

***Εφαρμογή και αξιολόγηση του
μοντέλου 3D U-Net
για την κατάτμηση εικόνων αξονικής
τομογραφίας***

Θεανώ-Ειρήνη Κεβρεκίδου

Μεταπτυχιακή Εργασία Εξειδίκευσης



Ιωάννινα, Ιούλιος 2025



ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ Η/Υ & ΠΛΗΡΟΦΟΡΙΚΗΣ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF IOANNINA

Εφαρμογή και αξιολόγηση του μοντέλου 3D U-Net για την κατάρτιση εικόνων αξονικής τομογραφίας

Η Μεταπτυχιακή Διπλωματική Εργασία

υποβάλλεται στην ορισθείσα

από τη Συνέλευση

του Τμήματος Μηχανικών Η/Υ και Πληροφορικής

Εξεταστική Επιτροπή

από την

Θεανώ-Ειρήνη Κεβρεκίδου

ως μέρος των υποχρεώσεων για την απόκτηση του

ΔΙΠΛΩΜΑΤΟΣ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΣΤΗ ΜΗΧΑΝΙΚΗ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΥΠΟΛΟΓΙΣΤΙΚΩΝ
ΣΥΣΤΗΜΑΤΩΝ

ΜΕ ΕΙΔΙΚΕΥΣΗ
ΣΤΗΝ ΕΠΙΣΤΗΜΗ ΚΑΙ ΜΗΧΑΝΙΚΗ ΔΕΔΟΜΕΝΩΝ

Πανεπιστήμιο Ιωαννίνων

Πολυτεχνική Σχολή

Ιωάννινα 2025

Εξεταστική Επιτροπή:

- **Νίκου Χριστόφορος**, Καθηγητής, Τμήμα Μηχανικών Η/Υ και Πληροφορικής, Πανεπιστήμιο Ιωαννίνων (Επιβλέπων)
- **Κόντης Λυσίμαχος-Πάυλος**, Καθηγητής, Τμήμα Μηχανικών Η/Υ και Πληροφορικής, Πανεπιστήμιο Ιωαννίνων
- **Μανής Γεώργιος**, Αναπληρωτής Καθηγητής, Τμήμα Μηχανικών Η/Υ και Πληροφορικής, Πανεπιστήμιο Ιωαννίνων

ΑΦΙΕΡΩΣΗ

Αφιερώνω τη διπλωματική αυτή εργασία στην οικογένειά μου και ιδιαίτερα στη μητέρα μου, Ευγενία.

ΕΥΧΑΡΙΣΤΙΕΣ

Ολοκληρώνοντας τον κύκλο των σπουδών μου με την παρούσα διπλωματική εργασία, η οποία εκπονήθηκε στο πλαίσιο του μεταπτυχιακού προγράμματος σπουδών «Μηχανική Δεδομένων και Υπολογιστικών Συστημάτων» του Τμήματος Μηχανικών Ηλεκτρονικών Υπολογιστών και Πληροφορικής, στην Πολυτεχνική Σχολή του Πανεπιστημίου Ιωαννίνων, θα ήθελα να εκφράσω τις ειλικρινείς μου ευχαριστίες σε όλους όσους με στήριξαν ουσιαστικά και ηθικά καθ' όλη τη διάρκεια των σπουδών μου.

Πρώτα απ' όλα, θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα καθηγητή μου, κ. Χριστόφορο Νίκου, Καθηγητή του Τμήματος Μηχανικών Ηλεκτρονικών Υπολογιστών και Πληροφορικής, για την ανάθεση του θέματος, καθώς και για την πολύτιμη καθοδήγηση, τις χρήσιμες παρατηρήσεις και τη διαρκή υποστήριξή του κατά την εκπόνηση της εργασίας.

Θερμές ευχαριστίες απευθύνω επίσης στην κ. Μαρίνα Πλησίτη, για την ουσιαστική συμβολή και την πρόθυμη βοήθειά της καθ' όλη τη διάρκεια της διπλωματικής μου. Η υποστήριξή της υπήρξε πολύτιμη σε όλα τα στάδια της υλοποίησης.

Τέλος, θα ήθελα να εκφράσω την ευγνωμοσύνη μου στην οικογένειά μου και στους φίλους μου, για την αμέριστη ψυχολογική στήριξη και κατανόηση που μου παρείχαν καθ' όλη τη διάρκεια των σπουδών μου.

ΠΕΡΙΕΧΟΜΕΝΑ

Κατάλογος Σχημάτων	iii
Κατάλογος Πινάκων	iv
Περίληψη	v
Extended Abstract	vii
1 Εισαγωγή	1
1.1 Ιατρική απεικόνιση και ανάγκη για αυτόματη τμηματοποίηση	1
1.2 Το πρόβλημα της τμηματοποίησης σε δεδομένα KiTS19	2
1.3 Νεότερες μέθοδοι και προσεγγίσεις τμηματοποίησης	3
1.4 Σκοπός και στόχοι της παρούσας εργασίας	4
2 Μηχανική Μάθηση και Βαθιά Νευρωνικά Δίκτυα	6
2.1 Μηχανική μάθηση	6
2.1.1 Επιβλεπόμενη Μάθηση (Supervised Learning)	8
2.1.2 Νευρωνικά Δίκτυα	9
2.1.3 Βαθιά Μάθηση (Deep Learning)	12
2.1.4 Συνελικτικό Επίπεδο	13
2.1.5 Συνάρτηση Ενεργοποίησης (Activation Function)	15
2.2 Χρήση Strided Convolution για Υποδειγματοληψία	16
2.2.1 Επίπεδο Εξόδου	17
2.2.2 Βελτιστοποιητές	17
2.2.3 Μέθοδος Οπισθοδιάδοσης Σφάλματος (Backpropagation)	19
2.2.4 Συναρτήσεις Απώλειας για Εκπαίδευση	21
2.2.5 Μέθοδος Επιπέδων (Level Set) και LSF Loss	22

2.2.6	Εισαγωγή στην Geodesic Active Contour και τη δική μας προ- σέγγιση	25
2.2.7	Μετρικές Αξιολόγησης (Evaluation Metrics)	28
2.2.8	Εκπαίδευση (Training)	29
3	Μεθοδολογία: Αλγόριθμος και Αρχιτεκτονικές Μοντέλων	31
3.1	Σύνολο Δεδομένων και Προεπεξεργασία	31
3.1.1	Σύνολο Δεδομένων	31
3.1.2	Προεπεξεργασία Δεδομένων	32
3.2	Αρχιτεκτονική Μοντέλου	36
3.2.1	Κωδικοποιητής (Encoder)	37
3.2.2	Επίπεδο Συστολής (Bottleneck)	38
3.2.3	Αποκωδικοποιητής (Decoder)	38
3.2.4	Διάγραμμα Αρχιτεκτονικής 3D U-Net	39
3.2.5	Συναρτήσεις Κόστους (Loss Functions)	39
4	Ανάλυση Πειραμάτων	41
4.1	Υπολογιστικό Περιβάλλον	41
4.2	Ανάλυση Μεθόδου	42
4.3	Διαδικασία Εκπαίδευσης Αλγορίθμων	43
4.3.1	Αξιολόγηση Απόδοσης	43
5	Αποτελέσματα	44
5.1	Αποτελέσματα και Συζήτηση	44
5.1.1	Συνάρτηση κόστους εκπαίδευσης	45
5.1.2	Συνάρτηση κόστους επικύρωσης	46
5.1.3	Δείκτης Dice για υπόβαθρο(background)	47
5.1.4	Δείκτης Dice για το νεφρό	48
5.1.5	Δείκτης Dice για τον όγκο	49
5.1.6	Δείκτης Dice για Kidney + Tumor	50
5.2	Ποσοτική Σύγκριση των Μεθόδων	50
5.3	Συγκριτική Οπτικοποίηση Αποτελεσμάτων	51
6	Συμπεράσματα και Μελλοντική Εργασία	54
	Βιβλιογραφία	56

ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ

2.1	Ενδεικτική αρχιτεκτονική τεχνητού νευρωνικού δικτύου με επίπεδο εισόδου, κρυφά επίπεδα και επίπεδο εξόδου.	10
2.2	Παράδειγμα πράξης συνέλιξης (convolution) με φίλτρο 3×3 σε εικόνα 4×4	14
3.1	Παράδειγμα 10 δισδιάστατων τομών CT νεφρού από το KiTS19 dataset από διαφορετικά cases.	32
3.2	Διάγραμμα της αρχιτεκτονικής 3D U-Net με τις διαστάσεις των feature maps και τις skip connections.	39
5.1	Πορεία του συνολικού training loss κατά τη διάρκεια της εκπαίδευσης για τα δύο μοντέλα.	45
5.2	Συνολική πορεία του Validation Loss για τα δύο μοντέλα.	46
5.3	Dice score στην περιοχή υποβάθρου (background) στο validation set. .	47
5.4	Εξέλιξη του Dice Score για την τμηματοποίηση του νεφρού κατά την επικύρωση.	48
5.5	Dice Score στην περιοχή του όγκου.	49
5.6	Dice score στο σύνολο των κατηγοριών νεφρό και όγκος στο validation set.	50
5.7	Σύγκριση των διαφορετικών loss function του 3D U-Net με βάση την αρχική εικόνα : (α) CT Image, (β) Ground Truth, (γ) CE + DICE Loss, (δ) LSF + GAC Loss	52

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

5.1 Δείκτες Dice ανά κλάση για τα δύο μοντέλα	51
---	----

ΠΕΡΙΛΗΨΗ

Θεανώ-Ειρήνη Κεβρεκίδου, Δ.Μ.Σ. στη Μηχανική Δεδομένων και Υπολογιστικών Συστημάτων, Τμήμα Μηχανικών Η/Υ και Πληροφορικής, Πολυτεχνική Σχολή, Πανεπιστήμιο Ιωαννίνων, 2025.

Εφαρμογή και αξιολόγηση του μοντέλου 3D U-Net για την κατάτμηση εικόνων αξονικής τομογραφίας.

Επιβλέπων: Νίκου Χριστόφορος, Καθηγητής.

Η τμηματοποίηση ιατρικών εικόνων αποτελεί κρίσιμη πρόκληση στη σύγχρονη βιοϊατρική ανάλυση, απαιτώντας ακριβή διαχωρισμό ανατομικών δομών σε δεδομένα υψηλής διάστασης και πολυπλοκότητας. Η εφαρμογή νευρωνικών δικτύων είναι αναγκαία για την αυτοματοποιημένη και αποτελεσματική τμηματοποίηση, λόγω της ικανότητας τους να μαθαίνουν σύνθετα χωρικά πρότυπα από πολυδιάστατα σύνολα δεδομένων. Παρά την επιτυχία τους, οι βασικές συναρτήσεις απώλειας (loss functions) όπως η Cross-Entropy και η Dice Loss, αν και αποδοτικά, δεν εκμεταλλεύονται πάντα πλήρως τις μορφολογικές ιδιότητες, περιορίζοντας την ομαλότητα και προσέγγιση των ορίων. Για το λόγο αυτό, πρόσφατες προσεγγίσεις έχουν εισαγάγει μορφολογικούς περιορισμούς μέσω ειδικών loss functions όπως οι Level Set Loss και Geodesic Active Contour Loss.

Στην παρούσα εργασία μελετήθηκε η εφαρμογή και συγκριτική αξιολόγηση του 3D U-Net σε προβλήματα τρισδιάστατης τμηματοποίησης ιατρικών εικόνων, με ειδική εστίαση στο δημόσιο dataset KiTS19 ένα από τα πιο απαιτητικά datasets που αφορά segmentation νεφρού και όγκου από αξονικές τομογραφίες. Για την αντιμετώπιση αυτού του προβλήματος υλοποιήθηκαν και συγκρίθηκαν δύο παραλλαγές του μοντέλου 3D U-Net. Η βασική παραλλαγή αξιοποιεί τις κλασικές συναρτήσεις κόστους Cross-Entropy και Dice, ενώ η προτεινόμενη εκδοχή ενσωματώνει επιπλέον μία παραλλαγή των συναρτήσεων Level Set Loss (LSF) και Geodesic Active Contour Loss (GAC). Οι δύο αυτές συναρτήσεις αποδείχθηκε ότι βοήθησαν στην πιο

ομαλή και ακριβή τμηματοποίηση. Συγκεκριμένα, η Level Set Loss (LSF) διαμορφώνεται ώστε το δίκτυο να παράγει απευθείας συναρτήσεις απόστασης (*signed distance functions*) για κάθε κλάση αντί απλών πιθανοτήτων, ενθαρρύνοντας τον σαφή διαχωρισμό προσκηνίου και φόντου και βελτιώνοντας την ευκρίνεια των ορίων. Η Geodesic Active Contour Loss (GAC) αποτελεί μια απλουστευμένη αλλά αποτελεσματική υλοποίηση της κλασικής ενεργειακής μεθόδου των ενεργών περιγραμμάτων (*active contours*), συνδυάζοντας δύο όρους: έναν όρο μήκους καμπύλης και έναν όρο περιοχής.

Η αξιολόγηση των μοντέλων διεξήχθη με βάση τον δείκτη Dice και τις αντίστοιχες καμπύλες εκπαίδευσης/επικύρωσης για κάθε κατηγορία (υπόβαθρο, νεφρό, όγκος). Τα αποτελέσματα αποκάλυψαν την προφανή υπεροχή του εμπλουτισμένου μοντέλου (DICE+CE+LSF+GAC), το οποίο πέτυχε ανώτερες τιμές Dice για τις απαιτητικές κατηγορίες νεφρού και όγκου. Συμπερασματικά, η εργασία αναδεικνύει τη σημασία της επιλογής κατάλληλου πολυπαραγοντικού loss στην τμηματοποίηση ιατρικών εικόνων. Με βάση τα παραπάνω, η μελλοντική έρευνα θα μπορούσε να επικεντρωθεί στη βελτίωση των loss functions και το εμπλουτισμό των δεδομένων. Η μελέτη αυτή αποδεικνύει ότι η ενσωμάτωση προχωρημένων συναρτήσεων απώλειας σε αρχιτεκτονικές 3D U-Net μπορεί να προσδώσει ουσιαστική βελτίωση στην ακρίβεια και τη γενικευσιμότητα των ιατρικών τμηματοποιήσεων.

EXTENDED ABSTRACT

Medical image segmentation is a critical challenge in modern biomedical analysis, which requires accurate separation of anatomical structures in complex, high-dimensional data. The application of neural networks is essential for automated and efficient segmentation, due to their ability to learn complex spatial patterns from multidimensional datasets. Despite their success, basic loss functions such as Cross-Entropy and Dice Loss, while effective, do not always fully exploit morphological properties, limiting the smoothness and accuracy of boundaries. Consequently, newer methodologies have included morphological limitations via specific loss functions, such as Level Set Loss and Geodesic Active Contour Loss.

This study examined the implementation and comparative assessment of the 3D U-Net model for 3D medical image segmentation tasks, specifically targeting the public KiTS19 dataset, which is recognized as one of the most challenging datasets for kidney and tumor segmentation from computed tomography (CT) scans. Two variants of the 3D U-Net model were developed and evaluated to tackle this issue. The baseline variation employed traditional loss functions, namely Cross-Entropy and Dice Loss, while the proposed version integrated supplementary types of Level Set Loss (LSF) and Geodesic Active Contour Loss (GAC). These two loss functions exhibited enhancements in generating smoother and more precise segmentations. Specifically, the Level Set Loss (LSF) is formulated to enable the network to directly produce signed distance functions for each class instead of simple probabilities, encouraging clear foreground-background separation and enhancing boundary sharpness. The Geodesic Active Contour Loss (GAC) is a simplified yet effective implementation of the classical active contour energy method, combining two terms: a length term and a region term.

The models were evaluated based on the Dice coefficient and corresponding training/validation curves for each category (background, kidney, tumor). The results re-

vealed the clear superiority of the enhanced model (DICE+CE+LSF+GAC), which achieved higher Dice scores for the challenging kidney and tumor categories. This study underscores the significance of selecting the appropriate combination of loss functions for medical image segmentation. Employing sophisticated loss functions that use both shape and regional information enhances the model's ability to generate more precise and smoother borders. This is particularly crucial when addressing complex circumstances such as renal and tumor segmentation. Based on these findings, future research could focus on improving loss functions and enriching the dataset. This study demonstrates that integrating advanced loss functions into 3D U-Net architectures can provide significant improvements in accuracy and generalizability for medical image segmentation.

ΚΕΦΑΛΑΙΟ 1

ΕΙΣΑΓΩΓΗ

-
- 1.1 Ιατρική απεικόνιση και ανάγκη για αυτόματη τμηματοποίηση
 - 1.2 Το πρόβλημα της τμηματοποίησης σε δεδομένα KiTS19
 - 1.3 Νεότερες μέθοδοι και προσεγγίσεις τμηματοποίησης
 - 1.4 Σκοπός και στόχοι της παρούσας εργασίας
-

Στο πρώτο κεφάλαιο αναλύονται οι θεμελιώδεις έννοιες της τμηματοποίησης εικόνων στην ιατρική απεικόνιση, καθώς και οι ευρέως χρησιμοποιούμενες μέθοδοι για αυτή τη διαδικασία. Αναφέρεται η ανάγκη για αυτοματοποιημένη και ακριβή τμηματοποίηση ιατρικών εικόνων, εστιάζοντας στην εφαρμογή νευρωνικών δικτύων τύπου U-Net, ιδιαιτέρως στην τρισδιάστατη μορφή τους (3D U-Net). Εξηγείται επίσης η σημασία της βάσης δεδομένων KiTS19, οι ιδιαιτερότητες της, και οι προκλήσεις κατά την τμηματοποίηση νεφρών και όγκων σε εικόνες CT.

1.1 Ιατρική απεικόνιση και ανάγκη για αυτόματη τμηματοποίηση

Η ιατρική απεικόνιση αποτελεί ακρογωνιαίο λίθο της σύγχρονης διάγνωσης και θεραπευτικής αντιμετώπισης παθήσεων, επιτρέποντας τη μη επεμβατική παρατήρηση και αξιολόγηση της εσωτερικής ανατομίας και λειτουργίας του ανθρώπινου σώματος. Μέθοδοι όπως η αξονική τομογραφία (CT), η μαγνητική τομογραφία (MRI) και η υπερηχοτομογραφία (Ultrasound) παράγουν πλούσια δεδομένα εικόνες, τα οποία

όμως απαιτούν λεπτομερή ανάλυση και ερμηνεία από εξειδικευμένους ιατρούς. Η τμηματοποίηση εικόνας (image segmentation) είναι η διαδικασία κατά την οποία απομονώνονται και αναγνωρίζονται περιοχές ενδιαφέροντος εντός της εικόνας, όπως είναι τα όργανα, οι ιστοί ή οι όγκοι. Όπως ορίζεται από τους Litjens et al. (2017)[1], η τμηματοποίηση περιλαμβάνει την ανάθεση κάθε pixel ή voxel μιας ιατρικής εικόνας σε μία συγκεκριμένη κατηγορία ιστού ή ανατομικής δομής. Η διαδικασία αυτή είναι κρίσιμη για την εξαγωγή ποσοτικών πληροφοριών, τη διάγνωση, την παρακολούθηση της εξέλιξης παθήσεων και τον σχεδιασμό θεραπευτικών προσεγγίσεων.

Παραδοσιακά, η τμηματοποίηση εκτελείται χειροκίνητα από ακτινολόγους ή άλλους επαγγελματίες υγείας, γεγονός που καθιστά τη διαδικασία χρονοβόρα, απαιτητική και επιρρεπή σε σφάλματα ή υποκειμενικές εκτιμήσεις. Η συνεχώς αυξανόμενη παραγωγή ιατρικών δεδομένων σε συνδυασμό με την ανάγκη για ακριβή, επαναλήψιμα και αποδοτικά αποτελέσματα καθιστά επιτακτική την ανάπτυξη αυτόματων μεθόδων. Τα τελευταία χρόνια, παρατηρείται αυξημένη ερευνητική δραστηριότητα σχετικά με την εφαρμογή τεχνικών βαθιάς μάθησης στην ιατρική εικόνα.

Ειδικότερα, τα βαθιά νευρωνικά δίκτυα (CNNs) έχουν αποδειχθεί ιδιαίτερα αποτελεσματικοί για προβλήματα ταξινόμησης και τμηματοποίησης εικόνων, με τα δίκτυα τύπου U-Net να βρίσκονται στο επίκεντρο αυτών των προσεγγίσεων. Η αρχιτεκτονική 3D U-Net [2], που προσαρμόζει το κλασικό U-Net για ογκομετρικά δεδομένα, αποτελεί μια από τις πλέον διαδεδομένες επιλογές για την τμηματοποίηση σε δεδομένα τριών διαστάσεων, όπως οι εικόνες από CT τομογραφίες. Η υψηλή χωρική συνοχή που προσφέρει, σε συνδυασμό με την ικανότητά της να διαχειρίζεται μεγάλους όγκους δεδομένων, καθιστά την 3D U-Net ιδανική για την επίλυση του προβλήματος της αυτόματης τμηματοποίησης οργάνων και παθολογικών περιοχών σε ιατρικές εικόνες.

1.2 Το πρόβλημα της τμηματοποίησης σε δεδομένα KiTS19

Η ακριβής τμηματοποίηση των νεφρών και των όγκων τους σε απεικονιστικά δεδομένα αποτελεί πρόκληση ιδιαίτερης σημασίας για την ουρολογική ογκολογία και τη χειρουργική σχεδίαση. Η αυτόματη εξαγωγή των ορίων αυτών των ανατομικών δομών συμβάλλει καθοριστικά τόσο στην αξιολόγηση της σοβαρότητας της νόσου όσο και στον προγραμματισμό επεμβάσεων, όπως η μερική ή ολική νεφρεκτομή.

Παρ' όλα αυτά, η ποικιλομορφία στη μορφολογία των όγκων, οι διαφορές στην ένταση των εικόνων CT και η ύπαρξη θορύβου ή artefacts καθιστούν το πρόβλημα εξαιρετικά απαιτητικό.

