



Τεχνολογικό Εκπαιδευτικό Ίδρυμα Κρήτης

Σχολή Τεχνολογικών Εφαρμογών

Τμήμα Μηχανικών Πληροφορικής

Πτυχιακή Εργασία

Τίτλος: 75.ΑΝΑΓΝΩΡΙΣΗ ΚΑΡΚΙΝΩΜΑΤΟΣ

Συνάπαλος Δημήτριος (ΑΜ:2051)

Επιβλέπων Καθηγητής: Παπαδάκης Νικόλαος

Επιτροπή αξιολόγησης:

Ημερομηνία παρουσίασης:

Ευχαριστίες

Η συγγραφή του παρόντος έργου δεν θα ήταν δυνατή χωρίς τη συνδρομή των καθηγητών κ.κ. Παπαδάκη Νικόλαο και Κατρίνη Χρυσάνθη, που διέθεσαν τον απαραίτητο χρόνο και υπομονή για την καθοδήγησή μου.

Ιδιαίτερες ευχαριστίες απευθύνονται στην οικογένειά μου για το κουράγιο που μου έδωσε σε δύσκολες μέρες, στον Νίκο για τα απογεύματα που περάσαμε συζητώντας λεπτομέρειες και κλίσεις, όπως και στους συναδέλφους μου για την Τεχνική μου Υποστήριξη.

Εις μνήμην του συνονόματού μου, που δεν πρόλαβε να δει την εργασία ολοκληρωμένη.

Abstract

The presented project, engages the topic of computer assisted detection and categorization of cancerous masses in mammograms. Starting on the cited papers, the cutting-edge techniques and technologies are mentioned and summarily applied in the construction of the detector.

Use of the MathWorks MATLAB suite and digital image processing technics allows the transformation of the selected database into a suitable and helpful form, as well as, enhancing data quality.

Following is extensive mention of the characteristics embodied in popular Deep Neural Network structures, the application of Transfer Learning techniques and the necessary parameterization in order to apply said structures in producing Faster R-CNN based detectors.

After presenting and selecting the appropriate training parameters, the networks offering prime results undergo further elaboration and structural redesign in order to achieve the best returns.

Finalizing the project, an analysis of the influence of detector retraining on performance and associated results.

Keywords – Tumor Recognition, Machine Learning, Neural Networks, Computer Assisted Medicine, Digital Image Processing, Faster R-CNN, Transfer Learning, Contrast Augmentation

Περίληψη

Στην παρούσα εργασία ασχολείται με το ζήτημα του υπολογιστικά υποβοηθούμενου εντοπισμού και κατηγοριοποίησης καρκινικών όγκων σε μαστογραφίες. Με αφετηρία τις δημοσιεύσεις της βιβλιογραφίας εντοπίζονται οι τεχνικές και τεχνολογίες αιχμής οι οποίες θα χρησιμοποιηθούν στην συνέχεια για την υλοποίηση του συστήματος εντοπισμού.

Χρησιμοποιώντας τη σουίτα εφαρμογών MathWorks MATLAB αξιοποιούνται τεχνικές ψηφιακής επεξεργασίας εικόνας ώστε η βάση δεδομένων που επιλέχθηκε να μετασχηματιστεί ανάλογα με τις απαιτήσεις της εργασίας και να ενισχυθεί η ποιότητα των δεδομένων.

Στην συνέχεια γίνεται εκτενής αναφορά στα χαρακτηριστικά που διαθέτουν δημοφιλείς δομές βαθιών νευρωνικών δικτύων, χρησιμοποιούνται τεχνικές μεταφοράς γνώσης και επιχειρείτε η κατάλληλη παραμετροποίηση τους ώστε να αξιοποιηθούν στην υλοποίηση ανιχνευτών με χρήση της τεχνολογίας Faster R-CNN.

Αφού γίνει παράθεση αποτελεσμάτων και επιλογή των κατάλληλων τιμών παραμέτρων εκπαίδευσης, τα δίκτυα που εμφάνισαν τα ικανότερα αποτελέσματα υφίστανται περαιτέρω επεξεργασία και δομικό επανασχεδιασμό με σκοπό την επίτευξη της μέγιστης δυνατής απόδοσης.

Ολοκληρώνοντας την εργασία, γίνεται ανάλυση της επιρροής που επιφέρει η επανεκπαίδευση των δικτύων στην απόδοσή τους και παρατίθενται τα σχετικά αποτελέσματα.

Λέξεις κλειδιά – Αναγνώριση Καρκίνωματος, Μηχανική Μάθηση, Νευρωνικά Δίκτυα, Υπολογιστικά Υποβοηθούμενη Ιατρική, Ψηφιακή Επεξεργασία Εικόνας, Faster R-CNN, Μεταφορά Γνώσης, Ενίσχυση Αντίθεσης

Πίνακας Περιεχομένων

Ευχαριστίες	<i>i</i>
Abstract	<i>ii</i>
Περίληψη	<i>iii</i>
Πίνακας Περιεχομένων	<i>iv</i>
Πίνακας Εικόνων	<i>vi</i>
Πίνακας Πινάκων	<i>viii</i>
Πίνακας Εξισώσεων	<i>ix</i>
1 Εισαγωγή στην Αναγνώριση Καρκινωμάτων	1
1.1 Αναφορά στα καρκινώματα και τον καρκίνο του μαστού	1
1.1.1 Υφιστάμενες μέθοδοι διάγνωσης	1
1.1.2 Κίνδυνοι και θεραπεία καρκίνων	1
1.1.3 Πρόληψη (Αποτελεσματικότερη από τη θεραπεία)	2
1.2 Τεχνολογικό υπόβαθρο εντοπισμού και αναγνώρισης καρκινωμάτων	2
1.2.1 Ιστορική εξέλιξη Υπολογιστικά Υποβοηθούμενης Ιατρικής	2
1.2.2 Βαθιά Νευρωνικά Δίκτυα	3
1.3 Σχέδιο Δράσης Εργασίας	5
1.3.1 Στόχοι Εργασίας	5
1.3.2 Μεθοδολογία ανάλυσης και βηματισμού	5
1.4 Πειραματικός Εξοπλισμός	6
1.4.1 Hardware	6
1.4.2 Software	6
2 Παρουσίαση και Επεξεργασία Πειραματικών Δεδομένων	7
2.1 Αναφορά στη Βάση Δεδομένων CBIS-DDSM	7
2.2 Επεξεργασία και μορφοποίηση εικόνων	9
2.2.1 Μετατροπή τύπου, διάνοιξη και αποκοπή	9
2.2.2 Μετασχηματισμός	10
2.2.3 Ενίσχυση αντίθεσης	11
2.3 Συλλογή δεδομένων και καταγραφή αρχείων Ground Truth	12
3 Μέθοδοι Αξιολόγησης Ανιχνευτών	14
4 Αρχιτεκτονική Βαθιών Νευρωνικών Δικτύων	17
4.1 Εισαγωγή στην οικογένεια R-CNN	17
4.2 Βασικές Αρχιτεκτονικές Δομές	19
4.2.1 AlexNet	19
4.2.2 R-FCN	23
4.2.3 GoogLeNet	25
4.2.4 YOLO	28
4.2.5 ResNet	30
4.3 Ανασκόπηση αποτελεσμάτων βασικών δομών	32
5 Παράμετροι Εκπαίδευσης Πειραμάτων	34

5.1 Αναφορά στην λειτουργία της Μηχανικής Μάθησης	34
5.2 Επεξήγηση αλγόριθμων κατάβασης απώλειας	35
5.2.1 SGDM	35
5.2.2 RMSPPROP	36
5.2.3 ADAM	37
5.3 Τυποποίηση παραμέτρων εκπαίδευσης	39
5.3.1 Διάρκεια Εκπαίδευσης	39
5.3.2 Ρυθμός Μεταβολής Μάθησης	42
5.3.3 Τομή-προς-Ένωση Θετικών-Αρνητικών Δειγμάτων	44
6 Επιπλέον Πειραματισμοί σε Δομές και Παραμέτρους	46
6.1 Πειράματα Δομών Αρχιτεκτονικής	46
6.1.1 Πειραματισμοί σε βάση δικτύου AlexNet	46
6.1.2 Πειραματισμοί σε βάση δικτύου GoogLeNet	49
6.1.3 Σύνοψη αποτελεσμάτων πειραματικών δομών	51
6.2 Πειραματισμοί επανεκπαίδευσης δικτύων	54
6.2.1 Επανεκπαίδευση Short AlexNet	54
6.2.2 Επανεκπαίδευση GoogLeNet	56
6.2.3 Επανεκπαίδευση Long GoogLeNet	57
6.2.4 Σύνοψη αποτελεσμάτων επανεκπαιδεύσεων	59
Βιβλιογραφία	60

