



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ  
ΔΙΠΜΣ ΕΠΙΣΤΗΜΗ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

---

# Αντιστοίχιση Ιατρικών Εικόνων PET/CT Αξιοποιώντας τις Σύγχρονες Μεθόδους Deep Learning

---

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

Μυρτώ-Άννης Γκιρκίζας

Επιβλέπων: Γεώργιος Ματσόπουλος,  
Καθηγητής ΕΜΠ ΣΗΜΜΥ

Αθήνα, Μάρτιος 2024



## Abstract

Image registration, also known as image fusion or image matching, is the process of aligning two or more images based on their representation. Medical image registration aims to find an optimal spatial transformation that best matches the underlying anatomical structures. It is used in many clinical applications such as image guidance, motion tracking, segmentation, image reconstruction, clinical status comparison, and so on. Typical registration methods optimize an objective function independently for each pair of images, which can be time-consuming for large data sets. However, modern registration methods based on Deep Learning achieve similar or better results than optimization methods developed for the same purpose. Once pretrained, during the application phase, they only require the computation of the output with given network weights and input, which significantly reduces time. The goal of this thesis is the registration of medical PET/CT images utilizing both modern Deep Learning methods and traditional registration methods. This registration aims to facilitate the diagnosis of patients with metastatic melanoma through the use of both imaging modalities in locating cancerous tumors. Therefore, the aim of the thesis is to create a methodology based on Deep Learning for the alignment of PET/CT image pairs of the same patient and compare it with classical registration methods.

**Keywords:** Deep Learning, Medical Image Registration, Voxelmorph, SimpleITK

## Περίληψη

Η ευθυγράμμιση εικόνας (registration), γνωστή και ως σύντηξη εικόνας (image fusion) ή αντιστοίχιση εικόνας, είναι η διαδικασία ευθυγράμμισης δύο ή περισσότερων εικόνων με βάση την απεικόνισή τους. Η ευθυγράμμιση ιατρικής εικόνας επιδιώκει να βρει μια βέλτιστη χωρική μεταμόρφωση που αντιστοιχίζει καλύτερα τις υποκείμενες ανατομικές δομές. Χρησιμοποιείται σε πολλές κλινικές εφαρμογές όπως καθοδήγηση εικόνας (image guidance), παρακολούθηση κίνησης (motion tracking), κατάτμηση (segmentation), ανακατασκευή εικόνας (image reconstruction), σύγκριση της κλινικής κατάστασης και ούτω καθεξής. Οι τυπικές μέθοδοι ευθυγράμμισης βελτιστοποιούν μια αντικειμενική συνάρτηση ανεξάρτητα για κάθε ζεύγος εικόνων, η οποία και μπορεί να είναι χρονοβόρα για μεγάλα δεδομένα. Οι σύγχρονες, όμως, μέθοδοι ευθυγράμμισης που βασίζονται στην Βαθιά Μάθηση πετυχαίνουν ανάλογα ή και καλύτερα αποτελέσματα από τις μεθόδους βελτιστοποίησης που έχουν αναπτυχθεί για τον ίδιο σκοπό. Εφόσον, μάλιστα, έχουν εκπαιδευτεί εκ των προτέρων, κατά το στάδιο της εφαρμογής χρειάζονται μόνο τον υπολογισμό της εξόδου με δεδομένα τα βάρη του δικτύου και την είσοδο, πράγμα που απαιτεί σημαντικά λιγότερο χρόνο. Ο στόχος αυτής της εργασίας είναι η αντιστοίχιση ιατρικών εικόνων PET/CT αξιοποιώντας τόσο τις σύγχρονες μεθόδους της Βαθιάς Μάθησης (Deep Learning), όσο και τις τυπικές μεθόδους ευθυγράμμισης. Η αντιστοίχιση αυτή έχει ως σκοπό την διευκόλυνση της διάγνωσης του ασθενούς με μεταστατικό μελάνωμα μέσω της αξιοποίησης και τον δύο μέσων απεικόνισης στον εντοπισμό των καρκινικών όγκων. Σκοπός, λοιπόν της διπλωματικής είναι η δημιουργία μίας μεθοδολογίας βασισμένης στη Βαθιά Μάθηση για την ευθυγράμμιση ζευγών εικόνων PET/CT του ίδιου ασθενούς κι η σύγκρισή της με τις κλασικές μεθόδους ευθυγράμμισης.

**Λέξεις Κλειδιά:** Βαθιά Μάθηση, Ευθυγράμμιση Ιατρικών Εικόνων, Voxelmorph, SimpleITK

*Στους γονείς μου*

## Ευχαριστίες

Αρχικά, οφείλω να ευχαριστήσω τον επιβλέποντα καθηγητή μου κ. Μασσόπουλο, ο οποίος μου έδωσε την ευκαιρία να εκπονήσω τη συγκεκριμένη διπλωματική εργασία. Έπειτα, οφείλω να ευχαριστήσω τον Υποψήφιο Διδάκτορα Θεόδωρο Βαγενά, ο οποίος με καθοδήγησε καθ' όλη τη διάρκεια εκπόνησης της εργασίας και με στήριξε με πολύτιμες συμβουλές κι ιδέες. Τέλος, θα ήθελα να ευχαριστήσω την οικογένειά μου και τους φίλους για την εμπιστοσύνη και την ενθάρρυνση που μου προσέφεραν σε όλη την ακαδημαϊκή μου πορεία.

## Σύνοψη

Στην παρούσα εργασία εξετάζεται η ευθυγράμμιση ιατρικών εικόνων με τη βοήθεια των κλασικών τεχνικών ευθυγράμμισης, όπως επίσης με τη βοήθεια της Βαθιάς Μάθησης.

Η ευθυγράμμιση ιατρικών εικόνων αφορά τη διαδικασία συγκριτικής ανάλυσης και ταιριάσματος διαφορετικών εικόνων, ώστε να επιτευχθεί η βέλτιστη αντιστοίχιση ανάμεσά τους. Είναι ζωτικής σημασίας για τη διαγνωστική ακρίβεια, τη θεραπευτική παρακολούθηση και την ερευνητική ανάλυση στον τομέα της υγείας, καθώς επιτρέπει την ακριβή εκτίμηση μεταβολών, αναγνώριση μορφολογικών χαρακτηριστικών και την παρακολούθηση της εξέλιξης ασθενειών. Αυτό είναι ιδιαίτερα σημαντικό στην ιατρική απεικόνιση, όπου διάφορες μέθοδοι απεικόνισης (όπως CT, MRI, PET) παρέχουν διαφορετικούς τύπους πληροφορίας για το ίδιο ανατομικό χαρακτηριστικό.

Η σχέση της με τη Βαθιά Μάθηση, και ειδικά με τη μη επιβλεπόμενη Βαθιά Μάθηση, είναι στενή διότι τεχνολογίες όπως η Βαθιά Μάθηση (Deep Learning) μπορούν να αναλύσουν μεγάλους όγκους δεδομένων από ιατρικές εικόνες, ανακαλύπτοντας πρότυπα και χαρακτηριστικά που δεν είναι εύκολα αντιληπτά από το ανθρώπινο μάτι. Αυτό επιτρέπει την αυτοματοποιημένη και ακριβέστερη ευθυγράμμιση, ανοίγοντας νέους δρόμους στη διαχείριση και θεραπεία ασθενειών. Η μη επιβλεπόμενη μάθηση έχει αποδειχθεί ιδιαίτερα ωφέλιμη στην ευθυγράμμιση εικόνων, καθώς μοντέλα Βαθιάς Μάθησης μπορούν να εκπαιδευτούν να αναγνωρίζουν και να ταιριάζουν αντιστοιχίες μεταξύ εικόνων χωρίς ρητή επισήμανση. Αυτό επιτρέπει την αυτόματη ευθυγράμμιση μεγάλων συνόλων δεδομένων, αυξάνοντας την αποτελεσματικότητα και μειώνοντας το χρόνο επεξεργασίας. Επιπλέον, η επιβλεπόμενη μάθηση προσφέρει τη δυνατότητα για πιο εξειδικευμένη ευθυγράμμιση όταν τα δεδομένα εκπαίδευσης είναι προσεκτικά επισημασμένα. Οι συμβατικές μέθοδοι, όπως η αμοιβαία πληροφορία και οι μετρήσεις βασισμένες σε χαρακτηριστικά, εξακολουθούν να χρησιμοποιούνται αλλά υπολείπονται σε αποδοτικότητα και αυτοματοποίηση σε σύγκριση με τις μεθόδους Βαθιάς Μάθησης. Παρά την αποτελεσματικότητά τους σε συγκεκριμένες εφαρμογές, συχνά περιορίζονται από την απαιτούμενη χρονοβόρα επεξεργασία και την ανάγκη για εξειδικευμένες γνώσεις. Η συνεχής ανάπτυξη στην τεχνολογία Βαθιάς Μάθησης υπόσχεται περαιτέρω βελτιώσεις στην ακρίβεια, την ταχύτητα και την προσβασιμότητα της ευθυγράμμισης ιατρικών εικόνων, κάτι που θα συνδράμει τα μάλα στην εξέλιξή της.

Στα δεδομένα που εξετάζονται στην παρούσα διπλωματική, η ευθυγράμμιση γίνεται πάνω σε εικόνες διαφορετικών τρόπων λήψεως. Οι σταθερές εικόνες είναι αποτέλεσμα λήψεως αξονικής τομογραφίας και οι κινούμενες εικόνες είναι αποτέλεσμα λήψεως ποζιτρονικής τομογραφίας. Η αξονική τομογραφία (CT) και η ποζιτρονική τομογραφία (PET) είναι δύο σημαντικές απεικονιστικές τεχνικές στην ιατρική. Η CT χρησιμοποιεί ακτίνες X για τη δημιουργία λεπτομερειών εικόνων τού εσωτερικού τού σώματος, επιτρέποντας την ακριβή εκτίμηση της μορφής και του μεγέθους των οργάνων και των δομών. Αντίθετα, η PET εκμεταλλεύεται την εκπομπή ποζιτρονίων από ραδιενεργές ουσίες για την απεικόνιση της μεταβολικής δραστηριότητας των ιστών, προσφέροντας μοναδικές πληροφορίες σχετικά με τη λειτουργία τους. Η κύρια διαφορά μεταξύ των δύο μεθόδων είναι ότι η CT παρέχει ανατομικές πληροφορίες, ενώ η PET εστιάζει στη μεταβολική δραστηριότητα και λειτουργία των ιστών.

Οι τεχνικές που ακολουθούνται στην εργασία αναπτύσσουν τόσο συμβατικές μεθόδους, όσο και μεθόδους Βαθιάς Μάθησης. Συγκεκριμένα, στην πρώτη κατηγορία υλοποιείται η ευθυγράμμιση με χρήση της SimpleITK, ενώ στη δεύτερη κατηγορία χρησιμοποιείται το μοντέλο VoxelMorph. Η SimpleITK είναι μια διεπαφή για τη βιβλιοθήκη ITK, η οποία διευκολύνει την επεξεργασία και ανάλυση ιατρικών εικόνων σε πολλές γλώσσες προγραμματισμού. Παρέχει εργαλεία για την ανάγνωση, κατάτμηση, επεξεργασία και ευθυγράμμιση εικόνων. Το VoxelMorph είναι ένας αλγόριθμος Βαθιάς Μάθησης, που αναπτύχθηκε για την αυτοματοποιημένη εγγραφή εικόνων, χρησιμοποιώντας μη επιβλεπόμενες μεθόδους, με σκοπό να παράγει γρήγορα και ακριβή αποτελέσματα, εξοικονομώντας πολύτιμο χρόνο και πόρους σε ιατρικές εφαρμογές.

Για την εγγραφή με τη SimpleITK αξιοποιούνται δύο μέθοδοι - τα φίλτρα Demons και η ImageRegistration-

Method. Για τα Demons χρησιμοποιείται μια διαδικασία πυραμίδας, όπου η εικόνα διαιρείται, και κάθε υποσύνολό της βρίσκεται σε διαφορετική ανάλυση. Η ευθυγράμμιση γίνεται διαδοχικά από το χαμηλότερο επίπεδο στο υψηλότερο. Κάθε εξαχθέν πεδίο μετασχηματισμού, λειτουργεί ως είσοδος στην επόμενη ευθυγράμμιση, μέχρις ότου επιτευχθεί το ανώτατο επίπεδο ανάλυσης. Το τελευταίο επίπεδο μετασχηματισμού αποτελεί το πεδίο παραμόρφωσης που θα εφαρμοστεί στην κινούμενη εικόνα, ώστε να γίνει η αντιστοίχιση με τη σταθερή. Δοκιμάζονται διαφορετικές παράμετροι για την αρχικοποίηση της ευθυγράμμισης, που οδηγούν σε διαφορετικά αποτελέσματα ευθυγράμμισης. Για την ImageRegistrationMethod, διεξάγεται μια διαδοχική έρευνα ώστε να επιτευχθεί η καλύτερη ευθυγράμμιση. Ξεκινώντας με την αρχική ευθυγράμμιση, συνεχίζοντας με τους βελτιστοποιητές, τη συνάρτηση παρεμβολής, τον αριθμό των καλαθιών (bins) και ολοκληρώνοντας με την επιλογή ποσοστού υποδειγματισμού (sampling percentage), εξακριβώνεται ο συνδυασμός των παραμέτρων, ο οποίος θα επιφέρει τα πιο αποδοτικά αποτελέσματα ευθυγράμμισης.

Για την σύντηξη με το Voxelmorph αξιοποιείται το μοντέλο Unet για την εξαγωγή των πεδίων παραμόρφωσης και το μοντέλο STN για τη εφαρμογή αυτών στην κινούμενη εικόνα. Το Unet έχει τη μορφή κωδικοποιητή-αποκωδικοποιητή, με τον κωδικοποιητή να χρησιμεύει στην εξαγωγή χαρακτηριστικών των εικόνων και τον αποκωδικοποιητή να ενσωματώνει χωρική πληροφορία στα εξαχθέντα πεδία. Η έξοδος είναι ένας τρισδιάστατος πίνακας, ο οποίος σε κάθε όγκο του φωλιάζει ένα επίσης τρισδιάστατο διάνυσμα. Το διάνυσμα αυτό, δείχνει την αλλαγή που πρέπει να εφαρμοστεί σε κάθε ογκοστοιχείο της αρχικής κινούμενης εικόνας. Προκειμένου να γίνει η αλλαγή αποτελεσματικά, το πεδίο μετασχηματισμού περνάει από συγκεκριμένες διαδικασίες (υποβάθμιση, ολοκλήρωση, αναβάθμιση) για να διασφαλιστεί η ομαλότητα και η συνέχειά του, αποφεύγοντας μη ρεαλιστικές παραμορφώσεις. Έπειτα από αυτές τις διαδικασίες, η κινούμενη εικόνα παρεμβάλλεται σύμφωνα με το πλέγμα που έχει δημιουργηθεί από το διανυσματικό πεδίο, έτσι ώστε να αντιστοιχιστεί στις αρχικές θέσεις των ογκοστοιχείων της σταθερής εικόνας. Για το μοντέλο Voxelmorph εξετάζονται διαφορετικές τιμές υπερπαραμέτρων και συναρτήσεων κόστους, ώστε να γίνει όσο το δυνατόν καλύτερη η εγγραφή των εικόνων.

Οι ευθυγραμμίσεις που δοκιμάζονται παράγουν ποικίλα αποτελέσματα. Η SimpleITK δυσκολεύεται να ευθυγραμμίσει τις εικόνες, παρόλο που ως ένα βαθμό το πετυχαίνει. Η δυσκολία αυτή οφείλεται κυρίως στην πολυτροπικότητα των δεδομένων, δεδομένου ότι οι εικόνες απεικονίζουν διαφορετικές φυσικές ιδιότητες και κατανομές έντασης. Έτσι, αν και φαίνεται ότι υπάρχει μια σύντηξη των εικόνων, αυτή δεν είναι η βέλτιστη, καθώς επίσης δεν υπάρχουν μεγάλα περιθώρια βελτίωσης. Στον αντίποδα, το Voxelmorph μαθαίνει τα δεδομένα και μπορεί δυνητικά να πετύχει καλύτερη ευθυγράμμιση. Η συνάρτηση κόστους μειώνεται σταδιακά, μολαταύτα δε συγκλίνει. Αυτό είναι απόρροια του μικρού αριθμού των εποχών. Επιλέγοντας τιμές για τον παράγοντα ρύθμισης  $\lambda$ , διαφοροποιούνται τα αποτελέσματα. Αν και η ευθυγράμμιση είναι ημιτελής, με τις τιμές Mutual Information (MI), μάλιστα, να σημειώνουν μικρότερη κλίμακα σε σχέση με τη SimpleITK, διαφαίνονται αρκετά περιθώρια καλυτέρευσης και βελτιστοποίησης για το Voxelmorph. Τα συμπεράσματα συνάδουν με αυτά της βιβλιογραφίας, καθώς οι νέες μέθοδοι ευθυγράμμισης που βασίζονται στη Βαθιά Μάθηση, είναι αποτελεσματικότερες σε τέτοιου είδους εγχειρήματα [48].





# Contents

<b>1</b>	<b>Μέρος Α</b>	<b>13</b>
1.1	Σημαντικότητα MIR	13
1.2	Ταξινόμηση της Ευθυγράμμισης Εικόνων (Image Registration - IR)	13
1.3	Τεχνική Ποζιτρονικής Τομογραφίας (Positron emission tomography - PET)	18
1.4	Τεχνική Αξονικής Τομογραφίας (Computed tomography - CT)	18
1.5	MIR και Βαθιά Μάθηση (Deep Learning - DL)	19
1.5.1	Μη Επιβλεπόμενη Μάθηση (Unsupervised Learning)	22
1.5.2	Μη Επιβλεπόμενη Μάθηση στην MIR	23
1.6	Μετρικές ομοιότητας/αξιολόγησης και MIR	23
1.7	Βιβλιογραφική Ανασκόπηση	25
1.8	Unet	27
1.8.1	Βασικές διαδικασίες μοντέλων	27
1.8.2	Αρχιτεκτονική μοντέλου Unet	34
1.9	Voxelmorph	38
1.9.1	Αρχιτεκτονική μοντέλου Voxelmorph	39
<b>2</b>	<b>Μέρος Β</b>	<b>45</b>
2.1	SimpleITK βιβλιοθήκη	45
2.2	Περιβάλλον προγραμματισμού και χρήσιμα εργαλεία	46
2.3	Σύνολο Δεδομένων	46
2.4	Προεπεξεργασία δεδομένων	48
2.5	Ευθυγράμμιση με SimpleITK	51
2.6	Εκπαίδευση με το Voxelmorph	51
2.7	Αποτελέσματα Ευθυγράμμισης	52
2.7.1	SimpleITK με Demons φίλτρα	53
2.7.2	SimpleITK με ImageRegistrationMethod	54
2.7.3	Συμπεράσματα για την ευθυγράμμιση με SimpleITK	57
2.7.4	Voxelmorph	58
2.7.5	Συμπεράσματα για την ευθυγράμμιση με Voxelmorph	61
2.7.6	Σύγκριση μεταξύ των δύο τεχνικών (SimpleITK, Voxelmorph)	61
<b>3</b>	<b>Συμπεράσματα</b>	<b>63</b>
<b>4</b>	<b>Μελλοντική Έρευνα</b>	<b>64</b>

## List of Figures

1	Ταξινόμηση της Ευθυγράμμισης Εικόνων [14].	14
2	(a) Επεμβατική διαδικασία χρησιμοποιώντας ένα στερεοτακτικό σκελετό (stereotactic frame) τοποθετημένο στο κεφάλι ασθενούς. (b) Μη επεμβατική διαδικασία χρησιμοποιώντας επιδερμικά στίγματα (skin-mounted markers). Οι εικόνες δημιουργήθηκαν από το DALL-E image generation model του OpenAI.	15
3	Οπτική σύγκλιση γραμμών τρένου λόγω προοπτικής. Η εικόνα δημιουργήθηκε από το DALL-E image generation model του OpenAI.	16
4	Παραδείγματα 2D μετασχηματισμών.[14]	17

5	Διάγραμμα λειτουργίας Pet. Ο πυρήνας τού ακτινοϊσοτροπικού στοιχείου εκπέμπει ένα ποζιτρόνιο (θετικό ηλεκτρόνιο). Το ποζιτρόνιο συγκρούεται με ένα ηλεκτρόνιο στον ιστό τού σώματος και κατά τη σύγκρουση μετασχηματίζει τη μάζα σε ενέργεια με τη μορφή δύο φωτονίων. Η κάμερα PET χρησιμοποιεί κρυστάλλους απόσβεσης τοποθετημένους γύρω από το αντικείμενο για να εντοπίζει αυτά τα φωτόνια. Τέλος, οι κρυστάλλοι απορροφούν τα φωτόνια, παράγοντας φως, το οποίο μετατρέπεται σε ηλεκτρικό σήμα [49]. . . . .	18
6	Διάγραμμα λειτουργίας CT [50]. . . . .	19
7	Διάγραμμα ροής ευθυγράμμισης εικόνων [2]. . . . .	20
8	Σύνοψη των επτά κατηγοριών μεθόδων βασισμένων στη Βαθιά Μάθηση για τη MIR.[11] . . . . .	22
9	Αριστερά η αρχική εικόνα. Στη μέση η εικόνα που της έχει εφαρμοστεί ο left sobel πυρήνας. Δεξιά η εικόνα που της έχει εφαρμοστεί ο right sobel πυρήνας. . . . .	29
10	(a) Γέμισμα με μηδέν. (b) Γέμισμα με αναπαραγωγή (c) Γέμισμα με αντανάκλαση. . . . .	31
11	(a) Αρχική εικόνα, επιπεδοποιημένη (flattened) εικόνα, πυρήνας (kernel), unrolled kernel. (b) Η εικόνα δείχνει πώς προκύπτει ο unrolled kernel. Τα κουτάκια με κίτρινο χρώμα δείχνουν τη θέση τού πυρήνα πάνω στην εικόνα. Ο πυρήνας καλύπτει τις θέσεις 1,2,4,5. Οι υπόλοιπες θέσεις είναι μηδέν (zero padding). Με τον ίδιο τρόπο υπολογίζεται κι ο υπόλοιπος πίνακας. . . . .	34
12	Διαδικασία αντίστροφης συνέλιξης. . . . .	34
13	Αρχιτεκτονική κωδικοποιητή Unet.[20] . . . . .	36
14	Αρχιτεκτονική αποκωδικοποιητή Unet [20]. . . . .	37
15	Αρχιτεκτονική Unet [20]. . . . .	38
16	Αρχιτεκτονική Unet για το Voxelmorph. Κάθε ορθογώνιο αναπαριστά έναν τρισδιάστατο πίνακα που παράχθηκε από τον προηγούμενο πίνακα χρησιμοποιώντας τρισδιάστατα συνελικτικά επίπεδα. Η χωρική ανάλυση κάθε όγκου-πίνακα καταγράφεται κάτωθεν των ορθογωνίων. Στον αποκωδικοποιητή χρησιμοποιούνται αρκετές συνελιξεις με 32 πυρήνες, ακολουθούμενοι από αντίστροφα συνελικτικά επίπεδα ώστε να φέρουν τον όγκο-πίνακα στην αρχική του ανάλυση. Τα βέλη αναπαριστούν τις συνδέσεις παράβλεψης, οι οποίες συγχωνεύονται με τα χαρακτηριστικά τού αποκωδικοποιητή. Έχοντας φτάσει στην αρχική ανάλυση, χρησιμοποιούνται επιπλέον συνελιξεις μέχρι την έξοδο [30]. . . . .	39
17	Με κόκκινο το σημείο $x = (1.5, 2.5)$ , με πράσινο τα κοντινότερα σημεία $(1, 2), (1, 3), (2, 2), (2, 3)$ . . . . .	41
18	B-splines μηδενικής τάξης. Ο x άξονας αναπαριστά τις τιμές 0 με 4 και ο y άξονας αναπαριστά την τιμή της B-spline συνάρτησης. Κάθε B-spline μοιάζει με κουτί κι έχει ύψος 1 μεταξύ των ορίων του, ενώ γίνεται 0 έξω από αυτά. . . . .	42
19	Δεδομένων μηδενικής τάξεως B-splines ορισμένες στο διαστήματα $[0, 1), [1, 2), [2, 3), [3, 4)$ η πρώτης τάξεως B-spline προερχόμενη από τη συνέλιξη των θα εμφανίσει μια γραμμική αύξηση, μια κορυφή και μια γραμμική μείωση μεταξύ δύο διαδοχικών διαστημάτων λόγω της διαδικασίας τής συνέλιξης. . . . .	43
20	Συνολική αρχιτεκτονική Voxelmorph.[30] . . . . .	44
21	Siemens Biograph mCT [44]. . . . .	47
22	CT παράδειγμα ασθενούς. . . . .	48
23	PET παράδειγμα ασθενούς. . . . .	48
24	Παράδειγμα αποτελέσματος προεπεξεργασίας εικόνων για ένα αντικείμενο CT. Η οπτικοποίηση έγινε με τη χρήση τού 3D Slicer 5.2.2. (a) Αρχική μη επεξεργασμένη CT εικόνα. (b) Εικόνα CT έπειτα από την υποβάθμιση. (c) Εικόνα CT έπειτα από την κανονικοποίηση. (d) Εικόνα CT έπειτα από την εξομάλυνση. . . . .	50
25	Παράδειγμα αποτελέσματος προεπεξεργασίας εικόνων για ένα αντικείμενο PET. Η οπτικοποίηση έγινε με τη χρήση τού 3D Slicer 5.2.2. (a) Αρχική μη επεξεργασμένη PET εικόνα. (b) Εικόνα PET έπειτα από την υποβάθμιση. (c) Εικόνα PET έπειτα από την κανονικοποίηση. (d) Εικόνα PET έπειτα από την εξομάλυνση. . . . .	50
26	Τιμές αμοιβαίας πληροφορίας για όλα τα ζευγάρια σταθερών, κινούμενων εικόνων πριν την ευθυγράμμιση. . . . .	53

27	Αποτελέσματα ευθυγράμμισης με SimpleITK για το μοντέλο με Demons φίλτρο SymmetricForces-DemonsRegistrationFilter, επαναλήψεις 110 τυπική απόκλιση 1.0 και συνάρτηση παρεμβολής των Κοντινότερο Γείτονα. Αριστερά η σταθερή εικόνα, στη μέση η κινούμενη και δεξιά η ευθυγραμμισμένη.	54
28	Αποτελέσματα ευθυγράμμισης με SimpleITK για το μοντέλο με Demons φίλτρο SymmetricForces-DemonsRegistrationFilter, επαναλήψεις 120 τυπική απόκλιση 0.5 και συνάρτηση παρεμβολής των Κοντινότερο Γείτονα. Αριστερά η σταθερή εικόνα, στη μέση η κινούμενη και δεξιά η ευθυγραμμισμένη.	54
29	Αποτελέσματα ευθυγράμμισης για τους διαφορετικούς τρόπους αρχικοποίησης. . . . .	55
30	Αποτελέσματα ευθυγράμμισης για τους διαφορετικούς βελτιστοποιητές. . . . .	55
31	Αποτελέσματα ευθυγράμμισης για τις διαφορετικές συναρτήσεις παρεμβολής. . . . .	56
32	Αποτελέσματα ευθυγράμμισης για διαφορετικά bins και ποσοστά υποδειγματισμού. . . . .	57
33	Αποτελέσματα ευθυγράμμισης με SimpleITK για το καλύτερο μοντέλο της ImageRegistration-Method. Αριστερά η σταθερή εικόνα, στη μέση η κινούμενη και δεξιά η ευθυγραμμισμένη. . . . .	57
34	Αποτελέσματα εκπαίδευσης Voxelmorph με συνάρτηση κόστους (loss function) το MSE, αριθμό εποχών 80 και $\lambda = 0.01$ . Στην εικόνα (a) φαίνεται η εξέλιξη της συνάρτησης κόστους σε σχέση με τις εποχές, ενώ στην εικόνα (b) φαίνεται η εξέλιξη της μετρικής MI με την πρόοδο των εποχών. . .	58
36	Αποτελέσματα ευθυγράμμισης με Voxelmorph για το κόστος MSE με $\lambda = 0.01$ . Αριστερά η σταθερή εικόνα, στη μέση η κινούμενη και δεξιά η ευθυγραμμισμένη. . . . .	58
35	Αποτελέσματα εκπαίδευσης Voxelmorph με συνάρτηση κόστους (loss function) το NCC, αριθμό εποχών 80 και $\lambda = 0.01$ . Στην εικόνα (a) φαίνεται η εξέλιξη της συνάρτησης κόστους σε σχέση με τις εποχές, ενώ στην εικόνα (b) φαίνεται η εξέλιξη της μετρικής MI με την πρόοδο των εποχών. . .	59
37	Αποτελέσματα ευθυγράμμισης με Voxelmorph για το κόστος NCC με $\lambda = 0.01$ . Αριστερά η σταθερή εικόνα, στη μέση η κινούμενη και δεξιά η ευθυγραμμισμένη. . . . .	59
38	Αποτελέσματα εκπαίδευσης Voxelmorph με συνάρτηση κόστους (loss function) το NCC, αριθμό εποχών 80 και $\lambda = 0.1$ . Στην εικόνα (a) φαίνεται η εξέλιξη της συνάρτησης κόστους σε σχέση με τις εποχές, ενώ στην εικόνα (b) φαίνεται η εξέλιξη της μετρικής MI με την πρόοδο των εποχών. . .	60
39	Αποτελέσματα εκπαίδευσης Voxelmorph με συνάρτηση κόστους (loss function) το MSE και $\lambda = 0.6$ . Στην εικόνα (a) φαίνεται η ευθυγράμμιση στις 50 εποχές, ενώ στην εικόνα (b) φαίνεται η ευθυγράμμιση για τις 80 εποχές. Για τις δύο περιπτώσεις, αριστερά είναι η σταθερή εικόνα, στη μέση η κινούμενη και δεξιά η μετασχηματισμένη. . . . .	60
40	Δομή μοντέλου Voxelmorph. . . . .	70

## List of Tables

1	Παραδείγματα τεχνικών για κάθε ένα τύπο τρόπου λήψης [14]. . . . .	18
2	Επίπεδα Unet, διαστάσεις πινάκων χαρακτηριστικών και αριθμός παραμέτρων προς εκπαίδευση. . . . .	52
3	Παραμετροποιήσεις ευθυγράμμισης με τα Demons. . . . .	53
4	Παραμετροποιήσεις ευθυγράμμισης για τους βελτιστοποιητές. . . . .	56
5	Παραμετροποιήσεις ευθυγράμμισης για το Voxelmorph. . . . .	60



## Εισαγωγή

Η ιατρική απεικόνιση (medical imaging) είναι από τα σημαντικότερα διαγνωστικά εργαλεία για πάμπολλες ασθένειες. Οι πληροφορίες που λαμβάνονται από τις ιατρικές εικόνες υποστηρίζουν τόσο τη λήψη αποφάσεων για τις θεραπείες των ασθενών, όσο και για την ίδια την πρόγνωση και την πρόβλεψη των ίδιων των ασθενειών. Τα στάδια στα οποία υπεισέρχεται η βοήθειά τους μπορεί να είναι, φυσικά, η θεραπεία, η ανίχνευση, ο χαρακτηρισμός, η ανταπόκριση στη θεραπεία, η παρακολούθηση επανεμφάνισης της νόσου και τέλος η καθοδήγηση χειρουργικών επεμβάσεων κι ακτινοθεραπείας. Ο αριθμός των ιατρικών εικόνων για κάθε περίπτωση ασθενούς αυξάνεται σε πολυπλοκότητα, καθώς εικόνες δισδιάστατες, πλέον γίνονται τρισδιάστατες και συμπεριλαμβανομένου τού χρόνου τετραδιάστατες [45]. Προκειμένου να αναλυθούν αυτές οι εικόνες με όσο το δυνατόν μεγαλύτερη ευλάβεια είναι αναγκαία η συμπερίληψη μεθόδων, όπως η Βαθιά Μάθηση, η οποία τυγχάνει ραγδαίας ανάπτυξης τα τελευταία χρόνια, με αποτέλεσμα να διαπλάτυνει τα πεδία έρευνας κι εφαρμογής της.

Στα πλαίσια της ανάλυσης των ιατρικών εικόνων, η ευθυγράμμιση (Registration) αποτελεί ένα δομικό στοιχείο. Στις περισσότερες ιατρικές παρεμβάσεις, υπάρχουν περιπτώσεις στις οποίες χρειάζεται να ληφθούν εικόνες για διαγνωστικούς, προγνωστικούς, θεραπευτικούς ή παρακολουθητικούς σκοπούς. Αυτές οι εικόνες ίσως διαφέρουν ως προς το χρόνο λήψης, τις διαστάσεις ή τη φύση των. Η σύντηξή τους προκαλεί συνέργεια πληροφοριών κι είναι αρχικώς σημαντική για την καθοδήγηση κι επικουρία των ιατρών στη λήψη αποφάσεων. Η ευθυγράμμιση ιατρικών εικόνων (Medical Image Registration) επιλύει το σκόπελο της διαφορετικότητάς τους, αντιστοιχίζοντάς τις με βάση ένα ταυτόσημο γεωμετρικό σύστημα συντεταγμένων. Σκοπός είναι να βρεθεί η βέλτιστη χωρική μετατροπή, η οποία ευθυγραμμίζει τις δομές ενδιαφέροντος με τον πιο αποτελεσματικό τρόπο [46].

Όλο και περισσότερα μοντέλα Μηχανικής Μάθησης χρησιμοποιούνται για να αυτοματοποιήσουν τη διαδικασία της ευθυγράμμισης. Τα μοντέλα Βαθιάς Μάθησης, ιδίως τα συνελικτικά νευρωνικά δίκτυα (CNN) έχουν επιδείξει αξιοσημείωτη ικανότητα στο να μαθαίνουν περίπλοκα μοτίβα και χαρακτηριστικά από ιατρικές απεικονίσεις, επιτρέποντας την αυτοματοποιημένη αντιστοίχιση των εικόνων με υψηλή ακρίβεια. Τα μοντέλα αυτά μαθαίνουν τη συμβατότερη χωρική μετατροπή απευθείας από τα δεδομένα, αμβλύνοντας την ανάγκη ανθρώπινης παρέμβασης. Η εφαρμογή της Βαθιάς Μάθησης στην ευθυγράμμιση ιατρικών εικόνων όχι μόνο βελτίωσε την ακρίβεια και την ευρωστία των αλγορίθμων ευθυγράμμισης, αλλά επίσης μείωσε σημαντικά τον υπολογιστικό χρόνο, καθιστώντας τις εφαρμογές σε πραγματικό χρόνο πιο εφικτές. Για να επιτύχουν την ευθυγράμμιση τα μοντέλα Βαθιάς Μάθησης, συνήθως υπολογίζουν ένα πεδίο μετασχηματισμού (deformation field), το οποίο αντιστοιχίζει εικονοστοιχεία (pixel) μεταξύ των εικόνων ενδιαφέροντος [47]. Διαφορετικές ανατομικές δομές ή ιστοί συνήθως εκδηλώνονται ως ξεχωριστές εντάσεις εικονοστοιχείων, επομένως η αντιστοίχιση μεταξύ ομοίων εντάσεων εικονοστοιχείων στις εικόνες είναι κοινή πρακτική για την ευθυγράμμιση των ίδιων δομών [47].

Παρ' όλη την πρόοδο που επιφέρει η Βαθιά Μάθηση, οι συμβατικές μέθοδοι ευθυγράμμισης παραμένουν στην επιφάνεια, ιδιαίτερα όταν τα μοντέλα Βαθιάς Μάθησης δεν είναι εφαρμόσιμα ή όταν χρειάζεται απόλυτη εποπτεία της διαδικασίας ευθυγράμμισης. Εργαλεία όπως το SimpleITK, ITK (Insight Segmentation and Registration Toolkit) και το Elastix προσφέρουν εύρωστες, ευέλικτες πλατφόρμες για την ευθυγράμμιση ιατρικών εικόνων, παρέχοντας ένα ευρύ φάσμα αλγορίθμων τόσο για άκαμπτους (rigid) όσο και για μη άκαμπτους μετασχηματισμούς.

Σκοπός της παρούσης διπλωματικής είναι η εμβάθυνση στην ευθυγράμμιση των ιατρικών εικόνων και στις τεχνικές που ακολουθούνται για να την επιτύχουν. Ακόμη, στόχος είναι ο πειραματισμός πάνω σε ιατρικά δεδομένα διαφορετικής τεχνικής λήψεως (multimodal data), τόσο με την εφαρμογή αλγοριθμικών μοντέλων Βαθιάς Μάθησης, όσο και με την αξιοποίηση παραδοσιακών εργαλείων ευθυγράμμισης. Η διαδικασία που προτείνεται αφορά μη επιβλεπόμενη μάθηση, κάτι που οξύνει τη δυσκολία της ευθυγράμμισης και της εξαγωγής αποτελεσμάτων και για τις δύο μεθόδους.

Η συγκεκριμένη εργασία απαρτίζεται από τέσσερα κεφάλαια:

- Μέρος Α: στο πρώτο κεφάλαιο αναλύεται όλο το θεωρητικό υπόβαθρο που χρειάζεται για την πλήρη κατανόηση της φύσης και των τεχνικών ευθυγράμμισης. Εξετάζεται ο ρόλος κι η σημασία τής ευθυγράμμισης, οι τρόποι που αυτή κατηγοριοποιείται, η σχέση της με τη Βαθιά Μάθηση, διάφορες μετρικές αξιολόγησης της ευθυγράμμισης, όπως επίσης γίνεται μια εκτενής αναφορά στα μοντέλα Unet και Voxelmorph.
- Μέρος Β: στο κεφάλαιο δεύτερο παρουσιάζεται το πειραματικό μέρος. Γίνεται αναφορά στη βιβλιοθήκη SimpleITK και τα προγραμματιστικά εργαλεία που χρησιμοποιούνται. Περιγράφονται τα δεδομένα τού πειράματος κι η προεπεξεργασία τους. Αναγράφεται η δομή και τα χαρακτηριστικά εκπαίδευσης του Voxelmorph και της εκπαίδευσης με τη SimpleITK, ενώ συγκρίνονται τα αποτελέσματα από τις δύο τεχνικές.
- Συμπεράσματα: στο τρίτο κεφάλαιο αναλύονται τα αποτελέσματα που αναφέρθηκαν στο κεφάλαιο δύο και εξάγονται τα τελικά συμπεράσματα.
- Μελλοντική Έρευνα: στο τέταρτο και τελευταίο κεφάλαιο γίνονται προτάσεις για περαιτέρω επέκταση και μελέτη τού παρόντος θέματος.

# 1 Μέρος Α

Η ιατρική απεικόνιση διαδραματίζει σημαντικό ρόλο στη σύγχρονη υγειονομική περίθαλψη [10], επιτρέποντας με μη παρεμβατικό τρόπο την οπτικοποίηση ανατομικών δομών και παθολογικών καταστάσεων εντός του ανθρώπινου σώματος. Μολαταύτα, για να αξιοποιηθούν πλήρως οι πληροφορίες που προέρχονται από τις ιατρικές απεικονίσεις, είναι συχνά απαραίτητο να εφαρμοστεί η διαδικασία της ευθυγράμμισης ιατρικών εικόνων (Medical Image Registration - MIR). Η εν λόγω διαδικασία εσωκλείνει την ευθυγράμμιση (alignment) και τη συγχώνευση διαφορετικών εικόνων ιδίων ή διαφορετικών ανατομών, παρμένες σε διαφορετικές χρονικές στιγμές, από διαφορετικές οπτικές ή αξιοποιώντας διαφορετικές τεχνικές. Η ευθυγράμμιση διευκολύνει τους ειδικούς στο να εκμαιεύουν χρήσιμες πληροφορίες, οι οποίες βοηθούν στις διαγνώσεις, τη θεραπεία ή την παρακολούθηση των διαφόρων ασθενειών [1].

