



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ & ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ & ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

Αυτόματη Τμηματοποίηση Οργάνων Σε Δεδομένα PET-CT

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΤΟΥ

Ανδρέα Α. Βεζάκη

Επιβλέπων: Γεώργιος Ματσόπουλος
Καθηγητής ΕΜΠ

Αθήνα, Δεκέμβριος 2021



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ & ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ & ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

Αυτόματη Τμηματοποίηση Οργάνων Σε Δεδομένα PET-CT

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ
ΤΟΥ
Ανδρέα Α. Βεζάκη

Επιβλέπων: Γεώργιος Ματσόπουλος
Καθηγητής ΕΜΠ

Εγκρίθηκε από την κάτωθι τριμελή επιτροπή την 9^η Δεκεμβρίου 2021.

Γεώργιος Ματσόπουλος
Καθηγητής

Δημήτριος Κουτσούρης
Ομ. Καθηγητής

Παναγιώτης Τσανάκας
Αναπληρωτής Καθηγητής

Αθήνα, Δεκέμβριος 2021

Βεζάκης Α. Ανδρέας

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών ΕΜΠ

Copyright © Βεζάκης Α. Ανδρέας, 2021.

Με την επιφύλαξη παντός δικαιώματος. All rights reserved.

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

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

ΔΗΛΩΣΗ ΜΗ ΛΟΓΟΚΛΟΠΗΣ ΚΑΙ ΑΝΑΛΗΨΗΣ ΠΡΟΣΩΠΙΚΗΣ ΕΥΘΥΝΗΣ

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

Περίληψη

Η εφαρμογή της μεθόδου PET/CT αποτελεί μία ριζική καινοτομία η οποία προσφέρει πολύτιμες πληροφορίες για την χαρτογράφηση των καρκινικών εστιών που υπάρχουν σε όργανα σε ολόκληρο σώμα. Εφαρμόζεται, τόσο κατά την αρχική διάγνωση της νόσου όσο και κατά την διάρκεια της θεραπείας, ώστε να εκτιμηθεί το αποτέλεσμα ανταπόκρισης στη θεραπεία αλλά και μετά την ολοκλήρωση της για να διερευνηθεί αν υπάρχουν υπολειπόμενες καρκινικές εστίες. Η τμηματοποίηση οργάνων σε τέτοια δεδομένα απαιτείται, για την εξαγωγή βιοδεικτών ή για τον διαχωρισμό τους από τις καρκινικές περιοχές. Ωστόσο, αυτό αποτελεί μία δύσκολη και χρονοβόρα διαδικασία λόγω της λεπτομέρειας που απαιτείται. Η αυτοματοποίηση της συγκεκριμένης διαδικασίας γίνεται με την σημασιολογική κατάτμηση. Η Σημασιολογική Κατάτμηση (Semantic Segmentation) είναι ένα από τα θεμελιώδη θέματα της Όρασης Υπολογιστών. Πιο συγκεκριμένα, είναι η διαδικασία ανάθεσης μιας κατηγορίας σε κάθε pixel μιας εικόνας. Τα Βαθιά Συνελκτικά Νευρωνικά Δίκτυα (Deep Convolutional Neural Networks) έχουν πετύχει πρόσφατα state-of-the-art επιδόσεις σε υψηλής κλίμακας προβλήματα αναγνώρισης όπως και στη σημασιολογική κατάτμηση. Αυτό έχει σαν αποτέλεσμα τέτοια μοντέλα να χρησιμοποιούνται πλέον σε πλείστες εφαρμογές αιχμής, μία εκ των οποίων είναι και η Επεξεργασία Ιατρικών Εικόνων.

Λέξεις-κλειδιά: PET/CT, Σημασιολογική Κατάτμηση, Βαθιά Συνελκτικά Νευρωνικά Δίκτυα, U-Net.

Abstract

The application of the PET/CT method is a radical innovation that offers valuable information for the mapping of cancer foci that appear in organs throughout the body. It is used both at the initial diagnosis of the disease and during treatment to assess the response to treatment and after treatment to investigate whether there are any remaining residual foci. However, segmentation of organs in such data is a difficult and time-consuming process due to the level of details required. The automation of this process is done by semantic segmentation.

Semantic Segmentation is one of the fundamental topics in Computer Vision. Specifically, it is the process of assigning a category to each pixel of an image. Deep Convolutional Neural Networks (DCN) have recently achieved state-of-the-art performance in high-scale recognition problems such as semantic segmentation. As a result, such models are now being used in most cutting-edge applications, one of which is in Medical Image Processing.

Keywords: PET/CT, Semantic Segmentation, Deep Convolutional Neural Networks, U-Net.

Ευχαριστίες

Τελειώνοντας τη διπλωματική μου εργασία, αισθάνομαι υποχρέωσή μου να ευχαριστήσω τον επιβλέποντα καθηγητή κ. Γεώργιο Ματσόπουλο για την εμπιστοσύνη που έδειξε στο πρόσωπό μου και για την ευκαιρία που μου έδωσε προκειμένου να εκπονήσω την εν λόγω εργασία στο εργαστήριό του. Επίσης, θα ήθελα να εκφράσω τις θερμές ευχαριστίες μου στην υπόλοιπη τριμελή επιτροπή, καθ. κ. Δημήτριο Κουτσούρη και καθ. κ. Παναγιώτη Τσανάκα. Τέλος, θα ήθελα να ευχαριστήσω και τον κ. Θεόδωρο Βαγενά ο οποίος συνεπέβλεπε την εργασία μου, καθώς και τον αδερφό μου κ. Ιωάννη Βεζάκη ο οποίος με ενέπνευσε και αποτέλεσε καταλυτικό παράγοντα ακόμα και για την εισαγωγή μου στην συγκεκριμένη σχολή.

Περιεχόμενα

1	Εισαγωγή	15
1.1	Σημσιολογική κατάνηση οργάνων σε PET δεδομένα	17
2	PET/CT	19
2.1	Αξονική Τομογραφία (CT)	19
2.1.1	Εξέλιξη της αξονικής τομογραφίας	20
2.1.2	Ανακατασκευή της εικόνας	24
2.2	Πυρηνική Ιατρική	27
2.2.1	PET	28
2.3	PET/CT	31
3	Βαθιά Μάθηση	35
3.1	Νευρωνικά Δίκτυα	35
3.1.1	Δομή	35
3.1.2	Convolutional Neural Networks	39
3.2	Εκπαίδευση Νευρωνικών Δικτύων	44
3.2.1	Αλγόριθμοι Βελτιστοποίησης	46
3.2.2	Backpropagation	48
3.2.3	Συνάρτηση Κόστους (Loss Function)	49
3.3	Σημσιολογική Κατάνηση (Semantic Segmentation)	52
3.3.1	Πριν την Βαθιά Μάθηση	53
3.3.2	Χρησιμοποιώντας την Βαθιά Μάθηση	54
3.3.3	Διάσημα Μοντέλα	55
4	Προτεινόμενη Μέθοδος	63
4.1	Κεντρική Ιδέα	63
4.2	Προεπεξεργασία	64
4.3	Εκπαίδευση	66
4.4	Δεδομένα	72
4.5	Αποτελέσματα και Συγκρίσεις	74
5	Συμπεράσματα	81

Κατάλογος Σχημάτων

- 1.1 (α) Τομογραφία εκπομπής ποζιτρονίων (β) Αξονική τομογραφία (γ) PET/CT [1] 16
- 2.1 Από τους ct scanners τελευταίας γενιάς μπορούν να πραγματοποιηθούν τομογραφίες ολόκληρου του σώματος, απεικονίζοντας με ακρίβεια κάθε μικρό αγγείο μέχρι και στα πόδια του ασθενή. . . . 20
- 2.2 (α) Η συμβατική ακτινογραφία παράγει εικόνες με χαμηλή αντίθεση (contrast) όπως παρουσιάζεται στη αριστερή εικόνα με την ακτινογραφία κρανίου, όπου λεπτομέρειες σχετικά με την δομή στο χώρο δεν γίνονται αντιληπτές και στη δεξιά εικόνα με το γόνατο το οποίο παρουσιάζεται θολό. Η χαμηλή αντίθεση είναι αναμενόμενη αφού χρησιμοποιεί averaging για την δημιουργία της εικόνας. (β) Στην εικόνα b παρουσιάζεται η διαδικασία του Grossman τομογράφου ο οποίος απεικονίζει ένα μόνο slice του γονάτου. 21
- 2.3 Στην συγκεκριμένη εικόνα παρουσιάζονται οι 4 γενιές αξονικών τομογράφων (α) Ο πρώτης γενιάς CT εκπέμπει μία λεπτή ακτίνα σε έναν μόνο ανιχνευτή (β) Ο δεύτερης γενιάς CT εκπέμπει ριποειδή δέσμη σε συστοιχία ανιχνευτών (γ) Ο τρίτης γενιάς CT εκπέμπει δέσμη, η οποία καλύπτει ολόκληρο τον κορμό του ασθενή, σε ακόμα περισσότερους ανιχνευτές (δ) Ο τέταρτης γενιάς CT διαθέτει πια στάσιμο δακτύλιο ανιχνευτών. 23
- 2.4 Η εξέλιξη των αξονικών τομογράφων. Οι (α) και (β) αποτελούν αξονικούς τομογράφους πρώτης γενιάς και πραγματοποιούν τομογραφίες μόνο στο κεφάλι. Ο (γ) είναι δεύτερης γενιάς και ο (δ) είναι τρίτης γενιάς. 24
- 2.5 Από τα ακατέργαστα δεδομένα στην ανακατασκευασμένη εικόνα. Σινόγραμμα τιμών εξασθένισης ή προβολής (α), τιμές προβολής φιλτραρισμένης με υπερπερατό φίλτρο (β), οπισθοπροβολή της φιλτραρισμένης προβολής (γ), συσσώρευση όλων των οπισθοπροβολών (δ) και της εγγενούς προβολής (ε) 25
- 2.6 CT scans του θώρακα για διαφορετικές τιμές HU threshold. (α) Για 0 HU (β) Για 60 HU (γ) Για -700 HU (δ) Για -1000 HU. 27
- 2.7 PET scanner με τους δακτύλιους να εντοπίζουν τις εξαυλώσεις [3]. . . 29

2.8	Σχηματική παρουσίαση της αρχής λειτουργίας TOF PET. (A) Το φαινόμενο εξαΰλωσης συμβαίνει σε τυχαία σημεία με απόσταση d από το κέντρο αντικειμένου με διάμετρο D . Οι χρόνοι σύμπτωσης είναι t_1 και t_2 , (B) Ένας όχι TOF-scanner δημιουργεί μία ομοιόμορφη κατανομή πιθανότητας πάνω στην LOR για το που βρίσκεται το επιθυμητό σημείο δημιουργώντας με αυτό τον τρόπο θόρυβο σε κάποιο κομμάτι της εικόνας που κατασκευάζεται, (C) Σε έναν TOF-scanner η θέση εξαΰλωσης περιορίζεται σε ένα μικρό τμήμα της LOR με ακρίβεια που καθορίζεται από Γκαουσιανή κατανομή Δx , (D) Με τον αναβαθμισμένο εντοπισμό των σημείων πάνω στις LOR μειώνεται ο θόρυβος που δημιουργόταν προηγουμένως στη (B)	30
2.9	Στο σχήμα παρουσιάζονται οι 3 περιπτώσεις σύμπτωσης. Στην (a) αληθής σύμπτωση, (b) τυχαία σύμπτωση και (c) σύμπτωση λόγω σκέδασης	31
2.10	Το PET/CT Discovery IQ 3-ring	33
3.1	Είδη Επιπέδων Νευρωνικών Δικτύων. Κάθε κύκλος αποτελεί έναν νευρώνα.	36
3.2	Γραφική Παράσταση της Heavyside Step Function.	37
3.3	Γραφική Παράσταση της ReLU.	38
3.4	Το χαρακτηριστικό S της σιγμοειδής συνάρτησης.	39
3.5	Συνέλιξη εικόνας f με πυρήνα h	40
3.6	Διαφορετικά είδη φίλτρων, οι αντίστοιχοι πυρήνες τους και το αποτέλεσμα της εφαρμογής τους [15].	41
3.7	Είδη pooling [17].	41
3.8	P είναι το σημείο που θέλουμε να προβλέψουμε την καινούρια τιμή. Τα Q είναι τα 4 γειτονικά σημεία [20].	43
3.9	Batch Normalization [22].	44
3.10	Dropout [23].	45
3.11	Γραφική παράσταση τρισδιάστατης συνάρτησης [25].	47
3.12	Dice Loss [30].	50
3.13	Intersection over Union [31].	51
3.14	Παράδειγμα thresholding χρησιμοποιώντας τον αλγόριθμο του Otsus [33].	54
3.15	FNC πραγματοποιεί σημασιολογική κατάτμηση.	55
3.16	Απεικόνιση των διαφορετικών FCN αρχιτεκτονικών. (a) FCN-32 αρχιτεκτονική, (b) FCN-16 αρχιτεκτονική, (c) FCN-8 αρχιτεκτονική [34]	56
3.17	Atrous convolution, με διαφορετικά rates.	57
3.18	ASPP.	58
3.19	DeepLabv3.	58
3.20	Η αρχιτεκτονική του DeconvNet.	58

3.21	Οι διαδικασίες pooling και unpooling. Στην διαδικασία unpooling χρησιμοποιούνται οι προηγούμενες θέσεις των μεγίστων.	59
3.22	59
3.23	Η αρχιτεκτονική του SegNet	59
3.24	U-Net architecture.	60
3.25	U-Net architecture.	62
4.1	Η αρχιτεκτονική του τρισδιάστατου Attention U-Net μοντέλου μας. .	64
4.2	Η αρχιτεκτονική του δισδιάστατου U-Net μοντέλου μας.	65
4.3	66
4.4	Η συγκεκριμένη εικόνα παρουσιάζει την αφινική αυτό-ομοιότητα. Κάθε φύλλο έχει την δυνατότητα να μετατραπεί σε κάποιο άλλο με αφινικό μετασχηματισμό. Για παράδειγμα το κόκκινο φύλλο μπορεί να μετατραπεί σε οποιοδήποτε άλλο με μετατροπή της γωνίας και του μεγέθους [42].	67
4.5	Η αντίθεση στο αριστερό μέρος της εικόνας είναι μικρότερη από την αντίθεση στο δεξιό μέρος [43].	68
4.6	Εφαρμογή Γκαουσιανού Blur για διαφορετικές τιμές τυπικής απόκλισης [44].	69
4.7	Precision και Recall [48]	71
4.8	Ενδεικτική παρουσίαση των δεδομένων μέσω του προγράμματος 3D Slicer.	73
4.9	Κατηγοριοποίηση του εγκεφάλου. Στην πρώτη φαίνεται το αποτέλεσμα του ensembled μοντέλου, στην δεύτερη του δισδιάστατου U-Net, στην τρίτη του τρισδιάστατου Attention U-Net και στην τέταρτη είναι το επιθυμητό αποτέλεσμα.	76
4.10	Κατηγοριοποίηση της αριστερή κοιλίας της καρδιάς. Στην πρώτη φαίνεται το αποτέλεσμα του ensembled μοντέλου, στην δεύτερη του δισδιάστατου U-Net, στην τρίτη του τρισδιάστατου Attention U-Net και στην τέταρτη είναι το επιθυμητό αποτέλεσμα.	77
4.11	Κατηγοριοποίηση του ήπατος. Στην πρώτη φαίνεται το αποτέλεσμα του ensembled μοντέλου, στην δεύτερη του δισδιάστατου U-Net, στην τρίτη του τρισδιάστατου Attention U-Net και στην τέταρτη είναι το επιθυμητό αποτέλεσμα.	78
4.12	Κατηγοριοποίηση του δεξιού νεφρού. Στην πρώτη φαίνεται το αποτέλεσμα του ensembled μοντέλου, στην δεύτερη του δισδιάστατου U-Net, στην τρίτη του τρισδιάστατου Attention U-Net και στην τέταρτη είναι το επιθυμητό αποτέλεσμα.	79
4.13	Κατηγοριοποίηση της ουροδόχου κύστης. Στην πρώτη φαίνεται το αποτέλεσμα του ensembled μοντέλου, στην δεύτερη του δισδιάστατου U-Net, στην τρίτη του τρισδιάστατου Attention U-Net και στην τέταρτη είναι το επιθυμητό αποτέλεσμα.	80

Κεφάλαιο 1

Εισαγωγή

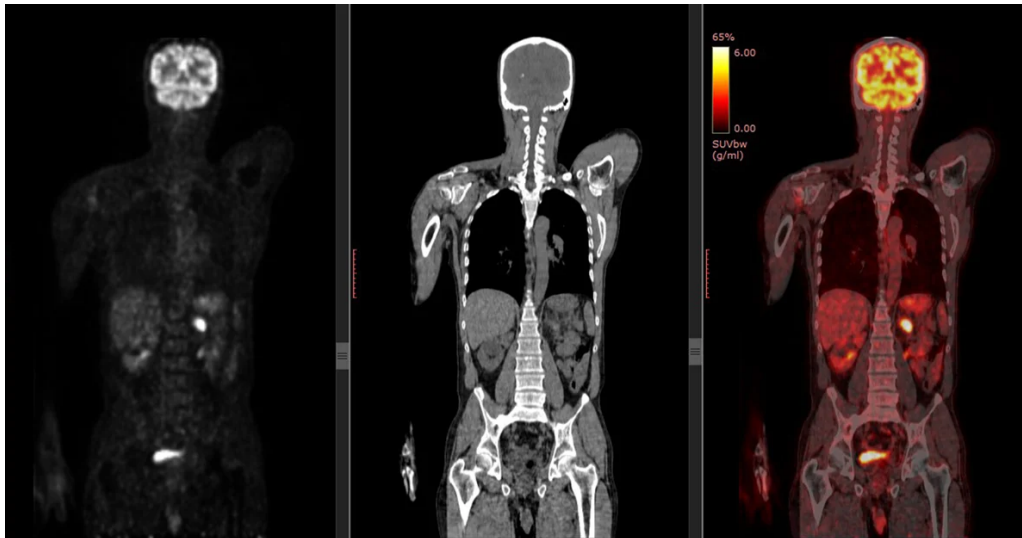
Η (PET/CT) είναι μια τεχνική ιατρικής απεικόνισης που συνδυάζει δύο μεθόδους, την τομογραφία εκπομπής ποζιτρονίων και την αξονική τομογραφία, για την παραγωγή λεπτομερών εικόνων του ανθρώπινου σώματος. Η PET είναι μια τεχνική απεικόνισης πυρηνικής ιατρικής που χρησιμοποιεί μια μικρή ποσότητα ραδιενεργού υλικού, για να παράγει εικόνες της μεταβολικής δραστηριότητας του σώματος. Η αξονική τομογραφία, από την άλλη πλευρά, χρησιμοποιεί ακτίνες X και έπειτα επεξεργασία των δεδομένων αυτών σε υπολογιστή για την παραγωγή λεπτομερών εικόνων της ανατομίας του σώματος.

Συνδυάζοντας τις δύο μεθόδους, η PET/CT παρέχει μια μοναδική και ιδιαίτερα κατατοπιστική εικόνα των εσωτερικών δομών και λειτουργιών του σώματος. Οι πληροφορίες αυτές μπορούν να χρησιμοποιηθούν για τη διάγνωση ενός ευρέος φάσματος παθήσεων, όπως ο καρκίνος, οι καρδιαγγειακές παθήσεις και οι νευρολογικές διαταραχές. Επιπλέον, η PET/CT μπορεί να βοηθήσει στην παρακολούθηση της εξέλιξης της νόσου και στην αξιολόγηση της αποτελεσματικότητας των θεραπειών. Είναι μια μη επεμβατική και ανώδυνη διαδικασία, η οποία συνήθως εκτελείται σε εξωνοσοκομειακή βάση. Η εξέταση διαρκεί συνήθως περίπου μία ώρα και είναι ανεκτή από τους περισσότερους ασθενείς. Τα αποτελέσματα μιας σάρωσης PET/CT μπορούν να παρέχουν πολύτιμες πληροφορίες που βοηθούν στην καθοδήγηση των θεραπευτικών αποφάσεων και στη βελτίωση της υγείας των ασθενών.

Όπως γίνεται αντιληπτό η παρατήρηση καθώς και η απεικόνιση του εσωτερικού του ανθρώπινου σώματος αποτελεί ένα βασικό κομμάτι της ιατρικής τόσο για την κατανόηση όσο και για τη διάγνωση παθολογιών. Με την εξέλιξη της τεχνολογίας η παρατήρηση και η απεικόνιση οδεύουν στην αυτοματοποίηση τους με την βοήθεια της τεχνητής νοημοσύνης.

Η Τεχνητή Νοημοσύνη (Artificial Intelligence, AI) είναι ένας ταχέως αναπτυσσόμενος τομέας που περιλαμβάνει την ανάπτυξη υπολογιστικών συστημάτων και αλγορίθμων που μπορούν να εκτελούν εργασίες, οι οποίες εκτελούνταν μόνο από ανθρώπους. Αυτό περιλαμβάνει εργασίες όπως η αναγνώριση εικόνων, η επεξεργασία φυσικής γλώσσας και η λήψη αποφάσεων.

Τα τελευταία χρόνια, η τεχνητή νοημοσύνη αποκτά ολοένα και μεγαλύτερη σημα-



Σχήμα 1.1: (α) Τομογραφία εκπομπής ποζιτρονίων (β) Αξονική τομογραφία (γ) PET/CT [1]

σία σε διάφορους κλάδους, συμπεριλαμβανομένης της ιατρικής. Στον τομέα της ιατρικής απεικόνισης, οι αλγόριθμοι AI μπορούν να χρησιμοποιηθούν για την ανάλυση και την ερμηνεία μεγάλου όγκου ιατρικών δεδομένων και να παρέχουν πληροφορίες που μπορούν να βοηθήσουν στη διάγνωση και τη θεραπεία.

Για παράδειγμα, οι αλγόριθμοι AI μπορούν να χρησιμοποιηθούν για τον εντοπισμό και κατηγοριοποίηση μοτίβων και ανωμαλιών σε ιατρικές εικόνες που μπορεί να υποδηλώνουν την παρουσία μιας ασθένειας. Χρησιμοποιούνται επίσης για την ποσοτικοποίηση της σοβαρότητας μιας νόσου και την παρακολούθηση της εξέλιξής της με την πάροδο του χρόνου. Επιπλέον, οι αλγόριθμοι τεχνητής νοημοσύνης γίνεται να χρησιμοποιηθούν για να βοηθήσουν στη λήψη αποφάσεων, όπως η πρόβλεψη της καλύτερης θεραπευτικής αγωγής για έναν συγκεκριμένο ασθενή με βάση το ιατρικό ιστορικό του και τα αποτελέσματα των απεικονίσεων.

Συνολικά, η τεχνητή νοημοσύνη έχει τη δυνατότητα να φέρει επανάσταση στον τομέα της ιατρικής, παρέχοντας ακριβέστερη, αποτελεσματικότερη και εξατομικευμένη φροντίδα για τους ασθενείς. Καθώς η τεχνολογία αυτή συνεχίζει να εξελίσσεται, είναι πιθανό να διαδραματίσει όλο και πιο σημαντικό ρόλο στον τομέα της ιατρικής και να βελτιώσει τα αποτελέσματα των ασθενών.

Στην παρούσα διπλωματική ελέγχεται η εφαρμογή της τεχνητής νοημοσύνης για την κατηγοριοποίηση των οργάνων καρδιά, ήπαρ, εγκέφαλο, νεφρά και ουροδόχο κύστη σε δεδομένα PET/CT. Δηλαδή, η αποτελεσματικότητα των μοντέλων βαθιάς μάθησης στον εντοπισμό και στην κατηγοριοποίηση μοτίβων και ανωμαλιών σε ιατρικά δεδομένα.

1.1 Σημαιολογική κατάτμηση οργάνων σε PET δεδομένα

Η σηματολογική κατάτμηση αναφέρεται στη διαδικασία ταξινόμησης κάθε εικονοστοιχείου σε μια εικόνα σε μία από διάφορες προκαθορισμένες κατηγορίες. Στο πλαίσιο των δεδομένων PET (Τομογραφία εκπομπής ποζιτρονίων), η σηματολογική τμηματοποίηση μπορεί να αποτελέσει χρήσιμο εργαλείο για την ανάλυση και ερμηνεία των εικόνων.

Η τμηματοποίηση των καρκινικών μεταστάσεων σε PET δεδομένα εμφανίζει πολλά ψευδώς θετικά αποτελέσματα, επομένως υπάρχουν φυσιολογικά όργανα τα οποία μπορούν να εξαιρεθούν από τον εντοπισμό των μεταστάσεων. Η χειροκίνητη εξαίρεση αυτών των φυσιολογικών προσλήψεων και γενικά η τμηματοποίηση των δεδομένων, από ειδικούς, είναι αρκετά χρονοβόρα, κουραστική και επιρρεπής σε σφάλματα. Επιπλέον, οι τμηματοποιημένες εικόνες PET μπορούν να χρησιμοποιηθούν για τη δημιουργία τρισδιάστατων απεικονίσεων των περιοχών ενδιαφέροντος, οι οποίες μπορούν να βοηθήσουν στον χειρουργικό σχεδιασμό, στον σχεδιασμό ακτινοθεραπείας και σε άλλες ιατρικές διαδικασίες. Αυτές οι απεικονίσεις μπορούν επίσης να βοηθήσουν τους ασθενείς να κατανοήσουν καλύτερα την κατάστασή τους. Επίσης, μπορούν να διευκολύνουν την έρευνα για διάφορες πτυχές της νόσου, όπως την παρατήρηση της εξέλιξης της, την ανταπόκριση στη θεραπεία και την επίδραση διαφόρων παραγόντων στην έκβαση της.

Κεφάλαιο 2

PET/CT

Η εξέταση PET/CT είναι ένα υβριδικό σύστημα απεικόνισης, το οποίο συνδυάζει δύο μεθόδους, την ποζιτρονική τομογραφία (PET) και την αξονική τομογραφία (CT), σε ένα μόνο μηχάνημα. Η αξονική τομογραφία (CT) είναι ανατομική απεικονιστική μέθοδος, δηλαδή παρουσιάζει το σχήμα και το μέγεθος οργάνων και δομών του σώματος. Η PET εντάσσεται στο τομέα της πυρηνικής ιατρικής και απεικονίζει μεταβολικές αλλαγές σε κυτταρικό επίπεδο, εξετάζοντας πως λειτουργούν όργανα και ιστοί.

2.1 Αξονική Τομογραφία (CT)

Η αξονική τομογραφία (Computed Tomography CT) προέρχεται από τις ελληνικές λέξεις τόμος που σημαίνει τμήμα και (ράδιο)γραφία και είναι η πρώτη μη επεμβατική απεικονιστική μέθοδος και πιο ευρέως χρησιμοποιημένη. Συγκεκριμένα, χρησιμοποιούνται ακτίνες X και λαμβάνονται εικόνες για όλα σχεδόν τα μέρη του σώματος, από μαλακούς ιστούς και οστά, μέχρι τον εγκέφαλο, την καρδιά και τα αγγεία. Χρησιμοποιείται τόσο για τη διάγνωση όσο και για την παρακολούθηση διαφόρων παθήσεων. Η αποτελεσματικότητά της σε συνδυασμό με την ευκολία χρήσης της, την καθιστά από τις πιο σημαντικές εξετάσεις. Γι' αυτό το λόγο υπάρχουν CT μηχανήματα σε shock rooms και trauma clinics για την γρήγορη εξέταση των ασθενών σε περίπτωση ανάγκης. Ωστόσο, υπάρχουν και άλλες ανταγωνιστικές μέθοδοι με τις πιο σημαντικές να είναι η MRI και το PET.

Η αξονική τομογραφία απεικονίζει όλες τις πληροφορίες και έχει πολύ υψηλότερη αντίθεση (contrast) σε σύγκριση με την συμβατική ακτινογραφία. Το πρόβλημα της x-ray είναι ότι προσπαθεί να παράξει μία μονοδιαστάσιατη εικόνα από ένα τρισδιάστατο αντικείμενο. Αυτό το επιτυγχάνει αναπαριστώντας την μέση πληροφορία του αντικειμένου, χάνοντας με αυτόν τον τρόπο αρκετή πληροφορία και μειώνοντας σημαντικά το contrast, καθιστώντας δύσκολη την δουλειά του ειδικού να πραγματοποιήσει την διάγνωση.



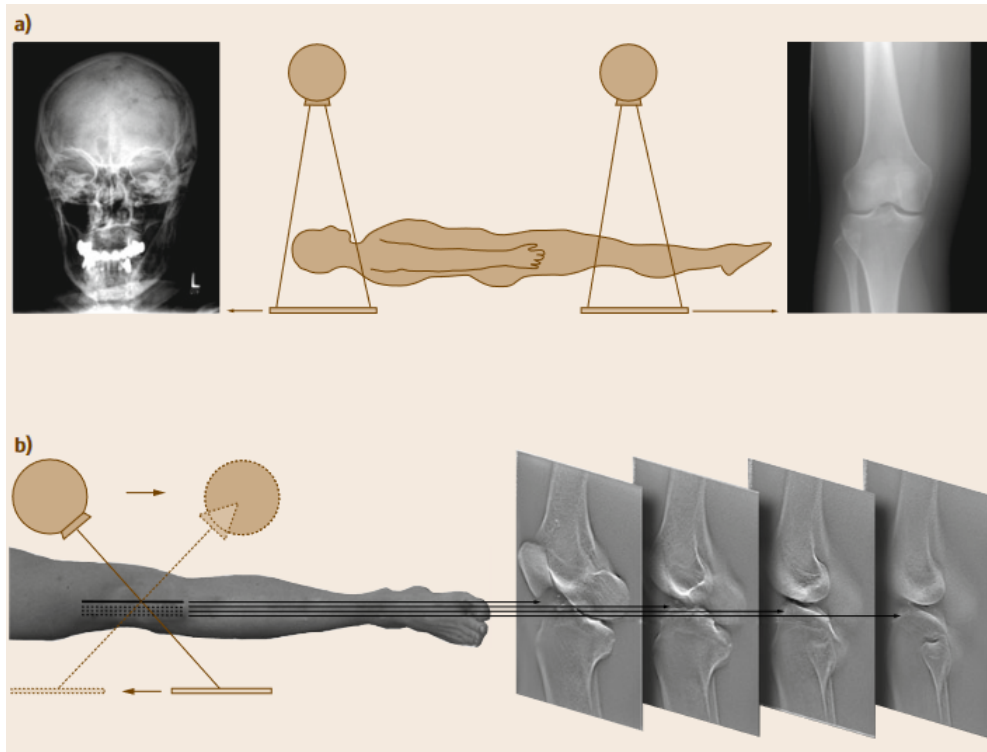
Σχήμα 2.1: Από τους ct scanners τελευταίας γενιάς μπορούν να πραγματοποιηθούν τομογραφίες ολόκληρου του σώματος, απεικονίζοντας με ακρίβεια κάθε μικρό αγγείο μέχρι και στα πόδια του ασθενή.

