Εθνικο Μετσοβιο Πολυτεχνειο
Σχολη Ηλεκτρολογων Μηχανικων και Μηχανικων Υπολογιστων
Τομεας Τεχνολογιας Πληροφορικης και Υπολογιστων

Τμηματοποίηση Ιατρικής Εικόνας

Μελέτη και αξιολόγηση σύγχρουων προσεγγίσεων

Διπλωματικη Εργασια

της

Ναταλίας Ε. Σαλπέα

Επιβλέπων: Στέφανος Κόλλιας
Καθηγητής

Αθήνα, Ιούλιος 2022

Εθνικο Μετσοβιο Πολυτεχνειο
Σχολη Ηλεκτρολογων Μηχανικων και Μηχανικων Υπολογιστων
Τομεας Τεχνολογιας Πληροφορικης και Υπολογιστων

Τμηματοποίηση Ιατρικής Εικόνας

Μελέτη και αξιολόγηση σύγχρουων προσεγγίσεων

Διπλωματικη Εργασια

της

Ναταλίας Ε. Σαλπέα

Επιβλέπων: Στέφανος Κόλλιας
Καθηγητής

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 13η Ιουλίου 2022.

(Υπογραφή)	*(Υπογραφή)*	*(Υπογραφή)*
.....	.....	.....
Στέφανος Κόλλιας	Ανδρέας-Γεώργιος Σταφυλοπάτης	Γιώργος Στάμου
Καθηγητής	Καθηγητής	Καθηγητής

Αθήνα, Ιούλιος 2022

Εθνικο Μετσοβιο Πολυτεχνειο
Σχολη Ηλεκτρολογων Μηχανικων και Μηχανικων Υπολογιστων
Τομεας Τεχνολογιας Πληροφορικης και Υπολογιστων

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)

.....
Ναταλία Ε. Σαλπέα
Διπλωματούχος
Ηλεκτρολόγος Μηχανικός
και Μηχανικός
Υπολογιστών Ε.Μ.Π

13 Ιουλίου 2022

Περίληψη

Η τμηματοποίηση ιατρικής εικόνας περιλαμβάνει τον εντοπισμό περιοχών ενδιαφέροντος σε εικόνες ιατρικού περιεχομένου. Στη σύγχρονη εποχή, υπάρχει μεγάλη ανάγκη ανάπτυξης εύρωστων αλγορίθμων όρασης υπολογιστών για την εκτέλεση του συγκεκριμένου έργου με σκοπό την μείωση του χρόνου και του κόστους διάγνωσης και επομένως την γρήγορη πρόληψη και θεραπεία ποικίλων ασθενειών. Οι προσεγγίσεις που έχουν παρουσιαστεί μέχρι στιγμής, κυρίως ακολουθούν την αρχιτεκτονική τύπου U που προτάθηκε με το μοντέλο UNet, υλοποιούν αρχιτεκτονικές τύπου κωδικοποιητή-αποκωδικοποιητή με πλήρως συνελικτικά δίκτυα, αλλά και αρχιτεκτονικές μετασχηματιστή, αξιοποιώντας ταυτόχρονα και μηχανισμούς προσοχής αλλά και υπολειμματικής μάθησης, και δίνοντας έμφαση στην συγκέντρωση πληροφοριών σε διαφορετικές κλίμακες ανάλυσης. Πολλές από αυτές τις παραλλαγές αρχιτεκτονικών, πετυχαίνουν εμφανή βελτίωση των ποσοτικών και ποιοτικών αποτελεσμάτων σχετικά με το πρωτοπόρο UNet, ενώ μερικές δεν καταφέρνουν να ξεπεράσουν την επίδοση του. Στην παρούσα διπλωματική εργασία, εκπαιδεύονται και εξετάζονται 11 μοντέλα σχεδιασμένα για τμηματοποίηση ιατρικής εικόνας αλλά και άλλους τύπους τμηματοποίησης, με κριτήρια συγκεκριμένες μετρικές αξιολόγησης, σε 4 δημόσια διαθέσιμα σύνολα δεδομένων που σχετίζονται με γαστρικούς πολύποδες και πυρήνες κυττάρων, τα οποία πρώτα επαυξάνονται ώστε να αυξηθεί το μέγεθός τους σε μια προσπάθεια να αντιμετωπιστεί το πρόβλημα της έλλειψης μεγάλου πλήθους ιατρικών δεδομένων. Επιπλέον εξετάζεται η ικανότητα γενίκευσης τους καθώς και η επίδραση της επαύξησης των δεδομένων στα scores των πειραμάτων. Τέλος παρατίθενται τα συμπεράσματα περί της επίδοσης των μοντέλων και συζητούνται μελλοντικές επεκτάσεις που μπορούν να βελτιώσουν την επίδοση τους στο έργο της τμηματοποίησης ιατρικής εικόνας.

Λέξεις Κλειδιά

τμηματοποίηση ιατρικής εικόνας, πολύποδες, πυρήνες κυττάρων, δίκτυα κωδικοποιητή-αποκωδικοποιητή, μετασχηματιστές, διεσταλμένη συνέλιξη, συνελικτικά δίκτυα, squeeze and excitation, υπολειμματική μάθηση, μηχανισμός προσοχής, shape stream,UNet, Vnet, Attention UNet, TransUNet, SwinUNet, ResUNet, ResUNet++, DeepLabv3+, ResUNet-a,R2U-Net, MSRF-Net, CVC-ClinicDB, Kvasir-SEG, 2018 Data Science Bowl, SegPC

Abstract

Medical image segmentation involves identifying regions of interest in medical images. In modern times, there is a great need to develop robust computer vision algorithms to perform this task in order to reduce the time and cost of diagnosis and thus to aid quicker prevention and treatment of a variety of diseases. The approaches presented so far, mainly follow the U-type architecture proposed along with the UNet model, implement encoder-decoder type architectures with fully convolutional networks,and also transformer architectures, exploiting both attention mechanisms and residual learning, and emphasizing information gathering at different resolution scales. Many of these architectural variants achieve significant improvements in quantitative and qualitative results in comparison to the pioneer UNet, while some fail to outperform it. In this thesis, 11 models designed for medical image segmentation, and other types of segmentation, are trained and tested, evaluated on specific evaluation metrics, on 4 publicly available datasets related to gastric polyps and cell nuclei, which are first augmented to increase their size in an attempt to address the problem of the lack of a large amount of medical data. In addition, their generalizability and the effect of data augmentation on the scores of the experiments are also examined. Finally, conclusions on the performance of the models are provided and future extensions that can improve their performance in the task of medical image segmentation are discussed.

Keywords

medical image segmentation, polyps, cell nuclei, encoder-decoder, transformers, atrous convolution, convolutional neural networks, squeeze and excitation, residual learning, attention, shape stream, UNet, DeepLabv3+, Vnet, Attention UNet, TransUNet, SwinUNet, ResUNet, ResUNet++, ResUNet-a,R2U-Net, MSRF-Net, CVC-ClinicDB, Kvasir-SEG, 2018 Data Science Bowl, SegPC

στους γονείς μου

Ευχαριστίες

Θα ήθελα καταρχήν να ευχαριστήσω τον καθηγητή κ. Στέφανο Κόλλια για την επίβλεψη αυτής της διπλωματικής εργασίας και για την ευκαιρία που μου έδωσε να την εκπονήσω στο Εργαστήριο Συστημάτων Τεχνητής Νοημοσύνης και Μάθησης. Επίσης ευχαριστώ ιδιαίτερα την κ. Παρασκευή Τζούβελη για την καθοδήγησή της και την εξαιρετική συνεργασία που είχαμε. Τέλος θα ήθελα να ευχαριστήσω τους γονείς μου για την καθοδήγηση και την ηθική συμπαράσταση που μου προσέφεραν όλα αυτά τα χρόνια.

Αθήνα, Ιούλιος 2022

Ναταλία Ε. Σαλπέα

Περιεχόμενα

Κατάλογος Σχημάτων

Κατάλογος Εικόνων

Κατάλογος Πινάκων

Πρόλογος

Η παρούσα διπλωματική εργασία εκπονήθηκε στην Αθήνα, το έτος 2022, στο Εργαστήριο Συστημάτων Τεχνητής Νοημοσύνης και Μάθησης που ανήκει στον τομέα Τομέας Τεχνολογίας Πληροφορικής και Υπολογιστών της Σχολής Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Εθνικού Μετσόβιου Πολυτεχνείου, με επιβλέποντες τον Καθηγητή κ. Στέφανο Κόλλια και την κ. Παρασκευή Τζούβελη.

Κεφάλαιο 1

Εισαγωγή

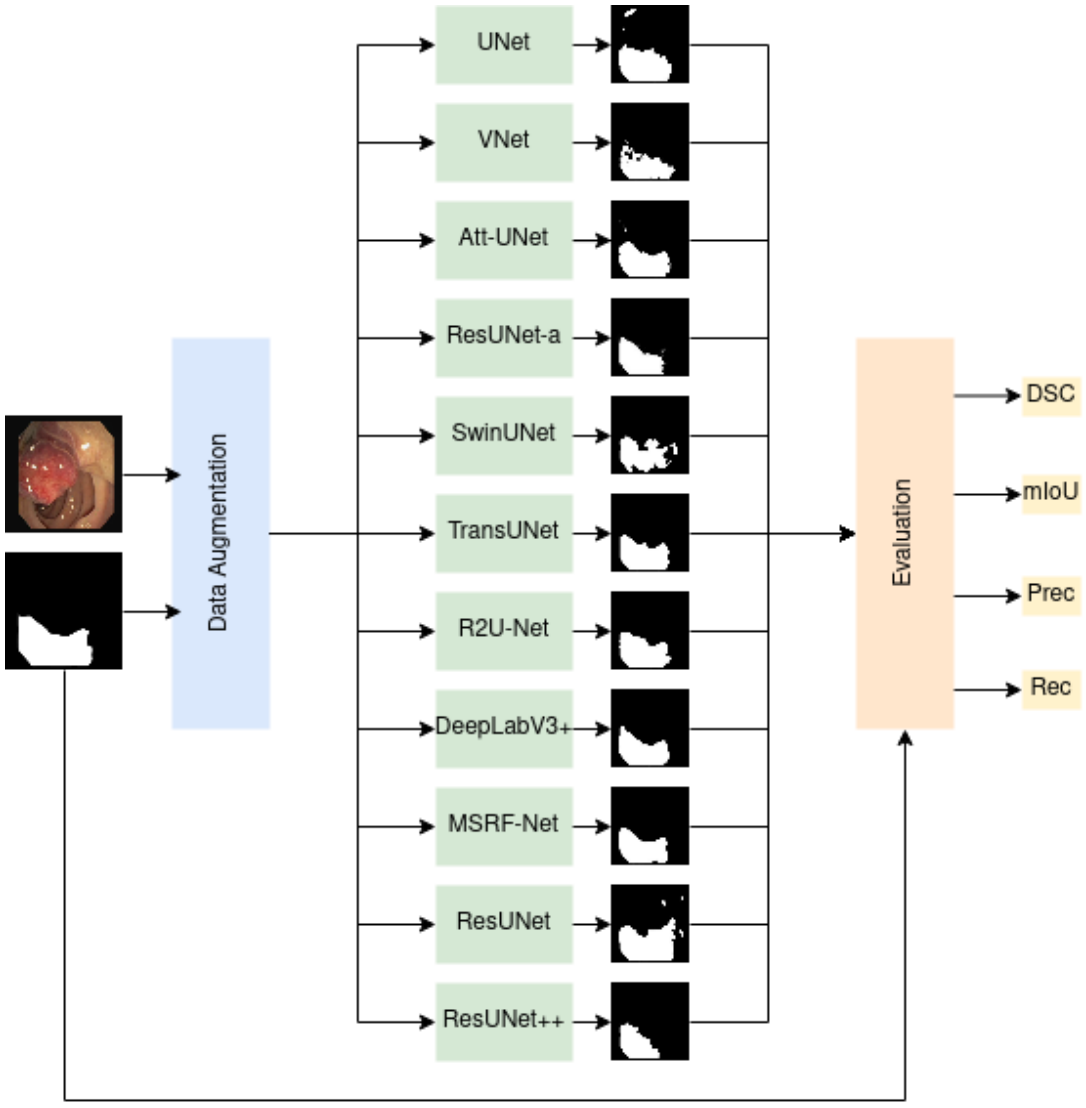
1.1 Αντικείμενο της διπλωματικής

Στην παρούσα εργασία εκτελείται μια ενδελεχής μελέτη των σημαντικότερων αρχιτεκτονικών που έχουν παρουσιαστεί στον τομέα της τμηματοποίησης ιατρικής εικόνας, έλεγχος και αξιολόγηση τους πάνω σε 4 δημόσια διαθέσιμα και ευρέως χρησιμοποιούμενα σύνολα δεδομένων και σύγκριση των αποτελεσμάτων για εξαγωγή συμπερασμάτων για το ποιές αρχιτεκτονικές είναι πιο ευνοϊκές στο συγκεκριμένο έργο και ποιες είναι πιο υποσχόμενες για μελλοντικές εξελίξεις. Η διαδικασία που ακολουθήθηκε απεικονίζεται στο σχήμα 1.1. Συγκεκριμένα, αφότου συγκεντρώθηκαν τα δεδομένα, χωρίστηκαν σε σύνολα εκπαίδευσης, ελέγχου και αξιολόγησης και μεταβλήθηκαν κατάλληλα ώστε να βρεθούν στη μορφή που απαιτούν τα μοντέλα, υποβλήθηκαν σε διαδικασία επαύξησης δεδομένων προκειμένου να αυξηθεί το πλήθος τους και να εισαχθεί ένας παράγοντας επιπλέον ποικιλίας. Στη συνέχεια, τα 11 μοντέλα, που αναλύονται σε παρακάτω κεφάλαιο, εκπαιδεύονται στα επαυξημένα δεδομένα, στα μη επαυξημένα δεδομένα και σε υπερσύνολα που προκύπτουν από τον συνδυασμό συνόλων, με κατάλληλα επιλεγμένες υπερπαραμέτρους, μετρικές σφάλματος και αξιολόγησης, και αποθηκεύονται τα στιγμιότυπα τους που πετυχαίνουν τα μέγιστα score στις μετρικές αξιολόγησης. Έπειτα τα μοντέλα ελέγχονται πάνω στα υποσύνολα ελέγχου, παράγονται οι μάσκες τμηματοποίησης οι οποίες συγκρίνονται ποιοτικά με τις πραγματικές μάσκες και υπολογίζονται οι επιλεγμένες μετρικές αξιολόγησης. Εφόσον συγκεντρώνονται όλα αυτά τα δεδομένα, ακολουθεί ποιοτική και ποσοτική σύγκριση των αρχιτεκτονικών και σχολιασμός των συνόλων δεδομένων. Επιπλέον εκτελούνται πειράματα και συγκρίσεις για τον έλεγχο της επίδρασης της επαύξησης των δεδομένων στην απόδοση των μοντέλων, καθώς και έλεγχος της ικανότητας γενίκευσης των μοντέλων μέσω της εκπαίδευσης και αξιολόγησης σε διαφορετικά σύνολα. Τέλος, παρατίθεται κάποια συμπεράσματα και κάποιες ιδέες για μελλοντικές επεκτάσεις της παρούσας εργασίας, που θα μπορούσαν να οδηγήσουν σε βελτίωση των State-of-the-art scores.

1.2 Οργάνωση του τόμου

Η εργασία αυτή είναι οργανωμένη σε 7 κεφάλαια:

1. Στο Κεφάλαιο 1 γίνεται η εισαγωγή στην παρούσα διπλωματική.



Σχήμα 1.1: *Ολοκληρωμένο pipeline της εργασίας*

2. Στο Κεφάλαιο 2 παρουσιάζεται αναλυτικά η θεματική της εργασίας καθώς και μια ιστορική αναδρομή στις προηγούμενες σχετικές εργασίες στον τομέα.

3. Στο Κεφάλαιο 3 δίνεται το θεωρητικό υπόβαθρο των βασικών αρχιτεκτονικών και των βασικών μηχανισμών που έχουν χρησιμοποιηθεί στην εργασία.

4. Στο Κεφάλαιο 4 γίνεται παρουσίαση των συνόλων δεδομένων καθώς και οι διαδικασίες προεπεξεργασίας και επαύξησης που υπέστησαν.

5. Στο Κεφάλαιο 5 παρατίθενται λεπτομέρειες για τα μοντέλα και τις αρχιτεκτονικές που χρησιμοποιήθηκαν στα πειράματα της εργασίας, για της μετρικές αξιολόγησης καθώς και για τις παραμέτρους και το υπολογιστικό σύστημα.

6. Στο Κεφάλαιο 6 περιγράφονται τα πειράματα που εκτελέστηκαν και παρουσιάζονται τα αποτελέσματα τους, αλλά γίνεται και μελέτη της ικανότητας γενίκευσης των μοντέλων και της επίδρασης της επαύξησης δεδομένων στην απόδοση τους.

7. Στο Κεφάλαιο 7 βρίσκεται ο επίλογος της εργασίας, όπου εξάγονται συμπεράσματα για τις αρχιτεκτονικές που παρουσιάστηκαν σε παραπάνω κεφάλαια και συζητούνται μελλοντικές πιθανές επεκτάσεις.

8. Στο Παράρτημα Α αναλύονται κάποιες ειδικότερες έννοιες, για την πληρότητα του θεωρητικού υποβάθρου.

Μέρος Ι

Θεωρητικό Μέρος

Κεφάλαιο 2

Περιγραφή θέματος

Σε αυτό το κεφάλαιο περιγράφεται το θέμα της παρούσας εργασίας που αφορά την τμηματοποίηση ιατρικής εικόνας και γίνεται μια ιστορική αναδρομή στις αρχιτεκτονικές και τους αλγορίθμους που έχουν χρησιμοποιηθεί στον τομέα αυτό.

2.1 Τμηματοποίηση Ιατρικής Εικόνας

Η τμηματοποίηση ιατρικής εικόνας είναι ένα απαραίτητο έργο για τη γρηγορότερη και καλύτερη διάγνωση ασθενειών και για την εξασφάλιση γρηγορότερης και καλύτερης θεραπείας μεγάλου αριθμού ασθενών ταυτόχρονα, μειώνοντας την ανάγκη για ανθρώπινη παρέμβαση καθώς και τα ανθρώπινα λάθη, τη χρονική διάρκεια και το κόστος. Υπάγεται στο χώρο της σημασιολογικής τμηματοποίησης (semantic segmentation) και περιλαμβάνει τον εντοπισμό περιοχών ενδιαφέροντος σε ιατρικά δεδομένα όπως δισδιάστατες εικόνες, τρισδιάστατες μαγνητικές τομογραφίες εγκεφάλου (MRI [35, 36, 37]), αξονικές τομογραφίες [38, 39], υπέρηχους και ακτινογραφίες [40], μεταξύ άλλων. Τα αποτελέσματα της τμηματοποίησης μπορούν να βοηθήσουν στον εντοπισμό περιοχών που σχετίζονται με ασθένειες, όπως πολύποδες στο έντερο, προκειμένου να αναλυθούν και να αποφανθεί αν είναι καρκινικοί και χρειάζεται να αφαιρεθούν. Η διάγνωση αυτή μπορεί να αποβεί κριτικής σημασίας για τη ζωή ασθενών, καθώς μπορεί να συμβάλει στην πρόληψη σοβαρών μορφών καρκίνου. Λόγω της αργής και κουραστικής διαδικασίας της χειροκίνητης τμηματοποίησης [41], εμφανίζεται μεγάλη ανάγκη για αλγορίθμους όρασης υπολογιστών που θα μπορούν να εκτελέσουν τμηματοποίηση γρήγορα και με ακρίβεια χωρίς την απαίτηση για ανθρώπινη εμπλοκή. Άλλες προκλήσεις στον τομέα της ιατρικής εικόνας είναι το μικρό πλήθος διαθέσιμων δεδομένων, και η ανάγκη για επαγγελματίες εξειδικευμένους στον συγκεκριμένο τομέα κατά την παραγωγή των δεδομένων. Κατά το annotation πολυπόδων σε εικόνες για παράδειγμα, τα πρωτόκολλα και οι κατευθυντήριες γραμμές ορίζονται από τον ειδικό που εκτελεί το annotation εκείνη την στιγμή. Όμως, υπάρχουν αποκλίσεις μεταξύ ειδικών, όπως για παράδειγμα στην απόφαση αν μια περιοχή είναι καρκινική ή όχι. Επιπλέον, λόγω της αρκετά σημαντικής απόκλισης μεταξύ των δεδομένων λόγω της διαφορετικής ποιότητας απεικόνισης, του λογισμικού με το οποίο γίνεται annotation, του ειδικού που κάνει το annotation, της φωτεινότητας των εικόνων και των γενικότερων βιολογικών ποικιλομορφιών οι αλγόριθμοι τμηματοποίησης πρέπει να είναι εύρωστοι και να έχουν υψηλή ικανότητα γενίκευσης.

2.2 Σχετικές εργασίες

Τα τελευταία χρόνια τα Συνελικτικά Νευρωνικά Δίκτυα (CNN) έχουν επικρατήσει στον τομέα της τμηματοποίησης σε διάφορους τύπους ιατρικών δεδομένων, αλλά και γενικότερα [42, 43, 44]. Οι πιο μοντέρνες αρχιτεκτονικές για την semantic και την instance τμηματοποίηση είναι συνήθως δίκτυα κωδικοποιητή-αποκωδικοποιητή, η επιτυχία των οποίων οφείλεται στα skip-connections, που επιτρέπουν την προώθηση σημασιολογικά σημαντικών και πυκνών χαρτών χαρακτηριστικών από τον κωδικοποιητή στον αποκωδικοποιητή και μειώνουν το σημασιολογικό κενό μεταξύ των 2 υποδικτύων. Οι παράμετροι της υλοποίησης αυτού του τύπου μοντέλων διαφέρουν ανάλογα τη βιοιατρική εφαρμογή. Επιπλέον ο μικρός αριθμός δεδομένων δυσχεραίνει το έργο των μοντέλων και απαιτεί εξειδικευμένο σχεδιασμό προκειμένου να αντιμετωπιστεί ως πρόβλημα. Μια από τις πρώτες και πιο δημοφιλείς προσεγγίσεις για το έργο της σημασιολογικής τμηματοποίησης ιατρικής εικόνας που αντιμετωπίζει το παραπάνω πρόβλημα, είναι το UNet [8], το οποίο μετέτρεψε το Fully Convolutional Network (FCN) του [3] που έχει μόνο συνελικτικά στρώματα, σε μια αρχιτεκτονική κωδικοποιητή-αποκωδικοποιητή U-τύπου στην οποία τα χαρακτηριστικά χαμηλού και υψηλού επιπέδου συνδυάζονται μέσω skip-connections. Ωστόσο, μπορεί να προκύψει σημασιολογικό κενό μεταξύ των εν λόγω χαρακτηριστικών. Προκειμένου να γεφυρωθεί το κενό προτείνεται στο [45] η προσθήκη συνελικτικών μονάδων στα skip-connections. Στο [11] προτείνεται μια εκδοχή του UNet με μηχανισμό attention ο οποίος συμβάλλει στην υλοποίηση ενός μηχανισμού pruning για την παράλειψη μη απαραίτητων χωρικών χαρακτηριστικών που περνούν από τα skip-connections. Το UNet δεν περιορίστηκε μόνο στον τομέα της ιατρικής εικόνας αλλά εμφανίστηκαν και εκδοχές του για διαφορετικούς τύπους τμηματοποίησης, όπως το [14]. Η τρισδιάστατη εκδοχή του UNet, το VNet [9] εκτελεί τμηματοποίηση σε αραιά annotated ογκομετρικές εικόνες, και αποτελείται από ένα FCN με υπολειπόμμενες συνδέσεις. Οι υπολειπόμμενες συνδέσεις βελτιώνουν την ροή της πληροφορίας και ταυτόχρονα μειώνουν τον χρόνο σύγκλισης, γιάυτό και προτιμήθηκαν σε αρκετές προσεγγίσεις, όπως στο [12]. Στο [13], στο μοντέλο ResUNet προστίθενται μπλοκ Squeeze and Excitation, attention μηχανισμοί καθώς και Atrous Spatial Pyramid Pooling (ASPP). Στο [10] οι υπολειμματικές συνδέσεις συνδυάζονται σε ένα επαναληπτικό δίκτυο το οποίο οδηγεί σε καλύτερες αναπαραστάσεις χαρακτηριστικών. Ο μηχανισμός ASPP εμφανίστηκε στο [46] για να συσσωρευθούν global χαρακτηριστικά, και εξελίχθηκε στο [17] σε μία αρχιτεκτονική που χρησιμοποιεί skip-connections μεταξύ του κωδικοποιητή και του αποκωδικοποιητή. Τα μπλοκ Squeeze and Excitation προτάθηκαν στο [21] και μοντελοποιούν της αλληλοεξαρτήσεις μεταξύ καναλιών και εξάγουν global χάρτες που δίνουν έμφαση στα σημαντικά χαρακτηριστικά και αποσβαίνουν τα λιγότερο σχετικά χαρακτηριστικά. Η ιδέα του gated shape stream εμφανίστηκε στο [47] και εξυπηρετεί την παραγωγή πιο λεπτομερών χαρτών τμηματοποίησης λαμβάνοντας υπόψιν το σχήμα της περιοχής ενδιαφέροντος. Για την διατήρηση της υψηλής ανάλυσης των αναπαραστάσεων, προτάθηκε η ανταλλαγή χαρακτηριστικών χαμηλού και υψηλού επιπέδου, μεταξύ διαφορετικών κλιμάκων ανάλυσης, η οποία αποδεικνύεται στο [48] ότι οδηγεί σε πιο χωρικά ακριβείς τελικούς χάρτες τμηματοποίησης. Τέλος, στο [18] συνδυάζονται πολλές από τις παραπάνω τεχνικές όπως Squeeze and Excitation μπλοκ, gated shape stream, attention μηχανισμός, με έμφαση στην ανταλλαγή χαρακτηριστικών μεταξύ κλιμάκων ανάλυσης

(multi-scale fusion), σε ένα ισχυρό μοντέλο που αποδίδει σταθερά στους τύπους ιατρικών δεδομένων που περιέχονται στην παρούσα εργασία.

Κεφάλαιο 3

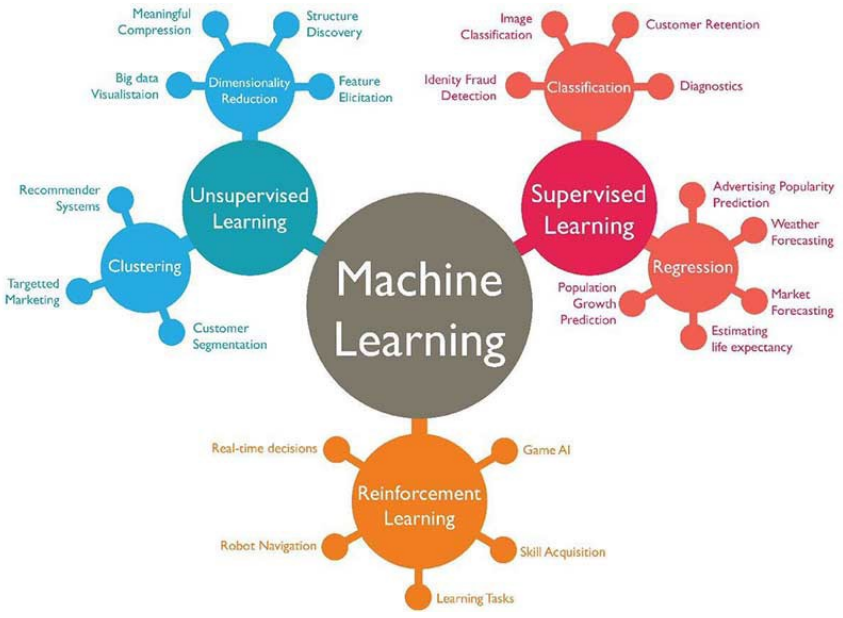
Θεωρητικό υπόβαθρο

Σε αυτό το κεφάλαιο δίνεται το θεωρητικό υπόβαθρο για τις βασικές αρχιτεκτονικές και τους μηχανισμούς που αξιοποιούνται στην παρούσα εργασία.

3.1 Μηχανική Μάθηση και Νευρωνικά Δίκτυα

3.1.1 Μηχανική Μάθηση

Η Μηχανική Μάθηση υπάγεται στο πεδίο της Τεχνητής Νοημοσύνης και σύμφωνα με τον Arthur Samuel, επιτρέπει στους υπολογιστές να μαθαίνουν από δεδομένα, να κάνουν ακριβείς προβλέψεις, ακόμα και να τις βελτιώνουν αυτόματα, χωρίς να είναι ρητά προγραμματισμένοι να το κάνουν. Οι αλγόριθμοι μηχανικής μάθησης τροφοδοτούνται με δεδομένα εισόδου, εκτελούν στατιστική ανάλυση για να προβλέψουν μια έξοδο, και ενημερώνονται κατάλληλα όσο εμφανίζονται νέα δεδομένα. Ενώ η μηχανική μάθηση ανήκει στον τομέα της επιστήμης υπολογιστών, διαφέρει αρκετά από τις παραδοσιακές υπολογιστικές προσεγγίσεις, στις οποίες οι αλγόριθμοι είναι ρητά προγραμματισμένοι με συγκεκριμένες οδηγίες να λύνουν ένα συγκεκριμένο πρόβλημα.



Εικόνα 3.1: *Τομείς της μηχανικής μάθησης [25]*

Η σύγχρονη τεχνολογία έχει ωφεληθεί αρκετά από τη μηχανική μάθηση. Για παράδειγμα, τεχνολογίες αναγνώρισης προσώπου, αναγνώρισης ψηφίων και γραμμάτων, εντοπισμού αντικειμένων, συστήματα προτάσεων, ακόμα και αυτο-οδηγούμενα οχήματα και διάγνωση ασθενειών, είναι μόνο μερικούς από τους τομείς που έχει προσφέρει σημαντικά εφόδια η μηχανική μάθηση, και είναι ένα συνεχώς εξελισσόμενο πεδίο που έχει πολλά να προσφέρει ακόμα στο μέλλον.

Οι αλγόριθμοι μηχανικής μάθησης χωρίζονται σε 3 κατηγορίες:

- Επιβλεπόμενης μάθησης (Supervised Learning)
- Μη επιβλεπόμενης μάθησης (Unsupervised Learning)
- Ενισχυτικής μάθησης (Reinforcement Learning)

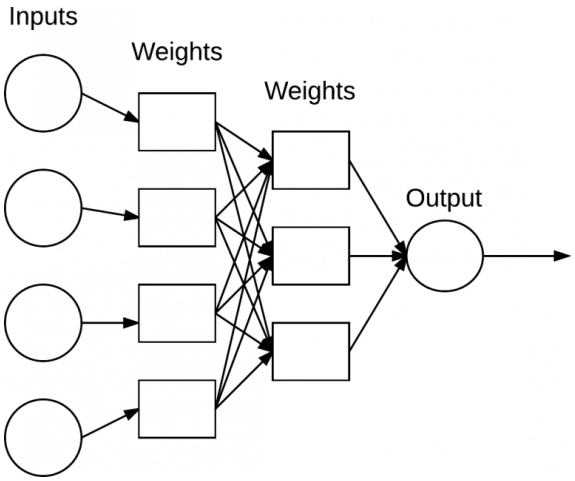
Στην επιβλεπόμενη μάθηση, τα συστήματα τροφοδοτούνται με δεδομένα το καθένα από τα οποία αντιστοιχίζεται με μια ετικέτα [49]. Ο σκοπός είναι να προσεγγιστεί η συνάρτηση αντιστοίχισης με τόση ακρίβεια ώστε όταν παρουσιαστεί μια καινούργια είσοδος στο μοντέλο, να μπορεί να γίνει σωστή πρόβλεψη της εξόδου. Παραδείγματα δοκιμασιών που υπάγονται στην επιβλεπόμενη μάθηση είναι η κατηγοριοποίηση (classification) στην οποία αποδίδονται κατηγορίες στα δεδομένα εισόδου και η παλινδρόμηση (regression), στην οποία προσεγγίζεται μια συνεχής τιμή, για παράδειγμα η τιμή μιας μετοχής σε κάποια χρονική στιγμή στο μέλλον.

Στη μη επιβλεπόμενη μάθηση, το μοντέλο τροφοδοτείται με δεδομένα χωρίς ετικέτα ή κατηγορία και δρά σε αυτά χωρίς προηγούμενη εκπαίδευση. Παραδείγματα δοκιμασιών σε αυτόν τον τομέα αποτελεί η ομαδοποίηση (clustering), στην οποία εντοπίζονται εσωτερικές ομοιότητες μεταξύ των δεδομένων με βάση των οποίων δημιουργούνται ομάδες, αλλά και η συσχέτιση (association), στην οποία αναζητούνται κανόνες που υπαγορεύουν συγκεκριμένη συσχέτιση μεταξύ των δεδομένων, για παράδειγμα αν τα άτομα που αγοράζουν το προιόν X θα αγοράσουν και το προιόν Y.

Στην ενισχυτική μάθηση, αλγόριθμος ή αλλιώς πράκτορας (agent), μαθαίνει μέσω της αλληλεπίδρασης με το περιβάλλον του. Λαμβάνει επιβράβευση όταν εκτελεί ένα σωστό βήμα και τιμωρία όταν εκτελεί ένα λανθασμένο βήμα. Επιχειρέι να μεγιστοποιήσει την επιβράβευση του και να ελαχιστοποιήσει την τιμωρία, χωρίς ανθρώπινη παρέμβαση.

3.1.2 Νευρωνικά Δίκτυα

Τα νευρωνικά δίκτυα είναι οι δομικοί λίθοι των μοντέλων μηχανικής μάθησης. Πολλές δοκιμασίες που αφορούν νοημοσύνη, αναγνώριση προτύπων και εντοπισμό αντικειμένων είναι δύσκολο να αυτοματοποιηθούν όμως εκτελούνται εύκολα από ανθρώπους, ακόμα και από μικρά παιδιά και ζώα. Τα βιολογικά νευρωνικά δίκτυα που υπάρχουν στους εγκεφάλους των ζωντανών οργανισμών με νοημοσύνη μπορούν και εκτελούν αυτά τα περίπλοκα έργα, και τα τεχνητά νευρωνικά δίκτυα της μηχανικής μάθησης επιχειρούν να τα μοντελοποιήσουν ώστε να τα μιμηθούν, από όπου προκύπτει και το όνομα τους. Τα τεχνητά νευρωνικά δίκτυα περιλαμβάνουν δομές κατευθυνόμενου γράφου όπου κάθε κόμβος (νευρώνας) εκτελεί έναν υπολογισμό. Κάθε σύνδεση μεταφέρει ένα σήμα, μαζί με ένα βάρος που υποδεικνύει αν το σήμα ενισχύεται ή αποσβαίνεται και καθορίζει τη σημασία του για την έξοδο.



Σχήμα 3.1: *Παράδειγμα Τεχνητού Νευρωνικού Δικτύου [1]*

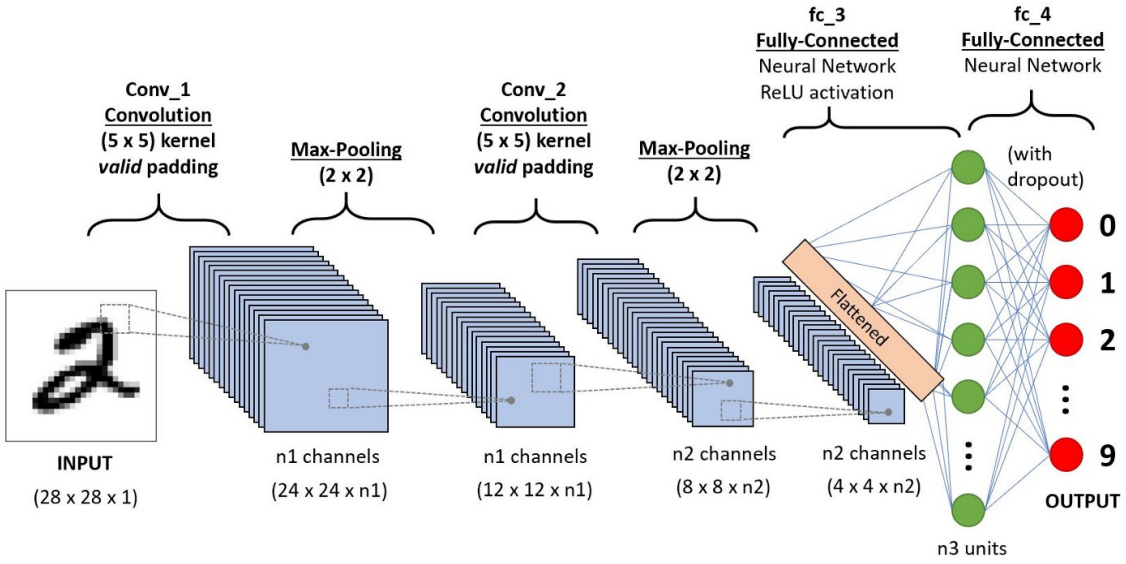
Οι δομικοί λίθοι των τεχνητών νευρωνικών δικτύων είναι οι νευρώνες, όπως και στον εγκέφαλο. Στον εγκέφαλο, κάθε νευρώνας είναι συνδεδεμένος με περίπου 10.000 άλλους νευρώνες, από τους οποίους λαμβάνουν και στους οποίους στέλνουν ηλεκτροχημικά σήματα τα οποία οδηγούν σε ενεργοποίηση του νευρώνα αν είναι αρκετά ισχυρά. Η ενεργοποίηση αυτή είναι μια δυαδική κατάσταση που μπορεί να μοντελοποιηθεί από έναν υπολογιστή. Έτσι γεννήθηκε η ιδέα των τεχνητών νευρωνικών δικτύων [1].

