

Εθνικο Μετσοβιο Πολυτεχνειο
Σχολη Ηλεκτρολογων Μηχανικων και Μηχανικων Υπολογιστων
Τομεας Τεχνολογιας Πληροφορικης και Υπολογιστων

Ενίσχυση Μοντέλων Κατάστασης Χώρου με Μετα-Επεξεργασία για Γενίκευση Πεδίου στον Ιατρικό Τομέα

Διπλωματικη Εργασια

του

ΔΗΜΗΤΡΙΟΥ Γ. ΛΙΑΚΗ

Επιβλέπων: Αθανάσιος Βουλόδημος
Αναπληρωτής Καθηγητής ΕΜΠ

Αθήνα, Ιούλιος 2026

Εθνικο Μετσοβιο Πολυτεχνειο
Σχολη Ηλεκτρολογων Μηχανικων και Μηχανικων Υπολογιστων
Τομεας Τεχνολογιας Πληροφορικης και Υπολογιστων

Ενίσχυση Μοντέλων Κατάστασης Χώρου με Μετα-Επεξεργασία για Γενίκευση Πεδίου στον Ιατρικό Τομέα

Διπλωματικη Εργασια

του

ΔΗΜΗΤΡΙΟΥ Γ. ΛΙΑΚΗ

Επιβλέπων: Αθανάσιος Βουλόδημος
Αναπληρωτής Καθηγητής ΕΜΠ

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 8 Ιουλίου 2026.

(Υπογραφή)	*(Υπογραφή)*	*(Υπογραφή)*
.....	.....	.....
Αθανάσιος Βουλόδημος	Γεώργιος Στάμου	Χρυσούλα Ζέρβα
Αναπληρωτής Καθηγητής ΕΜΠ	Καθηγητής ΕΜΠ	Επίκουρη Καθηγήτρια ΕΜΠ

Αθήνα, Ιούλιος 2026



Εθνικο Μετσοβιο Πολυτεχνειο
Σχολη Ηλεκτρολογων Μηχανικων και Μηχανικων Υπολογιστων
Τομεας Τεχνολογιας Πληροφορικης και Υπολογιστων

Copyright © – All rights reserved. Με την επιφύλαξη παντός δικαιώματος.
Δημήτριος Λιάκης, 2026.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα.

Το περιεχόμενο αυτής της εργασίας δεν απηχεί απαραίτητα τις απόψεις του Τμήματος, του Επιβλέποντα, ή της επιτροπής που την ενέκρινε.

ΔΗΛΩΣΗ ΜΗ ΛΟΓΟΚΛΟΠΗΣ ΚΑΙ ΑΝΑΛΗΨΗΣ ΠΡΟΣΩΠΙΚΗΣ ΕΥΘΥΝΗΣ

Με πλήρη επίγνωση των συνεπειών του νόμου περί πνευματικών δικαιωμάτων, δηλώνω ενυπογράφως ότι είμαι αποκλειστικός συγγραφέας της παρούσας Διπλωματικής Εργασίας, για την ολοκλήρωση της οποίας κάθε βοήθεια είναι πλήρως αναγνωρισμένη και αναφέρεται λεπτομερώς στην εργασία αυτή. Έχω αναφέρει πλήρως και με σαφείς αναφορές, όλες τις πηγές χρήσης δεδομένων, απόψεων, θέσεων και προτάσεων, ιδεών και λεκτικών αναφορών, είτε κατά κυριολεξία είτε βάσει επιστημονικής παράφρασης. Αναλαμβάνω την προσωπική και ατομική ευθύνη ότι σε περίπτωση αποτυχίας στην υλοποίηση των ανωτέρω δηλωθέντων στοιχείων, είμαι υπόλογος έναντι λογοκλοπής, γεγονός που σημαίνει αποτυχία στην Διπλωματική μου Εργασία και κατά συνέπεια αποτυχία απόκτησης του Τίτλου Σπουδών, πέραν των λοιπών συνεπειών του νόμου περί πνευματικών δικαιωμάτων. Δηλώνω, συνεπώς, ότι αυτή η Διπλωματική Εργασία προετοιμάστηκε και ολοκληρώθηκε από εμένα προσωπικά και αποκλειστικά και ότι, αναλαμβάνω πλήρως όλες τις συνέπειες του νόμου στην περίπτωση κατά την οποία αποδειχθεί, διαχρονικά, ότι η εργασία αυτή ή τμήμα της δεν μου ανήκει διότι είναι προϊόν λογοκλοπής άλλης πνευματικής ιδιοκτησίας.

(Υπογραφή)

.....
Δημήτριος Λιάκης

Ιούλιος 2026

στην οικογένειά μου

Περίληψη

Η κατάτμηση ιατρικής εικόνας αποτελεί μια κομβική διαδικασία στον τομέα της ιατρικής απεικόνισης και της υπολογιστικής υποβοήθησης διαγνώσεων. Ωστόσο, η ανάπτυξη εύρωστων υπολογιστικών συστημάτων παρεμποδίζεται από δύο θεμελιώδεις προκλήσεις: αφενός την σημαντική δυσκολία ακριβούς οριοθέτησης οργάνων με ασαφή ανατομικά όρια, όπως το πάγκρεας, και αφετέρου τη χρόνια έλλειψη τρισδιάστατων (3D) ιατρικών δεδομένων με πυκνό σχολιασμό από ειδικούς.

Στόχος της παρούσας διπλωματικής εργασίας είναι η ανάπτυξη και η συγκριτική αξιολόγηση ενός συστήματος βαθιάς μάθησης για ημι-επιβλεπόμενη κατάτμηση παγκρέατος υπό αυστηρά περιορισμένους υπολογιστικούς πόρους. Ως βασική αρχιτεκτονική (backbone) μελετάται το Swin-UMamba, το οποίο ενσωματώνει Μοντέλα Χώρου Καταστάσεων (State Space Models — SSMs) με γραμμική υπολογιστική πολυπλοκότητα, και αντιπαραβάλλεται με το καθιερωμένο συνελικτικό δίκτυο U-Net3D. Αμφότερα τα δίκτυα εκπαιδεύονται υπό το ημι-επιβλεπόμενο πλαίσιο DyCON, ενώ στη φάση της αξιολόγησης εφαρμόζεται το κλασικό βήμα μετα-επεξεργασίας Largest Connected Component (LCC) για την απαλοιφή ψευδώς-θετικών συνεκτικών συνιστωσών.

Το DyCON αξιολογεί δυναμικά την αβεβαιότητα των προβλέψεων μέσω της εντροπίας και εφαρμόζει αντιθετική μάθηση σε επίπεδο ογκοστοιχείων, αξιοποιώντας τις μη-σχολιασμένες τομογραφίες. Η κύρια συνεισφορά της εργασίας είναι η συστηματική συγκριτική μελέτη αρχιτεκτονικών SSM και CNN για ημι-επιβλεπόμενη κατάτμηση υπό περιορισμένη χωρική ανάλυση, καθώς και η ανάδειξη της επίδρασης της μετα-επεξεργασίας LCC στην ακρίβεια των ανατομικών ορίων.

Η πειραματική αξιολόγηση πραγματοποιήθηκε στο σύνολο δεδομένων Pancreas-CT. Υπό τους αυστηρούς περιορισμούς υλικού (τμήματα $64 \times 64 \times 64$), το LCC βελτιώνει δραστικά τις μετρικές ορίων (HD95, ASD) αμφότερων των αρχιτεκτονικών. Ο συνδυασμός U-Net3D + DyCON + LCC αποδίδει την υψηλότερη και σταθερότερη ακρίβεια, επιτυγχάνοντας με μόλις 12 σχολιασμένες τομογραφίες επιδόσεις ισοδύναμες με την πλήρως επιβλεπόμενη εκπαίδευση στις 62. Στο συγκεκριμένο καθεστώς περιορισμένης ανάλυσης, το συνελικτικό δίκτυο υπερτερεί του Swin-UMamba.

Λέξεις Κλειδιά

Κατάτμηση Ιατρικής Εικόνας, Βαθιά Μάθηση, Μοντέλα Χώρου Καταστάσεων (SSMs), Ημι-επιβλεπόμενη Μάθηση, Mamba, DyCON, Πάγκρεας

Abstract

Medical image segmentation is a pivotal process in the field of medical imaging and computer-aided diagnosis. However, the development of robust computational systems is hindered by two fundamental challenges: the considerable difficulty of accurately delineating organs with ambiguous anatomical boundaries, such as the pancreas, and the chronic scarcity of densely annotated 3D medical datasets.

The objective of this diploma thesis is the development and comparative evaluation of a deep learning system for semi-supervised pancreas segmentation under strict computational constraints. The primary backbone studied is Swin-UMamba, which integrates State Space Models (SSMs) with linear computational complexity, and is compared against the established convolutional network U-Net3D. Both networks are trained under the semi-supervised DyCON framework, while at inference a classical Largest Connected Component (LCC) post-processing step is applied to remove false-positive connected components.

DyCON dynamically evaluates prediction uncertainty via entropy and applies voxel-level contrastive learning, exploiting the unlabeled scans. The main contribution of this work is a systematic comparison of SSM and CNN architectures for semi-supervised segmentation under limited spatial resolution, together with an analysis of the impact of LCC post-processing on boundary accuracy.

The experimental evaluation was conducted on the Pancreas-CT dataset. Under strict hardware constraints (64x64x64 patches), LCC drastically improves the boundary metrics (HD95, ASD) of both architectures. The U-Net3D + DyCON + LCC combination yields the highest and most stable accuracy, reaching with only 12 annotated scans performance equivalent to fully supervised training on 62 scans. In this limited-resolution regime, the convolutional network outperforms Swin-UMamba.

Keywords

Medical Image Segmentation, Deep Learning, State Space Models (SSMs), Semi-Supervised Learning, Mamba, DyCON, Pancreas

Ευχαριστίες

Θα ήθελα καταρχάς να εκφράσω τις θερμές μου ευχαριστίες στον επιβλέποντα καθηγητή μου, κ. Αθανάσιο Βουλόδημο, για την ευκαιρία που μου έδωσε να εκπονήσω την παρούσα διπλωματική εργασία, καθώς και για την πολύτιμη καθοδήγηση, την υποστήριξη και την εμπιστοσύνη του σε όλη τη διάρκεια της προσπάθειας αυτής.

Επίσης, θα ήθελα να ευχαριστήσω ιδιαίτερα την κα Παρασκευή Τζούβελη και τον κ. Νικόλαο Σπανό για την αμέριστη βοήθειά τους, τις ουσιαστικές συμβουλές τους και την εξαιρετική συνεργασία που είχαμε σε κάθε στάδιο της υλοποίησης και της συγγραφής. Η συμβολή τους ήταν καθοριστική για την ολοκλήρωση αυτού του έργου.

Τέλος, ένα μεγάλο και ειλικρινές ευχαριστώ ανήκει στους γονείς μου και την οικογένειά μου για τη συνεχή ηθική συμπαράσταση, την υπομονή και την ενθάρρυνση που μου προσέφεραν απλόχερα καθ' όλη τη διάρκεια των σπουδών μου.

Αθήνα, Μάιος 2026

Δημήτριος Λιάκης

Περιεχόμενα

Περίληψη 7

Abstract 9

Ευχαριστίες 11

Πρόλογος 19

1 Εισαγωγή 21
1.1 Αντικείμενο της διπλωματικής . . . 22
1.2 Οργάνωση του τόμου . . . 23

I Θεωρητικό Μέρος 25

2 Θεωρητικό Υπόβαθρο και Μαθηματική Θεμελίωση 27
2.1 Κατάτμηση Ιατρικής Εικόνας . . . 27
2.1.1 Τρόποι Ιατρικής Απεικόνισης . . . 27
2.1.2 Ιστορική Εξέλιξη . . . 28
2.2 Προκλήσεις στην Κατάτμηση Ιατρικής Εικόνας . . . 28
2.2.1 Προκλήσεις Σχετιζόμενες με το Σύνολο Δεδομένων . . . 28
2.2.2 Προκλήσεις Σχετιζόμενες με τα Μοντέλα Βαθιάς Μάθησης . . . 29
2.3 Μοντέλα Βασισμένα σε Συνελικτικά Νευρωνικά Δίκτυα . . . 29
2.3.1 Η Πράξη Συνέλιξης . . . 30
2.3.2 Δομικά Επίπεδα CNN . . . 30
2.3.3 Εξέλιξη Αρχιτεκτονικών και Υπολειπόμενα Δίκτυα . . . 31
2.3.4 Η Αρχιτεκτονική U-Net . . . 31
2.3.5 Το Μοντέλο 3D U-Net . . . 33
2.4 Vision Transformers και Μηχανισμοί Προσοχής . . . 33
2.4.1 Swin Transformer: Ιεραρχικά Παράθυρα . . . 34
2.5 State Space Models (SSMs) και η Αρχιτεκτονική Mamba . . . 36
2.5.1 Θεμελίωση και Διακριτοποίηση . . . 36
2.5.2 Το Μοντέλο S4 — Δομημένα SSMs . . . 36
2.5.3 Ο Επιλεκτικός Μηχανισμός Mamba (S6) . . . 37
2.6 Το SwinUMamba και η Μετα-επεξεργασία LCC . . . 38
2.6.1 Μετα-επεξεργασία Largest Connected Component (LCC) . . . 38
2.7 Το Πλαίσιο Ημι-επιβλεπόμενης Μάθησης DyCON . . . 39

2.7.1 Αρχιτεκτονική Δασκάλου-Μαθητή (Mean Teacher) . . . 39
2.7.2 Δυναμική Επίγνωση Αβεβαιότητας και $\mathcal{L}_{\text{UnCL}}$ . . . 40
2.7.3 Focal Entropy-aware Contrastive Loss (FeCL) . . . 41
2.7.4 Ο Συνολικός Αγωγός: SwinUMamba, DyCON και LCC . . . 41

3 Σχετικές Εργασίες και Βιβλιογραφική Ανασκόπηση 43
3.1 Ιστορική Εξέλιξη Αρχιτεκτονικών στην Ιατρική Κατάτμηση . . . 43
3.1.1 Η Εποχή των Συνελικτικών Δικτύων (CNNs) . . . 43
3.1.2 Vision Transformers (ViTs) και Υβριδικά Μοντέλα . . . 43
3.1.3 State Space Models (SSMs) και Mamba . . . 44
3.2 Ημι-επιβλεπόμενη Μάθηση (SSL) στην Ιατρική Εικόνα . . . 44
3.2.1 Μάθηση Συνέπειας και Αντιθετική Μάθηση . . . 44
3.2.2 Διαχείριση Αβεβαιότητας και το Πλαίσιο DyCON . . . 44

II Πρακτικό Μέρος 45

4 Σύνολα Δεδομένων και Υλοποίηση 47
4.1 Σύνολο Δεδομένων Pancreas-CT . . . 47
4.1.1 Δημογραφικά και Κλινικά Χαρακτηριστικά . . . 47
4.1.2 Τεχνικές Προδιαγραφές και Ground-Truth . . . 47
4.1.3 Διαχωρισμός Δεδομένων (Data Split) . . . 48
4.2 Υπολογιστικοί Πόροι και Λεπτομέρειες Υλοποίησης . . . 48
4.2.1 Αντιμετώπιση Περιορισμών Μνήμης (OOM Mitigations) . . . 48
4.2.2 Υπερπαράμετροι και Προ-εκπαίδευση (Pre-training) . . . 49

5 Υλοποίηση 51
5.1 Λεπτομέρειες Υλοποίησης . . . 51
5.1.1 Περιβάλλον Ανάπτυξης και Βιβλιοθήκες Λογισμικού . . . 51
5.1.2 Δομή Κώδικα και Οργάνωση Συστήματος . . . 51
5.1.3 Ροή Εκπαίδευσης και Αλγοριθμική Διαχείριση . . . 52
5.1.4 Μηχανική Βελτιστοποίησης και Πόρων Μνήμης . . . 53

6 Παρουσίαση και Σχολιασμός Πειραματικών Αποτελεσμάτων 55
6.1 Μεθοδολογία Ελέγχου . . . 55
6.2 Αναλυτική Παρουσίαση Ελέγχου . . . 56
6.2.1 Μελέτη Αποδόμησης (Ablation Study) . . . 56
6.2.2 Σχολιασμός των Ωμών Προβλέψεων (RAW Αποτελέσματα) . . . 58
6.2.3 Επίδραση της Μετα-επεξεργασίας LCC . . . 58
6.2.4 Αποδοτικότητα του Πλαισίου DyCON . . . 58

III Επίλογος 61

7 Επίλογος 63
7.1 Συμπεράσματα . . . 63

7.2 Περιορισμοί της Μελέτης 64
7.3 Μελλοντικές Επεκτάσεις 65

Βιβλιογραφία **70**

Συντομογραφίες - Αρκτικόλεξα - Ακρωνύμια **71**

Απόδοση ξενόγλωσσων όρων **73**

Κατάλογος Σχημάτων

2.1 Τυπική αρχιτεκτονική ενός Συνελικτικού Νευρωνικού Δικτύου (CNN). 31
2.2 Η συμμετρική αρχιτεκτονική U-Net με τις χαρακτηριστικές συνδέσεις παράκαμψης. 33
2.3 Η ογκομετρική αρχιτεκτονική 3D U-Net. 34

4.1 Απεικόνιση του παγκρέατος στο σύνολο δεδομένων Pancreas-CT. Από αριστερά προς τα δεξιά : Τρισδιάστατη αναπαράσταση επιφάνειας (surface rendering) και δισδιάστατες προβολές στο αξονικό (axial), στεφανιαίο (coronal) και οβελιαίο (sagittal) επίπεδο. 48

Κατάλογος Πινάκων

6.1 Συγκεντρωτικά Αποτελέσματα Κατάτμησης Παγκρέατος (Ωμές Προβλέψεις — RAW / Πριν το LCC) 56
6.2 Συγκεντρωτικά Αποτελέσματα Κατάτμησης Παγκρέατος (Τελικά Αποτελέσματα μετά από LCC) 56
6.3 Μελέτη Αποδόμησης (Ablation Study): Συνεισφορά κάθε συνιστώσας στην απόδοση κατάτμησης παγκρέατος. Το σύμβολο ✓ υποδηλώνει ενεργοποίηση του αντίστοιχου μηχανισμού. 57

Πρόλογος

Η παρούσα διπλωματική εργασία εκπονήθηκε στο Εργαστήριο Τεχνητής Νοημοσύνης της Σχολής Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Εθνικού Μετσόβιου Πολυτεχνείου, υπό την επίβλεψη του κ. Αθανάσιου Βουλόδημου, κατά το ακαδημαϊκό έτος 2025–2026.

Το ενδιαφέρον για το συγκεκριμένο ερευνητικό αντικείμενο γεννήθηκε από την πεποίθηση ότι η Τεχνητή Νοημοσύνη μπορεί να διαδραματίσει καθοριστικό ρόλο στη βελτίωση της ιατρικής διαγνωστικής και, κατ' επέκταση, στην ποιότητα ζωής των ασθενών. Η αυτόματη κατάτμηση ιατρικών εικόνων αποτελεί ένα από τα πλέον απαιτητικά και ουσιώδη προβλήματα στον τομέα αυτό, καθώς η ακριβής οριοθέτηση ανατομικών δομών αποτελεί προϋπόθεση για την έγκαιρη και αξιόπιστη διάγνωση σοβαρών παθολογιών.

Η εργασία ξεκίνησε τον Σεπτέμβριο του 2025 και ολοκληρώθηκε τον Μάιο του 2026. Στο διάστημα αυτό, μελετήθηκαν σε βάθος οι σύγχρονες αρχιτεκτονικές βαθιάς μάθησης για ιατρική απεικόνιση, με έμφαση στα Μοντέλα Χώρου Καταστάσεων και στις μεθόδους ημιεπιβλεπόμενης μάθησης, και υλοποιήθηκε ένα ολοκληρωμένο σύστημα κατάτμησης που συνδυάζει πρωτότυπα στοιχεία από τη διεθνή ερευνητική κοινότητα.

Αθήνα, Μάιος 2026

Δημήτριος Λιάκης

Κεφάλαιο 1

Εισαγωγή

Η υπολογιστική ανάλυση και η αυτόματη επεξεργασία ιατρικών εικόνων αποτελούν έναν από τους πλέον ραγδαία αναπτυσσόμενους και κρίσιμους τομείς της σύγχρονης βιοιατρικής μηχανικής και της τεχνητής νοημοσύνης. Μέσα από την ανάπτυξη εξελιγμένων αλγορίθμων, οι κλινικοί γιατροί αποκτούν ισχυρά εργαλεία για την υποβοήθηση της διάγνωσης, τον ακριβή σχεδιασμό σύνθετων θεραπευτικών ή χειρουργικών σχημάτων, καθώς και την παρακολούθηση της εξέλιξης σοβαρών παθολογιών.