## 1.1 Σημαντικότητα MIR

Η διαδικασία της ευθυγράμμισης ιατρικών εικόνων είναι καθοριστική στον τομέα της υγείας, εφόσον ενισχύει τη διαγνωστική ακρίβεια, διευκολύνει την κατάτμηση των εικόνων, βοηθάει στη δημιουργία άτλα και βελτιώνει τη θεραπεία και την παρακολούθηση των ασθενειών. Η σημαντικότητα της MIR διαφαίνεται από τη χρήση της σε:

1. Διαγνώσεις: η MIR βοηθάει στην καλύτερευση των διαγνώσεων, δεδομένου ότι καθιστά εφικτή τη συγχώνευση πληροφοριών προερχόμενες από διαφορετικές τεχνικές (λόγου χάρι CT, PET), μειώνοντας το κόστος και την έκθεση στην ακτινοβολία [4]. Αντιστοιχίζει διαφορετικά σύνολα εικόνων σε κοινό χώρο, επιτρέποντας, έτσι, μια ακριβέστερη και πληρέστερη ανάλυση. Αυτή η διαδικασία είναι άκρως σημαντική για την ορθή διάγνωση των ασθενειών, εφόσον καθιστά εφικτή τη σύγκριση των εικόνων με άλλες, αποκτημένες σε προγενέστερο χρόνο ή συγχρίνοντάς τες με προκαθορισμένους άτλαντες.
2. Θεραπείες και παρακολούθηση ασθενειών: η MIR συμβάλλει στην αποτελεσματικότερη θεραπεία των ασθενών κι επιπλέον, διαδραματίζει ζωτικό ρόλο στο σχεδιασμό και την παρακολούθηση θεραπειών, συμπεριλαμβανομένης της ακτινοθεραπείας, όπου η ακριβής ευθυγράμμιση εξασφαλίζει στοχευμένη χορήγηση θεραπείας, ελαχιστοποιώντας τη βλάβη σε υγιείς ιστούς. Παράδειγμα αποτελούν οι έρευνες των M. Foskey et al., οι οποίοι αναπτύσσουν μια μέθοδο για να καλύτερεύσουν την ακρίβεια του στόχου στην ακτινοθεραπεία (radiation therapy), βελτιστοποιούν την χρονική ευθυγράμμιση των εικόνων των ασθενών, ακόμα κι όταν αυτές υπόκεινται σε αρκετές αλλαγές λόγω ανατομίας [5]. Επιπλέον παράδειγμα αποτελούν οι έρευνες των van der Hoorn et al., οι οποίες προσεγγίζουν την ακρίβεια και την αξιοπιστία της MIR στην καταγραφή προόδου εγκεφαλικών καρκινωμάτων σε σχέση με το χρόνο [6].
3. Άτλας και κατάτμηση εικόνων: εκτός των άλλων, η MIR ενισχύει την απόδοση της κατάτμησης των εικόνων και υποστηρίζει τη δημιουργία άτλαντα. Στην κατάτμηση εικόνων, η εγγραφή βοηθά στην οριοθέτηση συγκεκριμένων δομών για λεπτομερή μελέτη ή σχεδιασμό θεραπείας. Η δημιουργία άτλαντα επωφελείται από την εγγραφή συντάσσοντας μεμονωμένες εικόνες σε ένα ολοκληρωμένο μοντέλο αναφοράς. Οι Lorenzen et al. παρουσιάζουν μεθοδολογία για την ευθυγράμμιση διαφορετικών τύπων ιατρικών εικόνων και τη δημιουργία άτλα [7]. Οι Gooya et al., παρουσιάζουν μια μεθοδολογία για την ευθυγράμμιση εικόνων γλοιώματος (τύπος καρκίνου στον εγκέφαλο) [8].

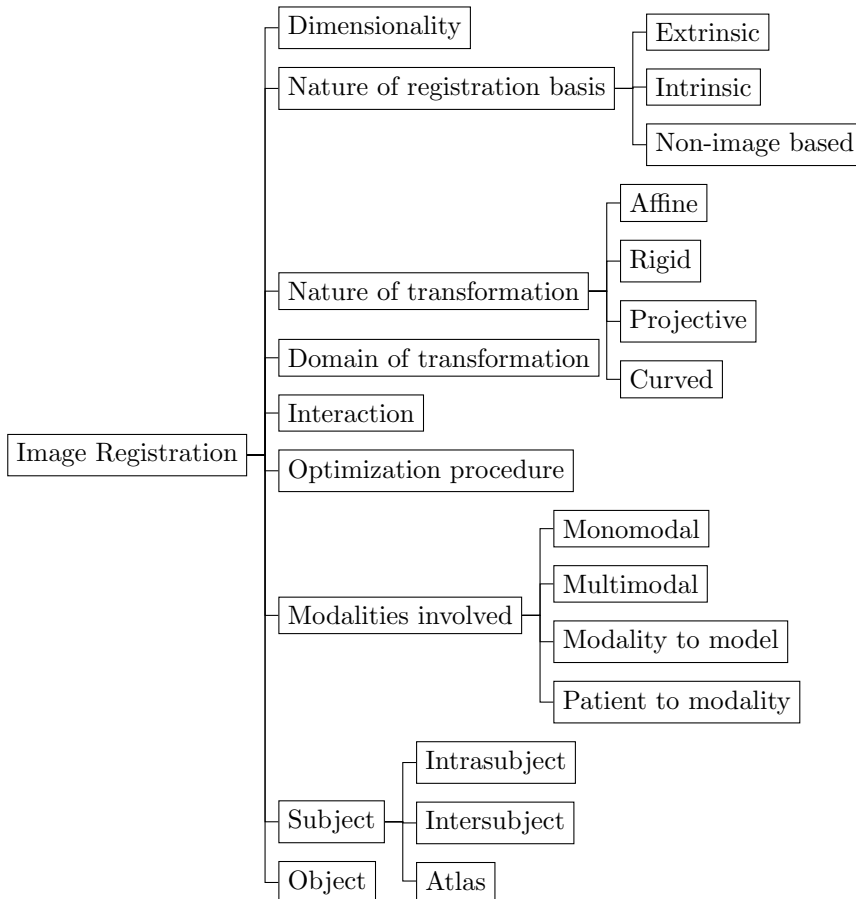
Συμπερασματικά, η ευθυγράμμιση ιατρικών εικόνων είναι θεμελιώδης για την προώθηση της εξατομικευμένης ιατρικής, βελτιώνοντας τα αποτελέσματα των ασθενών μέσω ακριβούς διάγνωσης και σχεδιασμού θεραπείας.

## 1.2 Ταξινόμηση της Ευθυγράμμισης Εικόνων (Image Registration - IR)

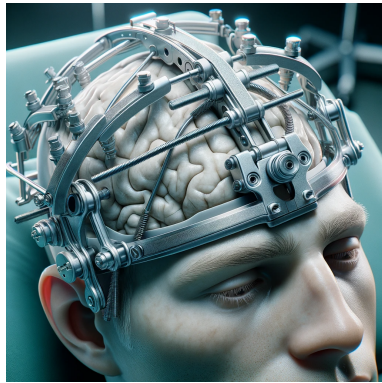
Η σύντηξη εικόνων μπορεί να ταξινομηθεί με πολλούς διαφορετικούς τρόπους. Στο παρόν κείμενο ακολουθείται η ταξινόμηση των Maintz and Viergever [14]. Η διαδικασία της ευθυγράμμισης μπορεί να αποδομηθεί σε τρία κύρια μέρη: το πρόβλημα (problem statement), τη φύση την ευθυγράμμισης (registration paradigm) και τη διαδικασία βελτιστοποίησης (optimization procedure). Ο συνδυασμός αυτών προσφέρει μια μοναδική κατάταξη



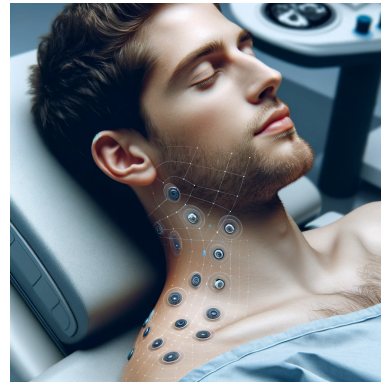
σύμφωνα με εννιά κριτήρια, όπως φαίνονται στην Εικόνα 1. Παρότι τα κριτήρια μεταξύ τους πιθανόν να εμπλέκονται ή να αλληλοεπηρεάζονται, μπορεί να διακριθεί ότι το πρόβλημα ορίζει την ταξινόμηση σύμφωνα με τα κριτήρια διαφορετικότητα στον τρόπο λήψης των εικόνων (modality), θέμα (subject), αντικείμενο (object), ενώ έχει άμεση σχέση με τα διαστασιοποίηση (dimensionality) και φύση τής βάσεως ευθυγράμμισης (nature of registration basis). Η φύση τής ευθυγράμμισης επηρεάζει τα κριτήρια φύση τής βάσεως ευθυγράμμισης, φύση μετασχηματισμού (nature of transformation), τομέας μετασχηματισμού (domain of transformation) και αλληλεπίδραση (interaction). Η διαδικασία βελτιστοποίησης επηρεάζει το κριτήριο αλληλεπίδρασης και ελέγχει τη διαδικασία βελτιστοποίησης (optimization procedure) [14].



Εικόνα 1: Ταξινόμηση της Ευθυγράμμισης Εικόνων [14].



(a) invasive



(b) non-invasive

Εικόνα 2: (a) Επεμβατική διαδικασία χρησιμοποιώντας ένα στερεοτακτικό σκελετό (stereotactic frame) τοποθετημένο στο κεφάλι ασθενούς. (b) Μη επεμβατική διαδικασία χρησιμοποιώντας επιδερμικά στίγματα (skin-mounted markers). Οι εικόνες δημιουργήθηκαν από το DALL-E image generation model του OpenAI.

Η κατηγορία διαστασιοποίηση αναφέρεται στις διαστάσεις των δύο εικόνων, ή στην ύπαρξη επιπλέον διάστασης, του χρόνου. Για παράδειγμα, η ευθυγράμμιση μπορεί να είναι  $3D-3D$  σε περιπτώσεις δύο συνόλων δεδομένων τομογραφίας λόγου χάρη, ή μπορεί να είναι χρονικές όπως για παράδειγμα η παρακολούθηση της ανάπτυξης των κοκκάλων σε παιδιά [14]. Η κατηγορία φύση τής βάσεως ευθυγράμμισης σχετίζεται με τη φύση τής ευθυγράμμισης και κατηγοριοποιείται ως εξωγενής (extrinsic), εσωγενής (intrinsic) και μη βασισμένη στην εικόνα (non - image based). Έτσι μια εξωγενής διαδικασία βασίζεται σε τεχνητά αντικείμενα που συνδέονται με τον ασθενή, αντικείμενα τα οποία έχουν σχεδιαστεί έτσι ώστε να είναι ευδιάκριτα και να εντοπίζονται με ακρίβεια σε όλες τις σχετικές απεικονιστικές μεθόδους. Μια εξωγενής διαδικασία μπορεί περαιτέρω να κατηγοριοποιηθεί σε επεμβατική (invasive) (πχ Εικόνα 2 (a)) και μη επεμβατική (non-invasive) (πχ Εικόνα 2 (b)). Μια εσωγενής διαδικασία εξαρτάται μόνο από εικόνες των ασθενών. Η ευθυγράμμιση μπορεί να βασίζεται σε ένα περιορισμένο σύνολο ταυτοποιημένων σημείων (landmarks), στην αντιστοίχιση τμηματοποιημένων δυαδικών δομών (segmentation based), ή σε μετρήσιμα που υπολογίζονται από την ένταση του γκρι στην εικόνα (voxel property based). Οι δύο πρώτες υποκατηγορίες (εξωγενής, εσωγενής) βασίζονται στην ύπαρξη εικόνας για την ευθυγράμμιση. Η τρίτη (μη βασισμένη στην εικόνα), ωστόσο, δε βασίζεται στις εικόνες, αλλά είναι δυνατή εάν τα συστήματα συντεταγμένων απεικόνισης των δύο σαρωτών έχουν κάπως βαθμονομηθεί μεταξύ τους. Αυτό, συνήθως, απαιτεί τους σαρωτές να βρίσκονται στην ίδια φυσική τοποθεσία και λαμβάνει δεδομένο ότι ο ασθενής παραμένει ακίνητος μεταξύ των δύο λήψεων. Δύσκολα εκπληρώνονται οι άνωθεν προϋποθέσεις σε οποιαδήποτε άλλη εφαρμογή εκτός αυτών που περιλαμβάνουν τη χρήση υπερήχων. Η επόμενη κατηγορία, φύση μετασχηματισμού, αφορά το πώς μετασχηματίζεται η μία εικόνα στην άλλη, δηλαδή καθορίζει το πώς τα σημεία στην κινούμενη εικόνα αντιστοιχίζονται στα σημεία στη σταθερή εικόνα. Ο διαχωρισμός σε αυτήν την κατηγορία έχει ως εξής:

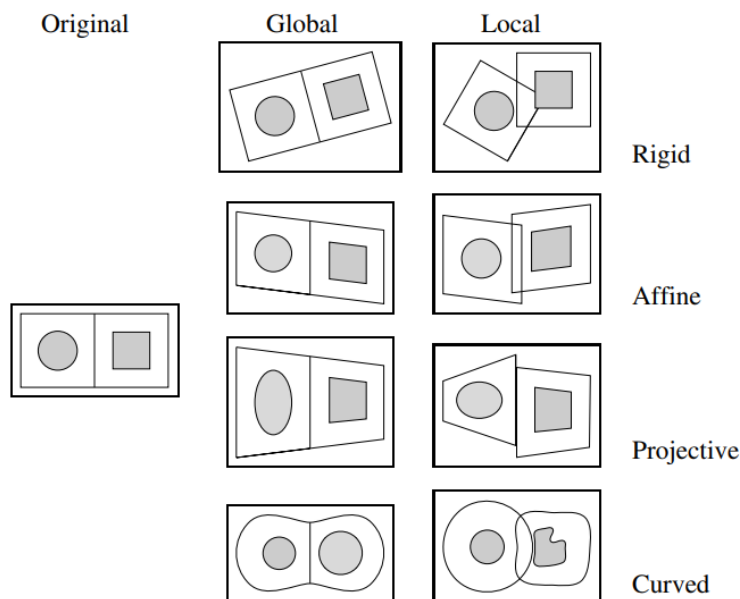
1. Άκαμπτος (rigid): γεωμετρικοί μετασχηματισμοί που διατηρούν όλες τις αποστάσεις. Αυτοί οι μετασχηματισμοί διατηρούν επίσης την ευθύτητα των γραμμών (και την επιπεδότητα των επιφανειών) καθώς και όλες τις γωνίες μεταξύ των ευθειών γραμμών. Υπάρχουν αρκετοί μέθοδοι για την επισήμανση των άκαμπτων μετασχηματισμών, κάθε μία από αυτές διεπόμενη από μια μεταφορά (translation) και μια περιστροφή (rotation). Η μεταφορά περιγράφεται ως ένα τρισδιάστατο διάνυσμα  $t$ , το οποίο προσδιορίζεται δίνοντας τις τρεις συντεταγμένες του  $t_x, t_y, t_z$  σε αντιστοιχία με τους  $x, y, z$  άξονες του Καρτεσιανού συστήματος ή δίνοντας το μήκος του και δύο γωνίες ώστε να προσδιοριστεί η κατεύθυνσή του στο πολικό σύστημα συντεταγμένων. Η περιστροφή παραμετροποιείται σε σχέση με τρεις γωνίες περιστροφής  $\theta_x, \theta_y, \theta_z$  (γωνίες Όιλερ) [15]. Στο πλαίσιο της ευθυγράμμισης εικόνων, ένας άκαμπτος μετασχηματισμός σημαίνει ότι η γενική δομή τής εικόνας θα μείνει αμετάβλητη. Τα σχήματα μέσα στην εικόνα μπορεί να μετακινηθούν ή να περιστραφούν, όμως δε θα παραμορφωθούν, χάνοντας τα βασικά γεωμετρικά χαρακτηριστικά των.

2. Αφινικός (affine): γεωμετρικός μετασχηματισμός που διατηρεί την ευθύτητα των γραμμών και την επιπεδότητα των επιφανειών καθώς και την παραλληλότητα. Επιτρέπει τις μη μηδενικές γωνίες μεταξύ των γραμμών να μεταβάλλονται. Αυτός ο μετασχηματισμός είναι κατάλληλος όταν κατά τη διάρκεια συλλογής των δεδομένων η εικόνα έχει εκταθεί (skewed), όπως για παράδειγμα, όταν η γωνία γκάντρι (gantry) της CT καταγράφεται λανθασμένα [15]. Όπως σε έναν άκαμπτο μετασχηματισμό, έτσι και σε έναν αφινικό η σχετική θέση των γραμμών που ήταν παράλληλες πριν το μετασχηματισμό διατηρείται, εντούτοις στον αφινικό μπορούν να αλλάξουν οι γωνίες κι οι αποστάσεις των σχημάτων.
3. Προοπτικός (projective): γεωμετρικοί μετασχηματισμοί που επιτρέπουν αλλαγή στην προοπτική. Απόρροια αυτού, η δυνατότητα να μοντελοποιούν τον τρόπο που μια εικόνα αλλάζει λόγω οπτικής γωνίας ή λόγω απόστασης παρατήρησης. Για παράδειγμα, οι παράλληλες ευθείες εμφανίζονται να συγκλίνουν σε μια εικόνα (λόγω προοπτικής βλ. Εικόνα 3). Αυτός ο μετασχηματισμός είναι σημαντικός, διότι δύναται να αντιπαρέλθει τα προβλήματα που δημιουργούνται όταν η εικόνα έχει αποκτηθεί από διαφορετική οπτική γωνία.
4. Καμπυλωτός (curved): γεωμετρικοί μετασχηματισμοί που δε διατηρούν την ευθύτητα των γραμμών [15]. Αυτοί οι μετασχηματισμοί αλλάζουν τοπικά συγκεκριμένα σημεία των εικόνων, επιτρέποντάς τους να τεταθούν, να λυγίσουν ή να παραμορφωθούν εν γένει. Χρησιμοποιούνται ευρέως στην ευθυγράμμιση ιατρικών εικόνων, λόγω του ότι ανατομικές δομές σε διαφορετικές εικόνες χρειάζεται να προσαρμοστούν για παραμορφώσεις (πχ εικόνες προγενέστερες μιας χειρουργικής επέμβασης με εικόνες μεταγενέστερης της χειρουργικής επέμβασης).



Εικόνα 3: Οπτική σύγκλιση γραμμών τρένου λόγω προοπτικής. Η εικόνα δημιουργήθηκε από το DALL-E image generation model του OpenAI.

Στην Εικόνα 4 φαίνονται και σχεδιαστικά οι διαφορετικές κατηγορίες μετασχηματισμών. Η κατηγορία τομέας μετασχηματισμού σχετίζεται με το κατά πόσο ο μετασχηματισμός αφορά ολόκληρη την εικόνα ή μέρος αυτής. Ένας μετασχηματισμός θεωρείται καθολικός (global), όταν εφαρμόζεται σε όλην την εικόνα, ενώ λέγεται τοπικός, όταν εφαρμόζεται σε κάθε κομμάτι της εικόνας (μπορεί το κάθε κομμάτι να μετασχηματίζεται διαφορετικά) [14]. Η αλληλεπίδραση σχετίζεται με το κατά πόσο αυτοματοποιημένη είναι η αλγοριθμική διαδικασία της ευθυγράμμισης. Υπάρχουν τρία επίπεδα αυτοματοποίησης, στο μεν πρώτο, αυτόματο (automatic), ο χρήστης επεμβαίνει μόνο για να τροφοδοτήσει τον αλγόριθμο με τις εικόνες και πιθανώς με πληροφορίες για τη συλλογή των. Στο δε δεύτερο, διαδραστικό (interactive), ο χρήστης εμπλέκεται στη διαδικασία κάνοντας την ευθυγράμμιση μόνος του, βοηθούμενος από κάποιο πρόγραμμα, το οποίο επιστρέφει μία εικόνα ή μια αριθμητική απεικόνιση του τρεχούμενου μετασχηματισμού και ίσως μια αρχική εκτίμηση για το μετασχηματισμό. Στο τρίτο επίπεδο, ημιαυτοματοποιημένο (semi - automatic), η αλληλεπίδραση που χρειάζεται παίρνει δύο διαφορετικές μορφές. Στην πρώτη ο χρήστης χρειάζεται να αρχικοποιήσει τον αλγόριθμο (πχ με την κατάτμηση των δεδομένων) ή χρειάζεται να κατευθύνει (πχ καταρρίπτοντας ή μη τις προτεινόμενες υποθέσεις ευθυγράμμισης). Οι εξωγενείς μέθοδοι είναι συνήθως αυτοματοποιημένες, παρ' όλα αυτά μπορεί να είναι ημιαυτοματοποιημένες (μερικές φορές οι χρήστες πρέπει να επιδείξουν την περιοχή με τα στίγματα). Οι εσωγενείς διαδικασίες είναι συνήθως ημιαυτοματοποιημένες όταν αφορούν κομμάτια ταυτοποιημένων σημείων (landmarks), βασισμένα στην κατάτμηση (segmentation-based), ενώ είναι αυτοματοποιημένες όταν πρόκει-



Εικόνα 4: Παραδείγματα 2D μετασχηματισμών.[14]

ται για βασισμένες στις ιδιότητες των ογκοστοιχείων (voxel property based) διαδικασίες [14]. Η διαδικασία βελτιστοποίησης σχετίζεται με τις διαδικασίες που ακολουθούνται για τη βελτιστοποίηση της ευθυγράμμισης. Οι παράμετροι που χρησιμοποιούνται στους μετασχηματισμούς, είτε υπολογίζονται άμεσα, για παράδειγμα ορίζονται αποκλειστικά από τα διαθέσιμα δεδομένα, είτε μπορούν να αναζητηθούν, δηλαδή να καθοριστούν βρίσκοντας το βέλτιστο κάποιας συνάρτησης ορισμένης στο χώρο των παραμέτρων. Οι υπολογιστικές μέθοδοι που χρησιμοποιούνται για την εύρεση γενικευμένων μετασχηματισμών, περιορίζονται κυρίως σε εφαρμογές που βασίζονται σε σποραδική πληροφόρηση. Στον αντίποδα, μπορεί να υπολογιστεί η τοπική μετατόπιση απευθείας από τα διαθέσιμα δεδομένα της εικόνας. Η άλλη περίπτωση των μεθόδων βελτιστοποίησης αναζητά τη βελτιστοποίηση μιας μαθηματικής συνάρτησης, η οποία έχει οριστεί στη φάση του σχεδιασμού του προβλήματος βελτιστοποίησης [14]. Μερικές γνωστές τεχνικές είναι οι downhill simplex [16], Newton–Raphson επανάληψη [17], gradient descent μέθοδοι [18] κ.ά. Οι τρόποι λήψεως αναφέρονται στους διαφορετικούς τύπους των τεχνικών ευθυγράμμισης που χρησιμοποιούνται για να παράξουν τις εικόνες. Κάθε τρόπος προσφέρει ένα μοναδικό τρόπο οπτικοποίησης και κατανόησης των δομών, τη λειτουργία ή την παθολογία των ιστών και των οργάνων μέσα στο σώμα. Η ευθυγράμμιση συνήθως περιλαμβάνει την αντιστοίχιση των εικόνων από τον έναν τρόπο στον άλλον, με σκοπό το συνδυασμό των πληροφοριών που παρέχει το καθένα. Ενυπάρχουν τέσσερις κατηγορίες, μονοτροπική (monomodal), πολυτροπική (multimodal), τρόπος λήψης σε μοντέλο (modality to model), ασθενής σε τρόπο λήψης (patient to modality) [14]. Στη μονοτροπική, οι εικόνες που πρέπει να ευθυγραμμιστούν ανήκουν στον ίδιο τρόπο λήψης, τουναντίον στην πολυτροπική οι εικόνες ανήκουν σε δύο διαφορετικούς τρόπους λήψης. Στις εναπομένουσες περιπτώσεις μόνο μία εικόνα εμπλέκεται και ο άλλος "τρόπος λήψης" είναι ο ίδιος ο ασθενής ή ένα μοντέλο. Ένα παράδειγμα για τη τρόπο λήψης σε μοντέλο είναι η ευθυγράμμιση μίας MR εγκεφαλικής εικόνας σε ένα μαθηματικοποιημένο μοντέλο διαχωρισμού (compartmental) των εγκεφαλικών δομών. Στην ακτινοθεραπεία ο ασθενής μπορεί να τοποθετηθεί με τη βοήθεια της ευθυγράμμισης εικόνων της εσω-θέσης του προσομοιωτή ακτίνων X (in-position X - Ray simulator images) σε μία ανατομική εικόνα προθεραπείας. Η διαδικασία τοποθέτησης ασθενούς είναι παράδειγμα ευθυγράμμισης ασθενής σε τρόπο λήψης [14].

Το θέμα χωρίζεται σε εσωθεματική ευθυγράμμιση (intrasubject) όταν όλες οι εικόνες που εμπλέκονται στη διαδικασία της ευθυγράμμισης έχουν παρθεί από τον ίδιο ασθενή. Αν η διαδικασία ευθυγράμμισης επιτυγχάνεται με δύο εικόνες δύο διαφορετικών ασθενών (ή ασθενούς και μοντέλου), τότε πρόκειται για διαθεματική (intersubject) ευθυγράμμιση.

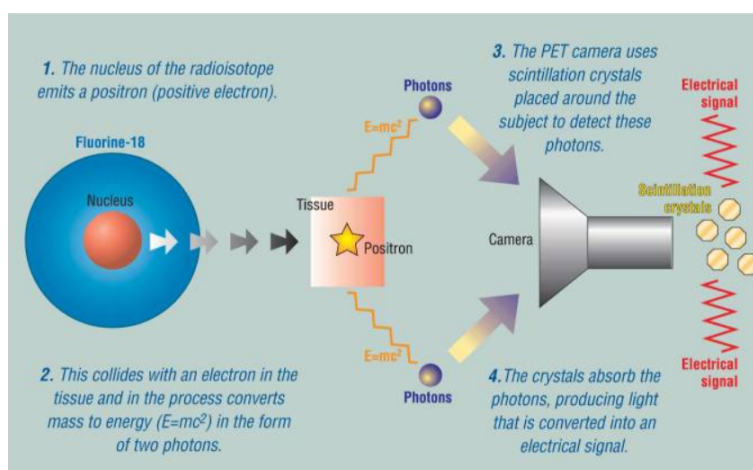
Μονοτροπική	Πολυτροπική	τρόπος λήψης σε μοντέλο	Ασθενής σε τρόπο λήψης
Auto-radiographic	CT-MR	CT	CT
CT	CT-PET	MR	MR
MR	CT-SPECT	SPECT	PET
Portal	DSA-MR	X-RAY	Portal

Πίνακας 1: Παραδείγματα τεχνικών για κάθε ένα τύπο τρόπου λήψης [14].

Αν η μία εικόνα χρησιμοποιείται από έναν ασθενή κι η άλλη εξάγεται χρησιμοποιώντας κάποια βάση δεδομένων που έχει δημιουργηθεί από πολλαπλούς ασθενείς, τότε η ευθυγράμμιση γίνεται με άτλα [14]. Τελευταία κατηγορία το αντικείμενο, η οποία αναφέρεται στο κομμάτι του σώματος που εξετάζει η ευθυγράμμιση. Αυτό μπορεί να είναι κάποιο μέρος στο κεφάλι, στο θώρακα, στην κοιλιακή χώρα, στη σπονδυλική στήλη κλπ [14].

### 1.3 Τεχνική Ποζιτρονικής Τομογραφίας (Positron emission tomography - PET)

Η Ποζιτρονική Τομογραφία είναι μια τεχνική, η οποία μετράει τη φυσιολογική λειτουργία του ασθενούς ελέγχοντας τη ροή του αίματος, το μεταβολισμό, τους νευροδιαβιβαστές και τα ραδιοσημασμένα φάρμακα. Η PET προσφέρει ποσοτικές αναλύσεις, επιτρέποντας έτσι, τις αλλαγές που γίνονται σε μια χρονική περίοδο να παρακολουθούνται (για παράδειγμα την εξέλιξη μιας νόσου ή την αντίδραση σε ένα συγκεκριμένο ερέθισμα. Η τεχνική βασίζεται στην ανίχνευση της ραδιενέργειας που εκπέμπεται μετά την έγχυση μικρής ποσότητας ενός ραδιενεργού δείκτη σε περιφερειακή φλέβα. Ο δείκτης χορηγείται ως ενδοφλέβια έγχυση, συνήθως σηματοδοτημένη με οξυγόνο-15, φθόριο-18, άνθρακα-11, ή αζώτο-13. Η συνολική ραδιενεργή δόση είναι παρόμοια με τη δόση που χρησιμοποιείται στην αξονική τομογραφία. Μια εξέταση PET χρειάζεται δέκα με σαράντα λεπτά για να ολοκληρωθεί [49].

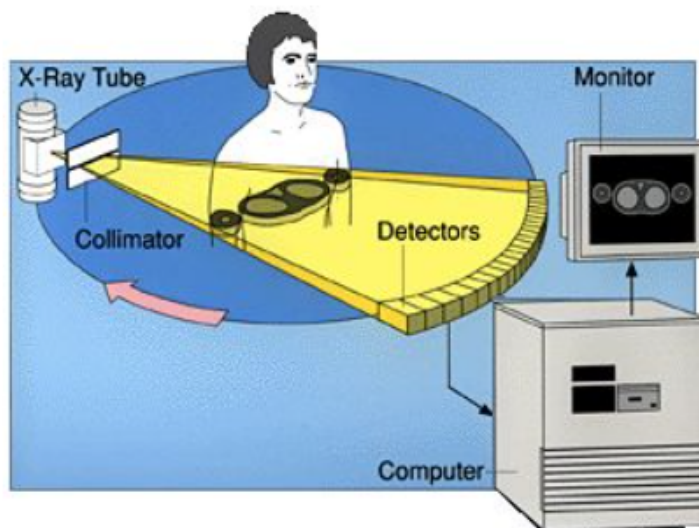


Εικόνα 5: Διάγραμμα λειτουργίας Pet. Ο πυρήνας του ακτινοϊσοτροπικού στοιχείου εκπέμπει ένα ποζιτρόνιο (θετικό ηλεκτρόνιο). Το ποζιτρόνιο συγκρούεται με ένα ηλεκτρόνιο στον ιστό του σώματος και κατά τη σύγκρουση μετασχηματίζει τη μάζα σε ενέργεια με τη μορφή δύο φωτονίων. Η κάμερα PET χρησιμοποιεί κρυστάλλους απόσβεσης τοποθετημένους γύρω από το αντικείμενο για να εντοπίζει αυτά τα φωτόνια. Τέλος, οι κρυστάλλοι απορροφούν τα φωτόνια, παράγοντας φως, το οποίο μετατρέπεται σε ηλεκτρικό σήμα [49].

### 1.4 Τεχνική Αξονικής Τομογραφίας (Computed tomography - CT)

Η αξονική τομογραφία (CT) είναι μια τεχνική, η οποία λειτουργεί με τη χρήση X-ακτίνων για τη δημιουργία λεπτομερών εικόνων διατομής του σώματος. Ένας ακτινολογικός σωλήνας εκπέμπει X-ακτίνες, περνώντας από το σώμα προς έναν ανιχνευτή στο αντίθετο μέρος. Το άνοιγμα για τον ασθενή είναι εξήντα με εβδομήντα εκατοστά σε διάμετρο. Στο εσωτερικό ενός CT βρίσκεται ένα περιστρεφόμενο κάδρο, στο οποίο από τη μία πλευρά τοποθετείται μια σωλήνα ακτίνων X και από την απέναντι, ένας ανιχνευτής σε σχήμα μπανάνας. Μια βεντάλια δέσμης ακτίνων X

δημιουργείται, καθώς το περιστρεφόμενο κάδρο κινεί τη σωλήνα ακτίνων X και τον ανιχνευτή γύρω από τον ασθενή. Κάθε φορά που η σωλήνα ακτίνων X και ο ανιχνευτής κάνουν μια περιστροφή 360 μοιρών, λαμβάνεται μια εικόνα ή "τομή". Αυτή η τομή επικεντρώνεται σε ένα πάχος ενός έως δέκα χιλιοστών, χρησιμοποιώντας μολύβδινα κλείστρα εμπροσθεν τού σωλήνα ακτίνων X και του ανιχνευτή ακτίνων X. Καθώς ο σωλήνας κι ο ανιχνευτής κάνουν αυτήν την περιστροφή, ο ανιχνευτής καταγράφει πολλαπλά στιγμιότυπα (προφίλ) τής ακτίνας X. Κάθε προφίλ διαιρείται χωρικά (διαιρείται σε τμήματα) από τους ανιχνευτές και τροφοδοτείται σε περίπου 700 μεμονωμένα κανάλια. Κάθε προφίλ στη συνέχεια ανακατασκευάζεται αντίστροφα (ή "προβάλλεται πίσω") από έναν ειδικό υπολογιστή σε μια δισδιάστατη εικόνα τής τομής που σαρώθηκε [50].

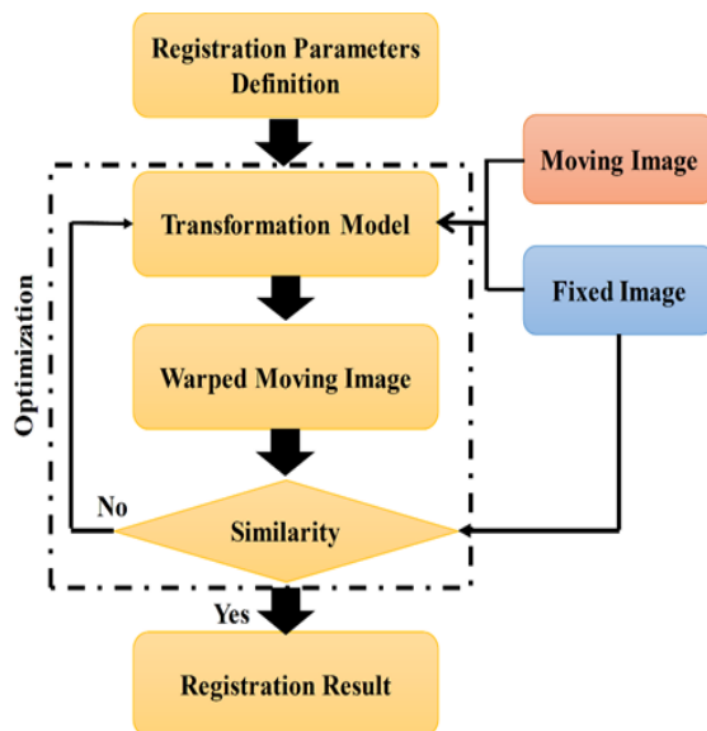


Εικόνα 6: Διάγραμμα λειτουργίας CT [50].

## 1.5 MIR και Βαθιά Μάθηση (Deep Learning - DL)

Μέχρι προσφάτως, η πλειοψηφία τής ευθυγράμμισης των εικόνων εφαρμοζόταν χειροκίνητα από τους ιατρούς, κάτι που εν δυνάμει μπορούσε να προβεί κλινικά καθοριστικό για την ποιότητα ορισμένων MIR. Η αυτοματοποίηση της διαδικασίας MIR υιοθετήθηκε για την αποφυγή σκοπέλων, ενώ η εντρυφήση κι αναδιαμόρφωση του ερευνητικού πεδίου ήρθε ως επακόλουθο της ανάπτυξης της Βαθιάς Μάθησης (Deep Learning -DL) [2]. Η συνήθης τεχνική ευθυγράμμισης εικόνων είναι μια επαναληπτική διαδικασία όπως φαίνεται στην Εικόνα 7 [9]. Η διαδικασία έχει ως είσοδο δύο τύπους εικόνων, την κινούμενη (moving) και τη σταθερή (fixed). Η κάλλιστη ευθυγράμμιση δύναται να επιτευχθεί μετακινώντας την κινούμενη εικόνα πάνω στη σταθερή επαναλαμβανόμενα. Επιλέγοντας ένα μέτρο ομοιότητας, αρχικά, διευκρινίζεται ο βαθμός εφαρμογής τής κινούμενης εικόνας στη σταθερή εικόνα σε κάθε επανάληψη. Έπειτα, προκειμένου να υπολογιστεί η νέα θέση τής κινούμενης εικόνας σε σχέση με τη σταθερή, χρησιμοποιείται ένας μηχανισμός βελτιστοποίησης των παραμέτρων. Η διαδικασία συνεχίζεται έως ότου επιτευχθεί η απαραίτητη ευθυγράμμιση (βάσει κριτηρίων) [2],[19].

Η MIR όσον αφορά το DL μπορεί να κατηγοριοποιηθεί με πολλαπλούς τρόπους, βάσει την αρχιτεκτονική των δικτύων (CNN, RL, GAN κλπ), με βάση τη διαδικασία μάθησης (επιβλεπόμενη, μη επιβλεπόμενη κλπ), τύποι πρόβλεψης (inference types) πχ επαναληπτικός ή άπαξ, το μέγεθος των εικόνων που λειτουργούν ως είσοδος (κομμάτια, ολόκληρη), με βάση τον τύπο εξόδου κά. Η κατηγοριοποίηση που γίνεται στην παρούσα διπλωματική για το DL είναι σύμφωνη με την έρευνα των Yabo et al. και φαίνεται παρακάτω [11].



Εικόνα 7: Διάγραμμα ροής ευθυγράμμισης εικόνων [2].

- Μέθοδοι Ενισχυτικής Μάθησης (Reinforcement learning based methods): η ενισχυτική μάθηση (RL) είναι ένας τύπος μηχανικής μάθησης που επικεντρώνεται στην πρόβλεψη των καλύτερων ενεργειών που πρέπει να ληφθούν δεδομένης της τρέχουσας κατάστασης σε ένα περιβάλλον. Συνήθως μοντελοποιείται ως μια Μαρκοβιανή διαδικασία απόφασης, χρησιμοποιώντας ένα σύνολο περιβαλλοντικών καταστάσεων και ενεργειών. Εκπαιδεύει έναν τεχνητό πράκτορα, ώστε να μεγιστοποιεί τις συνολικές αναμενόμενες ανταμοιβές του, ενώ η διαδικασία εκπαίδευσης συχνά περιλαμβάνει έναν συμβιβασμό μεταξύ εξερεύνησης και εκμετάλλευσης. Η εξερεύνηση σημαίνει την αναζήτηση σε ολόκληρο τον χώρο για τη συλλογή περισσότερων πληροφοριών, ενώ η εκμετάλλευση σημαίνει την αναζήτηση στις υποσχόμενες περιοχές βάσει των τρεχουσών πληροφοριών. Η Q-learning είναι ένας μη μοντελοποιημένος αλγόριθμος RL, που στοχεύει να μάθει μια συνάρτηση Q, η οποία μοντελοποιεί τη σχέση δράσης-ανταμοιβής. Η συνάρτηση Bellman συνήθως χρησιμοποιείται στη Q μάθηση για τον υπολογισμό της ανταμοιβής. Η εξίσωση Bellman υπολογίζει τη μέγιστη μελλοντική ανταμοιβή ως την άμεση ανταμοιβή που λαμβάνει ο πράκτορας εισερχόμενος στην τρέχουσα κατάσταση συν μια σταθμισμένη μέγιστη μελλοντική ανταμοιβή για την επόμενη κατάσταση. Για την επεξεργασία εικόνας, η συνάρτηση Q μοντελοποιείται συχνά ως CNN, η οποία μπορεί να κωδικοποιήσει τις εικόνες εισόδου ως καταστάσεις και να μάθει τη συνάρτηση Q μέσω επιβλεπόμενης εκπαίδευσης.
- Μέθοδοι Βαθιάς Ομοιότητας (Deep similarity based methods): οι συμβατικοί δείκτες ομοιότητας βασισμένοι στην ένταση περιλαμβάνουν την απόσταση αθροίσματος τετραγώνων (SSD), τη μέση τετραγωνική απόσταση (MSD), την (κανονικοποιημένη) διασταυρούμενη συσχέτιση (CC), και την (κανονικοποιημένη) αμοιβαία πληροφορία (MI). Γενικά, οι συμβατικοί δείκτες ομοιότητας λειτουργούν αρκετά καλά για την εγγραφή εικόνας μονοτροπικού (monomodal) τύπου, όπου το ζεύγος εικόνων μοιράζεται την ίδια κατανομή έντασης, όπως στην εγγραφή εικόνας CT-CT, MR-MR. Ωστόσο, ο θόρυβος και οι παραμορφώσεις σε εικόνες, που συνοδεύουν κάποιες τεχνικές εξαγωγής εικόνων ασθενών, συχνά επηρεάζουν τις συμβατικές μεθόδους ομοιότητας, με αποτέλεσμα να μην αντιπροσωπεύουν το πραγματικό αποτέλεσμα, ακόμη και στην εγγραφή εικόνας μονοτροπικού τύπου. Επίσης, μετρικές όπως οι SSD και MSD δε λειτουργούν για την εγγραφή εικόνας πολυμορφικού (multimodal) τύπου. Προκειμένου να καλυφθεί το κενό που αφήνουν οι προηγούμενες

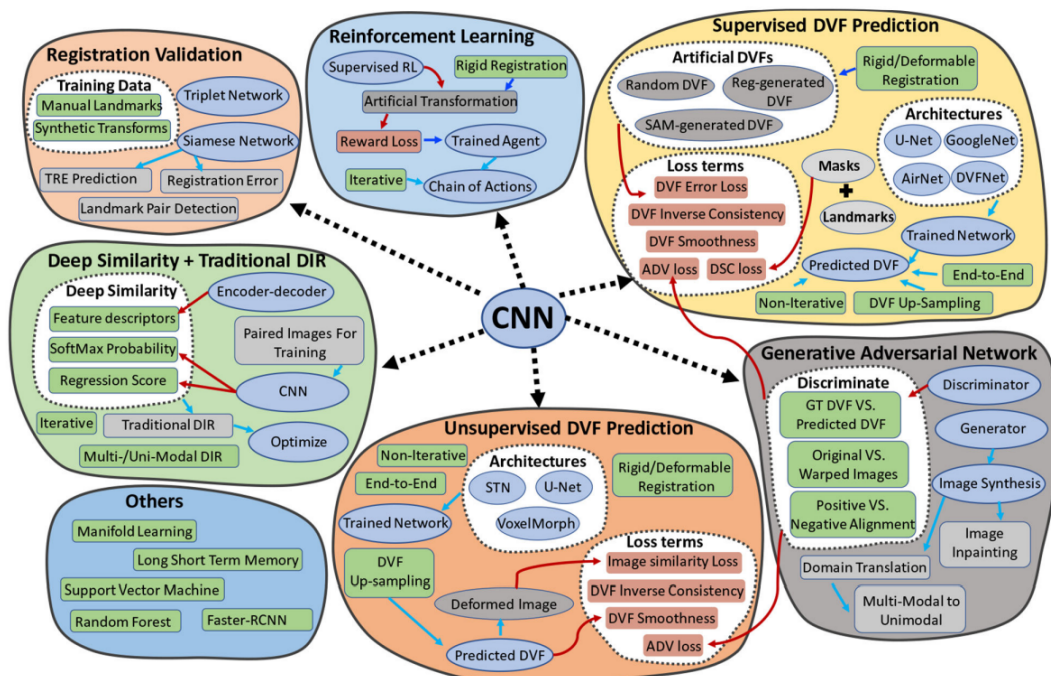
τεχνικές, προτάθηκαν καινούργιες μετρικές όπως η αμοιβαία πληροφορία (Mutual Information - MI). Λόγω της επιτυχίας των Συνελικτικών Νευρικών Δικτύων σε προβλήματα ταξινόμησης και κατάτμησης, έγινε μια απόπειρα να χρησιμοποιηθούν μέθοδοι βαθιάς ομοιότητας ως μετρικές ομοιότητας. Οι μέθοδοι βασισμένοι σε βαθιά ομοιότητα για την ευθυγράμμιση εικόνων αξιοποιούν νευρωνικά δίκτυα για να μάθουν περίπλοκους πίνακες αντιστοίχισης μεταξύ εικόνων, επικεντρώνοντας στον συγχρονισμό τους με βάση μετρήσεις ομοιότητας που έχουν μάθει. Αυτές οι προσεγγίσεις έχουν δείξει εντυπωσιακή επιτυχία σε διάφορες εφαρμογές ιατρικής απεικόνισης, προσφέροντας βελτιωμένη ακρίβεια και ανθεκτικότητα σε σύγκριση με τις παραδοσιακές μεθόδους.

