



Εθνικό Μετσόβιο Πολυτεχνείο

Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Δ.Π.Μ.Σ.: Επιστήμη Δεδομένων και Μηχανική Μάθηση

Μεταπτυχιακή Διπλωματική Εργασία:

# Ευθυγράμμιση Εικόνων FDG-PET (3D) Καρκινοπαθών Ασθενών με τεχνικές Βαθιάς Μάθησης

Στεφανής Μιχαήλ

Επιβλέπων: Γεώργιος Ματσόπουλος

Σεπτέμβριος, 2022

## Περιεχόμενα

<b>1</b>	<b>Εισαγωγή</b>	<b>13</b>
<b>2</b>	<b>Προαπαιτούμενες Ιατρικές Γνώσεις</b>	<b>18</b>
2.1	Ο καρκίνος ως μία σύγχρονη ασθένεια . . . . .	18
2.2	Τομογραφίες τύπου PET και FDG-PET . . . . .	23
2.2.1	Το ποζιτρόνιο . . . . .	23
2.2.2	Τομογραφίες Εκπομπής Ποζιτρονίων . . . . .	24
2.2.3	Τομογραφίες FDG-PET . . . . .	25
2.3	Οι υπό-μελέτη τύποι καρκίνων . . . . .	26
<b>3</b>	<b>Ευθυγράμμιση Εικόνων</b>	<b>28</b>
3.1	Κλασσικές Μέθοδοι Ευθυγράμμισης . . . . .	28
3.2	Ευθυγράμμιση με Βαθιά Μάθηση . . . . .	30
3.2.1	Το μοντέλο U-Net . . . . .	31
3.2.2	Το μοντέλο VoxelMorph . . . . .	34
3.2.3	Συναρτήσεις Κόστους . . . . .	36
3.2.4	Παραλλαγές του μοντέλου VoxelMorph . . . . .	42
<b>4</b>	<b>Προ-επεξεργασία των Εικόνων</b>	<b>47</b>
4.1	Βήμα 1: Μετατροπή των αρχείων DICOM σε αρχεία NIfTI . . . . .	47
4.1.1	Τύπος Αρχείων DICOM . . . . .	47
4.1.2	Τύπος Αρχείων NIfTI . . . . .	48
4.1.3	Από DICOM σε NIfTI . . . . .	49
4.2	Βήμα 2: Μετατροπή των αρχείων PET.nii.gz σε SUV.nii.gz . . . . .	50
4.3	Βήμα 3: Μετατροπή των αρχείων .nii.gz σε Arrays και προ-επεξεργασία . . . . .	51
<b>5</b>	<b>Επισκόπηση των Δεδομένων</b>	<b>55</b>
5.1	Πηγή των δεδομένων . . . . .	55
5.2	Εικόνες SUV . . . . .	56
5.3	Εικόνες SEG . . . . .	57
<b>6</b>	<b>Σχεδιασμός και Περιγραφή Πειραμάτων</b>	<b>61</b>
6.1	Χαρακτηριστικά του Εξοπλισμού . . . . .	61
6.2	Προ-επεξεργασία των δεδομένων . . . . .	61
6.3	Υπερπαραμέτροι . . . . .	62
6.4	Μοντέλα . . . . .	64
6.5	Πίνακας πειραμάτων . . . . .	67
<b>7</b>	<b>Αποτελέσματα</b>	<b>70</b>
7.1	Αξιολόγηση της εκπαίδευσης των μοντέλων . . . . .	70
7.2	Αξιολόγηση των μοντέλων στο σύνολο ελέγχου . . . . .	80
7.3	Οπτική επαλήθευση της ευθυγράμμισης στο σύνολο ελέγχου . . . . .	84

8 Σχολιασμοί και Συμπεράσματα	91
Βιβλιογραφία	93

## Ευρετήριο Εικόνων

1	Σχηματική αναπαράσταση της πορείας του καρκίνου . . . . .	19
2	Τομογράφος PET-CT [9] . . . . .	25
3	Μοριακή δομή της γλυκόζης και του FDG . . . . .	26
4	Δύο διαφορετικές διαγνωστικές μέθοδοι για τον καρκίνο του πνεύμονα . . . . .	27
5	Σχηματική αναπαράσταση του μοντέλου U-Net . . . . .	31
6	Σχηματική αναπαράσταση της πράξης της αποσυνέλιξης (1) . . . . .	33
7	Σχηματική αναπαράσταση της πράξης της αποσυνέλιξης (2) . . . . .	33
8	Σχηματική αναπαράσταση του VoxelMorph [31] . . . . .	35
9	Επίδραση του όρου ομαλοποίησης στο διανυσματικό πεδίο μετατόπισης . . . . .	40
10	Δίπλωμα (folding) ενός διανυσματικού πεδίου [26] . . . . .	41
11	Σχηματική αναπαράσταση του μοντέλου βασισμένου στα GANs . . . . .	43
12	Σχηματική αναπαράσταση του μοντέλου ViT-V-Net . . . . .	44
13	Σχηματική αναπαράσταση ενός μετασχηματιστή όρασης . . . . .	44
14	Σχηματική αναπαράσταση του μοντέλου TransMorph . . . . .	45
15	Σχηματική αναπαράσταση του μετασχηματιστή Swin . . . . .	45
16	Σχηματική Αναπαράσταση της προ-επεξεργασίας των δεδομένων . . . . .	47
17	Προβολή αρχείου DICOM με την χρήση του λογισμικού MicroDicom [33] . . . . .	48
18	Μετατροπή των αρχείων DICOM σε NIfTI [37] . . . . .	49
19	Κλασικές μέθοδοι προ-επεξεργασίας μίας εικόνα [43] . . . . .	53
20	Διαφορετικές όψεις της ίδιας εικόνας (τομή 170) . . . . .	56
21	Διαφορετικές όψεις της ίδιας εικόνας (τομή 200) . . . . .	57
22	Διαφορετικές όψεις της ίδιας μάσκας (τομή 190) . . . . .	58
23	Διαφορετικές όψεις της ίδιας μάσκας (τομή 200) . . . . .	58
24	Κόστος: SSIM, Ομαλοποίηση: $\mathcal{L}_2$ / Total Variation . . . . .	71
25	Κόστος: PCC, Ομαλοποίηση: $\mathcal{L}_2$ / Total Variation . . . . .	72
26	Κόστος: PCC_SSIM, Ομαλοποίηση: $\mathcal{L}_2$ / Total Variation . . . . .	73
27	Κόστος: SSIM, Ομαλοποίηση: $\mathcal{L}_2$ . . . . .	74
28	Κόστος: PCC, Ομαλοποίηση: $\mathcal{L}_2$ / Total Variation . . . . .	74
29	Κόστος: SSIM, Ομαλοποίηση: $\mathcal{L}_2$ . . . . .	75
30	Κόστος: PCC, Ομαλοποίηση: $\mathcal{L}_2$ / Total Variation . . . . .	76
31	Κόστος: PCC_SSIM, Ομαλοποίηση: $\mathcal{L}_2$ . . . . .	76
32	Κόστος: Parzen M.I., Ομαλοποίηση: $\mathcal{L}_2$ . . . . .	77
33	Κόστος: PCC., Ομαλοποίηση: $\mathcal{L}_2$ . . . . .	78
34	Κόστος: PCC_SSIM., Ομαλοποίηση: $\mathcal{L}_2$ . . . . .	78
35	Κόστος: PCC_SSIM, Ομαλοποίηση: $\mathcal{L}_2$ / Total Variation . . . . .	79
36	Σύγκριση της ευθυγράμμισης των μοντέλων (τομή 20) . . . . .	85
37	Σύγκριση της ευθυγράμμισης των μοντέλων (τομή 54) . . . . .	86
38	Σύγκριση της ευθυγράμμισης των μοντέλων (τομή 65) . . . . .	87
39	Σύγκριση της ευθυγράμμισης των μοντέλων (τομή 85) . . . . .	88
40	Σύγκριση της ευθυγράμμισης των μοντέλων (τομή 120) . . . . .	89
41	Διαγραμματική απεικόνιση του μοντέλου U-NET . . . . .	97

42	Διαγραμματική απεικόνιση του μοντέλου CNN . . . . .	98
43	Διαγραμματική απεικόνιση του μοντέλου U-2-NET . . . . .	99
44	Διαγραμματική απεικόνιση του μοντέλου D-UNET . . . . .	100

## Ευρετήριο Πινάκων

1	Συνηθέστερες εφαρμογές μηχανικής μάθησης σε εταιρίες παγκοσμίως . . . . .	14
2	Συνηθέστερες αιτίες θανάτου στις Ηνωμένες Πολιτείες το 2020 . . . . .	18
3	Παραδείγματα τύπων καρκίνου σε διάφορους ιστούς . . . . .	19
4	Θνησιμότητα συχνών τύπων καρκίνου . . . . .	20
5	Κατηγοριοποίηση καρκινογόνων ουσιών [7] . . . . .	21
6	Παραδείγματα καρκινογόνων ουσιών . . . . .	22
7	Τεχνικά χαρακτηριστικά του εξοπλισμού που χρησιμοποιήθηκε . . . . .	61
8	Υπερπαράμετροι των πειραμάτων . . . . .	64
9	Συνδυασμοί μοντέλων και ποινών που μελετήθηκαν . . . . .	68
10	Μετρικές ομοιότητας στο σύνολο ελέγχου πριν την ευθυγράμμιση . . . . .	80
11	Επίδραση διαφόρων μοντέλων στην μετρική ομοιότητας SSIM . . . . .	81
12	Επίδραση διαφόρων μοντέλων στην μετρική ομοιότητας PCC . . . . .	81
13	Επίδραση διαφόρων μοντέλων στην μετρική ομοιότητας MSE . . . . .	82
14	Επίδραση διαφόρων μοντέλων στην μετρική ομοιότητας Mutual Information . . . . .	82
15	Επίδραση διαφόρων μοντέλων στην μετρική ομοιότητας Dice (SUV) . . . . .	83
16	Επίδραση διαφόρων μοντέλων στην μετρική ομοιότητας Dice (μάσκες) . . . . .	83

## Περίληψη

Η ευθυγράμμιση εικόνας, γνωστή και ως σύντηξη εικόνας 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. Όταν λέμε χωρική μεταμόρφωση, εννοούμε έναν πίνακα με διαστάσεις ίδιες με αυτές των εικόνων που θέλουμε να ευθυγραμμίσουμε και παίρνει την ερμηνεία ενός διανυσματικού πεδίου παραμόρφωσης. Η ευθυγράμμιση εικόνων χρησιμοποιείται σε πολλές κλινικές εφαρμογές όπως καθοδήγηση εικόνας, παρακολούθηση κίνησης, κατάτμηση, ανακατασκευή εικόνας, σύγκριση της κλινικής κατάστασης και ούτω καθεξής. Οι τυπικές (κλασσικές) 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. Αφού μάλιστα έχουν εκπαιδευτεί εκ των προτέρων, κατά το στάδιο της εφαρμογής απαιτούν μόνο τον υπολογισμό της εξόδου με δεδομένα τα βάρη του δικτύου και την είσοδο, πράγμα που απαιτεί σημαντικά λιγότερο χρόνο. Η ευθυγράμμιση με βαθιά μάθηση αποσκοπεί δηλαδή στην ανάπτυξη ενός μοντέλου μηχανικής μάθησης που θα μπορεί αυτόματα και δυναμικά, εφόσον προφανώς έχει εκπαιδευτεί, να ευθυγραμμίσει ταχύτατα ένα ζεύγος εικόνων. Ο στόχος αυτής της εργασίας είναι η ευθυγράμμιση ιατρικών εικόνων FDG-PET μεταξύ των διαφορετικών χρονικών στιγμών των απεικονίσεων αξιοποιώντας τις σύγχρονες μεθόδους της βαθιάς μάθησης. Η αντιστοίχιση αυτή έχει ως σκοπό την διευκόλυνση της σύγκρισης της κατάστασης ασθενών με καρκίνο μεταξύ των απεικονίσεων διαφορετικών χρονικών στιγμών. Με την σειρά της, η σύγκριση αυτή μπορεί να μας βοηθήσει, για παράδειγμα, να καταλάβουμε αν ένας ασθενής έχει ανταποκριθεί σε κάποια θεραπεία ή αν η κατάστασή του έχει χειροτερέψει. Στα πλαίσια της εργασίας αυτής μελετήθηκαν παραλλαγές ενός φημισμένου και σύγχρονου μοντέλου ευθυγράμμισης εικόνων με βαθιά μάθηση, του μοντέλου *VoxelMorph*. Το *VoxelMorph* είναι ίσως το πρώτο και το πιο φημισμένο μοντέλο ευθυγράμμισης ιατρικών εικόνων. Στόχος του είναι η παραγωγή του λεγόμενου διανυσματικού πεδίου παραμόρφωσης. Το πεδίο αυτό είναι μία εικόνα διαστάσεων ίδιων με αυτές που θέλουμε να ευθυγραμμίσουμε. Το νευρωνικό δίκτυο δέχεται δύο τομογραφίες PET, την *κινούμενη εικόνα* και την *εικόνα στόχο*. Στην έξοδο μας δίνει το πεδίο παραμόρφωσης. Ο στόχος της εκπαίδευσης είναι το μοντέλο μας να μάθει να παράγει ένα διανυσματικό πεδίο το οποίο, όταν το εφαρμόσουμε στην κινούμενη εικόνα, να πάρουμε μία νέα εικόνα που να είναι *κατά το δυνατόν όμοια* με την εικόνα στόχο (ως προς κάποια μετρική ομοιότητας). Σημειώνεται εδώ ότι, με βάση της βιβλιογραφία, η ευθυγράμμιση εικόνων με βαθιά μάθηση θεωρείται ένα πρόβλημα μη-επιβλεπόμενης μάθησης. Θα δούμε και με ποιον τρόπο μπορεί να γίνει η μάθηση επιβλεπόμενη, παρόλο που στην περίπτωση μας θα ασχοληθούμε μόνο με την μη-επιβλεπόμενη έκδοση. Η δομή της διπλωματικής εργασίας έχει ως εξής: Ξεκινάμε προφανώς με μία εισαγωγή (Κεφάλαιο 1) που αποσκοπεί περισσότερο στο να μας εισάγει σε βασικές έννοιες που θα συναντήσουμε κατά την διάρκεια της εργασίας. Σειρά έχουν οι προαπαιτούμενες θεωρητικές ιατρικές γνώσεις (Κεφάλαιο 2) που θα μας βοηθήσουν όχι μόνο να τοποθετηθούμε στα ιατρικά πλαίσια της εργασίας αλλά και να κατανοήσουμε την σημαν-

τικότητα της έρευνας της μηχανικής μάθησης στον ιατρικό τομέα. Παρατίθενται δεδομένα σχετικά με την νόσο του καρκίνου καθώς επίσης και με τα διαγνωστικά εργαλεία γνωστά ως τομογραφίες τύπου FDG-PET. Τέλος περιγράφουμε συνοπτικά το είδος των καρκίνων που απεικονίζονται στις εικόνες που χρησιμοποιούμε στην μελέτη μας. Συνεχίζουμε με την μαθηματική θεμελίωση της κλασικής ευθυγράμμισης εικόνων (Κεφάλαιο 3.1), δηλαδή χωρίς μηχανική μάθηση. Η σύντομη αυτή παρουσίαση θα μας βοηθήσει όχι μόνο να κατανοήσουμε καλύτερα το πρόβλημα που πάμε να επιλύσουμε, αλλά επίσης να εκτιμήσουμε την πρόοδο της βαθιάς μάθησης στην επίλυση ενός πολύ δύσκολου προβλήματος βελτιστοποίησης πολλών μεταβλητών. Έπειτα θα δούμε περιγράφουμε αναλυτικά, με όλες τις μαθηματικές λεπτομέρειες, το προαναφερόμενο μοντέλο VoxelMorph (Κεφάλαιο 3.2). Θα δούμε με ποιον τρόπο επεξεργάζεται την πληροφορία (δηλαδή τις εικόνες FDG-PET) έτσι ώστε να παράξει το πεδίο παραμόρφωσης. Στο ίδιο κεφάλαιο περιγράφουμε τις πιο διαδιδόμενες και ευρέως χρησιμοποιούμενες συναρτήσεις ομοιότητας εικόνων και φυσικά της αντίστοιχες συναρτήσεις κόστους με τις οποίες θα εκπαιδεύσουμε τα μοντέλα μας. Περιγράφονται επίσης και ορισμένες συναρτήσεις κανονικοποίησης που παίζουν καθοριστικό ρόλο στην ποιότητα του παραγόμενου από τα μοντέλα πεδίου παραμόρφωσης. Τελειώνουμε το κεφάλαιο αυτό παραθέτοντας ορισμένες παραλλαγές του μοντέλου VoxelMorph καθώς επίσης και την σχετική βιβλιογραφία. Οι παραλλαγές αυτές σχετίζονται με το είδος του νευρωνικού δικτύου που παράγει το πεδίο παραμόρφωσης. Έχοντας περιγράψει το βασικό μοντέλο που θα διαχειριστούμε θα πρέπει φυσικά να αφιερωθεί μία μεγάλη ενότητα στην προ-επεξεργασία ιατρικών εικόνων (Κεφάλαιο 4). Παραθέτουμε όλες τις κλασικές τεχνικές προ-επεξεργασίας καθώς επίσης και μία υψηλού επιπέδου περιγραφή της μεθοδολογίας που θα εφαρμόσουμε στην εργασία αυτή. Η στρατηγική που ακολουθήσαμε στην παρούσα εργασία αναλύεται στο κεφάλαιο 6. Μετά από όλα αυτά είναι καιρός επιτέλους να γνωρίσουμε τα δεδομένα μας (Κεφάλαιο 5). Τα δεδομένα διαφοροποιούνται σε εικόνες SUV (κανονικοποιημένες αξονικές τομογραφίες) και SEG (Segmentation Masks). Οι εικόνες SUV χρησιμοποιούνται για την εκπαίδευση, την επικύρωση αλλά και τον έλεγχο του μοντέλου. Οι εικόνες SEG χρησιμοποιούνται κυρίως για την αξιολόγηση (έλεγχο) του μοντέλου αλλά μπορούν να αξιοποιηθούν και κατά την διάρκεια της εκπαίδευσης, μετατρέποντας το πρόβλημα της ευθυγράμμισης σε ένα πρόβλημα επιβλεπόμενης μάθησης. Παρατίθενται κατά το δυνατόν περισσότερες πληροφορίες σχετικά με την πηγή και την φύση των δεδομένων καθώς επίσης μία γενική οπτική παρουσίαση αυτών. Έχοντας πλέον και τα μοντέλα και τα δεδομένα μπορούμε να περιγράφουμε αναλυτικά τα πειράματα που θα εκτελέσουμε καθώς επίσης και το υπολογιστικό περιβάλλον στο οποίο θα εκτελεστούν (Κεφάλαιο 6). Συγκεκριμένα εκπαιδεύουμε τέσσερις διαφορετικές θεμελιώδεις αρχιτεκτονικές νευρωνικών δικτύων προς παραγωγή του πεδίου παραμόρφωσης. Για κάθε αρχιτεκτονική δοκιμάζουμε διαφορετικές συναρτήσεις κόστους και κανονικοποίησης. Οι διάφορες υπερπαραμέτροι, καθώς επίσης και περισσότερες λεπτομέρειες περί των μοντέλων παρατίθενται στο ίδιο κεφάλαιο. Στο κεφάλαιο 7 αναλύονται τα αποτελέσματα των πειραμάτων. Βλέπουμε πώς μεταβάλλονται οι διάφορες μετρικές ομοιότητας μετά την ευθυγράμμιση των εικόνων καθώς επίσης και το προφίλ της συνάρτησης κόστους κατά την διάρκεια της εκπαίδευσης. Γίνεται επίσης αξιολόγηση του μοντέλου, τόσο κατά την εκπαίδευση (μέσω του συνόλου επικύρωσης), αλλά φυσικά και μετά το πέρας της εκπαίδευσης στο σύνολο ελέγχου. Στο τελευταίο κεφάλαιο (Κεφάλαιο 8) κά-



νομε έναν τελευταίο εκτενή σχολιασμό των αποτελεσμάτων, δίνουμε οδηγίες-παροτρύνσεις για βελτίωση των αποτελεσμάτων και για περαιτέρω έρευνα. Τελειώνουμε φυσικά με την βιβλιογραφία που σχετίζεται με την παρούσα εργασία.

## Summary

*Image Registration*, also known as Image Fusion, is the procedure of aligning two or more images based on their depiction. Medical image registration aims to find an optimal *spatial transformation* that better maps subjacent anatomical structures. By spatial transformation we mean a matrix that has the same dimensions with the images that we wish to register and to this matrix we give the meaning of a *deformation vector field*. Image registration is used in many clinical applications like image guidance, movement monitoring, image reconstruction and segmentation, clinical condition comparison and so on. The typical (classical) image registration methods work by optimizing an objective function for each registration pair, a task that can be very time-consuming for large data. Most modern registration techniques are based on deep learning and can achieve similar, or even better results, comparing to classical optimization methods. When the training of the deep learning models is complete, during the application, the only step that we need to perform is to compute the output of the models. This output is a function of the adjusted (trained) weights of the neural network and the input and its computation is significantly faster. What deep learning-based image registration tries to accomplish is to develop a *machine learning model* that will be able to automatically and dynamically (after, of course, the end of the training process) register a give pair of images. What this thesis aims at, is the registration of medical FDG-PET images that have been produced at different dates using modern deep learning methods. This mapping between images can help a medical expert compare the medical condition of a cancer-inflicted patient between the two dates that the images have been produced. This comparison can in turn help us, for example, to deduce whether or not a patient has shown signs of improvement after some medication. In the present thesis we studied some variations of the deep learning registration model called *Voxelmorph*. Voxelmorph is the first and maybe the most renowned model of deep learning registration. The purpose of this model is to produce the deformation vector field, which is a matrix with the same dimensions as the images. The neural network takes as input two PET tomographies, the *moving* image and the *target* image. These images are processed by the network and the output is the desired deformation field. The purpose of the training phase is for our network to learn to produce a deformation field which, when applied to the moving image, produces a new image that is *as similar as possible* with the target image (according to some similarity metric). We have to note here that, according to the international bibliography, registration with deep learning is considered to be an unsupervised machine learning problem. We will see how we can transform it into a supervised version, although during our simulations we will only consider the unsupervised case. The structure of this thesis is as follows: We will obviously begin with a short introduction (Chapter 1) that will make us more familiar with certain concepts that we will need to understand and confront during the whole thesis. We continue with some required medical knowledge (Chapter 2), which is needed not only to comprehend the medical part of the thesis but also appreciate the importance of research of machine learning in the medical field. We describe some facts related with cancer and with certain diagnostic tools known as FDG-PET tomographies. Finally, we briefly describe the types of cancer that the patients

were histologically proven with. Next, we state the mathematical formulation of classical registration methods (Chapter 3.1) that are not based on deep learning. This short presentation will not only help us better understand the problem we are trying to solve but also make us appreciate the progress of deep learning techniques at solving a very difficult multi-variable optimization problem. We shall proceed to the deep learning part. A very thorough, with all the mathematical details needed, description of the Voxelmorph model is given (Chapter 3.2). We will examine exactly how the information of its input (the FDG-PET images) is processed in order to produce the deformation field. In the same chapter we formulate and describe the most widespread and widely used similarity metrics and, of course, the corresponding loss function with which we will train our models. Some regularization functions are also described. These mainly affect the quality of the deformation vector field. We end this chapter with some state-of-the-art variations of Voxelmorph and the related references. These variations are based on the structure of the neural network that produces the deformation field. Having described the basic model that we will work with, we must certainly dedicate a whole chapter to the preprocessing of our images (Chapter 4). We analyze all the classical preprocessing techniques that are related to image processing and also provide as a high-level description of the preprocessing pipeline that we are going to follow. The details of the strategy we chose is described in more detail in Chapter 6. After all these chapters, it is finally time to meet our data (Chapter 5). They fall into two big categories, the SUV images (normalized tomographies) and the SEG images (Segmentation Masks). The SUVs are used during the training, validation and test phase of the model. The SEG images are used mostly for the model evaluation, as part of the test set. They can also be used during training and when introducing the masks in the training process we transform the problem into a supervised machine learning problem. We give as much information, related to the source and the nature of the images, as possible and also a visual presentation. Now that we have the models and the data in our hands we can thoroughly describe the simulation we are going to perform and also the computational environment we are going to use (Chapter 6). More specifically, we train four different fundamental neural network architectures that will produce the deformation field for us. For each architecture we try different cost and regularization functions. The various hyperparameters, as well as more details regarding the neural networks, are given in the same chapter. In Chapter 7 we analyze the results of our experiments. We examine how some similarity metrics change after the registration as well as the loss function profile during training. We also evaluate our model, not only during the training phase (using the validation set), but also after the training using the test set. At the last chapter (Chapter 8) we make more extensive comments related to the results and also give some guidance for further improvements and research. The thesis ends with the relevant bibliography.

## Ευχαριστίες

Ευχαριστώ θερμά τον υποψήφιο διδάκτορα Θωδωρή-Παναγιώτη Βαγενά καθώς επίσης και τον καθηγητή Γεώργιο Μασόπουλο για την συνεχή υποστήριξη και καθοδήγηση καθ'όλη την διάρκεια εκπόνησης της παρούσας διπλωματικής εργασίας.

# 1 Εισαγωγή

Είναι αδιαμφισβήτητο πώς στις μέρες μας η μηχανική μάθηση και η τεχνητή νοημοσύνη παίζουν καθοριστικό ρόλο στην ανάπτυξη και την πρόοδο της τεχνολογίας και άρα και της κοινωνίας. Μέσα σε όλο αυτό το κλίμα ανάπτυξης είναι εύλογο να αναρωτηθεί κανείς τι ακριβώς είναι η μηχανική μάθηση και ποια είναι η διαφορά με την τεχνητή νοημοσύνη. Με απλά λόγια, η μηχανική μάθηση είναι ένα σύνολο υπολογιστικών τεχνικών που λαμβάνουν δεδομένα εισόδου καθώς και δεδομένα εξόδου και μας επιστρέφουν την λογική-αλγόριθμο με την οποία μπορούμε από τα δεδομένα εισόδου να φτάσουμε στα δεδομένα εξόδου. Υποσύνολο της μηχανικής μάθησης είναι και η λεγόμενη βαθιά μάθηση, με την οποία θα ασχοληθούμε προφανώς και στην παρούσα εργασία. Η υπολογιστική τεχνική που χρησιμοποιεί η βαθιά μάθηση είναι τα τεχνητά νευρωνικά δίκτυα, που δεν είναι τίποτα άλλο παρά αριθμητικές δομές που μπορούν να προσεγγίσουν αυθαιρέτως καλά μία συνεχή συνάρτηση (για τον λόγο αυτόν συνήθως αναφερόμαστε στα νευρωνικά δίκτυα ως συναρτήσεις). Η τεχνητή νοημοσύνη είναι συνήθως κάποιο πρόγραμμα που λαμβάνει τα δεδομένα εισόδου καθώς επίσης και την ίδια την λογική με την οποία θα θέλαμε από τα δεδομένα εισόδου να πάρουμε δεδομένα εξόδου και μας επιστρέφει τα επιθυμητά δεδομένα εξόδου. Εδώ δηλαδή θα πρέπει να δηλώσουμε εμείς ρητά στο λογισμικό μας με ποιον τρόπο θέλουμε τα δεδομένα μας να υποστούν επεξεργασία. Επειδή, όπως είπαμε, εδώ θα ασχοληθούμε με την μηχανική μάθηση θα περιγράψουμε συνοπτικά τους διάφορους κλάδους της μηχανικής μάθησης καθώς επίσης θα παραθέσουμε και διάφορα στατιστικά σχετικά με την κολοσιαία ανάπτυξη και εφαρμογή της.

Η μηχανική μάθηση, πέρα από την πρακτικότητά της, αποτελεί και έναν συνεχώς αναπτυσσόμενο ερευνητικό κλάδο. Ερευνητές και μηχανικοί παγκοσμίως αναλύουν και αναπτύσσουν συνεχώς καινούριους αλγόριθμους μηχανικής μάθησης με σκοπό όχι μόνο της αυτοματοποίησης πολλών εφαρμογών της καθημερινότητας αλλά και στην επιτυχή πρόβλεψη καθημερινών φαινομένων, όπως ο καιρός, η οικονομία και η υγεία. Όπως και κάθε άλλο ερευνητικό πεδίο, έτσι και η μηχανική μάθηση έχει την δικιά της αγορά, η οποία συνεχώς αυξάνεται. Το 2021 υπολογίζεται ότι περίπου 16 δισ. δολάρια ξοδεύτηκαν στην έρευνα που αφορά την μηχανική μάθηση. Το 2022 το ποσό αυτό ανέβηκε στα 21 δισ. ενώ μέχρι το 2029 προβλέπεται ότι το ποσό θα αυξηθεί στα 209 δισ. [1]. Παρακάτω παραθέτουμε έναν πίνακα με εφαρμογές της μηχανικής μάθησης καθώς επίσης και το ποσοστό υλοποίησης της κάθε εφαρμογής από εταιρείες παγκοσμίως [1]:

Εφαρμογή	Ποσοστό Υλοποίησης (%)
Βελτίωση εμπειρίας πελατών	57
Παραγωγή γνώσης από πελάτες	50
Αλληλεπίδραση με πελάτες	48
Ανίχνευση απάτης	46
Παράταση συμμετοχής πελατών	44
Αύξηση αφοσίωσης πελατών	40
Απόκτηση νέων πελατών	34
Προώθηση μάρκας	31
Συστήματα συστάσεων	27
Άλλα	1

Πίνακας 1: Συνηθέστερες εφαρμογές μηχανικής μάθησης σε εταιρίες παγκοσμίως

Όπως γίνεται προφανές από τον πίνακα 1, ο λόγος εφαρμογής τεχνικών μηχανικής μάθησης στις εταιρίες είναι κυρίως η προσέγγιση νέων πελατών αλλά και η διατήρηση των ήδη υπάρχοντων πελατών. Για παράδειγμα, το Netflix κατάφερε την τελευταία διετία να σώσει 1 δισ. δολάρια χάρη σε αλγορίθμους μηχανικής μάθησης που σχετίζονται με συστήματα συστάσεων. Άλλα ενδιαφέροντα επιτεύγματα της μηχανικής μάθησης είναι [2]:

- Το 97 % των χρηστών κινητών τηλεφώνων χρησιμοποιούν κάποια εφαρμογή φωνητικής βοήθειας
- Λογισμικό της Google βασισμένο στην μηχανική μάθηση μπορεί να εντοπίσει με ακρίβεια 89 % καρτίνο του μαστού
- Μέχρι το 2025, τα 3/4 όλων των υπηρεσιών φροντίδας ηλικιωμένων θα εκτελούνται από συστήματα μηχανικής μάθησης
- Η ακρίβεια πρόβλεψης των τιμών των μετοχών από αλγορίθμους μηχανικής μάθησης φτάνει το 65 %

Κάτι που δεν φαίνεται είναι το γεγονός ότι η μηχανική μάθηση αλλά και οι συναφείς κλάδοι δημιουργούν και νέες θέσεις εργασίας για πολλούς νέους. Μηχανικοί μηχανικής μάθησης αλλά και επιστήμονες δεδομένων είναι μόνο δύο από τις πολλές επιλογές που έχουν όσοι ασχολούνται με την μηχανική μάθηση. Συνεχίζουμε τώρα με τις διάφορες κατηγορίες μηχανικής μάθησης:

- **Επιβλεπόμενη Μάθηση**

Αποτελεί ένα σύνολο αλγορίθμων που εκτός από τα δεδομένα εισόδου απαιτούν επίσης και *ετικέτες*, δηλαδή αριθμούς που να δηλώνουν σε ποια κατηγορία ανήκει το κάθε δεδομένο. Τα δεδομένα εισόδου, μαζί με τις αντίστοιχες ετικέτες τροφοδοτούνται στους αλγορίθμους επιβλεπόμενης μάθησης και εν τέλει παίρνουμε έναν αλγόριθμο με τον οποίο μπορούμε σε πολύ μικρό χρόνο να αντιστοιχίσουμε τα δεδομένα μας σε κάποια κατηγορία. Στους αλγορίθμους μηχανικής μάθησης υπάγονται οι μηχανές διανυσματικής υποστήριξης (Support Vector Machines), η γραμμική και λογιστική παλινδρόμηση (Linear and Logistic Regression), τα δέντρα αποφάσεων (Decision Trees), οι K-κοντινότεροι γείτονες (K-nearest neighbors) και τα νευρωνικά δίκτυα. Τα νευρωνικά δίκτυα, παρόλο που μπορούν να χρησιμοποιηθούν στα πλαίσια της επιβλεπόμενης μάθησης, χρησιμοποιούνται ευρέως και στις άλλες κατηγορίες που ακολουθούν παρακάτω.

- **Μη-Επιβλεπόμενη Μάθηση**

Στην περίπτωση αυτή οι αλγόριθμοι δέχονται μόνο τα δεδομένα εισόδου απουσία ετικετών. Συνήθως εφαρμογές μη-επιβλεπόμενης μάθησης σχετίζονται με την ανακάλυψη κρυφών μοτίβων στα δεδομένα εισόδου ή με την ανακατασκευή των δεδομένων. Στην μη-επιβλεπόμενη μάθηση υπάγεται ένα μεγάλο σύνολο αλγορίθμων γνωστό ως *συσταδοποίηση* (clustering). Οι αλγόριθμοι αυτοί έχουν ως στόχο να εκμεταλλευτούν κάποια ενδογενή γεωμετρία ή μοτίβα που μπορεί να διαθέτουν τα δεδομένα μας έτσι ώστε να τα κατηγοριοποιήσουν κατάλληλα. Οι πιο γνωστοί αλγόριθμοι συσταδοποίησης είναι οι K-μέσοι (K-means) και η ιεραρχική συσταδοποίηση (hierarchical clustering). Πέρα από την συσταδοποίηση, έχουμε επίσης και την Ανάλυση σε Κύριες Συνιστώσες (Principal Component Analysis) και την Ανάλυση σε Ανεξάρτητες Συνιστώσες (Independent Component Analysis). Τέλος, τα τελευταία χρόνια, με την ραγδαία ανάπτυξη των βαθιών νευρωνικών δικτύων, έχουν αναπτυχθεί δομές που μπορούν να χρησιμοποιηθούν στα πλαίσια της μη-επιβλεπόμενης μάθησης. Τέτοια δίκτυα είναι, για παράδειγμα, οι αυτοκωδικοποιητές (autoencoders), οι μηχανές Boltzmann (Boltzmann machines) και τα δίκτυα βαθιάς πεποίθησης (Deep Belief Networks).

- **Ενισχυτική Μάθηση**

Η φιλοσοφία της ενισχυτικής μάθησης είναι τελείως διαφορετική από τις προηγούμενες δύο κατηγορίες. Οι αλγόριθμοι ενισχυτικής μάθησης βασίζονται στην αλληλεπίδραση ενός *πράκτορα* (agent) με ένα *περιβάλλον*. Στόχος του πράκτορα είναι να ξεκινήσει από μία αρχική κατάσταση στο περιβάλλον και, εκτελώντας "νόμιμες" κινήσεις εντός του περιβάλλοντος, να φτάσει σε μία επιθυμητή τελική κατάσταση με τον βέλτιστο δυνατό τρόπο. Ο βέλτιστος τρόπος αυτός ισοδυναμεί με το να λάβει ο πράκτορας το μέγιστο δυνατό *κέρδος*. Το κέρδος αυτό το ορίζει ο προγραμματιστής του αλγορίθμου. Για παράδειγμα, ο πράκτορας μπορεί να λάβει κέρδος +1 αν εκτελέσει μία κίνηση που τον φέρνει πιο κοντά στην επιθυμητή τελική

κατάσταση και 0 ή -1 αν εκτελέσει μία κακή κίνηση που τον απομακρύνει από τον τελικό του στόχο. Η μαθηματική θεμελίωση της ενισχυτικής μάθησης γίνεται μέσω των *Διαδικασιών Απόφασης κατά Markov* (Markov Decision Process). Η ενισχυτική μάθηση συνδυάζεται με υπέροχο τρόπο με τα βαθιά νευρωνικά δίκτυα και παίρνουμε έτσι την *Βαθιά ενισχυτική μάθηση* (Deep Reinforcement Learning), ένα σύνολο αλγορίθμων που παράγουν εκπληκτικές εφαρμογές.

Οι παραπάνω τρεις κατηγορίες είναι οι πιο συνηθισμένες κατηγορίες μηχανικής μάθησης. Η επιβλεπόμενη και η μη-επιβλεπόμενη μάθηση βρίσκουν, λόγω της φύσης των προβλημάτων που λύνουν, μεγαλύτερη εφαρμογή στην βιομηχανία, την οικονομία και την υγεία. Η ενισχυτική μάθηση χρησιμοποιείται για πιο σύγχρονες εφαρμογές, όπως ο σχεδιασμός συστημάτων αυτόματης ρύθμισης και η αυτοματοποίηση συστημάτων, όπως τα αυτοκίνητα. Στην παρούσα εργασία δεν θα ασχοληθούμε με την ενισχυτική μάθηση, αλλά με την εφαρμογή των νευρωνικών δικτύων στα πλαίσια της μη-επιβλεπόμενης μάθησης για την επίλυση ενός προβλήματος που σχετίζεται με τον κλάδο της ιατρικής.

Πριν από μερικά χρόνια ίσως να αποτελούσε έκπληξη ότι ένας κλάδος της πληροφορικής θα μπορούσε να βοηθήσει τόσο την ιατρική επιστήμη στο να αναπτυχθεί και να βοηθήσει την ανθρωπότητα. Παρόλο που στον τομέα της υγείας οι προδιαγραφές και οι απαιτήσεις είναι πολύ μεγαλύτερες και τα δεδομένα είναι πολύ πιο δύσκολο να παραχθούν σε σχέση με άλλες επιστήμες, η έρευνα δεν σταματά καθώς τα μακροπρόθεσμα οφέλη θεωρείται ότι θα είναι μεγάλα. Τα πιο συνηθισμένα προβλήματα που τίθεται να λύσει η μηχανική μάθηση στον τομέα της υγείας είναι η ταξινόμηση και η ανάλυση εικόνας. Όταν λέμε ταξινόμηση εννοούμε με απλά λόγια την πρόβλεψη του πιο πιθανού αποτελέσματος, από κάποιο σύνολο διακριτών επιλογών, με βάση κάποια χαρακτηριστικά. Για παράδειγμα, αν τα χαρακτηριστικά μας είναι διάφοροι δείκτες του οργανισμού ενός ασθενούς, τότε ενδεχομένως να θέλουμε να μάθουμε εάν ο ασθενής αυτός πάσχει από κάποια ασθένεια ή όχι. Αυτό θα ήταν ένα δυαδικό πρόβλημα λογιστικής παλινδρόμησης (που στην πραγματικότητα αποτελεί αλγόριθμο ταξινόμησης και όχι παλινδρόμησης). Η ανάλυση εικόνας στα πλαίσια της ιατρικής μπορεί να σημαίνει είτε ταξινόμηση, είτε κατάτμηση ή ευθυγράμμιση. Η ταξινόμηση των εικόνων σχετίζεται συνήθως με την διάγνωση ενός ασθενούς. Η εικόνα μπορεί να είναι, για παράδειγμα, μία ακτινογραφία θώρακος και το αποτέλεσμα να είναι "πνευμονία" ή "όχι πνευμονία". Η κατάτμηση αφορά την αναγνώριση διακεκριμένων περιοχών της εικόνας, όπως για παράδειγμα τα όργανα του σώματος ή όγκοι. Τέλος η ευθυγράμμιση, με την οποία προφανώς θα ασχοληθούμε και εμείς, αποσκοπεί στο να μετασχηματίσουμε μία εικόνα, έτσι ώστε να συμπέσει κατά το δυνατόν καλύτερα πάνω σε μία άλλη. Ο μετασχηματισμός αυτός μπορεί να μας βοηθήσει να ποσοτικοποιήσουμε κατά πόσο μία εικόνα άλλαξε από μία χρονική στιγμή σε μία άλλη. Αυτή η αλλαγή με την σειρά της μπορεί να μας βοηθήσει να καταλάβουμε αν ο οργανισμός έχει ανταποκριθεί σε μία θεραπεία ή αν μία ασθένεια έχει προχωρήσει. Τονίζεται εδώ ότι η ευθυγράμμιση δεν αποτελεί τεχνική διάγνωσης, όπως η ταξινόμηση. Η ευθυγράμμιση είναι το βήμα πριν την διάγνωση. Ο ειδικός θα εκτελέσει πρώτα την ευθυγράμμιση μεταξύ δύο ιατρικών εικόνων και σε μετέπειτα στάδιο θα βγάλει συμπεράσματα για την υγεία του ασθενούς. Εμείς εδώ δεν θα ασχοληθούμε με την διάγνωση, αλλά μόνο



με τις τεχνικές ευθυγράμμισης από την πλευρά της πληροφορικής και συγκεκριμένα από την πλευρά της βαθιάς μάθησης.

## 2 Προαπαιτούμενες Ιατρικές Γνώσεις

### 2.1 Ο καρκίνος ως μία σύγχρονη ασθένεια

Στις μέρες μας, είναι σχεδόν σίγουρο ότι όλοι μας έχουμε έρθει, είτε άμεσα είτε έμμεσα, σε επαφή με την ασθένεια (ή ακόμα ορθότερα με την κατηγορία ασθενειών) του καρκίνου. Συνεπώς όλοι γνωρίζουμε περί τίνος πρόκειται και σε πόσο σοβαρή και βάνουση κατάσταση μπορεί να φέρει τον ασθενή. Το πρώτο βήμα στην κατανόηση της ασθένειας είναι να δούμε πώς πυροδοτείται, πώς δηλαδή ξεκινά η ανάπτυξη του καρκίνου στο ανθρώπινο σώμα.