Στο επίκεντρο αυτής της ερευνητικής προσπάθειας βρίσκεται η διαδικασία της κατάτμησης (medical image segmentation), η οποία αποσκοπεί στην αυτόματη και ακριβή οριοθέτηση ανατομικών δομών, οργάνων ή παθολογικών περιοχών ενδιαφέροντος (όπως όγκοι ή βλάβες) μέσα σε σύνθετες απεικονίσεις. Παρά την αλματώδη πρόοδο των βαθιών νευρωνικών δικτύων, η κατάτμηση συγκεκριμένων οργάνων της κοιλιακής χώρας —με χαρακτηριστικότερο παράδειγμα το πάγκρεας— παραμένει ένα εξαιρετικά απαιτητικό πρόβλημα υπολογιστικής όρασης [1]. Η υψηλή μορφολογική μεταβλητότητα του οργάνου ως προς το σχήμα και τη θέση του από ασθενή σε ασθενή, σε συνδυασμό με την εξαιρετικά χαμηλή αντίθεση (low contrast) και τα ασαφή όρια που παρουσιάζει με τους γειτονικούς ιστούς, καθιστούν τις κλασικές μεθόδους επιρρεπείς σε σημαντικά σφάλματα οριοθέτησης.

Παράλληλα, η επιτυχής εκπαίδευση σύγχρονων μοντέλων βαθιάς μάθησης (Deep Learning) προσκρούει σε έναν θεμελιώδη περιορισμό: την ανάγκη για μεγάλους όγκους δεδομένων με πυκνό και λεπτομερή σχολιασμό (ground-truth masks). Στην ιατρική απεικόνιση, η χειροκίνητη σχεδίαση αυτών των μασκών τομή προς τομή είναι μια διαδικασία εξαιρετικά χρονοβόρος, δαπανηρή και απαιτεί εξειδικευμένη κλινική εμπειρία. Για τον λόγο αυτό, η έρευνα στρέφεται έντονα προς την Ημι-επιβλεπόμενη Μάθηση (Semi-Supervised Learning), η οποία επιτρέπει στα μοντέλα να εκπαιδεύονται αξιοποιώντας έναν πολύ μικρό αριθμό σχολιασμένων εικόνων και ένα σημαντικά μεγαλύτερο πλήθος μη-σχολιασμένων τομογραφιών [2].

Η ανάγκη για μοντελοποίηση μακροπρόθεσμων εξαρτήσεων (long-range dependencies) χωρίς το απαγορευτικό υπολογιστικό κόστος των Vision Transformers (ViTs) οδήγησε πρόσφατα στην ενσωμάτωση των Μοντέλων Χώρου Καταστάσεων (State Space Models — SSMs) στην υπολογιστική όραση. Τα SSMs υιοθετήθηκαν στην παρούσα εργασία ακριβώς επειδή παρουσιάζουν βελτιωμένη ικανότητα σύλληψης μακροπρόθεσμων εξαρτήσεων σε σύγκριση με τους Vision Transformers: ενώ ο μηχανισμός αυτο-προσοχής των ViTs κοστίζει τετραγωνικά $O(N^2)$ ως προς το μήκος της ακολουθίας — καθιστώντας απαγορευτική τη διαχείριση μεγάλων

ογκομετρικών δεδομένων — τα SSMs διατηρούν παγκόσμιο πεδίο αποδοχής με γραμμική πολυπλοκότητα $O(N)$, επιτρέποντας τη μοντελοποίηση πολύ μακρύτερων εξαρτήσεων υπό τους ίδιους υπολογιστικούς πόρους. Το πλαίσιο Mamba [3] αποτελεί ένα τέτοιο εργαλείο, προσφέροντας παγκόσμιο πεδίο αποδοχής με γραμμική υπολογιστική πολυπλοκότητα $O(N)$. Ωστόσο, η επιπεδοποίηση (flattening) του τρισδιάστατου χώρου σε μια μονοδιάστατη ακολουθία για τις ανάγκες σάρωσης του Mamba τείνει να διαρρηγνύει τη φυσική χωρική γειτνίαση των ογκοστοιχείων, υποβαθμίζοντας τις λεπτομέρειες υψηλής συχνότητας στα ανατομικά όρια.

Αξιοποιώντας τις τεχνολογίες των State Space Models σε συνδυασμό με σύγχρονες τεχνικές ημι-επιβλεπόμενης μάθησης, καθίσταται δυνατή η ανάπτυξη συστημάτων ικανών να εκτελούν υψηλής ακρίβειας ιατρική κατάτμηση, ακόμη και σε περιβάλλοντα με εξαιρετικά περιορισμένα διαθέσιμα δεδομένα και υπολογιστικούς πόρους.

1.1 Αντικείμενο της διπλωματικής

Το βασικό ζήτημα που προκύπτει για τα συστήματα κατάτμησης 3Δ ιατρικής εικόνας υψηλής ανάλυσης είναι πώς τα μοντέλα μπορούν να συλλαμβάνουν το παγκόσμιο πλαίσιο, διατηρώντας ταυτόχρονα την τοπική ακρίβεια των ακμών υπό περιορισμένους υπολογιστικούς πόρους. Δύο κύριες προσεγγίσεις έχουν προταθεί στη βιβλιογραφία:

1. Η πρώτη προσέγγιση βασίζεται στα κλασικά τρισδιάστατα συνελικτικά δίκτυα, όπως το 3D U-Net [4]. Τα δίκτυα αυτά διατηρούν εγγενώς την ισχυρή χωρική σχέση μεταξύ των τομών, αλλά περιορίζονται από το τοπικό τους πεδίο αποδοχής (receptive field), αδυνατώντας να κατανοήσουν τη σφαιρική ανατομική δομή χωρίς να αυξήσουν δραματικά το βάθος και τον αριθμό των παραμέτρων τους.

2. Η δεύτερη προσέγγιση εισάγει μηχανισμούς αυτο-προσοχής (self-attention) μέσω αρχιτεκτονικών Transformers, όπως το UNETR [5]. Παρά την επιτυχία τους στην κατανόηση του συνολικού πλαισίου της εικόνας, η τετραγωνική πολυπλοκότητα $O(N^2)$ του πλήρους μηχανισμού αυτο-προσοχής καθιστά την εκπαίδευσή τους εξαιρετικά απαιτητική σε πόρους μνήμης γραφικών (VRAM), προκαλώντας συχνά σφάλματα υπερχείλισης (Out-Of-Memory) σε ογκομετρικές εικόνες υψηλής ανάλυσης.

Αντικείμενο της παρούσας διπλωματικής εργασίας είναι η σχεδίαση, η υλοποίηση και η πειραματική αξιολόγηση ενός συστήματος βαθιάς μάθησης για την ημι-επιβλεπόμενη κατάτμηση του παγκρέατος υπό περιορισμένους υπολογιστικούς πόρους.

Ως βασική αρχιτεκτονική (backbone) μελετάται η ιεραρχική αρχιτεκτονική Swin-UMamba [6], η οποία εκπαιδεύεται υπό το ημι-επιβλεπόμενο πλαίσιο DyCON [7]. Το DyCON υπολογίζει δυναμικά την αβεβαιότητα των προβλέψεων μέσω της εντροπίας Shannon, φιλτράροντας τον θόρυβο των ψευδο-ετικετών. Στη φάση της αξιολόγησης, εφαρμόζεται το κλασικό βήμα μετα-επεξεργασίας Largest Connected Component (LCC), το οποίο διατηρεί μόνο τη μεγαλύτερη συνεκτική συνιστώσα της πρόβλεψης, απομακρύνοντας τις ψευδώς-θετικές νησίδες. Για τη συγκριτική αποτίμηση, η ίδια διάταξη (DyCON + LCC) εφαρμόζεται και στο καθιερω-

μένο συνελικτικό δίκτυο U-Net3D, ώστε να διερευνηθεί η συμπεριφορά των δύο οικογενειών αρχιτεκτονικών υπό περιορισμένη χωρική ανάλυση.

Ένα κρίσιμο χαρακτηριστικό της εργασίας είναι η αξιολόγηση του συστήματος στο κλινικό σύνολο δεδομένων Pancreas-CT [8] υπό αυστηρούς περιορισμούς υλικού, προσομοιώνοντας ένα ρεαλιστικό σενάριο περιορισμένων υπολογιστικών πόρων. Μέσα από τη μείωση της χωρικής ανάλυσης σε 64×64×64 και την εφαρμογή συσσώρευσης κλίσης (gradient accumulation), μελετάται η ικανότητα του συστήματος να επιτυγχάνει υψηλή ακρίβεια κατάτμησης χρησιμοποιώντας μόλις 12 σχολιασμένες τομογραφίες.

1.2 Οργάνωση του τόμου

Η εργασία αυτή είναι οργανωμένη σε επτά κεφάλαια:

- Στο **Κεφάλαιο 2** δίνεται το θεωρητικό υπόβαθρο και η μαθηματική θεμελίωση των βασικών τεχνολογιών που σχετίζονται με τη διπλωματική. Αναλύεται η μαθηματική διατύπωση των CNNs, των Transformers, των μοντέλων SSMs (Mamba), καθώς και η διατύπωση των συναρτήσεων απώλειας του πλαισίου DyCON και η περιγραφή της μετα-επεξεργασίας LCC.

- Στο **Κεφάλαιο 3** παρατίθεται η βιβλιογραφική ανασκόπηση των σχετικών εργασιών, αποτυπώνοντας την εξέλιξη των αρχιτεκτονικών και των μεθόδων ημι-επιβλεπόμενης μάθησης.

- Στο **Κεφάλαιο 4** περιγράφονται αναλυτικά τα δημογραφικά και τεχνικά χαρακτηριστικά του συνόλου δεδομένων Pancreas-CT, ο διαχωρισμός των δεδομένων εκπαίδευσης, καθώς και οι στρατηγικές βελτιστοποίησης μνήμης που εφαρμόστηκαν για την αποφυγή σφαλμάτων OOM.

- Στο **Κεφάλαιο 5** παρατίθενται οι λεπτομέρειες υλοποίησης του συστήματος, περιγράφοντας τα προγραμματιστικά εργαλεία, τις βιβλιοθήκες, τη ροή εκπαίδευσης και τη δομή του κώδικα.

- Στο **Κεφάλαιο 6** παρουσιάζεται η ποσοτική αξιολόγηση του συστήματος με βάση τις μετρικές Dice, HD95 και ASD, συνοδευόμενη από εκτενή σχολιασμό για τη συμβολή κάθε συνιστώσας.

- Τέλος, στο **Κεφάλαιο 7** αποτιμώνται τα τελικά συμπεράσματα της διπλωματικής εργασίας και προτείνονται συγκεκριμένες κατευθύνσεις για μελλοντικές ερευνητικές επεκτάσεις του συστήματος.

Μέρος I

Θεωρητικό Μέρος

Κεφάλαιο 2

Θεωρητικό Υπόβαθρο και Μαθηματική Θεμελίωση

Στο κεφάλαιο αυτό παρουσιάζονται οι βασικές θεωρητικές αρχές που είναι απαραίτητες για την κατανόηση της διπλωματικής εργασίας. Αρχικά, αναλύεται η έννοια της κατάτμησης ιατρικής εικόνας και οι τρόποι ιατρικής απεικόνισης. Στη συνέχεια, παρατίθενται οι βασικές προκλήσεις του πεδίου και η εις βάθος μαθηματική θεμελίωση των αρχιτεκτονικών βαθιάς μάθησης που χρησιμοποιούνται: από τα παραδοσιακά συνελικτικά δίκτυα και τους Vision Transformers, μέχρι τα Μοντέλα Χώρου Καταστάσεων (SSMs), το ιεραρχικό μοντέλο SwinUMamba με τη μετα-επεξεργασία LCC, και το ημι-επιβλεπόμενο πλαίσιο DyCON.

2.1 Κατάτμηση Ιατρικής Εικόνας

Η κατάτμηση ιατρικής εικόνας (medical image segmentation) αποτελεί μια κομβική τεχνική στον τομέα της ιατρικής απεικόνισης, η οποία περιλαμβάνει τη διαδικασία διαχωρισμού των ιατρικών εικόνων σε νοηματοδοτημένες και διακριτές περιοχές. Η οριοθέτηση αυτών των συγκεκριμένων περιοχών επιτρέπει στους επαγγελματίες υγείας να επιτύχουν ακριβή ανάλυση ανατομικών δομών και παθολογικών ευρημάτων, γεγονός απαραίτητο για την αποτελεσματική διάγνωση, τον σχεδιασμό θεραπείας και την παρακολούθηση εξέλιξης νόσων.

2.1.1 Τρόποι Ιατρικής Απεικόνισης

Τα σύγχρονα συστήματα ιατρικής απεικόνισης παράγουν δεδομένα σε ποικιλία μορφών, καθεμία με διαφορετικές φυσικές αρχές και κλινική αξία:

- **Αξονική Τομογραφία (CT):** Χρησιμοποιεί ακτινοβολία X από πολλαπλές γωνίες για να κατασκευάσει τρισδιάστατες εικόνες. Αποδίδει εξαιρετική αντίθεση για οστά και ασβεστοποιημένες δομές, ενώ αποτελεί τη βασική μέθοδο για απεικόνιση εσωτερικών οργάνων στο θώρακα και την κοιλία.

- **Μαγνητική Τομογραφία (MRI):** Βασίζεται στον πυρηνικό μαγνητικό συντονισμό. Προσφέρει ανώτερη ανάλυση μαλακών ιστών συγκριτικά με την CT και αποτελεί τη μέθοδο

εκλογής για νευροαπεικόνιση, ογκολογία και μυοσκελετική απεικόνιση. Η πολυπαραμετρική φύση του MRI (λόγω διαφορετικών αλληλουχιών T1, T2, FLAIR, DWI) ενισχύει την κλινική πληροφορία αλλά αυξάνει την πολυπλοκότητα της ανάλυσης.

- **Υπερηχογράφημα (Ultrasound):** Χρησιμοποιεί υπερηχητικά κύματα για απεικόνιση σε πραγματικό χρόνο. Χαρακτηρίζεται από υψηλό επίπεδο θορύβου (speckle noise) και τεχνημάτων, γεγονός που καθιστά την αυτόματη κατάτμηση ιδιαίτερα απαιτητική.

- **Τομογραφία Εκπομπής Ποζιτρονίων (PET):** Μετρά τη μεταβολική δραστηριότητα των ιστών και συνδυάζεται συχνά με CT ή MRI (PET/CT, PET/MRI) για ταυτόχρονη ανατομική και λειτουργική απεικόνιση.

2.1.2 Ιστορική Εξέλιξη

Ιστορικά, η κατάτμηση ιατρικής εικόνας πέρασε από τρεις εποχές. Στη χειρωνακτική εποχή, ιατροί και ακτινολόγοι σχεδίαζαν μόνοι τους τα όρια των δομών ενδιαφέροντος, μια διαδικασία εξαιρετικά χρονοβόρα και ανθρωποκεντρική. Στην ημι-αυτόματη εποχή, εμφανίστηκαν αλγόριθμοι βασισμένοι σε κατωφλίωση, ενεργά περιγράμματα (active contours), υδροχύτη (watershed) και ταξινόμηση k-means, που μείωσαν τον ανθρώπινο φόρτο αλλά απαιτούσαν χειροκίνητη ρύθμιση παραμέτρων. Η εποχή της βαθιάς μάθησης, που ξεκίνησε ουσιαστικά με το U-Net [9] το 2015, επέτρεψε την πλήρη αυτοματοποίηση της κατάτμησης σε κλινική κλίμακα, με αποτελέσματα που συχνά εξισώνονται ή ξεπερνούν την ανθρώπινη επίδοση [1].

2.2 Προκλήσεις στην Κατάτμηση Ιατρικής Εικόνας

Η εφαρμογή αλγορίθμων βαθιάς μάθησης (deep learning algorithms) παρουσιάζει μοναδικές προκλήσεις, οι οποίες κατηγοριοποιούνται σε δύο ομάδες: προκλήσεις συνόλου δεδομένων και προκλήσεις μοντέλων [2].

2.2.1 Προκλήσεις Σχετιζόμενες με το Σύνολο Δεδομένων

Περιορισμένη Διαθεσιμότητα και Αραιοί Σχολιασμοί

Παρά τη συσσώρευση δεδομένων στα συστήματα PACS (Picture Archiving and Communication Systems), οι ιατρικές εικόνες συνοδεύονται κυρίως από μη δομημένες αναφορές ελεύθερου κειμένου. Τα μοντέλα βαθιάς μάθησης απαιτούν μεγάλους όγκους λεπτομερώς σχολιασμένων δεδομένων (annotated data), κάτι εξαιρετικά δαπανηρό και χρονοβόρο, καθώς ο σχολιασμός απαιτεί εξειδικευμένη ιατρική γνώση. Στην κλινική πράξη, πραγματοποιείται συχνά αραιός σχολιασμός (sparse annotation) σε επιλεγμένες μόνο τομές ενός τρισδιάστατου ογκομετρικού δεδομένου. Για την ελάφρυνση αυτού του βάρους, επιστρατεύονται τεχνικές επαύξησης (augmentation) και πλαίσια ημι-επιβλεπόμενης μάθησης (semi-supervised learning).

Θόρυβος Ετικετών και Ασυμμετρία Κλάσεων

Οι αποκλίσεις στις ερμηνείες διαφορετικών ακτινολόγων εισάγουν θόρυβο στις ετικέτες (label noise). Η ενδο-παρατηρητική και δια-παρατηρητική μεταβλητότητα (intra- and inter-observer variability) είναι ιδιαίτερα εμφανής σε μαλακούς ιστούς με ασαφή όρια, όπως ο εγκέφαλος ή το πάγκρεας. Παράλληλα, τα δεδομένα υποφέρουν από έντονη ασυμμετρία κλάσεων (class imbalance), όπου ο παθολογικός ιστός (π.χ. ένας μικρός όγκος) υποεκπροσωπείται δραματικά έναντι του υγιούς φόντου. Ένα μοντέλο μπορεί να εγκλωβιστεί σε τοπικά ελάχιστα (local minima) προβλέποντας διαρκώς την πλειοψηφική κλάση, ελαχιστοποιώντας έτσι την απώλεια διασταυρούμενης εντροπίας αλλά αποτυγχάνοντας στην κλινικά κρίσιμη κλάση.

Ανομοιογένειες Έντασης και Ασαφή Όρια

Τα ιατρικά δεδομένα παρουσιάζουν συχνά τεχνήματα (artifacts) και ανομοιογένειες έντασης λόγω κίνησης ασθενούς, μαγνητικής παρεμβολής ή αναπνευστικής κίνησης. Όργανα όπως το πάγκρεας διαθέτουν εξαιρετικά ασαφή όρια (weak boundaries) και παρόμοια επίπεδα γκρι με τους γειτονικούς ιστούς, καθιστώντας τον διαχωρισμό τους ένα από τα πλέον δύσκολα προβλήματα υπολογιστικής όρασης στον τομέα.

2.2.2 Προκλήσεις Σχετιζόμενες με τα Μοντέλα Βαθιάς Μάθησης

Οι κύριες προκλήσεις στο επίπεδο των μοντέλων αφορούν:

- **Υπερεκπαίδευση (Overfitting):** Λόγω μικρών συνόλων δεδομένων, τα δίκτυα τείνουν να απομνημονεύουν τα δείγματα εκπαίδευσης αντί να γενικεύουν. Η χρήση Dropout, Batch Normalization και τεχνικών επαύξησης αποτελεί την κύρια αντιμέτρωση.
- **Περιορισμοί Μνήμης Γραφικών (VRAM Out-of-Memory):** Κατά την επεξεργασία πλήρους ογκομετρικού 3D CT/MRI, ο όγκος των δεδομένων υπερβαίνει συχνά τη διαθέσιμη μνήμη GPU. Η λύση εκπαίδευσης σε τμήματα (patches) εισάγει νέο πρόβλημα: το δίκτυο «βλέπει» μόνο τοπικό τμήμα του ογκομετρικού χώρου, χάνοντας παγκόσμια ανατομική πληροφορία.
- **Εξαφάνιση Κλίσης (Vanishing Gradient):** Σε πολύ βαθιά δίκτυα, η κλίση που οπισθοδιαδίδεται γίνεται εκθετικά μικρή, εμποδίζοντας την αποτελεσματική εκπαίδευση των αρχικών επιπέδων. Η εισαγωγή υπολειπόμενων συνδέσεων (residual connections) από το ResNet [10] αποτελεί την κλασική λύση.

