



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

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Σημασιολογική κατάτμηση
μεταστατικού μελανώματος σε
δεδομένα PET/CT

Σπουδαστής:
Ευάγγελος Μάρκος

Επιβλέπων:
Γεώργιος Ματσόπουλος,
Καθηγητής ΕΜΠ

Αθήνα, Σεπτέμβριος 2023



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

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Σημασιολογική κατάτμηση
μεταστατικού μελανώματος σε
δεδομένα PET/CT

Σπουδαστής:
Ευάγγελος Μάρκος

Επιβλέπων:
Γεώργιος Ματσόπουλος,
Καθηγητής ΕΜΠ

Εγκρίθηκε από την κάτωθι τριμελή επιτροπή την 15^η Σεπτεμβρίου 2023.

Γεώργιος Ματσόπουλος
Καθηγητής ΕΜΠ

Παναγιώτης Τσανάκας
Καθηγητής ΕΜΠ

Αθανάσιος Δ.
Παναγόπουλος
Καθηγητής ΕΜΠ

Αθήνα, Σεπτέμβριος 2023

.....
Ευάγγελος Θ. Μάρκος
Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright © Ευάγγελος Μάρκος, 2023
Με 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. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

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.

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

Περίληψη

Τομέας Συστημάτων Μετάδοσης Πληροφορίας και Τεχνολογίας Υλικών
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Σημσιολογική κατάτμηση μεταστατικού μελανώματος σε δεδομένα PET/CT

Ευάγγελος Μάρκος

Η σημσιολογική κατάτμηση αποτελεί έναν κλάδο της επιστήμης της όρασης των υπολογιστών με σκοπό την αυτόματη εξαγωγή κρίσιμων πληροφοριών από εικόνες. Σε αντιδιαστολή με την κατηγοριοποίηση των εικόνων, η σημσιολογική κατάτμηση κατηγοριοποιεί κάθε εικονοστοιχείο της εικόνας, προσφέροντας με αυτόν τον τρόπο πρωτοφανείς δυνατότητες αυτόματης χαρτογράφησης σε αυτές. Οι εφαρμογές εμφανίζονται σε πολλούς κλάδους και ιδιαίτερα σε αυτόν της ιατρικής και συγκεκριμένα της ιατρικής απεικόνισης. Η εξέλιξη των ιατρικών απεικονιστικών συστημάτων τα τελευταία χρόνια έχει δημιουργήσει τεχνικές οι οποίες αποτελούν πλέον αναπόσπαστο κομμάτι της διάγνωσης και της παρακολούθησης πολλών ασθενειών αλλά και την ανάγκη παρακολούθησης μεγάλου όγκου πληροφορίας που δημιουργείται από αυτές. Οι πρόσφατες εξελίξεις της τεχνητής νοημοσύνης και συγκεκριμένα της βαθιάς μάθησης, προσφέρουν εργαλεία που μπορεί να αποβούν χρήσιμα στην ανάγκη αυτή. Η σημσιολογική κατάτμηση αποτελεί ένα τέτοιο παράδειγμα και μπορεί να εφαρμοστεί σε ιατρικές εικόνες για την αυτόματη τμηματοποίηση κρίσιμων σημείων τους. Σε αυτήν την εργασία θα γίνει μια επίδειξη αυτής της τεχνικής, όπου θα εφαρμοστούν αρχιτεκτονικές και μεθοδολογίες νευρωνικών δικτύων όπως το U-net, η "προσοχή" και το ensemble learning, για να επιτευχθεί αυτόματη κατάτμηση καρκινικών περιοχών από μεταστατικό μελάνωμα σε εικόνες PET/CT.

Λέξεις κλειδιά: U-net, Σημσιολογική κατάτμηση, Βαθιά μάθηση, PET/CT, Μεταστατικό Μελάνωμα

NATIONAL TECHNICAL UNIVERSITY OF ATHENS

Abstract

Division of Information Transmission Systems and Material Technology
School of Electrical and Computer Engineering

Semantic segmentation of metastatic melanoma in PET/CT images

Evangelos Markos

Semantic segmentation is a field of computer vision aiming at the automated extraction of crucial information from images. In contrast to the classification of the images, semantic segmentation classifies each pixel of an image, hence offering unique prospects for their analysis. The applications can be found in many fields, especially on the one of medical image analysis. The development of medical imaging systems has resulted in techniques that are now inseparable from the diagnosis and monitoring of many diseases, but also in the need for the analysis of the great number of data that come out of them. The recent advent of artificial intelligence, and specifically deep learning, pose this field a good candidate in the aid of that need. Semantic segmentation, which stems from deep learning, can be implemented in medical images for the automatic segmentation of crucial parts of them. This work is a display of this technique, making use of different specialized neural network architectures and methodologies such as U-net, "attention" and ensemble learning, to automatically segment cancer regions of metastatic melanoma in PET/CT images.

Keywords: U-net, Semantic Segmentation, Deep Learning, PET/CT, Metastatic Melanoma

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον κ. Γεώργιο Ματσόπουλο, καθηγητή και επιβλέπων της παρούσας εργασίας για την άψογη συνεργασία και την εμπιστοσύνη που μου έδειξε έτσι ώστε να μπορώ να αναλάβω αυτήν την εργασία. Θα ήθελα επίσης να ευχαριστήσω και τους κυρίους Παναγιώτη Τσανάκα και Αθανάσιο Δ. Παναγόπουλο, οι οποίοι σαν μέλη της υπόλοιπης τριμελούς επιτροπής συνέβαλαν επίσης στο πέρας αυτού του έργου. Δεν θα έπρεπε να παραλείψω τον υποψήφιο διδάκτορα Θεόδωρο Π. Βαγενά, ο οποίος ενεπλάκη πολύ ενεργά σε αυτήν την εργασία, δίνοντας κατευθύνσεις, ιδέες και προτάσεις για την επίλυση πολλών προβλημάτων που προέκυψαν, πάντα με υπομονή και ευγένεια. Τέλος, θα ήθελα να ευχαριστήσω τους κοντινούς μου ανθρώπους οι οποίοι μπορεί να μην συνέβαλαν άμεσα στην εκπόνηση αυτής της εργασίας ή των σπουδών μου γενικότερα, αλλά έμμεσα βοήθησαν πολύ, αποτελώντας πυλώνες στήριξης και έμπνευσης για μένα.

Contents

Περίληψη	iii
Abstract	v
Ευχαριστίες	vii
1 Εισαγωγή	1
2 Εικόνες PET/CT και μεταστατικό μελάνωμα	3
2.1 Τομογραφία εκπομπής ποζιτρονίου - PET	3
2.1.1 Φυσικές αρχές	3
2.1.2 Διάταξη PET	4
2.1.3 Ανακατασκευή εικόνας	5
2.2 Υπολογιστική αξονική τομογραφία - CT	6
2.2.1 Φυσικές αρχές	7
2.2.2 Διάταξη CT	7
2.2.3 Ανακατασκευή εικόνας	10
2.3 Συνδυασμός εικόνων PET και CT: PET/CT	12
2.4 Εισαγωγή στο μεταστατικό μελάνωμα	13
2.5 Εντοπισμός μεταστατικού μελανώματος με χρήση PET/CT εικόνων	18
3 Νευρωνικά Δίκτυα και Βαθιά μάθηση	21
3.1 Βασικές έννοιες στα Νευρωνικά Δίκτυα	21
3.1.1 Εισαγωγή	21
3.1.2 Κεντρική ιδέα	22
3.1.3 Πρόβλεψη Νευρωνικών Δικτύων	23
3.1.4 Εκπαίδευση Νευρωνικών Δικτύων	24
3.1.5 Συναρτήσεις Κόστους	26
Mean Square Error - MSE	27
Mean Absolute Error - MAE	27
Cross Entropy	27
3.1.6 Συναρτήσεις Ενεργοποίησης	28
Sigmoid	28
Tanh	28
Softmax	28
Rectified Linear Unit - Relu	29
3.1.7 Βελτιστοποιητές	29
Gradient Descent	29
Stochastic Gradient Descent	30
Βελτιστοποιητές με χρήση τεχνικών Momentum	31
Adaptive Βελτιστοποιητές	32
Adam	33
3.2 Νευρωνικά Δίκτυα και Βαθιά Μάθηση με Εικόνες	35

3.2.1	Εισαγωγή	35
3.2.2	Κεντρική ιδέα	36
3.2.3	Εξαγωγή χαρακτηριστικών με χρήση Συνελικτικών Νευρωνικών Δικτύων	36
	Συνελικτικά επίπεδα	37
	Downsampling	38
3.2.4	Παράδειγμα: Lenet-5	39
3.2.5	Παράδειγμα: AlexNet	40
3.3	Σημασιολογική κατάτμηση με Νευρωνικά Δίκτυα	41
3.3.1	Εισαγωγή	41
3.3.2	Κεντρική ιδέα	42
3.3.3	Fully Convolutional Networks - FCN	42
3.3.4	Ανακατασκευή εικόνας	43
	Transposed Convolution	43
	Nearest-neighbor Interpolation	44
	Bilinear Interpolation	44
	Max Unpooling	45
3.3.5	Παράδειγμα: FCN	46
3.3.6	Παράδειγμα: SegNet	48
3.3.7	Συναρτήσεις Κόστους - Μετρικές για εφαρμογές Σημασιολογικής Κατάτμησης	48
	F1 score - Dice Coefficient	49
	Intersection over Union - IoU (Jaccard Index)	50
	Tversky Index	51
	Focal Tversky Loss	51
3.3.8	Σημασιολογική κατάτμηση Ιατρικών Εικόνων με χρήση του U-Net	51
	U-net	51
	Attention U-net	54
	Residual Attention U-net	56
4	Υλοποίηση	59
4.1	Dataset	59
4.2	Προεπεξεργασία εικόνων	60
4.3	U-net, Attention U-net, Residual Attention U-net: Βασική υλοποίηση	62
4.4	Προτεινόμενη μέθοδος	67
4.4.1	Attention U-net: Εύρεση όλων των όγκων	68
4.4.2	Residual Attention U-net: Μείωση των False Positives	69
4.4.3	Ensemble Learning: Συνδυασμός αποτελεσμάτων	71
4.5	Σχολιασμός αποτελεσμάτων	75
4.6	Προτάσεις για βελτίωση	78
	Bibliography	83

List of Figures

2.1	PET structure	4
2.2	PET scintillator and photosensor	5
2.3	Coincide Events	5
2.4	All Events	6
2.5	CT Tomography Parts	8
2.6	CT Generations	9
2.7	5th Generation of CT	10
2.8	Fourier Slice Theorem	11
2.9	PET/CT image example	13
2.10	PET/CT Stages of metastatic melanoma	14
2.11	PET/CT metastatic melanoma stage 0	14
2.12	PET/CT metastatic melanoma stage IA	15
2.13	PET/CT metastatic melanoma stage IB	15
2.14	PET/CT metastatic melanoma stage IIA	16
2.15	PET/CT metastatic melanoma stage IIB	16
2.16	PET/CT metastatic melanoma stage IIC	16
2.17	PET/CT metastatic melanoma stage III	17
2.18	PET/CT metastatic melanoma stage IV	17
2.19	PET bright spots	18
2.20	PET missing tumors in certain regions	19
2.21	PET false positives	19
2.22	PET false negatives	20
3.1	Neural Network Layers Structure	22
3.2	Neural Network Neuron Working	23
3.3	Learning rate different values	26
3.4	Gradient Descent critical points	30
3.5	Stochastic Gradient Descent	31
3.6	Stochastic Gradient Descent with momentum	32
3.7	AdaGrad convergence	33
3.8	Adam optimizer convergence	34
3.9	Basic CNN architecture	36
3.10	Convolution Example	37
3.11	Edge detection convolution	38
3.12	Lenet-5 architecture	39
3.13	MNIST dataset sample	39
3.14	AlexNet architecture	40
3.15	Fully Convolutional Networks architecture	43
3.16	Nearest-neighbor Interpolation	44
3.17	Bilinear Interpolation	45
3.18	Max Unpooling	46
3.19	FCN architecture	46
3.20	Upsampling in FCN	47

3.21	Compare Upsampling in FCN	47
3.22	Segnet architecture	48
3.23	Dice Coefficient as area sets calculation	50
3.24	IoU as area sets calculation	50
3.25	U-net architecture	52
3.26	U-net architecture in detail	53
3.27	U-net overlap-tile strategy	54
3.28	Attention U-net architecture	55
3.29	Attention gate architecture	56
3.30	Residual connections	57
3.31	U-net with residual connections	58
4.1	MICCAI Dataset Sample	59
4.2	Input data number frequency of cancer cells	60
4.3	Augmentation transforms	61
4.4	U-net boxplot results	62
4.5	U-net barplot results	63
4.6	Attention U-net boxplot results	64
4.7	Attention U-net barplot results	64
4.8	Residual Attention U-net boxplot results	65
4.9	Residual Attention U-net barplot results	66
4.10	Attention U-net all cancers barplot results	68
4.11	Attention U-net all cancers example images	69
4.12	Residual Attention U-net reduce false positives barplot results	70
4.13	Residual Attention U-net reduce false positives example images	71
4.14	Ensemble U-net boxplot results for all images	72
4.15	Ensemble U-net barplot results for all images	73
4.16	Ensemble U-net boxplot results for final set of images	73
4.17	Ensemble U-net example images	74
4.18	Ensemble U-net example images, compare image 1	75
4.19	Ensemble U-net example images, compare image 2	76
4.20	Ensemble U-net example images, false positives in organs	77
4.21	U-net++ nested skip connections	79
4.22	nnU-net framework	79
4.23	Expanded semantic classes	81

List of Tables

2.1	Ισότοποι ραδιενεργοί πυρήνες στην PET τομογραφία	4
3.1	Χαρακτηριστικά του Lenet-5	40
3.2	Χαρακτηριστικά του AlexNet	41
4.1	Αποτελέσματα των υλοποιήσεων	74

Κεφάλαιο 1

Εισαγωγή

Η τεχνητή νοημοσύνη τα τελευταία χρόνια με την ραγδαία εξέλιξή της έχει εισχωρήσει σε πολλές πτυχές της καθημερινότητάς μας. Η εφαρμογή της σε ιατρικές εφαρμογές δεν αποτελεί εξαίρεση. Ενώ η τεχνητή νοημοσύνη θεωρείται ότι είναι ακόμα στα αρχικά στάδια της ανάπτυξής της, έχει εισχωρήσει και ήδη συμβάλει σε έναν πολύ μεγάλο αριθμό εφαρμογών στην ιατρική, από τη διάγνωση μέχρι και την έρευνα. Η ικανότητά των συστημάτων τεχνητής νοημοσύνης να επεξεργαστούν μεγάλο όγκο δεδομένων και να εξάγουν συμπεράσματα γίνεται ιδιαίτερα χρήσιμη ειδικότερα στην εποχή μας όπου το διαδίκτυο, τα υπολογιστικά συστήματα και η πρόοδος των ιατρικών συστημάτων παρακολούθησης των ασθενών συμβάλλουν στην συσσώρευση τεράστιου όγκου δεδομένων, ο οποίος αυξάνεται με ρυθμούς πολύ μεγαλύτερους από αυτούς που θα μπορούσε ο άνθρωπος να τον παρακολουθήσει. Αυτή τη στιγμή, συστήματα τεχνητής νοημοσύνης είναι ικανά να παρακολουθήσουν ιατρικά δεδομένα και να προβλέψουν την εμφάνιση ή την εξέλιξη διαφόρων ασθενειών. Αξιοποιούνται σε εφαρμογές προσωποποιημένης παρακολούθησης καθημερινής χρήσης αλλά και σε εφαρμογές υποβοήθησης των γιατρών στην κατάστρωση πλάνων θεραπειών.

Τα συστήματα τεχνητής νοημοσύνης όμως είναι χρήσιμα και σε εφαρμογές προσομοίωσης, προσφέροντας πλέον δυνατότητες που δεν υπήρχαν παλιά σε αυτές τις εφαρμογές. Έτσι μπορούν να γίνουν ιδιαίτερα χρήσιμα στην έρευνα, υποβοηθώντας μεταξύ άλλων στην δημιουργία νέων φαρμάκων αλλά και σε διάφορες βιολογικές έρευνες. Παραθέτουμε το χαρακτηριστικό παράδειγμα του AlphaFold "*Highly accurate protein structure prediction with AlphaFold*" (Jumper et al., 2021), ενός συστήματος τεχνητής νοημοσύνης και συγκεκριμένα βαθιάς μάθησης, το οποίο κατάφερε να προσομοιώσει (προβλέψει) μοντέλα τριών διαστάσεων πρωτεϊνικών δομών, ένα από τα κεντρικά προβλήματα της βιολογίας το οποίο είχε παραμείνει άλυτο επί πενήντα χρόνια. Ένας κλάδος όπου η τεχνητή νοημοσύνη έχει φανεί ιδιαίτερα χρήσιμη τα τελευταία χρόνια είναι αυτός της αυτόματης διάγνωσης ασθενειών από ιατρικές εικόνες. Συστήματα τεχνητής νοημοσύνης μπορούν να αναγνωρίσουν ασθένειες, να βοηθήσουν ειδικούς στη διάγνωση και τον εντοπισμό ανωμαλιών (Computer Aided Detection and Diagnosis-CAD) και να αναλύσουν ένα εύρος ιατρικών εικόνων. Όλα αυτά μπορούν να τα επιτύχουν γιατί κατέχουν δυνατότητες εξαγωγής μοτίβων και πληροφορίας από εικόνες και είναι ικανά κατηγοριοποίησης και τμηματοποίησης αυτών. Ως τμηματοποίηση ή κατάτμηση ονομάζεται ο διαχωρισμός μιας εικόνας σε υποσύνολα τα οποία παρουσιάζουν κάποια κοινά χαρακτηριστικά. Αυτά τα χαρακτηριστικά μπορούν να είναι πολλά, όπως η φωτεινότητα, η αντίθεση, το χρώμα, αλλά και πιο γενικές και αφηρημένες έννοιες όπως τα όργανα σε μια ιατρική εικόνα ή οι καρκινικές περιοχές. Συγκεκριμένα με την αυτόματη κατάτμηση περιοχών μεταστατικού μελανώματος σε ιατρικές εικόνες PET/CT με χρήση τεχνικών μηχανικής μάθησης θα καταπιαστούμε σε αυτήν την διπλωματική εργασία.

Η διάρθρωση της εργασίας ακολουθεί τους εμπλεκόμενους κλάδους για το πέρας της υλοποίησης. Στην αρχή γίνεται μια παρουσίαση των αρχών των μεθόδων ιατρικής απεικόνισης των PET, CT και PET/CT εικόνων καθώς και των χαρακτηριστικών

του μεταστατικού μελανώματος, όπως και των εφαρμογών των μεθόδων αυτών στην παρακολούθηση του μεταστατικού μελανώματος. Στην συνέχεια παρουσιάζονται οι αρχές της βαθιάς μάθησης με έμφαση σε εφαρμογές της όρασης υπολογιστών και ειδικότερα της σημασιολογικής κατάτμησης ιατρικών εικόνων. Τέλος παρουσιάζουμε την υλοποίησή μας, τα διάφορα πειράματα, την μέθοδο που ακολουθήσαμε, τα αποτελέσματα που επιτύχαμε όπως και τις προτάσεις μας για μελλοντικές υλοποιήσεις οι οποίες θα μπορούσαν να επιφέρουν βελτίωση στα αποτελέσματα.

Κεφάλαιο 2

Εικόνες PET/CT και μεταστατικό μελάνωμα

2.1 Τομογραφία εκπομπής ποζιτρονίου - PET

Η τομογραφία εκπομπής ποζιτρονίων (Positron Emission Tomography-PET) αποτελεί μια μέθοδο ιατρικής απεικόνισης για την αναπαράσταση διεργασιών φυσιολογίας του ανθρώπινου οργανισμού. Συγκεκριμένα παρακολουθεί, μεταξύ άλλων, λειτουργικές (μεταβολικές) δραστηριότητες, με αποτέλεσμα να επιτρέπει την έγκαιρη διάγνωση παθολογικών καταστάσεων που εμφανίζουν διαταραχές περισσότερο σε λειτουργικό απ' ό τι σε ανατομικό επίπεδο. Στις εικόνες της τομογραφίας, οι περιοχές ενδιαφέροντος (περιοχές μεγαλύτερης μεταβολικής δραστηριότητας) απεικονίζονται εντονότερα από άλλες, δίνοντας έτσι την δυνατότητα εποπτείας διαφόρων διεργασιών όπου άλλοτε μπορεί να είναι αναμενόμενες ή μη. Αποτελεί χρήσιμη μέθοδο για τον εντοπισμό και την παρακολούθηση πλήθους παθήσεων, όπως νευρωλογικών ή καρδιακών αλλά ακόμα και του καρκίνου. Η διαδικασία της τομογραφίας στηρίζεται στην έγχυση ιχνηθέτη ο οποίος απορροφάται από τα σημεία ενδιαφέροντος και παράγει ακτινοβολία η οποία ανιχνεύεται έτσι ώστε να ανακατασκευαστεί η εικόνα υπολογιστικά. Σε αυτό το σημείο σημειώνουμε ότι παραθέτουμε υλικό όπως θεωρία, πίνακες και μαθηματικούς τύπους από έναν από τους κεντρικούς οδηγούς μας για αυτό το κεφάλαιο, το σύγγραμμα "Ιατρικά Απεικονιστικά Συστήματα" των Δ. Κουσούρη, Κ.Νικήτα και Σ. Παυλόπουλου.

2.1.1 Φυσικές αρχές

Η τομογραφία εκπομπής ποζιτρονίου στον πυρήνα της αξιοποιεί το φαινόμενο της εξαύλωσης για την χαρτογράφηση διαφόρων περιοχών του ανθρώπινου οργανισμού. Συγκεκριμένα αξιοποιείται η εξαύλωση μεταξύ ηλεκτρονίων και ποζιτρονίων, τα οποία όταν βρεθούν σε μια κοντινή απόσταση χάνουν τη μάζα τους και εκπέμπουν δύο ακτίνες γ ενέργειας 511 keV σε αντιδιαμετρικές κατευθύνσεις. Για να προκληθεί αυτό το φαινόμενο παρασκευάζεται ένα ραδιοφάρμακο το οποίο εμπεριέχει ισότοπους ραδιενεργούς πυρήνες (βλέπε πίνακα 2.1). Αυτό το φάρμακο εγχύεται στις περιοχές που θέλουμε να αναπαράστησουμε και στη συνέχεια τα ποζιτρόνια που εμπεριέχονται σε αυτό εξαυλώνονται με τα ηλεκτρόνια που βρίσκονται στον οργανισμό του ασθενή. Οι ακτίνες γ ανιχνεύονται από μια ανιχνευτική διάταξη και στη συνέχεια ανακατασκευάζεται η εικόνα, η οποία μπορεί να είναι και τριών διαστάσεων.

Η εικόνα που παράγεται από την τομογραφία δίνει πληροφορίες για τον τρόπο με τον οποίο μεταβολίζεται ο ιχνηθέτης ο οποίος μπορεί να δώσει πληροφορίες για διάφορες φυσιολογικές και βιοχημικές διαδικασίες.

TABLE 2.1: Ισότοποι ραδιενεργοί πυρήνες στην PET τομογραφία

Ραδιοϊσότοπο	Χρόνος Ημιζωής	Εφαρμογές
Carbon-11	20.3 min	Μελέτη μεταβολικών διαδικασιών
Fluorine-18	109.7 min	Μελέτη κατανάλωσης γλυκόζης Απεικόνιση νευροϋποδοχέων
Gallium-68	68 min	Βαθμονόμηση
Iodine-122	3.76 min	Μελέτη αιματικής ροής
Iron-52	8.2 hr	Απεικόνιση μυελού των οστών
Nitrogen-13	9.9 min	Μελέτη έκχυσης μυοκαρδίου
Oxygen-15	123 sec	Μελέτη αιματικής ροής
Rubidium-82	1.2 min	Μελέτη έκχυσης μυοκαρδίου

2.1.2 Διάταξη PET

Η διάταξη του PET, υπεύθυνη για την ανίχνευση των παραγόμενων ακτίνων γ που παράγονται, αποτελείται από κυρίως δύο συστατικά: τον κρυσταλλικό σπινθηριστή και τον φωτοπολλαπλασιαστή. Πιο συγκεκριμένα από κυκλικές αλυσίδες κρυσταλλικών σπινθηριστών και φωτοπολλαπλασιαστών οι οποίοι περικλείουν τον ασθενή έτσι ώστε να μπορούν να ανιχνεύσουν αντιδιαμετρικά ζεύγη ακτίνων γ σε όλες τις κατευθύνσεις. Αυτή η δακτυλική διάταξη διασχίζει όλο το μήκος του ασθενή έτσι ώστε να επιτρέψει την παραγωγή εικόνων ολόκληρου του σώματος.

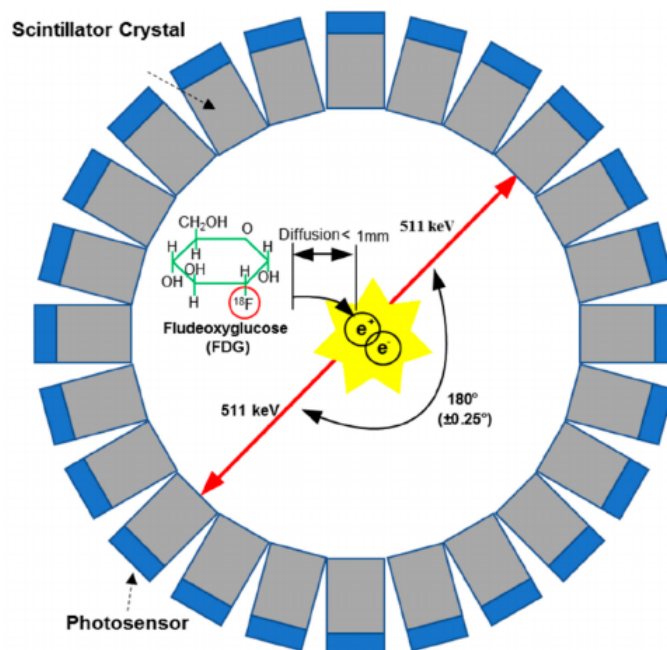


FIGURE 2.1: Η Δομή του PET αποτελείται από κυρίως δύο μέρη: τον κρυσταλλικό σπινθηριστή και τον φωτοπολλαπλασιαστή

Ο σπινθηριστής μετατρέπει τα υψηλής ενέργειας φωτόνια σε φωτόνια της ορατής περιοχής του ηλεκτρομαγνητικού φάσματος ενώ ο φωτοπολλαπλασιαστής παράγει έναν ενισχυμένο παλμό ρεύματος ανάλογο του ποσού των φωτονίων που προσπίπτουν στην

κάθοδο. Ο σπινθηριστής και ο φωτοπολλαπλασιαστής μαζί είναι υπεύθυνοι για την ανίχνευση και αποθήκευση τριών πληροφοριών: τη θέση που συλλέγεται το γ φωτόνιο, τον χρόνο συλλογής του και την ενέργεια που φέρει.

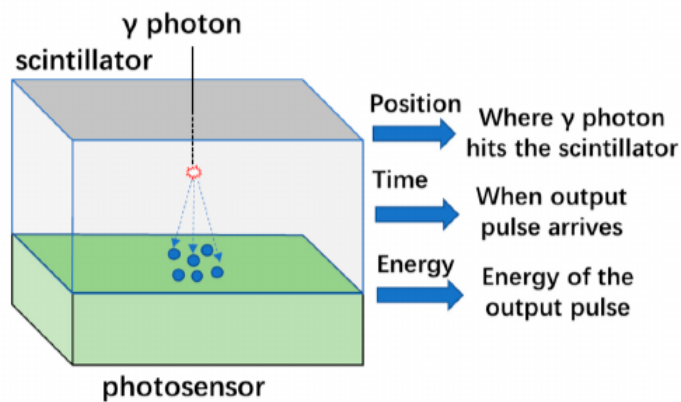


FIGURE 2.2: Η λειτουργία του κρυσταλλικού σπινθηριστή και του φωτοπολλαπλασιαστή

2.1.3 Ανακατασκευή εικόνας

Όπως προαναφέρθηκε ο κάθε ανιχνευτής θα λαμβάνει σήματα, τα οποία μπορούν να χρησιμοποιηθούν για την ανακατασκευή της εικόνας. Για την χαρτογράφηση της εικόνας από τα σήματα αυτά είναι απαραίτητο να προσδιοριστεί η χρονική και χωρική προέλευση της πηγής τους. Δεδομένου του γεγονότος ότι μετά το φαινόμενο της εξαύλωσης οι ακτίνες γ ταξιδεύουν σε αντιδιαμετρικές κατευθύνσεις, αντιδιαμετρικοί ανιχνευτές αναμένεται να εντοπίσουν σε κοντινές χρονικές στιγμές τα αντίστοιχα σήματα. Η Ευθεία πάνω στην οποία είναι συνευθειακά το προς αναζήτηση σημεία εξαύλωσης και οι δύο ανιχνευτές ονομάζεται line of response (LoR) και η καταγραφή σημάτων στις ίδιες χρονικές στιγμές από αντιδιαμετρικούς ανιχνευτές coincidence events.

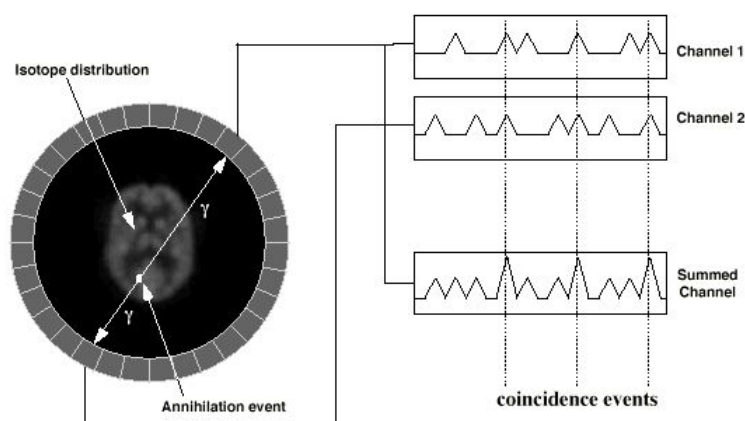


FIGURE 2.3: Τα φαινόμενα σύμπτωσης (Coincidence events) στην PET τομογραφία

Η συλλογή όλων των coincide events για όλα τα αντιδιαμετρικά ζεύγη ανιχνευτών αποτελούν αρκετά δεδομένα για την ανακατασκευή της εικόνας. Οι μέθοδοι που χρησιμοποιούνται για την ανακατασκευή της εικόνας βρίσκουν χρήση και σε άλλες τομογραφίες (βλέπε CT). Κάποιες από αυτές είναι η απλή οπισθοπροβολή (simple backprojection), η φιλτραρισμένη οπισθοπροβολή (filtered backprojection) αλλά και οι επαναληπτικοί αλγόριθμοι ανακατασκευής.

Η υπόθεση βέβαια ότι αντιδιαμετρικοί ανιχνευτές σε παρόμοιες χρονικές στιγμές ανίχνευσης σήματος ανιχνεύουν πάντα το σημείο εξαύλωσης δεν έχει πάντα ισχύ. Αρχικά τα σήματα σε παραπλήσιες χρονικές στιγμές αντιδιαμετρικών ανιχνευτών μπορεί να οφείλονται σε διαφορετικά σημεία εξαύλωσης που έτυχε να βρίσκονται στο σωστό σημείο την σωστή στιγμή ώστε να δώσουν το καθένα ένα σήμα που θα ανιχνευθεί από τον κάθε ανιχνευτή (στην ίδια LoR) σε παραπλήσια χρονική στιγμή (random coincidence events). Τα random coincidence events θα οδηγήσουν σε λανθασμένες υποθέσεις σημείων εξαύλωσης και δημιουργούν θόρυβο στην εικόνα. Υπάρχει όμως και το ενδεχόμενο κάποια σημεία εξαύλωσης να μην βρεθούν καθώς μπορεί οι κατευθύνσεις των ακτίνων γ μετά από μια εξαύλωση να αλλάξουν λόγω φαινομένων σκέδασης (scattered coincidence). Μέθοδοι όπως το attenuation correction μπορούν να εφαρμοστούν για την βελτίωση της ποιότητας της εικόνας σε αυτές τις περιπτώσεις.

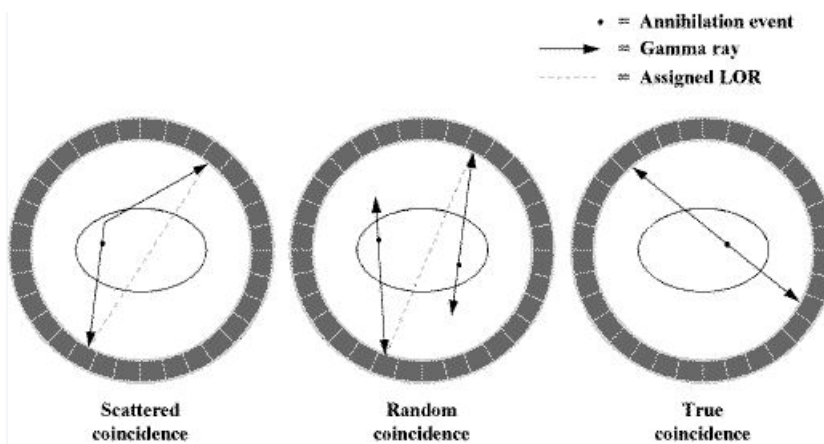


FIGURE 2.4: Όλα τα φαινόμενα σύμπτωσης στην PET τομογραφία

2.2 Υπολογιστική αξονική τομογραφία - CT

Η αξονική τομογραφία ή υπολογιστική τομογραφία (Computed Tomography - CT) αποτελεί μια από τις πιο διαδεδομένες μεθόδους ιατρικής απεικόνισης. Μπορεί να δώσει υψηλής ποιότητας εικόνες ανατομικών δομών του ανθρώπινου σώματος σε τρεις διαστάσεις. Η κεντρική ιδέα της αξονικής τομογραφίας είναι η υπολογιστική ανακατασκευή (υπολογιστική τομογραφία) της εικόνας συνδυάζοντας λήψεις ακτίνων X από πολλαπλές γωνίες. Ο ασθενής δέχεται αυτές τις ακτίνες X και το ποσοστό εξασθένισης τους μπορεί να καταδείξει το είδος των ιστών του, καθώς διαφορετικά είδη ιστών παρουσιάζουν διαφορετικές τιμές απορρόφησης των ακτίνων. Ο συνυπολογισμός από πολλαπλές γωνίες αυτών των εξασθενίσεων είναι δυνατό να ανακατασκευάσει την εικόνα των ιστών του ασθενή.

2.2.1 Φυσικές αρχές

Η φυσική αρχή που κρύβεται πίσω από τη λειτουργία της υπολογιστικής τομογραφίας είναι η απορρόφηση των ακτίνων X κατά τη διέλευσή τους από ένα αντικείμενο. Για ένα ομογενές αντικείμενο που χαρακτηρίζεται από ένα σταθερό συντελεστή απορρόφησης μ και μήκος l , αν I_d είναι η εξερχόμενη ένταση φωτονίων ανά μονάδα χρόνου και I_o η εισερχόμενη, θα ισχύει ότι:

$$I_d = I_o e^{-\mu l} \quad (2.1)$$

Για αντικείμενα τα οποία δεν είναι ομογενή (παράδειγμα ανθρώπινος οργανισμός) ο συντελεστής απορρόφησης σε κάθε σημείο τους θα είναι διαφορετικός. Συνεπώς μπορεί να περιγραφεί σαν μια συνάρτηση χωρική, όπου σε κάθε σημείο του χώρου αντιστοιχίζεται ένα αντικείμενο. Για πρόσπτωση ακτίνων X σε μια διδιάστατη τομή θα είναι $\mu(x,y)$. Τότε η ανωτέρα σχέση γίνεται:

$$\int_L \mu(x,y) ds = \ln \frac{I_o}{I_d} \quad (2.2)$$

Όπως καταλαβαίνουμε, για γνωστές εντάσεις φωτονίων ανά μονάδα χρόνου I_o και I_d μπορεί να γίνει προσπάθεια να προσδιοριστεί ο συντελεστής απορρόφησης ενός αντικειμένου σε κάθε σημείο του. Στην περίπτωση της υπολογιστικής τομογραφίας και του ανθρώπου ασθενή, οι διάφοροι ιστοί του χαρακτηρίζονται από διαφορετικές τιμές στο χώρο του συντελεστή απορρόφησης μ , καθώς γνωρίζουμε ότι απορροφούν διαφορετικά τις ακτίνες X, και ο προσδιορισμός τους μπορεί να οδηγήσει στον προσδιορισμό των ιστών. Οι τιμές I_o και I_d θα οδηγήσουν στον υπολογισμό του μ αφού είναι γνωστές. Γνωρίζουμε με πόση ακτινοβολία τροφοδοτούμε τον ασθενή (I_o) και μετράμε πόση ακτινοβολία εξέρχεται από αυτόν (I_d). Σημειώνουμε σε αυτό το σημείο πως για να καταστεί δυνατός ο προσδιορισμός του μ είναι απαραίτητη η λήψη από πολλές γωνίες μετρήσεων και στην υπολογιστική τομογραφία δημιουργούνται διατάξεις που κάνουν αυτό. Με την λήψη των δεδομένων αυτών στη συνέχεια γίνεται η ανακατασκευή της εικόνας με διάφορους υπολογιστικούς αλγόριθμους που υπάρχουν.

2.2.2 Διάταξη CT

Τα κύρια μέρη ενός συστήματος αξονικής τομογραφίας αποτελούν την πηγή ακτίνων X, τους ανιχνευτές ακτίνων X, το σύστημα απόκτησης δεδομένων, την κονσόλα ελέγχου, τον ηλεκτρονικό υπολογιστή και την κάμερα. Η πηγή των ακτίνων, οι ανιχνευτές και το σύστημα απόκτησης δεδομένων ονομάζεται gantry.

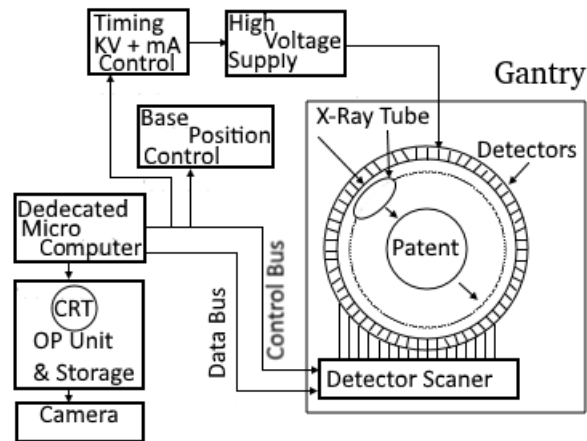


FIGURE 2.5: Τα κύρια μέρη της υπολογιστικής αξονικής τομογραφίας

Η λειτουργία της διάταξης είναι σχετικά απλή. Αφού ο ασθενής εισαχθεί στο gantry, η πηγή ακτίνων X ξεκινάει την παραγωγή των ακτίνων. Αυτές διαπερνούν τον ασθενή και συλλέγονται από τους ανιχνευτές. Στη συνέχεια το σύστημα απόκτησης δεδομένων αφού εκτελέσει τις μετρήσεις της εξερχόμενης έντασης I_d τις μετατρέπει σε ψηφιακή μορφή και τις μεταφέρει στον υπολογιστή. Στον υπολογιστή εκτελούνται οι υπολογισμοί για την ανακατασκευή της εικόνας, η οποία στη συνέχεια παρουσιάζεται στην οθόνη ή αποθηκεύεται.

Γεωμετρίες απόκτησης δεδομένων

Παραπάνω παρουσιάσαμε τα βασικά στοιχεία που εμπλέκονται στη διάταξη ενός τομογράφου. Φυσικά, υπάρχει πληθώρα υλοποιήσεων αξονικών τομογράφων και αυτοί μπορεί να παρουσιάζουν διαφοροποιήσεις μεταξύ τους, ειδικά όσο η τεχνολογία εξελίσσεται. Ιδιαίτερο ενδιαφέρον παρουσιάζει η διαφοροποίηση και η εξέλιξη του συστήματος της πηγής των ακτίνων και των ανιχνευτών τους. Η καλή σχεδίαση του συστήματος αυτού μπορεί να επηρεάσει κρίσιμες παραμέτρους της τομογραφίας όπως την ταχύτητα της σάρωσης, την ποιότητα της εικόνας και την δοσολογία της ακτινοβολίας στον ασθενή. Στην εξέλιξη της υπολογιστικής τομογραφίας έχουν υπάρξει διάφορες γενιές διατάξεων πηγών και ανιχνευτών, με την κάθε μια να επιφέρει σημαντική βελτίωση στις προηγούμενες μετρικές. Στη συνέχεια θα παρουσιάσουμε κάποιες από αυτές.

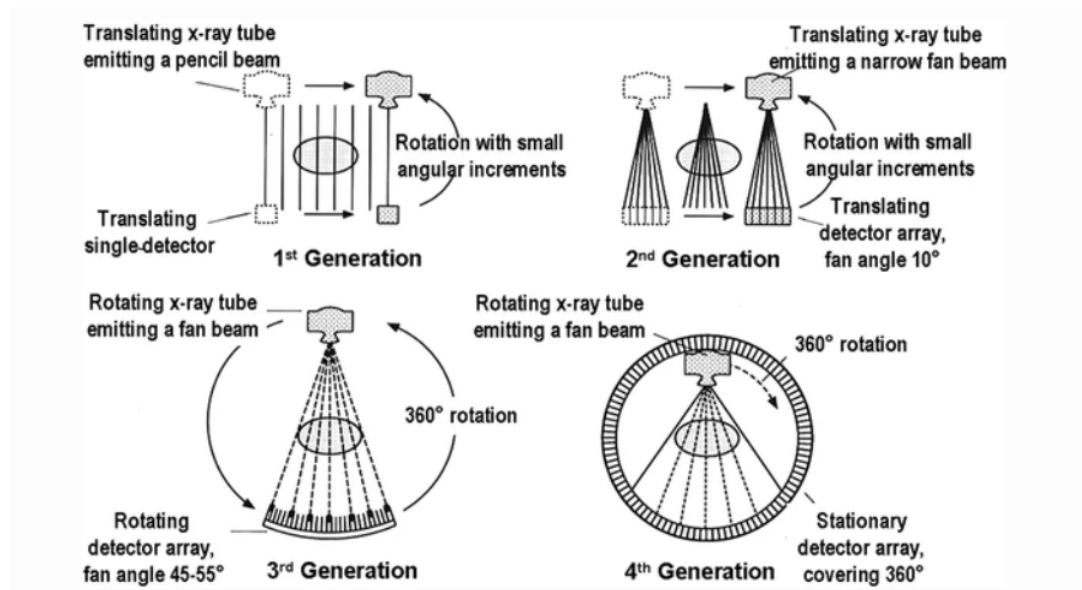


FIGURE 2.6: Οι τέσσερις πρώτες γενιές της υπολογιστικής αξονικής τομογραφίας

1η γενιά

Στην πρώτη γενιά είχε γίνει χρήση ενός ζεύγους από ένα x-ray tube και έναν ανιχνευτή. Για να καλύψουν όλη την επιφάνεια λειτουργούσαν επαναληπτικά με μετατόπιση. Στη συνέχεια επαναλαμβάνονταν η ίδια διαδικασία για διαφορετικές γωνίες, όπως επιβάλλεται για να είναι δυνατή η ανακατασκευή της εικόνας. Ο περιορισμένος αριθμός των στοιχείων και η σειριακή-επαναληπτική φύση της λειτουργίας τους είχε σαν αποτέλεσμα τον αργό συγκριτικά χρόνο σάρωσης και εικόνες χαμηλής ποιότητας.

