί αλγόριθμοι κατάτμησης έχουν χρησιμοποιηθεί στο παρελθόν για αντιμετώπιση παρόμοιων προβλημάτων, όπως τεχνητά νευρωνικά δίκτυα (Artificial Neural Networks - ANN), support vector machine (SVM), δέντρων απόφασης (decision tree) και άλλοι[3]. Πιο πρόσφατες έρευνες έχουν αναγνωρίσει την χρήση της Βαθιάς μάθησης για σημασιολογική κατάτμηση ως τον πιο ακριβή τρόπο για κατηγοριοποίηση των εικονοστοιχείων[4].

Η Βαθιά μηχανική μάθηση είναι ουσιαστικά μια μέθοδος, η οποία χρησιμοποιεί συνελκτικά νευρωνικά δίκτυα (CNN) για την εκπαίδευση μοντέλων και την κατάτμηση εικόνων. Αυτά με την σειρά τους έχουν διάφορες αρχιτεκτονικές. Για την βελτιστοποίηση των αρχιτεκτονικών αυτών, όσον αφορά τα αποτελέσματα και τον χρόνο περάτωσής τους, χρησιμοποιούνται συναρτήσεις όπως διόρθωσης γραμμικής μονάδας (ReLU), παραμετρικής διόρθωσης γραμμικής μονάδας (Parametric rectified linear Unit – PreLU)[5], dropout[6] και άλλες, και συναρτήσεις βελτιστοποίησης δικτύου όπως stochastic gradient descent (SGD)[7], και Adam Optimizer[8]. Ένα σύνηθες CNN αποτελείται από ένα επίπεδο εισόδου, διάφορα συνελκτικά επίπεδα, διάφορα επίπεδα δειγματοληψίας, ένα πλήρως συνδεδεμένο επίπεδο και ένα τελευταίο επίπεδο εξαγωγής με μια συνάρτηση soft-max.

Τα CNN παρόλη την αποτελεσματικότητά τους στην κατάτμηση αντιμετωπίζουν κάποια προβλήματα. Για την επίλυση των προβλημάτων αυτών δημιουργήθηκαν το 2015 τα πλήρως συνελκτικά δίκτυα (Fully Convolutional Network – FCN)[9]. Αυτά τα δίκτυα απαρτίζονται από δύο μέρη τους κωδικοποιητές (encoders) και τους αποκωδικοποιητές (decoders). Το μέρος της κωδικοποίησης έχει τον ρόλο της επεξεργασίας των εικόνων για την εξαγωγή χαρτών χαρακτηριστικών (feature maps) και το μέρος της αποκωδικοποίησης με την σειρά του μεταμορφώνει αυτούς τους χάρτες σε συμπαγείς χάρτες κατηγοριών με ίδιο μέγεθος με τις αρχικές εικόνες. Μια από τις συνηθέστερες τέτοιου είδους αρχιτεκτονικές που χρησιμοποιείται και στην συγκεκριμένη μελέτη είναι η ResNet[10] σε διάφορες μορφές της και με χρήση κορμού της αρχιτεκτονικής DeepLabV3[11].

Πέρα από τα απλά FCN η ανάγκη για παραπάνω πληροφορία και η βαθύτερη κατανόηση των δικτύων αυτών έδωσε την ιδέα για δημιουργία δικτύων τα οποία θα μπορούν να χρησιμοποιούν τα γειτονικά αντικείμενα για καλύτερη αναγνώριση των στοιχείων με αποτέλεσμα την καλύτερη δυνατή κατάτμηση των εικόνων. Αυτό έφερε στην επιφάνεια τις αρχιτεκτονικές με την χρήση συνδέσεων παράβλεψης (skip connections), τα οποία έχουν την δυνατότητα της χρήσης χαμηλότερου και υψηλότερου επιπέδου για ακριβέστερες προβλέψεις. Αυτή η σκέψη σήμανε και

την γέννηση της αρχιτεκτονικής U-Net [12], στην οποία κάθε επίπεδο δημιουργεί ανεξάρτητες προβλέψεις και με χρήση διαφόρων τεχνικών βελτιστοποίησης η αρχιτεκτονική αυτή μαθαίνει στο μοντέλο μοτίβα. Η αρχιτεκτονική αυτή χρησιμοποιήθηκε για εξέταση ιατρικών εικόνων και σαρώσεων και σε μερικές μελέτες για κατάτμηση εικόνων πόλεων[13]. Μια ακόμη αρχιτεκτονική που βασίστηκε στην ανάπτυξη των συμμετρικών αυτών αρχιτεκτονικών και πολύ διαδεδομένη είναι η αρχιτεκτονική SegNet[14], η οποία βελτίωσε την περιγραφή των κατηγοριών με ελάχιστες παραμέτρους επαναχρησιμοποιώντας δειγματοληπτικούς δείκτες.

3. Θεωρητικό Υπόβαθρο

Στην σύγχρονη εποχή ο κλάδος της τεχνητής νοημοσύνης είναι ένας ραγδαία αναπτυσσόμενος κλάδος της επιστήμης της πληροφορικής με τεράστια ερευνητική και πρακτική σημασία. Οι δύο κατηγορίες αυτής είναι:

- Η συμβολική τεχνητή νοημοσύνη, η οποία χρησιμοποιεί αλγοριθμικές τεχνικές (σύμβολα και λογικούς κανόνες) για την επίλυση προβλημάτων τεχνητής νοημοσύνης και
- Η υποσυμβολική τεχνητή νοημοσύνη, η οποία προσπαθεί να αντιγράψει την ανθρώπινη λογική με χρήση αριθμητικών μοντέλων που ακολουθούν την βιολογική λειτουργία του ανθρώπινου εγκεφάλου.

Η ικανότητα ενός νοούμενου (AI) συστήματος να αποκτά από μόνο του γνώση, εξάγοντας πρότυπα ή/και χαρακτηριστικά σημεία από τα δεδομένα, είναι γνωστή ως Μηχανική Μάθηση (Machine Learning - ML).

Στόχος του κεφαλαίου αυτού είναι η επεξήγηση τεχνικών και αλγορίθμων Μηχανικής Μάθησης και κυρίως των Βαθέων Νευρωνικών Δικτύων (Deep Neural Networks – DNN). Τα στοιχεία των δικτύων αυτών αποτελούν βάση στην μελέτη των νευρωνικών δικτύων συνέλιξης (CNN) και την χρήση αυτών στην επίλυση του προβλήματος όρασης υπολογιστών για την κατάτμηση εικόνων (Semantic Segmentation of Images).

3.1 Εισαγωγή στη Μηχανική Μάθηση

Χάρη στην εξέλιξη της μηχανικής μάθησης και της εισαγωγής αυτής στους υπολογιστές πολλές επιστήμες γνώρισαν ραγδαία ανάπτυξη κατά την τελευταία δεκαετία. Η εισαγωγή ειδικά αυτής στην πληροφορική έδωσε την δυνατότητα επίλυσης λογικών προβλημάτων σε υπολογιστικά συστήματα χωρίς την ανθρώπινη επέμβαση, καθώς και την δυνατότητα σε αυτά να αποφασίζουν για την πορεία επίλυσης.

Συνεπώς οι αλγόριθμοι ML επιτρέπουν σε αυτά τα ‘έξυπνα’ συστήματα να προσαρμόζονται και να μαθαίνουν, χωρίς σημαντική ανθρώπινη επέμβαση, νέες πληροφορίες για τον πραγματικό κόσμο. Για παράδειγμα ένας αλγόριθμος που «ξέρει» ήδη να αναγνωρίζει χειρόγραφα νούμερα μπορεί επίσης να εκπαιδευτεί να αναγνωρίζει και χειρόγραφα γράμματα χωρίς να χρειάζεται η ανθρώπινη παρέμβαση[15].



Εικόνα 3.1: Βασικές Εφαρμογές της Τεχνητής Νοημοσύνης¹

Η Μηχανική μάθηση χωρίζεται σε τέσσερις κατηγορίες αλγορίθμων επίλυσης προβλημάτων:

- **Supervised Learning algorithms** – Επιβλεπόμενοι Αλγόριθμοι Μάθησης: Αυτοί οι αλγόριθμοι είναι ικανοί να εφαρμόζουν ό,τι έχουν μάθει ήδη σε νέα δεδομένα με την βοήθεια κατηγοριοποιημένων παραδειγμάτων. Αυτή η διαδικασία πραγματοποιείται για μελλοντικές προβλέψεις. Τέτοιοι αλγόριθμοι ξεκινούν την ανάλυση των ήδη γνωστών δεδομένων εκπαίδευσης (training datasets) και ύστερα παράγουν χαρακτηριστικά για να πραγματοποιήσουν προβλέψεις για τα δεδομένα επεξεργασίας. Τα συστήματα αυτά μπορούν να παρέχουν στόχους για νέες εισαγωγές μόνο αφού έχουν εκπαιδευτεί αρκετά. Τέλος οι αλγόριθμοι αυτοί μπορούν να συγκρίνουν τα δεδομένα που εξήγαγαν με τα επιθυμητά αποτελέσματα, έτσι ώστε να βρουν σφάλματα και να τροποποιήσουν το τελικό μοντέλο με βάση αυτά.
- **Unsupervised Learning algorithms** – Μη επιβλεπόμενοι Αλγόριθμοι Μάθησης: Αυτοί χρησιμοποιούνται όταν τα δεδομένα εκπαίδευσης δεν έχουν κάποια κατηγοριοποίηση. Η μάθηση χωρίς επίβλεψη είναι βασισμένη σε συστήματα που αναλύουν πρότυπα σε μη κατηγοριοποιημένα δεδομένα. Τα συστήματα αυτά δεν φτάνουν σε κάποιο αποτέλεσμα που είναι λάθος ή σωστό, αλλά εξερευνούν τα δεδομένα και εξάγουν χαρακτηριστικά από τα δεδομένα που τους δόθηκαν.
- **Semi-supervised Learning Algorithms** – Αλγόριθμοι Ημι-επιβλεπόμενης Μάθησης: Είναι αλγόριθμοι οι οποίοι ακολουθούν και τις δυο παραπάνω διαδικασίες μαζί. Συνήθως με

¹ Πηγή: <https://electricalfundablog.com/artificial-intelligence-ai-introduction-types-application-areas/>

μικρό όγκο κατηγοριοποιημένων δεδομένων και μεγάλο μη κατηγοριοποιημένων. Συστήματα σαν και αυτά μπορούν να αυξήσουν την ακρίβεια μάθησης σημαντικά. Τις περισσότερες φορές γίνεται χρήση αυτών, όταν τα κατηγοριοποιημένα δεδομένα συνδυάζονται με πηγές για να βοηθήσουν την μάθηση του συστήματος.

- Reinforced Learning Algorithms – Ενισχυμένοι Αλγόριθμοι Μάθησης (δια της ανταμοιβής): Αυτού του τύπου οι αλγόριθμοι αλληλοεπιδρούν με το περιβάλλον με πράξεις που αργότερα ανακαλύπτουν σφάλματα ή επιβραβεύσεις. Αυτή η καθυστερημένη ανταμοιβή ή η δοκιμή και το σφάλμα (trial and error) είναι τα κλειδιά για την μάθηση αυτή. Η μέθοδος αυτή επιτρέπει σε ‘έξυπνα’ συστήματα να κρίνουν αυτόματα την καλύτερη πορεία σε ένα συγκεκριμένο πλαίσιο για να αυξήσουν την απόδοσή τους και να έχουν τα καλύτερα δυνατά αποτελέσματα. Για την διαδικασία αυτή απαραίτητη είναι η ύπαρξη του σήματος αποδοχής όταν οι πράξεις τους είναι οι ορθότερες.

Πέρα από τις παραπάνω κατηγορίες, οι αλγόριθμοι ML χωρίζονται και σε άλλες κατηγορίες με βάση την λειτουργία που εκτελούν. Οι πιο σημαντικές κατηγορίες είναι:

Regression

Αφορά στην στατιστική μέθοδο μοντελοποίησης της σχέσης μεταξύ εξαρτώμενων και ανεξάρτητων μεταβλητών με μια ή περισσότερες ανεξάρτητες μεταβλητές. Πιο συγκεκριμένα βοηθάει την κατανόηση του πώς μια εξαρτημένη μεταβλητή αλλάζει σύμφωνα με μια ανεξάρτητη μεταβλητή όταν όλες οι υπόλοιπες ανεξάρτητες μεταβλητές παραμένουν σταθερές.

- Linear Regression
- Polynomial Regression
- Logistic Regression
- Lasso Regression
- Ridge Regression

Instance-based

Αυτοί οι αλγόριθμοι χρησιμοποιούν ένα μέγεθος ομοιότητας μεταξύ των μεταβλητών, το οποίο χρησιμοποιείται για την κατηγοριοποίηση αυτών. Ύστερα χρησιμοποιώντας την κατηγοριοποίηση και το μέγεθος ομοιότητας ο αλγόριθμος κρίνει με βάση τις επόμενες εισαγωγές ποια χαρακτηριστικά να κρατήσει και ποια είναι πλέον ασήμαντα[16].

- k-Nearest Neighbor (kNN)
- Learning Vector Quantization (LVQ)
- Self-Organizing Map (SOM)
- Locally Weighted Learning (LWL)

Decision Trees

Διαγράμματα ροής τα οποία χρησιμοποιούνται για την κατασκευή μοντέλων αποφάσεων, με κάθε κλάδο τους να αναπαριστά ένα πρόβλημα χρησιμοποιώντας τις πραγματικές τιμές των χαρακτηριστικών των δεδομένων.

- Conditional Decision Trees
- M5

Bayesian

Αλγόριθμοι που εφαρμόζουν το θεώρημα του Bayes για την επίλυση προβλημάτων ταξινόμησης και συνέχειας.

- Naïve Bayes
- Gaussian Naïve Bayes

Clustering

Αφορά στις αυτοματοποιημένες διαδικασίες εύρεσης φυσικών στοιχείων ταξινόμησης των δεδομένων.

- k-Means
- k-Medians
- Hierarchical Clustering

Artificial Neural Networks (ANN)

Μοντέλα αναπαράστασης και αντιγραφής της λειτουργίας και της δομής των βιολογικών νευρωνικών δικτύων.

- Perceptron
- Back-Propagation

Deep Learning (DL)

Οι αλγόριθμοι DL πραγματοποιήθηκαν χάρη στην εξέλιξη των ANN με την ραγδαία ανάπτυξη των υπολογιστικών συστημάτων και της επεξεργαστικής τους ισχύος.

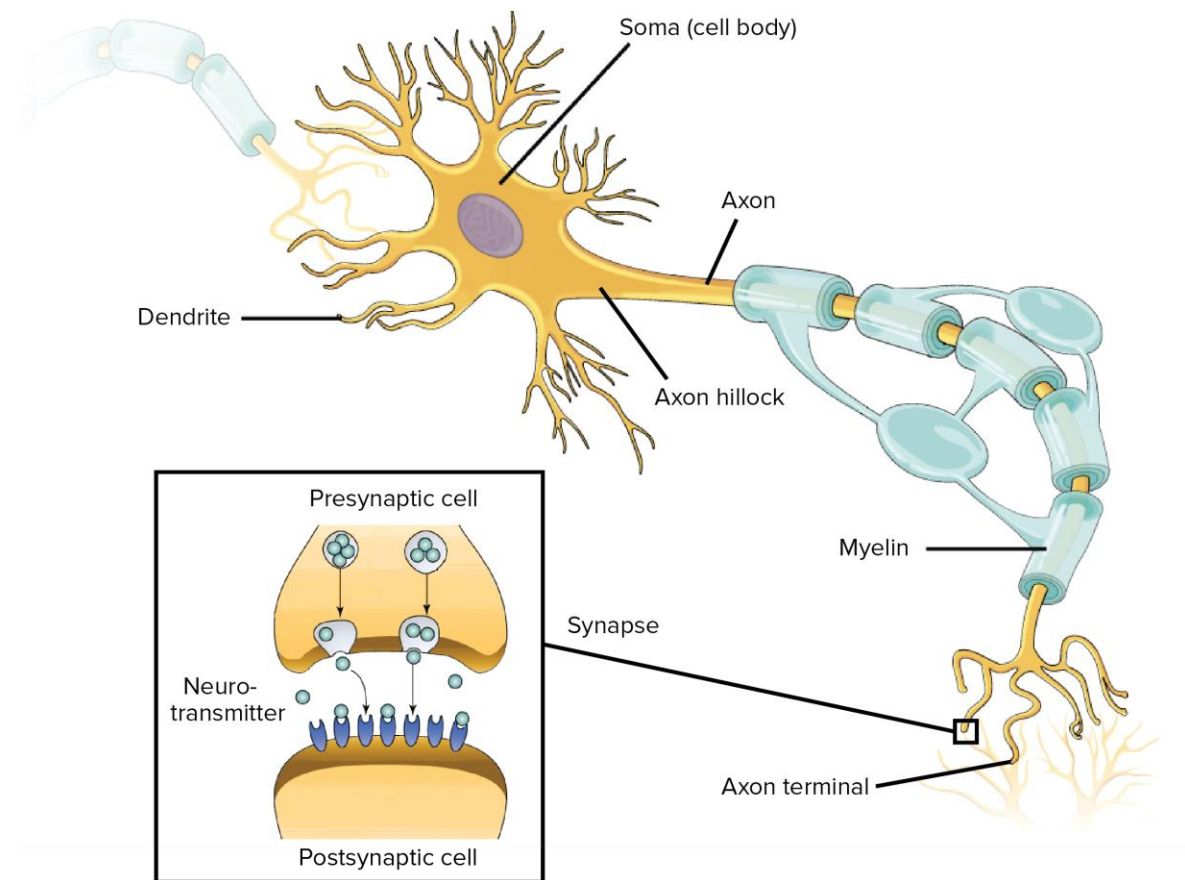
- Autoencoder
- Multilayer Perceptron (MLP)
- Convolutional Neural Networks (CNN)
- Recurrent Neural Network (RNN)

Οι αλγόριθμοι DL είναι και αυτοί που θα απασχολήσουν κατά κύριο λόγο στην συνέχεια της μελέτης με κύριους αλγορίθμους νευρωνικών δικτύων συνέλιξης (CNNs).

3.2 Νευρωνικά Δίκτυα

Τα Νευρωνικά Δίκτυα είναι προγραμματιστικά μοντέλα αναπαράστασης του βιολογικού νευρικού συστήματος του ανθρώπου. Το κάθε νευρικό κύτταρο αποτελεί μονάδα επεξεργασίας του ανθρώπινου εγκεφάλου. Ένας τέτοιος νευρώνας αποτελείται από:

- Δενδρίτες (Dendrites): Σημείο πολλαπλών εισόδων προς επεξεργασία.
- Κυτταρικό σώμα (cell body): Κύριο σώμα όπου γίνεται η επεξεργασία των εισόδων και αποφασίζεται η πράξη που θα πραγματοποιηθεί.
- Νευρικός Άξονας (Axon): Εκεί μεταφέρεται η πληροφορία της πράξης που θα πραγματοποιηθεί αν έχει υπάρξει ενεργοποίηση του νευρικού συστήματος και με την βοήθεια της μυελίνης θα γίνει η μεταφορά αυτής από τον άξονα στις
- Νευροαξονικές απολήξεις (Synapses): Από τις οποίες η πληροφορία θα περάσει στα επόμενα νευρικά κύτταρα μέσω των νευρομεταδοτών.



Εικόνα 3.2: Δομή Νευρικού Κυττάρου²

²Πηγή: <https://www.khanacademy.org/science/biology/human-biology/neuron-nervous-system/a/overview-of-neuron-structure-and-function>

Η βιολογική αυτή διαδικασία οδήγησε στο μαθηματικό μοντέλο το οποίο χρησιμοποιήθηκε για την εξέλιξη σύνθετων τεχνητών νευρωνικών δικτύων. Το μοντέλο αυτό ουσιαστικά παρουσιάζει τις εισαγωγές στους δενδρίτες ως δεδομένα (x_i) τα οποία πολλαπλασιάζονται με τον αριθμό που ίδια δεδομένα (με κάποια απόκλιση πάντα) παρουσιάστηκαν ως εισαγωγές σε άλλους δενδρίτες. Οι αριθμοί αυτοί παρουσιάζονται ως βάρη καθώς αποτελούν τις μεταβλητές παραμέτρους ενός νευρώνα. Η τιμές αυτές επιδρούν πολλαπλασιαστικά με τα δεδομένα στην λήψη των αποφάσεων ($w_i * x_i$). Σε συνδυασμό με κάποιο συστηματικό σφάλμα (bias) δίνεται η εξίσωση που δίνει την τιμή επίδρασης μεταξύ των νευρώνων. Τέλος σε συνδυασμό με κάποια συνάρτηση ενεργοποίησης f (π.χ. Σιγμοειδής), η οποία ελέγχει την ροή της πληροφορίας και δίνει την ικανότητα χρήσης της εξίσωσης αυτής σε μη γραμμικές σχέσεις, δημιουργείται η τελική σχέση υπολογισμού της εξόδου από ένα νευρώνα:

$$a = f\left(\sum_{i=0}^N (w_i x_i + b)\right)$$

Με την παραπάνω λογική λειτουργούν και τα τεχνητά νευρωνικά δίκτυα.

3.2.1 Αλγόριθμος Back-propagation

Ο σημαντικότερος αλγόριθμος στον προγραμματισμό νευρωνικών δικτύων ονομάζεται back-propagation. Πρωτοεμφανίστηκε το 1960 και έγινε διάσημο το 1989 από το paper των Rumelhart et.al. (1989) με ονομασία “Learning representations be back-propagating errors”[17].

Ο αλγόριθμος αυτός χρησιμοποιείται για την εμπειριστατωμένη εκπαίδευση ενός νευρωνικού δικτύου μέσω της μεθόδου του νόμου της αλυσίδας (chain rule). Απλοποιημένα, ο αλγόριθμος αυτός κάνει ένα ορθό πέρασμα από το δίκτυο, και ύστερα ένα ανάποδο προσαρμόζοντας τα βάρη και τις αποκλίσεις (παράμετροι) του μοντέλου.

Στο ορθό πέρασμα ο αλγόριθμος πρέπει να αξιολογήσει την έξοδο που προβλέφθηκε σε σχέση με την πραγματική έξοδο. Η πραγματική έξοδος βρίσκεται στα δεδομένα εκπαίδευσης του αλγορίθμου. Η αξιολόγηση αυτή γίνεται με την αξιοποίηση μίας συνάρτησης κόστους (MSE, cross-entropy, κ.λπ.). Βασισμένος στην τιμή κόστους ο αλγόριθμος ‘ξέρει’ πόσο πρέπει να προσαρμόσει τις παραμέτρους για να έρθει όσο πιο κοντά γίνεται στην πραγματική έξοδο.

Το μέγεθος αυτών των προσαρμογών των παραμέτρων καθορίζεται από τις κλίσεις (gradients) της συνάρτησης κόστους σε σχέση με τις παραμέτρους. Για τον υπολογισμό αυτών γίνεται χρήση του νόμου της αλυσίδας.

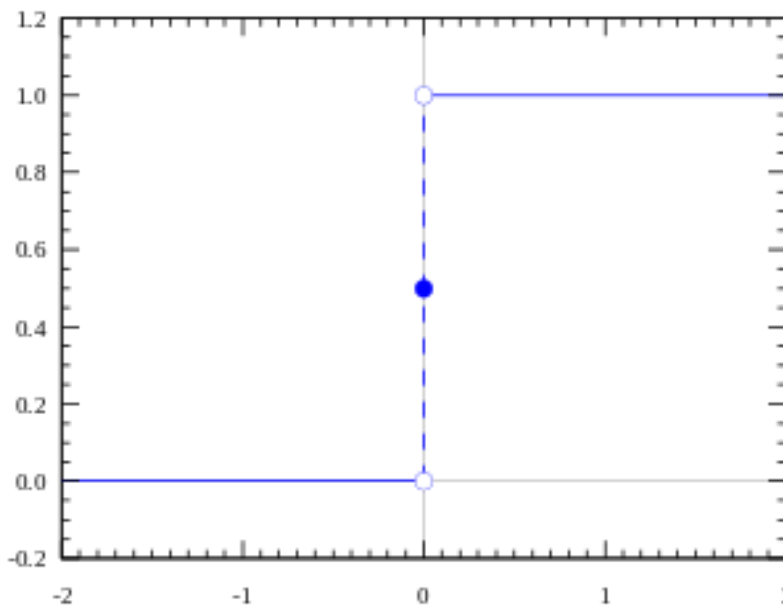
$$\frac{dy}{dx} = \frac{dy}{du} \frac{du}{dx}$$

3.2.2 Συναρτήσεις Ενεργοποίησης

Οι συναρτήσεις ενεργοποίησης προσδιορίζουν το αποτέλεσμα μίας σειράς εισόδων που έχουν δεχτεί. Στα νευρωνικά δίκτυα τα αποτελέσματα αυτά αναπαριστούν κυρίως κάποια απόφαση. Οι πιο συχνές συναρτήσεις ενεργοποίησης είναι:

3.2.2.1 Βαθμιδωτή Συνάρτηση (Step Function)

Η βαθμιδωτή συνάρτηση λειτουργεί με την λογική της επιλογής θέτοντας ένα κατώφλι (threshold). Αν η τιμή που εισέρθει σε αυτή την συνάρτηση είναι πάνω από αυτό το κατώφλι τότε εμφανίζεται η τιμή 1 ως έξοδος της συνάρτησης, αλλιώς η συνάρτηση εξάγει την τιμή 0.



Εικόνα 3.3: Γραφική Παράσταση Βαθμιδωτής Συνάρτησης

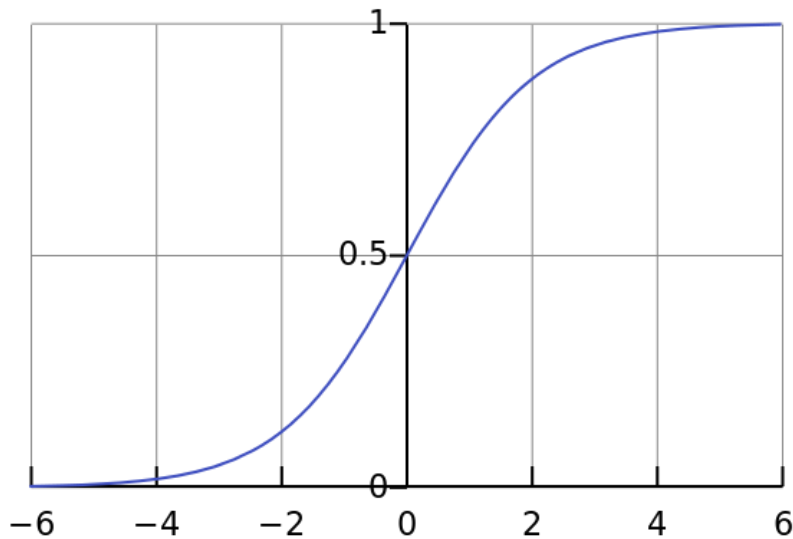
Βέβαια η συνάρτηση αυτή έχει το μειονέκτημα ότι είτε θα ενεργοποιήσει τον τεχνητό νευρώνα είτε θα τον απενεργοποιήσει χωρίς να δίνει περιθώρια σφάλματος και δημιουργώντας έτσι προβλήματα σε περιπτώσεις που η εκπαίδευση απαιτεί περισσότερες από μια κατηγορίες.

3.2.2.2 Γραμμική Συνάρτηση (Linear Function)

Η γραμμική συνάρτηση είναι η πιο απλή από τις μη δυαδικές εξισώσεις ενεργοποίησης. Είναι της μορφής $f = ax$. Το μειονέκτημά της είναι ότι είναι γραμμική εξίσωση. Συνεπώς αν συνδεθούν πολλοί γραμμικοί νευρώνες μαζί δεν θα αλλάξει ποτέ το αποτέλεσμα καθώς τα βάρη θα παραμείνουν σταθερά. Αυτό συμβαίνει διότι για τον επαναπροσδιορισμό των βαρών χρησιμοποιείται ο αλγόριθμος back-propagation, ο οποίος χρησιμοποιεί τον κανόνα της αλυσίδας (chain rule). Συνεπώς τα βάρη παραμένουν σταθερά με αποτέλεσμα, όσοι νευρώνες, και αν συνδεθούν, εφόσον αυτοί παραμένουν γραμμικοί δεν θα δημιουργούνται νέα βάρη άρα θα είναι σαν να χρησιμοποιείται μόνο ένας νευρώνας.

3.2.2.3 Σιγμοειδής Συνάρτηση (Sigmoid Function)

Η σιγμοειδής συνάρτηση είναι μια μη γραμμική και μη δυαδική συνάρτηση. Αποτέλεσμα αυτής της ιδιότητας είναι να μπορεί να έχει παραπάνω από 2 αποτελέσματα ενεργοποίησης (πολλαπλές κατηγορίες επιλογής) και επίσης να μπορεί να χρησιμοποιηθεί σε συνδυασμό με άλλους νευρώνες τέτοιου τύπου καθώς δεν θα έχει σταθερά βάρη λόγω της μη γραμμικότητάς της. Είναι της μορφής $f(x) = \frac{1}{1+e^{-x}}$ και παρατηρώντας την γραφική της παράσταση γίνεται αντιληπτό ότι ακόμη ένα από τα πλεονεκτήματά της είναι ότι οι τιμές εξόδου περιορίζονται μεταξύ 0 και 1 με αποτέλεσμα να μην υπάρχουν τιμές εξόδου οι οποίες να μην μπορούν να υπολογιστούν ή να δημιουργούν προβλήματα λόγω του μεγέθους τους.

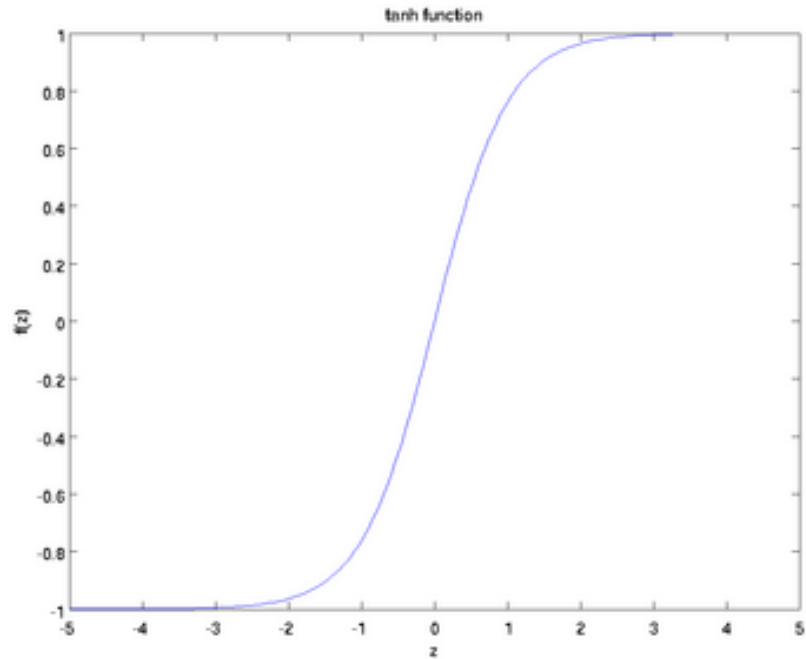


Εικόνα 3.4: Γραφική παράσταση Σιγμοειδής Συνάρτησης

Το μόνο πρόβλημα που παρατηρείται με αυτή την συνάρτηση είναι ότι στα άκρα της η συνάρτηση θα είναι δύσκολο να επιλέξει κάποια τιμή ή θα έχει τεράστια καθυστέρηση ανάλογα με την επεξεργαστική δυνατότητα του συστήματος όταν φτάσει στο όριο των δεκαδικών που μπορεί να επεξεργαστεί. Για το πρόβλημα αυτό υπάρχουν βεβαίως λύσεις, στις οποίες δεν θα γίνει αναφορά. Η συνάρτηση αυτή είναι μια από τις πιο διάσημες εξισώσεις ενεργοποίησης, ειδικά σε προβλήματα ταξινόμησης.