3.2 Συνελικτικά Νευρωνικά Δίκτυα (CNN)

Τα Συνελικτικά Νευρωνικά Δίκτυα αποτελούν έναν αλγόριθμο βαθιάς μάθησης που λαμβάνει ως είσοδο εικόνες και εντοπίζει σε αυτές αντικείμενα αποδίδοντας τους σημασία μέσω βαρών [2]. Προορίζονται κατά βάση για το έργο της κατηγοριοποίησης (classification) αλλά είναι πολύ ισχυρά στην εξαγωγή χαρακτηριστικών. Έχουν την δυνατότητα να εντοπίζουν χωρικές και χρονικές εξαρτήσεις σε εικόνες μέσω της εφαρμογής κατάλληλων φίλτρων, μειώνοντας τον αριθμό των παραμέτρων σε σύγκριση με προηγούμενες προσεγγίσεις. Ο ρόλος τους είναι να φέρνουν τις εικόνες σε μια απλούστερη μορφή που είναι πιο εύκολη στην επεξεργασία, χωρίς να χάνονται σημαντικές πληροφορίες. Αποτελούνται από 3 κύρια νευρωνικά στρώματα, το συνελικτικό στρώμα (convolutional layer), το στρώμα ομαδοποίησης (pooling layer) και το πυκνό πλήρως συνδεδεμένο στρώμα (fully connected layer), όπως φαίνεται και στο σχήμα 3.2.

Σκοπός του συνελικτικού στρώματος είναι να εξάγει χαρακτηριστικά υψηλού επιπέδου [50] όπως ακμές ή άλλα οπτικά στοιχεία, εκτελώντας την πράξη της συνέλιξης με διάφορους πυρήνες στην εικόνα-εισόδο. Δεν υπάρχει μόνο ένα τέτοιο στρώμα στα συνελικτικά δίκτυα, και ο συνδυασμός πολλών τέτοιων στρωμάτων οδηγεί σε όλική κατανόηση της εικόνας και όλων των χαρακτηριστικών της, υψηλού και χαμηλού επιπέδου, όπως για παράδειγμα οι ακμές, οι γωνίες και το χρώμα.

Το στρώμα ομαδοποίησης μειώνει το χωρικό μέγεθος των χαρακτηριστικών που προκύπτουν από τα συνελικτικά στρώματα, εκτελώντας υποδειγματοληψία, μειώνοντας έτσι και τις υπολογιστικές απαιτήσεις του δικτύου. Επιπλέον είναι χρήσιμο στην εξαγωγή κυρίαρχων



Σχήμα 3.2: *Παράδειγμα Συνελικτικού Νευρωνικού Δικτύου (CNN) για την κατηγοριοποίηση εικόνων ψηφίων [2]*

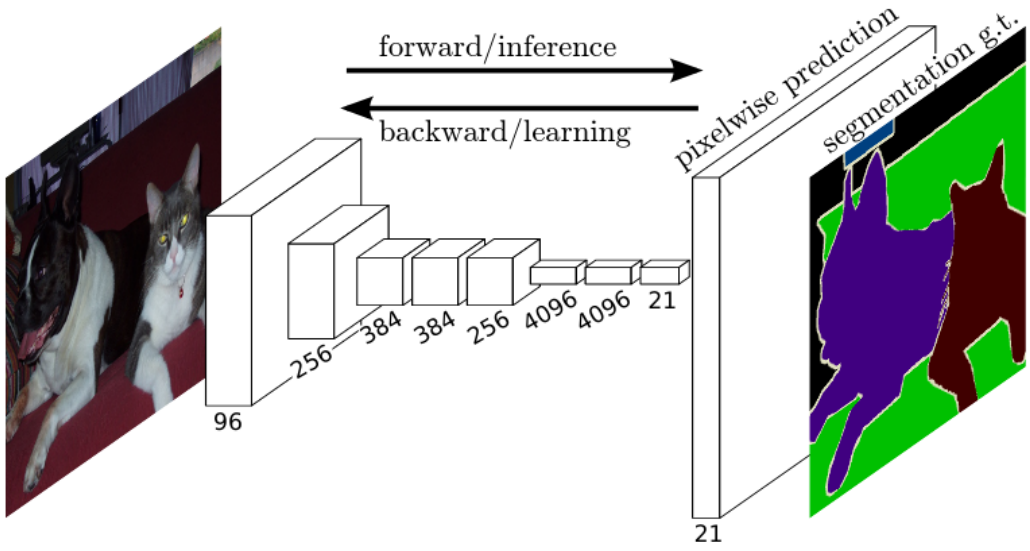
χαρακτηριστικών που είναι αναλλοίωτα περιστροφής και θέσης, διατηρώντας την ποιότητα της εκπαίδευσης του μοντέλου. Υπάρχουν δύο τύποι ομαδοποίησης, το Max Pooling που επιστρέφει τη μέγιστη τιμή του κομματιού της εικόνας που καλύπτει ο πυρήνας πετυχαίνοντας μείωση των διαστάσεων καθώς και αποθορυβοποίηση, και το Average Pooling που επιστρέφει τη μέση τιμή του κομματιού της εικόνας που καλύπτει ο πυρήνας, πετυχαίνοντας μείωση των διαστάσεων.

Το πλήρως συνδεδεμένο στρώμα μαθαίνει μη γραμμικούς συνδυασμούς των χαρακτηριστικών υψηλού επιπέδου που προκύπτουν από το τελικό συνελικτικό στρώμα. Τα μειονεκτήματα των συνελικτικών δικτύων είναι ότι δεν μπορούν να λειτουργήσουν με δεδομένα διαφορετικών μεγεθών, και επίσης δεν μπορούν να χρησιμοποιηθούν στο έργο της τμηματοποίησης αφού ο αριθμός των αντικειμένων/περιοχών ενδιαφέροντος δεν είναι συγκεκριμένος και επομένως το μέγεθος του στρώματος εξόδου δεν μπορεί να είναι σταθερό.

3.3 Εξ Ολοκλήρου Συνελικτικά Δίκτυα (FCN)

Στα εξόλοκλήρου συνελικτικά δίκτυα υπάρχουν μόνο συνελικτικά στρώματα. Η διαφορά τους από τα απλά συνελικτικά δίκτυα είναι ότι το τελευταίο πυκνό πλήρως συνδεδεμένο στρώμα αντικαθίσταται από ένα πλήρως συνελικτικό στρώμα, όπως φαίνεται στο σχήμα 3.3. Προτάθηκαν στο [3] και μπορούν να παράγουν στην έξοδο τους χωρικούς χάρτες τμηματοποίησης και πυκνές προβλέψεις σε επίπεδο εικονοστοιχείων για ολόκληρη την εικόνα και όχι για κομμάτια της (patches).

Χρησιμοποιούν skip-connections που υπερδειγματοληπτούν τους πίνακες χαρακτηριστικών από το τελευταίο στρώμα και τους συνδυάζουν με τους πίνακες χαρακτηριστικών προηγούμενων στρωμάτων. Έτσι παράγουν λεπτομερείς χάρτες τμηματοποίησης. Ωστόσο εμφανίζουν κάποιους περιορισμούς όπως ότι είναι αρκετά αργά μοντέλα για real-time infer-



Σχήμα 3.3: *Παράδειγμα Πλήρως Συνελικτικού Δικτύου (FCN) για σημασιολογική τμηματοποίηση [3]*

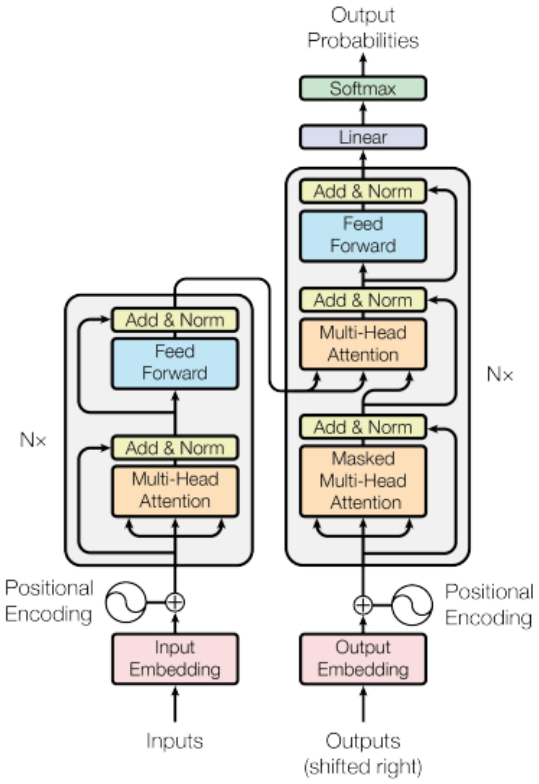
ence και ότι μειώνουν πολύ την ανάλυση των χαρακτηριστικών που περνούν από πολλαπλά επίπεδα συνέλιξης και ομαδοποίησης, με αποτέλεσμα να παράγουν χαμηλής ανάλυσης προβλέψεις με μη ακριβή περιγράμματα αντικειμένων.

3.4 Μετασχηματιστές (Transformers)

Οι μετασχηματιστές σχεδιάστηκαν αρχικά για το έργο της μετάφρασης γλωσσών. Επιτρέπουν τη μοντελοποίηση μακριών εξαρτήσεων μεταξύ ακολουθιών εισόδου και υποστηρίζουν παράλληλη επεξεργασία ακολουθιών σε αντίθεση με επαναληπτικά δίκτυα. Επιπλέον ο σχεδιασμός τους επιτρέπει την επεξεργασία διαφορετικών τύπων δεδομένων, όπως εικόνες, βίντεο, κείμενο και φωνή. Γιαυτό το λόγο έχουν επικρατήσει τελευταία ως αρχιτεκτονικές και εξαπλώνονται σε όλο και περισσότερους τομείς πέρα από την Επεξεργασία Φυσικής Γλώσσας, με προοπτική να αντικαταστήσουν ολοκληρωτικά τα Επαναληπτικά Νευρωνικά Δίκτυα (RNN) και τα Συνελικτικά Νευρωνικά Δίκτυα (CNN) [51].

Στον πυρήνα τους αποτελούνται από μια στίβα στρωμάτων κωδικοποιητή και αποκωδικοποιητή. Ο κωδικοποιητής περιέχει το πολύ σημαντικό στρώμα Self-Attention που υπολογίζει τη σχέση μεταξύ διαφορετικών μερών της ακολουθίας εισόδου καθώς και ένα στρώμα Feed-forward. Περιλαμβάνει επίσης υπολειμματικές συνδέσεις καθώς και στρώματα κανονικοποίησης. Ο μηχανισμός Self-Attention του κωδικοποιητή δημιουργεί μια αναπαράσταση για κάθε token της ακολουθίας εισόδου που περιγράφει την σημασιολογική ομοιότητα του με τα υπόλοιπα tokens [52]. Η αρχιτεκτονική του παρουσιάζεται στο σχήμα 3.4 .

Ο αποκωδικοποιητής παράγει ένα token τη φορά, λαμβάνοντας υπόψιν τόσο τις διανυσματικές παραστάσεις εισόδου (embeddings) όσο και αυτές τις εξόδου μέχρι στιγμής. Ο



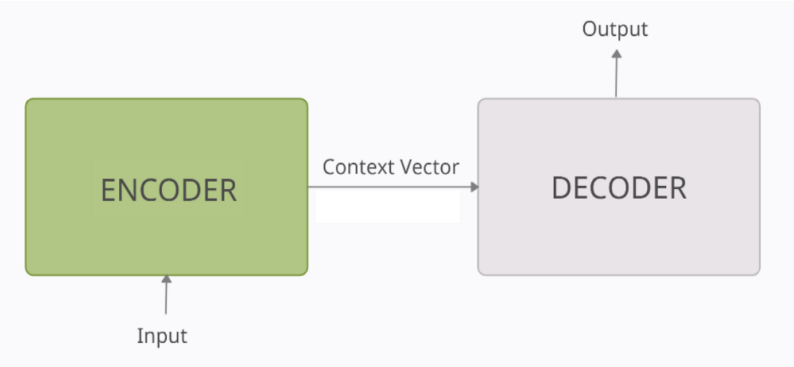
Σχήμα 3.4: *Παράδειγμα αρχιτεκτονικής μετασχηματιστή [4]*

μηχανισμός Self-Attention δεν ενδιαφέρεται για τη θέση των tokens και τα αποτελέσματα του είναι αναλλοίωτα σε διαδικασίες ανακατέματος. Ωστόσο σε διαδικασίες όπως η τμηματοποίηση εικόνας, είναι σημαντικές οι χωρικές πληροφορίες και η σχετική θέση των αντικειμένων/περιοχών ενδιαφέροντος. Ο τρόπος να εισαχθούν αυτές οι πληροφορίες στη διαδικασία εκπαίδευσης είναι να προσαρτηθούν σε κάθε token πριν αυτό περάσει στον μηχανισμό προσοχής, έτσι ώστε να συμμετέχουν στη διαδικασία μάθησης [53]. Το γεγονός αυτό καθιστά τους μετασχηματιστές πιο απαιτητικούς ως προς τα δεδομένα που χρειάζονται για την εκπαίδευση τους σε σχέση με τα CNN.

3.5 Δίκτυα Κωδικοποιητή-Αποκωδικοποιητή (Encoder-Decoder)

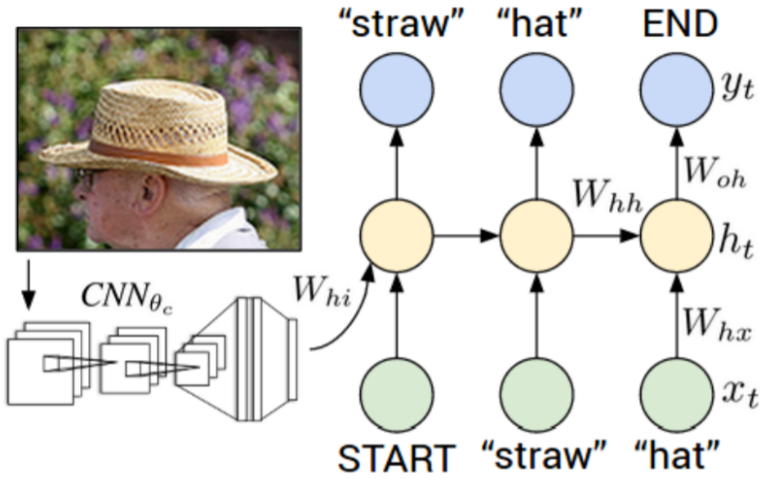
Τα δίκτυα κωδικοποιητή-αποκωδικοποιητή είναι στενά συνδεδεμένα με τους Μετασχηματιστές. Χρησιμοποιούνται σε δοκιμασίες όπως η παραγωγή λεζάντας σε εικόνες, η ανάλυση συναισθήματος και η μετάφραση. Σχεδιάστηκαν για την επίλυση προβλημάτων ακολουθιών, δηλαδή προβλήματα στα οποία η είσοδος και η έξοδος είναι ακολουθίες. Σε υψηλό επίπεδο ο κωδικοποιητής και ο αποκωδικοποιητής αποτελούν δύο μπλοκ που συνδέονται μέσω ενός διανύσματος συμφραζόμενων (context vector) [5], όπως φαίνεται στο σχήμα 3.5 .

Ο κωδικοποιητής επεξεργάζεται κάθε token της ακολουθίας εισόδου, συσσωρεύοντας όλη την πληροφορία για την είσοδο σε ένα διάνυσμα σταθερού μήκους. Το context vector περιέχει όλη τη σημασιολογική πληροφορία για την ακολουθία εισόδου, η οποία είναι απαραίτητη για να παράγει ο αποκωδικοποιητής με τη σειρά του τις προβλέψεις του, δηλαδή



Σχήμα 3.5: *Παράδειγμα αρχιτεκτονικής encoder-decoder [5]*

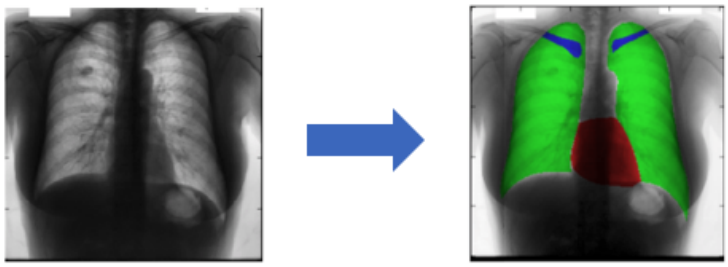
την ακολουθία εξόδου, token by token. Η εσωτερική αρχιτεκτονική των δύο μπλοκ μπορεί να διαφέρει ανάλογα με το έργο για το οποίο προορίζονται, για παράδειγμα για το έργο της μετάφρασης, τα δύο μπλοκ αποτελούνται από μονάδες LSTM. Ένα άλλο παράδειγμα διαφορετικής αρχιτεκτονικής, για το έργο της παραγωγής λεζάντας για εικόνα φαίνεται στο σχήμα 3.6 , στο οποίο η εικόνα αρχικά περνάει από ένα Συνελικτικό Δίκτυο (CNN) και τα χαρακτηριστικά από το τελευταίο πυκνό δίκτυο τροφοδοτούνται στο LSTM δίκτυο. Αυτό το στρώμα δρά σαν το διάνυσμα συμφραζόμενων αφού περιλαμβάνει όλη την ουσία της εικόνας [6].



Σχήμα 3.6: *Παράδειγμα αρχιτεκτονικής encoder-decoder για το έργο παραγωγής λεζάντας για εικόνα [6]*

3.6 Σημασιολογική Τμηματοποίηση (Semantic Segmentation)

Η σημασιολογική τμηματοποίηση εικόνας είναι το έργο της κατηγοριοποίησης κάθε εικονοστοιχείο σε μια από τις κλάσεις ενός προκαθορισμένου συνόλου [54]. Για παράδειγμα στην εικόνα 3.2 τα εικονοστοιχεία που ανήκουν στην καρδία έχουν κατηγοριοποίηθει στην κατηγορία ¨καρδιά¨ και ομοίως για τους πνεύμονες και τις κλείδες αλλά και για το παρασκήνιο (background). Η σημασιολογική τμηματοποίηση είναι διαφορετική από τον εντοπισμό αντικειμένων διότι δεν παράγει κουτιά οριοθέτησης για τα εντοπισμένα αντικείμενα, επίσης είναι



Εικόνα 3.2: *Παράδειγμα σημασιολογικής τμηματοποίησης σε εικόνα ακτινογραφίας στήθους στην οποία έχουν τμηματοποιηθεί η καρδιά (κόκκινο), οι πνεύμονες (πράσινο) και οι κλείδες (μπλέ) [26]*

διαφορετική από την τμηματοποίηση instance η οποία θα έδινε διαφορετική ετικέτα σε κάθε εμφάνιση του ίδιου τύπου αντικειμένου. Η σημασιολογική τμηματοποίηση τοποθετεί στην ίδια κατηγορία όλα τα αντικείμενα ίδιου τύπου όσες φορές και να εμφανίζονται στην εικόνα. Συνεπώς, η συγκεκριμένη τμηματοποίηση παράγει στην έξοδο μια εικόνα ίδιου μεγέθους με την αρχική (ενός καναλιού) που περιέχει ετικέτες για όλα τα εικονοστοιχεία ανάλογα με την κλάση στην οποία υπολογίστηκε πως ανήκουν.

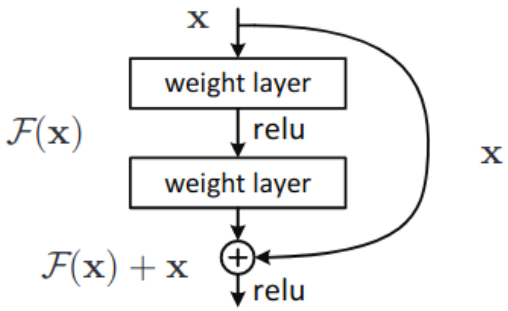
3.7 Μηχανισμός Προσοχής (Attention)

Στην ψυχολογία η προσοχή είναι η γνωστική διαδικασία της επιλεκτικής συγκέντρωσης σε ένα ή περισσότερα πράγματα ενώ τα υπόλοιπα αγνοούνται. Όπως τα νευρωνικά δίκτυα αποτελούν μια προσπάθεια μοντελοποίησης διεργασιών του ανθρώπινου εγκεφάλου έτσι και ο μηχανισμός προσοχής μιμείται την ανθρώπινη προσοχή, με το να συγκεντρώνεται σε μερικά σχετικά πράγματα αγνοώντας τα υπόλοιπα σε βαθιά νευρωνικά δίκτυα [55]. Η βασική ιδέα είναι ότι κάθε φορά που το μοντέλο προβλέπει ένα token εξόδου, αξιοποιεί μόνο τα μέρη της εισόδου που είναι συγκεντρωμένη η πιο σχετική πληροφορία αντί για ολόκληρη την ακολουθία. Ο μηχανισμός προσοχής συνδέει τον κωδικοποιητή με τον αποκωδικοποιητή και προσφέρει στον αποκωδικοποιητή πληροφορίες για τις κρυφές καταστάσεις του κωδικοποιητή με αποτέλεσμα το μοντέλο να μπορεί να δώσει περισσότερη προσοχή στα πιο σημαντικά κομμάτια της ακολουθίας εισόδου, αποδίδοντας τους scores. Ο μηχανισμός προσοχής εξυπηρετεί τη μοντελοποίηση των εξαρτήσεων μεγαλύτερης απόστασης μέσα στο μοντέλο, που αποτελεί σημαντικό πρόβλημα στα επαναληπτικά νευρωνικά δίκτυα. Ενώ εμφανίστηκε αρχικά για το έργο της μετάφρασης γλώσσας έχει εξαπλωθεί και σε τομείς όπως η όραση υπολογιστών [56]. Πριν προταθεί για πρώτη φορά στο [57], η μετάφραση γλώσσας στηριζόταν στη χρήση επαναληπτικών μοντέλων κωδικοποιητή-αποκωδικοποιητή (RNNs/LSTMs), τα οποία εμφανίζουν δυσκολία στην κατανόηση μακρύτερων ακολουθιών εισόδου και επομένως παράγουν μη ικανοποιητικά αποτελέσματα. Ακόμα και τα LSTMs (Long Short Term Memory) που σχεδιάστηκαν για να μοντελοποιούν μακρύτερες εξαρτήσεις τείνουν να ¨ξεχνούν¨ σημα-

ντικές πληροφορίες σε αρκετές περιπτώσεις. Υπάρχουν διαφορετικοί δημοφιλείς μηχανισμοί προσοχής όπως Content-base attention, Additive, Location-Base, General, Dot-Product, Scaled Dot-Product καθώς και κάποιες πιο γενικές κατηγορίες όπως Self-Attention, που αποδίδει βάρη στα tokens μιας μοναδικής ακολουθίας και παράγει την αναπαράσταση της, Multi-Head Self-Attention που συνδυάζει πληροφορίες από διαφορετικούς υποχώρους αναπαραστάσεων, Global/Soft Attention και Soft/Hard Attention [58].

3.8 Υπολειμματικές συνδέσεις (Residual connections)

Οι υπολειμματικές συνδέσεις εμφανίστηκαν προκειμένου να βοηθήσουν στη γρηγορότερη σύγκλιση κατά την εκπαίδευση βαθύτερων νευρωνικών δικτύων. Στα κλασσικά δίκτυα feed-forward οι πληροφορίες προχωρούν στα στρώματα σειριακά, η έξοδος ενός στρώματος είναι η είσοδος του επόμενου στρώματος. Οι υπολειμματικές συνδέσεις προσφέρουν ένα εναλλακτικό μονοπάτι για τα δεδομένα να φτάσουν επόμενα στρώματα του δικτύου. Αν F είναι η συνάρτηση που περιγράφει τα στρώματα του δικτύου, όπως απεικονίζονται στο σχήμα 3.7, τότε σε ένα κλασσικό feed-forward δίκτυο η έξοδος του δικτύου θα ήταν F(x), σε ένα δίκτυο με υπολειμματικές συνδέσεις όμως η έξοδος είναι F(x) + x εφόσον η είσοδος x περνάει κατευθείαν στην έξοδο. Κατά μήκος των υπολειμματικών συνδέσεων μπορούν να υπάρχουν και συναρτήσεις αντί για απλό πέρασμα στην έξοδο [59].



Σχήμα 3.7: *Η αρχιτεκτονική ενός υπολειμματικού μπλοκ (residual block)[7]*

Στα βαθιά νευρωνικά δίκτυα συνήθως εμφανίζονται τα προβλήματα των exploding και vanishing παραγώγων [60]. Οι υπολειμματικές συνδέσεις συμβάλλουν στην καταπολέμηση αυτών των προβλημάτων και προωθούν τη γρηγορότερη σύγκλιση των μοντέλων.

Μέρος II

Πρακτικό Μέρος

Κεφάλαιο 4

Δεδομένα

Στο κεφάλαιο αυτό παρουσιάζονται τα σύνολα δεδομένων που χρησιμοποιήθηκαν στην παρούσα εργασία, για την εκπαίδευση των μοντέλων καθώς και για τη δοκιμή τους. Για δοκιμασία της τμηματοποίησης ιατρικής εικόνας, δυστυχώς δεν υπάρχουν αρκετά δημόσια σύνολα δεδομένων ικανοποιητικού μεγέθους και ποιότητας. Για την αντιμετώπιση του πρώτου προβλήματος, δηλαδή του ανεπαρκή αριθμού δειγμάτων, η επαύξηση δεδομένων λειτουργεί καταλυτικά [61]. Ωστόσο όσον αφορά το δεύτερο πρόβλημα, αυτό της ποιότητας, η αντιμετώπιση του δεν κρίνεται ως εύκολη πρόκληση, διότι η ποιότητα καθορίζεται από τα μηχανήματα που χρησιμοποιεί το κάθε εργαστήριο, επομένως η ποιότητα των εικόνων μπορεί να βελτιωθεί μόνο όταν γίνει κάποια σημαντική εξέλιξη στην τεχνολογία των μηχανημάτων λήψης δειγμάτων.

Τα σύνολα που χρησιμοποιήθηκαν ανήκουν σε δύο μεγάλες κατηγορίες που αφορούν την τμηματοποίηση πολυπόδων και την τμηματοποίηση πυρήνων κυττάρων αντίστοιχα. Επιλέχθηκαν διαφορετικού είδους σύνολα προκειμένου να ελεγχθεί η ευρωστία, δηλαδή η ικανότητα του μοντέλου να αποδίδει ομοίως τόσο σε εύκολες όσο και σε δύσκολες εικόνες, αλλά και η γενίκευση των μοντέλων, δηλαδή η ικανότητα τους να αποδίδουν σωστά σε διαφορετικά σύνολα δεδομένων, όπως για παράδειγμα να έχουν εκπαιδευτεί σε δεδομένα ενός εργαστηρίου και να δοκιμάζονται σε δεδομένα άλλου εργαστηρίου που λήφθηκαν με διαφορετική τεχνολογία [62].

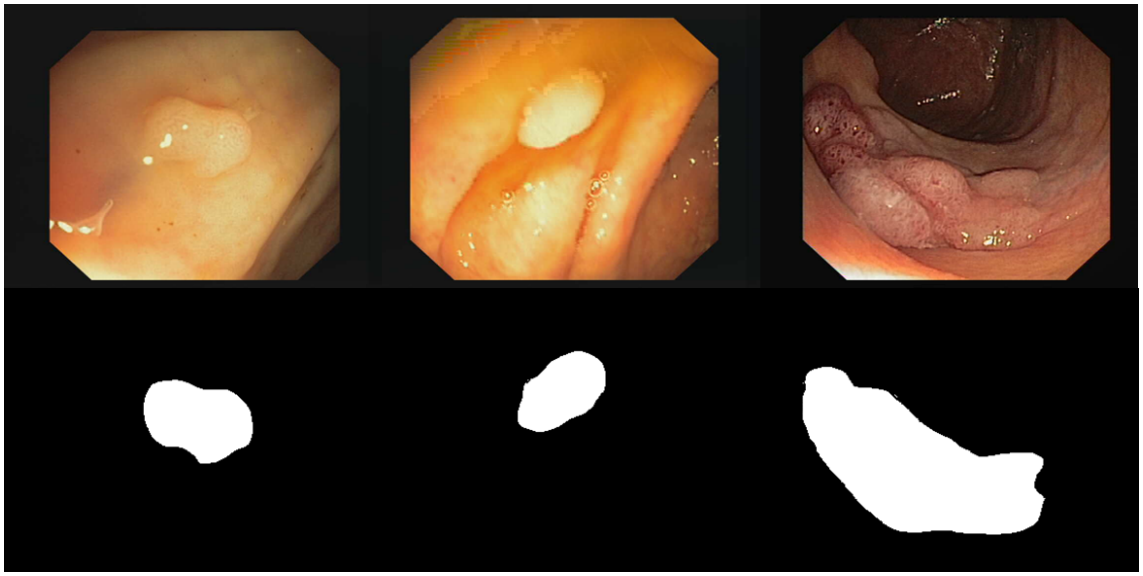
4.1 Σύνολα Δεδομένων

4.1.1 CVC-ClinicDB

Το ένα από τα σύνολα δεδομένων που χρησιμοποιήθηκαν στην παρούσα εργασία είναι το CVC-ClinicDB [27], το οποίο αποτελεί μια βάση δεδομένων με καρέ από βίντεο κολονοσκόπησης. Αυτά τα καρέ περιέχουν διάφορα παραδείγματα πολυπόδων. Το σύνολο περιέχει 612 εικόνες μεγέθους 384x288 pixels καθώς και τις αντίστοιχες μάσκες που αντιστοιχούν στις περιοχές που καλύπτονται από πολύποδες στις εικόνες, όπως φαίνεται στην εικόνα 4.1. Οι μάσκες είναι δυαδικές και εικονοστοιχεία που απεικονίζουν τον ιστό του πολύποδα είναι θετικά (λευκή μάσκα/προσκήνιο) ενώ η υπόλοιπη εικόνα είναι μαύρη.

Το σύνολο δεδομένων αποτελείται από δύο φακέλους, τον 'Original' που περιέχει τις εικόνες-καρέ σε μορφή .tiff, ο οποίος αποτελεί ιδιοκτησία του Hospital Clinic, Βαρκελώνη,

Ισπανία και τον 'Ground Truth' που περιέχει τις μάσκες των πολυπόδων στις εικόνες, επίσης σε μορφή .tiff, ο οποίος αποτελεί ιδιοκτησία του Computer Vision Center, Βαρκελώνη, Ισπανία. Στην εικόνα 4.2 φαίνεται η αντιστοιχία μεταξύ των εικόνων και των βίντεο. Το CVC-ClinicDB είναι το επίσημο σύνολο δεδομένων που χρησιμοποιήθηκε στη φάση εκπαίδευσης του διαγωνισμού MICCAI 2015 Sub-Challenge για τον Αυτόματο Εντοπισμό Πολυπόδων σε βίντεο κολονοσκόπησης. Η χρήση του είναι ελεύθερη για ερευνητικούς και εκπαιδευτικούς σκοπούς.



Εικόνα 4.1: *Παραδείγματα εικόνων-μασκών από το σύνολο CVC-ClinicDB [27]*

Sequence	1	2	3	4	5	6	7	8
Frames	1-25	26-50	51-67	68-78	79-103	104-126	127-151	152-177
Sequence	9	10	11	12	13	14	15	16
Frames	178-199	200-205	206-227	228-252	253-277	278-297	298-317	318-342
Sequence	17	18	19	20	21	22	23	24
Frames	343-363	364 -383	384 -408	409 -428	429 -447	448 -466	467 -478	479 -503
Sequence	25	26	27	28	29			
Frames	504-528	529 -546	547 -571	572-591	592-612			

Εικόνα 4.2: *Αντιστοιχία εικόνων-βίντεο για το σύνολο CVC-ClinicDB*

4.1.2 Kvasir-SEG

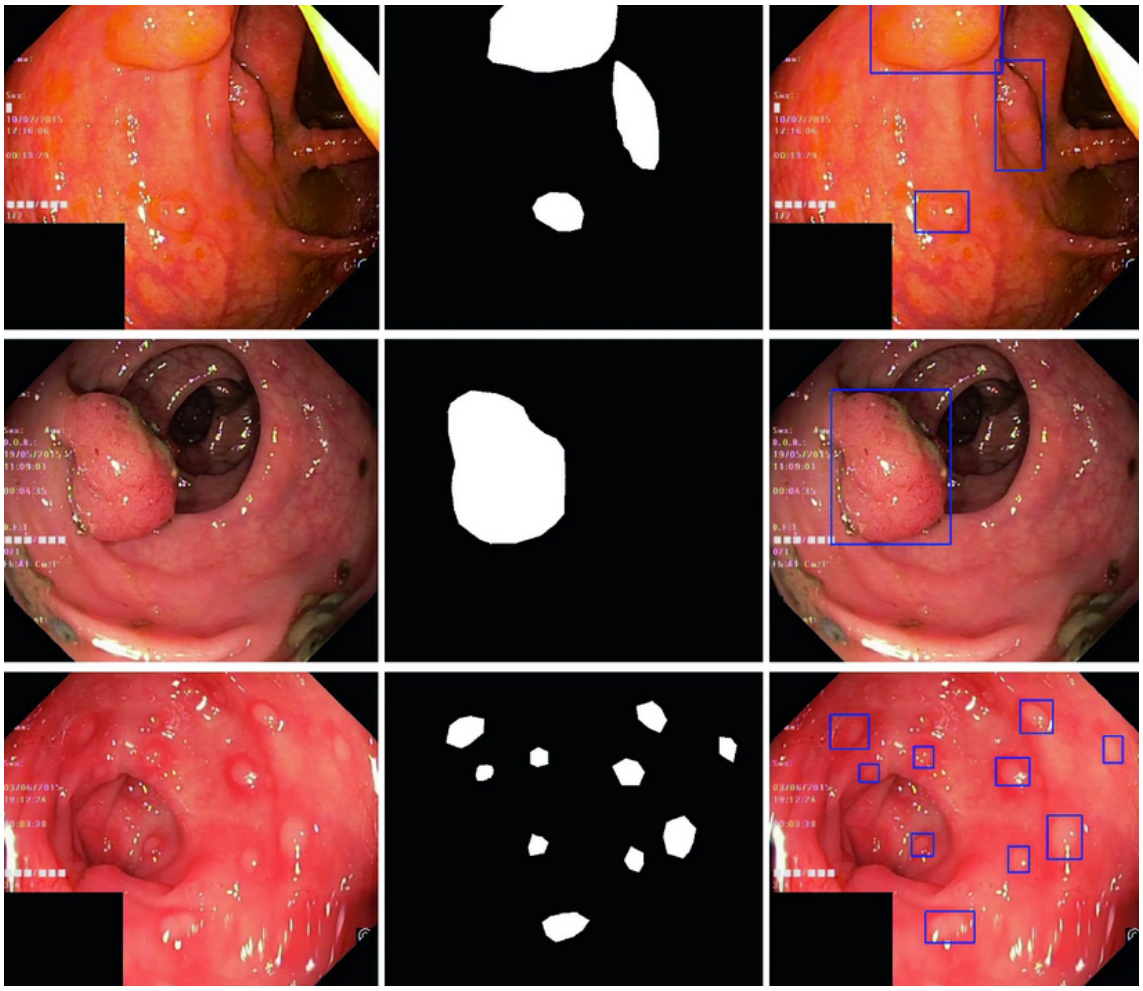
Ένα άλλο σύνολο δεδομένων που χρησιμοποιήθηκε σε αυτή την εργασία είναι το Kvasir-SEG [28]. Το Kvasir-SEG είναι ένα σύνολο δεδομένων ανοιχτής πρόσβασης με εικόνες γαστρεντερικών πολύποδων και αντίστοιχες μάσκες τμηματοποίησης, με μη αυτόματο σχολιασμό και επαλήθευση από έμπειρο γαστρεντερολόγο. Προσθέτοντας μάσκες τμηματοποίησης στο σύνολο δεδομένων Kvasir [63], το οποίο μέχρι σήμερα αποτελούνταν μόνο από

σχολιασμούς κατά πλαίσιο, δίνεται δυνατότητα στους ερευνητές όρασης υπολογιστών να συνεισφέρουν στον τομέα της τμηματοποίησης πολυπόδων και της αυτόματης ανάλυσης βίντεο κολονοσκόπησης. Η πρώιμη ανίχνευση και αξιολόγηση αυτών των πολυπόδων με επακόλουθη βιοψία και αφαίρεση των πολυπόδων έχει τεράστιο αντίκτυπο στην επιβίωση από τον καρκίνο του παχέος εντέρου. Αρκετές μελέτες έχουν δείξει ότι οι πολύποδες συχνά παραβλέπονται κατά τις κολονοσκοπήσεις, με ποσοστά απώλειας πολύποδων 14%-30% ανάλογα με τον τύπο και το μέγεθος των πολυπόδων. Η αύξηση της ανίχνευσης των πολυπόδων έχει αποδειχθεί ότι μειώνει τον κίνδυνο καρκίνου του παχέος εντέρου. Έτσι, η αυτόματη ανίχνευση περισσότερων πολυπόδων σε πρώιμο στάδιο μπορεί να διαδραματίσει κρίσιμο ρόλο στη βελτίωση τόσο της πρόληψης όσο και της επιβίωσης από τον καρκίνο του παχέος εντέρου. Αυτό είναι το κύριο κίνητρο πίσω από την ανάπτυξη του συνόλου δεδομένων Kvasir-SEG.