Πίνακας Εικόνων

Εικόνα 1 – Ενδεικτική μαστογραφία και οπτικοποίηση βαρών πρώτου επίπεδου συνέλιξης.....	3
Εικόνα 2 – Ενεργοποιήσεις πρώτου επίπεδου συνέλιξης και ισχυρότερο κανάλι ενεργοποίησης...	4
Εικόνα 3 – Ενδεικτικές εικόνες συλλογής CBIS-DDSM	7
Εικόνα 4 – Ενδεικτικά διαθέσιμες απεικονίσεις για την ασθενή με κωδικό P_00038	8
Εικόνα 5 – Βήματα μετασχηματισμού από αρχική σε επεξεργασμένη εικόνα.....	10
Εικόνα 6 – Βήματα ενίσχυσης αντίθεσης	12
Εικόνα 7 – Ενδεικτική απεικόνιση πίνακα αλήθειας εκπαίδευσης.....	13
Εικόνα 8 – Βήματα παραγωγής αποτελεσμάτων.....	14
Εικόνα 9 – Κατηγοριοποίηση αποτελεσμάτων.....	15
Εικόνα 10 – Ενδεικτικά σχήματα Miss Rate-FPPI και Precision-Recall.....	16
Εικόνα 11 – Ενδεικτικά σχήματα απώλειας εκπαίδευσης	16
Εικόνα 12 – Δομή RCNN [58]	17
Εικόνα 13 – Δομή Fast RCNN [58].....	17
Εικόνα 14 – Δομή Faster RCNN [58].....	18
Εικόνα 15 – Στρώματα που απαρτίζουν το Faster RCNN σε υλοποίηση Matlab	18
Εικόνα 16 – Αρχική δομή AlexNet.....	19
Εικόνα 17 – Δομή AlexNet με την προσθήκη επιπλέον Fully Connected στρώματος.....	20
Εικόνα 18 – Ανάλυση ροής δικτύου AlexNet που έχει υποστεί επεξεργασία.....	21
Εικόνα 19 – Απώλεια εκπαίδευσης δομής AlexNet	22
Εικόνα 20 – Σχέδιο απόδοσης δομής AlexNet	22
Εικόνα 21 – Δομή FCN AlexNet.....	23
Εικόνα 22 – Σχέδιο απόδοσης δομής FCN AlexNet.....	24
Εικόνα 23 – Απώλεια εκπαίδευσης δομής FCN AlexNet.....	24
Εικόνα 24 – Δομή GoogLeNet	25
Εικόνα 25 – Το RPN τμήμα της υλοποίησης GoogLeNet.....	26
Εικόνα 26 – Απώλεια εκπαίδευσης δομής GoogLeNet.....	27
Εικόνα 27 – Σχέδιο απόδοσης δομής GoogLeNet	27
Εικόνα 28 – Δομή YOLO	28
Εικόνα 29 – Απώλεια εκπαίδευσης δομής YOLO.....	29
Εικόνα 30 – Σχέδιο απόδοσης δομής YOLO.....	29
Εικόνα 31 – Δομή ResNet50 με σύνδεση RPN στη ReLU Activation 40.....	30
Εικόνα 32 – Απώλεια εκπαίδευσης δομής ResNet	31
Εικόνα 33 – Σχέδιο απόδοσης δομής ResNet	31
Εικόνα 34 – Σχέδιο απόδοσης Βασικών Δομών	32
Εικόνα 35 – Απώλεια εκπαίδευσης Βασικών Δομών	33
Εικόνα 36 – Κατάβαση απώλειας SGDM	36
Εικόνα 37 – Κατάβαση απώλειας RMSPROP	37
Εικόνα 38 – Κατάβαση απώλειας ADAM.....	38
Εικόνα 39 – Συγκριτικό σχήμα κατάβασης απώλειας	39
Εικόνα 40 - Κατάβαση Απώλειας σε 500 εποχές	41
Εικόνα 41 – Σχήμα ποσοστιαίας διαφοράς Απώλειας προς Χρόνο Εκπαίδευσης.....	42
Εικόνα 42 – Συγκεντρωτικό σχήμα σχέσης Loss με learning rate.....	43
Εικόνα 43 – Σχήμα απώλειας τιμών Negative Overlap	45
Εικόνα 44 – Δομή Short AlexNet	46
Εικόνα 45 – Διαφορά αποτελεσμάτων AlexNet / Short AlexNet.....	47
Εικόνα 46 – Δομή Inception AlexNet.....	48
Εικόνα 47 – Δομή Inception AlexNet+	48
Εικόνα 48 – Δομές Short & Long GoogLeNet	49
Εικόνα 49 – Δομές X-Connect GoogLeNet.....	50
Εικόνα 50 – Αποτελέσματα πειραματικών δομών	51
Εικόνα 51 – Απώλεια κατάβασης πειραμάτων	52
Εικόνα 52 – Σχήμα απόδοσης και απώλειας επανεκπαιδεύσεων Short AlexNet	55

Εικόνα 53 – Σχήμα απόδοσης και απώλειας επανεκπαιδεύσεων GoogLeNet.....	56
Εικόνα 54 – Ενδεικτικά αποτελέσματα δικτύων Short AlexNet	56
Εικόνα 55 – Ενδεικτικά αποτελέσματα δικτύων GoogLeNet.....	57
Εικόνα 56 – Ενδεικτικά αποτελέσματα δικτύων Long GoogLeNet	57
Εικόνα 57 – Σχήμα απόδοσης και απώλειας επανεκπαιδεύσεων Long GoogLeNet	58

Πίνακας Πινάκων

Πίνακας 1 – Σύνθεση συλλογών σετ εκπαίδευσης/ελέγχου και αναλογία σε μορφολογία όγκων .	13
Πίνακας 2 – Απόδοση δομής AlexNet.....	22
Πίνακας 3 – Απόδοση δομής FCN AlexNet.....	24
Πίνακας 4 – Απόδοση δομής GoogLeNet	27
Πίνακας 5 – Απόδοση δομής YOLO	29
Πίνακας 6 – Απόδοση δομής ResNet	31
Πίνακας 7 – Πίνακας απόδοσης Βασικών Δομών	32
Πίνακας 8 – Πίνακας Απώλειας ανά Εποχή	40
Πίνακας 9 – Πίνακας ποσοστιαίας διαφοράς Απώλειας προς Χρόνο Εκπαίδευσης.....	41
Πίνακας 10 – Πίνακας αποτελεσμάτων Τομής ως προς Ένωσης.....	44
Πίνακας 11 – Αποτελέσματα διαφορετικών τιμών Negative Overlap	45
Πίνακας 12 – Αποτελέσματα δομής Short AlexNet	46
Πίνακας 13 – Αποτελέσματα δομών Inception AlexNet.....	47
Πίνακας 14 – Αποτελέσματα πειραμάτων δομών GoogLeNet.....	51
Πίνακας 15 – Αποτελέσματα πειραματικών δομών	53
Πίνακας 16 – Συγκριτική απόδοση πειραματικών δομών	54
Πίνακας 17 – Αποτελέσματα επανεκπαίδευσης δομών.....	59
Πίνακας 18 – Συγκριτική απόδοση επανεκπαίδευσης δομών.....	59

Πίνακας Εξισώσεων

Εξίσωση 1 – Αποτέλεσμα νευρώνος με σιγμοειδή λειτουργία ενεργοποίησης.....	34
Εξίσωση 2 – Υπολογισμός μέσου κόστους συστήματος.....	34
Εξίσωση 3 – Λόγος μεταβολής κόστους/βάρους.....	34
Εξίσωση 4 – Παράγωγος μεταβολής.....	34
Εξίσωση 5 – Κλίση κατάβασης.....	35
Εξίσωση 6 – Στοχαστική Κλίση Κατάβασης με Ορμή.....	35
Εξίσωση 7 – Απόφαση κλίσης RPROP.....	36
Εξίσωση 8 – Απόφαση κλίσης RMSPROP.....	37
Εξίσωση 9 – Πρόβλεψη δύο στιγμιότυπων (moments).....	37
Εξίσωση 10 – Υπολογισμός μέσων όρων προβλέψεων.....	38
Εξίσωση 11 – Γενικευμένος μέσος όρος προβλέψεων.....	38
Εξίσωση 12 – Υπολογισμός προβλέψεων λαμβάνοντας υπόψιν την προκατάληψη.....	38
Εξίσωση 13 – Υπολογισμός βαρών ADAM.....	39

1 Εισαγωγή στην Αναγνώριση Καρκινωμάτων

Στο παρόν κεφάλαιο παρουσιάζονται οι έννοιες με τις οποίες αναμετράται το σύνολο της πτυχιακής εργασίας στην προσπάθεια σχεδιασμού, υλοποίησης και εκπαίδευσης ενός συστήματος μηχανικής μάθησης το οποίο εστιάζει στον εντοπισμό και τη κατηγοριοποίηση καρκινωμάτων σε ιατρικές ψηφιακές εικόνες.