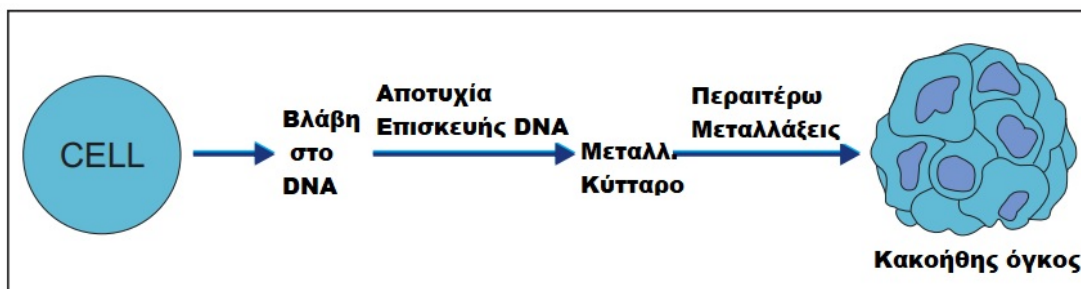
Ο καρκίνος, εκτός από μία πολύ σοβαρή πάθηση, είναι και μία πλέον ιδιαίτερα συχνή πάθηση. Σύμφωνα με τον αμερικάνικο οργανισμό CDC (Centers for Disease Control and Prevention) ο καρκίνος είναι πλέον η δεύτερη συνηθέστερη αιτία θανάτου στις Ηνωμένες Πολιτείες, όπως φαίνεται και από τον παρακάτω πίνακα [5]:

Αιτία Θανάτου	Θάνατοι ανά 100,000 άτομα
Καρδιακές Ασθένειες	168.2
Καρκίνος	144.1
COVID-19	85
Ατυχήματα	57.6
Εγκεφαλικό	38.8
Χρόνια αναπνευστική ασθένεια	36.4
Νόσος Alzheimer	32.4
Διαβήτης	24.8
Γρίπη και πνευμονία	13
Νεφρικές ασθένειες	12.7

Πίνακας 2: Συνηθέστερες αιτίες θανάτου στις Ηνωμένες Πολιτείες το 2020

Όταν αναφερόμαστε σε "καρκίνο", εννοούμε την *ανώμαλη ανάπτυξη μίας ομάδας κυττάρων*. Από την στοιχειώδη βιολογία γνωρίζουμε ότι στο εσωτερικό των κυττάρων και, πιο συγκεκριμένα στον πυρήνα τους, βρίσκεται η γενετική πληροφορία του οργανισμού, δηλαδή το DNA. Πρόκειται για ένα ιδιαίτερα περίπλοκο μόριο που κωδικοποιεί όλες τις πληροφορίες-οδηγίες που χρειάζεται ο οργανισμός για να επιτελέσει όλες τις λειτουργίες του. Το λεγόμενο *κεντρικό δόγμα της βιολογίας* μας λέει ότι η φυσική πορεία του DNA είναι η *αντιγραφή, η μεταγραφή και η μετάφραση*. Συνοπτικά, τα βήματα αυτά συνιστούν την παραγωγή δύο νέων μορίων DNA από το αρχικό μητρικό μόριο (αντιγραφή), την μετατροπή του ενός από τα δύο σε *αξιοποιήσιμη πληροφορία*, το λεγόμενο mRNA (μεταγραφή) και τέλος στην παραγωγή πρωτεϊνών (μετάφραση). Από τα τρία αυτά στάδια, το πιο "ευάλωτο" σε λάθη αλλά και κρίσιμο για τον οργανισμό είναι η αντιγραφή. Αν, για οποιονδήποτε λόγο, γίνει κάποιος λάθος στην αντιγραφή του DNA και η βλάβη δεν επισκευασθεί από τους μηχανισμούς του κυττάρου τότε το νέο *μεταλλαγμένο* μόριο DNA (και άρα το μεταλλαγμένο κύτταρο στο οποίο στεγάζεται) θα δίνει λάθος οδηγίες στα υπόλοιπα οργανίδια του κυττάρου με συνέπεια να παράγονται λάθος (και σε πολλές περιπτώσεις επικίνδυνες) πρωτεΐνες.

Τα μεταλλαγμένα κύτταρα με την σειρά τους πολλαπλασιάζονται, ενδεχομένως προσθέτοντας επιπλέον μεταλλάξεις στο ήδη μεταλλαγμένο DNA και καταλήγουν να σχηματίζουν αυτό που ξέρουμε ως *κακοήθη όγκο*, γνωστό και ως *καρκινικό όγκο*. Η παραπάνω συνοπτική περιγραφή φαίνεται και στην παρακάτω εικόνα:



Εικόνα 1: Σχηματική αναπαράσταση της πορείας του καρκίνου

Παρακάτω δίνουμε επίσης έναν πίνακα ονοματολογίας με κάποιες κατηγορίες ιστών και τα αντίστοιχα είδη όγκων που μπορούν να εμφανισθούν [4]:

Ιστός	Τύπος Κυττάρων	Καλοήθης Όγκος	Κακοήθης Όγκος
Δέρμα	Πλακώδες επιθήλιο	Ακανθοκυτταρικό θηλώμα	Πλακώδες καρκίνωμα Κακοήθες μελάνωμα
Πνεύμονας	Αναπνευστικό επιθήλιο		Αδενοκαρκίνωμα Πλακώδες καρκίνωμα Μικροκυτταρικό καρκίνωμα
Ουροποιητικό Σύστημα	Ουροθήλιο (μεταβατικό επιθήλιο)	Θηλώμα μεταβατικών κυττάρων	Καρκίνωμα μεταβατικών κυττάρων
Κόκαλα	Οστεοκύτταρα	Οστέωμα	Οστεοσάρκωμα
Αιμοφόρα αγγεία	Ενδοθηλιακά κύτταρα	Αιμαγγείωμα	Αγγειοσάρκωμα
Αιμοποιητικό σύστημα	Ερυθρά Αιμοσφαίρια		Οξεία μυελογενής λευχαιμία Χρόνια μυελογενής λευχαιμία
Ανοσοποιητικό σύστημα	Λεμφοειδή κύτταρα		Οξεία λεμφοβλαστική λευχαιμία Χρόνια λεμφοβλαστική λευχαιμία Λέμφωμα Hodgkin
Κεντρικό Νευρικό Σύστημα	Αστροκύτταρα και ολιγοδενδροκύτταρα		Αστροκύτωμα και ολιγοδενδρογλοίωμα

Πίνακας 3: Παραδείγματα τύπων καρκίνου σε διάφορους ιστούς

Προφανώς ο πίνακας 3 είναι ελλιπής και ο σκοπός του είναι να δώσει μία ένδειξη για την τεράστια ποικιλία που περιλαμβάνει το σύνολο των παθήσεων που ονομάζουμε καρκίνο. Μία

ομάδα κυττάρων που δεν εμφανίζεται στον παραπάνω πίνακα -και όχι τυχαία- είναι αυτή των καρδιακών κυττάρων. Πράγματι σπάνια ακούει κανείς για καρκίνο στην καρδιά. Ο καρκίνος μπορεί να ξεκινήσει από ένα και μόνο κύτταρο στο DNA του οποίου συνέβη λάθος αντιγραφή. Από το γεγονός αυτό καταλαβαίνουμε ότι είναι πιο πιθανό να εμφανισθεί καρκίνος σε ομάδες κυττάρων που πολλαπλασιάζονται πιο συχνά και πιο γρήγορα σε σχέση με άλλες. Στους ανθρώπους, η ικανότητα των περισσότερων καρδιακών μυικών κυττάρων να αναπαράγονται εξαφανίζεται λίγο μετά την γέννηση. Συνεπώς, σε κύτταρα που δεν αντιγράφονται δεν μπορούμε να έχουμε λάθη αντιγραφής και, άρα, καρκίνο. Παρόλα αυτά, τα καρκινικά κύτταρα μπορούν να υποστούν την λεγόμενη *μετάσταση*, δηλαδή να μεταναστεύσουν σε κοντινούς ή μακρινούς υγιείς ιστούς. Τα αιμοφόρα αγγεία είναι ευάλωτα στον καρκίνο και συνεπώς, σε πολύ σπάνιες περιπτώσεις, μπορεί να επηρεάσουν και την καρδιά.

Από τον πίνακα 2 γίνεται άμεσα αντιληπτό ότι (όχι μόνο στις Ηνωμένες Πολιτείες) ο καρκίνος αφαιρεί την ζωή από πάρα πολλούς ανθρώπους ετησίως. Έχει ενδιαφέρον να μελετήσει κανείς ποιοι τύποι καρκίνων είναι πιο συνηθισμένοι στον πληθυσμό. Ο οργανισμός NIH (National Cancer Institute) παραθέτει τον παρακάτω πίνακα που περιλαμβάνει τον τύπο καρκίνου, τα εκτιμώμενα νέα περιστατικά και τους εκτιμώμενους θανάτους το 2022 [6]:

Τύπος Καρκίνου	Εκτιμώμενα Περιστατικά	Εκτιμώμενοι Θάνατοι	Θνησιμότητα (%)
Στήθος (Γυναίκες-Άντρες)	287,850 - 2,710	43,250 - 530	15.02-19.56
Προστάτης	268,490	34,500	12.85
Πνεύμονας	236,740	130,180	54.98
Παχέος εντέρου και ορθού	151,030	52,580	34.81
Μελάνωμα	99,780	7,650	7.66
Κύστη	81,180	17,100	21.06
Μη-Hodgkin Λέμφωμα	80,470	20,250	25.16
Νεφρά	79,000	13,920	17.62
Ενδομήτριο	65,950	12,550	19.02
Πάγκρεας	62,210	49,830	80.09
Λευχαιμία (κάθε είδους)	60,650	24,000	39.57
Θυροειδής	43,800	2,230	5.09
Ήπαρ και Ενδοηπατικός Χολικός πόρος	41,260	30,520	73.96

Πίνακας 4: Θνησιμότητα συχνών τύπων καρκίνου

Παρατηρούμε ότι παρόλο που ο καρκίνος του στήθους (κυρίως στις γυναίκες) και του προστάτη είναι οι πιο συχνά εμφανιζόμενοι τύποι, δεν είναι οι πιο θανατηφόροι. Οι τύποι καρκίνου με τις μεγαλύτερες (%) θνησιμότητες είναι του παγκρέατος (80.09%), του ήπατος (73.96%) και του πνεύμονα (54.98%). Τα στατιστικά αυτά δεν είναι τυχαία και για να

κατανοήσουμε το μοτίβο που φανερώνει ο πίνακας 4 θα πρέπει να μελετήσουμε αν υπάρχουν μη-ανθρώπινοι (δηλαδή τεχνητοί) παράγοντες που μπορεί να συμβάλλουν στην εμφάνιση καρκινικών κυττάρων μέσω της βλάβης που μπορεί να προκαλέσουν στο DNA ή στην διεργασία αντιγραφής του.

Υπάρχουν πολλοί λόγοι για τους οποίους θα μπορούσε να γίνει κάποιο λάθος στην αντιγραφή του DNA. Παρόλα αυτά θα πρέπει να σκεφτούμε το εξής, ίσως μακάβριο, γεγονός που γνωρίζουν όλοι οι γιατροί και οι συναφείς επιστήμονες: Καθώς το DNA μας αντιγράφεται συνεχώς, χωρίς καμία παύση μέχρι το θάνατό μας, λάθη στην αντιγραφή του γίνονται σχεδόν συνεχώς. Αυτό δεν είναι και τόσο παράλογο (πιθανοκρατικά τουλάχιστον) αν σκεφτούμε πόσο τεράστιο είναι ένα μόριο DNA και πόσες φορές επιτελείται η διεργασία της αντιγραφής κατά την διάρκεια της ζωής μας. Παρόλα αυτά το σώμα μας έχει διορθωτικούς μηχανισμούς που είναι διαθέσιμοι ανά πάσα στιγμή να δράσουν, έτσι ώστε να διορθώσουν το λάθος της αντιγραφής. Οι μηχανισμοί αυτοί έχουν απίστευτα μεγάλη απόδοση και αυτό το έχουμε καταλάβει μετρώντας κατά μέσο όρο πόσα λάθη γίνονται καθημερινά στην αντιγραφή του DNA μας το οποίο, αν δεν επισκευαζόταν άμεσα, θα εμφανίζαμε όλοι μας προβλήματα σε πολύ μικρή ηλικία. Η απόδοσή τους όμως προφανώς δεν είναι τέλεια. Αυτό σημαίνει ότι ο καθένας από εμάς, αν ζούσε ικανοποιητικά αρκετά χρόνια, ακόμα και υπό τις πιο "βέλτιστες" δυνατές συνθήκες, θα εμφάνιζε στην ζωή του κάποια μορφή καρκίνου. Και όταν λέμε βέλτιστες συνθήκες εννοούμε μακριά από εξωγενείς παράγοντες που βοηθούν στην εμφάνιση καρκίνου, οι γνωστοί μας *καρκινογόνοι παράγοντες*. Τέτοιοι παράγοντες είναι βλαβερές συνήθειες όπως το κάπνισμα, το αλκοόλ, η ακτινοβολία και διάφορα τοξικά μόρια (π.χ. ο άσβεστος, βενζοπυρένια και οξείδια του αιθυλενίου). Αυτοί οι παράγοντες αυξάνουν την πιθανότητα εμφάνισης καρκίνου και δυστυχώς συνήθειες όπως το κάπνισμα και η εκτεταμένη κατανάλωση αλκοόλ μαστίζουν ακόμα και την σημερινή κοινωνία.

Με βάση τον Παγκόσμιο Οργανισμό Υγείας (World Health Organization) οι καρκινογόνες ουσίες κατατάσσονται σε τέσσερις κατηγορίες, ανάλογα με την πιθανότητα να προκαλέσουν καρκίνο στον άνθρωπο. Οι κατηγορίες αυτές είναι οι εξής:

<b>Ομάδα 1</b>	Καρκινογόνα για τους ανθρώπους	122 ουσίες
<b>Ομάδα 2A</b>	Μεγάλη πιθανότητα πρόκλησης καρκίνου	93 ουσίες
<b>Ομάδα 2B</b>	Μικρή πιθανότητα πρόκλησης καρκίνου	319 ουσίες
<b>Ομάδα 3</b>	Όχι καρκινογόνα για τους ανθρώπους	501 ουσίες

Πίνακας 5: Κατηγοριοποίηση καρκινογόνων ουσιών [7]

Η ομάδα 3 αναφέρεται σε ουσίες που είναι μεν καρκινογόνες αλλά όχι για τους ανθρώπους. Παρακάτω δίνουμε μερικά παραδείγματα καρκινογόνων ουσιών από την κάθε

ομάδα.

<b>Ομάδα 1</b>	Βίνυλο-Χλωρίδιο Κάδμιο Αρσενικό Βενζένιο Αιθανόλη Κυκλοσπορίνη Ταμοξιφένη Φορμαλδεύδη Άσβεστος Χ, γ ακτινοβολία Ιός Epstein-Bar Κάπνισμα
<b>Ομάδα 2A</b>	Αναβολικά Στεροειδή Ιός HPV Στυρένιο DDT Βίνυλο-Βρωμίδιο Ανιλίνη Κοβάλτιο
<b>Ομάδα 2B</b>	Αλόε Βέρα Πυριδίνη Χλωροπρένιο Διγοξίνη Ακεταμίδη Οξαζεπάμη Ισοπρένιο Νικέλιο
<b>Ομάδα 3</b>	Τσάι Χλωραμίνη Φαινόλες Προπυλένιο Ευλένιο Αμάρανθος Πολυστυρένιο

Πίνακας 6: Παραδείγματα καρκινογόνων ουσιών

Όλη αυτή η υπό-ενότητα αφιερώθηκε σε μία συνοπτική αναφορά στους διάφορους τύπους καρκίνου, στα αίτια που μπορεί να τον προκαλέσουν καθώς επίσης και σε μερικά ενδιαφέροντα στατιστικά δεδομένα σε ό,τι αφορά την συχνότητα εμφάνισης και την θνησιμότητα. Κάτι στο οποίο δεν αναφερθήκαμε καθόλου είναι προφανώς η θεραπεία. Ακόμα και σήμερα ο καρκίνος αποτελεί μίας από τις πιο δύσκολα αντιμετωπίσιμες ασθένειες παγκοσμίως καθώς δεν υπάρχει κάποια πλήρως αποτελεσματική θεραπεία, ειδικά αν ο ασθενής είναι σε προχωρημένο

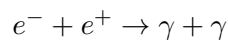
στάδιο. Παρόλα αυτά, όπως και με τις περισσότερες ασθένειες, η καλύτερη θεραπεία είναι η πρόληψη. Η δράση που δίνει τη μεγαλύτερη πιθανότητα επιβίωσης σε καρκινοπαθείς ασθενείς είναι η έγκαιρη διάγνωση, δηλαδή διάγνωση στα αρχικά στάδια της νόσου και φυσικά η οποιαδήποτε αγωγή προτείνουν οι γιατροί την εκάστοτε στιγμή. Παρόλο που στα πλαίσια της διπλωματικής αυτής, όπως αναφέρθηκε στην εισαγωγή, ο τελικός στόχος δεν είναι η διάγνωση ενός ασθενούς, θα πρέπει να κάνουμε και μία σύντομη αναφορά σε ένα πολύ σημαντικό διαγνωστικό εργαλείο του καρκίνου, δηλαδή τις τομογραφίες τύπου FDG-PET. Τέτοιες εικόνες θα χειριστούμε και στα πειράματα ευθυγράμμισης εικόνων και άρα θα βοηθήσουμε να έχουμε μία επίγνωση των τομογραφιών αυτών.

## 2.2 Τομογραφίες τύπου PET και FDG-PET

Το ακρωνύμιο PET σημαίνει *Positron Emission Tomography*, δηλαδή *Τομογραφία Εκπομπής Ποζιτρονίων*. Αποτελούν ιατρικές απεικονίσεις και συνιστούν ένα πολύτιμο διαγνωστικό εργαλείο. Παράγονται σε κατάλληλο εξοπλισμό που προφανώς ονομάζεται *Τομογράφος PET*. Το ακρωνύμιο FDG σημαίνει *18F-fluoro-deoxyglucose* και θα αναφερθούμε σε αυτό εφόσον κατανοήσουμε τον τρόπο λειτουργίας των PET τομογραφιών.

### 2.2.1 Το ποζιτρόνιο

Αρχικά θα κάνουμε μία σύντομη αναφορά στο *ποζιτρόνιο*. Το ποζιτρόνιο, γνωστό και ως *άντι-ηλεκτρόνιο*, είναι ένα υπό-ατομικό σωματίδιο που έχει ακριβώς την ίδια μάζα και σπιν με το ηλεκτρόνιο αλλά αντίθετο φορτίο. Ως γνωστόν, το ηλεκτρόνιο έχει αρνητικό φορτίο, ίσο με  $-1e = -1.602 \cdot 10^{-19} C$ . Το ποζιτρόνιο έχει λοιπόν φορτίο ίσο με  $+1e = +1.602 \cdot 10^{-19} C$ . Πρόκειται δηλαδή για ένα *αντι-σωματίδιο*. Τα ποζιτρόνια παράγονται φυσικά μέσω κάποιων ραδιενεργών διεργασιών αλλά και τεχνητά στο εργαστήριο. Μία ιδιότητα που έχουν τα αντι-σωματίδια είναι ότι όταν αλληλεπιδρούν με τα ομόλογα σωματίδιά τους εξαυλώνονται. Από την διεργασία αυτή παράγεται ενέργεια με την μορφή φωτονίων. Αν  $e^-$ ,  $e^+$ ,  $\gamma$  συμβολίζει ένα ηλεκτρόνιο, ποζιτρόνιο και φωτόνιο αντίστοιχα, τότε η διεργασία που λαμβάνει χώρα είναι η εξής:



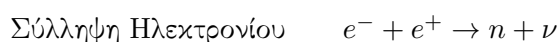
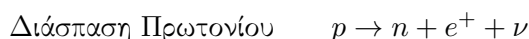
και ως γνωστόν ένα χβάντο ηλεκτρομαγνητισμού, δηλαδή ένα φωτόνιο, έχει ενέργεια ίση με  $hf$ , όπου  $h$  η σταθερά του Planck και  $f$  η συχνότητα του ηλεκτρονίου. Ο λόγος που παράγονται δύο φωτόνια είναι επειδή η παραπάνω διεργασία, όπως και κάθε άλλη φυσική διεργασία, θα πρέπει να υπακούει στους γνωστούς νόμους διατήρησης (ηλεκτρικού φορτίου, ορμής και στροφορμής). Με βάση αυτούς τους νόμους και φυσικά με βάση την ισοδυναμία μάζας-ενέργειας  $E = mc^2$ , προκύπτει ότι, σε χαμηλές ενέργειες, παράγονται δύο φωτόνια, καθένα από τα οποία έχει ενέργεια  $0.511 MeV$ . Σε υψηλές ενέργειες μπορεί να παραχθούν και άλλα σωματίδια που όμως δεν μας ενδιαφέρουν στις περισσότερες εφαρμογές. Θα δούμε

παρακάτω πώς αυτή η διεργασία εξαύλωσης αξιοποιείται στις τομογραφίες PET.

## 2.2.2 Τομογραφίες Εκπομπής Ποζιτρονίων

Οι τομογραφίες εκπομπής ποζιτρονίων είναι μία μέθοδος ποσοτικοποίησης βιοχημικών και φυσιολογικών διεργασιών *in vivo*<sup>1</sup>, χρησιμοποιώντας ραδιοφαρμακευτικές ουσίες που περιέχουν ραδιοπυρήνες που μπορούν να εκπέμπουν ποζιτρόνια, όπως <sup>11</sup>C, <sup>13</sup>N, <sup>15</sup>O, <sup>18</sup>F [8]. Ο χρόνος ημιζωής των ισοτόπων αυτών είναι 20,10,2 και 110 λεπτά αντίστοιχα. Παράγονται στο εργαστήριο μέσω μία διάταξης που ονομάζεται κύκλοτρο. Για τα προαναφερόμενα ισότοπα, ένα ικανοποιητικό ενεργειακό εύρος για την παραγωγή τους είναι τα 10-20 MeV. Υπάρχουν και άλλα ισότοπα που μπορούν να χρησιμοποιηθούν, όπως τα <sup>52</sup>Fe και <sup>55</sup>Co, που έχουν μεγαλύτερο χρόνο ημιζωής όμως είναι αρκετά πιο δύσκολο και ακριβό να παραχθούν.

Σε έναν τομογράφο PET μπορούν να λάβουν χώρα δύο διεργασίες, η διάσπαση πρωτονίου και η σύλληψη ηλεκτρονίου:



Το σωματίδιο  $\nu$  ονομάζεται νεutrίνο, έχει μηδενική μάζα ηρεμίας και δεν αλληλεπιδρά με την ύλη. Η διάσπαση πρωτονίου είναι δυνατό να λάβει χώρα αν υπάρχει αρκετή διαθέσιμη ενέργεια. Αν δεν υπάρχει, τότε θα λάβει χώρα σύλληψη ηλεκτρονίου.

Οι εξετάσεις τομογραφιών PET έχουν γίνει πλέον φυσικά θέμα ρουτίνας, λόγω των ποικίλων διαγνωστικών εφαρμογών. Υπάρχουν διάφορα είδη εξετάσεων PET, οι πιο συνηθισμένες από τις οποίες είναι οι εξής:

- Στατική Σάρωση: Η πιο συνηθισμένη εξέταση από την οποία παράγεται ένα σύνολο από εγχάρσιες τομές του οργανισμού
- Δυναμική Σάρωση: Εδώ παράγεται ένα σύνολο από διαδοχικές εικόνες στον χρόνο. Η χρονική αλληλουχία των σαρώσεων μπορεί προφανώς να δώσει πολύτιμες πληροφορίες για το πώς ο οργανισμός του ασθενούς ανταποκρίνεται σε μία διέγερση. Η χρονική διαφορά μεταξύ διαδοχικών σαρώσεων εξαρτάται από την εκάστοτε εφαρμογή και αποτελεί ακόμα και σήμερα ένα ενεργό ερευνητικό πεδίο
- Ολόσωμη Σάρωση: Στην εξέταση αυτή παράγεται μία τρισδιάστατη τομογραφία που περιέχει όλο το ανθρώπινο σώμα. Χρησιμοποιούνται κυρίως σε ογκολογικές εφαρμογές

---

<sup>1</sup>*In Vivo*: μία διεργασία που λαμβάνει χώρα σε έναν ζωντανό οργανισμό



Παρακάτω φαίνεται ένας τυπικός τομογράφος PET:

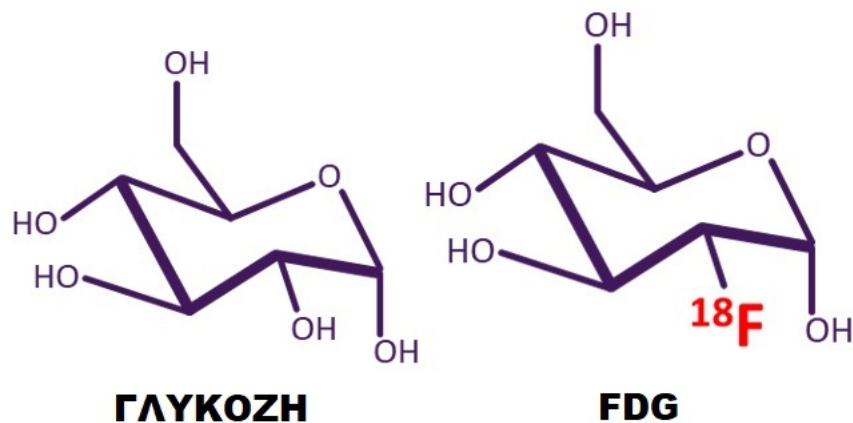


Εικόνα 2: Τομογράφος PET-CT [9]

Έχοντας πάρει τώρα μία ιδέα για το πώς λειτουργεί ένα τομογράφος PET, θα προχωρήσουμε σε μία ειδικότερη εφαρμογή της επιστήμης της ιατρικής απεικόνισης, τις τομογραφίες τύπου FDG-PET.

### 2.2.3 Τομογραφίες FDG-PET

Το FDG (ή  $^{18}\text{F}$ -FDG) είναι ένα ράδιο-φαρμακευτικό χημικό μόριο που ονομάζεται  $^{18}\text{F}$ -2-φθόρο-2-δεοξύ-D-γλυκόζη. Η μοριακή του δομή είναι πάρα πολύ παρόμοια με αυτή της γλυκόζης, όπως φαίνεται στην εικόνα 3. Η μόνη διαφορά είναι ότι ένα υδροξύλιο (-OH) έχει αντικατασταθεί με ένα φθόριο, που αποτελεί το ραδιενεργό κομμάτι του μορίου. Αυτό παράγει τα προ-αναφερόμενα ποζιτρόνια τα οποία εξαϊλώνονται με την αλληλεπίδρασή τους με ηλεκτρόνια προς παραγωγή ενέργειας.



Εικόνα 3: Μοριακή δομή της γλυκόζης και του FDG

Η χρήση του FDG στις τομογραφίες στηρίζεται στο γεγονός ότι μπορεί και προσλαμβάνεται από τους ιστούς που παρουσιάζουν αυξημένη μεταβολική δραστηριότητα. Τέτοιοι ιστοί είναι και οι καρκινικοί ιστοί. Επειδή τα καρκινικά κύτταρα έχουν την τάση να πολλαπλασιάζονται με πολύ γοργό ρυθμό, απαιτούν και τον μεταβολισμό μεγάλων ποσοτήτων γλυκόζης και συνεπώς αποτελούν ιδανικό στόχο για το FDG. Το FDG "προσβάλλει" τα καρκινικά κύτταρα, παραμένει σε αυτά και ακτινοβολεί ραδιενεργά. Συνεπώς δρα κατά μία έννοια σαν "φάρος" ακτινοβολίας. Σημεία του σώματος που ακτινοβολούν έντονα λόγω του FDG ανιχνεύονται από τους τομογράφους και απεικονίζονται ως φωτεινά σημεία στην τελική απεικόνιση.

Στο σημείο αυτό κλείνει και το κεφάλαιο που σχετίζεται με το ιατρικό κομμάτι της εργασίας. Παρόλο που δεν έγινε ιδιαίτερη εμβάθυνση στις ιατρικές και τεχνικές λεπτομέρειες του μηχανισμού του καρκίνου και των τομογραφιών, σίγουρα ο αναγνώστης έχει πάρει μία γενική ιδέα των εννοιών που χρειάζεται κανείς να ξέρει για την κατανόηση της παρούσας μελέτης.

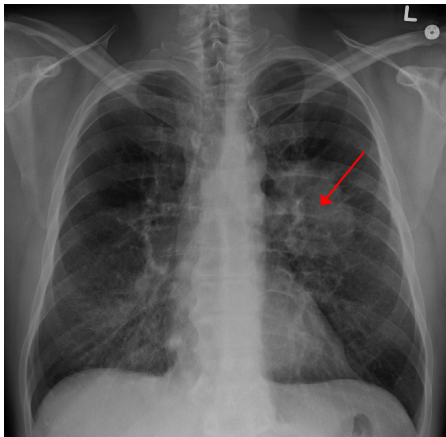
### 2.3 Οι υπό-μελέτη τύποι καρκίνων

Το σύνολο δεδομένων-εικόνων που θα μελετήσουμε στα επόμενα κεφάλαια περιλαμβάνει τρία είδη καρκίνων [44]: το *κακοήθες μελάνωμα*, το *λέμφωμα* και ο *καρκίνος του πνεύμονα*. Έχουμε επίσης και μερικούς ασθενείς που ανήκουν στο λεγόμενο σύνολο *αρνητικού ελέγχου*, δηλαδή ασθενείς που δεν παραλαμβάνουν κάποια φαρμακευτική αγωγή κατά την διάρκεια της μελέτης. Παρακάτω δίνουμε μία συνοπτική περιγραφή των υπό-μελέτη καρκίνων:

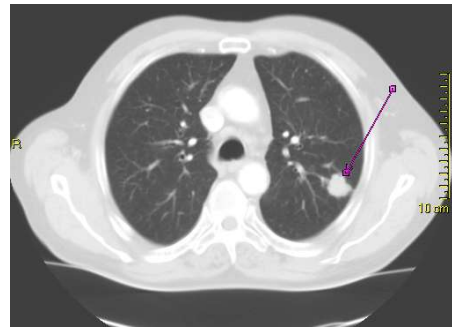
- **Κακοήθες Μελάνωμα:** Αποτελεί τον πιο σοβαρό τύπο καρκίνου του δέρματος και αναπτύσσεται στα κύτταρα που παράγουν μελανίνη. Μπορεί επίσης σπανίως να

αναπτυχθεί στα μάτια, στη μύτη και στο λαιμό. Ο πιο συνήθης παράγοντας που προκαλεί αυτόν τον καρκίνο είναι η υπεριώδης ακτινοβολία, η οποία προκαλεί βλάβη στο DNA.

- **Λέμφωμα:** Αποτελεί καρκίνο του λεμφικού συστήματος, που είναι μέρος του ανοσοποιητικού συστήματος. Το λεμφικό σύστημα περιλαμβάνει τους λεμφαδένες, τη σπλήνα, τον θυρεοειδή αδένα, τον θύμο αδένα και τον μυελό των οστών. Το λέμφωμα μπορεί να προσβάλει όλες αυτές τις περιοχές και, πιο συγκεκριμένα, τα κύτταρα που προσβάλλονται είναι τα λεμφοκύτταρα. Οι κυριότερες υπό-κατηγορίες του λεμφώματος είναι το λέμφωμα κατά *Hodgkin* και το λέμφωμα κατά *μη-Hodgkin*.
- **Καρκίνος του πνεύμονα:** Ίσως μία από τις πιο συνήθεις και οικείες σε όλους κατηγορία καρκίνου. Προσβάλει προφανώς τους πνεύμονες και σχετίζεται κυρίως με το κάπνισμα και τα αέρια τοξικά απόβλητα. Παρακάτω βλέπουμε μία ακτινογραφία (αριστερά) και μία αξονική τομογραφία (δεξιά) στις οποίες με βέλος φαίνεται ο κακοήθης όγκος στον πνεύμονα:



(a) Ακτινογραφία πνεύμονα με κακοήθη όγκο



(b) Αξονική τομογραφία πνεύμονα με κακοήθη όγκο

Εικόνα 4: Δύο διαφορετικές διαγνωστικές μέθοδοι για τον καρκίνο του πνεύμονα

### 3 Ευθυγράμμιση Εικόνων

Όπως αναφέρθηκε ήδη, το πρόβλημα της ευθυγράμμισης μίας εικόνας συνίσταται στο να ευθυγραμμίσουμε τους άξονες μίας εικόνας με δεδομένους άξονες. Οι άξονες αυτοί μπορεί να είναι οι άξονες μίας άλλης εικόνας και άρα το πρόβλημα υποβιβάζεται στο να μετατρέψουμε, με κατάλληλο μετασχηματισμό, την αρχική εικόνα στην δεύτερη. Είναι ένα σχετικά παλιό πρόβλημα που έχει μελετηθεί αρκετά, τόσο λόγω του θεωρητικού όσο και του πρακτικού ενδιαφέροντος. Υπάρχει μία πληθώρα τεχνικών που σχετίζονται με την ευθυγράμμιση εικόνων. Οι τεχνικές αυτές διακρίνονται σε κλασσικές και μη-κλασσικές. Όταν αναφερόμαστε σε κλασσικές τεχνικές ευθυγράμμισης εννοούμε αλγόριθμους που αποσκοπούν στην αριθμητική ελαχιστοποίηση μίας αντικειμενικής συνάρτησης κόστους. Μη-κλασσικές θεωρούνται οι τεχνικές που συνδυάζουν το μαθηματικό πλαίσιο της ευθυγράμμισης εικόνων με την μηχανική -και πιο συγκεκριμένα την βαθιά- μάθηση. Στην ενότητα αυτή θα περιγράψουμε σύντομα το μαθηματικό πλαίσιο των κλασσικών μεθόδων ευθυγράμμισης καθώς επίσης και ορισμένες μεθοδολογίες ευθυγράμμισης με βαθιά μάθηση.

#### 3.1 Κλασσικές Μέθοδοι Ευθυγράμμισης

Σε αυτήν την παράγραφο θα δούμε με ποιον τρόπο γίνεται η ευθυγράμμιση των εικόνων με κλασσικές μεθόδους βελτιστοποίησης. Προφανώς λόγω του όγκου των δεδομένων που παράγονται καθημερινά, η επιστημονική κοινότητα έχει εγκαταλείψει τις μεθόδους αυτές επειδή είναι εξαιρετικά αργές. Παρόλα αυτά, μελετώντας κανείς αυτές τις παλαιότερες μεθόδους μπορεί κανείς να εκτιμήσει το άλμα στην επίδοση σε προβλήματα ευθυγράμμισης που έφερε η βαθιά μάθηση. Για τον λόγο αυτόν θα κάνουμε μία σύντομη αναφορά εδώ. Ο πλήρης μαθηματικός φορμαλισμός του προβλήματος που περιγράφεται παρακάτω μπορεί να βρεθεί στο [10].

Με μαθηματικούς όρους, ο στόχος της ευθυγράμμισης είναι ο προσδιορισμός μίας συνάρτησης  $\phi : \Omega \rightarrow \Omega$ , όπου  $\Omega \subset \mathbb{R}^n$ .  $\Omega$  είναι το πεδίο στο οποίο ορίζονται τα δεδομένα μας. Αν  $n = 2$  τότε έχουμε διδιάστατα δεδομένα, ενώ αν  $n = 3$  τριδιάστατα. Τα δεδομένα μας, δηλαδή οι εικόνες, μπορούν να θεωρηθούν ως συναρτήσεις  $I : \Omega \rightarrow \mathbb{R}^d$ . Ο αριθμός  $d$  είναι και ο αριθμός των καναλιών της εικόνας. Έστω  $I_i, I_f$  η εικόνα πρότυπο και η εικόνα στόχος αντίστοιχα. Θέλουμε ιδανικά ο μετασχηματισμός  $\phi$  να απεικονίζει με το μικρότερο δυνατό σφάλμα την εικόνα  $I_i$  στην εικόνα  $I_f$ . Ο μετασχηματισμός της εικόνας  $I_i$  υπό την  $\phi$  ορίζεται ως  $I_i \circ \phi = I_i(\phi^{-1})$ . Ο τελεστής ο δηλώνει την παραμόρφωση (warping) της εικόνας  $I_i$  υπό την επίδραση του πεδίου  $\phi$ . Με απλά λόγια αυτό σημαίνει ότι η  $\phi$  περιγράφει με ποιον τρόπο πρέπει να αλλάξει το κάθε εικονοκύτταρο (pixel) της εικόνας  $I_i$ . Ο συμβολισμός  $\phi^{-1}$  υπονοεί προφανώς ότι πρέπει να ορίζεται το αντίστροφο πεδίο της  $\phi$ . Στην συνέχεια θα δούμε ότι για να διατυπώσουμε ένα καλώς-ορισμένο πρόβλημα βελτιστοποίησης θα πρέπει να υποθέσουμε κάτι ακόμα πιο ισχυρό από την ύπαρξη του αντίστροφου πεδίου, δηλαδή η  $\phi$  να είναι ένας διαφορομορφισμός.

Ο μετασχηματισμός  $\phi$  γραμμικοποιείται γύρω από το σύστημα συντεταγμένων της εικόνας προτύπου και παράγεται από ένα διανυσματικό πεδίο μετατόπισης (displacement vector

field)  $u : \Omega \rightarrow \mathbb{R}^n$ , τέτοιο ώστε  $\phi(x) = x + u(x)$ ,  $\forall x \in \Omega$ . Προφανώς προσδιορίζοντας το διανυσματικό πεδίο  $u$  μπορούμε άμεσα να προσδιορίσουμε και τον μετασχηματισμό  $\phi$ . Με βάση αυτά έχουμε ότι η εικόνα πρότυπο  $I_i$  μετασχηματίζεται στην  $I_i(x - u(x))$ . Προφανώς χρειαζόμαστε και μία συνάρτηση κόστους με βάση την οποία θα αξιολογήσουμε τον μετασχηματισμό αυτόν. Η συνάρτηση κόστους θα έχει την εξής μορφή:

$$C_2(I_i, I_f, \phi) = \|I_i \circ \phi^{-1} - I_f\|_2^2$$

όπου  $\|\cdot\|_2$  είναι η γνωστή  $L^2$  νόρμα των τετραγωνικά ολοκληρώσιμων συναρτήσεων, δηλαδή  $\|f\|_2^2 = \int_{\Omega} |f(x)|^2 dx$ .

Ένας βέλτιστος μετασχηματισμός είναι προφανώς εκείνος που ελαχιστοποιεί την συνάρτηση κόστους, δηλαδή ένας που μας στέλνει σε ένα από τα τοπικά ελάχιστα της. Το ολικό βέλτιστο είναι φυσικά εξαιρετικά δύσκολο να επιτευχθεί. Θα θέλαμε επίσης ο βέλτιστος μετασχηματισμός να έχει μία επιπλέον ιδιότητα, να είναι κατά το δυνατόν πιο λείος. Το πρόβλημα είναι τώρα να ποσοτικοποιήσουμε την ομαλότητα μίας συνάρτησης. Η ομαλότητα του διανυσματικού πεδίου μετατόπισης ορίζεται ως:

$$C_1(u) = \|Lu\|_2^2$$

όπου ο  $L$  είναι ένας διαφορικός τελεστής της μορφής  $L = (-\alpha\Delta + \gamma)I_{n \times n}$ . Πρόκειται δηλαδή για έναν όρο ομαλοποίησης (regularization). Συνεπώς η εύρεση του βέλτιστου διανυσματικού πεδίου  $u$  έγκειται στην επίλυση του προβλήματος βελτιστοποίησης:

$$u^* = \arg \min_u \frac{1}{\sigma^2} \|I_i \circ \phi^{-1} - I_f\|_2^2 + \|Lu\|_2^2 \quad (3.1.1)$$

Το παραπάνω πρόβλημα μπορεί να θεωρηθεί ως ένα πρόβλημα λογισμού μεταβολών και άρα μπορεί να επιλυθεί με την χρήση των εξισώσεων *Euler-Lagrange*. Για να καταλήξουμε όμως εκεί θα κάνουμε ένα βήμα παραπέρα στους ορισμούς μας έτσι ώστε να φροντίσουμε ο μετασχηματισμός  $\phi$  να είναι ένας διαφορομορφισμός, δηλαδή τόσο η  $\phi$  όσο και η  $\phi^{-1}$  να είναι συνεχώς παραγωγίσιμες συναρτήσεις.

Για τον σκοπό αυτόν θεωρούμε την μονοπαραμετρική οικογένεια μετασχηματισμών  $\Phi_t : \Omega \times [0, 1] \rightarrow \Omega$ , τέτοια ώστε  $\Phi_0 = Id_{\Omega}$ , καθώς επίσης και το χρονικά εξαρτημένο διανυσματικό  $u_t : \Omega \times [0, 1] \rightarrow \mathbb{R}^n$  που ορίζεται από την συνήθη διαφορική εξίσωση  $\dot{\Phi}_t = u_t(\Phi_t)$ . Σε αυτό το πλαίσιο, ο επιθυμητός μετασχηματισμός,  $\phi$ , παράγεται από την μονοπαραμετρική οικογένεια  $\Phi_t$  ως  $\phi = \Phi_1 = \Phi_0 + \int_0^1 u_t(\Phi_t) dt$ . Με βάση τους παραπάνω ορισμούς, το πρόβλημα λογισμού μεταβολών είναι τώρα το εξής:

$$u^* = \arg \min_{u: \Phi_t = u_t(\Phi_t)} \frac{1}{\sigma^2} \|I_i \circ \Phi_1^{-1} - I_f\|_2^2 + \int_0^1 \|u_t\|_V^2 dt \quad (3.1.2)$$

όπου  $V$  είναι ο συναρτησιακός χώρος στον οποίο ανήκουν τα μέλη της οικογένειας  $u_t$  και  $\sigma > 0$ . Με βάση τα παραπάνω παίρνουμε το παρακάτω θεώρημα [?]:

**Θεώρημα:** Το συναρτησοειδές ενέργειας:

$$E(u) = \int_0^1 \|u_t\|_V^2 dt + \frac{1}{\sigma^2} \|I_i \circ \Phi_{1,0}^u - I_f\|_2^2$$

ικανοποιεί την παρακάτω εξίσωση Euler-Lagrange:

$$2u_t - K\left(\frac{2}{\sigma^2} |D(\Phi_{t,1}^u)| \nabla J_t^0 (J_t^0 - J_t^1)\right) = 0$$

όπου  $\Phi_{s,t} = \Phi_t \circ \Phi_s^{-1}$ ,  $D\Phi_{s,t}$  ο Ιακωβιανός πίνακας,  $J_t^0 = I_i \circ \Phi_{t,0}$ ,  $J_t^1 = I_f \circ \Phi_{t,1}$  και  $K : L^2(\Omega, \mathbb{R}^d) \rightarrow V$  ένας συμπαγής αυτοσυζυγής τελεστής που ορίζεται από την σχέση  $\langle a, b \rangle_{L^2} = \langle Ka, b \rangle_V$ . Ο εκθέτης  $u$  αναφέρεται στο γεγονός ότι το διανυσματικό πεδίο  $\Phi$  είναι ρητά συνδεδεμένο με την συνάρτηση  $u$ .

