ΔΙΕΘΝΕΣ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΤΗΣ ΕΛΛΑΔΟΣ

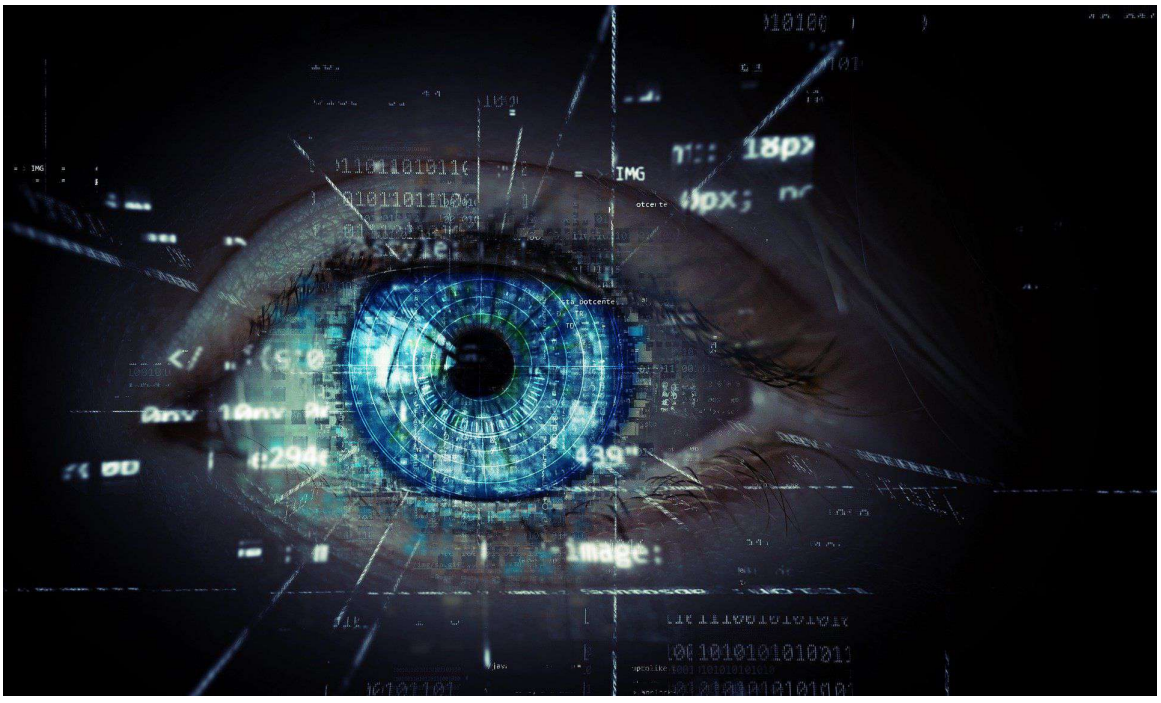
ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ

ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΗΛΕΚΤΡΟΝΙΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Σύγκριση αρχιτεκτονικών βαθιάς μάθησης για την κατάτμηση εικόνων



Του φοιτητή
Μερσίνη Χρήστου
Αρ. Μητρώου: 174985

Επιβλέπων
Δρ. Διαμαντάρας Κωνσταντίνος
Καθηγητής

23 Ιανουαρίου 2026

Τίτλος Δ.Ε. Σύγκριση αρχιτεκτονικών βαθιάς μάθησης για την κατάτμηση εικόνων
Κωδικός Δ.Ε. ....
Ονοματεπώνυμο φοιτητή Μερσίνης Χρήστος
Ονοματεπώνυμο εισηγητή Διαμαντάρας Κωνσταντίνος
Ημερομηνία ανάληψης Δ.Ε. ....
Ημερομηνία περάτωσης Δ.Ε. ....

Βεβαιώνω ότι είμαι ο συγγραφέας αυτής της εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, έχω καταγράψει τις όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών, εικόνων και κειμένου, είτε αυτές αναφέρονται ακριβώς είτε παραφρασμένες. Επιπλέον, βεβαιώνω ότι αυτή η εργασία προετοιμάστηκε από εμένα προσωπικά, ειδικά ως διπλωματική εργασία, στο Τμήμα Μηχανικών Πληροφορικής και Ηλεκτρονικών Συστημάτων του ΔΙ.ΠΑ.Ε.

Η έγκριση της διπλωματικής εργασίας από το Τμήμα Μηχανικών Πληροφορικής και Ηλεκτρονικών Συστημάτων του Διεθνούς Πανεπιστημίου της Ελλάδος, δεν υποδηλώνει απαραιτήτως και αποδοχή των απόψεων του συγγραφέα, εκ μέρους του Τμήματος.

«Στους αγαπημένους μου»

Πρόλογος

Η επιλογή του συγκεκριμένου θέματος προέκυψε από το έντονο ενδιαφέρον και ενθουσιασμό μου για το πεδίο της Υπολογιστικής Όρασης και, ειδικότερα, για τις μεθόδους Βαθιάς Μάθησης που το υποστηρίζουν. Στόχος μου είναι να εμβαθύνω θεωρητικά και πειραματικά στους αλγορίθμους βαθιάς μάθησης, κατανοώντας σε βάθος τα πλεονεκτήματα, τους περιορισμούς και τα κριτήρια σύγκρισης διαφορετικών μοντέλων. Παράλληλα, η ενασχόλησή μου με την πανεπιστημιακή ομάδα *drAIve*, η οποία δραστηριοποιείται σε εφαρμογές τεχνητής νοημοσύνης και αυτόνομης οδήγησης, ενίσχυσε την απόφασή μου, καθώς εκτιμώ ότι τα ευρήματα της παρούσας εργασίας μπορούν να συμβάλουν ουσιαστικά στην ερευνητική δραστηριότητα της ομάδας. Η εργασία αυτή, επομένως, λειτουργεί συνδυαστικά ως μέσο εμβάθυνσης σε ένα σύγχρονο και ραγδαία εξελισσόμενο επιστημονικό πεδίο και ως βάση για την παραγωγή εφαρμόσιμης γνώσης σε πραγματικά προβλήματα Υπολογιστικής Όρασης.

Περίληψη

Η παρούσα διπλωματική εργασία πραγματοποιεί συγκριτική αξιολόγηση τριών σύγχρονων αρχιτεκτονικών βαθιάς μάθησης για σημασιολογική τμηματοποίηση εικόνων: U-Net, DeepLabV3+ και SegNet. Η σημασιολογική τμηματοποίηση, αποτελεί θεμελιώδη εργασία υπολογιστικής όρασης με κρίσιμες εφαρμογές στην ιατρική απεικόνιση, την αυτόνομη οδήγηση, τη γεωργία και την εναέρια παρακολούθηση. Μετά από εκτενή θεωρητική ανασκόπηση που καλύπτει την εξέλιξη από κλασικές τεχνικές τμηματοποίησης έως τις σύγχρονες αρχιτεκτονικές, όπως FCN, PSPNet και Vision Transformers, η πειραματική μεθοδολογία βασίστηκε στο σύνολο δεδομένων Cityscapes με 5.000 εικόνες υψηλής ποιότητας από αστικές σκηνές και 30 κλάσεις αντικειμένων. Τα μοντέλα αξιολογήθηκαν με τυποποιημένες μετρικές, Intersection over Union (IoU), F1 Score και Pixel Accuracy, υλοποιημένα στο PyTorch framework υπό ελεγχόμενες παραμέτρους εκπαίδευσης. Τα πειραματικά αποτελέσματα αποδεικνύουν την ανωτερότητα του U-Net (Dice: 0.8127, Mean IoU: 0.6844) έναντι του DeepLabV3+ (Dice: 0.8016, Mean IoU: 0.669) και του SegNet (Dice: 0.7915, Mean IoU: 0.6549). Η υπεροχή του U-Net αποδίδεται στις συνδέσεις παράκαμψης που διατηρούν λεπτομερείς χωρικές πληροφορίες, επιτρέποντας ανώτερη απόδοση σε μικρά αντικείμενα όπως του Human (F1: 0.6622) και object (F1: 0.3261). Το DeepLabV3+ υπερτερεί σε μεγάλα ομοιογενή αντικείμενα, Flat (F1: 0.9545) και Sky (F1: 0.9238) χάρη στην αρχιτεκτονική ASPP, ενώ το SegNet παρουσιάζει υπολογιστική αποδοτικότητα με περιορισμένη χωρική ανακατασκευή. Η εργασία παρέχει πρακτικές κατευθύνσεις για επιλογή αρχιτεκτονικής βάσει εφαρμογής και προτείνει μελλοντικές κατευθύνσεις όπως επέκταση συνόλων δεδομένων, υβριδικές αρχιτεκτονικές και βελτιστοποίηση πραγματικού χρόνου.

Comparison of deep learning architectures for image segmentation

Mersinis Christos

Abstract

This thesis presents a comparative evaluation of three state-of-the-art deep learning architectures for semantic image segmentation: U-Net, DeepLabV3+, and SegNet. Semantic segmentation constitutes a fundamental computer vision task with critical applications in medical imaging, autonomous driving, precision agriculture, and aerial surveillance. Following an extensive theoretical review spanning classical segmentation techniques to contemporary architectures including FCN, PSPNet, and Vision Transformers, the experimental methodology employs the Cityscapes dataset comprising 5,000 high-quality urban scene images across 30 object classes. Models were evaluated using standardized metrics including Intersection over Union (IoU), F1 Score, and Pixel Accuracy, implemented in PyTorch framework under controlled training parameters. Experimental results demonstrate the superiority of U-Net (Dice: 0.8127, Mean IoU: 0.6844) over DeepLabV3+ (Dice: 0.8016, Mean IoU: 0.669) and SegNet (Dice: 0.7915, Mean IoU: 0.6549). U-Net's superior performance is attributed to its skip connections that preserve detailed spatial information during upsampling, enabling enhanced accuracy for small objects such as Human (F1: 0.6622) and Object (F1: 0.3261) classes. DeepLabV3+ excels in large homogeneous objects like Flat (F1: 0.9545) and Sky (F1: 0.9238) due to its ASPP architecture, while demonstrating limitations in classes requiring fine spatial precision. SegNet ranks third in overall performance, offering memory efficiency advantages through pooling indices but exhibiting restricted spatial reconstruction capabilities, particularly for complex classes. These findings provide practical guidelines for architecture selection based on application-specific requirements, with recommendations for future research including dataset expansion, hybrid architecture exploration, and real-time optimization techniques.

Ευχαριστίες

Ολοκληρώνοντας την παρούσα διπλωματική εργασία, θα ήθελα να εκφράσω τις ειλικρινείς μου ευχαριστίες σε όσους συνέβαλαν, με τον δικό τους μοναδικό τρόπο, στην επιτυχή ολοκλήρωσή της. Πρωτίστως, οφείλω ένα μεγάλο ευχαριστώ στον επιβλέποντα καθηγητή μου, Δρ. Διαμαντάρα Κωνσταντίνο, για την πολύτιμη καθοδήγηση, την επιστημονική του αρτιότητα και την εμπιστοσύνη που μου έδειξε. Οι εύστοχες παρατηρήσεις του και η συνεχής υποστήριξή του υπήρξαν καθοριστικής σημασίας για την πορεία της έρευνάς μου. Ευχαριστώ θερμά τη μητέρα μου, για την αμέριστη αγάπη, την ηθική στήριξη και την ενθάρρυνση που μου προσέφερε καθ'όλη τη διάρκεια εκπόνησης της διπλωματικής μου. Χωρίς την ακλόνητη πίστη της στις δυνατότητές μου, και τη συνεχή συναισθηματική της υποστήριξη, αυτό το ταξίδι θα ήταν πολύ πιο δύσκολο. Ένα ξεχωριστό ευχαριστώ στην σύντροφό μου, Αθανασία, για την απέραντη υπομονή, τη κατανόηση και την αγάπη της, ιδίως κατά τις απαιτητικές περιόδους της συγγραφής. Η παρουσία της αποτέλεσε για εμένα πηγή ισορροπίας και έμπνευσης. Τέλος, θα ήθελα να ευχαριστήσω τον καλό μου φίλο, Στάθη, για τη σταθερή του φιλία, τις γόνιμες συζητήσεις και την αισιοδοξία του στις ικανότητες μου, που με βοήθησαν να ξεπεράσω κάθε εμπόδιο και ανασφάλεια.

Περιεχόμενα

Κατάλογος Σχημάτων

Κατάλογος Πινάκων

Συντομογραφίες

FCN	Fully Convolutional Networks
CNN	Convolutional Neural Networks
IoU	Intersection Over Union
UAV	Unmanned Aerial Vehicle
SGD	Stochastic Gradient Descent
ReLU	Leaky Rectified Linear Unit
MLP	Multi Layer Perceptron
RGB	Red Green Blue
BN	Batch Normalization
RNN	Recurrent Neural Network
LSTM	Long Short-Term Memory
GAN	Generative Adversarial Network
PCA	Principal Component Analysis
VGG	Visual Geometry Group
ASPP	Atrous Spatial Pyramid Pooling
CRF	Conditional Random Field
ViTs	Visiion Transformers
GPU	Graphics Processing Units
PA	Pixel Accuracy
MAE	Mean Absolute Error

Κεφάλαιο 1ο: Εισαγωγή στην Σημασιολογική τμηματοποίηση

1.1 Εισαγωγή

Η σημασιολογική κατάτμηση εικόνας (semantic segmentation) αποτελεί μια από τις πιο απαιτητικές και κρίσιμες εργασίες στον τομέα της υπολογιστικής όρασης, καθώς στοχεύει στην κατηγοριοποίηση κάθε εικονοστοιχείου (pixel) μιας εικόνας σε προκαθορισμένες κλάσεις. Η δυνατότητα ακριβούς εντοπισμού και ερμηνείας αντικειμένων σε επίπεδο pixel έχει ζωτική σημασία για πλήθος εφαρμογών, όπως η αυτόνομη οδήγηση, η ανάλυση ιατρικών εικόνων, η παρακολούθηση γεωχωρικών δεδομένων και η υπολογιστική όραση.

Τα τελευταία χρόνια, οι αλγόριθμοι βαθιάς μάθησης (deep learning), και ειδικότερα τα συνελικτικά νευρωνικά δίκτυα (CNNs), έχουν επιφέρει σημαντικές προόδους στην ακρίβεια και στην αποδοτικότητα των συστημάτων σημασιολογικής κατάτμησης. Αρχιτεκτονικές όπως το U-Net, το DeepLab, και το SegNet, προσφέρουν δυνατότητες ακριβούς κατηγοριοποίησης αντικειμένων με υψηλό επίπεδο χωρικής πληροφορίας. Ωστόσο, κάθε αλγόριθμος παρουσιάζει πλεονεκτήματα και περιορισμούς ανάλογα με τη φύση των δεδομένων, τη διαχείριση των περιθωρίων των αντικειμένων, τη γενίκευση σε άγνωστες εικόνες και την υπολογιστική του αποδοτικότητα.

Η παρούσα εργασία επικεντρώνεται στην αναλυτική μελέτη και αξιολόγηση σύγχρονων deep learning αλγορίθμων για την κατάτμηση εικόνων. Πιο συγκεκριμένα, συγκρίνονται μοντέλα βάσει της ακρίβειας, της ικανότητας γενίκευσης, της πολυπλοκότητας, καθώς και της συμπεριφοράς τους σε διαφορετικά σύνολα δεδομένων. Η ανάλυση πραγματοποιείται με τη βοήθεια κατάλληλων μετρικών αξιολόγησης, όπως το Intersection over Union (IoU).

Μέσα από τη συστηματική μελέτη αυτών των μοντέλων, η εργασία φιλοδοξεί να συμβάλει στην κατανόηση των δυνατοτήτων και των περιορισμών τους, και να παρέχει πρακτικές κατευθύνσεις για την επιλογή του κατάλληλου αλγορίθμου σε εφαρμογές με διαφορετικές απαιτήσεις.

Οι επόμενες ενότητες περιλαμβάνουν ανασκόπηση της σχετικής βιβλιογραφίας, περιγραφή της μεθοδολογίας που ακολουθήθηκε, παρουσίαση των πειραματικών αποτελεσμάτων και συζήτηση των ευρημάτων, με στόχο την εξαγωγή χρήσιμων συμπερασμάτων για το πεδίο της σημασιολογικής κατατμησης μέσω βαθιάς μάθησης

1.2 Εφαρμογές Σημασιολογικής Τμηματοποίησης

Η τμηματοποίηση εικόνας διαδραματίζει καθοριστικό ρόλο σε διάφορους τομείς, όπου ο στόχος είναι η κατάτμηση εικόνων σε νοηματικές περιοχές, καθιστώντας την αναγνώριση αντικειμένων πιο αποδοτική. Σε αυτή την ενότητα, εξερευνούμε τους κύριους τομείς εφαρμογής όπου οι τεχνικές τμηματοποίησης εικόνας έχουν ασκήσει σημαντική επιρροή. Κάθε υποενότητα θα παρέχει μια επισκόπηση του τομέα και θα συζητά πώς οι τεχνικές τμηματοποίησης συμβάλλουν

στην πρόοδο του εκάστοτε πεδίου.

1.2.1 Ιατρική

Η ιατρική απεικόνιση αποτελεί έναν από τους πιο σημαντικούς τομείς εφαρμογής της τμηματοποίησης εικόνας. Η ακριβής τμηματοποίηση ανατομικών δομών σε ιατρικές εικόνες είναι κρίσιμη για τη διάγνωση, τον σχεδιασμό θεραπειών και την παρακολούθηση ασθενειών. Η τμηματοποίηση βοηθά στην αναγνώριση και διαφοροποίηση οργάνων, ιστών και βλαβών στην αξονική τομογραφία (CT) και τη μαγνητική τομογραφία (MRI). Αυτό επιτρέπει στους ακτινολόγους να διαγνώσουν με ακρίβεια καταστάσεις όπως όγκους, κατάγματα και εσωτερική αιμορραγία [48].

Ένας άλλος αναπτυσσόμενος τομέας στην ιατρική απεικόνιση είναι η τμηματοποίηση καρδιακής μαγνητικής τομογραφίας (MRI), όπου τα μοντέλα χρησιμοποιούνται για την τμηματοποίηση των θαλάμων της καρδιάς, επιτρέποντας λεπτομερή ανάλυση καρδιακών παθήσεων. Τεχνικές όπως το U-Net έχουν σημειώσει μεγάλη επιτυχία σε αυτόν τον τομέα, και η χρήση τους έχει επεκταθεί σε διάφορες μεθόδους απεικόνισης, συμπεριλαμβανομένης της απεικόνισης του αμφιβληστροειδούς για την ανίχνευση διαβητικής αμφιβληστροειδοπάθειας. [48].

1.2.2 Αυτόνομη Οδήγηση

Τα συστήματα αυτόνομης οδήγησης βασίζονται σε μεγάλο βαθμό στην τμηματοποίηση σε πραγματικό χρόνο για ασφαλή πλοήγηση. Για παράδειγμα, η ανίχνευση δρόμων και λωρίδων κυκλοφορίας είναι μία από τις κύριες εργασίες όπου χρησιμοποιούνται μοντέλα τμηματοποίησης [47] όπως τα FCNs (Πλήρως Συνελικτικά Δίκτυα) για να διακρίνουν τις οδικές επιφάνειες που είναι προσπελάσιμες από εκείνες που δεν είναι. Η τμηματοποίηση διαδραματίζει επίσης κρίσιμο ρόλο στην ανίχνευση πεζών και στην αποφυγή αντικειμένων. Αυτό ισχύει ιδιαίτερα για συστήματα όπως το Mask R-CNN, το οποίο μπορεί να εκτελεί instance segmentation για να διακρίνει μεταξύ πολλών δυναμικών αντικειμένων στην κυκλοφορία.

1.2.3 Εναέρια απεικόνιση

Στον τομέα της δορυφορικής και εναέριας απεικόνισης, τα μοντέλα τμηματοποίησης εφαρμόζονται για την ταξινόμηση καλύψεων γης, την ανίχνευση αλλαγών στη βλάστηση, στους υδάτινους όγκους και στις αστικές δομές. Αυτό είναι ιδιαίτερα χρήσιμο για την περιβαλλοντική παρακολούθηση και τον αστικό σχεδιασμό. Για παράδειγμα, η απεικόνιση με χρήση UAV (μη επανδρωμένα αεροσκάφη) χρησιμοποιείται συχνά στη γεωργία ακριβείας, όπου μοντέλα όπως το AgriSegNet [4] εκτελούν τμηματοποίηση σε εναέριες εικόνες για την παρακολούθηση της υγείας των καλλιεργειών, την ανίχνευση ασθενειών και τη διαχείριση της άρδευσης με μεγαλύτερη αποτελεσματικότητα.

Επιπλέον, τα μοντέλα μηχανικής μάθησης για απεικόνιση με UAV χρησιμοποιούνται όλο και

περισσότερο για την παρακολούθηση αλλαγών στην κάλυψη γης ή για την ανίχνευση περιοχών που έχουν πληγεί από φυσικές καταστροφές, όπως πλημμύρες ή δασικές πυρκαγιές. Σε μία ανασκόπηση, ο Zualkernan και οι συνεργάτες του διερεύνησαν πώς τα UAVs, σε συνδυασμό με τεχνικές μηχανικής μάθησης, συμβάλλουν στην ταυτοποίηση και αξιολόγηση της υγείας της βλάστησης από εναέριες εικόνες, συνεισφέροντας σε πιο αποτελεσματικές στρατηγικές περιβαλλοντικής διατήρησης.

1.2.4 Γεωργία

Στη γεωργία, η τμηματοποίηση χρησιμοποιείται για εργασίες όπως η ανίχνευση ζιζανίων, τη παρακολούθηση της υγείας των καλλιεργειών και στην εκτίμηση της γεωργικής απόδοσης. Για παράδειγμα, οι Fawakherji και συνεργάτες [23] διερεύνησαν μία προσέγγιση τμηματοποίησης σε επίπεδο εικονοστοιχείου, που επιτρέπει την ακριβή ταξινόμηση ζιζανίων και καλλιεργειών, βοηθώντας στη μείωση της ανάγκης για ζιζανιοκτόνα και στην βελτίωση της σοδειάς.

Η ενσωμάτωση ρομποτικών συστημάτων στη γεωργία γίνεται επίσης όλο και πιο διαδεδομένη [40]. Αυτά τα ρομπότ, εξοπλισμένα με κάμερες και αισθητήρες, χρησιμοποιούν μοντέλα τμηματοποίησης για να περιηγούνται αυτόνομα στις γεωργικές εκτάσεις, να παρακολουθούν την ανάπτυξη των φυτών και να εκτελούν εργασίες όπως ο ψεκασμός και η συγκομιδή. Μελέτες έχουν δείξει ότι τα ρομπότ μπορούν να βελτιώσουν σημαντικά την παραγωγικότητα, παρέχοντας εικόνες υψηλής ποιότητας για τη λήψη αποφάσεων βασισμένων στην τμηματοποίηση.

1.3 Ανάλυση προβλήματος

Παρά τις σημαντικές προόδους στον τομέα της τμηματοποίησης εικόνας, εξακολουθούν να υπάρχουν πολυάριθμες προκλήσεις στην εφαρμογή της σε σύνθετα και μεταβαλλόμενα περιβάλλοντα. Η υψηλή ακρίβεια και η αξιοπιστία είναι δύσκολο να επιτευχθούν σε πραγματικές συνθήκες, εξαιτίας παραγόντων όπως ο θόρυβος, οι αποκρύψεις αντικειμένων, οι διαφορετικές συνθήκες φωτισμού, δηλαδή λόγω της ετερογένειας των σκηνών. Οι περιορισμοί αυτοί, καθιστούν την ανάπτυξη αποδοτικών και αλγορίθμων κατάτμησης ικανών να γενικεύουν, μια ανοικτή πρόκληση στον τομέα της υπολογιστικής όρασης.

Στην ιατρική απεικόνιση, η σημασιολογική κατάτμηση είναι κρίσιμη για τον εντοπισμό παθολογικών περιοχών και την υποστήριξη της κλινικής διάγνωσης. Ωστόσο, η ποιότητα των εικόνων (χαμηλή αντίθεση, θόρυβος) και η ανάγκη για ακρίβεια σε επίπεδο εοικονοστοιχείου, καθιστούν το έργο δύσκολο. Επιπλέον, η χειροκίνητη σήμανση των δεδομένων (image annotation), που είναι απαραίτητη για την εκπαίδευση των αλγορίθμων, είναι εξαιρετικά χρονοβόρα και επιρρεπής σε υποκειμενικά σφάλματα.

Στον τομέα της αυτόνομης οδήγησης, απαιτείται σημασιολογική κατάτμηση σε πραγματικό χρόνο για την αναγνώριση δρόμων, πεζών, οχημάτων και εμποδίων. Ωστόσο, η απαίτηση για

υψηλή ακρίβεια σε πραγματικό χρόνο, υπό διαφορετικές και απρόβλεπτες συνθήκες (όπως βροχή, σκιές ή χαμηλός φωτισμός), δημιουργεί σημαντικές τεχνικές προκλήσεις, τόσο σε επίπεδο μοντελοποίησης όσο και υπολογιστικής απόδοσης.

Ανάλογες δυσκολίες παρουσιάζονται και σε τομείς όπως η γεωργία ακριβείας και η δορυφορική απεικόνιση, όπου οι αλγόριθμοι κατάτμησης καλούνται να διαχειριστούν μεγάλο όγκο δεδομένων, συχνά με διαφορετικά χαρακτηριστικά και χαμηλή ποιότητα. Η ποικιλομορφία των εικόνων και των εφαρμογών απαιτεί μοντέλα με υψηλή ικανότητα γενίκευσης, τα οποία να μπορούν να προσαρμοστούν σε νέα δεδομένα χωρίς σημαντική πτώση της απόδοσης.

Παρότι οι αρχιτεκτονικές βαθιάς μάθησης όπως τα U-Net, DeepLab και SegNet έχουν επιτύχει εντυπωσιακά αποτελέσματα σε πολλές εργασίες κατάτμησης, κάθε μία παρουσιάζει διαφορετικά χαρακτηριστικά, πλεονεκτήματα και περιορισμούς. Ορισμένα μοντέλα υπερέχουν σε ακρίβεια, άλλα σε ταχύτητα, ενώ άλλα παρουσιάζουν ισορροπία μεταξύ των δύο. Συνεπώς, προκύπτει η ανάγκη για συστηματική ανάλυση, συγκριτική αξιολόγηση και κατανόηση των δυνατοτήτων αυτών των μοντέλων, ώστε να μπορούν να αξιοποιηθούν σωστά στα διαφορετικά σενάρια που μπορούν να προκύψουν.

Η παρούσα εργασία εστιάζει ακριβώς σε αυτό το πρόβλημα: στη συγκριτική μελέτη και αξιολόγηση σύγχρονων deep learning αλγορίθμων για την κατάτμηση εικόνων, σε διαφορετικούς τομείς εφαρμογής. Η έρευνα στοχεύει στο να εντοπίσει τα χαρακτηριστικά εκείνα που καθιστούν έναν αλγόριθμο πιο κατάλληλο από έναν άλλο, αναλόγως των απαιτήσεων της εκάστοτε εφαρμογής.

Κεφάλαιο 2ο: Θεωρητικό Υπόβαθρο

2.1 Ιστορική Αναδρομή

Η έννοια της εφεύρεσης μιας μηχανής η οποία έχει την ικανότητα της σκέψης, υπήρχε από την αρχαιότητα. Συγκεκριμένα, στην ελληνική μυθολογία, ο Τάλως ήταν ένα τεράστιο γιγαντιαίο άγαλμα κατασκευασμένο από τον θεό Ήφαιστο το οποίο διακατείχε νοητικές ικανότητες. Με την εφεύρεση των υπολογιστών, και το διάσημο ρητό του Alan Turing — τον εφευρέτη του πρώτου λειτουργικού ηλεκτρονικού υπολογιστή — «μπορούν οι μηχανές να σκεφτούν;» οι συζητήσεις στην επιστημονική κοινότητα για την ικανότητα των μηχανών να αναπαράγουν την ανθρώπινη (αποκτήσουν) νοημοσύνη ξεκίνησαν. Αργότερα, το 1956, επινοήθηκε ο όρος «τεχνητή νοημοσύνη» και ξεκίνησαν οι έρευνες με στόχο την εφεύρεση μηχανών, αρκετά ικανές έτσι ώστε να μιμούνται την ανθρώπινη νοημοσύνη. Το 1997, το πρόγραμμα «Deep Blue» της IBM κατάφερε να νικήσει τον παγκόσμιο πρωταθλητή σκακιού, Gary Kasparov. (αποδεικνύοντας τα λεγόμενα του μαθηματικού James Lighthill, — η ικανότητα των μηχανών να μαθαίνουν θα φτάσει μέχρι το επίπεδο ενός ερασιτέχνη— λανθασμένα). Στην πραγματικότητα, όλη αυτή η έρευνα που οδήγησε στην ανάπτυξη αλγορίθμων πανομοιότυπου με τον «Deep Blue», ονομάζονται συστήματα εμπειρογνώμων (System Experts). Αυτά τα συστήματα έχουν την ικανότητα λύσης ντετερμινιστικών προβλημάτων και η δυνατότητα τους σταματάει εκεί. Η πραγματική νοημοσύνη, βρίσκεται στην ικανότητα μάθησης πληροφοριών και στην εφαρμογή προσαρμοστικότητα, με σκοπό τη λύση πανομοιότυπων προβλημάτων, όχι στην ανάπτυξη ενός τεράστιου « αν ... αλλιώς .... » αλγορίθμου. Στην δεκαετία του 1940, ο Καναδός ψυχολόγος, Donald Hebb ανέπτυξε τη θεωρία πως η πραγματική τεχνητή νοημοσύνη βρίσκεται στην αναπαραγωγή του τρόπου λειτουργίας των νευρώνων στον ανθρώπινο εγκέφαλο. Αυτή η θεωρία, είχε ως αποτέλεσμα την έρευνα στα τεχνητά νευρωνικά δίκτυα, ένα πολύ σημαντικό κεφάλαιο της μηχανικής μάθησης, καθώς αποτελεί το βασικό θεμέλιο της τεχνητής νοημοσύνης που συναντάμε στη καθημερινότητα. Το 1969 όμως, η έρευνα πάνω στα τεχνητά νευρωνικά σταμάτησε, καθώς διαπίστωσαν ότι δεν έχουν στη διάθεση τους την υπολογιστική ισχύ που απαιτούσε αυτή η τεχνολογία. Αργότερα το 2015 δημιουργήθηκε ένα υποπεδίο των τεχνητών νευρωνικών δικτύων γνωστό ως «Βαθιά μάθηση» (Deep learning) το οποίο επανέφερε το επίκεντρο της έρευνας ξανά πάνω στα τεχνητά νευρωνικά δίκτυα, καθώς η δυνατότητα αυτής της τεχνολογίας φαινόταν πολλά υποσχόμενη. της Έξυπνοι αλγόριθμοι φωτογραφιών στα κινητά, σύνθεση και κατανόηση ομιλίας, αλλά και σε πιο εξειδικευμένες εφαρμογές όπως την δικτύωση υπολογιστών (δικτύωση καθορισμένη από λογισμικό) [30].

2.2 Μηχανική Μάθηση

Η μηχανική μάθηση, πιο συγκεκριμένα, η βαθιά μάθηση (Deep Learning), αποτελεί το επίκεντρο αυτής της διπλωματικής εργασίας καθώς έχει αποδειχθεί η αποδοτικότερη τεχνική στην αναγνώριση και κατηγοριοποίηση αντικειμένων από εικόνες. Για τη σωστότερη κατανόηση της

βαθιάς μάθησης θα πραγματοποιηθεί ανασκόπηση των βασικότερων θεμελίων που οδήγησαν στην δημιουργία του υποπεδίου αυτου.

Η μηχανική μάθηση αντιπροσωπεύει ένα καίριο τομέα έρευνας και ανάπτυξης που επιδιώκει να παρέχει στους υπολογιστές τη δυνατότητα να μαθαίνουν από τα δεδομένα και να προβλέπουν μελλοντικές πληροφορίες χωρίς να χρειάζεται ανθρώπινη παρέμβαση. Η έρευνα στη μηχανική μάθηση επικεντρώνεται στην ανάπτυξη αλγορίθμων και μοντέλων που είναι ικανά να αντλούν συμπεράσματα από τα δεδομένα, επιτρέποντας την αυτόματη προσαρμογή στις νέες πληροφορίες. Κλασικές μεθόδοι, όπως οι τεχνικές μάθησης με επίβλεψη και μάθησης χωρίς επίβλεψη, αποτελούν καίρια τμήματα του πεδίου. Ενδεικτικές εργασίες που συμβάλλουν στην κατανόηση και την εξέλιξη αυτών των μεθόδων περιλαμβάνουν την έρευνα του Bishop [13] στη βαθιά μάθηση, την ανάλυση των αλγορίθμων στην εργασία του Hastie και συνεργάτες. [32], καθώς και τη σημαντική έρευνα του LeCun και συνεργάτες. [38] στον τομέα της βαθιάς μάθησης.

2.2.1 Μάθηση με Επίβλεψη

Υπάρχουν διάφοροι αλγόριθμοι εκπαίδευσης ενός μοντέλου μηχανικής μάθησης. Η μάθηση με επίβλεψη, είναι από τους δημοφιλέστερους αλγορίθμους, καθώς, κατά κανόνα χρησιμοποιούνται στα τεχνητά νευρωνικά δίκτυα [6]. Για να είναι εφικτή η εκπαίδευση του, είναι απαραίτητη η ύπαρξη ενός συνόλου δεδομένων ταξινομημένο σε κλάσεις. Σκοπός του, είναι η σωστή κατηγοριοποίηση άγνωστων αντικειμένων στις κλάσεις του συνόλου δεδομένων. Η αποτελεσματικότητα αυτών των αλγορίθμων, εξαρτάται από το την ποιότητα του εκάστοτε συνόλου δεδομένων που υπάρχει στη διάθεση του. Αναλυτικότερα, η εκπαίδευση, υλοποιείται περνώντας σαν είσοδο όλα τα δεδομένα που έχουμε στη διάθεση μας στο τεχνικό νευρωνικό δίκτυο, με αποτέλεσμα να εξάγει τα κοινά χαρακτηριστικά μεταξύ των κλάσεων. Κατά κανόνα, η τεχνική εκπαίδευσης που εφαρμόζεται στα μοντέλα βαθιάς μάθησης είναι η μάθηση με επίβλεψη.