2η γενιά

Στην δεύτερη γενιά έγινε χρήση αποκλίνουσας δέσμης x-ray tube και συστοιχίας ανιχνευτών. Έτσι σε μία λήψη συλλέγονται δεδομένα από πολλαπλές γωνίες, γεγονός που αυξάνει την ποιότητα της εικόνας. Σημειώστε πως το εύρος της αποκλίνουσας ακτινοβολίας δεν ήταν αρκετά μεγάλο έτσι ώστε να καλύψει ολόκληρη την επιφάνεια του ασθενή. Για αυτό το λόγο η χρήση της μετατόπισης της πηγής και των ανιχνευτών για μια γωνία σάρωσης έγινε και σε αυτή τη γενιά. Παρά το γεγονός αυτό ο χρόνος της σάρωσης μειώθηκε σημαντικά.

3η γενιά

Η γραμμική μετατόπιση τελικά παρακάμθηκε στην τρίτη γενιά αξονικών τομογράφων. Σε αυτή αυξήθηκε η γωνία της αποκλίνουσας δέσμης όπως και ο αριθμός των ανιχνευτών στον βαθμό που ήταν δυνατό για μια γωνία σάρωσης να καλύπτεται ολόκληρη η επιφάνεια του ασθενή. Αυτός είναι και ο λόγος που δεν χρειαζόταν πλέον η γραμμική μετατόπιση, με αποτέλεσμα την ακόμα καλύτερη ποιότητα εικόνων σε ακόμα λιγότερους χρόνους σάρωσης.

4η γενιά

Σε αυτή τη γενιά ο αριθμός των ανιχνευτών αυξήθηκε ακόμα περισσότερο, δημιουργώντας έναν πλήρη δακτύλιο. Αυτό το χαρακτηριστικό επέτρεψε την αποφυγή της περιστροφής των ανιχνευτών, αφού πλέον αρκούσε η περιστροφή μόνο του x-ray tube. Η αποφυγή της περιστροφής τους αύξησε ακόμα περισσότερο την ταχύτητα σάρωσης του τομογράφου.

5η γενιά

Η πέμπτη γενιά τομογράφων χαρακτηρίζεται από μια ελικοειδής σάρωση. Σε αυτή τη γενιά ο ασθενής μετακινείται στο εσωτερικό του περιστρεφόμενου gantry. Οι γρήγορες πολλαπλές σαρώσεις που προκύπτουν επιτρέπουν την ανακατασκευή τρισδιάστατων απεικονήσεων, με χρήση αλγορίθμων που λαμβάνουν υπόψη την ελικοειδή τροχιά των ακτίνων X.

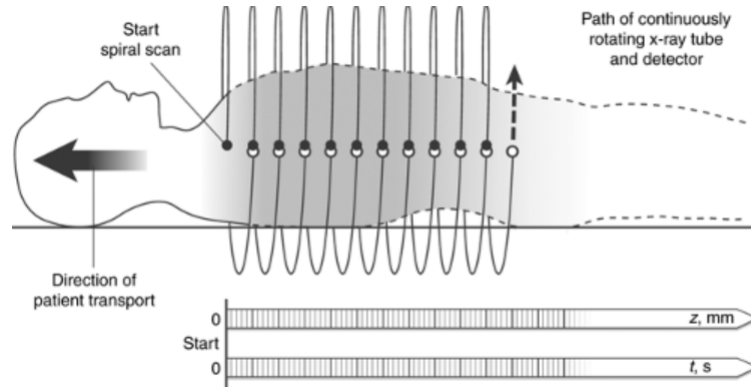


FIGURE 2.7: Η πέμπτη γενιά της υπολογιστικής αξονικής τομογραφίας και η ελικοειδής σάρωση

Επόμενες γενιές

Με το πέρασμα του χρόνου και την περαιτέρω εξέλιξη της τεχνολογίας της αξονικής τομογραφίας δημιουργούνται όλο και περισσότερες καινοτομίες. Μια πληθώρα από νέες ιδέες έχει εισαχθεί με πολλαπλά επίπεδα ανιχνευτών ή πολλαπλές πηγές οι οποίες λειτουργούν σε διαφορετικά επίπεδα ενέργειας οι οποίες μέρα με τη μέρα βελτιώνουν την ποιότητα των εικόνων και μειώνουν το επίπεδο έκθεσης του ασθενή σε ακτίνες X καθιστώντας την υπολογιστική αξονική τομογραφία όλο και πιο χρήσιμη και αποτελεσματική.

2.2.3 Ανακατασκευή εικόνας

Το πρόβλημα της ανακατασκευής της εικόνας αποτελεί τον ακρογωνιαίο λίθο όλων των μεθόδων των ιατρικών απεικονήσεων. Η φιλοδοξία της απεικόνισης των εσωτερικών ιστών ενός ανθρώπινου οργανισμού χωρίς την επέμβαση σε αυτούς δημιουργεί την ανάγκη επίλυσης ενός προβλήματος δύσκολου. Στην υπολογιστική τομογραφία, όπως προαναφέραμε, οι πολλαπλές λήψεις από διαφορετικές γωνίες των ακτίνων X είναι δυνατό να ανακατασκευάσει την εικόνα του ασθενή. Αυτή η κεντρική ιδέα της υπολογιστικής τομογραφίας και ο τρόπος με βάση τον οποίο ανακατασκευάζεται η εικόνα στηρίζονται στο θεώρημα τομής Fourier (Fourier slice theorem ή projection-slice theorem ή central slice theorem).

Θεώρημα τομής Fourier

Έστω μια συνάρτηση δύο διαστάσεων $f(x,y)$ η οποία αποτελεί το απεικονιστικό περιεχόμενο στην περίπτωση της τομογραφίας. Η συνάρτηση $f(x,y)$ είναι ο άγνωστος ο οποίος θέλουμε να προσδιορίσουμε. Έστω επίσης $P(\theta, t) = \int_L f(x,y) ds$ το επικαμπύλιο ολοκλήρωμα της $f(x,y)$ κατά μήκος μιας γραμμής L παράλληλης ως προς άξονα s , οριζόμενο από τη γωνία θ . Το $P(\theta, t)$ αποτελεί την προβολή της $f(x,y)$ στη γωνία θ και στην περίπτωση της τομογραφίας αποτελεί τιμή γνωστή, αφού προσδιορίζεται από την εξασθένηση των ακτίνων X.

Το θεώρημα τομής Fourier (βλέπε εικόνα 2.8) μας λέει ότι ο μετασχηματισμός Fourier $\mathcal{F}\{P(\theta, wt)\}$ αποτελεί την τομή του δισδιάστατου μετασχηματισμού Fourier $\mathcal{F}(wx, wy)$ της συνάρτησης $f(x,y)$ για τη γωνία θ :

$$\mathcal{F}\{P(\theta, wt)\} = \mathcal{F}(wx, wy), \text{ για } \theta \in (0, 180) \quad (2.3)$$

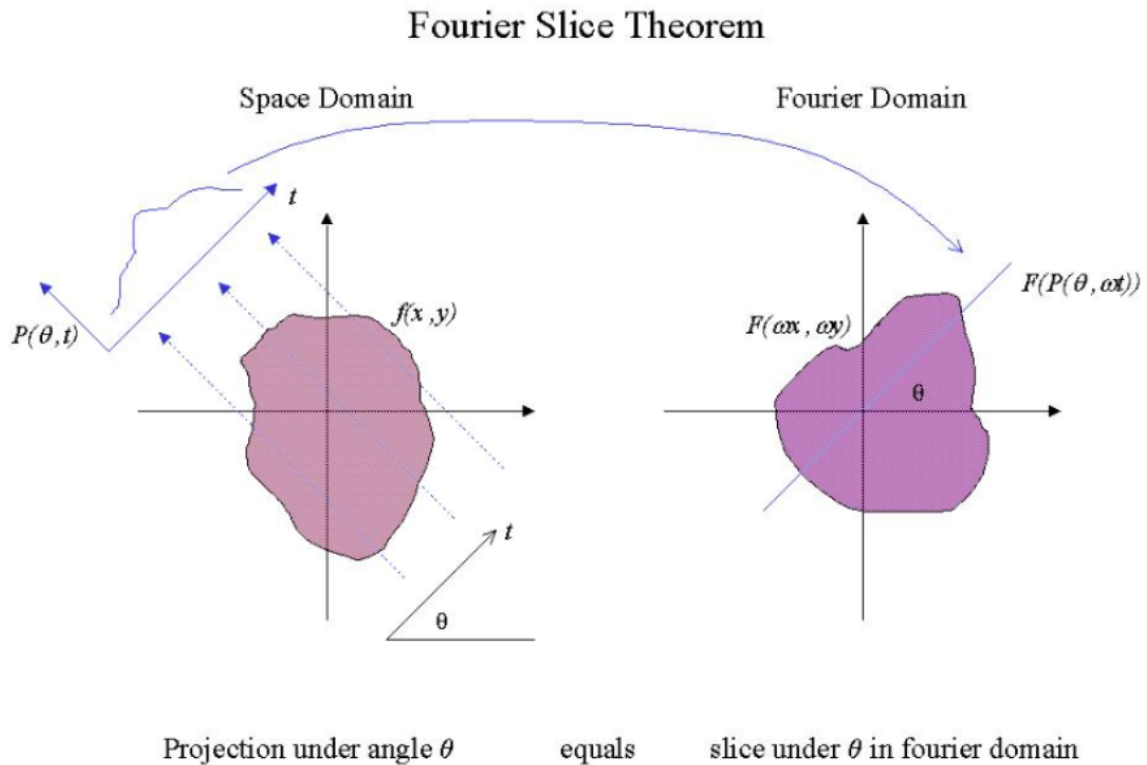


FIGURE 2.8: Το θεώρημα τομής Fourier

Οι τομές του $\mathcal{F}(wx, wy)$ για όλες τις γωνίες θ μπορούν να προσδιορίσουν όλο τον $\mathcal{F}(wx, wy)$ και αυτός με τη σειρά του την $f(x,y)$ (αντίστροφος μετασχηματισμός Fourier). Αυτός είναι και ο λόγος που λαμβάνονται λήψεις από πολλαπλές γωνίες. Προφανώς όσο μεγαλύτερος είναι ο αριθμός των λήψεων για διάφορες γωνίες θ τόσο περισσότερες τομές του μετασχηματισμού της $f(x,y)$ βρίσκουμε και τόσο καλύτερα την προσεγγίζουμε.

Αλγόριθμοι Backprojection

Γνωρίζοντας τα παραπάνω μπορούν να εφαρμοστούν υπολογιστικοί αλγόριθμοι για την ανακατασκευή μιας εικόνας. Αυτοί οι αλγόριθμοι ονομάζονται αλγόριθμοι οπισθοπροβολής (Backprojection). Στους αλγόριθμους οπισθοπροβολής η εικόνα προσεγγίζεται σαν ένα άθροισμα προβολών για κάθε γωνία:

$$\hat{f}(x, y) = \sum_{j=1}^n P(y \cos \theta_j + x \sin \theta_j) \Delta \theta_j \quad (2.4)$$

Όπου θ_j είναι η j -οστή γωνία προβολής και n ο συνολικός αριθμός των όψεων.

Πολλές φορές γίνεται και εφαρμογή φίλτρων στην κάθε προβολή πριν το άθροισμά τους για την βελτίωση της ποιότητας της εικόνας (Filtered Backprojection - FBP). Αν

$h(t)$ το φίλτρο, η κάθε προβολή $p(t, \theta)$ μετασχηματίζεται στην $p'(t, \theta)$ με συνέλιξη με το φίλτρο $h(t)$:

$$p'(t, \theta) = p(t, \theta) * h(t) \quad (2.5)$$

Σημειώνουμε πως για λόγους απόδοσης μπορεί να εφαρμοστεί το φίλτρο στο πεδίο της συχνότητας. Για το λόγο αυτό θα υπολογιστεί (με FFT) ο μετασχηματισμός Fourier του p , θα υπολογιστεί ο πολλαπλασιασμός με το μετασχηματισμό Fourier του φίλτρου $H(w)$ και τέλος θα υπολογιστεί ο αντίστροφος μετασχηματισμός Fourier της p' .

Επαναληπτικοί αλγόριθμοι

Υπάρχουν βέβαια και άλλες οικογένειες αλγορίθμων ανακατασκευής εικόνας, οι επαναληπτικοί αλγόριθμοι ανακατασκευής (Algebraic Reconstruction Technique - ART, Simultaneous Algebraic Reconstruction Technique - SART, Ordered Subset Expectation Maximization - OSEM). Στις επαναληπτικές μεθόδους ανακατασκευής η εύρεση της $f(x,y)$ αντιμετωπίζεται σαν ένα σύστημα εξισώσεων όπου ο αριθμός των αγνώστων ισούται με τον αριθμό των pixel της εικόνας και ο αριθμός των εξισώσεων με τον αριθμό των προβολών επί των όψεων.

Για παράδειγμα στην αλγεβρική τεχνική ανακατασκευής (ART) το σύστημα των εξισώσεων είναι το:

$$A^{mn} x^n = b^m \quad (2.6)$$

Στην παραπάνω εξίσωση είναι:

- n ο αριθμός των pixel
- m ο αριθμός των προβολών
- A ο πίνακας στατιστικών βαρών που αντιστοιχούν στη συνεισφορά κάθε pixel σε κάθε ακτίνα
- x οι τιμές των pixel
- b οι προβολές

Στην ART υπολογίζονται οι τιμές x επαναληπτικά με βάση την εξίσωση:

$$x^{k+1} = x^k + \lambda (b_i - A_i^T x^k) / \|A_i\|^2 \quad (2.7)$$

Όπου λ ο συντελεστής χαλάρωσης που παίρνει τιμές μεταξύ του 0 και του 2 και καθορίζει την ταχύτητα σύγκλισης.

2.3 Συνδυασμός εικόνων PET και CT: PET/CT

Είναι σκόπιμο καταλάβουμε ότι οι εικόνες PET και CT αποτελούν εναλλακτικές οι οποίες προσφέρουν διαφορετικές πληροφορίες. Οι PET εικόνες μπορούν να αναδείξουν εύκολα λειτουργική πληροφορία η οποία είναι ιδιαίτερα λεπτεπίλεπτη έτσι ώστε να ληφθεί από τη μέθοδο του αξονικού τομογράφου. Οι PET εικόνες όμως δεν μπορούν να δώσουν καλές πληροφορίες πέραν αυτών που σχετίζονται με μεταβολικές διαδικασίες. Η γεωμετρία των ιστών αλλά ακόμα και η τοποθεσία των ενδεχομένων καρκινικών κυττάρων που φαίνονται στις PET εικόνες στην γενικότερη ανατομία του ασθενή δεν μπορούν να γίνουν εύκολα γνωστές από τις PET εικόνες. Τις περισσότερες φορές οι πληροφορίες αυτές

είναι απαραίτητες για την παρακολούθηση και τη διάγνωση του ασθενή. Ο συνδυασμός λοιπόν των πληροφοριών και των PET και των CT εικόνων είναι πολλές φορές απαραίτητος. Στην αρχή έγινε μια απόπειρα να συνδυαστεί πληροφορία η οποία προερχόταν από ξεχωριστές-ανεξάρτητες PET και CT τομογραφίες. Αυτό δεν ευδοκίμησε καθώς παρουσιάστηκαν σημαντικές δυσκολίες στον συνδυασμό των εικόνων αυτών. Η λύση στα παραπάνω προβλήματα επήλθε από την από κοινού λήψη πληροφορίας ακτινών X και γ σε μια τομογραφία, την PET/CT. Στο ίδιο gantry υπάρχουν τα μέρη για και για τις δύο τομογραφίες και τελικά παράγεται μια εικόνα σύνθεση των δύο αυτών τομογραφιών. Παραθέτουμε ένα παράδειγμα μιας εικόνας από την τομογραφία (βλέπε εικόνα 2.9).

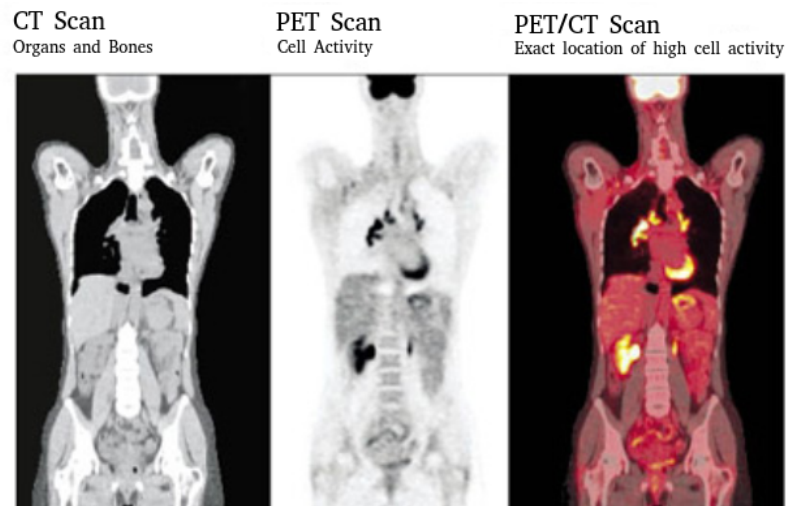


FIGURE 2.9: Ένα παράδειγμα εικόνας PET/CT

2.4 Εισαγωγή στο μεταστατικό μελάνωμα

Μεταστατικό μελάνωμα ονομάζουμε την μετάδοση καρκινικών κυττάρων από το δέρμα σε κάποιο άλλο σημείο του οργανισμού. Γενικότερα ως μετάσταση ονομάζουμε την διάδοση καρκινικών κυττάρων από μια αρχική πηγή σε άλλους ιστούς. Η δημιουργία δευτερευόντων (μεταστατικών) όγκων αποτελεί ένα επικίνδυνο φαινόμενο καθώς το μεγαλύτερο μέρος των θανάτων από καρκίνο προέρχεται από μεταστάσεις. Το μελάνωμα είναι κακοήθης όγκος του δέρματος. Προκαλείται μετά από κακοήθεια των μελανοκυττάρων. Τα μελανοκύτταρα βρίσκονται κυρίως στο δέρμα και είναι υπεύθυνα για την παραγωγή της μελανίνης, η οποία αποτελεί άμυνα απέναντι στις υπεριώδεις ακτίνες του ηλίου. Το μελάνωμα όταν βρίσκεται στο δέρμα ή στις άλλες περιοχές που παράγεται δεν αποτελεί ιδιαίτερο κίνδυνο για τον ασθενή. Μπορεί όμως να μεταναστεύσει σε άλλα κύτταρα και ιστούς. Η μετανάστευση αυτή μπορεί να γίνει μέσω του αίματος και των λεμφαδένων και το μέγεθος της μετανάστευσης αυτής καθορίζει και το προσδόκιμο ζωής των ασθενών. Αυτός είναι και ο λόγος για τον οποίο είναι κρίσιμη η ακριβής μελέτη της διάδοσης του μελανώματος. Ορίζονται λοιπόν πέντε στάδια μελανώματος (βλέπε εικόνα 2.10).

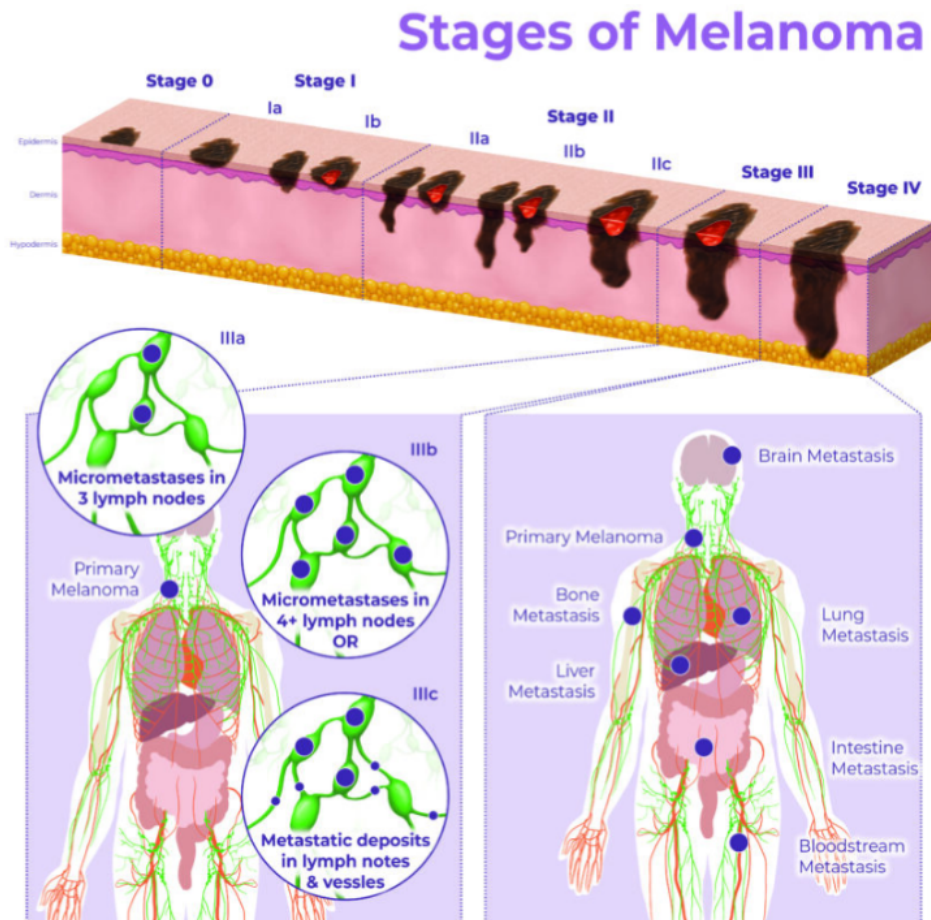


FIGURE 2.10: Τα στάδια του μεταστατικού μελάνωματος

Στάδιο 0

Σε αυτό το στάδιο το μελάνωμα είναι περιορισμένο στην επιδερμίδα. Δεν έχει εισχωρήσει βαθύτερα στο δέρμα ούτε σε άλλα γειτονικά κύτταρα.

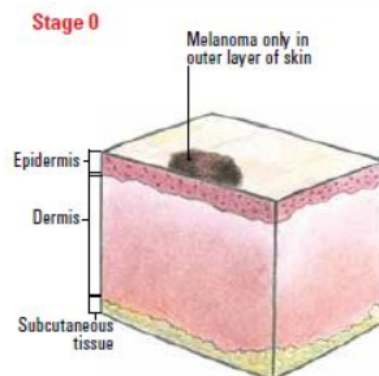


FIGURE 2.11: Το στάδιο 0 του μεταστατικού μελάνωματος

Στάδιο I

Το μελάνωμα έχει εισχωρήσει σε βαθύτερα επίπεδα της επιδερμίδας αλλά παραμένει να είναι τοπικό και μη επειληγτικό. Το στάδιο I κατηγοριοποιείται περισσότερο σε 2 στάδια: το IA και το IB.

Στο στάδιο IA το μελάνωμα δεν έχει ξεπεράσει το 1 mm σε πάχος ούτε έχει προκαλέσει έλκος στο δέρμα (skin ulceration).

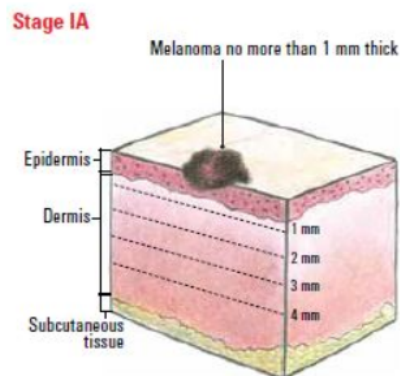


FIGURE 2.12: Το στάδιο IA του μεταστατικού μελανώματος

Στο στάδιο IB το μελάνωμα μπορεί είτε να έχει ξεπεράσει το 1 mm (όχι όμως περισσότερο από 2mm) χωρίς έλκος του δέρματος, είτε μην έχει ξεπεράσει το 1 mm αλλά να έχει προκαλέσει έλκος.

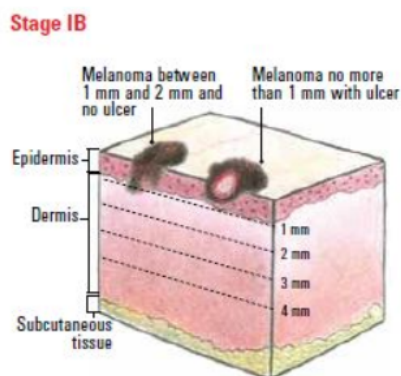


FIGURE 2.13: Το στάδιο IB του μεταστατικού μελανώματος

Στάδιο II

Σε αυτό το στάδιο το μελάνωμα έχει εισχωρήσει σε ακόμα βαθύτερο επίπεδο, στο δέρμα και ίσως σε κάποιους λεμφαδένες και άλλα τοπικά κύτταρα σε ένα βαθμό αλλά όχι εκτεταμένο. Το στάδιο II χωρίζεται σε 3 στάδια: το IIA, το IIB και το IIC.

Στο στάδιο IIA το μελάνωμα είτε έχει πάχος μεταξύ 2 mm και 4 mm χωρίς έλκος, είτε έχει πάχος μεταξύ 1 mm και 2 mm με έλκος.

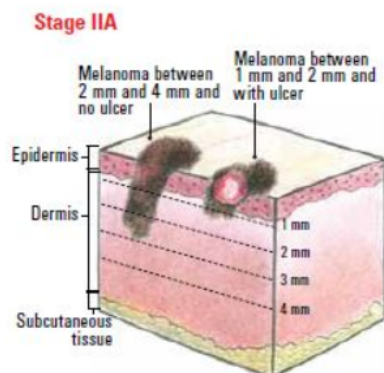


FIGURE 2.14: Το στάδιο IIA του μεταστατικού μελανώματος

Στο στάδιο IIB το μελάνωμα είτε έχει πάχος μεγαλύτερο από 4 mm χωρίς έλκος, είτε έχει πάχος μεταξύ 2 mm και 4 mm με έλκος.

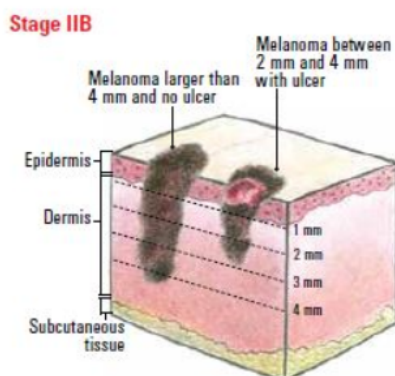


FIGURE 2.15: Το στάδιο IIB του μεταστατικού μελανώματος

Στο στάδιο IIC το μελάνωμα έχει ξεπεράσει τα 4 mm πάχους με παρουσία έλκους.

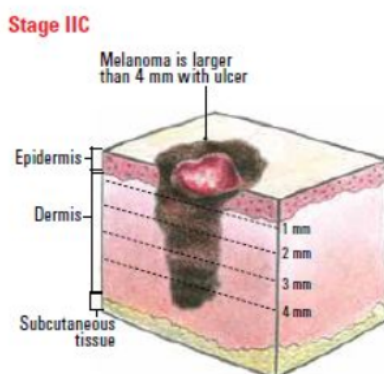


FIGURE 2.16: Το στάδιο IIC του μεταστατικού μελανώματος

Στάδιο III

Εδώ το μελάνωμα έχει επεκταθεί και σε λεμφαδένες και κύτταρα πέραν της αρχικής του περιοχής. Οι περιοχές αυτής της επέκτασης είναι κυρίως γύρω από την αρχική

περιοχή του μελανώματος αλλά θεωρούνται ξένες αυτής. Το στάδιο αυτό χωρίζεται σε 3 στάδια: το IIIA, το IIIB και το IIIC. Στο IIIA έχουν γίνει μικρομεταστάσεις σε όχι πάνω από 3 λεμφαδένες. Στο στάδιο IIIB έχουν γίνει είτε μακρομεταστάσεις σε όχι πάνω από 3 λεμφαδένες είτε παρουσιάζονται μεταστάσεις σε λεμφαγγεία αλλά όχι σε λεμφαδένες. Στο στάδιο IIIC είτε έχουν γίνει μακρομεταστάσεις σε πάνω από 4 λεμφαδένες είτε παρουσιάζονται μεταστάσεις και σε λεμφαγγεία και σε λεμφαδένες.

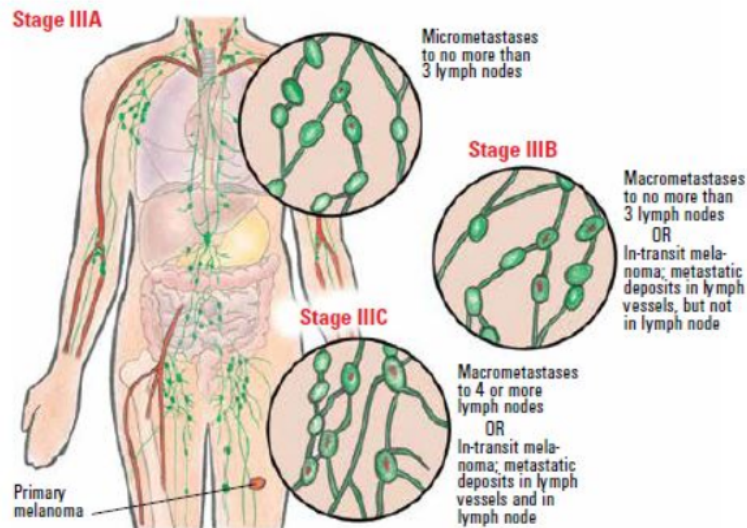


FIGURE 2.17: Το στάδιο III του μεταστατικού μελανώματος

Στάδιο IV

Το μελάνωμα έχει κάνει μετάσταση σε όργανα και λεμφαδένες μακριά της αρχικής αλλά και γειτονικής στην αρχική περιοχές. Αποτελεί το τελευταίο και πιο επικίνδυνο στάδιο της κατηγοριοποίησης του μελανώματος. Κάποιες από τις περιοχές αυτές μπορεί να είναι οι πνεύμονες, το συκώτι, ο εγκέφαλος και τα κόκκαλα.

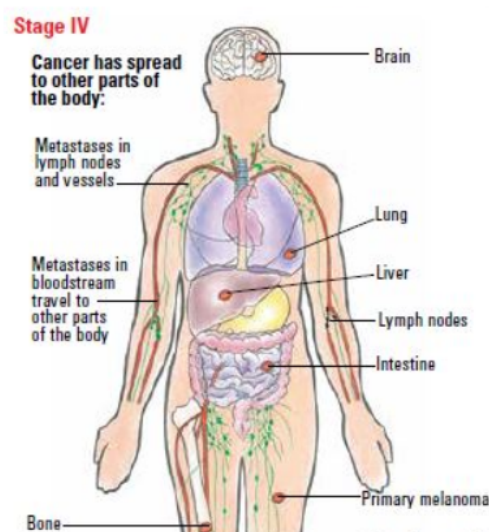


FIGURE 2.18: Το στάδιο IV του μεταστατικού μελανώματος

2.5 Εντοπισμός μεταστατικού μελανώματος με χρήση PET/CT εικόνων

Η PET/CT τομογραφία χαρακτηρίζεται μια από τις πιο αποτελεσματικές μέθοδος για τον εντοπισμό του μεταστατικού μελανώματος. Χρησιμοποιείται σε πολλές θεραπείες και στο στάδιο του σχεδιασμού τους αλλά και κατά τη διάρκειά τους για την επίβλεψη της προόδου τους. Τα καρκινικά κύτταρα παρουσιάζουν διαφορετικό μεταβολισμό από τα κανονικά και ως εκ τούτου στις PET/CT εικόνες εμφανίζονται (συνήθως) με μεγαλύτερη ένταση (βλέπε εικόνα 2.19)(Bisschop et al., 2020).

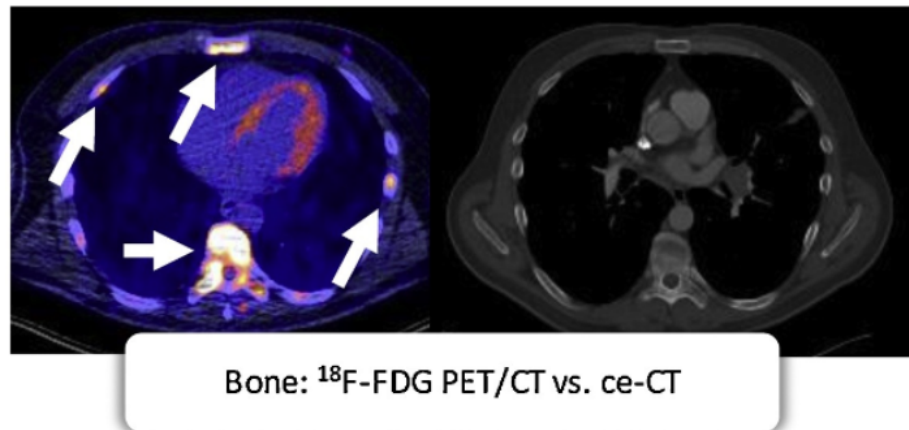


FIGURE 2.19: Τα σημεία ενδιαφέροντος στις PET εικόνες εμφανίζονται πιο έντονα (βλέπε βέλη)

Τις πιο πολλές φορές οι περιοχές ενδιαφέροντος πράγματι είναι δυνατό να γίνουν διακριτές. Άλλες φορές όχι. Εκτός από τους περιορισμούς της ποιότητας των εικόνων, υπάρχουν και άλλοι περιορισμοί. Τα καρκινικά κύτταρα δεν είναι τα μόνα που παρουσιάζουν διαφορετικές διαδικασίες μεταβολισμού. Συγκεκριμένα κύτταρα του ανθρώπινου οργανισμού έχουν από τη φύση τους διαφορετικούς μεταβολισμούς και ως εκ τούτου κάνουν δύσκολη την διάκριση από τα καρκινικά. Για να γίνει αυτό κατανοητό παραθέτουμε την εικόνα 2.20 όπου παρουσιάζονται μια τομογραφία PET/CT και μια τομογραφία μαγνητική, στην περιοχή του εγκεφάλου τα κύτταρα του οποίου αποτελούν ένα χαρακτηριστικό παράδειγμα. Το λευκό βέλος στη μαγνητική τομογραφία καταδεικνύει μια περιοχή καρκινική, την οποία αποτυγχάνει να καταδείξει το ίδιο καλά (αν καταδεικνύει έστω και λίγο) η PET/CT εικόνα.

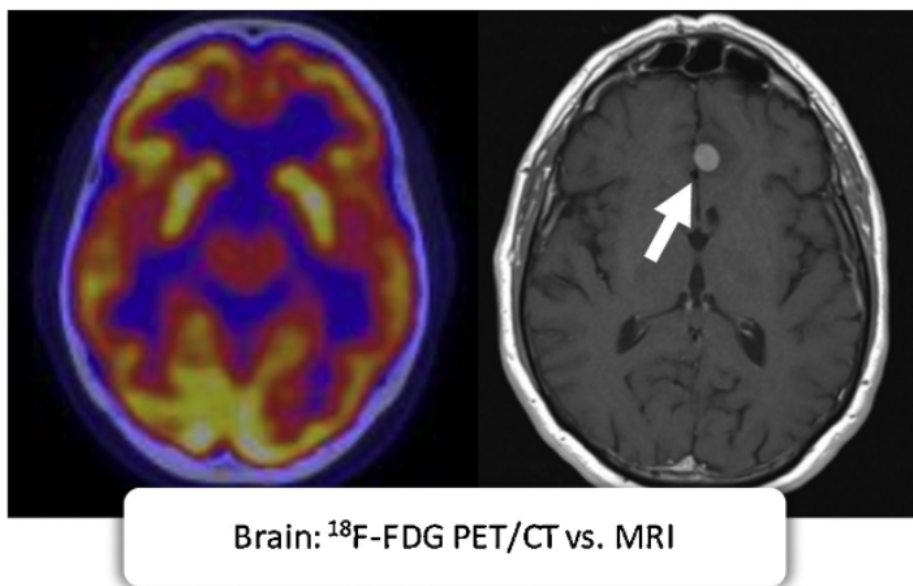


FIGURE 2.20: Η PET εικόνα αποτυγχάνει να καταδείξει τον όγκο στον εγκέφαλο (όπως φαίνεται στην MRI εικόνα)

Άλλοι παράγοντες που μπορούν να κάνουν δύσκολη τη διάγνωση αποτελούν διάφορες παθήσεις οι οποίες επίσης μπορούν επίσης να παρουσιάζουν μεγαλύτερη απορρόφηση των ραδιοϊσοτόπων. Στην εικόνα 2.21 (Aide et al., 2022) παρουσιάζεται μια περίπτωση μιας σαραντατετράχρονης γυναίκας στην οποία σημάδια ακμής στο πρόσωπο (A),(B) και στον κορμό (C) μπορούν εύκολα να παρερμηνευτούν ως καρκινικά (false positives).

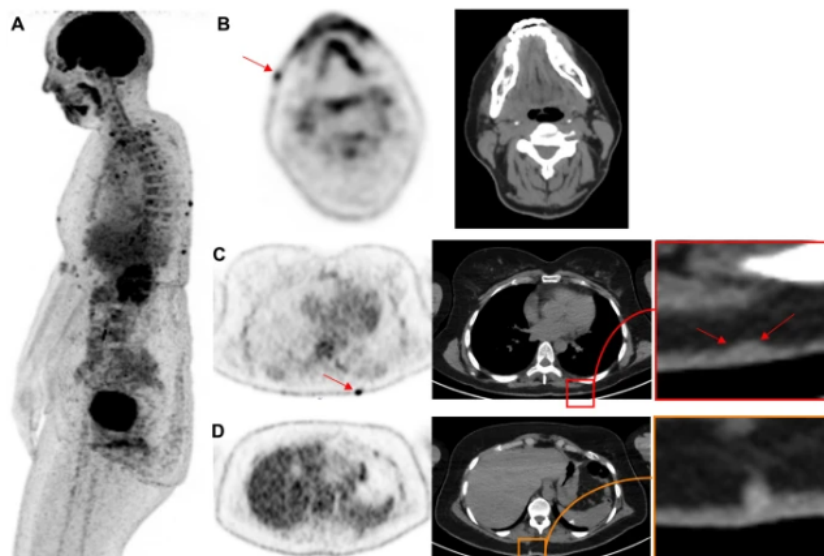


FIGURE 2.21: PET false positives. Σημάδια ακμής στο πρόσωπο (A),(B) και στον κορμό (C) μοιάζουν με καρκινικά

Υπάρχουν όμως και περιπτώσεις μεταστάσεων οι οποίες δεν εμφανίζονται στις PET/CT

(false negatives). Στην εικόνα 2.22 παρουσιάζεται η περίπτωση μιας εξηντατριάχρονης γυναίκας η οποία στην πρώτη τομογραφία που της έγινε φαίνεται να μην παρουσιάζει ιδιαίτερη συγκέντρωση καρκινικών κυττάρων (A), με τα μαύρα βελόνια να δείχνουν κάποιες περιοχές που είναι σχεδόν το ίδιο έντονες με το background. Μετά από τρεις μήνες ακολούθησε και άλλη εξέταση η οποία έδειξε ότι τελικά το πρόβλημα ήταν μεγαλύτερο απ' ό τι αρχικά φαινόταν (D).

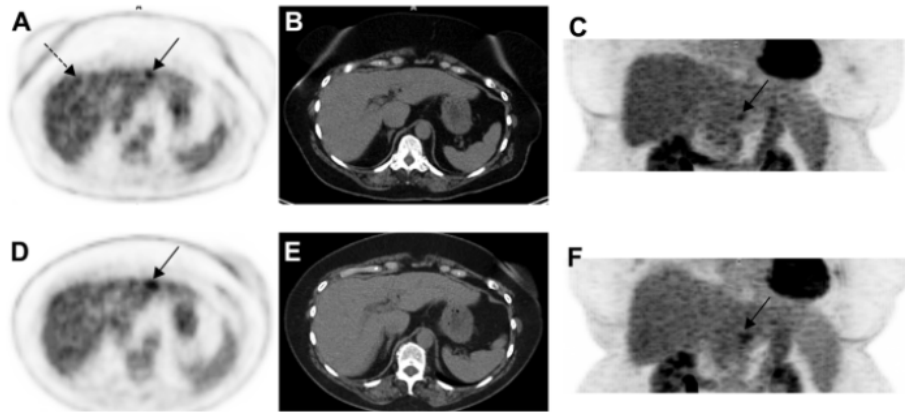


FIGURE 2.22: PET false negatives. Τα καρκινικά κύτταρα (D) δεν φαίνονται πάντα στις PET εικόνες (A)

Σύμφωνα με τα παραπάνω γίνεται σαφές πως η PET/CT τομογραφία είναι ένα πολύ ισχυρό εργαλείο στη διάγνωση του μεταστατικού μελανώματος αλλά χωρίς αυτό να σημαίνει ότι καλύπτει απόλυτα όλες τις προϋποθέσεις έτσι ώστε να προκύπτουν σωστά συμπεράσματα με μεμονωμένες αξιολογήσεις εικόνων. Σε πολλές περιπτώσεις πρέπει να συνυπολογιστεί μια πληθώρα άλλων παραγόντων όπως πληροφορίες σχετικές με το ιστορικό του προς εξέταση ασθενή ή τα σημεία στα οποία μπορεί να μη δίνονται ιδιαίτερα αξιόπιστες συγκριτικές εντάσεις στις εικόνες. Σε αυτές τις περιπτώσεις θα πρέπει να γίνεται διασταύρωση όλων αυτών των στοιχείων με το περιεχόμενο των PET/CT εικόνων αλλά ακόμα και με άλλες τομογραφικές εικόνες που έχουν αποδείξει μεγαλύτερη αξιοπιστία σε συγκεκριμένα σημεία ή εφαρμογές (MRI στον εγκέφαλο, υπέρηχοι στους λεμφαδένες κλπ).

