Εθνικο Μετσοβιο Πολυτεχνειο
Σχολη Ηλεκτρολογων Μηχανικων και Μηχανικων Υπολογιστων
ΔΠΜΣ Επιστημη Δεδομενων και μηχανικη μαθηση

Εντοπισμός Καρκινικών Όγκων με Αλγορίθμους Βαθιάς Μάθησης

Σύγκριση Συνελικτικών Νευρωνικών Δικτύων και Εμπλουτισμός τους

Διπλωματικη Εργασια

του

Γεωργίου Σ. Παπαδάκη

Επιβλέπων: Γεώργιος Ματσόπουλος
Καθηγητής ΕΜΠ ΣΗΜΜΥ

Αθήνα, Ιούλιος 2023

Εθνικο Μετσοβιο Πολυτεχνειο
Σχολη Ηλεκτρολογων Μηχανικων και Μηχανικων Υπολογιστων
ΔΠΜΣ Επιστημη Δεδομενων και μηχανικη μαθηση

Εντοπισμός Καρκινικών Όγκων με Αλγορίθμους Βαθιάς Μάθησης

Σύγκριση Συνελικτικών Νευρωνικών Δικτύων και Εμπλουτισμός τους

Διπλωματικη Εργασια

του

Γεωργίου Σ. Παπαδάκη

Επιβλέπων: Γεώργιος Ματσόπουλος
Καθηγητής ΕΜΠ ΣΗΜΜΥ

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 10 Ιουλίου 2023.

(Υπογραφή)	*(Υπογραφή)*	*(Υπογραφή)*
.....	.....	.....
Γεώργιος Ματσόπουλος	Γεώργιος Στάμου	Παναγιώτης Τσανάκας
Καθηγητής ΕΜΠ ΣΗΜΜΥ	Καθηγητής ΕΜΠ ΣΗΜΜΥ	Καθηγητής ΕΜΠ ΣΗΜΜΥ

Αθήνα, Ιούλιος 2023

Εθνικο Μετσοβιο Πολυτεχνειο
Σχολη Ηλεκτρολογων Μηχανικων και Μηχανικων Υπολογιστων
ΔΠΜΣ Επιστημη Δεδομενων και μηχανικη μαθηση

ΔΗΛΩΣΗ ΜΗ ΛΟΓΟΚΛΟΠΗΣ ΚΑΙ ΑΝΑΛΗΨΗΣ ΠΡΟΣΩΠΙΚΗΣ ΕΥΘΥΝΗΣ

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

(Υπογραφή)

.....
Γεώργιος Παπαδάκης

10 Ιουλίου 2023

Περίληψη

Ο καρκίνος αποτελεί μια από τις πιο θανάσιμες ασθένειες της δεκαετίας. Προβλέπεται πως θα συνεχίσει και την επόμενη δεκαετία να ανήκει στις 5 πιο θανάσιμες ασθένειες που θα μας προσβάλλουν. Για την επιτυχημένη θεραπεία του και την αποτελεσματική του αντιμετώπιση, το πρώτο και πιο καθοριστικό βήμα είναι αυτό του εντοπισμού του, μιας και οι τρόποι αφαίρεσης ενός όγκου, καθώς και της μετέπειτα παρακολούθησης του ασθενή είναι πάρα πολλοί. Ωστόσο, ο εντοπισμός των καρκινικών όγκων είναι ένα αρκετά περίπλοκο ζήτημα.

Το πρόβλημα εμφανίζεται συνήθως στην κυτταρική μορφολογία του όγκου, μιας και σε πολλές περιπτώσεις, η περιοχή στην οποία εμφανίζεται ο καρκινικός όγκος, αποτελείται από κύτταρα και ιστούς, τα οποία υπερκαλύπτουν και εν μέρει καταφέρνουν να αποκρύψουν τον όγκο, με αποτέλεσμα τον δύσκολο εντοπισμό του. Συνέπεια αυτού είναι η ραγδαία εξέλιξη του καρκινικού όγκου, η οποία οδηγεί τις περισσότερες φορές στην μετάσταση του σε περισσότερες από μια περιοχές του σώματος του ασθενή.

Σκοπός αυτής της διπλωματικής εργασίας είναι η θεωρητική θεμελίωση και η προγραμματιστική ανάπτυξη αλγορίθμων Βαθιάς Μάθησης, οι οποίοι έχουν ως σκοπό τον εντοπισμό καρκινικών όγκων, μέσα από εξετάσεις οι οποίες αποτελούνται από εικόνες ευπαθών περιοχών. Πιο συγκεκριμένα, θα μελετήσουμε δύο βασικούς επεξηγηματικούς αλγορίθμους, οι οποίοι έχουν ως στόχο της χαρτογράφηση του καρκινικού όγκου πάνω σε μια εικόνα εξέτασης. Ο πρώτος υπό μελέτη αλγόριθμος είναι ο UNET, που αποτελεί τον βασικότερο αλγόριθμο για τέτοια ζητήματα.[1] Στην συνέχεια, θα μελετηθεί μια από τις παραλλαγές του, ο αλγόριθμος UNET with Attention. Έπειτα, θα συγκρίνουμε την δομή και τα αποτελέσματα τους με σκοπό τον εντοπισμό κύριων διαφορών ανάμεσα τους. Τέλος θα δοκιμάσουμε διάφορες μορφές της Attention δομής, και θα συγκρίνουμε τα συνολικά αποτελέσματα, ώστε να αναπτυχθεί ένα μοντέλο ικανό να φέρει εις πέρας ακόμα και τις πιο απαιτητικές περιπτώσεις.

Λέξεις Κλειδιά

Βαθιά Μάθηση, Συνελικτικά Νευρωνικά Δίκτυα, UNET, UNET with Attention, Attention

Abstract

Cancer is one of the deadliest diseases of the decade. It is predicted that it will continue to belong in the next decade in the 5 most deadly diseases that will affect us. For successful treatment and its effective response, the first and most decisive step it is that of its localization, since the ways of removing a tumor, as well as of the subsequent monitoring of the patient are too many. However identifying cancerous tumors is a rather complicated issue.

The problem usually occurs in the cellular morphology of the tumor, since in many cases, the area in which the cancerous tumor appears, consists of cells and tissues, which overlay and partially manage to conceal the tumor, with the result that its infrequent identification. A consequence of this is the rapid progression of the cancerous tumor, which leads in most cases to its metastasis to more than one area of his body Patient.

This diploma thesis aims at the theoretical foundation and the programming development of Deep Learning algorithms, which aim to identify cancerous tumors, through tests that consist of images of vulnerable areas. More specifically, we will study two main ones explanatory algorithms, which aim to map the cancerous tumor on an examination picture. The first algorithm under study is the UNET, which is the most basic algorithm for such issues. [1] Next, one of its variants, the UNET with Attention algorithm, will be studied. Next, we will compare the structure and their results with a view to identifying main differences between them. Finally we will test different forms of attention structure, and compare the overall results, in order to develop a model capable of carrying out even the most demanding cases.

Keywords

Deep Learning, Convolution Neural Networks, UNET, UNET with Attention, Attention

Στους γονείς μου
και τους φίλους μου

Ευχαριστίες

Θα ήθελα καταρχήν να ευχαριστήσω τον καθηγητή κ. Ματσόπουλο για την επίβλεψη αυτής της διπλωματικής εργασίας και για την ευκαιρία που μου έδωσε να την εκπονήσω στο BIO-Medical Informatics Group. Επίσης ευχαριστώ ιδιαίτερα τον Υποψήφιο Διδάκτορα Δημήτρη Μπίνα και τον Θοδωρή Οικονομόπουλο για την καθοδήγησή τους, τις πολύτιμες συμβουλές τους και την εξαιρετική συνεργασία που είχαμε. Τέλος θα ήθελα να ευχαριστήσω τους γονείς και τους φίλους μου για την ηθική συμπαράσταση που μου προσέφεραν όλα αυτά τα χρόνια, την υπομονή που έδειξαν και τις δυνάμεις που μου έδωσαν σε αυτή την περίοδο.

Αθήνα, Ιούλιος 2023

Γεώργιος Παπαδάκης

Περιεχόμενα

Κατάλογος Σχημάτων

Κατάλογος Εικόνων

Κατάλογος Πινάκων

Κεφάλαιο 1

Εισαγωγή

Ο καρκίνος είναι ένας γενικός όρος για μια μεγάλη ομάδα ασθενειών που μπορούν να επηρεάσουν οποιοδήποτε μέρος του σώματος. Άλλοι όροι που χρησιμοποιούνται είναι κακοήθεις όγκοι και νεοπλάσματα. Ένα καθοριστικό χαρακτηριστικό του καρκίνου είναι η ταχεία δημιουργία ανώμαλων κυττάρων που αναπτύσσονται πέρα από τα συνηθισμένα όριά τους και τα οποία μπορούν στη συνέχεια να εισβάλουν σε γειτονικά μέρη του σώματος και να εξαπλωθούν σε άλλα όργανα. η τελευταία διαδικασία αναφέρεται ως μετάσταση. Οι εκτεταμένες μεταστάσεις είναι η κύρια αιτία θανάτου από καρκίνο [3],[4].

Σύμφωνα με στατιστικές μελέτες, το 2020, περίπου 20 εκατομμύρια νέες περιπτώσεις καρκίνου έλαβαν χώρα παγκοσμίως [5]. Προκάλεσε περίπου 10 εκατομμύρια θανάτους ή 15,8% του συνόλου των ανθρωπίνων θανάτων [5],[6]. Για τους άνδρες, οι πιο συνηθισμένοι τύποι καρκίνου είναι ο καρκίνος του πνεύμονα, ο καρκίνος του προστάτη, ο ορθοκολικός καρκίνος και ο καρκίνος στο στομάχι. Για τις γυναίκες, οι πιο συνηθισμένοι τύποι καρκίνου είναι ο καρκίνος του μαστού, ο ορθοκολικός καρκίνος, ο καρκίνος στον πνεύμονα και ο καρκίνος του τραχήλου της μήτρας [5]. Όσο αφορά τα παιδιά, η οξεία λεμφοβλαστική λευχαιμία και οι όγκοι στον εγκέφαλο είναι οι πιο συνηθισμένες μορφές καρκίνου, εκτός από την Αφρική, όπου εκεί το μη-Hodgkin λέμφωμα είναι πιο συχνό [5].

Ο λόγος που βασιζόμαστε σε δεδομένα του 2020 είναι διότι από το 2021 και μετά, τα συγκεκριμένα δεδομένα είναι επηρεασμένα από την πανδημία του κορονοϊού Covid-19. Η εξάπλωση αυτής της συγκεκριμένης ίωσης είχε ως αποτέλεσμα πολλοί καρκινοπαθής να καταλήξουν λόγω του ιού και όχι λόγο του καρκίνου.

1.1 Αντικείμενο της διπλωματικής

Πολλά είδη καρκίνου μπορούν να προληφθούν με διάφορους τρόπους, όπως η προσοχή στην καθημερινή διατροφή, η ήπια άθληση και η διακοπή του καπνίσματος, ωστόσο η πρόληψη αυτή δεν αποτελεί εγγύηση για την αποφυγή του [7]. Επομένως το επόμενο αποτελεσματικό βήμα το οποίο μπορεί να ληφθεί είναι ο γρήγορος εντοπισμός τέτοιων όγκων. Ένας αρχικός προληπτικός τρόπος ώστε κάποιος να μπορέσει να διαγνωστεί με την ύπαρξη πιθανών καρκινικών όγκων είναι με την χρήση αιματολογικών εξετάσεων, τις λεγόμενες Liquid Biopsies [8]. Ωστόσο για τον πλήρη εντοπισμό τους, απαιτείται στην συνέχεια η λήψη μαγνητικής τομογραφίας, μέσα από την οποία μπορεί να γίνει ο πλήρης εντοπισμός και

αναγνώριση της θέσης και των διαστάσεων του καρκινικού όγκου [5]. Αυτό είναι το πρώτο και ίσως το πιο απαιτητικό από τα στάδια της αντιμετώπιση του καρκίνου, μιας και το να θεραπευτεί στην συνέχεια είναι δυνατό με πολλές τεχνικές όπως η θεραπεία με ακτινοβολία, χειρουργείου, χημειοθεραπείας και στοχευμένης θεραπείας καθώς και με συνδυασμό των παραπάνω.

Αντικείμενο της διπλωματικής είναι η βιβλιογραφική επισκόπηση και η περαιτέρω έρευνα της εφαρμογής Συνελικτικών Νευρωνικών Δικτύων τα οποία έχουν ως στόχο τον εντοπισμό καρκινικών όγκων. Τα δεδομένα τα οποία θα επεξεργαστούν αυτά τα δίκτυα αποτελούνται από απεικονιστικές εξετάσεις ασθενών, σε συγκεκριμένες περιοχές όπου υπάρχει υποψία ύπαρξης όγκου.

Ο πλήρης εντοπισμός ωστόσο τέτοιων όγκων γενικά είναι ένα δύσκολο πρόβλημα, μιας και οι παράγοντες που επηρεάζουν τον εντοπισμό είναι πολλοί. Κατά κύριο λόγο, ο χρωματισμός των αποτελεσμάτων της εξέτασης, οι συνθήκες προετοιμασίας του ιστού που υποβάλλεται σε εξέταση καθώς και η συγκέντρωση των κυτταρικών μονάδων στην περιοχή της εξέτασης, μπορούν να επηρεάσουν τα τελικά αποτελέσματα.

Η εξέλιξη όμως της Βαθιάς Μάθησης μαζί ταυτόχρονα με την ανάπτυξη κατάλληλων επεξεργαστών που προσφέρουν γρηγορότερη επεξεργασία δεδομένων, κάνουν την λύση του προβλήματος να υλοποιείται ακόμα περισσότερο. Τα μοντέλα αυτά έχουν ως κύριο σκοπό τον πλήρη εντοπισμό της καρκινικής μάζας και όχι απλά το να αναφέρουν αν υπάρχει ή όχι μια τέτοια μάζα. Επιπροσθέτως, θα δοκιμάσουμε και θα συγκρίνουμε τέτοια μοντέλα, στα οποία θα αναλυθεί η αρχιτεκτονική τους καθώς και θα μελετηθούν οι διαφορές τους, τόσο στην δομή, όσο και στα αποτελέσματα που προσφέρουν.

1.2 Οργάνωση του τόμου

Η οργάνωση της διπλωματικής είναι η εξής:

- Θεωρητικό Υπόβαθρο: Εκεί γίνεται μια εισαγωγή στις δομικές βάσεις των Μοντέλων Βαθιάς Μηχανικής Μάθησης, όπου περιγράφουμε τα Τεχνητά Νευρωνικά Δίκτυα, το μοντέλο Neocognitron και τα Συνελικτικά Δίκτυα.

- Περιγραφή θέματος: Εκεί γίνεται η ανάλυση των Μοντέλων Βαθιάς Μηχανικής Μάθησης με τα οποία θα μελετηθούν ιατρικά δεδομένα απεικονιστικών εξετάσεων. Παρουσιάζεται η δομή τους και ο τρόπος λειτουργίας τους.

- Πρακτικό Μέρος: Εκεί γίνεται μια περαιτέρω περιγραφή των δεδομένων που διαθέτουμε, ο τρόπος εκπαίδευσης των μοντέλων, τα αποτελέσματα τους και η τελική σύγκριση τους.

- Επίλογος: Εκεί παρουσιάζουμε τα τελικά συμπεράσματα και τις μελλοντικές επεκτάσεις των μοντέλων μας.

Μέρος Ι

Θεωρητικό Μέρος

Κεφάλαιο 2

Θεωρητικό υπόβαθρο

Στο κεφάλαιο αυτό θα παρουσιάσουμε αναλυτικά τις βασικές αρχιτεκτονικές στις οποίες βασίστηκαν τα μοντέλα με τα οποία θα προσπαθήσουμε να αντιμετωπίσουμε το πρόβλημα του εντοπισμού των καρκινικών όγκων.

2.1 Τεχνητά Νευρωνικά Δίκτυα

Τα Τεχνητά Νευρωνικά Δίκτυα (ΤΝΔ) ή αλλιώς Artificial Neural Networks (ANNs) είναι υπολογιστικά μοντέλα τα οποία έχουν ως βασική λειτουργία την μίμηση της λειτουργίας των βιολογικών νευρών που διαθέτει ο άνθρωπος στον εγκέφαλό του και γενικότερα την λειτουργία του Κεντρικού Νευρικού Συστήματος(ΚΝΣ)[9]. Για αυτό τον λόγο, θα παρουσιάσουμε πρώτα την πιο βασική βιολογική μονάδα του νευρικού συστήματος, τον νευρώνα.

2.1.1 Ο βιολογικός νευρώνας

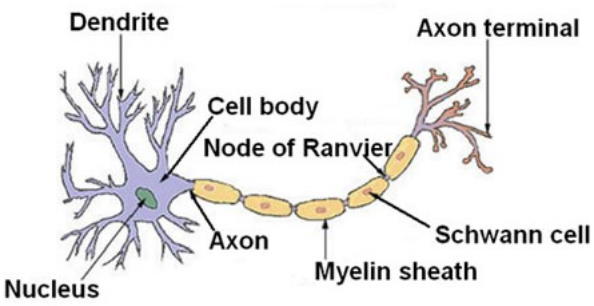
Νευρώνας ή νευρικό κύτταρο είναι το είδος των κυττάρων τα οποία αποτελούν την λειτουργική μονάδα του νευρικού συστήματος [10]. Κάθε νευρώνας αποτελείται από 3 βασικά μέρη:

- *Κυτταρικό σώμα*: Αποτελείται από τον πυρήνα του νευρικού κυττάρου καθώς και τα βασικά οργανίδια που αποτελούνται γενικά τα κύτταρα. Επιπλέον προεξέχουν από το κυτταρικό σώμα και αρκετές αποφυάδες, οι δενδρίτες. Οι δενδρίτες είναι το μέρος του σώματος που συλλέγει τις νευρικές ώσεις από τα υπόλοιπα νευρικά κύτταρα.

- *Νευράξονας*: Αποτελεί το μέρος του νευρικού κυττάρου το οποίο είναι υπεύθυνο για την μετάδοση της νευρικής ώσης από το κυτταρικό σώμα. Καλύπτεται από μυελίνη, που ως σκοπό έχει την μόνωση του νευράξονα για την καλύτερη μεταφορά της νευρικής ώσης στο εσωτερικό του κυττάρου.

- *Νευρικές απολήξεις*: Αποτελούν το τέλος του νευρικού κυττάρου και οι νευρικές ώσεις που καταλήγουν εκεί μεταδίδονται σε άλλα νευρικά κύτταρα.

Στο παρακάτω σχήμα μπορούμε να δούμε μια σχηματική του αναπαράσταση.



Εικόνα 2.1: *Η δομή ενός νευρικού κυττάρου*

2.1.2 Δομή Νευρωνικού Δικτύου

Η δομή των νευρωνικών δικτύων, όπως προαναφέρθηκε, βασίζεται στην δομή του ΚΝΣ. Ένα νευρωνικό δίκτυο αποτελείται από απλούς υπολογιστικούς κόμβους, τους νευρώνες, οι οποίοι είναι συνδεδεμένοι μεταξύ τους.

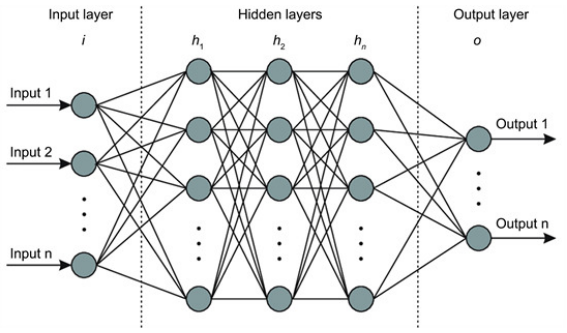
Οι νευρώνες αποτελούν τα δομικά στοιχεία του ΤΝΔ. Κάθε τέτοιος κόμβος, δέχεται ένα σύνολο αριθμητικών εισόδων, συνήθως από άλλους νευρώνες ή από το περιβάλλον του, εκτελεί έναν υπολογισμό με αυτές τις τιμές και στην συνέχεια επιστρέφει μια έξοδο [9]. Αυτές οι έξοδοι είτε κατευθύνονται στο περιβάλλον, είτε αποτελούν είσοδο για τους επόμενους νευρώνες. Υπάρχουν τρία είδη νευρώνων:

1. *Νευρώνες εισόδου*: Οι νευρώνες αυτοί δεν επιτελούν κάποιο υπολογισμό. Ο ρόλος τους είναι συνδετικός, μιας και έχουν ως σκοπό την μετάδοση πληροφοριών από το περιβάλλον προς τους υπόλοιπους νευρώνες.

2. *Υπολογιστικοί νευρώνες*: Οι νευρώνες αυτοί δέχονται τιμές από νευρώνες εισόδου ή από άλλους άλλους υπολογιστικούς νευρώνες και ο σκοπός τους είναι η εκτέλεση την υπολογιστικής ρουτίνας, την οποία θα αναλύσουμε στην συνέχεια.

3. *Νευρώνες εξόδου*: Οι νευρώνες αυτοί έχουν το αντίστροφο σκοπό από τους νευρώνες εισόδου. Αποτελούν τους νευρώνες εκείνους που διοχετεύουν στο περιβάλλον τις τελικές αριθμητικές τιμές που έχουν υπολογιστεί από τους υπολογιστικούς νευρώνες.

Με βάση τα παραπάνω είδη νευρώνων, καταλήγουμε στην τυπική δομή ενός ΤΝΔ, το οποίο αποτελείται από επίπεδα νευρώνων, όπως φαίνεται και στο παρακάτω σχήμα.



Σχήμα 2.1: *Η δομή ενός ΤΝΔ*

Το επίπεδο εισόδου αποτελείται αποκλειστικά από νευρώνες εισόδου, οι οποίοι συνδέονται και μεταφέρουν πληροφορία στους νευρώνες του επόμενου επιπέδου, του πρώτου κρυφού επιπέδου. Τα κρυφά επίπεδα αποτελούνται από υπολογιστικούς νευρώνες, οι οποίοι δεν είναι συνδεδεμένοι μεταξύ τους. Οι έξοδοι ενός κρυφού επιπέδου, γίνονται οι είσοδοι για τους νευρώνες του επόμενου κρυφού επιπέδου. Τέλος, το επίπεδο εξόδου, αποτελείται μόνο από νευρώνες εξόδου οι οποίοι δέχονται τις αριθμητικές τιμές του τελευταίου κρυφού επιπέδου[9].

Μαθηματική Δομή του ΤΝΔ

Όπως είδαμε παραπάνω, η καρδία της υπολογιστικής δύναμης ενός ΤΝΔ αποτελείται από τους υπολογιστικούς νευρώνες. Οι νευρώνες αυτοί αποτελούν το σημείο στο οποίο ένα νευρωνικό δίκτυο επεξεργάζεται την πληροφορία που δέχεται από το περιβάλλον του και εξάγει την απαραίτητη γνώση, πάλι σε μορφή πληροφορίας, πίσω σε αυτό.

Έστω x_{ki} η i-οστή είσοδος του k νευρώνα, w_{ki} το i-οστό συναπτικό βάρος του k νευρώνα και $\varphi(\cdot)$ η συνάρτηση ενεργοποίησης του νευρωνικού δικτύου. Τότε η έξοδος y_k του k νευρώνα δίνεται από την εξίσωση:

$$y_k = \varphi\left(\sum_{i=0}^{N} w_{ki} x_{ki}\right) \tag{2.1}$$

Τα δύο βασικά συστατικά της (2.1), είναι τα συναπτικά βάρη καθώς και η συνάρτηση ενεργοποίησης [9].

Συναπτικά Βάρη Νευρώνα

Το στοιχείο που δίνει στα ΤΝΔ την δυνατότητα να μπορέσουν να εκπαιδευτούν και να λύσουν τα προβλήματα που τους ζητούνται είναι τα συναπτικά βάρη κάθε νευρώνα.