Το θεώρημα αυτό μας λέει ότι προκειμένου να βρούμε το πεδίο  $u$ , θα πρέπει να λύσουμε την παραπάνω εξίσωση Euler-Lagrange. Αν δεν είναι προφανές από την έκφραση της εξίσωσης αυτό είναι εξαιρετικά δύσκολο και αργό. Αν σκεφτούμε ότι μία εικόνα μπορεί να είναι μία συνάρτηση χιλιάδων μεταβλητών, το να βρούμε ένα πεδίο  $u$  που να ικανοποιεί την παραπάνω εξίσωση δεν είναι καθόλου τετριμμένη υπόθεση. Μία επιπλέον δυσκολία είναι η επιλογή της παραμέτρου  $\sigma$  καθώς και του τελεστή  $K$ . Στην πράξη η εύρεση μίας αναλυτικής λύσης είναι αδύνατη και άρα καλούμαστε να χρησιμοποιήσουμε κάποια αριθμητική μέθοδο. Οσοδήποτε γρήγορη ή μαθηματικά εντυπωσιακή είναι μία αριθμητική μέθοδος, πάσχει από το πρόβλημα της μη-γενίκευσης. Αυτό σημαίνει ότι για κάθε νέο ζεύγος εικόνων  $I_i, I_f$  θα πρέπει να επιλυθεί ένα πολύ δύσκολο αριθμητικό πρόβλημα. Ίδανικά θα θέλαμε να δημιουργήσουμε ένα μοντέλο μηχανικής μάθησης, το οποίο θα πρέπει αρχικά φυσικά να εκπαιδευτεί σε ένα σύνολο εικόνων, αλλά στην συνέχεια θα μπορεί να ευθυγραμμίσει κατά το δυνατόν βέλτιστα οποιοδήποτε ζεύγος εικόνων που προέρχονται από παρόμοια κατανομή. Σε αυτό το σημείο μπαίνει στο παιχνίδι η βαθιά μάθηση.

### 3.2 Ευθυγράμμιση με Βαθιά Μάθηση

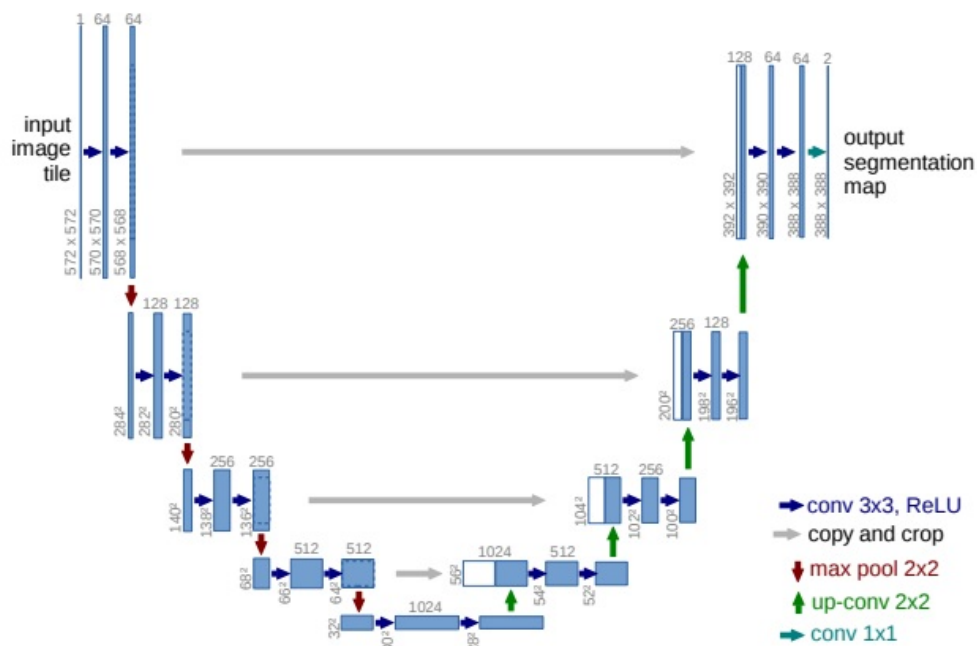
Το πιο φημισμένο μοντέλο ευθυγράμμισης ιατρικών εικόνων που κυκλοφορεί πλέον στην επιστημονική κοινότητα είναι το μοντέλο *VoxelMorph* [12]. Η δομή και η λειτουργία του *VoxelMorph* είναι αρκετά περίπλοκη. Ο "σκελετός" του, δηλαδή η υποκείμενη δομή νευρωνικού δικτύου που περιέχει είναι αυτή των συνελικτικών νευρωνικών δικτύων. Συνήθως, το *VoxelMorph* αποτελείται από ένα συνελικτικό νευρωνικό δίκτυο γνωστό στην διεθνή

βιβλιογραφία ως μοντέλο *U-Net* [11]. Συνεπώς για να καταλάβουμε την λειτουργία του *VoxelMorph*, θα πρέπει πρώτα να μελετήσουμε το *U-Net*. Σημειώνεται εδώ ότι η χρήση του *U-Net* δεν είναι απαραίτητη. Η μόνη απαίτηση που υπάρχει από το νευρωνικό δίκτυο, όπως θα δούμε παρακάτω, είναι η έξοδος του δικτύου να έχει ίδιες διαστάσεις με την εικόνα που θέλουμε να ευθυγραμμίσουμε.

Θα ξεκινήσουμε με μία σύνοψη του *U-Net*, ακολουθούμενη από την περιγραφή του μοντέλου *VoxelMorph*.

### 3.2.1 Το μοντέλο U-Net

Το *U-Net* αποτελεί στην ουσία μία αλληλουχία από συνελίξεις (convolution) και αποσυνελίξεις (deconvolution ή up-sampling ή transpose-convolution), μαζί με κάποιες ενδιάμεσες ενεργοποιήσεις και ενέργειες που θα περιγραφούν παρακάτω. Είναι δηλαδή ένα συνελικτικό νευρωνικό δίκτυο. Χρησιμοποιείται κυρίως για προβλήματα κατάτμησης (segmentation) αλλά τα τελευταία χρόνια έχει αποτελέσει τον βασικό δομικό λίθο για πιο περίπλοκα μοντέλα που αντιμετωπίζουν διάφορα άλλα είδη προβλημάτων, όπως η ευθυγράμμιση εικόνων. Στην παρακάτω εικόνα φαίνεται η σχηματική αναπαράσταση της βασικής δομής του *U-Net*:



Εικόνα 5: Σχηματική αναπαράσταση του μοντέλου *U-Net*

Όπως φαίνεται και στην εικόνα 5 τα μπλε βελόνια δηλώνουν την πράξη της συνελίξης με ενεργοποίηση ReLU, τα κόκκινα βελόνια την πράξη MaxPooling, τα γκρι βελόνια την πράξη

της συνένωσης (concatenation) και τα πράσινα βελόκια την πράξη της αποσυνέλιξης. Αυτό που κάνει δηλαδή το U-Net είναι να συμπίεσει την είσοδο-εικόνα μέσω διαδοχικών συνελίξεων σε μία εικόνα αρκετά μικρότερων διαστάσεων και στην συνέχεια την ανακατασκευάζει, αυξάνοντας και πάλι τις διαστάσεις, μέσω διαδοχικών αποσυνελίξεων. Η έξοδος του δικτύου είναι συνήθως μία εικόνα ίδιων διαστάσεων (πλάτος, μήκος και ύψος) με την είσοδο αν και μπορεί να είναι και μικρότερη. Το U-Net είναι δηλαδή μία ειδική περίπτωση αυτοκωδικοποιητή (autoencoder).

Τα περισσότερα συστατικά στοιχεία του U-Net που φαίνονται στην εικόνα 5 είναι πολύ γνωστά στην επιστημονική κοινότητα που ασχολείται με την βαθιά μάθηση. Μία πολύ καλή και αναλυτική περιγραφή των συνελικτικών νευρωνικών δικτύων μπορεί να βρεθεί, για παράδειγμα στα [13], [14]. Αυτό που θα περιγράψουμε εδώ είναι η πράξη της αποσυνέλιξης που μπορεί να μην είναι οικεία σε έναν όλοιο. Η αποσυνέλιξη είναι, όπως υποδηλώνει και το όνομα, η αντίστροφη πράξη της συνέλιξης<sup>1</sup>. Στο σημείο αυτό θα κάνουμε μία διάκριση ανάλογα με το επιστημονικό πεδίο εφαρμογής.

Στα πλαίσια των σημάτων και συστημάτων, δεδομένων πινάκων-εικόνων  $g$  και  $h$  (και ίσως και ενός πίνακα-θορύβου  $\epsilon$ ), στόχος της αποσυνέλιξης είναι να βρούμε έναν πίνακα-εικόνα που να ικανοποιεί την εξίσωση:

$$f * g = h + \epsilon$$

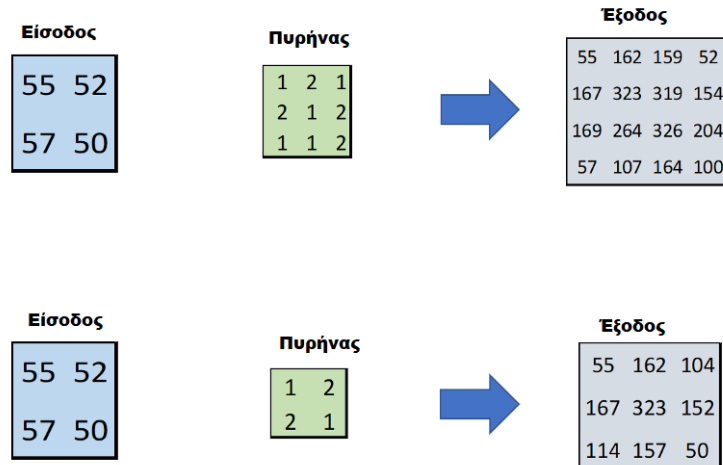
Η εξίσωση αυτή τις περισσότερες φορές δεν έχει μοναδική λύση ακόμα και απουσία θορύβου και ο τρόπος με τον οποίο λύνεται είναι με την μέθοδο της αποσυνέλιξης κατά Wiener (Wiener Deconvolution) [15]. Η μεθοδολογία στην ουσία περιλαμβάνει τον μετασχηματισμό Fourier της παραπάνω εξίσωσης, αλγεβρική λύση στο πεδίο των συχνοτήτων και τέλος τον αντίστροφο μετασχηματισμό Fourier της λύσης.

Στα πλαίσια της μηχανικής μάθησης και άρα της βαθιάς μάθησης, η αποσυνέλιξη έχει διαφορετικό χαρακτήρα. Όπως είδαμε και στην εικόνα 5 στόχος μας είναι να εφεύρουμε μία πράξη η οποία να παίρνει ως είσοδο έναν πίνακα και με κάποιον συστηματικό τρόπο να αυξάνει τις διαστάσεις του. Θέλουμε επίσης αυτή η πράξη να περιλαμβάνει παραμέτρους, τις οποίες προφανώς θα πρέπει να μάθουμε κατά την διάρκεια της εκπαίδευσης του δικτύου. Ομοίως με τα συνελικτικά στρώματα, έτσι και εδώ θα έχουμε πίνακες-πυρήνες (kernels) που θα δρουν στην είσοδο και, με τον τρόπο που θα ορίσουμε, θα παράγεται μία έξοδος μεγαλύτερων διαστάσεων. Σχηματικά, η δράση της αποσυνέλιξης φαίνεται στην παρακάτω εικόνα:

Θα περιγράψουμε τώρα πώς, δεδομένης της εισόδου και του πυρήνα, παράγεται η έξοδος. Θα συγκεντρωθούμε στο κάτω παράδειγμα της εικόνας 6 χωρίς να έχει σημασία μιας και

<sup>1</sup>Υπενθυμίζουμε εν συντομία ότι η συνέλιξη μεταξύ δύο ακολουθιών  $\{x_n\}_{n=1}^m$  και  $\{y_n\}_{n=1}^m$  είναι η ακολουθία με τύπο  $(x * y)_n = \sum_{k=1}^m y_{n-k}x_k$





Εικόνα 6: Σχηματική αναπαράσταση της πράξης της αποσυνέλιξης (1)

τα ίδια ισχύουν και για το πάνω. Ξεκινάμε με το στοιχείο  $A_{00} = 55$  της εισόδου. Πολλαπλασιάζουμε τον πυρήνα με το στοιχείο αυτό και παίρνουμε έναν νέο  $2 \times 2$  πίνακα, τον οποίο "τοποθετούμε" στην πάνω αριστερή γωνία ενός κενού  $3 \times 3$  πίνακα (το πορτοκαλί περίγραμμα της εικόνας 7). Κάνουμε το ίδιο με το στοιχείο  $A_{01} = 52$  της εισόδου και τον νέο πίνακα τον τοποθετούμε στην πάνω δεξιά γωνία του κενού πίνακα (Αριστερό περίγραμμα της εικόνας 7). Παρόλα αυτά τώρα θα υπάρχει επικάλυψη με τον προηγούμενο πίνακα που τοποθετήσαμε. Πολύ απλά, σε περίπτωση επικαλύψεων, αθροίζουμε τους εκάστοτε αριθμούς. Κάνουμε το ίδιο για όλα τα στοιχεία  $A_{ij}$  της εισόδου.



Εικόνα 7: Σχηματική αναπαράσταση της πράξης της αποσυνέλιξης (2)

Στην παραπάνω διαδικασία υποθέσαμε σιωπηλά ότι το βήμα (stride) του πυρήνα πάνω στην εικόνα εισόδου είναι ίσο με 1, κάτι που στην γενική περίπτωση δεν είναι ανάγκη να ισχύει. Στο παράδειγμά μας, αν η εικόνα εισόδου είχε διαστάσεις, για παράδειγμα,  $3 \times 3$  το βήμα του πυρήνα θα μπορούσε προφανώς να είναι 2 (και στις δύο διαστάσεις). Τότε όμως η έξοδός μας δεν θα είχε μεγαλύτερες διαστάσεις, αλλά θα ήταν μία εικόνα  $2 \times 2$ . Γενικά, αν η είσοδος έχει διαστάσεις  $(C_{in}, H_{in}, W_{in})$  όπου  $C_{in}$  ο αριθμός των καναλιών, τότε η έξοδος θα έχει διαστάσεις  $(C_{out}, H_{out}, W_{out})$ , όπου [16]:

$$H_{out} = (H_{in} - 1) \times Stride[0] - 2 \times Padding[0] + Dilation[0] \times (Kernel\_Size[0] - 1) +$$

$$Output\_Padding[0] + 1$$

$$W_{out} = (W_{in} - 1) \times Stride[1] - 2 \times Padding[1] + Dilation[1] \times (Kernel\_Size[1] - 1) + Output\_Padding[1] + 1$$

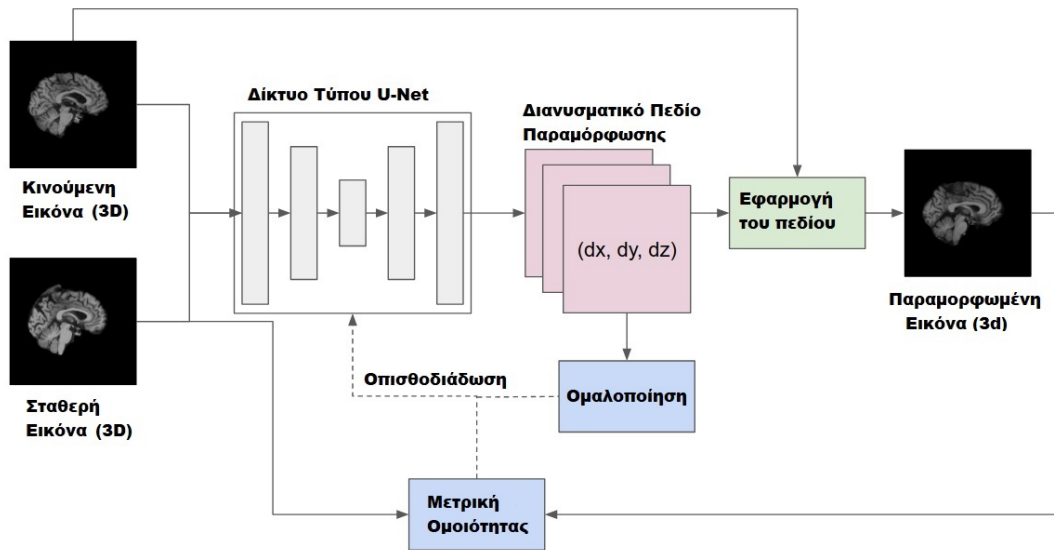
$$C_{out} = C_{in}$$

Οι ενέργειες Padding, Dilation και Outer Padding είναι οι ίδιες που συναντάμε στα κλασικά συνελικτικά νευρωνικά δίκτυα και παραπέμπουμε στην γνωστή βιβλιογραφία ([13], [14]). Γίνεται αμέσως κατανοητό ότι, προκειμένου η έξοδός μας να έχει τις επιθυμητές διαστάσεις, θα πρέπει να επιλέξουμε κατάλληλο μέγεθος πυρήνα, βήματος καθώς και των υπόλοιπων πράξεων, βάσει πάντα των παραπάνω τύπων.

Έχοντας τώρα καταλάβει με ποιον τρόπο μεταδίδεται η πληροφορία από την είσοδο προς την έξοδο του U-Net, είμαστε έτοιμοι να δούμε πώς αυτό μπορεί να χρησιμοποιηθεί σε ένα γενικότερο μοντέλο που τελικά προορίζεται να λύσει το πρόβλημα της ευθυγράμμισης των εικόνων.

### 3.2.2 Το μοντέλο VoxelMorph

Το VoxelMorph [12] είναι ένα μοντέλο-πλαίσιο βαθιάς μάθησης γενικού σκοπού που τα τελευταία χρόνια χρησιμοποιείται κατά κόρον για ευθυγράμμιση ιατρικών εικόνων, τόσο διδιάστατων όσο και τρισδιάστατων, με ένα ή και περισσότερα κανάλια. Η βασική νευρωνική δομή του είναι αυτή του U-Net. Περιληπτικά, το VoxelMorph παίρνει ως είσοδο την συνένωση (concatenation) δύο εικόνων ίδιων διαστάσεων, μίας *κινούμενης εικόνας* (moving image),  $I_m$ , και μίας *σταθερής εικόνας* (fixed image),  $I_f$  και στην έξοδο παράγει έναν πίνακα που έχει την ερμηνεία του πεδίου *παραμόρφωσης* (deformation field),  $\phi$ , του οποίου οι διαστάσεις είναι ίδιες με την κινούμενη (ή την σταθερή) εικόνα. Στην συνέχεια το πεδίο αυτό εφαρμόζεται πάνω στην κινούμενη εικόνα, μέσω μίας διαδικασίας παρεμβολής (interpolation) και τελικά παράγεται μία νέα, *παραμορφωμένη εικόνα* (deformed image ή moved images). Η παραμορφωμένη αυτή εικόνα συγκρίνεται, μέσω κάποιας συνάρτησης ομοιότητας,  $\mathcal{L}_{sim}$ , με την αρχική σταθερή εικόνα. Στόχος της εκπαίδευσης του μοντέλου είναι προφανώς να μάθει να παράγει ένα βέλτιστο πεδίο παραμόρφωσης, έτσι ώστε η ομοιότητα μεταξύ της παραμορφωμένης εικόνας και της σταθερής εικόνας να είναι μέγιστη. Αυτή είναι συνοπτικά η βασική λειτουργία του VoxelMorph. Σε αυτό το μοντέλο έχουν βασιστεί πάρα πολλές μελλοντικές (σε σχέση με το VoxelMorph) δημοσιεύσεις που σχετίζονται με ευθυγράμμιση εικόνων (ενδεικτικά παραπέμπουμε στις εργασίες [17], [18]). Πριν προχωρήσουμε στις τεχνικές λεπτομέρειες της υλοποίησης του VoxelMorph, παραθέτουμε μία εικόνα που δείχνει την σχηματική αναπαράσταση αυτών που περιγράφηκαν μέχρι τώρα.



Εικόνα 8: Σχηματική αναπαράσταση του VoxelMorph [31]

Βλέπουμε ότι η πρόσθια μετάδοση της πληροφορίας περιλαμβάνει την συνένωση της κινούμενης και της σταθερής εικόνας, πέρασμα μέσα από το νευρωνικό δίκτυο προς παραγωγή του πεδίου παραμόρφωσης  $\phi$  και τέλος εφαρμογή του πεδίου παραμόρφωσης στην αρχική κινούμενη εικόνα. Η οπισθοδιάδωση αναφέρεται προφανώς στην ανανέωση των βαρών του δικτύου μέσω της ελαχιστοποίησης μίας συνάρτησης κόστους. Η συνάρτηση αυτή περιλαμβάνει την μετρική ομοιότητας,  $\mathcal{L}_{sim}$ , (ή ποιο σωστά την ανομοιότητα,  $1 - \mathcal{L}_{sim}$  ή  $-\mathcal{L}_{sim}$ , ανάλογα με την περίπτωση) καθώς επίσης και έναν όρο ομαλοποίησης,  $\mathcal{L}_{smooth}$ , που φροντίζει το παραγόμενο πεδίο παραμόρφωσης να είναι κατά το δυνατόν λείο (smooth).

Κατά τα γνωστά, θεωρούμε την κινούμενη και την σταθερή εικόνα ως συναρτήσεις  $I_m, I_f : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}^d$  (συνήθως ότι δουλεύουμε με ιατρικές εικόνες έχουμε ένα κανάλι, δηλαδή  $d = 1$ ). Όπως και στην περίπτωση της κλασσικής ευθυγράμμισης, το νευρωνικό δίκτυο δεν θα παράγει άμεσα το διανυσματικό πεδίο  $\phi$ , αλλά ένα δευτερεύον πεδίο  $u$ , που θα αποτελεί την γραμμικοποίηση του επιθυμητού πεδίου, δηλαδή  $\phi(x) = x + u(x)$ . Αν  $\theta$  είναι το διάνυσμα βαρών του νευρωνικού δικτύου τότε χρησιμοποιούμε τον συμβολισμό  $V_\theta(I_m, I_f) = u$ . Αυτή είναι η μοντελοποίηση του πεδίου  $u$ . Η ερμηνεία που δίνουμε στο  $u$  είναι ότι για κάθε εικονοκύτταρο  $p \in \Omega$ , το  $u(p)$  δηλώνει μία μετατόπιση, τέτοια ώστε το  $f(p)$  και το  $(I_m \circ \phi)(p)$  να πέφτουν σε κατά το δυνατόν κοντινότερες ανατομικές περιοχές. Αυτή είναι και η έννοια της ευθυγράμμισης, να κουνήσουμε τα εικονοκύτταρα της κινούμενης εικόνας με κατάλληλο τρόπο έτσι ώστε να πέσουν πάνω στα εικονοκύτταρα της σταθερής εικόνας.

Με βάση τα παραπάνω καταλαβαίνουμε ότι ο στόχος μας είναι η ελαχιστοποίηση της "διαφοράς" μεταξύ των εικόνων  $I_f$  και  $I_m \circ \phi$ . Αυτό προφανώς θα γίνει στο στάδιο της

οπισθοδιάδωσης με την εφαρμογή κάποιας μεθόδου ελαχιστοποίησης κατάλληλης συνάρτησης κόστους. Συνεπώς, επειδή θα πρέπει να υπολογίσουμε παραγώγους, θα πρέπει να κατασκευάσουμε έναν διαφορικό τελεστή για να σχηματίσουμε την εικόνα  $I_m \circ \phi$ . Αυτό γίνεται με τον εξής τρόπο:

Ορίζουμε το εικονοκύτταρο,  $(I_m \circ \phi)(p)$  να ισούται με την γραμμική παρεμβολή των γειτονικών εικονοκυττάρων που δίνεται από τον τύπο:

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

όπου  $p' = p + u(p)$ ,  $\mathcal{Z}(p')$  είναι η γειτονιά του εικονοκυττάρου  $p'$  και το γράμμα  $d$  αναφέρεται προφανώς στις τρεις χωρικές διαστάσεις  $x,y,z$ . Ο παραπάνω τύπος εξυπηρετεί δύο σκοπούς: αφενός μας δίνει έναν διαφορίσιμο τύπο για τον υπολογισμό των απαιτούμενων παραγώγων και αφετέρου φροντίζει έτσι ώστε οι τιμές  $p'$  να είναι ακέραιες (μιας και αντιστοιχούν σε θέσεις των εικονοκυττάρων).

Στην επόμενη υποενότητα θα περιγράψουμε κάποιες τεχνικές λεπτομέρειες της εκπαίδευσης που περιλαμβάνουν τις πιθανές συναρτήσεις κόστους. Η συνάρτηση κόστους είναι προφανώς μία πολύ σημαντική σχεδιαστική παράμετρος και θα πρέπει να αντικατοπτρίζει τις προδιαγραφές που θέλουμε να έχει το μοντέλο μας.

### 3.2.3 Συναρτήσεις Κόστους

Κάθε φορά που θα αναφερόμαστε στην συνάρτηση κόστους θα εννοούμε το άθροισμα της συνάρτησης ανομοιότητας και του όρου ομαλοποίησης. Θα μελετήσουμε τους κάθε όρους ξεχωριστά.

1. **Συναρτήσεις Ανομοιότητας:** Οι συναρτήσεις αυτές μας βοηθούν να ποσοτικοποιήσουμε την διαφορά μεταξύ της σταθερής εικόνας και της παραμορφωμένης εικόνας.

- Μέσο Τετραγωνικό Σφάλμα (Mean Square Error): Πρόκειται για το γνωστό σε όλους μας σφάλμα που δίνεται από τον τύπο:

$$MSE(I_f, I_m \circ \phi) = \frac{1}{\Omega} \sum_{p \in \Omega} [I_f(p) - (I_m \circ \phi)(p)]^2$$

Πρόκειται δηλαδή για την μέση διαφορά της έντασης του κάθε εικονοκυττάρου της σταθερής και της παραμορφωμένης εικόνας. Είναι μία καλή αρχική επιλογή για το σχεδιασμό πειραμάτων αλλά χαρακτηρίζεται από χαμηλή ευρωστία, όπως έχουν δείξει διάφορες μελέτες.

- Κανονικοποιημένη Διασταυρούμενη Συσχέτιση (Normalized Cross Correlation): Για την συνάρτηση αυτή συμβολίζουμε με  $I_{\hat{f}}(p)$  και  $(I_{\hat{m}\circ\phi})(p)$  την μέση τοπική ένταση:  $I_{\hat{f}}(p) = \frac{1}{n^3} \sum_{p_i} I_f(p_i)$  και  $(I_{\hat{m}\circ\phi})(p) = \frac{1}{n^3} \sum_{p_i} [I_{m\circ\phi}](p_i)$ , όπου τα εικονοκύτταρα  $p_i$  διατρέχουν την γειτονιά όγκου  $n^3$  του  $p$ . Με βάση τα παραπάνω η Κανονικοποιημένη Διασταυρούμενη Συσχέτιση ορίζεται ως:

$$CC(I_f, I_m \circ \phi) = \frac{\left( \sum_{p_i} (I_f(p_i) - I_{\hat{f}}(p))(I_{m\circ\phi}(p_i) - [I_{\hat{m}\circ\phi}](p)) \right)^2}{\left( \sum_{p_i} (I_f(p_i) - I_{\hat{f}}(p))^2 \right) \left( \sum_{p_i} I_{m\circ\phi}(p_i) - [I_{\hat{m}\circ\phi}](p) \right)^2}$$

Από τον παραπάνω τύπο παρατηρούμε ότι μεγαλύτερες τιμές του δείκτη CC υποδηλώνει καλύτερη ευθυγράμμιση, αφού οι μέσες εντάσεις τείνουν να γίνουν ίσες παντού. Συνεπώς η αντίστοιχη συνάρτηση ανομοιότητας μπορεί να ορισθεί ως η -CC.

- Δείκτης Δομικής Ομοιότητας (Structural Similarity Index): Ο δείκτης αυτός υπολογίζεται σε διάφορα παράθυρα ίδιου μεγέθους  $N \times N$  της εικόνας. Αν  $x, y$  είναι δύο τέτοια παράθυρα, τότε:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

όπου  $\mu_x(\mu_y)$  η μέση τιμή των εικονοκυττάρων του  $x$  ( $y$ ),  $\sigma_x^2(\sigma_y^2)$  η διακύμανση του  $x$  ( $y$ ),  $\sigma_{xy}$  η συνδιακύμανση των  $x$  και  $y$ ,  $c_1 = (k_1L)^2$  και  $c_2 = (k_2L)^2$ . Οι  $c_1$  και  $c_2$  είναι σταθερές, ο ρόλος των οποίων είναι να σταθεροποιήσουν την διαίρεση με τον παρονομαστή. Οι τιμές των  $k_1$  και  $k_2$  θέτονται συνήθως ως 0.01 και 0.03 αντίστοιχα. Το  $L$  ονομάζεται δυναμικό εύρος των εικονοκυττάρων και τίθεται ίσο με  $2^{\#bitsperpixel} - 1$ . Ο δείκτης SSIM κυμαίνεται από το -1 (καμία ομοιότητα) έως το 1 (πλήρης ομοιότητα). Η αντίστοιχη συνάρτηση ανομοιότητα ορίζεται ως  $1 - SSIM$ .

- Συντελεστής Συσχέτισης Pearson (Pearson's Correlation Coefficient): Συνήθως ο δείκτης αυτό χρησιμοποιείται σε προβλήματα γραμμικής παλινδρόμησης αλλά μπορεί να ερμηνεύσει και την απόκλιση μεταξύ των εντάσεων δύο εικόνων. Ορίζεται, για δύο εικόνες,  $X, Y$ , με ένα κανάλι, ως [19]:

$$PCC(X, Y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

όπου  $\bar{x}(\bar{y})$  η μέση ένταση των εικονοκυττάρων της εικόνας X (Y) και φυσικά  $x_i(y_i)$  η τιμή της έντασης του εικονοκυττάρου της εικόνας X(Y) στην θέση  $i$ . Μπορούμε να φανταστούμε ότι, για τον υπολογισμό του παραπάνω τύπου, οι εικόνες μας που έχουν διάσταση  $N \times N$  έχουν μετατραπεί σε διανύσματα μήκους  $N^2 = n$ . Ο δείκτης Pearson κυμαίνεται από το -1 (αρνητική συσχέτιση) έως το 1 (θετική συσχέτιση). Η αντίστοιχη συνάρτηση ανομοιότητας ορίζεται ως  $1 - PCC$ .

- Σταθμισμένοι Δείκτες PPC και SSIM (Weighted PCC + SSIM): Όπως υποδηλώνει και το όνομα, μπορούμε να σχηματίσουμε μία συνάρτηση ανομοιότητας που να αποτελείται από ένα σταθμισμένο άθροισμα των δεικτών PCC και SSIM. Με βάση όσα έχουν διατυπωθεί παραπάνω, η αντίστοιχη συνάρτηση ανομοιότητας θα είναι η:

$$\mathcal{W}(I_f, I_{mof}) = \alpha(1 - PCC(I_f, I_{mof})) + \beta(1 - SSIM(I_f, I_{mof}))$$

όπου  $\alpha, \beta$  θετικές σταθερές που αποτελούν την συνεισφορά των αντίστοιχων όρων που πολλαπλασιάζουν.

- Αμοιβαία Πληροφορία (Mutual Information): Αν A, B είναι δύο εικόνες, τότε η δεσμευμένη εντροπία  $H(A, B)$  ορίζεται ως  $H(A, B) = - \sum_{a,b} p_{AB}(a, b) \log p_{AB}(a, b)$ ,

όπου  $p_{AB}(a, b) = \frac{1}{\Omega} \sum_{p \in \Omega} \delta(a - A(p)) \delta(b - B(p))$  η δεσμευμένη κατανομή πιθανότητας των εικόνων. Η εντροπία μιας εικόνας ορίζεται προφανώς ως  $H(A) = - \sum_x p_X(x) \log p_X(x)$ . Τότε, η αμοιβαία πληροφορία μεταξύ των εικόνων A και B ορίζεται ως [20]:

$$MI(A, B) = H(A) + H(B) - H(A, B)$$

Από τον παραπάνω τύπο καταλαβαίνουμε ότι το να μεγιστοποιήσουμε την αμοιβαία πληροφορία μεταξύ δύο εικόνων οδηγεί στο να εντοπίσουμε περιοχές των εικόνων που αλληλοεπικαλύπτονται. Χρησιμοποιείται ευρέως σε εφαρμογές που περιλαμβάνουν ιατρικές εικόνες και είναι αρκετά εύρωστος δείκτης. Παρόλα αυτά μπορεί κανείς να παρατηρήσει αμέσως ένα πρόβλημα με την παραπάνω συνάρτηση. Η δεσμευμένη πιθανότητα περιλαμβάνει τον πολλαπλασιασμό συναρτήσεων  $\delta$  του

Dirac, που δεν είναι παραγωγίσιμες κι άρα μπορεί να προκαλέσουν προβλήματα στην εκπαίδευση, στην φάση της οπισθοδιάδωσης. Προκειμένου να λύσουμε αυτό το πρόβλημα προσεγγίζουμε την πιθανότητα αυτή με την μέθοδο των παραθύρων Parzen [24]. Γράφουμε δηλαδή:

$$p_{AB}(a, b) = \alpha \sum_{p \in \Omega} \psi(a - A(p))\chi(b - B(p))$$

όπου  $\alpha$  σταθερά κανονικοποίησης, τέτοια ώστε  $\sum_{a,b} p_{AB}(a, b) = 1$ . Οι συναρτήσεις  $\psi$  και  $\chi$  ονομάζονται συναρτήσεις πυρήνες των παραθύρων Parzen. Οι πιο συνήθεις πυρήνες είναι Γκαουσιανοί. Με βάση αυτά μπορούμε να γράψουμε την εξής προσέγγιση της δεσμευμένης πιθανότητας:

$$p_{AB}(a, b) = \alpha \sum_{p \in \Omega} \frac{1}{\Delta w_A \Delta w_B \pi} \exp\left\{-\frac{(a - A(p))^2}{2\Delta w_A^2}\right\} \exp\left\{-\frac{(b - B(p))^2}{2\Delta w_B^2}\right\}$$

όπου  $\Delta w_A$  και  $\Delta w_B$  είναι τα εύρη των καλαθιών των εντάσεων των εικόνων A και B. Ένα μεγαλύτερο εύρος συνήθως οδηγεί σε μεγαλύτερη υπολογιστική απόδοση.

Πλέον έχουμε έναν ολοκληρωμένο τύπο για την αμοιβαία πληροφορία, MI. Η αντίστοιχη συνάρτηση ανομοιότητας ορίζεται ως η  $-MI$ .

**2. Συναρτήσεις Ομαλοποίησης:** Οι συναρτήσεις αυτές προστίθενται στην συνάρτηση κόστους αφενός έτσι ώστε το πεδίο παραμόρφωσης να είναι κατά το δυνατόν λείο και αφετέρου για την αντιμετώπιση ενδεχόμενης υπερπροσαρμογής. Προφανώς υπάρχουν πάρα πολλές συναρτήσεις ομαλοποίησης που δρουν η καθεμία με διαφορετικό τρόπο στην εκπαίδευση του μοντέλου. Εδώ θα παρουσιάσουμε εκείνες τις τεχνικές που χρησιμοποιούνται ευρέως σε εφαρμογές με ιατρικές εικόνες και κυρίως σε εφαρμογές ευθυγράμμισης.

- Ομαλοποίηση τύπου  $\mathcal{L}_2$  ή Ομαλοποίηση Ridge: Αποτελεί τον γνωστό σε όλους μας όρο ομαλοποίησης του οποίου ο ρόλος είναι να κρατάει τα βάρη του δικτύου σε μία κατά το δυνατόν μικρότερη γειτονιά του 0, χωρίς να τα μηδενίζει. Η δράση αυτή καταπολεμά την υπερπροσαρμογή, καθώς απαγορεύει τα βάρη του δικτύου από το να πάρουν οποιαδήποτε τιμή. Αν  $\mathbf{w}$  είναι το διάνυσμα των παραμέτρων του μοντέλου, τότε η συνάρτηση που περιγράφει τον όρο αυτόν είναι η:

$$\mathcal{R}(\mathbf{w}) = \lambda \sum_{i=1}^n w_i^2$$

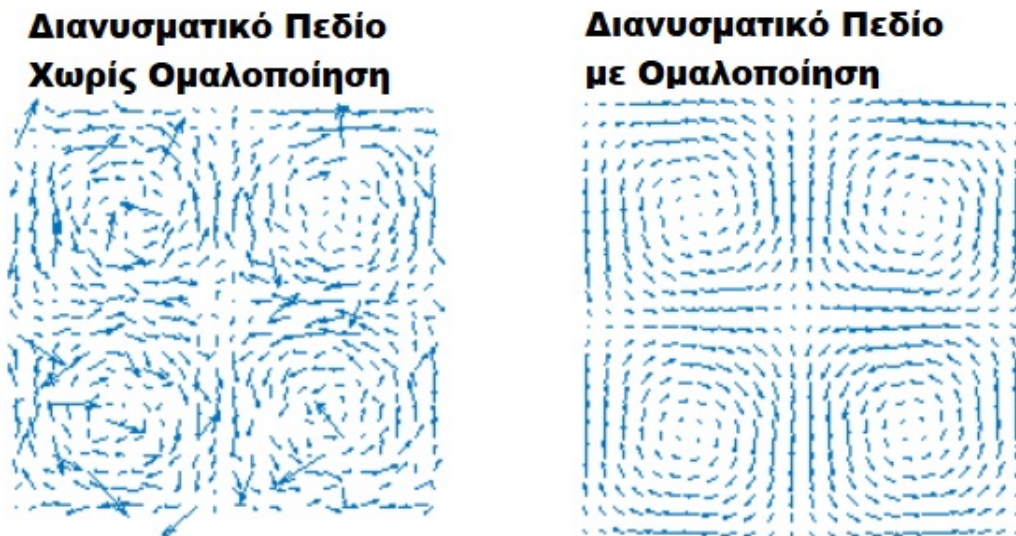
όπου  $\lambda$  μία θετική σταθερά, συνήθως μεταξύ του 0 και του 1. Μεγάλες τιμές της σταθεράς αυτής οδηγούν σε ισχυρότερη δράση ομαλοποίησης ενώ μικρές τιμές σε ασθενέστερη δράση. Ο όρος αυτός δεν συνεισφέρει στην εξομάλυνση του πεδίου παραμόρφωσης, αλλά δεν παύει να είναι χρήσιμος σε πολλές εφαρμογές.

- Ομαλοποίηση της Κλίσης του Πεδίου (Gradient of Field Regularization): Ο όρος αυτός αποτελεί ένα ομαλοποιητή της διάχυσης του πεδίου  $u$  (και άρα του πεδίου  $\phi$ ). Αφορά τις χωρικές παραγώγους του πεδίου  $u$  και ορίζεται ως [12]:

$$\mathcal{R}(u(\mathbf{w})) = \lambda \sum_{p \in \Omega} \|\nabla u(p)\|_2^2$$

Οι χωρικές παράγωγοι (δηλαδή το διάνυσμα κλίσης του πεδίου  $u$ ,  $\nabla u(p) = (\frac{\partial u(p)}{\partial x}, \frac{\partial u(p)}{\partial y}, \frac{\partial u(p)}{\partial z})$ ), υπολογίζονται με πρόσθιες πεπερασμένες διαφορές. Έχουμε,

για παράδειγμα:  $\frac{\partial u(p)}{\partial x} \sim u((p_x + 1, p_y, p_z)) - u((p_x, p_y, p_z))$ . Ελαχιστοποιώντας τον όρο αυτόν, ως μέρος της συνάρτησης κόστους οδηγεί, διαισθητικά, σε μικρότερη διάχυση του διανυσματικού πεδίου, που σημαίνει ότι θα υπάρχουν πολύ λιγότερα σημεία στα οποία το πεδίο θα έχει απότομες μεταβολές. Σημαντικά μπορεί κανείς να δει την παρακάτω εικόνα:



Εικόνα 9: Επίδραση του όρου ομαλοποίησης στο διανυσματικό πεδίο μετατόπισης

- Ομαλοποίηση Ολικής Απόκλισης (Total Variation Regularizer): Ο όρος αυτός είναι αρκετά παρόμοιος με τον όρο ομαλοποίησης της κλίσης, μόνο που σε αυτήν

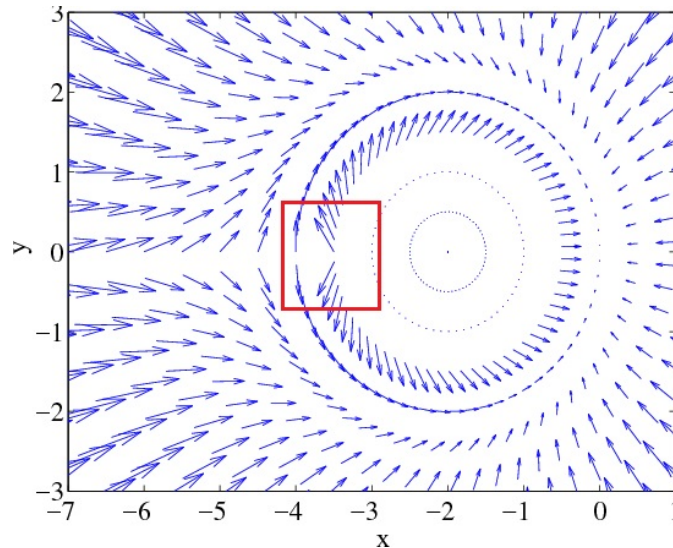


την περίπτωση χρησιμοποιούμε την 1-νόρμα της κλίσης, περιγράφεται δηλαδή από την σχέση:

$$\mathcal{R}(u(\mathbf{w})) = \lambda \sum_{p \in \Omega} \|\nabla u(p)\|_1$$

Για αναλυτική περιγραφή της δράσης και των ιδιοτήτων αυτού του είδους ομαλοποίησης παραπέμπουμε στο [25]. Διαισθητικά έχει παρόμοια δράση με την ομαλοποίηση της κλίσης.