Κεφάλαιο 3

Νευρωνικά Δίκτυα και Βαθιά μάθηση

3.1 Βασικές έννοιες στα Νευρωνικά Δίκτυα

3.1.1 Εισαγωγή

Η ιστορία των νευρωνικών δικτύων είναι ίσως μεγαλύτερη από ό,τι μπορεί να περιμέναμε. Οι απαρχές τους ξεκίνησαν με το έργο των Warren McCulloch (Νευροφυσιολόγος) και Walter Pitts (Λογικολόγος) "*A Logical Calculus of Ideas Immanent in Nervous Activity*" το 1943 (McCulloch and Pitts, 1943). Στόχος τους ήταν να δημιουργήσουν τεχνητά συστήματα τα οποία θα μπορούσαν να εκτελούν λειτουργίες παρόμοιες αυτών των βιολογικών νευρώνων. Αποτέλεσμα του έργου τους αποτέλεσε το perceptron το οποίο υλοποιήθηκε από τον Frank Rosenblatt το 1957 "*The Perceptron—a perceiving and recognizing automaton*" (Rosenblatt, 1957). Το perceptron αποτελούσε έναν νευρώνα ο οποίος μπορούσε να εκπαιδευτεί ώστε να μπορεί να κατηγοριοποιήσει διάφορες εισόδους σε δύο κλάσεις.

Το 1969 οι Marvin Minsky και Seymour Papert δημοσίευσαν το βιβλίο ονόματι "*Perceptrons*", το οποίο κατέδειξε αδυναμίες που είχε το perceptron. Το perceptron δεν μπορούσε να διαχωρίσει κλάσεις οι οποίες δεν ήταν γραμμικά διαχωρίσιμες. Αυτό περιόριζε πολύ την χρησιμότητα του perceptron καθώς οι γραμμικά διαχωρίσιμες κλάσεις αποτελούν ένα πολύ μικρό σύνολο προβλημάτων. Ακόμα και το απλό πρόβλημα του υπολογισμού της συνάρτησης XOR, όπως κατέδειξαν οι Minsky και Papert, δεν μπορούσε να κατηγοριοποιηθεί από το perceptron λόγω της μη γραμμικής φύσης του. Για την επίλυση προβλημάτων μη γραμμικά διαχωρίσιμων κλάσεων άρχισε να γίνεται χρήση πολλαπλών perceptron (MultiLayer Perceptron - MLP). Η χρήση των MLP σε συνδυασμό και η εκπαίδευσή τους με τον αλγόριθμο οπισθοδιάδοσης (backpropagation) μπορεί να θεωρηθεί η αρχή των νευρωνικών δικτύων.

Παρόλο που ακόμα και από την δεκαετία του 1980 είχαν αρχίσει να δημιουργούνται ιδέες για εφαρμογές των νευρωνικών δικτύων σε πολλούς κλάδους όπως κατηγοριοποίηση εικόνων και αναγνώριση ομιλίας, έλειπαν διάφορες ιδέες βελτιστοποίησης αλλά και τα υπολογιστικά μέσα έτσι ώστε να γίνουν δυνατές οι αντίστοιχες πρακτικές υλοποιήσεις. Πλέον έχουν αναπτυχθεί και αξιοποιούνται υπολογιστικά συστήματα (GPUs, TPUs, και πολλά άλλα) τα οποία μπορούν να υποστηρίξουν τον μεγάλο αριθμό υπολογισμών που απαιτούνται. Συνεχίζουν μάλιστα να αναπτύσσονται μαζί με ολόκληρα περιβάλλοντα ανάπτυξης που καθιστούν δυνατή την εύκολη εκμετάλλευσή τους για την δημιουργία νέων εφαρμογών. Η ραγδαία ταυτόχρονη εξέλιξη όλων αυτών των κλάδων έχει καρποφορήσει ώστε σήμερα να έχουμε λειτουργικές εφαρμογές σε πολλούς κλάδους όπως στην όραση υπολογιστών, την επεξεργασία φυσικής γλώσσας, τα συστήματα συστάσεων, τα αυτόνομα συστήματα και την ιατρική διάγνωση.

3.1.2 Κεντρική ιδέα

Η κεντρική ιδέα των νευρωνικών δικτύων έγκειται στην δημιουργία συστημάτων τα οποία σε μια δεδομένη είσοδο παράγουν έξοδο η οποία αποτελεί μια εκτίμηση για τα δεδομένα της εισόδου. Αξίζει να εστιάσουμε περισσότερο στις έννοιες “είσοδος”, “δημιουργία” και “έξοδος”.

Είσοδος

Είσοδος στα νευρωνικά δίκτυα είναι τα δεδομένα που θέλουμε να αναλύσουμε σε μια μορφή που αυτά μπορούν να την επεξεργαστούν. Τα δεδομένα αυτά μπορούν να αναπαριστούν ένα πλήθος πληροφοριών όπως αριθμούς, εικόνες, ήχους, γραπτό λόγο.

Δημιουργία

Η δημιουργία των νευρωνικών δικτύων μπορεί να χωριστεί σε δύο κατηγορίες, αυτές της σχεδίασης της δομής τους και της εκπαίδευσής τους.

Ο σχεδιασμός της δομής του νευρωνικού δικτύου αποτελεί τον ορισμό των δομικών στοιχείων και του τρόπου της διασύνδεσής τους ώστε να καθοριστεί το είδος των υπολογισμών που το νευρωνικό θα εκτελέσει ώστε από την είσοδο που του δίνεται να παράξει την έξοδο. Η τυπική διάταξη της δομής των νευρωνικών δικτύων αποτελείται από ένα επίπεδο εισόδου, από τα εσωτερικά επίπεδα νευρώνων που εκτελούν τους υπολογισμούς (hidden layers) και από ένα επίπεδο εξόδου (βλέπε εικόνα 3.1). Συνήθως κάθε νευρώνας ενός επιπέδου του νευρωνικού δικτύου συνδέεται με όλους τους νευρώνες του προηγούμενου και του επόμενου επιπέδου (Fully Connected).

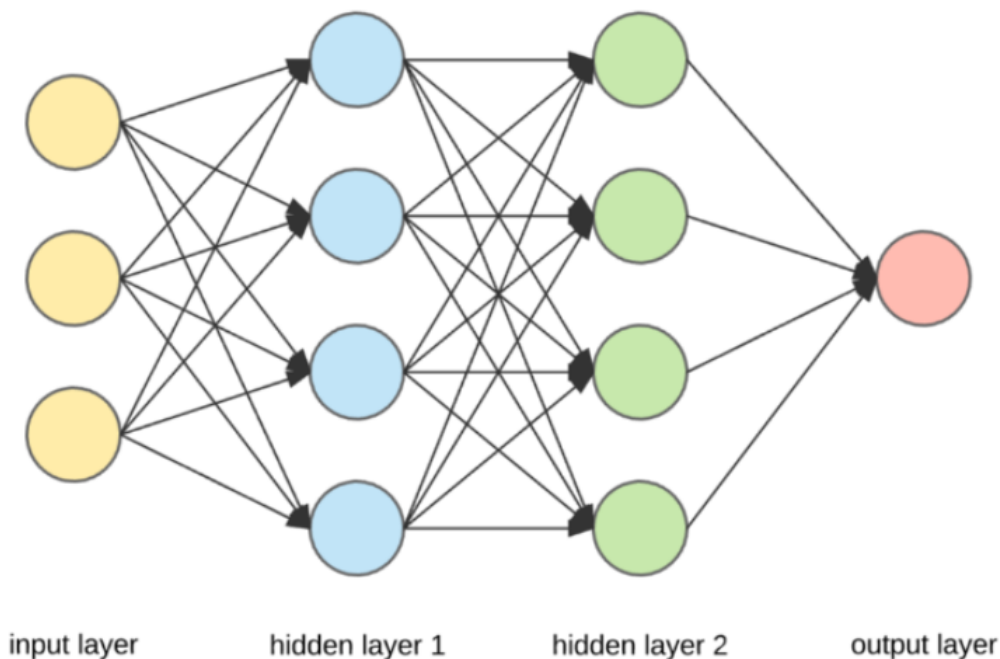


FIGURE 3.1: Η δομή των επιπέδων των νευρωνικών δικτύων. Αποτελείται από ένα επίπεδο εισόδου, ένα επίπεδο εξόδου και έναν αριθμό από εσωτερικά επίπεδα όπου οι νευρώνες των γειτονικών επιπέδων συνδέονται μεταξύ τους

Η εκπαίδευση του δικτύου αφορά τη διαδικασία η οποία θα καθορίσει τις τιμές με βάση τις οποίες αυτό θα εκτελέσει τους υπολογισμούς ώστε από την είσοδο που του δίνεται να παράξει την έξοδο. Οι τιμές αυτές ονομάζονται εκπαιδευσιμες παράμετροι (learnable parameters) του νευρωνικού. Ο υπολογισμός των τιμών αυτών στην εκπαίδευση

Ξεκινάει από τα τελευταία επίπεδα του δικτύου με στόχο τη μείωση της απόκλισης της εκτίμησης από τις επιθυμητές τιμές και διαδίδεται προς τα πρώτα με μια διαδικασία που ονομάζεται οπισθοδιάδοση (Backpropagation). Σημειώστε ότι είναι δυνατό ένα νευρωνικό με την ίδια είσοδο και με την ίδια δομή να παράξει διαφορετική έξοδο, εφόσον είναι διαφορετικά εκπαιδευμένο και τελικά κάνει διαφορετικούς υπολογισμούς. Οι εκπαιδευσιμες παράμετροι αποτελούνται από τα βάρη (weights) και τις πολώσεις (biases). Τα βάρη καθορίζουν τους αριθμούς με τους οποίους ένας νευρώνας πολλαπλασιάζει τις εισόδους του και οι πολώσεις τον αριθμό με τον οποίο προστίθενται (βλέπε εικόνα 3.2). Περισσότερες λεπτομέρειες για τα βάρη και τις πολώσεις θα δοθούν στη συνέχεια.

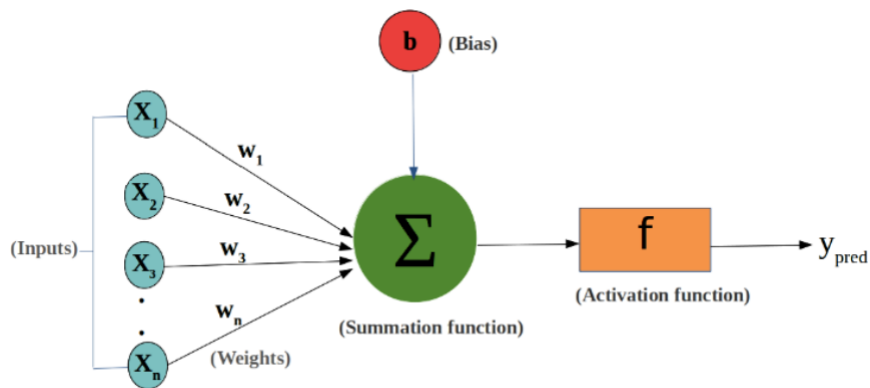


FIGURE 3.2: Η δομή και λειτουργία ενός νευρώνα. Τα βάρη w_n πολλαπλασιάζονται με τις εισόδους X_n και στη συνέχεια προστίθεται η πόλωση b . Στη συνέχεια, η έξοδος του νευρώνα τυπικά περνάει από μια συνάρτηση ενεργοποίησης

Έξοδος

Όπως προαναφέρθηκε η έξοδος του νευρωνικού αποτελεί μια εκτίμηση στα δεδομένα εισόδου που του δίνονται. Ανάλογα τα δεδομένα εισόδου και το είδος του προβλήματος η φύση της εκτίμησης αυτής μπορεί να είναι μια πρόβλεψη κλάσης για ένα δυαδικό πρόβλημα, μια πιθανότητα κατηγοριοποίησης, μια τιμή παλινδρόμησης (Regression) για την πρόβλεψη μιας τιμής μιας εξαρτημένης μεταβλητής από ένα σύνολο ανεξάρτητων μεταβλητών, ή μια παραγόμενη τιμή (δημιουργία κειμένου, δημιουργία εικόνων).

3.1.3 Πρόβλεψη Νευρωνικών Δικτύων

Όπως προαναφέρθηκε, δεδομένης μιας εισόδου και ενός εκπαιδευμένου νευρωνικού αυτό παράγει μια έξοδο-εκτίμηση. Η διαδικασία αυτή ονομάζεται πρόβλεψη. Οι υπολογισμοί που κάνει ένα νευρωνικό δίκτυο ξεκινάνε από τα αρχικά επίπεδα της δομής του και διαδοχικά προχωρούν στα βαθύτερα, μέχρι να καταλήξουν στην έξοδο. Η διαδικασία αυτή ονομάζεται εμπροσθοδρόμηση (Forward Propagation).

Κάθε νευρώνας, για να εκτελέσει τους υπολογισμούς του αξιοποιεί τις εκπαιδευσιμες παραμέτρους του οι οποίες είναι τα βάρη (weights) και οι πολώσεις (bias) του. Πιο συγκεκριμένα, κάθε νευρώνας ενός επιπέδου του νευρωνικού δικτύου πολλαπλασιάζει κάθε τιμή που του δίνεται από προηγούμενο νευρώνα (είσοδος) με ένα βάρος. Όσο μεγαλύτερο είναι αυτό το βάρος τόσο ισχυρότερη θεωρείται η συγκεκριμένη σύνδεση. Στη συνέχεια προσθέτει αυτά τα γινόμενα από όλες τις εισόδους του και στο τελικό αυτό αποτέλεσμα προσθέτει και την πόλωση. Πριν την τελική έξοδο του νευρώνα το μέχρι τώρα αποτέλεσμα περνάει μέσα από τη συνάρτηση ενεργοποίησης. Η χρήση

της συνάρτησης ενεργοποίησης είναι κρίσιμη και θα περιγραφεί στη συνέχεια. Η πόλωση αποτελεί επίσης ένα κρίσιμο κομμάτι στους υπολογισμούς του κάθε νευρώνα αφού επιτρέπει την μετατόπιση της συνάρτησης ενεργοποίησης αλλά παρέχει και γενικότερα μια καλή προσαρμοστικότητα στα νευρωνικά δίκτυα.

Όταν οι νευρώνες των τελευταίων σταδίων του νευρωνικού έχουν έτοιμες τις εισόδους τους από τα προηγούμενα στάδια, θα εκτελέσουν και αυτοί με τη σειρά τους τους υπολογισμούς τους και θα παράξουν την έξοδο. Η έξοδος ενός νευρωνικού δικτύου αποτελεί την πρόβλεψή του για τα δεδομένα τις εισόδου. Αν το πεδίο ενδιαφέροντός μας είναι ή περιγράφεται από μια συνάρτηση y , ο στόχος της σχεδίασης των νευρωνικών δικτύων είναι η πρόβλεψή τους, έστω y_{pred} , να προσεγγίζει όσο περισσότερο γίνεται τις πραγματικές τιμές της συνάρτησης y . Οι y και y_{pred} μπορούν να είναι οποιοσδήποτε συναρτήσεις και το αναφέρουμε αυτό για να καταδείξουμε τη δύναμη των νευρωνικών δικτύων.

3.1.4 Εκπαίδευση Νευρωνικών Δικτύων

Στην προηγούμενη ενότητα περιγράψαμε πώς όταν τα νευρωνικά δίκτυα έχουν έτοιμες τις παραμέτρους τους μπορούν να παράξουν τις εξόδους τους, δηλαδή να προβλέψουν. Η προετοιμασία των παραμέτρων αυτών συνιστά την εκπαίδευση του νευρωνικού δικτύου. Η εκπαίδευση ενός νευρωνικού δικτύου αποτελεί μια δύσκολη και χρονοβόρα διαδικασία, πολύ περισσότερο από την πρόβλεψη. Η διαδικασία είναι χρονοβόρα γιατί η εκπαίδευση των νευρωνικών δικτύων αποτελεί μια **επαναληπτική διαδικασία** στην οποία αυτά:

1. **Προβλέπουν** στην είσοδο που τους δίνεται την έξοδο.
2. **Υπολογίζουν** πόσο λάθος είναι οι προβλέψεις τους (απόκλιση).
3. **Προσαρμόζονται**, ώστε να μειώσουν τα λάθη στις προβλέψεις τους.

Ο υπολογισμός των λαθών και η προσαρμογή των νευρωνικών δικτύων αποτελούν βήματα τα οποία εισάγονται στην εκπαίδευση τους. Για να μπορέσουν τα νευρωνικά δίκτυα να υπολογίσουν λάθη και να προσαρμοστούν πρέπει να έχουν ένα σημείο αναφοράς. Έτσι θα μπορούν να συγκρίνουν πόσο μακριά είναι οι προβλέψεις τους από αυτό το σημείο αναφοράς. Η πρακτική της τροφοδότησης ενός νευρωνικού με δεδομένα τα οποία αποτελούν ένα σημείο αναφοράς για το τι πρέπει να μάθει (ground truth-labels) ονομάζεται επιβλεπόμενη μάθηση (supervised learning). Πέραν της επιβλεπόμενης μάθησης υπάρχει η μη-επιβλεπόμενη μάθηση (unsupervised learning) όπως και η ενισχυτική μάθηση (reinforcement learning). Το είδος της μάθησης αφορά τη φύση του προβλήματος που καλείται να επιλύσει ένα νευρωνικό δίκτυο. Η μη-επιβλεπόμενη μάθηση για παράδειγμα ενδείκνυται σε προβλήματα παραγωγής δεδομένων. Στο πλαίσιο της διπλωματικής αυτής θα αναλύσουμε την επιβλεπόμενη μάθηση και όταν αναφερόμαστε σε εκπαίδευση θα αναφερόμαστε στη διαδικασία τροφοδότησης του νευρωνικού με δεδομένα εισόδου (inputs) αλλά και αναμενόμενης - "σωστής" εξόδου (labels). Στη συνέχεια θα παρουσιάσουμε με περισσότερη λεπτομέρεια το κάθε βήμα της εκπαίδευσης.

Πρόβλεψη

Η πρόβλεψη σαν στάδιο της εκπαίδευσης είναι ίδια με την διαδικασία που περιγράφηκε παραπάνω. Η διαφορά στην εκπαίδευση είναι ότι όταν το νευρωνικό προβλέπει δεν θεωρούμε ότι η διαδικασία έληξε. Το νευρωνικό στην εκπαίδευση θα κάνει προβλέψεις συχνά ώστε να μπορούμε να γνωρίζουμε πόσο καλά έχει μάθει κάτι ώστε στη συνέχεια να μπορεί να βελτιωθεί περισσότερο.

Υπολογισμός απόκλισης

Αφού το νευρωνικό έχει κάνει την πρόβλεψή του υπολογίζεται στις εξόδους του η απόκλιση από τις αναμενόμενες τιμές που του δίνουμε. Ο υπολογισμός της απόκλισης από

τις αναμενόμενες τιμές γίνεται με τις συναρτήσεις κόστους. Υπάρχουν πολλές συναρτήσεις κόστους και ανάλογα το είδος του προβλήματος μπορούν να αποδειχθούν κρίσιμες για την σωστή εκπαίδευση του νευρωνικού. Θα παρουσιάσουμε κάποιες συναρτήσεις κόστους παρακάτω μαζί με τα μειονεκτήματα και τα πλεονεκτήματά τους, μπορούμε όμως σε αυτό το σημείο να αναφέρουμε ότι σε γενικές γραμμές μια καλή συνάρτηση κόστους πρέπει να διευκολύνει τους αριθμητικούς υπολογισμούς που πρέπει να γίνουν, είτε αυτό σημαίνει μεγαλύτερη ταχύτητα υπολογισμών είτε σημαίνει δημιουργία ευκολότερων υπολογισμών.

Προσαρμογή

Αφού υπολογιστούν οι αποκλίσεις στην έξοδο του νευρωνικού ακολουθεί η διαδικασία προσαρμογής του. Στόχος είναι όλες οι εκπαιδευόμενες παράμετροι του (βάρη και πόλωσεις) να πάρουν νέες τιμές, τέτοιες ώστε οι αποκλίσεις από τις αναμενόμενες τιμές να γίνουν μικρότερες. Ο αλγόριθμος που υλοποιείται για την επίτευξη αυτού του στόχου ονομάζεται οπισθοδιάδοση (Backpropagation).

Για να γίνει δυνατή η ενημέρωση των παραμέτρων του νευρωνικού, πρέπει αρχικά να μεταδοθεί η πληροφορία της απόκλισης από το τελευταίο επίπεδο του (υπολογίστηκε στο προηγούμενο βήμα) στα υπόλοιπα. Η μετάδοση αυτής της πληροφορίας γίνεται διαδοχικά από τα τελευταία επίπεδα στα πρώτα και ονομάζεται οπισθοδιάδοση των αποκλίσεων (backpropagation of errors). Κάθε νευρώνας παίρνει από τον επόμενο συνδεδεμένο σε αυτόν την απόκλιση και την πολλαπλασιάζει με το βάρος της ακμής που τους συνδέει. Σημειώστε ότι αυτή τη φορά η διάδοση της πληροφορίας έχει την αντίθετη φορά από την εμπροσθοδρόμηση. Αφού πολλαπλασιάσει όλες τις αποκλίσεις με το κάθε βάρος τις αθροίζει και μετά στο άθροισμα αυτό προσθέτει και την πόλωση. Αυτή η διαδικασία συνεχίζεται μέχρι οι αποκλίσεις να φτάσουν στα αρχικά επίπεδα του νευρωνικού.

Παράλληλα με τον υπολογισμό και τη διάδοση των αποκλίσεων λαμβάνει χώρα και ο υπολογισμός της κλίσης των αποκλίσεων (Gradient). Η κλίση των αποκλίσεων αποτελεί οδηγό για την νέα τιμή των παραμέτρων. Μεγάλη κλίση αποκλίσεων σημαίνει μεγάλη απόκλιση και άρα μεγάλη ενημέρωση στην παράμετρο. Η κλίση καταδεικνύει και την κατεύθυνση της προσαρμογής που απαιτείται για την ελαχιστοποίηση των αποκλίσεων (παράδειγμα μείωση ή αύξηση). Σημειώστε πως η κλίση αφορά τον υπολογισμό μερικών παραγώγων και την εφαρμογή του κανόνα της αλυσίδας λόγω της τοπολογικής διασύνδεσης των νευρώνων.

Αφού λάβουν χώρα τα προηγούμενα βήματα γίνεται τελικά και ο υπολογισμός της νέας τιμής των παραμέτρων. Εδώ γίνεται η χρήση της κλίσης που ελαχιστοποιεί τη συνάρτηση κόστους. Σε αυτό το σημείο σκόπιμο είναι να αναφέρουμε μια ακόμα παράμετρο, τον ρυθμό μάθησης (learning rate). Ο ρυθμός μάθησης πολλαπλασιάζεται με την υπολογισμένη παράμετρο αλλαγής των βαρών έτσι ώστε να αυξήσει η να μειώσει την απόλυτη τιμή της αλλαγής στην ενημέρωση των παραμέτρων. Μικρές τιμές του ρυθμού μάθησης οδηγούν σε αργές εκπαιδεύσεις ενώ μεγάλες σε αδυναμία εκπαίδευσης (βλέπε εικόνα 3.3). Οι μεγάλες τιμές του ρυθμού μάθησης θα οδηγούν σε μεγάλες αλλαγές στις παραμέτρους, τόσο μεγάλες που θα είναι αδύνατη η σύγκλιση στη βέλτιστη τιμή. Η σωστή επιλογή λοιπόν αυτής της υπερπαραμέτρου είναι κρίσιμη για την επιτυχή εκπαίδευση ενός νευρωνικού δικτύου.



FIGURE 3.3: Η επιρροή της επιλογής ρυθμού μάθησης στην εκπαίδευση των νευρωνικών δικτύων

Επανάληψη

Η εκπαίδευση των νευρωνικών δικτύων περιλαμβάνει την παραπάνω διαδικασία πρόβλεψης, υπολογισμού αποκλίσεων και κλίσεων και προσαρμογής επαναληπτικά. Αυτός ο κύκλος επαναλαμβάνεται για κάποιες εισόδους που θα δοθούν στο νευρωνικό και αυτός ο (υπερ)κύκλος επαναλαμβάνεται για έναν καθορισμένο αριθμό φορών που ονομάζονται εποχές (epochs) της εκπαίδευσης. Ο αριθμός των εποχών που θα εκπαιδευτεί ένα νευρωνικό δίκτυο αποτελεί μια από τις πιο κρίσιμες παραμέτρους της εκπαίδευσης. Μικρός αριθμός εποχών θα έχει σαν αποτέλεσμα το νευρωνικό να μην προλάβει να δημιουργήσει τις βέλτιστες τιμές των παραμέτρων του. Ένας πολύ μεγάλος αριθμός εποχών όμως μπορεί επίσης να επιδράσει αρνητικά στην εκπαίδευση του δικτύου. Ένα δίκτυο εκπαιδευμένο πολύ σε ένα σύνολο δεδομένων εισόδου προσαρμόζεται υπερβολικά στα συγκεκριμένα δεδομένα και χάνει τη δυνατότητα γενίκευσής του. Το φαινόμενο αυτό ονομάζεται υπερμοντελοποίηση (overfitting).

Ένα πράγμα που πρέπει να τονιστεί όσον αφορά τα παραπάνω, είναι ότι η κυκλική φύση των υπολογισμών στην εκπαίδευση ενώ είναι δεδομένη μπορεί να παρουσιάζει διαφοροποίηση σε κάποιες λεπτομέρειες της. Για παράδειγμα δεν είναι απαραίτητο η ενημέρωση των παραμέτρων να γίνεται αφού δοθούν όλα τα δεδομένα που δίνονται στην είσοδο του δικτύου. Συγκεκριμένα αυτή η προσέγγιση στους υπολογισμούς ονομάζεται Gradient Descent. Υπάρχουν πολλοί αλγόριθμοι για το πώς και πότε θα ενημερώνονται οι παράμετροι και θα παρουσιάσουμε κάποιους από αυτούς στη συνέχεια.

3.1.5 Συναρτήσεις Κόστους

Όπως προαναφέρθηκε ο ρόλος των συναρτήσεων κόστους είναι ο υπολογισμός της απόκλισης των εξόδων του νευρωνικού από τις αναμενόμενες τιμές (ground truth-labels). Υπάρχουν πολλοί τρόποι να γίνουν αυτοί οι υπολογισμοί και η σωστή επιλογή της συνάρτησης κόστους μπορεί να αποβεί κρίσιμη στην εκπαίδευση του νευρωνικού. Κάποιες συναρτήσεις κόστους είναι πιο αποδοτικές από άλλες και κάποιες συναρτήσεις κόστους μπορεί να είναι πιο κατάλληλες για συγκεκριμένα προβλήματα. Στη συνέχεια θα παρουσιάσουμε κάποιες βασικές και στο κεφάλαιο της σημασιολογικής κατάτμησης θα παρουσιάσουμε κάποιες που ενδείκνυται για τα προβλήματα της συγκεκριμένης κατηγορίας.

Mean Square Error - MSE

Το μέσο τετραγωνικό σφάλμα αποτελεί μια απλή αλλά ίσως όχι πάρα πολύ αποδοτική συνάρτηση κόστους. Υπολογίζεται ως η μέση τετραγωνική διαφορά των προβλέψεων \hat{Y}_i από τις πραγματικές Y_i τιμές:

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2, \text{ για } n \text{ αριθμό προβλέψεων} \quad (3.1)$$

Η ύψωση στο τετράγωνο γίνεται ώστε να υπολογίζεται σφάλμα ανεξάρτητο των προσήμων της διαφοράς των αποκλίσεων. Σημειώστε πως η συνάρτηση αυτή είναι δι-αφορίσιμη, γεγονός που την καθιστά χρήσιμη σε κάποιους υπολογισμούς όπως για παράδειγμα στον υπολογισμό της gradient descent. Σημειώστε επίσης ότι η ύψωση στο τετράγωνο μεγαλώνει ακόμα περισσότερο τις αποκλίσεις, καθιστώντας το μέσο τετραγωνικό σφάλμα πιο ευαίσθητο στις διαφορές.

Mean Absolute Error - MAE

Το μέσο απόλυτο σφάλμα αποτελεί επίσης μια βασική συνάρτηση κόστους. Υπολογίζεται ως η μέση απόλυτη διαφορά των προβλέψεων \hat{Y}_i από τις πραγματικές Y_i τιμές:

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i|, \text{ για } n \text{ αριθμό προβλέψεων} \quad (3.2)$$

Σημειώστε πως αυτή η συνάρτηση κόστους, σε αντίθεση με το μέσο τετραγωνικό σφάλμα, δεν είναι τόσο ευαίσθητη στις αποκλίσεις.

Cross Entropy

Η διασταυρούμενη εντροπία αποτελεί μια συνάρτηση κόστους βασισμένη στην έννοια της εντροπίας. Η εντροπία αποτελεί μια έννοια η οποία καταδικνύει το μέτρο αβεβαιότητας ενός συστήματος. Ως εκ τούτου η συνάρτηση κόστους της διασταυρούμενης εντροπίας για έναν νευρώνα δίνεται ως:

$$L_{CE} = - \sum_{i=1}^n t_i \log(p_i), \text{ για } n \text{ αριθμό κλάσεων} \quad (3.3)$$

Σαν t_i συμβολίζεται η τιμή αληθείας (truth value) της κλάσης i και σαν p η πιθανότητα ενσωμάτωσης στην κλάση i . Σημειώνουμε ότι τυπικά τα νευρωνικά δίκτυα υπολογίζουν τον μέσο όρο αυτών τιμών για όλα τα δεδομένα.

Αξίζει να μελετήσουμε συγκεκριμένα και τη δυαδική περίπτωση κλάσεων:

$$L_{CE} = - \sum_{i=1}^2 t_i \log(p_i) = -[t_1 \log(p_1) + t_2 \log(p_2)] = -[t \log(p) + (1-t) \log(1-p)] \quad (3.4)$$

Εύκολα μπορούμε να παρατηρήσουμε ότι η διασταυρούμενη εντροπία παίρνει μεγάλες τιμές όταν η πρόβλεψη ενός νευρώνα είναι λάθος στο δυαδικό πρόβλημα ($t = 0$ και $p \rightarrow 1$ ή $t = 1$ και $p \rightarrow 0$) και μικρές όταν είναι σωστή ($t = 0$ και $p \rightarrow 0$ ή $t = 1$ και $p \rightarrow 1$). Το γεγονός αυτό καθιστά την διασταυρούμενη εντροπία μια συνάρτηση κόστους ιδιαίτερα αποτελεσματική σε προβλήματα όπου η κατηγοριοποίηση των κλάσεων παρουσιάζει ανισορροπία. Άλλες συναρτήσεις κόστους θα δώσουν καλά αποτελέσματα για προβλέψεις οι οποίες τα πηγαίνουν καλά σε μια μόνο κλάση, αρκεί αυτή να αποτελεί

μια συντριπτική πλειοψηφία. Αυτό αποφεύγεται με την διασταυρούμενη εντροπία.

Προφανώς οι παραπάνω παρατηρήσεις γενικεύονται και στο πρόβλημα μεγαλύτερου αριθμού κλάσεων από 2.

3.1.6 Συναρτήσεις Ενεργοποίησης

Οι συναρτήσεις ενεργοποίησης εφαρμόζονται στην έξοδο των νευρώνων και καθορίζουν αν ένας νευρώνας θα είναι ενεργοποιημένος ή όχι. Αυτό το επιτυγχάνουν μετασχηματίζοντας τις τιμές των νευρώνων σε ένα μερικώς ή ολικώς φραγμένο σύνολο τιμών. Η χρήση τους διευρύνει σημαντικά την χρησιμότητα των νευρωνικών δικτύων καθώς εισάγουν σε αυτά μη γραμμικότητα. Όπως έχουμε αναφέρει στην ενότητα με τα perceptrons τα περισσότερα σύνθετα προβλήματα είναι μη γραμμικά. Εκτός από το μεγαλύτερο εύρος προβλημάτων οι συναρτήσεις ενεργοποίησης αυξάνουν δραματικά και την ικανότητα υπολογισμών αφού επιτρέπουν την δημιουργία σύνθετων συσχετίσεων στο πέρασμα της πληροφορίας στο εσωτερικό ενός νευρωνικού δικτύου. Στη συνέχεια θα παρουσιάσουμε κάποιες ευρέως χρησιμοποιημένες.

Sigmoid

Η Σιγμοειδής συνάρτηση παίρνει το όνομά της από το σχήμα της το οποίο θυμίζει το γράμμα S. Γνωστή και ως λογιστική συνάρτηση (logistic function), έχει σύνολο τιμών το διάστημα (0,1). Στο $+\infty$ τείνει στο 1 ενώ στο $-\infty$ τείνει στο 0. Στο 0 παίρνει την τιμή 0.5. Χρησιμοποιείται συχνά σαν συνάρτηση ενεργοποίησης στο τελευταίο επίπεδο των νευρωνικών δικτύων που κάνουν δυαδική ταξινόμηση. Έτσι παράγουν μια έξοδο η οποία λαμβάνει τιμές από το 0 μέχρι το 1 η οποία είναι μια πιθανότητα κατηγοριοποίησης όπου όσο πιο κοντά στο 0 είμαστε τόσο πιθανότερο να ανήκουμε στην κλάση 0 και αντίστοιχα για την 1.

Ορίζεται ως:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (3.5)$$

Tanh

Η Tanh συνάρτηση έχει τα ίδια ποιοτικά χαρακτηριστικά με την Σιγμοειδή με τη διαφορά ότι το σύνολο τιμών της είναι το διάστημα (-1,1).

Ορίζεται ως:

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (3.6)$$

Softmax

Η Softmax αποτελεί μια ακόμα συνάρτηση ενεργοποίησης που συναντώνται στα τελευταία επίπεδα των νευρωνικών δικτύων. Ο σκοπός της επίσης είναι να δώσει πιθανότητες κατηγοριοποίησης αλλά αυτή τη φορά σε αριθμό κλάσεων που μπορεί να ξεπερνάει τις δύο. Δέχεται σαν είσοδο ένα διάνυσμα \mathbf{z} μεγέθους K όσο τον αριθμό των προς κατηγοριοποίηση κλάσεων και παράγει ένα διάνυσμα πιθανοτήτων όπου το άθροισμά τους ισούται με τη μονάδα. Οι μεγαλύτερες τιμές θα λάβουν μεγαλύτερες πιθανότητες και οι μικρότερες αντίστοιχα. Η κάθε πιθανότητα στο διάστημα εξόδου αντιστοιχεί στην εκτίμηση κατηγοριοποίησης του νευρωνικού στην κάθε κλάση.

Ορίζεται ως:

$$\sigma(\mathbf{z})_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}}, \text{ για } i = 1, 2, \dots, K \text{ και } \mathbf{z} = (z_1, z_2, \dots, z_K) \in \mathbb{R}^K \quad (3.7)$$

Rectified Linear Unit - Relu

Η Relu αποτελεί μια συνάρτηση ενεργοποίησης η οποία βρίσκει εφαρμογή στους νευρώνες των εσωτερικών επιπέδων των δικτύων. Ρόλος της είναι η ενεργοποίηση ή μη των νευρώνων. Λειτουργεί αφήνοντας απείραχτες θετικές τιμές και κόβοντας τις αρνητικές (γυρνάει 0). Η χρήση τους έχει αυξηθεί λόγω της απλότητάς τους και της προσφοράς τους στην καταπολέμηση του προβλήματος των εξαφανιζόμενων κλίσεων (vanishing gradient problem).

Ορίζεται ως:

$$f(x) = \begin{cases} x, & \text{Αν } x > 0 \\ 0, & \text{Αλλιώς} \end{cases} \quad (3.8)$$

3.1.7 Βελτιστοποιητές

Ο τρόπος με τον οποίο θα εκπαιδευτεί το δίκτυο μπορεί να έχει μεγάλη επίδραση στα αποτελέσματά του. Όπως προαναφέρθηκε, θα μπορούσε μια επανάληψη για την ενημέρωση των παραμέτρων να αφορά το πέρασμα όλων των διαθέσιμων εισόδων. Θα μπορούσε επίσης να γίνει σε ένα υποσύνολο αυτών. Οι βελτιστοποιητές (optimizers) είναι οι αλγόριθμοι οι οποίοι καθορίζουν αυτές τις λεπτομέρειες στον τρόπο εκπαίδευσης των δικτύων και η προσεκτική επιλογή τους αποτελεί κρίσιμη παράμετρο στην εκπαίδευση των νευρωνικών δικτύων.

Gradient Descent

Η Gradient Descent αποτελεί τον στοιχειώδη αλγόριθμο για την εκπαίδευση των νευρωνικών δικτύων. Στην Gradient Descent τα βάρη ενημερώνονται αφού το δίκτυο εκτεθεί σε όλες τις διαθέσιμες εισόδους, στο τέλος του κύκλου που ονομάζουμε εποχή. Πιο συγκεκριμένα τα βήματα στην Gradient Descent αφορούν τα εξής:

1. Forward pass. Το δίκτυο προβλέπει την έξοδό του για μια συγκεκριμένη είσοδο.
2. Backward pass. Το δίκτυο υπολογίζει από το τέλος προς την αρχή τις αποκλίσεις και τις κλίσεις τους.
3. Συνυπολογισμός κλίσεων. Για κάθε εικόνα συνυπολογίζονται οι κλίσεις (άθροισμα, μέσος όρος κλπ) έτσι ώστε το δίκτυο να βρεί μια τιμή κλίσεων η οποία θα είναι βέλτιστη για το σύνολο των εισόδων που δόθηκαν.
4. Ενημέρωση των βαρών. Αφού εκτελεστούν τα παραπάνω βήματα για όλες τις εισόδους, υπολογίζονται τα βάρη με βάση τις προηγούμενες τιμές τους και την σταθμισμένη (συνυπολογισμένη) κλίση ως προς τις αποκλίσεις τους με σκοπό την δημιουργία νέων τιμών οι οποίες θα παράξουν μικρότερες αποκλίσεις.

Η Gradient Descent αν και αποτελεί τον στοιχειώδη αλγόριθμο για την εκπαίδευση των νευρωνικών δικτύων πάσχει από αδυναμίες. Η πρώτη και πιο προφανής αποτελεί τον ιδιαίτερα μεγάλο αριθμό υπολογισμών που δημιουργούνται. Δεν είναι απαραίτητη η επαναληπτική τροφοδότηση με όλα τα δεδομένα εισόδου, ούτε εφικτή σε περιπτώσεις όπου τα δεδομένα εισόδου είναι πάρα πολλά.

Η δεύτερη αδυναμία της Gradient Descent αφορά την αποτυχία σύγκλισης στη βέλτιστη ελάχιστη τιμή της συνάρτησης κόστους, σε πολλές περιπτώσεις. Οι συναρτήσεις κόστους σε πραγματικά προβλήματα (βλέπε σχήμα 3.4) καταλήγουν να έχουν γραφικές παραστάσεις οι οποίες παρουσιάζουν πλειάδα τοπικών ελαχίστων και saddle points (non convex functions). Η κλίση σε αυτά τα σημεία πράγματι προσεγγίζει το μηδέν, αλλά χωρίς να καταδεικνύει την κατεύθυνση για το πραγματικό σημείο βέλτιστης ελαχιστοποίησης της συνάρτησης κόστους. Η εκπαίδευση τελικά θα επιφέρει αποτελέσματα προσέγγισης αυτών των μη βέλτιστων σημείων. Καταλαβαίνουμε λοιπόν ότι η Gradient Descent δεν μπορεί να εγγυηθεί την σύγκλιση στο πραγματικά ελάχιστο σημείο της συνάρτησης κόστους.

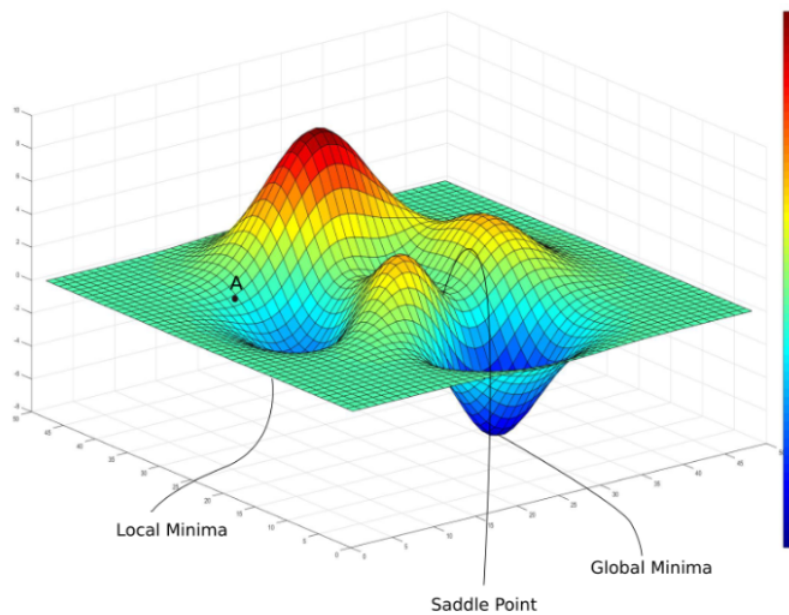


FIGURE 3.4: Τα κρίσιμα σημεία μιας συνάρτησης για την εφαρμογή της Gradient Descent

Stochastic Gradient Descent

Για την πιο αποτελεσματική αντιμετώπιση των προβλημάτων που προέκυψαν από την εφαρμογή της Gradient Descent, μπορούμε να εισάγουμε στοχαστικές μεθόδους στην επιλογή των δειγμάτων. Πιο συγκεκριμένα, θα αποφύγουμε την χρήση όλων των δεδομένων εισόδου και θα επιλέγουμε τυχαία κάποιο ή κάποια από αυτά (mini batch Stochastic Gradient Descent).