Ο σκοπός του συναπτικού βάρους w_{ml} ανάμεσα σε έναν νευρώνα k σε ένα επίπεδο h_i και σε έναν νευρώνα l στο επόμενο επίπεδο h_{i+1}, είναι να δώσει το κατάλληλο βάρος στην πληροφορία που στέλνει ο πρώτος νευρώνας στον δεύτερο. Αυτή η ποσότητα είναι η προς εκπαίδευση ποσότητα σε κάθε νευρώνα, την οποία κατά την διαδικασία της εκπαίδευσης, το ΤΝΔ καλείται να ρυθμίσει και ουσιαστικά να βελτιστοποιήσει. Για να γίνει αυτό κατανοητό, θα πρέπει να αναφέρουμε το τυπικό πλαίσιο εκπαίδευσης ενός ΤΝΔ.[9]

Η εκπαίδευση ενός ΤΝΔ γίνεται με την χρήση των δεδομένων εκπαίδευσης. Αυτά τα δεδομένα έρχονται σε ζευγάρια της μορφής (x_{input}, y_{true}), όπου το διάνυσμα εισόδου είναι το x_{input} και το επιθυμητό αποτέλεσμα που θα θέλαμε το ΤΝΔ να εξάγει στο επίπεδο εξόδου είναι το y_{true}. Κατά την διάρκεια της εκπαίδευσης, για κάθε συγκεκριμένη είσοδο x_{input} στο δίκτυο μας, ζητάμε να παραχθεί η τιμή $y_{predicted}$ η οποία θέλουμε να προσεγγίζει την y_{true} όσο το περισσότερο γίνεται, για όλα τα ζευγάρια που διαθέτουμε στα δεδομένα εκπαίδευσης. Για να μπορέσουμε ωστόσο να το καταφέρουμε αυτό, όταν μια έξοδος $y_{predicted}$ δεν είναι όσο ίδια επιζητούμε με την y_{true}, αλλάζουν τα βάρη κατάλληλα ώστε το τέλος να επέλθει το επιθυμητό αποτέλεσμα [9].

Συνάρτηση Ενεργοποίησης

Το στοιχείο, που επηρεάζει δραστικά την κατάλληλη προσέγγιση των συναπτικών βαρών κατά την εκπαίδευση είναι η συνάρτηση ενεργοποίησης. Ο βασικός της ρόλος είναι να διαμορφώνει την ληφθείσα πληροφορία από τους προηγούμενους νευρώνες σε συνδυασμό με τα συναπτικά βάρη και να προσδίδει την έξοδο του νευρώνα. Όπως αναφέρει και το όνομα της, αν η ληφθείσα πληροφορία έχει αρκετά μεγάλη επιρροή, τότε η συνάρτηση ενεργοποίησης λαμβάνει κατάλληλες τιμές, ώστε να θεωρηθεί ως ενεργός ο νευρώνας στην δομή του ΤΝΔ. Αυτό σημαίνει πως ο νευρώνας επηρεάζει την διαμόρφωση της τιμής εξόδου $y_{predicted}$ του δικτύου. Αν η πληροφορία δεν είναι τόσο σημαντική όσο πρέπει, η συνάρτηση ενεργοποίησης λαμβάνει μικρές τιμές, οι οποίες κάνουν τον νευρώνα ουσιαστικά ανενεργό.[9]

Ο τρόπος με τον οποίο η συναρτήσεις ενεργοποίησης επηρεάζουν τα συναπτικά βάρη κατά την εκπαίδευση του ΤΝΔ είναι με την παραγωγισιμότητα τους. Κατά την εκπαίδευση του μοντέλου μας, η σύγκριση ανάμεσα στα $y_{predicted}$ και y_{true} μέσω μια συνάρτησης κόστους, καθορίζει και το πόσο πρέπει να γίνουν αλλαγές στα συναπτικά βάρη. Ωστόσο δεν μπορούν αυτές οι αλλαγές να γίνουν αυθαίρετα. Για αυτόν ακριβώς τον λόγο, η παράγωγος της συνάρτησης ενεργοποίησης είναι αυτή η οποία επηρεάζει το πόσο μεγάλη θα είναι η αλλαγή, σε κάθε βάρος συγκεκριμένα. Προφανώς αν η διαφορά ανάμεσα στα $y_{predicted}$ και y_{true} είναι μικρή, τότε η συνάρτηση ενεργοποίησης θα εμποδίσει την μεγάλη αλλαγή του συναπτικού βάρους, μιας και θεωρεί πως έχει προσεγγιστεί σωστά, ώστε η $y_{predicted}$ να λαμβάνει τιμές ίδιες ή σχετικά κοντά στην y_{true} [9].

Μερικές από τις πιο βασικές συναρτήσεις ενεργοποίησης είναι οι εξής:

- *Binary Step - Βηματική Συνάρτηση*: Ο τύπος της είναι ο εξής:

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

  Με βάση αυτό, βλέπουμε πως η βηματική συνάρτηση μιμείται κατά πολύ τον τρόπο λειτουργίας των βιολογικών νευρώνων, μιας και δίνει την τιμή 1 αν η είσοδός της είναι θετική και 0 αν είναι αρνητική. Ωστόσο λόγω της μορφής της, η παράγωγος της είναι πάντα 0 σε όλα τα σημεία της και έτσι δεν την εφαρμόζουμε τόσο στην δομή των ΤΝΔ.

- *Linear Function - Γραμμική Συνάρτηση*: Ο τύπος της είναι ο εξής

$$\phi(x) = x$$

Η συνάρτηση αυτή αποδίδει την τιμή του γινομένου που δέχεται και την επιστρέφει αυτούσια. Ωστόσο το μειονέκτημα αυτής της συνάρτησης ενεργοποίησης είναι ότι η παράγωγός της είναι σταθερή και ίση με την μονάδα. Αυτό έχει ως συνέπεια την ομοιόμορφη επίδραση της συνάρτησης στα στοιχεία της και κατά επέκταση στην αλλαγή των συναπτικών βαρών των νευρώνων.

- *Sigmoid Function - Συγμοειδής Συνάρτηση*: Ο τύπος της είναι ο εξής

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

Η συνάρτηση αυτή είναι ίσως η πιο γνωστή συνάρτηση ενεργοποίησης και είναι ευρέως διαδεδομένη για τα αποτελέσματα που επιφέρει στην επιτυχημένη δόμηση των ΤΝΔ. Οι τιμές της για μεγάλες θετικές τιμές τείνουν στο 1 ενώ για μεγάλες αρνητικές τιμές η τιμή της τείνει στο 0.

- *Hyperbolic tangent - Υπερβολική Εφαπτομένη*: Ο τύπος της είναι ο εξής

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

Η συνάρτηση αυτή αποτελεί μια γενίκευση της σιγμοειδούς συνάρτησης, μιας και η τιμή της για μεγάλους θετικούς αριθμούς τείνει στο 1. ενώ για μεγάλους αρνητικούς αριθμούς τείνει στο -1.

- *ReLU - Rectified Linear Unit*: Ο τύπος της είναι ο εξής

$$\phi(x) = max(0, x)$$

Η συνάρτηση αυτή είναι ίσως η πιο γνωστή συνάρτηση ενεργοποίησης στα Βαθιά Νευρωνικά Δίκτυα. Όπως φαίνεται και από τον τύπο της, λαμβάνει μόνο θετικές τιμές, και αν η είσοδος της είναι αρνητική, τότε την μηδενίζει. Ο ρόλος της είναι καθοριστικός όπως θα δούμε παρακάτω στα Συνελικτικά Νευρωνικά Δίκτυα.

Στην συνέχεια θα ασχοληθούμε με την εξέλιξη των ΤΝΔ, τα Συνελικτικά Νευρωνικά Δικτυα.

2.2 Neocognitron: Η βάση των CNNs

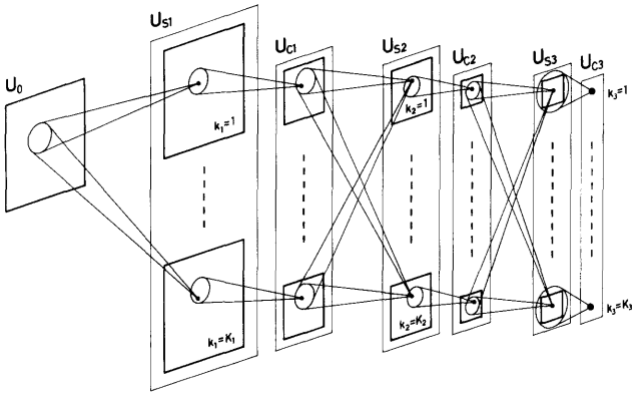
Τα Συνελικτικά Νευρωνικά Δίκτυα (ΣΝΔ) ή αλλιώς Convolutional Neural Networks (CNNs), δεν εμφανίστηκαν τυχαία ως μοντέλα. Πριν την ανάπτυξή τους, είχε προηγηθεί βαθιά και λεπτομερής μελέτη στην οποία οι επιστήμονες της δεκαετίας του '70, προσπαθούσαν να αναπαραστήσουν τις λειτουργίες του οπτικού φλοιού με διάφορα μοντέλα. Ίσως το πιο πετυχημένο το οποίο καταφέρνει να ανάπτυξη την δομή των κυττάρων του οπτικού φλοιού και των λειτουργιών του, είναι το μοντέλο Neocognitron[11].

2.2.1 Neocognitron

Η δομή του neocognitron βασίζεται στο γεγονός πως ο οπτικός φλοιός αποτελείται από 2 βασικά είδη νευρικών κυττάρων, τα Simple cells και τα Complex cells[12], [13].

Τα Simple cells είναι ένα είδος νευρικών κυττάρων τα οποία ενεργοποιούνται σε οπτικά ερεθίσματα τα οποία περιέχουν άκρες και πλέγματα. Με αυτό τον τρόπο, τα κύτταρα αυτά είναι υπεύθυνα για την μερική αντίληψη αντικειμένων στο οπτικό μας πεδίο. Τα Complex cells από την άλλη, είναι ένα είδος νευρικών κυττάρων τα οποία ενεργοποιούνται με ερεθίσματα τα οποία προέρχονται από Simple cells αλλά και από οπτικά ερεθίσματα. Λόγω αυτών των ερεθισμάτων, τα κύτταρα αυτά αποκτούν την ικανότητα να ενεργοποιούνται όταν τα αντικείμενα τα οποία λαμβάνουν βρίσκονται υπό συγκεκριμένες γωνίες και θέση, με αποτέλεσμα να αποκτούν μια χωρική διακριτικότητα [12], [13].

Με βάση τα παραπάνω κύτταρα, το Neocognitron δομείται σε μια ιεραρχική δομή, η οποία προσπαθεί να μιμηθεί την λειτουργική δομή του οπτικού φλοιού.



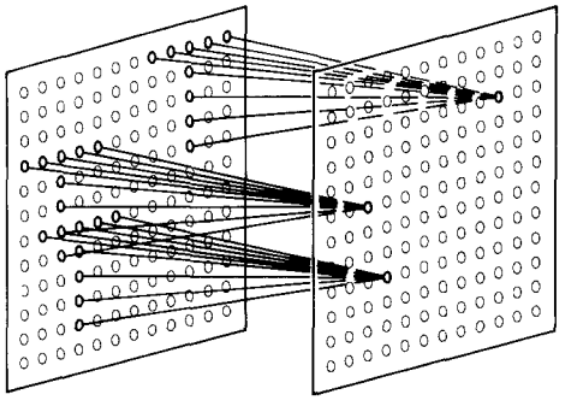
Σχήμα 2.2: *Η δομή του Neocognitron*

Όπως παρατηρούμε και στο παραπάνω σχήμα, η δομή του αποτελείται από συστάδες - ομάδες οι οποίες αποτυπώνουν την δομή Simple Cell - Complex Cell. Οι συστάδες αυτές αποτελούνται από επίπεδα (layers), το επίπεδο U_S που προσπαθεί να εξάγει χαρακτηριστικά όμοια με αυτά που εξάγει το Simple Cell και το U_C που προσπαθεί να εξάγει χαρακτηριστικά όμοια με αυτά που εξάγει το Complex Cell [11].

Κάθε επίπεδο, ανεξαρτήτως είδους, συγκροτείται από σημεία τα οποία αποτελούν την διδιάστατη αναπαράσταση των συντεταγμένων των σημείων που συμβολίζουν την θέση της εισόδου πάνω στο " οπτικό πεδίο " του κάθε κυττάρου - επιπέδου (the two-dimentional co-

ordinates representing the position of the cell's receptive field) [11]. Ο τρόπος με τον οποίο αρχικά γίνεται αυτό, είναι με την χρήση ενός επιπλέον επιπέδου στην αρχή του δικτύου, το contrast extraction layer U_G, το οποίο έχει ως ρόλο την σωστή κωδικοποίηση της εικόνας, σε σημεία του επιπέδου, έτσι ώστε να μπορεί το μοντέλο μας να επεξεργαστεί τις πληροφορίες αυτές.

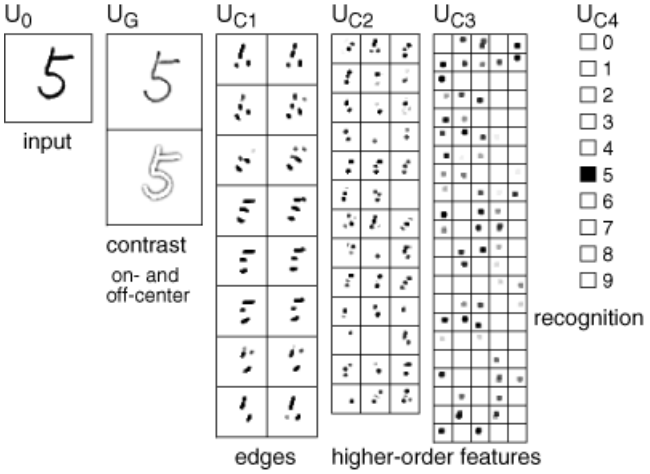
Στην συνέχεια, η πληροφορία αυτή λαμβάνεται από την επόμενη συστάδα και γίνεται μια " φιλτραρισμένη επεξεργασία ". Στην επεξεργασία αυτή, το κάθε στοιχείο ενός επιπέδου είναι συνδεδεμένο με τα στοιχεία του επόμενου επιπέδου, όπως φαίνεται στην παρακάτω εικόνα. Ο τρόπος με τον οποίο τα στοιχεία αυτά είναι συνδεδεμένα, φαίνεται να είναι ίδιος με αυτόν των ΤΝΔ, όπου κάθε νευρώνας του επιπέδου U_C είναι συνδεδεμένος με όλους τους νευρώνες του επιπέδου U_S, με την χρήση συναπτικών βαρών.



Σχήμα 2.3: *Ο τρόπος σύνδεσης των νευρώνων ενός επιπέδου U_S (αριστερά) με του επιπέδου U_C (δεξιά)*

Τέλος, ο δραστικός ρόλος των συστάδων όπως έχουμε προαναφέρει, είναι η δυνατότητά τους να εξάγουν κατάλληλα χαρακτηριστικά, ώστε στο τέλος να επιτευχθεί η ορθή αναγνώριση της αρχικής εικόνας εισόδου. Οι αρχικές συστάδες εξάγουν χαρακτηριστικά τα οποία είναι απλά στην φύση τους, όπως για παράδειγμα την θέση των ακμών ενός αντικειμένου και του περιγράμματός του. Οι επόμενες συστάδες, μέσω της επεξεργασίας τους, εξάγουν πιο αφαιρετικά στην νόηση χαρακτηριστικά, τα οποία σε χώρους μεγαλύτερης διάστασης είναι διακριτά [11]. Εν κατακλείδι, το τελευταίο επίπεδο, είναι αυτό το οποίο τελικά θα προσφέρει την απάντηση στο ερώτημα της αναγνώρισης του αντικειμένου.

Παράδειγμα 2.1. *Σε αυτό το παράδειγμα, η αρχική είσοδος αποτελείται από μια εικόνα ενός χειρόγραφου ψηφίου, του αριθμού 5. Αρχικά, στην συστάδα U_0, γίνεται η προεπεξεργασία της εικόνας ώστε να μπορεί να έρθει σε κατάλληλη μορφή για την επεξεργασία της από τις επόμενες συστάδες. Η πρώτη συστάδα U_{C1}, παρατηρούμε πως εξάγει χαρακτηριστικά τα οποία αντιστοιχούν σε χαρακτηριστικά τα οποία αντιστοιχούν σε τμήματα του ψηφίου, για αυτό λέμε πως εξάγει χαρακτηριστικά λεγόμενα ως edges. Οι επόμενες συστάδες, U_{C1}, U_{C2} εξάγουν τα χαρακτηριστικά μεγαλύτερης τάξης της εικόνας, τα higher - order feautures. Τέλος, η τελευταία συστάδα, αποδίδει την κλάση, δηλαδή το ποίο είναι το ψηφίο, όπου αποδίδει ορθά την απάντηση "5".*



Σχήμα 2.4: *Η διαδικασία αναγνώρισης του μοντέλου σε μια εικόνα χειρόγραφου ψηφίου*

2.2.2 Η ανάγκη για εξέλιξη

Η εξέλιξη της τεχνολογίας υλικού καθώς και η σχετική έρευνα σε θεωρητικό επίπεδο, έφερε την εξέλιξη του μοντέλου Neocognitron και το έθεσε σαν βάση για τα Συνελικτικά Νευρωνικά Δίκτυα.

Ωστόσο όλα αυτά συμβάδισαν μαζί με το πρώτο κύριο πρόβλημα που είχαν να αντιμετωπίσουν τα απλά μοντέλα ΤΝΔ καθώς και το Neocognitron. Πιο συγκεκριμένα, τα αρχικά αυτά μοντέλα, ενώ μπορούν να διαχειριστούν με ευκολία δεδομένα απλού τύπου, όπως σχετικά μικρά διανύσματα μερικών εκατοντάδων θέσεων, σε πιο σύνθετες μορφές δεδομένων όπως οι εικόνες, αντιμετωπίζουν αρκετά υπολογιστικά προβλήματα. Μια εξήγηση αυτού του προβλήματος μπορεί να φανεί με το παρακάτω παράδειγμα:

Παράδειγμα 2.2. *Έστω πως διαθέτουμε ένα απλό ΤΝΔ το οποίο θέλουμε να το εκπαιδεύσουμε πάνω στο γνωστό σε όλους μας, σύνολο δεδομένων χειρόγραφων ψηφίων, MNIST Dataset. Κάθε στοιχείο του συνόλου αποτελεί μια εικόνα ενός ψηφίου διάστασης* $28 \times 28 \times 1$. *Επομένως, για να αναπαρασταθεί σε μια μορφή διανύσματος, κατάλληλη για την εισαγωγή της σε ένα ΤΝΔ, θα πρέπει να ανασκευαστεί σε ένα διάνυσμα διαστάσεων* 784×1. *Αυτό σημαίνει πως κάθε νευρώνας του πρώτου επιπέδου, θα έχει 784 συναπτικά βάρη. Επομένως, αν θεωρήσουμε το πρώτο κρυφό επίπεδου του ΤΝΔ μας να αποτελείται από 10 νευρώνες, ήδη το πλήθος των παραμέτρων που πρέπει να υπολογιστούν ξεπερνούν τις 7.500. Αυτό είναι αρνητικό κομμάτι για την εκπαίδευση του ΤΝΔ, μιας και το μεγάλο πλήθος των παραμέτρων καθιστούν χρονοβόρα την εκπαίδευση και του δικτύου μας. Επομένως, η απλότητα της δομής των ΤΝΔ, δεν μπορεί να αποφέρει ικανοποιητικά αποτελέσματα σε εικόνες που είναι πιο πιθανές να βρεθούν σε ένα ρεαλιστικό πρόβλημα, όπου για παράδειγμα να έχουμε εικόνες 3 χρωματικών φίλτρων με διαστάσεις της τάξης* $1024 \times 1024 \times 3$. *Τέλος, ακόμα και αν είχαμε αρκετούς υπολογιστικούς πόρους, η χρήση ΤΝΔ με περισσότερα κρυφά επίπεδα και νευρώνες θα μπορούσε να επιτευχθεί, ωστόσο λόγω της μεγέθυνσης του μοντέλου θα είχαμε προβλήματα όπως αυτό της Υπεροπροσαρμογής (Overfitting). Αυτό σημαίνει πως το μοντέλο μας, θα εκπαιδευόταν έτσι ώστε να επιτυγχάνει υψηλά επίπεδα ακρίβειας στα στοιχεία του συνόλου εκπαίδευσης (training set), αλλά θα αποτύγχανε να δώσει σωστές προβλέψεις στα στοιχεία του συνόλου δοκιμής*

(test set) [14].

2.3 Συνελικτικά Νευρωνικά Δίκτυα - CNNs

Τα Συνελικτικά Νευρωνικά Δίκτυα (ΣΝΔ) ή αλλιώς Convolutional Neural Networks (CNNs), αποτελούν την εξέλιξη των ΤΝΔ, και πιο συγκεκριμένα βασίζεται στην δομή του Neocognitron [11].

Ο τρόπος με τον οποίο τα ΣΝΔ καταφέρνουν να ξεχωρίσουν από τα απλά ΤΝΔ είναι στην δομή του όλου μοντέλου, αλλά και στον τρόπο σύνδεσης του κάθε στοιχείου του δικτύου μεταξύ τους. Το όνομα που έχουν συνδέεται άμεσα με τον δεύτερο παράγοντα που προαναφέρθηκε. Ο λόγος που ονομάζονται ” Συνελικτικά ”, πηγάζει από το γεγονός πως το συγκεκριμένο είδος δικτύου κάνει χρήση της μαθηματικής πράξης της *Συνέλιξης*, αντί για την χρήση του κλασικού πολλαπλασιασμού πινάκων, ανάμεσα στα επίπεδα του. [15]

2.3.1 Συνέλιξη

Η συνέλιξη, είναι μια μαθηματική πράξη όπου δύο συναρτήσεις , f και g, παράγουν μια τρίτη, νέα συνάρτηση, την συνέλιξη $f * g$, η οποία περιγράφει το πως το σχήμα της συνάρτηση f επηρεάζεται από το σχήμα της συνάρτησης g. Ο τρόπος με τον οποίο ορίζεται η συνέλιξη των δύο συναρτήσεων, είναι ο εξής:

$$(f * g)(t) := \int_{-\infty}^{+\infty} f(a)g(t-a)da$$

Συνήθως, η συνάρτηση f ονομάζεται συνάρτηση εισόδου ή απλά είσοδος (input) και η συνάρτηση g ονομάζεται συνάρτηση πυρήνα ή απλά πυρήνας (kernel).

Η συνέλιξη δύο συναρτήσεων μπορεί να οριστεί και με διακριτό τρόπο, ο οποίος είναι προφανώς προτιμότερος για υπολογιστικές εφαρμογές, όπως αυτές στις οποίες χρειαζόμαστε στην Βαθιά Μηχανική Μάθηση. Ο τρόπος με τον οποίο ορίζεται η διακριτή συνέλιξη είναι ο εξής:

$$(f * g)(t) := \sum_{a=-\infty}^{+\infty} f(a)g(t-a)$$

Προφανώς ωστόσο, η συνέλιξη θέλουμε να εφαρμόζεται και σε δεδομένα περισσότερων διαστάσεων. Παραδείγματος χάρη, στην παρούσα εργασία, τα δεδομένα τα οποία διαθέτουμε είναι εικόνες, οι οποίες αναπαριστώνται με την χρήση πινάκων. Επομένως, ο τρόπος όπου ορίζεται η διακριτή διδιάστατη συνέλιξη, ανάμεσα σε μια είσοδο σε μορφή πίνακα, έστω I και ενός πυρήνα σε μορφή πίνακα, έστω K, είναι ο εξής:

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

Επιπροσθέτως, η πράξη της συνέλιξης είναι αντιμεταθετική:

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

Τέλος, οι περισσότερες βιβλιοθήκες Μηχανική Μάθησης, χρησιμοποιούν μια παραλλαγή της συνέλιξης, η οποία βασίζεται στην αντιμεταθετικότητα, την πράξη cross - correlation, η οποία στην μορφή που εφαρμόζεται ορίζεται ως εξής:

$$(K * I)(i,j) := \sum_m \sum_n I(i+m, j+n)K(m,n)$$

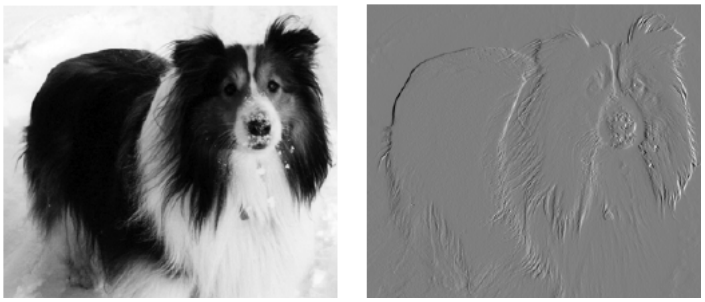
Ο λόγος που εφαρμόζεται η cross - correlation από την συνέλιξη, είναι διότι αλγοριθμικά είναι πιο εύχρηστη και ο κώδικας υλοποίησης της είναι ευκολότερο να υλοποιηθεί, λόγω της πράξης της πρόσθεσης. [15]

2.3.2 Δομικά Μέρη ΣΝΔ

Η βασική ιδέα πίσω από την αρχιτεκτονική του βασικού μοντέλο των ΣΝΔ βασίζεται στην λειτουργικότητα του μοντέλου του Neocognitron. Όπως προαναφέρθηκε, ένα Neocognitron βασίζεται στις συστάδες που το αποτελούν, οι οποίες αποτελούνται από επίπεδα που με την χρήση του κλασικού μοντέλου ΤΝΔ, επεξεργάζονται τα δεδομένα. Αυτή η επεξεργασία όπως είχαμε δει στα ΤΝΔ βασίζεται στον γραμμικό συνδυασμό των μοντέλων επί των βαρών των νευρώνων και στην συνέχεια το πέρασμα αυτού του συνδυασμού μέσα από μια συνάρτηση ενεργοποίησης. Προφανώς, το Neocognitron, επειδή δουλεύει με εισόδους δυο διαστάσεων, κάνει την χρήση του πολλαπλασιασμού πινάκων για να επιτύχει τους γραμμικούς συνδυασμούς.