3.2.2.4 Συνάρτηση Εφαπτομένης Υπερβολής (Hyperbolic tangent function)

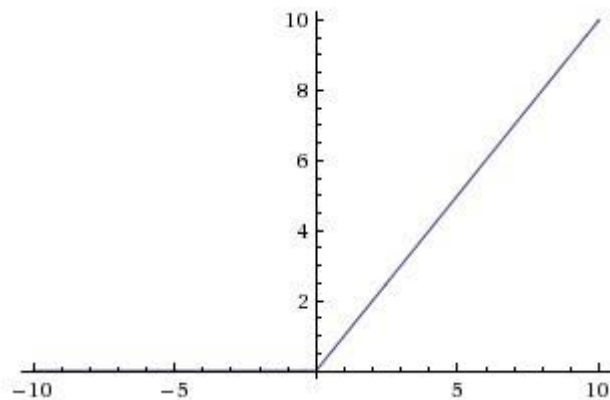
Αλλιώς ονομαζόμενη και tanh function είναι ουσιαστικά μια αναπροσαρμοσμένη σιγμοειδής συνάρτηση με ίδια χαρακτηριστικά αλλά πιο εξομαλυσμένη ως προς τον άξονα των X για αποφυγή του προβλήματος της σιγμοειδούς.



Εικόνα 3.5: Γραφική παράσταση Συνάρτησης Εφαπτομένης Υπερβολής

3.2.2.5 Συνάρτηση Διορθωμένης Γραμμικής Μονάδας (Rectified Linear Unit Function)

Γνωστή και ως ReLu είναι η συνάρτηση της μορφής $f(x) = \max(0, x)$ και δίνει την τιμή του x αν το x είναι θετικό αλλιώς 0.



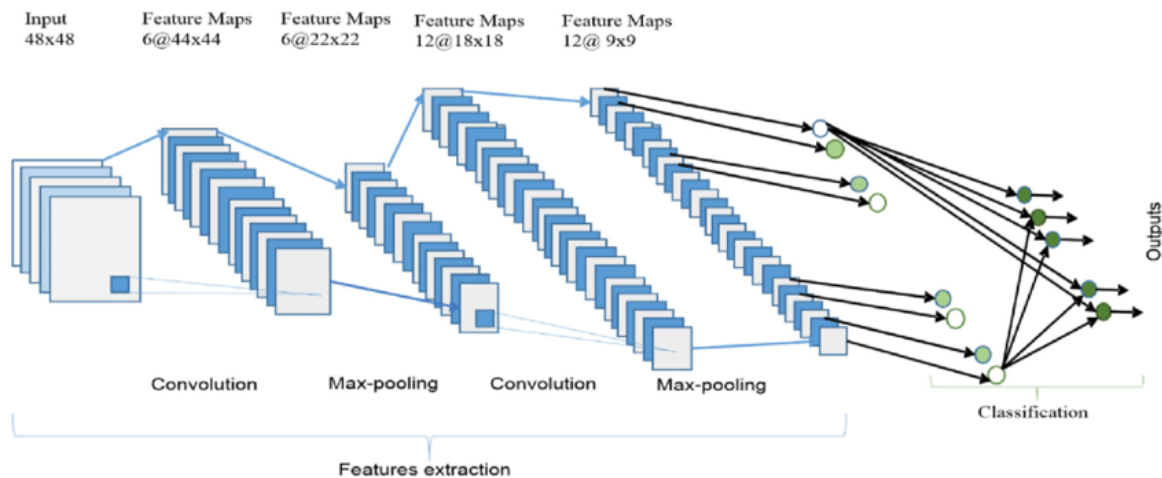
Εικόνα 3.6: Γραφική παράσταση Συνάρτησης Διορθωμένης Μονάδας

Είναι μια μη γραμμική συνάρτηση, παρόλη την εμφάνιση της γραφικής της παράστασης, και έχει το πλεονέκτημα να έχει μικρότερη υπολογιστική ανάγκη σε σχέση με σιγμοειδή και tanh, λόγω πιο απλών πράξεων, και επίσης λόγω της φύσης της ενεργοποιεί λιγότερους νευρώνες καθώς ορίζει από την αρχή ότι ό,τι είναι αρνητικό ή 0 παίρνει την τιμή 0 και συνεπώς εκεί δεν υπάρχει ενεργοποίηση του νευρώνα. Βέβαια έχει και αυτή τα μειονεκτήματά της. Λόγω της εμβέλειάς της $[0, \infty]$ η συνάρτηση μπορεί να σταματήσει να λειτουργεί.

3.3 Συνελκτικά Νευρωνικά Δίκτυα

Σε αυτή την παράγραφο θα γίνει ανάλυση των Συνελκτικών Νευρωνικών δικτύων. Τα συγκεκριμένα μοντέλα θα απασχολήσουν στην συνέχεια της μελέτης, καθώς είναι αυτά που χρησιμοποιούνται σήμερα για την αναγνώριση αντικειμένων σε εικόνες και την κατάτμηση αυτών.

Τα Συνελκτικά Νευρωνικά δίκτυα (CNN) αποτελούνται από πολλά επίπεδα. Κάθε επίπεδο αποτελείται από έναν αριθμό νευρώνων, οι οποίοι δέχονται μια είσοδο, ύστερα πολλαπλασιάζουν αυτή την είσοδο με κάποια βάρη και το αποτέλεσμα της διαδικασίας περνά από μια (συνήθως μη γραμμική) συνάρτηση ενεργοποίησης. Ανάλογα με τις ενεργοποιήσεις και τα βάρη που έχουν δημιουργηθεί στο πρώτο επίπεδο προχωρά η διαδικασία στο επόμενο επίπεδο με εισόδους τις εξόδους του προηγούμενου επιπέδου. Τέλος στο τελευταίο επίπεδο το οποίο είναι πλήρως συνδεδεμένο (FCL- Fully Connected Layer) τα δεδομένα υπόκειντο σε μια συνάρτηση σφάλματος.

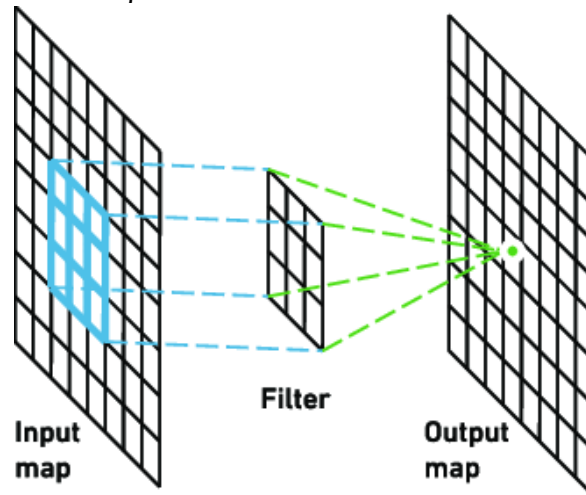


Εικόνα 3.7: Συνελκτικό Νευρωνικό Δίκτυο[18]

Τα επίπεδα σε κάθε τέτοιο μοντέλο έχουν κάποια χαρακτηριστικά και ανάλογα με τις ιδιότητές τους χωρίζονται σε 3 κατηγορίες οι οποίες περιγράφονται στην συνέχεια.

3.3.1 Επίπεδο Συνέλιξης (Convolutional Layer)

Είναι συνήθως το πρώτο επίπεδο και εφαρμόζεται για την εξαγωγή διαφόρων χαρακτηριστικών από τις εικόνες- εισόδους. Σε αυτό το επίπεδο πραγματοποιείται η μαθηματική διαδικασία της συνέλιξης ανάμεσα στην εικόνα και σε ένα φίλτρο $M \times M$. Το φίλτρο περνάει από όλα τα σημεία της εικόνας και δίνεται το τελικό προϊόν.



Εικόνα 3.8: Διαδικασία Συνέλιξης[19]

Η διαδικασία αυτή όπως φαίνεται και στο παραπάνω σχήμα δίνει μια νέα εικόνα με πληροφορία για τις γωνίες, τις ακμές και άλλα χαρακτηριστικά. Ύστερα αυτός ο χάρτης χαρακτηριστικών εισάγεται σε άλλα επίπεδα για την εξαγωγή νέων σημαντικών χαρακτηριστικών.

Το επίπεδο συνέλιξης έχει τέσσερις υπερπαραμέτρους:

- Τον αριθμό των φίλτρων K
- Το μέγεθος F φίλτρων. Κάθε φίλτρο έχει διαστάσεις $F \times F \times D$ εικονοστοιχείων, όπου D είναι ο αριθμός των καναλιών των εισαγόμενων εικόνων
- Το S είναι το βήμα που κινείται το φίλτρο επάνω στην εικόνα
- Και τέλος το γέμισμα με μηδενικά P , όπου αποφασίζεται το γέμισμα περιμετρικά της εικόνας με μαύρα εικονοστοιχεία για να μπορεί να πραγματοποιηθεί η διαδικασία της συνέλιξης σε ολόκληρη την εικόνα.

Σε κάθε εικόνα με μέγεθος $W \times H \times D$, το επίπεδο συνέλιξης επιστρέφει ένα πίνακα $W_c \times H_c \times D_c$, όπου:

$$W_c = \frac{W - F + 2P}{S} + 1$$

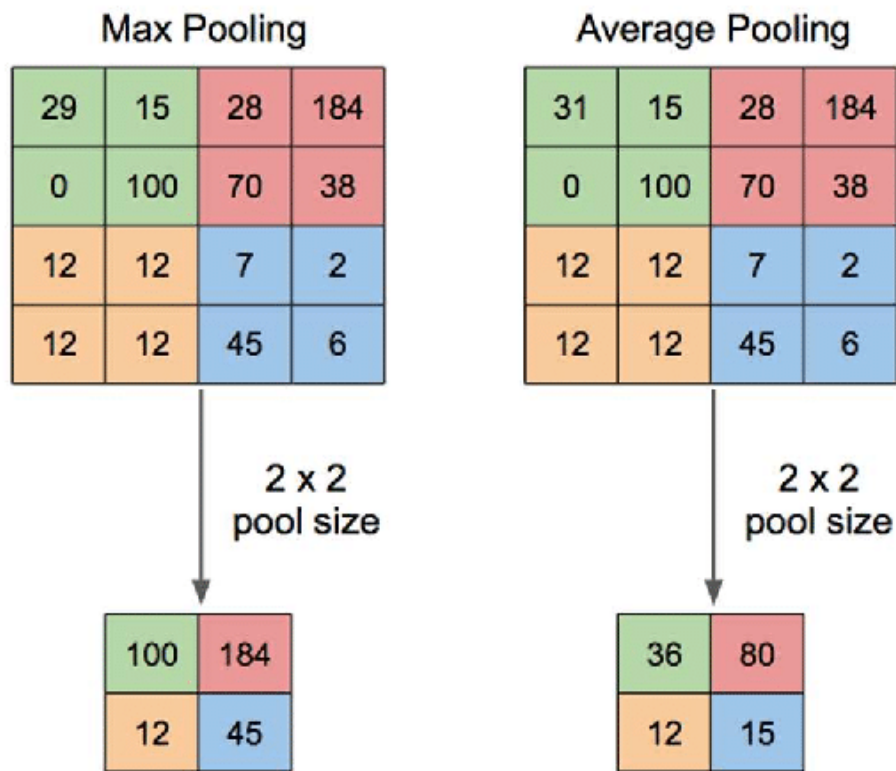
$$H_c = \frac{H - F + 2P}{S} + 1$$

$$D_c = K$$

3.3.2 Επίπεδο Υπό-Δειγματοληψίας (Pooling Layer)

Στις περισσότερες περιπτώσεις ένα επίπεδο συνέλιξης ακολουθείται από ένα επίπεδο υπό-δειγματοληψίας. Κύριος σκοπός αυτού του επιπέδου είναι να μειωθεί το υπολογιστικό κόστος. Αυτό πραγματοποιείται με την μείωση συνδέσμων μεταξύ επιπέδων και ανεξάρτητων εφαρμογών σε κάθε χάρτη χαρακτηριστικών (feature map). Ανάλογα με την μέθοδο, υπάρχουν πολλοί τύποι υπό-δειγματοληψίας.

- Υπό-Δειγματοληψία Μέγιστου (Max Pooling): το μεγαλύτερο στοιχείο παίρνεται από τον χάρτη χαρακτηριστικών
- Υπό-Δειγματοληψία Μέσου (Average Pooling): γίνεται ο υπολογισμός του μέσου όρου σε ένα κομμάτι της εικόνας με ορισμένο μέγεθος



Εικόνα 3.9: Μέθοδοι Υπό-Δειγματοληψίας[20]

Συνήθως το επίπεδο Υπό-Δειγματοληψίας συνδέει ένα επίπεδο συνέλιξης (Conv) με το πλήρως συνδεδεμένο (FC).

Το επίπεδο Υπό-Δειγματοληψίας έχει δύο υπερπαράμετρους:

- Το μέγεθος F των κελιών, η εικόνα είναι χωρισμένη σε τετράγωνα μεγέθους $F \times F$ εικονοστοιχείων.

- Το βήμα S , τα κελιά είναι χωρισμένα το ένα από το άλλο με S εικονοστοιχεία.

Σε κάθε εικόνα με μέγεθος $W \times H \times D$, το επίπεδο συνέλιξης επιστρέφει ένα πίνακα $W_p \times H_p \times D_p$, όπου:

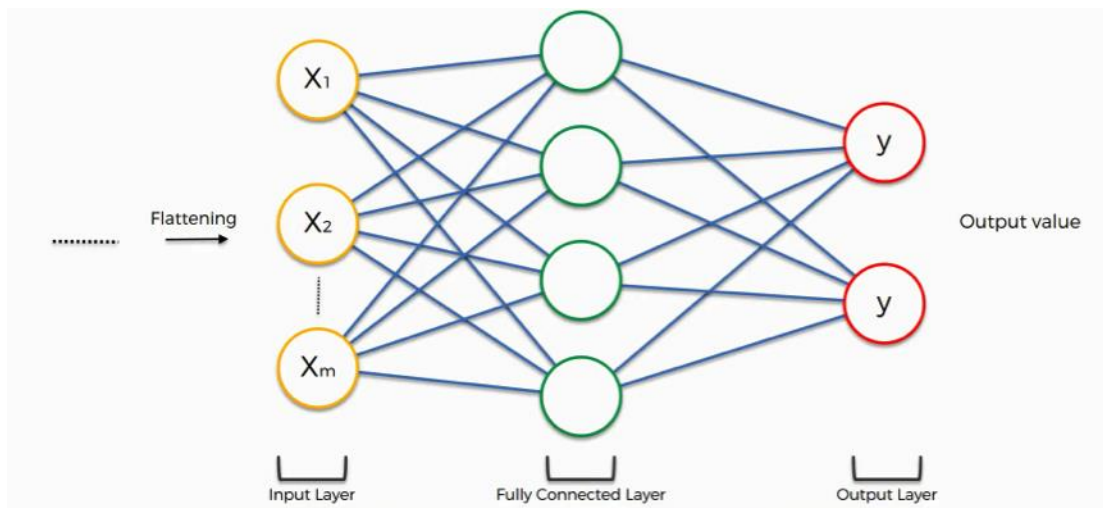
$$W_p = \frac{W - F}{S} + 1$$

$$H_p = \frac{H - F}{S} + 1$$

$$D_p = D$$

3.3.3 Πλήρως Συνδεδεμένο Επίπεδο (Fully Connected Layer)

Το πλήρως συνδεδεμένο δίκτυο (FC) είναι συνήθως το τελευταίο επίπεδο ενός συνελκτικού νευρωνικού δικτύου. Αποτελείται από μια γραμμική συνάρτηση και μια συνάρτηση ενεργοποίησης. Με αυτές τις συναρτήσεις το FC κατηγοριοποιεί τις εικόνες και επιστρέφει ένα διάνυσμα μεγέθους N , όπου N είναι ο αριθμός κατηγοριών που έχει επιλεχθεί για το πρόβλημα που έχει κληθεί ο αλγόριθμος να αντιμετωπίσει.



Εικόνα 3.10: Πλήρως Συνδεδεμένο Επίπεδο³

3.4 Αρχιτεκτονικές Βαθιάς Μηχανικής Μάθησης και η Επιστήμη της Μηχανικής Όρασης

Σήμερα η Βαθιά Μηχανική Μάθηση (deep learning), και κυρίως τα CNNs, χρησιμοποιείται εκτενώς σε υπολογιστικά συστήματα που έχουν εμπνευστεί από την βιολογία. Οι αρχιτεκτονικές

³ Πηγή: <https://www.superdatascience.com/blogs/convolutional-neural-networks-cnn-step-4-full-connection>

της Βαθιάς Μηχανικής Μάθησης αφορούν στα επίπεδα που χρησιμοποιούνται και στην σειρά που εκτελούνται για την αποτελεσματικότερη αντιμετώπιση του εκάστοτε προβλήματος.

Η επιστήμη της Μηχανικής Όρασης (Computer Vision) στοχεύει στην επίλυση προβλημάτων όπως η αναγνώριση προτύπων, η αναγνώριση αντικειμένων και η κατάτμηση εικόνων σε επιμέρους στοιχεία. Αυτό προσπαθεί να πετύχει εισάγοντας αναλυτικά μαθηματικά μοντέλα ή και μοντέλα πιθανοτήτων, αλλά βασικότερο για την επίλυση αυτών των στοιχείων τα τελευταία χρόνια είναι η ένταξη της Βαθιάς Μηχανικής Μάθησης.

Οι πρώτες αρχιτεκτονικές συνελκτικών νευρωνικών δικτύων άρχισαν να εμφανίζονται την περίοδο 1980 με 1999. Πρώτο μοντέλο αρχιτεκτονικής ήταν το ConpNet. Κατασκευάστηκε το 1989 από τον LeCun και έδειξε μεγάλη επιτυχία στην αναγνώριση χειρόγραφων αριθμών και ταχυδρομικών κωδικών[21]. Αυτή η αρχιτεκτονική έθεσε τα θεμέλια στα μοντέρνα δισδιάστατα αρχιτεκτονικά μοντέλα CNN. Ύστερα ο LeCun αποφάσισε να εξελίξει την αρχιτεκτονική αυτή δημιουργώντας μία πανίσχυρη για την εποχή αρχιτεκτονική που επίλυε πολλά προβλήματα όρασης υπολογιστών και αναγνώρισης αποτυπωμάτων. Αυτή ονομάστηκε LeNet-5 [22] και παρόλες τις επιτυχίες στην αναγνώριση οπτικών χαρακτήρων, δεν λειτουργούσε το ίδιο καλά και σε άλλα προβλήματα αναγνώρισης εικόνων.

Τα τέλη της δεκαετίας του 90 και οι αρχές του 2000 θεωρούνται η εποχή της παρακμής της Βαθιάς Μηχανικής Μάθησης. Αυτό συνέβη καθώς υπήρχαν πολύ λίγα στοιχεία για τον τρόπο λειτουργίας των CNN και η υπολογιστικές μονάδες της εποχής ήταν δύσκολο να πραγματοποιήσουν την εκπαίδευση με τόσο χρονοβόρες και σύνθετες αρχιτεκτονικές και τα δεδομένα που υπήρχαν τότε ήταν λίγα και δεν αρκούσαν για την σωστή εκπαίδευση ενός συστήματος. Αυτό είχε ως αποτέλεσμα την χρήση πιο απλών αλγορίθμων όπως ήταν ο αλγόριθμος Support Vector Machine (SVM) καθώς ήταν μία πολύ υψηλής απόδοσης στατιστική μέθοδος η οποία είχε χαμηλό υπολογιστικό κόστος[23].

Από τα τέλη 2006 και έπειτα πραγματοποιείται επάνοδος της Βαθιάς Μηχανικής Μάθησης σε μια νέα πιο σύγχρονη μορφή της με την χρήση καρτών γραφικών (GPU) για την επιτάχυνση της εκπαίδευσης των αρχιτεκτονικών CNN. Το 2007 η NVIDIA έβγαλε για πρώτη φορά την πλατφόρμα CUDA. Αυτή επέτρεπε την παράλληλη χρήση των επεξεργαστικών δυνατοτήτων των καρτών γραφικών[24].

Το 2015 μέχρι και σήμερα πραγματοποιούνται οι μεγαλύτερες καινοτομίες στις αρχιτεκτονικές CNN και κυρίως η δημιουργία αρχιτεκτονικών οι οποίες έχουν πολλές νέες παραμέτρους για την επίλυση απλών μέχρι και σύνθετων προβλημάτων. Τέτοιες μεγάλες αρχιτεκτονικές είναι οι VGG, ResNet, SegNet, U-Net, κ.α. Σε μερικές από αυτές τις αρχιτεκτονικές, που χρησιμοποιήθηκαν για την μελέτη, γίνεται εκτενέστερη αναφορά στα επόμενα κεφάλαια.

4. Υλοποίηση

Η συγκεκριμένη μελέτη αφορά στον αυτόματη κατάτμηση εικόνων με χρήση επιβλεπόμενων μεθόδων βαθιάς μηχανικής μάθησης και στοχεύει στον έλεγχο της αρχιτεκτονικής με τα λιγότερα σφάλματα σε ορθοφωτογραφίες RGB και RGB-D (Πρόβλημα Μηχανικής Όρασης).

Σε αυτό το κεφάλαιο γίνεται ανάλυση των δεδομένων (dataset) και των λογισμικών που χρησιμοποιήθηκαν για την αρχική επεξεργασία αυτού, των αρχιτεκτονικών CNN που χρησιμοποιήθηκαν για επιλογή της καλύτερης εξ αυτών για το συγκεκριμένο πρόβλημα και των αλγόριθμων που χρησιμοποιήθηκαν για τον τελικό έλεγχο και την σημασιολογική κατάτμηση των εικόνων.

4.1 Δεδομένα και Λογισμικά προεπεξεργασίας

Οι εικόνες που χρησιμοποιήθηκαν για την εκπαίδευση των διαφόρων αρχιτεκτονικών είναι ορθοφωτογραφίες από τον Αρχαιολογικό Χώρο της Πολιόχνης, καθώς και εικόνες βάθους της συγκεκριμένης περιοχής. Τα λογισμικά επεξεργασίας αυτών είναι το labelme κώδικας ανοιχτού λογισμικού για την δημιουργία κατηγοριοποιημένων εικόνων και ο κώδικας προσθήκης καναλιού βάθους σε εικόνα Adding_Depth_to_RGB, ο οποίος αναπτύχθηκε για την εκπόνηση της διπλωματικής εργασίας.

4.1.1 Ο Οικισμός της Πολιόχνης

Η Πολιόχνη είναι ένας αρχαιολογικός τόπος στην ανατολική ακτή της Λήμνου. Χρονολογείται ότι χτίστηκε στις αρχές της νεολιθικής περιόδου (4^η-5^η χιλιετία π.Χ.).

Ο οικισμός αυτός ανακαλύφθηκε από ανασκαφές της Ιταλικής αρχαιολογικής Σχολής Αθηνών το 1930, με επικεφαλής τον τότε διευθυντή Αλεσσάντρο Ντέλα Σέτα και τον βοηθό του Μπερνάμπο Μπρέα[25].

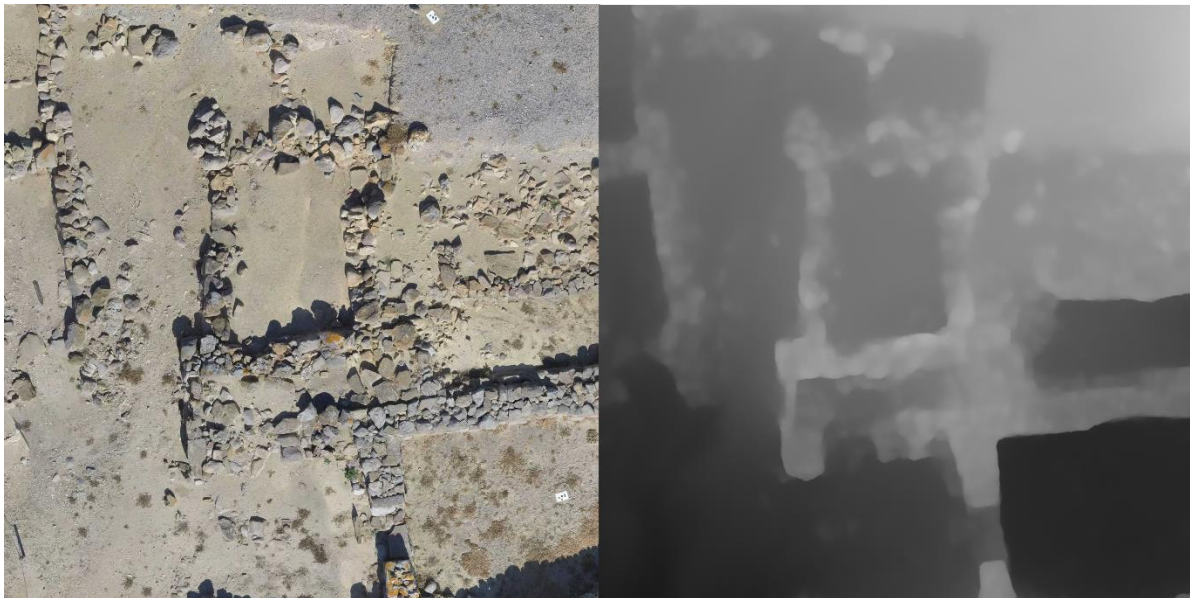
Περιβάλλεται από τείχη, που έχουν επιβλητική μορφή από την δυτική πλευρά, όπου ήταν μεγαλύτερη η ανάγκη προστασίας των θεμελίων από την διάβρωση του κοντινού ποταμού. Στην πόλη ανθούσε ένας ειρηνικός πληθυσμός χιλίων περίπου κατοικιών. Επίσης στην Πολιόχνη ήταν κατασκευασμένο και το επονομαζόμενο 'Βουλευτήριο', όπου μπορούσαν να παρθούν αποφάσεις για το σκάψιμο των φρεατίων, με κτιστή επένδυση, σε βάθος που έφτανε πάνω από τα 9 μέτρα, η κατασκευή και η συντήρηση των τειχών, των δρόμων και των αγωγών για τη συλλογή των νερών της βροχής, η δαπέδωση πλατειών όπου συναντιόταν οι κάτοικοι για να περάσουν τον ελεύθερο χρόνο τους ή με την ευκαιρία εορτών.

Επίσης με την κατεργασία και διακίνηση μετάλλων αναδείχθηκε ως το αρχαιότερο αστικό κέντρο και ως το πρώτο εμπορικό λιμάνι της Ευρώπης. Τα ευρήματα της Πολιόχνης αποδεικνύουν ότι τα

νησιά του Β.Α. Αιγαίου αποτελούσαν έναν προϊστορικό πολιτισμό ανεξάρτητο από τους πιο γνωστούς (Κυκλαδικό, Μινωικό, Μυκηναϊκό).⁴

4.1.2 Ορθοφωτογραφίες

Οι εικόνες που χρησιμοποιήθηκαν για την μελέτη είναι ορθοφωτογραφίες, δηλαδή εικόνες οι οποίες έχουν αναχθεί διαφορεικά για την εξάλειψη παραμορφώσεων των αρχικών εικόνων, που οφείλονταν στην επίδραση των κλίσεων, στις εκτροπές λόγω ανάγλυφου κ.α. Τα πλεονεκτήματα αυτών είναι η ενιαία κλίμακα, η ποιοτική πληροφορία που περιέχει, η μετρητική αξιοπιστία και αποτελεί ευέλικτο ψηφιακό προϊόν. Η παραγωγή μιας ορθοφωτογραφίας είναι μια πλέον αυτοματοποιημένη ψηφιακή μονοεικονική φωτογραμμετρική διαδικασία, που στοχεύει στην απεικόνιση του χώρου σε ορθή προβολή. Οι συγκεκριμένες ορθοφωτογραφίες παρήχθησαν από το Εργαστήριο Φωτογραμμετρίας στο πλαίσιο σχετικού Ερευνητικού Προγράμματος με την Εφορεία Αρχαιοτήτων Λέσβου.[26]



Εικόνα 4.1: Ορθοφωτογραφία και χάρτης βάθους (DEM) τμήματος του Αρχαιολογικού Χώρου

4.1.3 Λογισμικά και κώδικες Προεπεξεργασίας

Τα λογισμικά και οι κώδικες που χρησιμοποιήθηκαν για την προεπεξεργασία των αεροφωτογραφιών και την τελική εμφάνιση των δεδομένων είναι:






LabelMe

Το labelme είναι λογισμικό ανοιχτού κώδικα το οποίο χρησιμοποιείται κυρίως για κατασκευή δεδομένων (dataset) τα οποία χρήζουν σημασιολογικής κατάταξης. Συνεπώς, όπως λέει και το

⁴ Πληροφορίες από <https://el.wikipedia.org/wiki/Πολιόγνη>

όνομά του, έχει την δυνατότητα να κατηγοριοποιεί μέρη της εικόνας με βάση τις επιλογές κατάτμησης που έχει κάνει ο χρήστης[27].

Στην προκειμένη περίπτωση οι κατηγορίες που επιλέχθηκαν για την ορθότερη εκπαίδευση του κώδικα ύστερα από διερεύνηση του προβλήματος και του κώδικα ανοιχτού λογισμικού (labelme) είναι οι εξής:

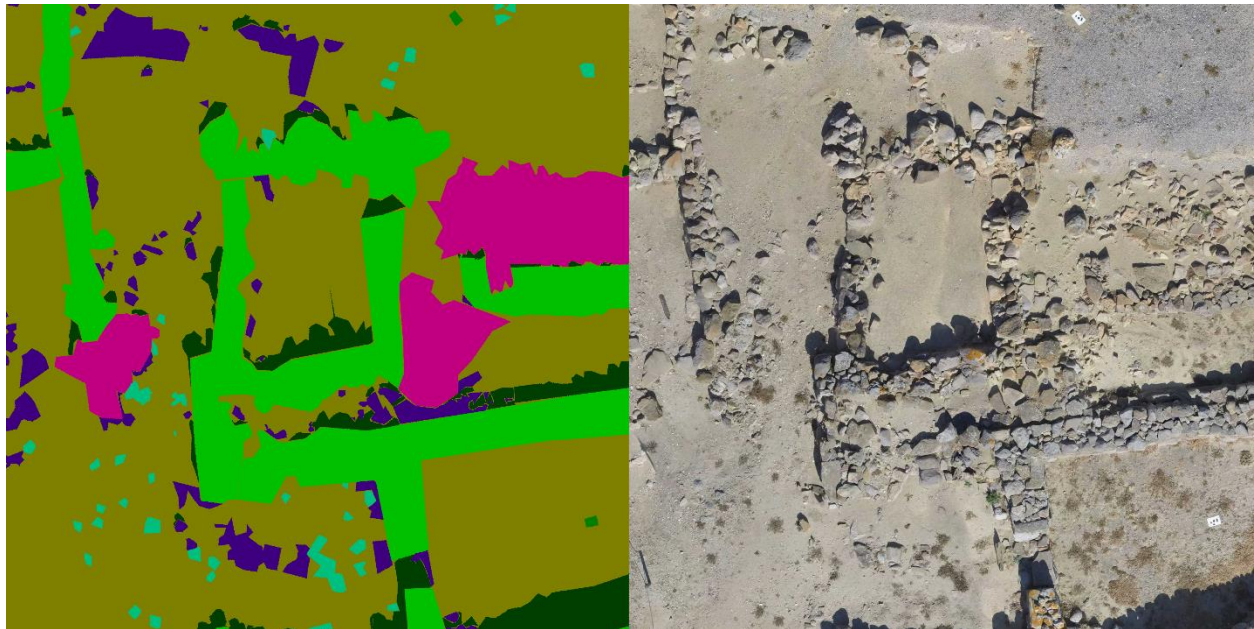
Εικόνα κατηγορίας	Όνομασία Κατηγορίας	Τιμή Χρωματικής απόδοσης Κατηγορίας (R,G,B)
	<p><u>Label: GCP</u> Σημεία γνωστών γεωδαιτικών συντεταγμένων στην εικόνα</p>	(0,128,0)
	<p><u>Label: Ground</u> Καθαρό έδαφος χωρίς εμπόδια όπως πέτρες φυτά και άλλα</p>	(128,128,0)
	<p><u>Label: Ruin</u> Ερείπιο του κτηρίου που πιθανώς βρισκόταν σε εκείνο το σημείο με μικρή διασπορά</p>	(192,0,128)
	<p><u>Label: Wall</u> Εμφανής λιθοδομή και αρκετά στέρεη ώστε να κρατάει την δομή της μέχρι σήμερα</p>	(0,192,0)
	<p><u>Label: Shadow</u> Σκιά η οποία δεν έχει αφαιρεθεί με κάποιο μέσο επεξεργασίας εικόνας</p>	(0,63,0)

	<p><u>Label: Dock</u> Εξέδρα/Τεχνητός διάδρομος που υπάρχει στην περιοχή μελέτης</p>	(128,0,0)
	<p><u>Label: Rocks</u> Πέτρες διάσπαρτες σε τυχαία σημεία της περιοχής μελέτης</p>	(63,0,128)
	<p><u>Label: Weeds</u> Ξερόχορτα διάσπαρτα σε τυχαία σημεία της περιοχής μελέτης</p>	(0,192,128)

Πίνακας 1: Πίνακας στοιχείων κατάταξης των εικόνων

Irfan View

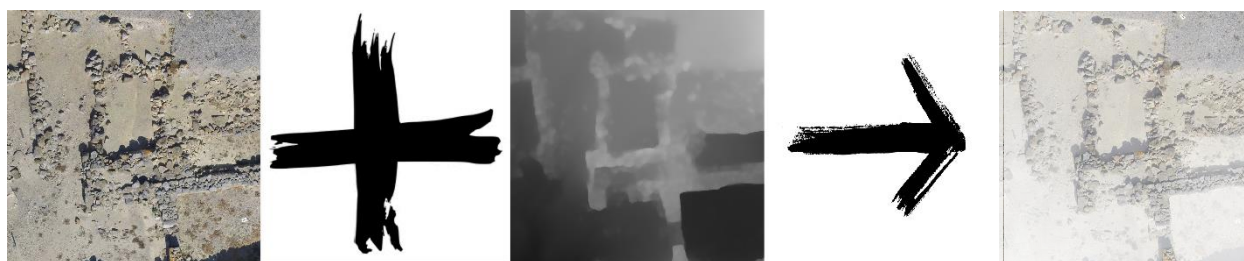
Το λογισμικό αυτό χρησιμοποιήθηκε για την επεξεργασία των εικόνων όσον αφορά τα χρώματα των εικόνων αλήθειας πεδίου (ground truth), καθώς και για την αποκοπή κομματιών εικόνας για την δημιουργία δεδομένων εκπαίδευσης (training dataset) μικρότερου όγκου για δοκιμές ρυθμίσεων στον κύριο κώδικα.