Η βάση δεδομένων KiTS19 (Kidney Tumor Segmentation 2019) σχεδιάστηκε στο πλαίσιο διεθνούς διαγωνισμού με σκοπό την αξιολόγηση αλγορίθμων αυτόματης τμηματοποίησης νεφρών και όγκων από αξονικές τομογραφίες [3]. Περιλαμβάνει 300 περιπτώσεις ασθενών με σχολιασμένες, voxel-wise, περιοχές των νεφρών και των αντίστοιχων όγκων, παρέχοντας ένα ποικιλόμορφο και κλινικά ρεαλιστικό σύνολο δεδομένων. Κάθε περίπτωση συνοδεύεται από τρισδιάστατες εικόνες CT και τις αντίστοιχες ετικέτες για την ανατομική περιοχή του νεφρού, καθώς και την παθολογική περιοχή του όγκου, διαμορφώνοντας ένα πλήρες πρόβλημα ιατρικής τμηματοποίησης [3].

Το σετ δεδομένων KiTS19 περιλαμβάνει περιπτώσεις με σημαντικές διαφορές στην ποιότητα της εικόνας, τη θέση, το μέγεθος και την υφή των όγκων. Οι δυσκολίες που προκύπτουν από αυτές τις διαφορές αντικατοπτρίζουν την πολυπλοκότητα της κλινικής πράξης και καθιστούν τη βάση δεδομένων ένα αυστηρό benchmark για την αξιολόγηση νέων μεθόδων. Επιπλέον, η ανάγκη για ακρίβεια όχι μόνο στη διάκριση νεφρού από φόντο, αλλά και στην απομόνωση του ίδιου του όγκου, επιβάλλει την υιοθέτηση πολύπλοκων αρχιτεκτονικών και συνδυαστικών loss functions ώστε να ενσωματωθεί η εννοιολογική, μορφολογική και τοπολογική πληροφορία κατά την εκπαίδευση του μοντέλου.

1.3 Νεότερες μέθοδοι και προσεγγίσεις τμηματοποίησης

Η εξέλιξη της υπολογιστικής όρασης τα τελευταία χρόνια έχει αναδείξει τη βαθιά μάθηση ως την κυρίαρχη προσέγγιση για την τμηματοποίηση ιατρικών εικόνων. Οι συνελκτικοί νευρωνικοί μηχανισμοί (Convolutional Neural Networks - CNNs) αποτελούν τη βάση πολλών σύγχρονων αρχιτεκτονικών, προσφέροντας εντυπωσιακή ικανότητα αναγνώρισης χωρικών προτύπων και λεπτομερειών σε δεδομένα υψηλής διάστασης. Ένα από τα πιο επιδραστικά μοντέλα στον τομέα αυτό είναι το U-Net, το οποίο αρχικά προτάθηκε για δισδιάστατη βιοϊατρική τμηματοποίηση και στη συνέχεια επεκτάθηκε σε τρισδιάστατες μορφές (3D U-Net) για την αντιμετώπιση ογκομετρικών δεδομένων, όπως οι εικόνες CT [4].

Ωστόσο, η απλή χρήση τυπικών loss functions, όπως η Dice Loss ή η Cross Entropy [5], δεν επαρκεί πάντα για την εκπαίδευση δικτύων σε προβλήματα όπου απαιτείται η ανάκτηση σύνθετων γεωμετρικών ή τοπολογικών χαρακτηριστικών. Στοχεύοντας σε μεγαλύτερη ακρίβεια και μορφολογική συνέπεια, έχουν προταθεί προηγμένες τεχνικές που ενσωματώνουν γνώσεις από γεωμετρικά μοντέλα, όπως τα Level Set Methods και τα Geodesic Active Contours (GAC). Οι μέθοδοι αυτές, αν και προέρχονται από το πεδίο της υπολογιστικής γεωμετρίας και της κλασικής επεξεργασίας εικόνας, μπορούν να ενσωματωθούν σε loss functions για την καθοδήγηση της εκπαίδευσης νευρωνικών δικτύων προς πιο συνεκτικά και ρεαλιστικά αποτελέσματα.

Η παρούσα εργασία αξιοποιεί αυτήν την προσέγγιση, συγκρίνοντας δύο παραλλαγές της ίδιας αρχιτεκτονικής 3D U-Net, οι οποίες διαφοροποιούνται αποκλειστικά ως προς τη συνάρτηση κόστους. Το πρώτο μοντέλο βασίζεται στον συνδυασμό Cross Entropy και Dice Loss, ενώ το δεύτερο επεκτείνεται με την ενσωμάτωση προτεινόμενων loss functions που προέρχονται από τις Level Set και Geodesic Active Contour μεθοδολογίες. Η συγκριτική μελέτη των δύο μοντέλων προσφέρει πολύτιμες ενδείξεις για την επίδραση της επιλογής της συνάρτησης απώλειας στην ποιότητα της παραγόμενης τμηματοποίησης.

1.4 Σκοπός και στόχοι της παρούσας εργασίας

Η παρούσα εργασία επικεντρώνεται στη μελέτη της επίδρασης διαφορετικών συναρτήσεων απώλειας (loss functions) στην απόδοση νευρωνικών δικτύων τύπου 3D U-Net για την τμηματοποίηση νεφρού και όγκου σε δεδομένα από αξονικές τομογραφίες της βάσης KiTS19. Στόχος είναι να αξιολογηθεί πώς η ενσωμάτωση γεωμετρικών πληροφοριών και μορφολογικών περιορισμών, μέσω προχωρημένων loss functions όπως οι Level Set και Geodesic Active Contour Loss, επηρεάζει την ακρίβεια και την ποιότητα των παραγόμενων αποτελεσμάτων.

Για τον σκοπό αυτό, υλοποιήθηκαν και εκπαιδεύτηκαν δύο μοντέλα που βασίζονται στην ίδια αρχιτεκτονική 3D U-Net. Το πρώτο μοντέλο χρησιμοποιεί ως loss function τον συνδυασμό Cross Entropy και Dice Loss, ενώ το δεύτερο εμπλουτίζεται με τις Level Set και GAC Loss, διαμορφώνοντας ένα πολυπαραγοντικό κριτήριο εκπαίδευσης που συνδυάζει γεωμετρικά χαρακτηριστικά. Η εκπαίδευση πραγμα-

τοποιοήθηκε σε περιβάλλον πενταπλής διασταυρούμενης επικύρωσης (5-fold cross-validation), και η αξιολόγηση βασίστηκε κυρίως στον δείκτη Dice, όπως συνηθίζεται σε προβλήματα τμηματοποίησης.

Ο βασικός σκοπός της εργασίας είναι η ανάδειξη της σημασίας της επιλογής της συνάρτησης κόστους στην απόδοση ενός συστήματος ιατρικής τμηματοποίησης. Παράλληλα, αναδεικνύεται η δυνατότητα ενσωμάτωσης παραδοσιακών γεωμετρικών μοντέλων (όπως τα Level Sets) σε πλαίσια σύγχρονης βαθιάς μάθησης, προσφέροντας έναν υβριδικό τρόπο βελτίωσης των αποτελεσμάτων. Η μελέτη αυτή φιλοδοξεί να συμβάλει στην κατεύθυνση της πιο ρεαλιστικής και αξιόπιστης τμηματοποίησης ιατρικών εικόνων μέσω βαθιάς μάθησης.

ΚΕΦΑΛΑΙΟ 2

ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ ΚΑΙ ΒΑΘΙΑ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

2.1 Μηχανική μάθηση

2.2 Χρήση Strided Convolution για Υποδειγματοληψία

Σε αυτό το κεφάλαιο, παρέχεται μια λεπτομερής επισκόπηση των θεμελιωδών αρχών και των σύγχρονων προσεγγίσεων στη μηχανική μάθηση και τα βαθιά νευρωνικά δίκτυα. Αναλύονται διεξοδικά οι κύριες κατηγορίες μηχανικής μάθησης, με έμφαση στην επιβλεπόμενη μάθηση, η οποία αποτελεί τη θεμελιώδη βάση της παρούσας εργασίας. Παρουσιάζεται η περιγραφή της δομής και της λειτουργίας των τεχνητών νευρωνικών δικτύων, δίνοντας έμφαση στα βαθιά νευρωνικά δίκτυα. Ιδιαίτερη εστίαση δίνεται στις συναρτήσεις ενεργοποίησης, τους βελτιστοποιητές και τις εξειδικευμένες συναρτήσεις κόστους που εφαρμόζονται κατά την εκπαίδευση μοντέλων τμηματοποίησης ιατρικών εικόνων.

2.1 Μηχανική μάθηση

Η μηχανική μάθηση αποτελεί κλάδο της τεχνητής νοημοσύνης και γενικότερα της επιστήμης υπολογιστών. Εστιάζει στην χρήση αλγορίθμων και δεδομένων για να προσομοιώσει τη διαδικασία της μάθησης του ανθρώπου. Η βασική της φιλοσοφία

έγκειται στην ανάπτυξη μαθηματικών μοντέλων που εκπαιδευόμενα πάνω σε σύνολα αντιπροσωπευτικών παραδειγμάτων, αποκτώντας τη δυνατότητα να γενικεύουν και να λαμβάνουν αποφάσεις για νέα άγνωστα δεδομένα.

Όπως χαρακτηριστικά ανέφερε ο Arthur Samuel-πρωτοπόρος στον τομέα τεχνητής νοημοσύνης- ήδη από το 1959, η μηχανική μάθηση είναι «το πεδίο μελέτης που δίνει στους υπολογιστές την ικανότητα να μαθαίνουν χωρίς να έχουν ρητά προγραμματιστεί» [6]. Αργότερα, ο Frank Rosenblatt σχεδίασε το perceptron το πρώτο υπολογιστικό νευρωνικό δίκτυο για να μιμηθεί τον ανθρώπινο εγκέφαλο. Έκτοτε, η μηχανική μάθηση γνώρισε ραγδαία ανάπτυξη, αξιοποιώντας τη σημαντική αύξηση της υπολογιστικής ισχύος και την ευρεία διαθεσιμότητα μεγάλων όγκων δεδομένων ενώ παράλληλα κλήθηκε να αντιμετωπίσει πλήθος προκλήσεων [7].

Οι αλγόριθμοι μηχανικής μάθησης μπορούν να διακριθούν στις τρεις παρακάτω κατηγορίες:

- **Επιβλεπόμενη Μάθηση (Supervised Learning):** Η διαδικασία εκμάθησης μιας συνάρτησης από ένα σύνολο δεδομένων, στο οποίο είναι γνωστές τόσο οι είσοδοι όσο και οι αντίστοιχες μεταβλητές-στόχοι (targets). Το μοντέλο εκπαιδεύεται σε αυτά τα δεδομένα και στη συνέχεια αξιολογείται ως προς την ικανότητά του να γενικεύει σε νέα, άγνωστα δεδομένα.
- **Μη Επιβλεπόμενη Μάθηση (Unsupervised Learning):** Σε αντίθεση με την επιβλεπόμενη μάθηση, εδώ δεν υπάρχουν ετικέτες (labels) για να καθοδηγήσουν την εκπαίδευση. Στόχος είναι η ανακάλυψη κρυφών μοτίβων και δομών μέσα στα δεδομένα, μέσω τεχνικών όπως η ομαδοποίηση (clustering), η εκτίμηση πυκνότητας και η ανάλυση συσχέτισης.
- **Ενισχυτική Μάθηση (Reinforcement Learning):** Είναι η διαδικασία κατά την οποία ένας “πράκτορας” (agent) μαθαίνει μέσω αλληλεπίδρασης με το περιβάλλον, λαμβάνοντας ανταμοιβές και προσαρμόζοντας τις ενέργειές του ώστε να μεγιστοποιηθούν.

Περισσότερη ανάλυση θα γίνει στην επιβλεπόμενη μάθηση, καθώς αποτελεί το πρακτικό υπόβαθρο στην παρούσα εργασία [8].

2.1.1 Επιβλεπόμενη Μάθηση (Supervised Learning)

Η επιβλεπόμενη μάθηση αποτελεί μία από τις θεμελιώδεις κατηγορίες της μηχανικής μάθησης και περιλαμβάνει τη διαδικασία εκπαίδευσης ενός μοντέλου με χρήση επισημασμένων δεδομένων. Στη συγκεκριμένη προσέγγιση, κάθε δείγμα εκπαίδευσης συνοδεύεται από μία επιθυμητή έξοδο (ή ετικέτα), επιτρέποντας στο μοντέλο να μάθει τη συσχέτιση μεταξύ εισόδων και εξόδων. Ο τελικός στόχος είναι η δημιουργία ενός συστήματος που να μπορεί να γενικεύει, δηλαδή να προβλέπει σωστά τις εξόδους για άγνωστα, μη εμφανισθέντα στο παρελθόν, δεδομένα εισόδου.

Η διαδικασία επιβλεπόμενης μάθησης [9] ξεκινά με τη συλλογή και την προεπεξεργασία δεδομένων. Η ποιότητα και η καταλληλότητα των δεδομένων διαδραματίζουν κρίσιμο ρόλο στην αποτελεσματικότητα του μοντέλου. Τα δεδομένα αυτά διαχωρίζονται συνήθως σε σύνολα εκπαίδευσης και δοκιμής. Το μοντέλο εκπαιδεύεται στο σύνολο εκπαίδευσης και αξιολογείται ως προς τη γενίκευσή του στο σύνολο δοκιμής. Τα πιο συνηθισμένα είδη προβλημάτων επιβλεπόμενης μάθησης είναι η ταξινόμηση (classification) και η παλινδρόμηση (regression).

Στην ταξινόμηση, η έξοδος ανήκει σε διακριτές κατηγορίες. Για παράδειγμα, ένα μοντέλο μπορεί να εκπαιδευτεί ώστε να αναγνωρίζει αν μία εικόνα περιέχει γάτα ή σκύλο. Αντίθετα, στην παλινδρόμηση, η έξοδος είναι συνεχής. Ένα παράδειγμα θα μπορούσε να είναι η πρόβλεψη της τιμής ενός ακινήτου βάσει χαρακτηριστικών όπως η τοποθεσία, το μέγεθος και η ηλικία του.

Οι αλγόριθμοι που χρησιμοποιούνται στην επιβλεπόμενη μάθηση ποικίλουν και περιλαμβάνουν μεθόδους όπως οι γραμμικές παλινδρομήσεις, οι μηχανές υποστήριξης διανυσμάτων (Support Vector Machines), τα νευρωνικά δίκτυα, τα δέντρα απόφασης (Decision Trees), καθώς και πιο σύνθετες προσεγγίσεις όπως οι τυχαία δάση (Random Forests) και οι ενισχυτές βαθμίδωσης (Gradient Boosting Machines) [10]. Η επιλογή του κατάλληλου αλγορίθμου εξαρτάται από τη φύση του προβλήματος, το μέγεθος και τα χαρακτηριστικά των δεδομένων, καθώς και από τον επιθυμητό βαθμό ερμηνευσιμότητας.

Ένα από τα σημαντικότερα πλεονεκτήματα της επιβλεπόμενης μάθησης είναι η υψηλή ακρίβεια που μπορεί να επιτευχθεί, εφόσον το μοντέλο εκπαιδευτεί σωστά και τα δεδομένα είναι αντιπροσωπευτικά. Ωστόσο, η κύρια πρόκληση εντοπίζεται στην ανάγκη ύπαρξης μεγάλου όγκου επισημασμένων δεδομένων, κάτι που σε πολλές περιπτώσεις απαιτεί σημαντικό κόπο και χρόνο για την απόκτηση και ε-

πιμέλεια. Επιπλέον, υπάρχει ο κίνδυνος υπερπροσαρμογής (overfitting), κατά τον οποίο το μοντέλο μαθαίνει υπερβολικά καλά τα δεδομένα εκπαίδευσης, εις βάρος της ικανότητάς του να γενικεύει.

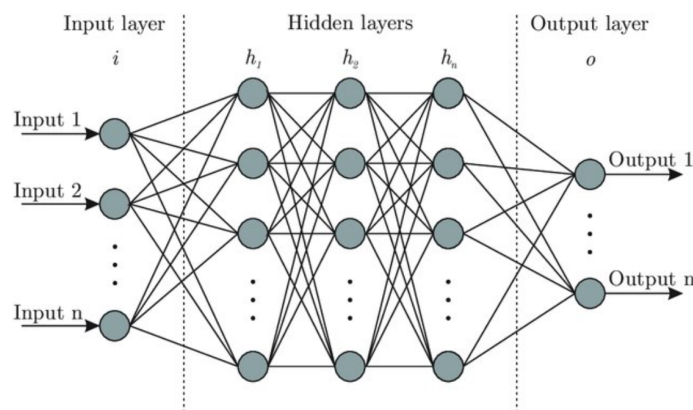
Η επιβλεπόμενη μάθηση αποτελεί μια ιδιαίτερα ισχυρή προσέγγιση στη μηχανική μάθηση, με ευρύ φάσμα εφαρμογών, από την αναγνώριση εικόνας και φωνής, μέχρι την ιατρική διάγνωση και την πρόβλεψη χρηματοοικονομικών τάσεων. Παρά τις προκλήσεις που ενδέχεται να παρουσιαστούν, η σωστή εφαρμογή της μπορεί να προσφέρει πολύτιμες λύσεις σε πραγματικά προβλήματα, καθιστώντας την αναπόσπαστο εργαλείο στη σύγχρονη ανάλυση δεδομένων.

2.1.2 Νευρωνικά Δίκτυα

Τα Τεχνητά Νευρωνικά Δίκτυα (ΤΝΔ) είναι υπολογιστικά μοντέλα εμπνευσμένα από τη δομή και τη λειτουργία του βιολογικού εγκεφάλου. Αποτελούνται από ένα σύνολο αλληλοσυνδεδεμένων κόμβων, που ονομάζονται τεχνητοί νευρώνες, οι οποίοι συνεργάζονται για την επεξεργασία πληροφοριών και την εξαγωγή συμπερασμάτων. Η θεμελιώδης φιλοσοφία πίσω από τα νευρωνικά δίκτυα είναι η προσπάθεια προσομοίωσης της πολύπλοκης δυναμικής του ανθρώπινου εγκεφάλου, ο οποίος αποτελείται από δισεκατομμύρια νευρώνες που επικοινωνούν μεταξύ τους μέσω ηλεκτροχημικών σημάτων. Σε αντίθεση με τον παραδοσιακό προγραμματισμό, όπου οι κανόνες και η λογική ορίζονται ρητά από τον προγραμματιστή, τα νευρωνικά δίκτυα “μαθαίνουν” από τα δεδομένα. Μέσω μιας διαδικασίας που ονομάζεται εκπαίδευση, τα δίκτυα προσαρμόζουν τις συνδέσεις μεταξύ των νευρώνων τους (τα λεγόμενα “βάρη”) βάσει παραδειγμάτων εισόδου και των αντίστοιχων επιθυμητών εξόδων. Αυτή η ικανότητα μάθησης τους επιτρέπει να αναγνωρίζουν πρότυπα, να κάνουν προβλέψεις και να λαμβάνουν αποφάσεις με έναν τρόπο που προσομοιάζει την ανθρώπινη διαίσθηση και εμπειρία. Η έμπνευση από τον εγκέφαλο δεν περιορίζεται μόνο στη δομή. Τα νευρωνικά δίκτυα υιοθετούν και τη φιλοσοφία της κατανεμημένης επεξεργασίας. Η “γνώση” σε ένα νευρωνικό δίκτυο δεν αποθηκεύεται σε μια κεντρική μονάδα, αλλά κατανέμεται σε όλο το δίκτυο, στις συνδέσεις μεταξύ των νευρώνων. Αυτό τα καθιστά ανθεκτικά σε μεμονωμένες βλάβες νευρώνων, μια ιδιότητα που παρατηρείται και στον βιολογικό εγκέφαλο.

Η αρχιτεκτονική ενός τυπικού νευρωνικού δικτύου αποτελείται από τρία βασικά είδη επιπέδων νευρώνων που συνεργάζονται για την επεξεργασία της πληροφορίας

[11]. Αρχικά, το επίπεδο εισόδου (Input Layer) είναι υπεύθυνο για την υποδοχή των αρχικών δεδομένων, τα οποία μπορεί να είναι οι τιμές των pixel μιας εικόνας, τα χαρακτηριστικά ενός ηχητικού σήματος ή οι αριθμητικές αναπαραστάσεις κειμένου, με κάθε νευρώνα να αντιστοιχεί σε ένα χαρακτηριστικό των δεδομένων. Στη συνέχεια, μεταξύ του επιπέδου εισόδου και εξόδου παρεμβάλλονται ένα ή περισσότερα κρυφά επίπεδα (Hidden Layers), όπου πραγματοποιείται το μεγαλύτερο μέρος της επεξεργασίας. Οι νευρώνες αυτών των επιπέδων λαμβάνουν ως είσοδο τις εξόδους των νευρώνων του προηγούμενου επιπέδου, τις επεξεργάζονται και τις προωθούν στο επόμενο. Ο αριθμός των κρυφών επιπέδων και των νευρώνων σε καθένα από αυτά καθορίζει την πολυπλοκότητα και την ικανότητα του δικτύου, με τα δίκτυα που διαθέτουν πολλά τέτοια επίπεδα να ονομάζονται “βαθιά” νευρωνικά δίκτυα (Deep Neural Networks). Τέλος, το επίπεδο εξόδου (Output Layer) είναι το τελευταίο στάδιο του δικτύου και παράγει το τελικό αποτέλεσμα. Ο αριθμός των νευρώνων στο επίπεδο εξόδου εξαρτάται από τη φύση του προβλήματος, όπως για παράδειγμα σε ένα πρόβλημα κατηγοριοποίησης, όπου θα μπορούσε να αντιστοιχεί ένας νευρώνας σε κάθε πιθανή κατηγορία. Η λειτουργία κάθε τεχνητού νευρώνα είναι σχετικά απλή: δέχεται πολλαπλές εισόδους, κάθε μία πολλαπλασιασμένη με ένα αντίστοιχο “βάρος”, αθροίζει τα γινόμενα και περνά το αποτέλεσμα μέσα από μια “συνάρτηση ενεργοποίησης” η οποία εισάγει την απαραίτητη μη-γραμμικότητα για την εκμάθηση πολύπλοκων σχέσεων και καθορίζει την τελική έξοδο του νευρώνα.



Σχήμα 2.1: Ενδεικτική αρχιτεκτονική τεχνητού νευρωνικού δικτύου με επίπεδο εισόδου, κρυφά επίπεδα και επίπεδο εξόδου.