Το πρόβλημα που εμφανίζεται στην χρήση του πολλαπλασιασμού πινάκων, και η λύση που επιφέρει η συνέλιξη, μπορεί να γίνει κατανοητό με το παρακάτω παράδειγμα [15]:

Παράδειγμα 2.3. *Θεωρούμε πως διαθέτουμε την αριστερή εικόνα και επιθυμούμε να κάνουμε edge detection πάνω σε αυτή, όπου το αποτέλεσμα φαίνεται στην δεξιά εικόνα.*



Εικόνα 2.2: *Edge Detection πάνω στην εικόνα ενός σκύλου*

Η εικόνα εισόδου είναι διαστάσεων $320 \times 280 \times 3$ *και η εικόνα εξόδου είναι διαστάσεις* $319 \times 280 \times 3$. *Η εξαγωγή αυτή βασίζεται στην εξής ιδέα: Αφαιρούμε το αριστερό γειτονικό pixel από κάθε ένα pixel της εικόνας.*
Αν προσπαθήσουμε να εφαρμόσουμε την παραπάνω ιδέα με συνέλιξη, το συνολικό κόστος σε πράξεις θα είναι ίσο με $319 \times 280 \times 3 = 267,960$ *πράξεις.*

Αν προσπαθήσουμε με πολλαπλασιασμό πινάκων, το συνολικό κόστος σε πράξεις θα είναι ίσο με $320 \times 280 \times 319 \times 280 = 8,003,072,000$ *πράξεις.*
Οι αριθμοί μιλούν από μόνοι τους, παρατηρούμε πως με την χρήση της συνέλιξης, χρειαστήκαμε μόνο το $1/29,866,667$ *περίπου των πράξεων. Επομένως η επεξεργασία εικόνων και η εξαγωγή χαρακτηριστικών γίνεται πολύ πιο αποδοτικά.*

Convolution Layer

Το πιο βασικό συστατικό, το οποίο αποτελεί την καρδιά των ΣΝΔ ειναι το Συνελικτικό επίπεδο (Convolution Layer). Σε αυτό το επίπεδο, με την χρήση της συνέλιξης, εξάγονται διάφορα χαρακτηριστικά τα οποία βοηθούν στην επίτευξη των εργασιών που έχει στόχο το ΣΝΔ να επιτύχει.

Για να επιτευχθεί ωστόσο η συνέλιξη χρειάζεται να οριστεί πρώτα ο κατάλληλος πυρήνας. Λόγω αυτού, το Convolution Layer είναι το σημείο όπου ορίζονται οι διαστάσεις του πυρήνα που θα εφαρμόσουμε. Τα στοιχεία του πυρήνα, είναι τα στοιχεία τα οποία το μοντέλο μας πρέπει να μάθει και να ορίσει κατάλληλα, μέσω της εκπαίδευσης.

Ο τρόπος με τον οποίο η συνέλιξη βοηθάει στην εξαγωγή χαρακτηριστικών είναι ο εξής:

Παράδειγμα 2.4. *Ας θεωρήσουμε πως έχουμε μια εικόνα* I*, με διαστάσεις* $3 \times 4 \times 1$*, η οποία αναπαριστάται από τον εξής πίνακα:*

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

και τον πυρήνα K*, ο οποίος τον ορίζουμε ως πίνακα με διαστάσεις* 2×2*, και αναπαριστάται από τον πίνακα:*

$$K = \begin{bmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{bmatrix}$$

Τότε το αποτέλεσμα της συνέλιξης $S = I * K$ *είναι ένας πίνακας διαστάσεων* 3×2 *ο οποίος είναι ο εξής:*

$$S = \begin{bmatrix} s_{11} & s_{12} & s_{13} \\ s_{21} & s_{22} & s_{23} \end{bmatrix}$$

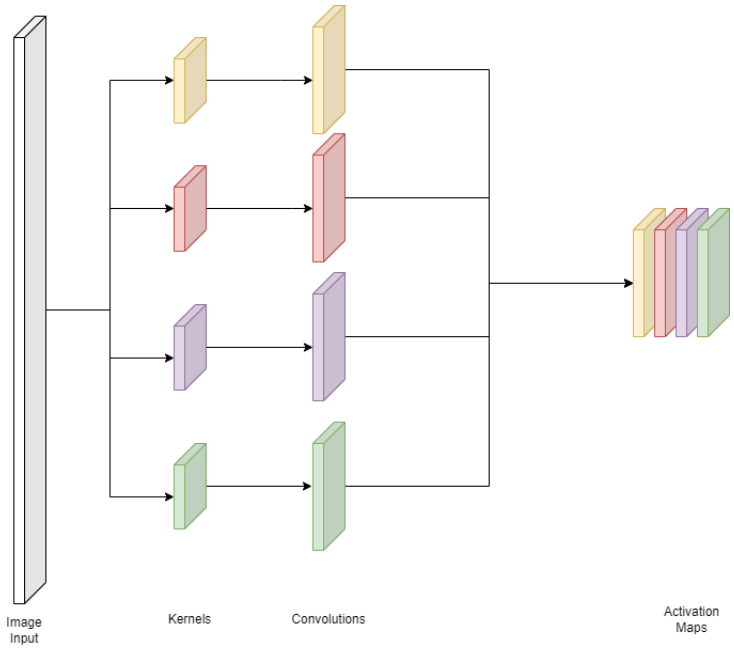
με στοιχεία

$$\begin{aligned} s_{11} &= a_{11}k_{11} + a_{12}k_{12} + a_{21}k_{21} + a_{22}k_{22} \\ s_{12} &= a_{12}k_{11} + a_{13}k_{12} + a_{22}k_{21} + a_{23}k_{22} \\ s_{13} &= a_{13}k_{11} + a_{14}k_{12} + a_{23}k_{21} + a_{24}k_{22} \\ s_{21} &= a_{21}k_{11} + a_{22}k_{12} + a_{31}k_{21} + a_{32}k_{22} \\ s_{22} &= a_{22}k_{11} + a_{23}k_{12} + a_{32}k_{21} + a_{33}k_{22} \\ s_{23} &= a_{23}k_{11} + a_{24}k_{12} + a_{33}k_{21} + a_{34}k_{22} \end{aligned}$$

Με βάση τα παραπάνω, βλέπουμε ότι το αποτέλεσμα της συνέλιξης ουσιαστικά είναι

αποτέλεσμα πράξεων μεταξύ γειτονικών στοιχείων της εικόνας. Αυτός είναι και ο λόγος, όπου τα συνελικτικά επίπεδα καταφέρνουν να παράξουν αποτελεσματικά χαρακτηριστικά πάνω στις εικόνες που δέχονται, τα οποία καταφέρνουν να επιτύχουν ορθές ταξινομήσεις σε ζητήματα ταξινόμησης.

Ωστόσο με την χρήση μόνο μιας συνέλιξης, τα χαρακτηριστικά αυτά σε εικόνες μεγαλύτερης διάστασης δεν θα είναι τόσο πλούσια. Για τον λόγο αυτό, λόγω του ότι η συνέλιξη είναι οικονομική πράξη, σε ένα Convolution Layer μπορούμε να ορίσουμε πολλούς πυρήνες, έτσι ώστε να δημιουργηθούν πολλές νέες εικόνες χαρακτηριστικών, τα λεγόμενα activation maps. Η σχετική ιδέα φαίνεται στην παρακάτω εικόνα.



Σχήμα 2.5: *Το Convolution Layer με χρήση πολλών πυρήνων*

Σε αυτή την εικόνα, βλέπουμε το παράδειγμα ενός Convolution Layer το οποίο αποτελείται από τέσσερις πυρήνες και το αποτέλεσμα του είναι το activation map που ορίζεται από τις τέσσερις συνελίξεις που προέκυψαν.

Το τελευταίο ζητήματα το οποίο μας έχουμε να εξετάσουμε στην δομή του Convolution Layer είναι το τι διαστάσεις θα έχουν τα τελικά αποτελέσματα του επιπέδου. Οι λόγοι που μας ενδιαφέρει ένα τέτοιο ζήτημα είναι δύο:

1. Κατά την διαδικασία του υπολογισμού της συνέλιξης του πίνακα εισόδου με τον πυρήνα, μπορούμε να εφαρμόσουμε την συνέλιξη με παραπάνω από ένα βήμα. Πιο πάνω, στο **Παράδειγμα 2.4** παρατηρούμε πως η συνέλιξη προκύπτει με ”σκανάρισμα” του πυρήνα κατά μια θέση δεξιά ανά νέο στοιχείο. Αυτό το βήμα ονομάζεται *Stride* και είναι και αυτή μια παράμετρος του επιπέδου.

2. Αναλόγως την τιμή του stride, υπάρχουν περιπτώσεις που η συνέλιξη δεν είναι εφικτή. Αυτό σημαίνει διότι οι διαστάσεις του πυρήνα και της εικόνας δεν συμβαδίζουν κατάλληλα, με αποτέλεσμα στο τέλος της εικόνας να μην επαρκούν τα στοιχεία της, ώστε να ολοκληρωθεί σωστά η συνέλιξη. Για τον λόγο αυτό, εφαρμόζουμε την τακτική

Zero Padding, η οποία αυξάνει τις διαστάσεις της εικόνας κατάλληλα, με την εισαγωγή μηδενικών pixel, τα οποία δεν προσφέρουν κάποια πληροφορία στο αποτέλεσμα και απλά βοηθούν στην ορθή ολοκλήρωση της συνέλιξης. Η παραπάνω λύση δίνεται σαν παράμετρος και αυτή στο επίπεδο, επονομαζόμενη ως *Padding*, όπου ορίζουμε πόσο θα μεγαλώσουμε περιμετρικά την εικόνα μας με εφαρμογή του Zero Padding.

Επομένως, συνοψίζουμε τις παραμέτρους και τα αποτελέσματα του Convolution Layer ως εξής:

Θεωρούμε πως έχουμε ένα Convolution Layer το οποίο δέχεται ως είσοδο μια εικόνα I διαστάσεων $W_I \times H_I \times D_I$, όπου W_I το πλάτος της εικόνας, H_I το μήκος της εικόνας και D_I το πλήθος των καναλιών της εικόνας. Επίσης, θεωρούμε πως το επίπεδο έχει τις εξής παραμέτρους:

- το πλήθος των πυρήνων K
- το μέγεθος των πυρήνων F
- το stride S
- το μέγεθος του Padding P.

Τότε το αποτέλεσμα που επιστρέφει το επίπεδο, μετά την πραγματοποίηση των συνελίξεων είναι ο πίνακας O διαστάσεων $W_O \times H_O \times D_O$, όπου οι διαστάσεις προκύπτουν ως εξής:

- Το πλάτος W_O είναι ίσο με $(W_I - F + 2P)/S + 1$
- Το μήκος H_O είναι ίσο με $(H_I - F + 2P)/S + 1$
- Το πλήθος των καναλιών της εξόδου D_O είναι ίσο με K

Το συνολικό πλήθος παραμέτρων που έχει να μάθει το ΣΝΔ σε ένα Convolution Layer είναι ίσο με $(F \times F \times D_I) \times K + K$, όπου το $(F \times F \times D_I) \times K$ είναι το πλήθος των βαρών των φίλτρων, και K είναι το πλήθος των biases που έχει να μάθει το μοντέλο, αν θεωρήσουμε πως χρησιμοποιεί [16].

Pooling Layer

Το επόμενο βασικό δομικό συστατικό στην δομή των ΣΝΔ είναι το Pooling Layer. Ο κύριος λόγος της εφαρμογής των Pooling Layers είναι για την βελτιστοποίηση της απόδοσης του ΣΝΔ καθώς και την αποφυγή του Overfitting.

Πιο συγκεκριμένα, είδαμε πιο πριν, πως το Convolution Layer δέχεται ως είσοδο μια εικόνα, και το παραγόμενο αποτέλεσμα, έχει πολύ μεγαλύτερες διαστάσεις, λόγω του πλήθους των φίλτρων που εφαρμόζονται πάνω στην εικόνα. Ταυτόχρονα, με την χρήση πολλών Convolution Layers, το πλήθος των παραμέτρων συνολικά σε όλο το δίκτυο θα μεγαλώσει ραγδαία, κάτι που σημαίνει είτε την αργή εκπαίδευση του μοντέλου, είτε το Overfitting του μοντέλου, μιας και το μεγάλο πλήθος βαρών, καταφέρνει να βρει όλα τα απαραίτητα χαρακτηριστικά ώστε να ταυτοποιήσει τα στοιχεία του training set.

Ο τρόπος με τον οποίο το Pooling Layer δίνει μια λύση στα παραπάνω προβλήματα, είναι με την μείωση της διαστατικότητας της εισόδου που δέχεται, στην περίπτωση μας, τα εξαγόμενα αποτελέσματα ενός Convolution Layer. Υπάρχουν δυο βασικοί τρόποι, με τους οποίους μπορεί να γίνει αυτό είναι το **Max Pooling** και το **Average Pooling**.

1. Η βασική ιδέα πίσω από το **Max Pooling** είναι η εξής:
   Θεωρούμε την είσοδο του επιπέδου να είναι ένας πίνακας με διαστάσεις 4×4, και είναι της μορφής
   $$I = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix}$$
   Ο τρόπος με τον οποίο το Max Pooling Layer διαχειρίζεται τις εισόδους του, είναι όμοιος με αυτόν της συνέλιξης. Αντί για την πράξη της συνέλιξης, τώρα εφαρμόζουμε την πράξη του μεγίστου, $max(\cdot)$, σε μια περιοχή του πίνακα εισόδου μεγέθους $F \times F$, με κάποιο stride S. Αν θεωρήσουμε πως $F = 2, S = 2$, τότε το αποτέλεσμα του Max Pooling είναι:
   $$P = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}$$
   με στοιχεία
   $$\begin{aligned} p_{11} &= max(a_{11}, a_{12}, a_{21}, a_{22}) \\ p_{12} &= max(a_{13}, a_{14}, a_{23}, a_{24}) \\ p_{21} &= max(a_{31}, a_{32}, a_{41}, a_{42}) \\ p_{22} &= max(a_{33}, a_{34}, a_{43}, a_{44}) \end{aligned}$$

2. Η βασική ιδέα πίσω από το **Average Pooling** είναι η εξής: Θεωρούμε την είσοδο του επιπέδου να είναι ένας πίνακας με διαστάσεις 4×4, και είναι της μορφής
   $$I = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix}$$
   Ο τρόπος με τον οποίο το Average Pooling Layer διαχειρίζεται τις εισόδους του, είναι όμοιος με αυτόν του Max Pooling Layer, τώρα εφαρμόζουμε την πράξη του μέσου όρου, $average(\cdot)$, σε μια περιοχή του πίνακα εισόδου μεγέθους $F \times F$, με κάποιο stride S. Αν θεωρήσουμε πως $F = 2, S = 2$, τότε το αποτέλεσμα του Max Pooling είναι:
   $$P = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}$$

με στοιχεία

$$p_{11} = average(a_{11}, a_{12}, a_{21}, a_{22})$$
$$p_{12} = average(a_{13}, a_{14}, a_{23}, a_{24})$$
$$p_{21} = average(a_{31}, a_{32}, a_{41}, a_{42})$$
$$p_{22} = average(a_{33}, a_{34}, a_{43}, a_{44})$$

Επομένως, συνοψίζουμε τα τελικά αποτελέσματα του Pooling Layer ως εξής:

Θεωρούμε πως έχουμε ένα Pooling Layer το οποίο δέχεται ως είσοδο έναν πίνακα I, διαστάσεων $W_I \times H_I \times D_I$. Επίσης, θεωρούμε πως το επίπεδο έχει τις εξής παραμέτρους:

- το μέγεθος του pooling F
- το stride S

Τότε το αποτέλεσμα που επιστρέφει το επίπεδο, μετά την πραγματοποίηση των συνελίξεων είναι ο πίνακας O διαστάσεων $W_O \times H_O \times D_O$, όπου οι διαστάσεις προκύπτουν ως εξής:

- Το πλάτος W_O είναι ίσο με $(W_I - F)/S + 1$
- Το μήκος H_O είναι ίσο με $(H_I - F)/S + 1$
- Το πλήθος των καναλιών της εξόδου D_O είναι ίσο με D_I

Το θετικό στην χρήση ενός Pooling Layer είναι πως δεν εισάγονται νέες παράμετροι εκμάθησης και πως ο υπολογισμός του είναι αρκετά ταχύς [16].

Fully Connected Layer

Το τελευταίο βασικό συστατικό στην δομή των ΣΝΔ είναι το Fully Connected Layer. Το επίπεδο αυτό δεν είναι τίποτα άλλο παρά ένα απλό ΤΝΔ, όπως αυτά που περιγράψαμε στο προηγούμενο κεφάλαιο.

Ένα Fully Connected Layer αποτελείται από συστάδες νευρώνων, οι οποίοι δέχονται τους πίνακες χαρακτηριστικών που έχουν παραχθεί μέσα από τα Convolution - Pooling Layers και με βάση αυτά, καταφέρνουν να επιτύχουν τα ζητούμενα που απαιτούνται. Οι βασικές υλοποιήσεις που τα κάνουν να ” διαφέρουν ” από τα ΤΝΔ είναι οι εξής:

1. Οι πίνακες χαρακτηριστικών που παράγονται από τα προηγούμενα επίπεδα έχουν ως τελική μορφή την μορφή πινάκων, με διαστάσεις $W_O \times H_O \times D_O$. Ο τρόπος με τον οποίο εισάγονται τα στοιχεία τους είναι στην μορφή ενός και μόνο διανύσματος, το οποίο είναι δισδιάστατο, με διαστάσεις $1 \times (W_O \cdot H_O \cdot D_O)$ και εισάγεται στο Input Layer του ΤΝΔ.

2. Για το Output Layer χρησιμοποιούμε συνήθως την συνάρτηση Softmax και ως πλήθος των νευρώνων έχουμε το πλήθος των κλάσεων όπου μελετάμε στο πρόβλημα μας. Αποτέλεσμα αυτού, είναι να έχουμε ως μια γενικότερη έξοδο, έναν πίνακα πιθανοτήτων ο οποίος δίνει τις πιθανότητες που υπάρχουν ,το στοιχείο που μελετάμε, να ανήκει στην διαθέσιμες κλάσεις.

Τελική Μορφή ΣΝΔ

Με βάση τα 3 παραπάνω επίπεδα που περιγράψαμε, μπορούμε να δούμε τώρα την βασική δομή από την οποία αποτελείται ένα ΣΝΔ.

Όπως είχε προαναφερθεί, η δομή ενός ΣΝΔ, βασίζεται στην δομή των Neocognitrons. Αυτό σημαίνει πως ένα ΣΝΔ αποτελείται από συστάδες, οι οποίες προσπαθούν να εξάγουν χαρακτηριστικά τα οποία είναι αρχικά πιο απλά, και στην συνέχεια πιο σύνθετα.

Μια συστάδα αποτελείται από τα εξής στοιχεία:

1. Ένα Convolution Layer, για την εξαγωγή των χαρακτηριστικών.
2. Ένα Pooling Layer, για μείωση της διαστατικότητας των χαρακτηριστικών και δυνατότητα γενίκευσης των χαρακτηριστικών.
3. Μερικές φορές, σε αυτό το σημείο μπορούν να προστεθούν και άλλα επίπεδα τα οποία προσφέρουν μια παραπάνω γενίκευση στα στοιχεία των χαρακτηριστικών. Μερικά από αυτά είναι τα εξής:

   - **Batch Normalization**: Μετασχηματίζουμε τα δεδομένα μας κανονικοποιόντας τα. Αν θεωρήσουμε τον πίνακα εισόδου I, και τις ποσότητες

   $$\mu_I = \frac{1}{m}\sum_{i=1}^{m} x_i$$

   $$\sigma_I^2 = \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu_I)^2$$

   όπου μ_I η μέση τιμή και σ_I η διασπορά του πίνακα. Τότε ο πίνακας μετασχηματίζεται στον νέο πίνακα Z, όπου

   $$Z = \gamma \cdot \frac{I - \mu_I}{\sqrt{\sigma_I^2 + \epsilon}} + \beta$$

   με το ϵ να είναι μια τυχαία παραγόμενη σταθερά με μικρή τιμή και γ, β παράμετροι οι οποίοι εκμαθαίνονται κατά την εκπαίδευση. [17]

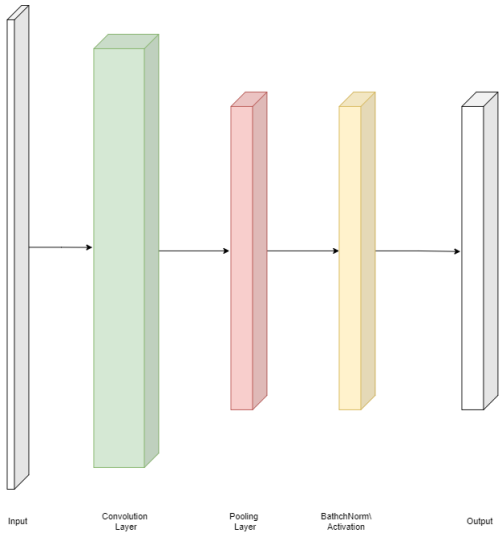
   - **Dropout**: Μειώνουμε την πληροφορία με τυχαίο τρόπο, θέτοντας τυχαία στοιχεία ίσα με 0. Αν θεωρήσουμε τον πίνακα εισόδου I και μια τυχαία μεταβλητή $r \sim Bernoulli(p)$, με διαστάσεις ίσες με τις διαστάσεις της εισόδου, τότε ο νέος πίνακας

   $$O = I \otimes r$$

   είναι ο πίνακας με στοιχεία $O(i,j,k) = I(i,j,k) \cdot r(i,j,k)$ [18]

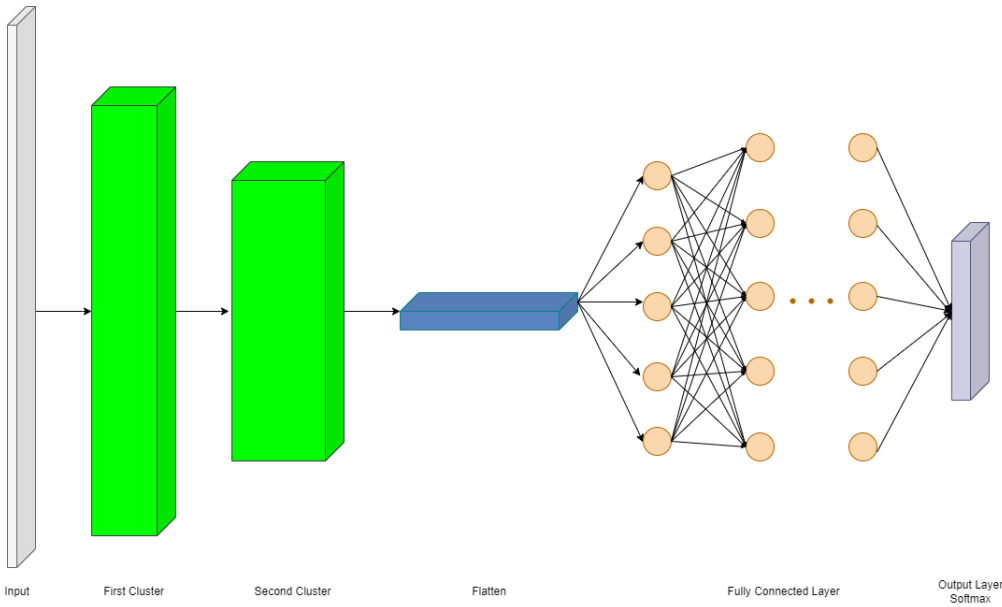
   - **Activation Layer**: Εφαρμόζουμε μια συνάρτηση ενεργοποίησης, πάνω στις τιμές της εισόδου. Συνήθως η συνάρτηση ενεργοποίησης που εφαρμόζεται είναι η *ReLu*.