Το συγκεκριμένο σύνολο δεδομένων αποτελείται από 1000 εικόνες πολυπόδων και τις αντίστοιχη μάσκα τμηματοποίησης των εντοπισμένων πολυπόδων. Η ανάλυση των εικόνων ποικίλει από 332x487 έως 1920x1072 pixels. Οι εικόνες και οι αντίστοιχες μάσκες αποθηκεύονται σε δύο ξεχωριστούς φακέλους με το ίδιο όνομα αρχείου. Τα αρχεία εικόνας κωδικοποιούνται με συμπίεση JPEG. Επιπλέον, οι συντεταγμένες των πλαισίων οριοθέτησης των πολυπόδων στις εικόνες είναι αποθηκευμένες σε ένα αρχείο JSON, το οποίο ωστόσο δεν αξιοποιήθηκε στην παρούσα εργασία.

Οι μάσκες τμηματοποίησης δημιουργήθηκαν μέσω του λογισμικού Labelbox, το οποίο είναι ένα εργαλείο που χρησιμοποιείται για την επισήμανση της περιοχής ενδιαφέροντος (ROI) σε καρέ εικόνων, δηλαδή των περιοχών πολύποδων για την περίπτωσή μας. Σημειώθηκαν χειροκίνητα οι πολύποδες και στις 1000 εικόνες με τη βοήθεια ειδικών γιατρών. Τα εικονοστοιχεία που απεικονίζουν τον ιστό του πολύποδα, την περιοχή ενδιαφέροντος, αντιπροσωπεύονται από το προσκήνιο (λευκή μάσκα), ενώ το φόντο (σε μαύρο) δεν περιέχει θετικά εικονοστοιχεία. Μερικές από τις αρχικές εικόνες περιέχουν την εικόνα του καθετήρα σήμανσης θέσης ενδοσκοπίου, ScopeGuide TM, Olympus Tokyo Japan, που βρίσκεται σε μία από τις κάτω γωνίες και φαίνεται ως ένα μικρό πράσινο κουτί. Καθώς αυτές οι πληροφορίες είναι περιττές για την εργασία τμηματοποίησης, έχουν αντικατασταθεί με μαύρα κουτιά.

Τα δεδομένα συλλέχθηκαν με χρήση ενδοσκοπικού εξοπλισμού στο Vestre Viken Health Trust (VV) στη Νορβηγία. Το VV αποτελείται από 4 νοσοκομεία. Ένα από αυτά τα νοσοκομεία (το Νοσοκομείο Bærum) διαθέτει ένα μεγάλο γαστρεντερολογικό τμήμα από το οποίο έχουν συλλεχθεί δεδομένα εκπαίδευσης. Επιπλέον, οι εικόνες σχολιάζονται προσεκτικά από έναν ή περισσότερους ειδικούς γιατρούς από το VV και το Μητρώο Καρκίνου της Νορβηγίας (CRN). Το CRN παρέχει νέες γνώσεις για τον καρκίνο μέσω της έρευνας για τον καρκίνο. Αποτελεί μέρος της Περιφερειακής Αρχής Υγείας της Νοτιοανατολικής Νορβηγίας και είναι οργανωμένο ως ανεξάρτητο ίδρυμα στο πλαίσιο του Πανεπιστημιακού Νοσοκομείου του Όσλο. Το CRN είναι υπεύθυνο για τα εθνικά προγράμματα προσυμπτωματικού ελέγχου του καρκίνου με στόχο την πρόληψη του θανάτου από καρκίνο ανακαλύπτοντας καρκίνους ή προκαρκινικές βλάβες όσο το δυνατόν νωρίτερα. Η χρήση του συγκεκριμένου συνόλου δεδομένων είναι ελεύθερη για εκπαιδευτικούς και ερευνητικούς σκοπούς.



Εικόνα 4.3: *Παραδείγματα εικόνων-μασκών-πλασίων οριοθέτησης για το σύνολο Kvasir-SEG [28]*

4.1.3 2018 Data Science Bowl

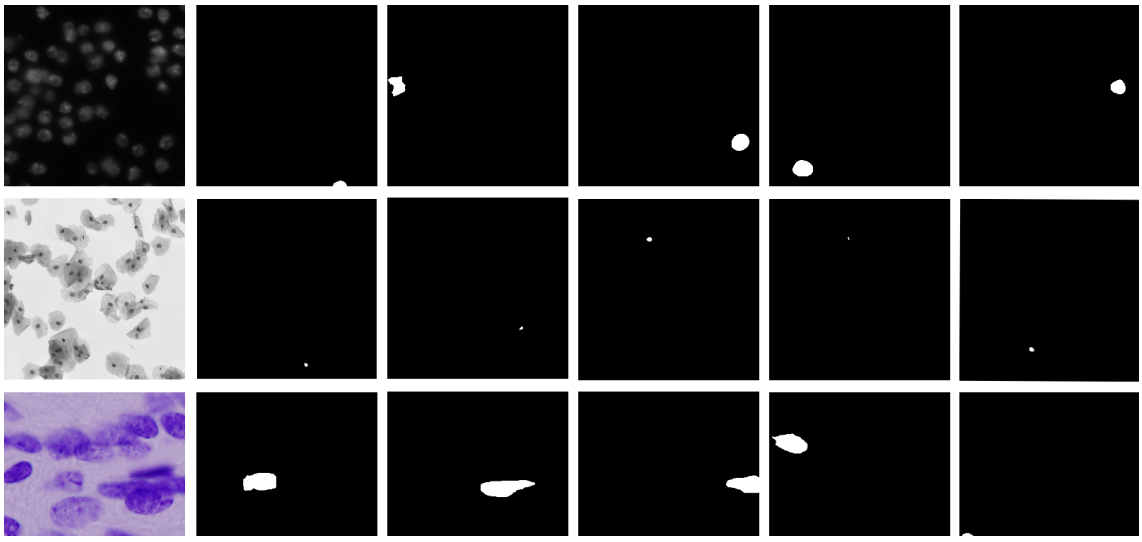
Το σύνολο δεδομένων 2018 Data Science Bowl [29] περιέχει μεγάλο αριθμό τμηματοποιημένων εικόνων πυρήνων κυττάρων, και δημιουργήθηκε στο πλαίσιο του ομώνυμου διαγωνισμού στην πλατφόρμα Kaggle. Οι εικόνες ελήφθησαν κάτω από ποικίλες συνθήκες και ποικίλλουν ως προς τον τύπο κυττάρου, τη μεγέθυνση και τον τρόπο απεικόνισης (bright-field έναντι fluorescence). Σκοπός του συνόλου δεδομένων είναι η δοκιμή και αξιολόγηση της ικανότητας ενός αλγορίθμου να αποδίδει σε αυτές τις παραλλαγές.

Κάθε εικόνα αντιπροσωπεύεται από ένα ImageId. Τα αρχεία που αντιστοιχούν σε μια εικόνα περιέχονται σε ένα φάκελο με αυτό το ImageId. Μέσα σε αυτόν τον φάκελο υπάρχουν δύο υποφάκελοι,ο ένας περιέχει το αρχείο εικόνας και ο άλλος περιέχει τις τμηματοποιημένες μάσκες κάθε πυρήνα. Κάθε μάσκα περιέχει έναν πυρήνα, έτσι σε κάθε εικόνα μπορούν να αντιστοιχούν παραπάνω από μία μάσκες. Οι μάσκες δεν επιτρέπεται να επικαλύπτονται, κανένα pixel δεν ανήκει σε δύο μάσκες.

Για τον διαγωνισμό, δημιουργήθηκε ένα σύνολο δεδομένων με 37.333 πυρήνες με μη αυτόματο σχολιασμό σε 841 2D εικόνες από περισσότερα από 30 πειράματα σε διαφορετικά δείγματα, κυτταρικές γραμμές, όργανα μικροσκοπίας, συνθήκες απεικόνισης, χειριστές,

ερευνητικές εγκαταστάσεις και πρωτόκολλα χρώσης. Οι σχολιασμοί έγιναν με μη αυτόματο τρόπο από μια ομάδα ειδικών βιολόγων του Broad Institute. Οι βιολόγοι αυτοί οριοθετούσαν χειροκίνητα κάθε αντικείμενο στις εικόνες χρησιμοποιώντας ένα από τα δύο εργαλεία: (1) ένα βοηθητικό εργαλείο σχολιασμού που προϋπολογίζει τμηματοποιήσεις superpixel για να διευκολύνει την επιλογή περιοχών στο προσκήνιο ή στο παρασκήνιο και (2) το λογισμικό επεξεργασίας εικόνων GIMP για τη δημιουργία μάσκες σχολιασμού χρωματίζοντας μεμονωμένα pixel που περιγράφουν κάθε πυρήνα. Συνολικά, το σύνολο δεδομένων περιέχει εικόνες από περισσότερα από 30 διαφορετικά βιολογικά πειράματα, τα οποία χωρίστηκαν σε 16 πειράματα για εκπαίδευση (670 εικόνες, 29,464 πυρήνες) και αξιολόγηση πρώτου σταδίου (65 εικόνες, 4,152 πυρήνες) και ακριβώς 15 πειράματα για την αξιολόγηση δεύτερου σταδίου (106 εικόνες, 3,717 πυρήνες). Στην παρούσα εργασία χρησιμοποιήθηκαν δεδομένα μόνο από την πρώτη φάση του διαγωνισμού, και μόνο το σύνολο εκπαίδευσης καθώς στο διαθέσιμο δοκιμαστικό σύνολο δεν παρέχονται οι μάσκες τμηματοποίησης.

Ο διαγωνισμός διεξήχθη για ένα σύνολο 3 μηνών κατά τους οποίους οι συμμετέχοντες είχαν πρόσβαση στο σύνολο εκπαίδευσης (με μάσκες στόχου) και στα σετ δοκιμών πρώτου σταδίου (με απουσία μάσκας στόχου). Ο κύριος στόχος του διαγωνισμού και του συνόλου δεδομένων ήταν η διερεύνηση γενικών στρατηγικών τμηματοποίησης που θα μπορούσαν να εφαρμοστούν αυτόματα σε πολλά πειράματα απεικόνισης χωρίς περαιτέρω παρέμβαση του χρήστη. Αυτή η προσέγγιση μπορεί να μειώσει τον χρόνο ποσοτικοποίησης των εικόνων, δίνοντας τη δυνατότητα στις μελλοντικές γενιές βιολόγων να υιοθετήσουν και να εκτελέσουν περισσότερα πειράματα ποσοτικής απεικόνισης για έρευνα και κλινική πρακτική.



Εικόνα 4.4: *Παραδείγματα εικόνων-μασκών για το σύνολο 2018 Data Science Bowl [29]*

4.1.4 SegPC

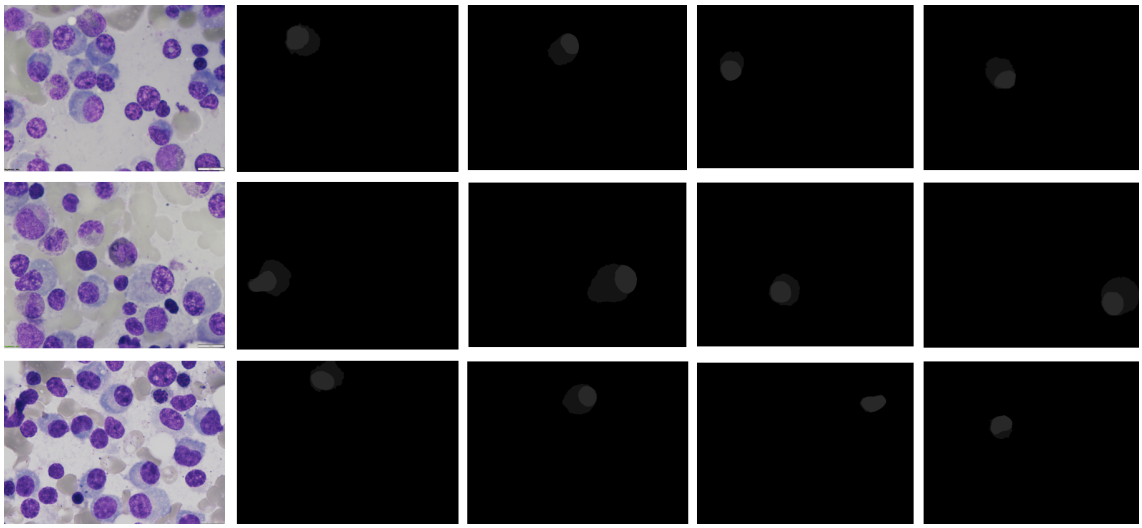
Για το συγκεκριμένο σύνολο δεδομένων, που ονομάζεται SegPC [30, 31, 32, 33] και συνοδεύει τον ομώνυμο διαγωνισμό στην πλατφόρμα Kaggle, μικροσκοπικές εικόνες καταγράφηκαν από αντικειμενοφόρους πλάκες αναρρόφησης μυελού των οστών ασθενών που είχαν διαγνωστεί με Πολλαπλό Μυέλωμα (ΜΜ), έναν τύπο καρκίνου του λευκού αίματος. Οι

αντικειμενοφόρες πλάκες χρωματίστηκαν χρησιμοποιώντας χρώση Jenner-Giemsa και τα πλασματοκύτταρα, τα οποία είναι κύτταρα ενδιαφέροντος, τμηματοποιήθηκαν. Οι εικόνες καταγράφηκαν σε ακατέργαστη μορφή BMP χρησιμοποιώντας δύο κάμερες, 1) με μέγεθος 2040x1536 pixel χρησιμοποιώντας το λογισμικό cellSens Έκδοση 2.1 (Olympus) συνδεδεμένο στο μικροσκόπιο και 2) σε μέγεθος 1920x2560 pixel από κάμερα Nikon συνδεδεμένη στο μικροσκόπιο.

Πρόσφατα, γίνονται προσπάθειες για την κατασκευή διαγνωστικών εργαλείων με τη βοήθεια υπολογιστή για τη διάγνωση του καρκίνου μέσω επεξεργασίας εικόνας. Τέτοια εργαλεία με τη βοήθεια υπολογιστή απαιτούν τη λήψη εικόνων, την κανονικοποίηση του χρώματος των εικόνων, την κατάτμηση των κυττάρων που ενδιαφέρουν και την ταξινόμηση για την καταμέτρηση των κακοήθων έναντι των υγιών κυττάρων. Αυτό το σύνολο δεδομένων σκοπεύει σε μια ισχυρή κατάτμηση των κυττάρων που είναι το πρώτο στάδιο για τη δημιουργία ενός τέτοιου εργαλείου για τον καρκίνο των κυττάρων πλάσματος, δηλαδή το πολλαπλό μυέλωμα (ΜΜ), το οποίο είναι ένας τύπος καρκίνου του αίματος. Οι εικόνες παρέχονται μετά την κανονικοποίηση του χρώματος.

Το πρόβλημα της τμηματοποίησης των κυττάρων πλάσματος στο ΜΜ είναι δύσκολο για πολλούς λόγους- 1) Υπάρχει ποικίλη ποσότητα πυρήνα και κυτταροπλάσματος από το ένα κύτταρο στο άλλο. 2) Τα κύτταρα μπορεί να εμφανίζονται σε ομάδες ή ως απομονωμένα μεμονωμένα κύτταρα. 3) Τα κύτταρα που εμφανίζονται σε συστάδες μπορεί να έχουν τρεις περιπτώσεις- (α) το κυτταρόπλασμα δύο κυττάρων αγγίζει το ένα το άλλο (β) το κυτταρόπλασμα ενός κυττάρου και ο πυρήνας ενός άλλου ακουμπούν ο ένας τον άλλον, (γ) ο πυρήνας των κυττάρων αγγίζει ο ένας τον άλλο. Δεδομένου ότι το κυτταρόπλασμα και ο πυρήνας έχουν διαφορετικά χρώματα, η κατάτμηση των κυττάρων μπορεί να δημιουργήσει προκλήσεις. 4) Μπορεί να υπάρχουν πολλά κύτταρα που αγγίζουν το ένα το άλλο στο σύμπλεγμα. 5) Μπορεί να υπάρχουν μη σεσημασμένα κύτταρα (stained cells), ας πούμε ένα ερυθρό αιμοσφαίριο κάτω από το κύτταρο ενδιαφέροντος, αλλάζοντας το χρώμα και τη σκιά του. 6) Το κυτταρόπλασμα ενός κυττάρου μπορεί να βρίσκεται κοντά στο φόντο ολόκληρης της εικόνας, καθιστώντας δύσκολη την αναγνώριση του ορίου του κυττάρου και την τμηματοποίησή του. Ως εκ τούτου, το πρόβλημα είναι πολύ προκλητικό και ενδιαφέρον. Αυτή είναι μια προσπάθεια για την κατασκευή μιας αυτοματοποιημένης διαδικασίας για την ανίχνευση καρκίνου στο πολλαπλό μυέλωμα.

Το σύνολο δεδομένων SegPC χρησιμοποιήθηκε στο σύνολο δεδομένων του διαγωνισμού ιατρικής εικόνας IEEE ISBI 2021.Το σύνολο δεδομένων εκπαίδευσης αποτελείται απο 298 εικόνες και τις αντίστοιχες τους μάσκες. Για κάθε εικόνα στον υποφάκελο x υπάρχει υπάρχουν οι μάσκες τμηματοποίησης μόνο των κυττάρων ενδιαφέροντος στον υποφάκελο y. Το σύνολο δεδομένων επικύρωσης αποτελείται απο 200 εικόνες και τις αντίστοιχες τους μάσκες, ενώ το δοκιμαστικό σύνολο αποτελείται απο 277 εικόνες χωρίς τις μάσκες τμηματοποίησης. Στις μάσκες έχουν σημειωθεί τόσο οι πυρήνες όσο και το κυτταρόπλασμα των κυττάρων, ως διαφορετικές κατηγορίες pixel. Για τον ίδιο λόγο με παραπάνω, στην παρούσα εργασία χρησιμοποιήθηκαν μόνο τα σύνολα εκπαίδευσης και επικύρωσης.



Εικόνα 4.5: *Παραδείγματα εικόνων-μασκών για το σύνολο SegPC [30, 31, 32, 33]*

4.2 Προεπεξεργασία Δεδομένων

Μερικά από τα σύνολα δεδομένων δεν βρίσκονταν εξαρχής στην μορφή που υπαγορεύει η διαδικασία των πειραμάτων, γιαυτό το λόγο χρειάστηκε μια διαδικασία προεπεξεργασίας. Συγκεκριμένα, όλα τα σύνολα χωρίστηκαν σε φακέλους εκπαίδευσης (train),επικύρωσης (validation) και δοκιμής (test) σε αναλογία 80%, 10% και 10% αντίστοιχα. Μέσα σε κάθε έναν από αυτούς τους φακέλους υπάρχουν δύο φάκελοι με τις εικόνες (images) και τις μάσκες (masks) αντίστοιχα. Για όσα σύνολα δεν υπήρχε το σύνολο δοκιμής (2018 Data Science Bowl, SegPC), αυτό δημιουργήθηκε εκ νέου από τα υπόλοιπα δεδομένα, εκπαίδευσης και επικύρωσης, στην αναλογία που αναφέρθηκε παραπάνω. Το πλήθος εικόνων σε κάθε φάκελο φαίνεται στον πίνακα 4.1 .

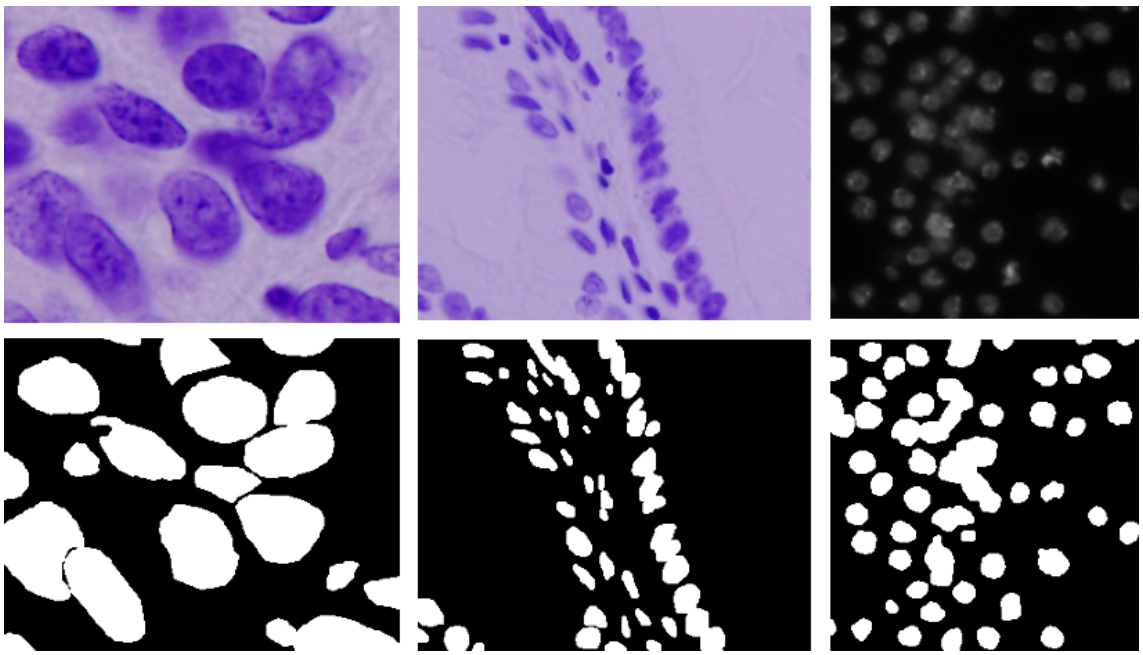
Για τα σύνολα 2018 Data Science Bowl και SegPC υπήρχε επιπλέον προεπεξεργασία διότι οι δοκιμασίες των διαγωνισμών που συνοδεύουν είναι κάπως διαφορετικές από τον σκοπό αυτής της εργασίας. Συγκεκριμένα, στο 2018 Data Science Bowl δίνεται μια ξεχωριστή μάσκα για κάθε πυρήνα που εντοπίζεται στην εικόνα. Οι μάσκες αυτές συγχωνεύθηκαν και έτσι για κάθε εικόνα υπάρχει μια μοναδική δυαδική μάσκα που περιλαμβάνει όλους τους πυρήνες. Ομοίως για το σύνολο SegPC στο οποίο για κάθε εικόνα δίνεται μια μάσκα τριών τιμών (μια τιμή για το προσκήνιο, μια τιμή για τον πυρήνα, μια τιμή για το κυτταρόπλασμα), οι μάσκες συγχωνεύονται σε μια τελική δυαδική μάσκα που περιέχει όλους τους πυρήνες της κάθε εικόνας (αγνοείται το κυτταρόπλασμα για να υπάρχει συμφωνία με το σύνολο 2018 Data Science Bowl). Οι τελικές μορφές των δύο συνόλων φαίνονται στα σχήματα 4.7 και 4.8.

Τέλος, ενώνοντας τα σύνολα Kvasir-SEG και CVC-CLinicDB δημιουργήθηκε ένα νέο μεγαλύτερο σύνολο που για ευκολία ονομάστηκε "Polyps dataset", αφού και τα δύο σύνολα προορίζονται για τμηματοποίηση πολυπόδων. Η δημιουργία αυτού του συνόλου κρίθηκε απαραίτητη για κάποια πειράματα που θα αναφερθούν σε επόμενα κεφάλαια και αφορούν την δυνατότητα γενίκευσης των μοντέλων. Ομοίως δημιουργήθηκε και το "Cells Dataset" με την συγχώνευση των 2018 Data Science Bowl και SegPC. Οι αριθμοί των δειγμάτων των

νέων αυτών συνόλων φαίνονται στον πίνακα 4.2 .

Dataset	Images	Train	Valid	Test
CVC-ClinicDB	612	489	61	62
Kvasir-SEG	1000	800	100	100
2018 Data Science Bowl	670	536	67	67
SegPC	497	397	50	50

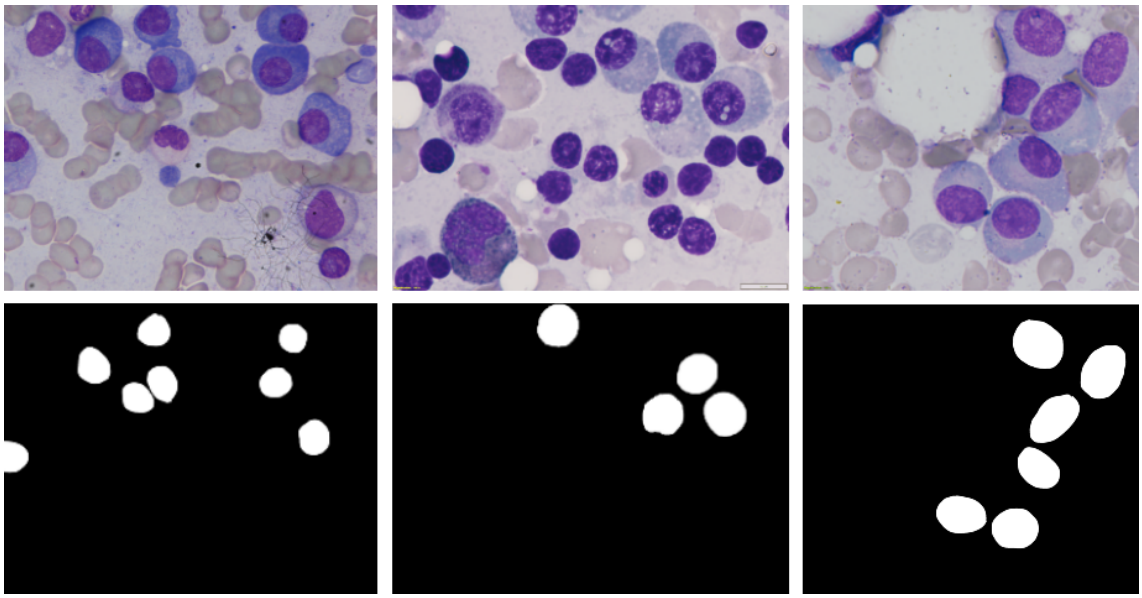
Πίνακας 4.1: *Αριθμός δειγμάτων σε κάθε σύνολο δεδομένων*



Εικόνα 4.6: *Παραδείγματα της νέας μορφής των δειγμάτων του συνόλου 2018 Data Science Bowl [29]*

Dataset	Images	Train	Valid	Test
Polyps Dataset	1612	1289	161	162
Cells Dataset	1167	933	117	117

Πίνακας 4.2: *Αριθμός δειγμάτων στα νέα σύνολα δεδομένων*



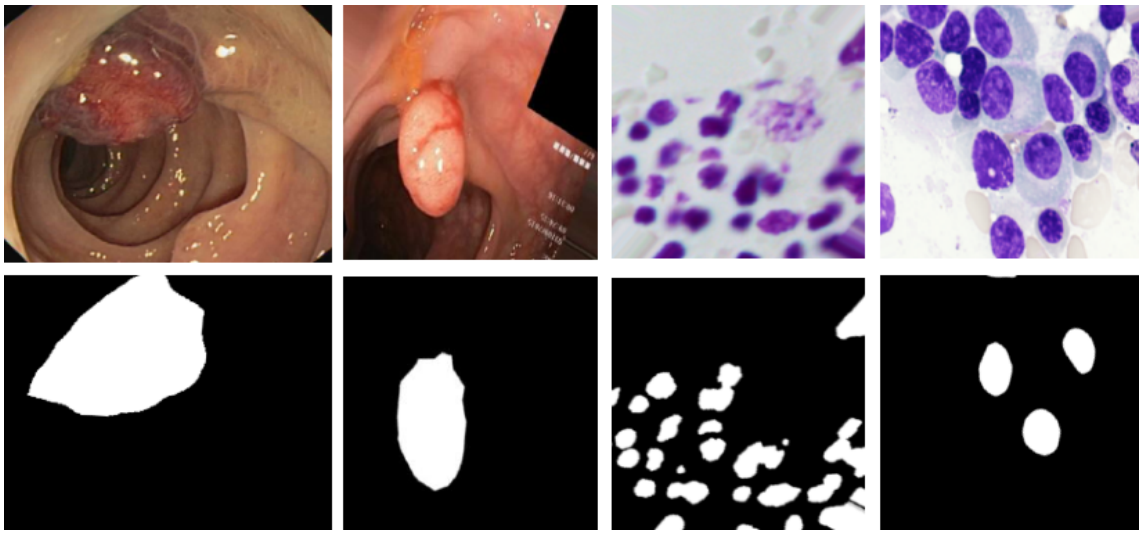
Εικόνα 4.7: *Παραδείγματα της νέας μορφής των δειγμάτων του συνόλου SegPC [30, 31, 32, 33]*

4.3 Επαύξηση Δεδομένων

Λόγω του σχετικά μικρού αριθμού δειγμάτων που υπάρχουν δημόσια διαθέσιμα για την δοκιμασία της τμηματοποίησης ιατρικής εικόνας στους τομείς που εξετάζονται στην παρούσα εργασία, κρίνεται απαραίτητη η επαύξηση των δεδομένων. Τεχνικές όπως τυχαία περιστροφή, διάτμηση, ζουμ, οριζόντια και κατακόρυφη ανατροπή εφαρμόστηκαν στα σύνολα εκπαίδευσης με αποτέλεσμα ο αριθμός των δειγμάτων να αυξηθεί αρκετά, μη ξεπερνώντας τα όρια, σε μνήμη και σε χρόνο λειτουργίας, των μηχανημάτων που είναι διαθέσιμα για την εκτέλεση των πειραμάτων της εργασίας. Χρησιμοποιήθηκε η βιβλιοθήκη Keras και το framework Tensorflow όπως και στην υπόλοιπη εργασία. Παραδείγματα των προσαυξημένων δεδομένων φαίνονται στο σχήμα 4.8 ενώ οι αριθμοί των νέων δειγμάτων φαίνονται στον πίνακα 4.3 .

Dataset	**Images**	**Train**	**Valid**	**Test**
CVC-ClinicDB	860	737 (+50%)	61	62
Kvasir-SEG	1496	1296 (+62%)	100	100
2018 Data Science Bowl	918	784 (+46%)	67	67
SegPC	745	645 (+62%)	50	50
Polyps Dataset	2356	2033 (+58%)	161	162
Cells Dataset	1664	1429 (+60%)	117	117

Πίνακας 4.3: *Αριθμός δειγμάτων στα νέα επαυξημένα σύνολα δεδομένων*



Εικόνα 4.8: *Παραδείγματα των δειγμάτων των προσαυξημένων συνόλων, από αριστερά προς τα δεξία, CVC-ClinicDB, Kvasir-SEG, 2018 Data Science Bowl, SegPC*

Κεφάλαιο 5

Υλοποίηση

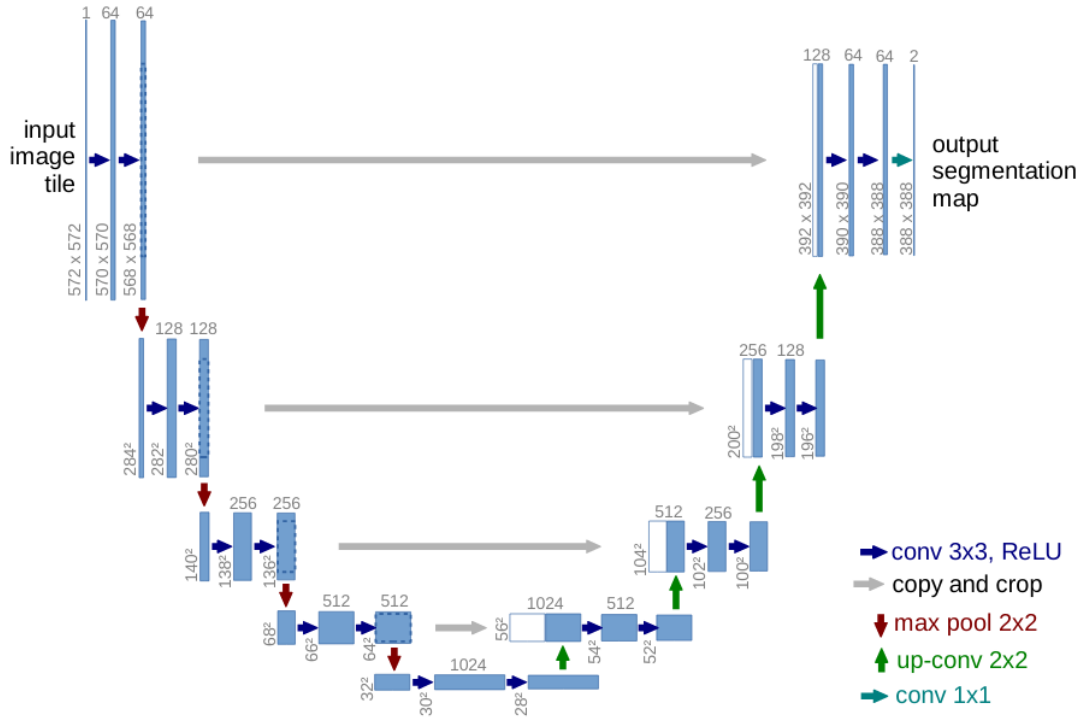
Στο κεφάλαιο αυτό παρουσιάζονται τα μοντέλα με τις αρχιτεκτονικές και τις υπερπαράμετρους τους, έτσι όπως χρησιμοποιήθηκαν στα πειράματα της παρούσας εργασίας. Επιλέχθηκαν αρκετά από τα πιο ευρέως χρησιμοποιούμενα μοντέλα στον χώρο της τμηματοποίησης ιατρικής εικόνας (και όχι μόνο) καθώς και κάποια State-of-the-Art. Επιπλέον αναφέρονται περαιτέρω λεπτομέρειες της υλοποίησης όπως οι μετρικές αξιολόγησης καθώς και το υπολογιστικό σύστημα στο οποίο εκπονήθηκε η εργασία. Ολόκληρη η εργασία έχει υλοποιηθεί με τη βιβλιοθήκη Keras και το framework Tensorflow.