2.3 Μοντέλα Βασισμένα σε Συνελικτικά Νευρωνικά Δίκτυα

Τα Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks — CNNs) συγκαταλέγονται στις πιο ισχυρές και ευρέως διαδεδομένες αρχιτεκτονικές στον χώρο της βαθιάς μάθησης, με ιδιαίτερη επιτυχία στις εφαρμογές υπολογιστικής όρασης. Η θεμελιώδης ιδέα των CNNs εισήχθη από τον Kunihiko Fukushima [11], εμπνευσμένη από την ιεραρχική δομή των πεδίων αποδοχής του οπτικού φλοιού, όπως περιγράφηκε από τους Hubel και Wiesel

[12]. Σημαντικό ορόσημο αποτέλεσαν οι Waibel et al. [13] με την εφαρμογή κοινών βαρών (shared weights) και οπισθοδιάδοσης (backpropagation), ενώ ο Yann LeCun [14] ανέπτυξε εξειδικευμένη αρχιτεκτονική CNN για αναγνώριση εγγράφων.

2.3.1 Η Πράξη Συνέλιξης

Η βασική λειτουργία ενός CNN βασίζεται στη συνέλιξη (convolution). Στον τρισδιάστατο (3D) χώρο ιατρικών δεδομένων, η συνέλιξη ενός τανυστή εισόδου I με ένα μαθητεύσιμο φίλτρο (πυρήνα) K διαστάσεων $D_K \times H_K \times W_K$ εκφράζεται μαθηματικά ως:

$$S(x, y, z) = (I * K)(x, y, z) = \sum_i \sum_j \sum_k I(x - i, y - j, z - k)\, K(i, j, k)$$

Δύο θεμελιώδεις ιδιότητες της πράξης καθιστούν τα CNNs ιδιαίτερα κατάλληλα για ανάλυση εικόνας: (α) η **ισοδυναμία μετατόπισης (translation equivariance)**, που σημαίνει ότι η ανίχνευση ενός χαρακτηριστικού δεν εξαρτάται από τη θέση του στην εικόνα, και (β) η **κοινή χρήση βαρών (weight sharing)**, που μειώνει δραματικά τον αριθμό παραμέτρων σε σύγκριση με τα πλήρως συνδεδεμένα δίκτυα.

2.3.2 Δομικά Επίπεδα CNN

Μια τυπική αρχιτεκτονική CNN αποτελείται από τρεις κύριους τύπους επιπέδων:

1. **Συνελικτικά Επίπεδα (Convolutional Layers):** Χρησιμοποιούν εκπαιδεύσιμους πυρήνες που ολισθαίνουν πάνω στα δεδομένα εισόδου για την εξαγωγή χαρακτηριστικών, όπως ακμές, υφές και μοτίβα.

2. **Επίπεδα Κανονικοποίησης Δέσμης (Batch Normalization) [15]:** Κανονικοποιούν τους χάρτες χαρακτηριστικών ανά μίνι-δέσμη, σταθεροποιώντας τη διαδικασία εκπαίδευσης. Για ένα χαρακτηριστικό $x^{(k)}$, η κανονικοποίηση ορίζεται ως:

$$\hat{x}^{(k)} = \frac{x^{(k)} - \mathbb{E}[x^{(k)}]}{\sqrt{\mathrm{Var}[x^{(k)}] + \epsilon}}, \quad y^{(k)} = \gamma^{(k)} \hat{x}^{(k)} + \beta^{(k)}$$

   όπου γ, β είναι μαθητεύσιμες παράμετροι κλιμάκωσης και μετατόπισης. Η Κανονικοποίηση Δέσμης επιτρέπει υψηλότερους ρυθμούς μάθησης και μειώνει σημαντικά το πρόβλημα της εξαφάνισης/έκρηξης της κλίσης.

3. **Επίπεδα Ενεργοποίησης (Activation Layers):** Εφαρμόζουν μη-γραμμικές συναρτήσεις (π.χ. ReLU: $f(x) = \max(0, x)$, ή LeakyReLU: $f(x) = \max(ax, x)$ για μικρό $a > 0$) ανά στοιχείο, επιτρέποντας στο δίκτυο να μοντελοποιεί πολύπλοκες μη-γραμμικές σχέσεις.

4. **Επίπεδα Συγκέντρωσης (Pooling Layers):** Συμπυκνώνουν την πληροφορία από μικρές περιοχές των χαρτών χαρακτηριστικών (π.χ. μέσω max pooling ή average pooling), μειώνοντας τις χωρικές διαστάσεις και επιτυγχάνοντας μερική αναλλοίωτη μετατόπισης (translation invariance).

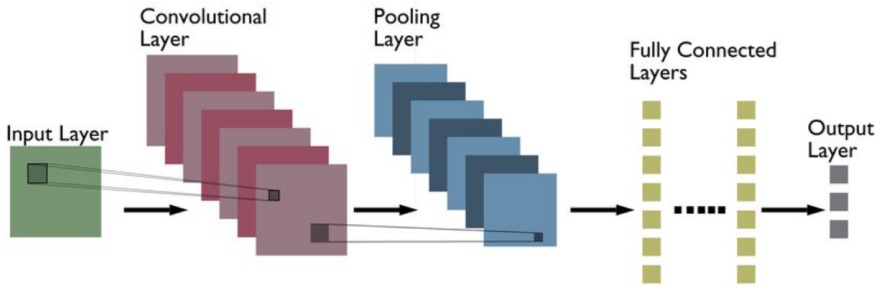
5. **Επίπεδα Dropout:** Κατά την εκπαίδευση, απενεργοποιούν τυχαία νευρώνες με πιθανότητα p. Αυτό λειτουργεί ως τεχνική κανονικοποίησης που αποτρέπει την υπερεκπαίδευση, εξαναγκάζοντας το δίκτυο να μαθαίνει πλεονάζουσες (redundant) αναπαραστάσεις.

2.3.3 Εξέλιξη Αρχιτεκτονικών και Υπολειπόμενα Δίκτυα

Αξιοσημείωτα ορόσημα αποτελούν το **AlexNet** [16] που ανανέωσε το ενδιαφέρον για βαθιά μάθηση, το **VGGNet** [17] με τα μικρά φίλτρα 3×3, και το **GoogLeNet/Inception** με τις παράλληλες συνελίξεις. Η πιο επαναστατική συνεισφορά ήλθε με το **ResNet** [10], που εισήγαγε τις **υπολειπόμενες συνδέσεις (residual connections)**:

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

Η σύνδεση ταυτότητας $\mathbf{x}$ παρακάμπτει ένα ή περισσότερα επίπεδα, επιτρέποντας στην κλίση να ρέει απευθείας σε βαθύτερα στρώματα και λύνοντας έτσι πρακτικά το πρόβλημα εξαφάνισης της κλίσης. Ένα τυπικό υπολειπόμενο μπλοκ αποτελείται από δύο συνελικτικά επίπεδα, δύο Batch Normalization και δύο ReLU, ενώ η σύνδεση ταυτότητας συχνά περιλαμβάνει γραμμική προβολή $W_s\mathbf{x}$ όταν οι διαστάσεις δεν ταιριάζουν.



Σχήμα 2.1: *Τυπική αρχιτεκτονική ενός Συνελικτικού Νευρωνικού Δικτύου (CNN).*

Παρά την τεράστια επιτυχία τους, τα CNNs περιορίζονται από το τοπικό τους πεδίο αποδοχής (receptive field). Για να «δει» ένα CNN ολόκληρη την εικόνα και να μοντελοποιήσει μακροπρόθεσμες εξαρτήσεις, απαιτείται τεράστιο βάθος δικτύου ή χρήση διευρυμένων συνελίξεων (dilated/atrous convolutions), που αυξάνουν τεχνητά το πεδίο αποδοχής εισάγοντας κενά στον πυρήνα.

2.3.4 Η Αρχιτεκτονική U-Net

Εξέλιξη και Κίνητρο

Πριν αναλυθεί η αρχιτεκτονική U-Net, είναι σημαντικό να κατανοηθεί η εξέλιξή της. Η πορεία ξεκίνησε με τον διαγωνισμό κατάτμησης ηλεκτρονικής μικροσκοπίας (EM) το 2012, όπου οι Cireșan et al. [18] ξεπέρασαν τις υπάρχουσες μεθόδους χρησιμοποιώντας CNNs με τεχνική εξαγωγής τμημάτων (image patching). Αν και η μέθοδος αύξησε τα δείγματα, εισήγαγε τεράστιο υπολογιστικό φόρτο λόγω επικάλυψης του συρόμενου παραθύρου (sliding window).

Το 2015, οι Ronneberger et al. [9] παρουσίασαν την αρχιτεκτονική U-Net, βασισμένη στο πλαίσιο των πλήρως συνελικτικών δικτύων (FCN) των Long et al. [19]. Το U-Net καθιέρωσε το πρότυπο στην ιατρική κατάτμηση μέσω μιας συμμετρικής δομής σχήματος «Y».

Μαθηματική Δομή

Το U-Net αποτελείται από:

(α) Κωδικοποιητής (Encoder / Contracting Path): Εφαρμόζει διαδοχικά μπλοκ από δύο 3×3 συνελίξεις, Batch Normalization, ReLU και 2×2 Max Pooling με βήμα 2. Σε κάθε βήμα οι χωρικές διαστάσεις υποδιπλασιάζονται και ο αριθμός καναλιών διπλασιάζεται (από 64 στο πρώτο επίπεδο).

(β) Αποκωδικοποιητής (Decoder / Expansive Path): Χρησιμοποιεί αντίστοιχα μπλοκ ανάδιπλωσης (up-convolution) 2×2 ή bilinear upsampling ακολουθούμενα από 3×3 συνελίξεις.

(γ) Συνδέσεις Παράκαμψης (Skip Connections): Αποτελούν τη σημαντικότερη καινοτομία. Ενώνουν τα επίπεδα εξαγωγής χαρακτηριστικών του κωδικοποιητή με τα αντίστοιχα επίπεδα ανακατασκευής του αποκωδικοποιητή, συνενώνοντας (concatenating) τη χωρική πληροφορία υψηλής ανάλυσης. Μαθηματικά, το χαρακτηριστικό F στο επίπεδο l του αποκωδικοποιητή υπολογίζεται ως:

$$F^{(l)}_{\text{decoder}} = \text{Conv}\Big(F^{(l)}_{\text{up}} \oplus F^{(l)}_{\text{encoder}}\Big)$$

όπου $\oplus$ συμβολίζει τη συνένωση κατά τη διάσταση των καναλιών (channel-wise concatenation).

Στρατηγική Εκπαίδευσης και Συναρτήσεις Απώλειας

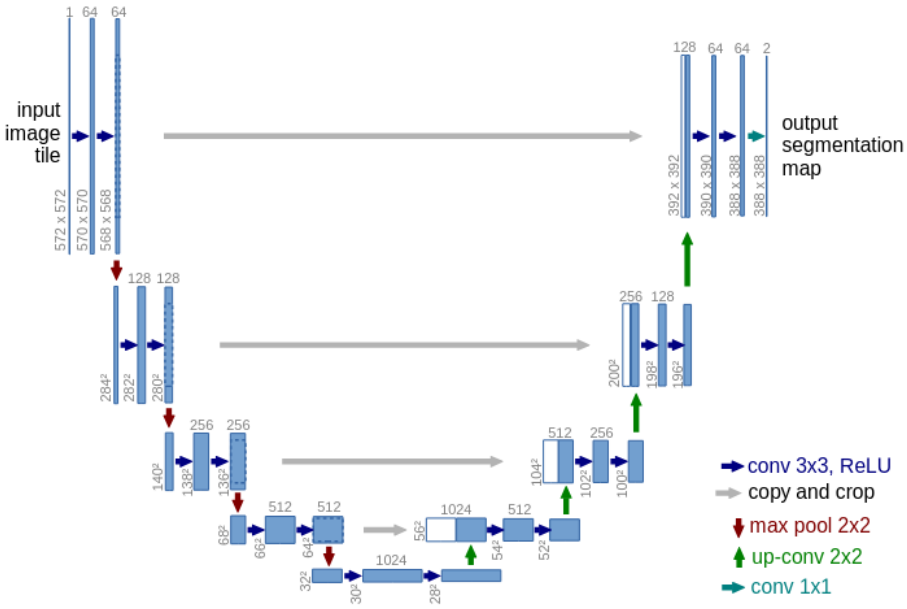
Στο αρχικό U-Net, η εκπαίδευση βασίστηκε σε εντατική επαύξηση δεδομένων μέσω ελαστικών παραμορφώσεων (elastic deformations), ειδικά σχεδιασμένων για ιατρικές εικόνες. Η τυπική συνάρτηση απώλειας για κατάτμηση είναι η σταθμισμένη διασταυρούμενη εντροπία (weighted cross-entropy), ενώ στην κλινική πράξη χρησιμοποιείται συχνά ο συνδυασμός:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{CE}} + \mathcal{L}_{\text{Dice}}$$

Η **σταθμισμένη διασταυρούμενη εντροπία** ορίζεται ως:

$$\mathcal{L}_{\text{CE}} = -\sum_{i} w(x_i)\, g_i \log p_i$$

όπου $w(x_i)$ είναι βάρος σχεδιασμένο να δίνει μεγαλύτερη έμφαση στα όρια μεταξύ επαφόμενων κυττάρων (touching cells) ή παθολογικών δομών. Ο συνδυασμός CE+Dice είναι ιδιαίτερα αποτελεσματικός σε δεδομένα με ασυμμετρία κλάσεων, καθώς η Dice απώλεια εξισορροπεί φυσικά τις κλάσεις.



Σχήμα 2.2: *Η συμμετρική αρχιτεκτονική U-Net με τις χαρακτηριστικές συνδέσεις παράκαμψης.*

2.3.5 Το Μοντέλο 3D U-Net

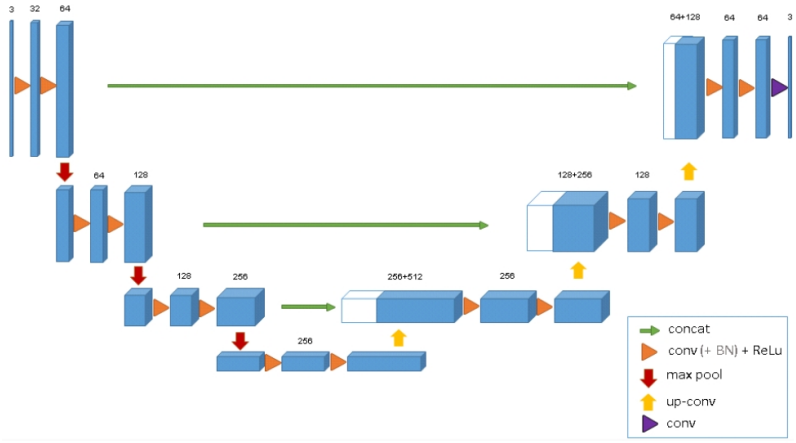
Για την ενίσχυση της αρχιτεκτονικής U-Net με πλουσιότερη τρισδιάστατη χωρική πληροφορία, οι Çiçek et al. πρότειναν το μοντέλο 3D U-Net [4]. Αυτό το μοντέλο παράγει πυκνές ογκομετρικές κατατμήσεις από έναν περιορισμένο αριθμό δισδιάστατων σχολιασμένων τομών, αντικαθιστώντας όλες τις 2D λειτουργίες (συνελίξεις, pooling, κανονικοποίηση) με αντίστοιχες 3D. Πιο συγκεκριμένα, τα μπλοκ κωδικοποιητή αποτελούνται από $3 \times 3 \times 3$ συνελίξεις, ενώ το max pooling γίνεται πλέον $2 \times 2 \times 2$.

Στρατηγικά, το 3D U-Net λειτουργεί σε δύο φάσεις: αρχικά, εκπαιδεύεται σε αραιά σχολιασμένες τομές, παράγοντας ψευδο-ετικέτες για το σύνολο του ογκομετρικού δεδομένου. Στη συνέχεια, εκπαιδεύεται εκ νέου στις πυκνές κατατμήσεις που παρήχθησαν. Σε μεταγενέστερες μελέτες, το 3D U-Net προσαρμόστηκε για διάφορες κλινικές εφαρμογές, ενώ οι Zeng et al. [20] διερεύνησαν την εφαρμογή πολυεπίπεδης βαθιάς επίβλεψης (multi-level deep supervision): προσθέτοντας μπλοκ αποσυνέλιξης (deconvolution) στα ενδιάμεσα επίπεδα, το μοντέλο παρήγαγε πολλαπλές εξόδους κατάτμησης, ενισχύοντας τη ροή της κλίσης (gradient flow).

Ένας αξιοσημείωτος περιορισμός του 3D U-Net είναι οι αυστηροί περιορισμοί στη μνήμη γραφικών (VRAM). Ο όγκος εισόδου δεν μπορεί να αυξηθεί αυθαίρετα, με αποτέλεσμα τη χρήση μικρότερων τμημάτων (patches), γεγονός που μπορεί να στερήσει από το δίκτυο τη δυνατότητα να συλλάβει πλήρως τις ευρύτερες ανατομικές δομές μιας ολόκληρης τομογραφίας.

2.4 Vision Transformers και Μηχανισμοί Προσοχής

Για την αντιμετώπιση του περιορισμένου πεδίου αποδοχής των παραδοσιακών συνελικτικών δικτύων, τα Vision Transformers (ViTs) [21] εισήγαγαν τον μηχανισμό αυτο-προσοχής



Σχήμα 2.3: *Η ογκομετρική αρχιτεκτονική 3D U-Net.*

(Self-Attention) στην ανάλυση εικόνας. Στην αρχιτεκτονική ViT, η εικόνα εισόδου διαιρείται σε ένα πλέγμα από μη-επικαλυπτόμενα τμήματα (patches) σταθερού μεγέθους (π.χ. 16×16). Τα τμήματα επιπεδοποιούνται, περνούν από γραμμική προβολή και, εμπλουτισμένα με κωδικοποιήσεις θέσης (positional embeddings), τροφοδοτούνται σε έναν Transformer Encoder.

Ο Encoder αποτελείται από επίπεδα Κανονικοποίησης (Layer Normalization), Πολυ-Κεφαλικής Αυτο-Προσοχής (Multi-Head Self-Attention — MSA) και Πολυστρωματικών Perceptrons (MLP). Ο μηχανισμός MSA προβάλλει γραμμικά την ακολουθία εισόδου $X \in \mathbb{R}^{N \times d}$ σε τρεις πίνακες — Ερωτήματα ($Q = XW_Q$), Κλειδιά ($K = XW_K$) και Τιμές ($V = XW_V$) — και ο υπολογισμός της προσοχής είναι:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V$$

Ο πίνακας $QK^\top$ υπολογίζει την ομοιότητα όλων των ζευγών τμημάτων, επιτρέποντας στο μοντέλο να αποτυπώσει παγκόσμιες εξαρτήσεις ανεξαρτήτως χωρικής απόστασης. Ωστόσο, η υπολογιστική πολυπλοκότητα $O(N^2)$ ως προς τον αριθμό τμημάτων N καθιστά την εφαρμογή των ViTs σε 3D ογκομετρικές εικόνες υψηλής ανάλυσης σχεδόν απαγορευτική.

2.4.1 Swin Transformer: Ιεραρχικά Παράθυρα

Για να γεφυρωθεί το χάσμα μεταξύ ViTs και CNNs, οι Liu et al. [22] παρουσίασαν τον **Swin Transformer**: μια ιεραρχική αρχιτεκτονική που διατηρεί την παγκόσμια μοντελοποίηση του μηχανισμού προσοχής, μειώνοντας παράλληλα την πολυπλοκότητα σε γραμμική $O(N)$ ως προς τον αριθμό των εικονοστοιχείων N.