2.2.2 Μάθηση Χωρίς Επίβλεψη

Μία άλλη κατηγορία αλγορίθμων, είναι οι αλγόριθμοι μάθησης χωρίς επίβλεψη. Αντιθέτως, με την προηγούμενη κατηγορία αλγορίθμων, δεν είναι η απαραίτητη η ύπαρξη συνόλου δεδομένων κατηγοριοποιημένο σε κλάσεις καθώς ο αλγόριθμος αποφασίζει τον τρόπο κατηγοριοποίησης του συνόλου δεδομένων. [41] Μία από τις δημοφιλέστερες τεχνικές σε αυτή την κατηγορία αλγορίθμων, είναι το λεγόμενο Clustering. Αναλυτικότερα, Αυτή η μέθοδος, εξάγει τα κοινά χαρακτηριστικά των δεδομένων και τα ομαδοποιεί σε γκρουπς, γνωστά ως clusters. Ένας από τους δημοφιλέστερους αλγορίθμους αυτής της κατηγορίας είναι ο k-means algorithm.

2.2.3 Υπερπαράμετροι

Οι υπερπαράμετροι αποτελούν κρίσιμο στοιχείο στην επιτυχημένη εφαρμογή της μηχανικής μάθησης. Αυτές οι παράμετροι προσδιορίζουν τη δομή και τη συμπεριφορά του μοντέλου,

καθορίζοντας την απόδοσή του κατά την εκπαίδευση και την εκτέλεση. Η επιλογή κατάλληλων υπερπαραμέτρων απαιτεί προσοχή, καθώς η μη σωστή ρύθμισή τους μπορεί να οδηγήσει σε υποεκπαίδευση (underfitting) ή υπερεκπαίδευση (overfitting) του μοντέλου. Η έρευνα του Bergstra και Bengio [12] ασχολείται με τη βελτιστοποίηση των υπερπαραμέτρων μέσω τυχαίας αναζήτησης έναντι των επικρατούμενων μεθόδων και αποδεικνύει πως η τυχαία αναζήτηση των υπερπαραμέτρων φέρνει καλύτερα αποτελέσματα. Μερικές από τις παραμέτρους είναι:

1. **Ρυθμός Μάθησης (Learning Rate)**: Ορίζει το μέγεθος των βημάτων που κάνει το μοντέλο κατά την εκπαίδευση. Ένας πολύ χαμηλός ρυθμός μάθησης μπορεί να οδηγήσει σε αργή σύγκλιση, ενώ ένας υψηλός μπορεί να προκαλέσει αστάθεια
2. **Παράμετροι Επίλυσης (Solver Parameters)**: Στο πλαίσιο αλγορίθμων βαθιάς μάθησης, όπως ένα νευρωνικό δίκτυο, οι παράμετροι επίλυσης όπως η βαρύτητα (weight decay) και η ορμή (momentum) επηρεάζουν τη σύγκλιση και την απόδοση του μοντέλου.
3. **Παράμετροι Κανονικοποίησης (Regularization Parameters)**: Παράμετροι όπως η πειραματική κανονικοποίηση (L1/L2 regularization) μπορούν να χρησιμοποιηθούν για τη μείωση της υπερεκπαίδευσης.

Συνεπώς, η εύστοχη επιλογή των υπερπαραμέτρων αποτελεί θεμέλιο πυλώνα για την αποτελεσματική εφαρμογή της μηχανικής μάθησης.

2.2.4 Υπερεκπαίδευση

Ένα από τα δημοφιλέστερα προβλήματα στην εκπαίδευση με επίβλεψη ενός νευρωνικού δικτύου, είναι η λεγόμενη υπερεκπαίδευση ή overfitting. Το overfitting, συναντάται όταν το νευρωνικό δίκτυο έχει αποστηθίσει τα δεδομένα με τα οποία εκπαιδεύτηκε, με αποτέλεσμα να μην έχει γενικεύσει το πρόβλημα, οδηγώντας στην ανικανότητα κατηγοριοποίησης νέων δεδομένων. Έχει παρατηρηθεί ότι όσο πιο μικρό είναι το σύνολο δεδομένων που έχει εκπαιδευτεί ένα τεχνικό νευρωνικό δίκτυο, τόσο πιο πιθανό είναι να εμφανισθεί το φαινόμενο του overfitting. Υπάρχουν διάφορες τεχνικές [37] για την αποφυγή αυτού του φαινομένου.

2.2.5 Υποεκπαίδευση

Η υποεκπαίδευση ή underfitting αποτελεί ένα σημαντικό πρόβλημα στον τομέα της επεξεργασίας εικόνας, ιδίως στο πλαίσιο των εργασιών σημασιολογικής τμηματοποίησης. Στη διαδικασία της εκπαίδευσης μοντέλων εικόνας, η υποεκπαίδευση εμφανίζεται όταν το μοντέλο δεν είναι σε θέση να εκμεταλλευτεί πλήρως την πληροφορία που περιέχεται στις εικόνες εκπαίδευσης, οδηγώντας σε ανεπαρκή γενίκευση και μειωμένη απόδοση στις εργασίες εντοπισμού αντικειμένων. Το φαινόμενο αυτό προκαλείται συχνά, από απλοποιημένα μοντέλα ή ανεπαρκή προεπεξεργασία των εικόνων, περιορίζοντας την ικανότητα του μοντέλου να αντιληφθεί πολύπλοκες δομές.

Η βελτίωση των τεχνικών προεπεξεργασίας μπορεί να συμβάλει στην αντιμετώπιση της υποεκπαίδευσης [27], ενισχύοντας την πολυπλοκότητα των εικόνων εκπαίδευσης και βελτιώνοντας την ικανότητα γενίκευσης των μοντέλων.

2.2.6 Οπισθοδρόμηση

Η οπισθοδρόμηση ή αλλιώς backpropagation, αποτελεί μια κεντρική διαδικασία στην εκπαίδευση των τεχνητών νευρωνικών δικτύων. Αυτή η μέθοδος επιτρέπει την υπολογιστική ανάδραση των κλίσεων της συνάρτησης κόστους ως προς τα βάρη του δικτύου, επιτρέποντας έτσι την αναβάθμισή τους με στόχο τη βελτιστοποίηση της απόδοσης του μοντέλου. Κατά τη διαδικασία της οπισθοδρόμησης, οι παράγωγοι της συνάρτησης κόστους υπολογίζονται με βάση τη διαφορική αλυσίδα, ξεκινώντας από την έξοδο του δικτύου προς τα πίσω, προς την είσοδο του δικτύου. Αυτή η πληροφορία σχετικά με το πόσο επηρεάζουν τα βάρη την επίδοση του δικτύου χρησιμοποιείται στη συνέχεια για να ενημερώσει τα βάρη μέσω της μεθόδου κατιούσας κατάβασης (gradient descent) ή κάποιου άλλου αλγορίθμου βελτιστοποίησης. Η οπισθοδρόμηση είναι ουσιαστικά η κύρια διαδικασία εκπαίδευσης των νευρωνικών δικτύων και είναι καθοριστική για την επίτευξη καλών επιδόσεων σε ποικίλες εφαρμογές μηχανικής μάθησης.

2.2.7 Βελτιστοποιητές

Οι βελτιστοποιητές ή optimizers, αποτελούν κρίσιμο στοιχείο κατά την εκπαίδευση των νευρωνικών δικτύων, καθώς επηρεάζουν τον τρόπο με τον οποίο τα βάρη του δικτύου ενημερώνονται κατά τη διάρκεια της εκπαίδευσης. Οι βελτιστοποιητές λειτουργούν με σκοπό την ελαχιστοποίηση της συνάρτησης κόστους, προσαρμόζοντας τα βάρη του δικτύου με βάση τη κλίση της συνάρτησης κόστους. Αυτός ο ρόλος των βελτιστοποιητών, είναι κρίσιμος για τη σύγκλιση του μοντέλου και τη βελτίωση της απόδοσής του. Μερικοί δημοφιλείς βελτιστοποιητές περιλαμβάνουν τον SGD (Stochastic Gradient Descent), τον Adam, τον RMSprop και τον Adagrad [68]. Κάθε ένας από αυτούς τους βελτιστοποιητές έχει μοναδικές ιδιότητες και παραμέτρους που επηρεάζουν τη σύγκλιση και τη σταθερότητα της εκπαίδευσης. Η επιλογή του κατάλληλου βελτιστοποιητή είναι κρίσιμη για την επίτευξη καλών αποτελεσμάτων κατά την εκπαίδευση ενός νευρωνικού δικτύου.

2.2.8 Συνάρτηση ενεργοποίησης

Η συνάρτηση ενεργοποίησης είναι ένα κρίσιμο στοιχείο στην αρχιτεκτονική των νευρωνικών δικτύων, καθώς καθορίζει τη μη γραμμική συμπεριφορά τους. Αυτή η συνάρτηση εφαρμόζεται στα αποτελέσματα ενός επιπέδου και δίνει ως έξοδο μια μη γραμμική αναπαράσταση των εισόδων. Η επιλογή της συνάρτησης ενεργοποίησης επηρεάζει την ικανότητα του μοντέλου να μάθει και να γενικεύει από τα δεδομένα. Κάποιες δημοφιλείς συναρτήσεις ενεργοποίησης περιλαμβάνουν τη σιγμοειδή, την υπερβολική εφαπτομένη γραμμική μονάδα (ReLU), την υπερβολική

εφαπτόμενη (tanh) και την ενεργοποίηση softmax. Κάθε μία από αυτές τις συναρτήσεις έχει μοναδικές ιδιότητες που μπορεί να την καθιστούν κατάλληλη για διαφορετικά είδη προβλημάτων και αρχιτεκτονικών δικτύων. Η επιλογή της κατάλληλης συνάρτησης ενεργοποίησης μπορεί να επηρεάσει σημαντικά την απόδοση και τη σύγκλιση του μοντέλου κατά τη διάρκεια της εκπαίδευσης.

2.2.9 Εξαφανιζόμενες/Εκτοξευμένες κλίσεις

Οι εξαφανιζόμενες και οι εκτοξευμένες κλίσεις είναι δύο προβλήματα που μπορούν να προκύψουν κατά την εκπαίδευση των νευρωνικών δικτύων και επηρεάζουν τη σύγκλιση του μοντέλου.

Οι εξαφανιζόμενες κλίσεις προκύπτουν όταν οι παράγωγοι της συνάρτησης κόστους γίνονται πολύ μικρές κατά την διαδικασία της οπισθοδρόμησης, ιδίως σε βαθιά νευρωνικά δίκτυα. Αυτό οφείλεται στην εκθετική μείωση της κλίσης κατά τη διάδοση προς τα πίσω, καθιστώντας την ενημέρωση των βαρών πολύ δύσκολη ή αδύνατη. Αυτό μπορεί να οδηγήσει σε πολύ αργή ή ακόμη και τον πρώιμο τερματισμό εκπαίδευσης του μοντέλου.

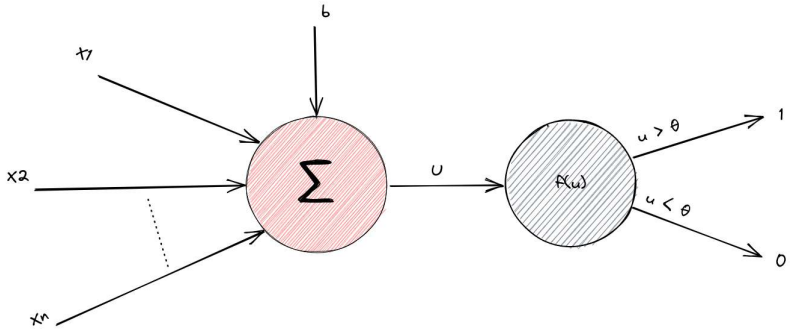
Αντίθετα, οι εκτοξευμένες Κλίσεις συμβαίνουν όταν οι παράγωγοι γίνονται πολύ μεγάλες κατά την διαδικασία της οπισθοδρόμησης αντίστοιχα, προκαλώντας αστάθεια κατά την εκπαίδευση. Αυτό συχνά συμβαίνει σε δίκτυα μεγάλου βάθους ή σε περιπτώσεις όπου οι παράμετροι του μοντέλου έχουν πολύ μεγάλες τιμές. Η ύπαρξη των εκτοξευμένων κλίσεων, μπορεί να οδηγήσει σε ασταθείς ενημερώσεις των βαρών, προκαλώντας το μοντέλο να μην συγκλίνει ή να χάσει τη συνοχή της εκπαίδευσής του.

Και τα δύο προβλήματα απαιτούν προσεκτικές τεχνικές για την αντιμετώπισή τους, όπως η χρήση κατάλληλων συναρτήσεων ενεργοποίησης, και η κανονικοποίηση των βαρών. Η κατανόηση αυτών των προβλημάτων και η ανάληψη κατάλληλων μέτρων για την αντιμετώπισή τους είναι κρίσιμες για την επιτυχή εκπαίδευση των νευρωνικών δικτύων.

2.2.10 Mcculloch-Pitts

Ένας βιολογικός νευρώνας αποτελείται από τρία μέρη, το κυρίως σώμα, τους δενδρίτες και τον άξονα. Οι δενδρίτες, αποτελούν την είσοδο ενός ηλεκτρικού σήματος στον νευρώνα. Το κυρίως σώμα επεξεργάζεται αυτή τη πληροφορία και αποφασίζει εάν είναι αρκετά σημαντική έτσι ώστε να τη προωθήσει στους γειτονικούς του νευρώνες. Τέλος, ο άξονας, είναι υπεύθυνος για τη μεταφορά αυτού του σήματος στους δενδρίτες του γειτονικού νευρώνα όποτε είναι απαραίτητο. Αυτό ακριβώς, προσομοιώνει και το μοντέλο Mcculloch-Pitts, ο πρώτος τεχνητός νευρώνας ο οποίος προτάθηκε από τους Warren S. Mcculloch και Walter Pitts [42]. Αναλυτικότερα, το μοντέλο δέχεται δυαδικές εισόδους Xn τις οποίες αθροίζει παράγοντας ένα αποτέλεσμα u, από το οποίο αφαιρείται η μεταβλητή b, γνωστή ως πόλωση. Έπειτα, το τελικό αποτέλεσμα αποτελεί την είσοδο στην συνάρτηση ενεργοποίησης. Η συνάρτηση ενεργοποίησης, ελέγχει εάν

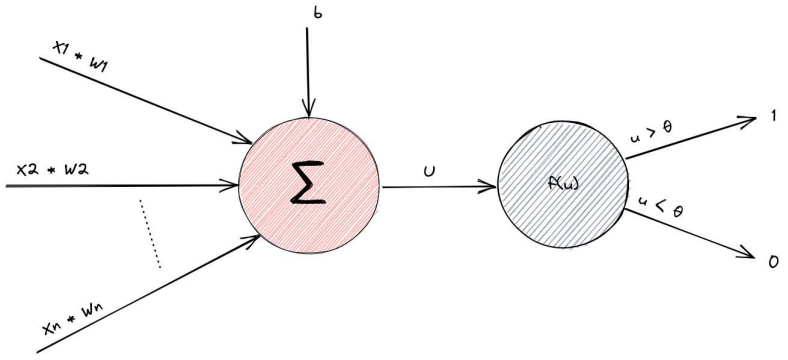
η διέγερση του νευρώνα είναι μεγαλύτερη από ένα κατώφλι θ. Εάν ισχύει, τότε ο νευρώνας ενεργοποιείται, αλλιώς παραμένει ανενεργός. Για λόγους απλούστευσης, η συνάρτηση ενεργοποίησης παράγει σαν έξοδο το αποτέλεσμα 1 εάν ο νευρώνας ενεργοποιήθηκε ή 0 στην αντίθετη περίπτωση.



Σχήμα 2.1: Νευρώνας Mcculloch-Pitts

2.2.11 Perceptron

Το μοντέλο Perceptron, προτάθηκε από τον Frank Rosenblatt to 1958 [56]. Η έρευνα του, ήταν ιδιαίτερα πρωτοπόρα, καθώς διατύπωσε μία διαφορετική θεωρία σχετικά με τον τρόπο που ο ανθρώπινος εγκέφαλος αποθηκεύει πληροφορίες. Οι επιστήμονες εκείνη την εποχή πίστευαν πως υπήρχε συσχέτιση ένα προς ένα ανάμεσα στο νευρικό σήμα, και την αποθηκευμένη πληροφορία στον εγκέφαλο. Ωστόσο, Ο Rosenblatt πίστευε πως η αποθήκευση της πληροφορίας πραγματοποιείται δημιουργώντας νέες συσχετίσεις μεταξύ των νευρώνων. Έτσι, δημιουργήθηκε το μοντέλο Perceptron, ένα πανομοιότυπο μοντέλο με το Mcculloch-Pitts, αλλά με μερικές διαφορές. Η βασική διαφορά των δύο μοντέλων είναι ότι οι είσοδοι Xn πολλαπλασιάζονται με κάποια βάρη Wn, προσομοιώνοντας την θεωρία του. Τέλος, οι είσοδοι του μοντέλου δεν είναι απαραίτητο να περιέχουν δυαδικές τιμές.



Σχήμα 2.2: Νευρώνας Perceptron

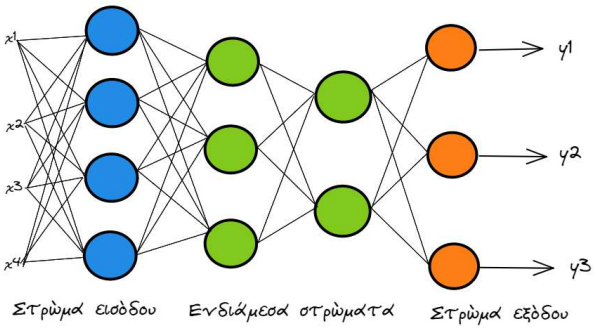
$$u = X_1W_1 \ + ... + XnWn + b$$

$$y = f(u) = \begin{cases} 1, & \text{αν } u \geq \theta \\ 0, & \text{αν } u < \theta \end{cases}$$

Το μοντέλο Perceptron, αν και πρωτοπόρο, είχε κάποιες αδυναμίες, η βασικότερη είναι πως δεν μπορεί να λύσει μη γραμμικά διαχωρίσιμα προβλήματα, όπως το πρόβλημα της λογικής πύλης XOR.

2.2.12 MLP

Τα MLP ή αλλιώς Multi Layer Perceptron επεκτείνουν τη λογική του perceptron, δίνοντας τη δυνατότητα στο νευρωνικό δίκτυο να επιλύει μη γραμμικά διαχωρίσιμα προβλήματα. Σε ένα MLP, υπάρχουν 3 ειδών κατηγορίες στρωμάτων. Το στρώμα εισόδου, χρησιμοποιείται για την είσοδο των δεδομένων στο νευρωνικό δίκτυο. Το ενδιάμεσο ή κρυφό στρώμα, αποτελεί την υπολογιστική ικανότητα του νευρωνικού δικτύου. Ως αποτέλεσμα αυτού, αυξάνοντας τον αριθμό των ενδιάμεσων στρωμάτων, ένα MLP έχει την δυνατότητα να λύσει περιπλοκότερα προβλήματα [50]. Τέλος, το τελευταίο στρώμα είναι το στρώμα εξόδου, λαμβάνει την έξοδο του τελευταίου ενδιάμεσου στρώματος και παράγει το τελικό αποτέλεσμα του μοντέλου. Η αρχιτεκτονική του, χαρακτηρίζεται ως Feedforward, σημαίνοντας, ότι δεν σχηματίζονται κύκλοι μεταξύ των νευρώνων και ότι το αποτέλεσμα των νευρώνων κινείται μόνο προς μια κατεύθυνση – μπροστά. Επιπρόσθετα, χαρακτηρίζεται ως πλήρως συνδεδεμένο δίκτυο επειδή όλα τα αποτελέσματα των νευρώνων αποτελούν είσοδο σε όλους τους νευρώνες του επόμενου στρώματος.



Σχήμα 2.3: Multi Layer Perceptron

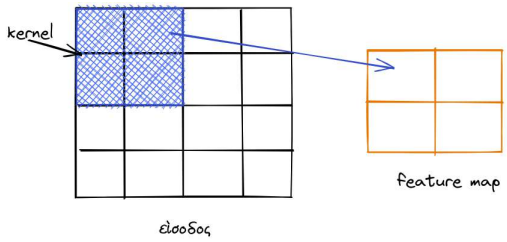
2.3 Deep Learning

Η βαθιά μάθηση, είναι σχετικά ένα νέο και υπό έρευνα υποπεδίο της μηχανικής μάθησης. Ένα από τα βασικά πλεονεκτήματα του έναντι της τυπικής μηχανικής μάθησης είναι η εκπληκτική ικανότητα εκμάθησης χωρίς κάποια προσεκτική προεπεξεργασία των δεδομένων. Η αρχιτεκτονική ενός μοντέλου βαθιάς μάθησης είναι παρόμοια με ενός MLP, με τη μόνη διαφορά να είναι ότι ο αριθμός των κρυφών στρωμάτων είναι αρκετά μεγαλύτερος. Τυπικά, όσο μεγαλύτερος είναι ο αριθμός των κρυφών στρωμάτων, τόσο πιο περίπλοκα προβλήματα μπορεί να λύσει ένα μοντέλο βαθιάς μάθησης. Για παράδειγμα, αναλύοντας μια εικόνα, τα αρχικά στρώματα θα

διακρίνουν την ύπαρξη γωνιών, τα μεσαία στρώματα μπορούν να εντοπίσουν διάφορα μοτίβα όπως κύκλους, κ.ο.κ. Γενικά, εμβαθύνοντας στα κρυφά στρώματα, παρατηρούμε ότι μπορούν να εντοπίσουν πιο αφηρημένα και περίπλοκα μοτίβα [27].

2.3.1 Convolutional Neural Networks

Τα συνελικτικά νευρωνικά δίκτυα ή αλλιώς CNN's, είναι μία συγκεκριμένη εφαρμογή βαθιάς μάθησης η οποία ειδικεύεται στον εντοπισμό μοτίβων σε εικόνες [29]. Είναι αρκετά αποδοτικά και ταυτόχρονα έχουν απλή αρχιτεκτονική. Τα CNN αποτελούνται από τριών ειδών στρώματα, τα πλήρως συνδεδεμένα, τα συνελικτικά και τα στρώματα υποδειγματοληψίας ή pooling layers. [46].



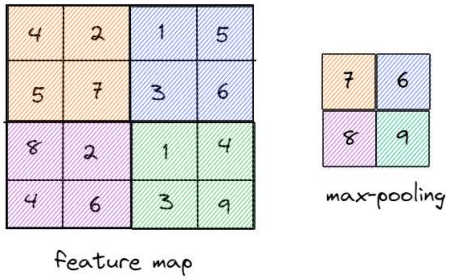
Σχήμα 2.4: Συνλέλιξη με βήμα 2

2.3.1.1 Συνελικτικό Στρώμα Το συνελικτικό στρώμα έχει σημαντικό ρόλο στην αρχιτεκτονική CNN. Η είσοδος του, μπορεί να απεικονιστεί ως ένας πίνακας τριών διαστάσεων. Εκ των οποίων είναι το πλάτος, (W) το ύψος, (H) και τέλος τα χρωματικά κανάλια της εικόνας (WxHxC). Αποτέλεσμα αυτού, είναι η ύπαρξη υπερβολικού αριθμού βαρών στο νευρωνικό δίκτυο. Παραδείγματος χάριν, για να συνδεθεί μια εικόνα με διαστάσεις 50χ50 με τρία χρωματικά κανάλια (RGB) από το στρώμα εισόδου με έναν νευρώνα, θα χειαστεί 7500 (50χ50χ3) βάρη [3]. Το συνελικτικό στρώμα, για την αποφυγή αυτού του προβλήματος χρησιμοποιεί έναν δισδιάστατο πίνακα γνωστό ως πυρήνα (kernel). Αναλυτικότερα, ο πυρήνας διασχίζει την εικόνα, πολλαπλασιάζοντας την είσοδο, με σκοπό την εξαγωγή χαρακτηριστικών, αποθηκεύοντας το αποτέλεσμα σε έναν δισδιάστατο πίνακα γνωστό ως χάρτη χαρακτηριστικών ή feature map [46]. Μία βασική υπερπαράμετρος είναι το βήμα ή αλλιώς stride το οποίο ορίζει τον αριθμό των εικονοστοιχείων που μεταπηδά το φίλτρο καθώς διασχίζει την είσοδο.

2.3.1.2 Στρώμα Υποδειγματοληψίας Το στρώμα υποδειγματοληψίας ή αλλιώς pooling layer, χρησιμοποιείται συνήθως μετά από κάθε στρώμα συνέλιξης και δρα πάνω στον χάρτη χαρακτηριστικών. Όπως και στο συνελικτικό στρώμα, χρησιμοποιεί έναν πυρήνα συνήθως με διαστάσεις 2χ2 και stride 2, μειώνοντας περαιτέρω την πολυπλοκότητα του νευρωνικού. Υπάρχουν δύο ειδών στρωμάτων υποδειγματοληψίας, το max-pooling το οποίο κρατάει την μέγιστη τιμή από τον χάρτη χαρακτηριστικών και το average-pooling το οποίο κρατάει την μέση τιμή. Συνή-

θως χρησιμοποιείται το max-pooling καθώς έτσι εξάγονται τα εντονότερα χαρακτηριστικά από τον χάρτη χαρακτηριστικών.

2.3.1.3 Πλήρως Συνδεδεμένο Στρώμα Το πλήρως συνδεδεμένο στρώμα, είναι πανομοιότυπο με τα στρώματα που βρίσκονται στα παραδοσιακά τεχνητά νευρωνικά δίκτυα. Συνεπώς, κάθε νευρώνας συνδέεται με το επόμενο στρώμα, δημιουργώντας έναν αρκετά υψηλό αριθμών παραμέτρων. Για αυτό το λόγο, το σημαντικότερο μειονέκτημα ενός πλήρους συνδεδεμένου στρώματος είναι η υπολογιστική ισχύ που απαιτείται για την εκπαίδευση τους [3]



Σχήμα 2.5: Στρώμα Υποδειγματοληψίας max-pooling 2χ2 με βήμα 2

2.3.2 Batch Normalization

Το Batch normalization (BN) είναι μια τεχνική που χρησιμοποιείται στα συνελικτικά νευρωνικά δίκτυα για την επιτάχυνση της εκπαίδευσης και τη βελτίωση της απόδοσης. Η ιδέα πίσω από το Batch Normalization είναι η κανονικοποίηση των εισόδων κάθε στρώματος, με την προσθήκη μιας παραμέτρου κανονικοποίησης, που συνήθως εφαρμόζεται μετά το στρώμα συνέλιξης ή το πλήρως συνδεδεμένο επίπεδο, και πριν την εφαρμογή της συνάρτησης ενεργοποίησης.

Το Batch normalization, λειτουργεί με τον υπολογισμό του μέσου όρου και της τυπικής απόκλισης των εισόδων του επιπέδου εντός κάθε batch, και στη συνέχεια κανονικοποιεί τα δεδομένα με βάση αυτές τις παραμέτρους [14]. Αυτό βοηθάει στην αποφυγή του προβλήματος της αστάθειας της διανυσματικής αναπαράστασης που μπορεί να προκύψει κατά την εκπαίδευση (exploding and vanishing descents), επιτρέποντας στο δίκτυο να εκπαιδευτεί πιο γρήγορα και να επιτύχει καλύτερη γενίκευση [36].

Ένα πλεονέκτημα του Batch Normalization είναι ότι μπορεί να επιτρέψει τη χρήση μεγαλύτερων ρυθμών μάθησης, καθώς μειώνει την πιθανότητα αστάθειας κατά την εκπαίδευση. Επιπλέον, μπορεί να λειτουργήσει ως ένα είδος κανονικοποίησης κατά την εκπαίδευση, βοηθώντας στην αποφυγή της υπερεκπαίδευσης.

Συνολικά, το Batch Normalization έχει αποδειχθεί ως ένα αποτελεσματικό εργαλείο για την επιτάχυνση της εκπαίδευσης και τη βελτίωση της απόδοσης των συνελικτικών νευρωνικών δικτύων.

2.3.3 Στρώμα Dropout

Το στρώμα Dropout είναι ένα από τα βασικά εργαλεία στον τομέα της μηχανικής μάθησης με επίβλεψης στα βαθιά νευρωνικά δίκτυα. Η λειτουργία ενός στρώματος Dropout, είναι να απενεργοποιεί τυχαία νευρώνες κατά την εκπαίδευση με πιθανότητα $\mathcal{P}$, με στόχο την αποτροπή της υπερεκπαίδευσης (overfitting). Παρόλα αυτά, κατά τη διαδικασία της δοκιμής του νευρωνικού σε νέα δεδομένα, όλοι οι νευρώνες είναι ενεργοί. Επιπρόσθετα, η τυχαιότητα στην απενεργοποίηση των νευρώνων κατά την εκπαίδευση λειτουργεί και ως ένα είδος κανονικοποίησης (regularization), επιτρέποντας στο μοντέλο να γενικεύει καλύτερα σε νέα δεδομένα. Το στρώμα Dropout μέσω της τυχαίας απενεργοποίησης ενός ποσοστού νευρώνων κατά την εκπαίδευση, μειώνει την εξάρτηση μεταξύ των νευρώνων και εξασφαλίζει ότι κάθε νευρώνας εκπαιδεύεται ανεξάρτητα. Με αυτόν τον τρόπο, το μοντέλο μπορεί να αναπτύξει κατανόηση σε πιο γενικευμένες αναπαραστάσεις και να αποφύγει την υπερεκπαίδευση. Ως αποτέλεσμα, το Dropout έχει γίνει ένα απαραίτητο εργαλείο για την ανάπτυξη αξιόπιστων μοντέλων νευρωνικών δικτύων στη σύγχρονη έρευνα και εφαρμογές στον τομέα της μηχανικής μάθησης [64].

2.3.4 Recurrent Neural Networks

Τα Recurrent Neural Networks (RNNs) χρησιμοποιούνται ευρέως για την επεξεργασία διαδοχικών δεδομένων, όπως η ομιλία, και τα βίντεο, όπου τα δεδομένα έχουν άμεση συσχέτιση αναμεταξύ τους. Μια σημαντική διαφορά στην αρχιτεκτονική τους με τα feed forward νευρωνικά δίκτυα, είναι ο τρόπος με τον οποίο η πληροφορία περνιέται μέσα στο νευρωνικό. Τα RNNs, σχηματίζουν κύκλους στα ενδιάμεσα στρώματα τους, επιτρέποντας να περαστεί η πληροφορία της προηγούμενης εισόδου , και να συνυπολογιστεί στην έξοδο του [28] [59]. Τα RNNs όπως τα feed forward νευρωνικά, έχουν πρόβλημα με vanishing ή exploding gradients. Τα οποία λύνουν τα νευρωνικά Long Short-Term Memory.

2.3.5 Νευρωνικά Long Short-Term Memory

Τα νευρωνικά δίκτυα Long Short-Term Memory (LSTM) είναι μία βελτιωμένη αρχιτεκτονική των Recurrent neural networks. Αναλυτικότερα, αποτελούνται από ένα δίκτυο συνδεδεμένων μνημονικών κυττάρων, το καθένα εξοπλισμένο με μηχανισμούς πύλης. Υπάρχουν τριών ειδών πύλες, η πύλη εισόδου, η πύλη λήθης, και η πύλη εξόδου, οι οποίες αποτελούν τον κρυφό μηχανισμό που επιτρέπει στο LSTM να διαχειριστεί πληροφορίες και να εκτελέσει τις απαραίτητες ενέργειες [61].

1. **Πύλη εισόδου**: Καθορίζει ποιες νέες πληροφορίες θα ενσωματωθούν στην εσωτερική κατάσταση του LSTM. Ενεργοποιείται από τα δεδομένα εισόδου και την προηγούμενη κρυφή κατάσταση
2. **Πύλη λήθης**: Ελέγχει ποια από τις προηγούμενες πληροφορίες πρέπει να διατηρηθούν ή να αφαιρεθούν από την εσωτερική κατάσταση. Βασίζεται στην προηγούμενη κρυφή

κατάσταση και την τρέχουσα είσοδο. Η πύλη αυτή, είναι ουσιαστική για την απόρριψη ασυσχέτιστης πληροφορίας.

3. **Πύλη εξόδου**: Καθορίζει την έξοδο του κυττάρου LSTM.

Οι πύλες αυτές επιτρέπουν στο LSTM να μαθαίνει τις μακροπρόθεσμες συσχετίσεις και να αντιμετωπίσει πολύπλοκες σειρές δεδομένων [65].

2.3.6 Μεταφορά Μάθησης

Η μεταφορά μάθησης επιτρέπει τη χρήση προ-εκπαιδευμένων δικτύων, τα οποία έχουν εκπαιδευτεί σε μεγάλα σύνολα δεδομένων, για την έναρξη της εκπαίδευσης σε νέες, συγκεκριμένες εργασίες. Αυτό μειώνει την ανάγκη για μεγάλα σύνολα δεδομένων με ετικέτες και επιταχύνει τη διαδικασία εκπαίδευσης. Για παράδειγμα, ένα δίκτυο που έχει εκπαιδευτεί στο σύνολο δεδομένων ImageNet μπορεί να χρησιμοποιηθεί ως βάση για την εκπαίδευση σε ένα πιο εξειδικευμένο σύνολο δεδομένων για τη σημασιολογική τμηματοποίηση [25] [21]. Η λεπτομερής προσαρμογή (fine-tuning) των υψηλότερων επιπέδων του δικτύου επιτρέπει την εξειδίκευση στις νέες κατηγορίες δεδομένων, ενώ οι χαμηλότερες στρώσεις παραμένουν ανεκπαίδευτες καθώς περιέχουν γενικότερα χαρακτηριστικά που είναι χρήσιμα σε πολλές εργασίες.

