



Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης
Πολυτεχνική Σχολή
Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Εντοπισμός και Ταξινόμηση επιμολυσμένων
ερυθροκυττάρων από το πλασμάτιο της ελονοσίας σε
εικόνες λεπτού επιχρίσματος αίματος μέσω τεχνικών βαθιάς
μάθησης.

Διπλωματική Εργασία
του
Κινού Αλέξανδρου

Επιβλέποντες

Καθηγητής κ. Χατζηλεοντιάδης Λεόντιος
Υποψήφιος Διδάκτωρ κ. Δημαρίδης Ιωάννης

Θεσσαλονίκη, Ιούνιος 2023

Περίληψη

Οι αξιόπιστες τεχνικές διάγνωσης της ελονοσίας ,που είναι κατάλληλες για δοκιμές σε περιοχές με υψηλή επιθάρυνση, είναι ζωτικής σημασίας για την αποτελεσματική θεραπεία και παρακολούθηση της νόσου. Η ταυτοποίηση των παρασίτων της ελονοσίας σε αντικειμενοφόρες πλάκες αίματος με χρώση Giemsa είναι σήμερα η πιο ευρέως αποδεκτή τεχνική, αλλά η διαθεσιμότητά της περιορίζεται από την ανάγκη για υψηλά καταρτισμένους εμπειρογνώμονες για την ερμηνεία των δεδομένων. Η παρούσα διπλωματική εργασία παρουσιάζει μια προσέγγιση μηχανικής μάθησης για την αυτοματοποιημένη ανίχνευση και ταξινόμηση των μολυσμένων από ελονοσία ερυθρών αιμοσφαιρίων (RBCs). Το πρώτο στάδιο της διαδικασίας χρησιμοποιεί τον αλγόριθμο YoloV7 για την ανίχνευση RBCs σε εικόνες λεπτών επιχρισμάτων αίματος, ο οποίος αποδεικνύεται ότι έχει τον χαμηλότερο χρόνο δοκιμών και παράλληλα την υψηλότερη ακρίβεια. Το στάδιο της ταξινόμησης χρησιμοποιεί ένα συνελικτικό νευρωνικό δίκτυο , η αρχιτεκτονική του οποίου βασίζεται στο LRNet, με ορισμένες προσαρμογές και βελτιώσεις που έγιναν για τους σκοπούς του προβλήματος. Η απόδοση της προτεινόμενης προσέγγισης αξιολογήθηκε με τη χρήση διαφορετικών συνόλων δεδομένων εικόνων λεπτών επιχρισμάτων αίματος και συγκρίθηκε με την υπάρχουσα σύγχρονη προσέγγιση. Τα αποτελέσματα καταδεικνύουν ότι η προτεινόμενη προσέγγιση μπορεί να βελτιώσει την ταχύτητα και την ακρίβεια της διάγνωσης της ελονοσίας, ενώ έχει υψηλή ευρωστία ,ειδικότητα και ευαισθησία.

Abstract

Reliable malaria diagnosis techniques that are suitable for testing in high burden areas, are vital for effective treatment and monitoring of the disease. Identification of malaria parasites in Giemsa stained blood slides is currently the most widely accepted technique, but its availability is limited by the need for highly trained experts to interpret the data. This thesis presents a machine learning approach for the automated detection and classification of malaria-infected Red blood cells (RBCs). The object detection stage utilizes YoloV7 algorithm to detect RBCs in thin blood smear images which is shown to have the least inference time while preserving high accuracy. The classification stage utilizes a Convolutional Neural Network, the architecture of which is based on LRNet, with adaptations and improvements made for this use case. The performance of the proposed approach was evaluated using different datasets of thin blood smear images and compared with the existing state of the art approach. The results demonstrate that the proposed approach can improve the speed and accuracy of malaria diagnosis while having high generalizability, specificity and sensitivity.

Εντοπισμός και Ταξινόμηση επιμολυσμένων
ερυθροκυττάρων απο το πλασμάδιο της ελονοσίας σε
εικόνες λεπτού επιχρίσματος αίματος μέσω τεχνικών βαθίας
μάθησης.

Κινούς Αλέξανδρος
akinous@ece.auth.gr

Περιεχόμενα

1 Εισαγωγή	3
1.1 Διατύπωση του προβλήματος	3
1.2 Σκοπός της Διπλωματικής	4
2 Θεωρητικό Υπόβαθρο	5
2.1 Μέθοδοι διάγνωσης ελονοσίας και περιορισμοί τους	5
2.1.1 Οπτική μικροσκοπία με χρώση επιχρίσματος	5
2.1.2 Μικροσκοπία Φθορισμού	6
2.1.3 Τέστ ταχείας διάγνωσης	6
2.1.4 PCR τεστ	6
2.2 Μηχανική Μάθηση	7
2.2.1 Βαθιά νευρωνικά δίκτυα με συνελεκτικά μπλοκ	7
2.2.2 Μηχανισμός προσοχής καναλιών	8
2.2.3 Forward and Back Propagation	9
2.2.4 Φαινομενο εξαφανιζομενων κλισεων και αντιγραφόμενων κλισεων	11
2.2.5 Συναρτησεις Ενεργοποίησης	12
2.3 Βιβλιογραφική Ανασκόπηση	14
3 Σύνολα Δεδομένων και χρήση κλασσικής επεξεργασίας εικόνας	15
3.1 Σύνολα δεδομένων	15
3.1.1 BCCD-Dataset	15
3.1.2 Archive-Dataset	16
3.1.3 NIH-Dataset	16
3.1.4 Custom-NIH-Dataset	16
3.1.5 Μοντέλο χρώματος HSV	17
3.1.6 Contrast Limited Adaptive Histogram Equalization (C.L.A.H.E)	17
3.2 Κατακερματισμός ερυθροκυττάρων μέσω τεχνικών επεξεργασίας εικόνας	18
3.2.1 Otsu's Thresholding	19
3.2.2 Μορφολογικοί μετασχηματισμοί δυαδικής μάσκας	20
3.2.3 Μέθοδος Watershed	21
3.2.4 Μετρική DICE	22
4 Υλοποίηση-Μεθοδολογία	23
4.1 Darknet και Yolo για ανίχνευση ερυθροκυττάρων σε εικόνες επιχρίσματος αίματος	23
4.1.1 Αρχιτεκτονική YoloV4	24
4.1.2 Αρχιτεκτονική YoloV7	27
4.1.3 Παράμετροι δικτύων	32
4.1.4 Επαναρχικοποίηση κελιών	33
4.1.5 Εκπαίδευση	33
4.2 Ταξινόμηση μεμονωμένων ερυθροκυττάρων σε επιμολυσμένα ή μη από τον ιό της ελονοσίας	34

4.2.1	Wide Multi-Kernel Network	34
4.2.2	EfficientNet	36
4.2.3	ConvNext	38
4.2.4	Ensemble Network	39
4.2.5	Εκπαίδευση	39
5	Αποτελέσματα Νευρωνικών δικτύων	40
5.1	Ανίχνευση ερυθροκυττάρων σε επίχρισμα αίματος	40
5.1.1	Μετρικές στο σύνολο εικόνων εκπαίδευσης	40
5.1.2	Μετρικές στο σύνολο εικόνων επαλήθευσης	42
5.1.3	Μετρικές στο σύνολο εικόνων δοκιμών	44
5.1.4	Μετάβαση από οριοθετούμενα πλαίσια σε δυαδικές μάσκες	46
5.2	Ταξινόμηση ερυθροκυττάρων σε επιμολυσμένα ή μη από το πλασμάδιο της ελονοσίας	47
5.2.1	Μετρικές στο σύνολο εικόνων εκπαίδευσης και επαλήθευσης	47
5.2.2	Μετρικές στο σύνολο εικόνων δοκιμών	50
5.2.3	Floating Point Operations (FLOPs) και μέσος χρόνος εκτέλεσης	53
5.2.4	Μετρικές σε εξωτερικά σύνολα δεδομένων BCCD και Archive	54
6	Σχολιασμός και συμπεράσματα	55
	Κατάλογος Σχημάτων	61

Κεφάλαιο 1

Εισαγωγή

1.1 Διατύπωση του προβλήματος

Η ελονοσία είναι μια ασθένεια που προκαλείται από το παράσιτο πλασμώδιο, το οποίο μεταδίδεται στον άνθρωπο με το τσίμπημα επιμολυσμένων κουνουπιών. Σύμφωνα με τον Παγκόσμιο Οργανισμό Υγείας (ΠΟΥ), η ελονοσία αποτελεί κύρια αιτία νοσηρότητας και θνησιμότητας παγκοσμίως, με εκτιμώμενα 229 εκατομμύρια κρούσματα και 409.000 θανάτους ετησίως, κυρίως σε παιδιά ηλικίας κάτω των 5 ετών που ζουν στην υποσαχάρια Αφρική. Η ταχεία διαγνωστική δοκιμή (RDT) είναι επί του παρόντος η προτιμώμενη μέθοδος για τη διάγνωση της ελονοσίας σε περιβάλλοντα περιορισμένων πόρων. Ωστόσο, οι RDT έχουν περιορισμούς, συμπεριλαμβανομένης της χαμηλής ευαισθησίας και ειδικότητας, ιδίως για λοιμώξεις χαμηλής πυκνότητας. Η μικροσκοπηση, ο χρυσός κανόνας για τη διάγνωση της ελονοσίας, είναι πιο ευαίσθητη και ειδική, αλλά απαιτεί εξειδικευμένο προσωπικό και εξοπλισμό καλής ποιότητας που μπορεί να είναι δύσκολο να βρεθεί σε τριτοκοσμικές χώρες στις οποίες μαστίζει η ασθένεια. Η τρέχουσα θεραπεία για την ελονοσία βασίζεται στη χορήγηση αντιελονοσιακών φαρμάκων, όπως οι συνδυασμένες θεραπείες με βάση την αρτεμισινίνη (ACT). Ωστόσο, η εμφάνιση ανθεκτικότητας σε αυτά τα φάρμακα αποτελεί μείζονα ανησυχία για τη δημόσια υγεία, αναδεικνύοντας την ανάγκη για βελτιωμένες διαγνωστικές μεθόδους. Η υπερβολική χρήση των εν λόγω φαρμάκων μπορεί να συμβάλει στην ανάπτυξη ανθεκτικότητας στα φάρμακα, γεγονός που καθιστά δυσκολότερη τη θεραπεία της νόσου στο μέλλον ενώ η χρήση τους σε υγιή άτομα λόγω λανθασμένης διάγνωσης μπορεί να προκαλέσει παρενέργειες, όπως στομαχικές διαταραχές και αλλεργικές αντιδράσεις έως και θάνατο. Ως εκ τούτου, είναι ζωτικής σημασίας η ανάπτυξη ακριβών διαγνωστικών μεθόδων και η κατάλληλη χρήση τους.

Η αυτοματοποιημένη ακριβής ανίχνευση και ταξινόμηση των μολυσμένων από ελονοσία ερυθρών αιμοσφαιρίων (RBC) σε εικόνες λεπτών επιχρισμάτων αίματος παραγομενες από χαμηλής ποιότητας εξοπλισμό μπορεί να αποτελέσει μια πολλά υποσχόμενη λύση για την υπέρβαση αυτών των περιορισμών και τη βελτίωση της ακρίβειας της διάγνωσης της ελονοσίας. Η χρήση τεχνικών υπολογιστικής όρασης και βαθιάς μηχανικής μάθησης μπορεί να βελτιώσει την ταχύτητα και την ακρίβεια της διάγνωσης της ελονοσίας, και σε αυτό επικεντρώνεται η παρούσα διπλωματική εργασία.

1.2 Σκοπός της Διπλωματικής

Κατά την ανάπτυξη μιας εναλλακτικής διαγνωστικής μεθόδου που είναι κατάλληλη για εξέταση σε σημείο φροντίδας σε περιβάλλοντα περιορισμένων πόρων, θα πρέπει ιδανικά να πληρεί τα ακόλουθα κριτήρια :

Diagnostic Test Requirements

1. Detection limit ≤ 50 parasites / μ l of blood, specificity $\geq 90\%$;
2. Ability to determine parasite count;
3. Ability to identify parasite species and stage;
4. Low cost of equipment, minimal use of electricity, etc.;
5. Minimal labour and skill required.

Σχήμα 1.1: Κριτήρια ενός αποδοτικού τρόπου ανίχνευσης της ελονοσίας με βάση τον Παγκόσμιο Οργανισμό Υγείας

Είναι σαφές ότι οποιαδήποτε μέθοδος διάγνωσης της ελονοσίας θα γίνει καταλληλότερη για τον στόχο όταν η εξάρτηση από έναν ανθρώπινο εμπειρογνώμονα περιορίζεται ή καταργείται εντελώς. Αυτό μπορεί επιτευχθεί μέσω της αυτοματοποίησης. Ο κύριος στόχος της παρούσας εργασίας είναι η ανάπτυξη μιας προσέγγισης μηχανικής μάθησης για την αυτοματοποιημένη ανίχνευση και ταξινόμηση ερυθρών αιμοσφαιρίων που έχουν μολυνθεί από ελονοσία σε εικόνες λεπτών επιχρισμάτων αίματος η οποία ιδανικά να πληρεί τις παραπάνω προϋποθέσεις. Η προτεινόμενη προσέγγιση χρησιμοποιεί το YOLO v7 για την ανίχνευση των ερυθρών αιμοσφαιρίων και έναν προσαρμοσμένο αλγόριθμο ταξινόμησης εξατομικευμένο για υψηλή γενικευσιμότητα, υψηλή ειδικότητα και ευαισθησία. Ο απώτερος στόχος αυτής της έρευνας είναι η δημιουργία ενός οικονομικώς αποδοτικού μικροσκοπίου που μπορεί να ανιχνεύει αυτόματα την παρασιταίμια, χωρίς την παρεμβολή ειδικού σε περιβάλλοντα περιορισμένων πόρων. Η παρούσα διπλωματική επικεντρώνεται στην ανάπτυξη του αγωγού από ακρο σε ακρο που λειτουργεί πίσω από ένα τέτοιο μικροσκόπιο και παραγει προβλεψεις. Επιπλέον αποσκοπεί στην αντιμετώπιση των περιορισμών των σημερινών διαγνωστικών μεθόδων και στη συμβολή στη μείωση της επιβάρυνσης από την ελονοσία παγκοσμίως.

Κεφάλαιο 2

Θεωρητικό Υπόβαθρο

2.1 Μέθοδοι διάγνωσης ελονοσίας και περιορισμοί τους

2.1.1 Οπτική μικροσκοπία με χρώση επιχρίσματος

Η ταυτοποίηση των παρασίτων μέσω της επιθεώρησης των επιχρισμάτων αίματος με φωτεινή μικροσκοπία είναι επί του παρόντος η συνιστώμενη μέθοδος για την ακριβή διάγνωση της ελονοσίας. Η συνήθης μικροσκοπική διαγνωστική διαδικασία, όπως συνιστάται από τον ΠΟΥ, αποτελείται από τα ακόλουθα βήματα [1]:

1. **Προετοιμασία επιχρίσματος:** Λαμβάνεται δείγμα περιφερικού αίματος του ασθενούς, συνήθως από το δάχτυλο. Το αίμα εφαρμόζεται σε μια μικροσκοπική πλάκα και δημιουργείται ένα παχύ ή λεπτό φιλμ. Τα παχιά υμένα αποτελούνται από πολλαπλά στρώματα αιμοσφαιρίων, ενώ τα λεπτά υμένα απλώνονται έτσι ώστε να υπάρχει μόνο ένα στρώμα ερυθροκυττάρων.
2. **Χρώση:** Για να καταστεί δυνατή η διάκριση των κυττάρων του αίματος και των παρασίτων, εφαρμόζεται μια χρώση. Οι αντικειμενοφόροι πλάκες μικροσκοπίου χρωματίζονται συνήθως με χρώση Giemsa, αλλά και άλλες χρώσεις τύπου Romanowsky, όπως οι χρώσεις Field και Leishman, μπορούν επίσης να χρησιμοποιηθούν. Η χρώση Field έχει το πλεονέκτημα του πολύ σύντομου χρόνου χρώσης, αλλά επιτυγχάνεται ελαφρώς χαμηλότερη ευαισθησία όταν αυτή χρησιμοποιείται [2]. Η χρώση Leishman διαρκεί το μισό χρόνο από τη χρώση Giemsa.
3. **Ερμηνεία δεδομένων:** Σε παχιά επιχρίσματα η πυκνότητα των παρασίτων προσδιορίζεται σε παράσιτα ανά μl αίματος, προσδιορίζοντας τον αριθμό των παρασίτων $\times 8000$, διαιρούμενο με τον αριθμό των παρασίτων ανά μl αίματος. των λευκών αιμοσφαιρίων. Στα λεπτά επιχρίσματα ο αριθμός των μολυσμένων και μη μολυσμένων ερυθρών αιμοσφαιρίων και η παρασαιμία εκφράζεται ως ποσοστό επί του συνόλου των μολυσμένων κυττάρων. Προσδιορίζεται επίσης το είδος και το στάδιο των παρασίτων.

Όταν ακολουθείται η συνιστώμενη από τον ΠΟΥ διαδικασία, το θεωρητικό όριο ανίχνευσης είναι 5 παράσιτα / μl αίματος, γεγονός που καθιστά την εξέταση πολύ ευαίσθητη. Αυτό, μαζί με τη δυνατότητα να λαμβάνονται πληροφορίες για το είδος και το στάδιο ζωής, είναι τα κύρια πλεονεκτήματα αυτής της διαγνωστικής μεθόδου. Ωστόσο, η προετοιμασία των χρωματισμένων αντικειμενοφόρων πιακών αίματος και η ερμηνεία των δεδομένων είναι επίπονη, εντατική και απαιτούν εξειδικευμένους εμπειρογνώμονες [3]. Στην πράξη, το θεωρητικό όριο ανίχνευσης σπάνια επιτυγχάνεται. Ένας μη ειδικός

μικροσκοπιστής μπορεί να διαπράξει πολλά λάθη κατά την ερμηνεία των δεδομένων. Η λανθασμένη ταυτοποίηση των ειδών των παρασίτων είναι συχνή, καθώς και η σοβαρή υποεκτίμηση των του αριθμού των παρασίτων [4]. Οι καλύτερες εκτιμήσεις για την ευαισθησία και την ειδικότητα της μικροσκοπικής διάγνωσης στα περιφερειακά νοσοκομεία και στα γενικά εργαστήρια των κέντρων υγείας στις υποσαχάριες χώρες είναι μόνο 82% και 85% αντίστοιχα [5].

2.1.2 Μικροσκοπία Φθορισμού

Ως εναλλακτική λύση στη οπτική μικροσκοπία, έχουν προταθεί διάφορες διαγνωστικές μέθοδοι που χρησιμοποιούν τη μικροσκοπία φθορισμού. Σε αυτές τις μεθόδους, ένα φθοριόχρωμα που προσκολλάται στα νουκλεϊκά οξέα του παρασίτου της ελονοσίας χρησιμοποιείται για τη χρώση ενός δείγματος αίματος [6]. Στη συνέχεια, το δείγμα αίματος εξετάζεται σε μικροσκόπιο φθορισμού, το οποίο εκπέμπει φως μόνο στο μήκος κύματος διέγερσης του φθοριοχρώματος. Το πλεονέκτημα της χρήσης της μικροσκοπίας φθορισμού έναντι της τυπικής μικροσκοπίας φωτός είναι ότι ο μικροσκοπιστής μπορεί εύκολα να εντοπίσει τα παράσιτα που εκπέμπουν φως και ότι η ερμηνεία μπορεί έτσι να γίνει ταχύτερα [7]. Το θεωρητικό όριο ανίχνευσης αυτής της τεχνικής είναι περίπου το ίδιο με εκείνο της φωτεινής μικροσκοπίας, αλλά σε αρκετά πειράματα πεδίου αποδείχθηκε ότι για χαμηλή πυκνότητα παρασιταίμιας υπερτερεί της φωτεινής μικροσκοπίας όσον αφορά την ευαισθησία. Όταν ο ασθενής είναι μολυσμένος με *P. falciparum*, η δοκιμασία αυτή προσφέρει εξαιρετική ειδικότητα (93%), αλλά για άλλα είδη, η ειδικότητα βρέθηκε πολύ χαμηλή σε σύγκριση με την οπτική μικροσκοπία (52%). Επιπλέον, η μέθοδος αυτή είναι τεχνικά απαιτητική και δεν είναι δυνατόν να προσδιοριστούν με ακρίβεια τα είδη των παρασίτων με αυτή τη δοκιμή.

2.1.3 Τέστ ταχείας διάγνωσης

Με σκοπό να ελαττωθεί η διαγνωστική πολυπλοκότητα των μικροσκοπικών μεθόδων ανίχνευσης της ελονοσίας, δημιουργήθηκαν τεστ ταχείας διάγνωσης. Το τεστ αυτό περιλαμβάνει μια μεμβράνη που περιέχει δύο μονοκλωνικά αντισώματα: ένα ειδικό για την πρωτεΐνη HRP-II στην περιοχή "Pf" και ένα ειδικό για την πλασμοδιακή γαλακτική δευτερογονάση (pLDH) στην περιοχή "Pan". Αν υπάρχουν οι εν λόγω πρωτεΐνες στο δείγμα αίματος, θα δεσμευτούν από τα αντίστοιχα αντισώματα στη μεμβράνη και θα παρατηρηθούν ορατές κόκκινες γραμμές στις αντίστοιχες περιοχές. Η απουσία γραμμών υποδηλώνει αρνητικό αποτέλεσμα. [8]. Το βασικό τους πλεονέκτημα είναι ότι δεν χρειάζεται η παρουσία εξειδικευμένου προσωπικού και σε πολλές περιπτώσεις μπορούν να χρησιμοποιηθούν άμεσα από τον ίδιο τον ασθενή καθώς και ο χρόνος παραγωγής αποτελεσμάτων που κοιμνεται στα 30 λεπτα [9].

Ωστόσο αυτά τα τεστ αποτελούν συμπληρωματική μέθοδο ως προς τη μέθοδο αναφοράς (μικροσκοπία σε επίχρισμα αίματος) και έχουν χαμηλή ειδικότητα και ευαισθησία. Επισημαίνεται δε ότι ένα αρνητικό αποτέλεσμα στη δοκιμή ταχείας διάγνωσης δεν αποκλείει την παρουσία ελονοσίας.

2.1.4 PCR τεστ

Η αλυσιδωτή αντίδραση πολυμεράσης (PCR) είναι μια τεχνική κατά την οποία πολλαπλασιάζεται γρήγορα μια συγκεκριμένη αλυσίδα DNA. Στις διαγνωστικές εξετάσεις ελονοσίας με βάση την PCR, το αίμα ενός ασθενούς εξετάζεται για την παρουσία του DNA ενός είδους πλασμώδιου, χρησιμοποιώντας μια αλυσίδα του παρασιτικού RNA ως εκκινητή. Αυτή η διαγνωστική μέθοδος έχει αποδειχθεί ότι είναι πιο ακριβής από

άλλες, τόσο σε εργαστηριακές συνθήκες όσο και στις επιβαρυνμένες από τον ιο περιοχές. Έχει εξαιρετική ευαισθησία και ειδικότητα καθώς τα ψευδώς θετικά αποτελέσματα μπορούν να αποκλειστούν όταν χρησιμοποιούνται δύο PCR. Μπορεί επίσης να χρησιμοποιηθεί για την ακριβή ανίχνευση μικτών λοιμώξεων και παρασίτων ανθεκτικών στα φάρμακα.

Η χρησιμότητα αυτής της τεχνικής, ιδίως σε περιβάλλοντα όπου η πρόσβαση σε εργαστηριακές εγκαταστάσεις είναι περιορισμένη, περιορίζεται ωστόσο από την πολυπλοκότητα της μεθοδολογίας, την ανάγκη για υψηλά εξειδικευμένου εξοπλισμού και προσωπικού, καθώς και η χρονική υστέρηση μεταξύ της συλλογής του δείγματος και της διάγνωσης που σε διάφορες μελέτες αγγίζει έως και εβδομάδες μετά τη λήψη του δείγματος αίματος, γεγονός που προφανώς καθιστά τα αποτελέσματα άχρηστα.

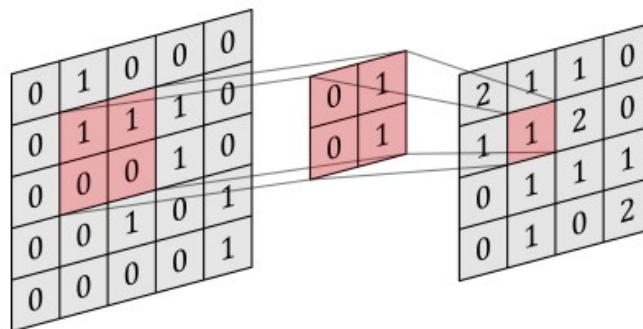
2.2 Μηχανική Μάθηση

Σε αυτό το υποκεφάλαιο θα γίνει αναφορά σε τεχνικές και αρχιτεκτονικές μηχανικής μάθησης που θα χρησιμοποιηθούν εκτενώς στα επόμενα τμήματα της εργασίας.

2.2.1 Βαθιά νευρωνικά δίκτυα με συνελεκτικά μπλοκ

Τα συνελεκτικά νευρωνικά δίκτυα (CNN) αποτελούν μια υποπεριοχή της μηχανικής μάθησης και εφαρμόζονται με επιτυχία στην ανάλυση και επεξεργασία εικόνων. Σε αντίθεση με τα πλήρως συνδεδεμένα δίκτυα όπου κάθε νευρώνας εξαρτάται από όλους τους νευρώνες του προηγούμενου επιπέδου, τα CNN εξαρτώνται μόνο από μια υποπεριοχή του προηγούμενου επιπέδου που προκύπτει από πράξεις συνελέξεων και ονομάζεται δεκτικό πεδίο (receptive field). Στο πλαίσιο των CNN, η συνέλεξη εμπλέκει μια διδιάστατη είσοδο, έναν πυρήνα ή αλλιώς φίλτρο, και την έξοδο. Η συνέλεξη μεταξύ μιας διδιάστατης εισόδου I και ενός διδιάστατου πυρήνα K υπολογίζεται από την εξίσωση:

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



Σχήμα 2.1: Γραφική αναπαράσταση της συνέλιξης μεταξύ διδιάστατης εισόδου και πυρήνα :

Στην περίπτωση της εικόνας, η οποία θεωρείται ως τριδιάστατος πίνακας λόγω των τριών καναλιών κόκκινου, πράσινου και μπλε (RGB), η συνέλιξη ονομάζεται πολλαπλών καναλιών (multi-channel) και υπολογίζεται ως εξής :

Το βασικό δομικό στοιχείο του εν λόγω δικτύου είναι το συνελεκτικό επίπεδο, το οποίο

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

αποτελείται από την είσοδο, έναν επιλεγόμενο αριθμό φίλτρων και μια έξοδο. Το βάθος της εξόδου ισούται με τον αριθμό των φίλτρων και υπολογίζεται από τη συνέλιξη του κάθε φίλτρου με την είσοδο. Η αρχική είσοδος του δικτύου, η οποία είναι η εικόνα, έχει ως βάθος τα κανάλια της εικόνας, συνήθως 3 σε περίπτωση εικόνας RGB. Ωστόσο, σε τυχαία είσοδο συνελκτικού επιπέδου εντός του δικτύου, ο αριθμός είναι διαφορετικός ανάλογα με τον αριθμό των φίλτρων που ορίζει η αρχιτεκτονική. Η έξοδος συνελκτικού επιπέδου απαντάται στη βιβλιογραφία και ως "χάρτης χαρακτηριστικών" υπό την έννοια ότι εξάγει διαφορετικά χαρακτηριστικά της εικόνας (ακμές, υφή) με βάση το μέγεθος και τις τιμές του κάθε φίλτρου. Όταν ένας χάρτης χαρακτηριστικών ακολουθείται από συνάρτηση ενεργοποίησης, τότε ονομάζεται "χάρτης ενεργοποίησης".

Η σχεδιαστική ιδιαιτερότητα ενός βαθιού συνελκτικού δικτύου είναι ότι αποτελείται από πολλά επίπεδα. Το βάθος των δικτύων προσδιορίζεται από το πλήθος των επιπέδων. Η προσθήκη περισσότερων και μεγαλύτερων (με περισσότερους νευρώνες) επιπέδων σε ένα νευρωνικό δίκτυο φαίνεται στην βιβλιογραφία ότι μπορεί να αναπαραστήσει πολύπλοκες συναρτήσεις και παίζει σημαντικό ρόλο σε εργασίες υπολογιστικής όρασης. [10], [11]

Τα σύγχρονα συνελκτικά δίκτυα, τα οποία παρουσιάζουν τα κορυφαία αποτελέσματα σε διάφορα προβλήματα ταξινόμησης ή ανίχνευσης, βασίζονται κατά κανόνα σε συνελκτικά μπλοκ όπως DenseBlock στην περίπτωση του DenseNet [12], MbConvBlock στην περίπτωση του MobileNet [13] κ.α. Πρόκειται ουσιαστικά για ένα επίπεδο το οποίο εσωτερικά απαρτίζεται από πολλαπλά συνελκτικά επίπεδα είτε ίδιου είτε διαφορετικού πυρήνα (multi-kernel blocks). Αυτή η μέθοδος υλοποίησης του δικτύου βοηθά στον αποτελεσματικότερο σχεδιασμό των διαδρομών των συνδέσεων εντός του δικτύου, για παράδειγμα μέσω των skip συνδέσεων όπως στο ResNet. Επιπλέον, σε κάθε επίπεδο εξάγονται πιο πλούσια και ποικίλες χαρακτηριστικά, αυξάνοντας την ακρίβεια των υπολογισμών τόσο σε προβλήματα ανίχνευσης όσο και ταξινόμησης.