2.1.1 Εξέλιξη της αξονικής τομογραφίας

Blurring Tomography

Η ανάγκη για βελτίωση της συμβατικής ακτινογραφίας που απεικονίζει την μέση πληροφορία οδήγησαν στο πρώτο τομογραφικό επιχείρημα. Το 1920 ο Grossman ήταν ο πρώτος που κατάφερε να απεικονίσει μία μόνο τομή (slice) του σώματος χρησιμοποιώντας τον Grossman τομογράφο [2].

Ο τομογράφος αποτελείται από έναν x-ray tube και ένα x-ray film, τα οποία κατά την διάρκεια της ανατασκευής της εικόνας μετακινούνται ταυτόχρονα γραμμικά προς μία κατεύθυνση και αντίθετα μεταξύ τους. Με αυτό τον τρόπο μόνο τα σημεία που βρίσκονται στο κέντρο περιστροφής απεικονίζονται με ακρίβεια. Τα υπόλοιπα σημεία τα οποία δεν βρίσκονται στο κέντρο δεν εξαφανίζονται, αλλά παρουσιάζονται θολά σαν μία γκρι ομίχλη. Όσο πιο μακριά από το κέντρο βρίσκεται ένα σημείο, τόσο πιο θολό απεικονίζεται. Από αυτό το γεγονός πήρε το όνομά της η διαδικασία Blurring Tomography ή αλλιώς tomosynthesis όταν η εικόνα επεξεργάζεται ψηφιακά. Σε σύγκριση με την συμβατική ακτινογραφία υπάρχει σημαντική αύξηση της αντίθεσης και χαρακτηριστικά είναι πολύ πιο εύκολα διαχωρίσιμα όπως παρουσιάζεται και στο Σχήμα 2.2.



Σχήμα 2.2: (a) Η συμβατική ακτινογραφία παράγει εικόνες με χαμηλή αντίθεση (contrast) όπως παρουσιάζεται στη αριστερή εικόνα με την ακτινογραφία κρανίου, όπου λεπτομέρειες σχετικά με την δομή στο χώρο δεν γίνονται αντιληπτές και στη δεξιά εικόνα με το γόνατο το οποίο παρουσιάζεται θολό. Η χαμηλή αντίθεση είναι αναμενόμενη αφού χρησιμοποιεί averaging για την δημιουργία της εικόνας. (b) Στην εικόνα b παρουσιάζεται η διαδικασία του Grossman τομογράφου ο οποίος απεικονίζει ένα μόνο slice του γονάτου.

Πρώτης Γενιάς CT

Ο πρώτης γενιάς αξονικός τομογράφος αποτελείται από έναν x-ray tube ο οποίος εκπέμπει μία λεπτή ακτίνα X, η οποία στη συνέχεια συλλέγεται από ένα x-ray cone μέσω ενός παραλληλιστή με οπές (pinhole collimator). Ένας μοναδικός ανιχνευτής τοποθετείται στην απέναντι μεριά του x-ray tube, και μετακινείται συγχρονισμένα μαζί του. Η συγκεκριμένη μετακίνηση είναι γραμμική και επαναλαμβάνεται για διαφορετικές γωνίες προβολής γ (projection angles). Η ακτίνα X, ανάλογα με τα χαρακτηριστικά των δομών του σώματος που διαπερνάει η ένταση της, εξασθενεί. Η συγκεκριμένη εξασθένιση μετρείται από τον ανιχνευτή και αποθηκεύεται.

Για κάθε γωνία προβολής το ραδιογράφημα είναι μίας διάστασης. Προκειμένου να παραχθεί η εικόνα πρέπει επαναληφθεί η διαδικασία για όλες τις γωνίες από 0° έως 180° . Ο Hounsfield το 1969 πραγματοποίησε αυτή τη διαδικασία που διήρκεσε 9 μέρες, λαμβάνοντας έτσι το πρώτο slice χωρίς averaging ή blurring [2].

Δεύτερης Γενιάς CT

Οι αξονικοί τομογράφοι δεύτερης γενιάς αποτελούνται από έναν x-ray tube ο οποίος εκπέμπει μια ριπιδοειδή δέσμη (fan-beam) (από την πηγή η ακτίνα σιγά σιγά ανοίγει με αποτέλεσμα να καταλήγει σε μεγαλύτερη επιφάνεια) και μία συστοιχία ανιχνευτών αποτελούμενη από 30 στοιχεία. Η γωνία της δέσμης είναι περίπου 30° και δεν μπορεί να καλύψει ολόκληρο το σώμα του ασθενή με μία σάρωση. Η διαδικασία πρέπει να επαναληφθεί για διαφορετικές γωνίες προβολής, και συνεπώς η λυχνία ακτίνων X και οι ανιχνευτές πρέπει να μετακινούνται γραμμικά.

Η γωνία της δέσμης, μείωσε σημαντικά τον χρόνο για την απεικόνιση ενός αντικείμενου, όχι όμως αρκετά για την απεικόνιση της κοιλιακής χώρας καθώς με την πάροδο του χρόνου μετακινείται και ο ίδιος ο ασθενής και τα όργανα του, δημιουργώντας artifacts στην εικόνα.

Οι τομογράφοι πρώτης και δεύτερης γενιάς χρησιμοποιούντουσαν κυρίως για την απεικόνιση του κρανίου [2].

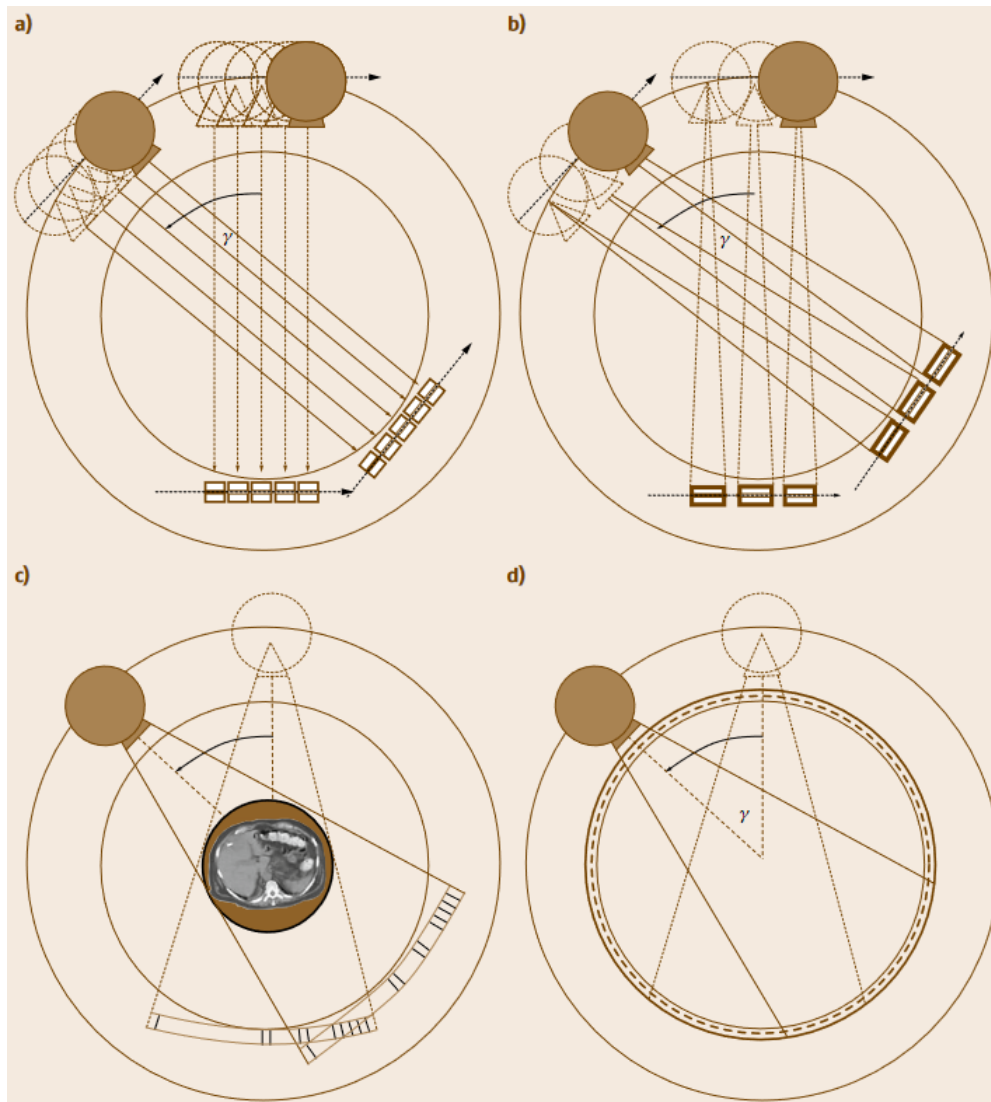
Τρίτης Γενιάς CT

Στους αξονικούς τομογράφους τρίτης γενιάς βελτιώνονται τα εξαρτήματα των τομογράφων της δεύτερης γενιάς. Η γωνία της ριπιδοειδούς δέσμης αυξάνεται και φτάνει τις γωνίες από 40° έως 60° , η συστοιχία ανιχνευτών έχει σχήμα τόξου και αποτελείται από 400 έως 1000 στοιχεία. Με αυτό τον τρόπο είναι δυνατόν να ακτινογραφηθεί ολόκληρος ο κορμός του ασθενή για κάθε γωνία προβολής γ χωρίς να απαιτείται γραμμική μετακίνηση του x-ray tube.

Με τις παραπάνω βελτιώσεις αλλά και εφαρμόζοντας συνεχόμενη περιστροφή, χωρίς διακοπές για γραμμική μετατόπιση, μειώνεται σημαντικά ο χρόνος που χρειάζεται για την απεικόνιση της εικόνας.

Τέταρτης Γενιάς CT

Οι αξονικοί τομογράφοι τέταρτης γενιάς διαφέρουν από τους ίδιους της τρίτης γενιάς στο ότι έχουν έναν στάσιμο δακτύλιο ανιχνευτών (detector ring) ακτίνων X με έως 5000 στοιχεία. Ο x-ray tube περιστρέφεται είτε μέσα από τον detector ring είτε απ' έξω. Στην περίπτωση που περιστρέφεται απ' έξω ο δακτύλιος παίρνει δυναμικά μία κλίση έτσι ώστε οι ακτίνες να μην διαπερνάνε τα ηλεκτρονικά κυκλώματα πίσω από το δακτύλιο και να περνάνε μόνο μέσα από τον ασθενή το κρεβάτι καταλήγοντας έπειτα στους ανιχνευτές.

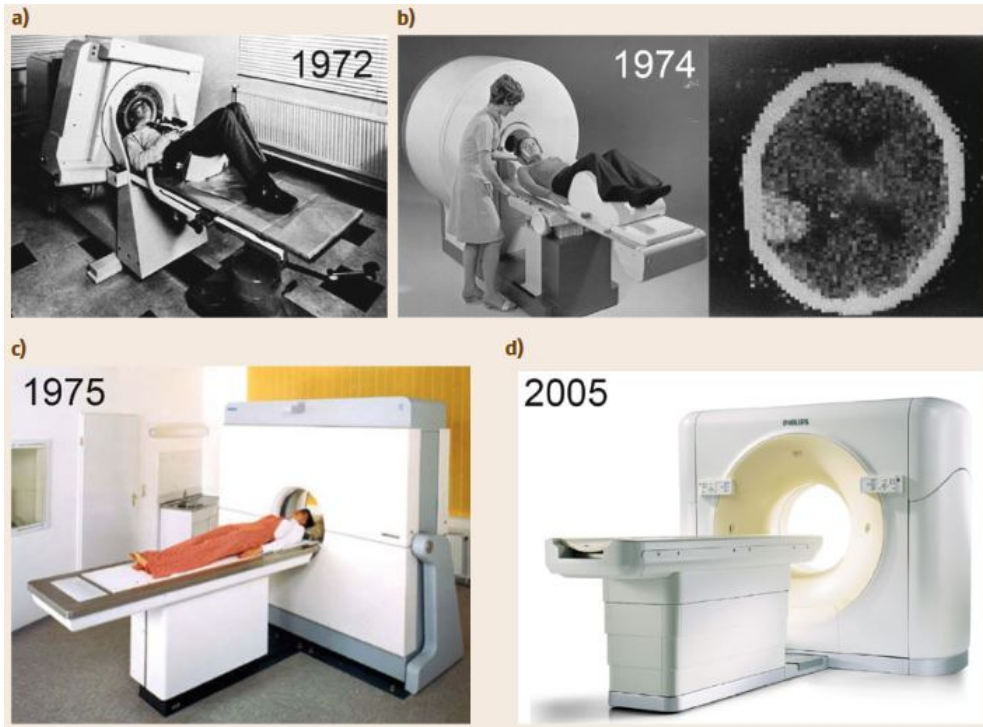


Σχήμα 2.3: Στην συγκεκριμένη εικόνα παρουσιάζονται οι 4 γενιές αξονικών τομογράφων (a) Ο πρώτης γενιάς CT εκπέμπει μία λεπτή ακτίνα σε έναν μόνο ανιχνευτή (b) Ο δεύτερης γενιάς CT εκπέμπει ριποειδή δέσμη σε συστοιχία ανιχνευτών (c) Ο τρίτης γενιάς CT εκπέμπει δέσμη, η οποία καλύπτει ολόκληρο τον κορμό του ασθενή, σε ακόμα περισσότερους ανιχνευτές (d) Ο τέταρτης γενιάς CT διαθέτει μια στάσιμο δακτύλιο ανιχνευτών.

Έκτης Γενιάς CT

Η τεχνολογία slip ring αποτελεί μία τεράστια αναβάθμιση στους αξονικούς τομογράφους τρίτης γενιάς, και γι' αυτό οι τομογράφοι που την φέρουν χαρακτηρίζονται ως έκτης γενιάς από τους Bushberg et al. Ένα σημαντικό πρόβλημα του x-ray tube είναι ότι πρέπει να τροφοδοτείται συνέχεια με ρεύμα. Η συγκεκριμένη παροχή γίνεται μέσω καλωδίων τα οποία περιορίζουν σημαντικά την κίνηση αυτής της λυχνίας ακτίνων X. Έτσι, η ταχύτητα με την οποία μπορεί να περιστρέφεται το μηχάνημα περιορίζεται αρκετά και επίσης πρέπει να σταματήσει και ξανααρχίσει από συγκεκριμένες γωνίες.

Με την τεχνολογία slip ring η ενέργεια παρέχεται μέσω sliding contacts μεταξύ του εξωτερικού της μονάδας δειγματοληψίας (gantry) και του περιστρεφόμενου τμήματος [2].



Σχήμα 2.4: Η εξέλιξη των αξονικών τομογράφων. Οι (a) και (b) αποτελούν αξονικούς τομογράφους πρώτης γενιάς και πραγματοποιούν τομογραφίες μόνο στο κεφάλι. Ο (c) είναι δεύτερης γενιάς και ο (d) είναι τρίτης γενιάς.

2.1.2 Ανακατασκευή της εικόνας

Το βασικό πρόβλημα της αξονικής τομογραφίας είναι η ανακατασκευή ενός αντικειμένου από τη σκιά του ή αλλιώς από τις προβολές του. Από τους τρίτης γενιάς και έπειτα CT-scanners η ακτίνα x-ray καλύπτει ολοκληρωτικά το αντικείμενο που εξετάζεται είτε είναι κάποιος ασθενής είτε ένα κρανίο από αρχαιολογικούς χώρους. Όπως είναι αναμενόμενο όταν η ακτίνα διαπερνά το αντικείμενο εξασθενεί σε διάφορα σημεία.

Η "σκιά" που παράγεται είναι μόνο μίας διάστασης και δεν παρέχει χωρικές πληροφορίες για το αντικείμενο. Προκειμένου να απεικονιστούν τα τρισδιάστατα χαρακτηριστικά, είναι απαραίτητο το αντικείμενο να ακτινοβοληθεί από όλες τις κατευθύνσεις.

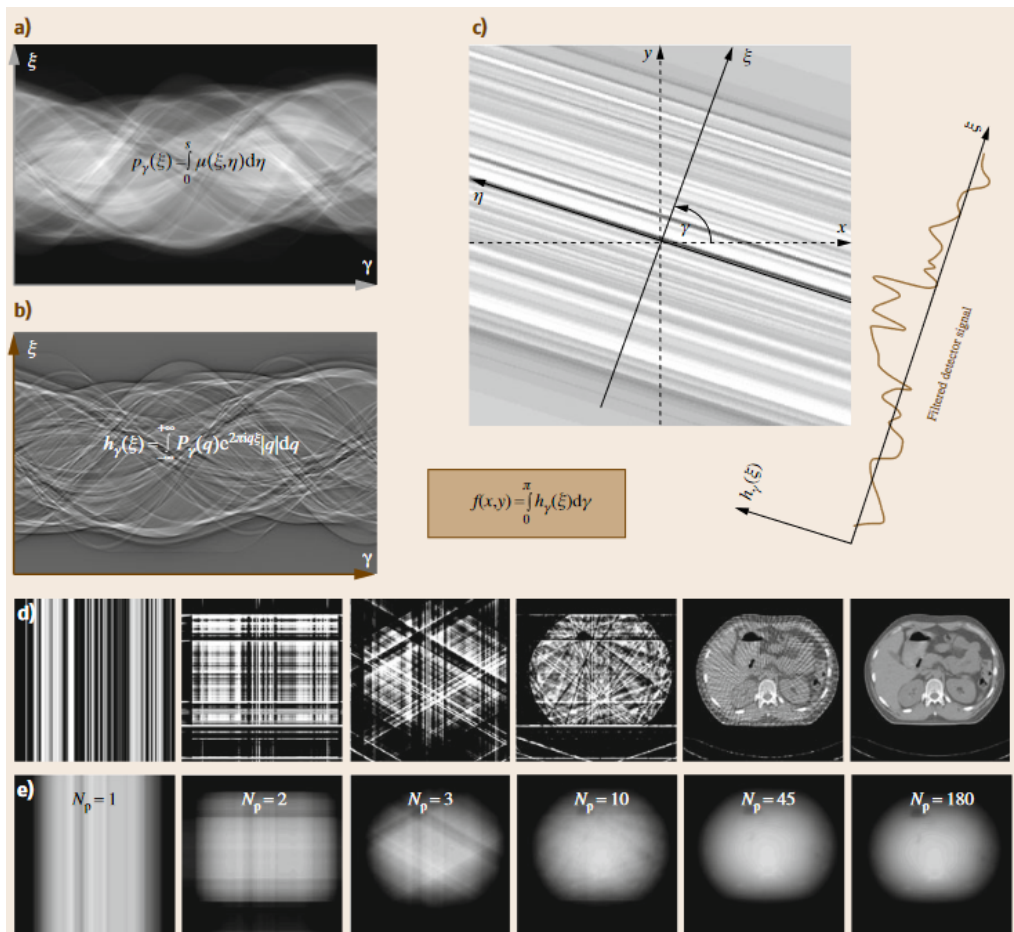
Οι ακτίνες X που χρησιμοποιούνται για την ακτινοβολήση του αντικειμένου αποτελούν ένα κομμάτι του ηλεκτρομαγνητικού φάσματος, με το φάσμα να περιλαμβάνει το ραντάρ, μικροκύματα, ορατό και υπεριώδες φως, χ και γ ακτίνες. Όταν οι ακτίνες X έρχονται σε επαφή με στερεά μεταλλική άνοδο, τα ηλεκτρόνια που εισέρχονται την άνοδο επιβραδύνουν με αποτέλεσμα να παράγουν ακτινοβολία. Επομένως, η ενέργεια της ακτινοβολίας είναι ανάλογη της ταχύτητας των ηλεκτρονίων,

η οποία εξαρτάται από το acceleration voltage.

Backprojection

Έστω, $p_{\gamma_i}(\xi)$ η συνάρτηση εξασθένησης της ακτίνας υπό γωνία προβολής γ_i με μεταβλητή την συντεταγμένη ξ . Στις δύο διαστάσεις οι συναρτήσεις $p_{\gamma_i}(\xi)$ για όλες τις γωνίες προβολής αποτελούν το Radon Space του αντικειμένου, δηλαδή είναι τα μη επεξεργασμένα δεδομένα.

Στη μέθοδο backprojection ή αλλιώς μέθοδο συνέλιξης, τα δεδομένα $p_{\gamma_i}(\xi)$ φιλτράρονται ψηφιακά. Ουσιαστικά εφαρμόζεται ένα υπερπερατό φίλτρο (high pass), χρησιμοποιώντας γραμμικά βάρη q στη συνάρτηση Fourier $P_{\gamma_i}(q)$ της συνάρτησης εξασθένησης $p_{\gamma_i}(\xi)$, λαμβάνοντας την $h_{\gamma_i}(\xi)$. Προφανώς, μόνο μία $h_{\gamma_i}(\xi)$ δεν αρκεί για την απεικόνιση του αντικειμένου. Η ανακατασκευή της εικόνας πραγματοποιείται από συνεχόμενο backprojection για κάθε γωνία προβολής από 0 έως 180.



Σχήμα 2.5: Από τα ακατέργαστα δεδομένα στην ανακατασκευασμένη εικόνα. Σι- νόγραμμα τιμών εξασθένισης ή προβολής (α), τιμές προβολής φιλτραρισμένης με υπερπερατό φίλτρο (β), οπισθοπροβολή της φιλτραρισμένης προβολής (γ), συσσώρευση όλων των οπισθοπροβολών (δ) και της εγγενούς προβολής (ε)

Τιμές Hounsfield

Στην αξονική τομογραφία, οι τιμές εξασθένησης μ παρουσιάζονται ως γκρι τιμές. Με δεδομένο αυτό, ο Hounsfield ανέπτυξε μία μέθοδο που χρησιμοποιείται ακόμη. Οι τιμές εξασθένησης μετατρέπονται σε αδιάστατη κλίμακα και σχετίζονται με τιμή εξασθένησης του νερού.

$$CT_{value} = \frac{\mu - \mu_{water}}{\mu_{water}} 1000 \quad (2.1)$$

$$|CT_{value}| = HU \quad (2.2)$$

όπου HU είναι ένα Hounsfield unit.

Σε αυτή τη κλίμακα το -1000 ανήκει στον αέρα και η τιμή 0 στο νερό. Θεωρητικά το πεδίο ορισμού είναι όλοι οι πραγματικοί αριθμοί, ωστόσο πρακτικά οι τιμές δεν ξεπερνάνε τα 3000HU. Αυτές οι τιμές μπορούν να αναπαρασταθούν αρκετά καλά από 12bit εικόνες σε γκρι κλίμακα. Όμως, η ανθρώπινη όραση δεν έχει την δυνατότητα να ξεχωρίσει και τα 4000 γκρι στάδια. Σε σύγχρονες έρευνες οι άνθρωποι μπορούν να ξεχωρίσουν από 700 έως 900 αποχρώσεις του γκρι. Γι' αυτό το λόγο στην πράξη παρουσιάζονται μόνο 256-512 αποχρώσεις στις συσκευές απεικόνισης. Ακόμη, για να γίνονται αντιληπτές οι διαφορές μεταξύ των οργάνων χρησιμοποιείται η γραμμική εξίσωση.

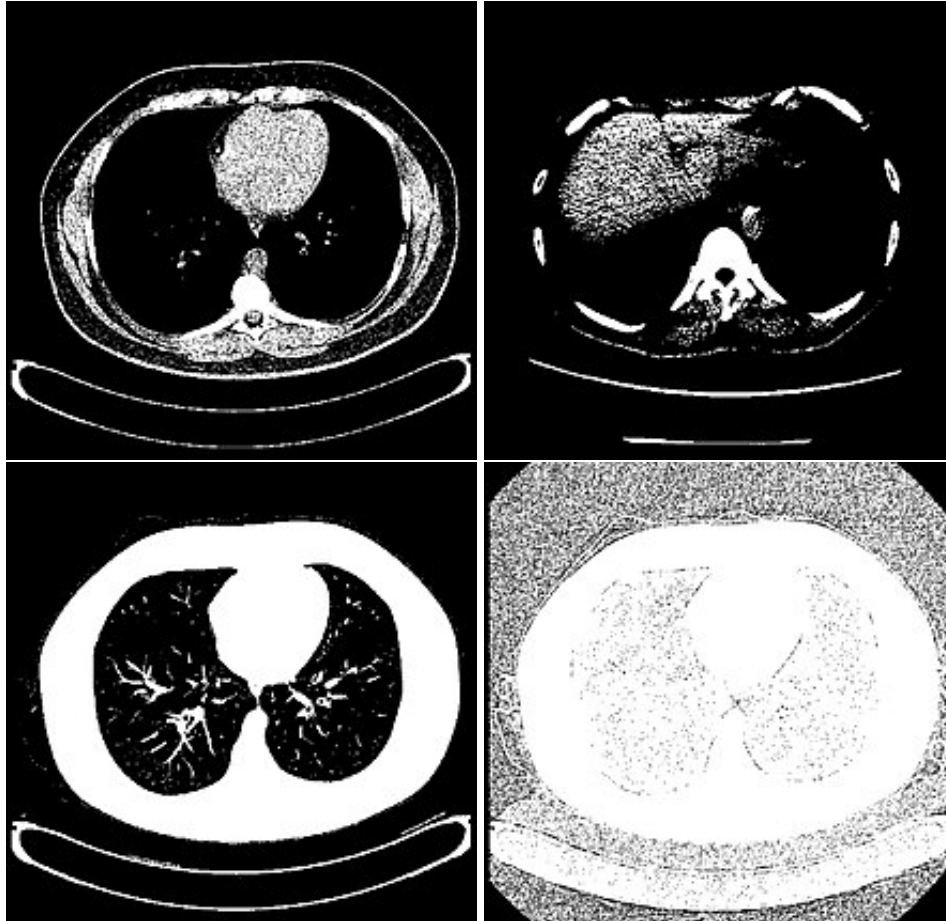
$$G = 255 \begin{cases} 0 & \text{for CT value} \leq WL - \frac{WW}{2} \\ WW^{-1} & \text{CT value} - WL + \frac{WW}{2} \\ 1 & \text{for CT value} \geq WL + \frac{WW}{2} \end{cases} \quad (2.3)$$

Επιπροσθέτως, για την αναπαράσταση της επιφάνειας, είναι απαραίτητη η επιλογή μίας τιμής threshold που θα αναπαριστά την επιφάνεια. Ο κάθε ειδικός έχει την δυνατότητα επιλογής όποιας τιμής επιθυμεί, αρκεί τα διαφορετικά αντικείμενα που θα απεικονίζονται να έχουν εμφανίσημα διαφορετικές Hounsfield τιμές μεταξύ τους. Στο Σχήμα 2.6 παρουσιάζονται αξονικές του θώρακα επιλέγοντας κάθε φορά διαφορετικές τιμές threshold.

Τρισδιάστατη απεικόνιση

Οι ιατρικές εφαρμογές της αξονικής τομογραφίας χρειάζονται τρισδιάστατες εικόνες. Ένας αρχικός τρόπος δημιουργίας αυτών των εικόνων είναι η ένωση πολλών δισδιάστατων εικόνων. Μία συνηθισμένη τεχνική είναι η μετακίνηση του ασθενή στο εγκάρσιο επίπεδο ή αλλιώς στον άξονα z του scanner. Το κρεβάτι τότε σταματάει και αποκτούνται τα μη επεξεργασμένα δεδομένα ενός slice. Η συγκεκριμένη τεχνική ονομάζεται secondary reconstruction.

Μία εναλλακτική επεικονιστική μέθοδος τρισδιάστατων δεδομένων είναι η MPR (multiplanar reformatting).



Σχήμα 2.6: CT scans του θώρακα για διαφορετικές τιμές HU threshold. (a) Για 0 HU (b) Για 60 HU (c) Για -700 HU (d) Για -1000 HU.

2.2 Πυρηνική Ιατρική

Η Πυρηνική Ιατρική χρησιμοποιεί ραδιοφάρμακα και κατάλληλους ανιχνευτές ακτινοβολίας για τη διάγνωση και θεραπεία μεγάλης ποικιλίας καλοήθων και κακοήθων νεοπλασματικών παθήσεων.

Τα ραδιοφάρμακα αποτελούν σύνθεση ενός ραδιοϊσοτόπου και ενός φαρμακευτικού τμήματος το οποίο καθορίζει την εξειδικευμένη κινητική και κατανομή του ραδιοφαρμάκου στους ιστούς. Το ραδιοϊσότοπο (ή ραδιονουκλίδιο), είναι η ραδιενεργή μορφή ενός στοιχείου που διασπάται εκπέμποντας ακτινοβολία κυματική (φωτόνια, ακτινοβολία γ) ή σωματιδιακή (σωματίδια β ή α). Ο χρόνος που απαιτείται για να διασπασθεί κατά 50% μια ραδιενεργή ποσότητα ονομάζεται χρόνος ημίσειας ζωής, ($T_{1/2}$) και είναι χαρακτηριστικός του ραδιοϊσοτόπου. Η εκπεμπόμενη από το ραδιοϊσότοπο ακτινοβολία καθορίζει και την επιλογή του ραδιοφαρμάκου ως θεραπευτικού ή διαγνωστικού, επειδή ραδιοφάρμακα με ραδιοϊσότοπο που εκπέμπει σωματιδιακή ακτινοβολία χρησιμοποιούνται για θεραπεία, ενώ αυτά, των οποίων το ραδιοϊσότοπο εκπέμπει ακτινοβολία γ για απεικόνιση ή μετρήσεις.