2.3.7 GANs

Τα γεννητικά αντιπαραθετικά δίκτυα ή GANs αποτελούν μια καινοτόμο προσέγγιση στη βαθιά μάθηση, όπου δύο δίκτυα ανταγωνίζονται μεταξύ τους: το δίκτυο γεννήτριας (generator) και το δίκτυο διακρίβωσης (discriminator). Το δίκτυο γεννήτριας, δημιουργεί συνθετικά δείγματα που προσπαθούν να μοιάζουν με τα πραγματικά δεδομένα, ενώ το δίκτυο διακρίβωσης, προσπαθεί να διακρίνει τα πραγματικά από τα συνθετικά δείγματα. Αυτός ο ανταγωνισμός οδηγεί τη γεννήτρια στη δημιουργία όλο και πιο ρεαλιστικών δειγμάτων. Στη σημασιολογική τμηματοποίηση, τα GANs μπορούν να χρησιμοποιηθούν για τη δημιουργία συνθετικών εικόνων που ενισχύουν τα σύνολα δεδομένων εκπαίδευσης, βελτιώνοντας την ακρίβεια και τη γενίκευση των μοντέλων [44].

2.3.8 Autoencoders

Οι Autoencoders είναι μια κατηγορία νευρωνικών η οποία χρησιμοποιεί τον αλγόριθμο μάθησης χωρίς επίβλεψη και έχουν σχεδιαστεί για να μαθαίνουν μια συμπιεσμένη αναπαράσταση (κωδικοποίηση) των εισερχόμενων δεδομένων. Τα νευρωνικά αυτά, χρησιμοποιούνται συνήθως για σκοπούς μείωσης διαστάσεων, εκμάθησης χαρακτηριστικών ή εξαλείφοντας τον θόρυβο από τα δεδομένα [11]. Οι autoencoders έχουν διάφορες εφαρμογές, μερικές από αυτές είναι:

1. **Μείωση διαστάσεων**: Στοχεύει στη συμπίεση του όγκου των δεδομένων προβάλλοντας τα σε έναν χώρο χαμηλότερης διάστασης (λανθάνων χώρος). Σε αντίθεση με τις γραμμικές μεθόδους όπως τη PCA, οι Autoencoders μπορούν να μάθουν μη γραμμικές μειώσεις διαστάσεων, διατηρώντας τη σημαντική πληροφορία και τη δομή των δεδομένων για σκοπούς οπτικοποίησης ή αποδοτικότερης αποθήκευσης.

2. **Εξαγωγή Χαρακτηριστικών:**: Εστιάζει στη χρησιμότητα της μαθημένης αναπαράστασης του κωδικοποιητή ως είσοδο για μεταγενέστερες εργασίες. Το δίκτυο μαθαίνει αυτόματα να αναγνωρίζει σύνθετα, αφηρημένα μοτίβα και χαρακτηριστικά από τα ακατέργαστα δεδομένα, τα οποία μπορούν να βελτιώσουν την απόδοση των ταξινομητών ή άλλων μοντέλων πρόβλεψης, λειτουργώντας ως ένας εκπαιδευόμενος προ-επεξεργαστής.

3. **Αφαίρεση Θορύβου**: Αφορά την ικανότητα του δικτύου να ανακατασκευάζει την αρχική, καθαρή είσοδο από δεδομένα που έχουν υποστεί αλλοίωση ή περιέχουν θόρυβο. Οι Denoising Autoencoders εκπαιδεύονται να αγνοούν τον στοχαστικό θόρυβο (Gaussian blur) και να συλλαμβάνουν μόνο τα δομικά χαρακτηριστικά των δεδομένων, λειτουργώντας ως φίλτρα αποκατάστασης σήματος ή εικόνας.

2.3.8.1 Κωδικοποιητής Ο κωδικοποιητής αποτελείται από ένα ή περισσότερα κρυφά στρώματα, τα οποία συμπιέζουν τα δεδομένα σε έναν χαμηλότερων διαστάσεων λανθάνοντα χώρο, καταγράφοντας τα πιο σημαντικά χαρακτηριστικά των δεδομένων. Αυτή η λανθάνουσα αναπαράσταση αναφέρεται συχνά ως ”bottleneck”, επειδή αναγκάζει το νευρωνικό δίκτυο να μάθει μια συμπαγή και ουσιαστική περίληψη των εισερχόμενων δεδομένων.

2.3.8.2 Αποκωδικοποιητής Ο αποκωδικοποιητής λαμβάνει αυτή τη συμπιεσμένη αναπαράσταση και ανακατασκευάζει τα δεδομένα πίσω στην αρχική τους μορφή. Ο στόχος είναι η έξοδος να είναι όσο το δυνατόν πιο κοντά στην αρχική είσοδο, διασφαλίζοντας έτσι ότι ο κωδικοποιητής έχει καταγράψει τα ουσιώδη χαρακτηριστικά των δεδομένων.

2.3.9 Αρχιτεκτοκική VGG

Η αρχιτεκτονική VGG (Visual Geometry Group) αποτελεί μία από τις θεμελιώδεις συνεισφορές στον χώρο των βαθιών νευρωνικών δικτύων για την επεξεργασία εικόνων. Η αρχιτεκτονική αυτή, η οποία αναπτύχθηκε από την Simonyan και Zisserman του πανεπιστημίου της Οξφόρδης [63], εισήγαγε την καινοτόμο προσέγγιση της χρήσης πολύ μικρών συνελικτικών φίλτρων διάστασης 3×3 συνδυασμένων με σημαντική αύξηση του βάθους του δικτύου.

Η VGG16, η οποία αποτελεί τη κεντρική δομή της αρχιτεκτονικής SegNet που εφαρμόζεται στην παρούσα ερευνητική εργασία, χαρακτηρίζεται από την εξαιρετική της ομοιομορφία και απλότητα στον σχεδιασμό. Η αρχιτεκτονική περιλαμβάνει συνολικά δεκαέξι στρώματα με εκπαι-

δεύσιμες παραμέτρους, τα οποία οργανώνονται σε πέντε διακριτές ομάδες συνελικτικών στρωμάτων. Κάθε ομάδα ακολουθείται από ένα στρώμα υποδειγματοληψίας max-pooling διάστασης 2×2 με βήμα ίσο με 2, γεγονός που επιτρέπει την προοδευτική μείωση της χωρικής διάστασης των χαρτών χαρακτηριστικών διατηρώντας παράλληλα την πληροφορία των χαρακτηριστικών.

Η θεωρητική βάση της VGG16 στηρίζεται στην υπόθεση ότι η χρήση μικρών φίλτρων 3×3 σε διαδοχικά στρώματα μπορεί να επιτύχει το ίδιο οπτικό πεδίο με μεγαλύτερα φίλτρα, ενώ παράλληλα προσφέρει σημαντικά πλεονεκτήματα. Συγκεκριμένα, δύο διαδοχικά 3×3 συνελικτικά στρώματα έχουν δραστικό οπτικό πεδίο 5×5, ενώ τρία τέτοια στρώματα αντιστοιχούν σε δραστικό οπτικό πεδίο 7×7. Αυτή η προσέγγιση επιτρέπει τη χρήση περισσότερων μη-γραμμικών συναρτήσεων ενεργοποίησης, αυξάνοντας την εκφραστικότητα του μοντέλου χωρίς αντίστοιχη αύξηση των παραμέτρων.

Στο πλαίσιο του SegNet, η αρχιτεκτονική VGG16 λειτουργεί ως κωδικοποιητής, εκτελώντας ιεραρχική εξαγωγή χαρακτηριστικών από τις εισερχόμενες εικόνες. Η αρχιτεκτονική ξεκινά με 64 χάρτες χαρακτηριστικών στο πρώτο στρώμα και διπλασιάζει αυτόν τον αριθμό σε κάθε στρώμα υποδειγματοληψίας max-pooling, φτάνοντας έως τους 512 χάρτες χαρακτηριστικών στα βαθύτερα στρώματα. Αυτή η προοδευτική αύξηση των καναλιών επιτρέπει στο δίκτυο να μάθει αρχικά απλά χαρακτηριστικά όπως γραμμές και άκρες, και στη συνέχεια να συνδυάζει αυτά τα χαρακτηριστικά για τη δημιουργία πιο σύνθετων αναπαραστάσεων.

2.3.10 ResNet Αρχιτεκτονική

Η αρχιτεκτονική ResNet αντιπροσωπεύει μία επαναστατική καινοτομία στον τομέα των βαθιών νευρωνικών δικτύων, καθώς εισήγαγε την έννοια των υπολειμματικών συνδέσεων (residual connections) ή συνδέσεων παράκαμψης (skip connections) [33]. Η συγκεκριμένη προσέγγιση επέλυσε ένα από τα πιο σημαντικά προβλήματα στην εκπαίδευση βαθιών δικτύων, το πρόβλημα των εξαφανιζόμενων κλίσεων, επιτρέποντας την επιτυχή εκπαίδευση αρχιτεκτονικών με πολύ βαθιά δομή.

Η αρχιτεκτονική ResNet, η οποία αποτελεί τον κεντρικό άξονα των μοντέλων U-Net και DeepLabv3+ στην παρούσα μελέτη, περιλαμβάνει τριάντα τέσσερα στρώματα οργανωμένα σε υπολειμματικά μπλοκ (residual blocks). Το θεμελιώδες στοιχείο αυτής της αρχιτεκτονικής είναι το υπολειμματικό μπλοκ (residual block), το οποίο υλοποιεί τη μαθηματική σχέση:

$$\mathbf{y} = \mathcal{F}(\mathbf{x}, \{W_i\}) + \mathbf{x} \tag{1}$$

όπου $\mathbf{x}$ και $\mathbf{y}$ αντιπροσωπεύουν τα διανύσματα εισόδου (input vectors) και εξόδου (output vectors) αντίστοιχα, και $\mathcal{F}(\mathbf{x}, \{W_i\})$ συμβολίζει την υπολειμματική απεικόνιση που το δίκτυο πρέπει να μάθει.

Η καινοτομία των υπολειμματικών συνδέσεων έγκειται στο γεγονός ότι αντί να εκπαιδεύεται το δίκτυο να μάθει απευθείας την επιθυμητή υποκείμενη απεικόνιση $H(\mathbf{x})$, εκπαιδεύεται να μάθει την υπολειμματική συνάρτηση ($\mathcal{F}(\mathbf{x}) = H(\mathbf{x}) - \mathbf{x}$. Αυτή η προσέγγιση βασίζεται στην υπόθεση ότι είναι ευκολότερο να βελτιστοποιηθεί η υπολειμματική απεικόνιση παρά η αρχική απεικόνιση, ιδιαίτερα όταν η βέλτιστη συνάρτηση είναι κοντά στην ταυτότητα.

Η δομή του ResNet αποτελείται από ένα αρχικό συνελικτικό στρώμα διάστασης 7×7 με 64 φίλτρα, ακολουθούμενο από τέσσερις ομάδες υπολειμματικών μπλοκ (residual blocks). Η πρώτη ομάδα περιλαμβάνει τρία μπλοκ με 64 φίλτρα, η δεύτερη τέσσερα μπλοκ με 128 κανάλια , η τρίτη έξι μπλοκ με 256 κανάλια, και η τέταρτη τρία μπλοκ με 512 κανάλια.

Στην υλοποίηση U-Net με ResNet34 ως κεντρικό άξονα (backbone) που χρησιμοποιείται στην έρευνα, η ResNet λειτουργεί ως κωδικοποιητής, παρέχοντας βαθμιαία εξαγωγή χαρακτηριστικών σε πολλαπλές κλίμακες. Τα χαρακτηριστικά που εξάγονται από τα διαφορετικά στάδια του μοντέλου, συνδέονται με τα αντίστοιχα στρώματα του αποκωδικοποιητή μέσω των υπολειμματικών συνδέσεων του U-Net, επιτρέποντας την αποτελεσματική ανάκτηση λεπτομερών πληροφοριών που χάνονται κατά τη διαδικασία της υποδειγματοληψίας.

Όσον αφορά τη χρήση της ResNet στο DeepLabv3+, η αρχιτεκτονική τροποποιείται για να υποστηρίξει διάτρητες συνελίξεις (atrous convolutions). Αυτή η τροποποίηση επιτρέπει τη διατήρηση υψηλότερης χωρικής ανάλυσης στα βαθύτερα στρώματα του δικτύου, γεγονός κρίσιμο για την αποτελεσματική σημασιολογική κατάτμηση.

Κεφάλαιο 3ο: Κλασικές Τεχνικές Κατάτμησης Εικόνων

Πριν την επικράτηση των μεθόδων βαθιάς μάθησης, η κατάτμηση εικόνων βασιζόταν σε τεχνικές που αξιοποιούσαν κυρίως στατιστικές και γεωμετρικές ιδιότητες των εικονοστοιχείων. Αυτές οι ”παραδοσιακές” μέθοδοι βασίζονταν στην ανάλυση του χρώματος, της υφής, των ακμών και της φωτεινότητας, με στόχο τη διάκριση των περιοχών ενδιαφέροντος στην εικόνα. Αν και παρουσιάζουν πλεονεκτήματα ως προς την απλότητα και τον υπολογιστικό χρόνο, στερούνται γενικευσιμότητας και είναι ευάλωτες σε θόρυβο και μεταβολές του φωτισμού.

Σε αυτό το κεφάλαιο θα αναλυθούν μερικές από τις προαναφερθέντες τεχνικές κατάτμησης εικόνας

3.1 Κατωφλίωση (Thresholding)

Η κατωφλίωση αποτελεί μία από τις απλούστερες μεθόδους κατάτμησης. Βασίζεται στην παραδοχή ότι οι περιοχές ενδιαφέροντος διαφέρουν σημαντικά ως προς τη φωτεινότητα από το φόντο. Ένα εικονοστοιχείο ταξινομείται ως αντικείμενο ή φόντο αναλόγως του αν η τιμή του ξεπερνά ή όχι ένα προκαθορισμένο κατώφλι τ.

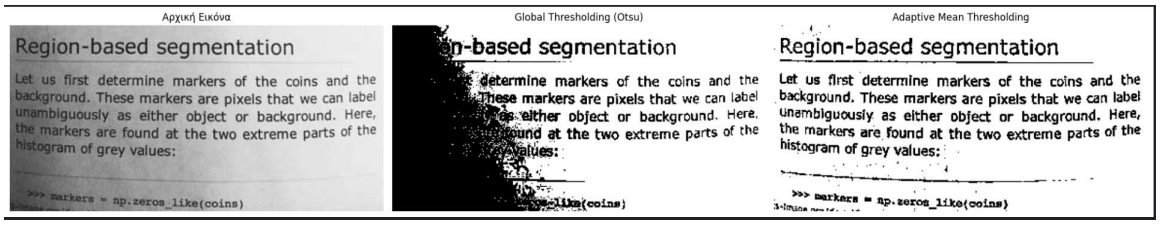
3.1.1 Global Thresholding

Η μέθοδος της παγκόσμιας κατωφλίωσης (global thresholding) εφαρμόζει ένα ενιαίο κατώφλι έντασης T σε ολόκληρη την εικόνα. Καθορίζει ότι ένα εικονοστοιχείο I(x,y) ανήκει στο προσκήνιο όταν I(x,y) > T, διαφορετικά ταξινομείται στο παρασκήνιο. Το εν λόγω κατώφλι μπορεί να επιλεγεί χειροκίνητα ή μέσω αυτοματοποιημένων μεθόδων, όπως η μέθοδος του Otsu [10] [26], που επιδιώκει να ελαχιστοποιήσει τη διακύμανση εντός των δύο τάξεων (Προσκηνίου-παρασκηνίου) και να μεγιστοποιήσει τη μεταξύ τους διασπορά.

Η παγκόσμια κατωφλίωση λειτουργεί ικανοποιητικά όταν η εικόνα διαθέτει ομοιογενή φωτισμό και σαφή διαχωρισμό των εντάσεων, αλλά αποτυγχάνει σε συνθήκες με μεταβλητό φωτισμό ή θόρυβο, καθώς το ίδιο κατώφλι εφαρμόζεται σε περιοχές με διαφορετικά χαρακτηριστικά [1]. Σε αυτές τις περιπτώσεις, προτιμούνται οι τοπικές προσαρμοστικές τεχνικές (Adaptive thresholding) όπως των Niblack, Sauvola, Bernsen κ.α, στις οποίες το όριο προσαρμόζεται τοπικά σε κάθε υποπεριοχή [72] [72].

3.1.2 Adaptive Thresholding

Η προσαρμοστική ή adaptive κατωφλίωση υπολογίζει την τιμή του κατωφλιού τοπικά, βασιζόμενη στις ιδιότητες του γειτονικού περιβάλλοντος κάθε εικονοστοιχείου. Συνήθως, υπολογίζει



Σχήμα 3.1: Αποτελέσματα Πειράματος

είτε μια αραιή μέση τιμή είτε μια σταθμισμένη (π.χ. Gaussian) μέσα σε ένα τοπικό παράθυρο, και προαιρετικά αφαιρεί μια σταθερά C εν ονόματι σταθερά προσαρμογής.

Η σταθερά C είναι ένας ρυθμιστικός όρος που ελέγχει το τελικό threshold. Αυξάνοντας τη μεταβλητή αυτή,το threshold γίνεται πιο αυστηρό, με αποτέλεσμα λιγότερα εικονοστοιχεία να ταξινομηθούν ως προσκήνιο. Αντιθέτως, όταν μειώνεται, το threshold γίνεται πιο επιεικές, και έχει ως αποτέλεσμα περισσότερα εικονοστοιχεία να περνούν το κατώφλι.

Ουσιαστικά, το C είναι μια σταθερά που αφαιρείται από τη μέση τιμή για να διορθώσει τη γραμμή διαχωρισμού προσκηνίου–παρασκηνίου.

Αυτή η μέθοδος, είναι καταλληλότερη για εικόνες με ετερογενές φωτισμό ή τοπικές διακυμάνσεις, όμως απαιτεί την εύρεση των σωστών τιμών των υπερπαραμέτρων, όπως το μέγεθος του παραθύρου και τη σταθερά C.

3.1.3 Πρακτική Εφαρμογή και Σύγκριση Thresholding Τεχνικών

Για την εμπεριστατωμένη κατανόηση των μεθόδων global και adaptive thresholding, πραγματοποιήθηκε ένα πολύ σύντομο σε μια πραγματική εικόνα εγγράφου. Το πείραμα υλοποιήθηκε σε περιβάλλον `Python`, κάνοντας χρήση των βιβλιοθηκών `OpenCV2` και `Matplotlib` για την απεικόνιση των διαγραμμάτων. Η εικόνα είναι μία τυπική εικόνα κειμένου με χρωματισμό κλίμακας του γκρι και προέρχεται από το σύνολο δεδομένων της βιβλιοθήκης `skimage`.

Το πείραμα, υλοποιεί τις δύο διαφορετικές τεχνικές κατωφλίωσης που αναλύθηκαν:

1. **Global thresholding με τη μέθοδο Otsu**, η οποία προσδιορίζει αυτόματα το βέλτιστο κατώφλι για το σύνολο της εικόνας όπως αναλύθηκε προηγουμένως.
2. `Adaptive Mean Thresholding`, η οποία υπολογίζει το κατώφλι με βάση τη μέση τιμή του εκάστοτε τοπικού παραθύρου.

Όπως φαίνεται και από τα αποτελέσματα, η τεχνική global thresholding αποδίδει ικανοποιητικά σε περιβάλλοντα με ομοιόμορφο φωτισμό, αλλά αποτυγχάνει σε περιπτώσεις σκιάσεων ή φωτεινών μεταβολών. Αντίθετα, οι adaptive μέθοδοι προσφέρουν μεγαλύτερη ακρίβεια σε πραγματικές συνθήκες, ειδικά όταν υπάρχει ανομοιογενής φωτισμός ή τοπικές παραμορφώσεις.

3.2 Εντοπισμός Ακμών

Ο εντοπισμός ακμών (edge detection) είναι μία θεμελιώδης διαδικασία στην επεξεργασία εικόνας και όραση υπολογιστών, που έχει ως στόχο την ανάδειξη των σημαντικών αλλαγών στη φωτεινότητα μιας εικόνας. Αυτές οι μεταβολές συνήθως αντιστοιχούν σε φυσικά όρια αντικειμένων, άκρες ή σημεία όπου αλλάζει η υφή ή η σκίαση [79]. Ουσιαστικά, οι ακμές αποτελούν τις περιοχές όπου παρατηρείται μεγάλη τοπική μεταβολή στην ένταση των εικονοστοιχείων.

Αναλυτικότερα, ο εντοπισμός των ακμών, στις περισσότερες τεχνικές όπως ο αλγόριθμος του Sobel [67] βασίζεται στον υπολογισμό των παραγώγων της έντασης της εικόνας, όπου μια υψηλή τιμή της παραγώγου υποδηλώνει απότομη μεταβολή — δηλαδή ακμή. Και αυτή η τεχνική όπως και το thresholding, αξιοποιούν ένα κατώφλι για την ταξινόμηση των εικονοστοιχείων, στη προκειμένη περίπτωση την ταξινόμηση της ακμής.

Ένας διαφορετικός, πιο εκλεπτυσμένος αλγόριθμος είναι ο Canny [19], ο οποίος εξομαλύνει πρώτα την εικόνα πριν υπολογίσει τους παραγώγους, αξιοποιώντας το Gaussian φίλτρο.

3.2.1 Πρακτική Εφαρμογή και Σύγκριση

Στα πλαίσια αυτής της υποενότητας, γίνεται μια απλή πρακτική και σύγκριση των τεχνικών εντοπισμού ακμών. Όπως και στο προηγούμενο παράδειγμα, χρησιμοποιήθηκε η γλώσσα python, και τα πακέτα `Matplotlib` για την απεικόνιση των διαγραμμάτων, και `skimage` για την δειγματοληπτική ψηφιακή εικόνα.

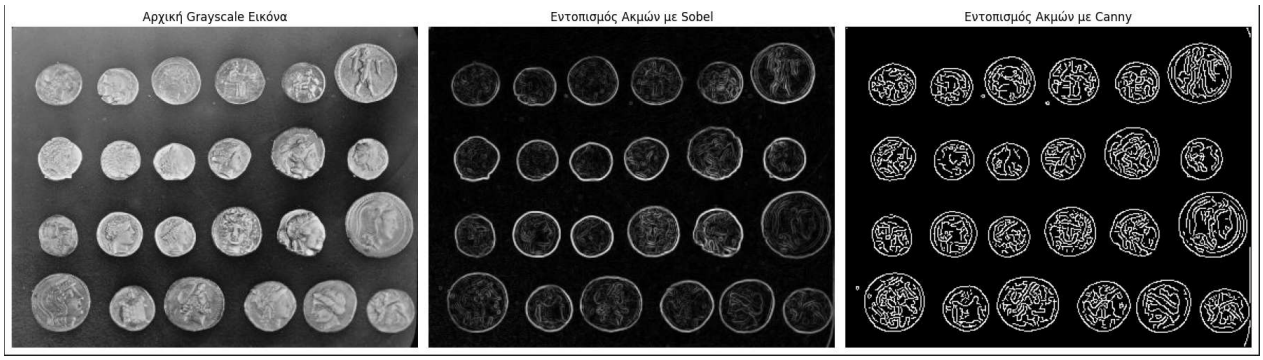
Αναλυτικότερα, χρησιμοποιήθηκε μια εικόνα με νομίσματα, ιδανική για την κατανόηση των αλγορίθμων εντοπισμού ακμών, εκ του αποτελέσματος.

Η σύγκριση των αποτελεσμάτων παρουσιάζεται σε κοινό γράφημα, αποκαλύπτοντας τις διαφορές ανάμεσα στις δύο προσεγγίσεις: ο Sobel παρέχει μια γενική εικόνα των ακμών, ενώ ο Canny παράγει ένα πιο καθαρό και ομαλό αποτέλεσμα, ιδανικό για εφαρμογές όπου απαιτείται ακριβέστερη απομόνωση των αντικειμένων. Το παράδειγμα αυτό υπογραμμίζει τη σημασία της επιλογής κατάλληλης τεχνικής ανάλογα με τις ανάγκες της εκάστοτε εφαρμογής στην επεξεργασία εικόνας.

3.3 Τεχνικές Region Based

Οι τεχνικές region based, χωρίζουν την εικόνα σε συνεκτικές περιοχές όπου τα εικονοστοιχεία έχουν παρόμοια χαρακτηριστικά (ένταση, χρώμα, υφή) και είναι χωρικά συνεχή.

Ένα παράδειγμα είναι η Seeded Region Growing όπου επιλέγονται ένα ή περισσότερα καίρια σημεία (seed εικονοστοιχεία) και επεκτείνουμε τη «περιοχή» γύρω τους προσθέτοντας γειτονικά εικονοστοιχεία που πληρούν κάποιο κριτήριο ομοιογένειας (π.χ. ίδιο χρώμα).



Σχήμα 3.2: Αποτελέσματα Πειράματος

Οι περιοχικές μέθοδοι παράγουν ομοιογενείς, λογικά συνεχή περιοχές (π.χ. «κομμάτια» αντικειμένων) χωρίς να απαιτούν προηγούμενη εκπαίδευση, και συχνά αποδίδουν καλά σε εικόνες όπου τα αντικείμενα παρουσιάζουν ομοιογένεια. Ωστόσο, είναι ευάλωτες αν η επιλογή των καίριων εικονοστοιχείων είναι κακή.

Για παράδειγμα, εάν εφαρμοστούν πολλά καίρια εικονοστοιχεία, μπορούν να δημιουργήσουν πάρα πολλές μικρές περιοχές, ενώ στην αντίθετη περίπτωση, μπορούν να συγχωνεύσουν ανεπιθύμητα αντικείμενα. Γενικά, αυτές οι μέθοδοι τείνουν να αποδίδουν καλά σε απλές εικόνες και χρησιμοποιούνται ευρέως σε εφαρμογές ιατρικής απεικόνισης [24].

3.4 Αλγόριθμος Watershed

Ο αλγόριθμος watershed αποτελεί κλασική τεχνική κατάτμησης εικόνων που αντιμετωπίζει αποτελεσματικά το πρόβλημα των επικαλυπτόμενων ή εφαπτόμενων αντικειμένων, όπου οι παραδοσιακές μέθοδοι thresholding αποτυγχάνουν [60] [43]. Η μέθοδος βασίζεται στην έννοια του υδρογραφικού διαχωρισμού, θεωρώντας την εικόνα ως τρισδιάστατο τοπογραφικό ανάγλυφο.

3.4.1 Θεωρητική Προσέγγιση

Στον αλγόριθμο watershed, η εικόνα αντιμετωπίζεται ως τοπογραφική επιφάνεια όπου η ένταση κάθε εικονοστοιχείου αντιστοιχεί σε ένα υψόμετρο. Η διαδικασία προσομοιώνει πλημμύρα που ξεκινάει από προκαθορισμένα σημεία-δείκτες (markers) και εξαπλώνεται μέχρι να συναντήσει όρια υψηλού gradient. Μαθηματικά, ο watershed μετασχηματισμός ορίζεται ως το σύνολο των σημείων που ανήκουν στα όρια μεταξύ διαφορετικών λεκανών απορροής.

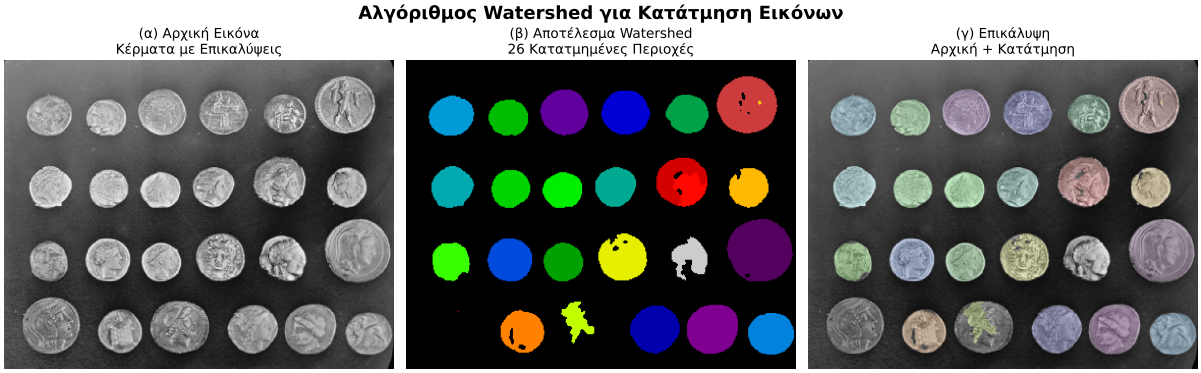
3.4.2 Πρακτική Υλοποίηση

Η υλοποίηση του αλγορίθμου περιλαμβάνει τέσσερα βασικά βήματα:

1. **Προεπεξεργασία**: Εφαρμογή median filter για μείωση θορύβου

2. **Gradient Υπολογισμός**: Χρήση Sobel operator για ανίχνευση ακμών
3. **Markers Δημιουργία**: μετασχηματισμός απόστασης για εύρεση κέντρων αντικειμένων
4. **Watershed Εφαρμογή**: Προσομοίωση πλημμύρας από τα σημεία-δείκτες

Η χρήση του μετασχηματισμού απόστασης για τη δημιουργία σημείων-δεικτών εξασφαλίζει ότι κάθε τοπικό μέγιστο αντιστοιχεί σε κέντρο αντικειμένου, επιτρέποντας τον αυτόματο εντοπισμό επικαλυπτόμενων δομών.



Σχήμα 3.3: Εφαρμογή αλγορίθμου watershed σε εικόνα με επικαλυπτόμενα κέρματα. (α) Αρχική εικόνα όπου παραδοσιακές μέθοδοι thresholding θα αναγνώριζαν ένα ενιαίο αντικείμενο, (β) Αποτέλεσμα watershed που επιτυγχάνει επιτυχή κατάτμηση σε 24 ξεχωριστές περιοχές, (γ) Επικάλυψη αρχικής εικόνας με τα αποτελέσματα κατάτμησης για οπτική επαλήθευση της ακρίβειας.

3.4.3 Αποτελέσματα και Αξιολόγηση

Η εφαρμογή του αλγορίθμου στην εικόνα κερμάτων επέτυχε την κατάτμηση σε 24 ξεχωριστές περιοχές, αντιμετωπίζοντας επιτυχώς το πρόβλημα των επικαλυπτόμενων αντικειμένων. Το αποτέλεσμα επιδεικνύει τη δυνατότητα του watershed να παράγει ακριβή όρια μεταξύ εφαπτόμενων αντικειμένων, κάτι που δεν είναι εφικτό με παραδοσιακές τεχνικές thresholding.

Η επιτυχία του αλγορίθμου οφείλεται στη χρήση του μετασχηματισμού απόστασης για την αυτόματη εύρεση σημείων-δεικτών, η οποία εξασφαλίζει ότι κάθε αντικείμενο έχει ένα μοναδικό σημείο εκκίνησης. Αυτή η προσέγγιση επιτρέπει τον διαχωρισμό αντικειμένων που μοιράζονται κοινά όρια, καθιστώντας τον watershed ιδανικό για εφαρμογές όπου η ακριβής κατάτμηση είναι κρίσιμη.

3.4.4 Πλεονεκτήματα και Περιορισμοί

Ο αλγόριθμος watershed παρουσιάζει σημαντικά πλεονεκτήματα στην αντιμετώπιση επικαλυπτόμενων αντικειμένων. Η ικανότητά του να παράγει συνεχείς γραμμές διαχωρισμού μονοψήφιου πλάτους εξασφαλίζει ακριβή οριοθέτηση, ενώ η αυτόματη εύρεση σημείων δεικτών μέσω

του μετασχηματισμού απόστασης, μειώνει την ανάγκη χειροκίνητης παρέμβασης. Επιπλέον, η μαθηματικά θεμελιωμένη προσέγγιση παρέχει αναπαραγωγίσιμα αποτελέσματα.

Παρόλα αυτά, ο αλγόριθμος εμφανίζει περιορισμούς σε εικόνες με έντονο θόρυβο, όπου μπορεί να προκύψει υπερκατάτμηση (over-segmentation). Επιπροσθέτως, η αποτελεσματικότητα εξαρτάται από την ποιότητα των σημείων-δεικτών, απαιτώντας προσεκτική προεπεξεργασία εικονών για βέλτιστα αποτελέσματα.

3.5 Συμπεράσματα

Οι παραδοσιακές τεχνικές κατάτμησης εικόνας αποτέλεσαν τους ακρογωνιαίους λίθους της υπολογιστικής όρασης πριν την εποχή του deep learning. Χαρακτηρίζονται από απλότητα και ερμηνευσιμότητα, προσφέροντας ταχεία λύση σε προβλήματα όπου τα αντικείμενα διαχωρίζονται καθαρά με βάση την ένταση τους.

Αν και οι παραδοσιακές μέθοδοι έχουν ιστορική και θεωρητική αξία, υστερούν στην αντιμετώπιση πολύπλοκων, και υψηλής ποικιλομορφίας σκηνών. Η έλλειψη δυνατότητας γενίκευσης οδήγησε στη στροφή προς τις μεθόδους βαθιάς μάθησης, οι οποίες μπορούν να μάθουν ισχυρά χαρακτηριστικά απευθείας από τα δεδομένα, ξεπερνώντας πολλούς περιορισμούς των κλασικών προσεγγίσεων.

Κεφάλαιο 4ο: Βιβλιογραφική Ανασκόπηση

Η εισαγωγή της βαθιάς μάθησης μεταμόρφωσε ριζικά την τμηματοποίηση εικόνων, επιτρέποντας την εκμάθηση ιεραρχικών αναπαραστάσεων χαρακτηριστικών με προσέγγιση από άκρο σε άκρο. Η πορεία της βαθιάς μάθησης στην εννοιολογική τμηματοποίηση ξεκίνησε με την ανάπτυξη των Συνελικτικών Νευρωνικών Δικτύων [46]. Πριν από την εμφάνιση των CNNs, οι εργασίες τμηματοποίησης βασίζονταν σε μεγάλο βαθμό στη χειροκίνητη εξαγωγή χαρακτηριστικών και σε ευρετικές μεθόδους. Η εξέλιξη των δικτύων για την εννοιολογική τμηματοποίηση ξεκίνησε από μια μικρή τροποποίηση των εκάστοτε κορυφαίων (state-of-the-art– SotA) μοντέλων ταξινόμησης, σηματοδοτώντας μία παραδειγματική αλλαγή προς τις εκμαθημένες αναπαραστάσεις, αντί των χειροποίητων χαρακτηριστικών.