Εικόνα 4.2: Κομμάτι δεδομένων αλήθειας πεδίου (Ground Truth) στα αριστερά και πραγματικής εικόνας στα δεξιά

Adding_Depth_to_RGB

Ο αλγόριθμος αυτός αναπτύχθηκε στο πλαίσιο της διπλωματικής αυτής για την εισαγωγή στις εικόνες ενός τέταρτου καναλιού, το οποίο δίνει το βάθος αυτών. Αυτό κατασκευάστηκε ύστερα από επεξεργασία του ψηφιακού μοντέλου εδάφους της περιοχής (DEM) και της μετατροπής αυτού σε ένα κανάλι του φάσματος του γκρι (grayscale channel). Με αυτόν τον τρόπο δημιουργήθηκαν οι RGB-D εικόνες που χρησιμοποιήθηκαν για την εκπαίδευση του αλγορίθμου.



Εικόνα 4.3: Προσθήκη βάθους σε RGB εικόνα και δημιουργία RGB-D εικόνας

4.2 Αρχιτεκτονικές CNN

Η βασική βιβλιοθήκη ανοιχτού λογισμικού που χρησιμοποιήθηκε για την ανάπτυξη των αρχιτεκτονικών CNN κατά την διάρκεια της έρευνας είναι η pytorch. Η βιβλιοθήκη αυτή της γλώσσας προγραμματισμού Python, πραγματοποιεί άμεση εκτέλεση δυναμικών τανυστών (tensor) για επιτάχυνση της μονάδας επεξεργασίας γραφικών (GPU), χωρίς να υστερεί σε απόδοση σε σχέση με τις πιο γρήγορες βιβλιοθήκες μηχανικής μάθησης[28].

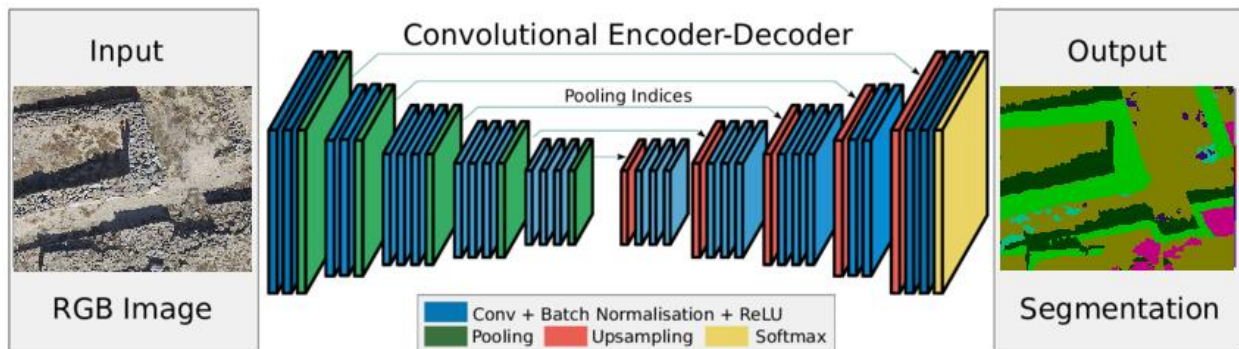
Η βιβλιοθήκη αυτή προσφέρει την δυνατότητα χρήσης κάποιων έτοιμων αρχιτεκτονικών CNN προς χρήση για εκπαίδευση, καθώς και την χρήση διαφόρων layer (pooling, convolutional, κ.α.) τα οποία είναι ήδη έτοιμα για χρήση χωρίς να χρήζουν κάποια περαιτέρω επεξεργασίας.

Όπως έχει ήδη αναφερθεί, η δομή των αρχιτεκτονικών CNN αποτελείται από διάφορα συνδεδεμένα επίπεδα τα οποία επεξεργάζονται τα δεδομένα που τους έχουν δοθεί και βγάζουν αποτελέσματα για την ορθότερη κατάτμηση, στην προκειμένη περίπτωση, των εικόνων. Οι αρχιτεκτονικές που χρησιμοποιήθηκαν στην μελέτη αυτή είναι οι SegNet, U-Net, ResNet50_fcn, ResNet50_dlv3 ResNet101_fcn,, ResNet101_dlv3.

4.2.1 SegNet

Αυτή η αρχιτεκτονική δημιουργήθηκε για την κατηγοριοποίηση εικονοστοιχείων εικόνων. Η βάση της είναι ένα δίκτυο κωδικοποιητών (encoder network) και ένα αντίστοιχο δίκτυο

αποκωδικοποιητών (decoder network). Το δίκτυο κωδικοποιητών είναι πανομοιότυπο με την αρχιτεκτονική VGG16[29]. Το δίκτυο κωδικοποιητών αποτελείται από 13 επίπεδα τα οποία είναι σχεδιασμένα για κατηγοριοποίηση. Κάθε ένα από τα στοιχεία των κωδικοποιητών πρέπει να έχει και ένα αντίστοιχο στοιχείο αποκωδικοποίησης, συνεπώς και το δίκτυο των αποκωδικοποιητών αποτελείται από 13 στοιχεία. Τέλος τα προϊόντα του τελευταίου στοιχείου τροφοδοτούνται σε μια συνάρτηση κατηγοριοποίησης πολλών κλάσεων γνωστή ως soft-max για να δημιουργηθούν πιθανότητες κατηγορίας για κάθε εικονοστοιχείο ξεχωριστά[14].

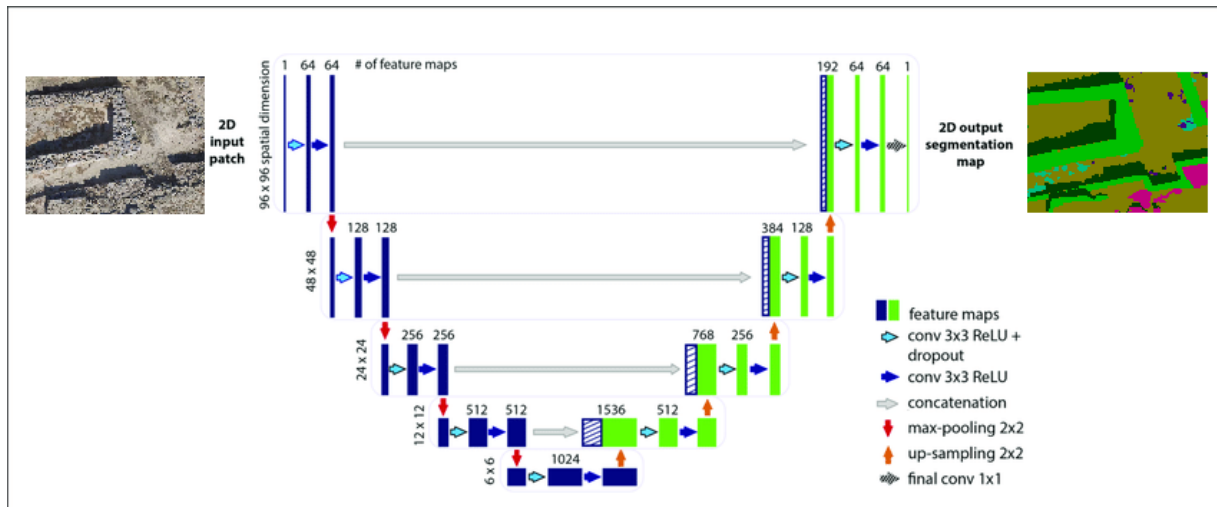


Εικόνα 4.4: Παράσταση Αρχιτεκτονικής SegNet με τροποποίηση ως προς τα δεδομένα εισαγωγής και εξαγωγής[14]

Όπως φαίνεται και στην Εικόνα 4.4 η αρχιτεκτονική SegNet περιλαμβάνει επίπεδα συνέλιξης, συναρτήσεις ενεργοποίησης και επίπεδα Υπό-δειγματοληψίας.

4.2.2 U-Net

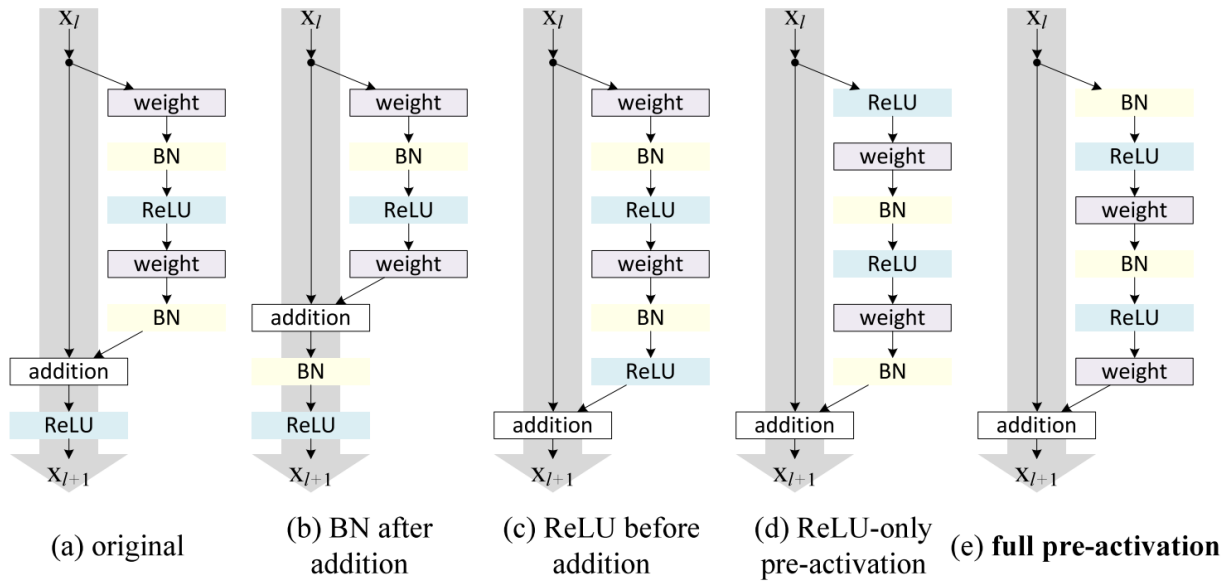
Η αρχιτεκτονική αυτή αποτελείται από ένα συστελλόμενο και ένα επεκτεινόμενο κομμάτι. Το συστελλόμενο κομμάτι, ακολουθώντας την τυπική αρχιτεκτονική συνελκτικού δικτύου, περιλαμβάνει επαναλαμβανόμενες εφαρμογές φίλτρων συνελίξεων, ακολουθούμενες, η κάθε μία, από μια συνάρτηση ενεργοποίησης ReLU και μια υπό-δειγματοληψία. Σε κάθε βήμα του συστελλόμενου κομματιού γίνεται διπλασιασμός των χαρακτηριστικών καναλιών. Το επεκτεινόμενο κομμάτι αποτελείται από επαναλήψεις αντίθετων βημάτων σε σχέση με το συστελλόμενο κομμάτι, Τέλος γίνεται εφαρμογή μίας 1x1 συνέλιξης για τα κατηγοριοποιηθεί κάθε εξηνατετραθέσιο διάνυσμα στον επιθυμητό αριθμό κλάσεων. Η αρχιτεκτονική αυτή αποτελείται συνολικά από 24 επίπεδα[12]. Μια μορφή αυτής της αρχιτεκτονικής παρουσιάζεται στην Εικόνα 4.5.



Εικόνα 4.5: Παράσταση Αρχιτεκτονικής U-Net με τροποποίηση ως προς τα δεδομένα εισαγωγής και εξαγωγής [12]

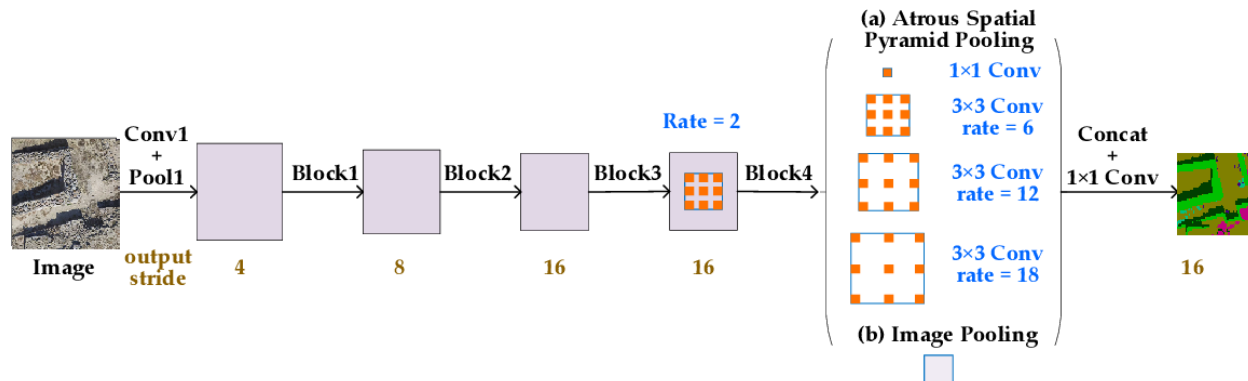
4.2.3 ResNet

Εμπνευσμένη η βάση της από το αρχιτεκτονικό μοντέλο της VGG-19[30] η αρχιτεκτονική ResNet χρησιμοποιεί συνδέσεις συντόμευσης που μετατρέπουν το δίκτυο στην υπολειπόμενη έκδοση αυτού. Οι συντομεύσεις ταυτοποίησης, που δίνονται από τον τύπο $y = F(x, (\{W_i\}) + x$, μπορούν να ενταχθούν απευθείας όταν οι έξοδοι ταιριάζουν ως προς τις διαστάσεις με τις εισόδους. Όταν υπάρχει αύξηση του μεγέθους των διαστάσεων τότε είτε χρησιμοποιούνται απλά επιπλέον μηδενικά γύρω από κάθε εικόνα για την μη ένταξη νέας μεταβλητής στον κώδικα, είτε χρησιμοποιείται μια γραμμική προβολή (W_s) στην συντόμευση ταυτοποίησης, με την συνάρτηση να διαμορφώνεται ως $y = F(x, (\{W_i\}) + W_s x$, για να ταιριάζουν τα μεγέθη των διαστάσεων με την χρήση συνελίξεων 1×1 . Και στις δυο περιπτώσεις το βήμα (stride) είναι 2, όταν παρουσιάζονται χάρτες χαρακτηριστικών (feature maps) 2 μεγεθών στο ενδιάμεσο από κάθε συντόμευση[10].



Εικόνα 4.6: Λογική των Δομών Συντόμευσης της ResNet[31]

Το ResNet χωρίζεται σε επιμέρους αρχιτεκτονικές που χαρακτηρίζονται από τις πόσες απλές δομές ή δομές συμφόρησης παρουσιάζονται στην αρχιτεκτονική και συνεπώς το πόσα επίπεδα απαρτίζουν την κάθε μια εξ αυτών. Στην έρευνα αυτή αξιοποιούνται οι αρχιτεκτονικές ResNet50, ResNet101 (50 και 100 επιπέδων αντίστοιχα), DeeplabV3-ResNet50, DeeplabV3-ResNet101. Το DeepLabV3[11] είναι μια ακόμη αρχιτεκτονική CNN η οποία χρησιμοποιείται σε συνδυασμό με τις ResNet50 και ResNet101 και εντάσσει σε αυτές τα στοιχεία atrous convolution (συνέλιξη με βήμα) και atrous spatial pyramid pooling (υπό-δειγματοληψία με πολλαπλές συνέλιξεις με βήμα).



Εικόνα 4.7: παράσταση αρχιτεκτονικής DeeplabV3 με Atrous Spatial Pyramid Pooling και Atrous Convolution με τροποποίηση ως προς τα δεδομένα εισαγωγής και εξαγωγής [32]

4.3 Κώδικας εκπαίδευσης

Ο κώδικας ανοιχτού λογισμικού για την ορθή κατάτμηση με βάση τις κατηγορίες που έχουν οριστεί μεταφορτώθηκε από την έρευνα “Deep Learning for Earth Observation” που υπάρχει στην

σελίδα Github (ιστοχώρος διάθεσης ελεύθερου και μη λογισμικού)⁵. Ο κώδικας αυτός διατίθεται μόνο για ερευνητική χρήση, συνεπώς αυτός ο κώδικας χρησιμοποιήθηκε για την εκπόνηση της εργασίας, ύστερα από μερικές τροποποιήσεις με βάση το dataset και τις ανάγκες αυτής.

Η ανάπτυξη αυτού του κώδικα πραγματοποιήθηκε με στόχο την εξερεύνηση νευρωνικών δικτύων σε βάθος για κατάτμηση πολύ υψηλής ανάλυσης τηλεπισκοπικών εικόνων πόλεων σε κατηγορίες. Η έρευνα αυτή έδειξε πόσο καλή είναι η λειτουργία των πλήρως συνδεδεμένων νευρωνικών δικτύων σε κατάτμηση εικόνων[33], για αυτό και επιλέχθηκε αυτός ο κώδικας, καθώς μπορεί να δείξει την αποτελεσματικότητα των διαφόρων αρχιτεκτονικών που εξετάστηκαν. Επίσης η δυνατότητα αυτού να εξάγει εικόνες οι οποίες είναι κατατμημένες, καθώς και στατιστικά στοιχεία για την ακρίβεια κατάτμησης αυτών, ύστερα από την εκπαίδευση με τις διάφορες αρχιτεκτονικές είναι άλλο ένα στοιχείο που τον έκρινε ιδανικό.

4.3.1 Περαιτέρω Επεξήγηση βασικών στοιχείων

Οι κώδικες που χρησιμοποιούνται για Βαθιά Μηχανική Μάθηση έχουν κάποια βασικά στοιχεία, τα οποία βοηθούν στην βελτιστοποίηση του αποτελέσματος του κάθε μοντέλου ανάλογα με κάποια χαρακτηριστικά του προβλήματος ή των δεδομένων (dataset) που μελετάται.

Μέγεθος παραθύρου (Window size)

Αφορά στο μέγεθος που θα κοπεί κάθε εικόνα για την εκπαίδευση του μοντέλου, καθώς η εκπαίδευση λόγω του μεγάλου όγκου μνήμης που χρειάζεται δεν μπορεί να εκπαιδευτεί το μοντέλο σε κάθε κύκλο με ολόκληρη την εικόνα σαν εισαγωγή.

Αριθμός εικόνων που θα χρησιμοποιηθούν (Batch size)

Η κάθε εικόνα μέσα στον κώδικα χωρίζεται σε μικρότερες με το window size, αλλά πρέπει να επιλεγθεί σε κάθε κύκλο εκπαίδευσης ο αριθμός εικόνων που θα χρησιμοποιηθούν για την εκπαίδευση (training) και τις δοκιμές (testing). Με αυτό τον τρόπο μικραίνει η ανάγκη σε προσωρινή μνήμη κατά την διάρκεια εκπαίδευσης, και βελτιστοποιεί το μοντέλο, καθώς δημιουργεί σημεία αναφοράς ύστερα από κάθε δοκιμή για τις καλύτερες επιλογές βαρών.

Βήμα Μάθησης (Learning Rate)

Σε κάθε κύκλο εκπαίδευσης πρέπει να ανανεωθούν τα βάρη, συνεπώς το βήμα μάθησης (learning rate) βοηθάει έτσι ώστε αυτή η μεταβολή των βαρών να μην γίνεται με εξωφρενικά γρήγορους ή αργούς ρυθμούς έτσι ώστε να υπάρχει μια ισορροπία στα τελικά αποτελέσματα. Επίσης για την ακόμη καλύτερη χρήση του βήματος μάθησης υπάρχουν μέθοδοι οι οποίες αλλάζουν αυτό το βήμα με το πέρας κάποιων κύκλων εκπαίδευσης. Αυτά τα βήματα μάθησης που χρησιμοποιούν αυτές τις μεθόδους ονομάζονται δυναμικά βήματα μάθησης[34].

Συναρτήσεις Βελτιστοποίησης (Optimizers)

⁵ Ιστοθέση κώδικα εκπαίδευσης <https://github.com/nshaud/DeepNetsForEO>

Η συνάρτηση βελτιστοποίησης (optimizer) είναι ίσως το βασικότερο κομμάτι του αλγορίθμου, λόγω του ότι από την επιλογή αυτής εξαρτάται η διάρκεια εκπαίδευσης και η ακρίβεια προβλέψεων του μοντέλου. Ουσιαστικά είναι διαφορικές συναρτήσεις απώλειας των οποίων τα διανύσματα δίνονται από την μερική παράγωγο της συνάρτησης αυτής. Αν οριστεί λοιπόν μια συνέχεια θ_i από επαναλήψεις που έχουν τοπικό ελάχιστο το θ^* , τότε η συνέχεια δομείται από ένα κανόνα ανανέωσης LR, το ιστορικό προηγούμενων επαναλήψεων και τιμών κλίσης (H_i) και κάποια παραμέτρους φ , έχοντας πάντα μια αρχική παραμετρική τιμή θ_0 . Συνεπώς δομείται η γενική συνάρτηση βελτιστοποίησης[35].

$$\theta_i = LR(H_i, \varphi_i)$$

Εποχές (Epochs)

Μια εποχή (epoch) είναι ο όρος που χρησιμοποιείται για να δειχθεί ο αριθμός επαναλήψεων που πραγματοποιήθηκαν σε ολόκληρο το αρχείο δεδομένων που αποτέλεσαν την είσοδο στον κώδικα. Αυτό συμβαίνει διότι συνηθίζεται τα αρχεία δεδομένων (datasets) να καταλαμβάνουν μεγάλο χώρο συνεπώς χωρίζονται σε μικρότερα κομμάτια (batches) και συνήθως ο όρος επανάληψη αφορά στις πιο απλές διεργασίες όπως η διαδικασία εκπαίδευσης ενός από τα μικρότερα αυτά κομμάτια.

4.3.2 Εμβάθυνση στη Λειτουργία του Κώδικα

Ο κώδικας απαρτίζεται από 6 κύρια μέρη:

- **Εισαγωγή δεδομένων και ρυθμίσεων:** Σε αυτό το κομμάτι γίνεται η εισαγωγή και ο έλεγχος των δεδομένων με μια πρώτη προβολή αυτών, καθώς και η εισαγωγή των ρυθμίσεων που θα χρειαστεί ο κώδικας για να πραγματοποιήσει την εκπαίδευση του μοντέλου. Οι ρυθμίσεις αυτές αφορούν στο μέγεθος των δεδομένων που θα εισάγονται σε κάθε επανάληψη εκπαίδευσης (batch size), το βήμα δοκιμών (stride) και το μέγεθος του κάθε κομματιού της εικόνας προς εκπαίδευση (window size).
- **Αρχιτεκτονικές CNN:** Εδώ εισάγονται οι αρχιτεκτονικές οι οποίες επιλέχθηκαν για την μελέτη, και γίνεται η επιλογή της κάθε μιας ανά τρέξιμο για την παραγωγή μοντέλου σημασιολογικής κατάτμησης από κάθε μια από αυτές.
- **Διαχωρισμός δεδομένων σε δεδομένα εκπαίδευσης και δεδομένα δοκιμής:** Ο διαχωρισμός αυτός γίνεται για να επιλεγεί ο αριθμός κομματιών της εικόνας ανάλογα με το batch size που θα χρησιμοποιηθούν για έλεγχο και ο αριθμός που θα χρησιμοποιηθεί για την εκπαίδευση. Αυτή η διαδικασία γίνεται για την καλύτερη επαλήθευση του μοντέλου, την αποφυγή της υπέρ-προσαρμογής (over-fitting).
- **Βήμα Εκμάθησης και Συνάρτηση βελτιστοποίησης:** Εδώ γίνεται επιλογή του βήματος εκμάθησης (learning rate), το οποίο στη προκειμένη περίπτωση είναι δυναμικό και βασικό για την βέλτιστη εκπαίδευση, καθώς και η συνάρτηση βελτιστοποίησης της εκπαίδευσης του μοντέλου. Στην περίπτωση αυτή επιλέχθηκε η συνάρτηση SGD[36]
- **Συνάρτηση εκπαίδευσης δικτύου (train):** Η συνάρτηση αυτή είναι ουσιαστικά η επαναληπτική διαδικασία που πραγματοποιείται για την εκπαίδευση και την κατασκευή

του μοντέλου πιθανοτήτων για την διεκπεραίωση των κατατμήσεων. Εδώ εισάγεται και ο αριθμός εποχών (epochs) που θα πραγματοποιηθούν για την εκπαίδευση του μοντέλου.

- **Συνάρτηση δοκιμής δικτύου (test):** Αποτελεί το τελευταίο στάδιο του κώδικα καθώς εδώ γίνεται δοκιμή σε εικόνες, οι οποίες έχουν υποστεί επεξεργασία κατάτμησης από κάποιο φυσικό πρόσωπο και συνεπώς υπάρχει για αυτές εικόνα αλήθειας πεδίου (ground truth), η οποία συγκρίνεται με την εικόνα πρόβλεψης (prediction) για την εύρεση της ακρίβειας με την οποία έγινε η τελική αναγνώριση.

Η εκπαίδευση των διαφόρων μοντέλων πραγματοποιήθηκε σε υπολογιστή με επεξεργαστή i7-9750H⁶, προσωρινή μνήμη μεγέθους 16 gigabyte και κάρτα γραφικών RTX2070⁷. Η διάρκεια της κάθε εκπαίδευσης εξαρτήθηκε από τον αριθμό εποχών (epochs) και για 5 εποχές διήρκησε από 1 έως 3 ώρες και για 50 εποχές από 9 έως 14 ώρες. Η διάρκεια δοκιμών ήταν από 1 έως 3 λεπτά.

⁶ Πληροφορίες Επεξεργαστή: <https://ark.intel.com/content/www/us/en/ark/products/191045/intel-core-i7-9750h-processor-12m-cache-up-to-4-50-ghz.html>

⁷ Πληροφορίες Κάρτας Γραφικών: <https://www.nvidia.com/en-eu/geforce/graphics-cards/rtx-2070/>

5. Εκπαίδευση και αποτελέσματα

Η εκπαίδευση του κώδικα των μοντέλων χωρίζεται σε 2 στάδια. Πρώτο στάδιο αφορά στην εκπαίδευση του κώδικα σε μικρότερη κλίμακα, δηλαδή λιγότερες εποχές, για το κάθε μοντέλο για επιλογή του βήματος μάθησης (learning rate). Δεύτερο στάδιο είναι η εκπαίδευση με τις τελικά επιλεγμένες ρυθμίσεις ανάλογα με τις ανάγκες των δεδομένων (dataset) και τις δυνατότητες των υπολογιστών που χρησιμοποιήθηκαν.


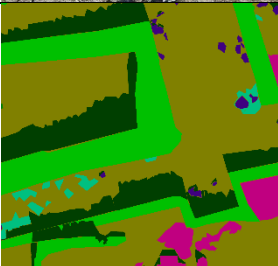
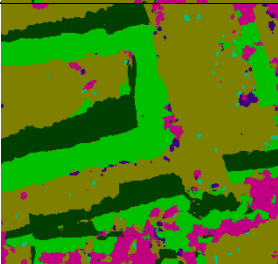
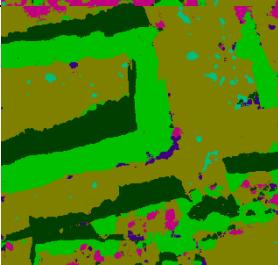
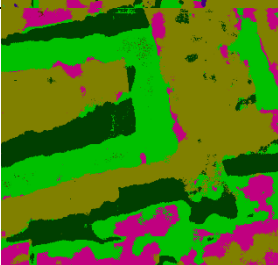
5.1 Στάδιο πρώτο

Η εκπαίδευση σε αυτό το στάδιο έγινε σε 5 εποχές για κάθε αρχιτεκτονική, καθώς και το κάθε dataset που χρησιμοποιήθηκε και με βήμα μάθησης 0.01, 0.001 και 0.0001. Η επιλογή των βημάτων αυτών έγινε διότι ο κώδικας δόθηκε με αυτές ως βέλτιστες παραμέτρους για την κατάτμηση των εικόνων και επειδή μέσα στον κώδικα γίνεται χρήση μια συνάρτησης δυναμικής αλλαγής βήματος μάθησης, συνεπώς δεν μπορεί να υφίσταται μικρότερο βήμα εκπαίδευσης από αυτά. Επίσης λόγω του ότι δεν βρέθηκε κάποια εφαρμογή των αρχιτεκτονικών ResNet για κατάτμηση εικόνων RGB-D αυτές εξετάστηκαν μόνο σε RGB εικόνες.

