



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ & ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ & ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

Αυτοματοποιημένη Κατάτμηση Οργάνων και Όγκων σε Αξονικές Τομογραφίες: Διερεύνηση της Σύνθετης Κλιμάκωσης στην Αρχιτεκτονική U-Net

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

Κυριακής Δ. Κολπετίνου

Επιβλέπων: Γεώργιος Ματσόπουλος

Καθηγητής ΕΜΠ

Αθήνα, Σεπτέμβριος 2024

Η σελίδα αυτή είναι σκόπιμα λευκή.



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ & ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ & ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

Αυτοματοποιημένη Κατάτμηση Οργάνων και Όγκων σε Αξονικές Τομογραφίες: Διερεύνηση της Σύνθετης Κλιμάκωσης στην Αρχιτεκτονική U-Net

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

Κυριακής Δ. Κολπετίνου

Επιβλέπων: Γεώργιος Ματσόπουλος

Καθηγητής ΕΜΠ

Εγκρίθηκε από την κάτωθι τριμελή επιτροπή την 30^η Σεπτεμβρίου 2024:

Γεώργιος Ματσόπουλος

Καθηγητής ΕΜΠ

Αθανάσιος Παναγόπουλος

Καθηγητής ΕΜΠ

Παναγιώτης Τσανάκας

Καθηγητής ΕΜΠ

Αθήνα, Σεπτέμβριος 2024

Κυριακή Δ. Κολπετίνου

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών ΕΜΠ

Copyright © Κυριακή Δ. Κολπετίνου, 2024

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.

Περίληψη

Η εφαρμογή προηγμένων τεχνικών κατάτμησης ιατρικών εικόνων αποτελεί σημαντική καινοτομία, προσφέροντας πολύτιμες πληροφορίες για την ανάλυση και τον εντοπισμό καρκινικών όγκων σε αξονικές τομογραφίες. Η κατάτμηση αυτών των εικόνων χρησιμοποιείται τόσο για τη διάγνωση όσο και για τη θεραπευτική παρακολούθηση των ασθενών. Η αυτοματοποίηση αυτής της διαδικασίας επιτυγχάνεται με τη σημασιολογική κατάτμηση. Η Μηχανική Μάθηση έχει φέρει επανάσταση στην ιατρική, επιτρέποντας την ανάπτυξη αλγορίθμων που μπορούν να εκπαιδευτούν σε σύνολα δεδομένων για την ανάλυση εικόνων και αναγνώριση περιοχών ενδιαφέροντος με μεγάλη ακρίβεια. Με τη χρήση αυτών των αλγορίθμων, είναι δυνατή η βελτίωση της ακρίβειας στη διάγνωση, η παρακολούθηση της εξέλιξης ασθενειών, και η εξατομικευμένη θεραπεία. Τα Συνελκτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks - CNNs) αποτελούν ένα από τα πιο ισχυρά εργαλεία της Μηχανικής Μάθησης και έχουν αποδειχθεί ιδιαίτερα αποτελεσματικά στην επεξεργασία ιατρικών εικόνων. Τα Συνελκτικά Νευρωνικά Δίκτυα αποδεικνύονται ιδιαίτερα αποτελεσματικά στην αναγνώριση χαρακτηριστικών και στην κατάτμηση εικόνων. Ειδικά στην κατάτμηση ιατρικών εικόνων, επιτρέπουν την ακριβή κατηγοριοποίηση των εικονοστοιχείων των εικόνων και την απομόνωση παθολογικών περιοχών, όπως καρκινικοί όγκοι. Στην παρούσα εργασία, χρησιμοποιήθηκε η αρχιτεκτονική U-Net, ένα συνελκτικό δίκτυο που έχει δημιουργηθεί για τη σημασιολογική κατάτμηση ιατρικών εικόνων με μεγάλη ακρίβεια. Το δίκτυο αυτό εκπαιδεύτηκε σε αξονικές τομογραφίες οργάνων και καρκινικών όγκων, σε συνδυασμό με την εφαρμογή της μεθόδου σύνθετης κλιμάκωσης του δικτύου EfficientNet με στόχο την διερεύνηση της επίδρασης της μεθόδου στην ακρίβεια του μοντέλου.

Λέξεις-Κλειδιά: Σημασιολογική κατάτμηση, Αξονικές τομογραφίες, U-Net, EfficientNet, Συνελκτικά Νευρωνικά Δίκτυα, Μηχανική Μάθηση.

Η σελίδα αυτή είναι σκόπιμα λευκή.

Abstract

The application of advanced medical image segmentation techniques is a significant innovation, offering valuable information for the analysis and detection of cancerous tumors in CT scans. The segmentation of these images is used both for diagnosis and for therapeutic monitoring of patients. The automation of this process is achieved through semantic segmentation. Machine Learning has revolutionized medicine, enabling the development of algorithms that can be trained on datasets to analyze images and recognize regions of interest with high accuracy. With the use of these algorithms, it is possible to improve diagnostic accuracy, monitor the progression of diseases, and provide personalized treatment. Convolutional Neural Networks (CNNs) are one of the most powerful tools in Machine Learning and have proven particularly effective in the processing of medical images. CNNs are particularly effective in feature recognition and image segmentation. Specifically, in medical image segmentation, they allow for the precise categorization of image pixels and the isolation of pathological regions, such as cancerous tumors. In this study, the U-Net architecture was used, a convolutional neural network designed for the semantic segmentation of medical images with high accuracy. This network was trained on CT scans of organs and cancerous tumors, combined with the application of the compound scaling method from EfficientNet, aiming to investigate the effect of this method on the model's accuracy.

Keywords: CT scans, semantic segmentation, U-Net, EfficientNet, Machine Learning, Convolutional Neural Networks.

Η σελίδα αυτή είναι σκόπιμα λευκή.

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή κ. Γεώργιο Ματσόπουλο, για την ευκαιρία που μου έδωσε να εκπονήσω την παρούσα εργασία στο εργαστήριο του. Θα ήθελα να εκφράσω τις θερμές ευχαριστίες μου και στην υπόλοιπη τριμελή επιτροπή, καθ. κ. Αθανάσιο Παναγόπουλο και καθ. κ. Παναγιώτη Τσανάκα. Επίσης, θα ήθελα να ευχαριστήσω τον Δρ. Ιωάννη Βεζάκη για την πολύτιμη βοήθεια και καθοδήγηση που μου προσέφερε καθ' όλη τη διάρκεια εκπόνησης της διπλωματικής αυτής εργασίας. Τέλος, νιώθω την ανάγκη να εκφράσω την ευγνωμοσύνη μου προς τον φίλο και συνάδελφο μου, Νικόλαο Ρομποτή, του οποίου η στήριξη σε όλη την διάρκεια των σπουδών μου υπήρξε καθοριστική για την επιτυχή ολοκλήρωση τους.

Περιεχόμενα

Περίληψη	5
Abstract.....	7
Ευρετήριο εικόνων.....	12
Ευρετήριο εξισώσεων	15
1. Εισαγωγή.....	19
2. Η αξονική τομογραφία	20
2.1 Ιστορική Αναδρομή.....	20
2.2 Αρχές Λειτουργίας	20
2.2.1 Παραγωγή ακτίνων Χ	21
2.3 Σύνθεση της εικόνας.....	23
2.3.1 Επιλογή παραμέτρων πρωτοκόλλου σάρωσης	23
2.3.2 Συλλογή αρχικού σήματος.....	23
2.3.3 Διορθώσεις και προ-επεξεργασία	24
2.3.4 Ανακατασκευή εικόνας.....	24
2.3.5 Περαιτέρω επεξεργασία για κλινική διάγνωση	26
2.4 Ασφάλεια αξονικής τομογραφίας	26
3. Καρκίνοι του πεπτικού συστήματος.....	28
3.1 Καρκίνος του παγκρέατος.....	28
3.1.1 Συμπτώματα και διάγνωση	28
3.1.2 Θεραπεία.....	29
3.2 Καρκίνος του ήπατος	30
3.2.1 Συμπτώματα και διάγνωση	30
3.2.2 Θεραπεία.....	31
4. Ανάλυση εικόνας με Μηχανική Μάθηση	33
4.1 Εισαγωγή στην Τεχνητή Νοημοσύνη.....	33
4.2 Μηχανική Μάθηση	33
4.2.1 Επιβλεπόμενη μάθηση	34
4.2.2 Μη επιβλεπόμενη μάθηση	34
4.2.3 Ενισχυτική μάθηση.....	34
4.3 Η Μηχανική Μάθηση στην Ιατρική	34
4.4 Ανάλυση εικόνας.....	36
4.4.1 Αναπαράσταση εικόνας.....	36
4.4.2 Δειγματοληψία.....	37

4.4.3	Κβάντιση.....	37
4.4.4	Χρωματικοί χώροι.....	38
4.5	Νευρωνικά Δίκτυα	43
4.5.1	Δομή.....	43
4.5.2	Συνελικτικά νευρωνικά δίκτυα	47
4.5.3	Προώθηση δεδομένων	50
4.5.4	Οπισθοδιάδοση	53
4.5.5	Τεχνικές κανονικοποίησης.....	55
4.6	Βαθιά Μάθηση.....	58
4.7	Τμηματοποίηση εικόνας.....	59
4.8	Δίκτυο U-Net	60
4.9	Δίκτυο EfficientNet.....	64
4.10	MONAI.....	68
5.	Υλικά και Μεθοδολογία	69
5.1	Κεντρική ιδέα.....	69
5.2	Σύνολα δεδομένων	69
5.3	Προεπεξεργασία δεδομένων	73
5.4	Αρχιτεκτονική του μοντέλου και εκπαίδευση	74
6.	Αποτελέσματα	81
7.	Συζήτηση	88
8.	Συμπεράσματα.....	91
	Βιβλιογραφία	92

Ευρετήριο εικόνων

Εικόνα 2.1 Η πρώτη κλινική αξονική τομογραφία πραγματοποιήθηκε τον Οκτώβριο του 1971 στο Νοσοκομείο Atkinson Morley στο Λονδίνο [5].	20
Εικόνα 2.2 Εικόνα ενός αξονικού τομογράφου που δείχνει την περιστροφή του συστήματος ακτίνων X και του ανιχνευτή γύρω από τον ασθενή. Το «gantry» φιλοξενεί τον σωλήνα ακτίνων X και τον ανιχνευτή, ενώ ο ασθενής βρίσκεται στο τραπέζι. Η δέσμη ακτίνων X διέρχεται από το σώμα του ασθενούς και ανιχνεύεται από τον ανιχνευτή για τη δημιουργία εικόνων [6].	21
Εικόνα 2.3 Εικόνα σωλήνα ακτίνων X. Η κάθοδος εκπέμπει ηλεκτρόνια τα οποία στη συνέχεια «βομβαρδίζουν» τον μεταλλικό στόχο (άνοδος), απελευθερώνοντας ενέργεια με τη μορφή θερμότητας και ακτίνων X [7].	23
Εικόνα 3.1 Αξονική τομογραφία που απεικονίζει αδενοκαρκίνωμα το οποίο βρίσκεται στη κεφαλή του παγκρέατος. Πηγή: Pancreatic Cancer - Clinical Features - Management - TeachMeSurgery	30
Εικόνα 4.1 Αναπαράσταση μικροπικής κάμερας. Πηγή: https://byjus.com/question-answer/an-image-formed-by-a-pinhole-camera-is-always/ .”	37
Εικόνα 4.2 Αναπαράσταση δειγματοληψίας εικόνας. Πηγή: Formal Powerpoint presentation (mitu.co.in)	37
Εικόνα 4.3 Κβάντιση εικόνας με διαφορετικά επίπεδα bits/pixel. Πηγή: MUClecture_2023_121254131.pdf (uomus.edu.iq)	38
Εικόνα 4.4 Ένα πρίσμα, σύμφωνα με τον Νεύτωνα, μπορεί να αποσυνθέσει το λευκό φως σε ένα συνεχές φάσμα διαφορετικών καθαρών χρωμάτων. Από όλο το ηλεκτρομαγνητικό φάσμα, ο άνθρωπος αντιλαμβάνεται το εύρος μηκών κύματος 380 - 740 nm [19].	39
Εικόνα 4.5 Αναπαράσταση καμπυλών των συναρτήσεων χρωματικού ταιριάσματος του χώρου RGB (αριστερά) και XYZ (δεξιά) [20].	41
Εικόνα 4.6 Διάγραμμα επιπέδων νευρωνικών δικτύων. Οι κύκλοι αντιπροσωπεύουν τους νευρώνες. Πηγή: (24) MUITI-LAYER PERCEPTRON LinkedIn	44
Εικόνα 4.7 Γραφική παράσταση της βηματικής συνάρτησης. Πηγή: https://en.wikipedia.org/wiki/Heaviside_step_function	45
Εικόνα 4.8 Σιγμοειδής συνάρτηση. Πηγή: Sigmoid function - Wikipedia	46
Εικόνα 4.9 Γραφική παράσταση της συνάρτησης ενεργοποίησης RELU. Πηγή: A Gentle Introduction to the Rectified Linear Unit (ReLU) - MachineLearningMastery.com	46
Εικόνα 4.10 Γραφική παράσταση συνάρτησης ενεργοποίησης Leaky RELU. Πηγή: Leaky ReLU Activation Function [with python code] - Vidyasheela	47
Εικόνα 4.11 Αναπαράσταση δισδιάστατης συνέλιξης. Πηγή: https://www.researchgate.net/figure/An-example-of-convolution-operation-in-2D-2_fig3_324165524	49
Εικόνα 4.12 Αναπαράσταση δύο ειδών pooling: Max pooling και Average pooling. Πηγή: https://www.researchgate.net/figure/Examples-for-max-and-average-pooling-layer_fig5_356816561	50
Εικόνα 4.13 Αναπαράσταση ενός απλού πλήρως συνδεδεμένου δικτύου.	51

Εικόνα 4.14 Γραφική αναπαράσταση αποτελέσματος μάθησης ενός μοντέλου. Στην πρώτη εικόνα το μοντέλο προσεγγίζει τις πραγματικές λύσεις, ενώ στην δεύτερη το μοντέλο έχει προσαρμοστεί υπερβολικά στα δεδομένα εκπαίδευσης, χωρίς να μπορεί να γενικεύσει τη γνώση σε νέα δεδομένα. Πηγή: What is Overfitting in Machine Learning? (freecodecamp.org).....	55
Εικόνα 4.15 Αναπαράσταση νευρωνικού δικτύου χωρίς χρήση της τεχνικής dropout (πρώτη εικόνα) και με dropout (δεύτερη εικόνα). Πηγή: https://medium.com/@amarbudhiraja/https-medium-com-amarbudhiraja-learning-less-to-learn-better-dropout-in-deep-machine-learning-74334da4bfc5	57
Εικόνα 4.16 Είδη δημιουργίας παραλλαγών μιας εικόνας εισόδου για ένα δίκτυο, με αποτέλεσμα την αύξηση των δεδομένων αλλά και την ποικιλία τους, για την μείωση του φαινομένου overfitting. Πηγή: What is image augmentation - Alumentations Documentation.....	57
Εικόνα 4.17 Η τεχνητή νοημοσύνη περιλαμβάνει τα πεδία μηχανικής μάθησης και βαθιάς μάθησης.....	58
Εικόνα 4.18 Αναπαράσταση αποτελέσματος σημασιολογικής κατάτμησης μιας εικόνας με Τεχνητή Νοημοσύνη. Πηγή: An overview of semantic image segmentation. (jeremyjordan.me)	59
Εικόνα 4.19 Η αρχιτεκτονική του μοντέλου U-Net [34].	60
Εικόνα 4.20 Αναπαράσταση Attention U-Net. Πηγή: Review: Attention U-Net — Learning Where to Look for the Pancreas (Biomedical Image Segmentation) by Sik-Ho Tsang Medium.....	63
Εικόνα 4.21 a) basic unit που χρησιμοποιείται στο U-Net και b) residual unit με identity mapping. Πηγή: Residual Unit in the ResU-Net Architecture [8]: (a) Plain neural unit... Download Scientific Diagram (researchgate.net).....	64
Εικόνα 4.22 Compound scaling method [37].	65
Εικόνα 4.23 Αρχιτεκτονική EfficientNet-B0. Πηγή: https://www.researchgate.net/figure/The-EfficientNet-B0-baseline-network-structure_tbl2_371935514	66
Εικόνα 4.24 Αποτελέσματα δικτύων EfficientNet-B0 έως EfficientNet-B7 [37].	67
Εικόνα 4.25 Medical Open Network for AI [39].	68
Εικόνα 5.1 Αξονικές τομογραφίες του συνόλου δεδομένων παγκρέατος και παγκρεατικών όγκων με τις αντίστοιχες μάσκες τμηματοποίησης του οργάνου και των όγκων. Με κίτρινο χρώμα απεικονίζεται το πάγκρεας και με κόκκινο χρώμα απεικονίζεται ο όγκος.	71
Εικόνα 5.2 Αξονικές τομογραφίες του συνόλου δεδομένων ήπατος και ηπατικών όγκων με τις αντίστοιχες μάσκες τμηματοποίησης του οργάνου και των όγκων. Με κίτρινο χρώμα απεικονίζεται το ήπαρ και με κόκκινο απεικονίζονται οι όγκοι.	72
Εικόνα 5.3 Αξονικές τομογραφίες του συνόλου δεδομένων ηπατικών αγγείων και ηπατικών όγκων με τις αντίστοιχες μάσκες τμηματοποίησης του οργάνου και των όγκων. Με κίτρινο χρώμα απεικονίζονται τα ηπατικά αγγεία ενώ με κόκκινο χρώμα απεικονίζονται οι όγκοι.....	73
Εικόνα 5.4 Εφαρμογή Αφινικού Μετασχηματισμού σε εικόνα. Πηγή: Random Affine Transformations in PyTorch by Mark Ai Code Aug, 2024 Medium	75
Εικόνα 5.5 Μετρικές Precision & Recall. Πηγή: https://en.wikipedia.org/wiki/Precision_and_recall	77

Εικόνα 5.6 Αποτελέσματα Αναζήτησης Νευρωνικών δικτύων για το σύνολο δεδομένων του παγκρέατος. Ο οριζόντιος άξονας είναι οι τιμές Dice score και ο κατακόρυφος άξονας είναι ο αριθμός των παραμέτρων των μοντέλων.	78
Εικόνα 6.1 Παράδειγμα τμηματοποίησης του μοντέλου B1. Η πρώτη εικόνα αποτελεί patch διαστάσεων 96x96 από την αρχική εικόνα. Η μεσαία εικόνα είναι η μάσκα τμηματοποίησης αυτού του patch. Η δεξιά εικόνα αποτελεί την πρόβλεψη για τμηματοποίηση από το μοντέλο.	82
Εικόνα 6.2 Παράδειγμα τμηματοποίησης του μοντέλου B6. Η πρώτη εικόνα αποτελεί patch διαστάσεων 96x96 από την αρχική εικόνα. Η μεσαία εικόνα είναι η μάσκα τμηματοποίησης αυτού του patch. Η δεξιά εικόνα αποτελεί την πρόβλεψη για τμηματοποίηση από το μοντέλο.	85
Εικόνα 6.3 Παράδειγμα τμηματοποίησης του μοντέλου B1. Η πρώτη εικόνα αποτελεί patch διαστάσεων 96x96 από την αρχική εικόνα. Η μεσαία εικόνα είναι η μάσκα τμηματοποίησης αυτού του patch. Η δεξιά εικόνα αποτελεί την πρόβλεψη για τμηματοποίηση από το μοντέλο.	87

Ευρετήριο εξισώσεων

Εξίσωση 2.1	25
Εξίσωση 2.2	25
Εξίσωση 2.3	25
Εξίσωση 4.1	40
Εξίσωση 4.2	40
Εξίσωση 4.3	41
Εξίσωση 4.4	41
Εξίσωση 4.5	41
Εξίσωση 4.6	42
Εξίσωση 4.7	42
Εξίσωση 4.8	43
Εξίσωση 4.9	43
Εξίσωση 4.10	45
Εξίσωση 4.11	45
Εξίσωση 4.12	46
Εξίσωση 4.13	47
Εξίσωση 4.14	48
Εξίσωση 4.15	48
Εξίσωση 4.16	48
Εξίσωση 4.17	51
Εξίσωση 4.18	51
Εξίσωση 4.19	51
Εξίσωση 4.20	52
Εξίσωση 4.21	52
Εξίσωση 4.22	52
Εξίσωση 4.23	52
Εξίσωση 4.24	53
Εξίσωση 4.25	53
Εξίσωση 4.26	53
Εξίσωση 4.27	54
Εξίσωση 4.28	54
Εξίσωση 4.29	54
Εξίσωση 4.30	54
Εξίσωση 4.31	54
Εξίσωση 4.32	54
Εξίσωση 4.33	54
Εξίσωση 4.34	56
Εξίσωση 4.35	56
Εξίσωση 4.36	56
Εξίσωση 4.37	56
Εξίσωση 4.38	56
Εξίσωση 4.39	61

Εξίσωση 4.40	61
Εξίσωση 4.41	61
Εξίσωση 4.42	63
Εξίσωση 4.43	65
Εξίσωση 4.44	65
Εξίσωση 4.45	65
Εξίσωση 5.1	74
Εξίσωση 5.2	76
Εξίσωση 5.3	76
Εξίσωση 5.4	76

Ευρετήριο πινάκων

Πίνακας 1 Δίκτυο U-Net B0 εκπαιδευμένο στο σύνολο δεδομένων του παγκρέατος. Αποτελείται από 5 στάδια στον encoder και 5 στάδια στον decoder. Η ανάλυση των εικόνων (patches) δίνεται σε pixels. Σε κάθε στάδιο του encoder η ανάλυση υποδιπλασιάζεται. Το κάθε επίπεδο του encoder και του decoder αποτελείται από 3 συνελκτικά επίπεδα εκτός από το τελευταίο επίπεδο που αποτελείται από 2.	78
Πίνακας 2 Αρχιτεκτονική U-Net B0 για το σύνολο δεδομένων του ήπατος και των ηπατικών όγκων. Αποτελείται από 5 στάδια στον encoder και 5 στάδια στον decoder. Η ανάλυση των εικόνων (patches) δίνεται σε pixels. Σε κάθε στάδιο του encoder η ανάλυση υποδιπλασιάζεται. Το κάθε επίπεδο του encoder και του decoder αποτελείται από 3 συνελκτικά επίπεδα εκτός από το τελευταίο επίπεδο που αποτελείται από 2. .	79
Πίνακας 3 Αρχιτεκτονική δικτύου U-Net B0 εκπαιδευμένο στο σύνολο δεδομένων των ηπατικών αγγείων και ηπατικών όγκων. Αποτελείται από 5 στάδια στον encoder και 5 στάδια στον decoder. Η ανάλυση των εικόνων (patches) δίνεται σε pixels. Σε κάθε στάδιο του encoder η ανάλυση υποδιπλασιάζεται. Το κάθε επίπεδο του encoder και του decoder αποτελείται από 3 συνελκτικά επίπεδα εκτός από το τελευταίο επίπεδο που αποτελείται από 2.	79
Πίνακας 4 Αποτελέσματα mean Dice score για το σύνολο δεδομένων του παγκρέατος και των παγκρεατικών όγκων.	81
Πίνακας 5 Αποτελέσματα μετρικών IoU, Recall, Precision για το σύνολο δεδομένων του παγκρέατος και των παγκρεατικών όγκων.	82
Πίνακας 6 Αποτελέσματα mean Dice score για το σύνολο δεδομένων του ήπατος και των ηπατικών όγκων.	83
Πίνακας 7 Αποτελέσματα μετρικών IoU, Recall, Precision για το σύνολο δεδομένων του ήπατος και των ηπατικών όγκων.	84
Πίνακας 8 Αποτελέσματα mean Dice score για το σύνολο δεδομένων ηπατικών αγγείων και ηπατικών όγκων.	85
Πίνακας 9 Αποτελέσματα μετρικών IoU, Recall, Precision για το σύνολο δεδομένων των ηπατικών αγγείων και ηπατικών όγκων.	86

Η σελίδα αυτή είναι σκόπιμα λευκή.

1. Εισαγωγή

Με τη συνεχή βελτίωση και αύξηση των υπολογιστικών πόρων και την μεγάλη διαθεσιμότητα δεδομένων, οι υπολογιστές, σήμερα, χρησιμοποιούνται σε ολοένα περισσότερες και περίπλοκες εφαρμογές του πραγματικού κόσμου. Η εισαγωγή της Τεχνητής Νοημοσύνης (AI), δηλαδή της ικανότητας του υπολογιστή να μαθαίνει με βάση το μοντέλο του ανθρώπινου εγκεφάλου, και να λύνει προβλήματα, έφερε την επανάσταση στον τρόπο που προσεγγίζονται πολλοί τομείς όπως η υγεία, οι επιχειρήσεις και η βιομηχανία. Η Τεχνητή Νοημοσύνη, και το υποπεδίο της, η Μηχανική Μάθηση, παρέχουν λύσεις σε πολύπλοκα προβλήματα μέσω αλγορίθμων που «μαθαίνουν» χωρίς να είναι πάντα απαραίτητη η ανθρώπινη παρέμβαση, κάνοντας την λήψη αποφάσεων πιο γρήγορη και ακριβή [1].

Ειδικότερα, στον τομέα της υγείας, η Τεχνητή Νοημοσύνη παρέχει υπολογιστικά συστήματα και αλγορίθμους ικανούς να βοηθήσουν τους ιατρούς, παρέχοντας πληροφορίες και ανάλυση δεδομένων, στη προσπάθεια για πιο γρήγορη και ακριβή διάγνωση και αντιμετώπιση ασθενειών. Για παράδειγμα, έχουν δημιουργηθεί αλγόριθμοι που αναλύουν ιατρικές εικόνες, όπως αξονικές τομογραφίες, και εντοπίζουν με μεγάλη ακρίβεια τους καρκινικούς όγκους καθώς και το σημείο στο οποίο βρίσκονται [2]. Επιπλέον, βελτιώνει την εμπειρία των ασθενών, παρέχοντας «έξυπνους βοηθούς» οι οποίοι μπορούν να κλείσουν ραντεβού σε ιατρούς εύκολα και γρήγορα, να καταγράψουν το ιατρικό ιστορικό, ακόμα και να προσφέρουν βασικές ιατρικές πληροφορίες [3]. Αξίζει να σημειωθεί ότι η Τεχνητή Νοημοσύνη, αυτή τη χρονική στιγμή, προορίζεται να λειτουργήσει βοηθητικά προς τους ιατρούς, να διευκολύνει και να βελτιώσει την ταχύτητα και την ακρίβεια της ιατρικής διαδικασίας, και όχι να αντικαταστήσει το ιατρικό προσωπικό.

Στο πλαίσιο αυτό, διεξάγονται συνεχώς έρευνες βελτίωσης αυτής της τεχνολογίας, ώστε να ενσωματώνονται σε περισσότερους τομείς της ιατρικής και να βελτιώνουν τις δυνατότητες διάγνωσης, θεραπείας, και διαχείρισης των ασθενειών. Μία από τις σημαντικές κατευθύνσεις είναι η περαιτέρω βελτίωση των αλγορίθμων ανάλυσης ιατρικών εικόνων, με στόχο την ακριβέστερη ανίχνευση ασθενειών σε πρώιμα στάδια, καθώς και η ανάπτυξη συστημάτων που μπορούν να κάνουν προβλέψεις σχετικά με την εξέλιξη ασθενειών. Παράλληλα, η πρόοδος στον τομέα της υπολογιστικής ισχύος και των δεδομένων μεγάλου όγκου (Big Data) επιτρέπει την επεξεργασία τεράστιων ποσοτήτων ιατρικών δεδομένων, τα οποία προέρχονται από διάφορες πηγές, όπως ιστορικά ασθενών, γενετικά δεδομένα, και αποτελέσματα κλινικών εξετάσεων [4]. Η επεξεργασία αυτών των δεδομένων με τη χρήση της Τεχνητής Νοημοσύνης προσφέρει πολύτιμες πληροφορίες που μπορούν να βοηθήσουν στη λήψη αποφάσεων για θεραπείες και στην ανάπτυξη εξατομικευμένης ιατρικής.

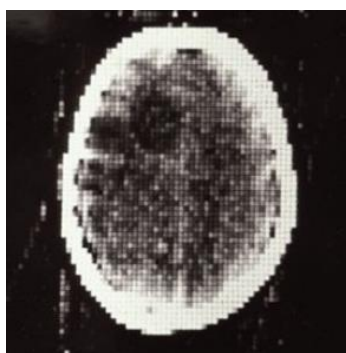
Στη παρούσα διπλωματική εργασία, θα αναφερθούμε στην Τεχνητή Νοημοσύνη και πως αυτή εφαρμόζεται, πιο συγκεκριμένα, στον τομέα της υγείας, και θα εξετάσουμε μια μεθοδολογία ανάλυσης εικόνων αξονικών τομογραφιών από καρκινικούς όγκους.

2. Η αξονική τομογραφία

Η αξονική τομογραφία ή υπολογιστική τομογραφία (Computed Tomography – CT) είναι μια ευρέως υιοθετημένη τεχνολογία απεικόνισης που χρησιμοποιείται για την επιθεώρηση της εσωτερικής μορφολογικής δομής των ανθρώπινων οργάνων. Πρόκειται για μια μη επεμβατική ιατρική εξέταση που χρησιμοποιεί εξειδικευμένο εξοπλισμό ακτίνων X και παράγει εικόνες του εσωτερικού των ασθενών ως εγκάρσιες τομές υψηλής ανάλυσης. Οι εγκάρσιες τομές αυτές (slices) συνθέτουν μια λεπτομερή, τρισδιάστατη εικόνα των οργάνων του ασθενή. Οι αξονικές τομογραφίες μπορούν να πραγματοποιηθούν σε κάθε περιοχή του σώματος για διάφορους λόγους (π.χ. διάγνωση, προγραμματισμός θεραπείας, προληπτικός έλεγχος).

2.1 Ιστορική Αναδρομή

Η αξονική τομογραφία εφευρέθηκε στις αρχές της δεκαετίας του 1970 από τον Βρετανό ηλεκτρολόγο μηχανικό Sir Godfrey Hounsfield και τον Νοτιοαφρικάνο φυσικό Allan Cormack. Τα μαθηματικά θεμέλια του αξονικού τομογράφου τέθηκαν από τον Johann Radon το 1917 με τον “Μετασχηματισμό Radon” (“Radon transform”) και την “Τεχνική Αλγεβρικής Ανασύνθεσης” (“Algebraic Reconstruction Technique”) του Stefan Kaczmarz το 1937. Οι Hounsfield και Cormack βασίστηκαν σε αυτές τις θεωρίες για να δημιουργήσουν μια επαναστατική ιατρική εξέλιξη. Πιο συγκεκριμένα, ανέπτυξαν τον πρώτο αξονικό τομογράφο το 1971. Η Εικόνα 2.1 δείχνει το αποτέλεσμα της πρώτης κλινικής σάρωσης κεφαλής ασθενούς που πραγματοποιήθηκε το ίδιο έτος [5]. Σε αναγνώριση των συνεισφορών τους, οι Hounsfield και Cormack βραβεύτηκαν με το Νόμπελ Φυσιολογίας και Ιατρικής το 1979. Η εργασία τους έθεσε τα θεμέλια για τις σύγχρονες τεχνικές ιατρικής απεικόνισης. Αρχικά, οι πρώτοι αξονικοί τομογράφοι είχαν περιορισμένη ποιότητα εικόνας και απαιτούσαν μεγάλο χρόνο σάρωσης. Από τότε, η τεχνολογία έχει προοδεύσει σημαντικά, οδηγώντας σε ταχύτερες σαρώσεις και βελτιωμένη ανάλυση εικόνας.



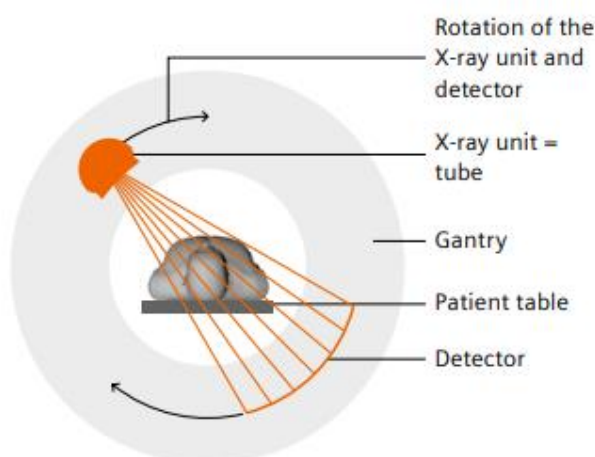
Εικόνα 2.1 Η πρώτη κλινική αξονική τομογραφία πραγματοποιήθηκε τον Οκτώβριο του 1971 στο Νοσοκομείο Atkinson Morley στο Λονδίνο [5].

2.2 Αρχές Λειτουργίας

Το άνοιγμα σε σχήμα δακτυλίου του αξονικού τομογράφου, είναι γνωστό στην ιατρική κοινότητα ως «gantry». Το gantry φιλοξενεί το σύστημα μέτρησης, που αποτελείται από τον σωλήνα ακτίνων X (X-ray tube) και τον ανιχνευτή (detector) απέναντί του (Εικόνα 2.2). Το σύστημα μέτρησης, το οποίο συνήθως ζυγίζει μεταξύ 400 και 1.600 κιλών, γυρίζει γύρω από τον ασθενή αρκετές φορές το δευτερόλεπτο. Κατά τη διάρκεια αυτής της διαδικασίας, ο

σωλήνας μεταδίδει μια δέσμη ακτίνων X σε σχήμα «βεντάλιας» (fan-shaped X ray beam). Όταν φτάνουν στον ανιχνευτή, οι ακτίνες χτυπούν έναν «σπινθηριστή» (scintillator) που μετατρέπει τις ανιχνευμένες ακτίνες X σε φως. Οι φωτοδιόδοι στη συνέχεια μετατρέπουν το φως σε ηλεκτρισμό και ένας μετατροπέας παράγει ψηφιακά δεδομένα από τα αναλογικά σήματα και τα μεταδίδει στον υπολογιστή για ανάλυση. Έπειτα, ο υπολογιστής μεταφράζει τις μετρήσεις σε μεμονωμένες εικόνες τομής ή ακόμα και σε τρισδιάστατο μοντέλο ολόκληρου του σώματος, όλα χωρίς αισθητή καθυστέρηση. Ο γιατρός μπορεί να περιστρέψει την εικόνα του σώματος στην οθόνη, να κάνει μεγέθυνση ή σμίκρυνση της εικόνας, ακόμη και να εξετάσει δομές όπως το έντερο [6].

Τα σημερινά συστήματα αξονικής τομογραφίας, των οποίων έγινε σύντομη περιγραφή της δομής και λειτουργίας τους παραπάνω, αναλύουν την ανατομία κάθε ασθενούς και υπολογίζουν τη βέλτιστη δόση ακτινοβολίας για κάθε σάρωση. Η έκθεση σε ακτινοβολία μετριέται σε millisieverts (mSv). Οι τυπικές δόσεις ακτινοβολίας κυμαίνονται από 2 έως 3 mSv [6].



Εικόνα 2.2 Εικόνα ενός αξονικού τομογράφου που δείχνει την περιστροφή του συστήματος ακτίνων X και του ανιχνευτή γύρω από τον ασθενή. Το «gantry» φιλοξενεί τον σωλήνα ακτίνων X και τον ανιχνευτή, ενώ ο ασθενής βρίσκεται στο τραπέζι. Η δέσμη ακτίνων X διέρχεται από το σώμα του ασθενούς και ανιχνεύεται από τον ανιχνευτή για τη δημιουργία εικόνων [6].

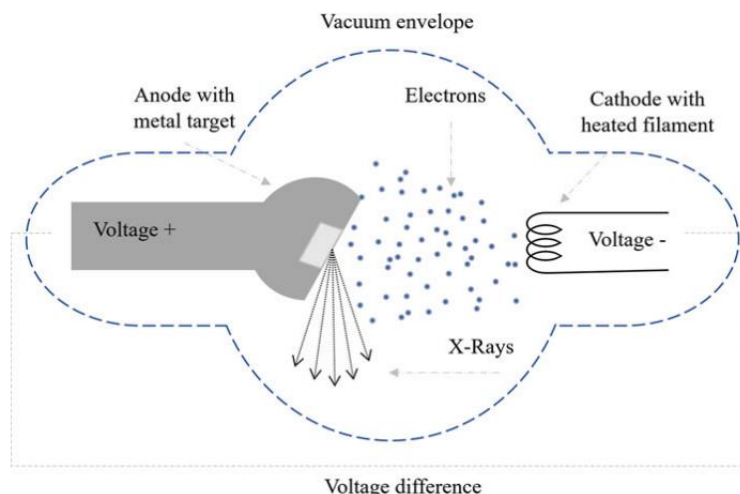
2.2.1 Παραγωγή ακτίνων X

Οι ακτίνες X στην ιατρική συνήθως παράγονται από σωλήνες ακτίνων X (X-ray tube). Μέσα σε έναν κενό σωλήνα, ηλεκτρόνια επιταχύνονται από μια κάθοδο (την πηγή ηλεκτρονίων) προς μια άνοδο (έναν μεταλλικό στόχο). Όταν τα ηλεκτρόνια χτυπούν την άνοδο, επιβραδύνονται απότομα και ένα μέρος της κινητικής τους ενέργειας μετατρέπεται σε ακτίνες X μέσω φυσικών αλληλεπιδράσεων (Εικόνα 2.3). Η εκπομπή των ηλεκτρονίων από την κάθοδο βασίζεται στην θερμοϊονική εκπομπή, δηλαδή τη θέρμανση ενός μεταλλικού νήματος. Η θερμοκρασία του νήματος ελέγχεται από ένα ηλεκτρικό ρεύμα. Αύξηση του ρεύματος οδηγεί σε αύξηση της θερμοκρασίας του νήματος, και ως αποτέλεσμα, εκπέμπονται περισσότερα ηλεκτρόνια. Συνήθως χρησιμοποιείται νήμα από βολφράμιο λόγω του υψηλού σημείου τήξης του. Τα ηλεκτρόνια που εκπέμπονται επιταχύνονται από ένα ηλεκτροστατικό πεδίο που δημιουργείται από τη διαφορά τάσης μεταξύ καθόδου και ανόδου. Ο αριθμός των επιταχυνόμενων ηλεκτρονίων είναι ανάλογος με τον αριθμό των φωτονίων ακτίνων X, και ο χρήστης μπορεί να

ελέγχει την ένταση της δέσμης των ακτίνων X αλλάζοντας το ηλεκτρικό ρεύμα του σωλήνα. Η αύξηση της τάσης μεταξύ ανόδου και καθόδου αυξάνει την κινητική ενέργεια των ηλεκτρονίων, αυξάνοντας έτσι και την μέγιστη ενέργεια των παραγόμενων ακτίνων X. Ο στόχος (άνοδος) μπορεί να είναι στατικός ή περιστρεφόμενος. Η περιστρεφόμενη άνοδος βοηθά στη διανομή της θερμότητας και επιτρέπει υψηλότερες εντάσεις ακτινοβολίας χωρίς θερμικές ζημιές. Ένα κύπελλο εστίασης (focusing cup) γύρω από το νήμα εμποδίζει τη διασπορά των ηλεκτρονίων, κατευθύνοντάς τα προς μια μικρή περιοχή στον στόχο, η οποία ονομάζεται εστιακό σημείο. Το μέγεθος του εστιακού σημείου επηρεάζει την τελική ανάλυση της εικόνας. Οι ακτίνες X που παράγονται δεν κατανομούνται ομοιόμορφα λόγω της γωνίας του στόχου, με αποτέλεσμα η ένταση της ακτινοβολίας να είναι μεγαλύτερη στην πλευρά της καθόδου. Τέλος, χρησιμοποιούνται συσκευές που ονομάζονται «collimators» για να διαμορφώσουν και να κατευθύνουν τη δέσμη ακτίνων X, απορρίπτοντας τα χαμηλής ενέργειας φωτόνια και μειώνοντας τη διάχυτη ακτινοβολία, βελτιώνοντας έτσι την ποιότητα της εικόνας.