- **Πρόβλεψη Επιβλεπόμενου Μετασχηματισμού (Supervised transformation prediction):** η πρόβλεψη επιβλεπόμενου μετασχηματισμού στη μηχανική μάθηση περιλαμβάνει την εκπαίδευση ενός μοντέλου με επισημασμένα σύνολα δεδομένων, όπου είναι γνωστές οι σωστές μετατροπές μεταξύ των εισόδων και των επιθυμητών εξόδων. Αν και η απευθείας πρόβλεψη μετατροπής δεν έχει ακόμα υπερτερήσει τις κορυφαίες παραδοσιακές μεθόδους DIR, η ακρίβεια εγγραφής έχει βελτιωθεί σημαντικά. Μερικές μέθοδοι έχουν επιτύχει συγκρίσιμη ακρίβεια εγγραφής με τις παραδοσιακές μεθόδους DIR. Η δημιουργία δεδομένων αληθείας θα συνεχίσει να παίζει σημαντικό ρόλο στην εκπαίδευση του δικτύου. Οι περιορισμοί της χρήσης τεχνητά δημιουργημένων ζευγών εικόνων με γνωστούς μετασχηματισμούς δεδομένων αλήθειας, περιλαμβάνουν τη μη αντανάκλαση της πραγματικής φυσιολογικής κίνησης, την ανεπάρκεια κάλυψης του ευρέως φάσματος των παραλλαγών πραγματικών σεναρίων ευθυγράμμισης εικόνων και τη διαφορά των τεχνητά δημιουργημένων ζευγών εικόνων στο στάδιο εκπαίδευσης από το πραγματικό ζευγάρι εικόνων στο στάδιο επαλήθευσης. Για τον πρώτο περιορισμό, μπορούν να χρησιμοποιηθούν διάφορα μοντέλα παραγωγής μετασχηματισμών. Για το δεύτερο περιορισμό, μπορούν να επαυξηθούν τα δεδομένα. Για τον τρίτο περιορισμό, μπορεί να προσαρμοστεί ο χώρος της ευθυγράμμισης, ώστε να συνάδει με το χώρο λήψης.
- **Πρόβλεψη Μη Επιβλεπόμενου Μετασχηματισμού (Unsupervised transformation prediction):** η συγκεκριμένη μέθοδος δε χρειάζεται την ύπαρξη δεδομένων αληθείας. Σε σύγκριση με την επιβλεπόμενη πρόβλεψη μετασχηματισμού, οι μη επιβλεπόμενες μέθοδοι αντιπαρέχονται αποτελεσματικά το πρόβλημα της έλλειψης συνόλων δεδομένων εκπαίδευσης. Έχουν προταθεί διάφοροι όροι κανονικοποίησης για την ενθάρρυνση πιθανών προβλέψεων μετασχηματισμού. Οι περισσότερες από αυτές τις μεθόδους επικεντρώνεται στη μονοτροπική (unimodal) ευθυγράμμιση. Με σκοπό την επιπλέον επίβλεψη, σε κάποιες περιπτώσεις συνδυάζονται οι δύο μέθοδοι.
- **Αναγεννητικά Ανταγωνιστικά Δίκτυα σε MIR (GAN in MIR):** αλγόριθμοι μηχανικής μάθησης, όπου δύο νευρωνικά δίκτυα, ένας παραγωγός και ένας διακριτής, ανταγωνίζονται μεταξύ τους για να παράγουν όσο το δυνατόν πιο ρεαλιστικά δείγματα δεδομένων. Χρησιμοποιούνται ευρέως στην παραγωγή νέων εικόνων, τη μεταφορά στυλ εικόνας και άλλες εφαρμογές όπου η παραγωγή νέων, πιστευτών δειγμάτων από ένα σύνολο δεδομένων είναι επιθυμητή. Τα GAN έχουν δείξει ότι είναι υποσχόμενα στην ευθυγράμμιση ιατρικών εικόνων μέσω είτε νέων αντίθετων κοστών είτε μεταφράσεων πεδίων εικόνας. Παρέχουν κανονικοποιήσεις βασισμένες σε δίκτυο που συμπληρώνουν τους παραδοσιακούς όρους κανονικοποίησης και μετατρέπουν την πολυτροπική εγγραφή σε μονοτροπική, διευκολύνοντας την εφαρμογή υπαρχόντων αλγορίθμων. Ωστόσο, η ακρίβεια απεικόνισης της απόλυτης έντασης των GAN τίθεται υπό διερεύνηση.
- **Ευθυγράμμιση και επαλήθευση χρησιμοποιώντας Βαθιά Μάθηση:** η απόδοση της εγγραφής εικόνων μπορεί να αξιολογηθεί με μετρικές ομοιότητας εικόνας, όπως η SSD, NCC και MI. Ωστόσο, αυτές οι μετρικές αξιολογούν μόνο τη γενική ευθυγράμμιση σε ολόκληρη την εικόνα. Για βαθύτερη κατανόηση της τοπικής ακρίβειας εγγραφής, συνήθως γίνεται επιλογή χειροκίνητων ζευγών σημαντικών σημείων, η οποία όμως είναι χρονοβόρα, υποκειμενική και επιρρεπής σε σφάλματα, ιδιαίτερα όταν πρέπει να επιλεγούν πολλά σημαντικά σημεία. Το ενδιαφέρον για τη χρήση Βαθιάς Μάθησης (DL) για την αξιολόγηση εγγραφής εικόνας αυξήθηκε σημαντικά το 2019, με τις περισσότερες μελέτες να αντιμετωπίζουν την πρόβλεψη σφάλματος εγγραφής ως πρόβλημα εποπτευόμενης παλινδρόμησης, επικεντρώνοντας κυρίως στην εγγραφή πνεύμονα λόγω διαθέσιμων σετ δεδομένων όπως το DIRLAB.



- Άλλες τεχνικές βασισμένες στη μάθηση: εκτός από αυτές που ήδη αναφέρθηκαν, υπάρχουν κι άλλες τεχνικές. Μερικές από αυτές είναι το LSTM (Long Short-Term Memory) -ένας τύπος αρχιτεκτονικής επαναλαμβανόμενων νευρωνικών δικτύων που σχεδιάστηκε για την επίλυση προβλημάτων πρόβλεψης ακολουθιών κι μεταφορική μάθηση (transfer learning), η οποία είναι μια μέθοδος μηχανικής μάθησης, όπου ένα μοντέλο που αναπτύχθηκε για ένα σκοπό χρησιμοποιείται ως αφετηρία για ένα μοντέλο σε ένα δεύτερο σκοπό.

Στην Εικόνα 8 φαίνεται μια συμπερίληψη όλων των MIR τεχνικών βάσει τη Βαθιά Μάθηση με τα αντίστοιχα στοιχεία τους. Οι μέθοδοι αυτές αποτελούν κομβικό σημείο στην προηγμένη επεξεργασία και ανάλυση ιατρικών εικόνων, βοηθώντας στην ακριβή ευθυγράμμιση και σύγκριση εικόνων από διάφορες ιατρικές εξετάσεις. Κάθε κατηγορία αντιπροσωπεύει διαφορετικές τεχνικές και προσεγγίσεις, συμβάλλοντας στην ακρίβεια, την αποδοτικότητα για την ολοκληρωμένη ανάλυση ιατρικών δεδομένων. Τα Συνελκτικικά Νευρωνικά Δίκτυα (Convolutional Neural Networks - CNN) σχεδιάστηκαν αρχικά για να επεξεργάζονται σύνθετα σύνολα δεδομένων, όπως εικόνες, που συνήθως εκφράζονται μέσω τακτικής δειγματοληψίας σημείων δεδομένων. Για αυτόν το λόγο, σχεδόν όλες οι αναφερόμενες μέθοδοι χρησιμοποιούν συνελκτικούς πυρήνες στον σχεδιασμό τους για Βαθιά Μάθηση. Αυτό εξηγεί γιατί το μοντέλο CNN βρίσκεται στο κέντρο της Εικόνας 8 [11].



Εικόνα 8: Σύνοψη των επτά κατηγοριών μεθόδων βασισμένων στη Βαθιά Μάθηση για τη MIR.[11]

### 1.5.1 Μη Επιβλεπόμενη Μάθηση (Unsupervised Learning)

Η μη επιβλεπόμενη μάθηση είναι ένα είδος τεχνητής μάθησης, όπου ο αλγόριθμος μαθαίνει μοτίβα από μη κατηγοριοποιημένα δεδομένα. Στη μη επιβλεπόμενη μάθηση η μηχανή δέχεται είσοδο  $x_1, x_2, \dots$ , χωρίς κάποια επιθυμητή έξοδο (επιβλεπόμενη μάθηση) ή επιβράβευση από το περιβάλλον (ενισχυτική μάθηση) [12]. Αντιθέτως, αναγνωρίζει ομοιότητες στα δεδομένα και αντιδρά βάσει παρουσίας ή μη αυτών των ομοιοτήτων σε κάθε δεδομένο. Η συγκεκριμένη μέθοδος είναι εφαρμόσιμη σε περιπτώσεις που έχουν πολλά δεδομένα αλλά όχι ξεκάθαρη εικόνα του τι μοτίβα ή συσχετίσεις μπορεί να υπάρχουν. Κυρίως χρησιμοποιείται στην ομαδοποίηση (clustering) και σε προβλήματα μείωσης διαστάσεων (dimensionality reduction) [12]. Η μη επιβλεπόμενη μάθηση προτιμάται έναντι των επιβλεπόμενης και ενισχυτικής μάθησης, όταν ο στόχος είναι να ερευνηθούν τα ενυπάρχοντα μοτίβα κι οι δομές στα δεδομένα ή όταν η επιθυμητή έξοδος δεν είναι πρότερα γνωστή.

### 1.5.2 Μη Επιβλεπόμενη Μάθηση στην MIR

Η μη επιβλεπόμενη μάθηση είναι ιδιαίτερα ταιριαστή σε κομμάτια ευθυγράμμισης εικόνων, αναλογιζόμενοι την έλλειψη των δεδομένων εκπαίδευσης (training datasets) [11]. Μετά τη δημοσίευση του Δικτυακού Μηχανισμού Χωρικής Μετατροπής (Spatial Transformation Network - STN) [13], το οποίο επιτρέπει τον υπολογισμό του κόστους (ομοιότητα εικόνων - image similarity) μέσα στη διαδικασία της εκπαίδευσης, όλο και περισσότερες έρευνες έχουν δημοσιευθεί σχετικά με τη μη επιβλεπόμενη εκπαίδευση στη MIR. Ένα τυπικό μη επιβλεπόμενο δίκτυο σε MIR δέχεται ως είσοδο ένα ζευγάρι εικόνων και έχει ως έξοδο ένα διανυσματικό πεδίο μετασχηματισμού (Deformation Vector Field - SVF), το οποίο εφαρμόζεται στην κινούμενη εικόνα για την προσομοίωση της σταθερής. Η νέα εικόνα συγκρίνεται με την σταθερή για να υπολογιστεί το κόστος (loss) [11]. Υπάρχει ολοένα και μεγαλύτερο ενδιαφέρον για τον τομέα της μη επιβλεπόμενης μάθησης όσον αφορά την ευθυγράμμιση ιατρικών εικόνων. Αρκετοί αναλυτές έχουν πετύχει επίπεδα ακρίβειας ίσα ή υψηλότερα με αυτά των παραδοσιακών μεθόδων [2]. Η μη επιβλεπόμενη πολυτροπική ευθυγράμμιση εικόνων έχει λάβει λιγότερη προσοχή στο παρελθόν και ως εκ τούτου απαιτεί περισσότερη μελέτη [11],[2]. Δεδομένου ότι δεν παρέχονται ανατομικά περιγράμματα ή αναφορές DVF, η μη επιβλεπόμενη MIR είναι εγγενώς δυσκολότερη από την επιβλεπόμενη. Προκειμένου να επιτευχθεί αποτελεσματικότερη μάθηση, απαιτούνται επιπλέον βήματα πέρα από την βασική εκπαίδευση δικτύου. Για παράδειγμα, χρησιμοποιήθηκε άκαμπτη ευθυγράμμιση πριν από τη χρήση του αλγορίθμου DL για τη μείωση των εντάσεων κίνησης στις ιατρικές εικόνες, ενώ δυαδικές μάσκες χρησιμοποιήθηκαν από άλλους ερευνητές για να επικεντρωθούν στις περιοχές ενδιαφέροντος. Η προεπεξεργασία των εικόνων μπορεί να βοηθήσει στην ακρίβεια της MIR, αλλά μπορεί επίσης να περιπλέξει την εκπαίδευση και να περιορίσει τη γενίκευση του μοντέλου. Τέλος, οι μη επιβλεπόμενες προσεγγίσεις για MIR είναι απλές στην εκμάθηση και αρκετά υποσχόμενες ως προς την ακρίβεια των αποτελεσμάτων τους. Ως εκ τούτου, αναμένεται να αυξηθεί το ερευνητικό ενδιαφέρον σε αυτόν τον τομέα [2].

## 1.6 Μετρικές ομοιότητας/αξιολόγησης και MIR

Ένα μοντέλο κρίνεται για την αποτελεσματικότητά του και την αποδοτικότητά του κατά τη διαδικασία της εκπαίδευσης σύμφωνα με κάποιες μετρικές αξιολόγησης. Στην ευθυγράμμιση των εικόνων, ένα κόστος ομοιότητας υπολογίζει την ομοιότητα ή την ανομοιότητα μεταξύ δυο εικόνων, μετρώντας την απόσταση μεταξύ των δύο εικόνων ή μεταξύ των χαρακτηριστικών των δύο εικόνων. Όσο πιο κοντά είναι το κόστος στο μηδέν, τόσο καλύτερα ευθυγραμμισμένες είναι οι εικόνες. Στις περισσότερες περιπτώσεις, οι αλγόριθμοι ευθυγράμμισης σχηματίζονται ως ένα πρόβλημα ελαχιστοποίησης [28]. Τέσσερις από τις μετρικές ομοιότητας που απαντώνται στη βιβλιογραφία είναι [27]:

1. Μέσο Τετραγωνικό Σφάλμα (Mean squared error - MSE): υπολογίζεται μεγιστοποιώντας την αρνητική λογαριθμική πιθανότητα (log-likelihood) μιας Γκαουσιανής κανονική κατανομής. Σε ένα πλαίσιο διακριτών σημείων  $p$  από ένα χώρο  $\Omega$  το MSE υπολογίζεται ως: [27]

$$\text{MSE}(I \circ \Phi, J) = \frac{1}{|\Omega|} \sum_{p \in \Omega} \|(I \circ \Phi)(p) - J(p)\|^2 \quad (1)$$

Τα  $I, J$  αναπαριστούν δύο εικόνες, όπου συνήθως το  $I$  αναφέρεται στη κινούμενη εικόνα, ενώ το  $J$  αναφέρεται στη σταθερή εικόνα. Το  $\Phi$  αναφέρεται στο χωρικό μετασχηματισμό που εφαρμόζεται στην κινούμενη εικόνα. Ο συμβολισμός  $I \circ \Phi$  αναπαριστά την εφαρμογή του μετασχηματισμού στην εικόνα  $I$ , δίνοντας τη νέα εικόνα που θα πρέπει να συνταυριάζει στην εικόνα  $J$ . Το  $p$  αναπαριστά τις θέσεις των εικονοστοιχείων (pixel) στο χώρο των εικόνων. Το  $\Omega$  είναι ο χώρος των εικόνων κι ειδικά όλο το σύνολο των θέσεων των εικονοστοιχείων στις δύο εικόνες  $I, J$ . Το  $|\Omega|$  είναι ο συνολικός αριθμός των εικονοστοιχείων στο  $\Omega$ , τα οποία χρησιμοποιούνται για να βρουν το μέσο όρο των τετραγωνισμένων σφαλμάτων. Ο όρος  $\|(I \circ \Phi)(p) - J(p)\|^2$  υπολογίζει την τετραγωνισμένη διαφορά στην ένταση μεταξύ των αντίστοιχων εικονοστοιχείων  $p$  στην μετασχηματισμένη εικόνα και στη σταθερή εικόνα. Συνεπώς, η μετρική MSE μετρά πόσο καλά η μετασχηματισμένη εικόνα αντιστοιχίζεται στη σταθερή εικόνα στόχο, βρίσκοντας το μέσο των τετραγωνικών διαφορών στις εντάσεις των εικονοστοιχείων σε ολόκληρο το όρο  $\Omega$ . Στόχος είναι να ελαχιστοποιηθεί αυτό το σφάλμα βρίσκοντας

τον κατάλληλο μετασχηματισμό  $\Phi$ .

2. Κανονικοποιημένη Συνάρτηση Διασυσχέτισης (Normalized cross correlation - NCC): συγκρίνει μια εικόνα πρότυπο (window) με μια εικόνα στόχο (target), μετακινώντας το πρότυπο πάνω από το στόχο κι υπολογίζοντας το συσχετισμό (correlation coefficient) για κάθε θέση. Ο συσχετισμός αυτός είναι ένα κανονικοποιημένο μέτρο, το οποίο ποσοτικοποιεί μια γραμμική σχέση μεταξύ μοτίβων έντασης στο πρότυπο και στο στόχο για μια επικαλυπτόμενη περιοχή. Ο παράγοντας κανονικοποίησης περιορίζει το συσχετισμό μεταξύ των τιμών -1 (τέλεια αρνητική γραμμική συσχέτιση) και 1 (τέλεια θετική γραμμική συσχέτιση) και με το 0 να δείχνει πως δεν υπάρχει συσχέτιση μεταξύ των δύο.[29] Το NCC είναι λιγότερο ευαίσθητο σε διαφοροποιήσεις τής φωτεινότητας και της αντίθεσης (contrast), κάτι το οποίο το κάνει δημοφιλές στις εικόνες που έχουν αποκτηθεί με διαφορετικές τεχνικές και πρωτόκολλα, ή και σε εικόνες με διαφορετικούς τρόπους λήψης. Για δύο κομμάτια (patches)  $A, B$  τα οποία αναπαριστώνται ως διανύσματα στήλης μεγέθους  $N$  με μέσες τιμές κομματιών  $\bar{A}, \bar{B}$  και απόκλιση  $\sigma_A^2, \sigma_B^2$  το NCC ορίζεται ως: [27]

$$\text{NCC}_{patch}(A, B) = \sum_{n=1}^N \frac{(A_n - \bar{A})(B_n - \bar{B})}{\sigma_A \sigma_B} \quad (2)$$

Έπειτα, όλες οι ομοιότητες των κομματιών που υπολογίστηκαν εξάγονται ως μέσος όρος στην εικόνα:

$$\text{NCC}(I \circ \Phi, J) = \frac{1}{|\Omega|} \sum_{p \in \Omega} \text{NCC}_{patch}(I_p \circ \Phi, J_p) \quad (3)$$

Το  $\text{NCC}_{patch}(A, B)$  δείχνει πόσο όμοια είναι τα δύο κομμάτια  $A, B$ . Τα  $A_n, B_n$  αντιπροσωπεύουν τις τιμές των εντάσεων των  $n$ -οστών εικονοστοιχείων στα αντίστοιχα κομμάτια. Τα  $\bar{A}, \bar{B}$  είναι οι μέσες τιμές εντάσεων όλων των εικονοστοιχείων στα  $A, B$ . Τα  $\sigma_A, \sigma_B$  είναι οι τυπικές αποκλίσεις των τιμών των εντάσεων στα  $A, B$ . Αυτές υπολογίζουν τον καταμερισμό των τιμών γύρω από τη μέση τιμή, υπογραμμίζοντας πόση είναι η απόκλιση από το μέσο όρο. Για κάθε εικονοστοιχείο μέσα στο κομμάτι υπολογίζεται το γινόμενο των διαφορών μεταξύ της έντασης κάθε εικονοστοιχείου και τη μέση τιμή τού κομματιού που ανήκει. Αυτό κανονικοποιείται και η διαδικασία επαναλαμβάνεται για όλα τα εικονοστοιχεία.

3. Κανονικοποιημένη Αμοιβαία Πληροφορία (Normalized mutual information -NMI): μοντελοποιεί η πιθανοκρατική συσχέτιση μεταξύ των εντάσεων των ογκοστοιχείων (voxel) των εικόνων. Είναι κατάλληλο για εφαρμογές, όπου δεν υφίσταται γραμμική συσχέτιση μεταξύ των εντάσεων των εικόνων, ιδανικό για πολυτροπικές περιπτώσεις [27]. Η σχέση μεταξύ μετασχηματισμένης εικόνας και σταθερής εικόνας είναι:

$$\text{NMI}(I \circ \Phi) = \frac{H(I \circ \Phi) + H(J)}{I(I \circ \Phi)} \quad (4)$$

,όπου η περιθωριακή εντροπία (marginal entropy)  $H(I) = - \int_{\mathbb{R}} p_I(x) \log(p_I(x)) dx$  και η αμοιβαία πληροφορία (mutual information)  $I(x, y) = - \int_{\mathbb{R}} p_{IJ}(x, y) \log(p_{IJ}(x, y)) dx dy$ . Για να υπολογιστεί η κοινή πιθανότητα (joint probability)  $p_{IJ}(x, y)$  και τα περιθώρια  $p_I(x), p_J(y)$ , οι κατανομές τής έντασης και στις δύο εικόνες προσεγγίζονται από ιστογράμματα, κάνοντάς τα μη διαφορίσιμα. Προκειμένου να χρησιμοποιηθούν σε DL εφαρμογές, χρησιμοποιείται το παράθυρο Parzen με Γκαουσιανούς πυρήνες για να προσεγγισθούν οι κατανομές. Δεδομένης μιας συνάρτησης παραθύρου Parzen  $w$ , το ενιαίο ιστόγραμμα για διακριτές τιμές  $x, y \in \mathbb{R}$  ορίζεται ως:

$$h_{IJ}(x, y) = \sum_{p \in \Omega} w((I \circ \Phi)(p) - x) w(J(p) - y) \quad (5)$$

, όπου η  $p_{IJ}(x, y)$  λαμβάνεται με κανονικοποίηση και τα  $p_{IJ}(x, y)$  με περιθωριοποίηση [27]. Προκειμένου να γίνει πιο κατανοητή η συγκεκριμένη μετρική ομοιότητας το επόμενο παράδειγμα παρατίθεται. Έστω, ότι δύο κομμάτια παζλ από διαφορετικά παζλ, τα οποία πρέπει να ευθυγραμμιστούν. Κάθε κομμάτι έχει μοναδικό μοτίβο (κατανομή έντασης). Η εντροπία ( $H$ ) μπορεί να εκφραστεί ως το πόσο περίπλοκο είναι το μοτίβο

τού κάθε κομματιού παζλ. Αν ένα κομμάτι απεικονίζει έναν μπλε ουρανό, τότε το κομμάτι αυτό, έχει χαμηλή εντροπία, τουναντίον αν ένα κομμάτι έχει ένα λεπτομερές τοπίο πόλης, τότε το κομμάτι αυτό είναι περισσότερο περίπλοκο και για αυτόν το λόγο έχει μεγαλύτερη εντροπία. Η αμοιβαία πληροφορία μετράει πόσο καλά τα μοτίβα στα δύο κομμάτια εφαρμόζουν. Εάν αντιστοιχιστεί ένα σύννεφο στον ουρανό σε δύο διαφορετικά παζλ, η αμοιβαία πληροφορία ποσοτικοποιεί το κατά πόσο το μοτίβο του σύννεφου στο ένα παζλ μπορεί να εξηγήσει το μοτίβο του σύννεφου στο άλλο παζλ. Η κανονικοποιημένη αμοιβαία πληροφορία προσαρμόζει αυτό το αντιστοιχισθέν αποτέλεσμα λαμβάνοντας υπόψη την πολυπλοκότητα του μοτίβου κάθε κομματιού, προτείνοντας ένα δίκαιο τρόπο σύγκρισης μεταξύ των διαφορετικών κομματιών παζλ. Τέλος, χρησιμοποιώντας τα ιστογράμματα και το παράθυρο Parzen τα κομμάτια παζλ σπάνε σε μικρότερα μέρη κι η σύγκριση γίνεται πάνω σε αυτά τα μέρη. Κάτι τέτοιο, επιτρέπει η σύγκριση να γίνει ομαλότερη κι ευπροσάρμοστη, καθιστώντας την συγκρίσιμη με τον τρόπο που μαθαίνουν οι υπολογιστές να αντιστοιχίζουν εικόνες.

4. Συντελεστής Dice (Dice Coefficient): η συγκεκριμένη μετρική κυμαίνεται μεταξύ 0 και 1, με το 1 να δεικνύει ότι οι δύο εικόνες είναι πανομοιότυπες και το μηδέν να σημαίνει ότι δεν υπάρχει επικάλυψη. Ορίζεται ως:

$$\text{Dice} = 2 * \frac{|A \cap B|}{(|A| + |B|)} \quad (6)$$

,όπου το  $|A|, |B|$  είναι ο αριθμός των στοιχείων στα σύνολα A,B αντίστοιχα. Το  $|A \cap B|$  αναπαριστά τον αριθμό των στοιχείων που υπάρχουν και στα δύο σύνολα. Ο Συντελεστής Dice δεν υπολογίζει άμεσα τη χωρική ακρίβεια των ευθυγραμμισμένων διανυσμάτων του πίνακα μετασχηματισμού. Χρησιμοποιείται κυρίως ως συμπληρωματική μετρική.

## 1.7 Βιβλιογραφική Ανασκόπηση

Χάρη στην πρόοδο των μοντέλων Βαθιάς Μάθησης, πολλοί ερευνητές έχουν επικεντρώσει τις μελέτες τους στο αντικείμενο της ευθυγράμμισης ιατρικών εικόνων, με στόχο να δώσουν λύσεις σε αυτόν τον τομέα τής ιατρικής.

Οι Fu et al., στο άρθρο τους "Biomechanically constrained non-rigid MR-TRUS prostate registration using deep learning based 3D point cloud matching" [53] παρουσιάζουν ένα πλαίσιο άκαμπτης ευθυγράμμισης εικόνων για παρεμβάσεις προστάτη. Στο άρθρο ενσωματώνονται τα συνελκτικά νευρωνικά δίκτυα για την κατάτμηση των εικόνων μαγνητικής τομογραφίας τού προστάτη (Magnetic Resonance - MR) και των υπερηχογραφήματων προστάτη (TransRectal UltraSound - TRUS), με ένα δίκτυο αντιστοίχισης βασισμένο σε τρισδιάστατα σύννεφα σημείων (3D point cloud network). Το σύστημα εκπαιδεύεται με ένα πεδίο παραμόρφωσης που προέρχεται από ανάλυση πεπερασμένων στοιχείων, ενσωματώνοντας βιομηχανικούς περιορισμούς. Η μέθοδος αξιολογείται χρησιμοποιώντας δεδομένα από πενήντα ασθενείς, με επίκεντρο την ακρίβεια ευθυγράμμισης των σχημάτων τού προστάτη μετά την εγγραφή. Η ακρίβεια ποσοτικοποιείται χρησιμοποιώντας διάφορες μετρήσεις, όπως ο δείκτης ομοιότητας Dice, η μέση απόσταση επιφάνειας (mean surface distance - MSD) κι η απόσταση Hausdorff, με το δίκτυο να εμφανίζει καλή απόδοση σε αυτές τις μετρήσεις. Επιπλέον, δοκιμάζεται η ανθεκτικότητα της μεθόδου στο θόρυβο του σύννεφου σημείων, δείχνοντας το δυναμικό αυτής της προσέγγισης σε κλινικές εφαρμογές για παρεμβάσεις που απαιτούν ακριβή ευθυγράμμιση εικόνων τού προστάτη από διαφορετικές μορφές λήψης. Τα αποτελέσματα υποδεικνύουν, ότι η προτεινόμενη μέθοδος μπορεί να εκτελέσει γρήγορα και αξιόπιστα την εγγραφή εικόνας, κάτι το οποίο είναι κρίσιμο για στοχευμένες θεραπείες του προστάτη.

Οι Fu et al., αναπτύσσουν μια μεθοδολογία Βαθιάς Μάθησης στην έρευνά τους "LungRegNet: An unsupervised deformable image registration method for 4D-CT lung" [54] με σκοπό τη γρήγορη, ακριβή κι ευέλικτη ευθυγράμμιση 4D εικόνων πνεύμονα αξονικής τομογραφίας. Η μέθοδος αυτή είναι μη επιβλεπόμενη κι ονομάζεται LungRegNet. Αποτελείται από δύο επιμέρους δίκτυα, τα CoarseNet (προβλέπει μεγάλες κινήσεις των πνευμόνων σε μια ευρύτερη κλίμακα) και FineNet (προβλέπει τοπικές κινήσεις των πνευμόνων σε μια τοπική κλίμακα). Οι γενικευμένοι πίνακες παραμόρφωσης προβλέπονται από το CoarseNet και χρησιμοποιούνται για να παραμορφώσουν

τα αρχικά κινούμενα κομμάτια της εικόνας χρησιμοποιώντας έναν χωρικό μετασχηματιστή. Η μετασχηματισμένη εικόνα συγκρίνεται, έπειτα, με τη σταθερή για τον υπολογισμό της ανομοιότητας μεταξύ των. Τόσο το CoarseNet, όσο και το FineNet περιλαμβάνουν επίσης ένα γεννήτορα (generator), ο οποίος εκπαιδεύεται για να προβλέπει απευθείας τους πίνακες μετασχηματισμού. Το CoarseNet και το FineNet περιλαμβάνουν και τα δύο έναν διακριτή (discriminator), ο οποίος εκπαιδεύεται χρησιμοποιώντας την παραμορφωμένη εικόνα και τη σταθερή εικόνα για την παραγωγή αντίθετων κοστών για επιπλέον κανονικοποίηση. Πρώτα εκπαιδεύεται το CoarseNet για να παράξει τη μετασχηματισμένη εικόνα, η οποία εν συνεχεία περνάει στο FineNet για εκπαίδευση. Προκειμένου να αυξηθεί η ακρίβεια εγγραφής του LungRegNet, δημιουργούνται εικόνες με ενισχυμένα αγγεία παράγοντας χάρτες πιθανότητας της πνευμονικής αγγειοδομής πριν από την πρόβλεψη του δικτύου. Τα δεδομένα αξιολογούνται με την τεχνική five-fold cross validation κι ελέγχονται σε διαφορετικά σύνολα δεδομένων χρησιμοποιώντας το TRE. Τα αποτελέσματα είναι ενθαρρυντικά, καθώς η προτεινόμενη μέθοδος πετυχαίνει καλύτερη ακρίβεια ευθυγράμμισης όσον αφορά το TRE, σε σύγκριση με άλλες μεθόδους Βαθιάς Μάθησης ελεγμένες στο ίδιο σύνολο δεδομένων.

Οι de Vos et al., παρουσιάζουν στην έρευνά τους "A Deep Learning Framework for Unsupervised Affine and Deformable Image Registration" [55] ένα μοντέλο Βαθιάς Μάθησης για μη επιβλεπόμενη αφινική (affine) και μετασχηματική (deformable) εγγραφή εικόνων. Με σκοπό να επιτύχουν μια ακριβή και γρήγορη ευθυγράμμιση, αναπτύσσουν ένα πλαίσιο, ονομαζόμενο Deep Learning Image Registration (DLIR), το οποίο χρησιμοποιεί τα δίκτυα ConvNet. Η συγκεκριμένη μεθοδολογία, αφορά τρεις τύπους ευθυγράμμισης, τον αφινικό, τον μετασχηματικό και το συνδυασμό τους ώστε να επιτευχθεί μία πολυεπίπεδη ευθυγράμμιση. Στην πρώτη περίπτωση, το ConvNet αναλύει τις κινούμενες και τις σταθερές εικόνες σε διαφορετικές ακολουθίες κι επομένως εξαλείφουν την ανάγκη για ψαλίδισμα και γέμισμα, ώστε να επιτευχθεί το ίδιο μέγεθος. Σε κάθε ακολουθία ο εξαγόμενος πίνακας χαρακτηριστικών έχει διαφορετικό μέγεθος. Έτσι, εφαρμόζεται ένα καθολική μέση συγκέντρωση (global average pooling) για να παραχθεί ένα χαρακτηριστικό ανά πίνακα χαρακτηριστικών, ως μέσος όρος των δύο πινάκων. Η συγκεκριμένη τεχνική εξαναγκάζει το δίκτυο να κωδικοποιήσει τις διευθύνσεις και τους αφινικούς μετασχηματισμούς γενικότερα. Έτσι, το δίκτυο μπορεί να συνδεθεί με ένα νευρωνικό δίκτυο, το οποίο θα αποκωδικοποιήσει τις σχετικές θέσεις των κινούμενων και των σταθερών εικόνων και θα τις μετατρέψει σε δώδεκα αφινικές παραμέτρους μετασχηματισμού. Στη δεύτερη περίπτωση, το ConvNet δέχεται ένα ζευγάρι σταθερών και κινούμενων εικόνων ίσου μεγέθους και τις συγχωνεύει. Έπειτα, εφαρμόζει υποβάθμιση (downsampling), βασισμένου στην επιλογή του χρήστη για το διάστημα του πλέγματος των B-spline καθορίζει την τιμή του υποβιβασμού. Ακολούθως, εφαρμόζονται αλληλάλληλα συνελκτικά επίπεδα και το τελευταίο προβλέπει τα B-spline σημεία ελέγχου. Το πεδίο μετασχηματισμού παράγεται από αυτά τα σημεία. Στην τρίτη περίπτωση, πολλαπλά επίπεδα ConvNet εφαρμόζονται το ένα μετά το άλλο, το κάθε ένα με το δικό του τρόπο ευθυγράμμισης. Κάθε επίπεδο εκπαιδεύεται για το συγκεκριμένο, δικό του τρόπο ευθυγράμμισης, διατηρώντας τα βάρη των ConvNet των προηγούμενων επιπέδων. Μετά την εκπαίδευση, το πολυεπίπεδο ConvNet χρησιμοποιείται ως ένα μοναδικό ConvNet. Τα δεδομένα που χρησιμοποιούνται για την ευθυγράμμιση, αφορούν μαγνητικές τομογραφίες για την κίνηση της καρδιάς κι αξονικές τομογραφίες τού στήθους. Το προτεινόμενο μοντέλο είναι συγκρίσιμο με τις συμβατικές μεθόδους ευθυγράμμισης, ενώ είναι αρκετά πιο γρήγορο.

Στο άρθρο των Zhao et al., "Unsupervised 3D End-to-End Medical Image Registration With Volume Tweening Network" [56] προτείνεται μία τεχνική για μη επιβλεπόμενη ευθυγράμμιση ιατρικών εικόνων το Volume Tweening Network (VTN). Με βάση αυτό, περιγράφονται τρία δομικά μέρη. Το πρώτο πραγματεύεται ένα σχήμα καταληκτικής αλληλουχίας που επιλύει μεγάλες μετατοπίσεις. Το δεύτερο μια αποδοτική ολοκλήρωση δικτύου εγγραφής αφινικής μετατόπισης. Το τρίτο ένα επιπρόσθετο κόστος αντιστρεψιμότητας, που ενθαρρύνει την προς τα πίσω συνέπεια. Το VTN, λοιπόν, αποτελείται από αρκετά υποδίκτυα ευθυγράμμισης έπειτα από κάθε ένα η κινούμενη εικόνα μετασχηματίζεται. Η μη επιβλεπόμενη εκπαίδευση των παραμέτρων τού δικτύου καθοδηγείται από τη διαφορά μεταξύ της σταθερής εικόνας και κάθε μίας από τις παραμορφωμένες εικόνες, με το κόστος κανονικοποίησης στις ροές που προβλέπονται από τα υποδίκτυα. Η διαδικασία τού μετασχηματισμού της εικόνας είναι διαφορίσιμη τόσο στην εισαγόμενη εικόνα, όσο και στο πεδίο ροής εισόδου λόγω της τριγραμμικής παρεμβολής. Αυτές οι λειτουργίες

παραμόρφωσης μπορούν να μεταδώσουν πίσω τις κλίσεις (back-propagate gradients) σε όλα τα προηγούμενα υποδίκτυα, κάτι που βαίνει κρίσιμο για την εκμάθηση από άκρο σε άκρο των διαδοχικών δικτύων. Τέλος, στο δίκτυο έχει συμπεριληφθεί ένα αρχικό επίπεδο ως το πρώτο υποδίκτυο, το οποίο εφαρμόζει έναν εύκαμπτο μετασχηματισμό. Τα αποτελέσματα της εν λόγω τεχνικής εξετάζονται σε σύνολα δεδομένων εικόνων για το συκώτι που έχουν παρθεί μέσα από την αξονική τομογραφία και σύνολα δεδομένων εικόνων για τον εγκέφαλο που έχουν παρθεί από μαγνητική τομογραφία. Αυτά δείχνουν ότι ο αλγόριθμος αυτός είναι ταχύτερος από τις παραδοσιακές μεθόδους βασισμένες σε βελτιστοποίηση και επιτυγχάνει την καλύτερη δυνατή απόδοση στην εγγραφή ιατρικών εικόνων.