- Ομαλοποίηση της Ορίζουσας της Ιακωβιανής (Determinant of Jacobian Regularization): Ο όρος αυτός δεν υπάρχει στην αρχική δημοσίευση του Voxel-Morph, αλλά αναπτύχθηκε σε μετέπειτα εργασίες. Παραπέμπουμε στις [22], [23]. Ας δούμε όμως την λογική πίσω από έναν τέτοιο όρο. Από τον λογισμό πολλών μεταβλητών, η ορίζουσα του Ιακωβιανού πίνακα μίας συνάρτησης (π.χ. ενός πεδίου) ποσοτικοποιεί την παραμόρφωση που προκαλεί η συνάρτηση αυτή. Στόχος μας εδώ είναι να επιβάλλουμε μία ποινή στα "δίπλωμα" (folding) της παραμόρφωσης καθώς και στα σημεία στα οποία δεν είναι αντιστρέψιμη. Με άλλα λόγια, επιβάλλουμε μία ποινή στις συνθήκες που καθιστούν το πεδίο παραμόρφωσης μη-λείο. Με "δίπλωμα" εννοούμε μία ανοιχτή περιοχή στην οποία ορίζεται το πεδίο αλλά παρουσιάζει απότομες μεταβολές. Κάτι τέτοιο φαίνεται στο κόκκινο περίγραμμα της παρακάτω εικόνας:



Εικόνα 10: Δίπλωμα (folding) ενός διανυσματικού πεδίου [26]

Στην περιοχή με το κόκκινο περίγραμμα παρατηρούμε ότι υπάρχει μία πολύ απότομη μεταβολή του πεδίου που σημαίνει ότι, τοπικά, το πεδίο δεν είναι παραγ-

ωγίσσιμο και άρα όχι λείο. Η δεύτερη συνθήκη είναι η μη-αντιστρεψιμότητα. Γνωρίζουμε ότι αν μία απεικόνιση πολλών μεταβλητών είναι λεία σε μία ανοιχτή περιοχή και έχει συνεχή αντίστροφη τότε, από το Θεώρημα Αντίστροφης Απεικόνισης, και η αντίστροφη θα είναι λεία στην αντίστοιχη περιοχή. Άρα αν το πεδίο μας είναι λείο και αντιστρέψιμο παντού, τότε και το αντίστροφο πεδίο θα είναι λείο. Θα είναι δηλαδή ένας *αμφιδιαφορισμός* (diffeomorphism), δηλαδή η καλύτερη ποιότητα πεδίου που μπορούμε να έχουμε. Για να ισχύουν οι παραπάνω δύο συνθήκες θα πρέπει η ορίζουσα του Ιακωβιανού πίνακα του πεδίου να είναι μικρότερη του 0. Με βάση την παραπάνω συλλογιστική πορεία κατασκευάζουμε τον ακόλουθο όρο ομαλοποίησης:

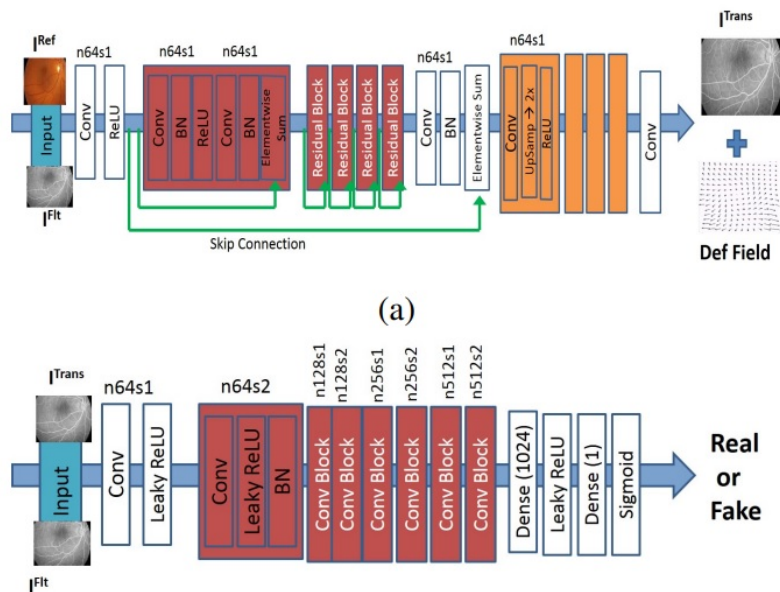
$$\mathcal{R}(\phi(\mathbf{w})) = \lambda \sum_{p \in \Omega} (|\det(\mathbf{J}_\phi(p))| - \det(\mathbf{J}_\phi(p)))$$

Έχουμε τελειώσει πλέον με την περιγραφή των πιο συνηθισμένων συναρτήσεων ανομοιότητας και ομαλοποίησης. Συνδυάζοντας τέτοιους όρους μέσω, για παράδειγμα, ενός σταθμισμένου αθροίσματος μίας συνάρτησης ανομοιότητας και μίας συνάρτησης ομαλοποίησης, σχηματίζουμε την συνάρτηση κόστους για την εκπαίδευση του μοντέλου. Αυτό που γίνεται στην πράξη είναι να επιλέγουμε μία συνάρτηση ανομοιότητας, έναν όρο ομαλοποίησης της ορίζουσας της Ιακωβιανής και έναν όρο ομαλοποίησης της κλίσης, χωρίς φυσικά να αποτελεί νόμο. Είναι θέμα δοκιμής και σφάλματος και πολλών πειραμάτων για να αποφασίσουμε ποια είναι η κατάλληλη επιλογή-συνδυασμός, ανάλογα με την φύση του προβλήματος και των δεδομένων. Τελειώνουμε το κεφάλαιο 3 με μία σύντομη αναφορά και περιγραφή διαφόρων state-of-the-art παραλλαγών του βασικού μοντέλου Voxelmorph.

### 3.2.4 Παραλλαγές του μοντέλου Voxelmorph

Είδαμε ότι το βασικό και θεμελιώδες μοντέλο Voxelmorph έχει σαν πυρήνα ένα νευρωνικό δίκτυο τύπου U-Net. Η επιλογή αυτή προφανώς δεν είναι ούτε μοναδική ούτε η "πιο σωστή". Οι ερευνητές που επινόησαν το Voxelmorph θα πρέπει να σκέφτηκαν ότι μέσω της αποσύνθεσης και την ανακατασκευής των εικόνων, το δίκτυο μαθαίνει τα ουσιώδη χαρακτηριστικά τους και μπορεί να παράξει μία απεικόνιση που να τις συνδέει. Στην εργασία αυτή, σε επόμενα κεφάλαια, θα μελετήσουμε και αυτήν την προσέγγιση αλλά και άλλες αρχιτεκτονικές που δεν έχουν δοκιμασθεί για χάρη της έρευνας. Προς το παρόν παραθέτουμε μερικές παραλλαγές που υπάρχουν και έχουν μελετηθεί:

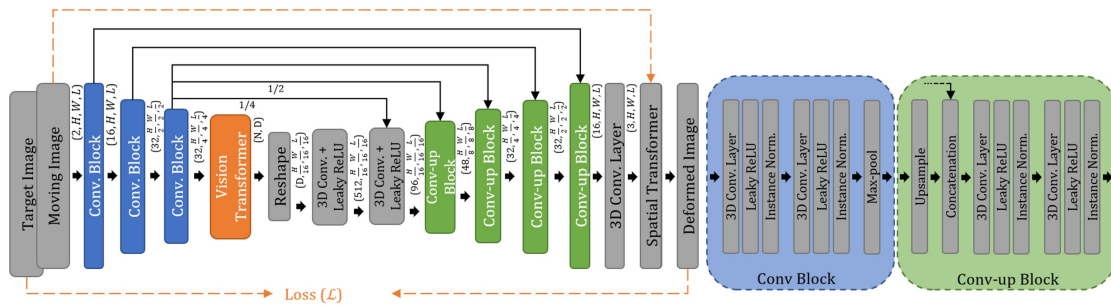
- **GAN [27]**: Το μοντέλο αυτό δεν έχει κάποιο φημισμένο όνομα όπως το VoxelMorph. Παρόλα αυτά η θεμελιώδης δομή του μοντέλου είναι τα Παραγωγικά Ανταγωνιστικά Δίκτυα (Generative Adversarial Networks). Η αρχιτεκτονική του μοντέλου φαίνεται παρακάτω:



Εικόνα 11: Σχηματική αναπαράσταση του μοντέλου βασισμένου στα GANs

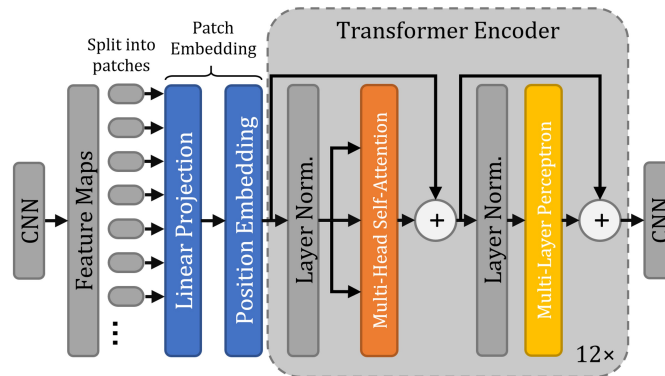
Τα παραγωγικά ανταγωνιστικά δίκτυα αποτελούν θεμελιώδεις δομές νευρωνικών δικτύων και χρησιμοποιούνται, μεταξύ άλλων, στην όραση υπολογιστών. Ένα GAN αποτελείται από δύο νευρωνικά δίκτυα, ένα με το όνομα παραγωγό (generator) και ένα με το όνομα διευκρινιστή (discriminator). Ο ρόλος του παραγωγού είναι να δημιουργήσει μία νέα εικόνα σχετική με το πρόβλημα. Με την σειρά του, ο διευκρινιστής θα πρέπει να κατηγοριοποιήσει την καινούρια αυτή εικόνα ως αληθινή (να προέρχεται δηλαδή όντως από το σύνολο δεδομένων) ή ψεύτικη (να έχει παραχθεί από τον παραγωγό). Η εκπαίδευση ενός GAN έγκειται στο να μπορέσει κάποια στιγμή ο παραγωγός να "ξεγελάσει" με μεγάλη πιθανότητα τον διευκρινιστή. Είναι μία λογική που προέρχεται από την θεωρία παιγνίων αλλά βρίσκει εφαρμογή και στην μηχανική μάθηση. Στα πλαίσια της ευθυγράμμισης εικόνων, ο παραγωγός πρέπει να μπορεί τελικά, μετά την εκπαίδευση, να παράγει μία εικόνα που να μοιάζει τόσο πολύ με την εικόνα στόχο ώστε ο διευκρινιστής να μην μπορεί να καταλάβει αν είναι όντως η εικόνα στόχος του συνόλου δεδομένων. Για επιπλέον πληροφορίες για την λειτουργία των GANs και την εφαρμογή τους στην ευθυγράμμιση εικόνων παραπέμπουμε στην εργασία [27].

- **ViT-V-Net [28]:** Το μοντέλο αυτό χρησιμοποιεί μία αρχιτεκτονική νευρωνικών δικτύων που ανήκει στην κλάση των *Μετασχηματιστών Όρασης* (Vision Transformers). Είναι δομές που χρησιμοποιούνται κατά κόρον στην όραση υπολογιστών. Η συνολική δομή του βασικού μοντέλου ViT-V-Net φαίνεται παρακάτω:



Εικόνα 12: Σχηματική αναπαράσταση του μοντέλου ViT-V-Net

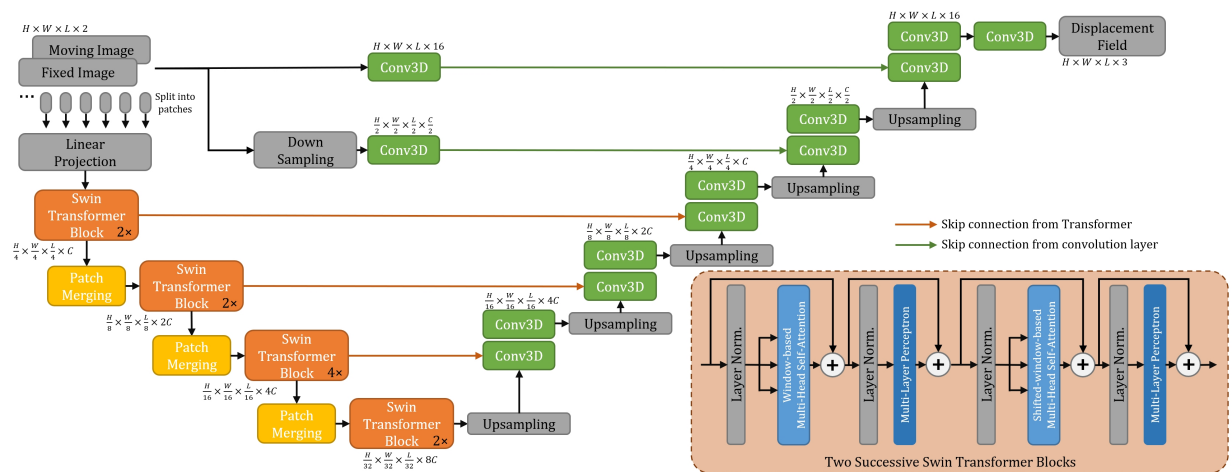
Στην παραπάνω εικόνα βλέπουμε αρκετά δομικά συστατικά που γνωρίσαμε και στο μοντέλο VoxelMorph. Η ουσιαστική διαφορά έγκειται στο ένα και μόνο κομμάτι που μας είναι άγνωστο και είναι φυσικά ο μετασχηματιστή όρασης. Το κομμάτι αυτό φαίνεται πιο αναλυτικά στην παρακάτω εικόνα:



Εικόνα 13: Σχηματική αναπαράσταση ενός μετασχηματιστή όρασης

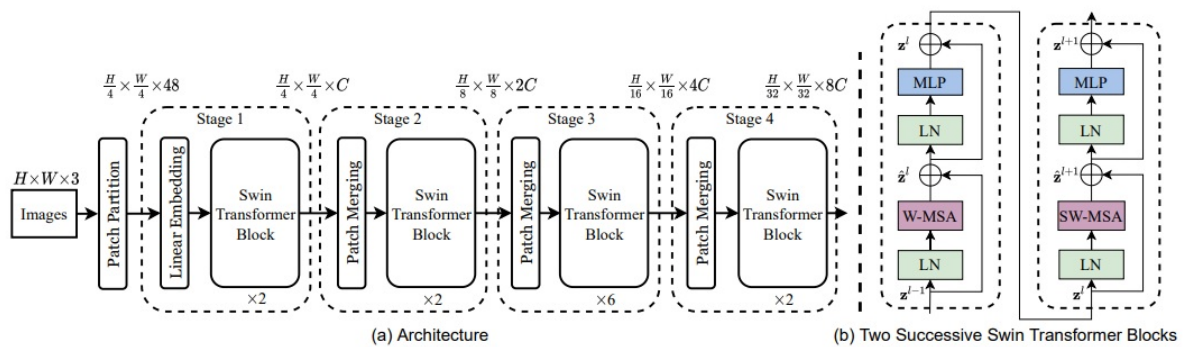
Οι μετασχηματιστές όρασης είναι σύγχρονα εργαλεία στην βαθιά μάθηση που συνδυάζουν τεχνικές ανάλυσης εικόνας (συνελικτικά νευρωνικά δίκτυα) και μετασχηματιστές. Μπορούν όχι μόνο να κάνουν εξαγωγή χαρακτηριστικών (μέσω των συνελίξεων) αλλά και να συσχετίζει τα χαρακτηριστικά ενός μέρους μίας εικόνας με τα χαρακτηριστικά ενός άλλου μέρους της ίδιας εικόνας. Για περαιτέρω λεπτομέρειες παραπέμπουμε στην εργασία [28].

- **TransMorph [29]:** Όπως μπορεί να προδίδει και το όνομα, το μοντέλο TransMorph χρησιμοποιεί στην καρδιά του τους λεγόμενους μετασχηματιστές (transformers). Παρακάτω βλέπουμε πώς μοιάζει το μοντέλο TransMorph:



Εικόνα 14: Σχηματική αναπαράσταση του μοντέλου TransMorph

Το βασικό κομμάτι του TransMorph είναι ο μετασχηματιστής Swin. Φαίνεται η δομή του πιο αναλυτικά στην παρακάτω εικόνα [30]:



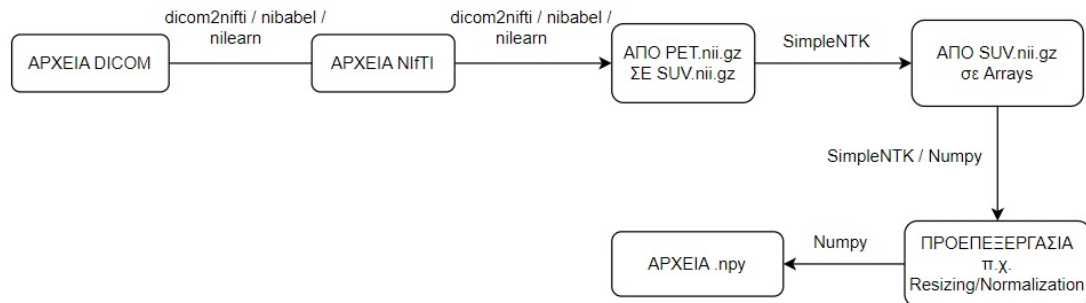
Εικόνα 15: Σχηματική αναπαράσταση του μετασχηματιστή Swin

Όλα τα παραπάνω αποτελούν πανίσχυρα state-of-the-art μοντέλα που χρησιμοποιούνται όχι μόνο στην ευθυγράμμιση εικόνων αλλά και σε πολλές άλλες εφαρμογές. Από τα διαγράμματα και μόνο μπορεί κανείς να καταλάβει όμως ότι η εκπαίδευση ενός μοντέλου τέτοιας κλίμακας απαιτεί τεράστια υπολογιστική ισχύ και είναι πολύ κοστο-και-χρονοβόρα. Σύγχρονες κάρτες γραφικών με πληθώρα μνήμης αξιοποιούνται για την παραλληλοποίηση των υπολογισμών που εκτελούνται κατά την εκπαίδευση μοντέλων βαθιάς μάθησης. Στην παρούσα εργασία θα μελετήσουμε πιο απλές παραλλαγές του VoxelMorph, που όμως δεν φαίνεται να έχουν δοκιμασθεί στην βιβλιογραφία, ώστε να δούμε αν μπορούμε με πιο "φτωχά" μοντέλα να πετύχουμε αποτελέσματα εξίσου (ή περίπου) καλά με πιο ισχυρά μοντέλα.

Σε αυτό το σημείο το κεφάλαιο που σχετίζεται με την ευθυγράμμιση εικόνων κλείνει. Είδαμε με ποιον τρόπο μπορεί να γίνει η ευθυγράμμιση τόσο με κλασσικές τεχνικές όσο και με νευρωνικά δίκτυα. Αφήνουμε προς το παρόν πίσω το αλγοριθμικό κομμάτι και θα προσηλωθούμε στα επόμενα δύο κεφάλαια στο κομμάτι των δεδομένων.

## 4 Προ-επεξεργασία των Εικόνων

Το πρώτο, το πιο σημαντικό αλλά συνήθως και το πιο δύσκολο κομμάτι της έρευνας που αφορά την μηχανική μάθηση είναι η προ-επεξεργασία των δεδομένων που έχουμε στην διάθεσή μας. Ανάλογα με την φύση των δεδομένων που έχουμε λάβει θα πρέπει να κάνουμε και τις κατάλληλες τροποποιήσεις για να μπορέσουμε να τα τροφοδοτήσουμε στους διάφορους αλγορίθμους μηχανικής μάθησης. Συνήθως, στην περίπτωση της βαθιάς μάθησης, τα δεδομένα εισέρχονται στο νευρωνικό δίκτυο σε μορφή πίνακα που είναι αποθηκευμένος σε ένα αρχείο τύπου .npy. Αυτό ακριβώς θα κάνουμε και εμείς. Τα δεδομένα μας, δηλαδή το σύνολο των FDG-PET εικόνων, μας δόθηκαν σαν αρχεία DICOM. Πρέπει με αφετηρία αυτά τα αρχεία να φτάσουμε σε τύπο δεδομένων .npy. Στην παρακάτω εικόνα φαίνεται σχηματικά η πορεία της προ-επεξεργασίας των εικόνων:



Εικόνα 16: Σχηματική Αναπαράσταση της προ-επεξεργασίας των δεδομένων

Στις επόμενες υποενότητες θα δώσουμε μία σύντομη περιγραφή των βημάτων.

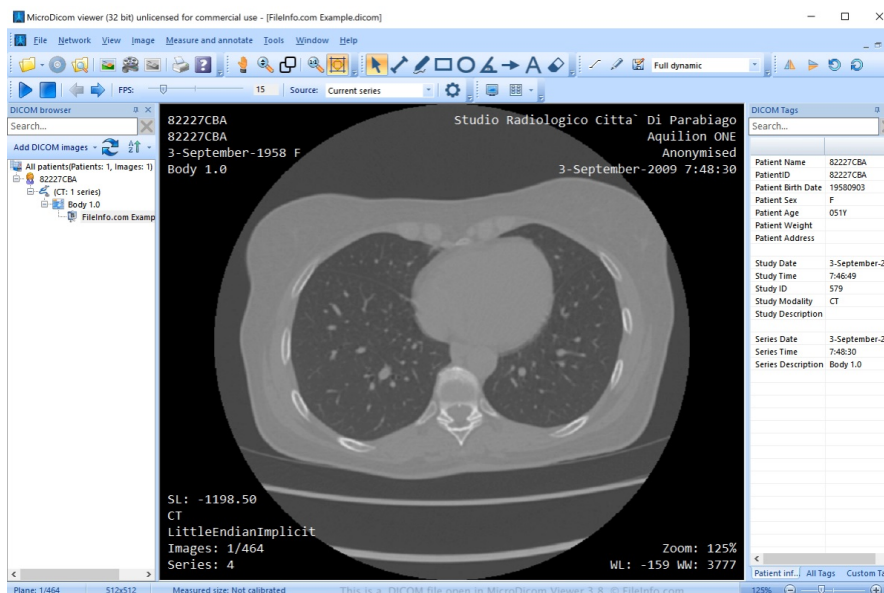
### 4.1 Βήμα 1: Μετατροπή των αρχείων DICOM σε αρχεία NIFTI

Θα ξεκινήσουμε με την παρουσίαση και την περιγραφή των τύπων αρχείων DICOM και NIFTI και στην συνέχεια θα δούμε σύντομα πώς γίνεται η μετατροπή από τον έναν τύπο στον άλλο.

#### 4.1.1 Τύπος Αρχείων DICOM

Όπως προαναφέρθηκε, οι ιατρικές εικόνες μας δόθηκαν σε μορφή αρχείων DICOM. Το DICOM είναι ένα ακρωνύμιο και σημαίνει *Digital Imaging and Communications in Medicine* [32]. Είναι ο πλέον καθιερωμένος τύπος αρχείων με τον οποίο αποθηκεύουμε ιατρικά δεδομένα και τα αναγνωρίζουμε από την κατάληξη .dcm. Τα δεδομένα αυτά είναι συνήθως εικόνες, όπως για παράδειγμα υπέρηχοι, αξονικές ή μαγνητικές τομογραφίες. Ο τύπος αυτός αναπτύχθηκε από την National Electrical Manufacturers Association (NEMA) [33]. Η πρώτη έκδοση των DICOM δημιουργήθηκε το 1983 ενώ συνεχίστηκαν να εξελίσσονται έως και την δεκαετία του 2000. Υπάρχουν διάφορα λογισμικά-εφαρμογές με τα οποία μπορεί

κάποιοι ερευνητές ή γιατροί να ανοίξει και να διαβάσει τα εν λόγω δεδομένα. Τέτοιες εφαρμογές είναι, μεταξύ άλλων, το MicroDicom, το ezDICOM, το OsiriX και το MeVisLab. Παρακάτω φαίνεται η διεπιφάνεια χρήστη του λογισμικού MicroDicom:



Εικόνα 17: Προβολή αρχείου DICOM με την χρήση του λογισμικού MicroDicom [33]

Το παραπάνω λογισμικό είναι προφανώς κατάλληλο για κάποιον ειδικό που ξέρει να διαβάζει τέτοιες εικόνες. Για εμάς όμως δεν έχει νόημα μία τέτοια απεικόνιση καθώς τα αρχεία DICOM δεν μπορούν να μπου σε ένα νευρωνικό δίκτυο. Το πρώτο βήμα είναι η μετατροπή τους σε αρχεία τύπου NIfTI.

#### 4.1.2 Τύπος Αρχείων NIfTI

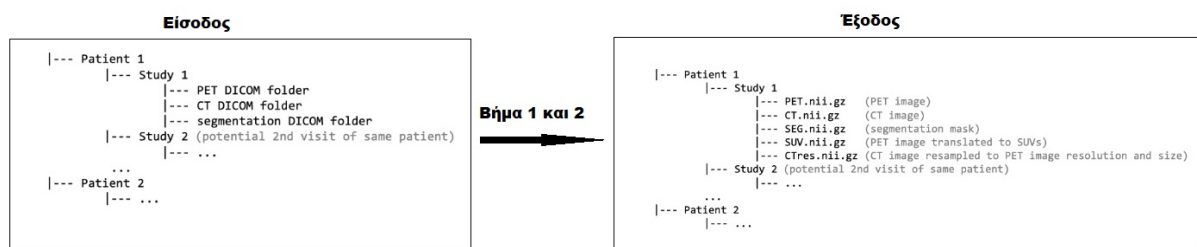
Το NIfTI είναι και αυτό ένα ακρωνύμιο και σημαίνει *Neuroimaging Informatics Technology Initiative* [34]. Είναι τύπος αρχείων που χρησιμοποιείται επίσης για την αποθήκευση ιατρικών δεδομένων, όπως fMRI και άλλου είδους εικόνων και τα αναγνωρίζουμε από την κατάληξη *.nii.gz*. Αναπτύχθηκαν από την National Institutes of Health (NIH) στις αρχές του 2000. Υπάρχουν δύο είδη αρχείων NIfTI, τα NIfTI-1 και τα NIfTI-2. Τα αρχεία NIfTI-2 αποτελούν μία σημαντική βελτίωση του τύπου NIfTI-1, καθώς επιτρέπουν την αποθήκευση μεγαλύτερου όγκου δεδομένων. Για περισσότερες τεχνικές πληροφορίες σε ότι αφορά τα αρχεία NIfTI παραπέμπουμε στους συνδέσμους [35] και [36].



### 4.1.3 Από DICOM σε NIFTI

Ο λόγος που πρέπει να γίνει η μετατροπή των αρχείων DICOM σε αρχεία NIFTI είναι καθαρά πρακτικός και αφορά τα διαθέσιμα πακέτα της γλώσσας Python. Ο τελικός μας στόχος είναι προφανώς το διάβασμα των δεδομένων και η απεικόνιση και η αποθήκευσή τους σε πίνακες (arrays). Η βιβλιοθήκη της Python που θα χρησιμοποιήσουμε για να διαβάσουμε τα δεδομένα μας και να τα ανοίξουμε σε μορφή πίνακα είναι η SimpleITK [38]. Το πακέτο αυτό παίρνει ως είσοδο μία εικόνα τύπου .nii.gz, την διαβάζει και ως έξοδο δίνει έναν κλασικό σε όλους μας πίνακα, τον οποίο προφανώς μπορούμε να απεικονίσουμε με κάποιο πακέτο όπως το matplotlib.

Ευτυχώς για εμάς, ο κώδικας που εκτελεί την απαραίτητη μετατροπή των αρχείων DICOM σε NIFTI είναι ανοιχτός σε όλους και μπορεί κανείς να τον βρει στο [37]. Μερικές από τις βιβλιοθήκες που αξιοποιούνται στην υλοποίηση του εν λόγω κώδικα είναι οι dicom2nifti, nibabel, nilearn και rydicom. Είναι σχεδιασμένες για να κάνουν, μεταξύ άλλων, αυτήν ακριβώς την μετατροπή. Για να δουλέψει ο συγκεκριμένος κώδικας θα πρέπει τα αρχεία μας να είναι οργανωμένα με έναν συγκεκριμένο τρόπο. Τα αρχεία εξόδου θα είναι επίσης οργανωμένα με συγκεκριμένο τρόπο που φαίνεται στην παρακάτω εικόνα:



Εικόνα 18: Μετατροπή των αρχείων DICOM σε NIFTI [37]

Οι παράμετροι εισόδου του κώδικα είναι το μονοπάτι (path) του φακέλου που περιέχει τα αρχεία DICOM και το μονοπάτι του φακέλου που θέλουμε να παραχθούν τα αρχεία NIFTI (συνεπώς μετά την εκτέλεση του κώδικα δεν χάνουμε τα αρχεία DICOM). Όπως βλέπουμε στην εικόνα 18, ο φάκελος εισόδου θα πρέπει να περιέχει μέσα φακέλους με τα ονόματα των ασθενών (ή κάποιο αναγνωριστικό αυτών). Εν συνέχεια, ο κάθε φάκελος ασθενούς θα πρέπει να περιέχει μέσα του φακέλους με τις χρονικές στιγμές (μέρες) στις οποίες πραγματοποιήθηκαν οι εξετάσεις του ασθενούς. Τέλος, ο κάθε χρονικός φάκελος θα πρέπει να περιέχει μέσα του τρία αρχεία DICOM, την τομογραφία PET και CT του ασθενούς καθώς επίσης και ένα αρχείο *κατάτμησης* (segmentation) που αποτελεί στην ουσία έναν φωτεινό χάρτη του σώματος του ασθενούς. Θα δούμε στην πορεία ποιον σκοπό εξυπηρετούν τα αρχεία αυτά.

Στην εικόνα 18 βλέπουμε ποια θα είναι η οργάνωση των αρχείων NIFTI στον φάκελο εξόδου. Ομοίως με πριν έχουμε φακέλους με τα αναγνωριστικά των ασθενών και μέσα

στον καθένα έχουμε φακέλους με τις χρονικές στιγμές των εξετάσεων. Τώρα όμως ο κάθε τέτοιος φάκελος θα περιέχει μέσα του πέντε διαφορετικά αρχεία NIfTI. Το PET.nii.gz (CT.nii.gz, SEG.nii.gz) είναι προφανώς το μετασχηματισμένο αρχείο της εικόνας (DICOM) PET (CT, segmentation). Το αρχείο CTres.nii.gz, όπως φαίνεται και στην εικόνα είναι ένα αντίγραφο της εικόνας CT.nii.gz, αλλά τροποποιημένο κατάλληλα έτσι ώστε να έχει τις ίδιες διαστάσεις με την εικόνα PET.nii.gz. Αυτό μπορεί να είναι χρήσιμο σε διάφορες εφαρμογές που θέλει κάποιος να χρησιμοποιήσει και τις δύο εξετάσεις, αλλά στην δικιά μας περίπτωση θα χρησιμοποιήσουμε τελικά μόνο τις εικόνες τύπου PET. Τέλος, παίρνουμε και ένα αρχείο SUV.nii.gz που είναι καθοριστικής σημασίας για την μελέτη μας. Συνεπώς, με βάση την εικόνα 16, ο ίδιος κώδικας εκτελεί το πρώτο και το δεύτερο βήμα της προ-επεξεργασίας. Παρόλα αυτά, επειδή δεν είναι προφανές τι είδους μετατροπή είναι αυτή από PET σε SUV, το βάλουμε σαν ξεχωριστό βήμα που θα περιγράψουμε σύντομα στην επόμενη υποενότητα.

## 4.2 Βήμα 2: Μετατροπή των αρχείων PET.nii.gz σε SUV.nii.gz

Η μετατροπή των εικόνων PET σε εικόνες SUV είναι επίσης ένα άκοπο βήμα που όμως είναι ύψιστης σημασίας. Το SUV είναι και αυτό ακρωνύμιο που σημαίνει *Standardized Uptake Values* και αποτελεί έναν δείκτη ήμι-ποσοτικής μέτρησης της απορρόφησης ενός ιστού. Οι εικόνες αυτές χρησιμοποιούνται πλέον ευρέως σε εφαρμογές κλινικής ογκολογικής απεικόνισης FDG-PET/CT. Ο ρόλος που διαδραματίζουν είναι να δώσουν μία ένδειξη της ανταπόκρισης ενός καρκινοπαθούς ασθενούς σε κάποια θεραπεία, καθώς επίσης να ξεχωρίσουμε κατά πόσο ένας όγκος είναι καλοήθης ή κακοήθης.

Πολλοί παράγοντες επηρεάζουν τον δείκτη SUV, όπως η δόση του ραδιενεργού υλικού, ο χρόνος που έχει περάσει από την έγχυση του υλικού μέχρι την απεικόνιση του ασθενούς, το βάρος του ασθενούς και τα επίπεδα γλυκόζης στο αίμα [40]. Ο βασικός τύπος υπολογισμού των τιμών SUV είναι ο εξής [39]:

$$SUV = \frac{r}{(\alpha/w)} \quad (4.2.1)$$

όπου  $r$  είναι η συγκέντρωση της ραδιενεργής ενέργειας, εκφρασμένη σε  $\frac{kBq}{mL}$ , που μετρείται από τον σαρωτή PET εντός μία περιοχής ενδιαφέροντος.  $\alpha$  είναι η διορθωμένη ποσότητα του εγχυμένου FDG, μετρημένο σε  $kBq$ , και  $w$  είναι το βάρος του ασθενούς σε γραμμάρια (g). Εάν το ραδιενεργό FDG κατανέμεται ομοιόμορφα μέσα στο σώμα του ασθενούς, τότε ο δείκτης SUV θα είναι παντού  $1\frac{g}{ml}$ , ανεξαρτήτως από την ποσότητα FDG και το βάρος του ασθενούς. Μπορεί κανείς να κάνει την παραδοχή ότι 1mL ιστού ζυγίζει 1g και τότε ο SUV θα αποτελεί μία αδιάστατη ποσότητα.

Θα περιγράψουμε τώρα συνοπτικά πώς γίνεται αυτή η μετατροπή της κάθε τιμής της εικόνας PET σε SUV σε επίπεδο κώδικα. Το πρώτο βήμα είναι ο υπολογισμός του παράγοντα διόρθωσης SUV. Αυτό γίνεται με την χρήση της βιβλιοθήκης pydicom [41]. Η βιβλιοθήκη αυτή υπολογίζει, με βάση κάποιους ευριστικούς κανόνες, την συνολική ποσότητα εγχυμένου ραδιενεργού υλικού,  $D$ , τον χρόνο ημιζωής του υλικού,  $\lambda$ , το βάρος του ασθενούς,  $w$ , καθώς επίσης και την χρονική διαφορά μεταξύ της έγχυσης του υλικού και της

ιατρικής απεικόνισης,  $\Delta\tau$ . Από αυτές τις ποσότητες υπολογίζεται η λεγόμενη *ενεργή δόση*, που είναι ίση με:

$$A = D \cdot 2^{-\frac{\Delta\tau}{\lambda}}$$

Τέλος, ο παράγοντας διόρθωσης SUV υπολογίζεται ως εξής:

$$SUV = 1000 \frac{w}{A}$$

Η σχέση αυτή μοιάζει πολύ με την εξίσωση 4.2.1 και περιέχει την ίδια ποιοτική πληροφορία. Ο πολλαπλασιασμός με 1000 γίνεται έτσι ώστε να πάρουμε τις κατάλληλες μονάδες. Ο όρος αυτός στην συνέχεια πολλαπλασιάζει το κάθε στοιχείο της εικόνας PET, δίνοντας έτσι μία νέα εικόνα SUV. Αυτές οι εικόνες, μετά από κατάλληλη προ-επεξεργασία, θα αποτελέσουν τα δεδομένα εισόδου στο βαθύ νευρωνικό δίκτυο που θα υλοποιήσουμε για την ευθυγράμμιση.

### 4.3 Βήμα 3: Μετατροπή των αρχείων .nii.gz σε Arrays και προ-επεξεργασία

Σε αυτό το βήμα είμαστε πλέον έτοιμοι να διαβάσουμε τα αρχεία .nii.gz ως πίνακες και να εφαρμόσουμε τις κλασσικές τεχνικές προ-επεξεργασίας εικόνων στην μηχανική μάθηση. Όπως προαναφέρθηκε, το διάβασμα των εικόνων .nii.gz γίνεται με την βιβλιοθήκη της Python, SimpleITK. Από εκεί και πέρα, τόσο με την SimpleITK όσο και με την numpy θα εκτελέσουμε τους κατάλληλους μετασχηματισμούς. Παρακάτω περιγράφουμε τους πιο συνηθισμένους και χρήσιμους μετασχηματισμούς που χρησιμοποιούνται σε εφαρμογές όπως την δική μας:

- **Αλλαγή Μεγέθους (Resizing):** Είναι συνήθως ο πρώτος από τους μετασχηματισμούς που εφαρμόζονται και είναι απολύτως απαραίτητος, σε αντίθεση με όλους τους υπόλοιπους. Το μοντέλο μας, είτε είναι ένα νευρωνικό δίκτυο είτε κάτι άλλο, θα είναι κατασκευασμένο να δέχεται εισόδους δεδομένου και σταθερού μεγέθους. Αυτή είναι μία γενική αρχή στην μηχανική μάθηση. Δεν είναι δυνατό η μία για παράδειγμα εικόνα να έχει διαστάσεις  $50 \times 50 \times 50$  και η άλλη  $60 \times 60 \times 60$  και να αποτελούν είσοδο στο ίδιο νευρωνικό δίκτυο. Θα πρέπει να τροποποιήσουμε την μία, να της αλλάξουμε δηλαδή το μέγεθος, έτσι ώστε να έχουν και οι δύο το ίδιο. Ο λόγος που τα δεδομένα που λαμβάνουμε είναι διαφορετικού μεγέθους, είναι ενδεχομένως το γεγονός ότι η κάθε εικόνα-εξέταση διενεργήθηκε με την χρήση διαφορετικών ιατρικών συσκευών. Μπορεί για παράδειγμα οι δύο τομογράφοι να έχουν διαφορετικό κατασκευαστή και να δίνουν στο τέλος της εξέτασης εικόνες διαφορετικού μεγέθους. Οπότε η αναγωγή όλων των δεδομένων σε ένα κοινό μέγεθος είναι αναπόφευκτο. Παρόλα αυτά θα πρέπει να προσέξουμε ένα κρίσιμο σημείο. Θα πρέπει να σκεφτούμε ποιο θέλουμε να

είναι το μέγεθος των εικόνων μας. Συνηθίζεται να πραγματοποιούμε σμίκρυνση των εικόνων έτσι ώστε να μειώσουμε και το υπολογιστικό κόστος της εκπαίδευσης. Θα πρέπει όμως η σμίκρυνση να γίνει έτσι ώστε η ποιότητα των εικόνων να μην αλλοιωθεί τελείως. Θα πρέπει δηλαδή "με το μάτι" να μπορούμε να κρίνουμε ότι η ανάλυση των εικόνων μας δεν έχει αλλοιωθεί σημαντικά μετά την αλλαγή μεγέθους. Αυτό γίνεται αφενός με την επιλογή όχι-και-τόσο μικρού τελικού μεγέθους και αφετέρου με την εφαρμογή φίλτρων, που αντιμετωπίζουν σε μεγάλο βαθμό την αλλοίωση των εικόνων. Τέτοια φίλτρα μπορεί να είναι γραμμικά ή μη-γραμμικά (Γκαουσιανά). Στην ουσία, τα φίλτρα είναι συναρτήσεις που υπολογίζουν σε τι βαθμό το κάθε εικονοκύτταρο της αρχικής εικόνας συνεισφέρει σε κάθε εικονοκύτταρο της τελικής εικόνας. Με βάση αυτές τις συνεισφορές, με μία διαδικασία παρεμβολής, υπολογίζονται οι τιμές των εικονοκυττάρων της τελικής εικόνας. Πακέτα όπως το Scikit-Image της Python έχουν ως προκαθορισμένο φίλτρο το Γκαουσιανό φίλτρο.