Η ακτινογραφία βασίζεται στο γεγονός ότι όταν ακτινοβολείται το σώμα με δέσμη ακτίνων X, οι διάφοροι ιστοί θα αλληλεπιδρούν διαφορετικά με τα φωτόνια των ακτίνων X. Αυτό το φαινόμενο αποδίδεται στις διαφορές στην ηλεκτρονιακή πυκνότητα κάθε τύπου ιστού. Όταν ακτινοβολείται ένας ασθενής με ακτίνες X, μέρος της ακτινοβολίας απορροφάται ή εξασθενείται από το σώμα και μέρος της ακτινοβολίας περνάει μέσα από το σώμα. Αυτή είναι η εναπομένουσα ακτινοβολία. Τοποθετώντας ένα ευαίσθητο στην ακτινοβολία φιλμ ή ψηφιακούς ανιχνευτές πίσω από το σώμα του ασθενούς για να απορροφήσουν τα εναπομείναντα φωτόνια, σχηματίζεται ένας χάρτης της απορρόφησης. Πρόκειται για μια δισδιάστατη εικόνα ή προβολή της εσωτερικής δομής του σαρωμένου σώματος. Οι διαφορετικές ιδιότητες απορρόφησης των διάφορων ιστών του σώματος είναι η πηγή της αντίθεσης της ακτινογραφίας. Αυτό είναι που δίνει τη δυνατότητα να διαφοροποιηθεί ένας ιστός από έναν άλλο.

Όπως το φως, έτσι και οι ακτίνες X είναι μέρος του ηλεκτρομαγνητικού φάσματος, αλλά δεν είναι ορατές με γυμνό μάτι. Μεταξύ των τμημάτων της υπεριώδους ακτινοβολίας και των ακτίνων γ του φάσματος, τα φωτόνια των ακτίνων X είναι αρκετά ενεργειακά ώστε να είναι ιονίζοντα. Γενικά, η ιονίζουσα ακτινοβολία μπορεί να σπάσει διατομικούς δεσμούς στην ύλη, συμπεριλαμβανομένου του ανθρώπινου ιστού. Ενώ τα φωτόνια του ορατού φωτός έχουν ενέργειες γύρω στα 1.6 - 3.1 eV («ηλεκτρονιοβόλτ»), οι ιατρικές ακτίνες X έχουν συνήθως ενέργειες που κυμαίνονται από 10 έως 150 keV («κιλοηλεκτρονιοβόλτ»). Δηλαδή, τα φωτόνια των ακτίνων X στην ιατρική απεικόνιση έχουν περίπου 10.000-100.000 φορές υψηλότερες ενέργειες από τα φωτόνια του ορατού φωτός [7].



Εικόνα 2.3 Εικόνα σωλήνα ακτίνων X. Η κάθοδος εκπέμπει ηλεκτρόνια τα οποία στη συνέχεια «βομβαρδίζουν» τον μεταλλικό στόχο (άνοδος), απελευθερώνοντας ενέργεια με τη μορφή θερμότητας και ακτίνων X [7].

2.3 Σύνθεση της εικόνας

Παρακάτω θα αναλυθούν τα στάδια που οδηγούν στη σύνθεση της εικόνας με τη διαδικασία της ιατρικής απεικόνισης με αξονική τομογραφία, από το στάδιο της επιλογής των κατάλληλων παραμέτρων του πρωτοκόλλου σάρωσης, μέχρι την ανακατασκευή και την επεξεργασία της εικόνας.

2.3.1 Επιλογή παραμέτρων πρωτοκόλλου σάρωσης

Η αρχική φάση της διαδικασίας της λήψης εικόνας με τη χρήση αξονικής τομογραφίας, περιλαμβάνει την επιλογή των σωστών παραμέτρων σάρωσης και ανακατασκευής, ώστε να καλύπτονται οι κλινικές ανάγκες της διαγνωστικής εικόνας. Συνήθως, οι μελέτες που αφορούν τη χρήση αξονικής τομογραφίας, ξεκινούν με μια σάρωση πολύ χαμηλής δόσης ακτινοβολίας που δεν περιλαμβάνει την περιστροφή του τμήματος του τομογράφου σε σχήμα δακτυλίου («gantry»), για την παραγωγή μιας αρχικής εικόνας του σώματος. Αυτή η εικόνα βοηθά στον καθορισμό των σημείων εκκίνησης και τερματισμού της κύριας τομογραφικής σάρωσης. Οι παράμετροι σάρωσης, όπως η υψηλότερη τάση που εφαρμόζεται στον σωλήνα ακτίνων X (kVp) και το ρεύμα του σωλήνα, επηρεάζουν την ποιότητα της εικόνας και τη δόση ακτινοβολίας που λαμβάνει ο ασθενής. Το οπτικό πεδίο (field-of-view), είναι μια ακόμα παράμετρος, η οποία αποτελεί την διάμετρο των εγκάρσιων τομών («slices») στο επίπεδο x-y. Η επιλογή φίλτρου «bowtie» είναι επίσης μια σημαντική παράμετρος, διότι μπορεί να μειώσει την δόση ακτινοβολίας που θα λάβει ο ασθενής [7]. Είναι σημαντικό να αναφερθεί ότι πρέπει να επιλέγονται σωστά οι παράμετροι σάρωσης, για να αποφεύγονται επιπλέον σαρώσεις που θα αύξαναν την έκθεση του ασθενή σε ακτινοβολία. Τέλος, οι παράμετροι που αφορούν την ανακατασκευή μπορούν να αλλαχθούν μετά τη σάρωση, εφόσον τα δεδομένα είναι διαθέσιμα.

2.3.2 Συλλογή αρχικού σήματος

Μόλις καθοριστούν οι παράμετροι, η διαδικασία ακτινοβολήσης ξεκινά, και συνήθως διαρκεί μόνο μερικά δευτερόλεπτα ανά σάρωση. Κατά τη διάρκεια αυτού του σταδίου, το σύστημα συλλέγει το ακατέργαστο σήμα I από κάθε ανιχνευτή. Ένας ειδικός ανιχνευτής, γνωστός και

ως ανιχνευτής αναφοράς ή κανονικοποίησης (reference or normalization detector), μετρά την έξοδο του σωλήνα, παρέχοντας την ένταση αναφοράς, I_0 , της σάρωσης. Αυτό το σήμα βάσης χρησιμοποιείται από το σύστημα για την κανονικοποίηση [7]

2.3.3 Διορθώσεις και προ-επεξεργασία

Στο στάδιο της προ-επεξεργασίας, το σήμα εξετάζεται για να εντοπιστούν τυχόν προβληματικές ή ελλείψεις αναγνώσεις. Για παράδειγμα, αν λείπει μια ανάγνωση από έναν συγκεκριμένο ανιχνευτή, αυτό μπορεί να διορθωθεί μέσω παρεμβολής (interpolation). Επιπλέον, εφαρμόζονται διορθώσεις για γνωστά φυσικά φαινόμενα πριν από το στάδιο της ανακατασκευής, χρησιμοποιώντας δεδομένα που συλλέγονται από σαρώσεις που πραγματοποιούνται υπό γνωστές συνθήκες και ρυθμίσεις, όπως σαρώσεις γνωστών αντικειμένων που προσομοιάζουν το ανθρώπινο σώμα (phantoms), με συγκεκριμένες παραμέτρους σάρωσης. Αυτά τα δεδομένα, παρέχουν πληροφορίες για την απόδοση και τις μετρήσεις του συστήματος πριν από τις κλινικές σαρώσεις. Για παράδειγμα, «σαρώσεις αέρα» γίνονται χωρίς αντικείμενο στη διαδρομή της δέσμης, με προκαθορισμένο και γνωστό ρεύμα, τάση και φίλτρο «bowtie». Η ανάλυση αυτών των σημάτων βοηθά στον υπολογισμό των αναγκαίων προσαρμογών για τις διαφορές στην ευαισθησία των στοιχείων του ανιχνευτή. Επίσης, μετράται η μοναδική ευαισθησία κάθε στοιχείου του ανιχνευτή. Αυτά τα δεδομένα χρησιμοποιούνται για την «κανονικοποίηση» του σήματος μεταξύ των διαφορετικών στοιχείων του ανιχνευτή και λειτουργούν ως βάση αναφοράς. Επιπλέον, υπάρχει ο ανιχνευτής αναφοράς που βοηθά στην κανονικοποίηση του συνολικού σήματος μεταξύ των διαφόρων καρτέ και στην αντιστάθμιση των αποκλίσεων στην έξοδο του σωλήνα κατά τη διάρκεια της σάρωσης. Ένα άλλο παράδειγμα είναι η σάρωση γνωστών αντικειμένων με τη χρήση των διαφορετικών φίλτρων «bowtie». Αυτό βοηθά στη συλλογή των απαραίτητων πληροφοριών σχετικά με τις αλλαγές στην κατανομή του φάσματος ακτίνων X που προκαλεί κάθε φίλτρο «bowtie» [7].

2.3.4 Ανακατασκευή εικόνας

Στους αξονικούς τομογράφους, οι εικόνες αποτελούνται από τετράγωνα πλέγματα με συγκεκριμένο αριθμό στοιχείων (εικονοστοιχείων ή αλλιώς, pixels) που μπορεί να επιλέξει ο χρήστης. Για παράδειγμα, ένα πλέγμα 512 x 512 εικονοστοιχείων για μια εγκάρσια τομή («slice») είναι συνηθισμένο. Ο χρήστης καθορίζει επίσης το επιθυμητό εγκάρσιο οπτικό πεδίο (field-of-view), όπως για παράδειγμα διάμετρο 50 cm, το οποίο καθορίζει το μέγεθος των εικονοστοιχείων. Στο συγκεκριμένο παράδειγμα, το μέγεθος θα ήταν 500 mm/512 εικονοστοιχεία, επομένως το πλάτος του εικονοστοιχείου είναι λίγο κάτω από 1 mm. Επιλέγοντας μεγαλύτερο μέγεθος πλέγματος ή μικρότερο οπτικό πεδίο, ο χρήστης μπορεί να μειώσει το μέγεθος που αντιπροσωπεύει κάθε εικονοστοιχείο.

Κατά τη συλλογή των δεδομένων, αυτά μπορούν να κανονικοποιηθούν και να υπολογιστεί το άθροισμα της διαφοράς απορρόφησης. Αυτή είναι η απορρόφηση των διαφόρων ιστών μ_{tissue}^i κατά μήκος της διαδρομής της δέσμης i . Επιπλέον, δεδομένης μιας βασικής έντασης I_0 , που μετράται από τις σαρώσεις αέρα και από τον ανιχνευτή αναφοράς, μπορούμε να ορίσουμε το άθροισμα του γινομένου του συντελεστή απορρόφησης μ_i , του πάχους Δx_i του ιστού που διαπερνάει η ακτινοβολία για κάθε διακριτό βήμα i^{th} , από n το πλήθος βήματα κατά μήκος της διαδρομής της δέσμης των ακτίνων X:

$$\ln\left(\frac{I_0}{I}\right) = \sum_{i=1}^n \mu_i \Delta x_i$$

Εξίσωση 2.1

Στις εικόνες CT, τα ανακατασκευασμένα εικονοστοιχεία (pixels) εικόνας αποδίδονται τιμές που ονομάζονται μονάδες Hounsfield (HU), και προκύπτουν από την εξίσωση:

$$CT\ number\ [HU] = \frac{\mu_{tissue} - \mu_{water}}{\mu_{water} - \mu_{air}} \cdot 1000$$

Εξίσωση 2.2

όπου μ_{tissue} είναι ο μετρούμενος συντελεστής απορρόφησης ενός συγκεκριμένου εικονοστοιχείου και τα μ_{water} και μ_{air} είναι οι συντελεστές απορρόφησης αναφοράς του νερού και του αέρα, αντίστοιχα. Δεδομένου ότι θεωρείται ότι ο αέρας δεν απορροφά (δηλαδή, $\mu_{air} = 0$), η Εξίσωση 2.2 μπορεί να γραφτεί ως:

$$CT\ number\ [HU] = \frac{\mu_{tissue} - \mu_{water}}{\mu_{water}} \cdot 1000$$

Εξίσωση 2.3

Είναι εύκολο να δει κανείς ότι σύμφωνα με αυτή τη μέθοδο κλιμάκωσης, ο αέρας έχει τιμή CT -1000 HU. Η παράμετρος μ_{water} είναι μια σταθερά που εξαρτάται από την ενέργεια, έτσι το σύστημα χρειάζεται να ισορροπήσει αυτή την τιμή ανεξάρτητα από τις παραμέτρους του τομογράφου: το νερό πάντα αποδίδεται με τιμή CT 0 HU. Είναι σημαντικό, ωστόσο, ότι η πραγματική μετρούμενη τιμή δεν είναι ακριβώς 0 HU για το νερό και -1000 HU για τον αέρα, αλλά μια τιμή πολύ κοντά σε αυτή. Αυτό οφείλεται στο θόρυβο, τη διασπορά και άλλες ατέλειες που μπορεί να μετατοπίσουν την τιμή συγκεκριμένων εικονοστοιχείων από την "πραγματική" τους τιμή. Από την Εξίσωση 2.3 μπορεί να φανεί ότι καθώς αυξάνεται η απορρόφηση του μέσου, αυξάνεται και η τιμή CT. Η τιμή του νερού και του αέρα είναι σταθερές, ανεξάρτητες από την ενέργεια της δέσμης ακτίνων X. Η απόδοση τιμών στα εικονοστοιχεία σε μια γνωστή και ισορροπημένη κλίμακα, ανεξάρτητα από τα διαφορετικά μέσα κατά μήκος της διαδρομής των ακτίνων X, είναι ένας σημαντικός παράγοντας για τη συνέπεια των τιμών της αξονικής τομογραφίας που αποδίδονται σε κάθε όργανο της εικόνας. Αυτό έχει τεράστια επίδραση στην ακρίβεια της τελικής διάγνωσης.

Κατά την ανακατασκευή, οι τελικές εικόνες παράγονται βρίσκοντας την συγκεκριμένη τιμή για κάθε εικονοστοιχείο που είναι συνεπής με το άθροισμα των τιμών κατά μήκος κάθε διαδρομής της δέσμης, όπως αντιπροσωπεύεται από τα δεδομένα προβολής που αποκτήθηκαν. Η ανακατασκευή των εικόνων περιλαμβάνει την μετατροπή ακατέργαστων δεδομένων των ακτίνων X σε μια οπτική απεικόνιση των εσωτερικών δομών του ανθρώπινου σώματος. Πιο συγκεκριμένα, χρησιμοποιούνται μέθοδοι όπως η μέθοδος Filtered Back Projection (FBP), κατά την οποία εφαρμόζεται ένα φίλτρο στα ακατέργαστα δεδομένα και ύστερα πραγματοποιείται οπισθοπροβολή (back projecting) για τη δημιουργία εικόνας. Είναι αρκετά αποτελεσματική μέθοδος αλλά μπορεί να οδηγήσει σε σφάλματα εικόνας (image artifacts). Η εικόνα που προκύπτει από την ανακατασκευή είναι ουσιαστικά ένας χάρτης της απορρόφησης κάθε εικονοστοιχείου, παρουσιασμένος σε τιμές HU. Αυτές οι εικόνες αποτελούν τη βάση για την μετέπειτα επεξεργασία.

2.3.5 Περαιτέρω επεξεργασία για κλινική διάγνωση

Σε αυτό το στάδιο, οι βασικές εικόνες εξόδου υφίστανται επεξεργασία και τροποποίηση. Ορισμένες από αυτές τις παραμέτρους περιλαμβάνονται στις επιλογές ανακατασκευής και επιλέγονται όπως απαιτείται. Για παράδειγμα, ο χρήστης μπορεί να επιλέξει τον πυρήνα ανακατασκευής που καθορίζει την ευκρίνεια των εικόνων. Ένας πυρήνας με υψηλή ευκρίνεια θα τονίσει τις άκρες, ενώ ένας πιο «μαλακός» πυρήνας θα λειάνει την εικόνα για να μειώσει τον θόρυβο.

Επιπλέον, ο χρήστης μπορεί να δημιουργήσει επιπλέον σύνολα εικόνων βασισμένα στο αρχικό σύνολο όταν είναι απαραίτητο. Για παράδειγμα, η εμφάνιση λεπτών τομών μπορεί να είναι κατάλληλη για την εξέταση των πνευμόνων, ενώ οι παχύτερες τομές μπορεί να είναι πιο κατάλληλες για την προβολή μεγαλύτερων δομών, μειώνοντας τον θόρυβο της εικόνας.

Με την επεξεργασία των ανακατασκευασμένων τρισδιάστατων δεδομένων, είναι δυνατόν να παρουσιαστούν οποιαδήποτε επίπεδα από οποιαδήποτε κατεύθυνση. Οι εγκάρσιες, οι λοξές και ακόμη και οι κυρτές προβολές μπορούν να εξαχθούν από τα δεδομένα όγκου. Αυτή η τεχνική ονομάζεται Multi-planar Reformatting - MPR (πολυεπίπεδη ανασύνθεση). Αυτό είναι πολύ χρήσιμο, για παράδειγμα, στην περίπτωση των σπονδύλων της σπονδυλικής στήλης, όπου είναι συχνά χρήσιμο να παράγονται εικόνες που είναι ακριβώς κάθετες στους κύριους άξονες των σπονδύλων. Οι διάφορες ανακατασκευές, χρησιμοποιώντας τα ίδια σαρωμένα δεδομένα αλλά διαφορετικές παραμέτρους ανακατασκευής, βελτιώνουν την ικανότητα του γιατρού να ανιχνεύει κλινικά ευρήματα και να παρέχει ακριβή διάγνωση [7].

2.4 Ασφάλεια αξονικής τομογραφίας

Η εφεύρεση της αξονικής τομογραφίας αποτελεί πολύτιμο εργαλείο για πολλές ιατρικές εξετάσεις και διαδικασίες. Χρησιμοποιείται για:

- τη διάγνωση ασθενειών και την παρακολούθηση της θεραπείας χωρίς επεμβάσεις και πόνο,
- τον σχεδιασμό ιατρικών και χειρουργικών θεραπειών, και
- την καθοδήγηση του ιατρικού προσωπικού κατά την εισαγωγή καθετήρων, στεντ ή άλλων συσκευών στο σώμα, τη θεραπεία όγκων ή την απομάκρυνση θρόμβων αίματος και άλλων αποφράξεων.

Ωστόσο, υπάρχουν και ορισμένοι σημαντικοί κίνδυνοι που αφορούν τη χρήση αξονικής τομογραφίας, η οποία χρησιμοποιεί ιονίζουσα ακτινοβολία για τη παραγωγή εικόνων του σώματος. Η ιονίζουσα ακτινοβολία είναι μιας μορφής ακτινοβολία που έχει αρκετή ενέργεια, με αποτέλεσμα να έχει τη δυνατότητα να προκαλέσει καταστροφές στο DNA του ανθρώπου. Η έκθεση σε ιονίζουσα ακτινοβολία φέρει κινδύνους όπως:

- Μια ελαφρώς αυξημένη πιθανότητα να αναπτύξει κάποιος καρκίνο αργότερα στη ζωή του, αν έχει εκτεθεί σε ακτίνες X.
- Επιπτώσεις στους ιστούς όπως ερυθρότητα του δέρματος και απώλεια μαλλιών, που συμβαίνουν σε υψηλά επίπεδα έκθεσης και είναι σπάνιες στις περισσότερες απεικονιστικές εξετάσεις.

Ο κίνδυνος ανάπτυξης καρκίνου από ιατρική απεικόνιση είναι γενικά μικρός και εξαρτάται από:

- Την δόση ακτινοβολίας: Ο κίνδυνος αυξάνεται με τη δόση και τον αριθμό των εξετάσεων.

- Την ηλικία: Ο κίνδυνος είναι μεγαλύτερος για νεότερους ασθενείς.
- Το φύλο: Οι γυναίκες έχουν μεγαλύτερο κίνδυνο από τους άνδρες για την ανάπτυξη καρκίνου λόγω ακτινοβολίας.
- Τη περιοχή σώματος: Ορισμένα όργανα είναι πιο ευαίσθητα στην ακτινοβολία.

Αυτές οι γενικεύσεις βασίζονται σε αναλύσεις μεγάλων πληθυσμιακών δεδομένων. Ενώ μπορεί να μην ταιριάζουν σε όλα τα άτομα, είναι χρήσιμες για την ανάπτυξη μιας συνολικής προσέγγισης στην ασφάλεια ακτινοβολίας, αναγνωρίζοντας πιθανές ομάδες κινδύνου. Η γνώση των τυπικών εκθέσεων ακτινοβολίας για διαφορετικές εξετάσεις είναι σημαντική για την επικοινωνία μεταξύ γιατρού και ασθενούς.

Παρόλο που το όφελος μιας κλινικής εξέτασης με ακτίνες Χ είναι συνήθως πολύ μεγαλύτερο από τον κίνδυνο, πρέπει να γίνονται προσπάθειες για να μειωθεί αυτός ο κίνδυνος περιορίζοντας την περιττή έκθεση σε ιονίζουσα ακτινοβολία. Για να μειωθεί ο κίνδυνος για τον ασθενή, οι εξετάσεις με ιονίζουσα ακτινοβολία πρέπει να εκτελούνται μόνο όταν είναι απαραίτητες για τη διάγνωση μιας ιατρικής κατάστασης, τη θεραπεία μιας ασθένειας ή την καθοδήγηση μιας ιατρικής διαδικασίας. Ωστόσο, αν μια συγκεκριμένη απεικονιστική διαδικασία είναι απαραίτητη και άλλες μέθοδοι χωρίς ή με λιγότερη ακτινοβολία δεν είναι κατάλληλες, τότε τα οφέλη υπερτερούν των κινδύνων και οι ανησυχίες για την ακτινοβολία δεν πρέπει να επηρεάζουν την απόφαση του γιατρού ή του ασθενούς.

3. Καρκίνοι του πεπτικού συστήματος

Το πεπτικό σύστημα, δηλαδή το σύστημα στο οποίο εκτελείται η διαδικασία της πέψης, περιλαμβάνει την γαστρεντερική οδό, η οποία αποτελείται από όργανα τα οποία συνιστούν την κύρια οδό απορρόφησης θρεπτικών ουσιών αλλά και από κάποια όργανα που σχετίζονται άμεσα με αυτή την οδό, δηλαδή το πάγκρεας, το ήπαρ και την χοληδόχο κύστη. Το πάγκρεας είναι ένας σύνθετος αδένας ο οποίος βρίσκεται πίσω από το στομάχι και έχει μήκος 10-15 cm. Παράγει το παγκρεατικό υγρό και την ινσουλίνη, η οποία ρυθμίζει τον μεταβολισμό των υδατανθράκων στον οργανισμό και άλλες ορμόνες. Το ήπαρ, ή αλλιώς συκώτι, είναι ο μεγαλύτερος αδένας στο σώμα μας. Τα ηπατικά κύτταρα παράγουν συνεχώς μικρές ποσότητες χολής, ένα υγρό απαραίτητο για την πέψη των λιπαρών στοιχείων της τροφής και την απορρόφηση ορισμένων βιταμινών, η οποία ρέει προς το δωδεκαδάκτυλο ή προς τη χοληδόχο κύστη όπου και αποθηκεύεται [8]. Σε αυτό το κεφάλαιο θα αναφερθούμε στον καρκίνο του παγκρέατος και του ήπατος, στους τρόπους διάγνωσης και τις διαθέσιμες θεραπείες.

3.1 Καρκίνος του παγκρέατος

Ο καρκίνος του παγκρέατος, αποτελεί αυτή τη στιγμή τον δυσκολότερο τύπο καρκίνου, με πολύ χαμηλό ποσοστό επιβίωσης πέντε ετών μετά τη διάγνωση [9]. Ο καρκίνος του παγκρέατος είναι μια ομάδα ασθενειών που περιλαμβάνουν:

- Τον καρκίνο του ενδοκρινικού συστήματος (καρκίνωμα των κυττάρων islet, νευροενδοκρινικό καρκίνωμα και καρκίνωμα των καρκινοειδών όγκων).
- Τον καρκίνο του εξοκρινικού συστήματος (αδενοκαρκίνωμα του παγκρεατικού πόρου).

Μεταξύ αυτών των τύπων καρκίνου, το αδενοκαρκίνωμα του παγκρέατος είναι ο πιο συχνός τύπος [10].

3.1.1 Συμπτώματα και διάγνωση

Οι ασθενείς με καρκίνο του παγκρέατος παρουσιάζουν συνήθως, τη στιγμή της διάγνωσης, τα παρακάτω συμπτώματα:

- Πόνο στην κοιλιακή χώρα
- Μη φυσιολογικές εξετάσεις λειτουργίας του συκωτιού
- Ίκτερο
- Εμφάνιση διαβήτη
- Δυσπεψία
- Ναυτία ή εμετούς
- Πόνο στη πλάτη
- Απώλεια βάρους
- Ανορεξία

Τα συμπτώματα που εμφανίζονται σε κάθε ασθενή σχετίζονται με τη περιοχή του παγκρέατος στην οποία έχει εμφανιστεί ο όγκος. Ωστόσο, αξίζει να σημειωθεί ότι πολλά από αυτά τα συμπτώματα πολλές φορές έχουν άκακες αιτίες, με αποτέλεσμα να μην γίνεται έγκαιρα η σωστή διάγνωση του καρκίνου [10].

Η διάγνωση βασίζεται κυρίως σε απεικονιστικά εργαλεία όπως η αξονική τομογραφία (CT), η μαγνητική τομογραφία (MRI), η τομογραφία εκπομπής ποζιτρονίων (PET) και το ενδοσκοπικό υπερηχογράφημα (EUS). Η αξονική τομογραφία χρησιμοποιείται συχνά για την διάγνωση του καρκίνου του παγκρέατος (Εικόνα 3.1). Πιο συγκεκριμένα, η αξονική τομογραφία πολλαπλών ανιχνευτών, ή multi-detector computed tomography (MDCT) είναι η

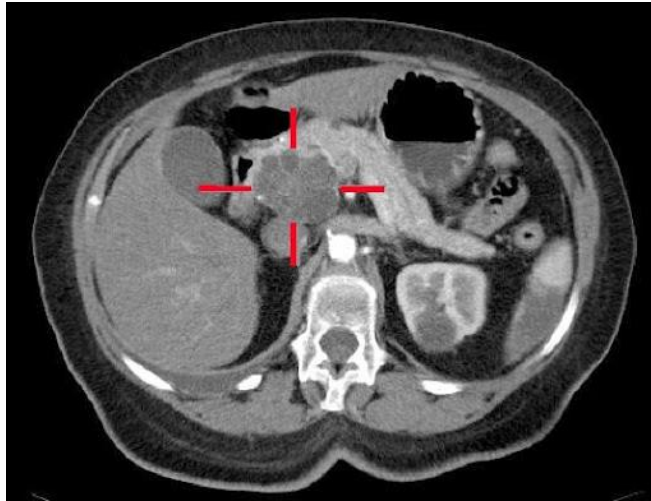
πιο ευρέως χρησιμοποιούμενη και αποτελεσματική απεικονιστική τεχνική για ασθενείς με αδενοκαρκίνωμα του παγκρέατος. Έχει αναφερθεί βελτίωση στην ανίχνευση του καρκίνου αυτού με πρόσφατες ενδείξεις για ευαισθησία ανίχνευσης έως και 96% για την τομογραφία MDCT [11]. Η μαγνητική τομογραφία (MRI) θεωρητικά επιτρέπει την ανίχνευση του καρκινικού όγκου σε πρώιμο στάδιο καθώς παρέχει μια ανάλυση των μορφολογικών αλλαγών του παγκρέατος, ωστόσο μετά από μετά-αναλύσεις, αποδεικνύεται ότι έχει παρόμοια ευαισθησία ανίχνευσης με αυτήν της αξονικής τομογραφίας, με ευαισθησία στο 89% και ειδικότητα 89%. Το ενδοσκοπικό υπερηχογράφημα (EUS), σε αντίθεση με την αξονική και μαγνητική τομογραφία, εκτελείται στον ασθενή όσο αυτός βρίσκεται υπό νάρκωση, και περιλαμβάνει ενδοσκοπική εξέταση του ανώτερου γαστρεντερικού συστήματος με τη χρήση ηχοενδοσκοπίου. Το ενδοσκοπικό υπερηχογράφημα με αναρρόφηση με λεπτή βελόνα (EUS – FNA) έχει αποδειχθεί πολύ ακριβής εξέταση, όχι μόνο για την διάγνωση του όγκου αλλά και για την εύρεση της αιτιολογίας του, με ευαισθησία 85% και ειδικότητα 96% [11]. Τέλος, ο συνδυασμός τομογραφίας ποζιτρονίων και αξονικής τομογραφίας (PET/CT), αν και έχει αποδειχθεί να λειτουργεί παρόμοια με την αξονική τομογραφία ως διαγνωστικό εργαλείο για τον καρκίνο του παγκρέατος, έρευνες αναφέρονται στην πιθανή χρησιμότητα της για εφαρμογή σε συνδυασμό με άλλες τεχνικές [12].

Επιπλέον, διάγνωση γίνεται και με ειδικές εξετάσεις αίματος για τον έλεγχο της ποσότητας συγκεκριμένων ουσιών στο αίμα, αλλά και με τη χρήση βιοψίας, δηλαδή την απομάκρυνση ιστού ή κυττάρων του παγκρέατος και έλεγχος αυτών υπό μικροσκόπιο από τον ιατρό.

3.1.2 Θεραπεία

Οι επιλογές για θεραπεία του καρκίνου του παγκρέατος περιλαμβάνουν:

- Χειρουργική επέμβαση: Επεμβάσεις απομάκρυνσης μερών του παγκρέατος, της χοληδόχου κύστης, του στομαχιού, του λεπτού εντέρου ή και ολόκληρου του παγκρέατος, ανάλογα με το μέγεθος του όγκου. Σε περίπτωση που ο καρκίνος έχει προχωρήσει και δεν μπορεί να αφαιρεθεί, τότε εκτελείται επέμβαση για την ανακούφιση των συμπτωμάτων και τη βελτίωση της ποιότητας ζωής, με την παράκαμψη των αποφράξεων που προκαλεί ο καρκίνος όπως στη χοληδόχο κύστη ή στο πεπτικό σύστημα, επιτρέποντας την καλύτερη ροή της χολής ή της τροφής.
- Ακτινοθεραπεία: Μέθοδος που χρησιμοποιεί ακτίνες X ή άλλα είδη ακτινοβολίας για να καταστρέψει τα καρκινικά κύτταρα ή να σταματήσει την περαιτέρω ανάπτυξη τους. Χρησιμοποιείται μια μηχανή έξω από το σώμα του ασθενούς η οποία στοχεύει την ακτινοβολία στο μέρος του σώματος του ασθενούς που υπάρχει ο όγκος.
- Χημειοθεραπεία: Χορήγηση φαρμάκων στον ασθενή τα οποία καταστέλλουν την ανάπτυξη των καρκινικών κυττάρων. Για τον καρκίνο του παγκρέατος, συνήθως οι ασθενείς δέχονται ενδοφλέβια την αγωγή αυτή, ή μέσω στόματος.
- Χημειοακτινοθεραπεία: Συνδυασμός χημειοθεραπείας και ακτινοθεραπείας για την αύξηση των επιδράσεων και των δύο μεθόδων.
- Στοχευμένη θεραπεία: Η μέθοδος αυτή χρησιμοποιεί φάρμακα ή άλλες ουσίες για την καταστροφή συγκεκριμένων καρκινικών κυττάρων.



Εικόνα 3.1 Αξονική τομογραφία που απεικονίζει αδενοκαρκίνωμα το οποίο βρίσκεται στη κεφαλή του παγκρέατος. Πηγή: [Pancreatic Cancer - Clinical Features - Management - TeachMeSurgery](#)

3.2 Καρκίνος του ήπατος

Ο καρκίνος του ήπατος αποτελεί την πιο συχνή αιτία θανάτου από καρκίνο παγκοσμίως και την πέμπτη πιο κοινή αιτία θανάτου στις Ηνωμένες Πολιτείες [13]. Διακρίνεται σε πρωτογενή και δευτερογενή τύπο καρκίνου. Ο πρωτογενής καρκίνος του ήπατος ξεκινά στο ήπαρ, ενώ ο δευτερογενής εξαπλώνεται στο ήπαρ μέσω άλλων μερών του σώματος. Ο πρωτογενής καρκίνος του ήπατος περιλαμβάνει τρεις τύπους καρκίνου:

- Ηπατοκυτταρικό καρκίνωμα (HCC): Ο πιο κοινός τύπος καρκίνου του ήπατος.
- Ενδοηπατικός καρκίνος (IHC): Αντιπροσωπεύει το 10% - 20% των περιπτώσεων πρωτογενούς τύπου καρκίνου του ήπατος, είναι ο καρκίνος των οδών του ήπατος που μεταφέρουν το υγρό της χολής.
- Αγγειοσάρκωμα: Το πιο σπάνιο είδος καρκίνου του ήπατος (1% των περιπτώσεων πρωτογενούς καρκίνου του ήπατος), ξεκινά στην επένδυση των αιμοφόρων αγγείων στο ήπαρ.

3.2.1 Συμπτώματα και διάγνωση

Τα συμπτώματα του καρκίνου του ήπατος περιλαμβάνουν:

- Ένα σκληρό εξόγκωμα στην δεξιά πλευρά του κορμού, ακριβώς κάτω από τον θώρακα
- Δυσφορία στην άνω κοιλιακή χώρα στη δεξιά πλευρά του κορμού
- Πρησμένη κοιλιακή χώρα
- Πόνος κοντά στη δεξιά ωμοπλάτη ή στην πλάτη
- Ίκτερος
- Εύκολη δημιουργία μελανιών ή αιμορραγία
- Ασυνήθιστη κούραση ή αδυναμία
- Ναυτία και εμετός
- Απώλεια όρεξης ή αίσθημα πληρότητας μετά από μικρό γεύμα
- Ανεξήγητη απώλεια βάρους
- Ανοιχτόχρωμες κενώσεις και σκούρα ούρα
- Πυρετός

Αν ο καρκίνος του ήπατος δεν διαγνωσθεί νωρίς, οι πιθανότητες επιβίωσης ενός ασθενούς είναι χαμηλές, καθώς η ασθένεια εξελίσσεται πολύ γρήγορα. Για παράδειγμα, στο Ηνωμένο

Βασίλειο, μόνο το 20% των ατόμων με καρκίνο του ήπατος εξακολουθούν να ζουν ένα χρόνο μετά τη διάγνωση. Η γρήγορη διάγνωση, λοιπόν, καθίσταται πολύ σημαντικός παράγοντας για την αντιμετώπιση της νόσου. Πιο συγκεκριμένα, η διάγνωση γίνεται με τους παρακάτω τρόπους:

- Εξετάσεις αίματος όπως εξέταση λειτουργίας του ήπατος, για τον έλεγχο των ενζύμων του ήπατος, των πρωτεϊνών και άλλων ουσιών που δείχνουν την κατάσταση της υγείας του ήπατος τους ασθενούς.
- Υπέρηχος: Απεικονίζει τους μαλακούς ιστούς του σώματος.
- Αξονική τομογραφία (CT): Απεικονίζει με λεπτομέρεια το συκώτι, παρέχοντας πληροφορίες για το μέγεθος και τη περιοχή που βρίσκεται ο όγκος.
- Μαγνητική τομογραφία (MRI): Απεικονίζει το εσωτερικό του σώματος χρησιμοποιώντας έναν μεγάλο μαγνήτη και ραδιοκύματα.
- Αγγειόγραμμα: Εξέταση των αιμοφόρων αγγείων του ήπατος μέσω έγχυσης σκιαγραφικού υλικού σε μια αρτηρία.
- Βιοψία: Ανάλυση ηπατικού ιστού υπό μικροσκόπιο.

3.2.2 Θεραπεία

Ο πρωτογενής καρκίνος του ήπατος, και ιδιαίτερα το ηπατοκυτταρικό καρκίνωμα, αποτελεί ένας τύπος καρκίνου δύσκολος στη θεραπεία.

Για τον καρκίνο του ήπατος αρχικού σταδίου, η μερική ηπατεκτομή, δηλαδή η αφαίρεση του μέρους του ήπατος στο οποίο βρίσκεται ο όγκος, μπορεί να αποτελέσει θεραπεία για τον ασθενή. Επιπλέον, για ηπατοκυτταρικό καρκίνωμα αρχικού σταδίου, η μεταμόσχευση ήπατος προσφέρει επίσης μια δυναμικά θεραπευτική επιλογή.