Οι Estienne et al., στο άρθρο τους με τίτλο "U-ResNet: Ultimate Coupling of Registration and Segmentation with Deep Nets" [57] προτείνουν ένα τρισδιάστατο νευρωνικό δίκτυο, το U-ResNet, το οποίο χρησιμεύει τόσο σε προβλήματα ευθυγράμμισης, όσο και σε προβλήματα κατάτμησης εικόνων. Το δίκτυο αυτό, αξιοποιεί δύο διακλαδώσεις με κοινές παραμέτρους. Κατά τη διάρκεια της εκπαίδευσης, το νευρωνικό χρησιμοποιεί ως είσοδο μια κινούμενη εικόνα και μια εικόνα αναφοράς, ενώ οι στόχοι είναι οι αντίστοιχες κατατμημένες μάσκες. Οι εικόνες αρχικά συγχωνεύονται και περνούν μέσα από συνελικτικά επίπεδα, ακολουθούμενα από μια συνάρτηση ενεργοποίησης. Το πρώτο κομμάτι μειώνει τις διαστάσεις εξάγοντας χαρακτηριστικά, ενώ το επόμενο κομμάτι, το οποίο χρησιμοποιεί αντίστροφα συνελικτικά δίκτυα αναδομεί την εικόνα. Το πρώτο αυτό δίκτυο, εξάγει ένα πίνακα αφινικού μετασχηματισμού και έναν πίνακα παραμορφωσιμότητας (deformable) μετασχηματισμού. Οι δύο αυτοί πίνακες εφαρμόζονται από κοινού στην αρχική εικόνα. Παράλληλα με αυτήν τη διαδικασία, εξάγονται οι μάσκες της κινούμενης εικόνας, χρησιμοποιώντας τον ίδιο κωδικοποιητή και εφαρμόζοντας ένα συμμετρικό αποκωδικοποιητή. Τέλος, εξάγονται οι μάσκες και για τα δύο αποτελέσματα (ευθυγράμμισης - κατάτμησης). Τα δεδομένα που χρησιμεύουν στην αξιολόγηση του μοντέλου αποτελούν εικόνες εγκέφαλου μαγνητικής τομογραφίας T1. Ο αλγόριθμος προβάλλει κάποια αρκούντως θετικά στοιχεία, με κύριο χαρακτηριστικό τη λιτή αρχιτεκτονική του, η οποία τον καθιστά εύκολα εκπαιδευσιμο με λιγότερα δεδομένα.

## 1.8 Unet

Το μοντέλο Unet προτάθηκε από τους Olaf Ronneberger, Philipp Fischer, και Thomas Brox το 2015 [20]. Λόγω της απόδοσής του στην κατάτμηση βιοϊατρικών εικόνων απέκτησε αρκετό ενδιαφέρον. Τη σημαντικότητά του στην ιατρική εξετάζουν οι Azad et al., οι οποίοι παρουσιάζουν μια εκτεταμένη έρευνα πάνω στο Unet και τις εξελικτικές μορφές του, τις ταξινομούν και πραγματοποιούν μια συγκριτική ανάλυση αναμεταξύ των [22]. Παρ' όλη την εξέλιξη στον τομέα της ευθυγράμμισης μέσω DL, το μοντέλο, συνεχίζει και πρωτεύει στο χώρο. Στο άρθρο των Jia et al. "U-Net vs Transformer: Is U-Net Outdated in Medical Image Registration?" οι ερευνητές δοκίμασαν δύο παραλλαγές του μοντέλου Unet και τις συνέκριναν με το μοντέλο TransMorph σε δύο διαφορετικά σύνολα δεδομένων, καταλήγοντας στο συμπέρασμα ότι το Unet ξεπερνά σε απόδοση τα μοντέλα που έχουν ως βάση τους αρχιτεκτονική transformer στα κομμάτια της εσωθεματικής (inter-subject) ευθυγράμμισης και της ευθυγράμμισης βασισμένης σε άτλα (atlas-based) 3D ιατρικών εικόνων [21].

### 1.8.1 Βασικές διαδικασίες μοντέλων

Προτού αναπτυχθεί η αρχιτεκτονική του μοντέλου Unet, εξηγούνται βασικές διαδικασίες των μοντέλων. Θα γίνει αναφορά στα επίπεδα συνέλιξης, στο βήμα μετακίνησης του πυρήνα, στο πλήθος των πυρήνων, στην προσαύξηση της εικόνας, στο αντίστροφο επίπεδο συνέλιξης, και στο επίπεδο συγκέντρωσης (pooling).

Ένα συνελικτικό επίπεδο (convolution layer) στο πλαίσιο της Βαθιάς Μάθησης είναι ένα θεμελιώδες μπλοκ, το οποίο χρησιμοποιεί τη συνέλιξη. Μεταξύ συναρτήσεων μπορούν να γίνουν πράξεις, όπως η πρόσθεση και ο πολλαπλασιασμός. Η συνέλιξη είναι ακόμα μια μαθηματική πράξη, η οποία περιγράφεται ως εξής:

$$(f * g)(t) := \int_{-\infty}^{\infty} f(x) \cdot g(t - x) dx \quad (7)$$

,όπου  $f$  είναι η είσοδος (input) και  $g$  ο πυρήνας (kernel).

#### Παράδειγμα 1.7.1

Έστω η ρίψη δύο ζαριών κι έστω ότι η πιθανότητα να εμφανιστεί μία από τις έξι πλευρές και για τα δύο ζάρια, δεν είναι ισοπίθανη. Η πιθανότητα για την εμφάνιση κάθε αριθμού με βάση τον συνδυασμό των ζαριών υπολογίζεται εάν αθροιστούν όλα τα γινόμενα πιθανοτήτων το ένα ζάρι να είναι αυτός ο αριθμός με την πιθανότητα του άλλου ζαριού να είναι αυτός ο αριθμός. Έτσι, λόγω χάρη η πιθανότητα να εμφανιστεί το πέντε υπολογίζεται ως [23]:

$$P(\text{ζάρι 1} + \text{ζάρι 2} = 5) = \alpha_1 \cdot b_4 + \alpha_2 \cdot b_3 + \alpha_3 \cdot b_2 + \alpha_4 \cdot b_1 \quad (8)$$

,όπου  $\alpha_1, \alpha_2, \alpha_3, \alpha_4$  είναι οι πιθανότητες να εμφανιστούν οι πλευρές 1,2,3,4 για το ζάρι 1 και  $b_4, b_3, b_2, b_1$  είναι οι πιθανότητες να εμφανιστούν οι πλευρές 1,2,3,4 για το ζάρι 2.

Επομένως, για τα δυο σύνολα αριθμών που δίνουν όλες οι πλευρές των ζαριών,  $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6$  και  $b_1, b_2, b_3, b_4, b_5, b_6$  η συνέλιξη μεταξύ των ορίζεται για κάθε πιθανό αριθμό:

$$(a * b)_n = \sum_{i+j=n} a_i \cdot b_j \quad (9)$$

ενώ το αποτέλεσμα της συνέλιξης, είναι το σύνολο των αποτελεσμάτων που προκύπτουν για κάθε πιθανό συνδυασμό.

#### Παράδειγμα 1.7.2

Έστω δύο σύνολα αριθμών (1, 2, 3) και (4, 5, 6). Η συνέλιξή τους δίνει:

$$(1, 2, 3) * (4, 5, 6) = (1 \cdot 4, 1 \cdot 5 + 2 \cdot 4, 1 \cdot 6 + 2 \cdot 5 + 3 \cdot 4, 2 \cdot 6 + 3 \cdot 5, 3 \cdot 6) = (4, 13, 28, 27, 18) \quad (10)$$

Όσον αφορά τα Νευρωνικά Δίκτυα, σε ένα συνελικτικό επίπεδο χρησιμοποιείται ένας πυρήνας (kernel), ο οποίος παράγει έναν πίνακα χαρακτηριστικών και χρησιμοποιείται σε εφαρμογές όπως η ταξινόμηση, η κατάτμηση κ.ά. Ο πυρήνας που χρησιμοποιείται έχει διαστάσεις, που μπορούν να βελτιστοποιηθούν για καλύτερη απόδοση.

#### Παράδειγμα 1.7.3

Έστω εικόνα δισδιάστατη  $[3 \times 4]$ :

$$\begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} & \alpha_{14} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} & \alpha_{24} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} & \alpha_{34} \end{bmatrix}$$

και πυρήνας  $[2 \times 2]$ :

$$\begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix}$$

Ο πυρήνας περνάει πάνω από όλη την εικόνα με βήμα 1 (stride) και δίνει τον τελικό πίνακα χαρακτηριστικών. Η επιλογή βήματος είναι κι αυτή μια παράμετρος των μοντέλων, η οποία διαλέγεται κατάλληλα για την βελτιστοποίησή τους. Ο πίνακας χαρακτηριστικών υπολογίζεται με πολλαπλασιασμό τού υποπίνακα της αρχικής εικόνας και του πυρήνα.

$$\begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \times \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix}$$

Με βάση αυτό, το συγκεκριμένο επίπεδο συνέλιξης δίνει τον εξής πίνακα χαρακτηριστικών.

$$\begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \end{bmatrix}$$

,όπου

$$c_{11} = \alpha_{11}\beta_{11} + \alpha_{12}\beta_{12} + \alpha_{21}\beta_{21} + \alpha_{22}\beta_{22}$$

$$c_{12} = \alpha_{12}\beta_{11} + \alpha_{13}\beta_{12} + \alpha_{22}\beta_{21} + \alpha_{23}\beta_{22}$$

$$c_{13} = \alpha_{13}\beta_{11} + \alpha_{14}\beta_{12} + \alpha_{23}\beta_{21} + \alpha_{24}\beta_{22}$$

$$c_{21} = \alpha_{21}\beta_{11} + \alpha_{22}\beta_{12} + \alpha_{31}\beta_{21} + \alpha_{32}\beta_{22}$$

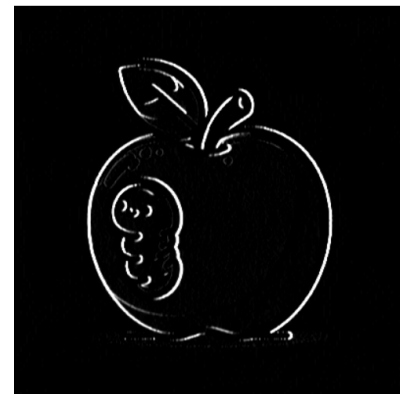
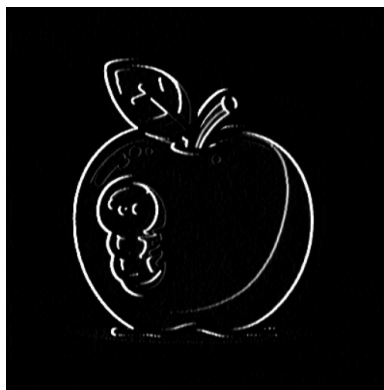
$$c_{22} = \alpha_{22}\beta_{11} + \alpha_{23}\beta_{12} + \alpha_{32}\beta_{21} + \alpha_{33}\beta_{22}$$

$$c_{23} = \alpha_{23}\beta_{11} + \alpha_{24}\beta_{12} + \alpha_{33}\beta_{21} + \alpha_{34}\beta_{22}$$

#### Παράδειγμα 1.7.4

Ανάλογα με την επιλογή τού πυρήνα, μπορούν να παραχθούν διαφορετικά χαρακτηριστικά τής εικόνας. Εδώ, χρησιμοποιούνται δύο διαφορετικοί πυρήνες. Οι δύο αυτοί πυρήνες αναγνωρίζουν τις άκρες τής εικόνας: ο πρώτος (left sobel) επικεντρώνεται στις ακμές προς την αριστερή πλευρά τής εικόνας, ο δεύτερος (right sobel) προς τη δεξιά. Αυτό που υπολογίζουν είναι η διαφορά στην ένταση μεταξύ δεξιάς και αριστερής μεριάς των εικονοστοιχείων. Αν η πιο φωτεινή πλευρά βρίσκεται στα αριστερά, τότε υπάρχει μια αριστερή ακμή, αν βρίσκεται στα δεξιά, μια δεξιά ακμή. Πρακτικά, εφαρμόζοντας αυτούς τους πυρήνες, οξύνονται οι κάθετες ακμές, κάτι που βοηθάει στην ανάδειξη κάθετων δομών κι ορίων που ενυπάρχουν στην εικόνα.

$$\text{Left sobel} = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}, \text{Right sobel} = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$



Εικόνα 9: Αριστερά η αρχική εικόνα. Στη μέση η εικόνα που της έχει εφαρμοστεί ο left sobel πυρήνας. Δεξιά η εικόνα που της έχει εφαρμοστεί ο right sobel πυρήνας.

Εάν υπάρχει πολυδιαστασιοποίηση των δεδομένων εισόδου, ο πυρήνας που εφαρμόζεται θα πρέπει να είναι τρισδιάστατος ή, ισοδύναμα, κάθε ένα από τα διαγράμματα χαρακτηριστικών θα συνελιχθεί με ένα διακριτό πυρήνα – και τα προκύπτοντα διαγράμματα θα συνοψιστούν ανά στοιχείο για να παράγουν τον τελικό πίνακα χαρακτηριστικών. Για παράδειγμα μια τρισδιάστατη συνέλιξη θα έχει ως πυρήνα έναν κύβο, ο οποίος θα μετατοπίζεται πάνω στους άξονες του ύψους, του βάθους και του μήκους [24]. Σε ένα συνελικτικό επίπεδο, συνήθως, χρησιμοποιούνται άνω του



ενός πυρήνες. Αυτό προκύπτει για το λόγω του ότι κάθε πυρήνας (όπως απεδείχθη και στο Παράδειγμα 1.7.4) παράγει διαφορετικά χαρακτηριστικά τής εικόνας. Εκτός, όμως, της ποικιλομορφίας που προσφέρει, ο μεγαλύτερος αριθμός πυρήνων συντελεί και σε άλλα καίρια σημεία. Αρχικά, αυξάνει την πολυπλοκότητα και το βάθος τού μοντέλου. Διαφορετικοί πυρήνες μπορούν να μάθουν να αναγνωρίζουν διαφορετικά επίπεδα. Στα αρχικά επίπεδα οι πυρήνες, αναγνωρίζουν ακμές κι υφές, ενώ σε βαθύτερα επίπεδα μπορούν κι αναγνωρίζουν πολυπλοκότερα μοτίβα ή συνδυασμούς χαρακτηριστικών που σχηματίζονται από τους πυρήνες που έχουν εξαχθεί σε ενωρίτερα στάδια. Επιπροσθέτως, ο μεγαλύτερος αριθμός πυρήνων βοηθάει στη γενίκευση του μοντέλου, δεδομένου ότι θα μάθει γενικότερα χαρακτηριστικά, αντί να υπερπροσαρμόσει τα δεδομένα εκπαίδευσης (overfitting). Με αυτόν τον τρόπο, το μοντέλο είναι προσαρμοστικότερο με δεδομένα, τα οποία δεν έχει δει. Κάτι τέτοιο, αφήνει στο μοντέλο μεγαλύτερες δυνατότητες μάθησης, αυξάνοντας τα περιθώρια βελτιστοποίησης της απόδοσής του. Τέλος, δίνει την ευκαιρία για παράλληλη επεξεργασία, με σκοπό να αυξηθεί η ταχύτητα εκπαίδευσης. Ο συνδυασμός όλων των διαφορετικών πινάκων χαρακτηριστικών, που παράγονται από τους πυρήνες ονομάζονται πίνακες ενεργοποίησης (activation maps). Ο βέλτιστος αριθμός πυρήνων συμπεραίνεται μέσα από τη μελέτη άλλων εμπειρικών μοντέλων, μέσω της υφιστάμενης βιβλιογραφίας και το βασικότερο μέσα από την ίδια την απόδοση του δικτύου, συνυπολογίζοντας την υπερπροσαρμογή του και την υπολογιστική αποδοτικότητα.

Οι πυρήνες μετατίθενται με βάση το βήμα που ορίζεται. Το βήμα (stride) είναι η απόσταση μεταξύ δύο συναπτών θέσεων πυρήνα πάνω στον άξονα κίνησης [24]. Το βήμα επηρεάζει το μέγεθος του πίνακα χαρακτηριστικών και το ποσοστό τής χωρικής επικάλυψης μεταξύ των αποκριτικών πεδίων όταν το φίλτρο εφαρμόζεται στην είσοδο. Όταν το βήμα είναι ένα, ο πυρήνας κινείται ανά ένα εικονοστοιχείο, αντιθέτως, για βήμα μεγαλύτερο του ενός ο πυρήνας παραβλέπει τόσα εικονοστοιχεία, όσα ορίζονται από το βήμα. Παραδείγματος χάρη, αν το βήμα είναι δύο, ο πυρήνας μετακινείται δύο εικονοστοιχεία δεξιά (και κάτω) μετά από κάθε διαδικασία. Αυξάνοντας το βήμα, μειώνονται οι χωρικές διαστάσεις τού εξαγόμενου πίνακα των χαρακτηριστικών, με αποτέλεσμα να πραγματώνεται ένα είδος υποβάθμισης (downsampling). Αυτό είναι επιθυμητό σε περιπτώσεις που η υπολογιστική πολυπλοκότητα δεν είναι αρκούντως μεγάλη ή σε περιπτώσεις που είναι επιθυμητή η ιεράρχηση των χαρακτηριστικών πάνω σε μεγαλύτερες χωρικές διαστάσεις.

Ακόμα μία παράμετρος που επηρεάζει το μέγεθος της εξαγόμενης εικόνας από τη διαδικασία τής συνέλιξης είναι το γέμισμα (padding). Η τεχνική γεμίσματος προσθέτει επιπλέον δεδομένα στα όρια τής εικόνας εισόδου, με σκοπό τη μεταχείριση των ύψος, πλάτος, μήκος τής εικόνας. Οι παραδοσιακές τεχνικές γεμίσματος, συμπεριλαμβάνουν τις γέμισμα με μηδέν (zero padding), γέμισμα με αναπαραγωγή (replication padding), και γέμισμα με αντανάκλαση (reflection padding). Οι δύο τελευταίες τεχνικές μεγαλώνουν την εικόνα με περιεχόμενο που δεν είναι ρεαλιστικό κι αυτό μπορεί να καταστρέψει την αρχική κατανομή. Η γέμισμα με μηδέν τεχνική υπερτερεί έναντι των άλλων δύο σε ταχύτητα κι ευκολία υπολογισμού. Οι τεχνικές αυτές είναι στατικές, εφαρμόζονται στην προεπεξεργασία των δεδομένων και δε βελτιστοποιούνται κατά τη διάρκεια της μάθησης [25].

- Γέμισμα με μηδέν: χρησιμοποιεί μηδενικές τιμές για να προσαυξήσει τις διαστάσεις τής εικόνας 10 (a).
- Γέμισμα με αναπαραγωγή: χρησιμοποιεί (αντιγράφει) τις τιμές των ορίων τής εικόνας για να προσαυξήσει τις διαστάσεις της 10 (b).
- Γέμισμα με αντανάκλαση: επαυξάνει την εικόνα αντανακλώντας την εικόνα στα άκρα της 10 (c).

Στα νευρωνικά δίκτυα, το επίπεδο συγκέντρωσης (pooling) χρησιμεύει στη μείωση των διαστάσεων της εικόνας. Κύρια αιτία χρησιμότητάς του είναι η σταδιακή μείωση των διαστάσεων, η μείωση των βαρών στο δίκτυο κι επομένως ο έλεγχος της υπερπροσαρμογής. Κάθε επίπεδο συγκέντρωσης λειτουργεί ανεξάρτητα σε κάθε τομή τής εισόδου, χρησιμοποιώντας ένα από τα δύο είδη συγκέντρωσης, το μέγιστο (max) ή το μέσο (average pooling). Το συγκεκριμένο επίπεδο, συνήθως επέρχεται του επιπέδου συνέλιξης. Το μέγιστο επίπεδο συγκέντρωσης κατατέμνει την είσοδο σε κομμάτια (patches) κι αποδίδει τη μέγιστη τιμή κάθε κομμάτι [24]. Το μέσο επίπεδο συγκέντρωσης αντί της μέγιστης τιμής αποδίδει ένα μέσο όρο των τιμών όλων των εικονοστοιχείων κάθε κομμάτι [24].

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	1	2	3	4	5	0	0	0
0	0	6	7	8	9	10	0	0	0
0	0	11	12	13	14	15	0	0	0
0	0	16	17	18	19	20	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

(a) zero padding

1	1	1	2	3	4	5	5	5	
1	1	1	2	3	4	5	5	5	
1	1	1	2	3	4	5	5	5	
6	6	6	7	8	9	10	10	10	
11	11	11	12	13	14	15	15	15	
16	16	16	17	18	19	20	20	20	
16	16	16	17	18	19	20	20	20	
16	16	16	17	18	19	20	20	20	

(b) replication padding

13	12	11	12	13	14	15	14	13	
8	7	6	7	8	9	10	9	8	
3	2	1	2	3	4	5	4	3	
8	7	6	7	8	9	10	9	8	
13	12	11	12	13	14	15	14	13	
18	17	16	17	18	19	20	19	18	
13	12	11	12	13	14	15	14	13	
8	7	6	7	8	9	10	9	8	

(c) reflection padding

Εικόνα 10: (a) Γέμισμα με μηδέν. (b) Γέμισμα με αναπαραγωγή (c) Γέμισμα με αντανάκλαση.

## Παράδειγμα 1.7.5

Έστω ο πίνακας χαρακτηριστικών  $[4 \times 4]$ :

$$\begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} & \alpha_{14} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} & \alpha_{24} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} & \alpha_{34} \\ \alpha_{41} & \alpha_{42} & \alpha_{43} & \alpha_{44} \end{bmatrix}$$

Η διαδικασία μέγιστης συγκέντρωσης ορίζεται παρόμοια με τη διαδικασία τής συνέλιξης. Επομένως, έστω ότι κάθε κομμάτι έχει διαστάσεις  $[2 \times 2]$  και μετατίθεται με βήμα δύο. Τότε, δημιουργείται ένας νέος πίνακας, ο οποίος έχει διαστάσεις  $[2 \times 2]$ :

$$\begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}$$

, όπου οι τιμές τους υπολογίζονται ως:

$$p_{11} = \max(\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22})$$

$$p_{12} = \max(\alpha_{13}, \alpha_{14}, \alpha_{23}, \alpha_{24})$$

$$p_{21} = \max(\alpha_{31}, \alpha_{32}, \alpha_{41}, \alpha_{42})$$

$$p_{22} = \max(\alpha_{33}, \alpha_{34}, \alpha_{43}, \alpha_{44})$$

## Παράδειγμα 1.7.6

Έστω ο πίνακας χαρακτηριστικών  $[4 \times 4]$ :

$$\begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} & \alpha_{14} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} & \alpha_{24} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} & \alpha_{34} \\ \alpha_{41} & \alpha_{42} & \alpha_{43} & \alpha_{44} \end{bmatrix}$$

Η διαδικασία μέσης συγκέντρωσης ορίζεται παρόμοια με τη διαδικασία τής συνέλιξης, όπως ακριβώς και η μέγιστη συγκέντρωση. Επομένως, έστω ότι κάθε κομμάτι έχει διαστάσεις  $[2 \times 2]$  και μετατίθεται με βήμα δύο. Τότε, δημιουργείται ένας νέος πίνακας, ο οποίος έχει διαστάσεις  $[2 \times 2]$ :

$$\begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}$$

, όπου οι τιμές τους υπολογίζονται ως:

$$p_{11} = avg(\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22})$$

$$p_{12} = avg(\alpha_{13}, \alpha_{14}, \alpha_{23}, \alpha_{24})$$

$$p_{21} = avg(\alpha_{31}, \alpha_{32}, \alpha_{41}, \alpha_{42})$$

$$p_{22} = avg(\alpha_{33}, \alpha_{34}, \alpha_{43}, \alpha_{44})$$

Το επίπεδο αντίστροφης συνέλιξης (up-conv ή transposed convolution) χρησιμοποιείται στο DL για να αυξήσει τις χωρικές διαστάσεις τής εισόδου, μια διαδικασία αντίστροφη από αυτήν της συνέλιξης. Αυτό το επίπεδο έχει χρησιμότητα σε επίπεδα αποκωδικοποιητή (decoder) ενός αυτοκωδικοποιητή (autoencoder) [24]. Η διαδικασία αντίστροφης συνέλιξης λειτουργεί εισάγοντας γέμισμα μεταξύ των μονάδων εισόδου, προσαυξάνοντας την είσοδο με ένα συνελικτικό τρόπο.

Παράδειγμα 1.7.7

Έστω μια εικόνα διαστάσεων  $[2 \times 2]$ :

$$\begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix}$$

κι έστω ο πίνακας που θα χρησιμοποιηθεί για τη συνέλιξη:

$$\begin{bmatrix} \beta_{11} & \beta_{12} & \beta_{13} & \beta_{14} \\ \beta_{21} & \beta_{22} & \beta_{23} & \beta_{24} \\ \beta_{31} & \beta_{32} & \beta_{33} & \beta_{34} \\ \beta_{41} & \beta_{42} & \beta_{43} & \beta_{44} \end{bmatrix}$$

,τότε υπολογίζεται ο επόμενος πίνακας:

$$\begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} \\ c_{21} & c_{22} & c_{23} & c_{24} \\ c_{31} & c_{32} & c_{33} & c_{34} \\ c_{41} & c_{42} & c_{43} & c_{44} \end{bmatrix}$$

,όπου

$$\begin{aligned}c_{11} &= \alpha_{11}\beta_{11} \\c_{12} &= \alpha_{11}\beta_{12} + \alpha_{12}\beta_{11} \\c_{13} &= \alpha_{11}\beta_{13} + \alpha_{12}\beta_{12} \\c_{14} &= \alpha_{12}\beta_{13} \\c_{21} &= \alpha_{11}\beta_{21} + \alpha_{21}\beta_{11} \\c_{22} &= \alpha_{11}\beta_{22} + \alpha_{12}\beta_{21} + \alpha_{21}\beta_{12} + \alpha_{22}\beta_{11} \\c_{23} &= \alpha_{11}\beta_{23} + \alpha_{12}\beta_{22} + \alpha_{21}\beta_{13} + \alpha_{22}\beta_{12} \\c_{24} &= \alpha_{12}\beta_{23} + \alpha_{22}\beta_{13} \\c_{31} &= \alpha_{11}\beta_{33} + \alpha_{21}\beta_{21} \\c_{32} &= \alpha_{11}\beta_{32} + \alpha_{12}\beta_{31} + \alpha_{21}\beta_{22} + \alpha_{22}\beta_{21} \\c_{33} &= \alpha_{11}\beta_{33} + \alpha_{12}\beta_{32} + \alpha_{21}\beta_{23} + \alpha_{22}\beta_{22} \\c_{34} &= \alpha_{12}\beta_{33} + \alpha_{22}\beta_{23} \\c_{41} &= \alpha_{21}\beta_{31} \\c_{42} &= \alpha_{21}\beta_{32} + \alpha_{22}\beta_{31} \\c_{43} &= \alpha_{21}\beta_{33} + \alpha_{22}\beta_{32} \\c_{44} &= \alpha_{22}\beta_{33}\end{aligned}$$

#### Παράδειγμα 1.7.8

Το συγκεκριμένο παράδειγμα βασίζεται στο άρθρο "Transposed Conv as Matrix Multiplication explained" [26]. Έστω ο πίνακας εισόδου που έχει προκύψει μετά από διαδικασία συνέλιξης κι ο πίνακας του πυρήνα:

$$\text{Είσοδος} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \text{Πυρήνας} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

Ο πίνακας εισόδου εκφράζεται κι ως μονοδιάστατος πίνακας:

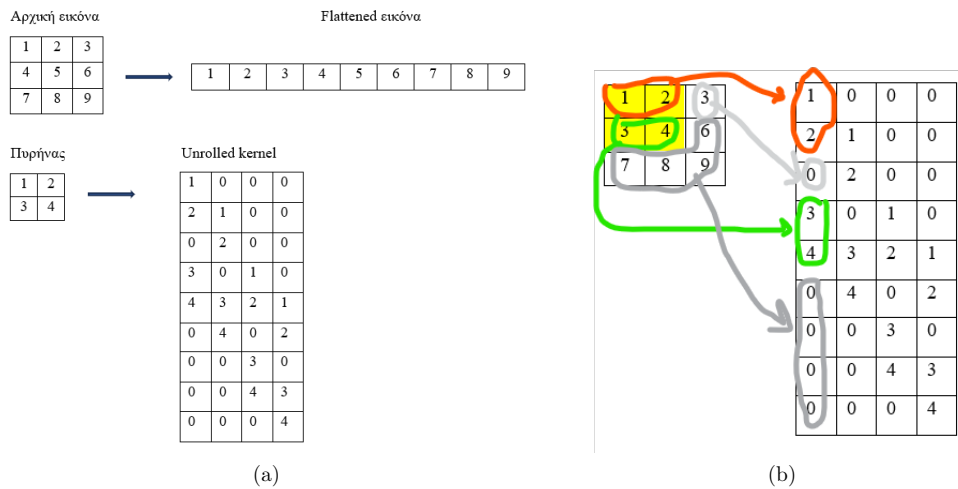
$$[1 \ 2 \ 3 \ 4]$$

, ενώ ο πίνακας πυρήνα εκφράζεται κι ως πίνακας που δείχνει όλες τις θέσεις τού πυρήνα στην αρχική εικόνα (transposed unrolled kernel):

$$\begin{bmatrix} 1 & 2 & 0 & 3 & 4 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 0 & 3 & 4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 2 & 0 & 3 & 4 & 0 \\ 0 & 0 & 0 & 0 & 1 & 2 & 0 & 3 & 4 \end{bmatrix}$$

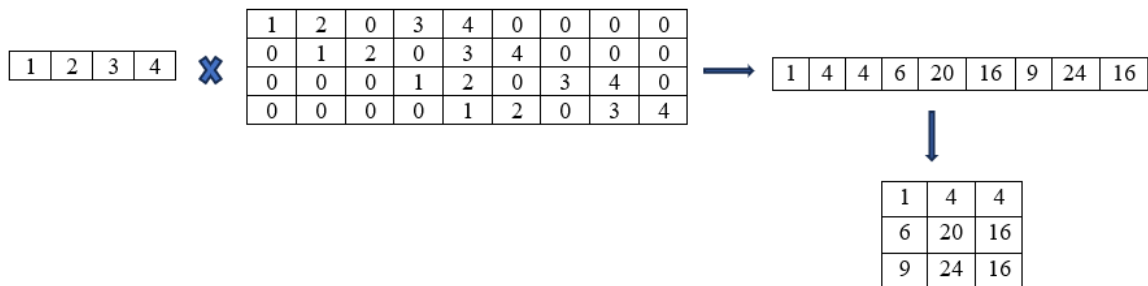
Οι διαστάσεις τού πίνακα transposed unrolled kernel αναφέρονται στις αρχικές διαστάσεις της εικόνας, αν αυτή γίνει ένας μονοδιάστατος πίνακας και στο πόσους διαφορετικούς πίνακες συνέλιξης μπορεί να εξάγει ο πυρήνας. Έτσι, μια εικόνα με αρχικές διαστάσεις  $[3 \times 3]$ , όταν γίνει μονοδιάστατος πίνακας έχει διαστάσεις  $[1 \times 9]$ . Στην αρχική εικόνα, ένας πυρήνας με διαστάσεις  $[2 \times 2]$  μπορεί να κάνει τέσσερις συνέλιξεις (Εικόνα 11). Επομένως, ο

unrolled kernel έχει διαστάσεις  $[4 \times 9]$ . Στο επίπεδο της αντίστροφης συνέλιξης, ο unrolled kernel αντιμετωπίζεται με αποτέλεσμα να έχει διαστάσεις  $[9 \times 4]$ .



Εικόνα 11: (a) Αρχική εικόνα, επιπεδοποιημένη (flattened) εικόνα, πυρήνας (kernel), unrolled kernel. (b) Η εικόνα δείχνει πώς προκύπτει ο unrolled kernel. Τα κουτάκια με κίτρινο χρώμα δείχνουν τη θέση του πυρήνα πάνω στην εικόνα. Ο πυρήνας καλύπτει τις θέσεις 1,2,4,5. Οι υπόλοιπες θέσεις είναι μηδέν (zero padding). Με τον ίδιο τρόπο υπολογίζεται κι ο υπόλοιπος πίνακας.

Για να εξαχθεί ο τελικός πίνακας μετά τη διαδικασία αντίστροφης συνέλιξης, πολλαπλασιάζονται ο μονοδιάστατος πίνακας εισόδου με τον πίνακα transposed unrolled kernel. Το αποτέλεσμα είναι ένας επίσης μονοδιάστατος πίνακας, ο οποίος εν συνεχεία μετασχηματίζεται στις διαστάσεις της αρχικής εικόνας (Εικόνα 12).



Εικόνα 12: Διαδικασία αντίστροφης συνέλιξης.

### 1.8.2 Αρχιτεκτονική μοντέλου Unet

Το μοντέλο έχει σχήμα U και η αρχιτεκτονική του είναι συμμετρική. Αποτελείται από δύο βασικά σημεία, το συστατικό μονοπάτι (αριστερό κομμάτι) και το διασταλτικό κομμάτι (δεξί κομμάτι). Το αριστερό μέρος της αρχιτεκτονικής, ο κωδικοποιητής (encoder) είναι δομημένο με συνελκτικές διαδικασίες, ενώ το δεξί μέρος, ο αποκωδικοποιητής (decoder) χρησιμοποιεί αντίστροφες συνελκτικές διαδικασίες.

#### Κωδικοποιητής

Το κομμάτι του κωδικοποιητή εκμαιεύει περιεχόμενο και συμπιέζει την εικόνα-είσοδο με το να μειώνει σταδιακά τις χωρικές της διαστάσεις. Η διαδικασία που ακολουθείται περιέχει επίπεδα συνέλιξης και διαδικασίες συγχέντρωσης. Κατά τη συγκεκριμένη διαδικασία του κωδικοποιητή, αποκτούνται γενικά χαρακτηριστικά, όπως σχήματα, υφές και μοτίβα, τα οποία ορίζουν διαφορετικές κλάσεις αντικειμένων εμφανιζόμενες στην εικόνα. Επίσης, μειώνονται

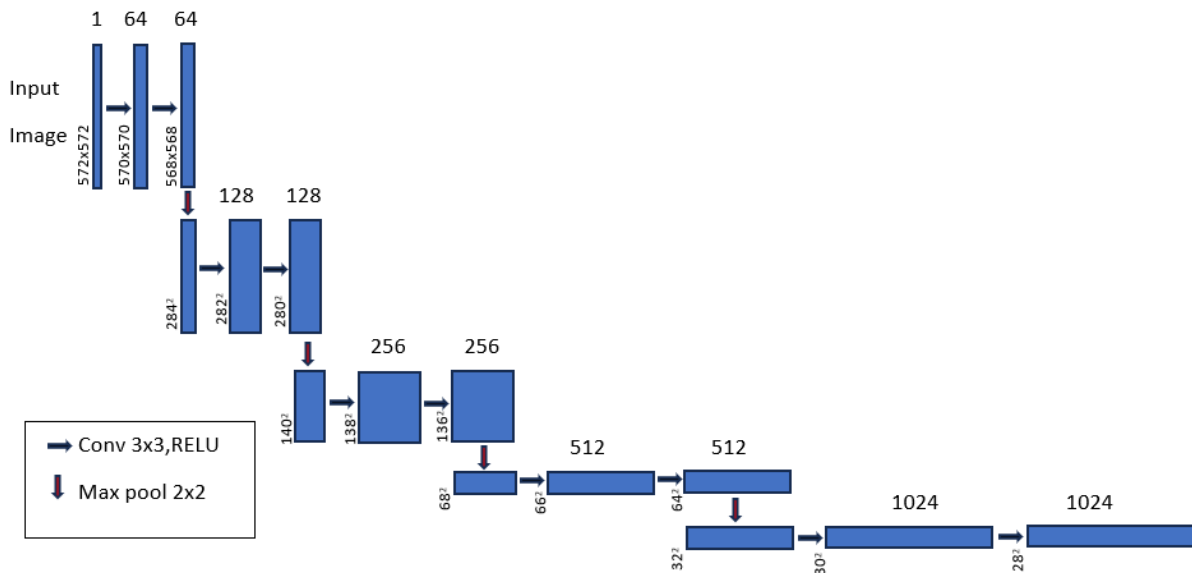
οι χωρικές διαστάσεις της, προετοιμάζοντας μία επιφανειακή αναπαράσταση χαρακτηριστικών για το μπλοκ του αποκωδικοποιητή, ώστε αυτός με τη σειρά του να την αναδομήσει. Όταν το Unet χρησιμοποιείται ως δομικό στοιχείο ενός μοντέλου ευθυγράμμισης, ο κωδικοποιητής αναλύει και συμπιέζει τόσο την κινούμενη, όσο και τη σταθερή εικόνα, σε μία συγχροτημένη, υψηλών διαστάσεων αναπαράσταση χαρακτηριστικών. Αυτή η αναπαράσταση είναι σημαντική για την κατανόηση των χωρικών σχέσεων και διαφορών μεταξύ των δύο εικόνων σε διάφορα επίπεδα. Η διαδικασία που ακολουθεί το άρθρο "U-Net: Convolutional Networks for Biomedical Image Segmentation" για την υλοποίηση του κωδικοποιητή παραχωρείται στην Εικόνα 13 [20]:

1. Αρχικά το μοντέλο δέχεται ως είσοδο μια εικόνα με διαστάσεις  $[572 \times 572 \times 1]$ . Περνάει από ένα επίπεδο συνέλιξης  $3 \times 3$  και μια ReLU με αποτέλεσμα να εξάγει μια εικόνα διαστάσεων  $[570 \times 570 \times 64]$ . Η RELU (Rectified Linear Unit) είναι μία συνάρτηση ενεργοποίησης η οποία ορίζεται ως:

$$\text{ReLU}(x) = \max(0, x) \quad (11)$$

Κύριος σκοπός που χρησιμοποιείται είναι για να εισάγει μη γραμμικότητα στο δίκτυο. Βοηθάει το μοντέλο να μάθει περίπλοκα μοτίβα και σχέσεις στα δεδομένα. Έστω, ότι ένα επίπεδο συνέλιξης εξάγει έναν πίνακα χαρακτηριστικών, ο οποίος εμπεριέχει τόσο θετικές όσο και αρνητικές τιμές. Όταν αυτός ο πίνακας περάσει από την ReLU, οι θετικές τιμές θα παραμείνουν ως έχουν, διατηρώντας τα χαρακτηριστικά που θεωρούνται χρήσιμα κι οι αρνητικές τιμές θα γίνουν μηδέν, παραγκωνίζοντας χαρακτηριστικά που δεν είναι χρήσιμα ή κρίσιμα για τη διαδικασία της εκμάθησης (όπως ο θόρυβος ή τα ασυσχέιστα μοτίβα).

2. Ο εξαχθέν πίνακας  $[570 \times 570 \times 64]$  περνάει από ένα ακόμη επίπεδο συνέλιξης ακολουθούμενο από τη ReLU με αποτέλεσμα τις νέες διαστάσεις  $[568 \times 568 \times 64]$ .
3. Έπειτα, εφαρμόζεται ένα επίπεδο max pool  $2 \times 2$  κι αυτό που προκύπτει είναι ένας καινούργιος πίνακας διαστάσεων  $[284 \times 284 \times 64]$ .
4. Επαναλαμβάνεται η πρώτη διαδικασία, κι ο πίνακας περνάει από ακόμη ένα επίπεδο συνέλιξης  $[3 \times 3]$  με μια ReLU, από τον οποίον και προκύπτει ο πίνακας με διαστάσεις  $[282 \times 282 \times 128]$ , όπου με τη σειρά του περνά για ακόμη μια φορά από το Conv  $[3 \times 3]$ , ReLU και προκύπτει ο πίνακας  $[280 \times 280 \times 128]$ .
5. Εφαρμόζεται ένα max pooling επίπεδο  $2 \times 2$  κι ο πίνακας έχει τώρα διαστάσεις  $[140 \times 140 \times 128]$ .
6. Η διαδικασία αυτή γίνεται άλλες δύο φορές επομένως: Conv  $3 \times 3$ , ReLU  $\implies [138 \times 138 \times 256]$  και Conv  $3 \times 3$ , ReLU  $\implies [136 \times 136 \times 256]$ .
7. Εφαρμόζεται ένα max pooling επίπεδο  $2 \times 2$  κι ο πίνακας έχει τώρα διαστάσεις  $[68 \times 68 \times 256]$ .
8. Conv  $3 \times 3$ , ReLU  $\implies [66 \times 66 \times 512]$  και Conv  $3 \times 3$ , ReLU  $\implies [64 \times 64 \times 512]$ .
9. Εφαρμόζεται το τελευταίο max pooling επίπεδο  $2 \times 2$  κι ο πίνακας έχει τώρα διαστάσεις  $[32 \times 32 \times 512]$ .
10. Επεξεργάζεται ακόμα δύο φορές ο πίνακας με Conv  $3 \times 3$ , ReLU  $\implies [30 \times 30 \times 1024]$  και Conv  $3 \times 3$ , ReLU  $\implies [28 \times 28 \times 1024]$ .