- **Απόκομμα Τιμών (Value Clipping):** Αποτελεί μία πολύ απλή ενέργεια που αποσκοπεί στο να περιορίσουμε τις τιμές ενός πίνακα. Η συνάρτηση αποκόμματος δέχεται ως είσοδο δύο αριθμούς, μία ελάχιστη τιμή,  $m$  και μία μέγιστη τιμή,  $M$ . Όσα στοιχεία του πίνακά μας έχουν τιμή μεταξύ των  $m$  και  $M$  παραμένουν ως έχουν. Όσα στοιχεία είναι μικρότερα από  $m$ , τα θέτουμε ίσα με  $m$ , ενώ όσα είναι μεγαλύτερα από  $M$  τα θέτουμε ίσα με  $M$ . Είναι δηλαδή ένας πολύ απλοϊκός τρόπος να περιορίσουμε τις τιμές του πίνακα. Παρόλα αυτά, προκειμένου η ενέργεια αυτή να έχει νόημα και να μην αλλοιώσει τον αρχικό μας πίνακα υπερβολικά πολύ, θα πρέπει να προσέξουμε οι τιμές  $m$  και  $M$  να μην είναι πολύ μακριά από την ελάχιστη και την μέγιστη τιμή του πίνακα. Θα πρέπει δηλαδή να ξέρουμε (κάτι που είναι εύκολο) σε ποιο εύρος κυμαίνονται τα στοιχεία του πίνακά μας. Για παράδειγμα, αν το ελάχιστο στοιχείο του πίνακά μας είναι 0 και η μέγιστη 50, τότε αν επιλέξουμε  $m = 20$  και  $M = 40$ , χάνουμε αρκετή πληροφορία που μπορεί να είναι σημαντική για τα δεδομένα μας.
- **Κανονικοποίηση (Normalization):** Στα πλαίσια της ανάλυσης εικόνας, η κανονικοποίηση είναι μία διαδικασία που αλλάζει την ένταση του κάθε εικονοκυττάρου της εικόνας με έναν συστηματικό τρόπο, περιορίζοντας τις εντάσεις αυτές. Έχει δηλαδή παρόμοια επίδραση με το απόκομμα τιμών, αλλά διατηρούνται οι ποιοτικές πληροφορίες της εικόνας. Όταν μιλάμε για κανονικοποίηση, συνήθως αναφερόμαστε στην γραμμική κανονικοποίηση, την οποία θα περιγράψουμε αμέσως. Έστω  $I$  μία  $n$ -διάστατη εικόνα  $I : \mathbb{R}^n \rightarrow [m, M]$ , όπου  $m$  και  $M$  η ελάχιστη και η μέγιστη τιμή της εικόνας αντίστοιχα. Στόχος μας είναι να πάρουμε μία νέα εικόνα  $I_N : \mathbb{R}^n \rightarrow [m', M']$ , μία εικόνα δηλαδή με διαφορετικό μέγιστο και ελάχιστο. Η γραμμική κανονικοποίηση της αρχικής εικόνας δίνεται από τον τύπο:

$$I_N = (I - m) \frac{M' - m'}{M - m} + m'$$

Η κανονικοποίηση μπορεί να είναι και μη-γραμμική, μέσω της χρήσης μίας μη-γραμμικής συνάρτησης, όπως είναι η σιγμοειδής. Σε αυτήν την περίπτωση θα πάρουμε την εικόνα [42]:

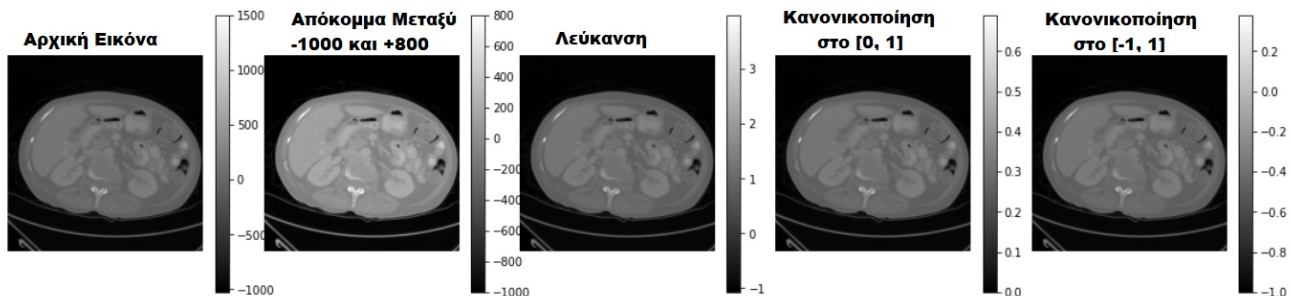
$$I_N = (M' - m') \frac{1}{1 + \exp\left(-\frac{I-\beta}{\alpha}\right)} + m'$$

Η παράμετρος  $\alpha$  αποτελεί το εύρος των τιμών της αρχικής εικόνας και η παράμετρος  $\beta$  ορίζει την ένταση γύρω από την οποία το εύρος κεντράρεται.

- Λεύκανση (Whitening): Είναι ακόμη μία ενέργεια που αποσκοπεί στον μετασχηματισμό των εντάσεων μίας εικόνας. Το αποτέλεσμα της λεύκανσης είναι μία νέα εικόνα, της οποίας η μέση τιμή των εντάσεων είναι 0 με διασπορά 1. Αυτό προφανώς επιτυγχάνεται αφαιρώντας από την αρχική εικόνα την μέση τιμή,  $\mu$  των εντάσεων και έπειτα διαιρώντας με την διασπορά τους,  $\sigma$ , δηλαδή:

$$I_N = \frac{I - \mu}{\sigma}$$

Στην παρακάτω εικόνα φαίνεται η επίδραση των παραπάνω μεθόδων προ-επεξεργασίας σε μία εικόνα:



Εικόνα 19: Κλασσικές μέθοδοι προ-επεξεργασίας μίας εικόνα [43]

Βεβαίως θα θέλαμε οι μετασχηματισμοί μας να μην αλλοιώνουν την πληροφορία των αρχικών μας εικόνων. Από την εικόνα 19 βλέπουμε ότι μόνο το απόκομμα τιμών φαίνεται να "ασπρίζει" λίγο την αρχική εικόνα, χωρίς όμως να αλλάζει ποιοτικά. Οι υπόλοιποι μετασχηματισμοί φαίνεται ότι αφήνουν την εικόνα ουσιαστικά απaráλλαχτη, τουλάχιστον στο ανθρώπινο μάτι.

Αφού λοιπόν οι ενέργειες που περιγράψαμε φαίνεται εκ πρώτης όψεως να μην αλλάζουν τις εικόνες, ίσως να αναρωτιέται κάποιος για ποιους λόγους να τις εφαρμόσουμε. Υπάρχουν δύο πολύ σημαντικοί λόγοι. Ο πρώτος και βασικότερος λόγος είναι η πηγή των δεδομένων μας. Όπως έχουμε ήδη αναφέρει, οι διαφορετικές τομογραφίες-εξετάσεις μπορεί να έχουν πραγματοποιηθεί σε διαφορετικά εργαστήρια με διαφορετικό εξοπλισμό. Συνεπώς οι εντάσεις των εικόνων αναμένονται να είναι αρκετά διαφορετικές, παρόλο που η ποιοτική πληροφορία είναι η ίδια. Άρα μετασχηματισμοί όπως η κανονικοποίηση και η λεύκανση ανάγουν όλες τις εικόνες στο ίδιο εύρος τιμών και άρα είναι καλύτερα συγκρίσιμες από τους αλγόριθμους μας. Έρχονται όλες δηλαδή σε μία κοινή βάση χωρίς να έχουν αλλάξει ποιοτικά. Ο δεύτερος λόγος είναι πιο τεχνικός -αλλά εξίσου σημαντικός- και έχει να κάνει με την φύση των υπολογισμών σε ένα νευρωνικό δίκτυο. Εν γένει, οι εντάσεις των εικόνων μπορεί να παίρνουν αρκετά μεγάλες θετικές ή αρνητικές τιμές. Οι τιμές αυτές, καθώς διαδίδονται στο νευρωνικό δίκτυο μέσω των πολλαπλασιασμών πινάκων μπορεί να οδηγήσουν σε προβλήματα υπερχειλίσης (overflow) και συνεπώς η εκπαίδευση να σταματήσει ή να επιβραδύνει σημαντικά. Μικρότερες τιμές οδηγούν εμπειρικά σε μία πιο ομαλή εκπαίδευση των βαθιών νευρωνικών δικτύων. Το απόκομμα τιμών ίσως να είναι ένα καλό πρώτο βήμα (αν εκτελεσθεί προσεχτικά) πριν την κανονικοποίηση ή την λεύκανση, αλλά συνήθως δεν απαιτείται.

Εφόσον τα δεδομένα μας είναι έτοιμα, δηλαδή έχουμε εκτελέσει όλα τα απαραίτητα βήματα προ-επεξεργασίας στις εικόνες μας, μπορούμε να τις αποθηκεύσουμε σε αρχεία τύπου `numpy`, δηλαδή αρχεία τύπου `.npy`. Τα αρχεία αυτά μπορούμε να τα διαβάζουμε και να τα φορτώνουμε κάθε φορά που θέλουμε να εκπαιδεύσουμε ή να επικυρώσουμε κάποιο μοντέλο χωρίς να τα αλλοιώσουμε. Η προ-επεξεργασία δηλαδή είναι μία πολύ κρίσιμη διαδικασία που θα πρέπει όμως ιδανικά να εκτελέσουμε μόνο μία φορά, εκτός φυσικά αν θέλουμε να δοκιμάσουμε πολλά διαφορετικά είδη προ-επεξεργασίας ή αν διαφορετικές υλοποιήσεις απαιτούν διαφορετική οργάνωση των αποθηκευμένων εικόνων.

## 5 Επισκόπηση των Δεδομένων

### 5.1 Πηγή των δεδομένων

Έχοντας μελετήσει την βασική δομή των μοντέλων ευθυγράμμισης εικόνων καθώς και τις πιθανές πορείες προ-επεξεργασίας των εικόνων, ήρθε η ώρα να γνωρίσουμε τα δεδομένα με τα οποία θα δουλέψουμε. Όπως έχει ήδη αναφερθεί, στην παρούσα εργασία θα χειριστούμε ιατρικές τομογραφίες FDG-PET. Η πηγή των δεδομένων μας είναι η ιστοσελίδα autoPET-Grand Challenge [44] καθώς επίσης και η [45]. Η ιστοσελίδα [44] αποτελεί μία διαδικτυακή αποθήκη ιατρικών εικόνων που προορίζονται για επεξεργασία με τεχνικές μηχανικής μάθησης με τελικό σκοπό την κατάτμησή ή την ευθυγράμμισή τους, ενώ στην [45] παρέχονται αναλυτικές πληροφορίες για τα δεδομένα μας. Περιέχει πάνω από χίλιες ολόσωμες τομογραφίες FDG-PET/CT από (ανώνυμους) ασθενείς από όλο τον κόσμο. Σχετικά με την κλινική κατάσταση των ασθενών, 501 τομογραφίες αφορούν ασθενείς με κακοήθες λέμφωμα, μελάνωμα και μεγάλου μεγέθους πνευμονικούς όγκους. Άλλες 503 τομογραφίες αφορούν ασθενείς με καλοήθεις όγκους. Όλες οι εξετάσεις-τομογραφίες έγιναν μεταξύ του 2014 και του 2018. Η εν λόγω συλλογή δεδομένων οφείλεται στις εξής δομές [46]:

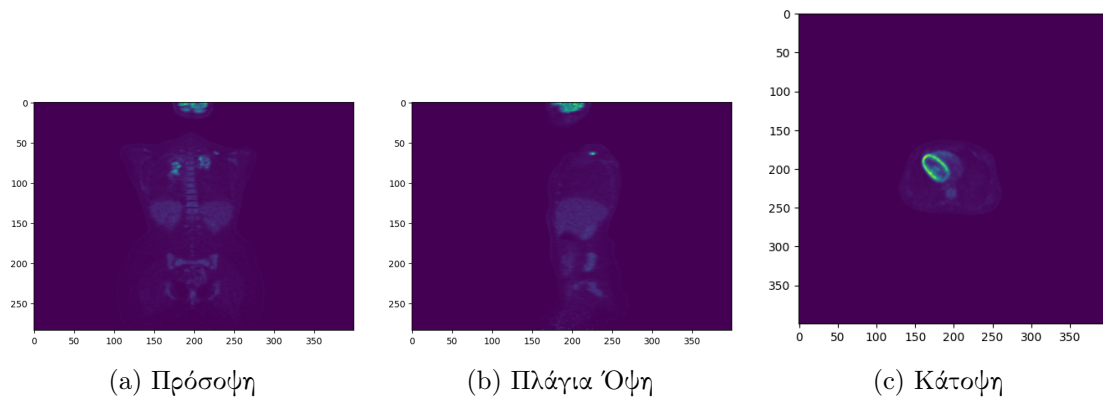
- Πανεπιστημιακό Νοσοκομείο του Tübingen, Γερμανία:
  - Christian La Fougère, MD του τμήματος πυρηνικής ιατρικής
  - Tobias Hepp, MD του τμήματος ραδιολογίας
  - Konstantin Nikolaou, MD του τμήματος ραδιολογίας
  - Christina Pfannenber, MD του τμήματος ραδιολογίας
- Πανεπιστημιακό Νοσοκομείο του Μονάχου, Γερμανία:
  - Clemens Cyran, του τμήματος ραδιολογίας
  - Michael Ingris, του τμήματος ραδιολογίας

Τα δεδομένα παρέχονται σε μορφή DICOM, που περιγράφηκε σε προηγούμενη ενότητα, και είναι οργανωμένα σε φακέλους ασθενών. Ο κάθε φάκελος ασθενούς περιέχει μέσα υπό-φακέλους με τις χρονικές στιγμές στις οποίες διενεργήθηκε η εκάστοτε εξέταση. Στην πηγή των δεδομένων υπάρχουν ασθενείς με μία μόνο χρονική στιγμή, δηλαδή μία μόνο τομογραφία FDG-PET. Παρόλα αυτά, για την δικιά μας εφαρμογή, χρειαζόμαστε τουλάχιστον

δύο χρονικές στιγμές-τομογραφίες για τον κάθε ασθενή προκειμένου να ευθυγραμμίσουμε την μία τομογραφία με την άλλη. Στην προκειμένη περίπτωση το εργαστήριο έχει πάρει την άδεια να κατεβάσει δεδομένα 81 ασθενών, κάθε ένας από τους οποίους περιέχει τουλάχιστον δύο τομογραφίες. Όπως αναφέραμε στην περίληψη και στην εισαγωγή της εργασίας οι εικόνες μας ανήκουν σε δύο κατηγορίες, τις SUV εικόνες και τις SEG εικόνες. Στα επόμενα δύο υπό-κεφάλαια θα γνωρίσουμε τις δύο αυτές κατηγορίες.

## 5.2 Εικόνες SUV

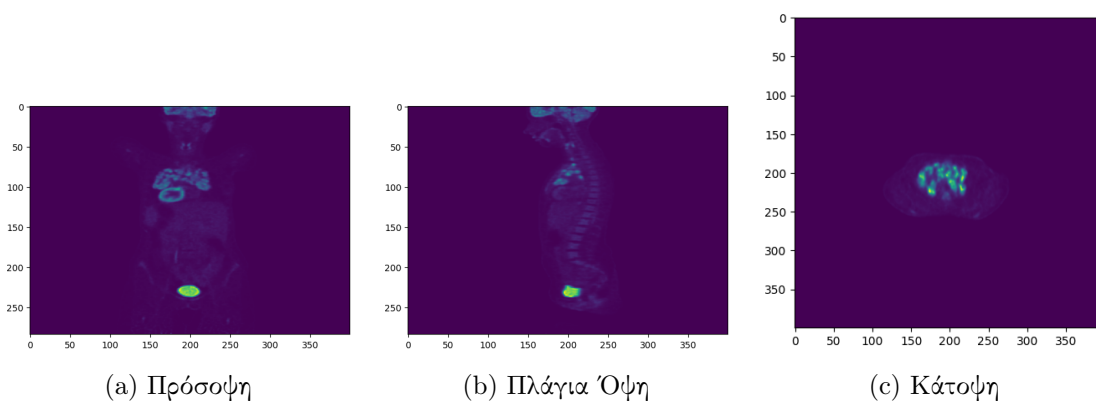
Όπως είδαμε στο κεφάλαιο 4, με τον όρο SUV εννοούμε ένα είδος κανονικοποίησης της έντασης (η οποία εκφράζει την συγκέντρωση της ραδιενέργειας στο σώμα) των εικονοστοιχείων των εικόνων. Οι SUV εικόνες αποτελούν το σύνολο εκπαίδευσης των μοντέλων μας. Θα οπτικοποιήσουμε τώρα μία από αυτές τις εικόνες για να πάρουμε μία ιδέα του πώς μοιάζουν. Είπαμε ότι έχουμε στα χέρια μας τρισδιάστατες εικόνες και, άρα, θα δείξουμε τρεις όψεις, την πρόσοψη, την κάτοψη και μία πλάγια όψη. Παρακάτω φαίνεται το πρώτο παράδειγμα. Η αρχική εικόνα έχει διαστάσεις (284, 400, 400) ενώ και για τις τρεις όψεις επιλέχθηκε η τομή 170 με χρωματική παλέτα την viridis του matplotlib:



Εικόνα 20: Διαφορετικές όψεις της ίδιας εικόνας (τομή 170)

Παρακάτω βλέπουμε τον ίδιο ασθενή, αλλά αυτή την φορά η σάρωση έχει προχωρήσει μέχρι την τομή 200:





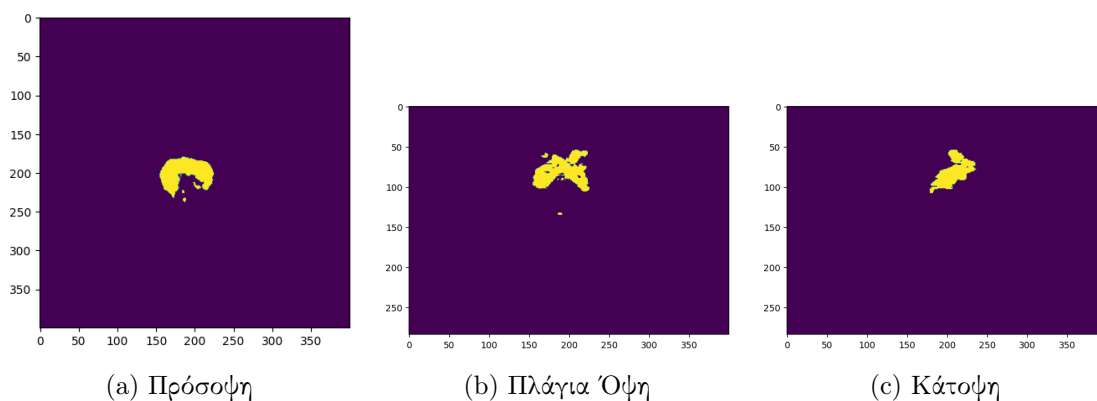
Εικόνα 21: Διαφορετικές όψεις της ίδιας εικόνας (τομή 200)

Η απόχρωση των παραπάνω εικόνων έγινε καθαρά για λόγους οπτικοποίησης. Στην πράξη, ο αριθμός των καναλιών των εικόνων μας θα είναι 1, δηλαδή θα πρόκειται για εικόνες, κάθε εικονοστοιχείο των οποίων θα χαρακτηρίζεται από έναν και μόνο αριθμό. Έτσι φαίνονται λοιπόν τα δεδομένα SUV. Ο κάθε ασθενής αντιστοιχίζεται σε δύο ή περισσότερες τρισδιάστατες εικόνες-τομογραφίες. Στόχος μας είναι να πάρουμε τις τομογραφίες που μας δίνονται και με κάποιον συστηματικό τρόπο να σχηματίσουμε ζεύγη εικόνων τα οποία θα ευθυγραμμίσουμε. Στην επόμενη ενότητα θα περιγράψουμε με ποιον τρόπο γίνεται αυτός ο διαχωρισμός των εικόνων σε ζεύγη.

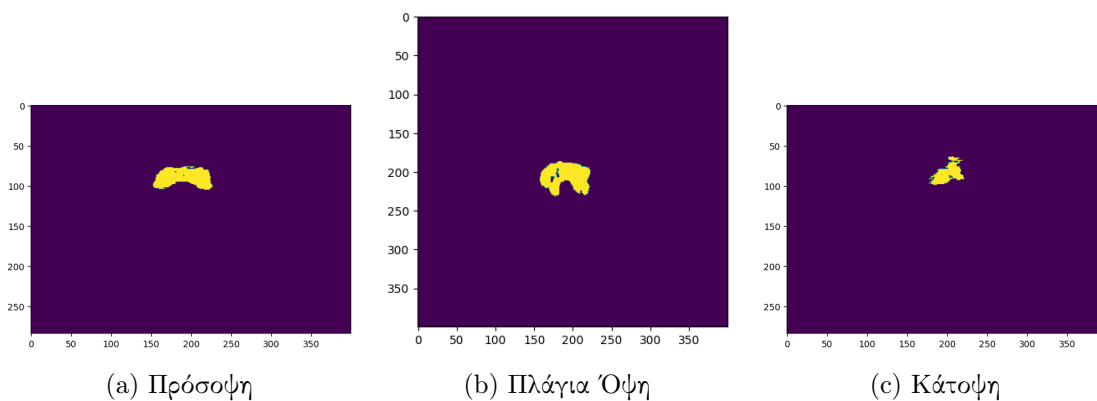
### 5.3 Εικόνες SEG

Για τις εικόνες αυτές δεν έχουμε μιλήσει πολύ. Το όνομά τους προέρχεται από τον όρο segmentation (κατάτμηση). Είναι και αυτές τρισδιάστατες εικόνες που στην ουσία αποτελούν έναν "χάρτη" του σώματος. Ένα εικονοστοιχείο μίας εικόνας SEG παίρνει τιμή 1, εάν ανήκει σε μία "ενδιαφέρουσα δομή" της εικόνας, αλλιώς 0. Στην δικιά μας εφαρμογή, μία εικόνα SEG έχει άσσους σε εκείνες τις περιοχές που αντιστοιχούν είτε σε όργανα είτε σε καρκινικούς όγκους του ασθενούς.

Για την οπτικοποίηση των μασκών χρησιμοποιήθηκε ο ίδιος ασθενής με τις εικόνες 20 και 21, με την διαφορά ότι εδώ πήραμε τις τομές 190 και 200. Αυτές φαίνονται παρακάτω:



Εικόνα 22: Διαφορετικές όψεις της ίδιας μάσκας (τομή 190)



Εικόνα 23: Διαφορετικές όψεις της ίδιας μάσκας (τομή 200)

Χρησιμοποιήθηκε η ίδια χρωματική παλέτα του matplotlib όπως και στην περίπτωση των SUV εικόνων. Οι κίτρινες περιοχές είναι τα εικονοστοιχεία των εικόνων που παίρνουν τιμή 1 και όπως φαίνεται αντιστοιχούν σε όργανα του ασθενούς (ενδεχομένως και σε όγκους). Όπως έχει ήδη αναφερθεί, οι μάσκες χρησιμοποιούνται για την αξιολόγηση του μοντέλου και ανήκουν τυπικά στο σύνολο ελέγχου. Θεωρητικά και, όπως έχει αποδειχθεί και στην πράξη, ένα μοντέλο που μπορεί αποδοτικά να ευθυγραμμίζει δύο SUV εικόνες θα πρέπει να είναι σε θέση να ευθυγραμμίζει αποδοτικά και δύο μάσκες. Μετά το πέρας της εκπαίδευσης δηλαδή δίνουμε στο μοντέλο μία *κινούμενη μάσκα* και μία *μάσκα στόχο*. Αυτές επεξεργάζονται με τον τρόπο που είδαμε, παράγεται το πεδίο παραμόρφωσης και αυτό με την σειρά του εφαρμόζεται πάνω στην κινούμενη μάσκα. Το αποτέλεσμα είναι μία νέα μάσκα που συγκρίνεται με την μάσκα στόχο. Η μετρική με την οποία τις συγκρίνουμε είναι ο δείκτης *Sørensen–Dice* (ή απλά *Dice*). Αν  $X$  και  $Y$  είναι δύο οποιαδήποτε σύνολα, τότε ο δείκτης *Dice* ορίζεται ως:

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|}$$

Εκφράζει δηλαδή τι ποσοστό των δύο συνόλων αλληλοεπικαλύπτεται. Μεγάλο δείκτης Dice (κοντά στο 1 δηλαδή) σημαίνει μεγάλη επικάλυψη και στην ουσία σημαίνει ότι τα δύο σύνολα ταυτίζονται (πέφτει το ένα πάνω στο άλλο). Η περίπτωση των εικόνων είναι μία ειδική περίπτωση του παραπάνω τύπου. Για να υπολογίσουμε τον δείκτη Dice μεταξύ δύο εικόνων αρχικά υπολογίζουμε το "λογικό και" μεταξύ των δύο εικόνων, δηλαδή σε πόσα εικονοστοιχεία έχουν και οι δύο εικόνες 1. Αυτός είναι ο αριθμητής (χωρίς το 2 φυσικά). Στην συνέχεια, για τον παρονομαστή, υπολογίζουμε το "λογικό ή" μεταξύ των δύο εικόνων (δηλαδή σε πόσα εικονοστοιχεία τουλάχιστον μία εικόνα έχει 1) και σε αυτό προσθέτουμε το "λογικό και". Διαιρούμε αριθμητή και παρονομαστή και παίρνουμε έτσι τον δείκτη Dice για δύο εικόνες. Στο σημείο αυτό θα δείξουμε κάτι που έχουμε υποσχεθεί από την αρχή της εργασίας, δηλαδή πώς μετατρέπουμε το πρόβλημα μη-επιβλεπόμενης μάθησης σε πρόβλημα επιβλεπόμενης μάθησης. Αυτό γίνεται προσθέτοντας στην οποιαδήποτε συνάρτηση κόστους έχουμε επιλέξει (από αυτές που περιγράφονται στο κεφ. 3.2.3) ένα επιπλέον όρο ποινής που σχετίζεται με τις μάσκες. Ο όρος αυτός ονομάζεται *ποινή Dice* (dice loss) και ισούται με  $1 - DSC$ . Προφανώς αν το νευρωνικό τροφοδοτείται με κάποια δέσμη (batch) δεδομένων, τότε ο επιπλέον αυτός όρος είναι ο αριθμητικός μέσος όρος των ποινών Dice για όλη την δέσμη. Στην επιβλεπόμενη έκδοση του προβλήματος δηλαδή, η συνολική συνάρτηση κόστους έχει την μορφή:

$$\mathcal{L} = \mathbb{E}_{(X_m, X_f) \in SUV \times SUV} [(1 - SIM(X_m, X_f))] + \mathbb{E}_{(Y_m, Y_f) \in SEG \times SEG} [(1 - DSC(Y_m, Y_f))]$$

Η εκδοχή της επιβλεπόμενης μάθησης έχει ξεχωριστό ενδιαφέρον για να το μελετήσει κανείς. Παρόλα αυτά, όπως αναφέρθηκε ήδη, στην παρούσα εργασία θα μελετήσουμε μόνο το πρόβλημα της μη-επιβλεπόμενης μάθησης και ο λόγος είναι η ποιότητα των διαθέσιμων δεδομένων. Στην περίπτωσή μας οι μάσκες δεν είναι κατάλληλες για την αξιολόγηση της ευθυγράμμισης των μοντέλων μας γιατί περιλαμβάνουν μόνο όγκους (δηλαδή μη-μηδενικές τιμές) σε ένα μικρό μέρος της εικόνας με πολλές μηδενικές τιμές και δεν έχουν σχεδιασμένες άλλες ανατομικές δομές, όπως τα όργανα, που θα ήταν πιο αντιπροσωπευτικές για την αξιολόγηση των μοντέλων. Συνεπώς η προσθήκη ενός επιπλέον όρου ποινής μασκών δεν θα έχει κάποιο νόημα στην δικιά μας εφαρμογή. Αυτό θα φανεί και από τον μέσο όρο των δεικτών Dice τους που θα υπολογίσουμε παρακάτω πριν την ευθυγράμμισή τους. Παρόλα αυτά θα πάρουμε από το σύνολο δεδομένων όσα ζευγάρια μασκών περιέχουν εικόνες που αμφότερες έχουν μη-μηδενικές τιμές και θα προσπαθήσουμε να κάνουμε μία ενδεικτική αξιολόγηση του μοντέλου με βάση τις μάσκες. Θα πάρουμε δηλαδή ένα εκπαιδευμένο μοντέλο και θα ακολουθήσουμε την διαδικασία που περιγράφηκε λίγο παραπάνω με όσα ζευγάρια μασκών μπορούμε. Φυσικά όλα τα μοντέλα που θα εκπαιδευσουμε θα αξιολογηθούν με τις SUV εικόνες του συνόλου ελέγχου, απλώς με τις μάσκες θα είχαμε μία παραπάνω ισχυρή ένδειξη της ποιότητας των μοντέλων.

Στο σημείο αυτό έχουμε τελειώσει με όλη την απαιτούμενη θεωρία. Σειρά έχει το πειραματικό μέρος της εργασίας που έχει φυσικά και το μεγαλύτερο ενδιαφέρον. Στο επόμενο κεφάλαιο θα περιγράψουμε με κατά το δυνατόν μεγαλύτερη λεπτομέρεια τα πειράματα που θα εκτελέσουμε, δηλαδή τις αρχιτεκτονικές των νευρωνικών δικτύων και τις διάφορες υπερπαραμέτρους τους.

## 6 Σχεδιασμός και Περιγραφή Πειραμάτων

### 6.1 Χαρακτηριστικά του Εξοπλισμού

Όλα τα πειράματα εκπονήθηκαν σε υπολογιστή του εργαστηρίου BioMed της σχολής Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Ε.Μ.Π.. Τα χαρακτηριστικά του υπολογιστή φαίνονται στο παρακάτω πίνακάκι:

Εξοπλισμός	Χαρακτηριστικά
Επεξεργαστής	12th Gen. Intel Core i7-12700, 2.1GHz
Κάρτα γραφικών	Nvidia GeForce RTX 3060 12GB
Μνήμη RAM	32GB

Πίνακας 7: Τεχνικά χαρακτηριστικά του εξοπλισμού που χρησιμοποιήθηκε

### 6.2 Προ-επεξεργασία των δεδομένων

Το επόμενο βήμα είναι να περιγράψουμε την γραμμή προ-επεξεργασίας των δεδομένων μας. Αρχικά τα δεδομένα τύπου DICOM μετατράπηκαν σε αρχεία τύπου NIFTI, με βάση το σχεδιάγραμμα της εικόνας 18. Σε αυτά τα δεδομένα PET.nii.gz έγινε η κανονικοποίηση SUV που περιγράφηκε στο κεφάλαιο 4. Σε αυτό το σημείο τα δεδομένα μας είναι οργανωμένα κατά ασθενή. Κάθε φάκελος ασθενούς περιέχει μέσα τόσους φακέλους, όσες είναι και οι τομογραφίες που εκπονήθηκαν, μαζί με τις αντίστοιχες χρονικές στιγμές. Συνεπώς, προκειμένου να σχηματισθούν ζευγάρια εικόνων που θα ευθυγραμμισθούν, θα πρέπει να χωρίσουμε κατάλληλα τους φακέλους αυτούς. Αυτό έγινε ως εξής: Ας υποθέσουμε, για παράδειγμα, ότι ένας φάκελος ασθενούς περιέχει μέσα του τρεις φακέλους-εξετάσεις που εκπονήθηκαν τις χρονικές στιγμές  $T_1, T_2$  και  $T_3$  (με την συνήθη χρονική διάταξη  $T_1 < T_2 < T_3$ ). Με βάση αυτές τις ημερομηνίες χωρίζουμε τα δεδομένα σε δύο περαιτέρω υπό-φακέλους ( $T_1 < T_2$ ) και ( $T_2 < T_3$ ). Με αυτόν τον τρόπο, από τρεις εικόνες, έχουμε πάρει δύο ζευγάρια εικόνων που θα ευθυγραμμίσουμε. Σε κάθε ζευγάρι, την εικόνα με την μικρότερη ημερομηνία την χρησιμοποιούμε ως κινούμενη εικόνα και την άλλη ως εικόνα στόχο. Αυτή είναι σχεδιαστική επιλογή και δεν έχει ιδιαίτερη σημασία. Στην συνέχεια πηγαίνουμε σε κάθε φάκελο με ημερομηνία  $T_i$  και κρατάμε μόνο τα αρχεία SUV.nii.gz (οι τομογραφίες) και SEG.nii.gz (οι μάσκες).

Τώρα τα αρχεία αυτά (SUV.nii.gz) θα πρέπει να υποστούν κατάλληλη επεξεργασία προκειμένου να μπορούν να τροφοδοτηθούν στα μοντέλα που θα επιλέξουμε. Όπως προαναφέραμε, θα πρέπει να επιλέξουμε ένα κοινό μέγεθος για όλες τις τρισδιάστατες εικόνες. Η επιλογή αυτή γίνεται με δύο βασικά κριτήρια. Το πρώτο κριτήριο είναι προφανώς να μην αλλοιωθεί σε σημαντικό βαθμό η ποιότητα των εικόνων. Το δεύτερο είναι κατά πόσο μας περιορίζει ο εξοπλισμός που χρησιμοποιούμε. Επειδή τόσο τα ίδια τα δεδομένα όσο και τα μοντέλα που εκπαιδεύονται είναι πολύ μεγάλα σε μέγεθος, η μνήμη του υπολογιστή (τόσο η

μνήμη RAM όσο και η μνήμη της GPU) αποτελεί καθοριστικό παράγοντα για την επιλογή του μεγέθους των εικόνων. Μετά από μερικές δοκιμές με μοντέλα και μεγέθη, αποφασίσαμε το μέγεθος των εικόνων να είναι  $128 \times 128 \times 128 \times 1$  (η τελευταία διάσταση είναι προφανώς ο αριθμός των καναλιών και επειδή οι εικόνες είναι σε κλίμακα του γκρι (grayscale), ισούται με 1). Η μετατροπή αυτή έγινε με την εντολή `resize` της βιβλιοθήκης `scikit-image`, όπου σαν επιπλέον όρισμα δόθηκε η παράμετρος `"anti_aliasing = True"` που βοηθά στο να μην αλλοιώνεται σε μεγάλο βαθμό η ποιότητα των εικόνων. Πέρα από το σχήμα των εικόνων, έγινε και κανονικοποίηση έτσι ώστε όλα τα εικονοστοιχεία των δεδομένων μας να παίρνουν τιμές στο  $[0, 1]$ . Αυτό γίνεται έτσι ώστε τα νευρωνικά δίκτυα να μην δέχονται μεγάλους αριθμούς στην είσοδό τους (προκαλώντας έτσι προβλήματα υπερχειλίσης) αλλά επίσης και για να εξαλείψουμε τυχόν ποιοτικές διαφορές μεταξύ των εικόνων. Όπως είπαμε, οι τομογραφίες προέρχονται ενδεχομένως από διαφορετικά μηχανήματα που μπορεί να παράγουν τις απεικονίσεις με διαφορετικό τρόπο. Με την (κοινή για όλες τις εικόνες) κανονικοποίηση καταφέρνουμε να φέρουμε όλες τις εικόνες σε μία "κοινή βάση" και αναμένουμε τυχόν ποιοτικές διαφορές να μην οφείλονται στις εντάσεις των εικονοστοιχείων.

Σε ό,τι αφορά τις μάσκες, θα πρέπει και αυτές να υποστούν παρόμοια προ-επεξεργασία. Αλλάζουμε το μέγεθός τους σε  $128 \times 128 \times 128 \times 1$  αλλά χωρίς την κανονικοποίηση. Όπως αναφέραμε στο κεφάλαιο 5, δυστυχώς η ποιότητα των μασκών δεν είναι καλή και δεν μπορούν να χρησιμοποιηθούν για να έχουμε μία πλήρη και αξιόπιστη αξιολόγηση των μοντέλων. Παρόλα αυτά θα τις χρησιμοποιήσουμε για να έχουμε μία απλή ένδειξη της απόδοσής τους.

Μετά την παραπάνω προ-επεξεργασία θα πρέπει να μετατρέψουμε τα αρχεία `.nii.gz` σε αρχεία `.npy` (`numpy`), με την βοήθεια της βιβλιοθήκης `SimpleITK`, μιας και οι γνωστές βιβλιοθήκες νευρωνικών δικτύων απαιτούν σαν είσοδο τέτοιου είδους αρχεία.

### 6.3 Υπερπαράμετροι

Όλα τα μοντέλα που εκπαιδεύτηκαν σχεδιάστηκαν με την χρήση της βιβλιοθήκης `TensorFlow` [47]. Η βιβλιοθήκη αυτή δίνει άμεση πρόσβαση στην GPU για ταχύτατους υπολογισμούς και καθιστά τον σχεδιασμό της αρχιτεκτονικής των μοντέλων, καθώς και την εκπαίδευσή τους, τετριμμένη. Η έκδοση του `TensorFlow` που χρησιμοποιήθηκε ήταν η 2.1.1.. Όπως είναι γνωστό, η εκπαίδευση ενός μοντέλου μηχανικής μάθησης συνήθως απαιτεί από τον προγραμματιστή να ορίσει εκείνος κάποιες σταθερές που σχετίζονται με τον εκάστοτε αλγόριθμο, τις λεγόμενες υπερπαράμετρους. Πρόκειται για παραμέτρους που, ως επί το πλείστον, δεν αλλάζουν κατά την διάρκεια της εκπαίδευσης (με λίγες εξαιρέσεις, όπως ο ρυθμός μάθησης, που μπορεί να μεταβάλλεται κατά την διάρκεια της εκπαίδευσης με ρυθμό που καθορίζει και πάλι ο προγραμματιστής) και φυσικά κατά την διάρκεια της επικύρωσης και του ελέγχου.

Σε ό,τι αφορά την προ-επεξεργασία των δεδομένων η μόνη υπερπαράμετρος είναι το μέγεθος των εικόνων που, όπως αναφέραμε παραπάνω, επιλέχθηκε ίσο με  $128 \times 128 \times 128 \times 1$ .

Συνεχίζουμε με τις υπερπαραμέτρους της εκπαίδευσης των μοντέλων. Σχεδόν όλα τα μοντέλα εκπαιδεύτηκαν για 600 εποχές, με εξαίρεση κάποια από τα τελευταία πειράματα με το μοντέλο D-UNET που εκπαιδεύτηκαν για 300 εποχές για λόγους χρονικού περιορισμού. Ο ρυθμός μάθησης (learning rate), δηλαδή το μέγεθος του βήματος ανανέωσης των παραμέτρων των νευρωνικών δικτύων τίθεται ίσος και σταθερός κατά την εκπαίδευση με 0.001. Ως βελτιστοποιητής (optimizer) που θα αναλαμβάνει τον υπολογισμό των παραγώγων και την ανανέωση των βαρών του δικτύου επιλέχθηκε ο Adam (Adaptive Momentum, [48]). Αποτελεί τον πιο ευρέως χρησιμοποιούμενο αλγόριθμο βελτιστοποίησης καθώς είναι ιδιαίτερα εύρωστος και γρήγορος. Με την σειρά του και αυτός έχει μέσα του κάποιες υπερπαραμέτρους, τις  $\beta_1$ ,  $\beta_2$  και  $\epsilon$ . Τέθηκαν ίσες με 0.9, 0.999 και  $10^{-7}$  αντίστοιχα. Σαν συναρτήσεις κόστους δοκιμάστηκαν οι: Δείκτης δομικής ομοιότητας (SSIM), Συντελεστής συσχέτισης Pearson (PCC), Σταθμισμένοι PCC και SSIM και Αμοιβαία Πληροφορία κατά Parzen. Οι τεχνικές λεπτομέρειες περί αυτών των συναρτήσεων δίνονται στο κεφάλαιο 3.2.3.. Το βάρος διαμοιρασμού του σταθμισμένου PCC/SSIM κόστους τίθεται ίσος με 0.5 και για τους δύο όρους. Ο αριθμός των καλαθιών (bins) για τον υπολογισμό της αμοιβαίας πληροφορίας κατά Parzen είναι ίσος με 128. Στην ποινή του δικτύου προστέθηκε σε κάθε πείραμα και ένας όρος ομαλοποίησης. Επιλέξαμε να δοκιμάσουμε τους όρους ομαλοποίησης  $L_2$  και Ολικής Απόκλισης (παραπέμπουμε και πάλι στο κεφάλαιο 3.2.3.). Ο όρος της "δύναμης" της ομαλοποίησης, που συνήθως συμβολίζεται με  $\lambda$  και πολλαπλασιάζει την συνάρτηση ομαλοποίησης, ορίστηκε ίσος με 0.5. Όλοι οι συνδυασμοί που δοκιμάστηκαν παρατίθενται παρακάτω.

Όπως αναφέραμε στο κεφάλαιο 5, έχουμε πρόσβαση σε 81 ασθενείς. Από αυτούς επιλέχθηκαν τυχαία οι 56 από αυτούς για να σχηματίσουμε το σύνολο εκπαίδευσης και επικύρωσης. Από τους 56 αυτούς, οι 12 ανήκουν στο σύνολο επικύρωσης. Συνεπώς έχουμε 44 ασθενείς για το σύνολο εκπαίδευσης, 12 για το σύνολο επικύρωσης και 25 για το σύνολο ελέγχου. Σε κάθε εποχή, το μέγεθος της δέσμης (batch size) που εισέρχεται στο νευρωνικό δίκτυο (και στην εκπαίδευση/επικύρωση και στον έλεγχο) είναι αναγκαστικά ίσο με 1 (δηλαδή κάθε φορά το νευρωνικό δίκτυο παίρνει στην είσοδό του ένα ζευγάρι κινούμενης-στόχου εικόνων). Λόγω του μεγέθους, τόσο των εικόνων όσο και των μοντέλων, στις περισσότερες εφαρμογές ευθυγράμμιση ιατρικών εικόνων το μέγεθος της δέσμης είναι είτε 1 ή 2, ανάλογα με την διαθέσιμη μνήμη.