Τα νευρωνικά δίκτυα μπορούν να κατηγοριοποιηθούν με βάση την αρχιτεκτονική τους και τον τρόπο με τον οποίο ρέει η πληροφορία μέσα τους. Μια θεμελιώδης

διάκριση είναι μεταξύ των Δικτύων Πρόσθιας Τροφοδότησης (Feedforward Neural Networks) και των Ανατροφοδοτούμενων Νευρωνικών Δικτύων (Recurrent Neural Networks - RNNs). Στην πρώτη κατηγορία, η πληροφορία κινείται αυστηρά προς μία κατεύθυνση, από το επίπεδο εισόδου προς το επίπεδο εξόδου, χωρίς να σχηματίζονται κύκλοι ή βρόχοι. Αυτά αποτελούν την απλούστερη και πιο διαδεδομένη μορφή νευρωνικών δικτύων, έχοντας ευρεία εφαρμογή σε προβλήματα ταξινόμησης και παλινδρόμησης. Αντιθέτως, τα Ανατροφοδοτούμενα Νευρωνικά Δίκτυα διαθέτουν συνδέσεις που σχηματίζουν βρόχους, επιτρέποντας στην πληροφορία να “επιμένει” και να ανακυκλώνεται μέσα στο δίκτυο. Αυτή η ιδιότητα τους προσδίδει ένα είδος “μνήμης”, καθιστώντας τα εξαιρετικά αποτελεσματικά για την επεξεργασία δεδομένων που έχουν ακολουθιακό χαρακτήρα.

Οι εφαρμογές των νευρωνικών δικτύων έχουν διαδοθεί σε ένα εντυπωσιακά ευρύ φάσμα τομέων, επιφέροντας σημαντικές προόδους και σε πολλές περιπτώσεις, επανάσταση. Στον τομέα της αναγνώρισης προτύπων, είναι η κινητήρια δύναμη πίσω από την αναγνώριση εικόνας και βίντεο, που επιτρέπει εφαρμογές όπως η αναγνώριση προσώπων και η αυτόνομη οδήγηση, καθώς και πίσω από την αναγνώριση ομιλίας και την επεξεργασία φυσικής γλώσσας, που τροφοδοτούν συστήματα μηχανικής μετάφρασης και ανάλυσης συναισθήματος. Στον κρίσιμο τομέα της ιατρικής, συμβάλλουν στην ιατρική διάγνωση μέσω της αυτόματης ανάλυσης ιατρικών εικόνων όπως ακτινογραφίες και μαγνητικές τομογραφίες. Η συνεχής αύξηση της υπολογιστικής ισχύος και η πρωτοφανής διαθεσιμότητα τεράστιων όγκων δεδομένων (Big Data) συνεχίζουν να τροφοδοτούν την εξέλιξη, οδηγώντας σε ολοένα και πιο ισχυρά νευρωνικά δίκτυα.

Η θεμελιώδης διαφορά των νευρωνικών δικτύων από τον παραδοσιακό προγραμματισμό έγκειται σε μια αντιστροφή της ίδιας της προγραμματιστικής λογικής. Στον κλασικό προγραμματισμό, ο άνθρωπος-προγραμματιστής παρέχει στο σύστημα δύο βασικά συστατικά: τα δεδομένα (data) και μια ρητά ορισμένη συνάρτηση (function) ή ένα σύνολο κανόνων που περιγράφουν βήμα προς βήμα πώς πρέπει να επεξεργαστούν αυτά τα δεδομένα. Το σύστημα, ακολουθώντας πιστά αυτές τις οδηγίες, εκτελεί τη συνάρτηση πάνω στα δεδομένα και παράγει ένα αποτέλεσμα (result). Η διαδικασία είναι πλήρως καθορισμένη και αιτιοκρατική. Τα νευρωνικά δίκτυα, ωστόσο, ακολουθούν μια ριζικά διαφορετική προσέγγιση που ονομάζεται μάθηση. Εδώ, ο προγραμματιστής δεν παρέχει τη συνάρτηση, αλλά τα δεδομένα εισόδου μαζί με το επιθυμητό αποτέλεσμα. Για παράδειγμα, δίνουμε στο δίκτυο χιλιάδες εικόνες

σκύλων (δεδομένα) και του λέμε για κάθε μία “αυτό είναι σκύλος” (αποτέλεσμα). Ο ρόλος του νευρωνικού δικτύου είναι να “ανακαλύψει” από μόνο του την υποκείμενη συνάρτηση, δηλαδή τις εσωτερικές παραμέτρους και τη δομή που συνδέουν με συνέπεια τα συγκεκριμένα δεδομένα με τα συγκεκριμένα αποτελέσματα. Το τελικό προϊόν της διαδικασίας εκπαίδευσης δεν είναι ένα απλό αποτέλεσμα, αλλά το ίδιο το πρόγραμμα—ένα εκπαιδευμένο μοντέλο ικανό να γενικεύει τη γνώση του και να παράγει σωστά αποτελέσματα για νέα, άγνωστα δεδομένα.

2.1.3 Βαθεία Μάθηση (Deep Learning)

Η Βαθεία Μάθηση (Deep Learning) αποτελεί έναν προηγμένο υποτομέα της μηχανικής μάθησης που αξιοποιεί Τεχνητά Νευρωνικά Δίκτυα με πολυεπίπεδη αρχιτεκτονική, γνωστά ως Βαθεία Νευρωνικά Δίκτυα (Deep Neural Networks - DNNs) [12]. Η θεμελιώδης διαφορά που αναβαθμίζει τα απλά νευρωνικά δίκτυα σε “βαθεία” είναι η αρχιτεκτονική τους πολυπλοκότητα. Ένα απλό νευρωνικό δίκτυο, γνωστό και ως “ρηχό” (shallow), μπορεί να αποτελείται από ένα επίπεδο εισόδου, ένα μόνο κρυφό επίπεδο επεξεργασίας, και ένα επίπεδο εξόδου. Αν και είναι αποτελεσματικά για βασικά προβλήματα ταξινόμησης, η ικανότητά τους να κατανοούν πολύπλοκες σχέσεις είναι περιορισμένη. Η βαθιά μάθηση, από την άλλη πλευρά, εισάγει πολλαπλά κρυφά επίπεδα μεταξύ της εισόδου και της εξόδου. Αυτό το “βάθος” είναι που αλλάζει δραματικά τις δυνατότητες του μοντέλου, επιτρέποντάς του να χτίσει μια πολύ πιο πλούσια και πολυδιάστατη κατανόηση των δεδομένων.

Ο ρόλος των κρυφών επιπέδων είναι να λειτουργούν ως διαδοχικά φίλτρα επεξεργασίας, δημιουργώντας μια ιεραρχία αφαίρεσης. Κάθε επίπεδο λαμβάνει ως είσοδο την έξοδο του προηγούμενου και μαθαίνει να αναγνωρίζει όλο και πιο σύνθετα χαρακτηριστικά. Για παράδειγμα, στην αναγνώριση μιας εικόνας, το πρώτο κρυφό επίπεδο μπορεί να μάθει να εντοπίζει απλές φόρμες, όπως άκρες και γωνίες. Το επόμενο επίπεδο συνδυάζει αυτές τις άκρες για να αναγνωρίσει πιο σύνθετες δομές, όπως μάτια ή μύτες. Ένα ακόμα βαθύτερο επίπεδο μπορεί να συνθέσει αυτά τα χαρακτηριστικά για να σχηματίσει ολόκληρα πρόσωπα. Αυτή η ιεραρχική προσέγγιση επιτρέπει στο δίκτυο να “σπάει” ένα περίπλοκο πρόβλημα σε απλούστερα, διαχειρίσιμα μέρη και να κατανοεί έννοιες με έναν τρόπο που προσομοιάζει την ανθρώπινη αντίληψη.

Η προσθήκη πολλών κρυφών επιπέδων, ωστόσο, έχει τόσο θετικές όσο και αρνη-

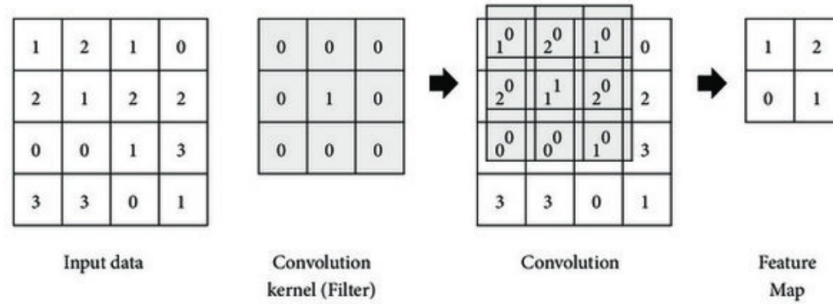
τικές πτυχές. Το κυριότερο πλεονέκτημα είναι η δραματική αύξηση της εκφραστικής δύναμης του μοντέλου. Τα βαθιά δίκτυα μπορούν να μοντελοποιήσουν εξαιρετικά πολύπλοκες και μη-γραμμικές σχέσεις στα δεδομένα, επιτυγχάνοντας κορυφαίες επιδόσεις σε δύσκολα προβλήματα.

Παρά τα σημαντικά τους πλεονεκτήματα, τα βαθιά νευρωνικά δίκτυα παρουσιάζουν και προκλήσεις. Η αυξημένη αρχιτεκτονική πολυπλοκότητα συνεπάγεται υψηλές υπολογιστικές απαιτήσεις και παρατεταμένο χρόνο εκπαίδευσης. Επιπλέον, απαιτούν μεγάλα, ποιοτικά σύνολα δεδομένων ώστε να αποφευχθεί η υπερπροσαρμογή (overfitting), δηλαδή η αποστήθιση αντί για γενίκευση. Τέλος, όσο αυξάνεται το βάθος ενός δικτύου, τόσο εντείνεται το πρόβλημα της ερμηνευσιμότητας.

Στο πλαίσιο της παρούσας εργασίας, η Βαθιά Μάθηση χρησιμοποιείται για την κατάτμηση ιατρικών εικόνων και συγκεκριμένα νεφρών και όγκων, με έμφαση σε αρχιτεκτονικές όπως το **3D U-Net**, που έχουν σχεδιαστεί ειδικά για εφαρμογές στην ιατρική απεικόνιση.

2.1.4 Συνελικτικό Επίπεδο

Το συνελικτικό επίπεδο (Convolutional Layer) αποτελεί τον πυρήνα της λειτουργίας των Συνελικτικών Νευρωνικών Δικτύων (Convolutional Neural Networks – CNNs), τα οποία έχουν αναδειχθεί ως η κυρίαρχη αρχιτεκτονική για προβλήματα επεξεργασίας εικόνας [13]. Η βασική αρχή πίσω από το συνελικτικό επίπεδο είναι η χρήση φίλτρων ή πυρήνων (kernels), τα οποία εφαρμόζονται στα δεδομένα εισόδου μέσω της διαδικασίας της συνέλιξης. Κατά τη συνέλιξη, κάθε φίλτρο «σαρώνει» την εικόνα (ή γενικά τον πίνακα εισόδου) και υπολογίζει το εσωτερικό γινόμενο μεταξύ των τιμών του φίλτρου και του αντίστοιχου υποπίνακα της εισόδου. Το αποτέλεσμα είναι ένας νέος πίνακας, ο οποίος ονομάζεται χάρτης ενεργοποίησης (feature map).



Σχήμα 2.2: Παράδειγμα πράξης συνέλιξης (convolution) με φίλτρο 3×3 σε εικόνα 4×4 .

Το βασικό πλεονέκτημα της συνέλιξης είναι η ικανότητα εντοπισμού τοπικών χαρακτηριστικών, όπως ακμές, γωνίες ή υφές, με τρόπο που διατηρεί την τοπολογική δομή της εικόνας. Τα φίλτρα μαθαίνονται αυτόματα κατά την εκπαίδευση του δικτύου και προσαρμόζονται ώστε να ανταποκρίνονται στις σημαντικότερες οπτικές ιδιότητες των δεδομένων. Η συνελκτική πράξη επιτρέπει την αφαίρεση σημαντικών γνωρισμάτων, τα οποία γίνονται σταδιακά πιο αφηρημένα όσο προχωρούμε στα επόμενα επίπεδα του δικτύου. Η χρήση του συνελκτικού επιπέδου συνοδεύεται συχνά από άλλες τεχνικές, όπως η συμπλήρωση με μηδενικά (zero-padding) και η ρύθμιση του βήματος (stride), οι οποίες επηρεάζουν τις διαστάσεις των παραγόμενων χαρτών ενεργοποίησης. Το padding χρησιμοποιείται για να διατηρηθεί το μέγεθος της εξόδου ίσο με την είσοδο ή για να ελεγχθεί ο ρυθμός μείωσης της χωρικής πληροφορίας. Από την άλλη, η παράμετρος stride καθορίζει πόσο μετακινείται το φίλτρο κατά τη διάρκεια της σάρωσης, επηρεάζοντας έτσι τον αριθμό των εξόδων που παράγονται.

Επιπλέον, ένα συνελκτικό επίπεδο μπορεί να περιλαμβάνει πολλαπλά φίλτρα, επιτρέποντας την ταυτόχρονη ανίχνευση διαφορετικών χαρακτηριστικών. Το αποτέλεσμα είναι ένα σύνολο πολλών χαρτών ενεργοποίησης, τα οποία συλλογικά σχηματίζουν έναν τρισδιάστατο πίνακα εξόδου. Αυτοί οι πίνακες περνούν στη συνέχεια από μη γραμμικές συναρτήσεις ενεργοποίησης (όπως ReLU), προκειμένου να εισαχθεί μη γραμμικότητα στο μοντέλο. Στην πράξη, τα συνελκτικά επίπεδα ενσωματώνονται σε πολυστρωματικά δίκτυα, επιτρέποντας την ιεραρχική κατανόηση των εισόδων. Τα πρώτα επίπεδα ανιχνεύουν απλά χαρακτηριστικά, ενώ τα επόμενα συνθέτουν πιο σύνθετες αναπαραστάσεις. Η σωστή διαμόρφωση και παραμετροποίηση των συνελκτικών επιπέδων είναι καθοριστικής σημασίας για την απόδοση του

νευρωνικού μοντέλου.

2.1.5 Συνάρτηση Ενεργοποίησης (Activation Function)

Οι συναρτήσεις ενεργοποίησης (activation functions) διαδραματίζουν κρίσιμο ρόλο στη λειτουργία των τεχνητών νευρωνικών δικτύων, καθώς καθορίζουν την έξοδο κάθε νευρώνα με βάση το σταθμισμένο άθροισμα των εισόδων του. Στόχος τους είναι η εισαγωγή μη γραμμικότητας στο δίκτυο, επιτρέποντας στο μοντέλο να μάθει και να προσεγγίσει σύνθετες συναρτήσεις, πέρα από γραμμικούς μετασχηματισμούς. Χωρίς τη χρήση μη γραμμικών συναρτήσεων ενεργοποίησης, ακόμα και ένα βαθύ νευρωνικό δίκτυο θα είχε περιορισμένη εκφραστική δύναμη, καθώς θα κατέρρεε σε μια απλή γραμμική συναρτησιακή σύνθεση.

Μια από τις πιο διαδεδομένες συναρτήσεις είναι η Rectified Linear Unit (ReLU), η οποία ορίζεται ως:

$$\text{ReLU}(x) = \max(0, x).$$

Η απλότητά της και η αποδοτική υλοποίηση επιταχύνουν τη σύγκλιση και περιορίζουν το πρόβλημα του εξαφανιζόμενου βαθμωτού (vanishing gradient). Ωστόσο, η ReLU μπορεί να παράγει «νεκρούς» νευρώνες (dead neurons) όπου οι έξοδοι κάποιων νευρώνων μηδενίζονται μόνιμα [14].

Παραλλαγές της έχουν προταθεί για την αντιμετώπιση αυτού του φαινομένου:

- **Leaky ReLU:** $\text{LeakyReLU}(x) = \begin{cases} x, & x \geq 0 \\ \alpha x, & x < 0 \end{cases}$, με $\alpha \in (0, 1)$ μικρή σταθερά, ώστε να διατηρείται ροή πληροφορίας για αρνητικές τιμές [14].

- **Parametric ReLU (PReLU):** $\text{PReLU}(x) = \begin{cases} x, & x \geq 0 \\ px, & x < 0 \end{cases}$, όπου p είναι παράμετρος που μαθαίνεται κατά την εκπαίδευση [14].

- **ELU:** Exponential Linear Unit, που συνδυάζει τα πλεονεκτήματα της ReLU με ομαλότερη μεταβολή για $x < 0$, ενισχύοντας τη γενίκευση.

Η softmax χρησιμοποιείται σχεδόν αποκλειστικά στο επίπεδο εξόδου για πολυκατηγορική ταξινόμηση (multi-class segmentation), καθώς μετατρέπει τις εξόδους σε πιθανότητες με άθροισμα 1.

$$p_c(x) = \frac{e^{z_c(x)}}{\sum_k e^{z_k(x)}}, \quad c \in \{0, 1, 2\},$$

Στην παρούσα εργασία χρησιμοποιήθηκε για την εκπαίδευση και εξαγωγή συμπερασμάτων η **Leaky ReLU** για τα κρυφά στρώματα και η **Softmax** στο επίπεδο εξόδου. Η σωστή επιλογή των συναρτήσεων ενεργοποίησης είναι κρίσιμη για τη σταθερότητα της εκπαίδευσης, την αποφυγή προβλημάτων βαθμού και την ποιότητα της κατάτμησης.

2.2 Χρήση Strided Convolution για Υποδειγματοληψία

Παραδοσιακά, η υποδειγματοληψία (downsampling) σε συνελικτικά νευρωνικά δίκτυα (CNNs) επιτυγχάνεται μέσω του *max pooling*, το οποίο επιλέγει τη μέγιστη τιμή σε κάθε τοπική περιοχή του χαρακτηριστικού χάρτη. Αν και είναι υπολογιστικά αποδοτικό, το *max pooling* είναι μη παραμετρικό και δεν επιτρέπει στο δίκτυο να μάθει βέλτιστες αναπαραστάσεις κατά τη διάρκεια της εκπαίδευσης.

Αντιθέτως, η τεχνική του *strided convolution* ενοποιεί τη συνέλιξη και την υποδειγματοληψία σε ένα ενιαίο παραμετρικό επίπεδο, επιτρέποντας στο δίκτυο να μάθει πώς να μειώνει τη χωρική πληροφορία μέσω εκπαίδευσης. Έρευνες υποδεικνύουν ότι αυτή η μέθοδος ενδέχεται να προσφέρει μεγαλύτερη αποτελεσματικότητα σε σύγκριση με το *max pooling*, τόσο στην απόδοση όσο και στη γενίκευση σε περίπλοκα δεδομένα [15].

Σε ένα συνελικτικό νευρωνικό δίκτυο (CNN), η τιμή του βήματος (*stride*) καθορίζει πόσο μετακινείται το φίλτρο συνέλιξης πάνω στην είσοδο. Στην απλή περίπτωση, όταν το *stride* είναι ίσο με 1, το φίλτρο μετακινείται κατά ένα voxel (ή pixel) κάθε φορά, δημιουργώντας έξοδο ίδιου ή παρόμοιου μεγέθους με την είσοδο. Όταν το *stride* είναι μεγαλύτερο από 1, η έξοδος γίνεται μικρότερη, δηλαδή, το *stride* επιφέρει χωρική υποδειγματοληψία (*spatial downsampling*).

Ο τύπος που καθορίζει το μέγεθος της εξόδου μιας συνέλιξης είναι:

$$D_{\text{out}} = \left\lfloor \frac{D_{\text{in}} + 2P - K}{S} \right\rfloor + 1$$

όπου:

- D_{in} είναι η διάσταση της εισόδου (σε κάθε άξονα),
- K kernel size,
- P Padding,

- S stride,
- D_{out} η διάσταση της εξόδου.

Για παράδειγμα, με $K = 3$, $P = 1$, και $S = 2$, η είσοδος $D_{\text{in}} = 64$ παράγει έξοδο $D_{\text{out}} = 32$. Αυτό συνδυάζει τη συνέλιξη και την υποδειγματοληψία σε ένα μόνο βήμα, χωρίς την ανάγκη ξεχωριστού επιπέδου *pooling*.

2.2.1 Επίπεδο Εξόδου

Το επίπεδο εξόδου ενός νευρωνικού δικτύου καθορίζει τη μορφή της τελικής πρόβλεψης και εξαρτάται από τη φύση του προβλήματος. Στην παρούσα εργασία, το ζητούμενο είναι η πολυκατηγορική ταξινόμηση κάθε voxel σε μία από τις διαθέσιμες κατηγορίες (π.χ. νεφρό, όγκος, υπόβαθρο), και για το σκοπό αυτό χρησιμοποιείται η συνάρτηση softmax στο επίπεδο εξόδου.

Σε προβλήματα πολυκατηγορικής ταξινόμησης, το επίπεδο εξόδου περιλαμβάνει k νευρώνες, όσοι και οι διαφορετικές κλάσεις, και εφαρμόζεται η συνάρτηση softmax. Μετατρέπει τα ακατέργαστα σκορ (logits) σε κανονικοποιημένες πιθανότητες, οι οποίες αθροίζονται στο 1. Η κλάση με τη μέγιστη πιθανότητα επιλέγεται ως η τελική πρόβλεψη. Η συνάρτηση ορίζεται ως:

$$\text{softmax}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^k e^{z_j}}$$

όπου z_i είναι το σκορ του i -οστού νευρώνα. Η softmax ευνοεί την ενίσχυση της κυρίαρχης κατηγορίας, ενώ κρατά υπόψη την πληροφορία των υπολοίπων, γεγονός που την καθιστά ιδανική για προβλήματα όπου απαιτείται πιθανολογική ερμηνεία.

2.2.2 Βελτιστοποιητές

Οι βελτιστοποιητές (optimizers) αποτελούν το βασικό μηχανισμό εκμάθησης των παραμέτρων ενός νευρωνικού δικτύου μέσω επαναληπτικής ελαχιστοποίησης της συνάρτησης απώλειας. Ο στόχος ενός βελτιστοποιητή είναι να επιλέγει σε κάθε βήμα εκπαίδευσης την κατάλληλη κατεύθυνση και το μέγεθος ενημέρωσης των παραμέτρων του μοντέλου (βάρη και προκαταλήψεις), με σκοπό τη μεγιστοποίηση της απόδοσης και τη σταθερή σύγκλιση σε ένα σημείο ελάχιστης απώλειας.