Επομένως, η δομή μιας συστάδας που αποτελείται από τα παραπάνω στοιχεία μπορεί να αναπαρασταθεί από την παρακάτω εικόνα:



Σχήμα 2.6: *Η δομή μιας συστάδας (cluster) ενός ΣΝΔ*

και η τελική αρχιτεκτονική δομή ενός ΣΝΔ μπορεί να αναπαρασταθεί από την παρακάτω εικόνα:



Σχήμα 2.7: *Η δομή ενός ΣΝΔ με 2 συστάδες*

Κεφάλαιο 3

Περιγραφή Θέματος

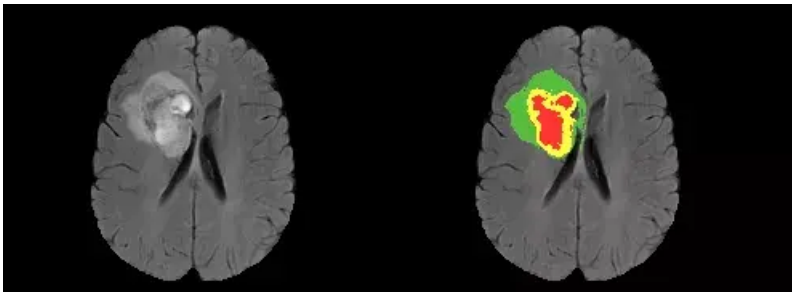
Στο κεφάλαιο αυτό θα αναλύσουμε τα μοντέλα τα οποία αποτελούν την καρδιά αυτής της εργασίας, τα U-NETs και τα U-NETs with Attention. Αρχικά θα αναφερθούμε στην αρχική τους δομή, την ονομαζόμενη Fully Convolutional Networks ή αλλιώς FCN. Στην συνέχει, θα αναλύσουμε την αρχιτεκτονική των U-NETs και U-NETs with Attention και τα μέρη από τα οποία αποτελούνται και θα μελετήσουμε τις διαφορές τους με τα προηγούμενα μοντέλα στα οποία βασίστηκαν.

3.1 Πρόβλημα: Image Segmentation

Όπως είδαμε στο προηγούμενο κεφάλαιο, τα CNNs αποτελούν μια πάρα πολύ παραγωγική και αποτελεσματική μορφή μοντέλου, τα οποία μπορούν να επεξεργαστούν με επιτυχία δεδομένα εικόνων και να επιλύσουν προβλήματα ταξινόμησης όσο αφορά το τι αντικείμενο αντιπροσωπεύει μια εικόνα. Ωστόσο σε μια εικόνα, μπορούν να υπάρχουν πάνω από 2 αντικείμενα, τα οποία θα θέλαμε όχι μόνο απλά να ξέρουμε την ύπαρξη τους, αλλά και το τι τμήμα της εικόνας καταλαμβάνουν. Αυτά τα ζητήματα αποτελούν κομμάτι του γενικότερου κλάδου του **Image Segmentation**.

Το πρόβλημα που υπάρχει με την δομή των CNNs είναι η μείωση της διαστατικότητας που προκαλούν. Πιο συγκεκριμένα, οι συνεχείς εφαρμογές των convolution layers και pooling layers μειώνουν την αρχική διάσταση της εικόνας, ενώ εμείς θα θέλαμε να φτάσουμε σε ένα αποτέλεσμα το οποίο θα είναι ουσιαστικά μια εικόνα, της οποίας τα αντίστοιχα pixels που αντιστοιχούν σε ένα από τα αντικείμενα που υπάρχουν στην εικόνα, να ξεχωρίζουν χρωματικά από τα αντίστοιχα pixels των άλλων αντικειμένων.

Ένα παράδειγμα ενός τέτοιου ζητούμενου φαίνεται στην παρακάτω εικόνα.



Εικόνα 3.1: *Παράδειγμα προβλήματος Image Segmentation [2]*

Όπως παρατηρούμε παραπάνω, αριστερά έχουμε μια εικόνα η οποία αποτελεί ένα στοιχείο εισόδου. Η εικόνα αυτή αναπαριστά έναν εγκέφαλο ο οποίος έχει έναν καρκινικό όγκο. Στην δεξιά εικόνα έχουμε την ίδια εικόνα, μόνο που η περιοχή του όγκου είναι χρωματισμένη, ως ένδειξη του εντοπισμού του.

3.2 Fully Convolutional Networks - FCN

Την αδυναμία των CNNs πάνω στο πρόβλημα του Image Segmentation έρχονται να το λύσουν τα Fully Convolutional Networks - FCN.

Τα FCNs είναι μια παραλλαγή των CNNs η οποία βασίζεται στην τεχνική του *Upsampling*. Η κεντρική ιδέα είναι πως από την στιγμή που τα CNNs μέσω των Convolution Layers μπορούν να εξάγουν χρήσιμα χαρακτηριστικά, μέσα από τα οποία μπορούν να διακριθούν τα διάφορα στοιχεία μια εικόνας, μπορούμε να τα χρησιμοποιήσουμε ώστε να ανακατασκευάσουμε την αρχική εικόνα. Ο τρόπος με τον οποίο γίνεται αυτό εφικτό είναι με την αντικατάσταση του Fully Connected Layer στην δομή του CNN, με νέα Convolution Layers και νέα επίπεδα Pooling, τα οποία αντί να μειώνουν τις διαστάσεις των εξόδων των συνελικτικών επιπέδων, να τις αυξάνουν, τα λεγόμενα Decovolution Layers και Max Unpooling Layers και ταυτόχρονα τα αποτελέσματα αυτών των επιπέδων να συνδυάζονται με αντίστοιχα αποτελέσματα των Convolution Layers από πιο πριν (περισσότερες λεπτομέρειες για τους μηχανισμούς Upsampling θα αναλύσουμε στο Κεφάλαιο 3.3.1).

Έτσι το τελικό αποτέλεσμα αποδίδεται ως μια εικόνα, η οποία με τις πληροφορίες από τα Downsampling Layers και τα Upsampling Layers καταφέρνει να διατυπώσει τις διακρίσεις που κάνει ανάμεσα σε στοιχεία μιας εικόνας από το τοπικό επίπεδο, σε ένα ολικό συμπέρασμα [19].

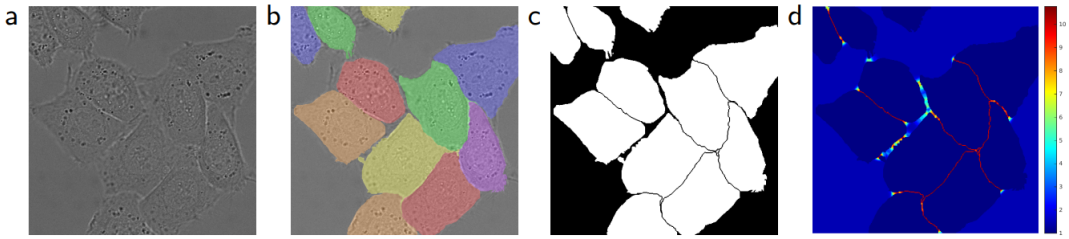
3.3 U-NETs

Η μετεξέλιξη των FCN έρχεται έναν χρόνο μετά, με την εισαγωγή των U-Nets. Τα FCN ενώ καταφέρνουν να επιλύσουν μερικώς το πρόβλημα του Image Segmentation, αδυνατούν να δουλέψουν σε προβλήματα στα οποία απαιτείται μια πιο λεπτομερής διαχώριση των στοιχείων μιας εικόνας. Άλλο ένα ζήτημα το οποίο έρχονται να λύσουν τα μοντέλα αυτά ο τρόπος εισαγωγής των εικόνων στο μοντέλων. Τα πιο κλασσικά μοντέλα FCN περισσότερες φορές πατάνε σε γενικότερα μοντέλα όπως το ImageNet, το οποίο έχει ιδιαίτερες απαιτήσεις στην εισαγωγή των δεδομένων [1].

Επομένως ήταν επιτακτική η ανάγκη για ένα πιο ισχυρό μοντέλο, το οποίο θα μπορούσε να δουλέψει με ένα πιο γενικό εύρος όσο αφορά τις διαστάσεις των δεδομένων, αλλά και ταυτόχρονα να μπορεί να διαχειριστεί το τι αντιπροσωπεύουν τα δεδομένα αυτά. Αναφερόμαστε ιδιαίτερα σε αυτό, διότι ο σκοπός των U-Nets είναι να γίνεται ένα ικανοποιητικό Image Segmentation πάνω σε εικόνες ιατρικού και βιολογικού χαρακτήρα [1].

Ένα παράδειγμα μιας τόσο απαιτητικής εικόνας είναι το παρακάτω:

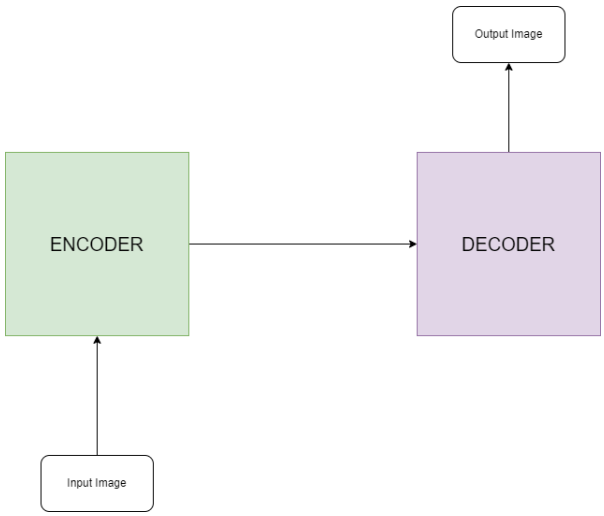
Παράδειγμα 3.5. *Θεωρούμε ένα ζεύγος κυττάρων HeLa, των οποίων την απεικόνιση την έχουμε λάβει μέσω DIC μικροσκοπίου, όπως φαίνεται στην εικόνα a. Στην εικόνα b απεικονίζεται η ground truth απεικόνιση του κάθε κυττάρου ξεχωριστά, το σωστό Image Segmentation. Στόχος του U-Net είναι να κατασκευάσει ένα όσο καλύτερο αποτέλεσμα, κοντά στο ground truth, κάτι που καταφέρνει να κάνει όπως φαίνεται στα αποτελέσματα που έχουμε στις εικόνες c, d από το [1] .*



Εικόνα 3.2: *Παράδειγμα κυττάρων και του Segmentation που ζητείται πάνω στην αρχική εικόνα.(Από το paper [1])*

3.3.1 Αρχιτεκτονική U-NETs

Η αρχιτεκτονική ενός U-Net μπορούμε να πούμε πως αποτελείται από 2 βασικά σύνολα επιπέδων, τα οποία λειτουργούν με την δομή Encoder - Decoder.



Σχήμα 3.1: *Η βασική δομή Encoder - Decoder.*

Η γενική ιδέα πίσω από την κλασσική αυτή δομή μοντέλου είναι η εξής:

- ***Encoder Block***: Το επίπεδο αυτό, δέχεται ως είσοδο την αρχική εικόνα και διενεργεί την βασική ανάλυση της. Μέσω αυτής της επεξεργασίας, εξάγονται ως έξοδος, τα κεντρικά χαρακτηριστικά της, τα οποία αποτελούν την πηγή πληροφορίας του επόμενου μπλοκ.

- ***Decoder Block***: Το επίπεδο αυτό, δέχεται ως είσοδο τα χαρακτηριστικά που παράχθηκαν από το Encoder Block και με βάση αυτά, ανακατασκευάζει την εικόνα και προσπαθεί να παράγει το τελικό αποτέλεσμα που επιθυμούμε.

Στην συνέχεια θα μελετήσουμε σε βάθος την εσωτερική δομή αυτών των επιπέδων.

Encoder Block

Όπως προαναφέραμε, το Encoder Block έχει ως στόχο την εξαγωγή των χαρακτηριστικών της εικόνας εισόδου. Επομένως η δομή του θα είναι όμοια με αυτή ενός CNN και ενός FCN και θα αποτελείται από συστάδες που επεξεργάζονται τα χαρακτηριστικά με την βοήθεια των Convolution Layers και Pooling Layers.

Πιο συγκεκριμένα, στο [1], στόχος είναι η διαδοχική μείωση της διαστατικότητας των δεδομένων, ώστε να καταλήξουμε σε χαρακτηριστικά αρκετής μικρής διάστασης $W \times H$, αλλά με αρκετά μεγάλο βάθος D, το οποίο κρατάει την πληροφορία συγκεντρωμένη.

Τα βασικά επίπεδα τα οποία θα χρησιμοποιήσουμε στην συγκεκριμένη δομή είναι τα εξής:

1. **Conv 3 x 3, ReLU**: Συνελικτικό επίπεδο με μέγεθος πυρήνα 3, stribe ίσο με 1 και την εφαρμογή της συνάρτηση ενεργοποίησης ReLU πάνω στα αποτελέσματα του Συνελικτικού επιπέδου.

2. **Max pool 2 x 2**: Pooling επίπεδο με μέγεθος πυρήνα 2 και stribe ίσο με 1.

Η δομή που προτάθηκε στο [1], με αρχική εικόνα εισόδου I διαστάσεων $572 \times 572 \times 1$, είναι η εξής:

- Ξεκινάμε με την εισαγωγή της εικόνας I στο μοντέλο και την επεξεργασία της από ένα Conv 3 x 3, ReLU με αποτέλεσμα την παραγωγή της συνέλιξης διαστάσεων $570 \times 570 \times 64$ και στην συνέχεια την εφαρμογή ξανά ενός Conv 3 x 3, ReLU με αποτέλεσμα την παραγωγή την τελική συνέλιξης O_1 διαστάσεων $568 \times 568 \times 64$.

- Εφαρμογή ενός Max pool 2 x 2 στον πίνακα O_1 με αποτέλεσμα τον πίνακα O_2 με διαστάσεις $284 \times 284 \times 64$.

- Επεξεργασία του πίνακα O_2 από ένα Conv 3 x 3, ReLU με αποτέλεσμα την παραγωγή της συνέλιξης διαστάσεων $282 \times 282 \times 128$ και στην συνέχεια την εφαρμογή ξανά ενός Conv 3 x 3, ReLU με αποτέλεσμα την παραγωγή την τελική συνέλιξης O_3 διαστάσεων $280 \times 280 \times 128$.

- Εφαρμογή ενός Max pool 2 x 2 στον πίνακα O_3 με αποτέλεσμα τον πίνακα O_4 με διαστάσεις $140 \times 140 \times 128$.

- Επεξεργασία του πίνακα O_4 από ένα Conv 3 x 3, ReLU με αποτέλεσμα την παραγωγή της συνέλιξης διαστάσεων $138 \times 138 \times 256$ και στην συνέχεια την εφαρμογή ξανά ενός Conv 3 x 3, ReLU με αποτέλεσμα την παραγωγή την τελική συνέλιξης O_5 διαστάσεων $136 \times 136 \times 256$.

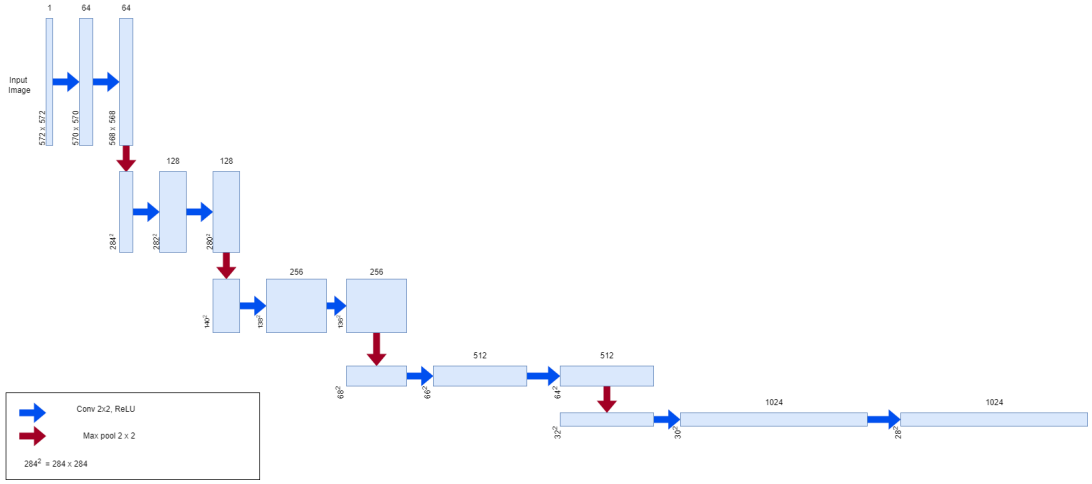
- Εφαρμογή ενός Max pool 2 x 2 στον πίνακα O_5 με αποτέλεσμα τον πίνακα O_6 με διαστάσεις $68 \times 68 \times 128$.

- Επεξεργασία του πίνακα O_6 από ένα Conv 3 x 3, ReLU με αποτέλεσμα την παραγωγή της συνέλιξης διαστάσεων $66 \times 66 \times 512$ και στην συνέχεια την εφαρμογή ξανά ενός Conv 3 x 3, ReLU με αποτέλεσμα την παραγωγή την τελική συνέλιξης O_7 διαστάσεων $64 \times 64 \times 512$.

- Εφαρμογή ενός Max pool 2 x 2 στον πίνακα O_7 με αποτέλεσμα τον πίνακα O_8 με διαστάσεις $32 \times 32 \times 512$.

- Επεξεργασία του πίνακα O_8 από ένα Conv 3 x 3, ReLU με αποτέλεσμα την παραγωγή της συνέλιξης διαστάσεων $30 \times 30 \times 1024$ και στην συνέχεια την εφαρμογή ξανά ενός Conv 3 x 3, ReLU με αποτέλεσμα την παραγωγή την τελική συνέλιξης O_9 διαστάσεων $28 \times 28 \times 1024$.

Μια οπτική αναπαράσταση της δομής του Encoder, η οποία περιγράφει την παραπάνω διαδικασία είναι η εξής:



Σχήμα 3.2: *U-NET : Encoder*

Decoder Block

Μετά την επεξεργασία της εικόνας I και την τελική εξαγωγή των χαρακτηριστικών της, που αναπαριστώνται από τον πίνακα O_9 από Encoder Block, σειρά λαμβάνει το Decoder Block. Ο στόχος του Decoder Block είναι η αναπαραγωγή της τελικής εικόνας, η οποία αποτελείται από τα αντικείμενα της εικόνας, χρωματισμένα σε κατάλληλα χρώματα, ώστε να φαίνεται το αντίστοιχο Segmentation. Επομένως η δομή του θα είναι όμοια αλλά ανάστροφη με αυτή ενός CNN και ενός FCN και θα αποτελείται από συστάδες που επεξεργάζονται τα χαρακτηριστικά μικρότερης διάστασης και θα προβάλουν σε μεγαλύτερες διαστάσεις με την βοήθεια κατάλληλων Convolution Layers που θα αυξάνουν τις διαστάσεις τους.

Τα βασικά επίπεδα τα οποία θα χρησιμοποιήσουμε στην συγκεκριμένη δομή είναι τα εξής:

1. **Conv 3 x 3, ReLU**: Συνελικτικό επίπεδο με μέγεθος πυρήνα 3, stribe ίσο με 1 και την εφαρμογή της συνάρτηση ενεργοποίησης ReLU πάνω στα αποτελέσματα του Συνελικτικού επιπέδου.
2. **Conv 1 x 1**: Συνελικτικό επίπεδο με μέγεθος πυρήνα 1, stribe ίσο με 1 .
3. **Up Conv 2 x 2**: Συνελικτικό επίπεδο με μέγεθος πυρήνα 2 και stribe ίσο με 1, το οποίο βοηθάει στην αύξηση της διαστατικότητας.
4. **Copy and Crop**: Αντιγραφή πίνακα δεδομένων και συνένωσή του με άλλο πίνακα δεδομένων.

Παρατηρούμε πως σε αυτό το επίπεδο, χρησιμοποιούμε ένα νέο είδος Συνελικτικού Επιπέδου, το Up Conv. Ο τρόπος με τον οποίο λειτουργεί το συγκεκριμένο επίπεδο, είναι με την πράξης **Transposed Convolution** [20]. Ο σκοπός αυτού του είδους συνέλιξης, είναι να αυξήσει την διαστατικότητα ενός αρχικού πίνακα, με την βοήθεια της κλασσικής συνέλιξης και ταυτόχρονου Padding. Ένα παράδειγμα της πράξης αυτής είναι το εξής:

Παράδειγμα 3.6. *Θεωρούμε πως έχουμε μια εικόνα I, με διαστάσεις 2×2 η οποία αναπαρίσταται από τον εξής πίνακα:*

$$I = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

και τον πυρήνα K, ο οποίος τον ορίζουμε ως πίνακα με διαστάσεις 3×3, και αναπαριστάται από τον πίνακα:

$$K = \begin{bmatrix} k_{11} & k_{12} & k_{13} \\ k_{21} & k_{22} & k_{23} \\ k_{31} & k_{32} & k_{33} \end{bmatrix}$$

Τότε το αποτέλεσμα της Transoped Convolution TS με stride $= 1$ είναι ένας πίνακας διαστάσεων 4×4 ο οποίος είναι ο εξής:

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

με στοιχεία

$$\begin{aligned}
ts_{11} &= a_{11}k_{11} \\
ts_{12} &= a_{11}k_{12} + a_{12}k_{11} \\
ts_{13} &= a_{11}k_{13} + a_{12}k_{12} \\
ts_{14} &= a_{12}k13 \\
ts_{21} &= a_{11}k_{21} + a_{21}k_{11} \\
ts_{22} &= a_{11}k_{22} + a_{12}k_{21} + a_{21}k_{12} + a_{22}k_{11} \\
ts_{23} &= a_{11}k_{23} + a_{12}k_{22} + a_{21}k_{13} + a_{22}k_{12} \\
ts_{24} &= a_{12}k_{23} + a_{22}k_{13}
\end{aligned}$$

$$
\begin{aligned}
ts_{31} &= a_{11}k_{33} + a_{21}k_{21} \\
ts_{32} &= a_{11}k_{32} + a_{12}k_{31} + a_{21}k_{22} + a_{22}k_{21} \\
ts_{33} &= a_{11}k_{33} + a_{12}k_{32} + a_{21}k_{23} + a_{22}k_{22} \\
ts_{34} &= a_{12}k_{33} + a_{22}k_{23} \\
ts_{11} &= a_{21}k_{31} \\
ts_{12} &= a_{21}k_{32} + a_{22}k_{31} \\
ts_{13} &= a_{21}k_{33} + a_{22}k_{32} \\
ts_{14} &= a_{22}k33
\end{aligned}
$$

Το Transposed Convolution μπορεί να γίνει και με την χρήση Zero Padding, όπως η κανονική συνέλιξη.[20]

Η δομή που προτάθηκε στο [1], με αρχικά δεδομένα O_9 από το Encoder Block με διαστάσεις $28 \times 28 \times 1024$, είναι η εξής:

- Εφαρμογή του UpConv 2 x 2 πάνω στον πίνακα δεδομένων 0_9 για την παραγωγή του πίνακα δεδομένων O_10, με διαστάσεις $56 \times 56 \times 512$.
- Εφαρμογή του Copy and Crop ώστε ο πίνακας δεδομένων O_7 ώστε να λάβουμε ένα αντίγραφό του με βάθος αντί για διαστάσεις 64×64×512 να έχει διαστάσεις 56×56×512 και ένωσής του με τον πίνακα O_10, για την κατασκευή του πίνακα O_11, με διαστάσεις $56 \times 56 \times 1024$
- Εφαρμογή του Conv 3 x 3, ReLU πάνω στον πίνακα O_11 με αποτέλεσμα την παραγωγή της συνέλιξης διαστάσεων $54 \times 54 \times 512$ και στην συνέχεια την εφαρμογή ξανά ενός Conv 3 x 3, ReLU με αποτέλεσμα την παραγωγή της τελικής συνέλιξης O_12 διαστάσεων $52 \times 52 \times 512$.
- Εφαρμογή του UpConv 2 x 2 πάνω στον πίνακα δεδομένων 0_12 για την παραγωγή του πίνακα δεδομένων O_13, με διαστάσεις $104 \times 104 \times 256$.
- Εφαρμογή του Copy and Crop ώστε ο πίνακας δεδομένων O_5 ώστε να λάβουμε ένα αντίγραφό του με βάθος αντί για διαστάσεις $136 \times 136 \times 256$ να έχει διαστάσεις $104 \times 104 \times 256$ και ένωσής του με τον πίνακα O_13, για την κατασκευή του πίνακα O_14, με διαστάσεις $104 \times 104 \times 256$
- Εφαρμογή του Conv 3 x 3, ReLU πάνω στον πίνακα O_14 με αποτέλεσμα την παραγωγή της συνέλιξης διαστάσεων $102 \times 102 \times 256$ και στην συνέχεια την εφαρμογή ξανά ενός Conv 3 x 3, ReLU με αποτέλεσμα την παραγωγή της τελικής συνέλιξης O_15 διαστάσεων $100 \times 100 \times 256$.
- Εφαρμογή του UpConv 2 x 2 πάνω στον πίνακα δεδομένων 0_15 για την παραγωγή του πίνακα δεδομένων O_16, με διαστάσεις $200 \times 200 \times 128$.
- Εφαρμογή του Copy and Crop ώστε ο πίνακας δεδομένων O_3 ώστε να λάβουμε ένα αντίγραφό του με βάθος αντί για διαστάσεις $280 \times 280 \times 128$ να έχει διαστάσεις $200 \times$

200×128 και ένωσής του με τον πίνακα O_16, για την κατασκευή του πίνακα O_17, με διαστάσεις $200 \times 200 \times 256$

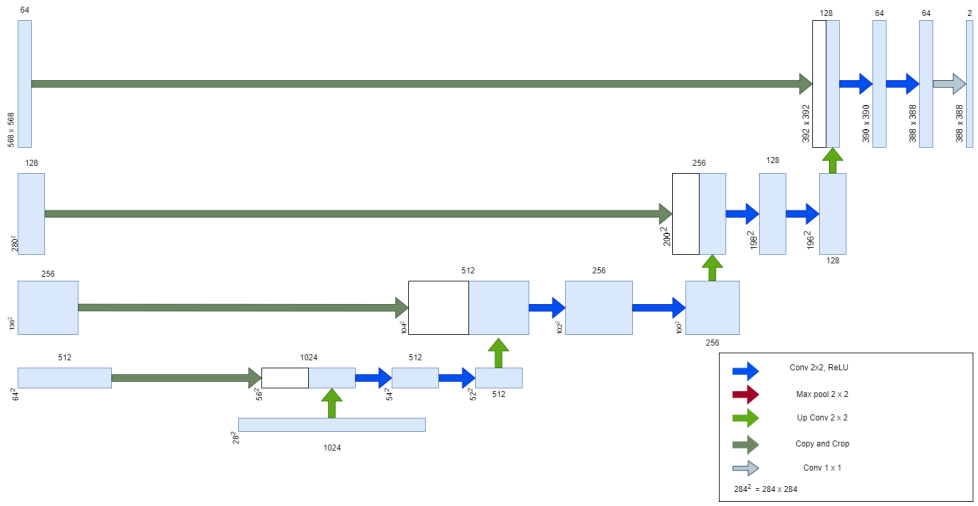
- Εφαρμογή του Conv 3 x 3, ReLU πάνω στον πίνακα O_17 με αποτέλεσμα την παραγωγή της συνέλιξης διαστάσεων $198 \times 198 \times 128$ και στην συνέχεια την εφαρμογή ξανά ενός Conv 3 x 3, ReLU με αποτέλεσμα την παραγωγή της τελικής συνέλιξης O_18 διαστάσεων $196 \times 196 \times 128$.