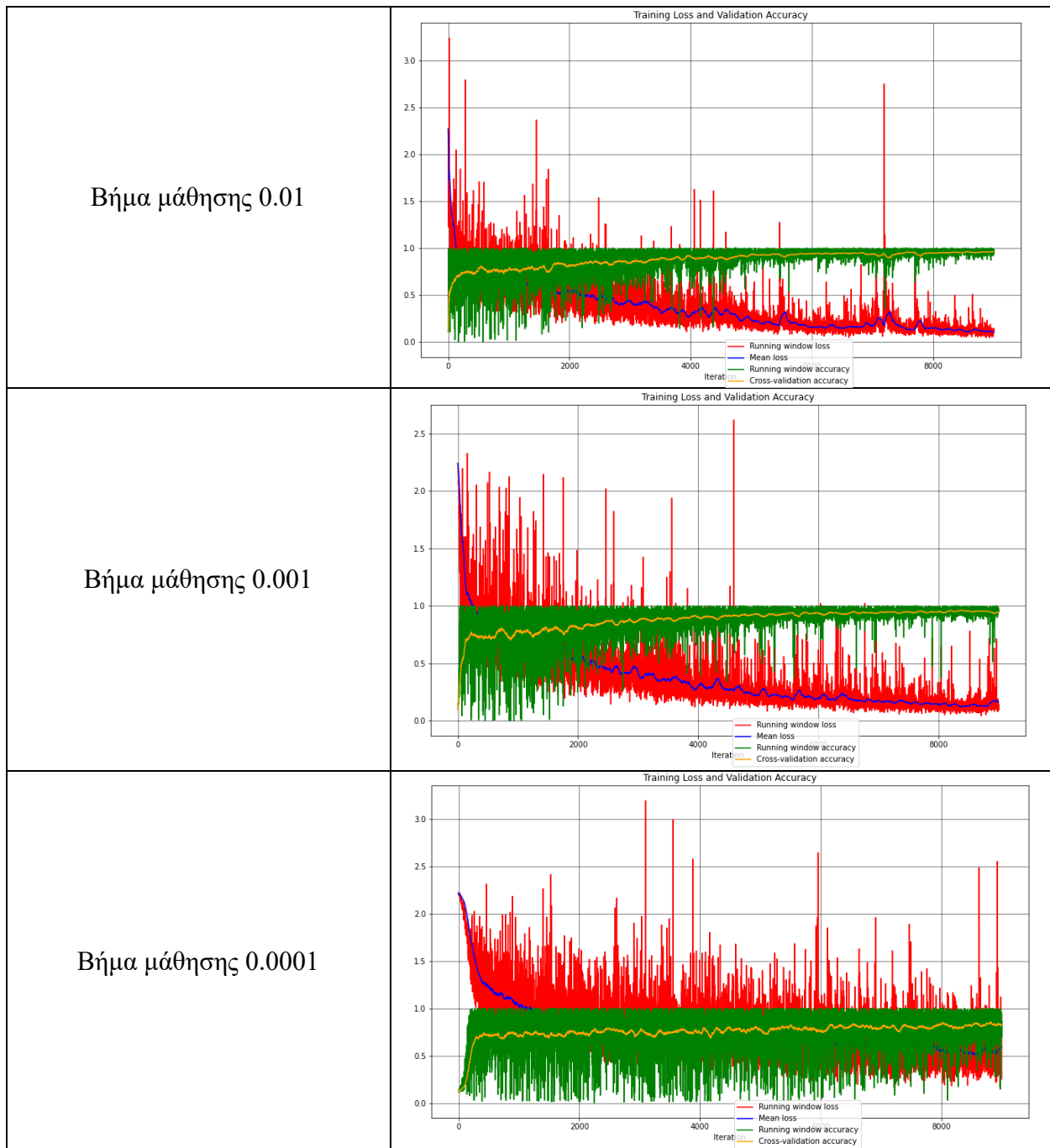
Κατά τη διαδικασία των εκπαιδεύσεων για την καλύτερη επιλογή των τελικών μοντέλων του πρώτου σταδίου λήφθηκαν υπόψη οι εξής παράμετροι:.

- Η διασταυρούμενη ακρίβεια εκπαίδευσης η οποία δίνεται από το ποσοστό ακρίβειας αναγνώρισης στις εικόνες πρόβλεψης (prediction images) σε σχέση με τις εικόνες αλήθειας πεδίου (ground truth images), για τον υπολογισμό αυτής χρησιμοποιήθηκε η μέση απώλεια, η οποία υπολογίστηκε με χρήση της συνάρτησης της διασταυρούμενης-εντροπίας (cross-entropy)[37].
- Η διασταυρούμενη ακρίβεια δοκιμών, η οποία δεν χρησιμοποιήθηκε για την εκπαίδευση του μοντέλου, αλλά για τις έλεγχο της κατάτμησης στην εικόνα δοκιμής.
- Η ομαλότητα των γραφικών παραστάσεων που δημιουργήθηκαν με αναπαράσταση των στοιχείων της μέσης απώλειας ανά συγκεκριμένο αριθμό επαναλήψεων, των τοπικών απωλειών, των τοπικών ακριβειών και της μέσης διασταυρούμενης ακρίβειας εκπαίδευσης. Η ομαλότητα αυτή ορίζεται ως το κατά πόσο η διασταυρούμενη ακρίβεια τείνει προς την τιμή 1 κατά το πέρασμα των επαναλήψεων χωρίς την ύπαρξη μεγάλων διαφοροποιήσεων σε όλο το εύρος αυτής, καθώς και η τιμή των μέσων απωλειών πόσο ομαλά τείνει προς την τιμή 0 χωρίς αποκλίσεις σε όλο το εύρος αυτής.

5.1.1 SegNet RGB

	Εικόνες	Διασταυρούμενες Ακρίβειες εκπαίδευσης	Διασταυρούμενες ακρίβειες δοκιμών
Αρχική εικόνα		-	-
Εικόνα αλήθειας πεδίου		-	-
Βήμα Μάθησης 0.01		78.8503%	81.4204%
Βήμα μάθησης 0.001		80.2754%	82.0589%
Βήμα μάθησης 0.0001		72.3578%	74.7409%

Πίνακας 2: Αποτελέσματα εκπαίδευσης και δοκιμών SegNet σε εικόνες RGB με διάρκεια 5 εποχών


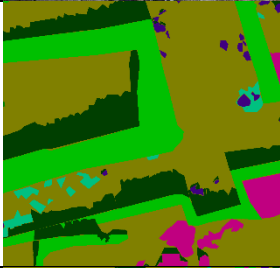
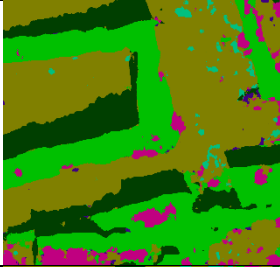
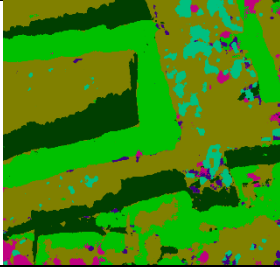


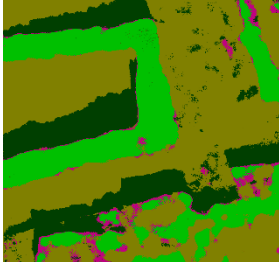
Πίνακας 3: Γραφικές παραστάσεις διασταυρούμενων ακριβειών και απωλειών για αρχιτεκτονική SegNet σε εικόνες RGB με διάρκεια 5 εποχών

Με βάση τα παραπάνω στοιχεία (Πίνακας 2, Πίνακας 3) η επιλογή που γίνεται για το βήμα μάθησης (learning rate) του παραπάνω μοντέλου είναι η τιμή 0.001. Εφόσον τα στοιχεία δείχνουν ότι έχει την πιο καλή ακρίβεια στα τελικά αποτελέσματα εκπαίδευσης και δοκιμών, καθώς οι ακρίβειες έχουν τις υψηλότερες τιμές για κάθε μια από αυτές, και επειδή η γραφική παράσταση είναι η πιο ομοιόμορφη όσον αφορά τις απώλειες και την πορεία των διασταυρούμενων ακριβειών. Στις άλλες δυο γραφικές παραστάσεις παρατηρούνται μεγάλες αποκλίσεις σε διάφορα σημεία κυρίως σε αυτή με βήμα ίσο με 0.01 και πολύ αργή μάθηση σε όλη την διάρκεια εκπαίδευσης 5

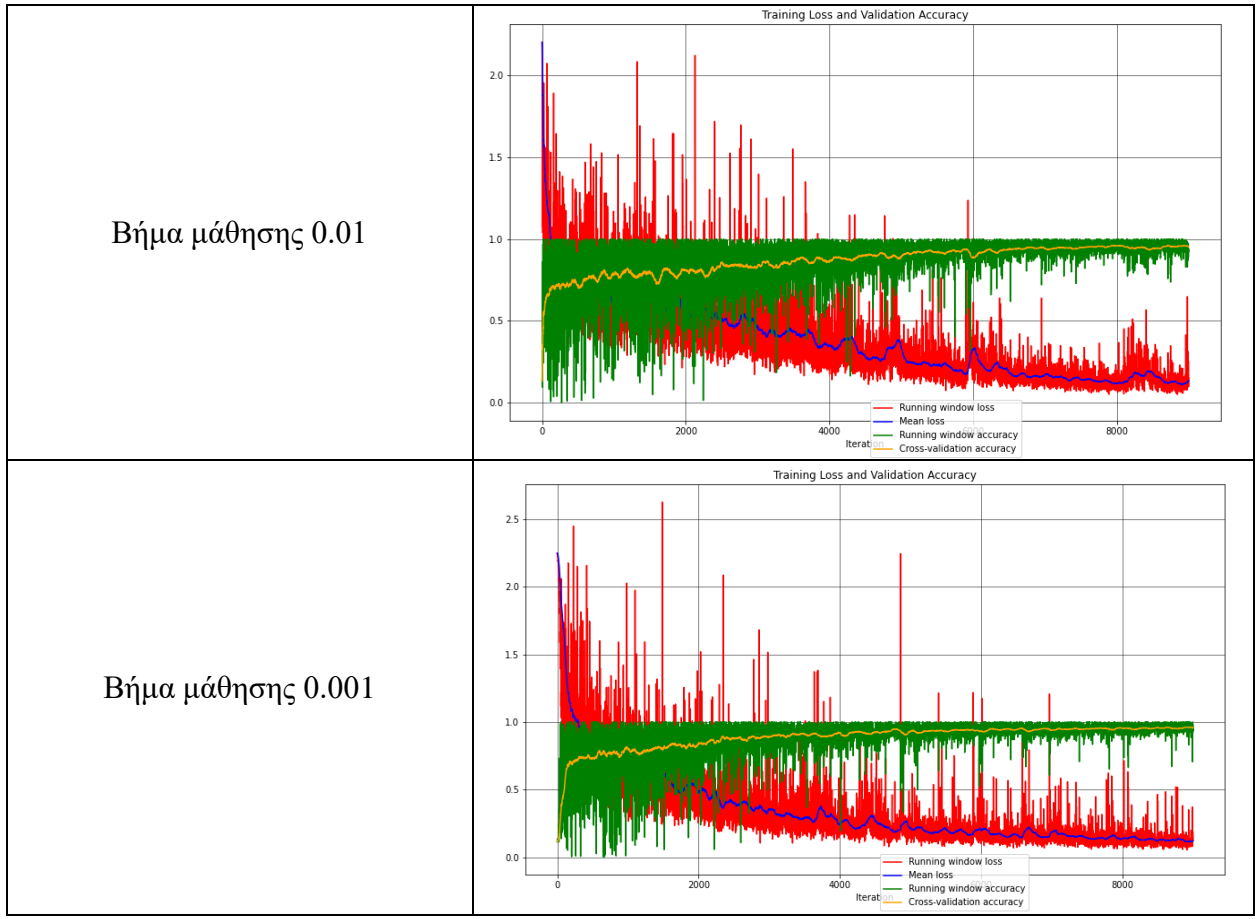
εποχών σε αυτή με βήμα μάθησης ίσο με 0.0001. Η αργή μάθηση μπορεί να συμβεί σε περίπτωση πολύ μικρού βήματος μάθησης, αυτό έχει ως αποτέλεσμα οι ακρίβειες, όσο μικραίνει το βήμα μάθησης, στον ίδιο αριθμό εποχών να μειώνεται η ταχύτητα μάθησης, με αποτέλεσμα τις τιμές που παρατηρούνται (Πίνακας 2). Τέλος παρατηρώντας τις εικόνες πρόβλεψης αυτού του μοντέλου για τα διάφορα βήματα η εικόνα με βήμα μάθησης 0.001 είναι και αυτή που έχει εξάγει το μεγαλύτερο πλήθος των κατηγοριών σε αντίθεση με αυτές των δυο άλλων βημάτων που δεν παρουσιάζουν καθόλου την ύπαρξη λίθων.

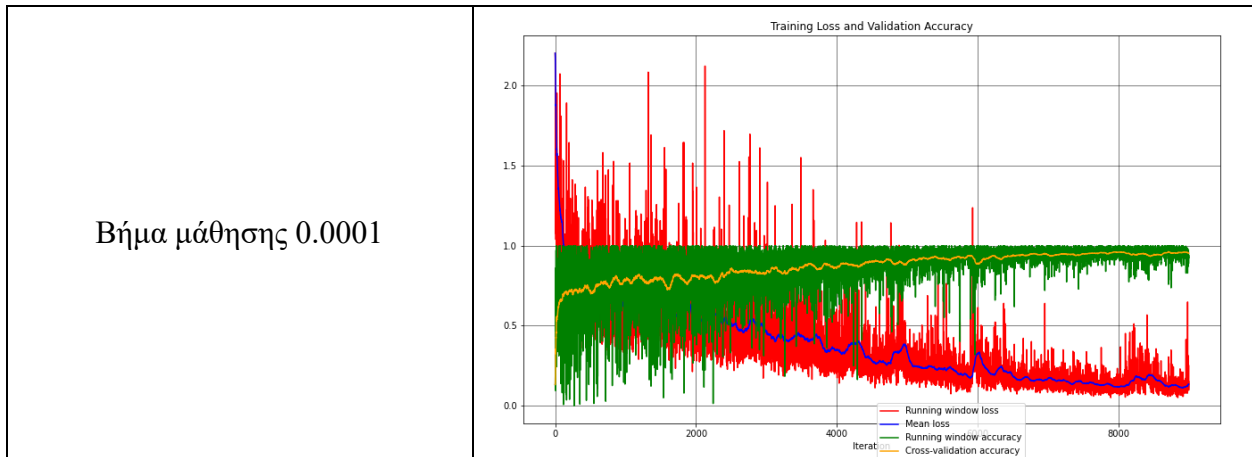
5.1.2 SegNet RGB-D

	Εικόνες	Διασταυρούμενες Ακρίβειες εκπαίδευσης	Διασταυρούμενες Ακρίβειες δοκιμών
Αρχική εικόνα		-	-
Εικόνα αλήθειας πεδίου		-	-
Βήμα Μάθησης 0.01		76.4369%	78.3630%
Βήμα μάθησης 0.001		76.7548%	78.7184%

<p>Βήμα μάθησης 0.0001</p>		<p>79.9021%</p>	<p>82.4639%</p>
--------------------------------	---	-----------------	-----------------

Πίνακας 4: Αποτελέσματα εκπαίδευσης και δοκιμών SegNet σε εικόνες RGB-D με διάρκεια 5 εποχών



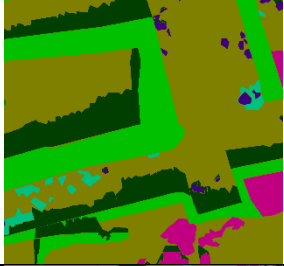
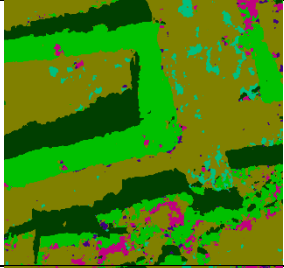
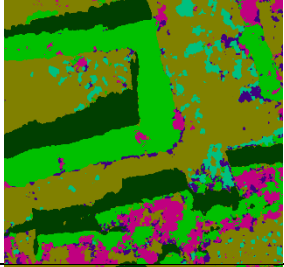
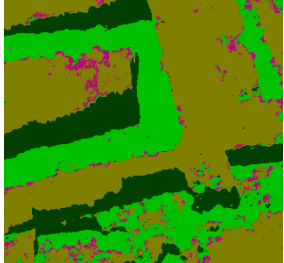


Πίνακας 5: Γραφικές παραστάσεις διασταυρούμενων ακριβειών και απωλειών για αρχιτεκτονική SegNet σε εικόνες RGB-D με διάρκεια 5 εποχών

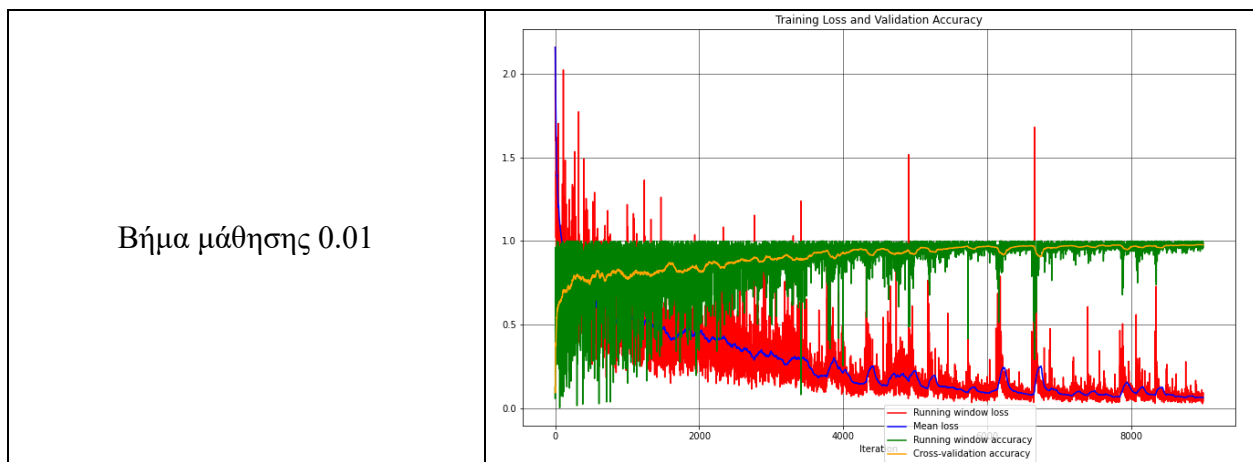
Στην συγκεκριμένη περίπτωση επιλέγεται το βήμα 0.001, παρ' όλο που το καλύτερο αποτέλεσμα ακρίβειας ανήκει στην εκπαίδευση σε βήμα 0.0001, καθώς λόγω της ομοιομορφίας των στοιχείων και των μικρότερων αποκλίσεων στην γραφική παράσταση (Πίνακας 5), σε σχέση με τις άλλες και κυρίως στις μέσες τιμές, πιστεύεται ότι για εκπαίδευση σε περισσότερες εποχές θα ήταν καλύτερο να επιλεγεί η πιο σταθερή περίπτωση. Αυτό φαίνεται και από την διαφορά που παρουσιάζεται στην ακρίβεια μεταξύ εκπαίδευσης και δοκιμής (Πίνακας 4), όπου το μοντέλο με το μικρότερο ποσοστό διαφοράς ακρίβειας είναι αυτό με το βήμα μάθησης 0.001. Επίσης ποιοτικά το βήμα μάθησης 0.001 εμφανίζει όλες της κατηγορίες που υπάρχουν στην εικόνα αλήθειας πεδίου (ground truth), καθώς με την χρήση του βάθους έχουμε περισσότερα στοιχεία και για λίθους και για άλλα χαρακτηριστικά που μπορεί να χρειάζονται κάποια επιπλέον πληροφορία για να καταταμηθούν.

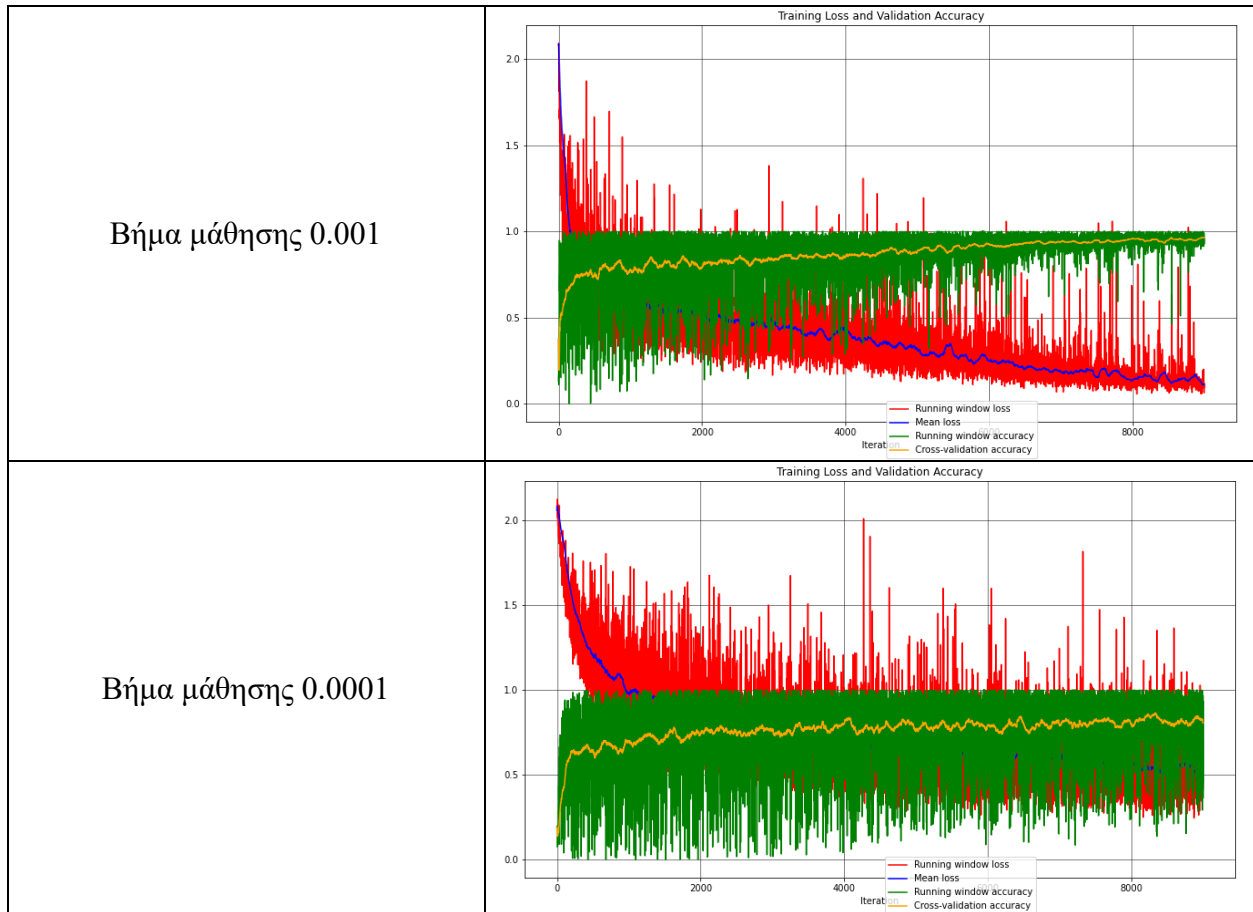
5.1.3 U-Net RGB

	Εικόνες	Διασταυρούμενες Ακρίβειες εκπαίδευσης	Διασταυρούμενες ακρίβειες δοκιμών
Αρχική εικόνα		-	-

Εικόνα αλήθειας πεδίου		-	-
Βήμα Μάθησης 0.01		80.3590%	81.6267%
Βήμα μάθησης 0.001		74.3414%	75.3913%
Βήμα μάθησης 0.0001		80.8388%	81.5057%

Πίνακας 6: Αποτελέσματα εκπαίδευσης και δοκιμών U-Net σε εικόνες RGB με διάρκεια 5 εποχών


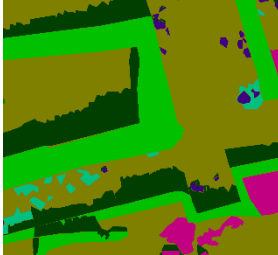
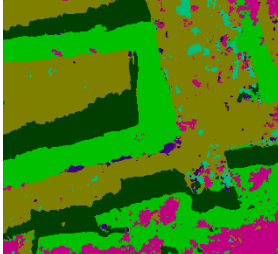
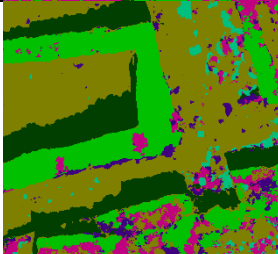
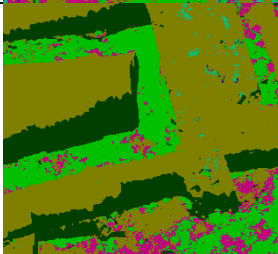




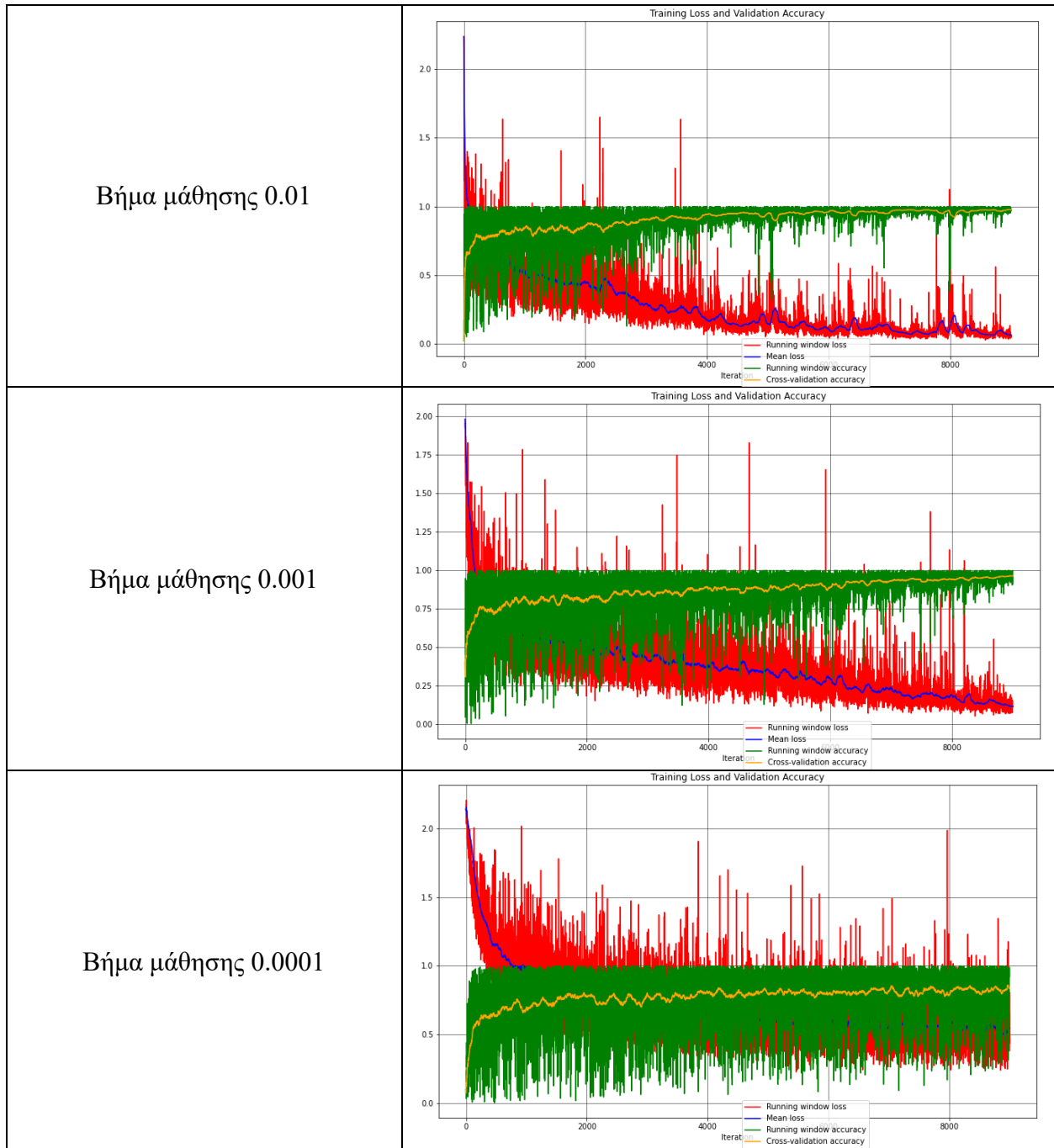
Πίνακας 7: Γραφικές παραστάσεις διασπαυρούμενων ακριβειών και απωλειών για αρχιτεκτονική U-Net σε εικόνες RGB με διάρκεια 5 εποχών

Και εδώ το βήμα που επιλέγεται είναι το 0.001 διότι είναι η πιο σταθερή από τις παραπάνω περιπτώσεις όσον αφορά στις απώλειες και την ακρίβεια καθώς δεν έχει μεγάλες διαφοροποιήσεις σε όλο το εύρος της γραφικής παράστασης (Πίνακας 7). Αν και τα αποτελέσματα αυτής δεν είναι πολύ καλά όσον αφορά στην ακρίβεια κατάτμησης στην διαδικασία εκπαίδευσης και δοκιμών (Πίνακας 6), αυτό διορθώνεται με την εκπαίδευση σε περισσότερες εποχές. Επίσης λόγω της χρήσης δυναμικού βήματος μάθησης στον κώδικα ύστερα από κάποιες εποχές το βήμα αυτό μειώνεται με συνέπεια την πιο σταθερή μάθηση, άρα και την καλύτερη ακρίβεια. Στο κομμάτι της ποιοτικής μελέτης των εικόνων της αρχιτεκτονικής αυτής παρατηρείται το πόσο αργά μαθαίνει ένας κώδικας με πολύ μικρό βήμα μάθησης αν εξεταστεί σε βάθος η εικόνα βήματος 0.0001, καθώς φαίνεται ότι το μοντέλο αδυνατεί να αναγνωρίσει βασικά χαρακτηριστικά όπως οι λίθο και τα ξερόχорта.