Τμηματοποίηση και Ιεραρχία Χαρακτηριστικών

Ο Swin Transformer ξεκινά με μια **ενότητα ενσωμάτωσης τμημάτων (Patch Partition + Linear Embedding)**: η εικόνα $H \times W \times 3$ χωρίζεται σε μη-επικαλυπτόμενα τμήματα 4×4, τα οποία προβάλλονται σε C διαστάσεις. Η αρχιτεκτονική διαθέτει τέσσερα ιεραρχικά στάδια:

- Στάδιο 1: χαρτογράφηση $\frac{H}{4} \times \frac{W}{4} \times C$

- Στάδιο 2: **Συγχώνευση Τμημάτων (Patch Merging)** $\rightarrow \frac{H}{8} \times \frac{W}{8} \times 2C$
- Στάδιο 3: $\frac{H}{16} \times \frac{W}{16} \times 4C$
- Στάδιο 4: $\frac{H}{32} \times \frac{W}{32} \times 8C$

Κάθε Patch Merging συνενώνει 2×2 γειτονικά τμήματα, εφαρμόζει γραμμική προβολή και μειώνει τις χωρικές διαστάσεις κατά $2\times$, ενώ διπλασιάζει τα κανάλια. Αυτή η ιεραρχία χαρακτηριστικών είναι πολύτιμη για την κατάτμηση ιατρικής εικόνας, καθώς επιτρέπει τη σύλληψη τόσο λεπτομερών τοπικών πληροφοριών (πρώτα στάδια) όσο και ευρύτερων ανατομικών σχέσεων (τελευταία στάδια).

Αυτο-Προσοχή Παραθύρου (W-MSA) και Μετατοπισμένου Παραθύρου (SW-MSA)

Η βασική καινοτομία του Swin Transformer είναι ο περιορισμός του υπολογισμού προσοχής εντός μη-επικαλυπτόμενων τοπικών παραθύρων (windows) σταθερού μεγέθους $M \times M$. Κάθε παράθυρο περιέχει M^2 τμήματα, και η προσοχή υπολογίζεται ανεξάρτητα εντός κάθε παραθύρου:

$$\text{W-MSA complexity} = 4hwC^2 + 2M^2hwC \implies O(NM^2)$$

αντί για $O(N^2)$ του κλασικού ViT (όπου $N = hw$). Για $M \ll \sqrt{N}$, η μείωση πολυπλοκότητας είναι δραστική.

Ωστόσο, ο υπολογισμός εντός ανεξάρτητων παραθύρων αποτρέπει τη ροή πληροφορίας μεταξύ διαφορετικών περιοχών. Για τη δημιουργία διασυνδέσεων παραθύρων, χρησιμοποιείται εναλλακτικά η **Αυτο-Προσοχή Μετατοπισμένου Παραθύρου (SW-MSA)**: τα παράθυρα μετατοπίζονται κατά $\lfloor M/2 \rfloor$ θέσεις σε κάθε άξονα, δημιουργώντας νέα παράθυρα που καλύπτουν τα όρια των παλιών. Αποτελεσματικά, ένα ζεύγος διαδοχικών μπλοκ (W-MSA $\rightarrow$ SW-MSA) επιτρέπει διασυνδέσεις παραθύρων με ελάχιστο υπολογιστικό κόστος. Η εναλλαγή αυτή συνοψίζεται στις εξισώσεις ενός ζεύγους Swin blocks:

$$\begin{aligned}
\hat{z}^l &= \text{W-MSA}(\text{LN}(z^{l-1})) + z^{l-1} \\
z^l &= \text{MLP}(\text{LN}(\hat{z}^l)) + \hat{z}^l \\
\hat{z}^{l+1} &= \text{SW-MSA}(\text{LN}(z^l)) + z^l \\
z^{l+1} &= \text{MLP}(\text{LN}(\hat{z}^{l+1})) + \hat{z}^{l+1}
\end{aligned}$$

Τέλος, ο τύπος της προσοχής τροποποιείται ώστε να περιλαμβάνει **σχετική κωδικοποίηση θέσης (Relative Position Bias)** B:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}} + B\right)V$$

Τα βάρη B παραμετροποιούνται από έναν πίνακα $\hat{B}$ και μαθαίνονται κατά την εκπαίδευση, παρέχοντας πλουσιότερη χωρική πληροφορία σε σύγκριση με τις απόλυτες κωδικοποιήσεις θέσης του κλασικού ViT.

2.5 State Space Models (SSMs) και η Αρχιτεκτονική Mamba

Τα State Space Models γεφυρώνουν το χάσμα μεταξύ της καθολικής προσοχής και της υπολογιστικής αποδοτικότητας, προσφέροντας επιδόσεις αντίστοιχες των Transformers αλλά με τη γραμμική κλιμάκωση των RNNs.

2.5.1 Θεμελίωση και Διακριτοποίηση

Ένα συνεχές SSM απεικονίζει μια μονοδιάστατη ακολουθία εισόδου $x(t)$ σε μια έξοδο $y(t)$ μέσω μιας κρυφής κατάστασης (μνήμης) $h(t) \in \mathbb{R}^N$:

$$h'(t) = \mathbf{A}h(t) + \mathbf{B}x(t)$$

$$y(t) = \mathbf{C}h(t) + \mathbf{D}x(t)$$

Ο πίνακας $\mathbf{A} \in \mathbb{R}^{N\times N}$ είναι ο πίνακας κατάστασης (state matrix), καθοριστικός για τη μοντελοποίηση μακροχρόνιων εξαρτήσεων. Ο πίνακας $\mathbf{B} \in \mathbb{R}^{N\times 1}$ ελέγχει πώς η είσοδος επηρεάζει την κατάσταση, ο $\mathbf{C} \in \mathbb{R}^{1\times N}$ πώς η κατάσταση παράγει έξοδο, και ο $\mathbf{D}$ τον άμεσο δρόμο (skip connection) από είσοδο σε έξοδο.

Για να εφαρμοστεί σε διακριτά δεδομένα (ακολουθίες πεπερασμένου βήματος), το σύστημα διακριτοποιείται χρησιμοποιώντας τον κανόνα Συγκράτησης Μηδενικής Τάξης (Zero-Order Hold — ZOH) με βήμα δειγματοληψίας Δ:

$$\bar{\mathbf{A}} = \exp(\Delta\mathbf{A}), \qquad \bar{\mathbf{B}} = (\Delta\mathbf{A})^{-1}(\exp(\Delta\mathbf{A}) - \mathbf{I}) \cdot \Delta\mathbf{B}$$

Η αναδρομική εξίσωση:

$$h_k = \bar{\mathbf{A}}h_{k-1} + \bar{\mathbf{B}}x_k, \qquad y_k = \mathbf{C}\,h_k + \mathbf{D}\,x_k$$

εξασφαλίζει **γραμμική πολυπλοκότητα** $O(L)$ ως προς το μήκος ακολουθίας L.

2.5.2 Το Μοντέλο S4 — Δομημένα SSMs

Για να επιτευχθεί αποτελεσματική μοντελοποίηση μακροχρόνιων εξαρτήσεων, οι Gu et al. [23] ανέπτυξαν το **S4 (Structured State Space Sequence model)**. Το S4 βασίζεται σε δύο κύριες συνεισφορές:

(α) Αρχικοποίηση HiPPO: Ο πίνακας $\mathbf{A}$ αρχικοποιείται βάσει του πλαισίου HiPPO (High-order Polynomial Projection Operators) [24], που βελτιστοποιεί το SSM για συμπίεση συνεχών σημάτων σε πολυώνυμη βάση Legendre. Συγκεκριμένα, ο πίνακας HiPPO-LegS ορίζεται ως:

$$A_{nk} = -\begin{cases} (2n+1)^{1/2}(2k+1)^{1/2} & \text{αν } n > k \\ n+1 & \text{αν } n = k \\ 0 & \text{αν } n < k \end{cases}$$

Αυτή η δομή εγγυάται ότι το σύστημα «θυμάται» βέλτιστα το παρελθόν αναπαριστώντας το ως άθροισμα πολυωνύμων, ξεπερνώντας το πρόβλημα βραχύβιας μνήμης των RNNs.

(β) Δομημένη Παραμετροποίηση (NPLR/DPLR): Ο πίνακας $\mathbf{A}$ εκφράζεται ως άθροισμα κανονικού και χαμηλής τάξης (Normal Plus Low-Rank) πίνακα:

$$\mathbf{A} = \mathbf{V}\Lambda\mathbf{V}^{*} - \mathbf{P}\mathbf{Q}^{\top}$$

Αυτή η δομή επιτρέπει τον υπολογισμό του πυρήνα συνέλιξης $\bar{K} = (\mathbf{C}\bar{B}, \mathbf{C}\bar{A}\bar{B}, \mathbf{C}\bar{A}^{2}\bar{B}, \ldots)$ αποτελεσματικά μέσω ταχείας μετασχηματισμού Fourier, με πολυπλοκότητα $O(L \log L)$ κατά την εκπαίδευση και $O(1)$ ανά βήμα κατά την αναφορά (inference). Με αυτόν τον τρόπο, το S4 υπολογίζει επαρκώς μακροχρόνιες εξαρτήσεις σε ακολουθίες έως 10^4 βημάτων, ξεπερνώντας τους Transformers σε αρκετά benchmarks μακρόχρονων εξαρτήσεων, ενώ διατηρεί γραμμική χρονική πολυπλοκότητα.

2.5.3 Ο Επιλεκτικός Μηχανισμός Mamba (S6)

Στα κλασικά SSMs (και στο S4), οι πίνακες $\bar{\mathbf{A}}, \bar{\mathbf{B}}, \mathbf{C}$ παραμένουν σταθεροί ανεξαρτήτως εισόδου (Linear Time-Invariant — LTI). Αυτό σημαίνει ότι το σύστημα προσδίδει ίδια βαρύτητα σε όλα τα στοιχεία της ακολουθίας, ανεξαρτήτως περιεχομένου. Η αρχιτεκτονική Mamba [3] υπερβαίνει αυτόν τον περιορισμό εισάγοντας έναν **επιλεκτικό μηχανισμό σάρωσης (Selective Scan — S6)**: οι παράμετροι $\mathbf{B}$, $\mathbf{C}$ και Δ γίνονται συναρτήσεις της εκάστοτε εισόδου x_k:

$$\mathbf{B}_k = \text{Linear}_N(x_k), \qquad \mathbf{C}_k = \text{Linear}_N(x_k), \qquad \Delta_k = \text{Softplus}(\text{Parameter} + \mathbf{W}_\Delta x_k)$$

Η δυναμική παράμετρος Δ_k λειτουργεί ως πύλη πληροφορίας (information gate):

- **Μικρό** Δ_k**:** $\bar{\mathbf{A}} \approx \mathbf{I}, \bar{\mathbf{B}} \approx \mathbf{0} \rightarrow$ το σύστημα διατηρεί την προηγούμενη κατάσταση αγνοώντας το x_k (π.χ. αδιάφορο φόντο εικόνας).
- **Μεγάλο** Δ_k**:** το σύστημα «ξεχνά» ταχύτατα την προηγούμενη κατάσταση και ανανεώνεται από το x_k (π.χ. κλινικά σημαντική ανατομική δομή).

Επιπλέον, το Mamba επιτυγχάνει εξαιρετική ταχύτητα μέσω **αλγορίθμου με επίγνωση υλικού (Hardware-Aware Algorithm)**: κατά τη σάρωση, ο υπολογισμός των καταστάσεων h_k εκτελείται απευθείας στη γρήγορη μνήμη SRAM της κάρτας γραφικών, αποφεύγοντας περιττές εγγραφές και αναγνώσεις από την αργή μνήμη HBM (High Bandwidth Memory), σε πλήρη αναλογία με τον μηχανισμό FlashAttention [25] για τους Transformers. Αυτό μεταφράζεται σε απόδοση εκτέλεσης έως 5× ταχύτερη και 5× μικρότερη χρήση μνήμης GPU σε σύγκριση με υλοποιήσεις αναφοράς [3].

Ένα πλήρες **μπλοκ Mamba** αποτελείται από: γραμμική προβολή ανόδου (linear expansion) της εισόδου σε δύο παράλληλες διαδρομές, μια σύντομη 1D συνέλιξη (depth-wise convolution) που αποτρέπει την ανεξάρτητη επεξεργασία κάθε στοιχείου, εφαρμογή του επιλεκτικού SSM στην πρώτη διαδρομή, έξοδο που ελέγχεται (gated) από τη δεύτερη διαδρομή μέσω ενός μη-γραμμικού μηχανισμού SiLU/Swish, και τέλος γραμμική προβολή εξόδου.

Αξίζει να σημειωθεί από νωρίς ότι τα θεωρητικά αυτά πλεονεκτήματα — παγκόσμιο πεδίο αποδοχής με γραμμική πολυπλοκότητα — εκδηλώνονται στο έπακρο όταν το μοντέλο επεξερ-

γάζεται αρκετά μακριές ακολουθίες, ώστε οι μακροπρόθεσμες εξαρτήσεις να υπερτερούν της τοπικής πληροφορίας. Όπως θα φανεί στο Κεφάλαιο 6, υπό αυστηρό περιορισμό χωρικής ανάλυσης (μικρά τμήματα $64 \times 64 \times 64$) η ισορροπία αυτή αντιστρέφεται, και τα συνελικτικά δίκτυα με ισχυρό τοπικό πεδίο αποδοχής αναδεικνύονται ανταγωνιστικότερα.

2.6 Το SwinUMamba και η Μετα-επεξεργασία LCC

Επειδή ο μηχανισμός Mamba είναι εγγενώς μονοδιάστατος (σχεδιασμένος για κείμενο ή ήχο), δεν διαθέτει χωρική αντίληψη. Το SwinUMamba [6] ενσωματώνει τα πλεονεκτήματα του Mamba στα 3D ιατρικά δεδομένα εικόνας μέσω τριών καινοτομιών.

(α) Ιεραρχικός Κωδικοποιητής μέσω Patch Merging: Δανειζόμενο από τον Swin Transformer [22], το SwinUMamba οργανώνει τα χαρακτηριστικά σε τέσσερα ιεραρχικά στάδια. Σε κάθε στάδιο, η Patch Merging μονάδα συνενώνει $2 \times 2 \times 2$ γειτονικά τεμαχίδια και εφαρμόζει γραμμική προβολή για να μειώσει τις χωρικές διαστάσεις κατά $2\times$ και να διπλασιάσει τα κανάλια, παράγοντας ιεραρχικές αναπαραστάσεις πολλαπλών κλιμάκων.

(β) Ενότητα Σταυροειδούς Σάρωσης (Cross-Scan Module — CSM): Για τη διαχείριση της χωρικής πληροφορίας ενός 3D τανυστή, το CSM ξεδιπλώνει τον ογκομετρικό τανυστή σε μονοδιάστατες ακολουθίες σαρώνοντάς τον ταυτόχρονα προς έξι κατευθύνσεις (πολυκατευθυντικά) κατά μήκος των τριών αξόνων: N→S, E→W, D→U και οι αντίστροφες κατευθύνσεις. Για κάθε κατεύθυνση $d \in \{1, \ldots, 6\}$:

$$y_d = \mathrm{SSM}_d(\mathrm{Flatten}_d(F))$$

Τα αποτελέσματα y_d συγχωνεύονται στο τέλος με πρόσθεση ή συνένωση, διασφαλίζοντας ότι η χωρική συνάφεια δεν χάνεται παρά την επιπεδοποίηση.

(γ) Τεχνική Largest Connected Component (LCC): Σημειώνεται ότι, ακόμη και με το CSM, η επιπεδοποίηση ενός ογκομετρικού τανυστή $H \times W \times D$ σε ακολουθία μήκους $L = H \cdot W \cdot D$ μπορεί να διαρρήξει τη φυσική χωρική γειτνίαση των ογκοστοιχείων, αποτελώντας εγγενή περιορισμό του μηχανισμού Mamba που επηρεάζει την ακρίβεια στα ανατομικά όρια. Ανεξάρτητα από αυτόν τον αρχιτεκτονικό περιορισμό, στην παρούσα εργασία εφαρμόζεται ως διακριτό βήμα μετα-επεξεργασίας η τεχνική Largest Connected Component (LCC), η οποία περιγράφεται στη συνέχεια.

2.6.1 Μετα-επεξεργασία Largest Connected Component (LCC)

Μετά την παραγωγή της δυαδικής πρόβλεψης από το δίκτυο, εφαρμόζεται ένα κλασικό βήμα μετα-επεξεργασίας (post-processing) γνωστό ως Largest Connected Component (LCC). Στόχος του είναι η απαλοιφή των ψευδώς-θετικών «νησίδων» — μικρών, αποσπασματικών συνεκτικών συνιστωσών που το μοντέλο εσφαλμένα ταξινομεί ως πάγκρεας σε απομακρυσμένες περιοχές της κοιλιακής χώρας.

Η διαδικασία έχει ως εξής: η δυαδική μάσκα αναλύεται στις συνεκτικές της συνιστώσες (connected components) στον 3D χώρο μέσω ανάλυσης γειτνίασης (26-συνεκτικότητα). Για κάθε συνιστώσα C_c υπολογίζεται ο όγκος της $|C_c|$ (πλήθος ογκοστοιχείων), και διατηρείται αποκλειστικά η συνιστώσα με τον μεγαλύτερο όγκο, η οποία αντιστοιχεί στο όργανο-στόχο:

$$M_{\text{LCC}} = \left\{ v : \text{label}(v) = \arg\max_{c} |C_c| \right\}$$

Όλες οι υπόλοιπες συνιστώσες μηδενίζονται. Το LCC είναι αμιγώς γεωμετρικό/τοπολογικό, δεν περιλαμβάνει μαθησιακές παραμέτρους και εφαρμόζεται μόνο κατά τη φάση της αξιολόγησης (inference), ανεξάρτητα από την υποκείμενη αρχιτεκτονική.

Αν και απλό, το LCC έχει δυσανάλογα μεγάλη επίδραση στις μετρικές ορίων: τόσο η απόσταση Hausdorff (HD95) όσο και η Μέση Επιφανειακή Απόσταση (ASD) κυριαρχούνται από τα πιο απομακρυσμένα εσφαλμένα ογκοστοιχεία. Αφαιρώντας τις ψευδώς-θετικές νησίδες, το LCC μειώνει δραστικά αυτές τις αποστάσεις, όπως επιβεβαιώνεται στα πειραματικά αποτελέσματα. Η προσέγγιση προϋποθέτει τη συνεκτικότητα του οργάνου-στόχου· για το πάγκρεας, που αποτελεί ένα ενιαίο όργανο, η υπόθεση αυτή είναι κατά κανόνα έγκυρη.

2.7 Το Πλαίσιο Ημι-επιβλεπόμενης Μάθησης DyCON

Το πλαίσιο DyCON [7] (εγκρίθηκε στο CVPR 2025) αντιμετωπίζει δύο ριζικά προβλήματα της ημι-επιβλεπόμενης κατάτμησης 3D ιατρικών εικόνων: (α) την **ασυμμετρία κλάσεων** μεταξύ παθολογικών και υγιών περιοχών, και (β) την **υψηλή αβεβαιότητα** λόγω παθολογικής μεταβλητότητας και ασαφών ορίων. Η αποτελεσματικότητα του DyCON αποδείχθηκε μέσω αξιολόγησης σε τέσσερα κλινικά σύνολα δεδομένων: εγκεφαλικά εμφράγματα (ISLES'22), όγκους εγκεφάλου (BraTS'19), αριστερό κόλπο (LA) και πάγκρεας (Pancreas).

2.7.1 Αρχιτεκτονική Δασκάλου-Μαθητή (Mean Teacher)

Το DyCON βασίζεται στην αρχιτεκτονική **Μέσου Δασκάλου (Mean Teacher)** [26]. Διατηρούνται δύο αντίγραφα του δικτύου: ο **μαθητής (student)** εκπαιδεύεται άμεσα μέσω οπισθοδιάδοσης, ενώ ο **δάσκαλος (teacher)** εκπαιδεύεται έμμεσα μέσω του Εκθετικού Κινητού Μέσου Όρου (Exponential Moving Average — EMA) των βαρών του μαθητή:

$$\theta_{\text{teacher}}^{(t)} = \alpha\, \theta_{\text{teacher}}^{(t-1)} + (1-\alpha)\, \theta_{\text{student}}^{(t)}$$

Η τιμή του συντελεστή α (συνήθως $\alpha = 0.99$ ή 0.999) διασφαλίζει ότι ο δάσκαλος αποτελεί μια «ευσταθέστερη» εκδοχή του μαθητή, παράγοντας πιο σταθερές και αξιόπιστες ψευδο-ετικέτες (pseudo-labels).

