



ΠΑΝΕΠΙΣΤΗΜΙΟ
ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ

Τμήμα Μηχανικών
Βιομηχανικής Σχεδίασης & Παραγωγής

ΣΗΜΑΣΙΟΛΟΓΙΚΗ ΚΑΤΑΤΜΗΣΗ ΕΙΚΟΝΩΝ

από

ΜΑΡΙΑ Β. ΠΛΑΚΙΑ

A.M. 18389253

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Τμήμα Μηχανικών Βιομηχανικής Σχεδίασης και Παραγωγής

Πανεπιστήμιο Δυτικής Αττικής

Εισηγητής: **ΝΙΚΟΛΑΟΥ ΓΡΗΓΟΡΙΟΣ**

ΙΟΥΛΙΟΣ 2023

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Σημασιολογική Κατάτμηση Εικόνων

**Μαρία Β. Πλακιά
Α.Μ. 18389253**

Εξεταστική Επιτροπή:

Γρηγόριος Νικολάου, Λέκτορας

Σουλτάνα Βασιλειάδου, Επίκουρη Καθηγήτρια

Χρήστος Δρόσος, Ε.ΔΙ.Π

ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΠΤΥΧΙΑΚΗΣ/ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

Ο/η κάτωθι υπογεγραμμένος/η ΠΛΑΚΙΑ ΜΑΡΙΑ του ΒΑΣΙΛΕΙΟΥ, με αριθμό μητρώου 18389253 φοιτήτρια του Τμήματος Βιομηχανική Σχεδίασης και Παραγωγής της Σχολής Μηχανικών του Πανεπιστημίου Δυτικής Αττικής, δηλώνω ότι:

«Είμαι συγγραφέας αυτής της πτυχιακής/διπλωματικής εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της, είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, οι όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε ακριβώς είτε παραφρασμένες, αναφέρονται στο σύνολό τους, με πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο. Επίσης, βεβαιώνω ότι αυτή η εργασία έχει συγγραφεί από μένα αποκλειστικά και αποτελεί προϊόν πνευματικής ιδιοκτησίας τόσο δικής μου, όσο και του Ιδρύματος.

Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του πτυχίου μου».

Η Δηλούσα



Περιεχόμενα

ΚΕΦΑΛΑΙΟ 1	12
ΕΙΣΑΓΩΓΗ.....	12
1.1. Περιγραφή του αντικειμένου	12
1.2. Ιστορική αναδρομή.....	12
1.3. Ανασκόπηση της διπλωματικής εργασίας.....	13
ΚΕΦΑΛΑΙΟ 2	15
ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ.....	15
2.1. Εισαγωγή.....	15
2.2. Είδη της μηχανικής μάθησης.....	16
2.2.1. Μάθηση με επίβλεψη.....	17
2.2.2. Μάθηση χωρίς επίβλεψη	18
2.2.3. Ενισχυτική μάθηση	18
2.3. Τεχνητά Νευρωνικά Δίκτυα- ΤΝΔ	20
2.4. Βαθιά μάθηση	20
ΚΕΦΑΛΑΙΟ 3	23
ΣΗΜΑΣΙΟΛΟΓΙΚΗ ΚΑΤΑΤΜΗΣΗ ΕΙΚΟΝΩΝ.....	23
3.1. Εισαγωγή.....	23
3.2. Κατάτμηση εικόνας (image segmentation).....	24
3.3. Σημασιολογική κατάτμηση (semantic segmentation)	25
3.4. Εφαρμογές	27
3.4.1. Ρομποτική	27
3.4.2. Ιατρική.....	28
3.4.3. Παιχνίδι	28
3.4.4. Αναγνώριση προσώπου	29
3.5. Τεχνικές σημασιολογικής κατάτμησης εικόνας	29
3.5.1. CNN	30
3.5.2. Πλήρως συνελκτικά δίκτυα (Fully convolutional neural network)	37
3.5.3. Κωδικοποιητές-Αποκωδικοποιητές	38
ΚΕΦΑΛΑΙΟ 4	39
ΜΕΣΑ ΥΛΟΠΟΙΗΣΗΣ.....	39
4.1. Εισαγωγή.....	39

4.2. Αρχιτεκτονική Δικτύου	39
4.2.1. U-NET	39
4.2.2. DEEPLAB.....	41
4.2.3. VGG16.....	43
4.2.4. ResNet50	44
4.3. Περιβάλλον υλοποίησης.....	45
4.3.1. Google Colab.....	45
4.4. Βιβλιοθήκες	46
4.4.1. TENSORFLOW	46
4.4.2. KERAS.....	47
4.4.3. NUMPY.....	47
4.4.4. MATPLOTLIB	47
4.4.5. PANDAS.....	48
4.5. Αλγόριθμοι.....	48
4.5.1. Πλησιέστεροι γείτονες (nearest neighbors).....	48
4.5.2. Αλγόριθμος Βελτιστοποίησης ADAM.....	49
4.6. Μετρήσεις αξιολόγησης.....	49
4.7. Dataset.....	51
ΚΕΦΑΛΑΙΟ 5.....	55
ΥΛΟΠΟΙΗΣΗ	55
5.1. Επιλογή δεδομένων.....	55
5.2. Εισαγωγή βιβλιοθηκών.....	56
5.3. Προετοιμασία δεδομένων.....	56
5.4. Προεπεξεργασία δεδομένων	58
5.5. Αρχιτεκτονική	61
5.5.1. UNET – VGG16	61
5.5.2 DeepLabV3+ – ResNet50.....	63
5.6. Εκπαίδευση.....	64
5.7. Αποτελέσματα.....	67
5.7.1. Αποτελέσματα 1 ^{ου} μοντέλου:.....	67
5.7.2. Αποτελέσματα 2 ^{ου} μοντέλου :.....	77
5.8. Συγκρίσεις.....	84
ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΠΡΟΟΠΤΙΚΕΣ.....	97

ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

Εικόνα 1: Κατηγορίες AI.....	16
Εικόνα 2:Είδη της μηχανικής μάθησης.....	19
Εικόνα 3: Μηχανική και Βαθία μάθηση.....	22
Εικόνα 4:Η εικόνα με τον αντίστοιχο σχολιασμό της.....	23
Εικόνα 5:Είδη κατάτμησης εικόνας.....	25
Εικόνα 6:Αρχιτεκτονική για κατάτμηση εικόνων.....	31
Εικόνα 7: Μέθοδος Transfer Learning.....	36
Εικόνα 8: Αρχιτεκτονική Unet.....	41
Εικόνα 9:Μοντέλο Deeplabv3+.....	43
Εικόνα 10:Αρχιτεκτονική VGG16.....	44
Εικόνα 11:Αρχιτεκτονική ResNet50.....	45
Εικόνα 12: Αλγόριθμος Nearest Neighbor.....	49
Εικόνα 13:Παράδειγμα μέτρησης IoU.....	50
Εικόνα 14: Η εικόνα με την αντίστοιχη μάσκα 1.....	52
Εικόνα 15:Η εικόνα με την αντίστοιχη μάσκα 2.....	52
Εικόνα 16:Η εικόνα με την αντίστοιχη μάσκα 3.....	53
Εικόνα 17:Η εικόνα με την αντίστοιχη μάσκα 4.....	53
Εικόνα 18:Η εικόνα με την αντίστοιχη μάσκα 5.....	53
Εικόνα 19:Η εικόνα με την αντίστοιχη μάσκα 6.....	54
Εικόνα 20:Η εικόνα με την αντίστοιχη μάσκα 7.....	54
Εικόνα 21:Η εικόνα με την αντίστοιχη μάσκα 8.....	54
Εικόνα 22:Η εικόνα με την αντίστοιχη μάσκα 9.....	55
Εικόνα 23: Πρόβλεψη εικόνας από το Μοντέλο 1(Unet).....	74
Εικόνα 24:Πρόβλεψη εικόνας από το Μοντέλο 1(Unet).....	74
Εικόνα 25:Πρόβλεψη εικόνας από το Μοντέλο 1(Unet).....	75
Εικόνα 26:Πρόβλεψη εικόνας από το Μοντέλο 2(Deeplabv3+).....	81
Εικόνα 27:Πρόβλεψη εικόνας από το Μοντέλο 2(Deeplabv3+).....	81
Εικόνα 28:Πρόβλεψη εικόνας από το Μοντέλο 2(Deeplabv3+).....	81
Εικόνα 29:Πρόβλεψη εικόνας από το Μοντέλο 2(Deeplabv3+).....	84
Εικόνα 30:Πρόβλεψη εικόνας από το Μοντέλο 1(Unet).....	84
Εικόνα 31:Πρόβλεψη εικόνας από το Μοντέλο 2(Deeplabv3+).....	84
Εικόνα 32:Πρόβλεψη εικόνας από το Μοντέλο 1(Unet).....	85
Εικόνα 33:Πρόβλεψη εικόνας από το Μοντέλο 2(Deeplabv3+).....	85
Εικόνα 34:Πρόβλεψη εικόνας από το Μοντέλο 1(Unet).....	85
Εικόνα 35:Πρόβλεψη εικόνας από το Μοντέλο 2(Deeplabv3+).....	86
Εικόνα 36:Πρόβλεψη εικόνας από το Μοντέλο 1(Unet).....	86
Εικόνα 37:Πρόβλεψη εικόνας από το Μοντέλο 2(Deeplabv3+).....	86

Εικόνα 38:Πρόβλεψη εικόνας από το Μοντέλο 1(Unet).....	86
Εικόνα 39:Πρόβλεψη εικόνας από το Μοντέλο 2(Deeplanv3+).....	87
Εικόνα 40:Πρόβλεψη εικόνας από το Μοντέλο 1(Unet).....	87
Εικόνα 41:Πρόβλεψη εικόνας από το Μοντέλο 2(Deeplanv3+).....	87
Εικόνα 42:Πρόβλεψη εικόνας από το Μοντέλο 1(Unet).....	87
Εικόνα 43:Πρόβλεψη εικόνας από το Μοντέλο 2(Deeplanv3+).....	88
Εικόνα 44:Πρόβλεψη εικόνας από το Μοντέλο 1(Unet).....	88
Εικόνα 45:Πρόβλεψη εικόνας από το Μοντέλο 2(Deeplanv3+).....	88
Εικόνα 46:Πρόβλεψη εικόνας από το Μοντέλο 1(Unet).....	88
Εικόνα 47:Πρόβλεψη εικόνας από το Μοντέλο 2(Deeplanv3+).....	89
Εικόνα 48:Πρόβλεψη εικόνας από το Μοντέλο 1(Unet).....	89
Εικόνα 49:Πρόβλεψη εικόνας από το Μοντέλο 2(Deeplanv3+).....	89
Εικόνα 50:Πρόβλεψη εικόνας από το Μοντέλο 1(Unet).....	89
Εικόνα 51:Πρόβλεψη εικόνας από το Μοντέλο 2(Deeplanv3+).....	90
Εικόνα 52:Πρόβλεψη εικόνας από το Μοντέλο 1(Unet).....	90
Εικόνα 53:Πρόβλεψη εικόνας από το Μοντέλο 2(Deeplanv3+).....	90
Εικόνα 54:Πρόβλεψη εικόνας από το Μοντέλο 1(Unet).....	90
Εικόνα 55:Πρόβλεψη εικόνας από το Μοντέλο 2(Deeplanv3+).....	91
Εικόνα 56:Πρόβλεψη εικόνας από το Μοντέλο 1(Unet).....	91
Εικόνα 57:Πρόβλεψη εικόνας από το Μοντέλο 2(Deeplanv3+).....	91
Εικόνα 58:Πρόβλεψη εικόνας από το Μοντέλο 1(Unet).....	91
Εικόνα 59:Πρόβλεψη εικόνας από το Μοντέλο 2(Deeplanv3+).....	92
Εικόνα 60:Πρόβλεψη εικόνας από το Μοντέλο 1(Unet).....	92
Εικόνα 61:Πρόβλεψη εικόνας από το Μοντέλο 2(Deeplanv3+).....	92
Εικόνα 62:Πρόβλεψη εικόνας από το Μοντέλο 1(Unet).....	92

Κατάλογος Σχημάτων και Πινάκων

Πίνακας 1: Χρωμάτων.....	74
Πίνακας 2:Μέσος όρος αποτελεσμάτων για κάθε κατηγορία για το μοντέλο 1(Unet)	75
Πίνακας 3:Μέσος όρος αποτελεσμάτων για κάθε κατηγορία για το μοντέλο 2(Deeplanv3+).....	82
Πίνακας 4:Επιπλέον αποτελέσματα για το μοντέλο 2(Deeplanv3+).....	83
Πίνακας 5:Σύγκριση των αποτελεσμάτων και για τα δύο μοντέλα.....	93

Διάγραμμα 1: Μέσος όρος αποτελεσμάτων για το μοντέλο 1 (Unet).....	76
Διάγραμμα 2:Μέσος όρος αποτελεσμάτων για το μοντέλο 2 (Deeplanv3+).....	82
Διάγραμμα 3: Σύγκριση precision των δύο μοντέλων για κάθε κατηγορία.....	94
Διάγραμμα 4:Σύγκριση IoU των δύο μοντέλων για κάθε κατηγορία.....	95
Διάγραμμα 5:Σύγκριση RECALL των δύο μοντέλων για κάθε κατηγορία.....	95

Διάγραμμα 6: Σύγκριση precision των δύο μοντέλων για κάθε κατηγορία..... 96

Γράφημα 1: Συνάρτηση sigmoid και Tanh	32
Γράφημα 2: Συνάρτηση ReLU	33
Γράφημα 3: Συναρτήσεις ενεργοποίησης Sigmoid, Softmax	34
Γράφημα 4: Training loss.....	68
Γράφημα 5: Training Accuracy για το μοντέλο 1(Unet).....	69
Γράφημα 6: Validation Loss για το μοντέλο 1(Unet)	70
Γράφημα 7: Validation accuracy για το μοντέλο 1(Unet)	71
Γράφημα 8: Validation dice coefficient για το μοντέλο 1(Unet).....	72
Γράφημα 9: dice coefficient για το μοντέλο 1(Unet)	73
Γράφημα 10: Training Loss για το μοντέλο 2 (Deeplabv3+)	77
Γράφημα 11: Training accuracy για το μοντέλο 2 (Deeplabv3+).....	78
Γράφημα 12: Validation Loss για το μοντέλο 2 (Deeplabv3+)	79
Γράφημα 13: Validation accuracy για το μοντέλο 2 (Deeplabv3+).....	80

ΠΕΡΙΛΗΨΗ

Η παρούσα διπλωματική εργασία ασχολείται με την σημασιολογική κατάτμηση εικόνων που αποτελεί βασικό πεδίο στην επιστήμη της υπολογιστικής όρασης και εκφράζει την ταξινόμηση σε επίπεδο εικονοστοιχείων μιας εικόνας, με την απόδοση σε κάθε εικονοστοιχείο μιας συγκεκριμένης κατηγορίας αντικειμένου ή φόντου. Σκοπός της εργασίας αποτελεί η δημιουργία μιας μεθόδου σημασιολογικής κατάτμησης εικόνων με συνδυασμό πλήρως συνελκτικών δικτύων (FCN), αλγορίθμων μεταφοράς μάθησης και μεθόδους επεξεργασίας για όσο το δυνατόν μεγαλύτερη βελτίωση των αποτελεσμάτων, ώστε να είναι δυνατή η πρόβλεψη σε επίπεδο εικονοστοιχείων της δομής μιας εικόνας και των πληροφοριών που περιέχονται σ' αυτήν. Απώτερος στόχος αποτελεί η χρήση της πειραματικής αυτής προσέγγισης σε ένα ρομπότ ως η λειτουργία που θα του δώσει την δυνατότητα να αναγνωρίσει και να ταξινομήσει αντικείμενα στο χώρο. Το βασικό θεωρητικό πλαίσιο του ευρύτερου πεδίου της μηχανικής μάθησης, της βαθιάς μάθησης, της σημασιολογικής κατάτμησης μετά των μεθόδων και τεχνικών που χρησιμοποιούνται για την επίλυση του προβλήματος, η ανάλυση των μέσων που έδωσαν την δυνατότητα να υλοποιηθεί η εργασία αυτή, τα στάδια υλοποίησης δύο διαφορετικών προσεγγίσεων μετά των αποτελεσμάτων, συγκρίσεων και συμπερασμάτων συνθέτουν την παρούσα εργασία.

ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ: Βαθιά Μάθηση

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: Μηχανική μάθηση, βαθιά μάθηση, νευρωνικά δίκτυα, σημασιολογική κατάτμηση εικόνας, υπολογιστική όραση

ABSTRACT

The present thesis concerns the development of semantic segmentation of images which is a key field in the science of computer vision and expresses the classification at the level of pixels of an image, with the assignment to each pixel of a specific class of object or background. The aim of the work is to create a method of semantic image segmentation with a combination of fully convolutional networks (FCN), transfer learning algorithms and processing methods to improve the results as much as possible, so that it is possible to predict at the pixel level the structure of an image and the information contained in it. The ultimate goal is to use this experimental approach in a robot as the function that will enable it to recognize and classify objects in space. The basic theoretical framework of the wider field of machine learning, deep learning, semantic segmentation after the methods and techniques used to solve the problem, the analysis of the means that made it possible to implement this work, the implementation stages of two different approaches after the results, comparisons and conclusions compose the present work.

SUBJECT AREA: Deep Learning

KEY WORDS: Machine Learning, deep learning, neural networks, semantic segmentation of images, computer vision

ΕΥΧΑΡΙΣΤΙΕΣ

Η παρούσα διπλωματική εργασία ολοκληρώθηκε μετά από πολλές και επίμονες προσπάθειες, στο γνωστικό αντικείμενο της σημασιολογικής κατάτμησης εικόνων που αποδείχτηκε ιδιαίτερα ενδιαφέρον. Ευχαριστώ για την πολύτιμη υποστήριξη και τις συμβουλές του, τον επιβλέποντα καθηγητή κο Γρηγόριο Νικολάου.

Ακόμα θα ήθελα να ευχαριστήσω τους φίλους και την οικογένεια μου που με στήριξαν με τον δικό τους τρόπο όλο αυτό το διάστημα.

ΣΥΝΤΟΜΟΓΡΑΦΙΕΣ

ΤΝΔ Τεχνητό Νευρωνικό Δίκτυο

CNN Convolutional Neural Network (Συνελικτικό Νευρωνικό Δίκτυο)

FCN Fully convolutional neural network (πλήρως συνελικτικό νευρωνικό δίκτυο)

ReLU Rectified Linear Unit

RGB Red, Green, Blue color

IoU Intersection over Union

ΚΕΦΑΛΑΙΟ 1

ΕΙΣΑΓΩΓΗ

1.1. Περιγραφή του αντικειμένου

Η σημασιολογική κατάτμηση εικόνων αποτελεί ένα σημαντικό επιστημονικό πεδίο στο πλαίσιο της επιστήμης των υπολογιστών και ιδιαίτερα της υπολογιστικής όρασης που συνδυάζει την επεξεργασία εικόνων με την ανάπτυξη τεχνικών και αλγορίθμων της βαθιάς μηχανικής μάθησης. Στόχος της είναι η αναγνώριση και η ανάλυση αντικειμένων και στοιχείων περιβάλλοντος χώρου σε μια εικόνα, αποδίδοντας σε κάθε εικονοστοιχείο της εικόνας μια σημασιολογική ετικέτα που ουσιαστικά αποτελεί την περιγραφή του τύπου ή της κατηγορίας του αντικειμένου.

Οι εφαρμογές της είναι πολλές και ποικίλες όπως στην ρομποτική, στην αυτόνομη οδήγηση, στην ιατρική, στην ασφάλεια, στην ανίχνευση ανωμαλιών, στην επεξεργασία εικόνων και βίντεο, στην εικονική πραγματικότητα.

Οι προοπτικές της είναι ενθαρρυντικές καθώς ολοένα και πιο βελτιωμένα μοντέλα και μέθοδοι αναπτύσσονται σε συνδυασμό με αύξηση παραγωγής των δεδομένων εικόνων, ώστε να δίνεται η δυνατότητα για πιο ακριβή και αποτελεσματικά αποτελέσματα σημασιολογικής κατάτμησης και περαιτέρω πρόοδο και καινοτομία στο τομέα αυτό.

1.2. Ιστορική αναδρομή

Στην δεκαετία του '80 το πρόβλημα της σημασιολογικής κατάτμησης εικόνων αντιμετωπιζόταν με τεχνικές και μεθόδους χειροκίνητης επεξεργασίας. Η κατηγοριοποίηση των εικονοστοιχείων ήταν βασισμένη συνήθως στο χρώμα, στην υφή ή σε κάποια άλλα χαρακτηριστικά. Οι μέθοδοι αυτές αλλάζουν ριζικά στα τέλη της δεκαετίας του '80 και αρχές του '90 όταν εμφανίζονται τα νευρωνικά δίκτυα. Την τελευταία δεκαετία η σημασιολογική κατάτμηση σημειώνει εξαιρετική πρόοδο με την χρήση των βαθιών νευρωνικών δικτύων

δηλαδή των συνελκτικών (CNNs) και ιδιαίτερα των πλήρως συνελκτικών δικτύων (FCNs). Οι μέθοδοι που βασίζονται στα δίκτυα αυτά έδωσαν την δυνατότητα να αναγνωρίζονται και να κατηγοριοποιούνται πολλά αντικείμενα σε πραγματικό χρόνο με αποτέλεσμα να αναπτυχθούν πολλές πρακτικές εφαρμογές.[52]

1.3. Ανασκόπηση της διπλωματικής εργασίας

Η παρούσα εργασία ερευνά το πεδίο της σημασιολογικής κατάτμησης εικόνων τόσο σε θεωρητικό πλαίσιο όσο και σε πρακτικό καθώς επιχειρείται η δημιουργία ενός κώδικα που θα έχει την δυνατότητα να επεξεργαστεί ένα σύνολο δεδομένων (εικόνων) και να επιδείξει πόσο αποτελεσματικά μπορεί να επιλύσει το πρόβλημα της σημασιολογικής κατάτμησης των εικόνων ώστε η χρήση του σε μια εφαρμογή ρομποτικής π.χ. ένα ρομπότ να λειτουργεί αποδοτικά και να είναι εφικτή η αναγνώριση αντικειμένων και πλαισίων στο χώρο.

Η εργασία είναι δομημένη σε πέντε κεφάλαια συμπεριλαμβανομένης της εισαγωγής που αποτελεί το πρώτο κεφάλαιο. Στο 2^ο και 3^ο κεφάλαιο αναπτύσσεται το θεωρητικό υπόβαθρο των επιστημονικών πεδίων που βασίζεται η εργασία. Πιο συγκεκριμένα στο δεύτερο κεφάλαιο εξετάζονται οι βασικές έννοιες της μηχανικής μάθησης που αποτελεί σημαντική περιοχή της τεχνητής νοημοσύνης, τα είδη της, η τεχνική των νευρωνικών δικτύων και η έννοια της βαθιάς μηχανικής μάθησης, στις τεχνικές της οποίας βασίζεται η ανάπτυξη μοντέλων για σημασιολογική κατάτμηση εικόνων. Το τρίτο κεφάλαιο επικεντρώνεται στην έννοια της σημασιολογικής κατάτμησης, στα είδη της, στις εφαρμογές που ευρέως γίνεται χρήση της, στις τεχνικές που χρησιμοποιούνται για την εφαρμογή της όπως τα βαθιά νευρωνικά δίκτυα CNN, FCN. Το τέταρτο κεφάλαιο αποτελεί την σύνδεση θεωρητικού και πρακτικού μέρους με την ανάλυση των εργαλείων που χρησιμοποιήθηκαν προκειμένου να υλοποιηθούν δύο διαφορετικές προσεγγίσεις για την εφαρμογή της σημασιολογικής κατάτμησης εικόνων.

Στο πέμπτο κεφάλαιο, το πρακτικό μέρος, περιγράφονται τα στάδια υλοποίησης της εργασίας, με παρουσίαση των αποτελεσμάτων που επιτεύχθηκαν, τις συγκρίσεις των αξιολογήσεων και των αποδόσεων τους. Ως επίλογος της εργασίας, θα παρατεθούν τα συμπεράσματα που εξήχθησαν, οι προκλήσεις που αντιμετωπίστηκαν κατά την διάρκεια της υλοποίησης, πιθανές βελτιώσεις για μελλοντικές επεκτάσεις της εργασίας και γενικότερες προοπτικές του επιστημονικού πεδίου της σημασιολογικής κατάτμησης εικόνας.

ΚΕΦΑΛΑΙΟ 2

ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

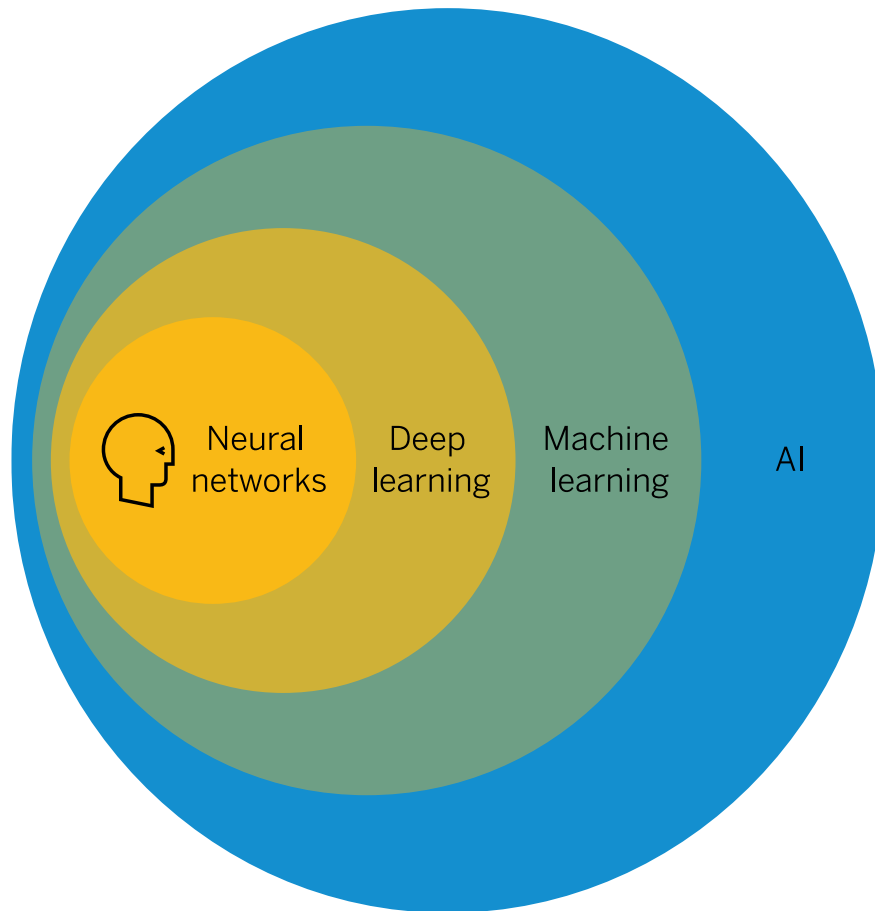
2.1. Εισαγωγή

Η μηχανική μάθηση ερευνά τις διαδικασίες που τόσο οι άνθρωποι όσο και οι μηχανές χρησιμοποιούν ώστε να αποκτήσουν γνώση και εκπαίδευση προκειμένου να δημιουργήσουν νέες τεχνολογίες. Η μηχανική μάθηση αποτελεί υποομάδα της Τεχνητής Νοημοσύνης. Αποσκοπεί στην εκπαίδευση των υπολογιστών μέσα από τα δεδομένα που τους παρέχονται και στην βελτίωση τους, μέσω της εμπειρίας που αποκτείται, χωρίς προγραμματισμό.

Οι εκπαιδευμένοι αλγόριθμοι στην μηχανική μάθηση βρίσκουν συσχετισμούς και μοτίβα σε μεγάλο όγκο δεδομένων, με αποτέλεσμα να παίρνουν τις πλέον ενδεδειγμένες αποφάσεις και να προβλέπουν κατάλληλα, με βάση την ανάλυση που επιτυγχάνεται. Όσο περισσότερα είναι τα δεδομένα, στα οποία υπάρχει πρόσβαση των υπολογιστών, τόσο μεγαλύτερη βελτίωση και ακρίβεια έχουν οι εφαρμογές μηχανικής μάθησης στην χρήση τους. Θα μπορούσε λοιπόν να οριστεί ότι μηχανική μάθηση είναι η «δημιουργία μοντέλων ή προτύπων με βάση ένα σύνολο δεδομένων, από ένα υπολογιστικό σύστημα» (Βλαχάβας κ.α. 2020)

Ο τεράστιος όγκος δεδομένων (π.χ. επιχειρήσεων, διαδικτύου των πραγμάτων, υγείας, κοινωνικών δικτύων, Παγκόσμιου Ιστού, φωτογραφιών, βίντεο κ.α.), οι εξελιγμένοι αλγόριθμοι, η συνεχώς αυξημένη ζήτηση για εφαρμογές που καλύπτουν το μεγαλύτερο μέρος της κοινωνικής δραστηριότητας, είναι μόνο κάποιοι από τους παράγοντες στους οποίους οφείλεται η ραγδαία ανάπτυξη που παρουσιάζει τα τελευταία χρόνια η μηχανική μάθηση.[1]

Σχετικές εφαρμογές πλέον έχουν εξαπλωθεί παντού γύρω μας, στα σπίτια, στο αστικό περιβάλλον, στα μέσα ψυχαγωγίας, στην περίθαλψη κ.α.



Εικόνα 1: Κατηγορίες AI

[Αναφορά: [What is machine learning? | Definition, types, and examples | SAP Insights](#)]

2.2. Είδη της μηχανικής μάθησης

Τρία είναι κυρίως τα είδη της μηχανικής μάθησης, στα οποία κατηγοριοποιούνται οι διάφορες τεχνικές της, ανάλογα της φύσης των προβλημάτων που καλούνται να επιλύσουν: Μάθηση με επίβλεψη 'supervised learning', μάθηση χωρίς επίβλεψη 'unsupervised learning', ενισχυτική μάθηση 'reinforcement learning'. Εδώ θα πρέπει να αναφερθεί ότι συνδυασμός των ανωτέρω τεχνικών μπορεί να χρησιμοποιηθεί ανάλογα με το πρόβλημα, όπως π.χ. η ημιεπιβλεπόμενη μάθηση (semi-supervised).

2.2.1. Μάθηση με επίβλεψη

Στην υπό επίβλεψη μάθηση ή αλλιώς εποπτευόμενη, κάποιος λειτουργεί ως επιτηρητής του συστήματος και επισημαίνει με την σωστή τιμή την έξοδο της συνάρτησης, που καλείται να εκπαιδευτεί με ένα σύνολο δεδομένων, ώστε να δημιουργηθεί ένα μοντέλο. Το μοντέλο αυτό, θα μπορέσει να λειτουργήσει στην συνέχεια και να προβλέψει τις ετικέτες εξόδου για νέα δεδομένα που θα εισαχθούν.[1] Τα δεδομένα εκπαίδευσης είναι ένα σύνολο από παραδείγματα, την ανάλυση των οποίων αναλαμβάνει ο εποπτευόμενος αλγόριθμος προκειμένου να δημιουργήσει την συνάρτηση που θα χρησιμοποιηθεί και για νέα παραδείγματα. Σημαντική θεωρείται η απεικόνιση των χαρακτηριστικών εισόδου της συνάρτησης. Η ακρίβεια της βασίζεται στην επιλογή αυτών των χαρακτηριστικών που περιγράφουν καλύτερα το αντικείμενο εισόδου, χωρίς ο αριθμός των χαρακτηριστικών να είναι μεγάλος. Στην συνέχεια η επιλογή παραμέτρων ελέγχου, εφόσον χρειάζεται, και η μέτρηση της απόδοσης με σύνολο δοκιμών θα αξιολογήσει την ακρίβεια της συνάρτησης και την επίτευξη του στόχου που τέθηκε.[2]

Οι πιο συνήθεις υποκατηγορίες των εποπτευόμενων εργασιών είναι η «ταξινόμηση» (classification) και η «παρεμβολή» (regression).

Στην ταξινόμηση το σύστημα μαθαίνει να διαχωρίζει τα δεδομένα σε μία από πολλές κατηγορίες με βάση περιπτώσεις εκπαίδευσης με γνωστή την κατηγορία που ανήκουν π.χ. η διάγνωση ασθένειας για ασθενείς με παρόμοια συμπτώματα με άλλους ήδη διαγνωσμένους. Χρησιμοποιούνται μοντέλα όπως τα δέντρα απόφασης, οι K-πλησιέστεροι γείτονες, ταξινομητές μηχανικής μάθησης, τεχνητά νευρωνικά δίκτυα. [3] Στην παρεμβολή το σύστημα δημιουργεί μοντέλο πρόβλεψης μιας παραμέτρου με αριθμητική τιμή ή με συνεχή τιμή βάσει άλλων παραμέτρων που είναι διαθέσιμες όπως π.χ. η τιμή ενός οικοπέδου σύμφωνα με τα χαρακτηριστικά του (έκταση, θέση κ.α.) , εκτίμηση ισοτιμίας νομισμάτων. Χρησιμοποιούνται μοντέλα όπως αλγόριθμοι γραμμικής ή μη γραμμικής παρεμβολής , τεχνητά νευρωνικά δίκτυα. [4]

2.2.2. Μάθηση χωρίς επίβλεψη

Η μάθηση χωρίς επίβλεψη εστιάζει στην ανάλυση δεδομένων χωρίς ετικέτα και χωρίς να παρεμβαίνει ανθρώπινη παρουσία. Οι ιδιότητες των δεδομένων είναι αυτές που θα βοηθήσουν να ανακαλυφθούν σχέσεις και ομάδες και να παραχθούν σαν αποτελέσματα «πρότυπα» που θα δώσουν τις απαραίτητες πληροφορίες. Η μέθοδος αυτή προτιμάται όταν αναζητείται προσδιορισμός διαφόρων τάσεων, για γενετικά χαρακτηριστικά, για ομαδοποιήσεις και γενικότερα για ερευνητικούς σκοπούς.

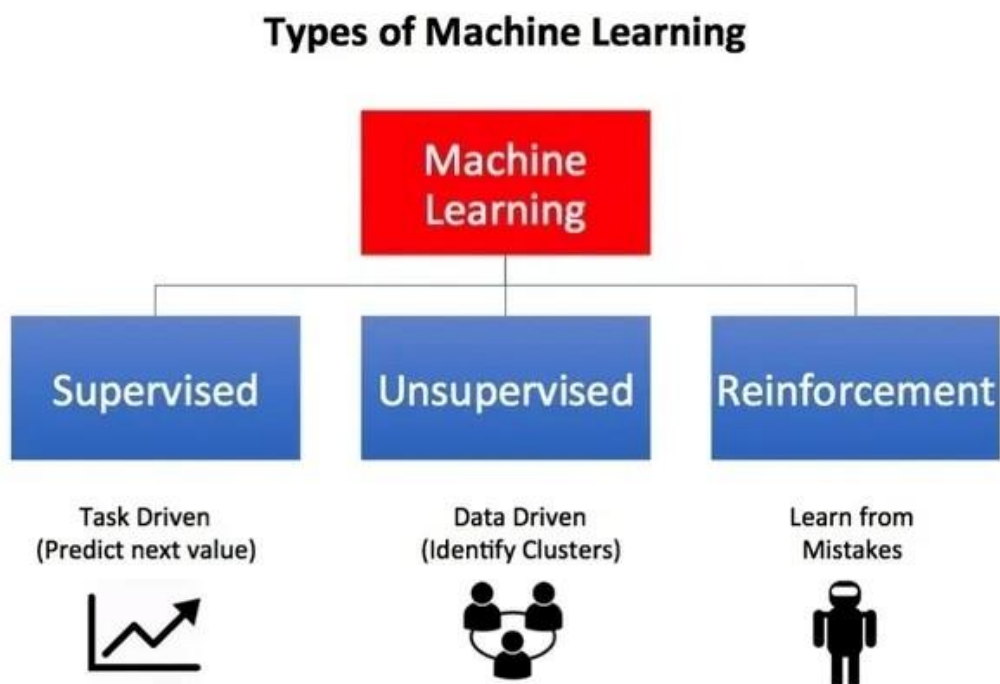
Συνήθεις εργασίες στην κατηγορία αυτή αποτελούν η ομαδοποίηση και οι κανόνες συσχέτισης. [5] Η ομαδοποίηση (clustering) διαχωρίζει τα πολύπλοκα δεδομένα σε όσο το δυνατόν διαφορετικές ομάδες που να διαφέρουν αρκετά μεταξύ τους. Παράδειγμα η ανάλυση θεμάτων από ένα μεγάλο σύνολο κειμένων όπου η ομαδοποίηση βοηθά στον εντοπισμό συγκεκριμένων θεμάτων όπως πολιτική, οικονομία κλπ. ώστε να αναπτυχθούν συστήματα ανάκτησης πληροφορίας ή να διευκολυνθεί η ανάγνωση και κατανόηση κειμένων. Κάποιοι από τους αλγόριθμους ομαδοποίησης είναι ο K-means και ο ιεραρχικής ομαδοποίησης. Οι κανόνες συσχέτισης (association rules) επιλέγονται μεταξύ εκείνων που υπάρχουν σε έναν όγκο δεδομένων κύρια με δεδομένα συναλλαγών, σε μια διαδικασία που προσπαθεί να ανακαλύψει τους πιο σημαντικούς και ενδιαφέροντες. Συνήθως χρησιμοποιούνται στην ανάλυση των προτιμήσεων των καταναλωτών σε αγοραστικά είδη και προϊόντα. Σχετικοί αλγόριθμοι για αυτή την κατηγορία είναι οι Apriori και Growth. [6] Ιδιαίτερη προσέγγιση στην κατηγορία αυτή αποτελούν τα Τεχνητά Νευρωνικά Δίκτυα (ΤΝΔ) καθώς και τα 'βαθιά νευρωνικά δίκτυα' (deep neural networks) που χρησιμοποιούνται σε πλήθος εφαρμογών.