- Εφαρμογή του UpConv 2 x 2 πάνω στον πίνακα δεδομένων 0_18 για την παραγωγή του πίνακα δεδομένων O_19, με διαστάσεις $392 \times 392 \times 64$.

- Εφαρμογή του Copy and Crop ώστε ο πίνακας δεδομένων O_1 ώστε να λάβουμε ένα αντίγραφό του με βάθος αντί για διαστάσεις $568 \times 568 \times 64$ να έχει διαστάσεις $392 \times 392 \times 128$ και ένωσής του με τον πίνακα O_19, για την κατασκευή του πίνακα O_20, με διαστάσεις $392 \times 392 \times 128$

- Εφαρμογή του Conv 3 x 3, ReLU πάνω στον πίνακα O_20 με αποτέλεσμα την παραγωγή της συνέλιξης διαστάσεων $390 \times 390 \times 64$, στην συνέχεια την εφαρμογή ξανά ενός Conv 3 x 3, ReLU με αποτέλεσμα την παραγωγή της συνέλιξης διαστάσεων $388 \times 388 \times 64$ και τέλος την εφαρμογή ενός Conv 1 x 1 για την παραγωγή του τελικού αποτελέσματος O_21 με διαστάσεις $388 \times 388 \times 2$

Μια οπτική αναπαράσταση της δομής του Decoder, η οποία περιγράφει την παραπάνω διαδικασία είναι η εξής:

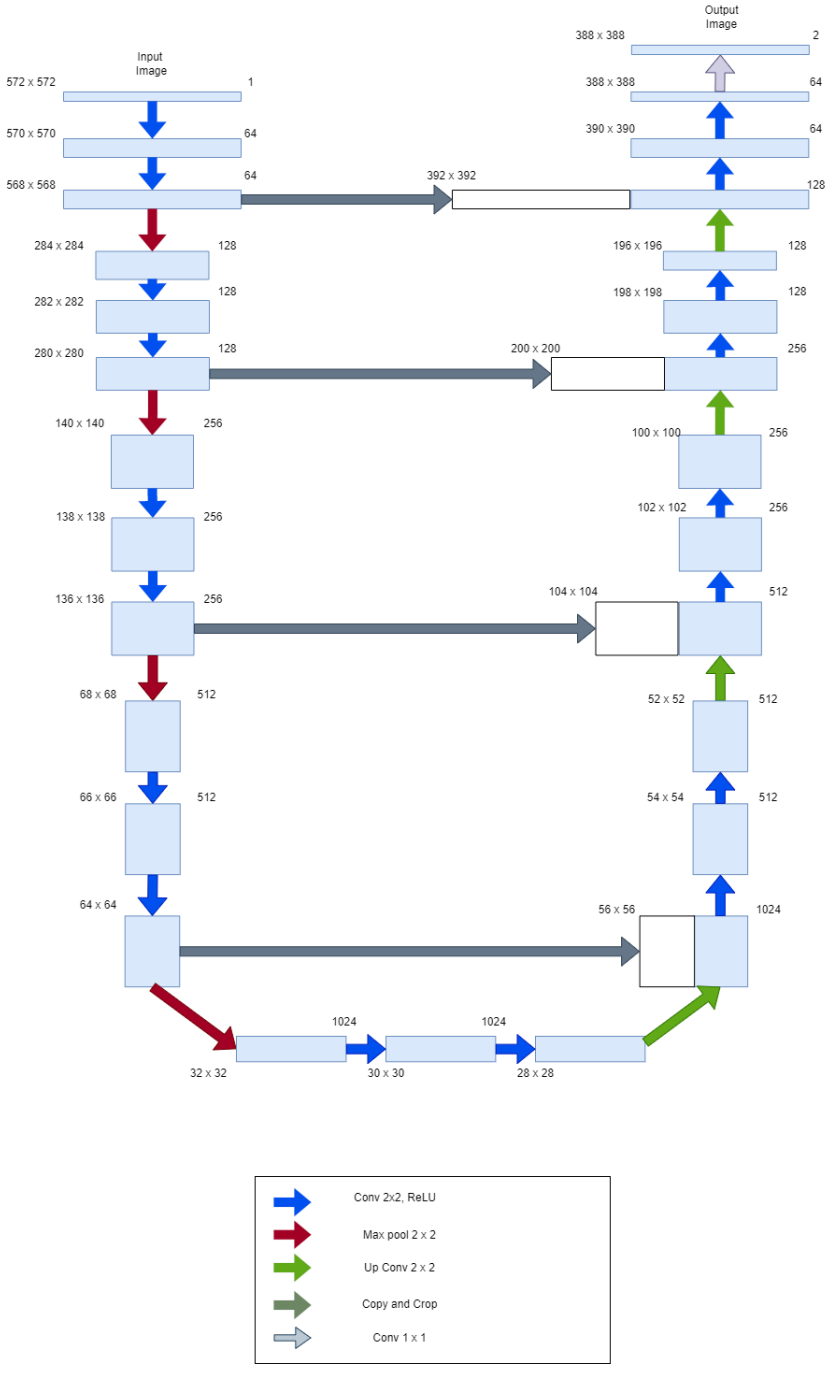


Σχήμα 3.3: *U-Net: Decoder*

Τελική Δομή

Με βάση τις 2 παραπάνω δομές που αναπτύξαμε, μπορούμε να συνθέσουμε την τελική δομή του U-Net.

Με βάση και την παρακάτω εικόνα, παρατηρούμε πως η αρχική εικόνα εισάγεται στον Encoder, αναλύεται σε χαρακτηριστικά με την βοήθεια των Convolution Layers και στην συνέχεια ο Decoder με την βοήθεια τον UpConvolution Layers και του Transposed Convolution ανακατασκευάζει την εικόνα και δημιουργεί το καταλληλο Segmentation.



Σχήμα 3.4: *Η πλήρης δομή του U-Net*

3.4 U-NETs with Attention

Όπως έχουμε αναφέρει στο προηγούμενο κεφάλαιο της εργασίας μας, η χρήση των U-Nets σε προβλήματα εντοπισμού αντικειμένων πάνω σε μια εικόνα επέφερε πολλά θετικά αποτελέσματα. Η κύρια χρήση τους εντοπίζεται στον Βιοϊατρικό κλάδο, όπου συνήθως τα χρησιμοποιούμε για τον εντοπισμό σύνθετων δομών πάνω σε μια γενικότερη εικόνα.

Ωστόσο η φύση δεν είναι τόσο απλή. Οι βιολογικές δομές είναι όπως γνωρίζουμε αρκετά περίπλοκες και η δομή τους αρκετά σύνθετη. Αυτό έχει ως αποτέλεσμα να υπάρχει δυσκολία στην αντιμετώπισή τους από την πλευρά των U-Nets. Ωστόσο το ότι οι δομές αυτές είναι περίπλοκες, δεν τις κάνει να περιέχουν βασικά χαρακτηριστικά τα οποία μετά από την επεξεργασία τους, να μην φανερωθούν.

Για τον λόγο αυτό, θα επιστρατεύσουμε μια βασική ιδέα, που εφαρμόζεται σε έναν άλλο συναφή κλάδο μελέτης της Τεχνητής Νοημοσύνης, στον κλάδο της Επεξεργασίας Φυσικής Γλώσσας (Natural Language Process), η οποία είναι ο μηχανισμός **Attention** [21]. Με την χρήση του Attention, θα ενισχύσουμε τα μοντέλα μας και εν τέλη θα ορίσουμε τα U - Nets with Attention.

3.4.1 Ο μηχανισμός Attention

Ο μηχανισμός Attention προέκυψε από την Επεξεργασία Φυσικής Γλώσσας, με βάση το εξής γεγονός: όταν συνθέτουμε μια γλωσσική έκφραση, υπάρχουν διαφορετικές σχέσεις - αλληλεπιδράσεις ανάμεσα σε κάθε λέξη που την αποτελεί.

Για παράδειγμα, στην πρόταση ” Το ποντίκι λατρεύει να τρώει τυρί”, το άρθρο ” Το ” αναφέρεται στο ουσιαστικό ” ποντίκι ” και όχι στο ουσιαστικό ”τυρί”. Αυτό σημαίνει η σχέση *Το - ποντίκι* είναι πιο ισχυρή νοηματικά από την σχέση *Το - τυρί*.

Αυτή η ιδέα ωστόσο, η ύπαρξη συγκεκριμένης αλληλεξάρτησης ανάμεσα σε στοιχεία ενός γενικότερου συνόλου αντικειμένων επεκτάθηκε αβίαστα και στον τομέα της εικόνας [22]. Έτσι, αναπτύχθηκαν πολλές προσπάθειες, ώστε να αποδοθεί όσο καλύτερα γίνεται αυτό ο μηχανισμός ανάμεσα στα pixel μιας εικόνας.

Το βασικό Attention Module

Ο γενικός μηχανισμός που υλοποιήθηκε στο paper «Attention Is All You Need»[21], αποτελεί και την βασική μορφή όσο αφορά την δομή του για σχεδόν όλες τις μορφές Attention που έχουν αναπτυχθεί για μοντέλα οπτικής απεικόνισης.

Τα δεδομένα μας αποτελούν μια συλλογή αντικειμένων, τα οποία είναι οργανωμένα σε μορφή διανύσματος, έστω το διάνυσμα I. Στόχος είναι η δημιουργία ενός διανύσματος $\mathcal{A}$, το διάνυσμα Attention, και ο ρόλος του θα είναι να λειτουργεί σαν διάνυσμα βαρών. Έτσι, μέσα από τους κατάλληλους πολλαπλασιασμούς του διανύσματος $\mathcal{A}$ με το αρχικό διάνυσμα I, το αποτέλεσμα θα είναι τέτοιο ώστε να φαίνεται ποια στοιχεία του διανύσματος I είναι ” εξαρτημένα ” μεταξύ τους.

Πιο συγκεκριμένα, ο τρόπος με τον οποίο ορίζεται ο πίνακας $\mathcal{A}$ είναι ο εξής:

Αρχικά, αντιγράφουμε το I, σε 3 νέα διανύσματα, το Query Q, το Keys K και το Values V. Στην συνέχεια, υπολογίζουμε το διάνυσμα $\mathcal{A}$ ως εξής:

$$Attention(Q, K, V) = \mathcal{A} = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

όπου με d_k συμβολίζουμε την διάσταση του διανύσματος K.

Η ιδέα αυτή ωστόσο δεν μας δίνει την δυνατότητα της ¨εκπαίδευσης¨ άμεσα. Σκοπός μας είναι και ο μηχανισμός Attention να μπορεί να εκπαιδευτεί, ώστε κατά την εκπαίδευσή του, να μπορεί να εμφανίζει καταλληλότερα αποτελέσματα. Για τον λόγω αυτό, επεκτείνουμε την ιδέα αυτή, με την χρήση κατάλληλων διανυσμάτων βαρών για τα αντίστοιχα διανύσματα Q, K και V, τα W^Q, W^K και W^V, με αποτέλεσμα εν τέλη τον υπολογισμό του εξής τύπου

$$\mathcal{A} = Attention(QW^Q, KW^K, VW^V)$$

[21].

Attention Gates

Με βάση λοιπόν τον παραπάνω ορισμό του μηχανισμού Attention, σειρά έχει να επεκτείνουμε αυτή την ιδέα και στην δομή των U - Nets. Ο τρόπος με τον οποίο καταφέρνουμε να το κάνουμε αυτό είναι με την εισαγωγή ενός νέο επιπέδου το οποίο δέχεται σαν είσοδο πίνακες χαρακτηριστικών και επιστρέφει σαν αποτέλεσμα το γινόμενο του πίνακα χαρακτηριστικών που μας ενδιαφέρει να εφαρμόσουμε τον μηχανισμό επί των βαρών του Attention. Το νέο επίπεδο θα το ονομάζουμε Attention Gate [23].

Η γενική ιδέα πίσω από το Attention Gate είναι η εξής:

Θέλουμε να χρησιμοποιήσουμε τα έτοιμα χαρακτηριστικά τα οποία προκύπτουν από τα Convolution Layers και να προσπαθήσουμε να εστιάσουμε σε pixels τα οποία φαίνεται να είναι σημαντικά πάνω σε αυτά. Επομένως η είσοδος μας θα είναι ένας πίνακας χαρακτηριστικών g και θα παίζει τον ρόλο του Query. Ωστόσο αυτό δεν είναι αρκετό. Όπως είδαμε, στα U - Nets το κύριο κομμάτι που βοηθάει στην παραγωγή του Segmentation είναι αυτό του Decoder. Στο κομμάτι αυτό, είδαμε πως χαρακτηριστικά από τα κατώτερ επίπεδα, συνδυάζονται με χαρακτηριστικά του Encoder ώστε να διατηρηθεί η αρχική δομή της πληροφορίας και ταυτόχρονα να επαυξηθεί η αρχική με βάση τα βαθύτερα χαρακτηριστικά. Για τον λόγο αυτό, θα χρησιμοποιήσουμε τα χαρακτηριστικά του Encoder για τον ρόλο του Key και Value. Ωστόσο δεν μπορούμε έτσι απλά τους προηγούμενους τύπους και να βγάλουμε καλά αποτελέσματα. Για τον λόγο αυτό, στο [23] γίνεται ο κατάλληλος ορισμός της δομής του Attention Gate.

Ένα Attention Gate ορίζεται ως εξής:

Θεωρούμε έναν πίνακα χαρακτηριστικών x του Encoder, με διαστάσεις $H_x \times W_x \times D_x$, και έναν πίνακα χαρακτηριστικών g από τον Decoder, με διαστάσεις $H_g \times W_g \times D_g$, τα οποία στην απλή δομή του U - Net, θα τα ενώναμε μεταξύ τους με την χρήση των Copy and Crop και UpConv Layer. Το πρώτο πράγμα που έχει να κάνει το Attention Gate, είναι να εισάγει κατάλληλα βάρη, ώστε να μπορέσουμε να κάνουμε εκπαιδεύσιμη την δομή μας, και ταυτόχρονα να αλλάξουμε τις διαστάσεις των πινάκων x και g, ώστε να είναι ίδιες, μιας και ισχύει ότι $H_x = 2H_g$, $W_x = 2W_g$ και $D_x = 2D_g$. Για τον λόγω αυτό, εφαρμόζουμε τις συνελίξεις W_x και W_g αντίστοιχα, οι οποίες είναι ουσιαστικά Convolution Layers με Kernel, οι οποίες

διαθέτουν πυρήνα διαστάσεων 1×1 και ίσο αριθμό φίλτρων, ίσο με D_x και με την W_x να έχει stride = 2 και την W_g να έχει stride = 1. Αυτό έχει ως αποτέλεσμα την παραγωγή των εξόδων O_x και O_g, οι οποίες έχουν διαστάσεις ίσες με $H_g \times W_g \times D_x$.

Στην συνέχεια, υπολογίζουμε το άθροισμα $S_0 = O_x + O_g$. Με αυτό το άθροισμα, καταφέρνουμε να επιτύχουμε παραπάνω ενίσχυση στα στοιχεία των πινάκων που είναι με σχετικά ίδιες τιμές και ταυτόχρονα να αποδυναμώσουμε τα στοιχεία των πινάκων τα οποία δεν είναι σχετικά κοντά μεταξύ τους.

Έπειτα, εφαρμόζουμε την συνάρτηση ενεργοποίησης ReLU ώστε να φιλτράρουμε τα δεδομένα μας από τον πίνακα S_O και να προκύψει ο πίνακας $S_{O+} = ReLU(S_O)$ έχουμε μόνο τα στοιχεία που είναι θετικά. Αυτό έχει ως συνέπεια να δοθεί ένα παραπάνω βάρος σε στοιχεία που φαίνεται να έχουν μεγάλη αξία πάνω στην δομή των χαρακτηριστικών.

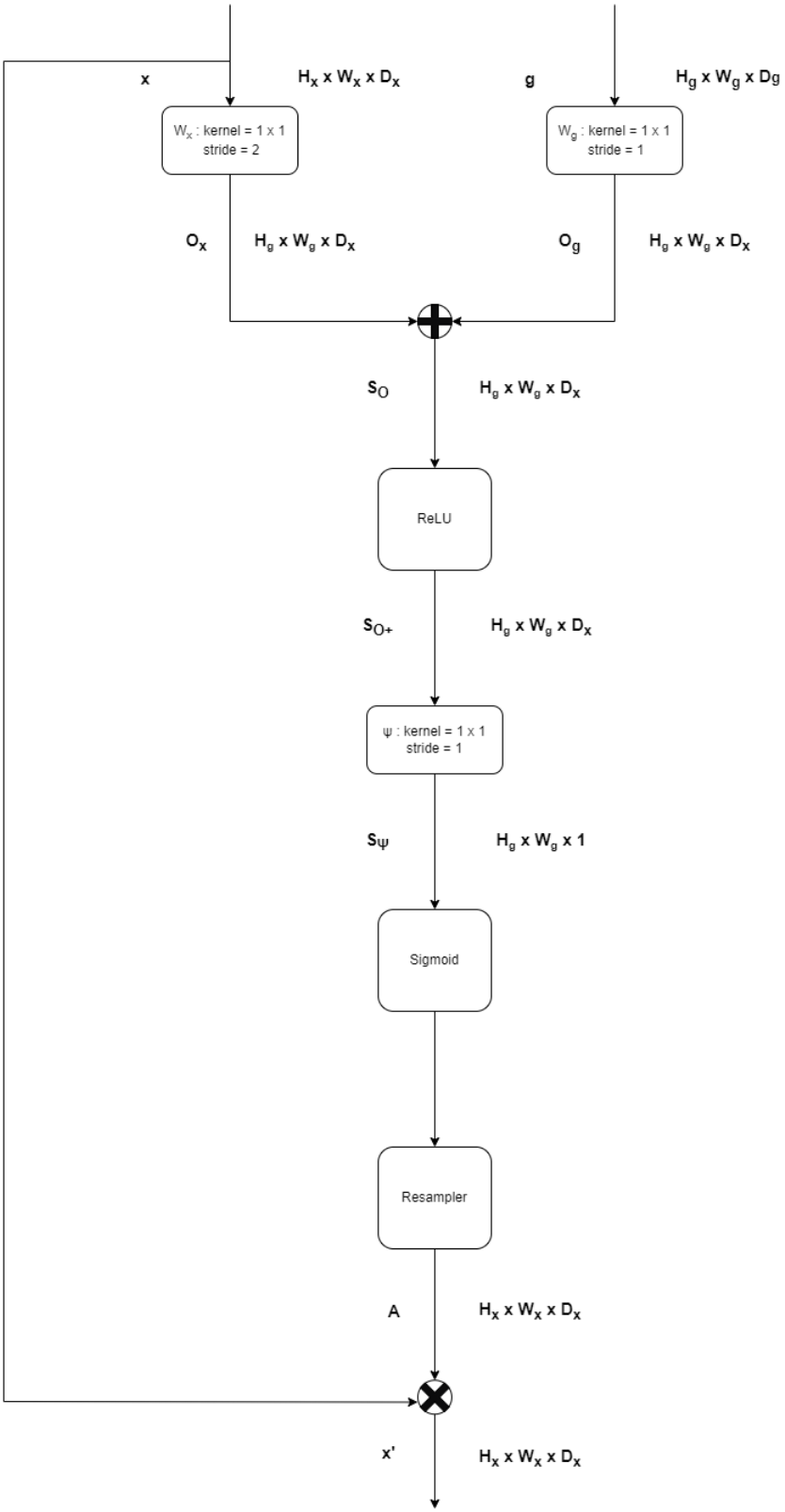
Μετά το φιλτράρισμα αυτό, σειρά έχει η εφαρμογή μιας ακόμα συνέλιξης, της ψ, με πυρήνα διαστάσεων 1×1 και αριθμό φίλτρων ίσο με 1, ώστε να συμπτύξουμε το βάθος D του πίνακα S_{O+} και να προκύψει ένας πίνακας $S_\psi = S_{O+} * \psi$, με διαστάσεις $H_{S_\psi} \times W_{S_\psi} \times D_{S_\psi} = H_g \times W_g \times 1$.

Σαν τελικό φιλτράρισμα πάνω στον πίνακα S_ψ εφαρμόζουμε μια σιγμοειδή συνάρτηση ενεργοποίησης, ώστε να περιορίσουμε το εύρος των τιμών μας στο πεδίο $(0, 1)$ και στην συνέχεια εφαρμόζουμε έναν Resampler, ο οποίος θα επαναφέρει τις διαστάσεις στις διαστάσεις $H_x \times W_x \times 1$ και εν τέλη ο τελικός πίνακας $\mathcal{A} = Resampler(sigmoid(S_\psi))$ θα είναι ο πίνακας Attention

Τέλος, για το τελικό αποτέλεσμα του Attention Gate, έχουμε το γινόμενο Hadamard του πίνακα Attention $\mathcal{A}$ με τα χαρακτηριστικά x, δηλαδή:

$$x' = \mathcal{A} \otimes x$$

Μια πιο γενική εικόνα ενός Attention Gate δίνεται από το επόμενο σχήμα:

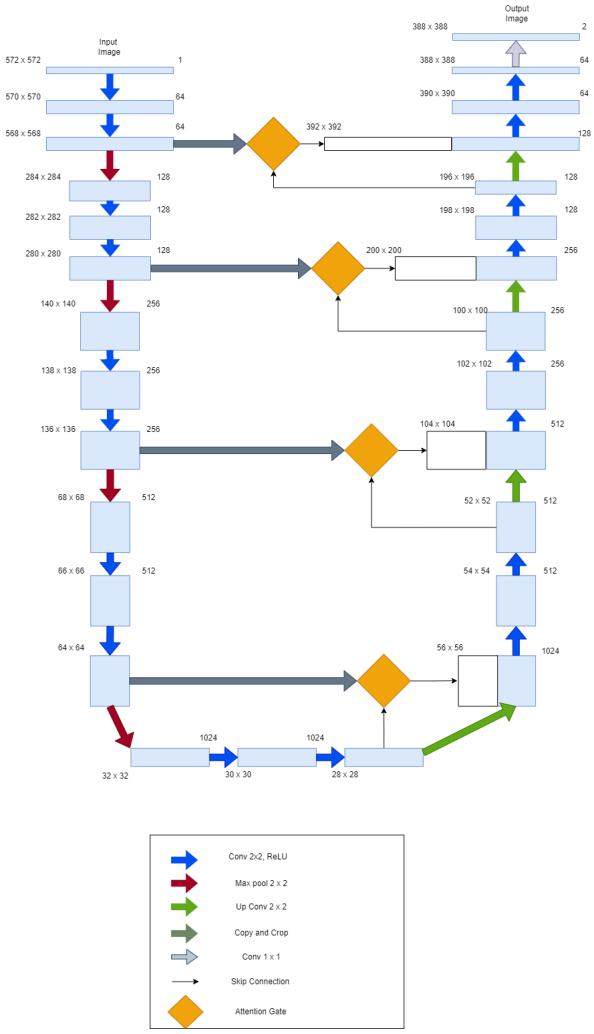


Σχήμα 3.5: *Attention Gate*

Αρχιτεκτονική U-NETs with Attention

Πλέον είμαστε σε θέση να μπορέσουμε να εντάξουμε την δομή του Attention Gate στην δομή του U - Net. Αυτό γίνεται πολύ απλά, με την εισαγωγή ενός Attention Gate κατά την διαδικασία του Copy and Crop, όπου όπως προαναφέρθηκε, όπου για x θεωρούμε τον πίνακα χαρακτηριστικών από τον Encoder και για g θεωρούμε τον πίνακα χαρακτηριστικών από τον Decoder.