Οι υπερπαραμέτροι που περιγράφηκαν παραπάνω παρατίθενται συνοπτικά στον παρακάτω πίνακα:

Υπερπαράμετρος	Τιμή
Μέγεθος εικόνων	128 × 128 × 128 × 1
Ρυθμός Μάθησης	0.001
Μέγεθος Δέσμης (Εκπαίδευση)	1
Μέγεθος Δέσμης (Επικύρωση)	1
Εποχές / Πείραμα	600
Μέγεθος Συνόλου Εκπαίδευσης	44
Μέγεθος Συνόλου Επικύρωσης	12
Μέγεθος Συνόλου Ελέγχου	25
Βελτιστοποιητής	Adam
$\beta_1$	0.9
$\beta_2$	0.999
$\epsilon$	$10^{-7}$
$\lambda$	0.5
$w_{PCC}$	0.5
Καλάθια Parzen	128

Πίνακας 8: Υπερπαράμετροι των πειραμάτων

## 6.4 Μοντέλα

Το μόνο που μένει πλέον είναι περιγράψουμε κατά το δυνατόν πιο λεπτομερώς την θεμελιώδη αρχιτεκτονική των μοντέλων-νευρωνικών δικτύων που χρησιμοποιήσαμε καθώς επίσης και τους συνδυασμούς συναρτήσεων κόστους και ομαλοποίησης με καθεμία από αυτές. Με "θεμελιώδη δομή" εννοούμε περισσότερο τον γενικό σκελετό των δικτύων και τον τρόπο με τον οποίο οι εικόνες επεξεργάζονται από αυτά. Εμείς δοκιμάσαμε τέσσερις θεμελιώδεις δομές. Αυτές ονομάστηκαν και, έτσι θα αναφερόμαστε από εδώ και στο εξής: U-NET, CNN, U-2-NET και D-UNET (Dense U-NET). Το TensorFlow περιλαμβάνει ένα πακέτο δι-αγραμματικής απεικόνισης των νευρωνικών δικτύων που ορίζει ο προγραμματιστής. Παρόλα αυτά, επειδή τα μοντέλα είναι τεράστια, τα διαγράμματα δεν μπορούν να διαβασθούν εύκολα. Παρατίθενται στο παράρτημα για όποιον ενδιαφέρεται. Παρακάτω περιγράφουμε την δομή των μοντέλων:

- **Το μοντέλο U-NET**

Το δίκτυο αυτό έχει την κλασική δομή ενός νευρωνικού δικτύου τύπου U-NET, όπως περιγράφηκε στο κεφάλαιο 3.2.1. και φαίνεται στην εικόνα 41. Επειδή ο σκελετός του μοντέλου είναι πολύ μεγάλος και ενδεχομένως στην εικόνα αυτή να μην μπορεί



κανείς να διακρίνει εύκολα τις λεπτομέρειές του, θα τις περιγράψουμε εδώ. Το δίκτυο αποτελείται από τέσσερα "μπλοκ" συνελίξεων. Το κάθε ένα από αυτά τα μπλοκ με την σειρά του αποτελείται από δύο στρώματα τρισδιάστατων συνελίξεων (convolutional layers) που ακολουθούνται από ένα στρώμα Max Pooling. Το στρώμα αυτό στην ουσία μειώνει την διάσταση της εικόνας εισόδου στο μισό. Ο αριθμός των φίλτρων που περιέχεται σε κάθε στρώμα συνελίξεως (από την αρχή του δικτύου και όσο πιο βαθιά πάμε) 32, 18, 8 και 8 αντίστοιχα. Προκειμένου να μην μειωθεί η διάσταση των εικόνων από την ενέργεια των συνελίξεων, δίνουμε και μία επιπλέον παράμετρο "Padding=Same". Μετά από τα τέσσερα αυτά μπλοκ έχουμε άλλα δύο τρισδιάστατα συνελικτικά στρώματα με 8 φίλτρα το καθένα. Το δίκτυο τώρα ξεκινά να "ανεβαίνει" με τα στρώματα από-συνελίξεως (de-convolution). Έχουμε και πάλι τέσσερα μπλοκ, καθένα από τα οποία έχει ένα στρώμα από-συνελίξεως και δύο στρώματα συνελίξεως. Ο αριθμός των φίλτρων από-συνελίξεως (συνελίξεως) σε κάθε μπλοκ είναι 256 (8), 128 (8), 64 (16) και 32 (32). Φυσικά γίνονται και, όπου απαιτείται, οι συνενώσεις (concatenations) των ενδιάμεσων χαρακτηριστικών. Η συνάρτηση ενεργοποίησης είναι σε όλο το δίκτυο η γνωστή ReLU. Το νευρωνικό δίκτυο τελειώνει με ένα στρώμα συνελίξεως που θα πρέπει κατ' ανάγκη να έχει τρία φίλτρα, προκειμένου η έξοδος του δικτύου να έχει διαστάσεις  $128 \times 128 \times 128 \times 3$ . Πρόκειται δηλαδή για μία νέα εικόνα (την οποία ονομάσαμε *διανυσματικό πεδίο παραμόρφωσης*) που διαφέρει ως προς την διάσταση μόνο στον αριθμό των καναλιών με τις εικόνες εισόδου. Ο αριθμός των καναλιών εδώ είναι 3, και οι τρεις αυτοί αριθμοί υποδεικνύουν σε κάθε εικονοστοιχείο της κινούμενης εικόνας με ποιον τρόπο πρέπει να "κινηθεί" σε κάθε μία από τις τρεις διαστάσεις έτσι ώστε το τελικό αποτέλεσμα να είναι μία εικόνα όμοια με την εικόνα στόχο (βλ. κεφάλαιο 3). Το τρισδιάστατο αυτό πεδίο εφαρμόζεται στην κινούμενη εικόνα, μέσω της διαδικασίας παρεμβολής που περιγράφηκε στο κεφάλαιο 3, και το αποτέλεσμα που προκύπτει αποτελεί την έξοδο του μοντέλου. Η μεθοδολογία αυτή είναι κοινή για όλα τα μοντέλα οπότε από εδώ και στο εξής θα περιγράφεται μόνο η δομή του μοντέλου. Το μοντέλο U-NET έχει 1.019.123 παραμέτρους (βάρη).

- **Το μοντέλο CNN**

Η ιδέα πίσω από το μοντέλο CNN είναι η εξής: Να μην υποβιβάζεται η διάσταση των εικόνων κατά την προώθηση της πληροφορίας μέσω του δικτύου. Έχουμε δηλαδή μόνο συνελικτικά στρώματα χωρίς Max Pooling. Αυτήν την φορά τα τέσσερα βασικά μπλοκ του δικτύου αποτελούνται από ένα συνελικτικό στρώμα (Padding=Same) που ακολουθείται από ένα στρώμα κανονικοποίησης δέσμης (Batch Normalization) με ενεργοποίηση ReLU παντού. Τα φίλτρα των συνελίξεων είναι σε αριθμό 64, 64, 48 και 16 αντίστοιχα. Τελειώνουμε με ένα στρώμα συνελίξεως με τρία φίλτρα. Για να έχουμε κοινή βάση σύγκρισης της ικανότητας των μοντέλων, προσπαθήσαμε το μοντέλο U-NET και CNN να έχουν παραπλήσιο αριθμό βαρών. Το μοντέλο CNN έχει 1.014.579 βάρη.

- **Το μοντέλο U-2-NET**

Από αυτό το σημείο και έπειτα, η λογική ήταν να σκεφτούμε τρόπους να μειώσουμε το μέγεθος των μοντέλων (με την έννοια του αριθμού των βαρών), χωρίς όμως να τα αποδυναμώσουμε. Αυτό έγινε με στόχο όχι μόνο μικρότερους χρόνους εκπαίδευσης αλλά και ενδεχομένως εκπαίδευσης με μεγαλύτερες εικόνες. Η πρώτη σκέψη είναι η εξής: Αντί να έχουμε ένα τεράστιο νευρωνικό δίκτυο τύπου U-NET που να επεξεργάζεται ταυτόχρονα ολόκληρες τις εικόνες (κινούμενη και στόχο), μπορούμε να σχεδιάσουμε δύο μικρότερα U-NETs, καθένα από τα οποία θα επεξεργάζεται την μισή κινούμενη και την μισή στόχο εικόνα. Μπορούμε να φανταστούμε τις δύο εικόνες ως δύο κύβους με ίδιο προσανατολισμό. Κόβουμε το "πάνω μισό" και από τις δύο εικόνες και αυτές τις δίνουμε σε ένα μικρό νευρωνικό δίκτυο και έτσι παίρνουμε το μισό από το συνολικό διανυσματικό πεδίο παραμόρφωσης. Κάνουμε το ίδιο με το "κάτω μισό" των δύο εικόνων και ένα άλλο μικρό δίκτυο. Παίρνουμε στο τέλος τα δύο πεδία παραμόρφωσης, τα συνενώνουμε με κατάλληλο τρόπο-προσανατολισμό (για να έχει τελικά διάσταση  $128 \times 128 \times 128 \times 3$ ) και αυτό εφαρμόζεται σε ολόκληρη προφανώς την κινούμενη εικόνα. Το αποτέλεσμα είναι η έξοδος του μοντέλου. Το σκεπτικό πίσω από την ιδέα αυτή είναι ότι αν ένα νευρωνικό δίκτυο έχει ως στόχο να ευθυγραμμίσει μία μικρότερη εικόνα, τότε αφενός θα χρειάζεται και λιγότερα βάρη και αφετέρου η ευθυγράμμιση ίσως να είναι πιο επιτυχημένη. Επίσης, δεν σημαίνει ότι επειδή οι επιμέρους εικόνες έχουν το μισό μέγεθος, το νευρωνικό δίκτυο θα πρέπει να έχει τα μισά βάρη. Προσπαθήσαμε το κάθε μικρότερο νευρωνικό δίκτυο να έχει αριθμό βαρών λιγότερο από το μισό των μοντέλων U-NET και CNN. Με αυτόν τον τρόπο μπορούμε να διαπιστώσουμε πόση "οικονομία" μπορούμε να κάνουμε στην πολυπλοκότητα των μοντέλων, χωρίς να θυσιάσουμε την επίδοση και απόδοσή τους. Τώρα θα περιγράψουμε την δομή του καθενός από τα δύο (ίδια) μικρότερα νευρωνικά δίκτυα του μοντέλου U-2-NET που φαίνεται στην εικόνα 43.

Τώρα έχουμε τρία μπλοκ συνελίξεων. Κάθε μπλοκ έχει πάλι δύο στρώματα συνελίξεων (Padding=Same) με 28 φίλτρα σε κάθε στρώμα. Στο τέλος του κάθε μπλοκ γίνεται και Max Pooling. Ακολουθούν άλλα δύο στρώματα συνελίξεων με 28 φίλτρα το καθένα. Τώρα ξεκινούν οι από-συνελίξεις. Έχουμε δύο μπλοκ, καθένα από τα οποία έχει ένα στρώμα από-συνελίξης και δύο στρώματα συνελίξεων. Ο αριθμός των φίλτρων των από-συνελίξεων (συνελίξεων) είναι 8 (28) και 8 (28) αντίστοιχα. Καταλήγουμε φυσικά σε ένα στρώμα συνελίξεων με τρία φίλτρα. Η συνάρτηση ενεργοποίησης είναι η ReLU παντού. Μπορεί κανείς προφανώς να παρατηρήσει την σημαντική διαφορά στον αριθμό των παραμέτρων. Πράγματι, το μοντέλο U-2-NET έχει συνολικά 822.370 παραμέτρους. Αυτό σημαίνει, εφόσον τα δύο δίκτυα είναι ίδια, ότι το κάθε μικρότερο νευρωνικό δίκτυο έχει 411.185 βάρη, δηλαδή λιγότερο από το μισό των μοντέλων U-NET και CNN. Θα δούμε τι μπορεί να πετύχει ένα τέτοιο μοντέλο (ειδικά σε σύγκριση με το μοντέλο U-NET) στο κεφάλαιο 7.

- **Το μοντέλο D(ense)-UNET**

Το μοντέλο αυτό είναι άλλη μία απόπειρα να σχεδιάσουμε ένα μοντέλο με σημαντικά

λιγότερες παραμέτρους. Η ιδέα είναι η εξής: Σε κάθε μπλοκ του προ-αναφερόμενου U-NET το νευρωνικό δίκτυο στην ουσία μαθαίνει τα χαρακτηριστικά των εικόνων, μέσω των συνελικτικών στρώματων. Αυτό που κάνουμε εδώ είναι να μειώσουμε τον αριθμό των φίλτρων σε κάθε συνελικτικό στρώμα και να προσθέσουμε, στο τέλος του κάθε μπλοκ, ένα πλήρως συνδεδεμένο, πυκνό (fully-connected, dense) στρώμα. Το σκεπτικό είναι ότι τα παραγόμενα από τις συνελίξεις χαρακτηριστικά θα επεξεργάζονται περαιτέρω από τα πυκνά στρώματα προς σχηματισμό ακόμα καλύτερων χαρακτηριστικών, χωρίς να υπάρχει ανάγκη να μάθει το δίκτυο έναν μεγάλο αριθμό φίλτρων.

Το νευρωνικό δίκτυο έχει και εδώ τέσσερα βασικά μπλοκ και φαίνεται στην εικόνα 44. Το κάθε ένα από αυτά έχει δύο στρώματα συνελίξεων (Padding=Same) που ακολουθείται από ένα πυκνό στρώμα και ένα Max Pooling στρώμα. Ο αριθμός των φίλτρων (μονάδων) σε κάθε συνελικτικό (πυκνό) στρώμα είναι 32 (32), 32 (32), 16 (32) και 16 (32). Ακολουθεί άλλα δύο στρώματα συνελίξεων με 8 φίλτρα το καθένα και ένα πυκνό στρώμα με 64 μονάδες. Τώρα ξεκινούν οι από-συνελίξεις. Το κάθε μπλοκ από-συνέλιξης περιέχει ένα στρώμα από-συνέλιξης, δύο στρώματα συνελίξεων και ένα πυκνό στρώμα. Ο αριθμός των φίλτρων των από-συνελίξεων (συνελίξεων) και των [μονάδων των πυκνών στρωμάτων] είναι 8 (8) [32], 16 (16) [32], 32 (32) [32] και 32 (32) [32]. Τελειώνουμε με ένα συνελικτικό στρώμα με τρία φίλτρα. Και εδώ έχουμε ως συνάρτηση ενεργοποίησης παντού την ReLU. Το πιο σημαντικό χαρακτηριστικό του μοντέλου D-UNET είναι ότι περιέχει μόνο 557.515 βάρη. Έχουμε δηλαδή μόνο ένα νευρωνικό δίκτυο που επεξεργάζεται ολόκληρες τις εικόνες εισόδου και έχει σχεδόν την μισή πολυπλοκότητα από το μοντέλο U-NET και CNN. Θα έχει ενδιαφέρον να δούμε τις διαφορές στην απόδοσή τους.

## 6.5 Πίνακας πειραμάτων

Τώρα που έχουμε όλα τα δομικά στοιχεία που θα χρησιμοποιήσουμε στα πειράματα, συνοψίζουμε όλους τους συνδυασμούς μοντέλων και συναρτήσεων κόστους και ομαλοποίησης που δοκιμάστηκαν στον παρακάτω πίνακα:

Μοντέλο	Κόστος + Ομαλοποίηση
U-NET	SSIM + $\mathcal{L}2$
	SSIM + Total Variation
	PCC + $\mathcal{L}2$
	PCC + Total Variation
	PCC_SSIM + $\mathcal{L}2$
	PCC_SSIM + Total Variation
CNN	SSIM + $\mathcal{L}2$
	SSIM + Total Variation
	PCC + $\mathcal{L}2$
	PCC + Total Variation
U-2-NET	SSIM + $\mathcal{L}2$
	PCC + $\mathcal{L}2$
	PCC + Total Variation
	PCC_SSIM + $\mathcal{L}2$
	Parzen M.I. + $\mathcal{L}2$
D-UNET	PCC_SSIM + $\mathcal{L}2$
	PCC + $\mathcal{L}2$
	Parzen M.I. + $\mathcal{L}2$
	Parzen M.I. + Total Variation

Πίνακας 9: Συνδυασμοί μοντέλων και ποινών που μελετήθηκαν

Το κάθε ένα από αυτά τα πειράματα χρειάζεται λίγες (από 2 μέχρι και 5) ημέρες για να πραγματοποιηθεί στον υπολογιστή τα χαρακτηριστικά του οποίου δίνονται στο κεφάλαιο 6.1.. Κατά την εκπαίδευση των μοντέλων παρακολουθούμε τις εξής μετρικές: Προφανώς καταγράφεται η πορεία της συνάρτησης ποινής (κόστος + ομαλοποίηση) κατά το πέρασμα των εποχών, τόσο στο σύνολο εκπαίδευσης όσο και στο σύνολο επικύρωσης. Επίσης καταγράφεται και ο δείκτης Dice, παρόλο που δεν αποτελεί μέρος της συνολικής ποινής.

Είναι γνωστό ότι ο δείκτης Dice έχει νόημα κυρίως για μάσκες κατάτμησης (segmentation masks), δηλαδή για εικόνες που έχουν τιμή 1 στα εικονοστοιχεία ενδιαφέροντος (π.χ. όγκοι και όργανα) και 0 αλλού. Παρόλα αυτά, επειδή όπως αναφέραμε στην περίπτωση μας οι μάσκες είναι κακής ποιότητας και δεν μπορούν να συμπεριληφθούν με αξιοπιστία στην εκπαίδευση και στον έλεγχο, μετράμε τον δείκτη Dice στις SUV εικόνες μιας και αναμένει κανείς μετά την ευθυγράμμιση των εικόνων να είναι μεγαλύτερος. Τα εικονοστοιχεία των εικόνων αυτών έχουν συνεχείς τιμές οπότε θα περίμενε κανείς ότι ο δείκτης Dice δεν έχει νόημα. Παρόλα αυτά υπάρχουν δημοσιεύσεις που ερευνούν τον υπολογισμό του δείκτη Dice και διάφορες παραλλαγές αυτού σε εικόνες με συνεχείς τιμές εικονοστοιχείων, παραπέμπουμε στην [49]. Στην προκειμένη περίπτωση θεωρήθηκε ενδιαφέρον να μετρηθεί ο κλασικός δείκτης Dice στις εικόνες SUV, εφόσον διαπιστώθηκε κιόλας ότι έχει μία καλά ορισμένη τιμή (μεταξύ του 0 και του 1) τόσο για τις κινούμενες εικόνες όσο και για τις

εικόνες στόχο. Ο τύπος που δίνει τον δείκτη Dice για δύο εικόνες με συνεχείς τιμές είναι ο εξής:

$$DSC(X, Y) = \frac{2|X \cap Y|}{|X| + |Y|} = \frac{\sum_i X_i Y_i}{\sum_i X_i + \sum_i Y_i}$$

Μετά την εκπαίδευση ενός μοντέλου το αξιολογούμε μέσω του συνόλου ελέγχου. Στο σύνολο ελέγχου μετράμε όλες τις μετρικές που μας ενδιαφέρουν πριν και μετά την ευθυγράμμιση των εικόνων που το απαρτίζουν για να κάνουμε την απαραίτητη σύγκριση.

## 7 Αποτελέσματα

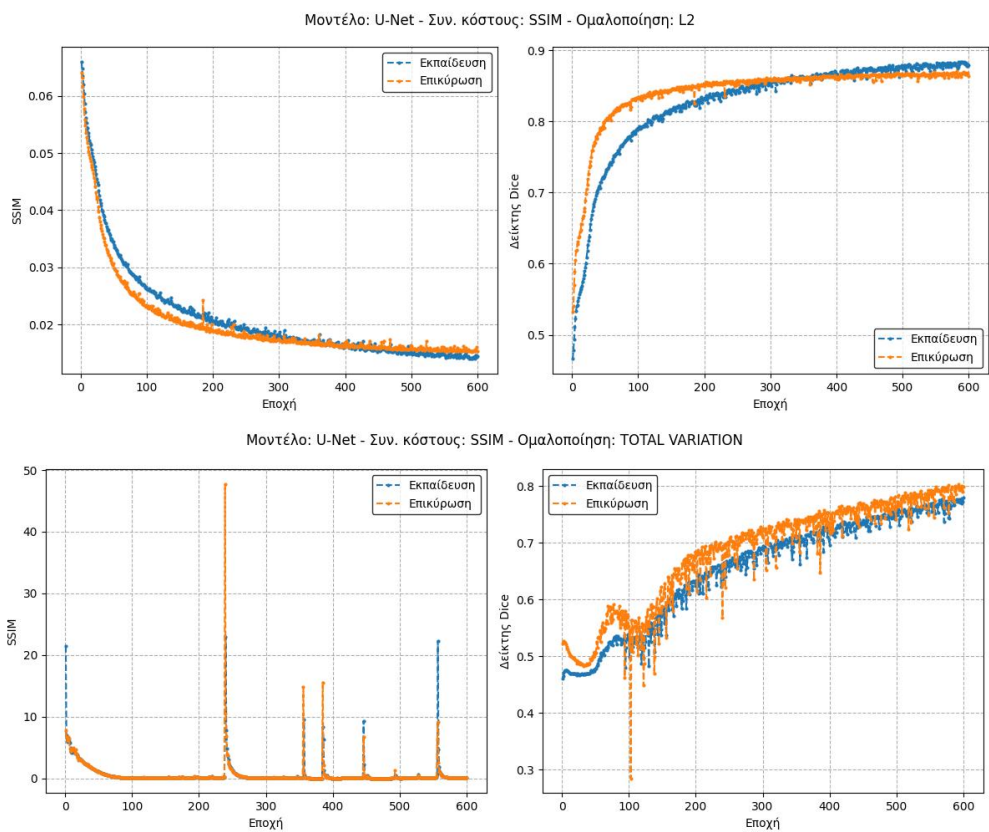
Στο κεφάλαιο αυτό παρουσιάζουμε τα ποιοτικά και ποσοτικά αποτελέσματα όλων των πειραμάτων που πραγματοποιήθηκαν. Αυτά περιλαμβάνουν τα εξής:

- Διαγράμματα του προφίλ της συνολικής ποινής και του δείκτη Dice στο σύνολο εκπαίδευσης και επικύρωσης, που θα μας αποδείξουν κατά πόσο η εκπαίδευση ήταν επιτυχής
- Πίνακα με τις μέσες μετρικές ομοιότητας (και την τυπική απόκλιση) στο σύνολο ελέγχου μετά την ευθυγράμμιση με το εκάστοτε μοντέλο
- Οπτική επαλήθευση της ευθυγράμμισης τυχαίων εικόνων του συνόλου ελέγχου τυπώνοντας τομές (slices) της κινούμενης εικόνας, της εικόνας στόχου, του πεδίου παραμόρφωσης και της παραμορφωμένης κινούμενης εικόνας

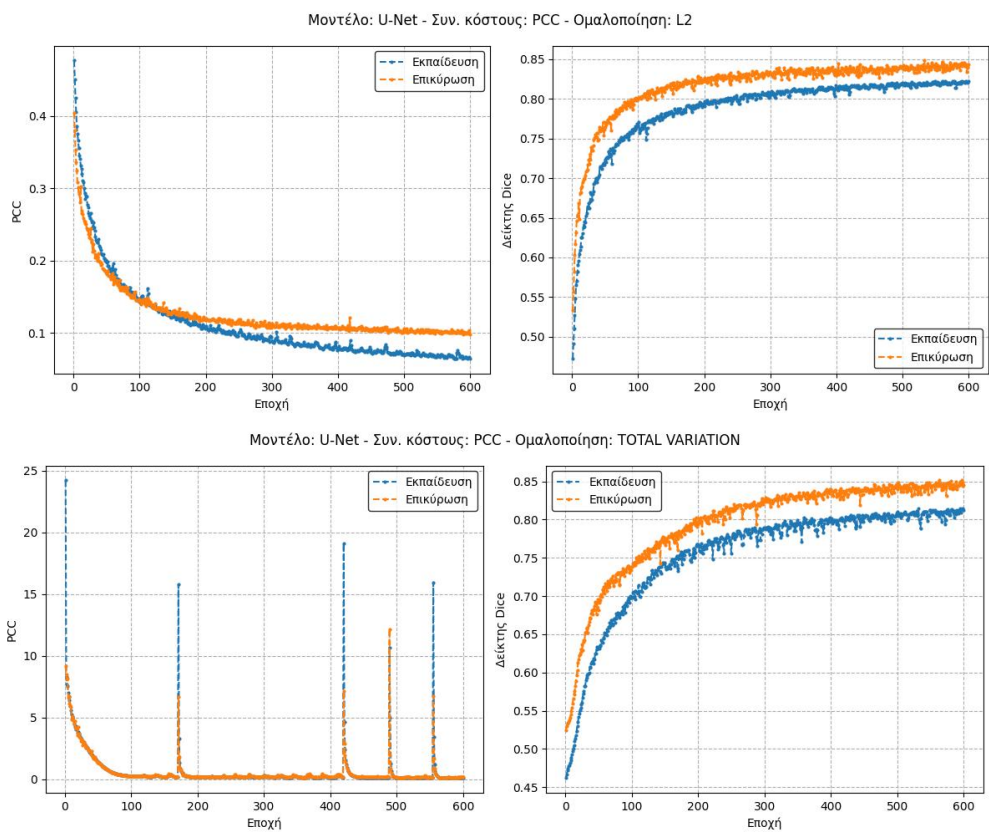
### 7.1 Αξιολόγηση της εκπαίδευσης των μοντέλων

#### Μοντέλο U-NET

Για το μοντέλο U-NET έγιναν αρχικά έξι πειράματα. Τα διαγράμματα της εκπαίδευσης φαίνονται στις παρακάτω τρεις εικόνες:

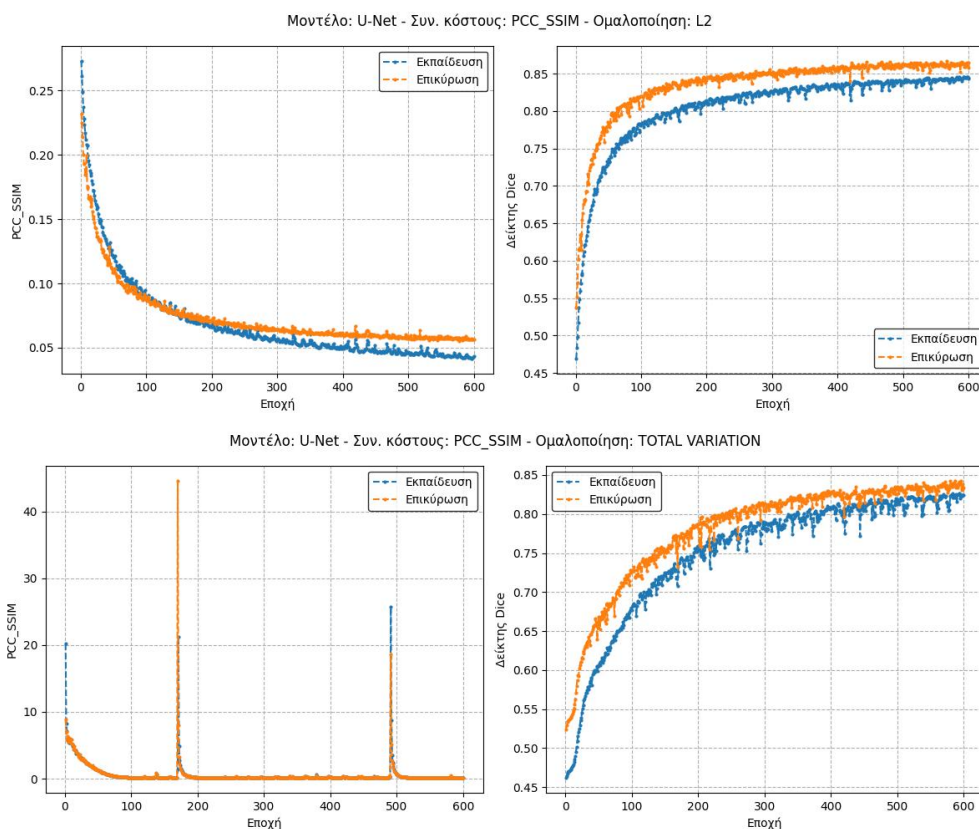


Εικόνα 24: Κόστος: SSIM, Ομαλοποίηση:  $\mathcal{L}_2$  / Total Variation



Εικόνα 25: Κόστος: PCC, Ομαλοποίηση:  $\mathcal{L}_2$  / Total Variation



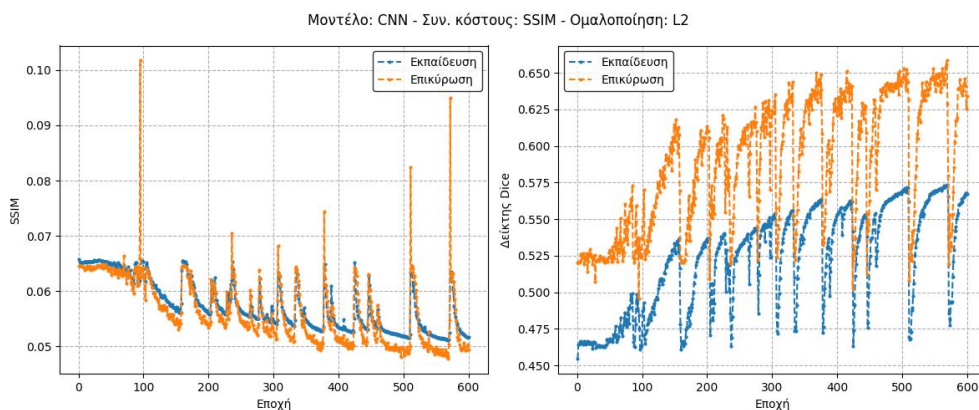


Εικόνα 26: Κόστος: PCC\_SSIM, Ομαλοποίηση:  $L2$  / Total Variation

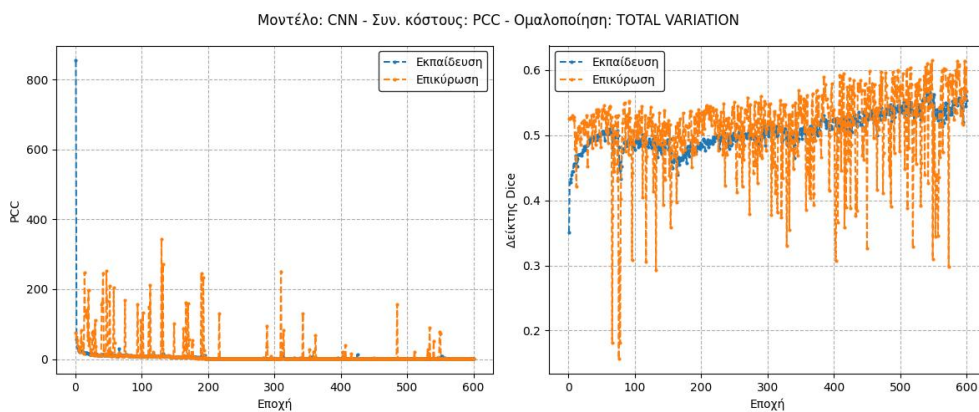
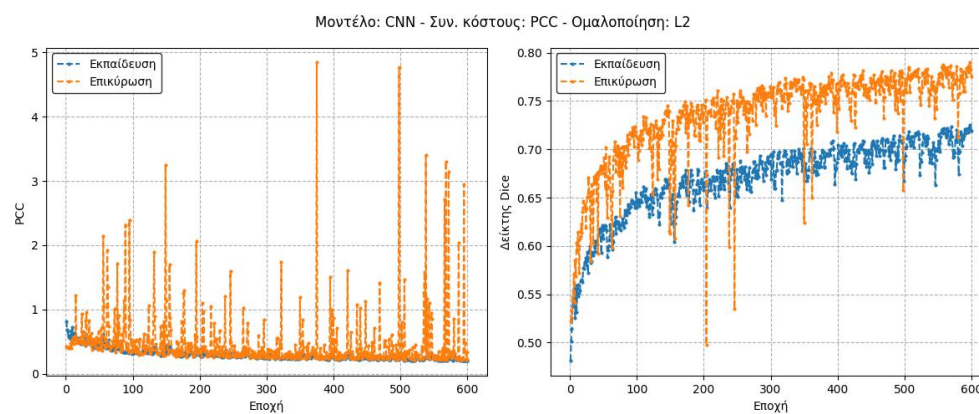
Παρατηρούμε ότι το κλασικό μοντέλο UNET πετυχαίνει με ικανοποιητικό τρόπο τον στόχο του όταν εκπαιδεύεται με οποιαδήποτε συνάρτηση κόστους και ομαλοποίηση τύπου  $L2$ . Περίεργες και ακανόνιστες συμπεριφορές παρουσιάζονται όταν για συνάρτηση ομαλοποίησης χρησιμοποιήσουμε την συνάρτηση ολικής απόκλισης. Στην περίπτωση αυτή εμφανίζονται ξαφνικά κατά την διάρκεια της εκπαίδευσης ορισμένες κορυφές που εκτινάσσουν την συνάρτηση κόστους στα ύψη. Φαίνεται δηλαδή ότι η συγκεκριμένη συνάρτηση ομαλοποίησης προκαλεί μία μικρή αστάθεια στη εκπαίδευση του μοντέλου, χωρίς αυτό να σημαίνει ότι δεν δίνει καλά αποτελέσματα στο τέλος. Άλλωστε φαίνεται ότι καθ'όλη την διάρκεια της εκπαίδευσης η συνάρτηση κόστους μειώνεται σχετικά ομαλά ενώ ο δείκτης Dice αυξάνεται. Συνολικά το μοντέλο UNET είναι ικανοποιητικό, όπως αναμέναμε άλλωστε.

## Μοντέλο CNN

Συνεχίζουμε με το μοντέλο CNN με τρία πειράματα:



Εικόνα 27: Κόστος: SSIM, Ομαλοποίηση:  $\mathcal{L}_2$



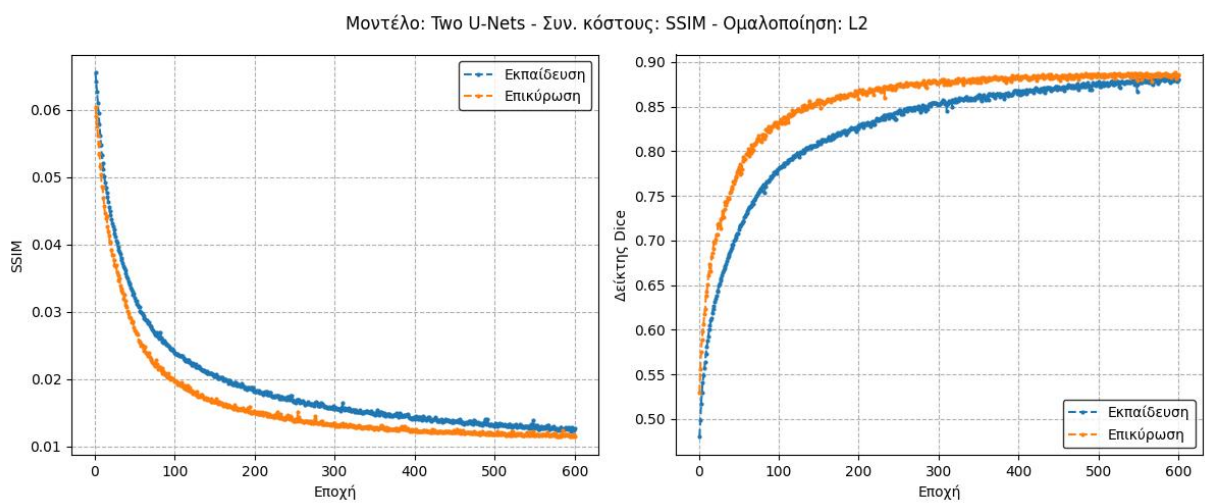
Εικόνα 28: Κόστος: PCC, Ομαλοποίηση:  $\mathcal{L}_2$  / Total Variation

Ο λόγος που εδώ έγιναν μόνο τρία πειράματα είναι τα παραπάνω διαγράμματα. Τα αποτελέσματα που δίνει το μοντέλο CNN είναι πολύ άσχημα και καθόλου ικανοποιητικά. Τόσο η συνάρτηση κόστους όσο και ο δείκτης Dice παρουσιάζουν μεγάλη αστάθεια και

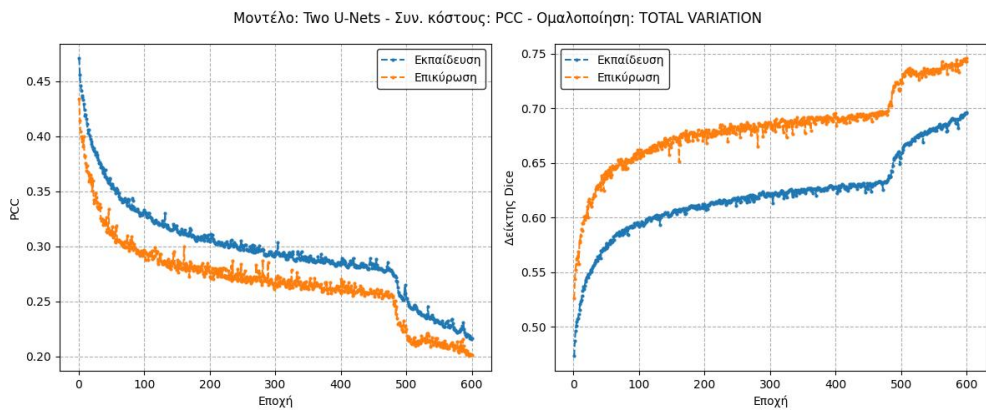
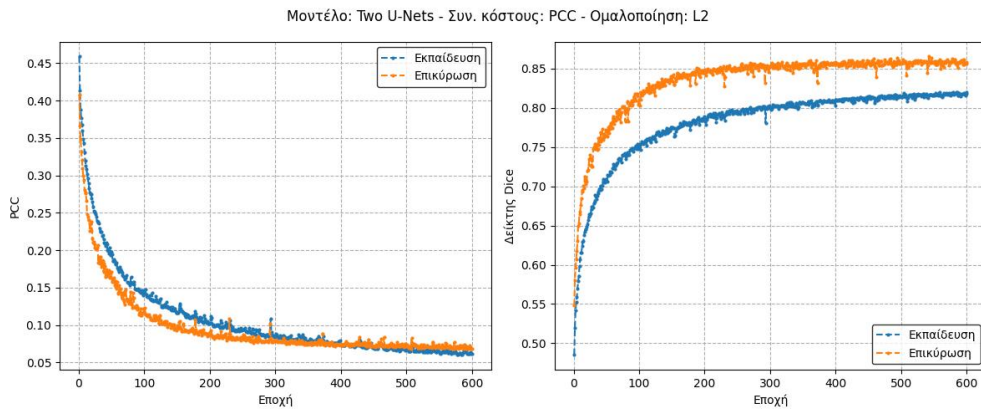
παρόλο που εμφανίζουν τις αναμενόμενες κατά μέσο όρο τάσεις (φθίνουσα για το κόστος και αύξουσα για το Dice) δεν μπορούν να θεωρηθούν αξιόπιστα και αποδεκτά αποτελέσματα. Φαίνεται λοιπόν ότι η δομή του UNET που μειώνει την διάσταση της εικόνας σε κάθε βήμα καθώς και η ανακατασκευή της είναι κρίσιμα σημεία για την ευστάθεια της εκπαίδευσης και την ποιότητα της ευθυγράμμισης. Με το σκεπτικό αυτό προχωρήσαμε στα επόμενα μοντέλα.