2.2.3. Ενισχυτική μάθηση

Η ενισχυτική μάθηση αφορά τον τρόπο με το οποίο ένα υπολογιστικό σύστημα ή αλλιώς 'πράκτορας', κατά την αλληλεπίδραση του με το περιβάλλον του, έχει την δυνατότητα να μαθαίνει συγκεκριμένες συμπεριφορές. Ο αλγόριθμος

κάθε χρονική περίοδο έχει την εικόνα σε ποια κατάσταση βρίσκεται το περιβάλλον του και από τις διαθέσιμες, σύμφωνα με τις πιθανότητες, επιλογές που έχει, ακολουθεί την 'πολιτική' του και την μεταβάλλει ανάλογα με την αποκτώμενη εμπειρία. Απώτερος στόχος αποτελεί η ανταμοιβή που μπορεί να πάρει από το περιβάλλον για τον τρόπο που ενήργησε και όσο το δυνατόν η μεγιστοποίηση της. Στην ενισχυτική μάθηση δεν υπάρχει γνωστή έξοδος, ούτε ανθρώπινη παρεμβολή αλλά οι μόνες πληροφορίες που λαμβάνει το σύστημα είναι μια αριθμητική ανταμοιβή. Σαν αποτέλεσμα μέσα από αυτή την διαδικασία, το σύστημα εκπαιδεύεται να επιλέγει τις πιο σωστές ενέργειες προς το περιβάλλον του. Σχετικοί αλγόριθμοι ενισχυτικής μάθησης οι Q-Learning, Deep Q-Network, Policy Gradient κ.α.[7]

Παράδειγμα ενισχυτικής μάθησης αποτελεί ο έλεγχος κίνησης ρομπότ, η εκμάθηση διαφόρων παιχνιδιών, η αυτόματη οδήγηση κ.α.



Εικόνα 2:Είδη της μηχανικής μάθησης

[Αναφορά: <https://towardsdatascience.com/what-are-the-types-of-machine-learning-e2b9e5d1756f>]

2.3. Τεχνητά Νευρωνικά Δίκτυα- ΤΝΔ

Το Τεχνητό Νευρωνικό Δίκτυο (ΤΝΔ) είναι ένα μοντέλο επεξεργασίας πληροφοριών που παίρνει έμπνευση από το τρόπο με τον οποίο τα βιολογικά νευρικά συστήματα του ανθρώπινου εγκεφάλου επεξεργάζονται πληροφορίες. Το βασικό στοιχείο αυτού του μοντέλου είναι η πρωτότυπη δομή του συστήματος επεξεργασίας δεδομένων το οποίο αποτελείται από ένα μεγάλο αριθμό πυκνά διασυνδεδεμένων στοιχείων επεξεργασίας τα οποία λέγονται νευρώνες. Οι κόμβοι, όπως ονομάζονται οι τεχνητοί νευρώνες, σε ομάδες πολλαπλών στρωμάτων, λειτουργούν παράλληλα, ώστε όταν ένας λαμβάνει αριθμητικό σήμα, προχωρά σε επεξεργασία και στέλνει σήμα και στους υπόλοιπους συνδεδεμένους. Ακριβώς όπως στον άνθρωπο, το αποτέλεσμα που παρέχεται είναι η βελτίωση της αναγνώρισης προτύπων, της εμπειρίας και της γενικότερης μάθησης. [8]

Οι νευρώνες λειτουργούν με ανεξάρτητο τρόπο μεταξύ τους. Δέχονται σήματα εισόδου που πολλαπλασιάζονται με τα ονομαζόμενα βάρη. Στον αθροιστή, αθροίζονται οι επηρεασμένες από τα βάρη είσοδοι και στην συνέχεια η τιμή του αθροίσματος εισάγεται σε συνάρτηση ενεργοποίησης ώστε να παραχθεί η έξοδος. Η έξοδος μπορεί να καταλήγει στο περιβάλλον ή να αποτελεί είσοδο σε άλλους νευρώνες ή επίπεδα του δικτύου.

Τα Τεχνητά Νευρωνικά Δίκτυα χρησιμοποιούνται γενικότερα όπου απαιτείται έλεγχος, ταξινόμηση ή πρόβλεψη [9]

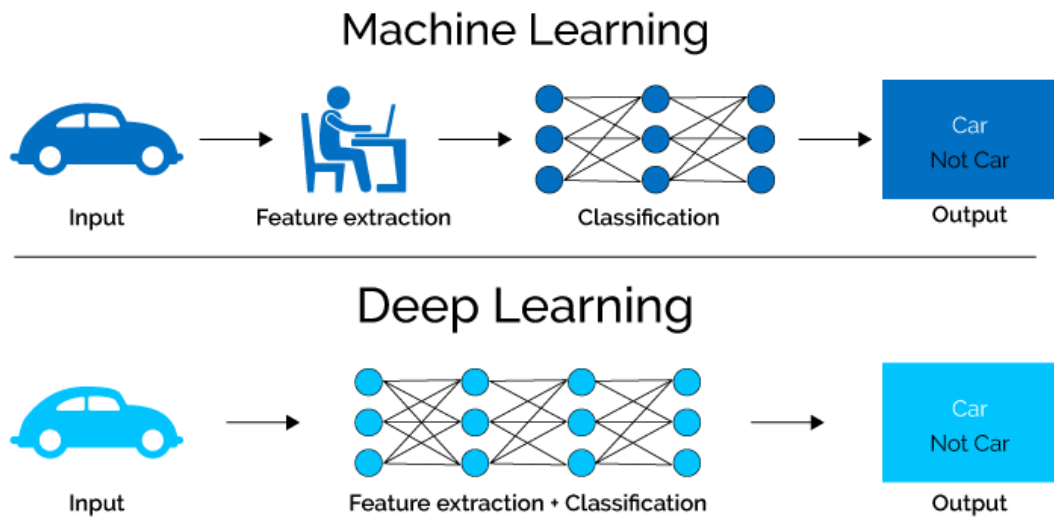
2.4. Βαθιά μάθηση

Μέσα στο υποσύνολο της μηχανικής μάθησης, που με τους αλγόριθμους της επιτρέπει στη τεχνητή νοημοσύνη να γίνει πιο έξυπνη χωρίς προγραμματισμό, βρίσκεται η βαθιά μάθηση. Η βαθιά μάθηση εμπνέεται από τις λειτουργίες του ανθρώπινου εγκεφάλου που αντιλαμβάνεται και επεξεργάζεται πολύμορφες και πολύπλοκες πληροφορίες από μεγάλο εύρος δεδομένων. Το πρότυπο της τεχνικής αυτής αποτελούν οι παραστάσεις και πληροφορίες που λαμβάνονται μέσω των ματιών, περνούν στον αμφιβληστροειδή χιτώνα, στην συνέχεια μέσω του οπτικού νεύρου φτάνουν στον εγκέφαλο. Με τον ίδιο τρόπο το

υπολογιστικό σύστημα μαθαίνει μέσα από πολλά επίπεδα επεξεργασίας να αναπαριστά δεδομένα επιχειρώντας να μάθει αφαιρέσεις υψηλού επιπέδου με την χρήση ιεραρχικών αρχιτεκτονικών.[10,11] Πιο απλά, στην βαθιά μάθηση ένα δίκτυο είναι συνδεδεμένο με πολλά επίπεδα επεξεργασίας, που βρίσκονται το ένα πίσω από το άλλο, ώστε τα δεδομένα εισόδου να περνούν υποχρεωτικά σε καθένα απ' αυτά διαδοχικά. [12] Σε κάθε επίπεδο το σύστημα εκπαιδεύεται να αναγνωρίζει κάποια από τα χαρακτηριστικά των δεδομένων, ξεκινώντας ιεραρχικά από τα πιο απλά έως και τα πιο σύνθετα. Μετά την εκμάθηση των χαρακτηριστικών που προέρχονται από αυτά τα ετερογενή δεδομένα, πολλαπλών πηγών, έχει την δυνατότητα να τα αναπαριστά ενοποιημένα.[13]

Τα τελευταία χρόνια η βαθιά μάθηση γνωρίζει τεράστια ανάπτυξη καθώς ανταπεξέρχεται με μεγάλη πρόοδο και επιτυχία σε ποικίλα προβλήματα όπως η υπολογιστική όραση, η παρακολούθηση κινήσεων, η αναγνώριση αντικειμένων και ενεργειών, η σημασιολογική τμηματοποίηση κ.α.

Τα μοντέλα βαθιάς μάθησης βασίζονται στα βαθιά νευρωνικά δίκτυα (deep neural networks) εκατοντάδων επιπέδων και ποικίλων συνδυασμένων αρχιτεκτονικών που παρουσιάζουν εκπληκτικές δυνατότητες μάθησης από πλήθος δεδομένων ήχου, εικόνας, κειμένου.[1]



Εικόνα 3: Μηχανική και Βαθία μάθηση

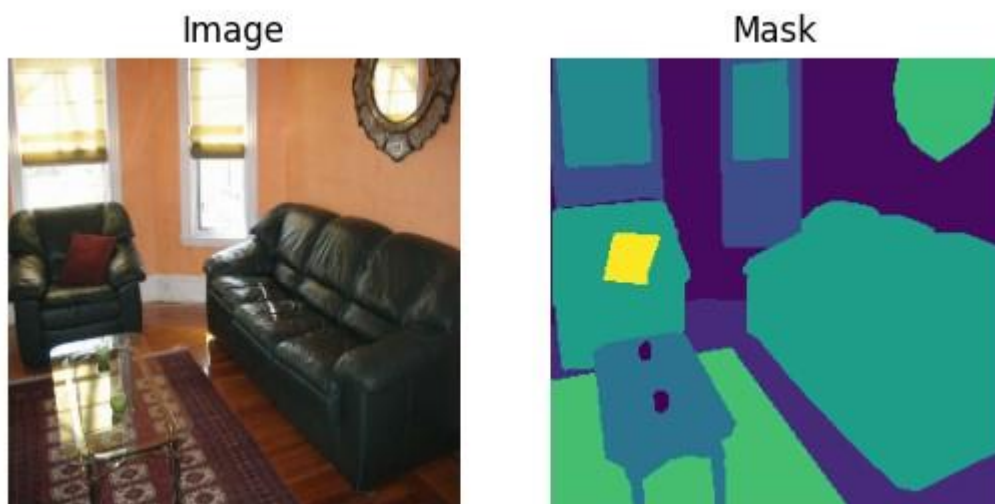
[Αναφορά: [Deep Learning Spreads \(semiengineering.com\)](http://semiengineering.com)]

ΚΕΦΑΛΑΙΟ 3

ΣΗΜΑΣΙΟΛΟΓΙΚΗ ΚΑΤΑΤΜΗΣΗ ΕΙΚΟΝΩΝ

3.1. Εισαγωγή

Το πρόβλημα της σημασιολογικής κατάτμησης εικόνας (image semantic segmentation) είναι αυτό που ενδιαφέρει πλέον όλο και περισσότερο τους ερευνητές της μηχανικής μάθησης και κυρίως αυτούς της υπολογιστικής όρασης. Ο σκοπός της σημασιολογικής κατάτμησης είναι η κατανόηση του οπτικού κόσμου από τους υπολογιστές, όπως ακριβώς τον κατανοούν οι άνθρωποι και μπορούν να αναγνωρίζουν στις εικόνες τα αντικείμενα και τα όρια τους. Πολλές είναι οι εφαρμογές που βασίζονται σε αποτελεσματικούς και ακριβείς υπολογισμούς κατάτμησης εικόνας όπως π.χ. στην ρομποτική, στην εικονική πραγματικότητα, στην αυτόνομη οδήγηση, στην πλοήγηση σε εσωτερικούς χώρους, στην ανάλυση ιατρικών εικόνων κ.α. Οι εφαρμογές αυτές ολοένα και αυξάνονται με την ανάπτυξη των μεθόδων βαθιάς μάθησης σε κάθε πεδίο δραστηριότητας και ιδιαίτερα σε όσα σχετίζονται με την όραση υπολογιστών περιλαμβανομένης και της σημασιολογικής κατάτμησης. [14,52]



Εικόνα 4: Η εικόνα με τον αντίστοιχο σχολιασμό της [49]

3.2. Κατάτμηση εικόνας (image segmentation)

Αρχικά ο όρος κατάτμηση ή τμηματοποίηση (segmentation) αναφέρεται στην διαδικασία με την οποία μια εικόνα μπορεί να διαιρεθεί σε πολλά τμήματα ή περιοχές. Τα επιμέρους αυτά τμήματα μπορούν να δώσουν την ανάλυση διαφορετικών μερών και να εξάγουν πληροφορίες ή χαρακτηριστικά από καθένα απ'αυτά. Η κατάτμηση περιλαμβάνει :

- Την σημασιολογική κατάτμηση εικόνας (image semantic segmentation), την ταξινόμηση δηλαδή των pixel μιας εικόνας σε σημασιολογικές κατηγορίες ανεξάρτητα από άλλες πληροφορίες του περιβάλλοντος. Για παράδειγμα θα μπορούσε να αναφερθεί, πως τα άτομα που εμφανίζονται σε μια εικόνα τμηματοποιούνται ως ένα αντικείμενο ενώ το περιβάλλον τους ως άλλο.
- Την κατάτμηση στιγμιότυπου (instance segmentation), που αφορά διαχωρισμό κάθε διαφορετικού αντικειμένου της εικόνας, λαμβάνοντας υπ' όψη τα όρια του, προσδιορίζοντας για κάθε εικονοστοιχείο σε ποιο αντικείμενο ανήκει. Στο προηγούμενο παράδειγμα τα άτομα θα εμφανίζονται το καθένα σαν ξεχωριστό αντικείμενο χωρίς όμως να μπορεί να προβλεφθεί τι ακριβώς είναι.
- Την πανοπτική κατάτμηση (panoptic segmentation) που αποτελεί συνδυασμό σημασιολογικής και κατάτμησης στιγμιότυπου και ενώ κατηγοριοποιεί τα εικονοστοιχεία όπως στην σημασιολογική, μπορεί να ξεχωρίσει διαφορετικά είδη της ίδιας κατηγορίας π.χ. την εμφάνιση δύο διαφορετικών ανθρώπων [15,16]

Semantic Segmentation vs. Instance Segmentation vs. Panoptic Segmentation



(a) Image



(b) Semantic Segmentation



(c) Instance Segmentation



(d) Panoptic Segmentation

Εικόνα 5:Είδη κατάτμησης εικόνας [15]

3.3. Σημασιολογική κατάτμηση (semantic segmentation)

Πιο αναλυτικά ο όρος σημασιολογική κατάτμηση, με την οποία ασχολείται η παρούσα εργασία, αναφέρεται στην κατηγοριοποίηση σε κλάσεις των εικονοστοιχείων (pixel), από τα οποία αναπαρίσταται μια εικόνα π.χ. δέντρο, άνθρωπος, σπίτι κλπ .Ουσιαστικά είναι η τεχνική που αποσκοπεί στην διαίρεση της εικόνας σε πολλά τμήματα που καθένα απ'αυτά ανήκει σε κάποια κατηγορία, οπότε σε κάθε εικονοστοιχείο της εικόνας αποδίδεται μια 'ετικέτα'. Η σημασιολογική κατάτμηση διαφέρει από την ανίχνευση αντικειμένων (object detection) για τον λόγο ότι δεν γίνεται διαχωρισμός διαφορετικών ειδών του ίδιου αντικειμένου, αλλά όσα ανήκουν στο συγκεκριμένο αντικείμενο λαμβάνουν την ίδια ετικέτα.

Η σημασιολογική κατάτμηση απαρτίζεται από τρεις επιμέρους εργασίες, την ταξινόμηση συγκεκριμένων αντικειμένων στην εικόνα, τον εντοπισμό των αντικειμένων και εύρεση των ορίων τους και την ομαδοποίηση των pixels που έχει σαν αποτέλεσμα μια 'μάσκα' κατάτμησης που οριοθετεί τα αντικείμενα της εικόνας ανάλογα με την κλάση που τους έχει δωθεί.[16]

Τα βασικά πλεονεκτήματα της σημασιολογικής κατάτμησης θα μπορούσαν να συνοψιστούν ως ακολούθως:

- Από μία εικόνα είναι δυνατή η εξαγωγή συγκεκριμένων πληροφοριών με μεγάλη ακρίβεια. Ο χωρισμός της εικόνας σε μικρότερα τμήματα διευκολύνει την διαχείριση της και επιτρέπει τον εντοπισμό των ζητούμενων αντικειμένων, με απώτερο στόχο οι πληροφορίες που θα ληφθούν να χρησιμοποιηθούν για προβλέψεις ή για να παρθούν σημαντικές αποφάσεις βασισμένες στο περιεχόμενο της. Ιδιαίτερα είναι σημαντικός ο ρόλος της όταν οι αποφάσεις αυτές αφορούν περιβάλλον που περιέχει πολλά αντικείμενα ή επικρατούν δύσκολες συνθήκες όπως περιορισμένη ορατότητα.
- Προσδιορίζεται η δομή τόσο των εικόνων όσο και των αντικειμένων ώστε να γίνεται πιο εύκολη η κατανόηση των περιεχομένων και η ανάλυση τους.
- Δίνεται η δυνατότητα η ποιότητα της εικόνας να βελτιωθεί, καθώς με την τμηματοποίηση της, απομονώνονται οι περιοχές που ενδιαφέρουν και εφαρμόζονται τεχνικές επεξεργασίας όπως π.χ. η ανίχνευση των άκρων ώστε να αφαιρεθούν όποια τυχόν ανεπιθύμητα στοιχεία υπάρχουν, μείωση του θορύβου, τακτοποίηση της αντίθεσης, του φόντου κλπ.
- Ανοίγει πολλά πεδία εφαρμογών στην υπολογιστική όραση καθώς δίνει την δυνατότητα να εντοπίζονται και να παρακολουθούνται σε πραγματικό χρόνο αντικείμενα που ενδιαφέρουν σε μια εικόνα [17]

Στους σχετικούς αλγόριθμους που χρησιμοποιούνται για το πρόβλημα της σημασιολογικής κατάτμησης περιλαμβάνονται τα πλήρως συνελκτικά δίκτυα (FCN), το U-NET, το Mask R-CNN και το DeepLab που με μηχανισμούς βαθιάς μάθησης μαθαίνουν τα χαρακτηριστικά των εικόνων και ταξινομούν σε κατηγορίες τα εικονοστοιχεία ανάλογα με αυτά τα χαρακτηριστικά.

3.4. Εφαρμογές

Η σημασιολογική κατάτμηση εικόνων είναι μια τεχνολογία που έχει αποδειχθεί πολύτιμη για την βελτίωση της αυτοματοποίησης στην καθημερινότητα μας. Ένα πλήθος εφαρμογών βασίζονται στην σημασιολογική κατάτμηση εικόνων που καλύπτει ένα μεγάλο εύρος δραστηριοτήτων και τομέων της καθημερινής ζωής. Ολοένα περισσότερες χρήσεις και εφαρμογές αναμένονται μελλοντικά να εμφανιστούν δεδομένης της συνεχούς τεχνολογικής ανάπτυξης.

3.4.1. Ρομποτική

Η ρομποτική έχει εισβάλει για τα καλά στην καθημερινότητα μας. Χρησιμοποιείται στον βιομηχανικό κλάδο, στην περίθαλψη, στον κλάδο της εφοδιαστικής κ.α. Η αποτελεσματικότητα εξαρτάται από την ικανότητα των ρομπότ να αντιλαμβάνονται με ακρίβεια τι υπάρχει γύρω τους.

Η σημασιολογική κατάτμηση εικόνων επιτρέπει στα ρομπότ να αναγνωρίζουν τα αντικείμενα που τα περιβάλλουν, να τα αναλύουν και να διαχωρίζουν σε κατηγορίες τα διάφορα μέρη τους ώστε να ενεργούν κατάλληλα π.χ. η απρόσκοπτη αυτόνομη πλοήγηση των ρομπότ στο οικιακό περιβάλλον αλλά και στο αστικό αφού μπορούν να διαφοροποιούν τύπους περιοχών και αντικειμένων όπως έπιπλα, κτίρια, πεζοδρόμια, δρόμους, βλάστηση κ.α. Για παράδειγμα, σε ένα πιθανό γεγονός συμμετοχής σε διεθνή διαγωνισμό ενός ρομπότ, η επιλογή της σημασιολογικής κατάτμησης εικόνων ως εφαρμογή στις λειτουργίες του, είναι ιδιαίτερα σημαντική έναντι άλλων τεχνικών καθώς βοηθά το ρομπότ να αναγνωρίσει και να ταξινομήσει τα διάφορα αντικείμενα στο περιβάλλον του έτσι ώστε να διευκολύνεται η αυτόνομη πλοήγηση του με την αναγνώριση των τυχόν εμποδίων που υπάρχουν ή την αναγνώριση αντικειμένων που θα πρέπει πιθανά να αποφύγει ή να αλληλεπιδράσει με αυτά.

Μια από τις σημαντικές επίσης εφαρμογές στο τομέα αυτό αποτελεί η αυτόνομη οδήγηση όπου το αυτόνομο όχημα χρησιμοποιώντας τα δεδομένα που του παρέχουν κάμερες και αισθητήρες, κατανοεί και ερμηνεύει το περιβάλλον του και με την ακρίβεια που του παρέχει η κατάτμηση εικόνας,

που μπορεί να αναγνωρίζει και να ταξινομεί τα στοιχεία που υπάρχουν στο δρόμο, όπως π.χ. αυτοκίνητα, ανθρώπους, πινακίδες κλπ., μπορεί να κινείται με ασφάλεια και να διεκπεραιώνει αποτελεσματικά το έργο του.

Στον τομέα αυτό συμβάλλει και η ανίχνευση των αντικειμένων αλλά και η τμηματοποίηση λωρίδας ώστε το όχημα να μπορεί να διαχωρίζει τις διαφορετικές λωρίδες και να ακολουθεί την ζητούμενη διαδρομή. [17]

3.4.2. Ιατρική

Μπορεί στην ιατρική, να εφαρμόζονται εδώ και χρόνια τεχνικές απεικόνισης όπως η ακτινογραφία, ο υπέρηχος, οι τομογραφίες κ.λπ. ωστόσο η κατάτμηση εικόνας έχει συμβάλει πάρα πολύ στην διάγνωση των ασθενειών. Οι γιατροί πλέον μπορούν να ανιχνεύσουν και να εντοπίσουν με ακρίβεια τη θέση και το μέγεθος όγκων ή οζιδίων ανάμεσα σε ιστούς που παρεμβάλλονται με πολύ καλύτερα αποτελέσματα από την απλή απεικόνιση. Με τις πληροφορίες αυτές μπορούν να αποφασίσουν καλύτερα ποια μέθοδο θεραπείας θα ακολουθήσουν. Ένα σημαντικό παράδειγμα αποτελεί και η εφαρμογή της κατάτμησης εικόνας στην απεικόνιση του ανθρώπινου εγκεφάλου. Η ακρίβεια και η ταχύτητα με τις οποίες λαμβάνονται οι πληροφορίες για αυτή την τόσο πολύπλοκη δομή με τους διαφορετικούς ιστούς συντελούν πέρα από την βελτίωση στην διάγνωση διάφορων εγκεφαλικών παθήσεων και στην βελτίωση της αντίδρασης στην υγειονομική περίθαλψη που απαιτείται.

3.4.3. Παιχνίδι

Η σημασιολογική κατάτμηση εικόνας χρησιμοποιείται στα βιντεοπαιχνίδια, σε παιχνίδια προσομοίωσης ώστε να επιτυγχάνεται ρεαλιστικό περιβάλλον παιχνιδιού. Ο διαχωρισμός φόντου, χαρακτήρων, αντικειμένων διευκολύνει τους προγραμματιστές να επεξεργάζονται ανεξάρτητα κάθε στοιχείο ώστε το παιχνίδι να γίνεται πιο διαδραστικό. Η δημιουργία εικονικού κόσμου και ο διαχωρισμός πραγματικών και εικονικών αντικειμένων με την χρήση της κατάτμησης εικόνας έχει επίσης αποτέλεσμα οι χρήστες να βιώνουν μια ανεπανάληπτη εμπειρία.[17]

3.4.4. Αναγνώριση προσώπου

Η σημασιολογική κατάτμηση στην αναγνώριση προσώπου κατηγοριοποιεί τα χαρακτηριστικά όπως μάτια, στόμα, μύτη. Χρησιμοποιείται στις εφαρμογές υπολογιστικής όρασης με στόχο να μπορούν να διακριθούν η ηλικία, έκφραση εθνικότητα κ.α. Μεγάλη βοήθεια προσφέρει σε προβλήματα που προκύπτουν σε περιβάλλοντα ασφάλειας όπως αεροδρόμια, σταθμούς, χώρους εργασίας κ.α. όπου κρίνεται αναγκαία η αναγνώριση και ταυτοποίηση ατόμου. [18]

Μια σειρά άλλων εφαρμογών θα μπορούσε να αναφερθεί όπως :

- η ανάλυση δορυφορικών εικόνων για τμηματοποίηση διάφορων τύπων γης ή για αυτόματη χαρτογράφηση της,
- η κατηγοριοποίηση ειδών ένδυσης όπως η εφαρμογή της Amazon 'Virtual Mirror', ή άλλων καταναλωτικών ειδών στο ηλεκτρονικό εμπόριο
- η αναγνώριση προϊόντων, κατά την παραγωγή, αυτόματα και αξιολόγηση της ποιότητάς τους κ.α.

3.5. Τεχνικές σημασιολογικής κατάτμησης εικόνας

Στο παρελθόν ερευνητικές εργασίες έχουν χρησιμοποιήσει διάφορα μοντέλα και τεχνικές για να επιλύσουν το πρόβλημα της σημασιολογικής κατάτμησης εικόνας . Πιο γνωστές προσεγγίσεις αποτελούν αυτές που βασίζονται :

- στο κατώφλι (κατανομή των εικονοστοιχείων σε κλάσεις ανάλογα της τιμής τους σε σχέση με την τιμή κατωφλίου)
- στην περιοχή που λειτουργούν (αναζήτηση ομοιοτήτων μεταξύ κοντινών εικονοστοιχείων)
- στις ακμές (ταξινόμηση εικονοστοιχείων σύμφωνα με την ανίχνευση ακμών με χρήση σχετικών φίλτρων)
- στην ομαδοποίηση (εικονοστοιχείων με κοινά χαρακτηριστικά).
- Στους μηχανισμούς βαθιάς μάθησης και ιδιαίτερα στα συνελκτικά δίκτυα.[19]

Πλέον οι σύγχρονες τεχνικές βαθιάς μάθησης που χρησιμοποιούνται για το πρόβλημα της σημασιολογικής κατάτμησης εικόνας προέρχονται κυρίως από τα συνελκτικά δίκτυα (CNN) και ιδιαίτερα από τα πλήρως συνελκτικά δίκτυα (FCN) που αναδεικνύονται σε ισχυρά εργαλεία για τις συγκεκριμένες εργασίες.

3.5.1. CNN

Το CNN είναι ο πιο δημοφιλής μηχανισμός βαθιάς μάθησης για προβλήματα υπολογιστικής όρασης καθώς και για την σημασιολογική κατάτμηση αφού η έξοδος του είναι εικόνα αντί για διάνυσμα σταθερού μήκους.

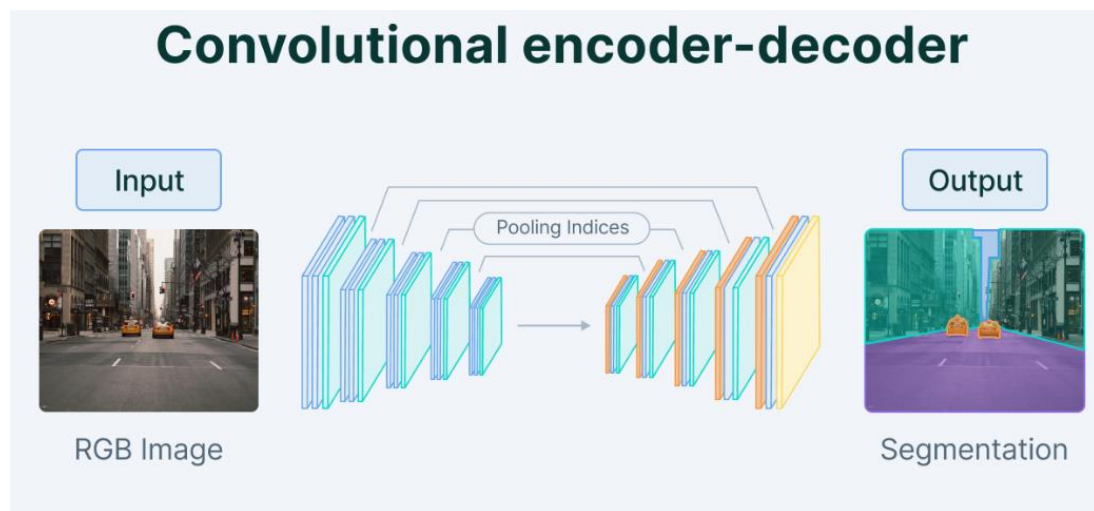
Το CNN αποτελεί περίπτωση βαθύς νευρωνικού δικτύου που μπορεί να επεξεργαστεί δεδομένα εισόδου και με μορφή εικόνων και να επιτύχει μεγάλες επιδόσεις στην οπτική αναγνώριση.

Το CNN ακολουθεί τη δομή ενός ΤΝΔ πολλών επιπέδων, ωστόσο διαφοροποιείται στην διαχείριση του σήματος εισόδου ώστε να επιτύχει καλύτερο αποτέλεσμα σε εικόνες. Η σύνθεση του αποτελείται από ένα επίπεδο εισόδου, ένα εξόδου και πολλά κρυφά. Τα κρυφά επίπεδα που χρησιμοποιεί είναι το συνελικτικό (convolutional), το συγκεντρωτικό (pooling) και το πλήρως συνδεδεμένο (fully connected).

Στο συνελικτικό στρώμα εφαρμόζονται μια σειρά φίλτρων για να εξαχθούν τα διάφορα χαρακτηριστικά της εικόνας. Η έξοδος λέγεται χάρτης χαρακτηριστικών και δίνει τις πληροφορίες για την εικόνα όπως γωνίες και άκρες. Στην συνέχεια ο χάρτης τροφοδοτεί άλλα επίπεδα για να μάθει και άλλα χαρακτηριστικά της εικόνας. Οι νευρώνες στο στρώμα αυτό έχουν τρεις διαστάσεις, πλάτος, ύψος, βάθος ώστε τρισδιάστατος όγκος εισόδου να μπορεί να μετατραπεί σε ένα όγκο εξόδου.

Στο επίπεδο συγκέντρωσης γίνεται προσπάθεια να μειωθεί το μέγεθος του χάρτη και συγκεκριμένα το ύψος και το πλάτος, ενώ το βάθος διατηρείται, με απώτερο στόχο την ελάττωση του αριθμού των υπολογισμών που απαιτούνται. Εδώ γίνεται η σύνοψη των χαρακτηριστικών από κάθε συνελικτικό επίπεδο και εξάγονται τα κυριότερα. Είδη στρωμάτων συγκέντρωσης είναι το max pooling και το average pooling. Στο max pooling εξάγεται η μέγιστη τιμή των χαρακτηριστικών του χάρτη για το μέρος της εικόνας που καλύπτει το φίλτρο ενώ στο average pooling εξάγεται η μέση τιμή.

Στο πλήρως συνδεδεμένο στρώμα γίνονται συνήθως οι μαθηματικές συναρτήσεις και ξεκινά η διαδικασία ταξινόμησης. Το στρώμα χρησιμοποιείται και για να συνδεθούν οι νευρώνες μεταξύ διαφορετικών στρωμάτων πριν το επίπεδο εξόδου καθώς περισσότερα πλήρως συνδεδεμένα στρώματα αποδίδουν καλύτερα από ένα.[20,21]



Εικόνα 6:Αρχιτεκτονική για κατάτμηση εικόνων [15]

Μία συνάρτηση ενεργοποίησης είναι απαραίτητη για την λειτουργία των νευρωνικών δικτύων και εκφράζει το πώς σ'ένα επίπεδο δικτύου το σταθμισμένο άθροισμα εισόδου γίνεται έξοδος από ένα ή πολλούς κόμβους. Η σωστή επιλογή της συνάρτησης ενεργοποίησης δίνει μεγάλα ποσοστά απόδοσης του δικτύου. Διαφορετικές συναρτήσεις ενεργοποίησης μπορούν να επιλεγούν για κάθε μέρος του δικτύου. Συνήθως στα κρυφά επίπεδα χρησιμοποιείται ίδια συνάρτηση ενεργοποίησης και διαφοροποιείται στο επίπεδο εξόδου, που η επιλογή εκεί εξαρτάται από τον τύπο της πρόβλεψης. Οι πιο δημοφιλείς συναρτήσεις ενεργοποίησης είναι οι σιγμοειδείς συναρτήσεις (sigmoid):, λογιστική (logistic) και υπερβολικής εφασπτομένης (tanh), ReLU και Softmax [1]

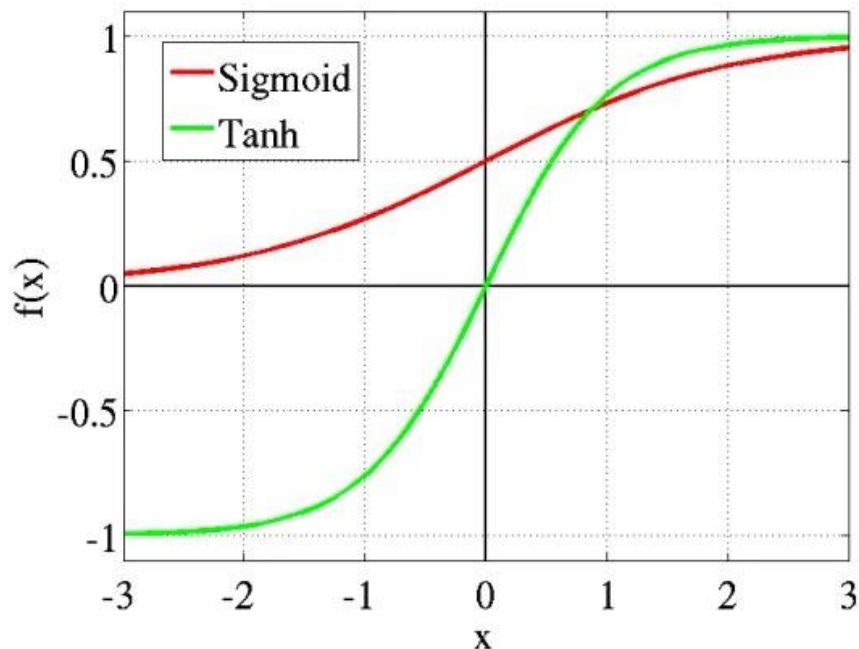
Η σιγμοειδής λογιστική συνάρτηση ενεργοποίησης λαμβάνει στην είσοδο τιμές πραγματικές, στην έξοδο 0 έως 1, οπότε όσο πιο θετικές τιμές παίρνει η είσοδος, η έξοδος θα είναι κοντά στο 1 ενώ, όσο πιο αρνητικές τιμές παίρνει η

είσοδος, η έξοδος θα είναι κοντά στο 0. Έχει τη μορφή (1) (όπου e :μαθηματική σταθερά) και το σχήμα της, απ' όπου προέρχεται και το όνομα της, είναι S.

$$\sigma(x) = \frac{1}{1+e^{-x}} \quad (1)$$

Η συνάρτηση Tanh στην είσοδο έχει πραγματικές τιμές και στην έξοδο από -1 έως 1. Η μορφή της (2) είναι (όπου e : μαθηματική σταθερά). [22]

$$f(x) = \frac{(e^x - e^{-x})}{(e^x + e^{-x})} \quad (2)$$



Γράφημα 1: Συνάρτηση sigmoid και Tanh

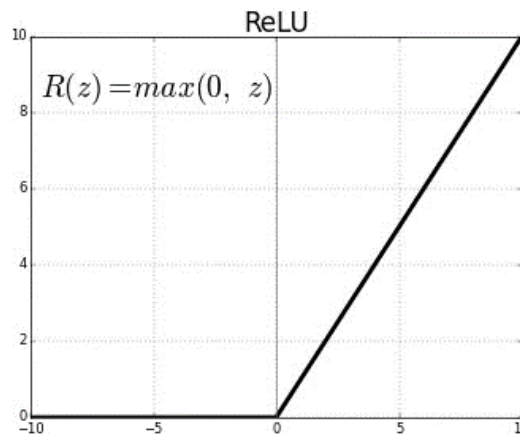
[Αναφορά:<https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6>]

Πιο συχνά χρησιμοποιούμενη συνάρτηση ενεργοποίησης στα κρυφά επίπεδα των σύγχρονων νευρωνικών δικτύων με κοινές αρχιτεκτονικές, όπως το CNN και το MLP (Multilayer Perceptron,) είναι η ReLU.