5.1 Υλοποίηση με διαφορετικά μοντέλα

5.1.1 UNet

Το UNet [8] είναι ίσως το πιο βασικό εκ των μοντέλων που αφορούν την πρόκληση της τμηματοποίησης ιατρικής εικόνας, μιας και αποτέλεσε την βάση για την σχεδίαση των περισσότερων μοντέλων που ακολουθούν παρακάτω, και επιπλέον είναι πολύ επιτυχημένο. Είναι ένα βαθύ συνελικτικό δίκτυο που σχεδιάστηκε ώστε να αντιμετωπίσει το πρόβλημα του μικρού αριθμού διαθέσιμων δεδομένων, αξιοποιώντας τα δεδομένα πιο αποδοτικά.

Η αρχιτεκτονική, που απεικονίζεται στο σχήμα 5.1, αποτελείται από μια συσταλτική διαδρομή (contractive path) για την αποτύπωση του γενικου πλαισίου (context) και μια συμμετρική επεκτεινόμενη διαδρομή (expansive path) που επιτρέπει τον ακριβή εντοπισμό (localization). Η συσταλτική διαδρομή ακολουθεί την τυπική αρχικεκτονική ενός συνελικτικού δικτύου. Αποτελείται από την επαναληπτική εφαρμογή δύο συνελίξεων 3x3 η καθεμία ακολουθούμενη από μια διορθωμένη γραμμική μονάδα (ReLU) και μια διαδικασία max pooling 2x2 με βήμα 2 για υποδειγματοληψία(downsampling). Σε κάθε στάδιο υποδειγματοληψίας διπλασιάζεται ο αριθμός των καναλιών χαρακτηριστικών (feature channels). Κάθε βήμα της επεκτεινόμενης διαδρομής αποτελείται από μια υπερδειγματοληψία (upsampling) του χάρτη χαρακτηριστικών ακολουθούμενο από μια συνέλιξη 2x2 που μειώνει τα κανάλια χαρακτηριστικών στα μισά, μια συνένωση με τον αντίστοιχα περικομμένο χάρτη χαρακτηριστικών από τη συσταλτική διαδρομή και δύο συνελίξεις 3x3 ακολουθούμενες από ReLU. Στο τέλος υπάρχει ένα τελευταίο στρώμα με μια συνέλιξη 1x1 ώστε να αντιστοιχηθεί το κάθε διάνυσμα χαρακτηριστικών μεγέθους 64 στον επιθυμητό αριθμό κλάσεων (στην περίπτωση μας, οι κλάσεις είναι δύο). Συνολικά το δίκτυο διαθέτει 23 συνελικτικά στρώματα. Η σημαντική



Σχήμα 5.1: *Η αρχιτεκτονική του μοντέλου UNet [8]*

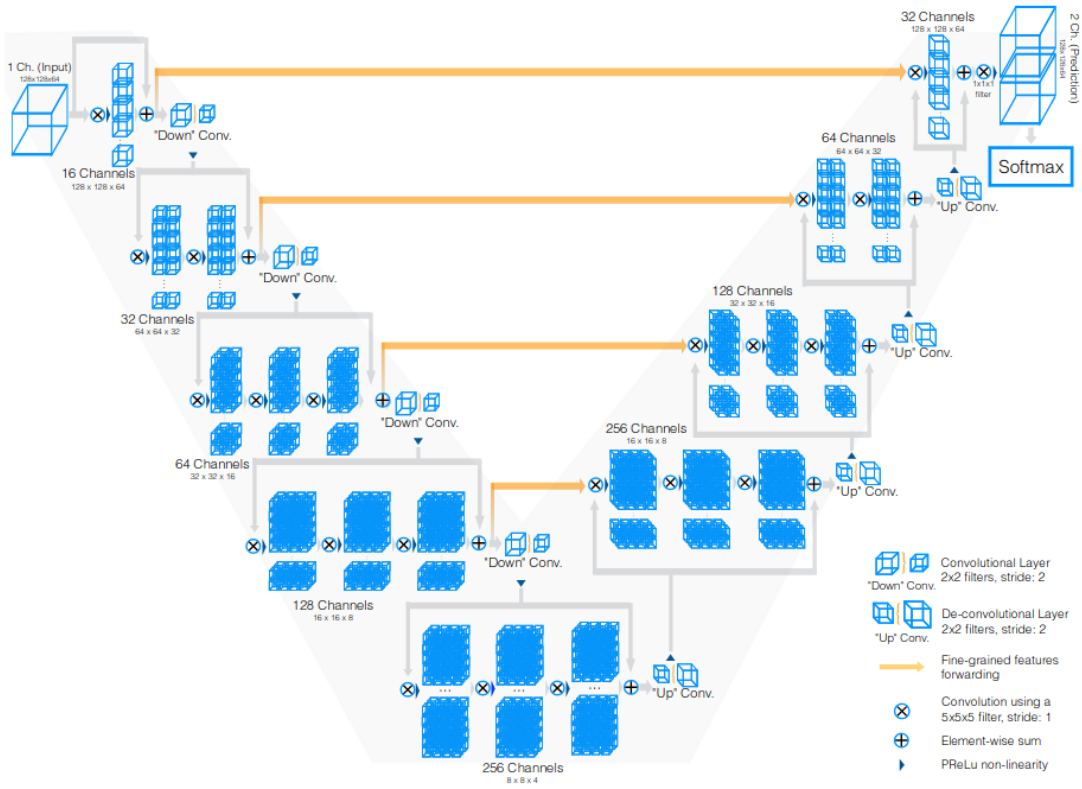
διαφοροποίηση αυτού του μοντέλου είναι ο μεγάλος αριθμός καναλιών χαρακτηριστικών στο κομμάτι της επεκτεινόμενης διαδρομής που επιτρέπει στο δίκτυο να μεταδόσει πληροφορίες για το γενικό πλαίσιο (context) σε στρώματα με μεγαλύτερη ανάλυση (resolution).

5.1.2 VNet

Το VNet [9] αποτελεί ένα βαθύ συνελικτικό δίκτυο το οποίο σχεδιάστηκε για τον τομέα της τμηματοποίησης ιατρικών δεδομένων, εμπνευσμένο από το UNet, όμως σε αντίθεση με τις υπόλοιπα συνελικτικά μοντέλα που παρουσιάζονται, το συγκεκριμένο λειτουργεί σε τρισδιάστατα (3D) δεδομένα (MRI scans) αντί για δισδιάστατες (2D) εικόνες (ωστόσο στη δική μας περίπτωση τα δεδομένα εισόδου είναι εικόνες, οπότε χρησιμοποιήθηκε μια τροποποιημένη μορφή του VNet).

Η αρχιτεκτονική του, που φαίνεται στο σχήμα 5.2, μοιάζει αρκετά με αυτή του UNet καθώς αποτελείται από ένα συσταλτικό μονοπάτι και ένα επεκτατικό μονοπάτι. Το αριστερό κομμάτι του δικτύου είναι χωρισμένο σε διαφορετικά στάδια που λειτουργούν σε διαφορετική ανάλυση (resolution) και αποτελούνται από ένα έως τρία συνελικτικά στρώματα. Όπως και στο [64] τα στάδια τροποποιούνται ώστε να μαθαίνουν μια υπολειπόμενη συνάρτηση (residual function), συγκεκριμένα η είσοδος κάθε σταδίου εκτός από το ότι περνάει από τα συνελικτικά στρώματα επιπλέον περνάει στην έξοδο του σταδίου αυτούσια. Η τροποποίηση αυτή μειώνει τον χρόνο σύγκλισης.

Στις συνελίξεις κάθε σταδίου χρησιμοποιούνται πυρήνες μεγέθους 5x5x5. Από το ένα στάδιο στο επόμενο, η ανάλυση των δεδομένων μειώνεται μέσω συνέλιξης με πυρήνες μεγέθους 2x2x2 και βήμα 2. Επειδή η λειτουργία αυτή μειώνει στο μισό το μέγεθος των χαρ-



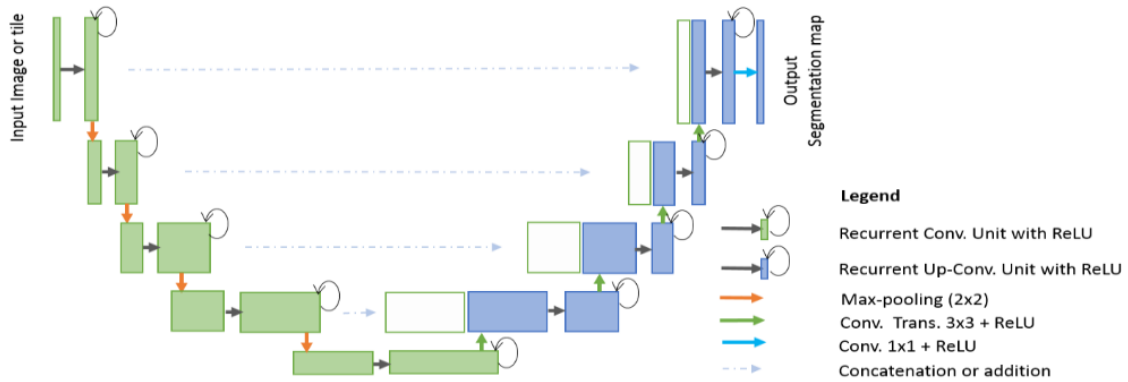
Σχήμα 5.2: *Η αρχιτεκτονική του μοντέλου VNet [9]*

τών χαρακτηριστικών (feature maps), προτείνεται συνελικτικό στρώμα για την αντικατάσταση των στρωμάτων pooling. Αποδεικνύεται, ότι αυτή η αλλαγή οδηγεί σε δίκτυα με μικρότερες απαιτήσεις μνήμης, μιας και η διαδικασία αποσυνέλιξης (deconvolution) είναι πιο απλή από την διαδικασία απο-ομαδοποίησης (un-pooling). Κάθε στάδιο του αριστερού κομματιού του δικτύου υπολογίζει διπλάσιο αριθμό χαρακτηριστικών από το προηγούμενο του. PReLU και μη-γραμμικά στρώματα χρησιμοποιούνται σε όλα τα στάδια. Το δεξί κομμάτι του δικτύου εξάγει χαρακτηριστικά και αξιοποιώντας τους πίνακες χαρακτηριστικών (feature maps) των προηγούμενων σταδίων χαμηλότερης ανάλυσης παράγει τελικά μια ογκομετρική τμηματοποίηση η οποία μετατρέπεται στην έξοδο σε πιθανοτική τμηματοποίηση του προσκηνίου και του παρασκηνίου μέσω μιας Soft-max συνάρτησης που εφαρμόζεται στα ογκοστοιχεία (voxel). Αντίστοιχα με το αριστερό κομμάτι, ανάμεσα στα στάδια εκτελείται συνέλιξη για να αυξηθεί το μέγεθος των δεδομένων. Ομοίως με το UNet, χαρακτηριστικά που εξάγονται από τα πρώτα στάδια του αριστερού κομματιού του δικτύου περνούν στο δεξί κομμάτι, και έτσι σώζεται πολύ λεπτή (fine-grained) πληροφορία που διαφορετικά θα χανόταν στο συσταλτικό μονοπάτι. Αυτές οι συνδέσεις βοηθούν επιπλέον το μοντέλο στην γρηγορότερη σύγκλιση.

5.1.3 R2U-Net

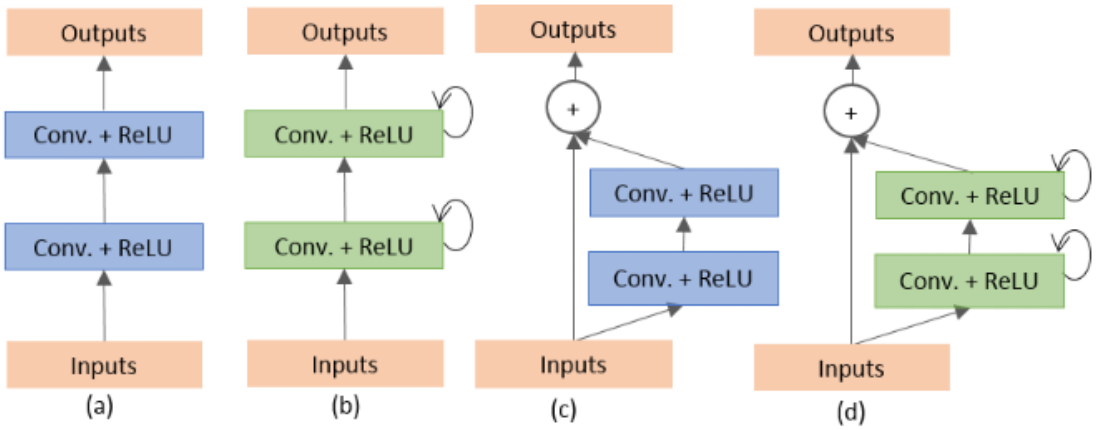
Στην δημοσίευση [10], παρουσιάζονται δύο μοντέλα, το RU-Net που αποτελεί ένα Επαναληπτικό Συνελικτικό Δίκτυο (RCNN) βασισμένο στο UNet, και το R2U-Net που αποτελεί ένα Επαναληπτικό Υπολειμματικό Συνελικτικό Δίκτυο (RRCNN) βασισμένο και αυτό στο UNet. Τα μοντέλα αυτά συνδυάζουν τα δυνατά σημεία του UNet, των υπολειμματικών δι-

κτύων (Residual Networks) και των RCNN, και έτσι εμφανίζουν αρκετά πλεονεκτήματα στην διαδικασία τμηματοποίησης ιατρικής εικόνας. Συγκεκριμένα, τα υπολειμματικά κομμάτια συμβάλλουν σε γρηγορότερη εκπαίδευση [64], η συσσώρευση χαρακτηριστικών που προσφέρουν τα RRCNN [65] εξασφαλίζει καλύτερες αναπαραστάσεις των χαρακτηριστικών, και τέλος επιτρέπεται η μείωση των παραμέτρων του δικτύου σε σχέση με το UNet.



Σχήμα 5.3: *Η αρχιτεκτονική του μοντέλου R2U-Net [10]*

Η αρχιτεκτονική των μοντέλων φαίνεται στο σχήμα 5.3. Πιο συγκεκριμένα η αρχιτεκτονική που απεικονίζεται είναι του RU-Net, και η διαφορά με το R2U-Net είναι ότι μαζί με τα επαναληπτικά συνελικτικά στρώματα (RCL) υπάρχουν και υπολειμματικές μονάδες (residual units). Εξετάζονται τέσσερις διαφορετικές παραλλαγές αρχιτεκτονικών στο πλαίσιο της δημοσίευσης [10], οι οποίες φαίνονται στο σχήμα 5.4 , και τονίζουν τις διαφορές των νέων μοντέλων που προτείνονται έναντι των προγενέστερων τους.



Σχήμα 5.4: *Διαφορετικές αρχιτεκτονικές που ελέγχθηκαν (α) Εμπρός συνελικτικές μονάδες, (β) Επαναληπτικό συνελικτικό μπλοκ (γ) Υπολειμματική συνελικτική μονάδα και (δ) Επαναληπτικές Υπολειμματικές συνελικτικές μονάδες (RRCU) [10]*

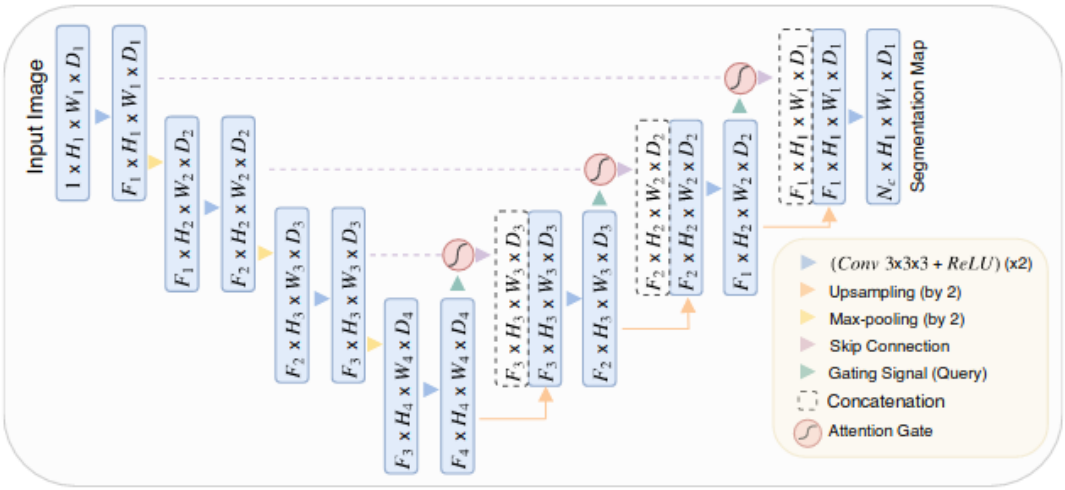
Η πρώτη από αυτές είναι το UNet με εμπρός συνελικτικά στρώματα και συνένωση χαρακτηριστικών, αντί για την μέθοδο περικοπής που είχε εφαρμοστεί στην αρχική μορφή του UNet (σχήμα 5.3(α)) και η δεύτερη είναι το UNet με εμπρός συνελικτικά στρώματα με υπολειμματική σύνδεση, που συνήθως αποκαλείται ResUNet (σχήμα 5.3(γ)). Η τρίτη

αρχιτεκτονική είναι το UNet με εμπρός επαναληπτικά συνελικτικά στρώματα που ονομάζεται RU-Net (σχήμα 5.4(β)). Η τέταρτη και τελευταία είναι το UNet με επαναληπτικά συνελικτικά δίκτυα με υπολειμματική σύνδεση, που ονομάζεται R2U-Net (σχήμα 5.4(δ)).

Στην υλοποίηση των RU-Net και R2U-Net εφαρμόζεται συνένωση των χαρακτηριστικών (concatenation) από τον κωδικοποιητή στον αποκωδικοποιητή. Σύμφωνα με την παραπάνω ανάλυση παρατηρείται ότι τα νέα μοντέλα μοιάζουν στην βάση τους με το γνωστό UNet, με την διαφορά ότι χρησιμοποιούνται επαναληπτικά συνελικτικά στρώματα (RCL) με ή χωρίς υπολειμματικές μονάδες αντί για κανονικά εμπρός συνελικτικά στρώματα. Έτσι δημιουργούνται βαθύτερα μοντέλα τα οποία είναι επίσης και πιο αποδοτικά λόγω της συσσώρευσης χαρακτηριστικών από διάφορα στάδια. Εκτός από τον μικρότερο χρόνο σύγκλισης, η νέες αυτές αρχιτεκτονικές οδηγούν σε καλύτερη εκπαίδευση και καλύτερα αποτελέσματα, σε δυνατότερες αναπαραστάσεις χαρακτηριστικών, χωρίς να αυξάνεται η πολυπλοκότητα του μοντέλου και ο αριθμός των παραμέτρων.

5.1.4 Attention UNet

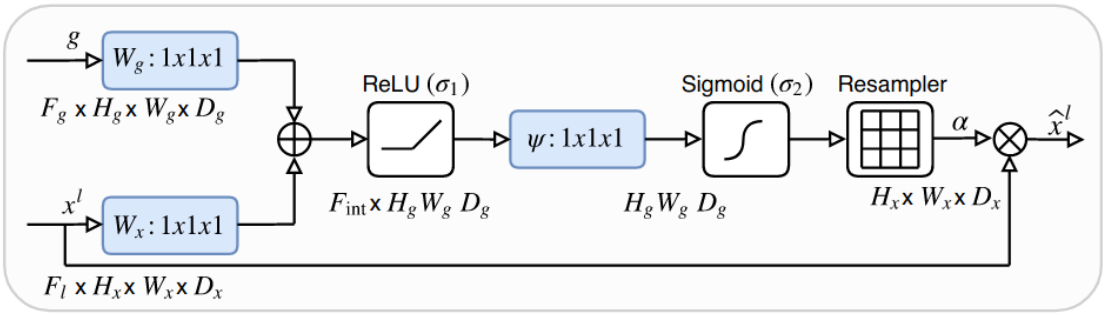
Το Attention UNet [11] είναι ένα μοντέλο με πύλες attention για ιατρικές εικόνες που μαθαίνει αυτόματα να εστιάζει σε δομές-στόχους διαφορετικού μεγέθους και σχήματος. Τα μοντέλα που διαθέτουν αυτές τις πύλες μαθαίνουν να αγνοούν μικρής σημασίας περιοχές και να εστιάζουν στα σημαντικά χαρακτηριστικά ανάλογα με την διεργασία που εκτελούν. Επιπλέον αφαιρούν την ανάγκη για εξωτερικά CNN modules που ασχολούνται με το localization, μειώνοντας την πολυπλοκότητα. Οι πύλες attention μπορούν να ενσωματωθούν έυκολα σε CNN μοντέλα όπως το UNet χωρίς την απαίτηση επιπλέον πόρων, ταυτόχρονα αυξάνοντας την ευαισθησία και την ακρίβεια του μοντέλου.



Σχήμα 5.5: *Η αρχιτεκτονική του μοντέλου Attention UNet [11]*

Η αρχιτεκτονική του Attention UNet παρουσιάζεται στο σχήμα 5.5. Οι εικόνες που αποτελούν την είσοδο φιλτράρονται και υποδειγματολειπτούνται με συντελεστή 2 σε κάθε στάδιο του κομματιού του μοντέλου που εκτελεί την κωδικοποίηση. Οι πύλες attention φιλτράρουν τα χαρακτηριστικά που περνούν από τα skip connections. Η αρχιτεκτονική

των attention πυλών φαίνεται στο σχήμα 5.6 . Οι πύλες αυτές ενσωματώνονται στο UNet προκειμένου να δώσουν έμφαση στα σημαντικά χαρακτηριστικά που περνούν από τα skip connections.



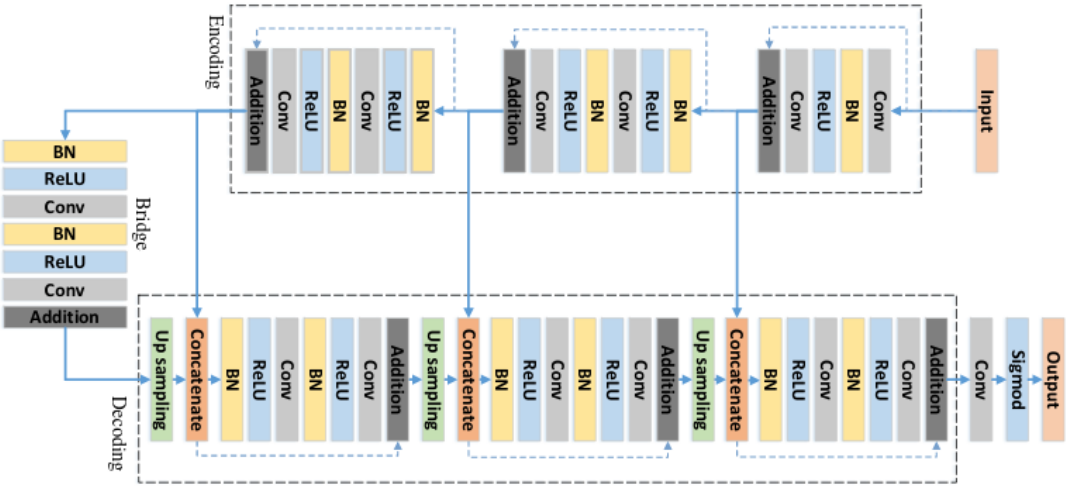
Σχήμα 5.6: *Η αρχιτεκτονική των Attention Gates [11]*

Πληροφορίες από τραχείς κλίμακες (coarse) αξιοποιούνται στις πύλες για να αποσαφηνίσουν θορυβώδεις αποκρίσεις στα skip connections. Αυτό συμβαίνει πριν από την συνένωση για να διατηρηθούν μόνο σχετικές ενεργοποιήσεις. Επιπλέον οι πύλες φιλτράρουν τις ενεργοποιήσεις των νευρώνων τόσο στο forward όσο και στο backward pass. Χρησιμοποιείται βαθιά επίβλεψη προκειμένου να διασφαλιστεί ότι τα attention units στις διαφορετικές κλίμακες έχουν τη δυνατότητα να επηρεάσουν τις αποκρίσεις.

5.1.5 ResUNet

Το ResUNet [12] αποτελεί άλλο ένα μοντέλο εμπνευσμένο από το UNet. Συνδυάζει τα δυνατά χαρακτηριστικά της υπολειπόμενης μάθησης (residual learning) και του UNet και σχεδιάστηκε για το έργο τμηματοποίησης οδικής περιοχής. Όπως αναφέρθηκε και προηγουμένως, ο υπολειμματικός χαρακτήρας βοηθά αρκετά στην εκπαίδευση των μοντέλων. Αυτό που το διαφοροποιεί από το UNet είναι η χρήση υπολειμματικών μονάδων (residual units) αντί για απλές νευρωνικές μονάδες ως βασικό block. Επιπλέον, το κομμάτι που περιλαμβάνει την περικοπή των εικόνων (cropping) αφαιρείται μιας και δεν κρίνεται απαραίτητο για την πρόκληση που αντιμετωπίζει.

Η αρχιτεκτονική του ResUNet, η οποία φαίνεται στο σχήμα 5.7, αποτελείται από 7 επίπεδα και απαρτίζεται από 3 μέρη, την κωδικοποίηση, την γέφυρα και την αποκωδικοποίηση. Το πρώτο μέρος κωδικοποιεί την είσοδο σε συμπαγείς αναπαραστάσεις, ενώ το τελευταίο μέρος επαναφέρει τις αναπαραστάσεις σε κατηγοριοποίηση επιπέδου pixel. Η γέφυρα ένωνει τα δύο μέρη. Και τα 3 μέρη αποτελούνται από residual blocks που με την σειρά τους αποτελούνται από δύο 3x3 συνελικτικά μπλοκ και ένα identity mapping. Τα συνελικτικά μπλοκ αποτελούνται από ένα στρώμα batch normalization, ένα στρώμα ενεργοποίησης ReLU και ένα συνελικτικό στρώμα. Το identity mapping συνδέει την είσοδο και την έξοδο της κάθε μονάδας. Το μονοπάτι της κωδικοποίησης διαθέτει 3 υπολειμματικές μονάδες, στις οποίες δεν εκτελείται pooling αλλά εισάγεται βήμα 2 στην συνέλιξη ώστε να μειωθεί ο χάρτης χαρακτηριστικών στο μισό μέγεθος. Αντίστοιχα, το μονοπάτι της αποκωδικοποίησης αποτελείται από 3 υπολειπόμμενες μονάδες, ανάμεσα στις οποίες εκτελείται υπερδειγματοληψία



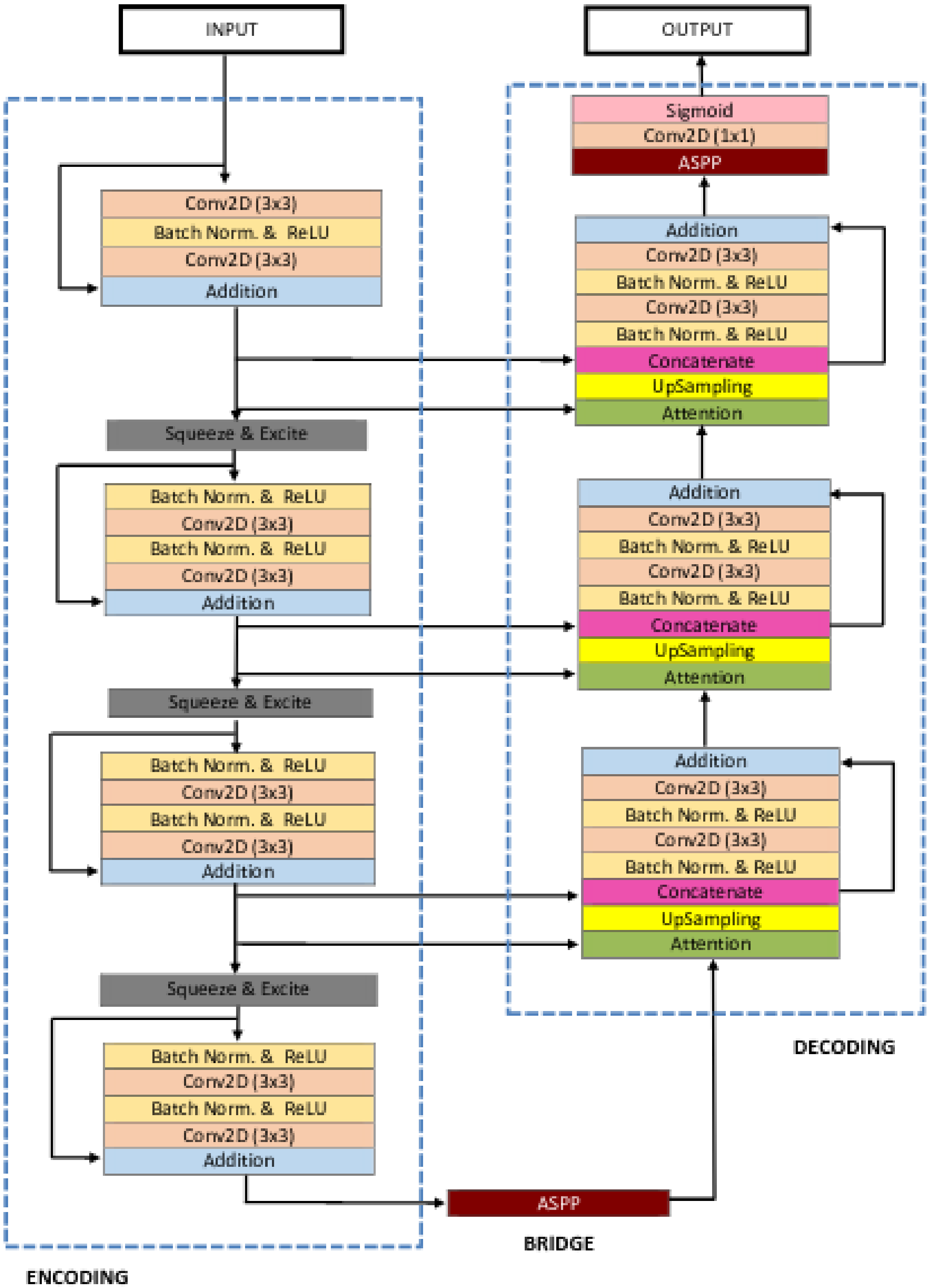
Σχήμα 5.7: *Η αρχιτεκτονική του ResUNet [12]*

των χαρακτηριστικών και συνένωση με τα χαρακτηριστικά του αντίστοιχου σταδίου του μονοπατιού κωδικοποίησης. Στο τελευταίο επίπεδο του μονοπατιού αποκωδικοποίησης υπάρχει μια τελευταία συνέλιξη 1x1 και μια σιγμοειδής ενεργοποίηση προκειμένου να προβληθούν τα χαρακτηριστικά στο επίπεδο των τελικών κλάσεων. Συγκριτικά με τα 23 στρώματα του UNet το ResUNet αποτελείται από 15 συνελικτικά στρώματα.

5.1.6 ResUNet++

Το ResUNet++ [13] παρουσιάστηκε ως μια βελτιωμένη εκδοχή του ResUNet για το έργο της τμηματοποίησης εικόνων κολονοσκόπησης. Το μοντέλο αυτό εκτελεί semantic segmentation και αξιοποιεί υπολειπόμμενες μονάδες, squeeze and excitation μονάδες, Atrous Spatial Pyramidal Pooling (ASSP) και attention blocks. Τα residual blocks μεταφέρουν την πληροφορία μεταξύ των στρωμάτων φτιάχνοντας ένα βαθύ δίκτυο και βελτιώνοντας τις εξαρτήσεις ανάμεσα στα κανάλια και μειώνοντας το υπολογιστικό κόστος.

Η αρχιτεκτονική του μοντέλου, που απεικονίζεται στο σχήμα 5.8, αποτελείται από ένα αρχικό μπλοκ ακολουθούμενο από 3 μπλοκ κωδικοποίησης, μια γέφυρα με ASPP, και 3 μπλοκ αποκωδικοποίησης. Όπως και παραπάνω, οι υπολειπόμμενες μονάδες (residual units) αποτελούν έναν συνδυασμό batch normalization, ενεργοποίησης ReLU και συνελικτικών στρωμάτων. Κάθε μπλοκ κωδικοποίησης αποτελείται από 2 συνελικτικά μπλοκ 3x3 και ένα identity mapping. Η έξοδος των μπλοκ κωδικοποίησης περνάει αρχικά από ένα squeeze -and-excitation μπλοκ, σκοπός του οποίου είναι να διασφαλίσει ότι το δίκτυο μπορεί να βελτιώσει την ευαισθησία του στα σχετικά χαρακτηριστικά και να υποβαθμίσει τα μη απαραίτητα χαρακτηριστικά, και επιπλέον βελτιώνει το generalization του μοντέλου. Στη συνέχεια τα δεδομένα περνούν από το ASPP που λειτουργεί σαν γέφυρα και αναδειγματοληπτεί τα χαρακτηριστικά σε διάφορες κλίμακες προσφέροντας έτσι σημαντική πληροφορία στο μοντέλο και μεγενθύνοντας το field-of-view των φίλτρων, δηλαδή το μέγιστο εύρος δεδομένων στο οποίο μπορούν να εκτεθούν τα φίλτρα. Αντίστοιχα, το μονοπάτι αποκωδικοποίησης α-



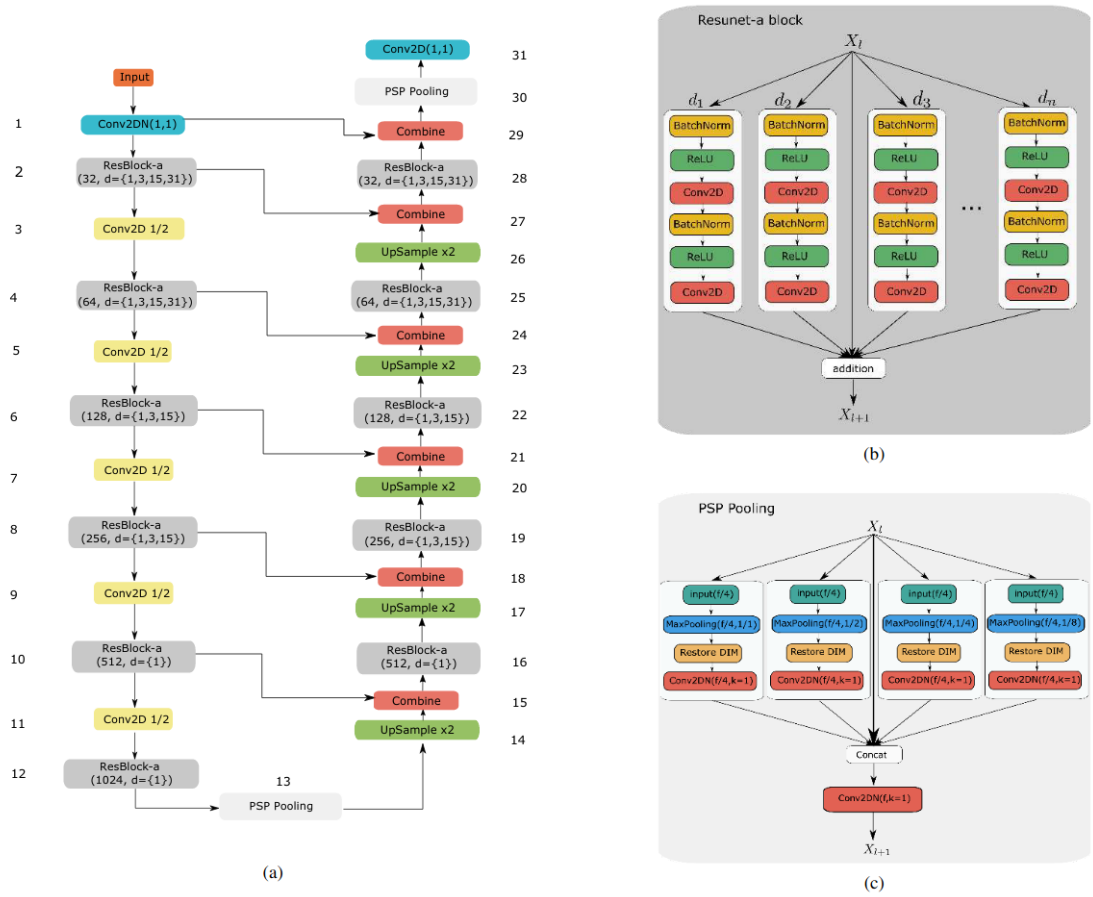
Σχήμα 5.8: *Η αρχιτεκτονική του ResUNet++ [13]*

ποτελείται και αυτό από residual μπλοκ. Πριν από κάθε μπλοκ, το μπλοκ attention αυξάνει την αποτελεσματικότητα των χαρακτηριστικών. Στη συνέχεια ακολουθεί υπερδειγματοληψία κοντινότερου γείτονα, και συνένωση με τα χαρακτηριστικά του αντίστοιχου σταδίου του μονοπατιού κωδικοποίησης. Η έξοδος στη συνέχεια περνάει από ASPP και τελικά από μια

συνέλιξη 1x1 και μια σιγμοειδή ενεργοποίηση, που δίνει τον τελικό χάρτη τμηματοποίησης.