Η χρήση ενός υποσυνόλου της εισόδου για την εκπαίδευση αρχικά μειώνει δραματικά τον αριθμό των υπολογισμών που απαιτείται να γίνουν για τις κλίσεις. Αντιμετωπίζεται το πρόβλημα επίσης της παλινδρόμησης στα τοπικά ελάχιστα της συνάρτησης κόστους. Η εισαγωγή της τυχαιότητας (βλέπε σχήμα 3.5) συνεπάγεται την χαρτογράφηση σε διαφορετικά σημεία της συνάρτησης κόστους, με αποτέλεσμα την κάθοδο χωρίς κίνδυνο εγκλωβισμού σε τοπικά ελάχιστα.

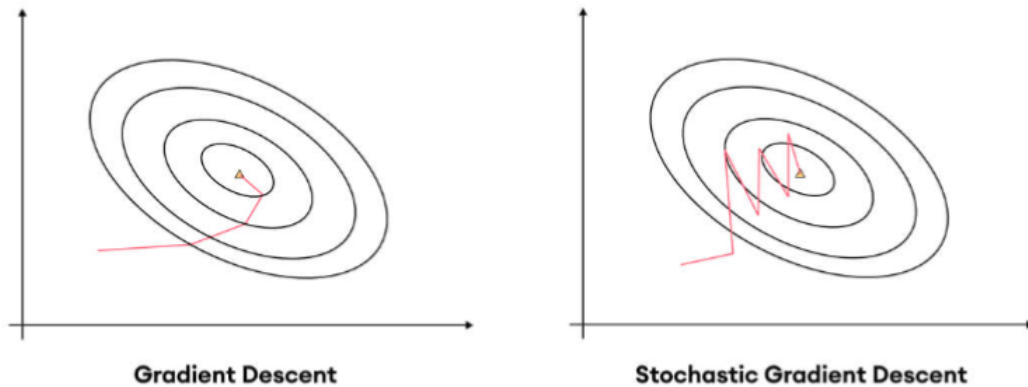


FIGURE 3.5: Η Stochastic Gradient Descent προσεγγίζει με στοχαστικές μεθόδους την ελαχιστοποίηση της συνάρτησης κόστους

Βελτιστοποιητές με χρήση τεχνικών Momentum

Για την εξασφάλιση γρηγορότερης σύγκλισης μπορεί να γίνει χρήση τεχνικών Momentum. Σε αυτές, η ενημέρωση των παραμέτρων γίνεται συνυπολογίζοντας το μέτρο της αλλαγής της προηγούμενης ενημέρωσης. Έτσι μια κλήση η οποία θα επιφέρει μεγάλη αλλαγή θα συνεισφέρει και σε μελλοντικές ενημερώσεις και τελικά η συνεισφορά όλων των αλλαγών αυτών θα συμβάλει σε συνολικά γρηγορότερη σύγκλιση (βλέπε σχήμα 3.6). Τα νέα βάρη υπολογίζονται ως:

$$w = w - \eta \nabla Q_i(w) + \alpha \Delta w \quad (3.9)$$

Ο όρος $w - \eta \nabla Q_i(w)$ αποτελεί τον κλασικό όρο υπολογισμού των βαρών που εφαρμόζεται με τον ίδιο τρόπο και στις Gradient Descent και Stochastic Gradient Descent. Το η είναι ο ρυθμός εκπαίδευσης και Q_i η i -οστή παρατήρηση της συνάρτησης κόστους (για το i -οστό δείγμα). Ο όρος $\alpha \Delta w$ αποτελεί την συνεισφορά της προηγούμενης αλλαγής (momentum). Το α αποτελεί μια παράμετρο η οποία καθορίζει το ποσό της συνεισφοράς που επιθυμούμε να έχει το momentum.

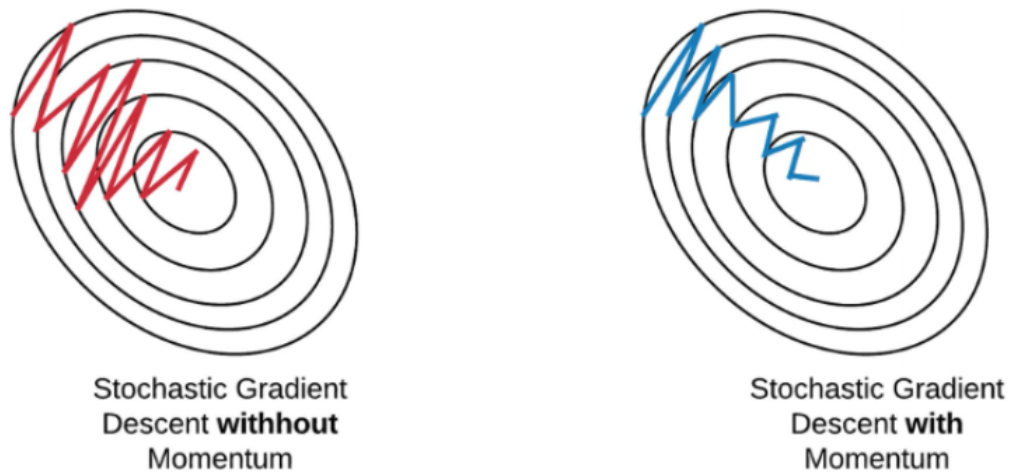


FIGURE 3.6: Οι τεχνικές momentum μπορούν να συμβάλλουν σε γρηγορότερη σύγκλιση

Adaptive Βελτιστοποιητές

Οι Adaptive βελτιστοποιητές προσαρμόζουν την συνεισφορά του ρυθμού μάθησης για κάθε παράμετρο, συνυπολογίζοντας την παρελθοντική συνεισφορά καιθεμίας από αυτές. Ο υπολογισμός αυτής της συνεισφοράς μπορεί να γίνει με διάφορους τρόπους (Adagrad, Adadelata κλπ), αλλά πάντα το τελικό αποτέλεσμα είναι η γρηγορότερη σύγκλιση λόγω της καλύτερης αξιοποίησης των σημαντικών παραμέτρων στην εκμάθηση (βλέπε εικόνα 3.7). Οι adaptive βελτιστοποιητές παρουσιάζουν για παράδειγμα καλύτερες αποδόσεις σε προβλήματα κατηγοριοποίησης κλάσεων όπου κάποιες από αυτές είναι σημαντικά λιγότερο πολυπληθείς. Πιο συγκεκριμένα, θα μείωναν την συνεισφορά της πολυπληθής κλάσης και θα αύξαναν την συνεισφορά της λιγότερο πολυπληθής κλάσης. Έτσι θα εξασφαλίζεται και η καλύτερη σύγκλιση αφού αλλιώς η συνεισφορά της λιγότερο πολυπληθής κλάσης θα χανόταν στους συνολικούς υπολογισμούς. Στη συνέχεια θα παρουσιάσουμε ένα παράδειγμα adaptive βελτιστοποιητή, αυτό του Adagrad (Adaptive Gradient).

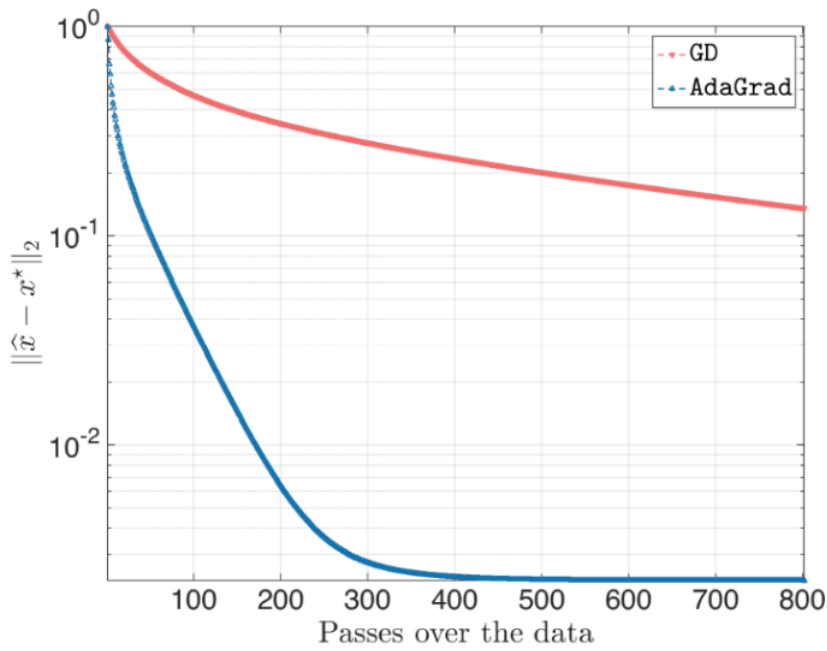


FIGURE 3.7: Ο AdaGrad συγκλίνει πολύ πιο γρήγορα από την Gradient Descent

Στον Adagrad τροποποιείται ο ρυθμός μάθησης με τον εξής τρόπο:

$$\eta'_i = \frac{\eta}{\sqrt{G_{t,ii} + \varepsilon}}, \text{ για την παράμετρο } i \quad (3.10)$$

Ο όρος ε έχει σαν ρόλο την αποτροπή του μηδενισμού του παρονομαστή και συνήθως είναι της τάξης 10^{-8} . Παρατηρούμε λοιπόν, ότι ο ρυθμός μάθησης είναι διαφορετικός για την κάθε παράμετρο.

Ο όρος $G_{t,ii}$ αποτελεί έναν ορθογώνιο πίνακα όπου το κάθε διαγώνιο στοιχείο του στη θέση (i,i) αποτελεί το άθροισμα των τετραγώνων των κλίσεων ως προς w_i για κάθε χρονική στιγμή μέχρι την t :

$$G_{t,ii} = \nabla_{w_{1,i}}^2 Q(w_{1,i}) + \nabla_{w_{2,i}}^2 Q(w_{2,i}) + \dots + \nabla_{w_{t,i}}^2 Q(w_{t,i}) \quad (3.11)$$

Οι παράμετροι w_i τελικά, στο βήμα t υπολογίζονται ως:

$$w_{t+1,i} = w_{t,i} - \frac{\eta}{\sqrt{G_{t,ii} + \varepsilon}} \nabla Q_{i,t}(w) \quad (3.12)$$

Adam

Ο αλγόριθμος Adam (Adaptive Moment Estimation) αποτελεί μια προσέγγιση που συνδυάζει τις ιδέες του momentum και των adaptive βελτιστοποιητών. Αυτό το επιτυγχάνει αποθηκεύοντας τον εκθετικά μειούμενο μέσο προηγούμενων κλίσεων m_t (momentum) όπως και τον εκθετικά μειούμενο μέσο των τετραγωνισμένων κλίσεων u_t (adaptive).

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) \nabla Q(t) \quad (3.13)$$

$$u_t = \beta_2 u_{t-1} + (1 - \beta_2) (\nabla Q(t))^2 \quad (3.14)$$

Ο υπολογισμός των εκθετικά μειούμενων μέσων τετραγωνισμένων κλίσεων αποτελεί μια παραλλαγή του υπολογισμού των τετραγωνικών κλίσεων που υλοποιείται στον `adagrad` βελτιστοποιητή που παρουσιάσαμε με σκοπό την αποφυγή του φαινομένου της σύγκλισης στο 0 για πολλά βήματα του προσαρμοζόμενου ρυθμού μάθησης.

Τα β_1 και β_2 αποτελούν παράγοντες ελέγχου του ποσού της συνεισφοράς του momentum και του adaptation αντίστοιχα.

Τα m_t και u_t πάντως υπολογίζονται με μια τροποποίηση. Καθώς αρχικοποιούνται σαν μηδενικά διανύσματα τείνουν να δίνουν τιμές κοντά στο 0, ειδικά όταν τα β_1 και β_2 είναι κοντά στο 1. Η τροποποίηση αυτή λοιπόν είναι η:

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad (3.15)$$

$$\hat{u}_t = \frac{u_t}{1 - \beta_2^t} \quad (3.16)$$

Τελικά οι παράμετροι υπολογίζονται ως:

$$w_{t+1} = w_t - \frac{\eta}{\sqrt{\hat{u}_t} + \epsilon} \hat{m}_t \quad (3.17)$$

Η τροποποίηση των m_t και u_t σε συνδυασμό με προτεινόμενες τιμές $\beta_1 = 0.9$, $\beta_2 = 0.999$ και $\epsilon = 10^{-8}$, έχει δείξει εμπειρικά ότι αποφέρει πολύ καλά αποτελέσματα.

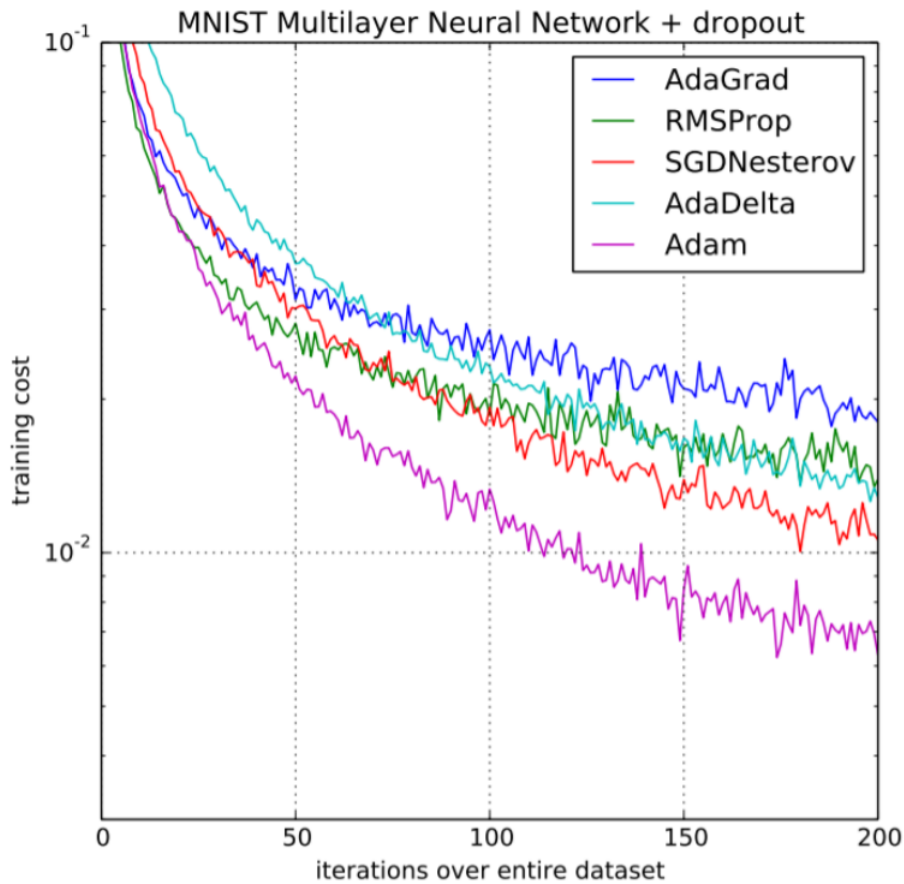


FIGURE 3.8: Ο Adam μέχρι στιγμής θεωρείται εν γέννη ο καλύτερος βελτιστοποιητής

3.2 Νευρωνικά Δίκτυα και Βαθιά Μάθηση με Εικόνες

3.2.1 Εισαγωγή

Από την σύλληψη των perceptron, διάφορες απόπειρες αναγνώρισης εικόνων είχαν δοκιμαστεί. Οι προσπάθειες αυτές όμως δεν είχαν οδηγήσει κάπου καθώς η πρόωμη γνώση εκείνης της εποχής δεν ήταν αρκετή για να δώσει αποτελέσματα σε ένα τόσο δύσκολο πρόβλημα. Η πρώτη σημαντική συνεισφορά στο πεδίο της αναγνώρισης εικόνας ήρθε το 1980 από τον Kunihiko Fukushima στο έργο του *"Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position"* (Fukushima, 1980). Το Neocognitron του Fukushima εισήγαγε τις έννοιες των συνελικτικών επιπέδων, του downsampling και του weight sharing. Και οι τρεις αυτές έννοιες είναι κεντρικές στη βαθιά μάθηση με εικόνες. Αναφέρουμε επίσης ότι ο Fukushima εφηύρε την relu, την οποία περιγράψαμε παραπάνω, το 1969 *"Visual feature extraction by a multilayered network of analog threshold elements"* (Fukushima, 1969). Όπως και το perceptron, οι συνεισφορές του Fukushima ήταν εμπνευσμένες από την βιολογική νευρολογία, συγκεκριμένα από το έργο των νευροφυσιολόγων David Hubel και Torsten Wiesel και τη μελέτη τους στους οπτικούς φλοιούς.

Ένα άλλο, επίσης σημαντικό ορόσημο αποτέλεσε το Lenet-5 *"Gradient-Based Learning Applied to Document Recognition"* (Lecun et al., 1998), ένα συνελικτικό νευρωνικό δίκτυο το οποίο χρησιμοποίησε τις ιδέες που προαναφέρθηκαν αλλά με τον αλγόριθμο backpropagation για την εκπαίδευσή του. Το Lenet-5, εκπαιδευμένο στο MNIST dataset, κατάφερε επιτυχημένα να αναγνωρίσει χειρόγραφα ψηφία. Τα επόμενα περίπου 10 χρόνια αποτέλεσαν μια περίοδο στασιμότητας στην χρήση των νευρωνικών δικτύων σε εφαρμογές αναγνώρισης μοτίβων σε εικόνες. Οι κύριοι λόγοι για το φαινόμενο αυτό ήταν η έλλειψη δυνατών υπολογιστικών μηχανών και η έλλειψη ενός επαρκούς dataset το οποίο θα έπρεπε να είναι αρκετά μεγάλο έτσι ώστε τα μοντέλα να μπορούν πραγματικά να μάθουν. Άλλες τεχνικές μηχανικής μάθησης όπως τα Support Vector Machines ήταν πιο διαδεδομένες.

Οι πραγματικές αλλαγές στο χώρο έμελλαν να ξεκινήσουν από το 2009 και μετά όπου δημοσιεύτηκε το ImageNet στο Conference on Computer Vision and Pattern Recognition (CVPR). Το ImageNet αποτελεί ένα dataset το οποίο περιέχει πάνω από 14 εκατομμύρια εικόνες σε 20.000 κατηγορίες. Το ImageNet project από το 2010 και μετά οργανώνει κάθε χρόνο το ImageNet Large Scale Visual Recognition Challenge (ILSVRC), έναν διαγωνισμό κατηγοριοποίησης εικόνων. Στον ILSVRC λοιπόν, το 2012 συμμετείχε το AlexNet *"ImageNet Classification with Deep Convolutional Neural Networks"* (Krizhevsky, Sutskever, and Hinton, 2012). Το AlexNet κέρδισε τον διαγωνισμό με πολλή μεγάλη διαφορά από τον δεύτερο διαγωνιζόμενο, επιφέροντας ριζική αλλαγή στον κλάδο. Μετά το AlexNet το ενδιαφέρον για την χρήση των νευρωνικών δικτύων σε εφαρμογές αναγνώρισης εικόνας εκτοξεύθηκε και τα επόμενα χρόνια ακολούθησε πλειάδα υλοποιήσεων και ερευνών, με αποτελέσματα εντυπωσιακά. Η αρχιτεκτονική του AlexNet δεν ήταν ριζικά επαναστατική. Θύμιζε πολύ αυτή του Lenet-5. Η διαφορά ήταν στο μέγεθος του δικτύου. Το AlexNet απέδειξε ότι για πολλές εφαρμογές, το μέγεθος των δικτύων είναι κρίσιμος παράγοντας. Έτσι άρχισε να χρησιμοποιείται συχνά και ο όρος βαθιά μάθηση (deep learning). Ένα κρίσιμο χαρακτηριστικό που είχε το AlexNet και του επέτρεψε να έχει μεγάλο μέγεθος ήταν οι υπολογιστικοί πόροι που χρησιμοποιήθηκαν για τους υπολογισμούς του. Το AlexNet γράφτηκε σε CUDA και έτρεχε σε GPU. Έκτοτε κάθε βαθύ νευρωνικό δίκτυο τρέχει σε υπολογιστικές μονάδες οι οποίες επιτρέπουν με κάποιο τρόπο μεγαλύτερο παραλληλισμό στους υπολογισμούς. Εκτός από GPUs διαδεδομένη χρήση γίνεται και των TPUs (Tensor Processing Unit),

οι οποίες αναπτύχθηκαν ειδικά για τον σκοπό υπολογισμών νευρωνικών δικτύων. Ερευνητικά, αλλά όχι ακόμα πολύ διαδεδομένα, μπορεί να γίνει χρήση και άλλων μονάδων επεξεργασίας όπως των FPGA (Field Programmable Gate Arrays).

3.2.2 Κεντρική ιδέα

Η αναγνώριση προτύπων σε εικόνες ενέχει ιδιαίτερες δυσκολίες. Οι εικόνες αποτελούν δεδομένα μεγάλου μεγέθους και πολλών διαστάσεων. Το γεγονός αυτό μπορεί να οδηγήσει σε πολυάριθμους και σύνθετους υπολογισμούς. Ένα άλλο πρόβλημα έγκειται στη φύση της αναπαράστασης των εννοιών που εμπεριέχονται στις εικόνες. Μια εικόνα που αναπαριστά ένα αντικείμενο, στην πραγματικότητα αναπαριστά μια χωροταξική προβολή αυτού του αντικειμένου τη στιγμή που δημιουργήθηκε η εικόνα. Οι χωροταξικές προβολές του ίδιου αντικειμένου μπορούν να είναι πολυάριθμες. Μια άλλη εικόνα για παράδειγμα μπορεί να το δείχνει μεγαλύτερο, περιστραμμένο ή με άλλη φωτεινότητα. Μπορεί επίσης να υπάρχουν διαφορετικά αντικείμενα σε διαφορετικές εικόνες που όμως συμβολίζουν το ίδιο μοτίβο (παράδειγμα πολλές γάτες). Τα νευρωνικά δίκτυα πρέπει επομένως να μπορούν να αναγνωρίζουν χαρακτηριστικά ανεξάρτητα της μοναδικής στην εικόνα γεωμετρίας τους (spatial invariance), να μπορούν να αναγνωρίζουν από συγκεκριμένα χαρακτηριστικά όπως τις ακμές, μέχρι πιο αφηρημένα όπως ένα τραπέζι, ή έναν άνθρωπο (Hierarchical Representation Learning - HARP) και όλα αυτά σε μεγάλες και σύνθετες εισόδους. Έννοιες όπως τα συνελκτικά επίπεδα και το downsampling εισήχθησαν για να επιλύσουν αυτά τα προβλήματα και θα τις παρουσιάσουμε στη συνέχεια.

3.2.3 Εξαγωγή χαρακτηριστικών με χρήση Συνελκτικών Νευρωνικών Δικτύων

Το σύνθετο πρόβλημα της αναγνώρισης μιας εικόνας μπορεί να απλοποιηθεί αν αναλογιστούμε ότι οποιαδήποτε εικόνα αποτελεί εν δυνάμει μια σύνθεση απλούστερων διακεκριμένων χαρακτηριστικών. Τα χαρακτηριστικά αυτά μπορεί να παρουσιάζουν διαφορετικό επίπεδο αφάιρησης αλλά σίγουρα αποτελούν δομικά στοιχεία πολλών εικόνων. Ο κύκλος για παράδειγμα αποτελεί δομικό στοιχείο όλων των ματιών και τα μάτια δομικό στοιχείο όλων των ανθρώπων. Ένα νευρωνικό λοιπόν δεν χρειάζεται να βλέπει μια εικόνα ενός ανθρώπου σαν ένα τελείως διαφορετικό πρόβλημα από μια εικόνα μιας γάτας. Μπορεί να μάθει να αναγνωρίζει συγκεκριμένα χαρακτηριστικά και το πώς αυτά συντίθενται για να παράξουν μια μοναδική εικόνα.

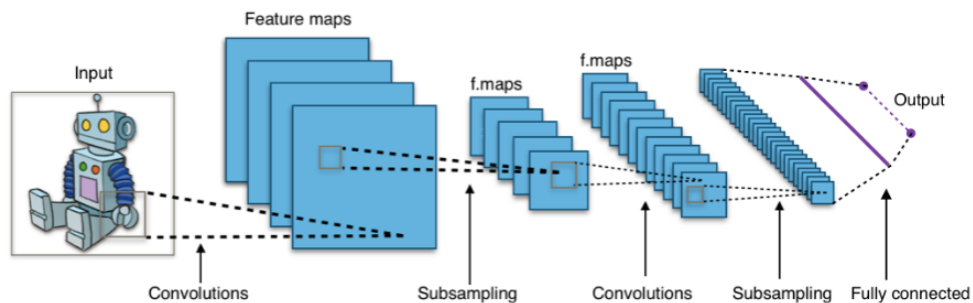


FIGURE 3.9: Η βασική αρχιτεκτονική των CNN. Γίνεται εισαγωγή συνελκτικών και subsampling επιπέδων

Συνελικτικά επίπεδα

Στόχος των συνελικτικών επιπέδων είναι να παράξουν από μια εικόνα ένα σύνολο χαρακτηριστικών. Αυτό το επιτυγχάνουν συγκρίνοντας χαρακτηριστικά της εικόνας με ένα προκαθορισμένο σύνολο χαρακτηριστικών. Η έξοδος των συνελικτικών επιπέδων (feature maps) αποτελεί μια ένδειξη ταύτισης της εικόνας με το κάθε προκαθορισμένο χαρακτηριστικό. Ο αριθμός των προκαθορισμένων χαρακτηριστικών ταυτίζεται λοιπόν με τον αριθμό των feature maps και καθορίζεται από την αρχιτεκτονική του νευρωνικού. Το ποιά θα είναι αυτά τα χαρακτηριστικά καθορίζονται από τα φίλτρα (filters ή kernels). Τα φίλτρα αποτελούν εκπαιδευσιμες παραμέτρους των δικτύων και είναι αυτά τα οποία εκτελούν την συνέλιξη με την εικόνα, ώστε να αναδειχθεί το επίπεδο ταύτισης της εικόνας με το χαρακτηριστικό που αναπαρίσταται από αυτά τα φίλτρα. Η συνέλιξη γίνεται σε βήματα όπου υπολογίζεται το εσωτερικό γινόμενο κομματιών της εικόνας με τα φίλτρα. Το μέγεθος του φίλτρου και του κομματιού της εικόνας πρέπει να έχουν την ίδια διάσταση και ονομάζεται receptive field. Σημειώστε σε αυτό το σημείο ότι οι τιμές των φίλτρων για το ίδιο feature map είναι οι ίδιες (parameter sharing) πάνω στην υπόθεση ότι το ίδιο χαρακτηριστικό αναζητείται σε όλα τα μέρη της εικόνας από το ίδιο feature map. Αυτό το χαρακτηριστικό αποτελεί σημαντική παράμετρο στην αποδοτικότητα των δικτύων.

Η πράξη της συνέλιξης καταλήγει να είναι ο αχρογωνιαίος λίθος των συνελικτικών νευρωνικών δικτύων. Μπορεί να αποτελέσει ένα μέτρο ταύτισης δύο σημάτων. Ορίζεται ως:

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} * \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mn} \end{bmatrix} = \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} x_{(m-i)(n-j)} y_{(1+i)(1+j)} \quad (3.18)$$

Ένα παράδειγμα υπολογισμού συνέλιξης παρουσιάζεται στην εικόνα 3.10.

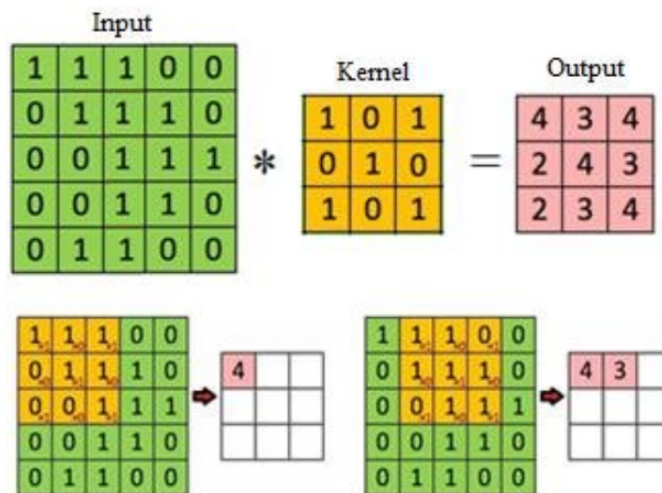


FIGURE 3.10: Ένα παράδειγμα συνέλιξης με είσοδο έναν πίνακα 5x5 και ένα φίλτρο 3x3

Στην εικόνα 3.11 παραθέτουμε και ένα παράδειγμα συνέλιξης μιας εικόνας με ένα φίλτρο αναγνώρισης καθέτων ακμών:

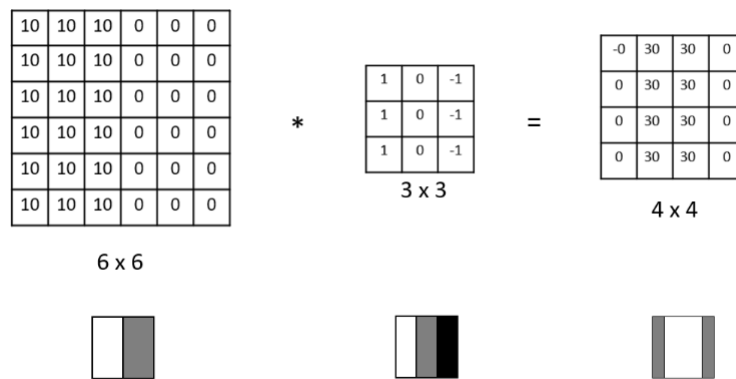


FIGURE 3.11: Ένα παράδειγμα συνέλιξης με είσοδο έναν πίνακα 6x6 και ένα ειδικό φίλτρο 3x3 για αναγνώριση ακμών

Downsampling

Το downsampling αποτελεί μια τεχνική η οποία βοηθάει τα νευρωνικά δίκτυα να αναγνωρίζουν πληροφορίες πιο αφαιρετικού επιπέδου. Πιο αφαιρετικού επιπέδου θα μπορούσε να σημαίνει πληροφορίες ανεξάρτητες της χωρικής κατανομής, γεωμετρικών διαστρεβλώσεων αλλά και πρωτόγονων χαρακτηριστικών όπως ακμές. Αυτό το επιτυγχάνει με μια ίσως αναπάντεχα απλή λειτουργία: αφαιρεί μέρη της εισόδου. Δίνει έτσι στα δίκτυα μεγάλες δυνατότητες γενίκευσης, ενώ παράλληλα τα απελευθερώνει και από μεγάλο φόρτο υπολογισμών. Οι λειτουργίες που ακολουθούνται ονομάζονται Pooling. Συνήθως χρησιμοποιούνται δύο είδη Pooling, το max pooling και το average pooling.

Max pooling

Σε αυτό η εικόνα διασπάται σε μέρη όπου από το κάθε μέρος επιλέγεται μόνο η μέγιστη τιμή. Η διατήρηση μόνο των μεγάλων τιμών έχει σαν αποτέλεσμα την διατήρηση σημαντικών χαρακτηριστικών όπως πιο έντονες ακμές ή γενικά πιο ευδιάκριτων χαρακτηριστικών.

Average Pooling

Σε αυτό η εικόνα διασπάται σε μέρη όπου από το κάθε μέρος επιλέγεται μόνο η μέση τιμή. Αυτό έχει σαν αποτέλεσμα την εξομάλυνση των χαρακτηριστικών που διατηρούνται. Ο θόρυβος αν δεν είναι συνεχής μπορεί να απαλειφθεί καλύτερα με αυτή τη τεχνική. Σημειώστε όμως ότι ενδέχεται να χαθούν σημαντικά χαρακτηριστικά με αυτή τη μέθοδο.

3.2.4 Παράδειγμα: Lenet-5

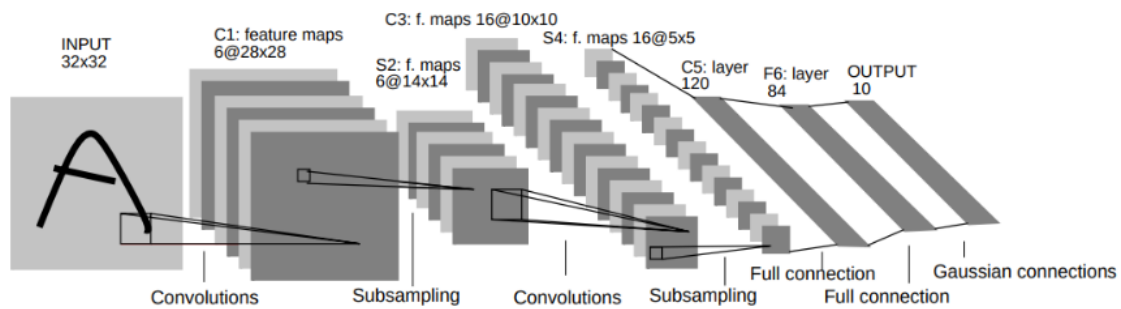


FIGURE 3.12: Η αρχιτεκτονική του Lenet-5

Το Lenet-5 έθεσε τα θεμέλια της κλασικής αρχιτεκτονικής των συνελικτικών νευρωνικών δικτύων. Δημιουργήθηκε για την αναγνώριση χειρόγραφων αριθμητικών ψηφίων (0-9) και εκπαιδεύτηκε στο MNIST dataset (βλέπε εικόνα 3.13). Η αρχιτεκτονική του αποτελείται από δύο συνελικτικά επίπεδα που ακολουθούνται από subsampling, ένα fully connected επίπεδο και ένα τελικό fully connected με συνάρτηση ενεργοποίησης τη softmax. Το subsampling δεν υλοποιείται με κλασικές pooling τεχνικές (max, average). Αντ' αυτού υλοποιείται μια τεχνική (κοντά στο average pooling) όπου 4 είσοδοι προστίθενται, μετα πολλαπλασιάζονται με μια εκπαιδύσιμη παράμετρο και μετά προστίθενται επίσης με μια εκπαιδύσιμη παράμετρο, όπου τέλος γίνεται χρήση και της softmax σαν συνάρτηση ενεργοποίησης. Σαν συνάρτηση ενεργοποίησης γίνεται χρήση της tanh, εκτός από τη softmax στα σημεία που αναφέρθηκαν.



FIGURE 3.13: Κάποιες εικόνες από το MNIST Dataset

TABLE 3.1: Χαρακτηριστικά του Lenet-5

Layer	Layer	Feature Map	Size	Kernel size	Stride	Activation
input	image	1	32x32	-	-	-
1	convolution	6	28x28	5x5	1	tanh
2	average pooling	6	14x14	2x2	2	tanh
3	convolution	16	10x10	5x5	1	tanh
4	average pooling	16	5x5	2x2	2	tanh
5	convolution	120	1x1	5x5	1	tanh
6	FC	-	84	-	-	tanh
output	FC	-	10	-	-	softmax

3.2.5 Παράδειγμα: AlexNet

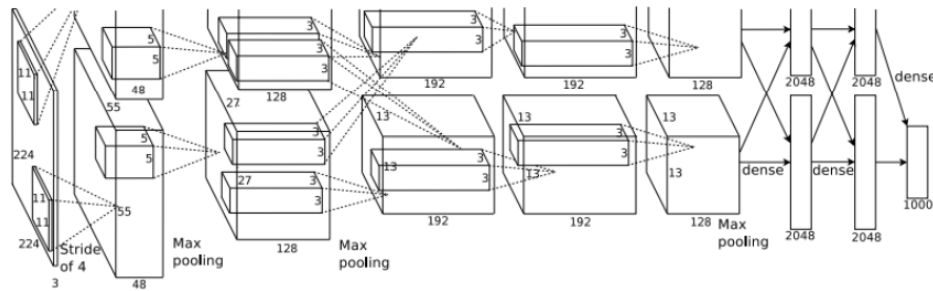


FIGURE 3.14: Η αρχιτεκτονική του AlexNet

Το AlexNet δεν ήταν το πρώτο συνελκτικό δίκτυο που δημιουργήθηκε αλλά ήταν το πρώτο που με την αρχιτεκτονική του και τη χρήση GPU κατάφερε να επιτύχει πολύ καλά αποτελέσματα και τελικά να ανοίξει τον δρόμο για την ευρεία έρευνα και χρήση των νευρωνικών δικτύων σε εφαρμογές όρασης υπολογιστών. Εισηγήαγε ή διέδωσε πολλές τεχνικές οι οποίες σήμερα θεωρούνται καθιερωμένες στα νευρωνικά δίκτυα.

Οι τεχνικές αυτές ήταν:

GPU για παραλληλοποίηση υπολογισμών. Στην εικόνα που παρατίθεται της αρχιτεκτονικής του AlexNet βλέπουμε ότι η αρχιτεκτονική του χωρίζεται σε δύο μέρη, το πάνω και το κάτω, τα οποία λειτουργούν σε διαφορετικές GPU. Αυτή η αρχιτεκτονική επέτρεψε καλύτερη αξιοποίηση των δυνατοτήτων των GPU, παρακάμπτοντας έτσι περιορισμούς τους όπως τη μικρή RAM που τουλάχιστον τότε είχαν.

ReLU σαν συνάρτηση ενεργοποίησης. Έγινε χρήση της ReLU αντί της tanh η οποία επέτρεψε πολύ γρηγορότερη εκπαίδευση του δικτύου.

Overlapping Pooling. Αποδείχθηκε ότι το overlapping pooling βοήθησε στην καλύτερη αποφυγή του overfitting.

Dropout. Αποτελεί μια τεχνική για την τυχαία απενεργοποίηση νευρώνων, οι οποία επίσης βοηθάει στην αποφυγή του overfitting.

Data augmentation. Αποτελεί μια τεχνική όπου με μετασχηματισμούς σε εικόνες δημιουργούνται περισσότερα παραδείγματα εισόδων. Το δίκτυο έτσι έχει περισσότερες εισόδους στην εκπαίδευσή του και αποφεύγεται το overfitting.

TABLE 3.2: Χαρακτηριστικά του AlexNet

Layer	Filter Number	Feature Map	Kernel Size	Stride	Padding	Activation
input	-	227x227x3	-	-	-	-
Conv 1	96	55x55x96	11x11	4	-	ReLU
Max Pool 1	-	27x27x96	3x3	2	-	-
Conv 2	256	27x27x256	5x5	1	2	ReLU
Max Pool 2	-	13x13x256	3x3	2	-	-
Conv 3	384	13x13x384	3x3	1	1	ReLU
Conv 4	384	13x13x384	3x3	1	1	ReLU
Conv 5	256	13x13x256	3x3	1	1	ReLU
Max Pool 3	-	6x6x256	3x3	2	-	-
Dropout 1	rate = 0.5	6x6x256	-	-	-	-

3.3 Σημασιολογική κατάτμηση με Νευρωνικά Δίκτυα

3.3.1 Εισαγωγή

Η σημασιολογική κατάτμηση αφορά το πρόβλημα της κατηγοριοποίησης σε κλάσεις (σημασιολογικές κλάσεις) μιας εικόνας στο επίπεδο των εικονοστοιχείων της. Αυτό σημαίνει ότι σε κάθε εικονοστοιχείο μιας εικόνας αποδίδεται μια κλάση όπου η κάθε κλάση ορίζεται από εμάς να έχει μια σημασία. Σε αντιδιαστολή λοιπόν με τα νευρωνικά δίκτυα που κατηγοριοποιούν μια εικόνα σε μια κλάση, τα δίκτυα σημασιολογικής κατάτμησης κατηγοριοποιούν όλα τα περιεχόμενα μιας εικόνας σε ένα σύνολο σημασιολογικών κλάσεων. Αυτή η φύση τους καθιστά τα συγκεκριμένα δίκτυα ικανά να χωρίσουν μια εικόνα σε περιοχές ενδιαφέροντος (κατάτμηση). Οι εφαρμογές τους είναι πολυάριθμες. Από την αυτόνομη οδήγηση, τη ρομποτική ακόμα και στην ιατρική απεικόνιση.

Η σημασιολογική κατάτμηση αποτελεί ένα πολύ σημαντικό πρόβλημα και απόπειρες επίτευξής του είχαν γίνει πριν την εκτεταμένη διάδοση των νευρωνικών δικτύων. Η πιο βασική από αυτές αποτελεί την κατάτμηση χρησιμοποιώντας κατωφλίωση (thresholding) με βάση κάποιο κριτήριο όπως το χρώμα, την ένταση ή την υφή των επιμέρους στοιχείων της εικόνας για την κατάτμηση της εικόνας σε background και foreground. Έχουν δημιουργηθεί μάλιστα και αλγόριθμοι αυτόματης εύρεσης της τιμής του κατωφλίου "*A threshold selection method from gray-level histograms*" (Otsu, 1979). Έχουν δημιουργηθεί και πιο φιλόδοξες προσεγγίσεις στην αυτόματη κατάτμηση εικόνας. Αναφέρουμε επιγραμματικά μεθόδους που στηρίζονται σε μεθόδους ελαχιστοποίησης ενέργειας (Markov Random Fields, Graph cut αλγόριθμοι, Active Contour Models-Snakes), τα Conditional Random Fields (στατιστικός μοντελισμός), αλλά και ανάλυσης της υφής των εικόνων (texture filters, co-occurrence matrices, Gabor filters, wavelets).

Η χρήση των νευρωνικών δικτύων σε προβλήματα σημασιολογικής κατάτμησης αποτελεί ένα σύγχρονο φαινόμενο. Η πρώτη ουσιαστική πρόοδος επήλθε το 2014 με την σύλληψη των Fully Convolutional Networks - FCN στο έργο "*Fully Convolutional Networks for Semantic Segmentation*" (Long, Shelhamer, and Darrell, 2015). Εκεί παρουσιάστηκε μια τροποποίηση των ήδη επιτυχημένων σε προβλήματα κατηγοριοποίησης Convolutional Neural Networks για την επίλυση προβλημάτων σημασιολογικής κατάτμησης. Η τροποποίηση αφορούσε την αντικατάσταση των fully connected επιπέδων με συνελκτικά επίπεδα. Τα αποτελέσματα αποδείχθηκαν εξαιρετικά. Όπως αναφέρεται και στο ίδιο

το έργο “We show that convolutional networks by themselves, trained end-to-end, pixels-to-pixels, exceed the state-of-the-art in semantic segmentation”.