Για τον καρκίνο του ήπατος μεσαίου σταδίου, υπάρχει η επιλογή της δια-αρτηριακής εμβολής (Transarterial embolization), κατά την οποία ένας καθετήρας εισέρχεται σε μια αρτηρία μέσω μιας σχισμής στον μηρό του ασθενούς, και φθάνει στην ηπατική αρτηρία. Στη συνέχεια, μικρά σωματίδια εγχέονται στην αρτηρία για να την αποφράξουν, μπλοκάροντας το οξυγόνο και τα βασικά θρεπτικά συστατικά από τον όγκο. Επιπλέον, υπάρχει και η μέθοδος της δια-αρτηριακής εμβολής με χημειοθεραπεία (Transarterial chemoembolization), κατά την οποία η μέθοδος της εμβολής συνδυάζεται με χημειοθεραπεία. Με την μέθοδο αυτή, η χημειοθεραπεία εγχέεται κατευθείαν στην αρτηρία του ασθενούς και μετά ακολουθεί η απόφραξη της ώστε τα φάρμακα της χημειοθεραπείας να παραμένουν κοντά στον όγκο. Μια ακόμα μέθοδος αποτελεί η ραδιοεμβολή (Radioembolization) η οποία συνδυάζει την εμβολή με την ακτινοθεραπεία, κατά την οποία εγχέονται μικρά σφαιρίδια στην ηπατική αρτηρία τα οποία εγκλωβίζονται κοντά στον όγκο και εκπέμπουν μικρές ποσότητες ακτινοβολίας για μέρες. Η ακτινοβολία αυτή επιδρά κυρίως στον όγκο και δεν εξαπλώνεται στα υπόλοιπα όργανα. Τέλος, η εμβολή χημειοθεραπείας με σφαιρίδια που εκλύουν φάρμακο (Drug-eluting bead chemoembolization) συνδυάζει την δια-αρτηριακή εμβολή με χημειοθεραπεία με σφαιρίδια που εκλύουν φάρμακα χημειοθεραπείας, όπου αρχικά εγχέεται το φάρμακο της χημειοθεραπείας και η αρτηρία αποφράσσεται αφού έχουν εγχυθεί και τα σφαιρίδια με το φάρμακο. Η χημειοθεραπεία βρίσκεται κοντά στον όγκο και τα σφαιρίδια εκλύουν πιο αργά το φάρμακο της χημειοθεραπείας, με αποτέλεσμα τα καρκινικά κύτταρα να είναι πιο πιθανό να καταστραφούν [14].

Για τον καρκίνο του ήπατος τελικού σταδίου, η πρώτη στοχευμένη θεραπεία είναι η χορήγηση του φαρμάκου Sorafenib για ασθενείς με προχωρημένο ηπατοκυτταρικό καρκίνωμα. Για τους ασθενείς που δεν ανταποκρίνονται σε αυτή τη θεραπεία, μερικές φορές χορηγείται χημειοθεραπεία, αν και στις περισσότερες περιπτώσεις η χημειοθεραπεία αποφεύγεται διότι το ηπατοκυτταρικό καρκίνωμα θεωρείται ανθεκτικό στην χημειοθεραπεία, και η συστηματική

χημειοθεραπεία συνήθως δεν είναι ανεκτή από ασθενείς με υποκείμενη ηπατική δυσλειτουργία [14]. Μια άλλη μέθοδος προσπάθειας καταπολέμησης του καρκίνου του ήπατος τελικού σταδίου, είναι η ανοσοθεραπεία, με την οποία χορηγούνται φάρμακα στον ασθενή για να βοηθήσουν το ανοσοποιητικό του σύστημα να βρει και να καταστρέψει τα καρκινικά κύτταρα. Τέλος, μια ανοσοθεραπεία που αναφέρεται ως «Oncolytic Virus Therapy» χρησιμοποιεί ιούς, οι οποίοι ονομάζονται ογκολυτικοί ιοί, για να μολύνουν και να καταστρέψουν τα καρκινικά κύτταρα. Η ανοσοθεραπεία αυτή αποτελεί μια ελπιδοφόρα προσέγγιση για τη θεραπεία του καρκίνου καθώς αυτοί οι φυσικοί ιοί μπορούν να τροποποιηθούν ώστε να μη μολύνουν τόσο τα υγιή κύτταρα και να μεταφέρουν θεραπευτικά φορτία στους όγκους, να παράγουν μόρια που ενισχύουν το ανοσοποιητικό αφού μολύνουν τα καρκινικά κύτταρα και να απελευθερώσουν καρκινικά αντιγόνα μετά τη καταστροφή των καρκινικών κυττάρων [15].

4. Ανάλυση εικόνας με Μηχανική Μάθηση

4.1 Εισαγωγή στην Τεχνητή Νοημοσύνη

Ο Άγγλος μαθηματικός Alan Turing, ο οποίος σήμερα θεωρείται «ο πατέρας της επιστήμης υπολογιστών», ήταν ο πρώτος που εισήγαγε το σενάριο της χρήσης του υπολογιστή για τη προσομοίωση έξυπνης συμπεριφοράς και κριτικής σκέψης, το 1950. Στο βιβλίο «Υπολογιστές και Νοημοσύνη» (“Computers and Intelligence”), περιέγραψε ένα απλό τεστ, το οποίο αργότερα ονομάστηκε «τεστ Turing», για να εξετάσει αν μια μηχανή είναι σε θέση να «σκέφτεται». Μετά από 6 χρόνια, ο John McCarthy περιέγραψε τον όρο «Τεχνητή Νοημοσύνη» ως την επιστήμη και την μηχανική της δημιουργίας έξυπνων μηχανών. Η τεχνητή νοημοσύνη ξεκίνησε ως μια σειρά από «αν, τότε» κανόνες, και με τα χρόνια έχει εξελιχθεί να περιλαμβάνει πιο πολλούς και περίπλοκους αλγορίθμους που δρουν παρόμοια με τον ανθρώπινο εγκέφαλο [16]. Σήμερα, η τεχνητή νοημοσύνη είναι το πεδίο της επιστήμης υπολογιστών που αναπτύσσει υπολογιστικά συστήματα που μπορούν να εκτελούν καθήκοντα που απαιτούν ανθρώπινη νοημοσύνη. Πιο συγκεκριμένα, τα συστήματα αυτά είναι ικανά για καθήκοντα όπως: μάθηση, προσαρμοστικότητα, εξαγωγή συμπερασμάτων, επίλυση προβλημάτων και λήψη αποφάσεων.

Μερικές από τις εφαρμογές της τεχνητής νοημοσύνης σήμερα περιλαμβάνουν:

- Στον τομέα των Μεταφορών: Ανάπτυξη αυτόνομων οχημάτων, όπως τα αυτοκίνητα που κατασκευάζει η εταιρία Tesla.
- Στον τομέα της Υγείας: Ανάπτυξη μοντέλων ανάλυσης ιατρικών εικόνων και διάγνωσης ασθενών, συστήματα παρακολούθησης ασθενών που καταγράφουν ιατρικά δεδομένα σε πραγματικό χρόνο.
- Στον τομέα της Διαφήμισης: Εξατομίκευση των διαφημίσεων, ανάλυση της συμπεριφοράς των καταναλωτών και βελτιστοποίηση των εκστρατειών μάρκετινγκ. Εφαρμογές όπως οι έξυπνοι βοηθοί και τα chatbots παρέχουν προτάσεις και βελτιώνουν την εμπειρία του πελάτη.
- Στο τομέα των Τραπεζών: Ασφάλεια και ταχύτητα στις τραπεζικές συναλλαγές, μέσω εφαρμογής που αναγνωρίζει την ψηφιακή υπογραφή.

Όσον αφορά τη τεχνητή νοημοσύνη συγκεκριμένα στον κλάδο της Ιατρικής, οι εφαρμογές της έχουν αυξηθεί με την εισαγωγή της Μηχανικής Μάθησης (Machine Learning) και της Βαθιάς Μάθησης (Deep Learning). Η Τεχνητή Νοημοσύνη μπορεί να βελτιώσει την ακρίβεια των διαγνώσεων, να αυξήσει την αποδοτικότητα στις κλινικές λειτουργίες και τη ροή εργασίας, να διευκολύνει την παρακολούθηση ασθενειών και θεραπειών και να βελτιώσει την ακρίβεια των ιατρικών διαδικασιών. Αυτό συμβαίνει, διότι τα συστήματα Τεχνητής Νοημοσύνης μπορούν να μαθαίνουν και να προσαρμόζονται πολύ πιο γρήγορα από τον άνθρωπο, έχοντας πρόσβαση σε τεράστιες ποσότητες δεδομένων και επεξεργάζοντας τα σε πραγματικό χρόνο. Αυτό τους επιτρέπει να γίνονται ικανά σε νέους τομείς της ιατρικής πολύ πιο γρήγορα από έναν ανθρώπινο ιατρό.

4.2 Μηχανική Μάθηση

Η Μηχανική Μάθηση αποτελεί υποπεδίο της Τεχνητής Νοημοσύνης. Αναφέρεται σε μοντέλα ή αλγορίθμους που έχουν την ικανότητα μάθησης από δεδομένα, χωρίς απαραίτητα να χρειάζονται την εισαγωγή κανόνων από τον άνθρωπο. Η επιστήμη αυτή προκύπτει από τη τομή που δημιουργούν η στατιστική, που επιδιώκει να μάθει τις «σχέσεις» μεταξύ των δεδομένων,

και της επιστήμης των υπολογιστών, που δίνει έμφαση σε αποδοτικούς υπολογιστικούς αλγορίθμους[17]. Οι εργασίες Μηχανικής Μάθησης συνήθως ταξινομούνται σε τρεις μεγάλες κατηγορίες: την επιβλεπόμενη μάθηση (supervised learning), την μη επιβλεπόμενη μάθηση (unsupervised learning) και την ενισχυτική μάθηση (reinforcement learning).

4.2.1 Επιβλεπόμενη μάθηση

Το σύστημα δέχεται συγκεκριμένα δεδομένα εκπαίδευσης για ένα πρόβλημα, τα οποία αποτελούνται από σύνολα εισόδου (π.χ. εικόνες) και τις αντίστοιχες τιμές εξόδου (π.χ. ετικέτα ή «label» της κάθε εικόνας). Βάσει αυτών των δεδομένων, το σύστημα καλείται να μάθει γενικούς κανόνες με τους οποίους θα μπορέσει να προβλέψει τις τιμές εξόδου για άγνωστα δεδομένα, παρόμοιων προβλημάτων. Πιο συγκεκριμένα, στο σύστημα δίνεται ένα σύνολο δεδομένων εκπαίδευσης $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, όπου «x» οι εισόδοι, και «y» οι αντίστοιχοι εξόδοι για το πρόβλημα (οι ετικέτες των εισόδων x), και «i» ο αριθμός του κάθε δεδομένου εκπαίδευσης. Κατά την επιβλεπόμενη μάθηση, δίνεται στο σύστημα η είσοδος x_i , και έπειτα το σύστημα παράγει την έξοδο \tilde{y}_i . Στη συνέχεια, η έξοδος που δημιουργήσε το σύστημα, \tilde{y}_i , συγκρίνεται με την αληθινή έξοδο (ground truth), y_i , και υπολογίζεται η διαφορά μεταξύ των εξόδων. Η διαφορά αυτή ονομάζεται σφάλμα (error). Με βάση το σφάλμα αυτό, το σύστημα αλλάζει κάποιες εσωτερικές παραμέτρους, και επαναλαμβάνει την ίδια διαδικασία ώσπου να ελαχιστοποιηθεί η διαφορά μεταξύ της αληθινής εξόδου και αυτής που παράγει το σύστημα για κάθε είσοδο.

4.2.2 Μη επιβλεπόμενη μάθηση

Σε αντίθεση με την επιβλεπόμενη μάθηση, το σύστημα σε αυτή τη κατηγορία μάθησης δέχεται εισόδους (x_i) ως δεδομένα εκπαίδευσης για ένα πρόβλημα, αλλά όχι τις αντίστοιχες εξόδους τους (y_i). Το σύστημα καλείται να μάθει μόνο του να αναλύει τα δεδομένα, να δημιουργεί σχέσεις μεταξύ τους και να εξάγει τα απαραίτητα συμπεράσματα. Στη περίπτωση αυτή δεν υπάρχει κάποιο σφάλμα το οποίο το σύστημα μπορεί να χρησιμοποιήσει για να βελτιώσει τις προβλέψεις του.

4.2.3 Ενισχυτική μάθηση

Στο συγκεκριμένο τύπο μάθησης, ένας «πράκτορας» προσπαθεί να μάθει πως να συμπεριφέρεται σε ένα περιβάλλον, όπου η μόνη ανατροφοδότηση που λαμβάνει είναι ένα αριθμητικό σήμα αμοιβής (reward signal). Η ενισχυτική μάθηση δεν πρέπει να θεωρείται μια συγκεκριμένη κατηγορία μεθόδων μάθησης, αλλά ένα πρόβλημα μάθησης. Ο στόχος του πράκτορα είναι να εκτελεί ενέργειες που μεγιστοποιούν το σήμα αμοιβής μακροπρόθεσμα και, καθώς δεν υπάρχει κάποια καθοδήγηση, πρέπει να ανακαλύψει μόνος του ποιες ενέργειες θα αποφέρουν αυτό το αποτέλεσμα.

4.3 Η Μηχανική Μάθηση στην Ιατρική

Οι εφαρμογές της Μηχανικής Μάθησης, και γενικότερα της Τεχνητής Νοημοσύνης, στον τομέα της Ιατρικής βοηθούν τους ιατρούς να παρέχουν πιο ακριβείς και αποτελεσματικές υπηρεσίες υγείας, μειώνοντας το κόστος και βελτιώνοντας τα αποτελέσματα για τους ασθενείς. Αξίζει να σημειωθεί, ωστόσο, ότι τα συστήματα Τεχνητής Νοημοσύνης αυτή τη στιγμή, δεν προορίζονται

για να αντικαταστήσουν τους ιατρούς, αλλά να βοηθήσουν στην ιατρική διαδικασία. Παρακάτω παρουσιάζονται κάποια από τα παραδείγματα τέτοιων τεχνολογιών σε διάφορους τομείς της υγειονομικής περίθαλψης.

Λήψη ιστορικού

Ένας «έξυπνος βοηθός» που ονομάζεται DAX copilot, της εταιρίας Nuance, καταγράφει αυτόματα τις συναντήσεις των ιατρών με τους ασθενείς με ακρίβεια, και τις μετατρέπει σε ιατρικές σημειώσεις για να τις διαβάσει ο εκάστοτε ιατρός [18]. Μια άλλη αντίστοιχη τεχνολογία, της εταιρείας Suki [19], βοηθά τους ιατρούς στην καταγραφή σημειώσεων, διατηρεί ηλεκτρονικούς φακέλους υγείας και απαντά σε ερωτήσεις.

Διάγνωση

Σε μια πρόσφατη μελέτη, η διάγνωση της νόσου του Πάρκινσον μέσω χαρακτηριστικών σήματος φωνής προτάθηκε με τη χρήση μεθόδων Μηχανικής Μάθησης και Βαθιάς Μάθησης, με τα μοντέλα MLP (Multilayer Perceptron) και SVM (Support Vector Machine) να δείχνουν υψηλές ακρίβειες 98.3% και 95% αντίστοιχα [20]. Σε άλλη μελέτη, έγινε χρήση Τεχνητής Νοημοσύνης για τη διάγνωση του ιού COVID-19 μέσω ήχων βήχα, με στόχο τη μείωση του κόστους θεραπείας και την πρόληψη της ασθένειας. Χρησιμοποιήθηκαν αλγόριθμοι ταξινόμησης επιβλεπόμενης μάθησης, όπως Μηχανή Υποστήριξης Διανυσμάτων (SVM), τυχαίο δάσος (Random Forest) και διάφοροι τύποι Νευρωνικών Δικτύων (Neural Networks), όπως τα ANN (Artificial Neural Networks) και τα LSTM (Long short-term memory networks). Τα μοντέλα πέτυχαν κατά μέσο όρο ακρίβεια της τάξης του 83%, με το δίκτυο LSTM να φτάνει την υψηλότερη ακρίβεια στο 95% [21]. Επιπλέον, έχουν δημιουργηθεί εργαλεία που χρησιμοποιούν Μηχανική Μάθηση όπως αυτό της εταιρίας Aidoc [22], το οποίο βοηθάει στην ανάλυση ακτινογραφιών, αξονικών τομογραφιών και μαγνητικών τομογραφιών, για τη διάγνωση διαφόρων παθήσεων, και το εργαλείο της εταιρείας quire.ai [23] το οποίο χρησιμοποιεί τεχνολογία Βαθιάς Μάθησης για την ανάλυση ακτινογραφιών.

Πρόγνωση

Η πρόγνωση είναι η πρόβλεψη της εξέλιξης μιας ασθένειας. Είναι σημαντική στην υγειονομική περίθαλψη, καθώς βοηθά τους γιατρούς να αποφασίζουν για τις κατάλληλες θεραπείες και τη μελλοντική διαχείριση της υγείας των ασθενών. Προς αυτή τη κατεύθυνση, επιστήμονες στο εργαστήριο Τεχνητής Νοημοσύνης Google DeepMind, ανέπτυξαν ένα μοντέλο Βαθιάς Μάθησης ικανό να προβλέψει την εμφάνιση οξείας νεφρικής βλάβης έως και 48 ώρες νωρίτερα, χρησιμοποιώντας δεδομένα ασθενών [24].

Θεραπεία

Στο τομέα της εύρεσης θεραπειών, έχουν δημιουργηθεί συστήματα υποστήριξης κλινικών αποφάσεων (CDSS), λογισμικά που χρησιμοποιούν αλγόριθμους Τεχνητής Νοημοσύνης για να παρέχουν στους ιατρούς πληροφορίες και συστάσεις σε πραγματικό χρόνο για τη φροντίδα των ασθενών [25]. Τα CDSS ενσωματώνονται συνήθως στα συστήματα ηλεκτρονικών φακέλων υγείας (EHR) και προσφέρουν μια σειρά από λειτουργίες υποστήριξης αποφάσεων για την βελτίωση της υγειονομικής περίθαλψης. Για παράδειγμα, ένα τέτοιο εργαλείο είναι το IBM Watson for Oncology (WFO) [26], το οποίο χρησιμοποιεί Μηχανική Μάθηση για την ανάλυση δομημένων, και μη, δεδομένων και παρέχει τεκμηριωμένες συστάσεις θεραπεία για καρκινοπαθείς. Σύμφωνα με μετά-αναλύσεις, το εργαλείο αυτό, όταν χρησιμοποιείται συμπληρωματικά με τις γνώσεις και την εμπειρία του ιατρού, μπορεί να προσφέρει χαμηλότερο ποσοστό σφαλμάτων στην επιλογή θεραπευτικών σχημάτων χημειοθεραπείας, βελτίωση στην αποδοτικότητα της εργασίας των ιατρών και γενικότερα καλύτερη επικοινωνία μεταξύ ιατρού και ασθενή [26].

4.4 Ανάλυση εικόνας

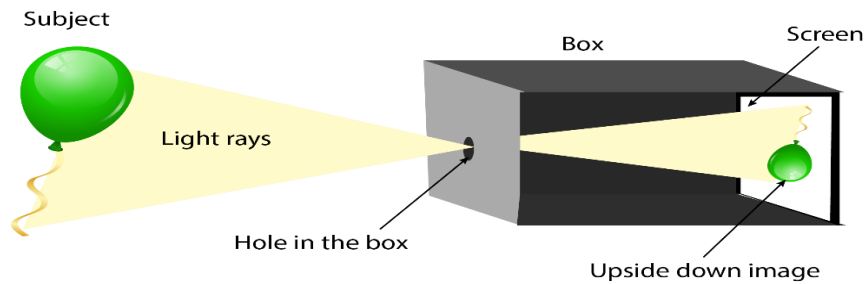
Η όραση επιτρέπει στους ανθρώπους να αντιλαμβάνονται και να κατανοούν τον κόσμο γύρω τους, ενώ η υπολογιστική όραση στοχεύει στην αναπαραγωγή του αποτελέσματος της ανθρώπινης όρασης με την ηλεκτρονική αντίληψη και κατανόηση μιας εικόνας. Η υπολογιστική όραση (Computer Vision) αποτελεί επιστημονικό πεδίο της Τεχνητής Νοημοσύνης το οποίο έκανε την εμφάνιση του μετά το έτος 1980. Μερικές από τις εφαρμογές της επιστήμης αυτής, είναι:

- Ταξινόμηση εικόνων (Image Classification): Τεχνική κατά την οποία ο υπολογιστής ταξινομεί την εικόνα σε μια μεγαλύτερη κατηγορία παρόμοιων εικόνων, με βάση τα περιεχόμενα της.
- Ανίχνευση αντικειμένων (Object Detection): Αναγνώριση συγκεκριμένων αντικειμένων στην εικόνα, και προσδιορισμός της ακριβούς τους θέσης μαζί με την ετικέτα τους.
- Αναγνώριση οπτικών χαρακτήρων (Optical Character Recognition): Αναγνώριση γραμμάτων και αριθμών σε εικόνες και έγγραφα.

4.4.1 Αναπαράσταση εικόνας

Το μάτι λειτουργεί σαν μια φωτογραφική μηχανή ακριβείας μέσω του οποίου αντιλαμβανόμαστε τον τρισδιάστατο κόσμο. Όταν οι ακτίνες φωτός εισέρχονται στο μάτι μέσω της κόρης, περνούν από τον κερατοειδή και τον φακό, οι οποίοι τις εστιάζουν στον αμφιβληστροειδή, δημιουργώντας μια ανεστραμμένη εικόνα. Ο αμφιβληστροειδής, που αποτελείται από φωτοευαίσθητα κύτταρα, μετατρέπει αυτή την εικόνα σε ηλεκτρικά σήματα. Αυτά τα σήματα μεταφέρονται μέσω του οπτικού νεύρου στον εγκέφαλο, ο οποίος επεξεργάζεται και αναστρέφει την εικόνα, επιτρέποντάς μας να αντιλαμβανόμαστε τον κόσμο σωστά.

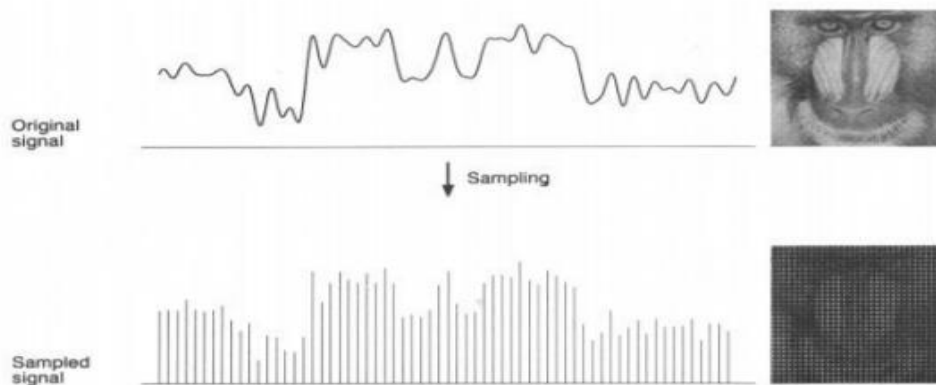
Η λειτουργία μιας κάμερας είναι παρόμοια με αυτή της ανθρώπινης όρασης. Ας θεωρήσουμε μια «μικροπτική κάμερα», η οποία ιστορικά, είναι η αρχή της φωτογραφίας, που αποτελείται από μια μικρή οπή στο κέντρο μιας έδρας ενός εσωτερικά σκοτεινού θαλάμου και στην απέναντι έδρα να βρίσκεται ένα φωτοευαίσθητο φιλμ ή χαρτί. Τότε, οι ακτίνες φωτός εισέρχονται μέσω του μικρού ανοίγματος και σχηματίζουν μια αντεστραμμένη εικόνα στην επιφάνεια απέναντι από το άνοιγμα. Μια τέτοια κατασκευή απεικονίζεται στην εικόνα 4.1. Έτσι, σχηματίζεται η δισδιάστατη (2D) εικόνα από το τρισδιάστατο (3D) αντικείμενο του πραγματικού κόσμου. Στις παλαιότερες κάμερες, που χρησιμοποιούνταν φιλμ, η φωτογραφία αποτυπωνόταν ως αναλογικό σήμα, δηλαδή ένα συνεχές σήμα. Σήμερα, για την επεξεργασία των φωτογραφιών από τον υπολογιστή, απαιτείται η μετατροπή του αναλογικού σε ψηφιακό σήμα, δηλαδή σε διακριτό σήμα. Οι σημερινές φωτογραφικές μηχανές, αντί για φιλμ, χρησιμοποιούν αισθητήρες ψηφιακής απεικόνισης (CCD/CMOS), που μετατρέπουν το αναλογικό σήμα σε διακριτό. Για να δημιουργήσουμε μια ψηφιακή εικόνα, πρέπει να μετατρέψουμε τα συνεχή δεδομένα σε ψηφιακή μορφή. Αυτό περιλαμβάνει δύο διαδικασίες - τη δειγματοληψία και την κβάντιση. Μια εικόνα μπορεί να είναι συνεχής ως προς τις συντεταγμένες x και y , καθώς και ως προς το πλάτος. Για να την μετατρέψουμε σε ψηφιακή μορφή, πρέπει να δειγματοληψήσουμε τη συνάρτηση τόσο στις συντεταγμένες όσο και στο πλάτος. Η ψηφιοποίηση των τιμών των συντεταγμένων ονομάζεται δειγματοληψία και αποτελεί το πρώτο στάδιο της μετατροπής συνεχούς σήματος σε διακριτό, ενώ το δεύτερο βήμα, η ψηφιοποίηση των τιμών του πλάτους, είναι η κβάντιση.



Εικόνα 4.1 Αναπαράσταση μικροπικής κάμερας. Πηγή: <https://byjus.com/question-answer/an-image-formed-by-a-pinhole-camera-is-always/>.”

4.4.2 Δειγματοληψία

Η δειγματοληψία είναι η διαδικασία λήψης δειγμάτων από ένα συνεχές σήμα σε καθορισμένα χρονικά διαστήματα ή διαστήματα συντεταγμένων. Στην περίπτωση των εικόνων, η δειγματοληψία εφαρμόζεται στις συντεταγμένες (x, y) , όπως προαναφέρθηκε, για να δημιουργηθεί ένα πλέγμα από διακριτά σημεία ή εικονοστοιχεία (pixels), όπως φαίνεται στην εικόνα 4.2. Έτσι, ένα συνεχές σήμα $f(x,y)$ μετατρέπεται σε έναν διδιάστατο πίνακα (matrix) με πλήθος M γραμμές και N στήλες. Όσο περισσότερα δείγματα παίρνουμε από το αρχικό σήμα, τόσα περισσότερα pixels θα έχουμε στον τελικό πίνακα. Ως αποτέλεσμα, η τελική εικόνα θα έχει καλύτερη ποιότητα.



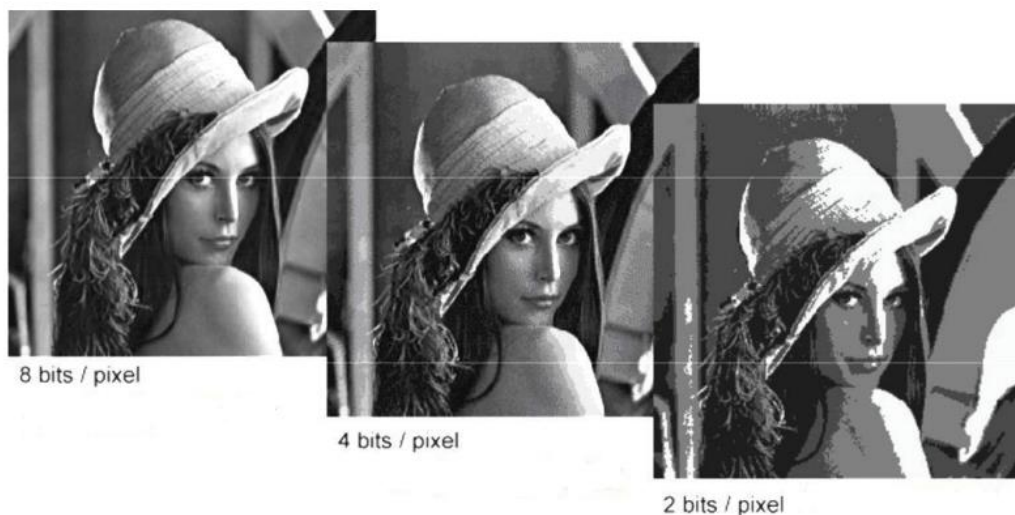
Εικόνα 4.2 Αναπαράσταση δειγματοληψίας εικόνας. Πηγή: [Formal Powerpoint presentation \(mitu.co.in\)](https://www.mitu.co.in/)

4.4.3 Κβάντιση

Η κβάντιση είναι το δεύτερο βήμα στη διαδικασία μετατροπής εικόνων από αναλογική μορφή σε ψηφιακή, και ακολουθεί τη δειγματοληψία. Αφού οι συνεχείς συντεταγμένες της εικόνας έχουν δειγματοληφθεί και μετατραπεί σε ένα πλέγμα από εικονοστοιχεία, οι τιμές έντασης (ή αλλιώς, φωτεινότητας) αυτών των εικονοστοιχείων πρέπει να κβαντιστούν. Ένα επίπεδο φωτεινότητας ονομάζεται και γκριζα διαβάθμιση (gray level).

Κατά την διαδικασία της κβάντισης, οι συνεχείς τιμές έντασης κάθε εικονοστοιχείου μετατρέπονται σε διακριτές τιμές. Πιο συγκεκριμένα, οι τιμές της έντασης αντιστοιχίζονται στους πλησιέστερους αριθμούς από ένα πεπερασμένο σύνολο, το οποίο αποτελείται από 2^B

επίπεδα έντασης, όπου B ο αριθμός των bits/ pixel. Για παράδειγμα, σε μια εικόνα που κβαντίζεται με 8-bit/pixel, οι τιμές έντασης περιορίζονται σε 256 διακριτά επίπεδα, από 0 έως 255. Αυτή είναι μια τυπική περίπτωση μονόχρωμης εικόνας. Στα συνήθη format ψηφιακών εικόνων (π.χ. JPEG), οι μονόχρωμες εικόνες κβαντίζονται με 8 bits/ pixel ενώ οι έγχρωμες εικόνες (RGB) με 24 bits/ pixel. Αν ο αριθμός των bit/ pixel είναι μικρός, π.χ. 2 bits/ pixel, παρατηρούνται αλλοιώσεις στην εικόνα, όπως για παράδειγμα εμφάνιση ψευδών περιγραμμάτων (false contours). Το φαινόμενο αυτό παρατηρείται στην εικόνα 4.3.

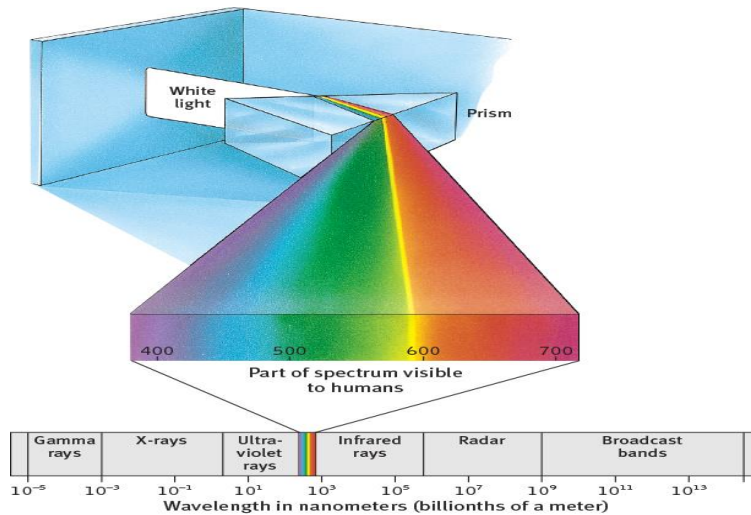


Εικόνα 4.3 Κβάντιση εικόνας με διαφορετικά επίπεδα bits/pixel. Πηγή: [MUC_Lecture_2023_121254131.pdf](https://uomus.edu.iq/MUCLecture_2023_121254131.pdf) (uomus.edu.iq)

Συνοψίζοντας, μια μονόχρωμη ψηφιακή εικόνα αποτελεί έναν δισδιάστατο πίνακα $M \times N$, τα στοιχεία του οποίου ονομάζονται εικονοστοιχεία, ή αλλιώς pixels. Η ένταση (ή φωτεινότητα) του κάθε pixel αποτελούν ένα σύνολο ακέραιων αριθμών με εύρος $(0, \dots, 2^B - 1)$.

4.4.4 Χρωματικοί χώροι

Οι περισσότερες πηγές φωτός εκπέμπουν φως σε διάφορα μήκη κύματος, το οποίο ανακλάται από τις επιφάνειες των αντικειμένων που έχουν και αυτές διαφορετική συμπεριφορά σε διαφορετικά μήκη κύματος. Ως «χρώμα της πηγής» εννοούμε το φάσμα των διαφόρων μηκών κύματος στα οποία η πηγή εκπέμπει φως, και αποτελεί την κατανομή έντασης σε μήκη κύματος. Ο άνθρωπος αντιλαμβάνεται μόνο ένα περιορισμένο κομμάτι του ηλεκτρομαγνητικού φάσματος, από μήκος κύματος περίπου 380 nm έως 740 nm, όπως απεικονίζεται στην εικόνα 4.4.



Εικόνα 4.4 Ένα πρίσμα, σύμφωνα με τον Νεύτωνα, μπορεί να αποσυνθέσει το λευκό φως σε ένα συνεχές φάσμα διαφορετικών καθαρών χρωμάτων. Από όλο το ηλεκτρομαγνητικό φάσμα, ο άνθρωπος αντιλαμβάνεται το εύρος μηκών κύματος 380 - 740 nm [19].

Αξίζει να αναφερθούμε και στον μηχανισμό του ματιού που μας επιτρέπει να βλέπουμε χρώματα γύρω μας. Όταν οι ακτίνες φωτός εισέρχονται στο μάτι μέσω της κόρης, περνούν από τον κερατοειδή και τον φακό και οδηγούνται στον αμφιβληστροειδή χιτώνα. Ο αμφιβληστροειδής έχει δύο διαφορετικούς τύπους κυττάρων που ανιχνεύουν και αντιδρούν στο φως — τα ραβδία και τα κωνία. Τα κύτταρα αυτά, που είναι ευαίσθητα στο φως, ονομάζονται φωτοϋποδοχείς. Τα ραβδία ενεργοποιούνται σε χαμηλό ή αμυδρό φως. Αντίθετα, τα κωνία διεγείρονται σε φωτεινότερα περιβάλλοντα. Οι περισσότεροι από εμάς έχουμε περίπου 6 εκατομμύρια κωνία και 110 εκατομμύρια ραβδία. Τα κωνία περιέχουν φωτοχρωστικές ουσίες, ή αλλιώς, μόρια ανίχνευσης χρώματος. Τα κωνία κατηγοριοποιούνται σε τρεις τύπους με βάση το εύρος του μήκους κύματος που ανιχνεύουν: S-κωνία, (short) με μέγιστη ευαισθησία περίπου στα 440 nm, M-κωνία (medium) περίπου στα 540 nm, και L-κωνία (long) περίπου στα 580 nm. Σε ένα πιο σκοτεινό περιβάλλον, το φως που αντανακλάται από ένα αντικείμενο θα διεγείρει μόνο τα ραβδία των ματιών. Ως αποτέλεσμα, εάν ενεργοποιούνται μόνο τα ραβδία, δεν βλέπουμε χρώμα, μόνο αποχρώσεις του γκρι .

Ωστόσο, η ανθρώπινη όραση είναι επιρρεπής σε διάφορες ψευδαισθήσεις. Το αντιληπτό χρώμα επηρεάζεται και από τα χρώματα και την ερμηνεία της σκηνής που περιβάλλει το παρατηρούμενο χρώμα. Επιπλέον, η προσαρμογή των ματιών σε μεταβαλλόμενες συνθήκες φωτισμού δεν είναι πολύ γρήγορη και η αντίληψη επηρεάζεται από αυτήν την προσαρμογή. Παρ' όλα αυτά, στην χρωματομετρία, την επιστήμη μέτρησης του χρώματος, όπως αυτή ορίζεται από την CIE (International Commission on Illumination) υποθέτουμε ότι δεν παίζει ρόλο το τμήμα της αντίληψης του χρώματος που εξαρτάται από το περιβάλλον. Το 1931, η CIE δημοσίευσε τους χρωματικούς χώρους CIE 1931, οι οποίοι ορίζουν τη σχέση μεταξύ του ορατού φάσματος και της οπτικής αίσθησης συγκεκριμένων χρωμάτων από την ανθρώπινη χρωματική όραση. Πιο συγκεκριμένα, οι χρωματικοί χώροι CIE είναι μαθηματικά μοντέλα που δημιουργούν έναν «τυπικό παρατηρητή», ο οποίος επιχειρεί να προβλέψει την αντίληψη μοναδικών αποχρώσεων χρώματος. Αυτοί οι χρωματικοί χώροι είναι βασικά εργαλεία που παρέχουν τα θεμέλια για τη μέτρηση του χρώματος σε διάφορους βιομηχανικούς τομείς, όπως μελάνια, βαφές και χρώματα.

Για να περιγράψουμε το χρώμα, μπορούμε να επιλέξουμε ως κύρια χρώματα τρία ανεξάρτητα πρωταρχικά χρώματα $P_1(\lambda)$, $P_2(\lambda)$, $P_3(\lambda)$, έστω το κόκκινο, το πράσινο και το μπλε. Τότε, το άγνωστο χρώμα $F(\lambda)$ θα είναι ο γραμμικός συνδυασμός των τριών πρωταρχικών χρωμάτων:

$$F(\lambda) = \alpha_1 P_1(\lambda) + \alpha_2 P_2(\lambda) + \alpha_3 P_3(\lambda),$$

Εξίσωση 4.1

όπου $(\alpha_1, \alpha_2, \alpha_3)$ τα ποσοστά μείξης των πρωταρχικών χρωμάτων και ο τελεστής $\overset{H}{=}$ αποτελεί οπτικό ταιρίασμα από ανθρώπινο παρατηρητή ενός τεστ χρώματος με μείγμα των πρωταρχικών, και όχι αριθμητική ισότητα.