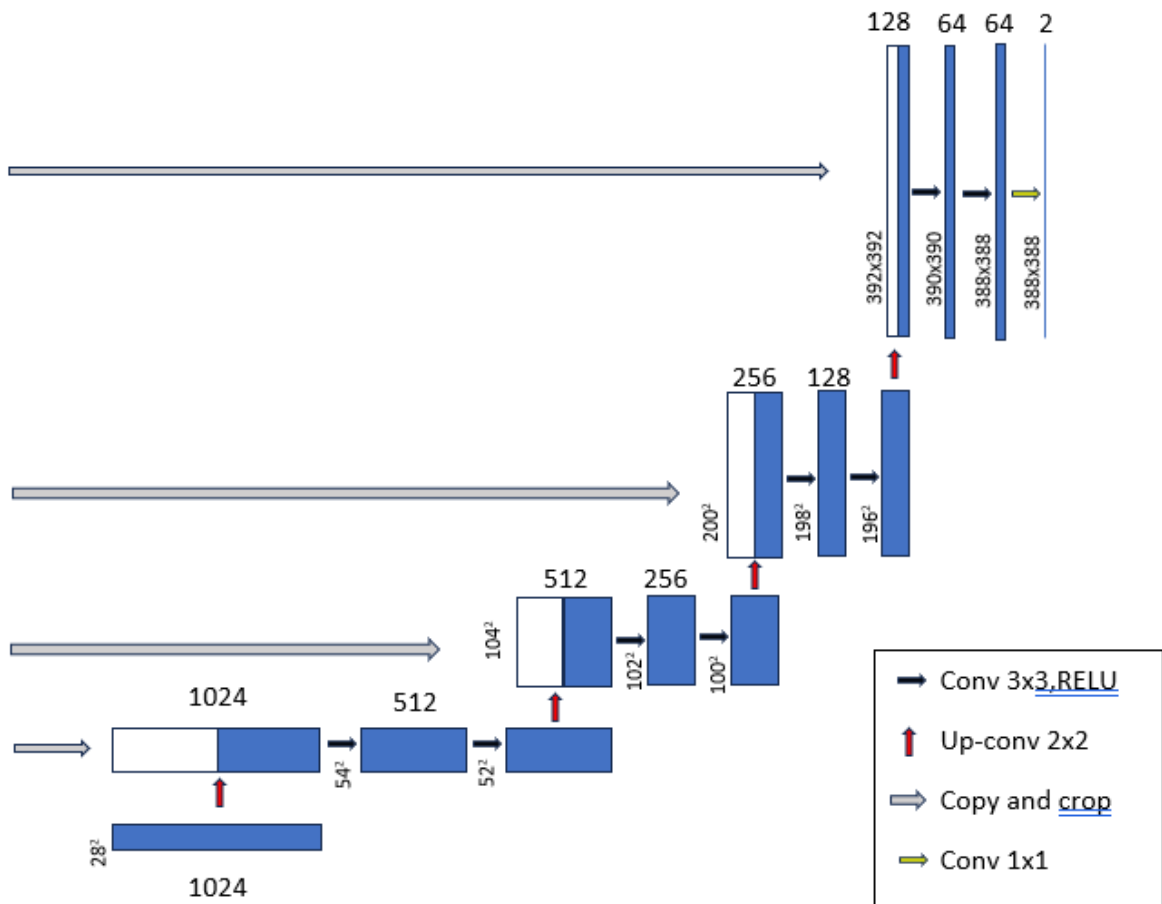


Εικόνα 13: Αρχιτεκτονική κωδικοποιητή Unet.[20]

### Αποκωδικοποιητής

Ο αποκωδικοποιητής χτίζει την έξοδο της εκάστοτε εφαρμογής, από το συμπιεσμένο πίνακα που εξήχθη από τον κωδικοποιητή. Αφού η εικόνα έχει σταδιακά συμπιεστεί, χάνοντας χωρική πληροφορία, αλλά κερδίζοντας πληροφορία για το περιεχόμενό της, ο ρόλος του αποκωδικοποιητή είναι να υπερδειγματίσει αυτήν την αναπαράσταση χαρακτηριστικών, αποκαθιστώντας τις χωρικές λεπτομέρειες. Ο αποκωδικοποιητής για να επιτύχει αυτό το αποτέλεσμα, κάνει χρήση των συνδέσεων παράβλεψης (skip connections). Οι συνδέσεις παράβλεψης προωθούν απευθείας πληροφορίες από το ένα επίπεδο στο άλλο, συνήθως παραβλέποντας ένα ή περισσότερα ενδιάμεσα επίπεδα. Στο μοντέλο Unet, οι συνδέσεις παράβλεψης διατηρούν χωρικές πληροφορίες που σε κάθε άλλη περίπτωση θα χανόταν λόγω υποβάθμισης (downsampling). Κάθε στάδιο υποβάθμισης, αντιστοιχίζεται με ένα στάδιο αναβάθμισης (upsampling). Οι συνδέσεις παράβλεψης για κάθε στάδιο προωθούν τη χωρική πληροφορία κι έπειτα αυτή ενσωματώνεται στον πίνακα χαρακτηριστικών που έχει προκύψει από τη διαδικασία αναβάθμισης. Το τελικό αποτέλεσμα της σύμπτυξης περνάει από επίπεδα συνέλιξης. Στην περίπτωση που το Unet χρησιμοποιείται ως μοντέλο σε εφαρμογές ευθυγράμμισης, ο αποκωδικοποιητής παράγει έναν πίνακα παραμόρφωσης (deformation matrix). Αυτό σημαίνει, ότι η έξοδος του αποκωδικοποιητή θα πρέπει να συμβαδίζει με τις επιθυμητές διαστάσεις της εικόνας που θα ευθυγραμμιστεί, δηλαδή αν πρόκειται για μια εικόνα τρισδιάστατη, τότε ο πίνακας θα πρέπει να παράγει την παραμόρφωση  $\Delta x, \Delta y, \Delta z$  (αντιστοιχεί στην παραμόρφωση που θα πρέπει να έχει κάθε ένα ογκοστοιχείο).

Η διαδικασία που ακολουθεί το άρθρο "U-Net: Convolutional Networks for Biomedical Image Segmentation" για την υλοποίηση του αποκωδικοποιητή παραχωρείται στην Εικόνα 14 [20]:



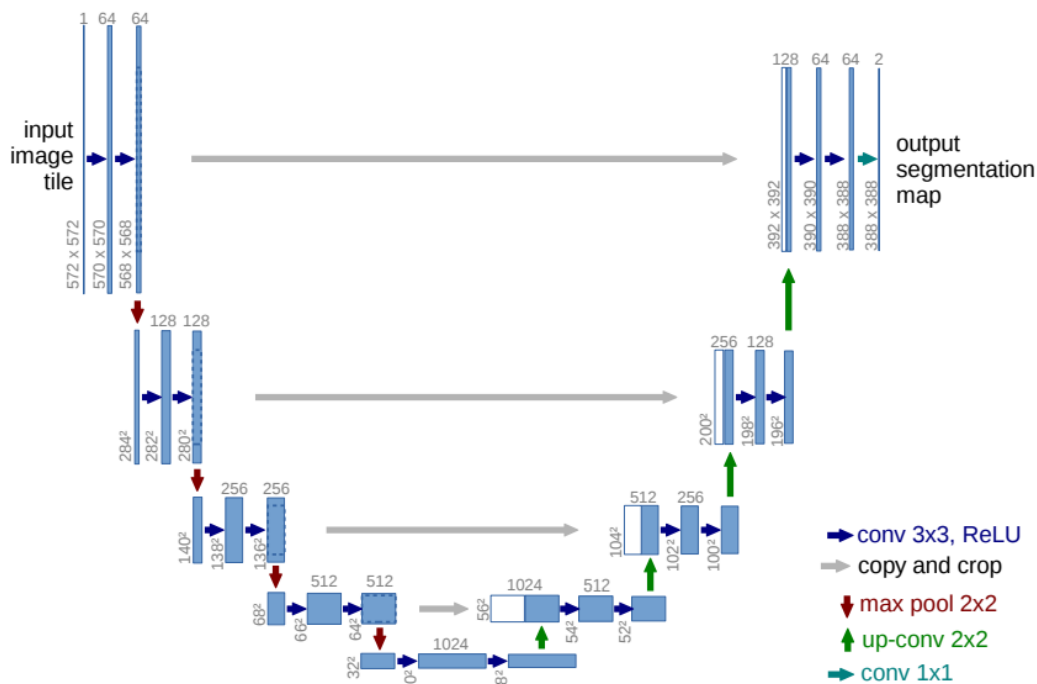
Εικόνα 14: Αρχιτεκτονική αποκωδικοποιητή Unet [20].

1. Ο αποκωδικοποιητής δέχεται ως είσοδο την έξοδο από τον κωδικοποιητή, επομένως έναν πίνακα χαρακτηριστικών με διαστάσεις  $[28 \times 28 \times 1024]$ .
2. Αυτός ο πίνακας περνάει από ένα επίπεδο αντίστροφης συνέλιξης  $2 \times 2$  με αποτέλεσμα έναν πίνακα διαστάσεων  $[56 \times 56 \times 512]$ .
3. Ο καινούργιος πίνακας, δέχεται από τις συνδέσεις παράβλεψης έναν άλλον πίνακα (τον πίνακα χαρακτηριστικών που έχει παραχθεί κι αντιγραφεί από το αντίστοιχο επίπεδο του κωδικοποιητή). Ο πίνακας αυτός ψαλιδίζεται (crop), ώστε να έχει τις ίδιες διαστάσεις με τον πίνακα από το επίπεδο αντίστροφης συνέλιξης και συνενώνεται μαζί του (concat). Μετά τη διαδικασία της συνένωσης προκύπτει ο πίνακας με διαστάσεις  $[56 \times 56 \times 1024]$ .
4. Επόμενο βήμα είναι η εφαρμογή ενός επιπέδου συνέλιξης  $3 \times 3$  ακολουθούμενο από μία ReLU και προκύπτει ο  $[54 \times 54 \times 512]$  πίνακας. Η διαδικασία αυτή επαναλαμβάνεται με αποτέλεσμα τις τελικές διαστάσεις  $[52 \times 52 \times 512]$ .
5. Εφαρμόζεται ξανά η αντίστροφη συνέλιξη  $2 \times 2$  με αποτέλεσμα τον  $\Rightarrow [104 \times 104 \times 256]$ .
6. Όπως και πριν, εφαρμόζεται η διαδικασία αντιγραφής και περικοπής και το αποτέλεσμα ενώνεται με τον πίνακα από την αντίστροφη συνέλιξη. Προκύπτει ο  $[104 \times 104 \times 512]$ .
7. Συνεχίζεται η διαδικασία με Conv  $3 \times 3$ , ReLU  $\Rightarrow [102 \times 102 \times 256]$  και ξανά Conv  $3 \times 3$ , ReLU  $\Rightarrow [100 \times 100 \times 256]$ .



8. Up-conv  $2 \times 2 \implies [200 \times 200 \times 128]$ .
9. Αντιγραφή, περικοπή και συνένωση  $\implies [200 \times 200 \times 256]$ .
10. Conv  $3 \times 3$ , ReLU  $\implies [198 \times 198 \times 128]$  και ξανά Conv  $3 \times 3$ , ReLU  $\implies [196 \times 196 \times 128]$ .
11. Up-conv  $2 \times 2 \implies [392 \times 392 \times 64]$ .
12. Αντιγραφή, περικοπή και συνένωση  $\implies [392 \times 392 \times 128]$ .
13. Τελευταίες δύο εφαρμογές επιπέδων συνέλιξης Conv  $3 \times 3$ , ReLU  $\implies [390 \times 390 \times 64]$  και ξανά Conv  $3 \times 3$ , ReLU  $\implies [388 \times 388 \times 64]$ .
14. Τελευταία επεξεργασία είναι η εφαρμογή ενός επιπέδου συνέλιξης  $1 \times 1$  με αποτέλεσμα τον τελικό πίνακα χαρακτηριστικών  $[388 \times 388 \times 2]$ .

Στην Εικόνα 15 φαίνεται η ολοκληρωμένη αρχιτεκτονική τού μοντέλου Unet, όπως αυτή προτάθηκε στο άρθρο των Ronneberger και Brox [20].



Εικόνα 15: Αρχιτεκτονική Unet [20].

## 1.9 Voxelmorph

Το Voxelmorph πρωτοσυστήθηκε από τους Balakrishnan et al. το 2019 [30]. Είναι ένα μοντέλο σχεδιασμένο για την ευθυγράμμιση ιατρικών εικόνων. Το συγκεκριμένο είναι πλήρως μη επιβλεπόμενο, που σημαίνει ότι δε χρειάζεται κάποια προορισμένη χωρική μεταμόρφωση, αλλά μαθαίνει να αντιστοιχίζει από ζευγάρια εικόνων χωρίς να υπάρχει κάποιο αληθές (ground truth) πεδίο παραμόρφωσης. Το Voxelmorph χρησιμοποιεί ένα CNN -συγκεκριμένα η αρχιτεκτονική του μοιάζει με του Unet- για να προβλέψει το πεδίο παραμόρφωσης, το οποίο συνταιριάζει κάλλιστα μια κινούμενη εικόνα σε μια σταθερή. Βρίσκει εφαρμογή τόσο σε δισδιάστατες, όσο και σε τρισδιάστατες εικόνες, ενώ έχει χρησιμοποιηθεί σε άπυλλες διαφορετικές ανατομικές περιπτώσεις. Η έξοδος είναι μια χωρική παραμόρφωση, η οποία προσδιορίζει πώς πρέπει να μετακινηθεί κάθε εικονοστοιχείο/ογκοστοιχείο στην κινούμενη εικόνα, ώστε να αντιστοιχίζεται με τη σταθερή εικόνα. Το παραχθέν πεδίο, έπειτα, εφαρμόζεται επί τής κινούμενης εικόνας για να παραχθεί η αντιστοιχισμένη εικόνα.

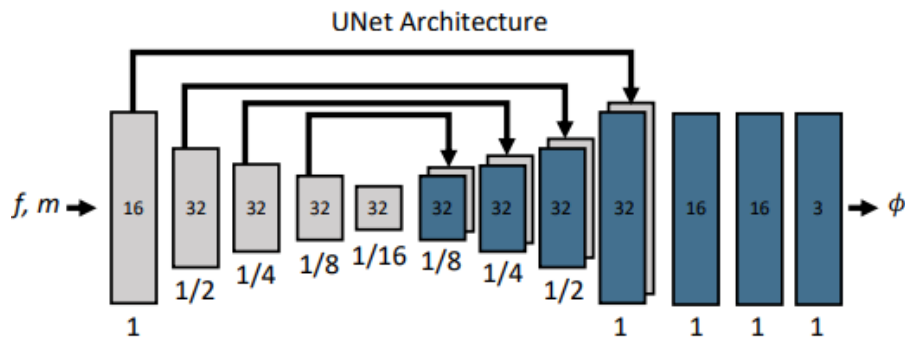
### 1.9.1 Αρχιτεκτονική μοντέλου Voxelmorph

Το Voxelmorph βασίζεται σε μια αρχιτεκτονική κωδικοποιητή, αποκωδικοποιητή με συνδέσεις παράβλεψης, όπως το Unet. Η είσοδος του μοντέλου αποτελείται από δύο εικόνες (κινούμενη, σταθερή) συγχωνευμένες. Μετά τη συγχώνευση δημιουργείται μια τρισδιάστατη εικόνα με δύο κανάλια. Όπως και στο Unet, ο κωδικοποιητής και ο αποκωδικοποιητής αποτελούνται από επίπεδα συνέλιξης με πυρήνα  $3 \times 3$  ακολουθούμενα από μία παραλλαγή της ReLU, την Leaky ReLU με παράμετρο 0.2. Η συγκεκριμένη παραλλαγή επιτρέπει ένα μικρό μη μηδενικό gradient, όταν ο νευρώνας δεν είναι ενεργός και η είσοδος είναι λιγότερη από μηδέν. Χάρη σε αυτήν τη διαφοροποίηση αποφεύγεται το πρόβλημα των νεκρών νευρώνων (όταν γίνονται ανενεργοί και σταματούν να προσφέρουν στην εκπαίδευση). Η Leaky ReLU ορίζεται ως:

$$\text{LeakyReLU}(x) = \begin{cases} x, & \text{if } x > 0 \\ \alpha x, & \text{if } x \leq 0 \end{cases} \quad (12)$$

,όπου το  $x$  είναι η είσοδος στο νευρώνα και το  $\alpha$  είναι μια πολύ μικρή μεταβλητή, που προσδιορίζει την κλίση της συνάρτησης για  $x \leq 0$ . Έτσι, για παράδειγμα αν η είσοδος στο νευρώνα είναι  $-2$  και το  $\alpha = 0.01$  η ReLU θα είχε ως έξοδο 0 και επομένως δε θα συντελούσε στη διαδικασία της οπισθοδιάδοσης (backpropagation), ενώ η Leaky ReLU θα είχε ως έξοδο  $0.01 \cdot -2 = -0.02$ .

Για τον κωδικοποιητή, οι χωρικές διαστάσεις της εικόνας υποδιπλασιάζονται μετά από κάθε συνέλιξη λόγω του ότι το βήμα ισούται με δύο. Οι συνέλιξεις χρησιμοποιούνται για να εντοπιστούν ιεραρχίες στο ζευγάρι εικόνων, που βοηθούν στην εκτίμηση του πεδίου παραμόρφωσης  $\phi$ . Στο κομμάτι του αποκωδικοποιητή χρησιμοποιούνται εναλλάξ up-convn επίπεδα, συνέλιξεις ακολουθούμενες από Leaky ReLU και συγχωνεύσεις των αντιγράφων πινάκων χαρακτηριστικών που δημιουργήθηκαν στον κωδικοποιητή με τις συνδέσεις παράβλεψης. Στην Εικόνα 16 φαίνεται η αρχιτεκτονική στο κομμάτι του κωδικοποιητή-αποκωδικοποιητή του Voxelmorph.



Εικόνα 16: Αρχιτεκτονική Unet για το Voxelmorph. Κάθε ορθογώνιο αναπαριστά έναν τρισδιάστατο πίνακα που παράχθηκε από τον προηγούμενο πίνακα χρησιμοποιώντας τρισδιάστατα συνελκτικά επίπεδα. Η χωρική ανάλυση κάθε όγκου-πίνακα καταγράφεται κάτωθεν των ορθογωνίων. Στον αποκωδικοποιητή χρησιμοποιούνται αρκετές συνέλιξεις με 32 πυρήνες, ακολουθούμενοι από αντίστροφα συνελκτικά επίπεδα ώστε να φέρουν τον όγκο-πίνακα στην αρχική του ανάλυση. Τα βέλη αναπαριστούν τις συνδέσεις παράβλεψης, οι οποίες συγχωνεύονται με τα χαρακτηριστικά του αποκωδικοποιητή. Έχοντας φτάσει στην αρχική ανάλυση, χρησιμοποιούνται επιπλέον συνέλιξεις μέχρι την έξοδο [30].

Προκειμένου να γίνει η παραμορφοποίηση (warping) της κινούμενης εικόνας χρησιμοποιείται μια Spatial Transformation συνάρτηση. Για κάθε ογκοστοιχείο  $p$ , υπολογίζεται μια υποπεριοχή (υπο-εικονοστοιχείο) στην κινούμενη εικόνα  $p' = p + u(p)$ , όπου  $u(p)$  είναι ένα διανυσματικό πεδίο, όπου κάθε διάνυσμα αντιπροσωπεύει τη μετατόπιση που θα πρέπει να εφαρμοστεί στο αντίστοιχο ογκοστοιχείο  $p$ . Αυτή η μετατόπιση μπορεί να έχει μια υπο-εικονοστοιχειακή ακρίβεια, που δείχνει ότι το ογκοστοιχείο θα πρέπει να μετακινηθεί, όχι μόνο σε μια άλλη θέση ενός ογκοστοιχείου, αλλά πιθανώς σε μια ενδιάμεση θέση μεταξύ ογκοστοιχείων. Το  $p'$  είναι η καινούργια θέση

τού ογκοστοιχείου κι υπολογίζεται αθροίζοντας τη μετατόπιση με την αρχική θέση  $p$  τού ογκοστοιχείου. Αυτό έχει ως αποτέλεσμα μια υπο-εικονοστοιχειακή περιοχή  $p'$ . Η διαδικασία τού Spatial Transformation χρησιμοποιείται για να μεταμορφώσει μια εικόνα βασισμένη στο πεδίο παραμόρφωσης. Για κάθε ογκοστοιχείο στην κινούμενη εικόνα η μεταμόρφωση δείχνει ακριβώς πού θα έπρεπε να είναι το ογκοστοιχείο για να υπάρχει αντιστοίχιση με τη σταθερή εικόνα. Ακριβώς επειδή η καινούργια θέση  $p'$  μπορεί να μην υπάρχει στο πλέγμα (grid) τής εικόνας, χρειάζεται συχνά η εφαρμογή παρεμβολής (interpolation), ώστε να μπορέσει να χτιστεί η μετασχηματισμένη εικόνα. Μερικές τεχνικές παρεμβολής εξηγούνται παρακάτω:

- Κοντινότερος Γείτονας (Nearest neighbor - NN): μια από τις απλούστερες και γρήγορες τεχνικές παρεμβολής. Αναθέτει σε κάθε ογκοστοιχείο στην μετασχηματισμένη εικόνα την τιμή έντασης του πιο κοντινού ογκοστοιχείου αρχικής εικόνας, βασισμένο στην Ευκλείδεια απόσταση. Έστω ένας χώρος  $\tilde{\Omega}$ , ο οποίος αναπαριστά το διακριτό χώρο τής εικόνας και έστω  $\Omega$  ο συνεχής χώρος, ενώ  $\tilde{x}, x$  είναι σημεία στους αντίστοιχους χώρους, τότε [32]:

$$I_{nearest}(\tilde{I}, x) = \tilde{I}([x]) \quad (13)$$

,όπου  $[x]$  αντιστοιχεί στο κοντινότερο σημείο τού πλέγματος. Παρόλο που αυτή η τεχνική είναι εύκολη και γρήγορη, χάνει σε ακρίβεια και για αυτό σπανίως χρησιμοποιείται σε ευθυγράμμιση εικόνων.

- Γραμμική (Linear): θεωρώντας ότι οι εντάσεις των τιμών αλλάζουν γραμμικώς από το ένα εικονοστοιχείο στο άλλο, η συγκεκριμένη τεχνική υπολογίζει το σταθμισμένο μέσο (weighted average) τού πιο κοντινού γειτονικού εικονοστοιχείου [32]:

$$I_{linear}(\tilde{I}, x) = \sum_{k \in \{0,1\}^d} \left[ \tilde{I}([x] + k) \prod_{l=1}^d \xi_l^{k_l} (1 - \xi_l)^{(1 - k_l)} \right] \quad (14)$$

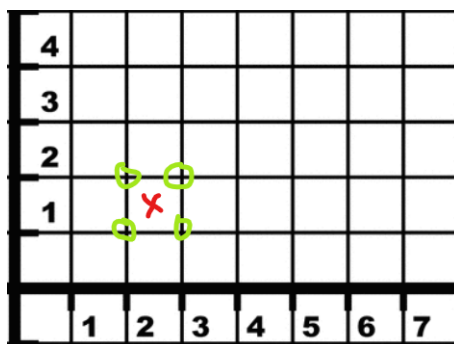
,όπου  $[x]$  αντιστοιχεί στο κοντινότερο σημείο πλέγματος, μικρότερο του  $\tilde{x}$ , το  $d$  αντιστοιχεί στις διαστάσεις τής εικόνας και το  $\xi = x - [x]$ . Για μια δεδομένη θέση  $x$ , προσδιορίζονται τα ακέραια σημεία πλέγματος που σχηματίζουν ένα μοναδιαίο κύβο ή τετράγωνο (για τρισδιάστατες ή δισδιάστατες εικόνες αντίστοιχα) γύρω από το  $x$ . Έπειτα, υπολογίζονται τα βάρη για αυτά τα σημεία βασισμένα στην απόστασή τους από το  $x$ . Αυτά τα βάρη λαμβάνονται από τα δεκαδικά κομμάτια  $\xi_l$ , τα οποία χρησιμεύουν στο να υπολογίζουν τη συμβολή των σημείων. Τα κοντινότερα σημεία έχουν μεγαλύτερη επιρροή στην παρεμβάλουσα τιμή σε σύγκριση με τα πιο απομακρυσμένα. Τέλος, αθροίζονται οι συμβολές όλων των γωνιών τού κύβου (ή του τετραγώνου) επιβαρυσμένα με την αντίστοιχη απόσταση από το  $x$ , ώστε να υπολογιστεί η τελική τιμή παρεμβολής. Για  $d = 2$  η τεχνική ονομάζεται διγραμμική (bilinear), για  $d = 3$  η τεχνική ονομάζεται τριγραμμική (trilinear).

#### Παράδειγμα 1.8.1.1

Έστω μια δισδιάστατη εικόνα με διαστάσεις  $[4 \times 4]$ :

$$\begin{bmatrix} 10 & 20 & 30 & 40 \\ 15 & 25 & 35 & 45 \\ 20 & 30 & 40 & 50 \\ 25 & 35 & 45 & 55 \end{bmatrix}$$

και μια μη ακέραια συντεταγμένη  $x = (1.5, 2.5)$ . Η διγραμμική παρεμβολή χρησιμοποιεί τα τέσσερα κοντινότερα ακέραια σημεία γύρω από το  $x$ , τα οποία είναι τα  $(1, 2)$ ,  $(1, 3)$ ,  $(2, 2)$ ,  $(2, 3)$  (Εικόνα 17).



Εικόνα 17: Με κόκκινο το σημείο  $x = (1.5, 2.5)$ , με πράσινο τα κοντινότερα σημεία  $(1, 2)$ ,  $(1, 3)$ ,  $(2, 2)$ ,  $(2, 3)$ .

Εφόσον το  $x = (1.5, 2.5)$ , το  $\xi_1 = 0.5$  για το  $x$  άξονα και το  $\xi_2 = 0.5$  για το  $y$  άξονα. Σύμφωνα με την εξίσωση 14 για διγραμμική παρεμβολή:

$$\begin{aligned} I_{linear}(\tilde{I}, x) &= \tilde{I}(1, 2) \cdot (1 - 0.5) \cdot (1 - 0.5) + \tilde{I}(1, 3) \cdot (1 - 0.5) \cdot 0.5 + \tilde{I}(2, 2) \cdot 0.5 \cdot (1 - 0.5) + \tilde{I}(2, 3) \cdot 0.5 \cdot 0.5 \\ &= 35 \cdot 0.25 + 45 \cdot 0.25 + 40 \cdot 0.25 + 50 \cdot 0.25 \\ &= 42.5 \end{aligned}$$

κι επομένως, η παρεμβάουσα τιμή στο  $x = (1.5, 2.5) = 43.5$ . Το  $k$  αντιπροσωπεύει ένα διάνυσμα που δεικνύει τις γωνίες τού τετραγώνου γύρω από το σημείο  $x$ . Το σύνολο  $\{0, 1\}^2$  δείχνει όλους τους δυνατούς συνδυασμούς τού  $k$  σε ένα δισδιάστατο χώρο, όπου το  $k$  παίρνει τη μορφή  $k = (k_x, k_y)$ . Εδώ, τα  $k_x, k_y$  είναι δυαδικοί δείκτες (0 ή 1) που επιλέγουν μεταξύ των χαμηλότερων και των μέγιστων ορίων των σημείων τού πλέγματος στις  $x, y$  διαστάσεις αντιστοίχως. Έτσι, το σύνολο  $\{0, 1\}^2$  απαρτίζεται από όλα τα ζευγάρια 0 και 1 σε δύο διαστάσεις, δηλαδή:  $(0, 0)$ ,  $(0, 1)$ ,  $(1, 0)$ ,  $(1, 1)$ . Κάθε συνδυασμός  $k$  αναπαριστά μία από τις τέσσερις γωνίες:

1.  $(0,0)$ : κάτω αριστερή
2.  $(0,1)$ : πάνω αριστερή
3.  $(1,0)$ : κάτω δεξιά
4.  $(1,1)$ : πάνω δεξιά

Η γενική ιδέα πίσω από τη συνάρτηση 14, όπως προαναφέρθηκε είναι να χρησιμοποιήσει ένα σταθμισμένο μέσο των τεσσάρων κοντινότερων εικονοστοιχείων κι ότι τα βάρη βασίζονται στην απόσταση ( $\xi_l$ ) μεταξύ τού σημείου στόχου και των γειτόνων. Έτσι, αν θεωρηθεί ένα σημείο  $x = (x_1, x_2)$  και τα δεκαδικά  $\xi_1, \xi_2$  σε κάθε διάσταση, η παρεμβολή για κάθε ένα από τα τέσσερα περιβάλλοντα σημεία (σηματοδοτημένα από το  $k = (k_1, k_2)$ ) είναι:

$$\sum_{k \in \{0,1\}^2} \tilde{I}([x] + k) \cdot (\xi_1^{k_1} \cdot (1 - \xi_1)^{(1-k_1)}) \cdot (\xi_2^{k_2} \cdot (1 - \xi_2)^{(1-k_2)})$$

Εφόσον, τα  $\xi_1, \xi_2$  είναι οι αποστάσεις από το σημείο στόχο με το κοντινότερο μικρότερο σημείο σε κάθε άξονα και τα  $k_1, k_2$  επαναλαμβάνονται πάνω στο  $\{0, 1\}$  αντικατοπτρίζοντας δύο πιθανές σχετικές θέσεις σε κάθε άξονα, η συνάρτηση προσαρμόζει τα βάρη σε κάθε γείτονα. Για παράδειγμα, για το γείτονα  $[x] + (0, 0)$  όπου και  $k_1 = k_2 = 0$  η συνεισφορά των βαρών έχει ως:

$$(1 - \xi_1)^1 \cdot (1 - \xi_2)^1 = (1 - 0.5) \cdot (1 - 0.5) = 0.25$$

,οπότε όταν το  $k_l = 0$  έχουμε  $(1 - \xi_l)$  επειδή ο γείτονας είναι στην χαμηλότερη πλευρά σύμφωνα με τη θέση

τού σημείου στόχου κι όταν το  $k_l = 1$  έχουμε  $\xi_l$  επειδή ο γείτονας είναι στην πάνω πλευρά. Η συγκεκριμένη μέθοδος εξασφαλίζει ότι τα βάρη μειώνονται με την απόσταση από το σημείο στόχο.

- Splines: μία spline βαθμού  $n$  είναι μία συνεχής πολυωνυμική συνάρτηση βαθμού  $n$  μια πραγματικής μεταβλητής με παραγωγούς τάξης μέχρι και  $n-1$ . Μοντελοποιεί τη φυσική διαδικασία σχεδιασμού μιας ομαλής καμπύλης. Μια ευρύτερη κατηγορία των splines είναι η B-splines. Έστω,  $n$  ένας μη αρνητικός ακέραιος αριθμός. Η κανονικοποιημένη B-spline συνάρτηση τάξης  $n$  γραφομένης  $\beta^n$  ορίζεται αναδρομικά ως [31]:

$$\beta^0(x) = \begin{cases} 1, & -\frac{1}{2} < x < \frac{1}{2} \\ \frac{1}{2}, & x = \pm\frac{1}{2} \\ 0, & \text{αλλού} \end{cases} \quad (15)$$

,

$$\text{και για } n \geq 0, \beta^{(n+1)} = \beta^{(n)} * \beta^{(0)}$$

,όπου  $*$  είναι μια συνέλιξη. Για κάθε  $n \geq 0$  μια συνάρτηση B-spline τάξης  $n+1$ ,  $\beta^{(n+1)}$  αποκτάται με συνέλιξη της  $n$ στής τάξης B-spline συνάρτησης,  $\beta^{(n)}$  με τη μηδενικής τάξης B-spline συνάρτησης  $\beta^{(0)}$ . Η συνέλιξη με το  $\beta^{(0)}$  προοδευτικά ομαλοποιεί ή διευρύνει τη συνάρτηση B-spline όσο το  $n$  μεγαλώνει. Αθροίζει αποτελεσματικά τις συνεισφορές των  $\beta^{(n)}$  σε ένα εύρος μετατοπίσεων που καθορίζονται από το  $\beta^{(0)}$ , κάτι που έχει ως αποτέλεσμα ομαλότερες μεταβάσεις κι ευρύτερη υποστήριξη καθώς αυξάνεται η τάξη. Κάθε αύξηση στην τάξη  $n$  αποδίδει σε μια συνάρτηση B-spline, η οποία είναι ένα βαθμό μεγαλύτερη σε ομαλότητα. Συγκεκριμένα, οι  $\beta^{(n)}$  συναρτήσεις είναι συνεχείς μέχρι και τη  $(n-1)$  παράγωγό τους.

#### Παράδειγμα 1.8.1.2

Έστω ένα ομοιόμορφο διάλυσμα κόμβων  $t = \{0, 1, 2, 3, 4\}$ . Η μηδενική B-splines,  $N_{i,0}(x)$  η οποία σχετίζεται με αυτό το διάλυσμα ορίζεται ως:

$$N_{i,0}(x) = \begin{cases} 1 & \text{αν } t_i \leq x < t_{i+1} \\ 0 & \text{αλλού} \end{cases}$$

Για αυτό το διάλυσμα προκύπτουν τέσσερις B-splines μηδενικής τάξης που αντιστοιχούν στα διαλύσματα  $[0, 1), [1, 2), [2, 3), [3, 4)$ . Οι μηδενικές B-spline ονομάζονται και βηματικές συναρτήσεις κι ορίζονται από την τιμή τους που είναι 1 στο διάστημα μεταξύ των διαδοχικών κόμβων και 0 παντού αλλού. Έτσι, για τα διαστήματα που ορίστηκαν παραπάνω (Εικόνα 18):

- $N_0 \implies N_{0,0}(x)$  και είναι 1 μόνο μεταξύ των κόμβων 0 και 1
- $N_1 \implies N_{1,0}(x)$  και είναι 1 μόνο μεταξύ των κόμβων 1 και 2
- $N_2 \implies N_{2,0}(x)$  και είναι 1 μόνο μεταξύ των κόμβων 2 και 3
- $N_3 \implies N_{3,0}(x)$  και είναι 1 μόνο μεταξύ των κόμβων 3 και 4



Εικόνα 18: B-splines μηδενικής τάξης. Ο x άξονας αναπαριστά τις τιμές 0 με 4 και ο y άξονας αναπαριστά την τιμή της B-spline συνάρτησης. Κάθε B-spline μοιάζει με κουτί κι έχει ύψος 1 μεταξύ των ορίων του, ενώ γίνεται 0 έξω από αυτά.

Μία B-spline πρώτης τάξης  $N_{i,1}(x)$  αποκτάται με τη συνέλιξη δύο γειτονικών μηδενικής τάξεως B-spline. Η συνέλιξη για πρώτης τάξεως B-splines δοσμένων των ομοιόμορφων αποστάσεων στο διάλυμα κόμβων, απλοποιείται σε μια γραμμική παρεμβολή μεταξύ των γειτονικών B-splines μηδενικής τάξης.

$$N_{i,1}(x) = (x - t_i)N_{i,0}(x) + (t_{i+2} - x)N_{i+1,0}(x)$$

Η απόσταση μεταξύ τους είναι 1 κι επομένως η παρεμβολή  $N_{1,1}(x)$  μεταξύ για παράδειγμα των κόμβων 1 και 2 είναι:

Για  $x$  στο διάστημα  $[1, 2)$ :

Από  $t_1$  σε  $t_2$  ( $1 \leq x < 2$ ):

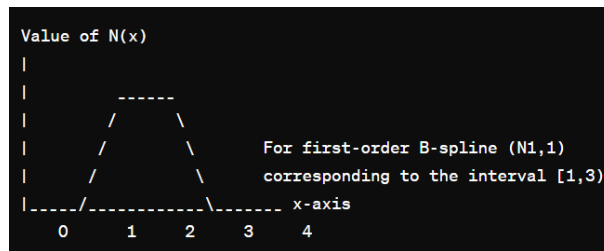
- $N_{1,0}(x) = 1$  για  $x \in [1, 2)$  επειδή  $t_1 \leq x < t_2$
- $N_{2,0}(x) = 1$  για  $x \in [2, 3)$ , αλλά επειδή το διάστημα ενδιαφέροντος είναι το  $[1, 2)$  θεωρείται η συμβολή του στο σχηματισμό τής καμπύλης μετά ακριβώς από το  $x = 2$

Για  $x$  στο διάστημα  $[1, 2)$ ,  $N_{1,1}(x)$  περνάει γραμμικά:

- $N_{1,1}(x) = (x - 1)$  επειδή το  $N_{1,0}(x) = 1$  συμβάλει στην κλίση αυξάνοντάς την από 0 στο  $x = 1$  σε 1 στο  $x = 2$
- Εξαιτίας της απλούστευσης η γραμμική μετάβαση για το  $N_{1,1}(x)$  διορθώνεται ως:  $N_{1,1}(x) = (x - 1)$  για  $1 \leq x < 2$  συμβολίζοντας μια γραμμική αύξηση.

Για το  $x$  στο διάστημα  $[2, 3)$ :

- $N_{1,1}(x) = (3 - x)$ , το οποίο παριστάνει μια γραμμική μείωση από το 1 στο σημείο  $x = 2$  στο 0 στο σημείο  $x = 3$ . Η οπτικοποίηση τής πρώτης τάξης B-spline φαίνεται στην Εικόνα 19.



Εικόνα 19: Δεδομένων μηδενικής τάξεως B-splines ορισμένες στο διαστήματα  $[0, 1)$ ,  $[1, 2)$ ,  $[2, 3)$ ,  $[3, 4)$  η πρώτης τάξεως B-spline προερχόμενη από τη συνέλιξη των θα εμφανίσει μια γραμμική αύξηση, μια κορυφή και μια γραμμική μείωση μεταξύ δύο διαδοχικών διαστημάτων λόγω της διαδικασίας τής συνέλιξης.

Στο άρθρο τους οι Balakrishnan et al. [30] χρησιμοποιούν τριγραμμική παλινδρόμηση:

$$m \circ \phi(p) = \sum_{q \in \mathcal{Z}(p')} m(q) \prod_{d \in \{x,y,z\}} (1 - |p'_d - q_d|) \quad (16)$$

,όπου το  $\mathcal{Z}(p')$  είναι τα γειτονικά ογκοστοιχεία τού  $p'$  και το  $d$  παίρνει διαστάσεις τού  $\Omega$ . Όπως φαίνεται και στην εξίσωση, η διαδικασία αναθέτει την τιμή τού μετασχηματισμένου voxel  $p'$  στην τιμή τού ογκοστοιχείου  $p$ . Το  $p'$  επηρεάζεται (με βάρος) από τα κοντινότερα γειτονικά εικονοστοιχεία  $q$  στις 8 κατευθύνσεις. Καθώς η έξοδος του Unet δεν είναι απαραίτητα ακέραιος, γίνεται η παρεμβολή με βάση αυτά τα 8 σημεία. Αν το  $q$  είναι ακέραιος, τότε το βάρος είναι 1 και τα υπόλοιπα σημεία 0.

Μετά τη διαδικασία τής παρεμβολής, όπου εξάγεται η τελική μετασχηματισμένη εικόνα, ενεργοποιείται η συνάρτηση κόστους. Υπάρχουν δύο είδη συναρτήσεων κόστους στο Voxelmorph, μια μη επιβλεπόμενη  $L_{us}$ , η οποία αξιολογεί το μοντέλο χρησιμοποιώντας μόνο τους αρχικούς όγκους και το παραγόμενο πίνακα παραμόρφωσης κι ένα επαυξητικό επιβλεπόμενο κόστος, το οποίο χρησιμοποιεί ανατομικές κατατιμήσεις στη διαδικασία τής εκπαίδευσης [30].