2.2.2 Μηχανισμός προσοχής καναλιών

Τα συνελκτικά φίλτρα είναι υπεύθυνα για την κατασκευή των χαρτών χαρακτηριστικών με βάση τα μαθημένα βάρη εντός των εν λόγω φίλτρων. Ενώ ορισμένα φίλτρα μαθαίνουν ακμές, άλλα μαθαίνουν υφές, και συλλογικά μαθαίνουν διαφορετικές αναπαραστάσεις χαρακτηριστικών της πληροφορίας μέσα στην εικόνα που ενσωματώνεται από τον τανυστή εισόδου. Με γνώμονα το γεγονός ότι κάποιοι χαρτες χαρακτηριστικών είναι πιο σημαντικοί από άλλους, θα ήταν λογικό να κατανεμηθεί στους "πιο σημαντικούς" χαρτες χαρακτηριστικών υψηλότερος βαθμός σπουδαιότητας σε σύγκριση με τους υπόλοιπους.

Squeeze and Excitation Module (SE)

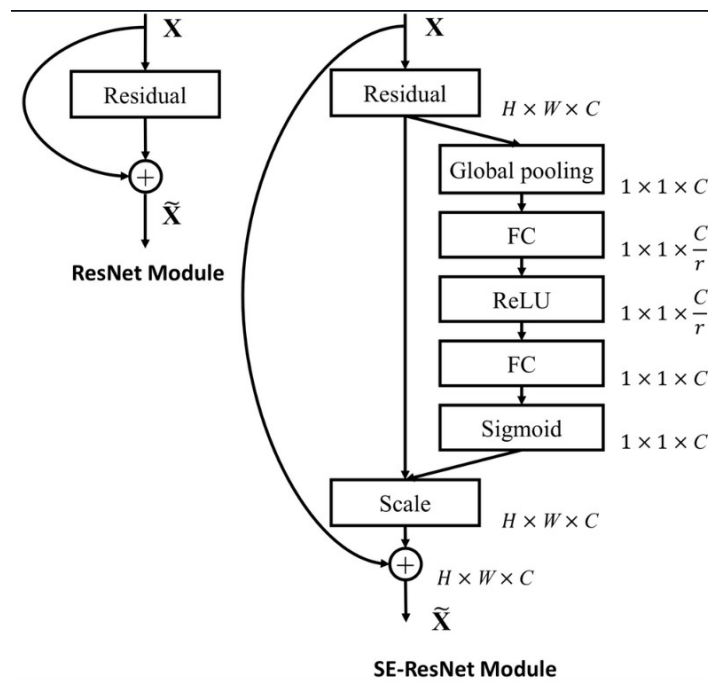
Οι Hu et al. (2018) [14] προτείνουν τον μηχανισμό Squeeze and Excitation με βασική ιδέα την προσθήκη παραμέτρων στο κανάλι κάθε συνελκτικού μπλοκ, έτσι ώστε το δίκτυο να μπορεί να τροποποιεί προσαρμοστικά το βάρος κάθε χάρτη χαρακτηριστικών χωρίς ωστόσο να επιβαρύνεται υπολογιστικά το συνολικό δίκτυο.

Το επίπεδο "**Squeeze**" του μπλοκ SE δέχεται ως είσοδο τον τανυστή μεγεθούς $C \times W \times H$ και συμπιέζει τις συνολικές χωρικές πληροφορίες σε μια περιγραφή καναλιού. Για την παροχή στατιστικών στοιχείων ανά κανάλι, το στάδιο συμπίεσης αποτελείται από ένα

επιπεδο GlobalAveragePooling στις χωρικές διαστάσεις $H \times W$ και εξάγει ένα διάνυσμα μεγέθους $C \times 1 \times 1$.

Το επίπεδο **Excitation** μετατρέπει την διανυσματική έξοδο του επιπέδου συμπίεσης σε ένα διάνυσμα ενεργοποιήσεων μέσω σιγμοειδούς συνάρτησης, το οποίο στη συνέχεια πολλαπλασιάζεται με την αρχική είσοδο και χρησιμοποιείται για την κλιμάκωση των χαρτών χαρακτηριστικών.

Το εν λόγω δίκτυο μπορεί να χρησιμοποιηθεί σε κάθε αρχιτεκτονική αυξάνοντας την ακρίβεια χωρίς να προσθέτει περαιτέρω πολυπλοκότητα στο δίκτυο. Συγκεκριμένα στην περίπτωση του ResNet-50 στο σύνολο δεδομένων ImageNet η χρήση των μπλοκ SE αυξάνει την ακρίβεια κατά 1.1% χωρίς να επηρεάζει τα συνολικά GFLOPS του δικτύου.

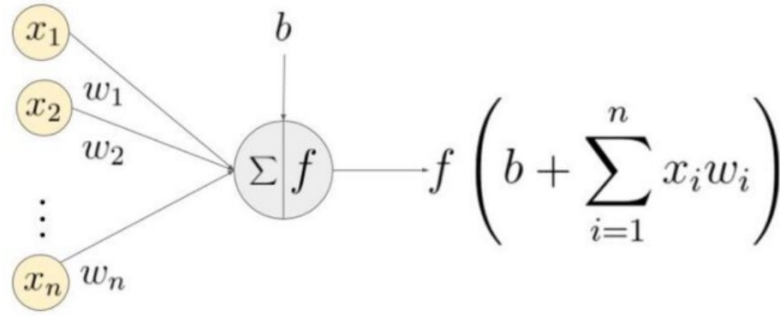


Σχήμα 2.2: Ο μηχανισμός SE εντός της αρχιτεκτονικής ResNET.

2.2.3 Forward and Back Propagation

Διάδοση προς τα εμπρός (Forward Propagation)

Η βασική δομή ενός νευρωνικού δικτύου είναι ένα επίπεδο εισόδου, ένα ή περισσότερα κρυφά επίπεδα και ένα μόνο επίπεδο εξόδου. Τα βάρη του δικτύου αρχικοποιούνται τυχαία κατά τη διάρκεια της προς τα εμπρός διάδοσης. Τα χαρακτηριστικά εισόδου πολλαπλασιάζονται με τα αντίστοιχα βάρη σε κάθε κόμβο του κρυφού επιπέδου και μια "προκατάληψη" (bias) προστίθεται σε κάθε κόμβο. Αυτή η τιμή μετατρέπεται στη συνέχεια στην έξοδο του κόμβου χρησιμοποιώντας μια συνάρτηση ενεργοποίησης. Για τη δημιουργία της εξόδου του νευρωνικού δικτύου, η έξοδος του κρυφού επιπέδου πολλαπλασιάζεται με τις τιμές των βαρών συν τις τιμές της προκατάληψης και το σύνολο μετατρέπεται χρησιμοποιώντας μια άλλη συνάρτηση ενεργοποίησης. Αυτή θα είναι η προβλεπόμενη τιμή του νευρωνικού δικτύου για μια δεδομένη τιμή εισόδου.



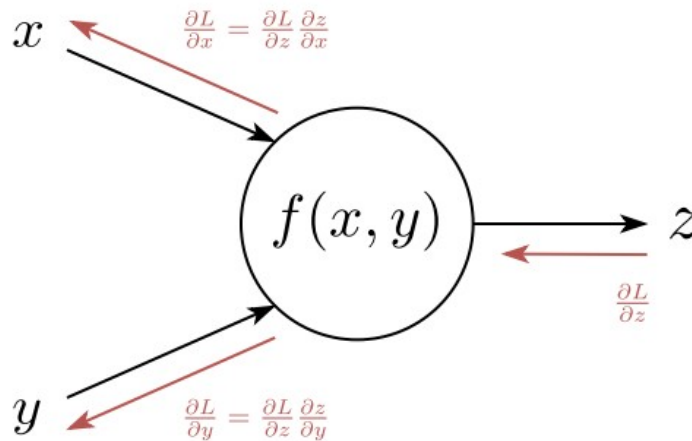
Σχήμα 2.3: Γραφική αναπαράσταση της προς τα εμπρός διαδοσης στην έξοδο ενός νευρώνα.

Οπισθοδιάδοση (Back Propagation)

Καθώς το δίκτυο παράγει μια έξοδο, η συνάρτηση απώλειας (C) δείχνει πόσο καλά προέβλεψε την έξοδο. Το δίκτυο εκτελεί αντίστροφη διάδοση για να ελαχιστοποιήσει την απώλεια. Η μέθοδος οπισθοδιάδοσης ελαχιστοποιεί τη συνάρτηση απώλειας προσαρμόζοντας τα βάρη και τις προκαταλήψεις του νευρωνικού δικτύου. Σε αυτή τη μέθοδο, η κλίση της συνάρτησης απώλειας υπολογίζεται σε σχέση με κάθε βάρος στο δίκτυο. Στην αντίστροφη διάδοση, το νέο βάρος (W_{new}) ενός κόμβου υπολογίζεται χρησιμοποιώντας το παλιό βάρος (W_{old}) και το γινόμενο του ρυθμού μάθησης (η) και της κλίσης της συνάρτησης απώλειας ($\frac{\partial C}{\partial w}$).

$$(W_{new} = W_{old} - \eta * \frac{\partial C}{\partial w})$$

Με τον αλυσιδωτό κανόνα των μερικών παραγώγων, μπορεί να αναπαρασταθεί η κλίση της συνάρτησης απώλειας ως γινόμενο των κλίσεων όλων των συναρτήσεων ενεργοποίησης των κόμβων ως προς τα βάρη τους. Επομένως, τα ενημερωμένα βάρη των κόμβων του δικτύου εξαρτώνται από τις κλίσεις των συναρτήσεων ενεργοποίησης κάθε κόμβου.



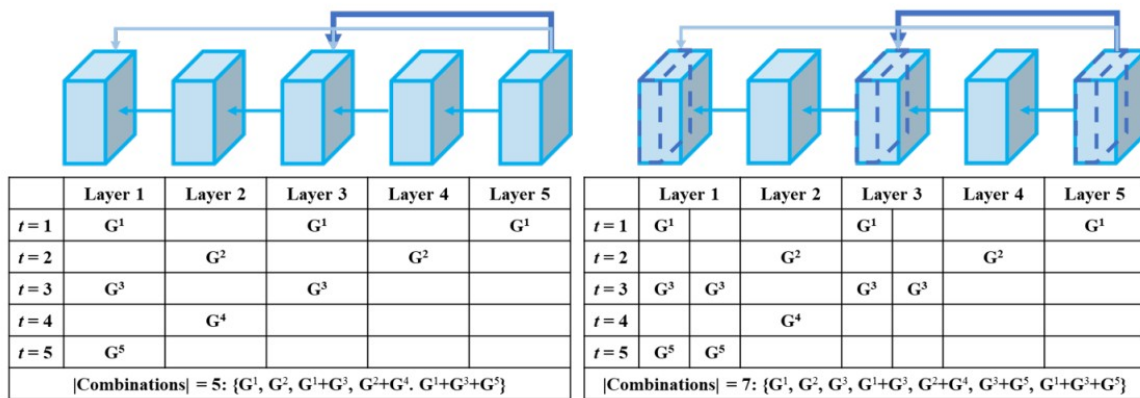
Σχήμα 2.4: Γραφική αναπαράσταση της οπισθοδιάδοσης από την έξοδο ενός νευρώνα στις εισόδους ή τις παραμέτρους του, με βάση τον κανόνα της αλυσίδας.

2.2.4 Φαινόμενο εξαφανιζόμενων κλίσεων και αντιγραφόμενων κλίσεων

Το φαινόμενο των εξαφανιζόμενων κλίσεων εμφανίζεται όταν τα στοιχεία της κλίσης, δηλαδή οι μερικές παράγωγοι ως προς τις παραμέτρους του νευρωνικού δικτύου, γίνονται εκθετικά μικρές κατά την οπισθοδιάδοση. Αυτό έχει ως αποτέλεσμα η ενημέρωση των παραμέτρων χρησιμοποιώντας τις κλίσεις του δικτύου να γίνεται σχεδόν ασήμαντη και, κατ' επέκταση, να δυσκολεύει την εκπαίδευσή του δικτύου. Αυτό συμβαίνει διότι οι κλίσεις υπολογίζονται μέσω του κανόνα της αλυσίδας, ο οποίος υπονοεί πολλαπλασιασμό κλίσεων, και είναι φανερό ότι, πολλαπλασιάζοντας μικρές τιμές, το αποτέλεσμα μπορεί να πλησιάσει το μηδέν. Ειδικότερα με χρήση κορεσμένων συναρτήσεων με άνω φραγμένη πρώτη παράγωγο όπως η υπερβολική εφαιπτομένη ή η σιγμοειδής το πρόβλημα γίνεται ακόμη εντονότερο. Το πρόβλημα αυτό επιλύεται με χρήση παρακάμψεων για την διαδοχή των κλίσεων τύπου ResNet [15], χρήση άλλων συναρτήσεων ενεργοποίησης, όπως RELU, LeakyRelu, Swish κ.α καθώς και με αποδοτικότερη αρχικοποίηση των βαρών παραδειγματος χάρη με gloriot initialization [16].

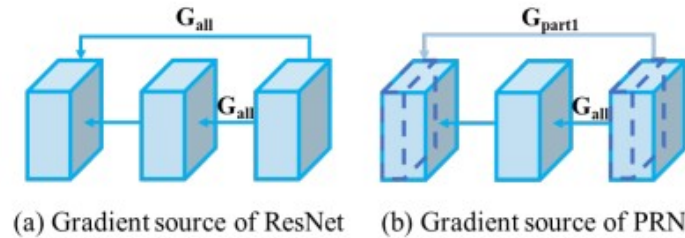
Gradient Timestamp και Gradient Source

Για να αναλυθεί η πορεία της ροής των κλίσεων εντός ενός δικτύου χρειάζονται δύο παράμετροι. Αυτοί είναι το Gradient Timestamp, δηλαδή ο 'χρόνος' που χρειάζεται μια κλίση για να φτάσει στο εκάστοτε επίπεδο ξεκινώντας από την έξοδο μέσω του κανόνα της αλυσίδας. Αν αναλυθεί ένα μεμονωμένο timestamp, τότε οι συγκεκριμένες πορείες των κλίσεων, δηλαδή από πού-εώς πού, ονομάζεται Gradient Source.



Σχήμα 2.5: Στο αριστερό γράφημα τα timestamp ενός ResNet και δεξιά τα timestamp ενός Partial-ResNet

Την στιγμή $\text{time}=1$ υπολογίζοντας το σφάλμα στο επίπεδο 5 (Layer 5) η κλίση διαδίδεται άμεσα μέσω των παρακάμψεων στα επίπεδα 1 και 3. Ο υπόλοιπος πίνακας παράγεται μέσω του κανόνα της αλυσίδας. Να τονιστεί ότι ο συμβολισμός G^i υπονοεί την κλίση που διαδόθηκε στο timestamp i . Παρακάτω θα παρουσιαστούν πηγές των κλίσεων στο timestamp $\text{time}=1$.



Σχήμα 2.6: Στο αριστερό γράφημα τα Sources ενός ResNet και δεξιά τα Sources ενός Partial-ResNet στο timestamp ,time =1

Από τα γραφήματα timestamp και sources παρατηρείται ότι διαφορετικά επίπεδα του ResNet μοιράζονται κλίσεις ίδιου timestamp και ίδιου gradient source. Σε πολύ βαθιά δίκτυα μορφής ResNet είναι πιθανό το δίκτυο να αρχίσει να μαθαίνει τις ίδιες, δυνητικά άχρηστες πληροφορίες και αυτό να αποτελέσει εμπόδιο για την πορεία της εκπαίδευσης. Το πρόβλημα αυτό ονομάζεται φαινόμενο αντιγραφόμενων κλίσεων.

2.2.5 Συναρτήσεις Ενεργοποίησης

Οι συναρτήσεις ενεργοποίησης επιτελούν αναπόσπαστο ρόλο στα νευρωνικά δίκτυα εισάγοντας μη γραμμικότητα. Αυτή η μη γραμμικότητα επιτρέπει στα νευρωνικά δίκτυα να αναπτύσσουν πολύπλοκες αναπαραστάσεις με βάση τις εισόδους, οι οποίες δεν θα ήταν δυνατές με ένα απλό μοντέλο γραμμικής παλινδρόμησης. Στην ιστορία των νευρωνικών δικτύων έχουν προταθεί πολλές διαφορετικές μη γραμμικές συναρτήσεις ενεργοποίησης και οι πιο συνηθισμένες που συναντώνται στην βιβλιογραφία είναι οι εξής

ReLU

Η συνάρτηση αυτή είναι μηδενική αριστερά από μία τιμή κατωφλίου, και δεξιά από αυτό είναι γραμμική. Η συνάρτηση αυτή συναντάται συχνά στα τεχνητά νευρωνικά δίκτυα, καθώς αντιμετωπίζει τα προβλήματα κορεσμού που παρουσιάζουν κάποιες άλλες συναρτήσεις, όπως οι σιγμοειδείς, όπου οι αρνητικές τιμές στην είσοδο μεταφράζονται σε μηδενική τιμή εξόδου, ενώ οι πολύ μεγάλες τιμές εισόδου δεν δίνουν ποτέ τιμή μεγαλύτερη της μονάδας στην έξοδο.

$$f(x) = x^+ = \max(0, x)$$

Σιγμοειδής Συναρτηση

Η σιγμοειδής συνάρτηση δίνει στην έξοδό της τιμές που κυμαίνονται στο διάστημα $[0, 1]$, και είναι συνεχής και παραγωγίσιμη σε όλο το διάστημά της. Βρίσκει χρήση στην έξοδο νευρωνικών δικτύων δυαδικής κατηγοριοποίησης όπου πρέπει να υπολογιστεί πιθανότητα.

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

Συναρτηση SWISH

Η συνάρτηση Swish [17] είναι μία σχετικά νεότερη συνάρτηση μεταφοράς που κατασκευάστηκε από ερευνητές της Google. Πρόκειται για μία μη μονοτονική συνάρτηση και ένας από τους κύριους λόγους που χρησιμοποιείται στα βαθιά νευρωνικά δίκτυα είναι ότι πρόκειται για μία από τις λίγες συναρτήσεις που φαίνεται να δίνουν καλύτερα αποτελέσματα από την συνάρτηση ReLU.

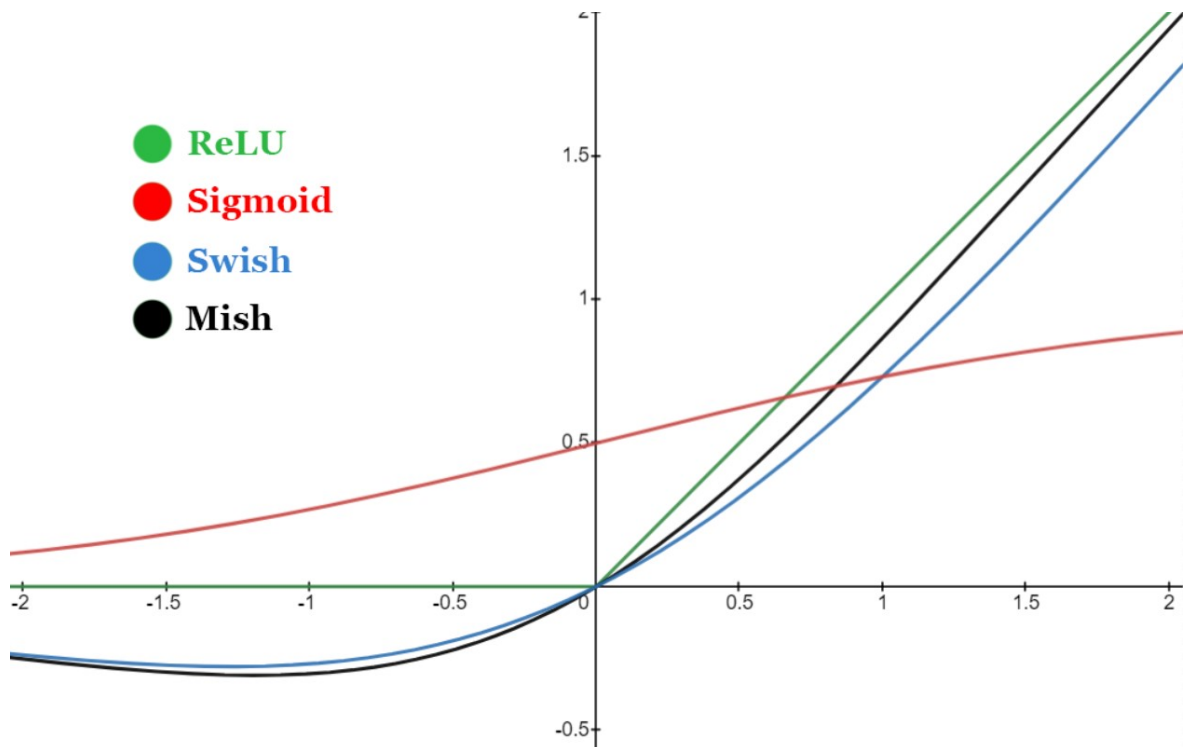
$$f(x) = x \cdot \text{sigmoid}(\beta x)$$

Συναρτηση Mish

Παρομοια συνάρτηση με την swish και σε επίπεδο πρώτων και δευτέρων παραγώγων. Ως επί το πλείστον εμφανίζουν πανομοιοτυπία αποτελεσματα ωστόσο οι ερευνητές υποστηρίζουν ότι σε προβλήματα πολύ βαθιάς μάθησης σε συγκεκριμένα dataset η συνάρτηση mish [18] εμφανίζει ελαφρώς ανώτερα αποτελέσματα.

$$f(x) = x \cdot \tanh \text{softplus}(x)$$

$$\text{softplus}(x) = \ln(1 + e^x)$$



Σχήμα 2.7: Γραφικές παραστάσεις των συναρτήσεων ενεργοποίησης που παρουσιάστηκαν παραπάνω

2.3 Βιβλιογραφική Ανασκόπηση

Η προσφατη βιβλιογραφια στο ζητημα της ανιχνευσης και ταξινομησης του ιου της ελονοσιας εχει στραφει σε τεχνικες βαθιας μαθησης παρα σε τεχνικες κλασσικης επεξεργασιας εικονας. Σε αυτο το κομματι της εργασιας θα παρατεθουν εργασιες και αποτελεσματα ερευνητων στο παρων προβλημα.

Παλαιότερες προσεγγίσεις επιλυανε το προβλημα του κατακερματισμου εικονων με μεθοδους επεξεργασιας εικονας τυπου αλγοριθμος watershed και εφαρμοζαν τεχνικες μηχανικης μαθησης στο κομματι της ταξινομησης των κυτταρων ωστοσο απο το 2017 οι ερευνητες ξεκινησαν να δοκιμαζουν αμιγως συνελικτικα δικτυα .

.Οι Dong et al.(2017) [19] συνέκριναν τις επιδόσεις του SVM και προ-εκπαιδευμένων μοντέλων βαθιάς μάθησης, συμπεριλαμβανομένου του LeNet [20] (LeCun et al., 1998), AlexNet [21] και GoogLeNet [22] προς την ταξινόμηση παρασιτικών και μη μολυσμένων κυττάρων. Οι συγγραφείς μηχανοποίησαν τα ερυθρά αιμοσφαίρια (RBCs) από εικόνες λεπτών επιχρισμάτων αίματος και τα χώρισαν τυχαία σε σύνολα εκπαίδευσης και δοκιμής. Συνολικά 25% των εικόνων εκπαίδευσης επιλέχθηκαν τυχαία για την επικύρωση για την επαλήθευση των μοντέλων. Οι Liang et. al. (2017) [23] πρότειναν ένα CNN 16 επιπέδων προς την κατεύθυνση της ταξινόμησης των μη μολυσμένων και παρασιτισμένων κυττάρων. Τα χαρακτηριστικά εξήχθησαν χρησιμοποιώντας το προ-εκπαιδευμένο AlexNet και ένας SVM ταξινομητής εκπαιδεύτηκε με βάση τα εξαγόμενα χαρακτηριστικά. Η απόδοση του προτεινόμενου μοντέλου συγκρίθηκε με εκείνη του προ-εκπαιδευμένου CNN. Η μελέτη ανέφερε ότι το προσαρμοσμένο μοντέλο ήταν πιο ακριβές, ευαίσθητο και ειδικό από το προ-εκπαιδευμένο μοντέλο. Οι εικόνες δειγματοληπτήθηκαν σε ανάλυση 44×44 πίξελ για να αντισταθμιστεί η έλλειψη υπολογιστικών πόρων. Οι Bibin, Nair & Punitha (2017) [24] πρότειναν ένα deep belief δικτυο 6 επιπέδων προς την ανίχνευση του παράσιτου της ελονοσίας σε εικόνες επιχρίσματος περιφερικού αίματος. Οι συγγραφείς ανέφεραν ακρίβεια 96,4% στην ταξινόμηση ενός συνόλου δεδομένων 4.100 κυττάρων με τυχαία τομή εκπαίδευσης/δοκιμής. Οι Gorakumar et al. (2018) [25] χρησιμοποίησαν ένα προσαρμοσμένο μοντέλο CNN για την ανάλυση βίντεο αντικειμενοφόρων πλακών με χρώση Leishman προς τη διαδικασία της αυτοματοποιημένης ανίχνευσης παρασίτων. Οι συγγραφείς χρησιμοποίησαν έναν προσαρμοσμένο φορητό σαρωτή αντικειμενοφόρων πλακών για την απόκτηση δεδομένων και κατέδειξαν ευαισθησία και ειδικότητα 97,06% και 98,50% αντίστοιχα. Δυσχεήματα της εποχής ήταν η έλλειψη υπολογιστικής δυναμής για την εκπαίδευση και δοκιμή τετοιων μοντελων καθώς και η έλλειψη μεγαλων βασεων δεδομενων ωστε να υπαρχει μια κοινή βάση συγκρίσεως των αποτελεσμάτων. Ερευνητες στο πανεπιστημιο Lister Hill δημιουργησαν μια τετοια βάση χρησιμοποιώντας καμερα κινητου και οι εικονες χρησιμοποιουνται ευρεως και σημερα για την αξιολογηση των εκπαιδευμενων μοντελων. Πολυ σημαντικη ειναι η ερευνα των Rajamaran et al [26], οπου χρησιμοποιώντας ρηχο συνελικτικο δικτυο τριων επιπεδων καταφερε αναλογη ακριβεια με τα δικτυα AlexNet, ResNet-50, GoogleNet της ταξεως του 95%.

Στην πολυ προσφατη βιβλιογραφια παρατηρειται χρηση μονο αλγοριθμων ανιχνευσης αντικειμενων ωστε να συμβαινει παραλληλα η ανιχνευση ερυθροκυτταρων στο επιχρισμα και η ταξινομηση τους ως επιμολυσμενα ή μη. Συγκεκριμενα οι Loh et al.(2021) [27] χρησιμοποισαν τον αλγοριθμο Mask-R-CNN επιτυχανωντας ακριβεια 94.57%. Οι Razin et al.(2022) [28] μεσω προσαρμοσμενου αλγοριθμου YoloV5 επιτυχανουν ακριβεια 96.26% σε βάση δεδομενων επιχρισματος που δημιουργησαν οι ιδιοι. Τελος οι Wojciech Silka et al. (2023) [29] μεσω προσαρμοσμενου αλγοριθμου μεγαλου βαθους που κατασκευασαν οι ιδιοι (Seg-Sem CNN) επιτυχανουν 99.67% ακριβεια στο Malaria Bounding Boxes Dataset του Kaggle.com.