Η ReLU (Rectified Linear Unit) (3), είναι περισσότερο δημοφιλής γιατί εφαρμόζεται με τρόπο απλό και αποτελεσματικό, δεν είναι τόσο ευαίσθητη για

να φέρνει κωλύματα στην εκπαίδευση και απαιτεί λιγότερους υπολογισμούς με αποτέλεσμα καλύτερες επιδόσεις. Η συνάρτηση ReLU έχει την μορφή $\max(0.0, x)$,

$$f(x) = \max(0.0, x), \quad x = w_x + b \quad (3)$$



Γράφημα 2: Συνάρτηση ReLU

[Αναφορά: <https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6>]

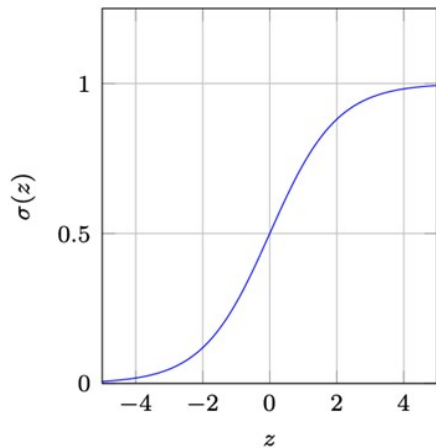
οπότε μια αρνητική τιμή εισόδου επιστρέφει με 0 ενώ η θετική τιμή παραμένει ως έχει και η παράγωγος της είναι σταθερή με τιμές 0 ή 1. Το γεγονός αυτό της δίνει πλεονέκτημα έναντι της σιγμοειδούς, γιατί στο λεγόμενο πρόβλημα 'vanishing gradient' που οι κλίσεις στους νευρώνες εξαφανίζονται και τα 'βάρη' μένουν στατικά, έχει μεγαλύτερες τιμές στις κλίσεις. Στην περίπτωση όπως που η τιμή της είναι 0 υπάρχει ο λεγόμενος κίνδυνος των 'νεκρών νευρώνων'. [22,23])

Όσον αφορά την συνάρτηση ενεργοποίησης για έξοδο, αυτή επιλέγεται ανάλογα της πρόβλεψης που επιζητείται. Για προβλήματα ταξινόμησης, δυαδικής ή πολλαπλών ετικετών επιλέγεται συνήθως η σιγμοειδής, ενώ για προβλήματα ταξινόμησης πολλαπλών κλάσεων η softmax.

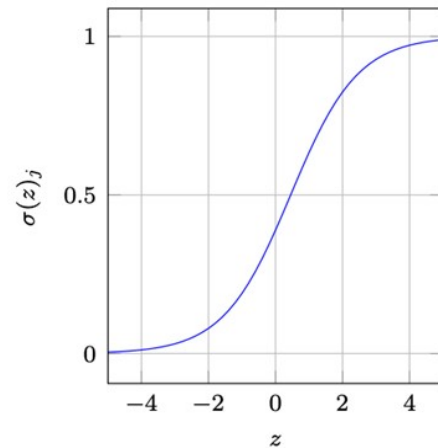
Η συνάρτηση ενεργοποίησης softmax (4) μπορεί να μετατρέψει ένα αριθμητικό διάνυσμα σε διάνυσμα πιθανοτήτων. Ουσιαστικά κανονικοποιεί τις

εξόδους και τις μεταβάλει από τιμές ενός σταθμισμένου αθροίσματος σε πιθανότητες του αθροίσματος σε μία .Κάθε τιμή εξόδου softmax θεωρείται σαν πιθανότητα να συμμετέχει σε κάθε κλάση. [22]

$$\sigma(x_i) = \frac{e^{x_i}}{\sum_{j=1}^n e^{x_j}} \quad (4)$$



(a) Sigmoid activation function.



(b) Softmax activation function.

Γράφημα 3: Συναρτήσεις ενεργοποίησης Sigmoid, Softmax

[Αναφορά:[Difference between Sigmoid and Softmax activation function? - Nomid!](#)]

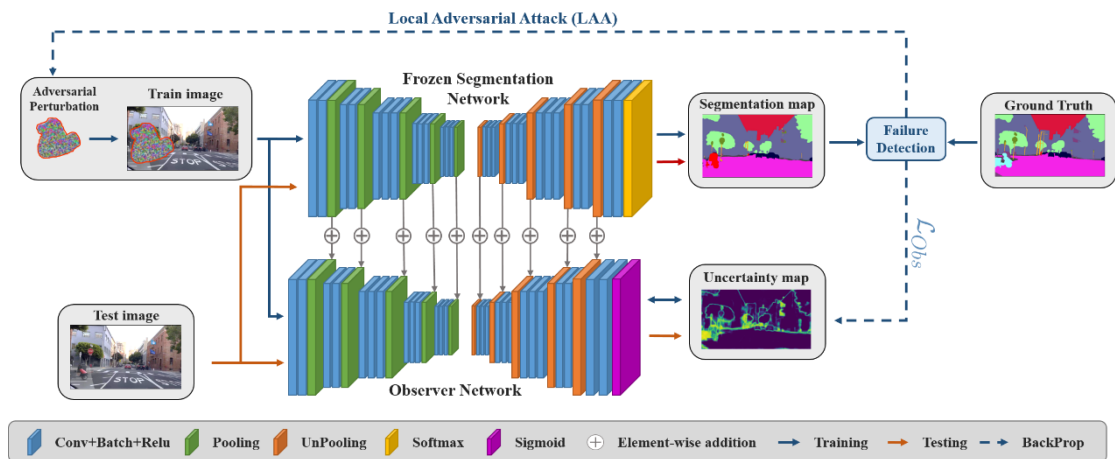
Ένα σύνολο από τεχνικές και αλγόριθμους απαιτούνται για την εκπαίδευση των CNN. Περιλαμβάνουν την επιλογή της καταλληλότερης αρχιτεκτονικής, και των ιδιαίτερων παραμέτρων όπως τον αριθμό των επιπέδων, των αριθμό των φίλτρων και των συντελεστών μάθησης. Επίσης, οι τεχνικές εκπαίδευσης συνελκτικικών δικτύων περιλαμβάνουν την επιλογή του αλγορίθμου βελτιστοποίησης, καθώς και τη χρήση τεχνικών όπως η αύξηση δεδομένων, η χρήση προεκπαιδευμένων μοντέλων, το dropout, η κανονικοποίηση, ο προγραμματισμός ρυθμού εκπαίδευσης, η πρόωρη διακοπή, η αρχικοποίηση βάρους.

Ο αλγόριθμος βελτιστοποίησης αποσκοπεί στην ελαχιστοποίηση της συνάρτησης κόστους ή απώλειας, κατά τη διάρκεια της εκπαίδευσης του μοντέλου.

Η απώλεια (loss) ουσιαστικά είναι ένα μέτρο αξιολόγησης και εκπαίδευσης του μοντέλου και δίνει το σύνολο των σφαλμάτων που μπορεί να υπάρχει σε κάθε εκπαιδευτική επανάληψη ή αξιολόγηση. Η απώλεια υποδεικνύει πόσο καλή ή όχι είναι η συμπεριφορά του μοντέλου μετά από τις επαναλήψεις βελτιστοποίησης.

Αύξηση Δεδομένων (Data augmentation): Η τεχνική αυτή αποσκοπεί στην τεχνητή επαύξηση του συνολικού μεγέθους της εκπαίδευσης μετασχηματίζοντας τα αρχικά δεδομένα, ώστε π.χ. υπάρχουσες εικόνες να θεωρούνται διαφορετικά δεδομένα. Η μέθοδος αυτή βελτιώνει το μοντέλο και βοηθά στο να αποφευχθεί η υπερβολική προσαρμογή (overfitting), φαινόμενο που παρουσιάζεται όταν τα βάρη έχουν υπερβολικά προσαρμοστεί στα δεδομένα εκπαίδευσης και στις ιδιομορφίες και χαρακτηριστικά που μπορεί να έχουν και δεν αποδίδουν πλέον σε νέα δεδομένα.[24]

Μεταφερόμενη μάθηση (Transfer Learning): Η απόδοση των CNN στην αναγνώριση αντικειμένων από εικόνες είναι πολύ υψηλή, ωστόσο η εκπαίδευση τους είναι αργή και δύσκολη και απαιτεί μεγάλο όγκο δεδομένων. Επίσης η επιλογή επιπέδων, σειράς στρωμάτων, κόμβων και μιας σειράς απαραίτητων παραμέτρων για την δημιουργία τους εκ του μηδενός, απαιτεί πολλές και χρονοβόρες δοκιμές. Τα προεκπαιδευμένα μοντέλα χρησιμοποιούνται ευρέως πλέον με την λογική ότι εφ' όσον έχουν ήδη εκπαιδευτεί σε παρόμοιο πρόβλημα τροποποιείται η δομή του για τη νέα εργασία, εφαρμόζεται το πάγωμα (freezing) των επιπέδων του καινούργιου μοντέλου και δεν τροποποιούνται τα βάρη. Στην εκπαίδευση που γίνεται σε νέα δεδομένα, γίνεται ενημέρωση βαρών μόνο στα νέα ή σε όσα επίπεδα έχουν τροποποιηθεί. Στην βαθιά μάθηση τα τελευταία επίπεδα συνήθως είναι αυτά που έχουν τροποποιηθεί. [25] Με τον τρόπο αυτό η εκπαίδευση γίνεται ταχύτερη και βελτιώνεται η ακρίβεια του μοντέλου.



Εικόνα 7: Μέθοδος Transfer Learning

[Αναφορά: [Triggering Failures: Out-Of-Distribution detection by learning from local adversarial attacks in Semantic Segmentation | valeo.ai blog \(valeoai.github.io\)](https://vleao.ai/blog/vleaoai.github.io)]

Dropout (Εγκατάλειψη): Η τεχνική αυτή συντελεί στην αποφυγή του overfitting αφήνοντας του κόμβους τυχαία απενεργοποιημένους στην διάρκεια της εκπαίδευσης, βελτιώνει την απόδοση, υλοποιείται με απλό τρόπο, ωστόσο σε μικρά δίκτυα περιορισμένων πόρων καθυστερεί την εκπαίδευση ή/και μειώνει την απόδοση αν η εκπαίδευση δεν επιτευχθεί μέχρι την σύγκλιση, οπότε στην περίπτωση αυτή θα πρέπει να αποφεύγεται.[26]

Κανονικοποίηση (Normalization): Πρόκειται για την τεχνική που ομαλοποιεί τις ενεργοποιήσεις των ενδιάμεσων στρωμάτων βαθένων νευρωνικών δικτύων προκειμένου η διαδικασία εκπαίδευσης να είναι σταθερότερη.[27]

Προγραμματισμός ρυθμού εκπαίδευσης (Learning Rate Scheduling): Η τεχνική αυτή σταδιακά μειώνει τον ρυθμό της μάθησης κατά την εκπαίδευση ώστε να επιτευχθεί μεγαλύτερη ακρίβεια.

Αρχικοποίηση βάρους (weight initialization): Είναι η τεχνική που καθορίζει τα βάρη σε τυχαίες μικρές τιμές που ορίζουν το σημείο που θα ξεκινήσει ο αλγόριθμος βελτιστοποίησης στο χώρο των τιμών βαρών.

Πρόωρη διακοπή (early stopping): Είναι η τεχνική κατά την οποία η εκπαίδευση διακόπτεται νωρίς όταν το σφάλμα επαλήθευσης μεγαλώνει προκειμένου να αποφευχθεί η υπερπροσαρμογή και να επιτευχθεί καλύτερη γενίκευση. [1]

3.5.2. Πλήρως συνελκτικά δίκτυα (Fully convolutional neural network)

Στην σημασιολογική κατάτμηση οι χωρικές πληροφορίες είναι απαραίτητες, για τον λόγο αυτό πλέον χρησιμοποιούνται τα πλήρως συνελκτικά δίκτυα. Τα FCN αποτελούν ένα τύπο νευρωνικού δικτύου που μπορεί να δεχθεί στην είσοδο μια εικόνα και να δώσει στην έξοδο μια εικόνα σε ίδιο μέγεθος, το κάθε εικονοστοιχείο της οποίας αντιπροσωπεύει το αντίστοιχο της εικόνας-εισόδου. Ο σχεδιασμός τους στοχεύει στην επίλυση των προβλημάτων κατάτμησης εικόνας, αντίθετα με τα παραδοσιακά δίκτυα που χρησιμοποιούνται για προβλήματα ταξινόμησης.

Η δομή τους μοιάζει με αυτή του CNN, ωστόσο τα τελευταία πλήρως συνδεδεμένα επίπεδα του CNN αντικαθίστανται με συνελκτικά, με απώτερο στόχο την παραγωγή ολοκληρωμένων εικόνων. Με τον τρόπο αυτό παράγεται μια 'μάσκα' για την έξοδο που μπορεί να επιλύσει το πρόβλημα της σημασιολογικής κατάτμησης αρκεί να επιτευχθεί η έξοδος της εικόνας στις ίδιες διαστάσεις με αυτής της εισόδου. Η πρόσθεση ενός επιπέδου ακόμα που βελτιώνει την ανάλυση της εικόνας εξόδου ως την επιθυμητή είναι το ονομαζόμενο επίπεδο upsampling, που μπορεί να επιτευχθεί με αντιστροφή της ενέργειας που πραγματοποιείται στα συγκεντρωτικά επίπεδα (unpooling) και με την ανάστροφη συνέλιξη. Με το unpooling αντιστρέφεται το αποτέλεσμα των συγκεντρωτικών επιπέδων που προηγήθηκαν, αυξάνοντας το μέγεθος της εικόνας, ενώ με την ανάστροφη συνέλιξη το δίκτυο εκπαιδεύεται να αυξάνει την ανάλυση της εικόνας χρησιμοποιώντας την συνέλιξη. Η τελική έξοδος περνά συνήθως από την συνάρτηση ενεργοποίησης softmax ώστε να παραχθεί κατανομή πιθανότητας σε κλάσεις για κάθε εικονοστοιχείο στην εικόνα-εξόδου.

Κύριο πλεονέκτημα των FCN αποτελεί ότι ολόκληρο το δίκτυο εκπαιδεύεται με την χρήση μιας μόνο συνάρτησης απώλειας που υπολογίζει την διαφορά πρόβλεψης με αλήθειας για κάθε εικονοστοιχείο στην έξοδο. Επίσης στην εκπαίδευση μπορούν να χρησιμοποιηθούν τεχνικές για αύξηση δεδομένων (αναστροφή, τυχαία περικοπή) που βελτιώνουν σημαντικά την αποτελεσματικότητά τους.[28]

3.5.3. Κωδικοποιητές-Αποκωδικοποιητές

Στην σημασιολογική κατάτμηση, οι αρχιτεκτονικές που χρησιμοποιούνται μπορούν συχνά να θεωρηθούν ως δίκτυα κωδικοποιητή-αποκωδικοποιητή. Ένα τέτοιο δίκτυο περιλαμβάνει αρχικά τον κωδικοποιητή (encoder) που προσπαθεί μια εισηγμένη εικόνα να την συμπιέσει με μορφή κώδικα και στην συνέχεια περιλαμβάνει τον αποκωδικοποιητή (decoder) που, αντίθετα, προσπαθεί να αναδομήσει την εικόνα με την χρήση των τεχνικών *unsampling*. Ο κωδικοποιητής είναι συχνά ένα δίκτυο ταξινόμησης προεκπαιδευμένο π.χ. VGG, ResNet που ακολουθείται από τον αποκωδικοποιητή που μπορεί να είναι και διαφορετικής δομής προκειμένου να καλύπτονται εφαρμογές σε μεγάλο εύρος. Ο αποκωδικοποιητής στοχεύει στην σημασιολογική ανάδειξη των διακριτικών χαρακτηριστικών ώστε να λάβει πυκνές εξόδους. Στην σημασιολογική κατάτμηση ο αποκωδικοποιητής προβάλλει τα διακριτικά χαρακτηριστικά που μαθαίνει ο κωδικοποιητής σε διαφορετικά στάδια σε επίπεδο εικονοστοιχείων. Ανάλογα την προσέγγιση που επιχειρείται χρησιμοποιούνται και διάφοροι μηχανισμοί αποκωδικοποίησης. [23]

ΚΕΦΑΛΑΙΟ 4

ΜΕΣΑ ΥΛΟΠΟΙΗΣΗΣ

4.1. Εισαγωγή

Τα μέσα που απαιτούνται προκειμένου να υλοποιηθεί μια πιθανή προσέγγιση, που μπορεί όσον το δυνατό πιο αποδοτικά να επιλύσει το πρόβλημα της κατάτμησης εικόνας, είναι κύρια η αρχιτεκτονική δικτύου που επιλέγεται, το περιβάλλον, οι βιβλιοθήκες, οι αλγόριθμοι, οι μετρήσεις και τα δεδομένα που χρησιμοποιούνται. Παρακάτω παρουσιάζονται η αρχιτεκτονική του δικτύου U-NET και του DEEPLAB, τα προεκπαιδευμένα μοντέλα VGG16 και ResNet50, το περιβάλλον GOOGLE COLAB, οι βιβλιοθήκες TENSORFLOW, KERAS, NUMPY, MATPLOTLIP και PANDAS, ο αλγόριθμος nearest neighbors και ο αλγόριθμος βελτιστοποίησης ADAM, οι μετρήσεις αξιολόγησης και το αρχείο δεδομένων ADE20K που επιλέχθηκε.

4.2. Αρχιτεκτονική Δικτύου

Για την αρχιτεκτονική του δικτύου που θα χρησιμοποιηθεί στην παρούσα εργασία, επιλέχθηκε αρχικά το δίκτυο U-NET σε συνδυασμό με το προεκπαιδευμένο μοντέλο VGG16 και στην συνέχεια για την υλοποίηση μιας δεύτερης προσπάθειας, ώστε να είναι εφικτή σύγκριση και αξιολόγηση αποτελεσμάτων, επιλέχθηκε το δίκτυο Deeplabv3+ σε συνδυασμό με το προεκπαιδευμένο μοντέλο ResNet50.

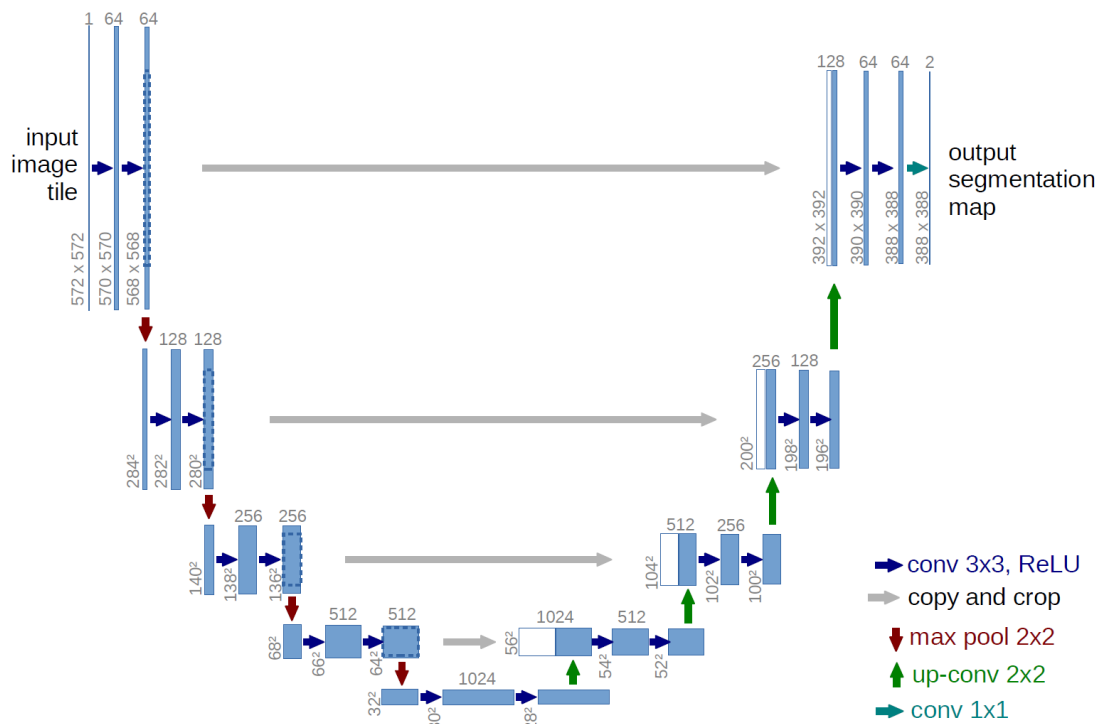
4.2.1. U-NET

Η δημιουργία του U-NET οφείλεται στους O.Ronneberger, P. Fischer, και T. Brox το 2015 προκειμένου να επιλυθεί σημασιολογική κατάτμηση εικόνων σχετικών με την ιατρική. Ουσιαστικά η αρχιτεκτονική του, είναι αυτή του δικτύου κωδικοποιητή-αποκωδικοποιητή. Είναι ένα καθαρά συνελικτικό δίκτυο που λαμβάνει εικόνες κάθε μεγέθους. Δύο διαδρομές ακολουθούνται στο δίκτυο, αριστερά μία συρρίκνωσης (κωδικοποιητής, με τέσσερα επίπεδα x 2 συνελίξεις το καθένα) και δεξιά μια εκτεταμένη (αποκωδικοποιητής, με

τέσσερα επίπεδα x 2 συνελίξεις το καθένα) που συνδέονται μέσω μιας γέφυρας (2 συνελίξεις) ώστε να δημιουργείται το σχήμα U.

Στην πρώτη διαδρομή οι χωρικές διαστάσεις μειώνονται κατά το ήμισυ ενώ διπλασιάζεται ο αριθμός των φίλτρων. Στην διαδρομή αυτή, τα διαδοχικά επίπεδα συνέλιξης ακολουθούνται από μια συνάρτηση ενεργοποίησης ReLU (η έξοδος της οποίας ενεργεί σαν σύνδεση παράκαμψης για το αντίστοιχο επίπεδο αποκωδικοποιητή) και από επίπεδα μέγιστης συγκέντρωσης. Από τον κωδικοποιητή εξάγεται ένα αφηρημένο περιεχόμενο της εικόνας εισόδου. Στην γέφυρα που συνδέει τις δύο διαδρομές τα συνελικτικά επίπεδα ακολουθούνται επίσης από συνάρτηση ενεργοποίησης ReLU.

Στην δεύτερη διαδρομή, αυτή του αποκωδικοποιητή, αναζητείται η δημιουργία μιας μάσκας κατάτμησης, σημασιολογική. Αυτό επιτυγχάνεται με συνδυασμό πληροφοριών τόσο χαρακτηριστικών όσο και χώρου μέσω μιας υψηλής ανάλυσης που πραγματοποιείται από μετατιθέμενες και κανονικές συνελίξεις. Για μεγαλύτερη ακρίβεια στα διαδοχικά βήματα της δεύτερης διαδρομής εφαρμόζονται συνδέσεις παράλειψης ώστε η έξοδος των μετατιθέμενων συνελικτικών στρωμάτων να συνδεθεί με τους χάρτες χαρακτηριστικών της πρώτης διαδρομής στο ίδιο επίπεδο. Δύο κανονικές συνελίξεις η μία μετά την άλλη εφαρμόζονται μετά από κάθε σύνδεση, με στόχο το μοντέλο να επιφέρει όσο το δυνατόν ακριβή έξοδο. [29,30,31].



Εικόνα 8: Αρχιτεκτονική Unet [29]

Το U-NET χρησιμοποιείται ως επί το πλείστον σαν βάση για σχετικές εφαρμογές σημασιολογικής κατάτμησης γιατί επιδέχεται τροποποιήσεις που αφορούν είτε τα φίλτρα, είτε την ανάλυση εισόδου ή εφαρμογή άλλων αρχιτεκτονικών χωρίς ουσιαστικά να απαιτείται αλλαγή της βασικής του αρχιτεκτονικής.[23]

4.2.2. DEEPLAB

Το DeepLab αποτελεί επίσης ένα μοντέλο βαθιάς μάθησης που χρησιμοποιείται στην σημασιολογική κατάτμηση εικόνας. Το συγκεκριμένο μοντέλο παρουσιάστηκε από την Google και αποτελεί εξέλιξη του μοντέλου FCN.

Η αρχιτεκτονική του αρχικά βασίζεται στην συνέλιξη φίλτρων (atrous convolution) (5). Η συνέλιξη αυτή αποσκοπεί στον έλεγχο της ανάλυσης όπου γίνεται ο υπολογισμός των χαρακτηριστικών όπως επίσης στην μεγέθυνση των οπτικών πεδίων των φίλτρων ώστε να είναι δυνατή η ενσωμάτωση περιβάλλοντος, μεγαλύτερου από το συνηθισμένο, χωρίς να είναι ανάγκη να

αυξηθούν οι παράμετροι ή ο υπολογιστικός όγκος. Μαθηματικά θα μπορούσε να οριστεί με την παρακάτω συνάρτηση :[32]

$$y[i] = \sum_{k=1}^K x [i + \gamma \cdot k]w[k] \quad (5)$$

Όπου $y[i]$: atrous convolution

$x[i]$: μονοδιάστατο σήμα

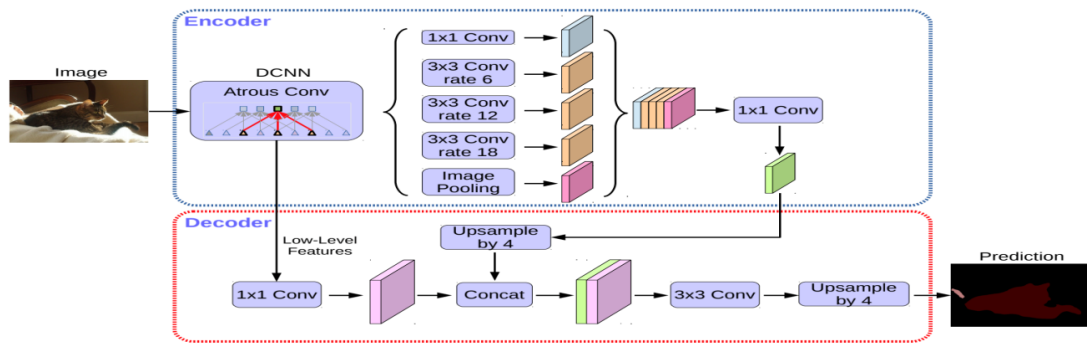
$w[k]$: φίλτρο

K : μήκος φίλτρου

R : ρυθμός διασκελισμού

Στην συνέχεια το μοντέλο βελτιώνεται με την έκδοση 2, με την εφαρμογή γνωστή ως ASPP (Atrous Spatial Pyramid Pooling) που έχει στόχο να κατηγοριοποιήσει τα αντικείμενα σε πολλές κλίμακες ώστε να καταγράψει υψηλότερης εμβέλειας πληροφορίες. Αυτό το επιτυγχάνει χρησιμοποιώντας φίλτρα πολλαπλών ρυθμών δειγματοληψίας και πεδία προβολής μεγάλης αποτελεσματικότητας σε επίπεδο συνελκτικών χαρακτηριστικών ώστε να επιτύχει την καταγραφή σε πολλαπλές κλίμακες των αντικειμένων αλλά και των πλαισίων της εικόνας. Παράμετροι ομαλοποίησης παρτίδας περιλαμβάνονται επίσης για να διευκολυνθεί η εκπαίδευση.

Στις επόμενες βελτιώσεις με την έκδοση 3 και 3+, για την αύξηση της ευκρίνειας των ορίων των αντικειμένων χρησιμοποιείται ένας αποτελεσματικός κωδικοποιητής και αποκωδικοποιητής, συνέλιξη atrous και συνδυασμός μεθόδων από DCNN (Deep Convolutional Neural Networks) και άλλα γραφικά μοντέλα. Ο κωδικοποιητής εφαρμόζει συνέλιξη atrous για την επεξεργασία πολλαπλής κλίμακας πληροφοριών ενώ ο αποκωδικοποιητής παρέχει βελτίωση των αποτελεσμάτων τμηματοποίησης στο μήκος ορίου των αντικειμένων.[33,34,35]

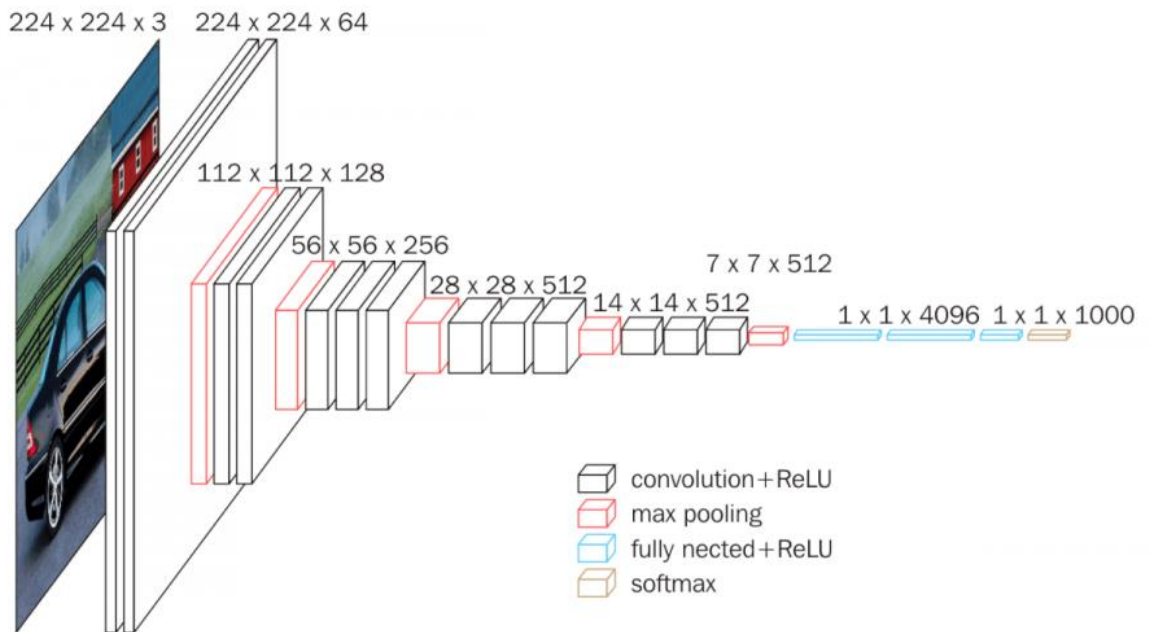


Εικόνα 9:Μοντέλο Deeplabv3+.

[αναφορά: <https://arxiv.org/abs/1802.02611>]

4.2.3. VGG16

Το VGG16 είναι ένα είδος CNN που κυρίως χρησιμοποιείται για την ανάλυση και ταξινόμηση εικόνων. Αποτελείται από 16 συνελκτικά επίπεδα που έχουν παραμέτρους με ικανότητα εκμάθησης (βάρη), από 5 επίπεδα max pooling και από 3 επίπεδα FCN. Με τις συνελίξεις του συντελεί στην εξαγωγή των χαρακτηριστικών από τις εικόνες. Στις συνελίξεις χρησιμοποιεί μικρά φίλτρα μεγέθους 3x3 και συνάρτηση ενεργοποίησης ReLU . Αν και η αρχιτεκτονική του χρησιμοποιείται ευρέως στην υπολογιστή όραση και στην ανάλυση εικόνων, σε προβλήματα σημασιολογικής κατάτμησης δεν έχει την δυνατότητα πρόβλεψης των εικονοστοιχείων μιας εικόνας. Για τον λόγο αυτό χρησιμοποιείται σε συνδυασμό με ειδικές αρχιτεκτονικές όπως είναι το U-NeT, το DEEPLAB, το FCN. Μπορεί να χρησιμοποιηθεί σαν προεκπαιδευμένο μοντέλο στις προαναφερόμενες αρχιτεκτονικές όπως π.χ. στο U-NeT όπου τα επίπεδα του VGG16 μπορούν να χρησιμεύσουν σαν συνδέσεις παραλείψων και να προωθήσουν πληροφορίες μεταξύ κωδικοποιητή και αποκωδικοποιητή [36,37].

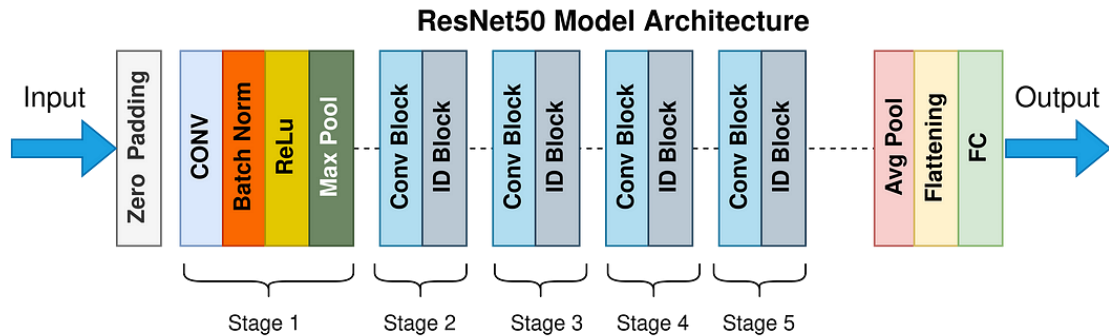


Εικόνα 10:Αρχιτεκτονική VGG16 [36]

4.2.4. ResNet50

Το ResNet50 είναι επίσης ένα είδος CNN, που αναπτύχθηκε από του K.He, X.Zhang, S.Ren και J.Sun το 2015 και αποτελεί και αυτό ένα δημοφιλές μοντέλο για την επίλυση προβλημάτων στο πεδίο της όρασης υπολογιστών. Η καινοτομία που εισάγει στην αρχιτεκτονική του αφορά τις συντομεύσεις (shortcuts) από την είσοδο στην έξοδο σε κάθε επίπεδο συνέλιξης. Αποτελείται από 50 συνελικτικά επίπεδα, συμπεριλαμβανομένων των συνελικτικών φίλτρων, των επιπέδων συγκέντρωσης και των πλήρως συνελικτικών επιπέδων, που οργανώνονται σε μπλοκ και επαναλαμβάνονται πολλές φορές. Το κύριο μπλοκ του ResNet είναι το 'Bottleneck block'. Οι συντομεύσεις δίνουν την δυνατότητα οι πληροφορίες να παρακάμπτουν πολλά συνελικτικά επίπεδα και να φτάνουν γρηγορότερα στα επόμενα. Η χρήση των μπλοκ και των συντομεύσεων βοηθά να αντιμετωπιστεί το πρόβλημα της εξαφάνισης των κλίσεων και βοηθά στην ανάπτυξη δικτύων περισσότερο αποδοτικών.

Μπορεί να χρησιμοποιηθεί σαν προεκπαιδευμένο μοντέλο και η ικανότητα του να εξαγάγει χαρακτηριστικά υψηλού επιπέδου βοηθά στην επίτευξη καλύτερης απόδοσης σε πολλές υπολογιστικές εργασίες. [50]



Εικόνα 11: Αρχιτεκτονική ResNet50

[Αναφορά: [The Annotated ResNet-50. Explaining how ResNet-50 works and why... | by Suvaditya Mukherjee | Towards Data Science](#)]

4.3. Περιβάλλον υλοποίησης

4.3.1. Google Colab

Το Google Colab είναι μια εφαρμογή cloud της Google, διαθέσιμη από το 2017. Είναι μια πλατφόρμα όπου μπορεί να συνταχθεί και να εκτελεστεί κώδικας Python με χρήση σημειωματάρων Jupyter. Οποιοσδήποτε χρήστης του διαδικτύου μπορεί να έχει δωρεάν πρόσβαση σε περιβάλλον Python χωρίς να απαιτείται εγκατάσταση άλλου λογισμικού και να πειραματιστεί στην τεχνητή νοημοσύνη χρησιμοποιώντας μεθόδους μηχανικής μάθησης και κωδικοποίηση.