Η μέτρηση του χρώματος γίνεται με την αναπαράσταση του ως διάνυσμα, συνήθως τρισδιάστατο, σε κατάλληλους διανυσματικούς χώρους χρώματος, γραμμικούς ή μη-γραμμικούς. Παρακάτω παρατίθενται ορισμένοι γραμμικοί χρωματικοί χώροι.

Χρωματικός χώρος RGB

Ο χώρος RGB προκύπτει αν κάνουμε χρήση μονοχρωματικών πρωταρχικών χρωμάτων, με μήκη κύματος στην κόκκινη (R), πράσινη (G) και μπλέ (B) ζώνη. Κάθε χρώμα που μπορεί να παρασταθεί από αυτό το μοντέλο είναι ένας μη-αρνητικός συνδυασμός των τριών πρωταρχικών χρωμάτων, άρα ονομάζουμε τον χώρο RGB «προσθετικό». Όταν οι δύο συνιστώσες έχουν μοναδιαίους συντελεστές και η τρίτη συνιστώσα έχει μηδενικό συντελεστή, τότε παίρνουμε τρία δευτερεύοντα χρώματα, δηλαδή το κυανό (Cyan = G + B), το κίτρινο (Yellow = R + G) και το φούξια (Magenta = R + B). Μπορούμε να υπολογίσουμε τις συντεταγμένες (R, G, B) ενός χρώματος $F(\lambda)$, χρησιμοποιώντας τις συναρτήσεις χρωματικού ταιριάσματος $\bar{r}(\lambda)$, $\bar{g}(\lambda)$, $\bar{b}(\lambda)$:

$$R = \int_0^\infty \bar{r}(\lambda) F(\lambda) d\lambda, \quad G = \int_0^\infty \bar{g}(\lambda) F(\lambda) d\lambda, \quad B = \int_0^\infty \bar{b}(\lambda) F(\lambda) d\lambda$$

Εξίσωση 4.2

Για να εξηγήσουμε τις συναρτήσεις χρωματικού ταιριάσματος, το διάγραμμα των οποίων απεικονίζεται στην εικόνα 4.5, ας υποθέσουμε ότι έχουμε τρεις φωτεινές πηγές, μια κόκκινη, μια πράσινη και μια μπλε με γνωστά μήκη κύματος λ . Αν κατευθύνουμε αυτές τις φωτεινές πηγές στο ίδιο σημείο σε έναν τοίχο, μπορούμε να ρυθμίσουμε την ισχύ κάθε πηγής μέχρι το σημείο να έχει ένα συγκεκριμένο χρώμα; Αυτή είναι ακριβώς η ερώτηση στην οποία μπορούν να απαντήσουν οι συναρτήσεις χρωματικού ταιριάσματος (CMFs). Οι συναρτήσεις αυτές, έχουν τρία βασικά χρώματα και κάθε συνάρτηση δίνει την ποσότητα του βασικού χρώματος που χρειάζεται για τη δημιουργία του εκάστοτε επιθυμητού χρώματος, όταν και τα τρία χρώματα αναμειχθούν. Τα βασικά μήκη κύματος των συναρτήσεων $\bar{r}(\lambda)$, $\bar{g}(\lambda)$, $\bar{b}(\lambda)$ είναι τα 700 nm, τα 546.1 nm και τα 435.8 nm αντίστοιχα.

Τέλος, ο χρωματικός χώρος RGB, χρησιμοποιείται ευρέως στη βιομηχανία, όπως στις τηλεοράσεις, τις οθόνες υπολογιστών και τα κινητά τηλέφωνα, για την απεικόνιση των χρωμάτων των εικόνων μέσω της πρόσθεσης των βασικών χρωμάτων.

Χρωματικός χώρος XYZ

Σε αυτόν τον χρωματικό χώρο, χρησιμοποιούνται μη-αρνητικές συναρτήσεις χρωματικού ταιριάσματος $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, $\bar{z}(\lambda)$, όπως φαίνονται από το διάγραμμα της εικόνας 4.5, που αντιστοιχούν σε φανταστικά πρωταρχικά χρώματα για να μετρήσουμε τις αντίστοιχες συντεταγμένες (X, Y, Z):

$$X = \int_0^{\infty} \bar{x}(\lambda) F(\lambda) d\lambda, \quad Y = \int_0^{\infty} \bar{y}(\lambda) F(\lambda) d\lambda, \quad Z = \int_0^{\infty} \bar{z}(\lambda) F(\lambda) d\lambda$$

Εξίσωση 4.3

Η συνιστώσα Y ορίζεται ως γκρίζα φωτεινότητα (luminance). Οι συνιστώσες X, Z μοιάζουν προσεγγιστικά στο κόκκινο και στο μπλε αντίστοιχα, όμως δεν είναι το ίδιο. Ο χώρος αυτός δημιουργήθηκε για τη μέτρηση χρώματος από φάσματα φωτός. Σήμερα, αν και θεωρείται ξεπερασμένος από επαγγελματίες χρωματομέτρους, αποτελεί χρήσιμος για ενδιάμεσες μετατροπές από και προς πολλές παραλλαγές του χώρου RGB. Οι συντεταγμένες (X, Y, Z) προκύπτουν από τον γραμμικό συνδυασμό των συντεταγμένων (R, G, B) με βάσει τον ακόλουθο μετασχηματισμό:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.412453 & 0.357580 & 0.180423 \\ 0.212671 & 0.715160 & 0.072169 \\ 0.019334 & 0.119193 & 0.950227 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}_{709}$$

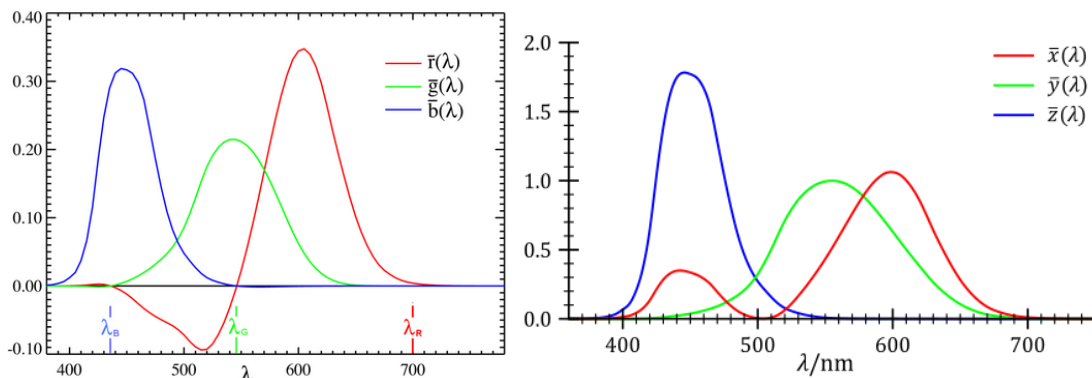
Εξίσωση 4.4

Επιπλέον, η κανονικοποίηση των συνιστωσών (X, Y, Z) γίνεται με βάση τη παρακάτω σχέση:

$$(x, y, z) = \left(\frac{X}{X+Y+Z}, \frac{Y}{X+Y+Z}, \frac{Z}{X+Y+Z} \right)$$

Εξίσωση 4.5

Οι κανονικοποιημένες συνιστώσες διαγράφουν ένα επίπεδο, αφού $x + y + z = 1$. Στο επίπεδο (x, y), το οποίο αποτελεί επίπεδο ανεξάρτητων χρωματικών συνιστωσών, μπορούμε να αναπαραστήσουμε τα αντιληπτά από τον άνθρωπο χρώματα. Το διάγραμμα αυτό είναι γνωστό ως CIE-xy χρωματικό διάγραμμα (chromaticity diagram). Η περιοχή με τα χρώματα πάνω στο διάγραμμα αυτό αναπαριστά όλες τις χρωματικότητες που είναι αντιληπτές από τον άνθρωπο, και ονομάζεται gamut of human vision (GHV).



Εικόνα 4.5 Αναπαράσταση καμπυλών των συναρτήσεων χρωματικού ταιριάσματος του χώρου RGB (αριστερά) και XYZ (δεξιά) [20].

Οι χώροι RGB και XYZ, ενώ έχουν αρκετά πλεονεκτήματα λόγω της αλγεβρικής δομής των γραμμικών χώρων, έχουν και ένα βασικό μειονέκτημα: Μετά την αρχική αίσθηση χρώματος, δεν αντιστοιχούν καλά στην αντίληψη του χρώματος από τον άνθρωπο. Έτσι,

έχουν δημιουργηθεί μη-γραμμικοί χώροι χρώματος, ορισμένοι από τους οποίους θα αναλυθούν στη συνέχεια, οι οποίοι σχετίζονται περισσότερο με την ανθρώπινη αντίληψη των χρωμάτων.

Χρωματικός χώρος YCbCr

Ο χρωματικός χώρος YCbCr χρησιμοποιείται ευρέως στην ψηφιακή φωτογραφία και το βίντεο, εξαιτίας της αποτελεσματικότητάς του στη συμπίεση και την αποθήκευση δεδομένων χρώματος. Η διαφορά μεταξύ των χρωματικών χώρων YCbCr και RGB είναι ότι με τον RGB αναπαριστούμε τα χρώματα ως συνδυασμούς κόκκινων, πράσινων και μπλε σημάτων, ενώ με τον χώρο YCbCr αναπαριστούμε τα χρώματα ως συνδυασμούς ενός σήματος φωτεινότητας και δύο χρωματικών σημάτων. Στον χώρο YCbCr, το Y είναι η φωτεινότητα (luma), το Cb είναι το μπλε μείον τη φωτεινότητα (B-Y) και το Cr είναι το κόκκινο μείον τη φωτεινότητα (R-Y). Το κανάλι φωτεινότητας, που συνήθως σημειώνεται ως Y (ή πιο σωστά Y', που υποδεικνύει ότι προκύπτει από την Y με τον μη-γραμμικό μετασχηματισμό του gamma-correction), προσεγγίζει το περιεχόμενο της εικόνας σε ασπρόμαυρο. Τα δύο χρωματικά κανάλια, Cb και Cr, είναι κανάλια χρωματικής διαφοράς. Γενικά, ο χώρος YCbCr προκύπτει από τον γραμμικό μετασχηματισμό των (R', G', B') που είναι οι gamma-corrected τιμές των (R, G, B). Παρ' όλο που προκύπτει από γραμμικό συνδυασμό, στο σύνολο του ο χώρος αυτός είναι μη-γραμμικός γιατί προηγείται μη-γραμμικός μετασχηματισμός gamma-correction των συνιστωσών του RGB.

Ο αντίστοιχος χώρος για αναλογικό βίντεο ονομάζεται YPbPr και ορίζεται:

$$Y' = k_r R' + (1 - k_r - k_b) G' + k_b B'$$

$$Pb = \frac{(B' - Y')}{2(1 - k_b)}$$

$$Pr = \frac{(R' - Y')}{2(1 - k_r)}$$

Εξίσωση 4.6

Οι σταθερές (k_r, k_b) εξαρτώνται από τα πρωταρχικά χρώματα R,G,B. Επιπλέον, οι τιμές των (Y', R', G', B') κυμαίνονται στο διάστημα [0,1] και οι αναλογικές τιμές των (Pb,Pr) ανήκουν στο διάστημα [- 0.5, 0.5].

Ο ψηφιακός χώρος YCbCr, με 8 bits/channel/pixel) προκύπτει από τον αναλογικό χώρο YPbPr με βάση τη παρακάτω εξίσωση:

$$(Y', Cb, Cr) = (16, 128, 128) + (219 \cdot Y', 224 \cdot Pb, 224 \cdot Pr)$$

Εξίσωση 4.7

Χρωματικός χώρος CIE LAB

Ο μη-γραμμικός χώρος CIE LAB, ή αλλιώς L*a*b*, χρησιμοποιείται για την καλύτερη και πιο ομοιόμορφη μοντελοποίηση της μη-γραμμικής φύσης της ανθρώπινης αντίληψης του

χρώματος. Έχει συνιστώσες (L^*, a^*, b^*) οι οποίες προκύπτουν από τον ακόλουθο μη-γραμμικό μετασχηματισμό του χώρου CIE XYZ:

$$\begin{aligned} L^* &= 116f\left(\frac{Y}{Y_n}\right) - 16 \\ a^* &= 500\left(f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right)\right) \\ b^* &= 200\left(f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right)\right), \end{aligned}$$

Εξίσωση 4.8

Στην εξίσωση αυτή, περιλαμβάνεται η συνάρτηση $f(t)$, όπου t οι όροι X/X_n , Y/Y_n , ή Z/Z_n . Η συνάρτηση $f(t)$ είναι η παρακάτω:

$$f(t) = \begin{cases} \sqrt[3]{t} & , \text{αν } t > \delta^3 \\ \frac{1}{3}t\delta^{-2} + \frac{4}{29} & , \text{αλλιού} \end{cases}$$

Εξίσωση 4.9

Οι όροι X, Y, Z περιγράφουν το χρωματικό ερέθισμα που εξετάζουμε, ενώ οι τιμές X_n, Y_n, Z_n περιγράφουν ένα συγκεκριμένο λευκό σημείο αναφοράς.

4.5 Νευρωνικά Δίκτυα

Τα νευρωνικά δίκτυα αποτελούν μια σημαντική και ταχέως αναπτυσσόμενη υποκατηγορία της Τεχνητής Νοημοσύνης, εμπνευσμένη από τον τρόπο λειτουργίας του ανθρώπινου εγκεφάλου. Το πρώτο βήμα για τη δημιουργία τους, έγινε το έτος 1943 όταν δύο επιστήμονες, οι Warren McCulloch και Walter Pitts, ανέπτυξαν τα πρώτα μοντέλα νευρωνικών δικτύων. Πολλά χρόνια αργότερα, μέχρι και σήμερα, έχει γίνει μεγάλη πρόοδος στο πεδίο αυτό – αρκετή ώστε να προσελκύσει μεγάλη προσοχή και να χρηματοδοτηθεί περαιτέρω έρευνα. Ένα νευρωνικό δίκτυο αποτελεί ένα σύστημα επεξεργασίας πληροφοριών που αποτελείται από έναν μεγάλο αριθμό στοιχείων επεξεργασίας τα οποία συνεργάζονται μεταξύ τους για την επίλυση συγκεκριμένων υπολογιστικών εργασιών. Παρόμοια με τον άνθρωπο, τα συστήματα αυτά επεξεργάζονται τις πληροφορίες και λύνουν προβλήματα, μαθαίνοντας από παραδείγματα.

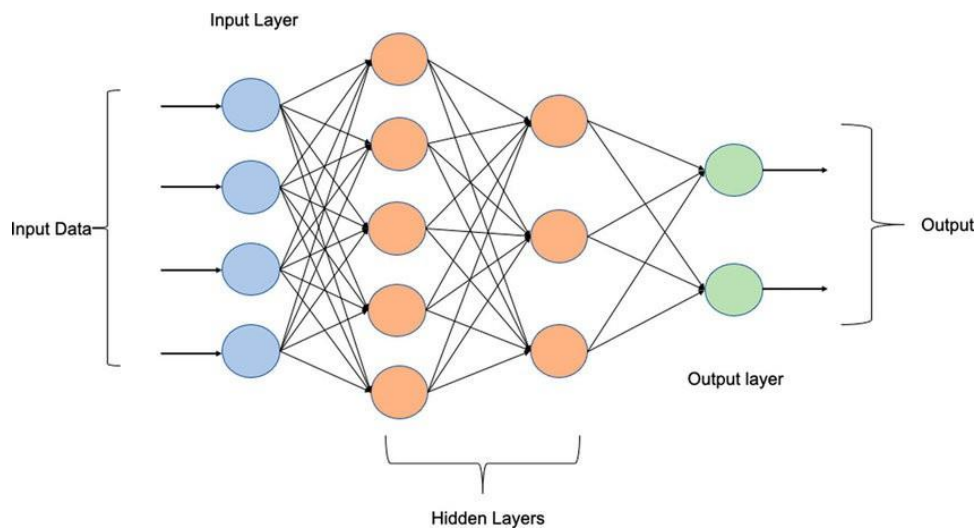
4.5.1 Δομή

Ένα νευρωνικό δίκτυο αποτελείται από πολλά επίπεδα (layers), καθένα από τα οποία περιλαμβάνουν κόμβους, ή αλλιώς νευρώνες, που τοποθετούνται ο ένας κάθετα πάνω από τον άλλον. Κάθε επίπεδο μπορεί να συνδεθεί με ένα άλλο επίπεδο, και με αυτόν τον τρόπο οι νευρώνες αλληλεπιδρούν μεταξύ τους. Όπως φαίνεται και από την εικόνα 4.6, υπάρχουν τρία είδη επιπέδων στα νευρωνικά δίκτυα:

- Το επίπεδο εισόδου (Input layer): Το πρώτο αυτό επίπεδο αποτελείται από νευρώνες που αντιπροσωπεύουν τα χαρακτηριστικά των δεδομένων εισόδου. Κάθε νευρώνας

αντιστοιχεί σε ένα χαρακτηριστικό και η τιμή του αντιπροσωπεύει την τιμή του χαρακτηριστικού.

- Τα ενδιάμεσα κρυμμένα επίπεδα (Hidden layers): Τα επίπεδα αυτά βρίσκονται μεταξύ της εισόδου και της εξόδου του δικτύου. Αυτά τα επίπεδα εκτελούν περίπλοκους υπολογισμούς χρησιμοποιώντας τα δεδομένα εισόδου και επιτρέπουν στο δίκτυο να μαθαίνει σύνθετα μοτίβα και αναπαραστάσεις στα δεδομένα. Κάθε νευρώνας σε ένα ενδιάμεσο επίπεδο δέχεται εισόδους από όλους τους νευρώνες από το προηγούμενο επίπεδο.
- Το επίπεδο εξόδου (Output layer): Το επίπεδο της εξόδου παράγει το τελικό αποτέλεσμα του δικτύου.



Εικόνα 4.6 Διάγραμμα επιπέδων νευρωνικών δικτύων. Οι κύκλοι αντιπροσωπεύουν τους νευρώνες. Πηγή: (24) MULTI-LAYER PERCEPTRON | LinkedIn

Κάθε νευρώνας σε ένα επίπεδο, δέχεται πολλαπλά σήματα ως εισόδους από τους νευρώνες του προηγούμενου επιπέδου. Αυτές οι εισοδοί πολλαπλασιάζονται με ένα βάρος, ύστερα τα σήματα αυτά αθροίζονται και τελικά περνούν μέσα από μια μη-γραμμική συνάρτηση η οποία δίνει το τελικό αποτέλεσμα, δηλαδή την τελική τιμή του νευρώνα. Ο νευρώνας, επομένως απαριθμεί 3 στάδια πριν την απόκτηση της τελικής του τιμής:

- Τις συνάψεις, που αντιπροσωπεύουν τις συνδέσεις μεταξύ των νευρώνων. Κάθε σύνδεση από έναν νευρώνα Α προς έναν νευρώνα Β περιλαμβάνει και ένα βάρος. Αυτά τα βάρη μαθαίνονται κατά τη διάρκεια της εκπαίδευσης του δικτύου και καθορίζουν την επίδραση της εξόδου του νευρώνα Α προς τον νευρώνα Β.
- Έναν αθροιστή, ο οποίος αθροίζει τις εξόδους όλων των συνάψεων. Σε αυτό το στάδιο, πολλές φορές χρησιμοποιείται και μια τιμή πόλωσης (bias term) η οποία προστίθεται στις εξόδους των συνάψεων, και αυτή είναι η τελική έξοδος του αθροιστή, η οποία διοχετεύεται στο επόμενο στάδιο.
- Τέλος, την συνάρτηση ενεργοποίησης η οποία καθορίζει την τελική τιμή ενεργοποίησης του νευρώνα. Η συνάρτηση αυτή εισάγει μη-γραμμικότητα, επιτρέποντας στο δίκτυο να μοντελοποιεί περίπλοκες, μη-γραμμικές σχέσεις μεταξύ των δεδομένων ενός προβλήματος. Είδη τέτοιων συναρτήσεων θα αναλυθούν στη συνέχεια.

Η συνάρτηση ενεργοποίησης

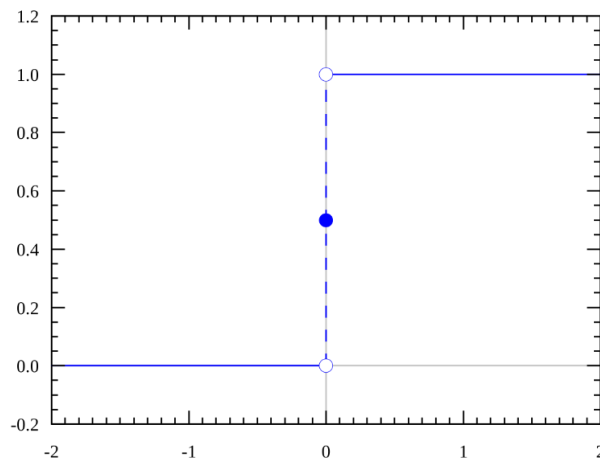
Ο σκοπός της συνάρτησης ενεργοποίησης είναι να εισάγει μη-γραμμικότητα στο δίκτυο. Χωρίς αυτή τη συνάρτηση, το δίκτυο μπορεί να μοντελοποιήσει μόνο γραμμικές σχέσεις, ενώ δεν μπορεί να μοντελοποιήσει τις μη-γραμμικές σχέσεις που υπάρχουν στα δεδομένα. Όμως, στα δεδομένα του πραγματικού κόσμου, οι περισσότερες σχέσεις είναι μη-γραμμικές. Επομένως, οι συναρτήσεις αυτές είναι απαραίτητες για την καλή λειτουργία των δικτύων. Επιπλέον, οι συναρτήσεις ενεργοποίησης είναι καθοριστικής σημασίας για την ικανότητα του δικτύου να μαθαίνει αποτελεσματικά και όσο γίνεται πιο γρήγορα.

➤ Βηματική συνάρτηση (Heavyside Step Function)

Αποτελεί την πιο απλή συνάρτηση ενεργοποίησης. Χρησιμοποιείται συχνά σε δυαδικά προβλήματα ταξινόμησης, όπου η έξοδος είναι 0 ή 1.

$$f(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Εξίσωση 4.10



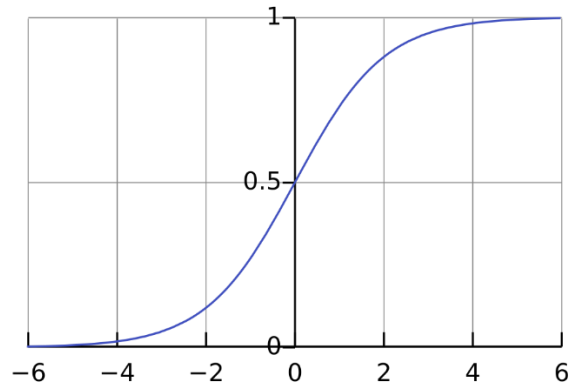
Εικόνα 4.7 Γραφική παράσταση της βηματικής συνάρτησης. Πηγή: https://en.wikipedia.org/wiki/Heaviside_step_function

➤ Σιγμοειδής συνάρτηση (Sigmoid Function)

Η συνάρτηση αυτή μετατρέπει τις εισόδους σε ένα εύρος μεταξύ 0 και 1. Επομένως, είναι χρήσιμη σε μοντέλα με τα οποία θέλουμε να προβλέψουμε την πιθανότητα ως αποτέλεσμα. Συμβολίζεται συχνά με sig(x) ή σ(x).

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

Εξίσωση 4.11



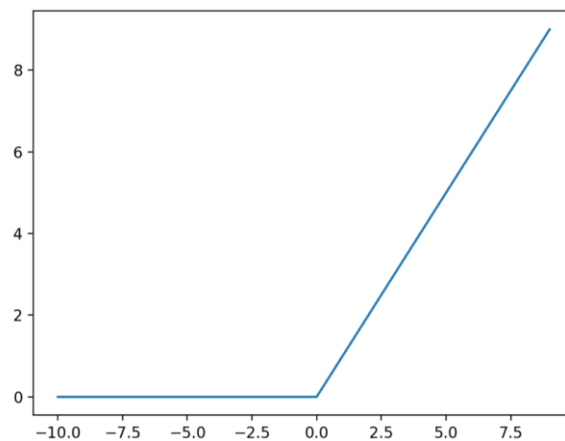
Εικόνα 4.8 Σιγμοειδής συνάρτηση. Πηγή: [Sigmoid function - Wikipedia](#)

➤ Συνάρτηση RELU (Rectified Linear Unit)

Αποτελεί την πιο διάσημη και συχνά χρησιμοποιούμενη μη-γραμμική συνάρτηση ενεργοποίησης στον κόσμο αυτή τη στιγμή, καθώς χρησιμοποιείται συχνά στα συνελκτικά νευρωνικά δίκτυα, μια κατηγορία νευρωνικών δικτύων που θα αναλυθεί στη συνέχεια του κεφαλαίου. Αποδείχθηκε, το 2011, πως είναι πιο αποτελεσματική στην εκπαίδευση βαθύτερων νευρωνικών δικτύων, συγκριτικά με άλλες συναρτήσεις ενεργοποίησης όπως η σιγμοειδής συνάρτηση [27]. Η συνάρτηση αυτή θέτει όλες τις αρνητικές τιμές ίσες με 0, στο μηδέν είναι ασυνεχής και θέτει τις θετικές τιμές ίσες με 1.

$$f(x) = \max(0, x) = \begin{cases} x, & \text{αν } x > 0 \\ 0, & \text{αλλιώς} \end{cases}$$

Εξίσωση 4.12



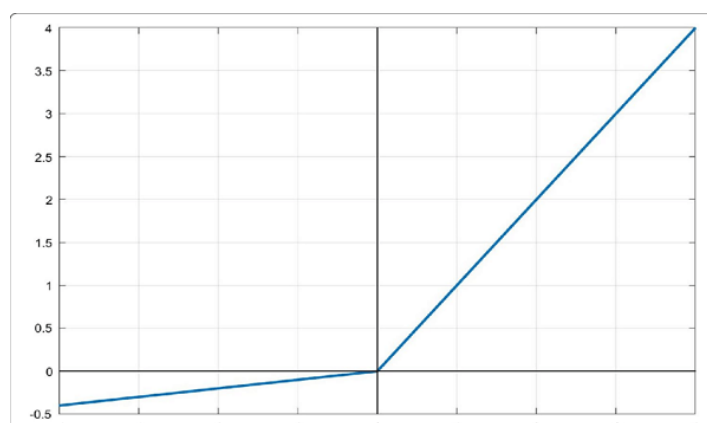
Εικόνα 4.9 Γραφική παράσταση της συνάρτησης ενεργοποίησης RELU. Πηγή: [A Gentle Introduction to the Rectified Linear Unit \(ReLU\) - MachineLearningMastery.com](#)

➤ Συνάρτηση Leaky RELU

Αποτελεί μια παραλλαγή της συνάρτησης ενεργοποίησης RELU που έχει σχεδιαστεί για να διορθώσει ένα από τα κύρια προβλήματα της, δηλαδή το γεγονός ότι η RELU μπορεί να κάνει μερικούς νευρώνες να «πεθαίνουν» κατά τη διάρκεια της εκπαίδευσης, όταν η έξοδος είναι μηδενική για όλες τις αρνητικές τιμές εισόδου. Λειτουργεί κανονικά όπως η RELU, αλλά προσθέτει μια μικρή κλίση στις αρνητικές τιμές εισόδου, αντί να τις ορίζει όλες σε 0. Αυτή η προσθήκη μπορεί να βοηθήσει στην εκπαίδευση των μοντέλων και να οδηγήσει σε καλύτερα αποτελέσματα των μοντέλων.

$$f(x) = \begin{cases} 0.01x, & x < 0 \\ x, & x \geq 0 \end{cases}$$

Εξίσωση 4.13



Εικόνα 4.10 Γραφική παράσταση συνάρτησης ενεργοποίησης Leaky RELU. Πηγή: [Leaky ReLU Activation Function \[with python code\] - Vidyasheela](#)

4.5.2 Συνελικτικά νευρωνικά δίκτυα

Τα συνελικτικά δίκτυα, ή αλλιώς Convolutional Neural Networks (CNNs), είναι μια ειδική κατηγορία νευρωνικών δικτύων και χρησιμοποιούνται συχνά για την αναγνώριση εικόνων και βίντεο, την ταξινόμηση εικόνων, την ανάλυση ιατρικών εικόνων, την επεξεργασία φυσικής γλώσσας. Ένα συνελικτικό νευρωνικό δίκτυο είναι ένας τύπος προωθητικού νευρωνικού δικτύου (feed-forward neural network) που μαθαίνει χαρακτηριστικά από μόνο του, μέσω της βελτιστοποίησης των φίλτρων [28], όπως θα αναλύσουμε στη συνέχεια. Με την έννοια προωθητικό δίκτυο, εννοούμε ότι η πληροφορία σε ένα τέτοιο δίκτυο, κατά το forward pass οδεύει μόνο προς τα «μπροστά», ξεκινώντας από την είσοδο του δικτύου, συνεχίζει μέσα από τα κρυμμένα επίπεδα και φτάνει στην έξοδο. Τέτοιου είδους δίκτυα, είναι συνήθως πλήρως συνδεδεμένα (fully connected), δηλαδή κάθε νευρώνας σε ένα επίπεδο ενώνεται με όλους τους άλλους νευρώνες στο επόμενο δίκτυο.

Τα συνελικτικά νευρωνικά δίκτυα αποτελούνται κυρίως από τριών ειδών επίπεδα: Από τα συνελικτικά επίπεδα (convolutional layers), τα επίπεδα «pooling» (pooling layers) και τα πλήρως συνδεδεμένα επίπεδα (fully-connected layers). Παρακάτω θα αναλυθούν τα τρία αυτά επίπεδα.

Συνελικτικά επίπεδα (Convolutional layers)

Για να γίνουν κατανοητά τα συνελκτικά επίπεδα σε ένα νευρωνικό δίκτυο, πρέπει πρώτα να εισάγουμε την έννοια της συνέλιξης. Η συνέλιξη, είναι η διαδικασία με την οποία εφαρμόζουμε ένα φίλτρο σε μια εικόνα. Μαθηματικά, η συνέλιξη ορίζεται ως:

$$(f * g)(x) = \int_{-\infty}^{+\infty} f(k) \cdot g(x-k) dk, \quad x \in \mathbb{R}$$

Εξίσωση 4.14

Αυτός ο ορισμός, όμως, αναφέρεται σε συνεχείς συναρτήσεις. Στον τομέα της όρασης υπολογιστών και των νευρωνικών δικτύων, όπως έχουμε αναφέρει και στο προηγούμενο κεφάλαιο, τα δεδομένα μας είναι σε διακριτή (ψηφιακή) μορφή. Επομένως, ο ορισμός της διακριτής συνέλιξης είναι ο εξής:

$$(f * g)(x) = \sum_{k=-\infty}^{+\infty} f(k) \cdot g(x-k), \quad \forall x \in \mathbb{Z}$$

Εξίσωση 4.15

Στον τομέα των νευρωνικών δικτύων, η είσοδος f συνήθως είναι μια εικόνα, και η συνάρτηση g είναι ένας πυρήνας (kernel) ενός φίλτρου. Ο πυρήνας είναι ένας μικρός πίνακας (matrix) που χρησιμοποιείται για την εφαρμογή φίλτρων σε μια εικόνα, μέσω της συνέλιξης του με την εικόνα, όπως ορίσαμε παραπάνω. Στα νευρωνικά δίκτυα, συνήθως η συνέλιξη εφαρμόζεται σε πίνακες πολλών διαστάσεων, αφού έχουμε αναλύσει ότι και οι εικόνες είναι δισδιάστατοι πίνακες $M \times N$. Άρα, αν θεωρήσουμε μια εικόνα I ως είσοδο στο νευρωνικό δίκτυο, μπορούμε να ορίσουμε την συνέλιξη σε 2 διαστάσεις, χρησιμοποιώντας και δισδιάστατο πυρήνα K :

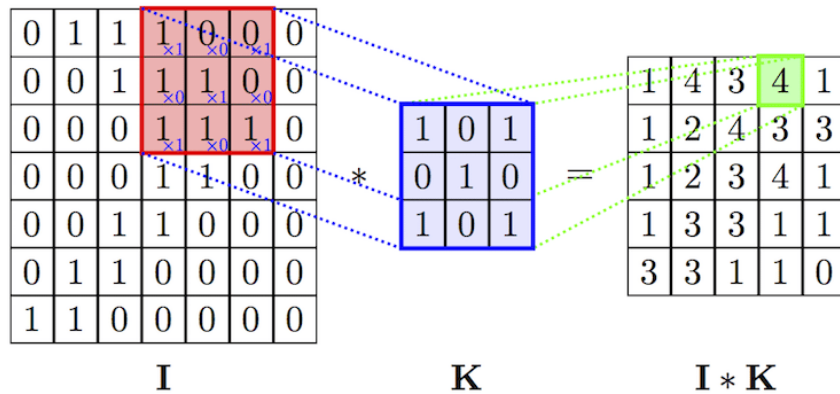
$$S(i,j) = (I * K)(i,j) = \sum_m \sum_n I(m,n) K(i-m, j-n)$$

Εξίσωση 4.16

Τα συνελκτικά επίπεδα χρησιμοποιούνται για την αναγνώριση και την εξαγωγή διαφόρων χαρακτηριστικών από τα δεδομένα εισόδου. Τα επίπεδα αυτά, εκτελούν συνέλιξη ενός πυρήνα με μια εικόνα. Πιο συγκεκριμένα, το κέντρο του πυρήνα τοποθετείται πάνω από καθένα pixel της εικόνας, όσα γειτονικά pixels καλύπτονται από τον πυρήνα πολλαπλασιάζονται με τον αριθμό του πυρήνα που τα καλύπτει, και έπειτα υπολογίζεται το άθροισμα των τελικών στοιχείων. Η τελική τιμή του pixel μετά το φιλτράρισμα, είναι αυτό το συνολικό άθροισμα. Η διαδικασία μιας δισδιάστατης συνέλιξης παρουσιάζεται στην εικόνα 4.11. Αν ο πυρήνας εφαρμοστεί στα άκρα της εικόνας, συχνά χρησιμοποιείται η μέθοδος «padding» με την οποία αυξάνεται το μέγεθος της εικόνας και έτσι ο πυρήνας δεν «πέφτει έξω» από τα προηγούμενα όρια της εικόνας. Το βήμα με το οποίο γίνεται η μετατόπιση του πυρήνα πάνω στην εικόνα ονομάζεται διασκελισμός (stride), δηλαδή αν ο διασκελισμός είναι 1 ο πυρήνας θα περάσει απ' όλα τα pixels της εικόνας, ενώ αν είναι 2 θα περάσει από τα μισά [29].

Από τον υπολογισμό της συνέλιξης της εικόνας με έναν πυρήνα, παίρνουμε ως αποτέλεσμα μια διαφορετική εικόνα, που στα νευρωνικά δίκτυα την ονομάζουμε χάρτη χαρακτηριστικών (feature map). Αυτοί οι χάρτες είναι πολύ σημαντικά στοιχεία ενός νευρωνικού δικτύου, καθώς είναι αρμόδιοι για την αναγνώριση διαφορετικών οπτικών χαρακτηριστικών και μοτίβων στα δεδομένα εισόδου. Σε ένα συνελκτικό επίπεδο, συνήθως εφαρμόζονται πάνω από ένα φίλτρα παράλληλα, καθένα από τα οποία ανιχνεύει διαφορετικά είδη χαρακτηριστικών από τις εικόνες, όπως γωνίες, υφές, ή ακόμα και σχήματα και αντικείμενα. Για την ακρίβεια, όσο προχωρούν τα επίπεδα ενός συνελκτικού νευρωνικού δικτύου, οι χάρτες χαρακτηριστικών γίνονται όλο και πιο περίπλοκοι. Στα αρχικά συνελκτικά επίπεδα ανιχνεύονται πιο απλά χαρακτηριστικά

όπως οι γωνίες, ενώ σε επόμενα, υψηλότερα επίπεδα ανιχνεύονται πιο σύνθετα μοτίβα, όπως για παράδειγμα και αντικείμενα.



Εικόνα 4.11 Αναπαράσταση δισδιάστατης συνέλιξης. Πηγή: https://www.researchgate.net/figure/An-example-of-convolution-operation-in-2D-2_fig3_324165524

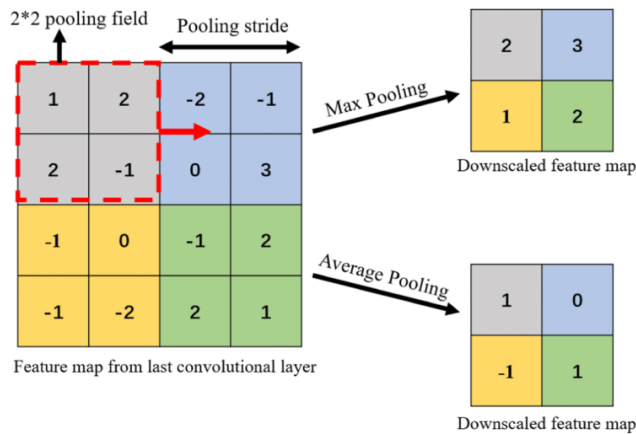
Pooling Layers

Τα επίπεδα αυτά είναι βασικά στοιχεία των συνελκτικών νευρωνικών δικτύων και χρησιμοποιούνται για τη μείωση των διαστάσεων των δεδομένων, διατηρώντας τα σημαντικά χαρακτηριστικά. Η μείωση των διαστάσεων βοηθά στη μείωση του αριθμού των παραμέτρων του δικτύου, και ως αποτέλεσμα και των υπολογιστικών απαιτήσεων. Επιπλέον, βοηθούν στην αποφυγή του φαινομένου της υπερπροσαρμογής στα δεδομένα (overfitting), το οποίο θα αναλυθεί σε επόμενο κεφάλαιο. Αυτά τα επίπεδα ακολουθούν τα συνελκτικά επίπεδα, και δρουν στους χάρτες χαρακτηριστικών. Η διαδικασία ενός επιπέδου pooling περιλαμβάνει την επιλογή μιας λειτουργίας pooling (pooling operation), το μέγεθος της οποίας είναι μικρότερο από το μέγεθος του χάρτη χαρακτηριστικών. Συγκεκριμένα, σχεδόν πάντα το μέγεθος της λειτουργίας αυτής είναι 2x2 pixels, και εφαρμόζεται με διασκελισμό (stride) 2. Αυτό σημαίνει ότι μετά το επίπεδο αυτό, κάθε χάρτης χαρακτηριστικών θα έχει μειωθεί στο μισό σε μέγεθος, μειώνοντας και τον αριθμό των pixels στο ένα τέταρτο του αρχικού.