2.2.1 PET

Η τομογραφία εκπομπής ποζιτρονίων αποτελεί μία μη επεμβατική, ποσοτική απεικονιστική τεχνική που κάνει ορατές βιοχημικές και βιολογικές διαδικασίες. Πρόκειται για εξέταση που απεικονίζει την κατάσταση του μεταβολισμού του ασθενή. Η απεικόνιση του PET βασίζεται στην ανίχνευση εκμηδενισμένων φωτονίων που παράγονται μετά από εκπομπή ποζιτρονίων από ένα ραδιοσημασμένο ιχνηθέτη, που ακολουθεί ένα συγκεκριμένο βιοχημικό μονοπάτι. Γενικά, η απεικόνιση PET περιλαμβάνει ισότοπα, όπως F-18, C-11, N-13, Ga-68 που εκπέμπουν, ποζιτρόνια.

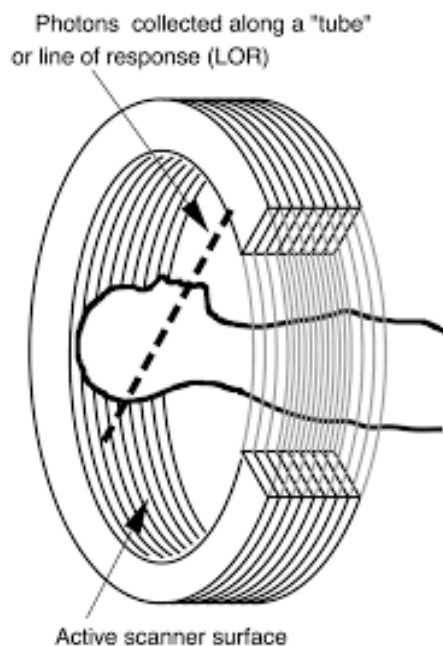
Το ποζιτρόνιο είναι ένα αντισωματίδιο ηλεκτρονίου με ίδια μάζα και αντίθετο φορτίο, με αυτά του ηλεκτρονίου. Μετά την εκπομπή του, το ποζιτρόνιο ταξιδεύει με ταχύ του υλικού και του νέφους ηλεκτρονίων μέχρι να χάσει σχεδόν όλη του την ενέργεια. Τότε το ποζιτρόνιο αλληλεπιδρά με ένα ελεύθερο ηλεκτρόνιο και αλληλοεξουδετερώνονται (εξαύλωση). Αυτό έχει ως αποτέλεσμα την παραγωγή δύο φωτονίων γάμμα ακτινοβολίας, με ενέργεια 511keV το κάθε ένα, ίση με την μάζα ηλεκτρονίων και ποζιτρονίων που εκπέμπονται προς αντίθετες κατευθύνσεις.

Απεικόνιση

Για την απεικόνιση, χρησιμοποιούνται ανιχνευτές οι οποίοι ανιχνεύουν φωτόνια και ονομάζονται φωτοπολλαπλασιαστές. Πολλαπλοί ανιχνευτές τοποθετούνται διαδοχικά σε κυκλική διάταξη, σχηματίζοντας έναν δακτύλιο, και συνδέονται μεταξύ τους μέσω ενός κυκλώματος που ονομάζεται κύκλωμα σύμπτωσης (coincidence circuit). Κατά την διάρκεια του φαινομένου της εξαύλωσης, εκπέμπονται δύο φωτόνια με αντίθετη φορά, συγκεκριμένα γ-φωτόνια 511KeV το κάθε ένα. Επομένως, όταν δύο αντιδιαμετρικά τοποθετημένοι ανιχνευτές ανιχνεύσουν αυτό το σήμα, σχεδόν ταυτόχρονα (εντός ενός μικρού χρονικού παραθύρου το οποίο ονομάζεται συμπτωτικό παράθυρο), γνωρίζουν ότι έχει συμβεί εξαύλωση κάπου στην ευθεία που τους ενώνει (ευθεία απόκρισης, line of response, LOR). Στη συνέχεια, από το ίδιο σημείο θα συμβούν και άλλα φαινόμενα εξαύλωσης, τα οποία θα ενεργοποιήσουν διαφορετικούς ανιχνευτές καθώς η κατεύθυνση των φωτονίων είναι τυχαία. Τώρα ο εντοπισμός του σημείου προέλευσης του φαινομένου γίνεται δυνατός από τον υπολογισμό της εγκάρσιας τομής όλων ευθειών.

Οι αντιδιαμετρικές ακτίνες όπως ειπώθηκε προηγουμένως φθάνουν στους ανιχνευτές με μικρή χρονική διαφορά. Οι παλαιότεροι κρύσταλλοι οι οποίοι έχουν μεγάλους χρόνους απόκρισης έκαναν δύσκολο τον υπολογισμό αυτής της διαφοράς. Με τους νεότερους κρυστάλλους γίνεται δυνατή η αρχή λειτουργίας της TOF PET, κατά την οποία εντοπίζεται η θέση εξαύλωσης στην πορεία της LOR από το χρόνο διαδρομής των ακτίνων γ (time of flight, TOF).

Όπως είναι αναμενόμενο πολλές φορές η ευθεία απόκρισης που εντοπίζεται είναι λανθασμένη κάτι το οποίο υποβαθμίζει την εικόνα. Αυτό οφείλεται είτε σε σκέδα-



Σχήμα 2.7: PET scanner με τους δακτύλιους να εντοπίζουν τις εξαϋλώσεις [3].

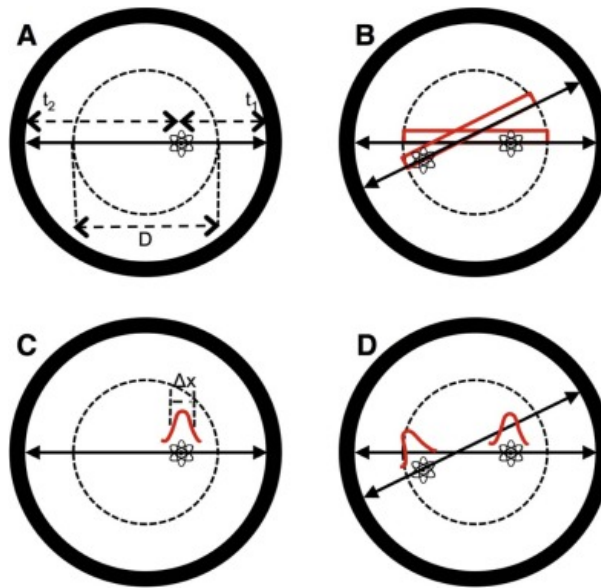
ση είτε σε τυχαία ταυτόχρονη πρόσπτωση γ-φωτονίων σε δύο απέναντι ανιχνευτές. Μειώνοντας την τιμή του συμπτωτικού παραθύρου είναι πιθανή η απόρριψη τέτοιων τυχαίων φαινομένων, όμως η πολύ μικρή τιμή απορρίπτει και αληθείς συμπτώσεις. Ακόμη, υπάρχει αβεβαιότητα για την ακριβή θέση του ισοτόπου καθώς τα ποζιτρόνια διανύουν μεγάλη απόσταση μέχρι να εξαϋλωθούν με κάποιο ηλεκτρόνιο. Όση περισσότερη ενέργεια έχει το ποζιτρόνιο τόσο μεγαλύτερη απόσταση διανύει.

Συχνά εμφανίζονται πολλά λάθη στην απεικόνιση λόγω ατελειών του μηχανήματος. Οι ανιχνευτές κρυστάλλου, ο τύπος και ο αριθμός φωτοπολλαπλασιαστών, τα ηλεκτρικά κυκλώματα καθώς και ο τρόπος λήψης δεδομένων επηρεάζουν την διακριτική ικανότητα.

Για την λήψη δεδομένων χρησιμοποιείται περισσότερο ο ποσοτικός δείκτης SUV (standardized uptake value). Η SUV εκφράζει τον λόγο της μετρούμενης ραδιενεργού ενεργότητας σε μία περιοχή της εικόνας, σε μία δεδομένη χρονική στιγμή μετά την χορήγηση του ραδιοϊχνηθέντη, προς την χορηγείσα δόση, αναγόμενο στο σωματικό βάρος του ασθενούς.

$$SUV = \frac{C_{img}(t)}{ID/BW} \quad (MBq/mL) \quad (2.4)$$

Όπου $C_{img}(t)$ είναι η ενεργότητα σε χρόνο t , BW το σωματικό βάρος του ασθενή, ID η χορηγούμενη δόση.



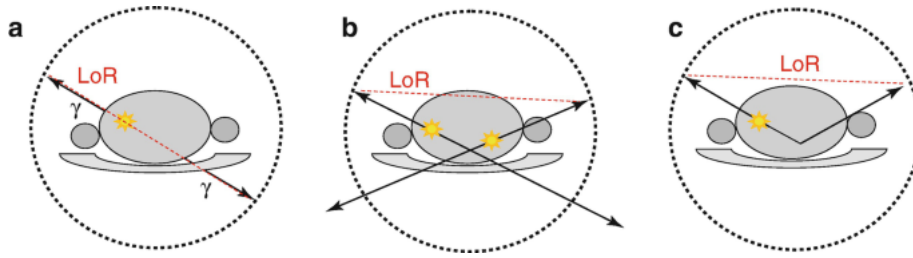
Σχήμα 2.8: Σχηματική παρουσίαση της αρχής λειτουργίας TOF PET. (A) Το φαινόμενο εξαΰλωσης συμβαίνει σε τυχαία σημεία με απόσταση d από το κέντρο αντικειμένου με διάμετρο D . Οι χρόνοι σύμπτωσης είναι t_1 και t_2 . (B) Ένας όχι TOF-scanner δημιουργεί μία ομοιόμορφη κατανομή πιθανότητας πάνω στην LOR για το που βρίσκεται το επιθυμητό σημείο δημιουργώντας με αυτό τον τρόπο θόρυβο σε κάποιο κομμάτι της εικόνας που κατασκευάζεται. (C) Σε έναν TOF-scanner η θέση εξαΰλωσης περιορίζεται σε ένα μικρό τμήμα της LOR με ακρίβεια που καθορίζεται από Γκαουσιανή κατανομή Δx . (D) Με τον αναβαθμισμένο εντοπισμό των σημείων πάνω στις LOR μειώνεται ο θόρυβος που δημιουργόταν προηγουμένως στη (B)

[4].

Λήψη Δεδομένων

Η ψηφιοποίηση του PET ξεκίνησε με την ανάπτυξη καινούριων κρυστάλλων, ημιαγωγών και την χρήση φωτοδιόδων και φωτοπολλαπλασιαστών στερεάς κατάστασης. Οι σύγχρονες PET κάμερες φέρουν πολλαπλούς επάλληλους δακτυλίους, που περιέχουν ομάδες ανιχνευτών, για τον ταυτόχρονο εντοπισμό πολλαπλών εγκάρσιων τομών. Μεταξύ των δακτυλίων υπάρχουν φύλλα μολύβδου (διαφράγματα), τα οποία περιορίζουν την παρείσφηση γ -φωτονίων από γειτονικά εγκάρσια επίπεδα σε άλλους εκτός ενός δακτυλίου. Για την αύξηση της ευαισθησίας ανίχνευσης διάταξης, τα φύλλα μολύβδου αφαιρούνται και χρησιμοποιούνται μέθοδοι ηλεκτρονικής κατεύθυνσης (electronic collimation), χωρίς να μειώνεται σημαντικά η διακριτική ικανότητα.

Η σάρωση του σώματος γίνεται, μετακινώντας την εξεταστική κλίνη οριζόντια κατά βήματα (beds) μέχρις ότου να περάσει από ολόκληρο το σώμα. Ανάλογα με το πλήθος των δακτυλίων και την ευαισθησία των ανιχνευτών η εξέταση μπορεί να διαρκήσει από 5 έως 20 λεπτά.



Σχήμα 2.9: Στο σχήμα παρουσιάζονται οι 3 περιπτώσεις σύμπτωσης. Στην (a) αληθής σύμπτωση, (b) τυχαία σύμπτωση και (c) σύμπτωση λόγω σκέδασης [5].

Ανιχνευτής	Πυκνότητα (g/cm^3)	Ενεργός Ατομικός Αριθμός (Z)	Απόδοση (KeV)	Χρόνος Απόκρισης (ns)
NaI	3.67	51	45	230
BGO	7.13	74	4	300
GSO	6.71	59	8	60
LSO	7.40	66	26	40
LYSO	5.37	54	33	36

Πίνακας 2.1: Τα χαρακτηριστικά κρυσταλλικών ανιχνευτών PET, καθώς και του NaI που χρησιμοποιείται στην γ -κάμερα.

2.3 PET/CT

Κύριο πρόβλημα της πυρηνικής φυσικής είναι η έλλειψη ανατομικών στοιχείων στην απεικόνιση. Συγκεκριμένα στο PET τα σημεία εξαύλωσης που εντοπίζονται είναι πολύ δύσκολο να προσδιοριστούν στο χώρο. Το πρόβλημα αυτό το επιλύει ο συνδυασμός του PET με την αξονική τομογραφία (CT) η οποία εξυπηρετεί κυρίως εντοπιστικούς σκοπούς. Στα πρώτα στάδια γίνονταν προσπάθειες συνδυασμού των δύο εικόνων ενώ οι εξετάσεις πραγματοποιούνταν σε διαφορετικά μηχανήματα. Ωστόσο, η ευθυγράμμιση των δύο εξετάσεων είναι πολύ δύσκολη, καθώς τα μηχανήματα μπορεί να είχαν διαφορετικές γωνίες στα κρεβάτια και ο ασθενής καθώς και τα όργανα του μετακινούνται κατά την διάρκεια των αρκετών λεπτών που διαρκεί η εξέταση. Η ανάγκη συνδυασμού των δύο μεθόδων οδήγησε στην ενσωμάτωση των δύο μηχανών, γ -κάμερα και PET-κάμερα σε ένα μηχανήμα τον PET/CT scanner. Με την χρήση αυτού του σαρωτή οι νέες εικόνες εμφανίζονται στο ίδιο μέγεθος και πλήρως αντιστοιχισμένες μεταξύ τους, και επομένως η ένωση τους είναι εύκολη. Ακόμη, με αυτό το τρόπο διορθώνεται και η ιστική εξασθένηση της ακτινοβολίας.

Ο PET/CT scanner στην αρχική του μορφή, ο αξονικός τομογράφος, με περιστρεφόμενη λυχνία και μία διάταξη ανιχνευτών, μίας ή ολίγων ταυτόχρονων τομών, ενσωματωνόταν στο στηρικτικό σύστημα της γ -κάμερας. Αυτή η διάταξη καθυστερεί την ολοκλήρωση της αξονικής τομογραφίας και ως απόρροια μειώνεται η ποιότητα των εικόνων λόγω κίνησης, τόσο του ασθενούς όσο και των οργάνων του.

Οι σύγχρονοι PET/CT scanners διαθέτουν ξεχωριστό στηρικτικό σύστημα για την αξονική τομογραφία, η οποία είναι ελικοειδής πολλαπλών τομών (6,8,16 ή και 64).

Ωστόσο, αν και η CT τώρα είναι ταχεία, διαρκεί λιγότερο από ένα λεπτό, η PET παραμένει αργή όπως και με ξεχωριστό scanner από 10 έως 20 λεπτά. Οι δύο εξετάσεις δεν πραγματοποιούνται ταυτόχρονα, αλλά ξεχωριστά. Δεν υπάρχει συγκεκριμένη σειρά για το ποια λήψη θα ολοκληρωθεί πρώτη.



Σχήμα 2.10: Το PET/CT Discovery IQ 3-ring

Κεφάλαιο 3

Βαθιά Μάθηση

3.1 Νευρωνικά Δίκτυα

Η έμπνευση για κάθε νευρωνικό δίκτυο προέρχεται από την βιολογία. Οι περισσότεροι ζώντες οργανισμοί έχουν ένα νευρικό σύστημα, το οποίο είναι υπεύθυνο για διεργασίες όπως η μάθηση και η μνήμη. Το νευρικό σύστημα των οργανισμών αποτελείται από πολλά νευρικά δίκτυα τα οποία είναι εξειδικευμένα για την κάθε διεργασία και κάθε δίκτυο συγκροτείται από ένα μεγάλο αριθμό νευρώνων (neurons). Οι νευρώνες συνεχώς και ασταμάτητα επεξεργάζονται πληροφορίες, λαμβάνοντας και στέλνοντας ηλεκτρικά σήματα σε άλλους νευρώνες. Όπως γίνεται αντιληπτό οι διεργασίες που επιτελούνται από τα βιολογικά νευρικά δίκτυα είναι πολύ περίπλοκες και χρήσιμες, και ο εγκέφαλος τις εκτελεί με ελάχιστη προσπάθεια. Μία τέτοια απλή διεργασία είναι η αναγνώριση ενός αριθμού σε μία εικόνα. Έχοντας δει αρκετές εικόνες αριθμών, ο άνθρωπος έχει μάθει τις διαφορές μεταξύ τους και μπορεί να τους ξεχωρίσει.

Με αυτό τον τρόπο εισάγεται η έννοια των νευρωνικών δικτύων. Ένας αλγόριθμος μπορεί να μαθαίνει όπως ακριβώς και ο άνθρωπος απλά δείχνοντας του αρκετή πληροφορία.

3.1.1 Δομή

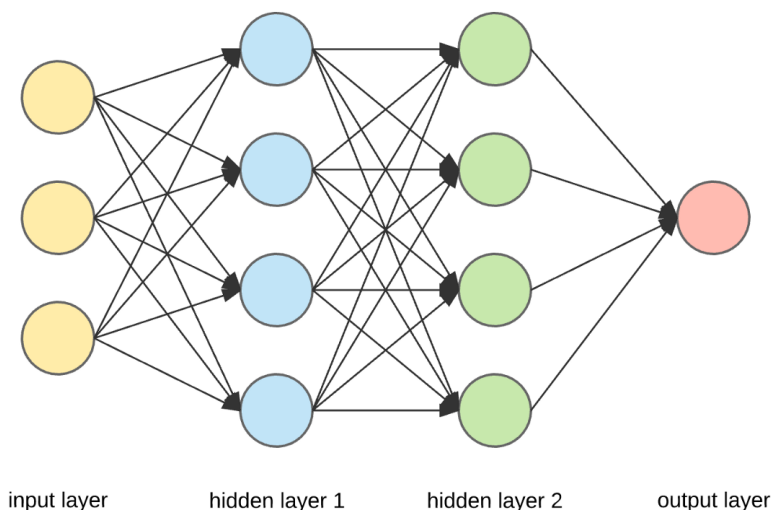
Το βασικό δομικό στοιχείο ενός νευρωνικού δικτύου είναι ο νευρώνας (neuron). Οι νευρώνες τοποθετούνται κάθετα ο ένας πάνω από τον άλλον, δημιουργώντας ένα επίπεδο (Layer). Κάθε επίπεδο μπορεί να συνδεθεί με ένα άλλο επίπεδο και επομένως ο κάθε νευρώνας μπορεί και αλληλεπιδρά με τους υπόλοιπους, έτσι ώστε άλλους να τους ενεργοποιεί λιγότερο, και άλλους περισσότερο.

Υπάρχουν τρία είδη επιπέδων στα νευρωνικά δίκτυα . [6]

- **Input Layer:** Είναι το πρώτο επίπεδο, το οποίο δέχεται την είσοδο και την προωθεί στο υπόλοιπο δίκτυο.
- **Hidden Layer:** Το κρυφό επίπεδο βρίσκεται ανάμεσα στο πρώτο και το τελευταίο επίπεδο του δικτύου. Συνήθως, είναι ένα ή περισσότερα σε αριθμό και

είναι υπεύθυνα για την καλή απόδοση των νευρωνικών δικτύων. Εκτελούν πολλές μαθηματικές πράξεις και οι συνδυασμοί τους σχηματίζουν μαθηματικές εξισώσεις ώστε να βγάλουν στην έξοδο το αναμενόμενο αποτέλεσμα.

- **Output Layer:** Το τελευταίο επίπεδο του δικτύου είναι το επίπεδο εξόδου. Στο συγκεκριμένο επίπεδο ο πιο ενεργοποιημένος νευρώνας είναι το τελικό αποτέλεσμα του προβλήματος.



Σχήμα 3.1: Είδη Επιπέδων Νευρωνικών Δικτύων. Κάθε κύκλος αποτελεί έναν νευρώνα.

Κάθε νευρώνας περιέχει έναν αριθμό. Ο αριθμός αυτός ονομάζεται τιμή ενεργοποίησης, και τον αποκτά με την λήψη πολλαπλών σημάτων ως εισόδους, τον πολλαπλασιασμό κάθε εισόδου με ένα βάρος, την άθροιση των σημάτων αυτών και την είσοδο του αθροίσματος σε μία μη-γραμμική συνάρτηση η οποία δίνει και το τελικό αποτέλεσμα. Ας θεωρήσουμε ότι κυμαίνεται μεταξύ των τιμών 0 και 1. Όσο πιο κοντά στο 1 είναι η τιμή, τόσο περισσότερο ενεργοποιημένος είναι ο νευρώνας και αντίστροφα.

Ένας νευρώνας αποτελείται, λοιπόν, από 3 βασικά μέρη [6]:

- Τις συνάψεις, μέσω των οποίων συνδέεται με άλλους νευρώνες, και οι οποίες χαρακτηρίζονται από μία τιμή βάρους, η οποία είναι ξεχωριστή για την κάθε μία. Έστω ο νευρώνας j στο επίπεδο $k - 1$ με έξοδο o_j^{k-1} συνδέεται με τον νευρώνα i στο επίπεδο k μέσω της σύναψης με βάρος w_{ji}^k . Ο τελευταίος νευρώνας δέχεται ως είσοδο την τιμή $w_{ji}^k o_j^{k-1}$.
- Έναν αθροιστή ο οποίος αθροίζει τις εξόδους όλων των συνάψεων, τα σήματα των οποίων δέχεται ως εισόδους ο νευρώνας.
- Τη μη-γραμμική συνάρτηση ενεργοποίησης η οποία ορίζει την τελική τιμή ενεργοποίησης του νευρώνα και την φράσσει μεταξύ δύο πεπερασμένων τιμών (στο παραπάνω παράδειγμα 0 και 1).

Συμπληρωματικά, τις περισσότερες φορές χρησιμοποιείται και μία τιμή πόλωσης b , η οποία αθροίζεται στην τελική τιμή του αθροιστή, και αυτός ο αριθμός δίνεται ως είσοδος στην συνάρτηση ενεργοποίησης του νευρώνα.

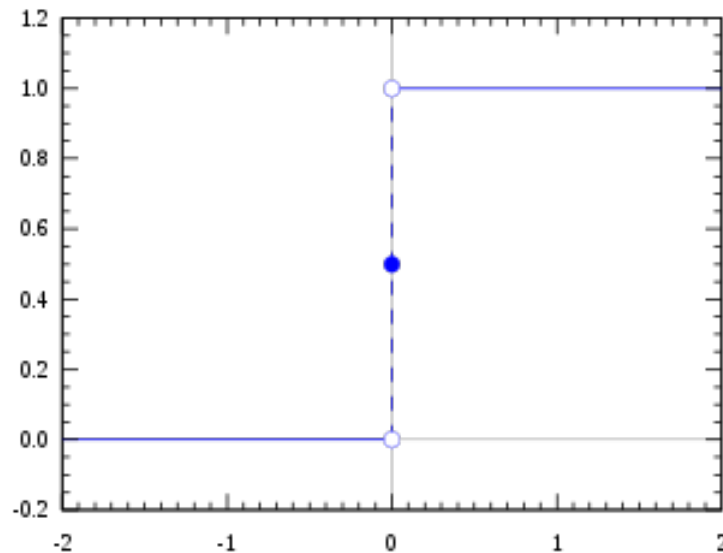
Η συνάρτηση Ενεργοποίησης

Ο κύριος σκοπός της συνάρτησης ενεργοποίησης είναι να ελέγχει την έξοδο του κάθε νευρώνα φράζοντας την έξοδο του αθροιστή μέσα σε ένα καθορισμένο όριο τιμών. Επομένως, η συνάρτηση είναι απαραίτητα μη-γραμμική καθώς στην αντίθετη περίπτωση οι τιμές των νευρώνων θα συσχετιζόντουσαν γραμμικά περιορίζοντας την πολυπλοκότητα τους. Μία απλή γραμμική συνάρτηση κάνει εύκολα overfit στα δεδομένα και δεν αποδίδει καλά σε καινούρια. Ακόμη, η μη γραμμικότητα βοηθάει στην αποφυγή του προβλήματος των εξαφανιζόμενων κλίσεων (vanishing gradient) κατά την διάρκεια της οπισθοδιάδοσης (backpropagation). Υπάρχουν πλείστες τέτοιες συναρτήσεις, ωστόσο οι πιο βασικές είναι οι εξής:

- Heavyside Step Function, Βηματική Συνάρτηση

Η βηματική συνάρτηση αποτελεί την απλούστερη συνάρτηση ενεργοποίησης. Μπορεί να ερμηνευτεί ως μία ευθεία που χωρίζει τις εισόδους σε δύο κλάσεις 0 και 1 [7].

$$f(x) = \begin{cases} 1 & \text{if } \vec{w} \cdot \vec{x} + b > 0 \\ 0 & \text{if } \vec{w} \cdot \vec{x} + b \leq 0 \end{cases} \quad (3.1)$$



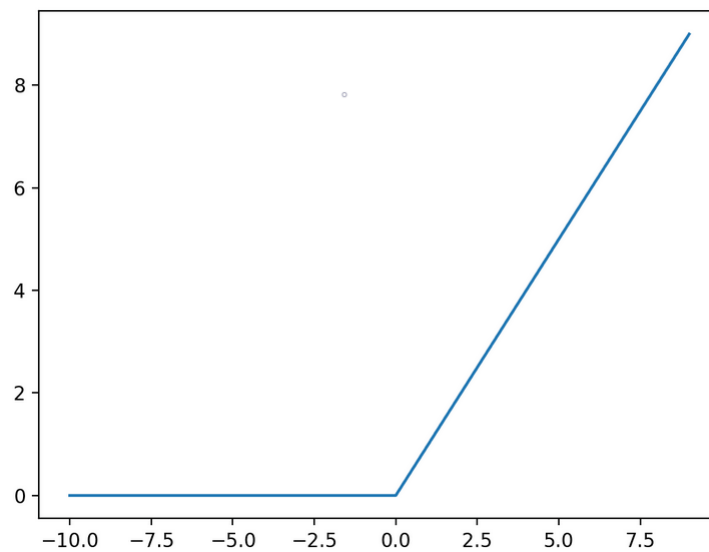
Σχήμα 3.2: Γραφική Παράσταση της Heavyside Step Function.

- Rectified Linear Unit (ReLU), Ανορθωμένη γραμμική συνάρτηση ράμπας
Η ReLU είναι από τις πιο βασικές συναρτήσεις για τη μηχανική μάθηση. Αν και είναι μη γραμμική έχει αρκετά χαρακτηρισικά γραμμικότητας και είναι

βοηθητική για τη βελτιστοποίηση γραμμικών μοντέλων. Θέτει όλες τις αρνητικές τιμές ίσες με μηδέν, στο μηδέν είναι ασυνεχής και τις θετικές ίσες με ένα. Το πιο σημαντικό χαρακτηριστικό της είναι ότι δεν διαθέτει κάποιο σημείο κορεσμού για τις θετικές εισόδους.

$$ReLU = \begin{cases} x & \text{if } \vec{w} \cdot \vec{x} + b > 0 \\ 0 & \text{if } \vec{w} \cdot \vec{x} + b \leq 0 \end{cases} \quad (3.2)$$

Ορισμένες φορές υπάρχει η πιθανότητα η συνάρτηση ράμπας να οδηγήσει ορισμένους νευρώνες σε τιμές βαρών, που να τους αποτρέπουν από το να ενεργοποιηθούν [8].



Σχήμα 3.3: Γραφική Παράσταση της ReLU.

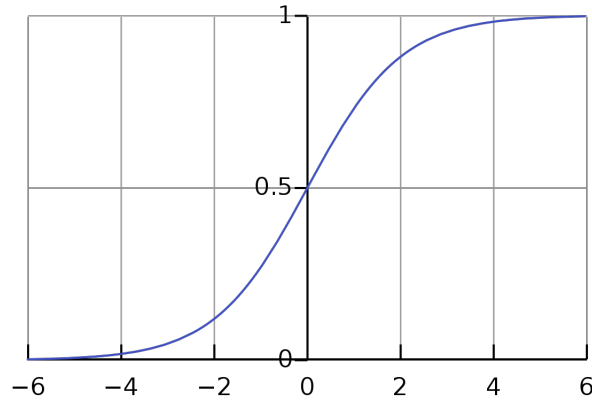
- Sigmoid Function, Σιγμοειδής Συνάρτηση

Η Σιγμοειδής Συνάρτηση πρόκειται για την πιο κοινή συνάρτηση ενεργοποίησης που χρησιμοποιείται στα νευρωνικά δίκτυα προσεγγίζοντας γραμμικά και μη γραμμικά προβλήματα. Αποτελεί ειδική περίπτωση της λογιστικής παλινδρόμησης (logistic function), δηλαδή η γραφική της παράσταση έχει σχήμα S. Συνήθως, συμβολίζεται με $\sigma(x)$ ή $sig(x)$. Η τιμή της συνάρτησης στο γραμμικό όριο είναι $\sigma(0) = 0.5$. Οι εισοδοί \vec{x} που βρίσκονται κοντά στο όριο πλησιάζουν την τιμή 0.5, ενώ οι εισοδοί που είναι πιο απομακρυσμένοι λαμβάνουν τιμές πολύ κοντά στο 0 ή στο 1 [9].

$$\sigma(x) = \frac{1}{1 + \exp(-ax)} \quad (3.3)$$

όπου a είναι η παράμετρος κλίσης της συνάρτησης.

- Softmax



Σχήμα 3.4: Το χαρακτηριστικό S της σιγμοειδής συνάρτησης.

Η Softmax συναντάται συνήθως στο τελευταίο επίπεδο του νευρωνικού δικτύου. Η διαφορά με τη σιγμοειδή συνάρτηση έγκειται στο γεγονός ότι η softmax εξαναγκάζει το άθροισμα των τιμών εξόδου να ισούται πάντα με τη μονάδα, έτσι ώστε κάθε τιμή εξόδου να περιέχει την πιθανότητα η είσοδος που δόθηκε στο δίκτυο να ανήκει στην αντίστοιχη κλάση. Χωρίς τη χρήση της softmax στο επίπεδο εξόδου του δικτύου, οι εξοδοί των νευρώνων είναι απλά αριθμητικές τιμές, με την υψηλότερη ένδειξη να αντιστοιχεί στην επικρατούσα κλάση [10].