Ουσιαστικά ο κώδικας εκτελείται σε εικονική μηχανή, που είναι όμως ιδιωτική και συνδέεται με τον λογαριασμό του χρήστη. Οι χρήστες έχουν επίσης την δυνατότητα να αποθηκεύσουν την έρευνα τους και να την μοιράζονται με άλλους. Η εφαρμογή έχει την δυνατότητα συνδυασμού κειμένου, εικόνων, γραφημάτων, HTML και διάφορων άλλων γραφικών. Το Google Colab είναι ένα βολικό εργαλείο για ερευνητές και επιστήμονες που μπορούν να συνδεθούν σε ένα υπολογιστικό περιβάλλον από οποιαδήποτε συσκευή έχει

σύνδεση διαδικτύου. Υποστηρίζει επίσης δημοφιλείς βιβλιοθήκες όπως π.χ. TensorFlow, PyTorch, Pandas, NumPy ώστε να γίνει εύκολο και προσιτό το ξεκίνημα μιας ερευνητικής προσπάθειας.[38]

4.4. Βιβλιοθήκες

4.4.1. TENSORFLOW

Το TensorFlow είναι μια δωρεάν βιβλιοθήκη εκμάθησης ανοικτού κώδικα για τεχνητή νοημοσύνη και μηχανική μάθηση. Χρησιμοποιείται για την δημιουργία και εκπαίδευση μοντέλων μηχανικής μάθησης που αφορούν σε εργασίες ευρέος φάσματος όπως π.χ. η αναγνώριση ομιλίας, η αναγνώριση εικόνας, η επεξεργασία φυσικής γλώσσας κ.α. εστιάζοντας ιδιαίτερα στην εκμάθηση και σε εξαγωγή συμπερασμάτων στα βαθιά νευρωνικά δίκτυα. Βασικό εργαλείο του Tensorflow για την αναπαράσταση και την διαχείριση των δεδομένων αποτελούν οι τανυστές (tensors). Οι τανυστές είναι πολυδιάστατοι πίνακες με μετρήσιμες τιμές, μπορούν να έχουν οποιοσδήποτε διαστάσεις και τύπους δεδομένων και μπορούν να εκτελούνται σε μονάδες επεξεργασίας GPUs και επεξεργαστές TPUs για να επιταχύνονται οι υπολογισμοί. Οι υπολογισμοί έτσι, απεικονίζονται ως υπολογιστικά γραφήματα ροής δεδομένων ώστε μεγάλα σύνολα δεδομένων να μπορούν να επεξεργαστούν με εκτέλεση πολύπλοκων μαθηματικών πράξεων και σχετικών γραφημάτων. Μια σειρά εργαλείων είναι διαθέσιμα επίσης για αξιολόγηση των μοντέλων και για οπτικοποίηση. Επίσης ένα βασικό του πλεονέκτημα είναι ότι διαθέτει προεκπαιδευμένα μοντέλα που μπορούν να αποτελέσουν μέγιστη βοήθεια για τους νέους χρήστες. Βασικά χαρακτηριστικά του αποτελούν :

- Η αυτόματη διαφοροποίηση (Auto Differentiation) όπου υπολογίζεται αυτόματα το διάνυσμα κλίσης του μοντέλου σχετικά με τις παραμέτρους του.
- Η πρόθυμη εκτέλεση , ο έλεγχος δηλαδή σε κάθε βήμα με ένα πρόγραμμα εντοπισμού σφαλμάτων και όχι καθυστερημένα.
- Η διανομή των υπολογισμών σε πολλές συσκευές βάση διαφόρων στρατηγικών.
- Η παροχή συναρτήσεων απώλειας όπως π.χ. η δυαδική διασταυρούμενη εντροπία (BCE), το μέσο τετραγωνικό σφάλμα (MSE).

- Η πρόσβαση σε μετρήσεις όπως η διασταύρωση πάνω από την ένωση (IoU), η ακρίβεια κ.α.
- Η παροχή μιας ενότητας λειτουργιών όπως διαφόρων συνελίξεων (atrous κλπ), συναρτήσεων ενεργοποίησης (softmax κλπ)
- Η παροχή βεστιστοποιητών όπως ο ADAM, ο ADAGRAD κλπ. [39]

4.4.2. KERAS

Το Keras είναι επίσης μια βιβλιοθήκη ανοιχτού κώδικα με γλώσσα προγραμματισμού Python και χρησιμοποιείται στα τεχνητά νευρωνικά δίκτυα. Είναι διεπαφή της TensorFlow και προσφέρει μια σειρά πλεονεκτημάτων όπως ταχύτητα στον εντοπισμό σφαλμάτων, μικρή, εύκολη στην επανάληψη και ευανάγνωστη βάση κώδικα, δυνατότητα ανάπτυξης και συντήρησης κώδικα, παροχή βελτιστοποιήσεων, ανάπτυξη σε πολλαπλές επιφάνειες. [40]

4.4.3. NUMPY

Το NUMPY είναι ένα λογισμικό ανοιχτού κώδικα για την γλώσσα προγραμματισμού Python. Είναι δημοφιλές για τον πολυδιάστατο πίνακα του, που επιτρέπει επεξεργασία και αποθήκευση τεράστιου όγκου δεδομένων, με το πλεονέκτημα της εκτέλεσης διανυσματικών λειτουργιών σε ολόκληρους πίνακες. Περιέχει πλήθος συναρτήσεων που δίνουν την δυνατότητα εκτέλεσης μαθηματικών πράξεων για την επίτευξη αποτελεσματικών υπολογισμών. Είναι απαραίτητο εργαλείο για εργασίες αριθμητικών δεδομένων στην Python.[41]

4.4.4. MATPLOTLIB

Το MATPLOTLIB είναι μια βιβλιοθήκη Python, ανοιχτού κώδικα, που έχει τα απαραίτητα εργαλεία για την οπτικοποίηση των δεδομένων προκειμένου να κατανοηθούν καλύτερα. Για τον σκοπό αυτό παρέχονται μια σειρά γραφημάτων όπως γραμμές, ράβδοι, διαγράμματα διασποράς, ιστογράμματα κ.α. ώστε να δημιουργηθούν απεικονίσεις που μπορεί να είναι στατικές, κινούμενες ή διαδραστικές. Χρησιμοποιείται και σε συνδυασμό με τις υπόλοιπες βιβλιοθήκες Python για να επιτύχει πιο σύνθετες απεικονίσεις και προηγμένες αναλύσεις δεδομένων.[42]

4.4.5. PANDAS

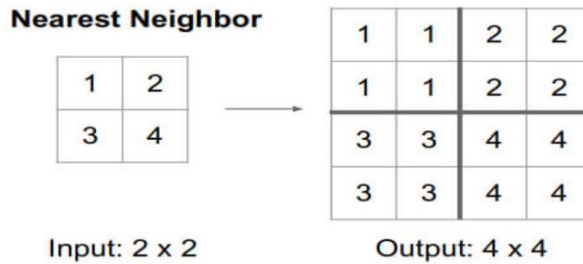
Το PANDAS αποτελεί και αυτό βιβλιοθήκη ανοιχτού κώδικα Python. Χρησιμοποιείται για την ανάλυση των δεδομένων και την επεξεργασία τους, χρησιμοποιώντας χρήσιμα εργαλεία όπως οι πίνακες και οι σειρές. Αποσκοπεί στην εξαγωγή συμπερασμάτων βάσει στατιστικών στοιχείων. Βασικό πλεονέκτημα ο καθαρισμός των δεδομένων με μια σειρά λειτουργιών όπως αντιμετώπιση δεδομένων που λείπουν, που επαναλαμβάνονται, φιλτράρισμα βάσει ορισμένων κριτηρίων σειρών ή στηλών κλπ., και η αποθήκευση τους σε άλλη βάση δεδομένων ή αρχείο.[43]

4.5. Αλγόριθμοι

4.5.1. Πλησιέστεροι γείτονες (nearest neighbors)

Ο αλγόριθμος nearest neighbors χρησιμοποιείται στην σημασιολογική κατάτμηση εικόνων για την αντιστοίχιση κάθε pixel μιας εικόνας σε προκαθορισμένη κατηγορία με βάση το σκεπτικό ότι εικονοστοιχεία που έχουν παρόμοια χαρακτηριστικά ανήκουν στην ίδια κλάση/κατηγορία. Για να εφαρμοστεί η τεχνική αυτή προαπαιτείται η επεξεργασία των δεδομένων εκπαίδευσης ώστε να εξαχθούν τα διάφορα χαρακτηριστικά από τις εικόνες π.χ. γωνίες, χρώμα κ.α. και να δωθούν σε κάθε pixel ετικέτες. Αφού ολοκληρωθεί αυτή η επεξεργασία ο αλγόριθμος βρίσκει με βάση την απόσταση τα γειτονικά pixel από το σύνολο των δεδομένων εκπαίδευσης και τους δίνει την ίδια ετικέτα. Είναι ένας αλγόριθμος εύκολος να εφαρμοστεί, ειδικά όταν το σύνολο εκπαίδευσης δεν είναι πολύ μεγάλο.[44]

Εικόνα 12: Αλγόριθμος Ne



arest Neighbor [15]

4.5.2. Αλγόριθμος Βελτιστοποίησης ADAM

Ο ADAM έχει σκοπό να βελτιστοποιήσει την απόδοση ενός μοντέλου μηχανικής μάθησης με το εναρμονίζει την ταχύτητα εκμάθησης κάθε βάρους του δικτύου (παράμετρος) σύμφωνα με τα ιστορικά στοιχεία τους κατά την φάση της εκπαίδευσης. Χρησιμοποιείται αντί του αλγορίθμου βελτιστοποίησης απότομης καθόδου (gradient descent) και αποτελεί συνδυασμό των αλγορίθμων βελτιστοποίησης momentum και RMSprop συμπληρώνοντας τους με την διόρθωση της αδράνειας (bias-correction) ώστε να βελτιωθεί η σύγκλιση της εκπαίδευσης. Κατά την διάρκεια της εκπαίδευσης υπολογίζει τον μέσο όρο των κινητών τιμών για να εκτιμήσει την 1^η και 2^η ώρα της κλίσης αντίστοιχα και στην συνέχεια υπολογίζει διόρθωση για την αδράνεια των τιμών αυτών. Στο τέλος ενημερώνει σχετικά τις παραμέτρους του μοντέλου.[45]

4.6. Μετρήσεις αξιολόγησης

Οι μετρήσεις που χρησιμοποιούνται στην αξιολόγηση ενός μοντέλου μηχανικής μάθησης δίνουν τα απαιτούμενα στοιχεία για το πόσο καλά έχει λειτουργήσει το μοντέλο, βοηθούν στην ρύθμιση των παραμέτρων του αλλά και στην σύγκριση διαφορετικών μοντέλων. Στα προβλήματα σημασιολογικής κατάτμησης οι κυριότεροι δείκτες μετρήσεων είναι οι ακόλουθοι:

- IoU (Intersection over Union) (6): Ο δείκτης αυτός μετρά σε τι ποσοστό υπάρχει επικάλυψη μεταξύ της πραγματικής και της προβλεπόμενης μάσκας κατάτμησης. Όσο ψηλότερη είναι η τιμή του δείκτη τόσο πιο ακριβής είναι η

κατάτμηση. Ουσιαστικά εκφράζει το ποσοστό της κοινής περιοχής μεταξύ των δύο μασκών.

$$IoU = \frac{TP}{TP+FP+FN} \quad (6)$$

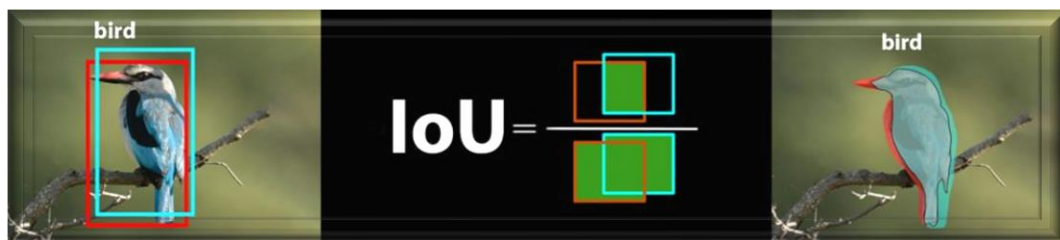
Όπου:

TP: true positive ο αριθμός των πραγματικών θετικών που έχουν εντοπιστεί σωστά

FP: false positive, ο αριθμός των ψευδών θετικών που ανιχνεύονται αλλά δεν είναι πραγματικά

TN: true negative, ο αριθμός των πραγματικών αρνητικών που έχουν εντοπιστεί σωστά

FN: false negative, ο αριθμός των ψευδών αρνητικών, πραγματικών που δεν ανιχνεύονται σωστά



Εικόνα 13: Παράδειγμα μέτρησης IoU

[Αναφορά: <https://learnopencv.com/intersection-over-union-iou-in-object-detection-and-segmentation/>]

- Pixel Accuracy (7): Με τον δείκτη αυτό υπολογίζεται σε τι ποσοστό ανέρχεται η σωστή ταξινόμηση των pixel και εκφράζει το ποσοστό των ορθών προβλέψεων μεταξύ των δύο μασκών.

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (7)$$

- Recall (Ανάκληση) (8): είναι μια μέτρηση που εστιάζει στην ικανότητα του μοντέλου να βρει μεταξύ των πραγματικά θετικών παραδειγμάτων τις σωστές

προβλέψεις. Ουσιαστικά υπολογίζεται σαν ο λόγος του αριθμού των θετικών προβλέψεων προς αυτών των πραγματικών θετικών παραδειγμάτων. Όσο μεγαλύτερη είναι η τιμή τόσο περισσότερο το μοντέλο είναι ικανό να εντοπίζει τις θετικές προβλέψεις.

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

- Precision (Ακρίβεια) (9): είναι ο δείκτης που μετρά την ακρίβεια των θετικών προβλέψεων και υπολογίζεται ως ο λόγος του αριθμού των σωστών θετικών προβλέψεων που μπορεί το μοντέλο να ανιχνεύσει προς τον αριθμό του συνόλου των προβλεπόμενων. Μετρά την αξιοπιστία του μοντέλου όσον αφορά την κατάταξη μιας περιοχής χωρίς ιδιαίτερες λανθασμένες προβλέψεις

$$Precision = \frac{TP}{TP+FP} \quad (9)$$

- F1score (10): αποτελεί ένα συνολικό δείκτη αξιολόγησης και ουσιαστικά ένα συνδυασμό των δύο παραπάνω δεικτών της ανάκλησης και της ακρίβειας με το σκεπτικό ότι για την καλή απόδοση του μοντέλου αυτοί οι δείκτες πρέπει να βρίσκονται σε ισορροπία. [48]

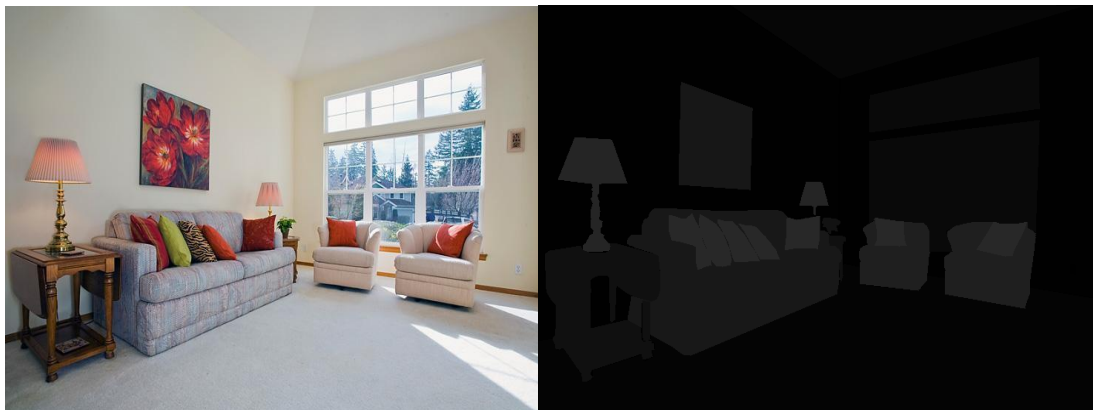
$$F_1 = \frac{2TP}{2TP+FP+FN} \quad (10)$$

4.7. Dataset

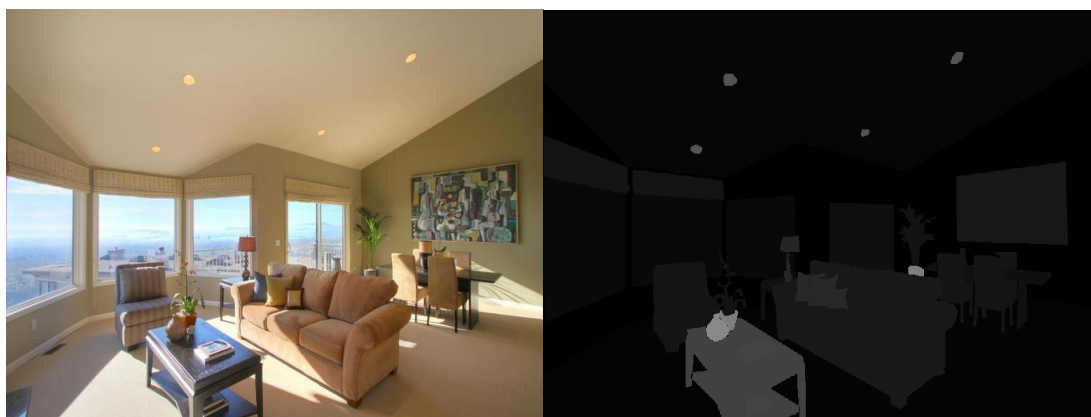
Το αρχείο δεδομένων που χρησιμοποιήθηκε για την υλοποίηση αυτής της εργασίας είναι το ADE20K και κυρίως το υποσύνολο MIT Scene Parsing Benchmark.

Το ADE20K είναι ένα αρχείο που περιλαμβάνει πλούσιες πληροφορίες σχετικές με το περιβάλλον. Τα δεδομένα αποτελούνται από εικόνες με ετικέτες, οι οποίες αφορούν την σημασιολογική περιγραφή ανθρώπων, αντικειμένων, σκηνών και γενικότερα διαφόρων πτυχών που μπορεί να αποτελούν μέρος του περιβάλλοντος. Με το αρχείο αυτό παρέχονται ετικέτες για 150 κατηγορίες διαφόρων αντικειμένων, στις οποίες συμπεριλαμβάνονται οι άνθρωποι, έπιπλα, αυτοκίνητα, συσκευές, δέντρα, δρόμοι, γραφεία κ.α. προκειμένου το

μοντέλο να εκπαιδευτεί να διακρίνει και να αναγνωρίζει τα αντικείμενα που περιέχονται στις εικόνες. Ουσιαστικά αποτελείται από περίπου 20.000 εικόνες και κάθε εικόνα έχει ετικέτα που αντιστοιχεί σε κάθε εικονοστοιχείο της εικόνας και περιγράφει το αντικείμενο ή το περιβάλλον του εικονοστοιχείου. Για να δημιουργηθεί το σύνολο των ετικετών το ADE20K έχει χρησιμοποιήσει 'annotation tools', τους σχολιασμούς δηλαδή που βοηθούν στην επιλογή της ετικέτας και στην αντιστοίχιση με κάθε εικονοστοιχείο σε μορφή png (Portable Network Graphics). Το αρχείο δεδομένων ADE20K χρησιμοποιείται στην σημασιολογική κατάτμηση, στην αναγνώριση αντικειμένων και σκηνών, στην ανάλυση εικόνων κ.α. [49]. Το MIT Scene Parsing Benchmark αποτελείται από μέρος των εικόνων και των ετικετοποιημένων масκών του αρχικού συνόλου και χρησιμοποιείται για την ανάλυση σκηνής και την σημασιολογική κατάτμηση.



Εικόνα 14: Η εικόνα με την αντίστοιχη μάσκα 1



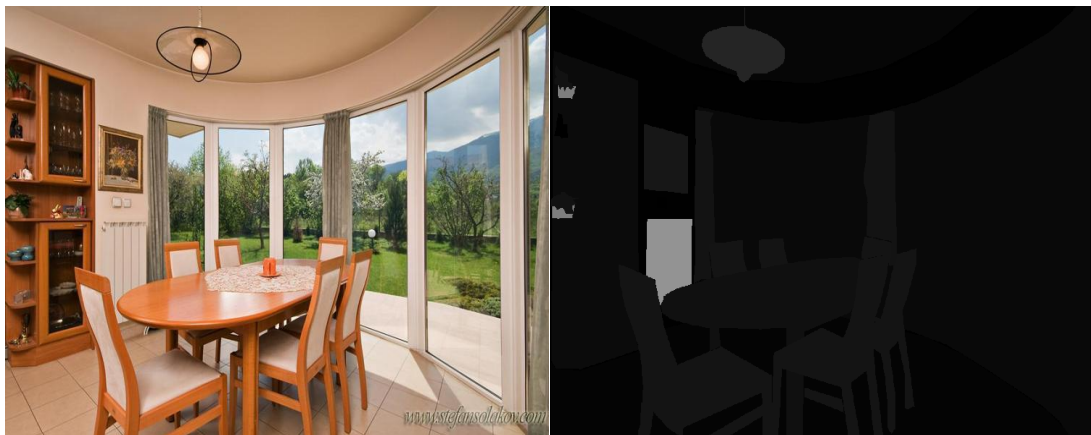
Εικόνα 15: Η εικόνα με την αντίστοιχη μάσκα 2



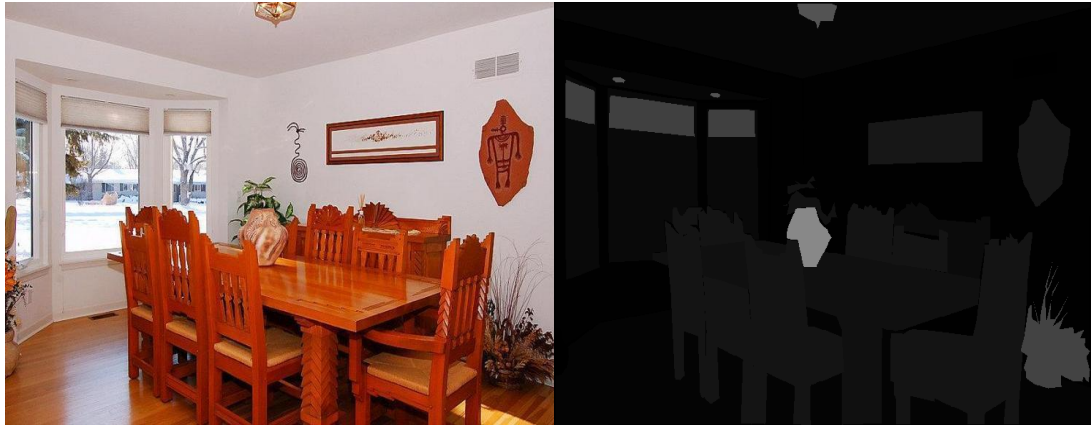
Εικόνα 16: Η εικόνα με την αντίστοιχη μάσκα 3



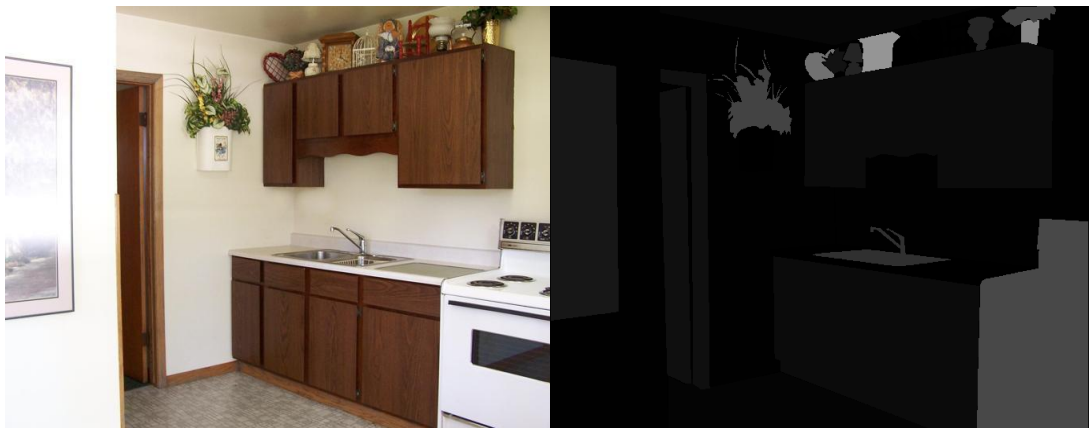
Εικόνα 17: Η εικόνα με την αντίστοιχη μάσκα 4



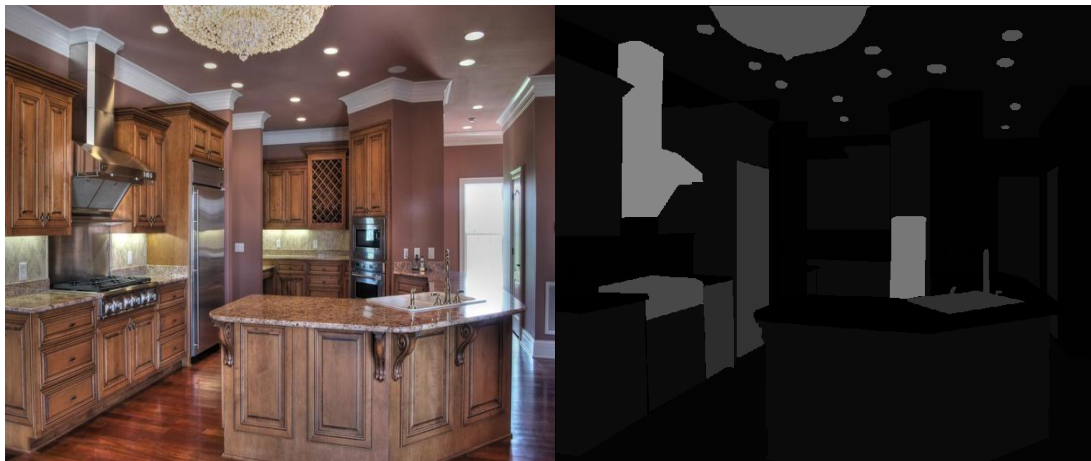
Εικόνα 18: Η εικόνα με την αντίστοιχη μάσκα 5



Εικόνα 19: Η εικόνα με την αντίστοιχη μάσκα 6



Εικόνα 20: Η εικόνα με την αντίστοιχη μάσκα 7



Εικόνα 21: Η εικόνα με την αντίστοιχη μάσκα 8



Εικόνα 22: Η εικόνα με την αντίστοιχη μάσκα 9 [49]

ΚΕΦΑΛΑΙΟ 5

ΥΛΟΠΟΙΗΣΗ

Ο τρόπος και η διαδικασία υλοποίησης της εφαρμογής αποτελεί το περιεχόμενο του παρόντος κεφαλαίου. Πιο αναλυτικά θα παρουσιαστεί η επιλογή των δεδομένων από το αρχείο δεδομένων που χρησιμοποιήθηκε, η εισαγωγή των βιβλιοθηκών, η προετοιμασία και η προεπεξεργασία των δεδομένων, η αρχιτεκτονική των μοντέλων, η εκπαίδευση, τα αποτελέσματα και οι συγκρίσεις της απόδοσης αυτών.

5.1. Επιλογή δεδομένων

Για την συγκεκριμένη εργασία το αρχείο δεδομένων MIT Scene Parsing Benchmark (υποσύνολο του ADE20K) επεξεργάστηκε και ελήφθη ένα υποσύνολο 1807 φωτογραφιών σχετικές με εσωτερικό χώρο για την εκπαίδευση του μοντέλου. Οι 696 εξ' αυτών αφορούν τον χώρο του καθιστικού, 541 από το χώρο της τραπεζαρίας και 570 από υπόλοιπους εσωτερικούς χώρους. Επίσης χρησιμοποιήθηκαν 171 φωτογραφίες για επικύρωση (validation) και 100 φωτογραφίες για δεδομένα ελέγχου από όλους τους χώρους.

5.2. Εισαγωγή βιβλιοθηκών

Ακολουθεί η εισαγωγή των βιβλιοθηκών που απαιτούνται για την εκτέλεση των εργασιών για την επίλυση του προβλήματος της σημασιολογικής κατάτμησης εικόνων όπως αναφέρονται στο κεφάλαιο «Μέσα Υλοποίησης».

```
import os
import fnmatch
import tensorflow as tf
import matplotlib.pyplot as plt
import cv2
import pandas as pd
from tqdm import tqdm
from tabulate import tabulate

#modelo Unet
from keras.applications.vgg16 import VGG16, preprocess_input
from keras.models import Model
from keras.layers import (Activation, BatchNormalization, Conv2D, concatenate, Conv2DTranspose)
#modelo Deeplabv3
from tensorflow import keras
from tensorflow.keras import layers
```

5.3. Προετοιμασία δεδομένων

Για την εισαγωγή των δεδομένων στο μοντέλο αρχικά απαιτείται ο διαχωρισμός τους σε δεδομένα εκπαίδευσης (training), σε δεδομένα επικύρωσης (validation) και δεδομένα ελέγχου (testing). Για τον σκοπό αυτό δημιουργούνται 6 λίστες συνολικά για δεδομένα εκπαίδευσης, επικύρωσης και ελέγχου, 3 διαδρομές των εικόνων με την επέκταση 'jpg' και 3 διαδρομές των масκών με επέκταση 'png' ώστε να εισαχθούν σωστά στο σύστημα.

```
train_folder = r"C:\Users\plaki\Desktop\DATASET\training"
vali_folder = r"C:\Users\plaki\Desktop\DATASET\validation"
test_folder = r"C:\Users\plaki\Desktop\DATASET\testing"

images_folder = os.path.join(train_folder, 'images')
val_images_folder = os.path.join(vali_folder, 'images')
test_image_folder = os.path.join(test_folder, 'images')

masks_folder = os.path.join(train_folder, 'annotation')
val_masks_folder = os.path.join(vali_folder, 'annotation')
test_masks_folder = os.path.join(test_folder, 'annotation')
```

```

#Collect image paths
image_paths = []
for root, dirnames, filenames in os.walk(images_folder):
    for filename in fnmatch.filter(filenames, '*.jpg'):
        path = os.path.join(root, filename)

        image_paths.append(path)

# Collect mask paths
masks_paths = []
for root, dirnames, filenames in os.walk(masks_folder):
    for filename in fnmatch.filter(filenames, '*.png'):
        path = os.path.join(root, filename)

        masks_paths.append(path)

# Collect validation image paths
val_image_paths = []
for root, dirnames, filenames in os.walk(val_images_folder):
    for filename in fnmatch.filter(filenames, '*.jpg'):
        path = os.path.join(root, filename)

        val_image_paths.append(path)

# Collect validation mask paths
val_masks_paths = []
for root, dirnames, filenames in os.walk(val_masks_folder):
    for filename in fnmatch.filter(filenames, '*.png'):
        path = os.path.join(root, filename)

        val_masks_paths.append(path)

test_image_paths = []
for root, dirnames, filenames in os.walk(test_image_folder):
    for filename in fnmatch.filter(filenames, '*.jpg'):
        path = os.path.join(root, filename)

        test_image_paths.append(path)

# Collect validation mask paths
test_masks_paths = []
for root, dirnames, filenames in os.walk(test_masks_folder):
    for filename in fnmatch.filter(filenames, '*.png'):
        path = os.path.join(root, filename)

        test_masks_paths.append(path)

image_paths.sort()
masks_paths.sort()

val_image_paths.sort()
val_masks_paths.sort()
test_image_paths.sort()
test_masks_paths.sort()

```

Δεδομένου ότι το αρχείο δεδομένων MIT Scene Parsing Benchmark.

περιλαμβάνει 150 κατηγορίες από όλες τις πτυχές του περιβάλλοντος, ενώ η εκπαίδευση του μοντέλου επιλέχθηκε να γίνει σε εσωτερικό χώρο, ακολουθεί ο έλεγχος των μοναδικών εικονοστοιχείων που υπάρχουν στο σύνολο των δεδομένων που επιλέχθηκαν για να εντοπιστούν οι κατηγορίες που είναι

παρούσες στις εικόνες ώστε να είναι δυνατή η δημιουργία ενός νέου χάρτη τμηματοποίησης.

```
#understanding data..identify the classes present in the masks
SIZE_X=256
def read_mask(mask_path):
    mask = tf.io.read_file(mask_path)
    mask = tf.image.decode_jpeg(mask, channels=1)

    resized_mask = tf.image.resize(images=mask, size=[SIZE_X, SIZE_X],method='nearest')
    resized_mask = tf.squeeze(resized_mask, axis=-1)

    return resized_mask

unique_values_in_train = set()
unique_values_in_validation = set()

for path in tqdm(masks_paths):
    mask = read_mask(path)
    unique_mask_values, _ = tf.unique(tf.reshape(mask, [-1]))
    unique_values_in_train.update(set(unique_mask_values.numpy()))

for path in tqdm(val_masks_paths):
    mask = read_mask(path)
    unique_mask_values, _ = tf.unique(tf.reshape(mask, [-1]))
    unique_values_in_validation.update(set(unique_mask_values.numpy()))

counter_train = 0
counter_val = 0
list_train = []
list_val = []
for i in range(150):
    if i not in unique_values_in_train:
        list_train.append(i)
        counter_train+=1
    if i not in unique_values_in_validation:
        list_val.append(i)
        counter_val+=1

print("number of classes not contained in my train dataset",counter_train)
print('list with the values not contained', list_train)
print()
print("number of classes not contained in my validation dataset",counter_val)
print('list with the values not contained', list_val)
```

Δημιουργείται λοιπόν ένας νέος χάρτης τμηματοποίησης. Για τις απούσες κατηγορίες γίνεται ομαδοποίηση σε μία κατηγορία η οποία επισημαίνεται με την τιμή '0' ώστε να μην λαμβάνεται υπ' όψη.

5.4. Προεπεξεργασία δεδομένων

Μετά την ολοκλήρωση της προετοιμασίας των δεδομένων, απαιτείται η αποκωδικοποίηση των συνολικά επιλεγέντων δεδομένων σε μορφή τανυστή. Προκειμένου να καθοριστούν τα κανάλια χρώματος στην κλίμακα RGB color (Red, Green, Blue color) επιλέγεται η παράμετρος 'channels=3'. Η παράμετρος 'channels=1' επιλέγεται ώστε το χρώμα των μασκών να δηλωθεί

σε grayscale. Επειδή τα δεδομένα που χρησιμοποιούνται είναι διαφορετικού μεγέθους, η συνάρτηση `resize` αλλάζει το μέγεθος τους σε ένα προκαθορισμένο μέγεθος που στην συγκεκριμένη περίπτωση ορίζεται σε `256x256`.

Στις μάσκες η αλλαγή του μεγέθους πραγματοποιείται με την χρήση του αλγορίθμου `nearest neighbours` ώστε να γίνει σωστή αντιστοίχιση των εικονοστοιχείων με τις ετικέτες. Στην συνέχεια ενημερώνονται οι τιμές των μασκών με βάση τον νέο χάρτη και διαβάζονται εικόνες και μάσκες.

Για τις μάσκες ακολουθείται η τεχνική μετατροπής δεδομένων `'one hot encoding'` που μετατρέπει κάθε μεταβλητή σε δυάνυσμα δυαδικών μεταβλητών όπου μία μόνο έχει τιμή 1 και όλες οι υπόλοιπες 0 ώστε να προκύπτει μια αναπαράσταση των διακριτών κατηγοριών των δεδομένων. Η τεχνική αυτή αποτελείται από τα εξής βήματα:

- Δημιουργία διανύσματος με μηδενικά στοιχεία ίσο με τον αριθμό των κατηγοριών
- Εντοπισμός της θέσης που αντιστοιχεί σε κάθε κατηγορία
- Επισήμανση με τιμή 1 κάθε θέσης που εντοπίστηκε προηγούμενα και με τιμή 0 στις υπόλοιπες θέσεις.

(Η τεχνική αυτή ακολουθείται μόνο στο 1^ο μοντέλο και παραλείπεται στο 2^ο μοντέλο γιατί εκεί χρησιμοποιείται άλλη συνάρτηση απώλειας)

Ακολουθεί η φόρτωση των δεδομένων σε `data generator` τόσο για την εκπαίδευση όσο και για την επικύρωση και η ανακατεύθυνση (`shuffle`) των δεδομένων ώστε να διασφαλίζεται η τυχαία παρουσίαση των δεδομένων κατά την εκπαίδευση για να αποφευχθεί το `'overfitting'`. Στο τέλος της προεπεξεργασίας επιλέγεται ο αριθμός των δειγμάτων που μπορούν να εκπαιδευτούν παράλληλα (`batch`). Το `batch size` αφορά τον αριθμό των δειγμάτων που κατά την εκπαίδευση περνούν μέσα από ένα μοντέλο ταυτόχρονα. Επιδρά στην απόδοση και την μνήμη που απαιτείται στην εκπαίδευση. Ένα μεγαλύτερο `batch size` βελτιώνει την ταχύτητα αλλά μπορεί

να επηρεάσει αρνητικά την απόδοση του μοντέλου. Στο συγκεκριμένο μοντέλο έχει οριστεί σε 4 batch size.