5.1.7 ResUNet-a

Το ResUNet-a [14] είναι ένα βαθύ πλήρως συνελικτικό δίκτυο (FCN) που σχεδιάστηκε για την σημασιολογική τμηματοποίηση (semantic segmentation) εναέριων εικόνων πολύ μεγάλης ανάλυσης. Το ResUNet-a διαθέτει ένα backbone κωδικοποιητή-αποκωδικοποιητή σαν το UNet, σε συνδυασμό με υπολειπόμμενες συνδέσεις (residual connections), τραχείς συνελίξεις (atrous convolutions) [46] και pyramid scene parsing pooling [66]. Το ResUNet-a παράγει σειριακά το περίγραμμα των αντικειμένων, τον distance μετασχηματισμό των μασκών τμηματοποίησης, τις μάσκες τμηματοποίησης, και μια έγχρωμη ανακατασκευή της εισόδου.



Σχήμα 5.9: *(α) Η αρχιτεκτονική του ResUNet-a (β) Το building block του ResUNet-a (ResBlock-a) (γ) Το στρώμα pyramid scene parse pooling (PSPP) [14]*

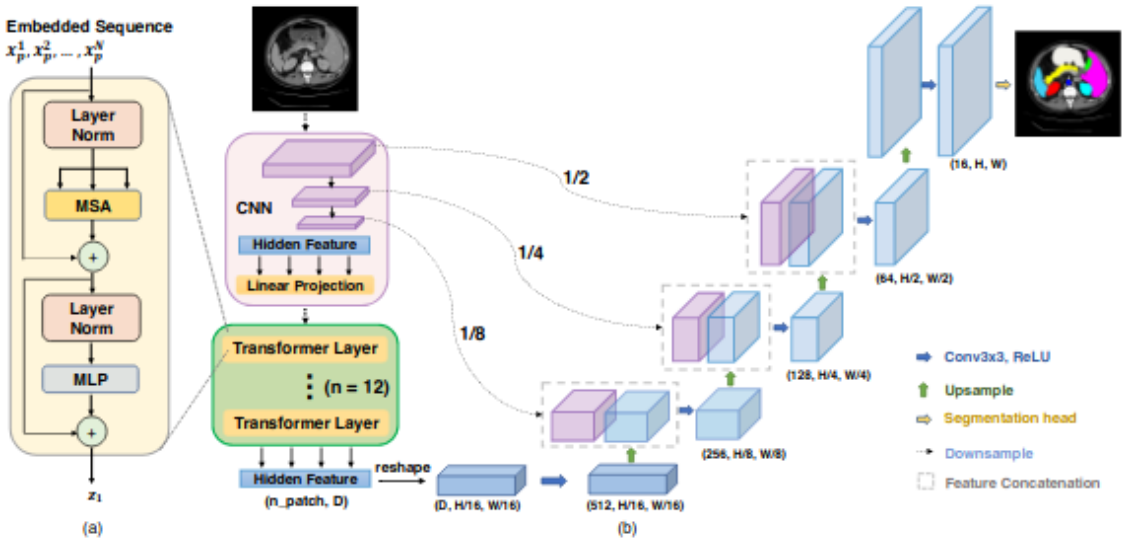
Η αρχιτεκτονική του δικτύου, που φαίνεται στο σχήμα 5.9, αποτελείται από τα μονοπάτια κωδικοποίησης και αποκωδικοποίησης του UNet, όπου τα μπλοκ του UNet αντικαθίστανται από residual συνελικτικά μπλοκ, τα οποία λύνουν το πρόβλημα των vanishing και exploding gradients. Επιπλέον, πολλαπλές τραχείς συνελίξεις (atrous) με διαφορετικούς ρυθμούς διαστολής εκτελούνται μέσα στα residual μπλοκ βοηθώντας στην καλύτερη κατανόηση σε πολλές κλίμακες. Για να βελτιωθεί η απόδοση του μοντέλου συμπεριλαμβάνοντας πληροφορίες για το παρασκήνιο, χρησιμοποιείται pyramid scene parsing pooling. Παρατίθενται δύο

μοντέλα ResUNet-a, το d6 και το d7 που διαφέρουν στο βάθος τους, στο ένα ο κωδικοποιητής αποτελείται από 6 ResBlock-a και ένα PSP Pooling στρώμα και το άλλο από 7 ResBlock-a. Η είσοδος αρχικά περνάει από μια 1x1 συνέλιξη για να αυξηθεί ο αριθμός των χαρακτηριστικών στο επιθυμητό μέγεθος του φίλτρου. Έπειτα ακολουθεί μια αλληλουχία από ResBlock-a. Σε κάθε μπλοκ χρησιμοποιήθηκαν έως και 3 παράλληλες τραχείς συνελίξεις εκτός από τις κλασσικές 2 συνελίξεις των residual μπλοκ. Στη συνέχεια η έξοδος προστίθεται στην είσοδο. Από κάθε μπλοκ στο επόμενο, η έξοδος υποδειγματοληπτείται μέσω συνέλιξης 1x1 και βήματος 2. Στο τέλος τόσο του κωδικοποιητή όσο και του αποκωδικοποιητή υπάρχει ένα στρώμα pyramid scene parsing pooling. Σε αυτό το στρώμα η αρχική είσοδος χωρίζεται σε 4 κομμάτια στον χώρο χαρακτηριστικών (feature space), στη συνέχεια εκτελείται max pooling σε διαδοχικές διχοτομήσεις, σε 1,4,16,64 κομμάτια. Στον αποκωδικοποιητή, η υπερδειγματοληψία γίνεται με την τεχνική κοντινότερου γείτονα και ακολουθείται από μια συνέλιξη 1x1 και batch normalization. Στρώματα του κωδικοποιητή και του αποκωδικοποιητή συνενώνονται μέσω του στρώματος Combine, και περνούν από μια συνέλιξη που φέρνει τον αριθμό των χαρακτηριστικών στην επιθυμητή τιμή.

5.1.8 TransUNet

Το TransUNet [15] αποτελεί ένα μοντέλο που συνδυάζει transformers και το UNet, και αποτελεί μια δυνατή εναλλακτική στο έργο της τμηματοποίησης ιατρικής εικόνας. Ενώ η u-αρχιτεκτονική με συνελικτικά δίκτυα του UNet έχει επικρατήσει στον τομέα της ιατρικής εικόνας, παρουσιάζει κάποιους περιορισμούς στην μοντελοποίηση εξαρτήσεων μεγάλης εμβέλειας [67]. Οι μετασχηματιστές έχουν εμφανιστεί ως εναλλακτικές αρχιτεκτονικές με εσωτερικούς μηχανισμούς global attention, με το μειονέκτημα ότι μπορεί να οδηγήσουν σε χαμηλή ικανότητα localization λόγω έλλειψης λεπτομερειών χαμηλού επιπέδου. Ο μετασχηματιστής στο TransUNet κωδικοποιεί tokenized patches εικόνων από ένα χάρτη χαρακτηριστικών (feature map) συνελικτικού νευρωνικού δικτύου (CNN) ως ακολουθία εισόδου για να εξάγει global contexts. Ο αποκωδικοποιητής υπερδειγματοληπτεί τα κωδικοποιημένα χαρακτηριστικά που συνδυάζονται με τα υψηλής ανάλυσης χαρακτηριστικά του CNN για να επιτευχθεί ακριβές localization. Αποδεικνύεται ότι οι μετασχηματιστές μπορούν να λειτουργήσουν αποδοτικά ως κωδικοποιητές, και σε συνδυασμό με το UNet που ενισχύει τις λεπτομέρειες επαναφέροντας δεδομένα localization.

Η αρχιτεκτονική του TransUNet απεικονίζεται στο σχήμα 5.10 . Η διαφοροποίηση του από την κλασσική u-αρχιτεκτονική φαίνεται στο μονοπάτι κωδικοποίησης, το οποίο έχει αντικατασταθεί από έναν υβριδικό μετασχηματιστή-CNN. Το CNN χρησιμοποιείται για την εξαγωγή χαρακτηριστικών της εισόδου. Αρχικά εκτελείται tokenization της εισόδου σε μια ακολουθία από flattened 2D patches. Τα διανυσματικά patches προβάλλονται σε latent space. Για την κωδικοποίηση της χωρικής πληροφορίας των patches, γίνεται εκμάθηση της θέσης των embeddings. Ο μετασχηματιστής-κωδικοποιητής αποτελείται από στρώματα από Multihead Self-Attention (MSA) και Multi-Layer Perceptron (MLP) μπλοκ. Patch embedding εφαρμόζεται σε patches 1x1 που προκύπτουν από τον χάρτη χαρακτηριστικών του CNN, αντί για τις εικόνες εισόδου. Αυτή η τεχνική προτιμάται διότι βοηθάει στην αξιοποίηση των χαρακτηριστικών υψηλής ανάλυσης του CNN στην αποκωδικοποίηση, και



Σχήμα 5.10: *(a) Η αρχιτεκτονική του στρώματος του Μετασχηματιστή (β) Η αρχιτεκτονική του TransUNet [15]*

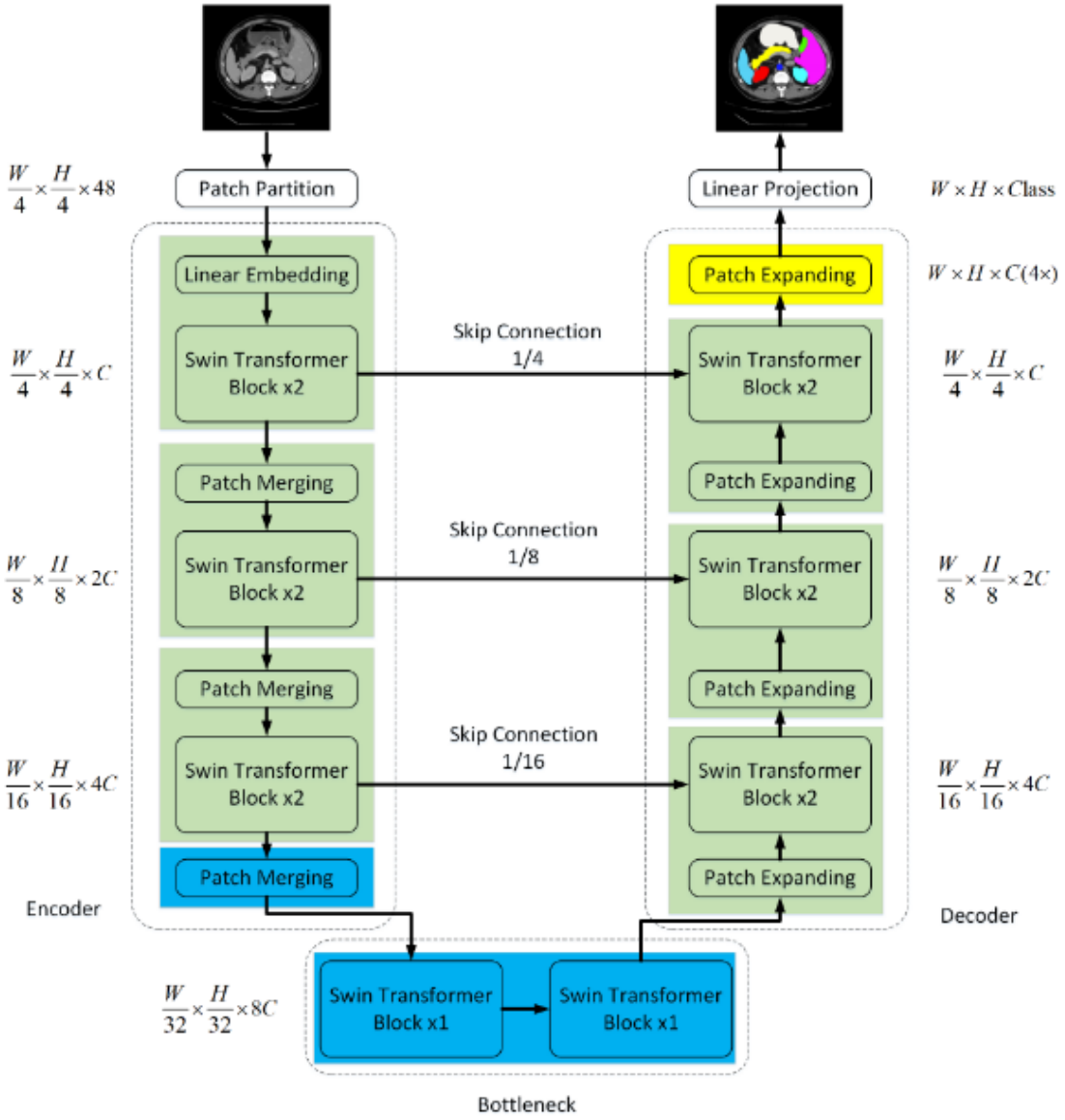
επιπλέον οδηγεί σε καλύτερα αποτελέσματα από ότι με απλό μετασχηματιστή χωρίς CNN. Στο κομμάτι του αποκωδικοποιητή, παρουσιάζεται ο Cascaded Upsampler (CUP) που αποτελείται από πολλαπλά βήματα upsampling και αποκωδικοποιεί τα χαρακτηριστικά. Το κάθε μπλοκ αποτελείται από μια 2 x upsampling διαδικασία, ένα στρώμα συνέλιξης 3x3 και ένα στρώμα ReLU. Ο CUP μαζί με τον υβριδικό Transformer-CNN κωδικοποιητή σχηματίζουν μια u-αρχιτεκτονική.

5.1.9 SwinUNet

Το SwinUNet [16] αποτελεί άλλο ένα μοντέλο u-αρχιτεκτονικής που βασίζεται στην χρήση transformers και επιχειρεί να αντιμετωπίσει το πρόβλημα των μεγάλης εμβέλειας εξαρτήσεων που υπάρχει στα συνελικτικά δίκτυα. Οι μετασχηματιστές αυτού το μοντέλου είναι ιεραρχικοί Swin Transformers [68] με shifted windows ως κωδικοποιητές οι οποίοι εξάγουν τα χαρακτηριστικά, και συμμετρικά ως αποκωδικοποιητές με patch expanding στρώμα για να εκτελέσουν upsampling και να επαναφέρουν την χωρική ανάλυση των χαρτών χαρακτηριστικών.

Η αρχιτεκτονική του SwinUNet φαίνεται στο σχήμα 5.11. Αποτελείται από τον κωδικοποιητή, το bottleneck, τον αποκωδικοποιητή και skip-connections. Το βασικό μπλοκ του μοντέλου είναι το Swin Transformer block. Οι εικόνες εισόδου χωρίζονται σε μη επικαλυπτόμενα patches, καθένα από τα οποία αντιμετωπίζεται σαν token και περνάει από τα Swin Transformer blocks και patch merging στρώματα του μετασχηματιστή-κωδικοποιητή που παράγει τις βαθιές αναπαραστάσεις χαρακτηριστικών. Τα patch merging στρώματα είναι υπεύθυνα για το downsampling και την αύξηση των διαστάσεων.

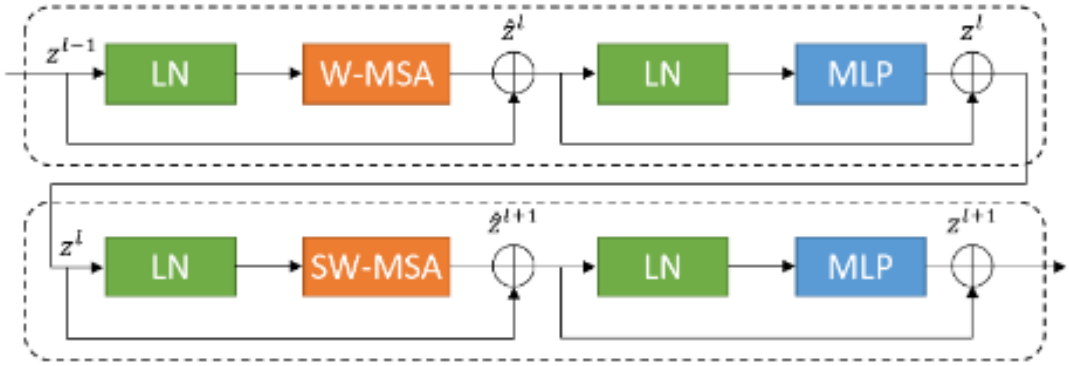
Όσον αφορά το κομμάτι του αποκωδικοποιητή, τα εξαγόμενα χαρακτηριστικά γίνονται upsample από τον αποκωδικοποιητή με patch expanding στρώμα, το οποίο επιτυγχάνει upsampling και αύξηση της διάστασης των χαρακτηριστικών χωρίς συνελίξεις και interpola-



Σχήμα 5.11: *Η αρχιτεκτονική του SwinUNet [16]*

tion, και συνδυάζονται με τα πολλαπλής κλίμακας χαρακτηριστικά του κωδικοποιητή μέσω skip-connections. Ένα τελικό στρώμα linear projection μετατρέπει τα upsampled χαρακτηριστικά στον τελικό χάρτη τμηματοποίησης.

Τα Swin Transformer blocks, που απεικονίζονται στο σχήμα 5.12, είναι βασισμένα σε shifted windows, και αποτελούνται από ένα Layer Norm στρώμα, ένα multi-head self-attention module, υπολειπόμμενες συνδέσεις (residual connections) και ένα Multi Layer Perceptron (MLP) 2 στρωμάτων με μη γραμμικότητα GELU. Το window-based multi-head self attention module (W-MSA) και το shifted multi-head self attention module (SW-MSA) εφαρμόζονται στα δύο διαδοχικά μπλοκ. Στον κωδικοποιητή τα tokenized inputs περνούν από τα 2 διαδοχικά Swin Transformer blocks για να γίνει μάθηση των αναπαραστάσεων. Το patch merging στρώμα μειώνει τον αριθμό των tokens (2 x downsampling) και αυξανει την διάσταση των χαρακτηριστικών στο διπλάσιο της αρχικής. Η διαδικασία αυτή επανα-



Σχήμα 5.12: *Η αρχιτεκτονική του Swin transformer block [16]*

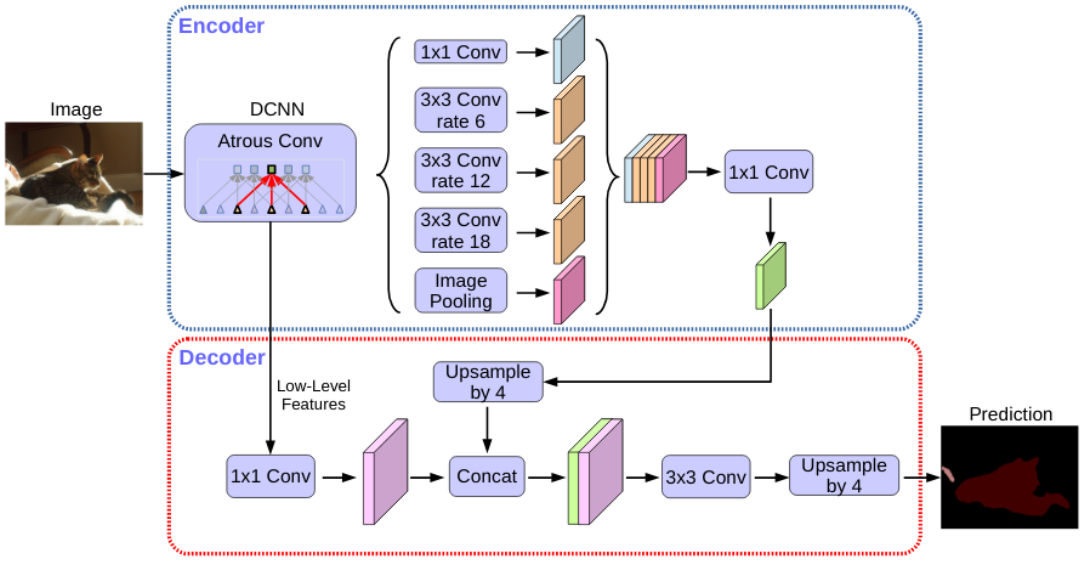
λαμβάνεται 3 φορές. Το bottleneck αποτελείται από 2 διαδοχικά Swin Transformer blocks. Ο συμμετρικός αποκωδικοποιητής είναι και αυτός βασισμένος στο Swin Transformer block και συνδυάζεται με το patch expanding στρώμα που κάνει upsample τα εξαγόμενα βαθιά χαρακτηριστικά και ανασχηματίζει τα feature maps γειτονικών διαστάσεων σε ένα feature map υψηλότερης ανάλυσης ενώ μειώνει τη διάσταση των χαρακτηριστικών στο μισό. Ομοίως με το UNet τα skip-connections συνδέουν τα χαρακτηριστικά πολλαπλών κλιμάκων του κωδικοποιητή με τα upsampled χαρακτηριστικά. Τα ρηχά και τα βαθιά χαρακτηριστικά συνενώνονται για να μειωθεί η απώλεια χωρικών πληροφοριών που προκαλείται από το downsampling.

5.1.10 DeepLabv3+

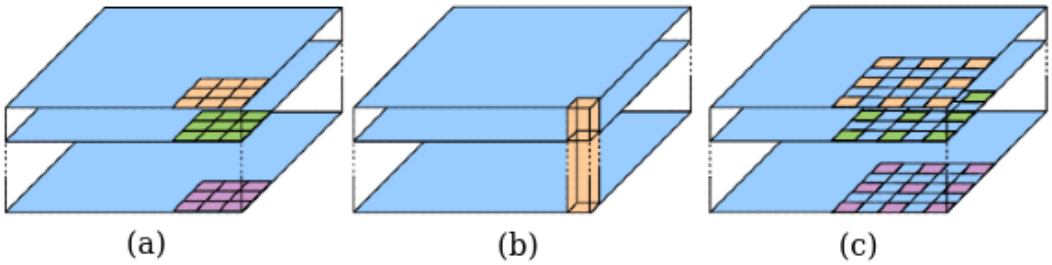
Οι αρχιτεκτονικές spatial pyramid pooling και encoder-decoder χρησιμοποιούνται ευρέως στα βαθιά νευρωνικά δίκτυα στο έργο της semantic τμηματοποίησης, και εμφανίζουν αρκετά πλεονεκτήματα. Συγκεκριμένα, οι πρώτες έχουν την δυνατότητα να κωδικοποιούν contextual πληροφορίες πολλαπλών κλιμάκων μέσω φίλτρων και διεργασιών pooling πάνω στα χαρακτηριστικά, ενώ οι δεύτερες μπορούν να οδηγήσουν σε πιο ακριβή περιγράμματα αντικειμένων με την σταδιακή επαναφορά των χωρικών πληροφοριών. Το μοντέλο DeepLabv3+ [17] συνδυάζει τα δυνατά χαρακτηριστικά και των δύο αρχιτεκτονικών. Το DeepLabv3+ αποτελεί επέκταση του DeepLabv3 [46], και διαθέτει ένα επιπλέον απλό αλλά αποδοτικό αποκωδικοποιητή που βελτιώνει τα αποτελέσματα της τμηματοποίησης όσον αφορά τα περιγράμματα των αντικειμένων. Στην έξοδο του DeepLabv3 είναι κωδικοποιημένες πλούσιες semantic πληροφορίες, και οι τραχείς συνελίξεις (atrous convolutions) επιτρέπουν τον έλεγχο της πυκνότητας των χαρακτηριστικών ανάλογα με τους διαθέσιμους υπολογιστικούς πόρους.

Οι atrous συνελίξεις [69, 70, 71, 72] επιτρέπουν τον έλεγχο της ανάλυσης των χαρακτηριστικών και ρυθμίζουν το field-of-view του φίλτρου για να καταγράψουν πολλαπλής κλίμακας πληροφορίες. Στην περίπτωση δισδιάστατων σημάτων, για κάθε θέση i στο feature map εξόδου y με φίλτρο συνέλιξης w, η atrous συνέλιξη εφαρμόζεται στο feature map εισόδου x ως εξής:

$$y[i] = \sum_k x[i + r \cdot k]w[k] \tag{5.1}$$



Σχήμα 5.13: *Η αρχιτεκτονική του DeepLabv3+[17]*



Σχήμα 5.14: *(a) Depthwise συνέλιξη (β) Pointwise συνέλιξη (γ) Atrous Depthwise συνέλιξη [17]*

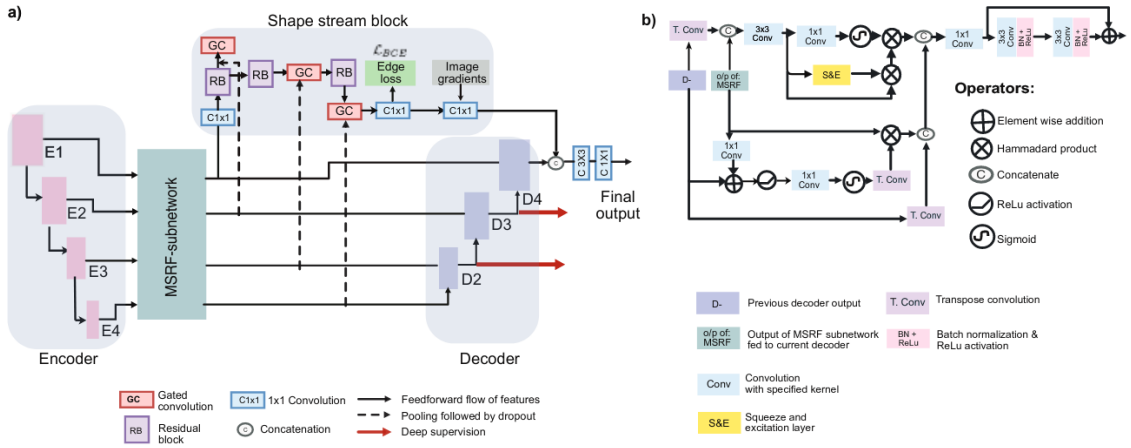
όπου ο συντελεστής atrous r ορίζει το βήμα δειγματοληψίας του σήματος εισόδου. Οι depthwise separable συνελίξεις [73, 74, 75], αποτελούν κλασσικές συνελίξεις που μετατρέπονται σε depthwise συνελίξεις ακολουθούμενες από pointwise συνελίξη, και μειώνουν σημαντικά την υπολογιστική πολυπλοκότητα. Συγκεκριμένα οι depthwise συνελίξεις εκτελούν χωρικές συνελίξεις ανεξάρτητα για κάθε κανάλι εισόδου και οι pointwise συνελίξεις συνδυάζουν την έξοδο των depthwise. Οι atrous separable συνελίξεις που αξιοποιούνται σε αυτό το μοντέλο παρουσιάζονται στο σχήμα 5.14 και προκύπτουν από τις depthwise συνελίξεις με rate=2. Αποδεικνύεται ότι μειώνουν σημαντικά την υπολογιστική πολυπλοκότητα χωρίς να μειώνουν την ποιότητα.

Η αρχιτεκτονική του DeepLabv3+ παρουσιάζεται στο σχήμα 5.13 . Ως κωδικοποιητής λειτουργεί το μοντέλο DeepLabv3. Αξιοποιεί atrous συνελίξεις για να εξάγει χαρακτηριστικά σε ορισμένη ανάλυση, τα οποία έχουν υπολογιστεί από βαθιά συνελικτικά νευρωνικά δίκτυα. Για την semantic τμηματοποίηση το output stride (που αποτελεί τον λόγο τις χωρικής ανάλυσης της εισόδου προς της εξόδου), μπορεί να έχει τιμή 16 ή και 8 για πιο πυκνά χαρακτηριστικά, αφαιρόντας το striding στο τελευταίο ή και το προτελευταίο μπλοκ και εφαρμόζοντας atrous συνέλιξη. Ως έξοδος του κωδικοποιητή χρησιμοποιείται ο τελευτα-

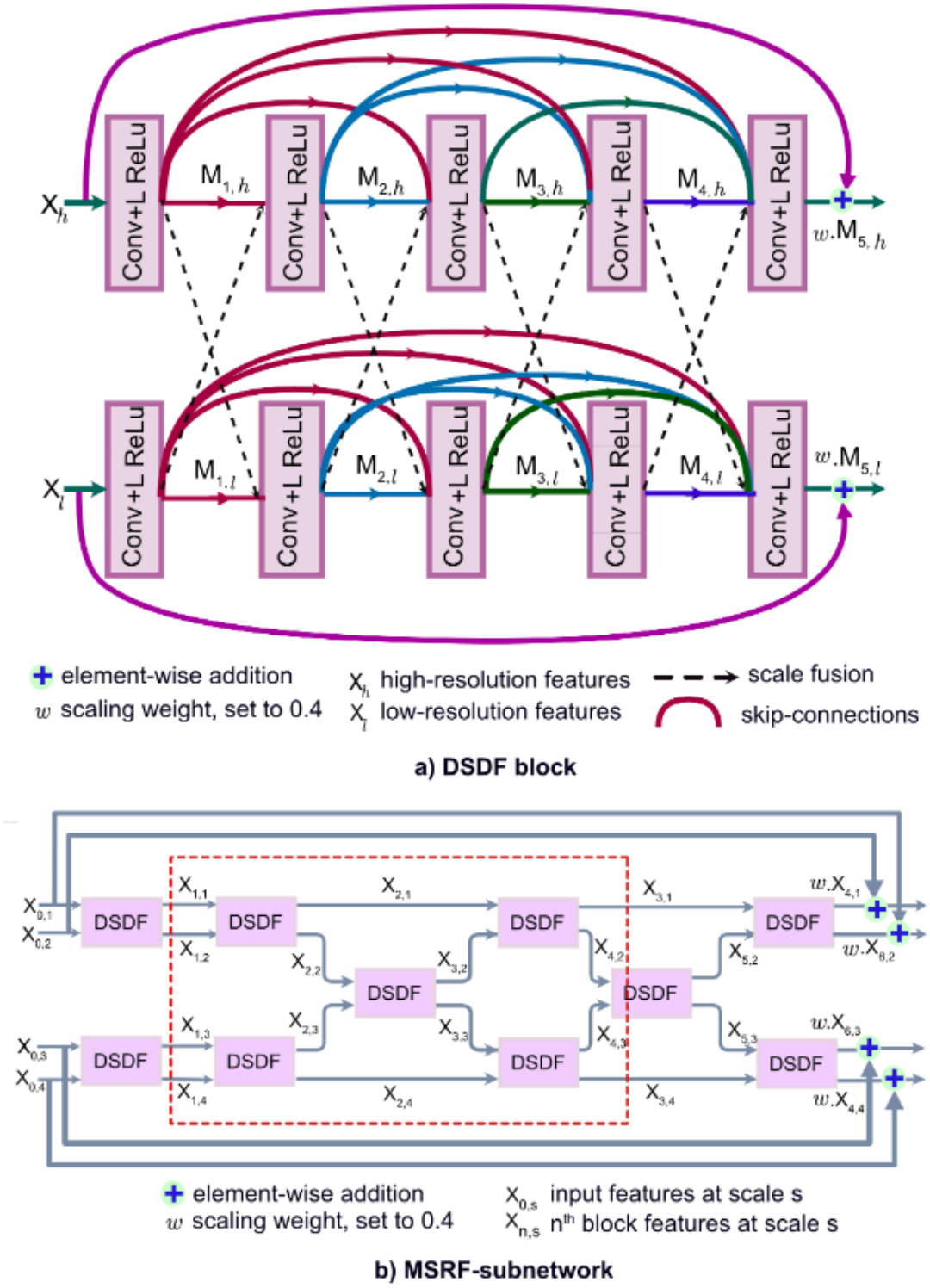
ίος χάρτης χαρακτηριστικών πριν τα logits, που αποτελείται από 256 κανάλια και πλούσια semantic πληροφορία. Στην μεριά του αποκωδικοποιητή, αρχικά, τα χαρακτηριστικά από τον κωδικοποιητή γίνονται bilinearly upsampled με παράγοντα 4 και συνενώνονται με τα αντίστοιχα low-level χαρακτηριστικά από το backbone του δικτύου που έχουν ίδια χωρική ανάλυση. Εφαρμόζεται μια συνέλιξη 1x1 στα χαρακτηριστικά low-level για να μειωθεί ο αριθμός των καναλιών ο οποίος είναι αρκετά μεγάλος (π.χ. 256 ή 512) και μπορεί να υπερκαλύψουν τα χαρακτηριστικά του κωδικοποιητή και να δυσκολέψουν την εκπαίδευση. Μετά την συνένωση (concatenation) εφαρμόζονται μερικές συνελίξεις 3x3 για να αυξηθεί η ποιότητα των χαρακτηριστικών και έπειτα άλλο ένα bilinear upsampling με παράγοντα 4. Η ποιότητα των αποτελεσμάτων αποδεικνύεται πως είναι πολύ καλύτερη όταν χρησιμοποιείται output stride = 8 αντί για 16, θυσιάζοντας υπολογιστική πολυπλοκότητα. Στην παρούσα εργασία, το μοντέλο DeepLabv3+ που χρησιμοποιήθηκε διαθέτει δίκτυο ResNet50 [64] pretrained στο ImageNet [76] ως backbone.

5.1.11 MSRF-Net

Οι μέθοδοι που βασίζονται στα συνελικτικά νευρωνικά δίκτυα ενώ έχουν βελτιώσει την απόδοση στο έργο της τμηματοποίησης ιατρικής εικόνας, εμφανίζουν αδυναμία στην τμηματοποίηση αντικειμένων μεταβλητού μεγέθους και επίσης εκπαιδεύονται σε μικρά και biased σύνολα δεδομένων. Υπάρχουν μέθοδοι που υλοποιούν multi-scale fusion όμως συνήθως χρησιμοποιούν πολύπλοκα μοντέλα που είναι σχεδιασμένα για το γενικό έργο της semantic τμηματοποίησης. Το Multi-Scale Residual Fusion Network (MSRF-Net) [18] είναι μια νέα αρχιτεκτονική σχεδιασμένη συγκεκριμένα για τμηματοποίηση ιατρικής εικόνας, η οποία έχει τη δυνατότητα να ανταλλάσει χαρακτηριστικά πολλαπλών κλιμάκων και receptive πεδίων χρησιμοποιώντας το Dual-Scale Dense Fusion (DSDF) μπλοκ. Το DSDF μπλοκ μπορεί να ανταλλάσει πληροφορίες εύρωστα ανάμεσα σε διαφορετικές κλίμακες ανάλυσης, και το MSRF-Net το χρησιμοποιεί στο πλαίσιο ενός υποδικτύου για να επιτύχει multi-scale fusion. Η διαδικασία αυτή επιτρέπει την διατήρηση της ανάλυσης και της χωρικής ακρίβειας, βελτιώνει την ροή των πληροφοριών και το propagation τόσο των low όσο και των high level χαρακτηριστικών και πετυχαίνει ακριβή αποτελέσματα τμηματοποίησης.