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}} \quad (3.4)$$

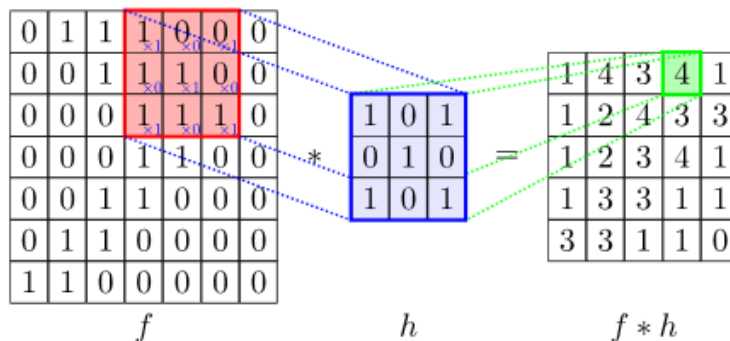
3.1.2 Convolutional Neural Networks

Τα συνελκτικά νευρωνικά δίκτυα (CNNs) χρησιμοποιούνται ευρέως στην όραση υπολογιστών για την αναγνώριση εικόνων και βίντεο. Αποτελούνται από νευρώνες με εκπαιδευσιμα βάρη και τιμή πόλωσης παρόμοια με τα κλασσικά νευρωνικά δίκτυα. Στο τελευταίο (fully-connected) επίπεδο περιέχεται η συνάρτηση σφάλματος (loss function). Τα CNNs είναι πολύ πιο αποτελεσματικά σε σύγκριση με άλλες αρχιτεκτονικές νευρωνικών δικτύων όσον αφορά προβλήματα ανάλυσης εικόνας [11]. Αυτό οφείλεται στη χωρική συσχέτιση (spatial correlation) των εικόνων. Τα CNNs συνήθως αποτελούνται από πλείστα επίπεδα το κάθε ένα με τις δικές του δυνατότητες. Χαμηλού επιπέδου χαρακτηριστικά όπως το χρώμα, οι γωνίες και σχήματα εντοπίζονται στα αρχικά επίπεδα. Όσο το μοντέλο εκπαιδευείται και το τελικό επίπεδο κάνει προβλέψεις, τόσο περισσότερα τα επίπεδα μαθαίνουν πιο πολύπλοκα χαρακτηριστικά [12].

Ένα συνελκτικό δίκτυο συνήθως αποτελείται από συνελκτικά, pooling, activation και fully connected επίπεδα. Ταυτόχρονα στο δίκτυο γίνονται και πολλές επεξεργασίες όπως upsampling, concatenation, dropout, nearest neighbour και bilinear interpolation [6].

Συνέλιξη

Η κάθε εικόνα περιγράφεται σε ένα πίνακα ανάλογου μεγέθους με της εικόνας με την μορφή αριθμών οι οποίοι είναι ανάλογοι της φωτεινότητας κάθε σημείου. Κάθε τέτοιο σημείο ονομάζεται pixel. Μία δισδιάστατη εικόνα έχει διαστάσεις $(h \times w \times d)$, όπου h είναι το ύψος, w το πλάτος και d το βάθος των χρωματικών καναλιών. Με την συνέλιξη, είναι δυνατή η εφαρμογή φίλτρων σε μία εικόνα. Αυτό πραγματοποιείται με την συνέλιξη της εικόνας f με μία μάσκα ή αλλιώς πυρήνα του φίλτρου h (kernel). Για να υπολογιστεί το αποτέλεσμα της εφαρμογής του φίλτρου σε ένα pixel, τοποθετείται το κέντρο του πυρήνα πάνω από το συγκεκριμένο σημείο, όσα pixel καλύπτονται από τον πυρήνα πολλαπλασιάζονται με τον αριθμό που τα καλύπτει και έπειτα υπολογίζεται το άθροισμα των τελικών στοιχείων. Το συνολικό άθροισμα είναι η τιμή του pixel μετά το φιλτράρισμα. Ο διασκελισμός (stride) καθορίζει το βήμα με το οποίο γίνεται η μετατόπιση της μάσκας πάνω στην εικόνα. Με βήμα ένα το φίλτρο εφαρμόζεται σε όλη την εικόνα, με βήμα δύο εφαρμόζεται στη μισή [13]. Όταν ο πυρήνας εφαρμόζεται στα pixel στην άκρη της εικόνας, μέρος του βρίσκεται έξω απ' αυτήν όπου δεν υπάρχουν τιμές pixel. Μία λύση είναι η μη εφαρμογή του φίλτρου στα συγκεκριμένα σημεία, το οποίο θα μειώσει τις διαστάσεις της τελικής εικόνας. Συνήθως για να αποφευχθεί μείωση των διαστάσεων, αυξάνονται οι διαστάσεις της εικόνας με μία τεχνική που ονομάζεται padding [14].







Σχήμα 3.5: Συνέλιξη εικόνας f με πυρήνα h .

Στην εικόνα 3.6 παρουσιάζονται διαφορετικά είδη kernel και το αποτέλεσμα τους έπειτα από την εφαρμογή τους πάνω στην εικόνα. Τα συγκεκριμένα φίλτρα χρησιμοποιούνται πολύ στην όραση υπολογιστών. Συγκεκριμένα, το Gaussian Blur κάνει πιο λεία την εικόνα, το sharpen αυξάνει το βαθμό όξυνσης και γίνεται πιο ξεκάθαρο το βάθος της εικόνας. Τέλος, με το edge detection εντοπίζουμε τις ακμές.

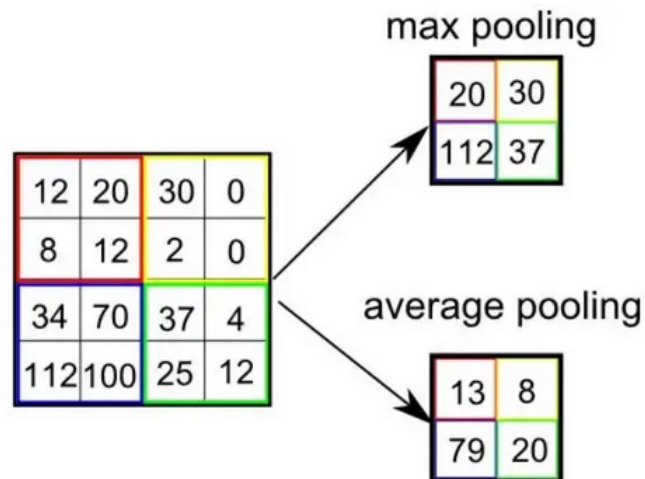
Pooling

Τα pooling layers χρησιμοποιούνται για να μειώσουν τις διαστάσεις των εικόνων, και επομένως τον αριθμό των παραμέτρων του δικτύου το οποίο με την σειρά του θα μειώσει την υπολογιστική πολυπλοκότητα. Ακόμη, χρησιμοποιώντας pooling

Original	Gaussian Blur	Sharpen	Edge Detection
$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	$\frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}$	$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$
			

Σχήμα 3.6: Διαφορετικά είδη φίλτρων, οι αντίστοιχοι πυρήνες τους και το αποτέλεσμα της εφαρμογής τους [15].

layers αποφεύγεται το overfitting στα δεδομένα εκπαίδευσης. Υπάρχουν διάφοροι τρόποι pooling, οι πιο συχνοί είναι το max και average pooling. Στην πρώτη περίπτωση στην τελική εικόνα εισάγεται η μεγαλύτερη τιμή από τους αριθμούς των pixels που βρίσκονται μέσα στο kernel και στη δεύτερη εισάγεται ο μέσος όρος των αριθμών, όπως παρουσιάζεται στο παρακάτω Σχήμα 3.7 [16].



Σχήμα 3.7: Είδη pooling [17].

Downsampling - Upsampling

Η διαδικασία του Downsampling χρησιμοποιείται για την μείωση της ανάλυσης μίας εικόνας εισόδου, ενώ ταυτόχρονα διατηρεί όση περισσότερη πληροφορία γίνεται. Upsampling είναι το αντίστροφο της πρώτης διαδικασίας, και παράγει μία εικόνα με μεγαλύτερη ανάλυση από την αρχική. Ο απώτερος στόχος και των δύο είναι η δημιουργία μίας αξιόπιστης εικόνας διατηρώντας μεγάλη λεπτομέρεια χωρίς ανεπιθύμητα artifacts. Ωστόσο, το upsampling αποτελεί μία αρκετά δύσκολη διαδικασία καθώς απαιτεί την εισαγωγή καινούριων δεδομένων στην εικόνα (Interpolation). Οι δύο πιο διάσημες μέθοδοι για upsampling είναι η Παρεμβολή Πλη-

σιέστερου Γείτονα (Nearest Neighbour Interpolation) και η Διγραμμική Παρεμβολή (Bilinear Interpolation) [18].

Παρεμβολή Πλησιέστερου Γείτονα

Σε γενικές γραμμές, το interpolation προσπαθεί να υπολογίσει την τιμή του z στο καινούριο σημείο x χρησιμοποιώντας ένα συγκεκριμένο δείγμα δεδομένων $(z_1, z_2, \dots,$

$z_n)$ στις θέσεις (x_1, x_2, \dots, x_n) . Ο αλγόριθμος του κοντινότερου γείτονα, ψάχνει την τιμή του i έτσι ώστε το $|x_i - x|$ να είναι το ελάχιστο δυνατό, τότε $z = z_i$.

Διγραμμική Παρεμβολή

Η διγραμμική παρεμβολή είναι η επέκταση της γραμμικής παρεμβολής στις 2 διαστάσεις. Χρησιμοποιεί γραμμική παρεμβολή δύο φορές, μία για κάθε διάσταση. Συγκεκριμένα, η διαδικασία υπολογίζει την τιμή στο καινούριο pixel ως γραμμικός συνδυασμός των τεσσάρων πλησιέστερων pixel στην αρχική εικόνα [19].

Έστω άγνωστη συνάρτηση f που γνωρίζουμε την τιμή της στα σημεία $Q_{11} = (x_1, y_1)$, $Q_{12} = (x_1, y_2)$, $Q_{21} = (x_2, y_1)$, $Q_{22} = (x_2, y_2)$. Εφαρμόζουμε linear interpolation για την x κατεύθυνση.

$$f(x, y_1) = \frac{x_2 - x}{x_2 - x_1} f(Q_{11}) + \frac{x - x_1}{x_2 - x_1} f(Q_{21}) \quad (3.5)$$

$$f(x, y_2) = \frac{x_2 - x}{x_2 - x_1} f(Q_{12}) + \frac{x - x_1}{x_2 - x_1} f(Q_{22}) \quad (3.6)$$

Για την y κατεύθυνση.

$$\begin{aligned} f(x, y) &= \frac{y_2 - y}{y_2 - y_1} f(x, y_1) + \frac{y - y_1}{y_2 - y_1} f(x, y_2) = \dots = \\ &= \frac{1}{(x_2 - x_1)(y_2 - y_1)} \begin{bmatrix} x_2 - x & x - x_1 \end{bmatrix} \begin{bmatrix} f(Q_{11}) & f(Q_{12}) \\ f(Q_{21}) & f(Q_{22}) \end{bmatrix} \begin{bmatrix} y_2 - y \\ y - y_1 \end{bmatrix} \end{aligned} \quad (3.7)$$

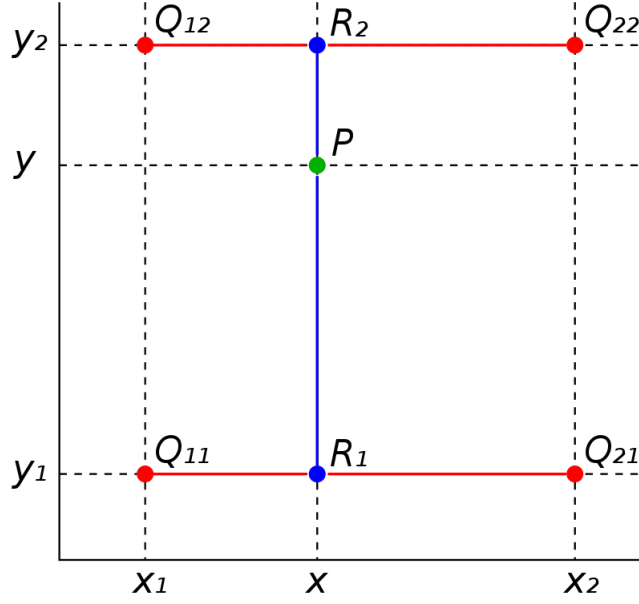
Το ίδιο αποτέλεσμα θα εκλάβουμε σε περίπτωση που ακολουθήσουμε την ίδια μεθοδολογία αλλά πρώτα για την y και έπειτα για την x .

Ακόμη, μπορούμε να υπολογίσουμε την f και ως συνδυασμό μέσων βαρών με την $f(Q)$:

$$f(x, y) \approx w_{11}f(Q_{11}) + w_{12}f(Q_{12}) + w_{21}f(Q_{21}) + w_{22}f(Q_{22}) \quad (3.8)$$

Το άθροισμα των βαρών είναι ίσο με 1, και επομένως έχουμε το γραμμικό σύστημα [20]:

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ x_1 & x_1 & x_2 & x_2 \\ y_1 & y_2 & y_1 & y_2 \\ x_1y_1 & x_1y_2 & x_2y_1 & x_2y_2 \end{bmatrix} \begin{bmatrix} w_{11} \\ w_{12} \\ w_{21} \\ w_{22} \end{bmatrix} = \begin{bmatrix} 1 \\ x \\ y \\ xy \end{bmatrix} \quad (3.9)$$



Σχήμα 3.8: P είναι το σημείο που θέλουμε να προβλέψουμε την καινούρια τιμή. Τα Q είναι τα 4 γειτονικά σημεία [20].

$$\begin{bmatrix} w_{11} \\ w_{12} \\ w_{21} \\ w_{22} \end{bmatrix} = \frac{1}{(x_2 - x_1)(y_2 - y_1)} \begin{bmatrix} x_2 y_2 & -y_2 & -x_2 & 1 \\ -x_2 y_1 & y_1 & x_2 & -1 \\ -x_1 y_2 & y_2 & x_1 & -1 \\ x_1 y_1 & -y_1 & -x_1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ x \\ y \\ xy \end{bmatrix} \quad (3.10)$$

$$\begin{cases} w_{11} = \frac{(x_2 - x)(y_2 - y)}{(x_2 - x_1)(y_2 - y_1)} \\ w_{12} = \frac{(x_2 - x)(y - y_1)}{(x_2 - x_1)(y_2 - y_1)} \\ w_{21} = \frac{(x - x_1)(y_2 - y)}{(x - x_1)(y_2 - y_1)} \\ w_{22} = \frac{(x - x_1)(y - y_1)}{(x - x_1)(y_2 - y_1)} \end{cases} \quad (3.11)$$

Batch Normalization

Το batch normalization είναι μία μέθοδος που επιταχύνει και σταθεροποιεί την εκπαίδευση των νευρωνικών δικτύων [21]. Περιέχει διανύσματα κανονικοποίησης στα κρυφά επίπεδα, χρησιμοποιώντας τον μέσο όρο και την διακύμανση του κάθε batch. Σε κάθε κρυφό επίπεδο μετατρέπει το σήμα με μη γραμμική συνάρτηση πριν (ή μετά) το βήμα κανονικοποίησης [21].

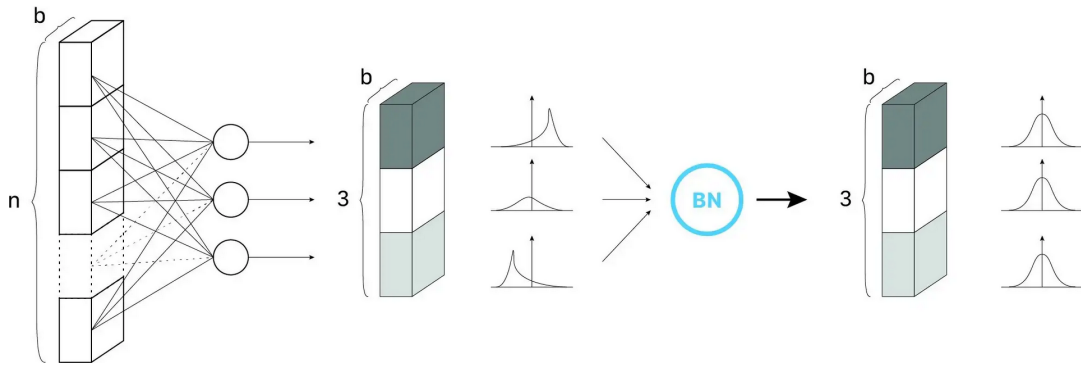
$$\mu = \frac{1}{n} \sum_i Z^{(i)} \quad (3.12)$$

$$\sigma^2 = \frac{1}{n} \sum_i (Z^{(i)} - \mu)^2 \quad (3.13)$$

$$Z_{norm}^{(i)} = \frac{Z^i - \mu}{\sqrt{\sigma^2 - \epsilon}} \quad (3.14)$$

$$\hat{Z} = \gamma * Z_{norm}^{(i)} + \beta \quad (3.15)$$

Με τις εξισώσεις 3.12 και 3.13 υπολογίζονται ο μέσος όρος και η συνδιακύμανση του κάθε batch. Το διάνυσμα ενεργοποίησης κανονικοποιείται από την 3.14. Με αυτό τον τρόπο οι έξοδοι κάθε νευρώνα ακολουθούν καθορισμένη κανονική κατανομή. Η έξοδος του κάθε επιπέδου \hat{Z} υπολογίζεται γραμμικά με δύο εκπαιδευσιμες παραμέτρους β και γ . Αλλάζοντας την τιμή αυτών των δύο, μέσω του αλγορίθμου της απότομης καθόδου, το μοντέλο βρίσκει την καλύτερη κατανομή για κάθε επίπεδο [22].



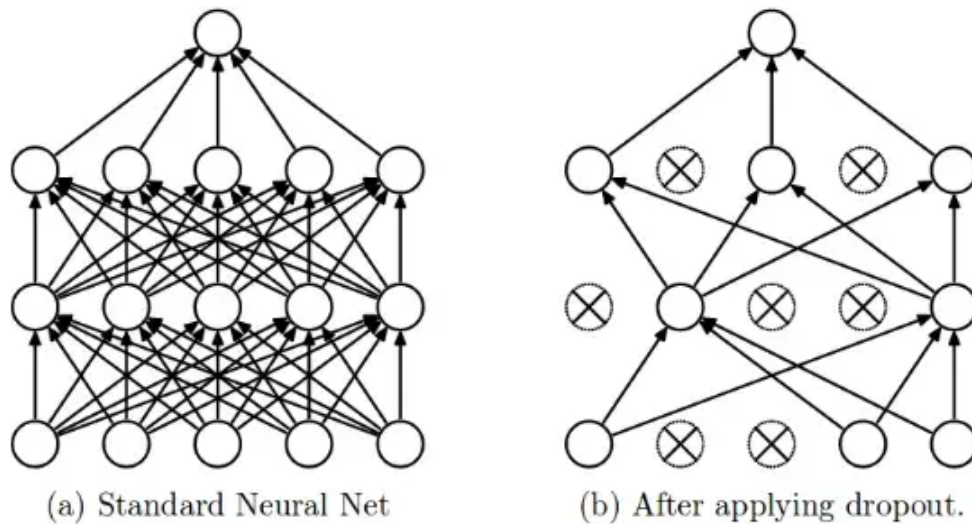
Σχήμα 3.9: Batch Normalization [22].

Dropout

Το dropout, είναι η πιο συχνή μέθοδος κανονικοποίησης στα νευρωνικά δίκτυα. Ακόμα, και στα πιο εξελιγμένα μοντέλα, προσθέτοντας dropout η ακρίβεια μπορεί να αυξηθεί σημαντικά. Το κύριο χαρακτηριστικό του είναι ότι αποτρέπει το overfitting στα δεδομένα. Εφαρμόζεται μόνο στα κρυφά επίπεδα και στα επίπεδα εισόδου, κατά την διάρκεια της εκπαίδευσης. Ο ρόλος του είναι να απενεργοποιεί (*drop*) τυχαία ένα κομμάτι νευρώνων με μία πιθανότητα p σε κάθε επανάληψη. Με αυτό τον τρόπο οι έξοδοι και οι εισοδοι των νευρώνων θα είναι ομοίως απενεργοποιημένοι. Στην επόμενη επανάληψη θα είναι με την ίδια πιθανότητα ενεργοί ή όχι. Συνήθως η πιθανότητα είναι 50% ($p = 0.5$) [23].

3.2 Εκπαίδευση Νευρωνικών Δικτύων

Εκτός από την αρχιτεκτονική των νευρωνικών δικτύων πολύ σημαντική είναι και η ανάλυση του τρόπου με τον οποίο ορίζονται τα βάρη των συνάψεων και η πόλωση σε κάθε νευρώνα ώστε η έξοδος του μοντέλου να είναι η καλύτερη επιθυμητή.



Σχήμα 3.10: Dropout [23].

Συγκεκριμένα, η τιμή των βαρών μεταβάλλεται εφαρμόζοντας κάποιον κανόνα εκμάθησης, οι οποίοι κανόνες χωρίζονται σε 3 κατηγορίες [6]:

- Επιβλεπόμενη μάθηση (supervised learning)
- Μη-επιβλεπόμενη μάθηση (unsupervised learning)
- Ενισχυτική μάθηση (reinforcement learning)

Επιβλεπόμενη μάθηση

Στις περισσότερες περιπτώσεις μηχανικής μάθησης χρησιμοποιείται επιβλεπόμενη μάθηση, στις οποίες παρέχονται δεδομένα input με τα αντίστοιχα labels που κατηγοριοποιούν τις εισόδους. Για παράδειγμα η είσοδος είναι 2D X-ray του σώματος ενός ασθενή και το label η δυαδική κατηγορία άρρωστος-υγιής. Κάθε ζευγάρι είσοδος-label εισάγεται στο δίκτυο, το οποίο με υπολογισμούς στο κάθε επίπεδο παράγει μία έξοδο. Αυτή η έξοδος συγκρίνεται μετά με το label, και εφαρμόζεται ο κανόνας εκμάθησης για να μεταβάλλει την έξοδο πιο κοντά στην επιθυμητή.

Μη-Επιβλεπόμενη μάθηση

Στη μη-επιβλεπόμενη μάθηση, τα βάρη και οι πολώσεις μεταβάλλονται αποκλειστικά βάσει των εισόδων στο δίκτυο. Συγκεκριμένα, χρησιμοποιούνται αλγόριθμοι που επιτελούν ομαδοποίηση των δεδομένων (clustering). Μαθαίνουν, δηλαδή, να κατηγοριοποιούν τις εισόδους σε ένα πεπερασμένο αριθμό από ομάδες, ή αλλιώς, κλάσεις.

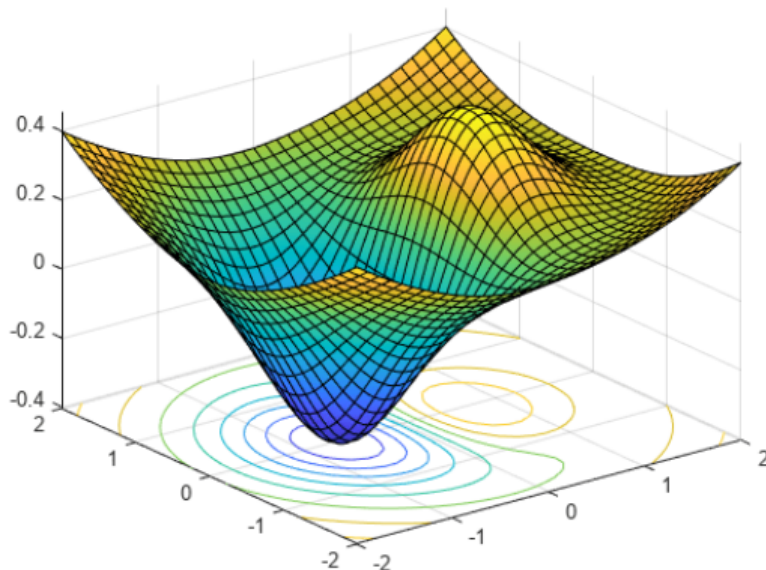
Ενισχυτική μάθηση

Η ενισχυτική μάθηση μοιάζει αρκετά με την επιβλεπόμενη μάθηση, όμως αντί να δίνει κάθε φορά την σωστή έξοδο για την αντίστοιχη είσοδο, δίνει ένα βαθμό. Ο βαθμός (grade ή score) είναι ένα μέτρο της απόδοσης του δικτύου για μία αλληλουχία εισόδων. Πρόκειται για έναν τύπο εκμάθησης ο οποίος χρησιμοποιείται πολύ πιο σπάνια από την επιβλεπόμενη μάθηση, και βρίσκει εφαρμογή κυρίως σε προβλήματα συστημάτων ελέγχου.

3.2.1 Αλγόριθμοι Βελτιστοποίησης

Αλγόριθμος απότομης καθόδου (Gradient Descent)

Κατά την δημιουργία ενός νευρωνικού δικτύου τα βάρη και οι πολώσεις επιλέγονται τυχαία και ως αποτέλεσμα η έξοδος του μοντέλου είναι λανθασμένη. Έστω συνάρτηση κόστους που ορίζει πόσο απέχει η έξοδος από την πραγματική λύση (ground truth), τα είδη αυτών των συναρτήσεων θα αναλυθούν στη συνέχεια. Το πρόβλημα στην εκπαίδευση του νευρωνικού δικτύου είναι η εύρεση βαρών και πολώσεων τα οποία ελαχιστοποιούν τη συνάρτηση κόστους. Πρόκειται, δηλαδή, για ένα πρόβλημα εύρεσης ελαχίστου. Ιδανικά θα θέλαμε να βρούμε το ολικό ελάχιστο, κάτι όμως το οποίο αποδεικνύεται εξαιρετικά δύσκολο. Συνήθως αρκεί να βρούμε ένα τοπικό ελάχιστο της συνάρτησης. Ξεκινώντας από ένα τυχαίο σημείο της συνάρτησης (∇C) υπολογίζεται η κλίση (gradient) η οποία δίνει την κατεύθυνση που πρέπει να κινηθεί ανάλογα το πρόσημο. Θετική τιμή σημαίνει ότι είναι αύξουσα και αρνητική ότι είναι φθίνουσα. Καθώς σκοπός είναι η εύρεση τοπικού ελαχίστου σε σημείο $\nabla C(x, y)$, μετακινούμαστε πάντα προς την κατεύθυνση του σημείου μικρότερης τιμής, κατά ένα μικρό βήμα. Έπειτα από μετακίνηση γίνεται εκ νέου υπολογισμός του gradient. Εάν η κλίση έχει διαφορετικό πρόσημο από αυτό που είχε το προηγούμενο σημείο τότε το βήμα είναι μεγάλο και προσπεράσαμε το επιθυμητό σημείο και το βήμα πρέπει να προσαρμοστεί [24].



Σχήμα 3.11: Γραφική παράσταση τρισδιάστατης συνάρτησης [25].

Αν φανταστούμε πως κάθε βάρος του δικτύου αντιστοιχεί και σε μία τιμή, ή συνιστώσα, του διανύσματος του gradient για τη συνάρτηση κόστους, το πρόσημο της τιμής αυτής μας δείχνει σε ποιά κατεύθυνση πρέπει να αλλαχθεί ο αριθμός ώστε να βρεθεί ένα τοπικό ελάχιστο, ενώ το μέτρο της τιμής μας δείχνει πόσο θα επηρεαστεί το τελικό αποτέλεσμα από την μεταβολή αυτού του βάρους ή πόλωσης. Ο αλγόριθμος με τον οποίο υπολογίζεται αποδοτικά το gradient ονομάζεται back-propagation (οπισθοδιάδοση).

Αλγόριθμος Στοχαστικής Σταδιακής Καθόδου (Stochastic Gradient Descent (SGD))

Η στοχαστική σταδιακή κάθοδος (SGD) είναι ένας δημοφιλής αλγόριθμος βελτιστοποίησης που χρησιμοποιείται για την ελαχιστοποίηση της συνάρτησης κόστους ή απώλειας ενός μοντέλου μηχανικής μάθησης. Πρόκειται για έναν τύπο αλγορίθμου απότομης καθόδου όπου αντί να ενημερώνονται οι παράμετροι του μοντέλου με βάση τον μέσο όρο των κλίσεων σε ολόκληρο ή τμήμα του συνόλου δεδομένων εκπαίδευσης, οι παράμετροι ενημερώνονται με βάση την κλίση ενός μόνο τυχαία επιλεγμένου δείγματος από το σύνολο δεδομένων εκπαίδευσης.

Η βασική ιδέα πίσω από τον SGD είναι η εκτίμηση της κλίσης της συνάρτησης κόστους για ένα τυχαίο παράδειγμα και η προσαρμογή των παραμέτρων του μοντέλου προς την κατεύθυνση της αρνητικής κλίσης. Αυτή η διαδικασία επαναλαμβάνεται για πολλές επαναλήψεις μέχρι η συνάρτηση κόστους να φτάσει σε μια ελάχιστη τιμή.

Ο SGD έχει πολλά πλεονεκτήματα σε σύγκριση με τον αλγόριθμο απότομης καθόδου, όπως είναι η ταχύτερη σύγκλιση, η δυνατότητα διαφυγής από τοπικά ελάχιστα και η δυνατότητα αποτελεσματικής επεξεργασίας μεγάλων συνόλων δεδομένων,

καθώς οι παράμετροι μπορούν να ενημερώνονται μετά την επεξεργασία κάθε παραδείγματος. Ωστόσο, η στοχαστική φύση του SGD μπορεί επίσης να οδηγήσει σε ταλαντώσεις και αργή σύγκλιση, οπότε είναι σημαντικό να χρησιμοποιείται κατάλληλη ρύθμιση των υπερπαραμέτρων, όπως ο καθορισμός ενός αλγορίθμου για την αλλαγή του ρυθμού μάθησης κατά την διάρκεια των εποχών, για να έχουμε καλή απόδοση [26].

Ο SGD χρησιμοποιείται ευρέως για την εκπαίδευση μοντέλων μηχανικής μάθησης, όπως η γραμμική παλινδρόμηση, η λογιστική παλινδρόμηση και τα νευρωνικά δίκτυα.