Έτσι, η τελική μορφή του U-Net with Attention φαίνεται στην παρακάτω εικόνα:



Σχήμα 3.6: *U-Net with Attention*

Μέρος II

Πρακτικό Μέρος

Κεφάλαιο 4

Δομή Δεδομένων Προβλήματος και Μετρικές Σύγκρισης

Στο κεφάλαιο αυτό θα παρουσιάσουμε την δομή των δεδομένων τα οποία θα βοηθήσουν στην εκπαίδευση των αλγορίθμων Βαθιάς Μάθησης που αναλύσαμε στο Θεωρητικό Πλαίσιο που προηγήθηκε. Θα εξετάσουμε τα βασικά χαρακτηριστικά τους και την φιλοσοφία πίσω από την οποία τα μοντέλα που θα αναπτύξουμε στην συνέχεια θα μπορέσουν να εκπαιδευτούν κατάλληλα με αυτά. Επιπλέον, θα μελετήσουμε τις μετρικές που θα εφαρμόσουμε κατά την εκπαίδευση.

4.1 Γενική Μορφή Δεδομένων

Το πρόβλημα του Image Segmentation απαιτεί την επεξεργασία και την κατάτμηση των εικόνων που δίνονται, έτσι ώστε να διαφοροποιηθούν τα διάφορα στοιχεία που απεικονίζονται σε αυτές. Ο τρόπος με τον οποίο μια εικόνα ωστόσο γίνεται αντιληπτή στην προκειμένη περίπτωση, είναι με την χρήση πινάκων δύο διαστάσεων.

Πιο συγκεκριμένα, μια εικόνα της μορφής *RBG* διαστάσεων $W \times H$ ουσιαστικά αποτελείται από 3 διδιάστατους πίνακες ιδίων διαστάσεων, όπου κάθε ένας από αυτούς αντιπροσωπεύει κάθε ένα από τα βασικά χρώματα, το κόκκινο, το μπλε και το πράσινο. Έτσι, στην θέση (i, j) κάθε ενός από τους παραπάνω πίνακες, αντιστοιχεί η ποσότητα του κάθε χρώματος η οποία αντιστοιχεί σε κάθε ένα τέτοιο pixel της συνολικής εικόνας που διαθέτουμε.

Αυτό το είδους εικόνες είναι το κύριο υλικό, πάνω στο οποίο τα μοντέλα μας θα επεξεργαστούν και θα εκπαιδευτούν και θα καθορίσουν κατά ένα μεγάλο ποσοστό την αποτελεσματικότητα του μοντέλου μας. Αυτό γίνεται διότι μέσω της συγκεκριμένης μορφής τους, τα Συνελικτικά επίπεδα, παράγουν τις εξόδους τους με βάση τα στοιχεία των παραπάνω πινάκων.

Το υπόλοιπο βασικό μέρος της εκπαίδευσης του μοντέλου επέρχεται από τον στόχο που έχουν τα μοντέλα μας. Πιο συγκεκριμένα, τα μοντέλα θέλουμε να παράγουν εικόνες οι οποίες θα απεικονίζουν κατάλληλα τις περιοχές των αντικειμένων που μας ενδιαφέρουν. Για να καταφέρουν να το κάνουν αυτό, κατά την εκπαίδευση τους, τα αποτελέσματα τους συγκρίνονται με εικόνες που απεικονίζουν ήδη το σωστό αποτέλεσμα της διάκρισης των στοιχείων, τις λεγόμενες ” αληθείς εικόνες ” ή αλλιώς μάσκες. Μέσω της διαφοράς των εικόνων εξόδου με τις είδη γνωστές μάσκες των εικόνων, τα μοντέλα αλλάζουν κατάλληλα

τις παραμέτρους τους ώστε πλησιάσει όσο το δυνατόν περισσότερο στην γνωστή μάσκα, αλλά ταυτόχρονα να το καταφέρνουν αυτό για το μεγαλύτερο ποσοστό εικόνων, ώστε να μην έχουμε overfitting.

Επομένως, για να καταφέρουμε να εκπαιδεύσουμε τα μοντέλα μας χρειαζόμαστε:

1. **Εικόνες εκπαίδευσης**: Οι εικόνες με τις οποίες θα εκπαιδευτεί το μοντέλο μας και τις δέχεται ως είσοδο.

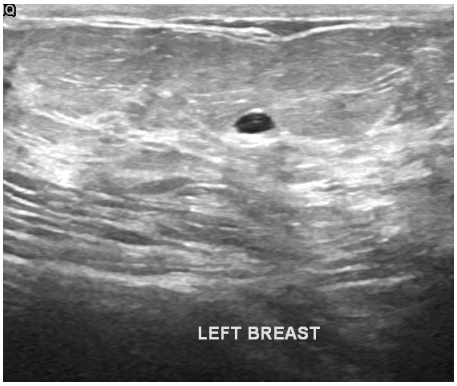
2. **Αληθείς Εικόνες / Μάσκες**: Τα τελικά αποτελέσματα τα οποία θέλουμε να προσεγγίσουμε με τα μοντέλα μας.

4.2 Το Σύνολο Δεδομένων

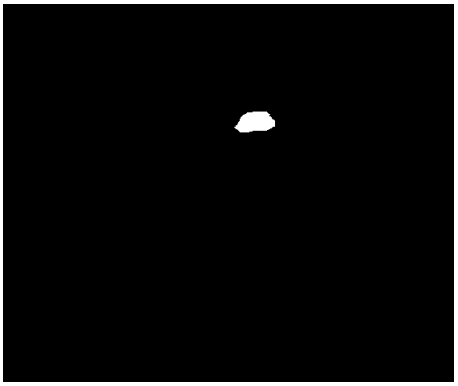
Για την συγκεκριμένη εργασία, θα χρησιμοποιήσουμε ως σύνολο δεδομένων το *Dataset Of Breast Ultrasound Images* [24]. Το συγκεκριμένο dataset δημιουργήθηκε το 2018-19 από κλινικές μελέτες οι οποίες έγιναν πάνω σε 600 γυναίκες, οι οποίες υποβλήθηκαν σε εξέταση υπερήχου στην περιοχή του στήθους. Η ηλικία των συμμετεχόντων κυμαίνεται από 25 εως 75 ετών. Συνολικά υπάρχουν 780 αρχικές εικόνες, όπου για την καθεμία υπάρχει και η αντίστοιχη μάσκα, στην οποία αντιστοιχίζεται η κατάλληλη περιοχή η οποία υποδηλώνει την ύπαρξη καρκινικού όγκου. Η συλλογή των δεδομένων έγινε από το Baheya hospital, όπου η επεξεργασία των αρχικών εικόνων και η δημιουργία των μασκών έγινε από το προσωπικό του νοσοκομείου και τους αρμόδιους καθηγητές του Πανεπιστημίου του Καίρου. Επιπροσθέτως οι εικόνες κατηγοριοποιούνται και σε 3 κατηγορίες, ανάλογα με το είδος των αποτελεσμάτων του υπερήχου. Οι κατηγορίες αυτές είναι:

- Benign: Ο όγκου είναι καλοήθης.
- Malignant: Ο όγκος είναι κακοήθης.
- Normal: Δεν υπάρχει όγκος.

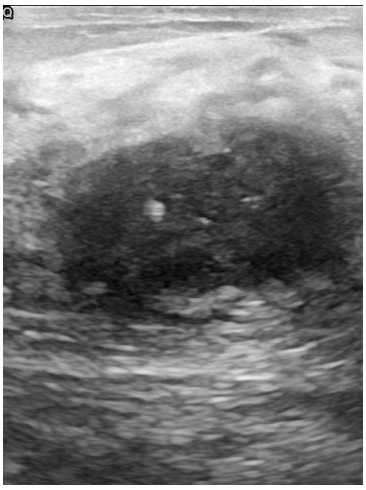
Μερικά παραδείγματα ανά κατηγορία από τις εικόνες που διαθέτουμε είναι τα εξής:



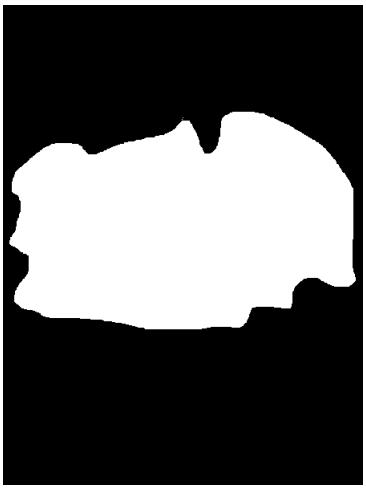
Εικόνα 4.1: *Εικόνα Κατηγορίας Bening*



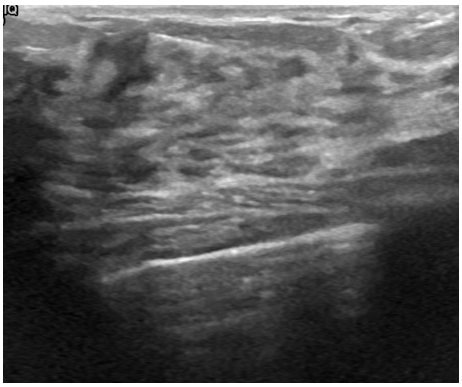
Εικόνα 4.2: *Μάσκα εικόνας*



Εικόνα 4.3: *Εικόνα Κατηγορίας Maligant*



Εικόνα 4.4: *Μάσκα εικόνας*



Εικόνα 4.5: *Εικόνα Κατηγορίας Normal*

Εικόνα 4.6: *Μάσκα εικόνας*

4.3 Μετρικές Εκπαίδευσης

Για να μπορέσει ένας αλγόριθμος Τεχνητής Νοημοσύνης να εκπαιδευτεί κατάλληλα, θα πρέπει τα αποτελέσματα που παράγει με βάση το σύνολο εκπαίδευσης, να είναι όσο το δυνατό όμοια με τα ήδη γνωστά αποτελέσματα που γνωρίζουμε για αυτά [25]. Για να υπολογίσουμε αυτή την ομοιότητα, χρησιμοποιούμε τις λεγόμενες *Μετρικές Συναρτήσεις (Metrics Functions)*.

Για την κατηγορία του θέματος μας, θέλουμε οι μετρικές που θα χρησιμοποιήσουμε, να συγκρίνουν σε επίπεδο pixel, το κατά πόσο η μάσκα που παράχθηκε από το μοντέλο μας, ταυτίζεται με την μάσκα που αληθεύει για την εικόνα που έλαβε. Έτσι χρησιμοποιήσαμε δύο βασικές μετρικές, την **Intersection over Union (IoU)** και την **Dice Coefficient**.

4.3.1 Intersection over Union (IoU)

Η βασική ιδέα στην δομή της πατάει πάνω στην γενικότερη μετρική **Jaccard Index** ή αλλιώς **Jaccard Similarity Coefficient**, με την οποία μετράμε την ομοιότητα που έχουν δύο σύνολα στοιχείων.

Ο τρόπος με τον οποίο ορίζεται ο Jaccard Index δύο συνόλων A, B είναι ο εξής:

$$JaccardIndex(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

Ο τρόπος έκφρασης αυτής της μετρικής της δίνει την ιδιότητα να λαμβάνει τιμές ανάμεσα στο σύνολο [0, 1]. Τιμές κοντά στο 0 σημαίνουν πως τα σύνολα δεν είναι αρκετά όμοια, ενώ αντιθέτως, τιμές κοντά στο 1 υποδεικνύουν την ομοιότητα των συνόλων μας. Αν η μετρική είναι ίση με 0 τότε τα σύνολα είναι ξένα, ενώ αν είναι ίση με 1, τότε τα σύνολα ταυτίζονται.

Αυτή την δομή και λειτουργικότητα προσπαθεί να υλοποιήσει και η μετρική IoU. Αν διαθέτουμε δύο μάσκες A και B για μια εικόνα, τότε προσπαθεί να βρει, κατά πόσο οι δυο αυτές μάσκες είναι ίδιες και " συμβολίζουν " το ίδιο μέρος της εικόνας.

Ο τρόπος με τον οποίο ορίζεται η IoU για δύο μάσκες A και B είναι ο εξής:

$$IoU = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

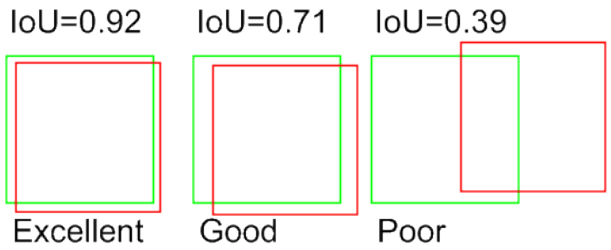
όπου η τομή και η ένωση των μασκών μπορεί να υπολογιστεί σε επίπεδων pixel.

Με όρους στατιστικής, ο τρόπος με τον οποίο ορίζεται η IoU είναι ο εξής:

$$IoU = \frac{TP}{TP + FP + FN}$$

όπου ως TP = True Positive θεωρούμε τα pixels της μάσκας που υποδεικνύουν σωστά την παρουσία όγκου, FP = False Positive θεωρούμε τα pixels της μάσκας που υποδεικνύουν εσφαλμένα παρουσία όγκου, ενώ αυτός δεν υπάρχει και FN = False Negative θεωρούμε τα pixels της μάσκας που υποδεικνύουν εσφαλμένα απουσία όγκου, ενώ αυτός υπάρχει.

Ένα παράδειγμα την μετρικής και των αποτελεσμάτων της είναι το εξής:



Εικόνα 4.7: *Παραδείγματα IoU*

4.3.2 Dice Coefficient

Η δεύτερη μετρική με την οποία θα αξιολογηθούν τα αποτελέσματα των μοντέλων μας είναι η Dice Coefficient και πρόκειται για μια μετρική όμοια με την IoU, ωστόσο η διαχείριση της είναι ευκολότερη.

Ο τρόπος με τον οποίο ορίζεται η Dice Coefficient δύο συνόλων A, B είναι ο εξής:

$$DiceCoefficient(A, B) = \frac{2|A \cap B|}{|A| + |B|}$$

όπου η τομή των μασκών μπορεί να υπολογιστεί σε επίπεδων pixel.

Με όρους στατιστικής, ο τρόπος με τον οποίο ορίζεται η Dice Coefficient είναι ο εξής:

$$DiceCoefficient = \frac{2TP}{2TP + FP + FN}$$

με όμοια σημασία στον συμβολισμό όπως πριν.

Όπως μπορούμε να καταλάβουμε με βάση τους τύπους που παραθέσαμε παραπάνω, οι δύο μετρικές είναι θετικά συσχετισμένες. Αυτό σημαίνει πως αν έχουμε 2 ταξινομητές A και B και συγκριθούν με την μετρική IoU και αποφανεί πως ο ταξινομητής A παράγει καλύτερα αποτελέσματα από τον ταξινομητή B, το ίδιο ακριβώς αποτέλεσμα θα παραχθεί και από την σύγκρισή τους με την μετρική Dice Coefficient.

Ωστόσο η χρήση της Dice Coefficient είναι ευκολότερη, διότι είναι πιο ελαστική στον τρόπο με τον οποίο διαχειρίζεται της εσφαλμένες προβλέψεις. Αυτό συμβαίνει λόγω των παρανομαστών που έχουμε. Στην περίπτωση της IoU, ο παρανομαστής γίνεται αρκετά μεγάλος όταν οι δύο παραγόμενες μάσκες έχουν κοινά σημεία και σε συνδυασμό με την ήδη μικρή τιμή του αριθμητή, λόγω της εσγαλμένης πρόβλεψης, η μετρική παίρνει αρκετά μικρές τιμές. Αντιθέτως, στην περίπτωση της Dice Coefficient, ο παρανομαστής δεν επηρεάζεται από το παραπάνω γεγονός, με αποτέλεσμα ο παρανομαστής να μην οδηγεί σε πολύ μικρές τιμές τις εσφαλμένες προβλέψεις.

Κεφάλαιο 5

Εκπαίδευση Αλγορίθμων Βαθιάς Μάθησης

Στο κεφάλαιο αυτό περιγράφεται η υλοποίηση των αλγορίθμων, με βάση την μελέτη που παρουσιάσαμε στο προηγούμενο κεφάλαιο. Αρχικά παρουσιάζονται τα κατάλληλα προγραμματιστικά εργαλεία τα οποία χρησιμοποιήθηκαν. Στη συνέχεια θα αναλύσουμε αναλυτικά την δομή των αλγορίθμων καθώς και χρήσιμες λεπτομέρειες για την εκπαίδευση τους. Τέλος, θα συγκρίνουμε την απόδοση των μοντέλων μας

5.1 Προγραμματιστικά Εργαλεία

Για την παρούσα εργασία, υλοποιήσαμε τα μοντέλα μας με βασικό εργαλείο την γλώσσα προγραμματισμού Python. Στην συνέχεια χρησιμοποιήσαμε κατάλληλες βιβλιοθήκες, με τις οποίες επεξεργάστηκαν τα στοιχεία του συνόλου δεδομένων και κατασκευάστηκαν - εκπαιδεύτηκαν οι αλγόριθμοι Βαθιάς Μάθησης που μελετούνται. Αυτές οι βιβλιοθήκες είναι οι εξής:

- Γενικές βιβλιοθήκες:
  - Os
  - Glob
- Για την επεξεργασία δεδομένων:
  - Numpy
  - Pandas
- Για την απεικόνιση εικόνων και αποτελεσμάτων:
  - Matplotlib
- Για την κατασκευή των μοντέλων:
  - Tensorflow
  - Keras

5.2 Το μοντέλο U-Net

Το πιο βασικό μοντέλο που υπάρχει για προβλήματα Image Segmentation πάνω σε ιατρικές εικόνες είναι το U - Net. Θα μελετήσουμε την δομή που υλοποιήσαμε ώστε να εκπαιδευτεί για το παραπάνω σύνολο δεδομένων και θα παρουσιάσουμε τα αποτελέσματά του.

5.2.1 Δομή του U-Net

Στην παρούσα εργασία, όπως αναφέρθηκε και στο θεωρητικό κομμάτι, το μοντέλο μας αποτελείται από τα 2 βασικές δομές, αυτήν του *Κωδικοποιητή Encoder* και αυτή του *Αποκωδικοποιητή Decoder*.

Η δομή του Κωδικοποιητή είναι υπεύθυνη για την εξαγωγή των κατάλληλων χαρακτηριστικών τα οποία εστιάζονται σε τοπικές περιοχές της αρχικής εικόνας και εξάγουν κατάλληλα χαρακτηριστικά από αυτή.

Η δομή του Αποκωδικοποιητή είναι υπεύθυνη για την σύνθεση του ζητούμενου αποτελέσματος, με τον κατάλληλο συνδυασμό χαρακτηριστικών τα οποία προέρχονται από δυο διαφορετικές αναλύσεις της εικόνας. Αυτό έχει ως στόχο την δυνατότητα γενίκευσης των πιο εξειδικευμένων χαρακτηριστικών μεγαλύτερων διαστάσεων, να προβληθούν κατάλληλα στις μικρότερες διαστάσεις, με σκοπό την τελική απεικόνιση στην ζητούμενη μάσκα.

Η λεπτομερής σύσταση του συνολικού μοντέλου είναι η εξής:

Πίνακας 5.1: *Δομή Κωδικοποιητή (Encoder Block)*

ΑΑ	Είδος Επιπέδου	Επίπεδο Λήψης Εισόδου
1	Convolution Layer 1	Input Layer
2	Dropout Layer	Convolution Layer 1
3	Convolution Layer 2	Dropout Layer
4	Pooling Layer	Convolution Layer 2

Πίνακας 5.2: *Δομή Αποκωδικοποιητή (Decoder Block)*

ΑΑ	Είδος Επιπέδου	Επίπεδο Λήψης Εισόδου
1	Upsample Layer	Input from linked Decoder Block
2	Concatenate	Input from linked Encoder Block & Upsample Layer
3	Encoder Block	Concatenate

Πίνακας 5.3: *Δομή Μοντέλου U-Net*

Δομή	Είδος επιπέδου	Μέγεθος Εξόδου	# Παραμέτρων	Επίπεδο Λήψης Εισόδου
Κωδικοποιητής	Input Layer	(256,256,3)	0	-
	Encoder 1	(128,128,32)	10.144	Input Layer
	Encoder 2	(64,64,64)	55.424	Encoder 1
	Encoder 3	(32,32,128)	221.440	Encoder 2
	Encoder 4	(16,16,256)	885.248	Encoder 3
	Encoder 5	(16,16,512)	3.539.968	Encoder 4
Αποκωδικοποιητής	Decoder 1	(32,32,256)	2.359.808	Encoders 4 - 5
	Decoder 2	(64,64,128)	590.080	Decoder 1- Encoder 3
	Decoder 3	(128,128,64)	147.584	Decoder 2- Encoder 2
	Decoder 4	(256,256,32)	36.928	Decoder 3- Encoder 1
Έξοδος	Convolution Layer	(256,256,1)	32	Decoder 4

Το σύνολο των παραμέτρων είναι 7.846.657 εκ των οποίων όλοι είναι προς εκπαίδευση.

5.3 Το μοντέλο Attention U-Net

Όπως είναι λογικό, τα σημεία του υπερηχογραφήματος τα οποία έχουν καρκινικούς όγκους, αποτελούνται από pixels τα οποία όσα είναι συσχετισμένα μεταξύ τους, μιας και αποτελούν το ίδιο αντικείμενο, τον όγκο. Λόγω αυτού, ο μηχανισμός Attention εισέρχεται στο βασικό μοντέλο U-Net για να φανερώσει τις συσχετίσεις αυτές και έτσι η μάσκα που το μοντέλο προβλέπει να είναι πιο ακριβής.

5.3.1 Δομή του Attention U-Net

Όπως αναφέρθηκε και παραπάνω, ο μηχανισμός Attention εντάσσεται στο μοντέλο U-Net. Το σημείο στο οποίο θέλουμε να κάνουμε πιο φανερές αυτές τις ιδιότητες των στοιχείων της εικόνας μας, είναι κατά την ανακατασκευή της εικόνας. Λόγω αυτού, ο μηχανισμός αυτός μπαίνει στο σημείο όπου προσπαθούμε να συνενώσουμε τα χαρακτηριστικά του Encoder με αυτά του Decoder, έτσι ώστε να γίνει η κατάλληλη επεξεργασία των pixels και να κατασκευαστεί ένα καλύτερο αποτέλεσμα του segmentation.

Οι δομές του Encoder και του Decoder είναι ίδιες με αυτές στο U-Net. Η λεπτομερής δομή του μηχανισμού Attention καθώς και του μοντέλου είναι οι εξής:

Πίνακας 5.4: *Δομή Μηχανισμού Attention (Attention Gate)*

ΑΑ	Είδος Επιπέδου	Επίπεδο Λήψης Εισόδου
1.1	Convolution Layer 1.1	Input from linked Decoder Block
1.2	Convolution Layer 1.2	Input from linked Encoder Block
2	Add	Convolution Layers 1.1 & 1.2
3	Convolution Layer 2	Add Layer
4	Upsample Layer	Convolution Layer 2
5	Multiply Layer	Convolution Layer 2 & Encoder Block
6	BachNorm Layer	Multply Layer

Πίνακας 5.5: *Δομή Μοντέλου*

Δομή	Είδος επιπέδου	Μέγεθος Εξόδου	# Παραμέτρων	Επίπεδο Λήψης Εισόδου
Κωδικοποιητής	Input Layer	(256,256,3)	0	-
	Encoder 1	(128,128,32)	10.144	Input Layer
	Encoder 2	(64,64,64)	55.424	Encoder 1
	Encoder 3	(32,32,128)	221.440	Encoder 2
	Encoder 4	(16,16,256)	885.248	Encoder 3
	Encoder 5	(16,16,512)	3.539.968	Encoder 4
Αποκωδικοποιητής	Attention 1	(32,32,256)	1.771.265	Encoders 4 - 5
	Decoder 1	(32,32,256)	2.359.808	Attention 1 - Encoders 5
	Attention 2	(64,64,128)	443.265	Encoders 3 - Decoder 1
	Decoder 2	(64,64,128)	590.080	Attention 2 - Encoder 3
	Attention 3	(128,128,64)	111.041	Encoders 2 - Decoder 2
	Decoder 3	(128,128,64)	147.584	Attention 3 - Encoder 2
	Attention 4	(256,256,32)	27.873	Encoder 1 - Decoder 3
	Decoder 4	(256,256,32)	36.928	Encoder 1 - Attention 4
Έξοδος	Convolution Layer	(256,256,1)	32	Decoder 4

Πίνακας 5.6: *Δομή Μοντέλου Attention U-Net*

Το σύνολο των παραμέτρων είναι 10.200.101, εκ των οποίων οι 10.199.141 είναι προς εκπαίδευση, και οι υπόλοιποι 960 είναι σταθερές.