Το σχήμα του τελικού data generator (train_dataset, val_dataset) για το 1^ο

μοντέλο έχει ως εξής :

Images: (4, 256, 256, 3), dtype=tf.float32

Annotation:(4, 256, 256, 105), dtype=tf.float32

Val Dataset:

Images:(4, 256, 256, 3), dtype=tf.float32

Annotation:(4, 256, 256, 105), dtype=tf.float32

Και για το 2^ο **μοντέλο** ως ακολούθως:

Images: (4, 256, 256, 3), dtype=tf.float32

Annotation: (4, 512, 512, 1), dtype=tf.int32

Val Dataset:

Images:(4, 256, 256, 3), dtype=tf.float32

Annotation: (4, 512, 512, 1),dtype=tf.int32

```

def read_image(image_path, mask=False):
    image = tf.io.read_file(image_path)
    if mask:
        image = tf.image.decode_png(image, channels=1)
        image.set_shape([None, None, 1])
        image = tf.image.resize(images=image, size=[IMAGE_SIZE, IMAGE_SIZE],method='nearest')
    else:
        image = tf.image.decode_jpeg(image, channels=3)
        image.set_shape([None, None, 3])
        image = tf.image.resize(images=image, size=[IMAGE_SIZE, IMAGE_SIZE])
        #image = tf.keras.applications.resnet50.preprocess_input(image)
    return image

# Function to update mask values based on the mapping
class_mapping_tensor = tf.constant(list(new_labels.keys()))
mapped_classes_tensor = tf.constant(list(new_labels.values()))
lookup_table = tf.lookup.StaticHashTable(
    tf.lookup.KeyValueTensorInitializer(class_mapping_tensor, mapped_classes_tensor),
    default_value=0
)
def update_mask(mask):
    mask = tf.cast(mask, dtype=tf.int32)
    updated_mask = lookup_table.lookup(mask)
    return updated_mask

def load_data(image_path, mask_path):
    image = read_image(image_path)
    mask = read_image(mask_path,mask=True)
    mask = update_mask(mask)
    return image, mask

def data_generator(image_paths, mask_paths):
    dataset = tf.data.Dataset.from_tensor_slices((image_paths, mask_paths))
    dataset = dataset.shuffle(buffer_size=len(image_paths))
    dataset = dataset.map(load_data, num_parallel_calls=tf.data.AUTOTUNE)
    dataset = dataset.batch(BATCH_SIZE, drop_remainder=True)

    return dataset
train_dataset = data_generator(image_paths, masks_paths)
val_dataset = data_generator(val_image_paths, val_masks_paths)
test_dataset = data_generator(test_image_paths, test_masks_paths)

```

5.5. Αρχιτεκτονική

Για την υλοποίηση της συγκεκριμένης εργασίας επιλέχθηκαν δυο διαφορετικοί συνδυασμοί αρχιτεκτονικών με απώτερο στόχο την σύγκριση και την αξιολόγηση των τελικών αποτελεσμάτων :

5.5.1. UNET – VGG16

Η πρώτη πειραματική προσέγγιση βασίζεται σε μια βελτιωμένη έκδοση αρχιτεκτονικής U-Net για την ανάκτηση πληροφοριών από εικόνες και στη χρήση ενός προ-εκπαιδευμένου μοντέλου VGG16 ως βάση για τον κωδικοποιητή. Η επιλογή του VGG16 βασίζεται στην αξιοσημείωτη απόδοση του σε εργασίες υπολογιστικής όρασης όπως η αναγνώριση αντικειμένων.

Έχει αποδειχθεί εξαιρετικό εργαλείο για την εξαγωγή χρήσιμων χαρακτηριστικών από εικόνες καθώς έχει εκπαιδευτεί σε μεγάλο αριθμό εικόνων από το ImageNet dataset. Τα αρχικά επίπεδα του VGG16 που αποτελούν και το κύριο μέρος του προεκπαιδευμένου μοντέλου πραγματοποιούν 'downsampling' (αποδειγματοληψία), μείωση δηλαδή του μεγέθους των εικόνων με μείωση της ανάλυσης ώστε να αφαιρεθεί πληροφορία από τις αρχικές εικόνες προκειμένου να μειωθεί ο όγκος των δεδομένων. Τα τελευταία επίπεδα πραγματοποιούν 'upsampling' (αναδειγματοληψία) αύξηση δηλαδή του μεγέθους των εικόνων με αύξηση της ανάλυσης ή της συχνότητας δειγματοληψίας ώστε να επαναφερθούν οι χαμένες πληροφορίες από το προηγούμενο στάδιο. Στο συγκεκριμένο μοντέλο τα πρώτα 10 επίπεδα του προεκπαιδευμένου VGG16 έχουν απενεργοποιηθεί για εκπαίδευση ενώ τα επόμενα 5 παραμένουν εκπαιδευσιμα για το νέο πρόβλημα. Για την υλοποίηση του αποκωδικοποιητή επιλέχθηκαν οι έξοδοι από συγκεκριμένα επίπεδα του VGG16, οι 'block 4 out', 'block 3 out' και 'block 2 out' για τον λόγο ότι τα επίπεδα αυτά παρέχουν υψηλές πληροφορίες για την αναπαράσταση των εικόνων και η συνδυαστική τους χρήση βοηθά στην ανάκτηση σημαντικών πληροφοριών ώστε να δημιουργηθεί μια πιο ακριβής ανακατασκευή των εικόνων. Έτσι τα 'block 4 out', 'block 3 out', 'block 2 out' αποθηκεύουν τις εξόδους από το 13°, 10°, και 6° επίπεδο αντίστοιχα και οι έξοδοι αυτοί συνδέονται με τις αντίστοιχες των από κωδικοποιητών του U-Net. Ο κώδικας συνεχίζει με την ανάπτυξη των αντίστοιχων αποκωδικοποιητών με την χρήση συνενώσεων και συνελίξεων με συναρτήσεις ενεργοποίησης ReLu και διαδοχικής παραμόρφωσης (batch normalization). Στο τελικό επίπεδο συνελίξεων χρησιμοποιείται μια συνάρτηση ενεργοποίησης softmax για την παραγωγή των προβλέψεων του μοντέλου. Ουσιαστικά η προσέγγιση της αρχιτεκτονικής που χρησιμοποιείται αποτελεί μια τεχνική της μεταφοράς μάθησης (transfer learning) που, όπως ήδη έχει αναφερθεί στο θεωρητικό μέρος, επιτρέπει να αξιοποιηθούν πληροφορίες από προηγούμενα μοντέλα, με εκπαίδευση σε μεγάλο όγκο δεδομένων, σε νέα πλέον προβλήματα, ώστε να επιτευχθούν καλύτερες επιδόσεις κατά την εκπαίδευση του νέου μοντέλου.

5.5.2 DeepLabV3+ – ResNet50

Η δεύτερη πειραματική προσέγγιση βασίζεται στο μοντέλο DeepLabV3+ σε συνδυασμό με το προεκπαιδευμένο μοντέλο ResNet50. [40]

Ο κώδικας περιλαμβάνει τις εξής βασικές συναρτήσεις:

- Συνάρτηση 'convolution_block': Η συνάρτηση αυτή δημιουργεί ένα μπλοκ συνελικτικό, που αποτελείται από ένα επίπεδο συνελικτικό, μια κανονικοποίηση παρτίδας (batch normalization) και μια συνάρτηση ενεργοποίησης ReLu.
- Συνάρτηση 'DilatedSpatialPyramidPooling': η τεχνική αυτή χρησιμοποιείται για να εξαχθούν αποτελεσματικά, τα χαρακτηριστικά από τις εικόνες σε διάφορες κλίμακες. Η είσοδος εδώ υποβάλλεται σε μια σειρά συνελικτικών μπλοκ και η έξοδος από αυτά συγχωνεύεται σε ένα επίπεδο συνένωσης (concatenation).
- Συνάρτηση 'DeeplabV3Plus': Στη συνέχεια, ορίζεται η συνάρτηση DeeplabV3Plus, που δέχεται το μέγεθος εικόνας (image_size) και τον αριθμό των κλάσεων (num_classes) ως είσοδο. Αρχικά, δημιουργείται μία είσοδος μοντέλου (model_input) με τις διαστάσεις του εικονοστοιχείου. Στη συνέχεια, δημιουργείται ένα μοντέλο ResNet50, με βάση το προ-εκπαιδευμένο δίκτυο ResNet50, που δέχεται το model_input ως είσοδο. Από αυτό το μοντέλο, λαμβάνεται η έξοδος του επιπέδου "conv4_block6_2_relu" και αποθηκεύεται στη μεταβλητή x. Στη συνέχεια, η μεταβλητή x περνά από τη συνάρτηση DilatedSpatialPyramidPooling για την εξαγωγή χαρακτηριστικών. Έπειτα, δημιουργούνται δύο εισόδοι input_a και input_b, οι οποίοι υπολογίζονται με βάση την x και τα επίπεδα του μοντέλου ResNet50. Το input_a περνά από ένα επίπεδο UpSampling2D για να έχει τις ίδιες διαστάσεις με την x, ενώ το input_b περνά από μια συνάρτηση convolution_block με παραμέτρους num_filters=48 και kernel_size=1. Έπειτα, οι μεταβλητές input_a και input_b συνενώνονται με τη χρήση της συνάρτησης Concatenate και ακολουθούν επιπλέον επίπεδα convolution_block. Τέλος, οι τελικές εξόδοι περνούν μέσω επιπλέον επιπέδων UpSampling2D και τελικά οδηγούν σε ένα τελικό επίπεδο Conv2D που παράγει την έξοδο του μοντέλου, η οποία έχει τις ίδιες διαστάσεις με την είσοδο εικόνας και αριθμό καναλιών ίσο με τον αριθμό των κλάσεων (num_classes).

5.6. Εκπαίδευση

Για την εκπαίδευση των δεδομένων απαιτείται αρχικά ο ορισμός κάποιων παραμέτρων που είναι βασικές για το στάδιο αυτό. Οι παράμετροι αυτές βοηθούν στην εκπαίδευση και στην αξιολόγηση του μοντέλου. Η ρύθμιση τους επηρεάζει την ακρίβεια και την ταχύτητα της εκπαίδευσης, για το λόγο αυτό είναι σημαντική η σωστή επιλογή για να εμφανιστούν τα επιθυμητά αποτελέσματα.

Κοινά και για τα 2 μοντέλα:

- Epoch: Η παράμετρος αυτή αναφέρεται στο γεγονός ότι το μοντέλο έχει εκπαιδευτεί να δει τα δεδομένα και να τα επεξεργαστεί μια φορά. Κατά την διάρκεια του epoch μετρείται η απόδοση του μοντέλου και με την ολοκλήρωση της το μοντέλο αναβαθμίζεται με βάση τον αλγόριθμο εκπαίδευσης που χρησιμοποιείται, ώστε να βελτιωθεί. Η επιλογή του αριθμού επηρεάζει και την απόδοση του μοντέλου καθώς θα πρέπει να επιτευχθεί μια ισορροπία ώστε τόσο η εκπαίδευση να είναι επαρκής αλλά και να μην προκληθεί overfitting. Στο πρώτο μοντέλο ο αριθμός των epoch ορίστηκε σε 25 ενώ στο δεύτερο σε 20.
- Learning Rate (ρυθμός εκμάθησης): Ο ρυθμός εκμάθησης μπορεί να ορίσει την ταχύτητα προσαρμογής των βαρών κατά την εκπαίδευση. Ένα μικρότερο learning rate μπορεί να είναι καλύτερη επιλογή για τον λόγο ότι μπορεί να επιτρέπει μια πιο αργή προσέγγιση στην καλύτερη απόδοση, ωστόσο βοηθά στην επίτευξη καλύτερης ακρίβειας. Εδώ, η τιμή του LR ορίζεται ως 0.001.
- Optimizer (Βελτιστοποιητής): Ο βελτιστοποιητής Adam για τον οποίο πληροφορίες αναφέρθηκαν στο κεφάλαιο των μέσων υλοποίησης χρησιμοποιείται και στα δύο μοντέλα.
- Accuracy (Ακρίβεια): Αποτελεί επίσης ένα μέτρο αξιολόγησης κατά την εκπαίδευση και αφορά το ποσοστό των σωστών προβλέψεων που πραγματοποιούνται σε σχέση με το σύνολο των προβλέψεων.

Για το 1^ο μοντέλο (UNET):

- Dice Loss (11): Η συνάρτηση απώλειας Dice χρησιμοποιείται στο πρώτο μοντέλο κυρίως γιατί έχει καλά αποτελέσματα σε προβλήματα σημασιολογικής

κατάτμησης εικόνας. Με την dice loss υπολογίζεται το κόστος και ενημερώνονται τα βάρη του μοντέλου. Ο υπολογισμός του Dice score αφορά την απόκλιση που παρατηρείται μεταξύ των πραγματικών συνόλων εικονοστοιχείων με τα προβλεπόμενα με στόχο την ελαχιστοποίηση της συνάρτησης, ώστε να επιτευχθεί η καλύτερη πρόβλεψη στην κατάτμηση αντικειμένων. Η συνάρτηση dice loss είναι αντίθετη του συντελεστή dice και επιδιώκει την μεγιστοποίηση του συντελεστή και αντίστοιχα την ελαχιστοποίηση της συνάρτησης. [46]

$$Loss_{DL} = 1 - 2 \frac{\sum_{i \in L} \sum_{i \in N} y_i^{(l)} \hat{y}_j^{(l)} + \epsilon}{\sum_{i \in L} \sum_{i \in N} (y_i^{(l)} + \hat{y}_i^{(l)}) + \epsilon} \quad (11)$$

όπου $y_i^{(l)}$: προβλεπόμενες τιμές

$y_j^{(l)}$: πραγματικές τιμές

ϵ : μαθηματική σταθερά

- Dice Coefficient (12): Η συνάρτηση Dice Coefficient χρησιμοποιείται για την αξιολόγηση της απόδοσης του μοντέλου κατά την εκπαίδευση. Με την συνάρτηση αυτή υπολογίζεται ο συντελεστής Dice score που μετρά την ομοιότητα μεταξύ των μασκών κατάτμησης, πραγματικής και προβλεπόμενης. Όσο μεγαλύτερος είναι ο δείκτης τόσο πιο ακριβές το μοντέλο.[47]

$$Dicecoef = 2 \frac{\sum_{i \in L} \sum_{i \in N} y_i^{(l)} \hat{y}_j^{(l)} + \epsilon}{\sum_{i \in L} \sum_{i \in N} (y_i^{(l)} + \hat{y}_i^{(l)}) + \epsilon} \quad (12)$$

όπου $y_i^{(l)}$: προβλεπόμενες τιμές

$y_j^{(l)}$: πραγματικές τιμές

ϵ : μαθηματική σταθερά

```

def dice_loss(y_true, y_pred, epsilon=1e-6):
    # skip the batch and class axis for calculating Dice score
    axes = tuple(range(1, len(y_pred.shape)-1))
    numerator = 2. * tf.reduce_sum(y_pred * y_true, axes)
    denominator = tf.reduce_sum(tf.square(y_pred) + tf.square(y_true), axes)

    return 1 - tf.reduce_mean((numerator + epsilon) / (denominator + epsilon)) # average over classes

def dice_coef(y_true, y_pred, epsilon=1e-6):
    axes = tuple(range(1, len(y_pred.shape)-1))
    numerator = 2. * tf.reduce_sum(y_pred * y_true, axes)
    denominator = tf.reduce_sum(tf.square(y_pred) + tf.square(y_true), axes)

    dice_scores = (numerator + epsilon) / (denominator + epsilon)

    return dice_scores

LR=0.001
opt = tf.keras.optimizers.Adam(LR)
metrics = [dice_coef, 'accuracy']
loss = dice_loss

```

Για το 2^ο μοντέλο (DeepLabV3+):

- Sparse Categorical Crossentropy Loss (13): Η συνάρτηση απώλειας sparse categorical crossentropy χρησιμοποιείται στο δεύτερο μοντέλο και υπολογίζεται με βάση τον προσδιορισμό που αναμένεται (ground truth) και τις προβλέψεις του μοντέλου. Αντί για μια μεταβλητή εξόδου σε κάθε κατηγορία όπως συμβαίνει στην απώλεια crossentropy, χρησιμοποιείται μια ακέραια μεταβλητή που αναπαριστά την ετικέτα της πραγματικής κατηγορίας. Ορίζεται ως :

$$Loss_i = -\sum_{i \in N} \sum_{l \in L} w_i y_i^{(l)} \log \hat{y}_i^{(l)} \quad (13)$$

όπου i : το δείγμα

l : κατηγορία επιλογής

$y_i^{(l)}$: πραγματικό αποτέλεσμα (ground truth)

$\hat{y}_i^{(l)}$: πρόβλεψη μοντέλου

w_i : βάρη

Στην περίπτωση που η ακέραιη ετικέτα αντιστοιχεί στην πραγματική κατηγορία, αυξάνεται η απώλεια. Όταν δεν αντιστοιχεί, η απώλεια μένει χαμηλή. [51]

```
loss = keras.losses.SparseCategoricalCrossentropy(from_logits=True)
LR=0.001
opt = tf.keras.optimizers.Adam(LR)
metrics=['accuracy']
```

Οι εντολές 'compile' και 'fit' είναι από τις πιο σημαντικές μεθόδους στο εκπαιδευτικό κύκλο της μηχανικής μάθησης. Η 'compile' ορίζει την αρχιτεκτονική του μοντέλου και το προετοιμάζει για την εκπαίδευση. Περιλαμβάνει τρεις σημαντικούς ορισμούς, τον βελτιστοποιητή (optimizer), την συνάρτηση απώλειας (loss) και τις μετρικές αξιολόγησης (metrics). Η 'fit' χρησιμοποιείται για την εκπαίδευση του μοντέλου με τα δεδομένα εκπαίδευσης. Περιλαμβάνει τους εξής ορισμούς: 'train dataset' το σύνολο δεδομένων εκπαίδευσης, 'validation data' το σύνολο δεδομένων επικύρωσης και το 'epochs' τον αριθμό των εποχών.

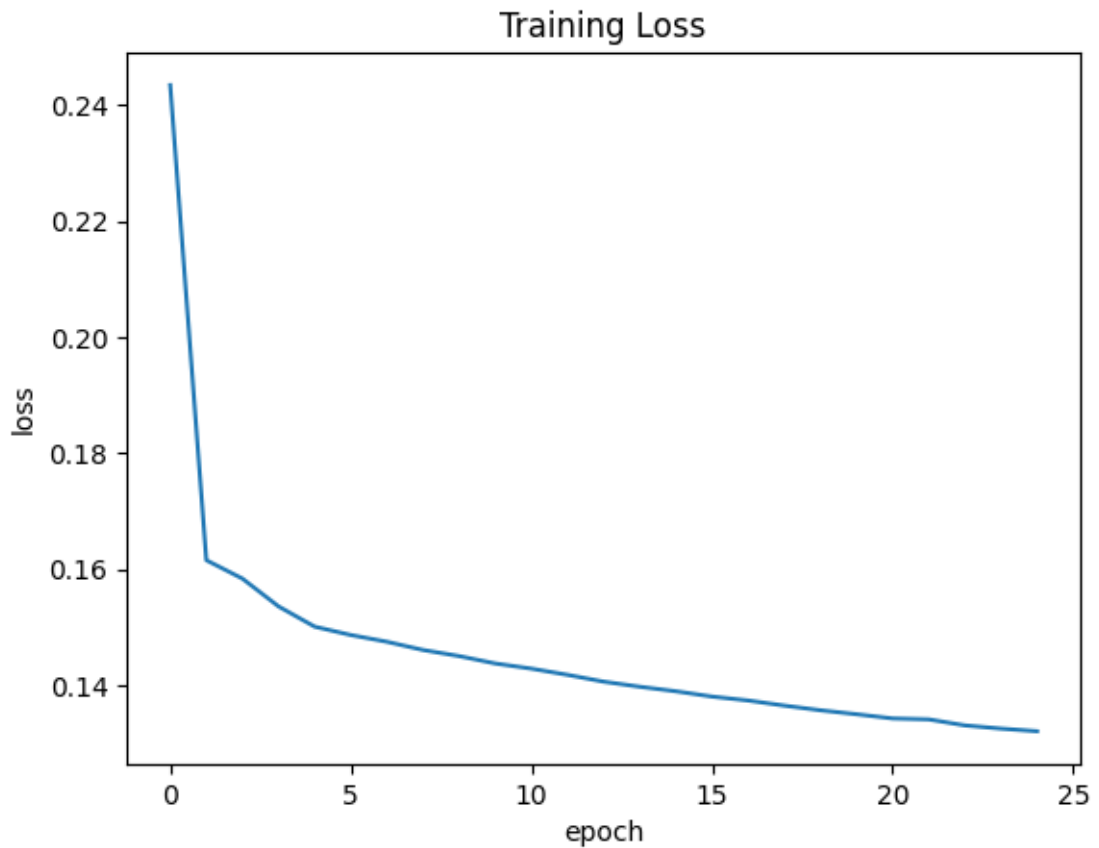
```
model.compile(optimizer=opt,
              loss = loss,
              metrics=metrics)
history = model.fit(train_dataset, validation_data=val_dataset, epochs=EPOCHS)
```

5.7. Αποτελέσματα

Μετά το τέλος της εκπαίδευσης, είναι δυνατή η παρουσίαση των αποτελεσμάτων για κάθε μοντέλο ξεχωριστά :

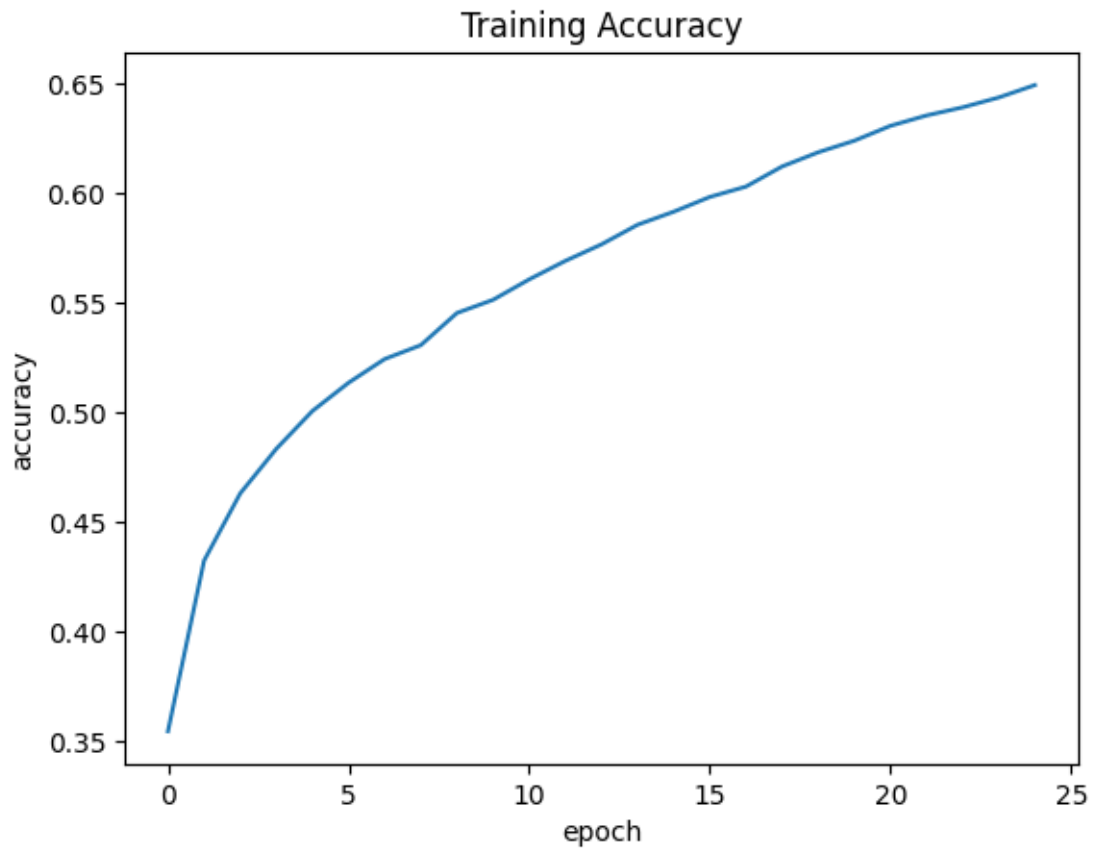
5.7.1. Αποτελέσματα 1^{ου} μοντέλου:

Τα αποτελέσματα παρουσιάζονται μέσω των ακόλουθων γραφικών παραστάσεων:



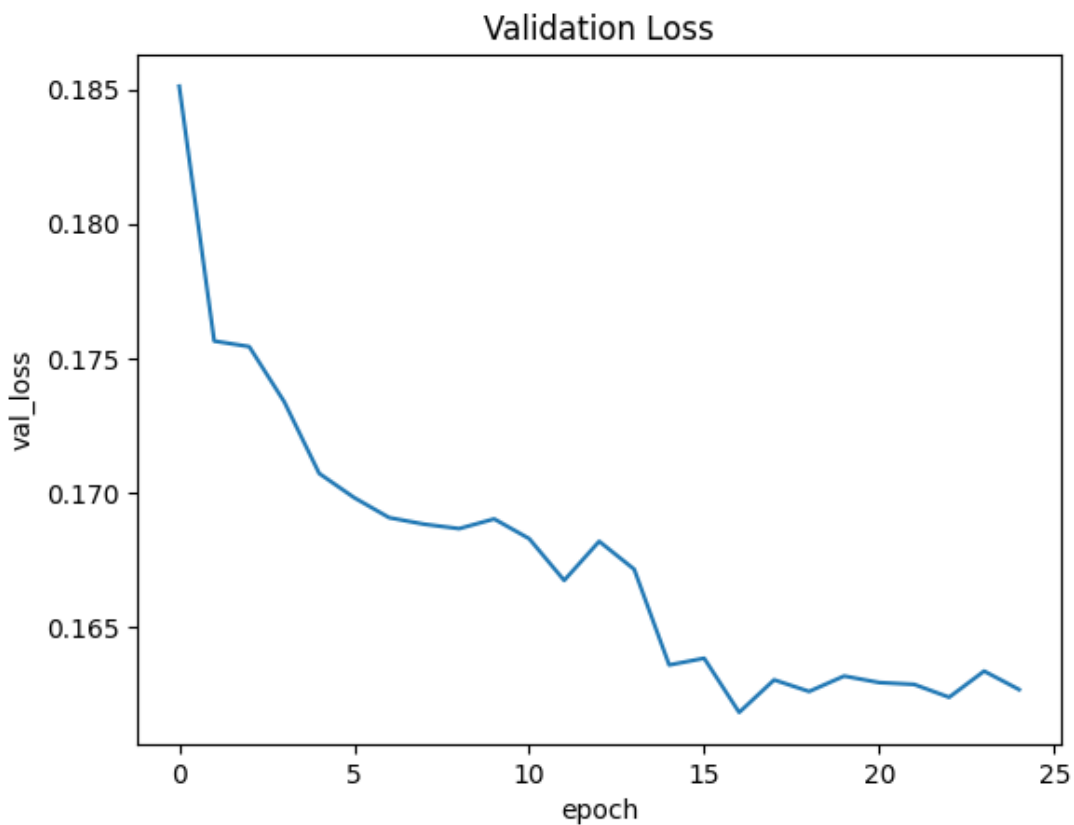
Γράφημα 4: Training loss

Το διάγραμμα 'training loss' εκφράζει την μεταβολή της απώλειας (loss) σε όλη την διάρκεια της εκπαίδευσης. Από το διάγραμμα παρατηρείται η σταδιακή μείωση της απώλειας όσο επαναλαμβάνονται οι εποχές, γεγονός που υποδεικνύει την προσαρμογή του μοντέλου στα δεδομένα εκπαίδευσης και στην βελτίωση της ακρίβειας των προβλέψεων του.



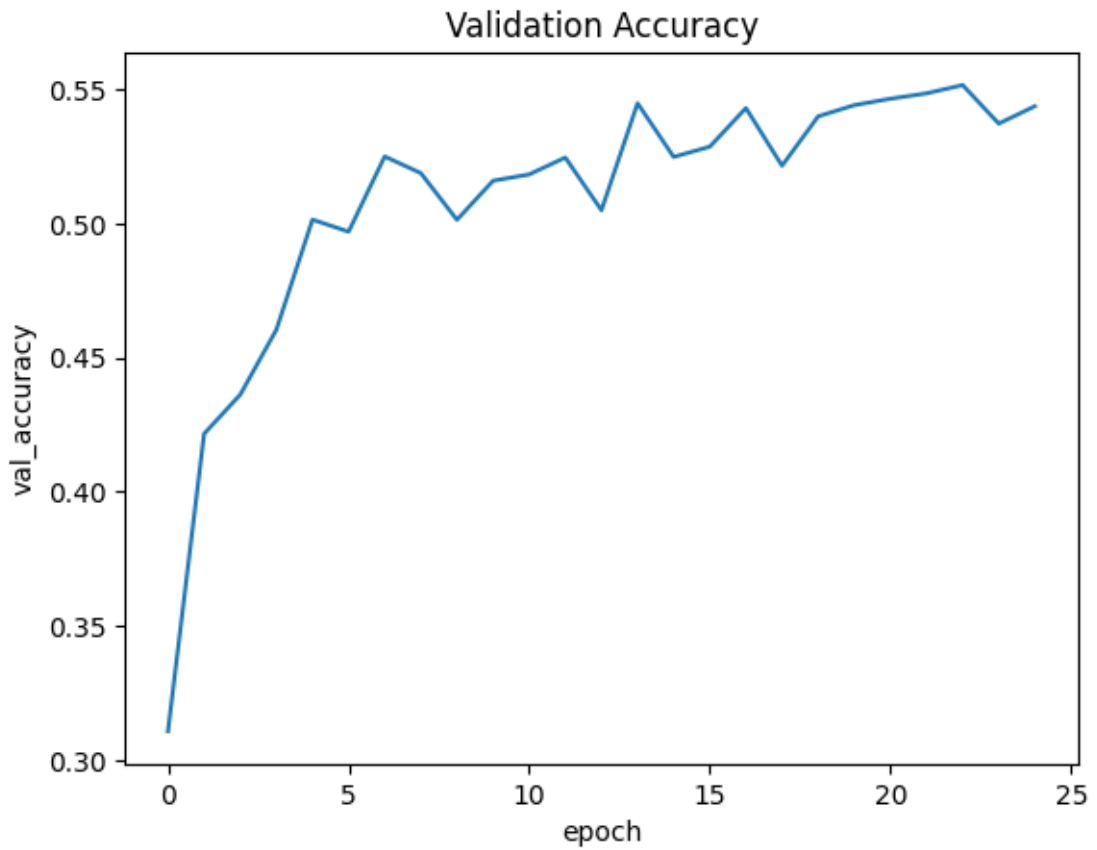
Γράφημα 5: Training Accuracy για το μοντέλο 1(Unet)

Το διάγραμμα 'Training Accuracy' εκφράζει την μεταβολή της ακρίβειας κατά την εκπαίδευση. Η ακρίβεια αυξάνεται όσο αυξάνονται οι εποχές βελτιώνοντας την απόδοση του μοντέλου καθώς οι προβλέψεις γίνονται πιο σωστές όσο εκπαιδεύεται.



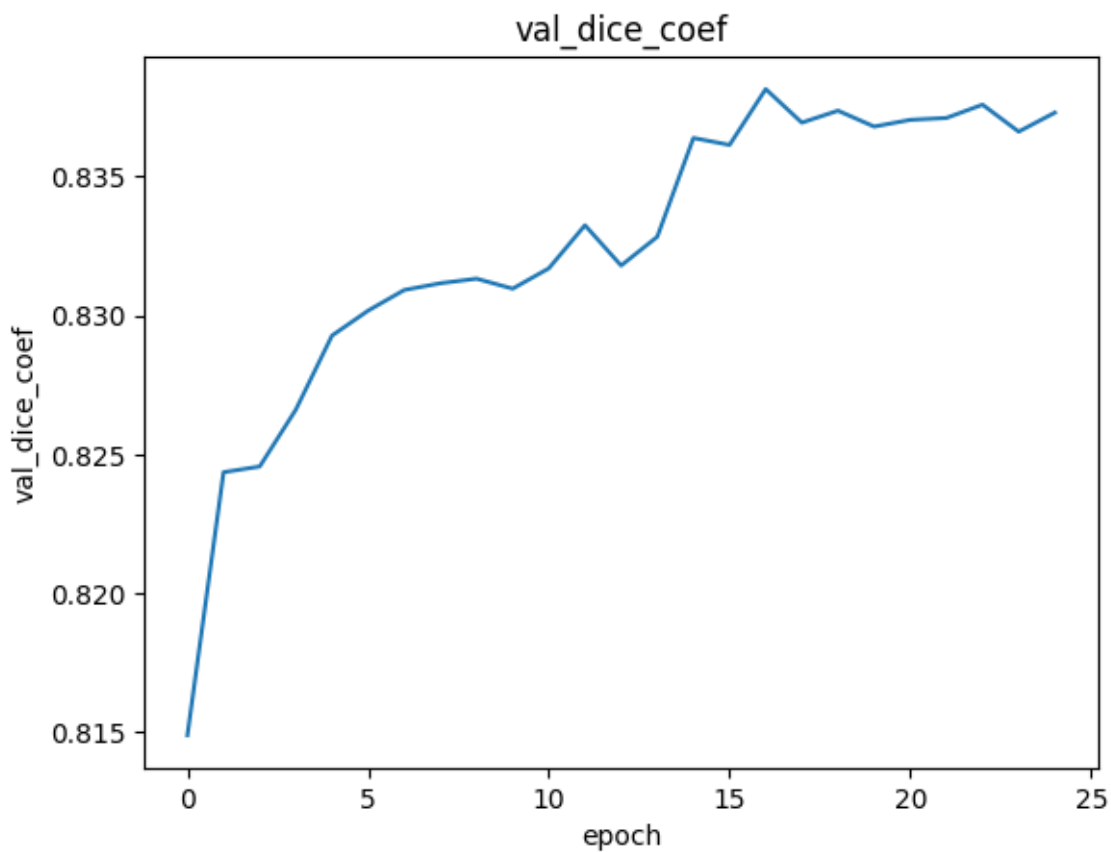
Γράφημα 6: Valodation Loss για το μοντέλο 1(Unet)

Το διάγραμμα 'Validation Loss' δείχνει την μεταβολή της απώλειας (loss) στο σύνολο επικύρωσης κατά την εκπαίδευση. Η μείωση που παρατηρείται υποδεικνύει ότι το μοντέλο συμπεριφέρεται καλά σε νέα δεδομένα και δεν παθαίνει υπερπροσαρμογή μέχρι αυτό το σημείο. Σε περίπτωση όμως που η εκπαίδευση συνεχιστεί σε περισσότερες εποχές είναι πολύ πιθανό να παρατηρηθεί το φαινόμενο του 'overfitting'.



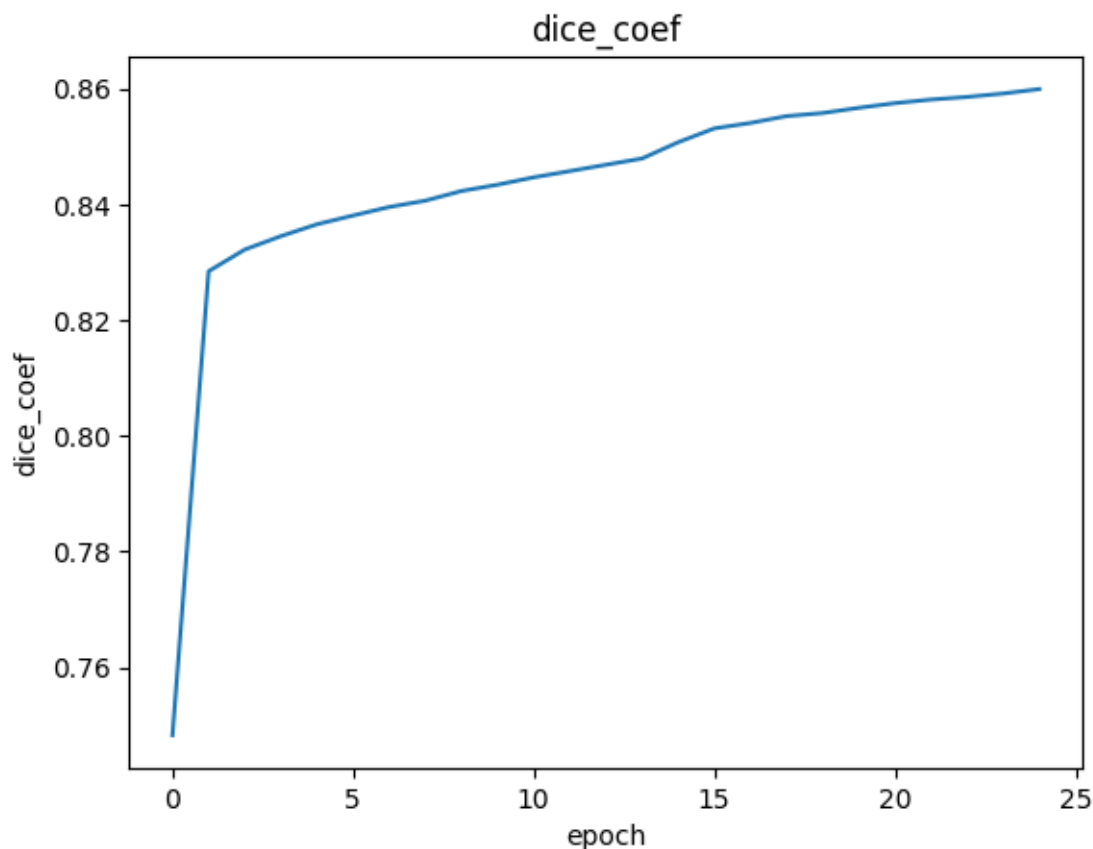
Γράφημα 7: Validation accuracy για το μοντέλο 1(U-net)

Το διάγραμμα 'Validation Accuracy' δείχνει την μεταβολή της ακρίβειας στο σύνολο επικύρωσης κατά την εκπαίδευση. Η αύξηση που παρατηρείται υποδεικνύει ότι βελτιώνεται η απόδοση του μοντέλου σε νέα δεδομένα.