5.1.4 U-Net RGB-D

	Εικόνες	Διασταυρούμενες Ακρίβειες εκπαίδευσης	Διασταυρούμενες Ακρίβειες δοκιμών
Αρχική εικόνα		-	-
Εικόνα αλήθειας πεδίου		-	-
Βήμα Μάθησης 0.01		73.0717%	74.0437%
Βήμα μάθησης 0.001		74.9028%	75.6790 %
Βήμα μάθησης 0.0001		75.3108%	75.8109%

Πίνακας 8: Αποτελέσματα εκπαίδευσης και δοκιμών U-Net σε εικόνες RGB-D με διάρκεια 5 εποχών


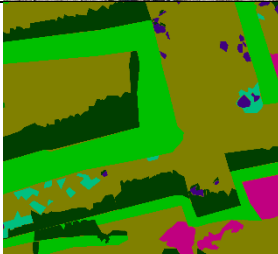
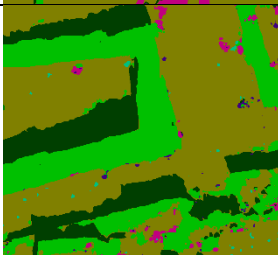
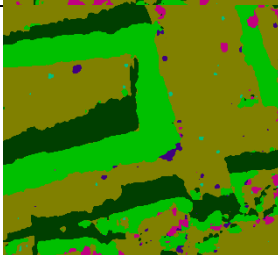


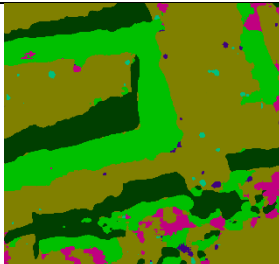
Πίνακας 9: Γραφικές παραστάσεις διασταυρούμενων ακριβειών και απωλειών για αρχιτεκτονική U-Net σε εικόνες RGB-D με διάρκεια 5 εποχών

Εδώ για άλλη μια φορά επιλέγεται το βήμα μάθησης 0.001, καθώς είναι το πιο ομοίμορφο όσον αφορά στην γραφική παράσταση τιμών απώλειας και ακρίβειας (Πίνακας 9). Επίσης το αποτέλεσμα της ακρίβειας στην κατάτμηση της εικόνας δοκιμής του μοντέλου (Πίνακας 8) δεν είναι το καλύτερο αλλά λόγω της πολύ αργής μάθησης που παρατηρείται στην γραφική παράσταση του βήματος 0.0001 δεν μπορεί να γίνει η επιλογή αυτού, παρ' όλο που έχει τα καλύτερα ποσοστά αναγνώρισης. Ποιοτικά επίσης από τις εικόνες (Πίνακας 8) παρουσιάζονται προβλήματα ως προς την αναγνώριση αρκετών κατηγοριών ειδικά στην εικόνα πρόβλεψης

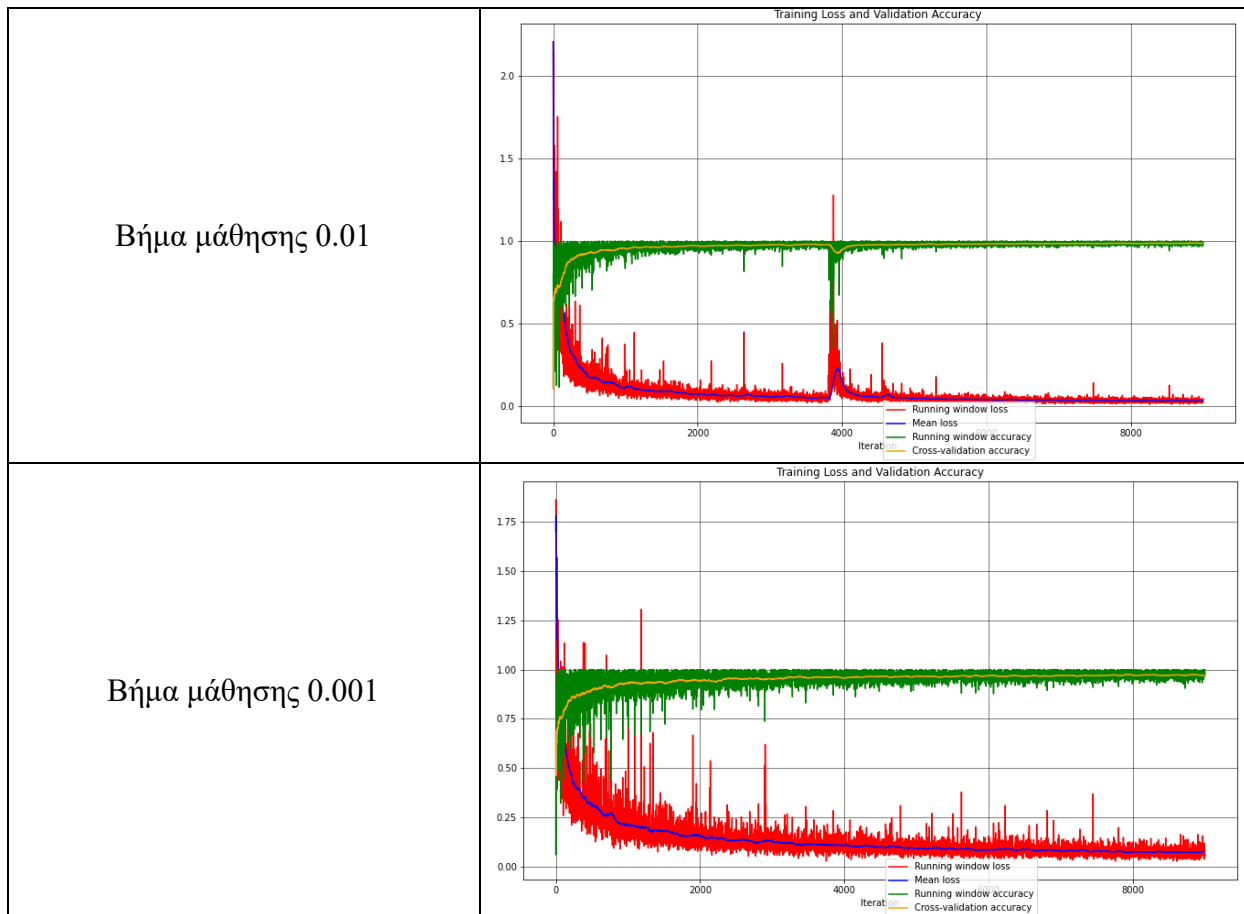
βήματος 0.0001, όπου λόγω της αργής μάθησης του μοντέλου δεν είναι εφικτή η ανάλυση των διαφορών στοιχείων, με αποτέλεσμα την μη ορθή περιγραφή αυτών. Σε αντίθεση με την εικόνα πρόβλεψης με βήμα 0.001 όπου η περιγραφή των στοιχείων είναι πολύ πιο κοντά στην πραγματικότητα παρ' όλο το φαινομενικά μικρότερο ποσοστό ακρίβειας της δοκιμής.

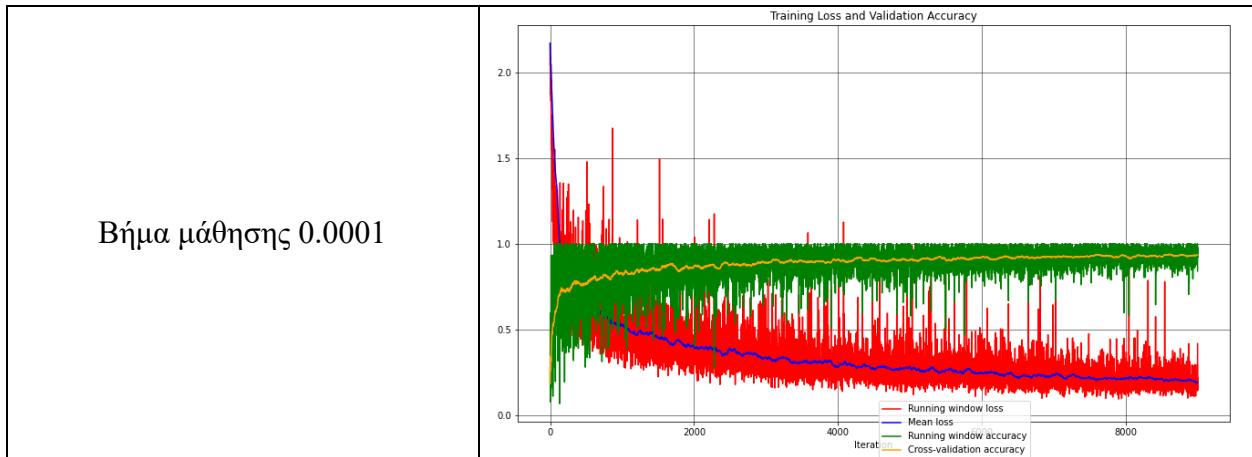
5.1.5 ResNet50_fcn RGB

	Εικόνες	Διασταυρούμενες Ακρίβειες εκπαίδευσης	Διασταυρούμενες ακρίβειες δοκιμών
Αρχική εικόνα		-	-
Εικόνα αλήθειας πεδίου		-	-
Βήμα Μάθησης 0.01		84.7380%	85.3461%
Βήμα μάθησης 0.001		83.7242%	84.8405%

<p>Βήμα μάθησης 0.0001</p>		<p>82.1670%</p>	<p>84.1764%</p>
--------------------------------	---	-----------------	-----------------

Πίνακας 10: Αποτελέσματα εκπαίδευσης και δοκιμών ResNet50_fcn σε εικόνες RGB με διάρκεια 5 εποχών



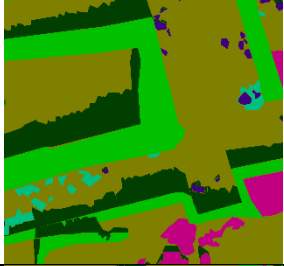
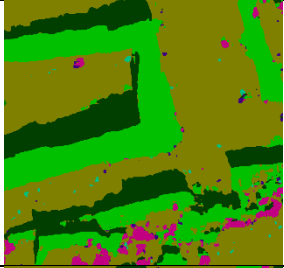
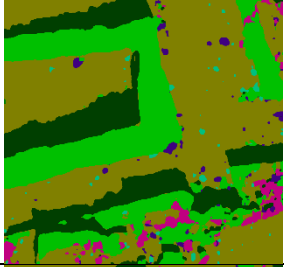
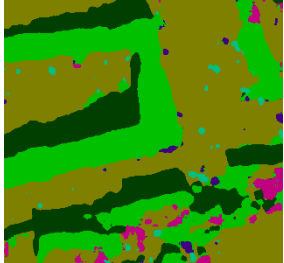


Πίνακας 11: Γραφικές παραστάσεις διασταυρούμενων ακριβειών και απωλειών για αρχιτεκτονική ResNet50_fcn σε εικόνες RGB με διάρκεια 5 εποχών

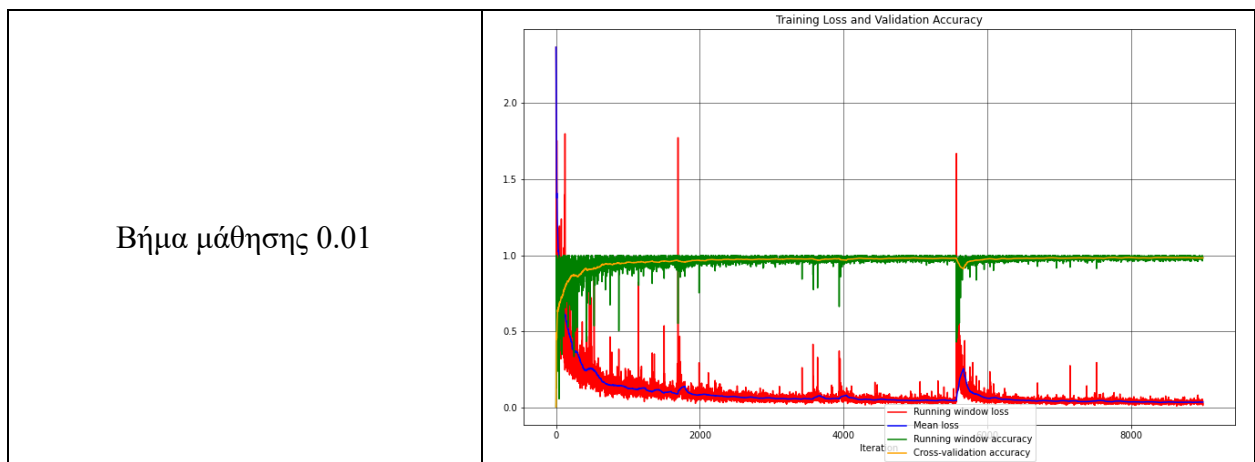
Στο συγκεκριμένο μοντέλο τα αποτελέσματα δεν είναι εντελώς σαφή, καθώς το βήμα μάθησης 0.01 παρουσιάζει μια ιδιομορφία με όλα τα υπόλοιπα σημεία να δείχνουν μια πολύ καλή ομοιομορφία στην γραφική παράσταση (Πίνακας 11) και ένα αποτέλεσμα στην κατάτμηση της εικόνας δοκιμής και εκπαίδευσης που σπάει το φράγμα του 80%, όσον αφορά τις τιμές ακρίβειας (Πίνακας 10), παρ' όλη την μικρή διάρκεια μάθησης. Το μοντέλο με βήμα 0.001 παρουσιάζει μια πολύ ομαλή γραφική παράσταση με ένα πολύ καλό επίσης αποτέλεσμα ακριβειών κατάτμησης, αλλά όχι τόσο καλό όσο στον βήμα 0.01. Συνεπώς αυτά τα δυο επιλέγονται ως βήματα για έλεγχο στις 50 εποχές όπου το αποτέλεσμα θα είναι πιο εμφανές. Η προβληματική που παρατηρείται βέβαια με μια πρώτη ματιά στο συγκεκριμένο μοντέλο είναι η οριακά μη αναγνώριση στοιχείων τα οποία πλησιάζουν υψομετρικά το έδαφος (λίθοι, ξερόχορτα), κάτι που συμβαίνει λόγω της έλλειψης στοιχείων για την ομαλότερη εκπαίδευση του μοντέλου.

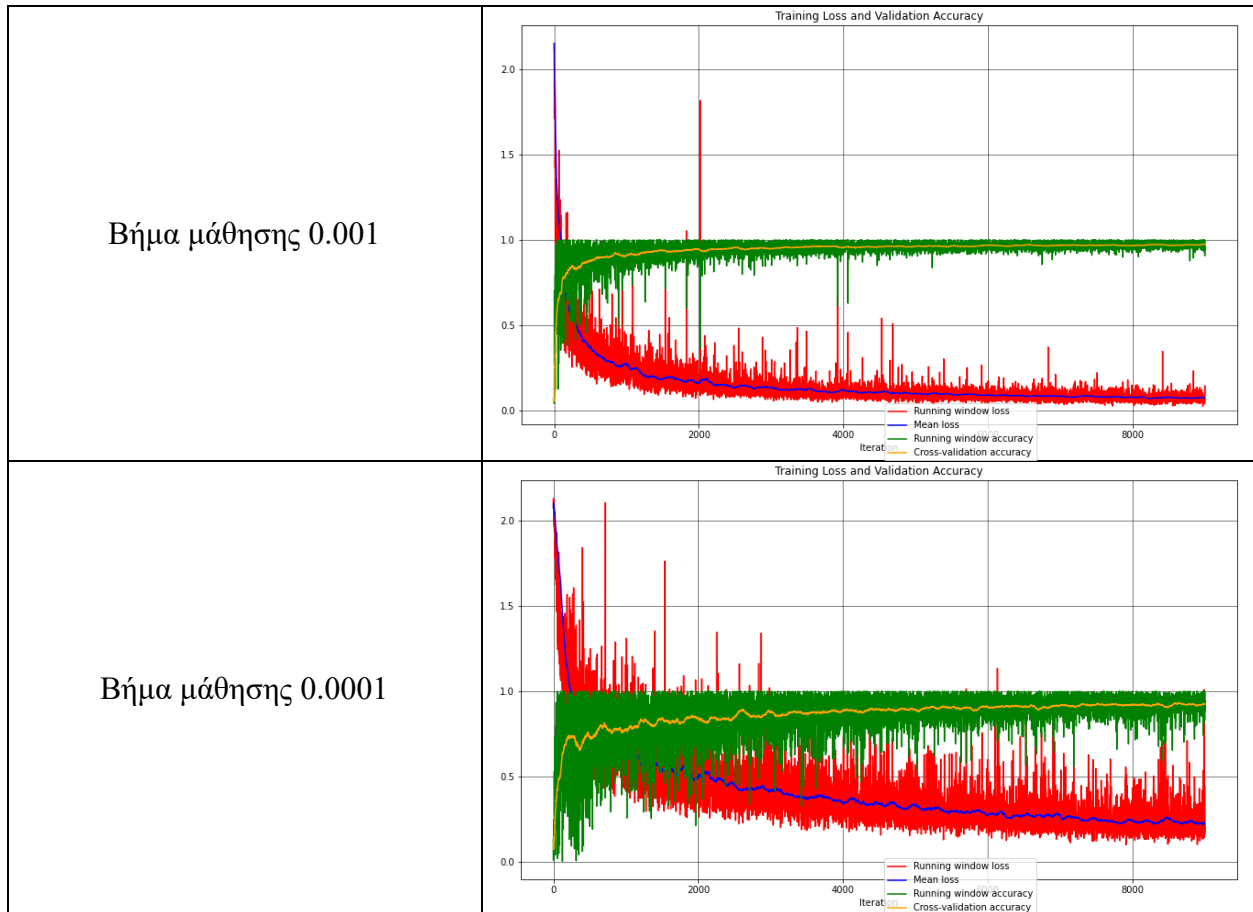
5.1.6 ResNet50_dlv3 RGB

	Εικόνες	Διασταυρούμενες Ακρίβειες εκπαίδευσης	Διασταυρούμενες ακρίβειες δοκιμών
Αρχική εικόνα		-	-

Εικόνα αλήθειας πεδίου		-	-
Βήμα Μάθησης 0.01		84.9575 %	86.2845%
Βήμα μάθησης 0.001		83.4368%	85.1175%
Βήμα μάθησης 0.0001		81.6374%	83.8736%

Πίνακας 12: Αποτελέσματα εκπαίδευσης και δοκιμών ResNet50_dln3 σε εικόνες RGB με διάρκεια 5 εποχών


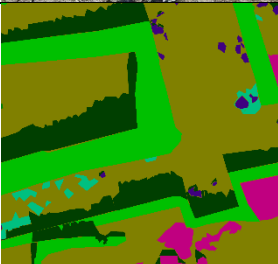
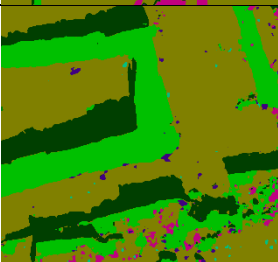
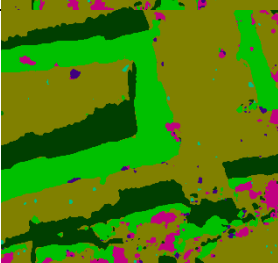
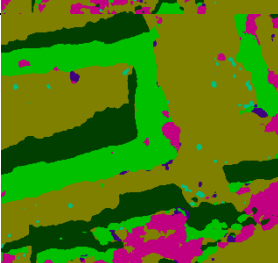




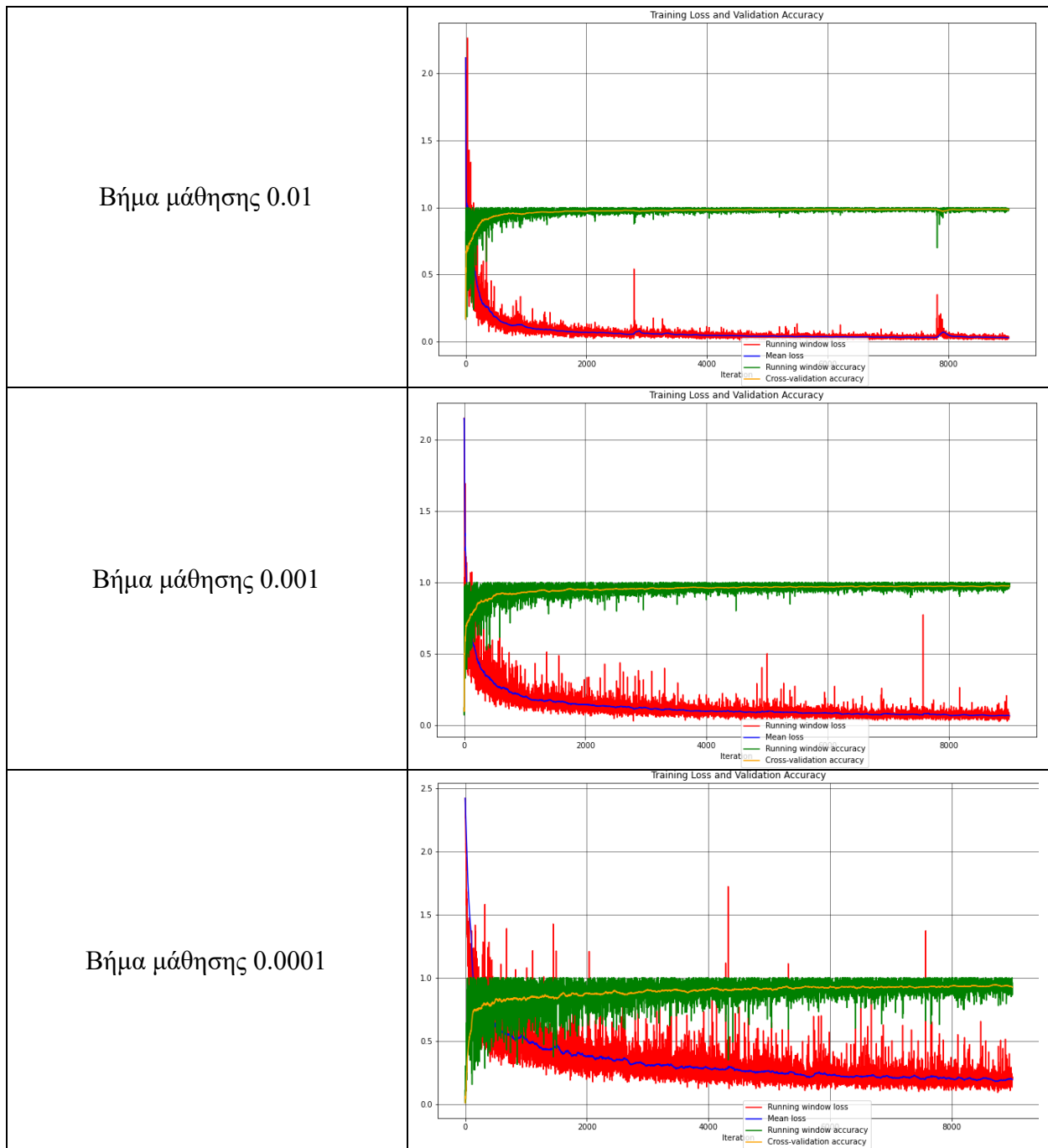
Πίνακας 13: Γραφικές παραστάσεις διασταυρούμενων ακριβειών και απωλειών για αρχιτεκτονική ResNet50_dlv3 σε εικόνες RGB με διάρκεια 5 εποχών

Με βάση τα παραπάνω στοιχεία το βήμα που επιλέγεται για την αρχιτεκτονική αυτή είναι το 0.001, καθώς παρατηρείται ομοιομορφία ως προς την γραφική παράσταση (Πίνακας 13) χωρίς μεγάλες αποκλίσεις οι οποίες εξαλείφονται κατά το πέρας των επαναλήψεων. Επίσης οι τιμές ακριβειών (Πίνακας 12) είναι αρκετά ικανοποιητικές ακόμη και στην μικρή αυτή εκπαίδευση των 5 εποχών, Θα μπορούσε και το βήμα 0.01 να χαρακτηριστεί ως αξιόλογο προς μελέτη αλλά λόγω των διακυμάνσεων των απωλειών θεωρήθηκε ότι δεν πρέπει να συνεχίσει στην εκπαίδευση με 50 εποχές, καθώς δεν μπορεί να προβλεφθεί αν οι 50 εποχές θα διορθώσουν ή θα αυξήσουν αυτές τις ανομοιομορφίες. Τα ποιοτικά στοιχεία του μοντέλου αυτού παρουσιάζουν τις ίδιες ιδιομορφίες με αυτές του προηγούμενου. Αυτό είναι λογικό να συμβαίνει όταν μιλάμε για μοντέλα με την ίδια αρχιτεκτονική ως βάση.

5.1.7 ResNet101_fcn RGB

	Εικόνες	Διασταυρούμενες Ακρίβειες εκπαίδευσης	Διασταυρούμενες ακρίβειες δοκιμών
Αρχική εικόνα		-	-
Εικόνα αλήθειας πεδίου		-	-
Βήμα Μάθησης 0.01		85.4069%	86.1444%
Βήμα μάθησης 0.001		84.0405%	85.1875%
Βήμα μάθησης 0.0001		81.6572%	83.32827%

Πίνακας 14: Αποτελέσματα εκπαίδευσης και δοκιμών ResNet101_fcn σε εικόνες RGB με διάρκεια 5 εποχών


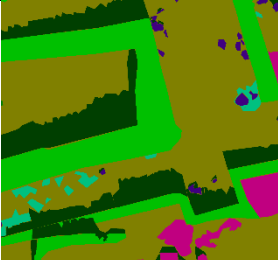
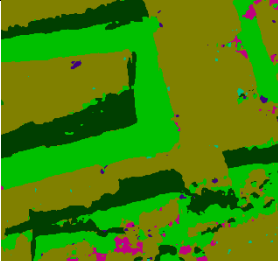
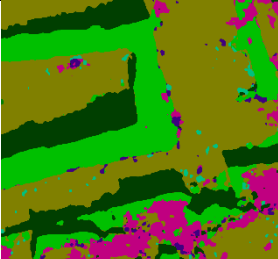


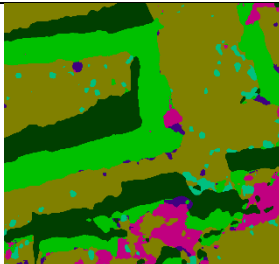
Πίνακας 15: Γραφικές παραστάσεις διασταυρούμενων ακριβειών και απωλειών για αρχιτεκτονική ResNet101_fcn σε εικόνες RGB με διάρκεια 5 εποχών

Γίνεται σαφές ότι σε αυτή την αρχιτεκτονική θα επιλεγούν δύο βήματα μάθησης προς ανάλυση. Το βήμα μάθησης 0.01 και το βήμα μάθησης 0.001. Αυτό συμβαίνει διότι και τα δυο αυτά βήματα φέρουν ομαλότητα ως προς τις τιμές απωλειών (τείνουν ομαλά προς το 0) και ακριβειών (τείνουν ομαλά προς το 1), όπως παρουσιάζεται στις επιμέρους γραφικές παραστάσεις (Πίνακας 15). Τα ποσοστά συνολικών ακριβειών επίσης παρουσιάζουν πολύ καλά αποτελέσματα για τα δύο αυτά βήματα (Πίνακας 14). Άρα για να γίνει εκτενέστερη μελέτη στο πιο βήμα είναι καλύτερο για το μοντέλο θα πρέπει να εξεταστούν και τα δυο με 50 εποχές. Το βήμα 0.0001 παρουσιάζει μεγάλες

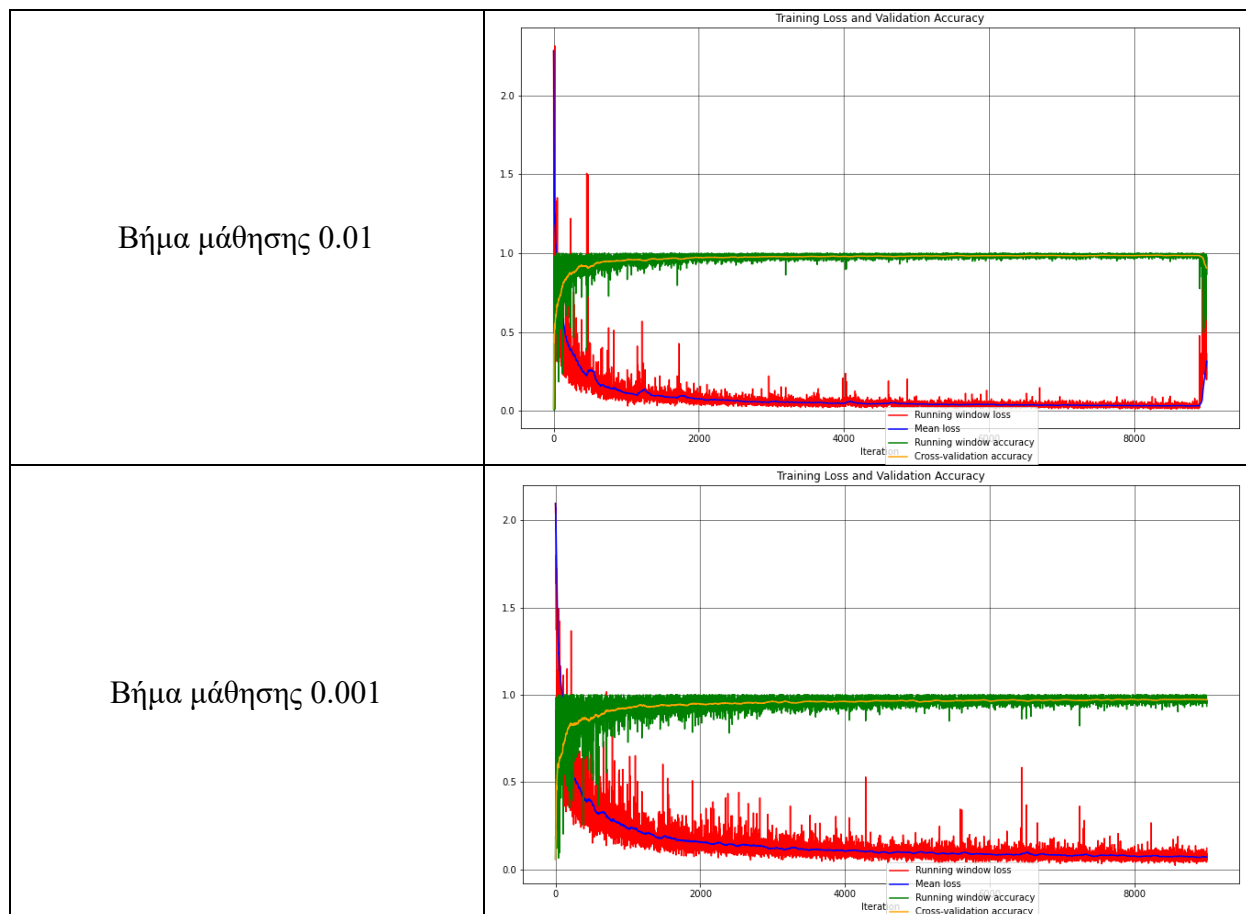
τοπικές διακυμάνσεις και αργή φαινομενικά εκπαίδευση, άρα δεν αξίζει περαιτέρω μελέτης. Όσο για τα ποιοτικά χαρακτηριστικά των εικόνων παρατηρείτε μια έλλειψη σε κατηγορίες μικρης υψομετρικής διαφοράς, αλλά πιστεύεται ότι θα ενισχυθεί η ύπαρξη αυτών με την χρήση περισσότερων εποχών.

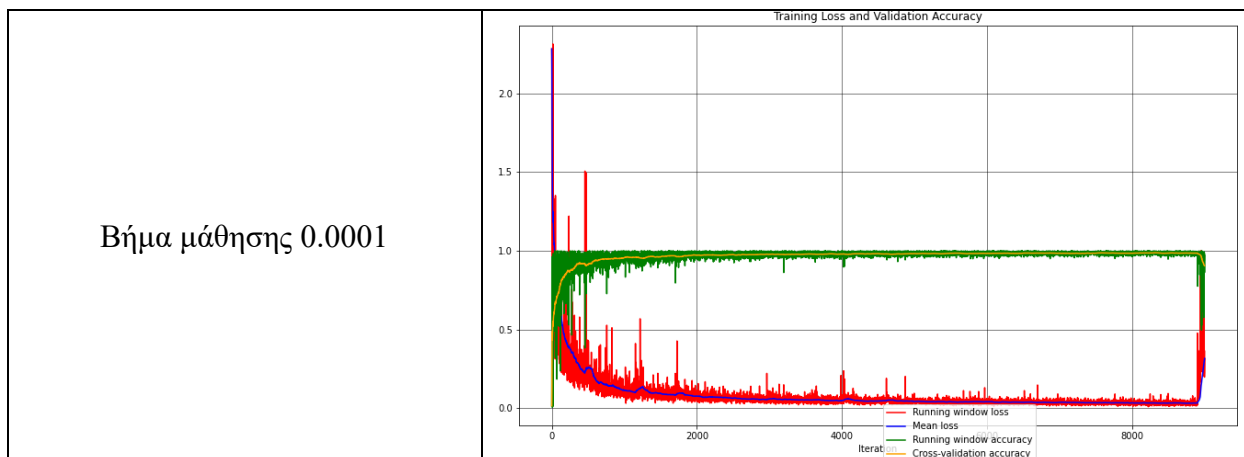
5.1.8 ResNet101_dlv3 RGB

	Εικόνες	Διασταυρούμενες Ακρίβειες εκπαίδευσης	Διασταυρούμενες ακρίβειες δοκιμών
Αρχική εικόνα		-	-
Εικόνα αλήθειας πεδίου		-	-
Βήμα Μάθησης 0.01		84.3468%	85.4623%
Βήμα μάθησης 0.001		80.4901%	82.7485%

<p>Βήμα μάθησης 0.0001</p>		<p>81.0033%</p>	<p>82.9486%</p>
--------------------------------	---	-----------------	-----------------

Πίνακας 16: Αποτελέσματα εκπαίδευσης και δοκιμών ResNet101_d1v3 σε εικόνες RGB με διάρκεια 5 εποχών





Πίνακας 17: Γραφικές παραστάσεις διασταυρούμενων ακριβειών και απωλειών για αρχιτεκτονική ResNet101_dlv3 σε εικόνες RGB με διάρκεια 5 εποχών

Εδώ επιλέγεται ως καλύτερο βήμα μάθησης το βήμα 0.001, καθώς έχει την μεγαλύτερη ομοιομορφία όσον αφορά την γραφική παράσταση των απωλειών και ακριβειών (Πίνακας 17), χωρίς όμως να έχει το καλύτερο αποτέλεσμα στην μέση ακρίβεια απόδοσης της κατάρτισης στην εκπαίδευση και την δοκιμή (Πίνακας 16) κάτι που θα βελτιωθεί με το πέρας μερικών ακόμη εποχών εκπαίδευσης. Οι άλλες δυο τιμές παρουσίασαν μια διακύμανση αρκετά σημαντική, προς το δεξί άκρο, στις επιμέρους γραφικές παραστάσεις τους, οι οποίες τις κρίνουν ακατάλληλες για να συνεχίσουν στο επόμενο στάδιο της μελέτης, καθώς δημιουργούν μεγάλη ανομοιομορφία ως προς τις μέσες τιμές κυρίως. Σε αυτή την περίπτωση ποιοτικά στις εικόνες παρουσιάζονται περισσότερα χαρακτηριστικά όσον αφορά ειδικά τα στοιχεία τα οποία είναι παρόμοια με το έδαφος και χρωματικά και υψομετρικά. Συνεπώς σε αυτό το μοντέλο με περισσότερες επαναλήψεις μπορεί να υπάρξει και μια καλύτερη αναγνώριση του συνόλου των κατηγοριών.