3.2.2 Backpropagation

Ο αλγόριθμος backpropagation ή αλλιώς μέθοδος οπισθοδρόμησης είναι από τους πιο σημαντικούς αλγορίθμους για την εκπαίδευση των νευρωνικών δικτύων. Έπειτα από κάθε προς τα εμπρός διάδοση (δηλαδή αφού το μοντέλο βγάλει μία έξοδο) η μέθοδος οπισθοδρόμησης πραγματοποιεί μία προς τα πίσω διάδοση αλλάζοντας ταυτόχρονα τα βάρη και τις πολώσεις του κάθε νευρώνα. Σκοπός του αλγορίθμου είναι αλλάζοντας τις συγκεκριμένες παραμέτρους να ελαχιστοποιήσει την συνάρτηση κόστους. Οι αλλαγές στις παραμέτρους καθορίζονται από το gradient της συνάρτησης κόστους συναρτήσει των βαρών [27].

Έστω η συνάρτηση κόστους C , και ένα βάρος w_{jk}^l , τότε σύμφωνα με τον κανόνα αλυσίδας (chain rule) ισχύει:

$$\frac{\partial C}{\partial w_{jk}^l} = \frac{\partial C}{\partial z_j^l} \frac{\partial z_j^l}{\partial w_{jk}^l} \quad (3.16)$$

Για κάθε έξοδο νευρώνα ισχύει:

$$z_j^l = \sum_{k=1}^m w_{jk}^l + b_j^l \quad (3.17)$$

όπου m ο αριθμός νευρώνων στο επίπεδο $l - 1$.

Παραγωγίζοντας ως προς το βάρος w_{jk}^l έχουμε:

$$\frac{\partial z_j^l}{\partial w_{jk}^l} = a_k^{l-1} \quad (3.18)$$

Αντικαθιστώντας στην 3.16:

$$\frac{\partial C}{\partial w_{jk}^l} = \frac{\partial C}{\partial z_j^l} a_k^{l-1} \quad (3.19)$$

Με όμοιο τρόπο υπολογίζεται και η μερική παράγωγος της C ως προς κάποιο *bias* b_j^l :

$$\frac{\partial C}{\partial b_j^l} = \frac{\partial C}{\partial z_j^l} \frac{\partial z_j^l}{\partial b_j^l} \quad (3.20)$$

$$\frac{\partial z_j^l}{\partial b_j^l} = 1 \quad (3.21)$$

$$\frac{\partial C}{\partial b_j^l} = \frac{\partial C}{\partial z_j^l} \quad (3.22)$$

Ο κοινός όρος των 3.19 και 3.22 ονομάζεται σφάλμα ή αλλιώς local gradient και συμβολίζεται με :

$$\delta_j^l = \frac{\partial C}{\partial z_j^l} \quad (3.23)$$

$$\delta_j^l = \sum_{i=1}^{r_{l+1}} \frac{\partial C}{\partial z_i^{l+1}} \frac{\partial z_i^{l+1}}{\partial z_j^l} = \sum_{i=1}^{r_{l+1}} \delta_i^{l+1} \frac{\partial z_i^{l+1}}{\partial z_j^l} \quad (3.24)$$

Συνεπώς, ο αλγόριθμος backpropagation ξεκινώντας από το τελευταίο επίπεδο υπολογίζει κάθε φορά τις μερικές παραγώγους χρησιμοποιώντας τον κανόνα αλυσίδας. Όπως φαίνεται από την 3.24 το local gradient ενός επιπέδου k εξαρτάται από το local gradient του επόμενου επιπέδου $k + 1$. Επομένως, τα σφάλματα ρέουν προς τα πίσω, από το τελευταίο επίπεδο στο πρώτο. Αυτός ακριβώς είναι και ο λόγος που ο αλγόριθμος ξεκινά από το τελευταίο επίπεδο και εκεί οφείλεται και το όνομα της μεθόδου [28].

3.2.3 Συνάρτηση Κόστους (Loss Function)

Η συνάρτηση κόστους υπολογίζει την διαφορά μεταξύ της εξόδου του νευρωνικού δικτύου και του αναμενόμενου αποτελέσματος (label). Χωρίζεται σε δύο κατηγορίες. Η πρώτη είναι κατηγορία κατηγοριοποίησης (classification) που έχει ως αποτέλεσμα διακριτές τιμές (0,1,2,...) και η δεύτερη κατηγορία λογιστικής παλινδρόμησης (regression) με συνεχείς τιμές.

Mean Square Error (MSE)

Το μέσο τετραγωνικό σφάλμα είναι ίσως η πιο απλή μορφή συνάρτησης κόστους. Είναι η μέση τετραγωνική διαφορά μεταξύ της εξόδου του μοντέλου και της πραγματικής λύσης (ground truth) σε όλο το dataset.

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (3.25)$$

Mean Absolute Error (MAE)

Το μέσο απόλυτο σφάλμα διαφέρει με το MSE πολύ λίγο όπως φαίνεται και από το όνομα. Είναι η μέση απόλυτη διαφορά μεταξύ της εξόδου του μοντέλου και της πραγματικής λύσης (ground truth) σε όλο το dataset.

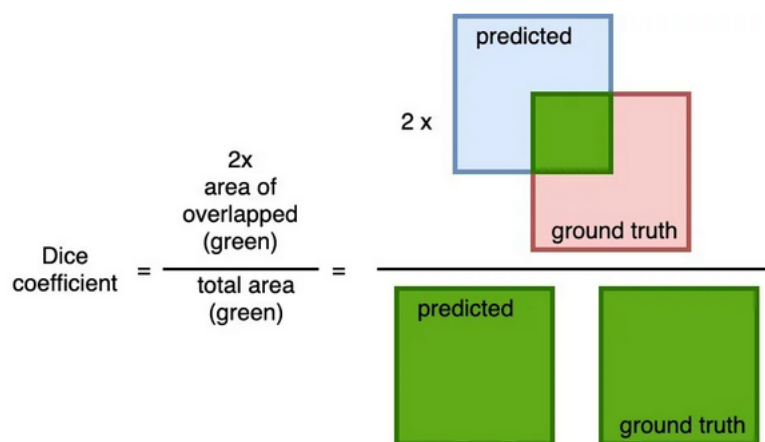
$$MSE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (3.26)$$

Dice loss

Η dice loss συνάρτηση κόστους προέρχεται από το 1940 για την εύρεση της διαφοράς μεταξύ δύο δειγμάτων [29]. Ξεκίνησε να χρησιμοποιείται τα τελευταία χρόνια σε προβλήματα νευρωνικών δικτύων για την διαφορά μεταξύ εικόνων, συγκεκριμένα για κατηγοριοποίηση ιατρικών εικόνων.

$$D = \frac{2 \sum_i^N p_i g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2} \quad (3.27)$$

Όπου p_i η έξοδος του μοντέλου και g_i το ground truth. Σημαντική παρατήρηση είναι πως η συνάρτηση δέχεται μόνο δυαδικές τιμές 0 και 1.



Σχήμα 3.12: Dice Loss [30].

Intersection over Union (IoU)

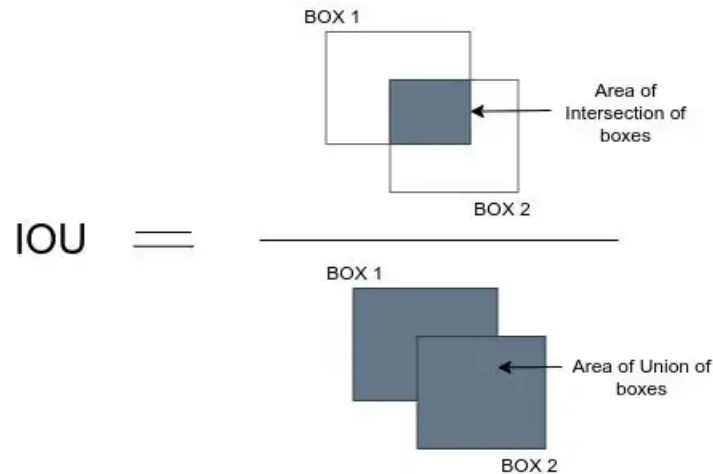
Η συνάρτηση κόστους Intersection over Union (IoU) είναι μια κοινή συνάρτηση που χρησιμοποιείται στην εκπαίδευση μοντέλων όρασης υπολογιστών, ειδικά για προβλήματα όπως η σημασιολογική κατάτμηση και η ανίχνευση αντικειμένων. Η μετρική IoU μετράει την επικάλυψη μεταξύ των επιθυμητών μασκών και της εξόδου του δικτύου και ορίζεται ως ο λόγος της τομής τους προς την ένωση τους.

$$IoU = \frac{A \cap B}{A \cup B} \quad (3.28)$$

Η συνάρτηση κόστους IoU είναι ιδιαίτερα χρήσιμη για προβλήματα όρασης υπολογιστών, καθώς παρέχει μια πιο αναλυτική αξιολόγηση των προβλέψεων του μοντέλου σε σύγκριση με άλλες συναρτήσεις κόστους, όπως το μέσο τετραγωνικό σφάλμα ή η δυαδική διασταυρούμενη εντροπία (Binary Cross-Entropy). Για παράδειγμα, στην ανίχνευση αντικειμένων, η IoU τιμωρεί με διαφορετικό τρόπο τα ψευδώς αρνητικά και τα ψευδώς θετικά αποτελέσματα, βάσει του βαθμού επικάλυψης μεταξύ των επιθυμητών μασκών και της εξόδου του δικτύου. Αυτό βοηθά το

μοντέλο να μάθει να προβλέπει τα πλαίσια οριοθέτησης που έχουν υψηλή επικάλυψη με τα πλαίσια της βασικής αλήθειας (ground truth).

Στην πράξη, η συνάρτηση κόστους IoU χρησιμοποιείται συχνά σε συνδυασμό με άλλες συναρτήσεις κόστους για να παρέχει μια πιο ολοκληρωμένη αξιολόγηση της απόδοσης του μοντέλου. Για παράδειγμα, η συνολική συνάρτηση κόστους μπορεί να είναι ένα βεβαρημένο άθροισμα της απώλειας IoU, της δυαδικής διασταυρούμενης εντροπίας και της απώλειας μέσου τετραγωνικού σφάλματος, ανάλογα με τις ειδικές απαιτήσεις του προβλήματος.



Σχήμα 3.13: Intersection over Union [31].

Cross-entropy Loss

Η συνάρτηση Cross-entropy υπολογίζει την διαφορά μεταξύ δύο συναρτήσεων κατανομής πιθανότητας. Η cross-entropy της κατανομής q με την κατανομή p ορίζεται ως η εξής:

$$H(p, q) = -E_p[\log q] \quad (3.29)$$

Για διακριτές κατανομές πιθανότητες p και q στο ίδιο πεδίο ορισμού X

$$H(p, q) = - \sum_{x \in X} p(x) \log q(x) \quad (3.30)$$

Υπάρχουν πολλά διαφορετικά είδη της συνάρτηση κόστους Cross-entropy που το κάθε ένα λύνει και διαφορετικά προβλήματα.

- Binary Cross-Entropy:

$$L_{BCE}(y, \hat{y}) = -(y \log \hat{y} + (1 - y) \log(1 - \hat{y})) \quad (3.31)$$

όπου το y είναι η πραγματική τιμή και \hat{y} είναι η έξοδος του μοντέλου.

- Weighted Binary Cross-Entropy:

$$L_{w-BCE}(y, \hat{y}) = -(\beta * y \log \hat{y} + (1 - y) \log(1 - \hat{y})) \quad (3.32)$$

Όπου β μεταβλητή που χρησιμοποιείται για την βελτίωση του μοντέλου μειώνοντας τα λάθη καθώς και τα false positives. Η συγκεκριμένη περίπτωση χρησιμοποιείται πολύ σε μη συμμετρικά δεδομένα.

- Balanced Cross-Entropy

$$L_{BCE}(y, \hat{y}) = -(\beta * y \log \hat{y} + (1 - \beta)(1 - y) \log(1 - \hat{y})) \quad (3.33)$$

Όπου β στην ανάλυση εικόνας ορίζεται ως $1 - \frac{y}{H * W}$

Exponential Loss

Η εκθετική συνάρτηση κόστους είναι κυρτή και αυξάνεται εκθετικά για αρνητικές τιμές. Επομένως, είναι ευαίσθητη στις αρκαίες τιμές. Αρχικά σχεδιάστηκε για τον αλγόριθμο Adaboost.

$$Exp_loss = \frac{\sum e^{-y*\hat{y}}}{m} \quad (3.34)$$

Hinge Loss

Το Hinge Loss function δημιουργήθηκε για να διορθώσει το υπερεπίπεδο του αλγορίθμου SVM σε προβλήματα κατηγοριοποίησης. Είναι ισοδύναμο με την ελαχιστοποίηση του loss function $L(y, f) = [1 - yf]_+$

$$Hinge = \max(0, 1 - y * f(x)) \quad (3.35)$$

Kullback Leibler Divergence Loss

Η συνάρτηση κόστους KL Divergence είναι το αποτέλεσμα δύο διαφορετικών συναρτήσεων κατανομής πιθανότητας. Η KL Divergence δύο PDF $q(x)$ και $p(x)$ συμβολίζεται με $D_{KL}(q||p)$ όπου $||$ σημαίνει divergence.

$$D_{KL}(q||p) = - \sum q(x) \log \frac{p(x)}{q(x)} \quad (3.36)$$

3.3 Σημασιολογική Κατάτμηση (Semantic Segmentation)

Ιδιαίτερο ενδιαφέρον έχουν τα προβλήματα στα οποία εκτός από το γενικό συμπέρασμα μιας εικόνας μας ενδιαφέρει η επεξεργασία και η εξαγωγή χαρακτηριστικών από πολλά σημεία στην εικόνα. Υπάρχουν τρεις κατηγορίες μεθόδων για την ανάλυση μιας εικόνας:

- Image level classification: Συνήθως η εικόνα περιέχει μόνο ένα αντικείμενο που προσπαθούμε να κατηγοριοποιήσουμε.

- Αναγνώριση αντικειμένου εντός της εικόνας (Object Detection): Η εικόνα περιέχει διάφορα αντικείμενα και προϋποθέτει τον εντοπισμό και κατηγοριοποίηση των σημαντικών αντικειμένων συνήθως με κάποιο τετράγωνο (bounding box).
- Σημασιολογική Κατάτμηση (Semantic Segmentation): Όλα τα pixel στην εικόνα είναι ταξινομημένα ορίζοντας το σχήμα όλων των κατηγοριών αντικειμένων. Η τμηματοποίηση μπορεί επίσης να περιπλέκεται αφού χρειάζεται η κατηγοριοποίηση και του background. Το πλεονέκτημα αυτής της μεθόδου σε σύγκριση με το object detection είναι πως οι μάσκες δεν επικαλύπτονται ενώ τα bounding boxes ειδικά σε περιπτώσεις κατηγοριοποίησης πολλών ίδιων αντικειμένων που είτε έχουν περίεργο σχήμα είτε βρίσκονται πολλά μαζί στην ίδια περιοχή.

Η κατηγοριοποίηση χρησιμοποιείται σε πρώτο στάδιο από έξυπνα συστήματα τα οποία χρησιμοποιούν όραση για να κάνουν σημαντικές αποφάσεις. Για παράδειγμα στην αυτόνομη οδήγηση το σύστημα αρχικά συλλέγει φωτογραφίες από το περιβάλλον γύρω του και είναι ικανό να κατηγοροποιήσει τα αντικείμενα που υπάρχουν όπως ένα πεζός άνθρωπος ή άλλο αυτοκίνητο για να πάρει γρήγορα την επόμενη απόφαση του. Άλλες εφαρμογές της κατηγοριοποίησης στην όραση υπολογιστών είναι η ανάλυση ιατρικών εικόνων ή συστήματα ασφάλειας.

3.3.1 Πριν την Βαθιά Μάθηση

Η σημασιολογική κατάτμηση είναι η τεχνική ανάλυσης της εικόνας έως το επίπεδο των pixel, δηλαδή κατηγοριοποιούμε το κάθε pixel σε πολλές διαφορετικές κλάσεις. Πριν την βαθιά μάθηση χρησιμοποιούνταν διάφορες τεχνικές όπως το Image Thresholding, Conditional Random Fields (CRF) και K-means clustering.

Image Thresholding

Το image thresholding είναι ίσως η πιο παλιά και βασική μέθοδος για σημασιολογική κατάτμηση. Αποτελεί μία τεχνική η οποία χωρίζει την εικόνα σε δύο ή περισσότερα μέρη, συνήθως σε foreground και background. Αρχικά η εικόνα μετατρέπεται σε γκρι κλίμακα (grayscale) και έπειτα εφαρμόζεται η τεχνική του threshold, η οποία μετατρέπει την εικόνα σε δυαδική μορφή. Συγκεκριμένα, όταν η τιμή του pixel ξεπερνάει την τιμή του threshold τότε θεωρείται background και η τιμή τίθεται σε 1 (άσπρο). Στην αντίθετη περίπτωση που η τιμή είναι μικρότερη από την προκαθορισμένη τότε θεωρείται ως αντικείμενο και θέτεται σε 0 (μαύρο). Η τιμή του threshold επιλέγεται είτε από τον μηχανικό που σχεδιάζει το μοντέλο, είτε αυτόματα.

Από το 1979, ο Nobuyuki Otsu εφηύρε τον αλγόριθμο γνωστό ως Otsus's technique ο οποίος χρησιμοποιείται από τότε για να καθορίζεται αυτόματα η τιμή του

threshold. Η τεχνική ψάχνει όλες τις τιμές του threshold μέχρι να εντοπίσει την τιμή που ελαχιστοποιεί την intra-class διακύμανση (variance) [32].

Έστω ότι διχοτομούμε την εικόμα σε δύο κλάσεις C_0 και C_1 :

$$\sigma_{\omega}^2(t) = \omega_0(t)\sigma_0^2(t) + \omega_1(t)\sigma_1^2(t) \quad (3.37)$$

$$T = \min_t(\sigma_{\omega}^2(t)) \quad (3.38)$$

Όπου $\sigma_{\omega}^2(t)$ η intra-class variance και ω_0 , $\sigma_0^2(t)$ και ω_1 , $\sigma_1^2(t)$ είναι η πιθανότητα εμφάνισης της κάθε κλάσης και οι διακυμάνσεις των κλάσεων αντίστοιχα.

$$\omega_0 = Pr(C_0) = \sum_{i=1}^k p_i = \omega(k) \quad (3.39)$$

$$\omega_1 = Pr(C_1) = \sum_{i=k+1}^L p_i = 1 - \omega(k) \quad (3.40)$$



Σχήμα 3.14: Παράδειγμα thresholding χρησιμοποιώντας τον αλγόριθμο του Otsu [33].

3.3.2 Χρησιμοποιώντας την Βαθιά Μάθηση

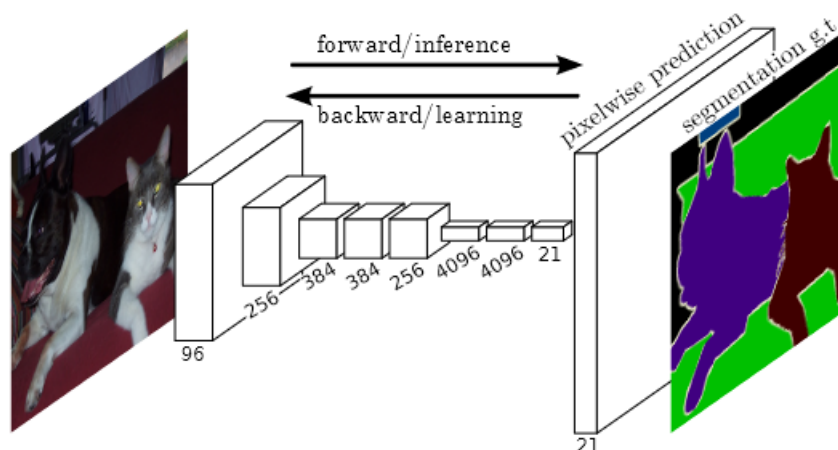
Η βαθιά μάθηση έχει κάνει σημαντική πρόοδο στην όραση υπολογιστών. Τα νευρωνικά δίκτυα έχουν καταφέρει state of the art αποδόσεις σε όλα τα προβλήματα ανάλυσης εικόνας, όπως η σημασιολογική κατάτμηση. Μία από τις αρχικές συνελκτικές μεθόδους για semantic segmentation είναι η FCN. Έπειτα, προτάθηκαν αρκετές μέθοδοι βασισμένες στη πρώτη οι οποίες την βελτίωναν, όπως είναι οι U-Net, DeepLab, Segnet και Deconvnet. Οι πιο αποδοτικές από αυτές αποδείχτηκε να είναι εκείνες με αρχιτεκτονική encoder-decoder. Ο encoder απαρτίζεται από διαδοχικά convolutional και pooling layers τα οποία ουσιαστικά μαθαίνουν να εξάγουν τα πιο σημαντικά χαρακτηριστικά της εικόνας και καταλήγουν στο bottleneck layer το οποίο δίνεται στον decoder. Ο decoder αυξάνει την διάσταση της εικόνας και την φέρνει στην επιθυμητή μορφή της εξόδου [11]. Για το μέρος του encoder πολλές φορές χρησιμοποιούνται pre-trained μοντέλα όπως είναι το VGG και το ResNet.

3.3.3 Διάσημα Μοντέλα

Fully Convolutional Network (FCN)

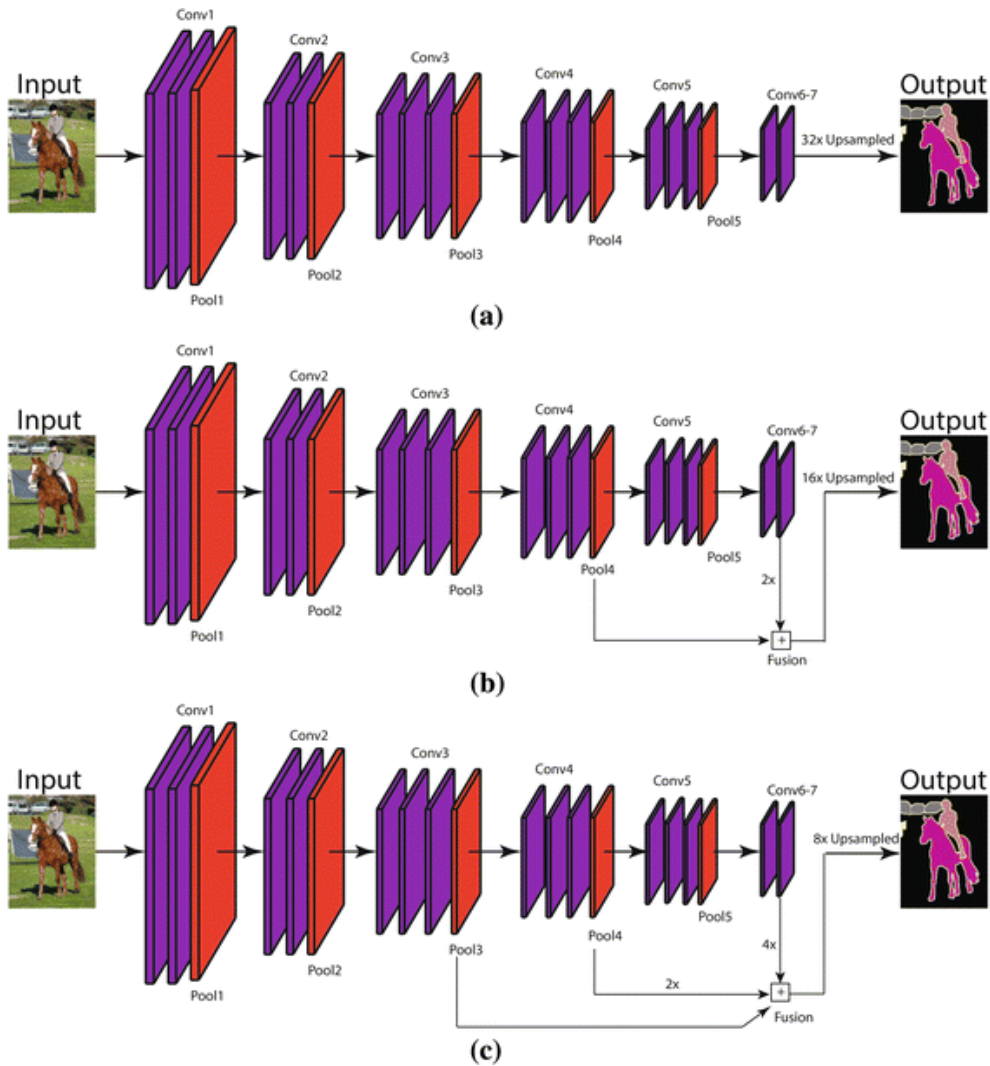
Το Fully Convolutional Network για semantic segmentation είναι το πρώτο μοντέλο που αλλάζει την αρχιτεκτονική και πραγματοποιεί fine tuning σε μοντέλα κατηγοριοποίησης για να τα προσανατολίσει προς την σημασιολογική κατάτμηση. Τα βασικά δίκτυα που χρησιμοποιήθηκαν είναι τα AlexNet, VGGNet και GoogleNet προ-εκπαιδευμένα στα ILSVRC δεδομένα. Το κάθε δίκτυο μετατράπηκε σε FCN αγνοώντας το τελευταίο classifier layer, μετατρέποντας όλα τα fully connected layers σε συνελκτικά και προσθέτοντας μία 1x1 συνέλιξη με 21 κανάλια ένα για κάθε κλάση του PASCAL dataset.

Κάθε ένα μοντέλο FCN-VGG16, FCN-AlexNet και FCN-GoogleNet, έδινε ικανοποιητικές αποδόσεις, με καλύτερο να αποδεικνύεται το FCN-VGG16 με state-of-the-art, φτάνοντας απόδοση το 52.6 mean IoU στο PASCAL 2011 test set. Η εκπαίδευση του σε περισσότερα δεδομένα, αύξησε την απόδοση σε 59.4 mean IoU.



Σχήμα 3.15: FNC πραγματοποιεί σημασιολογική κατάτμηση.

Ωστόσο, αν και οι αποδόσεις στις μετρικές είναι καλές, οι έξοδοι δεν είναι αρκετά λεπτομερείς. Στο τελευταίο prediction layer χρησιμοποιείται 32 pixel stride το οποίο μειώνει την λεπτομέρεια στην έξοδο. Για να λυθεί αυτό το πρόβλημα, χρησιμοποιήσαν skip connection για να συνδυάσουν το τελευταίο prediction layer με χαμηλότερα επίπεδα με "καλύτερα" strides και εφάρμοσαν bilinear interpolation για να κάνουν upsample την έξοδο και να είναι πιο πυκνή σε pixel. Χρησιμοποιήθηκαν αρκετοί συνδυασμοί, κάποιιοι από τους οποίους είναι οι FCN-8s, FCN-16s και FCN32s όπως φαίνονται στην 3.16. Ο FCN-8s βελτιώνει την απόδοση του δικτύου φτάνοντας το 62.7 mean IoU στο VOC2011 test set και 62.2 στο VOC2012 test set [34].



Σχήμα 3.16: Απεικόνιση των διαφορετικών FCN αρχιτεκτονικών. (a) FCN-32 αρχιτεκτονική, (b) FCN-16 αρχιτεκτονική, (c) FCN-8 αρχιτεκτονική [34]

DeepLab

Τα βαθιά συνελκτικά νευρωνικά δίκτυα (DCNNs) έχουν αποδειχθεί να είναι πολύ χρήσιμα στην σημασιολογική κατάτμηση. Ωστόσο, η χρήση τους εμφανίζει τρία προβλήματα, τα οποία προσπαθεί να λύσει το Deeplab:

- Η μειωμένη η ανάλυση των χαρακτηριστικών, λόγω της επανειλημμένης χρήσης maxpooling και striding σε πολλά συνεχόμενα επίπεδα.
- Η εμφάνιση των αντικειμένων σε πολλές διαστάσεις
- Η μειωμένη ακρίβεια εντοπισμού εξαιτίας των αντικειμένων που είναι αμετάβλητα σε spatial transformations.

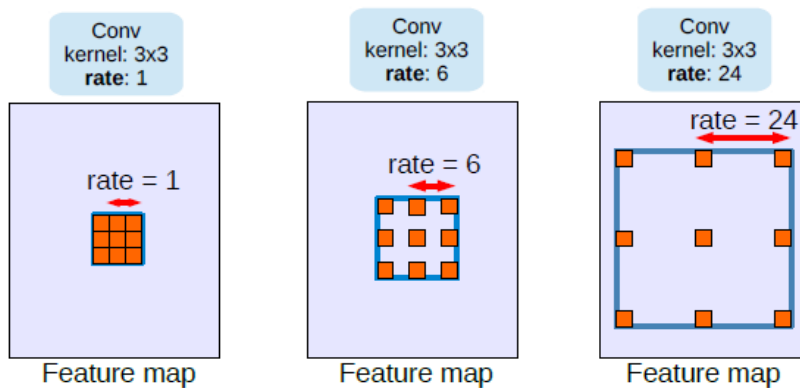
Για το πρώτο πρόβλημα, χρησιμοποίησαν τον αλγόριθμο atrous [35]. Η συνέλιξη atrous είναι πολύ ικανή στον έλεγχο των χαρακτηριστικών που παράγουν τα DCNNs. Όταν το atrous rate είναι ίσο με ένα τότε η atrous συνέλιξη ταυτίζεται με την

κανονική. Αυξάνοντας το rate αυξάνεται και το οπτικό πεδίο (field-of-view) του μοντέλου. Έστω, σήματα μίας διάστασης, η συνέλιξη atrous $y[i]$ ενός σήματος $x[i]$ με φίλτρο $w[k]$ μήκους K και $rate = r$ ορίζεται ως:

$$y[i] = \sum_{k=1}^K x[i + rk]w[k] \quad (3.41)$$

Για το δεύτερο πρόβλημα, δημιούργησαν τμήματα που χρησιμοποιούν την συνέλιξη atrous είτε σε σειρά είτε παράλληλα ώστε το κάθε αντικείμενο να αναγνωρίζεται σε πολλές διαστάσεις.

Για το τελευταίο, χρησιμοποιούν ένα fully connected conditional Random Field (CRF) το οποίο είναι αποδοτικό υπολογιστικά και εντοπίζει αρκετές λεπτομέρειες διατηρώντας μεγάλο οπτικό πεδίο.



Σχήμα 3.17: Atrous convolution, με διαφορετικά rates.