4.1 Πλήρως Συνελικτικά Δίκτυα (Fully Convolutional Networks - FCN)

Τα Πλήρως Συνελικτικά Δίκτυα (FCN) αποτελούν κομβικό σημείο στην εξέλιξη της σημασιολογικής κατάτμησης, καθώς εισήγαγαν την έννοια της άμεσης εκμάθησης pixel-to-pixel απεικονίσεων [39]. Το κύριο καινοτόμο στοιχείο των FCN έγκειται στην αντικατάσταση των πλήρως

συνδεδεμένων στρωμάτων των παραδοσιακών CNN με συνελικτικά στρώματα, επιτρέποντας την επεξεργασία εικόνων αυθαίρετου μεγέθους.

Οι τρεις κύριες παραλλαγές των FCN - FCN-32, FCN-16 και FCN-8 - διαφοροποιούνται ως προς τον τρόπο με τον οποίο συνδυάζουν πληροφορίες από διαφορετικά επίπεδα του δικτύου για την επίτευξη βελτιωμένης χωρικής ανάλυσης.

4.1.1 FCN-32 - Η Βασική Αρχιτεκτονική

Το FCN-32 αποτελεί την πιο απλή μορφή των πλήρως συνελικτικών δικτύων, όπου η τελική πρόβλεψη πραγματοποιείται με άμεση αναδειγματοληψία (upsampling) κατά τον συντελεστή 32 από το τελικό στρώμα χαρακτηριστικών, παράγωντας έξοδο κατά 32 φορές μικρότερη από την αρχική είσοδο. Αυτή η προσέγγιση χρησιμοποιεί μόνο το βαθύτερο στρώμα του δικτύου για την παραγωγή της σημασιολογικής κατάτμησης, το οποίο περιέχει υψηλού επιπέδου σημασιολογικές πληροφορίες αλλά χαμηλή χωρική ανάλυση [39].

Η αρχιτεκτονική FCN-32 υφίσταται από σημαντικούς περιορισμούς λόγω του μεγάλου βήματος (stride) 32 εικονοστοιχεία στο τελικό στρώμα πρόβλεψης, γεγονός που περιορίζει την κλίμακα λεπτομερειών στην τελική έξοδο. Παρά την απλότητά της, η αρχιτεκτονική αυτή παρέχει αποδεκτά αποτελέσματα για εφαρμογές όπου η ακριβής χωρική ανάλυση δεν είναι κρίσιμη.

4.1.2 FCN-16 Ενσωμάτωση Ενδιάμεσων Χαρακτηριστικών

Το FCN-16 εισάγει την έννοια των συνδέσεων παράκαμψης για τη βελτίωση της χωρικής ακρίβειας της κατάτμησης. Η αρχιτεκτονική αυτή συνδυάζει προβλέψεις από ένα ενδιάμεσο στρώμα του κωδικοποιητή με τις προβλέψεις του τελικού βαθιού στρώματος.

Η διαδικασία περιλαμβάνει τη δημιουργία επιπλέον προβλεψιακών στρωμάτων επάνω στα ενδιάμεσα στρώματα χαρακτηριστικών, τα οποία στη συνέχεια συνενώνονται με τις προβλέψεις του βαθύτερου στρώματος μέσω πρόσθεσης. Αυτή η προσέγγιση επιτρέπει στο μοντέλο να συνδυάζει σημασιολογικές πληροφορίες υψηλού επιπέδου με χωρικές πληροφορίες μέσου επιπέδου.

Πειραματικά αποτελέσματα καταδεικνύουν ότι το FCN-16 επιτυγχάνει σημαντική βελτίωση στην απόδοση σε σχέση με το FCN-32, παρουσιάζοντας καλύτερη ισορροπία μεταξύ σημασιολογικής ακρίβειας και χωρικής λεπτομέρειας.

4.1.3 FCN-8 Η Εξελιγμένη Αρχιτεκτονική

Το FCN-8 αντιπροσωπεύει την πιο εξελιγμένη εκδοχή των FCN αρχιτεκτονικών, ενσωματώνοντας πληροφορίες από πολλαπλά επίπεδα του κωδικοποιητή μέσω σύνθετων συνδέσεων πα-

ράκαμψης. Η αρχιτεκτονική αυτή συνδυάζει προβλέψεις από ρηχότερα στρώματα με υψηλή χωρική ανάλυση με τις ήδη συνενωμένες προβλέψεις από βαθύτερα στρώματα.

Η πολυεπίπεδη σύντηξη χαρακτηριστικών επιτυγχάνεται μέσω σειράς προοδευτικών συνδυασμών, όπου κάθε βήμα ενσωματώνει πληροφορίες από ένα επιπλέον επίπεδο ανάλυσης. Αυτή η προσέγγιση επιτρέπει την ενσωμάτωση χαρακτηριστικών από πολλαπλές κλίμακες, συνδυάζοντας βαθιές σημασιολογικές πληροφορίες με ρηχές, λεπτομερείς πληροφορίες εμφάνισης.

4.1.4 Συγκριτική ανάλυση

Οι τρεις αρχιτεκτονικές παρουσιάζουν διαφορετικά χαρακτηριστικά ως προς την υπολογιστική πολυπλοκότητα και την ποιότητα των αποτελεσμάτων. Το FCN-32 προσφέρει την απλούστερη υλοποίηση με τη χαμηλότερη υπολογιστική απαίτηση, ενώ το FCN-8 παρέχει την υψηλότερη χωρική ακρίβεια με αυξημένο υπολογιστικό κόστος.

Η προοδευτική βελτίωση από το FCN-32 στο FCN-8 αποδεικνύεται σε διάφορες εφαρμογές, συμπεριλαμβανομένων της ιατρικής απεικόνισης και της αυτόνομης οδήγησης. Η επιλογή μεταξύ των τριών αρχιτεκτονικών εξαρτάται από τις συγκεκριμένες απαιτήσεις της εφαρμογής ως προς την ισορροπία μεταξύ ακρίβειας και υπολογιστικής αποδοτικότητας.

4.1.5 Τεχνικές Υλοποίησης και Βελτιστοποιήσεις

Οι σύγχρονες υλοποιήσεις των FCN αρχιτεκτονικών ενσωματώνουν προηγμένες τεχνικές βελτιστοποίησης για τη βελτίωση της αποδοτικότητας. Η εφαρμογή ελαφρών συνελικτικών μονάδων και μηχανισμών προσοχής επιτρέπει τη διατήρηση υψηλής απόδοσης με μειωμένο υπολογιστικό κόστος.

Η ενσωμάτωση τεχνικών μεταφοράς γνώσης από προ-εκπαιδευμένα δίκτυα ταξινόμησης αποτελεί κρίσιμο παράγοντα για την επίτευξη βέλτιστης απόδοσης σε όλες τις παραλλαγές FCN. Οι μελλοντικές κατευθύνσεις περιλαμβάνουν την ανάπτυξη υβριδικών αρχιτεκτονικών που συνδυάζουν τα πλεονεκτήματα κάθε παραλλαγής για εξειδικευμένες εφαρμογές [39].

4.1.6 Μαθηματική Διατύπωση

Για μια είσοδο $X \in \mathbb{R}^{H \times W \times C}$, το FCN παράγει έξοδο $Y \in \mathbb{R}^{H \times W \times N}$ όπου N είναι ο αριθμός των κλάσεων. Η διαδικασία upsampling μπορεί να περιγραφεί ως:

$$Y_{up} = f_{upsample}(f_{conv}(X)) \tag{2}$$

όπου f_{conv} είναι η λειτουργία συνέλιξης και $f_{upsample}$ η λειτουργία αύξησης ανάλυσης.

Τα skip connections ενσωματώνουν πληροφορίες από διαφορετικά επίπεδα:

$$Y_{final} = f_{combine}(Y_{up}, F_{skip}) \tag{3}$$

όπου F_{skip} αντιπροσωπεύει τα χαρακτηριστικά από ενδιάμεσα στρώματα.

4.1.7 Πλεονεκτήματα και Περιορισμοί FCN

Τα Πλήρως Συνελικτικά Δίκτυα παρουσιάζουν σημαντικά πλεονεκτήματα στην άμεση εκμάθηση κατάτμησης σε επίπεδο εικονοστοιχείου, χωρίς την ανάγκη επεξεργασίας των αποτελεσμάτων. Η αρχιτεκτονική τους επιτρέπει την αποδοτική επεξεργασία εικόνων αυθαίρετου μεγέθους, καθώς η απουσία πλήρως συνδεδεμένων στρωμάτων εξαλείφει τους περιορισμούς σταθερού μεγέθους εισόδου. Επιπροσθέτως, η δυνατότητα μεταφοράς γνώσης από προεκπαιδευμένα μοντέλα αποτελεί κρίσιμο πλεονέκτημα, επιτρέποντας την αξιοποίηση χαρακτηριστικών που έχουν μαθευτεί από μεγάλα σύνολα δεδομένων. Η σχετικά απλή αρχιτεκτονική και υλοποίηση καθιστούν τα FCN προσβάσιμα για ερευνητικούς και βιομηχανικούς σκοπούς.

Παρά τα σημαντικά πλεονεκτήματα, τα FCN εμφανίζουν ουσιώδεις περιορισμούς που επηρεάζουν την απόδοσή τους σε σύνθετες εφαρμογές. Η περιορισμένη ικανότητα σύλληψης ολόκληρης της εικόνας αποτελεί τη βασικότερη αδυναμία, καθώς η αρχιτεκτονική εστιάζει κυρίως σε τοπικά χαρακτηριστικά. Η απώλεια χωρικών λεπτομερειών λόγω της υποδειγματοληψίας στα βαθύτερα στρώματα δημιουργεί προκλήσεις στην ακριβή οριοθέτηση λεπτών δομών. Επιπλέον, η δυσκολία στην κατάτμηση αντικειμένων με σημαντικές διαφορές κλίμακας και ο ανεπαρκής χειρισμός σύνθετων χωρικών σχέσεων αποτελούν περιορισμούς που οδήγησαν στην ανάπτυξη προηγμένων αρχιτεκτονικών.

4.2 Δίκτυα Ανάλυσης Πυραμιδικής Σκηνής: Η Αρχιτεκτονική PSPNet

Το Δίκτυο Ανάλυσης Πυραμιδικής Σκηνής [77] αντιπροσωπεύει μια σημαντική εξέλιξη στον τομέα της σημασιολογικής κατάτμησης εικόνων, εισάγοντας καινοτόμες προσεγγίσεις για την αξιοποίηση παγκόσμιων πληροφοριών περιεχομένου. Η θεμελιώδης καινοτομία του PSPNet [77] έγκειται στην ικανότητα εκμετάλλευσης πληροφοριών παγκόσμιου περιεχομένου μέσω συγκέντρωσης περιεχομένου βασισμένης σε διαφορετικές περιοχές, χρησιμοποιώντας την προτεινόμενη μονάδα πυραμιδικής συγκέντρωσης.

Η αρχιτεκτονική PSPNet αντιμετωπίζει τις προκλήσεις της ανάλυσης σκηνών που χαρακτηρίζονται από απεριόριστο ανοικτό λεξιλόγιο και ποικίλες σκηνές. Η παγκόσμια προτεραία αναπαράσταση του PSPNet αποδεικνύεται αποτελεσματική για την παραγωγή υψηλής ποιότητας αποτελεσμάτων στην εργασία ανάλυσης σκηνών, παρέχοντας ανώτερο πλαίσιο για εργασίες πρόβλεψης σε επίπεδο εικονοστοιχείου.

4.2.1 Πυραμιδική Μονάδα Συγκέντρωσης

Η πυραμιδική μονάδα συγκέντρωσης αποτελεί τον πυρήνα της αρχιτεκτονικής PSPNet, εκμεταλλευόμενη χαρακτηριστικά από διαφορετικές κλίμακες για την κατασκευή μιας ιεραρχικής αναπαράστασης. Η μονάδα αυτή πραγματοποιεί συγκέντρωση σε πολλαπλές κλίμακες, δημιουργώντας έναν χάρτη χαρακτηριστικών που ενσωματώνει τοπικές και παγκόσμιες πληροφορίες με συστηματικό τρόπο.

Η λειτουργικότητα της πυραμιδικής συγκέντρωσης βασίζεται στη δημιουργία τεσσάρων διαφορετικών επιπέδων με κλίμακες 1×1, 2×2, 3×3 και 6×6, επιτρέποντας την αποτύπωση χαρακτηριστικών σε διαφορετικές χωρικές αναλύσεις. Αυτή η προσέγγιση παρέχει στο μοντέλο την ικανότητα κατανόησης του περιεχομένου τόσο σε τοπικό όσο και σε παγκόσμιο επίπεδο, βελτιώνοντας την ακρίβεια της σημασιολογικής κατάτμησης.

4.2.2 Αρχιτεκτονικά Χαρακτηριστικά

Το PSPNet υιοθετεί μια αρχιτεκτονική κωδικοποιητή-αποκωδικοποιητή που ενσωματώνει προεκπαιδευμένα βαθιά δίκτυα ως τον κορμό (backbone) εξαγωγής χαρακτηριστικών. Η αρχιτεκτονική αυτή επιτυγχάνει αποτελεσματική εξισορρόπηση μεταξύ της διατήρησης λεπτομερών χωρικών πληροφοριών και της αξιοποίησης σημασιολογικών πληροφοριών υψηλού επιπέδου.

4.2.3 Βελτιωμένες Παραλλαγές και Εφαρμογές

Οι σύγχρονες παραλλαγές του PSPNet ενσωματώνουν προηγμένες τεχνικές για τη βελτίωση της απόδοσης σε εξειδικευμένες εφαρμογές. Το Shift Pooling PSPNet [74] εισάγει τη μονάδα μετατόπισης πυραμιδικής συγκέντρωσης για την αντιμετώπιση του προβλήματος της μη πλήρους εξαγωγής τοπικών χαρακτηριστικών στις άκρες του πλέγματος.

Το DSCA-PSPNet επιτρέπει την προσαρμοστική εστίαση σε σχετικές χωρικές και καναλιακές πληροφορίες [75]. Αυτές οι βελτιώσεις ενισχύουν την ικανότητα του μοντέλου να χειρίζεται πολύπλοκες σκηνές με μεταβλητές περιβαλλοντικές συνθήκες και φασματικές ομοιότητες.

4.3 Συναρτήσεις Απώλειας για Σημασιολογική Κατάτμηση

Η σημασιολογική τμηματοποίηση απαιτούν εξειδικευμένες συναρτήσεις απωλειών, οι οποίες αντιμετωπίζουν την ανισορροπία μεταξύ κλάσεων και δίνουν έμφαση στην ακρίβεια των ορίων. Η συνάρτηση απώλειας cross-entropy παραμένει η καθιερωμένη επιλογή, ωστόσο έχουν αναπτυχθεί παραλλαγές όπως η weighted cross-entropyη και η Dice loss, ώστε να αντιμετωπιστούν οι ιδιαίτερες προκλήσεις που εμφανίζονται στις εργασίες τμηματοποίησης.

4.3.1 Συνάρτηση Απώλειας Cross-Entropy

Η συνάρτηση απώλειας cross-entropy παραμένει η θεμελιώδης και ευρέως υιοθετημένη συνάρτηση απώλειας για εργασίες σημασιολογικής κατάτμησης, λειτουργώντας ως η βάση για προβλήματα ταξινόμησης σε επίπεδο εικονοστοιχείου. Μαθηματικά ορίζεται ως $L_{CE} = -\sum\sum\sum y_{i,j,k} \log(\hat{y}_{i,j,k})$, όπου το $y_{i,j,k}$ αντιπροσωπεύει την πραγματική ετικέτα one-hot encoded και το $\hat{y}_{i,j,k}$ δηλώνει την προβλεπόμενη πιθανότητα για την κλάση k στη θέση εικονοστοιχείου (i, j). Η συνάρτηση απώλειας cross-entropy, μετρά τη διαφορά μεταξύ των προβλεπόμενων κατανομών πιθανότητας και των πραγματικών ετικετών σε όλες τις χωρικές θέσεις. Η συνάρτηση τιμωρεί αποτελεσματικά τις σίγουρες εσφαλμένες προβλέψεις πιο σοβαρά από τις αβέβαιες προβλέψεις, παρέχοντας σταθερές κλίσεις που διευκολύνουν την αποδοτική βελτιστοποίηση μέσω της οπισθοδρομής. Η συνάρτηση απώλειας cross-entropy επιδεικνύει ιδιαίτερη αποτελεσματικότητα σε ισορροπημένα σύνολα δεδομένων όπου όλες οι κλάσεις είναι περίπου ίσα αντιπροσωπευόμενες, καθώς αντιμετωπίζει κάθε εικονοστοιχείο και κλάση ομοιόμορφα χωρίς εγγενείς μηχανισμούς διόρθωσης μεροληψίας. Ωστόσο, ο κύριος περιορισμός της έγκειται στην ανικανότητά της να αντιμετωπίσει ζητήματα ανισορροπίας κλάσεων που συναντώνται συνήθως σε εργασίες κατάτμησης, όπου τα εικονοστοιχεία του φόντου συχνά ξεπερνούν κατά πολύ τα εικονοστοιχεία των αντικειμένων προσκηνίου, οδηγώντας σε μεροληψία βελτιστοποίησης προς τις κλάσεις πλειονότητας και υποβέλτιστη απόδοση στις κλάσεις μειονότητας ενδιαφέροντος [73].

4.3.2 Σταθμισμένη Συνάρτηση Απώλειας Cross-Entropy

Η σταθμισμένη συνάρτηση απώλειας cross-entropy αντιπροσωπεύει μια άμεση επέκταση της τυπικής συνάρτησης απώλειας cross-entropy που σχεδιάστηκε για να αντιμετωπίσει την ανισορροπία κλάσεων με την ενσωμάτωση ειδικών βαρών για κάθε κλάση που προσαρμόζουν τη συμβολή κάθε κλάσης στον συνολικό υπολογισμό απώλειας. Η μαθηματική διατύπωση $L_{\text{WCE}} = -\sum_i \sum_j \sum_k w_k \, y_{i,j,k} \log(\hat{y}_{i,j,k})$ εισάγει βάρη w_k για κάθε κλάση k, που συνήθως υπολογίζονται ως το αντίστροφο των συχνοτήτων κλάσεων ή μέσω πιο εξελιγμένων σχημάτων όπως ο αποτελεσματικός αριθμός δειγμάτων. Αυτός ο μηχανισμός στάθμισης επιτρέπει στη συνάρτηση απώλειας να δίνει έμφαση σε υποεκπροσωπούμενες κλάσεις κατά την εκπαίδευση, εξισορροπώντας τη φυσική μεροληψία προς τις κλάσεις πλειονότητας που εμφανίζεται με την τυπική συνάρτηση απώλειας cross-entropy. Η στρατηγική επιλογής βαρών επηρεάζει σημαντικά την απόδοση, με κοινές προσεγγίσεις να περιλαμβάνουν τη στάθμιση αντίστροφης συχνότητας ($w_k = 1/f_k$ όπου f_k είναι η συχνότητα της κλάσης k), την εξισορρόπηση διάμεσης συχνότητας και τη λογαριθμική στάθμιση συχνότητας. Εμπειρικές μελέτες έχουν αποδείξει ότι η σταθμισμένη cross-entropy μπορεί να βελτιώσει την απόδοση κατάτμησης σε ανισορροπημένα σύνολα δεδομένων, ιδιαίτερα για εφαρμογές όπως η κατάτμηση ιατρικών εικόνων όπου τα εικονοστοιχεία βλαβών ή ανατομικών δομών αποτελούν ένα μικρό κλάσμα των συνολικών εικονοστοιχείων της εικόνας, αν και η βέλτιστη στρατηγική στάθμισης παραμένει εξαρτώμενη από την εργασία και απαιτεί προσεκτική ρύθμιση υπερπαραμέτρων για την επίτευξη βέλτιστων αποτελεσμάτων [8].

4.3.3 Συνάρτηση Απώλειας Dice

Η συνάρτηση απώλειας Dice, που προέρχεται από τον συντελεστή ομοιότητας Dice που χρησιμοποιείται συνήθως ως μετρική αξιολόγησης, βελτιστοποιεί άμεσα την επικάλυψη μεταξύ προβλεπόμενων και πραγματικών κατατμήσεων, καθιστώντας την ιδιαίτερα κατάλληλη για εφαρμογές που απαιτούν ακριβή οριοθέτηση ορίων και στιβαρή απόδοση υπό συνθήκες ανισορροπίας κλάσεων.

Η συνάρτηση απώλειας εκφράζεται μαθηματικά ως

$$L_{\text{Dice}} = 1 - \frac{2 \sum \hat{y} y}{\sum \hat{y} + \sum y}$$

για δυαδική κατάτμηση ή

$$L_{\text{Dice}} = 1 - \frac{2 \sum \sum y_{i,j,k} \hat{y}_{i,j,k}}{\sum \sum y_{i,j,k} + \sum \sum \hat{y}_{i,j,k}}$$

για σενάρια πολλαπλών κλάσεων, όπου οι αθροίσεις εκτελούνται σε χωρικές διαστάσεις και κλάσεις αντίστοιχα.

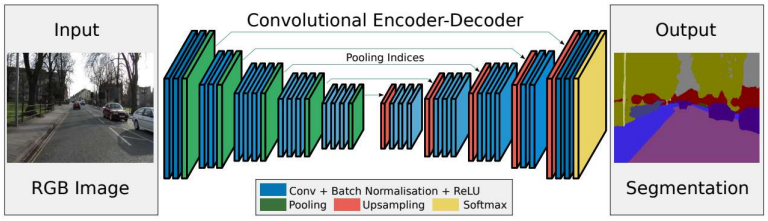
Σε αντίθεση με την συνάρτηση απώλειας cross-entropy, που λειτουργεί σε επίπεδο εικονοστοιχείου, η συνάρτηση απώλειας Dice παρέχει ένα παγκόσμιο μέτρο ποιότητας κατάτμησης που χειρίζεται εγγενώς την ανισορροπία κλάσεων εστιάζοντας στην επικάλυψη μεταξύ προβλεπόμενων και πραγματικών περιοχών αντί για μεμονωμένες ταξινομήσεις εικονοστοιχείων.

Αυτή η προσέγγιση παγκόσμιας βελτιστοποίησης καθιστά τη Dice loss ιδιαίτερα αποτελεσματική για εργασίες κατάτμησης ιατρικών εικόνων όπου η ακριβής οριοθέτηση ανατομικών δομών ή παθολογικών περιοχών είναι κρίσιμη, καθώς βελτιστοποιεί άμεσα τη μετρική πιο σχετική για κλινική αξιολόγηση.

Ωστόσο, η συνάρτηση απώλειας Dice μπορεί να εμφανίσει ασταθείς κλίσεις όταν ο συντελεστής Dice πλησιάζει το μηδέν, και μπορεί να αντιμετωπίσει δυσκολίες με την αποδοτικότητα βελτιστοποίησης σε σύγκριση με απώλειες επιπέδου εικονοστοιχείου, οδηγώντας πολλούς ερευνητές να συνδυάζουν τη συνάρτηση απώλειας Dice με την cross-entropy για να αξιοποιήσουν τα οφέλη και των δύο προσεγγίσεων [73].

4.4 SegNet

Το SegNet είναι μια αρχιτεκτονική συνελικτικού νευρωνικού δικτύου η οποία προτάθηκε για το πρόβλημα της κατηγοριοποίησης κάθε εικονοστοιχείου σε εικόνες. Η αρχιτεκτονική αποτελείται από δύο συμμετρικά τμήματα: τον κωδικοποιητή και τον αποκωδικοποιητή. Ο κωδικοποιητής βασισμένος στη δομή του VGG-16, —αναλυτικότερα στα 13 πρώτα συνελικτικά του στρώματα— αποτελείται από διαδοχικά στρώματα συνελίξεων, ακολουθούμενο από ένα στρώμα κανονικοποίησης (BN), ενεργοποίησης και υποδειγματοληψίας max-pooling [9]. Κατά τη



Σχήμα 4.1: Αρχιτεκτονική SegNet

φάση της υποδειγματοληψίας, αποθηκεύονται οι δείκτες (indices) των μέγιστων τιμών κάθε περιοχής, οι οποίοι στη συνέχεια χρησιμοποιούνται από τον αποκωδικοποιητή για την ακριβή επαναφορά της χωρικής πληροφορίας κατά τη διαδικασία του upsampling.

Ο αποκωδικοποιητής, αξιοποιεί τις αποθηκευμένες θέσεις των max-pooling για να εκτελέσει μη γραμμική υπερδειγματοληψία (non-linear upsampling), μειώνοντας την ανάγκη για εκμάθηση πολύπλοκων φίλτρων υπερδειγματοληψίας, και βελτιώνοντας τη χωρική ακρίβεια της εξόδου. Το τελικό επίπεδο του δικτύου αποτελείται από ένα στρώμα softmax, το οποίο παράγει προβλέψεις σε επίπεδο εικονοστοιχείου για κάθε κατηγορία.

Σε αντίθεση με άλλες αρχιτεκτονικές, το SegNet δεν περιλαμβάνει πλήρως συνδεδεμένα επίπεδα, γεγονός που μειώνει το υπολογιστικό του κόστος και καθιστά την αρχιτεκτονική κατάλληλη για εφαρμογές όπου απαιτείται ταχύτερη επεξεργασία και χαμηλότερη κατανάλωση μνήμης. Επιπλέον, η χρήση των δεικτών max-pooling οδηγεί σε βελτιωμένη διατήρηση των ορίων των αντικειμένων κατά την επανασύνθεση της εικόνας [9].

4.5 Αρχιτεκτονικές DeepLab

Η οικογένεια μοντέλων **DeepLab** αποτελεί μία από τις πιο επιδραστικές και ευρέως χρησιμοποιούμενες προσεγγίσεις για την κατάτμηση εικόνας σε επίπεδο εικονοστοιχείου (semantic segmentation). Αναπτύχθηκε από την ομάδα της Google Research [15] και εξελίχθηκε μέσω διαδοχικών εκδόσεων από την *DeepLabv1* έως και την *DeepLabv3+*, κάθε μία από τις οποίες εισήγαγε σημαντικές τεχνολογικές βελτιώσεις με στόχο την αύξηση της ακρίβειας και της αποδοτικότητας στην κατανόηση του περιεχομένου των εικόνων.

Η βασική φιλοσοφία πίσω από την ανάπτυξη των `DeepLab` μοντέλων είναι η διατήρηση της υψηλής χωρικής ανάλυσης και τη κατανόηση συμφραζομένων σε διαφορετικές κλίμακες, στοιχεία τα οποία κρίνονται καθοριστικά για ακριβή κατάτμηση. Η υιοθέτηση τεχνικών όπως οι διογκωμένες συνελίξεις (atrous convolutions), το *Atrous Spatial Pyramid Pooling (ASPP)* και η χρήση αποκωδικοποιητών σηματοδοτούν την εξέλιξη της αρχιτεκτονικής αυτής.

Στη συνέχεια παρουσιάζεται η αναλυτική περιγραφή κάθε έκδοσης του DeepLab, με έμφαση στα κύρια χαρακτηριστικά, τις διαφορές ως προς τις προηγούμενες εκδόσεις, καθώς και τις

βελτιώσεις στην απόδοση.

4.5.1 DeepLabv1

Το DeepLabv1 [15] αποτέλεσε το πρώτο μοντέλο της οικογένειας DeepLab, προτείνοντας μια αρχιτεκτονική που στοχεύει στη βελτίωση της ακρίβειας στην κατάτμηση εικόνας, μέσω της διατήρησης της χωρικής ανάλυσης και της εκμετάλλευσης του συμφραζόμενου περιεχομένου. Το βασικό χαρακτηριστικό της αρχιτεκτονικής ήταν η αντικατάσταση των συμβατικών συνελικτικών επιπέδων με διογκωμένες συνελίξεις (atrous convolutions), οι οποίες επιτρέπουν τον έλεγχο της ανάλυσης χαρακτηριστικών χωρίς αύξηση του πλήθους των παραμέτρων ή του υπολογιστικού κόστους.

Οι διογκωμένες συνελίξεις (atrous convolutions), γνωστές και ως "hole algorithm" στη βιβλιογραφία, αποτελούν μια τροποποίηση των κλασικών συνελικτικών πυρήνων όπου εισάγονται κενά (holes) μεταξύ των στοιχείων του φίλτρου. Η παράμετρος διαστολής ή dilation rate r καθορίζει το μέγεθος αυτών των κενών, με $r = 1$ να αντιστοιχεί στη κλασική συνέλιξη. Μαθηματικά, για έναν μονοδιάστατο πυρήνα $w[k]$ και ρυθμό διαστολής r, η διογκωμένη συνέλιξη υπολογίζεται ως:

$$y[i] = \sum_k x[i + r \cdot k]w[k]$$

Αυτή η τεχνική επιτρέπει την επέκταση του οπτικού πεδίου χωρίς αύξηση των παραμέτρων, καθώς ο πυρήνας "δειγματοληπτεί" πιο αραιά την είσοδό του. Στο DeepLabv1, οι διογκωμένες συνελίξεις atrous εφαρμόζονται για τη διατήρηση υψηλότερης χωρικής ανάλυσης ενώ παράλληλα διευρύνουν το οπτικό πεδίο για την καλύτερη σύλληψη της εισόδου.

Επιπλέον, για τη βελτίωση της ευκρίνειας στα όρια των αντικειμένων, η DeepLabv1 χρησιμοποίησε Fully Connected Conditional Random Fields (CRFs) ως επεξεργασία μετά την έξοδο του νευρωνικού δικτύου [15]. Τα CRFs ενισχύουν τη συνέπεια στην κατάτμηση, αξιοποιώντας τόσο τη χρωματική πληροφορία όσο και τη σχετική θέση των εικονοστοιχείων.

Παρότι το μοντέλο προσφέρει βελτιωμένη απόδοση σε σχέση με τα πλήρως συνελικτικά δίκτυα χωρίς διογκωμένες συνελίξεις, παρουσιάζει περιορισμούς στη σύλληψη πολυκλιμακωτής πληροφορίας, καθώς βασίζεται σε έναν μόνο ρυθμό διαστολής.

4.5.2 DeepLabv2

Η DeepLabv2 [16] εισήγαγε σημαντικές βελτιώσεις σε σχέση με την πρώτη έκδοση, με σκοπό τη βελτίωση της κατανόησης του γενικού πλαισίου και την ενίσχυση της γενίκευσης. Η πιο σημαντική καινοτομία ήταν η εισαγωγή του Atrous Spatial Pyramid Pooling (ASPP) [16], το οποίο αποτελεί έναν μηχανισμό εξαγωγής πολυκλιμακωτής πληροφορίας, εφαρμόζοντας πολλαπλές διογκωμένες συνελίξεις με διαφορετικούς ρυθμούς διαστολής σε παράλληλα μονοπάτια.

Το ASPP module αποτελείται από τέσσερις παράλληλους κλάδους διογκωμένων συνελίξεων με διαφορετικούς ρυθμούς διαστολής (συνήθως r = 6, 12, 18, 24), καθένας από τους οποίους συλλαμβάνει πληροφορίες σε διαφορετική χωρική κλίμακα. Οι χαμηλότεροι ρυθμοί διαστολής εστιάζουν σε τοπικές λεπτομέρειες και μικρά αντικείμενα, ενώ οι υψηλότεροι ρυθμοί επεκτείνουν το οπτικό πεδίο για τη σύλληψη ευρύτερου πλαισίου και μεγάλων δομών. οι χάρτες χαρακτηριστικών από όλους τους παράλληλους κλάδους συνενώνονται μέσω πρόσθεσης και στη συνέχεια επεξεργάζονται μέσω συνέλιξης 1×1 για τη δημιουργία ενοποιημένων χαρακτηριστικών που ενσωματώνουν πολυκλιμακωτή πληροφορία.

Επιπλέον βελτιώσεις του αλγορίθμου DeepLabv2 περιλαμβάνουν τη χρήση βαθύτερων κωδικοποιητών όπως το ResNet-101 αντί του VGG-16, καθώς και την επέκταση της εκπαίδευσης στο MS-COCO dataset για καλύτερη γενίκευση. η CRF μετά-επεξεργασία βελτιώθηκε με την υιοθέτηση του DenseCRF για πιο αποδοτικά αποτελέσματα.

Ωστόσο, παρά τις βελτιώσεις, η αρχιτεκτονική δεν περιλαμβάνει αποκωδικοποιητή και συνεπώς εξακολουθεί να παρουσιάζει αδυναμία στην ακριβή ανακατασκευή λεπτών δομών και ορίων, πρόβλημα που θα αντιμετωπιστεί στις μεταγενέστερες εκδόσεις της οικογένειας DeepLab.

4.5.3 DeepLabv3

Το DeepLabv3 [17] διατήρησε την κεντρική ιδέα του ASPP, την οποία ενίσχυσε περαιτέρω με σημαντικές αρχιτεκτονικές βελτιώσεις. Η νέα εκδοχή ενσωμάτωσε επιπλέον παραμέτρους διαστολής και αξιοποίησε τη παγκόσμια average pooling στην κατασκευή του ASPP, επιτρέποντας στο μοντέλο να λαμβάνει υπόψη του ευρύτερα σημασιολογικά πλαίσια της εικόνας.

Το βελτιωμένο ASPP τμήμα του DeepLabv3 αποτελείται από πέντε παράλληλους κλάδους: μία συνέλιξη 1×1, τρεις διογκωμένες συνελίξεις 3×3 με ρυθμούς διαστολής 6, 12, και 18, και ένα τμήμα image pooling που χρησιμοποιεί παγκόσμια average pooling για τη σύλληψη του γενικού πλαισίου. Κάθε συνελικτικό επίπεδο παράγει 256 χάρτες χαρακτηριστικών και περιλαμβάνει στρώμα batch normalization για βελτιωμένη σταθερότητα εκπαίδευσης. Τα αποτελέσματα από όλους τους κλάδους συνενώνονται και επεξεργάζονται μέσω μίας τελικής συνέλιξης 1×1 με 256 κανάλια πριν τη δημιουργία των τελικών προβλέψεων κλάσεων.

Μια ακόμη σημαντική αλλαγή είναι η κατάργηση των CRFs, υποδηλώνοντας εμπιστοσύνη στις ικανότητες του ίδιου του δικτύου να μοντελοποιήσει τις λεπτομέρειες των ορίων. Αυτή η απλοποίηση, οδήγησε σε μειωμένο υπολογιστικό χρόνο και χρόνο συμπεράσματος, ενώ παράλληλα διατήρησε ή και βελτίωσε την ακρίβεια κατάτμησης. Η DeepLabv3 βασίζεται εξ ολοκλήρου σε διογκωμένες συνελίξεις και έχει πλήρως συνελικτική μορφή, χωρίς την ύπαρξη αποκωδικοποιητή.