Γράφημα 8: Validation dice coefficient για το μοντέλο 1(Unet)

Το διάγραμμα 'val dice coef' απεικονίζει την μέτρηση 'dice coefficient' στο σύνολο επικύρωσης κατά την εκπαίδευση. Η αύξηση δείχνει ότι παράγονται προβλέψεις πιο ακριβείς .



Γράφημα 9: dice coefficient για το μοντέλο 1(Unet)

Το διάγραμμα 'dice coef' απεικονίζει την μέτρηση 'dice coefficient' στο σύνολο της εκπαίδευσης, κατά την εκπαίδευση. Η αύξηση δείχνει την βελτίωση της ικανότητας του μοντέλου για αναγνώριση και παραγωγή ακριβών προβλέψεων.

Από την συνολική παρατήρηση των διαγραμμάτων φαίνεται ότι το μοντέλο βελτιώνει την απόδοση του όσο αυξάνονται οι εποχές εκπαίδευσης. Η μείωση της απώλειας (loss) και αντίθετα η αύξηση της ακρίβειας και των μετρήσεων 'dice coefficient' δείχνουν την εκμάθηση του μοντέλου και την προσαρμογή του τόσο στα δεδομένα εκπαίδευσης όσο και επικύρωσης με αποτέλεσμα την βελτίωση της ικανότητας πρόβλεψης.

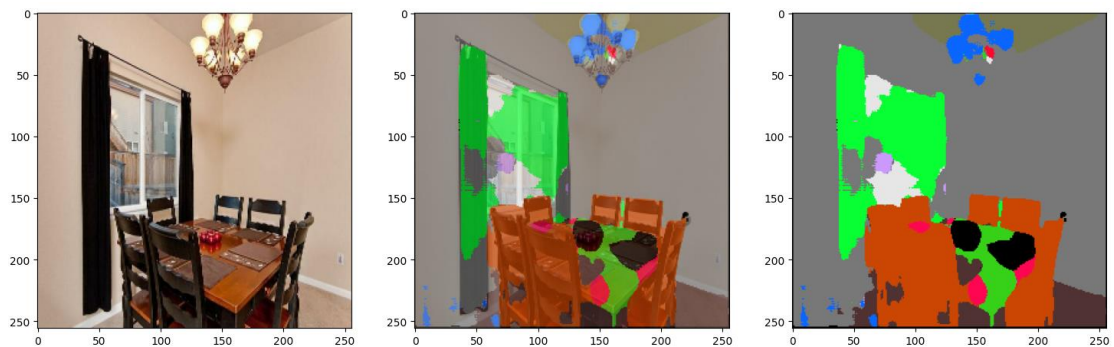
Για να οπτικοποιηθούν τα αποτελέσματα σχεδιάζονται ως μάσκες κατάτμησης RGB, ώστε κάθε εικονοστοιχείο που αντιστοιχεί σε κάθε μία από τις κατηγορίες επισημαίνεται με ένα μοναδικό χρώμα, όπως φαίνεται από την

ανάλυση του πίνακα που εμφανίζεται παρακάτω. Επίσης δημιουργήθηκε ένα csv αρχείο όπου στην κάθε κατηγορία προστέθηκε το όνομα της επικέτας.

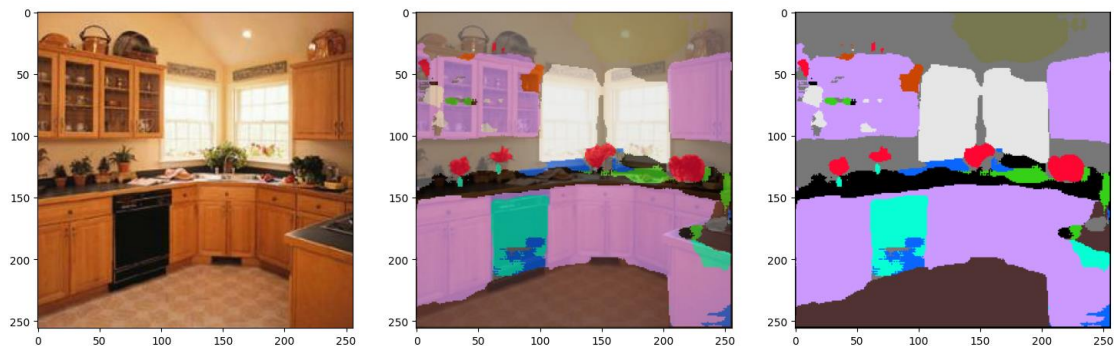
wall	floor	ceiling	window	cabinet	
table	chair	painting	sofa	lamp	
smalltable					
door	curtain	cushion	sink	fireplace	refrigerator
flower	stove	kitchen	chandelier	oven	microwave
dishwasher					

Πίνακας 1: Χρωμάτων

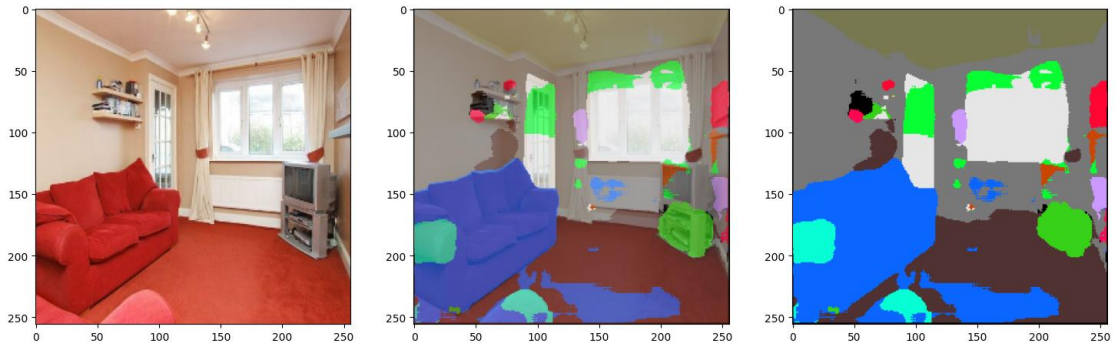
Οι εικόνες που παρατίθενται παρακάτω, σε οριζόντια σειρά αφορούν πρώτα την πραγματική εικόνα, στην συνέχεια την μάσκα κατάμησης και τέλος την πρόβλεψη.



Εικόνα 23: Πρόβλεψη εικόνας από το Μοντέλο 1(Unet)



Εικόνα 24: Πρόβλεψη εικόνας από το Μοντέλο 1(Unet)

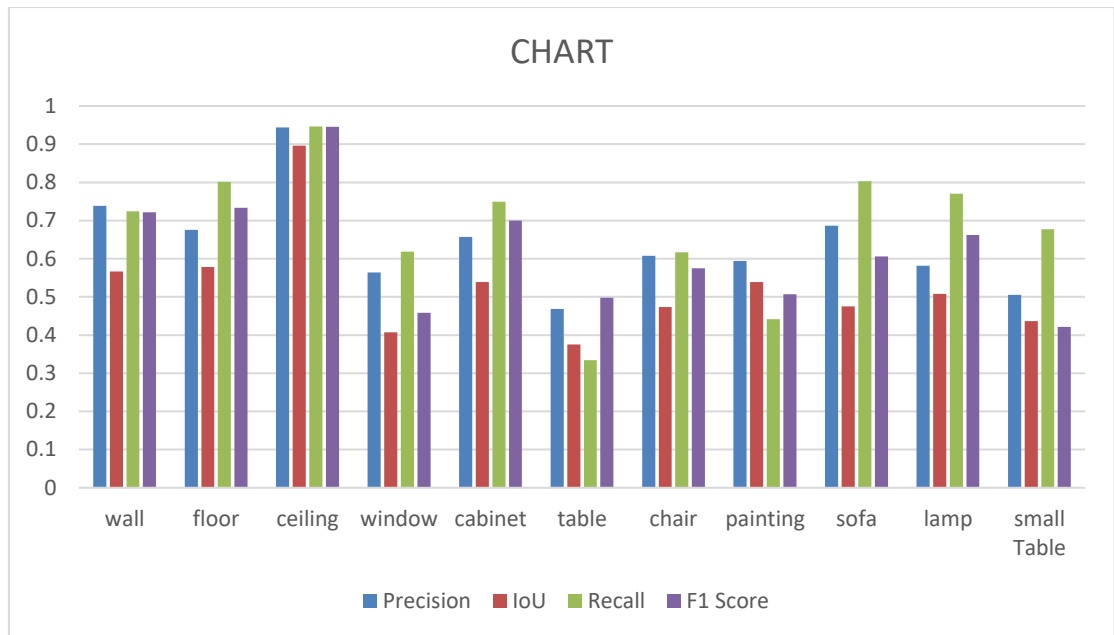


Εικόνα 25: Πρόβλεψη εικόνας από το Μοντέλο 1(Unet)

Παρατηρείται οπτικά ότι προβλέπεται με ευκρίνεια το μεγαλύτερο μέρος της εικόνας, ωστόσο τα οπτικά αποτελέσματα θα πρέπει να συνδυαστούν με τις μετρήσεις που έγιναν ώστε να είναι δυνατή η αξιολόγηση της απόδοσης του μοντέλου.

Class Name	Precision	IoU	Recall	F1 Score
Wall	0,738426	0,566821	0,724251	0,721326
Floor	0,675446	0,578519	0,801252	0,732989
ceiling	0,944133	0,896343	0,946547	0,945338
window	0,56422	0,407496	0,618891	0,458569
cabinet	0,657221	0,538775	0,749342	0,700264
table	0,468806	0,375103	0,334361	0,498021
chair	0,60756	0,473155	0,617217	0,57464
painting	0,594473	0,539242	0,441387	0,506617
Sofa	0,686596	0,474767	0,803218	0,606045
lamp	0,58171	0,508251	0,770165	0,662116
small Table	0,505693	0,43677	0,676914	0,421181

Πίνακας 2: Μέσος όρος αποτελεσμάτων για κάθε κατηγορία για το μοντέλο 1(Unet)



Διάγραμμα 1: Μέσος όρος αποτελεσμάτων για το μοντέλο 1 (Unet)

- Από τις μετρήσεις για κάθε συγκεκριμένη κατηγορία μπορούν να σχολιαστούν τα παρακάτω:

Από την μέτρηση της ακρίβειας, προκύπτει υψηλή ακρίβεια για την κατηγορία 'ceiling', ενώ χαμηλότερη ακρίβεια έχουν οι κατηγορίες 'table' και 'small table', ενώ οι υπόλοιπες κατηγορίες κινούνται σε ενδιάμεσα επίπεδα

Από την επικάλυψη (IoU) παρατηρείται ότι υπάρχει υψηλή επικάλυψη στην κατηγορία 'ceiling' με τις κατηγορίες 'window' και 'table' να έχουν χαμηλότερη. Από την ανάκληση (recall) προκύπτει ότι το μοντέλο παρουσιάζει υψηλή ανάκληση στις κατηγορίες 'ceiling', 'sofa' και 'floor' ενώ οι κατηγορίες 'table' και 'painting' έχουν χαμηλότερη ανάκληση.

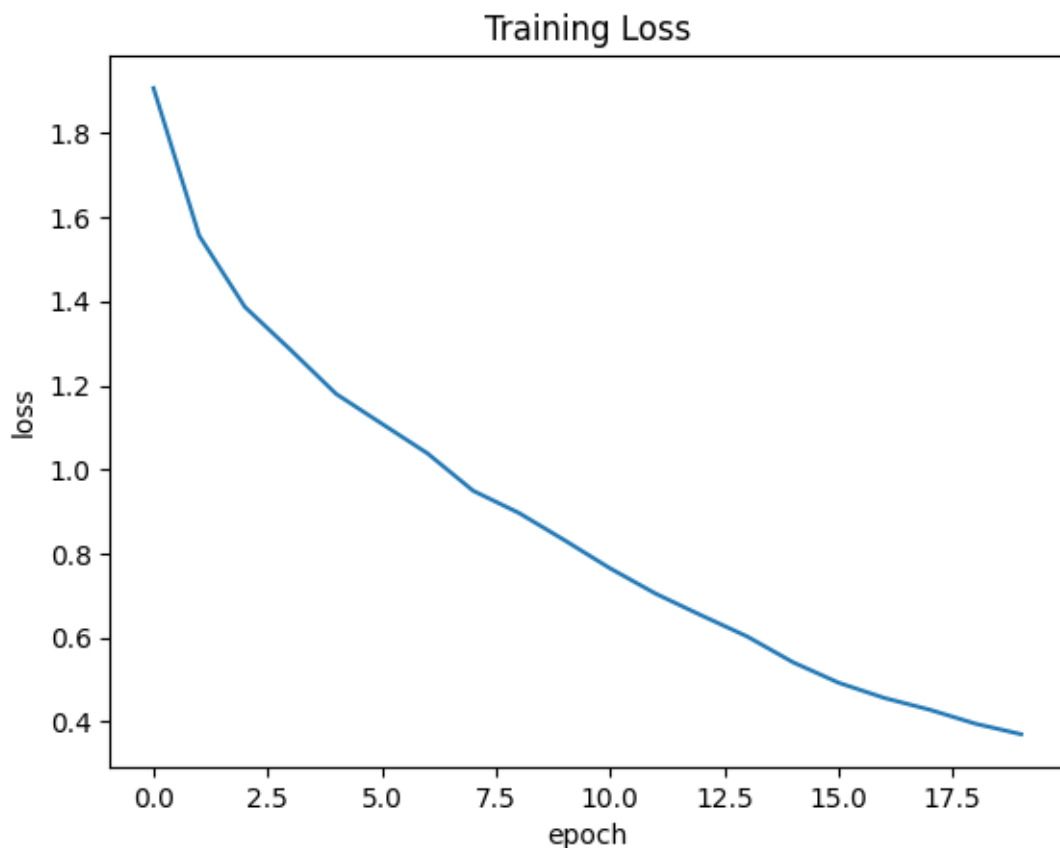
Από το F1score παρατηρείται ότι υψηλό F1score έχει η κατηγορία 'ceiling' και το χαμηλότερο οι κατηγορίες 'small table' και 'window'.

Γενικότερα παρατηρείται ότι η κατηγορία 'ceiling' παρουσιάζει εξαιρετική απόδοση σε όλες τις μετρήσεις, οι κατηγορίες 'wall' και 'floor' τα αποτελέσματα είναι αρκετά ικανοποιητικά στις μετρήσεις 'precision', 'recall' και 'F1score', ωστόσο με μέτρια αποτελέσματα στην επικάλυψη 'IoU', σε αντίθεση με τις υπόλοιπες κατηγορίες που τα αποτελέσματα είναι λιγότερο ικανοποιητικά και οι μετρήσεις παρουσιάζουν ένα ευρύ φάσμα τιμών, άλλες κοντά στο ικανοποιητικό επίπεδο και άλλες στο πιο μέτριο ή χαμηλό, γεγονός

που υποδεικνύει ότι η ανίχνευση των κλάσεων χρειάζεται περαιτέρω βελτίωση.

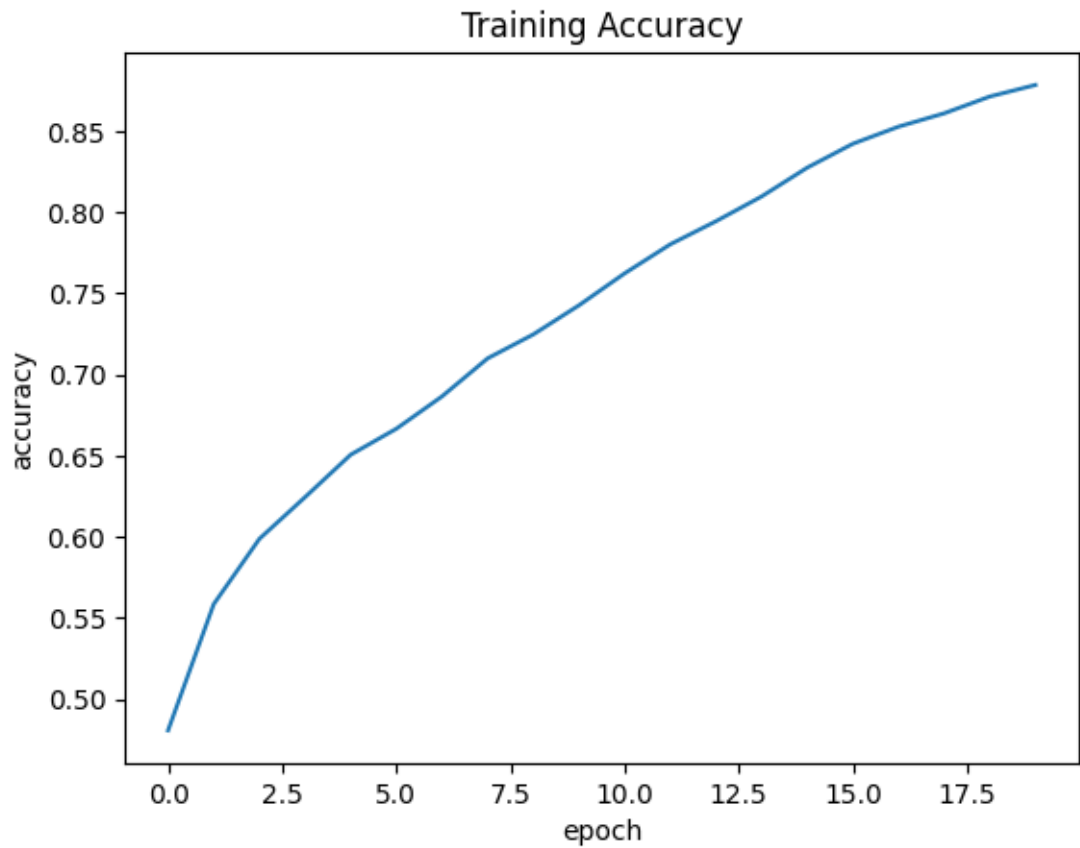
5.7.2. Αποτελέσματα 2^{ου} μοντέλου :

Τα ακόλουθα διαγράμματα παρουσιάζουν τα αποτελέσματα του 2^{ου} μοντέλου για το σύνολο εκπαίδευσης και επικύρωσης :



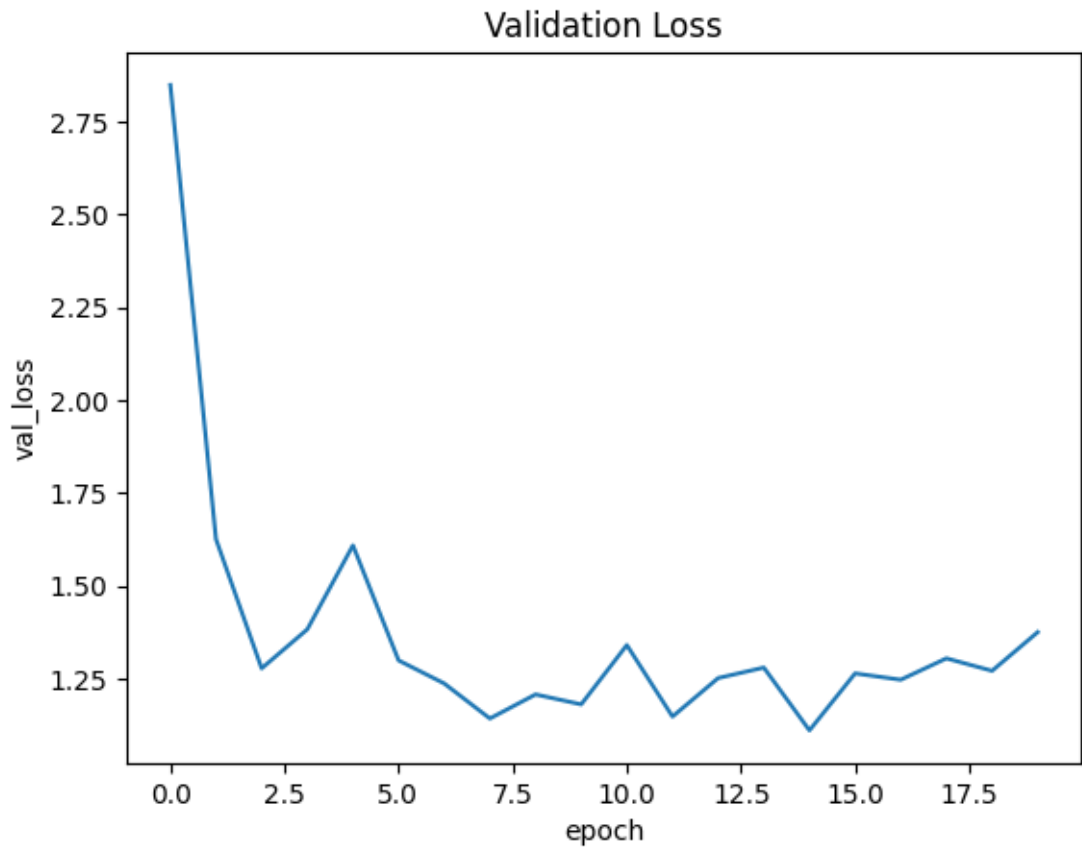
Γράφημα 10: Training Loss για το μοντέλο 2 (Deeplabv3+)

Παρατηρείται ότι η απώλεια στο σύνολο εκπαίδευσης μειώνεται σταδιακά κατά τη διάρκεια των εποχών, οπότε το μοντέλο βελτιώνεται και προσαρμόζεται καλύτερα στα δεδομένα εκπαίδευσης καθώς προχωρεί η εκπαίδευση.



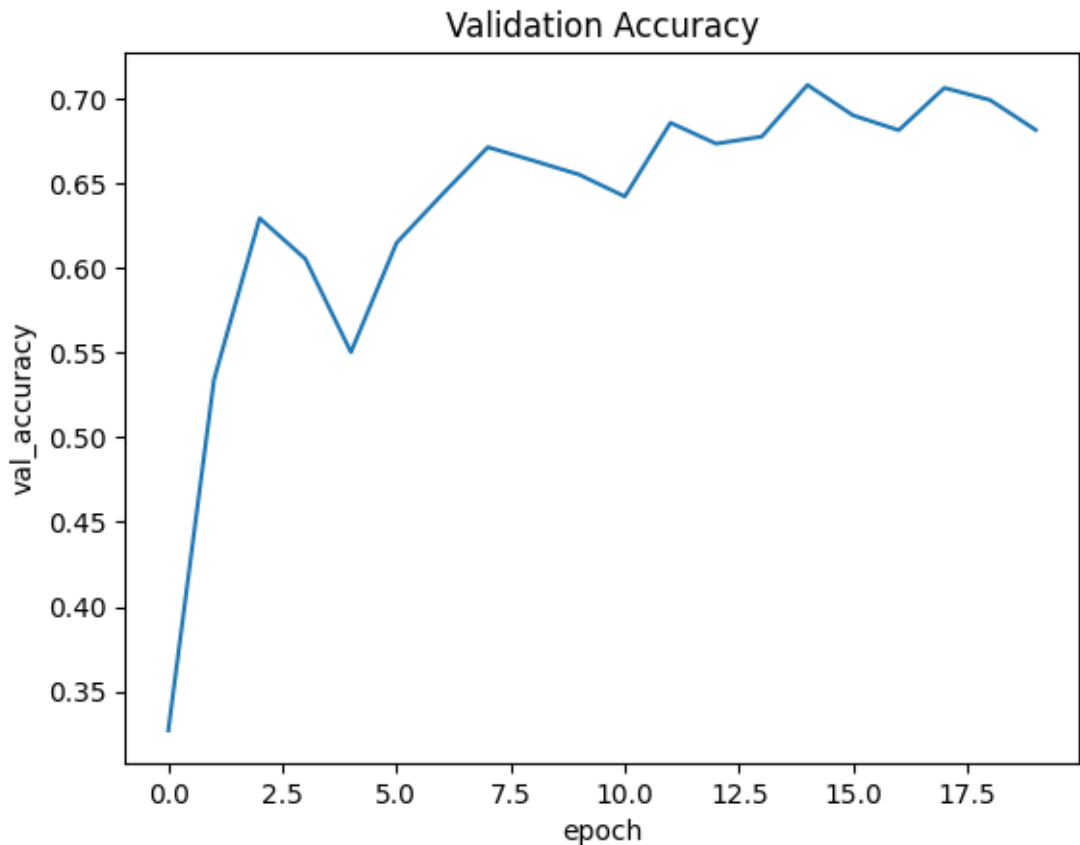
Γράφημα 11: Training accuracy για το μοντέλο 2 (Deeplabv3+)

Στο διάγραμμα αυτό παρατηρείται ότι η ακρίβεια στο σύνολο εκπαίδευσης αυξάνεται καθώς προχωρεί η εκπαίδευση, οπότε αυτό δείχνει ότι το μοντέλο μαθαίνει να κατηγοριοποιεί σωστά και βελτιώνεται στην πρόβλεψη των σωστών κατηγοριών.



Γράφημα 12: Validation Loss για το μοντέλο 2 (Deeplabv3+)

Στο διάγραμμα αυτό παρατηρείται ότι η απώλεια στο σύνολο επικύρωσης έχει κατασταλάξει και δεν μειώνεται περισσότερο μετά από ορισμένες εποχές.

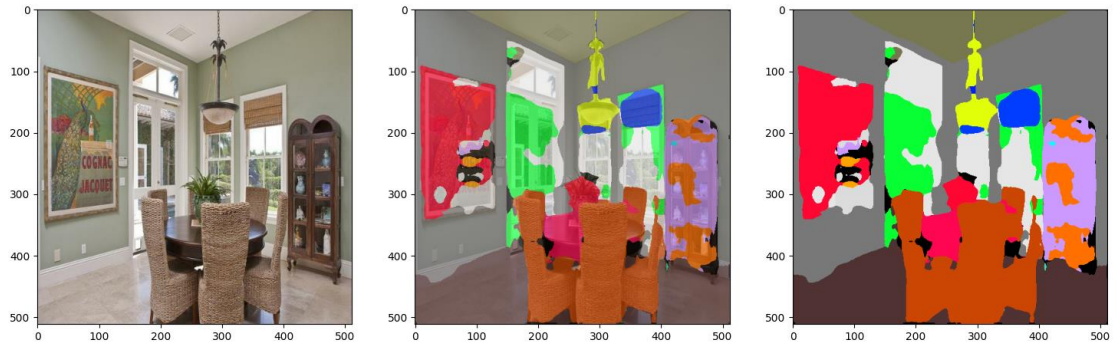


Γράφημα 13: Validation accuracy για το μοντέλο 2 (Deeplabv3+)

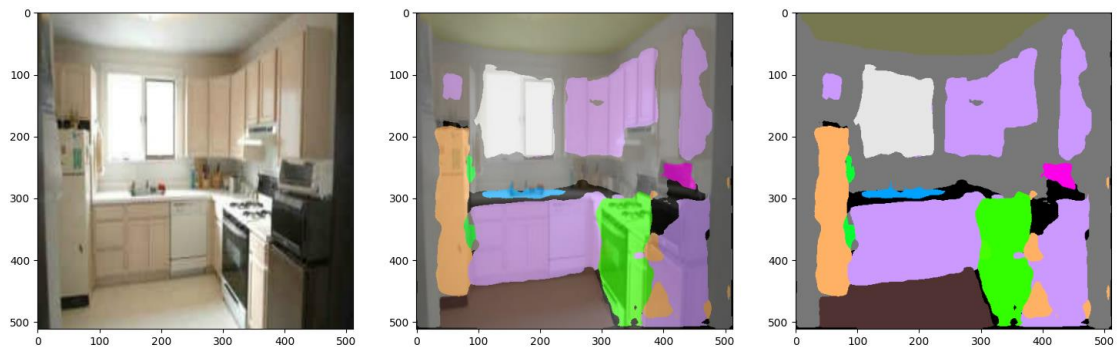
Στο τελευταίο διάγραμμα, παρατηρείται ότι η ακρίβεια στο σύνολο επικύρωσης συνεχίζει να αυξάνεται καθώς προχωρεί η εκπαίδευση, αν και με πιο χαμηλό ρυθμό σε σχέση με την ακρίβεια του συνόλου εκπαίδευσης. Αυτό δείχνει ότι βελτιώνεται η ικανότητα του μοντέλου να γενικεύει και να κατηγοριοποιεί σωστά και νέα δεδομένα.

Από την συνολική παρατήρηση των διαγραμμάτων φαίνεται ότι το μοντέλο εκπαιδεύεται και βελτιώνεται καθώς προχωρεί η εκπαίδευση. Ωστόσο η σύγκλιση και η απόκλιση της απώλειας και της ακρίβειας στο σύνολο επικύρωσης υποδεικνύουν ίσως την πιθανότητα υπερπροσαρμογής.

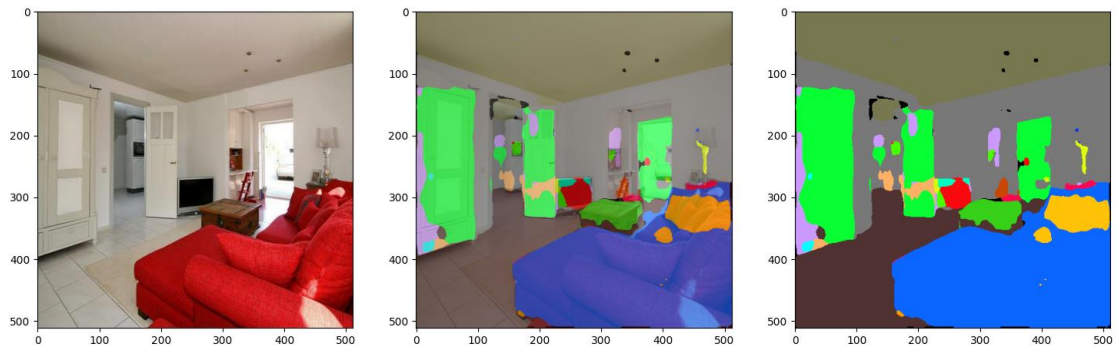
Οπτικοποίηση αποτελεσμάτων σύμφωνα με το ίδιο csv αρχείο που χρησιμοποιήθηκε και στο 1^ο μοντέλο :



Εικόνα 26: Πρόβλεψη εικόνας από το Μοντέλο 2(Deerlan3+)



Εικόνα 27: Πρόβλεψη εικόνας από το Μοντέλο 2(Deerlan3+)

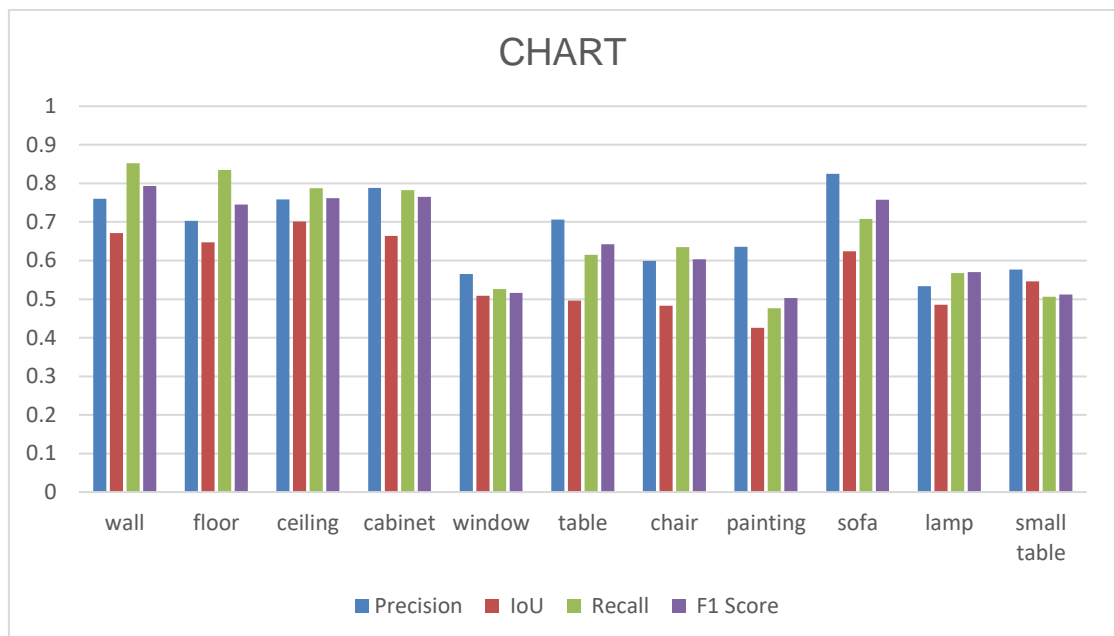


Εικόνα 28: Πρόβλεψη εικόνας από το Μοντέλο 2(Deerlan3+)

Από τα παραπάνω παρατηρείται ότι στο μεγαλύτερο μέρος της οι εικόνες προβλέπονται με ευκρίνεια. Ακολουθούν τα αποτελέσματα των μετρήσεων:

Category	Precision	IoU	Recall	F1 Score
wall	0,760274	0,671542	0,85211	0,793205
floor	0,702464	0,646918	0,834495	0,745439
ceiling	0,758064	0,700978	0,787166	0,761733
cabinet	0,788714	0,66349	0,78242	0,764666
window	0,565302	0,508442	0,525759	0,516306
table	0,706078	0,496348	0,615149	0,642694
chair	0,598841	0,482682	0,634845	0,603651
painting	0,635703	0,425915	0,476376	0,502713
sofa	0,825026	0,623884	0,708207	0,757928
lamp	0,533615	0,485802	0,567869	0,569847
small table	0,576723	0,545993	0,506337	0,511838

Πίνακας 3: Μέσος όρος αποτελεσμάτων για κάθε κατηγορία για το μοντέλο 2 (Deerlabv3+)



Διάγραμμα 2: Μέσος όρος αποτελεσμάτων για το μοντέλο 2 (Deerlabv3+)

Από τις μετρήσεις σε κάθε συγκεκριμένη κατηγορία μπορούν να σχολιαστούν τα παρακάτω :

Από την μέτρηση της ακρίβειας (precision) προκύπτει πολύ καλή ακρίβεια για τις κατηγορίες 'sofa' και 'cabinet' και χαμηλή για τις κατηγορίες 'window' και 'lamp'.

Από την επικάλυψη IoU παρατηρείται ότι υπάρχει αρκετά καλή επικάλυψη στις κατηγορίες 'ceiling', 'wall', 'cabinet', 'floor' και 'sofa' και πιο χαμηλή να παρατηρείται στην κατηγορία 'painting'.

Από την ανάκληση (recall) προκύπτει ότι παρουσιάζεται υψηλή ανάκληση στις κατηγορίες 'wall', 'floor' και πιο χαμηλή στις κατηγορίες 'painting' και 'small table'.

Από το F1score παρατηρείται ότι καλό F1score έχουν οι κατηγορίες 'wall', 'ceiling', 'cabinet', 'sofa' και 'floor' με χαμηλότερο τις κατηγορίες 'painting' και 'window' και 'small table'.

Γενικότερα πολύ καλή απόδοση παρουσιάζουν οι κατηγορίες 'wall', 'ceiling', 'cabinet', 'sofa' και 'floor'. Η κατηγορία 'table' παρουσιάζει μέτρια απόδοση σε όλες τις μετρήσεις ενώ οι υπόλοιπες κατηγορίες 'painting', 'window', 'chair', 'lamp' και 'small table' βρίσκονται σε χαμηλότερο μεν επίπεδο αλλά κοντά στο ικανοποιητικό.

Category	Precision	IoU	Recall	F1 Score
microwave	0,67101	0,589418	0,665273	0,658339
stove	0,52918	0,402897	0,573157	0,500398
refrigerator	0,594471	0,469697	0,543767	0,542971
chandelier	0,372468	0,421207	0,536567	0,450767
flower	0,336776	0,40635	0,764484	0,448503
fireplace	0,356531	0,438429	0,601612	0,438301

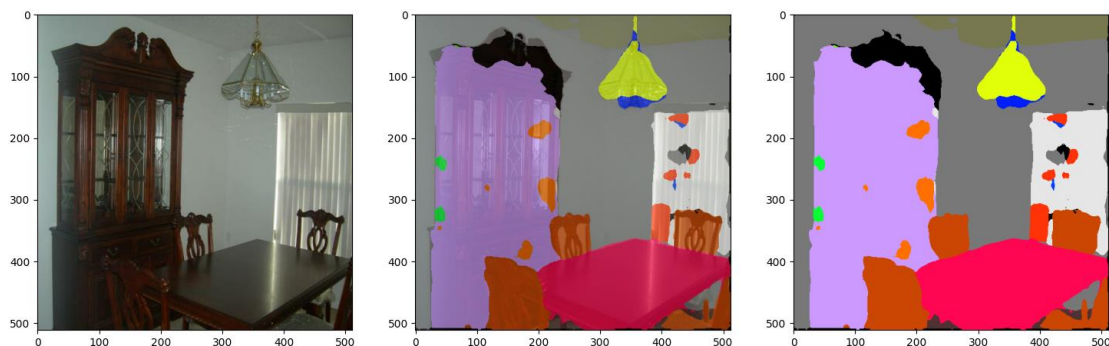
Πίνακας 4:Επιπλέον αποτελέσματα για το μοντέλο 2(Deeplabv3+)

Στο μοντέλο αυτό παρατηρείται ότι ανιχνεύονται κατηγορίες για τις οποίες τα δεδομένα ήταν ελάχιστα σε σύγκριση με τις κύριες κατηγορίες που ενδιέφεραν για εκπαίδευση. Όπως φαίνεται από τον παραπάνω πίνακα οι αποδόσεις στις μετρήσεις αρκετών από αυτών κινούνται σε μέτριο έως ικανοποιητικό επίπεδο.