Κεφάλαιο 3

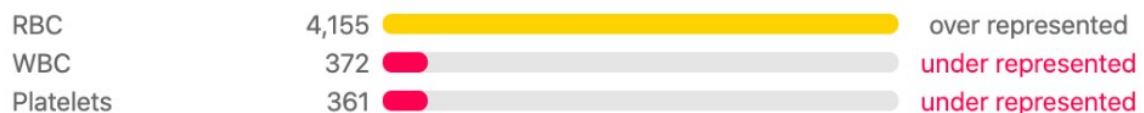
Σύνολα Δεδομένων και χρήση κλασσικής επεξεργασίας εικόνας

3.1 Σύνολα δεδομένων

Σε αυτό το κεφάλαιο θα εξεταστούν τα διαφορετικά σύνολα δεδομένων που θα χρησιμοποιηθούν κατά τη διαδικασία της εκπαίδευσης των νευρωνικών δικτύων τόσο της ανίχνευσης όσο και της κατηγοριοποίησης των επιμολυσμένων κυττάρων

3.1.1 BCCD-Dataset

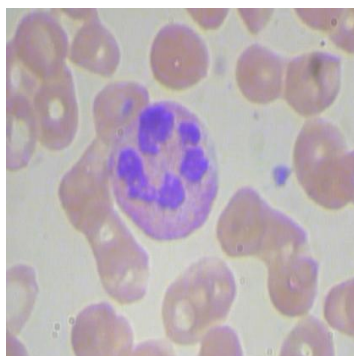
Πρόκειται για ένα σύνολο 364 εικόνων επιχρίσματος αίματος μικρής κλίμακας 416x416 πίξελ όπου η εκάστοτε εικόνα αποτελείται από ερυθρά αιμοσφαίρια, λευκά αιμοσφαίρια και αιμοπετάλια. Κάθε εικόνα συνοδεύεται από ένα αντίστοιχο αρχείο κειμένου, στο οποίο βρίσκονται οι συντεταγμένες οριοθετημένων πλαισίων μέσα στα οποία βρίσκονται τα επιμερους κυττάρων.



Σχήμα 3.1: Στο BCCD-Dataset παρατηρείται έντονη ανισορροπία μεταξύ των κλάσεων, κάτι που δεν δημιουργεί πρόβλημα καθώς στα πλαίσια της εργασίας ανιχνεύονται μόνο τα ερυθρά αιμοσφαίρια.

Αυτό το σύνολο εικόνων θα χρησιμοποιηθεί για εκπαίδευση αλγορίθμων ανίχνευσης αντικειμένων με σκοπό να διαχωριστούν μόνο τα ερυθροκύτταρα από την υπολοιπή εικόνα. Το 90% των εικόνων θα χρησιμοποιηθεί για εκπαίδευση, το 10% για επαλήθευση της πορείας της εκπαίδευσης και το τελευταίο 10% για αξιολόγηση του αλγορίθμου και εξαγωγή μετρικών και αποτελεσμάτων.

Αυτό το σύνολο εικόνων επιλέχτηκε διότι τα κυττάρων εμφανίζουν έντονη επικάλυψη μεταξύ τους και έτσι μπορεί να διερευνηθεί η δυνατότητα του εκάστοτε αλγορίθμου να ανιχνεύει επικαλυπτόμενα ερυθροκύτταρα.



Σχήμα 3.2: Παραδειγμα εικονας επιχρισματος αιματος απο το BCCD Dataset

3.1.2 Archive-Dataset

Αυτο το συνολο εικονων περιεχει 100 εικονες επιχρισματος αιματος μικρης κλιμακας 256x256 πιξελ οπου η καθε εικονα αποτελειται απο ερυθρα και λευκα αιμοσφαιρια. Συνοδευεται με ενα αρχειο excel στο οποιο εμπεριεχονται οι συντεταγμενες οριοθετημενων πλαισιων των δυο κλασσεων κυτταρων. Για τις αναγκες της εκπαίδευσης θα γίνει μια τομή 90/10/10 των δεδομένων. Το συνολο επιλεχθηκε καθως τα κυτταρα παρουσιαζουν μικροτερη επικαλυψη μεταξυ τους σε σχεση με το BCCD-Dataset και αποτελει ενα διαφορετικο σημειο αναφορας για τους αλγοριθμους ανιχνευσης.

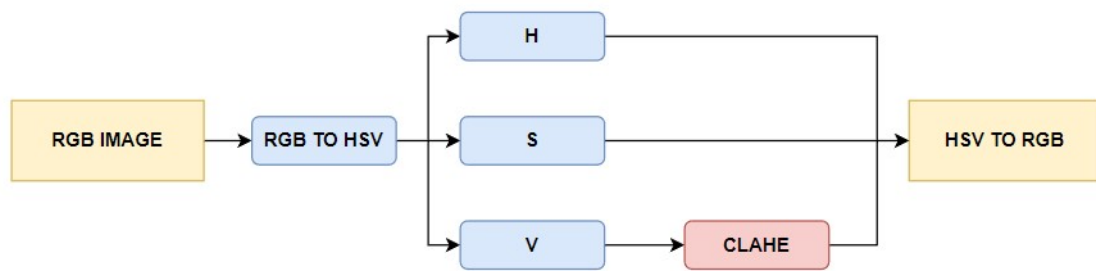
3.1.3 NIH-Dataset

Με σκοπό να διερευνήσουν τεχνικές βαθιάς μάθησης στον εντοπισμό της ελονοσίας σε χώρες τρίτου κόσμου, ερευνητές στο Εθνικό κέντρο Lister Hill δημιούργησαν το σύνολο εικόνων NIH. Αποτελείται από επιχρίσματα αίματος 150 επιμολυσμένων ασθενών και 50 υγιών όπου φωτογραφήθηκαν με κάμερα κινητού και επειτά περικοπήκαν στα οριοθετημένα πλαίσια των κυττάρων. Το τελικό σύνολο απαρτίζεται από δύο κλάσεις μεμονωμένων πλέον κυττάρων που διαχωρίζονται σε υγιή και επιμολυσμένα από τον ιό της ελονοσίας.

Οι εικονες αυτες θα χρησιμοποιηθουν αμιγως για την εκπαιδευση των δικτυων ταξινομισης ερυθροκυτταρων

3.1.4 Custom-NIH-Dataset

Οι εικόνες υγιών ερυθροκυττάρων στο NIH-Dataset παρουσιάζουν πολύ χαμηλά επίπεδα αντίθεσης χρωμάτων και υψηλά επίπεδα θολότητας. Αυτό δύναται να δημιουργήσει προβλήματα στην γενίκευση του αλγορίθμου ταξινόμησης σε περαιτέρω εικόνες που δεν μοιάζουν με αυτές στις οποίες έχει εκπαιδευθεί. Για αυτό το λόγο θα γίνει μια προεπεξεργασία των εικόνων με σκοπό την αύξηση των λεπτομερειών της εικόνας και τη βελτίωση της αντίθεσής τους. Αυτό θα αποτελέσει ένα νέο σύνολο δεδομένων στο οποίο θα εμπεριέχονται και οι αρχικές εικόνες, καθώς και οι προεπεξεργασμένες.



Σχήμα 3.3: Στο παραπάνω σχήμα παρουσιάζεται η διαδικασία προεπεξεργασίας των εικόνων

3.1.5 Μοντέλο χρώματος HSV

HUE

Η απόχρωση - hue (H) ενός χρώματος είναι μια ιδιότητα που περιγράφει πόσο κοντά είναι ένα χρώμα σε ένα καθαρό χρώμα όπως μπλε, κοκκίνο. Οι αποχρώσεις μπορούν να αναπαρασταθούν με έναν αριθμό που καθορίζει τη θέση του αντίστοιχου καθαρού χρώματος στον τροχό χρώματος.

Saturation

Ο κορεσμός - saturation (S) περιγράφει πόσο κορεσμένο (ή εντονο) είναι το χρώμα. Ένα καθαρό κόκκινο είναι πλήρως κορεσμένο, με κορεσμό 1, οι αποχρώσεις του κόκκινου έχουν κορεσμούς μικρότερους από 1 και το λευκό έχει κορεσμό 0.

Value

Η τιμή - value (V) ενός χρώματος, που ονομάζεται επίσης φωτεινότητα- lightness, περιγράφει πόσο σκοτεινό ή φωτεινό είναι ένα χρώμα. Η τιμή 0 αντιστοιχεί στο μαύρο, με αυξανόμενη φωτεινότητα που απομακρύνεται από το μαύρο. Η τιμή 100 αντιστοιχεί στο λευκό.

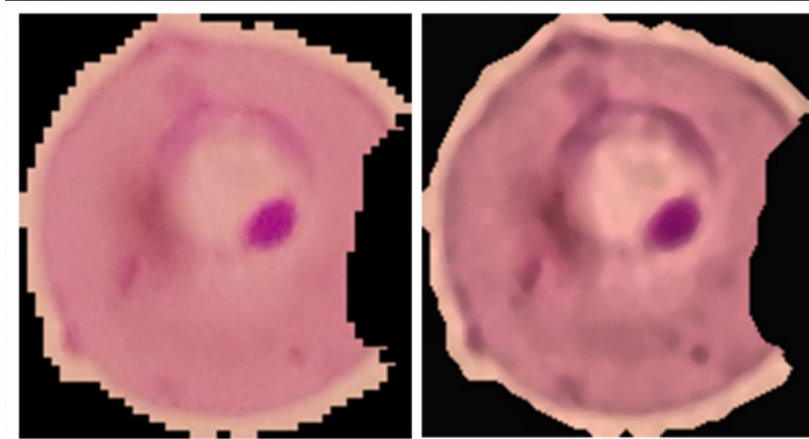
3.1.6 Contrast Limited Adaptive Histogram Equalization (C.L.A.H.E)

Σε μια ψηφιακή εικόνα τα ιστογράμματα επεξηγούν, με μορφή γραφήματος, τη φωτεινότητα και τα χαρακτηριστικά αντίθεσης μιας εικόνας, δηλαδή αν υπάρχουν και πόσα pixel με κάποια δεδομένη τιμή έντασης. Σε ένα ιστόγραμμα, ο X άξονας αντιπροσωπεύει την κλίμακα έντασης (0 έως 255 σε ένα σύστημα των 8 bit). Ο Y άξονας μετρά τον αριθμό από pixel στην εικόνα που έχουν μια ορισμένη τιμή έντασης.

Ισοσταθμίζοντας το ιστόγραμμα μιας εικόνας ώστε η χαμηλότερη τιμή έντασης να είναι το μηδέν, η μέγιστη το 255 και οι ενδιάμεσες τιμές να αυξομειώνονται ανάλογα, επιτυγχάνεται αύξηση της αντίθεσης της εικόνας. Καθώς όμως η αύξηση της αντίθεσης είναι γενική και όχι τοπική, ανάλογα αυξάνεται και ο θόρυβος στην εικόνα. Το Contrast Limited AHE (CLAHE) είναι μια παραλλαγή προσαρμοστικής εξισορρόπησης ιστογράμματος που περιορίζει την ενίσχυση αντίθεσης για τη μείωση της ενίσχυσης θορύβου. Το CLAHE εκτελεί εξισορρόπηση ιστογράμματος σε μικρά τμήματα της εικόνας με υψηλή

ακρίβεια και περιορισμό αντίθεσης και κατ' αναλογία τον θορυβο. Προτείνεται δε ότι σε εγχρωμες εικόνες το CLAHE θα πρέπει να υπολογισθεί επάνω στο κανάλι V μιας HSV χρωματικής αναπαράστασης για τα βέλτιστα δυνατά αποτελέσματα.

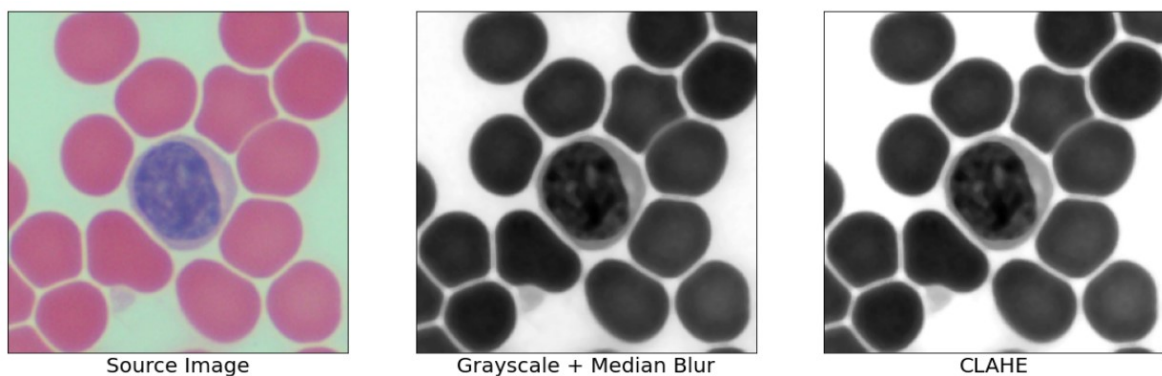
Για τον περαιτέρω περιορισμό του παραγόμενου θορύβου από την ισοσταθμισμό ιστογράμματος χρησιμοποιείται και ένα φίλτρο διάμεσης θολώσης το οποίο λειτουργεί αντικαθιστώντας κάθε τιμή έντασης της εικόνας με την διάμεσο των γειτονικών τιμών της.



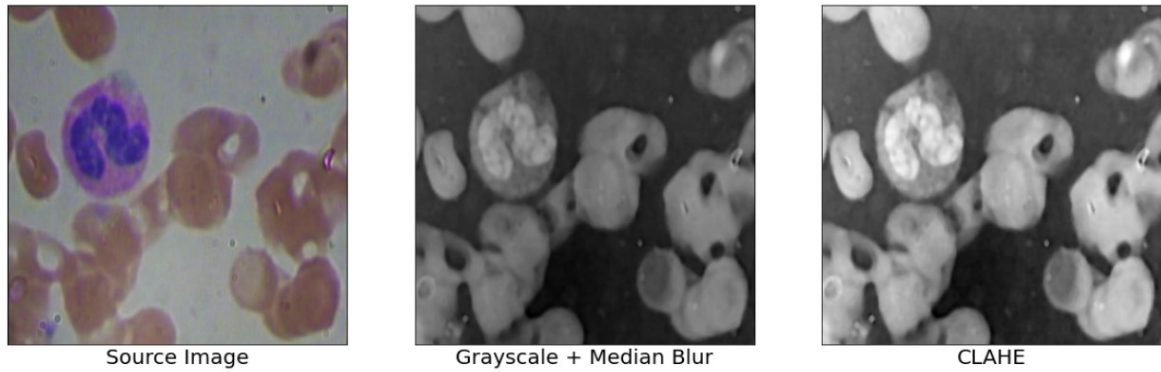
Σχήμα 3.4: Αριστερά ένα υγιές ερυθροκύτταρο του NIH Dataset πριν την προεπεξεργασία και δεξιά μετά την εφαρμογή του αλγορίθμου που περιγράφηκε παραπάνω

3.2 Κατακερματισμός ερυθροκυττάρων μέσω τεχνικών επεξεργασίας εικόνας

Σε αυτό το σημείο θα αναφερθεί συνοπτικά η χρήση του αλγορίθμου watershed με σκοπό τον κατακερματισμό ερυθροκυττάρων από την υπόλοιπη εικόνα του επίχρισμα-τος. Το πρώτο βήμα της διαδικασίας είναι η μετατροπή της RGB εικόνας σε Grayscale, η εφαρμογή ενός φίλτρου θολώσης για την απαλοιφή του θορύβου και η εξισορρόπηση ιστογράμματος. Για τις ανάγκες του προβλήματος θα χρησιμοποιηθεί ένα φίλτρο ενδιάμεσης θολώσης και ο αλγόριθμος CLAHE για την εξισορρόπηση του ιστογράμματος των εικόνων όπως παρουσιάστηκαν στο προηγούμενο υποκεφάλαιο. Ακολουθεί επείτα η δυαδική μετατροπή της εικόνας (thresholding) όπου θα εξεταστεί η μέθοδος κατωφλίωσης Otsu.



Σχήμα 3.5: Προεπεξεργασία τυχαίας εικόνας από το Archive Dataset



Σχήμα 3.6: Προεπεξεργασία τυχαίας εικόνας απο το BCCD Dataset

3.2.1 Otsu's Thresholding

Η μέθοδος του Otsu ακολουθεί πέντε βήματα για την δυαδική μετατροπή της εικόνας

1. **Υπολογισμός Ιστογράμματος** : Το ιστόγραμμα κλίμακας του γκρι μιας εικόνας δείχνει την κατανομή των τιμών έντασης των εικονοστοιχείων σε αυτήν. Το ιστόγραμμα υπολογίζεται μετρώντας τον αριθμό των εικονοστοιχείων σε κάθε επίπεδο έντασης. Ως εκ τούτου, το ιστόγραμμα κλίμακας του γκρι $H(i)$ μιας εικόνας I με επίπεδο έντασης i ορίζεται από την ακόλουθη εξίσωση :

$$H(i) = \sum_{x=1}^M \sum_{y=1}^N [I(x, y) = i]$$

2. **Υπολογισμός συνάρτησης αθροιστικής κατανομής** : Η αθροιστική συνάρτηση κατανομής (CDF) αντιπροσωπεύει την πιθανότητα ένα εικονοστοιχείο στην εικόνα να έχει τιμή έντασης κλίμακας του γκρι μικρότερη ή ίση με ένα συγκεκριμένο επίπεδο. Για τον υπολογισμό της CDF από ένα ιστόγραμμα κλίμακας του γκρι, αθροίζονται οι τιμές του ιστογράμματος μέχρι κάθε επίπεδο έντασης.

$$C(i) = \sum_{j=0}^i \frac{H(j)}{M \times N}$$

3. **Υπολογισμός της μέσης τιμής έντασης της εικόνας** : Η μέση τιμή έντασης της κλίμακας του γκρι της εικόνας είναι η διαδικασία του μέσου όρου των τιμών έντασης της κλίμακας του γκρι όλων των εικονοστοιχείων της εικόνας

$$\mu = \frac{1}{M \times N} \sum_{x=1}^M \sum_{y=1}^N I(x, y)$$

4. **Υπολογισμός της διακύμανσης μεταξύ των κατηγοριών για κάθε πιθανή τιμή κατωφλίου** : Η διακύμανση μεταξύ των κλάσεων μετράει το διαχωρισμό μεταξύ των περιοχών του προσκηνίου και του φόντου. Μπορούμε να υπολογίσουμε τη διακύμανση μεταξύ των κλάσεων ως το γινόμενο του τετραγώνου της διαφοράς

μεταξύ των μέσων τιμών έντασης του γκρι και των πιθανοτήτων των περιοχών (P0 και P1) οι οποίες ορίζονται μέσω της αθροιστικής κατανομής.

$P_0(T) = C(T)$ $P_1(T) = 1 - C(T)$	$m_0(T) = \frac{\sum_{i=0}^{T-1} i \cdot H(i)}{P_0(T) \cdot MN}$
$\text{var}(T) = P_0(T) \cdot P_1(T) \cdot (m_0(T) - m_1(T))^2$	

5. **Εύρεση της τιμής κατωφλίου που μεγιστοποιεί τη διακύμανση μεταξύ των κατηγοριών** : Αφού προσδιοριστεί η βέλτιστη τιμή κατωφλίου, μπορεί να εφαρμοστεί στην εικόνα της κλίμακας του γκρι για να ληφθεί μια δυαδική εικόνα. Σε αυτή τη δυαδική εικόνα, τα εικονοστοιχεία με τιμές έντασης πάνω από το κατώφλι αντιστοιχίζονται στην περιοχή του προσκηνίου. Αντίθετα, τα εικονοστοιχεία με τιμές έντασης κάτω από το κατώφλι αντιστοιχίζονται στην περιοχή φόντου.

3.2.2 Μορφολογικοί μετασχηματισμοί δυαδικής μάσκας

Οι μορφολογικοί μετασχηματισμοί αποτελούν λειτουργίες που βασίζονται στο σχήμα της εικόνας. Συνήθως εκτελούνται σε δυαδικές εικόνες. Χρειάζονται δύο είσοδοι, η μία είναι η αρχική εικόνα, η δεύτερη ονομάζεται δομικό στοιχείο ή πυρήνας που καθορίζει τη φύση της λειτουργίας.

Erosion

Ο πυρήνας ολισθαίνει μέσα στην εικόνα (όπως στη συνέλιξη 2D). Ένα εικονοστοιχείο της αρχικής εικόνας (είτε 1 είτε 0) θα θεωρηθεί 1 μόνο εάν όλα τα εικονοστοιχεία κάτω από τον πυρήνα είναι 1, διαφορετικά θα διαβρωθεί (θα γίνει μηδέν). Όλα τα εικονοστοιχεία κοντά στο όριο θα απορριφθούν ανάλογα με το μέγεθος του πυρήνα. Έτσι, το πάχος ή το μέγεθος του αντικειμένου του προσκηνίου μειώνεται ή απλά μειώνεται η λευκή περιοχή στην εικόνα. Είναι χρήσιμο για την αφαίρεση μικρών θορύβων (όπως, αποκόλληση δύο συνδεδεμένων αντικειμένων κ.λπ.

Dilation

Είναι ακριβώς το αντίθετο του Erosion. Εδώ, ένα στοιχείο εικονοστοιχείου είναι "1" εάν τουλάχιστον ένα εικονοστοιχείο κάτω από τον πυρήνα είναι "1". Έτσι αυξάνεται η λευκή περιοχή στην εικόνα ή αυξάνεται το μέγεθος του αντικειμένου του προσκηνίου. Συνήθως, σε περιπτώσεις όπως η αφαίρεση θορύβου, η διάβρωση (erosion) ακολουθείται από τη διαστολή (dilation). Επειδή, η διάβρωση αφαιρεί τους λευκούς θορύβους, αλλά επίσης συρρικνώνει το αντικείμενο. Έτσι, το διαστέλλουμε.

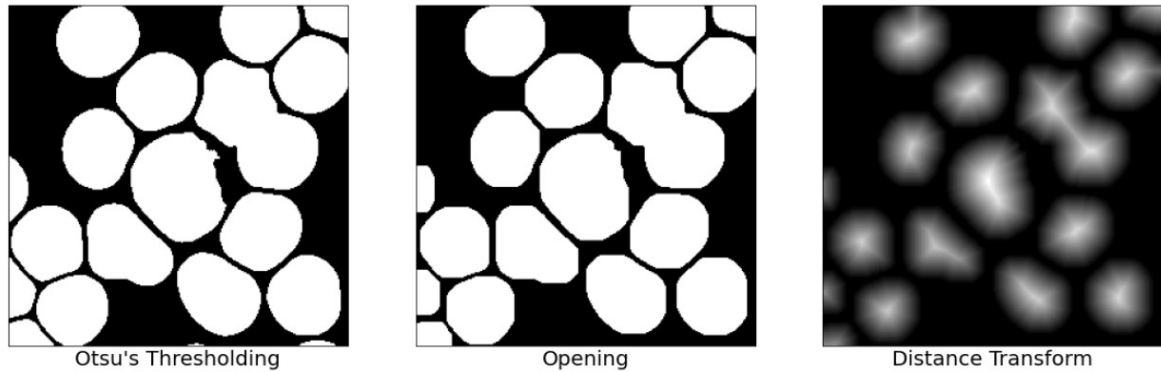
Opening

Το άνοιγμα είναι απλώς μια άλλη ονομασία της διάβρωσης που ακολουθείται από διαστολή. Είναι χρήσιμο για την αφαίρεση του θορύβου, όπως παρουσιάστηκε παραπάνω.

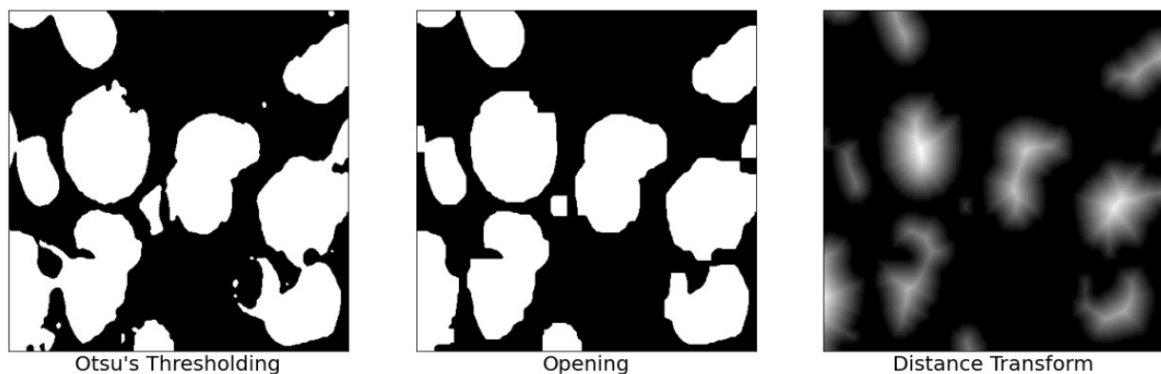
Μετασχηματισμός απόστασης

Ο μετασχηματισμός απόστασης είναι ένας τελεστής που συνήθως εφαρμόζεται μόνο σε δυαδικές εικόνες. Το αποτέλεσμα του μετασχηματισμού είναι μια εικόνα σε επίπεδο

γκρίζου που μοιάζει με την εικόνα εισόδου, εκτός από το ότι οι εντάσεις του επιπέδου γκριζου των σημείων εντός των περιοχών του προσκηνίου αλλάζουν για να δείξουν την απόσταση από το πλησιέστερο όριο από κάθε σημείο. Ο μετασχηματισμός απόστασης παρέχει μια μετρική ή μέτρο του διαχωρισμού των σημείων στην εικόνα. Υπολογίζεται αντικαθιστώντας κάθε εικονοστοιχείο μιας δυαδικής εικόνας με την απόσταση από το πλησιέστερο εικονοστοιχείο φόντου. Εάν το ίδιο το εικονοστοιχείο είναι ήδη μέρος του φόντου, τότε αυτή είναι μηδέν.



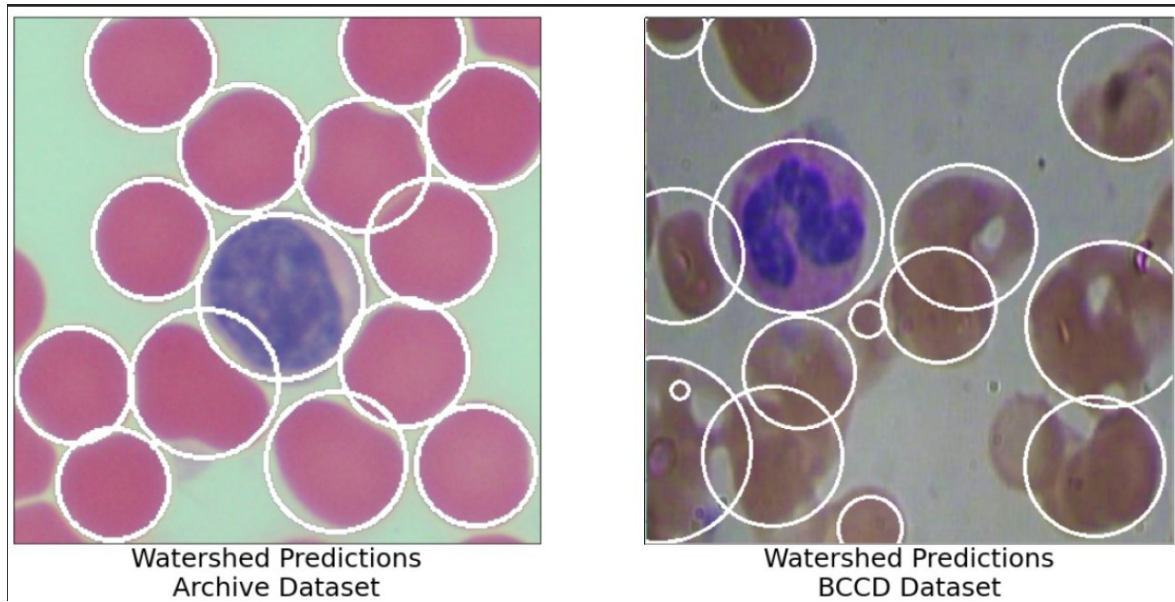
Σχήμα 3.7: Η συνέχεια της διαδικασίας στην εικόνα του Archive Dataset



Σχήμα 3.8: Η συνέχεια της διαδικασίας στην εικόνα του BCCD Dataset

3.2.3 Μέθοδος Watershed

Ο αλγόριθμος watershed δεχεται ως εισοδο την εικόνα μετασχηματισμού αποστασης και τοποθετεί εναν δείκτη στα τοπικά ελάχιστα της εικόνας. Αυτοί οι δείκτες αποτελούν τα σημεία εκκίνησης για τη διαδικασία. Θα πρέπει ιδανικά να αντιστοιχούν στα αντικείμενα ή στις περιοχές ενδιαφέροντος της εικόνας. Ξεκινώντας από τους δείκτες, ο αλγόριθμος αυξάνει επαναληπτικά κάθε περιοχή εξετάζοντας τα γειτονικά εικονοστοιχεία. Συγκρίνει τις εντάσεις των γειτονικών εικονοστοιχείων για να καθορίσει την ομοιότητά τους. Εάν ένα γειτονικό εικονοστοιχείο έχει παρόμοια ένταση με την τρέχουσα περιοχή, προστίθεται στην περιοχή. Η διαδικασία αυτή συνεχίζεται μέχρι να μην μπορούν να προστεθούν άλλα εικονοστοιχεία στην περιοχή. Το τελικό αποτέλεσμα του αλγορίθμου watershed περιλαμβάνει τόσο τις τμηματοποιημένες περιοχές όσο και τα όρια μεταξύ τους. Τα όρια αυτά αναφέρονται ως γραμμές watershed. Αντιπροσωπεύουν τα όρια των περιοχών όπου οι διαφορές έντασης μεταξύ γειτονικών εικονοστοιχείων είναι έντονες.