Δύο κοινές συναρτήσεις που χρησιμοποιούνται στη διαδικασία pooling είναι:

- Average pooling: Μέθοδος η οποία υπολογίζει την μέση τιμή για κάθε αντίστοιχο υποπίνακα 2x2 στον χάρτη χαρακτηριστικών, και κάθε μια από αυτές τις τιμές συνθέτουν τον νέο πίνακα.
- Max pooling: Υπολογίζεται η μέγιστη τιμή για κάθε υποπίνακα 2x2 του χάρτη χαρακτηριστικών, και κάθε μια από αυτές τις τιμές συνθέτουν τον νέο πίνακα.

Οι δύο αυτές μέθοδοι pooling αναπαρίστανται στην παρακάτω εικόνα (4.12):



Εικόνα 4.12 Αναπαράσταση δύο ειδών pooling: Max pooling και Average pooling. Πηγή: https://www.researchgate.net/figure/Examples-for-max-and-average-pooling-layer_fig5_356816561

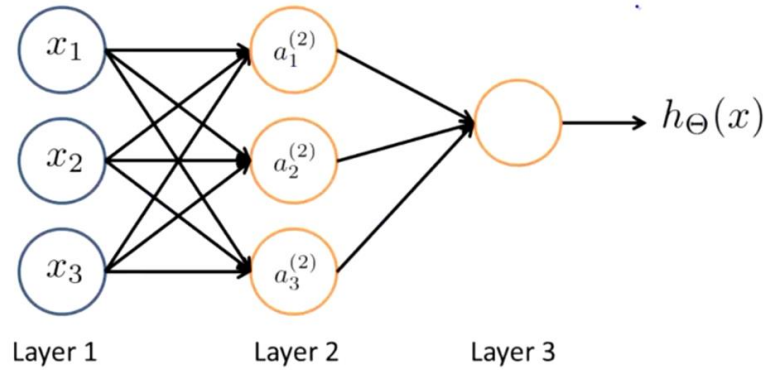
Πλήρως συνδεδεμένα επίπεδα (Fully-connected layers)

Τα πλήρως συνδεδεμένα επίπεδα είναι ένα σημαντικό στοιχείο στα συνελκτικά νευρωνικά δίκτυα και χρησιμοποιούνται για να συνδυάσουν και να εξάγουν τελικά τα χαρακτηριστικά που έχουν ανιχνευθεί από τα προηγούμενα συνελκτικά και pooling επίπεδα. Σε ένα πλήρως συνδεδεμένο επίπεδο, όπως έχουμε ξανά αναφέρει παραπάνω, κάθε νευρώνας είναι συνδεδεμένος με κάθε νευρώνα του προηγούμενου επιπέδου. Κάθε νευρώνας εφαρμόζει έναν γραμμικό μετασχηματισμό στο διάνυσμα εισόδου x μέσω ενός πίνακα βαρών W . Στο άθροισμα αυτό προστίθεται και μια τιμή πόλωσης b . Στη συνέχεια, εφαρμόζεται μια μη-γραμμική συνάρτηση ενεργοποίησης f στον μετασχηματισμό αυτό. Έτσι, παράγεται η τιμή του νευρώνα, η οποία θα διοχετευθεί στη συνέχεια στο επόμενο επίπεδο νευρώνων και θα επαναληφθεί η ίδια διαδικασία, μέχρι και την έξοδο του δικτύου. Θα αναλύσουμε αυτήν την διαδικασία με περισσότερη λεπτομέρεια, στη συνέχεια.

4.5.3 Προώθηση δεδομένων

Προώθηση (forward pass)

Η προώθηση (forward pass) είναι η διαδικασία κατά την οποία τα δεδομένα διέρχονται μέσα από το συνελκτικό νευρωνικό δίκτυο, κατευθυνόμενα από την είσοδο προς την έξοδο. Παρακάτω, θα αναλύσουμε αυτή τη διαδικασία, παίρνοντας ως παράδειγμα ένα απλό δίκτυο, με 3 νευρώνες εισόδου x_1, x_2, x_3 , ένα ενδιάμεσο επίπεδο (hidden layer) με 3 νευρώνες, και μια έξοδο h (Εικόνα 4.13).



Εικόνα 4.13 Αναπαράσταση ενός απλού πλήρως συνδεδεμένου δικτύου.

Οι νευρώνες εισόδου, περιέχουν τα δεδομένα εισόδου τα οποία μπορεί να είναι από μια τιμή μέχρι και πολυδιάστατοι πίνακες. Έστω W^l ο πίνακας των βαρών ενός επιπέδου l (από την αγγλική λέξη layer). Μεταξύ επιπέδου l και επιπέδου $l+1$, και μεταξύ επιπέδου $l+1$ και επιπέδου $l+2$, έχουμε τους πίνακες βαρών W^l , W^{l+1} αντίστοιχα. Οι πίνακες αυτοί περιέχουν τα βάρη για κάθε σύναψη από έναν νευρώνα προς έναν άλλον. Όπως έχουμε αναλύσει και παραπάνω, για να βρούμε τη τιμή του κάθε νευρώνα, πολλαπλασιάζουμε κάθε είσοδο που φτάνει στον νευρώνα από το προηγούμενο επίπεδο, με την τιμή του βάρους της αντίστοιχης σύναψης, αθροίζουμε αυτά τα γινόμενα, προσθέτουμε μια τιμή πόλωσης b^l (bias) και τέλος, περνάμε το αποτέλεσμα από την μη-γραμμική συνάρτηση ενεργοποίησης, έστω f . Στη συγκεκριμένη περίπτωση, έχουμε μια τιμή πόλωσης b^1 για το 2^ο επίπεδο του δικτύου. Επομένως, για το 2^ο επίπεδο η τιμή ενός νευρώνα, πριν την εφαρμογή της συνάρτησης ενεργοποίησης θα είναι:

$$z^{(2)} = W^1 x + b^1$$

Εξίσωση 4.17

Έπειτα, με την εφαρμογή της συνάρτησης ενεργοποίησης f οι νευρώνες στο 2^ο επίπεδο θα λάβουν τις τελικές τιμές:

$$a^{(2)} = f(z^{(2)})$$

Εξίσωση 4.18

Τέλος, στην έξοδο, στη συγκεκριμένη περίπτωση στον τελευταίο μοναδικό νευρώνα, πραγματοποιείται ο πολλαπλασιασμός βαρών κάθε σύναψης με τις αντίστοιχες εισόδους από τους νευρώνες του προηγούμενου επιπέδου:

$$h = W^2 a^{(2)}$$

Εξίσωση 4.19

Το τελευταίο βήμα στη διαδικασία της διάδοσης προς τα εμπρός (forward pass) είναι η αξιολόγηση του αποτελέσματος της εξόδου, δηλαδή της πρόβλεψης του δικτύου με την πραγματική τιμή (ground truth). Η αξιολόγηση αυτή γίνεται μέσω της συνάρτησης κόστους (Loss function).

Συνάρτηση κόστους (Loss function)

Η αξιολόγηση του δικτύου, για το πόσο καλά μπορεί να μοντελοποιήσει το σύνολο δεδομένων εισόδου, γίνεται με τη συνάρτηση κόστους. Η συνάρτηση κόστους συγκρίνει τις τιμές που προβλέπει το δίκτυο με τις αναμενόμενες (ground truth). Αν οι προβλέψεις του δικτύου στην

έξοδο του απέχουν πολύ από τις αναμενόμενες τότε η συνάρτηση κόστους παράγει μεγάλες τιμές στην έξοδο της, ενώ αν οι προβλέψεις συνάδουν με τις αναμενόμενες, η συνάρτηση κόστους παράγει μικρές τιμές. Βασικές συναρτήσεις κόστους παρατίθενται παρακάτω, με τα σύμβολα \hat{y}_i, y_i να συμβολίζουν την πρόβλεψη του δικτύου και την αναμενόμενη τιμή (ground truth), αντίστοιχα.

➤ Μέσο τετραγωνικό σφάλμα (Mean Squared Error - MSE)

Μια από τις πιο δημοφιλείς συναρτήσεις κόστους. Υπολογίζεται ως η μέση τετραγωνική διαφορά μεταξύ της πρόβλεψης του δικτύου και της πραγματικής αναμενόμενης τιμής (ground truth) σε όλο το σύνολο δεδομένων.

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

Εξίσωση 4.20

➤ Μέσο απόλυτο σφάλμα (Mean Absolute Error – MAE)

Με αυτή τη συνάρτηση κόστους, υπολογίζεται η μέση απόλυτη διαφορά μεταξύ της πρόβλεψης του δικτύου και της πραγματικής τιμής σε όλο το σύνολο δεδομένων.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

Εξίσωση 4.21

➤ Cross-Entropy Loss

Χρησιμοποιείται σε προβλήματα ταξινόμησης, όπου οι προβλέψεις είναι πιθανότητες, δηλαδή η έξοδος του νευρωνικού δικτύου πρέπει να είναι σε ένα εύρος μεταξύ 0 και 1.

$$L(\theta) = - \sum_{i=0}^n \hat{y}_i \cdot \log(y_i)$$

Εξίσωση 4.22

➤ Dice Loss

Η συνάρτηση κόστους Dice Loss προέρχεται από τον συντελεστή Dice score coefficient (DSC), ο οποίος είναι ένα μέτρο ομοιότητας μεταξύ δύο συνόλων. Χρησιμοποιείται συνήθως σε προβλήματα τμηματοποίησης εικόνας (image segmentation) για την αξιολόγηση της επικάλυψης μεταξύ μιας μάσκας τμηματοποίησης που έχει προβλέψει το δίκτυο και της επιθυμητής μάσκας τμηματοποίησης. Υπολογίζεται ξεχωριστά για κάθε κατηγορία (class) και αναφέρεται ως τελικό αποτέλεσμα ο μέσος όρος όλων των κατηγοριών.

$$Dice\ Coefficient = \frac{2|Y \cap T|}{|Y| + |T|},$$

Εξίσωση 4.23

όπου Y η πρόβλεψη του δικτύου για τη μάσκα τμηματοποίησης και T η πραγματική μάσκα για μια κατηγορία. Η συνάρτηση κόστους Dice προτάθηκε από τους Milletari et al. [30] και μπορεί να οριστεί ως:

$$D = \frac{2 \sum_i^N p_i g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2},$$

Εξίσωση 4.24

όπου p_i είναι η πρόβλεψη της μάσκας του δικτύου και g_i η πραγματική μάσκα (ground truth label).

Η συνάρτηση κόστους αυτή είναι ιδιαίτερα χρήσιμη σε προβλήματα τμηματοποίησης εικόνας όπου τα αντικείμενα ενδιαφέροντος σε μια εικόνα είναι μικρότερα ή σπανιότερα σε σχέση με τα υπόλοιπα στοιχεία μιας εικόνας [31].

4.5.4 Οπισθοδιάδοση

Η μέθοδος οπισθοδιάδοσης (backpropagation) αποτελεί πολύ σημαντικό παράγοντα για την εκπαίδευση των συνελκτικών νευρωνικών δικτύων. Έπειτα από κάθε προώθηση (forward pass) των δεδομένων, η μέθοδος οπισθοδιάδοσης πραγματοποιεί μια διάδοση προς τα πίσω, αλλάζοντας παράλληλα τα βάρη (weights) και τις πολώσεις (biases) του κάθε νευρώνα, με βάση το σφάλμα, δηλαδή την διαφορά μεταξύ πρόβλεψης του δικτύου στην έξοδο και της πραγματικής τιμής. Πιο συγκεκριμένα, στην οπισθοδιάδοση υπολογίζεται η κλίση (gradient) της συνάρτησης κόστους (loss function) σε σχέση με το κάθε βάρος. Αυτός ο υπολογισμός βοηθά το δίκτυο να τροποποιήσει και να ενημερώσει τα βάρη ώστε να ελαχιστοποιηθεί το σφάλμα πρόβλεψης του δικτύου. Θα χρησιμοποιήσουμε το παράδειγμα του απλού νευρωνικού δικτύου της εικόνας 4.13 για να εξηγήσουμε τη διαδικασία της οπισθοδιάδοσης, υποθέτοντας συνάρτηση κόστους C.

Αρχικά, η κλίση (gradient) μιας συνάρτησης C ως προς x, είναι ένα διάνυσμα των μερικών παραγώγων της C ως προς x. Η κλίση, ουσιαστικά, δείχνει πόσο πρέπει να αυξηθεί ή να μειωθεί η μεταβλητή x για να ελαχιστοποιηθεί η συνάρτηση C. Ορίζεται ως:

$$\frac{\partial C}{\partial x} = \left[\frac{\partial C}{\partial x_1}, \frac{\partial C}{\partial x_2}, \dots, \frac{\partial C}{\partial x_m} \right]$$

Εξίσωση 4.25

Ο υπολογισμός αυτών των παραγώγων γίνεται με τον κανόνα της αλυσίδας (chain rule):

$$\frac{\partial C}{\partial w_{jk}^l} = \frac{\partial C}{\partial z_j^l} \cdot \frac{\partial z_j^l}{\partial w_{jk}^l} \quad (\text{chain rule}),$$

Εξίσωση 4.26

όπου w_{jk}^l τα βάρη των συνάψεων μεταξύ των νευρώνων από ένα επίπεδο στο επόμενο, b_j^l οι τιμές πόλωσης των επιπέδων, z_j^l οι τιμές των νευρώνων πριν την εφαρμογή της συνάρτησης ενεργοποίησης, και a_k^{l-1} οι είσοδοι σε έναν νευρώνα από όλους τους προηγούμενους νευρώνες του προηγούμενου επιπέδου. Οι τιμές των νευρώνων z_j^l ορίζονται στην εξίσωση 4.27, όπως τους ορίσαμε και στο προηγούμενο κεφάλαιο. Ο αριθμός των νευρώνων στο επίπεδο (l-1) είναι m.

$$z_j^l = \sum_{k=1}^m w_{jk}^l a_k^{l-1} + b_j^l$$

Εξίσωση 4.27

Αν παραγωγίσουμε ως προς το βάρος w_{jk}^l παίρνουμε:

$$\frac{\partial z_j^l}{\partial w_{jk}^l} = a_k^{l-1}$$

Εξίσωση 4.28

Επομένως, αντικαθιστώντας στην εξίσωση 4.26, έχουμε την τελική τιμή:

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

Εξίσωση 4.29

Όμοια, μπορούμε να εφαρμόσουμε τις ίδιες εξισώσεις για τις τιμές πόλωσης b_j^l :

$$\frac{\partial C}{\partial b_j^l} = \frac{\partial C}{\partial z_j^l} \frac{\partial z_j^l}{\partial b_j^l} \quad (\text{chain rule})$$

Εξίσωση 4.30

Παραγωγίζοντας την τιμή του νευρώνα z_j^l ως προς την τιμή πόλωσης b_j^l παίρνουμε:

$$\frac{\partial z_j^l}{\partial b_j^l} = 1$$

Εξίσωση 4.31

Τελικά, αντικαθιστώντας στην εξίσωση 4.30 παίρνουμε τη τελική τιμή:

$$\frac{\partial C}{\partial b_j^l} = \frac{\partial C}{\partial z_j^l} 1$$

Εξίσωση 4.32

Αξίζει να σημειωθεί, ότι το κοινό κομμάτι των εξισώσεων 4.29 και 4.32 ονομάζεται local gradient και συμβολίζεται ως:

$$\delta_j^l = \frac{\partial C}{\partial z_j^l}$$

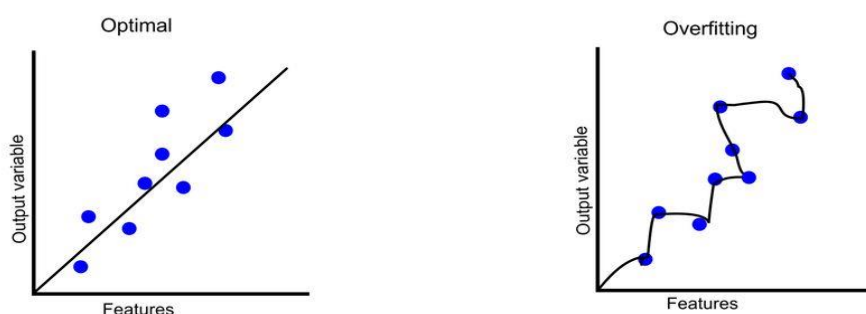
Εξίσωση 4.33

Συνοψίζοντας, κατά τη διαδικασία της οπισθοδιάδοσης το σφάλμα που υπολογίζεται από τη συνάρτηση κόστους, διαδίδεται στο δίκτυο προς τα πίσω, ξεκινώντας από το τέλος του δικτύου προς την αρχή, υπολογίζοντας την κλίση της συνάρτησης κόστους σε σχέση με κάθε βάρος. Αυτός ο υπολογισμός καθορίζει πως θα τροποποιηθούν τα βάρη και οι πολώσεις του δικτύου για να ελαχιστοποιηθεί, τελικά, το σφάλμα. Τέλος, η ενημέρωση των βαρών και πολώσεων γίνεται με βάση τον ρυθμό μάθησης (learning rate), μια παράμετρο που διαμορφώνει πόσο μεγάλη ή μικρή θα είναι η προσαρμογή των βαρών σε κάθε επανάληψη της διαδικασίας αυτής.

Αυτή η διαδικασία ακολουθεί έναν αλγόριθμο βελτιστοποίησης, όπως οι SGD (Stochastic gradient descent), Adam (Adaptive Moment Estimation), RMSprop (Root Mean Square Propagation).

4.5.5 Τεχνικές κανονικοποίησης

Ένα κοινό πρόβλημα στη μηχανική μάθηση, είναι η υπερπροσαρμογή στα δεδομένα εκπαίδευσης, ή αλλιώς overfitting. Συμβαίνει όταν ένα μοντέλο αποδίδει καλά στα δεδομένα εκπαίδευσης, αλλά δεν κάνει σωστές προβλέψεις για καινούρια δεδομένα του ίδιου προβλήματος. Αυτό σημαίνει ότι το μοντέλο έχει απομνημονεύσει τα συγκεκριμένα μοτίβα και τον θόρυβο στα δεδομένα εκπαίδευσης, αλλά δεν μπορεί να προσαρμοστεί σε νέες καταστάσεις ή παραλλαγές (εικόνα 4.14). Αυτό μπορεί να συμβεί σε περιπτώσεις όπως όταν έχουμε μικρό σύνολο δεδομένων εισόδου ή όταν το μοντέλο είναι υπερβολικά περίπλοκο για ένα σύνολο δεδομένων. Η υπερπροσαρμογή αυτή μειώνει την ακρίβεια και την αξιοπιστία του μοντέλου και μπορεί να οδηγήσει σε κακές αποφάσεις και αποτελέσματα. Για να αποφύγουμε την υπερπροσαρμογή, πρέπει να χρησιμοποιήσουμε τεχνικές κανονικοποίησης που περιορίζουν την πολυπλοκότητα του μοντέλου και εισάγουν κάποια τυχαιότητα στην διαδικασία μάθησης.



Εικόνα 4.14 Γραφική αναπαράσταση αποτελέσματος μάθησης ενός μοντέλου. Στην πρώτη εικόνα το μοντέλο προσεγγίζει τις πραγματικές λύσεις, ενώ στην δεύτερη το μοντέλο έχει προσαρμοστεί υπερβολικά στα δεδομένα εκπαίδευσης, χωρίς να μπορεί να γενικεύσει τη γνώση σε νέα δεδομένα. Πηγή: [What is Overfitting in Machine Learning? \(freecodecamp.org\)](https://www.freecodecamp.org/learn/machine-learning-in-depth/what-is-overfitting-in-machine-learning/)

Τεχνική κανονικοποίησης L1 και L2

Η L1 και η L2 είναι οι πιο κοινές τεχνικές κανονικοποίησης. Στην συνάρτηση κόστους προστίθεται ένας ακόμα όρος γνωστός ως όρος κανονικοποίησης (regularization term). Λόγω της προσθήκης αυτού του όρου κανονικοποίησης, οι τιμές των πινάκων των βαρών του μοντέλου μειώνονται και έτσι το μοντέλο απλοποιείται. Επομένως, μειώνεται η υπερπροσαρμογή σε μεγάλο βαθμό.

Γενικά, οι νόρμες L_p ($p \geq 1$) αναγκάζουν τις νόρμες των πινάκων των βαρών να είναι μικρές. Η νόρμα L_p ενός διανύσματος x σε n -διαστάσεις ορίζεται ως:

$$L_p(x) = \|x\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}}$$

Εξίσωση 4.34

Η νόρμα L1: Όταν $p=1$, τότε έχουμε το άθροισμα των απόλυτων τιμών του διανύσματος x .

$$L_1(x) = \|x\|_1 = \sum_{i=1}^n |x_i|$$

Εξίσωση 4.35

Η νόρμα L2: Όταν $p=2$, τότε υπολογίζουμε την Ευκλείδεια απόσταση του σημείου από την αρχή στον n -διάστατο διανυσματικό χώρο.

$$L_2(x) = \|x\|_2 = \left(\sum_{i=1}^n |x_i|^2 \right)^{\frac{1}{2}}$$

Εξίσωση 4.36

Αφού ορίσαμε τις νόρμες, έστω L και \tilde{L} οι συναρτήσεις κόστους χωρίς και με κανονικοποίηση, αντίστοιχα. Για την τεχνική κανονικοποίησης L1 παίρνουμε:

$$\tilde{L}(w) = L(w) + a\|w\|_1,$$

Εξίσωση 4.37

όπου a είναι μια σταθερά κανονικοποίησης. Η μέθοδος αυτή αναγκάζει τα βάρη να πάνε προς το μηδέν, αποκλείοντας έτσι ορισμένα χαρακτηριστικά από το μοντέλο. Αυτό το φαινόμενο είναι ιδιαίτερα χρήσιμο όταν έχουμε δεδομένα πολλών διαστάσεων, όπου κάποια χαρακτηριστικά μπορεί να είναι περιττά [32].

Όμοια, η τεχνική κανονικοποίησης L2 ορίζεται ως:

$$\tilde{L}(w) = L(w) + \frac{a}{2} \|w\|_2^2$$

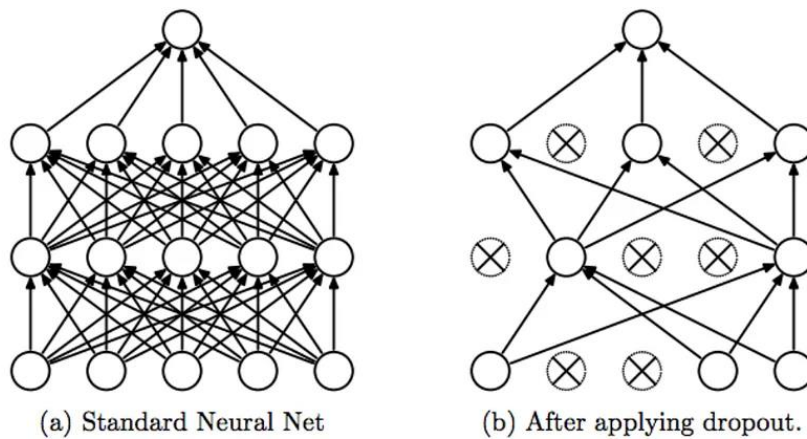
Εξίσωση 4.38

Σε αντίθεση με την τεχνική κανονικοποίησης L1 που αναγκάζει τα βάρη να μηδενίζονται, η κανονικοποίηση L2 συρρικνώνει τα βάρη, δεν τα μηδενίζει, ενώ διασφαλίζει ότι τα σημαντικά στοιχεία του διανύσματος βαρών είναι μεγαλύτερα από τα υπόλοιπα.

Dropout

Αυτή είναι μία από τις πιο ενδιαφέρουσες τεχνικές για μείωση της υπερπροσαρμογής στα δεδομένα εκπαίδευσης. Παράγει επίσης πολύ καλά αποτελέσματα και, ως εκ τούτου, είναι η πιο συχνά χρησιμοποιούμενη τεχνική κανονικοποίησης στον τομέα της βαθιάς μάθησης.

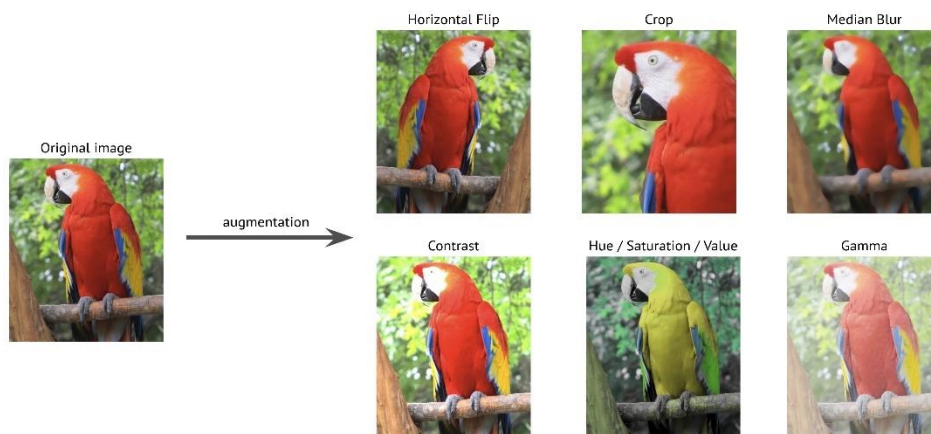
Με τη χρήση αυτής της τεχνικής, σε κάθε επανάληψη, το δίκτυο επιλέγει τυχαία ορισμένους νευρώνες και τους αφαιρεί μαζί με όλες τις εισερχόμενες και εξερχόμενες συνδέσεις τους, όπως φαίνεται στην εικόνα 4.15. Οπότε, σε κάθε επανάληψη χρησιμοποιούνται διαφορετικοί νευρώνες και αυτό έχει ως αποτέλεσμα κάθε φορά να υπάρχουν διαφορετικές έξοδοι. Έτσι, εισάγεται τυχαιότητα στο δίκτυο. Η τεχνική αυτή συνήθως προτιμάται σε μεγαλύτερα, πιο βαθιά δίκτυα. Με τη τεχνική αυτή διπλασιάζεται ο αριθμός των επαναλήψεων που απαιτούνται για σύγκλιση του δικτύου. Ωστόσο, ο χρόνος εκπαίδευσης για κάθε εποχή γίνεται μικρότερος.



Εικόνα 4.15 Αναπαράσταση νευρωνικού δικτύου χωρίς χρήση της τεχνικής dropout (πρώτη εικόνα) και με dropout (δεύτερη εικόνα). Πηγή: <https://medium.com/@amarbudhiraja/https-medium-com-amarbudhiraja-learning-less-to-learn-better-dropout-in-deep-machine-learning-74334da4bfc5>

Επαύξηση δεδομένων (Data augmentation)

Η πιο απλή και συχνά χρησιμοποιούμενη μέθοδος για την μείωση της υπερπροσαρμογής των νευρωνικών δικτύων στα δεδομένα, είναι η δημιουργία τροποποιημένων εκδοχών των υπαρχόντων δεδομένων εκπαίδευσης. Συχνά, έχουμε περιορισμένο αριθμό δεδομένων για την εκπαίδευση ενός δικτύου, με αποτέλεσμα το δίκτυο να απομνημονεύει τις σχέσεις μεταξύ τους και να μη μπορεί να γενικεύσει καλά σε καινούρια δεδομένα. Η απόκτηση περισσότερων δεδομένων, για το ίδιο πρόβλημα, δεν είναι πάντα εύκολη. Έτσι, μπορούμε να δημιουργήσουμε τυχαίες εκδοχές των δεδομένων που μας είναι διαθέσιμα, για να παρέχουμε στο δίκτυο δεδομένα με μεγαλύτερη ποικιλομορφία ώστε αυτό, έχοντας δει αρκετές παραλλαγές εικόνων ενός προβλήματος, να είναι ικανό να γενικεύσει τη γνώση σε άγνωστα δεδομένα. Τρόποι τροποποίησης των εικόνων περιλαμβάνουν αντιστροφή, αλλαγή χρωμάτων και αντίθεσης, περιστροφή, κλιμάκωση των διαστάσεων της εικόνας. Παραδείγματα τέτοιων τρόπων επαύξησης δεδομένων φαίνονται στην εικόνα 4.16.



Εικόνα 4.16 Είδη δημιουργίας παραλλαγών μιας εικόνας εισόδου για ένα δίκτυο, με αποτέλεσμα την αύξηση των δεδομένων αλλά και την ποικιλία τους, για την μείωση του φαινομένου overfitting. Πηγή: [What is image augmentation - Albumentations Documentation](#)

4.6 Βαθιά Μάθηση

Η βαθιά μάθηση (deep learning) είναι εμπνευσμένη από τη λειτουργία του εγκεφάλου και η γέννηση αυτού του κλάδου υπολογίζεται γύρω στο 1950 [33]. Η βαθιά μάθηση αναφέρεται σε μια υποκατηγορία της μηχανικής μάθησης, όπως φαίνεται από την εικόνα 4.17, και πιο συγκεκριμένα στη χρήση νευρωνικών δικτύων τα οποία διαθέτουν τουλάχιστον δύο ή τρία κρυμμένα επίπεδα (hidden layers).

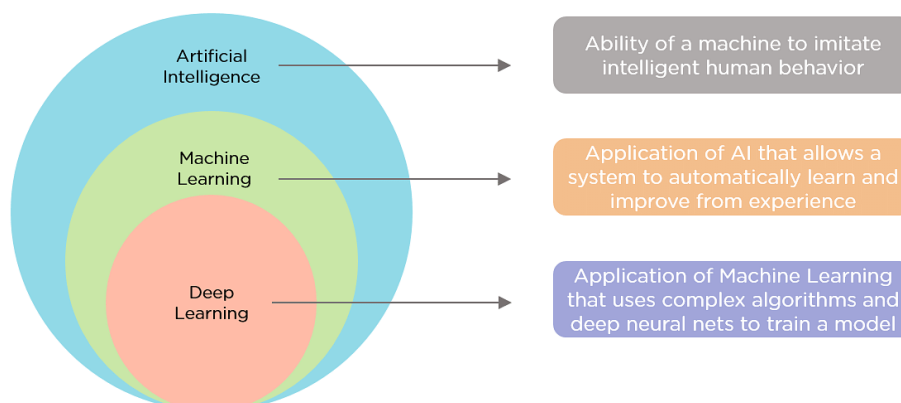
Τόσο οι αλγόριθμοι μηχανικής μάθησης όσο και οι αλγόριθμοι βαθιάς μάθησης μπορούν να εκπαιδευτούν με τη χρήση επιβλεπόμενης και μη επιβλεπόμενης μάθησης. Η βαθιά μάθηση εφαρμόζεται σε εργασίες όπως η αναγνώριση εικόνας, η αναγνώριση ομιλίας και η επεξεργασία φυσικής γλώσσας. Επιπλέον, η βαθιά μάθηση συχνά ξεπερνά την μηχανική μάθηση σε πιο περίπλοκες εργασίες αναγνώρισης μοτίβων, όπως η ταξινόμηση εικόνων (image classification) και η ανίχνευση αντικειμένων (object detection), λόγω της ικανότητάς της να μαθαίνει ιεραρχικές αναπαραστάσεις δεδομένων. Πιο συγκεκριμένα, τα μοντέλα βαθιάς μάθησης μπορούν:

- Να μάθουν σύνθετες σχέσεις μεταξύ των χαρακτηριστικών στα δεδομένα.
- Να εκπαιδευθούν με μεγάλα σύνολα δεδομένων, με αποτέλεσμα να είναι ικανά να μάθουν μεγάλη ποικιλία χαρακτηριστικών και να κάνουν πιο ακριβείς προβλέψεις για ένα πρόβλημα, επιτυγχάνοντας state-of-the-art αποδόσεις.
- Να μαθαίνουν με έναν τρόπο που βασίζεται περισσότερο στα δεδομένα, και λιγότερο στην παρέμβαση του ανθρώπου για την εκπαίδευσή τους.

Στον αντίποδα, μερικά μειονεκτήματα της βαθιάς μάθησης είναι:

- Λόγω του μεγάλου βάθους των μοντέλων, απαιτείται μεγαλύτερος αριθμός δεδομένων για εκπαίδευση. Ωστόσο, σε ορισμένα προβλήματα όπως η ταξινόμηση ιατρικών εικόνων, τα δεδομένα είναι πολλές φορές περιορισμένα και δεν αρκούν.
- Υπάρχει μεγαλύτερος κίνδυνος υπερπροσαρμογής στα δεδομένα, λόγω της περιπλοκότητας των μοντέλων.
- Απαιτείται μεγαλύτερος χρόνος εκπαίδευσης των μοντέλων, και συχνά χρειάζονται περισσότερους υπολογιστικούς πόρους.

Τέλος, μερικά από τα μοντέλα βαθιάς μάθησης αποτελούν τα συνελκτικά νευρωνικά δίκτυα (CNN) και τα Recurrent νευρωνικά δίκτυα (RNN).



Εικόνα 4.17 Η τεχνητή νοημοσύνη περιλαμβάνει τα πεδία μηχανικής μάθησης και βαθιάς μάθησης.

4.7 Τμηματοποίηση εικόνας

Η τμηματοποίηση εικόνας (image segmentation) είναι μια θεμελιώδης τεχνική στην επεξεργασία ψηφιακής εικόνας και την υπολογιστική όραση, που αφορά την διαίρεση μιας ψηφιακής εικόνας σε πολλαπλά τμήματα (περιοχές ή αντικείμενα) με σκοπό την απλοποίηση και ανάλυση της. Χρησιμοποιείται σε ένα ευρύ φάσμα εφαρμογών, όπως στην ανάλυση ιατρικών εικόνων για τον εντοπισμό καρκινικών όγκων ή σε αυτόνομα οχήματα για την ανίχνευση πεζών ή εμποδίων στον δρόμο. Με την τεχνική αυτή, αναθέεται σε κάθε pixel μιας εικόνας μια ετικέτα (label), η οποία είναι συνήθως ένας ακέραιος αριθμός, και έτσι κατηγοριοποιούνται τα pixels σε προκαθορισμένες κλάσεις και αναγνωρίζονται συγκεκριμένα αντικείμενα, όπως φαίνεται από το παράδειγμα της εικόνας 4.18.

Μια από τις υποκατηγορίες της τεχνικής της τμηματοποίησης εικόνας, είναι η σημασιολογική κατάτμηση (semantic image segmentation), κατά την οποία αναθέτονται στα pixels μιας εικόνας ετικέτες για τη κατηγοριοποίηση τους σε κλάσεις, με βάση τα χαρακτηριστικά τους. Έτσι, παράγεται ένας χάρτης τμηματοποίησης (segmentation map), ο οποίος αποτελεί μια ανακατασκευή της πραγματικής εικόνας στην οποία όμως το κάθε pixel έχει χρωματιστεί σύμφωνα με την σημασιολογική κλάση στην οποία ανήκει, με σκοπό τη δημιουργία μάσκων τμηματοποίησης (segmentation masks). Οι μάσκες αυτές είναι τμήματα της εικόνας που διαφοροποιούνται από άλλα τμήματα της ίδιας εικόνας. Για παράδειγμα, ένας χάρτης τμηματοποίησης μιας εικόνας ενός δέντρου θα περιλάμβανε πιθανώς τρεις μάσκες τμηματοποίησης: μία για το δέντρο, μία για το έδαφος και μια για τον ουρανό στο πίσω μέρος της εικόνας (background).



Εικόνα 4.18 Αναπαράσταση αποτελέσματος σημασιολογικής κατάτμησης μιας εικόνας με Τεχνητή Νοημοσύνη. Πηγή: [An overview of semantic image segmentation. \(jeremyjordan.me\)](http://An%20overview%20of%20semantic%20image%20segmentation.%20(jeremyjordan.me))

Παραδοσιακές τεχνικές για τμηματοποίηση εικόνας αποτελούν:

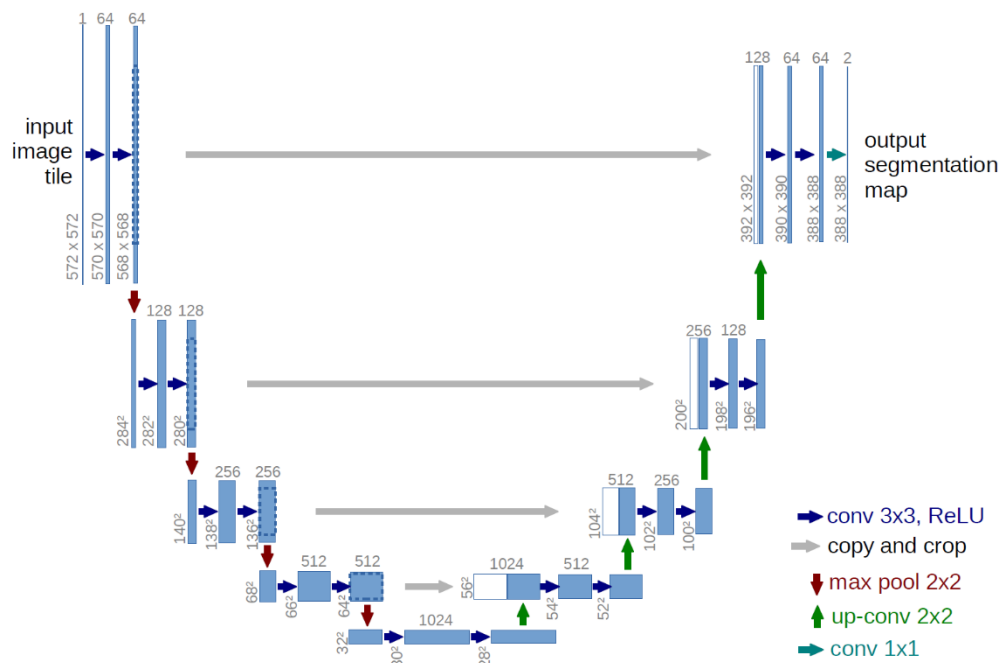
- **Thresholding:** Στην μέθοδο αυτή επιλέγεται μια τιμή κατωφλίου και τα pixels της εικόνας κατηγοριοποιούνται μεταξύ του foreground και του background με βάση τις τιμές έντασης τους.
- **Edge Detection:** Χρησιμοποιούνται αλγόριθμοι για την ανίχνευση ακμών, με βάση την αναγνώριση αλλαγών στην ένταση της εικόνας.
- **Clustering Algorithms:** Αλγόριθμοι όπως ο K-means, ομαδοποιούν τα pixels της εικόνας που αντιστοιχούν σε διαφορετικά αντικείμενα, σε clusters, με βάση κοινά χαρακτηριστικά τους όπως το χρώμα.