Μια βελτίωση στη χρήση νευρωνικών δικτύων επήλθε το 2014 στο έργο *"Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs"* (Chen et al., 2014), όπου εισήχθη η ιδέα των dilated (atrous) συνελιζέων και της χρήσης ενός Fully Connected CRF (Conditional Random Field) στο τελευταίο επίπεδο. Τα CRF όπως αναφέραμε παραπάνω αποτελούν μια μέθοδο μηχανικής μάθησης ανεξάρτητη της βαθιάς μάθησης, η οποία έχει χρησιμοποιηθεί μεταξύ άλλων και σε σημασιολογική κατάτμηση.

Δύο ακόμα περιπτώσεις που αξίζει να αναφερθούν αποτελούν το SegNet *"SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation"* (Badrinarayanan, Kendall, and Cipolla, 2015) και το U-net *"U-Net: Convolutional Networks for Biomedical Image Segmentation"* (Ronneberger, Fischer, and Brox, 2015), που δημοσιοποιήθηκαν το 2015. Η αρχιτεκτονική τους είναι παρόμοια. Και τα δύο υλοποιούν encoder-decoder αρχιτεκτονική με σύνδεση των encoder και decoder επιπέδων για αποδοτική ανακατασκευή της εικόνας συνδυάζοντας πληροφορίες εξαγόμενων χαρακτηριστικών αλλά και χωρικών χαρακτηριστικών της εικόνας εισόδου για την παραγωγή της εικόνας εξόδου. Ο τρόπος που συνδυάζουν τις δύο πληροφορίες διαφέρει, όπως και κάποιες άλλες λεπτομέρειες στη δομή της αρχιτεκτονικής τους. Το SegNet αποτελεί μια καλή επιλογή για γενικές εφαρμογές σημασιολογικής κατάτμησης ενώ το U-net ειδικεύεται περισσότερο σε εφαρμογές σημασιολογικής κατάτμησης σε ιατρικές εικόνες, χωρίς αυτοί να είναι απόλυτοι κανόνες.

3.3.2 Κεντρική ιδέα

Όπως προαναφέρθηκε, στόχος των νευρωνικών δικτύων σημασιολογικής κατάτμησης είναι η έξοδος μιας εικόνας προβλέψεων σε επίπεδο εικονοστοιχείου από ένα σύνολο σημασιολογικών κλάσεων. Για την πρόβλεψη είναι απαραίτητη η εξαγωγή χαρακτηριστικών από τις εικόνες εισόδου. Τα συνελκτικά δίκτυα είχαν αποδείξει τις δυνατότητές τους στο συγκεκριμένο πρόβλημα. Η χρήση όμως των διαδεδομένων διαθέσιμων αρχιτεκτονικών δεν θα μπορούσε να υλοποιηθεί καθώς απαντούσαν σε άλλο πρόβλημα, αυτό της σε επίπεδο εικόνας κατηγοριοποίησης. Η χρήση όμως συνελκτικών επιπέδων μπορεί να γίνει δυνατή για την εξαγωγή χαρακτηριστικών με κατάλληλες τροποποιήσεις πριν την έξοδο ώστε οι προβλέψεις να λαμβάνουν χώρα σε επίπεδο εικονοστοιχείου. Η προσέγγιση αυτή, όμως, ενέχει προβλήματα που πρέπει να επιλυθούν. Η χρήση των συνελκτικών επιπέδων σε συνδυασμό με το downsampling που ακολουθεί για την εξαγωγή των χαρακτηριστικών μειώνει τις διαστάσεις της εικόνας και αφαιρεί χαρακτηριστικά της. Εγείρεται λοιπόν το ερώτημα της ανακατασκευής της εικόνας ώστε η έξοδος να αποτελεί ένα πιστό αντίγραφο αυτής, συνοδευόμενο από τις προβλέψεις για την εφαρμογή σημασιολογικής κατάτμησης. Η ανακατασκευή αυτή είναι δυνατή και μπορεί να γίνει με πολλούς τρόπους. Στη συνέχεια θα παρουσιάσουμε, μεταξύ άλλων, κάποιους από αυτούς.

3.3.3 Fully Convolutional Networks - FCN

Η δομή των Fully Convolutional Networks μπορεί να ποικίλει από υλοποίηση σε υλοποίηση, παρόλα αυτά παρουσιάζει κάποια γενικά χαρακτηριστικά (βλέπε εικόνα 3.15). Τα δίκτυα αυτά αποτελούνται από ένα μέρος που εξάγει τα χαρακτηριστικά της εικόνας με χρήση συνελκτικών και downsampling επιπέδων, και από ένα μέρος που ανακατασκευάζει την εικόνα με χρήση upsampling τεχνικών. Τα δύο αυτά διακεκριμένα μέρη και οι λειτουργίες που εκτελούν χαρακτηρίζουν τα δίκτυα αυτά σαν encoder-decoder δίκτυα. Για την διευκόλυνση της ανακατασκευής της εικόνας, ένα άλλο συχνό χαρακτηριστικό αποτελούν

τα skip connections. Τα skip connections συνδέουν συνήθως επίπεδα των δικτύων που έχουν εικόνες μεγάλων διαστάσεων με επίπεδα που έχουν εικόνες μικρών διαστάσεων. Έτσι τα επίπεδα της ανακατασκευής μπορούν να αντλήσουν πληροφορία χωρικής φύσης από τα επίπεδα κωδικοποίησης.

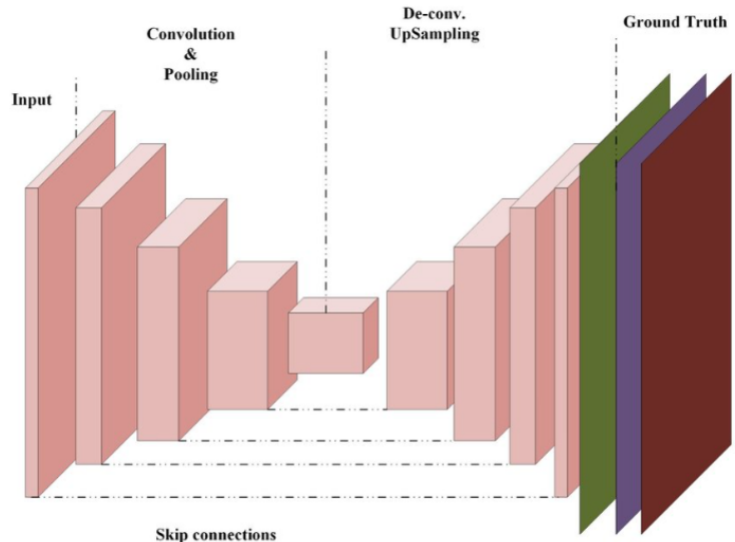


FIGURE 3.15: Η αρχιτεκτονική των Fully Convolutional Neural Networks

3.3.4 Ανακατασκευή εικόνας

Ο συνδυασμός της κωδικοποίησης και της αποκωδικοποίησης της εικόνας αποτελεί μια έννοια κεντρική στα Fully Convolutional Networks. Η πρώτη είναι υπεύθυνη για την εξαγωγή των χαρακτηριστικών της εικόνας και η δεύτερη για την ανακατασκευή της. Για την εξαγωγή χαρακτηριστικών όπως προαναφέραμε γίνεται χρήση τεχνικών συνέλιξης και pooling. Η ανακατασκευή μπορεί να επιτευχθεί με πολλούς τρόπους.

Transposed Convolution

Η αντίστροφη συνέλιξη αποτελεί μια διαδεδομένη μέθοδο ανακατασκευής εικόνας. Χρησιμοποιείται για την αύξηση των διαστάσεων των (συρρικνωμένων) feature maps όπως προκύπτουν από τα επίπεδα του κωδικοποιητή. Η λειτουργία της είναι συνυφασμένη με την έννοια της εκπαίδευσης, με την έννοια ότι η πορεία της εκπαίδευσης επηρεάζει τα αποτελέσματα που θα δώσει και αντίστροφα η πράξη αυτή επηρεάζει την εκπαίδευση. Στην πραγματικότητα αποτελεί μια πράξη συνέλιξης, με τη διαφορά ότι πριν την εφαρμογή εκτελεί έναν μετασχηματισμό στην είσοδό της. Ο μετασχηματισμός αυτός αφορά την επέκταση με μηδενικά (padding) σε επιλεγμένα σημεία με τρόπο που εξαρτάται από τον τρόπο που εκτελέστηκε η πρώτη συνέλιξη.

Πιο συγκεκριμένα αν s , p και k τα μεγέθη του stride, padding και kernel(φίλτρο) της αρχικής συνέλιξης ορίζουμε τις τιμές: $Z = s - 1$, $P' = k - p - 1$, $S' = 1$.

Το Z είναι ο αριθμός των μηδενικών γραμμών και στηλών που θα ενταχθούν ανάμεσα στις τιμές της εισόδου ενώ τα P' και S' οι νέες τιμές για το padding και το stride.

Αφού εφαρμόσουμε τον μετασχηματισμό σύμφωνα με το Z εκτελούμε τη συνέλιξη σύμφωνα με τα P' και S' .

Nearest-neighbor Interpolation

Η παρεμβολή πλησιέστερου γείτονα ακολουθεί μια απλή διαδικασία. Για τον υπολογισμό μιας νέας τιμής σε ένα ορισμένο σημείο του χώρου αντιγράφεται η τιμή του κοντινότερου διαθέσιμου δείγματος.

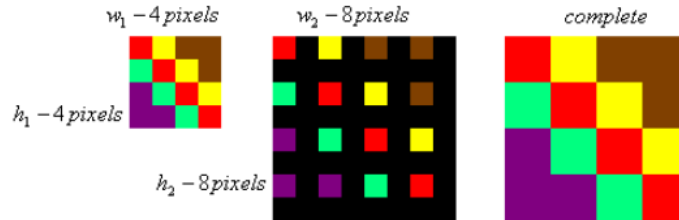


FIGURE 3.16: Παρεμβολή πλησιέστερου γείτονα. Αντιγραφή του κοντινότερου διαθέσιμου δείγματος για τον υπολογισμό των νέων τιμών

Bilinear Interpolation

Η διγραμμική παρεμβολή υπολογίζει την τιμή του νέου στοιχείου ως τον γραμμικό συνδυασμό των τεσσάρων πλησιέστερών του από την αρχική εικόνα. Αν $f(x, y)$ η τιμή που θα προκύψει μέσω της παρεμβολής στο σημείο (x, y) και $Q_{11} = (x_1, y_1)$, $Q_{12} = (x_1, y_2)$, $Q_{21} = (x_2, y_1)$, $Q_{22} = (x_2, y_2)$ οι τιμές των γειτονικών του σημείων η διγραμμική παρεμβολή θα βρεθεί υλοποιώντας διαδοχικά γραμμική παρεμβολή στην κατεύθυνση x και y .

Για την γραμμική παρεμβολή στην κατεύθυνση x θα είναι:

$$f(x, y_1) = \frac{x_2 - x}{x_2 - x_1} f(Q_{11}) + \frac{x - x_1}{x_2 - x_1} f(Q_{21}) \quad (3.19)$$

$$f(x, y_2) = \frac{x_2 - x}{x_2 - x_1} f(Q_{12}) + \frac{x - x_1}{x_2 - x_1} f(Q_{22}) \quad (3.20)$$

Μετά εφαρμόζουμε και την γραμμική παρεμβολή στην κατεύθυνση y :

$$f(x, y) = \frac{y_2 - y}{y_2 - y_1} f(x, y_1) + \frac{y - y_1}{y_2 - y_1} f(x, y_2)$$

$$f(x, y) = \frac{y_2 - y}{y_2 - y_1} \left(\frac{x_2 - x}{x_2 - x_1} f(Q_{11}) + \frac{x - x_1}{x_2 - x_1} f(Q_{21}) \right) + \frac{y - y_1}{y_2 - y_1} \left(\frac{x_2 - x}{x_2 - x_1} f(Q_{12}) + \frac{x - x_1}{x_2 - x_1} f(Q_{22}) \right)$$

Και τελικά μετά τις πράξεις καταλήγουμε:

$$f(x, y) = \frac{1}{(x_2 - x_1)(y_2 - y_1)} [x_2 - x \quad x - x_1] \begin{bmatrix} f(Q_{11}) & f(Q_{12}) \\ f(Q_{21}) & f(Q_{22}) \end{bmatrix} \begin{bmatrix} y_2 - y \\ y - y_1 \end{bmatrix} \quad (3.21)$$

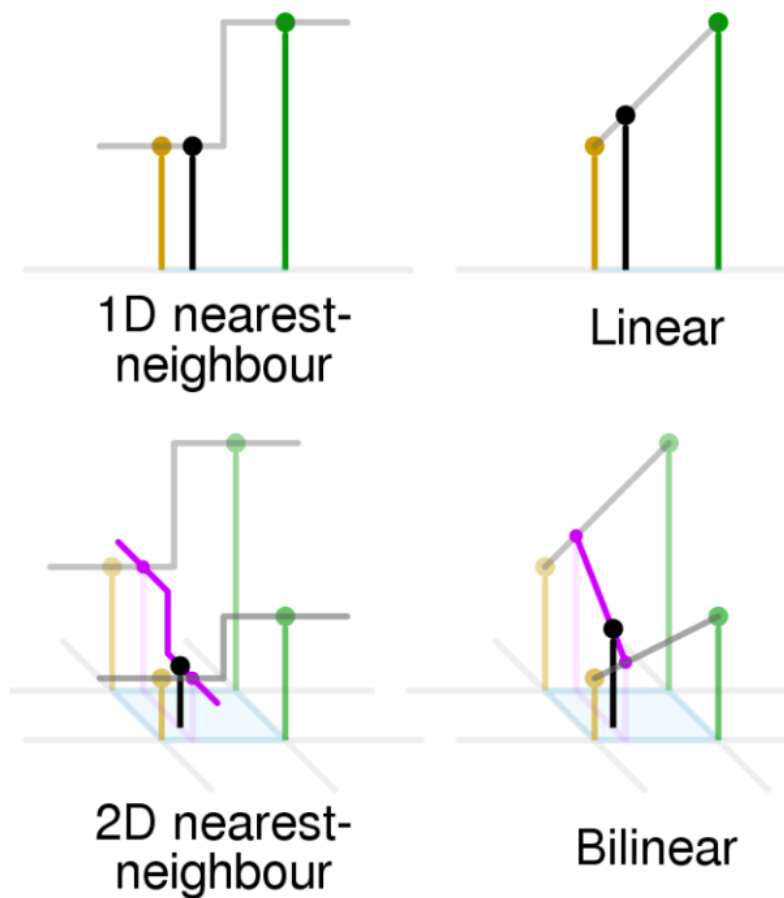


FIGURE 3.17: Γραμμική και διγραμμική παρεμβολή σε αντιδιαστολή με την παρεμβολή πλησιέστερου γείτονα

Max Unpooling

Στο Max Unpooling ανακατασκευάζουμε την εικόνα με βάση τις αρχικές τοποθεσίες των μέγιστων τιμών. Προφανώς οι τιμές που προσδιορίζουν τη θέση της μέγιστης τιμής πρέπει να αποθηκευτούν πριν εκτελεστεί το pooling. Οι μέγιστες τιμές τοποθετούνται στις αρχικές τους θέσεις ενώ οι υπόλοιπες συμπληρώνονται με μηδενικά.

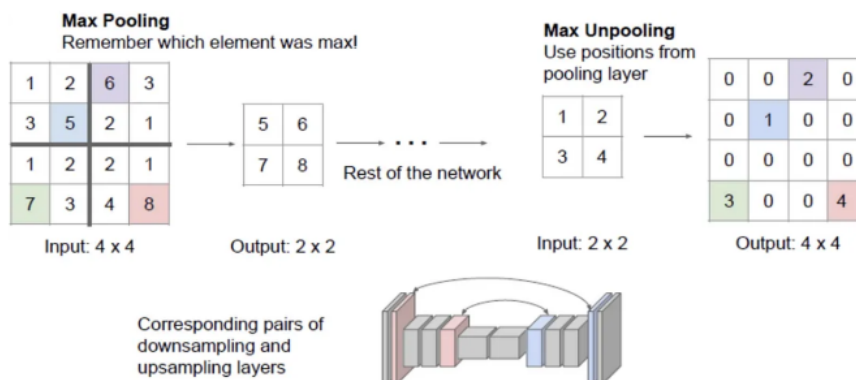


FIGURE 3.18: Στο Max Unpooling ανακατασκευάζεται η εικόνα με βάση τις αρχικές τοποθεσίες των μέγιστων τιμών

3.3.5 Παράδειγμα: FCN

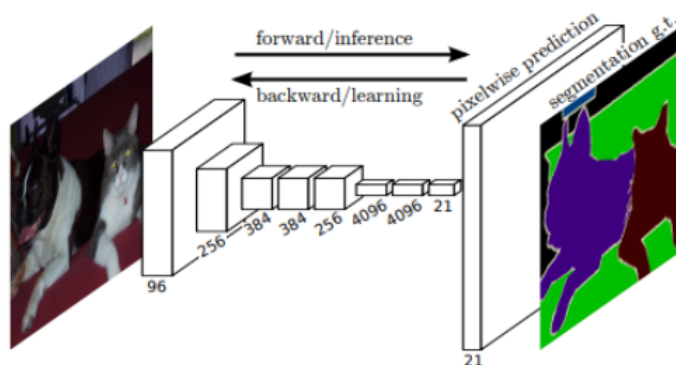


FIGURE 3.19: Η αρχιτεκτονική του FCN αποτελείται μόνο από συνελκτικά επίπεδα

Το FCN που παρουσιάζεται στο έργο “Fully Convolutional Networks for Semantic Segmentation”, αποτελεί ένα δίκτυο το οποίο αντικαθιστά τα fully connected επίπεδα με συνελκτικά για την εφαρμογή σημασιολογικής κατάτμησης σε εικόνες. Υλοποιείται στα τελευταία επίπεδα transposed convolution για την ανακατασκευή της εικόνας, συνδυάζοντας την πληροφορία από τα τελευταία επίπεδα (βλέπε εικόνα 3.20). Πιο συγκεκριμένα από όλα τα επίπεδα συνελίξεων χρησιμοποιείται η πληροφορία από τα τελευταία 3 με transposed convolutions, προβλέψεις και αθροίσματα.

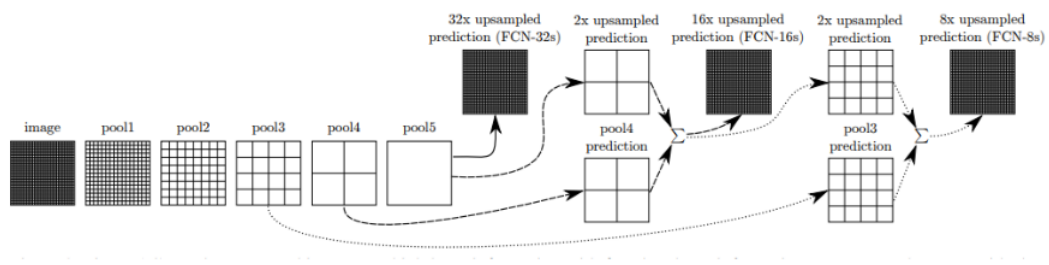


FIGURE 3.20: Στο FCN μπορούν να χρησιμοποιηθούν 3 μέθοδοι ανακατασκευής της εικόνας ανάλογα την επιλογή του αριθμού των τελευταίων επιπέδων

Κάθε upsampled prediction αποτελεί μια πρόβλεψη της εικόνας αξιοποιώντας διαφορετικό αριθμό τελευταίων επιπέδων (FCN-32, FCN-16, FCN-8). Όσο περισσότερα προηγούμενα επίπεδα χρησιμοποιούνται τόσο καλύτερη τείνει να είναι η κατάτμηση της εικόνας, με το FCN-16 να μην έχει βέβαια ιδιαίτερα μεγάλες διαφορές στην απόδοση από το FCN-8.

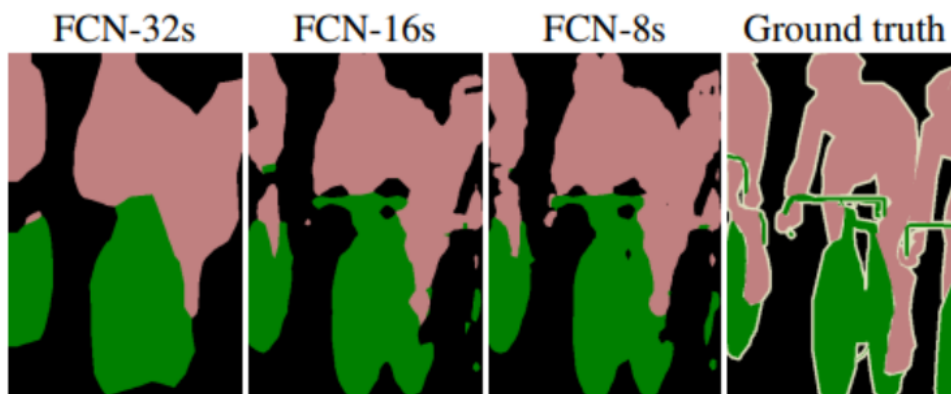


FIGURE 3.21: Τα αποτελέσματα των διαφορετικών μεθόδων ανακατασκευής του FCN. Τα FCN-16 και FCN-8 παράγουν αισθητά καλύτερα αποτελέσματα από το FCN-32

3.3.6 Παράδειγμα: SegNet

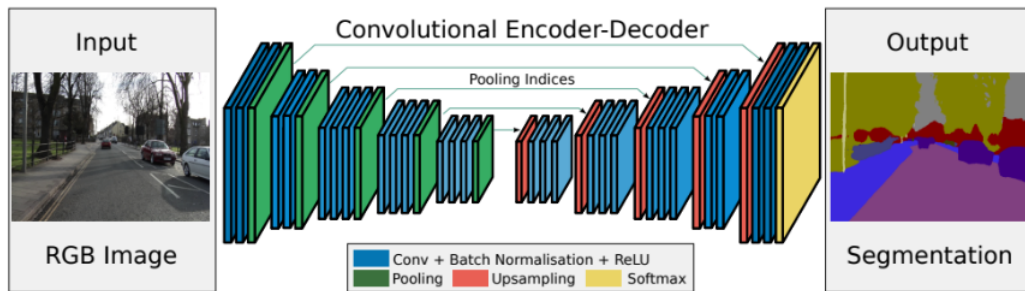


FIGURE 3.22: Η αρχιτεκτονική κωδικοποιητή-αποκωδικοποιητή του SegNet

Το SegNet αποτελεί ένα δίκτυο για σημασιολογική κατάτμηση με δύο χαρακτηριστικά που το κάνουν να ξεχωρίζει από τα υπόλοιπα. Το ένα είναι η συμμετρική αρχιτεκτονική encoder-decoder, με τον encoder να υπεύθυνο για την εξαγωγή των χαρακτηριστικών και τον decoder υπεύθυνο για την ανακατασκευή της εικόνας και το άλλο είναι ο τρόπος με τον οποίο υλοποιεί την ανακατασκευή της εικόνας. Για την ανακατασκευή της εικόνας χρησιμοποιεί Max Unpooling. Το δίκτυο αποθηκεύει τις θέσεις των μεγίστων στο μέρος του encoder και τις περνάει στον decoder όπου χρησιμοποιούνται για να γίνει το upsampling. Αυτός ο τρόπος λειτουργίας επιτρέπει στο SegNet να επιτυγχάνει αρκετά καλά αποτελέσματα χωρίς να απαιτεί μεγάλο αριθμό υπολογισμών (συγκριτικά με άλλες υλοποιήσεις όπως το U-net το οποίο μεταφέρει όλη την εικόνα από τον encoder στον decoder).

3.3.7 Συναρτήσεις Κόστους - Μετρικές για εφαρμογές Σημασιολογικής Κατάτμησης

Τα προβλήματα σημασιολογικής κατάτμησης χαρακτηρίζονται από ιδιαιτερότητες οι οποίες απαιτούν διαφορετική αντιμετώπιση για την σωστή αξιολόγηση της απόδοσής τους από άλλες εφαρμογές νευρωνικών δικτύων όπως για παράδειγμα σε προβλήματα κατηγοριοποίησης. Ένα δίκτυο σημασιολογικής κατάτμησης για μια εικόνα δεν παράγει μια εκτίμηση αλλά πολλές, τόσες όσο είναι το σύνολο των εικονοστοιχείων της εικόνας εξόδου. Το γεγονός αυτό εγείρει το ερώτημα να αναλογιστούμε τι θεωρούμε καλή απόδοση. Η πρώτη εκτίμηση που ίσως γίνει από εμάς δεν θα είναι καλή. Το αναφέρουμε αυτό σε περίπτωση που η πρώτη εκτίμηση ήταν η: “όσο περισσότερα εικονοστοιχεία εκτιμήθηκαν σωστά τόσο καλύτερη ήταν η απόδοση του μοντέλου”. Η προσέγγιση αυτή ονομάζεται ακρίβεια και συγκεκριμένα ακρίβεια σε επίπεδο εικονοστοιχείου (pixel accuracy). Η ακρίβεια εικονοστοιχείου PA για k κλάσεις ορίζεται ως:

$$PA = \frac{\sum_{j=1}^k n_{jj}}{\sum_{j=1}^k t_j} \quad (3.22)$$

Όπου n_{jj} είναι ο αριθμός των σωστά εκτιμημένων pixel για την κλάση j (true positives) και t_j είναι ο αριθμός των πραγματικών στοιχείων της κλάσης.

Για να αναδείξουμε την αδυναμία αυτής της προσέγγισης ας αναλογιστούμε το ενδεχόμενο δύο κλάσεων $K = 0, 1$ με $t_0 \gg t_1$

Τότε θα είναι:

$$PA = \frac{n_{00} + n_{11}}{t_0 + t_1} \simeq \frac{n_{00} + n_{11}}{t_0} \quad (3.23)$$

Η απαλοιφή του παράγοντα t_1 από τον παρονομαστή είναι προβληματική. Σημαίνει ότι η δεύτερη κλάση όντας μειοψηφία έχει χάσει την επιρροή της στην εκτίμηση απόδοσης του δικτύου.

Αν ένα δίκτυο απλά εκτιμούσε πάντα την πιο πολυπληθή κλάση στο συγκεκριμένο πρόβλημα θα ήταν $n_{00} = t_0, n_{11} = 0$ και τελικά:

$$PA \simeq \frac{t_0 + 0}{t_0} = 1 \quad (3.24)$$

Η εκτίμηση αυτή μας αναφέρει ότι το δίκτυο τα πήγε εξαιρετικά (σχεδόν τέλεια), ενώ αυτό παραμέλησε τελείως την κλάση που αποτελούσε τη μειοψηφία. Πολλές φορές σε προβλήματα σηματολογικής κατάτμησης θέλουμε να χωρίσουμε μια εικόνα σε μέρη όπου κάποιο θα είναι πάντα μεγαλύτερο από τα άλλα (για παράδειγμα background). Στη συνέχεια θα παρουσιάσουμε κάποιες μετρικές που μεριμνούν για το πρόβλημα αυτό και δίνουν καλύτερες εκτιμήσεις για την αποτελεσματικότητα των δικτύων.

F1 score - Dice Coefficient

Το F1 score υπολογίζεται σαν αρμονικός μέσος του Precision και του Recall. Το Precision και το Recall αποτελούν δύο διαφορετικές πτυχές εκτίμησης της επιτυχίας ενός δικτύου.

Συγκεκριμένα ορίζονται ως:

$$Precision = \frac{TP}{TP + FP} \quad (3.25)$$

$$Recall = \frac{TP}{TP + FN} \quad (3.26)$$

Καταλαβαίνουμε ότι το Precision εκτιμά την επιτυχία του δικτύου με βάση το πόσα λίγα False Positives δίνει ενώ το Recall με βάση το πόσα λίγα False Negatives. Για την συνεκτίμηση και των δύο το F1-score υπολογίζεται ως ο αρμοτικός τους μέσος:

$$Dice = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} \quad (3.27)$$

ή

$$Dice = \frac{2PrecisionRecall}{Precision + Recall} \quad (3.28)$$

Αρκετά διαδεδομένη είναι η μορφή του F1 score, υπολογισμένο με βάση τα TP, FN, FP. Με αντικατάσταση:

$$Dice = \frac{2TP}{2TP + FP + FN} \quad (3.29)$$

Μπορούμε να καταλάβουμε τον υπολογισμό και με μια προσέγγιση συνόλων. Το πλήθος $TP+FP$ αποτελεί το σύνολο της πρόβλεψης θετικών τιμών ενώ το πλήθος $TP+FN$ αποτελεί τις πραγματικές θετικές τιμές.

Έτσι το F1 score υπολογίζεται ως:

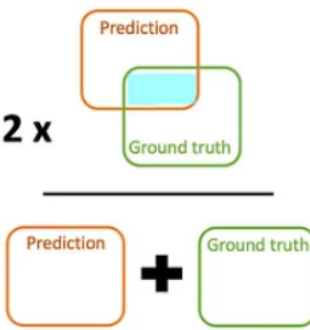
$$\text{Dice} = \frac{2 \times \text{Area of overlap}}{\text{Total area}} = \frac{2 \times \text{Prediction} \cap \text{Ground truth}}{\text{Prediction} \cup \text{Ground truth}}$$


FIGURE 3.23: Το Dice coefficient υπολογίζεται ως ο λόγος της επικάλυψης (επί δύο) προς την συνολική περιοχή

Καταλαβαίνουμε λοιπόν ότι όσο μεγαλύτερο είναι το σύνολο της τομής συγκριτικά με συνολικό σύνολο η μετρική θα είναι όλο και μεγαλύτερη με μέγιστη τιμή τη μονάδα.

Παρατηρήστε πως η εφαρμογή αυτής της μετρικής στο προηγούμενο παράδειγμα θα έδινε αποτέλεσμα μηδενικό, που είναι και η τιμή που θα θέλαμε να δούμε καταδεικνύοντας την καταλληλότητα της συγκεκριμένης μετρικής σε εφαρμογές σημασιολογικής κατάταξης.

Intersection over Union - IoU (Jaccard Index)

Το intersection over Union αποτελεί μια μετρική παρόμοια με το F1 score.

Υπολογίζεται ως:

$$\text{IoU} = \frac{TP}{TP + FP + FN} \quad (3.30)$$

Και σε αυτή τη μετρική μπορούμε να δοκιμάσουμε μια προσέγγιση συνόλων. Η IoU υπολογίζεται σαν ο λόγος της τομής προς την ένωση των συνόλων της πρόβλεψης και των πραγματικών τιμών, όπως αναφέρει και το όνομα της μετρικής. Η ένωση ορίζεται ως: $(TP + FP) + (TP + FN) - TP = TP + FP + FN$

Άρα τελικά:

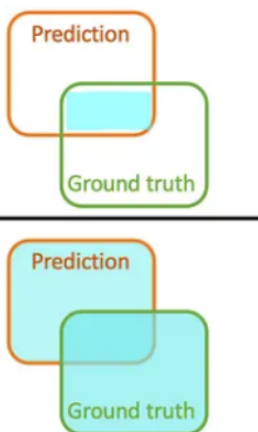
$$\text{IoU} = \frac{\text{Area of overlap}}{\text{Area of union}} = \frac{\text{Prediction} \cap \text{Ground truth}}{\text{Prediction} \cup \text{Ground truth}}$$


FIGURE 3.24: Το IoU υπολογίζεται ως ο λόγος της επικάλυψης προς την ένωση

Tversky Index

Το Tversky Index αποτελεί μια παραλλαγή των παραπάνω μετρικών με σκοπό τον καλύτερο έλεγχο της ισορροπίας μεταξύ των False Positives και των False Negatives.

Ορίζεται ως:

$$Tverskyindex = \frac{TP}{TP + \alpha FP + (1 - \alpha)FN} \quad (3.31)$$

Παρατηρούμε εδώ ότι εισάγεται ένας παράγοντας ελέγχου, το α . Αν $\alpha < 0.5$ δίνεται μεγαλύτερη έμφαση στα FN ενώ αλλιώς δίνεται στα FP.

Το Tversky index μπορεί να χρησιμοποιηθεί και στην εκπαίδευση ενός δικτύου σαν συνάρτηση κόστους ως $(1 - Tversky Index)$ έτσι ώστε το δίκτυο στην εκπαίδευσή του να εστιάζει περισσότερο στη βελτίωση ενός εκ των FP και FN, σε περιπτώσεις όπου το πρόβλημα της τμηματοποίησης το απαιτεί.

Focal Tversky Loss

Μια παραλλαγή του Tversky index σαν συνάρτηση κόστους αποτελεί και η Focal Tversky Loss. Εισάγει έναν ακόμη παράγοντα ελέγχου με σκοπό την αποτελεσματικότερη κατάτμηση κλάσεων με ανισόρροπο πλήθος.

Ορίζεται ως:

$$FocalTverskyloss = (1 - Tverskyindex)^\gamma \quad (3.32)$$

Η ύψωση στον παράγοντα γ έχει σαν αποτέλεσμα την περαιτέρω αύξηση της σχετικής διαφοράς αποκλίσεων για την κάθε κλάση. Έτσι το δίκτυο στην εκπαίδευσή του μπορεί να εστιάζει περισσότερο στην βελτιστοποίηση των αποτελεσμάτων και για τις λιγότερο πολυπληθείς κλάσεις.

3.3.8 Σημαιολογική κατάτμηση Ιατρικών Εικόνων με χρήση του U-Net

Το U-net αποτελεί μια αρχιτεκτονική νευρωνικού δικτύου σηματολογικής κατάτμησης που δημοσιεύθηκε το 2015 από τους Olaf Ronneberger, Philipp Fischer, και Thomas Brox στο έργο "U-Net: Convolutional Networks for Biomedical Image Segmentation". Την ίδια χρονιά, κέρδισε τον ISBI (International Symposium on Biomedical Imaging) διαγωνισμό σηματολογικής κατάτμησης κυττάρων με μεγάλη διαφορά. Όπως αναφέρεται και στον τίτλο της δημοσίευσης, το U-net βρίσκει εφαρμογή σε προβλήματα σηματολογικής κατάτμησης ιατρικών εικόνων, χωρίς βέβαια να περιορίζεται υποχρεωτικά μόνο σε αυτά. Από τη στιγμή της δημοσίευσής του, βρέθηκε στο επίκεντρο της προσοχής. Η χρήση του διαδόθηκε και ακολούθησαν έργα βασισμένα σε αυτό επεκτείνοντας και βελτιώνοντας τις δυνατότητές του. Στη συνέχεια θα παρουσιάσουμε την αρχιτεκτονική του.

U-net

Αρχιτεκτονική

Η αρχιτεκτονική του U-net, όπως και άλλες αρχιτεκτονικές σηματολογικής κατάτμησης, ακολουθεί τις αρχές των Fully Convolutional Networks και των encoder decoder δομών. Το encoder block είναι υπεύθυνο για την εξαγωγή των χαρακτηριστικών της εικόνας και το decoder για την ανακατασκευή της εικόνας. Τα δύο αυτά blocks συνδέονται με skip connections για την μετάδοση χωρικών πληροφοριών από τον encoder στον decoder για

την υποβοήθηση της ποιοτικής ανακατασκευής της εικόνας. Λόγω της φύσης της λειτουργίας τους ο encoder και ο decoder ονομάζονται και contracting και expanding path αντίστοιχα. Μια ιδιαιτερότητα του U-net μεταξύ άλλων είναι η συμμετρία που έχει στην αρχιτεκτονική του, με την έννοια ότι το contracting και το expanding path έχουν τις ίδιες διαστάσεις. Άλλες, προγενέστερες αρχιτεκτονικές δεν είχαν στα επίπεδα ανατασκευής τόσο μεγάλο αριθμό από feature maps. Η συμμετρική φύση του contracting και του expanding path είναι αυτή που δίνει στο U-net και το όνομά του από την χαρακτηριστική U μορφή που σχηματίζεται. Στο σχήμα 3.25 παρουσιάζεται η αρχιτεκτονική του U-net.

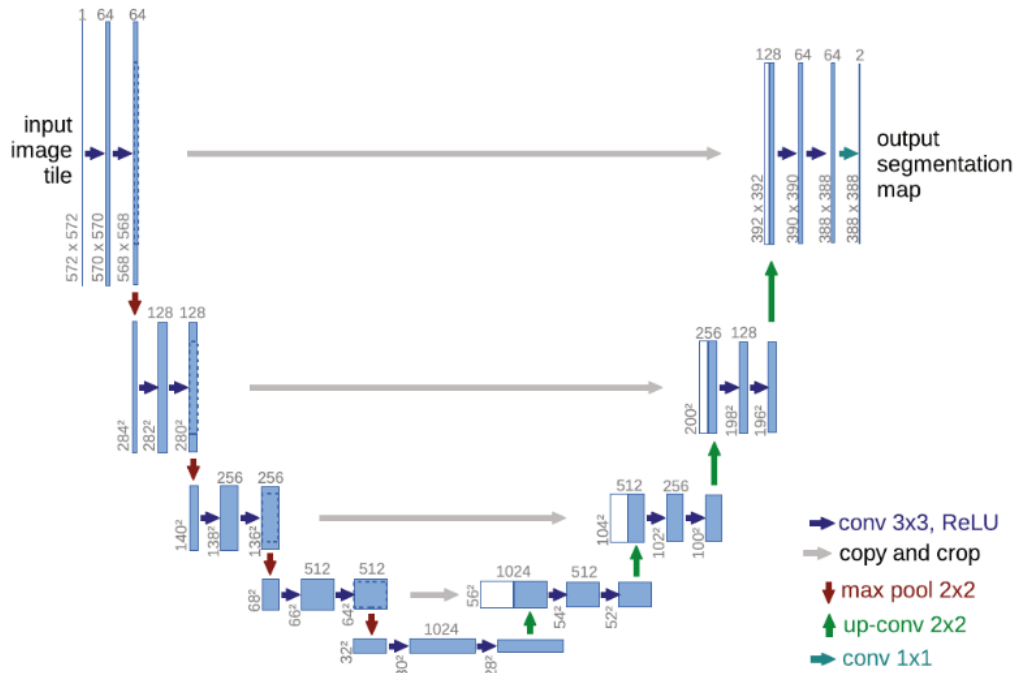


FIGURE 3.25: Η αρχιτεκτονική του U-net. Αποτελείται από τον κωδικοποιητή και τον αποκωδικοποιητή οι οποίοι είναι συμμετρικοί. Η συμμετρικότητά τους θυμίζει τελικά το σύμβολο U

Πιο συγκεκριμένα, το contracting path αποτελείται από διαδοχικές εφαρμογές δύο 3x3 συνελίξεων (χωρίς padding) με χρήση ενεργοποίησης ReLU, η κάθε μια ακολουθούμενη από 2x2 max pooling. Το expansive path αποτελείται από περισσότερα βήματα. Αρχικά γίνεται upsampling του κάτω feature map ακολουθούμενο από ένα inverse convolution 2x2. Το βήμα αυτό αυξάνει τις διαστάσεις τις εικόνας αλλά μειώνει τις διαστάσεις του feature map. Αυτό είναι το σημείο που γίνεται concatenation με τα feature map του encoder από το skip connection του δικτύου. Στη συνέχεια ακολουθούν δύο συνελίξεις 3x3 όπως στο επίπεδο του encoder. Στο τελευταίο επίπεδο του δικτύου υλοποιείται μια 1x1 συνέλιξη έτσι ώστε από το 64 επιπέδων feature map να προκύψει η πρόβλεψη στο κάθε εικονοστοιχείο. Παρουσιάζουμε και ένα πιο αναλυτικό σχήμα της αρχιτεκτονικής του U-net (3.26).

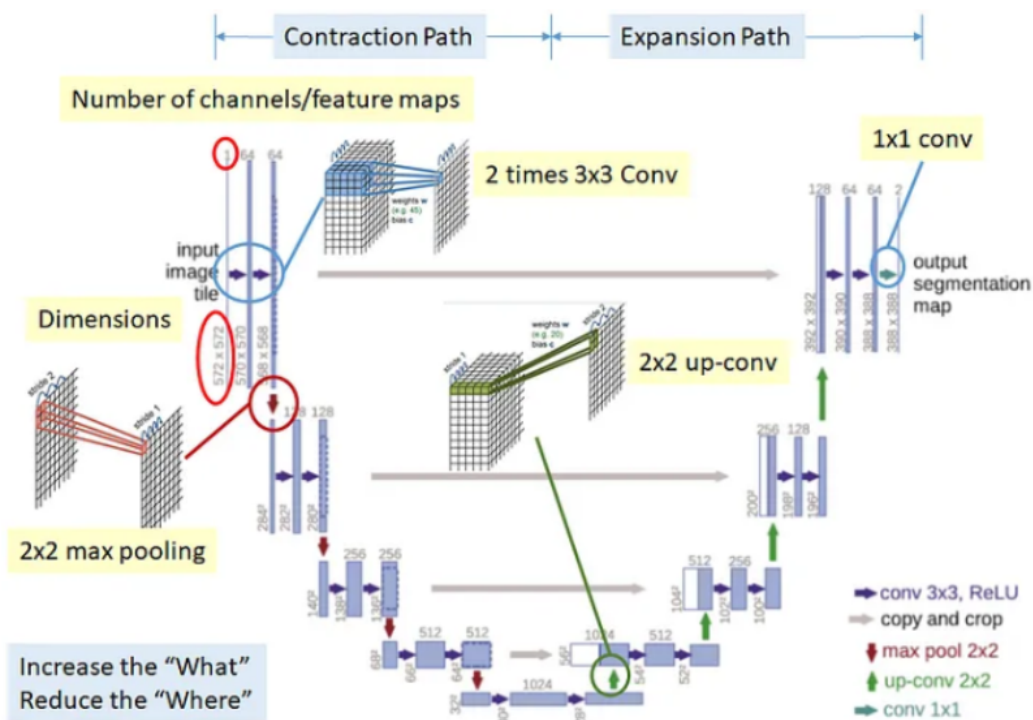


FIGURE 3.26: Η αρχιτεκτονική του U-net με περισσότερη λεπτομέρεια

Τμηματοποίηση εικόνων μεγάλων διαστάσεων με την **Overlap-Tile Strategy**

Για κατάτμηση εικόνων μεγαλύτερων διαστάσεων από την είσοδο του U-net μπορεί να εφαρμοστεί η στρατηγική **Overlap-Tile** (βλέπε εικόνα 3.27). Η εικόνα χωρίζεται σε επικαλυπτούμενα (**Overlap**) μέρη (**Tiles**) και δίνεται στο U-net για να παραχθεί η κατάτμηση αυτού του μέρους της εικόνας. Το U-net θα παράξει την έξοδό του η οποία θα έχει διάσταση μικρότερη από την είσοδό του, αφού οι συνελίξεις υλοποιούνται χωρίς padding. Αυτή η διαδικασία θα συνεχιστεί και για τα υπόλοιπα μέρη της εικόνας. Τα μέρη της εικόνας εισόδου είναι επικαλυπτούμενα έτσι ώστε να υπάρχει συνέχεια στην συνολική και ενοποιημένη έξοδο. Η μείωση των διαστάσεων της εξόδου του U-net δημιουργεί το πρόβλημα της έλλειψης δεδομένων εισόδου για τα περιμετρικά μέρη της εικόνας εξόδου. Για να μπορέσουν να παραχθούν και αυτά, μπορεί να γίνει μια επέκταση της εικόνας εισόδου με τη μέθοδο του **mirroring**.