1. Μη επιβλεπόμενο κόστος: αποτελείται από δύο παράγοντες τον  $L_{sim}$  και τον  $L_{smooth}$ :

$$L_{us}(f, m, \phi) = L_{sim}(f, m \circ \phi) + \lambda L_{smooth}(\phi) \quad (17)$$

Ο πρώτος παράγοντας μετράει το πόσο διαφέρει οπτικά η σταθερή εικόνα από τη μετασχηματισμένη εικόνα, ενώ ο δεύτερος χρησιμοποιεί μια παράμετρο κανονικοποίησης  $\lambda$  για να βρει τι διαφορές στις χωρικές διαφορές. Το  $L_{sim}$  μπορεί να είναι οποιαδήποτε συνάρτηση κόστους. Στο άρθρο χρησιμοποιείται η MSE κι η NCC (οι οποίες αναλύθηκαν παραπάνω). Το  $L_{smooth}$  εισάγεται ώστε να ομαλοποιήσει το πεδίο παραμόρφωσης. Ο συγκεκριμένος παράγοντας, εξασφαλίζει ότι το προβλεπόμενο πεδίο παραμόρφωσης είναι ρεαλιστικά εφικτό και δεν εισάγει μη ρεαλιστικές στρεβλώσεις στη μετασχηματισμένη εικόνα. Το  $L_{smooth}$  ορίζεται ως:

$$L_{smooth}(\phi) = \sum_{p \in \Omega} \|\nabla u(p)\|^2 \quad (18)$$

,όπου  $\nabla u(p) = (\frac{\partial u(p)}{\partial x}, \frac{\partial u(p)}{\partial y}, \frac{\partial u(p)}{\partial z})$  και  $\frac{\partial u(p)}{\partial x} \sim u(p_x + 1, p_y, p_z) - u(p_x, p_y, p_z)$  (ομοίως για τα υπόλοιπα).

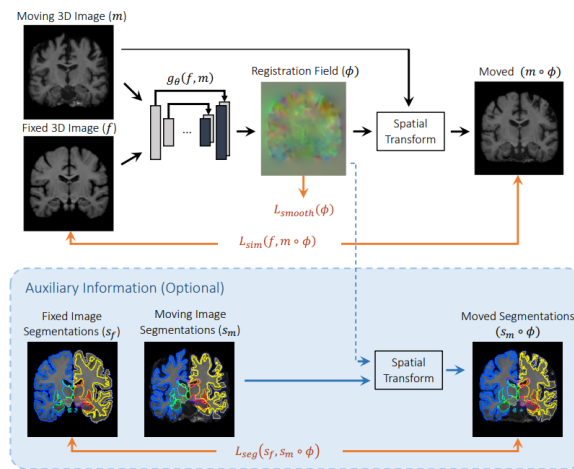
2. Επιβλεπόμενο κόστος: το Voxelmorph δύναται να εξάγει επιπρόσθετες πληροφορίες κατά τη διάρκεια της εκπαίδευσης. Οι πίνακες κατάτμησης ατομικών χαρακτηριστικών μπορούν να βρεθούν κατά την εκπαίδευση και να υποσημειωθούν (labeling) από ανθρώπους ή αλγόριθμους. Ένας τέτοιος πίνακας αναθέτει κάθε ογκοστοιχείο σε μία κατακερματισμένη δομή. Αν το πεδίο παραμόρφωσης  $\phi$  αναπαριστά ακριβή ανατομικές εικόνες, οι περιοχές στη σταθερή εικόνα κι η μετασχηματισμένη εικόνα που αντιστοιχούν στη ίδια ανατομική δομή θα πρέπει να επικαλύπτονται [30]. Για να γίνει αυτή η εξέταση της επικάλυψης χρησιμοποιείται η μετρική Dice score, η οποία αναλύθηκε παραπάνω. Το κόστος κατάτμησης  $L_{seg}$  όλων των δομών  $k \in [1, K]$  ορίζεται ως:

$$L_{seg}(s_f, s_m \circ \phi) = -\frac{1}{K} \sum_{k=1}^K \text{Dice}(s_f, s_m \circ \phi) \quad (19)$$

,όπου  $s_f, s_m \circ \phi$  είναι τα ογκοστοιχεία στις δομές  $k$  για τη σταθερή και τη μετασχηματισμένη εικόνα αντίστοιχα. Επειδή, από μόνη της η  $L_{seg}$  δεν εισάγει ομαλότητα κι έλεγχο ομοιότητας εικόνας, γίνεται ένας συνδυασμός των 19 και 17 με το  $\gamma$  να είναι μια παράμετρος κανονικοποίησης [30]:

$$L_a(f, m, s_f, s_m, \phi) = L_{us}(f, m, \phi) + \gamma L_{seg}(s_f, s_m \circ \phi) \quad (20)$$

Στην Εικόνα 20 φαίνεται η συνολική αρχιτεκτονική του μοντέλου Voxelmorph.



Εικόνα 20: Συνολική αρχιτεκτονική Voxelmorph.[30]

## 2 Μέρος Β

Με σκοπό την εξέταση του θεωρητικού μέρους σε πρακτικό επίπεδο, αναπτύχθηκαν συγκεκριμένες τεχνικές ευθυγράμμισης εικόνων και δοκιμάστηκαν πάνω σε ένα σύνολο δεδομένων. Οι τεχνικές αυτές αφορούν το μοντέλο Βαθιάς Μάθησης Voxmorph και τη βιβλιοθήκη SimpleITK. Οι δύο τεχνικές υλοποιήθηκαν χρησιμοποιώντας τη γλώσσα προγραμματισμού Python 3.10.12, στο περιβάλλον Colab Pro. Στα επερχόμενα υποκεφάλαια αναπτύσσονται αναλυτικά τα δεδομένα εκπαίδευσης, η προεπεξεργασία των δεδομένων εκπαίδευσης, οι προσαρμογές των τεχνικών ευθυγράμμισης, το περιβάλλον προγραμματισμού κι οι βιβλιοθήκες που χρησιμοποιούνται.

### 2.1 SimpleITK βιβλιοθήκη

Η SimpleITK βιβλιοθήκη είναι μια απλοποιημένη, ευρέως διαθέσιμη (open source) διεπαφή στην εργαλειοθήκη Insight Toolkit (ITK). Η ITK είναι κι αυτή μια διαθέσιμη σε όλους βιβλιοθήκη αναπτυγμένη στη γλώσσα προγραμματισμού C++, η οποία χρησιμοποιείται κατά κόρον στην ακαδημαϊκή έρευνα και στη βιομηχανία. Η SimpleITK είναι διαθέσιμη για χρήση σε οχτώ γλώσσες προγραμματισμού, συμπεριλαμβανομένων των C++, Python, R, Java, C#, Lua, Ruby, and TCL. Υφίστανται διαθέσιμες προμεταγλωττισμένες εκδόσεις της SimpleITK βιβλιοθήκης, έτοιμες να χρησιμοποιηθούν χωρίς την ανάγκη μεταγλώττισης του αρχικού πηγαίου κώδικα για τα λειτουργικά συστήματα Linux, OS X, και Windows [33],[34],[35].

Η διεπαφή SimpleITK δημιουργήθηκε στην προσπάθεια απλοποίησης της ITK, ώστε να γίνει ευκολότερα προσβάσιμη σε μεγαλύτερο κοινό. Η αρχική χρηματοδότηση δόθηκε από την Αμερικανική Εθνική Ιατρική Βιβλιοθήκη (United States National Library of Medicine - NLM) υπό την αιγίδα του Αμερικανικού νόμου Ανάκαμψης και Επανεπένδυσης (American Recovery and Reinvestment Act - ARRA). Η αρχική έκδοση ανεπτύχθη με τη συνεργασία των Κλινική Mayo (Mayo Clinic) [36], Kitware Inc [37], Πανεπιστήμιο της Αϊόβα (University of Iowa) [38] και του ενδοσχολικού ερευνητικού προγράμματος του NLM [39]. Η πρώτη βασική έκδοση της SimpleITK πραγματοποιήθηκε τον Απρίλιο-Μάιο του 2017. Μεταξύ 2013 και 2019 η ανάπτυξη υποστηρίχθηκε από ερευνητικό πρόγραμμα του NLM σε συνεργασία με το Πανεπιστήμιο της Αϊόβα και το Πανεπιστήμιο Monash [40]. Έκτοτε, η ανάπτυξή του υποστηρίζεται από το Γραφείο Κυβερνοϋποδομής και Υπολογιστικής Βιολογίας (Office of Cyber Infrastructure and Computational Biology - OCICB) του Εθνικού Ινστιτούτου Αλλεργιών και Λοιμωδών Νοσημάτων [41] (National Institute of Allergy and Infectious Diseases) [42].

Η SimpleITK αφορά τόσο κομμάτια κατάτμησης εικόνων, όσο κι ευθυγράμμισης εικόνων. Η διαδικασία της ευθυγράμμισης υλοποιείται μέσω της μεθόδου ImageRegistrationMethod και των φίλτρων Demons.

Προκειμένου να δημιουργηθεί μια διαδικασία ευθυγράμμισης χρησιμοποιώντας την ImageRegistrationMethod, επιβάλλεται η επιλογή συγκεκριμένων χαρακτηριστικών [43]:

- Μετασχηματισμός (Transformation): ορίζει ποια θα είναι η αντιστοίχιση μεταξύ των δύο εικόνων.
- Μετρική Ομοιότητας (Similarity metric): ορίζει ποια θα είναι η συσχέτιση μεταξύ των εντάσεων των εικονοστοιχείων των εικόνων.
- Βελτιστοποιητής (Optimizer): εντοπίζει τις παραμέτρους μετασχηματισμού που αντιστοιχίζουν καλύτερα την κινούμενη εικόνα με τη σταθερή.
- Συνάρτηση Παρεμβολής (Interpolator): επιλέγεται κατάλληλα για το μετασχηματισμό της ευθυγραμμισμένης εικόνας.

Τα φίλτρα Demons επαναληπτικά ελαχιστοποιούν τη διαφορά μεταξύ δύο εικόνων (μια σταθερή εικόνα και μια κινούμενη εικόνα) θεωρώντας ότι η κινούμενη εικόνα παραμορφώνεται υπό την επίδραση ενός συναγόμενου πεδίου δύναμης. Αυτό το πεδίο δύναμης προέρχεται από τις διαφορές έντασης μεταξύ των εικόνων, με στόχο την αντιστοίχιση των χαρακτηριστικών και χρησιμοποιείται για την επαναληπτική ενημέρωση του πεδίου παραμόρφωσης, ωθώντας την κινούμενη εικόνα προς τη σταθερή εικόνα. Ο όρος "πεδίο δύναμης" αναφέρεται σε ένα θεωρητικό μοντέλο και περιγράφει πώς κάθε εικονοστοιχείο/ογκοστοιχείο στην κινούμενη εικόνα θα εξωθηθεί/τραβηχτεί προς το αντίστοιχο σημείο στη σταθερή εικόνα. Η συνάρτηση ακολουθεί τα εξής βήματα:



- Ο αλγόριθμος ξεκινάει υπολογίζοντας τη διαφορά στις εντάσεις των αντίστοιχων εικονοστοιχείων μεταξύ των εικόνων. Η διαφορά αυτή επισημαίνει πόσο "μακριά" είναι οι εικόνες.
- Μαζί με τις διαφορές στις εντάσεις, ο αλγόριθμος υπολογίζει τη διαφορά στη φωτεινότητα της κινούμενης εικόνας (gradient). Αυτό βοηθάει στο να οριστεί σε ποια κατεύθυνση θα πρέπει να κινηθούν τα εικονοστοιχεία.
- Συνδυάζοντας τις διαφορές στις εντάσεις με τη φωτεινότητα (gradient), υπολογίζεται ένα διάνυσμα στην κινούμενη εικόνα. Αυτό είναι και το εξαγόμενο πεδίο μετασχηματισμού ή πεδίο δύναμης.
- Η κινούμενη εικόνα μετασχηματίζεται σύμφωνα με αυτό το πεδίο. Αυτό το βήμα επαναλαμβάνεται, με το πεδίο δύναμης να επανυπολογίζεται σε κάθε επανάληψη, ώστε να επιτευχθεί βελτίωση στην ευθυγράμμιση.
- Για να εξασφαλιστεί ότι ο μετασχηματισμός είναι ομαλός, συνήθως εισάγεται ένα βήμα κανονικοποίησης, το οποίο εφαρμόζεται στο πεδίο δύναμης. Με αυτόν τον τρόπο τα διανύσματα εξομαλύνονται, διασφαλίζοντας ότι τα γειτονικά εικονοστοιχεία κινούνται με συνοχή.

Τα φίλτρα Demons είναι ιδιαίτερα αξιοσημείωτα για την εφαρμογή τους σε σενάρια όπου η εγγραφή περιλαμβάνει παραμορφωτικούς μετασχηματισμούς λόγω της ικανότητάς τους να προσαρμόζονται σε τοπικές παραμορφώσεις.

## 2.2 Περιβάλλον προγραμματισμού και χρήσιμα εργαλεία

Για την υλοποίηση της ευθυγράμμισης χρησιμοποιήθηκε το περιβάλλον προγραμματισμού Google Colab Pro. Η γλώσσα προγραμματισμού είναι η Python 3.10.12, ενώ οι βιβλιοθήκες που αξιοποιήθηκαν τόσο για την προεπεξεργασία των δεδομένων, την κατασκευή των μοντέλων όσο και για την οπτικοποίηση των αποτελεσμάτων είναι οι Numpy, Pandas, Nibabel, SimpleITK, Matplotlib, Os, Torchvision.

## 2.3 Σύνολο Δεδομένων

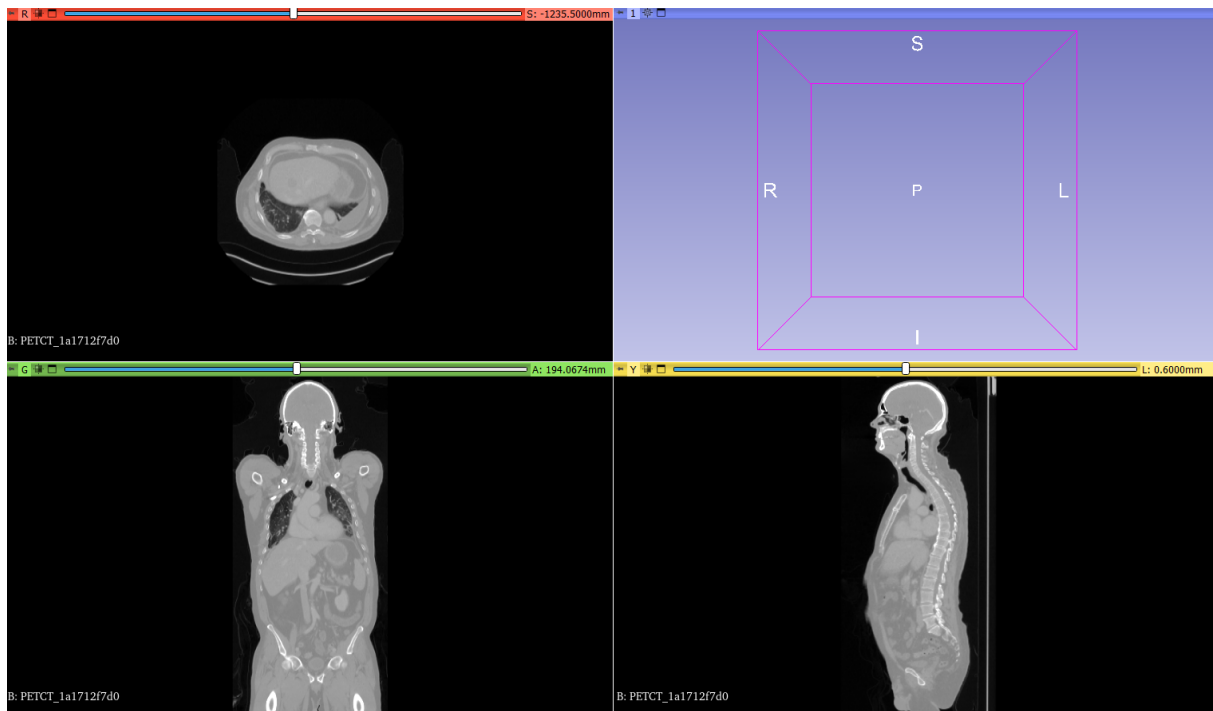
Το σύνολο των δεδομένων που χρησιμοποιούνται για την ευθυγράμμιση αποτελείται από απρόσωπα δεδομένα, συμβαδίζοντα με την τοπική επιτροπή ηθικής και τον υπεύθυνο προστασίας δεδομένων. Συμπεριελήφθησαν 501 διαδοχικά σύνολα δεδομένων ολόσωμης FDG-PET/CT [58] ασθενών με κακοήθη λέμφωμα, μελάνωμα και μη μικροκυτταρικό καρκίνο του πνεύμονα (NSCLC), καθώς και 513 σύνολα δεδομένων χωρίς PET-θετικές κακοήθειες πληγές (αρνητικοί έλεγχοι) που εξετάστηκαν μεταξύ 2014 και 2018 στο Πανεπιστημιακό Νοσοκομείο του Τίμπινγκεν. Όλες οι εξετάσεις πραγματοποιήθηκαν σε έναν μοναδικό, σύγχρονο σαρωτή PET/CT (Siemens Biograph mCT 21). Το πρωτόκολλο απεικόνισης περιλαμβάνει μια διαγνωστική σάρωση CT (κυρίως από τη βάση του κρανίου έως το μέσο του μηρού) με ενδοφλέβια ενίσχυση αντίθεσης στις περισσότερες περιπτώσεις, εκτός από ασθενείς με αντενδείξεις. Χρησιμοποιήθηκαν οι ακόλουθοι παράμετροι CT: αναφορική δόση 200 mAs, τάση σωλήνα 120 kV, ανακατασκευή επαναλαμβανόμενη με πάχος τομής 2 - 3 mm. Επιπλέον, πραγματοποιήθηκε ολόσωμη σάρωση FDG-PET 60 λεπτά μετά την ενδοφλέβια έγχυση 300-350 MBq 18F-FDG. Τα δεδομένα PET ανακατασκευάστηκαν χρησιμοποιώντας έναν αλγόριθμο προσδοχιών μεγιστοποίησης υποσυνόλου (OSEM) με 21 υποσύνολα και 2 επαναλήψεις και έναν Γκαουσιανό πυρήνα 2 mm και μέγεθος πλέγματος 400 x 400.



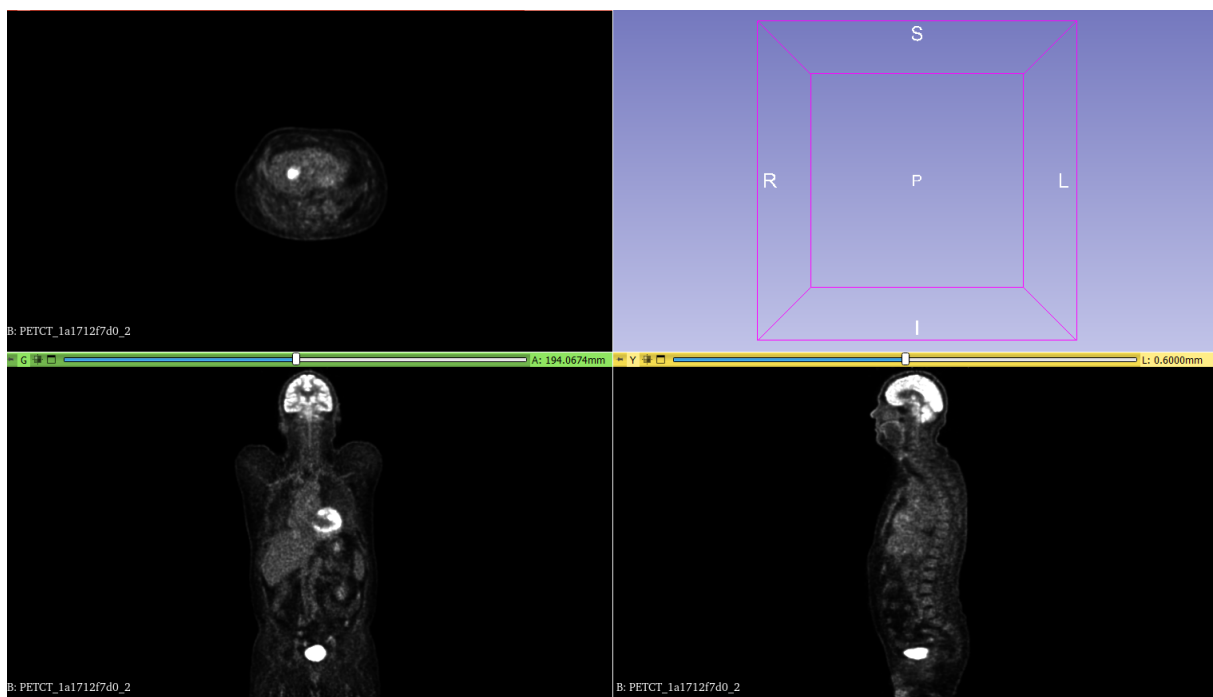
Εικόνα 21: Siemens Biograph mCT [44].

Όλα τα σύνολα δεδομένων αναλύθηκαν σε κλινικό πλαίσιο από έναν ακτινολόγο και έναν ιατρό πυρηνικής ιατρικής με συναίνεση, προσδιορίζοντας πρωτογενείς όγκους και μεταστάσεις σε κάθε σύνολο δεδομένων. Όλες οι πληγές FDG που ταυτίστηκαν ως κακοήθειες με βάση το ιστορικό του ασθενούς και προηγούμενες εξετάσεις χαρτογραφήθηκαν χειροκίνητα στις εικόνες PET με τρόπο τομή προς τομή από έναν αναγνώστη χρησιμοποιώντας ειδικό λογισμικό (πλατφόρμα απεικόνισης NORA, Πανεπιστήμιο του Φράιμπουργκ, Γερμανία).

Για τους σκοπούς της εργασίας από το σύνολο των δεδομένων αξιοποιήθηκαν 84 εικόνες (84 CT, 84 PET). Σκοπός είναι η ευθυγράμμιση των PET στις CT, για αυτό οι εικόνες CT αποτελούν το σύνολο των σταθερών εικόνων που ζευγαρώνουν με τις αντίστοιχες κινούμενες PET εικόνες ώστε να γίνει η ευθυγράμμισή τους. Παραδείγματα εικόνας CT και εικόνας PET για το ίδιο αντικείμενο μελέτης παρουσιάζονται παρακάτω. Οι εικόνες προβάλλονται με τη βοήθεια του προγράμματος Slicer 5.2.2:



Εικόνα 22: CT παράδειγμα ασθενούς.



Εικόνα 23: PET παράδειγμα ασθενούς.

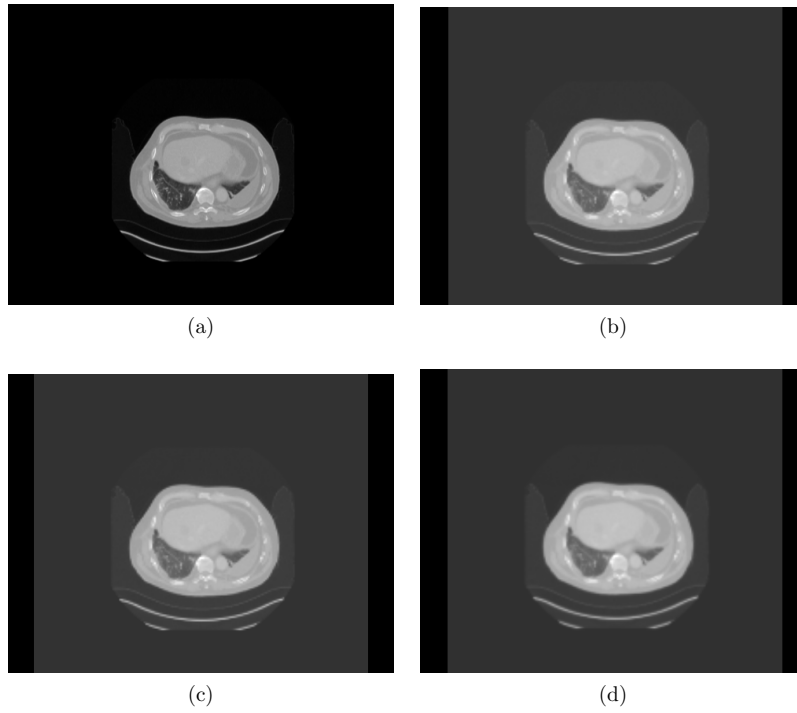
## 2.4 Προεπεξεργασία δεδομένων

Οι αρχικές εικόνες για να χρησιμοποιηθούν ως δεδομένα εκπαίδευσης, επεξεργάστηκαν και μετασχηματίστηκαν. Η αποτελεσματικότητα της ευθυγράμμισης στηρίζεται αρκετά στην ποιότητα των εικόνων εισόδου. Τα βήματα της προεπεξεργασίας των εικόνων, όπως τα υποβάθμιση (downsampling), κανονικοποίηση (normalizing) και εξομάλυνση (smoothing) καθιστούν τις εικόνες έτοιμες για τη διαδικασία της ευθυγράμμισης επισπεύδοντάς την, δεδομένου ότι ο

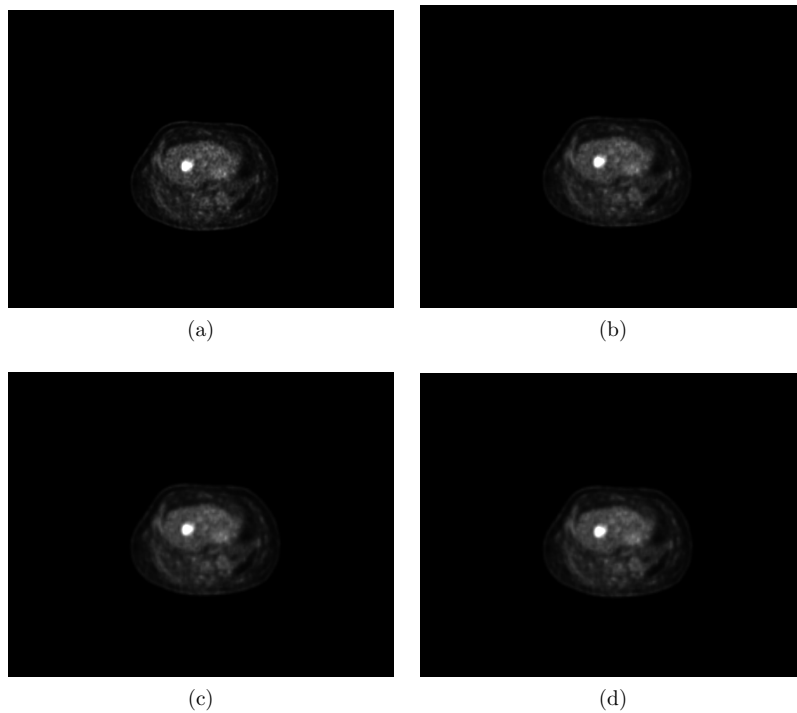
αλγόριθμος αναγνωρίζει ευκολότερα τα μοτίβα και τα αντιστοιχίζόμενα σημεία των. Επιπροσθέτως, στο πλαίσιο της ευθυγράμμισης ιατρικών εικόνων η διαδικασία αυτή είναι χρήσιμη διότι συντελεί στον μετριάσμό των επιπτώσεων των παραλλαγών που είναι εγγενείς στις διαδικασίες απεικόνισης, ιδιαίτερα όταν οι εικόνες προέρχονται από διαφορετικές χρονικές στιγμές, συνθήκες ή τρόπους (π.χ. CT, MRI, PET). Με τη βελτίωση της ομοιογένειας και τη μείωση του θορύβου τού συνόλου δεδομένων, αυτά τα βήματα προεπεξεργασίας ενισχύουν την ευρωστία και την αξιοπιστία των αποτελεσμάτων ευθυγράμμισης. Στο παρόν πείραμα, η διαδικασία ετοιμασίας των εικόνων για την εισαγωγή στους αλγόριθμους έγινε με τη βιβλιοθήκη SimpleITK. Ακολουθήθηκαν τα εξής βήματα:

1. Υποβάθμιση: προκειμένου να μειωθεί το υπολογιστικό κόστος και να εφαρμοστεί μια ομοιογένεια στα δεδομένα, εφαρμόστηκε η μέθοδος υποβιβασμού των εικόνων. Η συγκεκριμένη εφαρμογή είναι απαραίτητη καθώς τα δεδομένα εκπαίδευσης αποτελούνται από τρισδιάστατες ιατρικές εικόνες, υψηλών διαστάσεων με αποτέλεσμα να μη φτάνει η διαθέσιμη GPU στο περιβάλλον τού Google Colab Pro. Επίσης, δεδομένου ότι τα σύνολα των σταθερών και των κινούμενων εικόνων απεκτήθηκαν με διαφορετικούς τρόπους, η υποβάθμιση σταθεροποιεί την ανάλυση σε όλες τις εικόνες. Αυτό απορρέει σημαντικό για τη βελτίωση της αποτελεσματικότητας των μοντέλων. Οι εικόνες υποβαθμίστηκαν στο ίδιο μέγεθος (192,192,192) διατηρώντας τις αρχικές ανατομικές πληροφορίες όσο το δυνατόν περισσότερο, αλλά με μειωμένη την ανάλυση. Προκειμένου να γίνει αυτό προσδιορίστηκε το κέντρο τής εικόνας και χρησιμοποιήθηκε ως το κέντρο για το μετασχηματισμό Όιλερ. Ως μέθοδος μετασχηματισμού χρησιμοποιήθηκε η γραμμική παρεμβολή.
2. Εξομάλυνση: ένα ακόμη σημαντικό βήμα είναι η εξομάλυνση. Η συγκεκριμένη τεχνική είναι καίρια λόγω του ότι περιορίζει το θόρυβο στις εικόνες. Ο εν λόγω θόρυβος δύναται να έχει προκληθεί, για παράδειγμα, από ατέλειες των ανιχνευτών, από τις περιβαλλοντικές συνθήκες ή κι από λάθη εκπομπών. Μειώνοντας το θόρυβο επισημαίνονται σημαντικά χαρακτηριστικά στις εικόνες, όπως οι γωνίες ή οι περιοχές ενδιαφέροντος. Με αυτόν τον τρόπο βελτιώνεται η απόδοση των μοντέλων, διότι ο θόρυβος που εσφαλμένα αναγνωρίζεται ως χαρακτηριστικό έχει μειωθεί. Στις εικόνες τού πειράματος εφαρμόζεται η Γκαουσιανή εξομάλυνση με τον Γκαουσιανό πυρήνα να έχει τυπική απόκλιση δύο εικονοστοιχεία.
3. Κανονικοποίηση: τελευταίο στη διαδικασία τής προεπεξεργασίας των δεδομένων είναι η κανονικοποίηση των τιμών των εντάσεων των εικόνων. Χάρη στην κανονικοποίηση μπορεί να βελτιωθεί σημαντικά η σύγκλιση του αλγορίθμου, εφόσον χωρίς αυτή τα χαρακτηριστικά με τις μεγαλύτερες τιμές ποδηγετούν τη διαδικασία εκπαίδευσης με αποτέλεσμα ο αλγόριθμος να συγκλίνει πολύ αργότερα ή να συγκλίνει σε υποδεέστερες λύσεις. Επιπλέον, εξασφαλίζει ότι κάθε χαρακτηριστικό συντελεί σχεδόν ισότιμα στην τελική πρόβλεψη. Τέλος, διευκολύνει τη γενίκευση του μοντέλου, ήτοι το μοντέλο δε μαθαίνει ασυσχέτιστα μοτίβα από τα δεδομένα, ενώ επικεντρώνεται στην εκμάθηση των δομών και των μοτίβων που ενυπάρχουν στις εικόνες. Οι εικόνες στο παρόν σύνολο κανονικοποιήθηκαν έτσι ώστε οι τιμές έντασης να κυμαίνονται μεταξύ μηδέν κι ένα.

Η διαδικασία που αναγράφεται παραπάνω, ενσωματώνεται στον κώδικα όταν δημιουργείται η βάση δεδομένων για την κατασκευή των δεδομένων εκπαίδευσης και των δεδομένων επαλήθευσης. Επομένως, όταν ένας ζεύγος κινούμενης, σταθερής εικόνας εξετάζεται από αυτήν τη βάση, προτού δοθεί στον κατάλληλο Dataloader μορφοποιείται με τα παραπάνω βήματα. Στις Εικόνες 24,25 δίνονται οπτικά παραδείγματα αυτής της προεπεξεργασίας δεδομένων ξεκινώντας από την αρχική μη επεξεργασμένη εικόνα και οδεύοντας στην τελική μορφή μετά την εξομάλυνση, για το ίδιο αντικείμενο σε CT και PET.



Εικόνα 24: Παράδειγμα αποτελέσματος προεπεξεργασίας εικόνων για ένα αντικείμενο CT. Η οπτικοποίηση έγινε με τη χρήση τού 3D Slicer 5.2.2. (a) Αρχική μη επεξεργασμένη CT εικόνα. (b) Εικόνα CT έπειτα από την υποβάθμιση. (c) Εικόνα CT έπειτα από την κανονικοποίηση. (d) Εικόνα CT έπειτα από την εξομάλυνση.



Εικόνα 25: Παράδειγμα αποτελέσματος προεπεξεργασίας εικόνων για ένα αντικείμενο PET. Η οπτικοποίηση έγινε με τη χρήση τού 3D Slicer 5.2.2. (a) Αρχική μη επεξεργασμένη PET εικόνα. (b) Εικόνα PET έπειτα από την υποβάθμιση. (c) Εικόνα PET έπειτα από την κανονικοποίηση. (d) Εικόνα PET έπειτα από την εξομάλυνση.

## 2.5 Ευθυγράμμιση με SimpleITK

Η υλοποίηση με τη βιβλιοθήκη SimpleITK πραγματοποιήθηκε με μία πολυεπίπεδη Demons ευθυγράμμιση. Η διαδικασία αυτή, αντιστοιχίζει δύο εικόνες εφαρμόζοντας μια σειρά μετασχηματισμών, ώστε να ελαχιστοποιήσουν τη διαφορά μεταξύ των εικόνων. Η συγκεκριμένη μέθοδος χτίζει πυραμίδες των σταθερών και κινούμενων εικόνων με τη διαδοχική εξομάλυνση κι υποβάθμισή τους. Έτσι, δημιουργείται μια σειρά εικόνων με μειούμενες αναλύσεις. Αυτές οι πυραμίδες επιτρέπουν στη διαδικασία τής ευθυγράμμισης να ξεκινά με πιο τραχιές προσαρμογές σε χαμηλότερες αναλύσεις και σταδιακά να βελτιώνει το μετασχηματισμό σε υψηλότερες αναλύσεις. Η διαδικασία ευθυγράμμισης αρχίζει από το χαμηλότερο επίπεδο ανάλυσης εικόνας και χρησιμοποιώντας φίλτρα Demons με παραμέτρους τον αριθμό των επαναλήψεων, αν θα εξομαλυνθεί το πεδίο μετασχηματισμού και την τυπική απόκλιση για την εξομάλυνση. Αφού γίνει η ευθυγράμμιση στο χαμηλότερο επίπεδο, το εξαχθέν πεδίο μετασχηματισμού αναβαθμίζεται στην επόμενη ανάλυση της πυραμίδας. Η ευθυγράμμιση επαναλαμβάνεται σε αυτήν την ανάλυση, βελτιώνοντας το μετασχηματισμό σε αυτό το επίπεδο ανάλυσης. Η διαδικασία αυτή επαναλαμβάνεται μέχρι την κορυφή τής πυραμίδας. Το τελευταίο επίπεδο μετασχηματισμού, το οποίο αποκτάται στο υψηλότερο επίπεδο ανάλυσης ενσωματώνει τις εκτενείς μετατοπίσεις κατά εικονοστοιχείο, που απαιτούνται για την ευθυγράμμιση της κινούμενης εικόνας με τη σταθερή εικόνα σε ολόκληρη την εικόνα. Ως αλγόριθμος ευθυγράμμισης ορίζεται κι αρχικοποιείται ένα Demons φίλτρο. Εκτός από την ευθυγράμμιση με τα φίλτρα Demons έγινε ευθυγράμμιση αξιοποιώντας την κλάση ImageRegistrationMethod. Προκειμένου να επιλεγεί ο συνδυασμός εκείνος που θα ευθυγραμμίσει την κινούμενη εικόνα, διεξάγονται πολλαπλά διαδοχικά πειράματα, λαμβάνοντας υπόψη παραμετροποιήσεις των δομικών στοιχείων (αρχική ευθυγράμμιση, βελτιστοποιητής, συνάρτηση παρεμβολής, αριθμός καλάθων, ποσοστό υποδειγματισμού).

## 2.6 Εκπαίδευση με το Voxelmorph

Το Voxelmorph που υλοποιήθηκε πλαισιώνει το μοντέλο που περιγράφει στο θεωρητικό μέρος τής εργασίας [30]. Αρχικά, δημιουργείται μια βάση, η οποία εμπεριέχει την προεπεξεργασία των εικόνων και το διαχωρισμό τους σε δεδομένα εκπαίδευσης κι επαλήθευσης. Επιλέγεται το 70% τού συνόλου δεδομένων να αποτελεί τα δεδομένα εκπαίδευσης και το 30% να αποτελεί τα δεδομένα επαλήθευσης, ενώ ο αριθμός των κομματιών (batch) στους Dataloaders είναι ένα. Από τη διαδικασία αυτή προκύπτουν ζεύγη σταθερών, κινούμενων εικόνων κάθε μία από τις οποίες έχει διαστάσεις [1, 1, 192, 192], όπου [Κομμάτια,Κανάλια,Ύψος,Μήκος,Βάθος]. Οι δύο εικόνες (σταθερή, κινούμενη) συγχωνεύονται επάνω στα κανάλια (με αποτέλεσμα η νέα εικόνα να έχει διαστάσεις [1, 2, 192, 192, 192]) και τροφοδοτούνται ως είσοδος στο πρώτο μοντέλο, το Unet. Το Unet δομείται από τον κωδικοποιητή και τον αποκωδικοποιητή. Ο κωδικοποιητής αποτελείται από πέντε συναπτά συνελικτικά επίπεδα, ακολουθούμενα από μία συνάρτηση ενεργοποίησης LeakyReLU. Κάθε ένα συνελικτικό επίπεδο εφαρμόζει μια συνέλιξη με πυρήνα [3 × 3 × 3] και βήμα ένα. Η υποβάθμιση γίνεται με μέγιστη συγκέντρωση (Max Pooling) με πυρήνα [2 × 2 × 2]. Ο αποκωδικοποιητής αποτελείται από έξι αντίστροφα συνελικτικά επίπεδα ακολουθούμενα από μια LeakyReLU, εμπλουτισμένα με τις συνδέσεις παράβλεψης (skip connections) τού κωδικοποιητή. Το τελικό επίπεδο είναι μια συνέλιξη, η οποία μειώνει το βάθος των καναλιών σε τρία λόγω των τριών διαστάσεων των εικόνων (τελικός πίνακας μετασχηματισμού, κάθε ογκοστοιχείο δείχνει τη θέση που  $x, y, z$ ). Ο κωδικοποιητής συμβάλει στην αύξηση του βάθους των πινάκων χαρακτηριστικών και μειώνει τις χωρικές διαστάσεις, ενώ ο αποκωδικοποιητής μειώνει τους πίνακες χαρακτηριστικών κι αυξάνει τις χωρικές διαστάσεις. Στον Πίνακα 2 αναγράφεται όλη η δομή τού Unet, με τις αντίστοιχες διαστάσεις των εικόνων έπειτα από κάθε επίπεδο, καθώς κι οι παράμετροι προς εκπαίδευση. Οι συνολικές παράμετροι εκπαίδευσης είναι 326, 083.