## Μοντέλο U-2-NET

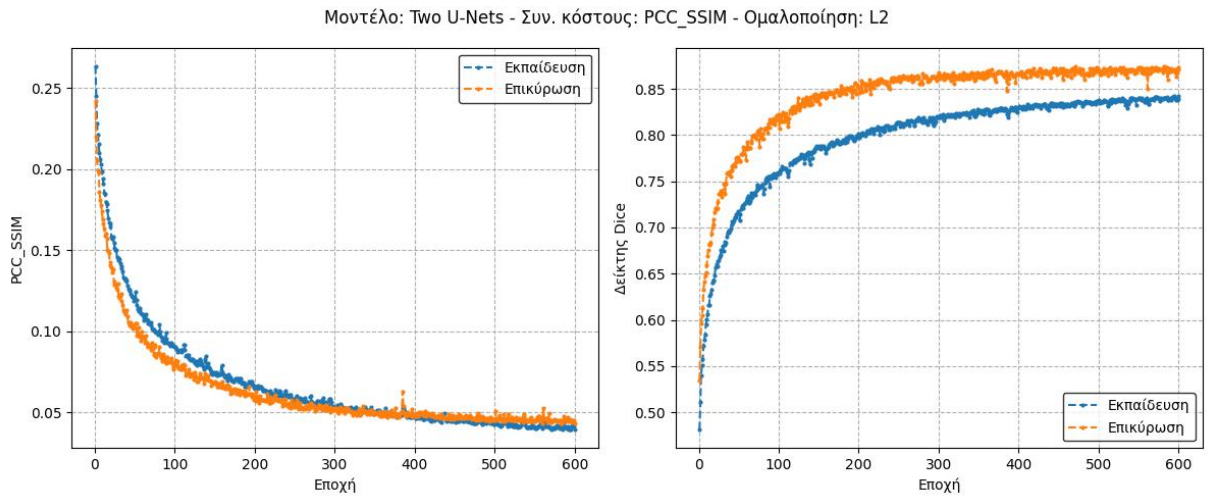
Μετά την αποτυχία του μοντέλου CNN, συνεχίζουμε με το U-2-NET, ελπίζοντας καλύτερα αποτελέσματα μιας και είναι βασισμένο στην δομή του U-NET:



Εικόνα 29: Κόστος: SSIM, Ομαλοποίηση:  $\mathcal{L}_2$

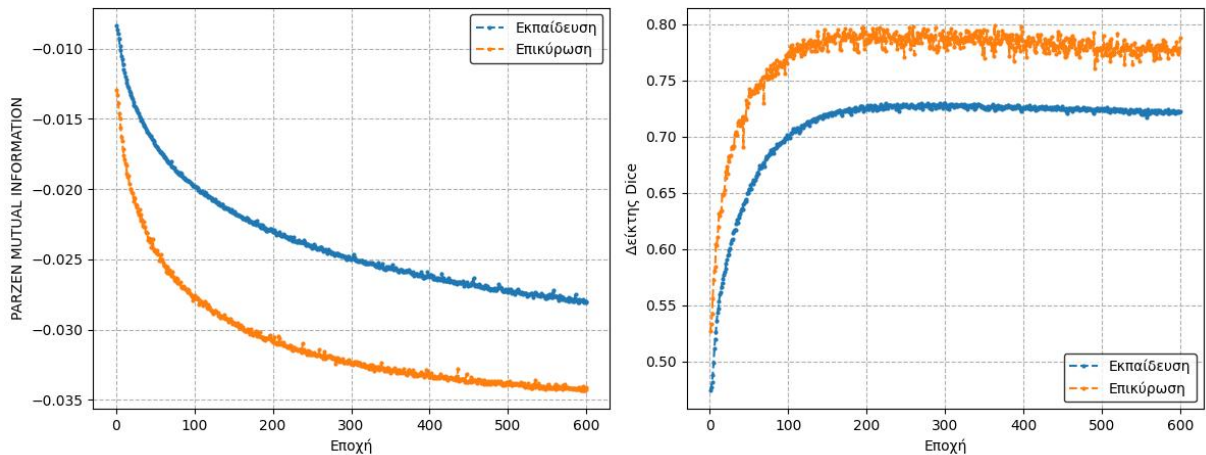


Εικόνα 30: Κόστος: PCC, Ομαλοποίηση:  $L_2$  / Total Variation



Εικόνα 31: Κόστος: PCC\_SSIM, Ομαλοποίηση:  $L_2$

Μοντέλο: Two U-Nets - Συν. κόστους: PARZEN MUTUAL INFORMATION - Ομαλοποίηση: L2



Εικόνα 32: Κόστος: Parzen M.I., Ομαλοποίηση:  $L_2$

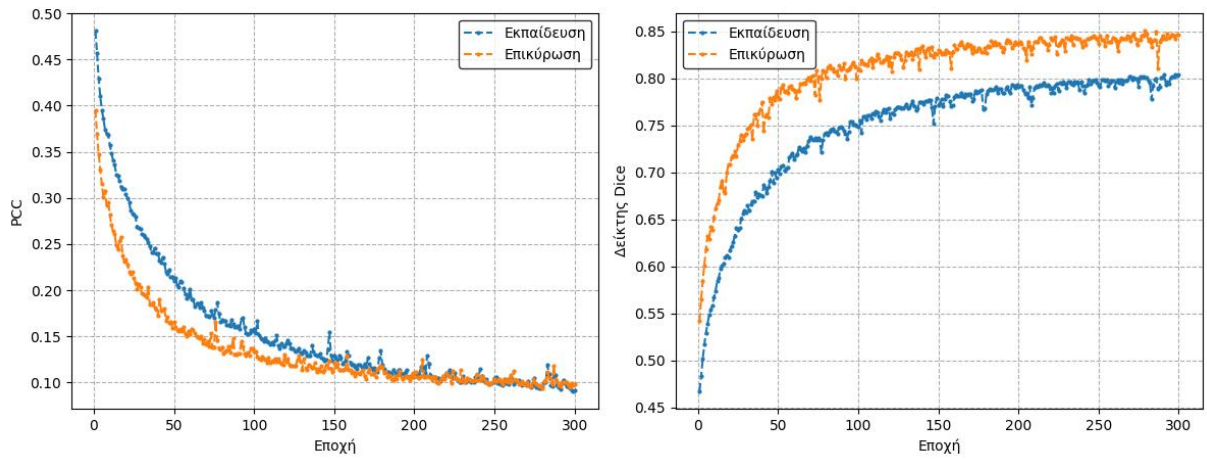
Τα αποτελέσματα αυτή τη φορά είναι σαφώς πολύ καλύτερα και, όπως θα δούμε και στους τελικούς πίνακες με τις μετρικές στην επόμενη υπό-ενότητα, είναι ακόμη καλύτερα από του UNET. Σε όλες σχεδόν τις περιπτώσεις η συνάρτηση κόστους έχει μια σταθερή, ομαλή φθίνουσα πορεία και ο δείκτης Dice μία αύξουσα σταθερή πορεία. Μοναδική εξαίρεση είναι το πείραμα με συνάρτηση κόστους την PCC και ομαλοποίηση την ολική απόκλιση. Φαίνεται ότι περίπου στην εποχή 480 η συνάρτηση κόστους παρουσιάζει μία απότομη μείωση που φυσικά είναι καλό σημάδι σε ό,τι αφορά την εκπαίδευση. Παρόλα αυτά η τελική τιμή της συνάρτησης κόστους δεν είναι τόσο χαμηλή όσο στην περίπτωση της  $L_2$  ομαλοποίησης. Ενδεχομένως να έχει ενδιαφέρον να τρέξουμε μερικά πειράματα με αυτόν τον συνδυασμό παραμέτρων για περισσότερες όμως εποχές.

Φαίνεται λοιπόν ότι ο σχεδιασμός δύο αρκετά μικρότερων νευρωνικών δικτύων που θα ευθυγραμμίζουν το καθένα την μισή εικόνα είναι μία καλή επιλογή. Βγάζει νόημα άλλωστε ότι όσο μικρότερη είναι η εικόνα (και άρα τόσο μικρότερη η πολυπλοκότητά της) τόσο πιο εύκολο θα είναι να ευθυγραμμιστεί. Βλέπουμε επίσης ότι παρόλο που το μέγεθος της εικόνας που εισέρχεται σε κάθε δίκτυο είναι το μισό της αρχικής, η πολυπλοκότητα (τα βάρη) του κάθε δικτύου είναι μικρότερη από το μισό του UNET. Συνεπώς η σχέση της πολυπλοκότητας του μοντέλου και του μεγέθους της εικόνας (για δεδομένη ποιότητα ευθυγράμμισης) δεν είναι ακριβώς γραμμική. Μπορούμε να κόψουμε την εικόνα στην μέση αλλά το νευρωνικό δίκτυο σε ακόμα μικρότερα κομμάτια και να πετύχουμε εξίσου καλή ή καλύτερα ευθυγράμμιση.

## Μοντέλο D-UNET

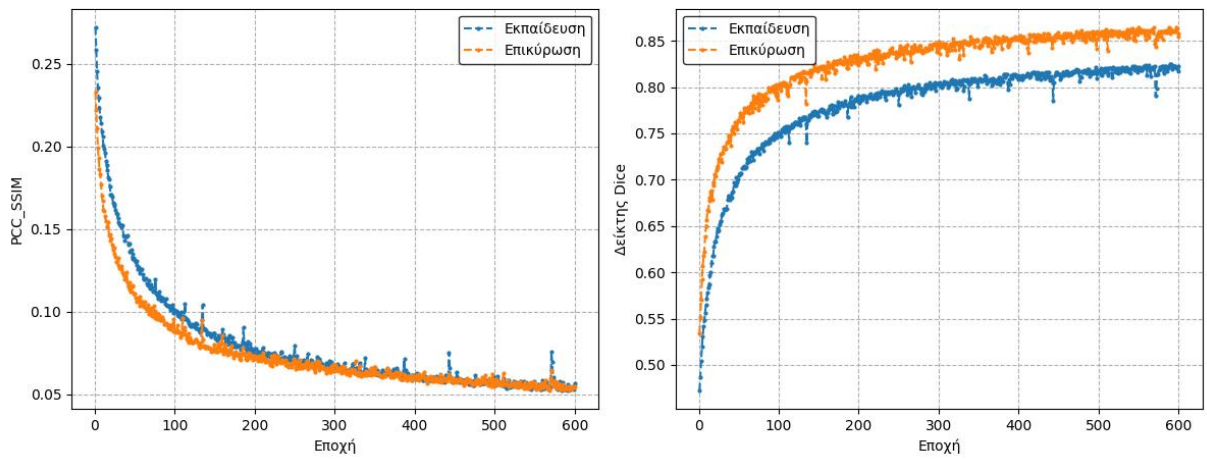
Τελειώνουμε με την αξιολόγηση του μοντέλου D-UNET στην εκπαίδευσή του.

Μοντέλο: Dense U-Net - Συν. κόστους: PCC - Ομαλοποίηση: L2

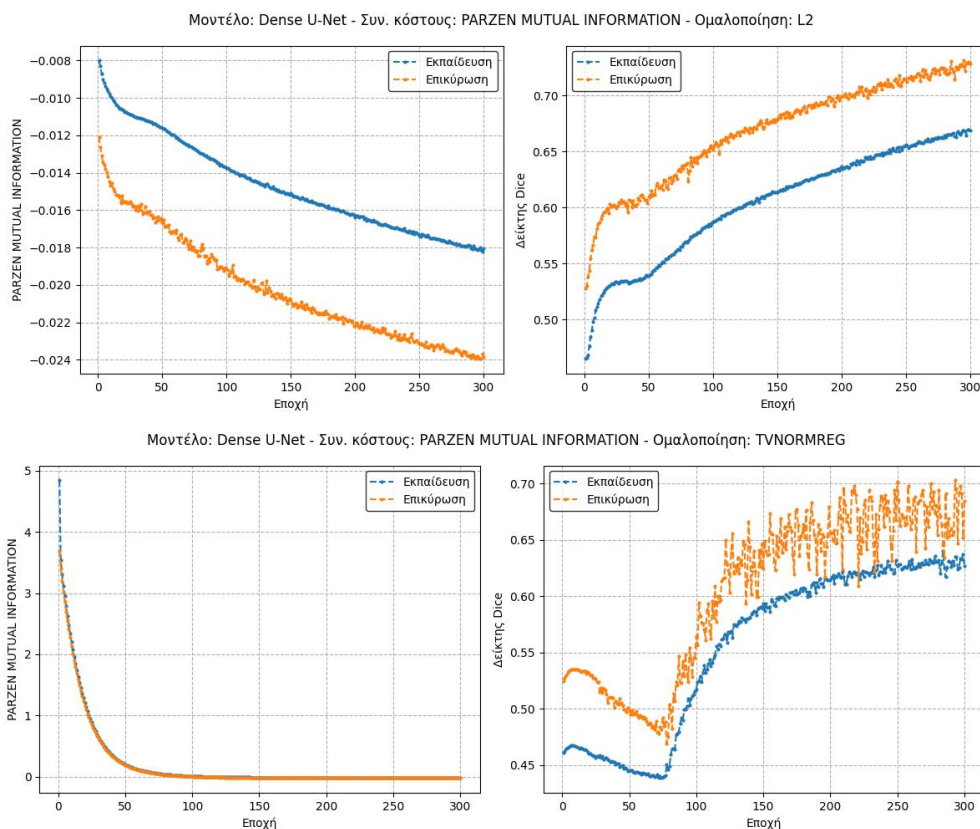


Εικόνα 33: Κόστος: PCC., Ομαλοποίηση:  $\mathcal{L}_2$

Μοντέλο: Dense U-Net - Συν. κόστους: PCC\_SSIM - Ομαλοποίηση: L2



Εικόνα 34: Κόστος: PCC\_SSIM., Ομαλοποίηση:  $\mathcal{L}_2$



Εικόνα 35: Κόστος: PCC\_SSIM, Ομαλοποίηση:  $L2$  / Total Variation

Όπως αναφέρθηκε και στην προηγούμενη υπό-ενότητα, για αυτό το μοντέλο τα πειράματα έτρεξαν για 300 εποχές (με εξαίρεση τον συνδυασμό PCC\_SSIM &  $L2$ ) λόγω χρονικού περιορισμού. Καταρχάς ας παρατηρήσουμε τον συνδυασμό PCC\_SSIM + Ολική Απόκλιση. Και εδώ (ομοίως με την περίπτωση του UNET) η συνάρτηση της ολικής απόκλισης προκαλεί μεγάλη αστάθεια στην εκπαίδευση και αυτό φαίνεται από τον δείκτη Dice. Στο διάγραμμα της συνάρτησης κόστους δεν φαίνεται καθόλου γιατί η αρχική τιμή είναι ασυνήθιστα μεγάλη οπότε η αστάθεια που παρουσιάζεται στις μικρότερες τιμές δεν είναι εμφανής λόγω της κλίμακας του κατακόρυφου άξονα. Στις υπόλοιπες περιπτώσεις η εκπαίδευση είναι σχετικά ομαλή και οι τελικές τιμές είναι εξίσου καλές με τα μοντέλα UNET και U-2-NET. Μάλιστα παρατηρούμε κάτι εξαιρετικά ενδιαφέρον: Στις εικόνες 33 και 34 βλέπουμε δύο πειράματα που διαφέρουν κατά 300 εποχές. Παρόλα αυτά ο τελικός δείκτης Dice είναι περίπου ο ίδιος. Συνεπώς, με βάση αυτή την μετρική, το μοντέλο που εκπαιδεύτηκε για 300 εποχές τα πήγε εξίσου καλά με το μοντέλο που εκπαιδεύτηκε για 600 εποχές. Αυτό μπορεί ενδεχομένως να σημαίνει ότι για την συγκεκριμένη εφαρμογή και το συγκεκριμένο μοντέλο, η συνάρτηση κόστους PCC με ομαλοποίηση  $L2$  είναι ένας πολύ καλός συνδυασμός για να εκπαιδευτεί το μοντέλο. Θα είχε ενδιαφέρον να αφήσουμε ένα τέτοιο μοντέλο να εκπαιδευτεί για 600 εποχές για να μπορέσουμε να έχουμε μία κοινή βάση σύγκρισης.

## 7.2 Αξιολόγηση των μοντέλων στο σύνολο ελέγχου

Προκειμένου να δούμε κατά πόσο το κάθε μοντέλο κατάφερε να ευθυγραμμίσει τις εικόνες του συνόλου ελέγχου, θα πρέπει να ξέρουμε ποια είναι η αφετηρία μας, δηλαδή ποια είναι η μέση τιμή των μετρικών ομοιότητας στο σύνολο ελέγχου πριν την ευθυγράμμιση. Οι τιμές αυτές δίνονται στον παρακάτω πίνακα:

Μετρική	Μέση τιμή	Τυπική απόκλιση
SSIM	0.942	0.0212
PCC	0.555	0.1431
MSE	$4.7 \cdot 10^{-4}$	$3.3 \cdot 10^{-4}$
Parzen M.I.	0.0938	0.0510
SUV Dice	0.5327	0.1518
SEG Dice	$9.5 \cdot 10^{-3}$	$3.47 \cdot 10^{-2}$

Πίνακας 10: Μετρικές ομοιότητας στο σύνολο ελέγχου πριν την ευθυγράμμιση

Από τις παραπάνω μετρικές, οι SSIM, PCC και Dice είναι κανονικοποιημένες, δηλαδή έχουν μέγιστη τιμή 1. Παρατηρούμε ότι η μετρική SSIM έχει ήδη τιμή αρκετά κοντά στο 1 ακόμα και πριν την ευθυγράμμιση. Αυτό πιθανότατα οφείλεται στην κανονικοποίηση που κάναμε ως προ-επεξεργασία, μιας και ο δείκτης αυτός συσχετίζει τις μέσες εντάσεις των εικόνων σε διάφορα παράθυρα των εικόνων. Εφόσον όλες οι εντάσεις έχουν παραπλήσια τιμή μετά την κανονικοποίηση, ο δείκτης αυτός προέκυψε αρκετά μεγάλος και χωρίς την επεξεργασία από τα μοντέλα. Αρκετά χαμηλό είναι επίσης και το μέσο τετραγωνικό σφάλμα των εικόνων. Θα δούμε κατά πόσο μπορούμε να βελτιώσουμε τις ήδη αυτές καλές μετρικές. Η αμοιβαία πληροφορία κατά Parzen ξεκινά αρκετά χαμηλή σε σχέση με τις υπόλοιπες μετρικές, αλλά θα πρέπει να θυμηθούμε ότι δεν είναι κανονικοποιημένη. Συνεπώς στόχος των μοντέλων δεν είναι να την κάνει να πλησιάζει όσο πιο κοντά γίνεται στο 1, αλλά να την αυξήσει κατά το δυνατόν περισσότερο σε σχέση με την αρχική τιμή. Τέλος, παρατηρούμε ότι ο μέσος δείκτης Dice στις μάσκες ελέγχου είναι πολύ κοντά στο 0 και αυτό φυσικά οφείλεται στην πολύ κακή ποιότητα των δεδομένων που έχουμε στην διάθεσή μας. Παρόλα αυτά θα υπολογίσουμε και τον μέσο δείκτη Dice με τα εκπαιδευμένα μας μοντέλα και θα δούμε κατά πόσο μπορεί να βελτιωθεί.

Στόχος των μοντέλων δεν είναι μόνο μία καλύτερη μέση τιμή της εκάστοτε μετρικής, αλλά και μία μικρότερη τυπική απόκλιση. Θα πρέπει να συνυπολογίσουμε όλες τις πιθανές επιδράσεις που έχει ένα μοντέλο στις μετρικές μας προκειμένου να επιλέξουμε τα καλύτερα. Θα σχηματίσουμε έναν πίνακα για κάθε μετρική ομοιότητας και θα υπολογίζουμε πόσο αλλάζει (και σε απόλυτη τιμή και σε % μεταβολή) μετά την ευθυγράμμιση με το κάθε μοντέλο. Για κάθε έναν τέτοιο πίνακα θα επιλέγουμε τα 3-4 καλύτερα μοντέλα έτσι ώστε τελικά να



αποφασίσουμε συνολικά ποια είχαν την καλύτερη επίδοση σε όλες τις μετρικές. Η επιλογή των μοντέλων θα γίνεται κάθε φορά όχι μόνο με βάση της επίδοσή τους, αλλά και με βάση την πολυπλοκότητά τους. Μεταξύ δύο μοντέλων που έχουν, για παράδειγμα, παραπλήσια επίδραση στον δείκτη PCC, θα επιλέξουμε εκείνο με την μικρότερη πολυπλοκότητα, δηλαδή εκείνο με τις λιγότερες παραμέτρους-βάρη. Ας ξεκινήσουμε να βλέπουμε τους προαναφερόμενους πίνακες με τον ήδη ανεβασμένο δείκτη SSIM:

Μετρική (Αρχική Τιμή)	Μοντέλο + Κόστος + Ομαλοποίηση	Τελική Μέση Τιμή	Διαφορά	% Μεταβολή
SSIM (0.942 ± 0.0212)	U-NET + SSIM + $\mathcal{L}2$	0.988 ± 0.0063	0.0453	+4.8
	U-NET + SSIM + T.V.	0.9822 ± 0.0072	0.0395	+4.19
	U-NET + PCC_SSIM + $\mathcal{L}2$	0.9836 ± 0.0085	0.0409	+4.3
	U-NET + PCC_SSIM + T.V.	0.9792 ± 0.0083	0.0365	+ 3.88
	CNN + SSIM + $\mathcal{L}2$	0.9574 ± 0.017	0.0148	+1.56
	<b>U-2-NET + SSIM + <math>\mathcal{L}2</math></b>	<b>0.9901 ± 0.0054</b>	<b>0.0474</b>	<b>+5.03</b>
	U-2-NET + PCC_SSIM + $\mathcal{L}2$	0.9840 ± 0.0085	0.0413	+4.37
	<b>D-UNET + PCC_SSIM + <math>\mathcal{L}2</math></b>	<b>0.9808 ± 0.0095</b>	<b>0.0381</b>	<b>+4.04</b>
	D-UNET + PCC + $\mathcal{L}2$	0.9735 ± 0.012	0.0308	+3.27
D-UNET + Parzen M.I. + $\mathcal{L}2$	0.9492 ± 0.019	0.0065	+0.685	

Πίνακας 11: Επίδραση διαφόρων μοντέλων στην μετρική ομοιότητας SSIM

Μετρική (Αρχική Τιμή)	Μοντέλο + Κόστος + Ομαλοποίηση	Τελική Μέση Τιμή	Διαφορά	% Μεταβολή
PCC (0.5327 ± 0.1518)	U-NET + PCC + $\mathcal{L}2$	0.89 ± 0.10	0.3351	+60.37
	U-NET + PCC + T.V.	0.8842 ± 0.099	0.3291	+59.29
	U-NET + PCC_SSIM + $\mathcal{L}2$	0.8951 ± 0.0983	0.340	+61.26
	U-NET + PCC_SSIM + T.V.	0.8735 ± 0.098	0.3184	+57.36
	<b>U-2-NET + SSIM + <math>\mathcal{L}2</math></b>	<b>0.907 ± 0.0858</b>	<b>0.3523</b>	<b>+63.48</b>
	U-2-NET + PCC + $\mathcal{L}2$	0.9003 ± 0.1015	0.3452	+62.19
	U-2-NET + PCC + T.V.	0.778 ± 0.112	0.222	+40.06
	U-2-NET + PCC_SSIM + $\mathcal{L}2$	0.905 ± 0.096	0.35	+63.08
	U-2-NET + Parzen M.I. + $\mathcal{L}2$	0.8153 ± 0.094	0.26	+46.88
	<b>D-UNET + PCC_SSIM + <math>\mathcal{L}2</math></b>	<b>0.8971 ± 0.1006</b>	<b>0.342</b>	<b>+61.61</b>
	D-UNET + PCC + $\mathcal{L}2$	0.8862 ± 0.1009	0.331	+59.69
D-UNET + Parzen M.I. + $\mathcal{L}2$	0.779 ± 0.114	0.224	+40.4	

Πίνακας 12: Επίδραση διαφόρων μοντέλων στην μετρική ομοιότητας PCC

Μετρική (Αρχική Τιμή $\cdot 10^{-4}$ )	Μοντέλο + Κόστος + Ομαλοποίηση	Τελική Μέση Τιμή $\cdot 10^{-4}$	Διαφορά $\cdot 10^{-4}$	% Μεταβολή
MSE ( $4.7 \pm 3.3$ )	U-NET + SSIM + $\mathcal{L}2$	$0.83 \pm 0.65$	3.9	-82.60
	U-NET + SSIM + T.V.	$1.2 \pm 0.77$	3.5	-75.09
	U-NET + PCC + $\mathcal{L}2$	$0.886 \pm 0.693$	3.88	-81.40
	U-NET + PCC + T.V.	$0.885 \pm 0.576$	3.88	-81.43
	U-NET + PCC_SSIM + $\mathcal{L}2$	$0.782 \pm 0.570$	3.895	-83.60
	U-NET + PCC_SSIM + T.V.	$0.943 \pm 0.609$	3.823	-80.21
	CNN + SSIM + $\mathcal{L}2$	$2.4 \pm 1.8$	1.3	-72.34
	<b>U-2-NET + SSIM + <math>\mathcal{L}2</math></b>	<b><math>0.661 \pm 0.562</math></b>	<b>4.105</b>	<b>-86.12</b>
	U-2-NET + PCC + $\mathcal{L}2$	$0.807 \pm 0.609$	3.96	-83.06
	U-2-NET + PCC + T.V.	$1.78 \pm 1.109$	2.984	-62.60
	<b>U-2-NET + PCC_SSIM + <math>\mathcal{L}2</math></b>	<b><math>0.742 \pm 0.566</math></b>	<b>4.02</b>	<b>-84.42</b>
	U-2-NET + Parzen M.I. + $\mathcal{L}2$	$2.54 \pm 1.53$	2.22	-46.67
	D-UNET + PCC_SSIM + $\mathcal{L}2$	$0.803 \pm 0.603$	3.964	-83.16
	D-UNET + PCC + $\mathcal{L}2$	$0.875 \pm 0.476$	3.825	-81.38
D-UNET + Parzen M.I. + $\mathcal{L}2$	$2.403 \pm 1.602$	2.297	-48.87	

Πίνακας 13: Επίδραση διαφόρων μοντέλων στην μετρική ομοιότητας MSE

Μετρική (Αρχική Τιμή)	Μοντέλο + Κόστος + Ομαλοποίηση	Τελική Μέση Τιμή	Διαφορά	% Μεταβολή
Parzen M.I. ( $0.0938 \pm 0.0510$ )	U-NET + SSIM + $\mathcal{L}2$	$0.1525 \pm 0.081$	0.0587	+62.64
	U-NET + SSIM + T.V.	$0.1390 \pm 0.0746$	0.0453	+48.29
	U-NET + PCC + $\mathcal{L}2$	$0.1365 \pm 0.074$	0.0428	+45.63
	U-NET + PCC + T.V.	$0.1327 \pm 0.0727$	0.0389	+41.54
	U-NET + PCC_SSIM + $\mathcal{L}2$	$0.1421 \pm 0.077$	0.04834	+51.54
	U-NET + PCC_SSIM + T.V.	$0.1359 \pm 0.073$	0.0421	+44.89
	CNN + SSIM + $\mathcal{L}2$	$0.11 \pm 0.06$	0.0159	+16.88
	<b>U-2-NET + SSIM + <math>\mathcal{L}2</math></b>	<b><math>0.1628 \pm 0.0831</math></b>	<b>0.0638</b>	<b>+64.52</b>
	U-2-NET + PCC + $\mathcal{L}2$	$0.1392 \pm 0.073$	0.040	+40.67
	U-2-NET + PCC + T.V.	$0.1238 \pm 0.063$	0.025	+25.15
	U-2-NET + PCC_SSIM + $\mathcal{L}2$	$0.1473 \pm 0.076$	0.0483	+48.86
	U-2-NET + Parzen M.I. + $\mathcal{L}2$	$0.1482 \pm 0.0738$	0.0493	+49.80
	<b>D-UNET + PCC_SSIM + <math>\mathcal{L}2</math></b>	<b><math>0.1450 \pm 0.0745</math></b>	<b>0.0461</b>	<b>+46.56</b>
	D-UNET + PCC + $\mathcal{L}2$	$0.1368 \pm 0.0708$	0.0378	+38.26
D-UNET + Parzen M.I. + $\mathcal{L}2$	$0.1227 \pm 0.062$	0.0237	+24.02	

Πίνακας 14: Επίδραση διαφόρων μοντέλων στην μετρική ομοιότητας Mutual Information

Μετρική (Αρχική Τιμή)	Μοντέλο + Κόστος + Ομαλοποίηση	Τελική Μέση Τιμή	Διαφορά	% Μεταβολή
Dice (SUV) (0.555 ± 0.143)	U-NET + SSIM + $\mathcal{L}2$	0.865 ± 0.098	0.3324	+62.40
	U-NET + SSIM + T.V.	0.798 ± 0.11	0.2653	+49.81
	U-NET + PCC + $\mathcal{L}2$	0.850 ± 0.118	0.3172	+59.54
	U-NET + PCC + T.V.	0.846 ± 0.116	0.3136	+58.86
	U-NET + PCC_SSIM + $\mathcal{L}2$	0.864 ± 0.110	0.3313	+62.19
	U-NET + PCC_SSIM + T.V.	0.840 ± 0.113	0.3076	+57.74
	CNN + SSIM + $\mathcal{L}2$	0.654 ± 0.17	0.1208	+22.68
	<b>U-2-NET + SSIM + <math>\mathcal{L}2</math></b>	<b>0.893 ± 0.0963</b>	<b>0.3602</b>	<b>+67.61</b>
	U-2-NET + PCC + $\mathcal{L}2$	0.852 ± 0.1215	0.3188	+59.85
	U-2-NET + PCC + T.V.	0.746 ± 0.1213	0.2133	+40.03
	U-2-NET + PCC_SSIM + $\mathcal{L}2$	0.870 ± 0.1121	0.337	+63.29
	U-2-NET + Parzen M.I. + $\mathcal{L}2$	0.769 ± 0.1077	0.2364	+44.38
	<b>D-UNET + PCC_SSIM + <math>\mathcal{L}2</math></b>	<b>0.857 ± 0.1161</b>	<b>0.3245</b>	<b>+60.92</b>
	D-UNET + PCC + $\mathcal{L}2$	0.849 ± 0.1206	0.3163	+59.38
D-UNET + Parzen M.I. + $\mathcal{L}2$	0.746 ± 0.1335	0.2128	+39.95	

Πίνακας 15: Επίδραση διαφόρων μοντέλων στην μετρική ομοιότητας Dice (SUV)

Μετρική (Αρχική Τιμή)	Μοντέλο + Κόστος + Ομαλοποίηση	Τελική Μέση Τιμή	Διαφορά	% Μεταβολή
Dice (SEG) ( $9.5 \cdot 10^{-3}$ $\pm 3.47 \cdot 10^{-2}$ )	U-NET + SSIM + $\mathcal{L}2$	0.071 ± 0.143	0.061	+642.11
	U-NET + SSIM + T.V.	0.032 ± 0.079	0.022	+234.97
	U-NET + PCC + $\mathcal{L}2$	0.189 ± 0.315	0.179	+1886.95
	U-NET + PCC + T.V.	0.272 ± 0.389	0.2632	+2766.08
	U-NET + PCC_SSIM + $\mathcal{L}2$	0.181 ± 0.2978	0.1714	+1801.47
	U-NET + PCC_SSIM + T.V.	0.2935 ± 0.3915	0.2840	+2984.86
	CNN + SSIM + $\mathcal{L}2$	0.014 ± 0.05	0.0042	+44.13
	U-2-NET + SSIM + $\mathcal{L}2$	0.085 ± 0.1466	0.076	+796.05
	U-2-NET + PCC + $\mathcal{L}2$	0.168 ± 0.28	0.1582	+1663.31
	U-2-NET + PCC + T.V.	0.1687 ± 0.3541	0.1591	+1672.96
	U-2-NET + PCC_SSIM + $\mathcal{L}2$	0.1405 ± 0.2248	0.1310	+1377.31
	U-2-NET + Parzen M.I. + $\mathcal{L}2$	0.0787 ± 0.1684	0.069	+727.58
	D-UNET + PCC_SSIM + $\mathcal{L}2$	0.1926 ± 0.329	0.1830	+1923.90
	D-UNET + PCC + $\mathcal{L}2$	0.2405 ± 0.3498	0.2310	+2427.63
D-UNET + Parzen M.I. + $\mathcal{L}2$	0.086 ± 0.1835	0.0763	+802.76	

Πίνακας 16: Επίδραση διαφόρων μοντέλων στην μετρική ομοιότητας Dice (μάσκες)

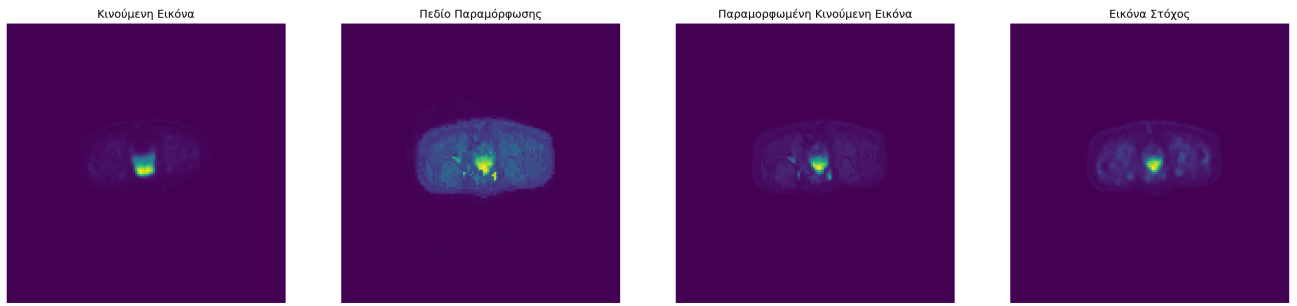
Πριν προχωρήσουμε στην αξιολόγηση θα παρατηρήσουμε το εξής: Στους παραπάνω πίνακες φαίνεται ότι λείπουν κάποια πειράματα που είδαμε στην προηγούμενη υπό-ενότητα όπου

αξιολογούσαμε την εκπαίδευση των μοντέλων. Αυτό οφείλεται στο εξής ατυχές γεγονός: Σε όλα τα πειράματα, στο τέλος, προφανώς έπρεπε να σωθεί το μοντέλο σε ένα αρχείο τύπου .h5 έτσι ώστε αυτό να φορτωθεί σε μεταγενέστερο χρόνο για ανάλυση και αξιολόγηση. Παρόλα αυτά, σε κάποιες περιπτώσεις (όπως στο CNN αλλά και στο D-UNET) τα αρχεία .h5 δεν ήταν δυνατόν να φορτωθούν και να διαβασθούν από το TensorFlow, λόγω του σφάλματος "Bad Marshal Code". Στην ουσία αυτό σημαίνει ότι τα αρχεία αυτά ήταν κατά κάποιο τρόπο αλλοιωμένα. Δοκιμάστηκαν διαφορετικές εκδόσεις Python και TensorFlow καθώς επίσης μελετήθηκε αναλυτικά το documentation του TensorFlow για εύρεση πιθανής λύσης. Παρόλα αυτά δεν βρέθηκε λύση. Επειδή τα αποτελέσματα της εκπαίδευσης του CNN δεν ήταν ικανοποιητικά (και επίσης η εκπαίδευση του CNN απεδείχθη ιδιαίτερα χρονοβόρα σε σχέση με τα άλλα μοντέλα) τα πειράματα δεν επαναλήφθηκαν. Το ίδιο πρόβλημα παρουσιάστηκε και στο μοντέλο D-UNET, αλλά επειδή το πρώτο πείραμα ήταν πολύ ικανοποιητικό, αποφασίσαμε να ξανατρέξουμε τρία πειράματα από την αρχή. Μετά από αυτήν την παρατήρηση συνεχίζουμε με την αξιολόγηση.

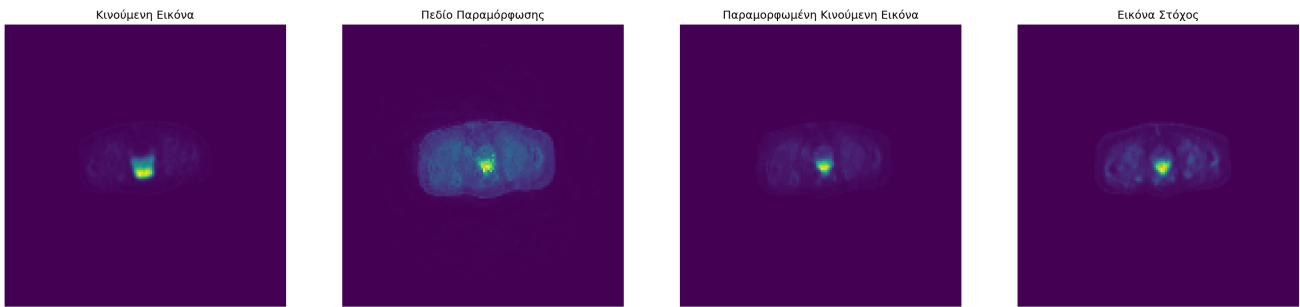
Σε κάθε έναν από τους παραπάνω πίνακες πηγαίνουμε και διαβάζουμε την κάθε γραμμή για να δούμε ποιο μοντέλο παρουσιάζει την μεγαλύτερη κατά απόλυτη τιμή % μεταβολή. Επίσης ελέγχουμε και κατά πόσο η τυπική απόκλιση της κάθε μετρικής μειώθηκε με το κάθε μοντέλο. Παρατηρούμε ότι μετά την ευθυγράμμιση όχι μόνο η μέση τιμή της μετρικής ομοιότητας αυξάνεται αλλά στις περισσότερες περιπτώσεις μειώνεται και η τυπική απόκλιση, πράγμα φυσικά επιθυμητό για ένα μοντέλο. Ο πίνακας που σχετίζεται με τις μάσκες (πίνακας 16) δεν χρησιμοποιήθηκε για την αξιολόγηση των μοντέλων, μιας και τα νούμερα δεν είναι ούτε ποιοτικά ούτε αξιόπιστα για τους λόγους που έχουμε ήδη αναφέρει. Ο παραπάνω πίνακας 16 παρατέθηκε καθαρά για λόγους πληρότητας. Με βάση αυτά, βλέπουμε ότι τα δύο καλύτερα μοντέλα (με βάση τόσο τις μετρικές αλλά και την συνολική πολυπλοκότητα) είναι το U-2-NET με κόστος SSIM και ομαλοποίηση L2 και το D-UNET με κόστος PCC\_SSIM και ομαλοποίηση L2. Το μοντέλο U-2-NET είναι σχεδόν σε κάθε πίνακα το βέλτιστο μοντέλο, σε αντίθεση με το D-UNET που υστερεί ελάχιστα στις μετρικές αλλά κερδίζει στο κριτήριο της πολυπλοκότητας. Στην επόμενη υπό-ενότητα θα χρησιμοποιήσουμε τα δύο αυτά μοντέλα για να κάνουμε μία σύντομη οπτικοποίηση της ευθυγράμμισης πάνω σε τομές (slices) τυχαίων ζευγαριών (κινούμενων-στόχων) εικόνων. Αυτή θα είναι και η απόλυτη επιβεβαίωση ότι τα μοντέλα έχουν πετύχει τον στόχο τους, μιας και οι μετρικές συνήθως δεν είναι εύκολα κατανοητές και ερμηνεύσιμες.

### 7.3 Οπτική επαλήθευση της ευθυγράμμισης στο σύνολο ελέγχου

Για τις εικόνες που ακολουθούν επιλέχθηκε ένας τυχαίος ασθενής από το σύνολο ελέγχου που, θυμίζουμε, αποτελείται από ένα ή περισσότερα ζεύγη εικόνων (κινούμενη και στόχο). Επιλέγουμε ένα τέτοιο ζεύγος και απεικονίζουμε σε μία γραμμή την τετράδα: Κινούμενη Εικόνα - Πεδίο Παραμόρφωσης - Παραμορφωμένη κινούμενη εικόνα - Εικόνα Στόχος. Βλέπουμε ότι το πεδίο παραμόρφωσης είναι και αυτό μία εικόνα, αλλά δεν μας λείπει πολλά από μόνο του. Με τον έλεγχο αυτό θέλουμε να δούμε (χυριολεκτικά με το μάτι) κατά πόσο η παραμορφωμένη κινούμενη εικόνα είναι παρόμοια με την εικόνα στόχο.