Σχήμα 3.9: Η συνέχεια της διαδικασίας στην εικόνα του BCCD Dataset

3.2.4 Μετρική DICE

Η βαθμολογία Dice, επίσης γνωστή ως συντελεστής ομοιότητας Dice, είναι ένα μέτρο της ομοιότητας μεταξύ δύο συνόλων δεδομένων, που συνήθως αναπαρίστανται ως δυαδικοί πίνακες. Στο πλαίσιο της τμηματοποίησης εικόνας, για παράδειγμα, το σκορ Dice μπορεί να χρησιμοποιηθεί για την αξιολόγηση της ομοιότητας μεταξύ μιας προβλεπόμενης μάσκας τμηματοποίησης και της αληθούς μάσκας τμηματοποίησης. Η βαθμολογία Dice κυμαίνεται από 0, που υποδηλώνει μηδενική επικάλυψη, έως 1, που υποδηλώνει τέλεια επικάλυψη.

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

Οι παραγόμενοι κύκλοι κατακερματισμού που προβλέφθηκαν από τον αλγόριθμο watershed μετατρέπονται σε οριοθετημένα πλαίσια προβλέψεων και συγκρίνονται με τα αληθή οριοθετημένα πλαίσια μέσω της μετρικής DICE. Το μέσο DICE επάνω στις εικόνες του archive dataset υπολογίστηκε στο 72.91%, ενώ στο BCCD Dataset ως 41.71%. Τα αποτελέσματα, κυρίως στο BCCD, δεν είναι ικανοποιητικά και αυτό αποδίδεται στις επικαλύψεις κυττάρων που δεν μπορούν να απεικονιστούν επαρκώς σε δυαδικές μάσκες παραγόμενες με την μέθοδο OTSU. Για αυτόν τον λόγο, θα δοκιμαστούν παρακάτω τεχνικές μηχανικής μάθησης με σκοπό τη βελτίωση των αποτελεσμάτων.

Κεφάλαιο 4

Υλοποίηση-Μεθοδολογία

Στόχος είναι να υλοποιηθεί κατάλληλο pipeline με στόχο την απομόνωση των επιμερών ερυθροκυττάρων από την εικόνα επιχρίσματος αίματος και στη συνέχεια την ταξινόμησή τους ως επιμολυσμένα ή μη από τον ιό της ελονοσίας. Θα χρησιμοποιηθούν διαφορετικές εκδόσεις του αλγορίθμου Yolo με σκοπό τη σύγκρισή τους και επείτα ένα custom συνελικτικό δίκτυο για την κατηγοριοποίηση των ερυθροκυττάρων. Για ολόκληρη τη διαδικασία της εκπαίδευσης αλλά και των δοκιμών χρησιμοποιήθηκε η πλατφόρμα Google Collab Pro με κάρτα γραφικών Nvidia Tesla T4.

4.1 Darknet και Yolo για ανίχνευση ερυθροκυττάρων σε εικόνες επιχρίσματος αίματος

Το Darknet [30] προκειται για ένα πλαίσιο νευρωνικών δικτύων ανοικτού κώδικα που έχει υλοποιηθεί με κύρια γλώσσα την C και CUDA από τον Joseph Redmon. Σε αντιθεση με το tensorflow ή το pytorch που επιτελούν ένα ευρύ φάσμα διεργασιών το darknet μαζί με τον αλγόριθμο Yolo εξειδικεύονται αποκλειστικά στην ανίχνευση αντικειμένων έχοντας ασύγκριτη ταχύτητα και ακρίβεια στους υπολογισμούς. Το darknet μπορεί να χρησιμοποιηθεί τόσο στη CPU όσο και στη GPU αλλά όπως αναφέρει ο δημιουργός του στην GPU μπορεί να είναι έως και 500 φορές γρηγορότερο καθώς εκμεταλεύεται την παραλληλοποίηση διεργασιών μέσω CUDA και CUDNN [31]. Λόγο των διαθεσιμών πόρων στην πλατφόρμα Google Collab PRO θα υλοποιηθεί η εκδοση για GPU.

Ο αλγόριθμος Yolo (συντομευση για "You Only Look Once") είναι ένας αλγόριθμος ανίχνευσης αντικειμένων πραγματικού χρόνου που παρουσιάστηκε πρώτη φορά το 2016 από τον δημιουργό του Darknet, Joseph Redmon [32]. Σχεδιάστηκε με γνώμονα τις υψηλές επιδόσεις σε ταχύτητα και ακρίβεια και γι' αυτό χρησιμοποιεί ένα νευρωνικό δίκτυο τόσο για την ανίχνευση όσο και για την κατηγοριοποίηση, σε αντίθεση με άλλες τεχνικές της εποχής όπως το R-CNN που βασίζονταν σε πολλαπλά επίπεδα για τη λειτουργία τους.

Ο αλγόριθμος Yolo λειτουργεί τμηματίζοντας την εκάστοτε εικόνα σε ένα πλέγμα από κελιά και προβλέποντας οριοθετούμενα πλαίσια, καθώς και έναν βαθμό σιγουριάς για κάθε κελί. Κάθε παραγόμενο οριοθετούμενο πλαίσιο απαρτίζεται από τις συντεταγμένες x, y (το κέντρο του πλαισίου), το πλάτος, το μήκος και τέλος έναν βαθμό εμπιστοσύνης που απεικονίζει την πιθανότητα το προς ανίχνευση αντικείμενο να βρίσκεται εντός αυτού του πλαισίου.

Η αρχική έκδοση του αλγορίθμου είχε αρκετούς περιορισμούς, όπως δυσκολία ανίχνευσης μικρών αντικειμένων, καθώς και δυσκολία ανίχνευσης αντικειμένων που βρίσκονται

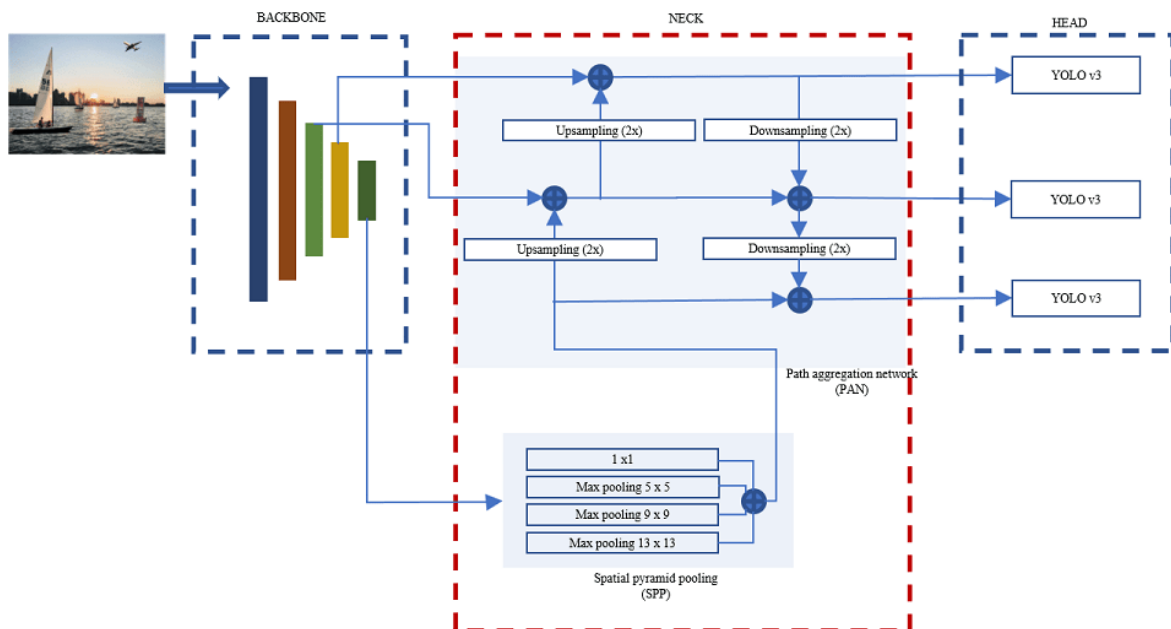
πολύ κοντά μεταξύ τους ή ακόμη χειρότερα παρουσιάζουν overlap. Οι περιορισμοί αυτοί είναι πολύ σημαντικοί για την παρούσα διπλωματική εργασία, καθώς οι εικόνες ερυθροκυττάρων συνήθως είναι αρκετά χαμηλού μεγέθους και τα κύτταρα αρκετές φορές είναι συμπηκνωμένα μεταξύ τους.

Γι' αυτό το λόγο θα χρησιμοποιηθούν οι μεταγενέστερες εκδόσεις του αλγορίθμου Yolo-V4 και YoloV7 που επιφέρουν λύσεις στα προαναφερθέντα ζητήματα, και θα συγκριθούν τα αποτελέσματα ταχύτητας και ακρίβειας των δύο.

4.1.1 Αρχιτεκτονική YoloV4

Η αρχιτεκτονική της τέταρτης έκδοσης του αλγορίθμου Yolo [33] απαρτίζεται από τρία βασικά στοιχεία :

- Ραχοκοκαλία (Backbone : CSPDarknet53.
- Λαιμός (Neck) : SPP,PAN.
- Κεφαλή (Head) : YoloV3.

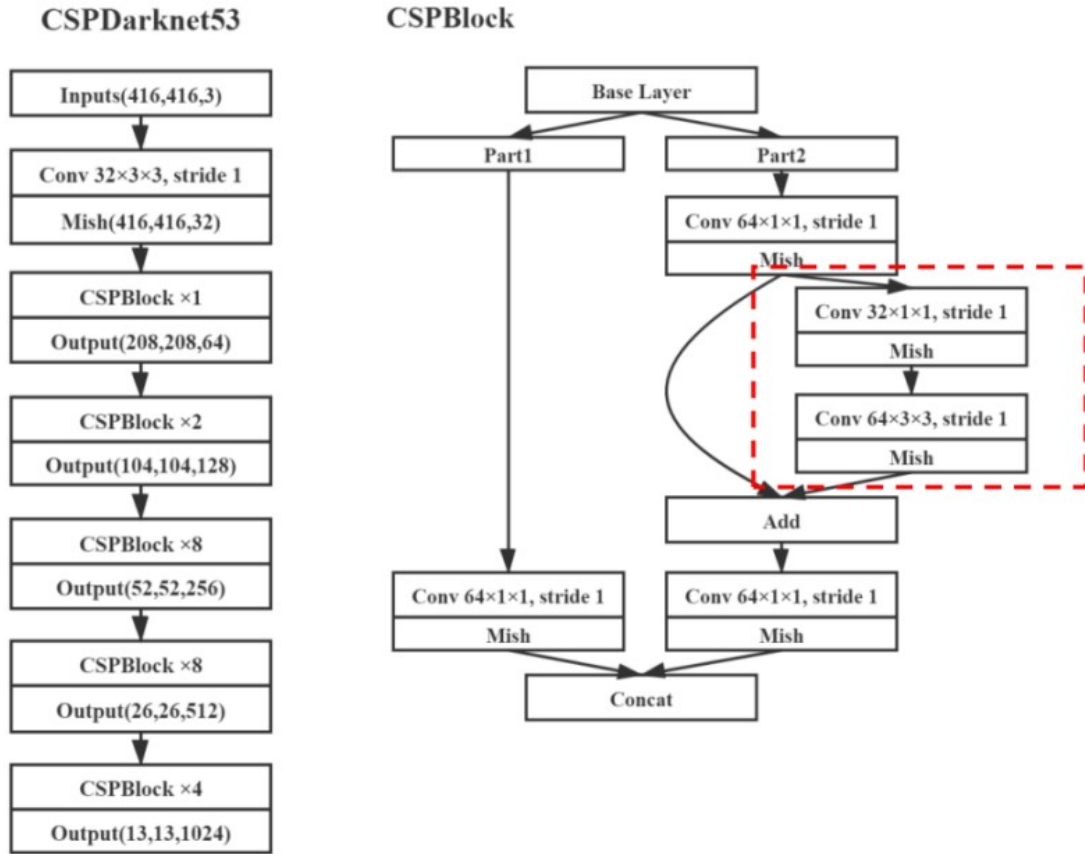


Σχήμα 4.1: Η αρχιτεκτονική του YoloV4

Ραχοκοκαλιά

Σε έναν αλγόριθμο ανίχνευσης, η ραχοκοκαλιά είναι συνήθως ένα δίκτυο προεκπαιδευμένο στο ImageNet [34] και χρησιμοποιείται με σκοπό την εξαγωγή χρήσιμων και πολύτιμων χαρακτηριστικών από την είσοδο του δικτύου. Μετά από εκτενείς πειραματισμούς, οι συγγραφείς του YoloV4 επέλεξαν το συνελκτικό δίκτυο CSPDarknet53 [35] για να επιτελέσει αυτόν τον σκοπό. Το εν λόγω δίκτυο είναι εμπνευσμένο από το CSPnet και εκμεταλλεύεται τη βασική ιδέα του διαχωρισμού της εισόδου σε δύο μέρη. Το ένα μέρος περνάει από το εκάστοτε συνελκτικό block, στην προκειμένη ένα darkblock, καθώς και ένα επίπεδο μετάβασης, και το δεύτερο μέρος συνδέεται απευθείας στην έξοδο κατ' αλληλουχία (concatenation) με το πρώτο. Αυτή η αρχιτεκτονική έχει ως αποτέλεσμα τη μείωση του φαινομένου εξαφάνισης κλίσεων και επαναλαμβανόμενων κλίσεων που ταλανίζουν την απόδοση πολύ βαθιών συνελκτικών δικτύων (Vanishing

Gradient Problem), ενισχύει τη διάδοση χαρακτηριστικών και μειώνει σημαντικά τις παραμέτρους του δικτύου, κάτι που είναι πολύ σημαντικό λόγω των περιορισμών του hardware.

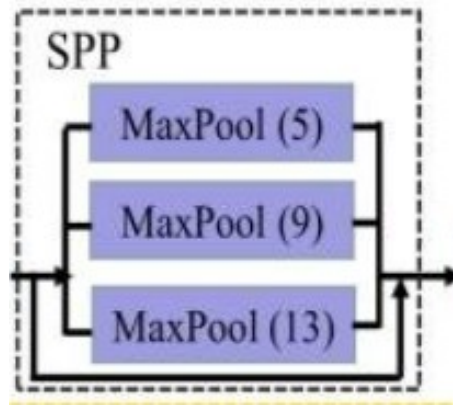


Σχήμα 4.2: Η αρχιτεκτονική της ραχοκοκαλιάς **CSPDarknet53**
(Στο κόκκινο πλαίσιο βλέπουμε ένα επίπεδο μετάβασης εντός του DarkBlock)

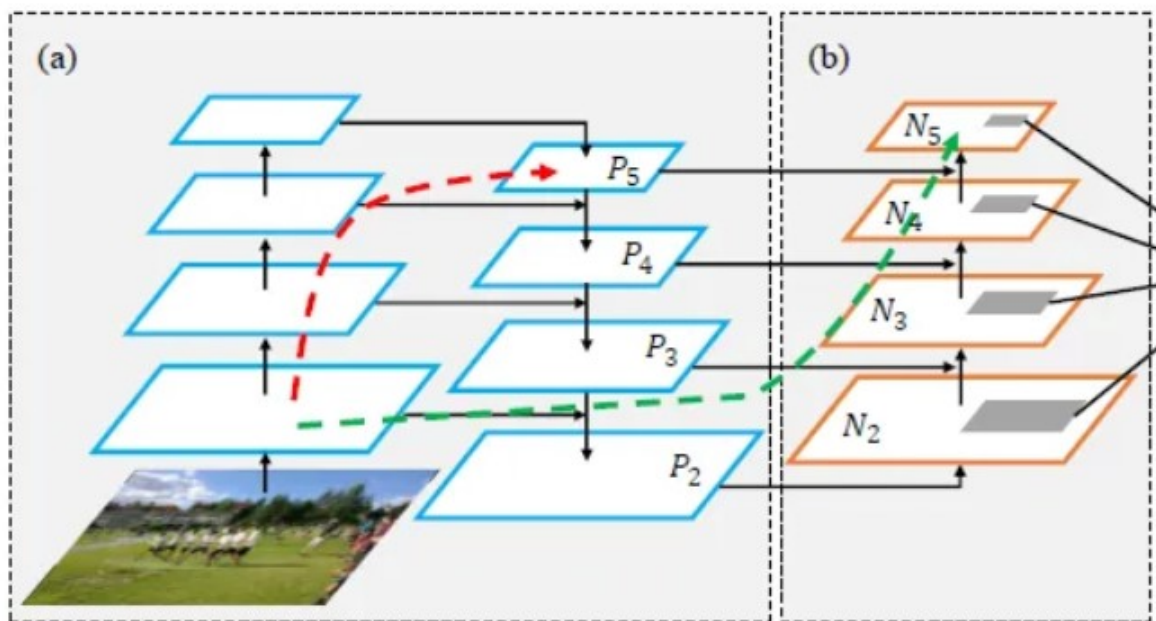
Λαιμός

Στο πρώτο στάδιο, ο λαιμός του δικτύου επιτελεί τη διαδικασία διαχείρισης και συνδυασμού των χαρακτηριστικών που έχουν εξαχθεί από τη ραχοκοκκαλιά. Χρησιμοποιείται ένα επίπεδο **Spatial Pyramid Pooling (SPP)** [36] με σκοπό την αύξηση του δεκτικού πεδίου, τη διατήρηση χωρικών χαρακτηριστικών και σχέσεων σε επίπεδο πιξελ που θα βοηθήσουν στην ακρίβεια των υπολογισμών. Αυτό επιτυγχάνεται μέσω της εξαγωγής χωρικών σχέσεων και χαρακτηριστικών μέσω πολλαπλών επιπέδων MaxPooling διαφορετικής κλίμακας. Ένα επίπεδο MaxPooling κλίμακας (2,2) διπλασιάζει το δεκτικό πεδίο του δικτύου.

Καθώς η εικόνα περνάει μέσα από τα πολλαπλά συνελκτικά επίπεδα του δικτύου, η πολυπλοκότητα των εξαγόμενων χαρακτηριστικών αυξάνεται και ανάλογα μειώνεται η χωρική ανάλυσή τους. Για την επίλυση αυτού του ζητήματος προστίθεται στον λαιμό το δίκτυο Path Aggregation Network (PA-net) [37]. Η βασική ιδέα είναι ότι χαρακτηριστικά εξαγόμενα από υψηλά επίπεδα έχουν και την αντίστοιχη υψηλή χωρική ανάλυση



Σχήμα 4.3: Το επιπεδο **SPP** στο δίκτυο YoloV4



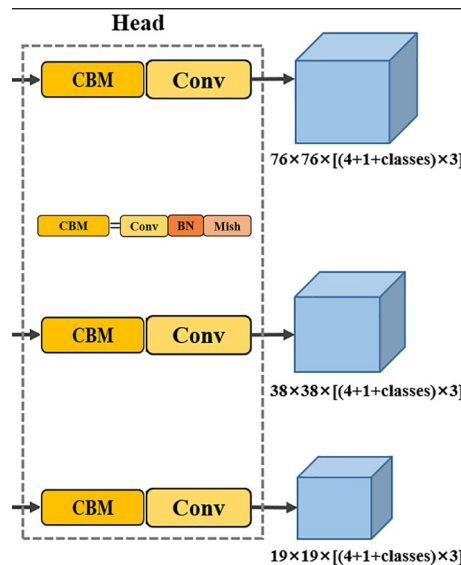
Σχήμα 4.4: Η ιεραρχία κατά πυραμίδες εντός του δικτύου **PANet** στον λαιμό του YoloV4

με έμφαση στις λεπτομέρειες της εικόνας, ενώ τα αντίστοιχα χαρακτηριστικά των χαμηλότερων επιπέδων έχουν μεγαλύτερο δεκτικό πεδίο και, κατά συνέπεια, περισσότερη γενική πληροφορία. Για να μπορέσει ο αλγόριθμος να εκμεταλλευθεί τις σημαντικές και χρήσιμες πληροφορίες τόσο των υψηλών όσο και των χαμηλότερων στρωμάτων, δημιουργούνται δύο ιεραρχίες χαρακτηριστικών σε μορφή πυραμίδας: μία από πάνω προς τα κάτω και μία από κάτω προς τα πάνω, και τα χαρακτηριστικά τους ανασυνδυάζονται.

Κεφαλή

Ο Yolon4 έχει τρεις κεφαλές ανίχνευσης όπου η κάθε μία λαμβάνει χαρακτηριστικά από διαφορετικά επίπεδα της πυραμίδας ιεραρχίας χαρακτηριστικών. Η κάθε κεφαλή YoloV4 παράγει ένα σύνολο προβλέψεων συντεταγμένων για κάθε πλέγμα κελιών στον χάρτη χαρακτηριστικών εξόδου.. Στην κάθε πρόβλεψη περιλαμβάνονται οι συντεταγμένες του πλαισίου που περιβάλλει το αντικείμενο (συντεταγμένες οριζόντιας και κάθετης άξονα), καθώς και ο βαθμός σιγουριάς (confidence score) για την ύπαρξη αντικειμένου σε αυτό

το πλαίσιο. Επίσης, κάθε πρόβλεψη περιλαμβάνει και τις πιθανότητες των κλάσεων των αντικειμένων που ανιχνεύονται. Η χρήση τριών κεφαλών ανίχνευσης επιτρέπει στο YoloV4 να ανιχνεύει αντικείμενα σε διαφορετικές κλίμακες και να παρέχει ακριβέστερες προβλέψεις για αντικείμενα διαφορετικού μεγέθους στην εικόνα.



Σχήμα 4.5: Η κεφαλή του δικτύου YoloV4

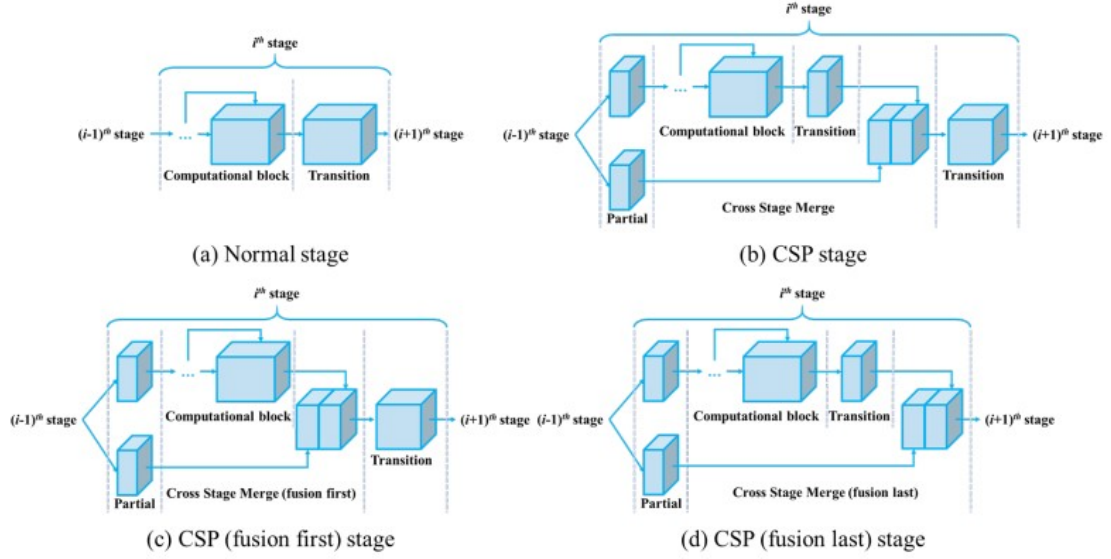
4.1.2 Αρχιτεκτονική YoloV7

Σκοπός της δημιουργίας του YoloV7 [38] ήταν η κατασκευή ενός αλγορίθμου τελευταίας τεχνολογίας ικανό να παράγει ικανοποιητικότερα αποτελέσματα από τις προηγούμενες εκδόσεις, κυρίως το ΨολοP [39], με καλύτερες ωστόσο χρονικές αποδόσεις. Και αυτός ο αλγόριθμος βασίζεται στην δομή ραχοκοκαλιά, λαιμός, κεφαλή ωστόσο με νέες αρχιτεκτονικές υλοποιήσεις που θα αναλυθούν στην συνέχεια

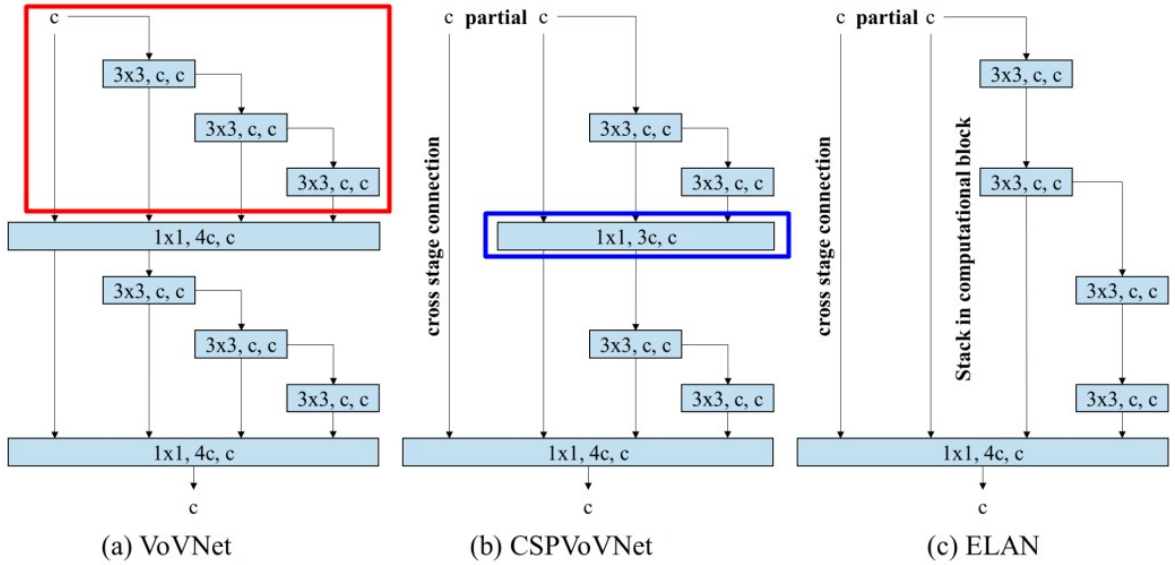
ELAN | E-ELAN

Ο βασικός σκοπός της κατασκευής της αρχιτεκτονικής ELAN είναι η επίλυση του ζητήματος ότι, καθώς αυξάνεται το βάθος και το πλάτος ενός δικτύου σε μεγάλη κλίμακα, η απόδοση χειροτερεύει αισθητά. Συγκεκριμένα, η αρχιτεκτονική ResNet 152 επιπέδων (ResNet-152) έχει μόλις 1% αύξηση ακρίβειας σε σχέση με την κλασική μορφή 50 επιπέδων αυτού του μοντέλου. Ακόμη χειρότερα, το VovNet 99 επιπέδων έχει μικρότερη ακρίβεια από το VovNet 39 επιπέδων [40]. Οι ερευνητές εικάζουν ότι αυτή η διαφορά έγκειται στο γεγονός ότι το ResNet, μέσω των πλευρικών συνδέσεων εισόδου-εξόδου, δεν αυξάνει την μικρότερη διαδρομή κλίσεων, σε αντίθεση με το VovNet που χρησιμοποιεί OSA modules και σε κάθε νέο στοιβάγμα αυξάνει την διαδρομή κατά ένα [41]. Μετά από πολλαπλά πειράματα επιβεβαιώνεται αυτή η υπόθεση και παρατηρείται ότι, καθώς συνδέονται πάνω από 80+ επίπεδα CSP (Cross Stage Partial), και το δίκτυο γίνεται πολύ βαθύ και πλατύ, η fusion last μορφή που χρησιμοποιεί λιγότερα επίπεδα μετάβασης αποδίδει καλύτερα αποτελέσματα από την κλασική μορφή CSP. Με γνώμονα το γεγονός ότι η μικρότερη διαδρομή κλίσεων δεν πρέπει να είναι αρκετά μεγάλη για να περιορίζεται το φαινόμενο εξαφάνισης κλίσεων και ότι η μεγαλύτερη διαδρομή κλίσεων δεν πρέπει να είναι αρκετά μικρή για να μπορεί το δίκτυο να εξάγει πλουσιότερα

χαρακτηριστικά, κατασκευάστηκε η συνδεσμολογία ELAN. Αποτελείται από δύο σκιπ συνδέσεις και ένα συνελκτικό block που απαρτίζεται εσωτερικά από συνελίζεις πυρήνα 3x3.