5.2 Στάδιο Δεύτερο

Σε αυτό το στάδιο γίνεται η παρουσίαση των αρχιτεκτονικών με βάση την εκπαίδευση και τα αποτελέσματα που έδωσαν στις δοκιμές με διαφορετικές εικόνες από αυτή της εκπαίδευσης. Τα αποτελέσματα αφορούν κατά κύριο λόγο στην ακρίβεια αναγνώρισης και κατάρτισης των εικόνων που δόθηκαν στο κάθε μοντέλο, την ομαλότητα των γραφικών παραστάσεων που παρουσιάζονται για κάθε μοντέλο και τα ποιοτικά στοιχεία που παρουσιάζονται μεταξύ των εικόνων αλήθειας πεδίου και πρόβλεψης. Η εκπαίδευση του κάθε μοντέλου έγινε ξεχωριστά για κάθε αρχιτεκτονική με διάρκεια 50 εποχών και βήμα μάθησης την τιμή που επιλέχθηκε για κάθε μοντέλο στο πρώτο στάδιο. Ο αριθμός των εποχών επιλέχθηκε κατά κύριο λόγο για ερευνητικούς σκοπούς, καθώς ένα μοντέλο με μικρό όγκο δεδομένων όπως εδώ μπορεί να εκπαιδευτεί και με λιγότερες επαναλήψεις. Τα κυριότερα στοιχεία που εξετάζονται για την επιλογή του καλύτερου μοντέλου είναι[38]:

- **Ανάκληση-Recall:** Παρουσιάζει το ποσοστό των στοιχείων που είναι σχετικά με την ορθή ταξινόμηση και έχουν επιλεγθεί. Δηλαδή από το σύνολο των παρατηρήσεων πόσες από αυτές έχουν προβλεφθεί από τον αλγόριθμο ορθά. Η τιμή της ανάκλησης είναι το πηλίκο

του συνόλου των αληθώς θετικών (true positive) αποτελεσμάτων διαιρούμενο με το άθροισμα του συνόλου των αληθώς θετικών και των λανθασμένα αρνητικών (false negative) αποτελεσμάτων[39].

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

- **Ακρίβεια-Precision:** Είναι το ποσοστό των επιλεγμένων στοιχείων που είναι σχετικά με την ορθή ταξινόμηση. Δηλαδή, από τον αριθμό παρατηρήσεων που είναι θετικοί, πόσα από αυτά έχουν προβλεφθεί από τον αλγόριθμο. Η τιμή της ακρίβειας είναι το πηλίκο του συνόλου των αληθώς θετικών (true positive) αποτελεσμάτων διαιρούμενο με το άθροισμα του συνόλου των αληθώς θετικών και των λανθασμένα θετικών (false positive) αποτελεσμάτων[39].

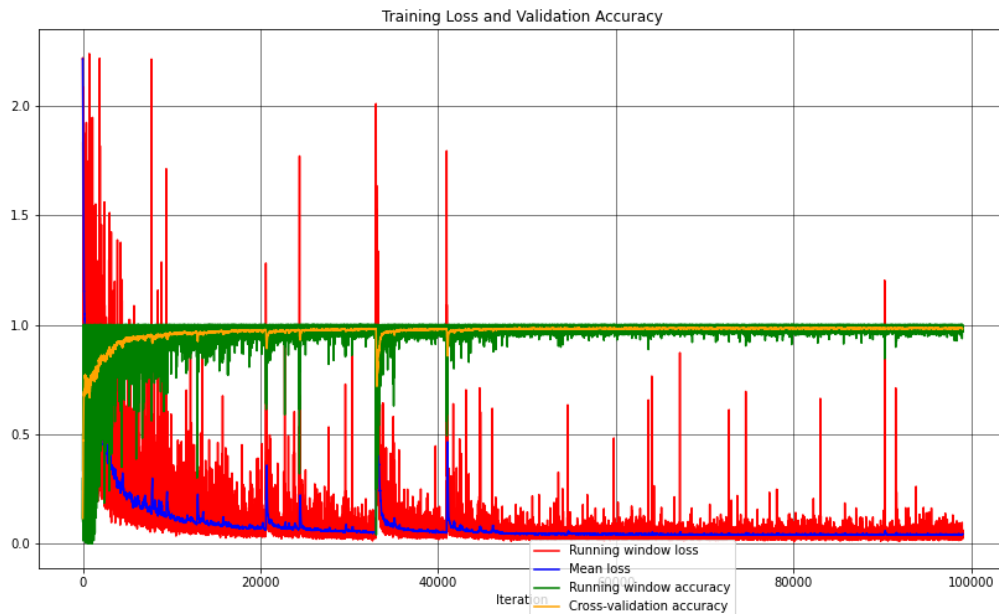
$$Precision = \frac{TP}{TP + FP}$$

- **Συνολική ακρίβεια-Accuracy:** Είναι η πρώτη και πιο διαδεδομένη τιμή που χρησιμοποιείται για την αξιολόγηση της απόδοσης ενός μοντέλου στην διαδικασία της κατάτμησης. Η τιμή της συνολικής ακρίβειας δίνεται από το πηλίκο των ορθά κατηγοριοποιημένων στοιχείων προς το σύνολο των στοιχείων.


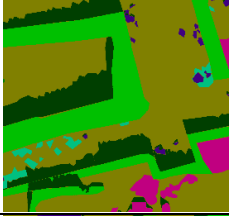
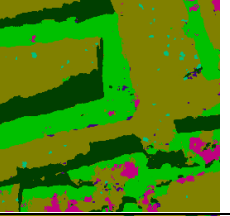

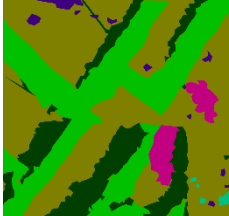
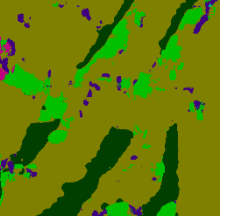


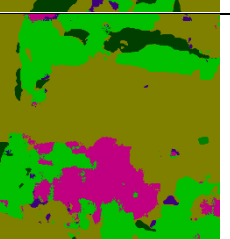
$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Για τον υπολογισμό των παραπάνω στοιχείων χρησιμοποιείται ο πίνακας σύγχυσης (Confusion Matrix), για τον οποίο περισσότερα στοιχεία παρουσιάζονται εδώ[40].

5.2.1 SegNet RGB



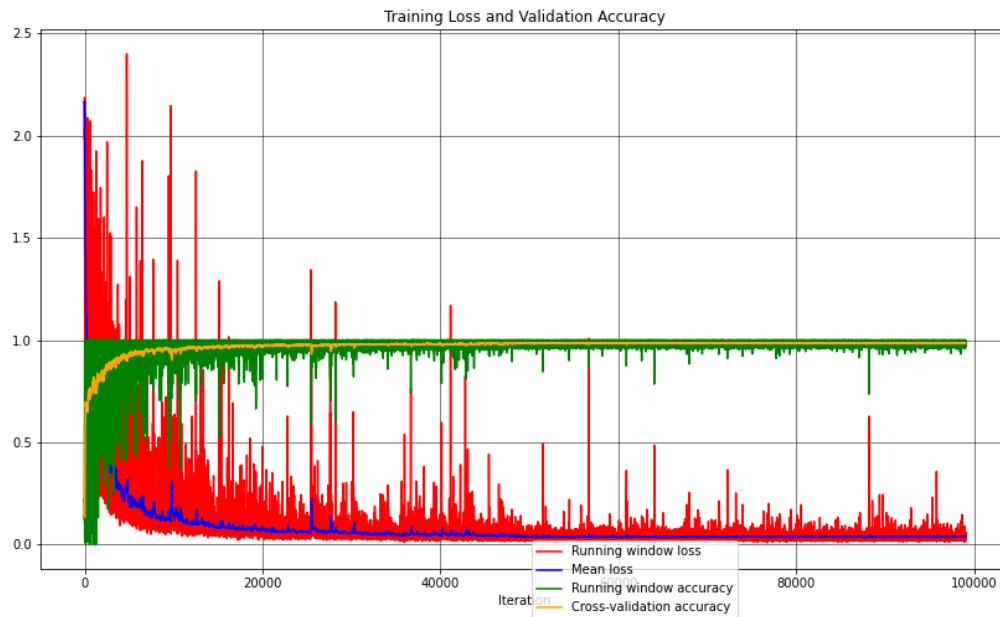
Εικόνα 5.1: Γραφική παράσταση Μέσης απόλειας(Mean loss) και διασταυρούμενης ακρίβειας(Cross-validation accuracy) για το μοντέλο αρχιτεκτονικής SegNet σε εικόνες RGB

	Αρχική Εικόνα	Εικόνα αλήθειας πεδίου	Εικόνα Πρόβλεψης	Ανάκληση-Recall	Ακρίβεια-Precision	Συνολική ακρίβεια-Accuracy
Δοκιμή 1				42.1616%	51.5415%	84.5858%
Δοκιμή 2				27.5382%	30.9805%	67.6598%
Δοκιμή 3				45.1213%	50.6398%	67.2612%

Πίνακας 18: Αποτελέσματα δοκιμών αρχιτεκτονικής SegNet με βήμα 0.001 για εικόνες RGB

Στο γράφημα (Εικόνα 5.1) παρατηρείται μια ανομοιομορφία μέχρι το σημείο των 40000 επαναλήψεων η οποία ύστερα ομαλοποιείται με την συνολική ακρίβεια να τείνει στο 1 και τις απώλειες να τείνουν στο 0. Στις δοκιμές(Πίνακας 18) παρατηρείται μια ανομοιομορφία ως προς τις εικόνες και την τελική κατάτμηση. Αυτό συμβαίνει κυρίως στην δοκιμή 2 όπου δεν θα έπρεπε σε ορισμένα σημεία όπου εμφανίζονται τοίχοι να χαρακτηρίζονται με το χρώμα του εδάφους στην εικόνα πρόβλεψης. Όσον αφορά στα μετρητικά στοιχεία και εκεί παρουσιάζεται μια ανομοιομορφία στην ανάκληση και την ακρίβεια της δοκιμής 2, καθώς τα νούμερα είναι κάτω του μετρίου με αποτέλεσμα να μην επιτρέπει στο μοντέλο της δοκιμής αυτής σωστή κατάτμηση της εικόνας. Στην δοκιμή 3 το αποτέλεσμα που παρουσιάζεται είναι λογικό καθώς στις εικόνες εκπαίδευσης δεν συναντώνται τόσο μεγάλα γκρεμίσματα, με αποτέλεσμα την ύπαρξη μέτριων αποτελεσμάτων στην ανάκληση και στην ακρίβεια, κάτι που συμβαίνει και στα υπόλοιπα μοντέλα. Στην δοκιμή 1, η οποία είναι και η μόνη δοκιμή με κομμάτι από την εικόνα που προήλθε και η εικόνα εκπαίδευσης, έχει τα καλύτερα αποτελέσματα, κάτι που είναι απόλυτα λογικό εφόσον είναι μια εικόνα η οποία παρουσιάζει τα ίδια ακριβώς τεχνικά χαρακτηριστικά με την εικόνα εκπαίδευσης.

5.2.2 SegNet RGB-D



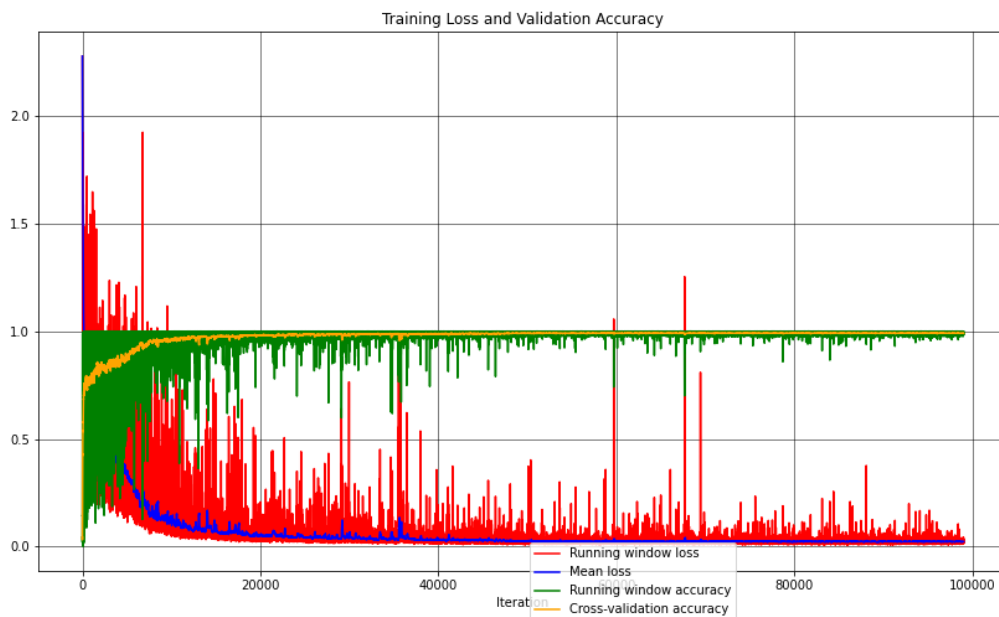
Εικόνα 5.2: Γραφική παράσταση Μέσης απώλειας(Mean loss) και διασταυρούμενης ακρίβειας(Cross-validation accuracy) για το μοντέλο αρχιτεκτονικής SegNet σε εικόνες RGB-D

	Αρχική Εικόνα	Εικόνα αλήθειας πεδίου	Εικόνα Πρόβλεψης	Ανάκληση-Recall	Ακρίβεια-Precision	Συνολική ακρίβεια-Accuracy
Δοκιμή 1				46.5851%	49.0638%	79.2979%
Δοκιμή 2				29.0094%	32.0057%	67.3989%
Δοκιμή 3				43.4351%	51.8353%	67.1016%


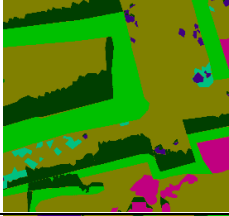
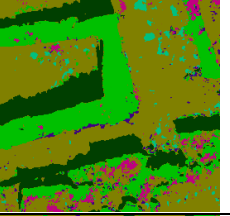

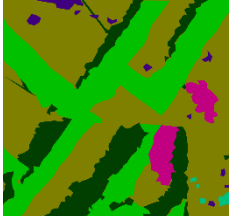
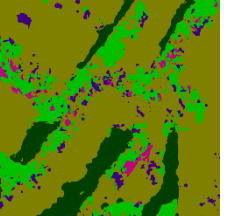


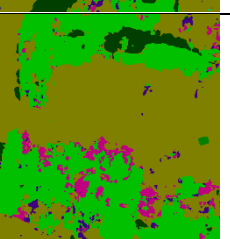
Πίνακας 19: Αποτελέσματα δοκιμών αρχιτεκτονικής SegNet με βήμα 0.001 για εικόνες RGB-D

Στο γράφημα αυτό (Εικόνα 5.2) παρατηρείται ομοιομορφία σε όλο το εύρος αυτής πέρα από κάποιες τοπικές ιδιορρυθμίες ως προς τις απώλειες. Στις δοκιμές (Πίνακας 19) παρατηρείται μια ανομοιομορφία στις τιμές της δοκιμής 2 στις τιμές της ανάκλησης και της ακρίβειας, οι οποίες μεταφέρονται και ποιοτικά στην εικόνα παρουσιάζοντας τοίχους ως γκρεμίσματα και γκρεμίσματα ως έδαφος. Μια παρόμοια συνθήκη επικρατεί και στην δοκιμή 3 στο ποιοτικό αποτέλεσμα όσον αφορά τα γκρεμίσματα και τους τοίχους. Αυτό δημιουργεί λάθη στην τελική κατάτμηση των εικόνων 2 και 3 τα οποία δεν θα ήταν αποδεκτά, δικαιολογώντας μόνο την δοκιμή 3 λόγω χρήσης μικρού φάσματος γκρεμισμάτων για την εκπαίδευση. Η τελική ακρίβεια παρουσιάζει μια αρκετά καλή ομοιομορφία και δικαιολογημένα η δοκιμή 1 έχει μεγαλύτερη ακρίβεια. Επίσης η κατάτμηση των εικόνων ειδικά στην δοκιμή 2 και 3 όσο αναφορά τις δομές και τα γκρεμίσματα παρουσιάζουν ένα ενδιαφέρον, καθώς χάρη στην ένταξη του τέταρτου καναλιού παρατηρείται ότι το μοντέλο προσπαθεί να ακολουθήσει τα δεδομένα που του δόθηκαν και με βάση την διαφοροποίηση του κάθε στοιχείου από το έδαφος δίνει περισσότερη λεπτομέρεια στα γκρεμίσματα, καθώς και στους τοίχους οι οποίοι είναι στην δεύτερη δοκιμή στην εσοχή του μνημείου.

5.2.3 U-Net RGB



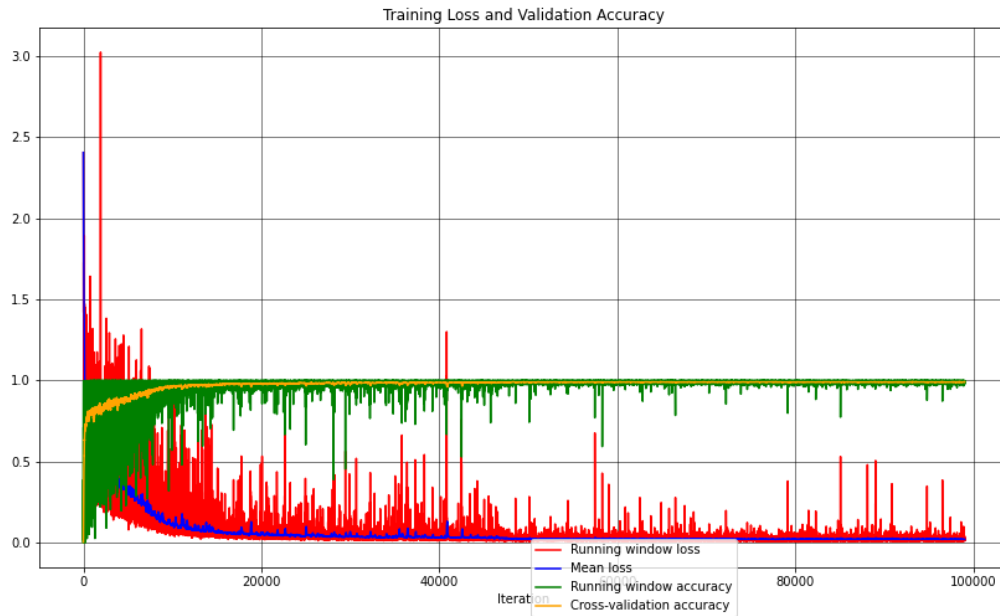
Εικόνα 5.3: Γραφική παράσταση Μέσης απώλειας(Mean loss) και διασταυρούμενης ακρίβειας(Cross-validation accuracy) για το μοντέλο αρχιτεκτονικής U-Net σε εικόνες RGB

	Αρχική Εικόνα	Εικόνα αλήθειας πεδίου	Εικόνα Πρόβλεψης	Ανάκληση-Recall	Ακρίβεια-Precision	Συνολική ακρίβεια-Accuracy
Δοκιμή 1				46.2564%	44.1144%	80.5712%
Δοκιμή 2				31.4439%	41.5346%	70.8192%
Δοκιμή 3				42.9459%	49.4065%	59.7685%

Πίνακας 20: Αποτελέσματα δοκιμών αρχιτεκτονικής U-Net με βήμα 0.001 για εικόνες RGB



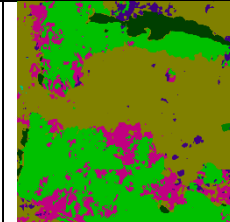
Η γραφική παράσταση (Εικόνα 5.3) φαίνεται να ομαλοποιείται και να τείνει προς το 0 για τις απώλειες και προς το 1 για την τελική ακρίβεια ύστερα από τις 40000 επαναλήψεις. Στον πίνακα δοκιμών (Πίνακας 20) παρατηρείται μια καλύτερη ομοιομορφία όσον αφορά την ανάκληση και την ακρίβεια των δοκιμών σε σχέση με τα προηγούμενα μοντέλα. Παρόλα αυτά συνολικά παρουσιάζει αποτελέσματα στις εικόνες πρόβλεψης όπου δεν υπάρχουν διακριτά στοιχεία για την κατάτμηση κατά κύριο λόγο στις δοκιμές 2 και 3, αλλά και στην 1 σε μικρότερο βαθμό. Επίσης οι τιμές τελικής ακρίβειας θα μπορούσαν να είναι και πιο ικανοποιητικές, ακόμη και μη λαμβάνοντας τόσο υπόψη την δικαιολογημένη χαμηλή συνολική ακρίβεια της δοκιμής 3. Μια γενικότερη παρατήρηση σε αυτό το μοντέλο είναι ότι προσπαθεί να πραγματοποιήσει την κατάτμηση σε μικρότερα χωρία της κάθε εικόνας κάτι που είναι λογικό καθώς οι κύριες χρήσεις αυτού είναι η εύρεση μικρότερων αντικειμένων σε ιατρικές εικόνες κυρίως, με αποτέλεσμα εδώ να δίνει κατηγορία σε μικρότερα αντικείμενα των εικόνων όπως λίθοι και ξερόχορτα. Ειδικά στην δοκιμή 1 η κατάτμηση που έχει πραγματοποιηθεί από τον αλγόριθμο είναι αν όχι καλύτερη της ίδια τάξης με την εικόνα αλήθειας πεδίου (ground truth).

5.2.4 U-Net RGB-D



Εικόνα 5.4: Γραφική παράσταση Μέσης απόλειας (Mean loss) και διασταυρούμενης ακρίβειας (Cross-validation accuracy) για το μοντέλο αρχιτεκτονικής U-Net σε εικόνες RGB-D

	Αρχική Εικόνα	Εικόνα αλήθειας πεδίου	Εικόνα Πρόβλεψης	Ανάκληση-Recall	Ακρίβεια-Precision	Συνολική ακρίβεια-Accuracy
Δοκιμή 1				48.3714%	43.3076%	77.3202%
Δοκιμή 2				30.7358%	31.1126%	61.7942%

Δοκιμή 3				43.1883%	47.5672%	59.6646%
-------------	---	---	---	----------	----------	----------


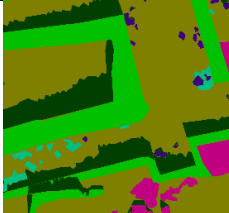
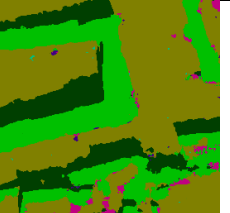

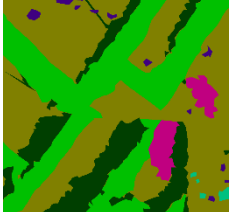
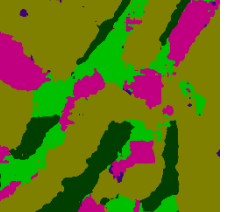


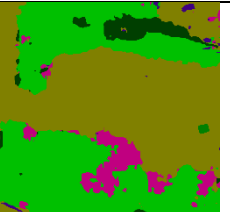
Πίνακας 21: Αποτελέσματα δοκιμών αρχιτεκτονικής U-Net με βήμα 0.001 για εικόνες RGB-D

Όπως και στα προηγούμενα μοντέλα έτσι και εδώ παρουσιάζεται ένα ομαλό γράφημα (Εικόνα 5.4) με τοπικές περισσότερο διακυμάνσεις ως προς τις τιμές απωλειών και ακριβειών. Στον πίνακα δοκιμών (Πίνακας 21) μόνο από τις εικόνες μπορεί να παρατηρήσει κανείς ότι υπάρχουν τεράστια προβλήματα στην κατάτμηση γκρεμισμάτων, λιθοδομών και εδάφους, δηλαδή των βασικών στοιχείων που συναντώνται σε έναν αρχαιολογικό χώρο. Επίσης τα αποτελέσματα της συνολικής ακρίβειας, ανάκλασης και ακρίβειας είναι πολύ χαμηλά στο συγκεκριμένο μοντέλο ακόμη και αν δεν ληφθεί καθόλου υπόψη η δοκιμή 3. Ειδικότερα η δοκιμή 2 παρουσιάζει ένα πολύ κακό αποτέλεσμα στο σύνολο των ποιοτικών και ποσοτικών αποτελεσμάτων, καθώς εμφανίζει λάθος κατάτμηση σε όλο της το εύρος. Η δοκιμή 3 στο συγκεκριμένο μοντέλο παρουσιάζει ένα ενδιαφέρον καθώς διαφαίνεται όπως και στην προηγούμενη δοκιμή με RGB-D εικόνες (SegNet RGB-D) μια προσπάθεια του μοντέλου να πραγματοποιήσει κατάτμηση της εικόνας βασισμένο στο τέταρτο κανάλι ξεχωρίζοντας έτσι με αρκετά ομαλό τρόπο γκρεμίσματα με τοίχους.

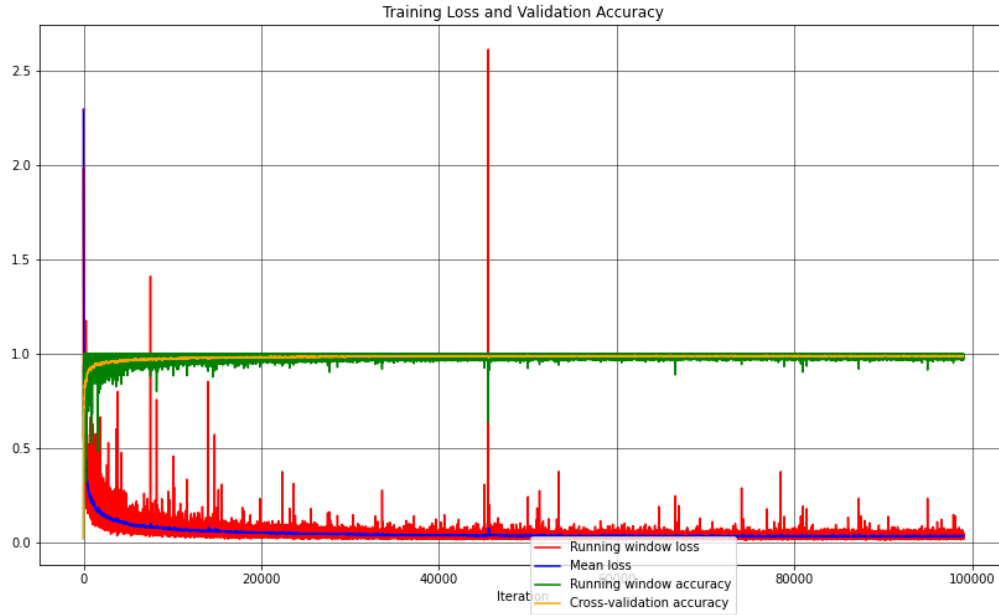
5.2.5 ResNet50_fcn RGB




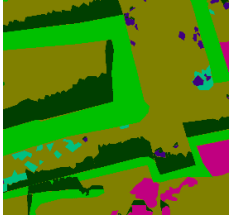
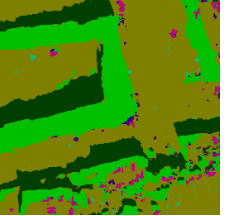

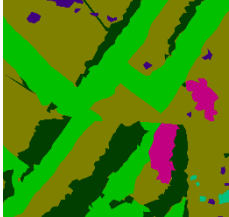
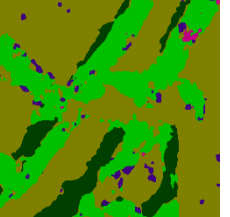


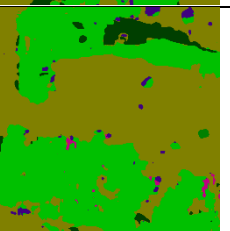
Εικόνα 5.5: Γραφική παράσταση Μέσης απώλειας(Mean loss) και διασταυρούμενης ακρίβειας(Cross-validation accuracy) για το μοντέλο αρχιτεκτονικής ResNet50_fcn με βήμα 0.01

	Αρχική Εικόνα	Εικόνα αλήθειας πεδίου	Εικόνα Πρόβλεψης	Ανάκληση-Recall	Ακρίβεια-Precision	Συνολική ακρίβεια-Accuracy
Δοκιμή 1				48.2695%	54.4931%	86.7169%
Δοκιμή 2				29.9369%	35.4368%	70.2288%
Δοκιμή 3				44.7309%	51.8750%	61.0849%

Πίνακας 22: Αποτελέσματα δοκιμών αρχιτεκτονικής ResNet50_fcn με βήμα 0.01



Εικόνα 5.6: Γραφική παράσταση Μέσης απώλειας(Mean loss) και διασταυρούμενης ακρίβειας(Cross-validation accuracy) για το μοντέλο αρχιτεκτονικής ResNet50_fcn με βήμα 0.001

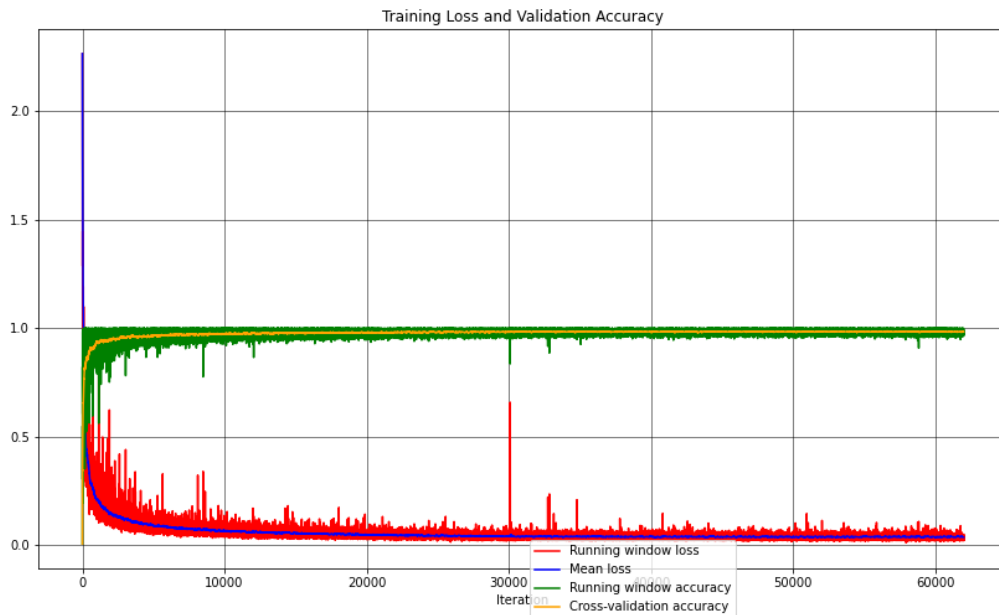
	Αρχική Εικόνα	Εικόνα αλήθειας πεδίου	Εικόνα Πρόβλεψης	Ανάκληση-Recall	Ακρίβεια-Precision	Συνολική ακρίβεια-Accuracy
Δοκιμή 1				48.0185%	49.3952%	84.9890%
Δοκιμή 2				31.1544%	33.8979%	79.2893%
Δοκιμή 3				44.1651%	50.3406%	58.4593%

Πίνακας 23: Αποτελέσματα δοκιμών αρχιτεκτονικής ResNet50_fcn με βήμα 0.001


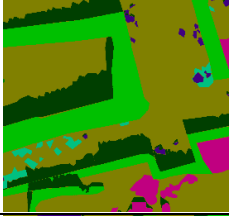
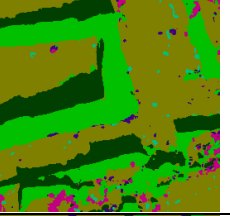

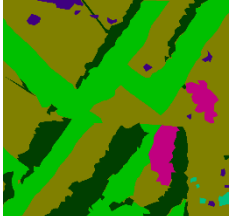
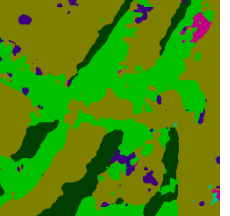


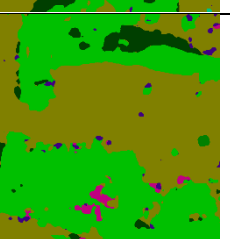
Συγκρίνοντας τα δυο αυτά γραφήματα υπάρχει μια σαφής διαφορά στην ομαλότητα του γραφήματος με βήμα 0.001(Εικόνα 5.6), καθώς παρατηρείται μόνο μία τοπική έκρηξη στις απώλειες σε όλο το εύρος αυτού. Από την άλλη πλευρά το γράφημα με βήμα 0.01 (Εικόνα 5.5)

παρουσιάζει αρκετές ιδιομορφίες ακόμη και στις μέσες τιμές απωλειών και ακρίβειας, οι οποίες ομαλοποιούνται μετά το μέσο περίπου του εύρους του γραφήματος. Ο πίνακας δοκιμών με βήμα 0.01 (Πίνακας 22) παρουσιάζει προβλήματα ως προς την ορθή αναγνώριση λιθοδομών κυρίως στην εικόνα 2 όπου εμφανίζει λιθοδομές ως γκρεμίσματα. Γενικά δεν φαίνεται να παρουσιάζει κάποια άλλο πρόβλημα, πέρα από την μετάφραση του παραπάνω ποιοτικού προβλήματος σε τιμές ανάκλησης και ακρίβειας. Στις υπόλοιπες δοκιμές παρατηρούνται αρκετά καλά αποτελέσματα, χωρίς να ληφθεί υπόψη σε μεγάλο βαθμό την δοκιμή 3 λόγω του μεγάλου γκρεμίσματος όπως προαναφέρθηκε. Στον πίνακα με βήμα 0.001 (Πίνακας 23) παρατηρούνται ομαλότερα αποτελέσματα όσον αφορά την αναγνώριση λιθοδομών, και ειδικότερα στην δοκιμή 2 ποιοτικά και ποσοτικά, χωρίς μεγάλες διαφοροποιήσεις στις άλλες δοκιμές. Και στα δυο αυτά μοντέλα παρατηρείτε το βασικό πρόβλημα της ελλιπούς αναγνώρισης γκρεμισμάτων, αλλά και λόγω των λίγων δεδομένων κατά την εκπαίδευση των μοντέλων δεν γίνεται η αναγνώριση αντικειμένων τα οποία έχουν μικρή διαφοροποίηση από το έδαφος όπως λίθοι και ξερόχορτα. Αυτά τα σφάλματα συμβαίνουν κυρίως στην περίπτωση του βήματος 0.01 και οφείλονται στο στοιχείο της βεβιασμένης εκπαίδευσης (μεγάλο βήμα εκπαίδευσης).