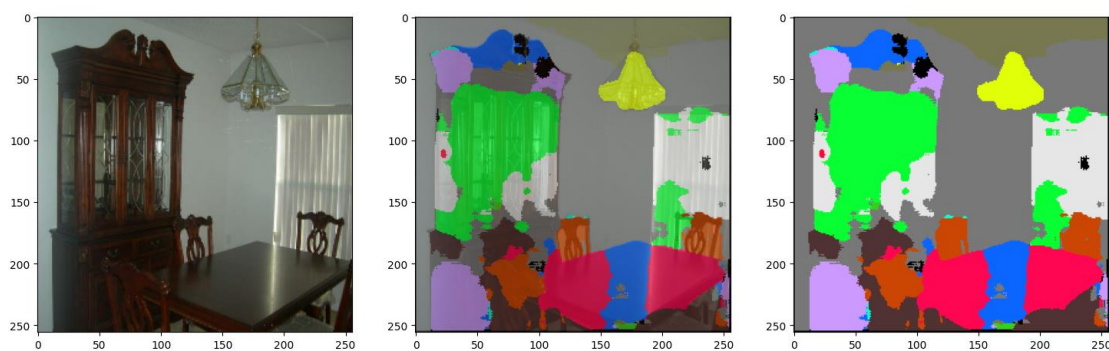
5.8. Συγκρίσεις

Προκειμένου να γίνουν οι συγκρίσεις των αποτελεσμάτων των 2 μοντέλων παρουσιάζονται παρακάτω :

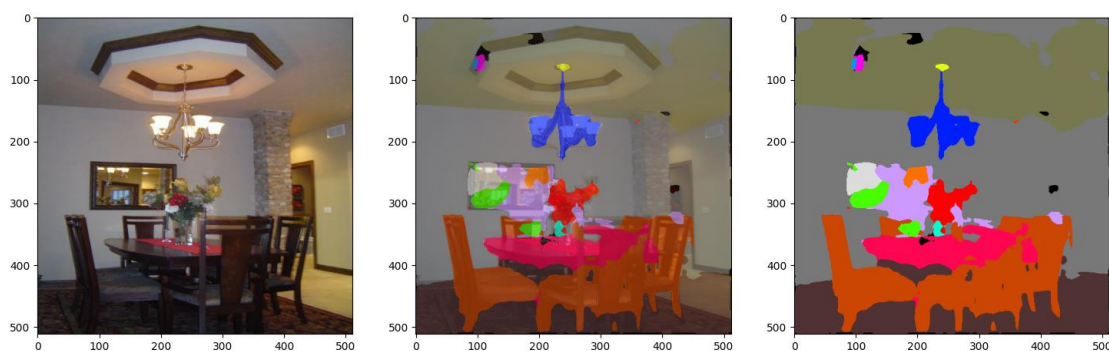
- Από την οπτικοποίηση των αποτελεσμάτων :



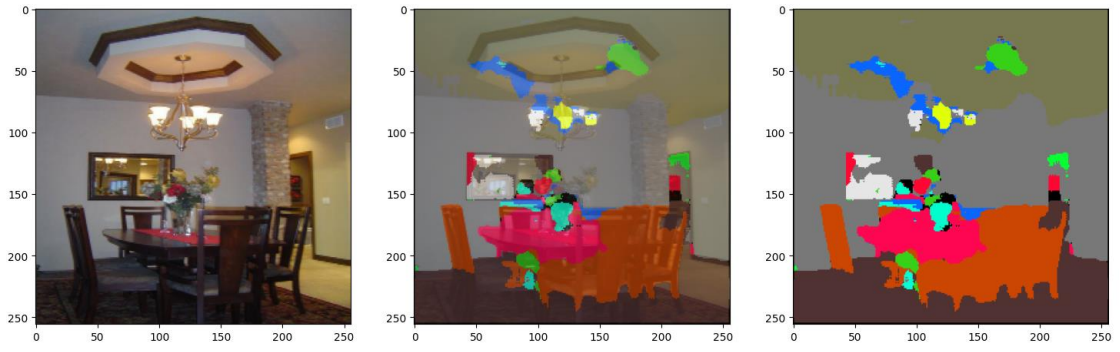
Εικόνα 29: Πρόβλεψη εικόνας από το Μοντέλο 2(Deerplan3+)



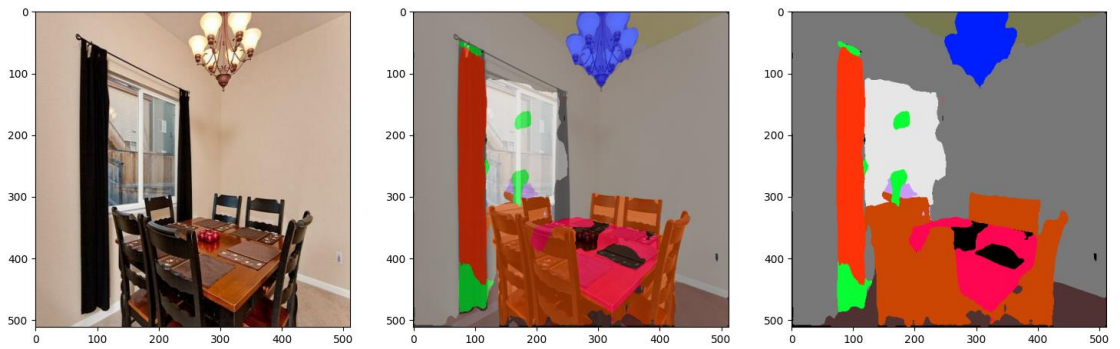
Εικόνα 30: Πρόβλεψη εικόνας από το Μοντέλο 1(Unet)



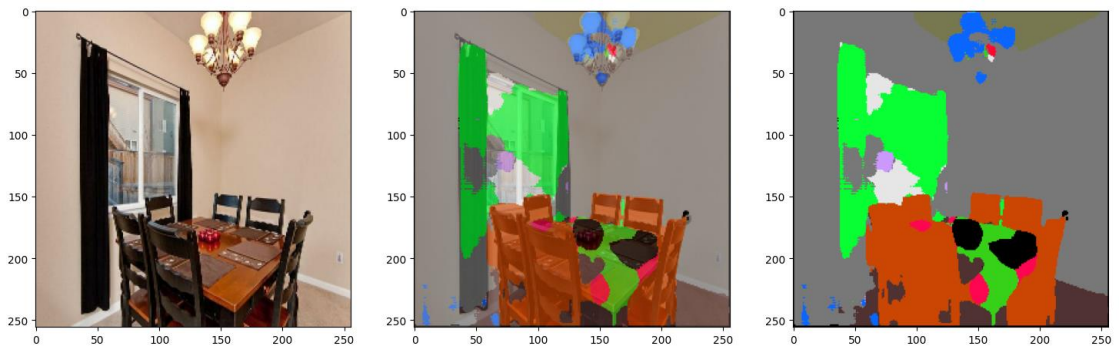
Εικόνα 31: Πρόβλεψη εικόνας από το Μοντέλο 2(Deerplan3+)



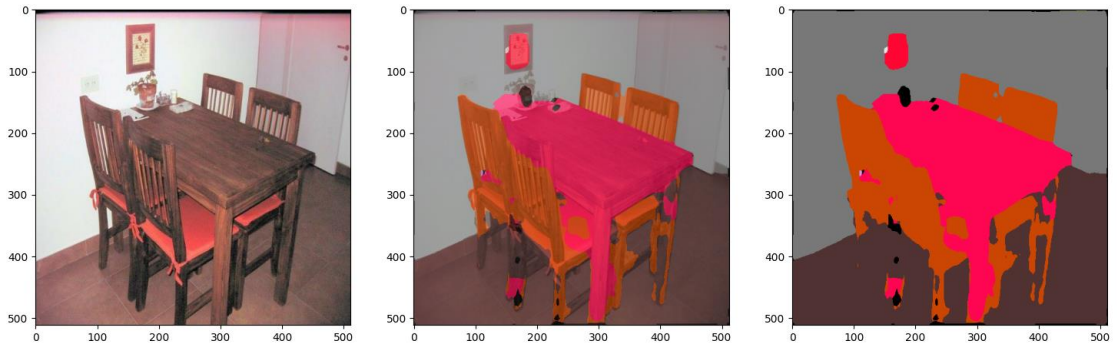
Εικόνα 32: Πρόβλεψη εικόνας από το Μοντέλο 1(Unet)



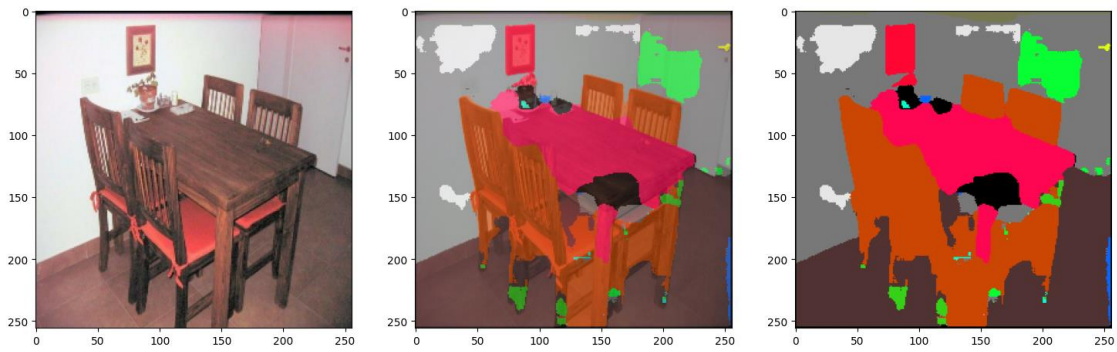
Εικόνα 33: Πρόβλεψη εικόνας από το Μοντέλο 2(Deerlanv3+)



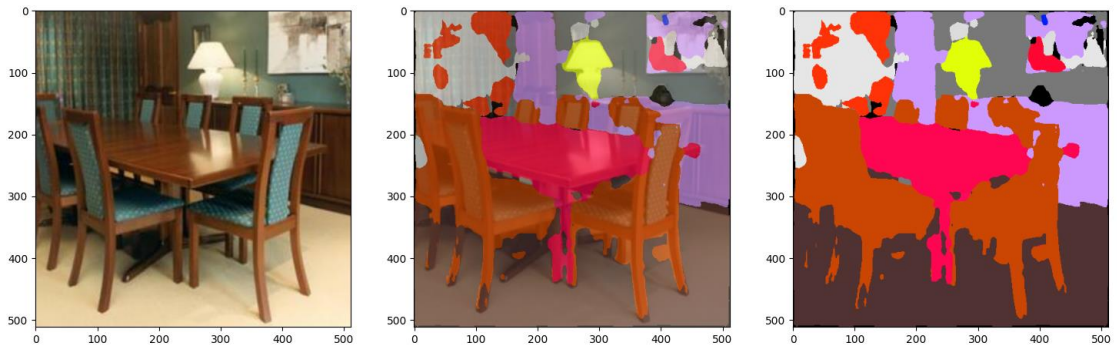
Εικόνα 34: Πρόβλεψη εικόνας από το Μοντέλο 1(Unet)



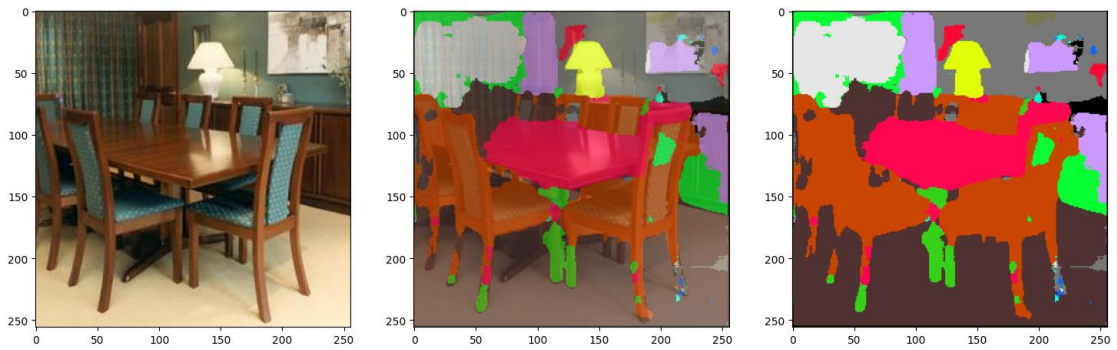
Εικόνα 35: Πρόβλεψη εικόνας από το Μοντέλο 2(DeeplabV3+)



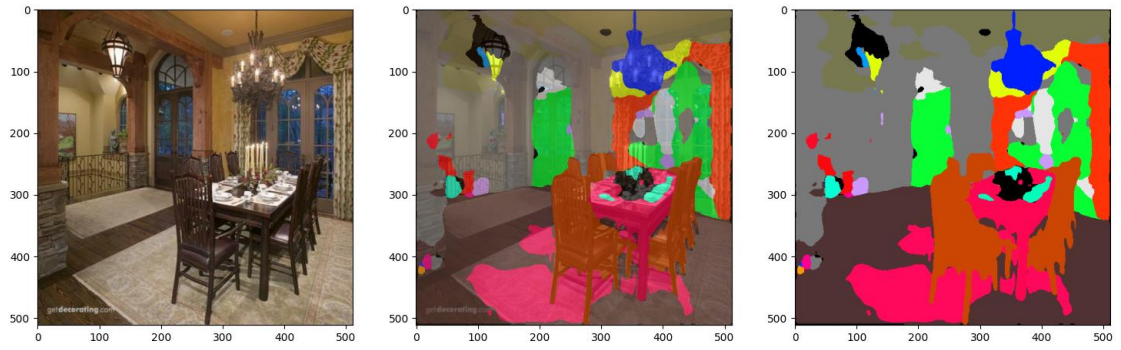
Εικόνα 36: Πρόβλεψη εικόνας από το Μοντέλο 1(Unet)



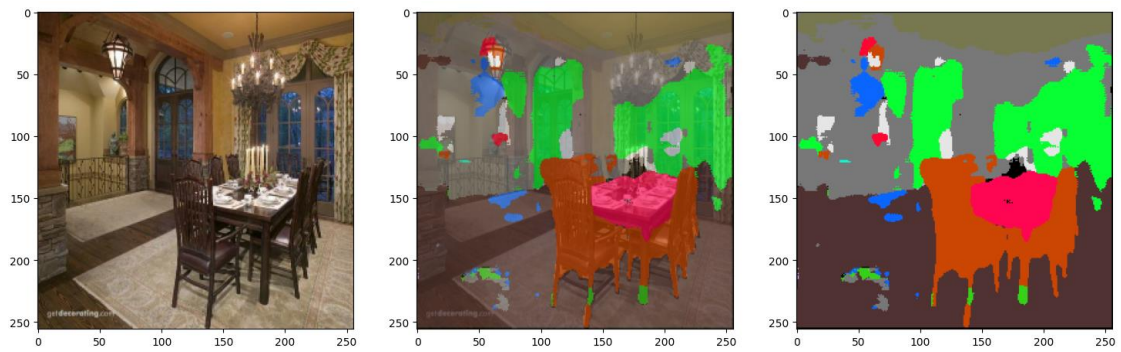
Εικόνα 37: Πρόβλεψη εικόνας από το Μοντέλο 2(DeeplabV3+)



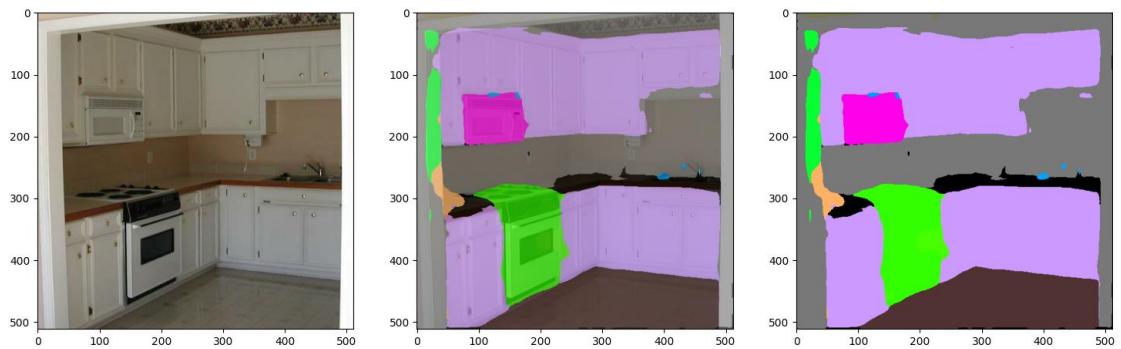
Εικόνα 38: Πρόβλεψη εικόνας από το Μοντέλο 1(Unet)



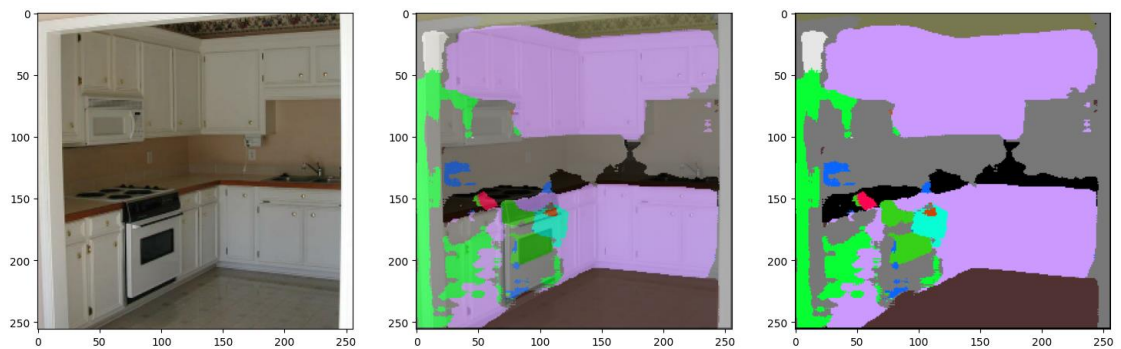
Εικόνα 39: Πρόβλεψη εικόνας από το Μοντέλο 2 (Deerlan3+)



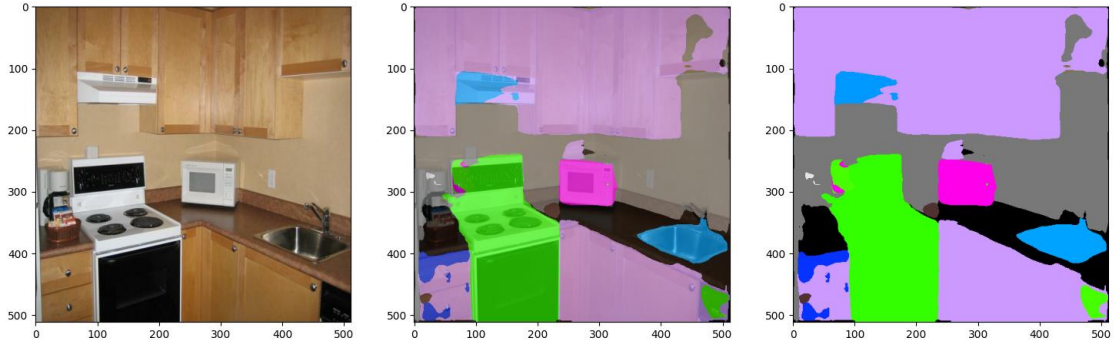
Εικόνα 40: Πρόβλεψη εικόνας από το Μοντέλο 1 (Unet)



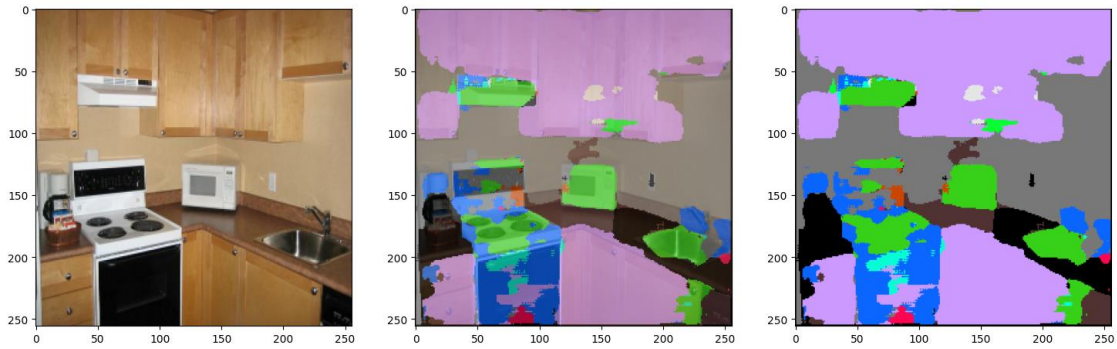
Εικόνα 41: Πρόβλεψη εικόνας από το Μοντέλο 2 (Deerlan3+)



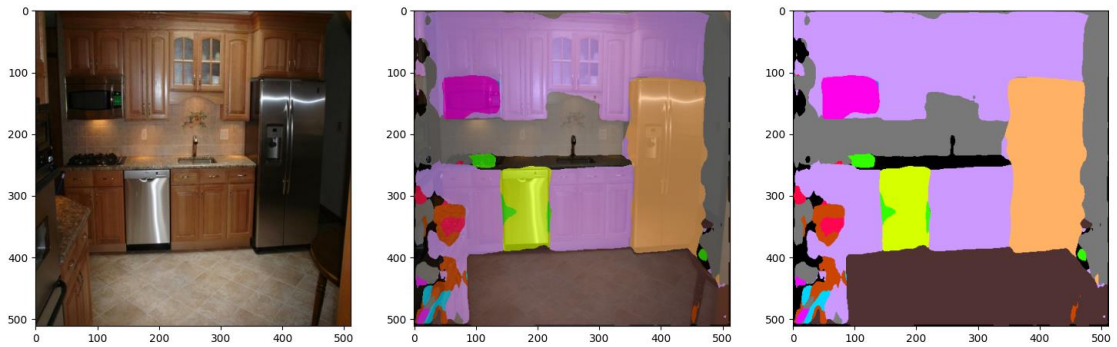
Εικόνα 42: Πρόβλεψη εικόνας από το Μοντέλο 1 (Unet)



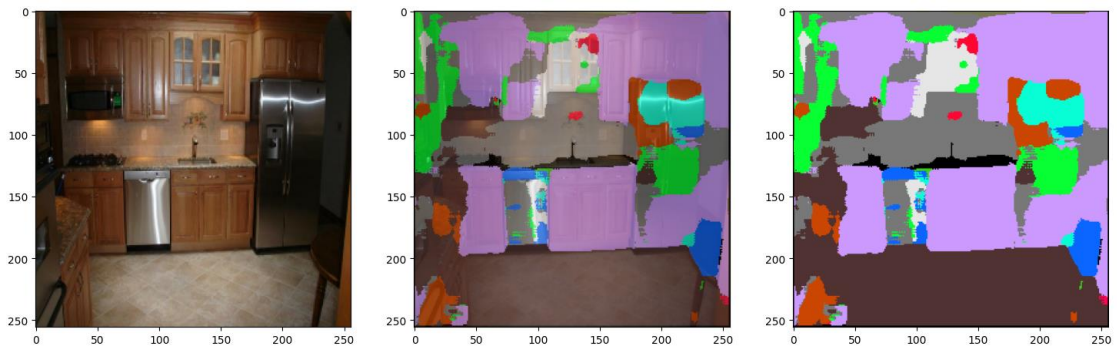
Εικόνα 43: Πρόβλεψη εικόνας από το Μοντέλο 2(Deeplabv3+)



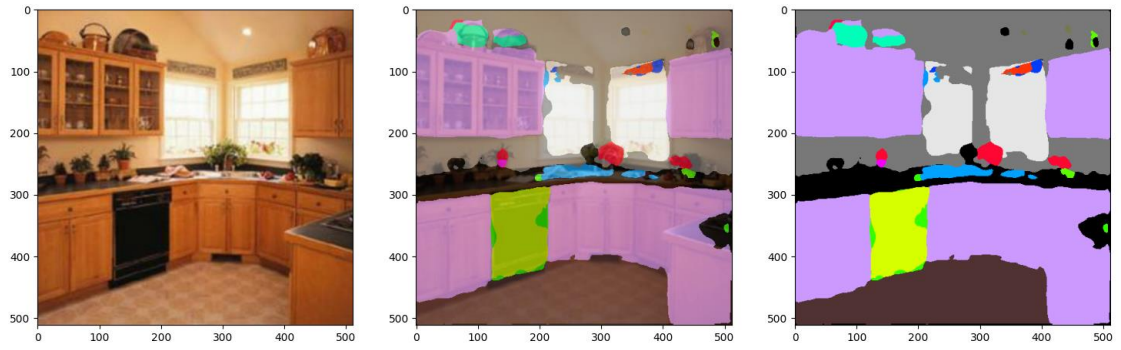
Εικόνα 44: Πρόβλεψη εικόνας από το Μοντέλο 1(Unet)



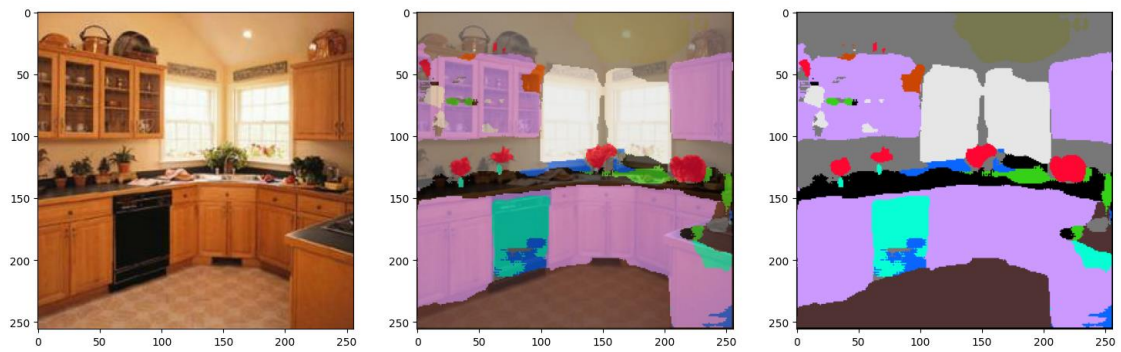
Εικόνα 45: Πρόβλεψη εικόνας από το Μοντέλο 2(Deeplabv3+)



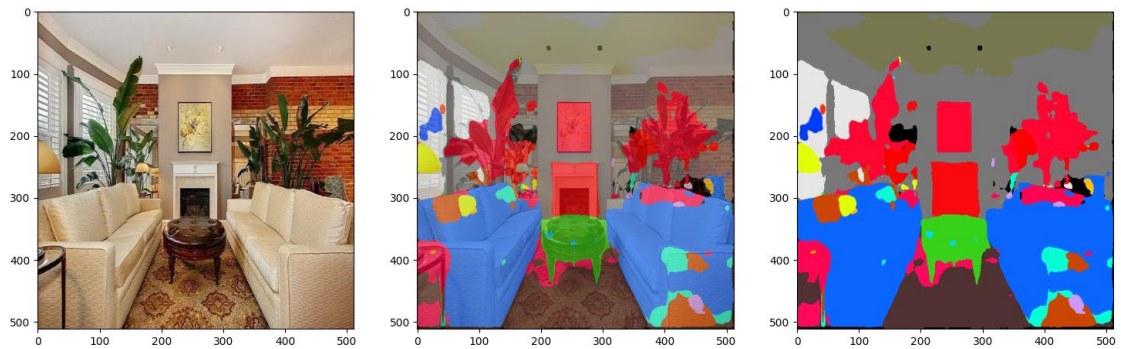
Εικόνα 46: Πρόβλεψη εικόνας από το Μοντέλο 1(Unet)



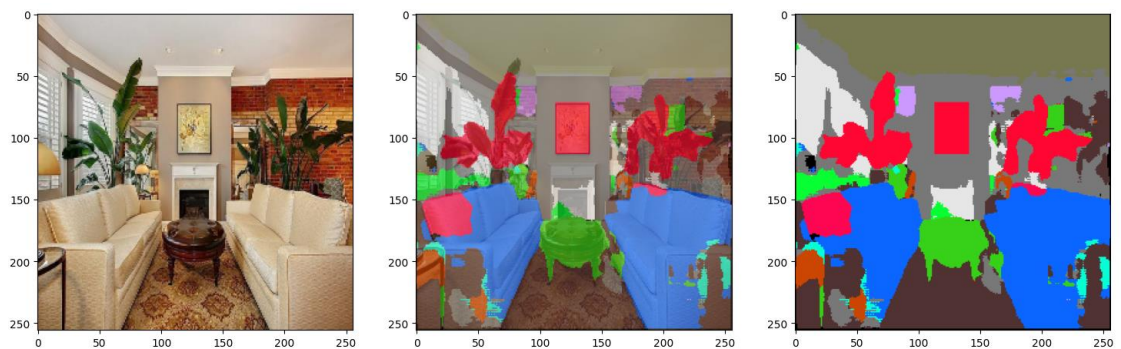
Εικόνα 47: Πρόβλεψη εικόνας από το Μοντέλο 2 (Deerlanv3+)



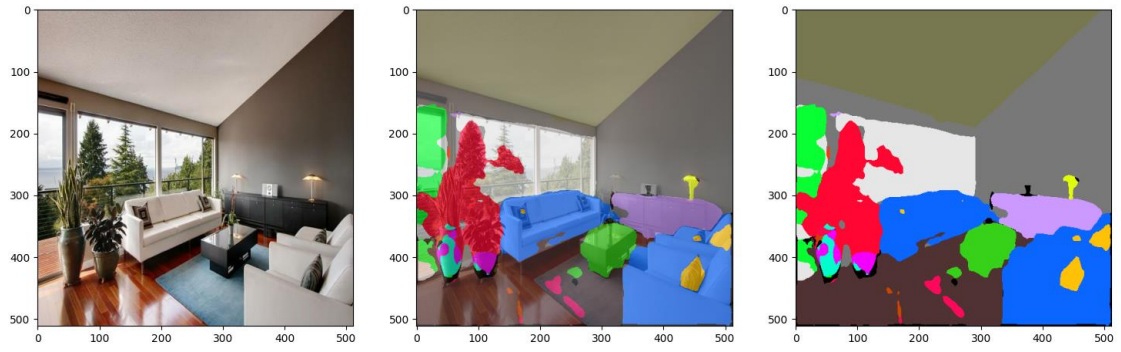
Εικόνα 48: Πρόβλεψη εικόνας από το Μοντέλο 1 (Unet)



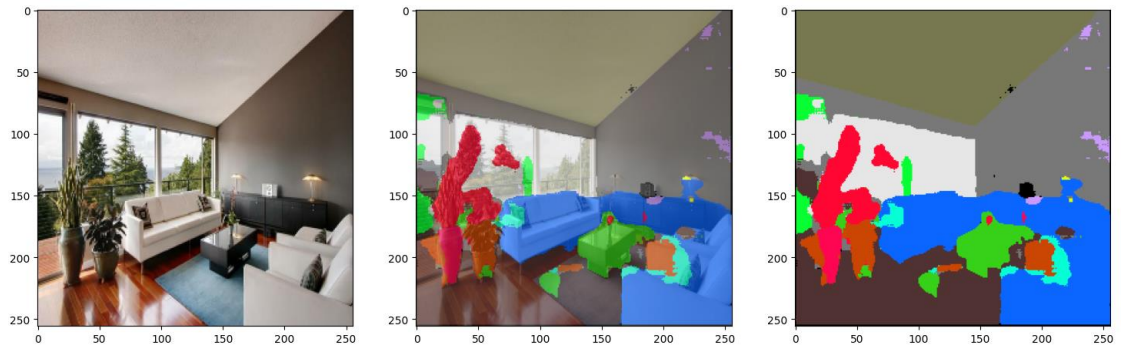
Εικόνα 49: Πρόβλεψη εικόνας από το Μοντέλο 2 (Deerlanv3+)



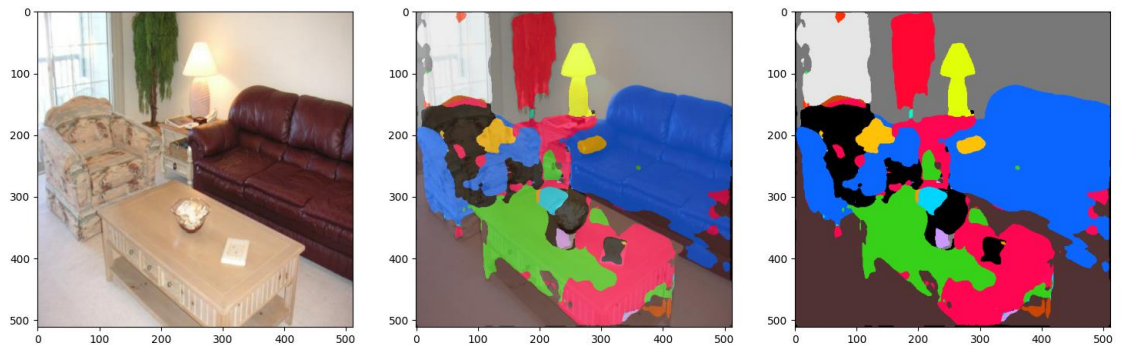
Εικόνα 50: Πρόβλεψη εικόνας από το Μοντέλο 1 (Unet)



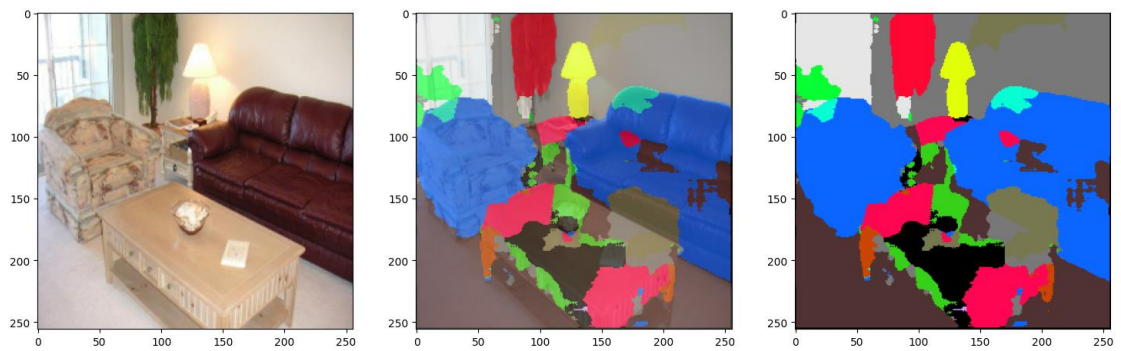
Εικόνα 51: Πρόβλεψη εικόνας από το Μοντέλο 2(Deeplabv3+)



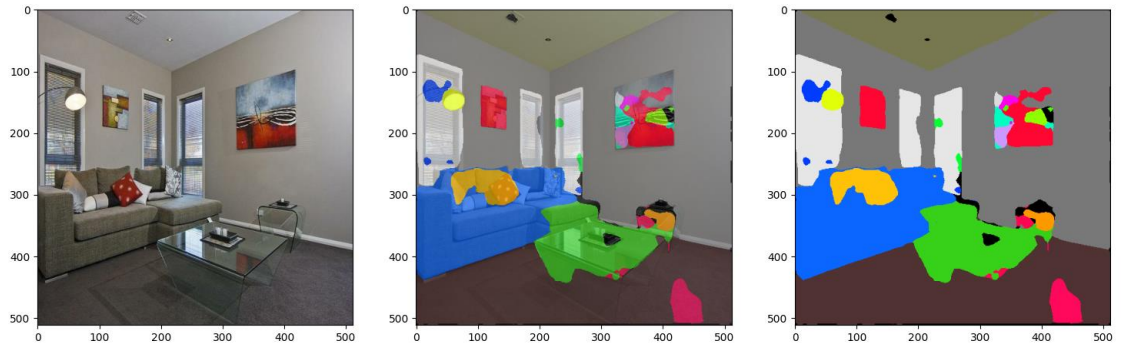
Εικόνα 52: Πρόβλεψη εικόνας από το Μοντέλο 1(Unet)



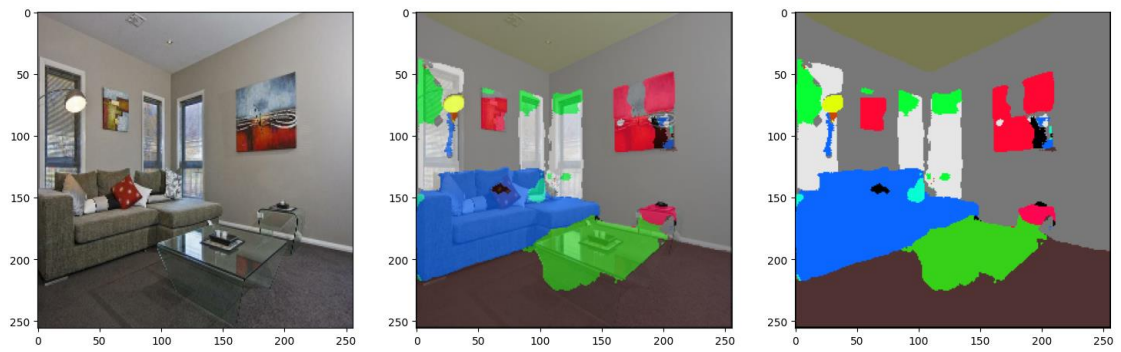
Εικόνα 53: Πρόβλεψη εικόνας από το Μοντέλο 2(Deeplabv3+)