Επιπρόσθετες βελτιώσεις περιλαμβάνουν τη χρήση βαθύτερων αρχιτεκτονικών για κωδικοποιητή όπως το ResNet-101, καθώς και βελτιστοποιημένες στρατηγικές για τη διαχείριση των διο-

γκωμένων συνελίξεων σε διαφορετικά επίπεδα του δικτύου. Παρότι η απόδοσή του ήταν ιδιαίτερα υψηλή, ιδιαίτερα σε σύνολα δεδομένων όπως το PASCAL VOC 2012, η απουσία μηχανισμού ανάκτησης λεπτομερειών σε πρώιμα στάδια προκαλούσε περιορισμούς στην ακρίβεια των ορίων των αντικειμένων, ζήτημα που θα αντιμετωπιστεί στην επόμενη έκδοση με την εισαγωγή της αρχιτεκτονικής κωιδκοποιητή-αποκωδικοοποιητή.

4.5.4 DeepLabv3+

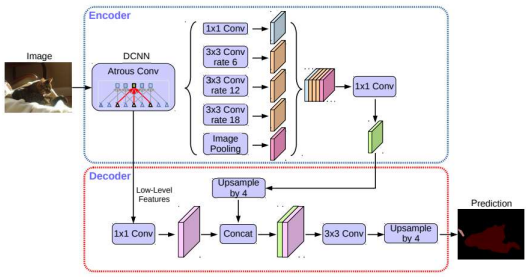
Το DeepLabv3+ [18] αποτέλεσε τη φυσική εξέλιξη της οικογένειας DeepLab, συνδυάζοντας τα πλεονεκτήματα του DeepLabv3 με την προσθήκη ενός αποκωδικοποιητή, με στόχο τη βελτίωση της χωρικής ακρίβειας. Η κεντρική καινοτομία του DeepLabv3+ είναι η υιοθέτηση μιας αρχιτεκτονικής κωδικοποιητή-αποκωδικοποιητή που διατηρεί το τμήμα ASPP ως κεντρική ενότητα του κωδικοποιητή ενώ εισάγει έναν απλό αλλά αποδοτικό αποκωδικοποιητή για την ανάκτηση των χωρικών λεπτομερειών.

Ο αποκωδικοποιητής λειτουργεί μέσω της σταδιακής αύξησης της ανάλυσης των χαρακτηριστικών εξόδου του τμήματος ASPP με συντελεστή 4, τα οποία στη συνέχεια συνδυάζονται με χαρακτηριστικά χαμηλού επιπέδου από τα πρώιμα στάδια του κωδικοποιητή μέσω άμεσων συνδέσεων παράκαμψης. Συγκεκριμένα, τα χαρακτηριστικά χαμηλού επιπέδου επεξεργάζονται μέσω συνέλιξης 1×1 για μείωση των διαστάσεων (συνήθως σε 48 κανάλια) πριν τη συνένωση με τα χαρακτηριστικά υψηλού επιπέδου αυξημένης ανάλυσης. Τα συνδυασμένα χαρακτηριστικά επεξεργάζονται μέσω συνελίξεων 3×3 και στη συνέχεια υφίστανται τελική αύξηση ανάλυσης κατά συντελεστή 4 για την παραγωγή της τελικής μάσκας κατάτμησης.

Επιπρόσθετη σημαντική βελτίωση αποτελεί η χρήση των διαχωρίσιμων συνελίξεων κατά βάθος (depthwise separable convolutions), ιδιαίτερα μέσω του δικτύου Xception ως αποκωδικοποιητή, το οποίο διαχωρίζει τη χωρική και την επεξεργασία κατά κανάλια σε ξεχωριστές διεργασίες. Αυτή η προσέγγιση μειώνει δραστικά τον αριθμό των παραμέτρων και το υπολογιστικό κόστος ενώ διατηρεί ή και βελτιώνει την ακρίβεια κατάτμησης. Το δίκτυο Xception, όταν συνδυάζεται με διογκωμένες συνελίξεις, παρέχει έναν ισχυρό εξαγωγέα χαρακτηριστικών που είναι τόσο αποδοτικός όσο και ακριβής.

Η προσέγγιση αυτή επιτρέπει στο δίκτυο να συνδυάζει την πολυκλιμακωτή κατανόηση που προσφέρει το ASPP με τη διατήρηση λεπτών χωρικών λεπτομερειών μέσω του αποκωδικοποιητή, κάτι που οδηγεί σε βελτιωμένες επιδόσεις, ιδίως στα όρια αντικειμένων. Η ισορροπία μεταξύ υπολογιστικής αποδοτικότητας και ακρίβειας κατάτμησης καθιστά το DeepLabv3+ ιδανικό για πρακτικές εφαρμογές όπου απαιτείται επεξεργασία σε πραγματικό χρόνο.

Το DeepLabv3+ αποτελεί έως σήμερα μια από τις πιο επιτυχημένες και ευρέως χρησιμοποιούμενες αρχιτεκτονικές στην κατάτμηση εικόνας, ενώ αποτέλεσε σημείο αναφοράς για μελλοντικές εξελίξεις με βάση τα μετασχηματιστές όρασης, καθώς η φιλοσοφία κωδικοποιητή-αποκωδικοποιητή του υιοθετήθηκε και προσαρμόστηκε σε νεότερες αρχιτεκτονικές.



Σχήμα 4.2: Αρχιτεκτονική DeepLabV3+

4.5.5 Σύνοψη των Αρχιτεκτονικών DeepLab

Η οικογένεια αλγορίθμων DeepLab αποτελεί μια από τις σημαντικότερες συνεισφορές στο πεδίο της σημασιολογικής κατάτμησης εικόνων, παρουσιάζοντας συνεχή εξέλιξη από το 2014 έως το 2018. Η προοδευτική ανάπτυξη των τεσσάρων εκδόσεων αντανακλά την τεχνολογική πρόοδο στη βαθιά μάθηση και την στοχευμένη αντιμετώπιση βασικών προκλήσεων της σημασιολογικής κατάτμησης.

Εξέλιξη και Βασικές Καινοτομίες

Το DeepLabv1 εισήγαγε δύο θεμελιώδεις καινοτομίες: τις διογκωμένες συνελίξεις (atrous convolutions) για τη διατήρηση της χωρικής ανάλυσης και τα πλήρως συνδεδεμένα τυχαία πεδία (Fully Connected CRFs) για την εξέλιξη των ορίων αντικειμένων. Αυτές οι τεχνικές αντιμετώπισαν το κεντρικό πρόβλημα της απώλειας λεπτομερειών που προκαλούσαν τα παραδοσιακά συνελικτικά επίπεδα με τους μηχανισμούς συγκέντρωσης (pooling operations).

Το DeepLabv2 επέκτεινε τις δυνατότητες με την εισαγωγή του Atrous Spatial Pyramid Pooling (ASPP), επιτρέποντας την εξαγωγή χαρακτηριστικών σε πολλαπλές κλίμακες μέσω παράλληλων διογκωμένων συνελίξεων με διαφορετικούς ρυθμούς διαστολής. Η προσθήκη βαθύτερων δικτύων κωδικοποιητών (ResNet-101) και η επέκταση εκπαίδευσης στο MS-COCO ενίσχυσαν περαιτέρω την απόδοση.

Το DeepLabv3 πραγματοποίησε καθοριστικές αρχιτεκτονικές αλλαγές με την κατάργηση των CRFs και την ενσωμάτωση παγκόσμιας average pooling στο βελτιωμένο ASPP module. Η μετάβαση σε αποκλειστικά πλήρως συνελικτικές προσεγγίσεις απλοποίησε την αρχιτεκτονική διατηρώντας παράλληλα υψηλή ακρίβεια.

Το DeepLabv3+ συνέθεσε την τελειοποιημένη μορφή με την προσθήκη αποκωδικοποιητή που συνδυάζει χαρακτηριστικά χαμηλού και υψηλού επιπέδου μέσω άμεσων συνδέσεων παράκαμψης. Η υιοθέτηση του Xception δίκτυου αποκωδικοποιητή με διαχωρίσιμες συνελίξεις κατά βάθος βελτίωσε την υπολογιστική αποδοτικότητα χωρίς απώλεια ακρίβειας.

Μετρήσιμη Πρόοδος Απόδοσης

Η προοδευτική εξέλιξη των DeepLab αρχιτεκτονικών αντικατοπτρίζεται στη σταδιακή βελτίωση της απόδοσης σε βασικά σύνολα αξιολόγησης. Στο PASCAL VOC 2012, η απόδοση mIoU βελτιώθηκε από 71.6% (DeepLabv1) σε 89.0% (DeepLabv3+), ενώ στο σύνολο δεδομένων Cityscapes το DeepLabv3+ πέτυχε 82.1% mIoU, καταδεικνύοντας την αποτελεσματικότητα των τεχνολογικών καινοτομιών.

Αρχιτεκτονική Συμβολή και Επιρροή

Η οικογένεια DeepLab καθιέρωσε βασικές αρχές που επηρέασαν ευρέως την ερευνητική κοινότητα: την αξιοποίηση διογκωμένων συνελίξεων για τη διατήρηση ανάλυσης, την πολυκλιμακωτή εξαγωγή χαρακτηριστικών μέσω ASPP, και την αποδοτική συνένωση υψηλού και χαμηλού επιπέδου πληροφοριών μέσω αρχιτεκτονικής κωδικοποιητή-αποκωδικοποιητή. Οι αρχές αυτές υιοθετήθηκε και προσαρμόστηκαν σε σύγχρονες αρχιτεκτονικές.

Η πρακτική εφαρμοσιμότητα των DeepLab μοντέλων επιβεβαιώνεται από την ευρεία υιοθέτησή τους σε διαφορετικούς τομείς, από την αυτόνομη πλοήγηση και τη γεωργική ρομποτική έως την ιατρική απεικόνιση και την τηλεπισκόπηση, καθιστώντας τα πρότυπα αναφοράς για την πρακτική εφαρμογή τεχνικών σημασιολογικής κατάτμησης.

4.6 Αρχιτεκτονική U-Net: Κωδικοποιητής-Αποκωδικοποιητής με Συνδέσεις Παράκαμψης

Η αρχιτεκτονική U-Net αποτελεί μια από τις πιο επιτυχημένες και ευρέως υιοθετημένες προσεγγίσεις στον τομέα της σημασιολογικής κατάτμησης, ιδιαίτερα στην ιατρική απεικόνιση [7] [62], όπως εισήχθη από τους Ronneberger, και συνεργάτες το 2015 [55]. Η θεμελιώδης καινοτομία της αρχιτεκτονικής έγκειται στη συμμετρική δομή κωδικοποιητή-αποκωδικοποιητή που συνδυάζεται με συνδέσεις παράκαμψης (skip connections), επιτρέποντας την αποτελεσματική διατήρηση χωρικών πληροφοριών σε πολλαπλές κλίμακες ανάλυσης.

Η ονομασία ”U-Net” προκύπτει από τη χαρακτηριστική με U-σχημα αρχιτεκτονική, η οποία αποτελείται από έναν συμμετρικό κωδικοποιητή που εκτελεί προοδευτική υποδειγματοληψία και έναν αποκωδικοποιητή που πραγματοποιεί αναδειγματοληψία για την αποκατάσταση της πρωτότυπης ανάλυσης. Αυτή η δομή επιτρέπει στο δίκτυο να συλλαμβάνει τόσο τοπικές όσο και παγκόσμιες πληροφορίες, καθιστώντας το ιδανικό για εργασίες που απαιτούν λεπτομερή χωρική ακρίβεια.

4.6.1 Δομικά Χαρακτηριστικά του Κωδικοποιητή

Ο κωδικοποιητής της αρχιτεκτονικής U-Net ακολουθεί μια προοδευτική στρατηγική υποδειγματοληψίας, όπου κάθε επίπεδο αποτελείται από δύο συνελίξεις 3×3 ακολουθούμενες από συνάρτηση ενεργοποίησης ReLU και μια λειτουργία υποδειγματοληψίας max-pooling 2×2. Αυτή

η διαδοχική διαδικασία επιτρέπει την εξαγωγή ιεραρχικών χαρακτηριστικών, ξεκινώντας από χαρακτηριστικά χαμηλού επιπέδου στα πρώτα στρώματα και προοδευτικά μεταβαίνοντας σε πιο αφηρημένες αναπαραστάσεις στα βαθύτερα στρώματα.

Η στρατηγική αύξησης του αριθμού των φίλτρων σε κάθε επίπεδο του κωδικοποιητή επιτρέπει την αποτύπωση όλο και πιο πολύπλοκων χαρακτηριστικών, ενώ παράλληλα μειώνει τη χωρική διάσταση των χαρτών χαρακτηριστικών. Αυτή η προσέγγιση διευκολύνει την εξαγωγή σημασιολογικά πλούσιων πληροφοριών που είναι απαραίτητες για την ακριβή κατάτμηση [71].

4.6.2 Μηχανισμός Συνδέσεων Παράκαμψης

Οι συνδέσεις παράκαμψης αποτελούν το κρίσιμο στοιχείο που διαφοροποιεί το U-Net από παραδοσιακές αρχιτεκτονικές κωδικοποιητή-αποκωδικοποιητή. Αυτές οι συνδέσεις μεταφέρουν χαρτές χαρακτηριστικών υψηλής ανάλυσης από τον κωδικοποιητή απευθείας στα αντίστοιχα επίπεδα του αποκωδικοποιητή μέσω πράξεων συνένωσης.

Η λειτουργία αυτή επιτρέπει την αποκατάσταση λεπτομερών χωρικών πληροφοριών που θα μπορούσαν να χαθούν κατά τη διαδικασία της υποδειγματοληψίας. Πρόσφατες μελέτες έχουν δείξει ότι η προσεκτική διαχείριση αυτών των συνδέσεων μπορεί να βελτιώσει σημαντικά την απόδοση της κατάτμησης, με τη βέλτιστη απόδοση να επιτυγχάνεται μέσω της επιλεκτικής αφαίρεσης της ανώτατης σύνδεσης παράκαμψης.

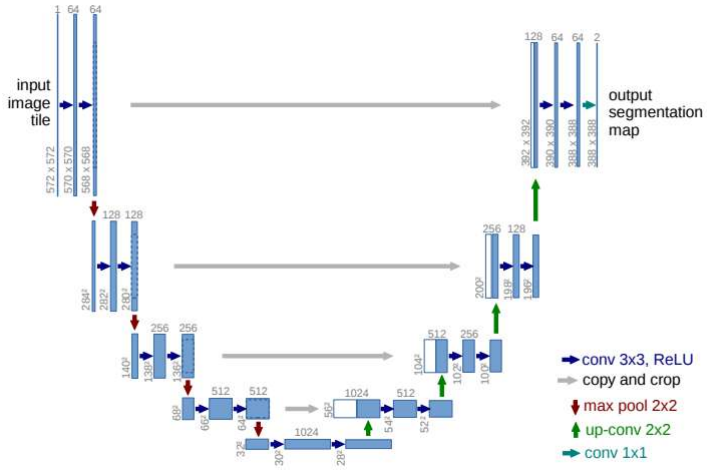
4.6.3 Δομή του Αποκωδικοποιητή

Ο αποκωδικοποιητής του U-Net πραγματοποιεί τη σταδιακή αποκατάσταση της χωρικής ανάλυσης μέσω μιας σειράς λειτουργιών αναδειγματοληψίας. Κάθε βήμα του αποκωδικοποιητή περιλαμβάνει μια λειτουργία αναδειγματοληψίας transposed ή bilinear που διπλασιάζει τη χωρική διάσταση, ακολουθούμενη από συνένωση με τον αντίστοιχο χάρτη χαρακτηριστικών από τον κωδικοποιητή.

Η διαδικασία ολοκληρώνεται με δύο συνελίξεις 3×3 και ReLU ενεργοποίηση, επιτρέποντας στο δίκτυο να συνδυάσει τις πληροφορίες υψηλής ανάλυσης από τον κωδικοποιητή με τις σημασιολογικές πληροφορίες που έχουν επεξεργαστεί από τον αποκωδικοποιητή. Αυτός ο συνδυασμός αποδεικνύεται κρίσιμος για την παραγωγή ακριβών κατατμήσεων με διατηρημένες λεπτομέρειες ακμών.

4.6.4 Εφαρμογές στην Ιατρική Απεικόνιση

Το U-Net έχει καθιερωθεί ως το κυρίαρχο εργαλείο για την κατάτμηση ιατρικών εικόνων, με εφαρμογές που εκτείνονται σε όλες τις μείζονες απεικονιστικές τεχνικές. Στον τομέα της ογκολογίας, προηγμένες παραλλαγές όπως το MCU-Net-GD έχουν επιτύχει Dice scores έως 0.95 για



Σχήμα 4.3: Αρχιτεκτονική U-Net

την κατάτμηση ολικού όγκου από MRI εικόνες εγκεφάλου , ενώ βελτιωμένες αρχιτεκτονικές με VGG-16 επιτυγχάνουν Dice coefficient 0.9153 έναντι του 0.8612 το οποίο επιτυγχάνεται από τη κλασική αρχιτεκτονική U-Net [52] [2].

Η εφαρμογή του σε ηπατικούς όγκους από CT εικόνες [70] αποδεικνύει την ευελιξία της αρχιτεκτονικής στη διαχείριση διαφορετικών απεικονιστικών τρόπων. Παρομοίως, στην κατάτμηση παγκρεατικών εικόνων CT [76], όπου η χαμηλή αντίθεση και η πολύπλοκη ανατομία δημιουργούν σημαντικές προκλήσεις, οι παραλλαγές του U-Net έχουν καταδείξει ανώτερη απόδοση.

4.6.5 Συγκριτική Απόδοση και Περιορισμοί

Η συγκριτική ανάλυση του U-Net έναντι άλλων αρχιτεκτονικών κατάτμησης αποκαλύπτει τα ισχυρά σημεία και τους περιορισμούς της προσέγγισης. Παρά την εξαιρετική απόδοση σε ιατρικές εφαρμογές, η βασική αρχιτεκτονική αντιμετωπίζει δυσκολίες στη διαχείριση εικόνων με μεταβλητή μορφολογία όγκων, θολές οριογραμμές και παρόμοιες κατανομές έντασης, όπως αναφέρεται στις μελέτες για υπερηχογραφικές εικόνες μαστού [5].

Η ανάπτυξη εξειδικευμένων παραλλαγών για συγκεκριμένες εφαρμογές, όπως η κατάτμηση μαστικών όγκων σε υπερηχογραφικές εικόνες, καταδεικνύει τη σημασία της προσαρμογής της αρχιτεκτονικής στα ιδιαίτερα χαρακτηριστικά κάθε προβλήματος. Η εξισορρόπηση μεταξύ πολυπλοκότητας μοντέλου και ερμηνευσιμότητας παραμένει ένα ανοιχτό ερευνητικό ζήτημα που επηρεάζει την κλινική υιοθέτηση αυτών των τεχνολογιών.

4.7 Vision Transformers

4.7.1 Εισαγωγή και Θεμελιώδεις Έννοιες

Οι Vision Transformers (ViTs) αποτελούν μια επαναστατική κατηγορία αρχιτεκτονικών βαθιάς μάθησης που εισήχθησαν το 2020 από τους Dosovitskiy et al [20], προσαρμόζοντας το μοντέλο Transformer — αρχικά σχεδιασμένο για επεξεργασία φυσικής γλώσσας— σε εργασίες υπολογιστικής όρασης [20]. Η θεμελιώδης καινοτομία των ViTs έγκειται στην αντικατάσταση των παραδοσιακών συνελικτικών δομών με μηχανισμούς αυτοπροσοχής (self-attention), επιτρέποντας στο δίκτυο να μοντελοποιεί μακρινές χωρικές εξαρτήσεις και το γενικότερο πλαίσιο ήδη από τα πρώτα επίπεδα επεξεργασίας.

Σε αντίθεση με τα συνελικτικά νευρωνικά δίκτυα (CNNs), τα οποία εξάγουν τοπική πληροφορία μέσω παραθύρων σταθερού μεγέθους και χτίζουν ιεραρχικές αναπαραστάσεις σταδιακά, οι Vision Transformers επεξεργάζονται την εικόνα ως ακολουθία μη επικαλυπτόμενων εικονοστοιχειομπαλωμάτων (image patches). Κάθε patch αντιμετωπίζεται σαν μια "λέξη" σε ένα κείμενο, επιτρέποντας στο μοντέλο να υπολογίζει συσχετίσεις μεταξύ όλων των χωρικών θέσεων ταυτόχρονα μέσω του μηχανισμού αυτοπροσοχής. [20, 31].

4.7.2 Αρχιτεκτονική και Μαθηματική Διατύπωση

Η αρχιτεκτονική των Vision Transformers ακολουθεί μια σαφώς καθορισμένη διαδικασία επεξεργασίας που αποτελείται από τρία βασικά στάδια: την προετοιμασία εισόδου, την επεξεργασία μέσω κωδικοποιητών Transformer, και την τελική πρόβλεψη.

Προετοιμασία Εισόδου (Patch Embedding)

Δεδομένης μιας εικόνας εισόδου $\mathbf{x} \in \mathbb{R}^{H \times W \times C}$, όπου H και W είναι το ύψος και το πλάτος αντίστοιχα και C ο αριθμός των καναλιών, η εικόνα χωρίζεται σε N μη επικαλυπτόμενα patches μεγέθους $P \times P$, όπου $N = HW/P^2$. Κάθε patch επιπεδοποιείται (flattened) και προβάλλεται γραμμικά σε ένα διάνυσμα διάστασης D:

$$\mathbf{x}_p^i = \text{Flatten}(\mathbf{x}_{patch}^i) \cdot \mathbf{E}, \quad i = 1, 2, \ldots, N$$

όπου $\mathbf{E} \in \mathbb{R}^{(P^2 \cdot C) \times D}$ είναι ο πίνακας προβολής που μαθαίνεται κατά την εκπαίδευση. Στο αρχικό άρθρο του ViT, χρησιμοποιήθηκαν patches μεγέθους 16×16 pixels, γεγονός που οδήγησε στην ονομασία "An Image is Worth 16x16 Words" [20].

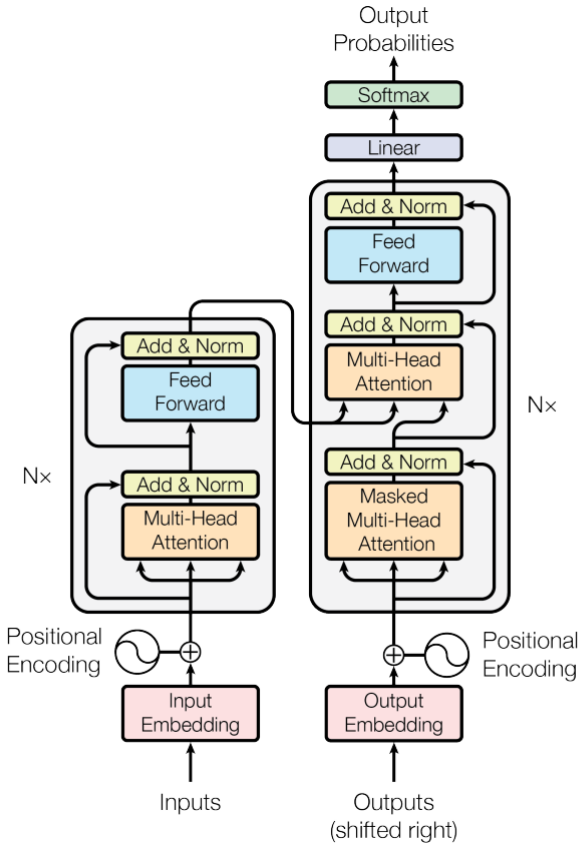
Για τη διατήρηση της χωρικής πληροφορίας, προστίθενται διανύσματα θέσης (positional embeddings) $\mathbf{E}_{pos} \in \mathbb{R}^{N \times D}$ στα patch embeddings. Επιπρόσθετα, ένα ειδικό διανύσμα κλάσης (class token) $\mathbf{x}_{cls}$ προστίθεται στην αρχή της ακολουθίας, το οποίο θα χρησιμοποιηθεί για την τελική πρόβλε-

ψη:

$$\mathbf{z}_0 = [\mathbf{x}_{cls}; \mathbf{x}_p^1; \mathbf{x}_p^2; \ldots; \mathbf{x}_p^N] + \mathbf{E}_{pos}$$

Transformer Encoder

Η ακολουθία των embeddings $\mathbf{z}_0$ τροφοδοτείται σε έναν τυπικό κωδικοποιητή Transformer που αποτελείται από L όμοια επίπεδα. Κάθε επίπεδο περιλαμβάνει δύο κύριες ενότητες: έναν μηχανισμό Multi-Head Self-Attention (MSA) και ένα πλήρως συνδεδεμένο δίκτυο τροφοδότησης προς τα εμπρός (Feed-Forward Network, FFN), με υπολειμματικές συνδέσεις και κανονικοποίηση επιπέδου (Layer Normalization) μεταξύ τους [69].



Σχήμα 4.4: Αρχιτεκτονική Vision Transformer

4.7.3 Επίδραση και Μελλοντικές Κατευθύνσεις

Οι Vision Transformers έχουν επιφέρει σημαντική αλλαγή στην έρευνα της υπολογιστικής όρασης, αμφισβητώντας τη δεσπόζουσα θέση των συνελικτικών δικτύων και ανοίγοντας νέες κατευθύνσεις για την ενοποίηση αρχιτεκτονικών μεταξύ διαφορετικών τομέων (όραση, γλώσσα, ήχος) [31]. Η δυνατότητα προεκπαίδευσης σε τεράστια σύνολα δεδομένων και μεταφοράς

μάθησης σε εξειδικευμένες εργασίες τους καθιστά ιδιαίτερα ελκυστικούς για πρακτικές εφαρμογές.

Μελλοντικές ερευνητικές κατευθύνσεις περιλαμβάνουν τη βελτίωση της αποδοτικότητάς τους για εφαρμογές πραγματικού χρόνου, την ανάπτυξη καλύτερων στρατηγικών προεκπαίδευσης με λιγότερα δεδομένα, και την ενσωμάτωση πρόσθετων επαγωγικών προκαταλήψεων που μπορούν να επιταχύνουν τη μάθηση διατηρώντας παράλληλα την ευελιξία και την ισχύ μοντελοποίησης των Transformers.

Κεφάλαιο 5ο: Μεθοδολογία και Πειραματική Διαδικασία

Το παρόν κεφάλαιο παρουσιάζει την περιεκτική μεθοδολογία που αναπτύχθηκε για τη συγκριτική αξιολόγηση των τριών αρχιτεκτονικών βαθιάς μάθησης: SegNet, U-Net και DeepLabV3+. Η πειραματική προσέγγιση σχεδιάστηκε με στόχο τη δίκαιη και αντικειμενική σύγκριση των μοντέλων υπό ελεγχόμενες συνθήκες, διασφαλίζοντας ταυτόχρονα την αναπαραγωγιμότητα των αποτελεσμάτων. Η μεθοδολογία ακολουθεί αυστηρά επιστημονικά πρωτόκολλα για την αξιολόγηση συστημάτων μηχανικής όρασης, λαμβάνοντας υπόψη τόσο την ποσοτική απόδοση όσο και τις υπολογιστικές απαιτήσεις κάθε αρχιτεκτονικής.

5.1 Επισκόπηση Πειραματικής Μεθοδολογίας

Το πειραματικό σχέδιο επικεντρώνεται στην αξιολόγηση των μοντέλων σε πραγματικές συνθήκες αστικών σκηνών μέσω του συνόλου δεδομένων Cityscapes, το οποίο θεωρείται ως ένα από τα πλέον αντιπροσωπευτικά benchmarks για σημασιολογική κατάτμηση. Η επιλογή του συγκεκριμένου συνόλου δεδομένων βασίστηκε στην ευρεία αποδοχή του από την επιστημονική κοινότητα και στην πλούσια ποικιλία σκηνών που περιλαμβάνει, επιτρέποντας την αξιολόγηση των μοντέλων σε διαφορετικές συνθήκες φωτισμού, καιρικές συνθήκες και αστικά περιβάλλοντα.

5.2 Σύνολο Δεδομένων Cityscapes

Το σύνολο δεδομένων Cityscapes αποτελεί έναν από τους πλέον διαδεδομένους και αυστηρούς δείκτες αξιολόγησης για αλγορίθμους σημασιολογικής κατάτμησης σε αστικές σκηνές. Το σύνολο δεδομένων, περιλαμβάνει σημασιολογικές ετικέτες για κάθε εικονοστοιχείο, 30 κλάσεις, 5.000 εικόνες με υψηλής ποιότητας ετικέτες, 20.000 εικόνες με μικρότερης ποιότητας ετικέτες από 50 διαφορετικές πόλεις. Η ποικιλομορφία των γεωγραφικών θέσεων διασφαλίζει την αντιπροσωπευτικότητα του συνόλου δεδομένων και την ικανότητα γενίκευσης των εκπαιδευμένων μοντέλων. Για εργασίες κατάτμησης, το Cityscapes παρέχει ετικέτες σε επίπεδο εικονοστοιχείου για 5.000 εικόνες με ανάλυση 1024×2048 pixel, προ-διαιρεμένες σε σύνολα εκπαίδευσης (2.975), επικύρωσης (500) και ελέγχου (1.525). Αυτή η τυποποιημένη διαίρεση επιτρέπει τη δίκαιη σύγκριση με προηγούμενες μελέτες και διασφαλίζει την αξιοπιστία των αποτελεσμάτων αξιολόγησης.

5.3 Κατηγορίες και Κλάσεις

Το σύνολο δεδομένων παρέχει σημασιολογικές ετικέτες για 30 κλάσεις ομαδοποιημένες σε 8 κατηγορίες (επίπεδες επιφάνειες, άνθρωποι, οχήματα, κατασκευές, αντικείμενα, φύση, ουρανός και κενό). Αυτή η ταξινόμηση καλύπτει το σύνολο των στοιχείων που συναντώνται σε τυπικές

αστικές σκηνές, από στοιχεία υποδομής όπως δρόμους και κτίρια έως δυναμικά αντικείμενα όπως οχήματα και πεζούς. Οι κύριες κατηγορίες περιλαμβάνουν:

- **Επίπεδες επιφάνειες:** Δρόμος, πεζοδρόμιο, σιδηροδρομικές γραμμές
- **Άνθρωποι:** Πεζοί, ποδηλάτες
- **Οχήματα:** Αυτοκίνητα, φορτηγά, λεωφορεία, μοτοσικλέτες
- **Κατασκευές:** Κτίρια, τοίχοι, φράχτες, γέφυρες
- **Αντικείμενα:** Πινακίδες κυκλοφορίας, φωτεινοί σηματοδότες, πόλοι
- **Φύση:** Βλάστηση, έδαφος, ουρανός
- **Κενό:** Περιοχές χωρίς σημασιολογική σημασία

5.4 Προεπεξεργασία δεδομένων

Η προεπεξεργασία των εικόνων αποτελεί κρίσιμο στάδιο για τη διασφάλιση της ποιότητας και συνέπειας των δεδομένων εισόδου. Λόγω της υψηλής ανάλυσης των αρχικών εικόνων (1024 × 2048), εφαρμόστηκε τεχνική κλιμάκωσης για τη μείωση της υπολογιστικής πολυπλοκότητας χωρίς σημαντική απώλεια πληροφορίας. Η εικόνα διαιρείται σε μικρότερα τμήματα με χωρική ανάλυση (256, 256), επειδή κάθε εικόνα έχει χωρική ανάλυση (1024, 2048).

Η διαδικασία προεπεξεργασίας περιλαμβάνει:

- **Κανονικοποίηση Pixel:** Οι τιμές των εικονοστοιχείων κανονικοποιούνται στο διάστημα [0, 1] διαιρώντας κατά 255, διασφαλίζοντας σταθερή σύγκλιση κατά την εκπαίδευση. Η διαίρεση κατά 255 είναι απαραίτητη διότι οι εικόνες RGB αποθηκεύονται συνήθως σε 8-bit μορφή, όπου κάθε κανάλι χρώματος (κόκκινο, πράσινο, μπλε) μπορεί να πάρει τιμές από 0 έως 255. Η κανονικοποίηση σε τιμές [0, 1] βελτιώνει την ταχύτητα σύγκλισης των αλγορίθμων βελτιστοποίησης.

  Επιπρόσθετα, εφαρμόζεται τυποποίηση σύμφωνα με τις στατιστικές παραμέτρους του ImageNet συνόλου δεδομένων, όπου σε κάθε κανάλι αφαιρείται ο μέσος όρος και διαιρείται με την τυπική απόκλιση για τα κανάλια RGB αντίστοιχα. Αυτή η προσέγγιση είναι ιδιαίτερα σημαντική όταν χρησιμοποιούνται προεκπαιδευμένα μοντέλα (transfer learning) που έχουν εκπαιδευτεί στο ImageNet, καθώς διασφαλίζει ότι η κατανομή των δεδομένων εισόδου παραμένει συνεπής με αυτή που αναμένει το προεκπαιδευμένο δίκτυο.

- **Διαχείριση Ετικετών:** Οι σημασιολογικές ετικέτες μετατρέπονται σε μορφή one-hot encoding για συμβατότητα με τις συναρτήσεις απώλειας των νευρωνικών δικτύων. Το

one-hot encoding αποτελεί μια τεχνική κωδικοποίησης που μετατρέπει κατηγοριακές μεταβλητές σε δυαδική μορφή, δημιουργώντας νέες στήλες για κάθε κατηγορία όπου η τιμή 1 υποδεικνύει την παρουσία της συγκεκριμένης κατηγορίας και η τιμή 0 την απουσία της. Για παράδειγμα, αν έχουμε τρεις κλάσεις (φόντο, αυτοκίνητο, δρόμος), η κλάση ”αυτοκίνητο” θα αναπαρασταθεί ως το διάνυσμα [0, 1, 0]. Αυτή η αναπαράσταση είναι απαραίτητη διότι τα νευρωνικά δίκτυα απαιτούν αριθμητικές εισόδους και εξόδους, ενώ παράλληλα αποτρέπει την εσφαλμένη ερμηνεία ιεραρχικών σχέσεων μεταξύ των κατηγοριών που θα μπορούσε να προκύψει από την απλή αρίθμηση των κλάσεων.