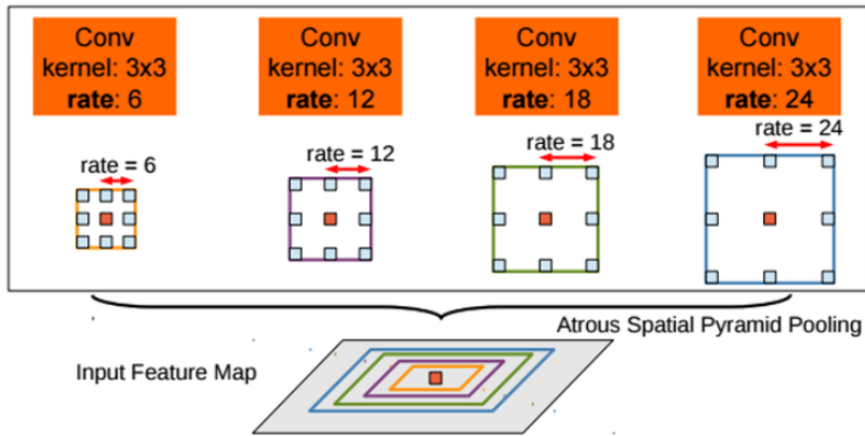
Οι ίδιοι που ανέπτυξαν το Deeplab, βελτίωσαν το δίκτυο τους (Deeplabv2) χρησιμοποιώντας την τεχνική ASPP (Atrous Spatial Pyramid Pooling) η οποία βοήθησε στην καλύτερη περιγραφή διαστάσεων. Με αυτή τη μέθοδο εφαρμόζουν παράλληλες dilated (atrous) συνέλιξεις με διαφορετικά rates στον input feature map και έπειτα τις ενώνουν όλες μαζί.

Σε μετέπειτα έρευνα τους βελτίωσαν το ASPP μοντέλο τους εφαρμόζοντας global average pooling στο τελευταίο feature map του μοντέλου, δίνοντας σε επίπεδο εικόνας χαρακτηριστικά σε μία 1x1 συνέλιξη με 256 φίλτρα με batch normalization και κάνοντας bilinear upsample τα χαρακτηριστικά στην σωστή διάσταση [36].

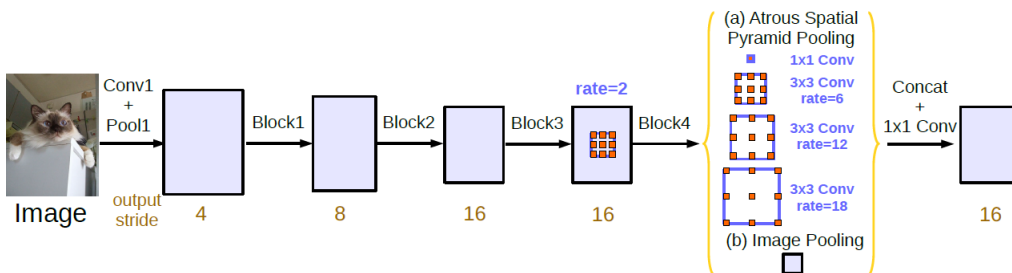
DeconvNet

Το DeconvNet αποτελείται από δύο μέρη. Το πρώτο μέρος του δικτύου είναι ένα συνελκτικό δίκτυο όπως τα συνηθισμένα FCN, με επίπεδα συνέλιξης και pooling. Το δεύτερο μέρος ονομάζεται deconvolution δίκτυο στο οποίο πραγματοποιείται unpooling.

Το unpooling είναι η αντίθετη διαδικασία του pooling αυξάνει δηλαδή τις διαστάσεις της εικόνας και αυτό το επιτυγχάνει αποθηκεύοντας τη θέση στην οπο-



Σχήμα 3.18: ASPP.



Σχήμα 3.19: DeepLabv3.

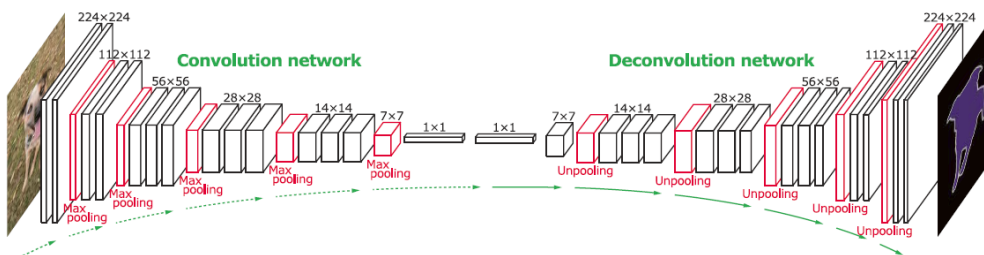
ία βρίσκεται η μέγιστη τιμή ενεργοποίησης όταν πραγματοποιείται max pooling, όπως φαίνεται και στο Σχήμα 3.21.

Στο Σχήμα 3.22 παρουσιάζεται το αποτέλεσμα έπειτα απο εφαρμογή deconvolution επιπέδου από 14x14 έως 224x224. Είναι φανερό ότι το τελευταίο επίπεδο καταφέρνει να κρατήσει τις ιδιότητες του σχήματος του αντικείμενο (ποδηλάτου στη συγκεκριμένη περίπτωση) και να το αναπαραγάγει.

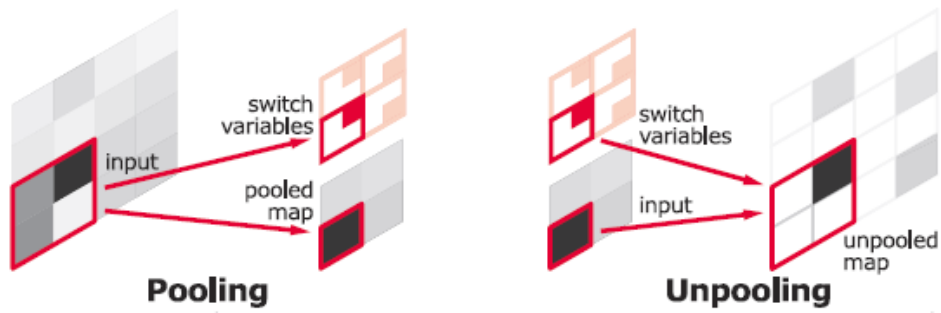
Για την εκπαίδευση του μοντέλου χρησιμοποιήθηκε batch normalization και για το πρώτο μέρος του δικτύου χρησιμοποιήθηκαν τα αρχικά βάρη του VGGNet [37].

SegNet

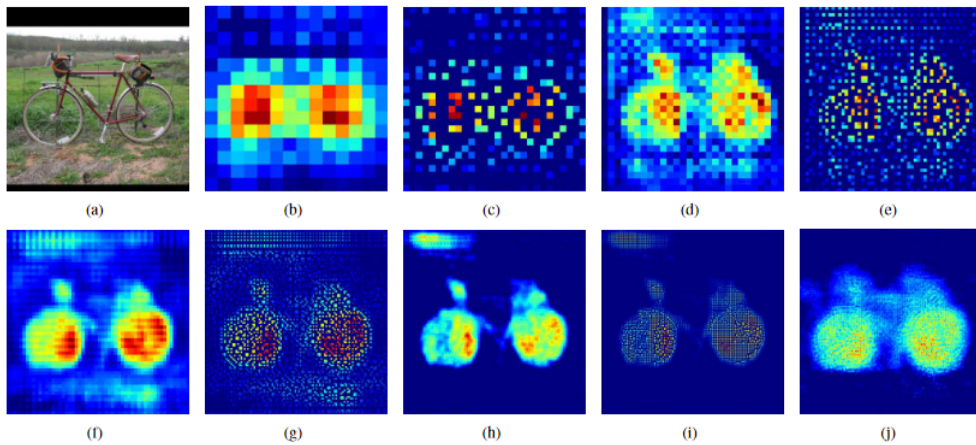
Ομοίως με το deconvNet το SegNet αποτελείται από δύο μέρη, τον encoder και τον decoder. Στο τέλος του decoder εφαρμόζεται ένα pixelwise classification επίπεδο.



Σχήμα 3.20: Η αρχιτεκτονική του DeconvNet.



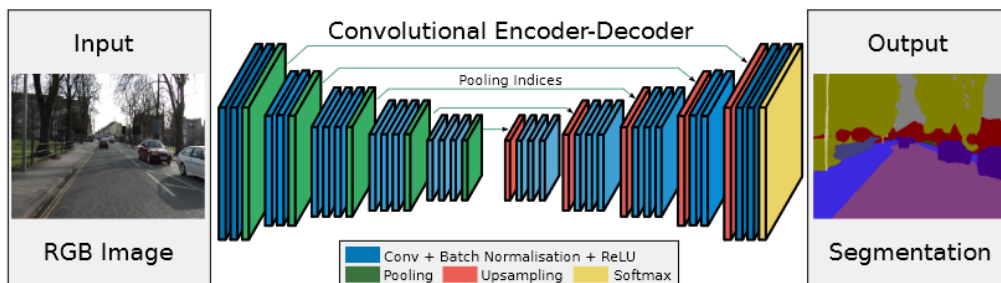
Σχήμα 3.21: Οι διαδικασίες pooling και unpooling. Στην διαδικασία unpooling χρησιμοποιούνται οι προηγούμενες θέσεις των μεγίστων.



Σχήμα 3.22

Το πρώτο μέρος είναι και αυτό ένα συνελκτικό δίκτυο με συνελίξεις και max pooling. Έχει 13 συνελκτικά επίπεδα από το VGG-16, και κατά την διάρκεια του max pooling αποθηκεύει τις θέσεις των μεγίστων στην εικόνα. Το δεύτερο μέρος εφαρμόζει upsampling, χρησιμοποιώντας τις θέσεις που αποθηκεύτηκαν προηγουμένως στο max pooling, και συνελίξεις. Για το τελικό αποτέλεσμα χρησιμοποιείται ένας K-class softmax classifier για να προβλέψει την κλάση του κάθε pixel.

Το SegNet διαφέρει από το DeconvNet και το U-Net στο ότι τα πλήρως συνελκτικά δίκτυα κάνουν το μοντέλο πιο μεγάλο και να απαιτεί περισσότερη μνήμη [38].

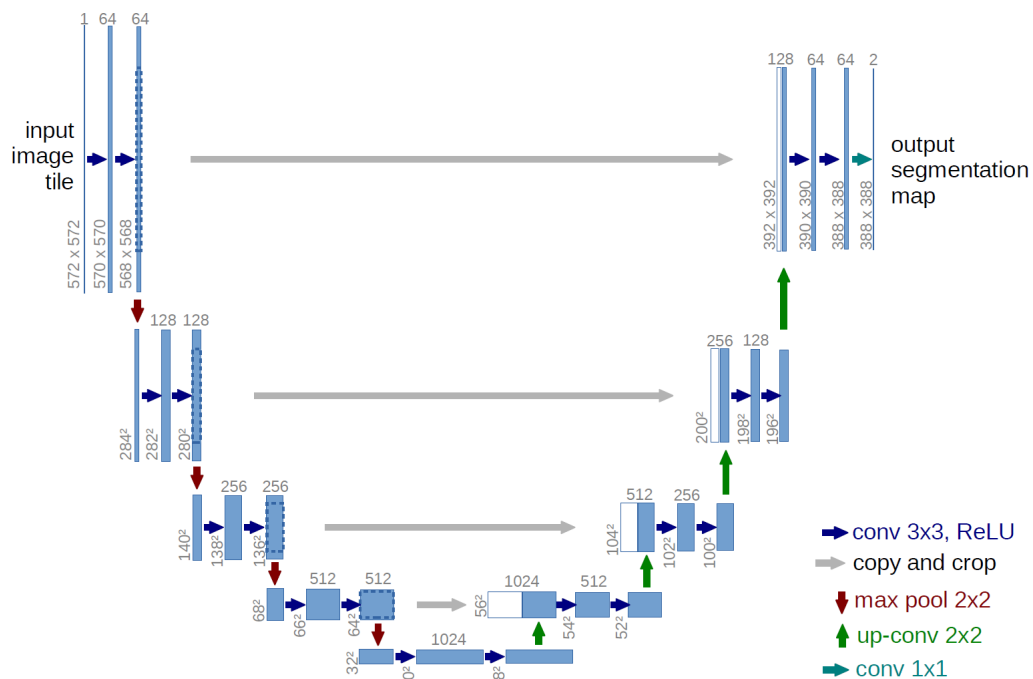


Σχήμα 3.23: Η αρχιτεκτονική του SegNet

U-Net

Τα συνελκτικά δίκτυα τυπικά χρησιμοποιούνταν για προβλήματα κατηγοριοποίησης, στα οποία η κάθε έξοδος αντιστοιχούσε ένα label σε κάθε εικόνα. Ωστόσο, σε προβλήματα ανάλυσης εικόνας (ειδικά σε ιατρικές) απαιτείται σε κάθε pixel να αντιστοιχεί ένα ξεχωριστό label. Το U-Net είναι από τα πιο διάσημα συνελκτικά δίκτυα για σημασιολογική κατάτμηση που χρησιμοποιούνται αυτή την εποχή και ο αρχικός σκοπός του όταν δημιουργήθηκε ήταν για εφαρμογή σε ιατρικές εικόνες. Το συγκεκριμένο μοντέλο χρησιμοποιήθηκε και σε αυτή την διπλωματική εξάγοντας πολύ ενδιαφέροντα αποτελέσματα. Το δίκτυο έχει το χαρακτηριστικό σχήμα U όπως παρουσιάζεται στο Σχήμα 3.24 και αποτελείται από έναν encoder (το αριστερό μέρος) και decoder (το δεξί μέρος).

Η αρχιτεκτονική θα χωριστεί σε blocks. Στο πρώτο μισό του U σχήματος τα blocks αποτελούνται από 2 συνεχόμενες 3x3 συνελίξεις, στο τέλος των οποίων εφαρμόζεται η συνάρτηση ενεργοποίησης ReLU και max pooling με stride 2 για την μείωση του μεγέθους της εικόνας στο μισό. Κάθε φορά που γίνεται αυτό το downsampling τα κανάλια χαρακτηριστικών διπλασιάζονται. Η διαφορά στον decoder (δεύτερο μισό του U σχήματος), είναι ότι στην αρχή εφαρμόζεται 2x2 transposed convolution αυξάνοντας τις διαστάσεις της εικόνας και υποδιπλασιάζοντας τις διαστάσεις στα κανάλια χαρακτηριστικών. Έπειτα πραγματοποιείται ένωση (concatenated) με την αντίστοιχων διαστάσεων εικόνα από τον encoder. Τέλος, εφαρμόζονται οι συνελίξεις και η ReLU, όπως και στον encoder. Για το τελικό αποτέλεσμα πραγματοποιείται συνέλιξη με τον επιθυμητό αριθμό φίλτρων [11].



Σχήμα 3.24: U-Net architecture.

Για την εκπαίδευση χρησιμοποιείται μία weighted cross entropy συνάρτηση κόστους.

Αρχικά ορίζουμε συνάρτηση soft-max

$$p_k(x) = \exp(a_k(x)) / \left(\sum_{k'=1}^K \exp(a_{k'}(x)) \right) \quad (3.42)$$

Όπου $a_k(x)$ είναι η συνάρτηση ενεργοποίησης των feature channels k στην θέση pixel x . K είναι ο αριθμός των κλάσεων και $p_k(x)$ είναι μία συνάρτηση μεγίστου, για την οποία ισχύει $p_k(x) = 1$ όταν για το k η $a_k(x)$ έχει την μέγιστη τιμή και $p_k(x) = 0$ για όλες τις άλλες τιμές. Η συνάρτηση cross entropy είναι η εξής [11]:

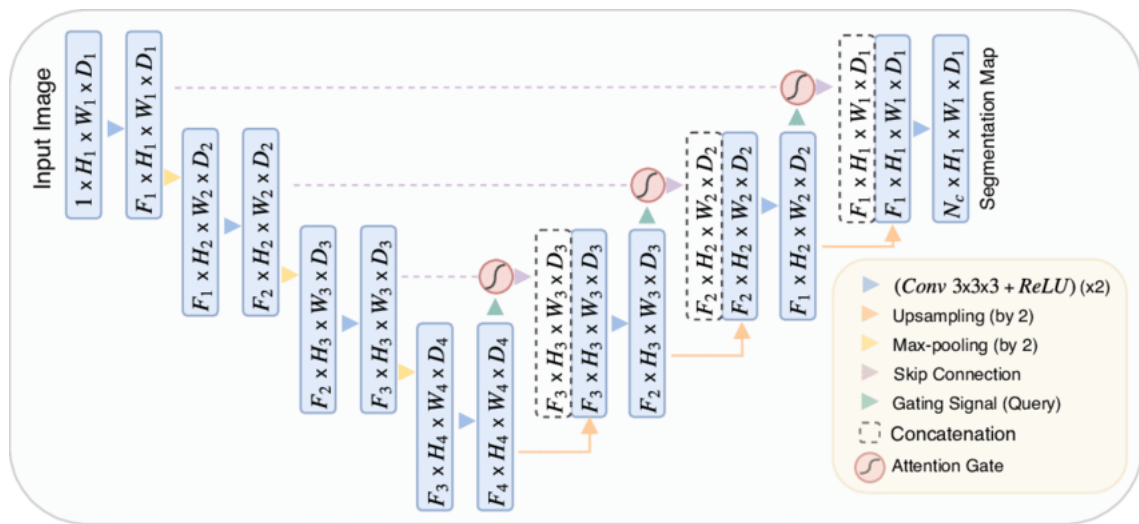
$$E = \sum_{x \in \Omega} w(x) \log(p_{l(x)}(X)) \quad (3.43)$$

$$w(x) = w_c(x) + w_0 \exp\left(-\frac{(d_1(x) + d_2(x))^2}{2\sigma^2}\right) \quad (3.44)$$

Attention U-Net

Το δίκτυο Attention U-Net βασίζεται πάνω στην αρχιτεκτονική του U-Net προσθέτοντας attention gates. Αποτελείται επομένως από έναν encoder και έναν decoder που συνδέονται με skip connections. Στα skip connections εισάγονται τα attention gates και επιτρέπουν στο δίκτυο να ενεργοποιεί ή να καταστέλλει επιλεκτικά ορισμένα χαρακτηριστικά ανάλογα με τη σημασία τους για την τμηματοποίηση. Αυτό βελτιώνει την ικανότητα του δικτύου να τμηματοποιεί με ακρίβεια τα αντικείμενα ενδιαφέροντος, ιδίως σε περιπτώσεις όπου τα αντικείμενα αυτά είναι μικρά ή αποκρύπτονται από άλλα αντικείμενα.

Τα attention gates βασίζονται στην έννοια της squeeze-and-excitation, η οποία παρουσιάστηκε για πρώτη φορά σε μια εργασία του 2018 από τους Hu et al [39]. Η τεχνική αυτή χρησιμοποιεί έναν εκπαιδευσιμο μηχανισμό πύλης για την επαναβαθμολόγηση των χαρτών χαρακτηριστικών, ο οποίος βοηθάει το δίκτυο να εστιάζει σε σημαντικά χαρακτηριστικά και να καταστέλλει τα άσχετα. Οι πύλες προσοχής υλοποιούνται χρησιμοποιώντας έναν συνδυασμό του global average pooling, πλήρως συνδεδεμένων επιπέδων και σιγμοειδών συναρτήσεων ενεργοποίησης [40]. Συνολικά, το Attention U-Net είναι μια ισχυρή αρχιτεκτονική για εργασίες σημασιολογικής τμηματοποίησης, ιδίως σε περιπτώσεις όπου απαιτείται ακριβής τμηματοποίηση μικρών ή καλυμμένων αντικειμένων. Με την ενσωμάτωση πυλών προσοχής, είναι σε θέση να εστιάζει επιλεκτικά στα σχετικά χαρακτηριστικά, οδηγώντας σε βελτιωμένη απόδοση τμηματοποίησης.



Σχήμα 3.25: U-Net architecture.

Κεφάλαιο 4

Προτεινόμενη Μέθοδος

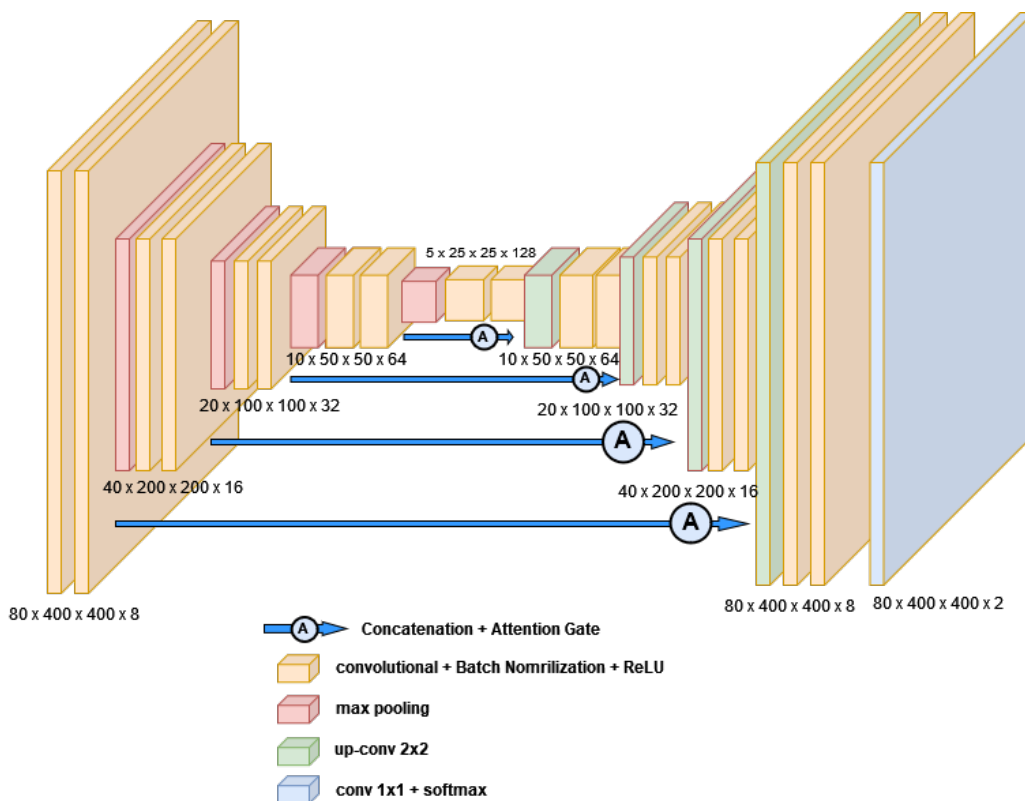
4.1 Κεντρική Ιδέα

Σε αυτή την ενότητα, εισάγεται μια νέα προσέγγιση για τη βελτίωση των σημασιολογικών προβλέψεων. Στην διπλωματική αυτή εργασία μελετάται η κατηγοριοποίηση των οργάνων (i) καρδιά, (ii) ήπαρ, (iii) εγκέφαλο, (iv) νεφρά και (v) ουροδόχος κύστη σε δεδομένα PET/CT.

Τα δεδομένα που χρησιμοποιήθηκαν για την συγκεκριμένη διπλωματική εργασία είναι τρισδιάστατα λόγω των voxels αξονικής τομογραφίας τα οποία θα αναλυθούν περαιτέρω στην ενότητα 4.4. Τα νευρωνικά δίκτυα που χρησιμοποιήθηκαν είναι τύπου U-Net, δηλαδή έχουν το χαρακτηριστικό σχήμα U όπως παρουσιάζεται στο Σχήμα 3.24 και αποτελείται από έναν encoder και decoder. Όπως φαίνεται στις ενότητες 4.1 και 4.2 η αρχιτεκτονική αποτελείται από blocks, που το κάθε ένα επιτελεί διαφορετικές διεργασίες. Για αρχή χρησιμοποιήθηκε ένα τρισδιάστατο Attention U-Net το οποίο τροποποιήθηκε κατάλληλα για τη καλύτερη δυνατή επίδοση. Συγκεκριμένα, χρησιμοποιήθηκαν 4 blocks τα οποία έφταναν μέγιστο αριθμό φίλτρων 128. Με την κατάλληλη προεπεξεργασία και εκπαίδευση τα αποτελέσματα ήταν αρκετά ικανοποιητικά ως προς την κατηγοριοποίηση του κάθε οργάνου. Ωστόσο, όπως ήταν αναμενόμενο λόγω των περιορισμένων δεδομένων οι έξοδοι του μοντέλου μας παρουσίαζαν πολλά ψευδώς θετικά αποτελέσματα τα οποία μειώνουν αρκετά την απόδοση.

Σε μετέπειτα δοκιμή διάφορων μοντέλων, δοκιμάστηκε η εκπαίδευση του δισδιάστατου U-Net. Αρχικά, μετατράπηκαν τα τρισδιάστατα δεδομένα, σε δισδιάστατα στο στεφανιαίο επίπεδο (coronal view), καθώς παρατηρήθηκε ότι το μοντέλο έχει καλύτερη απόδοση σε σύγκριση με το εγκάρσιο και το οβελιαίο. Εκπαιδευόμενος το μοντέλο σε δεδομένα που φαίνεται ολόκληρο το σώμα του ασθενή, μαθαίνει καλύτερα την θέση του κάθε οργάνου και δεν μπερδεύει άλλες δομές με αυτό όπως συμβαίνει στο τρισδιάστατο Attention U-Net που αναφέρθηκε πιο πάνω. Η τελική μορφή του δισδιάστατου U-Net είναι με 4 blocks και με μέγιστο αριθμό φίλτρων 1024. Για το τελικό αποτέλεσμα, ενώθηκαν όλες οι δισδιάστατες μάσκες σε μία, δημιουργώντας μία τρισδιάστατη, ίδιων διαστάσεων με την αρχική PET/CT τομο-

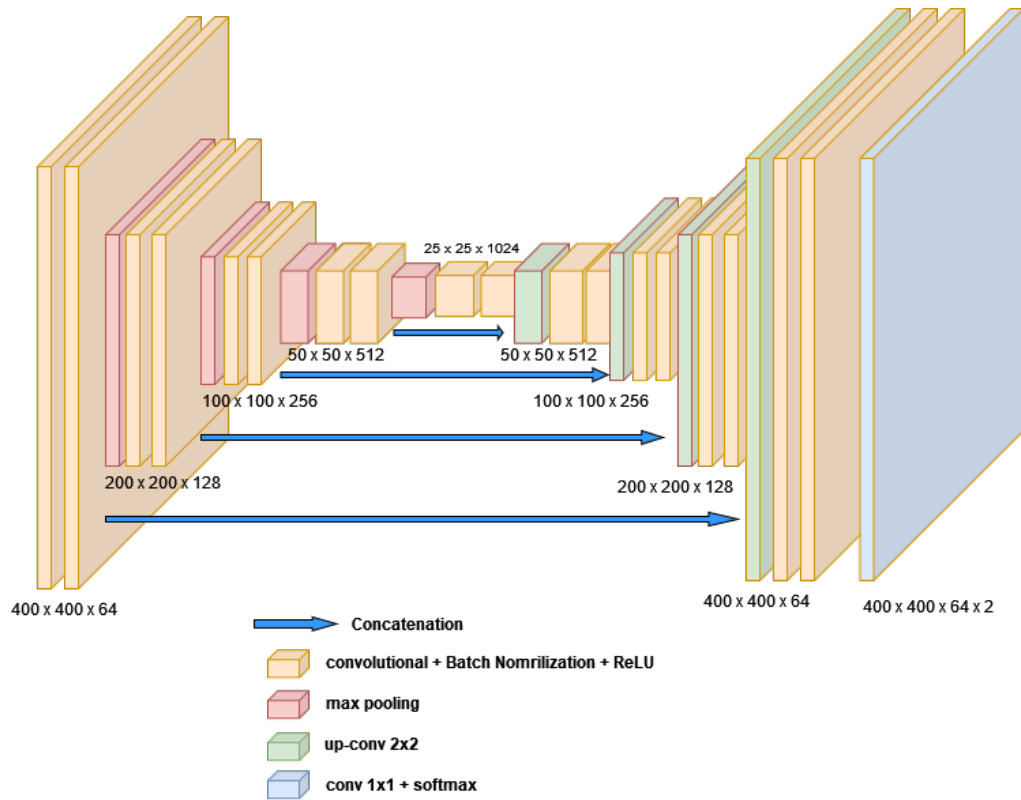
γραφία. Η απόδοση του συγκεκριμένου δικτύου είναι αρκετά υψηλή καθώς αναγνωρίζει το όργανο αλλά επίσης παρουσιάζει ελάχιστα έως μηδενικά ψευδώς θετικά αποτελέσματα. Οι λανθασμένες μάσκες που μπορεί να υπάρχουν είναι εύκολο να απορριφθούν διαλέγοντας και κρατώντας την μεγαλύτερη σε μέγεθος μάσκα. Όπως γίνεται αντιληπτό τα δύο αυτά μοντέλα συμπληρώνουν το ένα το άλλο. Το τρισδιάστατο κατηγοριοποιεί περισσότερη λεπτομέρεια του οργάνου και το δισδιάστατο βοηθάει στην πιο λεπτομερή κατηγοριοποίηση αλλά πιο σημαντικό είναι ότι προσδιορίζει τη τοποθεσία του οργάνου στο χώρο, αφαιρώντας με αυτό τον τρόπο τις ψευδώς θετικές μάσκες. Επομένως, ενώνοντας τα δύο μοντέλα, η απόδοση αυξάνεται έως και 20%.



Σχήμα 4.1: Η αρχιτεκτονική του τρισδιάστατου Attention U-Net μοντέλου μας.