Η γενική μορφή ενημέρωσης ενός συνόλου παραμέτρων θ (που περιλαμβάνει τα $W^{(l)}$ και $b^{(l)}$ όλων των επιπέδων) βασίζεται στον υπολογισμό της κλίσης της συνάρτησης απώλειας $L(\theta)$ ως προς τις παραμέτρους:

$$\theta_{t+1} = \theta_t - \eta_t \nabla_{\theta} L(\theta_t),$$

όπου:

- θ_t είναι το διάνυσμα παραμέτρων στη χρονική στιγμή t ,
- η_t ο ρυθμός μάθησης (learning rate),
- $\nabla_{\theta} L$ η κλίση της συνάρτησης κόστους ως προς τις παραμέτρους.

Αν και ο απλός **Gradient Descent** αποτελεί τη βάση όλων των μεθόδων, στην πράξη χρησιμοποιούνται παραλλαγές και πιο εξελιγμένοι αλγόριθμοι. Ένας από αυτούς είναι ο **Stochastic Gradient Descent (SGD)** με προσθήκη των τεχνικών **momentum** και **Nesterov Accelerated Gradient (NAG)**, όπως περιγράφονται στο άρθρο του Sutskever et al. (2013)[16].

Η κύρια ενημέρωση των παραμέτρων θ παρέχεται από:

$$\theta_{t+1} = \theta_t - \eta \nabla_{\theta} L(\theta_t),$$

όπου η αντιπροσωπεύει τον ρυθμό μάθησης και $\nabla_{\theta} L(\theta_t)$ είναι το gradient της συνάρτησης κόστους στο βήμα t .

Η τεχνική του momentum εισάγει το διάνυσμα ταχύτητας v_t , όπου κατευθύνει την ενημέρωση προς κατευθύνσεις όπου το gradient παραμένει σταθερά μειωτικό, επιταχύνοντας την εκπαίδευση και ελαττώνοντας τις ταλαντώσεις.

$$v_t = \mu v_{t-1} + \nabla_{\theta} L(\theta_{t-1}),$$

$$\theta_t = \theta_{t-1} - \eta v_t,$$

όπου $\mu \in [0, 1]$ είναι ο συντελεστής momentum. Η μέθοδος **Nesterov Accelerated Gradient** προάγει την ενημέρωση, υπολογίζοντας το gradient στην αναμενόμενη θέση μετά την εφαρμογή του βήματος momentum. Με αυτόν τον τρόπο, η ενημέρωση αποκτά “προγνωστική” κατεύθυνση, ενισχύοντας τη σταθερότητα και επιταχύνοντας σύγκλιση:

$$v_t = \mu v_{t-1} + \nabla_{\theta} L(\theta_{t-1} - \eta \mu v_{t-1}),$$

$$\theta_t = \theta_{t-1} - \eta v_t.$$

Η μέθοδος που περιγράφεται από τους Sutskever et al. (2013) [16] συνδυάζει την ταχύτητα του momentum με τη βελτιωμένη κατεύθυνση της ενημέρωσης, επιταχύνοντας τη σύγκλιση και μειώνοντας τις ταλαντώσεις, γεγονός που είναι κρίσιμο για την εκπαίδευση βαθιών δικτύων όπως το 3D U-Net.

2.2.3 Μέθοδος Οπισθοδιάδοσης Σφάλματος (Backpropagation)

Η μέθοδος οπισθοδιάδοσης σφάλματος (backpropagation) αποτελεί τον βασικό μηχανισμό εκπαίδευσης των τεχνητών νευρωνικών δικτύων μέσω του αλγορίθμου βελτιστοποίησης της μεθόδου καθοδικής κλίσης (gradient descent). Η διαδικασία στηρίζεται στον κανόνα της αλυσίδας από τον διαφορικό λογισμό, επιτρέποντας τον υπολογισμό των παραγώγων της συνάρτησης απώλειας ως προς τα βάρη και τις προκαταλήψεις του δικτύου, με στόχο την ελαχιστοποίησή της μέσω ενημέρωσης των παραμέτρων.

Έστω ένα νευρωνικό δίκτυο με L επίπεδα. Ορίζουμε τα εξής:

- $a^{(l)}$: οι ενεργοποιήσεις του επιπέδου l
- $z^{(l)}$: η είσοδος του επιπέδου l πριν την εφαρμογή της συνάρτησης ενεργοποίησης
- $W^{(l)}$: ο πίνακας βαρών του επιπέδου l
- $b^{(l)}$: το διάνυσμα προκαταλήψεων του επιπέδου l
- $f^{(l)}$: η συνάρτηση ενεργοποίησης στο επίπεδο l
- L : η συνολική συνάρτηση απώλειας

Η προώθηση (forward pass) ορίζεται επαναληπτικά ως:

$$z^{(l)} = W^{(l)} a^{(l-1)} + b^{(l)}, \quad a^{(l)} = f^{(l)}(z^{(l)}).$$

Το τελικό επίπεδο $a^{(L)}$ παράγει την έξοδο του δικτύου. Η συνάρτηση απώλειας $L(a^{(L)}, y)$ εξαρτάται από την πρόβλεψη και την αληθινή έξοδο y .

Στόχος: Να υπολογιστούν τα $\frac{\partial L}{\partial W^{(l)}}$ και $\frac{\partial L}{\partial b^{(l)}}$ για όλα τα επίπεδα l , ώστε να ενημερωθούν τα βάρη με τον κανόνα:

$$W^{(l)} = W^{(l)} - \eta \frac{\partial L}{\partial W^{(l)}}, \quad b^{(l)} = b^{(l)} - \eta \frac{\partial L}{\partial b^{(l)}}$$

όπου η ο ρυθμός μάθησης.

Οπισθοδιάδοση (Backpropagation): Ορίζουμε το σφάλμα στο επίπεδο l ως:

$$\delta^{(l)} = \frac{\partial L}{\partial z^{(l)}} = \frac{\partial L}{\partial a^{(l)}} \odot f'(z^{(l)})$$

όπου $\odot$ δηλώνει το γινόμενο κατά στοιχεία (Hadamard product) και f' η παράγωγος της συνάρτησης ενεργοποίησης.

Για το τελευταίο επίπεδο ($l = L$):

$$\delta^{(L)} = \nabla_{a^{(L)}} L \odot f'(z^{(L)}).$$

Για τα ενδιάμεσα επίπεδα ($l = L - 1, \dots, 1$), επαναληπτικά:

$$\delta^{(l)} = ((W^{(l+1)})^T \delta^{(l+1)}) \odot f'(z^{(l)}).$$

Παράγωγοι κόστους ως προς βάρη και προκαταλήψεις: Αφού υπολογιστούν τα $\delta^{(l)}$, οι επιθυμητές παράγωγοι δίνονται από:

$$\frac{\partial L}{\partial W^{(l)}} = \delta^{(l)} (a^{(l-1)})^T, \quad \frac{\partial L}{\partial b^{(l)}} = \delta^{(l)}.$$

Οι παραπάνω εξισώσεις επιτρέπουν την εφαρμογή της ενημέρωσης βαρών για όλα τα επίπεδα του δικτύου.

Η μέθοδος backpropagation είναι υπολογιστικά αποδοτική, καθώς ανακυκλώνει ενδιάμεσους υπολογισμούς και εφαρμόζεται μέσω δυναμικού προγραμματισμού[17]. Αποτελεί θεμέλιο της επιτυχίας των βαθιών νευρωνικών δικτύων, και η αποδοτική υλοποίησή της σε σύγχρονες GPU αποτελεί ακρογωνιαίο λίθο των βιβλιοθηκών TensorFlow, PyTorch και άλλων. Η κατανόηση της οπισθοδιάδοσης επιτρέπει επίσης τη βελτιστοποίηση των αρχιτεκτονικών μέσω τεχνικών όπως η κανονικοποίηση (batch/layer norm), το gradient clipping, καθώς και η χρήση κατάλληλων ενεργοποιήσεων ώστε να διατηρείται η σταθερότητα της διάδοσης των παραγώγων. Επιπλέον, η μέθοδος μπορεί να εφαρμοστεί εκτενώς σε σύνθετες αρχιτεκτονικές όπως 3D U-Net, χάρη στη δυνατότητα ταξινομημένης αναδρομής των παραγώγων σε πολλαπλά επίπεδα.

2.2.4 Συναρτήσεις Απώλειας για Εκπαίδευση

Η συνάρτηση απώλειας αποτελεί θεμελιώδες στοιχείο στην εκπαίδευση ενός νευρωνικού δικτύου, καθώς ποσοτικοποιεί τη διαφορά μεταξύ της προβλεπόμενης εξόδου του μοντέλου και της πραγματικής επιθυμητής τιμής. Η ενημέρωση των παραμέτρων πραγματοποιείται μέσω του αλγορίθμου οπισθοδιάδοσης και του αντίστοιχου βελτιστοποιητή. Η επιλογή της κατάλληλης συνάρτησης εξαρτάται από τη φύση του προβλήματος, όπως η ταξινόμηση ή η παλινδρόμηση.

Στην παρούσα εργασία, εφαρμόζεται ένας συνδυασμός συναρτήσεων απώλειας για τη βελτίωση της ακρίβειας στην τρισδιάστατη κατάτμηση, συγκεκριμένα:

1. Categorical Cross-Entropy Loss

Ο αρχικός τύπος της Cross-Entropy εισήχθη από τον Claude Shannon το 1948 [18] ως μέτρο της «απόκλισης» μεταξύ δύο κατανομών πιθανοτήτων:

$$H(p, q) = - \sum_x p(x) \log q(x)$$

όπου:

- $p(x)$: πραγματική κατανομή (ground truth).
- $q(x)$: προβλεπόμενη κατανομή.

Στην παρούσα εργασία υιοθετούμε την μέθοδο του Christopher M. Bishop [19] όπου επέκτεινε τον τύπο για πολυκατηγορική ταξινόμηση σε νευρωνικά δίκτυα. Ο υπολογισμός του cross-entropy loss γίνεται σε επίπεδο voxel με τον παρακάτω τύπο για το δίκτυό μας που έχει ως έξοδο την softmax. Έστω N το πλήθος των voxels, C ο αριθμός των κλάσεων, και $\hat{y}_{i,c}$ το logit της πρόβλεψης για την κλάση c στο voxel i . Αν y_i είναι η πραγματική κλάση του i -ού voxel, τότε:

$$\mathcal{L}_{CE} = -\frac{1}{N} \sum_{i=1}^N \log \left(\frac{\exp(\hat{y}_{i,c=y_i})}{\sum_{c=1}^C \exp(\hat{y}_{i,c})} \right).$$

2. Dice Loss

Η συνάρτηση Dice Loss προέρχεται από τον δείκτη Dice (Dice Coefficient), ο οποίος είναι ένας στατιστικός δείκτης που αξιολογεί την ομοιότητα μεταξύ δύο συνόλων. Στην τμηματοποίηση εικόνας, ο δείκτης Dice αξιολογεί τον βαθμό επικάλυψης μεταξύ της προβλεπόμενης και της πραγματικής μάσκας. Ορίζεται ως το μέγεθος της τομής των δύο μασκών διαιρούμενο με το συνολικό άθροισμα των μεγεθών τους.

Η Dice Loss μετρά τον βαθμό επικάλυψης μεταξύ της προβλεπόμενης και της πραγματικής ετικέτας. Προτάθηκε από τους Milletari et al., 2016 [20]. Ο ορισμός του δείκτη Dice δίνεται παρακάτω:

$$\text{Dice}(Y, T) = \frac{2|Y \cap T|}{|Y| + |T|} \quad (2.1)$$

Όπου Y είναι η δυαδική μάσκα πρόβλεψης και T η δυαδική μάσκα στόχου. Στην πράξη, συχνά εφαρμόζεται μια διαφορετική εκδοχή του δείκτη Dice η οποία είναι διαφορίσιμη και επιτρέπει τη χρήση της σε στοχαστικούς αλγορίθμους εκπαίδευσης (π.χ. με backpropagation) και ελαχιστοποιείται κατά την εκπαίδευση. Η Dice Loss ορίζεται ως:

$$\mathcal{L}_{\text{Dice}} = 1 - \frac{2 \sum_{i=1}^N p_i t_i + \epsilon}{\sum_{i=1}^N p_i + \sum_{i=1}^N t_i + \epsilon}$$

Όπου:

- $p_i \in [0, 1]$: η προβλεπόμενη πιθανότητα για το voxel i
- $t_i \in \{0, 1\}$: η ground truth ετικέτα στο voxel i
- ϵ : μια μικρή σταθερά για αποφυγή διαίρεσης με το μηδέν
- N : συνολικός αριθμός voxel.

Η υλοποίηση που εφαρμόζεται στην παρούσα διπλωματική εργασία υποστηρίζει πολυκατηγορική τμηματοποίηση (multi-class semantic segmentation). Για κάθε κατηγορία C υπολογίζεται ανεξάρτητα η απώλεια Dice και τελικά προκύπτει ο μέσος όρος ως:

$$\mathcal{L}_{\text{total}} = \frac{1}{C} \sum_{c=1}^C \mathcal{L}_{\text{Dice}}^{(c)}$$

2.2.5 Μέθοδος Επιπέδων (Level Set) και LSF Loss

Η μέθοδος επιπέδων (level set) αποτελεί ένα μαθηματικό πλαίσιο για την παρακολούθηση κινούμενων μετώπων (όπως τα περιγράμματα αντικειμένων) μέσω της αναπαράστασης αυτών ως το μηδενικό ισοδύναμο μιας συνάρτησης

$$\phi(x, t) \quad (2.2)$$

όπου $x = (x_1, \dots, x_n) \in \mathbb{R}^n$. είναι το σημείο στον χώρο (συντεταγμένες) και $t \in \mathbb{R}$ η παράμετρος χρόνου ή εξέλιξης. Το όριο του αντικειμένου δίνεται από το σύνολο

$$\{x \mid \phi(x, t) = 0\} \quad (2.3)$$

ενώ η ϕ ορίζεται ως signed distance ικανοποιώντας το παρακάτω:

$$\begin{cases} \phi(x, t) > 0, & x \in \Omega, \\ \phi(x, t) < 0, & x \notin \bar{\Omega}, \\ \phi(x, t) = 0, & x \in \partial\Omega = \Gamma(t). \end{cases} \quad (2.4)$$

Η συνάρτηση $\phi(x, t)$ ορίζεται έτσι ώστε να είναι θετική μέσα στην περιοχή Ω , αρνητική έξω από την κλειστή περιοχή $\bar{\Omega}$ και μηδέν ακριβώς στο όριο $\partial\Omega$, το οποίο συμβολίζεται με $\Gamma(t)$ [21]. Έτσι, το όριο του αντικειμένου δίνεται από το zero level set της ϕ , όπου και ικανοποιεί την παρακάτω διαφορική εξίσωση :

$$\frac{\partial \phi}{\partial t} + F |\nabla \phi| = 0 \quad (2.5)$$

όπου F η ταχύτητα διάδοσης κατά τον κανονικό άξονα της επιφάνειας. Το F ορίζεται ως το κανονικό της ταχύτητας $\vec{v}$, δηλαδή η προβολή της ταχύτητας πάνω στη μοναδιαία κανονική διεύθυνση $\frac{\nabla \phi}{|\nabla \phi|}$:

$$F = \vec{v} \cdot \frac{\nabla \phi}{|\nabla \phi|}, \quad (2.6)$$

όπου $\frac{\nabla \phi}{|\nabla \phi|}$ είναι η μοναδιαία κανονική διεύθυνση στην επιφάνεια. Τα πλεονεκτήματα είναι η ικανότητα να χειρίζονται αυτόματα τοπολογικές αλλαγές (πχ διαχωρισμός ή συγχώνευση) και ο υπολογισμός γεωμετρικών χαρακτηριστικών μέσω των παραγώγων της ϕ [22]. Στις εφαρμογές εικόνas, το πρόσημο της ϕ κατηγοριοποιεί τα pixel σε foreground και background.

Level Set Loss Υλοποίηση Βασιζόμενη στην ιδέα της level-set μεθόδου και χρησιμοποιώντας τις softmax πιθανότητες $p_{i,c}$ που προβλέπει το δίκτυο για κάθε voxel i και κλάση c , το δίκτυο εκπαιδεύεται ώστε να παράγει απευθείας level-set συναρτήσεις ϕ (signed distance functions) αντί για απλούς χάρτες πιθανοτήτων. Έτσι, η μηδενική στάθμη (zero level set) της ϕ οριοθετεί με σαφήνεια την περιοχή ενδιαφέροντος, δίνοντας έμφαση στην ακρίβεια των ορίων.

Βήματα Υλοποίησης:

- Από τα logits των προβλέψεων εφαρμόζουμε τη συνάρτηση *softmax* για να λάβουμε τον πίνακα πιθανοτήτων:

$$\mathbf{probs} \in \mathbb{R}^{N \times C \times H \times W \times D}$$

όπου N το μέγεθος παρτίδας (batch size), C οι κατηγορίες, και H, W, D οι χωρικές διαστάσεις ύψους, πλάτους και βάθους αντίστοιχα.

- Επιλέγουμε τα κανάλια δύο κύριων κλάσεων νεφρού και όγκου αντίστοιχα:

$$p_1 = \mathbf{probs}[:, 1, :, :, :], \quad p_2 = \mathbf{probs}[:, 2, :, :, :] \quad (2.7)$$

- Ορίζουμε τις αντίστοιχες level-set συναρτήσεις:

$$\phi_1 = p_1 - \tau, \quad \phi_2 = p_2 - \tau, \quad \text{με } \tau = 0.5 \quad (2.8)$$

- Μετατρέπουμε τις ετικέτες αναφοράς (ground truth) σε δυαδικές μάσκες:

$$M_1(x) = \begin{cases} 1, & \text{αν } gt(x) = 1 \\ 0, & \text{διαφορετικά} \end{cases} \quad M_2(x) = \begin{cases} 1, & \text{αν } gt(x) = 2 \\ 0, & \text{διαφορετικά} \end{cases} \quad (2.9)$$

- Υπολογίζουμε τις απώλειες L_i ως τη μέση τιμή της απόλυτης διαφοράς:

$$L_i = \frac{1}{N} \sum_x |\phi_i(x) - M_i(x)|, \quad \text{για } i = 1, 2 \quad (2.10)$$

- Ορίζουμε την τελική συνάρτηση απώλειας ως το άθροισμα των δύο όρων:

$$L_{\text{LSF}} = L_1 + L_2 \quad (2.11)$$

Η χρήση της LSF Loss ενθαρρύνει το δίκτυο να διαχωρίζει με μεγαλύτερη σαφήνεια τις περιοχές προσκηνίου από το φόντο, βελτιώνοντας την ποιότητα και τη σαφήνεια των ορίων στην τελική τμηματοποίηση.

Αλγόριθμος 2.1 Υπολογισμός της Level Set Loss (LSF)

Require: Προβλέψεις logits σχήματος (N, C, H, W, D) , αληθείς ετικέτες gt σχήματος (N, H, W, D) , κατώφλι $\tau = 0.5$

```
1:  $probs \leftarrow \text{softmax}(logits)$ 
2:  $L_{LSF} \leftarrow 0$ 
3: for κάθε foreground κλάση  $c$  (π.χ.  $c = 1$ : νεφρός,  $c = 2$ : όγκος) do
4:    $\phi_c \leftarrow probs[:, c, :, :, :] - \tau$ 
5:    $M_c \leftarrow (gt == c)$ 
6:    $L_c \leftarrow \text{mean}(|\phi_c - M_c|)$ 
7:    $L_{LSF} \leftarrow L_{LSF} + L_c$ 
8: end for
9: return  $L_{LSF}$ 
```

2.2.6 Εισαγωγή στην Geodesic Active Contour και τη δική μας προσέγγιση

Η μέθοδος Geodesic Active Contours (GAC) είναι μια κλασική ενεργειακή προσέγγιση στην κατάτμιση εικόνων, που αποσκοπεί στην ανακάλυψη του περιγράμματος αντικειμένων ως καμπύλης ελάχιστου μήκους, προσαρμοσμένης στην τοπική γεωμετρία και τις ακμές της εικόνας. Η κύρια έννοια είναι η ελαχιστοποίηση ενός ενεργειακού λειτουργικού που συνδυάζει το μήκος της καμπύλης με μια συνάρτηση ένδειξης ακμών, η οποία καθοδηγεί το περίγραμμα να συγκλίνει σε περιοχές με έντονο gradient (δηλαδή όρια αντικειμένων). Η εφαρμογή της signed distance function (λειτουργίας επιπέδου) διευκολύνει την ευέλικτη αναπαράσταση και εξέλιξη του περιγράμματος, επιτρέποντας τη διαχείριση σύνθετων τοπολογιών και γεωμετρικών περιορισμών [23]. Πιο συγκεκριμένα ορίζεται :

$$L_{GAC}(\phi) = \int_{\Omega} g_I(\mathbf{x}) \delta_{\epsilon}(\phi(\mathbf{x})) |\nabla \phi(\mathbf{x})| d\mathbf{x} \quad (2.12)$$

όπου:

- Ω ορίζεται ως ο χώρος της εικόνας,
- $\phi(\mathbf{x})$ η λειτουργία επιπέδου (level set function),
- $\nabla \phi$ το gradient της ϕ ,

- $g_I(\mathbf{x})$ η συνάρτηση ένδειξης ακμών, η οποία ορίζεται ως:

$$g_I(\mathbf{x}) = \frac{1}{1 + |\nabla (G_\sigma * I)(\mathbf{x})|^2} \quad (2.13)$$

με I την εικόνα εισόδου και G_σ ένα Gaussian φίλτρο με διασπορά σ .

$$\delta_\epsilon(\phi) = \frac{\epsilon}{\pi(\phi^2 + \epsilon^2)} \quad (2.14)$$

όπου ϵ είναι ένας μικρός θετικός αριθμός, και η συνάρτηση αυτή αποτελεί μια ομαλή προσέγγιση της δέλτα συνάρτησης (Dirac delta).

Παρά τα πλεονεκτήματα αυτά, η εκπαίδευση τέτοιων μοντέλων παραδοσιακά απαιτεί υψηλό υπολογιστικό κόστος και είναι ευαίσθητη στην αρχική τοποθέτηση του περιγράμματος.