Εφόσον παραχθεί ο πίνακας χαρακτηριστικών (με διαστάσεις [1, 3, 192, 192, 192]) εφαρμόζεται πάνω στην κινούμενη εικόνα για να παραχθεί το τελικό μετασχηματισμένο αποτέλεσμα. Αρχικά, ο πίνακας μετασχηματισμού υποβαθμίζεται με παράγοντα δύο. Αυτό σημαίνει ότι μειώνεται η ανάλυση τού πεδίου μετασχηματισμού. Κάθε ογκοστοιχείο στην κινούμενη εικόνα πρέπει να αντιστοιχιστεί με ένα ογκοστοιχείο στη σταθερή εικόνα. Επίσης, κάθε ογκοστοιχείο στο πεδίο μετασχηματισμού περιέχει ένα διάνυσμα, το οποίο προσδιορίζει την κατεύθυνση και το βεληνεκές για κάθε κίνηση. Επαναπροσδιορίζοντας το πεδίο μετασχηματισμού, για παράδειγμα, με παράγοντα δύο, κάθε διάσταση στο εξαχθέν πεδίο θα υποδιπλασιαστεί και θα έχει επιρροή σε μεγαλύτερο κομμάτι τής αρχικής εικόνας (η μετακίνηση

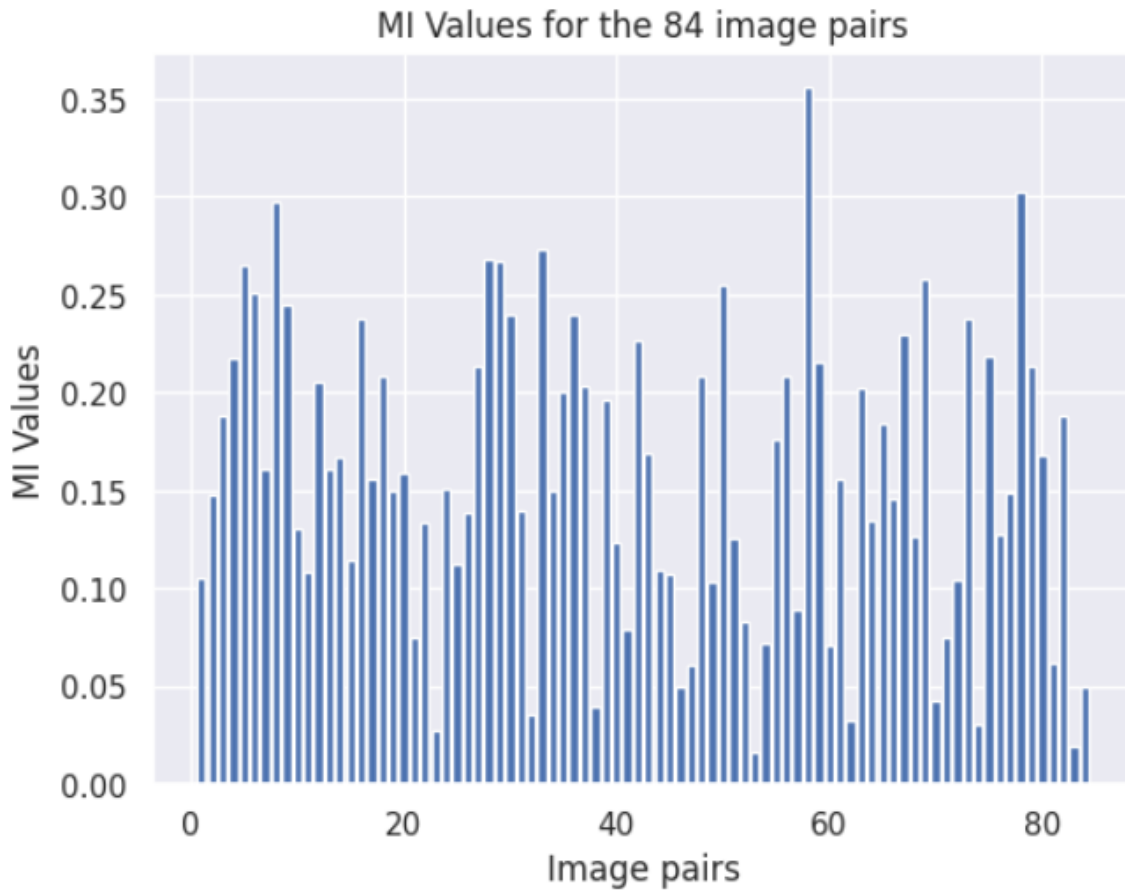
Επίπεδα	Διαστάσεις Εξόδου	Αριθμός Παραμέτρων
Είσοδος	[2,192,192,192]	0
Επίπεδο Κωδικοποιητή 1	[16,192,192,192]	880
Επίπεδο Κωδικοποιητή 2	[32,96,96,96]	13,856
Επίπεδο Κωδικοποιητή 3	[32,48,48,48]	27,680
Επίπεδο Κωδικοποιητή 4	[32,24,24,24]	27,680
Επίπεδο Κωδικοποιητή 5	[32,12,12,12]	27,680
Επίπεδο Αποκωδικοποιητή 1	[32,24,24,24]	55,328
Επίπεδο Αποκωδικοποιητή 2	[32,48,48,48]	55,328
Επίπεδο Αποκωδικοποιητή 3	[32,96,96,96]	55,328
Επίπεδο Αποκωδικοποιητή 4	[32,192,192,192]	41,504
Επίπεδο Αποκωδικοποιητή 5	[16,192,192,192]	13,840
Επίπεδο Αποκωδικοποιητή 6	[16,192,192,192]	6,928
Επίπεδο Εξόδου	[3,192,192,192]	51

Πίνακας 2: Επίπεδα Unet, διαστάσεις πινάκων χαρακτηριστικών κι αριθμός παραμέτρων προς εκπαίδευση.

πλέον τού ογκοστοιχείου στην κινούμενη εικόνα θα είναι το αποτέλεσμα της παρεμβολής των ογκοστοιχείων τής περιοχής). Για να γίνει πλήρως κατανοητό, έστω ότι ο πίνακας μετασχηματισμού πριν την υποβάθμιση αναπαριστά ένα γεωγραφικό χάρτη με όλα τα μονοπάτια σε ένα βουνό. Ο χάρτης αυτός δείχνει ακριβώς τις κινήσεις που πρέπει να γίνουν για να περιπλανηθεί ένας ορειβάτης. Όταν ο χάρτης υποβαθμιστεί, θα δείχνει μόνο τα κύρια μονοπάτια σε κομβικές πτυχές τού βουνού. Η λεπτομερής καθοδήγηση θα έχει χαθεί, όμως η κατεύθυνση υπάρχει κι είναι πιο εύκολο να ακολουθηθεί σε μια πιο ευρεία περιοχή. Εν συνεχεία, το καινούργιο πεδίο ολοκληρώνεται, με σκοπό να παραχθεί ένα ομαλό, συνεχές διανυσματικό πεδίο. Αυτό είναι σημαντικό ώστε να μην εμφανιστούν παρεμβολές, όπως οδοντωτές ακμές ή μη ρεαλιστικές παραμορφώσεις, δυσχεραίνοντας έτσι την απόδοση της ευθυγράμμισης. Μετά την ολοκλήρωση, το πεδίο επανέρχεται στις αρχικές του διαστάσεις ώστε με τη σειρά του να εφαρμοστεί (warp) στην κινούμενη εικόνα. Για να γίνει το warp, πρώτα δημιουργείται ένα πλέγμα που αναπαριστά τις συντεταγμένες κάθε ογκοστοιχείου στην εικόνα. Το πρωταρχικό πλέγμα αντιστοιχίζεται στις θέσεις των ογκοστοιχείων τής σταθερής εικόνας. Το πεδίο μετασχηματισμού εφαρμόζεται στο αρχικό αυτό πλέγμα προσθέτοντας τα διανύσματα μετασχηματισμού στις αρχικές συντεταγμένες. Το προσαρμοσμένο πλέγμα αναπαριστά τις καινούργιες θέσεις από τις οποίες δειγματίζονται τα ογκοστοιχεία για να δημιουργήσουν τη μετασχηματισμένη εικόνα. Τα σημεία αυτά, που έχουν προσαρμοστεί στο πλέγμα, δεν αντιστοιχούν γενικά στα σημεία τής σταθερής εικόνας, αλλά μπορούν να υπάρχουν ανάμεσά τους. Για το λόγο αυτόν, χρησιμοποιείται μια μέθοδος παρεμβολής, η διγραμμική (bilinear). Αυτή η διαδικασία επαναλαμβάνεται για κάθε μετασχηματισμένο σημείο πλέγματος, ώστε να παραχθούν οι τελικές τιμές ογκοστοιχείων στην έξοδο. Εν κατακλείδι, το STN πρώτα υπολογίζει ένα καινούργιο σύνολο σημείων βασισμένο στον πίνακα μετασχηματισμού τού Unet κι έπειτα παρεμβάλει την κινούμενη εικόνα σε αυτά τα σημεία για να παράξει την ευθυγραμμισμένη εικόνα. Στο Παράρτημα (Εικόνα 40) φαίνεται όλη η δομή τού μοντέλου Voxelmorph που χρησιμοποιήθηκε.

## 2.7 Αποτελέσματα Ευθυγράμμισης

Εξετάστηκαν δύο εκδοχές ευθυγράμμισης, η μεν πρώτη με μια συμβατική μέθοδο ευθυγράμμισης χρησιμοποιώντας τα Demons φίλτρα και την ImageRegistrationMethod τής SimpleITK, η δε δεύτερη χρησιμοποιώντας έναν αλγόριθμο Βαθιάς Μάθησης το Voxelmorph. Προκειμένου να υπάρξει μία βάση σύγκρισης, υπολογίζονται οι αμοιβαίες πληροφορίες για τις 84 εικόνες πριν την ευθυγράμμιση (Εικόνα 26). Η μέγιστη τιμή που παρατηρείται είναι η  $\sim 0.3557$  κι η ελάχιστη είναι  $\sim 0.0159$ . Ο μέσος όρος MI για τις 84 εικόνες ανέρχεται σε  $\sim 0.1565$ . Οι μετέπειτα αποφάνσεις περί ποιότητας σύντηξης των εικόνων, βασίζονται σε αυτόν το μέσο όρο.



Εικόνα 26: Τιμές αμοιβαίας πληροφορίας για όλα τα ζευγάρια σταθερών, κινούμενων εικόνων πριν την ευθυγράμμιση.

### 2.7.1 SimpleITK με Demons φίλτρα

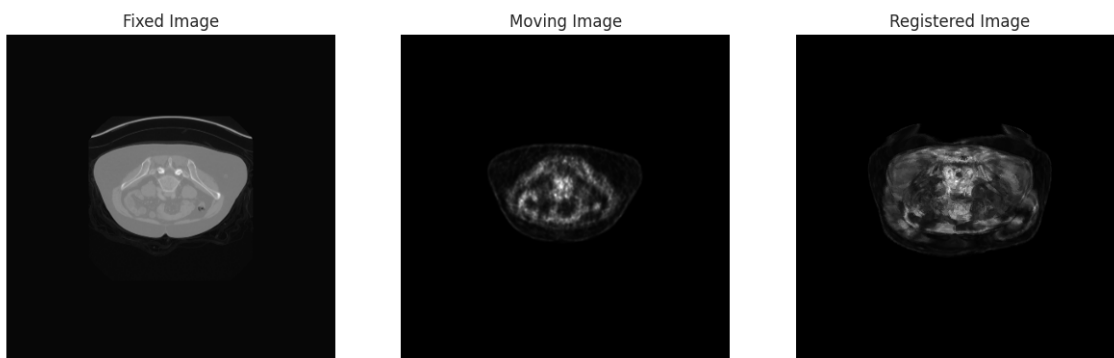
Για την πρώτη περίπτωση ευθυγράμμισης και τα Demons φίλτρα εξετάζονται διαφορετικές συνθήκες και παραμετροποιήσεις, οι οποίες φαίνονται στον Πίνακα 3.

Φίλτρο Demon	Επαναλήψεις	Εξομάλυνση	Τυπική απόκλιση	Παρεμβολή	MI
FastSymmetricForces	120	Ναι	1.5	Γραμμική	0.3938
FastSymmetricForces	120	Ναι	2.0	Γραμμική	0.3867
FastSymmetricForces	120	Ναι	2.5	Γραμμική	0.3605
FastSymmetricForces	120	Ναι	1.5	Κοντινότερος Γείτονας	0.3964
FastSymmetricForces	110	Ναι	1.5	Κοντινότερος Γείτονας	0.4077
DemonsRegistrationFilter	120	Ναι	1.5	Κοντινότερος Γείτονας	0.3984
DemonsRegistrationFilter	110	Ναι	1.5	Κοντινότερος Γείτονας	0.3864
Diffeomorphic	110	Ναι	1.5	Κοντινότερος Γείτονας	0.3816
SymmetricForces	110	Ναι	1.0	Κοντινότερος Γείτονας	0.4402
SymmetricForces	110	Ναι	1.5	Κοντινότερος Γείτονας	0.4310
SymmetricForces	120	Ναι	0.5	Κοντινότερος Γείτονας	0.4586

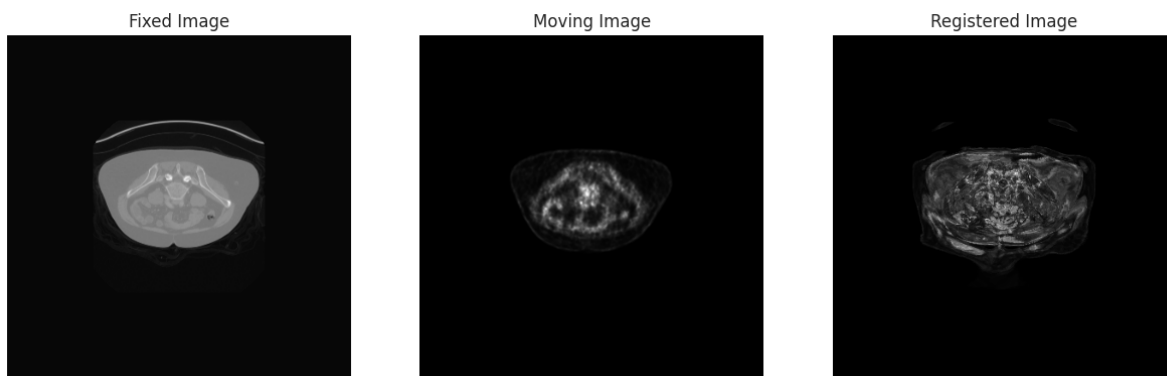
Πίνακας 3: Παραμετροποιήσεις ευθυγράμμισης με τα Demons.



Από όλους τους συνδυασμούς που ερευνήθηκαν καλύτερα αποτελέσματα έδωσε η χρήση τού φίλτρου `SymmetricForcesDemonsRegistrationFilter` με 120 επαναλήψεις, τυπική απόκλιση 0.5 και συνάρτηση παρεμβολής τον Κοντινότερο Γείτονα. Οι δύο καλύτερες ευθυγραμμίσεις φαίνονται στις Εικόνες 27, 28.



Εικόνα 27: Αποτελέσματα ευθυγράμμισης με SimpleITK για το μοντέλο με Demons φίλτρο `SymmetricForcesDemonsRegistrationFilter`, επαναλήψεις 110 τυπική απόκλιση 1.0 και συνάρτηση παρεμβολής τον Κοντινότερο Γείτονα. Αριστερά η σταθερή εικόνα, στη μέση η κινούμενη και δεξιά η ευθυγραμμισμένη.



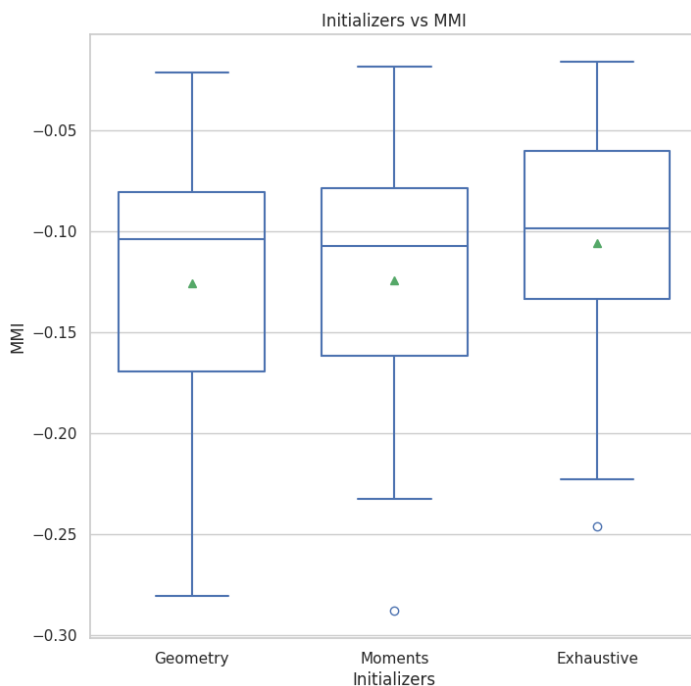
Εικόνα 28: Αποτελέσματα ευθυγράμμισης με SimpleITK για το μοντέλο με Demons φίλτρο `SymmetricForcesDemonsRegistrationFilter`, επαναλήψεις 120 τυπική απόκλιση 0.5 και συνάρτηση παρεμβολής τον Κοντινότερο Γείτονα. Αριστερά η σταθερή εικόνα, στη μέση η κινούμενη και δεξιά η ευθυγραμμισμένη.

### 2.7.2 SimpleITK με `ImageRegistrationMethod`

Για τη συμβατική μέθοδο ευθυγράμμισης με την `ImageRegistrationMethod` έγινε, επίσης, ένας έλεγχος παραμέτρων. Ο έλεγχος έγινε πάνω σε δεκαπέντε αντικείμενα μελέτης. Εξετάζονται διαδοχικά ο αρχικός μετασχηματισμός (`Initial transform`), ο βελτιστοποιητής (`Optimizer`), η συνάρτηση παρεμβολής (`Interpolator`), ο αριθμός των bins (`Bins`) και τέλος το ποσοστό υποδειγματισμού (`Sampling percentage`). Για κάθε ένα από αυτά, η διερεύνηση γίνεται πάνω στην κάλλιστη απόδοση της προηγούμενη εξέτασης.

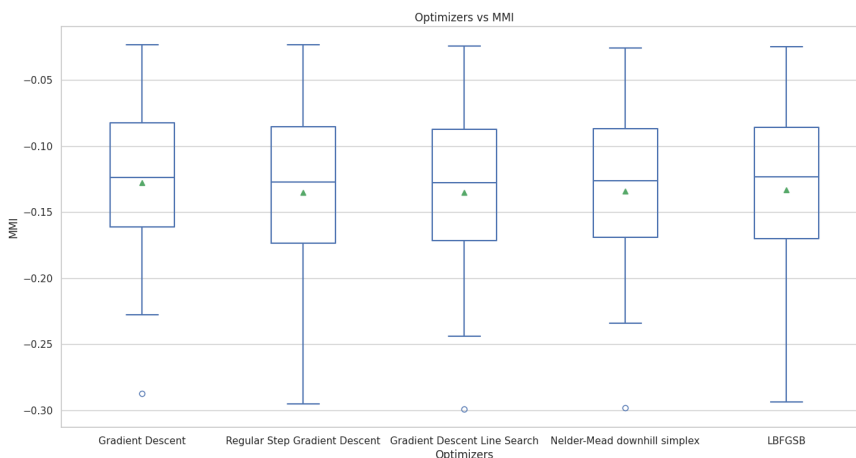
Ο αρχικός μετασχηματισμός είναι σημαντικός λόγω του ότι επηρεάζει το χρόνο ευθυγράμμισης και τη σύγκλιση της στο σωστό ελάχιστο. Ιδανικά, είναι επιθυμητό ο αρχικός μετασχηματισμός να είναι κοντά στο βέλτιστο τελικό μετασχηματισμό. ώστε να εξασφαλίζεται η σύγκλιση σε περιορισμένο χρονικό διάστημα [51]. Για το λόγο αυτό, δοκιμάζονται τρεις αρχικοποιήσεις, εκ των οποίων οι δύο βγαίνουν με τον `CenteredTransformInitializerFilter` κι αυτές είναι η μέθοδος `Geometry` (αντιστοιχίζει τα φυσικά κέντρα των εικόνων) κι η μέθοδος `Moments` (αντιστοιχίζει τη μέση τιμή των εντάσεων ογκοστοιχείων των εικόνων). Η τρίτη αρχικοποίηση συνακολουθεί με τη μέθοδο `Exhaustive` (υπολογίζει το μέτρο ομοιότητας σε ένα πλέγμα στο χώρο των παραμέτρων, κεντραρισμένο στις παραμέτρους τού αρχικού μετασχηματισμού). Ως μετρική ελέγχου ορίζεται η `MattesMutualInformation` (MMI). Οι αλγοριθμικές

διαδικασίες βασίζονται στα Notebooks των Yaniv et al. [51]. Στην Εικόνα 29 φαίνονται τα αποτελέσματα για αυτές τις τρεις μεθόδους. Οι Geometry, Moments κυμαίνονται στο ίδιο, περίπου, εύρος, με μέσες τιμές  $-0.1510$  και  $-0.1523$  αντιστοίχως. Επιλέγεται η Geometry ως μέθοδος αρχικοποίησης, λόγω των περισσότερο διευρυμένων άκρων της.



Εικόνα 29: Αποτελέσματα ευθυγράμμισης για τους διαφορετικούς τρόπους αρχικοποίησης.

Οι βελτιστοποιητές είναι εξίσου σημαντικοί για την ευθυγράμμιση. Ανάλογα με το είδος των εικόνων και την κατανομή των εντάσεων, η απόδοσή τους διαφέρει. Χρησιμοποιούνται τόσο Gradient Free όσο και Gradient μέθοδοι. Από την πρώτη κατηγορία χρησιμοποιείται ο Nelder-Mead downhill simplex (Amoeba), ενώ από τη δεύτερη οι Gradient Descent, Gradient Descent Line Search, Regular Step Gradient Descent και LBFGSB. Στον Πίνακα 4 φαίνονται όλοι οι συνδυασμοί για την κάθε περίπτωση και στην Εικόνα 30 φαίνονται τα αποτελέσματα για 15 ασθενείς. Από τα αποτελέσματα, ως καλύτερος βελτιστοποιητής επιλέγεται ο Gradient Descent Line Search.



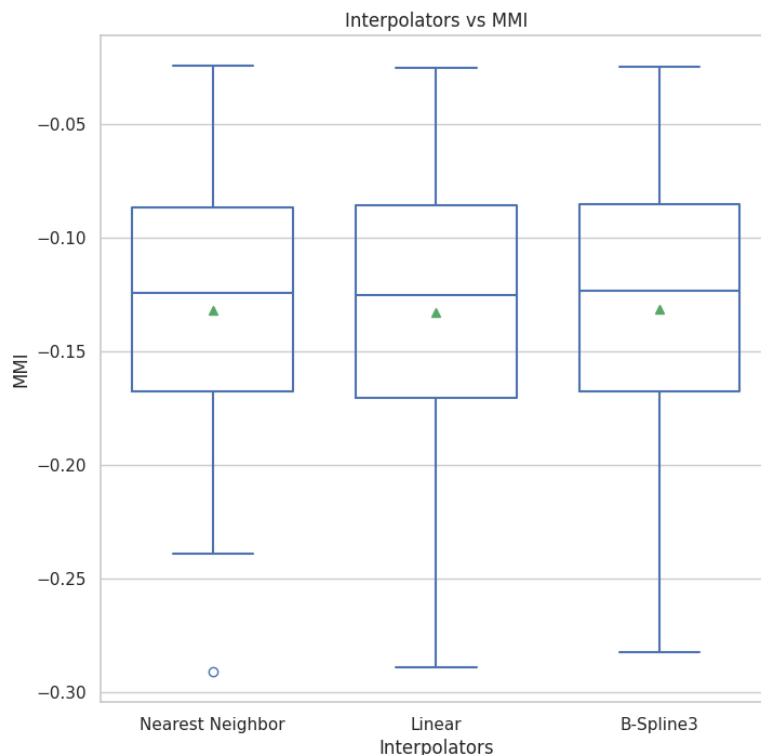
Εικόνα 30: Αποτελέσματα ευθυγράμμισης για τους διαφορετικούς βελτιστοποιητές.

Επόμενο βήμα είναι ο έλεγχος διαφορετικών συναρτήσεων παρεμβολής. Εξετάζονται οι Κοντινότερος Γείτονας (Nearest Neighbor), Γραμμική (Linear) και B-Spline3. Στην Εικόνα 31 φαίνονται οι αποδόσεις τους. Από

Βελτιστοποιητής	Παράμετροι	Τιμές
Nelder-Mead downhill simplex	simplexDelta	0.01
	Number of iterations	300
Gradient Descent	learningRate	1
	numberOfIterations	100
	convergenceMinimumValue	1e-5
	convergenceWindowSize	5
	estimateLearningRate	EachIteration
Gradient Descent Line Search	learningRate	1e-4
	numberOfIterations	100
	convergenceMinimumValue	1e-3
	convergenceWindowSize	5
	lineSearchLowerLimit=0.5,lineSearchUpperLimit	0.5
	lineSearchEpsilon	5
	estimateLearningRate	0.01
Regular Step Gradient Descent	learningRate	EachIteration
	numberOfIterations	1
	minStep	100
	,relaxationFactor	1e-3
	gradientMagnitudeTolerance	0.5
	estimateLearningRate	1e-4
LBFGSB	gradientConvergenceTolerance	EachIteration
	numberOfIterations	1e-5
	maximumNumberOfCorrections	100
	maximumNumberOfFunctionEvaluations	2
		100

Πίνακας 4: Παραμετροποιήσεις ευθυγράμμισης για τους βελτιστοποιητές.

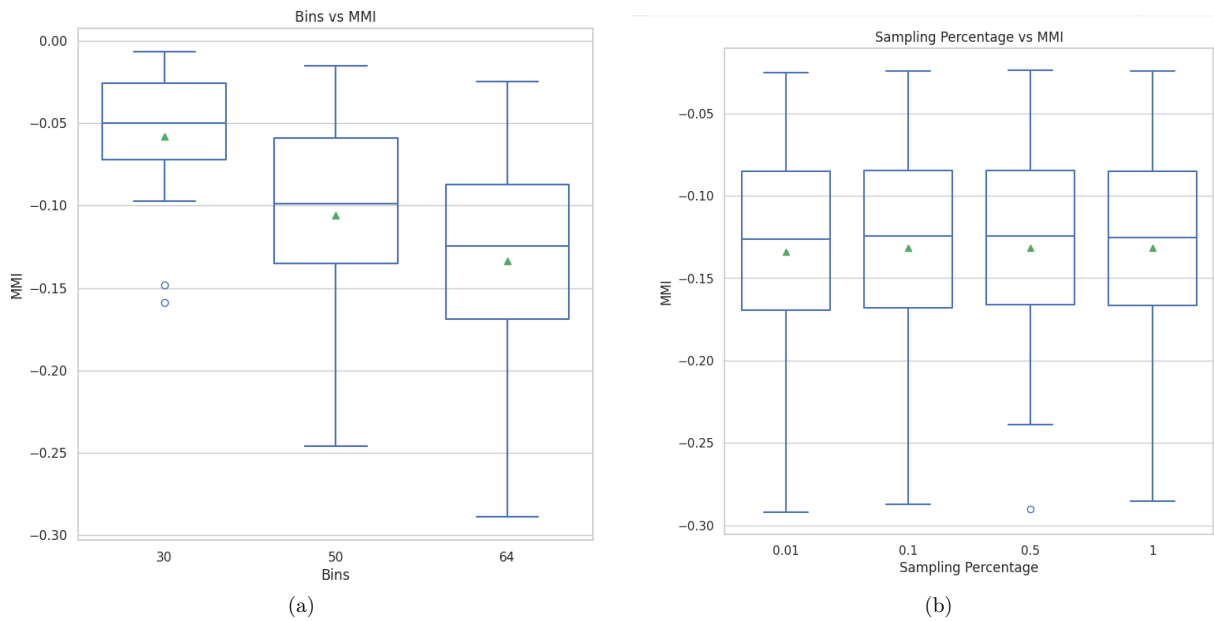
αυτούς, επιλέγεται η Γραμμική παρεμβολή. Τέλος, ερευνάται ο αριθμός των καλαθιών (Bins) και το ποσοστό



Εικόνα 31: Αποτελέσματα ευθυγράμμισης για τις διαφορετικές συναρτήσεις παρεμβολής.

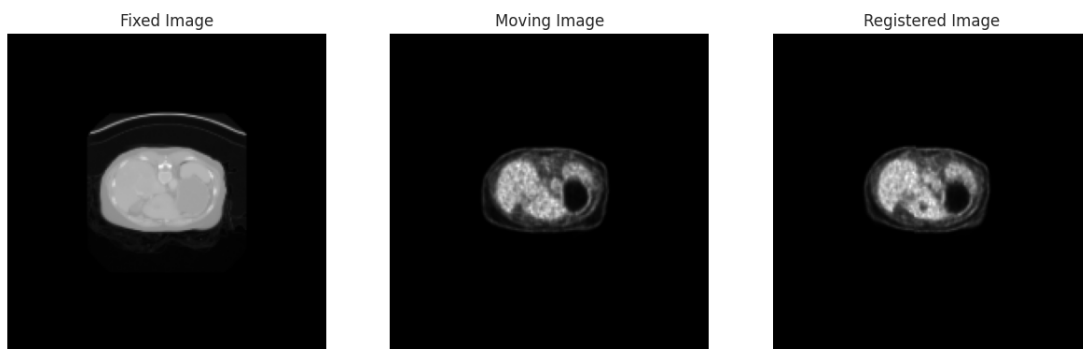
υποδειγματισμού (Sampling percentage). Όπως φαίνεται στην Εικόνα 32 (a) όσο αυξάνεται ο αριθμός των καλαθιών τόσο καλύτερεύει η ευθυγράμμιση. Έτσι, η τελική επιλογή είναι 64. Αντιθέτως, τα καλύτερα αποτελέσματα όσον

αφορά το ποσοστό εμφανίζονται με τιμή 0.01.



Εικόνα 32: Αποτελέσματα ευθυγράμμισης για διαφορετικά bins και ποσοστά υποδειγματισμού.

Στην Εικόνα 33 απεικονίζονται τα αποτελέσματα τού καλύτερου συνδυασμού ευθυγράμμισης με την ImageRegistrationMethod.



Εικόνα 33: Αποτελέσματα ευθυγράμμισης με SimpleITK για το καλύτερο μοντέλο τής ImageRegistrationMethod. Αριστερά η σταθερή εικόνα, στη μέση η κινούμενη και δεξιά η ευθυγραμμισμένη.

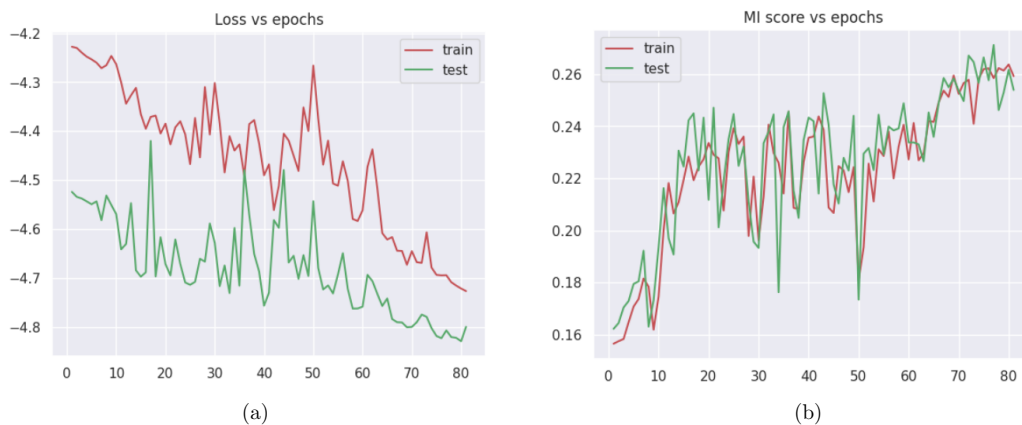
### 2.7.3 Συμπεράσματα για την ευθυγράμμιση με SimpleITK

Τα αποτελέσματα της ευθυγράμμισης με τη μέθοδο ImageRegistrationMethod είναι αρκετά υποδεέστερα της ευθυγράμμισης με τα φίλτρα Demons. Στην καλύτερη των περιπτώσεων, έπειτα από τη διεξοδική έρευνα των κομματιών που απαρτίζουν την ευθυγράμμιση με χρήση τής ImageRegistrationMethod, ο αλγόριθμος πετυχαίνει μία τιμή αμοιβαίας πληροφορίας ίση με 0.1986. Ακόμη, γίνεται αντιληπτό μέσα από τη διαδικασία τής παραμετροποίησης, ότι παρά τις διαφοροποιήσεις που γίνονται, η απόδοση δεν αλλάζει σε μεγάλο βαθμό. Αυτό υποδεικνύει πως η χρήση τής συγκεκριμένης τεχνικής δεν ταιριάζει στα δεδομένα. Εμφανώς, καλύτερα αποτελέσματα πετυχαίνονται με τη χρήση των φίλτρων Demons, καθώς εξειδικεύονται στις παραμορφωτικές (deformable) ευθυγραμμίσεις. Οι τιμές MI υποδηλώνουν ότι η αλγοριθμική διαδικασία μαθαίνει τα χαρακτηριστικά των εικόνων κι επιτυγχάνει καλύτερη ευθυγράμμιση (κυρίως για τα Demons φίλτρα). Ωστόσο, οι τιμές αυτές δεν είναι υψηλές. Αυτό οφείλεται στη φύση των δεδομένων, καθώς οι εικόνες ανήκουν σε διαφορετικούς τρόπους λήψης (modalities), κάτι που σημαίνει

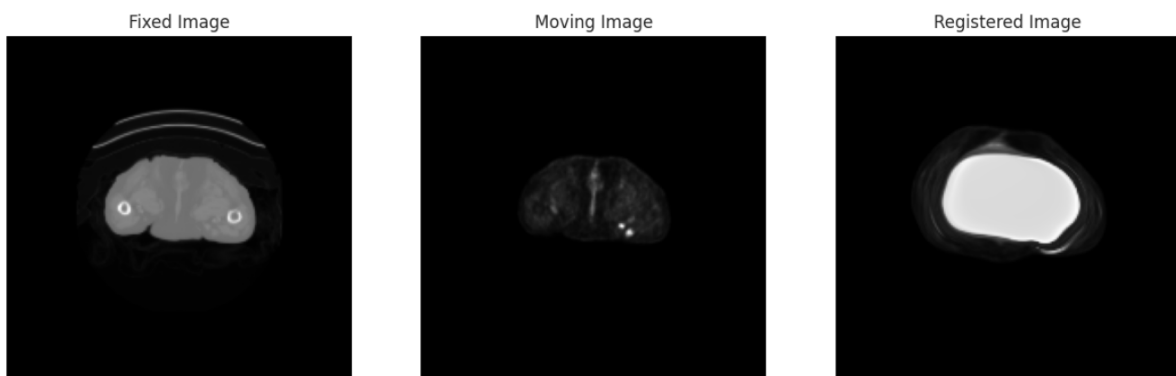
ότι απαθανατίζουν διαφορετικές φυσικές ιδιότητες, οδηγώντας σε εικόνες με πολύ διαφορετικές εμφανίσεις και κατανομές έντασης. Αυτή η μεταβλητότητα επηρεάζει σημαντικά την ευθυγράμμιση για διάφορους λόγους. Ο πρώτος λόγος έγκειται στην πολυτροπική φύση των δεδομένων. Οι εικόνες έχουν διαφορετικές εντάσεις για την ίδια ανατομική δομή κι αυτό δυσκολεύει τη χρήση απλών μετρικών βασισμένων στις εντάσεις. Έπειτα, η ταυτοποίηση χαρακτηριστικών ή δομών μεταξύ των εικόνων δυσχεραίνεται λόγω του τρόπου που αναπαρίσταται η κάθε δομή κι ο κάθε ιστός στους διαφορετικούς τρόπους λήψης των εικόνων (μία λήψη CT οπτικοποιεί καλύτερα τις ανατομικές δομές τού σώματος, ενώ μια λήψη PET παρατηρεί τις μεταβολικές διαδικασίες στο σώμα). Τέλος, οι πολυτροπικές εικόνες πιθανώς να έχουν αρκετές διαφορές όσον αφορά χαρακτηριστικά αντίθεσης κι υφής. Αυτές οι διαφορές εμποδίζουν την αποτελεσματικότητα των παραδοσιακών αλγορίθμων ευθυγράμμισης που βασίζονται σε άμεσες συγκρίσεις εντάσεων εικονοστοιχείων ή ογκοστοιχείων.

#### 2.7.4 Voxelmorph

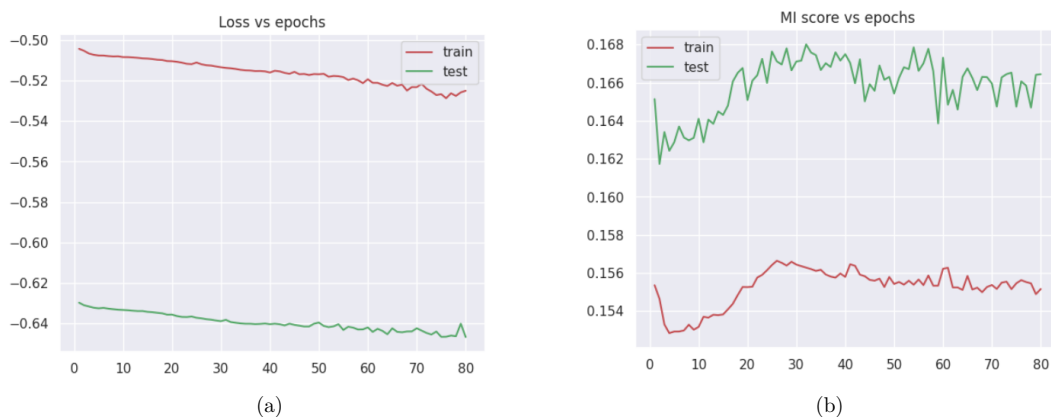
Για το Voxelmorph έτρεξαν δύο κύρια μοντέλα με 80 εποχές, βελτιστοποιητή (optimizer) τον Adam Optimizer, ρυθμό μάθησης (learning rate) ίσο με  $10^{-4}$ , παράγοντα ρύθμισης (regularization parameter)  $\lambda$  ίσο με 0.01 και συνάρτηση κόστους το MSE και το NCC, όπως αναλύονται στο Μέρος Α. Τα ποσοτικά αποτελέσματα των εκπαιδεύσεων φαίνονται στις Εικόνες 34 για το MSE και 35 για το NCC. Ποιοτικές απεικονίσεις φαίνονται στις Εικόνες 36 για το MSE και 37 για το NCC.



Εικόνα 34: Αποτελέσματα εκπαίδευσης Voxelmorph με συνάρτηση κόστους (loss function) το MSE, αριθμό εποχών 80 και  $\lambda = 0.01$ . Στην εικόνα (a) φαίνεται η εξέλιξη της συνάρτησης κόστους σε σχέση με τις εποχές, ενώ στην εικόνα (b) φαίνεται η εξέλιξη της μετρικής MI με την πρόοδο των εποχών.



Εικόνα 36: Αποτελέσματα ευθυγράμμισης με Voxelmorph για το κόστος MSE με  $\lambda = 0.01$ . Αριστερά η σταθερή εικόνα, στη μέση η κινούμενη και δεξιά η ευθυγραμμισμένη.



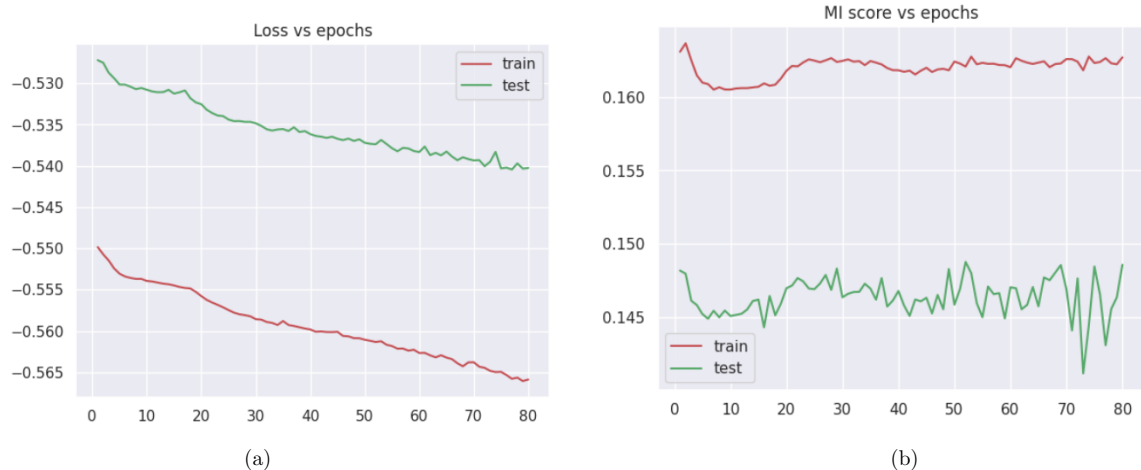
Εικόνα 35: Αποτελέσματα εκπαίδευσης Voxelmorph με συνάρτηση κόστους (loss function) το NCC, αριθμό εποχών 80 και  $\lambda = 0.01$ . Στην εικόνα (a) φαίνεται η εξέλιξη της συνάρτησης κόστους σε σχέση με τις εποχές, ενώ στην εικόνα (b) φαίνεται η εξέλιξη της μετρικής MI με την πρόοδο των εποχών.