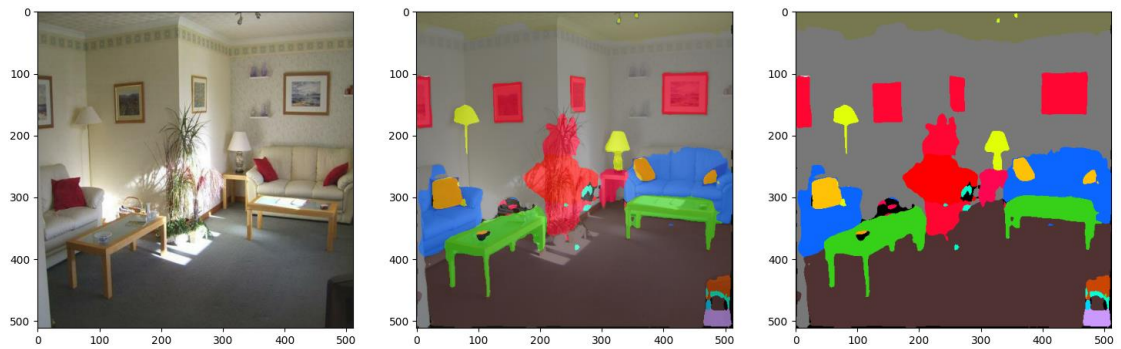
Εικόνα 54: Πρόβλεψη εικόνας από το Μοντέλο 1(Unet)



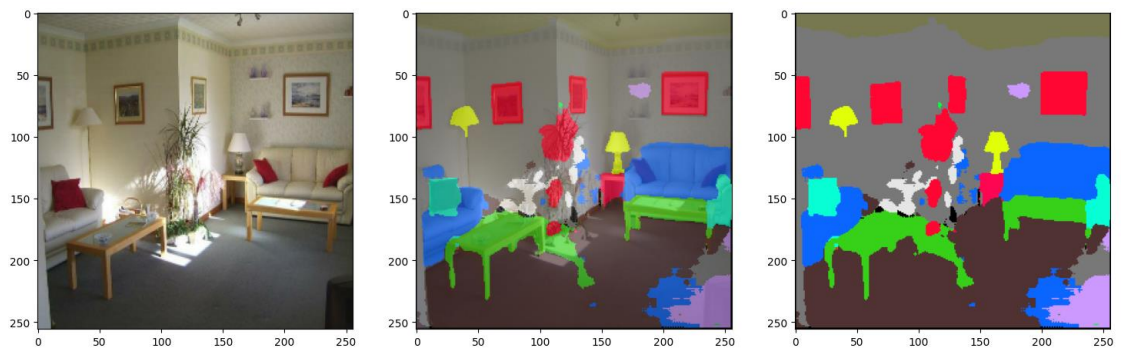
Εικόνα 55: Πρόβλεψη εικόνας από το Μοντέλο 2 (Deerlan3+)



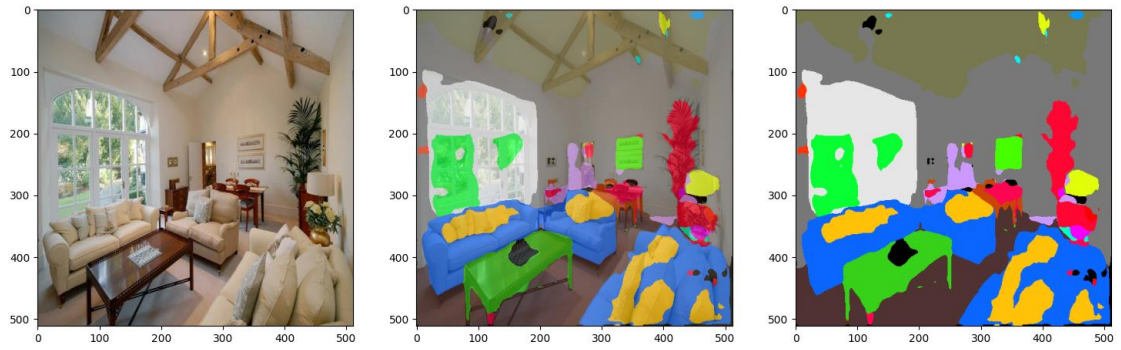
Εικόνα 56: Πρόβλεψη εικόνας από το Μοντέλο 1 (Unet)



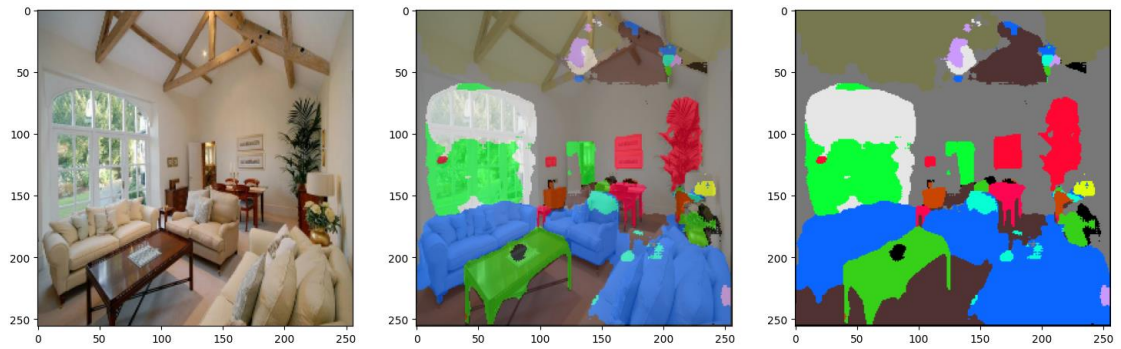
Εικόνα 57: Πρόβλεψη εικόνας από το Μοντέλο 2 (Deerlan3+)



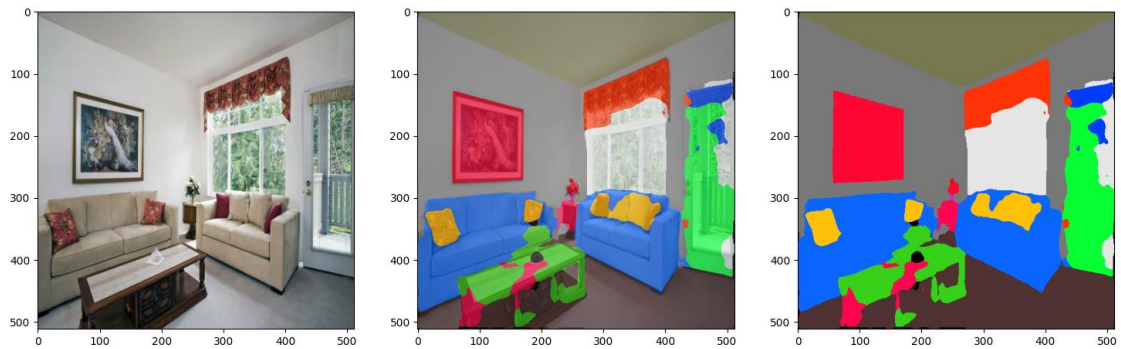
Εικόνα 58: Πρόβλεψη εικόνας από το Μοντέλο 1 (Unet)



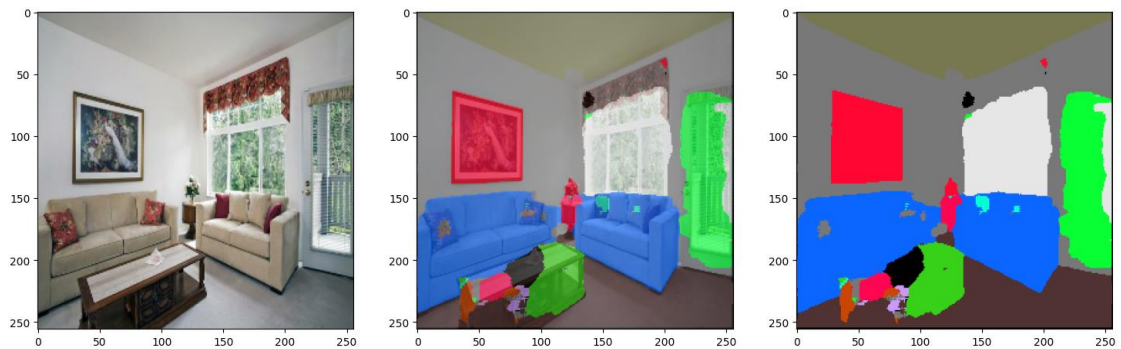
Εικόνα 59: Πρόβλεψη εικόνας από το Μοντέλο 2 (Deepplan3+)



Εικόνα 60: Πρόβλεψη εικόνας από το Μοντέλο 1 (Unet)



Εικόνα 61: Πρόβλεψη εικόνας από το Μοντέλο 2 (Deepplan3+)



Εικόνα 62: Πρόβλεψη εικόνας από το Μοντέλο 1 (Unet)

Από τα οπτικά αποτελέσματα, παρατηρείται ότι το 2^ο μοντέλο(DeepLabV3+) ανιχνεύει καλύτερα τις κατηγορίες, χωρίς να τις μπερδεύει με άλλες κατηγορίες. Για παράδειγμα στο 1^ο μοντέλο (UNET) η κατηγορία ‘table’ συχνά συγχέεται με αυτήν της ‘small table’ και η κατηγορία ‘window’ με αυτήν της ‘door’ (υποκατηγορία που περιλαμβάνεται στα δεδομένα), γεγονός που επηρεάζει την απόδοση του μοντέλου.

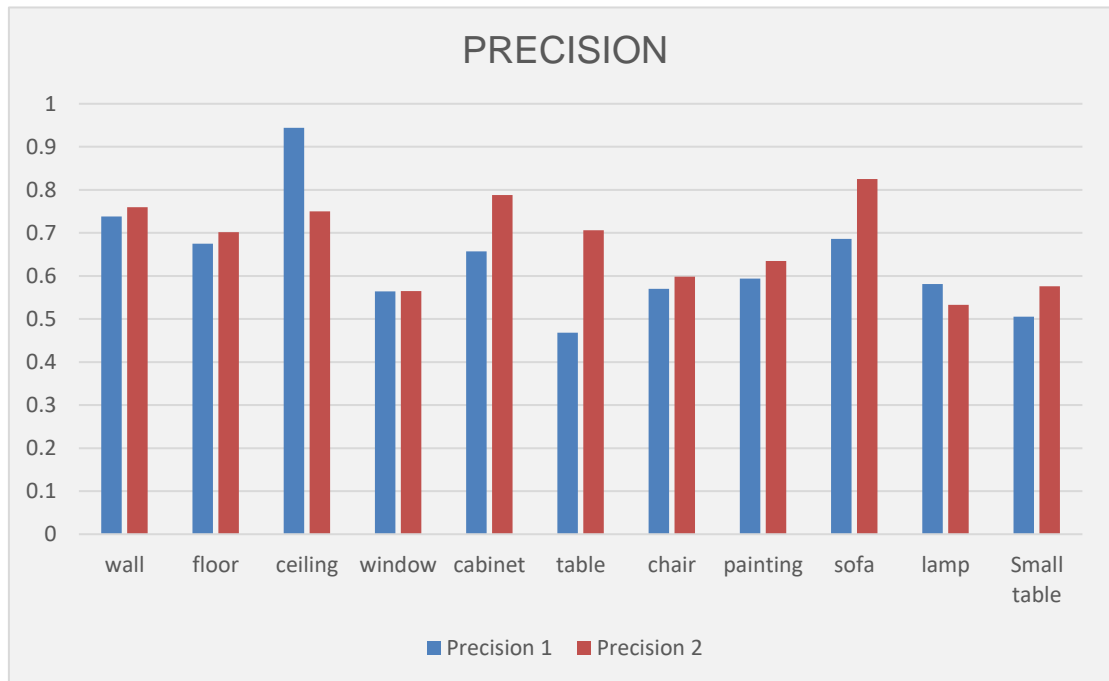
Επίσης το 2^ο μοντέλο ανιχνεύει περισσότερες λεπτομέρειες στις εικόνες οι οποίες είναι στις υποκατηγορίες των δεδομένων όπως π.χ. ‘cushion’, ‘refrigerator’, ‘stove’, ‘microwave’ κ.α.

- Από τις μετρήσεις :

Class Name	Precision 1	Precision 2	IoU 1	IoU 2	Recall 1	Recall 2	F1 Score 1	F1 Score 2
wall	0,738	0,760	0,566	0,671	0,724	0,852	0,72	0,793
floor	0,675	0,702	0,578	0,646	0,801	0,834	0,732	0,745
ceiling	0,944	0,75	0,896	0,700	0,946	0,787	0,945	0,761
window	0,564	0,565	0,40	0,508	0,618	0,525	0,458	0,516
cabinet	0,657	0,788	0,53	0,663	0,749	0,782	0,700	0,764
Table	0,468	0,706	0,375	0,496	0,334	0,615	0,498	0,642
Chair	0,57	0,598	0,473	0,482	0,617	0,634	0,57	0,603
painting	0,594	0,635	0,539	0,425	0,441	0,476	0,506	0,502
Sofa	0,686	0,825	0,474	0,623	0,803	0,708	0,606	0,757
Lamp	0,581	0,533	0,508	0,48	0,770	0,367	0,662	0,369
Small table	0,505	0,576	0,436	0,545	0,676	0,506	0,42	0,511

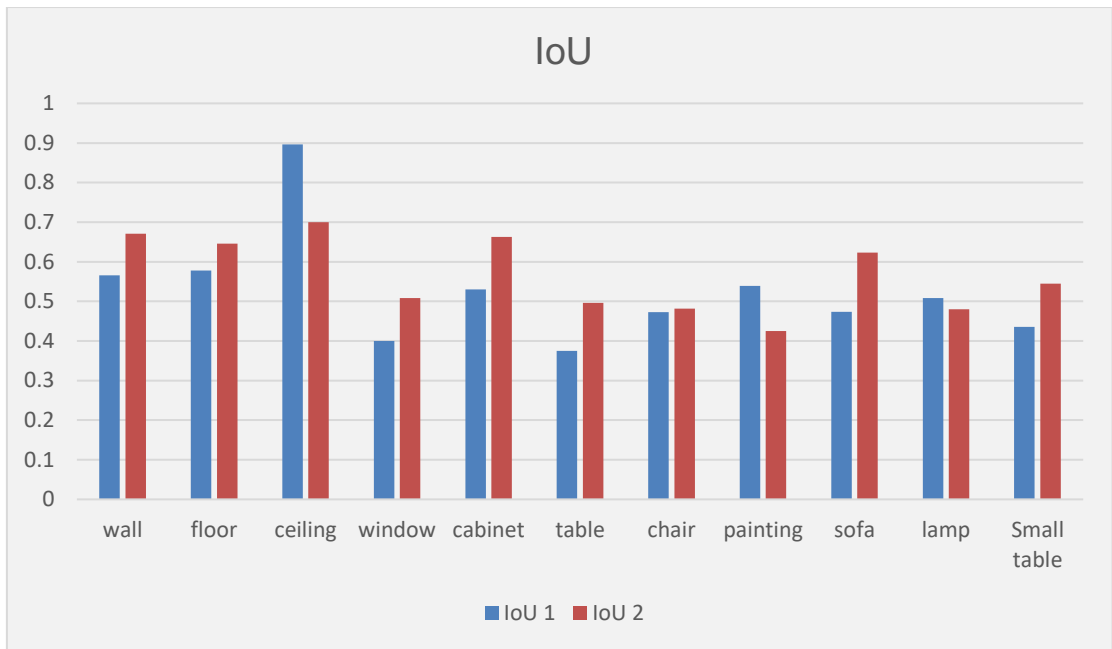
Πίνακας 5: Σύγκριση των αποτελεσμάτων και για τα δύο μοντέλα

- Και από τους συγκριτικούς πίνακες παρατηρούνται τα ακόλουθα:



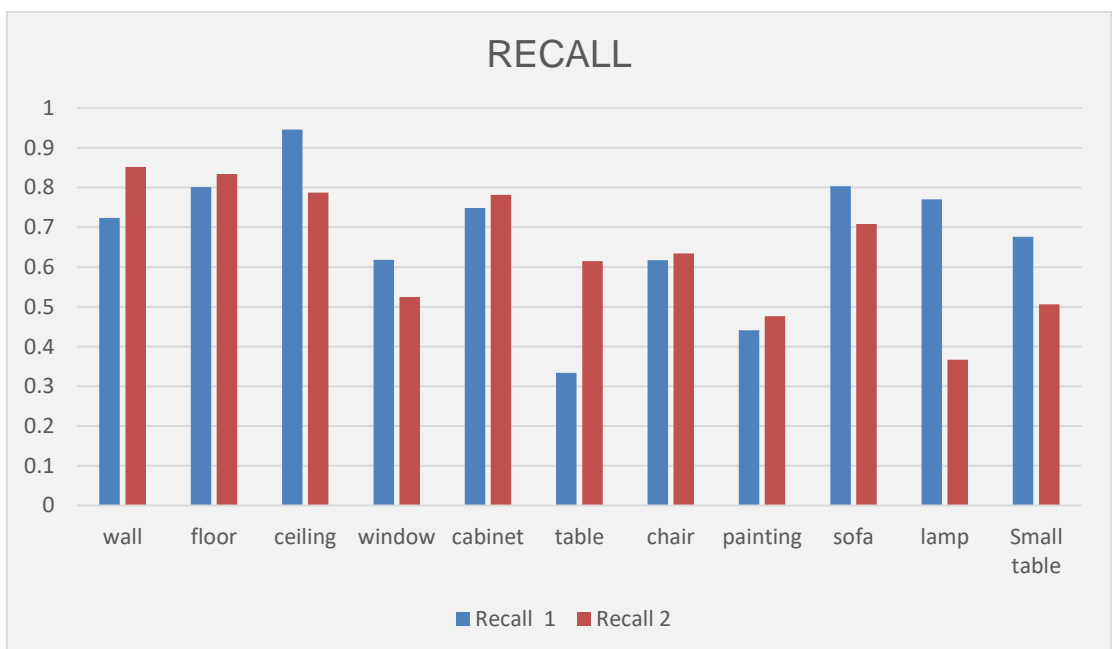
Διάγραμμα 3: Σύγκριση precision των δύο μοντέλων για κάθε κατηγορία

Το 2^ο μοντέλο παρουσιάζει υψηλότερη ακρίβεια από το 1^ο μοντέλο σε όλες τις κατηγορίες εκτός των κατηγοριών 'ceiling' και 'lamp' όπου το 1^ο έχει μεγαλύτερη ακρίβεια



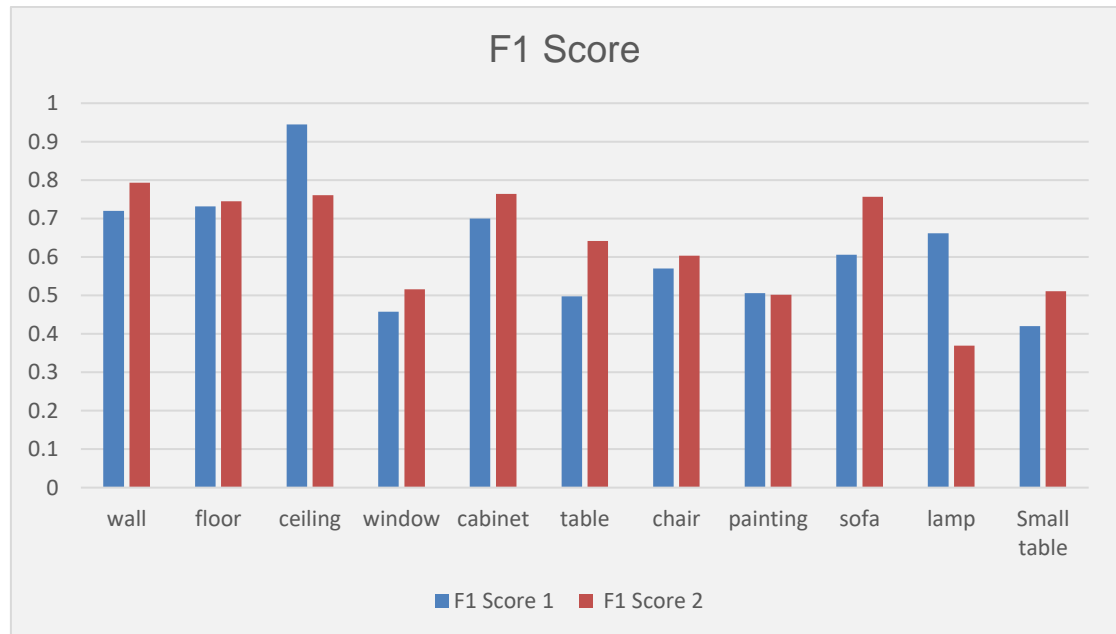
Διάγραμμα 4: Σύγκριση IoU των δύο μοντέλων για κάθε κατηγορία

Επίσης το 2^ο μοντέλο υπερτερεί στην επικάλυψη IoU σε όλες τις κατηγορίες εκτός των 'ceiling' 'painting' και 'lamp' όπου υπερτερεί το 1^ο.



Διάγραμμα 5: Σύγκριση RECALL των δύο μοντέλων για κάθε κατηγορία

Στην μέτρηση 'recall' η απόδοση των μοντέλων μοιράζεται με το 1^ο μοντέλο να έχει μεγαλύτερη στις κατηγορίες 'ceiling', 'window', 'sofa', 'lamp', και 'small table' ενώ το 2^ο υπερτερεί στις υπόλοιπες 6 κατηγορίες.



Διάγραμμα 6: Σύγκριση precision των δύο μοντέλων για κάθε κατηγορία

Τέλος στο F1 score το 2^ο μοντέλο είναι πιο αποδοτικό σε όλες τις κατηγορίες εκτός των 'ceiling' και 'lamp' όπου η απόδοση του 1^{ου} μοντέλου είναι καλύτερη με την κατηγορία 'painting' να είναι οριακά καλύτερη στο 1^ο μοντέλο.

ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΠΡΟΟΠΤΙΚΕΣ

Από τα πιο πάνω συγκριτικά αποτελέσματα, μπορούν να εξαχθούν τα ακόλουθα συμπεράσματα :

Με δεδομένο ότι οι γενικοί μέσοι όροι των κατηγοριών στις μετρήσεις που παρατέθηκαν είναι:

Precision 1: 0.629

Precision 2: 0.681

IoU 1: 0.516

IoU 2: 0.572

Recall 1: 0.696

Recall 2: 0.674

F1 Score 1: 0.631

F1 Score 2: 0.633

Από τα παραπάνω προκύπτει το 2^ο μοντέλο έχει καλύτερη ακρίβεια συνολικά σε σχέση με το 1^ο . Επίσης το 2^ο μοντέλο έχει υψηλότερη τιμή IoU προσφέροντας καλύτερη ακρίβεια στην αναγνώριση των αντικειμένων. Αντίθετα το 1^ο μοντέλο έχει λίγο υψηλότερη ανάκληση, καλύπτει δηλαδή καλύτερα την πλήρη αναγνώριση των αντικειμένων. Στο F1 score όπου συνδυάζεται η ακρίβεια με την ανάκληση και αποτελεί έναν αξιόπιστο τρόπο μέτρησης της απόδοσης, οι τιμές είναι παρόμοιες με το 2^ο μοντέλο να έχει μέση τιμή 0,633 ενώ το 1^ο μοντέλο να έχει μέση τιμή 0,631.

Συνολικά λόγω της υψηλότερης ακρίβειας και IoU, το 2^ο μοντέλο φαίνεται να έχει καλύτερη απόδοση από το 1^ο. Συνολικά η απόδοση και των δύο φαίνεται ικανοποιητική αν ληφθεί υπόψη ότι το δείγμα των δεδομένων ελέγχου ήταν αρκετά μεγάλο (100 φωτογραφίες) προκειμένου να είναι όσο το δυνατόν πιο αντιπροσωπευτικό.

Επίσης τα δείγματα αυτά (εικόνες) συχνά απεικονίζουν πολλαπλά αντικείμενα στην ίδια εικόνα, με αποτέλεσμα η πολυπλοκότητα που παρουσιάζεται σε σχήματα, γωνίες κλπ. να δυσκολεύουν την σωστή πρόβλεψη των

αντικειμένων. Εικόνες με λιγότερα αντικείμενα μπορεί να οδηγήσουν τα μοντέλα σε μεγαλύτερη ακρίβεια και γενικότερα καλύτερη απόδοση.

Η επίτευξη του στόχου για την λειτουργία τους σε ένα ρομπότ, μπορεί να επιφέρει ικανοποιητικά αποτελέσματα στην αναγνώριση των αντικειμένων αλλά ουσιαστικά και τα δύο μοντέλα επιδέχονται αρκετές βελτιώσεις τόσο ίσως στην αρχιτεκτονική τους, στην επιλογή διαφόρων παραμέτρων όσο και στις επιλογές των δειγμάτων από το αρχείο δεδομένων που ιδανικά θα μπορούσαν να αναδιαμορφώσουν τα αποτελέσματα και να βελτιώσουν την απόδοση τους. Οι συνδυασμοί των εναλλακτικών πειραματικών προσπαθειών θα μπορούσαν να είναι αρκετοί και ιδιαίτερα χρονοβόροι που ξεφεύγουν από τα χρονικά όρια που απαιτούνταν για την παράδοση αυτής της εργασίας. Γενικότερα παρόμοια μοντέλα ή και ακόμα μοντέλα με πολύ πιο ικανοποιητικά αποτελέσματα αναμφισβήτητα επιδέχονται συνεχείς βελτιώσεις καθώς πλέον στο ερευνητικό πεδίο της σημασιολογικής κατάτμησης εικόνων οι εξελίξεις είναι ραγδαίες και οι επιλογές τόσο στην αρχιτεκτονική όσο και στις επιμέρους παραμέτρους είναι πολλαπλές.

Οι δυσκολίες που αντιμετωπίστηκαν σε όλη την διάρκεια αυτής της εργασίας δεν ήταν αμελητέες. Το αρχείο δεδομένων που επιλέχθηκε προκειμένου να εξεταστούν προβλέψεις σε εσωτερικό χώρο, δημιούργησε πολλά προβλήματα λόγω του όγκου και της πολυπλοκότητας του. Η δημιουργία του κώδικα και για τις δύο περιπτώσεις, αποδείχτηκε πολύ ενδιαφέρουσα εργασία αλλά χρειαζόταν την απόλυτη προσοχή για την αποφυγή μικρών λαθών που προκάλεσαν αρκετές καθυστερήσεις.

Συνολικά η εμπειρία που αποκτήθηκε ήταν πολύτιμη και κράτησε αμείωτο το ενδιαφέρον σε όλη την διάρκεια αυτής της προσπάθειας.

Οι προοπτικές στο πεδίο της σημασιολογικής κατάτμησης εικόνων είναι πολλές για το μέλλον. Είναι ένας δυναμικός τομέας με αμέτρητες δυνατότητες για περαιτέρω ανάπτυξη και καινοτομία. Η ανάπτυξη πιο αποδοτικών τεχνικών και αλγορίθμων για την επεξεργασία δεδομένων, μεγάλου όγκου

καθώς και η έρευνα με νέες αρχιτεκτονικές ή η αξιοποίηση όλων των τεχνικών της βαθιάς μάθησης θα μπορούσαν μελλοντικά να οδηγήσουν σε πιο προηγμένες μεθόδους στο ερευνητικό πεδίο της σημασιολογικής κατάτμησης εικόνων.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] Ι. Βλαχάβα, Π. Κεφαλά, Ν. Βασιλειάδη, Φ. Κόκκορα, Η. Σακελλαρίου “Τεχνητή Νοημοσύνη” Δ’ Έκδοση 2020 Εκδόσεις Πανεπιστημίου Μακεδονίας
- [2] D Sharma, N Kumar A review on machine learning algorithms, tasks and applications International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 6, Issue 10, October 2017, ISSN: 2278 – 1323 1548 2017 IJARCET
- [3] Q.V.Le, M.Ranzato, R.Monga, M.Devin, K.Chen, G.Corrado, J.Dean, A.Y.Ng ‘Building high-level features using large scale unsupervised learning’ Proceedings of the International Conference on Machine Learning, 2012
- [4] J.Duchi, E.Hazan, Y.Singer ‘Adaptive subgradient methods for online learning and stochastic optimization’ Journal of Machine Learning Research 2011
- [5].Iqbal H. Sarker SN Computer Science volume 2, Article number: 160 (2021) Published: 22 March 2021 Machine Learning: Algorithms, Real-World Applications and Research Directions
- [6] Cios, Krzysztof J., et al. “Unsupervised learning: association rules. “Data Mining: A Knowledge Discovery Approach (2007): pp 289-306
- [7] Ι. Παρτάλας «Μέθοδοι Ενισχυτικής Μάθησης σε συστήματα πρακτόρων», *Διδακτορική Διατριβή*, Τμήμα Πληροφορικής, Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, 2009
- [8] SN Computer Science volume 2, Article number: 160 (2021) Πρόσβαση την 24.4.23
- [9] https://el.wikipedia.org/wiki/Νευρωνικό_δίκτυο Πρόσβαση την 24.4.23
- [10] Voulodimos, N. Doulamis, A. Doulamis, E Protopapadakis “Deep learning for computer vision: A brief review, Department of Informatics, Technological Educational Institute of Athens, 2018
- [11] A. Shrestha-A.Mahmood ‘Review of deep learning algorithms and architectures’ IEEE Access 2019 ieeexplore.ieee.org πρόσβαση την 25.4.23
- [12] <https://iguru.gr/einai-kai-pos-leitourgei-vathia-mathisi/> Πρόσβαση την 25.4.23

- [13] R.Mu, X.Zeng-KSII 'A review of deep learning research' Article 2019 koreascience.or.kr
- [14] [23] A.Garcia-Garcia, S.Orts-Escolano, S.Oprea, V. Villena-Martinez, P. Martinez-Gonzalez, J. Garcia-Rodriguez 'A survey on deep learning techniques for image and video semantic segmentation' *Article* Computing journal homepage: www.elsevier.com/locate/asoc, 2017 Πρόσβαση την 2.5.23
- [15]Hmrishav Bandyopadhyay. An Introduction to Image Segmentation: Deep Learning vs. Traditional [+Examples]. Beginner's Guide to Semantic Segmentation [2023] (v7labs.com), 2023. Πρόσβαση την : 2.5.2023
- [16]<https://divamgupta.com/image-segmentation/2019/06/06/deep-learning-semantic-segmentation-keras.html> Πρόσβαση την 3.5.2023
- [17]<https://deeplobe.ai/image-segmentation-the-most-interesting-applications> Πρόσβαση την 14.5.23
- [18]<https://keymakr.com/blog/semantic-segmentation-uses-and-applications/> Πρόσβαση την 14.5.23
- [19] Y. Mo, Y. Wu, X. Yang, F. Liu, Y. Liao 'Review the state-of-the-art technologies of semantic segmentation based on deep learning' *Article*, 2022
- [20] A. Krizhevsky, Ilya Sutskever, G.Hinton 'ImageNet classification with deep convolutional neural networks' *Advances in neural information processing systems*, 2012
- [21] <https://www.upgrad.com/blog/basic-cnn-architecture> Πρόσβαση την 25.4.23
- [22]<https://machinelearningmastery.com/choose-an-activation-function-for-deep-learning/> Πρόσβαση την 27.4.23
- [23]Χ. Παπαδόπουλος 'Βαθιά Μάθηση: Σημασιολογική κατάτμηση δορυφορικών εικόνων' *Πτυχιακή εργασία*, Τμήμα Πληροφορικής και Τηλεπικοινωνιών, Εθνικό Καποδιστριακό Πανεπιστήμιο Αθηνών, 2019
- [24]Γ.Γιδάκος, Π.Παπαδόπουλος 'Εφαρμογές τεχνητής νοημοσύνης στην Μηχανολογία- Μια μελέτη περίπτωσης τεχνητών νευρωνικών δικτύων' *Πτυχιακή εργασία* τμήμα Μηχανολόγων Μηχανικών Πανεπιστημίου Πελοποννήσου,2021

- [25] Α. Κωστόπουλος 'Θέμα Οπτική Αναγνώριση Αντικειμένων Σε Μεγάλες Βάσεις Δεδομένων' *Διπλωματική εργασία* Τμήμα Ηλεκτρολόγων Μηχανικών και Τεχνολογίας Υπολογιστών Πανεπιστημίου Πατρών, 2021
- [26] N.Srivastava, G.Hinton, A. Krizhevsky, I.Sutskever, R. Salakhutdinov 'Dropout: A Simple Way to Prevent Neural Networks from Overfitting' *Article Journal of Machine Learning Research* 15 (1): 1929-1958, 2014
- [27] N. Bjorck, CP Gomes, B. Selman, K. Weinberger 'Understanding batch normalization' Cornell University 2018
- [28] J.Long, E.Shelhamer, T.Darrell 'Fully Convolutional Networks for semantic segmentation', 2014 <https://arxiv.org/pdf/1411.4038.pdf> Πρόσβαση την 3.5.23
- [29]<https://towardsdatascience.com/understanding-semantic-segmentation-with-unet-6be4f42d4b47> πρόσβαση την 4.5.23
- [30]<https://medium.com/analytics-vidhya/what-is-unet-157314c87634> πρόσβαση την 4.5.23
- [31] O. Ronneberger, P. Fischer, and T. Brox, —U-net: Convolutional networks for biomedical image segmentation, in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015 -18th International Conference Munich, Germany, October 5 - 9, 2015, Proceedings, Part III* (N.Navab, J. Hornegger, W. M. W. III, and A. F. Frangi, eds.), vol. 9351 of *Lecture Notes in Computer Science*, pp. 234–241, Springer, 2015.
- [32]<https://towardsdatascience.com/the-evolution-of-deeplab-for-semantic-segmentation-95082b025571> πρόσβαση την 6.5.23
- [33] <http://liangchiehchen.com/projects/DeepLab.html> πρόσβαση την 6.5.23
- [34] https://keras.io/examples/vision/deeplabv3_plus/ πρόσβαση την 7.5.23
- [35] <https://modelzoo.co/model/deeplab> πρόσβαση την 6.5.23
- [36]<https://medium.com/@mygreatlearning/everything-you-need-to-know-about-vgg16-7315defb5918> Πρόσβαση την 30.5.23
- [37] K.Simonyan, A.Zisserman ' Very Deep Convolutional Networks for Large-Scale image recognition' Visual Geometry Group, Department of Engineering Science, University of Oxford <https://arxiv.org/pdf/1409.1556.pdf>

- [38] <https://www.androidpolice.com/google-colab-explainer/> πρόσβαση την 9.5.23
- [39] <https://en.wikipedia.org/wiki/TensorFlow#History> πρόσβαση την 9.5.23
- [40] <https://keras.io/> πρόσβαση την 9.5.23
- [41] C.Harris, K.Millman, S.Van der Walt, R.Gommers.....T.Oliphant 'Array Programming with NumPy' 2020, Nature 585, 357-362 <https://doi.org/10.1038/s41586-020-2649-2> Πρόσβαση την 9.5.23
- [42] Matplotlib.org Πρόσβαση την 10.5.23
- [43] <https://www.learn datasci.com/tutorials/python-pandas-tutorial-complete-introduction-for-beginners/> Πρόσβαση την 10.5.23
- [44] <https://www.superannotate.com/blog/guide-to-semantic-segmentation> Πρόσβαση την 30.5.23
- [45] Diederik P. Kingma & Jimmy Lei Ba 'ADAM: A Method for Stochastic Optimization' <https://Theberkeleyview.wordpress.com/2015/11/19/Berkeleyview-for-ADAM-a-method-for-stochastic-optimization> Πρόσβαση την 30.5.23
- [46] <https://pycad.co/the-difference-between-dice-and-dice-loss> Πρόσβαση την 30.5.23
- [47] <https://towardsdatascience.com/metrics-to-evaluate-your-semantic-segmentation-model-6bcb99639aa2> Πρόσβαση την 30.5.23
- [48] <https://www.v7labs.com/blog/f1-score-guide> Πρόσβαση την 30.5.23
- [49] B.Zhou, H.Zhao, X. Puig, T.Xiao, S. Fidler , A. Barriuso, A. Torralba 'Semantic Understanding of Scenes through the ADE20K Dataset',2018 Arxiv.1608.05442v2 [cs.CV]2018
- [50] <https://datagen.tech/guides/computer-vision/resnet-50/> Πρόσβαση την 17.6.23
- [51] Ajitesh Kumar 'Keras-Categorical Cross Entropy Loss Function' 2020 <https://vitalflux.com/keras-categorical-cross-entropy-loss-function/> Πρόσβαση την 17.6.23
- [52] <https://www.sciencedirect.com/science/article/abs/pii/S0925231222000054> Πρόσβαση την 3.5.23