Δημιουργία Ψευδο-Ετικετών: Σε κάθε επανάληψη εκπαίδευσης, μια μη-σχολιασμένη εικόνα x_j τροφοδοτείται στον δάσκαλο μέσω **αδύναμης επαύξησης (weak augmentation)** (τυχαίος αντικατοπτρισμός, μικρές περιστροφές), και η έξοδος $\hat{y}_{j,w} = \arg\max p_w$ αποτελεί ψευδο-ετικέτα. Η ίδια εικόνα τροφοδοτείται στον μαθητή μέσω **ισχυρής επαύξησης (strong augmentation)** (CutMix, τυχαία αποκοπή (random erasing), θόρυβος), και η απόκλιση από την ψευδο-ετικέτα αποτελεί σήμα εκπαίδευσης. Η συνολική απώλεια ορίζεται ως:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{sup}} + \lambda_{\text{cons}} \cdot \mathcal{L}_{\text{UnCL}} + \lambda_{\text{contra}} \cdot \mathcal{L}_{\text{FeCL}}$$

Η εποπτευόμενη απώλεια $\mathcal{L}_{\text{sup}}$ υπολογίζεται μόνο στα σχολιασμένα δεδομένα (Dice + CE), ενώ οι $\mathcal{L}_{\text{UnCL}}$ και $\mathcal{L}_{\text{FeCL}}$ εκμεταλλεύονται και τα μη-σχολιασμένα.

2.7.2 Δυναμική Επίγνωση Αβεβαιότητας και $\mathcal{L}_{\text{UnCL}}$

Κατά τη μάθηση συνέπειας (consistency learning), το σύστημα τροφοδοτείται με μια ασθενώς επαυξημένη (weakly augmented) και μια ισχυρά επαυξημένη (strongly augmented) εκδοχή της ίδιας εικόνας. Για να αποφευχθεί ο θόρυβος επιβεβαίωσης (confirmation bias), το DyCON υπολογίζει την αβεβαιότητα σε επίπεδο εικονοστοιχείου i μέσω της Εντροπίας Shannon τόσο για το δίκτυο του μαθητή ($\mathcal{H}_s$) όσο και για το δίκτυο του δασκάλου ($\mathcal{H}_t$).

Η **Uncertainty-aware Consistency Loss (UnCL)** ενσωματώνει άμεσα αυτή την αβεβαιότητα, αντικαθιστώντας την τυπική απώλεια με μια δυναμικά σταθμισμένη μορφή. Συγκεκριμένα, η απώλεια UnCL ορίζεται μαθηματικά [7] ως εξής:

$$\mathcal{L}_{\text{UnCL}} = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{\mathcal{L}(p_{i,s}, p_{i,t})}{\exp(\beta \cdot \mathcal{H}_s(p_{i,s})) + \exp(\beta \cdot \mathcal{H}_t(p_{i,t}))} + \beta(\mathcal{H}_s(p_{i,s}) + \mathcal{H}_t(p_{i,t}))\right]$$

όπου:

- $\mathcal{L}(p_{i,s}, p_{i,t})$ είναι μια συνάρτηση απόστασης, τυπικά το Μέσο Τετραγωνικό Σφάλμα (MSE), που ευθυγραμμίζει τις πιθανότητες του μαθητή ($p_{i,s}$) με τις πιθανότητες του δασκάλου ($p_{i,t}$).
- Ο παρονομαστής $\exp(\beta \cdot \mathcal{H}_s) + \exp(\beta \cdot \mathcal{H}_t)$ **αντιστρέφει το βάρος** της απώλειας: εικονοστοιχεία με υψηλή αβεβαιότητα (μεγάλη εντροπία) λαμβάνουν μικρότερο βάρος ευθυγράμμισης, ενώ τα βέβαια εικονοστοιχεία κυριαρχούν στην απώλεια.
- Ο όρος $\beta(\mathcal{H}_s + \mathcal{H}_t)$ λειτουργεί ως **κανονικοποίηση εντροπίας (entropy regularization)**, αποτρέποντας την πρόωρη σύγκλιση (premature convergence) και ενθαρρύνοντας το μοντέλο να εξερευνήσει δύσκολες ανατομικές περιοχές.

Το ιδιαίτερο χαρακτηριστικό του UnCL είναι η **δυναμική υπερπαράμετρος** $\beta(t)$, η οποία λειτουργεί ως συντελεστής κλιμάκωσης και μειώνεται εκθετικά με την πάροδο των επαναλήψεων εκπαίδευσης t:

$$\beta(t) = \max\left(\beta_{\min},\ \beta_{\max} \cdot \exp\left(-\lambda \cdot \frac{t}{T}\right)\right)$$

Αυτός ο **φθίνων δυναμικός προγραμματισμός** εξασφαλίζει ότι:

- **Στα πρώιμα στάδια εκπαίδευσης (υψηλό β):** Η αβεβαιότητα έχει ισχυρό αντίκτυπο. Επιτρέπει στο μοντέλο να «εξερευνά» δύσκολες παθολογικές περιοχές χωρίς αυστηρές ποινές σε περίπτωση διαφωνίας μαθητή-δασκάλου. Τα εικονοστοιχεία υψηλής αβεβαιότητας δεν φιλτράρονται εξ ολοκλήρου, αλλά συνεισφέρουν ήπια στην απώλεια.
- **Στα ύστερα στάδια εκπαίδευσης (χαμηλό β):** Η επιρροή της αβεβαιότητας μειώνεται σταδιακά, οδηγώντας το μοντέλο στην εκλέπτυνση των αποφάσεών του σε βέβαιες περιοχές και διασφαλίζοντας την παγκόσμια συνέπεια ολόκληρου του ογκομετρικού δεδομένου.

Η αντίθεση με τυπικές μεθόδους, όπως το UAMT [27], είναι ουσιαστική: εκείνες αποκλείουν οριστικά τα εικονοστοιχεία με υψηλή εντροπία χάνοντας πολύτιμη πληροφορία στα όρια, ενώ το DyCON τα ενσωματώνει δυναμικά.

2.7.3 Focal Entropy-aware Contrastive Loss (FeCL)

Η **Focal Entropy-aware Contrastive Loss (FeCL)** στοχεύει στην επίλυση της ασυμμετρίας κλάσεων ενσωματώνοντας έναν εστιακό μηχανισμό στην αντιθετική απώλεια. Η βασική ιδέα είναι να δοθεί μεγαλύτερη βαρύτητα στα «δύσκολα» εικονοστοιχεία — εκείνα δηλαδή που το μοντέλο δυσκολεύεται να ταξινομήσει ορθά ή που βρίσκονται κοντά στα όρια απόφασης.

Ορίζοντας ως $\text{sim}(z_i, z_j) = \exp(z_i \cdot z_j/\tau)$ τη συνάρτηση ομοιότητας, η FeCL ορίζεται ως:

$$\mathcal{L}_{\text{FeCL}} = -\frac{1}{N}\sum_{i=1}^{N}\log\left[\frac{\displaystyle\sum_{z^+\in P(i)}(1-\text{sim}(z_i,z^+))^{\gamma}\cdot\text{sim}(z_i,z^+)}{\displaystyle\sum_{z^+\in P(i)}\text{sim}(z_i,z^+)+\sum_{z^-\in N(i)}(1-\text{sim}(z_i,z^-))^{\gamma}\cdot\text{sim}(z_i,z^-)}\right]$$

Όπου:

- **Ο όρος $(1-\textbf{sim})^{\gamma}$:** Είναι ο εστιακός παράγοντας (focal factor) με υπερπαράμετρο $\gamma \geq 0$. Όταν η ομοιότητα sim είναι υψηλή (εύκολο δείγμα), ο όρος $(1-\text{sim})^{\gamma}$ τείνει στο μηδέν, μειώνοντας τη συνεισφορά του δείγματος στην απώλεια. Αντίθετα, για «δύσκολα» δείγματα (χαμηλή ομοιότητα), ο όρος μεγιστοποιείται, αναγκάζοντας το δίκτυο να εστιάσει την εκπαίδευσή του σε αυτά.

- **Εντροπική Στάθμιση:** Τα βάρη των ζευγών προσαρμόζονται επιπλέον μέσω της αβεβαιότητας $\mathcal{H}_i$: δείγματα με υψηλή εντροπία εισάγονται με μεγαλύτερη προτεραιότητα στον μηχανισμό focal, ενισχύοντας τη διακρισιμότητα στα ασαφή ανατομικά όρια.

- **Εξόρυξη Top-k:** Για τη διατήρηση της υπολογιστικής αποδοτικότητας, ο παρονομαστής δεν περιλαμβάνει το σύνολο των αρνητικών δειγμάτων $N(i)$, αλλά μόνο τα k «δυσκολότερα» αρνητικά ζεύγη που ικανοποιούν τη συνθήκη της εστιακής απόστασης, αποτρέποντας την υπερφόρτωση της GPU.

Αυτός ο μηχανισμός διασφαλίζει ότι το δίκτυο δεν «σπαταλά» την ικανότητα εκμάθησης σε εύκολα παραδείγματα (όπως ο καθαρός ιστός του παγκρέατος), αλλά κατευθύνει τους πόρους του στην οριοθέτηση των δύσκολων, θολών ορίων που αποτελούν την κύρια πηγή σφάλματος στην κατάτμηση.

2.7.4 Ο Συνολικός Αγωγός: SwinUMamba, DyCON και LCC

Είναι σημαντικό να τονιστεί ότι το DyCON και το LCC λειτουργούν σε διαφορετικά, ανεξάρτητα στάδια του αγωγού και δεν αλληλεπιδρούν μεταξύ τους. Το DyCON επενεργεί *κατά την εκπαίδευση*: αξιοποιεί τις μη-σχολιασμένες τομογραφίες, διαχειρίζεται δυναμικά την αβεβαιότητα μέσω της εντροπίας και βελτιώνει τις αναπαραστάσεις κατα τη διάρκεια της μάθησης, οδηγώντας σε ποιοτικότερες ωμές προβλέψεις. Το LCC επενεργεί *κατά την αξιολόγηση*: λαμβάνει την τελική δυαδική πρόβλεψη και αφαιρεί τις απομακρυσμένες ψευδώς-θετικές συνεκτικές συνιστώσες, εξυγιαίνοντας τα ανατομικά όρια.

Τα δύο στάδια είναι συμπληρωματικά αλλά αυτοτελή το DyCON δεν εξαρτάται από το LCC, και το LCC εφαρμόζεται με τον ίδιο τρόπο σε κάθε μοντέλο ανεξαρτήτως μεθόδου εκπαίδευσης. Όπως θα δειχθεί στο Κεφάλαιο 6, ο συνδυασμός τους αποδίδει τα βέλτιστα

τελικά αποτελέσματα — όχι μέσω κάποιας αλληλεπίδρασης σε επίπεδο χαρακτηριστικών, αλλά επειδή αντιμετωπίζουν δύο διακριτές πηγές σφάλματος (την έλλειψη επισημασμένων δεδομένων και τα διάσπαρτα ψευδώς-θετικά αντίστοιχα).

Κεφάλαιο 3

Σχετικές Εργασίες και Βιβλιογραφική Ανασκόπηση

Στο κεφάλαιο αυτό παρουσιάζεται η βιβλιογραφική ανασκόπηση του ερευνητικού πλαισίου γύρω από τις αρχιτεκτονικές βαθιάς μάθησης στην ιατρική εικόνα. Παρατίθενται οι σχετικές εργασίες που σηματοδότησαν τη μετάβαση από τα παραδοσιακά συνελικτικά δίκτυα στα προηγμένα Μοντέλα Χώρου Καταστάσεων, και αναλύεται η ιστορική εξέλιξη των μεθόδων ημι-επιβλεπόμενης μάθησης.

3.1 Ιστορική Εξέλιξη Αρχιτεκτονικών στην Ιατρική Κατάτμηση

Η έρευνα στον τομέα της αυτόματης κατάτμησης ιατρικής εικόνας έχει γνωρίσει ραγδαία εξέλιξη την τελευταία δεκαετία, με τις αρχιτεκτονικές να προσαρμόζονται συνεχώς στις αυξανόμενες απαιτήσεις των κλινικών δεδομένων.

3.1.1 Η Εποχή των Συνελικτικών Δικτύων (CNNs)

Η ιδέα των Συνελικτικών Νευρωνικών Δικτύων (CNNs) εισήχθη από τον Kunihiko Fukushima [11] και εξελίχθηκε από τους Waibel [13] και Yann LeCun [14]. Το πραγματικό ορόσημο, ωστόσο, ήρθε το 2012 με το AlexNet [16] και στη συνέχεια με αρχιτεκτονικές όπως το VGGNet [17] και το ResNet [10].

Στην ιατρική εικόνα, η επανάσταση σημειώθηκε το 2015 με την αρχιτεκτονική U-Net [9], η οποία καθιέρωσε το απόλυτο πρότυπο για την 2D ιατρική κατάτμηση, επιλύοντας το πρόβλημα του υπολογιστικού φόρτου των συρόμενων παραθύρων που χρησιμοποιούσαν οι Cireșan et al. [18]. Μεταγενέστερες βελτιώσεις, όπως το Attention U-Net [28], βελτιστοποίησαν περαιτέρω τα δίκτυα αυτά. Καθώς οι τομογραφίες CT και MRI είναι ογκομετρικές, η έρευνα στράφηκε στις 3D αρχιτεκτονικές. Οι Çiçek et al. [4] παρουσίασαν το 3D U-Net, ενώ οι Milletari et al. [29] εισήγαγαν το V-Net χρησιμοποιώντας τη συνάρτηση απώλειας Dice για την αντιμετώπιση της ασυμμετρίας κλάσεων.

3.1.2 Vision Transformers (ViTs) και Υβριδικά Μοντέλα

Για την υπέρβαση των περιορισμών του τοπικού πεδίου αποδοχής των CNNs, οι Dosovitskiy et al. [21] εισήγαγαν τα Vision Transformers (ViTs). Στον χώρο της ιατρικής απεικόνισης, αυτό οδήγησε σε πρωτοποριακά υβριδικά δίκτυα όπως το TransUNet [30] και καθαρές

Transformer αρχιτεκτονικές όπως το Swin-Unet [31]. Η προσαρμογή στο 3D περιβάλλον οδήγησε σε ισχυρά δίκτυα όπως το UNETR [5] και το Swin UNETR [32]. Παρ' όλα αυτά, το πρόβλημα της τετραγωνικής υπολογιστικής πολυπλοκότητας της προσοχής έθεσε σοβαρούς περιορισμούς μνήμης (VRAM), εμποδίζοντας την ευρεία εφαρμογή τους σε εικόνες υψηλής ανάλυσης.

3.1.3 State Space Models (SSMs) και Mamba

Για την επίλυση του απαγορευτικού υπολογιστικού κόστους, τα SSMs αναδείχθηκαν ως η πλέον υποσχόμενη εναλλακτική. Με αφετηρία το μοντέλο S4 [23], η πραγματική καινοτομία ήρθε με την αρχιτεκτονική Mamba από τους Gu και Dao [3]. Ακολούθησαν μοντέλα όπως το Vision Mamba [33] και το VMamba [34].

Στην ιατρική κατάτμηση, η εξέλιξη πέρασε από το U-Mamba [35] στο SwinUMamba [6], το οποίο συνδυάζει την ιεραρχική εξαγωγή χαρακτηριστικών με πολυκατευθυντική σάρωση. Η παρούσα διπλωματική βασίζεται στο SwinUMamba, αναγνωρίζοντας ωστόσο τα κενά της μονοδιάστατης σάρωσης, και εφαρμόζει επιπλέον το ανεξάρτητο βήμα μετα-επεξεργασίας LCC.

3.2 Ημι-επιβλεπόμενη Μάθηση (SSL) στην Ιατρική Εικόνα

Η χρόνια έλλειψη 3D ιατρικών δεδομένων με πυκνό σχολιασμό [2] κατέστησε επιτακτική τη χρήση Ημι-επιβλεπόμενης Μάθησης (Semi-Supervised Learning).

3.2.1 Μάθηση Συνέπειας και Αντιθετική Μάθηση

Οι παραδοσιακές προσεγγίσεις SSL βασίστηκαν στις ψευδο-ετικέτες. Μια μέθοδος-ορόσημο ήταν η αρχιτεκτονική Mean Teacher [26], η οποία σταθεροποίησε την εκπαίδευση χρησιμοποιώντας τον κινητό μέσο όρο των βαρών. Παράλληλα, η Αντιθετική Μάθηση, γνωστή από πλαίσια όπως το SimCLR [36], προσαρμόστηκε για να ενισχύσει τη διακρισιμότητα των ιστών στον λανθάνοντα χώρο.

3.2.2 Διαχείριση Αβεβαιότητας και το Πλαίσιο DyCON

Παρ' όλη την επιτυχία τους, στις ιατρικές εφαρμογές οι μέθοδοι αυτές πάσχουν συχνά από **θόρυβο επιβεβαίωσης (confirmation bias)**: το μοντέλο παράγει λανθασμένες ψευδο-ετικέτες στα ασαφή όρια των οργάνων και εκπαιδεύεται πάνω στα ίδια του τα λάθη. Για την επίλυση αυτού του κενού, προτάθηκε το πλαίσιο DyCON [7], το οποίο αξιολογεί δυναμικά την αβεβαιότητα κάθε πρόβλεψης. Η ικανότητα του DyCON να συνδυάζεται αρμονικά με ισχυρούς κορμούς όπως το SwinUMamba και μηχανισμούς όπως το LCC αποτελεί τον πυρήνα της παρούσας ερευνητικής προσέγγισης.

Μέρος II

Πρακτικό Μέρος

Κεφάλαιο 4

Σύνολα Δεδομένων και Υλοποίηση

Στο παρόν κεφάλαιο παρατίθενται οι λεπτομέρειες της πειραματικής διάταξης που χρησιμοποιήθηκε για την αξιολόγηση της προτεινόμενης αρχιτεκτονικής (SwinUMamba + DyCON + LCC). Αναλύονται το σύνολο δεδομένων, ο τρόπος διαχωρισμού του, οι υπολογιστικοί πόροι, καθώς και οι τεχνικές προδιαγραφές της εκπαίδευσης του μοντέλου.

4.1 Σύνολο Δεδομένων Pancreas-CT

Για την αξιολόγηση της απόδοσης του προτεινόμενου συστήματος χρησιμοποιήθηκε το ευρέως αναγνωρισμένο σύνολο δεδομένων Pancreas-CT [8]. Το συγκεκριμένο dataset δημιουργήθηκε από το Κλινικό Κέντρο των Εθνικών Ινστιτούτων Υγείας των ΗΠΑ (NIH) και φιλοξενείται στην πλατφόρμα The Cancer Imaging Archive (TCIA) [37].

4.1.1 Δημογραφικά και Κλινικά Χαρακτηριστικά

Το σύνολο δεδομένων περιλαμβάνει 82 τρισδιάστατες (3D) αξονικές τομογραφίες (CT) κοιλιακής χώρας με χορήγηση σκιαγραφικού, οι οποίες ελήφθησαν κατά την πυλαία φλεβική φάση (περίπου 70 δευτερόλεπτα μετά την ενδοφλέβια έγχυση). Η ομάδα των εξεταζόμενων αποτελείται από 53 άνδρες και 27 γυναίκες, ηλικίας 18 έως 76 ετών (με μέση ηλικία τα 46.8 ± 16.7 έτη).

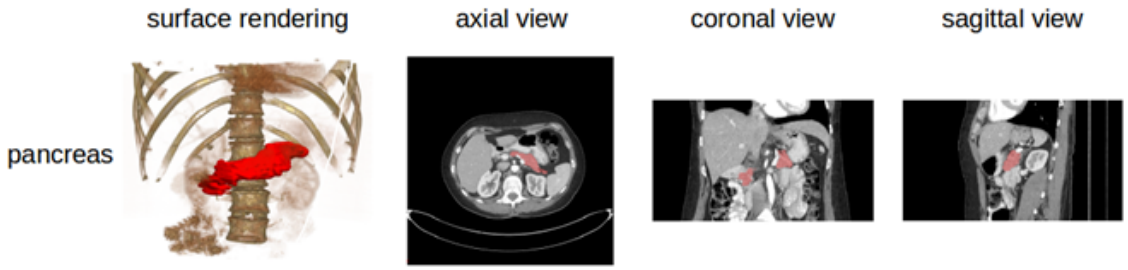
Οι ασθενείς επιλέχθηκαν με αυστηρά κριτήρια: 17 από αυτούς ήταν υγιείς δότες νεφρού που υποβλήθηκαν σε τομογραφία πριν από τη νεφρεκτομή, ενώ οι υπόλοιποι 65 επιλέχθηκαν από έμπειρο ακτινολόγο ώστε να μην παρουσιάζουν σοβαρές κοιλιακές παθολογίες ή καρκινικές βλάβες στο πάγκρεας. Η επιλογή μη παθολογικών παγκρεάτων βοηθά στη μοντελοποίηση της υγιούς ανατομίας του οργάνου.