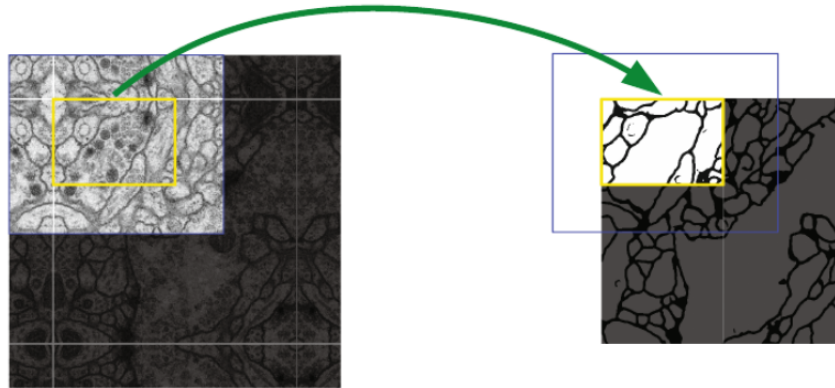


FIGURE 3.27: Η στρατηγική Overlap-Tile που χρησιμοποιήθηκε στο U-net

Εκπαίδευση του U-net

Η υλοποίηση της εκπαίδευσης πάντα αποτελεί ένα κρίσιμο κομμάτι στην δημιουργία ενός δικτύου. Στο U-net έγινε χρήση της Stochastic Gradient Descent. Οι εικόνες δόθηκαν ολόκληρες και το batch size αποτελούνταν από μια εικόνα. Ως προς την προετοιμασία των εικόνων, υλοποιήθηκε data augmentation με elastic transformations, ένα ταιριαστό είδος αφού το δίκτυο προοριζόταν για σημασιολογική κατάτμηση κυττάρων σε ιατρικές εικόνες και τα κύτταρα σε αυτές παρουσιάζουν γενικά ποικιλομορφία τέτοιας φύσεως. Σαν συνάρτηση κόστους χρησιμοποιήθηκε η cross entropy και σαν συνάρτηση ενεργοποίησης στο τελευταίο επίπεδο η softmax. Τέλος αναφέρουμε ότι για την καλύτερη σύγκλιση στην εκπαίδευση έγινε αρχικοποίηση των βαρών σύμφωνα με τη Γκαουσιανή κατανομή με τυπική απόκλιση $s = \sqrt{\frac{2}{N}}$, όπου N ο αριθμός εισόδων ενός νευρώνα.

Attention U-net

Η προσοχή στα νευρωνικά δίκτυα αποτελεί έναν μηχανισμό που τους δίνει τη δυνατότητα της εστίασης σε χρήσιμες πληροφορίες. Ο μηχανισμός αυτός αποτελεί μια ιδιαίτερα χρήσιμη τεχνική η οποία έχει ευρεία χρήση και στην επεξεργασία φυσικής γλώσσας και στην όραση υπολογιστών, ενώ σε συγκεκριμένες εφαρμογές έχει επιφέρει δραματικές βελτιώσεις. Στη σύγχρονη μορφή του εισήχθη το 2014 στο έργο "*Neural Machine Translation by Jointly Learning to Align and Translate*" (Bahdanau, Cho, and Bengio, 2014) σε εφαρμογή δημιουργίας νευρωνικού δικτύου μετάφρασης γλωσσών. Το συγκεκριμένο δίκτυο κατάφερε να παράξει σημαντικά ποιοτικότερες μεταφράσεις. Η προσοχή αποδείχθηκε, λοιπόν, ένα πολύ χρήσιμο εργαλείο από τότε, αλλά δεν είχε ακόμα εκμεταλλευτεί στο έπακρο. Το 2017 ο Ashish Vaswani στο έργο "*Attention is all you Need*" (Vaswani et al., 2017) εισήγαγε την αρχιτεκτονική των transformer, μια αρχιτεκτονική η οποία αντικατέστησε μηχανισμούς recurrence (RRNs) και μηχανισμούς συνελίξεων με μηχανισμούς προσοχής για εφαρμογές Επεξεργασίας Φυσικής Γλώσσας. Ο transformer που παρουσίασαν είχε ιδιαίτερα απλούστερη αρχιτεκτονική, ιδιαίτερα καλύτερα αποτελέσματα και ιδιαίτερα γρηγορότερους χρόνους εκπαίδευσης. Αναμφίβολα οι transformers επέφεραν την επανάσταση στην επεξεργασία φυσικής γλώσσας, ανοίγοντας το δρόμο για πληθώρα εντυπωσιακών εφαρμογών όπως για παράδειγμα των large-scale language

models. Η χρήση παρόλα αυτά της προσοχής δεν περιορίζεται μόνο σε αυτές τις εφαρμογές. Η προσοχή έχει συμβάλει και σε εφαρμογές της όρασης των υπολογιστών και στη συνέχεια θα παρουσιάσουμε ένα τέτοιο παράδειγμα.

Το Attention U-net (Oktay et al., 2018) αποτελεί μια παραλλαγή του U-net η οποία εισάγει μηχανισμούς προσοχής για την βελτίωση των επιδόσεων σε προβλήματα σηματολογικής κατάτμησης. Αυτό το επιτυγχάνει δίνοντας τη δυνατότητα στο δίκτυο να εστιάζει περισσότερο σε περιοχές που παρουσιάζουν ενδιαφέρον. Η προσοχή αυτή μάλιστα αποτελεί μέρος του δικτύου, χωρίς να απαιτείται κάποια εξωτερική παρέμβαση και χτίζεται από το ίδιο το δίκτυο κατά την εκπαίδευσή του (soft attention). Υλοποιείται στο στάδιο της αντιγραφής των χωρικών πληροφοριών από τον encoder, υποδεικνύοντας στο δίκτυο να ανακατασκευάσει την κατετμημένη εικόνα εστιάζοντας περισσότερο σε συγκεκριμένες περιοχές. Η πληροφορία για το που θα εστιάσει έρχεται από τα κάτω επίπεδα όπου εμπεριέχονται οι νοηματικές πληροφορίες, αυτές των χαρακτηριστικών που έχουν εξαχθεί.

Παρουσιάζουμε αρχικά την αρχιτεκτονική του δικτύου στο σχήμα (3.28).

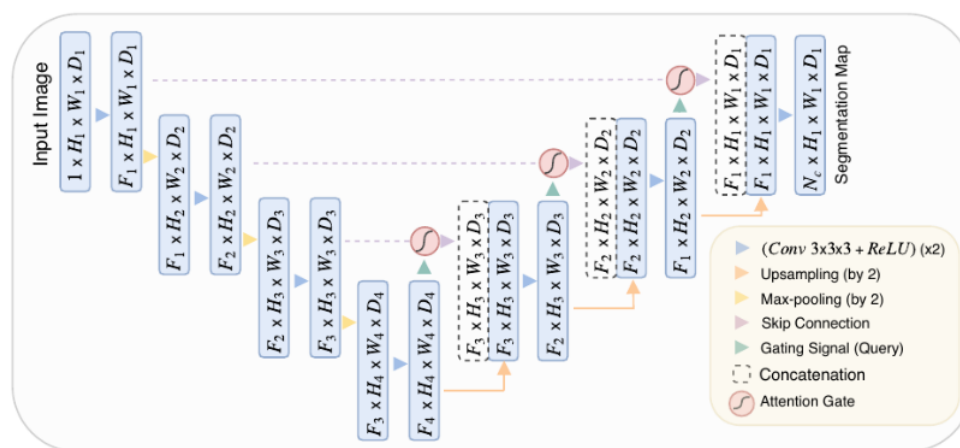


FIGURE 3.28: Η αρχιτεκτονική του attention U-net

Παρατηρούμε αρχικά ότι η αρχιτεκτονική του δικτύου δεν διαφέρει και πολύ από αυτή του U-net. Η πραγματική διαφορά εντοπίζεται στα Attention Gates που έχουν προστεθεί. Τα attention gates είναι υπεύθυνα, όπως προαναφέραμε για την ανακατασκευή της εικόνας με γνώμονα τις σημαντικές περιοχές που πρέπει να εστιάσει το δίκτυο. Η έξοδος των attention gate αποτελεί την αντεγγραμμένη πληροφορία από feature maps του encoder πολλαπλασιασμένη με τους συντελεστές προσοχής a , οι οποίοι μεγιστοποιούν ή ελαχιστοποιούν για την συνεισφορά του κάθε μέρους στο τελικό αποτέλεσμα, προσδίδοντας έτσι την προσοχή στο δίκτυο.

Attention Gates

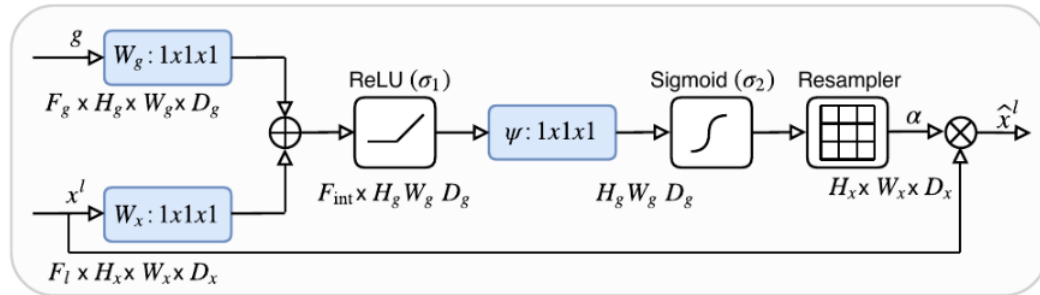


FIGURE 3.29: Η αρχιτεκτονική των attention gate

Ο τρόπος που συνδυάζουν τα attention gates (βλέπε σχήμα 3.29) την πληροφορία από τα σήματα εισόδου x και g στον πυρήνα του είναι απλός. Προσθέτουν τα βάρη των σημάτων. Αν αναλογιστούμε τι είναι ένα βάρος μπορούμε να καταλάβουμε ότι μεγάλες τιμές βαρών συνεπάγονται μεγαλύτερη συνεισφορά στους υπολογισμούς ενός δικτύου. Με άλλα λόγια τα μεγάλα βάρη καταδεικνύουν την σημαντικότερη πληροφορία σε ένα δίκτυο. Δεδομένου αυτού μπορούμε να δούμε την πρόσθεση των βαρών σαν έναν συνδυασμό της σημαντικότητας δύο σημάτων. Δύο μικρά βάρη θα δώσουν σαν αποτέλεσμα μια μικρή πρόσθεση βαρών, ένα μικρό και ένα μεγάλο μια μεσαία και δύο μεγάλα μια μεγάλη.

Όλα τα υπόλοιπα μέρη του attention gate δεδομένων των παραπάνω μπορούν εύκολα να πάρουν τη θέση τους. Οι $1 \times 1 \times 1$ (channel-wise) συνελίξεις εκτελούνται για να ετοιμάσουν τις διαστάσεις των g και x^l για την πρόσθεση. Στο σήμα x^l εκτελούνται strided συνελίξεις και οι διαστάσεις του μικραίνουν (ήταν μεγαλύτερες από το g). Γίνεται χρήση της ReLU καθώς όλα αυτά τα ενδιάμεσα-εσωτερικά βάρη συνεχίζουν να αποτελούν μέρος του δικτύου τα οποία εκπαιδεύονται αλλά μετέχουν και σε υπολογισμούς πρόβλεψης. Η συνέλιξη μετά τη ReLU γίνεται για να πέσουν οι διαστάσεις του σήματος ώστε μετά να περάσει από την Σιγμοειδή συνάρτηση ενεργοποίησης για κανονικοποίηση των τιμών. Η κανονικοποίηση είναι απαραίτητη γιατί ειδικά μετά την πρόσθεση, τα βάρη μπορεί να πάρουν πολύ μεγάλες τιμές. Τέλος το σήμα περνάει από τον resampler ο οποίος εφαρμόζει trilinear interpolation (επέκταση του bilinear interpolation που παρουσιάσαμε σε τρεις διαστάσεις), για να ανακατασκευάσει τα βάρη α στις κατάλληλες διαστάσεις ώστε να μπορούν να εφαρμοστούν πάνω στο σήμα x^l .

Residual Attention U-net

Τα residual δίκτυα εισήχθησαν το 2015, μια εποχή όπου τα βαθιά συνελικτικά δίκτυα είχαν δείξει τις δυνατότητες τους και είχε δημιουργηθεί μια ροπή προς αυτήν την κατεύθυνση, δικτύων που γινόντουσαν όλο και πιο βαθιά. Αν και πράγματι το μέγεθος στα νευρωνικά δίκτυα παίζει κρίσιμο ρόλο στην αποτελεσματικότητά τους, η αλόγιστη αύξησή του δεν συνεπάγεται πάντα βελτίωση στις αποδόσεις. Ο λόγος είναι ότι η αύξηση του μεγέθους των δικτύων δημιουργεί προβλήματα. Ένα από αυτά αποτελεί το vanishing gradient στο οποίο οι κλίσεις όπως υπολογίζονται από τα τελευταία προς τα πρώτα επίπεδα στην οπισθοδιάδοση μικραίνουν πάρα πολύ σε σημείο που χάνεται ουσιαστικά η πληροφορία τους. Αυτό συμβαίνει γιατί σύμφωνα με τον κανόνα της αλυσίδας όσο περισσότερο διαδίδεται στα πρώτα επίπεδα η κλίση, τόσο περισσότεροι πολλαπλασιασμοί με αριθμούς μικρότερους της μονάδας εκτελούνται οι οποίοι συνεχώς μειώνουν τους υπολογισμούς. Καταλαβαίνουμε λοιπόν ότι όσο βαθύτερο είναι ένα δίκτυο τόσο περισσότερο θα πάσχει από αυτό το πρόβλημα. Υπάρχουν πολλές μέθοδοι οι οποίες

μπορούν να βοηθήσουν σε αυτό το πρόβλημα (batch normalization, weight initialization, gradient clipping), αλλά μια πολύ αποτελεσματική παρουσιάστηκε το 2015 στο έργο "*Deep Residual Learning for Image Recognition*" (He et al., 2015) όπου εισήγαγε την έννοια των residual connections. Το ResNet, το οποίο προέκυψε από αυτό το έργο κέρδισε τον ILSVRC διαγωνισμό (μεταξύ άλλων) το 2015 επιτυγχάνοντας ποσοστό λάθους 3.57% στο ImageNet. Η φύση της ιδέας των residual connections επιτρέπει την χρήση τους όχι μόνο σε εφαρμογές κατηγοριοποίησης αλλά ούτε και της όρασης υπολογιστών γενικότερα, γεγονός που τα καθιστά μια χρήσιμη τεχνική σε όλες τις εν δυνάμει εφαρμογές των νευρωνικών δικτύων.

Residual connections

Η ιδέα πίσω από τα residual connections είναι πολύ απλή. Η είσοδος σε ένα γενικό επίπεδο βαρών συνδυάζεται με την έξοδο από αυτά. Πλέον δηλαδή η έξοδος από ένα επίπεδο βαρών αποτελεί το άθροισμα του μονοπατιού της εισόδου στα επίπεδα αυτά αλλά και της ίδιας της εισόδου.

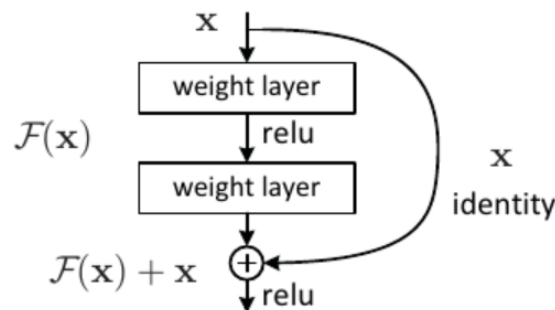


FIGURE 3.30: Η λειτουργία των residual connections

Η σύνδεση αυτή αυξάνει την συνεισφορά της εισόδου, επιτρέποντάς της να μην χάνει την επιρροή της όπως μπορεί να χανθεί στο μονοπάτι των βαρών. Προφανώς ένα μόνο επίπεδο δεν παρουσιάζει τέτοιο κίνδυνο, αλλά αξίζει να αναλογιστούμε τι συμβαίνει στην πολυπληθή εφαρμογή σειριακών μετασχηματισμών από επίπεδα βαρών. Η συνεισφορά της εισόδου μετά από πολλά τέτοια επίπεδα μειώνεται σε βαθμό κρίσιμης σημασίας. Τα βάρη, με την παρουσία των νέων συνδέσεων, μαθαίνουν να συνεισφέρουν μόνο την διαφορά της εισόδου και της εξόδου και γίνονται πιο ουσιαστικά στις αλλαγές που θα αποδώσουν. Αυτό συμβαίνει γιατί οι εναλλακτικές διαδρομές που δημιουργούνται από τα residual connections δίνουν ένα νέο πέρασμα για τις κλίσεις στο backpropagation κατά την εκπαίδευση. Η πληροφορία δεν χάνεται πλέον στο πέρασμα του κάθε επιπέδου καθώς υπάρχει ροή της από τις εναλλακτικές διαδρομές.

Όσον αφορά το U-net (ή το attention U-net), μπορούν να υλοποιηθούν residual συνδέσεις στα συνελικτικά blocks που αποτελούνται από τις δύο διαδοχικές συνελίξεις στον encoder και στον decoder του δικτύου (βλέπε εικόνα 3.31).

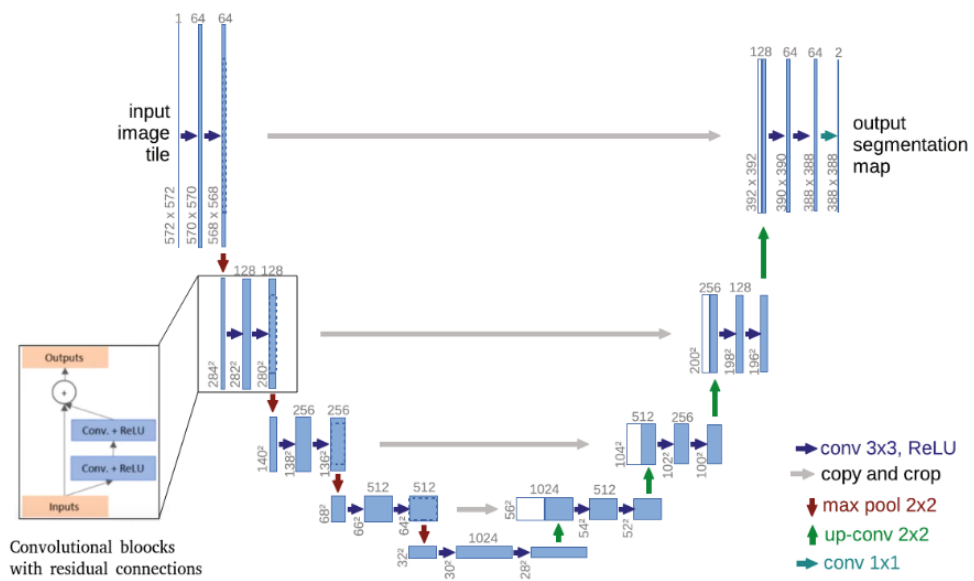


FIGURE 3.31: To U-net με residual connections

Κεφάλαιο 4

Υλοποίηση

4.1 Dataset

Το Dataset που χρησιμοποιήθηκε, είναι ένα υποσύνολο του dataset που δίνεται στον autoPET διαγωνισμό του International Conference on Medical Image Computing and Computer Assisted Intervention, MICCAI "A whole-body FDG-PET/CT dataset with manually annotated tumor lesions" (Gatidis et al., 2022). Από αυτό το dataset, έχουμε διαλέξει μόνο τις PET κανονικοποιημένες σε SUV εικόνες πρώτης εξέτασης ασθενών με μεταστατικό μελάνωμα. Οι εικόνες έχουν μέγεθος (326, 400, 400).

Οι εικόνες αποτελούνται από τις εισόδους και τα labels. Τα labels παίρνουν τις τιμές 0 και 1. Το πρόβλημα της τμηματοποίησης δηλαδή ήταν δυαδικό. Στις τιμές 1 αντιστοιχούσαν τα καρκινικά κύτταρα (ή έστω στα voxel που χαρακτηρίζονται καρκινικά) και στις τιμές 0 τα υπόλοιπα pixel. Στην εικόνα (4.1) παρουσιάζεται το Maximum Intensity Projection (MIP) μιας από τις εικόνες μαζί με τα αντίστοιχα labels.

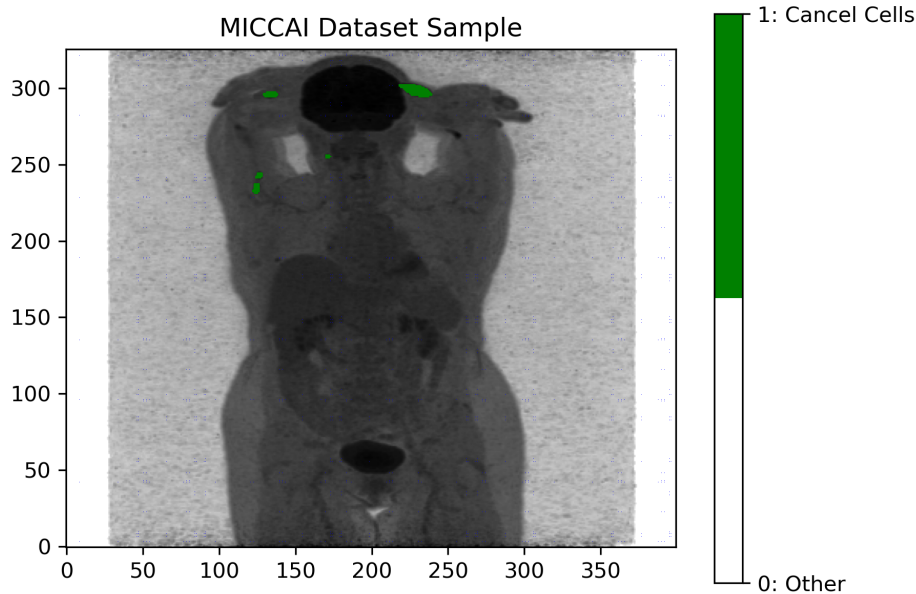


FIGURE 4.1: Ένα παράδειγμα των PET/CT εικόνων του MICCAI Dataset Sample. Παρουσιάζουμε το MIP (Maximum Intensity Projection) της εικόνας και των масκών (πράσινο στις καρκινικές περιοχές/διαφανές στις υπόλοιπες)

Σε αυτό το σημείο αναφέρουμε την κατανομή των label 1 στις εικόνες, καθώς αποτελεί μια πολύ σημαντική παράμετρο. Τα label 1 αποτελούν μια αρκετά μεμονωμένη μειωψηφία στο τελικό dataset το οποίο επιλέξαμε και διαμορφώσαμε.

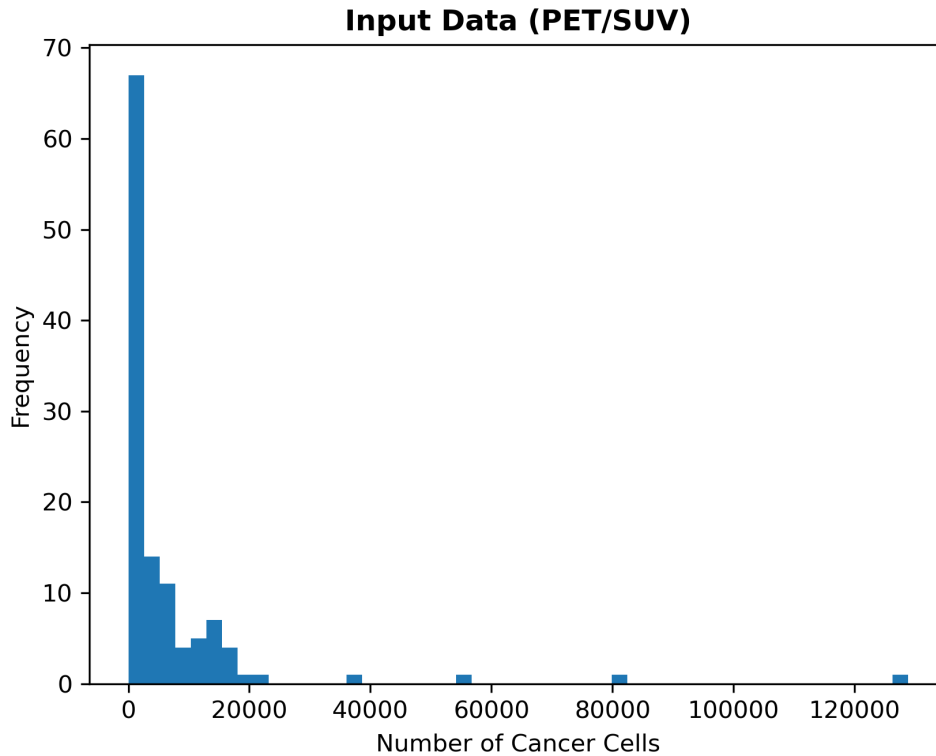


FIGURE 4.2: Η συχνότητα των καρκινικών κυττάρων στις εικόνες. Παρατηρούμε ότι οι περισσότερες εικόνες του dataset περιέχουν μικρό αριθμό καρκινικών κυττάρων

Το γεγονός αυτό καθιστά την τμηματοποίηση των εικόνων ένα αρκετά δύσκολο πρόβλημα καθώς χωρίς την απαραίτητη διαχείριση το δίκτυο θα τείνει να είναι προκατειλημμένο προς την πολυπληθή κλάση, παραμελώντας την μειωψηφία.

Στην εκπαίδευση το dataset το χωρίσαμε σε δύο σύνολα, το ένα χρησιμοποιήθηκε για την εκπαίδευση (118 εικόνες) και το άλλο για το evaluation (59 εικόνες) του δικτύου. Το σύνολο της εκπαίδευσης κατά τη διάρκεια της εκπαίδευσης επίσης χωρίστηκε 70-30.

4.2 Προεπεξεργασία εικόνων

Λόγω περιορισμένης μνήμης της GPU που χρησιμοποιήθηκε αναγκαστήκαμε να κάναμε δύο μετασχηματισμούς στις διαθέσιμες εικόνες. Ο πρώτος αφορά την αλλαγή του μεγέθους της εικόνας σε (256,256,256). Αυτός ο μετασχηματισμός έρχεται με ένα μικρό τίμημα, την απώλεια πληροφορίας σχετικής με την ανάλυση της εικόνας, όμως ήταν απαραίτητος. Ο δεύτερος μετασχηματισμός αφορά την διάσπαση της εικόνας σε μικρότερα κομμάτια (64,64,64), τα οποία ήταν αυτά που τελικά δόθηκαν σαν είσοδος στο δίκτυο. Από τα 64 μικρότερα κομμάτια (patches) που προέκυψαν, όμως, δεν δίνονταν όλα στο δίκτυο για εκπαίδευση. Ο λόγος ήταν ότι η συντριπτική πλειοψηφία αυτών για σχεδόν όλες τις εικόνες ήταν πλήρως μηδενικά. Αν δίνονταν όλα για την εκπαίδευση του

δικτύου όχι μόνο θα διαρκούσε η διαδικασία πολύ περισσότερο χρόνο, αλλά το δίκτυο θα κατέλειγε να έχει προκατάληψη (bias) προς τις μηδενικές τιμές.

Μια άλλη μέθοδος που ακολουθήθηκε ήταν αυτή του Data Augmentation. Στη μέθοδο αυτή, οι διαθέσιμες προς εκπαίδευση εικόνες υφίστανται διάφορους μετασχηματισμούς ώστε να παράξουν τεχνητές "νέες εικόνες" και τελικά την επαύξηση του διαθέσιμου dataset προς εκπαίδευση (βλέπε εικόνα 4.3). Αυτή η μέθοδος είναι ιδιαίτερα χρήσιμη, ειδικά στην περίπτωση μας όπου το dataset είναι ιδιαίτερα μικρό, με μεγάλο κίνδυνο την αδυναμία γενίκευσης του δικτύου. Οι μετασχηματισμοί που εφαρμόσαμε στις εικόνες ήταν δύο: περιστροφή και μεγέθυνση/σμίκρυνση. Οι μετασχηματισμοί αυτοί επιλέχθηκαν δεδομένης της φύσης των εικόνων που είχαμε στη διάθεσή μας. Τα κύτταρα των οργανισμών, οι ιστοί και τα όργανα των ανθρώπων παρουσιάζουν γενικότερα μια μεταβλητότητα στην φαινομενική περιστροφική τους κατεύθυνση όπως και στο μέγεθος, ειδικά αν λάβουμε υπόψη τη μεταβλητότητα που εισάγεται από την τομογραφία (κάποιες εικόνες ασθενών μπορεί να είναι πιο μεγεθυμένες από άλλες). Έτσι, οι μετασχηματισμοί που εφαρμόσαμε δημιούργησαν μια μεταβλητότητα η οποία μιμείται τη φυσική μεταβλητότητα που ούτως ή άλλως υπάρχει στις εικόνες, παράγοντας ένα επαυξημένο dataset που δεν θα απέκλινε από ένα πραγματικό μεγαλύτερο. Προφανώς αναφέρουμε ότι οι ίδιοι ακριβώς μετασχηματισμοί που έγιναν στις εικόνες εισόδου έγιναν και στα labels της εξόδου, έτσι ώστε να διατηρηθεί η πληροφορία των καρκινικών κυττάρων.

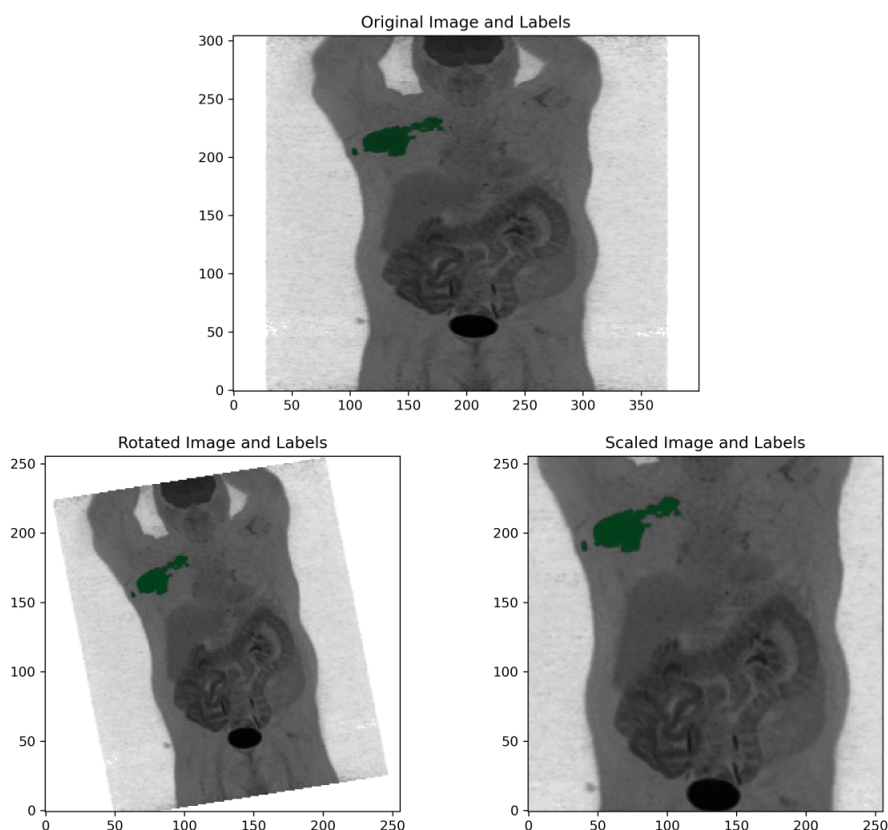


FIGURE 4.3: Οι μετασχηματισμοί που υλοποιήθηκαν για το augmentation των εικόνων του dataset: περιστροφή και μεγέθυνση/σμίκρυνση. Οι ίδιοι ακριβώς μετασχηματισμοί εφαρμόστηκαν στις εικόνες και στις μάσκες (πράσινο)

4.3 U-net, Attention U-net, Residual Attention U-net: Βασική υλοποίηση

Στην υλοποίηση έγιναν πολλά πειράματα όπου δοκιμάστηκαν διάφορες τεχνικές και διάφοροι συνδυασμοί από παραμέτρους και αρχιτεκτονικές των δικτύων. Η πειραματική προσέγγιση ήταν απαραίτητη καθώς αποτελούσε το μόνο τρόπο αξιολόγησης μιας ιδέας και της πραγματικής επιρροής της στα αποτελέσματα. Το κάθε πείραμα αλλά και το γενικό σύνολο αυτών μπορεί να καταδείξει επίσης ποιές είναι οι δυσκολίες του προβλήματος ώστε να προσπαθήσουμε να εστιάσουμε σε αυτές και να βελτιστοποιήσουμε όσο μπορούμε τα αποτελέσματα. Για τους λόγους αυτούς, θα παρουσιάσουμε κάποια από τα πειράματα αυτά και τα αποτελέσματά τους για την κάθε αρχιτεκτονική: U-net, Attention U-net, Residual Attention U-net. Τα βέλτιστα αποτελέσματα που επιτύχαμε θα παρουσιαστούν στην ενότητα της προτεινόμενης μεθόδου. Ο σχολιασμός και η σύγκριση των αποτελεσμάτων της βασικής υλοποίησης θα γίνει στο τέλος αυτής της ενότητας.

U-net

Σε αυτήν την υλοποίηση χρησιμοποιήθηκε η αρχιτεκτονική του U-net. Το δίκτυο εκπαιδεύτηκε με όλες τις διαθέσιμες εικόνες. Η συνάρτηση κόστους ήταν η focal tversky. Augmentation δεν χρησιμοποιήθηκε. Έχουμε παρατηρήσει ότι δεν επιφέρει ουσιαστική βελτίωση αν δεν εφαρμοστεί με συγκεκριμένο τρόπο. Στη συνέχεια θα δείξουμε έναν τρόπο εφαρμογής που θα επιφέρει αποτελέσματα. Το δίκτυο εκπαιδεύτηκε για 90 εποχές. Τα αποτελέσματα σε dice coefficient ήταν 0.30. Στη συνέχεια παρουσιάζουμε κάποια διαγράμματα των αποτελεσμάτων.

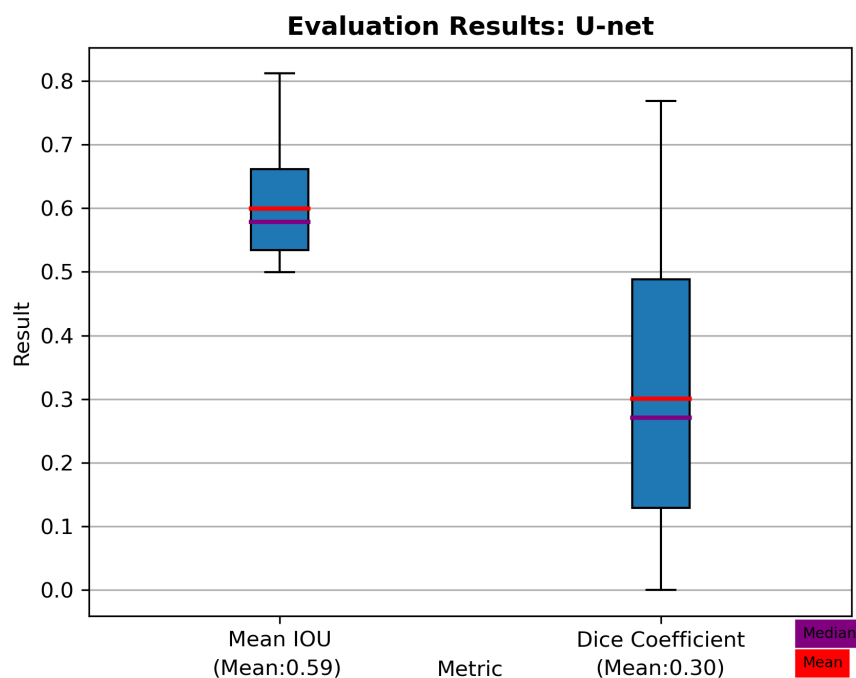


FIGURE 4.4: Τα αποτελέσματα της υλοποίησης του U-net

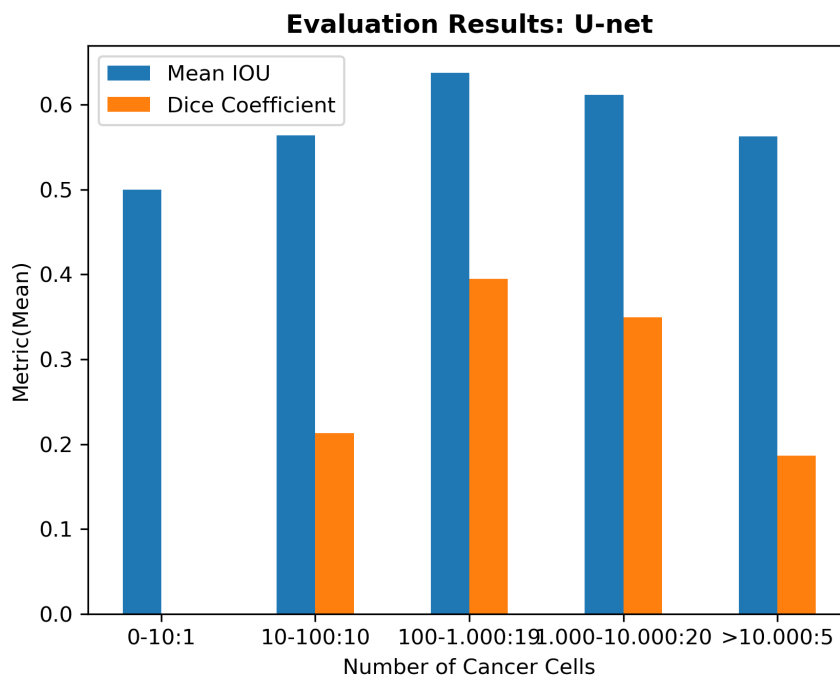


FIGURE 4.5: Τα αποτελέσματα της υλοποίησης του U-net για το κάθε διάστημα αριθμού καρκινικών κυττάρων στις εικόνες

Attention U-net

Σε αυτήν την υλοποίηση χρησιμοποιήθηκε η αρχιτεκτονική του Attention U-net. Το δίκτυο εκπαιδεύτηκε με εικόνες με αριθμό καρκινικών κυττάρων μεγαλύτερο του 100 και μικρότερο του 10000. Η συνάρτηση κόστους ήταν η dice coefficient loss (dice coefficient - 1). Augmentation δεν χρησιμοποιήθηκε. Το δίκτυο εκπαιδεύτηκε για 70 εποχές. Τα αποτελέσματα σε dice coefficient ήταν 0.31. Στη συνέχεια παρουσιάζουμε κάποια διαγράμματα των αποτελεσμάτων.

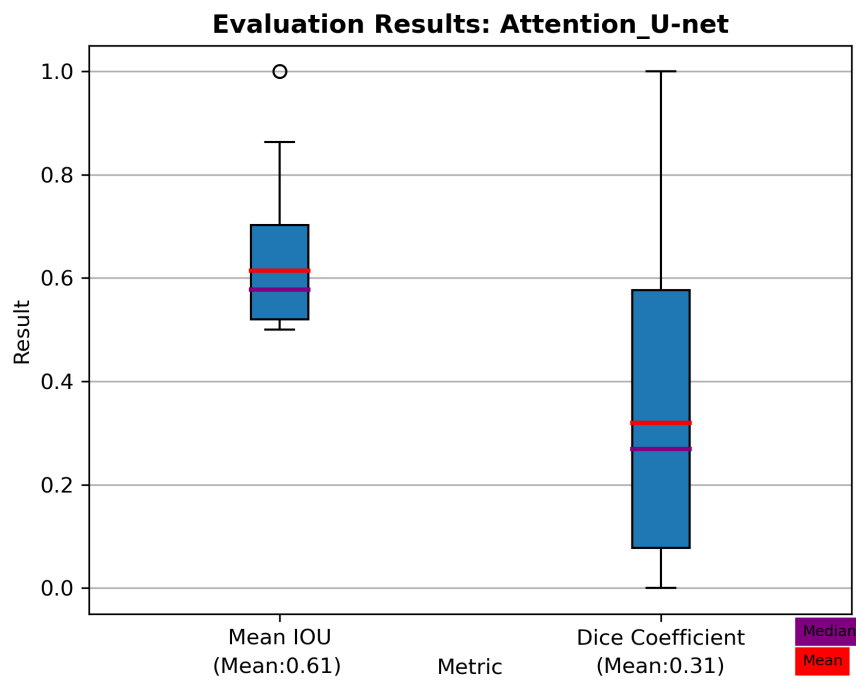


FIGURE 4.6: Τα αποτελέσματα της υλοποίησης του Attention U-net

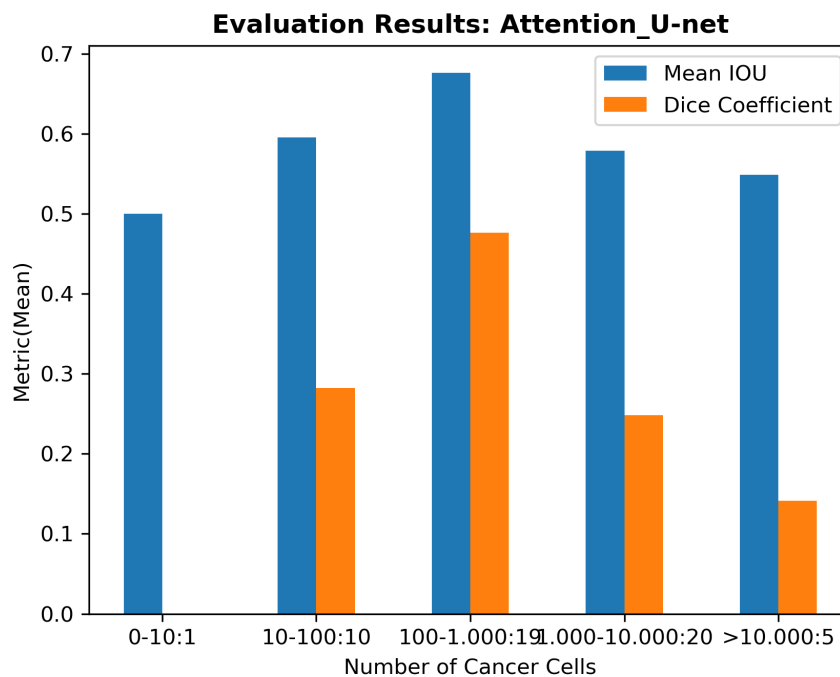


FIGURE 4.7: Τα αποτελέσματα της υλοποίησης του Attention U-net για το κάθε διάστημα αριθμού καρκινικών κυττάρων στις εικόνες

Residual Attention U-net

Σε αυτήν την υλοποίηση χρησιμοποιήθηκε η αρχιτεκτονική του Residual Attention U-net. Το δίκτυο εκπαιδεύτηκε με εικόνες με αριθμό καρκινικών κυττάρων μεγαλύτερο του 100 και μικρότερο του 10000. Η συνάρτηση κόστους ήταν η dice coefficient loss (dice coefficient - 1). Augmentation δεν χρησιμοποιήθηκε. Το δίκτυο εκπαιδεύτηκε για 70 εποχές. Τα αποτελέσματα σε dice coefficient ήταν 0.29. Στη συνέχεια παρουσιάζουμε κάποια διαγράμματα των αποτελεσμάτων.