Σχήμα 5.15: *Η αρχιτεκτονική του MSRF-Net [18] (β) Η αρχιτεκτονική του αποκωδικοποιητή*



Σχήμα 5.16: *(α) Η αρχιτεκτονική του DSDF μπλοκ (β) Η αρχιτεκτονική του υποδικτύου MSRF [18]*

Το DSDF μπλοκ, που απεικονίζεται στο σχήμα 5.16(α) παίρνει είσοδο από δύο διαφορετικές κλίμακες και μέσω ενός residual dense μπλοκ που ανταλάσσει πληροφορίες με άλλες κλίμακες μετά από κάθε στρώμα συνέλιξης. Το επαναλαμβανόμενο multi-scale fusion βοηθά

στην ενίσχυση των υψηλής ανάλυσης αναπαραστάσεων χαρακτηριστικών με τις πληροφορίες από αναπαραστάσεις χαμηλής ανάλυσης. Επιπλέον, στρώματα residual δικτύων επιτρέπουν σε περιττά DSDF μπλοκ να σταματήσουν να έχουν επίδραση, και μόνο τα σχετικά εξαγόμενα χαρακτηριστικά να συμβάλουν στο τελικό χάρτη τμηματοποίησης. Εκτός από τα παραπάνω κομμάτια υπάρχει και ένα συμπληρωματικό gated shape stream που αξιοποιεί τον συνδυασμό χαρακτηριστικών υψηλού και χαμηλού επιπέδου για να υπολογίσει περιγράμματα σχημάτων με ακρίβεια. Ένας κωδικοποιητής τροφοδοτεί το MSRF υποδίκτυο (σχήμα 5.16(β)), που αποτελείται από πολλαπλά DSDF μπλοκ, με αναπαραστάσεις χαρακτηριστικών. Έπειτα, στρώματα του αποκωδικοποιητή με skip-connections από το υποδίκτυο και τριπλό μηχανισμό attention επεξεργάζονται τα fused feature maps μαζί με το shape stream.

Η αρχιτεκτονική του δικτύου MSRF φαίνεται στο σχήμα 5.15 και αποτελείται από μπλοκ κωδικοποίησης, το MSRF υποδίκτυο, ένα shape stream μπλοκ και μπλοκ αποκωδικοποίησης. Τα μπλοκ κωδικοποίησης αποτελούνται από 2 διαδοχικές συνελίξεις ακολουθούμενες από μια μονάδα squeeze-and-excitation η οποία αυξάνει την αναπαραστατική δύναμη του δικτύου υπολογίζοντας τις αλληλοεξαρτήσεις μεταξύ καναλιών. Στο βήμα του squeezing τα χαρακτηριστικά γίνονται aggregate στην χωρική διάσταση των καναλιών μέσω global average pooling, ενώ στο βήμα του excitation παράγεται μια συλλογή από βάρη ανά κανάλι για να εκφράσουν τις εξαρτήσεις μεταξύ καναλιών. Σε κάθε στάδιο του κωδικοποιητή εκτελείται max pooling με βήμα 2 για την μείωση της ανάλυσης, και dropout για το regularization του μοντέλου. Το DSDF μπλοκ, το οποίο βοηθάει στην ανταλλαγή πληροφοριών μεταξύ κλιμάκων, έχει 2 παράλληλες ροές για 2 διαφορετικές κλίμακες. Αν ονομάσουμε CLR(·) την διεργασία μιας συνέλιξης 3x3 ακολουθούμενη από μια ενεργοποίηση LeakyReLU τότε κάθε ροή έχει ένα densely connected residual μπλοκ με 5 διεργασίες CLR(·). Το feature map $M_{d,h}$ της εξόδου υπολογίζεται από την υψηλής ανάλυσης είσοδο X_h ως εξής:

$$M_{d,h} = CLR(M_{d-1,h} \oplus M_{d-1,l} \oplus M_{d-2,h} \oplus \dots M_{0,h}), 1 \leq d \leq 5 \tag{5.2}$$

στην ροή υψηλής ανάλυσης, και ομοίως για την ροή χαμηλής ανάλυσης ως εξής:

$$M_{d,l} = CLR(M_{d-1,l} \oplus M_{d-1,h} \oplus M_{d-2,l} \oplus \dots M_{0,l}), 1 \leq d \leq 5 \tag{5.3}$$

Ο τελεστής $\oplus$ συμβολίζει το concatenation και τα σύμβολα h,l αντιπροσωπεύουν υψηλή και χαμηλή ανάλυση αντίστοιχα. Η έξοδος κάθε CLR(·) έχει k κανάλια που καθορίζουν τον παράγοντα growth ο οποίος ρυθμίζει τον αριθμό νέων χαρακτηριστικών που μπορεί να εξάγει και να προωθήσει το στρώμα στο υπόλοιπο δίκτυο. Προκειμένου να μην αυξηθεί η πολυπλοκότητα του μοντέλου επειδή ο παράγοντας growth αλλάζει για κάθε κλίμακα, μόνο 2 κλίμακες χρησιμοποιούνται κάθε στιγμή στα DSDF μπλοκ. Εφαρμόζεται επιπλέον residual μάθηση και residual scaling για να αποφευχθεί η αστάθεια [77, 78]. Άρα η τελική έξοδος του DSDF μπλοκ εκφράζεται ως εξής:

$$X_r = w \times M_{5,r} + X_r, r \in [h, l], 0 \leq w \leq 1 \tag{5.4}$$

Το υποδίκτυο MSRF αποτελείται από πολλά DSDF μπλοκ και επιτυγχάνει global multi-scale context μέσω του μηχανισμού dual-scale fusion. Αρχικά οριοθετούνται τα ζεύγη αναλύσε-

ων/κλιμάκων και τροφοδοτούνται στα αντίστοιχα DSDF μπλοκ. Ξεκινώντας από το πρώτο στρώμα που αποτελείται από 4 κλίμακες ανάλυσης, η συνάρτηση DSDF(·) εκτελεί feature fusion μεταξύ κλιμάκων στο DSDF μπλοκ. Ήδη μετά το τέταρτο στρώμα έχει επιτευχθεί ανταλλαγή χαρακτηριστικών σε όλες τις κλίμακες. Η μέθοδος αυτή μπορεί να υποστηρίξει και παραπάνω από 4 κλίμακες. Η έξοδος του τελευταίου στρώματος του υποδικτύου γίνεται scaled κατά παράγοντα w και προστίθεται στην είσοδο.

Το μοντέλο διαθέτει ένα gated shape stream [79] το οποίο κάνει πρόβλεψη σχήματος αξιοποιώντας high-level αναπαραστάσεις που εξάγονται από τα DSDF μπλοκ. Τα shape stream feature maps ορίζονται ως S_l όπου l ο αριθμός των στρωμάτων, και ως X ορίζεται η έξοδος από το υποδίκτυο MSRF, η οποία περνάει από bilinear interpolation ώστε να ταιριάζουν οι διαστάσεις της με του S_l. Ο χάρτης attention στην gated συνέλιξη υπολογίζεται ως εξής:

$$a_l = \sigma(Conv_{1\times1}(S_l X)) \tag{5.5}$$

όπου $\sigma(\cdot)$ είναι η σιγμοειδής ενεργοποίηση. Το S_{l+1} υπολογίζεται ως:

$$S_{l+1} = RB(S_l \times a_l) \tag{5.6}$$

όπου RB είναι ένα residual block με 2 CLR(·) διεργασίες ακολουθούμενες από ένα skip-connection. Η έξοδος του shape stream ενώνεται με τα gradients των εικόνων της εισόδου και αναμειγνύεται με το αρχικό stream τμηματοποίησης πριν την τελευταία CLR(·) ώστε να αυξηθεί η χωρική ακρίβεια.

Το μπλοκ αποκωδικοποίησης έχει skip-connections από το υποδίκτυο MSRF και την έξοδο του προηγούμενου αποκωδικοποιητή. Σε αυτό το κομμάτι υπάρχουν 2 μηχανισμοί attention, ο πρώτος εφαρμόζει channel και spatial attention και ο δεύτερος χρησιμοποιεί ένα μηχανισμό gating. Ένα μπλοκ squeeze-and-excitation εκτελεί τον υπολογισμό συντελεστών κλίμακας ανά κανάλι $X_{a_{se}}$ και επιπλέον υπολογίζεται το spatial attention αφού τα κανάλια μειωθούν μέσω μιας συνέλιξης 1x1. Η σιγμοειδής ενεργοποίηση τοποθετεί τις τιμές ανάμεσα σε 0 και 1 και έτσι σχηματίζει το activation map X_{a_s}. Έτσι η έξοδος του spatial και channel attention εκφράζεται ως εξής:

$$D_{sc} = (X_{as} + 1) \odot X_{a_{se}} \tag{5.7}$$

όπου ο συντελεστής $\odot$ αποτελεί το γινόμενο Hadamard. Όσον αφορά τον μηχανισμό gated attention [11], οι συντελεστές attention υπολογίζονται ως εξής:

$$D_{AG} = \Omega(\sigma(\Psi(\partial(X) + \phi(D^{-})))) \tag{5.8}$$

όπου X τα χαρακτηριστικά από το MSRF-Net, D^{-} η έξοδος του προηγούμενου μπλοκ αποκωδικοποίησης, $\partial(\cdot)$ ο τελεστής της συνέλιξης με βήμα 2 και μέγεθος πυρήνα 1 και G έξοδοι καναλιών, $\phi(\cdot)$ ο τελεστής συνέλιξης με βήμα 1 και μέγεθος πυρήνα 1x1, $\Psi(\cdot)$ ο τελεστής συνέλιξης με πυρήνα 1x1 που εφαρμόζεται στα συνδυασμένα χαρακτηριστικά $\partial(\cdot)$ και $\phi(\cdot)$, $\sigma(\cdot)$ ο τελεστής της σιγμοειδούς ενεργοποίησης και $\Omega(\cdot)$ ο τελεστής της συνέλιξης transpose. Οι συντελεστές D_{AG} περιέχουν contextual πληροφορίες και αναγνωρίζουν τις περιοχές στόχους

και τις δομές της εικόνας. Μέσω της διεργασίας $\tilde{D}_{AG} = D_{AG} \oplus X$ παραλείπονται μη σχετικά χαρακτηριστικά. Οι συντελεστές $\tilde{D}_{AG}$ γίνονται update ως εξής:

$$\tilde{D}_{AG} = \tilde{D}_{AG} \oplus \Omega(D^{-}) \tag{5.9}$$

Τελικά η έξοδος του τριπλού attention decoder block είναι

$$D_a = D_{sc} \oplus \tilde{D}_{AG} \tag{5.10}$$

και ακολουθείται από 2 CLR(·) διαδικασίες.

5.2 Μετρικές Αξιολόγησης

5.2.1 Συναρτήσεις απώλειας

Μια από τις μετρικές σφάλματος (loss) για το έργο της τμηματοποίησης εικόνας που έχουν επικρατήσει είναι η Binary Cross Entropy [80] που περιγράφεται από τον τύπο:

$$\mathscr{L}_{BCE} = (1-y)log(1-\hat{y}) + ylog\hat{y} \tag{5.11}$$

όπου y η αληθινή τιμή (ground truth) και $\hat{y}$ η υπολογισμένη τιμή (predicted value). Αποτελεί ειδική περίπτωση της μετρικής Cross Entropy η οποία υπολογίζει την απόδοση ενός μοντέλου κατηγοριοποίησης που παράγει ως έξοδο τιμές πιθανότητες 0 ως 1. Αντιμετωπίζει ισάξια και τους δύο τύπους λαθών (false positive, false negative) και αυξάνεται όσο η πιθανότητα που υπολογίζεται απομακρύνεται από την πραγματική τιμή. Εκτός από την Binary Cross Entropy, υπάρχει και η Dice [81] που είναι ένα μέτρο του overlap μεταξύ πραγματικής εικόνας και υπολογισμένης από το μοντέλο εικόνας και περιγράφεται από τον τύπο:

$$\mathscr{L}_{DCS} = 1 - \frac{2y\hat{y}+1}{y+\hat{y}+1} = 1 - DC \tag{5.12}$$

όπου y και $\hat{y}$ ορίζονται ομοίως με πριν και DC είναι ο Dice Coefficient που ορίζεται ως εξής:

$$DC = \frac{2|A \cap B|}{|A| + |B|} \tag{5.13}$$

όπου $|A \cap B|$ ο αριθμός των κοινών στοιχείων των συνόλων A, B και $|A|, |B|$ ο αριθμός των στοιχείων στα σύνολα A, B αντίστοιχα. Ο αριθμητής και των δύο σχέσεων ουσιαστικά αναφέρεται στις κοινές ενεργοποιήσεις μεταξύ της πραγματικής εικόνας και της υπολογισμένης από το μοντέλο εικόνας, και ο παρονομαστής αναφέρεται στον αριθμό των ενεργοποιήσεων στις δύο εικόνες ξεχωριστά. Σε όλα τα μοντέλα που παρουσιάστηκαν παραπάνω χρησιμοποιήθηκαν και οι δύο μετρικές σφάλματος σε ενδεικτικά πειράματα και για το καθένα επιλέχθηκε η μετρική σφάλματος που οδηγούσε σε καλύτερα αποτελέσματα. Στο MSRF-Net μοντέλο χρησιμοποιήθηκε ένας συνδυασμός μετρικών που ορίζεται ως εξής [18]:

$$\mathscr{L}_{MSRF} = \mathscr{L}_{comb} + \mathscr{L}_{comb}^{DS^0} + \mathscr{L}_{comb}^{DS^1} + \mathscr{L}_{BCE}^{SS} \tag{5.14}$$

όπου $\mathscr{L}_{comb} = \mathscr{L}_{DCS} + \mathscr{L}_{BCE}$, $\mathscr{L}_{comb}^{DS^0}$ και $\mathscr{L}_{comb}^{DS^1}$ οι μετρικές για τις δύο εξόδους του deep supervision, και $\mathscr{L}_{BCE}^{SS}$ η μετρική σφάλματος BCE από το shape stream του μοντέλου MSRF-Net.

5.2.2 Μετρικές ακριβείας

Εκτός από τις συναρτήσεις απώλειας που διαδραματίζουν σημαντικό ρόλο κατά την εκπαίδευση των μοντέλων, υπάρχουν και οι μετρικές ακριβείας που χρησιμεύουν στην αξιολόγηση των μοντέλων. Σε αυτή την εργασία χρησιμοποιήθηκαν οι μετρικές Dice Coefficient, (mean) Intersection Over Union, Precision και Recall.

Η μετρική (mean) Dice Coefficient (DSC) [81] έχει αναλυθεί στην προηγούμενη παράγραφο και ο τύπος που την περιγράφει είναι ο (5.13). Η τελική τιμή της προκύπτει ως η μέση τιμή από τα αποτελέσματα των υπολογισμών της Dice Coefficient για όλα τα ζεύγη πραγματικών και υπολογισμένων εικόνων.

Η μετρική (mean) Intersection Over Union (mIoU) [81] εκτιμά την ομοιότητα μεταξύ των εικόνων που υπολογίστηκαν από το μοντέλο και των πραγματικών εικόνων, και όπως φαίνεται από το ονομά της υπολογίζει το ποσοστό των κοινών ενεργοποιήσεων μεταξύ των δύο τύπων εικόνων (Intersection) προς την ένωση των ενεργοποιήσεων τους (Union). Η μέγιστη τιμή που λαμβάνει είναι 1 και όσο μικρότερη είναι η τιμή της τόσο χειρότερης ποιότητας είναι τα αποτελέσματα του μοντέλου. Υπολογίζεται ως εξής:

$$IoU = \frac{|A \cap B|}{|A \cup B|} \tag{5.15}$$

όπου όπου $|A \cap B|$ ο αριθμός των κοινών στοιχείων των συνόλων A, B που αποτελούν τις ενεργοποιήσεις των πραγματικών και των υπολογισμένων εικόνων αντίστοιχα, και $|A \cup B|$ ο αριθμός των στοιχείων στην ένωση των συνόλων A, B. Η τελική τιμή της μετρικής mIoU προκύπτει ως η μέση τιμή από τα αποτελέσματα των υπολογισμών της IoU για όλα τα ζεύγη πραγματικών και υπολογισμένων εικόνων.

Η μετρική Precision [82] υπολογίζει το ποσοστό των σωστών θετικών προβλέψεων προς όλες τις θετικές προβλέψεις, δηλαδή:

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives} \tag{5.16}$$

όπου True Positives είναι οι θετικές προβλέψεις που συμπίπτουν στις πραγματικές και τις υπολογισμένες εικόνες και False Positives είναι οι θετικές προβλέψεις στις υπολογισμένες εικόνες που όμως δεν υπάρχουν στις πραγματικές εικόνες. Λαμβάνει τιμές από 0 ως 1 και όσο μικρότερη είναι η τιμή της τόσο χειρότερη κρίνεται η αξιολόγηση του μοντέλου.

Τέλος, η μετρική Recall [82] υπολογίζει το ποσοστό των σωστών θετικών προβλέψεων προς τον συνολικό αριθμό σωστών θετικών προβλέψεων και λανθασμένων αρνητικών προβλέψεων, δηλαδή:

$$Recall = \frac{TruePositives}{TruePositives + FalseNegatives} \tag{5.17}$$

όπου True Positives ορίζονται όπως παραπάνω και False Negatives είναι οι αρνητικές προβλέψεις στις υπολογισμένες εικόνες που είναι θετικές στις πραγματικές εικόνες. Μοιάζει αρ-

κετά με την μετρική Precision όμως δίνει περισσότερη βάση στις χαμένες θετικές προβλέψεις. Η μετρική που χρησιμοποιήθηκε για την αξιολόγηση των μοντέλων κατά την διάρκεια της εκπαίδευσης (απόφαση αν το μοντέλο στην συγκεκριμένη εποχή θα σωθεί ή όχι) είναι η Dice Coefficient.

5.3 Λεπτομέρειες Υλοποίησης

5.3.1 Παράμετροι μοντέλων

Στο πλαίσιο των πειραμάτων της εργασίας, το μέγεθος των εικόνων και των μασκών από τα 4 σύνολα δεδομένων (CVC-ClinicDB, Kvasir-Seg, 2018 Data Science Bowl, SegPC) ορίστηκε σε 256x256 στα περισσότερα μοντέλα. Επιπλέον για τα σύνολα CVC-ClinicDB και SegPC χρησιμοποιήθηκε batch size = 8, στο Kvasir-Seg και στο 2018 Data Science Bowl χρησιμοποιήθηκε batch size=16, στα περισσότερα μοντέλα. Μερικά μοντέλα (π.χ. ResUNet-a) εμφάνισαν μεγάλες απαιτήσεις μνήμης που ξεπερνούσαν τους διαθέσιμους πόρους με τις συγκεκριμένες παραμέτρους, οπότε επιλέχθηκαν διαστάσεις 128x128 και batch size=4 προκειμένου να ολοκληρωθεί σωστά η εκπαίδευση. Ως optimizer χρησιμοποιήθηκε ο Adam [83] σε όλα τα μοντέλα, με learning rate = 0.0001. Ο αριθμός των εποχών τέθηκε στις 200, ωστόσο όλα τα μοντέλα φτάνουν στην μέγιστη απόδοση τους αρκετά πριν την τελευταία εποχή. Χρησιμοποιήθηκε κώδικας ανοιχτής πρόσβασης για όλα τα μοντέλα που παρουσιάζονται. Οι παράμετροι επιλέχθηκαν με βάση την βιβλιογραφία, και την διαίσθηση όπου η βιβλιογραφία δεν παρείχε την απάντηση.

Για το μοντέλο MSRF-Net ο scaling factor (w) για το μπλοκ DSDF και το MSRF υποδίκτυο έχει τιμή 0.4, και ο growth factor (k) τα τα ζεύγη resolution στο DSDF έχει τιμές 16, 32 και 64.

Το μοντέλο DeepLabv3+ έχει για backbone ένα μοντέλο ResNet50 το οποίο είναι pre-trained στο σύνολο δεδομένων ImageNet και χρησιμοποιούνται τα low level features από το επίπεδο conv4_block6_2_relu του backbone.

Στο μοντέλο UNet οι παράμετροι είναι οι εξής: 5 down και upsampling επίπεδα, 2 συνελικτικά στρώματα για κάθε downsampling επίπεδο, 1 συνελικτικό στρώμα για κάθε upsampling επίπεδο, ReLU activation, Sigmoid output activation, Max pooling, upsampling με reflective padding.

Στο μοντέλο VNet οι παράμετροι είναι οι εξής: 5 down και upsampling επίπεδα, ο αριθμός των συνελικτικών επιπέδων του residual μονοπατιού αυξάνεται από 1 σε 3 με επίπεδα downsampling (και συμμετρικά μειώνεται με επίπεδα upsampling), PReLU activation, Sigmoid output activation.

Στο μοντέλο Attention UNet οι παράμετροι είναι οι εξής: 4 down και upsampling επίπεδα, 2 συνελικτικά στρώματα ανά κάθε downsampling και upsampling επίπεδο, ReLU activation, additive attention, ReLU attention activation, Sigmoid output activation, Max pooling και upsampling με reflective padding.

Στο μοντέλο R2U-Net οι παράμετροι είναι οι εξής: 4 down και upsampling επίπεδα, 2 recurrent συνελικτικά στρώματα με 2 iterations ανά down και upsampling επίπεδο, ReLU activation, Sigmoid output activation, Max pooling και upsampling με reflective padding.

Στο μοντέλο ResUNet-a οι παράμετροι είναι οι εξής: 5 downsampling επίπεδα ακολουθούμενα από ένα στρώμα Atrous Spatial Pyramid Pooling (ASPP) με 256 φίλτρα, 5 upsampling επίπεδα ακολουθούμενα από στρώμα ASPP με 128 φίλτρα, dilation rates με τιμές 1,3,15,31, ReLU activation, Sigmoid output activation και upsampling με reflective padding.

Στο μοντέλο TransUNet οι παράμετροι είναι οι εξής: 4 down και upsampling επίπεδα, 2 συνελικτικά στρώματα ανά down και upsampling επίεπδο, 2 transformer blocks, 2 attention heads, 3072 MLP nodes ανά μετασχηματιστή, 768 διαστάσεις embedding, Gaussian Error Linear Unit (GeLU) activation για τα MLP, ReLU activation, Sigmoid output activation, Max pooling και upsampling μέσω bilinear interpolation.

Στο μοντέλο SwinUNet οι παράμετροι είναι οι εξής: 4 down και upsampling επίπεδα, 2 Swin Transformers ανά down και upsampling επίπεδο, μέγεθος patch (2,2), embedded patches σε 64 διαστάσεις, αριθμός attention heads για κάθε down και upsampling επίπεδο [4,8,8,8], μέγεθος attention window για κάθε down και upsampling επίπεδο [4,2,2,2], 512 MLP nodes ανά Swin Transformer, Shift Attention windows.

5.3.2 Υπολογιστικό Σύστημα

Τα πειράματα της παρούσας εργασίας εκπονήθηκαν εξ ολοκλήρου στο ARIS [84], που αποτελεί τον ελληνικό υπερυπολογιστή που αναπτύχθηκε και λειτουργεί από το ΕΔΥΤΕ (Εθνικό Δίκτυο Υποδομών Τεχνολογίας και Έρευνας, GRNET) στην Αθήνα. Το ARIS αποτελείται από 532 υπολογιστικά nodes χωρισμένα ως εξής:

- 426 thin nodes: κανονικά υπολογιστικά nodes χωρίς accelerator.
- 44 gpu nodes: “2 x NVIDIA Tesla k40m” accelerated nodes.
- 18 phi nodes: “2 x INTEL Xeon Phi 7120p” accelerated nodes.
- 44 fat nodes: Τα υπολογιστικά Fat nodes διαθέτουν μεγαλύτερο αριθμό πυρήνων και περισσότερη μνήμη ανά πυρήνα σε σχέση με τα thin nodes.
- 1 ml node: “8 x NVIDIA Volta V100” accelerators.

Σε αυτή την εργασία χρησιμοποιήθηκαν gpu nodes. Τα πειράματα εκτελούνται μέσω του τρεξίματος SLURM jobs το οποία έχουν περιορισμούς σε υπολογιστικούς πόρους, συγκεκριμένα το μέγιστο επιτρεπόμενο run time είναι 96 ώρες και η μέγιστη επιτρεπόμενη χρήση μνήμης είναι 56000 MB. Αυτοί οι περιορισμοί επηρέασαν την επιλογή των παραμέτρων της προηγούμενης παραγράφου.

Κεφάλαιο 6

Πειραματικά αποτελέσματα

Σε αυτό το κεφάλαιο παρουσιάζονται τα αποτελέσματα από τα πειράματα που εκτελέστηκαν με τα μοντέλα που αναλύθηκαν στο κεφάλαιο 5, στα σύνολα δεδομένων που παρουσιάστηκαν στο κεφάλαιο 4. Τα αποτελέσματα παρατίθεται και σε ποσοτική μορφή, στους πίνακες που ακολουθούν παρακάτω, όσο και σε ποιοτική μορφή με παραδείγματα εικόνων-μασκών που παράχθηκαν από τα μοντέλα που εκπαιδεύτηκαν, που συγκρίνονται με τις πραγματικές εικόνες-μάσκες. Επιπλέον εξετάζεται η επίδραση της επαύξησης δεδομένων στην βελτίωση των αποτελεσμάτων καθώς και η ικανότητα γενίκευσης των μοντέλων.

6.1 Παρουσίαση αποτελεσμάτων

Στον πίνακα 6.1 παρουσιάζονται τα αποτελέσματα των πειραμάτων στο σύνολο δεδομένων CVC-ClinicDB που όπως έχει αναλύθει προηγουμένως, περιλαμβάνει εικόνες από πολύποδες που προέρχονται από βίντεο κολονοσκόπησης. Το μοντέλο που αποδίδει καλύτερα με κριτήριο τις μετρικές Dice Coefficient (DSC) και mean Intersection over Union (mIoU) είναι το DeepLabv3+ ακολουθούμενο από τα MSRF-Net, R2U-net, TransUnet και UNet. Το μεγαλύτερο Precision πετυχαίνουν τα μοντέλα MSRF-Net και R2U-Net, ενώ το μεγαλύτερο Recall πετυχαίνει το DeepLabv3+. Η καλή απόδοση των μοντέλων αυτών επιβεβαιώνεται και από την εικόνα 6.1 και οι μάσκες που παρήχθησαν από αυτά μοιάζουν ικανοποιητικά με τις πραγματικές μάσκες. Μοντέλα όπως το SwinUNet, VNet και ResUNet φαίνεται να αποτυγχάνουν στο συγκεκριμένο σύνολο δεδομένων. Οι μάσκες που παρήχθησαν από αυτά εμφανίζουν σημαντικές ατέλειες σε σχέση με τις πραγματικές μάσκες.

Στον πίνακα 6.2 παρουσιάζονται τα αποτελέσματα των πειραμάτων στο σύνολο δεδομένων Kvasir-SEG το οποίο επίσης περιλαμβάνει εικόνες από γαστρεντερικούς πολύποδες. Τα πιο ισχυρά μοντέλα με κριτήριο τη μετρική mIoU είναι τα MSRF-Net και DeepLabv3+ ακολουθούμενα από τα UNet και TransUNet, ενώ με κριτήριο τη μετρική DSC το πιο ισχυρό μοντέλο είναι το DeepLabv3+ ακολουθούμενο από τα R2U-Net, TransUNet και UNet. Το μεγαλύτερο Precision πετυχαίνουν τα μοντέλα MSRF-Net και UNet, και το μεγαλύτερο Recall πετυχαίνει το μοντέλο DeepLabv3+. Παρατηρώντας τα οπτικά αποτελέσματα στην εικόνα 6.1, τα ισχυρά μοντέλα παράγουν αρκετά ποιότικές μάσκες, στις οποίες ο πολύποδας-στόχος εντοπίζεται, ωστόσο μπορεί να εντοπίζονται λανθασμένα ως πολύποδες και επιπλέον μικρές περιοχές. Τα μοντέλα που φαίνεται να αποτυγχάνουν είναι τα VNet, ResUNet-a και SwinUNet, το οποίο επιβεβαιώνεται και οπτικά από τις αδύναμες παραγόμενες μάσκες.

Στον πίνακα 6.3 παρατίθενται τα αποτελέσματα στο σύνολο δεδομένων 2018 Data Science Bowl το οποίο περιλαμβάνει εικόνες από πυρήνες διαφόρων τύπων κυττάρων. Μια αρχική παρατήρηση είναι ότι όλα τα μοντέλα πετυχαίνουν σημαντικά υψηλότερη απόδοση σχετικά με τα υπόλοιπα σύνολα δεδομένων. Σύμφωνα με την μετρική DSC τα ισχυρότερα μοντέλα είναι τα DeepLabv3+, R2U-Net, TransUNet, Attention UNet, SwinUNet, UNet και VNet ενώ σύμφωνα με την mIoU τα ισχυρότερα είναι τα TransUNet, UNet ακολουθούμενα από τα MSRF-Net, Attention UNet, R2U-Net, SwinUNet, VNet και DeepLabv3+. Μεγαλύτερο Precision πετυχαίνει το TransUNet και μεγαλύτερο Recall πετυχαίνουν τα MSRF-Net και R2U-Net. Το μόνο μοντέλο που εμφανίζει σημαντικά χαμηλότερη απόδοση από τα υπόλοιπα είναι το ResUNet-a. Στην εικόνα 6.1 φαίνεται η πολύ μεγάλη ομοιότητα των μασκών των μοντέλων με την πραγματική μάσκα όσον αφορά την grayscale εικόνα, ενώ όσον αφορά την έγχρωμη εικόνα οι μάσκες των μοντέλων που πετυχαίνουν τα μεγαλύτερα score στη μετρική mIoU είναι εμφανώς πιο ποιοτικές.

Τέλος στον πίνακα 6.4 παρατίθενται τα αποτελέσματα στο σύνολο δεδομένων SegPC το οποίο αποτελείται από εικόνες κυττάρων πλάσματος από ασθενείς που πάσχουν από Πολλαπλό Μυέλωμα. Οι αποδόσεις των μοντέλων για αυτό το σύνολο δεδομένων είναι χαμηλότερες σε σχέση με τα παραπάνω σύνολα δεδομένων, γεγονός που καθιστά το SegPC το πιο απαιτητικό σύνολο. Το ισχυρότερο μοντέλο με βάση την μετρική DSC είναι το DeepLabv3+ ακολουθούμενο από τα TransUNet,R2U-Net,Attention UNet και UNet, ενώ με βάση την μετρική mIoU το ισχυρότερο μοντέλο είναι το R2U-Net ακολουθούμενο από τα MSRF-Net και TransUNet. Το μεγαλύτερο Precision πετυχαίνει το μοντέλο R2U-Net και το μεγαλύτερο Recall πετυχαίνει το μοντέλο ResUNet. Τα αποτελέσματα αυτά επιβεβαιώνονται και οπτικά στην εικόνα 6.1 . Τα μοντέλα που φαίνεται να αποτυγχάνουν τόσο από τα ποσοτικά αποτελέσματα του πίνακα 6.4 καθώς και από τα οπτικά αποτελέσματα του σχήματος 6.1 είναι τα SwinUNet και ResUNet-a.

Method	**DSC**	**mIoU**	**Precision**	**Recall**
MSRF-Net	0.87	0.84	**0.95**	0.89
UNet	0.92	0.80	0.91	0.87
VNet	0.81	0.64	0.79	0.76
Att-UNet	0.87	0.69	0.92	0.73
ResUNet-a	0.86	0.71	0.85	0.81
SwinUNet	0.79	0.56	0.70	0.71
TransUNet	0.94	0.80	0.92	0.86
R2U-Net	0.94	0.83	**0.95**	0.88
DeepLabv3+	**0.99**	**0.89**	0.94	**0.94**
ResUNet	0.70	0.64	0.80	0.77
ResUNet++	0.78	0.72	0.90	0.78

Πίνακας 6.1: *Ποσοτικά αποτελέσματα αξιολόγησης των μοντέλων για το σύνολο δεδομένων CVC-ClinicDB με επάυξηση δεδομένων*

Method	DSC	mIoU	Precision	Recall
MSRF-Net	0.86	**0.73**	**0.89**	0.80
UNet	0.90	0.71	**0.89**	0.78
VNet	0.76	0.49	0.73	0.60
Att-UNet	0.86	0.62	0.82	0.72
ResUNet-a	0.78	0.49	0.73	0.60
SwinUNet	0.77	0.50	0.79	0.66
TransUNet	0.90	0.70	0.87	0.79
R2U-Net	0.88	0.69	0.85	0.79
DeepLabv3+	**0.96**	**0.73**	0.87	**0.82**
ResUNet	0.60	0.49	0.66	0.66
ResUNet++	0.65	0.55	0.74	0.69

Πίνακας 6.2: *Ποσοτικά αποτελέσματα αξιολόγησης μοντέλων για το σύνολο δεδομένων Kvasir-SEG με επάυξηση δεδομένων*

Method	DSC	mIoU	Precision	Recall
MSRF-Net	0.91	0.86	0.91	**0.95**
UNet	0.95	**0.87**	0.92	0.94
VNet	0.94	0.85	0.91	0.93
Att-UNet	0.95	0.86	0.92	0.93
ResUNet-a	0.91	0.76	0.90	0.83
SwinUNet	0.94	0.85	0.92	0.92
TransUNet	0.95	**0.87**	**0.93**	0.93
R2U-Net	0.95	0.86	0.90	**0.95**
DeepLabv3+	**0.98**	0.85	0.91	0.93
ResUNet	0.88	0.84	0.89	0.94
ResUNet++	0.89	0.83	0.89	0.93

Πίνακας 6.3: *Ποσοτικά αποτελέσματα αξιολόγησης μοντέλων για το σύνολο δεδομένων 2018 Data Science Bowl με επάυξηση δεδομένων*

Method	DSC	mIoU	Precision	Recall
MSRF-Net	0.80	0.68	0.82	0.79
UNet	0.88	0.66	0.81	0.78
VNet	0.85	0.57	0.75	0.71
Att-UNet	0.89	0.66	0.80	0.79
ResUNet-a	0.77	0.41	0.60	0.57
SwinUNet	0.80	0.45	0.63	0.62
TransUNet	0.90	0.68	0.81	0.80
R2U-Net	0.89	**0.69**	**0.85**	0.78
DeepLabv3+	**0.98**	0.64	0.78	0.78
ResUNet	0.77	0.64	0.74	**0.82**
ResUNet++	0.76	0.62	0.76	0.77

Πίνακας 6.4: *Ποσοτικά αποτελέσματα αξιολόγησης μοντέλων για το σύνολο δεδομένων SegPC με επάυξηση δεδομένων*

6.2 Μελέτη επίδρασης επαύξησης δεδομένων

Σε αυτό το κομμάτι αυτού του κεφαλαίου, εξετάζεται η επίδραση της επαύξησης των δεδομένων στα ποσοτικά αποτελέσματα των πειραμάτων. Συγκεκριμένα επαναλήφθηκαν όλα τα πειράματα που φαίνονται στους παραπάνω πίνακες, με τα αρχικά σύνολα δεδομένων χωρίς να έχουν υποστεί επαύξηση, επομένως περιέχουν μικρότερο αριθμό δειγμάτων. Αναμένεται τα αποτελέσματα χωρίς επαύξηση δεδομένων να είναι χειρότερα από τα αποτελέσματα με επαύξηση δεδομένων. Επιπλέον όσο μεγαλύτερη αύξηση στα scores παρατηρείται με την αύξηση του μεγέθους των συνόλων δεδομένων τόσο επιβεβαιώνεται η ισχύς των μοντέλων, αφού αποδεικνύεται ότι η απόδοση τους θα είναι αυξημένη στην περίπτωση που παραχθούν περισσότερα και μεγαλύτερα σύνολα δεδομένων ιατρικών εικόνων στο μέλλον.

Στους πίνακες 6.5-6.8 παρουσιάζονται τα αποτελέσματα χωρίς επαύξηση δεδομένων. Παρατηρείται ότι η επαύξηση δεδομένων προκαλεί κατά προσέγγιση 4-5% αύξηση στη μετρική mIoU, 2-3% αύξηση στη μετρική DSC, 1-2% αύξηση στη μετρική Precision και 3-4% αύξηση στην μετρική Recall. Μερικά μοντέλα ευνοούνται περισσότερο από τον μεγαλύτερο όγκο δεδομένων από κάποια άλλα, όπως για παράδειγμα το DeepLabv3+ πετυχαίνει αύξηση 2% στη μετρική mIoU ενώ το MSRF-Net πετυχαίνει 4% στο σύνολο δεδομένων Kvasir-SEG. Άλλα μοντέλα φαίνεται να μην πετυχαίνουν αύξηση μέσω της επάυξησης δεδομένων, όπως για παράδειγμα τα ResUNet, ResUNet++ στο σύνολο CVC-ClinicDB, και το ResUNet-a που μάλιστα πετυχαίνει σημαντική μείωση της mIoU στο σύνολο 2018 Data Science Bowl. Γενικότερα επιβεβαιώνεται η αυξητική τάση στην απόδοση των μοντέλων συναρτήσει της αύξησης του όγκου των δεδομένων. Μοντέλα όπως το MSRF-Net,R2U-Net,TransUNet φαίνεται να κλιμακώνουν πιο γρήγορα και σε μεγαλύτερο βαθμό με την επαύξηση δεδομένων.