Σχήμα 4.6: Cross stage partial δίκτυα. (a) κλασσικό δίκτυο, (b) CSP απλό: μετάβαση → αλληλουχία → μετάβαση, (c) CSP fusion first: αλληλουχία → μετάβαση, και (d) CSP fusion last: μετάβαση → αλληλουχία

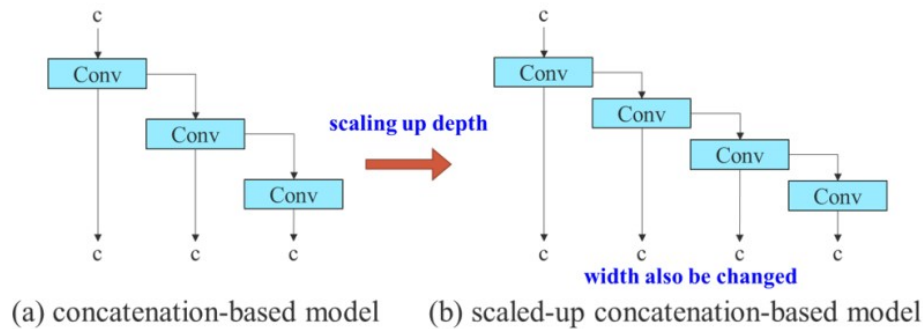


Σχήμα 4.7: (a) Η αρχιτεκτονική vonnet . Στο κόκκινο πλαίσιο ένα osa module (b) Η αρχιτεκτονική CSPvonnet . Εξέλιξη του vonnet καθώς εντός δικτύου χρησιμοποιεί cross stage partial σύνδεση . Στο μπλέ πλαίσιο βλέπουμε ένα στάδιο μετάβασης. (c) Τέλος η αρχιτεκτονική ELAN . Παρατηρείται η cross stage partial σύνδεση , η απουσία πολλαπλών επιπέδων μετάβασης και το συνελκτικό μπλόκ για εξαγωγή χαρακτηριστικών.

Κλιμακοποίηση δικτύων που βασίζονται σε αλληλουχία χαρακτηριστικών

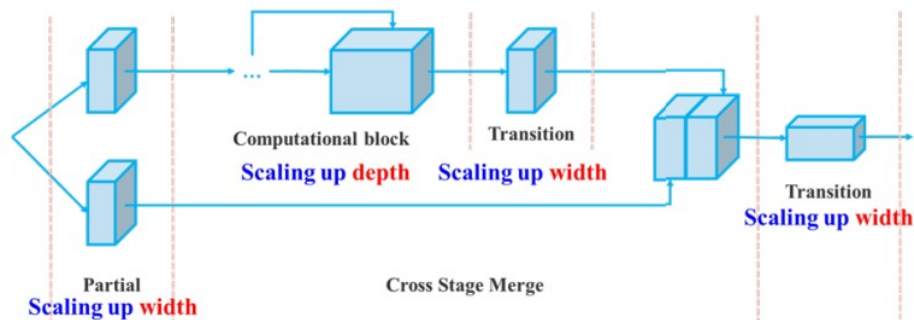
Στόχος της κλιμακοποίησης δικτύων είναι η ρύθμιση διάφορων παραμέτρων του δικτύου που το καθιστούν βαθύτερο ή πλατύτερο, ώστε να επιτευχθούν οι διαφορετικές

ανάγκες σε ταχύτητα και ακρίβεια που προκύπτουν κατά την επίλυση του εκάστοτε προβλήματος βαθιάς μάθησης. Αυξάνοντας όμως το βάθος ενός συνελκτικού μπλοκ που προηγείται από ένα επίπεδο αλληλουχίας χαρακτηριστικών και ένα επίπεδο μετάβασης, παράλληλα αυξάνεται και το πλάτος του μπλοκ. Αυτό συμβαίνει επειδή η κλίμακα του βάθους συνήθως αυξάνει τον αριθμό των φίλτρων ή/και των επιπέδων στο συνελκτικό μπλοκ. Κάθε επίπεδο μπορεί να περιέχει περισσότερες μονάδες, που οδηγούν σε αυξημένο πλήθος χαρακτηριστικών κατά την επεξεργασία της εικόνας. Ως αποτέλεσμα, το πλάτος του συνελκτικού μπλοκ αυξάνεται. Έτσι προτείνεται ένας ελεγχόμενος



Σχήμα 4.8: Παρατηρείται ότι αυξάνοντας το βάθος στο συνελκτικό μπλοκ με τη μέθοδο της εικόνας, τότε αυξάνεται παράλληλα και το πλάτος του.

τρόπος αύξησης βάθους και πλάτους, με τον οποίο η είσοδος του μπλοκ χωρίζεται σε δύο μέρη και η κλιμακοποίηση του βάθους εκτελείται μόνο στο συνελκτικό μπλοκ του ενός μέρους. Με αυτόν τον τρόπο, το δίκτυο διατηρεί την αρχική σχεδίαση και δομή του, και παράλληλα μπορεί να εξεταστεί η επίδραση κάθε κλιμακοποίησης στο αποτέλεσμα.

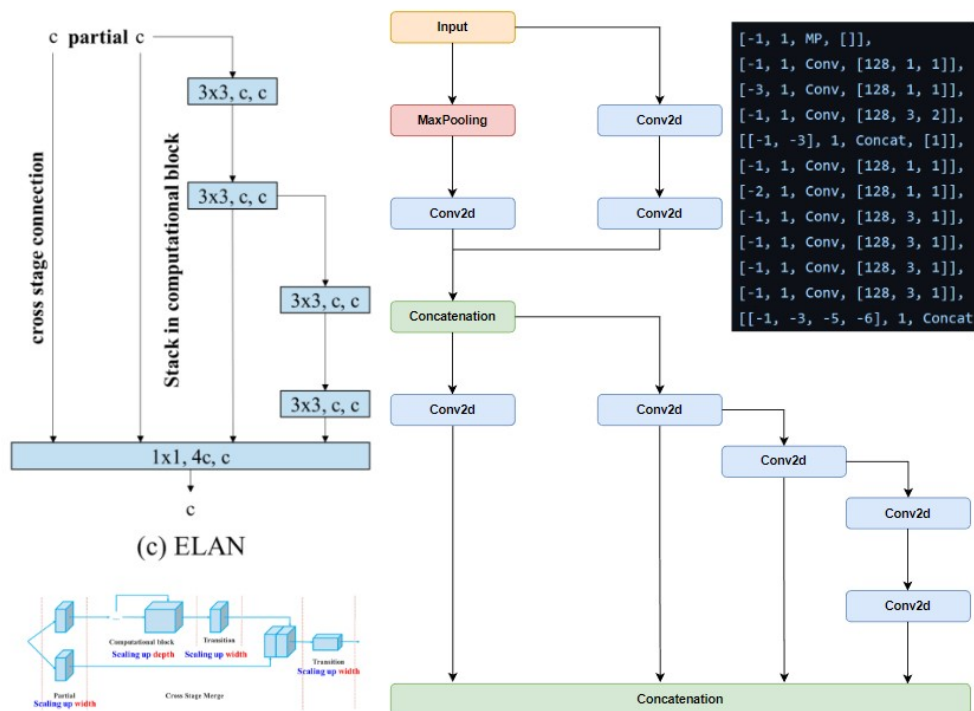


Σχήμα 4.9: Η προτεινόμενη μέθοδος κλιμακοποίησης του δικτύου.

Ραχοκοκαλιά

Καθώς υπάρχει μια ασάφεια στο αρχικό άρθρο σχετικά με τα δομικά στοιχεία της ραχοκοκκαλιάς, η αρχιτεκτονική θα επιβεβαιωθεί καλύτερα κοιτώντας τα αρχεία YAML [42] που διέπουν την κατασκευή του δικτύου. Η YAML είναι μια γλώσσα ευκολονόητη από ανθρώπους, που χρησιμοποιείται σε αρχεία διαμόρφωσης, στην προκειμένη για τις παραμέτρους του δικτύου.

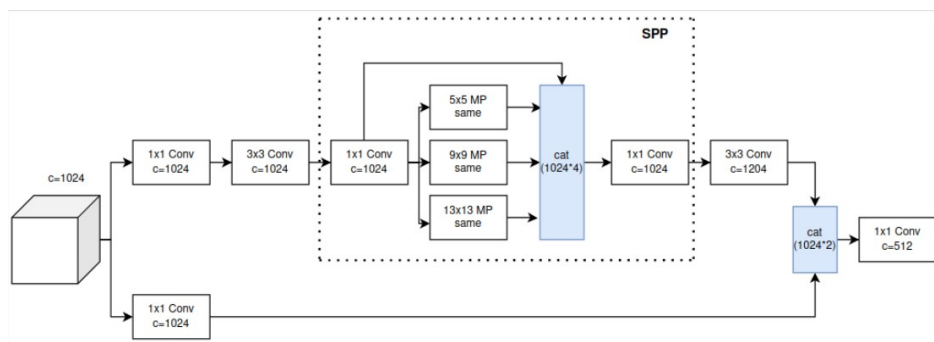
Παρατηρείται ότι σε τμήματα της ραχοκοκκαλιάς, η αρχιτεκτονική ELAN χρησιμοποιεί δύο συνελίζεις 1×1 που συνδέονται κατά αλληλουχία στην έξοδο, ακολουθούμενες από ένα συνελκτικό μπλοκ 3×3 συνελίζεων.



Σχήμα 4.10: Ένα κομμάτι της ραχοκοκαλιάς του δικτύου YoloV7

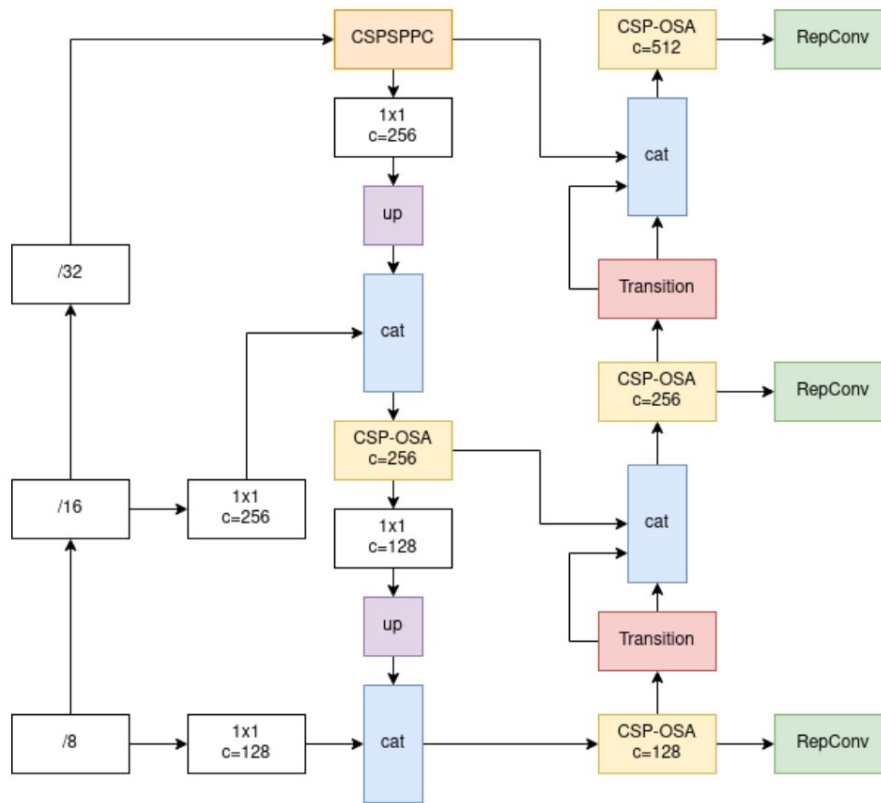
Λαιμός

Ο λαιμός του YoloV7 παρουσιάζει παρομοιες ιδέες με τον YoloV4 Σχετικά με την διαχείριση των χαρακτηριστικών της ραχοκοκαλιάς με διαφορετική ωστόσο υλοποίηση και αρχιτεκτονική. Αρχικά και πάλι εισάγονται χαρακτηριστικά στον λαιμό από διαφορετικά επίπεδα της ραχοκοκαλιάς. Για την αύξηση του δεκτικού πεδίου, η έβδομη χρησιμοποιεί μια βελτιωμένη έκδοση του SPP, την Cross-Stage-Partial-SPP (CSPSPP). [43] που εκμεταλλεύεται τον τρόπο επέκτασης πλάτους σε αρχιτεκτονικές με επίπεδα αλληλουχίας που αναλύθηκε προηγουμένως. .



Σχήμα 4.11: Η βελτιωμένη έκδοση του SPP χρησιμοποιεί τις ιδέες της αποτελεσματικότερης κλιμακοποίησης των κατά αλληλουχία δικτύων.

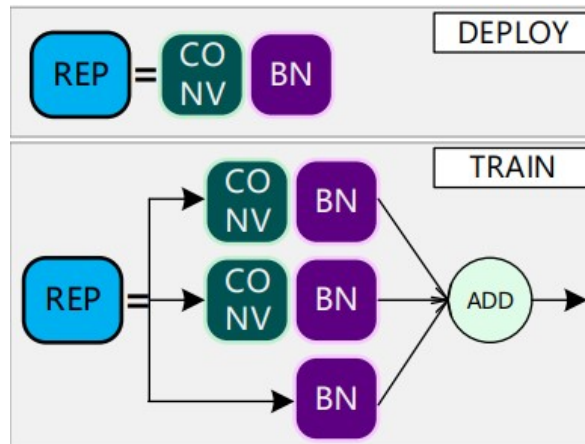
Και πάλι, υλοποιείται μια ιεραρχία στρώματων κατά πυραμίδες, ώστε ο αλγόριθμος να μπορέσει να εκμεταλλευθεί τις σημαντικές και χρήσιμες πληροφορίες τόσο από τα υψηλά όσο και από τα χαμηλά στρώματα, στις τρεις κεφαλές που ακολουθούν.



Σχήμα 4.12: Ο λαιμός του δικτύου YoloV7.

Κεφαλή

Στην κεφαλή του δικτύου υπάρχουν τρία επαναπαραμετροποιήσιμα μπλοκ που ακολουθούνται από συνελιξείς 1x1 λόγω της φύσης του δικτύου που είναι πλήρως συνελκτικό. Το επαναπαραμετροποιήσιμο μπλοκ αποτελείται κατά τη διάρκεια της εκπαίδευσης από συνελιξείς 3x3, 1x1 και επίπεδα κανονικοποίησης. Ωστόσο, κατά τη διάρκεια της χρήσης του δικτύου για ανίχνευση, το μπλοκ μετατρέπεται σε μια μόνο συνέλιξη ακολουθούμενη από επίπεδο κανονικοποίησης, εξού και η ονομασία του. Αυτή η αλλαγή δεν περιορίζει την ακρίβεια των υπολογισμών, ενώ παράλληλα μειώνει τα FLOPs (Floating Point Operations) και αυξάνει την ταχύτητα υπολογισμών. [44] καθώς μειώνονται οι παραμετροί του δικτύου. Έπειτα οι εξοδοί των τριών συνελιξεων που προέρχονται από διαφορετικά βαθμια του δικτύου συμψηφίζονται για να εξαχθεί το τελικό αποτέλεσμα.



Σχήμα 4.13: Το επαναπαραμετροποιήσιμο μπλοκ RepConv

4.1.3 Παράμετροι δικτύων

Κάθε έκδοση του αλγορίθμου Yolo εμπεριέχει ένα αρχείο διαμόρφωσης (.cfg) παραμέτρων και υπερπαραμέτρων το οποίο διεπεί την λειτουργία του δικτύου. Έτσι οι βασικοί υπερπαραμέτροι εκπαίδευσης θα επεξεργαστούν πανομοιότυπα σε κάθε αλγόριθμο για να υπάρχει δίκαιη σύγκριση με βάση το πρός επίλυση πρόβλημα. Παρακάτω θα παρουσιαστούν οι παράμετροι οι οποίοι θα υποστούν αλλαγές με βάση τις οδηγίες εκπαίδευσης του δικτύου :

Batch

Το μέγεθος της δέσμης αποτελεί τον αριθμό εικόνων που θα επεξεργαστούν πριν γίνει μια ανανέωση στις παραμέτρους του δικτύου. Προτείνεται δε να είναι πάντα πολλαπλάσιο του 8. Για το παρόν πρόβλημα επιλέχθηκε μέγεθος δέσμης 64.

Subdivisions

Το μέγεθος της δέσμης διαιρείται με τον αριθμό υποδιαιρέσεων και το νούμερο που προκύπτει αποτελεί τον αριθμό εικόνων που μπορούν να φορτωθούν παράλληλα στην κάρτα γραφικών. Έχει λοιπόν άμεση εξάρτηση από τους περιορισμούς υλικού και προτείνεται να είναι επίσης πολλαπλάσιο του 8. Ο χαμηλότερος αριθμός subdivisions που μπορεί να υπάρξει στο παρόν πρόβλημα χωρίς να προκύψουν προβλήματα μνήμης είναι 32. Άρα φορτώνονται παράλληλα 2 εικόνες.

Max Batches

Ο αριθμός των επαναληψεων που θα γίνουν μεχρι να ολοκληρωθει η εκπαιδευση. Προτείνεται να είναι 2000 φορές επι τις προς εκπαιδευση κλασεις. Ωστοσο στο παρον προβλημα λογο των περιορισμων υλικου θα σταματησει η εκπαιδευση στις 2000 επαναληψεις παρολο που το archive dataset εχει 2 κλασεις και βαση θεωριας θα επρεπε να τελειωσει το λιγοτερο στις 4000.

Steps

Πρόκειται για συγκεκριμένο αριθμό επαναλήψεων (βημάτων) στα οποία θα μεταβληθεί ο ρυθμός εκπαίδευσης κατα 0.1 και 0.01 αντίστοιχα. Για χρήση δύο βημάτων

προτείνεται να συμβεί στο 80% και 90% των συνολικών επαναλήψεων. Για το παρόν πρόβλημα τα βήματα είναι 1800 και 1900.

Classes

Ο αριθμός των κλάσεων. Για το archive dataset είναι 2 και για το BCCD dataset 1.

Filters

Τα φίλτρα ή αλλιώς οι χάρτες χαρακτηριστικών των συνελκτικών μπλόκ. Υπολογίζονται από τον τύπο : **Filters = (Classes + 5)*3** . Συνεπώς 21 φίλτρα όταν πρόκειται για ανίχνευση 2 κλάσεων και 18 φίλτρα όταν πρόκειται για ανίχνευση μιας κλάσης.

4.1.4 Επαναρχικοποίηση κελιών

Το πρώτο βήμα είναι ο υπολογισμός των νέων συντεταγμένων μήκους και πλάτους των οριοθετούμενων πλαισίων με βάση τις εικόνες του εκάστοτε προβλήματος μέσω αλγορίθμου ομαδοποίησης k-μέσων. Η ομαδοποίηση αυτή έχει ως στόχο να διαχωρίσει παρατηρήσεις, στην προκειμένη συντεταγμένες των πλαισίων, σε κ ομάδες, 9 στην περίπτωση του αλγορίθμου Yolo. Η διαδικασία αυτή γίνεται αυτόματα εντός του Darknet μέσω της εντολής

!./darknet detector calc_anchors Data num_of_clusters 9 width .. height ..

Ετσι παράγονται 9 ζεύγη συντεταγμένων ξ, ψ , τα οποία θα τοποθετηθούν στο αρχείο διαμόρφωσης με βάση τον κανόνα ότι η πρώτη κεφαλή ανίχνευσης δέχεται συντεταγμένες με γινόμενο μικρότερο των $30 * 30$ pixel , δεύτερη με γινόμενο μικρότερο των $60 * 60$ και η τρίτη τα υπόλοιπα που δεν υπάγονται στις ανωτέρες κατηγορίες. Ένα πιθανό ζήτημα που μπορεί να προκύψει μετά τον επανυπολογισμό είναι να μην υπάρξει τιμή μικρότερη του $30 * 30$ για την πρώτη κεφαλή, όπου σε αυτήν την περίπτωση εισάγεται μια αυθαίρετη μικρότερη τιμή και αλλάζει ο αριθμός των ζευγών συντεταγμένων από 9 σε 10. Συγκεκριμένα, οι παράμετροι εντός του αρχείου διαμόρφωσης που πρέπει να αλλάξουν για την ορθή επαναρχικοποίηση κελιών είναι οι εξής:

Num

Ο αριθμός των ζευγών συντεταγμένων είναι συνήθως 9, εκτός από τις περιπτώσεις όπου απαιτείται πρόσθεση ή αφαίρεση συντεταγμένων.

Anchors

Οι συντεταγμένες των αρχικών οριοθετημένων πλαισίων.

Mask

Πρόκειται για έναν δείκτη, επάνω στις συντεταγμένες (anchors), που δείχνει στην εκάστοτε κεφαλή Yolo ποιες συντεταγμένες θα χρησιμοποιηθούν για ανίχνευση.

4.1.5 Εκπαίδευση

Η εκπαίδευση υλοποιείται στο πλαίσιο Darknet για 2000 επαναλήψεις όπως έχει αναφερθεί με batch size 64. Μέσω της τεχνικής του Transfer Learning θα χρησιμοποιηθούν στην έναρξη τα προεκπαιδευμένα βάρη του εκάστοτε αλγορίθμου στο ImageNet ή στο MsCocoDataset ώστε να επιταχυνθεί η ορθή διαδικασία της εκπαίδευσης. Το αρχείο

παραμετρών του δικτύου επεξεργάζεται με τον τρόπο που ειπώθηκε στην παραπάνω ενότητα τόσο για την απλή εκδοση του αλγορίθμου όσο και για τα επαναρχικοποιημένα κελιά. Η εκπαίδευση ξεκινά με την εντολή : **!./darknet detector train data/obj.data cfg/yolov7_training.cfg yolov7.conv.132**

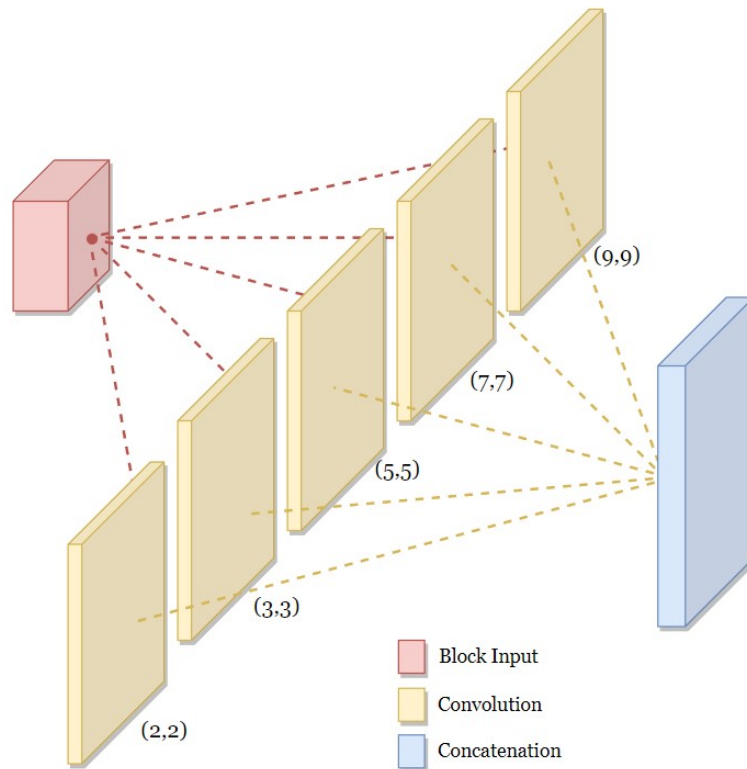
4.2 Ταξινόμηση μεμονωμένων ερυθροκυττάρων σε επιμολυσμένα ή μη από τον ιό της ελονοσίας

Σε αυτό το σημείο του γενικότερου pipeline, οι μεμονωμένες εικόνες κυττάρων που έχουν προκύψει από προηγούμενα βήματα πρέπει να αναλυθούν και να ταξινομηθούν ανάλογα με το αν είναι επιμολυσμένες από τον ιό της ελονοσίας. Θα μελετηθούν τρία συνελκτικά δίκτυα διαφορετικής αρχιτεκτονικής και σχεδιασμού ώστε να εξαχθεί και να μελετηθεί ένα ευρύ φάσμα αποτελεσμάτων. Το πρώτο δίκτυο πρόκειται για ένα προσαρμοσμένο μοντέλο βασισμένο σε συνελκτικά μπλοκ διαφορετικών πυρήνων (multi kernel blocks) εμπνευσμένο από το LNet [45] και τα ικανοποιητικά αποτελέσματα του σε εικόνες μικρής και χαμηλής ανάλυσης. Το επόμενο είναι μια ακόμη πιο αποδοτική και λιγότερων παραμετρών εκδοση του EfficientNet, το K-EffNet-B0 το οποίο είναι ιδανικό για δοκιμές σε εδγε συσκευές όπως το Jetson Nano. Έπειτα θα εξεταστεί το δίκτυο ConvNeXt ως ένα συνελκτικό δίκτυο εμπνευσμένο από τρανσφορμέρς και τέλος ένα ensemble δίκτυο του Multi Kernel και K-effnet-B0

4.2.1 Wide Multi-Kernel Network

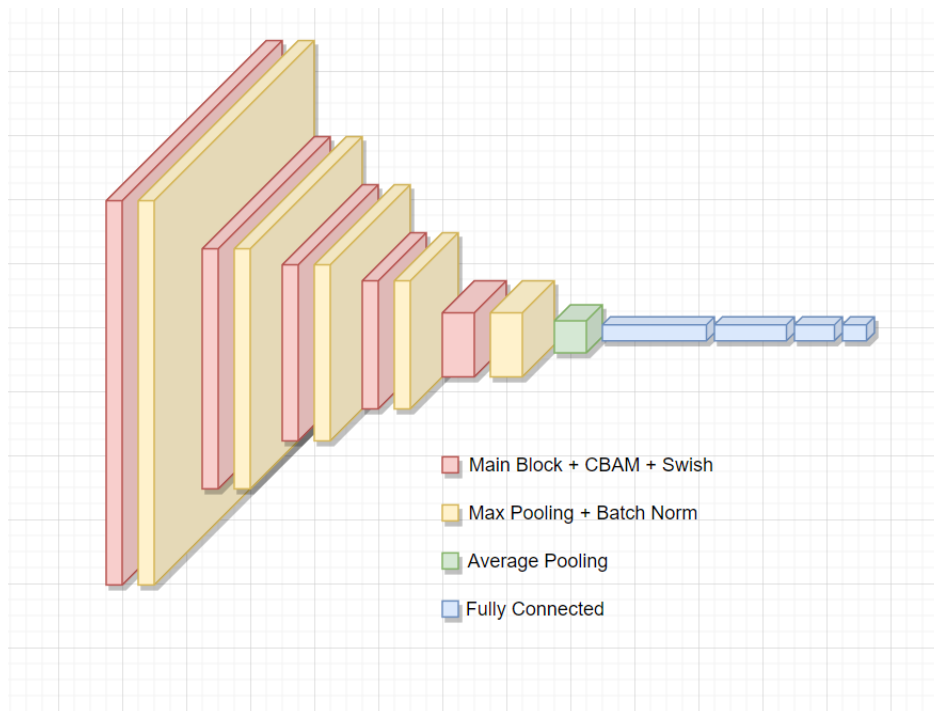
Η επιτυχία των συνελκτικών δικτύων στην κατηγοριοποίηση εικόνων και στην εξαγωγή χαρακτηριστικών τα καθιστά την πιο δημοφιλή αρχιτεκτονική για τέτοιες διεργασίες. Ωστόσο η χρήση μοντέλων τελευταίας τεχνολογίας, προεκπαιδευμένα σε βάσεις δεδομένων μεγάλων εικόνων όπως το ImageNet , για την ταξινόμηση θορυβώδων, μικρών και χαμηλής ποιότητας εικόνων δεν επιφέρει πολλές φορές τα επιθυμητά αποτελέσματα. Οι εικόνες που απαρτίζουν την βάση δεδομένων που θα χρησιμοποιηθεί, καθώς και γενικότερα οι εικόνες επιχρίσματος αίματος που δεν θα παραχθούν σε εργαστήριο με μικροσκόπια και μηχανισμούς τελευταίας τεχνολογίας, υπάγονται στην τελευταία κατηγορία. Για τις ανάγκες λοιπόν του προβλήματος υλοποιήθηκε μια custom αρχιτεκτονική , εμπνευσμένη από την έρευνα των Xiaohan Ding et al. [46] στους μεγαλύτερους πυρήνες και Sepidehsadat Hosseini et al. [47] στα πλατιά δίκτυα, βασισμένη σε blocks και σε διαφορετικών μεγεθών kernel ικανή να αντλεί περισσότερες πληροφορίες από κάθε εικόνα και να βελτιώνει τόσο τα αποτελέσματα επάνω στο testing dataset όσο και την γενίκευση του αλγορίθμου επάνω σε νέες εικόνες κυττάρων που δεν βρίσκονται στα δεδομένα εκπαίδευσης.