5.2.6 ResNet50_dlv3 RGB



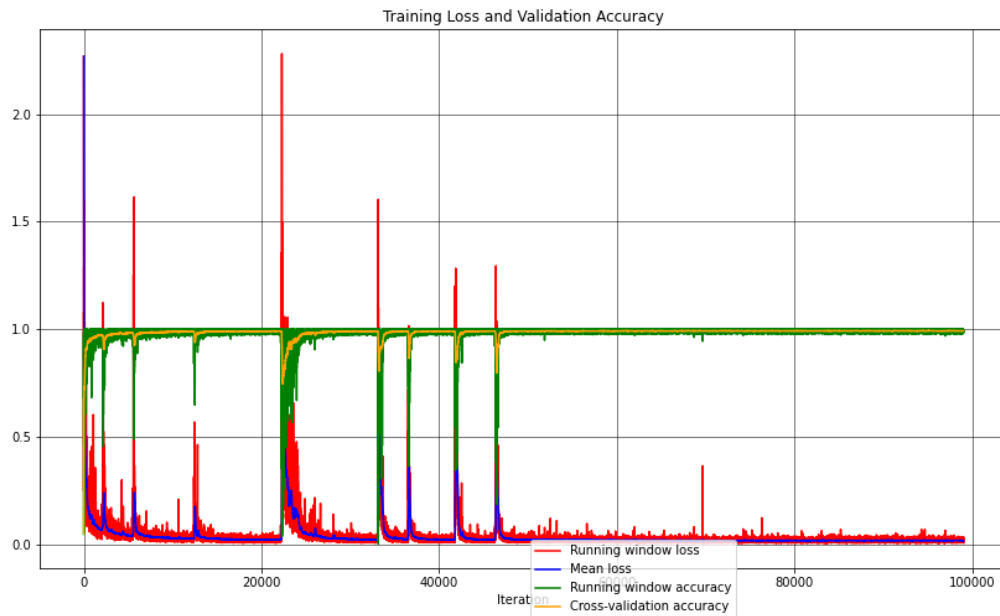
Εικόνα 5.7: Γραφική παράσταση Μέσης απόλειας(Mean loss) και διασταυρούμενης ακρίβειας(Cross-validation accuracy) για το μοντέλο αρχιτεκτονικής ResNet50_dlv3 με βήμα 0.001

	Αρχική Εικόνα	Εικόνα αλήθειας πεδίου	Εικόνα Πρόβλεψης	Ανάκληση-Recall	Ακρίβεια-Precision	Συνολική ακρίβεια-Accuracy
Δοκιμή 1				50.8489%	48.3915%	84.2689%
Δοκιμή 2				36.0671%	41.1594%	77.6019%
Δοκιμή 3				44.9941%	48.2062%	58.9724%

Πίνακας 24: Αποτελέσματα δοκιμών αρχιτεκτονικής ResNet50_dlv3 με βήμα 0.001

Στο μοντέλο αυτό κατά την μελέτη του γραφήματος (Εικόνα 5.7) παρατηρείται μια οριακά απόλυτη ομοιομορφία σε όλο το εύρος αυτού με τις απώλειες να τείνουν στο 0 και τις ακρίβειες να τείνουν στο 1. Στις δοκιμές (Πίνακας 24) επίσης παρουσιάζεται μια σχετική ομαλότητα στα αποτελέσματα των δοκιμών 2 και 3, καθώς τα αποτελέσματά τους στις εικόνες πρόβλεψης παρουσιάζουν μόνο το σύνηθες πρόβλημα των γκρεμισμάτων και των λιθοδομών. Οι τιμές επίσης στην ανάκληση και την ακρίβεια φαίνεται να αντικατοπτρίζουν τα τελικά αποτελέσματα των ποιοτικών χαρακτηριστικών στις δοκιμές 2 και 3 ειδικότερα. Η συνολική ακρίβεια μας δίνει μια καλή συσχέτιση με τις εικόνες αλήθειας πεδίου αν δεν ληφθούν τόσο υπόψη τα σημεία όπου εμφανίζονται περισσότερο γκρεμίσματα. Όπως και στα προηγούμενα μοντέλα της ίδιας αρχιτεκτονικής έτσι και εδώ ποιοτικά παρατηρείτε το πρόβλημα επίσης της μη κατάτμησης μικρών στοιχείων τα οποία δεν είναι εύκολο να αναγνωριστούν χωρίς μεγαλύτερο όγκο δεδομένων.

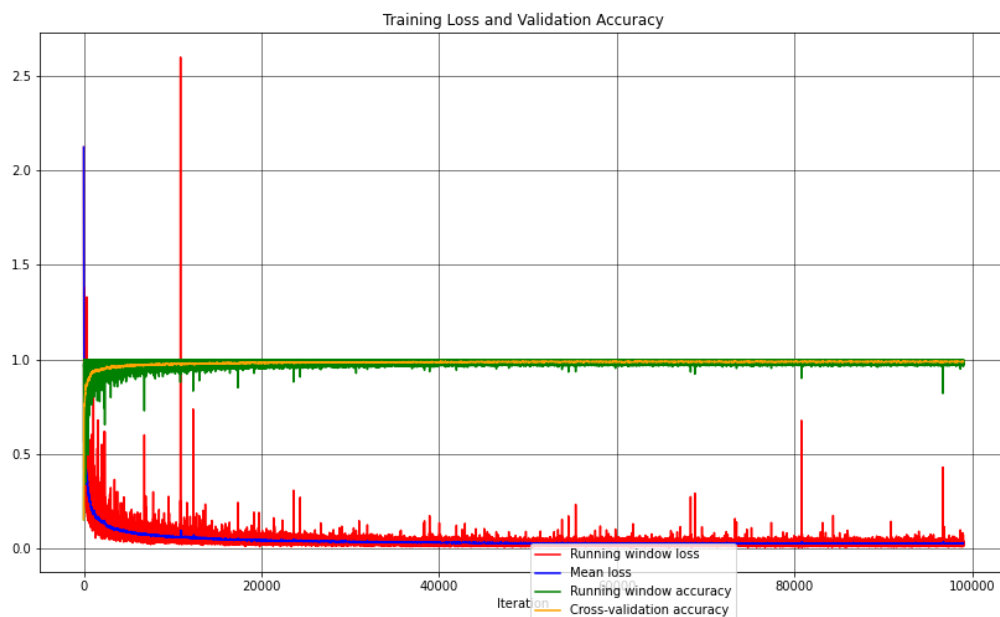
5.2.7 ResNet101_fcn RGB




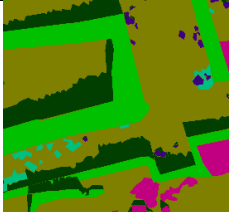
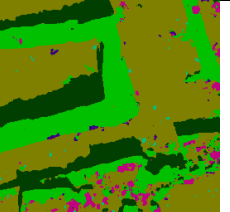

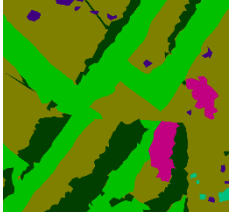
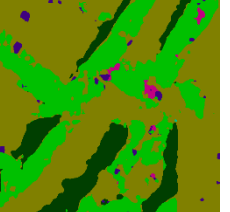


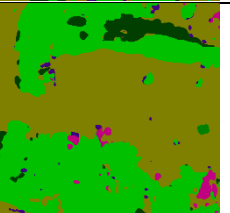
Εικόνα 5.8: Γραφική παράσταση Μέσης απόλειας(Mean loss) και διασταυρούμενης ακρίβειας(Cross-validation accuracy) για το μοντέλο αρχιτεκτονικής ResNet101_fcn με βήμα 0.01

	Αρχική Εικόνα	Εικόνα αλήθειας πεδίου	Εικόνα Πρόβλεψης	Ανάκληση-Recall	Ακρίβεια-Precision	Συνολική ακρίβεια-Accuracy
Δοκιμή 1				47.5398%	51.6928%	85.9088%
Δοκιμή 2				29.4079%	32.5985%	72.3898%
Δοκιμή 3				44.4356%	49.7396%	59.0259%

Πίνακας 25: Αποτελέσματα δοκιμών αρχιτεκτονικής ResNet101_fcn με βήμα 0.01



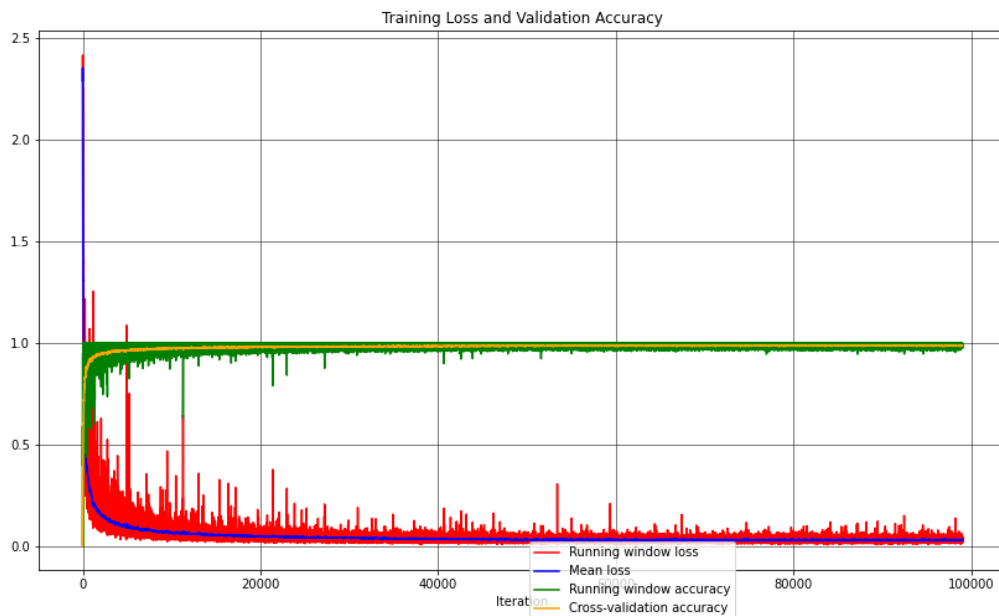
Εικόνα 5.9: Γραφική παράσταση Μέσης απόλειας(Mean loss) και διασταυρούμενης ακρίβειας(Cross-validation accuracy) για το μοντέλο αρχιτεκτονικής ResNet101_fcn με βήμα 0.001

	Αρχική Εικόνα	Εικόνα αλήθειας πεδίου	Εικόνα Πρόβλεψης	Ανάκληση-Recall	Ακρίβεια-Precision	Συνολική ακρίβεια-Accuracy
Δοκιμή 1				49.6265%	51.1594%	85.2574%
Δοκιμή 2				33.7051%	41.7966%	81.2188%
Δοκιμή 3				42.6088%	48.2611%	59.0355%


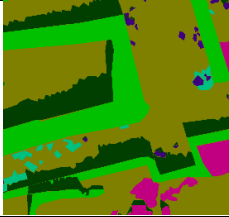
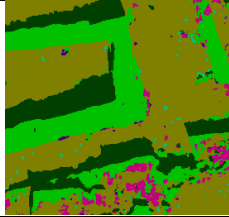

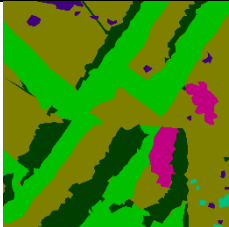
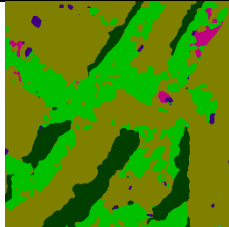

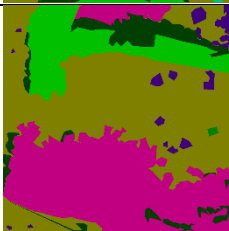
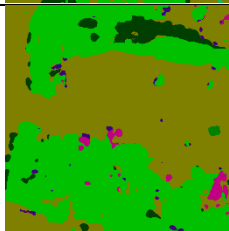
Πίνακας 26: Αποτελέσματα δοκιμών αρχιτεκτονικής ResNet101_fcn με βήμα 0.001

Στην αρχιτεκτονική αυτή και ειδικά στην γραφική παράσταση του βήματος 0.01 (Εικόνα 5.8) υπάρχουν σημαντικές διαφοροποιήσεις μέχρι το μέσο των συνολικών επαναλήψεων οι οποίες ομαλοποιούνται στην συνέχεια. Στη γραφική παράσταση με βήμα 0.001 (Εικόνα 5.9) οι διαφοροποιήσεις που παρατηρούνται είναι τοπικές σε αρχικές επαναλήψεις και μόνο στις απώλειες. Οι διαφοροποιήσεις αυτές επηρεάζουν σε οριακά μηδαμινό βαθμό το τελικό μοντέλο. Οι μέσες ακρίβειες, και στις δυο περιπτώσεις, όσον αφορά τις μέσες τιμές, είναι παρόμοιες με τις απώλειες. Στον πίνακα δοκιμών βήματος 0.01 (Πίνακας 25) και ειδικά στην δοκιμή 2 παρουσιάζεται ποιοτικά στην εικόνα μια σύγχυση στον κώδικα στην κατηγορία λιθοδομές καθώς δεν μπορεί να πραγματοποιήσει κατάτμηση με την σωστή χρωματική κατηγορία. Αυτό μεταφράζεται και ποσοτικά στις τιμές ακρίβειας και ανάκλησης, καθώς είναι πολύ χαμηλές. Στις άλλες δυο δοκιμές δεν παρουσιάζεται κάποιο ιδιαίτερο πρόβλημα καθώς στην δοκιμή 3 συναντάται το κλασικό πρόβλημα των γκρεμισμάτων και στην δοκιμή 1 παρατηρείται ένα αρκετά ικανοποιητικό αποτέλεσμα. Στον πίνακα του μοντέλου με βήμα 0.001 (Πίνακας 26) δεν παρατηρείται ούτε ποιοτικά, ούτε ποσοτικά κάποιο πρόβλημα στην κατάτμηση, πέρα από το πρόβλημα των γκρεμισμάτων. Ειδικά στην εικόνα τα δοκιμής 2 παρατηρείτε μια πολύ καλή κατάτμηση και μια προσπάθεια ορθότερης κατηγοριοποίησης όλων των στοιχείων με αποτέλεσμα να ξεπεραστεί το φράγμα, που είχε τεθεί μέχρι τώρα στη συνολική ακρίβεια της δοκιμής αυτής, του 80%. Τέλος αξίζει παρατήρησης ότι στις εικόνες του βήματος 0.01 η κατάτμηση μικρών αντικειμένων πραγματοποιήθηκε σε ένα μικρό βαθμό σε σχέση με τις εικόνες του βήματος 0.001, όπου στις δοκιμές 2 και 3 κυρίως οι περισσότεροι λίθοι περιγράφονται ως λιθοδομές το οποίο συμβαίνει λόγω της παρόμοιας φύσης των δυο στοιχείων.

5.2.8 ResNet101_dlv3 RGB



Εικόνα 5.10: Γραφική παράσταση Μέσης απώλειας(Mean loss) και διασταυρούμενης ακρίβειας(Cross-validation accuracy) για το μοντέλο αρχιτεκτονικής ResNet101_dlv3 με βήμα 0.001

	Αρχική Εικόνα	Εικόνα αλήθειας πεδίου	Εικόνα Πρόβλεψης	Ανάκληση-Recall	Ακρίβεια-Precision	Συνολική ακρίβεια-Accuracy
Δοκιμή 1				51.9360%	52.1215%	85.6249%
Δοκιμή 2				33.0083%	42.7554%	78.7922%
Δοκιμή 3				44.1406%	47.2742%	59.2545%

Πίνακας 27: Αποτελέσματα δοκιμών αρχιτεκτονικής ResNet101_dlv3 με βήμα 0.001

Το μοντέλο αυτό είναι και το τελευταίο μοντέλο στο οποίο πραγματοποιήθηκαν δοκιμές για την κατάτμηση των εικόνων. Εξετάζοντας την γραφική παράσταση αυτού (Εικόνα 5.10) διαπιστώθηκε πολύ καλός βαθμός ομοιομορφίας και στις τοπικές τιμές ανά επανάληψη, αλλά και στις μέσες τιμές, καθώς ακολουθούσαν το μοτίβο για την πορεία μάθησης. Για άλλη μια φορά βέβαια, όπως μπορεί να διακρίνει κανείς στον πίνακα δοκιμών (Πίνακας 27), υπήρξε η μη ορθή κατηγοριοποίηση των γκρεμισμάτων για του λόγους που προαναφέρθηκαν, καθώς και η μη ορθή κατηγοριοποίηση λιθοδομών στην δοκιμή 2. Παρόλα αυτά οι τιμές που εμφάνισαν οι δοκιμές στην ανάκληση και στην ακρίβεια ήταν αρκετά ικανοποιητικές με την συνολική ακρίβεια να φθίνει στην δοκιμή 2. Όπως και στα προηγούμενα μοντέλα που προήλθα από εκπαίδευση με την αρχιτεκτονική ResNet101 έτσι και εδώ παρουσιάζεται και το πρόβλημα της κατάτμησης λίθων ως λιθοδομές, και της μη κατηγοριοποίησης των μικρότερων αντικειμένων. Αυτό δείχνει την ύπαρξη ανάγκης για περισσότερα δεδομένα κατά την εκπαίδευση.

5.3 Τελική Επιλογή

Λαμβάνοντας υπόψη τα παραπάνω στοιχεία το μοντέλο που επιλέχθηκε ως το πιο κατάλληλο για την κατάτμηση των εικόνων RGB είναι το ResNet101_fcn με βήμα 0.001 και για κατάτμηση εικόνων RGB-D είναι το μοντέλο SegNet. Οι επιλογές αυτές έγιναν καθώς τα συγκεκριμένα μοντέλα παρουσίασαν ως μέσο όρο τα καλύτερα ποσοτικά στοιχεία ανάκλησης, ακρίβειας και συνολικής ακρίβειας. Επίσης ποιοτικά τα μοντέλα αυτά εμφάνισαν την καλύτερη δυνατή κατάτμηση των εικόνων σε σχέση με τα άλλα μοντέλα της κατηγορίας τους, προσπαθώντας να διορθώσουν το πρόβλημα της μη ορθής αναγνώρισης των λιθοδομών, όπου αυτό ήταν εφικτό. Επίσης το μοντέλο ResNet101_fcn με βήμα 0.001 είναι αυτό που κατάφερε να ξεπεράσει το φράγμα ποσοστού ορθής κατάτμησης της εικόνας (συνολικής ακρίβειας) του 80% για την δοκιμή 2. Στην λήψη αυτής της απόφασης δεν λήφθηκε τόσο υπόψιν η δοκιμή 3 καθώς είχε και το μεγαλύτερο πρόβλημα, λόγω του μεγάλου χωρίου γκρεμίσματος που περιλάμβανε, όσο αναφορά τις RGB εικόνες. Μη λαμβάνοντας υπόψη βέβαια το κομμάτι γκρεμίσματος της δοκιμής 3 παρατηρήθηκε άρτια κατάτμηση των υπόλοιπων μερών τις εικόνας. Σε αντίθεση με τις δοκιμές των RGB-D εικόνων, όπου το μοντέλο SegNet κατάφερε χάρη στην ύπαρξη της παραπάνω πληροφορίας του καναλιού βάθους να κατατμήσει με ομοιόμορφο τρόπο όλες τις εικόνες που του δόθηκαν ακόμη και των μικρότερων στοιχείων αυτής, παρουσιάζοντας βέβαια ορισμένα προβλήματα στην κατηγοριοποίηση των γκρεμισμάτων και των λιθοδομών.

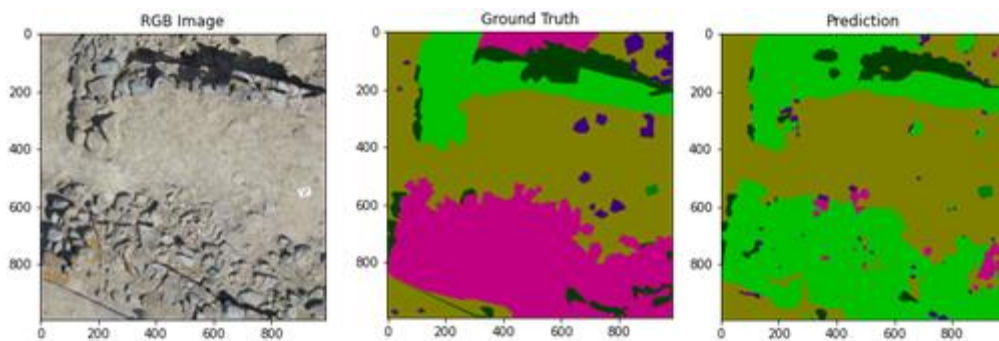
6. Συμπεράσματα

Το κεφάλαιο αυτό είναι μια αναφορά στην αξιολόγηση του τελικού αποτελέσματος, τα προβλήματα που αντιμετωπίστηκαν κατά την εκπόνηση της εργασίας και τέλος τις μελλοντικές χρήσεις που θα μπορούσε να έχει η έρευνα αυτή.

6.1 Αξιολόγηση


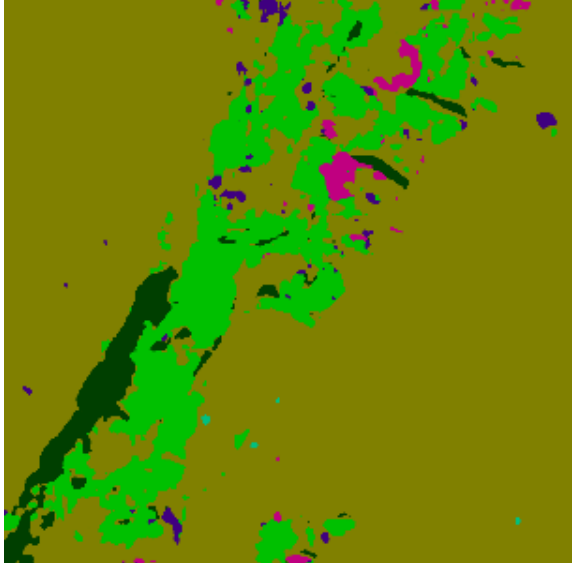

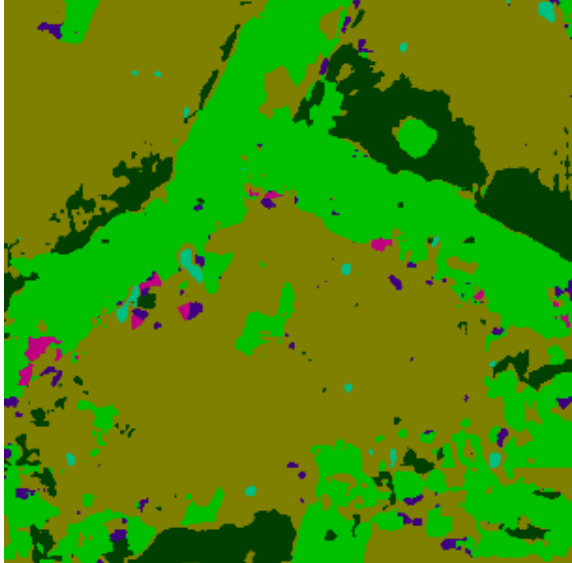
6.1.1 Αξιολόγηση Σημασιολογική Κατάτμησης σε RGB εικόνες

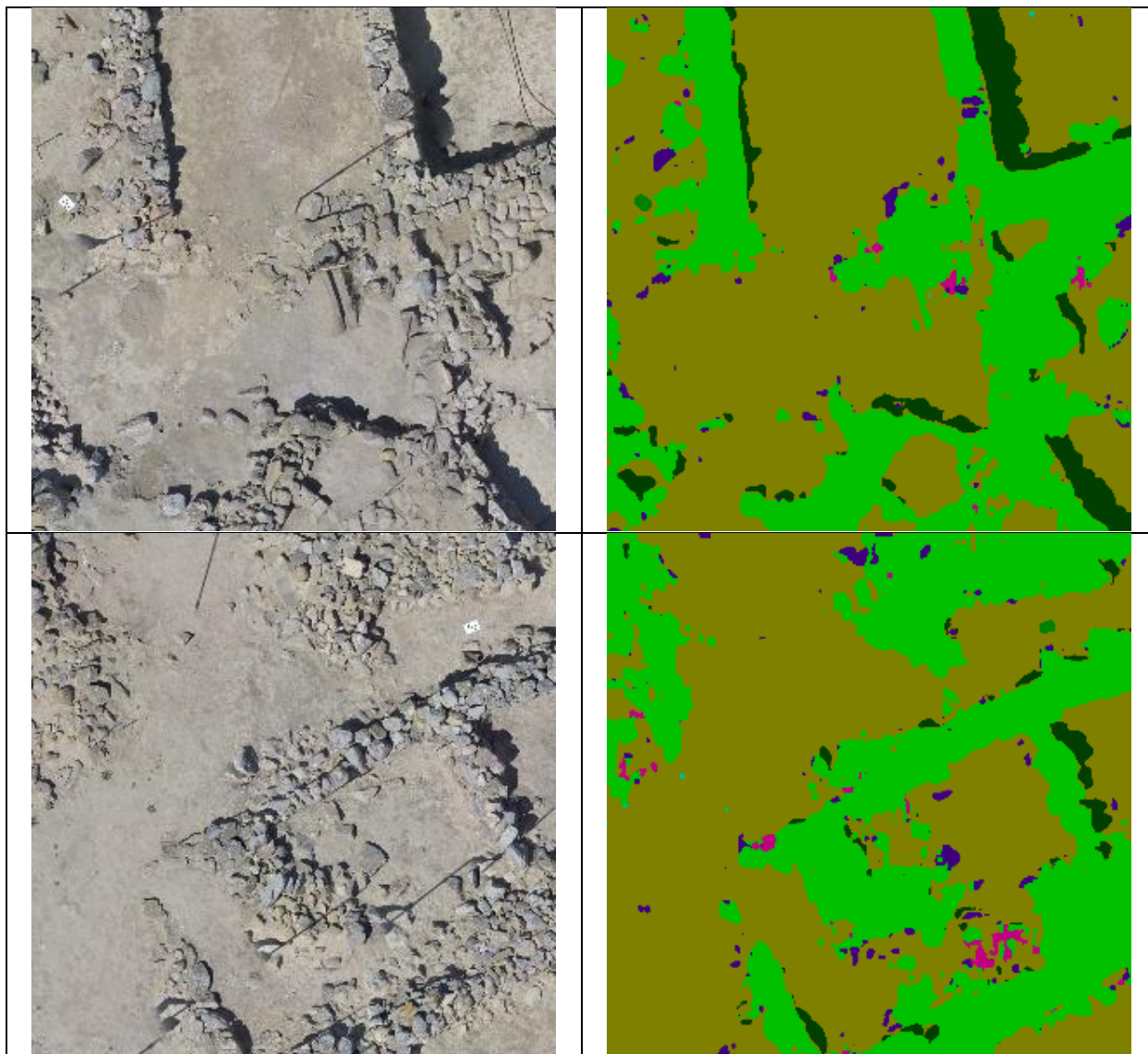
Το τελικό μοντέλο που επιλέχθηκε για τις RGB εικόνες είναι το ResNet101_fcn με βήμα μάθησης 0.001, διότι έδωσε τα καλύτερα αποτελέσματα συγκριτικά με τα υπόλοιπα μοντέλα που δοκιμάστηκαν στην συνθήκη των 50 εποχών. Ως μοντέλο στις εικόνες που του δόθηκαν για δοκιμή και επεξεργασία παρουσίασε μια ιδιομορφία ως προς την κατηγορία των γκρεμισμάτων, κάτι που αποδείχθηκε λογικό στο σύνολό του καθώς η εικόνα που χρησιμοποιήθηκε για την αρχική εκπαίδευση όλων των μοντέλων στην συνθήκη των 50 εποχών δεν περιλάμβανε τόσο έντονα το στοιχείο αυτό (Εικόνα 6.1). Επίσης στην ίδια εικόνα παρατηρείται ότι η κατηγορία των λίθων αναπαρίσταται με το χρώμα κατηγοριοποίησης της λιθοδομής κάτι που συμβαίνει καθώς οι λιθοδομές κυριαρχούν στις συγκεκριμένες εικόνες και οι λίθοι στο έδαφος έχουν πολύ κοντινά χαρακτηριστικά και ραδιομετρία με αυτούς που απαρτίζουν την λιθοδομή.