Στην παρούσα εργασία, προτείνεται μια απλουστευμένη αλλά αποτελεσματική υλοποίηση του Geodesic Active Contour Loss, η οποία συνδυάζει έναν όρο μήκους βασισμένο στο μέτρο του gradient και έναν όρο περιοχής που επιβάλλει ποινές για τις αποκλίσεις της πρόβλεψης από τις ετικέτες, διαχωρίζοντας το εσωτερικό και το εξωτερικό της περιοχής ενδιαφέροντος. Αυτή η προσέγγιση ενσωματώνεται ως όρος απώλειας σε ένα νευρωνικό δίκτυο, διευκολύνοντας την εκπαίδευση σε προβλήματα κατάτμισης εικόνων χωρίς την απαίτηση πολύπλοκων υπολογισμών. Το αποτέλεσμα είναι ένα απλό, σταθερό και υπολογιστικά αποδοτικό loss, το οποίο διατηρεί τα θεμελιώδη χαρακτηριστικά της ενεργού περιγράμματος, οδηγώντας σε ομαλές και ακριβείς προβλέψεις [24]. Πιο συγκεκριμένα, περιγράφεται η μεθοδολογία υλοποίησης και ο αλγόριθμος της προτεινόμενης μεθόδου:

Εισάγουμε την ενέργεια του μοντέλου, που αποτελείται από δύο όρους: **length term** και **region term**.

Πρώτα όμως ορίζουμε τον πυρήνα Sobel για τον υπολογισμό παραγώγων σε κάθε διάσταση της εικόνας:

$$\text{sobel} = \left[\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} -1 & -2 & -1 \\ -2 & -4 & -2 \\ -1 & -2 & -1 \end{bmatrix} \right] \quad (2.15)$$

Αυτός ο πυρήνας εφαρμόζεται και στους τρεις άξονες (x , y , z) με κατάλληλες μεταθέσεις (transposes), δημιουργώντας έτσι φίλτρα 3D για κάθε κατεύθυνση. Αυτό αντιστοιχεί στην εύρεση των ακμών σε κάθε κλάση.

Υπολογισμός όρων για length και region

- Ο όρος μήκους υπολογίζει το μέτρο του gradient της πρόβλεψης (δηλαδή το μήκος του περιγράμματος), και
- Ο όρος περιοχής μετρά την απόκλιση της πρόβλεψης από τις ετικέτες (label) τόσο στο εσωτερικό όσο και στο εξωτερικό του αντικειμένου.

Η συνάρτηση μήκους ορίζεται ως:

$$\text{length} = \frac{1}{N} \sum_{i=1}^N \frac{\sqrt{(\partial_x p_i)^2 + (\partial_y p_i)^2 + (\partial_z p_i)^2 + \epsilon}}{\max \left(\sqrt{(\partial_x p)^2 + (\partial_y p)^2 + (\partial_z p)^2 + \epsilon} \right) - \min \left(\sqrt{(\partial_x p)^2 + (\partial_y p)^2 + (\partial_z p)^2 + \epsilon} \right) + \epsilon} \quad (2.16)$$

όπου p_i είναι η πρόβλεψη της πιθανότητας για το voxel i , $\partial_x, \partial_y, \partial_z$ οι παράγωγοι κατά τις αντίστοιχες διαστάσεις, και ϵ μια μικρή σταθερά για αριθμητική σταθερότητα.

Ο όρος περιοχής διαμορφώνεται από δύο επιμέρους μέρη:

$$\text{region}_{in} = \left| \sum_i p_i (l_i - c_{in})^2 \right|, \quad \text{region}_{out} = \left| \sum_i (1 - p_i) (l_i - c_{out})^2 \right| \quad (2.17)$$

όπου l_i η πραγματική ετικέτα (label) του voxel i , και c_{in}, c_{out} είναι δύο σταθερές αναφοράς που ορίζονται ως:

$$c_{in} = 1, \quad c_{out} = 0 \quad (2.18)$$

Οι σταθερές αυτές αντιπροσωπεύουν τις ιδανικές τιμές μέσα στην περιοχή του αντικειμένου (c_{in}) και στο φόντο ή έξω από το αντικείμενο (c_{out}) αντίστοιχα.

Το τελικό region loss είναι:

$$\text{region} = \text{region}_{in} + \text{region}_{out} \quad (2.19)$$

Η συνολική απώλεια συνδυάζεται ως:

$$L = \alpha \times \text{region} + \text{length} \quad (2.20)$$

όπου α είναι υπερπαραμέτρος που ελέγχει τη σχετική βαρύτητα των δύο όρων.

Η προσέγγιση αυτή, αν και απλούστερη, επιτυγχάνει την προώθηση ομαλών και ακριβών περιγραμμάτων, ενώ παραμένει υπολογιστικά ελαφριά, κατάλληλη για εκπαίδευση νευρωνικών δικτύων χωρίς το κόστος που επιφέρει η πλήρης υλοποίηση της Geodesic Active Contour Loss.

Αλγόριθμος 2.2 Υπολογισμός Geodesic Active Contour Loss (LGAC)

Require: Προβλέψεις πιθανοτήτων p σχήματος (N, C, H, W, D) ,

Αληθείς ετικέτες label σχήματος (N, C, H, W, D) ,

Υπερπαραμέτρος βαρύτητας α

- 1: Ορισμός 3D Sobel φίλτρων για τους άξονες x, y, z
 - 2: $\text{grd}_x \leftarrow \text{Sobel}_x(p)$
 - 3: $\text{grd}_y \leftarrow \text{Sobel}_y(p)$
 - 4: $\text{grd}_z \leftarrow \text{Sobel}_z(p)$
 - 5: $\text{length} \leftarrow \text{mean} \left(\sqrt{\text{grd}_x^2 + \text{grd}_y^2 + \text{grd}_z^2 + 10^{-8}} \right)$
 - 6: $\text{label} \leftarrow \text{label.float}()$
 - 7: $c_{\text{in}} \leftarrow \text{ones_like}(p)$
 - 8: $c_{\text{out}} \leftarrow \text{zeros_like}(p)$
 - 9: $\text{region}_{\text{in}} \leftarrow \text{mean} (|p \cdot (\text{label} - c_{\text{in}})|^2)$
 - 10: $\text{region}_{\text{out}} \leftarrow \text{mean} (|(1 - p) \cdot (\text{label} - c_{\text{out}})|^2)$
 - 11: $\text{region} \leftarrow \text{region}_{\text{in}} + \text{region}_{\text{out}}$
 - 12: $L_{\text{GAC}} \leftarrow \alpha \times \text{region} + \text{length}$
 - 13: **return** L_{GAC}
-

2.2.7 Μετρικές Αξιολόγησης (Evaluation Metrics)

Οι μετρικές αξιολόγησης (evaluation metrics) αποτελούν ουσιώδες εργαλείο για την αποτίμηση της απόδοσης ενός μοντέλου μηχανικής μάθησης. Σε αντίθεση με τη συνάρτηση απώλειας που χρησιμοποιείται για την εκπαίδευση, οι μετρικές εφαρμόζονται κυρίως στη φάση της αξιολόγησης (validation/test) και επιτρέπουν τη σύγκριση μοντέλων, τη βελτίωση υπερπαραμέτρων και τη λήψη αποφάσεων σχετικά με την ανάπτυξη και παραγωγική χρήση του μοντέλου.

Η μετρική αξιολόγησης που χρησιμοποιείται συχνότερα για image segmentation tasks είναι το Dice Similarity Coefficient (Τύπος (2.1)) όπως ήδη αναφέρθηκε και στην προηγούμενη ενότητα.

Ο δείκτης Dice (Τύπος (2.1)) λαμβάνει τιμές στο διάστημα $[0, 1]$.

- $Dice = 0$: καμία επικάλυψη μεταξύ προβλεπόμενης και πραγματικής μάσκας.
- $Dice = 1$: τέλεια ταύτιση (πλήρης επικάλυψη).
- Στην πράξη, για 3D ιατρική τμηματοποίηση:
 - $Dice > 0.7$ θεωρείται ικανοποιητικό,
 - $Dice > 0.8$ καλό,
 - $Dice > 0.9$ εξαιρετικό.

2.2.8 Εκπαίδευση (Training)

Η διαδικασία εκπαίδευσης ενός νευρωνικού δικτύου αποτελεί τον πυρήνα της μηχανικής μάθησης και περιλαμβάνει την προσαρμογή των παραμέτρων του μοντέλου ώστε να ελαχιστοποιείται η συνάρτηση απώλειας. Η εκπαίδευση πραγματοποιείται μέσω επαναληπτικής ενημέρωσης των βαρών, με χρήση του αλγορίθμου οπισθοδιάδοσης και κάποιου βελτιστοποιητή, βάσει της απόκλισης μεταξύ της προβλεπόμενης και της πραγματικής τιμής.

Η διαδικασία διακρίνεται σε τρεις κύριες φάσεις: προπόνηση (training), επικύρωση (validation) και δοκιμή (testing). Κατά την προπόνηση, το μοντέλο εκπαιδεύεται σε δεδομένα που έχει «δει» και τα οποία χρησιμοποιούνται για τη βελτίωση των παραμέτρων. Η επικύρωση γίνεται σε ένα ξεχωριστό σύνολο που δεν συμμετέχει άμεσα στη μάθηση, με στόχο την αξιολόγηση της απόδοσης και τη ρύθμιση υπερπαραμέτρων. Τέλος, η δοκιμή αξιολογεί το μοντέλο σε εντελώς άγνωστα δεδομένα για να μετρηθεί η ικανότητά του στη γενίκευση.

Η προπόνηση οργανώνεται σε κύκλους που ονομάζονται epochs, κατά τους οποίους το δίκτυο διατρέχει ολόκληρο το σύνολο εκπαίδευσης μία φορά. Κάθε epoch αποτελείται από πολλαπλά βήματα, στα οποία το σύνολο δεδομένων χωρίζεται σε μικρές παρτίδες, γνωστές ως mini-batches. Το μέγεθος του batch καθορίζει τον αριθμό των δειγμάτων που χρησιμοποιούνται σε κάθε βήμα ενημέρωσης. Μικρότερα batches εισάγουν στοχαστικότητα και βοηθούν στη γενίκευση, ενώ μεγαλύτερα πα-
ρέχουν σταθερότερες εκτιμήσεις αλλά απαιτούν περισσότερη μνήμη.

Κατά τη διάρκεια της εκπαίδευσης είναι σημαντικό να εντοπίζονται φαινόμενα όπως η υπερπροσαρμογή (overfitting), όπου το μοντέλο αποδίδει εξαιρετικά στο σύνολο εκπαίδευσης αλλά αποτυγχάνει στο validation set. Το φαινόμενο αυτό υποδηλώνει ότι το μοντέλο έχει μάθει όχι μόνο τα μοτίβα αλλά και τον θόρυβο

των δεδομένων. Στον αντίποδα βρίσκεται η υποπροσαρμογή (underfitting), όπου το μοντέλο αποτυγχάνει να αποδώσει ακόμη και στα δεδομένα εκπαίδευσης, λόγω ανεπαρκούς χωρητικότητας ή εσφαλμένης αρχιτεκτονικής.

Για την αντιμετώπιση αυτών των προβλημάτων χρησιμοποιούνται στρατηγικές όπως η πρόωρη διακοπή εκπαίδευσης (early stopping), όπου η διαδικασία τερματίζεται όταν η απόδοση στο σύνολο επικύρωσης παύει να βελτιώνεται. Επιπλέον, εφαρμόζονται τεχνικές τακτικής όπως το dropout, η κανονικοποίηση (L1, L2), και η χρήση callbacks για δυναμικό έλεγχο παραμέτρων, όπως ο ρυθμός μάθησης.

Η παρακολούθηση της απόδοσης ανά epoch μέσω καμπυλών εκπαίδευσης (loss και accuracy curves) είναι κρίσιμη για τη διάγνωση του τρόπου με τον οποίο το μοντέλο μαθαίνει. Η διατήρηση χαμηλής απώλειας στο training set αλλά υψηλής στο validation set υποδεικνύει overfitting, ενώ σταθερά υψηλές απώλειες και στα δύο μπορεί να σημαίνουν underfitting ή εσφαλμένες υπερπαραμέτρους.

Η επιλογή κατάλληλου αριθμού epochs, η σωστή διαχείριση του learning rate και η ισορροπία μεταξύ χωρητικότητας και regularization συνιστούν τους βασικούς άξονες για την επιτυχή εκπαίδευση ενός μοντέλου. Συνολικά, η εκπαίδευση δεν είναι μία αυτόματη διαδικασία αλλά μια δυναμική διαδικασία βελτιστοποίησης, που απαιτεί συνεχή παρακολούθηση, ρύθμιση και κατανόηση της συμπεριφοράς του μοντέλου σε όλα τα στάδια.

ΚΕΦΑΛΑΙΟ 3

ΜΕΘΟΔΟΛΟΓΙΑ: ΑΛΓΟΡΙΘΜΟΣ ΚΑΙ ΑΡΧΙΤΕΚΤΟΝΙΚΕΣ ΜΟΝΤΕΛΩΝ

3.1 Σύνολο Δεδομένων και Προεπεξεργασία

3.2 Αρχιτεκτονική Μοντέλου

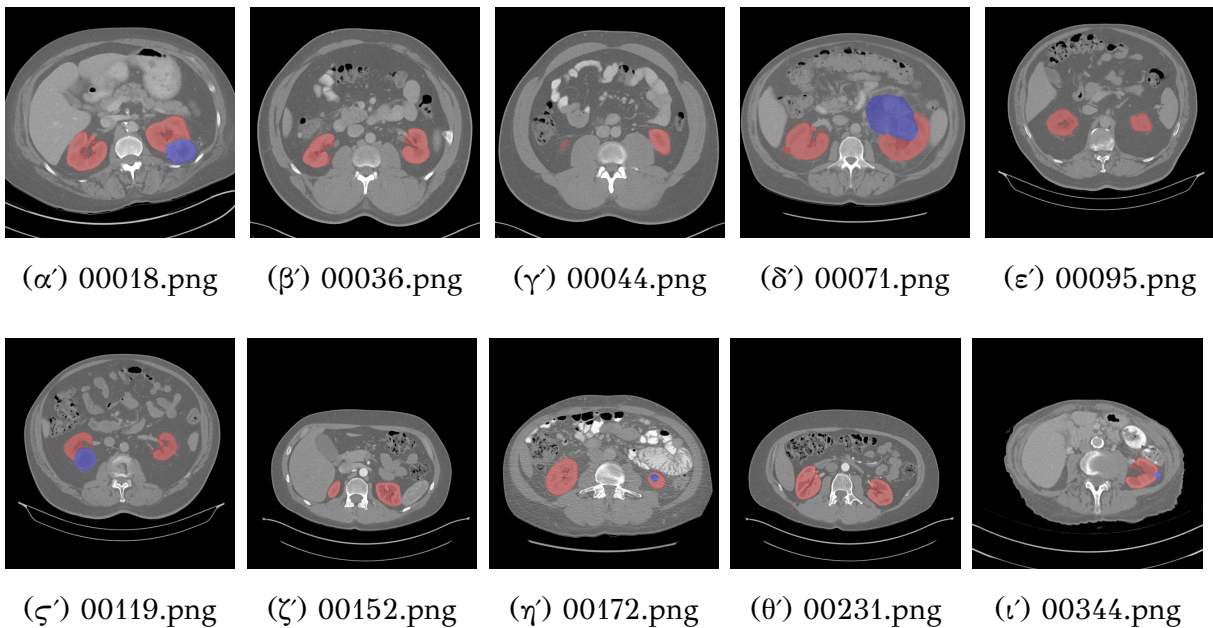
Στο κεφάλαιο αυτό αναλύονται λεπτομερώς το σύνολο δεδομένων που χρησιμοποιήθηκε και οι τεχνικές προεπεξεργασίας τους, καθώς και η αρχιτεκτονική του προτεινόμενου νευρωνικού δικτύου για την τρισδιάστατη τμηματοποίηση νεφρού και όγκου.

3.1 Σύνολο Δεδομένων και Προεπεξεργασία

3.1.1 Σύνολο Δεδομένων

Για την ανάπτυξη και αξιολόγηση του μοντέλου χρησιμοποιήθηκε το δημόσιο σύνολο δεδομένων **KiTS19 (Kidney Tumor Segmentation Challenge 2019)** [25]. Το εν λόγω dataset περιλαμβάνει 300 περιπτώσεις ασθενών με προεπεξεργασμένες τρισδιάστατες αξονικές τομογραφίες (CT scans), καθώς και αντίστοιχες χειροκίνητα επισημασμένες μάσκες (ground truth) που προσδιορίζουν την περιοχή του νεφρού και του όγκου. Από αυτές τις περιπτώσεις από 0 έως 209 διαθέτουν δημόσια προσβάσιμο ground truth και χρησιμοποιήθηκαν για την εκπαίδευση και επικύρωση

του μοντέλου. Αντίθετα, οι περιπτώσεις με από 210 έως 299 έχουν πρόσβαση στο αντίστοιχο ground truth μόνο οι διοργανωτές του διαγωνισμού και χρησιμοποιούνται αποκλειστικά για την τελική αξιολόγηση (challenge leaderboard), χωρίς να παρέχεται το αντίστοιχο ground truth στους συμμετέχοντες. Οι περιπτώσεις διαχωρίστηκαν σε σύνολο εκπαίδευσης και ελέγχου τυχαία μέσω της μεθόδου cross validation. Συγκεκριμένα, τα περιστατικά από 0 έως 209 χωρίστηκαν τυχαία όπου σε κάποια έγινε η εκπαίδευση και στα υπόλοιπα το test. Επιπρόσθετα, αποκλείστηκαν οι περιπτώσεις με αριθμούς αναγνώρισης 15 και 37 καθώς θεωρήθηκαν ελαττωματικά από τους διοργανωτές και τα περιστατικά 23, 68, 125 και 133 αφαιρέθηκαν με βάση τη δημοσίευση του Isensee et al. (2019) [26], καθώς διαπιστώθηκε σημαντική ασυμφωνία μεταξύ των προβλέψεων των νευρωνικών δικτύων και των παρεχόμενων annotations, καθιστώντας αδύνατο να κριθεί με αξιοπιστία η ορθότητα. Το σύνολο δεδομένων KiTS19 παρέχεται σε μορφή NIFTI (.nii.gz) και συνοδεύεται από με-
τα δεδομένα σε μορφή JSON για κάθε ασθενή. Παρακάτω παρουσιάζονται κάποιες εικόνες σε δισδιάστατη μορφή από το σύνολο δεδομένων KiTS19.



Σχήμα 3.1: Παράδειγμα 10 δισδιάστατων τομών CT νεφρού από το KiTS19 dataset από διαφορετικά cases.

3.1.2 Προεπεξεργασία Δεδομένων

Η προεπεξεργασία των δεδομένων κρίθηκε απαραίτητη προκειμένου να διασφαλιστεί η ομοιομορφία και η ποιότητα των εισόδων προς το νευρωνικό δίκτυο. Για την

εκπαίδευση ενός μοντέλου όπως το 3D U-Net σε αυτά τα δεδομένα, εφαρμόστηκε μια συγκεκριμένη αλληλουχία προεπεξεργασίας και τεχνικών data augmentation (ενίσχυσης δεδομένων), με στόχο την τυποποίηση των εισόδων και τη βελτίωση της ικανότητας γενίκευσης του μοντέλου. Ακολουθεί λεπτομερής περιγραφή των φάσεων αυτής της διαδικασίας.

Κανονικοποίηση Έντασης

Οι εικόνες CT υποβλήθηκαν σε κανονικοποίηση έντασης (intensity normalization) σε μονάδες Hounsfield και κυμαίνονται στο διάστημα $[-79, 304]$ σαν ένα “παράθυρο” που επικεντρώνεται στις σχετικές τιμές για τον νεφρό και αποκλείει εξαιρετικά χαμηλές ή υψηλές τιμές εκτός του ενδιαφέροντος. Στη συνέχεια, εκτελείται η κανονικοποίηση όπου αφαιρείται η μέση τιμή (mean) $\mu = 101.0$ και διαιρείται η τυπική απόκλιση (std) $\sigma = 76.9$. Με αυτό το τρόπο επιτυγχάνεται οι τιμές των εικονων να είναι σε συγκρίσιμη κλίμακα και ταυτόχρονα βοηθάει στη σταθερότερη και ταχύτερη σύγκλιση του μοντέλου.

Padding

Μετά την κανονικοποίηση, κάθε εικόνα επεκτείνεται περιφερειακά με σταθερό padding. Σκοπός είναι να εξασφαλιστεί συνεπές μέγεθος εισόδου και να αποτραπεί η απώλεια πληροφορίας κατά την επιλογή υποτομών κοντά στα όρια της εικόνας ή κατά την εφαρμογή περιστροφών και μετατοπίσεων. Οι νέες περιοχές που προστίθενται γύρω από την εικόνα συμπληρώνονται με σταθερή τιμή έντασης υπολογισμένη ως $\frac{\min - \mu}{\sigma}$. Κατά συνέπεια, κάθε κενό τμήμα εκτός των ορίων της αρχικής τομογραφίας (λόγω padding ή μετασχηματισμών) συμπληρώνεται με μια ουδέτερη τιμή φόντου(background), η οποία δεν προκαλεί τεχνητές υψηλές ή χαμηλές εντάσεις στο δίκτυο. Ανάλογο padding εφαρμόζεται και γύρω από τις ετικέτες, όπου σε αυτή την περίπτωση, χρησιμοποιείται η τιμή 0 (background label) για όλες τις προστιθέμενες περιοχές, ώστε οι επεκτάσεις εκτός της αρχικής περιοχής να θεωρούνται ως υπόβαθρο.