4.2 Προεπεξεργασία

Για όλα τα δεδομένα ισχύει ότι η αρχική τους μορφή δεν είναι η κατάλληλη για την εκπαίδευση ενός νευρωνικού δικτύου. Όταν συλλέγονται τα δεδομένα, αν και είναι ίδιου τύπου και συλλέγονται με την ίδια τεχνική, φέρουν πολλές εσωτερικές διαφορές οι οποίες είναι φανερές στο δίκτυο και μειώνουν την αποτελεσματικότητά του. Στην περίπτωση των εικόνων από δεδομένα PET/CT ένα βασικό χαρακτηριστικό είναι η απόσταση των pixel μεταξύ τους (spacing). Κατά την διάρκεια ανακατασκευής της εικόνας CT μία σειρά από δισδιάστατες εικόνες δημιουργούν την τρισδιάστατη τομογραφία. Οι τιμές απόστασης pixel υποδεικνύουν το πραγματι-



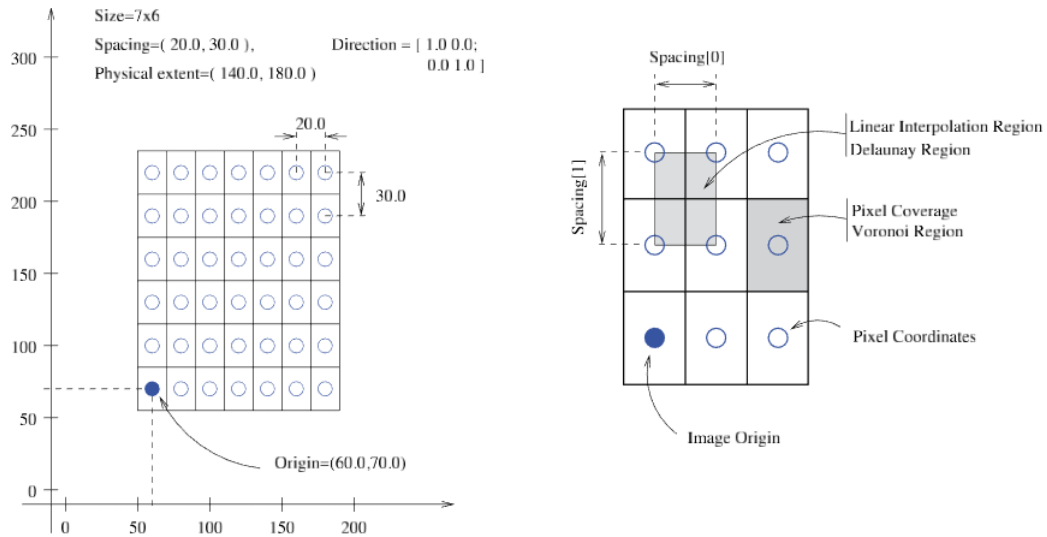
Σχήμα 4.2: Η αρχιτεκτονική του διδιάστατου U-Net μοντέλου μας.

κό μέγεθος ή την κλίμακα του κάθε pixel στην τρισδιάστατη εικόνα. Υπάρχουν τρεις τιμές απόστασης pixel, μία για κάθε άξονα. Στον Z υποδεικνύεται η απόσταση μεταξύ κάθε διδιάστατης εικόνας. Οι τιμές στους X και Y , προσδιορίζουν την απόσταση από το κέντρο του ενός pixel έως το κέντρο των γειτονικών pixel στους X και Y άξονες αντίστοιχα. Επομένως, πριν την εκπαίδευση του δικτύου τροποποιήθηκε η απόσταση pixel όλων των εικόνων σε τιμή $(1, 1, 1)$.

Στην βαθιά μάθηση είναι σημαντικό τα δεδομένα να βρίσκονται στην ίδια κλίμακα και στο ίδιο εύρος τιμών. Η διαδικασία που το επιτυγχάνει ονομάζεται κανονικοποίηση, και δημιουργεί καλοσχηματισμένες σχέσεις ώστε να μη δημιουργούνται προβλήματα συνέπειας, πλεονασμού και εγκυρότητας. Η κανονικοποίηση που χρησιμοποιήθηκε είναι ως προς μέση τιμή και τυπική απόκλιση, δηλαδή στα τελικά δεδομένα η μέση τιμή τους θα είναι 0 και η τυπική απόκλιση 1. Αυτό πραγματοποιείται από την 4.1

$$z = \frac{x - \mu}{\sigma} \quad (4.1)$$

όπου μ , σ είναι η μέση τιμή και η τυπική απόκλιση όλων των δεδομένων αντίστοιχα [41].



Σχήμα 4.3

4.3 Εκπαίδευση

Στην περίπτωση του τρισδιάστατου Attention U-Net, για την εκπαίδευση του δικτύου, λόγω περιορισμού μνήμης, κάθε εικόνα χωρίστηκε σε μέρη διαστάσεων 80x400x400 επιλέγοντας μόνο τα κομμάτια τα οποία περιέχουν μάσκα δηλαδή που εμφανίζεται το κάθε όργανο. Ωστόσο καθώς το δίκτυο μπορεί να χρειαστεί να εφαρμοστεί σε εικόνες που δεν θα περιέχουν το όργανο που θα ψάχνει, είναι σημαντικό να εκπαιδευτεί και σε δεδομένα χωρίς μάσκα. Επομένως, ένα στα τέσσερα κομμάτια χωρίς μάσκα, χρησιμοποιήθηκε στην εκπαίδευση του δικτύου.

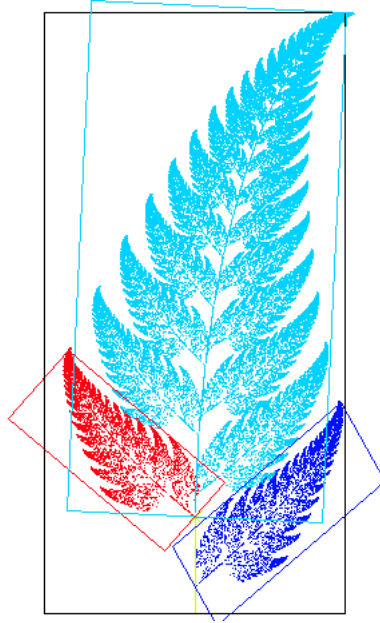
Ακόμη, λόγω των περιορισμένων δεδομένων χρησιμοποιήθηκε data augmentation. Data augmentation είναι η διαδικασία δημιουργίας καινούριων δεδομένων από τα ήδη υπάρχοντα δεδομένα εφαρμόζοντας διάφορες τεχνικές τροποποίησης. Συγκεκριμένα, στο δίκτυο μας στο κάθε δεδομένο εφαρμόστηκε μία τροποποίηση με πιθανότητα 50%. Οι τροποποιήσεις που εφαρμόζονται είναι οι εξής:

- Αφινικός Μετασχηματισμός (Affine transformation): Ο συγκεκριμένος μετασχηματισμός κρατάει ίδιες τις ευθείες και τον παραλληλισμό μεταξύ τους, αλλά όχι τις ευκλείδειες αποστάσεις και τις γωνίες, όπως παρουσιάζεται στην σχήμα 4.5.
- Αλλαγή στην αντίθεση (contrast) της εικόνας. Το contrast είναι η διαφορά στο χρώμα και στη φωτεινότητα που κάνουν ένα αντικείμενο ευδιάκριτο σε σύγκριση με ένα άλλο.
- Γκαουσιανό Blur: Ο συγκεκριμένος μετασχηματισμός είναι το αποτέλεσμα θόλωσης μίας εικόνας εφαρμόζοντας μία Γκαουσιανή συνάρτηση. Είναι φίλτρο το οποίο εφαρμόζει την Γκαουσιανή συνάρτηση σε κάθε pixel. Η συνάρτη-

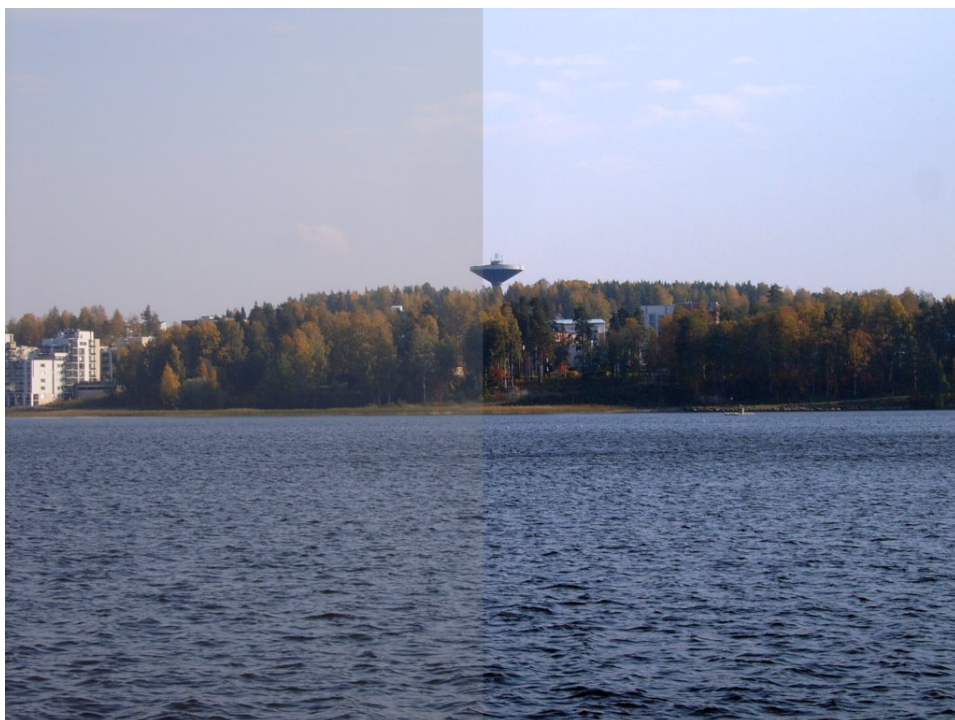
τηση στη μία διάσταση είναι η εξής:

$$G(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \quad (4.2)$$

όπου σ είναι η τυπική απόκλιση (standard deviation) της Γκαουσιανής κατανομής. Στο Σχήμα 4.6 φαίνεται η εφαρμογή του φίλτρου στην ίδια εικόνα για διαφορετικές τιμές της σ .



Σχήμα 4.4: Η συγκεκριμένη εικόνα παρουσιάζει την αφινική αυτό-ομοιότητα. Κάθε φύλλο έχει την δυνατότητα να μετατραπεί σε κάποιο άλλο με αφινικό μετασχηματισμό. Για παράδειγμα το κόκκινο φύλλο μπορεί να μετατραπεί σε οποιοδήποτε άλλο με μετατροπή της γωνίας και του μεγέθους [42].



Σχήμα 4.5: Η αντίθεση στο αριστερό μέρος της εικόνας είναι μικρότερη από την αντίθεση στο δεξιό μέρος [43].



Σχήμα 4.6: Εφαρμογή Γκαουσιανού Blur για διαφορετικές τιμές τυπικής απόκλισης [44].

Επιπλέον, σημαντικό για την εκπαίδευση του μοντέλου μας είναι η επιλογή κατάλληλου βελτιστοποιητή (optimizer) ο οποίος έχει πολύ σημαντικό ρόλο στην απόδοση του δικτύου. Ο βελτιστοποιητής είναι μέθοδος ή αλγόριθμος που χρησιμοποιείται για την ελαχιστοποίηση της συνάρτησης κόστους. Μία τέτοια μέθοδος που αναφέρθηκε στα προηγούμενα κεφάλαια είναι ο αλγόριθμος απότομης καθόδου.

Στην περίπτωση της παρούσας διπλωματικής επιλέχθηκε να χρησιμοποιηθεί ο αλγόριθμος Adam [45]. Σε αντίθεση με την μέθοδο στοχαστικής σταδιακής καθόδου που κρατάει σταθερό learning rate κατά την διάρκεια της εκπαίδευσης, ο Adam συνδυάζει τα πλεονεκτήματα δύο παραλλαγών της στοχαστικής σταδιακής καθόδου των Adaptive Gradient Algorithm (AdaGrad) [46] και Root Mean Square Propagation (RMSProp) [47]. Επεξεργάζεται διαφορετικά learning rates για διαφορετικά βάρη από εκτιμήσεις της πρώτης και της δεύτερης ροπής των κλίσεων (Μέση Τιμή και Διακύμανση). Συγκεκριμένα, για να εκτιμήσει τις ροπές, ο Adam, χρησιμοποιεί εκθετικά εναλλασσόμενους μέσους όρους της ροπής (μ_t) και την τετραγωνισμένη ροπή (u_t). Οι καινούριες παράμετροι $\beta_1, \beta_2 \in [0, 1)$ ελέγχουν τον ρυθμό μείωσης των εναλλασσόμενων μέσων. Τέλος, για την συνάρτηση που επιδιώκει να ελαχιστοποιήσει ο βελτιστοποιητής Adam επιλέχθηκε η dice loss, και το δίκτυο εκπαιδεύτηκε συνολικά για 50 εποχές για κάθε όργανο.

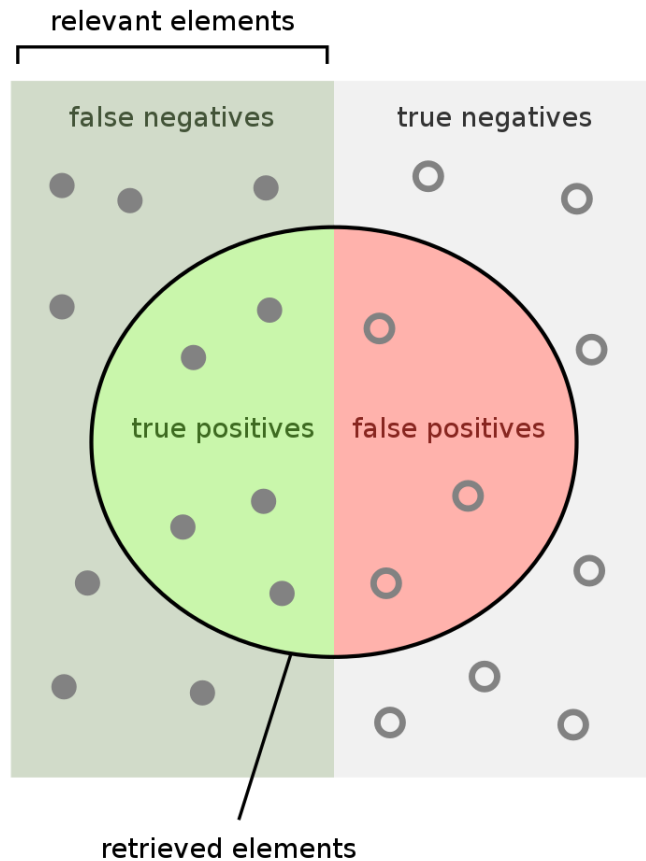
Στην περίπτωση του δισδιάστατου U-Net χρησιμοποιήθηκε ο ίδιος τρόπος εκπαίδευσης με την διαφορά ότι δεν χρησιμοποιήθηκε Data augmentation και το μέγεθος του batch είναι 10.

Για την αξιολόγηση των αποτελεσμάτων του μοντέλου χρησιμοποιήθηκαν 4 μετρικές. Η dice score, η Intersection over Union (IoU) και οι precision και recall. Οι δύο πρώτες αναλύθηκαν στα προηγούμενα κεφάλαια 3.2.3 και 3.2.3

Το precision και το recall είναι δύο σημαντικές μετρικές που χρησιμοποιούνται για την αξιολόγηση της απόδοσης των μοντέλων μηχανικής μάθησης, ιδίως για προβλήματα ταξινόμησης.

Το precision είναι ένα μέτρο της ακρίβειας των θετικών προβλέψεων που κάνει το μοντέλο. Ορίζεται ως ο αριθμός των αληθώς θετικών προβλέψεων διαιρούμενος με το άθροισμα των αληθώς θετικών προβλέψεων και των ψευδώς θετικών προβλέψεων. Το precision μετράει το ποσοστό των θετικών προβλέψεων που είναι πραγματικά σωστές.

Το recall, από την άλλη πλευρά, είναι ένα μέτρο της ικανότητας του μοντέλου να κατηγοροποιεί σωστά την περιοχή της κάθε κλάσης. Ορίζεται ως ο αριθμός των αληθώς θετικών προβλέψεων διαιρούμενος με το άθροισμα των αληθώς θετικών προβλέψεων και των ψευδώς αρνητικών προβλέψεων.



How many retrieved items are relevant?

$$\text{Precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}$$

How many relevant items are retrieved?

$$\text{Recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$

Σχήμα 4.7: Precision και Recall [48]

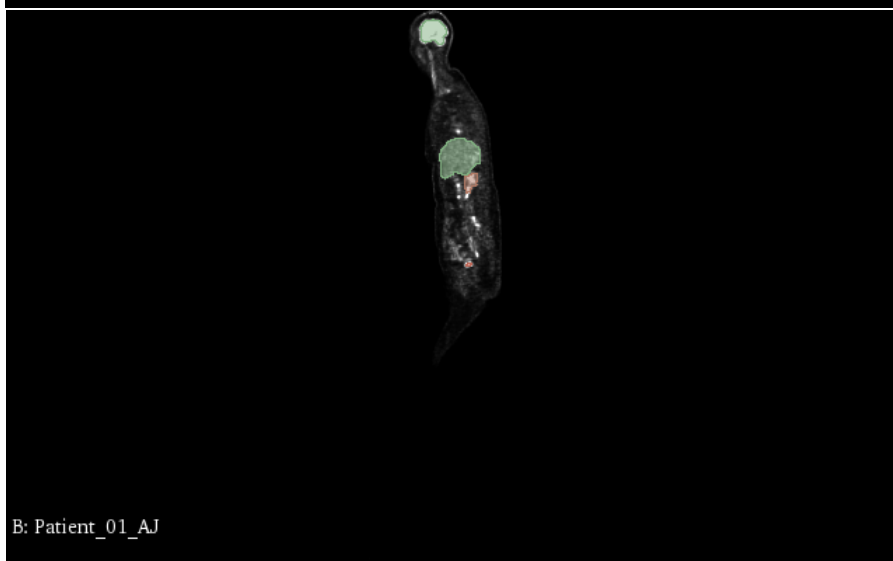
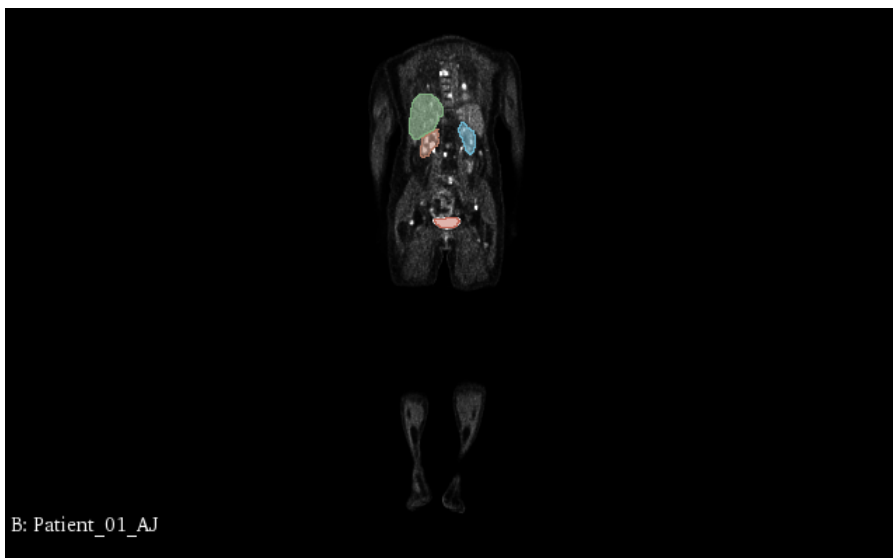
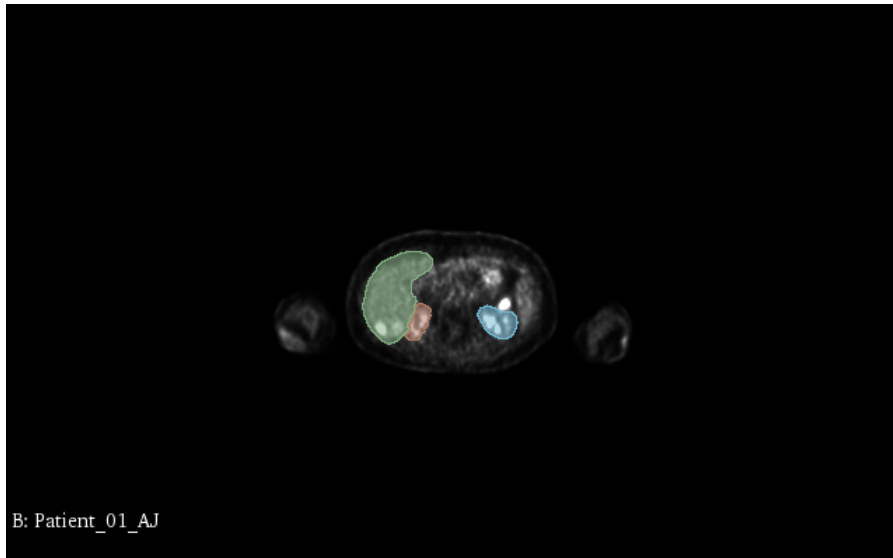
Μαζί οι δύο μετρικές παρέχουν μια πιο ολοκληρωμένη αξιολόγηση της απόδοσης ενός μοντέλου, καθώς μας δίνουν πληροφορίες τόσο για την ακρίβεια των θετικών προβλέψεων όσο και για την ικανότητα του μοντέλου να βρίσκει ολόκληρες τις μάσκες.

Στην πράξη, υπάρχει συχνά ένας συμβιβασμός μεταξύ precision και recall, καθώς η αύξηση της μιας μετρικής έχει συχνά ως αποτέλεσμα τη μείωση της άλλης. Επομένως, είναι σημαντικό να εξετάζονται και οι δύο μαζί κατά την αξιολόγηση της απόδοσης ενός μοντέλου και να επιλεγεί εκείνο που παρέχει την καλύτερη ισορροπία μεταξύ των δύο για ένα δεδομένο πρόβλημα.

4.4 Δεδομένα

Τα δεδομένα που χρησιμοποιήθηκαν για την εκπαίδευση και επαλήθευση των νευρωνικών μοντέλων είναι PET/CT από 19 ασθενείς, οι οποίοι πάσχουν από μεταστατικό μελάνωμα. Ο κάθε ασθενής υποβλήθηκε σε PET/CT ολόκληρου σώματος έπειτα από ενδοφλέβια δόση των $250 \text{ MBq}^{18}\text{F} - \text{FDG}$ 60 λεπτά μετά την έγχυση. Η απεικόνιση πραγματοποιήθηκε από το κεφάλι προς τα πόδια, διάρκειας 2 λεπτά για κάθε διαφορετική θέση του κρεβατιού. Χρησιμοποιήθηκε ένα εξειδικευμένο σύστημα (Biograph mCT, S128, Siemens Co., Erlangen, Germany), με εγκάρσιο οπτικό πεδίο των 21.6cm με TruePoint και TrueV. Ακόμη, χρησιμοποιήθηκε CT (120kV, 30mA) για την διόρθωση της απόσβεσης και ένωσης των εικόνων. Όλες οι εικόνες ήταν διορθωμένες ως προς την απόσβεση και χρησιμοποιήθηκε πίνακας μεγέθους (400x400) για την επαναληπτική ανακατασκευή εικόνας. Η επαναληπτική ανακατασκευή εικόνας βασίστηκε στον αλγόριθμο OSEM με 2 επαναλήψεις και 21 υποσύνολα καθώς και την TOF.





Σχήμα 4.8: Ενδεικτική παρουσίαση των δεδομένων μέσω του προγράμματος 3D Slicer.

4.5 Αποτελέσματα και Συγκρίσεις

Η αξιολόγηση των αποτελεσμάτων πραγματοποιήθηκε με τυχαία επιλογή στο 20% των δεδομένων, που αναφέρονται παραπάνω. Για την αξιολόγηση των αποτελεσμάτων του μοντέλου χρησιμοποιήθηκαν 4 μετρικές. Η dice score, η Intersection over Union (IoU) και οι precision και recall.

Στους πίνακες 4.1, 4.3, 4.4, 4.5 και 4.6 παρουσιάζονται τα αποτελέσματα της εφαρμογής του μοντέλου στα όργανα (i) αριστερή κοιλία της καρδιάς, (ii) ήπαρ, (iii) εγκέφαλο, (iv) νεφρά και (v) ουροδόχος κύστη αντίστοιχα. Στον πίνακα 4.7 αναφέρονται οι μέσοι όροι των συνολικών αποτελεσμάτων.

Μοντέλο	Dice Score	IoU	Precision	Recall
Attention U-Net	0.6550	0.5939	0.6072	0.7972
2D U-Net	0.8794	0.8114	0.9391	0.8345
Ensemble	0.9106	0.851	0.96	0.8715

Πίνακας 4.1: Αποτελέσματα εφαρμογής του μοντέλου στην αριστερή κοιλία της καρδιάς.

Μοντέλο	Dice Score	IoU	Precision	Recall
Attention U-Net	0.8117	0.7267	0.7537	0.9138
2D U-Net	0.9031	0.843	0.862	0.9601
Ensemble	0.972	0.9473	0.981	0.9643

Πίνακας 4.2: Αποτελέσματα εφαρμογής του μοντέλου στην ουροδόχο κύστη.

Μοντέλο	Dice Score	IoU	Precision	Recall
Attention U-Net	0.8037	0.7224	0.7751	0.8399
2D U-Net	0.698	0.6241	0.6306	0.9452
Ensemble	0.9324	0.882	0.9165	0.9528

Πίνακας 4.3: Αποτελέσματα εφαρμογής του μοντέλου στο ήπαρ.

Μοντέλο	Dice Score	IoU	Precision	Recall
Attention U-Net	0.7029	0.6263	0.6336	0.9361
2D U-Net	0.7999	0.7607	0.9932	0.7664
Ensemble	0.9721	0.9477	0.9913	0.9557

Πίνακας 4.4: Αποτελέσματα εφαρμογής του μοντέλου στον εγκέφαλο.

Μοντέλο	Dice Score	IoU	Precision	Recall
Attention U-Net	0.6845	0.6148	0.6273	0.8408
2D U-Net	0.9113	0.85	0.9194	0.9066
Ensemble	0.9201	0.8625	0.9135	0.9292

Πίνακας 4.5: Αποτελέσματα εφαρμογής του μοντέλου στον αριστερό νεφρό.

Μοντέλο	Dice Score	IoU	Precision	Recall
Attention U-Net	0.6384	0.5867	0.6151	0.674
2D U-Net	0.8912	0.826	0.9319	0.8682
Ensemble	0.8966	0.8335	0.9317	0.8792

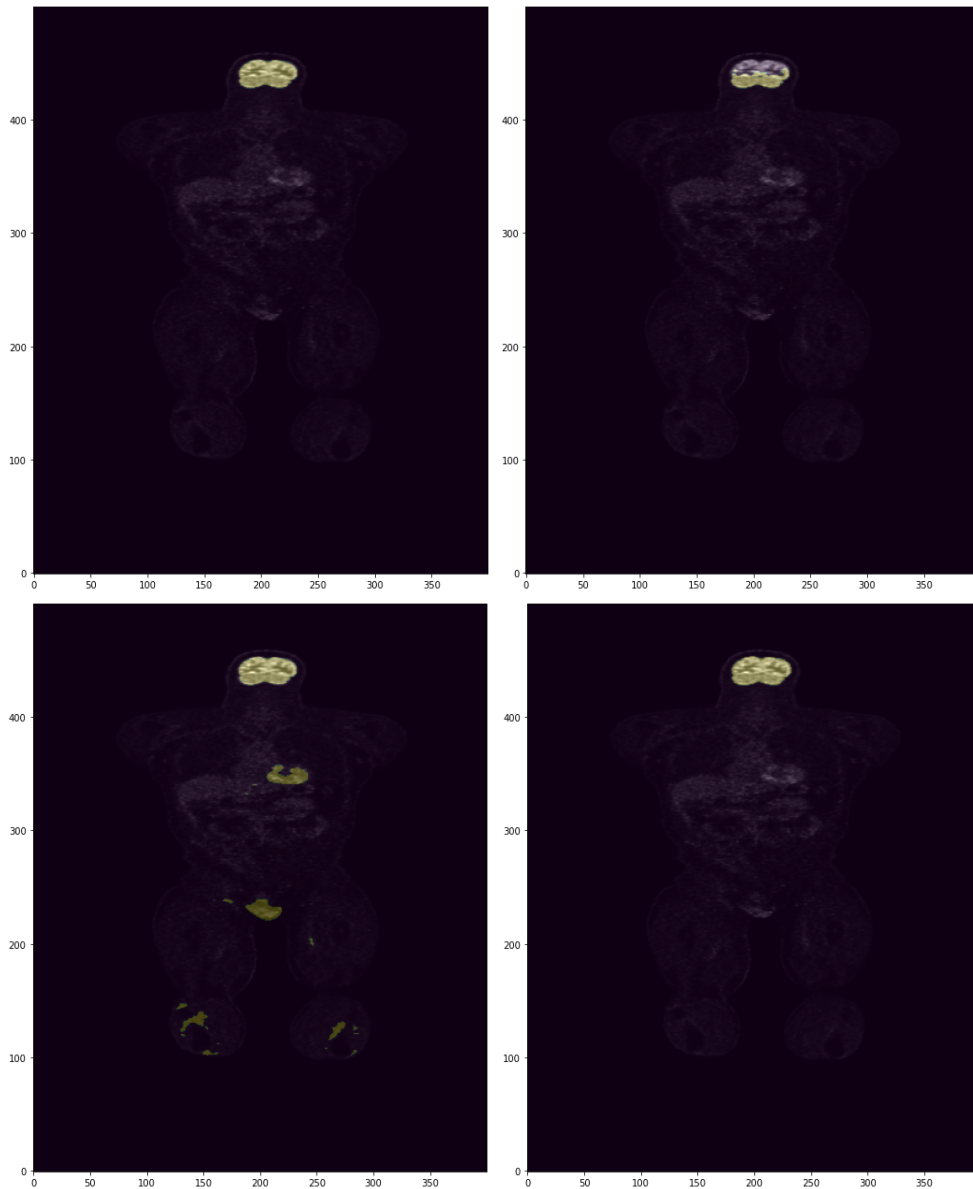
Πίνακας 4.6: Αποτελέσματα εφαρμογής του μοντέλου στον δεξιό νεφρό.