1.1 Αναφορά στα καρκινώματα και τον καρκίνο του μαστού

Καρκίνος^a είναι το όνομα που έχει δοθεί σε μια ομάδα ασθενειών με κύριο χαρακτηριστικό την πάρα φύσιν και ανομοιογενή αναπαραγωγή κυττάρων και τον σχηματισμό καρκινωμάτων, μαζών καρκινικών κυττάρων, που ονομάστηκαν έτσι λόγω του σχήματος τους. Οι όγκοι που σχηματίζονται από πολλές διαταραχές στην κυτταρική αναπαραγωγή χωρίζονται σε δύο κατηγορίες βάσει της μορφολογίας τους. Όγκοι με συγκεκριμένο σχήμα, γενικά αβλαβής και ασυμπτωματικοί, χωρίς την ικανότητα μετάστασης ονομάζονται καλοήθεις. Όγκοι με ακανόνιστο, ακανθώδες ή ανομοιογενές σχήμα και ικανότητα μετάστασης ονομάζονται καρκινικοί ή κακοήθεις. Οι καρκινικοί όγκοι χαρακτηρίζονται από τις παρακάτω ιδιότητες:

- Κυτταρική αναπαραγωγή χωρίς τα φυσιολογικά σήματα
- Συνεχής αναπαραγωγή ακόμη και παρουσία αντιθετικών σημάτων
- Αποφυγή του φυσιολογικού κυτταρικού θανάτου
- Απεριόριστος αριθμός αναπαραγωγής
- Προώθηση στένωσης αρτηριών
- Εισβολή άλλων ιστών και σχηματισμός μεταστάσεων

Οι διάφορες μορφές καρκίνου χαρακτηρίζονται από το μέρος του σώματος το οποίο προσβάλλουν και τον τύπο των κυττάρων που τραυματίστηκαν, αλλά από τη στιγμή που προσβληθούν άλλα όργανα η πάθηση ονομάζεται μεταστατικός καρκίνος. Λόγω της ανομοιογενούς φύσης της ασθένειας έχουν αναπτυχθεί ποικίλες τεχνικές και μέθοδοι, τόσο για την πρόληψη, τον εντοπισμό, τη διάγνωση και τη θεραπεία του καρκίνου.

Η παρούσα πτυχιακή εργασία εστιάζει στον καρκίνο του ανθρώπινου μαστού, ο οποίος μπορεί να εμφανιστεί σε οποιοδήποτε από τα στρώματα που απαρτίζουν τον μαστό, αλλά εμφανίζεται συχνότερα στα λιπώδη στρώματα (σάρκωμα) και κυρίως τους γαλακτοφόρους αδένες (πορογενές και λοβιακό καρκίνωμα). Ο καρκίνος του μαστού εμφανίζεται πλειοψηφικά σε γυναίκες, ωστόσο μπορεί να εμφανιστεί και στον άνδρα, κυρίως σε περιπτώσεις που υπάρχει ανεπτυγμένος μαστός λόγω ορμονικής ή άλλης διαταραχής.

1.1.1 Υφιστάμενες μέθοδοι διάγνωσης

Η διάγνωση των ασθενειών που περιγράφονται σαν καρκινικές γίνεται με ποικίλες μεθόδους, όπως προκύπτει από την ανάγκη να αντιμετωπιστεί μια τόσο ανομοιογενής ασθένεια. Εξαιρετικά συχνή είναι η χρήση απεικονιστικών μεθόδων όπως η ακτινογραφία (radiogram), η μαγνητική ή αξονική τομογραφία (MRI, CT) και η μαστογραφία (mammogram). Σε περιπτώσεις όπου δεν υπάρχει η τάση σχηματισμού όγκων, όπως οι μορφές τις λευχαιμίας, είναι απαραίτητη η χημική ανάλυση σωματικών ιστών ή υγρών όπως η ανάλυση αίματος. Σε άλλες περιπτώσεις είναι απαραίτητη η ψηλάφηση περιοχών του σώματος όπως στον μαστικό, πρωκτικό ή ορχικός καρκίνος, ενώ συχνά για την επιβεβαίωση της φύσης του καρκινώματος αξιοποιούνται βιοψίες των επηρεασμένων περιοχών.

1.1.2 Κίνδυνοι και θεραπεία καρκίνων

Η θεραπεία του καρκίνου αποτελεί ένα περίπλοκο και συναισθηματικά φορτισμένο ζήτημα. Ενώ έχουν σχεδιαστεί και εφαρμοστεί ποικίλες τεχνικές για την απονέκρωση των καρκινικών όγκων, η επιτυχία τους εξαρτάται σε μεγάλο βαθμό από ετερογενείς παράγοντες. Η επιτυχία της θεραπείας συχνά εξαρτάται από την έγκαιρη διάγνωση και σε περιπτώσεις όπου δεν έχει εμφανιστεί παθολογία ή κίνδυνος μετάστασης αξιοποιείται η χειρουργική αφαίρεση των επιβλαβών όγκων.

^a <https://www.cancer.gov/about-cancer/understanding/what-is-cancer>

Σε περιπτώσεις όπου τα καρκινικά κύτταρα είναι συγκεντρωμένα σε μία περιοχή αλλά η χειρουργική επέμβαση είναι επικίνδυνη ή και αδύνατη αξιοποιούνται μέθοδοι όπως η ραδιοθεραπεία ή χημικές μέθοδοι που εμποδίζουν τον μεταβολισμό των επιζήμιων κυττάρων.

Τέλος όταν έχει υπάρξει, ή θεωρείται υψηλός ο κίνδυνος μετάστασης, χρησιμοποιείται γενική χημειοθεραπεία η οποία σαν σκοπό έχει την καταστροφή μεμονωμένων κυττάρων που μπορεί να προκαλέσουν υποτροπή.

1.1.3 Πρόληψη (Αποτελεσματικότερη από τη θεραπεία)

Για κάποιες μορφές καρκίνων υπάρχουν ενέργειες στις οποίες μπορεί κάποιος να προβεί ώστε να μειώσει τις πιθανότητες εμφάνισης της ασθένειας όπως η αποφυγή καρκινογόνων ουσιών και δραστηριοτήτων που έχουν συνδεθεί με αυξημένη επικινδυνότητα όπως η υπερϊώδης ακτινοβολία και το κάπνισμα.

Για την αποφυγή του καρκίνου το μαστού και των επιπτώσεών του είναι καίριας σημασία η συντομότερη δυνατή διάγνωση της ασθένειας, προτού την εκδήλωση συμπτωμάτων ή του κινδύνου μετάστασης αν αυτό είναι δυνατό. Για να υπάρξει ο έγκαιρος εντοπισμός των όγκων είναι απαραίτητος ο συχνός έλεγχος μέσω ψηλάφησης του μαστού όπως και ο ακτινολογικός έλεγχος, όπου ένα ζεύγος μαστογραφιών ελέγχεται από ακτινολόγους οι οποίοι καταλήγουν στην διάγνωση.

Σε περιπτώσεις δυσδιάκριτοι ή ασαφής όγκοι είναι δυνατών να περάσουν απαρατήρητοι ή να συγχυθούν με τη φυσιολογική δομή του μαστού. Για αυτό το λόγο είναι ορθή πρακτική ο επαναληπτικός έλεγχος της μαστογραφίας από δεύτερο ειδικό, κάτι που εν τέλη είναι δαπανηρό και συχνά αποφεύγεται. Στη παρούσα εργασία θα επιχειρήσουμε να σχεδιάσουμε και να υλοποιήσουμε ένα σύστημα αυτόματου εντοπισμού αντικειμένων το οποίο θα μπορούσε να αξιοποιηθεί, συμπληρωματικά και υποβοηθητικά στον ρόλο του ακτινολόγου-διαγνώστη πιθανώς βελτιώνοντας την ασφάλεια και την ακρίβεια της πρώτης διάγνωσης αυξάνοντας έτσι τις πιθανότητες πλήρους θεραπείας.

1.2 Τεχνολογικό υπόβαθρο εντοπισμού και αναγνώρισης καρκινωμάτων

Στη παρούσα ενότητα λαμβάνει χώρα σύντομη αναφορά σε τεχνικές και τεχνολογίες αιχμής της πληροφορικής που χρησιμοποιούνται στα πλαίσια της υποβοήθησης των ιατρικών μεθόδων διάγνωσης και πρόληψης, με έμφαση στην τεχνολογία των βαθιών νευρωνικών δικτύων η οποία θα αξιοποιηθεί στα επόμενα κεφάλαια.