Τυχαίο και Κεντρικό Cropping για τα δεδομένα Εκπαίδευσης

Εξαιτίας του μεγάλου όγκου των 3D δεδομένων, το μοντέλο δεν εκπαιδεύεται στο πλήρες volume, αλλά σε μικρά υπο-volumes (patches) σταθερού μεγέθους $160 \times 160 \times 80$

voxels. Έχει αποδειχθεί ότι ένα τέτοιο μέγεθος (περίπου 80 τομές με 160×160 pixels ανά τομή) προσφέρει επαρκές οπτικό πεδίο για την πλήρη απεικόνιση του νεφρού και των παρακείμενων οργάνων, διατηρώντας ταυτόχρονα την αναγκαία λεπτομέρεια στην ανάλυση [26]. Επίσης να σημειωθεί ότι πριν από το cropping, εφαρμόζεται padding με τιμή φόντου για volumes μικρότερους από το επιθυμητό μέγεθος. Κατά τη διάρκεια της εκπαίδευσης, για κάθε εικόνα επιλέγεται τυχαία μια υποτομή, δηλαδή εκτελείται τυχαίο cropping $160 \times 160 \times 80$ από τη συνολική εικόνα που έχει προηγηθεί padding. Η επιλογή αυτή ενισχύει την ποικιλία των χωρικών περιοχών που αντιλαμβάνεται το δίκτυο, επιτρέποντάς του να εκπαιδευτεί σε διαφορετικά τμήματα του οργάνου και του περιβάλλοντός του κάθε φορά. Στη συνέχεια, σε κάθε επιλεγμένη υποτομή εφαρμόζονται οι μετασχηματισμοί data augmentation (που περιγράφονται κατωτέρω). Δεδομένου ότι ορισμένοι από αυτούς τους μετασχηματισμούς (π.χ. περιστροφές ή κλίμακα) ενδέχεται να μετατοπίσουν ή να περιστρέψουν το patch, μπορεί να προκύψουν κενά ή εξωτερικές περιοχές γύρω από αυτό. Για τον λόγο αυτό, μετά την εφαρμογή των augmentations, εκτελείται μια σταθερή κεντρική αποκοπή (center-cropping), επαναφέροντας το τελικό patch στις διαστάσεις $160 \times 160 \times 80$. Η κεντρική αποκοπή εξαλείφει τυχόν περιθώρια ή κενό χώρο που προκύπτει από τους μετασχηματισμούς, διασφαλίζοντας ότι το εισαγόμενο patch στο δίκτυο περιέχει κυρίως ουσιώδες περιεχόμενο και έχει το προκαθορισμένο μέγεθος. Αυτή η διαδικασία

(τυχαίο cropping $\rightarrow$ augmentation $\rightarrow$ κεντρικό cropping)

εξασφαλίζει ότι το δίκτυο εκπαιδεύεται σε ποικίλες ανατομικές περιοχές, διατηρώντας παράλληλα τη δομή των δεδομένων.

Προεπεξεργασία Συνόλου Επαλήθευσης (Validation)

Για το σύνολο επαλήθευσης (validation set) δεν εφαρμόζονται τυχαίες περικοπές ή augmentations, στόχος είναι η αξιολόγηση του μοντέλου σε δεδομένα ίδια με την είσοδο. Κάθε volume του validation υποβάλλεται σε μία μόνο κεντρική αποκοπή στις διαστάσεις $256 \times 256 \times 128$. Η εκτενέστερη αυτή περιοχή εισόδου επιλέχθηκε προκειμένου να καλύψει πλήρως την περιοχή ενδιαφέροντος (οι νεφροί και οι όγκοι εντοπίζονται συνήθως κεντρικά στο ογκώδες σύνολο. Οπότε, κατά την επαλήθευση, το δίκτυο αποκτά το πλήρες σχετικό τμήμα της τομογραφίας (κεντρικό crop) χωρίς

αυθαίρετες μεταβολές, διευκολύνοντας την αξιολόγηση της απόδοσης των segmentations στο πλήρες οργάνο.

Τεχνικές Εμπλουτισμού Δεδομένων (Data Augmentation)

Για την ενίσχυση της γενίκευσης του μοντέλου και την αποφυγή υπερπροσαρμογής εφαρμόστηκαν διάφορες τεχνικές που στοχεύουν στην τεχνητή αύξηση της ποικιλίας των δεδομένων και στη βελτίωση της γενίκευσης του μοντέλου. Συγκεκριμένα, χρησιμοποιήθηκαν οι ακόλουθοι μετασχηματισμοί (σε παρένθεση αναφέρεται η πιθανότητα εφαρμογής κάθε τεχνικής σε ένα συγκεκριμένο patch):

- **Τυχαία περιστροφή** ($\pm 15^\circ$, $p = 0.2$): Περιστρέφεται το patch κατά τυχαία γωνία έως $\pm 15^\circ$ μοιρών γύρω από τον κάθε έναν από τους τρεις άξονες. Η συγκεκριμένη τεχνική εφαρμόζεται με πιθανότητα ίση με 20%. Με αποτέλεσμα το δίκτυο να προσαρμόζεται καλύτερα σε μικρές στροφές ή σε διαφορετικές ευθυγραμμίσεις των νεφρών μεταξύ των ασθενών.
- **Τυχαία κλιμάκωση (Zoom)** (0.85–1.25, $p = 0.2$): Ομοιομορφική μεγέθυνση/σμίκρυνση με παράγοντα επιλεγμένο στο διάστημα $[0.85, 1.25]$ και με πιθανότητα 20%. Αυτή η αυξομείωση μεγέθους βοηθά το μοντέλο να γενικεύει σε περιπτώσεις όπου τα όργανα εμφανίζονται μεγαλύτερα ή μικρότερα από το μέσο όρο.
- **Γκαουσιανός θόρυβος (Gaussian Noise)** (variance 0–0.1, $p = 0.1$): Προσθήκη τυχαίου θορύβου θεωρούμενου από κανονική κατανομή με διασπορά στο εύρος $[0, 0.1]$. Αυτή η παρέμβαση ενισχύει την ανθεκτικότητα του δικτύου απέναντι σε θορυβώδη δεδομένα, αποτρέποντάς το από το να εστιάζει υπερβολικά σε λεπτομέρειες θορύβου.
- **Φωτεινότητα/Αντίθεση** (multiplier 0.75–1.25, $p = 0.15$): Πολλαπλασιασμός των εντάσεων μεταξύ του διαστήματος $[0.75, 1.25]$, προσομοιώνοντας αλλαγές σε brightness/contrast. Το μοντέλο αποκτά ανθεκτικότητα σε μεταβολές της γενικής φωτεινότητας και της κατανομής των εντάσεων.
- **Διόρθωση Gamma** (0.7–1.5, $p = 0.1$): Προσαρμόζει μη-γραμμικά την κατανομή των εντάσεων μέσω της διόρθωσης gamma. Ένας εκθέτης γ επιλέγεται τυχαία από το εύρος (0.7, 1.5) και οι τιμές έντασης του patch ανυψώνονται στη δύναμη

αυτή: για $\gamma < 1$ οι σκοτεινές περιοχές φωτίζονται, ενισχύοντας τις λεπτομέρειες σε σκιές, ενώ για $\gamma > 1$ οι φωτεινές περιοχές γίνονται πιο σκοτεινές. Πριν από την εφαρμογή της, υπάρχει η δυνατότητα αντιστροφής της εικόνας, όπου οι υψηλές και χαμηλές τιμές ανταλλάσσονται. Στις κανονικοποιημένες τιμές, αυτό ισοδυναμεί με την ανάκλαση των εντάσεων γύρω από το μηδέν, με τις φωτεινές περιοχές να γίνονται σκοτεινές και το αντίστροφο. Μετά την εφαρμογή του gamma transform και τυχόν αναστροφής, οι τιμές επανακλίμακονται για να διατηρηθούν τα βασικά στατιστικά της αρχικής εικόνας, όπως η μέση τιμή και η διασπορά, εξασφαλίζοντας ότι η συνολική φωτεινότητα και αντίθεση του patch παραμένει συγκρίσιμη με την αρχική. Με αυτόν τον τρόπο το μοντέλο αποκτά ανοχή σε διαφορετικές ρυθμίσεις συσκευών και βελτιωμένη αντίληψη λεπτομερειών σε σκοτεινές και φωτεινές περιοχές.

Τέλος, σημειώνεται ότι ο καθορισμένος αριθμός seed (777) εφαρμόστηκε και στους τρεις κρίσιμους τομείς που περιγράφονται παρακάτω:

- **Data Augmentation** κυρίως για cropping και γεωμετρικές/έντασης τροποποιήσεις
- **Στη διαδικασία εκπαίδευσης** συμπεριλαμβανομένης της σειράς επεξεργασίας δεδομένων
- **Στην αρχικοποίηση βαρών** του μοντέλου (μέσω Kaiming Normal)

Η ρύθμιση αυτή διασφαλίζει ότι τα τυχαία βήματα (από την προεπεξεργασία έως την εκπαίδευση) μπορούν να αναπαραχθούν επακριβώς, κάτι κρίσιμο για την αξιολόγηση και τη σύγκριση πειραμάτων.

3.2 Αρχιτεκτονική Μοντέλου

Για την τμηματοποίηση των ανατομικών περιοχών νεφρού και όγκου αξιοποιήθηκε η αρχιτεκτονική *3D U-Net*, και συγκεκριμένα μια απλοποιημένη εκδοχή της, όπως αυτή περιγράφεται στην υλοποίηση “An attempt at beating the 3D U-Net”[26]. Η επιλογή του συγκεκριμένου μοντέλου βασίστηκε στη δυνατότητά του να χειρίζεται δεδομένα τριών διαστάσεων, διατηρώντας παράλληλα την ικανότητα εξαγωγής χωρικής πληροφορίας σε πολλαπλά επίπεδα ανάλυσης. Το μοντέλο αποτελείται από

δύο συμμετρικά μέρη: το contracting path (αριστερή πλευρά), το οποίο λειτουργεί ως κωδικοποιητής (encoder) και το expansive path (δεξιά πλευρά), που περιέχει τον αποκωδικοποιητή (decoder).

Η αρχιτεκτονική του 3D U-Net [27] αποτελείται από ένα συμμετρικό encoder-decoder σχήμα, όπου τα χαρακτηριστικά εξάγονται σταδιακά μέσω διαδοχικών επιπέδων συνελίξεων και υποδειγματοληψίας (downsampling), και στη συνέχεια ανακατασκευάζονται χωρικά μέσω upsampling, διατηρώντας την τοπική πληροφόρηση μέσω skip connections. Το συγκεκριμένο σχήμα έχει αποδειχθεί ιδιαίτερα αποτελεσματικό για την τμηματοποίηση πολύπλοκων ανατομικών δομών, παρέχοντας υψηλή ακρίβεια και χωρική συνέπεια στις εξόδους του.

3.2.1 Κωδικοποιητής (Encoder)

Η διαδρομή του Encoder στο μοντέλο μας αποτελείται από 5 διαδοχικά επίπεδα επεξεργασίας, όπου τα δεδομένα εισόδου συμπιέζονται προοδευτικά σε μικρότερες χωρικές διαστάσεις και περισσότερα κανάλια χαρακτηριστικών. Κάθε επίπεδο εφαρμόζει διπλή συνελικτική επεξεργασία (Double Convolution): δύο διαδοχικές 3D συνελίξεις με πυρήνα $3 \times 3 \times 3$, καθεμία ακολουθούμενη από Instance Normalization και μη γραμμική ενεργοποίηση Leaky ReLU. Η χρήση Instance Normalization αντί Batch Normalization βελτιώνει τη συνέπεια των ενεργοποιήσεων μεταξύ των διαδικασιών training και test, προσφέροντας πιο σταθερή συμπεριφορά ειδικά σε ιατρικές εικόνες. Σε κάθε επίπεδο πραγματοποιείται υποδειγματοληψία (downsampling) μέσω συνελικτικού πυρήνα με βήμα stride αντί για pooling, προκειμένου να μειωθεί η ανάλυση του feature map και παράλληλα να αυξηθεί ο αριθμός των καναλιών. Ειδικότερα, ο πρώτος downsampling πυρήνας έχει stride (2,2,1) (μείωση κατά 2 στις χωρικές διαστάσεις ύψους/πλάτους, διατηρώντας το βάθος ίδιο), ενώ οι επόμενοι downsampling πυρήνες έχουν stride (2,2,2) (μείωση κατά 2 και στις τρεις διαστάσεις). Τα output feature maps κάθε encoder level διατηρούνται και ενώνονται (concatenate) με το αντίστοιχο decoder block, πριν από το διπλό convolution (Skip Connections). Ως συναρτήσεις ενεργοποίησης εντός των συνελικτικών blocks χρησιμοποιήθηκαν οι *Leaky ReLU*, οι οποίες υπερτερούν έναντι των απλών ReLU στη διαχείριση αρνητικών τιμών, περιορίζοντας το πρόβλημα των “νεκρών νευρώνων”. Στην έξοδο του δικτύου εφαρμόστηκε η συνάρτηση *softmax*, παρέχοντας πιθανοκρατική ερμηνεία για την ταξινόμηση κάθε voxel. Για την κανονικοποίηση της

εξόδου κάθε επιπέδου εφαρμόστηκε *Instance Normalization*. Ο αρχικοποιητής των βαρών του δικτύου ήταν ο *Kaiming Normal*, ο οποίος διασφαλίζει σταθερή διασπορά των ενεργοποιήσεων σε βάθος και διευκολύνει τη σύγκλιση του μοντέλου. Η συνολική αρχιτεκτονική έχει σχεδιαστεί έτσι ώστε να ισορροπεί μεταξύ υπολογιστικής αποδοτικότητας και υψηλής ακρίβειας, ενώ προσαρμόζεται κατάλληλα στα χαρακτηριστικά του dataset KiTS19.

Οι διαστάσεις των feature maps σε κάθε επίπεδο του encoder είναι οι εξής:

- Encoder block 0: `torch.Size([1, 30, 160, 160, 80])`
- Encoder block 1: `torch.Size([1, 60, 80, 80, 80])`
- Encoder block 2: `torch.Size([1, 120, 40, 40, 40])`
- Encoder block 3: `torch.Size([1, 240, 20, 20, 20])`
- Encoder block 4: `torch.Size([1, 320, 10, 10, 10])`

3.2.2 Επίπεδο Συστολής (Bottleneck)

Το bottleneck επίπεδο βρίσκεται στη βάση “U” του δικτύου αποτελεί το πιο συμπιεσμένο σημείο του δικτύου και λειτουργεί ως “σύνδεσμος” μεταξύ κωδικοποιητή και αποκωδικοποιητή έχει διαστάσεις:

- Bottleneck block: `torch.Size([1, 320, 5, 5, 5])`

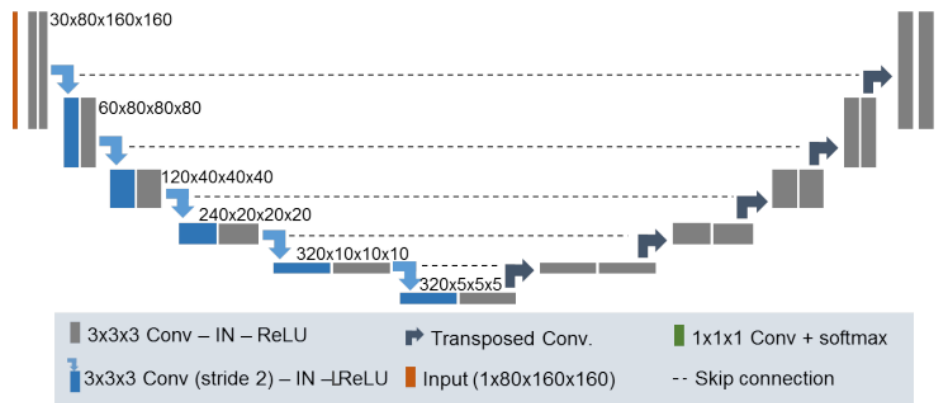
3.2.3 Αποκωδικοποιητής (Decoder)

Αντίστοιχα η διαδρομή του Decoder αποτελείται επίσης από 5 διαδοχικά επίπεδα επεξεργασίας, όπου τα δεδομένα “αποσυμπιέζονται” ώστε να επαναφέρουν την ανάλυση με αντίστροφα βήματα, συνδυάζοντας κάθε upsampling με τα αντίστοιχα feature maps του encoder μέσω skip connections. Κάθε επίπεδο εφαρμόζει transposed convolution όπου υλοποιεί το upsampling με τα ίδια strides όπως πριν αντίστροφα (πρώτα (2,2,2) και μετά (2,2,1)) για αύξηση διαστάσεων, στην συνέχεια ακολουθεί concatenation με τα αντίστοιχα αποθηκευμένα feature maps του encoder. Ακολουθεί μια διπλή συνελικτική επεξεργασία (double convolution). Η πορεία αυτή είναι αντιστρόφως συμμετρική της κωδικοποίησης. Στο τελευταίο επίπεδο εφαρμόζεται μία $1 \times 1 \times 1$ συνέλιξη όπου μετασχηματίζει τα 30 feature maps σε 3 κανάλια

εξόδου καθώς έχουμε 3 segmentation tasks. Στο τέλος της διαδικασίας τα feature maps περνούν απο τη συνάρτηση softmax. Οι διαστάσεις σε κάθε στάδιο του decoder είναι:

- Decoder block 0: Skip [1,320,10,10,10], Input [1,320,5,5,5]
- Decoder block 1: Skip [1,240,20,20,20], Input [1,320,10,10,10]
- Decoder block 2: Skip [1,120,40,40,40], Input [1,240,20,20,20]
- Decoder block 3: Skip [1,60,80,80,80], Input [1,120,40,40,40]
- Decoder block 4: Skip [1,30,160,160,80], Input [1,60,80,80,80]
- Final Conv: Output [1,3,160,160,80]

3.2.4 Διάγραμμα Αρχιτεκτονικής 3D U-Net



Σχήμα 3.2: Διάγραμμα της αρχιτεκτονικής 3D U-Net με τις διαστάσεις των feature maps και τις skip connections.

3.2.5 Συναρτήσεις Κόστους (Loss Functions)

Για την εκπαίδευση των μοντέλων επιλέχθηκαν σύνθετες συναρτήσεις κόστους, οι οποίες επιτρέπουν την ταυτόχρονη βελτιστοποίηση διαφορετικών πτυχών της απόδοσης. Συγκεκριμένα, αναπτύχθηκαν και συγκρίθηκαν δύο μοντέλα με διαφορετικό συνδυασμό loss functions. Το πρώτο μοντέλο συνδύασε την Cross-Entropy Loss με τη Dice Loss. Η Cross-Entropy είναι κατάλληλη για πολυταξινομητικά προβλήματα

και ευνοεί την ταχύτατη σύγκλιση, ενώ η Dice Loss στοχεύει στη βελτίωση της επικάλυψης μεταξύ των προβλεπόμενων και των πραγματικών περιοχών, κάτι που είναι ιδιαίτερα κρίσιμο σε προβλήματα τμηματοποίησης με μη ισομερώς κατανεμημένες κατηγορίες. Τα βάρη των δύο συνιστωσών ορίστηκαν ίσα, με τιμή 1.0, ώστε να επιτυγχάνεται ισόρροπη βαρύτητα μεταξύ ταξινόμησης και χωρικής συμφωνίας.

Το δεύτερο μοντέλο επεκτείνει περαιτέρω τον στόχο βελτιστοποίησης, ενσωματώνοντας επιπλέον δύο loss terms που βασίζονται σε γεωμετρικές και μορφολογικές ιδιότητες των δομών. Πιο συγκεκριμένα, προστέθηκαν οι δικές μας Level Set Loss και η Geodesic Active Contour Loss όπως τις ορίσαμε προηγουμένως. Η Level Set Loss λειτουργεί ως μορφολογικός περιορισμός, ευνοώντας τη συνεκτικότητα των ορίων των αντικειμένων, ενώ η Geodesic Active Contour Loss καθοδηγεί την πρόβλεψη προς γεωδαιτικές καμπύλες που προσεγγίζουν με μεγαλύτερη ακρίβεια τις πραγματικές ανατομικές οριοθετήσεις. Για τη σύνθεση της συνολικής loss function του δεύτερου μοντέλου χρησιμοποιήθηκαν τα εξής βάρη μετά από αρκετά πειράγματα : [1.0, 1.0, 0.001, 0.000005] αντίστοιχα για τις Cross-Entropy, Dice, Level Set και Geodesic Active Contour losses. Η επιλογή αυτών των συναρτήσεων κόστους δεν έγινε μόνο με βάση τη βιβλιογραφία, αλλά και με σκοπό την πειραματική διερεύνηση της επίδρασής τους στην ποιότητα της τμηματοποίησης σε σύνθετες ογκολογικές περιοχές όπως αυτές του KiTS19.

Συνοπτικά βλέπουμε των συνδυασμό των συναρτήσεων κόστους που χρησιμοποιήθηκαν ,για το πρώτο μοντέλο:

$$\mathcal{L}_{\text{total}} = 1.0 \cdot \mathcal{L}_{\text{CE}} + 1.0 \cdot \mathcal{L}_{\text{Dice}} \quad (3.1)$$

Για το δεύτερο προτεινόμενο μοντέλο :

$$\mathcal{L}_{\text{total}} = 1.0 \cdot \mathcal{L}_{\text{CE}} + 1.0 \cdot \mathcal{L}_{\text{Dice}} + 0.001 \cdot \mathcal{L}_{\text{LevelSet}} + 0.000005 \cdot \mathcal{L}_{\text{GAC}} \quad (3.2)$$

ΚΕΦΑΛΑΙΟ 4

ΑΝΑΛΥΣΗ ΠΕΙΡΑΜΑΤΩΝ

4.1 Υπολογιστικό Περιβάλλον

4.2 Ανάλυση Μεθόδου

4.3 Διαδικασία Εκπαίδευσης Αλγορίθμων

Σε αυτό το κεφάλαιο περιγράφεται το υπολογιστικό περιβάλλον που χρησιμοποιήθηκε για την εκπαίδευση και αξιολόγηση των μοντέλων τρισδιάστατης τμηματοποίησης (3D segmentation). Οι τεχνικές προδιαγραφές του συστήματος, η ανάλυση μεθοδολογίας εστιάζοντας στη συγκριτική αξιολόγηση δύο μοντέλων που διέφεραν αποκλειστικά στη συνάρτηση κόστους. Τέλος, η διαδικασία της εκπαίδευσης, οι τεχνικές επικύρωσης, οι παράμετροι βελτιστοποίησης και η αξιολόγηση μέσω του δείκτη Συντελεστή Ομοιότητας Dice (DSC).