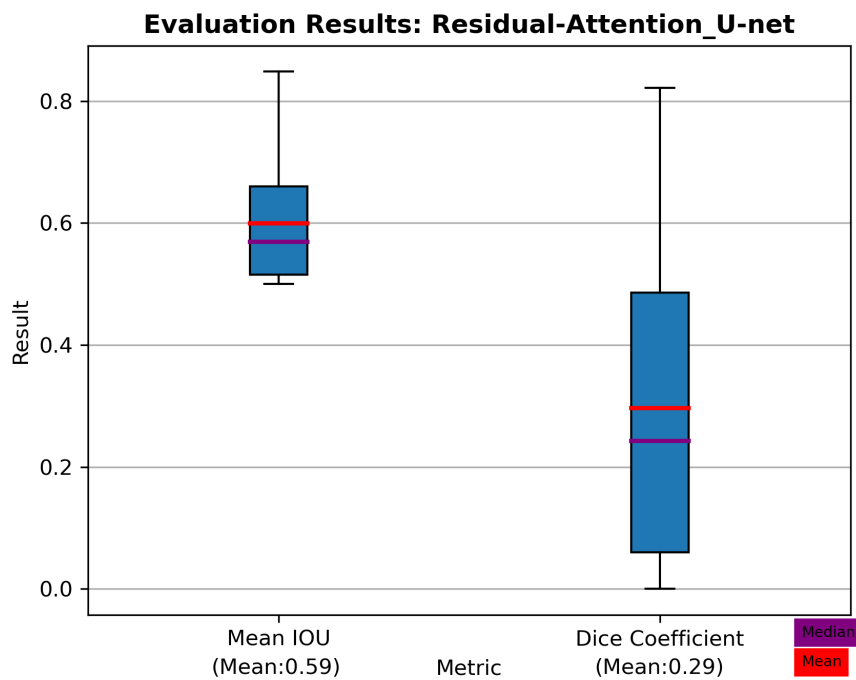


FIGURE 4.8: Τα αποτελέσματα της υλοποίησης του Residual Attention U-net

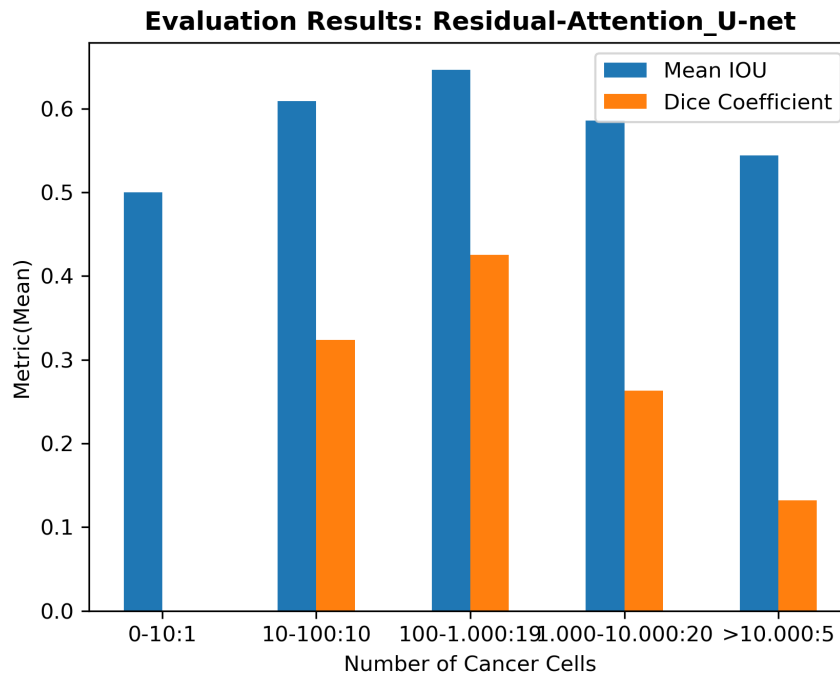


FIGURE 4.9: Τα αποτελέσματα της υλοποίησης του Residual Attention U-net για το κάθε διάστημα αριθμού καρκινικών κυττάρων στις εικόνες

Σύγκριση-Σχολιασμός αποτελεσμάτων

Με μία πρώτη ματιά τα αποτελέσματα των τριών υλοποιήσεων φαίνεται να είναι παρόμοια. Πράγματι είναι παραπλήσια, αλλά παρουσιάζουν κάποιες ενδιαφέρουσες διαφορές, ειδικά στα αποτελέσματα με τα διαιρεμένα εύρη με βάση τον αριθμό καρκινικών κυττάρων.

Το πρώτο πράγμα που παρατηρούμε είναι ότι το U-net εμφανίζει λίγο πιο ισορροπημένα αποτελέσματα από τις άλλες υλοποιήσεις. Πετυχαίνει σχετικά υψηλά αποτελέσματα και στις εικόνες με εύρος 100-1000 και στις εικόνες με εύρος 1000-10000 καρκινικών κυττάρων. Ειδικά στο δεύτερο σύνολο πετυχαίνει τα καλύτερα αποτελέσματα. Η ψαλίδα στο boxplot φαίνεται να είναι επίσης η μικρότερη, πετυχαίνοντας τα μεγαλύτερα ελάχιστα αποτελέσματα και τα μικρότερα μέγιστα. Μέρος της επίδρασης αυτής έχει και η χρήση της focal tversky σαν συνάρτηση κόστους η οποία επιτρέπει την εξομάλυνση των ανισοτήτων των κλάσεων και τον έλεγχο στην παρουσία των false positives έναντι των false negatives. Η καλή επίδραση του U-net στο εύρος 1000-10000 οφείλεται στον σχετικά μικρό αριθμό από false negatives, όπως επιλέξαμε με τις υπερπαραμέτρους της focal tversky.

Οι άλλες υλοποιήσεις παρατηρούμε να παρουσιάζουν καλύτερες αποδόσεις σε συγκεκριμένα εύρη με μεγαλύτερα μέγιστα αποτελέσματα και μικρότερα ελάχιστα. Σημειώστε σε αυτή τη φάση ότι το attention U-net τείνει να έχει την καλύτερη επίδραση στο εύρος 100-1000 και το residual attention U-net τείνει να έχει την καλύτερη επίδραση στο εύρος 10-100. Ο λόγος για τις δύο αυτές αποδόσεις είναι ότι το attention U-net τείνει να βρίσκει καλύτερα τα καρκινικά κύτταρα (true positives) και το residual attention U-net τείνει να παρουσιάζει λιγότερα false positives. Στο κάθε εύρος καρκινικών κυττάρων οι αντίστοιχες αποδόσεις έχουν τη μεγαλύτερη επίδραση. Αυτά τα χαρακτηριστικά των δικτύων θα αξιοποιηθούν στη συνέχεια, στην προτεινόμενη μέθοδο.

Παρατηρούμε ότι μια αδυναμία όλων των δικτύων είναι η επίδοση στις εικόνες με μεγάλο αριθμό καρκινικών κυττάρων. Η μεγάλη παρουσία εικόνων με πολύ λίγα καρκινικά κύτταρα σε σύγκριση με αυτές που έχουν πιο πολλά έχουν σαν αποτέλεσμα τα δίκτυα

να είναι προκατειλημμένα στην αρνητική κλάση (όχι καρκινικά). Η επίδραση αυτής της τάσης είναι να χάνονται πολλά καρκινικά κύτταρα στις προβλέψεις (false negatives). Όσο περισσότερο είναι το ποσοστό των καρκινικών κυττάρων στις εικόνες τόσο χειρότερη θα είναι η απόδοση ενός δικτύου προκατειλημμένου στην άλλη κλάση. Αυτό αποτελεί ένα από τα μεγαλύτερα προβλήματα της υλοποίησής μας και θα δείξουμε στην προτεινόμενη μέθοδο πώς θα το αντιμετωπίσουμε.

4.4 Προτεινόμενη μέθοδος

Οι απλές προσεγγίσεις που παρουσιάσαμε κατέδειξαν αποτελέσματα που δεν είναι ικανοποιητικά. Οι προκλήσεις που εμφανίζονται και οι περιορισμοί που υπάρχουν απαιτούν μια διαφορετική αντιμετώπιση. Όλες οι υλοποιήσεις αποτυγχάνουν να εντοπίσουν πολλές περιοχές καρκινικών κυττάρων ενώ παράλληλα παρατηρούμε την λανθασμένη κατηγοριοποίηση όχι λίγων περιοχών ως καρκινικές, σε δίκτυα που κατα τ' άλλα είναι βεβιασμένα προς την αρνητική κλάση. Τα δύο αυτά προβλήματα αποτελούν τον κεντρικό πυρήνα της υλοποίησής μας και η κατανόηση της πηγής τους μπορεί να συμβάλει στην βελτίωση των αποτελεσμάτων.

Η αποτυχία εντοπισμού περιοχών καρκινικών κυττάρων (false negatives) οφείλεται στην μεγάλη ανισορροπία των κλάσεων στα δεδομένα της εκπαίδευσης. Το δίκτυο τελικά καταλήγει να είναι προκατειλημμένο στην αρνητική κλάση και έτσι αποτυγχάνει να εντοπίσει πολλές καρκινικές περιοχές. Για την καταπολέμηση αυτού του προβλήματος δοκιμάσαμε δύο στρατηγικές: την χρήση της focal tversky σαν συνάρτηση κόστους και την εξομάλυνση του dataset. Όπως αναφέραμε παραπάνω η focal tversky αποτελεί μια συνάρτηση κόστους η οποία αυξάνει την επιρροή της κλάσης μειοψηφίας. Η εξομάλυνση του dataset έχει σαν σκοπό να φέρει μια ισορροπία στην παρουσία των δύο κλάσεων και μπορεί να γίνει με δύο τρόπους: την αφαίρεση εικόνων που παρουσιάζουν τεράστια ανισότητα στις κλάσεις τους και το επιλεκτικό augmentation μόνο για εικόνες οι οποίες παρουσιάζουν μεγάλο αριθμό από την κλάση μειοψηφία. Στην υλοποίηση (Attention U-net: Εύρεση όλων των όγκων) που θα παρουσιάσουμε παρακάτω εφαρμόσαμε όλες αυτές τις μεθόδους στο άκρο έτσι ώστε να καταφέρουμε να έχουμε μια υλοποίηση που δεν χάνει κανένα καρκινικό κύτταρο (με τίμημα την παρουσία όμως πολλών επίσης false positives).

Η παρουσία περιοχών λανθασμένης κατηγοριοποίησης ως καρκινικών (false positives) μπορεί να οφείλεται σε πολλούς παράγοντες. Ένας μπορεί να είναι η προκατάληψη του δικτύου μετά από εκπαίδευση σε δεδομένα που παρουσιάζουν μεγαλύτερα ποσοστά της θετικής κλάσης έναντι της αρνητικής. Για παράδειγμα, στην παραπάνω μέθοδο της εξομάλυνσης του dataset αν αφαιρέσουμε πολύ το στοιχείο της αρνητικής κλάσης και προσθέσουμε πολύ το στοιχείο της θετικής θα καταλήξουμε να έχουμε αρκετά false positives. Ένας άλλος παράγοντας μπορεί να είναι η παρόμοια φύση των μοτίβων στα δεδομένα εισόδου και πρόβλεψης. Μπορεί να υπάρχουν κύτταρα στις εικόνες που μπορεί να μοιάζουν πολύ με καρκινικά αλλά να μην είναι. Το δίκτυο αυτά μπορεί να τα κατηγοριοποιήσει λανθασμένα ως καρκινικά, όπως ίσως ένας άνθρωπος που δεν γνωρίζει καλά τις διαφορές των κυττάρων. Περισσότερα για αυτήν την κατηγορία των false positives θα αναφέρουμε στη συνέχεια, αλλά σε αυτή τη φάση αξίζει να παρατηρηθεί ότι στις εικόνες που παρουσιάστηκαν στις βασικές υλοποιήσεις υπάρχουν πάντα false positives στην περιοχή της κύστης των ασθενών.

Η ταυτόχρονη μείωση και των false positives και των false negatives είναι ένας στόχος απαραίτητος αλλά παράλληλα δύσκολος. Η βελτιστοποίηση του ενός φαίνεται να έρχεται στο κόστος του άλλου. Έτσι επιλέξαμε δεδομένων αυτών να βελτιστοποιήσουμε

το κάθε πρόβλημα ξεχωριστά, μιας και είναι πιο εύκολο, και στη συνέχεια να συνδυάσουμε με κάποιο τρόπο τα θετικά της κάθε υλοποίησης σε μια. Ο συγκεκριμένος τρόπος της υλοποίησης της βελτιστοποίησης των false positives, των false negatives και του συνδυασμού αυτών θα παρουσιαστεί στις επόμενες ενότητες.

4.4.1 Attention U-net: Εύρεση όλων των όγκων

Για την εύρεση όλων των όγκων (μείωση των false negatives) ακολουθήσαμε μια μέθοδο όπου γύραμε την αντιπροσώπευση του dataset προς την θετική κλάση των καρκινικών κυττάρων. Αυτό το επιτύχαμε αφαιρώντας τις εικόνες όπου είχαν λιγότερα από 2000 καρκινικά κύτταρα και κάνοντας augmentation μόνο στις εικόνες με τον μεγαλύτερο αριθμό καρκινικών κυττάρων. Μετά την αφαίρεση των εικόνων με λίγα καρκινικά κύτταρα οι εικόνες της εκπαίδευσης έγιναν 43 και το augmentation έγινε για το 10% των εικόνων με μεγαλύτερο αριθμό καρκινικών κυττάρων. Ο τελικός αριθμός των εικόνων έγινε 64. Το δίκτυο εκπαιδεύτηκε για 90 εποχές. Σαν συνάρτηση κόστους χρησιμοποιήθηκε η focal tversky. Βελτιστοποιητής ήταν ο Adam. Τέλος, το δίκτυο που χρησιμοποιήθηκε ήταν το Attention U-net.

Συνολικά η απόδοση του δικτύου δεν αυξήθηκε, καθώς η μέση τιμή του dice coefficient για όλες τις εικόνες ήταν 0.23. Παρόλα αυτά παρατηρούμε μεγάλη βελτίωση στις μετρικές για τις εικόνες με μεγάλο αριθμό καρκινικών κυττάρων, όπου στις βασικές υλοποιήσεις πετυχαίναμε τα χειρότερα ποσοστά, με μέσο dice coefficient να ξεπερνάει το 0.5.

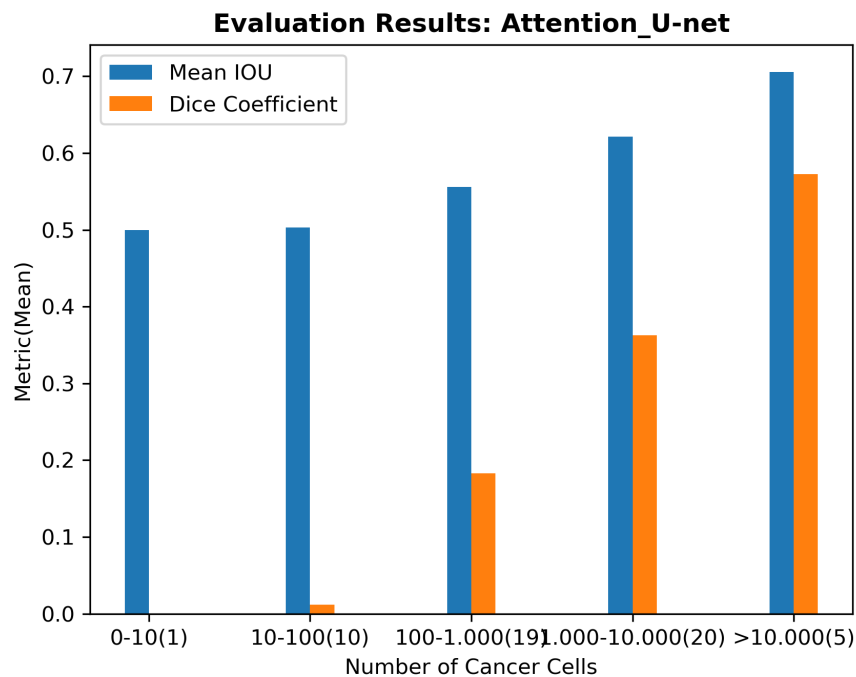


FIGURE 4.10: Τα αποτελέσματα της υλοποίησης του Attention U-net εύρεσης όλων των όγκων για το κάθε διάστημα αριθμού καρκινικών κυττάρων στις εικόνες

Το tradeoff της αύξησης των false positives γίνεται φανερό από την μείωση της μετρικής στις εικόνες με μικρό αριθμό καρκινικών κυττάρων. Σε αυτές τις εικόνες οι όγκοι βρίσκονται αλλά μειώνεται η μετρική λόγω της παρουσίας των false positives.

Παρουσιάζουμε στη συνέχεια και κάποιες ενδεικτικές εικόνες με τις αντίστοιχες μάσκες των labels και των προβλέψεων.

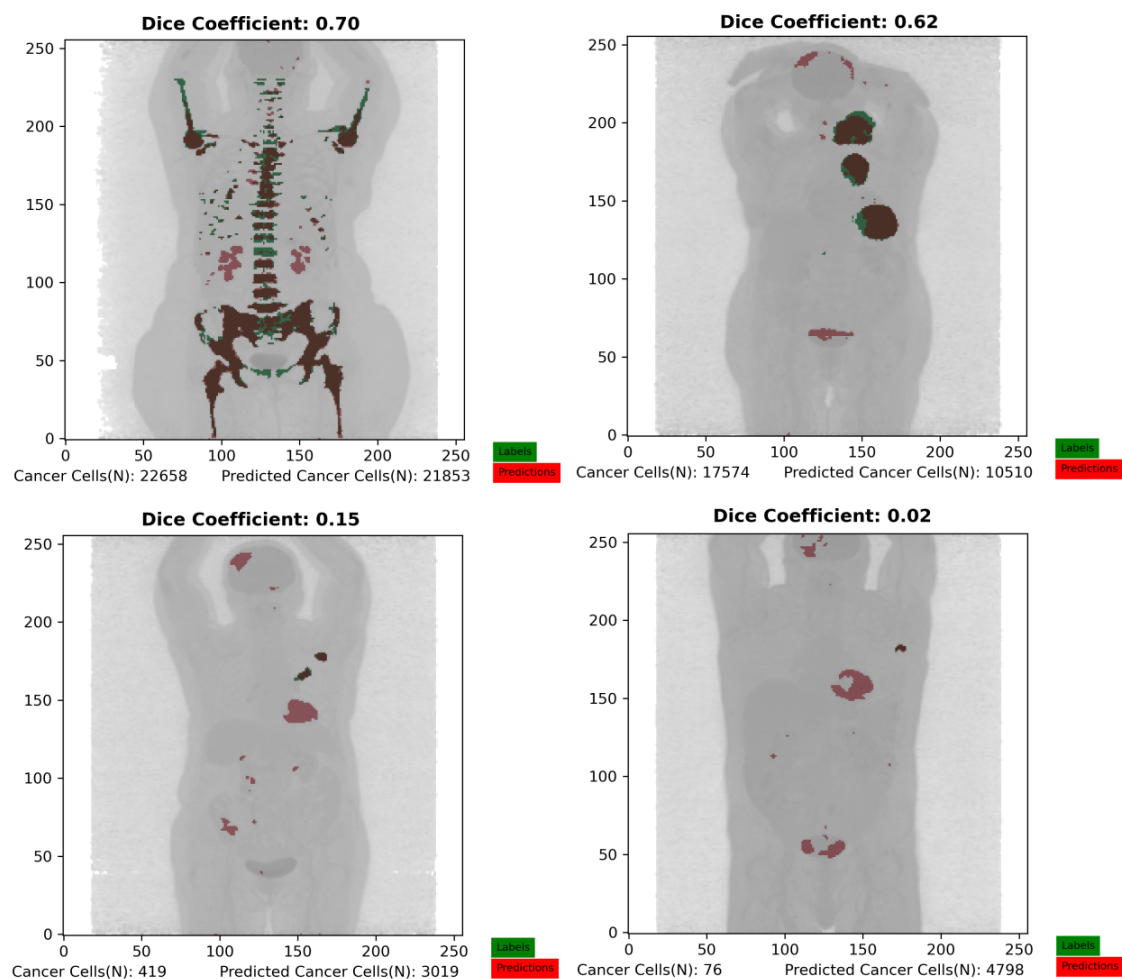


FIGURE 4.11: Κάποιες εικόνες της υλοποίησης του Attention U-net εύρεσης όλων των όγκων. Με κόκκινο παρουσιάζονται οι μάσκες της πρόβλεψης του δικτύου. Με πράσινο οι πραγματικές τιμές. Το σκούρο κόκκινο χρώμα στις μάσκες καταδεικνύει τις προβλέψεις που ήταν σωστές

4.4.2 Residual Attention U-net: Μείωση των False Positives

Για την μείωση των false positives ακολουθήσαμε την προηγούμενη μέθοδο με την διαφορά ότι αυτή τη φορά δεν αφαιρέσαμε τις εικόνες με μικρό αριθμό καρκινικών κυττάρων. Αυτό το κάναμε έτσι ώστε αυτή τη φορά το dataset να είναι πιο ισορροπημένο. Ο στόχος αυτής της υλοποίησης είναι να καταφέρει να βρει όλους τους όγκους έστω σε ένα ποσοστό με την μεγαλύτερη δυνατή ελαχιστοποίηση των false positives. Ο αριθμός των εικόνων της εκπαίδευσης ήταν 87 πριν το augmentation ενώ μετά το augmentation έγινε 130. Το δίκτυο εκπαιδεύτηκε για 190 εποχές. Σαν συνάρτηση κόστους χρησιμοποιήθηκε η focal tversky. Βελτιστοποιητής ήταν ο Adam. Τέλος, το δίκτυο που χρησιμοποιήθηκε ήταν το Residual Attention U-net. Παρατηρήσαμε ότι είχε λίγο καλύτερες επιδόσεις και ίσως αυτό να οφείλεται στον μεγαλύτερο αριθμό εικόνων στην εκπαίδευση.

Συνολικά η απόδοση του δικτύου ούτε αυτή τη φορά ήταν μεγάλη. Η μέση τιμή του dice coefficient για όλες τις εικόνες ήταν 0.26. Ο στόχος όμως της υλοποίησης αυτής δεν ήταν αυτός. Όπως προαναφέραμε ο στόχος αυτής της υλοποίησης είναι να καταφέρει να βρει όλους τους όγκους έστω σε ένα ποσοστό με την μεγαλύτερη δυνατή ελαχιστοποίηση των false positives. Αυτό επετεύχθη, όχι όμως με τα ίδια ποσοστά επιτυχίας στην προηγούμενη υλοποίηση. Στη συνέχεια θα αναφέρουμε τους λόγους της απόδοσης αυτής όπως και προτάσεις για βελτίωση. Η μείωση των false positives αντικατοπτρίζεται στα αποτελέσματα των εικόνων με λίγα καρκινικά κύτταρα. Πιο συγκεκριμένα οι εικόνες με αριθμό καρκινικών από 100 μέχρι 1000 έχουν μέσο dice coefficient πάνω από 0.4, αρκετά μεγαλύτερο από την προηγούμενη υλοποίηση. Στα αποτελέσματα επίσης αντικατοπτρίζεται η αδυναμία της ολικής εξάλειψης των false positives. Αν τα false positives είχαν εξλειφθεί οι εικόνες με χαμηλό αριθμό καρκινικών κυττάρων θα έπρεπε να έχουν ποσοστά υψηλά καθώς αν ο πραγματικός αριθμός των καρκινικών κυττάρων είναι μικρός, η απόδοση θα εξαρτάται όλο και περισσότερο από τον αριθμό των false positives. Παρατηρούμε όμως ότι αυτό στις εικόνες με αριθμό καρκινικών κυττάρων από 10 έως 100 αυτό δεν συμβαίνει καθώς το dice coefficient είναι μικρότερο από 0.05.

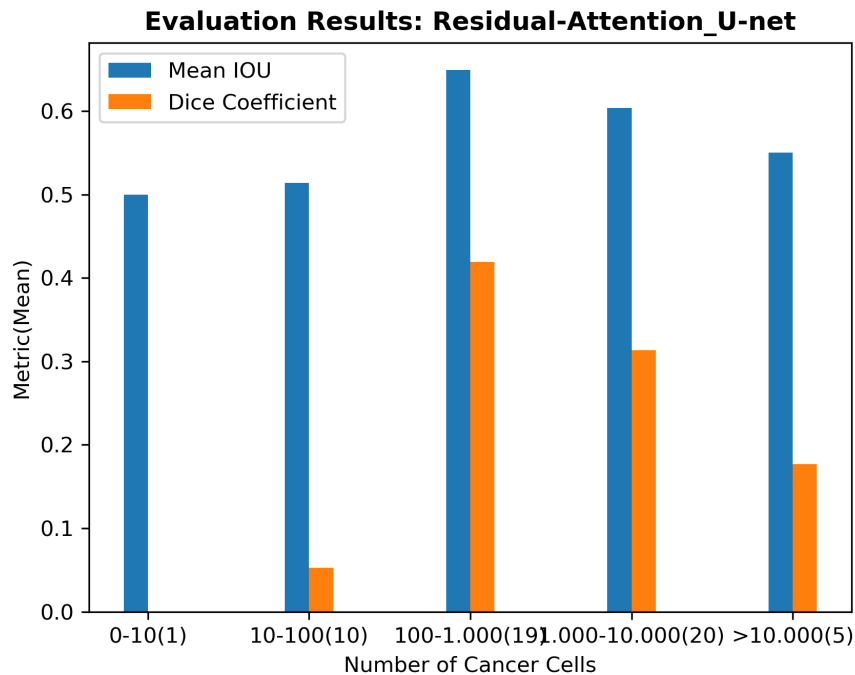


FIGURE 4.12: Τα αποτελέσματα της υλοποίησης του Residual Attention U-net μείωσης των false positives για το κάθε διάστημα αριθμού καρκινικών κυττάρων στις εικόνες

Στη συνέχεια παρουσιάζουμε και κάποιες ενδεικτικές εικόνες (τις ίδιες με την προηγούμενη υλοποίηση για λόγους σύγκρισης).

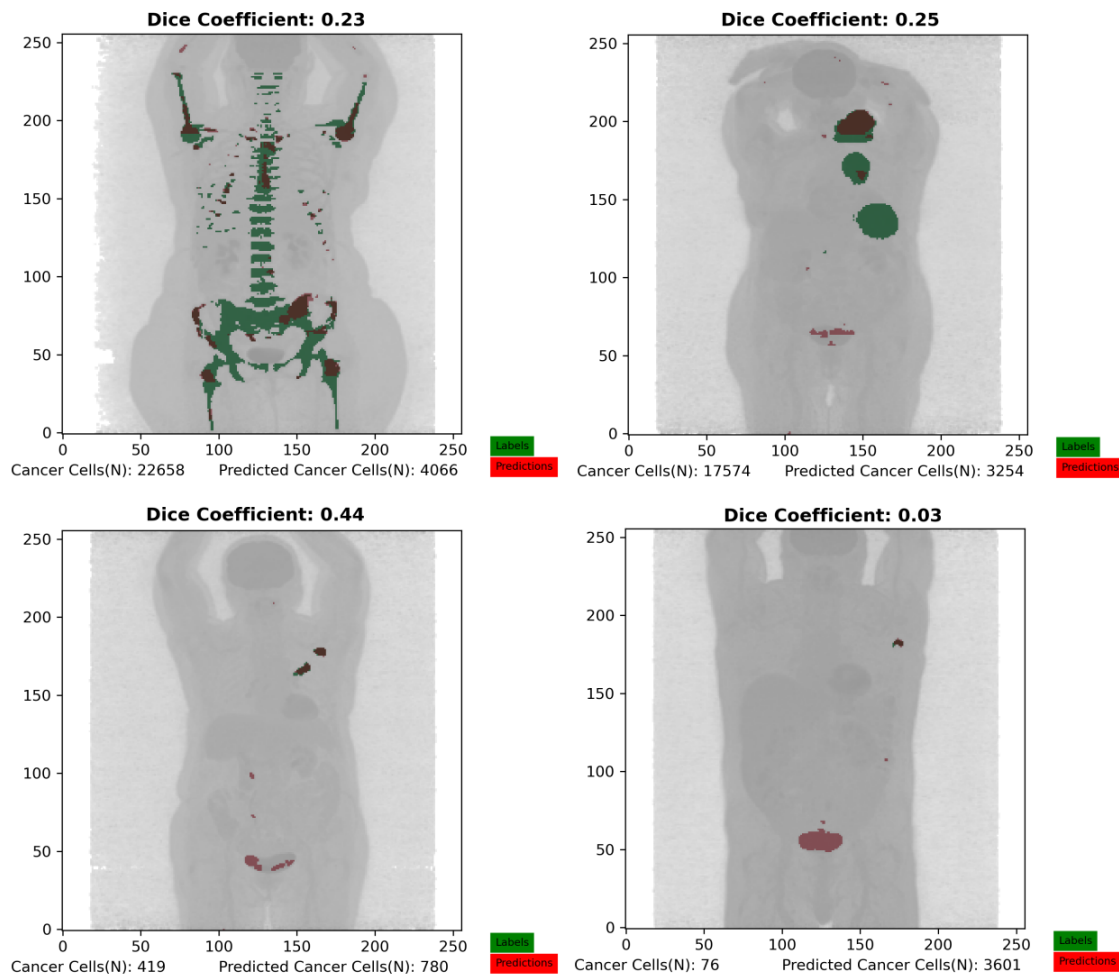


FIGURE 4.13: Κάποιες εικόνες της υλοποίησης του Residual Attention U-net μείωσης των false positives. Με κόκκινο παρουσιάζονται οι μάσκες της πρόβλεψης του δικτύου. Με πράσινο οι πραγματικές τιμές. Το σκούρο κόκκινο χρώμα στις μάσκες καταδεικνύει τις προβλέψεις που ήταν σωστές

Σε αυτό το σημείο αξίζει να παρατηρήσουμε συγκεκριμένα τα αποτελέσματα της τέταρτης εικόνας με τα 76 καρκινικά κύτταρα. Σε αυτήν την υλοποίηση καταφέραμε να αφαιρέσουμε τα false positives στον εγκέφαλο και την καρδιά σε σύγκριση με την προηγούμενη, με μείωση αυτών σε ποσοστό περίπου 33%. Παρόλα αυτά η μετρική είναι μόλις 0.03 ενώ η προηγούμενη ήταν 0.02. Οι μετρικές παραμένουν τόσο μικρές επειδή ο αριθμός των labels με τιμή 1 είναι τόσο μικρός που ακόμα και ο μειωμένος αριθμός των false positives είναι σχετικά αρκετά μεγαλύτερος. Τελικά, ενώ επιτύχαμε μείωση των false positives κατά 33% δεν αντικατοπτρίζεται η ουσιαστική βελτίωση στις μετρικές.

4.4.3 Ensemble Learning: Συνδυασμός αποτελεσμάτων

Στις δύο προηγούμενες ενότητες παρουσιάσαμε δυο υλοποιήσεις όπου η μια κατάφερε να βρει όλα τα καρκινικά κύτταρα αλλά με την παρουσία αρκετών false positives και η άλλη προσπάθησε να μειώσει τα false positives βρίσκοντας όσο περισσότερα καρκινικά κύτταρα όσο γινόταν. Στην ενότητα αυτή θα παρουσιάσουμε τον τρόπο με τον οποίο προσπαθήσαμε να συνδυάσουμε τα αποτελέσματα αυτά με έναν τρόπο που θα κρατήσει τα θετικά της κάθε υλοποίησης: λίγα false positives και λίγα false negatives.

Στην ενότητα που παρουσιάσαμε την υλοποίηση της ελαχιστοποίησης των false positives αναφέραμε ότι ο στόχος ήταν να ελαχιστοποιήσουμε τα false positives βρίσκοντας όσο περισσότερους όγκους έστω σε ένα ποσοστό. Το δεύτερο κομμάτι του στόχου είναι σημαντικό. Ο τρόπος με τον οποίο θα συνδυάσουμε τα παραπάνω αποτελέσματα είναι να πάρουμε τις προβλέψεις της υλοποίησης που βρίσκει όλους τους όγκους, να εξάγουμε τους όγκους αυτούς σαν αυτοτελή κομμάτια και αν για το καθένα από αυτά το δεύτερο μοντέλο παράξει επίσης έναν αριθμό καρκινικών κυττάρων πάνω από ένα κατώφλι να το θεωρήσουμε σαν όγκο, αλλιώς να το θεωρήσουμε σαν μη όγκο. Η προσέγγιση αυτή δεν είναι τέλεια καθώς μπορεί το δεύτερο μοντέλο να μην έχει παράξει αρκετά καρκινικά κύτταρα για μια περιοχή όγκου ώστε να ξεπεραστεί το κατώφλι και έτσι να χαθεί μια περιοχή όγκου ή μπορεί να θεωρηθεί λανθασμένα μια περιοχή σαν όγκος αν τα false positives της δεύτερης υλοποίησης ξεπερνούν το κατώφλι. Σε κάθε περίπτωση γίνεται φανερό ότι τα αποτελέσματα του συνδυασμού αυτού εξαρτώνται από τα αποτελέσματα των προηγούμενων υλοποιήσεων. Εμείς συγκεκριμένα καταφέραμε να τα πάμε πολύ καλά στην πρώτη υλοποίηση και λιγότερο στη δεύτερη. Το κατώφλι που χρησιμοποιήσαμε είχε την τιμή 5. Δηλαδή αν σε μια υποψήφια περιοχή υπάρχουν πάνω από 5 καρκινικά κύτταρα στην δεύτερη υλοποίηση την θεωρούμε καρκινική.

Η απόδοση αυτής της υλοποίησης είναι αισθητά βελτιωμένη από όλα τα προηγούμενα αποτελέσματα. Με μια πρώτη ματιά στις συνολικές μετρικές μπορεί να μην είναι προφανές. Το μέσο dice coefficient για όλες τις εικόνες είναι 0.33. Παραθέτουμε και ένα boxplot για μια καλύτερη εποπτεία της απόδοσης σε όλες τις εικόνες του evaluation.

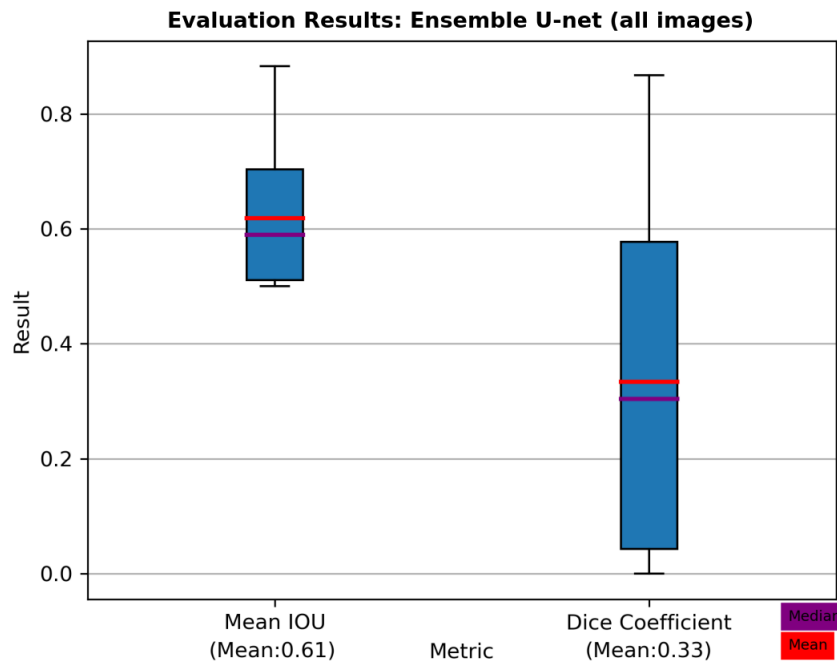


FIGURE 4.14: Τα αποτελέσματα της υλοποίησης του Ensemble U-net για όλες τις εικόνες

Η βελτίωση όμως των αποτελεσμάτων γίνεται φανερή αν παραθέσουμε τις μετρικές για το κάθε εύρος του αριθμού των καρκινικών κυττάρων στις εικόνες.

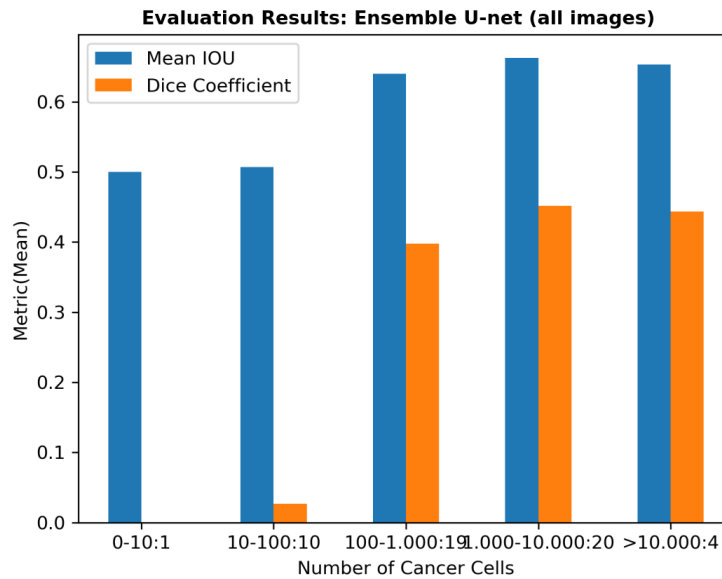


FIGURE 4.15: Τα αποτελέσματα της υλοποίησης του Ensemble U-net για το κάθε διάστημα αριθμού καρκινικών κυττάρων στις εικόνες

Παρατηρούμε δηλαδή ότι για τις εικόνες που έχουν έναν αντιπροσωπευτικό αριθμό καρκινικών κυττάρων (πάνω από 100) και αποφεύγεται η άνιση σύγκριση των false positives και των πολύ λίγων διαθέσιμων true positives τα αποτελέσματα χαρακτηρίζονται από μέσο dice coefficient 0.42. Παραθέτουμε και ένα boxplot για τις αποδόσεις του δικτύου μόνο για τις εικόνες που παρουσιάζουν αριθμό καρκινικών κυττάρων μεγαλύτερο από 100.

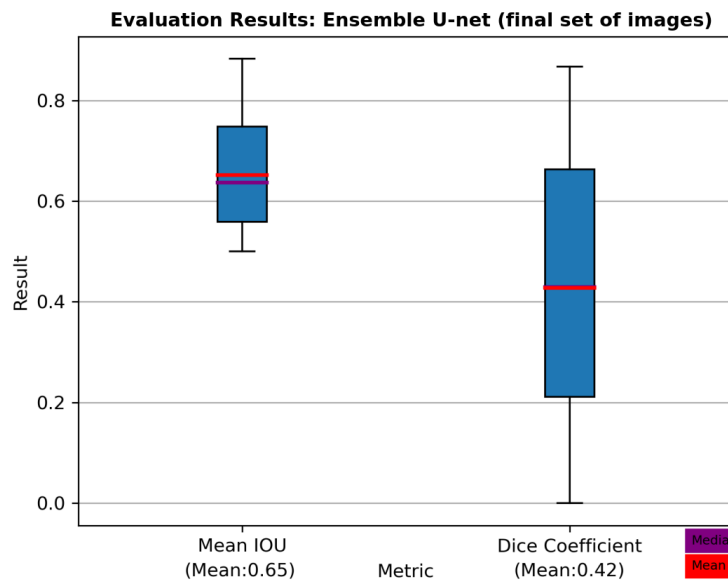


FIGURE 4.16: Τα αποτελέσματα της υλοποίησης του Ensemble U-net για το τελικό σύνολο εικόνων (εικόνες με αριθμό καρκινικών κυττάρων μεγαλύτερο του 100)

Παραθέτουμε επίσης και τις εικόνες με τις μάσκες που είχαμε παραθέσει και στις προηγούμενες υλοποιήσεις, για να δούμε την τελική απόδοση του δικτύου.