5.5 Πλατφόρμα υλοποίησης Και Τεχνικές Προδιαγραφές

Η υλοποίηση και εκτέλεση των πειραμάτων πραγματοποιήθηκε στην πλατφόρμα Kaggle, η οποία προσφέρει ισχυρούς υπολογιστικούς πόρους και προκαθορισμένα περιβάλλοντα ανάπτυξης βελτιστοποιημένα για εφαρμογές βαθιάς μάθησης. Η επιλογή της πλατφόρμας Kaggle βασίστηκε σε αρκετούς παράγοντες που συμβάλλουν στην αποδοτικότητα και αναπαραγωγιμότητα των πειραμάτων.

Η πλατφόρμα παρέχει πρόσβαση σε σύγχρονες GPU (Graphics Processing Units) με επαρκή μνήμη για την εκπαίδευση μοντέλων βαθιάς μάθησης υψηλής πολυπλοκότητας. Οι διαθέσιμες GPU περιλαμβάνουν NVIDIA Tesla P100 και T4 GPUs με 16GB και 15GB μνήμης αντίστοιχα. Τα Kaggle kernels προσφέρουν προεγκατεστημένες βιβλιοθήκες μηχανικής μάθησης και εργαλεία ανάλυσης δεδομένων, εξαλείφοντας προβλήματα συμβατότητας και επιταχύνοντας την ανάπτυξη λογισμικού. Η πλατφόρμα επίσης, επιτρέπει την άμεση πρόσβαση σε δημοφιλή σύνολα δεδομένων μέσω του ενσωματωμένου συστήματος διαχείρισης δεδομένων, απλοποιώντας τη διαδικασία φόρτωσης και εκπαίδευσης.

5.6 Framework Pytorch

Για την υλοποίηση των τριών αρχιτεκτονικών επιλέχθηκε το framework `PyTorch` λόγω της ευελιξίας, της ευκολίας χρήσης και της ευρείας υποστήριξης από την ερευνητική κοινότητα. Το PyTorch παρέχει δυναμικά υπολογιστικά γραφήματα που επιτρέπουν ευέλικτη αρχιτεκτονική σχεδίαση και εκαπίδευση.

- `torch` και `torchvision` για τα νευρωνικά δίκτυα και προεπεξεργασία εικόνων.
- `numpy` για αριθμητικούς υπολογισμούς.
- `matplotlib` για οπτικοποίηση αποτελεσμάτων.
- `tqdm` για την παρακολούθηση χρονοβόρων διαδικασιών όπως της εκπαίδευσης.

5.7 Μετρικές Αξιολόγησης για Σημασιολογική Τμηματοποίηση

Η αξιολόγηση της απόδοσης των αλγορίθμων σημασιολογικής τμηματοποίησης αποτελεί κρίσιμο στοιχείο για την αντικειμενική σύγκριση και επιλογή της καταλληλότερης μεθόδου. Οι μετρικές αξιολόγησης παρέχουν ποσοτικά μέτρα που επιτρέπουν την εκτίμηση διαφόρων πτυχών της απόδοσης των μοντέλων, όπως η ακρίβεια κατάτμησης, η ποιότητα ορίων και η γενική αξιοπιστία. Η παρούσα ενότητα αναλύει τις κυριότερες μετρικές αξιολόγησης που χρησιμοποιούνται από την επιστημονική κοινότητα.

5.7.1 Intersection over Union (IoU)

Η μετρική Intersection over Union (IoU), γνωστή και ως Jaccard Index, αποτελεί μια κρίσιμη μέτρηση στην αξιολόγηση της απόδοσης αλγορίθμων τμηματοποίησης και ανίχνευσης αντικειμένων. Ουσιαστικά, το IoU εκφράζει το λόγο της επιφάνειας της τομής μεταξύ της προβλεπόμενης περιοχής και της πραγματικής (ground truth) προς την επιφάνεια της ένωσης αυτών των δύο περιοχών [45].

$$\text{IoU} = \frac{|A \cap B|}{|A \cup B|} = \frac{\text{True Positives}}{\text{True Positives + False Positives + False Negatives}} \tag{4}$$

Η τιμή του IoU κυμαίνεται από 0 έως 1, όπου το 1 συμβολίζει τέλεια επικάλυψη και το 0 καμία επικάλυψη. Αυτή η μετρική επιτρέπει να εκτιμηθεί πόσο καλά η προβλεπόμενη περιοχή συμφωνεί με την αληθινή περιοχή, λαμβάνοντας υπόψη τόσο την ακριβή τοποθεσία όσο και το μέγεθος των περιοχών. Η χρήση της IoU είναι διαδεδομένη σε πεδία όπως η ιατρική απεικόνιση και η μηχανική όραση γενικότερα, καθώς επιτρέπει αξιόπιστη και ποσοτικοποιημένη αξιολόγηση συγκριτικά με άλλες απλούστερες μετρικές που δεν λαμβάνουν υπόψη τον ακριβή γεωμετρικό συνδυασμό μεταξύ προβλεπόμενων και πραγματικών περιοχών.

Η μέση IoU (mIoU) για προβλήματα με πολλές κλάσεις υπολογίζεται ως:

$$\text{mIoU} = \frac{1}{N} \sum_{i=1}^{N} \text{IoU}_i \tag{5}$$

5.7.2 Dice Coefficient

Ο δείκτης Dice Coefficient και η μετρική Intersection over Union (IoU) είναι δύο στενά συναφείς μετρικές που αξιολογούν την επικάλυψη μεταξύ της προβλεπόμενης και της πραγματικής περιοχής σε εργασίες τμηματοποίησης εικόνων. Ο Dice Coefficient υπολογίζεται ως δύο φορές το μέγεθος της τομής διαιρεμένο με το άθροισμα των μεγεθών των δύο περιοχών [45].

$$\text{Dice} = \frac{2 \times \text{IoU}}{1 + \text{IoU}} \tag{6}$$

Αυτό σημαίνει ότι ο Dice δίνει μεγαλύτερο βάρος στην περιοχή της τομής, οδηγώντας συχνά σε υψηλότερες τιμές σε σχέση με την IoU για παρόμοια επίπεδα επικάλυψης. Μια σημαντική λειτουργική διαφορά είναι ότι ο Dice Coefficient είναι διαφορικός και χρησιμοποιείται ευρέως ως συνάρτηση απώλειας (Dice loss) κατά την εκπαίδευση νευρωνικών δικτύων, ενώ η IoU, αν και μπορεί να χρησιμοποιηθεί ως αξιολογητική μετρική, δεν είναι εύκολα διαφορική. Μαθηματικά, οι τιμές τους συνδέονται μέσω απλών σχέσεων, και συνεπώς η γνώση της μίας μπορεί να οδηγήσει στην εύρεση της άλλης. Εν γένει, η IoU θεωρείται πιο αυστηρή μέτρηση της επικάλυψης και προτιμάται σε εφαρμογές όπου η ακριβής ανάλυση των ορίων είναι κρίσιμη, ενώ ο Dice είναι πιο ευαίσθητος στις μικρές αποκλίσεις και κατάλληλος για περιβάλλοντα όπου η σταθερότητα κατά την εκπαίδευση είναι σημαντική, όπως η ιατρική τμηματοποίηση.

5.7.3 Pixel Accuracy

Η Pixel Accuracy (PA) αποτελεί μια βασική μετρική απόδοσης στον τομέα της τμηματοποίησης εικόνων, η οποία εκφράζει το ποσοστό των εικονοστοιχείων που ταξινομούνται σωστά ως προς το σύνολο των εικονοστοιχείων σε μια εικόνα. Συγκεκριμένα, η PA υπολογίζεται ως ο λόγος του συνολικού αριθμού των σωστών προβλέψεων (True Positives και True Negatives) προς το συνολικό αριθμό εικονοστοιχείων της εικόνας.

$$\text{Pixel Accuracy} = \frac{\text{Correctly Classified Pixels}}{\text{Total Pixels}} \tag{7}$$

Η μετρική αυτή είναι εύκολη στην κατανόηση και στην εφαρμογή, παρέχοντας μια γρήγορη εκτίμηση της συνολικής απόδοσης του αλγορίθμου τμηματοποίησης. Ωστόσο, είναι σημαντικό να σημειωθεί ότι η Pixel Accuracy μπορεί να είναι παραπλανητική σε περιπτώσεις όπου υπάρχει έντονη ασυμμετρία στις κλάσεις, όπως συμβαίνει στην ιατρική απεικόνιση, όπου συνήθως η περιοχή ενδιαφέροντος καταλαμβάνει μικρό μόνο ποσοστό της εικόνας. Σε τέτοιες περιπτώσεις, η PA μπορεί να δίνει υπερβολικά υψηλές τιμές ακόμα και με πολύ απλές, μη ακριβείς προβλέψεις, λόγω της μεγάλης αναλογίας των εικονοστοιχείων χωρίς του παρασκηνίου [54].

5.7.4 Hausdorff Distance

Η μετρική απόδοσης Hausdorff Distance αποτελεί ένα σημαντικό μέτρο για την εκτίμηση της ομοιότητας μεταξύ δύο συνόλων σημείων, με ιδιαίτερη χρήση στην ιατρική απεικόνιση για την αξιολόγηση της ποιότητας των αλγορίθμων τμηματοποίησης. Συγκεκριμένα, μετρά την μέγιστη απόσταση μεταξύ των σημείων στα όρια των δύο συνόλων, αξιολογώντας πόσο ”μακριά” βρίσκεται το ένα σύνολο από το άλλο [66].

Εντούτοις, έχει αναγνωριστεί ότι η κλασική μορφή της Hausdorff Distance μπορεί να παρουσιάζει σφάλματα κατάταξης, τα οποία υπονομεύουν την αξιολόγηση της ποιότητας των τμηματοποιήσεων [49]. Για να αντιμετωπιστούν αυτά τα σφάλματα, έχουν προταθεί τροποποιήσεις όπως η balanced average Hausdorff Distance, που βελτιώνει την αξιοπιστία της κατάταξης και την ακρίβεια της μέτρησης, ειδικά σε πολύπλοκα δεδομένα ιατρικής απεικόνισης. Αυτή η τροποποιημένη μορφή διατηρεί σταθερό τον παρονομαστή στον υπολογισμό της απόστασης, αποφεύγοντας την επίδραση των μεταβολών στο μέγεθος των τμηματοποιήσεων, καθιστώντας τη μέτρηση πιο συνεπή και αξιόπιστη.

5.7.5 Mean Absolute Error

Η μετρική Mean Absolute Error (MAE) αποτελεί έναν από τους βασικούς δείκτες αξιολόγησης της απόδοσης μοντέλων πρόβλεψης και παλινδρόμησης (regression), καθώς υπολογίζει τον μέσο απόλυτο όρο της απόστασης μεταξύ των προβλεπόμενων τιμών και των πραγματικών παρατηρήσεων. Με άλλα λόγια, η MAE μετρά το μέσο μέγεθος του σφάλματος, δίνοντας μια σαφή και άμεσα ερμηνεύσιμη ένδειξη της ακρίβειας ενός μοντέλου χωρίς να δίνει μεγαλύτερο βάρος σε μεγάλες αποκλίσεις, όπως συμβαίνει με άλλες μετρικές π.χ. RMSE [34]. Η απλότητα και η ευκολία ερμηνείας της την καθιστούν ιδιαίτερα χρήσιμη σε πολλούς τομείς εφαρμογής, όπως η πρόβλεψη χρονικών σειρών και η μηχανική μάθηση. Επιπρόσθετα, έχουν προταθεί παραλλαγές όπως η σχετική MAE (relative MAE), που επιτρέπουν τη σύγκριση μεταξύ διαφορετικών μοντέλων με βάση την απόδοση [53], προσφέροντας σταθερότητα και αξιόπιστα αποτελέσματα στην αξιολόγηση πολλαπλών μοντέλων.

5.7.6 F1 Score

Το F1 Score αποτελεί μια από τις πιο σημαντικές μετρικές αξιολόγησης στα προβλήματα σημασιολογικής κατάτμησης, καθώς συνδυάζει την ακρίβεια (precision) και την ανάκληση (recall) σε μία ενοποιημένη μετρική που παρέχει ισορροπημένη εκτίμηση της απόδοσης του μοντέλου [51]. Η σημασία του F1 Score είναι ιδιαίτερα σημαντική σε περιπτώσεις ανισορροπημένων συνόλων δεδομένων, όπου οι παραδοσιακές μετρικές ακρίβειας μπορεί να παραπλανούν [57].

Το F1 Score ορίζεται ως ο αρμονικός μέσος όρος της ακρίβειας και της ανάκλησης:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{8}$$

όπου: - Precision = $\frac{\text{TP}}{\text{TP+FP}}$ - Recall = $\frac{\text{TP}}{\text{TP+FN}}$

Αντικαθιστώντας τις εκφράσεις της ακρίβειας και της ανάκλησης στην εξίσωση του F1 Score, προκύπτει:

$$F1 = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FP} + \text{FN}} \tag{9}$$

Η επιλογή του αρμονικού μέσου όρου αντί του αριθμητικού οφείλεται στο γεγονός ότι ο αρμονικός μέσος όρος επιβάλλει μεγαλύτερη ποινή όταν οι τιμές της ακρίβειας και της ανάκλησης διαφέρουν σημαντικά μεταξύ τους [58]. Αυτό καθιστά το F1 Score ιδιαίτερα χρήσιμο για την αξιολόγηση μοντέλων όπου είναι απαραίτητη η ισορροπία μεταξύ των δύο αυτών μετρικών.

Στο πλαίσιο των προβλημάτων πολλαπλών κλάσεων (multi-class), το F1 Score υπολογίζεται για κάθε κλάση ξεχωριστά και στη συνέχεια συνδυάζεται με διάφορες τεχνικές υπολογισμού μέσου όρου. Για τον υπολογισμό του συνολικού F1 Score χρησιμοποιείται η μέθοδος macro-averaging:

$$\text{Macro-F1} = \frac{1}{C} \sum_{i=1}^{C} F1_i \tag{10}$$

όπου C είναι ο αριθμός των κλάσεων και $F1_i$ το F1 Score της κλάσης i.

Το F1 Score κυμαίνεται στο εύρος [0, 1], όπου η τιμή 1 υποδηλώνει τέλεια απόδοση (100% ακρίβεια και ανάκληση), ενώ η τιμή 0 υποδηλώνει την χειρότερη δυνατή απόδοση. Γενικώς, τιμές F1 Score μεγαλύτερες του 0.8 θεωρούνται ικανοποιητικές για τα περισσότερα προβλήματα σημασιολογικής κατάτμησης, ενώ τιμές άνω του 0.9 χαρακτηρίζονται ως εξαιρετικές.

Η χρήση του F1 Score στη σημασιολογική κατάτμηση παρουσιάζει ιδιαίτερες προκλήσεις λόγω της φύσης των δεδομένων. Σε αντίθεση με τα προβλήματα ταξινόμησης, όπου κάθε δείγμα κατηγοριοποιείται σε μία μόνο κλάση, στη σημασιολογική κατάτμηση κάθε εικονοστοιχείο αντιστοιχίζεται σε μία κλάση. Αυτό σημαίνει ότι το F1 Score υπολογίζεται σε επίπεδο εικονοστοιχείου, γεγονός που μπορεί να οδηγήσει σε υπερεκτίμηση της απόδοσης σε περιπτώσεις όπου υπάρχουν μεγάλες περιοχές με σωστή ταξινόμηση και μικρές περιοχές με λανθασμένη ταξινόμηση.

Στην παρούσα εργασία, ο υπολογισμός του F1 Score, πραγματοποιείται αξιοποιώντας την κλάση MulticlassF1Score του πακέτου torchmetrics. Η διαδικασία περιλαμβάνει τον υπολογισμό της ακρίβειας και της ανάκλησης για κάθε κλάση και στη συνέχεια την εφαρμογή του τύπου του αρμονικού μέσου όρου

5.8 Επιλογή Μετρικών για την Παρούσα Μελέτη

Για την αξιολόγηση των αρχιτεκτονικών SegNet, U-Net και DeepLabV3+, επιλέχθηκαν οι ακόλουθες μετρικές βάσει διεθνών βέλτιστων πρακτικών:

1. **Mean Intersection over Union (mIoU)**: Η πλέον τυποποιημένη μετρική στη βιβλιογραφία σημασιολογικής τμηματοποίησης.
2. **Dice Coefficient**: Λόγω της αποτελεσματικότητάς του σε εργασίες σημασιολογικής τμηματοποίησης και της ευρείας χρήσης του απο την ακαδημαϊκή κοινότητα.
3. **F1 score**: Λόγω της σημαντικότητας του σε προβλήματα σημασιολογικής τμηματοποίησης.

5.8.1 Κριτήριο Επιλογής

Η επιλογή βασίστηκε σε τέσσερα βασικά κριτήρια. Πρώτον, την τυποποίηση, ώστε να διασφαλίζεται η συμβατότητα με την υπάρχουσα βιβλιογραφία και η συγκρισιμότητα των αποτελεσμάτων. Δεύτερον, τη συμπληρωματικότητα, καθώς κάθε μετρική εξετάζει διαφορετικές πτυχές της απόδοσης. Τρίτον, την αξιοπιστία, η οποία έχει αποδειχθεί σε πλήθος συγκριτικών μελετών. Τέλος, την υλοποιησιμότητα, δεδομένης της διαθεσιμότητας των μετρικών σε τυποποιημένες βιβλιοθήκες όπως της PyTorch διασφαλίζοντας την αξιοπιστία των μετρικών.

5.8.2 Μαθηματικοί Ορισμοί Επιλεγμένων Μετρικών

Για την παρούσα μελέτη, οι επιλεγμένες μετρικές ορίζονται ως εξής:

Για C κλάσεις με confusion matrix M:

$$\text{IoU}i = \frac{Mii}{\sum_{j=1}^{C} M_{ij} + \sum_{j=1}^{C} M_{ji} - M_{ii}} \tag{11}$$

$$\text{mIoU} = \frac{1}{C} \sum_{i=1}^{C} \text{IoU}_i \tag{12}$$

Υπολογίζεται από την αρμονική μέση των class-wise Precision και Recall:

$$\text{F1}_i = 2 \times \frac{\text{Precision}_i \times \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i} \tag{13}$$

5.8.3 Αιτιολόγηση Μη-Επιλεγμένων Μετρικών

Παρόλο που μετρικές όπως η Hausdorff Distance και το Mean Absolute Error παρέχουν πολύτιμες πληροφορίες για την ακρίβεια ορίων, δεν επιλέχθηκαν για την παρούσα μελέτη για τρεις βασικούς λόγους. Αρχικά, παρουσιάζουν σημαντική υπολογιστική πολυπλοκότητα, καθώς απαιτούν εκτεταμένους υπολογιστικούς πόρους, ειδικά όταν εφαρμόζονται σε μεγάλα σύνολα δεδομένων. Επιπλέον, η υλοποίησή τους εμφανίζει σημαντική μεταβλητότητα, γεγονός που

μπορεί να επηρεάσει την αξιοπιστία και τη συνέπεια των αποτελεσμάτων. Τέλος, η κατεύθυνση της παρούσας εργασίας εστιάζει κυρίως στη γενική απόδοση των μοντέλων και όχι αποκλειστικά στην ακρίβεια των ορίων, καθιστώντας έτσι την επιλογή άλλων μετρικών περισσότερο καταλληλότερη.

Κεφάλαιο 6ο: Πειραματική Αξιολόγηση και Ανάλυση Αποτελεσμάτων

Στο παρόν κεφάλαιο παρουσιάζεται η λεπτομερής πειραματική αξιολόγηση των τριών αρχιτεκτονικών βαθιάς μάθησης που μελετήθηκαν: U-Net, SegNet και DeepLabV3+. Η ανάλυση επικεντρώνεται στη συγκριτική απόδοση των μοντέλων τόσο από ποσοτική όσο και από ποιοτική άποψη, εξετάζοντας παράλληλα τους υπολογιστικούς περιορισμούς και τη γενικευτική τους ικανότητα. Δεδομένου ότι οι θεωρητικές βάσεις των αρχιτεκτονικών και των μετρικών αξιολόγησης έχουν αναλυθεί εκτενώς στα προηγούμενα κεφάλαια, η παρούσα ενότητα εστιάζει αποκλειστικά στην παρουσίαση και ερμηνεία των πειραματικών δεδομένων.

Η πειραματική διαδικασία σχεδιάστηκε με γνώμονα την επίτευξη αντικειμενικών και αναπαραγώγιμων αποτελεσμάτων. Όλα τα μοντέλα εκπαιδεύτηκαν υπό τις ίδιες συνθήκες, χρησιμοποιώντας κοινές υπερπαραμέτρους και την ίδια υπολογιστική πλατφόρμα. Αυτή η προσέγγιση διασφαλίζει τη δίκαιη σύγκριση των αρχιτεκτονικών και την αξιοπιστία των συμπερασμάτων που προκύπτουν από την ανάλυση.

6.1 Πειραματικός Σχεδιασμός και Μεθοδολογία

6.1.1 Πλατφόρμα Υλοποίησης

Η υλοποίηση και εκτέλεση των πειραμάτων πραγματοποιήθηκε στην πλατφόρμα Kaggle Notebooks, η οποία παρέχει πρόσβαση σε υπολογιστικούς πόρους υψηλών προδιαγραφών. Συγκεκριμένα, χρησιμοποιήθηκαν GPU NVIDIA Tesla P100 με 16GB μνήμης σε συνδυασμό με τον επεξεργσατήIntel(R) Xeon(R) CPU @ 2.00GHz και 32GB RAM συστήματος. Η επιλογή αυτής της πλατφόρμας εξασφαλίζει την αναπαραγωγιμότητα των αποτελεσμάτων και τη δυνατότητα επαλήθευσης από άλλους ερευνητές.

Το λογισμικό περιβάλλον βασίστηκε στο PyTorch 2.6.0 με CUDA 12.4 για την επιτάχυνση των υπολογισμών μέσω GPU. Επιπλέον, χρησιμοποιήθηκε η βιβλιοθήκη segmentation-models-pytorch [35] για την υλοποίηση των προεκπαιδευμένων μοντέλων. Η επιλογή αυτών των εργαλείων στηρίζεται στην ευρεία αποδοχή τους από την επιστημονική κοινότητα και στη σταθερότητά τους.

6.1.2 Παραμετροποίηση Εκπαίδευσης

Για τη διασφάλιση της δίκαιης σύγκρισης, όλα τα μοντέλα εκπαιδεύτηκαν με ίδιες υπερπαραμέτρους. Το μέγεθος του batch ορίστηκε στα 32 δείγματα, ισορροπώντας μεταξύ της αποδοτικής εκμετάλλευσης της μνήμης GPU και της σταθερότητας της εκπαίδευσης. Ο αριθμός των εποχών περιορίστηκε σε 15, επιλογή που βασίστηκε σε προκαταρκτικά πειράματα που έδειξαν ότι τα μοντέλα επιτυγχάνουν σύγκλιση εντός αυτού του χρονικού πλαισίου.

Ο ρυθμός μάθησης (learning rate) ορίστηκε σε 3×10^{-4}, τιμή που προκύπτει από βελτιστοποίηση βασισμένη σε grid search για παρόμοια προβλήματα στη βιβλιογραφία. Ως βελτιστοποιητής επιλέχθηκε ο βελτιστοποιητής Adam, ο οποίος συνδυάζει τα πλεονεκτήματα της προσαρμοστικής ρύθμισης του ρυθμού μάθησης με την αποτελεσματική εκμετάλλευση της ορμής. Επιπλέον, εφαρμόστηκε mixed precision training (FP16) για τη βελτίωση της υπολογιστικής απόδοσης χωρίς σημαντική επίδραση στην ακρίβεια των αποτελεσμάτων.

6.1.3 Προεπεξεργασία Δεδομένων

Η προεπεξεργασία των εικόνων ακολούθησε τις βέλτιστες πρακτικές για τη σημασιολογική κατάτμηση. Οι εικόνες κλιμακώθηκαν από την αρχική ανάλυση 2048×1024 εικονοστοιχείων, σε 256×256 εικονοστοιχεία, μείωση που αποσκοπεί στη βελτίωση της υπολογιστικής αποδοτικότητας καθώς και στην εξοικονόμηση υπολογιστικών πόρων, διατηρώντας παράλληλα την ουσιαστική πληροφορία των σκηνών. Η κανονικοποίηση των εικονοστοιχείων πραγματοποιήθηκε χρησιμοποιώντας τη μέση τιμή και τυπική απόκλιση που χρησιμοποιήθηκε από το σύνολο δεδομένων ImageNet. Πιο συγκεκριμένα, η μέση τιμή ανά κανάλι: [0.485, 0.456, 0.406] ενώ η τυπική απόκλιση: [0.229, 0.224, 0.225], αυτή η προσέγγιση συνάδει με τη χρήση προεκπαιδευμένων μοντέλων στο σύνολο δεδομένων ImageNet.

Δεν χρησιμοποιήθηκε καθόλου επαύξηση δεδομένων, όπως γεωμετρικούς μετασχηματισμούς και Gausian Blur, τεχνικές που θα μπορούσαν να αλλοιώσουν τη σημασιολογική πληροφορία των αστικών σκηνών. Αυτή η προσέγγιση εξασφαλίζει ότι τα αποτελέσματα αντικατοπτρίζουν την πραγματική απόδοση των μοντέλων και όχι την αποτελεσματικότητα των τεχνικών επαύξησης.

6.2 Αποτελέσματα Απόδοσης ανά Αρχιτεκτονική

6.2.1 Ανάλυση Απόδοσης U-Net

Η ανάλυση των αποτελεσμάτων του πίνακα 6.1 αποκαλύπτει σημαντικές πληροφορίες για την απόδοση του U-Net μοντέλου κατά τη διάρκεια της εκπαίδευσης. Το μοντέλο παρουσιάζει σταθερή βελτίωση σε όλες τις μετρικές εκπαίδευσης, με τον συντελεστή Dice να αυξάνεται από 0.6279 στην 1η εποχή σε 0.8702 στην 15η εποχή, υποδηλώνοντας αποτελεσματική μάθηση των χαρακτηριστικών του dataset.

Συμπεριφορά Μετρικών Εκπαίδευσης: Οι μετρικές εκπαίδευσης εμφανίζουν μονοτονική βελτίωση καθ' όλη τη διάρκεια των 15 εποχών. Συγκεκριμένα, το F1-score αυξάνεται από 0.5678 σε 0.8512, ενώ το Mean Intersection over Union (MIoU) βελτιώνεται από 0.4576 σε 0.7703. Παράλληλα, η συνάρτηση κόστους (Loss) παρουσιάζει συνεχή μείωση από 1.6008 σε 0.3587, επιβεβαιώνοντας την αποτελεσματική σύγκλιση του αλγορίθμου βελτιστοποίησης.

Πίνακας 6.1: Μετρικές Εκπαίδευσης και Επικύρωσης - U-Net

Epoch	Dice		F1		Miou		Loss	
	Training	Validation	Training	Validation	Training	Validation	Training	Validation
1	0.6279	0.7002	0.5678	0.6245	0.4576	0.5387	1.6008	0.5881
2	0.7295	0.7202	0.6485	0.6422	0.5742	0.5628	0.8092	0.4783
3	0.7432	0.7323	0.6651	0.6597	0.5914	0.5777	0.6837	0.4586
4	0.7707	0.7676	0.7104	0.7172	0.6269	0.6228	0.6219	0.4484
5	0.7938	0.7713	0.7418	0.7218	0.6581	0.6278	0.5824	0.4259
6	0.8062	0.7756	0.7554	0.7268	0.6753	0.6334	0.5386	0.44
7	0.8152	0.7854	0.7656	0.7378	0.6881	0.6467	0.5015	0.4197
8	0.8221	0.7842	0.7779	0.7429	0.6979	0.645	0.4839	0.4057
9	0.8323	0.7921	0.7965	0.7555	0.7127	0.6558	0.4643	0.4153
10	0.8401	0.7937	0.81	0.7603	0.7242	0.6579	0.4464	0.4518
11	0.8483	0.8033	0.8215	0.7734	0.7365	0.6713	0.4221	0.4246
12	0.8544	0.8015	0.8299	0.7719	0.7458	0.6688	0.404	0.4273
13	0.8589	0.8018	0.8362	0.7712	0.7527	0.6692	0.3906	0.4437
14	0.8652	0.8127	0.8446	0.7877	0.7625	0.6844	0.3729	0.428
15	0.8702	0.8089	0.8512	0.7794	0.7703	0.6792	0.3587	0.437

Απόδοση Επικύρωσης και Γενίκευση: Οι μετρικές επικύρωσης αποκαλύπτουν διαφορετική συμπεριφορά σε σύγκριση με την εκπαίδευση. Ενώ αρχικά υπερτερούν των αντίστοιχων μετρικών εκπαίδευσης (π.χ., Dice: 0.7002 έναντι 0.6279 στην 1η εποχή), σταδιακά παρουσιάζουν πλατωπέδιο μετά την 10η-11η εποχή. Το καλύτερο validation Dice επιτυγχάνεται στην 14η εποχή (0.8127), ενώ το F1-score κορυφώνεται επίσης στην 14η εποχή (0.7877).

Ενδείξεις Υπερπροσαρμογής: Η αυξανόμενη απόκλιση μεταξύ των μετρικών εκπαίδευσης και επικύρωσης υποδηλώνει την παρουσία υπερπροσαρμογής (overfitting). Στην τελευταία εποχή, το χάσμα για τον συντελεστή Dice φτάνει τις 0.0613 μονάδες (0.8702 έναντι 0.8089), ενώ για το MIoU η διαφορά είναι 0.0911 μονάδες. Αυτό το φαινόμενο είναι αναμενόμενο σε deep learning εφαρμογές και υποδηλώνει την ανάγκη για τεχνικές regularization ή early stopping.

Συνολική Αξιολόγηση Συνολικά, το U-Net μοντέλο επιδεικνύει ικανοποιητική απόδοση με το validation Dice να φτάνει το 0.8127, το οποίο θεωρείται υψηλή επίδοση για εργασίες image segmentation. Η σταθερή βελτίωση των μετρικών και η σχετικά μικρή διακύμανση στις τελευταίες εποχές επικύρωσης (0.8015-0.8127 για Dice) υποδηλώνουν σταθερότητα στην απόδοση του μοντέλου.

6.2.2 Ανάλυση Απόδοσης SegNet

Η ανάλυση των αποτελεσμάτων του πίνακα 6.2 αποκαλύπτει μεικτή απόδοση του SegNet μοντέλου, με χαμηλότερες επιδόσεις σε σύγκριση με τα U-Net και DeepLabV3+. Το μοντέλο ξεκινά από σημαντικά χαμηλότερες αρχικές τιμές με Dice coefficient 0.416 στην πρώτη εποχή, υποδηλώνοντας αργή αρχικοποίηση και περιορισμένη ικανότητα άμεσης προσαρμογής στο

Σχήμα 6.1: Αποτελέσματα Σημασιολογικής τμηματοποίησης του U-Net σε νέα δεδομένα

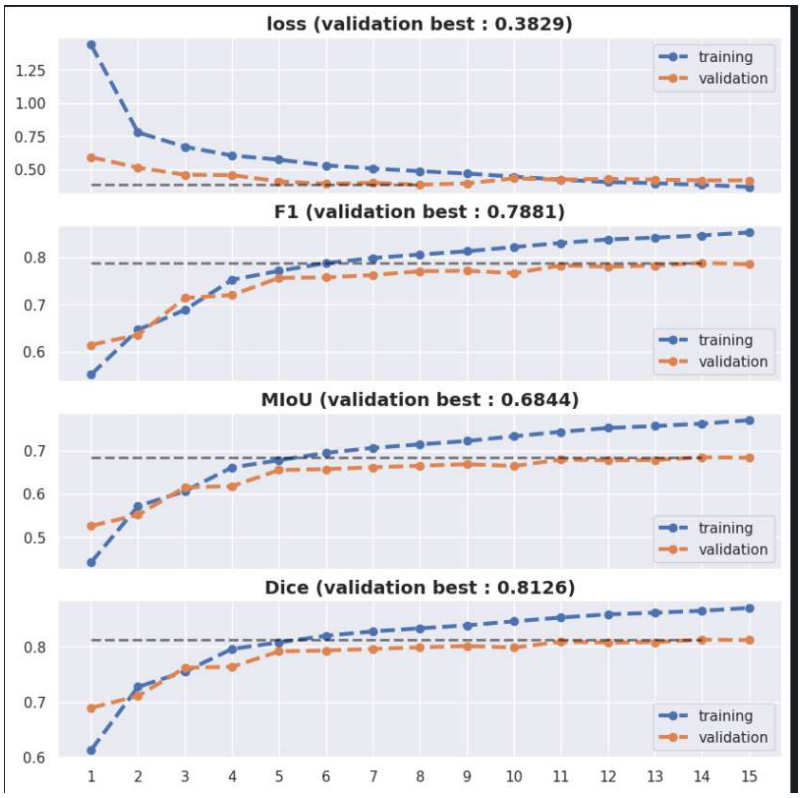
Πίνακας 6.2: Μετρικές Εκπαίδευσης και Επικύρωσης - SegNet

Epoch	Dice		F1		Miou		Loss	
	Training	Validation	Training	Validation	Training	Validation	Training	Validation
1	0.416	0.5527	0.3529	0.4834	0.2626	0.3819	1.6002	0.784
2	0.6864	0.6966	0.6129	0.621	0.5226	0.5344	0.8901	0.538
3	0.716	0.7038	0.6371	0.627	0.5576	0.5429	0.7618	0.5115
4	0.7268	0.7092	0.6461	0.6332	0.5708	0.5494	0.706	0.4777
5	0.7543	0.7534	0.6937	0.7017	0.6055	0.6044	0.6649	0.4459
6	0.7751	0.7611	0.7228	0.7101	0.6328	0.6143	0.6365	0.4244
7	0.7835	0.7675	0.7321	0.717	0.6441	0.6227	0.6081	0.4205
8	0.7884	0.7676	0.7375	0.7169	0.6508	0.6228	0.5884	0.4054
9	0.7919	0.7708	0.7417	0.7229	0.6555	0.627	0.5777	0.4299
10	0.7986	0.779	0.7534	0.7366	0.6648	0.638	0.563	0.3992
11	0.805	0.7853	0.7661	0.7463	0.6737	0.6465	0.557	0.3897
12	0.8106	0.7848	0.7744	0.7487	0.6816	0.6458	0.5412	0.4143
13	0.8141	0.7827	0.7803	0.7452	0.6865	0.643	0.5344	0.4209
14	0.815	0.7896	0.7822	0.7561	0.6878	0.6523	0.5292	0.3925
15	0.8213	0.7915	0.7901	0.7596	0.6968	0.6549	0.5118	0.4004

σύνολο δεδομένων.