4.1 Υπολογιστικό Περιβάλλον

Η εκπαίδευση και ο πειραματισμός των μοντέλων πραγματοποιήθηκαν στον server του Πανεπιστημίου Ιωαννίνων με όνομα stilvi, ο οποίος αποτελείται από ένα ισχυρό σύστημα GPU. Συγκεκριμένα, χρησιμοποιήθηκε η κάρτα γραφικών NVIDIA TITAN V (αρχιτεκτονικής Volta), η οποία διαθέτει 5120 CUDA πυρήνες και 640 Tensor πυρήνες για επιτάχυνση deep learning εργασιών, καθώς και 12 GB μνήμης. Αυτή

η GPU έχει σχεδιαστεί ειδικά για απαιτητικούς υπολογισμούς μηχανικής μάθησης, καθιστώντας την κατάλληλη για τρισδιάστατη ιατρική απεικόνιση.

Ο κώδικας που χρησιμοποιήθηκε βασίζεται στην υλοποίηση του αποθετηρίου [28]. Το αρχικό αποθετήριο προτείνει τις εξής απαιτήσεις για την εκτέλεση:

- Λειτουργικό Σύστημα: Ubuntu OS (δοκιμασμένο με Ubuntu 20.04 LTS)
- NVIDIA Driver (δοκιμασμένο με έκδοση 455.32.00)
- Docker (δοκιμασμένο με έκδοση 19.03.12)
- Docker Compose (δοκιμασμένο με έκδοση 1.27.4)
- NVIDIA Container Toolkit (δοκιμασμένο με έκδοση 2.4.0)

Ωστόσο, στην παρούσα εργασία, ο κώδικας εκτελέστηκε απευθείας στο τοπικό σύστημα, χωρίς τη χρήση Docker. Ως deep learning framework επιλέχθηκε το **PyTorch** το οποίο προσφέρει GPU επιταχυνόμενες δυνατότητες και είναι κατάλληλο για χρήση σε μοντέλα για segmentation τρισδιάστατων δεδομένων.

4.2 Ανάλυση Μεθόδου

Για την ακριβή αξιολόγηση και σύγκριση των αποτελεσμάτων (που θα παρουσιαστούν στο Κεφ. 5), εκπαιδεύτηκαν δύο μοντέλα με ταυτόσημη αρχιτεκτονική και παραμέτρους, διαφοροποιούμενα αποκλειστικά στη συνάρτηση κόστους (loss function). Συγκεκριμένα, σε κάθε μοντέλο διατηρήσαμε κοινή αρχιτεκτονική ως προς την επιλογή νευρωνικού δικτύου (3D-UNet), τη μέθοδο βελτιστοποίησης και τις υπερπαραμέτρους της εκπαίδευσης (ρυθμός εκμάθησης, βελτιστοποιητής κλπ.), προκειμένου να απομονωθεί ως μοναδική μεταβλητή η συνάρτηση κόστους (loss function). Έτσι, διασφαλίζεται ότι τυχόν αποκλίσεις στην απόδοση οφείλονται αποκλειστικά και μόνο στη συνάρτηση αυτή. Συνοπτικά, έγιναν τα παρακάτω βήματα για τη συστηματική σύγκριση των δύο αυτών μοντέλων:

Αρχικά και για τα δύο μοντέλα χρησιμοποιήθηκε κοινό πλαίσιο εκπαίδευσης με ακριβώς τα ίδια δεδομένα, χρησιμοποιώντας πανομοιότυπη αρχιτεκτονική δικτύου και σταθερές υπερπαραμέτρους. Πιο αναλυτικά, εξαιρέθηκαν οι ίδιες περιπτώσεις με αναγνωριστικά 15, 23, 37, 68, 125 και 133, σύμφωνα με την καθιερωμένη πρακτική στη βιβλιογραφία, λόγω προβλημάτων ποιότητας ή ασυνέπειας στις μάσκες

εδάφους-αλήθειας. Η εκπαίδευση πραγματοποιήθηκε με τη χρήση του αλγορίθμου Stochastic Gradient Descent (SGD), ενισχυμένου με Nesterov momentum με συντελεστή 0.99, ώστε να επιταχυνθεί η σύγκλιση και να αποφευχθούν τοπικά ελάχιστα. Ο αριθμός των εποχών ορίστηκε σε 1000, με εφαρμογή μηχανισμού early stopping για αποφυγή υπερπροσαρμογής, σταματώντας την εκπαίδευση όταν δεν παρατηρείται βελτίωση στην επικύρωση. Όπως φαίνεται παρακάτω το early stopping έγινε κοντά στα 300 epochs.

4.3 Διαδικασία Εκπαίδευσης Αλγορίθμων

Η βιβλιογραφία προτείνει τη μέθοδο διασταυρούμενης επικύρωσης τύπου k-fold (π.χ. $k=5$ ή 10) ως πρακτική αξιόπιστης εκτίμησης της απόδοσης των μοντέλων [29]. Παρ'όλα αυτά, στην παρούσα εργασία επιλέχθηκε η εκπαίδευση και αξιολόγηση σε ένα μόνο fold, κυρίως λόγω των εξαιρετικά υψηλών υπολογιστικών απαιτήσεων και των χρονικών περιορισμών. Έχει παρατηρηθεί ότι η αύξηση του αριθμού των folds συνεπάγεται πολλαπλή εκπαίδευση μοντέλων και, συνεπώς, αυξημένο υπολογιστικό κόστος, ειδικά στην προκείμενη φάση όπου χρειάστηκε η εκπαίδευση δύο μοντέλων.

4.3.1 Αξιολόγηση Απόδοσης

Η αξιολόγηση της απόδοσης των δύο μοντέλων πραγματοποιήθηκε με βάση τον δείκτη Dice Similarity Coefficient (DSC)[20], ο οποίος αποτελεί καθιερωμένο μέτρο για προβλήματα τμηματοποίησης σε ιατρικές εικόνες. Ο δείκτης Dice αποτιμά τον βαθμό επικάλυψης μεταξύ της προβλεπόμενης μάσκας και της αντίστοιχης ground truth, λαμβάνοντας τιμές από 0 (καμία επικάλυψη) έως 1 (τέλεια ταύτιση). Η επιλογή του συγκεκριμένου μετρικού δείκτη υπαγορεύεται από την ανάγκη για υψηλή ακρίβεια στα όρια των ανατομικών δομών, ιδίως σε περιπτώσεις όπου το μέγεθος του όγκου είναι μικρό σε σύγκριση με τον συνολικό όγκο της εικόνας.

ΚΕΦΑΛΑΙΟ 5

ΑΠΟΤΕΛΕΣΜΑΤΑ

5.1 Αποτελέσματα και Συζήτηση

5.2 Ποσοτική Σύγκριση των Μεθόδων

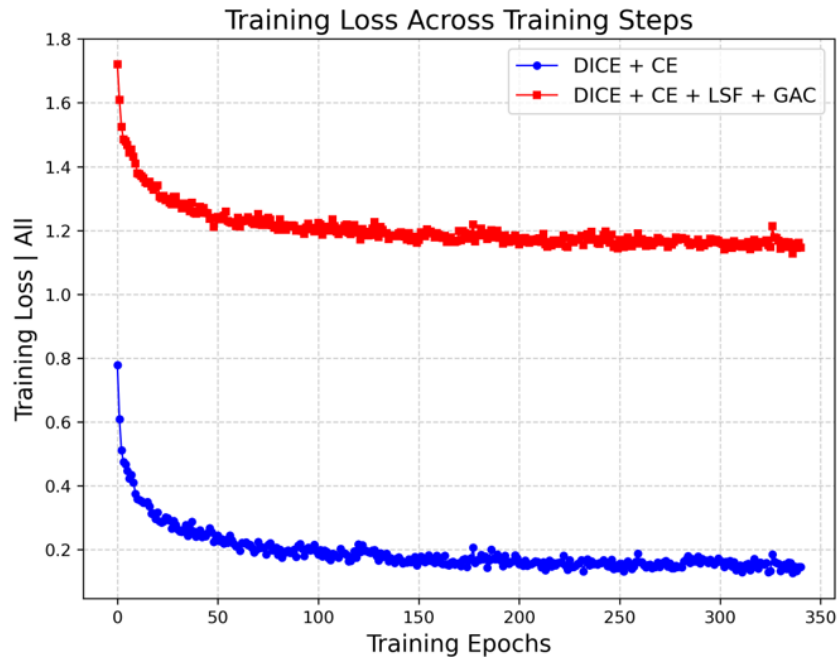
5.3 Συγκριτική Οπτικοποίηση Αποτελεσμάτων

Σε αυτό το κεφάλαιο αναλύονται τα αποτελέσματα των δύο αρχιτεκτονικών 3D U-Net. Αρχικά, γίνεται μια ποσοτική αξιολόγηση των μοντέλων μέσω του δείκτη Dice για κάθε κατηγορία τμηματοποίησης. Στην συνέχεια γίνεται σύγκριση της εξέλιξης των συναρτήσεων κόστους κατά την διάρκεια της εκπαίδευσης και επικύρωσης. Τέλος, παρατίθενται ενδεικτικά και αντιπροσωπευτικά παραδείγματα τμηματοποιήσεων για να αναδειχθεί η ποιοτική υπεροχή του προτεινόμενου μοντέλου στην αναγνώριση των πιο απαιτητικών ανατομικών δομών, όπως οι όγκοι.

5.1 Αποτελέσματα και Συζήτηση

Η αποδοτικότητα των δύο μοντέλων αξιολογήθηκε ως προς την πορεία εκπαίδευσης και επικύρωσης, με βάση τον δείκτη Dice και τις αντίστοιχες συναρτήσεις κόστους. Στην παρούσα ενότητα παρουσιάζονται αναλυτικά τα ευρήματα, ερμηνεύονται οι αποκλίσεις μεταξύ των μοντέλων και εξάγονται ποιοτικά συμπεράσματα αναφορικά με τη συμπεριφορά τους.

5.1.1 Συνάρτηση κόστους εκπαίδευσης



Σχήμα 5.1: Πορεία του συνολικού training loss κατά τη διάρκεια της εκπαίδευσης για τα δύο μοντέλα.

Στο Σχήμα 5.1 απεικονίζεται η εξέλιξη της συνάρτησης κόστους κατά τη διάρκεια της εκπαίδευσης και για τα δύο μοντέλα (βασικό DICE+CE και προτεινόμενο DICE+CE+LSF+GAC).

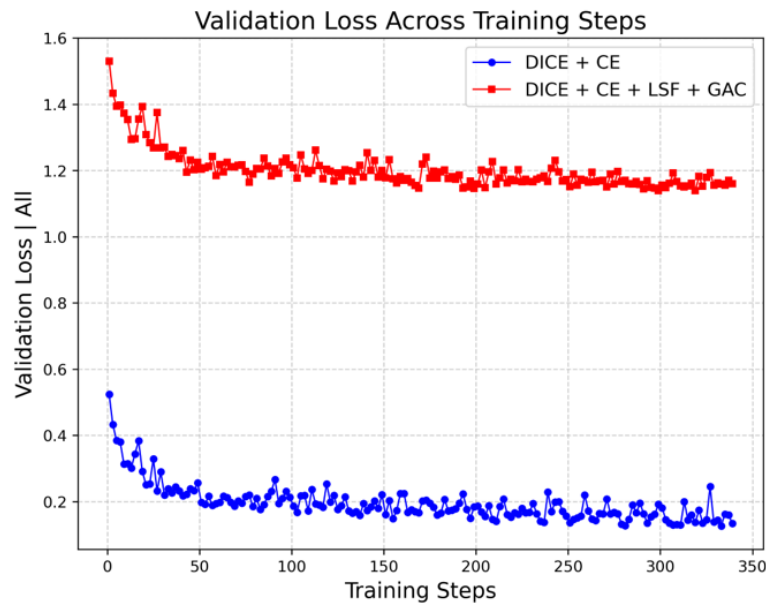
Όπως φαίνεται, το μοντέλο βάσης (μπλε καμπύλη) μειώνει το κόστος με ταχύτερο ρυθμό, επιτυγχάνοντας σε πολύ χαμηλές τιμές (περίπου 0.15 ήδη στις πρώτες 50 εποχές).

Αντίθετα, το προτεινόμενο μοντέλο (κόκκινη καμπύλη) ξεκινά με μεγαλύτερη τιμή κόστους και μειώνεται πιο σταδιακά (φτάνοντας περίπου στο 1.1 μετά από 300 εποχές). Η υψηλότερη αρχική τιμή και η πιο αργή πτώση της κόστους οφείλονται στους επιπλέον όρους LSF και GAC: αυτοί οι όροι προσθέτουν περιορισμούς στο αποτέλεσμα (π.χ. ομαλότητα και γεωδαισιακή ευθυγράμμιση των ακμών), ως συνέπεια η εκπαίδευση να είναι πιο αργή. Παρ' όλα αυτά, η σταδιακή αύξηση του κόστους στον προτεινόμενο αλγόριθμο υποδηλώνει ότι η βελτίωση των παραμέτρων συνεχίζεται σε σταθερό ρυθμό.

Συνοψίζοντας, το μοντέλο DICE+CE+LSF+GAC συγκλίνει πιο αργά, αλλά αυτό το αναμένουμε λόγω των σύνθετων συναρτήσεων απώλειας. Παρά το ότι το προτει-

νόμενο μοντέλο εμφανίζει υψηλότερη απώλεια, η σταθερή του συμπεριφορά μπορεί να συμβάλλει σε καλύτερη μορφολογική και τοπολογική σύλληψη των δομών στην τμηματοποίηση κάτι που θα φανεί παρακάτω από τα metrics στο validation/test set.

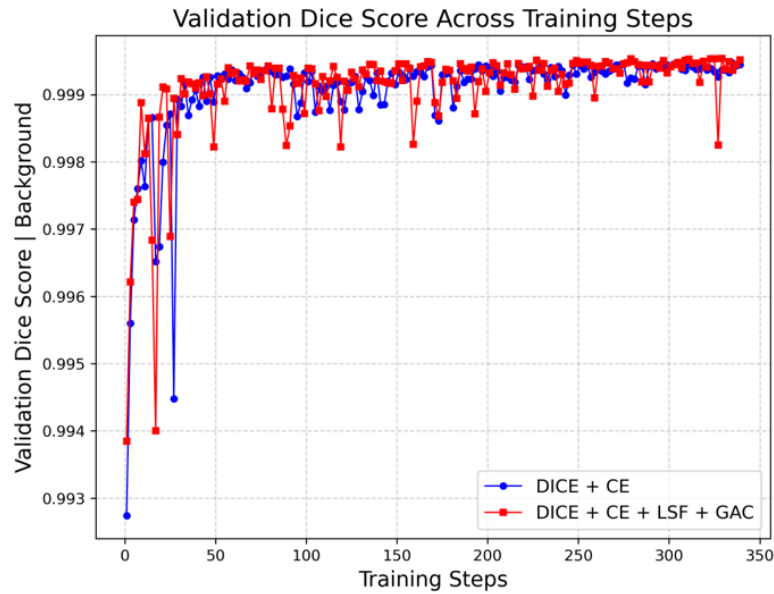
5.1.2 Συνάρτηση κόστους επικύρωσης



Σχήμα 5.2: Συνολική πορεία του Validation Loss για τα δύο μοντέλα.

Στο Σχήμα 5.2 εμφανίζεται η εξέλιξη του κόστους επικύρωσης για τα δύο μοντέλα. Όπως και προηγουμένως το “μπλέ” μοντέλο έχει χαμηλότερη τιμή κόστους επικύρωσης σε κάθε epoch, ενώ το προτεινόμενο (κόκκινο) μοντέλο έχει υψηλότερη. Αν και το validation loss είναι υψηλότερο, η κόκκινη καμπύλη παρουσιάζει πολύ ομαλή και σταθερή σύγκλιση χωρίς μεγάλες διακυμάνσεις. Αυτό υποδηλώνει ότι ο GAC όρος ενθαρρύνει την ευθυγράμμιση του περιγράμματος με τα πραγματικά όρια. Ενώ το απλό μοντέλο ελαχιστοποιεί πιο αποτελεσματικά την τιμή της συνάρτησης κόστους, το σύνθετο μοντέλο παρέχει πιο πειθαρχημένη προσαρμογή στα σημαντικά μορφολογικά χαρακτηριστικά (π.χ. σχήμα νεφρού/όγκου), οδηγώντας σε συγκρίσιμη ή ενδεχομένως βελτιωμένη απόδοση Dice, την οποία θα δούμε στη συνέχεια.

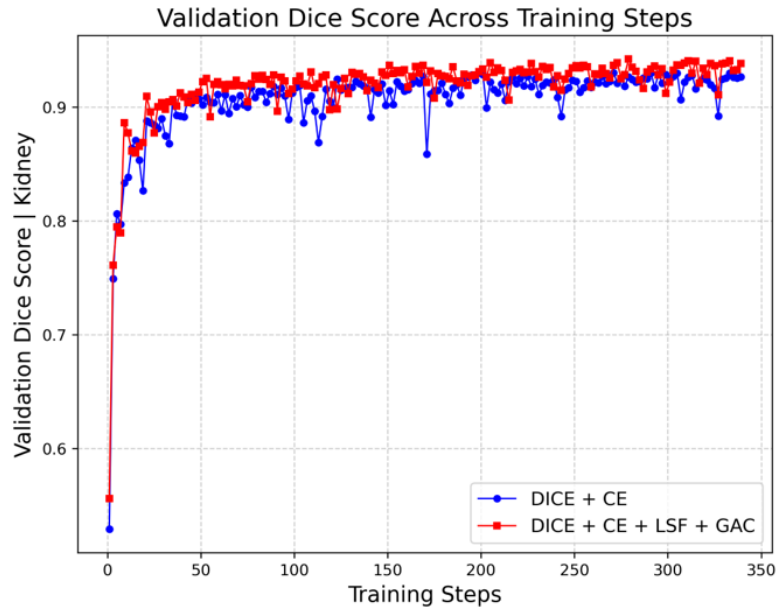
5.1.3 Δείκτης Dice για υπόβαθρο(background)



Σχήμα 5.3: Dice score στην περιοχή υποβάθρου (background) στο validation set.

Στο Σχήμα 5.3 παρουσιάζεται η πορεία του Dice Score για την περιοχή υποβάθρου (background class). Και τα δύο μοντέλα επιτυγχάνουν εξαιρετικά υψηλές επιδόσεις (άνω του 0.998) ήδη από τα πρώτα 30 epochs. Βέβαια, επηρεάζει το γεγονός ότι η μάθηση του υποβάθρου είναι σχετικά πιο εύκολη διαδικασία. Εντούτοις, η προσθήκη των LSF και GAC φαίνεται να συνεισφέρει στη σταθερότητα του μοντέλου: η καμπύλη του προτεινόμενου μοντέλου (κόκκινη) εμφανίζει μικρότερη διακύμανση και λιγότερα σημεία απότομης πτώσης, υποδηλώνοντας μεγαλύτερη συνέπεια στις προβλέψεις, ακόμη και κατά την εναλλαγή των batches. Το αποτέλεσμα αυτό ενισχύει την άποψη ότι οι γεωμετρικές συνιστώσες loss συμβάλλουν στη σταθεροποίηση της απόδοσης, ιδίως σε περιοχές χαμηλής πολυπλοκότητας.

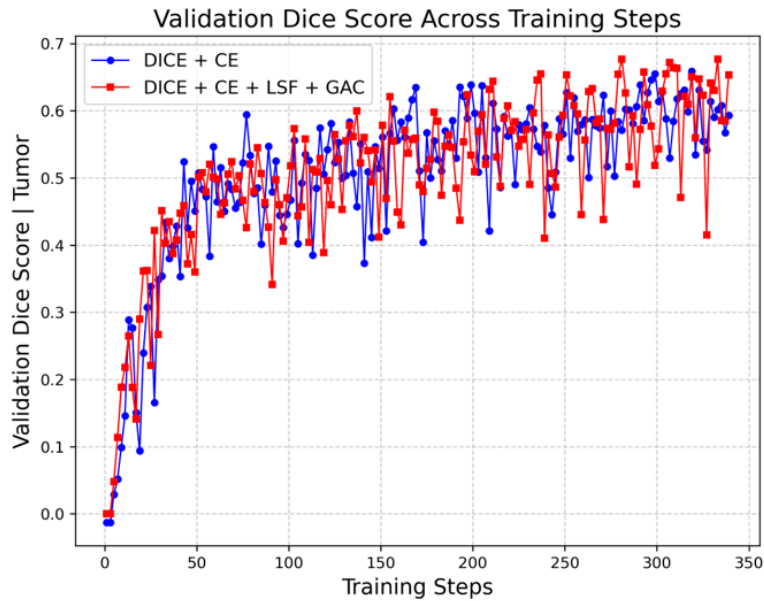
5.1.4 Δείκτης Dice για το νεφρό



Σχήμα 5.4: Εξέλιξη του Dice Score για την τμηματοποίηση του νεφρού κατά την επικύρωση.

Στο Σχήμα 5.4 παρουσιάζεται η εξέλιξη του Dice Score για την κατηγορία του νεφρού. Η ανάλυση της απόδοσης στο επίπεδο επιμέρους κατηγοριών αποκαλύπτει περαιτέρω διαφορές στη συμπεριφορά των μοντέλων. Όσον αφορά την κατηγορία του νεφρού, τα Dice Scores κινούνται σε υψηλά επίπεδα ήδη από τα αρχικά βήματα της εκπαίδευσης. Και τα δύο μοντέλα προσεγγίζουν τιμές άνω του 0.92, ωστόσο το μοντέλο με ενσωματωμένες τις LSF και GAC παρουσιάζει μια πιο σταθερή τροχιά και ελαφρώς υψηλότερες κορυφές, ιδιαίτερα από το μέσο της εκπαίδευσης και μετά. Αυτό υποδηλώνει ότι η γεωμετρική πληροφορία βοηθά στην ομαλότερη και πιο ακριβή αποτύπωση της ανατομικής περιοχής του νεφρού, η οποία χαρακτηρίζεται από σαφή όρια.