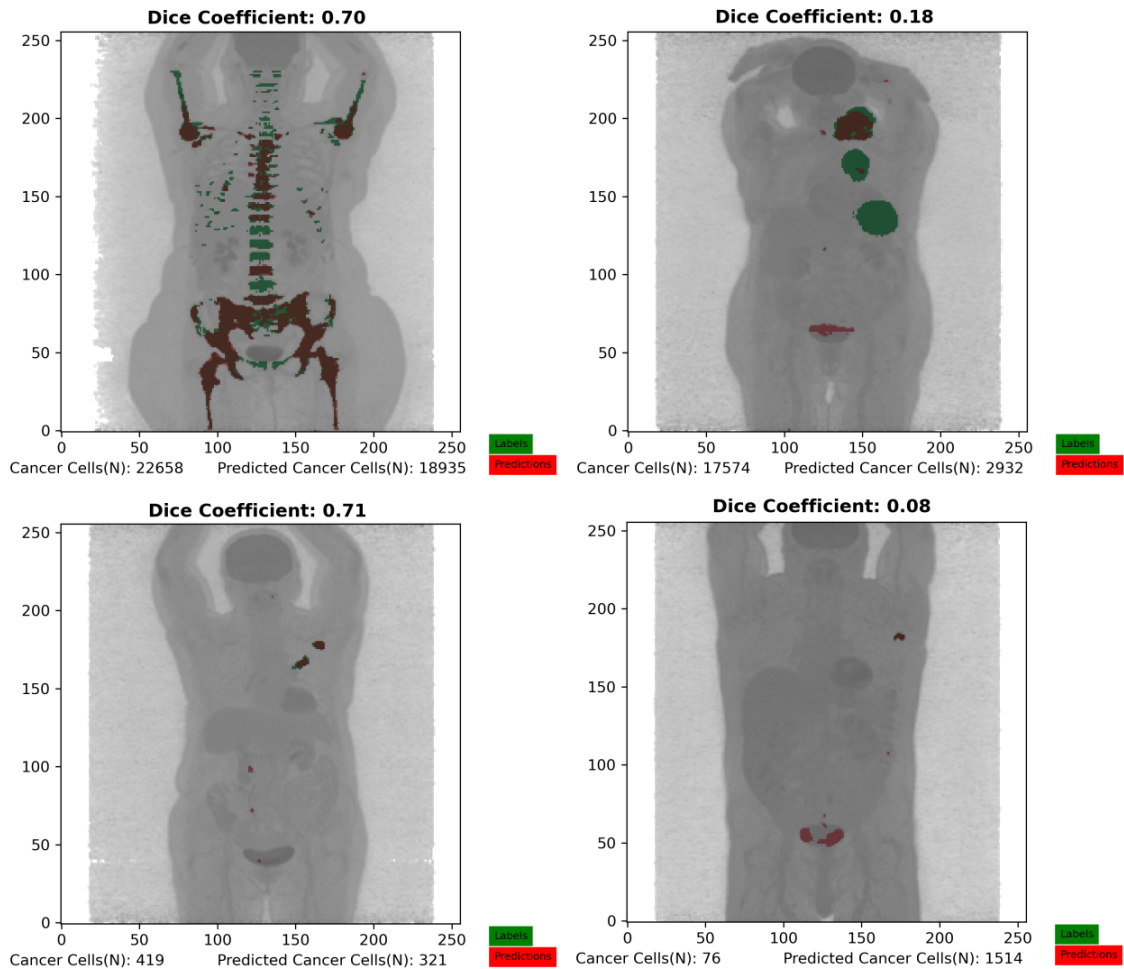


FIGURE 4.17: Κάποιες εικόνες της υλοποίησης του Ensemble U-net. Με κόκκινο παρουσιάζονται οι μάσκες της πρόβλεψης του δικτύου. Με πράσινο οι πραγματικές τιμές. Το σκούρο κόκκινο χρώμα στις μάσκες καταδεικνύει τις προβλέψεις που ήταν σωστές

Τέλος παρουσιάζουμε τα αποτελέσματα όλων των υλοποιήσεων συνοπτικά.

TABLE 4.1: Αποτελέσματα των υλοποιήσεων

Mean Dice Coefficient	U-net	Attention U-net	Residual Attention U-net	Εύρεση όλων των όγκων	Μείωση False Positives	Ensemble U-net
All images set	0.30	0.31	0.29	0.23	0.26	0.33
Final set	0.35	0.33	0.31	-	-	0.42

4.5 Σχολιασμός αποτελεσμάτων

Η προτεινόμενη μέθοδος προσπάθησε να συνδυάσει τα θετικά χαρακτηριστικά επιμέρους υλοποιήσεων οι οποίες επιτυγχάνουν έναν στόχο αλλά αποτυγχάνουν σε έναν άλλο. Η υλοποίηση η οποία ελαχιστοποιεί τα false negatives για να βρει όλους τους όγκους έχει σαν αρνητικό την μεγάλη παρουσία από false negatives. Το δίκτυο (Attention U-net: Εύρεση όλων των όγκων) φαίνεται να προκατειλημένο προς τη θετική κλάση των καρκινικών κυττάρων. Όταν προσπαθήσαμε να αλλάξουμε την ισορροπία των κλάσεων στην εκπαίδευση για την μείωση των false positives χάναμε όγκους πριν την μείωσή τους (των false positives). Αυτό συνέβαινε σε όλες τις υλοποιήσεις. Αυτός ήταν ο λόγος που επιλέξαμε να προσπαθήσουμε να μειώσουμε τα false positives, με τίμημα τη μείωση των false negatives, μέχρι το σημείο που οι όγκοι βρίσκονται έστω σε ένα ποσοστό ώστε συνδυαστικά μετά να τους βρούμε με τον τρόπο που αναφέραμε. Στις εικόνες (4.18) παρατηρούμε κάποια χαρακτηριστικά παραδείγματα της υλοποίησης αυτής.

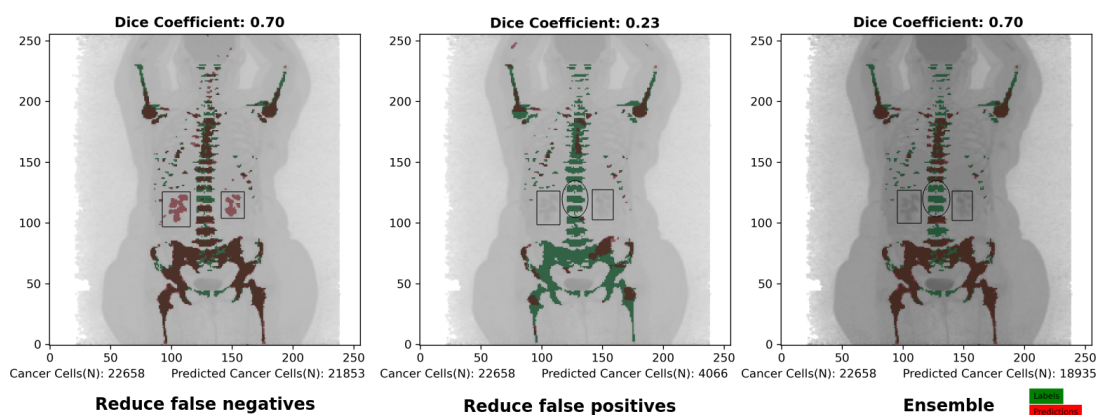


FIGURE 4.18: Κάποιες εικόνες της υλοποίησης του Ensemble U-net. Με κόκκινο παρουσιάζονται οι μάσκες της πρόβλεψης του δικτύου. Με πράσινο οι πραγματικές τιμές. Το σκούρο κόκκινο χρώμα στις μάσκες καταδεικνύει τις προβλέψεις που ήταν σωστές. Η περιοχή των νεφρών (ορθογώνια παραλληλόγραμμα) σωστά καταλήγει να μην προβλέπεται σαν καρκινική, ενώ κάποια σημεία της σπονδυλικής στήλης (ελλείψεις) χάνονται

Στις παραπάνω εικόνες παρατηρούμε ότι στην *Reduce false negatives* ενώ έχουν βρεθεί τα καρκινικά κύτταρα στο μεγαλύτερό τους μέρος έχουμε την παρουσία κάποιων false positives (περικλείονται από τα ορθογώνια παραλληλεπίπεδα). Αυτά στην υλοποίηση των *Reduce false positives* κατηγοριοποιούνται σωστά σαν μη καρκινικά κύτταρα. Χάνονται βέβαια κάποιες περιοχές (περικλείονται από τις ελλείψεις) αφού κατηγοριοποιούνται λανθασμένα ως μη καρκινικές. Αυτός είναι και ο λόγος που η Ensemble υλοποίηση δεν παρουσιάζει καλύτερα αποτελέσματα από την *Reduce false negatives* (Dice Coefficient: 0.70) αφού ενώ πράγματι δεν παρουσιάζει τα false positives έχει χάσει κάποιες καρκινικές περιοχές.

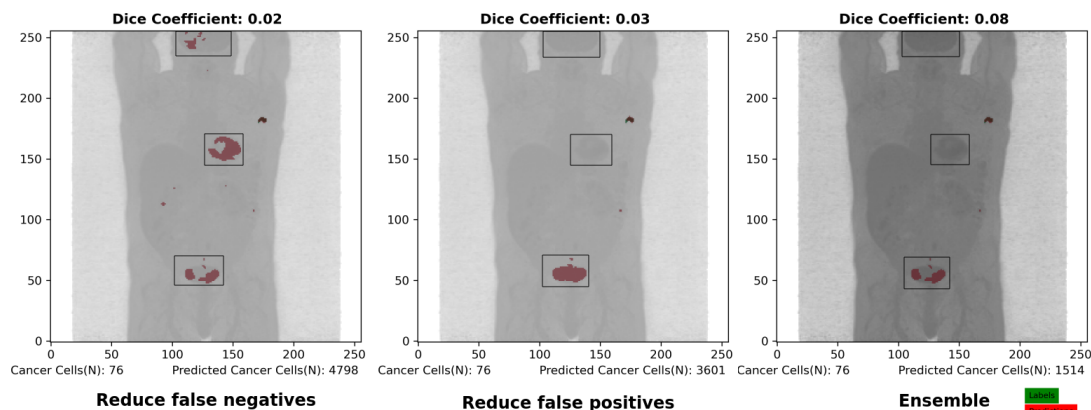


FIGURE 4.19: Κάποιες εικόνες της υλοποίησης του Ensemble U-net. Με κόκκινο παρουσιάζονται οι μάσκες της πρόβλεψης του δικτύου. Με πράσινο οι πραγματικές τιμές. Το σκούρο κόκκινο χρώμα στις μάσκες καταδεικνύει τις προβλέψεις που ήταν σωστές. Αφαίρεση των false positives του εγκεφάλου και της καρδιάς αλλά όχι και της ουροδόχου κύστης

Παρόμοια αποτελέσματα παρατηρούμε και στις παραπάνω εικόνες (4.19). Σε αυτή την περίπτωση καταφέραμε να αφαιρέσουμε τα false positives τις περιοχές τους εγκεφάλου και της καρδιάς και πράγματι είχαμε βελτίωση στα αποτελέσματα (Dice Coefficient: 0.08 - περίπου τέσσερις φορές μεγαλύτερο από τις άλλες υλοποιήσεις). Τα αποτελέσματα παραμένουν μικρά όμως επειδή δεν καταφέραμε να αφαιρέσουμε τα false positives της περιοχής της ουροδόχου κύστης, τα οποία είναι αναλογικά περισσότερα από τα ελάχιστα καρκινικά κύτταρα (76).

Οι παρατηρήσεις που έγιναν στις παραπάνω εικόνες έχουν ισχύ για όλο το σύνολο των εικόνων. Τα false positives δεν μπορούσαν να αφαιρεθούν εξ ολοκλήρου χωρίς να χαθούν και τα true positives. Αξίζει λοιπόν να εμβαθύνουμε σε αυτά και να καταλάβουμε τον λόγο της παρουσίας τους. Μια πολύ σημαντική παρατήρηση που θα μας βοηθήσει στην κατανόηση του προβλήματος είναι ότι τα false positives που εμφανίζονται δεν είναι τυχαία. Στις εικόνες που παρουσιάσαμε παραπάνω φαίνεται ότι τα false positives αφορούν συγκεκριμένες περιοχές του ανθρώπινου οργανισμού. Συγκεκριμένα αφορούν όργανα όπως τον εγκέφαλο, τους νεφρούς, την ουροδόχο κύστη και την καρδιά. Στην Ensemble υλοποίηση καταφέραμε να αφαιρέσουμε σε μεγάλο βαθμό τα false positives στον εγκέφαλο και τους νεφρούς αλλά όχι στην ουροδόχο κύστη και στην καρδιά. Παραθέτουμε ενδεικτικά κάποιες εικόνες (4.20) της Ensemble υλοποίησης.

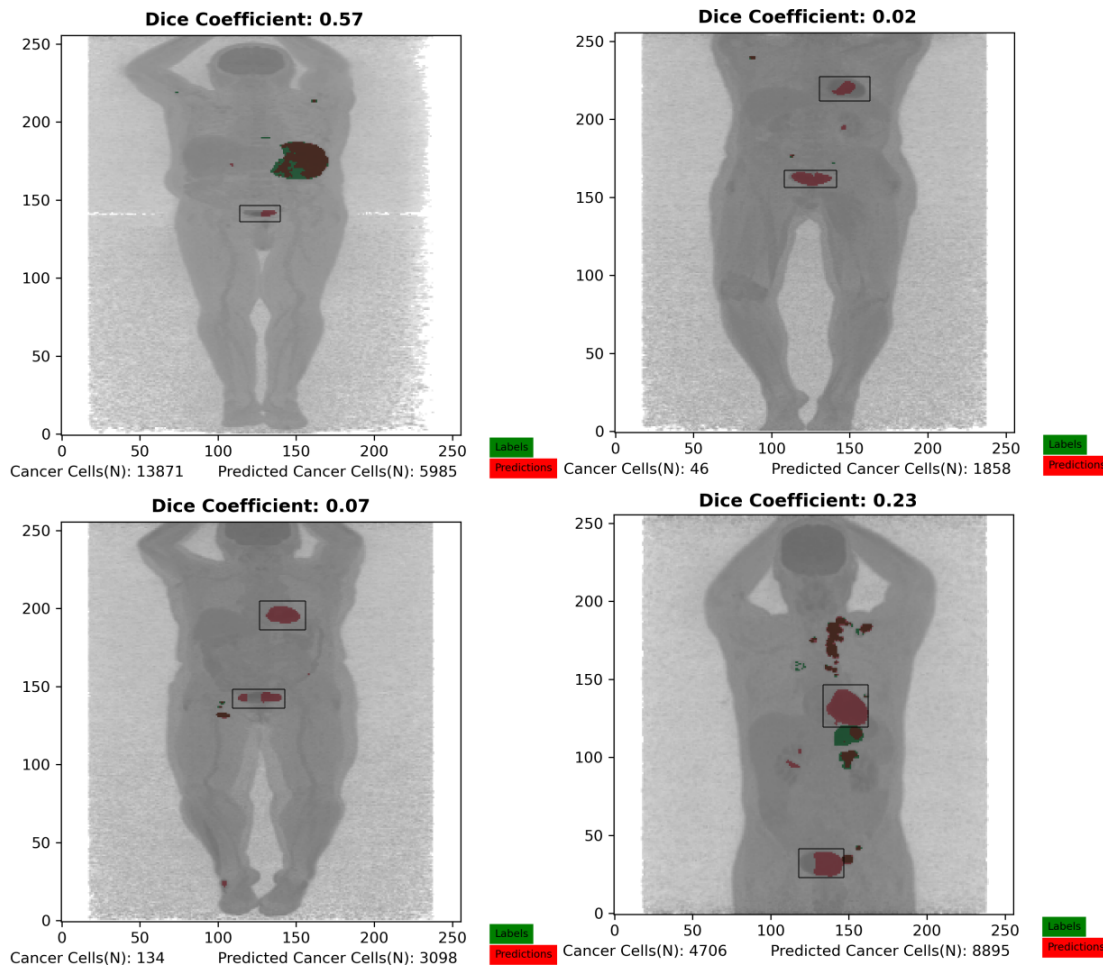


FIGURE 4.20: Κάποιες εικόνες της υλοποίησης του Ensemble U-net. Με κόκκινο παρουσιάζονται οι μάσκες της πρόβλεψης του δικτύου. Με πράσινο οι πραγματικές τιμές. Το σκούρο κόκκινο χρώμα στις μάσκες καταδεικνύει τις προβλέψεις που ήταν σωστές. False positives σε όργανα που παρουσιάζουν μεγαλύτερη μεταβολική δραστηριότητα

Ο λόγος που αυτές οι περιοχές παρουσιάζουν τόσο συχνά false positives σχετίζεται με τις βιολογικές ιδιότητες των συγκεκριμένων περιοχών και των PET εικόνων. Υπενθυμίζουμε ότι οι PET εικόνες ως στόχο έχουν να παρουσιάσουν τις περιοχές ενδιαφέροντος με μεγαλύτερη ένταση. Αυτό συμβαίνει επειδή οι ενδεχόμενες περιοχές ενδιαφέροντος χαρακτηρίζονται από διαφορετικές μεταβολικές λειτουργίες και τελικά απορροφούν σε μεγαλύτερα ποσοστά τον ραδιοσημασμένο ιχνηθέτη που χορηγείται στον ασθενή. Δυστυχώς όμως, διαφορετικές μεταβολικές λειτουργίες δεν παρουσιάζουν μόνο οι περιοχές ενδιαφέροντος (καρκινικά κύτταρα) αλλά και άλλες περιοχές, όπως η καρδιά, η ουροδόχος κύστη, οι νεφροί και ο εγκέφαλος. Τελικά αυτές οι περιοχές, όπως και των καρκινικών κυττάρων, παρουσιάζονται με μεγαλύτερη ένταση και έτσι τελικά το δίκτυο τις χαρακτηρίζει λανθασμένα ως καρκινικές. Στην επόμενη ενότητα θα παρουσιάσουμε κάποιες προτάσεις που ενδεχομένως θα προσπεράσουν αυτό το εμπόδιο.

4.6 Προτάσεις για βελτίωση

Παρά τις δυσκολίες και τους περιορισμούς που παρουσιάζονται υπάρχουν περιθώρια για βελτίωση των αποτελεσμάτων. Μια πλειάδα στρατηγικών μπορεί να επιστρατευτεί σε πολλά επίπεδα της υλοποίησης, από την προεπεξεργασία των δεδομένων μέχρι την αρχιτεκτονική του δικτύου αλλά και γενικότερα της υλοποίησης. Για να βρεθούν αυτές οι στρατηγικές πρέπει να γνωρίζουμε ποιά είναι τα εμπόδια που παρουσιάζονται και στις προηγούμενες ενότητες παρουσιάσαμε κάποια από αυτά. Στη συνέχεια θα παρουσιάσουμε κάποιες προτάσεις οι οποίες θα προσφέρουν εναλλακτικές που ξεπερνούν τα εμπόδια που συναντήσαμε και ενδεχομένως θα μπορούσαν να επιφέρουν βελτιώσεις στα αποτελέσματα. Οι προτάσεις αυτές είναι ανεξάρτητες μεταξύ τους, πράγμα που σημαίνει ότι θα μπορούσαν να εφαρμοστούν μεμονωμένα ή όλες μαζί.

Εναλλακτική αρχιτεκτονική δικτύου.

Προφανώς η αρχιτεκτονική του δικτύου αποτελεί έναν από τους πιο σημαντικούς παράγοντες και μια εναλλακτική σε αυτό το επίπεδο πάντα αποτελεί μια δοκιμή που αξίζει να γίνει. Από την στιγμή που δημοσιοποιήθηκε το U-net, έχουν γίνει πολλές απόπειρες βελτίωσης και εξέλιξής του, πολύ περισσότερες από αυτές που παρουσιάσαμε. Προφανώς, η κάθε αρχιτεκτονική παρουσιάζει διαφορετικά επίπεδα πολυπλοκότητας και απαιτεί διαφορετικούς υπολογιστικούς ή άλλους πόρους για να είναι δυνατό να υλοποιηθεί ή να αξιοποιηθεί στο έπακρο. Συνεπώς δεν είναι πάντα δυνατό να εφαρμοστεί οποιαδήποτε. Παρόλα αυτά λόγω της κεντρικής συνήθως θέσης που έχουν οι αρχιτεκτονικές των δικτύων στις υλοποιήσεις θα αναφέρουμε κάποιες.

Θα προτείνουμε δύο αρχιτεκτονικές:

U-net++. Αποτελεί μια αρχιτεκτονική (Zhou et al., 2018) η οποία στηρίζεται στο U-net και αλλάζει συγκεκριμένα το μέρος των skip connections (βλέπε εικόνα 4.21). Σε αυτήν την αρχιτεκτονική τα skip connections δεν μεταφέρουν μόνο την πληροφορία από το αντίστοιχο επίπεδο του encoder στον decoder αλλά έναν συνδυασμό από πολλά επίπεδα. Τα νέα skip connections ονομάζονται nested connections λόγω της συνδεσιμότητάς τους και έχουν σαν στόχο να γεφυρώσουν το σημασιολογικό κενό που δημιουργείται από τον συνδυασμό του σημασιολογικού και χωρικού περιεχομένου. Τελικά το αποτέλεσμα θα είναι η σημασιολογική κατάτμηση να γίνει πιο ακριβής, ένα χαρακτηριστικό που ίσως φανεί χρήσιμο καθώς μπορεί να επιτρέψει στο δίκτυο να έχει καλύτερη εποπτεία σε χαρακτηριστικά υψής των κυττάρων τα οποία διαφοροποιούν τα υγιή από τα καρκινικά αλλά είναι πιο λεπτεπίλεπτα και χρειάζονται μεγαλύτερη ακρίβεια για να εντοπιστούν. Περισσότερα για αυτό θα ειπωθούν στην επόμενη ενότητα.

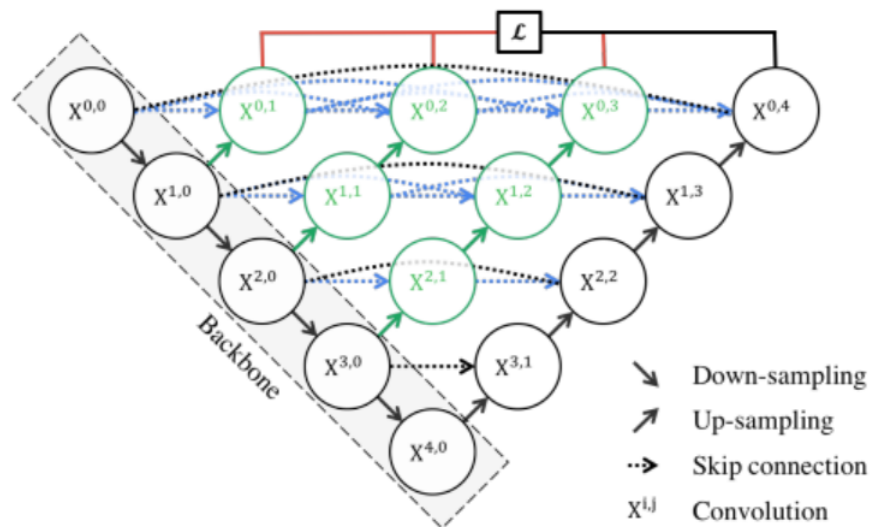


FIGURE 4.21: Τα nested connections του U-net++

nnU-net. Το nnU-net (No New-Net) (Fabian et al., 2021) αποτελεί κάτι παραπάνω από μια αρχιτεκτονική. Αποτελεί ένα framework (βλέπε εικόνα 4.22) το οποίο καλύπτει μια ευρεία γκάμα σταδίων από το pre-processing μέχρι το post-processing αξιοποιώντας υπάρχουσες αρχιτεκτονικές (εξού και το όνομα) για να επιτύχει state-of-the-art αποτελέσματα σε εφαρμογές κατάτμησης ιατρικών εικόνων. Εφαρμόζονται διάφορες υλοποιήσεις οι οποίες παρέχουν λύσεις σε προβλήματα που εμφανίζονται ειδικά σε εφαρμογές σημασιολογικής κατάτμησης ιατρικών εικόνων όπως εκτεταμένο data augmentation (κρίσιμο λόγω του μικρού αριθμού δεδομένων για εκπαίδευση), σειριακή χρήση δικτύων (cascade architecture) και εκπαίδευση δύο σταδίων. Ένα θετικό χαρακτηριστικό της συγκεκριμένης υπερ-αρχιτεκτονικής αποτελούν τα πολλά μέρη της καθιστώντας δυνατή την μερική υλοποίησή της ανάλογα τις ανάγκες και τους περιορισμούς μας.

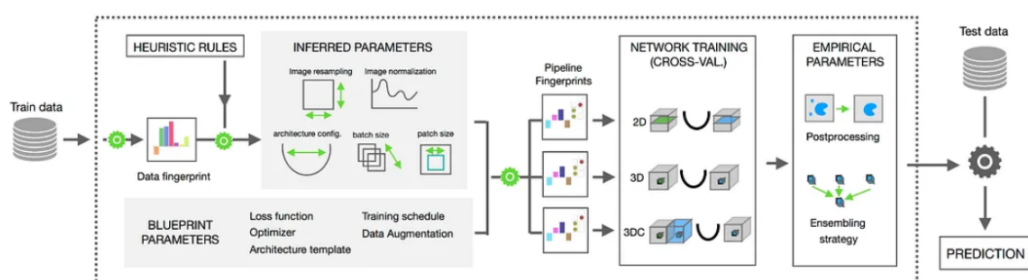


FIGURE 4.22: Το pipeline του nnU-net

Αύξηση ανάλυσης εικόνας ή patch size.

Το μεγαλύτερο εμπόδιο της υλοποίησής μας είναι όπως αναφέραμε η λανθασμένη κατηγοριοποίηση συγκεκριμένων οργάνων ως καρκινικές περιοχές. Ο λόγος είναι ότι οι περιοχές αυτές λόγω των βιολογικών τους ιδιοτήτων παρουσιάζουν μεγαλύτερη ένταση στις PET εικόνες όπως και οι καρκινικές με αποτέλεσμα το δίκτυο στηριζόμενο κυρίως σε αυτό τους το χαρακτηριστικό να μπερδεύεται. Είναι απαραίτητο λοιπόν να βρεθούν και

άλλα μοτίβα τα οποία θα χαρακτηρίζουν μονοσήμαντα τα καρκινικά κύτταρα έτσι ώστε το δίκτυο να μπορεί να τα αναγνωρίσει. Δύο μοτίβα που ενδεχομένως θα μπορούσαν να γίνουν κριτήρια διαχωρισμού είναι η υφή των ιστών και η χωρική τους τοποθεσία.

Τα καρκινικά κύτταρα και οι ιστοί παρουσιάζουν διαφορετικά χαρακτηριστικά υφής. Η υφή αποτελεί ένα σύνολο χωρικών, γεωμετρικών, χρωματικών και άλλων χαρακτηριστικών και συχνά αποτελεί ένα κριτήριο αναγνώρισης των καρκινικών κυττάρων. Όλα αυτά μπορούν να αναχθούν και σε πιο μακροσκοπικά μεγέθη όπως τους ιστούς, την εποπτεία των οποίων δίνουν οι εικόνες που έχουμε στη διάθεσή μας. Τα χαρακτηριστικά της υφής, όμως, είναι λεπτεπίλεπτα και για να μην χαθούν πρέπει οι εικόνες να έχουν καλή ανάλυση. Οι διαθέσιμες από το dataset εικόνες έχουν διαστάσεις (326,400,400) αλλά εμείς στην προεπεξεργασία τις μειώσαμε σε (256,256,256). Την μείωση αυτή την κάναμε επειδή η GPU που χρησιμοποιήσαμε δεν είχε αρκετή μνήμη για να υποστηρίξει τους διάφορους υπολογισμούς που ήταν απαραίτητοι να γίνουν στις εικόνες. Μια άλλη GPU ή μια άλλη στρατηγική στους υπολογισμούς θα μπορούσε να επιτρέψει τη χρήση εικόνων με μεγαλύτερη ανάλυση, η οποία με τη σειρά της θα είχε σαν ενδεχόμενο αποτέλεσμα καλύτερα αποτελέσματα κατάτμησης.

Τα χαρακτηριστικά της χωρικής τοποθεσίας των εικόνων θα μπορούσαν ενδεχομένως επίσης να βοηθήσουν. Υπενθυμίζουμε ότι το δίκτυο δεν δέχεται ολόκληρες τις εικόνες στην είσοδό του αλλά κομμάτια (patches) μεγέθους (64, 64, 64). Αυτό έχει σαν αποτέλεσμα διάφορες περιοχές της εικόνας είτε να κόβονται (παράδειγμα η περιοχή της καρδιάς να μην εμπεριέχεται σε ένα μόνο patch αλλά σε περισσότερα) είτε να παρουσιάζονται μεμονωμένα και να χάνεται ένα ευρύτερο πλαίσιο τοποθεσίας, μια πληροφορία που ενδεχομένως θα ήταν χρήσιμη. Για να δώσουμε ένα παράδειγμα, το δίκτυο ίσως τμηματοποιούσε καλύτερα την περιοχή της καρδιάς αν στο στάδιο της εκπαίδευσης είχε δει πολλά παραδείγματα τέτοιων περιοχών οι οποίες ενώ παρουσιάζουν μεγαλύτερη ένταση δεν αποτελούν καρκινικές. Για να είναι όμως αυτό εφικτό, πρέπει το δίκτυο να καταλαβαίνει ότι μια περιοχή είναι περιοχή καρδιάς και όσο μικρότερο είναι το μέγεθος του patch τόσο πιο δύσκολο γίνεται. Ο λόγος που εμείς επιλέξαμε το μέγεθος (64, 64, 64) ήταν πάλι σχετικός με τους περιορισμούς της μνήμης της GPU.

Σημασιολογική κατάτμηση με συνυπολογισμό των βιολογικών ιδιοτήτων επίμαχων περιοχών εισάγοντας ειδικές κλάσεις.

Στις προηγούμενες ενότητες παρουσιάσαμε κάποιες προτάσεις οι οποίες ενδεχομένως θα βελτίωναν τις επιδόσεις του δικτύου στις επίμαχες περιοχές (όργανα διαφορετικών μεταβολικών ιδιοτήτων), οι οποίες αποτελούν το κύριο εμπόδιο της κατάτμησης. Το κύριο αρνητικό αυτών των προτάσεων είναι ότι για να υλοποιηθούν απαιτείται καλύτερη υπολογιστική υποδομή η οποία είναι πιο κοστοβόρα και πολλές φορές αδύνατη. Στον πυρήνα των προηγούμενων προσεγγίσεων κρύβεται η πρακτική της αύξησης του bandwidth των υπολογισμών έτσι ώστε να διατηρηθεί χρήσιμη πληροφορία. Σε αυτήν την ενότητα θα παρουσιάσουμε μια πρόταση η οποία έχει ως σκοπό να διατηρήσει και να αξιοποιήσει την απαραίτητη πληροφορία (επίμαχες περιοχές) με έναν σειριακό τρόπο χωρίς να απαιτείται μεγάλο bandwidth υπολογισμών.

Η κεντρική ιδέα της πρότασης αυτής έγκειται στην εκπαίδευση του δικτύου σε ένα υπερσύνολο σηματολογικών κλάσεων όπου η κατάτμηση σε αυτές θα είναι πιο εύκολη από το αρχικό δυαδικό. Το υπερσύνολο αυτών κλάσεων θα προκύψει από κατάλληλη διαίρεση του αρχικού συνόλου, τέτοια ώστε η σηματολογική κατάτμηση να γίνει πιο εύκολη. Στην περίπτωση μας το δίκτυο δυσκολευόταν στην κατάτμηση των επίμαχων περιοχών οι οποίες αναφέραμε ότι είναι όργανα όπως η καρδιά και η ουροδόχος κύστη. Έτσι η διαίρεση θα γίνει με γνώμονα αυτές τις περιοχές. Θα δημιουργήσουμε έναν νέο άξονα διαχωρισμού των κλάσεων, αυτόν των επίμαχων περιοχών ο οποίος θα διαιρέσει τις περιοχές σε επίμαχες και μη επίμαχες. Το τελικό υπερσύνολο των κλάσεων θα

προκύψει από τον συνδυασμό λοιπόν των δύο αξόνων, αυτών των καρκινικών κυττάρων και των επίμαχων περιοχών παράγοντας συνολικά 4 κλάσεις (βλέπε εικόνα 4.23). Οι κλάσεις αυτές θα είναι τα μη καρκινικά κύτταρα και μη επίμαχα, τα καρκινικά και μη επίμαχα, τα μη καρκινικά και επίμαχα και τα καρκινικά και επίμαχα. Αφού επιτευχθεί η κατάτμηση σε αυτό το υπερσύνολο μπορούμε μετά εύκολα να δημιουργήσουμε ξανά το αρχικό αγνοώντας τον άξονα των επίμαχων περιοχών. Η κλάση των καρκινικών κυττάρων θα αποτελείται από τις κλάσεις καρκινικά και μη επίμαχα και καρκινικά και επίμαχα και η κλάση των μη καρκινικών θα αποτελείται από τις κλάσεις μη καρκινικά και μη επίμαχα και μη καρκινικά και επίμαχα.

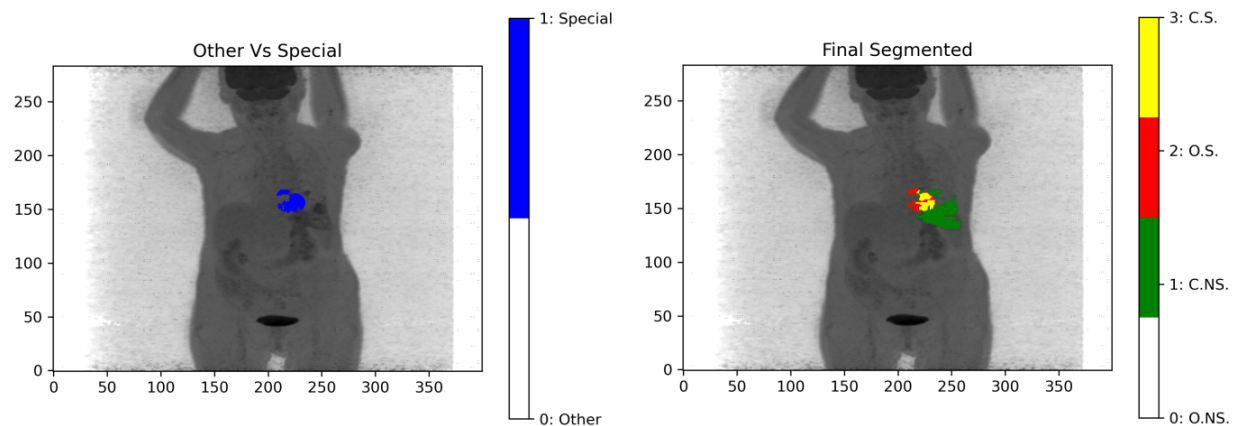


FIGURE 4.23: Το επεκτεταμένο σύνολο σημασιολογικών κλάσεων. O.N.S: Όχι καρκινικό κύτταρο/όχι επίμαχη περιοχή, C.N.S: καρκινικό κύτταρο/όχι επίμαχη περιοχή, O.S: όχι καρκινικό κύτταρο/επίμαχη περιοχή, C.S: καρκινικό κύτταρο/επίμαχη περιοχή

Στόχος της συγκεκριμένης πρότασης είναι το δίκτυο να αναπτύσσει διαφορετικούς συσχετισμούς για την κάθε περιοχή που θα προσπαθήσει να κατηγοριοποιήσει. Για παράδειγμα αν το δίκτυο εξετάζει μια περιοχή η οποία χαρακτηρίζεται ως επίμαχη και συνεπώς έχει μεγαλύτερη ένταση στις PET εικόνες να μην χαρακτηριστεί αμέσως ως καρκινική. Εάν υποθετικά μια επίμαχη καρκινική περιοχή παρουσιάζει ακόμα μεγαλύτερα ποσά έντασης από μια επίμαχη μη καρκινική περιοχή, θα χαρακτηριστεί ως καρκινική εφόσον ξεπερνάει ένα μεγαλύτερο κατώφλι από μια μη επίμαχη περιοχή. Η ένταση της περιοχής θα μπορούσε ενδεχομένως να είναι ένα κριτήριο αλλά θα μπορούσαν να υπάρχουν και άλλα. Σημασία έχει το δίκτυο να μπορεί να αντιμετωπίζει αλλιώς περιοχές οι οποίες εν γένει μπορούν να ομαδοποιηθούν μεταξύ τους παρουσιάζοντας παράλληλα παρόμοια χαρακτηριστικά μεταξύ τους και διαφορετικά από τις υπόλοιπες.

Για την διαίρεση των κλάσεων σε επίμαχες και μη επίμαχες μπορεί να γίνει χρήση επίσης ενός δικτύου σημασιολογικής κατάτμησης οργάνων. Γνωρίζουμε μάλιστα ότι αυτό το πρόβλημα είναι πιο εύκολο, οπότε έτσι αποφεύγονται δυσκολίες πολλαπλής βελτιστοποίησης του κάθε μέρους της συνολικής υλοποίησης ώστε τα αποτελέσματα να είναι καλά.

Τελικά η συνολική αρχιτεκτονική της πρότασης αποτελείται από δύο μέρη. Στην αρχή γίνεται μια πρώτη κατάτμηση για την εύρεση των επίμαχων περιοχών (όργανα όπως ουροδόχος κύστη, καρδιά, εγκέφαλος), μετά δημιουργείται το υπερσύνολο των κλάσεων συνδυάζοντας τις επίμαχες περιοχές και τα labels των καρκινικών κυττάρων για να υλοποιηθεί η δεύτερη κατάτμηση και τέλος συνδυάζονται οι υπερ-κλάσεις για να δημιουργηθούν οι απλές (καρκινικές περιοχές ή μη).

Bibliography

- Aide, Nicolas et al. (Jan. 2022). “PET/CT variants and pitfalls in malignant melanoma”. In: *Cancer imaging journal (part of springer nature)*, pp. 1–20. URL: <https://doi.org/10.1186/s40644-021-00440-4>.
- Badrinarayanan, Vijay, Alex Kendall, and Roberto Cipolla (Nov. 2015). “SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 2481–2495. URL: <https://api.semanticscholar.org/CorpusID:60814714>.
- Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio (Sept. 2014). “Neural Machine Translation by Jointly Learning to Align and Translate”. In: *ArXiv*. URL: <https://api.semanticscholar.org/CorpusID:11212020>.
- Bisschop, C. et al. (June 2020). “Rational use of 18F-FDG PET/CT in patients with advanced cutaneous melanoma: A systematic review”. In: *ScienceDirect* 29, pp. 1236–1239. URL: <https://www.sciencedirect.com/science/article/pii/S1040842820301815>.
- Chen, Liang-Chieh et al. (Dec. 2014). “Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs”. In: *arXiv*. URL: <https://api.semanticscholar.org/CorpusID:1996665>.
- Fabian, Isensee et al. (Feb. 2021). “nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation”. In: *Nature Methods*, 203–211. URL: <https://doi.org/10.1038/s41592-020-01008-z>.
- Fukushima, Kunihiko (Oct. 1969). “Visual Feature Extraction by a Multilayered Network of Analog Threshold Elements”. In: *IEEE Transactions on Systems Science and Cybernetics*, pp. 322–333. URL: <https://doi.org/10.1109/TSSC.1969.300225>.
- (Apr. 1980). “Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position”. In: *Biol. Cybernetics*. URL: <https://doi.org/10.1007/BF00344251>.
- Gatidis, Sergios et al. (Oct. 2022). “A whole-body FDG-PET/CT dataset with manually annotated tumor lesions”. In: *Scientific Data*. URL: <https://doi.org/10.1038/s41597-022-01718-3>.
- He, Kaiming et al. (Dec. 2015). “Deep Residual Learning for Image Recognition”. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778. URL: <https://api.semanticscholar.org/CorpusID:206594692>.
- Jumper, John et al. (July 2021). “Highly accurate protein structure prediction with AlphaFold”. In: *Nature* 15, pp. 4477–4479. URL: <https://www.nature.com/articles/s41586-021-03819-2>.
- Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E Hinton (Jan. 2012). “ImageNet Classification with Deep Convolutional Neural Networks”. In: *Neural Information Processing Systems*. URL: <https://doi.org/10.1145/3065386>.
- Lecun, Y. et al. (Nov. 1998). “Gradient-based learning applied to document recognition”. In: *Proceedings of the IEEE*, pp. 2278–2324. URL: <https://doi.org/10.1109/5.726791>.

- Long, J., E. Shelhamer, and T. Darrell (June 2015). “Fully convolutional networks for semantic segmentation”. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 193–202, pp. 3431–3440. URL: <https://doi.org/10.1109/CVPR.2015.7298965>.
- McCulloch, Warren S. and Walter Pitts (Dec. 1943). “A logical calculus of the ideas immanent in nervous activity”. In: *The bulletin of mathematical biophysics*, 115–133. URL: <https://doi.org/10.1007/BF02478259>.
- Oktay, Ozan et al. (May 2018). “Attention U-Net: Learning Where to Look for the Pancreas”. In: *ArXiv*. URL: <https://api.semanticscholar.org/CorpusID:4861068>.
- Otsu, Nobuyuki (Jan. 1979). “A threshold selection method from gray-level histograms”. In: *IEEE Transactions on Systems, Man, and Cybernetics*, pp. 62–66. URL: <https://doi.org/10.1109/TSMC.1979.4310076>.
- Ronneberger, Olaf, Philipp Fischer, and Thomas Brox (Oct. 2015). “U-Net: Convolutional Networks for Biomedical Image Segmentation”. In: *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference*, pp. 234–241. URL: <https://api.semanticscholar.org/CorpusID:3719281>.
- Rosenblatt, Frank (Jan. 1957). “The perceptron - A perceiving and recognizing automaton”. In: *Cornell Aeronautical Laboratory*. URL: https://www.researchgate.net/publication/262288029_Perceptron.
- Vaswani, Ashish et al. (June 2017). “Attention is all you Need”. In: *Advances in neural information processing systems*. URL: <https://api.semanticscholar.org/CorpusID:13756489>.
- Zhou, Zongwei et al. (June 2018). “UNet++: A Nested U-Net Architecture for Medical Image Segmentation”. In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support : 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, held in conjunction with MICCAI 2018*, pp. 3–11. URL: <https://api.semanticscholar.org/CorpusID:50786304>.