1.2.1 Ιστορική εξέλιξη Υπολογιστικά Υποβοηθούμενης Ιατρικής

Την πρώτη Απριλίου του 1972 οι de Dombal et al. [1] του Πανεπιστημίου του Leeds δημοσιεύουν στο British Medical Journal, την έρευνά τους με τίτλο Computer-aided Diagnosis of Acute Abdominal Pain. Σε αυτήν πραγματεύονται τα αποτελέσματα μίας έρευνας που αξιοποίησε υπολογιστικά συστήματα της εποχής στην κατηγοριοποίηση 304 περιπτώσεων κοιλιακού άλγους. Οι ερευνητές παρατηρούν τα υψηλά ποσοστά επιτυχίας του υπολογιστικού συστήματος στην πρόβλεψη των περιπτώσεων και περιγράφουν τους τομείς πάνω στους οποίους πρέπει να αξιολογηθεί οποιοδήποτε σύστημα υλοποιείται για λόγους υποβοήθησης ιατρικού προσωπικού, δίνοντας έμφαση στην Ακρίβεια, την Βεβαιότητα και την Ασφάλεια της πρόβλεψης του συστήματος.

Η ερευνά τους δεν αποτελεί προϊόν παρθενόγενεσης. Ήδη από τα τέλη του 19^{ου} Αιώνα παρατηρούμε την προσπάθεια να χρησιμοποιηθούν από την Ιατρική Επιστήμη οι μηχανές διάτρητων καρτών του Herman Hollerith [2], κυρίως στην δημιουργία βάσεων δεδομένων με στοιχεία όπως αρχεία ασθενών και στατιστικά θνησιμότητας, ενώ με χρήση των ανωτέρω οι Doll & Hill [3], το 1950, συνέδεσαν μορφές του καρκίνου του πνεύμονα με το κάπνισμα. Η χρήση συστημάτων διάτρητων καρτών αντικαταστάθηκε σταδιακά τις δεκαετίες 1950-1960 με πλήρως ηλεκτρονικά συστήματα βασισμένα στα συστήματα Colossus και ENIAC που αναπτύχθηκαν κατά τη διάρκεια του Β' Παγκοσμίου Πολέμου.

Στις δεκαετίες που ακολούθησαν η ραγδαία εξέλιξη της ιατρικής ανατροφοδοτήθηκε με την αντίστοιχα εντυπωσιακή εξέλιξη της τεχνολογίας της πληροφορικής. Η ψηφιοποίηση της πληροφορίας επέφερε ιδιαίτερα πλεονεκτήματα όπως η προσβασιμότητα στα δεδομένα (accessibility), η ασφαλής αποθήκευσή τους (redundancy) και η ευκολότερη επεξεργασία τους (manageability). Η διαρκώς αυξούμενη πυκνότητα δεδομένων (αυξημένη χωρητικότητα σε μειωμένο όγκο) επέτρεψε την δημιουργία αντίστοιχα μεγάλων βάσεων δεδομένων και η εκθετική ανάπτυξη της υπολογιστικής

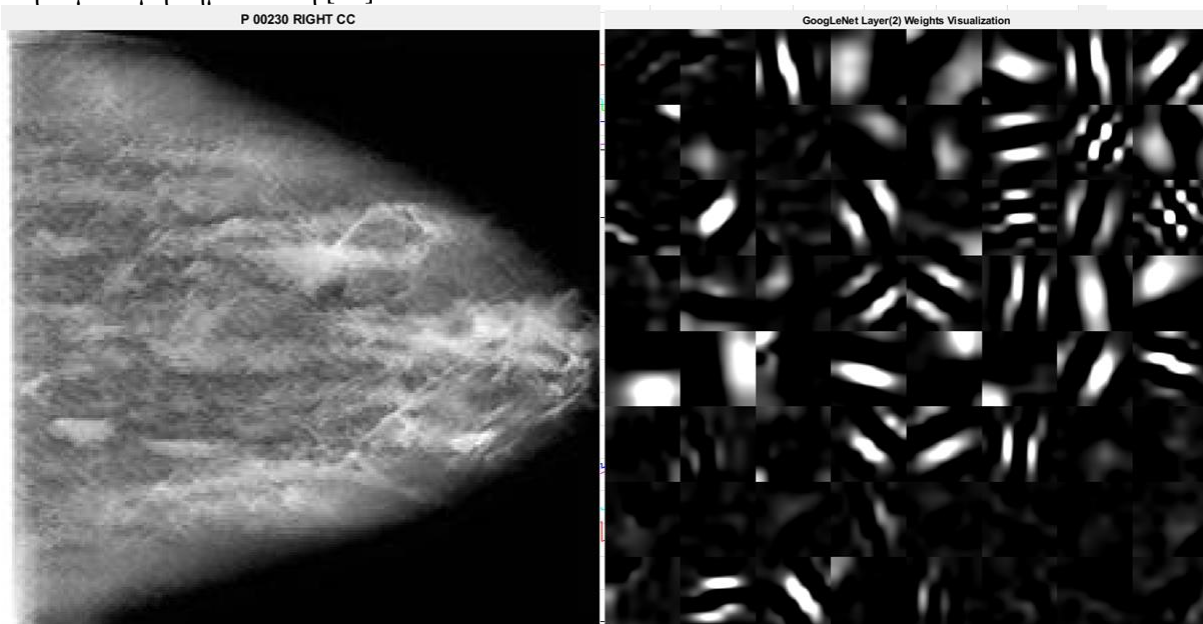
ικανότητας μέχρι τις αρχές του 21^{ου} Αιώνα, σε συνδυασμό με τον σχεδιασμό και την υλοποίηση του παγκόσμιου ιστού δημιουργούν την, ιστορικά πρωτοφανή, σημερινή εικόνα στην ιατρική και υπολογιστική έρευνα.

Κατά την πορεία ανάπτυξης των συστημάτων που αργότερα κατηγοριοποιήθηκαν σαν Υπολογιστικά Υποβοηθούμενου Εντοπισμού και Διάγνωσης (Computer Aided Detection & Diagnosis CADe & CADx) αξιοποιήθηκαν ιδιαίτερες και ενδιαφέρουσες τεχνικές όπως η χρήση Ασαφούς Λογικής [4], χρήση support vector machines για κατηγοριοποίηση δεδομένων [5] και η χρήση απλών συστημάτων αισθητήρων για μακροχρόνια παρακολούθηση [6].

Την τελευταία δεκαετία τεράστιο ερευνητικό ενδιαφέρον έχει προσελκύσει η χρήση Βαθιών Νευρωνικών Δικτύων (DNN) λόγω των εξαιρετικών αποτελεσμάτων που έχουν παρουσιάσει, κυρίως στην αναγνώριση και κατηγοριοποίηση εικόνων, αλλά και την πρόβλεψη ακολουθιών. Η παρούσα εργασία εστιάζει στην πρώτη ιδιότητα, αναφορικά με την υλοποίηση συστήματος εντοπισμού και κατηγοριοποίησης καρκινωμάτων σε μαστογραφίες.