Εικόνα 6.1: Εικόνα προβολής διαφορών όσον αφορά την κατηγορίας γκρεμισμάτων

Λόγω των προβλημάτων αυτών το μοντέλο θα μπορούσε να εξελιχθεί με την παροχή ενός μεγαλύτερου όγκου δεδομένων κατά την διαδικασία της εκπαίδευσης έτσι ώστε να εξαλειφθούν τέτοιου είδους προβλήματα, με την ενδιάμεση κατασκευή, κατά την εκπαίδευσης του μοντέλου, περισσότερων χαρτών χαρακτηριστικών (feature maps) με τις διαδικασίες της συνέλιξης. Βέβαια το μοντέλο αυτό για τα βασικά χαρακτηριστικά που υπάρχουν σε έναν αρχαιολογικό χώρο δουλεύει ικανοποιητικά, δημιουργώντας μερικά θέματα κατά την αναγνώριση γκρεμισμάτων.

Αρχική εικόνα	Εικόνα πρόβλεψης
	
	

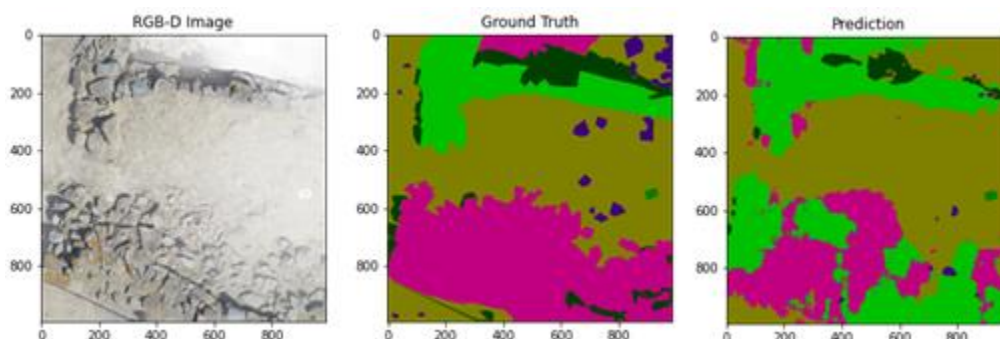


Πίνακας 28: Εικόνες δοκιμών χωρίς αληθινή βάση για αρχιτεκτονική ResNet101_fcn με βήμα 0.001

Για την καλύτερη αξιολόγηση του μοντέλου έγιναν δοκιμές και σε εικόνες, για τις οποίες δεν σχεδιάστηκε υπόβαθρο αλήθειας πεδίου, για την εξαγωγή περισσότερων ποιοτικών στοιχείων για το μοντέλο. Όπως παρατηρείται στον πίνακα αυτών των δοκιμών (Πίνακας 28) παρουσιάζεται μια αξιόλογη προσπάθεια κατάτμησης των εικόνων, καθώς τα αποτελέσματα είναι αρκετά ικανοποιητικά με μερικά προβλήματα στην ορθή περιγραφή των λιθοδομών, και ειδικότερα στην πρώτη εικόνα όπου δεν έχει πραγματοποιηθεί πλήρης εκσκαφή της περιοχής. Στην δεύτερη εικόνα παρατηρούνται τα προβλήματα της αναγνώρισης μικρότερων αντικειμένων και κυρίως λίθων ως λιθοδομές. Αυτό συμβαίνει λόγω του μικρού όγκου δεδομένων και ως ένα βαθμό στην έννοια της υπερπροσαρμογής (overfitting). Η υπερπροσαρμογή αναφέρετε σε ένα πρόβλημα των μοντέλων που έχουν παραχθεί από αρχιτεκτονικές νευρωνικών δικτύων όπου όταν έχουν πολλά στοιχεία για ένα αντικείμενο μπορεί να περιγράψουν άλλα με παρόμοια χαρακτηριστικά ως αυτό. Χαρακτηριστικό παράδειγμα εδώ είναι και οι λίθοι με τις λιθοδομές


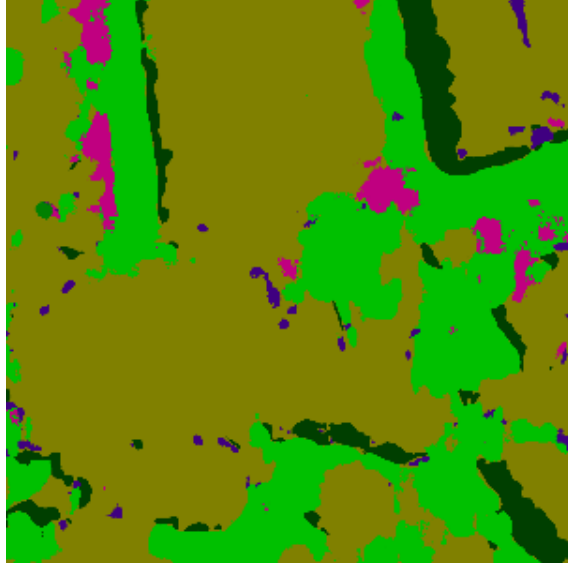

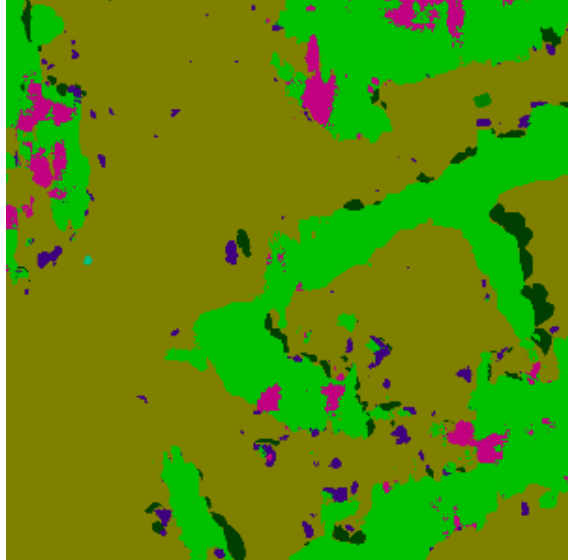
6.1.2 Αξιολόγηση Σημασιολογικής Κατάτμησης σε RGB-D εικόνες

Το τελικό μοντέλο που επιλέχθηκε για τις RGB-D εικόνες είναι το SegNet, διότι έδωσε όσον αφορά την κατάτμηση των εικόνων τα καλύτερα αποτελέσματα μεταξύ αυτού και του U-Net στην συνθήκη των 50 εποχών. Επίσης στο μοντέλο αυτό χάρη στην επιπλέον πληροφορία του βάθους παρατηρήθηκαν κάποια αξιοσημείωτα χαρακτηριστικά στην κατάτμηση των εικόνων, καθώς και ορισμένα προβλήματα τα οποία όμως μπορούν να ξεπεραστούν με εισαγωγή περισσότερων δεδομένων για την εκπαίδευση του μοντέλου. Το αξιοσημείωτο σε αυτό το μοντέλο είναι η μεγάλη επιτυχία στην κατάτμηση γκρεμισμάτων, καθώς τα δεδομένα που χρησιμοποιήθηκαν και ιδίως για γκρεμίσματα ήταν λιγοστά. Αυτό πιστεύεται ότι συνέβη χάρη στην ένταξη του βάθους, το οποίο έκανε δυνατή την διαφοροποίηση λιθοδομών από γκρεμίσματα χάρη στην υψομετρική διαφοροποίηση αυτών κατά την επεξεργασία των εικόνων κατά την εκπαίδευση του μοντέλου. Επίσης η ύπαρξη καναλιού βάθους φαίνεται να βοήθησε και στην ορθότερη κατάτμηση των μικρότερων αντικειμένων στην εικόνα, όπου τα μικρότερα στοιχεία όπως οι λίθοι και τα ξερόχορτα δεν εμφανίζονταν καθόλου ή εμφανίζονταν με χρωματική ετικέτα λιθοδομής, κυρίως όσον αφορά τους λίθους. Επίσης τα όρια στις εικόνες δοκιμών (Πίνακας 19) φαίνεται να έχουν πολύ καλύτερα όρια σε σχέση με τις δοκιμές των υπόλοιπων μοντέλων και πιθανολογείται ότι και εκεί η εισαγωγή του βάθους έπαιξε σημαντικό ρόλο. Όπως διακρίνεται και στην Εικόνα 6.2 τα παραπάνω συμπεράσματα παίρνουν μορφή και γίνονται πιο ξεκάθαρα.



Εικόνα 6.2: Εικόνα σύγκρισης πρόβλεψης (Prediction) και αλήθειας πεδίου (Ground Truth)

Και εδώ όμως τα προβλήματα που παρουσιάζονται όπως η μπερδεμένη κατάτμηση στοιχείων ιδίως στις λιθοδομές οφείλονται κατά κύριο λόγο στον μικρό όγκο εισαγμένων δεδομένων. Συνεπώς παρατέθηκαν παρακάτω εικόνες οι οποίες χρησιμοποιήθηκαν ως δοκιμές χωρίς να έχουν εικόνες αλήθειας πεδίου έτσι ώστε να υπάρχει μια ποιοτική προβολή των προβλέψεων σε πραγματικές συνθήκες (Πίνακας 29).

Αρχική εικόνα	Εικόνα πρόβλεψης
	
	


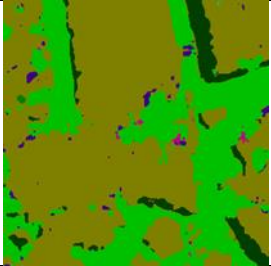
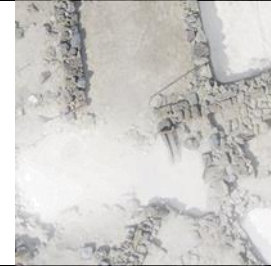
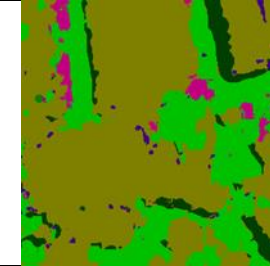

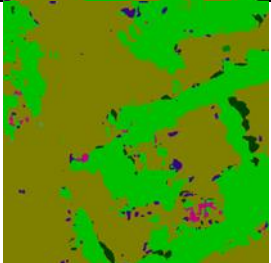

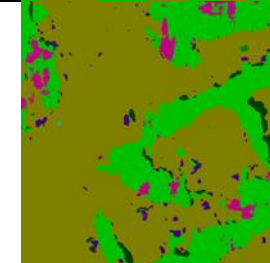
Πίνακας 29: Εικόνες δοκιμών χωρίς αληθινή βάση για αρχιτεκτονική SegNet με βήμα 0.001

Στις παραπάνω εικόνες (Πίνακας 29) διακρίνονται όλες οι κατηγορίες κατάτμησης που έχουν οριστεί για την εκπόνηση της εργασίας ως χρωματικές ετικέτες στις εικόνες πρόβλεψης. Συγκρίνοντας αυτές με τις αρχικές εικόνες παρατηρείτε ένα αποτέλεσμα κατάτμησης με πολύ καλά όρια και διαχωρισμό των λίθων και των γκρεμισμάτων σε αρκετά καλό βαθμό. Αυτή η συνθήκη μας δείχνει την αξία για άλλη μια φορά της μελέτης εικόνων RGB-D. Ακόμη και σε σημεία της εικόνας που χαρακτηριστικά στοιχεία δεν θα μπορούσαν να διακρίνουν διαφορές, η χρήση του μοντέλου αυτού βοηθάει στην κατάτμηση εδάφους από λίθους. Βέβαια και εδώ η κατάτμηση δεν έχει πραγματοποιηθεί άρτια όπως είναι λογικό καθώς το πρόβλημα της περιπλοκής των λιθοδομών με τα γκρεμίσματα παραμένει, σε μικρό βέβαια βαθμό. Τέλος το μοντέλο αυτό φαίνεται σε πιο ομοιόμορφες εικόνες, όσον αφορά την εκσκαφή που έχει πραγματοποιηθεί,

παρατηρείτε μια βελτίωση στην κατάτμηση. Αυτό είναι λογικό να συμβαίνει καθώς τα στοιχεία βάθους στο τέταρτο κανάλι δεν εμφανίζουν διαφορές, με αποτέλεσμα να «μπερδεύουν» το μοντέλο.

6.1.3 Σύγκριση Επιλεγμένων Μεθόδων Σημασιολογικής Κατάτμησης για RGB και RGB-D εικόνες

Τα δύο μοντέλα που επιλέχθηκαν ως τελικά για την κάθε κατηγορία RGB και RGB-D δίνουν αρκετά αξιόλογα αποτελέσματα στην κατάτμηση. Το μοντέλο ResNet101_fcn να παρουσιάζει το πρόβλημα της μη αναγνώρισης μικρών στοιχείων, καθώς και την απουσία στην κατάτμηση γκρεμισμάτων. Το μοντέλο SegNet επίσης παρουσίασε το πρόβλημα της μη ορθής κατηγοριοποίησης γκρεμισμάτων, η της ένταξης γκρεμισμάτων ως μέρη λιθοδομών.

Εικόνα RGB	ResNet101_fcn	Εικόνα RGB-D	SegNet
			
			

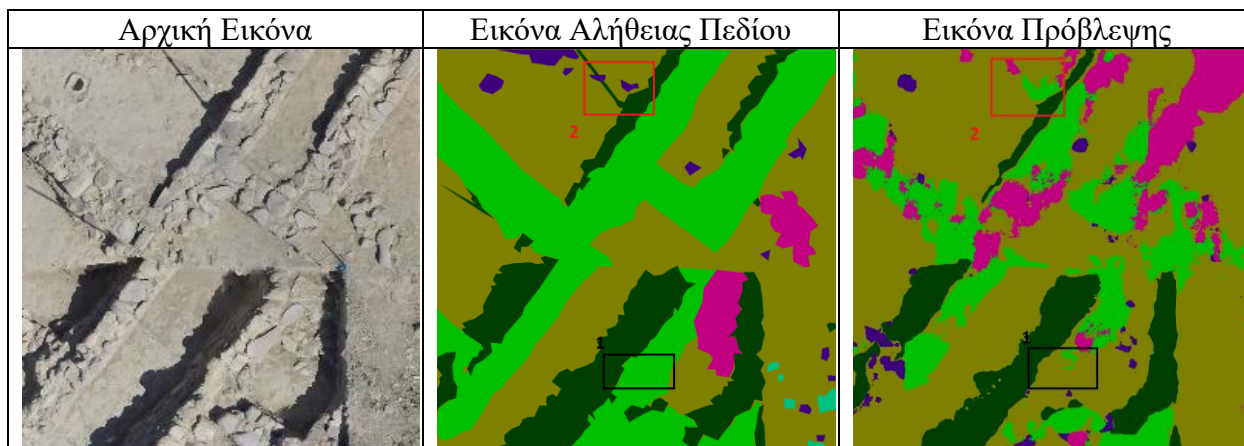
Πίνακας 30: Πίνακας σύγκρισης εικόνων από μοντέλα ResNet101_fcn και SegNet

Παρατηρώντας τον παραπάνω πίνακα (Πίνακας 30) είναι πιο εύκολο να γίνει σύγκριση των αποτελεσμάτων στις κοινές εικόνες πρόβλεψης που δημιουργήθηκαν από κάθε ένα από τα τελικά μοντέλα. Και οι δύο περιπτώσεις κατάτμησης είναι αρκετά ομαλές με το κυρίαρχο πρόβλημα του μοντέλου ResNet101_fcn να μην αναγνωρίζει σχεδόν καθόλου τα γκρεμίσματα ή να ορίζει λίθους και γκρεμίσματα ως λιθοδομές (Πίνακας 31 ετικέτα 1 και 3), καθώς και ραδιομετρικά και σχηματικά έχουν ίδια χαρακτηριστικά, εφόσον οι λιθοδομές και τα γκρεμίσματα απαρτίζονται από λίθους. Επίσης άλλο ένα πρόβλημα της κατάτμησης σε αυτό το μοντέλο είναι η οριακή ανυπαρξία κατάτμησης ξερόχortων και πετρών (Πίνακας 31 ετικέτα 2).



Πίνακας 31: Ποιοτικά Προβλήματα μοντέλου αρχιτεκτονικής ResNet101_fcn με βήμα 0.001

Αυτό το πρόβλημα με την ένταξη του βάθους φαίνεται να διορθώνεται αρκετά στο μοντέλο SegNet, λόγω του επιπλέον στοιχείου της υψομετρικής διαφοράς από το έδαφος που δίνεται χάρη στο τέταρτο κανάλι. Επίσης ανατρέχοντας στα ποσοτικά χαρακτηριστικά των δυο μοντέλων (Πίνακας 19 και Πίνακας 26) παρατηρούνται ακρίβειες, στον πίνακα του SegNet(Πίνακας 19), οι οποίες δεν συμβαδίζουν με την πραγματικότητα, καθώς η κατάτμηση της εικόνας από τον αλγόριθμο είναι ίσως και καλύτερη από αυτή της αλήθειας πεδίου (Ground Truth) στις δοκιμές 1 και 3, στην δοκιμή 2 η ύπαρξη αναχώματος δημιουργεί κάποια προβλήματα στον ορθό διαχωρισμό γκρεμισμάτων από λιθοδομές(Πίνακας 32 ετικέτα 1), καθώς δεν έχει λάβει υπόψη ο αλγόριθμος αυτή την ιδιομορφία βάθους, εφόσον δεν του δόθηκε παρόμοιο δεδομένο στην διαδικασία της εκπαίδευσης. Επίσης άλλο ένα πρόβλημα που παρουσιάζει η εισαγωγή βάθους είναι αυτό της μη ορθής κατάτμησης στοιχείων λόγω της μη ολοκληρωμένης εκσκαφής(Πίνακας 32 ετικέτα 2).



Πίνακας 32: Ποιοτικά Προβλήματα μοντέλου αρχιτεκτονικής SegNet

Στον πίνακα του ResNet101_fcn (Πίνακας 26) παρατηρείται ακριβώς η συνθήκη που θα έπρεπε. Δηλαδή ένα τελικό αποτέλεσμα με βάση τα ποσοστά αναγνώρισης επιτυχίας. Συνεπώς η τελική επιλογή σε αυτή την περίπτωση δεν μπορεί να είναι μια, καθώς δεν μπορούμε να συγκρίνουμε επαρκώς δυο μοντέλα τα οποία έχουν διαφορετικές μορφές δεδομένων εισόδου. Για να μπορεί να

βγει ένα τελικό αποτέλεσμα εδώ θα έπρεπε να μελετηθεί και ένα μοντέλο ResNet101_fcn με περισσότερα κανάλια για την ένταξη και του ασπρόμαυρου χάρτη βάθους. Συνεπώς ως τελικά αποτελέσματα θεωρούμε ότι παραμένουν και τα δύο μοντέλα.

6.2 Προβλήματα

Κυρίαρχο πρόβλημα κατά την διάρκεια της έρευνας ήταν η έλλειψη δεδομένων. Αυτό επηρέασε σημαντικά στο τελικό αποτέλεσμα, καθώς δεν υπήρχαν αρκετά δεδομένα για την άμεση εκπαίδευση του κώδικα. Επίσης ο χρόνος που κόστισε η κατασκευή αυτών ήταν σημαντικός, εφόσον υπήρχαν από την αρχή δεδομένα θα μπορούσαν να γίνουν δοκιμές με παραπάνω μεθόδους και απλούστερης μηχανικής μάθησης όπως είναι τα Random Forest (RF)[41], Support Vector Machines (SVM)[42] και άλλους.

Ένα ακόμη πρόβλημα που παρουσιάστηκε κατά την έρευνα ήταν η έλλειψη επαρκούς υπολογιστικής ισχύς. Ο υπολογιστής που χρησιμοποιήθηκε δεν είχε την απαραίτητη ισχύ για να φέρει εις πέρας τις διεργασίες που χρειάζονταν για την εκπαίδευση σε μικρό χρονικό διάστημα με αποτέλεσμα για κάθε μοντέλο η εκπαίδευση να διαρκεί από 8 έως 12 ώρες. Αυτή η συνθήκη του χρόνου δημιουργούσε και προβλήματα για την γρηγορότερη εκπόνηση, αλλά και για την ταχύτερη επιλογή ρυθμίσεων, καθώς για ρυθμίσεις όπως είναι η επιλογή βήματος μάθησης χρειάζεται να ληφθούν υπόψη τα δεδομένα προς επεξεργασία. Συνεπώς η επιλογή ορισμένων ρυθμίσεων έγιναν με βάση την επεξεργαστική ισχύ και όχι με βάση την βέλτιστη απόδοση του κώδικα για την παραγωγή του καλύτερου δυνατού μοντέλου.

6.3 Μελλοντικές Χρήσεις

Η έρευνα αυτή θα μπορούσε να εξελιχθεί με διάφορους τρόπους. Όπως με την δημιουργία ενός αλγόριθμου για αναγνώριση σε πραγματικό χρόνο (real time recognition) με απλούστερες εικόνες μνημείων για πιο απλές μελέτες. Επίσης θα μπορούσε να χρησιμοποιηθεί για την ανάπτυξη αλγόριθμου όπου δεν θα χρειάζεται πλέον κάποιος χρήστης να κατηγοριοποιεί τα στοιχεία ενός μνημείου, αλλά η κατάτμηση θα γίνεται αυτόματα από το ήδη εκπαιδευμένο μοντέλο σε συνδυασμό με τον κατάλληλο αλγόριθμο, καθώς και στην αυτοματοποίηση της αφαίρεσης σκιών, όπου αυτές δημιουργούν προβλήματα, μέσω του ίδιου αλγόριθμου. Ακόμη η έρευνα αυτή θα μπορούσε να χρησιμοποιηθεί για την διερεύνηση ενός νέου μοντέλου ή μιας νέας αρχιτεκτονικής, η οποία θα στοχεύει σε αυτό το πρόβλημα για καλύτερες κατατμήσεις ορθοφωτογραφιών μνημείων. Μια χρήσιμη μελέτη στην οποία θα μπορούσε να χρησιμοποιηθεί αυτή η έρευνα είναι η διερεύνηση της κατάτμησης ενός αρχαιολογικού χώρου σε εποχές με βάση την εκσκαφή και τη στρωματοποίηση που έχει δημιουργηθεί σε αυτή από τους αρχαιολόγους του συγκεκριμένου χώρου. Τέλος ενδιαφέρον θα είχε η αξιολόγηση και άλλων μοντέλων και κυρίως η επέκταση των μοντέλων ResNet που χρησιμοποιήθηκαν και στις RGB-D εικόνες, δίνοντας βάθος στον αριθμό καναλιών που αυτά μπορούν να χρησιμοποιούν.

Βιβλιογραφία

- [1] A. L. Samuel, “Some Studies in Machine Learning Using the Game of Checkers,” p. 21, 1959.
- [2] B. Neupane, T. Horanont, and J. Aryal, “Deep Learning-Based Semantic Segmentation of Urban Features in Satellite Images: A Review and Meta-Analysis,” *Remote Sens.*, vol. 13, no. 4, p. 808, Feb. 2021, doi: 10.3390/rs13040808.
- [3] S. S. Nath, G. Mishra, J. Kar, S. Chakraborty, and N. Dey, “A survey of image classification methods and techniques,” in *2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT)*, Kanyakumari District, India, Jul. 2014, pp. 554–557. doi: 10.1109/ICCICCT.2014.6993023.
- [4] H. Yang, B. Yu, J. Luo, and F. Chen, “Semantic segmentation of high spatial resolution images with deep neural networks,” *GIScience Remote Sens.*, vol. 56, no. 5, pp. 749–768, Jul. 2019, doi: 10.1080/15481603.2018.1564499.
- [5] K. He, X. Zhang, S. Ren, and J. Sun, “Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification,” in *2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile, Dec. 2015, pp. 1026–1034. doi: 10.1109/ICCV.2015.123.
- [6] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A Simple Way to Prevent Neural Networks from Overfitting,” p. 30.
- [7] M. G. Summa, L. Bottou, B. Goldfarb, F. Murtagh, C. Pardoux, and M. Touati, Eds., “Large-Scale Machine Learning with Stochastic Gradient Descent Léon Bottou,” in *Statistical Learning and Data Science*, 0 ed., Chapman and Hall/CRC, 2011, pp. 33–42. doi: 10.1201/b11429-6.
- [8] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” *ArXiv14126980 Cs*, Jan. 2017, Accessed: Oct. 10, 2021. [Online]. Available: <http://arxiv.org/abs/1412.6980>
- [9] E. Shelhamer, J. Long, and T. Darrell, “Fully Convolutional Networks for Semantic Segmentation,” *ArXiv160506211 Cs*, May 2016, Accessed: Oct. 10, 2021. [Online]. Available: <http://arxiv.org/abs/1605.06211>
- [10] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” *ArXiv151203385 Cs*, Dec. 2015, Accessed: Sep. 22, 2021. [Online]. Available: <http://arxiv.org/abs/1512.03385>
- [11] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, “Rethinking Atrous Convolution for Semantic Image Segmentation,” *ArXiv170605587 Cs*, Dec. 2017, Accessed: Sep. 22, 2021. [Online]. Available: <http://arxiv.org/abs/1706.05587>

- [12] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” *ArXiv150504597 Cs*, May 2015, Accessed: Sep. 21, 2021. [Online]. Available: <http://arxiv.org/abs/1505.04597>
- [13] Z. Guo *et al.*, “Semantic Segmentation for Urban Planning Maps Based on U-Net,” in *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, Valencia, Jul. 2018, pp. 6187–6190. doi: 10.1109/IGARSS.2018.8519049.
- [14] V. Badrinarayanan, A. Kendall, and R. Cipolla, “SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation,” *ArXiv151100561 Cs*, Oct. 2016, Accessed: Sep. 11, 2021. [Online]. Available: <http://arxiv.org/abs/1511.00561>
- [15] G. Cohen, S. Afshar, J. Tapson, and A. van Schaik, “EMNIST: an extension of MNIST to handwritten letters,” *ArXiv170205373 Cs*, Mar. 2017, Accessed: Sep. 13, 2021. [Online]. Available: <http://arxiv.org/abs/1702.05373>
- [16] D. W. Aha, D. Kibler, and M. K. Albert, “Instance-based learning algorithms,” *Mach. Learn.*, vol. 6, no. 1, pp. 37–66, Jan. 1991, doi: 10.1007/BF00153759.
- [17] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning representations by back-propagating errors,” p. 4, 1986.
- [18] M. Z. Alom *et al.*, “A State-of-the-Art Survey on Deep Learning Theory and Architectures,” *Electronics*, vol. 8, no. 3, p. 292, Mar. 2019, doi: 10.3390/electronics8030292.
- [19] H. Yakura, S. Shinozaki, R. Nishimura, Y. Oyama, and J. Sakuma, “Malware Analysis of Imaged Binary Samples by Convolutional Neural Network with Attention Mechanism,” in *Proceedings of the Eighth ACM Conference on Data and Application Security and Privacy*, Tempe AZ USA, Mar. 2018, pp. 127–134. doi: 10.1145/3176258.3176335.
- [20] M. Yani and B. Irawan, “Application of Transfer Learning Using Convolutional Neural Network Method for Early Detection of Terry’s Nail,” *J. Phys.*, p. 10, 2019.
- [21] Y. LeCun *et al.*, “Handwritten Digit Recognition with a Back-Propagation Network,” p. 9.
- [22] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998, doi: 10.1109/5.726791.
- [23] T. Joachims, “Text categorization with Support Vector Machines: Learning with many relevant features,” in *Machine Learning: ECML-98*, vol. 1398, C. Nédellec and C. Roudier, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998, pp. 137–142. doi: 10.1007/BFb0026683.
- [24] J. Nickolls, I. Buck, M. Garland, and K. Skadron, “Scalable parallel programming with CUDA,” in *ACM SIGGRAPH 2008 classes on - SIGGRAPH '08*, Los Angeles, California, 2008, p. 1. doi: 10.1145/1401132.1401152.

- [25] E. Photos-Jones, Y. Bassiakos, E. Filippaki, and A. Hein, Eds., *Proceedings of the 6th Symposium of the Hellenic Society for Archaeometry*. Ann Arbor, MI: University of Michigan Press, 2016. doi: 10.30861/9781407314303.
- [26] “Σημειώσεις για την Ορθοφωτογραφία σχολής Α.Τ.Μ.” Χαράλαμπος Ιωαννίδης, Καθηγητής ΕΜΠ.
- [27] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman, “LabelMe: A Database and Web-Based Tool for Image Annotation,” *Int. J. Comput. Vis.*, vol. 77, no. 1–3, pp. 157–173, May 2008, doi: 10.1007/s11263-007-0090-8.
- [28] A. Paszke *et al.*, “PyTorch: An Imperative Style, High-Performance Deep Learning Library,” *ArXiv191201703 Cs Stat*, Dec. 2019, Accessed: Sep. 19, 2021. [Online]. Available: <http://arxiv.org/abs/1912.01703>
- [29] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” *ArXiv14091556 Cs*, Apr. 2015, Accessed: Sep. 21, 2021. [Online]. Available: <http://arxiv.org/abs/1409.1556>
- [30] J. Xiao, J. Wang, S. Cao, and B. Li, “Application of a Novel and Improved VGG-19 Network in the Detection of Workers Wearing Masks,” *J. Phys. Conf. Ser.*, vol. 1518, p. 012041, Apr. 2020, doi: 10.1088/1742-6596/1518/1/012041.
- [31] K. He, X. Zhang, S. Ren, and J. Sun, “Identity Mappings in Deep Residual Networks,” *ArXiv160305027 Cs*, Jul. 2016, Accessed: Oct. 10, 2021. [Online]. Available: <http://arxiv.org/abs/1603.05027>
- [32] G. Chen *et al.*, “Fully Convolutional Neural Network with Augmented Atrous Spatial Pyramid Pool and Fully Connected Fusion Path for High Resolution Remote Sensing Image Segmentation,” *Appl. Sci.*, vol. 9, no. 9, p. 1816, May 2019, doi: 10.3390/app9091816.
- [33] N. Audebert, B. L. Saux, and S. Lefèvre, “Beyond RGB: Very High Resolution Urban Remote Sensing With Multimodal Deep Networks,” *ArXiv171108681 Cs*, Nov. 2017, Accessed: Sep. 24, 2021. [Online]. Available: <http://arxiv.org/abs/1711.08681>
- [34] Y. Wu *et al.*, “Demystifying Learning Rate Policies for High Accuracy Training of Deep Neural Networks,” *ArXiv190806477 Cs Stat*, Oct. 2019, Accessed: Sep. 24, 2021. [Online]. Available: <http://arxiv.org/abs/1908.06477>
- [35] D. Choi, C. J. Shallue, Z. Nado, J. Lee, C. J. Maddison, and G. E. Dahl, “On Empirical Comparisons of Optimizers for Deep Learning,” *ArXiv191005446 Cs Stat*, Jun. 2020, Accessed: Sep. 25, 2021. [Online]. Available: <http://arxiv.org/abs/1910.05446>
- [36] R. Sun, “Optimization for deep learning: theory and algorithms,” *ArXiv191208957 Cs Math Stat*, Dec. 2019, Accessed: Sep. 25, 2021. [Online]. Available: <http://arxiv.org/abs/1912.08957>
- [37] G. E. Nasr, E. A. Badr, and C. Joun, “Cross Entropy Error Function in Neural Networks: Forecasting Gasoline Demand,” p. 5.

- [38] M. Vakili, M. Ghamsari, and M. Rezaei, “Performance Analysis and Comparison of Machine and Deep Learning Algorithms for IoT Data Classification,” p. 13.
- [39] D. Powers, “Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation,” p. 25.
- [40] C. D. Brown and H. T. Davis, “Receiver operating characteristics curves and related decision measures: A tutorial,” *Chemom. Intell. Lab. Syst.*, vol. 80, no. 1, pp. 24–38, Jan. 2006, doi: 10.1016/j.chemolab.2005.05.004.
- [41] G. Biau, “Analysis of a Random Forests Model,” p. 33.
- [42] T. Evgeniou and M. Pontil, “Support Vector Machines: Theory and Applications,” in *Machine Learning and Its Applications*, vol. 2049, G. Paliouras, V. Karkaletsis, and C. D. Spyropoulos, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2001, pp. 249–257. doi: 10.1007/3-540-44673-7_12.