4.1.2 Τεχνικές Προδιαγραφές και Ground-Truth

Οι αξονικές τομογραφίες λήφθηκαν με τη χρήση τομογράφων των εταιρειών Philips και Siemens (με τάση σωλήνα στα 120 kVp). Κάθε τομή έχει ανάλυση 512×512 εικονοστοιχείων (pixels) με ποικίλο μέγεθος εικονοστοιχείου στο δισδιάστατο επίπεδο, ενώ το πάχος της τομής (slice thickness) κυμαίνεται μεταξύ 1,5 και 2,5 χιλιοστών (mm).

Για την εκπαίδευση και την ποσοτική αξιολόγηση των αλγορίθμων βαθιάς μάθησης, η ακρίβεια των ετικετών αναφοράς (ground-truth) είναι καθοριστική. Στο σύνολο Pancreas-

CT, η κατάτμηση του παγκρέατος πραγματοποιήθηκε αρχικά χειροκίνητα, τομή προς τομή, από φοιτητή ιατρικής. Στη συνέχεια, όλες οι μάσκες κατάτμησης επαληθεύτηκαν και τροποποιήθηκαν ενδελεχώς από έναν εξειδικευμένο ακτινολόγο, εξασφαλίζοντας υψηλή κλινική πιστότητα. Οπτικοποίηση της ανατομίας του παγκρέατος και του αντίστοιχου ground-truth σε πολλαπλά επίπεδα απεικόνισης φαίνεται στο Σχήμα 4.1.



Σχήμα 4.1: *Απεικόνιση του παγκρέατος στο σύνολο δεδομένων Pancreas-CT. Από αριστερά προς τα δεξιά: Τρισδιάστατη αναπαράσταση επιφάνειας (surface rendering) και δισδιάστατες προβολές στο αξονικό (axial), στεφανιαίο (coronal) και οβελιαίο (sagittal) επίπεδο.*

4.1.3 Διαχωρισμός Δεδομένων (Data Split)

Για τον σχεδιασμό των πειραμάτων, οι 82 διαθέσιμες τομογραφίες διαχωρίστηκαν σε ένα σύνολο εκπαίδευσης (training set) αποτελούμενο από 62 εικόνες και ένα σύνολο δοκιμής (test set) 20 εικόνων.

Για την αξιολόγηση του δικτύου στο πλαίσιο της Ημι-επιβλεπόμενης Μάθησης (Semi-Supervised Learning), από το σύνολο εκπαίδευσης χρησιμοποιήθηκαν μόνο 12 τομογραφίες με τις αντίστοιχες ετικέτες τους (labeled data), ενώ οι υπόλοιπες 50 τομογραφίες αντιμετωπίστηκαν ως μη-σχολιασμένες (unlabeled data). Αυτές τροφοδότησαν το πλαίσιο DyCON για την αξιοποίηση της πληροφορίας μέσω μάθησης συνέπειας και αντιθετικής μάθησης. Αντίθετα, στο σενάριο της Πλήρως Επιβλεπόμενης Μάθησης (Fully Supervised), το μοντέλο εκπαιδεύτηκε αξιοποιώντας και τις 62 διαθέσιμες ετικέτες.

4.2 Υπολογιστικοί Πόροι και Λεπτομέρειες Υλοποίησης

Η εκπαίδευση τρισδιάστατων (3D) αρχιτεκτονικών που ενσωματώνουν Μοντέλα Χώρου Καταστάσεων (SSMs), όπως το Swin-UMamba, είναι εξαιρετικά απαιτητική σε υπολογιστικούς πόρους και μνήμη γραφικών (VRAM). Λόγω των περιορισμών του υλικού (hardware limitations), τα πειράματα της παρούσας εργασίας υλοποιήθηκαν στο υπολογιστικό νέφος της πλατφόρμας Kaggle, το οποίο παρέχει δωρεάν αλλά αυστηρά περιορισμένη πρόσβαση σε μονάδες επεξεργασίας γραφικών (GPUs).

4.2.1 Αντιμετώπιση Περιορισμών Μνήμης (OOM Mitigations)

Για την αποφυγή σφαλμάτων υπερχείλισης μνήμης (Out-Of-Memory — OOM) κατά την εκπαίδευση, εφαρμόστηκε μια σειρά από αυστηρές στρατηγικές βελτιστοποίησης τόσο σε επίπεδο δεδομένων όσο και σε επίπεδο πλαισίου (framework):

1. **Υποβάθμιση Χωρικής Ανάλυσης (Patch Resizing):** Αντί της επεξεργασίας ολόκληρου του ογκομετρικού πλέγματος σε πλήρη ανάλυση, τα δεδομένα εισόδου υποβαθμίστηκαν και τεμαχίστηκαν σε τρισδιάστατα τμήματα (3D patches) διαστάσεων 64×64×64 ογκοστοιχείων. Αυτό μείωσε δραματικά το υπολογιστικό φορτίο των Mamba blocks, διατηρώντας ωστόσο επαρκή χωρική πληροφορία για την αναγνώριση του παγκρέατος.

2. **Συσσώρευση Κλίσης (Gradient Accumulation):** Λόγω του περιορισμού της μνήμης, το μέγεθος δέσμης (batch size) ορίστηκε αναγκαστικά στην πολύ μικρή τιμή του 2 (με 1 labeled και 1 unlabeled δείγμα ανά επανάληψη στο DyCON). Για να διασφαλιστεί η σταθερότητα της εκπαίδευσης, εφαρμόστηκε συσσώρευση κλίσης (accumulation steps) ίση με 4. Αυτό προσομοιώνει ένα ενεργό batch size 8 δειγμάτων πριν την ενημέρωση των βαρών, ομαλοποιώντας τα gradients.

3. **Διαχείριση Μνήμης PyTorch:** Στο επίπεδο του λειτουργικού συστήματος, ενεργοποιήθηκε η παράμετρος PYTORCH_ALLOC_CONF=expandable_segments:True, η οποία αποτρέπει τον κατακερματισμό της μνήμης (memory fragmentation) της GPU. Παράλληλα, γινόταν περιοδικός καθαρισμός της προσωρινής μνήμης (torch.cuda.empty_cache()) και διαγραφή παλαιών σημείων ελέγχου (checkpoints) από τον δίσκο του Kaggle για την εξοικονόμηση αποθηκευτικού χώρου.

4.2.2 Υπερπαράμετροι και Προ-εκπαίδευση (Pre-training)

Για την ταχύτερη σύγκλιση του μοντέλου και τη βελτίωση της γενίκευσης, ο κωδικοποιητής του Swin-UMamba αρχικοποιήθηκε χρησιμοποιώντας προ-εκπαιδευμένα βάρη (pre-trained weights). Συγκεκριμένα, αξιοποιήθηκαν τα βάρη της αρχιτεκτονικής VMamba [34], τα οποία είχαν εκπαιδευτεί στο εκτεταμένο σύνολο δεδομένων ImageNet-1K [38]. Η προσέγγιση αυτή είναι ιδιαίτερα κρίσιμη στο σενάριο της ημι-επιβλεπόμενης μάθησης, όπου τα επισημασμένα δεδομένα είναι ελάχιστα.

Η βελτιστοποίηση του δικτύου πραγματοποιήθηκε χρησιμοποιώντας τον αλγόριθμο AdamW με αρχικό ρυθμό μάθησης (base learning rate) 10^{-4}. Η εκπαίδευση διήρκεσε για μέγιστο αριθμό 20.000 επαναλήψεων (iterations). Σχετικά με το ημι-επιβλεπόμενο πλαίσιο DyCON, οι υπερπαράμετροι ορίστηκαν ως εξής:

- **Βάρος Μη-Επιβλεπόμενης Απώλειας (u_weight):** 0.5, καθορίζοντας τη συνεισφορά των unlabeled δεδομένων στο συνολικό σφάλμα.

- **Συντελεστής Συνέπειας (consistency):** 0.1, για τη διατήρηση σταθερών προβλέψεων μεταξύ των διαφορετικών επαυξήσεων (augmentations).

- **Θερμοκρασία (Temperature - temp):** 0.6, για τον έλεγχο της οξύτητας (sharpness) της κατανομής των ψευδο-ετικετών (pseudo-labels).

Κεφάλαιο 5

Υλοποίηση

Στο κεφάλαιο αυτό περιγράφεται η υλοποίηση του συστήματος, με βάση τη μελέτη που παρουσιάστηκε στο προηγούμενο κεφάλαιο. Αρχικά παρουσιάζεται η πλατφόρμα και τα προγραμματιστικά εργαλεία που χρησιμοποιήθηκαν. Στη συνέχεια δίνονται οι λεπτομέρειες υλοποίησης για τους βασικούς αλγορίθμους του συστήματος καθώς και η δομή του κώδικα.

5.1 Λεπτομέρειες Υλοποίησης

5.1.1 Περιβάλλον Ανάπτυξης και Βιβλιοθήκες Λογισμικού

Η υλοποίηση του προτεινόμενου συστήματος πραγματοποιήθηκε στη γλώσσα προγραμματισμού Python (έκδοση 3.10), η οποία αποτελεί το καθιερωμένο πρότυπο στην ανάπτυξη εφαρμογών Τεχνητής Νοημοσύνης και Μηχανικής Μάθησης. Ως κεντρικό πλαίσιο βαθιάς μάθησης επιλέχθηκε η βιβλιοθήκη PyTorch [39], η οποία προσφέρει εξαιρετική ευελιξία μέσω της δυναμικής διαχείρισης του γραφήματος υπολογισμών (dynamic computational graph) και της εγγενούς υποστήριξης επιτάχυνσης υλικού μέσω της πλατφόρμας NVIDIA CUDA.

Για τη διαχείριση, τη φόρτωση και την επεξεργασία των τρισδιάστατων ιατρικών δεδομένων ογκομετρικής απεικόνισης μορφότυπου NIfTI (.nii.gz) αξιοποιήθηκαν οι εξειδικευμένες βιβλιοθήκες SimpleITK και Nibabel. Οι βιβλιοθήκες αυτές επέτρεψαν την ομαλή ανάγνωση των πλεγμάτων και των μεταδεδομένων των τομογραφιών, καθώς και την εκτέλεση χωρικών μετασχηματισμών. Επιπλέον, χρησιμοποιήθηκε η βιβλιοθήκη MONAI (Medical Open Network for AI) [40] για την υλοποίηση προηγμένων τεχνικών επαύξησης δεδομένων (data augmentation), όπως τυχαίες περιστροφές, προσθήκη θορύβου και ελαστικές παραμορφώσεις, οι οποίες είναι κρίσιμες για τη σταθερότητα του ημι-επιβλεπόμενου πλαισίου.

5.1.2 Δομή Κώδικα και Οργάνωση Συστήματος

Η αρχιτεκτονική του κώδικα οργανώθηκε με τρόπο αρθρωτό (modular), διαχωρίζοντας σαφώς τη διαχείριση των δεδομένων, τον ορισμό των νευρωνικών μοντέλων, τη διαδικασία της εκπαίδευσης και την αξιολόγηση. Το κεντρικό σημείο εισόδου της εκπαίδευσης είναι το σενάριο train_DyCON_Pancreas.py, το οποίο αναλαμβάνει τον συντονισμό ολόκληρης της ροής εκπαίδευσης, ενώ η τελική αποτίμηση των μοντέλων πραγματοποιείται από ξεχωριστό

σενάριο αξιολόγησης.

Η δομή των φακέλων του έργου διαμορφώθηκε ως εξής:

- **code/models/:** Περιλαμβάνει την υλοποίηση των δικτύων που συγκρίνονται — του κορμού Swin-UMamba με τα επιλεκτικά blocks σάρωσης του Mamba, καθώς και του συνελικτικού U-Net3D.
- **code/dataloaders/:** Περιέχει τους προσαρμοσμένους φορτωτές δεδομένων (custom DataLoaders), οι οποίοι σχεδιάστηκαν για να υποστηρίζουν τη διπλή ροή δειγμάτων που απαιτεί το πλαίσιο DyCON.
- **code/losses/:** Περιλαμβάνει τις συναρτήσεις απώλειας, όπως τη συνεπή απώλεια με επίγνωση αβεβαιότητας (UnCL) και την αντιθετική απώλεια focal (FeCL).
- **code/utils/:** Περιλαμβάνει βοηθητικές ρουτίνες, μεταξύ των οποίων τη διαδικασία κατάτμησης ογκομετρικών εικόνων με ολίσθηση παραθύρου (sliding-window inference) κατά την αξιολόγηση.

Σημειώνεται ότι η μετα-επεξεργασία Largest Connected Component (LCC) δεν αποτελεί μέρος της αρχιτεκτονικής των δικτύων, αλλά υλοποιείται ως αυτοτελής συνάρτηση στο σενάριο αξιολόγησης. Η συνάρτηση αυτή λαμβάνει τη δυαδική πρόβλεψη του δικτύου, εντοπίζει τις συνεκτικές της συνιστώσες (connected components) στον 3D χώρο (μέσω της skimage.measure.label) και διατηρεί αποκλειστικά εκείνη με τον μεγαλύτερο όγκο, απομακρύνοντας τις ψευδώς-θετικές νησίδες πριν τον υπολογισμό των μετρικών Dice, HD95 και ASD.

5.1.3 Ροή Εκπαίδευσης και Αλγοριθμική Διαχείριση

Η εκπαίδευση του συστήματος οργανώθηκε γύρω από έναν επαναληπτικό βρόχο (training loop) μέγιστης διάρκειας 20.000 επαναλήψεων (iterations). Σε κάθε επανάληψη, ο φορτωτής δεδομένων τροφοδοτεί ταυτόχρονα το δίκτυο με μία δέσμη επισημασμένων δεδομένων (labeled batch size ίσο με 1) και μία δέσμη μη-επισημασμένων δεδομένων (unlabeled batch size ίσο με 1), διαμορφώνοντας ένα συνολικό batch size ίσο με 2.

Η ροή εκτέλεσης ανά επανάληψη ακολουθεί τα εξής βήματα:

1. **Επιβλεπόμενη Φάση:** Το επισημασμένο δείγμα εισάγεται στο δίκτυο του μαθητή (student network). Υπολογίζεται η επιβλεπόμενη απώλεια $\mathcal{L}_{\text{sup}}$ μέσω του συνδυασμού της διασταυρούμενης εντροπίας (Cross-Entropy) και της απώλειας Dice, προκειμένου να αντιμετωπιστεί η έντονη ασυμμετρία κλάσεων του παγκρέατος.
2. **Μη-Επιβλεπόμενη Φάση (Παραγωγή Ψευδο-ετικετών):** Το μη-επισημασμένο δείγμα υφίσταται ασθενή επαύξηση (weak augmentation) και εισάγεται στο δίκτυο του δασκάλου (teacher network). Οι παραγόμενες πιθανότητες χρησιμοποιούνται για τον υπολογισμό της εντροπίας Shannon σε επίπεδο ογκοστοιχείου, η οποία αποτιμά την αβεβαιότητα της πρόβλεψης. Οι τιμές αυτές μετατρέπονται σε σκληρές ψευδο-ετικέτες (hard pseudo-labels).

3. **Μη-Επιβλεπόμενη Φάση (Μάθηση Συνέπειας):** Το ίδιο μη-επισημασμένο δείγμα υφίσταται ισχυρά επαύξηση (strong augmentation) και εισάγεται στο δίκτυο του μαθητή. Υπολογίζεται η απώλεια συνέπειας UnCL σε συνδυασμό με την απώλεια του δασκάλου (`--use_teacher_loss 1`), σταθμισμένη δυναμικά από τον βαθμό αβεβαιότητας.

4. **Αντιθετική Φάση:** Εφαρμόζεται η απώλεια FeCL στον λανθάνοντα χώρο, ωθώντας τα χαρακτηριστικά των ογκοστοιχείων που ανήκουν στο πάγκρεας μακριά από τα χαρακτηριστικά του φόντου.

5.1.4 Μηχανική Βελτιστοποίησης και Πόρων Μνήμης

Λόγω των περιορισμών της GPU στην πλατφόρμα Kaggle, η συσσώρευση κλίσης (Gradient Accumulation) ρυθμίστηκε σε 4 βήματα (`--accumulation_steps 4`). Αυτό σημαίνει ότι οι κλίσεις προστίθενται για 4 διαδοχικές επαναλήψεις και η ενημέρωση των βαρών μέσω του βελτιστοποιητή AdamW [41] πραγματοποιείται κάθε 4 βήματα, προσομοιώνοντας έτσι ένα σταθερό ενεργό μέγεθος δέσμης (effective batch size) 8 δειγμάτων, γεγονός που εξασφαλίζει τη σύγκλιση και τη σταθερότητα των gradients.

Τέλος, επειδή ο διαθέσιμος αποθηκευτικός χώρος του Kaggle είναι περιορισμένος και τα αρχεία των βαρών (.pth) των τρισδιάστατων αρχιτεκτονικών είναι μεγάλα, υλοποιήθηκε μια παρασκηνιακή διεργασία (background process). Αυτή η διεργασία εκτελείται αυτόματα ανά 30 λεπτά, εντοπίζει τα παλαιότερα ενδιάμεσα σημεία ελέγχου (checkpoints) εκπαίδευσης και τα διαγράφει αυτόματα, αποτρέποντας την υπερχείλιση του δίσκου και την απότομη διακοπή της πειραματικής διαδικασίας.

Κεφάλαιο 6

Παρουσίαση και Σχολιασμός Πειραματικών Αποτελεσμάτων

Στο κεφάλαιο αυτό παρουσιάζεται ο έλεγχος καλής λειτουργίας του προτεινόμενου συστήματος και η αξιολόγηση της απόδοσής του. Μέσα από την ποσοτική ανάλυση των πειραματικών αποτελεσμάτων, αποτιμάται η συμβολή του ημι-επιβλεπόμενου πλαισίου DyCON και της μετα-επεξεργασίας LCC (Largest Connected Component), καθώς και η συγκριτική συμπεριφορά των αρχιτεκτονικών Swin-UMamba και U-Net3D υπό περιορισμένη χωρική ανάλυση.

6.1 Μεθοδολογία Ελέγχου

Η αξιολόγηση των μοντέλων πραγματοποιήθηκε με βάση τη σύγκριση των παραγόμενων χαρτών κατάτμησης (segmentation masks) με τους σχολιασμούς αναφοράς (ground-truth) των ειδικών ιατρών στο σύνολο δεδομένων Pancreas-CT.

Για την ολοκληρωμένη αποτίμηση της απόδοσης, η μεθοδολογία ελέγχου βασίστηκε σε τρεις κρίσιμες μετρικές:

1. **Συντελεστής Dice (Dice Similarity Coefficient — DSC):** Μετρά τη συνολική ομοιότητα και επικάλυψη των όγκων. Υψηλότερες τιμές (↑) υποδηλώνουν καλύτερη απόδοση.

2. **Απόσταση Hausdorff 95% (HD95):** Αξιολογεί την απόσταση μεταξύ των επιφανειών/ορίων των προβλέψεων και του ground-truth, αγνοώντας το 5% των πιο ακραίων σφαλμάτων (outliers). Χαμηλότερες τιμές (↓) υποδηλώνουν ακριβέστερα ανατομικά όρια.

3. **Μέση Επιφανειακή Απόσταση (Average Surface Distance — ASD):** Μετρά τη μέση απόσταση όλων των σημείων της επιφάνειας της πρόβλεψης από την πραγματική επιφάνεια. Χαμηλότερες τιμές (↓) υποδεικνύουν υψηλή ακρίβεια οριοθέτησης.

Τα πειράματα δομήθηκαν σε δύο άξονες: αφενός στη σύγκριση αρχιτεκτονικών (Swin-UMamba έναντι U-Net3D), και αφετέρου στη σύγκριση σεναρίων εκπαίδευσης (Ημι-επιβλεπόμενη με DyCON χρησιμοποιώντας 12 labels έναντι Πλήρως Επιβλεπόμενης χρησιμοποιώντας 62

labels). Σε όλες τις διατάξεις, η μετα-επεξεργασία LCC εφαρμόζεται κατά τη φάση της αξιολόγησης με τον ίδιο ακριβώς τρόπο και στις δύο αρχιτεκτονικές, διατηρώντας μόνο τη μεγαλύτερη συνεκτική συνιστώσα κάθε πρόβλεψης, ώστε η συνεισφορά της να αποτιμηθεί απομονωμένα.