5.4 Χαρακτηριστικά Εκπαίδευσης Μοντέλων

Η εκπαίδευση των μοντέλων έγινε με τα εξής βασικά στοιχεία:

- Ως Loss Function χρησιμοποιήσαμε την συνάρτηση Binary Crossentropy, μιας και το τελικό αποτέλεσμα αποτελείται από δύο στοιχεία, την περιοχή του όγκου και την υπόλοιπη περιοχή.
- Ως Optimizer χρησιμοποιήσαμε τον Adam Optimizer.
- Ως μετρικές χρησιμοποιούμε τις συναρτήσεις Accuracy, IoU και την Dice Coefficient.

Αρχικά χωρίζουμε το σύνολο των δεδομένων μας σε Training set και Validation set, με το 80% να είναι οι προς εκπαίδευση εικόνες και το υπόλοιπο να αποτελείται από τις εικόνες όπου θα δοκιμάσουμε την αποτελεσματικότητα του μοντέλου μας. Επίσης, κατά την εκπαίδευση εκπαίδευση, χρησιμοποιείται το 20% του Training set ως Validation set. Αυτό σημαίνει πως κατά την εκπαίδευση, από τις συνολικά 780 εικόνες, θα χρησιμοποιηθούν οι 624 για να εκπαιδευτούν και να ενημερωθούν τα βάρη των μοντέλων, και οι υπόλοιπες 156 θα χρησιμοποιηθούν μόνο για τον έλεγχο του μοντέλου, χωρίς να εκπαιδευτούν τα μοντέλα πάνω σε αυτές.

Τέλος, η εκπαίδευση των μοντέλων μας έγινε με ένα πιο σύνθετο τρόπο. Λόγω του μικρού μεγέθους που έχει το training set, ελλοχεύει ο κίνδυνος του overfitting. Αυτό συμβαίνει διότι η μεγάλη πολυπλοκότητα του μοντέλου μας, το κάνει ικανό να προσαρμόζει πάρα πολύ καλά τα βάρη των παραμέτρων του πάνω στο Training set. Άμεση συνέπεια αυτού, είναι η έλλειψη της ικανότητάς του για γενίκευση, και κατα επέκταση την αδυναμία του να παράγει πιο ακριβής μάσκες σε νέα δεδομένα που δέχεται, εκτός του Training set.

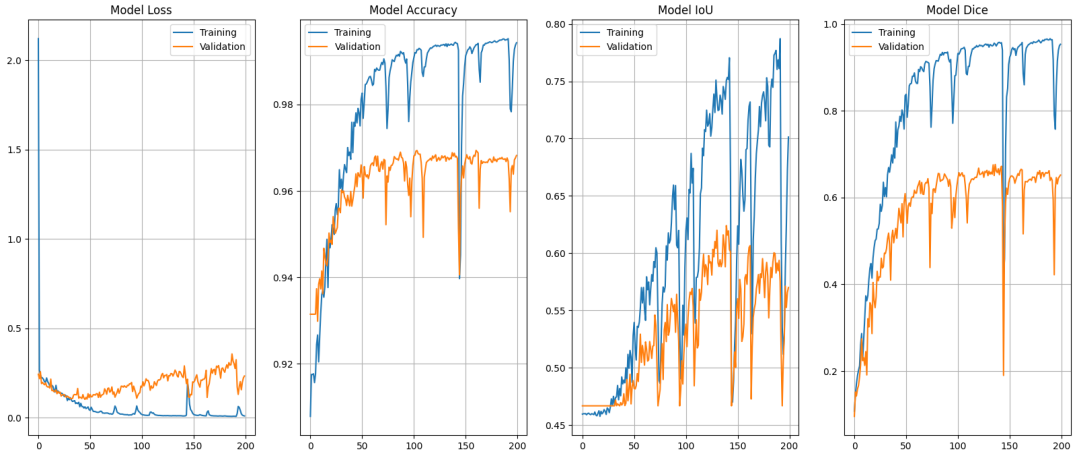
Λόγω αυτού, η κλασσική ρουτίνα εκπαίδευσης ολόκληρου του μοντέλου και εξέτασης αποτελεσμάτων στο Test set δεν είναι αποδοτική. Έτσι θα διαμερίσουμε την εκπαίδευση των μοντέλων μας, με βάση την δομή τους.

5.4.1 Χαρακτηριστικά Εκπαίδευσης του Μοντέλου UNET

Η αρχική εκπαίδευση του μοντέλου UNET γίνεται πάνω σε ολόκληρο το μοντέλο. Για 50 εποχές, εκπαιδεύουμε το μοντέλο, και έτσι ο Κωδικοποιητής, που εν γένει είναι ένα CNN, εκπαιδεύεται στο να επεξεργάζεται σωστά την εικόνα εισόδου, ωστόσο ο Αποκωδικοποιητής δεν είναι σε θέση να αποδώσει σωστές προβλέψεις. Αυτό το καταλαβαίνουμε από την σταθερή τιμή της μετρικής IoU πάνω στο Validation set.

Στην συνέχεια, αποτρέπουμε στον Κωδικοποιητή να ανανεώνει τα βάρη του και εκπαιδεύουμε το μοντέλο μας για 150 επιπλέον εποχές. Η δεύτερη αυτή εκπαίδευση, δίνει την δυνατότητα στον Αποκωδικοποιητή, τον πυρήνα της κατασκευής των μασκών, να καταφέρει να παράγει καλύτερα αποτελέσματα. Πράγματι, όπως φαίνεται και στην Εικόνα 5.1 , παρατηρούμε αύξηση στις τιμές των μετρικών μας, και έτσι το μοντέλο UNET εκπαιδεύεται ορθά.

Παρακάτω παραθέτουμε τα σχεδιαγράμματα με τις τιμές των μετρικών κατά την εκπαίδευση του μοντέλου UNET.



Εικόνα 5.1: *Σχεδιαγράμματα Εκπαίδευσης του UNET*

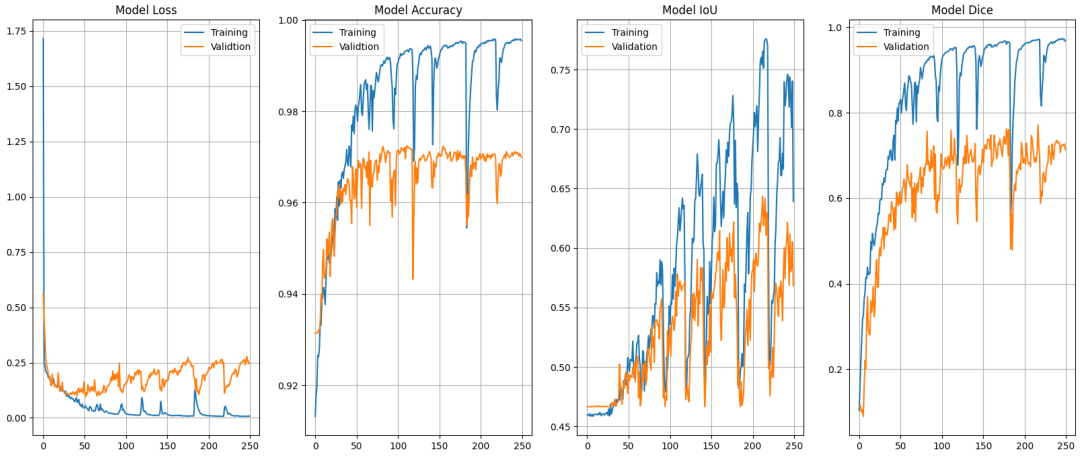
Τα ανεβοκατεβάσματα στις τιμές των μετρικών οφείλονται στις αλλαγές των βαρών του μοντέλου κατά την εκπαίδευση. Από αυτά, μαζί με το γεγονός ότι στην συνέχεια οι τιμές των μετρικών βελτιώνονται καλύτερα σε σχέση με τις προηγούμενες τιμές συνολικά, επαληθεύουν την πολυπλοκότητα του μοντέλου μας, μιας και οι παράμετροι που είναι προς εκπαίδευση είναι μεγάλοι το πλήθος όπως είδαμε.

5.4.2 Χαρακτηριστικά Εκπαίδευσης του Μοντέλου Attention UNET

Η αρχική εκπαίδευση του μοντέλου Attention UNET γίνεται πάνω σε ολόκληρο το μοντέλο. Για 50 εποχές, εκπαιδεύουμε το μοντέλο, και έτσι ο Κωδικοποιητής, συμπεριφέρεται όπως και στην εκπαίδευση του UNET Αυτό το καταλαβαίνουμε από την σταθερή τιμή της μετρικής IoU.

Στην συνέχεια, αποτρέπουμε στον Κωδικοποιητή να ανανεώνει τα βάρη του και εκπαιδεύουμε το μοντέλο μας για 150 επιπλέον εποχές. Η δεύτερη αυτή εκπαίδευση, δίνει την δυνατότητα στον Αποκωδικοποιητή και στα Attention Gates, να παράγουν καλύτερα αποτελέσματα. Πράγματι, όπως φαίνεται και στην Εικόνα 5.2 , παρατηρούμε αύξηση στις τιμές των μετρικών μας. Τέλος, για να βελτιώσουμε παραπάνω το μοντέλο μας, αποτρέπουμε και τον Αποκωδικοποιητή να ανανεώνει τα βάρη του, και έτσι μόνο τα Attention Gates ανανεώνουν τα βάρη τους, για 50 εποχές.

Παρακάτω παραθέτουμε τα σχεδιαγράμματα με τις τιμές των μετρικών κατά την εκπαίδευση του μοντέλου Attention UNET.



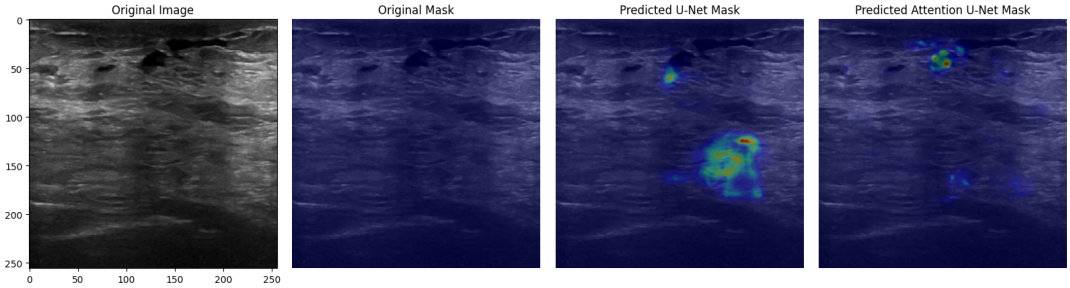
Εικόνα 5.2: *Σχεδιαγράμματα Εκπαίδευσης του Attention UNET*

Όμοια με πριν, οι απότομες μεταβολές στις τιμές των μετρικών οφείλονται στις αλλαγές των βαρών του μοντέλου κατά την εκπαίδευση. Και πάλι, επαληθεύεται η όμοια κλίμακας πολυπλοκότητα του μοντέλου Attention UNETλόγω των μεταβολών που φαίνονται στα παραπάνω γραφήματα, με αυτή του UNET. Ωστόσο ο μηχανισμός Attention φαίνεται πως είναι αποτελεσματικότερος για την παραγωγή καλύτερων αποτελεσμάτων.

5.5 Τελική Σύγκριση Μοντέλων

Στα παρακάτω παραδείγματα, παραθέτουμε τα αποτελέσματα και των δυο μοντέλων, μαζί με τις τιμές των μετρικών, ανά τις τρείς κατηγορίες δεδομένων που διαθέτει το Test set:

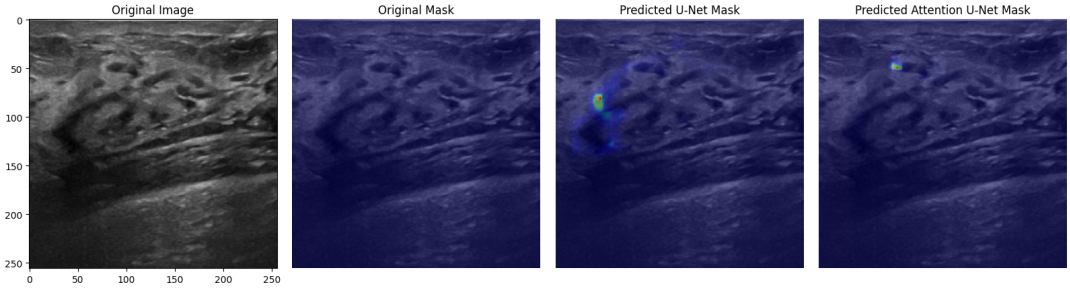
5.5.1 Εικόνες χωρίς καρκινικό όγκο (Normal Images)



Εικόνα 5.3: *Αποτελέσματα Εικόνας 1*

	Iou	Dice Coefficient
U-Net	0.00	0.95
Attention U-Net	0.00	**0.99**

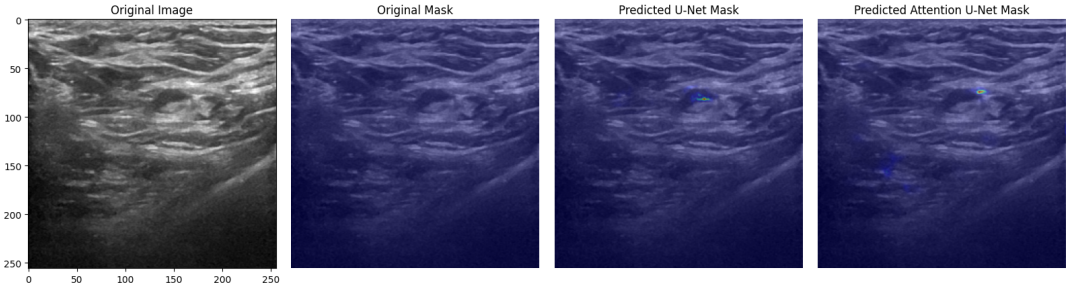
Πίνακας 5.7: *Μετρικές για την Εικόνα 1*



Εικόνα 5.4: *Αποτελέσματα Εικόνας 2*

	Iou	Dice Coefficient
U-Net	0.00	0.97
Attention U-Net	0.00	**0.99**

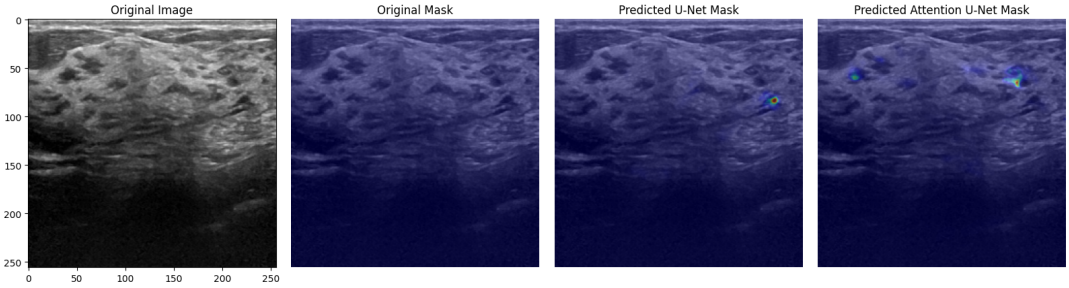
Πίνακας 5.8: *Μετρικές για την Εικόνα 2*



Εικόνα 5.5: *Αποτελέσματα Εικόνας 3*

	Iou	Dice Coefficient
U-Net	0.00	0.98
Attention U-Net	0.00	**0.99**

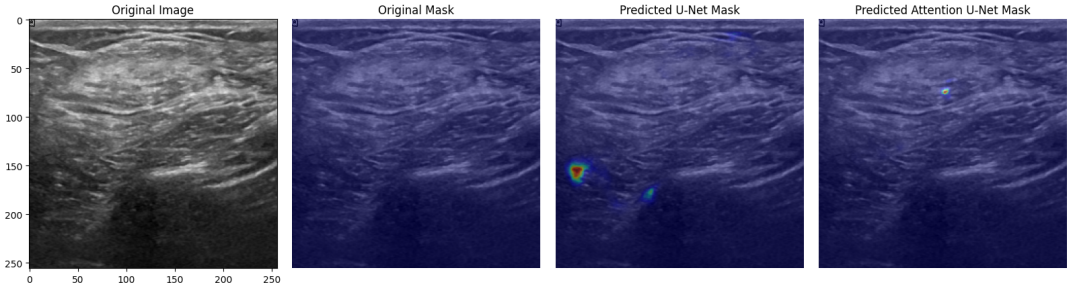
Πίνακας 5.9: *Μετρικές για την Εικόνα 3*



Εικόνα 5.6: *Αποτελέσματα Εικόνας 4*

	Iou	Dice Coefficient
U-Net	0.00	0.99
Attention U-Net	0.00	**0.99**

Πίνακας 5.10: *Μετρικές για την Εικόνα 4*

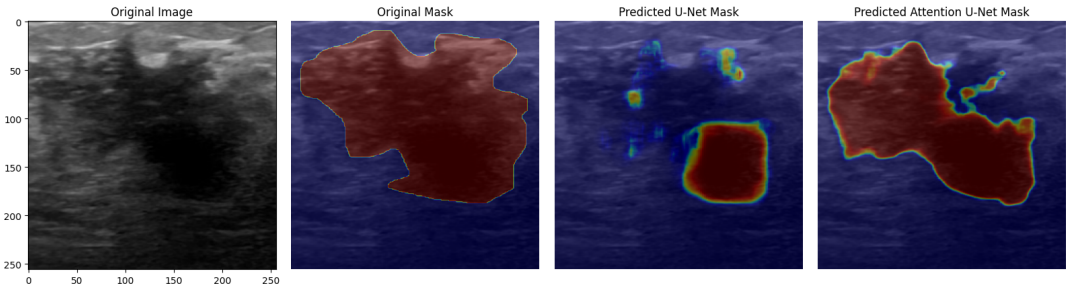


Εικόνα 5.7: *Αποτελέσματα Εικόνας 5*

	Iou	Dice Coefficient
U-Net	0.00	0.99
Attention U-Net	0.00	**0.99**

Πίνακας 5.11: *Μετρικές για την Εικόνα 5*

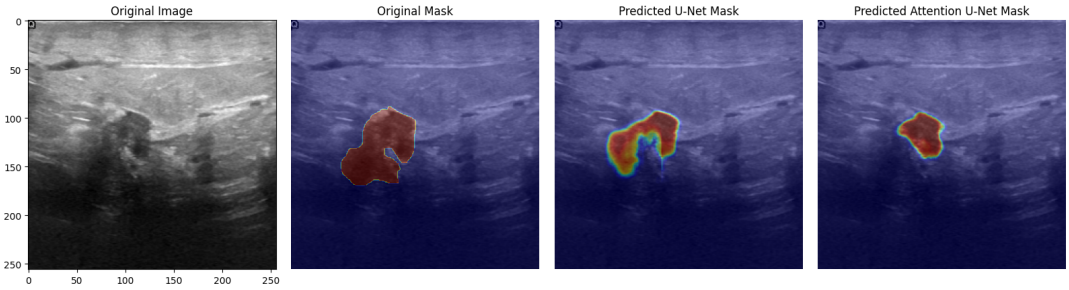
5.5.2 Εικόνες με έναν καρκινικό όγκο (Single Images)



Εικόνα 5.8: *Αποτελέσματα Εικόνας 6*

	Iou	Dice Coefficient
U-Net	0.23	0.38
Attention U-Net	**0.57**	**0.73**

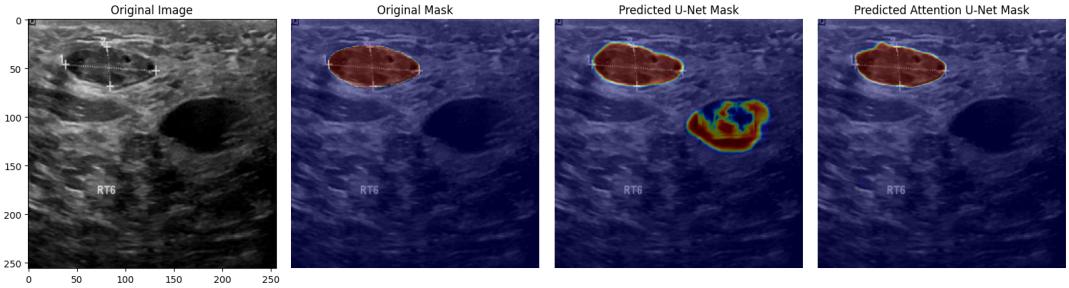
Πίνακας 5.12: *Μετρικές για την Εικόνα 6*



Εικόνα 5.9: *Αποτελέσματα Εικόνας 7*

	Iou	Dice Coefficient
U-Net	**0.44**	**0.61**
Attention U-Net	0.32	0.50

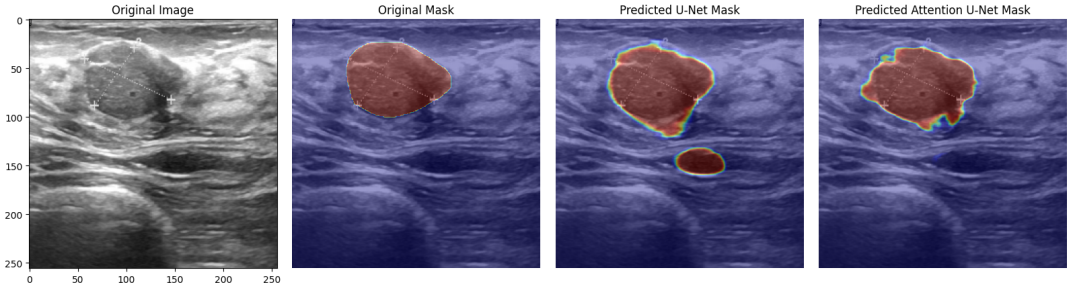
Πίνακας 5.13: *Μετρικές για την Εικόνα 7*



Εικόνα 5.10: *Αποτελέσματα Εικόνας 8*

	Iou	Dice Coefficient
U-Net	0.52	0.69
Attention U-Net	**0.90**	**0.95**

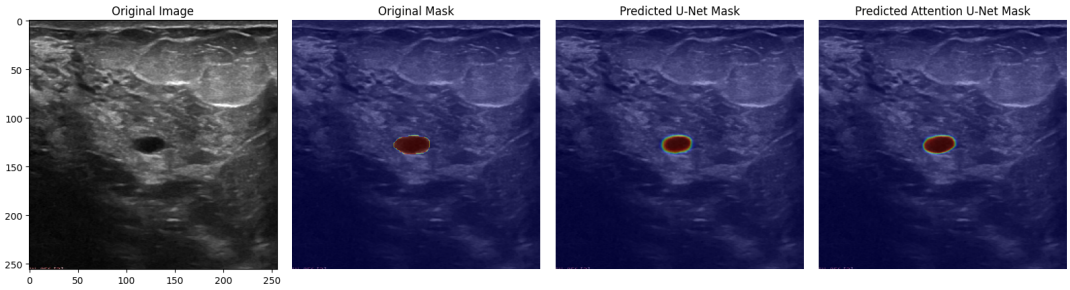
Πίνακας 5.14: *Μετρικές για την Εικόνα 8*



Εικόνα 5.11: *Αποτελέσματα Εικόνας 8*

	Iou	Dice Coefficient
U-Net	0.67	0.80
Attention U-Net	**0.75**	**0.86**

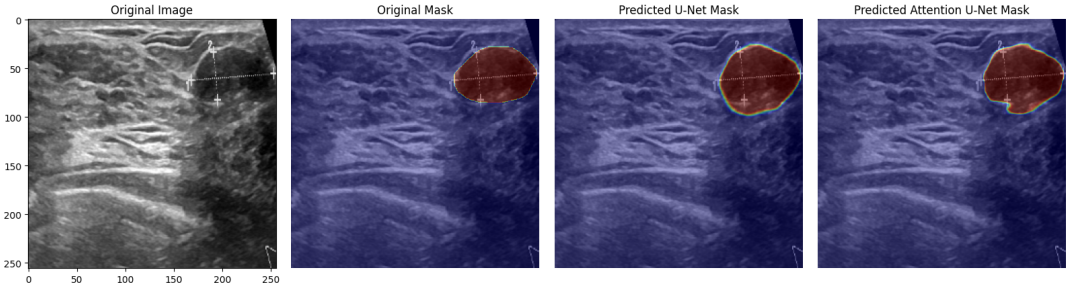
Πίνακας 5.15: *Μετρικές για την Εικόνα 8*