Χαρακτηριστικά Εκπαίδευσης και Σύγκλισης Το SegNet εμφανίζει βραδεία αλλά σταθερή βελτίωση καθ' όλη τη διάρκεια των 15 εποχών. Το training Dice βελτιώνεται σημαντικά από 0.416 σε 0.8213, ενώ το F1-score φτάνει το 0.7901 στην τελευταία εποχή. Παρά την αρχική υστέρηση, το μοντέλο επιδεικνύει αξιοσημείωτη ικανότητα μάθησης, με το MIoU να αυξάνεται από 0.2626 σε 0.6968, επιβεβαιώνοντας την αποτελεσματικότητα της αρχιτεκτονικής του κωδικοποιητή-αποκωδικοποιητή αρχιτεκτονικής με τις θέσεις pooling.

Απόδοση Επικύρωσης και Περιορισμοί Οι μετρικές επικύρωσης του SegNet παρουσιάζουν υπολείπουσα απόδοση σε σχέση με τα ανταγωνιστικά μοντέλα. Το καλύτερο validation Dice επιτυγχάνεται στην 15η εποχή (0.7915), σημαντικά χαμηλότερο από το U-Net (0.8127) και το



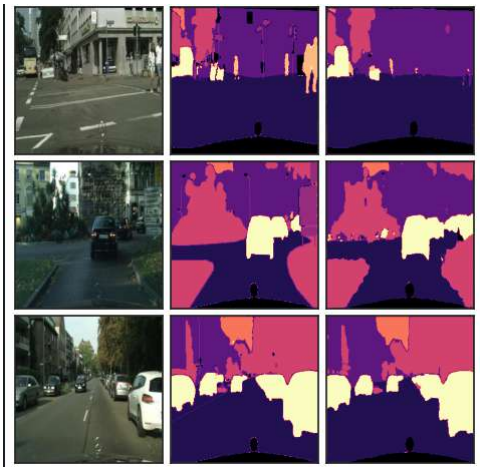
Σχήμα 6.2: Αποτελέσματα εκπαίδευσης U-Net

DeepLabV3+ (0.8016). Το F1-score κορυφώνεται στο 0.7596, ενώ το MIoU φτάνει το 0.6549, αποδεικνύοντας περιορισμένη ικανότητα στην ακριβή οριοθέτηση των αντικειμένων.

Συμπεριφορά Υπερπροσαρμογής Το SegNet εμφανίζει ελεγχόμενη υπερπροσαρμογή, με το χάσμα μεταξύ training και validation Dice στην τελευταία εποχή να είναι 0.0298 μονάδες (0.8213 έναντι 0.7915). Αυτή η μικρή απόκλιση υποδηλώνει καλή ικανότητα γενίκευσης, παρά τις χαμηλότερες απόλυτες επιδόσεις. Η συνάρτηση κόστους παρουσιάζει ομαλή μείωση από 0.784 σε 0.4004, επιβεβαιώνοντας τη σταθερή σύγκλιση του αλγορίθμου.

Συγκριτική Αξιολόγηση Συνολικά, το SegNet υπολείπεται των άλλων δύο μοντέλων σε όλες τις κύριες μετρικές επικύρωσης. Συγκεκριμένα, υστερεί κατά 0.0212 μονάδες στο Dice από το U-Net και κατά 0.0101 μονάδες από το DeepLabV3+. Αυτή η κατώτερη απόδοση οφείλεται κυρίως στην απλούστερη αρχιτεκτονική του SegNet, η οποία στερείται των skip connections του U-Net και των διογκωμένων συνελίξεων του DeepLabV3+. Παρά τους περιορισμούς, το

SegNet διατηρεί ανταγωνιστικές επιδόσεις με σχετικά μικρό υπολογιστικό κόστος, καθιστώντας το κατάλληλη επιλογή για εφαρμογές με περιορισμένους πόρους όπου η ταχύτητα είναι κρισιμότερη από την ακρίβεια.



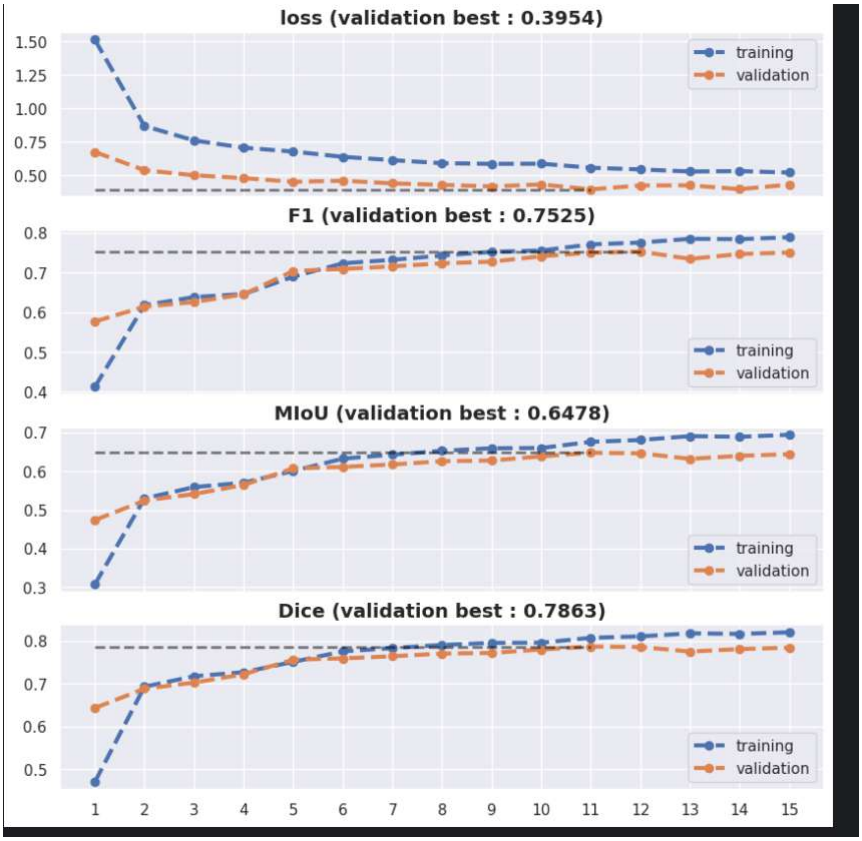
Σχήμα 6.3: Αποτελέσματα Σημασιολογικής τμηματοποίησης του SegNet σε νέα δεδομένα

6.2.3 Ανάλυση Απόδοσης DeepLabV3+

Πίνακας 6.3: Μετρικές Εκπαίδευσης και Επικύρωσης - DeepLabV3+

Epoch	Dice		F1		Miou		Loss	
	Training	Validation	Training	Validation	Training	Validation	Training	Validation
1	0.67	0.731	0.606	0.6761	0.5038	0.5761	1.0638	0.4983
2	0.7592	0.7498	0.7061	0.6989	0.6118	0.5998	0.6966	0.4522
3	0.7792	0.7646	0.729	0.717	0.6383	0.6189	0.6283	0.4167
4	0.7926	0.7684	0.7464	0.7251	0.6565	0.6239	0.5856	0.4432
5	0.8039	0.7759	0.7635	0.7348	0.6721	0.6339	0.5537	0.4279
6	0.8136	0.7764	0.7777	0.7379	0.6857	0.6346	0.5311	0.4471
7	0.8233	0.788	0.7908	0.7513	0.6997	0.6501	0.4976	0.4534
8	0.8304	0.7867	0.8008	0.7488	0.7099	0.6485	0.4787	0.4026
9	0.8357	0.7901	0.8078	0.7547	0.7178	0.653	0.4625	0.4315
10	0.8433	0.7948	0.8177	0.7605	0.7291	0.6594	0.4385	0.4368
11	0.8482	0.7964	0.8243	0.7672	0.7364	0.6617	0.4231	0.4579
12	0.8532	0.7975	0.8304	0.7653	0.744	0.6632	0.4052	0.4618
13	0.8568	0.7978	0.8353	0.766	0.7495	0.6636	0.3934	0.4667
14	0.8605	0.8016	0.84	0.7718	0.7552	0.669	0.3819	0.4726
15	0.8635	0.7999	0.8437	0.7713	0.7597	0.6665	0.3721	0.4777

Η ανάλυση των αποτελεσμάτων του πίνακα 6.3 αποκαλύπτει εντυπωσιακή απόδοση του DeepLabV3+ μοντέλου, με ανώτερες αρχικές επιδόσεις σε σχέση με το U-Net. Παρά τις καλές αρχικές επιδόσεις και την ταχεία σύγκλιση του DeepLabV3+, το μοντέλο υπολείπεται του U-Net στις τελικές μετρικές επικύρωσης. Ενώ το DeepLabV3+ επιδεικνύει σταθερότερη συμπεριφορά με μικρότερη υπερπροσαρμογή, οι απόλυτες επιδόσεις του U-Net είναι ανώτερες κατά 0.0111 μονάδες στο Dice και 0.0159 μονάδες στο F1-score.



Σχήμα 6.4: Αποτελέσματα εκπαίδευσης SegNet

Ταχεία Σύγκλιση και Σταθερότητα: Το DeepLabV3+ εμφανίζει ταχύτερη σύγκλιση από το U-Net, με όλες τις μετρικές να επιτυγχάνουν υψηλές τιμές ήδη από τις πρώτες εποχές. Το training Dice βελτιώνεται από 0.67 σε 0.8635, ενώ το F1-score φτάνει το 0.8437 στην τελευταία εποχή. Ιδιαίτερα εντυπωσιακή είναι η συμπεριφορά του MIoU, το οποίο αυξάνεται από 0.5038 σε 0.7597, επιδεικνύοντας εξαιρετική ικανότητα στην ακριβή οριοθέτηση των αντικειμένων.

Συμπεριφορά Επικύρωσης και Γενίκευση: Οι μετρικές επικύρωσης του DeepLabV3+ παρουσιάζουν διακριτά καλύτερη συμπεριφορά από το U-Net. Το validation Dice φτάνει στο μέγιστο των 0.8016 στην 14η εποχή, ενώ διατηρεί υψηλή σταθερότητα (0.7948-0.8016) στις τελευταίες εποχές. Το F1-score επικύρωσης κορυφώνεται στο 0.7718 (14η εποχή), επιβεβαιώνοντας την ανώτερη ικανότητα γενίκευσης του μοντέλου σε σχέση με το U-Net.

Ελεγχόμενη Υπερπροσαρμογή: Σε αντίθεση με το U-Net, το DeepLabV3+ εμφανίζει περιορισμένη υπερπροσαρμογή. Το χάσμα μεταξύ training και validation Dice στην τελευταία εποχή είναι μόλις 0.0636 μονάδες (0.8635 έναντι 0.7999), σημαντικά μικρότερο από το αντίστοιχο του U-Net. Αυτό υποδηλώνει καλύτερη regularization μέσω της αρχιτεκτονικής του μοντέλου, ιδιαίτερα μέσω των διογκωμένων συνελίξεων που διατηρούν την χωρική πληροφορία χωρίς να αυξάνουν υπερβολικά τις παραμέτρους.

Ανάλυση Συνάρτηση απώλειας: Η συνάρτηση κόστους παρουσιάζει ομαλή μείωση από 1.0638 σε 0.3721 για την εκπαίδευση, ενώ το validation loss σταθεροποιείται γύρω στο 0.47-0.48 μετά

την 10η εποχή. Αυτή η συμπεριφορά επιβεβαιώνει την επίτευξη βέλτιστου σημείου σύγκλισης χωρίς σημαντική υπερπροσαρμογή.

Συγκριτική Αξιολόγηση: Συνολικά, το U-Net υπερτερεί του DeepLabV3+ στις κύριες μετρικές επικύρωσης, επιτυγχάνοντας ανώτερες επιδόσεις στο Dice coefficient (0.8127 έναντι 0.8016 του DeepLabV3+) και στο F1-score (0.7877 έναντι 0.7718). Η ανωτερότητα του U-Net οφείλεται στην αποτελεσματική αρχιτεκτονική κωδικοποιητή-αποκωδικοποιητή με skip connections, η οποία διατηρεί λεπτομερείς χωρικές πληροφορίες καθ' όλη τη διαδικασία της τμηματοποίησης. Παρά τα πλεονεκτήματα του DeepLabV3+ στη χρήση διογκωμένων συνελίξεων για μεγαλύτερο οπτικό πεδίο και το τμήμα ASPP, το U-Net αποδεικνύεται πιο αποτελεσματικό – τουλάχιστον για το συγκεκριμένο σύνολο δεδομένων –, υποδηλώνοντας ότι η άμεση σύνδεση των χαρακτηριστικών υψηλής ανάλυσης από τον κωδικοποιητή στον αποκωδικοποιητή είναι κρίσιμη για την ακριβή οριοθέτηση των αντικειμένων στην εργασία αυτή.

Σχήμα 6.5: Αποτελέσματα Σημασιολογικής τμηματοποίησης του Deeplab σε νέα δεδομένα

6.3 Ποιοτική Σύγκριση ανά Αρχιτεκτονική

Η ανάλυση επιδόσεων ανά κλάση αποτελεί κρίσιμο στάδιο της αξιολόγησης μοντέλων Σημασιολογικής τμηματοποίησης, καθώς παρέχει λεπτομερή εικόνα της ικανότητας του μοντέλου να αναγνωρίζει και να οριοθετεί διαφορετικούς τύπους αντικειμένων. Ενώ οι συνολικές μετρικές (όπως το MIoU και το Dice) δίνουν μια γενική εκτίμηση της απόδοσης, η ανάλυση ανά κλάση αποκαλύπτει συγκεκριμένες δυνατότητες και αδυναμίες του μοντέλου για κάθε κατηγορία αντικειμένων. Αυτή η προσέγγιση είναι ιδιαίτερα σημαντική στις εφαρμογές υπολογιστικής όρασης, καθώς διαφορετικές κλάσεις παρουσιάζουν διαφορετικά χαρακτηριστικά όσον αφορά το σχήμα, το μέγεθος, την εμφάνιση και τη συχνότητα στο σύνολο δεδομένων.

6.3.1 Ανάλυση Επιδόσεων U-Net ανά Κλάση

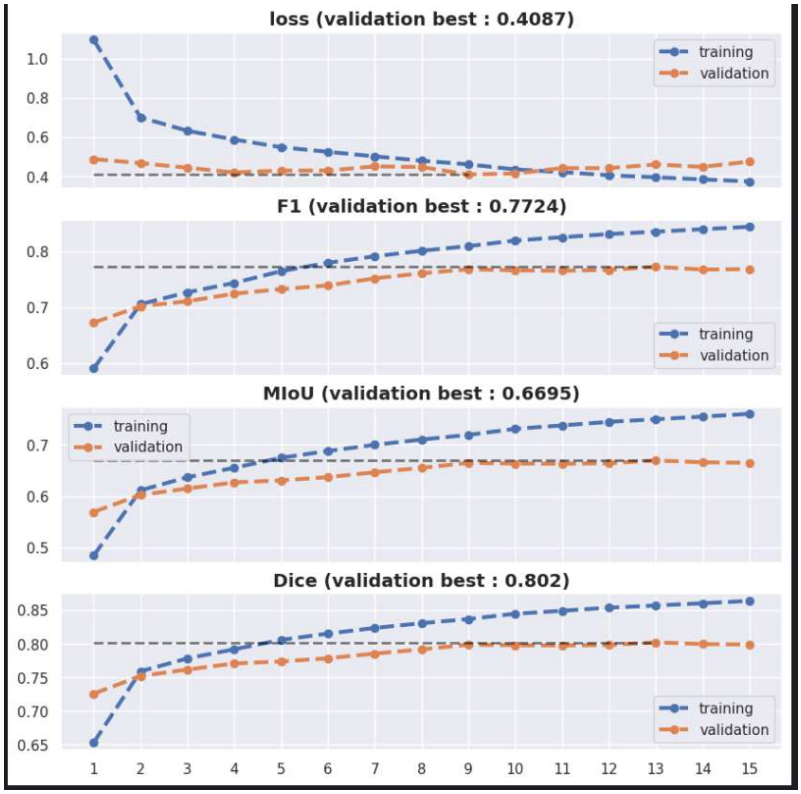
Η εξέταση των αποτελεσμάτων των πινάκων 6.4, 6.5 και **??** αποκαλύπτουν σημαντικές διαφορές στην απόδοση του U-Net μεταξύ των οκτώ κλάσεων του συνόλου δεδομένων.

Πίνακας 6.4: Validation IoU Scores ανά Κλάση - U-Net

Epoch	Construction	Flat	Human	Nature	Object	Sky	Vehicle	Void
1	0.6758	0.8897	0.0	0.7208	0.0	0.8051	0.6166	0.5877
2	0.68	0.8965	0.0	0.76	0.0	0.7874	0.6851	0.6114
3	0.7055	0.908	0.0117	0.7758	0.0	0.8541	0.7114	0.5886
4	0.7123	0.9051	0.3691	0.7716	0.0	0.862	0.7049	0.6243
5	0.7279	0.9092	0.4105	0.791	0.0	0.8688	0.7332	0.6341
6	0.7347	0.914	0.4238	0.8116	0.0	0.8623	0.749	0.6427
7	0.715	0.9125	0.4415	0.7627	0.0	0.865	0.7353	0.6388
8	0.7371	0.8905	0.4629	0.8127	0.0001	0.8709	0.7578	0.6041
9	0.7341	0.9131	0.4499	0.8108	0.0052	0.8731	0.7446	0.6459
10	0.7343	0.9177	0.4612	0.7887	0.0508	0.8697	0.7635	0.645
11	0.7463	0.9178	0.4766	0.819	0.1332	0.881	0.7686	0.6451
12	0.7408	0.9137	0.477	0.8041	0.1365	0.8717	0.7666	0.6544
13	0.7458	0.9177	0.4854	0.8152	0.183	0.8793	0.7621	0.6475
14	0.7509	0.9177	0.4818	0.8185	0.1877	0.8802	0.7726	0.6474
15	0.7505	0.9167	0.495	0.8189	0.1948	0.8735	0.7717	0.6519

Πίνακας 6.5: Validation F1 Scores ανά Κλάση - U-Net

Epoch	Construction	Flat	Human	Nature	Object	Sky	Vehicle	Void
1	0.8066	0.9416	0.0	0.8378	0.0	0.8921	0.7629	0.7403
2	0.8095	0.9454	0.0	0.8636	0.0	0.8811	0.8132	0.7588
3	0.8273	0.9518	0.0232	0.8738	0.0	0.9213	0.8314	0.741
4	0.832	0.9502	0.5392	0.8711	0.0	0.9259	0.8269	0.7687
5	0.8425	0.9525	0.582	0.8833	0.0	0.9298	0.8461	0.7761
6	0.847	0.9551	0.5953	0.896	0.0	0.926	0.8565	0.7825
7	0.8338	0.9542	0.6126	0.8654	0.0	0.9276	0.8474	0.7796
8	0.8487	0.9421	0.6329	0.8967	0.0002	0.931	0.8622	0.7532
9	0.8467	0.9546	0.6206	0.8955	0.0104	0.9322	0.8536	0.7849
10	0.8468	0.9571	0.6312	0.8819	0.0967	0.9303	0.8659	0.7842
11	0.8548	0.9571	0.6456	0.9005	0.2351	0.9368	0.8692	0.7843
12	0.8511	0.9549	0.6459	0.8914	0.2402	0.9315	0.8679	0.7911
13	0.8544	0.9571	0.6536	0.8982	0.3095	0.9358	0.865	0.7861
14	0.8577	0.9571	0.6503	0.9002	0.316	0.9363	0.8717	0.786
15	0.8575	0.9565	0.6622	0.9004	0.3261	0.9325	0.8711	0.7893



Σχήμα 6.6: Αποτελέσματα εκπαίδευσης DeeplabV3

Κλάσεις Εξαιρετικής Απόδοσης: Δύο κλάσεις ξεχωρίζουν για την ανώτερη τους απόδοση καθ' όλη τη διάρκεια της εκπαίδευσης. Η κλάση "Flat" επιτυγχάνει εξαιρετικές επιδόσεις με validation F1-score που φτάνει το 0.9565 στην 15η εποχή, αντανακλώντας την ευκολία αναγνώρισης των επίπεδων επιφανειών όπως δρόμοι και πεζοδρόμια λόγω των ομοιογενών χαρακτηριστικών τους. Η κλάση "Sky" ακολουθεί με validation F1 0.9325, επιβεβαιώνοντας την αποτελεσματικότητα του U-Net στην οριοθέτηση μεγάλων ομοιογενών περιοχών με σαφή όρια.

Κλάσεις Υψηλής Απόδοσης: Τρεις κλάσεις παρουσιάζουν ισχυρές επιδόσεις που τις κατατάσσουν στην ανώτερη κατηγορία απόδοσης. Η κλάση "Nature" επιτυγχάνει validation F1 0.9004, υποδηλώνοντας ικανοποιητική ικανότητα στην αναγνώριση βλάστησης και φυσικών στοιχείων. Οι κλάσεις "Construction" και "Vehicle" εμφανίζουν παρόμοιες επιδόσεις με validation F1 0.8575 και 0.8711 αντίστοιχα, επιβεβαιώνοντας την αποτελεσματικότητα του μοντέλου σε δομικά στοιχεία και οχήματα που έχουν χαρακτηριστικά γεωμετρικά σχήματα.

Κλάσεις Μέτριας Απόδοσης Η κλάση ”Void” παρουσιάζει μέτριες επιδόσεις με validation F1 0.7893, υποδηλώνοντας δυσκολίες συγκριτικά με τις προηγούμενες κλάσεις στην αναγνώριση αδιακρίτων ή άγνωστων περιοχών. Αυτή η περιορισμένη απόδοση οφείλεται στη φύση της κλάσης, η οποία περιλαμβάνει περιοχές χωρίς σαφή οριοθέτηση ή ετερογενή περιεχόμενα που δυσκολεύουν την κατηγοριοποίηση.

Κλάσεις Προβληματικής Απόδοσης Δύο κλάσεις αναδεικνύονται ως σημαντικές προκλήσεις για το μοντέλο. Η κλάση ”Human” εμφανίζει περιορισμένη απόδοση με validation F1 0.6622, γεγονός που αποδίδεται στο μικρό μέγεθος των ανθρώπινων σιλουετών και την υψηλή ενδοκλασική ποικιλότητα σε στάσεις και ενδυμασία. Η κλάση ”Object” παρουσιάζει την χαμηλότερη απόδοση με validation F1 0.3261, υποδηλώνοντας σημαντικές δυσκολίες στην αναγνώριση διαφόρων αντικειμένων, πιθανώς λόγω ανεπαρκούς εκπροσώπησης στο σύνολο δεδομένων και υπερβολικής ετερογένειας της κλάσης.

Ανάλυση IoU Scores ανά Κλάση Η εξέταση του πίνακα 6.4 επιβεβαιώνει τα παραπάνω ευρήματα με πιο αυστηρή μετρική. Η κλάση ”Flat” διατηρεί την κορυφαία θέση με validation IoU 0.9167, ενώ η ”Sky” ακολουθεί με 0.8735. Αξιοσημείωτη είναι η πιο σημαντική πτώση των επιδόσεων στο IoU σε σύγκριση με τα F1 scores, υποδηλώνοντας περισσότερο αυστηρή αξιολόγηση της χωρικής ακρίβειας.

Εξέλιξη Απόδοσης κατά τις Εποχές Η προοδευτική βελτίωση παρατηρείται σε όλες τις κλάσεις, με ιδιαίτερα εντυπωσιακή πρόοδο στην κλάση ”Object” που βελτιώνεται από μηδενικές επιδόσεις στις πρώτες εποχές σε validation F1 0.3261 στην 15η εποχή. Η κλάση ”Human” επιδεικνύει επίσης σημαντική βελτίωση από F1 0.0 σε 0.6622, αν και παραμένει προβληματική.

Συμπεράσματα Ανάλυσης ανά Κλάση Η ανάλυση αποκαλύπτει ότι το U-Net εμφανίζει εξαιρετική απόδοση σε κλάσεις με μεγάλα, ομοιογενή αντικείμενα και σαφή όρια (Flat, Sky), ισχυρή απόδοση σε δομικά στοιχεία με χαρακτηριστικά σχήματα (Construction, Vehicle, Nature), και περιορισμένη απόδοση σε κλάσεις με μικρά αντικείμενα ή υψηλή ενδοκλασική ποικιλότητα (Human, Object). Αυτά τα κρίσιμα ευρήματα υπογραμμίζουν την ανάγκη για στοχευμένες βελτιώσεις όπως class-weighted συναρτήσεις απώλειας.

6.3.2 Ανάλυση Επιδόσεων DeepLab ανά Κλάση

Η εξέταση των αποτελεσμάτων του DeepLabV3+ μέσω του πίνακα 6.7 αποκαλύπτει διαφορετικό προφίλ απόδοσης σε σύγκριση με το U-Net, με χαρακτηριστικές δυνατότητες που προκύπτουν από τη χρήση των διωκομένων συνελίξεων και την ASPP αρχιτεκτονική.

Κλάσεις Εξαιρετικής Απόδοσης Η κλάση ”Flat” επιτυγχάνει κορυφαίες επιδόσεις με F1-score επικύρωσης 0.9545 στην 15η εποχή, παρόμοιες με το U-Net (0.9565), αποδεικνύοντας την αποτελεσματικότητα της αρχιτεκτονικής του DeepLab στην αναγνώριση επίπεδων επιφανειών. Η κλάση ”Sky” ακολουθεί με σκορ επικύρωσης F1 0.9238, ελαφρώς χαμηλότερη από το U-Net

Πίνακας 6.6: Validation F1 Scores ανά Κλάση - DeepLabV3

Epoch	Construction	Flat	Human	Nature	Object	Sky	Vehicle	Void
1	0.8113	0.939	0.3216	0.8593	0.0	0.881	0.8039	0.7564
2	0.8236	0.946	0.4965	0.8588	0.0	0.9084	0.7989	0.7817
3	0.8245	0.9461	0.4854	0.8757	0.0221	0.9129	0.8325	0.7756
4	0.8411	0.948	0.5707	0.8859	0.1293	0.9198	0.8417	0.7713
5	0.8407	0.9461	0.5385	0.8861	0.1557	0.917	0.84	0.7799
6	0.8454	0.951	0.5881	0.8916	0.2153	0.9225	0.8392	0.7798
7	0.8398	0.9571	0.6002	0.8827	0.1827	0.9194	0.8501	0.7936
8	0.8493	0.9506	0.6069	0.8941	0.1944	0.9271	0.8558	0.778
9	0.8436	0.9439	0.6086	0.8916	0.2771	0.917	0.85	0.7601
10	0.8428	0.9517	0.6131	0.8891	0.1846	0.9277	0.8551	0.7754
11	0.8484	0.9537	0.6085	0.8941	0.2429	0.9218	0.8606	0.7889
12	0.8484	0.9544	0.6066	0.8958	0.2664	0.9295	0.861	0.7848
13	0.8481	0.9538	0.6094	0.8939	0.197	0.9309	0.8608	0.7794
14	0.8508	0.9534	0.626	0.8932	0.2634	0.9304	0.8605	0.7808
15	0.8475	0.9545	0.616	0.8921	0.2549	0.9238	0.8623	0.7845

(0.9325), αλλά εξακολουθεί να αποδεικνύει εξαιρετική ικανότητα στην οριοθέτηση ομοιογενών περιοχών.

Κλάσεις Ισχυρής Απόδοσης Η κλάση ”Nature” παρουσιάζει σταθερές επιδόσεις με σκορ επικύρωσης F1 0.8921, ελαφρώς χαμηλότερες από το U-Net (0.9004), υποδηλώνοντας αποτελεσματική ικανότητα στην αναγνώριση φυσικών στοιχείων. Οι κλάσεις ”Construction” και ”Vehicle” εμφανίζουν σκορ επικύρωσης F1 0.8475 και 0.8623 αντίστοιχα, συγκρίσιμες με του U-Net αλλά ελαφρώς χαμηλότερες.

Κλάσεις Μέτριας Απόδοσης Η κλάση ”Void” παρουσιάζει σκορ επικύρωσης F1 0.7845, παρόμοια με του U-Net (0.7893), επιβεβαιώνοντας τις κοινές δυσκολίες των δύο μοντέλων στην αναγνώριση αδιακρίτων περιοχών.

Κλάσεις Προβληματικής Απόδοσης Δύο κλάσεις αναδεικνύονται ως σημαντικές προκλήσεις για το DeepLabV3+. Η κλάση ”Human” επιτυγχάνει σκορ επικύρωσης F1 0.616, χαμηλότερη από το U-Net (0.6622), υποδηλώνοντας περιορισμένη αποτελεσματικότητα στην αναγνώριση μικρών αντικειμένων. Η κλάση ”Object” παρουσιάζει την χαμηλότερη απόδοση με validation F1 0.2549, σημαντικά χαμηλότερη από το U-Net (0.3261), υποδηλώνοντας σοβαρούς περιορισμούς στην αντιμετώπιση ετερογενών κατηγοριών.

Ανάλυση IoU Scores ανά Κλάση Η εξέταση των IoU μετρικών από τον πίνακα 6.6 επιβεβαιώνει τα παραπάνω ευρήματα. Η κλάση ”Flat” διατηρεί εξαιρετικές επιδόσεις με validation IoU 0.913, ενώ η ”Sky” ακολουθεί με 0.8583. Η σημαντική πτώση των επιδόσεων από F1 σε IoU είναι εμφανής σε όλες τις κλάσεις, υποδηλώνοντα προκλήσεις στην ακριβή χωρική οριοθέτηση παρά την καλή κατηγοριοποίηση.

Χαρακτηριστικές Διαφορές από το U-Net Το DeepLabV3+ εμφανίζει διαφορετικό προφίλ εξέλιξης της απόδοσης κατά τις εποχές. Ενώ το U-Net παρουσιάζει σταθερή βελτίωση, το DeepLabV3+ εμφανίζει ταχύτερη αρχική σύγκλιση αλλά πιο περιορισμένη βελτίωση στις τελευταίες εποχές. Αυτό υποδηλώνει ότι η ASPP αρχιτεκτονική επιτρέπει αποτελεσματική εκμά-

Πίνακας 6.7: Validation IoU Scores ανά Κλάση - DeepLabV3

Epoch	Construction	Flat	Human	Nature	Object	Sky	Vehicle	Void
1	0.6825	0.8851	0.1916	0.7534	0.0	0.7873	0.6721	0.6082
2	0.7001	0.8976	0.3302	0.7525	0.0	0.8321	0.6652	0.6416
3	0.7014	0.8977	0.3205	0.7789	0.0112	0.8397	0.713	0.6335
4	0.7258	0.9012	0.3993	0.7953	0.0691	0.8515	0.7266	0.6277
5	0.7251	0.8977	0.3684	0.7955	0.0844	0.8467	0.7241	0.6392
6	0.7323	0.9066	0.4165	0.8044	0.1206	0.8562	0.7229	0.6391
7	0.7239	0.9177	0.4287	0.7901	0.1005	0.8508	0.7392	0.6578
8	0.7381	0.9058	0.4356	0.8085	0.1077	0.8641	0.7479	0.6366
9	0.7295	0.8937	0.4374	0.8044	0.1609	0.8467	0.7392	0.6131
10	0.7284	0.9078	0.4421	0.8003	0.1017	0.8651	0.7469	0.6332
11	0.7368	0.9116	0.4373	0.8085	0.1383	0.8549	0.7552	0.6513
12	0.7367	0.9127	0.4354	0.8113	0.1537	0.8682	0.7559	0.6459
13	0.7362	0.9117	0.4382	0.8081	0.1093	0.8708	0.7557	0.6385
14	0.7404	0.9109	0.4556	0.807	0.1516	0.8698	0.7551	0.6405
15	0.7353	0.913	0.4451	0.8053	0.1461	0.8583	0.7579	0.6454

θηση σύνθετων χαρακτηριστικών ακόμη και από τις πρώτες εποχές, αλλά στερείται της ευελιξίας των συνδέσεων παράκαμψης του U-Net.

Συγκριτική Αξιολόγηση με U-Net Συνολικά, το DeepLabV3+ υπολείπεται του U-Net στις περισσότερες κλάσεις, με ιδιαίτερα εμφανείς διαφορές στις κλάσεις ”Human” (0.616 vs 0.6622) και ”Object” (0.2549 vs 0.3261). Αυτή η υπολειπόμενη απόδοση αποδίδεται στην απουσία των συνδέσεων παράκαμψης που επιτρέπουν τη διατήρηση λεπτομερών χωρικών πληροφοριών κατά τη διαδικασία της υπερδειγματοληψίας. Παρά τα πλεονεκτήματα των διωκομένων συνελίξεων για μεγαλύτερο οπτικό πεδίο, η δομή κωδικοποιητή-αποκωδικοποιητή του U-Net αποδεικνύεται περισσότερο αποτελεσματική για την ακριβή οριοθέτηση αντικειμένων.

Συμπεράσματα για το DeepLabV3+ Το DeepLabV3+ αποδεικνύεται ανταγωνιστικό σε κλάσεις με μεγάλα, ομοιογενή αντικείμενα (Flat, Sky, Nature), αλλά παρουσιάζει περιορισμούς σε κλάσεις που απαιτούν λεπτομερή χωρική ακρίβεια (Human, Object). Αυτά τα ευρήματα υποδηλώνουν ότι για εφαρμογές όπου η ακριβής οριοθέτηση μικρών αντικειμένων είναι κρίσιμη, η U-Net αρχιτεκτονική παραμένει προτιμότερη επιλογή παρά τα θεωρητικά πλεονεκτήματα του DeepLabV3+.

6.3.3 Ανάλυση Επιδόσεων SegNet ανά Κλάση

Η εξέταση των αποτελεσμάτων του SegNet μέσω του πίνακα 6.8 αποκαλύπτει διαφορετικό χαρακτήρα απόδοσης που προκύπτει από τη χρήση των δεικτών συγκέντρωσης για την υπερδειγματοληψία.