Τα πλατιά δίκτυα έχουν το πλεονέκτημα ότι διαθέτουν περισσότερα φίλτρα σε κάθε επίπεδο, γεγονός που τους επιτρέπει να μάθουν πιο σύνθετα χαρακτηριστικά από την εικόνα εισόδου. Αυτό μπορεί να είναι ιδιαίτερα επωφελές όταν χρησιμοποιούνται μικρές εικόνες, καθώς το δίκτυο έχει λιγότερα εικονοστοιχεία να επεξεργαστεί και μπορεί να χρειαστεί να εξάγει πιο λεπτομερή χαρακτηριστικά για να ταξινομήσει με ακρίβεια την εικόνα. Επιπλέον, τα ευρύτερα δίκτυα έχουν καλύτερες ιδιότητες γενίκευσης και μπορούν να είναι πιο ανθεκτικά στην υπερπροσαρμογή, η οποία αποτελεί κοινό πρόβλημα κατά την εργασία με μικρά σύνολα δεδομένων.



Σχήμα 4.14: Το συνελκτικό μπλόκ του Multi Kernel Network (MKN).

Για το λόγο αυτό, το κύριο μπλόκ του CNN αποτελείται από πέντε παράλληλα στρώματα συνελκτικού δικτύου, το καθένα με διαφορετικό μέγεθος πυρήνα που κυμαίνεται από (2,2) έως (9,9). Αυτά τα παράλληλα στρώματα συνελίξεων έχουν σχεδιαστεί για να εξάγουν διαφορετικά χαρακτηριστικά από την εικόνα εισόδου σε διαφορετικές κλίμακες. Οι μικρότεροι πυρήνες έχουν τη δυνατότητα να εξάγουν πιο λεπτομερή και τοπικά χαρακτηριστικά ενώ οι μεγαλύτεροι πυρήνες μια γενικότερη πληροφορία από την εικόνα. Οι έξοδοι αυτών των παράλληλων επιπέδων συνελίξεων στη συνέχεια συνενώνονται σε ένα επίπεδο αλληλουχίας, το οποίο συνδυάζει τα εξαγόμενα χαρακτηριστικά για να σχηματίσει μια πιο ολοκληρωμένη αναπαράσταση της εικόνας εισόδου. Αυτός ο συνδυασμένος χάρτης χαρακτηριστικών περνά στη συνέχεια από ένα επίπεδο προσοχής, το οποίο δίνει επιλεκτικά έμφαση σε ορισμένα χαρακτηριστικά και καταστέλλει άλλα και αποσκοπεί στην αύξηση της ακρίβειας και της ευρωστίας του δικτύου. Τέλος, η έξοδος του στρώματος προσοχής περνά από ένα επίπεδο ενεργοποίησης, το οποίο εφαρμόζει μια μη γραμμική συνάρτηση στην είσοδο, συγκεκριμένα την Swish, για να εισαγάγει μη γραμμικότητα στο δίκτυο.



Σχήμα 4.15: Η αρχιτεκτονική του δικτύου Wide MNIST.

4.2.2 EfficientNet

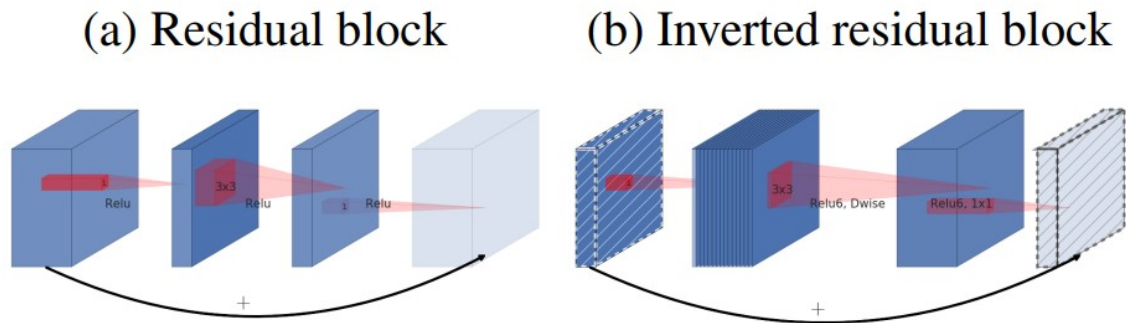
Το 2019, οι μηχανικοί Mingxing Tan και Quoc V. Le της Google Brain Team, δημοσίευσαν το άρθρο “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks” [48]. Βασικός στόχος της δημοσίευσης ήταν η εισαγωγή μιας αποτελεσματικής στρατηγικής κλιμάκωσης νευρωνικών δικτύων παρομοία με την στρατηγική που αναλύθηκε στο παραρτήμα του YoloV7, παράλληλα όμως εισήγαγαν μία νέα οικογένεια νευρωνικών δικτύων, τα EfficientNets (B0-B7). Τα EfficientNet σχεδιαστηκαν με βασικό γνώμονα την βελτιστη απόδοση και την ελαχιστη πολυπλοκότητα. Για παράδειγμα ο μοντέλο ResNet50 έχει 23.534.592 παραμέτρους συνολικά, και παρόλα αυτά, εξακολουθεί να υπολείπεται σε απόδοση το μικρότερο EfficientNet (που ονομάζεται EfficientNet-B0), το οποίο έχει μόνο 5.330.564 παραμέτρους συνολικά.

Η αρχιτεκτονική EfficientNet-B0 που αποτελεί το μοντέλο-βάση για την ανάπτυξη των υπολοίπων EfficientNets (μέσω κλιμακώσης), περιέχει έντεκα εκατομμύρια παραμέτρους προς εκπαίδευση και έχει ως δομικά στοιχεία τη συνάρτηση ενεργοποίησης Swiss, τα Inverted Residual Blocks και τα Squeeze & Excitation Blocks.

Inverted Residual Block

Το inverted residual block προτάθηκε στην αρχιτεκτονική MobileNet [49] η οποία χρησιμοποιεί depthwise separable convolution εντός του block. Προκειται για την ακραία περίπτωση συνελίξης κατά γκρουπ όπου ο αριθμός των γκρουπ ισούται με των αριθμω των καναλιών εισόδου και ακολουθείται από 1x1 συνελίξη για συνδυασμό των χαρακτηριστικών. Στα συμβατικά residual blocks τα skip connections χρησιμοποιούνται για την σύνδεση επιπέδων με μεγάλο βάθος, δηλαδή μεγάλο αριθμό καναλιών, ενώ εντός του block τα επίπεδα διαθέτουν μικρότερο βάθος. Τα inverted residual blocks, αντιθέτως, συνδέουν επίπεδα μικρού βάθους (bottlenecks). Μέσω αυτής της προσέγγισης μειώνονται σημαντικά οι παράμετροι προς εκπαίδευση. Αμέσως μετά το DSC ακολουθεί

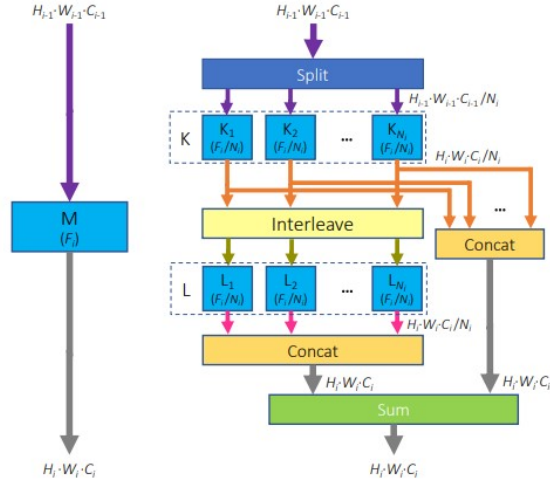
ο μηχανισμός προσοχής Squeeze-excitation με σκοπό να αποδωθεί ένας παραγοντας σημαντικότητας στους χάρτες χαρακτηριστικών πριν σμικρυνθούν ξανά στο τέλος του inverted block.



Σχήμα 4.16: Ένα Inverted Residual block. Τα σκιαγραφημένα τμήματα αναπαριστούν επίπεδα χωρίς συναρτήσεις ενεργοποίησης.

K-EffNet-B0

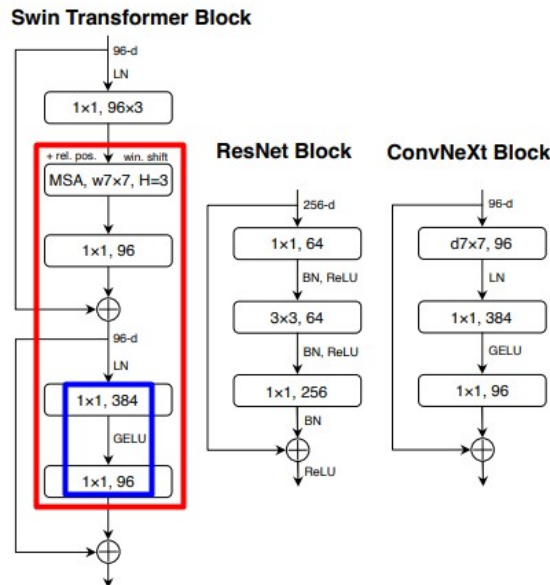
Καθώς το μεγαλύτερο τμήμα των παραμέτρων και της πολυπλοκότητας βρίσκεται εντός των πολλαπλών ποιντωσε (1x1) συνελίξεων, οι Schwarz Schuler et al. προτείνουν μια μέθοδο αντικατάστασής τους με σκοπό τη μείωση των παραμέτρων του δικτύου [50]. Συγκεκριμένα, κάθε pointwise συνέλιξη (1x1) αντικαθίσταται από ένα επίπεδο παράλληλων συνελίξεων κατά K γκρουπ, και οι εξαγόμενοι χάρτες χαρακτηριστικών αναμειγνύονται πριν περάσουν από ένα επίπεδο παράλληλων συνελίξεων κατά L γκρουπ. Όλοι οι χάρτες χαρακτηριστικών, τόσο από το επίπεδο K όσο και από το επίπεδο L, συνδέονται κατ' αλληλουχία και τελικά αθροίζονται όλοι μαζί. Οι συνελίξεις κατά γκρουπ πάντα μειώνουν σημαντικά την πολυπλοκότητα, και η χρήση δύο τέτοιων επιπέδων διατηρεί σε σημαντικό βαθμό την πληροφορία που χάνεται λόγω της αντικατάστασης της εκάστοτε pointwise συνελίξης. Οι παράμετροι αυτής της έκδοσης του EfficientNet-B0 είναι κατά 85% λιγότεροι, και συγκεκριμένα 639,702. Αυτή η σημαντική ελαχιστοποίηση της πολυπλοκότητας καθιστά αυτό το δίκτυο ιδανικό για δοκιμές σε edge συσκευές.



Σχήμα 4.17: Αριστερά μια συνέλιξη pointwise και δεξιά η αντικατάσταση που περιγράφηκε παραπάνω.

4.2.3 ConvNext

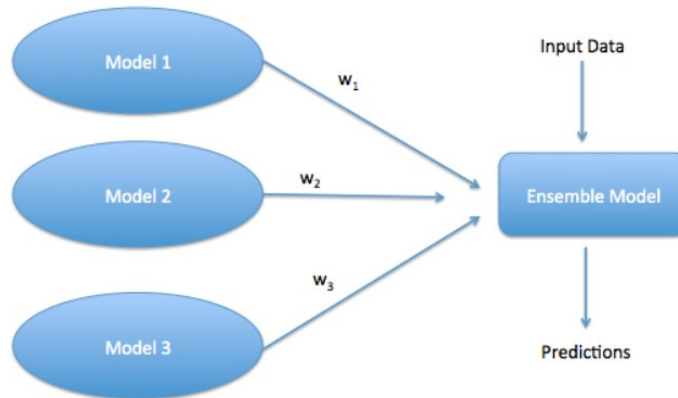
Με γνώμονα την τεράστια επιτυχία των transformers σε όλο το φάσμα της υπολογιστικής όρασης, οι ερευνητές της Meta Research θέλησαν να εκμοντερνίσουν την αρχιτεκτονική ενός απλού ResNet-50 δανειζόμενοι ιδέες από το Swin Transformer. Ένα σημαντικό χαρακτηριστικό του Swin Transformer είναι η μη τοπική προσοχή του δικτύου χρησιμοποιώντας μεγαλύτερους πυρήνες στις συνελιξείς επεκτείνοντας έτσι το δεκτικό πεδίο του μοντέλου. Έτσι σε κάθε συνελκτικό inverted block αντικαθιστώνται οι πυρήνες των depthwise separable συνελιξέων με πυρήνα 7x7 και μετακινείται στην κορυφή του μπλόκ για την μείωση της πολυπλοκότητας. Άλλες σχεδιαστικές ιδιαιτερότητες είναι η αντικατάσταση της RELU με GELU, η χρήση λιγότερων συναρτήσεων ενεργοποίησης εντός δικτύου και τέλος η χρήση λιγότερων επίπεδων κανονικοποίησης [51].



Σχήμα 4.18: Εσωτερική δομή του ConvNeXt Block και ομοιότητες του με την δομή του Swin Transformer.

4.2.4 Ensemble Network

Οι μέθοδοι ensemble στη μηχανική μάθηση συνδυάζουν τις γνώσεις που λαμβάνονται από πολλαπλά μοντέλα μάθησης για να διευκολύνουν τη λήψη ακριβών και βελτιωμένων προβλεψεων [52]. Στην προκειμένη περίπτωση τα δυο καλύτερα σε προβλεψεις δικτυα θα επανεκπαιδευθούν με ένα επιπεδο Average ώστε να ληφθεί η μεση τιμη των προβλεψεων των δυο μοντελων και δυνητικα να βελτιωθεί η ακριβεια των προβλεψεων.



Σχήμα 4.19: Δομή ενός δικτύου Ensemble που βασίζεται σε 3 επιμέρους μοντέλα.

4.2.5 Εκπαίδευση

Η εκπαίδευση όλων των μοντέλων υλοποιήθηκε στο πλαίσιο νευρωνικών δικτύων TensorFlow. Οι υπερπαράμετροι εκπαίδευσης επιλέχθηκαν ίδιοι για όλα τα μοντέλα. Ως συνάρτηση απώλειας επιλέχθηκε η Binary Cross Entropy loss, λόγω του ότι το πρόβλημα είναι δυαδικής ταξινόμησης, με αλγόριθμο βελτιστοποίησης Adam, αρχικό ρυθμό μάθησης $3e-4$ και μέγεθος δέσμης 64. Για τον προγραμματισμό του ρυθμού μάθησης επιλέχθηκε η τεχνική ReduceLROnPlateau, η οποία πολλαπλασιάζει τον ρυθμό μάθησης κατά έναν επιλεγόμενο παράγοντα αν η μετρική που παρακολουθείται μείνει σταθερή για έναν αυθαίρετο αριθμό epoch. Συγκεκριμένα, ο ρυθμός μάθησης πολλαπλασιάζεται με 0.3162 αν η ακρίβεια στο σύνολο επαλήθευσης μείνει σταθερή για 10 epoch. Αντί για μέθοδο πρόωρης διακοπής της εκπαίδευσης τύπου EarlyStopping, προτιμήθηκε η αποθήκευση των καλύτερων μοντέλων μέσω ModelCheckpoint.

```
[ ] metrics = ["accuracy", "Precision", "AUC", "FalsePositives", "FalseNegatives", "TruePositives", "TrueNegatives"]
    model.compile(loss = "binary_crossentropy", optimizer = Adam(learning_rate = 3e-4, amsgrad=True), metrics = metrics)

[ ] reduce_lr = ReduceLROnPlateau(monitor='val_accuracy', factor=0.3162,
    patience=10, min_lr=3e-6)

[ ] filepath = "/mydrive/malariadetect/finalmodels/" + "..."

    checkpoint = ModelCheckpoint(filepath, monitor = "val_accuracy", save_best_only=True, verbose = 1, mode = "max")

[ ] batch_size = 64

[ ] train_image_gen = image_gen.flow_from_directory(train_path, target_size = image_shape[:2],
    color_mode = "rgb",
    batch_size = batch_size,
    class_mode = "binary")
```

Σχήμα 4.20: Ενδεικτικά υπερπαράμετροι εκπαίδευσης των δικτύων.

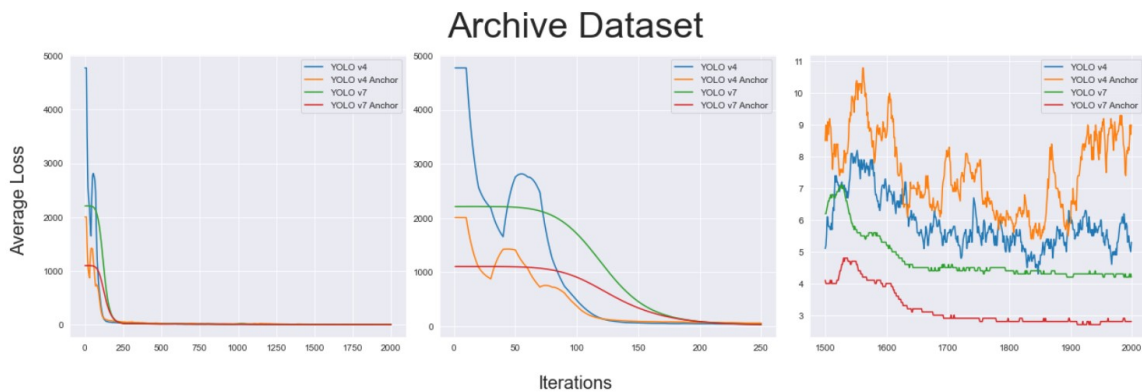
Κεφάλαιο 5

Αποτελέσματα Νευρωνικών δικτύων

5.1 Ανίχνευση ερυθροκυττάρων σε επίχρισμα αίματος

5.1.1 Μετρικές στο σύνολο εικόνων εκπαίδευσης

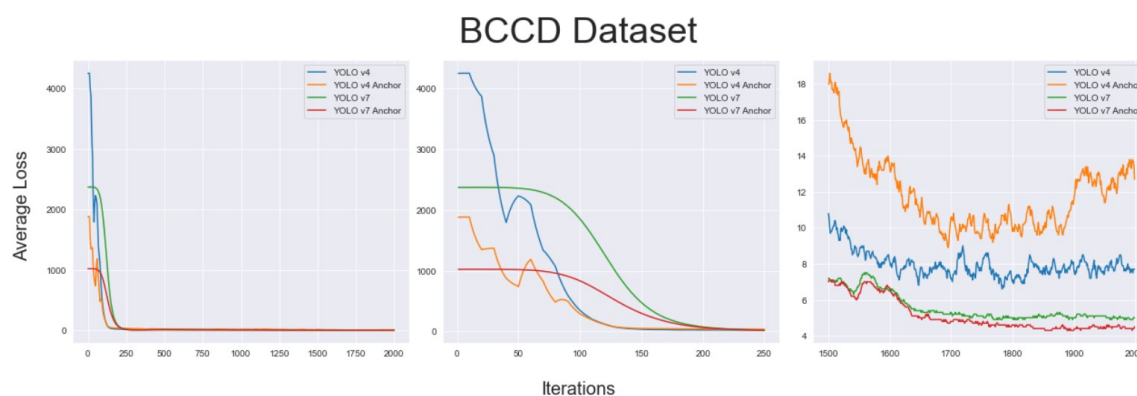
Σε αυτό το σημείο θα γίνει σύγκριση της καμπύλης απώλειας των τεσσάρων αρχιτεκτονικών που εκπαιδεύτηκαν στα δύο σύνολα δεδομένων "archive" και "BCCD dataset". Οι αρχιτεκτονικές είναι YoloV4, YoloV4 με επαναυπολογισμό αρχικών κελιών ανίχνευσης, YoloV7 και YoloV7 με επαναυπολογισμό αρχικών κελιών ανίχνευσης, εκπαιδευμένες για 2000 επαναλήψεις λόγω των περιορισμών υλικού, όπως αναφέρθηκε και στο κεφάλαιο 5.



Σχήμα 5.1: Καμπύλες απώλειας των τεσσάρων αρχιτεκτονικών Yolo στο Archive Dataset. Στο πρώτο διάγραμμα παρουσιάζεται η συνολική καμπύλη καθ' όλες τις 2000 επαναλήψεις. Στο δεύτερο οι πρώτες 250 και στο τρίτο οι τελευταίες 500.

Πίνακας 5.1: Μέσο σφάλμα στο Archive Dataset ανα 1000 επαναλήψεις

Method \ Iter	0	1000	2000
• YoloV4	4773.9	9.4	5.3
• YoloV4-A	2005.0	14.2	4.2
• YoloV7	2208.4	17.6	9.0
• YoloV7-A	1099.1	9.9	2.8



Σχήμα 5.2: Καμπύλες απώλειας των τεσσάρων αρχιτεκτονικών Yolo στο BCCD Dataset. Στο πρώτο διάγραμμα παρουσιάζεται η συνολική καμπύλη καθ' όλες τις 2000 επαναλήψεις. Στο δεύτερο οι πρώτες 250 και στο τρίτο οι τελευταίες 500.

Πίνακας 5.2: Μέσο σφάλμα στο BCCD Dataset ανα 1000 επαναληψεις

Method \ Iter	0	1000	2000
• YoloV4	4251.3	12.3	7.7
• YoloV4-A	1882.2	21.4	12.70
• YoloV7	2371.7	11.6	5.0
• YoloV7-A	1023.0	13.5	4.5

Όπως ήταν αναμενόμενο, οι εξατομικευμένες εκδόσεις τόσο του YoloV4 όσο και του YoloV7 συγκλίνουν γρηγορότερα σε ικανοποιητικές τιμές κελιών και κατά συνέπεια σε μικρότερο σφάλμα σε σύγκριση με τις αρχικές εκδόσεις τους. Τα επαναρχικοποιημένα κελιά των αλγορίθμων συμβάλλουν στην έναρξη της εκπαίδευσης με χαμηλότερες τιμές σφάλματος κατά περίπου 150% στην περίπτωση του YoloV4 και 100% στην περίπτωση του YoloV7.

Στο τρίτο, καθώς και ειδικότερα στο δεύτερο διάγραμμα, είναι εμφανής η ομαλότητα της καμπύλης απώλειας της έβδομης έκδοσης του αλγορίθμου σε σχέση με την τέταρτη.

Το φαινόμενο αυτό οφείλεται στην αποτελεσματικότερη ροή κλίσεων κατά την οπισθο-διαδότηση της αρχιτεκτονικής ELAN. Λόγω της προσεγγισμένης σχεδίασης της μικρότερης και μεγαλύτερης διαδρομής κλίσεων εντός του δικτύου, οι κλίσεις αλλάζουν ομαλά και με συνέπεια, καθώς ενημερώνονται οι παράμετροι του δικτύου. Μικρές αλλαγές στις παραμέτρους επιφέρουν μικρές και προβλέψιμες αλλαγές στις κλίσεις.

Στο τρίτο διάγραμμα είναι εμφανές ότι το μέσο σφάλμα του YoloV7 με επαναρχικοποιημένα κελιά έχει το μικρότερο μέσο σφάλμα από τις τέσσερις εκδόσεις κατά την ολοκλήρωση της εκπαίδευσης, με τιμή 2.8, περίπου 30% χαμηλότερη από την αντίστοιχη του YoloV4. Ειδικότερα, η τελική τιμή σύγκλισης του YoloV4 επιτυγχάνεται 35% πιο γρήγορα, δηλαδή σε περίπου 750 επαναλήψεις, από τον εξατομικευμένο YoloV7, κάτι που επιδεικνύει τη μεγάλη διαφορά στην απόδοση μετά το πέρας αρκετών επαναλήψεων. Το γεγονός ότι σε μικρότερο αριθμό επαναλήψεων (π.χ. 100) ο YoloV4 συγκλίνει γρηγορότερα σε χαμηλότερες τιμές μέσου σφάλματος αποδίδεται στον απλούστερο σχεδιασμό του, καθώς το YoloV7 δημιουργήθηκε με γνώμονα πολυπλοκότερα προβλήματα ανίχνευσης.

Αξίζει σε αυτό το σημείο να σημειωθεί ότι σε αυθαίρετο αριθμό επαναλήψεων όλοι οι αλγόριθμοι θα προσεγγίζουν μια τιμή μέσου σφάλματος μικρότερη του 1, με τον YoloV7 όμως να επιτυγχάνει αυτό το αποτέλεσμα πολύ γρηγορότερα και αποδοτικότερα.


Ο αλγόριθμος Yolo και ειδικότερα η έβδομη έκδοση επιφέρει ικανοποιητικά αποτελέσματα σε μόλις 2000 επαναλήψεις τόσο στο Archive Dataset όσο και στο BCCD Dataset. Η διαφορά των τελικών τιμών μέσου σφάλματος, δηλαδή 2.8 και 4.5 (περίπου 62%), αποδίδεται στις έντονες επικαλύψεις ερυθροκυττάρων που παρουσιάζει το BCCD Dataset και, κατά συνέπεια, δυσκολεύει την εκπαίδευση.

5.1.2 Μετρικές στο σύνολο εικόνων επαλήθευσης

Η βασική μετρική που χρησιμοποιείται για την αξιολόγηση αποτελεσμάτων ενός αλγορίθμου ανίχνευσης αντικειμένων είναι το Mean Average Precision (Μέση Ακρίβεια). Ο πολύπλοκος υπολογισμός του βασίζεται σε διάφορες μετρικές που πρέπει να αναλυθούν ξεχωριστά.

Intersection Over Union (IOU)

Πρόκειται για μια μετρική της επικάλυψης του προβλεπόμενου οριοθετημένου πλαισίου με το αληθές οριοθετούμενο πλαίσιο της εκάστοτε εικόνας του συνόλου δεδομένων. Υψηλότερο IOU υποδεικνύει ότι οι προβλεπόμενες συντεταγμένες παρομοιάζουν επιτυχέστερα τις αληθείς συντεταγμένες. Είναι το αποτέλεσμα του πηλίκου του εμβαδού επικάλυψης με το εμβαδόν τομής των δύο πλαισίων.


$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$

Precision

Η μετρική precision υπολογίζει τη δυνατότητα του αλγορίθμου να ανιχνεύει αληθώς θετικά αποτελέσματα ως προς τα συνολικά θετικά αποτελέσματα, δηλαδή το άθροισμα των αληθώς και ψευδώς θετικών.

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive}$$

Στο πλαίσιο των δικτύων ανίχνευσης αντικειμένων υπολογίζεται μέσω ενός κατωφλίου IOU (IOU threshold) που συνήθως είναι η τιμή 0.5. Για παράδειγμα, αν το IOU μεταξύ του οριοθετημένου πλαισίου μιας ανίχνευσης και του αντίστοιχου αληθούς οριοθετημένου πλαισίου είναι μεγαλύτερο του 0.5, τότε λογίζεται ως θετικό. Πρακτικά, το precision είναι μια μετρική για το πόσο σωστά προβλέπει ο αλγόριθμος και υποδεικνύει πόσο μπορεί να βασιστεί κανείς στις θετικές προβλέψεις του μοντέλου.

Recall

Η μετρική recall υπολογίζει τη δυνατότητα του αλγορίθμου να ανιχνεύει αληθώς θετικά αποτελέσματα ως προς τα συνολικά αποτελέσματα που θα έπρεπε να είναι θετικά, δηλαδή το άθροισμα των αληθώς θετικών και ψευδώς αρνητικών δειγμάτων.

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative}$$

Πρακτικά, το recall είναι μια μέτρηση του πόσο καλά ανιχνεύει ο αλγόριθμος όταν θα έπρεπε να ανιχνεύσει. Υποδεικνύει τις προβλεπτικές αστοχίες του μοντέλου.

Mean Average Precision

Όταν ένα μοντέλο παρουσιάζει υψηλό recall αλλά χαμηλή ακρίβεια, αυτό σημαίνει ότι κατηγοριοποιεί σωστά τα περισσότερα θετικά δείγματα, αλλά έχει πολλά ψευδώς θετικά αποτελέσματα. Αντίθετα, όταν ένα μοντέλο έχει υψηλή ακρίβεια και χαμηλό recall, αυτό σημαίνει ότι το μοντέλο είναι ακριβές στην κατηγοριοποίηση δειγμάτων ως θετικά, αλλά χάνονται πολλές περιπτώσεις θετικών δειγμάτων. Αυτή η συμπεριφορά περιγράφεται από την καμπύλη precision-recall του εκάστοτε μοντέλου.

Το εμβαδόν αυτής της καμπύλης είναι η μετρική Average Precision (AP) (μέση ακρίβεια). Ανάλογα με τα συμφραζόμενα του προβλήματος, το Average Precision υπολογίζεται πάνω σε όλες τις κλάσεις και από τον μέσο όρο προκύπτει το Mean Average Precision (mAP), ενώ σε άλλες περιπτώσεις το AP έχει ίδια τιμή με το mAP. Στην περίπτωση του αλγορίθμου Yolo, το mAP έχει την ίδια τιμή με το AP όταν έχουμε μία κλάση ανίχνευσης (όπως στο archive dataset), ενώ ισούται με τον μέσο όρο των εκάστοτε AP για περισσότερες κλάσεις (όπως στο BCCD dataset).