1.2.2 Βαθιά Νευρωνικά Δίκτυα

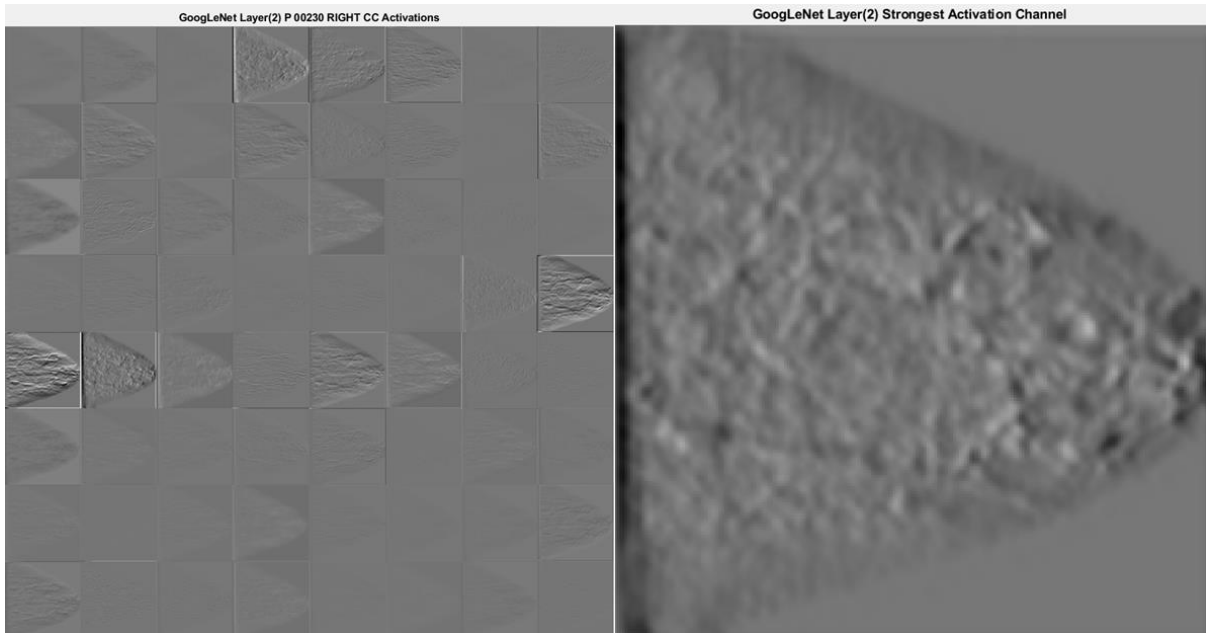
Το 1989 οι LeCun et al. [7] δημοσιεύουν μελέτη στην οποία υλοποιούν ένα νευρωνικό δίκτυο για την αναγνώριση των πέντε ψηφίων του ταχυδρομικού κώδικα από μια εικόνα με χειρόγραφα ψηφία χρησιμοποιώντας αλγόριθμο αναδρομής. Το δίκτυο που υλοποίησαν αποτελεί το άθροισμα τριών πλήρως συνδεδεμένων στρωμάτων και αποτελεί την εξέλιξη του δικτύου Perceptron [8] που είχε προταθεί ήδη από το 1958. Η πρωτοπορία των LeCun et al. εντοπίζεται στην οξύνει τους να συνδέσουν πολλαπλά στρώματα νευρώνων, συνδεδεμένα και ανατροφοδοτούμενα μέσω του αλγορίθμου Backpropagation [9], σύνδεση η οποία επέτρεψε στο δίκτυό τους να εξάγει το ίδιο τα διακριτικά χαρακτηριστικά των σχημάτων, σε αντίθεση με τεχνικές υποβοηθούμενης μάθησης που απαιτούν την χορήγηση χαρακτηριστικών στο σύστημα. Η εκπαίδευση και χρήση δικτύων με κρυμμένα στρώματα ενδιάμεσα των στρωμάτων εισόδου και εξόδου έμεινε γνωστή ως Deep Learning όπως αναφέρθηκε στην ομώνυμη δημοσίευση [10].



Εικόνα 1 – Ενδεικτική μαστογραφία και οπτικοποίηση βαρών πρώτου επίπεδου συνέλιξης

Η επόμενη εμβάθυνση στην οικογένεια των DNN προήλθε από την χρήση στρωμάτων συνέλιξεων τα οποία διοχετεύουν την πηγή εισόδου μέσα από φίλτρα συνέλιξεων τα οποία χαρτογραφούν χαρακτηριστικά των περιοχών που μετασχηματίζουν. Τα χαρακτηριστικά αυτά αποκτούν όλο και μεγαλύτερη αφαίρεση όσο βαθιάει το δυτικό με αποτέλεσμα ένα επαρκώς βαθύ νευρωνικό δίκτυο να μπορεί, χωρίς ανθρώπινη βοήθεια, να καταλήγει σε συμπεράσματα για περίπλοκους σχηματισμούς που δεν μπορούν εύκολα να περιγραφούν. Τα δίκτυα που χρησιμοποιούν στρώματα συνέλιξης στην αρχιτεκτονική τους ονομάζονται Convolutional Neural Networks – CNNs και εναλλάσσουν στρώματα συνέλιξης με στρώματα διορθωτικών γραμμικών μονάδων, Rectified Linear Units -ReLUs [11] οι οποίες επαναφέρουν το αποτέλεσμα των συνέλιξεων σε κανονικά πλαίσια. Επιπλέον των στρωμάτων

συνέλιξης και ενεργοποίησης αξιοποιούνται στρώματα συγκέντρωσης, τα οποία μειώνουν το μέγεθος των πινάκων που προκύπτουν από τις συνελίξεις, μέγεθος που έχει την τάση να αυξάνεται εκρηκτικά, συγκεντρώνοντας τις υψηλότερες, ή μέσες, τιμές ενός πίνακα. Τα στρώματα συνέλιξης των δικτύων αποτελούνται από φίλτρα -πίνακες βαρών τα οποία παρουσιάζονται ενδεικτικά στην Εικόνα 1 και έχουν σαν έξοδο πολλαπλά κανάλια με διαφορετικό βαθμό ενεργοποίησης που εμφανίζονται στην Εικόνα 2.



Εικόνα 2 – Ενεργοποιήσεις πρώτου επίπεδου συνέλιξης και ισχυρότερο κανάλι ενεργοποίησης

Τα DNN, και δη, τα δίκτυα συνέλιξης CNN έχουν χρησιμοποιηθεί κατά κόρον σε έρευνες σχετικά με τον εντοπισμό αντικειμένων σε ιατρικές εικόνες [12]. Ιδιαίτερο ενδιαφέρον έχει δοθεί στην έρευνα θολών εικόνων με ακανόνιστα αντικείμενα ενδιαφέροντος, όπως ο εντοπισμός πολύποδων παχέος εντέρου [13], πνευμονικών οζιδίων [14] [15] και, όπως θα απασχολήσει και εμάς στην παρούσα έρευνα, στην ανίχνευση όγκων σε μαστογραφίες [16] [17] [18] [19] [20] [21].

Στα ιδιαίτερα ελκυστικά χαρακτηριστικά των βαθιών νευρωνικών δικτύων κατατάσσεται και η ικανότητά τους να αξιοποιούν γνώση που έχει προκύψει από παλαιότερα δίκτυα που έχουν εκπαιδευτεί πιθανώς και σε εξαιρετικά διαφορετικά αντικείμενα από τη χρήση για την οποία προορίζονται, υπερβαίνοντας έτσι βασικές δυσκολίες που προκύπτουν από τους αντικειμενικούς περιορισμούς στην εκπαίδευση δικτύων μηχανικής μάθησης, τον χρόνο εκπαίδευσης και τη βάση δεδομένων. Η εκπαίδευση ενός νευρωνικού δικτύου εξ αρχής απαιτεί τεράστια βάση δεδομένων με αριθμό δειγμάτων που, πολλές φορές ξεπερνάει τις εκατοντάδες χιλιάδες [22].

Ο όγκος της αναγκαίας πληροφορίας πέρα από εξαιρετικά δαπανηρός (αγορά και συντήρηση αποθηκευτικού χώρου, ανάγκη επεξεργασίας από ανθρώπινο παράγοντα, πιθανά έξοδα που προκύπτουν από χαρακτηρισμό ως πνευματική ιδιοκτησία, των δεδομένων κατά τη συλλογή ή τη φύλαξή τους), πολλαπλασιάζει και τον αντίστοιχο χρόνο που είναι απαραίτητος για την επεξεργασία των δεδομένων και την χρήση τους στην εκπαίδευση των δικτύων. Τι θα μπορούσε, όμως, να επιτρέψει την χρήση μικρότερων βάσεων δεδομένων, χωρίς την αντίστοιχη απώλεια ακρίβειας που μπορεί να προέλθει από την υποκάλυψη (underfitting) του δικτύου;

Στο ερώτημα που τέθηκε απαντάει η ικανότητα των νευρωνικών δικτύων να εξάγουν συμπεράσματα για τα διακριτικά χαρακτηριστικά των δομών που επεξεργάζονται, χαρακτηριστικά που απεικονίζονται στους πίνακες βαρών των στρωμάτων συνέλιξης [23] και έτσι επιτρέπουν στα CNN να εμφανίζουν εξαιρετικά αποτελέσματα σε σχέση με άλλες μεθόδους μη εποπτευομένης μηχανικής μάθησης (unsupervised machine learning). Το άθροισμα αυτών των ιδιοτήτων οδηγεί στην παρατήρηση των Morgan et al. [24] και Tajbakhsh et al. [25] πως είναι εφικτό να εκπαιδεύσεις ένα βαθύ δίκτυο σε μία εκτεταμένη βάση δεδομένων, φαινομενικά άσχετη με το πραγματικό αντικείμενο της έρευνας και να αξιοποιήσεις τα χαρακτηριστικά που αποτυπώθηκαν στα εκπαιδευμένα βάρη για την αναγνώριση των διαφορετικών αντικειμένων. Αυτή την ικανότητα της μεταφοράς της γνώσης ενός δικτύου (transfer learning) αξιοποιούν οι Bar et al [26] ώστε να εφαρμόσουν τα συμπεράσματα που προέρχονται από

δίκτυα εκπαιδευμένα στις εικόνες της συλλογής ImageNet, σε ακτινογραφίες στήθους, μία τεχνική που θα αξιοποιήσουμε και εμείς στην πορεία της παρούσης εργασίας.