6.2 Αναλυτική Παρουσίαση Ελέγχου

Στους πίνακες που ακολουθούν παρουσιάζονται τα συγκεντρωτικά αποτελέσματα της κατάτμησης του παγκρέατος. Ο Πίνακας 6.1 αποτυπώνει τις ωμές προβλέψεις των μοντέλων (RAW) πριν την εφαρμογή του LCC, ενώ ο Πίνακας 6.2 παρουσιάζει την τελική απόδοση μετά τη μετα-επεξεργασία.

Πίνακας 6.1: *Συγκεντρωτικά Αποτελέσματα Κατάτμησης Παγκρέατος (Ωμές Προβλέψεις — RAW / Πριν το LCC)*

Αρχιτεκτονική	**Μέθοδος (Labels)**	**Dice (%) ↑**	**HD95 ↓**	**ASD ↓**
Ημι-Επιβλεπόμενη Μάθηση (Semi-Supervised — 12 Labels)				
Swin-UMamba	DyCON	46.08 ± 11.97	74.00 ± 18.72	32.39 ± 10.37
U-Net3D	**DyCON**	**63.60 ± 14.57**	**63.99 ± 14.25**	**21.55 ± 6.39**
Πλήρως Επιβλεπόμενη Μάθηση (Fully Supervised — 62 Labels)				
Swin-UMamba	Fully Supervised	55.56 ± 9.22	68.20 ± 18.29	25.80 ± 8.43
U-Net3D	**Fully Supervised**	**66.90 ± 7.87**	**54.16 ± 19.11**	**17.45 ± 6.32**

Πίνακας 6.2: *Συγκεντρωτικά Αποτελέσματα Κατάτμησης Παγκρέατος (Τελικά Αποτελέσματα μετά από LCC)*

Αρχιτεκτονική	**Μέθοδος (Labels)**	**Dice (%) ↑**	**HD95 ↓**	**ASD ↓**
Ημι-Επιβλεπόμενη Μάθηση (Semi-Supervised — 12 Labels)				
Swin-UMamba	DyCON	57.60 ± 24.09	36.41 ± 43.98	12.85 ± 27.09
U-Net3D	**DyCON**	**71.09 ± 17.63**	**20.51 ± 23.82**	**2.22 ± 1.44**
Πλήρως Επιβλεπόμενη Μάθηση (Fully Supervised — 62 Labels)				
Swin-UMamba	Fully Supervised	63.65 ± 11.69	27.61 ± 18.96	2.83 ± 2.52
U-Net3D	**Fully Supervised**	**71.89 ± 11.09**	**24.60 ± 26.52**	**1.78 ± 0.77**

6.2.1 Μελέτη Αποδόμησης (Ablation Study)

Ο Πίνακας 6.3 παρουσιάζει συστηματικά τη συνεισφορά κάθε συνιστώσας. Από την ανάλυση προκύπτουν τρία βασικά συμπεράσματα.

Αδυναμία της μονοδιάστατης σάρωσης υπό περιορισμένη ανάλυση. Στις ωμές προβλέψεις, το Swin-UMamba με DyCON υστερεί σημαντικά ακόμη και έναντι της βασικής γραμμής πλήρους επίβλεψης (46.08% έναντι 55.56%). Αυτό οφείλεται στη χωρική παραμόρφωση που εισάγει η μονοδιάστατη σάρωση του μηχανισμού Mamba σε τμήματα μικρής ανάλυσης ($64 \times 64 \times 64$), η οποία υποβαθμίζει την ποιότητα των χαρτών χαρακτηριστικών σε ένα δίκτυο που βασίζεται στο παγκόσμιο πλαίσιο. Επιπρόσθετα, η αστοχία αυτή πιθανώς επιτείνεται από μια ασυμβατότητα πεδίου (domain mismatch): ο κωδικοποιητής του

Πίνακας 6.3: *Μελέτη Αποδόμησης (Ablation Study): Συνεισφορά κάθε συνιστώσας στην απόδοση κατάτμησης παγκρέατος. Το σύμβολο ✓ υποδηλώνει ενεργοποίηση του αντίστοιχου μηχανισμού.*

Αρχιτεκτονική	**DyCON**	**LCC**	**Labels**	**Dice (%) ↑**	**HD95 ↓**	**ASD ↓**
Swin-UMamba						
Swin-UMamba	×	×	62	55.56 ± 9.22	68.20 ± 18.29	25.80 ± 8.43
Swin-UMamba	✓	×	12	46.08 ± 11.97	74.00 ± 18.72	32.39 ± 10.37
Swin-UMamba	✓	✓	12	57.60 ± 24.09	36.41 ± 43.98	12.85 ± 27.09
Swin-UMamba	×	✓	62	63.65 ± 11.69	27.61 ± 18.96	2.83 ± 2.52
U-Net3D						
U-Net3D	×	×	62	66.90 ± 7.87	54.16 ± 19.11	17.45 ± 6.32
U-Net3D	✓	×	12	63.60 ± 14.57	63.99 ± 14.25	21.55 ± 6.39
U-Net3D	✓	✓	12	**71.09 ± 17.63**	**20.51 ± 23.82**	**2.22 ± 1.44**
U-Net3D	×	✓	62	71.89 ± 11.09	24.60 ± 26.52	1.78 ± 0.77

Swin-UMamba αρχικοποιείται με βάρη VMamba προεκπαιδευμένα σε δισδιάστατες φυσικές εικόνες (ImageNet-1K), ενώ καλείται να επεξεργαστεί τρισδιάστατα ιατρικά δεδομένα· η μεταφορά αυτή των αναπαραστάσεων από το 2D στο 3D πεδίο είναι λιγότερο άμεση από ό,τι στα αμιγώς συνελικτικά δίκτυα, τα οποία εκπαιδεύονται από μηδενική βάση στα ίδια τα ογκομετρικά τμήματα.

Επίδραση της μετα-επεξεργασίας LCC. Η εφαρμογή του Largest Connected Component βελτιώνει αισθητά όλες τις μετρικές και στις δύο αρχιτεκτονικές, με ιδιαίτερα έντονη επίδραση στις μετρικές ορίων: για το Swin-UMamba (ημι-επιβλεπόμενο) το HD95 υποδιπλασιάζεται (από 74.00 σε 36.41) και το Dice αυξάνεται κατά 11.52 ποσοστιαίες μονάδες, ενώ για το U-Net3D το ASD καταρρέει από 21.55 σε 2.22. Το εύρημα αυτό είναι αναμενόμενο: το LCC αφαιρεί τις απομακρυσμένες ψευδώς-θετικές συνεκτικές συνιστώσες, στις οποίες οφείλονται κατά κύριο λόγο οι μεγάλες αποστάσεις HD95 και ASD. Τονίζεται ότι το LCC είναι κλασικό βήμα μετα-επεξεργασίας ανεξάρτητο της αρχιτεκτονικής, και όχι μαθησιακή συνιστώσα· γι' αυτόν τον λόγο ωφελεί εξίσου το συνελικτικό U-Net3D.

Αποδοτικότητα της διάταξης DyCON + LCC. Το σημαντικότερο εύρημα αφορά το U-Net3D με DyCON + LCC: χρησιμοποιώντας μόλις 12 σχολιασμένες τομογραφίες, επιτυγχάνει Dice 71.09%, εξαιρετικά κοντά στο 71.89% της πλήρως επιβλεπόμενης εκπαίδευσης με 62 ετικέτες — δηλαδή με πενταπλάσιο αριθμό σχολιασμένων δεδομένων.

Σταθερότητα και συνολική αποτίμηση. Πέραν των μέσων τιμών, η ανάλυση της διασποράς αποκαλύπτει ουσιώδεις διαφορές αξιοπιστίας. Το Swin-UMamba, ακόμη και μετά το LCC, παρουσιάζει πολύ υψηλή τυπική απόκλιση στο ημι-επιβλεπόμενο σενάριο (HD95 36.41 ± 43.98, ASD 12.85 ± 27.09). Αυτό υποδηλώνει ότι σε ορισμένες τομογραφίες το LCC αποτυγχάνει: όταν η μεγαλύτερη συνεκτική συνιστώσα της πρόβλεψης δεν αντιστοιχεί στο πάγκρεας αλλά σε κάποια εκτενή εσφαλμένη δομή, το βήμα διατηρεί τη λάθος περιοχή και απορρίπτει το όργανο-στόχο. Αντιθέτως, το U-Net3D με DyCON και LCC επιτυγχάνει όχι μόνο την υψηλότερη μέση ακρίβεια αλλά και σαφώς σταθερότερα αποτελέσματα (ASD 2.22 ± 1.44). Συνολικά, υπό το συγκεκριμένο καθεστώς περιορισμένης χωρικής ανάλυσης, ο συνδυασμός U-Net3D + DyCON + LCC αναδεικνύεται ως η πλέον εύρωστη επιλογή.

6.2.2 Σχολιασμός των Ωμών Προβλέψεων (RAW Αποτελέσματα)

Εξετάζοντας τον Πίνακα 6.1, καθίσταται σαφές ότι στις ωμές προβλέψεις (πριν το LCC), το αμιγώς συνελικτικό U-Net3D υπερτερεί σημαντικά του Swin-UMamba σε όλες τις μετρικές. Στο σενάριο ημι-επιβλεπόμενης μάθησης, το U-Net3D επιτυγχάνει Dice 63.60%, ενώ το Swin-UMamba περιορίζεται στο 46.08%.

Αυτή η υστέρηση του Swin-UMamba οφείλεται πιθανότατα όχι σε εγγενή αδυναμία της αρχιτεκτονικής, αλλά στους περιορισμούς του υλικού. Όπως αναφέρθηκε, λόγω των περιορισμών μνήμης (OOM) της πλατφόρμας Kaggle, το μέγεθος των τρισδιάστατων τμημάτων (3D patches) εισόδου μειώθηκε δραστικά στην ανάλυση $64 \times 64 \times 64$.

Σε τόσο μικρά ογκομετρικά τμήματα, τα συνελικτικά δίκτυα όπως το U-Net3D πλεονεκτούν εγγενώς: το μικρό και εστιασμένο τοπικό πεδίο αποδοχής τους (local receptive field) τους επιτρέπει να εξάγουν χρήσιμα γεωμετρικά χαρακτηριστικά ακόμη και από μικρά τμήματα της εικόνας. Αντίθετα, οι αρχιτεκτονικές που βασίζονται στην παγκόσμια προσοχή και τις μακροπρόθεσμες εξαρτήσεις, όπως το Swin-UMamba, απαιτούν μεγαλύτερο οπτικό πεδίο (global context) για να αποδώσουν τα μέγιστα. Σε συνδυασμό με την απώλεια χωρικής πληροφορίας που προκαλεί η μονοδιάστατη σάρωση (1D scanning) του μηχανισμού Mamba, η απόδοση του δικτύου στα μικρά patches μειώνεται αισθητά, κάτι που αντικατοπτρίζεται στο υψηλό σφάλμα HD95 (74.00).

6.2.3 Επίδραση της Μετα-επεξεργασίας LCC

Η εικόνα αλλάζει αισθητά με την εφαρμογή του Largest Connected Component (LCC), όπως καταγράφεται στον Πίνακα 6.2. Επειδή οι μετρικές HD95 και ASD κυριαρχούνται από τα πιο απομακρυσμένα εσφαλμένα ογκοστοιχεία, η αφαίρεση των διάσπαρτων ψευδώς-θετικών συνιστωσών έχει δυσανάλογα μεγάλη θετική επίδραση.

Για το Swin-UMamba στο ημι-επιβλεπόμενο σενάριο, το Dice αυξάνεται κατά 11.52 ποσοστιαίες μονάδες (από 46.08% σε 57.60%) και το σφάλμα ορίων HD95 υποδιπλασιάζεται (από 74.00 σε 36.41). Ανάλογη βελτίωση παρατηρείται και στο U-Net3D, όπου το ASD μειώνεται από 21.55 σε 2.22. Δεδομένου ότι το LCC εφαρμόζεται με τον ίδιο τρόπο και ανεξάρτητα από την αρχιτεκτονική, τα αποτελέσματα επιβεβαιώνουν ότι πρόκειται για ένα αποδοτικό, γενικής χρήσης βήμα εξυγίανσης των προβλέψεων και όχι για μηχανισμό ειδικό προς κάποια αρχιτεκτονική. Όπως σημειώθηκε όμως, η αποτελεσματικότητά του εξαρτάται από την υπόθεση ότι η μεγαλύτερη συνεκτική συνιστώσα αντιστοιχεί στο όργανο-στόχο — υπόθεση που, όταν παραβιάζεται, οδηγεί στην υψηλή διασπορά που παρατηρήθηκε για το Swin-UMamba.

6.2.4 Αποδοτικότητα του Πλαισίου DyCON

Ένα από τα πιο σημαντικά ευρήματα αφορά τη δυναμική του πλαισίου DyCON. Συγκρίνοντας τις προβλέψεις της ημι-επιβλεπόμενης μάθησης (με μόλις 12 labels) με αυτές της πλήρως επιβλεπόμενης μάθησης (με όλα τα 62 labels), παρατηρούμε ότι το DyCON καταφέρνει να γεφυρώσει το χάσμα.

Συγκεκριμένα, το U-Net3D με DyCON (μετά το LCC) επιτυγχάνει Dice 71.09%, το οποίο

είναι εξαιρετικά κοντά στο 71.89% που επιτυγχάνει το ίδιο δίκτυο έχοντας πρόσβαση στο 100% των σχολιασμένων δεδομένων. Το γεγονός αυτό πιστοποιεί την ικανότητα του DyCON να φιλτράρει δυναμικά την αβεβαιότητα κατά την εκπαίδευση, επιτρέποντας στο σύστημα να εξάγει πολύτιμη γνώση από τις μη-σχολιασμένες τομογραφίες χωρίς να παρασύρεται από θόρυβο ή λανθασμένες ψευδο-ετικέτες.

Μέρος III

Επίλογος

Κεφάλαιο 7

Επίλογος

7.1 Συμπεράσματα

Η παρούσα διπλωματική εργασία επικεντρώθηκε στην αντιμετώπιση δύο εκ των σημαντικότερων προκλήσεων στην ιατρική ανάλυση εικόνας: την ακριβή κατάτμηση οργάνων με ασαφή ανατομικά όρια, όπως το πάγκρεας, και την έλλειψη μεγάλου όγκου σχολιασμένων δεδομένων. Για την επίτευξη αυτού του στόχου, αναπτύχθηκε και αξιολογήθηκε ένα σύστημα βαθιάς μάθησης που συνδυάζει την αρχιτεκτονική Swin-UMamba, το ημι-επιβλεπόμενο πλαίσιο DyCON και τη μετα-επεξεργασία Largest Connected Component (LCC), ενώ ως βάση σύγκρισης χρησιμοποιήθηκε το καθιερωμένο συνελικτικό δίκτυο U-Net3D.

Μέσα από την ενδελεχή πειραματική διαδικασία και την ποσοτική ανάλυση των αποτελεσμάτων στο σύνολο δεδομένων Pancreas-CT, εξάγονται τα ακόλουθα βασικά συμπεράσματα:

- **Η Αδυναμία της Μονοδιάστατης Σάρωσης σε Μικρά Τμήματα:** Επιβεβαιώθηκε ότι τα Μοντέλα Χώρου Καταστάσεων (SSMs), παρά τη θεωρητική τους υπεροχή στη μοντελοποίηση μακροπρόθεσμων εξαρτήσεων με γραμμική πολυπλοκότητα, υστερούν όταν εφαρμόζονται σε περιορισμένα χωρικά τμήματα (patches $64 \times 64 \times 64$). Η μονοδιάστατη σάρωση (1D scanning) του μηχανισμού Mamba διαρρηγνύει τη χωρική συνοχή, οδηγώντας σε υποδεέστερες ωμές προβλέψεις συγκριτικά με το παραδοσιακό συνελικτικό δίκτυο U-Net3D, το οποίο εκμεταλλεύεται εγγενώς το ισχυρό τοπικό του πεδίο αποδοχής.
- **Η Επίδραση της Μετα-επεξεργασίας LCC:** Η εφαρμογή του κλασικού βήματος Largest Connected Component, το οποίο διατηρεί μόνο τη μεγαλύτερη συνεκτική συνιστώσα κάθε πρόβλεψης, αποδείχθηκε καθοριστική για την ακρίβεια των ορίων. Αφαιρώντας τις απομακρυσμένες ψευδώς-θετικές νησίδες, το LCC οδήγησε σε σημαντική αύξηση του συντελεστή Dice (κατά 11.52 ποσοστιαίες μονάδες στο Swin-UMamba) και σε δραστική μείωση των σφαλμάτων απόστασης (HD95 και ASD). Καθώς πρόκειται για γενικής χρήσης βήμα μετα-επεξεργασίας ανεξάρτητο της αρχιτεκτονικής, βελτίωσε εξίσου και το συνελικτικό U-Net3D· η αποτελεσματικότητά του όμως εξαρτάται από την υπόθεση ότι το όργανο-στόχος αποτελεί τη μεγαλύτερη συνεκτική δομή της πρόβλεψης.
- **Η Αποτελεσματικότητα της Ημι-επιβλεπόμενης Μάθησης (DyCON):** Το πλαίσιο DyCON απέδωσε εξαιρετικά αποτελέσματα, αξιοποιώντας τις μη-σχολιασμένες τομο-

γραφίες και φιλτράροντας δυναμικά την αβεβαιότητα ώστε να αποφευχθεί ο θόρυβος επιβεβαίωσης (confirmation bias). Ως αποτέλεσμα, ο ισχυρότερος συνδυασμός (U-Net3D + DyCON + LCC) εκπαιδεύτηκε επιτυχώς με μόλις 12 σχολιασμένες τομογραφίες, επιτυγχάνοντας Dice 71.09% — επίδοση ουσιαστικά ισοδύναμη με το 71.89% της πλήρως επιβλεπόμενης εκπαίδευσης στο σύνολο των 62 τομογραφιών.

- **Ο Ρόλος της Φύσης των Ιατρικών Εικόνων:** Οι απόλυτες επιδόσεις όλων των μοντέλων παραμένουν συγκρατημένες (~70% Dice). Πέρα από την περιορισμένη χωρική ανάλυση των $64 \times 64 \times 64$ τμημάτων, σημαντικό ρόλο παίζουν οι εγγενείς ιδιότητες των ίδιων των ιατρικών δεδομένων: χαμηλή αντίθεση, ασαφή και θολά ανατομικά όρια, θόρυβος ετικετών και δια-παρατηρητική μεταβλητότητα. Αντίθετα με τις φυσικές εικόνες (π.χ. από κάμερες), όπου τα αντικείμενα διαχωρίζονται από ευκρινείς ακμές, υφές και χρώμα, στην ιατρική απεικόνιση το σήμα του οργάνου-στόχου συχνά είναι σχεδόν αδιάκριτο από τον περιβάλλοντα ιστό. Η ενδογενής αυτή διαφορά στη φύση του σήματος δυσκολεύει σημαντικά την ακριβή οριοθέτηση, ανεξάρτητα από την επιλογή της αρχιτεκτονικής.

- **Η Δυσανάλογη Αξία των Μη-Μαθησιακών Τεχνικών:** Ακριβώς επειδή το σφάλμα δεν οφείλεται αποκλειστικά στο μοντέλο, ένα απλό, *μη-μαθησιακό* βήμα μετα-επεξεργασίας (LCC) — χωρίς καμία εκπαιδεύσιμη παράμετρο — βελτίωσε την απόδοση δυσανάλογα περισσότερο από πολλές αρχιτεκτονικές επιλογές: υποδιπλασίασε το HD95 και ανέβασε το Dice έως και 11.5 μονάδες. Το εύρημα αυτό υπογραμμίζει ότι, στον ιατρικό τομέα, η κατανόηση της δομής του προβλήματος και η κατάλληλη μετα-επεξεργασία μπορούν να αποδώσουν περισσότερα από την τυφλή αύξηση της πολυπλοκότητας του δικτύου.