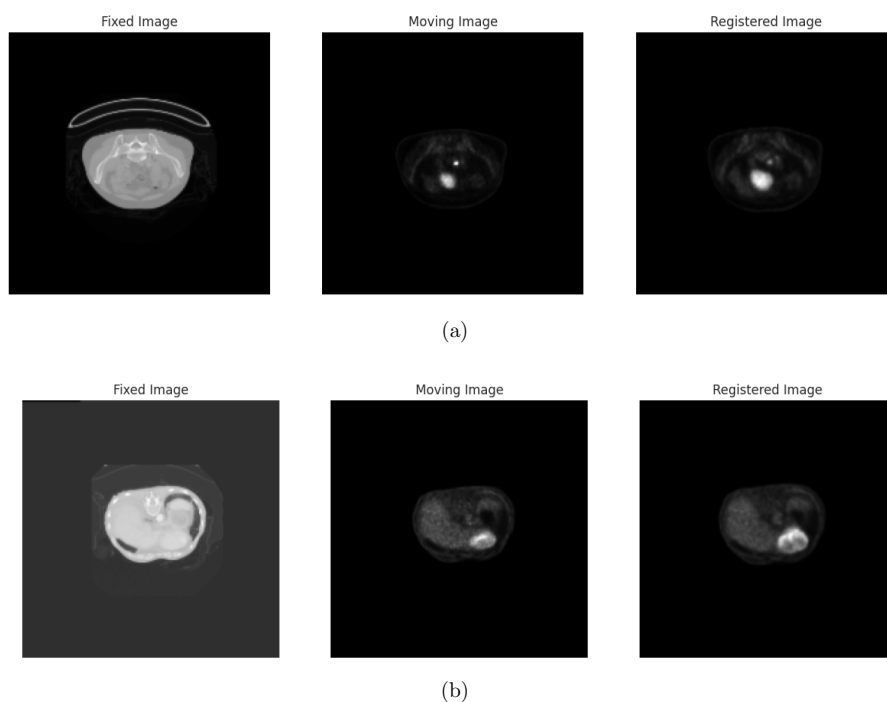
Εικόνα 37: Αποτελέσματα ευθυγράμμισης με Voxelmorph για το κόστος NCC με  $\lambda = 0.01$ . Αριστερά η σταθερή εικόνα, στη μέση η κινούμενη και δεξιά η ευθυγραμμισμένη.

Πέραν αυτών των δύο βασικών πειραμάτων, έγιναν κι άλλες δοκιμές, ώστε να διερευνηθεί ποιοι παράγοντες επηρεάζουν περισσότερο την απόδοση των μοντέλων και πώς αυτή αλλάζει, σύμφωνα με τις αλλαγές που γίνονται στις υπερπαραμέτρους. Οι αλλαγές επικεντρώνονται στον παράγοντα ρύθμισης και στον αριθμό των εποχών. Παρακάτω απαριθμούνται αυτές οι τροποποιήσεις:

1. Συνάρτηση κόστους NCC,  $\lambda$  ίσο με 0.1, εποχές 80, του οποίου τα αποτελέσματα οπτικοποιούνται ποσοτικά στην Εικόνα 38.
2. Συνάρτηση κόστους MSE,  $\lambda$  ίσο με 0.6, εποχές 50, των οποίων τα αποτελέσματα οπτικοποιούνται ποιοτικά στην Εικόνα 39 (a).
3. Συνάρτηση κόστους MSE,  $\lambda$  ίσο με 0.6, εποχές 80, των οποίων τα αποτελέσματα οπτικοποιούνται ποιοτικά στην Εικόνα 39 (b).



Εικόνα 38: Αποτελέσματα εκπαίδευσης Voxelmorph με συνάρτηση κόστους (loss function) το NCC, αριθμό εποχών 80 και  $\lambda = 0.1$ . Στην εικόνα (a) φαίνεται η εξέλιξη της συνάρτησης κόστους σε σχέση με τις εποχές, ενώ στην εικόνα (b) φαίνεται η εξέλιξη της μετρικής MI με την πρόοδο των εποχών.



Εικόνα 39: Αποτελέσματα εκπαίδευσης Voxelmorph με συνάρτηση κόστους (loss function) το MSE και  $\lambda = 0.6$ . Στην εικόνα (a) φαίνεται η ευθυγράμμιση στις 50 εποχές, ενώ στην εικόνα (b) φαίνεται η ευθυγράμμιση για τις 80 εποχές. Για τις δύο περιπτώσεις, αριστερά είναι η σταθερή εικόνα, στη μέση η κινούμενη και δεξιά η μετασχηματισμένη.

Συνάρτηση Κόστους	Εποχές	$\lambda$	MI
MSE	80	0.01	0.2292
NCC	80	0.01	0.1458
MSE	80	0.6	0.1603
MSE	50	0.6	0.1577
NCC	80	0.1	0.1526

Πίνακας 5: Παραμετροποιήσεις ευθυγράμμισης για το Voxelmorph.

### 2.7.5 Συμπεράσματα για την ευθυγράμμιση με Voxelmorph

Όπως φαίνεται από το διάγραμμα κόστους στην Εικόνα 34, το βασικό μοντέλο MSE εκπαιδεύεται κατά τις 80 εποχές. Η σταδιακή μείωση τόσο στα δεδομένα εκπαίδευσης, όσο και στα δεδομένα επαλήθευσης υπονοούν πως ο αλγόριθμος μαθαίνει κατά τη διάρκεια των εποχών. Δεν υπάρχει ένδειξη υπερεκπαίδευσης, διότι οι καμπύλες και των δύο κοστών κινούνται με παρόμοιο τρόπο. Επίσης, φαίνεται από τη σταδιακή μείωση της καμπύλης ότι το μοντέλο θα συγκλίνει, ωστόσο αυτό είναι κάτι που δεν έχει επιτευχθεί. Το προαναφερθέν συμβαίνει λόγω των λίγων εποχών εκπαίδευσης τού μοντέλου. Οι τιμές τής συνάρτησης κόστους είναι αρνητικές λόγω του συνδυασμένου MSE με τη μέθοδο gradient κανονικοποίησης (για την εξομάλυνση χωρικών διαστάσεων και πεδίου μετασχηματισμού). Η απόλυτη τιμή του ξεκινάει από το  $\sim 4.2$  και μειώνεται σταδιακά. Αυτή η τιμή αναπαριστά το σημείο εκκίνησης τής διαδικασίας βελτιστοποίησης βασισμένη στη συνάρτηση κόστους. Εδώ, αντιπροσωπεύει το βαθμό τής ανομοιότητας των εικόνων. Μια υψηλότερη απόλυτη τιμή υποδηλώνει μεγαλύτερη αρχική απόκλιση, την οποία το μοντέλο στοχεύει να ελαχιστοποιήσει μέσω της εκμάθησης κατά τη διάρκεια της εκπαίδευσης. Βασίζομένου ότι οι εικόνες έχουν κανονικοποιηθεί μεταξύ μηδέν κι ένα, η αρχική αυτή τιμή ίσως οφείλεται στα αρχικά υψηλά βάρη τού μοντέλου. Οι τιμές MI ξεκινούν από  $\sim 0.16$  κι αυξάνονται όσο αυξάνονται οι εποχές. Αν κι η ανώτατη τιμή ( $\sim 0.3$ ) είναι χαμηλή, αυτό δε σημαίνει ότι το μοντέλο δεν ταιριάζει στα δεδομένα και δεν μπορεί να μάθει. Το μοντέλο, όπως προαναφέρθηκε και κατά την εξέταση της συνάρτησης κόστους, χρειάζεται περισσότερες εποχές για να συγκλίνει κι επομένως να αυξήσει τη μετρική MI. Οι απότομες διακυμάνσεις στις τιμές τους οφείλονται στις μεταβολές των βαρών τού μοντέλου κατά την εκπαίδευση.

Όσον αφορά το βασικό μοντέλο NCC, διαπιστώνεται από την Εικόνα 35 πως υπάρχει κάποια δυσκολία στην εκπαίδευση, παρ' όλο που φαίνεται να μαθαίνει. Το MI για τα δεδομένα εκπαίδευσης ξεκινάει από μια τιμή  $\sim 0.155$  και κυμαίνεται ελαφρά ανοδικά προς το  $\sim 0.156$  κάνοντας ένα άλμα στο  $\sim 0.17$ . Αυτή η ανοδική τάση, δείχνει κάποια εκπαίδευση. Οι αλληπάλληλες διακυμάνσεις είναι φυσιολογικές κατά τη διάρκεια της εκπαίδευσης, καθώς το μοντέλο προσαρμόζει τις υπερπαραμέτρους του, προκειμένου να ελαχιστοποιήσει το κόστος. Κατόπιν αυτού, το MI στα δεδομένα επαλήθευσης συμβαδίζει με την τάση τού MI των δεδομένων εκπαίδευσης, κάτι που επισφραγίζει τη γενίκευση του μοντέλου σε δεδομένα που δεν έχει δει. Πάλι, οι διακυμάνσεις στις τιμές τού MI κι οι περιστασιακές αυξήσεις του, υποσημειώνουν ότι η απόδοση του αλγορίθμου ποικίλει μεταξύ διαφορετικών συνόλων εικόνων, όμως εν γένει βελτιώνεται. Όπως και στην εκπαίδευση με συνάρτηση κόστους το MSE, δε διαφαίνεται υπερπροσαρμογή των δεδομένων, λόγω της συνοχής των τάσεων των καμπυλών εκπαίδευσης και επαλήθευσης.

Τα διαφορετικά πειράματα που υλοποιήθηκαν επικεντρώθηκαν στην προσαρμογή τού παράγοντα ρύθμισης  $\lambda$  και στον αριθμό των εποχών. Η αύξηση του  $\lambda$  λειτουργεί θετικά και για τα δύο μοντέλα. Οι τιμές που επιλέγονται προσομοιώνουν τις βέλτιστες τιμές που περιγράφονται στο άρθρο των Hoopes et al. "Learning the Effect of Registration Hyperparameters with HyperMorph" [52]. Κατά αυτόν τον τρόπο, επιλέγεται η τιμή 0.1 για το NCC μοντέλο κι η τιμή 0.6 για το MSE μοντέλο. Η αύξηση των εποχών φαίνεται, επίσης, να επηρεάζει την ποιότητα της ευθυγράμμισης, δεδομένου ότι το ποιοτικό αποτέλεσμα για τις 80 εποχές είναι καλύτερο σε σύγκριση με τις 50 (Εικόνα 39). Τα αποτελέσματα από ένα αντικείμενο εξέτασης φαίνονται στον Πίνακα 5.

### 2.7.6 Σύγκριση μεταξύ των δύο τεχνικών (SimpleITK, Voxelmorph)

Κάνοντας μια απόλυτη σύγκριση τιμών μεταξύ των δύο τεχνικών, η SimpleITK υπερσχύει λόγω των αποτελεσμάτων με τα φίλτρα Demons. Ποιοτικά αυτό εξακριβώνεται από την Εικόνα 28, στην οποία φαίνεται η παραμορφοποίηση της κινούμενης εικόνας, ώστε να εφαρμοστεί στη σταθερή. Μολαταύτα, τόσο με τη χρήση των φίλτρων Demons, όσο και με τη χρήση τής ImageRegistrationMethod, οι εναλλαγές στις διάφορες μετρικές που απαρτίζουν και καθορίζουν τους εκάστοτε αλγορίθμους δε βελτιώνουν σημαντικά την ευθυγράμμιση. Έτσι, οι μεταβολές των τιμών MI σε κάθε διαφοροποίησή τους δεν αλλάζει εμφανώς, κάτι που υποδεικνύει την ύπαρξη πλαφόν στη βελτίωση της ευθυγράμμισης. Σε αντίθεση, η καλύτερη ευθυγράμμιση με το μοντέλο Βαθιάς Μάθησης βρίσκεται σε νεογιλά στάδια. Εμφανίζεται μια αρχική σύντηξη (Εικόνα 39 (b)), η οποία εντοπίζει κάποια βασικά χαρακτηριστικά τής σταθερής εικόνας, ωστόσο απέχει από μια ολοκληρωμένη ευθυγράμμιση. Είναι φανερό ότι η εκπαίδευση του μοντέλου αυτού είναι ημιτελής, μπορεί, όμως, κάλλιστα να βελτιωθεί. Τα αποτελέσματα δείχνουν μια θετική τάση προόδου, η οποία



θα επιτευχθεί με περισσότερες εποχές εκπαίδευσης κι εκτενέστερη προσαρμογή των υπερπαραμέτρων.

### 3 Συμπεράσματα

Στη συγκεκριμένη διπλωματική εξετάστηκε η ευθυγράμμιση ιατρικών εικόνων με τη βοήθεια του μοντέλου Βαθιάς Μάθησης Voxelmorph, το οποίο έχει ως βάση του το Unet και με τη βοήθεια των Demons φίλτρων και της ImageRegistrationMethod. Οι μεθοδολογίες αυτές αναφέρθηκαν κι αναλύθηκαν εκτενώς στα Μέρους Α, Μέρος Β στο Μέρος Α εξετάστηκε κυρίως το θεωρητικό υπόβαθρο, στο Μέρος Β η υλοποίησή τους για το παρόν πρόβλημα.

Γίνεται αντιληπτό στο Μέρος Β, ότι κι οι δύο τεχνικές αυτές βοηθούν στην ευθυγράμμιση των δεδομένων. Οι μεθοδολογίες εξετάζονται ως προς τη μετρική MI, οποία ποσοτικοποιεί τη στατιστική εξάρτηση ή την πληροφορία που μοιράζονται δύο εικόνες. Υψηλότερες τιμές MI υπονοούν καλύτερη αντιστοίχιση ή επικάλυψη του περιεχομένου των εικόνων.

Η SimpleITK εξετάζεται στην ευθυγράμμιση με δύο μεθόδους, τα φίλτρα Demons και την ImageRegistrationMethod. Η πρώτη μέθοδος πετυχαίνει μία τιμή γύρω στο  $\sim 0.4$  καταφέροντας μία μέτρια ευθυγράμμιση. Η δεύτερη λαμβάνει ανώτατη τιμή  $\sim 0.19$ , δυσκολευόμενη να συντήξει τις εικόνες. Καμία από τις δύο μεθόδους αντιστοιχίζει τις εικόνες στο έπακρο, λόγω των ξεχωριστών δομών τους που επηρεάζουν την ευθυγράμμιση. Η διαφοροποιήσεις στην τιμή MI και των δύο τεχνικών αυξομειώνεται σε μικρή κλίμακα, ανεξάρτητα από τις αλλαγές που γίνονται στην αρχικοποίηση της ευθυγράμμισης. Η καλύτερη των δύο τεχνική είναι μέτρια τόσο ως προς την ακρίβεια της ευθυγράμμισης, αλλά κι ως προς την αποδοτικότητα, δεδομένου ότι ο χρόνος εκπαίδευσης είναι μεγάλος, ιδίως όταν χρησιμοποιούνται πιο περίπλοκες μέθοδοι παρεμβολής όπως οι B-Splines κι όταν επιλέγεται μικρότερο διάστημα για την τυπική απόκλιση. Η βιβλιοθήκη αυτή, η οποία προσφέρει παραδοσιακές μεθόδους ευθυγράμμισης εικόνων δεν εντοπίζει τις πιο σύνθετες αποκλίσεις μεταξύ των εικόνων, με αποτέλεσμα μια όχι και τόσο καλή ευθυγράμμιση.

Όπως στην ευθυγράμμιση με τη SimpleITK, έτσι και στη Βαθιά Μάθηση εξετάζονται διαφορετικά μοντέλα. Το καλύτερο εξ αυτών έχει συνάρτηση κόστους το MSE. Όσο αυξάνονται οι εποχές εκπαίδευσης κι όσο βελτιστοποιούνται οι υπερπαραμέτροι που το διέπουν, τόσο καλύτερεύει η ευθυγράμμιση, έτσι συμπεραίνεται ότι το μοντέλο εκπαιδεύεται και μαθαίνει τα δεδομένα. Το Voxelmorph είναι μια προσέγγιση βασισμένη στη μάθηση και μπορεί να μοντελοποιεί πολύπλοκες παραμορφώσεις καλύτερα σε ένα σύνολο δεδομένων εκπαίδευσης, σε σχέση με μία κλασική τεχνική. Στο υπό μελέτη πρόβλημα καταφέρνει να φτάσει σε MI score κοντά στο 0.3 κάτι που επιβεβαιώνει την ικανότητά του να αντιστοιχίζει τις εικόνες. Μία παρατήρηση που μπορεί να γίνει στα αποτελέσματα του Voxelmorph είναι ότι φαίνεται πως το μοντέλο έχει περισσότερα περιθώρια εκπαίδευσης. Κάτι τέτοιο απορρέει από τη μη σύγκλιση της συνάρτησης κόστους στις δοθείσες εποχές εκπαίδευσης. Για να επαυξηθούν οι εποχές, χρειάζεται περισσότερη υπολογιστική ισχύς και διαθέσιμος χρόνος σε σχέση με τις δυνατότητες που δίνει το περιβάλλον του Google Colab Pro. Για τις 80 εποχές το μοντέλο δαπανά περίπου εννιά ώρες εκπαίδευσης και 96 υπολογιστικές μονάδες.

Συμπεριληπτικά, το μοντέλο Voxelmorph με τη σωστή εκπαίδευση δύναται να αντιστοιχίζει ακριβώς καλά την κινούμενη εικόνα στη σταθερή. Χάρη σε αυτό η ευθυγράμμιση ιατρικών εικόνων δύναται να γίνει ευκολότερη, καθώς προσφέρει μια γρήγορη κι ακριβή μέθοδο για τη σύγκριση και ανάλυση διαφόρων εικόνων. Αυτό είναι ιδιαίτερα χρήσιμο για την παρακολούθηση της εξέλιξης των ασθενειών, την αξιολόγηση της αποτελεσματικότητας των θεραπειών και τη λήψη εικόνων πριν και μετά από χειρουργικές επεμβάσεις. Επιπλέον, με την αυτοματοποίηση της διαδικασίας, οι ιατροί μπορούν να εξοικονομήσουν χρόνο και να εστιάσουν περισσότερο στη διάγνωση και τη θεραπεία, ενώ διασφαλίζεται η ακρίβεια και η συνέπεια των αποτελεσμάτων.

## 4 Μελλοντική Έρευνα

Τα μοντέλα κι οι διαδικασίες που εξετάστηκαν μπορούν να επεκταθούν με σκοπό τη βελτίωσή τους.

- Προεπεξεργασία δεδομένων: στο κομμάτι αυτό, μπορούν να εφαρμοστούν μέθοδοι εξαγωγής σημείων ενδιαφέροντος από τις εικόνες και των δύο τρόπων λήψης, τα οποία θα λειτουργήσουν επικουρικά στη διαδικασία τής εκπαίδευσης, εφόσον θα διευκολύνουν την επίβλεψή της. Για παράδειγμα, ο εντοπισμός των οργάνων στις εικόνες CT/PET κι η δημιουργία масκών για τα όργανα, κατατέμνοντας τις εικόνες με τη βοήθεια ενός μοντέλου Βαθιάς Μάθησης (λόγου χάρι Unet) ή με τη βοήθεια βιβλιοθηκών, ανοιχτών προς γενική χρήση (λόγου χάρι Slicer 5.2.2) είναι τεχνικές που εκπληρώνουν την προαναφερθείσα διαδικασία. Θα μπορέσουν, με αυτόν τον τρόπο να χρησιμοποιηθούν και περαιτέρω τεχνικές μέτρησης ακρίβειας, όπως το Dice Score ή το TRE.
- Αλλαγή παραμέτρων μοντέλου: όπως διαπιστώθηκε το μοντέλο εκπαιδεύεται, ωστόσο δε συγχλίνει. Για το λόγο αυτό, μπορούν να εφαρμοστούν περισσότερες εποχές σε συνδυασμό με callbacks, ώστε να ενεργοποιηθεί η πρόωρη παύση (early stopping) και να σταματήσει η εκπαίδευση εν ευθέτω χρόνω, ώστε να αποφευχθεί η υπερεκπαίδευση. Επίσης, μπορεί να γίνει μία διερεύνηση όσον αφορά τα αρχικά βάρη τής εκπαίδευσης και να προσαρμοστούν κατάλληλα σε μικρότερες ή μηδενικές τιμές με σύνθετες αρχικοποιημένες συναρτήσεις. Για παράδειγμα μπορεί να οριστεί μια συνάρτηση η οποία θέτει τα βάρη ίσα με το μηδέν ή χρησιμοποιεί μικρότερη κλίμακα για τυχαία αρχικοποίηση κι έπειτα να εφαρμοστεί στα επίπεδα του μοντέλου. Για την αρχικοποίηση στο μηδέν μπορεί να χρησιμοποιηθεί η συνάρτηση `torch.nn.init.constant_(layer.weight, 0)`. Αν κι αυτό ίσως βοηθήσει στην αρχική ανομοιότητα τού μοντέλου, δε συνιστάται για επίπεδα που χρησιμοποιούν το ρυθμό μάθησης, διότι μπορεί να παρεμποδίσουν το μοντέλο από το να μάθει. Ακριβώς επειδή υπάρχει αυτή η πιθανότητα είναι προτιμότερα να αρχικοποιηθούν τα βάρη σε μια αρκετά μικρή τιμή (`torch.nn.init.normal_(layer.weight, mean=0, std=0.01)`) για τα επίπεδα που επιδέχονται ρυθμό μάθησης. Με αυτόν τον τρόπο το μοντέλο μπορεί να ξεκινήσει από ένα ελεγχόμενο σημείο και κατά συνέπεια να οδηγήσει σε μια πιο σταθερή εκπαίδευση.
- Επέκταση μοντέλου: η ευθυγράμμιση με τη SimpleITK καταγράφει βασικά χαρακτηριστικά των εικόνων κι αυτό μπορεί να χρησιμεύσει ως προέκταση του μοντέλου Voxelmorph. Συγκεκριμένα, η χρήση των φίλτρων Demons εστιάζει στην σύντηξη με βάση τις εντάσεις, κάτι που μπορεί να υλοποιήσει μια χονδρική ευθυγράμμιση των εικόνων πριν την εισαγωγή τους στο μοντέλο Voxelmorph με το οποίο θα ακολουθήσει πιο λεπτομερειακή εγγραφή. Αυτή η προσέγγιση δύναται να βελτιώσει τη συνολική ακρίβεια τής ευθυγράμμισης, καθώς η αρχική ευθυγράμμιση που παρέχεται από τους Demons μπορεί να μειώσει την πολυπλοκότητα του μετασχηματισμού που χρειάζεται να μάθει το Voxelmorph, οδηγώντας ενδεχομένως σε καλύτερα αποτελέσματα εγγραφής.

## Βιβλιογραφικές Αναφορές

- [1] Oliveira FP, Tavares JM. Medical image registration: a review. *Comput Methods Biomech Biomed Engin.* 2014;17(2):73-93. doi: 10.1080/10255842.2012.670855. Epub 2012 Mar 22.
- [2] Bharati, S., Mondal, M. R. H., Podder, P., & Prasath, V. B. S. 2022. Deep Learning for Medical Image Registration: A Comprehensive Review. *International Journal of Computer Information Systems and Industrial, Management Applications (ISSN 2150-7988)*, Volume 14, pp. 173-190. doi: 10.48550/arXiv.2204.11341.
- [3] Goubran, M., Leuze, C., Hsueh, B. et al. Multimodal image registration and connectivity analysis for integration of connectomic data from microscopy to MRI. *Nat Commun* 10, 5504 (2019). doi: 10.1038/s41467-019-13374-0
- [4] Slomka PJ, Baum RP. Multimodality image registration with software: state-of-the-art. *Eur J Nucl Med Mol Imaging.* 2009 Mar;36 Suppl 1:S44-55. doi: 10.1007/s00259-008-0941-8. PMID: 19104803.
- [5] M. Foskey et al., "Large deformation three-dimensional image registration in image-guided radiation therapy," *Phys. Med. Biol.*, vol. 50, no. 24, pp. 5869–5892, Dec. 2005.
- [6] van der Hoorn, J. L. Yan, T. J. Larkin, N. R. Boonzaier, T. Matys, and S. J. Price, "Validation of a semi-automatic co-registration of MRI scans in patients with brain tumors during treatment follow-up," *NMR Biomed.*, vol. 29, no. 7, pp. 882–889, Jul. 2016.
- [7] Lorenzen P, Prastawa M, Davis B, Gerig G, Bullitt E, Joshi S. Multi-modal image set registration and atlas formation. *Med Image Anal.* 2006 Jun;10(3):440-51. doi: 10.1016/j.media.2005.03.002. PMID: 15919231; PMCID: PMC2430608.
- [8] Gooya A, Biros G, Davatzikos C. Deformable registration of glioma images using EM algorithm and diffusion reaction modeling. *IEEE Trans Med Imaging.* 2011 Feb;30(2):375-90. doi: 10.1109/TMI.2010.2078833. Epub 2010 Sep 27. PMID: 20876010; PMCID: PMC3245665.
- [9] R. Ahmed, K. E. K. Emon, and M. F. Hossain."Robust driver fatigue recognition using image processing." In 2014 International Conference on Informatics, Electronics & ision (ICIEV), pp. 1-6. IEEE, 2014
- [10] Zheng Y, Sui X, Jiang Y, Che T, Zhang S, Yang J, Li H. SymReg-GAN: Symmetric Image Registration With Generative Adversarial Networks. *IEEE Trans Pattern Anal Mach Intell.* 2022 Sep;44(9):5631-5646. doi: 10.1109/TPAMI.2021.3083543. Epub 2022 Aug 4. PMID: 34033536.
- [11] Fu Y, Lei Y, Wang T, Curran WJ, Liu T, Yang X. Deep learning in medical image registration: a review. *Phys Med Biol.* 2020 Oct 22;65(20):20TR01. doi: 10.1088/1361-6560/ab843e. PMID: 32217829; PMCID: PMC7759388.
- [12] Ghahramani, Z. (2004). Unsupervised Learning. In: Bousquet, O., von Luxburg, U., Rätsch, G. (eds) *Advanced Lectures on Machine Learning. ML 2003. Lecture Notes in Computer Science()*, vol 3176. Springer, Berlin, Heidelberg. [https://doi.org/10.1007/978-3-540-28650-9\\_5](https://doi.org/10.1007/978-3-540-28650-9_5)
- [13] Jaderberg M, Simonyan K, Zisserman A and Kavukcuoglu K 2015 Spatial transformer networks (arXiv:1506.02025).

- [14] Maintz JB, Viergever MA. A survey of medical image registration. *Med Image Anal.* 1998 Mar;2(1):1-36. doi: 10.1016/s1361-8415(01)80026-8. PMID: 10638851.
- [15] Fitzpatrick, J.M., Hill, D.L.G., Maurer, C.R. Jr., 2000. Image Registration. In: J.M. Fitzpatrick, M. Sonka (Eds.), *Handbook of Medical Imaging, Volume 2. Medical Image Processing and Analysis*. SPIE Press, pp. 376-429.
- [16] J. A. Nelder, R. Mead, A Simplex Method for Function Minimization, *The Computer Journal*, Volume 7, Issue 4, January 1965, Pages 308–313, doi: 10.1093/comjnl/7.4.308.
- [17] Dedieu, JP., 2015. Newton-Raphson Method. In: Engquist, B. (eds) *Encyclopedia of Applied and Computational Mathematics*. Springer, Berlin, Heidelberg. doi: 10.1007/978-3-540-70529-1\_374.
- [18] Ruder S., 2017. An overview of gradient descent optimization algorithms. Insight Centre for Data Analytics, NUI Galway. doi: 10.48550/arXiv.1609.04747.
- [19] Chen, M., Tustison, N.J., Jena, R., Gee, J.C. (2023). Image Registration: Fundamentals and Recent Advances Based on Deep Learning. In: Colliot, O. (eds) *Machine Learning for Brain Disorders*. *Neuroinformatics*, vol 197. Humana, New York, NY. doi: 10.1007/978-1-0716-3195-9\_14.
- [20] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 234-241). Springer, Cham. doi: 10.48550/arXiv.1505.04597.
- [21] Jia, X., Bartlett, J., Zhang, T., Lu, W., Qiu, Z., Duan, J. (2022). U-Net vs Transformer: Is U-Net Outdated in Medical Image Registration?. In: Lian, C., Cao, X., Rekić, I., Xu, X., Cui, Z. (eds) *Machine Learning in Medical Imaging. MLMI 2022. Lecture Notes in Computer Science*, vol 13583. Springer, Cham. doi: 10.1007/978-3-031-21014-3\_16.
- [22] Azad, R., Aghdam, E.K., Rauland, A., Jia, Y., Avval, A.H., Bozorgpour, A., ... Merhof, D. (2022). Medical Image Segmentation Review: The Success of U-Net. doi: 10.48550/arXiv.2211.14830.
- [23] Sanderson, G. [3Blue1Brown]. (2022, November 18). But what is a convolution? [Video]. YouTube. [https://www.youtube.com/watch?v=KuXjwB4LzSA&ab\\_channel=3Blue1Brown](https://www.youtube.com/watch?v=KuXjwB4LzSA&ab_channel=3Blue1Brown). Accessed [9/03/2024].
- [24] Dumoulin, V., & Visin, F. (2018, January 12). A guide to convolution arithmetic for deep learning. MILA, Université de Montréal; AIRLab, Politecnico di Milano. doi: arXiv:1603.07285v2.
- [25] Alrasheedi, F., Zhong, X., & Huang, P.-C. (2023). Padding module: Learning the padding in deep neural networks. *IEEE Access*. doi: 10.48550/arXiv.2301.04608v1.
- [26] Kwok, R., (2023, January 15). Transposed Conv as Matrix Multiplication explained.[Article]. Medium. <https://medium.com/@rmkwok/explain-implement-and-compare-2d-transposed-convolution-in-num-py-tensorflow-and-pytorch-2068d986ec5>. Accessed [9/03/2024].
- [27] Steffen Czolbe, Paraskevas Pegios, Oswin Krause, Aasa Feragen, (2023). Semantic similarity metrics for image registration. *Medical Image Analysis*, vol 87, 102830, ISSN 1361-8415. doi: 10.1016/j.media.2023.102830.

- [28] Song, J. H. (2017). Methods for evaluating image registration [Doctoral dissertation, University of Iowa]. University of Iowa. doi: 10.17077/etd.v0vailob.
- [29] Gaidhane, V. H., Hote, Y. V., and Singh, V. (2012). An efficient similarity measure technique for medical image registration. *Sadhanā*, 37(6), 709–721. <https://www.ias.ac.in/article/fulltext/sadh/037/06/0709-0721>. Accessed [9/03/2024].
- [30] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, “Voxelmorph: A learning framework for deformable medical image registration,” *IEEE Transactions on Medical Imaging*, vol. 38, pp. 1788–1800, Aug 2019.
- [31] Thibaud Briand, Pascal Monasse. Theory and Practice of Image B-Spline Interpolation. *Image Processing On Line*, 2018, 8, pp.99-141. ff10.5201/ipol.2018.221ff. fihal-01846912f.
- [32] Waldkirch, B. I. (2020). Methods for three-dimensional Registration of Multimodal Abdominal Image Data. Inaugural dissertation for the acquisition of Doctor scientiarum humanarum (Dr. sc. hum.) at the Medical Faculty Mannheim of the Ruprecht-Karls-University Heidelberg. Ludwigshafen am Rhein.
- [33] R. Beare, B. C. Lowekamp, Z. Yaniv, “Image Segmentation, Registration and Characterization in R with SimpleITK”, *J Stat Software*, 86(8), doi: 10.18637/jss.v086.i08, 2018.
- [34] Z. Yaniv, B. C. Lowekamp, H. J. Johnson, R. Beare, “SimpleITK Image-Analysis Notebooks: a Collaborative Environment for Education and Reproducible Research”, *J Digit Imaging.*, doi: 10.1007/s10278-017-0037-8, 31(3): 290-303, 2018.
- [35] B. C. Lowekamp, D. T. Chen, L. Ibáñez, D. Blezek, “The Design of SimpleITK”, *Front. Neuroinform.*, 7:45. doi: 10.3389/fninf.2013.00045, 2013.
- [36] Mayo Clinic. (n.d.). Retrieved from <https://www.mayoclinic.org/>. Accessed [17/02/2024].
- [37] Kitware Inc. (n.d.). <https://www.kitware.com/>. Accessed [17/02/2024].
- [38] University of Iowa. (n.d.). <https://uiowa.edu/>. Accessed [17/02/2024].
- [39] NLM intramural research program. (n.d.). <https://www.nlm.nih.gov/research/index.html>. Accessed [17/02/2024].
- [40] Monash University. (n.d.). <https://www.monash.edu/>. Accessed [17/02/2024].
- [41] Office of Cyber Infrastructure and Computational Biology. (n.d.). <https://www.niaid.nih.gov/about/cyber-infrastructure-computational-biology-contacts>. Accessed [17/02/2024].
- [42] SimpleITK (n.d.). <https://simpleitk.readthedocs.io/en/master/history.html>. Accessed [17/02/2024].
- [43] ImageRegistrationMethod. [https://simpleitk.org/doxygen/latest/html/classitk\\_1\\_1simple\\_1\\_1ImageRegistrationMethod.html](https://simpleitk.org/doxygen/latest/html/classitk_1_1simple_1_1ImageRegistrationMethod.html). Accessed [20/02/2024].

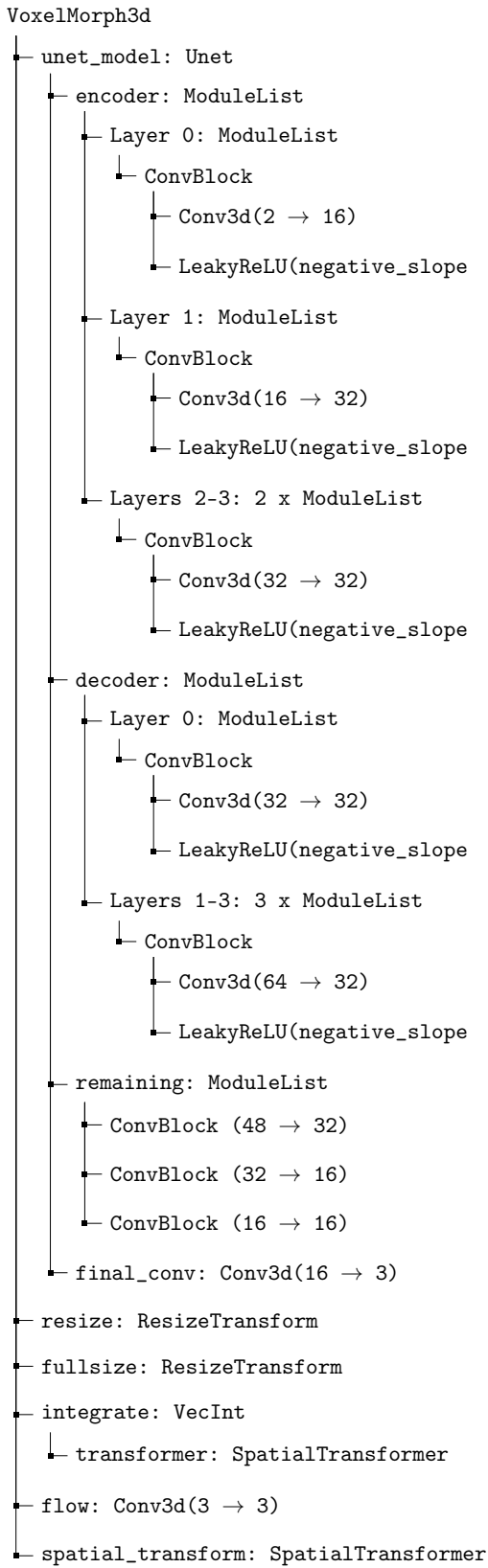
- [44] Siemens Biograph mCT. <https://www.siemens-healthineers.com/gr/molecular-imaging/pet-ct/biograph-mct>. Accessed [20/02/2024].
- [45] Chan HP, Samala RK, Hadjiiski LM, Zhou C. Deep Learning in Medical Image Analysis. *Adv Exp Med Biol.* 2020;1213:3-21. doi: 10.1007/978-3-030-33128-3\_1. PMID: 32030660; PMCID: PMC7442218.
- [46] Boveiri, H. R., Khayami, R., Javidan, R., & Mehdizadeh, A. R. (2020). Medical image registration using deep neural networks: A comprehensive review. *Computers & Electrical Engineering*, 87, 106767. doi: 10.1016/j.compeleceng.2020.106767.
- [47] Balakrishnan, G., Zhao, A., Sabuncu, M. R., Gutttag, J. V., & Dalca, A. V. (2018). An unsupervised learning model for deformable medical image registration. doi: 10.48550/arXiv.1802.02604.
- [48] Nazib, A., Fookes, C., & Perrin, D. (2018). A Comparative Analysis of Registration Tools: Traditional vs Deep Learning Approach on High Resolution Tissue Cleared Data. *ArXiv*, abs/1810.08315.
- [49] Berger A. How does it work? Positron emission tomography. *BMJ.* 2003 Jun 28;326(7404):1449. doi: 10.1136/bmj.326.7404.1449. PMID: 12829560; PMCID: PMC1126321.
- [50] How Does CT Work?. *imaginis*. Available from: <https://www.imaginis.com/ct-scan/how-does-ct-work?r>. Accessed [05/03/2024].
- [51] Z. Yaniv, B. C. Lowekamp, H. J. Johnson, R. Beare. SimpleITK Image-Analysis Notebooks: a Collaborative Environment for Education and Reproducible Research. *J Digit Imaging.*, doi: 10.1007/s10278-017-0037-8, 31(3): 290-303, 2018.
- [52] Hoopes A, Hoffmann M, Greve DN, Fischl B, Gutttag J, Dalca AV. Learning the Effect of Registration Hyperparameters with HyperMorph. *J Mach Learn Biomed Imaging.* 2022 Mar;1:003. Epub 2022 Apr 7. PMID: 36147449; PMCID: PMC9491317.
- [53] Fu, Y., Lei, Y., Wang, T., Patel, P., Jani, A. B., Mao, H., Curran, W. J., Liu, T., & Yang, X. (2021). Biomechanically constrained non-rigid MR-TRUS prostate registration using deep learning based 3D point cloud matching. *Medical Image Analysis*, 67, 101845. doi: 10.1016/j.media.2020.101845
- [54] Fu Y, Lei Y, Wang T, Higgins K, Bradley JD, Curran WJ, Liu T, Yang X. LungRegNet: An unsupervised deformable image registration method for 4D-CT lung. *Med Phys.* 2020 Apr;47(4):1763-1774. doi: 10.1002/mp.14065. Epub 2020 Feb 26. PMID: 32017141; PMCID: PMC7165051.
- [55] de Vos, B. D., Berendsen, F. F., Viergever, M. A., Sokooti, H., Staring, M., & Išgum, I. (2019). A Deep Learning Framework for Unsupervised Affine and Deformable Image Registration. *Medical Image Analysis*, 52, 128-143. doi: 10.48550/arXiv.1809.06130.
- [56] Zhao, S., Lau, T., Luo, J., Chang, E., & Xu, Y. (2019). Unsupervised 3D End-to-End Medical Image Registration With Volume Tweening Network. *IEEE Journal of Biomedical and Health Informatics*, 24(11), 1394-1404. doi: 10.1109/JBHI.2019.2951024.
- [57] Estienne, T., Vakalopoulou, M., Christodoulidis, S., Battistella, E., Lerousseau, M., Carre, A., Klaus-

ner, G., Sun, R., Robert, C., Mougiakakou, S., Paragios, N., & Deutsch, E. (2019). U-ResNet: Ultimate Coupling of Registration and Segmentation with Deep Nets. In ISBN 978-3-030-32247-2 (pp. 310-319). doi: 10.1007/978-3-030-32248-9\_35

[58] Gatidis S, Kuestner T. (2022) A whole-body FDG-PET/CT dataset with manually annotated tumor lesions (FDG-PET-CT-Lesions) [Dataset]. The Cancer Imaging Archive. doi: 10.7937/gkr0-xv29.



# Παράρτημα



Εικόνα 40: Δομή μοντέλου Voxelmorph.