Έχοντας αποκτήσει τη δυνατότητα εκπαίδευσης δικτύων με μια μικρή βάση δεδομένων, μερικών εκατοντάδων δειγμάτων, ο δεύτερος περιορισμός αφορά τον χρόνο εκπαίδευσης. Ακόμα και τα, εμφανώς μικρότερα σε όγκο, δεδομένα που καλούμαστε να διαχειριστούμε απαιτούν ένα ευμέγεθες χρονικό πλαίσιο για την εκπαίδευσή τους. Ο κεντρικός επεξεργαστής ενός υπολογιστή είναι ένα εξαιρετικά βαθύ σύστημα με ικανότητα να πραγματοποιεί πράξεις υψηλής περιπλοκότητας αλλά λειτουργεί σε μεγάλο βαθμό σειριακά. Με αυτό τον τρόπο δεν εξυπηρετεί για τον υπολογισμό απλών μικρών πράξεων συνέλιξης όπως αυτές που λαμβάνουν χώρα στα περισσότερα βαθιά δίκτυα. Για αυτό το λόγο αναπτύχθηκε η χρήση των επεξεργαστών μιας κάρτας γραφικών, αξιοποιώντας τις τεχνολογίες παράλληλου προγραμματισμού CUDA^b και OpenMP^c (γλώσσες που αξιοποιήθηκαν από τις εταιρίες Nvidia και AMD αντίστοιχα) για την παράλληλη επεξεργασία των υπολογισμών που απαιτούνται ώστε να τροφοδοτηθούν και να ανατροφοδοτηθούν τα εκατομμύρια των μεταβλητών που εμφανίζονται ακόμα και σε απλά δίκτυα όπως παρατήρησαν οι Huqani et al. [27] και οι Cengil et al. [28].

Οι ανωτέρω τεχνολογίες, η χρήση προ-εκπαιδευμένων δικτύων και ο συντονισμός του με τα καινούρια δεδομένα με την χρήση καταναλωτικών καρτών γραφικών (consumer-grade GPUs), ανοίγει ένα νέο πεδίο στην έρευνα πάνω στα νευρωνικά δίκτυα, επιτρέποντας σε εξοπλισμό αξίας μερικών εκατοντάδων ευρώ, να υλοποιήσουν προϊόντα (έρευνα, πειράματα και εκπαιδευμένους ανιχνευτές detectors), για την υλοποίηση των οποίων μέχρι τώρα ήταν απαραίτητη η επένδυση μεγάλων κεφαλαίων, καθώς τα πειράματα που πραγματοποιήσαμε απαιτήσαν από 20 έως και 140 ώρες έκαστο, όπως θα αναφερθεί και στα επόμενα κεφάλαια, αριθμός που δεν ανταποκρίνεται βέβαια σε χρονικά πλαίσια εκπαίδευσης που διαρκούν εβδομάδες ώστε να παραχθούν συστήματα εμπορικής αξίας.

Στη συνέχεια περιγράφονται τα απαραίτητα βήματα για την ανάλυση, σχεδιασμό και υλοποίησης των συστημάτων αναγνώρισης καρκινωμάτων στα οποία έχει γίνει αναφορά μέχρι αυτό το σημείο.

1.3 Σχέδιο Δράσης Εργασίας

1.3.1 Στόχοι Εργασίας

Στόχος της εργασίας είναι η υλοποίηση συστήματος μηχανικής μάθησης το οποίο κάνοντας χρήση τεχνολογιών αιχμής ψηφιακής επεξεργασίας εικόνας και νευρωνικών δικτύων θα εντοπίζει και κατηγοριοποιεί καρκινώματα σε παρεχόμενες εικόνες.

επιδιώξουμε την ανάλυση και υλοποίηση των δημοφιλέστερων αρχιτεκτονικών δομών βαθιών νευρωνικών δικτύων, επιχειρώντας τον σχεδιασμό, την υλοποίηση και εκπαίδευση ενός νευρωνικού δικτύου το οποίο, αξιοποιώντας τον συνδυασμό τεχνολογιών αιχμής, θα πετυχαίνει, με την μεγαλύτερη δυνατή ακρίβεια, τον εντοπισμό όγκων σε μαστογραφίες.

Το σύστημα θα εντοπίζει όγκους που διακρίνονται στις εικόνες που του δίδονται και θα τους κατηγοριοποιεί σε καλοήθεις και κακοήθεις. Μετά τον σχεδιασμό, την υλοποίηση και την εκπαίδευση του, θα αξιολογηθεί και η απόδοσή του θα συγκριθεί με τα αρχικά δημοφιλή δίκτυα, παραθέτοντας τα σχετικά διαγράμματα και σχήματα.

1.3.2 Μεθοδολογία ανάλυσης και βηματισμού

Ήδη έχει γίνει αναφορά σε τεχνολογίες που αξιοποιήθηκαν κατά τη διάρκεια της έρευνας όπως η έμφαση σε βαθιά νευρωνικά δίκτυα και η χρήση επεξεργαστή κάρτας γραφικών για την εκπαίδευση των δικτύων. Στην συνέχεια γίνεται εξέταση των δυνατοτήτων τεχνολογιών που προέρχονται από τις πηγές όπως παρατίθενται στη βιβλιογραφία του παραρτήματος και επιχειρείται η επαλήθευση των προτάσεών τους χρησιμοποιώντας τα ειδικά πειραματικά δεδομένα της εργασίας. Η ανάλυση πραγματοποιείται σε τρία επίπεδα όπως παρακάτω.

Στο Κεφάλαιο 2 αναπτύσσονται θέματα με την αρχιοθέτηση και επεξεργασία μιας κατάλληλης βάσης δεδομένων η οποία, αφενός θα παρέχει ποιοτικά δεδομένα με επαρκή ανάλυση για την εκπαίδευση και αξιολόγηση των συστημάτων και αφετέρου απλοποιεί τη διαχείριση των εκατοντάδων δειγμάτων που είναι διαθέσιμα, διευκολύνοντας την συνέχιση της εργασίας. Σε αυτό το στάδιο εφαρμόζονται οι κατάλληλοι μετασχηματισμοί στα δείγματά ώστε να μειωθεί το μέγεθος των εικόνων,

^b <https://developer.nvidia.com/cuda-zone>

^c <https://www.openmp.org/>

μειώνοντας αντίστοιχα τον απαραίτητο χρόνο επεξεργασίας και τις απαιτήσεις χώρου του συστήματος, διατηρώντας πάντα τη μέγιστη ανάλυση και εφαρμόζοντας τεχνικές αύξησης της αντίθεσης και ευκρίνειας των εικόνων.

Στην συνέχεια γίνεται εμβάθυνση στα δίκτυα τεχνολογίας RCNN και αναφέρονται πληθώρα συγγραμμάτων που προτείνουν τις δημοφιλέστερες σήμερα αρχιτεκτονικές βαθιών νευρωνικών δικτύων συνέλιξης. Οι μέθοδοι αυτές υλοποιούνται σε περιβάλλον MATLAB από τις οποίες και αντλούνται οι τεχνικές που εφαρμόζονται για την πολυδιάστατη εμβάθυνσή των δικτύων που υλοποιούν επιδιώκοντας τη διάκριση των πιο αποτελεσματικών όσον αφορά τον τύπο δεδομένων που διαχειρίζεται η έρευνά μας. Γίνεται αρχική εκτίμηση της αποτελεσματικότητας των δομών και περιγράφονται οι αποτελεσματικότερες.

Στο Κεφάλαιο 4 πραγματοποιείται αναλυτική αναφορά στις ρυθμιστικές παραμέτρους που οριοθετούν τα μεγέθη που χρησιμοποιούνται κατά την εκπαίδευση των δικτύων και παρουσιάζεται το καταλληλότερο φάσμα τιμών για την λεπτομερέστερη και πιο ολοκληρωμένη υλοποίηση των πειραμάτων, διατηρώντας πάντα υπόψιν τις δομικές διαφορές των ποικίλων δικτύων, όπως και την υποκειμενικότητα τόσο των δεδομένων, όσο και των αποτελεσμάτων.