Method	**DSC**	**mIoU**	**Precision**	**Recall**
MSRF-Net	0.89	0.83	0.94	0.80
UNet	0.90	0.78	0.91	0.85
VNet	0.79	0.59	0.78	0.71
Att-UNet	0.83	0.67	0.86	0.75
ResUNet-a	0.84	0.70	0.89	0.76
SwinUNet	0.78	0.54	0.70	0.71
TransUNet	0.91	0.76	0.90	0.83
R2U-Net	0.90	0.76	0.92	0.81
DeepLabv3+	**0.99**	**0.87**	**0.96**	**0.91**
ResUNet	0.66	0.64	0.81	0.76
ResUNet++	0.77	0.72	0.94	0.76

Πίνακας 6.5: *Ποσοτικά αποτελέσματα αξιολόγησης μουτέλων για το σύνολο δεδομένων CVC-ClinicDB χωρίς επάυξηση δεδομένων*

6.3 Μελέτη ικανότητας γενίκευσης των μοντέλων

Η ικανότητα γενίκευσης, δηλαδή η ικανότητα των μοντέλων να αποδίδουν ικανοποιητικά σε νέα σύνολα δεδομένων που μπορεί να έχουν ληφθεί με διαφορετική τεχνολογία σε σχέση

Method	DSC	mIoU	Precision	Recall
MSRF-Net	0.81	0.69	0.84	0.79
UNet	0.87	0.64	0.83	0.74
VNet	0.75	0.44	0.69	0.55
Att-UNet	0.84	0.58	0.84	0.66
ResUNet-a	0.75	0.46	0.64	0.62
SwinUNet	0.75	0.45	0.65	0.60
TransUNet	0.87	0.66	0.80	0.80
R2U-Net	0.87	0.66	0.85	0.75
DeepLabv3+	**0.96**	**0.75**	**0.89**	**0.83**
ResUNet	0.54	0.45	0.58	0.67
ResUNet++	0.60	0.50	0.70	0.63

Πίνακας 6.6: *Ποσοτικά αποτελέσματα αξιολόγησης μοντέλων για το σύνολο δεδομένων Kvasir-SEG χωρίς επάυξηση δεδομένων*

Method	DSC	mIoU	Precision	Recall
MSRF-Net	0.91	**0.86**	0.91	0.94
UNet	0.95	**0.86**	0.91	**0.95**
VNet	0.94	0.85	0.91	0.93
Att-UNet	0.95	**0.86**	0.91	0.94
ResUNet-a	0.93	0.83	0.88	0.94
SwinUNet	0.94	0.85	0.92	0.92
TransUNet	0.95	**0.86**	**0.95**	0.90
R2U-Net	0.95	**0.86**	0.92	0.93
DeepLabv3+	**0.98**	0.84	0.92	0.91
ResUNet	0.89	0.85	0.90	0.93
ResUNet++	0.89	0.84	0.89	0.94

Πίνακας 6.7: *Ποσοτικά αποτελέσματα αξιολόγησης μοντέλων για το σύνολο δεδομένων 2018 Data Science Bowl χωρίς επάυξηση δεδομένων*

Method	DSC	mIoU	Precision	Recall
MSRF-Net	0.79	**0.67**	**0.81**	0.78
UNet	0.86	0.63	0.78	0.77
VNet	0.84	0.57	0.75	0.71
Att-UNet	0.88	0.65	0.78	**0.80**
ResUNet-a	0.74	0.39	0.52	0.39
SwinUNet	0.77	0.43	0.56	0.65
TransUNet	0.88	0.63	0.75	**0.80**
R2U-Net	0.85	0.63	0.77	0.78
DeepLabv3+	**0.97**	0.61	0.76	0.76
ResUNet	0.70	0.56	0.69	0.76
ResUNet++	0.72	0.60	0.79	0.72

Πίνακας 6.8: *Ποσοτικά αποτελέσματα αξιολόγησης μοντέλων για το σύνολο δεδομένων SegPC χωρίς επάυξηση δεδομένων*

με το σύνολο δεδομένων πάνω στο οποίο έχουν εκπαιδευτεί, είναι ένα σημαντικό χαρακτηριστικό ενός μοντέλου και κρίνεται αναγκαίο να αξιολογηθεί. Προκειμένου να αξιολογηθεί η ικανότητα γενίκευσης, τα μοντέλα εκπαιδεύονται στο ένα από τα δύο σύνολα δεδομένων που περιλαμβάνουν γαστρεντερικούς πολύποδες (CVC-ClinicDB, Kvasir-SEG) και αξιολογούνται πάνω στο άλλο σύνολο. Ομοίως για τα δύο σύνολα δεδομένων που περιέχουν πυρήνες κυττάρων (2018 Data Science Bowl, SegPC). Τα αποτελέσματα των νέων πειραμάτων φαίνονται στους πίνακες 6.9-6.11 . Τα αποτελέσματα του πειράματος εκπαίδευσης πάνω στο SegPC και αξιολόγησης στο 2018 Data Science Bowl παραλείπονται διότι είναι εξαιρετικά χαμηλής ποιότητας, καθιστώντας το SegPC ένα όχι καλό σύνολο δεδομένων για εκπαίδευση, όσον αφορά στη γενίκευση.

Method	**DSC**	**mIoU**	**Precision**	**Recall**
MSRF-Net	0.41	0.30	0.42	0.52
UNet	0.53	0.23	0.24	0.84
VNet	0.59	0.26	0.28	0.80
Att-UNet	0.56	0.24	0.26	0.80
ResUNet-a	0.46	0.10	0.25	0.10
SwinUNet	0.55	0.24	0.25	0.76
TransUNet	0.57	0.25	0.27	0.82
R2U-Net	0.55	0.24	0.25	0.78
DeepLabv3+	**0.91**	**0.53**	**0.66**	0.72
ResUNet	0.35	0.22	0.23	**0.85**
ResUNet++	0.30	0.23	0.34	0.40

Πίνακας 6.9: *Ποσοτικά αποτελέσματα αξιολόγησης μοντέλων στο σύνολο Kvasir-SEG που έχουν εκπαιδευτεί στο σύνολο δεδομένων CVC-ClinicDB*

Method	**DSC**	**mIoU**	**Precision**	**Recall**
MSRF-Net	0.52	0.42	0.53	0.67
UNet	0.75	0.45	0.69	0.56
VNet	0.66	0.30	0.44	0.49
Att-UNet	0.72	0.42	0.78	0.47
ResUNet-a	0.55	0.14	0.19	0.38
SwinUNet	0.63	0.26	0.30	0.63
TransUNet	0.76	0.47	0.70	0.59
R2U-Net	0.74	0.46	0.77	0.53
DeepLabv3+	**0.97**	**0.66**	**0.93**	**0.70**
ResUNet	0.38	0.30	0.50	0.43
ResUNet++	0.51	0.37	0.52	0.57

Πίνακας 6.10: *Ποσοτικά αποτελέσματα αξιολόγησης μοντέλων στο σύνολο CVC-ClinicDB που έχουν εκπαιδευτεί στο σύνολο δεδομένων Kvasir-SEG*

Ενώ τα αποτελέσματα γενικά δεν είναι ικανοποιητικά, δηλαδή ένα μοντέλο εκπαιδευμένο σε ένα σύνολο δεδομένων δεν αποδίδει καλά όταν αξιολογείται σε διαφορετικό σύνολο δεδομένων, με τα score στη μετρική mIoU να μην ξεπερνούν το 53% στο σύνολο CVC-ClinicDB, το 66% στο σύνολο Kvasir-SEG και το 30% στο σύνολο 2018 Data Science Bowl, τα απο-

Method	DSC	mIoU	Precision	Recall
MSRF-Net	0.42	0.27	0.27	0.96
UNet	0.69	0.29	0.29	0.94
VNet	0.69	0.29	**0.30**	0.93
Att-UNet	0.68	0.28	0.29	0.95
ResUNet-a	0.54	0.11	0.23	0.14
SwinUNet	0.66	0.26	0.27	0.95
TransUNet	0.68	0.28	0.28	0.96
R2U-Net	0.66	0.26	0.27	**0.97**
DeepLabv3+	**0.87**	0.28	0.29	0.94
ResUNet	0.39	0.25	0.25	**0.97**
ResUNet++	0.45	**0.30**	**0.30**	0.96

Πίνακας 6.11: *Ποσοτικά αποτελέσματα αξιολόγησης μοντέλων στο σύνολο SegPC που έχουν εκπαιδευτεί στο σύνολο δεδομένων 2018 Data Science Bowl*

τελέσματα μπορούν να χρησιμοποιηθούν ως μέτρο σύγκρισης της ικανότητας γενίκευσης. Το μοντέλο DeepLabv3+ εμφανίζει τη μεγαλύτερη ικανότητα γενίκευσης πετυχαίνοντας τα μεγαλύτερα score σε DCS,mIoU, ενώ αντιθέτως το μοντέλο ResUNet-a αποτυγχάνει σε αυτόν τον τομέα. Μοντέλα όπως το MSRF-Net, VNet, TransUNet, R2U-Net, Attention UNet, UNet και ResUNet++ παρουσιάζουν αρκετά καλά αποτελέσματα σε αυτά τα πειράματα, επομένως κρίνεται πως μπορούν να γενικεύουν σε ικανοποιητικό βαθμό. Μια επιπλέον παρατήρηση είναι πως μοντέλα εκπαιδευμένα στο Kvasir-SEG αποδίδουν καλύτερα στο CVC-ClinicDB ενώ το αντίθετο δεν ισχύει.

Επιπλέον στο πλαίσιο της μελέτης ικανότητας γενίκευσης καθώς και στην προσπάθεια αύξησης των δεδομένων δημιουργήθηκαν τα σύνολα δεδομένων Polyps και Cells, εκ των οποίων το πρώτο αποτελεί συνένωση των CVC-ClinicDB και Kvasir-SEG και το δεύτερο αποτελεί συνένωση των 2018 Data Science Bowl και SegPC. Σε αυτά τα πειράματα, τα μοντέλα εκπαιδεύτηκαν πάνω στα νέα σύνολα Polyps, Cells και ελέγχθηκαν και αξιολογήθηκαν πάνω στα υποσύνολα που τα συνιστούν. Τα αποτελέσματα φαίνονται στους πίνακες 6.12-6.13 .

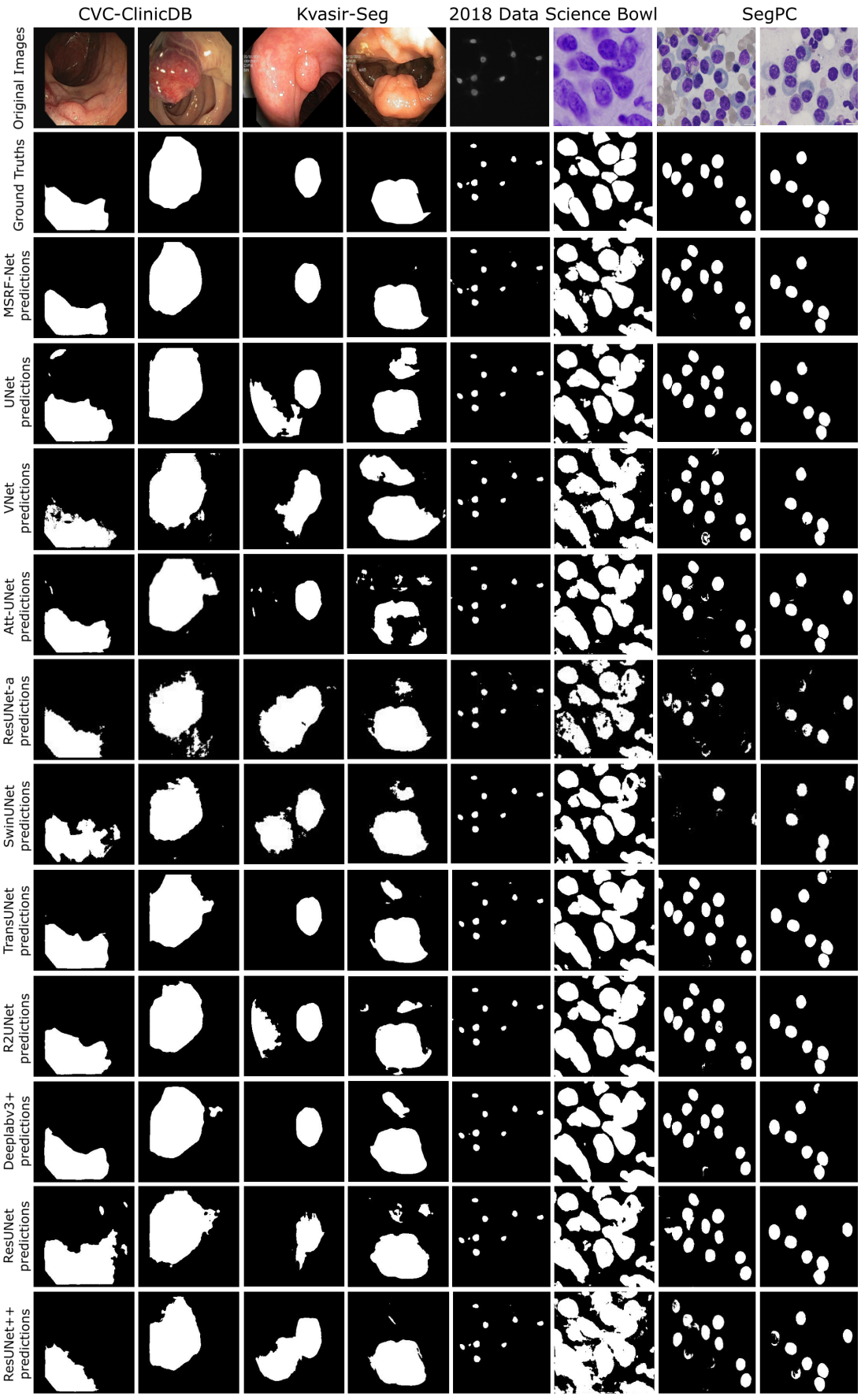
Σε αυτά τα πειράματα εξετάζεται παράλληλα και η καταλληλότητα των συνόλων δεδομένων Polyps και Cells για εκπαίδευση, καθώς και η επίδραση τους στη βελτίωση των αποτελεσμάτων. Σύμφωνα με τους πίνακες 6.12 και 6.13 αποδεικνύεται πως η εκπαίδευση των περισσότερων μοντέλων στο σύνολο Polyps οδηγεί σε αρκετά αυξημένα score κατά την αξιολόγηση στο σύνολο CVC-ClinicDB και λιγότερα αυξημένα score κατά την αξιολόγηση στο σύνολο Kvasir-SEG. Αντιθέτως, η εκπαίδευση των μοντέλων στο σύνολο Cells οδηγεί σε μειωμένα αποτελέσματα κατά την αξιολόγηση στα σύνολα 2018 Data Science Bowl και SegPC. Τα μοντέλα που αντιδρούν θετικά στην εκπαίδευση στα σύνολα Polyps και Cells σε σχέση με την εκπαίδευση στα μεμονωμένα σύνολα που τα συνιστούν, έχουν μεγαλύτερη ικανότητα γενίκευσης. Όπως αποδείχθηκε και στα παραπάνω πειράματα, μοντέλα όπως τα DeepLabv3+, MSRF-Net, VNet, Attention UNet, TransUNet, UNet και R2U-Net γενικεύουν καλύτερα από τα υπόλοιπα μοντέλα.

Tested on	**CVC-ClinicDB**				**Kvasir-SEG**			
Method	**DSC**	**mIoU**	**Precision**	**Recall**	**DSC**	**mIoU**	**Precision**	**Recall**
MSRF-Net	0.90	**0.86**	**0.94**	0.91	0.85	0.75	0.88	0.84
UNet	0.90	0.77	**0.94**	0.80	0.89	0.69	0.88	0.76
VNet	0.78	0.54	0.77	0.64	0.83	0.58	0.82	0.66
Att-UNet	0.91	0.74	**0.94**	0.78	0.87	0.64	0.82	0.75
ResUNet-a	0.81	0.61	0.92	0.65	0.79	0.53	0.77	0.63
SwinUNet	0.78	0.48	0.66	0.64	0.75	0.44	0.56	0.68
TransUNet	0.92	0.74	0.88	0.82	0.89	0.69	0.83	0.81
R2U-Net	0.89	0.75	**0.94**	0.79	0.90	0.73	**0.92**	0.78
DeepLabv3+	**0.99**	**0.86**	0.93	**0.92**	**0.96**	**0.76**	0.86	**0.87**
ResUNet	0.55	0.50	0.58	0.77	0.58	0.46	0.54	0.75
ResUNet++	0.76	0.68	0.93	0.71	0.68	0.59	0.83	0.68

Πίνακας 6.12: *Ποσοτικά αποτελέσματα αξιολόγησης μοντέλων στα σύνολα CVC-ClinicDB,Kvasir-SEG που έχουν εκπαιδευτεί στο σύνολο δεδομένων Polyps*

Tested on	**2018 Data Science Bowl**				**SegPC**			
Method	**DSC**	**mIoU**	**Precision**	**Recall**	**DSC**	**mIoU**	**Precision**	**Recall**
MSRF-Net	0.91	**0.87**	0.93	**0.93**	0.81	**0.70**	**0.84**	0.81
UNet	0.95	**0.87**	0.93	**0.93**	0.87	0.63	0.75	0.79
VNet	0.92	0.80	**0.94**	0.84	0.76	0.39	0.45	0.74
Att-UNet	0.94	0.86	**0.94**	0.91	0.87	0.63	0.70	0.86
ResUNet-a	0.94	0.84	0.91	0.92	0.75	0.37	0.43	0.73
SwinUNet	0.93	0.82	0.90	0.91	0.77	0.43	0.54	0.69
TransUNet	0.95	0.87	**0.94**	**0.93**	0.87	0.64	0.70	**0.88**
R2U-Net	0.94	0.87	0.92	**0.93**	0.87	0.64	0.74	0.83
DeepLabv3+	**0.98**	0.85	0.92	0.91	**0.97**	0.61	0.75	0.77
ResUNet	0.85	0.72	0.90	0.78	0.57	0.43	0.49	0.79
ResUNet++	0.86	0.74	0.82	0.89	0.69	0.54	0.71	0.69

Πίνακας 6.13: *Ποσοτικά αποτελέσματα αξιολόγησης μοντέλων στα σύνολα 2018 Data Science Bowl,SegPC που έχουν εκπαιδευτεί στο σύνολο δεδομένων Cells*



Εικόνα 6.1: *Παραδείγματα εικόνων που παράχθηκαν από τα μοντέλα, σε σύγκριση με τις πραγματικές εικόνες*

Μέρος III

Επίλογος

Κεφάλαιο 7

Συμπεράσματα και Μελλοντικές Επεκτάσεις

7.1 Συμπεράσματα

Τα μοντέλα που δοκιμάστηκαν στα πειράματα του προηγούμενου κεφαλαίου γενικά ακολουθούν την U-αρχιτεκτονική με παραλλαγές όπως για παράδειγμα τη χρήση μετασχηματιστών, την ύπαρξη μηχανισμού attention, την εκμετάλλευση residual μάθησης και την ανταλλαγή χαρακτηριστικών σε διαφορετικές κλίμακες. Ενώ το αρχικό μοντέλο που ενέπνευσε τη δημιουργία σχεδόν όλων των υπολοίπων, το UNet, φαίνεται να αποδίδει σταθερά αρκετά καλά σε όλα τα σύνολα δεδομένων, μερικές από τις παραλλαγές οδηγούν σε βελτιωμένα αποτελέσματα ενώ άλλες δεν είναι τόσο επιτυχημένες.

Μοντέλα όπως το ResUNet και το ResUNet-a που έχουν σχεδιαστεί για τομείς διαφορετικούς από την τμηματοποίηση ιατρικής εικόνας, δεν αποδίδουν αρκετά καλά στα σύνολα δεδομένων που αφορούν πολύποδες όμως πετυχαίνουν ικανοποιητικά scores στα σύνολα δεδομένων με πυρήνες κυττάρων, ομοίως και η 2D εκδοχή του VNet, υποδεικνύοντας πως το συγκεκριμένο έργο είναι αρκετά εξειδικευμένο και απαιτεί εξειδικευμένα μοντέλα για να επιτευχθούν State-of-the-art αποτελέσματα. Η αξιοποίηση των residual συνδέσεων σε συνδυασμό με επαναληπτικά δίκτυα κρίνεται πολύ αποτελεσματική στο δίκτυο R2U-Net ενώ ο συνδυασμός τους με Squeeze and excitation blocks, Atrous Spatial Pyramid Pooling και attention δεν οδηγεί σε κάποια βελτίωση σε σχέση με το UNet και ταυτόχρονα περιπλέκει πολύ την αρχιτεκτονική. Η αξιοποίηση του μηχανισμού attention δεν αποδίδει στα σύνολα δεδομένων με πολύποδες ωστόσο οδηγεί σε καλύτερο εντοπισμό των κυττάρων με πάθηση στο σύνολο SegPC και τον διαχωρισμό τους από τα υγιή κύτταρα, όπως συμβαίνει με το Attention UNet. Η εμφάνιση των μετασχηματιστών στις αρχιτεκτονικές τύπου U, ενώ δεν ήταν πολλά υποσχόμενη σαν ιδέα στην αρχή διότι εκ της φύσεως τους οι μετασχηματιστές δεν έχουν ικανότητα localization λόγω της έλλειψης λεπτομερειών χαμηλού επιπέδου, σε συνδυασμό με συνελικτικά δίκτυα, όπως συμβαίνει στο TransUNet, αποτελούν πολύ αποδοτικούς κωδικοποιητές επιφέροντας μικρή βελτίωση στα score του UNet. Αντιθέτως, χωρίς τη χρήση συνελικτικών δικτύων μαζί με τους μετασχηματιστές, τα τελικά αποτελέσματα δεν είναι ικανοποιητικά, όπως συμβαίνει με το SwinUNet. Τέλος τα μοντέλα που εστιάζουν περισσότερο στον συνδυασμό πληροφοριών από διαφορετικές κλίμακες και επομένως καταφέρνουν να μειώσουν το σημασιολογικό κενό ανάμεσα στα χαρακτηριστικά χαμηλού επιπέδου του κωδικοποιητή και στα χαρακτηριστικά υψηλού επιπέδου του αποκωδικοποιητή, όπως τα DeepLabv3+ και MSRF-Net, φαίνεται να πετυχαίνουν σημαντική βελτίωση στα αποτελέσμα-

τα στα σύνολα δεδομένων με πολύποδες που κρίνονται πιο απαιτητικά, καθώς και στο SegPC. Είναι σημαντικό να σημειωθεί ότι το μοντέλο MSRF-Net είναι αρκετά πιο απαιτητικό στην εκπαίδευσή του όσον αφορά υπολογιστικούς πόρους και χρόνο από τα υπόλοιπα μοντέλα. Τα R2U-Net,TransUNet πετυχαίνουν κοντινά scores με το MSRF-Net με πολύ λιγότερες απαιτήσεις. Επιπλέον, το προεκπαιδευμένο στο μεγάλου μεγέθους σύνολο δεδομένων ImageNet [85] ResNet50 backbone που διαθέτει το DeepLabv3+, του προσδίδει ισχύ και συμβάλλει στην γρήγορη εκπαίδευση.

Σε ρεαλιστικά σενάρια όπου τα μοντέλα αυτά θα εκτελούσαν τμηματοποίηση πολυπόδων πάνω σε εικόνες από διαφορετικά εργαστήρια, με διαφορετικές τεχνολογίες, από διαφορετικούς ασθενείς, η ικανότητα γενίκευσης κρίνεται ως απαραίτητο χαρακτηριστικό τους. Τα μοντέλα πρέπει να μπορούν να αποδίδουν ικανοποιητικά ανεξαρτήτως παραμέτρων όπως ο εξοπλισμός με τον οποίο απαθανατίστηκαν οι εικόνες ή ο ασθενής από τον οποίο προέρχονται οι εικόνες. Αποδείχθηκε από την μελέτη του προηγούμενου κεφαλαίου ότι τα μοντέλα που εμφανίζουν τα υψηλότερα scores έχουν και υψηλή ικανότητα γενίκευσης.

Όσον αφορά τα σύνολα δεδομένων, αυτά που σχετίζονται με πολύποδες είναι απαιτητικά και φέρουν μερικές ιδιαιτερότητες, όπως κάποια μαύρα πλαίσια που εμφανίζουν μεγάλη αντίθεση στις εικόνες και προκαλούν σύγχυση στα μοντέλα ως προς τις σημαντικές περιοχές της εικόνας, αλλά και μικρή γενικότερη αντίθεση στις εικόνες ιδιαίτερα στα σημεία των πολυπόδων η οποία δυσκολεύει το έργο του εντοπισμού τους. Στα σύνολα δεδομένων που σχετίζονται με πυρήνες κυττάρων κρίνεται αρκετά εύκολο το έργο του εντοπισμού των πυρήνων διότι υπάρχει σημαντική αντίθεση μεταξύ εκείνων και του απλού μονόχρωμου background, με δυσκολίες να παρουσιάζονται στον εντοπισμό συγκεκριμένων κυττάρων που πάσχουν από Πολλαπλό Μυέλωμα (στο SegPC) αλλά και στις εικόνες που οι πυρήνες απεικονίζονται πιο αχνοί ή υπάρχει σημαντική επικάλυψη μεταξύ τους. Αυτά τα σχόλια επιβεβαιώνονται από τα scores που παρουσιάστηκαν στο προηγούμενο κεφάλαιο.

7.2 Μελλοντικές Επεκτάσεις

Ενώ στην παρούσα εργασία εξετάστηκαν αρκετά μοντέλα με διαφορετικές αρχιτεκτονικές και επομένως οδήγησαν σε μια σφαιρική μελέτη του έργου της τμηματοποίησης ιατρικής εικόνας, υπάρχουν επιπλέον μοντέλα που θα μπορούσαν να ελεγχθούν, όπως τα DCSAU-Net [86], MCGU-Net [87] αλλά και UNet++ [88], U^2-Net [89], UNet3+ [90], τα οποία προσθέτουν καινούριες λεπτομέρειες και αλλαγές που μπορεί να ευνοήσουν σημαντικά τα αποτελέσματα των πειραμάτων. Πιο συγκεκριμένα το DCSAU-Net είναι ένα βαθύτερο και πιο συμπαγές δίκτυο με αρχιτεκτονική U-shape και split-attention που εξάγει σημαντικά χαρακτηριστικά χρησιμοποιώντας multi-scale split-attention και βαθύτερες συνελίξεις, και στόχο έχει να πετύχει υψηλή απόδοση σε πιο απαιτητικές εικόνες. Το MCGU-Net αποτελεί επέκταση του UNet με επιπλέον Squeeze and Excitation (SE) blocks, Bi-Directional Convolutional LSTM και τον μηχανισμό των πυκνών συνελίξεων. Το UNet++ αποτελεί ένα βαθιά επιβλεπόμενο δίκτυο κωδικοποιητή-αποκωδικοποιητή όπου τα υποδίκτυα κωδικοποιητή και αποκωδικοποιητή συνδέονται μέσω nested dense skip pathways που μειώνουν το σημασιολογικό κενό μεταξύ των αντίστοιχων χαρτών χαρακτηριστικών. Το Unet3+ εκμεταλλεύεται τα full-scale skip connections και τη βαθιά επίβλεψη, πετυχαίνοντας μικρότερη υπολογιστική πολυ-

πλοκότητα. Τα full-scale skip connections συνδυάζουν χαμηλών επιπέδων λεπτομέρειες με τα semantics από υψηλότερα επίπεδα σε διαφορετικές κλίμακες ενώ η βαθιά επίβλεψη βοηθά στην εκμάθηση ιεραρχικών αναπαραστάσεων από τους full-scale συγκεντρωμένους χάρτες χαρακτηριστικών. Το U^2Net σχεδιάστηκε για εντοπισμό salient αντικειμένων και η αρχιτεκτονική του αποτελείται από δύο επίπεδα της αρχιτεκτονικής UNet. Πετυχαίνει να καταγράψει περισσότερες contextual πληροφορίες σε διαφορετικές κλίμακες και αυξάνει το βάθος της αρχιτεκτονικής χωρίς τα προσθέτει επιπλέον υπολογιστικό κόστος.

Επιπλέον, στο σχήμα 6.1 εύκολα παρατηρείται πως σε κάποια παραδείγματα στα σύνολα δεδομένων CVC-ClinicDB και Kvasir-SEG, τα πιο αποδοτικά μοντέλα παράγουν μάσκες που εντοπίζουν λανθασμένα ως πολύποδες, περιοχές που δεν αποτελούν πολύποδες. Αυτό οδηγεί σε μείωση των scores στις μετρικές αξιολόγησης και σε λιγότερο ποιοτικά οπτικά αποτελέσματα. Οι ατέλειες αυτές μπορούν να μειωθούν μέσω του συνδυασμού των παραπάνω μοντέλων με κάποιο αρκετά ισχυρό μοντέλο εντοπισμού αντικειμένων (Object Detection) όπως για παράδειγμα το Detectron2 [91], το οποίο παράγει στην έξοδό του bounding boxes για το κάθε εντοπισμένο αντικείμενο. Αν συνδυαστούν οι μάσκες από τα μοντέλα που αναλύθηκαν στο κεφάλαιο 5 με τα νέα bounding boxes μπορούν να μειωθούν ή ακόμα και να εξαφανιστούν οι λανθασμένα ενεργοποιημένες (εντοπισμένες ως πολύποδες) περιοχές. Ο συνδυασμός αυτός μπορεί να γίνει σε δεύτερο χρόνο, δηλαδή αφού έχουν παραχθεί οι μάσκες και τα bounding boxes, ή και κατά την διάρκεια εκπαίδευσης του μοντέλου με την προϋπόθεση να μεταβληθεί η αρχιτεκτονική του ώστε να υποστηρίζει την εκμάθηση 2 μορφών labels (binary masks, bounding boxes), διαδικασία που ανήκει στον τομέα του Multimodal Learning.

Τέλος, μεγαλύτερη και ίσως πιο έντονη επαύξηση δεδομένων σίγουρα θα μπορούσε να επιφέρει βελτίωση των αποτελεσμάτων, εφόσον ήδη παρατηρήθηκε αυξητική τάση στα scores συναρτήσει του όγκου των δεδομένων.

Παραρτήματα

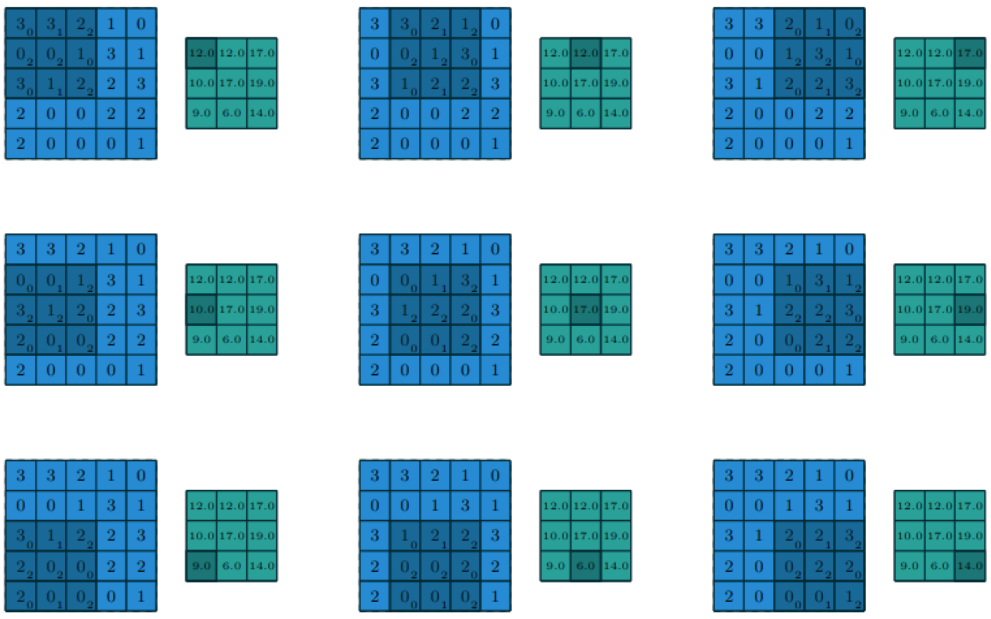
Παράρτημα Α΄

Θεωρητικό Υπόβαθρο - Ειδικότερες έννοιες

Α΄.1 Συνελίξεις

Α΄.1.1 Απλές Συνελίξεις

Οι συνελίξεις στις εικόνες λειτουργούν σαν φίλτρα. Είναι ένα είδος πράξης πινάκων, κατά την οποία ένας μικρός πίνακας με βάρη που ονομάζεται πυρήνας, περνάει από ολόκληρη την εικόνα εκτελώντας πολλαπλασιασμό στοιχείο προς στοιχείο με το κομμάτι της εικόνας το οποίο καλύπτει κάθε στιγμή, και αθροίζοντας τα αποτελέσματα σχηματίζοντας την έξοδο. Μέσω της διαδικασίας της συνέλιξης μειώνεται η διάσταση των χαρακτηριστικών σε ένα νευρωνικό δίκτυο [34].

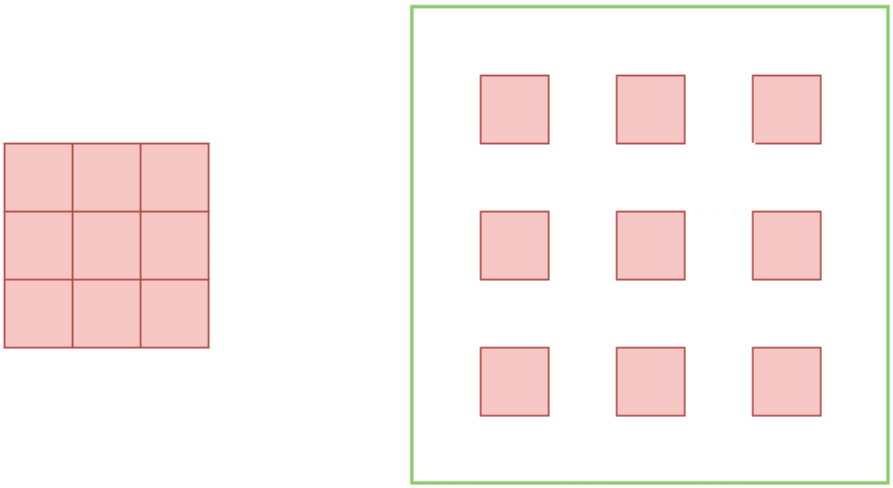


Εικόνα Α΄.1: *Παράδειγμα απλής συνέλιξης σε εικόνα [34]*

Α΄.1.2 Διεσταλμένες (dilated) Συνελίξεις

Οι διεσταλμένες (dilated/atrous) συνελίξεις εισάγουν μια παράμετρο στις απλές συνελίξεις, τον παράγοντα διαστολής. Αυτός καθορίζει το κενό ανάμεσα στις τιμές ενός πυρήνα,

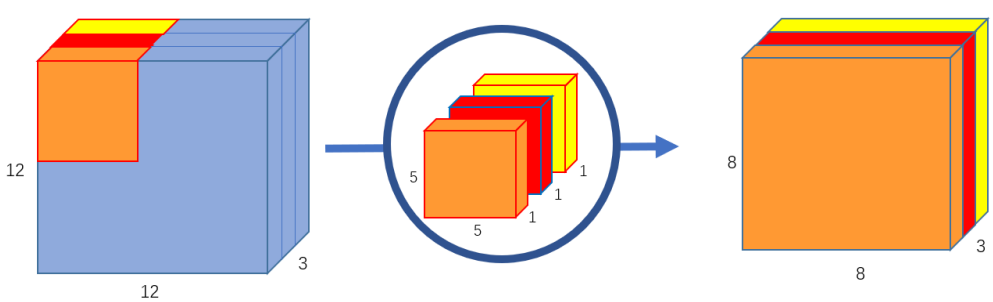
όπως φαίνεται στην εικόνα Α΄.2. Ένας πυρήνας 3x3 με παράγοντα διαστολής με τιμή 2 έχει το ίδιο receptive-field με έναν πυρήνα 5x5 χρησιμοποιώντας μόνο 9 παραμέτρους. Έτσι μεγαλώνει το receptive field των φίλτρων με ίδιο υπολογιστικό κόστος. Οι διεσταλμένες συνελίξεις συμβάλλουν στην διατήρηση της υψηλής χωρικής ανάλυσης των πινάκων χαρακτηριστικών κατά μήκος ενός συνελικτικού δικτύου οι οποίοι πίνακες περνούν από πολλαπλά στρώματα συνέλιξης και ομαδοποίησης που οδηγούν σε σημαντική αραίωση τους. Γιάυτό το λόγο χρησιμοποιούνται ευρέως στο έργο της τμηματοποίησης εικόνων σε πραγματικό χρόνο [20].