Το mAP υπολογίζεται αυτόματα κάθε 100 επαναλήψεις της εκπαίδευσης του μοντέλου. Ωστόσο, λόγω περιορισμών του υλικού (hardware), δεν ήταν δυνατός ο αυτόματος υπολογισμός κάθε 100 επαναλήψεις και έτσι υπολογίστηκε χειροκίνητα στις μόνο στις 1000 και 2000 επαναλήψεις

Πίνακας 5.3: mAP στο σύνολο επαλήθευσης του Archive Dataset

Method \ Iter	1000	2000
• YoloV4	98.79%	98.71%
• YoloV4-A	98.97%	98.73%
• YoloV7	93.09%	99.01%
• YoloV7-A	98.91%	99.47%

Πίνακας 5.4: mAP στο σύνολο επαλήθευσης BCCD Dataset

Method \ Iter	1000	2000
• YoloV4	91.71%	89.71%
• YoloV4-A	80.51%	87.48%
• YoloV7	79.71%	88.15%
• YoloV7-A	73.71%	84.47%

Από τα αποτελέσματα συμπαίρνουμε ότι τα μοντέλα που εκπαιδεύτηκαν για τη βάση δεδομένων BCCD χρήζουν περαιτέρω εκπαίδευσης για να μειωθεί το μέσο σφάλμα και παράλληλα να αυξηθεί το mAP. Η διαφορά στην ακρίβεια ανάμεσα στις προβλέψεις στο BCCD και στο Archive έγκειται στο γεγονός ότι τα ερυθροκύτταρα στο BCCD εμφανίζουν μεγάλη επικάλυψη μεταξύ τους, δυσκολεύοντας έτσι την προβλεπτική ικανότητα των μοντέλων. Να τονιστεί ξανά ότι λόγω περιορισμών στο υλικό δεν μπορούσε να υλοποιηθεί περαιτέρω εκπαίδευση, γιατί χωρίς τον παράλληλο υπολογισμό μετρικών στο σύνολο επαλήθευσης είναι αδύνατο να αξιολογηθεί η πορεία της εκπαίδευσης και να εξαχθούν τα μοντέλα που εμφανίζουν την υψηλότερη ακρίβεια προβλέψεων.

5.1.3 Μετρικές στο σύνολο εικονων δοκιμών

Το μοντελο προβλεπει αποτελεσματα στο συνολο εικονων δοκιμων παρουσιαζοντας και το ποσοστο εμπιστοσυνης , τα οποια εξαγονται σε ενα αρχειο συντεταγμενων JSON για περαιτερω αξιολογηση και αναλυση.Παλι θα χρησιμοποιηθει η μετρικη Mean Average Precision για την αξιολογηση και τον σχολιασμο των αποτελεσματος σε καθε μοντελο ξεχωριστα.Παραλληλα αναμενεται τα βελτιστα αποτελεσματα να προκυψουν απο το μοντελο που παρουσιασε το υψηλοτερο mAP στο συνολο εικονων επαλυθευσης δηλαδη το YoloV7 στην περιπτωση του Archive Dataset και το YoloV4 στην περιπτωση του BCCD Dataset

Πίνακας 5.5: mAP στο σύνολο δοκιμών του Archive Dataset

Method \ Iter	1000	2000
• YoloV4	90.06%	90.1%
• YoloV4-A	89.18%	91.18%
• YoloV7	89.81%	92.23%
• YoloV7-A	90.39%	91.1%

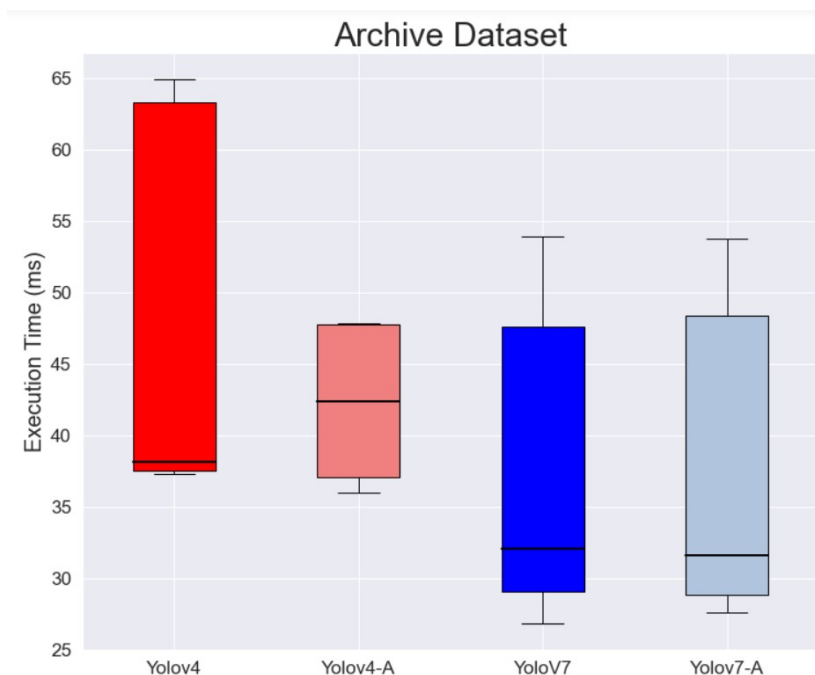
Πίνακας 5.6: mAP στο σύνολο δοκιμών BCCD Dataset

Method \ Iter	1000	2000
• YoloV4	90.14%	89.48%
• YoloV4-A	83.03%	89.53%
• YoloV7	83.01%	87.25%
• YoloV7-A	79.66%	83.81%

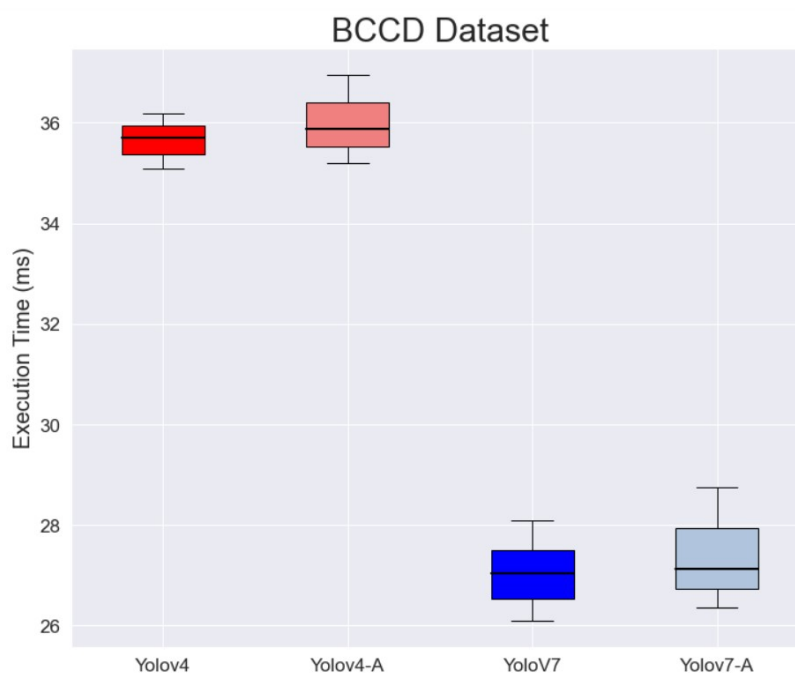
Τα αποτελέσματα του καλύτερου μοντέλου σε κάθε dataset εξάγονται σε μορφή συντεταγμένων Yolo. Στη συνέχεια, υπολογίζεται ο δείκτης Intersection over Union (IOU) για να συγκριθούν οι συντεταγμένες των προβλέψεων με τις συντεταγμένες του ground truth και να αντιστοιχιστεί κάθε πρόβλεψη με το σωστό κύτταρο. Στη συνέχεια, γίνεται περικοπή των εικόνων με βάση τις συντεταγμένες αυτές, ώστε το λεπτό επίχρισμα αίματος να διαιρείται σε επιμέρους κύτταρα για μετέπειτα πρόβλεψη στο κομμάτι των αλγορίθμων ταξινόμησης.

Χρόνος Εκτέλεσης

Σε αυτό το σημείο θα παρουσιαστούν οι χρόνοι εκτέλεσης των μοντέλων τόσο στο σύνολο δοκιμών του BCCD Dataset όσο και στο Archive Dataset όσον αφορά τον χρόνο που πήρε για προβλέψεις στην κάθε εικόνα. Οι ακραίες τιμές στην έναρξη των προβλέψεων απαλείφονται διότι αφορούν τον χρόνο που χρειάζεται η GPU για να ξεκινήσει να λειτουργεί στη μέγιστη δυνατότητά της.



Σχήμα 5.3: Χρονος εκτέλεσης των τεσσάρων μοντελών στο σύνολο δοκιμών του Archive Dataset



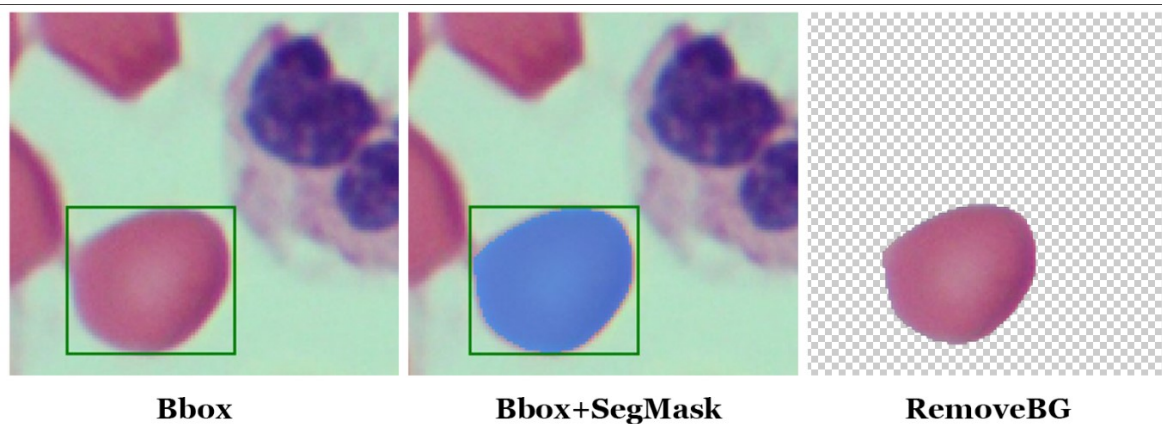
Σχήμα 5.4: Χρονος εκτέλεσης των τεσσάρων μοντελών στο σύνολο δοκιμών του BCCD Dataset

Όπως αναμενόταν, ο χρόνος εκτέλεσης δοκιμών (inference time) των πρόσφατων αλγορίθμων YOLOv7 είναι αισθητά μικρότερος από τον YOLOv4.

5.1.4 Μετάβαση από οριοθετούμενα πλαίσια σε δυαδικές μάσκες

Σε περίπτωση που οι ανάγκες του προβλήματος απαιτούν κατακερματισμό εικόνας σε επίπεδο δυαδικής μάσκας τότε μπορεί να χρησιμοποιηθεί ο αλγόριθμος SegmentAnything (SAM) [53] της ερευνητικής ομάδας Meta Research. Πρόκειται για έναν αλγόριθμο κατακερματισμού προεκπαιδευμένο σε πάνω από 11 εκατομμύρια εικόνες και σε

πάνω από 1 δισεκατομμύριο δυαδικές μάσκες. Η ιδιαιτερότητα αυτού του αλγορίθμου κατακερματισμού εγκείται στο γεγονός ότι δέχεται μια δευτερεύουσα είσοδο είτε σε φυσική γλώσσα είτε σε μορφή συντεταγμένων οριοθετημένων πλαισίων, και βάσει αυτής της παραπλησίας εισόδου συγκεκριμενοποιείται το τμήμα της εικόνας που θα υποστεί κατακερματισμό. Στην προκειμένη περίπτωση, οι παραχθείσες συντεταγμένες του εκάστοτε αλγορίθμου. Στην προκειμένη περίπτωση οι παραγομενες συντεταγμενες του εκαστοτε αλγοριθμου Yolo μπορούν να τροφοδοτηθούν ως δευτερευουσα εισοδο στο SAM μαζί με την εικονα και ετσι να δημιουργηθει η δυαδικη μασκα του κυτταρου.



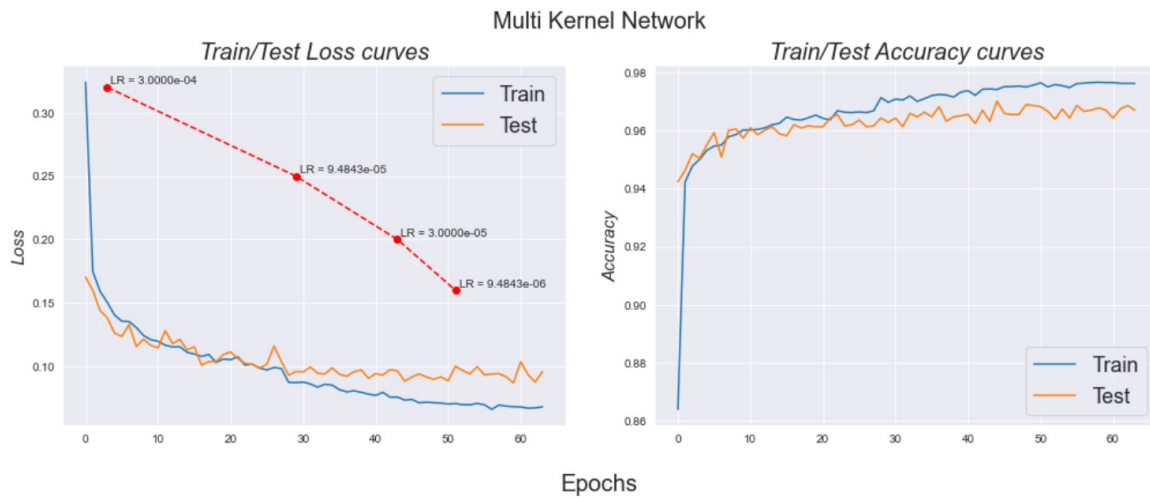
Σχήμα 5.5: Αριστερά βλέπουμε το παραγόμενο οριοθετούμενο πλαίσιο από τον YoloV7 ,στη μέση το οριοθετούμενο πλαίσιο μαζί με την δυαδική μάσκα που προέκυψε ως αποτέλεσμα του SAM και τέλος την απομονωση του μεμονομένου κυττάρου απο την υπόλοιπη εικόνα

5.2 Ταξινόμηση ερυθροκυττάρων σε επιμολυσμένα ή μη απο το πλασμάδιο της ελονοσίας

5.2.1 Μετρικές στο σύνολο εικονων εκπαιδευσης και επαληθευσης

Για τα τέσσερα δίκτυα που εκπαιδεύτηκαν σε αυτό το πρόβλημα, τα οποία είναι το Wide-Multi-kernel network, το k-Effnet-B0, το ConvNeXt και το Ensemble Network, θα συγκριθούν οι καμπύλες απωλειών καθώς και οι καμπύλες ακρίβειας (accuracy) στο σύνολο εικόνων εκπαίδευσης και επαλήθευσης, για να αξιολογηθεί η συμπεριφορά των μοντέλων.

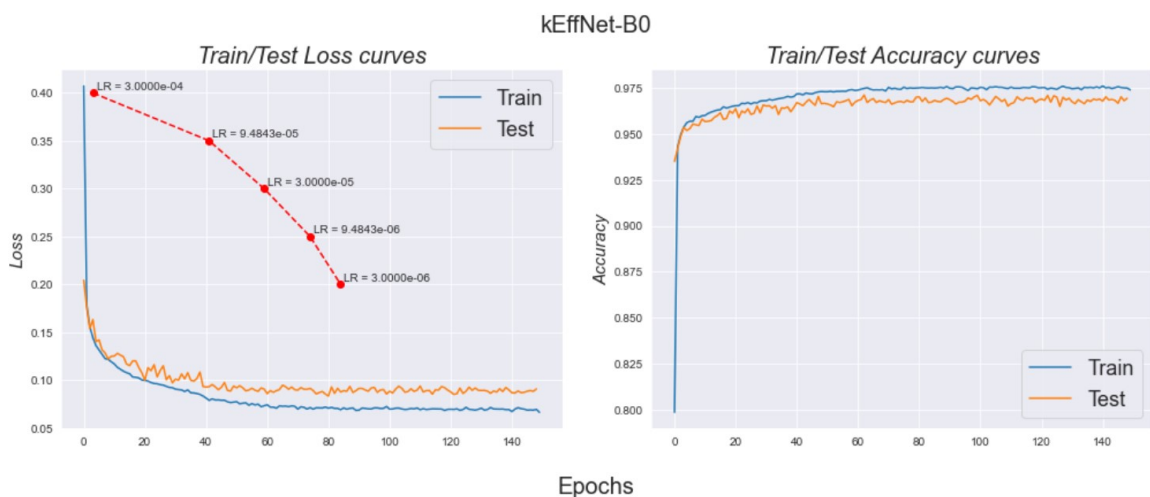
Multi Kernel Network



Σχήμα 5.6: Καμπύλες απώλειας και ακρίβειας για το μοντέλο Multi Kernel. Η κόκκινη ευθεία αναπαριστά την πτώση του ρυθμού μάθησης με βάση τον προγραμματισμό του.

Παρατηρείται ότι η καμπύλη σφάλματος δοκιμών παρουσιάζει ένα μικρό underfitting έως τα 10 epoch, έπειτα προσαρμόζεται ορθά στην καμπύλη σφάλματος εκπαίδευσης και μειώνεται παράλληλα με αυτήν. Ωστόσο μετά τα 40 epochs το σφάλμα επαλήθευσης παρουσιάζει ανοδική πορεία σε αντίθεση με το σφάλμα εκπαίδευσης που συνεχίζει να μειώνεται και η ίδια συμπεριφορά εμφανίζεται και στα διαγράμματα ακρίβειας. Αυτό αποτελεί ένδειξη overfitting και έτσι θα χρησιμοποιηθούν τα καλύτερα βάρη του μοντέλου που αποθηκεύτηκαν πριν την έναρξη αυτού του φαινομένου.

K-EffNet-B0

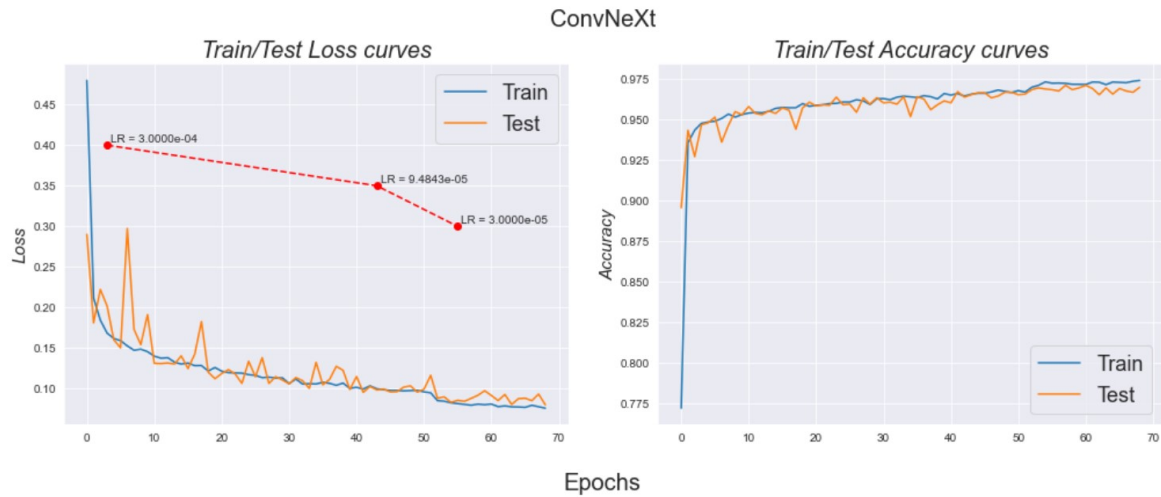


Σχήμα 5.7: Καμπύλες απώλειας και ακρίβειας για το μοντέλο K-EffNet-B0. Η κόκκινη ευθεία αναπαριστά την πτώση του ρυθμού μάθησης με βάση τον προγραμματισμό του.

Όπως φαίνεται, η καμπύλη ακρίβειας επαλήθευσης σταματά να βελτιώνεται μετά τα περίπου 60 epoch εκπαίδευσης, ενώ το σφάλμα επαλήθευσης αυξάνεται ελάχιστα μετά τα 40 epoch. Αυτή η συμπεριφορά στο σφάλμα υποδεικνύει ότι παρόλο που το μοντέλο είναι σε θέση να προβλέπει καλύτερα, ο δείκτης εμπιστοσύνης στις προβλέψεις του είναι

ελαφρώς χαμηλότερος, καθώς το σφάλμα γροσς εντροπψ αυξάνεται. Παρόλα αυτά, επειδή η απόκλιση είναι ελάχιστη και δεν παρουσιάζεται έντονο overfitting έως τα 60 epoch, θα χρησιμοποιηθούν για προβλέψεις τα καλύτερα βάρη έως τα 60 epoch.

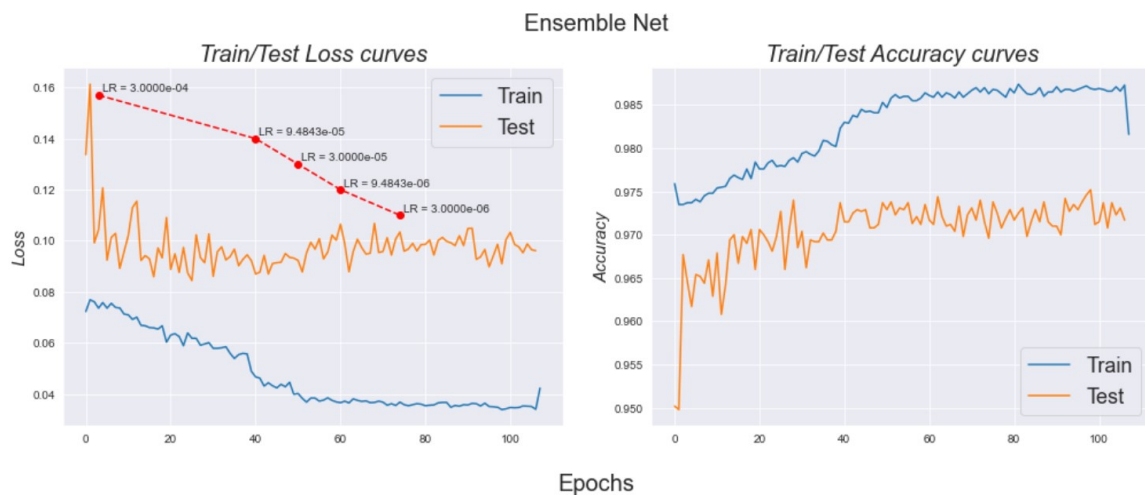
ConvNeXt



Σχήμα 5.8: Καμπύλες απώλειας και ακρίβειας για το μοντέλο ConvNeXt. Η κόκκινη ευθεία αναπαριστά την πτώση του ρυθμού μάθησης με βάση τον προγραμματισμό του.

Σχεδόν σε όλο τον πλήρη κύκλο της εκπαίδευσης παρατηρείται πολύ καλή προσαρμογή των καμπυλών σφάλματος/ακρίβειας, ενώ γύρω στα 60 epoch ξεκινά μια ανοδική πορεία του σφάλματος και μια καθοδική της ακρίβειας επαλήθευσης. Θα χρησιμοποιηθούν τα καλύτερα βάρη έως τα 60 epoch.

Ensemble Network



Σχήμα 5.9: Καμπύλες απώλειας και ακρίβειας για το μοντέλο Ensemble. Η κόκκινη ευθεία αναπαριστά την πτώση του ρυθμού μάθησης με βάση τον προγραμματισμό του.

5.2.2 Μετρικές στο σύνολο εικονών δοκιμών

Οι μετρικές που θα συγκριθούν ώστε να αντληθούν συμπεράσματα για την προβλεπτική ικανότητα των μοντέλων που εκπαιδεύτηκαν είναι το Precision, Accuracy, Recall, Confusion Matrix, AUROC, F1Score. Παρακάτω θα αναλυθούν οι μετρικές οι οποίες δεν εμφανίστηκαν σε προηγούμενα βήματα του pipeline.

Επαύξηση δεδομένων στον χρόνο δοκιμών

Η επαύξηση χρόνου δοκιμών είναι μια εφαρμογή επαύξησης δεδομένων στο σύνολο δεδομένων δοκιμής. Συγκεκριμένα, περιλαμβάνει τη δημιουργία πολλαπλών επαυξημένων αντιγράφων κάθε εικόνας του συνόλου δοκιμών, την πραγματοποίηση μιας πρόβλεψης από το μοντέλο για το καθένα από αυτά και, στη συνέχεια, την επιστροφή ενός συνόλου αυτών των προβλέψεων. Συχνά, εκτελείται μία απλή επαύξηση σε χρόνο δοκιμής, όπως μετατόπιση, περικοπή ή αναστροφή εικόνας. Στην προκειμένη περίπτωση, χρησιμοποιείται συνδυασμός αυτών των τριών μεθόδων.

Στην εργασία τους το 2015 με τίτλο "Very Deep Convolutional Networks for Large-Scale Image Recognition" [54], οι συγγραφείς πραγματοποιούν επαύξηση χρόνου δοκιμής με οριζόντια αναστροφή. Επίσης, στην εργασία τους το 2015 με τίτλο "Rethinking the Inception Architecture for Computer Vision" [55], οι συγγραφείς της Google χρησιμοποιούν επαύξηση χρόνου δοκιμής με περικοπή, την οποία αναφέρουν ως αξιολόγηση πολλαπλών περικοπών.

Η διαδικασία των δοκιμών επαναλαμβάνεται 10 φορές με επαύξηση δοκιμών, και στο τέλος λαμβάνονται οι μέσοι όροι των προβλέψεων.

Confusion Matrix (Πίνακας σύγχυσης)

Αποτελεί έναν πίνακα που αναπαριστά τον αριθμό των δειγμάτων που ανιχνεύθηκαν σωστά. Για τη δημιουργία του χρειάζονται τέσσερις μετρικές.

- True Positives (Αληθώς θετικά)
- True Negatives (Αληθώς αρνητικά)
- False Positives (Ψευδώς Θετικά)
- False Negatives (Ψευδώς αρνητικά)

Στο παρών πρόβλημα ως θετικό δείγμα (1) λογίζεται ένα παρασιτισμένο κύτταρο ενώ ως αρνητικό (0) ένα υγιές ερυθροκύτταρο

		True Class	
		Parasitized	Uninfected
Predicted Class	Parasitized	2505	65
	Uninfected	95	2535

Σχήμα 5.10: Πίνακας σύγχυσης του δικτύου **Wide Multi-Kernel** στο σύνολο δοκιμών του Custom NIH Dataset

		True Class	
		Parasitized	Uninfected
Predicted Class	Parasitized	2492	57
	Uninfected	108	2543

Σχήμα 5.11: Πίνακας σύγχυσης του δικτύου **K-EffNet-B0** στο σύνολο δοκιμών του Custom NIH Dataset

		True Class	
		Parasitized	Uninfected
Predicted Class	Parasitized	2504	95
	Uninfected	96	2505

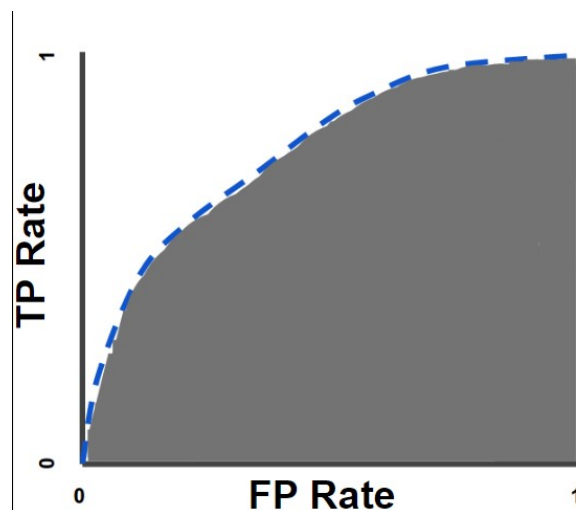
Σχήμα 5.12: Πίνακας σύγχυσης του δικτύου **ConvNeXt** στο σύνολο δοκιμών του Custom NIH Dataset

		True Class	
		Parasitized	Uninfected
Predicted Class	Parasitized	2525	65
	Uninfected	75	2535

Σχήμα 5.13: Πίνακας σύγχυσης του δικτύου **Ensemble** στο σύνολο δοκιμών του Custom NIH Dataset

AUROC

Η καμπύλη ROC (Receiver Operator Characteristic) είναι μια μετρική αξιολόγησης για προβλήματα δυαδικής ταξινόμησης. Πρόκειται για μια καμπύλη πιθανότητας που απεικονίζει το True Positive Rate (TPR) έναντι του False Positive Rate (FPR) σε κάθε τιμή κατωφλίου ταξινόμησης. Το TPR πρόκειται για την πιθανότητα ένα θετικό δείγμα να προβλεφθεί σωστά στη θετική κλάση ενώ το FPR για την πιθανότητα ένα αρνητικό δείγμα να προβλεφθεί εσφαλμένα στη θετική κλάση. Με άλλα λόγια, δείχνει την απόδοση ενός μοντέλου ταξινόμησης σε όλα τα κατώφλια ταξινόμησης. Η περιοχή κάτω από την καμπύλη (AUC) είναι το μέτρο της ικανότητας ενός δυαδικού ταξινομητή να διακρίνει μεταξύ κλάσεων και χρησιμοποιείται ως σύνοψη της καμπύλης ROC.