Εικόνα 5.12: *Αποτελέσματα Εικόνας 9*

	Iou	Dice Coefficient
U-Net	0.70	0.84
Attention U-Net	**0.77**	**0.88**

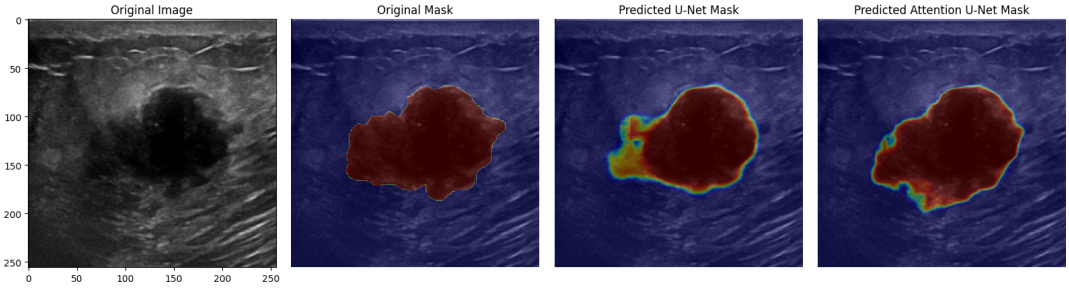
Πίνακας 5.16: *Μετρικές για την Εικόνα 9*



Εικόνα 5.13: *Αποτελέσματα Εικόνας 10*

	Iou	Dice Coefficient
U-Net	0.79	0.88
Attention U-Net	**0.82**	**0.90**

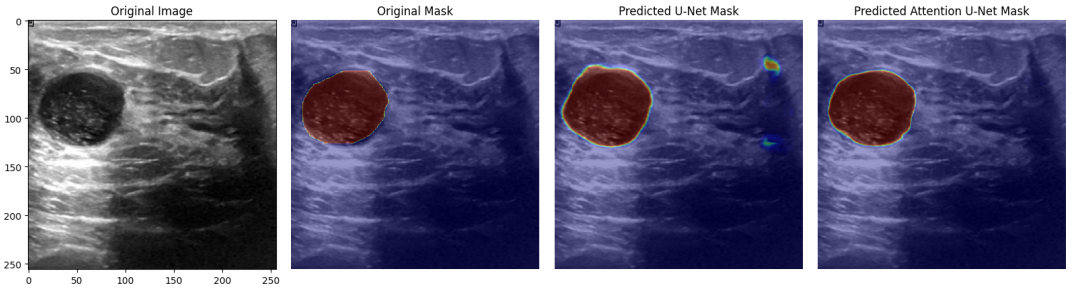
Πίνακας 5.17: *Μετρικές για την Εικόνα 10*



Εικόνα 5.14: *Αποτελέσματα Εικόνας 11*

	Iou	Dice Coefficient
U-Net	0.81	0.90
Attention U-Net	**0.83**	**0.91**

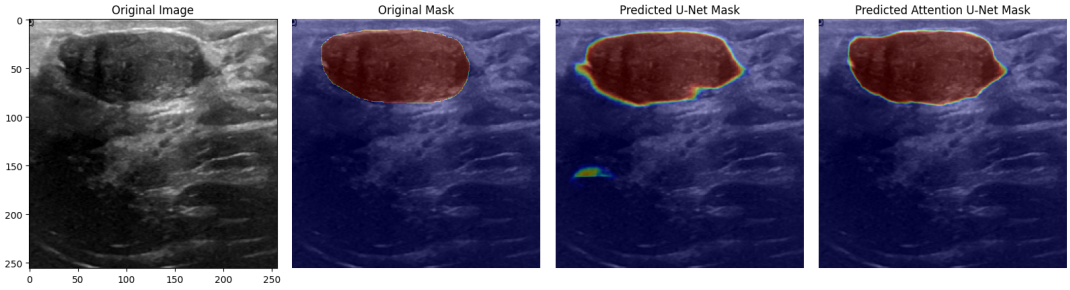
Πίνακας 5.18: *Μετρικές για την Εικόνα 11*



Εικόνα 5.15: *Αποτελέσματα Εικόνας 12*

	Iou	Dice Coefficient
U-Net	0.82	0.90
Attention U-Net	**0.88**	**0.94**

Πίνακας 5.19: *Μετρικές για την Εικόνα 12*

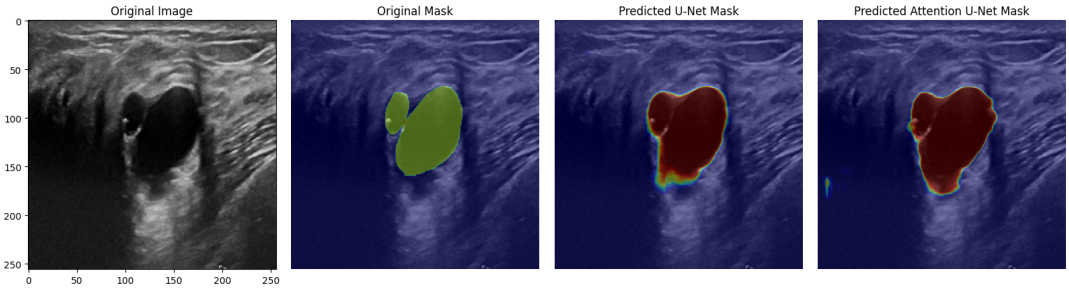


Εικόνα 5.16: *Αποτελέσματα Εικόνας 13*

	Iou	Dice Coefficient
U-Net	0.83	0.91
Attention U-Net	**0.91**	**0.96**

Πίνακας 5.20: *Μετρικές για την Εικόνα 13*

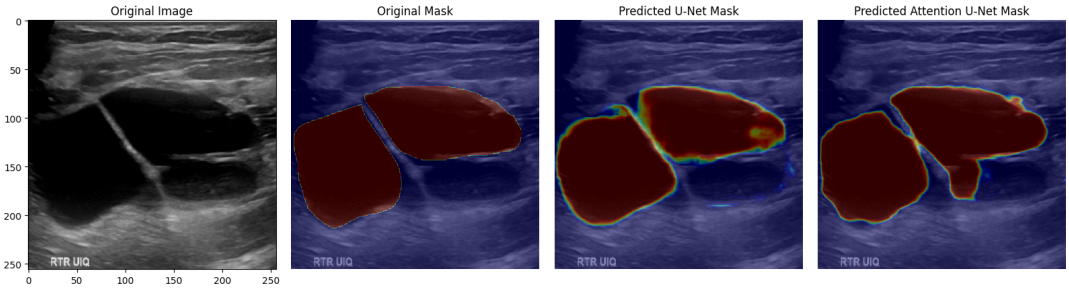
5.5.3 Εικόνες με δύο καρκινικούς όγκους (Double Images)



Εικόνα 5.17: *Αποτελέσματα Εικόνας 14*

	Iou	Dice Coefficient
U-Net	**0.47**	**0.64**
Attention U-Net	0.44	0.61

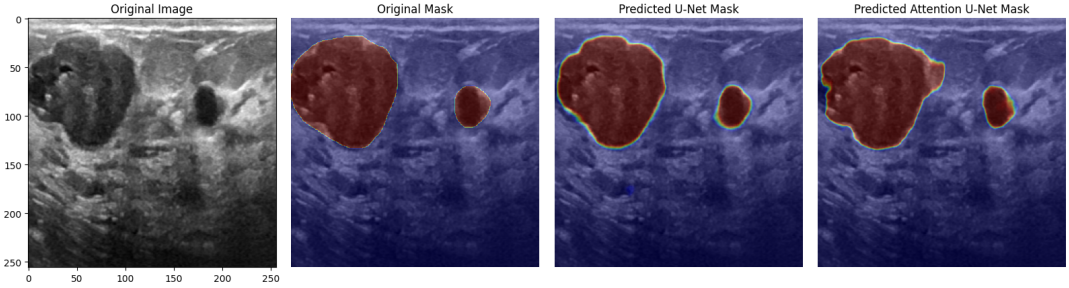
Πίνακας 5.21: *Μετρικές για την Εικόνα 14*



Εικόνα 5.18: *Αποτελέσματα Εικόνας 15*

	Iou	Dice Coefficient
U-Net	0.86	0.92
Attention U-Net	**0.86**	**0.92**

Πίνακας 5.22: *Μετρικές για την Εικόνα 15*



Εικόνα 5.19: *Αποτελέσματα Εικόνας 15*

	Iou	Dice Coefficient
U-Net	**0.88**	**0.94**
Attention U-Net	0.85	0.94

Πίνακας 5.23: *Μετρικές για την Εικόνα 15*

5.5.4 Συνολικά Συμπεράσματα και Αποτελέσματα

Με βάση τα παραπάνω αποτελέσματα, παρατηρούμε πως η υπεροχή του Attention U-Net έναντι του απλού μοντέλου U-Net είναι προφανής. Τα αποτελέσματα στις μετρικές, καθώς και στις εικόνες είναι καλύτερα, μιας και το Attention Layer κατάφερε να δώσει παραπάνω διακριτικές ικανότητες πάνω στις εικόνες που διαθέτουμε.

Παρακάτω, παραθέτουμε τα στατιστικά των μετρικών μας, πάνω σε ολόκληρο το Test set καθώς και με το Test set απομονωμένο από τις εικόνες οι οποίες ανήκουν μόνο στις κατηγορίες των καρκίνων.

	Μέση Τιμή IoU	Ελάχιστη Τιμή IoU	Μέγιστη Τιμή Iou
U-Net	0.58	0.0	0.92
Attention U-Net	0.58	0.0	0.93
	Μέση Τιμή Dice Coef.	Ελάχιστη Τιμή Dice Coef.	Μέγιστη Τιμή Dice Coef.
U-Net	0.73	0.01	0.99
Attention U-Net	0.74	0.01	0.99

Πίνακας 5.24: *Στατιστικά Μετρικών σε ολόκληρο το Test set*

Μπορεί οι μετρικές στα μοντέλα μας να είναι αριθμητικά ίδιες, ωστόσο υπάρχουν αρκετές διαφορές όσο αφορά τα αποτελέσματα που έχουμε ανά εικόνα. Είδαμε παραπάνω στα αποτελέσματα, πως υπάρχουν εικόνες όπου το απλό μοντέλο UNET παράγει καλύτερα αποτελέσματα από το πιο σύνθετο μοντέλο Attention UNET. Λόγω αυτών λοιπόν, οι αριθμητικές τιμές των μετρικών δεν φαίνεται να έχουν μεγάλες διαφορές.

Τέλος, αποτελέσματα αυτά μπορούμε να τα συγκρίνουμε με τα αποτελέσματα από δημοσιεύσεις που εκπαίδευσαν αντίστοιχα μοντέλα πάνω στο ίδιο σύνολο δεδομένων. Οι δημοσιεύσεις αυτές είναι οι εξής:

1. "U-Netmer: U-Net meets Transformer for medical image segmentation" by Sheng He, Rina Bao, P. Ellen Grant, Yangming Ou [26]

2. " Attention guided neural ODE network for breast tumor segmentation in medical images " by Jintao Ru, Beichen Lu, Buran Chen, Jialin Shi, Gaoxiang Chen, Meihao Wang, Zhifang Pan, Yezhi Lin, Zhihong Gao, Jiejie Zhou, Xiaoming Liu, Chen Zhang [27]

3. " CSwin-PNet: A CNN-Swin Transformer combined pyramid network for breast lesion segmentation in ultrasound images " by Haonan Yang , Dapeng Yang [28]

Στους παρακάτω πίνακες, μπορούμε να δούμε τις αποδόσεις των μοντέλων UNET και Attention UNET στις παραπάνω δημοσιεύσεις και των δικών μας μοντέλων:

Μοντέλο	IoU	Dice Coef.
Paper [26]	0.62	0.71
Paper [27]	0.64	0.70
Paper [28]	0.64	0.70
Το μοντέλο μας	0.58	0.73

Πίνακας 5.25: *Σύγκριση μοντέλου UNET*

Μοντέλο	IoU	Dice Coef.
Paper [26]	0.61	0.70
Paper [27]	0.62	0.71
Paper [28]	0.62	0.71
Το μοντέλο μας	0.58	0.74

Πίνακας 5.26: *Σύγκριση μοντέλου Attention UNET*

Παρατηρούμε πως για την μετρική IoU τα μοντέλα μας δεν είναι τόσο αποδοτικά όσο αυτών στις αντίστοιχες δημοσιεύσεις. Ωστόσο, η μετρική Dice Coef. λαμβάνει πολύ καλά αποτελέσματα. Αυτό υποστηρίζει και τα καλά αποτελέσματα που φαίνονται στις παραγόμενες μάσκες από τα μοντέλα μας. Έτσι τα μοντέλα που αναπάραγαμε είναι αποτελεσματικά, για την διάγνωση καρκινικών όγκων.

Μέρος III

Επίλογος

Κεφάλαιο 6

Επίλογος

Στο προηγούμενο κεφάλαιο παραθέσαμε τα αποτελέσματα των μοντέλων, πάνω στο δεδομένα μας, καθώς και μια αρχική σύγκριση μεταξύ τους. Σε αυτό το κεφάλαιο θα παρουσιάσουμε τα τελικά συμπεράσματα και τις μελλοντικές επεμβάσεις πάνω στα μοντέλα μας

6.1 Συμπεράσματα

Όπως φαίνεται και από τα αποτελέσματα του προηγούμενου κεφαλαίου, τα δυο μοντέλα καταφέρνουν με επιτυχία να εντοπίσουν στο μεγαλύτερο μέρος τους, τους καρκινικούς όγκους που υπάρχουν σε ένα εξεταζόμενο υπερηχογράφημα. Τα αποτελέσματα δείχνουν πως σε πολύ μεγάλο μέρος τους, γίνεται ένας καλός εντοπισμός ενός όγκου, με πάνω από το 70% του να αναγνωρίζεται από τους αλγορίθμους που διαθέτουμε.

Αυτό είναι ένα θετικό αποτέλεσμα, διότι από πρακτικής άποψης μπορεί να βοηθήσει την ιατρική κοινότητα σε ένα αρχικό στάδιο ως βοηθητικό διαγνωστικό εργαλείο. Με την διακριτική ικανότητα τους, οι αλγόριθμοι αυτοί μπορούν να φανερώσουν πιθανές περιοχές πάνω σε ένα υπερηχογράφημα που είναι πιθανά σημεία όγκου, οι οποίες στην συνέχεια θα εξεταστούν από τον ιατρό που είναι ειδικότερος στο συγκεκριμένο τομέα.

Παρατηρούμε ωστόσο την αδυναμία των μοντέλων αυτών, πάνω σε υγιής περιοχές οι οποίες εξετάζονται είτε σε περιοχές όπου η μυική τους σύσταση είναι πυκνότερη τοπικά είτε υπάρχουν μαστικοί αδένες. Είδαμε πως όταν έχουμε ένα υπερηχογράφημα το οποίο δεν έχει κάποιον όγκο πάνω του, παράγονται μικρές σε μέγεθος μάσκες οι οποίες από την μια δεν δείχνουν κάποιο συγκεκριμένο τμήμα της εικόνας ως όγκο, ωστόσο υπάρχει μια κάπως στόχευση περιοχών. Ωστόσο αυτό ένας ειδικός μπορεί να μη το λάβει υπόψιν, μιας και φαίνεται ήδη από τα αποτελέσματα, πως η στόχευση αυτή είναι αμυδρή.

Εν κατακλείδι, παρατηρούμε πως τα μοντέλα αυτά, μπορούν να έχουν ένα θετικό αντίκτυπο πάνω στον ιατρικό κλάδο της Ογκολογίας. Με την χρήση τους, μπορούμε να βοηθήσουμε το ιατρικό προσωπικό που είναι υπεύθυνο για τον εντοπισμό του όγκο ως προς την περιοχή που θα πρέπει να μελετήσουν ειδικότερα. Έτσι μπορεί να μειωθεί ο χρόνος εντοπισμού των καρκινικών όγκων, με αποτέλεσμα την αποδοτικότερη αντιμετώπισή τους και τον περιορισμό τους σε πιθανές μεταστάσεις, αν επρόκειτο κιόλας για κακοήθης όγκους.

6.2 Μελλοντικές Επεκτάσεις

Στις μελλοντικές επεκτάσεις του μοντέλου μας, μπορούμε να εφαρμόσουμε τις παρακάτω 2 προτεινόμενες τροποποιήσεις που ίσως βελτιώσουν τις αδυναμίες του μοντέλου μας:

1. Η πρώτη μετατροπή είναι η εφαρμογή κάποιων Activation Functions στο τέλος κάθε επιπέδου του Decoder είτε την χρήση ενός τελικού Activation Function πριν την τελική μάσκα που παράγουν τα μοντέλα μας. Αυτό θα έχει ως συνέπεια τον περιορισμό της αδυναμίας που έχουν τα μοντέλα μας να παράγουν σωστές μάσκες, σε υπερηχογραφήματα τα οποία δεν έχουν κάποιο όγκο.

   Αυτό θα βοηθήσει το μοντέλο να παράγει καλύτερες μάσκες σε υγιείς περιπτώσεις, με άμεσο αποτέλεσμα την αποφυγή την παραγωγή κάποιας εσφαλμένης μάσκας που σε άλλη περίπτωση ίσως να εμφάνιζε λανθασμένα κάποιον υγιή ιστό ως καρκινικό όγκο.

2. Με βάση την εξαγωγή των χαρακτηριστικών που κάνει το μέρος του Encoder, θα μπορούσαμε να επαυξήσουμε το μοντέλο μας με έναν μηχανισμό αναγνώρισης του είδους της εικόνας που έχουμε. Πιο συγκεκριμένα, τα μοντέλα στο σημείο όπου ο Encoder ετοιμάζει τα χαρακτηριστικά που έχει παράγει για να τα εισάγει στον Decoder θα μπορούσε να εφοδιαστεί με έναν απλό ΤΝΔ το οποίο θα ταξινομούσε αν η εικόνα αυτή ανήκει σε κατηγορία που έχει όγκου ή όχι. Αυτή η ταξινόμηση στην συνέχεια, θα επηρέαζε την ανασύνδεση της παραγόμενης μάσκας.

   Αυτό ίσως έκανε αποδοτικότερο τον μηχανισμό παραγωγής των μασκών, διότι αυτή την στιγμή, τα μοντέλα επηρεάζονται στην εκπαίδευση τους και από τις 2 περιπτώσεις(ύπαρξη - μη ύπαρξη) όγκου, ενώ στην επαυξημένο εκδοχή τους, οι μάσκες που θα εντόπιζαν όγκους, θα εκπαιδεύονταν μόνο σε εικόνες που έχουν όντως όγκο.

   Το αρνητικό αυτής της επαύξησης είναι η αύξηση των παραμέτρων εκπαίδευσης των μοντέλων μας, καθώς και την ύπαρξη τελείως λανθασμένης μάσκας. Αν κάποια εικόνα με όγκο ταξινομηθεί λανθασμένα σε εικόνα χωρίς όγκο, η μάσκα που θα παραγόταν θα είναι 100% εσφαλμένη μιας και δεν θα στοχοποιούσε κάποιον όγκο.

Βιβλιογραφία

[1] Philipp Fischer Olaf Ronneberger και Thomas Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. *MICCAI 2015*, σελίδες 234–241, 2015.

[2] Adel Kermi, Issam Mahmoudi και Mohamed Tarek Khadir. *Deep Convolutional Neural Networks Using U-Net for Automatic Brain Tumor Segmentation in Multimodal MRI Volumes*. *International MICCAI Brainlesion Workshop*, σελίδες 37–48. Springer, 2018.

[3] *Cancer Fact sheet No297*. https://www.who.int/en/news-room/fact-sheets/detail/cancer. Ημερομηνία πρόσβασης: 02-2014.

[4] *Defining Cancer*. http://www.cancer.gov/cancertopics/cancerlibrary/what-is-cancer. Ημερομηνία πρόσβασης: 10-06-2014.

[5] Various Authors. *World Cancer Report 2020*. *World Health Organization*, 2022.

[6] *Deaths 1951-2022*. Data published by United Nations - Population Division (2022).

[7] Marcin Rudzki Slawomir Rudzki Barbara Laskowska Anna Maria Lewandowska, Tomasz Lewandowski. *Cancer prevention - review paper*. *Ann Agric Environ Med. 2021;28(1):11-19*, 2021.

[8] Razelle Kurzrock Mina Nikanjam, Shumei Kato. *Liquid biopsy: current technology and clinical applications*. *Journal of Hematology and Oncology*, 2022.

[9] S. Haykin. *Neural Networks: A Comprehensive Foundation*. Prentice Hall, 1998.

[10] Richard S.Snell. *Clinical Neuroanatomy*. Lippincott Williams and Wilkins, 7η έκδοση, 2009.

[11] Kunihiko Fukushima. *Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position*. NHK Broadcasting Science Research Laboratories, Kinuta, Setagaya, Tokyo, Japan, 1980.

[12] Matteo Carandini. *What simple and complex cells compute*. *The Jornal of Physiology*, 2006.

[13] T. N. Wiesel D. H. Hubel. *Receptive fields of single neurones in the cat's striate cortex*. *The Jornal of Physiology*, 1959.

[14] Keiron O'Shea και Ryan Nash. *An Introduction to Convolutional Neural Networks*. *ArXiv e-prints*, 2015.

[15] Ian Goodfellow, Yoshua Bengio και Aaron Courville. *Deep Learning*. MIT Press, 2016. ηττπ://ωωω.δεεπλεαρνινγβοοκ.οργ.

[16] AILS Teaching Team. *Convolytion Neural Networks Slides*, 2021.

[17] Christian Szegedy Sergey Ioffe. *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*. *ArXiv e-prints*, 2015.

[18] Alex Krizhevsky Ilya Sutskever Ruslan Salakhutdinov Nitish Srivastava, Geoffrey Hinton. *Dropout: A Simple Way to Prevent Neural Networks from Overfitting*. *Journal of Machine Learning*, 2014.

[19] Trevor Darrell Jonathan Long, Evan Shelhamer. *Fully Convolutional Networks for Semantic Segmentation*. CVPR (2015), 2015.

[20] Francesco Visin Vincent Dumoulin. *A guide to convolution arithmetic for deep learning*. 23-3-2016.

[21] Niki Parmar Jakob Uszkoreit Llion Jones Aidan N. Gomez Lukasz Kaiser Ashish Vaswani, Noam Shazeer και Illia Polosukhin. *Attention Is All You Need*. NIPS 2017, 2017.

[22] *CBAM: Convolutional Block Attention Module*. ECCV 2018, 2018.

[23] Loic Le Folgoc Matthew Lee Mattias Heinrich Kazunari Misawa Kensaku Mori Steven McDonagh Nils Y Hammerla Bernhard Kainz Ben Glocker Daniel Rueckert Ozan Oktay, Jo Schlemper. *Attention U-Net: Learning Where to Look for the Pancrea*. MIDL 2018, 2018.

[24] Khaled H Fahmy A. Al-Dhabyani W, Gomaa M. *Dataset of breast ultrasound images*. 2019.

[25] Ameet Talwalkar Mehryar Mohri, Afshin Rostamizadeh. *Foundations of Machine Learning*. MIT Press, 2012.

[26] P. Ellen Grant Yangming Ou Sheng He, Rina Bao. *U-Netmer: U-Net meets Transformer for medical image segmentation*. arxiv, 2023.

[27] Buran Chen Jialin Shi Gaoxiang Chen Meihao Wang Zhifang Pan Yezhi Lin Zhihong Gao Jiejie Zhou Xiaoming Liu Chen Zhang Jintao Ru, Beichen Lu. *Attention guided neural ODE network for breast tumor segmentation in medical images*. Computers in Biology and Medicine, 2023.

[28] Dapeng Yang Haonan Yang. *CSwin-PNet: A CNN-Swin Transformer combined pyramid network for breast lesion segmentation in ultrasound images "*. Expert Systems with Applications, 2023.

Συντομογραφίες - Αρκτικόλεξα - Ακρωνύμια

Τεχνητά Νευρωνικά Δίκτυα	ΤΝΔ
Artificial Neural Networks	ANNs
Κεντρικό Νευρικό Σύστημα	ΚΝΣ
Συνελικτικά Νευρωνικά Δίκτυα	ΣΝΔ
Convolutional Neural Networks	CNNs