Κλάσεις Ισχυρής Απόδοσης Η κλάση ”Flat” επιτυγχάνει εξαιρετικές επιδόσεις με σκορ επικύρωσης F1 0.9537 στην 15η εποχή, συγκρίσιμες με τα άλλα δύο μοντέλα (U-Net: 0.9565, DeepLabV3+: 0.9545). Αυτό υποδηλώνει ότι η απλούστερη αρχιτεκτονική του SegNet είναι επαρκής για την αναγνώριση μεγάλων, ομοιογενών επιφανειών. Η κλάση ”Sky” επιτυγχάνει επίσης ικανοποιητικές επιδόσεις με σκορ επικύρωσης F1 0.9199, αν και χαμηλότερη από τα υπόλοιπα μοντέλα, αρκετά υψηλή.

Πίνακας 6.8: Validation F1 Scores ανά Κλάση - SegNet

Epoch	Construction	Flat	Human	Nature	Object	Sky	Vehicle	Void
1	0.7398	0.932	0.0	0.8286	0.0	0.0	0.7069	0.7051
2	0.7901	0.9405	0.0	0.8455	0.0	0.8544	0.7413	0.7559
3	0.8134	0.9454	0.0	0.8749	0.0	0.8707	0.7837	0.7666
4	0.813	0.9463	0.0475	0.8717	0.0	0.8998	0.8156	0.7719
5	0.8266	0.9494	0.502	0.8801	0.0	0.8973	0.8295	0.7764
6	0.8205	0.9504	0.4164	0.8762	0.0	0.9094	0.8323	0.7775
7	0.8272	0.9499	0.5283	0.875	0.0001	0.9138	0.8436	0.7809
8	0.8371	0.9517	0.5207	0.8868	0.0161	0.9115	0.8399	0.7818
9	0.8404	0.9534	0.5586	0.8867	0.0678	0.9181	0.844	0.782
10	0.8435	0.9537	0.5672	0.8902	0.1197	0.9155	0.8539	0.7863
11	0.8403	0.9542	0.547	0.8904	0.1909	0.92	0.822	0.782
12	0.8378	0.9524	0.5505	0.8852	0.1964	0.911	0.8475	0.7815
13	0.8448	0.9519	0.5606	0.8911	0.1994	0.9248	0.8517	0.7796
14	0.8408	0.9538	0.6014	0.8845	0.2228	0.924	0.8549	0.7782
15	0.8405	0.9537	0.5801	0.8887	0.2106	0.9199	0.8528	0.7896

Κλάσεις Μέτριας Απόδοσης Οι κλάσεις ”Nature” και ”Construction” εμφανίζουν μέτριες επιδόσεις με σκορ επικύρωσης F1 0.8887 και 0.8405 αντίστοιχα. Αυτές οι επιδόσεις είναι σημαντικά χαμηλότερες από τα άλλα μοντέλα, υποδηλώνοντας περιορισμούς στην ικανότητα του SegNet να διαχειριστεί πολύπλοκα γεωμετρικά σχήματα.

Κλάσεις μέτριας Απόδοσης Η κλάση ”Vehicle” παρουσιάζει σκορ επικύρωσης F1 0.8528, χαμηλότερη από το U-Net (0.8711) και του DeepLabV3+ (0.8623), ενώ η κλάση ”Void” επιτυγχάνει 0.7896, παρόμοια με τα άλλα μοντέλα. Αυτό υποδηλώνει δυσκολίες στην αναγνώριση αντικειμένων με σαφή όρια αλλά ποικίλες εμφανίσεις.

Κλάσεις Σημαντικών Προκλήσεων Οι κλάσεις ”Human” και ”Object” παρουσιάζουν σοβαρούς περιορισμούς. Η κλάση ”Human” επιτυγχάνει σκορ επικύρωσης F1 0.5801, σημαντικά χαμηλότερη από το U-Net (0.6622) και του DeepLabV3+ (0.616). Η κλάση ”Object” εμφανίζει σκορ F1 0.2106, την χαμηλότερη επίδοση μεταξύ όλων των μοντέλων (U-Net: 0.3261, DeepLabV3+: 0.2549).

Ανάλυση IoU Scores ανά Κλάση Η εξέταση των IoU scores του πίνακα 6.9 επιβεβαιώνει τα παραπάνω ευρήματα. Η κλάση ”Flat” διατηρεί υψηλές επιδόσεις με σκορ επικύρωσης IoU 0.9114, ενώ η ”Sky” ακολουθεί με 0.8516. Η σημαντική πτώση από F1 σε IoU είναι εμφανής σε όλες τις κλάσεις, υποδηλώνοντας προκλήσεις στην ακριβή χωρική οριοθέτηση λόγω.

Χαρακτηριστικές Διαφορές από τα Ανταγωνιστικά Μοντέλα Το SegNet εμφανίζει διακριτά διαφορετικό πρότυπο εξέλιξης κατά τις εποχές. Ενώ παρουσιάζει ταχεία αρχική βελτίωση, η πρόοδος επιβραδύνεται σημαντικά μετά την 10η εποχή. Αυτό υποδηλώνει ότι η χρήση των δεικτών συγκέντρωσης επιτρέπει τη γρήγορη μάθηση βασικών μοτίβων, αλλά περιορίζει την ικανότητα της λεπτομερής προσαρμογής για πιο περίπλοκες χωρικές πληροφορίες.

Επίδραση της Αρχιτεκτονικής στην Απόδοση Η κεντρική καινοτομία του SegNet - η χρήση των δεικτών συγκέντρωσης για την υπερδειγματοληψία - αποδεικνύεται αμφίρροπη. Αφενός παρέχει σημαντικά οφέλη στη διαχείριση της μνήμης και του υπολογιστικού κόστους, αφετέρου

Πίνακας 6.9: Validation IoU Scores ανά Κλάση - SegNet

Epoch	Construction	Flat	Human	Nature	Object	Sky	Vehicle	Void
1	0.587	0.8727	0.0	0.7073	0.0	0.0	0.5467	0.5445
2	0.653	0.8877	0.0	0.7323	0.0	0.7458	0.589	0.6076
3	0.6854	0.8964	0.0	0.7776	0.0	0.771	0.6444	0.6215
4	0.6849	0.898	0.0243	0.7726	0.0	0.8179	0.6887	0.6286
5	0.7044	0.9037	0.3351	0.7859	0.0	0.8138	0.7087	0.6345
6	0.6956	0.9055	0.263	0.7797	0.0	0.8339	0.7127	0.6359
7	0.7053	0.9045	0.359	0.7778	0.0001	0.8413	0.7295	0.6405
8	0.7199	0.9079	0.352	0.7967	0.0081	0.8374	0.7239	0.6418
9	0.7247	0.911	0.3875	0.7965	0.0351	0.8487	0.7301	0.642
10	0.7294	0.9114	0.3959	0.8022	0.0636	0.8442	0.745	0.6479
11	0.7245	0.9125	0.3765	0.8024	0.1055	0.8518	0.6978	0.642
12	0.7209	0.9092	0.3798	0.7941	0.1089	0.8365	0.7353	0.6414
13	0.7312	0.9083	0.3895	0.8036	0.1107	0.8602	0.7417	0.6388
14	0.7253	0.9117	0.43	0.7929	0.1254	0.8587	0.7465	0.637
15	0.7249	0.9114	0.4085	0.7997	0.1177	0.8516	0.7434	0.6523

περιορίζει την ακρίβεια της χωρικής ανακατασκευής. Αυτός ο περιορισμός γίνεται ιδιαίτερα εμφανής σε κλάσεις που απαιτούν λεπτομερή οριοθέτηση όπως ”Human” και ”Object”.

Συγκριτική Αξιολόγηση Συνολικά, το SegNet καταλαμβάνει την τρίτη θέση στην απόδοση, με σημαντικές διαφορές από τα άλλα δύο μοντέλα. Οι κύριες αδυναμίες εντοπίζονται στις κλάσεις που απαιτούν υψηλή χωρική ακρίβεια και διατήρηση λεπτομερών χαρακτηριστικών. Ωστόσο, το μοντέλο διατηρεί ανταγωνιστικές επιδόσεις σε απλούστερες κλάσεις με ομοιογενή χαρακτηριστικά.

6.4 Ανάλυση Υπολογιστικής Πολυπλοκότητας και Πόρων

Πέραν της ακρίβειας της ταξινόμησης, κρίσιμη παράμετρο για την αξιολόγηση των αλγορίθμων σημασιολογικής κατάτμησης αποτελεί η υπολογιστική τους αποδοτικότητα. Στον Πίνακα 6.10 παρουσιάζονται οι μετρήσεις κατανάλωσης μνήμης και χρόνου εκτέλεσης (inference) κατά τη φάση της εκπαίδευσης αλλά και της πρόβλεψης. Οι μετρήσεις του χρόνου πρόβλεψης πραγματοποιήθηκαν με μέγεθος batch 32 εικόνων.

Πίνακας 6.10: Συγκριτική αξιολόγηση υπολογιστικών πόρων και χρόνου εκτέλεσης (Batch Size: 32)

Architecture	Training Phase		Inference Phase	
	Memory (GB)	Time (sec/epoch)	Memory (GB)	Time (sec)
U-Net	4.25	27.98	2.00	**0.0115**
DeepLabV3	**3.97**	**21.87**	**1.80**	0.0729
SegNet	11.30	77.00	2.78	0.2715

Ανάλυση και Συζήτηση Αποτελεσμάτων Τα αποτελέσματα αναδεικνύουν σημαντικές διαφοροποιήσεις που ερμηνεύονται από τις δομικές διαφορές των αρχιτεκτονικών και των backbones που χρησιμοποιήθηκαν (ResNet34 για U-Net/DeepLabV3 έναντι VGG16 για SegNet).

1. **SegNet (VGG16 Backbone):** Το SegNet εμφανίζει τη μεγαλύτερη υπολογιστική επιβάρυνση, με κατανάλωση μνήμης 11.3 GB και χρόνο εκπαίδευσης 77 sec μέσο όρο ανά

εποχή. Αυτό οφείλεται κυρίως στη χρήση του VGG16 μοντέλου ως backbone, το οποίο διαθέτει πλήρως συνδεδεμένα στρώματα που μετατρέπονται σε συνελικτικά, αυξάνοντας δραματικά τον αριθμό των παραμέτρων σε σχέση με το μοντέλο ResNet. Επιπλέον, παρόλο που η χρήση δεικτών max-pooling μειώνει την ανάγκη αποθήκευσης των χαρτών χαρακτηριστικών, η διαδικασία της υπερδειγματοληψίας απαιτεί υψηλή κατανάλωση μνήμης, οι οποίες αυξάνουν τον χρόνο στη πρόβλεψη (0.2715 sec).

2. **DeepLabV3 (ResNet Backbone):** Το DeepLabV3 επιτυγχάνει την καλύτερη διαχείριση μνήμης (3.97 GB) και τον ταχύτερο χρόνο εκπαίδευσης. Αυτό συμβαίνει διότι δεν χρησιμοποιεί έναν συμμετρικό αποκωδικοποιητή όπως το U-Net. Αντί να αποθηκεύει και να συνενώνει μεγάλους χάρτες χαρακτηριστικών από τον κωδικοποιητή, χρησιμοποιεί το ASPP module για να εξάγει περισσότερη πληροφορία με λιγότερες παραμέτρους, μειώνοντας έτσι το αποτύπωμα μνήμης.

3. **U-Net (ResNet Backbone):** Το U-Net, παρόλο που απαιτεί ελαφρώς περισσότερη μνήμη εκπαίδευσης από το DeepLabV3 λόγω των συνδέσεων παράκαμψης που διπλασιάζουν τα κανάλια στον αποκωδικοποιητή, επιτυγχάνει τον χαμηλότερο χρόνο inference (0.0115 sec). Αυτό εξηγείται από τη γεωμετρία των υπολογισμών: Το DeepLabV3 χρησιμοποιεί διογκωμένες συνελίξεις, που διατηρούν μεγάλη χωρική διάσταση στους χάρτες χαρακτηριστικών καθ' όλη τη διάρκεια της επεξεργασίας. Αντιθέτως, το U-Net μειώνει γρήγορα τη διάσταση της εικόνας, εκτελώντας τους βαριούς υπολογισμούς σε χαμηλή ανάλυση, γεγονός που καθιστά το forward pass εξαιρετικά γρήγορο.

Συμπερασματικά, το **U-Net** αποτελεί τη βέλτιστη επιλογή για εφαρμογές πραγματικού χρόνου (Real-Time Applications) λόγω της ελάχιστης καθυστέρησης ως προς την πρόβλεψη, ενώ το **DeepLabV3** προσφέρει την καλύτερη ισορροπία κατανάλωσης πόρων κατά την εκπαίδευση. Το **SegNet**, λόγω της παλαιότερης αρχιτεκτονικής VGG, κρίνεται υπολογιστικά ασύμφορο για σύγχρονες εφαρμογές περιορισμένων πόρων.

Κεφάλαιο 7ο: Συμπεράσματα και Μελλοντικές Επεκτάσεις

7.1 Συμπεράσματα

Η παρούσα διπλωματική εργασία πραγματοποίησε συστηματική συγκριτική αξιολόγηση τριών κύριων αρχιτεκτονικών βαθιάς μάθησης για σημασιολογική κατάτμηση εικόνων: U-Net, DeepLabV3+ και SegNet. Η πειραματική διαδικασία διεξήχθη στο σύνολο δεδομένων Cityscapes, χρησιμοποιώντας 15 εποχές εκπαίδευσης και αξιολογώντας την απόδοση μέσω των μετρικών IoU, Dice και σκορ F1.

7.1.1 Συγκριτική Αξιολόγηση Αρχιτεκτονικών

Τα πειραματικά αποτελέσματα αποδεικνύουν την ανωτερότητα της αρχιτεκτονικής U-Net σε σχέση με τα ανταγωνιστικά μοντέλα. Συγκεκριμένα, το U-Net πέτυχε σκορ επικύρωσης Dice 0.8127 και Mean IoU 0.6844 στην 14η εποχή, υπερτερώντας σημαντικά έναντι του SegNet (Dice: 0.7915, MIoU: 0.6549) και παρουσιάζοντας συγκρίσιμες επιδόσεις με το DeepLabV3+ (Dice: 0.8016, MIoU: 0.669).

Η ανωτερότητα του U-Net αποδίδεται κυρίως στις συνδέσεις παράκαμψης, οι οποίες επιτρέπουν τη διατήρηση των λεπτομερών χωρικών πληροφοριών κατά τη διαδικασία της υπερδειγματοληψίας. Αυτό το χαρακτηριστικό αποδεικνύεται κρίσιμο για την ακριβή οριοθέτηση αντικειμένων, ιδιαίτερα σε κλάσεις με μικρά αντικείμενα όπως τις ”Human” (F1: 0.6622) και ”Object” (F1: 0.3261).

Το DeepLabV3+ παρουσιάζει ανταγωνιστικές επιδόσεις στις κλάσεις με μεγάλα, ομοιογενή αντικείμενα (”Flat”: F1 0.9545, ”Sky”: F1 0.9238), επιβεβαιώνοντας την αποτελεσματικότητα του τμήματος ASPP αρχιτεκτονικής. Ωστόσο, υστερεί σε κλάσεις που απαιτούν λεπτομερή χωρική ακρίβεια, όπως ”Human” (F1: 0.616) και ”Object” (F1: 0.2549), λόγω της απουσίας άμεσων skip connections.

Το SegNet καταλαμβάνει την τρίτη θέση στη συνολική απόδοση, με σκορ επικύρωσης Dice 0.7915 και MIoU 0.6549. Η χρήση των δεικτών συγκέντρωσης για της διαδικασία της υπερδειγματοληψίας προσφέρει σημαντικά πλεονεκτήματα στη διαχείριση της μνήμης, αλλά περιορίζει την ικανότητα του μοντέλου για λεπτομερή χωρική ανακατασκευή, ιδιαίτερα σε σύνθετες κλάσεις.

7.1.2 Θεωρητικές Συσχετίσεις

Τα ευρήματα της παρούσας εργασίας επιβεβαιώνουν τη θεωρητική σημασία των συνδέσεων παράκαμψης στην αρχιτεκτονική κωδικοποιητή-αποκωδικοποιητή. Η ικανότητα του U-Net να διατηρεί χαρακτηριστικά υψηλής ανάλυσης μέσω της άμεσης σύνδεσης των στρωμάτων

κωδικοποιητή-αποκωδικοποιητή αποδεικνύεται καθοριστική για την ακριβή κατηγοριοποίηση ανά εικονοστοιχείο, ιδιαίτερα στα όρια μεταξύ διαφορετικών κλάσεων.

Παράλληλα, η χρήση των διογκομενων συνελίξεων στην αρχιτεκτονική του DeepLabV3+ επιτρέπει την επέκταση του οπτικού πεδίου χωρίς την αύξηση των παραμέτρων, βελτιώνοντας την γενικότερη σημασιολογική κατανόηση του μοντέλου. Ωστόσο, η έλλειψη άμεσων χωρικών πληροφοριών από τα αρχικά στρώματα του κωδικοποιητή περιορίζει την ακρίβεια της λεπτομερής προσαρμογής σε σημασιολογικές εργασίες.

7.2 Περιορισμοί της Έρευνας

Η παρούσα μελέτη παρουσιάζει συγκεκριμένους περιορισμούς που πρέπει να ληφθούν υπόψη κατά την ερμηνεία των αποτελεσμάτων.

Η εκπαίδευση περιορίστηκε σε 15 εποχές, γεγονός που ενδέχεται να μην επιτρέπει την πλήρη σύγκλιση των μοντέλων. Ενώ τα αποτελέσματα δείχνουν σταθεροποίηση μετά την 10η-11η εποχή, επιπλέον εποχές εκπαίδευσης με learning rate scheduling και early stopping criteria θα μπορούσαν να βελτιώσουν περαιτέρω την απόδοση.

Η χρήση αποκλειστικά του συνόλου δεδομένων Cityscapes περιορίζει τη γενικευσιμότητα των συμπερασμάτων. Το Cityscapes, εστιάζει σε αστικές σκηνές από Ευρωπαϊκές πόλεις, επομένως η απόδοση των μοντέλων σε διαφορετικά γεωγραφικά περιβάλλοντα, καιρικές συνθήκες ή τύπους σκηνών παραμένει αδιευκρίνιστη.

Η κλιμάκωση των εικόνων από την αρχική ανάλυση 2048×1024 σε 256×256 εικονοστοιχεία, προκαλεί απώλεια λεπτομερειών, ιδιαίτερα για μικρά αντικείμενα. Αυτή η μείωση της ανάλυσης επιλέχθηκε λόγω υπολογιστικών περιορισμών, αλλά επηρεάζει αρνητικά την ακρίβεια σε κλάσεις όπως ”Human” και ”Object”.

7.3 Μελλοντικές επεκτάσεις

Βάσει των ευρημάτων και περιορισμών της παρούσας μελέτης, προτείνονται οι ακόλουθες κατευθύνσεις για μελλοντική έρευνα.

7.3.1 Επέκταση Συνόλων Δεδομένων

Η επέκταση της αξιολόγησης σε πολλαπλά σύνολα δεδομένα αποτελεί κρίσιμη προτεραιότητα για την αξιολόγηση της γενικευσιμότητας των μοντέλων. Συγκεκριμένα, προτείνεται η δοκιμή των μοντέλων στο ADE20K, το οποίο περιέχει 150 κλάσεις και πάνω από 20,000 εικόνες από ποικίλα περιβάλλοντα συμπεριλαμβανομένων εσωτερικών και εξωτερικών χώρων [78]. Επιπρόσθετα, το PASCAL VOC 2012 αποτελεί κλασικό σημείο αναφοράς με 21 κλάσεις για

συγκριτική αξιολόγηση [22]. Αυτή η αξιολόγηση σε πολλαπλά σύνολα δεδομένων, θα αποκαλύψει καλύτερα την ικανότητα γενίκευσης των μοντέλων και θα προσδιορίσει τις αδυναμίες τους σε διαφορετικές εργασίες σημασιολογικής τμηματοποίησης.

7.4 Επιστημονική Συμβολή

Η παρούσα διπλωματική εργασία συμβάλλει στην επιστημονική κοινότητα μέσω τριών κύριων διαστάσεων. Πρωτίστως, παρέχεται ολοκληρωμένη και αντικειμενική σύγκριση τριών θεμελιωδών αρχιτεκτονικών σημασιολογικής τμηματοποίησης υπό ταυτόσημες πειραματικές συνθήκες. Η συστηματική μεθοδολογία και η λεπτομερής τεκμηρίωση επιτρέπουν την αναπαραγωγή των αποτελεσμάτων και την αξιόπιστη σύγκριση με μελλοντικές μελέτες.

Δευτερευόντως, η λεπτομερής ανάλυση των επιδόσεων ανά κλάση αποκαλύπτει τις δυνατότητες και αδυναμίες κάθε αρχιτεκτονικής σε διαφορετικούς τύπους αντικειμένων. Αυτή η γνώση είναι ιδιαίτερα χρήσιμη για χρήστες που επιλέγουν μοντέλο βάσει των συγκεκριμένων απαιτήσεων της εφαρμογής τους.

Τέλος, παρέχονται συγκεκριμένες κατευθύνσεις για την επιλογή κατάλληλης αρχιτεκτονικής. Το U-Net συνιστάται για εφαρμογές που απαιτούν υψηλή ακρίβεια και λεπτομερή οριοθέτηση, το DeepLabV3+ προτείνεται για σκηνές όπου η ανάγκη για την καλύτερη κατανόηση του γενικού πλαισίου επιβάλλεται, ενώ το SegNet αποτελεί την ιδανική επιλογή για περιβάλλοντα με περιοριστικούς υπολογιστικούς πόρους όπου η διαχείρισης της μνήμης είναι κρίσιμη. Αυτές οι συστάσεις βασίζονται σε εμπειρικά δεδομένα και μπορούν να καθοδηγήσουν μελλοντικές έρευνες και βιομηχανικές εφαρμογές.

Αναφορές

[1] Kokeb A Abera, Kalehiwot Nega Manahiloh, and Mohammad Motalleb Nejad. The effectiveness of global thresholding techniques in segmenting two-phase porous media. *Construction and Building Materials*, 142:256–267, 2017.

[2] Muhammad Fuad Al Haris, I Ketut Eddy Purnama, Eko Mulyanto Yuniarno, and Heni Fatmawati. An optimized u-net encoder with vgg-16 and additional feature block for brain tumor segmentation on mri adc images. In *2024 International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM)*, pages 1–6. IEEE, 2024.

[3] Saad Albawi, Tareq Abed Mohammed, and Saad Al-Zawi. Understanding of a convolutional neural network. In *2017 International Conference on Engineering and Technology (ICET)*, pages 1–6, 2017.

[4] Tanmay Anand, Soumendu Sinha, Murari Mandal, Vinay Chamola, and Fei Richard Yu. Agrisegnet: Deep aerial semantic segmentation framework for iot-assisted precision agriculture. *IEEE Sensors Journal*, 21(16):17581–17590, 2021.

[5] Khadijeh Askaripour and Arkadiusz Zak. Breast mri segmentation by deep learning: Key gaps and challenges. *IEEE Access*, 11:117935–117946, 2023.

[6] Taiwo Oladipupo Ayodele. Types of machine learning algorithms. *New advances in machine learning*, 3(19-48):5–1, 2010.

[7] Reza Azad, Ehsan Khodapanah Aghdam, Amelie Rauland, Yiwei Jia, Atlas Haddadi Avval, Afshin Bozorgpour, Sanaz Karimijafarbigloo, Joseph Paul Cohen, Ehsan Adeli, and Dorit Merhof. Medical image segmentation review: The success of u-net. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

[8] Reza Azad, Moein Heidary, Kadir Yilmaz, Michael Hüttemann, Sanaz Karimijafarbigloo, Yuli Wu, Anke Schmeink, and Dorit Merhof. Loss functions in the era of semantic segmentation: A survey and outlook, 2023.

[9] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12):2481–2495, 2017.

[10] Sunil L Bangare, Amruta Dubal, Pallavi S Bangare, and Suhas Patil. Reviewing otsu's method for image thresholding. *International Journal of Applied Engineering Research*, 10(9):21777–21783, 2015.

[11] Dor Bank, Noam Koenigstein, and Raja Giryes. Autoencoders. *Machine learning for data science handbook: data mining and knowledge discovery handbook*, pages 353–374, 2023.

[12] James Bergstra and Yoshua Bengio. Random search for hyper-parameter optimization. *The journal of machine learning research*, 13(1):281–305, 2012.

[13] Christopher M Bishop and Nasser M Nasrabadi. *Pattern recognition and machine learning*, volume 4. Springer, 2006.

[14] Nils Bjorck, Carla P Gomes, Bart Selman, and Kilian Q Weinberger. Understanding batch normalization. *Advances in neural information processing systems*, 31, 2018.

[15] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille. Semantic image segmentation with deep convolutional nets and fully connected crfs, 2016.

[16] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs, 2017.

[17] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation, 2017.

[18] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation, 2018.

[19] Lijun Ding and Ardeshir Goshtasby. On the canny edge detector. *Pattern recognition*, 34(3):721–725, 2001.

[20] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale, 2021.

[21] Busra Emek Soylu, Mehmet Serdar Guzel, Gazi Erkan Bostanci, Fatih Ekinci, Tunc Asuroglu, and Koray Acici. Deep-learning-based approaches for semantic segmentation of natural scene images: A review. *Electronics*, 12(12):2730, 2023.

[22] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results. http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html.

[23] Mulham Fawakherji, Ali Youssef, Domenico Bloisi, Alberto Pretto, and Daniele Nardi. Crop and weeds classification for precision agriculture using context-independent pixel-wise segmentation. In *2019 Third IEEE International Conference on Robotic Computing (IRC)*, pages 146–152, 2019.

[24] Orazio Gambino, Salvatore Vitabile, Giuseppe Lo Re, Giuseppe La Tona, Santino Librizzi, Roberto Pirrone, Edoardo Ardizzone, and Massimo Midiri. Automatic volumetric liver segmentation using texture based region growing. In *2010 International Conference on Complex, Intelligent and Software Intensive Systems*, pages 146–152, 2010.

[25] Alberto Garcia-Garcia, Sergio Orts-Escolano, Sergiu Oprea, Victor Villena-Martinez, and Jose Garcia-Rodriguez. A review on deep learning techniques applied to semantic segmentation. *arXiv preprint arXiv:1704.06857*, 2017.

[26] Ta Yang Goh, Shafriza Nisha Basah, Haniza Yazid, Muhammad Juhairi Aziz Safar, and Fathinul Syahir Ahmad Saad. Performance analysis of image thresholding: Otsu technique. *Measurement*, 114:298–307, 2018.

[27] Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume 1. MIT press Cambridge, 2016.

[28] Stephen Grossberg. Recurrent neural networks. *Scholarpedia*, 8(2):1888, 2013.

[29] Jiuxiang Gu, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu, Xingxing Wang, Gang Wang, Jianfei Cai, et al. Recent advances in convolutional neural networks. *Pattern recognition*, 77:354–377, 2018.

[30] Michael Haenlein and Andreas Kaplan. A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. *California Management Review*, 61:000812561986492, 07 2019.

[31] Kai Han, Yunhe Wang, Hanting Chen, Xinghao Chen, Jianyuan Guo, Zhenhua Liu, Yehui Tang, An Xiao, Chunjing Xu, Yixing Xu, Zhaohui Yang, Yiman Zhang, and Dacheng Tao. A survey on vision transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PP:1–1, 2020.

[32] Trevor Hastie, Robert Tibshirani, Jerome Friedman, et al. The elements of statistical learning, 2009.

[33] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[34] Timothy O Hodson. Root mean square error (rmse) or mean absolute error (mae): When to use them or not. *Geoscientific Model Development Discussions*, 2022:1–10, 2022.

[35] Pavel Iakubovskii. Segmentation models pytorch. `https://github.com/qubvel/segmentation_models.pytorch`, 2019.

[36] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. pmlr, 2015.

[37] H Jabbar and Rafiqul Zaman Khan. Methods to avoid over-fitting and under-fitting in supervised machine learning (comparative study). *Computer science, communication and instrumentation devices*, 70(10.3850):978–981, 2015.

[38] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.

[39] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[40] Dimitrios Loukatos, Charalampos Templalexis, Diamanto Lentzou, Georgios Xanthopoulos, and Konstantinos G. Arvanitis. Enhancing a flexible robotic spraying platform for distant plant inspection via high-quality thermal imagery data. *Computers and Electronics in Agriculture*, 190:106462, 2021.

[41] T Soni Madhulatha. An overview on clustering methods. *arXiv preprint arXiv:1205.1117*, 2012.

[42] Warren S McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, 1943.

[43] Fernand Meyer and Serge Beucher. Morphological segmentation. *Journal of visual communication and image representation*, 1(1):21–46, 1990.

[44] Shervin Minaee, Yuri Boykov, Fatih Porikli, Antonio Plaza, Nasser Kehtarnavaz, and Demetri Terzopoulos. Image segmentation using deep learning: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 44(7):3523–3542, 2021.

[45] Dominik Müller, Iñaki Soto-Rey, and Frank Kramer. Towards a guideline for evaluation metrics in medical image segmentation, 2022.

[46] Keiron O'shea and Ryan Nash. An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*, 2015.

[47] Shivam K Panda, Yongkyu Lee, and M Khalid Jawed. Agronav: Autonomous navigation framework for agricultural robots and vehicles using semantic segmentation and semantic line detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6272–6281, 2023.

[48] Rajendra Patil and Renu Aggarwal. Comprehensive review on image segmentation applications. *Available at SSRN 5227277*, 2023.

[49] Gasper Podobnik and Tomaz Vrtovec. Understanding implementation pitfalls of distance-based metrics for image segmentation, 2025.

[50] Marius-Constantin Popescu, Valentina E Balas, Liliana Perescu-Popescu, and Nikos Mastorakis. Multilayer perceptron and neural networks. *WSEAS Transactions on Circuits and Systems*, 8(7):579–588, 2009.

[51] David MW Powers. Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. *arXiv preprint arXiv:2010.16061*, 2020.

[52] Rama Rani, Sukhjeet Kaur Ranade, and Chandan Singh. The multimodal mri brain tumor segmentation using modified connected u-net with a guided decoder. In *2024 4th International Conference on Mobile Networks and Wireless Communications (ICMNWC)*, pages 1–5. IEEE, 2024.

[53] Nicholas G Reich, Justin Lessler, Krzysztof Sakrejda, Stephen A Lauer, Sopon Iamsirithaworn, and Derek AT Cummings. Case study in evaluating time series prediction models using the relative mean absolute error. *The American Statistician*, 70(3):285–292, 2016.

[54] Intisar Rizwan I Haque and Jeremiah Neubert. Deep learning approaches to biomedical image segmentation. *Informatics in Medicine Unlocked*, 18:100297, 2020.

[55] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[56] Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386, 1958.

[57] Takaya Saito and Marc Rehmsmeier. The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. *PloS one*, 10(3):e0118432, 2015.

[58] Yutaka Sasaki. The truth of the f-measure. *Teach Tutor Mater*, 01 2007.

[59] Robin M Schmidt. Recurrent neural networks (rnns): A gentle introduction and overview. *arXiv preprint arXiv:1912.05911*, 2019.

[60] Jean Serra. *Image analysis and mathematical morphology*. Academic Press, Inc., 1983.

[61] Alex Sherstinsky. Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network. *Physica D: Nonlinear Phenomena*, 404:132306, 2020.

[62] Nahian Siddique, Paheding Sidike, Colin Elkin, and Vijay Devabhaktuni. U-net and its variants for medical image segmentation: theory and applications. *arXiv preprint arXiv:2011.01118*, 2020.

[63] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[64] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.

[65] Ralf C Staudemeyer and Eric Rothstein Morris. Understanding lstm–a tutorial into long short-term memory recurrent neural networks. *arXiv preprint arXiv:1909.09586*, 2019.

[66] Abdel Aziz Taha and Allan Hanbury. An efficient algorithm for calculating the exact hausdorff distance. *IEEE transactions on pattern analysis and machine intelligence*, 37(11):2153–2163, 2015.

[67] Manoj K Vairalkar and SU Nimbhorkar. Edge detection of images using sobel operator. *Int. J. Emerg. Technol. Adv. Eng*, 2(1):291–293, 2012.

[68] S Vani and TV Madhusudhana Rao. An experimental approach towards the performance assessment of various optimizers on convolutional neural network. In *2019 3rd international conference on trends in electronics and informatics (ICOEI)*, pages 331–336. IEEE, 2019.

[69] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2023.

[70] Biao Wang and Chunfeng Yang. Liver tumor segmentation method based on u-net architecture: a review. *EAI Endorsed Transactions on e-Learning*, 10, 2024.

[71] Frauke Wilm, Jonas Ammeling, Mathias Öttl, Rutger HJ Fick, Marc Aubreville, and Katharina Breininger. Rethinking u-net skip connections for biomedical image segmentation. *arXiv preprint arXiv:2402.08276*, 2024.

[72] Yan Xu, Rixiang Quan, Weiting Xu, Yi Huang, Xiaolong Chen, and Fengyuan Liu. Advances in medical image segmentation: A comprehensive review of traditional, deep learning and hybrid approaches. *Bioengineering*, 11(10):1034, 2024.

[73] M Yeung, E Sala, CB Schönlieb, and L Rundo. Unified focal loss: generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation. comput med imaging gr 95: 102026, 2021.

[74] Wei Yuan, Jin Wang, and Wenbo Xu. Shift pooling pspnet: rethinking pspnet for building extraction in remote sensing images from entire local feature pooling. *Remote sensing*, 14(19):4889, 2022.

[75] Yujian Yuan, Lina Yang, Kan Chang, Youju Huang, Haoyan Yang, and Jiale Wang. Dsca-pspnet: Dynamic spatial-channel attention pyramid scene parsing network for sugarcane field segmentation in satellite imagery. *Frontiers in Plant Science*, 14:1324491, 2024.

[76] Chaohui Zhang, Anusha Achuthan, and Galib Muhammad Shahriar Himel. State-of-the-art and challenges in pancreatic ct segmentation: a systematic review of u-net and its variants. *IEEE Access*, 12:78726–78742, 2024.

[77] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2881–2890, 2017.

[78] Bolei Zhou, Hang Zhao, Xavier Puig, Sanja Fidler, Adela Barriuso, and Antonio Torralba. Scene parsing through ade20k dataset. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 633–641, 2017.

[79] Djemel Ziou and Salvatore Tabbone. Edge detection techniques-an overview. *Распознавание образов и анализ изображен/Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications*, 8(4):537–559, 1998.