Συνοψίζοντας, η εργασία ανέδειξε ότι το ημι-επιβλεπόμενο πλαίσιο DyCON, σε συνδυασμό με ένα απλό βήμα μετα-επεξεργασίας, επιτρέπει υψηλής ακρίβειας κατάτμηση με ελάχιστα σχολιασμένα δεδομένα. Παράλληλα, κατέδειξε ότι, υπό αυστηρούς περιορισμούς χωρικής ανάλυσης, τα παραδοσιακά συνελικτικά δίκτυα παραμένουν εξαιρετικά ανταγωνιστικά έναντι των πιο σύγχρονων State Space Models — εύρημα που οριοθετεί με σαφήνεια τις συνθήκες υπό τις οποίες υπερτερεί κάθε οικογένεια αρχιτεκτονικών. Πάνω απ' όλα, όμως, η μελέτη κατέδειξε ότι το κρίσιμο εμπόδιο δεν είναι η επιλογή της αρχιτεκτονικής αλλά η ίδια η φύση των ιατρικών εικόνων: όταν το σήμα του στόχου είναι εγγενώς ασαφές, η στοχευμένη μετα-επεξεργασία που εκμεταλλεύεται τη δομή του προβλήματος αποδεικνύεται εξίσου — αν όχι περισσότερο — καθοριστική από την ίδια τη μαθησιακή ικανότητα του μοντέλου.

7.2 Περιορισμοί της Μελέτης

Τα ανωτέρω συμπεράσματα οφείλουν να ερμηνευθούν υπό το πρίσμα ορισμένων περιορισμών της πειραματικής διάταξης:

- **Περιορισμένη χωρική ανάλυση:** Λόγω των περιορισμών μνήμης του περιβάλλοντος Kaggle, όλα τα πειράματα εκτελέστηκαν σε μικρά τμήματα $64 \times 64 \times 64$. Συνεπώς,

η παρατηρούμενη υστέρηση του Swin-UMamba δεν συνιστά γενική κρίση για τα Μοντέλα Χώρου Καταστάσεων, αλλά αφορά αυστηρά το συγκεκριμένο καθεστώς χαμηλής ανάλυσης· σε πλήρη ανάλυση η σύγκριση ενδέχεται να αντιστραφεί.

- **Μικρό και μοναδικό σύνολο αξιολόγησης:** Η αποτίμηση στηρίχθηκε σε ένα μόνο σύνολο δεδομένων (Pancreas-CT) με μικρό σύνολο δοκιμής (20 τομογραφίες). Η υψηλή τυπική απόκλιση που παρατηρήθηκε — ιδίως για το Swin-UMamba — υπογραμμίζει ότι οι μέσες τιμές πρέπει να αντιμετωπίζονται με επιφύλαξη, ενώ η γενίκευση των ευρημάτων σε άλλα όργανα ή απεικονιστικές μεθόδους παραμένει ανοιχτή.

- **Εξάρτηση του LCC από την τοπολογία του στόχου:** Η αποτελεσματικότητα της μετα-επεξεργασίας προϋποθέτει ότι το όργανο-στόχος αποτελεί τη μεγαλύτερη συνεκτική συνιστώσα της πρόβλεψης. Όταν η υπόθεση αυτή παραβιάζεται, το LCC μπορεί να απορρίψει το ίδιο το όργανο, γεγονός που εξηγεί μέρος της παρατηρούμενης αστάθειας.

7.3 Μελλοντικές Επεκτάσεις

Παρά τα ενθαρρυντικά αποτελέσματα που επιτεύχθηκαν, η παρούσα έρευνα αποτελεί τη βάση για περαιτέρω διερεύνηση. Στο πλαίσιο των μελλοντικών επεκτάσεων του συστήματος, προτείνονται οι ακόλουθες κατευθύνσεις:

- **Αξιοποίηση Ισχυρότερων Υπολογιστικών Πόρων:** Λόγω των περιορισμών μνήμης γραφικών (OOM) του περιβάλλοντος Kaggle, το μέγεθος των 3D patches διατηρήθηκε μικρό ($64 \times 64 \times 64$). Μελλοντικές έρευνες σε ισχυρότερα συστήματα θα επέτρεπαν την εκπαίδευση σε μεγαλύτερα τμήματα (π.χ. $128 \times 128 \times 128$) ή και σε ολόκληρους τους ογκομετρικούς χάρτες, επιτρέποντας στο Swin-UMamba να αξιοποιήσει στο έπακρο το παγκόσμιο πεδίο αποδοχής του (global context) και πιθανώς να αντιστρέψει την υπεροχή του συνελικτικού δικτύου.

- **Μαθησιακή Τοπική Βαθμονόμηση εντός του Δικτύου:** Αντί του βήματος μετα-επεξεργασίας LCC, μια ενδιαφέρουσα κατεύθυνση θα ήταν η ενσωμάτωση ενός *μαθησιακού* μηχανισμού τοπικής βαθμονόμησης μέσα στο ίδιο το δίκτυο (π.χ. παραμένουσες τρισδιάστατες συνελίξεις βάθους επί των χαρτών χαρακτηριστικών του Mamba), ο οποίος θα αποκαθιστούσε την τοπική χωρική πληροφορία ήδη κατά την εκπαίδευση, αντί να εξυγιαίνει την πρόβλεψη εκ των υστέρων.

- **Πολυ-οργανική Κατάτμηση (Multi-organ Segmentation):** Η τρέχουσα μεθοδολογία επικεντρώθηκε αποκλειστικά στο πάγκρεας. Θα παρουσίαζε ιδιαίτερο ερευνητικό ενδιαφέρον η επέκταση του συστήματος για την ταυτόχρονη κατάτμηση πολλαπλών οργάνων της κοιλιακής χώρας (όπως το ήπαρ, ο σπλήνας και οι νεφροί). Σημειώνεται ότι σε αυτό το σενάριο η υπόθεση του LCC (ένα όργανο = μία συνεκτική συνιστώσα) παύει να ισχύει, και θα απαιτούσε γενίκευση σε διατήρηση πολλαπλών συνιστωσών ανά κλάση.

- **Διαπεδιακή Γενίκευση και Διαφορετικές Απεικονιστικές Μέθοδοι (Domain Generalization):** Το προτεινόμενο πλαίσιο θα μπορούσε να δοκιμαστεί σε σύνολα δεδομένων διαφορετικής φύσης, όπως η Μαγνητική Τομογραφία (MRI) ή οι Υπέρηχοι (Ultrasound). Η κατεύθυνση αυτή συνδέεται άμεσα με τη γενίκευση πεδίου: η εκπαίδευση σε μία απεικονιστική μέθοδο και η αξιολόγηση σε άλλη θα επέτρεπε τη μελέτη της ικανότητας του συστήματος να γενικεύει πέραν της κατανομής εκπαίδευσης, αξιοποιώντας την ευρωστία της Αντιθετικής Μάθησης (Contrastive Learning) του DyCON απέναντι στα διαφορετικά προφίλ θορύβου και τεχνημάτων (artifacts).

Βιβλιογραφία

[1] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A. W. M. van der Laak, Bram van Ginneken και Clara I. Sánchez. *A Survey on Deep Learning in Medical Image Analysis*. *Medical Image Analysis*, 42:60–88, 2017.

[2] Nima Tajbakhsh, Laura Jeyaseelan, Qian Li, Jeffrey N. Chiang, Zhihao Wu και Xiaowei Ding. *Embracing Imperfect Datasets: A Review of Deep Learning Solutions for Medical Image Segmentation*. *Medical Image Analysis*, 63:101693, 2020.

[3] Albert Gu και Tri Dao. *Mamba: Linear-Time Sequence Modeling with Selective State Spaces*. *Conference on Language Modeling (COLM)*, 2024.

[4] Özgün Çiçek, Ahmed Abdulkadir, Soeren S. Lienkamp, Thomas Brox και Olaf Ronneberger. *3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation*. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, σελίδες 424–432, 2016.

[5] Ali Hatamizadeh, Yucheng Tang, Vishwesh Nath, Dong Yang, Andriy Myronenko, Bennett Landman, Holger R. Roth και Daguang Xu. *UNETR: Transformers for 3D Medical Image Segmentation*. *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, σελίδες 574–584, 2022.

[6] Jiarun Liu, Hao Yang, Hong Yu Zhou, Yan Xi, Lequan Yu, Yizhou Yu, Yong Liang, Guangming Shi, Shaoting Zhang, Hairong Zheng και Shanshan Wang. *Swin-UMamba: Mamba-based UNet with ImageNet-based Pretraining*. *arXiv preprint arXiv:2402.03302*, 2024.

[7] Maregu Assefa, Muzammal Naseer, Iyyakutti Iyappan Ganapathi, Syed Sadaf Ali, Mohamed L. Seghier και Naoufel Werghi. *DyCON: Dynamic Uncertainty-aware Consistency and Contrastive Learning for Semi-supervised Medical Image Segmentation*. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025.

[8] Holger R. Roth, Le Lu, Amal Farag, Hoo Chang Shin, Jiamin Liu, Evrim B. Turkbey και Ronald M. Summers. *DeepOrgan: Multi-level Deep Convolutional Networks for Automated Pancreas Segmentation*. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, σελίδες 556–564, 2015.

[9] Olaf Ronneberger, Philipp Fischer και Thomas Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, σελίδες 234–241, 2015.

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren και Jian Sun. *Deep Residual Learning for Image Recognition*. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, σελίδες 770–778, 2016.

[11] Kunihiko Fukushima. *Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position*. *Biological Cybernetics*, 36(4):193–202, 1980.

[12] David H. Hubel και Torsten N. Wiesel. *Receptive Fields of Single Neurones in the Cat's Striate Cortex*. *The Journal of Physiology*, 148(3):574–591, 1959.

[13] Alex Waibel, Toshiyuki Hanazawa, Geoffrey Hinton, Kiyohiro Shikano και Kevin J. Lang. *Phoneme Recognition Using Time-Delay Neural Networks*. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(3):328–339, 1989.

[14] Yann LeCun, Léon Bottou, Yoshua Bengio και Patrick Haffner. *Gradient-based Learning Applied to Document Recognition*. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[15] Sergey Ioffe και Christian Szegedy. *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*. *International Conference on Machine Learning (ICML)*, σελίδες 448–456, 2015.

[16] Alex Krizhevsky, Ilya Sutskever και Geoffrey E. Hinton. *ImageNet Classification with Deep Convolutional Neural Networks*. *Advances in Neural Information Processing Systems (NeurIPS)*, σελίδες 1097–1105, 2012.

[17] Karen Simonyan και Andrew Zisserman. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. *International Conference on Learning Representations (ICLR)*, 2015.

[18] Dan C. Ciresan, Alessandro Giusti, Luca M. Gambardella και Jürgen Schmidhuber. *Deep Neural Networks Segment Neuronal Membranes in Electron Microscopy Images*. *Advances in Neural Information Processing Systems (NeurIPS)*, σελίδες 2843–2851, 2012.

[19] Jonathan Long, Evan Shelhamer και Trevor Darrell. *Fully Convolutional Networks for Semantic Segmentation*. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, σελίδες 3431–3440, 2015.

[20] Guodong Zeng, Xin Yang, Jing Li, Lequan Yu, Pheng Ann Heng και Guoyan Zheng. *3D U-Net with Multi-level Deep Supervision: Fully Automatic Segmentation of Proximal Femur in 3D MR Images*. *International Workshop on Machine Learning in Medical Imaging (MLMI)*, σελίδες 274–282, 2017.

[21] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit και Neil Houlsby. *An Image is Worth 16x16 Words:*

Transformers for Image Recognition at Scale. *International Conference on Learning Representations (ICLR)*, 2021.

[22] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin και Baining Guo. *Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows*. *IEEE/CVF International Conference on Computer Vision (ICCV)*, σελίδες 10012–10022, 2021.

[23] Albert Gu, Karan Goel και Christopher Ré. *Efficiently Modeling Long Sequences with Structured State Spaces*. *International Conference on Learning Representations (ICLR)*, 2022.

[24] Albert Gu, Tri Dao, Stefano Ermon, Atri Rudra και Christopher Ré. *HiPPO: Recurrent Memory with Optimal Polynomial Projections*. *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

[25] Tri Dao, Daniel Y. Fu, Stefano Ermon, Atri Rudra και Christopher Ré. *FlashAttention: Fast and Memory-Efficient Exact Attention with IO-Awareness*. *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

[26] Antti Tarvainen και Harri Valpola. *Mean Teachers Are Better Role Models: Weight-Averaged Consistency Targets Improve Semi-supervised Deep Learning Results*. *Advances in Neural Information Processing Systems (NeurIPS)*, σελίδες 1195–1204, 2017.

[27] Lequan Yu, Shujun Wang, Xiaomeng Li, Chi Wing Fu και Pheng Ann Heng. *Uncertainty-aware Self-ensembling Model for Semi-supervised 3D Left Atrium Segmentation*. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, σελίδες 605–613, 2019.

[28] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y. Hammerla, Bernhard Kainz, Ben Glocker και Daniel Rueckert. *Attention U-Net: Learning Where to Look for the Pancreas*. *Medical Imaging with Deep Learning (MIDL)*, 2018.

[29] Fausto Milletari, Nassir Navab και Seyed Ahmad Ahmadi. *V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation*. *International Conference on 3D Vision (3DV)*, σελίδες 565–571, 2016.

[30] Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L. Yuille και Yuyin Zhou. *TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation*. *arXiv preprint arXiv:2102.04306*, 2021.

[31] Hu Cao, Yueyue Wang, Joy Chen, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian και Manning Wang. *Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation*. *European Conference on Computer Vision Workshops (ECCVW)*, σελίδες 205–218, 2022.

[32] Ali Hatamizadeh, Vishwesh Nath, Yucheng Tang, Dong Yang, Holger R. Roth και Daguang Xu. *Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images*. *International MICCAI Brainlesion Workshop*, σελίδες 272–284, 2021.

[33] Lianghui Zhu, Bencheng Liao, Qian Zhang, Xinlong Wang, Wenyu Liu και Xinggang Wang. *Vision Mamba: Efficient Visual Representation Learning with Bidirectional State Space Model*. *International Conference on Machine Learning (ICML)*, 2024.

[34] Yue Liu, Yunjie Tian, Yuzhong Zhao, Hongtian Yu, Lingxi Xie, Yaowei Wang, Qixiang Ye και Yunfan Liu. *VMamba: Visual State Space Model*. *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.

[35] Jun Ma, Feifei Li και Bo Wang. *U-Mamba: Enhancing Long-range Dependency for Biomedical Image Segmentation*. *arXiv preprint arXiv:2401.04722*, 2024.

[36] Ting Chen, Simon Kornblith, Mohammad Norouzi και Geoffrey Hinton. *A Simple Framework for Contrastive Learning of Visual Representations*. *International Conference on Machine Learning (ICML)*, σελίδες 1597–1607, 2020.

[37] Kenneth Clark, Bruce Vendt, Kirk Smith, John Freymann, Justin Kirby, Paul Koppel, Stephen Moore, Stanley Phillips, David Maffitt, Michael Pringle, Lawrence Tarbox και Fred Prior. *The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository*. *Journal of Digital Imaging*, 26(6):1045–1057, 2013.

[38] Jia Deng, Wei Dong, Richard Socher, Li Jia Li, Kai Li και Li Fei-Fei. *ImageNet: A Large-Scale Hierarchical Image Database*. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, σελίδες 248–255, 2009.

[39] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai και Soumith Chintala. *PyTorch: An Imperative Style, High-Performance Deep Learning Library*. *Advances in Neural Information Processing Systems (NeurIPS)*, σελίδες 8024–8035, 2019.

[40] M. Jorge Cardoso, Wenqi Li, Richard Brown, Nic Ma, Eric Kerfoot, Yiheng Wang, Benjamin Murrey, Andriy Myronenko, Can Zhao, Dong Yang κ.ά. *MONAI: An Open-source Framework for Deep Learning in Healthcare*. *arXiv preprint arXiv:2211.02701*, 2022.

[41] Ilya Loshchilov και Frank Hutter. *Decoupled Weight Decay Regularization*. *International Conference on Learning Representations (ICLR)*, 2019.

Συντομογραφίες - Αρκτικόλεξα - Ακρωνύμια

κ.λπ.	και λοιπά
2D	Two-Dimensional
3D	Three-Dimensional
AdamW	Adam with Weight Decay
ASD	Average Surface Distance
BN	Batch Normalization
CE	Cross-Entropy
CNN	Convolutional Neural Network
CSM	Cross-Scan Module
CT	Computed Tomography
CUDA	Compute Unified Device Architecture
DSC	Dice Similarity Coefficient
DWConv	Depth-wise Convolution
DyCON	Dynamic Uncertainty-aware Consistency and Contrastive Learning
EMA	Exponential Moving Average
FCN	Fully Convolutional Network
FeCL	Focal Entropy-aware Contrastive Loss
GPU	Graphics Processing Unit
HBM	High Bandwidth Memory
HD95	95th percentile Hausdorff Distance
HiPPO	High-order Polynomial Projection Operators
LCC	Largest Connected Component
LTI	Linear Time-Invariant
MLP	Multi-Layer Perceptron
MONAI	Medical Open Network for AI
MRI	Magnetic Resonance Imaging
MSA	Multi-Head Self-Attention
MSE	Mean Squared Error
NIfTI	Neuroimaging Informatics Technology Initiative
NIH	National Institutes of Health
NPLR	Normal Plus Low-Rank
OOM	Out-of-Memory
PACS	Picture Archiving and Communication Systems
PET	Positron Emission Tomography
ReLU	Rectified Linear Unit

RNN	Recurrent Neural Network
S4	Structured State Space Sequence model
S6	Selective Scan (Mamba)
SiLU	Sigmoid Linear Unit
SRAM	Static Random Access Memory
SSL	Semi-Supervised Learning
SSM	State Space Model
SW-MSA	Shifted Window Multi-Head Self-Attention
TCIA	The Cancer Imaging Archive
UAMT	Uncertainty Aware Mean Teacher
UnCL	Uncertainty-aware Consistency Loss
ViT	Vision Transformer
VRAM	Video Random Access Memory
W-MSA	Window Multi-Head Self-Attention
ZOH	Zero-Order Hold

Απόδοση ξενόγλωσσων όρων

Απόδοση	Ξενόγλωσσος όρος
αβεβαιότητα	uncertainty
ακρίβεια ορίων	boundary accuracy
αναδειγματοληψία	upsampling
αντιθετική μάθηση	contrastive learning
αποκωδικοποιητής	decoder
αποσυνέλιξη	deconvolution
ασαφή όρια	weak boundaries
ασθενής επαύξηση	weak augmentation
ασυμμετρία κλάσεων	class imbalance
βαθιά μάθηση	deep learning
βασική αρχιτεκτονική	backbone
γενίκευση πεδίου	domain generalization
γραμμική προβολή	linear projection
δάσκαλος-μαθητής	teacher-student
διευρυμένη συνέλιξη	dilated convolution
εικονοστοιχείο	pixel
ελαστική παραμόρφωση	elastic deformation
ενεργά περιγράμματα	active contours
εντροπία	entropy
επαγωγική προκατάληψη	inductive bias
επαύξηση δεδομένων	data augmentation
επιλεκτική σάρωση	selective scan
επιπεδοποίηση	flattening
εστιακός παράγοντας	focal factor
θόρυβος επιβεβαίωσης	confirmation bias
θόρυβος ετικετών	label noise
κατάτμηση ιατρικής εικόνας	medical image segmentation
κλίση	gradient
κοινή χρήση βαρών	weight sharing
κρυφή κατάσταση	hidden state
κωδικοποίηση θέσης	positional embedding
κωδικοποιητής	encoder
λανθάνων χώρος	latent space
μακροπρόθεσμες εξαρτήσεις	long-range dependencies

μάθηση συνέπειας	consistency learning
μέγιστη συνεκτική συνιστώσα	largest connected component
μετα-επεξεργασία	post-processing
μη-σχολιασμένα δεδομένα	unlabeled data
μηχανισμός προσοχής	attention mechanism
ογκοστοιχείο	voxel
οπισθοδιάδοση	backpropagation
παράθυρο	window
πεδίο αποδοχής	receptive field
προ-εκπαίδευση	pre-training
πύλη πληροφορίας	information gate
σάρωση	scanning
συνεκτική συνιστώσα	connected component
συνέλιξη	convolution
συνέλιξη βάθους	depth-wise convolution
συσσώρευση κλίσης	gradient accumulation
σύνδεση παράκαμψης	skip connection
σχολιασμένα δεδομένα	labeled data
σχολιασμός αναφοράς	ground truth
τανυστής	tensor
υπερεκπαίδευση	overfitting
υπερπαράμετρος	hyperparameter
υπολειπόμενη σύνδεση	residual connection
χάρτης χαρακτηριστικών	feature map
ψευδο-ετικέτα	pseudo-label
ψευδώς-θετικό	false positive