Σε αυτό το σημείο, αφού έχουν τυποποιηθεί οι παράμετροι εκπαίδευσης που εγγυόνται τη μεγαλύτερη απόδοση των δικτύων, γίνεται αναφορά και ανάλυση των μεθόδων αξιολόγησης των δικτύων που έχουν υλοποιηθεί. Ακολουθώς απαριθμούνται μια σειρά πειραματισμών που αξιοποιούν τις βασικές αρχιτεκτονικές δομές όπως και επεμβάσεις και παραλλαγές σε αυτές ώστε να προκύψουν πρωτότυπα δίκτυα με απόδοση ικανοποιητικότερη των βασικών.

Τέλος περιγράφονται δυνατότητες επέκτασης των δικτύων, τόσο σε επίπεδο ικανοτήτων πάνω στο θέμα της παρούσης Πτυχιακής Εργασίας, όσο και σε άλλα επιστημονικά αντικείμενα.

1.4 Πειραματικός Εξοπλισμός

Ο σχεδιασμός, η υλοποίηση, εκπαίδευση και αξιολόγηση των δικτύων που αναλύθηκαν και παρουσιάζονται κατά τη διάρκεια της παρούσας εργασίας πραγματοποιήθηκε σε σύστημα με χρήση του ακόλουθου εξοπλισμού και λογισμικού.

1.4.1 Hardware

- Motherboard – Gigabyte Z370 HD3P (rev. 1.0)^d
- CPU – Intel i5 8400 6core 2.8GHz^e
- Memory – Corsair Vengeance RAM 16GB 3000MHz^f
- GPU – Asus GTX 1060 6GB 9Gbps^g
- Storage – Samsung 850 250GB^h & Western Digital Caviar Black 1TBⁱ
- PSU – Evga SuperNova 650 G3 Y2^j

1.4.2 Software

- Microsoft Windows 10 version 1803^k
- Mathworks MATLAB r2018b^l με τα παρακάτω πρόσθετα
 - Computer Vision System Toolbox
 - Deep Learning Toolbox
 - Image Processing Toolbox
- Microsoft Office 2016^m

^d <https://www.gigabyte.com/Motherboard/Z370-HD3P-rev-10>

^e <https://ark.intel.com/products/126687/Intel-Core-i5-8400-Processor-9M-Cache-up-to-4-00-GHz->

^f <https://www.corsair.com/eu/en/Memory-Size/vengeance-lpx-black/p/CMK16GX4M2B3000C15>

^g <https://www.asus.com/us/Graphics-Cards/GTX1060-O6G-9GBPS/>

^h <https://www.samsung.com/us/computing/memory-storage/solid-state-drives/ssd-850-evo-2-5-sata-iii-250gb-mz-75e250b-am/>

ⁱ <https://www.wd.com/products/internal-storage/wd-black-desktop.html#WD1003FZEX>

^j <https://eu.evga.com/products/product.aspx?pn=220-G3-0650-Y2>

^k <https://docs.microsoft.com/en-us/windows/whats-new/whats-new-windows-10-version-1803>

^l https://www.mathworks.com/products/new_products/latest_features.html

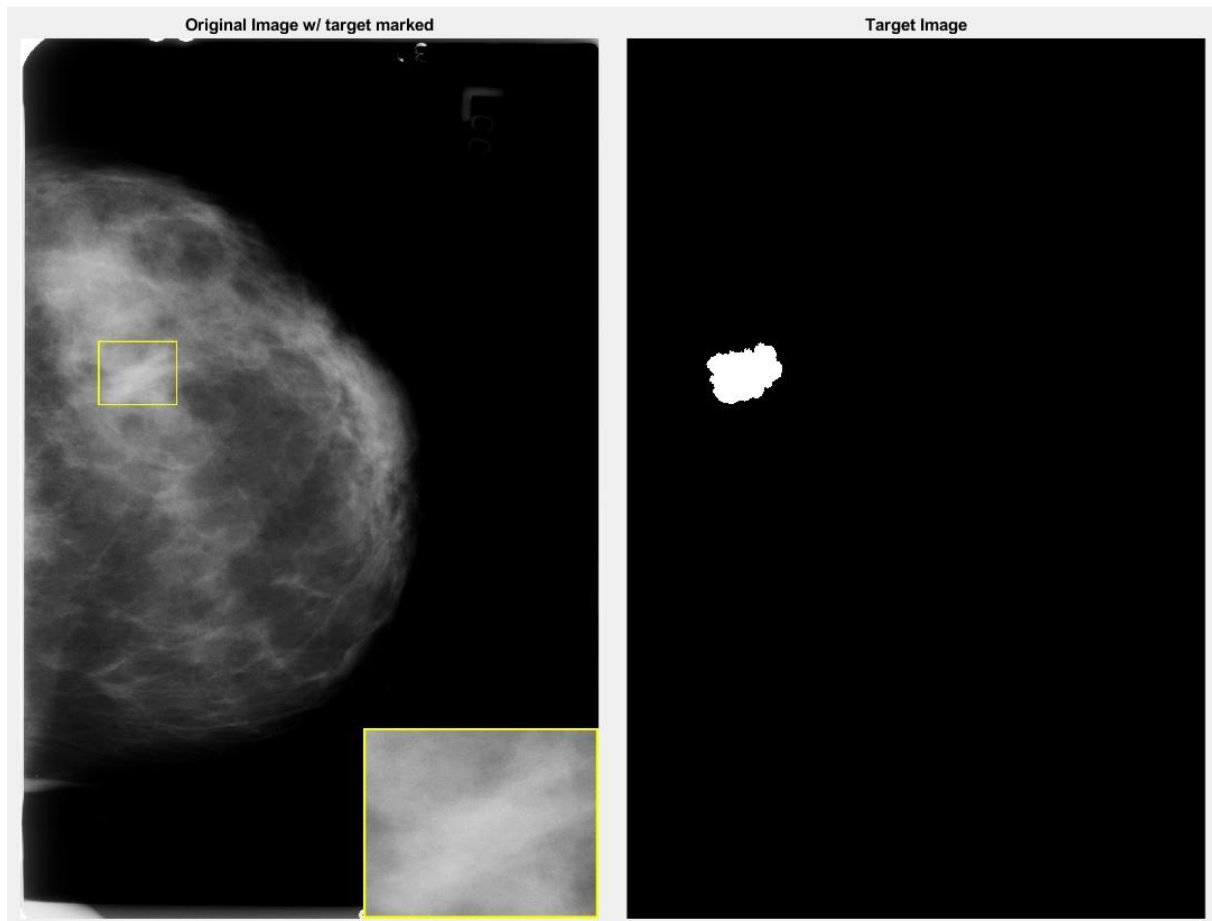
^m <https://products.office.com/el-GR/>

2 Παρουσίαση και Επεξεργασία Πειραματικών Δεδομένων

Για τις ανάγκες των πειραμάτων, που πραγματοποιήθηκαν κατά την ερευνητική διαδικασία που περιγράφεται στη παρούσα εργασία, είναι απαραίτητο να χρησιμοποιηθούν δεδομένα που θα πληρούν υψηλές προδιαγραφές ποιότητας, ώστε να διατηρηθεί η ακρίβεια των αποτελεσμάτων, ο βαθμός σημαντικότητας των ρυθμίσεων, όπως και η επαναληψιμότητα και αναπαραγωγιμότητα των πειραμάτων. Το υψηλό κόστος της επεξεργασίας εικόνων από ανθρώπινο παράγοντα και γαρ επιστήμονες, αυθεντίες σε ένα αντικείμενο όπως τομείς της ιατρικής είναι χαρακτηριστική τροχόπεδη στην έρευνα, προβληματισμός που, όπως προαναφέρθηκε απαντάται δια της τεχνολογικής προόδου, η οποία επιτρέπει την εύρυθμη διαχείριση μεγάλων όγκων δεδομένων, εφόσον αυτά είναι διαθέσιμα.

2.1 Αναφορά στη Βάση Δεδομένων CBIS-DDSM

Σε αυτή την ανάγκη παροχής ποιοτικών ερευνητικών δεδομένων απαντά ο ιστοχώρος Cancer Imaging Archive [29], μία υπηρεσία συλλογής και καταγραφής ιατρικών εικόνων καρκινικών διαγνώσεων. Από τις ποικίλες συλλογές του, που εκτείνονται από συλλογές μερικών δεκάδων περιπτώσεων σε βάσεις με ιατρικά δεδομένα χιλιάδων ασθενών, βάσεις όγκου δεκάδων Terabyte, επιλέξαμε για τους σκοπούς της εργασίας την συλλογή Curated Breast Imaging Subset of DDSM / CBIS-DDSM [30, 31] μία αναβαθμισμένη και κανονικοποιημένη έκδοση της Ψηφιακής Βάσης Μαστογραφιών / Digital Database for Screening Mammography (DDSM).