Για την εφαρμογή τμηματοποίησης εικόνας, και πιο συγκεκριμένα της σημασιολογικής κατάτμησης, σήμερα έχουν αναπτυχθεί και χρησιμοποιούνται μοντέλα βαθιάς μάθησης, τα οποία περιλαμβάνουν περίπλοκα νευρωνικά δίκτυα. Αν και τα βαθιά νευρωνικά δίκτυα

απαιτούν περισσότερο χρόνο, δεδομένα και υπολογιστικούς πόρους για να εκπαιδευτούν, ξεπερνούν τις παραδοσιακές μεθόδους σε απόδοση και ακρίβεια. Τέτοια δίκτυα αποτελούν το Fully Convolutional Network (FCN), SegNet, DeepLab και το U-Net, το οποίο θα αναλυθεί στο επόμενο κεφάλαιο.

4.8 Δίκτυο U-Net

Το δίκτυο U-Net αποτελεί ένα συνελκτικό νευρωνικό δίκτυο, που δημιουργήθηκε από τους Ronneberger et al. το 2015 [34] για σημασιολογική κατάτμηση ιατρικών εικόνων. Ένα από τα βασικά προβλήματα στην εκπαίδευση νευρωνικών δικτύων για την ανάλυση ιατρικών εικόνων, είναι ο περιορισμένος αριθμός επισημασμένων δεδομένων (labeled data), δηλαδή δεδομένων που συνοδεύονται από τις αντίστοιχες ετικέτες τους που υποδεικνύουν συγκεκριμένα χαρακτηριστικά ή κλάσεις. Το δίκτυο αυτό αντιμετωπίζει σε μεγάλο βαθμό αυτό το πρόβλημα, καθώς χρειάζεται μικρότερο αριθμό επισημασμένων δεδομένων για την εκπαίδευσή του, και παράλληλα παράγει τμηματοποιήσεις (segmentation masks) μεγάλης ακρίβειας.



Εικόνα 4.19 Η αρχιτεκτονική του μοντέλου U-Net [34].

Το δίκτυο αυτό ονομάζεται U-Net εξαιτίας της αρχιτεκτονικής του που είναι σε σχήμα «U» όπως φαίνεται από την εικόνα 4.19. Η αρχιτεκτονική του αποτελείται από ένα contracting path (αριστερό τμήμα) και ένα expansive path (δεξιό τμήμα). Το αριστερό τμήμα, αλλιώς γνωστό και ως encoder, αποτελείται από την επαναλαμβανόμενη εφαρμογή δύο συνεχόμενων 3x3 συνελίξεων στην εικόνα, στο τέλος των οποίων εφαρμόζεται η συνάρτηση ενεργοποίησης RELU και στη συνέχεια εκτελείται max pooling με stride 2 με αποτέλεσμα τον υποδιπλασιασμό των διαστάσεων της εικόνας. Η μείωση αυτή των διαστάσεων της εικόνας ονομάζεται υποδειγματοληψία (downsampling). Κάθε φορά που εκτελείται η υποδειγματοληψία, διπλασιάζονται τα κανάλια των χαρακτηριστικών (feature channels). Αυτή η διαδικασία επαναλαμβάνεται για όλα τα στάδια (blocks) του encoder, μέχρι και το τελευταίο

στάδιο πριν την μετάβαση στο δεξιό τμήμα του δικτύου. Το τελευταίο στάδιο, δηλαδή το 5^ο στάδιο του παραδείγματος της εικόνας 4.19, ονομάζεται «bottleneck». Το δεξιό τμήμα, αλλιώς γνωστό και ως decoder, δέχεται στο 1^ο του στάδιο ως είσοδο τους χάρτες χαρακτηριστικών από το στάδιο bottleneck, και εφαρμόζει 2x2 transposed convolutions (συνελίξεις) για τον διπλασιασμό των διαστάσεων των χαρτών και τον υποδιπλασιασμό των καναλιών χαρακτηριστικών. Στη συνέχεια, ενώνεται το αποτέλεσμα αυτής της διαδικασίας με τους αντίστοιχους χάρτες χαρακτηριστικών από τον encoder, με τη χρήση ενώσεων που ονομάζονται skip connections. Τέλος, εφαρμόζονται οι 3x3 συνελίξεις ακολουθούμενες από την συνάρτηση ενεργοποίησης RELU, όπως και στο αριστερό τμήμα του δικτύου. Αυτή η διαδικασία συνεχίζεται για όλα τα στάδια του decoder, όπου στο τελευταίο στάδιο εφαρμόζεται μια 1x1 συνέλιξη για να μετατρέψει κάθε διάνυσμα χαρακτηριστικών στο αντίστοιχο πλήθος των κατηγοριών του προβλήματος. Στη συγκεκριμένη περίπτωση της εικόνας 4.19 τα κανάλια χαρακτηριστικών είναι 64 και μετατρέπονται σε 2, δηλαδή παράγει δυαδικές μάσκες τμηματοποίησης όπου το κάθε pixel κατηγοριοποιείται σε μια από τις 2 κλάσεις. Το δίκτυο αυτό εκπαιδεύεται με τη χρήση της συνάρτησης κόστους cross-entropy:

$$E = \sum_{x \in \Omega} w(x) \log(p_{I(x)}(x)),$$

Εξίσωση 4.39

όπου η συνάρτηση p είναι η συνάρτηση soft-max ανά pixel, που εφαρμόζεται στον τελικό χάρτη χαρακτηριστικών, και w(x) ο χάρτης βαρών (weight map), που δημιουργήθηκε για να δοθούν σε ορισμένα pixels περισσότερη βαρύτητα από άλλα στον χάρτη χαρακτηριστικών:

$$p_k(x) = \frac{\exp(a_k(x))}{\sum_{k'=1}^K \exp(a_{k'}(x))}$$

Εξίσωση 4.40

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

Εξίσωση 4.41

Η δομή αυτή αποτελεί τη βασική αρχιτεκτονική του αρχικού δικτύου U-Net. Στη συνέχεια, αναπτύχθηκαν διάφορες παραλλαγές αυτής της αρχιτεκτονικής, οι οποίες προσαρμόστηκαν και βελτιώθηκαν για να καλύψουν διαφορετικές ανάγκες και εφαρμογές. Παρακάτω παρουσιάζονται ορισμένες από τις παραλλαγές του δικτύου U-Net.

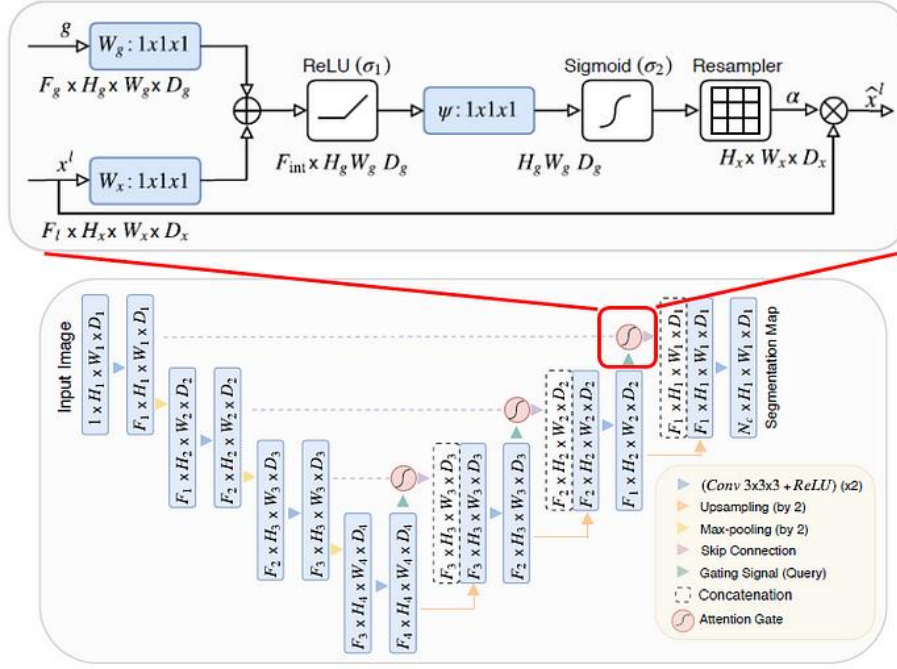
Attention U-Net

Η προσοχή (attention), στο τομέα της τμηματοποίησης εικόνας, είναι ένας τρόπος να τονιστούν μόνο οι ενεργοποιήσεις των νευρώνων που έχουν περισσότερη σημασία για ένα πρόβλημα κατά τη διάρκεια της εκπαίδευσης. Με αυτόν τον τρόπο μειώνεται η ανάγκη για χρήση περισσότερων υπολογιστικών πόρων άσκοπα για μη σημαντικές ενεργοποιήσεις, και παρέχεται στο δίκτυο καλύτερη ικανότητα γενίκευσης. Ουσιαστικά, το δίκτυο «δίνει προσοχή» σε

συγκεκριμένα χαρακτηριστικά της εικόνας που κρίνονται πιο σημαντικά από άλλα. Υπάρχουν 2 τύποι προσοχής:

- **Hard attention:** Η τεχνική αυτή λειτουργεί αναδεικνύοντας τις σχετικές περιοχές ενδιαφέροντος με την περικοπή της εικόνας ή προτείνοντας επαναλαμβανόμενα περιοχές. Μπορεί να επιλέγεται μόνο μια περιοχή της εικόνας τη φορά, και απαιτεί ενισχυτική μάθηση για την εκπαίδευση του δικτύου, με αποτέλεσμα να είναι δυσκολότερη η βελτιστοποίηση τέτοιων μοντέλων.
- **Soft attention:** Σε αυτή τη περίπτωση, αποδίδονται διαφορετικά βάρη σε διάφορα μέρη της εικόνας. Ειδικότερα, περιοχές υψηλής σημασίας πολλαπλασιάζονται με μεγαλύτερα βάρη, ενώ περιοχές με μικρότερη σημασία πολλαπλασιάζονται με μικρότερα βάρη. Έτσι, καθώς το μοντέλο εκπαιδεύεται, δίνεται περισσότερη προσοχή στις περιοχές με μεγαλύτερο βάρος. Σε αντίθεση με τη προηγούμενη τεχνική, τα βάρη εδώ μπορούν να αποδοθούν σε πολλές περιοχές μιας εικόνας. Όσο το μοντέλο εκπαιδεύεται, τα βάρη επίσης αλλάζουν ώστε το μοντέλο να παίρνει καλύτερες αποφάσεις για το σε ποιες περιοχές πρέπει να «δίνει σημασία».

Η χρήση αυτού του μηχανισμού της προσοχής, προσφέρει στο δίκτυο U-Net καλύτερη απόδοση, επειδή αντιμετωπίζει το πρόβλημα που εισάγουν τα skip connections. Κατά τη διαδικασία που οι εικόνες διπλασιάζονται σε μέγεθος στο στάδιο του decoder, δηλαδή στο δεξιό τμήμα του U-Net, η ανακατασκευή της χωρικής πληροφορίας δεν είναι ακριβής. Για την αντιμετώπιση αυτού του προβλήματος, χρησιμοποιούνται οι συνδέσεις που ονομάζονται skip connections, οι οποίες συνδυάζουν την χωρική πληροφορία του encoder με αυτή του decoder. Ωστόσο, το δίκτυο στα αρχικά επίπεδα του encoder αναγνωρίζει χαρακτηριστικά χαμηλού επιπέδου, τα οποία μεταφέρονται με τις συνδέσεις αυτές στα μεταγενέστερα στάδια του decoder, δημιουργώντας πλεονάζουσα και, ίσως όχι τόσο χρήσιμη, πληροφορία. Έτσι, η εφαρμογή της μεθόδου soft attention στις συνδέσεις αυτές μπορεί να καταστείλει τις ενεργοποιήσεις στις περιοχές που δεν κρίνονται τόσο σημαντικές, μειώνοντας τον αριθμό των χαρακτηριστικών που μεταφέρονται. Η μέθοδος αυτή χρησιμοποιήθηκε πρώτη φορά από τους Oktay et al, το 2018 [35]. Συνοψίζοντας, χρησιμοποιούνται πύλες προσοχής, ή αλλιώς attention gates πριν την ένωση των εισόδων του decoder με τις αντίστοιχες εικόνες του encoder, για να συγχωνευθούν τελικά οι πιο σχετικές πληροφορίες.



Εικόνα 4.20 Αναπαράσταση Attention U-Net. Πηγή: [Review: Attention U-Net — Learning Where to Look for the Pancreas \(Biomedical Image Segmentation\) | by Sik-Ho Tsang | Medium](#)

Όπως φαίνεται και από την εικόνα 4.20, η πύλη προσοχής δέχεται 2 εισόδους, την x η οποία είναι η είσοδος από το skip connection που προέρχεται από το στάδιο του encoder και το διάνυσμα g , αλλιώς γνωστό και ως gating vector, που προέρχεται από το στάδιο του decoder και περιέχει πληροφορίες οι οποίες έχουν συλλεχθεί από τα χαρακτηριστικά που παράγονται στον decoder. Η πληροφορία που περιέχει είναι βελτιωμένη σε σχέση με προηγούμενα επίπεδα του δικτύου και έτσι βοηθάει στην καθοδήγηση του μηχανισμού της προσοχής. Τα δύο αυτά σήματα, προστίθενται σε επίπεδο pixel, το διάνυσμα που προκύπτει περνάει μέσα από συνάρτηση ενεργοποίησης RELU και μια συνέλιξη 1×1 . Ύστερα, το διάνυσμα περνάει μέσα από τη σιγμοειδή συνάρτηση με την οποία παράγονται τελικά τα βάρη (attention coefficients) των οποίων οι τιμές βρίσκονται στο εύρος $[0,1]$. Τέλος, το διάνυσμα των βαρών δειγματοληπτείται ώστε να αποκτήσει τις διαστάσεις του διανύσματος x και έπειτα τα βάρη αυτά πολλαπλασιάζονται με το διάνυσμα x . Το διάνυσμα που προκύπτει περνάει μέσα από το skip connection. Η διαδικασία αυτή αναπαρίσταται με τις παρακάτω εξισώσεις:

$$q_{att}^l = \psi^T \left(\sigma_1 \left(W_x^T x_i^l + W_g^T g_i + b_g \right) \right) + b_\psi$$

$$a_i^l = \sigma_2 \left(q_{att}^l \left(x_i^l, g_i ; \Theta_{att} \right) \right),$$

Εξίσωση 4.42

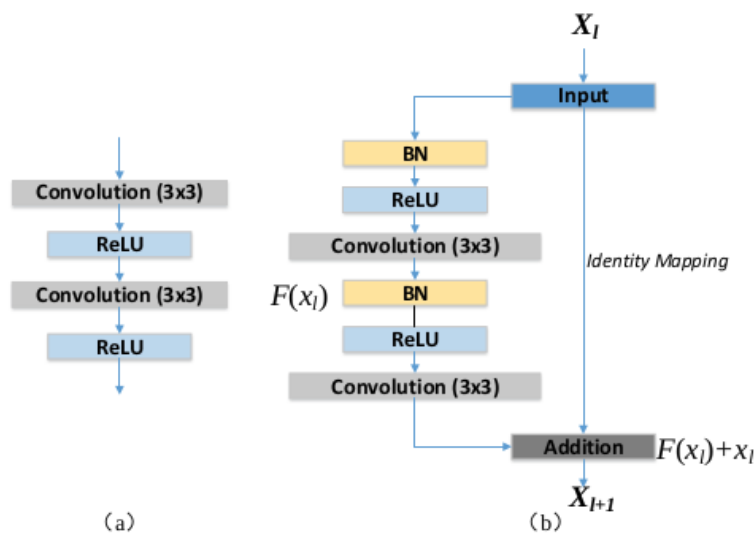
όπου x_i, g_i το διάνυσμα εισόδου και το gating vector αντίστοιχα, σ_1 η συνάρτηση ενεργοποίησης, σ_2 η σιγμοειδής συνάρτηση, W_x^T, W_g^T οι πίνακες βαρών των διανυσμάτων x και g αντίστοιχα, Θ_{att} οι παράμετροι που περιλαμβάνουν γραμμικές μετατροπές και τιμές πόλωσης b_ψ , και τέλος α_i^l τα attention coefficients.

Residual U-Net

Αυτός ο τύπος U-Net βασίζεται στην αρχιτεκτονική του δικτύου ResNet (Residual Network). Το δίκτυο ResNet το οποίο εισάχθηκε το 2015 [36], αντιμετωπίζει το πρόβλημα που είναι γνωστό ως diminishing gradients problem, σύμφωνα με το οποίο τα βαθύτερα δίκτυα αν και

μπορούν να «μαθαίνουν» ταχύτερα λόγω των περισσότερων επιπέδων, φτάνουν συχνά σε κορεσμό και οποιαδήποτε επιπλέον αύξηση των επιπέδων τους οδηγεί σε μείωση της απόδοσης τους. Το δίκτυο αυτό, λοιπόν, χρησιμοποιεί συνδέσεις ή αλλιώς skip connections, οι οποίες μεταφέρουν τους χάρτες χαρακτηριστικών από ένα επίπεδο σε ένα άλλο επίπεδο βαθύτερα στο δίκτυο. Η μέθοδος αυτή επιτρέπει στο δίκτυο να διατηρεί τους χάρτες χαρακτηριστικών ανεξάρτητα του βάθους του.

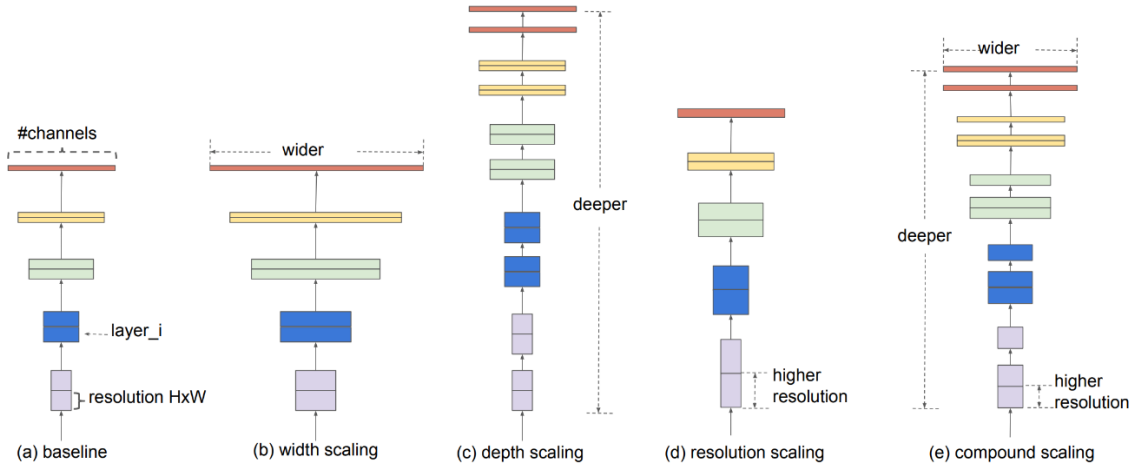
Το δίκτυο Residual U-Net ή ResUNet, αντί για συνελκτικά block σε κάθε επίπεδο, χρησιμοποιεί residual units, τα οποία αποτελούνται από δύο 3x3 συνελκτικά blocks και το identity mapping. Τα συνελκτικά blocks με τη σειρά τους αποτελούνται από ένα επίπεδο batch normalization, ένα επίπεδο συνάρτησης ενεργοποίησης RELU, και ένα συνελκτικό επίπεδο. Τέλος, το identity mapping ενώνει την είσοδο και την έξοδο του residual unit. Η αρχιτεκτονική αυτή παρουσιάζεται στην εικόνα 4.21 (b).



Εικόνα 4.21 a) basic unit που χρησιμοποιείται στο U-Net και b) residual unit με identity mapping. Πηγή: Residual Unit in the ResU-Net Architecture [8]: (a) Plain neural unit... | Download Scientific Diagram ([researchgate.net](https://www.researchgate.net))

4.9 Δίκτυο EfficientNet

Το δίκτυο EfficientNet, αποτελούν μια οικογένεια 7 δικτύων, που δημιουργήθηκαν βασισμένα στην ιδέα της κλιμάκωσης (scaling) των δικτύων για την βελτίωση της επίδοσης τους. Συχνά, για την βελτίωση της ακρίβειας των μοντέλων, κλιμακώνουμε τα δίκτυα, αυξάνοντας είτε τον αριθμό των επιπέδων τους, δηλαδή το βάθος του δικτύου, είτε το πλάτος των καναλιών τους (channels), είτε τις διαστάσεις των εικόνων που δίνουμε ως είσοδο στο μοντέλο (resolution). Η διαδικασία αυτή γίνεται πολλές φορές με τυχαίο τρόπο, δηλαδή άλλα δίκτυα κλιμακώνονται ως προς το βάθος, άλλα ως προς το πλάτος, και άλλα ως προς τις διαστάσεις των εικόνων, ή κάποιο συνδυασμό αυτών των τριών παραμέτρων. Βασισμένοι λοιπόν σε αυτήν την ιδέα, οι δημιουργοί των δικτύων EfficientNet, πρότειναν, το 2019, μια πιο συστηματική μέθοδο ταυτόχρονης κλιμάκωσης και των τριών παραμέτρων σε νευρωνικά δίκτυα, δηλαδή του βάθους, του πλάτους και των διαστάσεων εικόνας, με τη χρήση της μεθόδου «compound scaling» [37]. Η μέθοδος αυτή αναπαρίσταται στην εικόνα 4.22.



Εικόνα 4.22 Compound scaling method [37].

Η μέθοδος κλιμάκωσης «compound scaling» βασίζεται στην ταυτόχρονη κλιμάκωση του βάθους, του πλάτους και των διαστάσεων των εικόνων του δικτύου, με τη χρήση συστηματικών συντελεστών. Πιο συγκεκριμένα, η κλιμάκωση γίνεται με τη χρήση τριών συντελεστών, d για το βάθος, w για το πλάτος και r για τις διαστάσεις τις εικόνας:

$$\begin{aligned}
 d &= a^\phi \\
 w &= \beta^\phi \\
 r &= \gamma^\phi, \quad a \cdot b^2 \cdot \gamma^2 \approx 2 \quad \text{και} \quad \alpha \geq 1, \beta \geq 1, \gamma \geq 1
 \end{aligned}$$

Εξίσωση 4.43

Αν θεωρήσουμε ένα νευρωνικό δίκτυο N , κάθε επίπεδο του i μπορεί να αναπαρασταθεί ως μια συνάρτηση: $Y_i = F_i(X_i)$, όπου F_i ο τελεστής, Y_i η έξοδος και X_i η είσοδος στο δίκτυο, με διαστάσεις $\langle H_i, W_i, C_i \rangle$, όπου H_i, W_i οι διαστάσεις των εικόνων και C_i η διάσταση των καναλιών [37]. Έτσι, μπορούμε να ορίσουμε ένα νευρωνικό δίκτυο όπως παρακάτω:

$$N = \bigodot_{i=1 \dots s} F_i^{L_i} (X_{\langle H_i, W_i, C_i \rangle})$$

Εξίσωση 4.44

Εφαρμόζοντας την μέθοδο κλιμάκωσης του EfficientNet, η συνάρτηση αυτή γίνεται:

$$N(d, w, r) = \bigodot_{i=1 \dots s} \widehat{F}_i^{d \cdot L_i} (X_{\langle r \cdot \widehat{H}_i, r \cdot \widehat{W}_i, w \cdot \widehat{C}_i \rangle})$$

Εξίσωση 4.45

Η βασική ιδέα είναι ότι, όταν η εικόνα που δίνουμε ως είσοδο είναι μεγαλύτερη, το δίκτυο χρειάζεται περισσότερα επίπεδα για να μπορέσει να «δει» μεγαλύτερες περιοχές της εικόνας και περισσότερα κανάλια για να αναγνωρίσει πιο λεπτομερή χαρακτηριστικά μέσα σε αυτήν. Ωστόσο, η κλιμάκωση αυτή λαμβάνει υπόψιν και την περιορισμένη διαθέσιμη υπολογιστική ισχύ, επομένως οι παράμετροι d, w, r επιλέγονται με στόχο την απόκτηση μέγιστης δυνατής

ακρίβειας του μοντέλου με την διατήρηση των παραμέτρων κάτω από κάποια προκαθορισμένα όρια.

Αρχιτεκτονική των δικτύων EfficientNet

Η δημιουργία των δικτύων EfficientNet, ξεκίνησε με την κατασκευή ενός δικτύου βάσης (baseline network), το EfficientNet-B0, ένα νευρωνικό δίκτυο σε μέγεθος για χρήση σε κινητό τηλέφωνο (mobile-size network). Το δίκτυο αυτό αναπτύχθηκε εκτελώντας έναν αλγόριθμο αναζήτησης νευρωνικών δικτύων, αλλιώς γνωστό ως Neural Architecture Search (NAS), με τον οποίο αναζητήθηκε και βρέθηκε μια αρχιτεκτονική η οποία είχε τη καλύτερη δυνατή ακρίβεια αποτελεσμάτων, με τον λιγότερο δυνατό αριθμό FLOPS (Floating-point Operations per Second), ο οποίος αποτελεί τον αριθμό πράξεων κινητής υποδιαστολής ανά δευτερόλεπτο που εκτελεί ο υπολογιστής, όπως πρόσθεση και αφαίρεση πραγματικών αριθμών, και είναι μια μετρική της απόδοσης του H/Y. Το πρόβλημα βελτιστοποίησης της ακρίβειας παράλληλα με τον αριθμό FLOPS από τους δημιουργούς του EfficientNet, ήταν εμπνευσμένο από την δουλειά των Tan et al [38] οι οποίοι πρότειναν έναν αλγόριθμο αναζήτησης νευρωνικών δικτύων για κινητά, που στοχεύει στην εύρεση δικτύων με ένα καλό ισοζύγιο ακρίβειας και καθυστέρησης (latency).

Θέτοντας, λοιπόν, μέγιστο αριθμό πράξεων FLOPS 400 εκατομμύρια, και εκτελώντας την αναζήτηση, αναπτύχθηκε η παρακάτω αρχιτεκτονική ως το EfficientNet-B0:

Stage (Convolution Order)	Operator	Resolution	# of Channels	# of Layers
i	\hat{F}_i	$\hat{H}_i \times \hat{W}_i$	\hat{C}_i	\hat{L}_i
1	Conv3 × 3	224 × 224	32	1
2	MBCConv1, k3 × 3	112 × 112	16	1
3	MBCConv6, k3 × 3	112 × 112	24	2
4	MBCConv6, k5 × 5	56 × 56	40	2
5	MBCConv6, k3 × 3	28 × 28	80	3
6	MBCConv6, k5 × 5	14 × 14	112	3
7	MBCConv6, k5 × 5	14 × 14	192	4
8	MBCConv6, k3 × 3	7 × 7	320	1
9	Conv 1 × 1 & Pool & FC	7 × 7	1280	1

Εικόνα 4.23 Αρχιτεκτονική EfficientNet-B0. Πηγή: https://www.researchgate.net/figure/The-EfficientNet-B0-baseline-network-structure_tbl2_371935514

Στη συνέχεια, αφού έχει βρεθεί το δίκτυο βάσης, χρειαζόταν να βρεθούν οι παράμετροι α , β , γ οι οποίοι χρησιμοποιούνται για την κλιμάκωση του βάθους, του πλάτους και των διαστάσεων των εικόνων, όπως φαίνεται από την εξίσωση 4.43. Για την εύρεση των α , β , γ , ο εκθέτης ϕ ορίζεται σταθερός και ίσος με 1, και εκτελείται μια μικρού μεγέθους αναζήτηση, κλιμακώνοντας το δίκτυο B0, με βάση την εξίσωση 4.45, με διαφορετικές κάθε φορά τιμές στα α , β , γ ώστε να βρεθεί ο κατάλληλος συνδυασμός με βάση και την διαθέσιμη υπολογιστική ισχύ. Στη περίπτωση του EfficientNet, οι τιμές αυτές βρέθηκαν $\alpha = 1.2$, $\beta = 1.1$ και $\gamma = 1.15$.

Τέλος, τα α , β , γ ορίζονται ως αυτές οι σταθερές τιμές που βρέθηκαν από την προαναφερόμενη αναζήτηση, και τα δίκτυα κλιμακώνονται χρησιμοποιώντας κάθε φορά διαφορετικό αριθμό ϕ , με βάση την εξίσωση 4.43. Με αυτόν τον τρόπο, αναπτύχθηκαν τα επόμενα μοντέλα EfficientNet-B1 έως EfficientNet-B7. Τα αποτελέσματα των δικτύων αυτών,

συγκριτικά με άλλα δίκτυα παρόμοιας ακρίβειας εκπαιδευμένα στο ίδιο σύνολο δεδομένων, φαίνονται στη παρακάτω εικόνα.

Model	Top-1 Acc.	Top-5 Acc.	#Params	Ratio-to-EfficientNet	#FLOPS	Ratio-to-EfficientNet
EfficientNet-B0	76.3%	93.2%	5.3M	1x	0.39B	1x
ResNet-50 (He et al., 2016)	76.0%	93.0%	26M	4.9x	4.1B	11x
DenseNet-169 (Huang et al., 2017)	76.2%	93.2%	14M	2.6x	3.5B	8.9x
EfficientNet-B1	78.8%	94.4%	7.8M	1x	0.70B	1x
ResNet-152 (He et al., 2016)	77.8%	93.8%	60M	7.6x	11B	16x
DenseNet-264 (Huang et al., 2017)	77.9%	93.9%	34M	4.3x	6.0B	8.6x
Inception-v3 (Szegedy et al., 2016)	78.8%	94.4%	24M	3.0x	5.7B	8.1x
Xception (Chollet, 2017)	79.0%	94.5%	23M	3.0x	8.4B	12x
EfficientNet-B2	79.8%	94.9%	9.2M	1x	1.0B	1x
Inception-v4 (Szegedy et al., 2017)	80.0%	95.0%	48M	5.2x	13B	13x
Inception-resnet-v2 (Szegedy et al., 2017)	80.1%	95.1%	56M	6.1x	13B	13x
EfficientNet-B3	81.1%	95.5%	12M	1x	1.8B	1x
ResNeXt-101 (Xie et al., 2017)	80.9%	95.6%	84M	7.0x	32B	18x
PolyNet (Zhang et al., 2017)	81.3%	95.8%	92M	7.7x	35B	19x
EfficientNet-B4	82.6%	96.3%	19M	1x	4.2B	1x
SENet (Hu et al., 2018)	82.7%	96.2%	146M	7.7x	42B	10x
NASNet-A (Zoph et al., 2018)	82.7%	96.2%	89M	4.7x	24B	5.7x
AmoebaNet-A (Real et al., 2019)	82.8%	96.1%	87M	4.6x	23B	5.5x
PNASNet (Liu et al., 2018)	82.9%	96.2%	86M	4.5x	23B	6.0x
EfficientNet-B5	83.3%	96.7%	30M	1x	9.9B	1x
AmoebaNet-C (Cubuk et al., 2019)	83.5%	96.5%	155M	5.2x	41B	4.1x
EfficientNet-B6	84.0%	96.9%	43M	1x	19B	1x
EfficientNet-B7	84.4%	97.1%	66M	1x	37B	1x
GPipe (Huang et al., 2018)	84.3%	97.0%	557M	8.4x	-	-

Εικόνα 4.24 Αποτελέσματα δικτύων EfficientNet-B0 έως EfficientNet-B7 [37].

Όπως αποδεικνύεται, τα δίκτυα EfficientNet, επιτυγχάνουν αντίστοιχη αλλά και καλύτερη ακρίβεια από άλλα δίκτυα εκπαιδευμένα στο ίδιο σύνολο δεδομένων (ImageNet), με πολύ λιγότερες παραμέτρους. Συγκεκριμένα, το δίκτυο EfficientNet-B7 πέτυχε απόδοση state-of-the-art στο σύνολο δεδομένων ImageNet, με top-1 accuracy 84.4% και top-5 accuracy 97.1%, ξεπερνώντας το προηγούμενο καλύτερο νευρωνικό δίκτυο, από το οποίο το EfficientNet-B7 είναι 8.4 φορές μικρότερο σε μέγεθος.

Επιπλέον, για περαιτέρω υποστήριξη της έρευνας τους, οι δημιουργοί των δικτύων EfficientNet εφάρμοσαν την μέθοδο compound scaling στα δίκτυα MobileNets και ResNet, όπου συγκριτικά με άλλες μεθόδους κλιμάκωσης, η μέθοδος compound scaling οδήγησε στη βελτίωση της ακρίβειας τους. Τέλος, αρχικοποιώντας τα δίκτυα EfficientNet με βάρη προ-εκπαιδευμένων δικτύων στο σύνολο δεδομένων ImageNet, στη συνέχεια τα εκπαιδύσαν σε άλλα, μικρότερα σύνολα δεδομένων και αξιολόγησαν την ακρίβεια τους. Η διαδικασία αυτή ονομάζεται «μεταφορά μάθησης» ή transfer learning. Το αποτέλεσμα ήταν ότι στις περισσότερες περιπτώσεις, συγκριτικά με άλλα δίκτυα που είχαν state-of-the-art ακρίβειες σε αυτά τα μικρά σύνολα δεδομένων, τα δίκτυα EfficientNet έχουν μεγαλύτερη ακρίβεια αποτελεσμάτων με αρκετά λιγότερες παραμέτρους.

4.10 MONAI



Εικόνα 4.25 Medical Open Network for AI [39].

Το MONAI αποτελεί ένα εργαλείο, ή αλλιώς framework, που βασίζεται στη βιβλιοθήκη Pytorch [40] και είναι σχεδιασμένο για εργασίες βαθιάς μάθησης για την επεξεργασία ιατρικών εικόνων. Η κοινότητα του MONAI είναι πολύ ενεργή, καθώς είναι εργαλείο ανοιχτού κώδικα, που σημαίνει ότι οποιοσδήποτε μπορεί να έχει πρόσβαση στον κώδικα πίσω από αυτό το εργαλείο, καθώς και να συνδράμει στη βελτίωση του. Προσφέρει ένα ευρύ φάσμα εργαλείων και επιλογών για την επεξεργασία αρχείων ιατρικών εικόνων, συμπεριλαμβανομένων προκαθορισμένων μετρικών για την αξιολόγηση δικτύων, έτοιμων αρχιτεκτονικών νευρωνικών δικτύων, συναρτήσεων κόστους, καθώς και βελτιστοποιητών για την εκπαίδευση και βελτίωση των μοντέλων. Επιπλέον, παρέχει υψηλής ποιότητας, εύκολης χρήσης λογισμικό, όπως για παράδειγμα το MONAI Label, που διευκολύνει την εργασία και την συνεργασία των ερευνητών.

Το MONAI χρησιμοποιήθηκε στο πειραματικό μέρος αυτής της διπλωματικής εργασίας. Πιο συγκεκριμένα, χρησιμοποιήθηκε το U-Net μοντέλο του MONAI για την εκπαίδευση σε δεδομένα ιατρικών εικόνων, η συνάρτηση κόστους DiceCELoss, δηλαδή ένας συνδυασμός Dice και Cross Entropy Loss, και οι μετρικές Dice, Intersection Over Union (IoU), Recall, Precision, F1 για την αξιολόγηση των μοντέλων.

5. Υλικά και Μεθοδολογία

5.1 Κεντρική ιδέα

Η κλιμάκωση των νευρικών δικτύων, όπως έχει προαναφερθεί και στο κεφάλαιο 4.9, χρησιμοποιείται συχνά για την επίτευξη καλύτερης ακρίβειας αποτελεσμάτων των νευρωνικών δικτύων. Με την εισαγωγή της μεθόδου compound scaling των δικτύων EfficientNet, δημιουργήθηκε μια πιο συστηματική προσέγγιση για κλιμάκωση των νευρωνικών δικτύων, επιτυγχάνοντας ταυτόχρονα υψηλή ακρίβεια και σημαντικά μικρότερο μέγεθος μοντέλων σε σύγκριση με άλλα δίκτυα εκπαιδευμένα στο ίδιο σύνολο δεδομένων.

Η μέθοδος αυτή έχει εφαρμοστεί στα δίκτυα ResNet και MobileNetV1-V2 [37], τα οποία κατέγραψαν βελτιωμένη ακρίβεια σε σύγκριση με τα ίδια δίκτυα που κλιμακώθηκαν μόνο σε μια διάσταση, είτε με αύξηση του βάθους, είτε του πλάτους, είτε της ανάλυσης των εικόνων. Συγκεκριμένα για την αρχιτεκτονική U-Net, οι Baltruschat et al. [41] εμπνευσμένοι από τη μέθοδο compound scaling, κλιμάκωσαν τον αριθμό των συνελκτικών μπλοκ (convolutional blocks) για την αύξηση του βάθους, τον αριθμό των καναλιών του πρώτου επιπέδου για την αύξηση του πλάτους, και την ανάλυση των εικόνων. Με την ταυτόχρονη κλιμάκωση αυτών των παραμέτρων, κατάφεραν να πετύχουν βελτιωμένη απόδοση, με βάση τη μετρική IoU, στη τμηματοποίηση εικόνων μικροτομογραφίας SRμCT, σε σύγκριση με μεμονωμένη κλιμάκωση των παραμέτρων. Σε μια άλλη έρευνα, έγινε κλιμάκωση του βάθους με προσθήκη επιπλέον συνελκτικών επιπέδων σε κάθε συνελκτικό μπλοκ, παράλληλα με κλιμάκωση των καναλιών και της ανάλυσης των εικόνων σε U-Net συμμετρικά, δηλαδή όμοια στον encoder και στον decoder, αλλά και μη συμμετρικά, δηλαδή μόνο στον encoder [42]. Ως αποτέλεσμα, η μη συμμετρική κλιμάκωση οδήγησε σε καλύτερη ακρίβεια με λιγότερο αριθμό παραμέτρων.