Εικόνα Α΄.2: *Παράδειγμα διεσταλμένου πυρήνα [20]*

Α΄.1.3 Depthwise Συνελίξεις

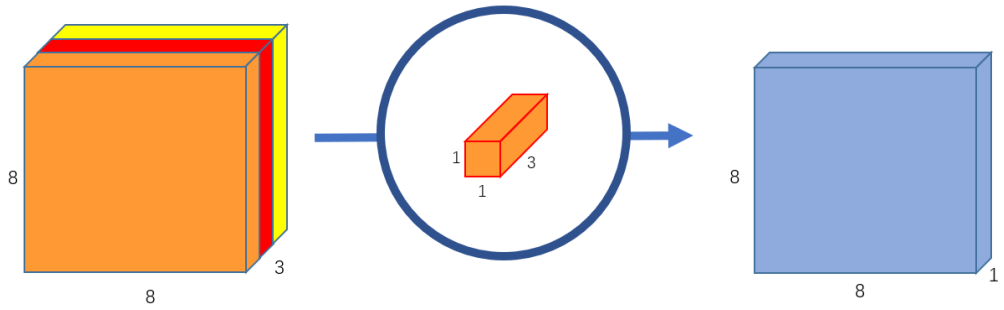
Στην depthwise συνέλιξη σε μία RGB εικόνα τριών καναλιών, το βάθος της εικόνας δεν μεταβάλλεται, δηλαδή η εικόνα της εξόδου έχει 3 κανάλια. Αυτό γίνεται με την χρήση τριών πυρήνων βάθους 1 αντί για την χρήση ενός πυρήνα βάθους 3. Κάθε πυρήνας εκτελεί συνέλιξη σε ένα κανάλι της εικόνας και τα αποτελέσματα από τις 3 συνελίξεις τοποθετούνται σε στίβα σχηματίζοντας την εικόνα εξόδου η οποία έχει 3 κανάλια [19].



Σχήμα Α΄.1: *Παράδειγμα depthwise συνέλιξης [19]*

Α΄.1.4 Pointwise Συνελίξεις

Η pointwise συνέλιξη ονομάζεται έτσι γιατί χρησιμοποιεί έναν πυρήνα 1x1 που περνάει από κάθε σημείο της εικόνας. Ο πυρήνας αυτός έχει τόσα κανάλια όσα και η εικόνα, άρα 3 για μία RGB εικόνα. επομένως στην έξοδο μιας τέτοιας συνέλιξης προκύπτει μια εικόνα ενός καναλιού με ίδιο μήκος και πλάτος με την αρχική εικόνα εισόδου [19].



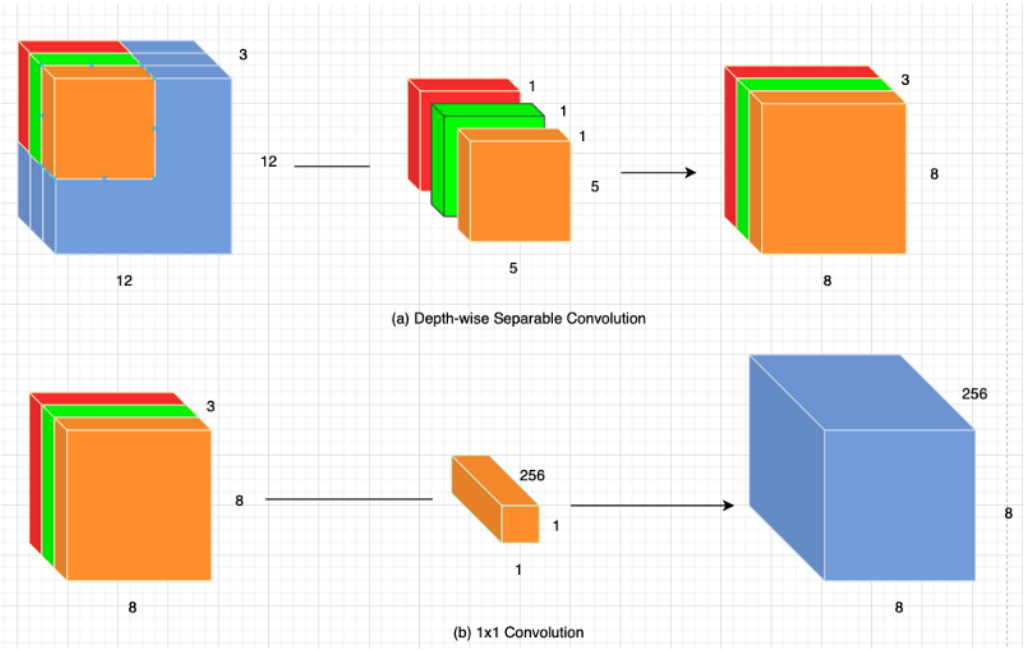
Σχήμα Α΄.2: *Παράδειγμα pointwise συνέλιξης [19]*

Α΄.1.5 Depthwise Separable Συνελίξεις

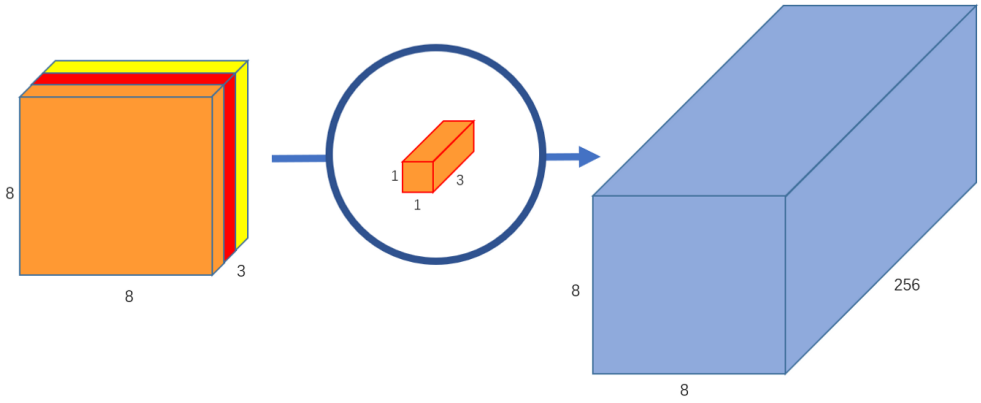
Στις depthwise separable συνελίξεις συνδυάζονται οι δύο παραπάνω τύποι συνέλιξης, όπως φαίνεται στο σχήμα Α΄.3. Μπορεί να αυξήσει τα κανάλια της εικόνας εξόδου με μικρότερη υπολογιστική πολυπλοκότητα από ότι μια απλή συνέλιξη. Στην αρχή εκτελείται μια depthwise συνέλιξη όπως στο σχήμα Α΄.1, και έπειτα εκτελούνται pointwise συνελίξεις με κατάλληλο αριθμό πυρήνων, ανάλογα με τον επιθυμητό αριθμό καναλιών εξόδου. Στο σχήμα Α΄.4 για παράδειγμα εκτελείται συνέλιξη με 256 πυρήνες 1x1 με αποτέλεσμα η εικόνα εξόδου να έχει 256 κανάλια. Το μεγάλο πλεονέκτημα που προσφέρουν οι depthwise separable συνελίξεις είναι η μικρή υπολογιστική πολυπλοκότητα. Αν έπρεπε να παραχθεί το ίδιο αποτέλεσμα με απλή συνέλιξη θα απαιτούνταν πολύ περισσότεροι πολλαπλασιασμοί και μετατροπές της εικόνας εισόδου. Αντιθέτως με την depthwise separable συνέλιξη ο αριθμός των πολλαπλασιασμών είναι σημαντικά μικρότερος και η διαδικασία αρκετά απλούστερη, επομένως αποτελεί μια πιο αποδοτική προσέγγιση [19].

Α΄.2 Squeeze and Excitation

Το μπλοκ Squeeze and Excitation είναι μια αρχιτεκτονική μονάδα που σχεδιάστηκε για την βελτίωση της αναπαραστατικής ικανότητας των δικτύων επιτρέποντας του να εκτελέσει δυναμικό επαναϋπολογισμό των χαρακτηριστικών ανά κανάλι. Το συγκεκριμένο μπλοκ περιλαμβάνει ένα συνελικτικό μπλοκ ως είσοδο. Στη συνέχεια κάθε κανάλι μετατρέπεται σε μια μοναδική αριθμητική τιμή μέσω average pooling (διαδικασία squeeze). Έπειτα ακολουθεί ένα πυκνό στρώμα με μια ReLU ενεργοποίηση και η πολυπλοκότητα καναλιών εξόδου μειώνεται. Ένα ακόμα πυκνό στρώμα με μια Sigmoid ενεργοποίηση προσδίδει σε κάθε κανάλι μια ομαλή συνάρτηση gating. Τέλος, με βάση αυτά τα αποτελέσματα αποδίδονται βάρη

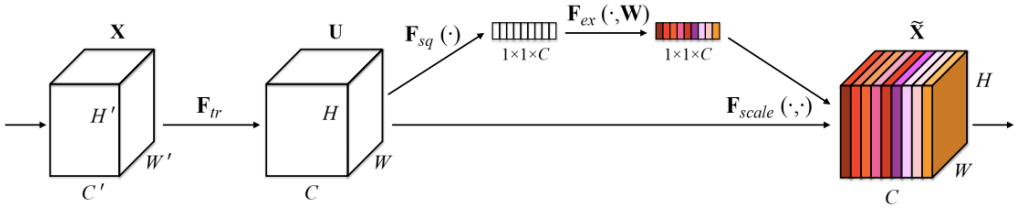


Σχήμα Α'.3: *Παράδειγμα depthwise separable συνέλιξης [20]*



Σχήμα Α'.4: *Παράδειγμα pointwise συνέλιξης με 256 πυρήνες 1x1 [20]*

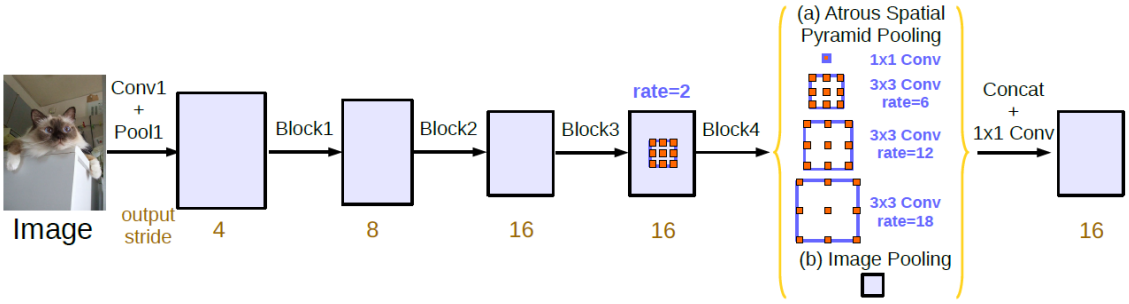
σε κάθε πίνακα χαρακτηριστικών του συνελικτικού μπλοκ (διαδικασία excitation) [21]. Η αρχιτεκτονική αυτή φαίνεται στο σχήμα Α'.5.



Σχήμα Α'.5: *Η αρχιτεκτονική Squeeze and Excitation [21]*

Α΄.3 Atrous Spatial Pyramid Pooling

Το Atrous Spatial Pyramid Pooling (διεσταλμένη χωρική ομαδοποίηση πυραμίδας) εξυπηρετεί την εξαγωγή πληροφοριών για πολλαπλές κλίμακες. Στον χάρτη χαρακτηριστικών που εξάγεται από το backbone εκτελούνται 4 παράλληλες διασταλμένες συνελίξεις (atrous) με διαφορετικούς παράγοντες διαστολής (συγκεκριμένα 1,6,12,18), ώστε να διαχειριστούν την τμηματοποίηση αντικειμένων σε διαφορετικές κλίμακες. Απεικονίζεται στο σχήμα Α΄.6 και προτάθηκε πρώτη φορά στο μοντέλο DeepLabv2 [46]. Τα αποτελέσματα από τις 4 διεσταλμένες συνελίξεις ομαδοποιούνται, συνενώνονται και περνούν από μια τελευταία συνέλιξη 1x1, πριν τα τελικά αποτελέσματα.



Σχήμα Α΄.6: *Ο μηχανισμός Atrous Spatial Pyramid Pooling [22]*

Α΄.4 Συναρτήσεις Ενεργοποίησης

Οι συναρτήσεις ενεργοποίησης τροποποιούν την έξοδο των νευρώνων φέρνοντας τις σε κατάλληλη μορφή ανάλογα με την συνέχεια της ροής της πληροφορίας στο υπόλοιπο δίκτυο, προβάλλοντας τις τιμές σε συγκεκριμένα διαστήματα. Παρακάτω παρουσιάζονται οι συναρτήσεις ενεργοποίησης που εμφανίζονται στην παρούσα εργασία.

Α΄.4.1 ReLU

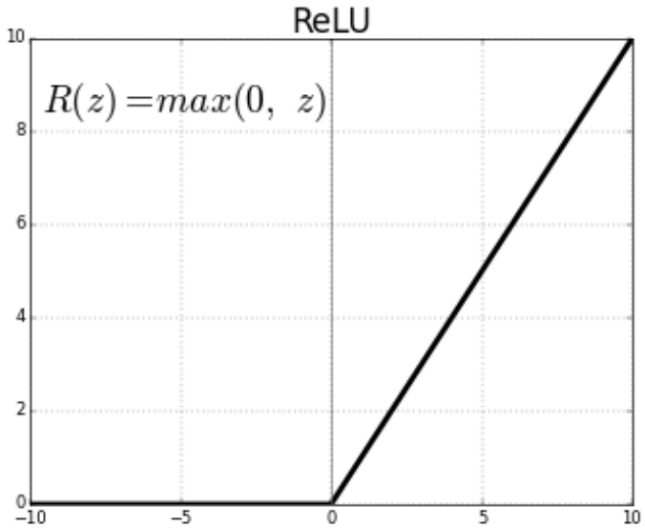
Η ReLU (Rectified Linear Unit είναι η πιο δημοφιλής συνάρτηση ενεργοποίησης [23]. Είναι μια κατά το ήμισυ ανορθωμένη συνάρτηση, δηλαδή οι αρνητικές τιμές μηδενίζονται ενώ οι θετικές τιμές παραμένουν αναλλοίωτες. Είναι μονότονη και έχει μονότονη παράγωγο. Το διάστημα τιμών είναι $[0, \infty]$. Περιγράφεται από τον τύπο $y(x) = max(0, x)$ και παρουσιάζεται στο σχήμα Α΄7.

Α΄.4.2 LeakyReLU

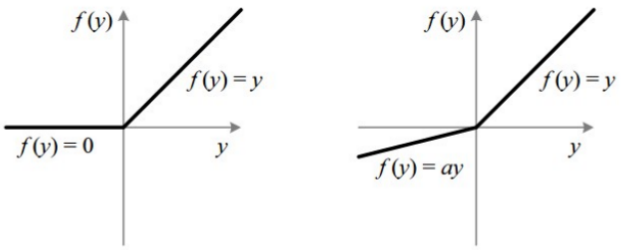
Η διαφορά της LeakyReLU από την ReLU είναι ότι οι αρνητικές τιμές δεν μηδενίζονται άλλα πολλαπλασιάζονται με έναν παράγοντα a που έχει την τιμή 0.01 [23]. Παρουσιάζεται στο σχήμα Α΄.8 και ο τύπος της είναι

$$y(x) = 0.01x, x < 0$$

$$y(x) = x, x \geq 0$$



Σχήμα Α΄.7: *Η συνάρτηση ενεργοποίησης ReLU [23]*



Σχήμα Α΄.8: *Η συνάρτηση ενεργοποίησης ReLU (αριστερά) και η LeakyReLU (δεξιά) [23]*

Α΄.4.3 PReLU

Η PReLU (Parametric ReLU) είναι ένας τύπος LeakyReLU όπου η παράμετρος a δεν έχει καθορισμένη τιμή αλλά υπολογίζεται από το δίκτυο [92]. Ο τύπος που την περιγράφει είναι ο εξής:

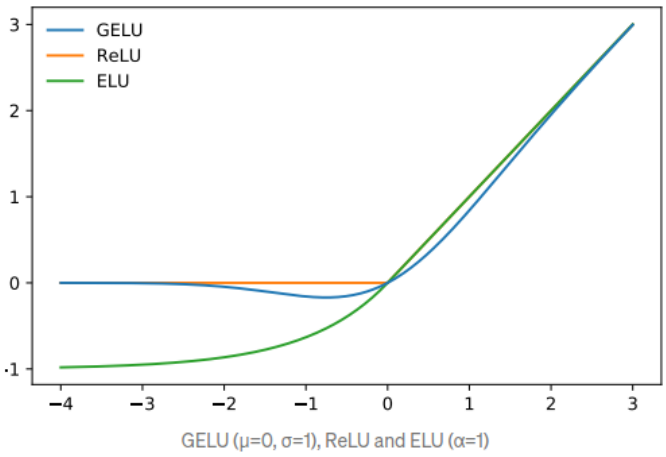
$$y(x) = ax, x < 0$$

$$y(x) = x, x \geq 0$$

Α΄.4.4 GELU

Η συνάρτηση GELU (Gaussian Error Linear Unit) συνδυάζει την γρήγορη σύγκλιση που προσφέρουν οι συναρτήσεις ReLU, PReLU, με την διαδικασία Dropout που μηδενίζει τυχαία μερικές ενεργοποιήσεις, και με την διαδικασία Zoneout που πολλαπλασιάζει στοχαστικά την είσοδο με την μονάδα. Συνεπώς στοχαστικά πολλαπλασιάζει την είσοδο με 0 ή 1 και λαμβάνει την τιμή της εξόδου ντετερμινιστικά [24]. Παρουσιάζεται στο σχήμα Α΄.9 και ο τύπος της είναι ο εξής:

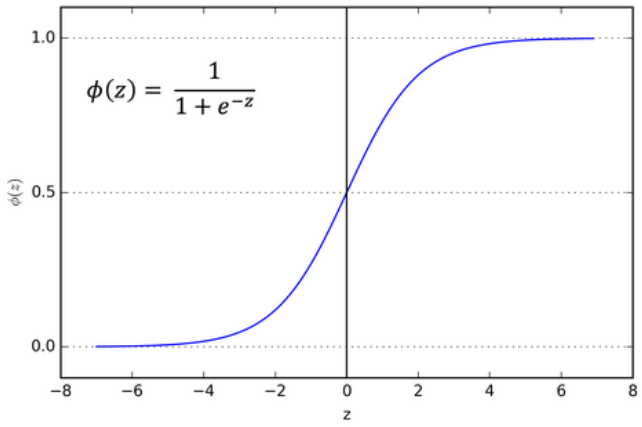
$$y(x) = 0.5x(1 + tanh(\sqrt{2/\pi}(x + 0.044715x^3)))$$



Σχήμα Α΄.9: *Η συνάρτηση ενεργοποίησης GELU (μπλέ γραμμή) [24]*

Α΄.4.5 Sigmoid

Η Sigmoid συνάρτηση ενεργοποίησης τοποθετεί τις τιμές στο διάστημα [0,1] και χρησιμοποιείται ευρέως σε μοντέλα που προβλέπουν πιθανότητες ως έξοδο. Είναι παραγωγήσιμη και μονότονη και η παράγωγος της δεν είναι μονότονη. Η συνάρτηση είναι η εξής $y(x) = \frac{1}{1+e^{-x}}$ [23]. Παρουσιάζεται στο σχήμα Α΄10.



Σχήμα Α΄.10: *Η συνάρτηση ενεργοποίησης Sigmoid [23]*

Βιβλιογραφία

[1] Adrian Rosebrock. *Introduction to neural networks*, 2021.

[2] Sumit Saha. *A comprehensive guide to Convolutional Neural Networks-the eli5 way*, 2018.

[3] Jonathan Long, Evan Shelhamer και Trevor Darrell. *Fully Convolutional Networks for Semantic Segmentation*, 2014.

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

[5] Kriz Moses. *Encoder-decoder Seq2Seq models, clearly explained*, 2021.

[6] Andrej Karpathy και Li Fei-Fei. *Deep Visual-Semantic Alignments for Generating Image Descriptions*, 2014.

[7] Kaiming He, Xiangyu Zhang, Shaoqing Ren και Jian Sun. *Deep Residual Learning for Image Recognition*, 2015.

[8] Olaf Ronneberger, Philipp Fischer και Thomas Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. *Lecture Notes in Computer Science*, σελίδες 234–241. Springer International Publishing, 2015.

[9] Fausto Milletari, Nassir Navab και Seyed Ahmad Ahmadi. *V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation*, 2016.

[10] Md Zahangir Alom, Mahmudul Hasan, Chris Yakopcic, Tarek M. Taha και Vijayan K. Asari. *Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation*, 2018.

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

[12] Zhengxin Zhang, Qingjie Liu και Yunhong Wang. *Road Extraction by Deep Residual U-Net*. 2017.

[13] Debesh Jha, Pia H. Smedsrud, Michael A. Riegler, Dag Johansen, Thomasde Lange, Pal Halvorsen και Havard D. Johansen. *ResUNet++: An Advanced Architecture for Medical Image Segmentation*, 2019.

[14] Foivos I. Diakogiannis, François Waldner, Peter Caccetta και Chen Wu. *ResUNet-a: a deep learning framework for semantic segmentation of remotely sensed data*. 2019.

[15] Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L. Yuille και Yuyin Zhou. *TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation*, 2021.

[16] Hu Cao, Yueyue Wang, Joy Chen, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian και Manning Wang. *Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation*, 2021.

[17] Liang Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff και Hartwig Adam. *Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation*, 2018.

[18] Abhishek Srivastava, Debesh Jha, Sukalpa Chanda, Umapada Pal, Håvard D. Johansen, Dag Johansen, Michael A. Riegler, Sharib Ali και Pål Halvorsen. *MSRF-Net: A Multi-Scale Residual Fusion Network for Biomedical Image Segmentation*. 2021.

[19] Chi Feng Wang. *A basic introduction to separable convolutions*, 2018.

[20] Aadhithya Sankar. *A primer on atrous convolutions and depth-wise separable convolutions*, 2021.

[21] Jie Hu, Li Shen, Samuel Albanie, Gang Sun και Enhua Wu. *Squeeze-and-Excitation Networks*, 2017.

[22] Sik Ho Tsang. *Review: Deeplabv3 - atrous convolution (semantic segmentation)*, 2019.

[23] Sagar Sharma. *Activation functions in neural networks*, 2021.

[24] Shaurya Goel. *Gelu (gaussian error linear unit)*, 2019.

[25] Ayush Pant. *Introduction to machine learning for beginners*, 2019.

[26] Alexey A. Novikov, Dimitrios Lenis, David Major, Jiri Hladůvka, Maria Wimmer και Katja Bühler. *Fully Convolutional Architectures for Multi-Class Segmentation in Chest Radiographs*, 2017.

[27] Jorge Bernal, F. Javier Sánchez, Gloria Fernández-Esparrach, Debora Gil, Cristina Rodríguez και Fernando Vilariño. *WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians*. *Computerized Medical Imaging and Graphics*, 43:99–111, 2015.

[28] Debesh Jha, Pia H Smedsrud, Michael A Riegler, Pl Halvorsen, Thomasde Lange, Dag Johansen και Hvard D Johansen. *Kvasir-seg: A segmented polyp dataset*. *International Conference on Multimedia Modeling*, σελίδες 451–462. Springer, 2020.

[29] Juan C. Caicedo, Allen Goodman, Kyle W. Karhohs, Beth A. Cimini, Jeanelle Ackerman, Marzieh Haghighi, CherKeng Heng, Tim Becker, Minh Doan, Claire McQuin, Mohammad Rohban, Shantanu Singh και Anne E. Carpenter. *Nucleus segmentation across imaging experiments: the 2018 Data Science Bowl*. *Nature Methods*, 16(12):1247–1253, 2019.

[30] Anubha Gupta, Pramit Mallick, Ojaswa Sharma, Ritu Gupta και Rahul Duggal. *PCSeg: Color model driven probabilistic multiphase level set based tool for plasma cell segmentation in multiple myeloma*. *PLOS ONE*, 13(12):e0207908, 2018.

[31] Anubha Gupta, Rahul Duggal, Shiv Gehlot, Ritu Gupta, Anvit Mangal, Lalit Kumar, Nisarg Thakkar και Devprakash Satpathy. *GCTI-SN: Geometry-inspired chemical and tissue invariant stain normalization of microscopic medical images*. *Medical Image Analysis*, 65:101788, 2020.

[32] Shiv Gehlot, Anubha Gupta και Ritu Gupta. *EDNFC-Net: Convolutional Neural Network with Nested Feature Concatenation for Nuclei-Instance Segmentation*. *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020.

[33] Anubha Gupta, Ritu Gupta, Shiv Gehlot και Shubham Gehlot. *SegPC-2021: Segmentation of Multiple Myeloma Plasma Cells in Microscopic Images*, 2021.

[34] Vincent Dumoulin και Francesco Visin. *A guide to convolution arithmetic for deep learning*, 2016.

[35] Athanasios Tagaris, Dimitrios Kollias και Andreas Stafylopatis. *Assessment of Parkinson's disease based on deep neural networks*. *International Conference on Engineering Applications of Neural Networks*, σελίδες 391–403. Springer, 2017.

[36] Ilianna Kollia, Andreas Georgios Stafylopatis και Stefanos Kollias. *Predicting Parkinson's disease using latent information extracted from deep neural networks*. *2019 International Joint Conference on Neural Networks (IJCNN)*, σελίδες 1–8. IEEE, 2019.

[37] James Wingate, Ilianna Kollia, Luc Bidaut και Stefanos Kollias. *Unified deep learning approach for prediction of Parkinson39;s disease*. *IET Image Processing*, 14(10):1980–1989, 2020.

[38] Dimitrios Kollias, Anastasios Arsenos, Levon Soukissian και Stefanos Kollias. *Mia-cov19d: Covid-19 detection through 3-d chest ct image analysis*. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, σελίδες 537–544, 2021.

[39] Dimitrios Kollias, Anastasios Arsenos και Stefanos Kollias. *Ai-mia: Covid-19 detection & severity analysis through medical imaging*. *arXiv preprint arXiv:2206.04732*, 2022.

[40] Dimitris Kollias, Y Vlaxos, M Seferis, Ilianna Kollia, Levon Sukissian, James Wingate και S Kollias. *Transparent adaptation in deep medical image diagnosis*. *International*

Workshop on the Foundations of Trustworthy AI Integrating Learning, Optimization and Reasoning, σελίδες 251–267. Springer, 2020.

[41] Athanasios Tagaris, Dimitrios Kollias, Andreas Stafylopatis, Georgios Tagaris και Stefanos Kollias. *Machine learning for neurodegenerative disorder diagnosis—survey of practices and launch of benchmark dataset*. *International Journal on Artificial Intelligence Tools*, 27(03):1850011, 2018.

[42] Dimitrios Kollias, Athanasios Tagaris, Andreas Stafylopatis, Stefanos Kollias και Georgios Tagaris. *Deep neural architectures for prediction in healthcare*. *Complex & Intelligent Systems*, 4(2):119–131, 2018.

[43] Dimitrios Kollias, Miao Yu, Athanasios Tagaris, Georgios Leontidis, Andreas Stafylopatis και Stefanos Kollias. *Adaptation and contextualization of deep neural network models*. *2017 IEEE symposium series on computational intelligence (SSCI)*, σελίδες 1–8. IEEE.

[44] Francesco Caliva, Fabio Sousa De Ribeiro, Antonios Mylonakis, Christophe Demazi'ere, Paolo Vinai, Georgios Leontidis και Stefanos Kollias. *A deep learning approach to anomaly detection in nuclear reactors*. *2018 International joint conference on neural networks (IJCNN)*, σελίδες 1–8. IEEE, 2018.

[45] Nabil Ibtehaz και M. Sohel Rahman. *MultiResUNet : Rethinking the U-Net Architecture for Multimodal Biomedical Image Segmentation*. 2019.

[46] Liang Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy και Alan L. Yuille. *DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs*, 2016.

[47] Towaki Takikawa, David Acuna, Varun Jampani και Sanja Fidler. *Gated-SCNN: Gated Shape CNNs for Semantic Segmentation*. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, 2019.

[48] Jingdong Wang, Ke Sun, Tianheng Cheng, Borui Jiang, Chaorui Deng, Yang Zhao, Dong Liu, Yadong Mu, Mingkui Tan, Xinggang Wang, Wenyu Liu και Bin Xiao. *Deep High-Resolution Representation Learning for Visual Recognition*, 2019.

[49] Fabio De Sousa Ribeiro, Francesco Caliv&39;a, Mark Swainson, Kjartan Gudmundsson, Georgios Leontidis και Stefanos Kollias. *Deep bayesian self-training*. *Neural Computing and Applications*, 32(9):4275–4291, 2020.

[50] Phivos Mylonas, Evaggelos Spyrou, Yannis Avrithis και Stefanos Kollias. *Using visual context and region semantics for high-level concept detection*. *IEEE Transactions on Multimedia*, 11(2):229–243, 2009.

[51] Pranoy Radhakrishnan. *Why transformers are slowly replacing cnns in Computer Vision?*, 2021.

[52] Salman Khan, Muzammal Naseer, Munawar Hayat, Syed Waqas Zamir, Fahad Shahbaz Khan και Mubarak Shah. *Transformers in Vision: A Survey*. *ACM Computing Surveys*, 2022.

[53] Stéphane d'Ascoli, Hugo Touvron, Matthew Leavitt, Ari Morcos, Giulio Biroli και Levent Sagun. *ConViT: Improving Vision Transformers with Soft Convolutional Inductive Biases*, 2021.

[54] Hack A BIT. *Semantic segmentation*, 2019.

[55] Analytics Vidhya. *Attention mechanism in deep learning*, 2020.

[56] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhutdinov, Richard Zemel και Yoshua Bengio. *Show, Attend and Tell: Neural Image Caption Generation with Visual Attention*, 2015.

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

[58] Lilian Weng. *Attention? Attention!* *lilianweng.github.io*, 2018.

[59] Wanshun Wong. *What is residual connection?*, 2021.

[60] Bashar Alhnaity, Stefanos Kollias, Georgios Leontidis, Shouyong Jiang, Bert Schamp και Simon Pearson. *An autoencoder wavelet based deep neural network with attention mechanism for multi-step prediction of plant growth*. *Information Sciences*, 560:35–50, 2021.

[61] Andreas Psaroudakis και Dimitrios Kollias. *MixAugment & Mixup: Augmentation Methods for Facial Expression Recognition*. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, σελίδες 2367–2375, 2022.

[62] Debesh Jha, Pia H. Smedsrud, Dag Johansen, Thomasde Lange, Havard D. Johansen, Pal Halvorsen και Michael A. Riegler. *A Comprehensive Study on Colorectal Polyp Segmentation With ResUNet, Conditional Random Field and Test-Time Augmentation*. *IEEE Journal of Biomedical and Health Informatics*, 25(6):2029–2040, 2021.

[63] Konstantin Pogorelov, Kristin Ranheim Randel, Carsten Griwodz, Sigrun Losada Eskeland, Thomasde Lange, Dag Johansen, Concetto Spampinato, Duc Tien Dang-Nguyen, Mathias Lux, Peter Thelin Schmidt, Michael Riegler και Pl Halvorsen. *KVASIR: A Multi-Class Image Dataset for Computer Aided Gastrointestinal Disease Detection*. *Proceedings of the 8th ACM on Multimedia Systems Conference*, MMSys'17, σελίδες 164–169, New York, NY, USA, 2017. ACM.

[64] Kaiming He, Xiangyu Zhang, Shaoqing Ren και Jian Sun. *Deep Residual Learning for Image Recognition*, 2015.

[65] Ming Liang και Xiaolin Hu. *Recurrent Convolutional Neural Network for Object Recognition*. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[66] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang και Jiaya Jia. *Pyramid Scene Parsing Network*. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[67] D Kollias, N Bouas, Y Vlaxos, V Brillakis, M Seferis, I Kollia, L Sukissian, J Wingate και S Kollias. *Deep Transparent Prediction through Latent Representation Analysis*. *arXiv preprint arXiv:2009.07044*, 2020.

[68] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin και Baining Guo. *Swin Transformer: Hierarchical Vision Transformer using Shifted Windows*, 2021.

[69] Kronland Martinet R. Morlet J. Tchamitchian P. Holschneider, M. *A real-timealgorithm for signal analysis with the help of the wavelet transform*. σελίδες 289–297, 1989.

[70] Pierre Sermanet, David Eigen, Xiang Zhang, Michael Mathieu, Rob Fergus και Yann LeCun. *OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks*, 2013.

[71] Alessandro Giusti, Dan C. Cireşan, Jonathan Masci, Luca M. Gambardella και Jürgen Schmidhuber. *Fast Image Scanning with Deep Max-Pooling Convolutional Neural Networks*. 2013.

[72] George Papandreou, Iasonas Kokkinos και Pierre Andre Savalle. *Modeling local and global deformations in Deep Learning: Epitomic convolution, Multiple Instance Learning, and sliding window detection*. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2015.

[73] François Chollet. *Xception: Deep Learning with Depthwise Separable Convolutions*, 2016.

[74] Laurent SIfre και Stéphane Mallat. *Rigid-Motion Scattering for Texture Classification*, 2014.

[75] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto και Hartwig Adam. *MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications*, 2017.

[76] Jia Deng, Wei Dong, Richard Socher, Li Jia Li, Kai Li και Li Fei-Fei. *ImageNet: A large-scale hierarchical image database*. *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2009.

[77] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah και Kyoung Mu Lee. *Enhanced Deep Residual Networks for Single Image Super-Resolution*, 2017.

[78] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke και Alex Alemi. *Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning*, 2016.

[79] Towaki Takikawa, David Acuna, Varun Jampani και Sanja Fidler. *Gated-SCNN: Gated Shape CNNs for Semantic Segmentation*, 2019.

[80] Daniel Godoy. *Understanding binary cross-entropy / log loss: A visual explanation*, 2019.

[81] Ekin Tiu. *Metrics to evaluate your semantic segmentation model*, 2020.

[82] Adam Shafi. *How to learn the definitions of precision and recall*, 2022.

[83] Diederik P. Kingma και Jimmy Ba. *Adam: A Method for Stochastic Optimization*, 2014.

[84] GRNET S.A. *ARIS Documentation - Hardware overview*.

[85] Jia Deng, Wei Dong, Richard Socher, Li Jia Li, Kai Li και Li Fei-Fei. *Imagenet: A large-scale hierarchical image database*. *2009 IEEE conference on computer vision and pattern recognition*, σελίδες 248–255. Ieee, 2009.

[86] Qing Xu, Wenting Duan και Na He. *DCSAU-Net: A Deeper and More Compact Split-Attention U-Net for Medical Image Segmentation*, 2022.

[87] Maryam Asadi-Aghbolaghi, Reza Azad, Mahmood Fathy και Sergio Escalera. *Multi-level Context Gating of Embedded Collective Knowledge for Medical Image Segmentation*, 2020.

[88] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh και Jianming Liang. *UNet: A Nested U-Net Architecture for Medical Image Segmentation*. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, σελίδες 3–11. Springer International Publishing, 2018.

[89] Xuebin Qin, Zichen Zhang, Chenyang Huang, Masood Dehghan, Osmar R. Zaiane και Martin Jagersand. *U^2-Net: Going Deeper with Nested U-Structure for Salient Object Detection*. 2020.

[90] Huimin Huang, Lanfen Lin, Ruofeng Tong, Hongjie Hu, Qiaowei Zhang, Yutaro Iwamoto, Xianhua Han, Yen Wei Chen και Jian Wu. *UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation*, 2020.

[91] Yuxin Wu, Alexander Kirillov, Francisco Massa, Wan Yen Lo και Ross Girshick. *Detectron2*. https://github.com/facebookresearch/detectron2, 2019.

[92] Danqing Liu. *A practical guide to relu*, 2017.