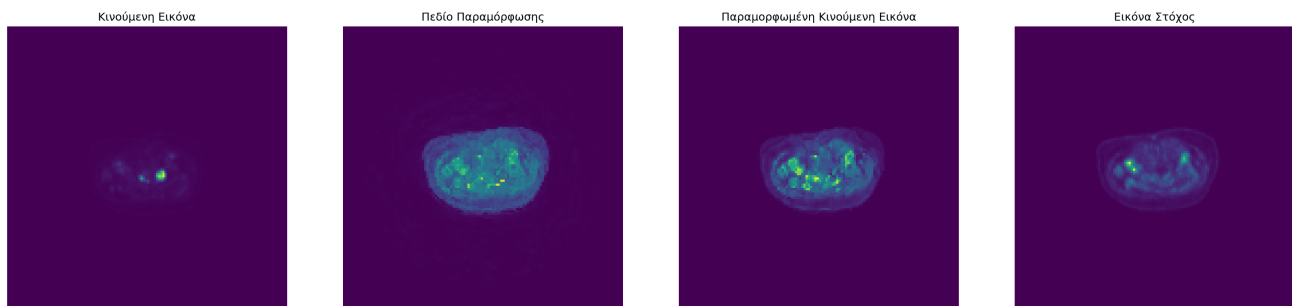


(a) Μοντέλο: U-2-NET

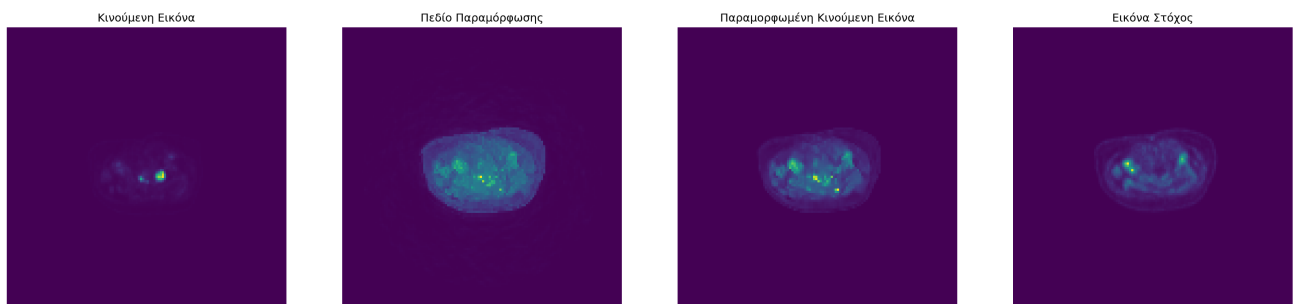


(b) Μοντέλο: D-UNET

Εικόνα 36: Σύγκριση της ευθυγράμμισης των μοντέλων (τομή 20)

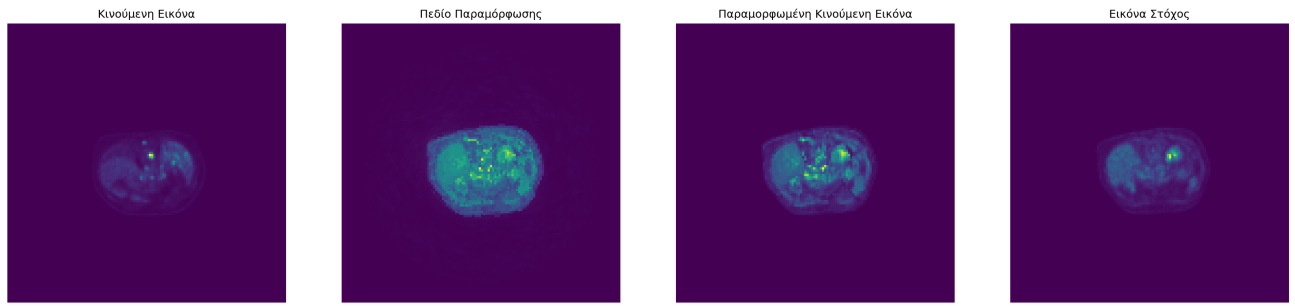


(a) Μοντέλο: U-2-NET

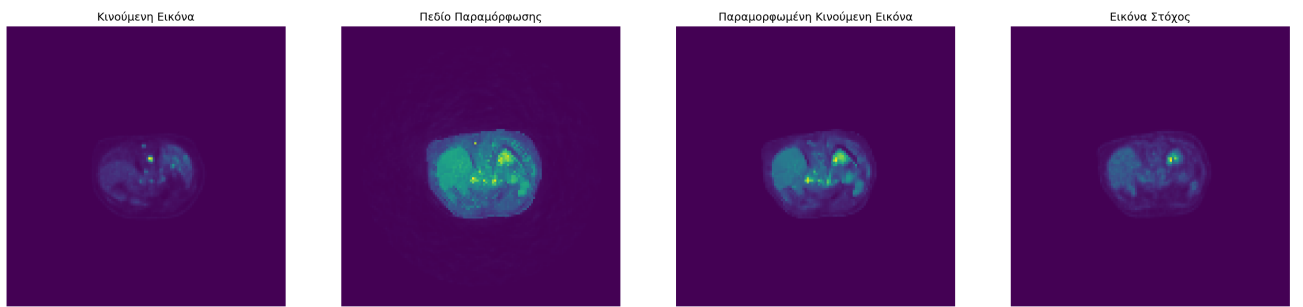


(b) Μοντέλο: D-UNET

Εικόνα 37: Σύγκριση της ευθυγράμμισης των μοντέλων (τομή 54)

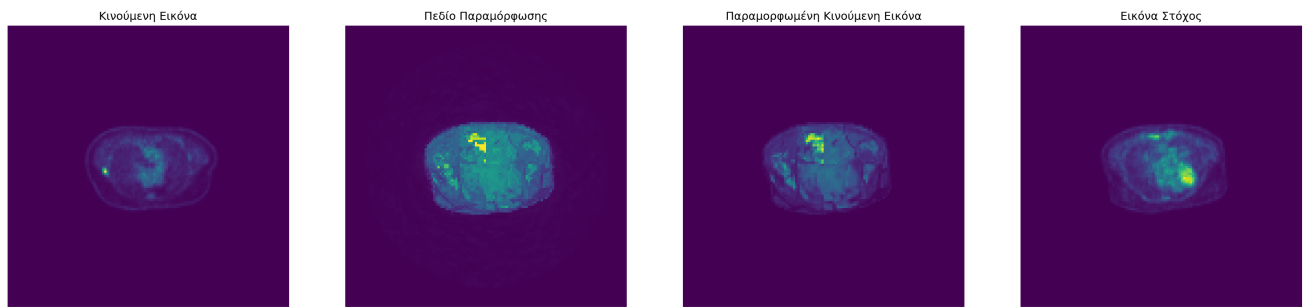


(a) Μοντέλο: U-2-NET

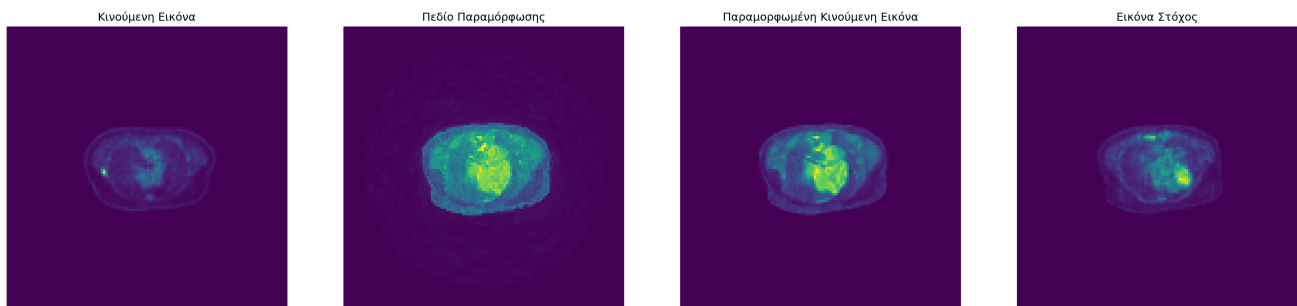


(b) Μοντέλο: D-UNET

Εικόνα 38: Σύγκριση της ευθυγράμμισης των μοντέλων (τομή 65)



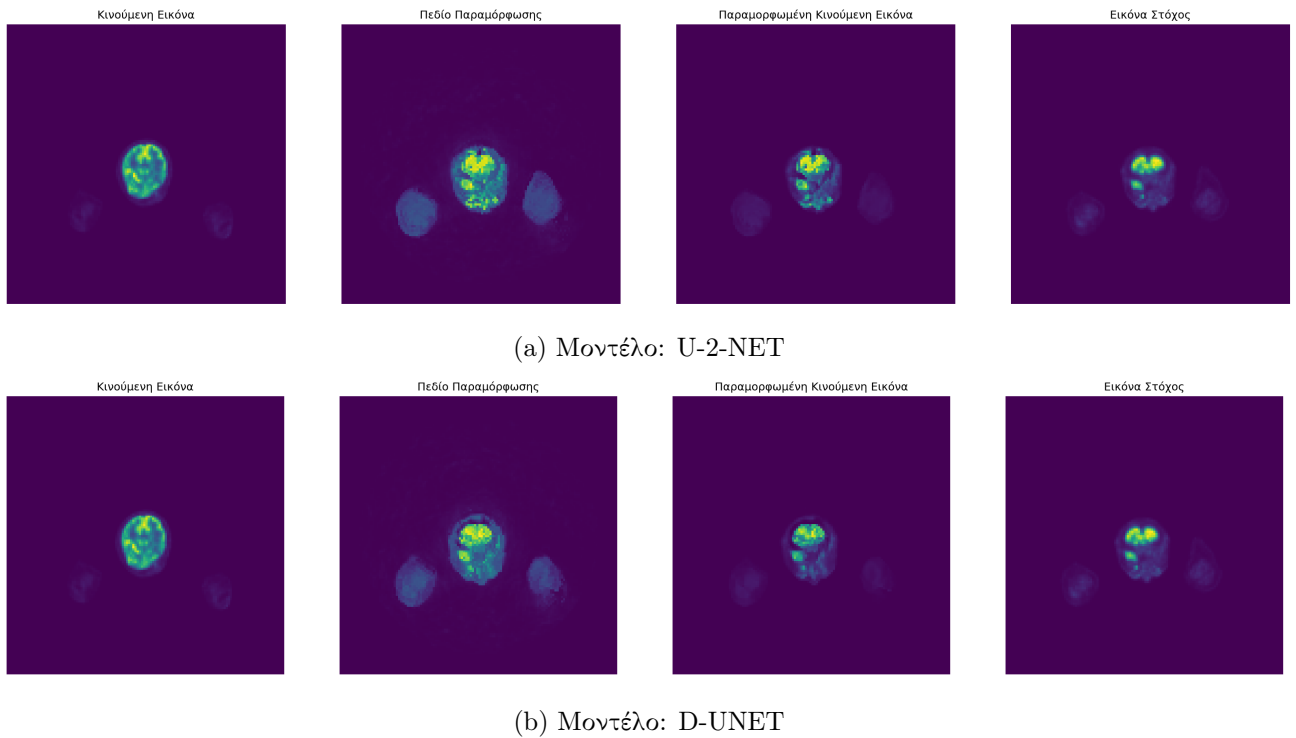
(a) Μοντέλο: U-2-NET



(b) Μοντέλο: D-UNET

Εικόνα 39: Σύγκριση της ευθυγράμμισης των μοντέλων (τομή 85)





Εικόνα 40: Σύγκριση της ευθυγράμμισης των μοντέλων (τομή 120)

Θα κάνουμε τώρα μερικά σχόλια για τις παραπάνω εικόνες. Ξεκινάμε με την σειρά, με την εικόνα 36. Παρατηρούμε ότι το μοντέλο U-2-NET παράγει κάποια περιττά φωτεινά σημεία αλλά πετυχαίνει καλύτερα το σχήμα της περιοχής του οργανισμού, σε αντίθεση με το μοντέλο D-UNET που φαίνεται ότι πετυχαίνει καλύτερα τις εντάσεις (συγκεκριμένα στην μέση της εικόνας, στο κίτρινο φωτεινό σημείο) αλλά χάνει λίγο στο σύνορο της περιοχής. Σε ό,τι αφορά τις εντάσεις παρατηρούμε τα ίδια πράγματα στις εικόνες 37, 38 αλλά και τα δύο μοντέλα πετυχαίνουν το σχήμα της εικόνας στόχο. Η μεγαλύτερη διαφορά στις εντάσεις φαίνεται στην εικόνα 39 όπου, παρόλο που και τα δύο μοντέλα δεν καταφέρνουν με τεράστια επιτυχία να απεικονίσουν την κινούμενη εικόνα στην εικόνα στόχο (σε ό,τι αφορά τις εντάσεις-χρώματα), βλέπουμε ότι το μοντέλο U-2-NET αστοχεί αρκετά σε σχέση με το D-UNET. Το D-UNET φαίνεται ότι "φωτίζει" ένα μεγαλύτερο μέρος της εικόνας σε σχέση με αυτό που υποδεικνύει η εικόνα στόχος. Και πάλι όμως και στις δύο περιπτώσεις η παραμορφωμένη κινούμενη εικόνα μοιάζει πολύ περισσότερο στην εικόνα στόχο σε σχέση με την κινούμενη εικόνα. Τελειώνουμε με την εικόνα 40. Και εδώ παρατηρούμε ότι το D-UNET όχι μόνο παράγει πιο ακριβείς εντάσεις, αλλά πετυχαίνει και το σχήμα της εικόνας, τόσο του εγκεφάλου, όσο και τα δύο "ποδαράκια".

Βλέπουμε λοιπόν ότι παρόλο που και τα δύο μοντέλα που επιλέξαμε για αξιολόγηση πετυχαίνουν τον στόχο τους, αν έπρεπε να επιλεγεί ένα από τα δύο ως "καλύτερο" τότε

αυτό θα ήταν το D-UNET. Πετυχαίνει καλύτερη απεικόνιση των εντάσεων σε όλες τις εικόνες και το σχήμα του απεικονιζόμενου χωρίου είναι επίσης πολύ ικανοποιητικό. Φαίνεται ότι η ενδιάμεση επεξεργασία των χαρακτηριστικών που παράγονται από τις συνελίξεις από τα πλήρως συνδεδεμένα στρώματα δίνει στο τέλος των συνελίξεων ένα πιο ουσιώδες και αντιπροσωπευτικό διάνυσμα το οποίο ανακατασκευάζεται με μεγαλύτερη επιτυχία μέσω των από-συνελίξεων.

## 8 Σχολιασμοί και Συμπεράσματα

Όπως ίσως φάνηκε από τα προηγούμενα κεφάλαια, το πρόβλημα της ευθυγράμμισης εικόνων δεν είναι καθόλου τετριμμένο. Απαιτείται ιδιαίτερη προσοχή στην επεξεργασία των εικόνων έτσι ώστε να αποφύγουμε το κλασικό πρόβλημα στην βαθιά μάθηση που λέγεται "Garbage in - Garbage out". Αυτό σημαίνει ότι αν δώσουμε κακής ποιότητας δεδομένα στο νευρωνικό δίκτυο θα πάρουμε και κακής ποιότητας αποτελέσματα. Στην περίπτωσή μας, οι αρχικές εικόνες συμπίεστηκαν όλες σε εικόνες διαστάσεων  $128 \times 128 \times 128$  για λόγους που σχετίζονται με περιορισμούς στην διαθέσιμη μνήμη GPU. Προφανώς όταν μειώνονται οι διαστάσεις μίας εικόνας η ανάλυσή της αλλοιώνεται έστω και σε μικρό βαθμό οπότε και χάνεται πληροφορία που θα μπορούσε να επεξεργασθεί από το νευρωνικό δίκτυο. Συνεπώς όσο λιγότερο συμπίεσουμε τις εικόνες μας, τόσο πιο αξιόπιστα αποτελέσματα θα πάρουμε σε ό,τι αφορά την ποιότητα της ευθυγράμμισης. Αν υπάρχει αφθονία μνήμης GPU, τότε προφανώς μειώνουμε τις εικόνες όσο λιγότερο μπορούμε ώστε να διατηρήσουμε κατά το δυνατόν περισσότερη πληροφορία αλλά και για να μειώσουμε τον υπολογιστικό χρόνο του κάθε πειράματος. Δυστυχώς στην παρούσα διπλωματική εργασία δεν μπορέσαμε να δουλέψουμε καθόλου με τις μάσκες κατάτμησης λόγω του πολύ μεγάλου αριθμού μηδενικών τιμών, όπως αναφέρθηκε στην σελίδα 59. Άλλη μία ερευνητική πορεία θα ήταν να μελετηθεί διεξοδικά η επιβλεπόμενη έκδοση του προβλήματος της ευθυγράμμισης, όπως περιγράφηκε στο τέλος του κεφαλαίου 5.

Από αυτό το σημείο και έπειτα, η πιο κρίσιμη μεταβλητή του προβλήματος είναι το ίδιο το νευρωνικό δίκτυο. Είδαμε ότι η δομή του παίζει καθοριστικό ρόλο στην ποιότητα της ευθυγράμμισης των εικόνων. Από τα πειράματα που εκτελέστηκαν φαίνεται ότι η δομή ενός δικτύου που αρχικά συμπιέζει την εικόνα και έπειτα την ανακατασκευάζει είναι η καταλληλότερη για το πρόβλημα της ευθυγράμμισης. Άλλωστε το μοντέλο VoxelMorph χρησιμοποιούσε δίκτυο τύπου U-NET οπότε και αναμέναμε ικανοποιητικά αποτελέσματα. Είδαμε και άλλα μοντέλα που σχετίζονται με την ευθυγράμμιση εικόνων αλλά προφανώς η επιλογή της δομής νευρωνικού δικτύου αποτελεί ανοιχτό -και εξαιρετικά ενδιαφέρον- ερευνητικό πρόβλημα. Από την δικιά μας μικρής κλίμακας έρευνα φαίνεται ότι μοντέλα όπως το D-UNET ή μοντέλα που σπάνε την εικόνα όπως το U-2-NET δείχνουν μία πολλά υποσχόμενη ερευνητική πορεία. Συγκεκριμένα, σε ό,τι αφορά το U-2-NET, θα μπορούσε κάλλιστα να γενικευθεί έτσι ώστε να περιλαμβάνει περισσότερα αλλά μικρότερα νευρωνικά δίκτυα και η εικόνα να σπάει σε μικρότερα κομμάτια. Μάλιστα παρατηρήθηκε ότι το μοντέλο U-2-NET απαιτούσε τον μικρότερο υπολογιστικό χρόνο για να ολοκληρωθεί η εκπαίδευσή του. Προφανώς θα πρέπει να κρατήσουμε στο μυαλό μας ότι δεν είναι επιθυμητό να σπάσουμε την εικόνα σε πάρα πολύ μικρά κομμάτια, αφού έτσι ενδέχεται να μην υπάρχει ικανοποιητικά αρκετή και ουσιαστική πληροφορία σε κάθε κομμάτι για να μπορεί να επεξεργασθεί από το εκάστοτε νευρωνικό δίκτυο. Επίσης μπορούν επίσης να γίνουν πειράματα που να συνδυάζουν την λογική του U-2-NET και του D-UNET, δηλαδή αντί για δύο μικρότερα δίκτυα τύπου UNET, να έχουμε δύο μικρότερα δίκτυα τύπου D-UNET. Οι συνδυασμοί προφανώς είναι άπειροι και οι καλύτεροι από αυτούς μπορούν να βρεθούν μόνο με δοκιμές και πειράματα. Σημειώνουμε εδώ επίσης ότι παρατηρήθηκε στο προηγούμενο κεφάλαιο, δηλαδή το γεγονός ότι το μοντέλο D-UNET φαίνεται να απεικονίζει με μεγαλύτερη ακρίβεια τις εντάσεις της

κινούμενης εικόνας σε αυτές της εικόνας στόχου σε σχέση με το μοντέλο U-2-NET. Αυτή η παρατήρηση θα πρέπει να διερευνηθεί με περισσότερα πειράματα.

Επιπλέον μεταβλητές είναι όπως είδαμε η συνάρτηση κόστους και ομαλοποίησης. Σε γενικές γραμμές φάνηκε ότι η συνάρτηση ολικής απόκλισης καθιστά την εκπαίδευση ασταθή ενώ τα τελικά μοντέλα δεν είναι τόσο ικανοποιητικά όσο αυτά που εκπαιδεύονται με ομαλοποίηση  $L2$ . Στο κεφάλαιο 3 περιγράφηκαν και άλλες συναρτήσεις κόστους και ομαλοποίησης. Και σε αυτήν την περίπτωση, ο λόγος που δεν δοκιμάστηκαν στα πειράματά μας είναι επειδή δεν υπήρχε αρκετή μνήμη. Όλες αυτές οι περίπλοκες συναρτήσεις εκτελούν έναν μεγάλο αριθμό πολλαπλασιασμών πινάκων μεγάλης διάστασης. Σε συνδυασμό με τις πράξεις που εκτελούνται κατά την οπισθοδιάσωση στο νευρωνικό δίκτυο, αυτό οδηγούσε σε σφάλματα τύπου OOM (Out Of Memory). Θα είχε ενδιαφέρον να γίνουν πειράματα με διάφορους συνδυασμούς ομαλοποίησης για να συγκριθεί η ποιότητα του παραγόμενου διανυσματικού πεδίου παραμόρφωσης.

Τελειώνουμε συνοψίζοντας τα παραπάνω σχόλια σε μία λίστα που περιλαμβάνει πιθανές μελλοντικές ερευνητικές πορείες - επεκτάσεις της παρούσας εργασίας:

- Διεξαγωγή πειραμάτων με διαφορετικές δομές νευρωνικών δικτύων με στόχο την μείωση της πολυπλοκότητας του τελικού μοντέλου χωρίς φυσικά να θυσιαστεί η ποιότητα της ευθυγράμμισης. Συγκεκριμένα ο συνδυασμός των μοντέλων U-2-NET και D-UNET φαίνεται ένα πολλά υποσχόμενο επόμενο βήμα
- Διεξαγωγή πειραμάτων με διαφορετικές συναρτήσεις ομαλοποίησης με στόχο την παραγωγή ενός κατά το δυνατόν ποιοτικού και αποτελεσματικού διανυσματικού πεδίου παραμόρφωσης
- Μελέτη του προβλήματος επιβλεπόμενης ευθυγράμμισης με την χρήση μασκών κατάτμησης. Για τον στόχο αυτόν θα πρέπει να χορηγηθούν καλής ποιότητας δεδομένα από αρμόδιους φορείς
- Γενίκευση του μοντέλου U-2-NET σε ένα μοντέλο που θα περιλαμβάνει περισσότερα από δύο μικρά νευρωνικά δίκτυα. Διερεύνηση της δυνατότητας παραλληλοποίησης του εν λόγω γενικευμένου μοντέλου, U-N-NET, σε  $N$  μονάδες επεξεργασίας GPU

## Βιβλιογραφία

- [1] <https://www.itransition.com/machine-learning/statistics>
- [2] <https://learn.g2.com/machine-learning-statistics>
- [3] <https://nix-united.com/blog/machine-learning-in-healthcare-12-real-world-use-cases-to-know/>
- [4] Knowles MA Selby P. Introduction to the Cellular and Molecular Biology of Cancer. New York: Oxford University Press; 2005.
- [5] Sherry L. Murphy, B.S., Kenneth D. Kochanek, M.A., Jiaquan Xu, M.D., and Elizabeth Arias, Ph.D. Mortality in the United States, 2020. NCHS Data Brief, No. 427, December 2021
- [6] Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. CA: A Cancer Journal for Clinicians 2022; 72(1):7-33. Last accessed May 10, 2022. [PubMed Abstract]
- [7] <https://monographs.iarc.who.int/agents-classified-by-the-iarc/>
- [8] A.M.J., Paans. Positron emission tomography. <https://cds.cern.ch/record/1005065/files/p363.pdf>
- [9] [https://en.wikipedia.org/wiki/Positron\\_emission\\_tomography#/media/File:ECAT-Exact-HR--PET-Scanner.jpg](https://en.wikipedia.org/wiki/Positron_emission_tomography#/media/File:ECAT-Exact-HR--PET-Scanner.jpg)
- [10] Mirza, F., Michael, I., Alain, T., and Laurent, Y. 2005. Computing Large Deformation Metric Mappings via Geodesic Flows of Diffeomorphisms. International Journal of Computer Vision 61(2), 139–157
- [11] Olaf, R., Philipp, F. and Thomas, B. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv:1505.04597 [cs.CV]
- [12] Guha, B., Amy, Z., Mert, R., John, G and Adrian, V. 2019. VoxelMorph: A Learning Framework for Deformable Medical Image Registration. arXiv:1809.05231 [cs.CV]
- [13] Ian Goodfellow, Yoshua Bengio and Aaron Courville. Deep Learning. MIT Press. 2016. <http://www.deeplearningbook.org>
- [14] Le Lu, Xiaosong Wan, Gustavo Carneiro and Lin Yang. 2019. Convolutional Neural Networks for Medical Imaging and Clinical Informatics. Springer; 1st ed. 2019 edition (1 Oct. 2019)
- [15] Norbert Wiener, Lambert M. Surhone and Mariam T. Ed (Author). 2010. Wiener Deconvolution. *betascript publishing*.

- [16] <https://pytorch.org/docs/stable/generated/torch.nn.ConvTranspose2d.html>
- [17] X. Hu, J. Yang and J. Yang. 2020. "A CNN-Based Approach for Lung 3D-CT Registration", in IEEE Access, vol. 8, pp. 192835-192843, 2020, doi: 10.1109/ACCESS.2020.3032612.
- [18] Stergios, C. et al. (2018). Linear and Deformable Image Registration with 3D Convolutional Neural Networks. In: , et al. Image Analysis for Moving Organ, Breast, and Thoracic Images. RAMBO BIA TIA 2018 2018 2018. Lecture Notes in Computer Science(), vol 11040. Springer, Cham. [https://doi.org/10.1007/978-3-030-00946-5\\_2](https://doi.org/10.1007/978-3-030-00946-5_2)
- [19] Mohapatra, S., Weisshaar, J.C. Modified Pearson correlation coefficient for two-color imaging in spherocylindrical cells. BMC Bioinformatics 19, 428 (2018). <https://doi.org/10.1186/s12859-018-2444-3>
- [20] Russakoff, D.B., Tomasi, C., Rohlfing, T., Maurer, C.R. (2004). Image Similarity Using Mutual Information of Regions. In: Pajdla, T., Matas, J. (eds) Computer Vision - ECCV 2004. ECCV 2004. Lecture Notes in Computer Science, vol 3023. Springer, Berlin, Heidelberg. [https://doi.org/10.1007/978-3-540-24672-5\\_47](https://doi.org/10.1007/978-3-540-24672-5_47)
- [21] Kuang, D., Schmah, T. (2019). FAIM – A ConvNet Method for Unsupervised 3D Medical Image Registration. In: Suk, HI., Liu, M., Yan, P., Lian, C. (eds) Machine Learning in Medical Imaging. MLMI 2019. Lecture Notes in Computer Science(), vol 11861. Springer, Cham. [https://doi.org/10.1007/978-3-030-32692-0\\_74](https://doi.org/10.1007/978-3-030-32692-0_74)
- [22] Mok, Tony and Chung, Albert. (2020). Fast Symmetric Diffeomorphic Image Registration with Convolutional Neural Networks. 4643-4652. 10.1109/CVPR42600.2020.00470.
- [23] Chen J, Li Y, Du Y, Frey EC. Generating anthropomorphic phantoms using fully unsupervised deformable image registration with convolutional neural networks. Med Phys. 2020 Dec;47(12):6366-6380. doi: 10.1002/mp.14545. Epub 2020 Nov 9. PMID: 33078422.
- [24] Mattes D, Haynor DR, Vesselle H, Lewellen TK, Eubank W. PET-CT image registration in the chest using free-form deformations. IEEE Trans Med Imaging. 2003 Jan;22(1):120-8. doi: 10.1109/TMI.2003.809072. PMID: 12703765.
- [25] V. Vishnevskiy, T. Gass, G. Szekely, C. Tanner and O. Goksel, "Isotropic Total Variation Regularization of Displacements in Parametric Image Registration," in IEEE Transactions on Medical Imaging, vol. 36, no. 2, pp. 385-395, Feb. 2017, doi: 10.1109/TMI.2016.2610583.
- [26] Fukui, Yoshiro and Takahiro Wada. "Passive velocity field control with discontinuous desired fields: Non-smooth potential gradient vector field." 2017 IEEE 56th Annual Conference on Decision and Control (CDC) (2017): 3301-3307.

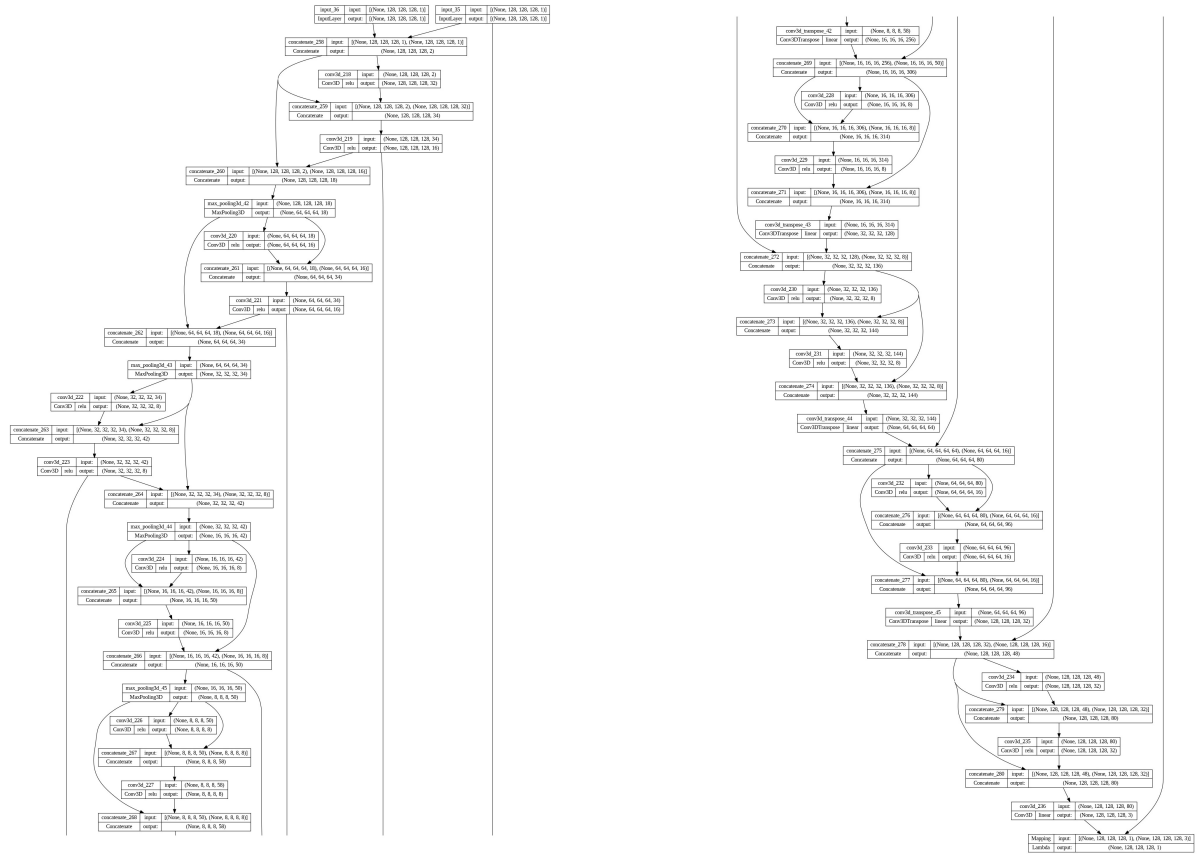
- [27] Dwarikanath Mahapatra, "GAN Based Medical Image Registration". (2019). eprint:1805.02369, arXiv:1805.02369, doi:<https://doi.org/10.48550/arXiv.1805.02369>
- [28] Junyu Chen and Yufan He and Eric Frey and Ye Li and Yong Du, "ViT-V-Net: Vision Transformer for Unsupervised Volumetric Medical Image Registration", in Medical Imaging with Deep Learning. (2021), url: <https://openreview.net/forum?id=h3HC1EU7AEz>
- [29] Junyu Chen and Eric C. Frey and Yufan He and William P. Segars and Ye Li and Yong Du, "TransMorph: Transformer for unsupervised medical image registration". Medical Image Analysis, p. 102615. (2022), <https://doi.org/10.1016/j.media.2022.102615>
- [30] Liu, Ze and Lin, Yutong and Cao, Yue and Hu, Han and Wei, Yixuan and Zhang, Zheng and Lin, Stephen and Guo, Baining, "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows", in Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), (2021)
- [31] <https://medium.com/@sarathchandra.knv31/deep-learning-based-3d-deformable-registration-of-t1-weighted-images-686777469d7c>
- [32] <https://en.wikipedia.org/wiki/DICOM>
- [33] <https://fileinfo.com/extension/dicom>
- [34] [https://docs.safe.com/fme/html/FME\\_Desktop\\_Documentation/FME\\_Readers\\_Writers/nifti/nifti.htm](https://docs.safe.com/fme/html/FME_Desktop_Documentation/FME_Readers_Writers/nifti/nifti.htm)
- [35] <https://brainder.org/2012/09/23/the-nifti-file-format/>
- [36] [https://docs.safe.com/fme/html/FME\\_Desktop\\_Documentation/FME\\_Readers\\_Writers/nifti/nifti.htm](https://docs.safe.com/fme/html/FME_Desktop_Documentation/FME_Readers_Writers/nifti/nifti.htm)
- [37] [https://github.com/lab-midas/TCIA\\_processing](https://github.com/lab-midas/TCIA_processing)
- [38] <https://simpleitk.org/doxygen/latest/html/index.html>
- [39] Kinahan PE, Fletcher JW. Positron emission tomography-computed tomography standardized uptake values in clinical practice and assessing response to therapy. Semin Ultrasound CT MR. 2010 Dec;31(6):496-505. doi: 10.1053/j.sult.2010.10.001. PMID: 21147377; PMCID: PMC3026294.
- [40] Gauri, M. and Jyotika, R. 2021. Protocol Handbook for Cancer Biology. Academic Press. DOI: <https://doi.org/10.1016/C2020-0-03032-X>
- [41] <https://pydicom.github.io/>
- [42] [https://en.wikipedia.org/wiki/Normalization\\_\(image\\_processing\)](https://en.wikipedia.org/wiki/Normalization_(image_processing))

- [43] [https://github.com/fitushar/3D-Medical-Imaging-Preprocessing-All-you-need/blob/master/Code/.ipynb\\_checkpoints/3DMedical-Imaging-Preprocessing-All-you-need-checkpoint.ipynb](https://github.com/fitushar/3D-Medical-Imaging-Preprocessing-All-you-need/blob/master/Code/.ipynb_checkpoints/3DMedical-Imaging-Preprocessing-All-you-need-checkpoint.ipynb)
- [44] <https://autopet.grand-challenge.org/>
- [45] 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
- [46] <https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=93258287>
- [47] <https://www.tensorflow.org/>
- [48] Diederik P. and Jimmy B., "Adam: A Method for Stochastic Optimization". 3rd International Conference for Learning Representations, San Diego, 2015, arxiv: [arXiv:1412.6980](https://arxiv.org/abs/1412.6980)
- [49] Reuben R., Yuval D., Jinyoung K., Guillermo S. and Noam H., "Continuous Dice Coefficient: a Method for Evaluating Probabilistic Segmentations". 2018, arxiv: [arXiv:1906.11031](https://arxiv.org/abs/1906.11031)



## Παράρτημα

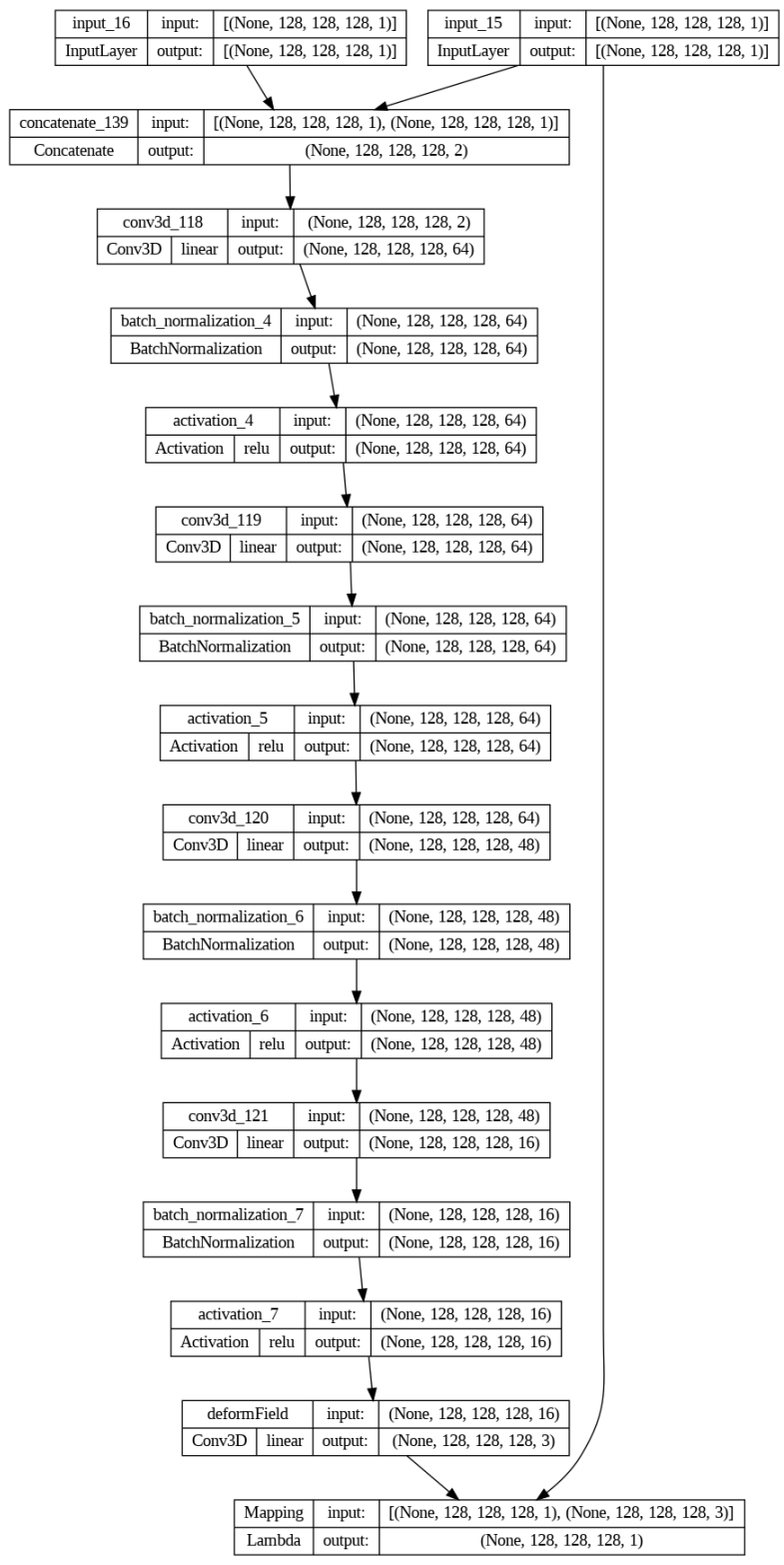
### Παράρτημα Α: Διαγραμματική απεικόνιση των μοντέλων στο Tensor-Flow



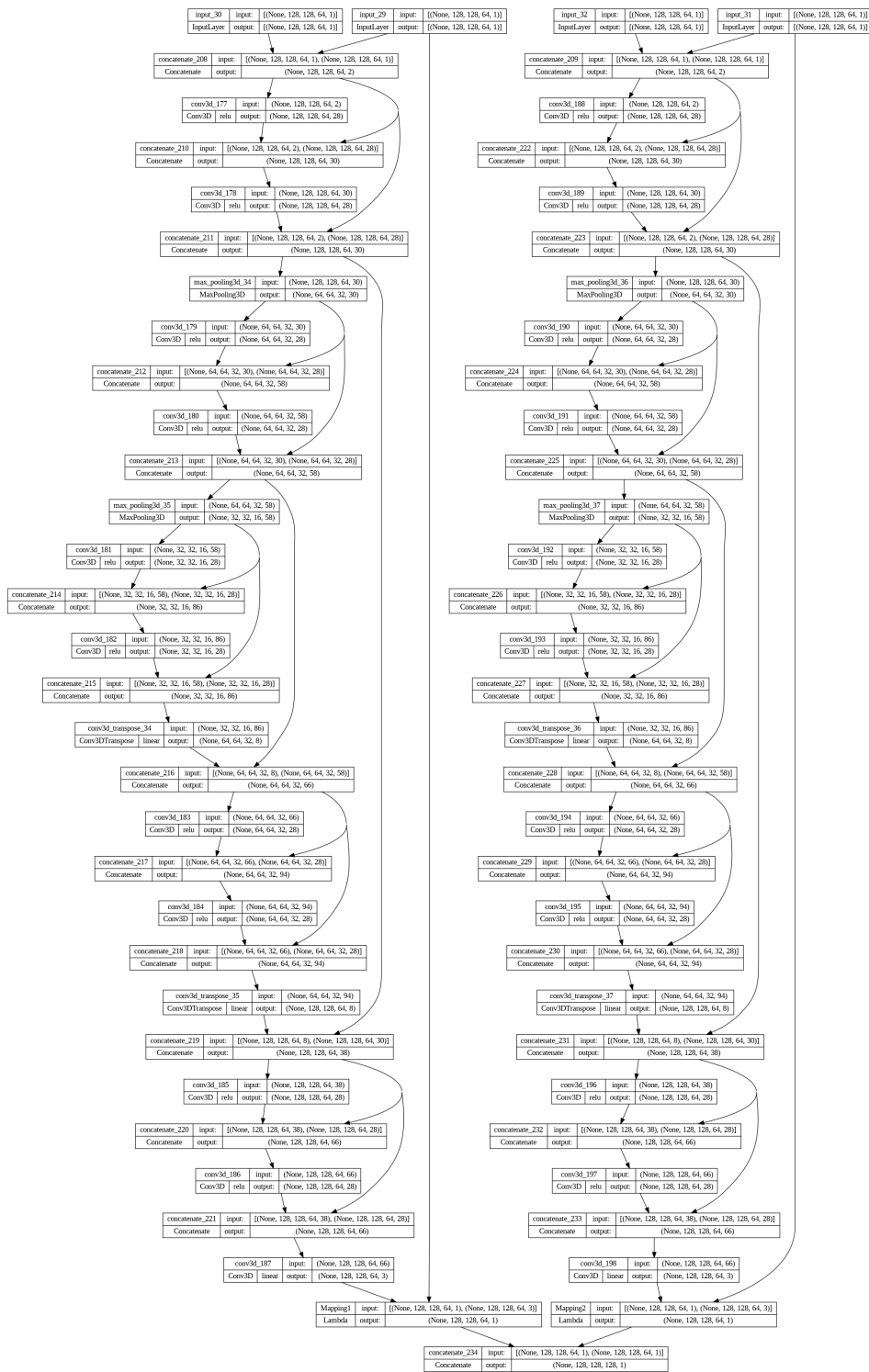
**Ανω μισό**

**Κάτω μισό**

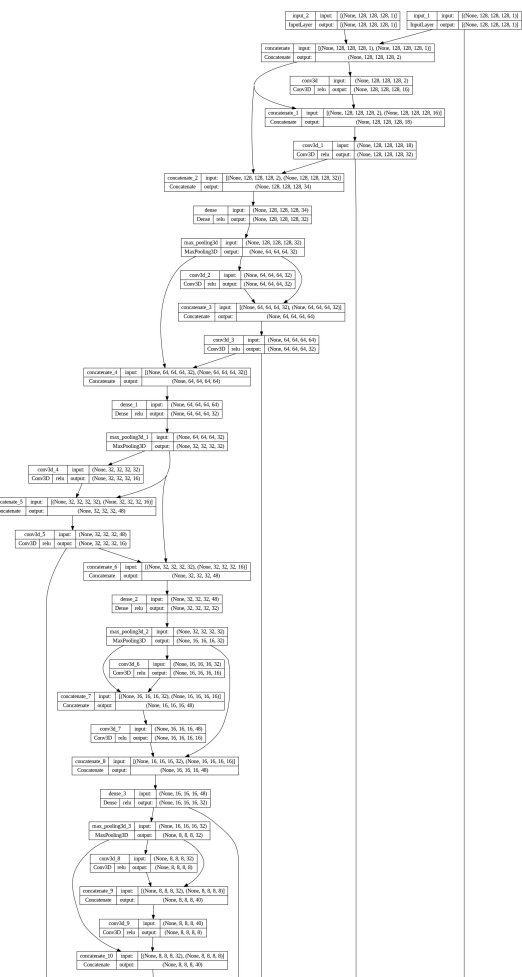
Εικόνα 41: Διαγραμματική απεικόνιση του μοντέλου U-NET



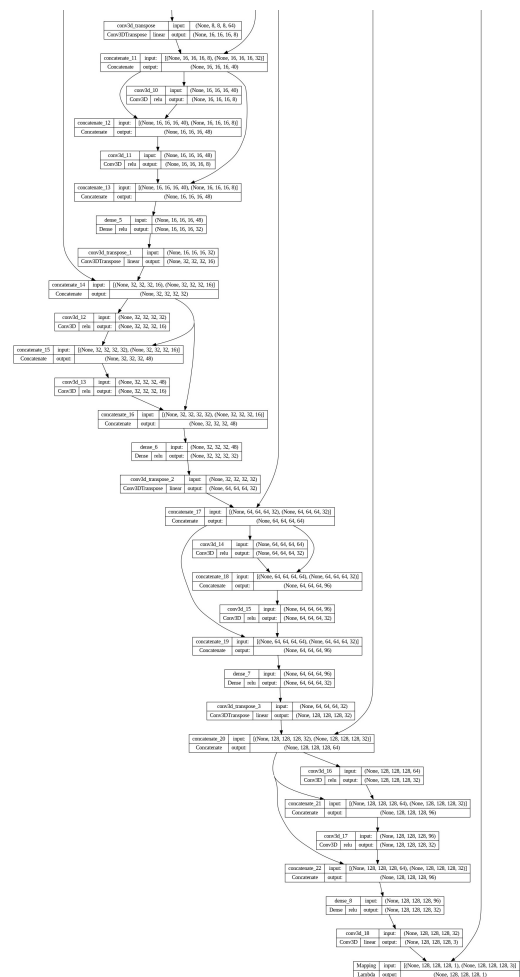
Εικόνα 42: Διαγραμματική απεικόνιση του μοντέλου CNN



Εικόνα 43: Διαγραμματική απεικόνιση του μοντέλου U-2-NET



**Άνω μισό**



**Κάτω μισό**

Εικόνα 44: Διαγραμματική απεικόνιση του μοντέλου D-UNET