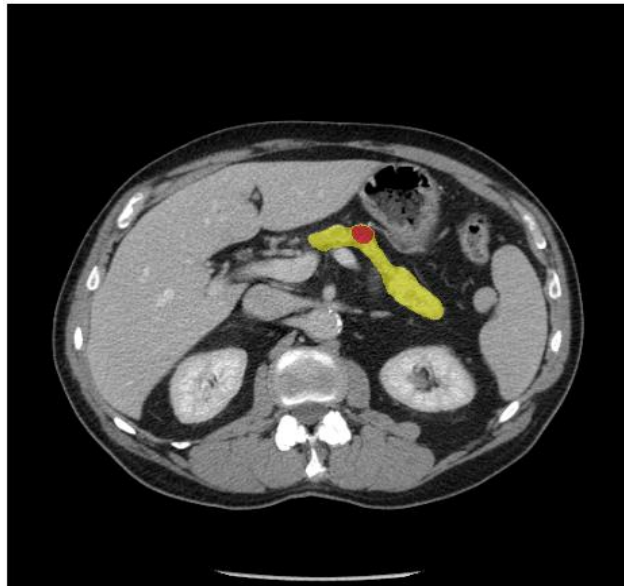
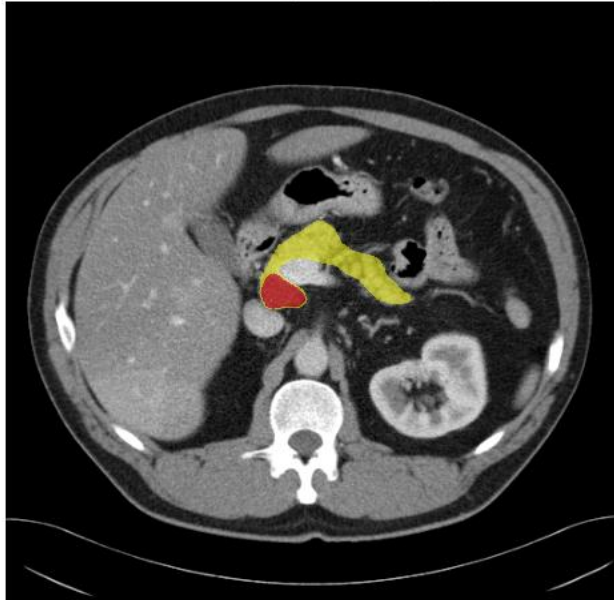
Το αντικείμενο αυτής της εργασίας είναι η εφαρμογή της μεθοδολογίας compound scaling του EfficientNet, όπως εξηγήθηκε στο κεφάλαιο 4.9, στην αρχιτεκτονική U-Net για τη σημασιολογική τμηματοποίηση αξονικών τομογραφιών ιατρικών εικόνων που περιλαμβάνουν καρκινικούς όγκους. Πιο συγκεκριμένα, διερευνάται κατά πόσο η εφαρμογή της μεθοδολογίας compound scaling μπορεί να οδηγήσει σε βελτίωση της ακρίβειας του U-Net στη διαδικασία κατάτμησης ιατρικών εικόνων αξονικών τομογραφιών.

5.2 Σύνολα δεδομένων

Για την εκπαίδευση των μοντέλων, χρησιμοποιήθηκαν 3 δημόσια διαθέσιμα σύνολα δεδομένων που παρέχει το Medical Segmentation Decathlon [43]:

- Αξονικές τομογραφίες παγκρέατος και παγκρεατικών όγκων:

Αποτελείται από 420 αξονικές τομογραφίες, μαζί με τις μάσκες τμηματοποίησης οργάνου (παγκρέατος) και όγκων, από ασθενείς που υποβλήθηκαν σε εκτομή παγκρεατικών όγκων, από τις οποίες οι 281 χρησιμοποιήθηκαν για την εκπαίδευση των μοντέλων. Οι περιοχές ενδιαφέροντος (ROIs) είναι το παγκρεατικό παρέγχυμα και οι παγκρεατικές μάζες (κύστες ή όγκοι). Τα δεδομένα αυτά αποκτήθηκαν στο Memorial Sloan Kettering Cancer Center, στη Νέα Υόρκη των ΗΠΑ. Ενδεικτικά παρουσιάζονται τρεις αξονικές τομογραφίες με τις μάσκες τμηματοποίησης τους παρακάτω (Εικόνα 5.1).

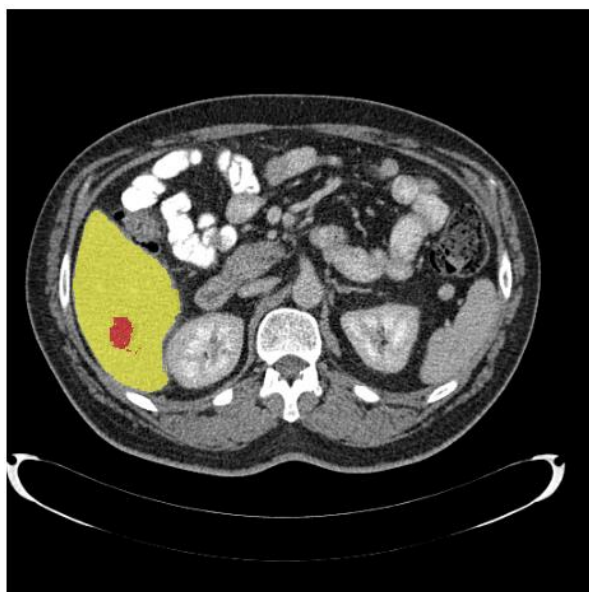


Εικόνα 5.1 Αξονικές τομογραφίες του συνόλου δεδομένων παγκρέατος και παγκρεατικών όγκων με τις αντίστοιχες μάσκες τμηματοποίησης του οργάνου και των όγκων. Με κίτρινο χρώμα απεικονίζεται το πάγκρεας και με κόκκινο χρώμα απεικονίζεται ο όγκος.

➤ Αξονικές τομογραφίες ήπατος και ηπατικών όγκων:

Αποτελείται από 201 αξονικές τομογραφίες από ασθενείς με πρωτοπαθή καρκίνο και μεταστατική νόσο στο ήπαρ, ως αποτέλεσμα καρκίνου του παχέος εντέρου, του μαστού και του πνεύμονα, μαζί με τις αντίστοιχες μάσκες τμηματοποίησης του οργάνου και των όγκων. Από το σύνολο αυτό, στη συγκεκριμένη εργασία επιλέχθηκαν οι 131 αξονικές για εκπαίδευση των μοντέλων. Οι περιοχές ενδιαφέροντος είναι το ήπαρ, και οι όγκοι εντός του ήπατος. Τα δεδομένα αποκτήθηκαν από το IRCAD Hôpitaux Universitaires, στο Στρασβούργο της Γαλλίας. Στη παρακάτω εικόνα (5.2) παρουσιάζονται ενδεικτικά ορισμένες αξονικές τομογραφίες που απεικονίζουν το ήπαρ μαζί με ηπατικούς όγκους, μαζί με τις μάσκες τμηματοποίησης τους.

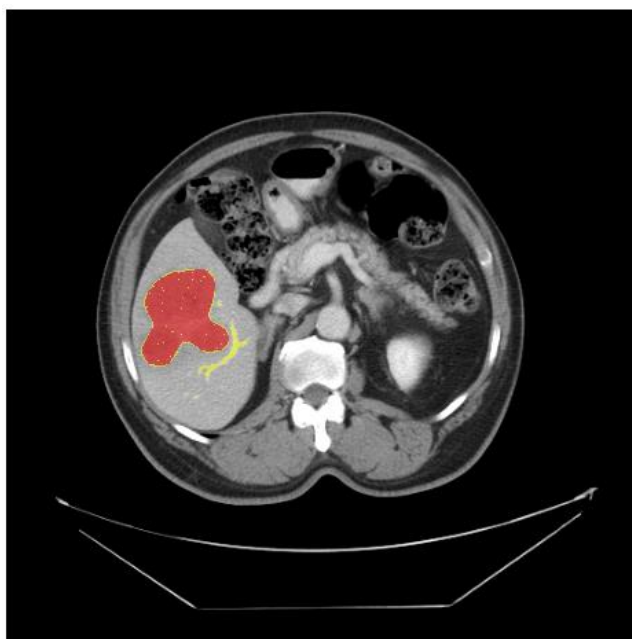


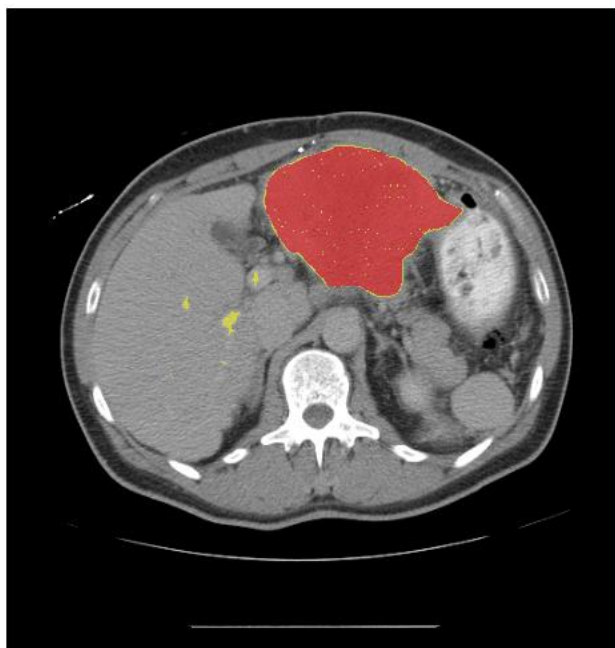


Εικόνα 5.2 Αξονικές τομογραφίες του συνόλου δεδομένων ήπατος και ηπατικών όγκων με τις αντίστοιχες μάσκες τμηματοποίησης του οργάνου και των όγκων. Με κίτρινο χρώμα απεικονίζεται το ήπαρ και με κόκκινο απεικονίζονται οι όγκοι.

➤ Αξονικές τομογραφίες ηπατικών αγγείων και ηπατικών όγκων:

Αποτελείται από 443 αξονικές τομογραφίες από ασθενείς με διάφορους πρωτοπαθείς και μεταστατικούς καρκίνους του ήπατος, μαζί με τις αντίστοιχες μάσκες τμηματοποίησης των αγγείων και των όγκων. Από αυτές, οι 303 χρησιμοποιήθηκαν για την εκπαίδευση των μοντέλων σε αυτή την εργασία. Οι περιοχές ενδιαφέροντος είναι τα αγγεία και οι όγκοι εντός του ήπατος. Τα δεδομένα αυτά αποκτήθηκαν από το Memorial Sloan Kettering Cancer Center στην Νέα Υόρκη των ΗΠΑ. Παρακάτω, στην εικόνα 5.3 παρουσιάζονται ενδεικτικά ορισμένες από τις αξονικές τομογραφίες ηπατικών αγγείων και ηπατικών όγκων, μαζί με τις μάσκες τμηματοποίησης τους.





Εικόνα 5.3 Αξονικές τομογραφίες του συνόλου δεδομένων ηπατικών αγγείων και ηπατικών όγκων με τις αντίστοιχες μάσκες τμηματοποίησης του οργάνου και των όγκων. Με κίτρινο χρώμα απεικονίζονται τα ηπατικά αγγεία ενώ με κόκκινο χρώμα απεικονίζονται οι όγκοι.

5.3 Προεπεξεργασία δεδομένων

Για να εξασφαλιστεί η καλύτερη δυνατή ποιότητα και συνοχή των δεδομένων, ακολουθήθηκε και για τα τρία σύνολα δεδομένων, η μεθοδολογία προεπεξεργασίας που χρησιμοποιήθηκε από τους δημιουργούς του μοντέλου nnU-Net [44].

Αρχικά, για το κάθε σύνολο δεδομένων, βρίσκουμε τις τιμές έντασης (intensity values) όλων των αξονικών οι οποίες αντιστοιχούν σε περιοχές ενδιαφέροντος, αποκλείοντας το background. Συγκεκριμένα, υπολογίζονται οι τιμές έντασης των εικόνων από τις περιοχές που αντιστοιχούν σε όργανο ή όγκους (με βάση τις ετικέτες), εξαιρώντας τις περιοχές γύρω από αυτά. Αυτές οι τιμές μπαίνουν σε μια λίστα, και χρησιμοποιούνται για τον υπολογισμό της μέσης τιμής μ και

της τυπικής απόκλισης σ του συνόλου δεδομένων. Με τη διαδικασία αυτή, το σύνολο δεδομένων κανονικοποιείται, ώστε τα δεδομένα να βρίσκονται στην ίδια κλίμακα και στο ίδιο εύρος τιμών, σύμφωνα με την παρακάτω εξίσωση:

$$Z = \frac{x - \mu}{\sigma}$$

Εξίσωση 5.1

Επιπλέον, πραγματοποιείται κατωφλίωση της έντασης όλων των εικονοστοιχείων των εικόνων εντός του διαστήματος των [0.005, 0.995] εκατοστημορίων, δηλαδή κανονικοποιούνται οι τιμές έντασης των εικονοστοιχείων μεταξύ του 0.5ου και του 99.5ου εκατοστημορίου. Αυτή η διαδικασία εφαρμόζεται για να απομακρυνθούν πιθανές ακραίες τιμές εντάσεων.

Στη συνέχεια, υπολογίζεται η διάμεση απόσταση των εικονοστοιχείων (σε χιλιοστά) στους άξονες X, Y και Z αντίστοιχα. Η απόσταση των εικονοστοιχείων στους άξονες X, Y αναφέρεται στην απόσταση του κέντρου του κάθε εικονοστοιχείου από το κέντρο των γειτονικών εικονοστοιχείων, ενώ στον άξονα Z είναι η απόσταση μεταξύ κάθε δισδιάστατης εικόνας, οι οποίες συνολικά συνθέτουν την τρισδιάστατη αξονική τομογραφία. Για το σύνολο δεδομένων του ήπατος και των ηπατικών όγκων, η διάμεση απόσταση βρέθηκε [0.7676, 0.7676, 1]. Για τα σύνολα δεδομένων του παγκρέατος και των ηπατικών αγγείων και όγκων, δηλαδή το πρώτο και το τρίτο σύνολο δεδομένων που χρησιμοποιήθηκε, λόγω ανισοτροπίας των δεδομένων, εκτελείται ένα επιπλέον βήμα που είναι η εύρεση του 10ου εκατοστημορίου για τις αποστάσεις στον άξονα Z. Έτσι, υπολογίζονται οι αποστάσεις [0.8, 0.8, 2.5] για το σύνολο δεδομένων του παγκρέατος και οι αποστάσεις [0.8, 0.8, 1] για το σύνολο δεδομένων των ηπατικών αγγείων και όγκων. Οι τιμές αυτές χρησιμοποιούνται για τον μετασχηματισμό των δεδομένων με χρήση παρεμβολής, για να εξασφαλιστεί ότι όλες οι εικόνες θα έχουν τις ίδιες αποστάσεις εικονοστοιχείων και στους τρεις άξονες.

Τέλος, εκτελείται περικοπή των εικόνων για να περιλαμβάνουν όσο γίνεται περισσότερο τις περιοχές ενδιαφέροντος, και όσο γίνεται λιγότερο background.

5.4 Αρχιτεκτονική του μοντέλου και εκπαίδευση

Αρχικά, ακολουθώντας την μεθοδολογία του EfficientNet, πρέπει να βρεθεί το δίκτυο U-Net B0, το οποίο θα λειτουργήσει ως το βασικό δίκτυο το οποίο θα κλιμακωθεί ως προς το βάθος, το πλάτος και την ανάλυση των εικόνων. Για το σκοπό αυτό, χρησιμοποιείται το δίκτυο U-Net του MONAI, το οποίο αντί για απλά συνελκτικά μπλοκ στον encoder και τον decoder, όπως αυτό που αναλύθηκε στο κεφάλαιο 4.8, αποτελείται από residual units. Για κάθε σύνολο δεδομένων, για την εύρεση του βασικού δικτύου B0, χρησιμοποιείται U-Net με 1 residual unit για κάθε επίπεδο και αποτελείται από 5 encoder blocks και 5 decoder blocks. Επιπλέον, για την εκπαίδευση και αξιολόγηση του μοντέλου, το κάθε σύνολο δεδομένων χωρίστηκε σε δύο μέρη: ένα σύνολο εκπαίδευσης (training set) που αποτελεί το 80% των εικόνων του συνόλου δεδομένων και ένα σύνολο αξιολόγησης (test set) που είναι το υπόλοιπο 20% των εικόνων. Το μέγεθος του batch που χρησιμοποιήθηκε είναι 8, η συνάρτηση κόστους είναι η DiceCELoss του MONAI, και ο βελτιστοποιητής (optimizer) που χρησιμοποιήθηκε για την ελαχιστοποίηση της συνάρτησης κόστους είναι ο AdamW. Επίσης, ο ρυθμός μάθησης (learning rate) που επιλέγεται είναι 0.0003.

Λόγω περιορισμένων υπολογιστικών πόρων, επιλέγονται τμήματα (patches) διαστάσεων 96x96x96 από τις αρχικές εικόνες. Η επιλογή των patches γίνεται με βάση την πιθανότητα εμφάνισης συγκεκριμένων ανατομικών δομών ή παθολογικών ευρημάτων. Συγκεκριμένα, για το σύνολο δεδομένων του παγκρέατος, επιλέγονται patches με πιθανότητα 30% να περιέχουν

πάγκρεας και 50% να περιέχουν παγκρεατικό όγκο. Αντίστοιχα, για το σύνολο δεδομένων του ήπατος, τα patches έχουν πιθανότητα 30% να περιέχουν ήπαρ και 50% να περιέχουν ηπατικό όγκο. Τέλος, για το σύνολο δεδομένων των ηπατικών αγγείων, επιλέγονται patches με πιθανότητες 40% για αγγεία και 40% για ηπατικούς όγκους. Αυτές οι πιθανότητες σχετίζονται με τη δυνατότητα κάθε patch να περιέχει τις συγκεκριμένες ανατομικές ή παθολογικές δομές, ώστε να διασφαλιστεί η επαρκής εκπροσώπηση τους στην εκπαίδευση του μοντέλου.

Επιπλέον, λόγω του περιορισμένου αριθμού δεδομένων και για την αποφυγή της υπερπροσαρμογής των μοντέλων στα δεδομένα εκπαίδευσης, χρησιμοποιούνται οι κάτωθι τρόποι επαύξησης των δεδομένων:

- Τυχαίος Αφινικός Μετασχηματισμός (Random Affine Transformation): Είναι γεωμετρικός μετασχηματισμός που περιλαμβάνει μετατόπιση των εικόνων, περιστροφή, κλιμάκωση. Παράδειγμα εφαρμογής αυτού του μετασχηματισμού αναπαρίσταται στην εικόνα 5.4. Στη συγκεκριμένη εργασία χρησιμοποιείται για τυχαία κλιμάκωση των εικόνων με παράγοντα κλιμάκωσης να κυμαίνεται στο διάστημα [0.8, 1.2] αλλά και για τυχαία περιστροφή των εικόνων έως και 20 μοίρες, σε οποιαδήποτε κατεύθυνση.
- Τυχαία Αναστροφή (Random Flip Transformation): Γίνεται αναστροφή της εικόνας κατά τους τρεις άξονες X, Y, Z με πιθανότητα 50% σε κάθε άξονα.



Εικόνα 5.4 Εφαρμογή Αφινικού Μετασχηματισμού σε εικόνα. Πηγή: [Random Affine Transformations in PyTorch | by Mark Ai Code | Aug, 2024 | Medium](#)

Για την αξιολόγηση των μοντέλων χρησιμοποιήθηκαν οι μετρικές:

- Dice score: Μετρά την ομοιότητα μεταξύ δύο συνόλων και κυμαίνεται στο διάστημα [0,1], όπως έχει αναλυθεί και στο κεφάλαιο 4.5.3. Σε αυτή την εργασία, ο μέσος δείκτης Dice (mean Dice score) υπολογίζεται ξεχωριστά για το όργανο και τον όγκο στο τέλος κάθε εποχής, λαμβάνοντας τον μέσο όρο όλων των εικόνων οργάνου και όγκου αντίστοιχα του συνόλου αξιολόγησης (test set).
- Intersection over Union (IoU): Υπολογίζει το ποσοστό επικάλυψης μεταξύ της προβλεπόμενης τιμής από το μοντέλο και της πραγματικής τιμής. Στη συγκεκριμένη εργασία, υπολογίζεται ο μέσος όρος του IoU για το όργανο και τον όγκο, λαμβάνοντας υπόψιν τον μέσο όρο από τις τιμές όλων των εικόνων του συνόλου αξιολόγησης (test set) στο τέλος κάθε εποχής.

$$IoU = \frac{A \cap B}{A \cup B}$$

Εξίσωση 5.2

- Precision: Μετρά πόσο συχνά ένα μοντέλο μηχανικής μάθησης προβλέπει σωστά την θετική κλάση (Εικόνα 5.5). Υπολογίζεται διαιρώντας τον αριθμό των σωστών θετικών προβλέψεων (true positives) με το σύνολο των περιπτώσεων που το μοντέλο πρόβλεψε ως θετικές (συμπεριλαμβανομένων των true positives και false positives). Στη συγκεκριμένη εργασία, υπολογίζεται ο μέσος όρος του Precision για το όργανο και τον όγκο, λαμβάνοντας υπόψιν τον μέσο όρο από τις τιμές όλων των εικόνων του συνόλου αξιολόγησης (test set) στο τέλος κάθε εποχής.

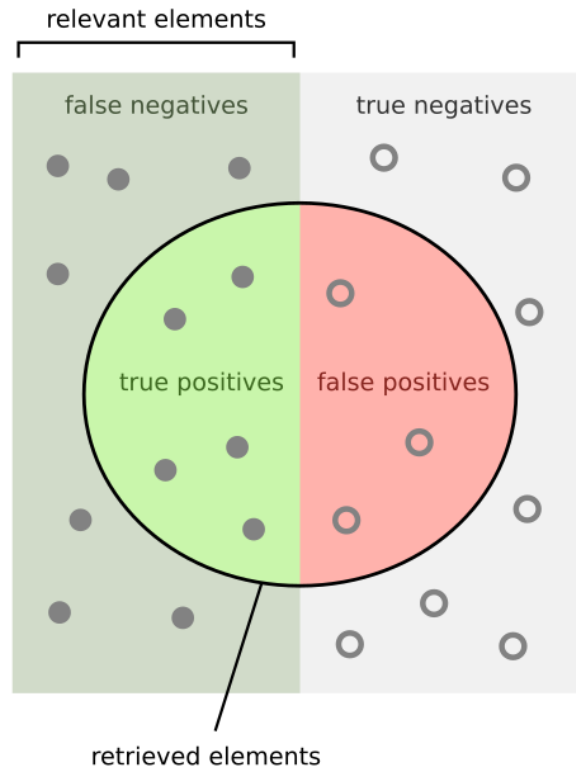
$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Εξίσωση 5.3

- Recall: Μετρά πόσο συχνά ένα μοντέλο μηχανικής μάθησης εντοπίζει σωστά τις θετικές περιπτώσεις (true positives) από όλα τα πραγματικά θετικά δείγματα στο σύνολο δεδομένων (Εικόνα 5.5). Υπολογίζεται διαιρώντας τον αριθμό των true positives με τον συνολικό αριθμό των θετικών περιπτώσεων. Το τελευταίο περιλαμβάνει τόσο τα true positives και τα false negatives (περιπτώσεις που το μοντέλο παρέλειψε να αναγνωρίσει ως θετικές). Στη συγκεκριμένη εργασία, υπολογίζεται ο μέσος όρος της μετρικής Recall για το όργανο και τον όγκο, λαμβάνοντας υπόψιν τον μέσο όρο από τις τιμές όλων των εικόνων του συνόλου αξιολόγησης (test set) στο τέλος κάθε εποχής.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Εξίσωση 5.4



How many retrieved items are relevant?

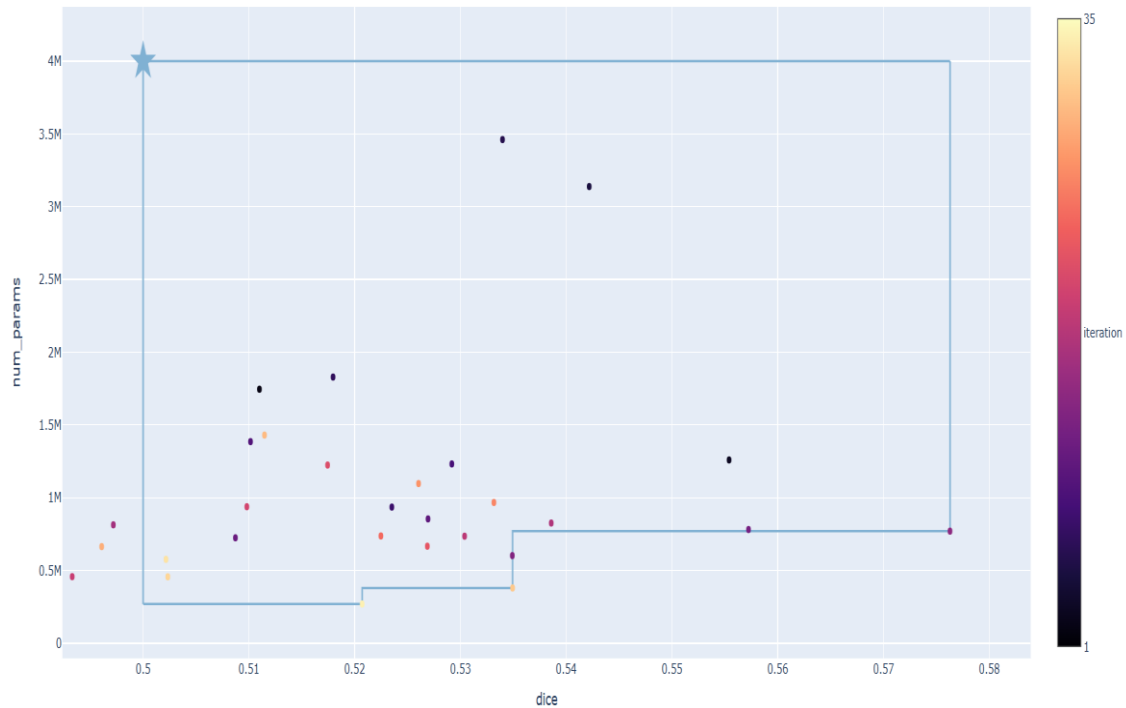
$$\text{Precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}$$

How many relevant items are retrieved?

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

Εικόνα 5.5 Μετρικές Precision & Recall. Πηγή: https://en.wikipedia.org/wiki/Precision_and_recall

Οι παραπάνω παράμετροι διατηρούνται σταθερές και εκτελείται Αναζήτηση Αρχιτεκτονικής Νευρωνικών Δικτύων (Neural Architecture Search) με χρήση της πλατφόρμας Ax [45]. Η διαδικασία επικεντρώνεται στην εύρεση των βέλτιστων τιμών για τα κανάλια (φίλτρα) του πρώτου επιπέδου του encoder, καθώς και των καναλιών των επόμενων επιπέδων. Η αναζήτηση γίνεται εκπαιδύοντας τα μοντέλα για 100 εποχές. Στόχος της αναζήτησης αυτής είναι να βρεθεί το βασικό μοντέλο, B0, το οποίο θα έχει το μέγιστο δυνατό Dice score, και ταυτόχρονα μέγιστο αριθμό παραμέτρων 4 εκατομμύρια (4M). Για το πρώτο σύνολο δεδομένων του παγκρέατος και των παγκρεατικών όγκων, η αναζήτηση οδηγεί στα παρακάτω αποτελέσματα:



Εικόνα 5.6 Αποτελέσματα Αναζήτησης Νευρωνικών δικτύων για το σύνολο δεδομένων του παγκρέατος. Ο οριζόντιος άξονας είναι οι τιμές Dice score και ο κατακόρυφος άξονας είναι ο αριθμός των παραμέτρων των μοντέλων.

Με βάση το παραπάνω γράφημα, επιλέγεται το παρακάτω μοντέλο U-Net το οποίο έχει 702 χιλιάδες (702K) παραμέτρους.

Στάδιο	Ανάλυση	Κανάλια	Συνελκτικά επίπεδα (Conv3x3)
1	96x96x96	20	3
2	48x48x48	28	3
3	24x24x24	40	3
4	12x12x12	60	3
5	6x6x6	89	2

Πίνακας 1 Δίκτυο U-Net B0 εκπαιδευμένο στο σύνολο δεδομένων του παγκρέατος. Αποτελείται από 5 στάδια στον encoder και 5 στάδια στον decoder. Η ανάλυση των εικόνων (patches) δίνεται σε pixels. Σε κάθε στάδιο του encoder η ανάλυση υποδιπλασιάζεται. Το κάθε επίπεδο του encoder και του decoder αποτελείται από 3 συνελκτικά επίπεδα εκτός από το τελευταίο επίπεδο που αποτελείται από 2.

Όμοια, εκπαιδύεται το U-Net στο σύνολο δεδομένων του ήπατος και των ηπατικών όγκων, και επιλέγεται, μετά την Αναζήτηση Αρχιτεκτονικής Νευρωνικών Δικτύων, η παρακάτω αρχιτεκτονική. Το δίκτυο έχει Dice score 66,5% και αριθμό παραμέτρων 768K.

Στάδιο	Ανάλυση	Κανάλια	Συνελκτικά επίπεδα (Conv3x3)
1	96x96x96	16	3
2	48x48x48	26	3
3	24x24x24	45	3
4	12x12x12	69	3
5	6x6x6	83	2

Πίνακας 2 Αρχιτεκτονική U-Net B0 για το σύνολο δεδομένων του ήπατος και των ηπατικών όγκων. Αποτελείται από 5 στάδια στον encoder και 5 στάδια στον decoder. Η ανάλυση των εικόνων (patches) δίνεται σε pixels. Σε κάθε στάδιο του encoder η ανάλυση υποδιπλασιάζεται. Το κάθε επίπεδο του encoder και του decoder αποτελείται από 3 συνελκτικά επίπεδα εκτός από το τελευταίο επίπεδο που αποτελείται από 2.

Τέλος, εκπαιδύοντας το U-Net στο σύνολο δεδομένων των ηπατικών αγγείων και καρκινικών ηπατικών όγκων, επιλέγεται η παρακάτω αρχιτεκτονική. Το δίκτυο έχει Dice score 46,3% και αριθμό παραμέτρων 537K.

Στάδιο	Ανάλυση	Κανάλια	Συνελκτικά επίπεδα (Conv3x3)
1	96x96x96	27	3
2	48x48x48	29	3
3	24x24x24	31	3
4	12x12x12	55	3
5	6x6x6	60	2

Πίνακας 3 Αρχιτεκτονική δικτύου U-Net B0 εκπαιδευμένο στο σύνολο δεδομένων την ηπατικών αγγείων και ηπατικών όγκων. Αποτελείται από 5 στάδια στον encoder και 5 στάδια στον decoder. Η ανάλυση των εικόνων (patches) δίνεται σε pixels. Σε κάθε στάδιο του encoder η ανάλυση υποδιπλασιάζεται. Το κάθε επίπεδο του encoder και του decoder αποτελείται από 3 συνελκτικά επίπεδα εκτός από το τελευταίο επίπεδο που αποτελείται από 2.

Στη συνέχεια, για κάθε σύνολο δεδομένων ξεχωριστά, με βάση τις εξισώσεις 4.43, ορίζεται ο συντελεστής $\varphi=1$, και αναζητούνται οι τιμές των παραμέτρων κλιμάκωσης α , β , γ με χρήση της πλατφόρμας Ax. Κατά την διαδικασία αυτή και πάλι, αναζητούνται οι παράμετροι με στόχο την επίτευξη βελτιωμένου Dice score, διατηρώντας μικρό αριθμό παραμέτρων. Σε όλες τις περιπτώσεις ισχύει ο περιορισμός της εξίσωσης 4.43.

- Για το σύνολο δεδομένων του παγκρέατος υπολογίζονται οι τιμές $\alpha=1.3$, $\beta=1.2$, $\gamma=1.1$.

- Για το σύνολο δεδομένων του ήπατος και των ηπατικών όγκων υπολογίζονται οι τιμές: $\alpha=1.3$, $\beta=1.3$, $\gamma=1.1$.
- Για το σύνολο δεδομένων των ηπατικών αγγείων και ηπατικών όγκων υπολογίζονται οι τιμές: $\alpha=1.2$, $\beta=1.1$, $\gamma=1.1$.

Τέλος, οι τιμές των α , β , γ για κάθε σύνολο δεδομένων ορίζονται σταθερές βάσει της αναζήτησης και με την διαφοροποίηση του συντελεστή φ κλιμακώνεται το βασικό δίκτυο B0 ως προς το βάθος, το πλάτος και την ανάλυση των εικόνων. Το φ επιλέγεται να ξεκινά από τη τιμή 0.5 και να φθάνει ως το 3 με βήμα 0.5.

6. Αποτελέσματα

Για την αξιολόγηση των αποτελεσμάτων στο 20% των εικόνων (test set) για κάθε σύνολο δεδομένων, χρησιμοποιούνται οι μετρικές Dice score, Precision, Recall, IoU, F1, όπως αυτές αναφέρθηκαν παραπάνω.

1^ο σύνολο δεδομένων – Πάγκρεας και παγκρεατικοί όγκοι

Για την κλιμάκωση του βασικού δικτύου, τα δίκτυα εκπαιδεύονται μέχρι 200 εποχές λόγω του μεγαλύτερου μεγέθους τους από το βασικό, με ρυθμό μάθησης 0.001 ο οποίος μειώνεται σταδιακά σύμφωνα με μια συνημιτονική συνάρτηση μέχρι 0.00001 ώστε το μοντέλο να έχει πιο σταθερή ακρίβεια αποτελεσμάτων. Επίσης, χρησιμοποιείται dropout λόγω της προοδευτικής αύξησης του βάθους και της πολυπλοκότητας των μοντέλων, για την αποφυγή της υπερπροσαρμογής τους στα δεδομένα εκπαίδευσης.

Παράλληλα, με σκοπό τη σύγκριση των αποτελεσμάτων εκπαιδεύεται και το U-Net με τη χρήση των default παραμέτρων που αναφέρθηκαν στο κεφάλαιο 4.8. Το δίκτυο εκπαιδεύτηκε για 200 εποχές, με ρυθμό μάθησης 0.001 ο οποίος μειώνεται σταδιακά σύμφωνα με συνημιτονική συνάρτηση έως τιμή 0.00001 για 200 εποχές. Το αποτέλεσμα φαίνεται στην τελευταία γραμμή των ακόλουθων πινάκων.

Μοντέλο	Mean Dice Score		
	Πάγκρεας	Όγκος	Μέσος όρος
B0	0.6066	0.4002	0.5034
B1	0.7007	0.5156	0.6081
B2	0.6987	0.4352	0.5669
B3	0.6928	0.4083	0.5505
B4	0.6895	0.4567	0.5731
B5	0.6963	0.4459	0.5711
B6	0.7092	0.4558	0.5825
Default U-Net	0.7127	0.3822	0.5475

Πίνακας 4 Αποτελέσματα mean Dice score για το σύνολο δεδομένων του παγκρέατος και των παγκρεατικών όγκων.

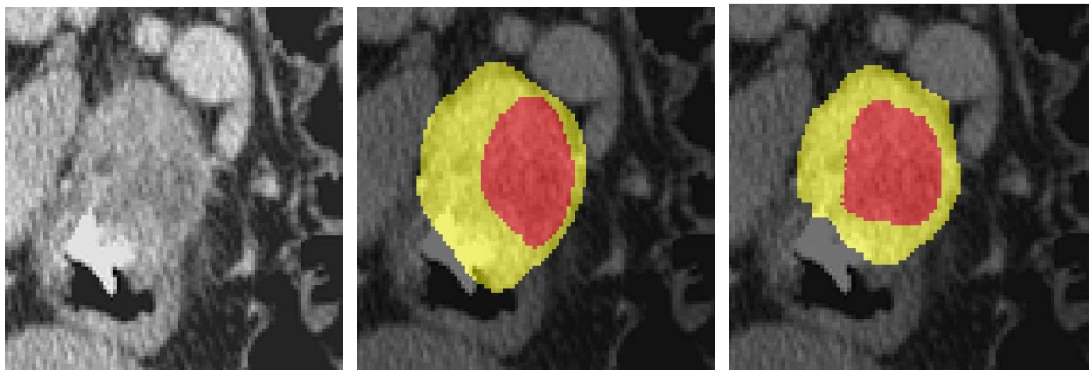
Από τα αποτελέσματα παρατηρείται ότι με την κλιμάκωση των δικτύων η μετρική Dice αυξάνεται. Το μοντέλο B1 φαίνεται να έχει την καλύτερη επίδοση μεταξύ των B0 – B6 όσον αφορά την αναγνώριση του παγκρέατος και του όγκου με τιμές 0.7007 για το πάγκρεας και 0.5156 για τον όγκο. Ωστόσο, για τα υπόλοιπα μοντέλα B2 - B6 δεν παρατηρείται ιδιαίτερη διαφοροποίηση της μετρικής και η βελτίωση τους συγκριτικά με το αρχικό μοντέλο B0 είναι πιο περιορισμένη.

Μοντέλο	IoU	Recall	Precision
B0	38,4%	59,9%	63,4%
B1	49,5%	71,2%	70,7%
B2	46%	78,1%	64,2%
B3	44%	76,3%	63,3%
B4	45,7%	83,3%	63,3%
B5	46,1%	77,5%	64,1%
B6	47,1%	77,7%	67,7%
Default U-Net	46%	73,1%	69,1%

Πίνακας 5 Αποτελέσματα μετρικών IoU, Recall, Precision για το σύνολο δεδομένων του παγκρέατος και των παγκρεατικών όγκων.