Μοντέλο	Dice Score	IoU	Precision	Recall
Attention U-Net	0.716	0.6451	0.6686	0.8337
2D U-Net	0.8471	0.7858	0.8793	0.8801
Ensemble	0.934	0.8873	0.949	0.9254

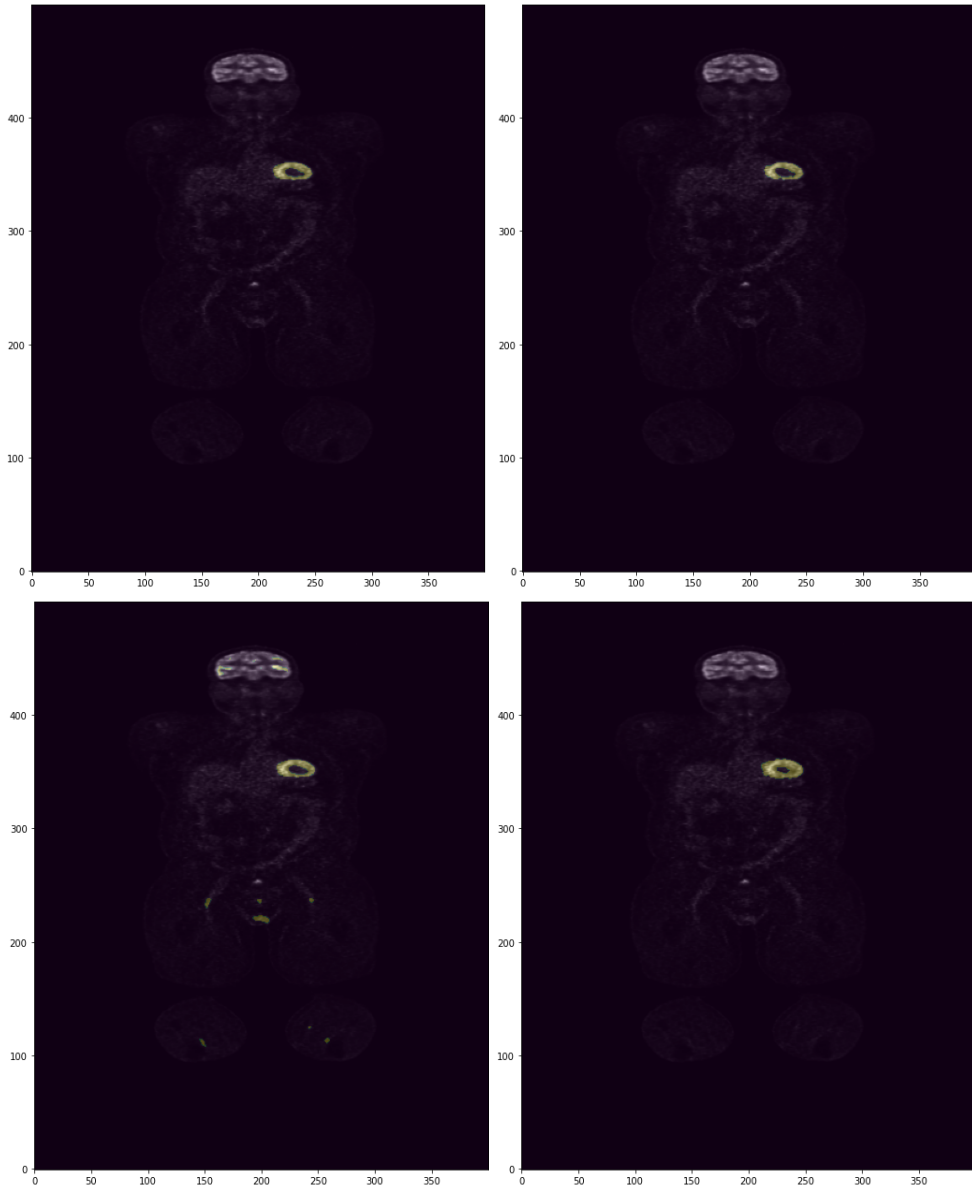
Πίνακας 4.7: Μέσος όρος αποτελεσμάτων εφαρμογής του μοντέλου σε όλα τα όργανα που αναφέρθηκαν παραπάνω.

Από τα αποτελέσματα γίνεται αντιληπτό ότι τα ψευδώς θετικά επηρεάζουν πολύ την απόδοση του τρισδιάστατου Attention U-Net, καθώς κατά μέσο όρο το precision είναι στο 66% ενώ το recall στο 83%. Στην περίπτωση του δισδιάστατου παρατηρούμε ότι κατηγοριοποιεί καλύτερα τα μικρότερα σε μέγεθος όργανα χωρίς ψευδώς θετικά στη τελική μάσκα. Όπως παρουσιάζεται και στο σχήμα 4.10 (με κίτρινο χρώμα απεικονίζεται η μάσκα), στην διδιάστατη κατηγοριοποίηση υπάρχουν μηδενικά ψευδώς θετικά αποτελέσματα. Αντίθετως, στα μεγαλύτερα σε μέγεθος όργανα παρατηρούνται ψευδώς θετικές μάσκες, οι οποίες όμως είναι εύκολο να απορριφθούν, καθώς είναι σίγουρα μικρότερες σε μέγεθος από την μάσκα του πραγματικού οργάνου. Αυτό είναι σίγουρο καθώς γίνεται εφαρμογή του μοντέλου σε δισδιάστατες εικόνες και οι ψευδώς θετικές μάσκες σε διαφορετικές εικόνες δεν θα έχουν συνοχή μεταξύ τους.

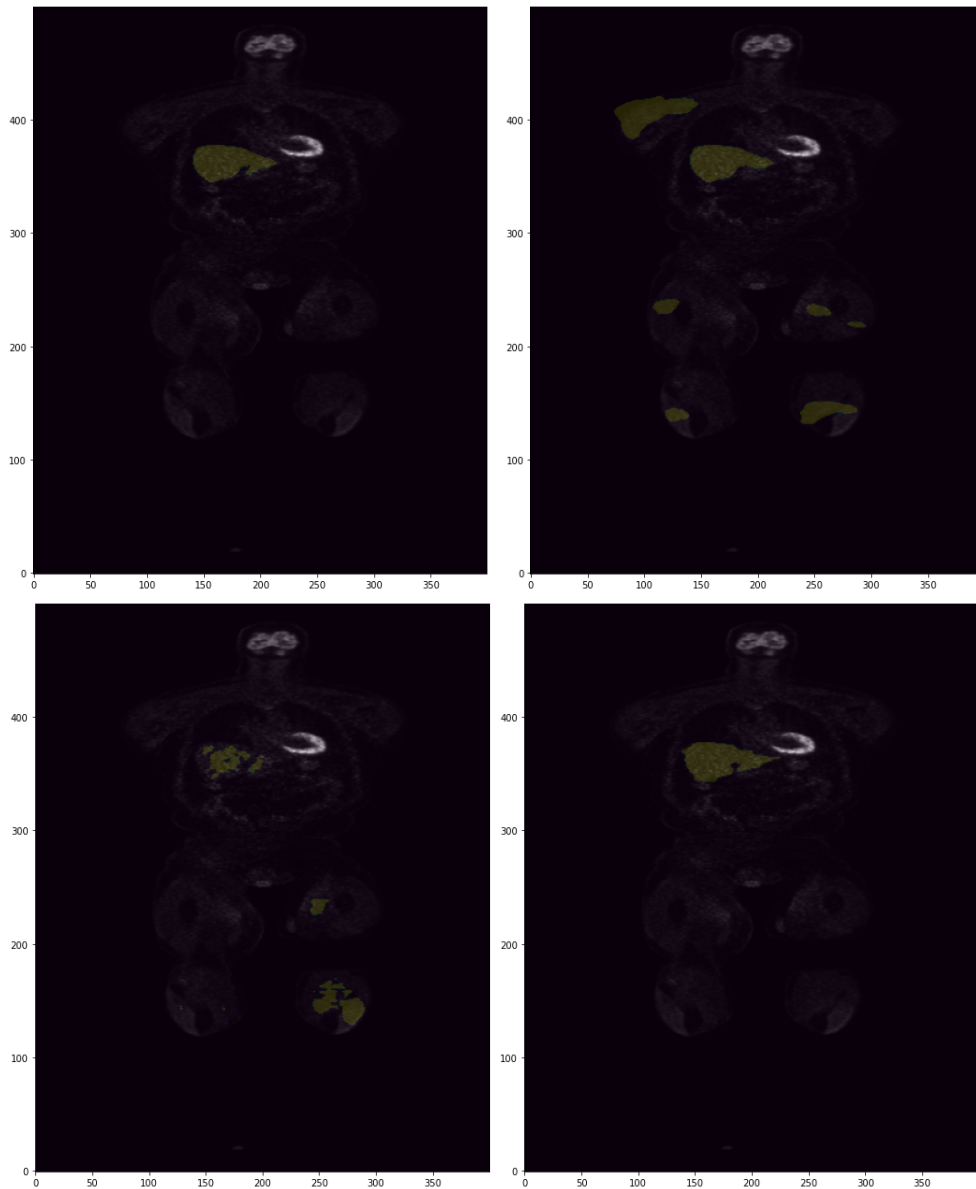
Όσον αφορά το dice score, του τρισδιάστατου Attention U-Net, οι τιμές του είναι μικρότερες από αυτές του δισδιάστατου U-Net, λόγω του χαμηλού precision, παρόλο που το recall έχει καλές αποδόσεις. Αντίθετα, το δισδιάστατο U-Net παρουσιάζει σταθερά υψηλές τιμές στο precision, εμφανίζει επομένως λίγα ψευδώς θετικά αποτελέσματα, κάτι που το καθιστά αρκετά αποδοτικό στον εντοπισμό του οργάνου. Συνδυάζοντας και τα δύο μοντέλα μαζί υπάρχει σαφής αύξηση της απόδοσης στις παραπάνω μετρικές.



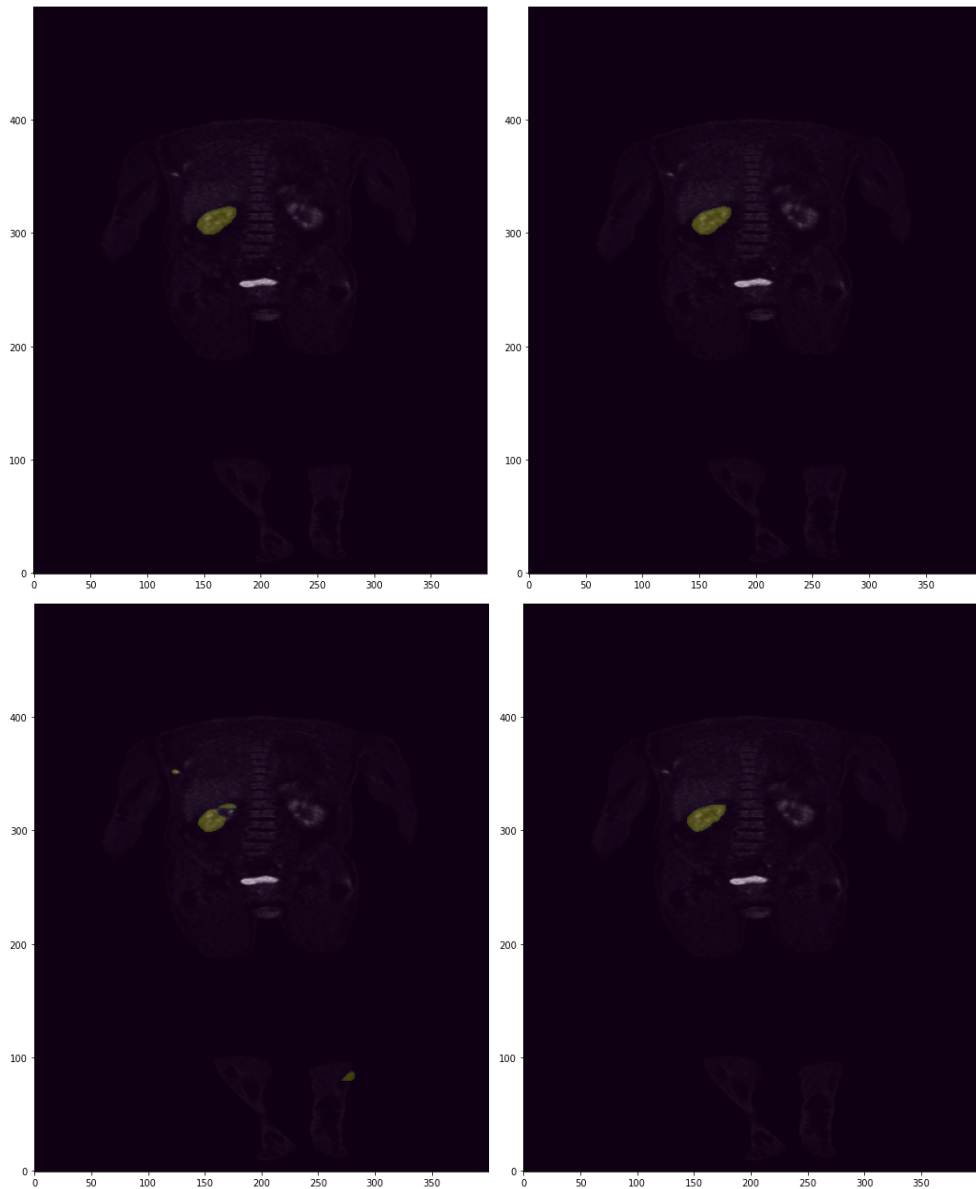
Σχήμα 4.9: Κατηγοριοποίηση του εγκεφάλου. Στην πρώτη φαίνεται το αποτέλεσμα του ensembled μοντέλου, στην δεύτερη του διδιάστατου U-Net, στην τρίτη του τρισδιάστατου Attention U-Net και στην τέταρτη είναι το επιθυμητό αποτέλεσμα.



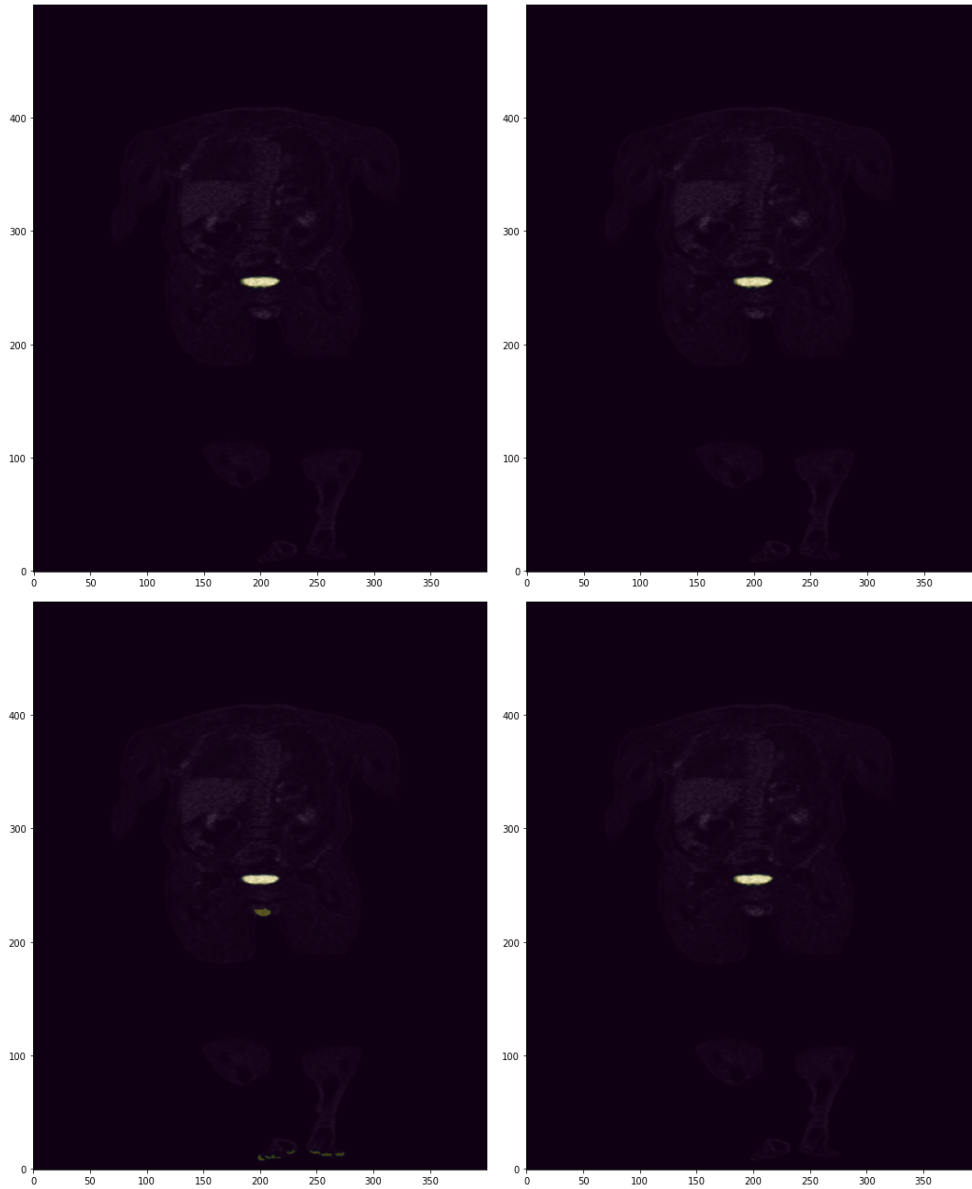
Σχήμα 4.10: Κατηγοριοποίηση της αριστερή κοιλίας της καρδιάς. Στην πρώτη φαίνεται το αποτέλεσμα του ensembled μοντέλου, στην δεύτερη του δισδιάστατου U-Net, στην τρίτη του τρισδιάστου Attention U-Net και στην τέταρτη είναι το επιθυμητό αποτέλεσμα.



Σχήμα 4.11: Κατηγοριοποίηση του ήπατος. Στην πρώτη φαίνεται το αποτέλεσμα του ensembled μοντέλου, στην δεύτερη του διδιάστατου U-Net, στην τρίτη του τρισδιάστατου Attention U-Net και στην τέταρτη είναι το επιθυμητό αποτέλεσμα.



Σχήμα 4.12: Κατηγοριοποίηση του δεξιού νεφρού. Στην πρώτη φαίνεται το αποτέλεσμα του ensembled μοντέλου, στην δεύτερη του διδιάστατου U-Net, στην τρίτη του τρισδιάστου Attention U-Net και στην τέταρτη είναι το επιθυμητό αποτέλεσμα.



Σχήμα 4.13: Κατηγοριοποίηση της ουροδόχου κύστης. Στην πρώτη φαίνεται το αποτέλεσμα του ensembled μοντέλου, στην δεύτερη του διδιάστατου U-Net, στην τρίτη του τρισδιάστου Attention U-Net και στην τέταρτη είναι το επιθυμητό αποτέλεσμα.

Κεφάλαιο 5

Συμπεράσματα

Στη διπλωματική αυτή εργασία παρουσιάζεται μία νέα μέθοδος συνδυασμού μοντέλων σημασιολογικής κατάτμησης οργάνων από δεδομένα PET/CT. Η μέθοδος αυτή χρησιμοποιεί διαφορετικά δίκτυα από την άποψη ότι δέχονται διαφορετικών διαστάσεων εικόνες. Το ένα δέχεται τρισδιάστατα και το άλλο δισδιάστατα δεδομένα. Ο συγκεκριμένος συνδυασμός τους, αυξάνει την απόδοση του δικτύου σε μερικές περιπτώσεις έως και 20%, όσον αφορά την dice coefficient.

Ο μέσος όρος των αποτελεσμάτων είναι:

Μοντέλο	Dice Score	IoU	Precision	Recall
Attention U-Net	0.716	0.6451	0.6686	0.8337
2D U-Net	0.8471	0.7858	0.8793	0.8801
Ensemble	0.934	0.8873	0.949	0.9254

Πολύ σημαντικό για την συγκεκριμένη μέθοδο είναι η υψηλή απόδοση του δισδιάστατου μοντέλου. Σε όλα τα μικρού μεγέθους όργανα το δίκτυο μας διατηρεί παρόμοιες τιμές. Στα μεγαλύτερα σε μέγεθος όργανα (ήπαρ, εγκέφαλος) οι τιμές διαφοροποιούνται. Στο ήπαρ παρουσιάζει χαμηλό precision και υψηλό recall, ενώ στον εγκέφαλο ακριβώς το αντίθετο.

Πολύ εύκολα όργανα στην κατηγοριοποίηση τους ήταν ο εγκέφαλος και η ουροδόχος κύστη καθώς φτάνουμε στο τελικό νευρωνικό 97% dice score απόδοση.

Στα νεφρά, το τρισδιάστατο δίκτυο πραγματοποίησε τις χειρότερες αποδόσεις, σε σύγκριση με τα υπόλοιπα όργανα, αλλά το δισδιάστατο είχε τις πιο υψηλές του.

Αντίθετα, στην περίπτωση του ήπατος, το δισδιάστατο μοντέλο επιτυγχάνει τις πιο χαμηλές του τιμές, και το τρισδιάστατο σχεδόν τις πιο υψηλές.

Για την περαιτέρω βελτίωση της μεθόδου, δηλαδή την αύξηση της απόδοσης, μπορούν να εξεταστούν περισσότερα δίκτυα που βελτιώνουν το U-Net. Συγκεκριμένα στα μικρότερου μεγέθους όργανα, όπως στην περίπτωση των νεφρών, το τρισδιάστατο Attention U-Net που χρησιμοποιήθηκε δεν είχε την δυνατότητα να τα κατηγοριοποιήσει με αρκετή ακρίβεια, ώστε να αυξήσει την τελική απόδοση του ensemble δικτύου. Επομένως, η πιο λεπτομερής κατηγοριοποίηση των τρισδιάστατων

εικόνων θα βελτιώνει πάντα το μοντέλο μας. Αντίθετα, η βελτίωση του δισδιάτατου νευρωνικού, είναι πιθανό να μην έχει ιδιαίτερη σημασία για το συγκεκριμένο ensemble. Αυτό επειδή, το κλασικό U-Net που χρησιμοποιήθηκε, εντοπίζει με επιτυχία την περιοχή που βρίσκεται το όργανο και η απόδοση του είναι πιο υψηλή από εκείνη του τρισδιάστατου.

Επιπροσθέτως, σημαντικό ρόλο στην απόδοση έχει η υπολογιστική ισχύς. Στην παρούσα διπλωματική λόγω των περιορισμένων πόρων, η εκπαίδευση των νευρωνικών δικτύων πραγματοποιήθηκε τροφοδοτώντας τις εικόνες σε κομμάτια και με batch 1. Όπως είναι γνωστό, το μέγεθος του αριθμού του batch είναι από τους κύριους παράγοντες που επηρεάζουν την απόδοση του δικτύου. Επομένως, με την αύξηση του συγκεκριμένου αριθμού περιμένουμε σαφή βελτίωση των αποτελεσμάτων, καθώς και αρκετή μείωση του χρόνου εκπαίδευσης.

Βιβλιογραφία

- [1] NeuroLogica. “PET scan vs. CT scan”. (), [Online]. Available: <https://www.neurologica.com/blog/pet-scan-vs-ct-scan> (visited on 02/17/2023).
- [2] T. M. Buzug, “Computed tomography”, in *Springer Handbook of Medical Technology*, ser. Springer Handbooks, R. Kramme, K.-P. Hoffmann, and R. S. Pozos, Eds., Berlin, Heidelberg: Springer, 2011, pp. 311-342.
- [3] R. E. Schmitz, A. M. Alessio, and P. E. Kinahan, “The physics of PET/CT scanners”,
- [4] S. Surti and J. S. Karp, “Advances in time-of-flight PET”, *Physica Medica*, vol. 32, no. 1, pp. 12-22, Jan. 1, 2016.
- [5] D. Carrick, J. Dickson, and A. Bradley, “Basic principles of PET/CT imaging”, in *PET/CT Imaging: Basics and Practice*, ser. Clinicians’ Guides to Radionuclide Hybrid Imaging, K. Agrawal, A. Skillen, A. Esmail, and S. Usmani, Eds., Cham: Springer International Publishing, 2022, pp. 1-12.
- [6] K. O’Shea and R. Nash, *An introduction to convolutional neural networks*, Dec. 2, 2015.
- [7] E. W. Weisstein. “Heaviside step function”. Publisher: Wolfram Research, Inc. (), [Online]. Available: <https://mathworld.wolfram.com/> (visited on 01/27/2023).
- [8] V. Nair and G. E. Hinton, “Rectified linear units improve restricted boltzmann machines”,
- [9] J. Han and C. Moraga, “The influence of the sigmoid function parameters on the speed of backpropagation learning”, in *From Natural to Artificial Neural Computation*, J. Mira and F. Sandoval, Eds., ser. Lecture Notes in Computer Science, Berlin, Heidelberg: Springer, 1995, pp. 195-201.
- [10] B. Gao and L. Pavel, *On the properties of the softmax function with application in game theory and reinforcement learning*, Aug. 20, 2018.
- [11] O. Ronneberger, P. Fischer, and T. Brox, *U-net: Convolutional networks for biomedical image segmentation*, May 18, 2015.

- [12] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, *et al.*, “Convolutional neural networks for medical image analysis: Full training or fine tuning?”, *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1299–1312, May 2016.
- [13] “convolutionSeparable”, 2007.
- [14] M. Hashemi, “Enlarging smaller images before inputting into convolutional neural network: Zero-padding vs. interpolation”, *Journal of Big Data*, vol. 6, no. 1, p. 98, Dec. 2019.
- [15] S. Saha. “A comprehensive guide to convolutional neural networks — the ELI5 way”, Medium. (Nov. 16, 2022), [Online]. Available: <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53> (visited on 11/29/2022).
- [16] M. Sun, Z. Song, X. Jiang, J. Pan, and Y. Pang, “Learning pooling for convolutional neural network”, *Neurocomputing*, vol. 224, pp. 96–104, Feb. 2017.
- [17] P. Mahajan. “Max pooling”, Medium. (Jul. 5, 2020), [Online]. Available: <https://poojamahajan5131.medium.com/max-pooling-210fc94c4f11> (visited on 01/27/2023).
- [18] A. Youssef, “Image downsampling and upsampling methods”,
- [19] P. Hurtik and N. Madrid, “Bilinear interpolation over fuzzified images: Enlargement”, in *2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, Istanbul, Turkey: IEEE, Aug. 2015, pp. 1–8.
- [20] *Bilinear interpolation*, in *Wikipedia*, Page Version ID: 1119799277, Nov. 3, 2022.
- [21] S. Santurkar, D. Tsipras, A. Ilyas, and A. Ma, “How does batch normalization help optimization?”,
- [22] J. Huber. “Batch normalization in 3 levels of understanding”, Medium. (Jan. 26, 2022), [Online]. Available: <https://towardsdatascience.com/batch-normalization-in-3-levels-of-understanding-14c2da90a338> (visited on 12/01/2022).
- [23] H. Yadav. “Dropout in neural networks”, Medium. (Jul. 5, 2022), [Online]. Available: <https://towardsdatascience.com/dropout-in-neural-networks-47a162d621d9> (visited on 12/01/2022).
- [24] S. Ruder, *An overview of gradient descent optimization algorithms*, Jun. 15, 2017.

- [25] “Tutorial for optimization toolbox - MATLAB & simulink - MathWorks deutschland”. (), [Online]. Available: <https://de.mathworks.com/help/optim/ug/optimization-toolbox-tutorial.html> (visited on 12/13/2022).
- [26] H. Robbins and S. Monro, “A stochastic approximation method”, *The Annals of Mathematical Statistics*, vol. 22, no. 3, pp. 400–407, Sep. 1951.
- [27] P. Werbos, “Backpropagation through time: What it does and how to do it”, *Proceedings of the IEEE*, vol. 78, no. 10, pp. 1550–1560, Oct. 1990, Conference Name: Proceedings of the IEEE.
- [28] “Backpropagation | brilliant math & science wiki”. (), [Online]. Available: <https://brilliant.org/wiki/backpropagation/> (visited on 01/27/2023).
- [29] *Sørensen-dice coefficient*, in *Wikipedia*, Page Version ID: 1138494124, Feb. 10, 2023.
- [30] Chopin. “Neural network probability output and loss function (example: Dice loss)”, Data Science Stack Exchange. (Jun. 9, 2020), [Online]. Available: <https://datascience.stackexchange.com/q/75708> (visited on 12/13/2022).
- [31] V. S. Subramanyam. “IOU(intersection over union)”, Analytics Vidhya. (Jan. 28, 2021), [Online]. Available: <https://medium.com/analytics-vidhya/iou-intersection-over-union-705a39e7acef> (visited on 02/09/2023).
- [32] N. Otsu, “A threshold selection method from gray-level histograms”, *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, Jan. 1979, Conference Name: IEEE Transactions on Systems, Man, and Cybernetics.
- [33] *Otsu’s method*, in *Wikipedia*, Page Version ID: 1120318148, Nov. 6, 2022.
- [34] “Illustration of different FCN architectures. a illustration of FCN-32...”, ResearchGate. (), [Online]. Available: https://www.researchgate.net/figure/Illustration-of-different-FCN-architectures-a-Illustration-of-FCN-32-architecture-b_fig2_325048059 (visited on 12/19/2022).
- [35] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, *DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs*, May 11, 2017.
- [36] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, *Rethinking atrous convolution for semantic image segmentation*, Dec. 5, 2017.

- [37] H. Noh, S. Hong, and B. Han, “Learning deconvolution network for semantic segmentation”, in *2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile: IEEE, Dec. 2015, pp. 1520–1528.
- [38] V. Badrinarayanan, A. Kendall, and R. Cipolla, *SegNet: A deep convolutional encoder-decoder architecture for image segmentation*, Oct. 10, 2016.
- [39] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, *Squeeze-and-excitation networks*, May 16, 2019.
- [40] O. Oktay, J. Schlemper, L. L. Folgoc, *et al.*, *Attention u-net: Learning where to look for the pancreas*, May 20, 2018.
- [41] R. Alam. “Normalization vs standardization explained”, Medium. (May 9, 2020), [Online]. Available: <https://towardsdatascience.com/normalization-vs-standardization-explained-209e84d0f81e> (visited on 02/01/2023).
- [42] *Affine transformation*, in *Wikipedia*, Page Version ID: 1125144549, Dec. 2, 2022.
- [43] *Contrast (vision)*, in *Wikipedia*, Page Version ID: 1136823048, Feb. 1, 2023.
- [44] *Gaussian blur*, in *Wikipedia*, Page Version ID: 1132518555, Jan. 9, 2023.
- [45] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, Jan. 29, 2017.
- [46] J. Duchi, E. Hazan, and Y. Singer, “Adaptive subgradient methods for online learning and stochastic optimization”,
- [47] “RMSProp - cornell university computational optimization open textbook - optimization wiki”. (), [Online]. Available: <https://optimization.cbe.cornell.edu/index.php?title=RMSProp> (visited on 02/02/2023).
- [48] *Precision and recall*, in *Wikipedia*, Page Version ID: 1122267443, Nov. 16, 2022.