F1Score

Η ακρίβεια και η ανάκληση παρουσιάζουν έναν συμβιβασμό μεταξύ τους, δηλαδή η μία μετρική επιβαρύνει την άλλη. Η μεγαλύτερη ακρίβεια συνεπάγεται έναν αυστηρότερο κριτή (ταξινομητή) που αμφισβητεί ακόμη και τα πραγματικά θετικά δείγματα από το σύνολο δεδομένων, μειώνοντας έτσι τη βαθμολογία ανάκλησης. Από την άλλη πλευρά, περισσότερη ανάκληση συνεπάγεται έναν χαλαρό κριτή που επιτρέπει να περάσει κάθε δείγμα που μοιάζει με θετική κλάση, γεγονός που καθιστά αρνητικά δείγματα οριακών περιπτώσεων ταξινομημένα ως "θετικά", μειώνοντας έτσι την ακρίβεια. Ιδανικά,

Θέλουμε να μεγιστοποιήσουμε και τις δύο μετρικές προεξιον και ρεσαλλ για τον τέλειο ταξινομητή. Η βαθμολογία F1 συνδυάζει την ακρίβεια και την ανάκληση χρησιμοποιώντας τον αρμονικό μέσο όρο τους και η αύξηση της βαθμολογίας F1 συνεπάγεται την ταυτόχρονη αύξηση τόσο της ακρίβειας όσο και της ανάκλησης.Ωστοσο να τονιστει ο-τι το F1score ειναι ιδανικη μετρικη για προβληματα που παρουσιαζουν ανισορροπια κλασεων.Στην προκειμενη περιπτωση οι κλασεις ειναι τελεια ισορροπημενες ωστοσο η μετρικη θα υπολογιστει.

$$F1 = 2. \frac{Precision \times Recall}{Precision + Recall}$$

Πίνακας 5.7: Μετρικές των τεσσάρων δικτύων στο σύνολο δοκιμών του Custom-NIH Dataset

Metric Model	Accuracy	Precision	Recall	F1Score	AUROC
• Wide Multi Kernel	96.92%	97.47%	96.34%	96.91%	99.37%
• K-Effnet-B0	96.83%	97.76%	95.84%	96.8%	99.43%
• ConvNeXt	96.33%	96.34%	96.31%	96.33%	99.12%
• Ensemble	97.31%	97.49%	97.11%	97.3%	99.21%

Πίνακας 5.8: Ευαισθησία και Ειδικότητα των τεσσάρων δικτύων στο σύνολο δοκιμών του Custom-NIH Dataset

Metric Model	Sensitivity	Specificity
• Wide Multi Kernel	96.35%	97.5%
• K-Effnet-B0	95.85%	97.81%
• ConvNeXt	96.31%	96.35%
• Ensemble	97.12%	97.5%

5.2.3 Floating Point Operations (FLOPs) και μέσος χρόνος εκτέλεσης

Οι πράξεις κινητής υποδιαστολής (FLOPs) είναι η μονάδα που εκφράζει το υπολογιστικό κόστος. Μπορεί επίσης να χρησιμοποιηθεί για την αξιολόγηση του κόστους εκπαίδευσης και εκφράζει πόσες βασικές μαθηματικές πράξεις θα εφαρμοστούν στον αριθμό κινητής υποδιαστολής κατά την εκτέλεση του αλγορίθμου. Η βιβλιογραφία ωστοσο υποδεικνύει οτι FLOPs και χρονος εκτελεσης δεν εχουν παντα θετικη συσχετιση μεταξυ τους υπο την εννοια οτι αν ενα μοντελο εχει λιγοτερα FLOPs απο ενα αλλο δεν ειναι απαραιτητα και γρηγοροτερο [44].Καθως οι δοκιμές εχουν επαναληφθεί 10 φορές για το κάθε μοντέλων λόγω της επαυξησης δεδομενων στον χρονο δοκιμων θα ληφθει ο μεσος χρονος εκτελεσης αφαιρώντας την αρχική ακραία τιμη οπως και στην περίπτωση των αλγοριθμων Yolo.

Πίνακας 5.9: FLOPs και χρόνος εκτέλεσης των τεσσάρων δικτύων στο σύνολο δοκιμών του Custom-NIH Dataset

Metric Model	FLOPs	Execution Time
• Wide Multi Kernel	14.5 G	30 s
• K-Effnet-B0	0.0326 G	28 s
• ConvNeXt	9.92 G	38 s
• Ensemble	14.5 G	33 s

Επιβεβαιώνεται το γεγονός ότι η μέτρηση των FLOPs δεν είναι αρκετή για την πρόβλεψη του χρόνου εκτέλεσης ενός μοντέλου.

5.2.4 Μετρικές σε εξωτερικά σύνολα δεδομένων BCCD και Archive

Τέλος θα συγκριθεί η ευρωστία των μοντέλων και η δυνατότητά τους να προβλέπουν με ακρίβεια σε εξωτερικά σύνολα δεδομένων. Πρόκειται για το τελευταίο βήμα του pipeline όπου τα κατακερματισμένα ερυθροκύτταρα από τον αλγόριθμο YoloV7-anchor θα ταξινομηθούν ως επιμολυσμένα ή υγιή. Καθώς όλα τα κύτταρα των ασθενών των δύο αυτών συνόλων δεδομένων είναι υγιή, θα διαπιστωθεί η ειδικότητα των μοντέλων που είναι πολύ σημαντική μετρική σε προβλήματα εντοπισμού ασθενειών όπως είδαμε σε προηγούμενα κεφάλαια. Όπως αναλύθηκε και στο κεφάλαιο 4 η επίτευξη υψηλής ειδικότητας είναι δυσκολότερη από αυτήν της υψηλής ευαισθησίας λόγω της φύσης των δεδομένων και έχουν γίνει κινήσεις βελτιστοποίησης των εικόνων (προεπεξεργασία) και των μοντέλων (multi-kernel network) με σκοπό την επίλυση αυτού του ζητήματος. Τα επιμερούς κατακερματισμένα κύτταρα είναι χωρισμένα σε υποφακέλους που μπορούν να θεωρηθούν ως επιμερούς ασθενείς για τις ανάγκες του προβλήματος. Η αξιολόγηση θα γίνει τόσο σε επίπεδο κυττάρων όσο και σε επίπεδο ασθενών. Η κατάληξη "-PL" υποδηλώνει μέτρηση ακρίβειας σε επίπεδο ασθενών.

Πίνακας 5.10: Ακρίβεια στα εξωτερικά σύνολα δεδομένων

Metric Method	Archive	BCCD	Archive-PL	BCCD-PL
• Wide Multi Kernel	91.24%	97.78%	86.11%	83.3%
• K-Effnet-B0	83.87%	91.45%	83.33%	50%
• ConvNeXt	85.71%	83.2%	83.33%	22.2%
• Ensemble	82.94%	68.03%	86.1%	18.4%

Φαίνεται ξεκαθαρα ότι μικρες διαφορες των μοντελων στην ακριβεια σε επιπεδο κυτταρων δυναται να προκαλεσει χαωδεις διαφορες σε επιπεδο ασθενων ,καθως εως και ενα κυτταρο να υπολογισθει ως λανθασμενα επιμολυσμενο στο επιχρισμα του ασθενους, ταυτοχρονα και ο ασθενης θεωρεται επιμολυσμενος.Συνολικα τα καλυτερα αποτελεσματα σε καθε κατηγορια τα παρουσιαζει το προσαρμοσμενο δικτυο ,γεγονος που ταιριαζει και με τα προηγουμενα αποτελεσματα ερευνητων οπως ειδαμε στην βιβλιογραφικη ανασκοπηση.

Κεφάλαιο 6

Σχολιασμός και συμπεράσματα

Από τα αποτελέσματα της πειραματικής διαδικασίας προκύπτει ότι η χρήση τεχνικών βαθειάς μάθησης είναι μια πολλά υποσχόμενη μέθοδος για τη διάγνωση της ελονοσίας σε τριτοκοσμικές χώρες που πλήττονται από αυτήν. Οι αλγόριθμοι ανίχνευσης παρουσίασαν υψηλή ακρίβεια πάνω από 90% σε δεδομένα εικόνων χαμηλής ποιότητας χωρίς να έχουν εκπαιδευθεί στο μέγιστο των δυνατοτήτων τους. Ειδικότερα, ο YoloV7 ταιριάζει ιδανικά στις ανάγκες του προβλήματος καθώς εμφανίζει τον χαμηλότερο χρόνο εκτέλεσης. Οι αλγόριθμοι ταξινόμησης και ειδικότερα το προσαρμοσμένο δίκτυο εμφανίζουν εξαιρετικά αποτελέσματα ευαισθησίας και ειδικότητας 95% τόσο στο σύνολο δοκιμών της βάσης δεδομένων που έχουν εκπαιδευθεί όσο και σε εξωτερικά σύνολα δεδομένων, κάτι το οποίο υποδεικνύει τη δυνατότητά τους να γενικεύουν σε νέα δεδομένα. Το K-Effnet-B0 παρουσιάζει συγκρίσιμα αποτελέσματα με το προσαρμοσμένο δίκτυο με πολύ λιγότερες παραμέτρους και FLOPs. Μελλοντικά θα πρέπει τα μοντέλα να δοκιμαστούν σε συσκευές εδγε χαμηλού κόστους και ενέργειας, ενώ παράλληλα το εκάστοτε δίκτυο θα πρέπει δυνητικά να διακρίνει και τον τύπο του πλασμοδίου ως ένα πρόβλημα ταξινόμησης πολλών κλάσεων, ώστε να πληρεί όλες τις προϋποθέσεις διαγνωστικής μεθόδου που έχουν θεσπιστεί από τον Παγκόσμιο Οργανισμό Υγείας (ΠΟΥ).

Bibliography

- [1] W. H. Organization, *Malaria microscopy quality assurance manual*, version 2 ed. World Health Organization, 2016.
- [2] F.-A. E. Mbassi, G. Mombo-Ngoma, W. N. Ndoumba, E. K. Yovo, K. A. Eberhardt, D. E. Mbassi, A. A. Adegnika, S. T. Agnandji, M. K. Bouyou-Akotet, M. Ramharter, and R. Zoleko-Manego, "Performance of field's stain compared with conventional giemsa stain for the rapid detection of blood microfilariae in gabon," *The American Journal of Tropical Medicine and Hygiene*, vol. 107, no. 2, pp. 383 – 387, 2022. [Online]. Available: <https://www.ajtmh.org/view/journals/tpmd/107/2/article-p383.xml>
- [3] N. Tangpukdee, C. Duangdee, P. Wilairatana, and S. Krudsood, "Malaria diagnosis: A brief review," *The Korean journal of parasitology*, vol. 47, pp. 93–102, 07 2009.
- [4] C. Ohrt, Purnomo, M. A. Sutamihardja, D. Tang, and K. C. Kain, "Impact of Microscopy Error on Estimates of Protective Efficacy in Malaria-Prevention Trials," *The Journal of Infectious Diseases*, vol. 186, no. 4, pp. 540–546, 08 2002. [Online]. Available: <https://doi.org/10.1086/341938>
- [5] S. Shillcutt, C. Morel, C. Goodman, P. Coleman, D. Bell, C. Whitty, and A. Mills, "Cost-effectiveness of malaria diagnostic methods in sub-saharan africa in an era of combination therapy," *Bulletin of the World Health Organisation*, vol. 86, pp. 101–110, 02 2008.
- [6] C. Wongsrichanalai, J. Pornsilapatip, V. Namsiripongpun, H. K. Webster, A. Luccini, P. Pansamdang, H. Wilde, and M. Prasittisuk, "Acridine orange fluorescent microscopy and the detection of malaria in populations with low-density parasitemia," *The American Journal of Tropical Medicine and Hygiene*, vol. 44, no. 1, pp. 17 – 20, 1991. [Online]. Available: <https://www.ajtmh.org/view/journals/tpmd/44/1/article-p17.xml>
- [7] G. Adeoye and I. Nga, "Comparison of quantitative buffy coat technique (qbc) with giemsa-stained thick film (gtf) for diagnosis of malaria," *Parasitology International*, vol. 56, no. 4, pp. 308–312, 2007. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1383576907000943>
- [8] A. Moody, "Rapid diagnostic tests for malaria parasites," *Clinical Microbiology Reviews*, vol. 15, no. 1, pp. 66–78, 2002. [Online]. Available: <https://journals.asm.org/doi/abs/10.1128/CMR.15.1.66-78.2002>
- [9] W. CJM, M. Armstrong, and R. H. Behrens, "Self-testing for falciparum malaria with antigen-capture cards by travelers with symptoms of malaria."

The American journal of tropical medicine and hygiene Am J Trop Med Hyg Am. J. Trop. Med. Hyg., vol. 63, no. 5, pp. 295 – 297, 2000. [Online]. Available: <https://www.ajtmh.org/view/journals/tpmd/63/5/article-p295.xml>

- [10] M. Najafabadi, F. Villanustre, T. Khoshgoftaar, N. Seliya, R. Wald, and E. Muharemagic, “Deep learning applications and challenges in big data analytics,” *Journal of Big Data*, vol. 2, 12 2015.
- [11] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” 2015.
- [12] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” 2018.
- [13] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, “Mobilenets: Efficient convolutional neural networks for mobile vision applications,” 2017.
- [14] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, “Squeeze-and-excitation networks,” 2019.
- [15] A. Veit, M. Wilber, and S. Belongie, “Residual networks behave like ensembles of relatively shallow networks,” 2016.
- [16] X. Glorot and Y. Bengio, “Understanding the difficulty of training deep feed-forward neural networks,” *Journal of Machine Learning Research - Proceedings Track*, vol. 9, pp. 249–256, 01 2010.
- [17] P. Ramachandran, B. Zoph, and Q. V. Le, “Searching for activation functions,” 2017.
- [18] D. Misra, “Mish: A self regularized non-monotonic activation function,” 2020.
- [19] Y. Dong, Z. Jiang, H. Shen, W. David Pan, L. A. Williams, V. V. B. Reddy, W. H. Benjamin, and A. W. Bryan, “Evaluations of deep convolutional neural networks for automatic identification of malaria infected cells,” in *2017 IEEE EMBS International Conference on Biomedical Health Informatics (BHI)*, 2017, pp. 101–104.
- [20] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in Neural Information Processing Systems*, F. Pereira, C. Burges, L. Bottou, and K. Weinberger, Eds., vol. 25. Curran Associates, Inc., 2012. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf
- [22] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” 2014.
- [23] Z. Liang, A. Powell, I. Ersoy, M. Poostchi, K. Silamut, K. Palaniappan, P. Guo, M. A. Hossain, A. Sameer, R. J. Maude, J. X. Huang, S. Jaeger, and G. Thoma, “Cnn-based image analysis for malaria diagnosis,” in *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2016, pp. 493–496.

- [24] D. Bibin, M. S. Nair, and P. Punitha, "Malaria parasite detection from peripheral blood smear images using deep belief networks," *IEEE Access*, vol. 5, pp. 9099–9108, 2017.
- [25] G. P. Gopakumar, M. Swetha, G. Sai Siva, and G. R. K. Sai Subrahmanyam, "Convolutional neural network-based malaria diagnosis from focus stack of blood smear images acquired using custom-built slide scanner," *Journal of Biophotonics*, vol. 11, no. 3, p. e201700003. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/jbio.201700003>
- [26] S. Rajaraman, S. K. Antani, M. Poostchi, K. Silamut, M. A. Hossain, R. J. Maude, S. Jaeger, and G. R. Thoma, "Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images," *PeerJ*, vol. 6, 2018.
- [27] D. R. Loh, W. X. Yong, J. Yapeter, K. Subburaj, and R. Chandramohanadas, "A deep learning approach to the screening of malaria infection: Automated and rapid cell counting, object detection and instance segmentation using mask r-cnn," *Computerized Medical Imaging and Graphics*, vol. 88, p. 101845, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0895611120301403>
- [28] W. R. W. M. Razin, T. S. Gunawan, M. Kartiwi, and N. M. Yusoff, "Malaria parasite detection and classification using cnn and yolov5 architectures," in *2022 IEEE 8th International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA)*, 2022, pp. 277–281.
- [29] W. Siłka, M. Wiecek, J. Siłka, and M. Woźniak, "Malaria detection using advanced deep learning architecture," *Sensors (Basel, Switzerland)*, vol. 23, no. 3, p. 1501, January 2023.
- [30] J. Redmon, "Darknet: Open source neural networks in c," 2013.
- [31] R. Guo, S. Li, and K. Wang, "Research on yolov3 algorithm based on darknet framework," *Journal of Physics: Conference Series*, vol. 1629, no. 1, p. 012062, sep 2020. [Online]. Available: <https://dx.doi.org/10.1088/1742-6596/1629/1/012062>
- [32] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," 2016.
- [33] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," 2020.
- [34] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.
- [35] C.-Y. Wang, H.-Y. M. Liao, I.-H. Yeh, Y.-H. Wu, P.-Y. Chen, and J.-W. Hsieh, "Cspnet: A new backbone that can enhance learning capability of cnn," 2019.
- [36] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," in *Computer Vision - ECCV 2014*. Springer International Publishing, 2014, pp. 346–361. [Online]. Available: https://doi.org/10.1007%2F978-3-319-10578-9_23

- [37] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," 2018.
- [38] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," 2022.
- [39] C.-Y. Wang, I.-H. Yeh, and H.-Y. M. Liao, "You only learn one representation: Unified network for multiple tasks," 2021.
- [40] Y. Lee, J. won Hwang, S. Lee, Y. Bae, and J. Park, "An energy and gpu-computation efficient backbone network for real-time object detection," 2019.
- [41] C.-Y. Wang, H.-y. Liao, and I.-H. Yeh, "Designing network design strategies through gradient path analysis," 11 2022.
- [42] Wikipedia contributors, "Yaml — Wikipedia, the free encyclopedia," 2023, [Online; accessed 23-June-2023]. [Online]. Available: <https://en.wikipedia.org/w/index.php?title=YAML&oldid=1161475936>
- [43] C. Dewi and H. Juli Christanto, "Combination of deep cross-stage partial network and spatial pyramid pooling for automatic hand detection," *Big Data and Cognitive Computing*, vol. 6, no. 3, 2022. [Online]. Available: <https://www.mdpi.com/2504-2289/6/3/85>
- [44] H. Bouzidi, O. Hamza, S. Niar, and A. Ait El Cadi, "Performance prediction for convolutional neural networks in edge devices," 10 2020.
- [45] A. Ganj, M. Ebadpour, M. Darvish, and H. Bahador, "Lr-net: A block-based convolutional neural network for low-resolution image classification," 2023.
- [46] X. Ding, X. Zhang, Y. Zhou, J. Han, G. Ding, and J. Sun, "Scaling up your kernels to 31x31: Revisiting large kernel design in cnns," 2022.
- [47] S. Hosseini, S. H. Lee, H. J. Kwon, H. I. Koo, and N. I. Cho, "Age and gender classification using wide convolutional neural network and gabor filter," in *2018 International Workshop on Advanced Image Technology (IWAIT)*, 2018, pp. 1–3.
- [48] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," 2020.
- [49] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," 2019.
- [50] J. P. Schwarz Schuler, S. Romání, D. Puig, H. Rashwan, and M. Abdel-Nasser, "An enhanced scheme for reducing the complexity of pointwise convolutions in cnns for image classification based on interleaved grouped filters without divisibility constraints," *Entropy*, vol. 24, p. 1264, 09 2022.
- [51] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A convnet for the 2020s," 2022.
- [52] A. Ibrahim, H. Abutarboush, A. Wagdy, M. Fouad, and E.-S. El-kenawy, "An optimized ensemble model for prediction the bandwidth of metamaterial antenna," *Computers, Materials and Continua*, vol. 71, pp. 199–213, 11 2021.
- [53] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollár, and R. Girshick, "Segment anything," *arXiv:2304.02643*, 2023.

- [54] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” 2015.
- [55] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” 2015.

Κατάλογος Σχημάτων

1.1 Κριτήρια ενός αποδοτικού τρόπου ανίχνευσης της ελονοσίας με βάση τον Παγκόσμιο Οργανισμό Υγείας	4
2.1 Γραφική αναπαράσταση της συνέλιξης μεταξύ διοδιάστατης εισόδου και πυρήνα:	7
2.2 Ο μηχανισμός SE εντός της αρχιτεκτονικής ResNET.	9
2.3 Γραφική αναπαράσταση της προς τα εμπρός διαδοσης στην έξοδο ενός νευρώνα.	10
2.4 Γραφική αναπαράσταση της οπισθοδιάδοσης από την έξοδο ενός νευρώνα στις εισόδους ή τις παραμέτρους του, με βάση τον κανόνα της αλυσίδας.	10
2.5 Στο αριστερό γράφημα τα τιμεσταμπς ενός ResNet και δεξιά τα τιμεσταμπς ενός Partial-ResNet	11
2.6 Στο αριστερό γράφημα τα Sources ενός ResNet και δεξιά τα Sources ενός Partial-ResNet στο timestamp ,time =1	12
2.7 Γραφικές παραστάσεις των συναρτήσεων ενεργοποίησης που παρουσιάστηκαν παραπάνω	13
3.1 Στο BCCD-Dataset παρατηρείται έντονη ανισορροπία μεταξύ των κλάσεων, κάτι που δεν δημιουργεί πρόβλημα καθώς στα πλαίσια της εργασίας ανιχνεύονται μόνο τα ερυθρά αιμοσφαίρια.	15
3.2 Παραδειγμα εικονας επιχρισματος αιματος απο το BCCD Dataset	16
3.3 Στο παραπάνω σχήμα παρουσιάζεται η διαδικασία προεπεξεργασίας των εικόνων	17
3.4 Αριστερά ένα υγιές ερυθροκύτταρο του NIH Dataset πριν την προεπεξεργασία και δεξιά μετά την εφαρμογή του αλγορίθμου που περιγράφηκε παραπάνω	18
3.5 Προεπεξεργασία τυχαίας εικονας απο το Archive Dataset	18
3.6 Προεπεξεργασία τυχαίας εικονας απο το BCCD Dataset	19
3.7 Η συνέχεια της διαδικασίας στην εικόνα του Archive Dataset	21
3.8 Η συνέχεια της διαδικασίας στην εικόνα του BCCD Dataset	21
3.9 Η συνέχεια της διαδικασίας στην εικόνα του BCCD Dataset	22
4.1 Η αρχιτεκτονική του YoloV4	24
4.2 Η αρχιτεκτονική της ραχοκοκαλιάς CSPDarknet53 (Στο κόκκινο πλαίσιο βλέπουμε ένα επίπεδο μετάβασης εντός του DarkBlock)	25
4.3 Το επιπεδο SPP στο δίκτυο YoloV4	26
4.4 Η ιεραρχία κατά πυραμίδες εντός του δικτύου PANet στον λαιμο του YoloV4	26
4.5 Η κεφαλή του δικτύου YoloV4	27

4.6	Cross stage partial δίκτυα. (a) κλασσικό δίκτυο, (b) CSP απλό: μετάβαση → αλληλουχία → μετάβαση, (c) CSP fusion first: αλληλουχία → μετάβαση, και (d) CSP fusion last: μετάβαση → αλληλουχία	28
4.7	(a) Η αρχιτεκτονική nonnet . Στο κόκκινο πλαίσιο ένα osa module (b) Η αρχιτεκτονική CSPnonnet . Εξέλιξη του nonnet καθώς εντός δικτύου χρησιμοποιεί cross stage partial σύνδεση . Στο μπλέ πλαίσιο βλέπουμε ένα στάδιο μετάβασης. (c) Τέλος η αρχιτεκτονική ELAN . Παρατηρείται η cross stage partial σύνδεση , η απουσία πολλαπλών επιπέδων μετάβασης και το συνελικτικό μπλόκ για εξαγωγή χαρακτηριστικών.	28
4.8	Παρατηρείται ότι αυξάνοντας το βάθος στο συνελικτικό μπλοκ με τη μέθοδο της εικόνας, τότε αυξάνεται παράλληλα και το πλάτος του.	29
4.9	Η προτεινόμενη μέθοδος κλιμακοποίησης του δικτύου.	29
4.10	Ένα κομμάτι της ραχοκοκαλιάς του δικτύου YoloV7	30
4.11	Η βελτιωμένη έκδοση του SPP χρησιμοποιεί τις ιδέες της αποτελεσματικότερης κλιμακοποίησης των κατά αλληλουχία δικτύων.	30
4.12	Ο λαιμός του δικτύου YoloV7.	31
4.13	Το επαναπαραμετροποιήσιμο μπλοκ RepConv	32
4.14	Το συνελικτικό μπλόκ του Multi Kernel Network (MKN).	35
4.15	Η αρχιτεκτονική του δικτύου Wide MKN.	36
4.16	Ένα Inverted Residual block. Τα σκιαγραφημένα τμήματα αναπαριστούν επίπεδα χωρίς συναρτήσεις ενεργοποίησης.	37
4.17	Αριστερά μια συνέλιξη pointwise και δεξιά η αντικατάσταση που περιγράφηκε παραπάνω.	38
4.18	Εσωτερική δομή του ConvNeXt Block και ομοιότητες του με την δομή του Swin Transformer	38
4.19	Δομή ενός δικτύου Ensemble που βασίζεται σε 3 επιμέρους μοντέλα. . .	39
4.20	Ενδεικτικά υπερπαραμέτροι εκπαίδευσης των δικτύων.	39
5.1	Καμπύλες απώλειας των τεσσάρων αρχιτεκτονικών Yolo στο Archive Dataset. Στο πρώτο διάγραμμα παρουσιάζεται η συνολική καμπύλη καθ'όλες τις 2000 επαναλήψεις. Στο δεύτερο οι πρώτες 250 και στο τρίτο οι τελευταίες 500.	40
5.2	Καμπύλες απώλειας των τεσσάρων αρχιτεκτονικών Yolo στο BCCD Dataset. Στο πρώτο διάγραμμα παρουσιάζεται η συνολική καμπύλη καθ'όλες τις 2000 επαναλήψεις. Στο δεύτερο οι πρώτες 250 και στο τρίτο οι τελευταίες 500.	41
5.3	Χρονos εκτελεσης των τεσσαρων μοντελων στο συνολο δοκιμων του Archive Dataset	46
5.4	Χρονos εκτελεσης των τεσσαρων μοντελων στο συνολο δοκιμων του BCCD Dataset	46
5.5	Αριστερά βλέπουμε το παραγώμενο οριοθετούμενο πλαίσιο από τον YoloV7 ,στη μέση το οριοθετούμενο πλαίσιο μαζί με την δυαδική μάσκα που προέκυψε ως αποτέλεσμα του SAM και τέλος την απομονωση του μεμονομένου κυττάρου απο την υπόλοιπη εικόνα	47
5.6	Καμπύλες απώλειας και ακρίβειας για το μοντέλο Multi Kernel. Η κόκκινη ευθεία αναπαριστά την πτώση του ρυθμού μάθησης με βάση τον προγραμματισμό του.	48
5.7	Καμπύλες απώλειας και ακρίβειας για το μοντέλο K-EffNet-B0. Η κόκκινη ευθεία αναπαριστά την πτώση του ρυθμού μάθησης με βάση τον προγραμματισμό του.	48

5.8 Καμπύλες απώλειας και ακρίβειας για το μοντέλο ConvNeXt. Η κόκκινη ευθεία αναπαριστά την πτώση του ρυθμού μάθησης με βάση τον προγραμματισμό του.	49
5.9 Καμπύλες απώλειας και ακρίβειας για το μοντέλο Ensemble. Η κόκκινη ευθεία αναπαριστά την πτώση του ρυθμού μάθησης με βάση τον προγραμματισμό του.	49
5.10 Πίνακας σύγκρισης του δικτύου Wide Multi-Kernel στο σύνολο δοκιμών του Custom NIH Dataset	51
5.11 Πίνακας σύγκρισης του δικτύου K-EffNet-B0 στο σύνολο δοκιμών του Custom NIH Dataset	51
5.12 Πίνακας σύγκρισης του δικτύου ConvNeXt στο σύνολο δοκιμών του Custom NIH Dataset	51
5.13 Πίνακας σύγκρισης του δικτύου Ensemble στο σύνολο δοκιμών του Custom NIH Dataset	52