Στον παραπάνω πίνακα παρατίθενται και οι υπόλοιπες μετρικές για την αξιολόγηση των κλιμακούμενων δικτύων καθώς και του δικτύου U-Net με τις default παραμέτρους στο σύνολο δεδομένων του παγκρέατος. Παρατηρείται ότι η κλιμάκωση επιφέρει βελτίωση των μετρικών. Το μοντέλο B1 παρουσιάζει αρκετή βελτίωση συγκριτικά με το B0, ωστόσο η βελτίωση δεν συνεχίζεται στα επόμενα δίκτυα, με εξαίρεση τον δείκτη Recall.

Παρακάτω, στο σχήμα 6.1 παρατίθεται ενδεικτικά ένα patch από τις αρχικές εικόνες του test set του παγκρέατος, διαστάσεων 96x96. Η μεσαία εικόνα αποτελεί την μάσκα τμηματοποίησης (ground truth label) αυτού του patch, όπου με κίτρινο χρώμα αναπαρίσταται το πάγκρεας και με κόκκινο χρώμα ο όγκος. Δεξιά, στην τελευταία εικόνα, αναπαρίσταται η πρόβλεψη τμηματοποίησης παγκρέατος και όγκου από το μοντέλο B1. Παρατηρείται ότι το μοντέλο έχει προβλέψει το όργανο και τον όγκο ωστόσο δεν οριοθετεί με μεγάλη ακρίβεια τις δύο περιοχές.



Εικόνα 6.1 Παράδειγμα τμηματοποίησης του μοντέλου B1. Η πρώτη εικόνα αποτελεί patch διαστάσεων 96x96 από την αρχική εικόνα. Η μεσαία εικόνα είναι η μάσκα τμηματοποίησης αυτού του patch. Η δεξιά εικόνα αποτελεί την πρόβλεψη για τμηματοποίηση από το μοντέλο.

2^ο σύνολο δεδομένων – Ήπαρ και ηπατικοί όγκοι

Για την κλιμάκωση του βασικού δικτύου, τα δίκτυα εκπαιδεύονται μέχρι 150 εποχές λόγω του μεγαλύτερου μεγέθους τους από το βασικό, με ρυθμό μάθησης 0.001 ο οποίος μειώνεται σταδιακά σύμφωνα με μια συνημιτονική συνάρτηση μέχρι 0.00001 ώστε το μοντέλο να έχει πιο σταθερή ακρίβεια αποτελεσμάτων. Επίσης, όπως και προηγουμένως, χρησιμοποιείται dropout λόγω της προοδευτικής αύξησης του βάθους και της πολυπλοκότητας των μοντέλων, για την αποφυγή της υπερπροσαρμογής τους στα δεδομένα εκπαίδευσης.

Παράλληλα, με σκοπό τη σύγκριση των αποτελεσμάτων εκπαιδεύεται και το U-Net με τη χρήση των default παραμέτρων που αναφέρθηκαν στο κεφάλαιο 4.8. Το δίκτυο εκπαιδεύτηκε για 200 εποχές, με ρυθμό μάθησης 0.001 ο οποίος μειώνεται σταδιακά σύμφωνα με συνημιτονική συνάρτηση έως τιμή 0.00001 για 200 εποχές. Το αποτέλεσμα φαίνεται στην τελευταία γραμμή των ακόλουθων πινάκων.

Μοντέλο	Mean Dice Score		
	Ήπαρ	Όγκος	Μέσος όρος
B0	0.806	0.456	0.631
B1	0.8351	0.5235	0.6792
B2	0.8517	0.462	0.6569
B3	0.832	0.3814	0.6067
B4	0.8068	0.5243	0.6656
B5	0.8298	0.4948	0.6623
B6	0.8429	0.5629	0.7029
Default U-Net	0.8689	0.4437	0.6563

Πίνακας 6 Αποτελέσματα mean Dice score για το σύνολο δεδομένων του ήπατος και των ηπατικών όγκων.

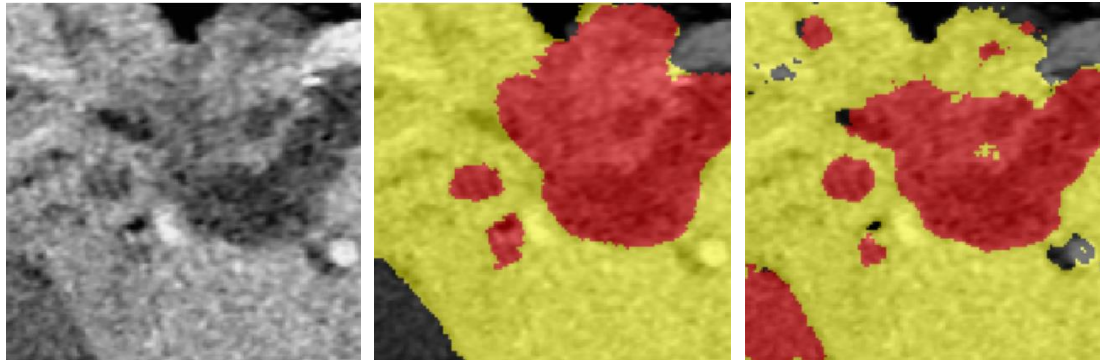
Από τα αποτελέσματα φαίνεται ότι με την κλιμάκωση των δικτύων για το σύνολο δεδομένων του ήπατος και των ηπατικών όγκων, η μετρική Dice αυξάνεται τόσο για το ήπαρ όσο και για τους όγκους. Την καλύτερη απόδοση από τα μοντέλα B1 – B6 στη συγκεκριμένη περίπτωση έχει το δίκτυο B6, με τιμή mean Dice 0.8429 για το ήπαρ και 0.5629 για τους όγκους.

Μοντέλο	IoU	Recall	Precision
B0	57,7%	86,6 %	83%
B1	68,3%	92,4%	91,6%
B2	65,3%	90,1%	89,1%
B3	59,5%	89,1%	88,7%
B4	64,4%	93,9%	93,3%
B5	68%	96%	91,5%
B6	66,8%	93,1%	92%
Default U-Net	68,9%	95,5%	93,3%

Πίνακας 7 Αποτελέσματα μετρικών IoU, Recall, Precision για το σύνολο δεδομένων του ήπατος και των ηπατικών όγκων.

Στον παραπάνω πίνακα φαίνονται και τα αποτελέσματα των υπόλοιπων μετρικών των κλιμακούμενων δικτύων και του U-Net με τις default παραμέτρους εκπαιδευμένα στο σύνολο δεδομένων του ήπατος. Παρατηρείται ότι τα μοντέλα B1 και B5 έχουν την καλύτερη απόδοση από αυτά που έχουν κλιμακωθεί από το B0, ωστόσο το U-Net έχει γενικά καλύτερη απόδοση από τα B0 – B6.

Παρακάτω, στο σχήμα 6.2 παρατίθεται ενδεικτικά ένα patch από τις αρχικές εικόνες του test set του ήπατος, διαστάσεων 96x96. Η μεσαία εικόνα αποτελεί την μάσκα τμηματοποίησης (ground truth label) αυτού του patch, όπου με κίτρινο χρώμα αναπαρίσταται το ήπαρ και με κόκκινο χρώμα ο όγκος. Δεξιά, στην τελευταία εικόνα, αναπαρίσταται η πρόβλεψη τμηματοποίησης ήπατος και όγκου από το μοντέλο B6. Παρατηρείται ότι το μοντέλο έχει προβλέψει τόσο το όργανο και τον όγκο, ωστόσο υπάρχουν αρκετές περιοχές που λανθασμένα προβλέπει ως όργανο και ως όγκο.



Εικόνα 6.2 Παράδειγμα τμηματοποίησης του μοντέλου B6. Η πρώτη εικόνα αποτελεί patch διαστάσεων 96x96 από την αρχική εικόνα. Η μεσαία εικόνα είναι η μάσκα τμηματοποίησης αυτού του patch. Η δεξιά εικόνα αποτελεί την πρόβλεψη για τμηματοποίηση από το μοντέλο.

3^ο σύνολο δεδομένων – Ηπατικά αγγεία και ηπατικοί όγκοι

Για την κλιμάκωση του βασικού δικτύου, τα δίκτυα εκπαιδεύονται μέχρι 200 εποχές λόγω του μεγαλύτερου μεγέθους τους από το βασικό, με ρυθμό μάθησης 0.001 ο οποίος μειώνεται σταδιακά σύμφωνα με μια συνημιτονική συνάρτηση μέχρι 0.00001 ώστε το μοντέλο να έχει πιο σταθερή ακρίβεια αποτελεσμάτων. Επίσης, και εδώ χρησιμοποιείται dropout λόγω της προοδευτικής αύξησης του βάθους και της πολυπλοκότητας των μοντέλων, για την αποφυγή της υπερπροσαρμογής τους στα δεδομένα εκπαίδευσης. Στη συγκεκριμένη περίπτωση, το δίκτυο B0 κλιμακώνεται ως το B4 και όχι ως το B6 όπως στις προηγούμενες περιπτώσεις, λόγω της επιλογής των συντελεστών α , β , γ και του φ , καθώς τα 2 εναπομείναντα δίκτυα θα ήταν ίδια με προηγούμενα.

Παράλληλα, με σκοπό τη σύγκριση των αποτελεσμάτων εκπαιδεύεται και το U-Net με τη χρήση των default παραμέτρων που αναφέρθηκαν στο κεφάλαιο 4.8. Το δίκτυο εκπαιδεύτηκε για 200 εποχές, με ρυθμό μάθησης 0.001 ο οποίος μειώνεται σταδιακά σύμφωνα με συνημιτονική συνάρτηση έως τιμή 0.00001 για 200 εποχές. Το αποτέλεσμα φαίνεται στην τελευταία γραμμή των ακόλουθων πινάκων.

Μοντέλο	Mean Dice Score		
	Ηπατικά αγγεία	Όγκος	Μέσος όρος
B0	0.4867	0.4367	0.4617
B1	0.5545	0.5966	0.5756
B2	0.5589	0.5426	0.5508
B3	0.5145	0.538	0.5263
B4	0.5504	0.4902	0.5203
Default U-Net	0.575	0.5511	0.5631

Πίνακας 8 Αποτελέσματα mean Dice score για το σύνολο δεδομένων ηπατικών αγγείων και ηπατικών όγκων.

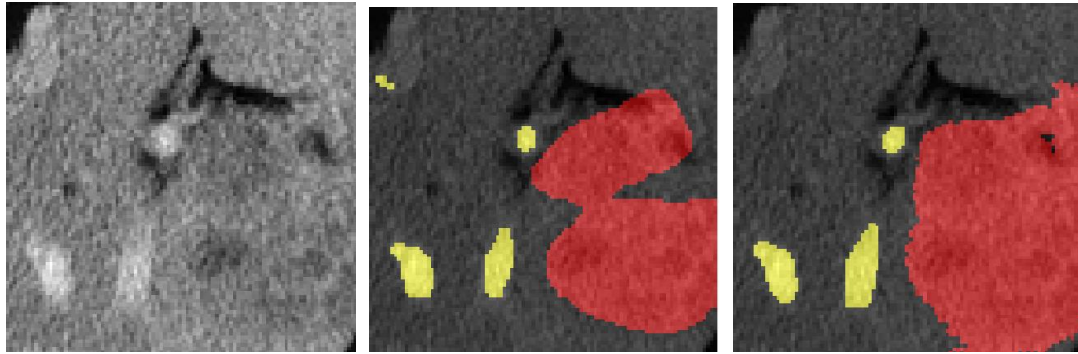
Από τα παραπάνω αποτελέσματα παρατηρείται ότι με την κλιμάκωση των δικτύων η μετρική mean Dice βελτιώνεται. Συγκεκριμένα το δίκτυο B1 επιτυγχάνει την καλύτερη απόδοση τόσο στην αναγνώριση ηπατικών αγγείων όσο και στην αναγνώριση όγκων μεταξύ των μοντέλων B1 – B6. Ωστόσο, παρατηρείται ότι η απόδοση είναι αρκετά κοντά με την απόδοση του U-Net με τις default παραμέτρους.

Μοντέλο	IoU	Recall	Precision
B0	34%	59,3%	67%
B1	43%	81,6%	77,1%
B2	40,4%	71,6%	75,3%
B3	37,7%	76,5%	70,7%
B4	38,9%	67%	67,2%
Default U-Net	42,6%	69,6%	76,5%

Πίνακας 9 Αποτελέσματα μετρικών IoU, Recall, Precision για το σύνολο δεδομένων των ηπατικών αγγείων και ηπατικών όγκων.

Στον παραπάνω πίνακα φαίνονται και τα αποτελέσματα από τις υπόλοιπες μετρικές για τα κλιμακούμενα δίκτυα αλλά και το U-Net με τις default παραμέτρους για το σύνολο δεδομένων των ηπατικών αγγείων και όγκων. Παρατηρείται ότι το μοντέλο B1 έχει την καλύτερη απόδοση σε όλες τις μετρικές, με μεγάλη αύξηση του Recall συγκριτικά με το B0.

Παρακάτω, στο σχήμα 6.3 παρατίθεται ενδεικτικά ένα patch από τις αρχικές εικόνες του test set του ήπατος, διαστάσεων 96x96. Η μεσαία εικόνα αποτελεί την μάσκα τμηματοποίησης (ground truth label) αυτού του patch, όπου με κίτρινο χρώμα αναπαρίστανται τα ηπατικά αγγεία και με κόκκινο χρώμα ο ηπατικός όγκος. Δεξιά, στην τελευταία εικόνα, αναπαρίσταται η πρόβλεψη τμηματοποίησης ηπατικών αγγείων και όγκου από το μοντέλο B1. Παρατηρείται ότι το μοντέλο προέβλεψε με μεγάλη ακρίβεια τα ηπατικά αγγεία, ωστόσο δεν έχει μεγάλη ακρίβεια στην οριοθέτηση του όγκου.



Εικόνα 6.3 Παράδειγμα τμηματοποίησης του μοντέλου B1. Η πρώτη εικόνα αποτελεί patch διαστάσεων 96x96 από την αρχική εικόνα. Η μεσαία εικόνα είναι η μάσκα τμηματοποίησης αυτού του patch. Η δεξιά εικόνα αποτελεί την πρόβλεψη για τμηματοποίηση από το μοντέλο.

7. Συζήτηση

Όλα τα πειράματα ακολουθούν τη μέθοδο της σύνθετης κλιμάκωσης (compound scaling) του EfficientNet, η οποία περιλαμβάνει την ταυτόχρονη αύξηση του βάθους, του πλάτους και της ανάλυσης εικόνων. Όλα τα μοντέλα, από το U-Net B0 έως το U-Net B6, αξιοποιούν αυτή την τεχνική.

Αρχικά, παρατηρείται ότι και για τα τρία σύνολα δεδομένων που χρησιμοποιήθηκαν, η κλιμάκωση οδήγησε σε βελτίωση των μετρικών Dice score, IoU, Recall, Precision. Συγκεκριμένα, για όλα τα σύνολα δεδομένων, υπήρξε σημαντική βελτίωση στις μετρικές στο μοντέλο B1, σε σχέση με το B0. Με βάση τον συντελεστή ομοιότητας Dice στα σύνολα δεδομένων του παγκρέατος και των ηπατικών αγγείων, το B1 είναι το μοντέλο με την καλύτερη απόδοση τόσο στην αναγνώριση οργάνου όσο και στην αναγνώριση των όγκων. Όσο τα μοντέλα γίνονται πολυπλοκότερα, δεν παρουσιάζεται ιδιαίτερη βελτίωση στον συντελεστή ομοιότητας, καθώς και στις υπόλοιπες μετρικές. Το γεγονός αυτό μπορεί να οφείλεται σε πιθανή υπερπροσαρμογή των βαθύτερων δικτύων στα σύνολα δεδομένων που χρησιμοποιούνται, με αποτέλεσμα να μην έχουν την ικανότητα να γενικεύσουν σωστά σε νέα δεδομένα. Ωστόσο, για το σύνολο δεδομένων του ήπατος και των ηπατικών όγκων, την καλύτερη απόδοση έχει το μοντέλο B6, με τα υπόλοιπα δίκτυα B1- B5 να έχουν παρόμοια αποτελέσματα.

Παράλληλα, όσον αφορά και τις υπόλοιπες μετρικές, στο σύνολο δεδομένων του παγκρέατος τα μοντέλα παρουσιάζουν χαμηλή μετρική IoU, η οποία βελτιώνεται με την κλιμάκωση. Η χαμηλή τιμή IoU για το συγκεκριμένο σύνολο δεδομένων δικαιολογείται καθώς το πάγκρεας είναι ένα δύσκολο όργανο για να το αναγνωρίσει ένα νευρωνικό δίκτυο λόγω του μεγέθους και της ιδιομορφίας του, όπως και οι όγκοι μέσα σε αυτό. Οι τιμές Recall έχουν κατά μέσο όρο τιμή 77,35% ενώ οι τιμές Precision 65,55% για τα δίκτυα B1 – B6 ενώ το B0 έχει αντίστοιχα τιμές 59,9% και 63,4% για Recall και Precision. Αυτό δείχνει ότι τα μοντέλα B1 – B6 έχουν κατά μέσο όρο καλύτερη ικανότητα να ανιχνεύουν πραγματικές θετικές περιπτώσεις (όργανο και όγκους) συγκριτικά με το βασικό δίκτυο B0, δηλαδή έχουν υψηλότερη ευαισθησία. Ωστόσο, αντιμετωπίζουν μεγαλύτερη δυσκολία στο να μειώσουν τις ψευδώς θετικές αναγνωρίσεις καθώς η βελτίωση στην ακρίβεια (Precision) σε σχέση με το B0 είναι μικρότερη.

Για το σύνολο δεδομένων του ήπατος και των ηπατικών όγκων, παρατηρούνται καλύτερες τιμές μετρικών από αυτές του συνόλου δεδομένων παγκρέατος. Αυτό δικαιολογείται καθώς το ήπαρ και οι ηπατικοί όγκοι είναι πιο εύκολα αναγνωρίσιμοι καθώς έχουν μεγαλύτερο μέγεθος. Η μετρική IoU έχει κατά μέσο όρο τιμή 65,38%, η μετρική Recall έχει κατά μέσο όρο τιμή 92,43% και η μετρική Precision 91,03% για τα μοντέλα B1 – B6. Αντίθετα το B0 έχει αρκετά μικρότερες τιμές, 57,7% IoU, 86,6% Recall και 83% Precision. Αυτό δείχνει ότι τα μοντέλα B1 – B6 δεν αναγνωρίζουν με μεγάλη ακρίβεια τις περιοχές ενδιαφέροντος, λόγω του χαμηλού IoU, ωστόσο οι υψηλές τιμές Recall και Precision υποδεικνύουν ότι τα μοντέλα μπορούν να εντοπίσουν τις περισσότερες πραγματικές περιοχές ενδιαφέροντος (υψηλό Recall), αλλά και να κάνουν ακριβείς προβλέψεις όσον αφορά την αναγνώριση των σωστών περιοχών ως όγκους ή όργανα (υψηλό Precision). Παρά το χαμηλό IoU, που υποδεικνύει ότι οι προβλέψεις των ορίων δεν επικαλύπτονται πλήρως με τις πραγματικές περιοχές, τα αποτελέσματα δείχνουν πως τα μοντέλα καταφέρνουν να εντοπίζουν πιο σωστά τις περισσότερες περιοχές ενδιαφέροντος, συγκριτικά με το βασικό δίκτυο B0.

Για το σύνολο δεδομένων των ηπατικών αγγείων και ηπατικών όγκων, παρατηρείται μέση τιμή IoU 40%, η οποία είναι αρκετά χαμηλή. Δικαιολογείται και σε αυτή τη περίπτωση καθώς τα μοντέλα καλούνται να αναγνωρίσουν πολύ μικρές περιοχές, όπως είναι τα ηπατικά αγγεία κι οι ηπατικοί όγκοι. Επιπλέον, η μέση τιμή ευαισθησίας είναι 74,18% και η μέση τιμή της μετρικής Precision είναι 72,58% για τα δίκτυα B1 – B6 συγκριτικά με 59,3% και 67% αντίστοιχα για το μοντέλο B0. Αυτό δείχνει ότι, παρόλο που τα μοντέλα καταφέρνουν να αναγνωρίσουν αρκετές από τις περιοχές ενδιαφέροντος (όπως φαίνεται από τη μέση τιμή ευαισθησίας), η ακρίβεια με την οποία αναγνωρίζουν αυτές τις περιοχές είναι επίσης

περιορισμένη, όπως υποδηλώνει η τιμή της μετρικής Precision. Το γεγονός ότι οι τιμές του Recall είναι υψηλότερες από τις τιμές της μετρικής Precision υποδεικνύει ότι τα μοντέλα τείνουν να κάνουν περισσότερες ψευδώς θετικές προβλέψεις, δηλαδή να αναγνωρίζουν περιοχές ως αγγεία ή όγκους που στην πραγματικότητα δεν είναι.

Επίσης, για την περαιτέρω αξιολόγηση της μεθόδου κλιμάκωσης, το U-Net με τις παραμέτρους που έχουν αναφερθεί στο κεφάλαιο 4.8 εκπαιδεύτηκε και στα τρία σύνολα δεδομένων. Και για τα τρία σύνολα δεδομένων, το default U-Net παρουσίασε Dice score οργάνου υψηλότερο από το αντίστοιχο κλιμακούμενο δίκτυο με την καλύτερη απόδοση, και ταυτόχρονα μικρότερο Dice όγκων. Όσον αφορά τις υπόλοιπες μετρικές, το δίκτυο B1 παρουσίασε καλύτερη απόδοση από το U-Net για το σύνολο δεδομένων του παγκρέατος και των ηπατικών αγγείων και όγκων, ωστόσο για το σύνολο δεδομένων του ήπατος, το U-Net είχε καλύτερη απόδοση από το αντίστοιχο καλύτερο κλιμακούμενο δίκτυο (B6). Προκύπτει ότι στην περίπτωση των δύο συνόλων δεδομένων η κλιμάκωση επέφερε καλύτερα αποτελέσματα, ωστόσο στο τρίτο σύνολο δεδομένων αυτό δεν ισχύει.

Συνοψίζοντας, οι τιμές των μετρικών με την κλιμάκωση βελτιώνονται από το δίκτυο B0 έως το δίκτυο B6. Ωστόσο δεν βελτιώνονται συνέχεια όσο τα δίκτυα κλιμακώνονται, καθώς στα δύο από τα τρία σύνολα δεδομένων το πρώτο κλιμακούμενο δίκτυο (B1) σημείωσε τη μεγαλύτερη βελτίωση, ενώ στο τρίτο σύνολο δεδομένων (ήπαρ) τα B1 – B5 παρουσίασαν παρόμοιες τιμές, με το B6 να σημειώνει τη καλύτερη γενικότερη απόδοση. Συγκρίνοντας και με το U-Net με τις default παραμέτρους, στα δύο από τα τρία σύνολα δεδομένων η κλιμάκωση οδήγησε σε καλύτερα αποτελέσματα, με τη χρήση πολύ λιγότερων παραμέτρων. Συγκριτικά με τα αποτελέσματα του EfficientNet, δεν παρατηρούμε παρόμοια συμπεριφορά, καθώς στο EfficientNet οι μετρικές αυξάνονταν συστηματικά καθώς τα δίκτυα γίνονταν πιο μεγάλα και πολύπλοκα, ενώ στην παρούσα εργασία αυτό δεν ισχύει. Η κλιμάκωση των δικτύων δεν οδήγησε σε συνεχή βελτίωση των μετρικών. Αυτό μπορεί να οφείλεται σε περιορισμούς που υπήρχαν στην παρούσα εργασία, και οι οποίοι θα αναλυθούν στην συνέχεια

Περιορισμοί

Στην διαδικασία εφαρμογής της μεθόδου κλιμάκωσης του EfficientNet στο U-Net στην παρούσα εργασία, υπήρχαν ορισμένοι περιορισμοί. Αρχικά, τα σύνολα δεδομένων περιέχουν περιορισμένο αριθμό δειγμάτων, ο οποίος είναι πιθανό να οδήγησε τα βαθύτερα δίκτυα σε υπερπροσαρμογή. Επιπλέον, λόγω περιορισμένης υπολογιστικής ισχύος όλα τα δίκτυα εκπαιδεύθηκαν με batch size 8, το οποίο μπορεί να επηρεάσει την σταθερότητα και την απόδοση των δικτύων. Λόγω των περιορισμένων υπολογιστικών πόρων, επίσης, ο συντελεστής ϕ που χρησιμοποιείται για την κλιμάκωση των δικτύων παίρνει τιμές στο διάστημα $[0.5, 3]$ με βήμα 0.5, γεγονός το οποίο δεν επιτρέπει πολύ μεγάλη κλιμάκωση των δικτύων, για την αξιολόγηση βαθύτερων δικτύων. Με γνώμονα την περιορισμένη υπολογιστική ισχύ επιλέγονται και οι συντελεστές α , β , γ . Ακόμα, λόγω περιορισμών στη διαθέσιμη μνήμη, οι εικόνες δεν χρησιμοποιήθηκαν στις αρχικές τους διαστάσεις. Αντ' αυτού, χρησιμοποιήθηκαν patches διαστάσεων $96 \times 96 \times 96$ από τις αρχικές εικόνες για την εκπαίδευση των μοντέλων. Αυτό το μικρό μέγεθος των patches ενδέχεται να επηρεάσει αρνητικά την απόδοση, ιδιαίτερα όταν πρόκειται για την αναγνώριση μεγαλύτερων ή πιο περίπλοκων δομών, όπως όργανα ή όγκοι.

Μελλοντικές προεκτάσεις

Για την περαιτέρω βελτίωση της μεθόδου, μπορεί να υλοποιηθεί μια παραμετροποιημένη εκδοχή του U-Net, χρησιμοποιώντας την κλασική αρχιτεκτονική με συνελκτικά επίπεδα αντί για τη χρήση του U-Net από το MONAI, το οποίο διαθέτει residual units, και πάνω σε αυτό να

βασιστεί η αναζήτηση του καλύτερου βασικού μοντέλου (B0). Στη συνέχεια, μπορεί να εφαρμοστεί η μέθοδος κλιμάκωσης για την περαιτέρω βελτίωση των μετρικών. Αυτή η προσέγγιση θα μπορούσε να επιτρέψει μεγαλύτερο έλεγχο στις παραμέτρους του δικτύου και στην κλιμάκωση τους. Επιπλέον, η χρήση μεγαλύτερου αριθμού batch σε συνδυασμό με patches μεγαλύτερων διαστάσεων θα βοηθούσε στην καλύτερη και πιο σταθερή απόδοση των δικτύων. Τέλος, η χρήση μεγαλύτερων συνόλων δεδομένων ή και περισσότερων τεχνικών επαύξησης τους, θα βοηθούσε στην αποφυγή της υπερπροσαρμογής των δικτύων και ως αποτέλεσμα στην ικανότητα γενίκευσης τους σε νέα δεδομένα.

8. Συμπεράσματα

Στην παρούσα εργασία εφαρμόστηκε η μεθοδολογία συστηματικής κλιμάκωσης compound scaling του EfficientNet στην αρχιτεκτονική U-Net. Τα μοντέλα U-Net εκπαιδεύτηκαν σε τρία σύνολα δεδομένων που περιλάμβαναν αξονικές τομογραφίες με όργανα και καρκινικούς όγκους. Επιπλέον, για σύγκριση, το U-Net με τις αρχικές παραμέτρους, όπως ειπώθηκαν στο κεφάλαιο 4.8, εκπαιδεύτηκε και αυτό στα ίδια τρία σύνολα δεδομένων.

Η κλιμάκωση είχε ως αποτέλεσμα την βελτίωση των μετρικών, ειδικά στα πρώτα επίπεδα κλιμάκωσης. Συγκεκριμένα, σε δύο από τα τρία σύνολα δεδομένων το πρώτο κλιμακούμενο δίκτυο, B1, είχε την καλύτερη απόδοση. Στο τρίτο σύνολο δεδομένων, το τελευταίο δίκτυο, B6 είχε την καλύτερη απόδοση. Τα αποτελέσματα δείχνουν ότι η κλιμάκωση μπορεί να βελτιώσει τις μετρικές, ωστόσο η αύξηση του βάθους των δικτύων μπορεί να οδηγήσει σε κορεσμό ή πτώση της απόδοσης τους. Αυτό μπορεί να οφείλεται σε πολλές παραμέτρους που έχουν ήδη αναφερθεί, όπως το μικρό μέγεθος των συνόλων δεδομένων που μπορεί να οδηγήσει τα δίκτυα σε υπερπροσαρμογή.

Επιπλέον, η σύγκριση με το κλασικό U-Net με τις default παραμέτρους με βάση το Dice score δείχνει ότι έχει συγκριτικά καλύτερη απόδοση στην κατάτμηση των οργάνων, αλλά υστερεί στην αναγνώριση των όγκων. Όσον αφορά τις υπόλοιπες μετρικές, δηλαδή IoU, Recall, Precision το δίκτυο B1 παρουσίασε καλύτερη απόδοση από το U-Net για το σύνολο δεδομένων του παγκρέατος και των ηπατικών αγγείων και όγκων, ωστόσο για το σύνολο δεδομένων του ήπατος, το U-Net είχε καλύτερη απόδοση από το αντίστοιχο καλύτερο κλιμακούμενο δίκτυο (B6).

Συνοψίζοντας, τα αποτελέσματα της παρούσας εργασίας επιβεβαιώνουν ότι η μέθοδος compound scaling του EfficientNet μπορεί να οδηγήσει σε σημαντικές βελτιώσεις στην απόδοση των δικτύων εκπαιδευμένα σε τρισδιάστατα δεδομένα αξονικών τομογραφιών. Ωστόσο, η συνεχής αύξηση του μεγέθους και της πολυπλοκότητας των μοντέλων δεν εγγυάται πάντοτε βελτιώσεις, καθώς παρατηρήθηκε ότι με την αύξηση της πολυπλοκότητας η απόδοση σταθεροποιείται ή μπορεί ακόμα και να μειωθεί. Το γεγονός αυτό τονίζει τη σημασία της σωστής ισορροπίας μεταξύ της πολυπλοκότητας του μοντέλου και του μεγέθους του συνόλου δεδομένων για την αποφυγή φαινομένων υπερπροσαρμογής.

Βιβλιογραφία

- [1] I. El, N. R. Li, and M. J. Murphy, “Theory and Applications Machine Learning in Radiation Oncology.”
- [2] X. Jiang, Z. Hu, S. Wang, and Y. Zhang, “Deep Learning for Medical Image-Based Cancer Diagnosis,” Jul. 01, 2023, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/cancers15143608.
- [3] “<https://www.lindy.ai/tools/ai-healthcare-assistant>.” Accessed on 20/08/2024.
- [4] T. B. Murdoch and A. S. Detsky, “The Inevitable Application of Big Data to Health Care.” [Online]. Available: <http://clarabridge.com/default.aspx?tabid=137&ModuleID=635>
- [5] A. Maier, S. Steidl, V. Christlein, and J. Hornegger, Eds., *Medical Imaging Systems*, vol. 11111. in *Lecture Notes in Computer Science*, vol. 11111. Cham: Springer International Publishing, 2018. doi: 10.1007/978-3-319-96520-8.
- [6] “The history of computed tomography at Siemens Healthineers.”
- [7] H. Azhari, J. A. Kennedy, N. Weiss, and L. Volokh, “From Signals to Image A Basic Course on Medical Imaging for Engineers.”
- [8] R. K. Buddington and V. Kuz'mina, “Digestive System,” 2000.
- [9] R. L. Siegel, K. D. Miller, and A. Jemal, “Cancer statistics, 2020,” *CA Cancer J Clin*, vol. 70, no. 1, pp. 7–30, Jan. 2020, doi: 10.3322/caac.21590.
- [10] J. D. Mizrahi, R. Surana, J. W. Valle, and R. T. Shroff, “Pancreatic cancer,” 2008. [Online]. Available: www.thelancet.com
- [11] L. Zhang, S. Sanagapalli, and A. Stoita, “Challenges in diagnosis of pancreatic cancer,” May 21, 2018, *Baishideng Publishing Group Co.* doi: 10.3748/wjg.v24.i19.2047.
- [12] S. Tang *et al.*, “Usefulness of 18F-FDG PET, combined FDG-PET/CT and EUS in diagnosing primary pancreatic carcinoma: A meta-analysis,” *Eur J Radiol*, vol. 78, no. 1, pp. 142–150, Apr. 2011, doi: 10.1016/j.ejrad.2009.09.026.
- [13] D. Anwanwan, S. Kumar Singh, S. Singh, V. Saikam, and R. Singh, “Challenges in liver cancer and possible treatment approaches,” 2019.
- [14] C. Y. Liu, K. F. Chen, and P. J. Chen, “Treatment of liver cancer,” *Cold Spring Harb Perspect Med*, vol. 5, no. 9, Sep. 2015, doi: 10.1101/cshperspect.a021535.
- [15] H. Fukuhara, Y. Ino, and T. Todo, “Oncolytic virus therapy: A new era of cancer treatment at dawn,” Oct. 01, 2016, *Blackwell Publishing Ltd.* doi: 10.1111/cas.13027.

- [16] V. Kaul, S. Enslin, and S. A. Gross, "History of artificial intelligence in medicine," Oct. 01, 2020, *Mosby Inc.* doi: 10.1016/j.gie.2020.06.040.
- [17] R. C. Deo, "Machine learning in medicine," *Circulation*, vol. 132, no. 20, pp. 1920–1930, Nov. 2015, doi: 10.1161/CIRCULATIONAHA.115.001593.
- [18] T. Haberle *et al.*, "The impact of nuance DAX ambient listening AI documentation: a cohort study," *Journal of the American Medical Informatics Association*, vol. 31, no. 4, pp. 975–979, Apr. 2024, doi: 10.1093/jamia/ocae022.
- [19] "https://www.suki.ai/." Accessed on 1/09/2024.
- [20] L.-C. Chang *et al.*, "Machine learning approaches to identify Parkinson's disease using voice signal features."
- [21] F. Napolitano *et al.*, "Coronavirus diagnosis using cough sounds: Artificial intelligence approaches," 2019.
- [22] "https://www.aidoc.com/solutions/radiology/." Accessed on 10/09/2024.
- [23] E. Zavaleta-Monestel *et al.*, "Revolutionizing Healthcare: Qure.AI's Innovations in Medical Diagnosis and Treatment," *Cureus*, Jun. 2024, doi: 10.7759/cureus.61585.
- [24] N. Tomašev *et al.*, "A clinically applicable approach to continuous prediction of future acute kidney injury," *Nature*, vol. 572, no. 7767, pp. 116–119, Aug. 2019, doi: 10.1038/s41586-019-1390-1.
- [25] R. T. Sutton, D. Pincock, D. C. Baumgart, D. C. Sadowski, R. N. Fedorak, and K. I. Kroeker, "An overview of clinical decision support systems: benefits, risks, and strategies for success," Dec. 01, 2020, *Nature Research*. doi: 10.1038/s41746-020-0221-y.
- [26] Z. Jie, Z. Zhiying, and L. Li, "A meta-analysis of Watson for Oncology in clinical application," *Sci Rep*, vol. 11, no. 1, Dec. 2021, doi: 10.1038/s41598-021-84973-5.
- [27] X. Glorot, A. Bordes, and Y. Bengio, "Deep Sparse Rectifier Neural Networks," 2011.
- [28] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," May 27, 2015, *Nature Publishing Group*. doi: 10.1038/nature14539.
- [29] N. Ketkar and J. Moolayil, *Deep learning with python: Learn Best Practices of Deep Learning Models with PyTorch*. Apress Media LLC, 2021. doi: 10.1007/978-1-4842-5364-9.
- [30] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation," Jun. 2016, [Online]. Available: <http://arxiv.org/abs/1606.04797>
- [31] R. Azad *et al.*, "Loss Functions in the Era of Semantic Segmentation: A Survey and Outlook," Dec. 2023, [Online]. Available: <http://arxiv.org/abs/2312.05391>
- [32] J. Kacprzyk, · Mostafa, E. · Valentina, and E. Balas, "International Conference on Advanced Intelligent Systems for Sustainable Development Volume 1-Advanced Intelligent Systems on Artificial Intelligence, Software, and Data Science Lecture Notes in Networks and Systems 637."

- [33] T. J. Sejnowski, "The unreasonable effectiveness of deep learning in artificial intelligence," *Proc Natl Acad Sci U S A*, vol. 117, no. 48, pp. 30033–30038, Dec. 2020, doi: 10.1073/pnas.1907373117.
- [34] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," May 2015, [Online]. Available: <http://arxiv.org/abs/1505.04597>
- [35] O. Oktay *et al.*, "Attention U-Net: Learning Where to Look for the Pancreas," Apr. 2018, [Online]. Available: <http://arxiv.org/abs/1804.03999>
- [36] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition." [Online]. Available: <http://image-net.org/challenges/LSVRC/2015/>
- [37] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," May 2019, [Online]. Available: <http://arxiv.org/abs/1905.11946>
- [38] M. Tan *et al.*, "MnasNet: Platform-Aware Neural Architecture Search for Mobile," Jul. 2018, [Online]. Available: <http://arxiv.org/abs/1807.11626>
- [39] Cardoso, M. J., Li, W., Brown, et al. (2022). MONAI: An open-source framework for deep learning in healthcare. <https://doi.org/https://doi.org/10.48550/arXiv.2211.02701>
- [40] K. G. Kanagachidambaresann, "EAI/Springer Innovations in Communication and Computing Programming with TensorFlow Solution for Edge Computing Applications." [Online]. Available: <http://www.springer.com/series/15427>
- [41] I. M. Baltruschat *et al.*, "Scaling the U-net: segmentation of biodegradable bone implants in high-resolution synchrotron radiation microtomograms," *Sci Rep*, vol. 11, no. 1, Dec. 2021, doi: 10.1038/s41598-021-03542-y.
- [42] I. O. Hmyria and N. S. Kravets, "CONVOLUTIONAL NEURAL NETWORK SCALING METHODS IN SEMANTIC SEGMENTATION," *Radio Electronics, Computer Science, Control*, no. 2, p. 52, Jun. 2024, doi: 10.15588/1607-3274-2024-2-6.
- [43] M. Antonelli *et al.*, "The Medical Segmentation Decathlon," *Nat Commun*, vol. 13, no. 1, Dec. 2022, doi: 10.1038/s41467-022-30695-9.
- [44] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation," *Nat Methods*, vol. 18, no. 2, pp. 203–211, Feb. 2021, doi: 10.1038/s41592-020-01008-z.
- [45] "<https://github.com/facebook/Ax>." Accessed on 15/09/2024.