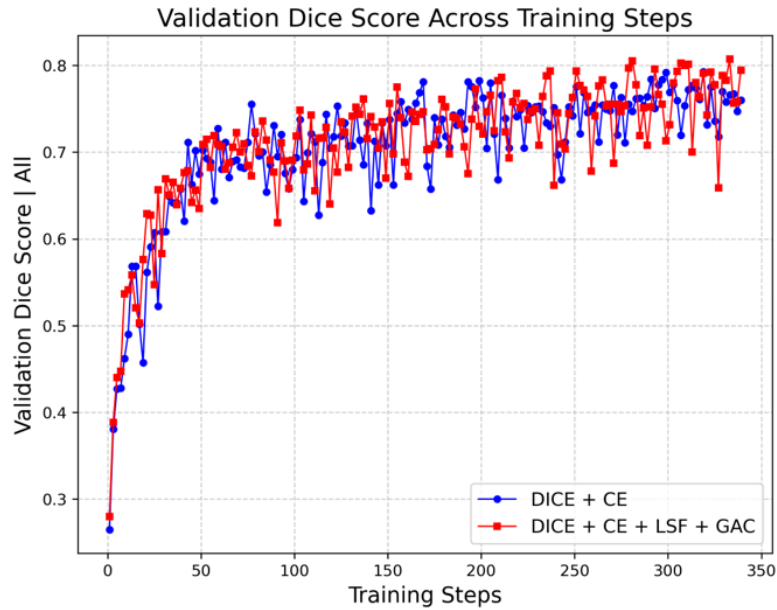
5.1.5 Δείκτης Dice για τον όγκο



Σχήμα 5.5: Dice Score στην περιοχή του όγκου.

Στο Σχήμα 5.5 φαίνεται η εξέλιξη του Dice Score του όγκου. Αυτό είναι από τα πιο απαιτητικά ως τώρα, αφού οι όγκοι μπορεί να είναι πολύ μικροί ή ακανόνιστοι. Τα Dice Scores ξεκινούν από πολύ χαμηλές τιμές, κάτι που είναι αναμενόμενο λόγω της μικρής έκτασης και της ακανόνιστης μορφής των όγκων, σε σχέση με το υπόλοιπο περιεχόμενο του όγκου εισόδου. Σε αυτή την περίπτωση, το μοντέλο που ενσωματώνει τις Level Set και Geodesic Active Contour Loss όχι μόνο συγκλίνει ταχύτερα αλλά διατηρεί και αισθητά υψηλότερη μέση τιμή Dice σε όλη τη διάρκεια της εκπαίδευσης. Επιπλέον, η γραμμή της απόδοσής του χαρακτηρίζεται από μικρότερες διακυμάνσεις, στοιχείο που υποδηλώνει καλύτερη γενίκευση στις περιοχές υψηλής πολυπλοκότητας. Αντίθετα, το βασικό μοντέλο παρουσιάζει συχνές απότομες πτώσεις απόδοσης, γεγονός που θα μπορούσε να συνδέεται με ευαισθησία σε μικρές μεταβολές των χαρακτηριστικών των patches. Τελικά, μοντέλο με τις γεωμετρικές απώλειες επιτυγχάνει σταθερά υψηλότερη απόδοση, με καλύτερη σύγκλιση και λιγότερη διακύμανση.

5.1.6 Δείκτης Dice για Kidney + Tumor



Σχήμα 5.6: Dice score στο σύνολο των κατηγοριών νεφρό και όγκος στο validation set.

Στο σχήμα 5.6 αποτυπώνεται η συνολική εικόνα της απόδοσης στο validation set, λαμβάνοντας υπόψη τις κλάσεις νεφρού και όγκου. Εδώ παρατηρείται ότι αν και το βασικό μοντέλο εμφανίζει αρχικά καλύτερη επίδοση, από τη μέση της εκπαίδευσης και μετά το μοντέλο με τις γεωμετρικές συνιστώσες τείνει να υπερέχει. Η εξέλιξη αυτή είναι ενδεικτική του γεγονότος ότι οι LSF και GAC loss ενισχύουν τη γενίκευση, οδηγώντας σε σταδιακά βελτιωμένες προβλέψεις, ιδίως σε περιοχές υψηλής πολυπλοκότητας όπως είναι οι όγκοι με ασαφή όρια. Συνολικά, τα αποτελέσματα καταδεικνύουν ότι το προτεινόμενο μοντέλο με των συνδυασμό των συναρτήσεων απώλειας μπορεί να προσφέρει βελτιώσεις στη σταθερότητα και στην ακρίβεια, παρά την αυξημένη υπολογιστική πολυπλοκότητα που συνεπάγεται. Τα αποτελέσματα αυτά θα ενισχυθούν περαιτέρω τόσο ποσοτικά όσο και οπτικά.

5.2 Ποσοτική Σύγκριση των Μεθόδων

Στον Πίνακα παρουσιάζονται τα ποσοτικά αποτελέσματα που προέκυψαν από τη χρήση των δύο διαφορετικών συναρτήσεων κόστους για την εκπαίδευση του 3D

U-Net.

Πίνακας 5.1: Δείκτες Dice ανά κλάση για τα δύο μοντέλα

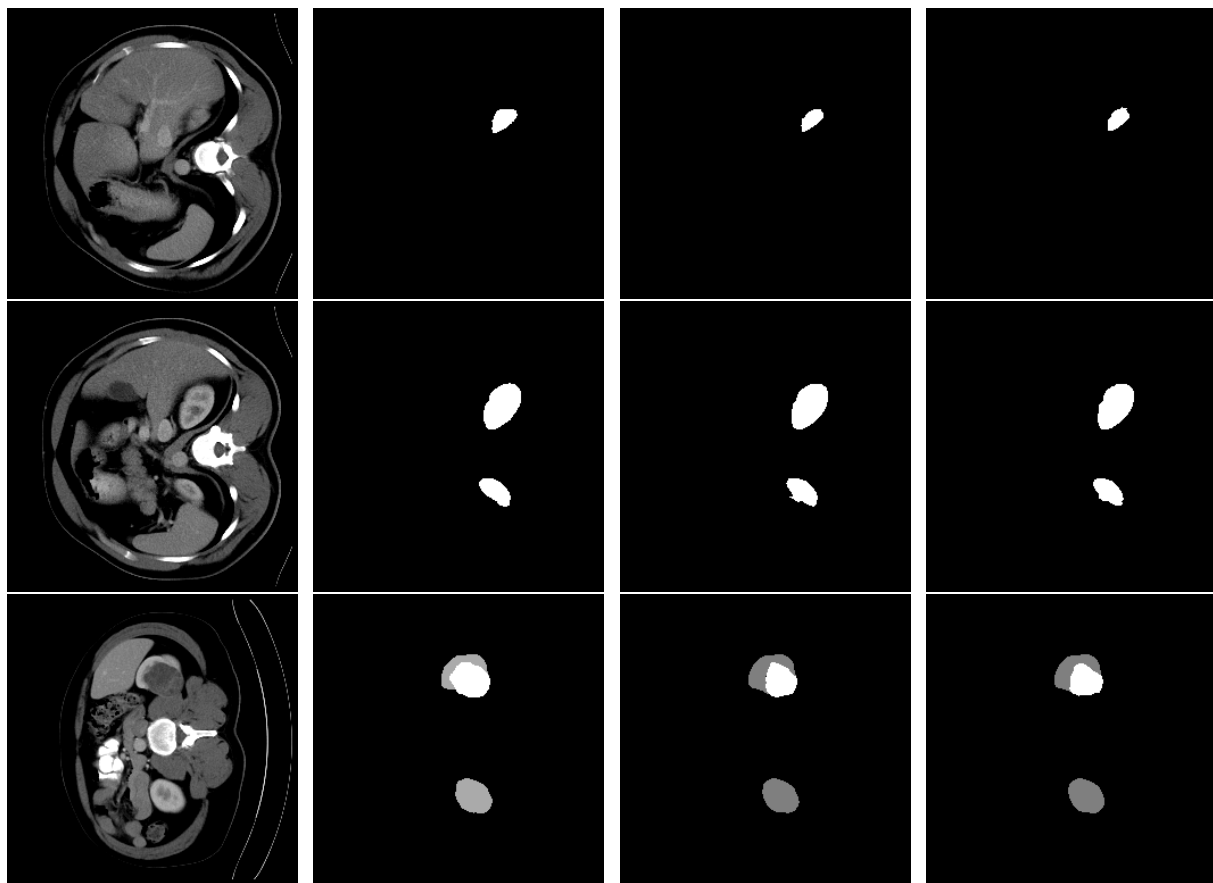
Κλάση	DICE + CE	DICE + CE + LSF + GAC
Υπόβαθρο (Background)	0,9994	0,9995
Νεφρό (Kidney)	0,9380	0,9480
Όγκος (Tumor)	0,6560	0,6840
Νεφρό + Όγκος (Kidney + Tumor)	0,7970	0,8160

- Στην κατηγορία **Νεφρού (Kidney)** παρατηρείται αύξηση του δείκτη Dice από 0,9380 σε 0,9480, κάτι που επιβεβαιώνει ακριβέστερη τμηματοποίηση των ορίων του νεφρού με την προτεινόμενη μέθοδο.
- Η μεγαλύτερη σχετική βελτίωση παρατηρείται στην κατηγορία **Όγκου (Tumor)**, όπου το Dice αυξάνεται από 0,6560 σε 0,6840, αναδεικνύοντας την αποτελεσματικότητα του προτεινόμενου μοντέλου στην καλύτερη ανίχνευση πολύπλοκων και ανομοιογενών δομών.
- Στην κατηγορία και των δύο (Νεφρό+Όγκος) από 0,7970 σε 0,8160, υποδηλώνοντας για ακόμη μια φορά καλύτερη αποτελεσματικότητα του προτεινόμενου μοντέλου.

5.3 Συγκριτική Οπτικοποίηση Αποτελεσμάτων

Προκειμένου να γίνει ποιοτική αξιολόγηση των αποτελεσμάτων, πραγματοποιήθηκε οπτικοποίηση δειγμάτων από το validation set. Η οπτικοποίηση έγινε σε μορφή δισδιάστατων τομών (slices), στις οποίες παρουσιάζονται συγκριτικά η αρχική εικόνα, η μάσκα (ground truth), καθώς και οι προβλέψεις των δύο μοντέλων για την ίδια τομή. Με τον τρόπο αυτό κατέστη δυνατή η απευθείας σύγκριση της ακρίβειας της τμηματοποίησης, τόσο ως προς τη γεωμετρία όσο και ως προς τα όρια των δομών.

Ενδεικτικά, παρατίθενται κάποια αποτελέσματα όπου σε κάποιες περιπτώσεις θα δούμε την ύπαρξη μόνο του kidney ενώ σε άλλες του kidney και tumor καθώς κάθε 3D περίπτωση αποτελείται από πολλά δισδιάστατα slices οπότε είναι φυσιολογικό να μην έχει δημιουργηθεί ακόμη ο όγκος σε κάποια από αυτά.



(α) CT Image

(β) Ground Truth

(γ) CE + Dice loss

(δ) LSF + GAC Loss

Σχήμα 5.7: Σύγκριση των διαφορετικών loss function του 3D U-Net με βάση την αρχική εικόνα : (α) CT Image, (β) Ground Truth, (γ) CE + DICE Loss, (δ) LSF + GAC Loss

Στο παραπάνω σχήμα παρατηρούμε η πρώτη στήλη αποτελεί την αρχική εικόνα, η δεύτερη την πραγματική περιοχή τμηματοποίησης, η τρίτη το αποτέλεσμα τμηματοποίησης πρώτου μοντέλου και η τέταρτη το αποτέλεσμα τμηματοποίησης του προτεινόμενου μοντέλου. Ακόμη οι περιοχές με βάση τους χρωματισμούς είναι οι εξής:

- **Λευκό:** περιοχή νεφρού
- **Γκρι:** περιοχή όγκου του νεφρού
- **Μαύρο:** φόντο (background)

Συνολικά, από την οπτική σύγκριση προκύπτει πως η ενσωμάτωση των επιπλέον συναρτήσεων απώλειας LSF και GAC (προτεινόμενο μοντέλο) βελτιώνει σημαντικά

την ακρίβεια τμηματοποίησης ειδικά στην απαιτητική περιοχή του όγκου. Ακόμη, παρέχει πιο συνεκτικό, ακριβές segmentation και σε μικρές ή ακανόνιστες περιοχές βελτιώνοντας το αποτέλεσμα της τμηματοποίησης, κάτι που επιβεβαιώνεται και από τα αντίστοιχα ποσοτικά αποτελέσματα. Θεωρείται λοιπόν πιο ικανό για τέτοιου είδους προβλήματα.

ΚΕΦΑΛΑΙΟ 6

ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΜΕΛΛΟΝΤΙΚΗ ΕΡΓΑΣΙΑ

Η παρούσα εργασία επικεντρώθηκε στη μελέτη και συγκριτική αξιολόγηση του μοντέλου 3D U-Net για την τρισδιάστατη τμηματοποίηση ιατρικών εικόνων και συγκεκριμένα του συνόλου δεδομένων KiTS19 όπου απαιτούσε τμηματοποίηση νεφρού και όγκου. Για την υλοποίηση αυτή διερευνήθηκαν και συγκρίθηκαν δύο εκδοχές του μοντέλου. Η βασική εκδοχή όπου το μοντέλο εκπαιδεύτηκε σε διαδεδομένες συναρτήσεις κόστους Cross-Entropy και Dice. Η δεύτερη και προτεινόμενη εκδοχή έχει μια πιο εμπλουτισμένη συνάρτηση απώλειας όπου, εκτός από τις δύο προηγούμενες, ενσωματώθηκαν και οι προτεινόμενες Level Set και Geodesic Active Contour.

Τα αποτελέσματα αξιολόγησης ανέδειξαν σαφή πλεονεκτήματα υπέρ του εμπλουτισμένου μοντέλου σε όρους ακρίβειας και σταθερότητας. Ειδικότερα, παρατηρήθηκε ανώτερη επίδοση του δείκτη Dice Score για την κατηγορία «νεφρό» καθώς και για την πιο απαιτητική κατηγορία «όγκος», σε σύγκριση με το βασικό μοντέλο. Αν και η συνολική τιμή της συνάρτησης κόστους ήταν αριθμητικά ανώτερη στο προτεινόμενο μοντέλο (λόγω των επιπρόσθετων όρων), οι επιδόσεις στο δείκτη Dice υποδήλωσαν βελτιωμένη γενίκευση του μοντέλου και πιο ακριβή αναπαράσταση των περιγραμμάτων. Παράλληλα, η στατιστική ανάλυση των αποτελεσμάτων ανά κατηγορία και κατά την πορεία της εκπαίδευσης ανέδειξε τη σημασία της επιλογής της συνάρτησης κόστους, όχι μόνο για τη συνολική απόδοση αλλά και για τη συμπεριφορά του μοντέλου σε κρίσιμες περιοχές ενδιαφέροντος.

Τα ευρήματα της παρούσας διπλωματικής ανοίγουν αρκετές κατευθύνσεις για περαιτέρω βελτίωση και διερεύνηση. Αρχικά, προτείνεται η βελτιστοποίηση των συναρτήσεων απώλειας μελλοντικά, για παράδειγμα, μέσω μιας «κανονικοποιημένης»

εκδοχής της GAC Loss που θα ενσωματώνει ρητά μια συνάρτηση ένδειξης ακμών (edge indicator) του αρχικού εικόνas. Μια τέτοια προσθήκη θα επιτρέψει στο μοντέλο να καθοδηγείται ακόμη πιο αποτελεσματικά προς τα πραγματικά μορφολογικά όρια των οργάνων, αξιοποιώντας άμεσα την πληροφορία των μοτίβων έντασης της εικόνas. Επίσης, θα ήταν εξίσου χρήσιμη η προσθήκη περισσότερων δεδομένων ώστε να επικρατήσει καλύτερη γενίκευση και αξιοπιστία στο μοντέλο. Περισσότερα και πιο ποικίλα δείγματα θα μειώσουν τον κίνδυνο υπερ-προσαρμογής σε συγκεκριμένες λεπτομέρειες του αρχικού dataset και θα καταστήσουν το μοντέλο πιο ανθεκτικό σε νέες και άγνωστες περιπτώσεις.

Συμπερασματικά, η ενσωμάτωση προχωρημένων μορφολογικών συναρτήσεων κόστους (LSF και GAC) στο πλαίσιο ενός 3D U-Net απέδειξαν σημαντική βελτίωση στην τρισδιάστατη τμηματοποίηση ιατρικής εικόνas. Οι προτεινόμενες κατευθύνσεις για μελλοντική εργασία αποσκοπούν τόσο στην περαιτέρω αύξηση της ακρίβειας όσο και στη βελτίωση της γενικευσιμότητας της προσέγγισης, μέσω ενίσχυσης των loss functions και εμπλουτισμού των δεδομένων,

ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, and C. I. Sanchez, “A survey on deep learning in medical image analysis,” *Medical image analysis*, vol. 42, pp. 60–88, 2017.
- [2] S. Qamar, H. Jin, R. Zheng, P. Ahmad, and M. Usama, “A variant form of 3d-unet for infant brain segmentation,” *Future Generation Computer Systems*, vol. 108, pp. 613–623, 2020.
- [3] N. Heller, F. Isensee, K. H. Maier-Hein, X. Hou, C. Xie, F. Li, Y. Nan, G. Mu, Z. Lin, M. Han *et al.*, “The state of the art in kidney and kidney tumor segmentation in contrast-enhanced ct imaging: Results of the kits19 challenge,” *Medical image analysis*, vol. 67, p. 101821, 2021.
- [4] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, “A survey of convolutional neural networks: analysis, applications, and prospects,” *IEEE transactions on neural networks and learning systems*, vol. 33, no. 12, pp. 6999–7019, 2021.
- [5] M. Yeung, E. Sala, C.-B. Schönlieb, and L. Rundo, “Unified focal loss: Generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation,” *Computerized Medical Imaging and Graphics*, vol. 95, p. 102026, 2022.
- [6] A. L. Samuel, “Some studies in machine learning using the game of checkers,” *IBM Journal of Research and Development*, vol. 44, no. 1, pp. 206–227, 1959.
- [7] R. Jain and S. Sharma, “Machine learning: The new language for applications,” *ResearchGate*, 2020.
- [8] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

- [9] T. Hastie, R. Tibshirani, J. Friedman, T. Hastie, R. Tibshirani, and J. Friedman, "Overview of supervised learning," *The elements of statistical learning: Data mining, inference, and prediction*, pp. 9–41, 2009.
- [10] Z. A. Ali, Z. H. Abduljabbar, H. A. Tahir, A. B. Sallow, and S. M. Almufti, "extreme gradient boosting algorithm with machine learning: A review," *Academic Journal of Nawroz University*, vol. 12, no. 2, pp. 320–334, 2023.
- [11] R. Qamar and B. A. Zardari, "Artificial neural networks: An overview," *Mesopotamian Journal of Computer Science*, vol. 2023, pp. 124–133, 2023.
- [12] K. Sharifani and M. Amini, "Machine learning and deep learning: A review of methods and applications," *World Information Technology and Engineering Journal*, vol. 10, no. 07, pp. 3897–3904, 2023.
- [13] X. Zhao, L. Wang, Y. Zhang, X. Han, M. Deveci, and M. Parmar, "A review of convolutional neural networks in computer vision," *Artificial Intelligence Review*, vol. 57, no. 4, p. 99, 2024.
- [14] S. R. Dubey, S. K. Singh, and B. B. Chaudhuri, "Activation functions in deep learning: A comprehensive survey and benchmark," *arXiv preprint arXiv:2109.14545*, 2021, available: <https://arxiv.org/abs/2109.14545>.
- [15] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, "Striving for simplicity: The all convolutional net," *arXiv preprint arXiv:1412.6806*, 2014.
- [16] I. Sutskever, J. Martens, G. Dahl, and G. E. Hinton, "On the importance of initialization and momentum in deep learning," in *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*. PMLR, 2013, pp. 1139–1147. [Online]. Available: <https://proceedings.mlr.press/v28/sutskever13.pdf>
- [17] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
- [18] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 3–4, pp. 379–423, 623–656, July–October 1948. [Online]. Available: <https://people.math.harvard.edu/~ctm/home/text/others/shannon/entropy/entropy.pdf>
- [19] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.

- [20] F. Milletari, N. Navab, and S.-A. Ahmadi, “V-net: Fully convolutional neural networks for volumetric medical image segmentation,” *arXiv preprint arXiv:1606.04797*, 2016.
- [21] S. Osher and J. A. Sethian, “Fronts propagating with curvature-dependent speed: Algorithms based on hamilton–jacobi formulations,” *Journal of Computational Physics*, vol. 79, no. 1, pp. 12–49, 1988.
- [22] S. Osher and R. P. Fedkiw, “Level set methods: An overview and some recent results,” *Journal of Computational Physics*, vol. 169, no. 2, pp. 463–502, November 2000. [Online]. Available: <https://physbam.stanford.edu/~fedkiw/papers/cam2000-08.pdf>
- [23] J. Ma, J. He, and X. Yang, “Learning geodesic active contours for embedding object global information in segmentation cnns,” *IEEE Transactions on Medical Imaging*, 2020. [Online]. Available: <https://doi.org/10.1109/TMI.2020.3022693>
- [24] X. Chen, X. Luo, Y. Zhao, S. Zhang, G. Wang, and Y. Zheng, “Learning euler’s elastica model for medical image segmentation,” *arXiv preprint arXiv:2011.00526*, 2020, preprint. [Online]. Available: <https://arxiv.org/abs/2011.00526>
- [25] N. Heller, N. Sathianathan, A. Kalapara, E. Walczak, K. Moore, H. Kaluzniak, J. Rosenberg, P. Blake, Z. Rengel, M. Oestreich, J. Dean, M. Tradewell, A. Shah, R. Tejpal, Z. Edgerton, M. Peterson, S. Raza, S. Regmi, N. Papanikolopoulos, and C. Weight, “The kits19 challenge data: 300 kidney tumor cases with clinical context, ct semantic segmentations, and surgical outcomes,” *arXiv preprint arXiv:1904.00445*, 2019. [Online]. Available: <https://arxiv.org/abs/1904.00445>
- [26] F. Isensee, P. F. J. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, “No new-net,” *arXiv preprint arXiv:1904.04248*, 2019. [Online]. Available: https://kits.lib.umn.edu/wp-content/uploads/2019/11/Isensee_1e.pdf
- [27] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, “3d u-net: learning dense volumetric segmentation from sparse annotation,” in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17–21, 2016, Proceedings, Part II 19*. Springer, 2016, pp. 424–432.

- [28] motokimura, “kits19_3d_segmentation,” GitHub repository, 2020, accessed: 2025-06-25. [Online]. Available: https://github.com/motokimura/kits19_3d_segmentation
- [29] S. Raschka, “Model evaluation, model selection, and algorithm selection in machine learning,” *arXiv preprint arXiv:1811.12808*, 2018.