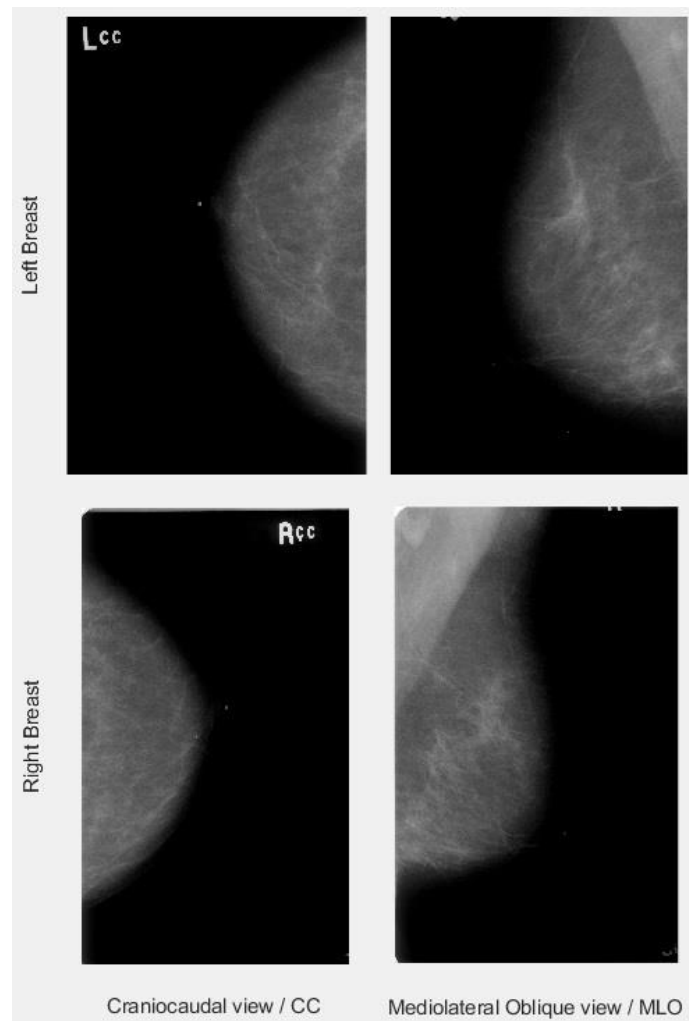


Εικόνα 3 – Ενδεικτικές εικόνες συλλογής CBIS-DDSM

Η CBIS-DDSM είναι ελεύθερη προς μη εμπορική χρήση και αποτελείται από 10239 εικόνες συνολικού όγκου 163.6GB. Οι εικόνες βρίσκονται σε μορφή Dicom (.dcm), μια κωδικοποίηση που απαντάται συχνά σε εικόνες ιατρικών δεδομένων καθώς προσφέρει την δυνατότητα εισαγωγής μετά-δεδομένων (meta-data) τα οποία κωδικοποιούν και μοναδικοποιούν την εκάστοτε εικόνα προστατεύοντας παράλληλα τα προσωπικά δεδομένα των ασθενών που απεικονίζονται.

Οι εικόνες της συλλογής κατατάσσονται σε τρεις κατηγορίες:

1. Πλήρης εικόνας μαστογραφίας στις οποίες έχει εντοπιστεί από εκπαιδευμένο ακτινολόγο τουλάχιστον ένα όγκος, καλοήθης ή κακοήθης και οποίος αποτελεί το αντικείμενο ενδιαφέροντος (Region Of Interest - ROI) που καλούνται να εντοπίσουν τα υλοποιημένα δίκτυα.
2. Εικόνα του αντικείμενου ενδιαφέροντος αποκομμένου (cropped) από την αρχική εικόνα, η συγκεκριμένη κατηγορία δεν εξυπηρετεί τους σκοπούς της παρούσας εργασίας.
3. Εικόνα του αντικείμενου ενδιαφέροντος σημειωμένου στην αρχική εικόνα ως λευκά εικονοστοιχεία σε μαύρο φόντο, η συγκεκριμένη εικόνα επιτρέπει την αυτοματοποίηση της δημιουργίας πινάκων αληθείας οι οποίοι εξυπηρέτησαν στην εκπαίδευση και έλεγχο των δικτύων που υλοποιήθηκαν.



Εικόνα 4 – Ενδεικτικά διαθέσιμες απεικονίσεις για την ασθενή με κωδικό P_00038

Η συνολική συλλογή απαρτίζεται από απεικονίσεις αντικειμένων τα οποία έχουν διαχωριστεί σε ασβεστώσεις (calcification) και όγκους (mass). Για τις ανάγκες της εργασίας χρησιμοποιείται μόνο η συλλογή των όγκων καθώς τα αντικείμενα ενδιαφέροντος των ασβεστώσεων είναι ιδιαίτερα πολύπλοκα, πολυποίκιλα και δυσθεώρητα, με μεγάλες διαφορές σε χαρακτηριστικά όπως όγκο, πυκνότητα και σχήμα, με αποτέλεσμα να μειώνουν την αποτελεσματικότητα των δικτύων στην ανίχνευση αντικειμένων.

Περαιτέρω διαχωρισμός προκύπτει βάση της γωνίας απεικόνισης της μαστογραφίας. Τα δεδομένα προέρχονται από τις πιο ενδιαφέρουσες διαγνωστικά απεικονίσεις, την Κεφαλουραία (Craniocaudal view / CC) και την Μεσοπλάγια λοξή (Mediolateral Oblique / MLO) όπως εμφανίζονται και στην Εικόνα 4. Από τα πειράματά μπορεί να παρατηρηθεί πως οι δύο προβολές έχουν κοινά χαρακτηριστικά και είναι επωφελές το να ενσωματωθούν αμφότερες σε μία κοινή συλλογή εκπαίδευσης και ελέγχου, ώστε να αυξηθεί το ωφέλιμο δείγμα των δεδομένων μας, σε αντίθεση του να χωριστούν σε δύο διακριτές λίστες.

Η συλλογή συνοδεύεται από αρχεία comma separated values - .csv τα οποία περιέχουν τα αποτελέσματα που έχουν σημειώσει οι διαγνώστες ακτινολόγοι και τα οποία θα αξιοποιηθούν ώστε να χαρακτηριστεί το κάθε αντικείμενο ενδιαφέροντος ως καλοήθες/Benign ή κακοήθες/Malignant. Τα αρχεία αυτά περιέχουν περισσότερες μορφολογικές λεπτομέρειες όπως το σχήμα (ωοειδές, με λοβούς ή ακανόνιστο) και τα όρια (περιτμημένα, αποκρυσμμένα, ακανθωτά ή ακαθόριστα) του αντικειμένου, χαρακτηρισμούς και κατηγοριοποιήσεις με τους οποίους μπορεί να εμπλουτιστεί το σύστημά σε επόμενη υλοποίησή του.

Τέλος η συλλογή έχει διαχωριστεί τυχαία σε δύο τμήματα, μία συλλογή 583 εικόνων, το 63% των συνολικά διαθέσιμων μαστογραφιών, για την εκπαίδευση των συστημάτων και μία δεύτερη 345 εικόνων για τον έλεγχο των επιδόσεών τους, γεγονός που εξασφαλίζει πως τα συστήματα θα αξιολογηθούν σε περιπτώσεις που δεν έχουν χρησιμοποιηθεί για την εκπαίδευσή τους.

2.2 Επεξεργασία και μορφοποίηση εικόνων

Στην συνέχεια γίνεται αναφορά στα βήματα που ακολουθήθηκαν για την επεξεργασία των δεδομένων, λαμβάνοντας σαν πρώτη ύλη τις εικόνες που παρέχονται μέσω της συλλογής CBIS-DDSM και με αποτέλεσμα ενισχυμένες εικόνες και πίνακες αληθείας που περιέχουν τις περιοχές ενδιαφέροντος που χρησιμοποιήθηκαν για την εκπαίδευση και τον έλεγχο των δικτύων που θα αναλυθούν στα επόμενα κεφάλαια.

2.2.1 Μετατροπή τύπου, διάνοιξη και αποκοπή

Ενώ τα αρχεία .dcm περιέχουν χρήσιμη πληροφορία για την αναφορά της κάθε περίπτωσης στην αρχική μελέτη DDSM και την ταυτοποίηση των περιπτώσεων, είναι περίπλοκα στον χειρισμό τους και δεν εξυπηρετούν για αυτούσια χρήση στους πειραματισμούς που θα υλοποιηθούν. Για αυτό τον λόγο κάθε αρχείο μετατράπηκε σε εικόνα jpg διατηρώντας τις αρχικές διαστάσεις ώστε να υπάρχει η δυνατότητα για περαιτέρω μετασχηματισμούς και ψηφιακής επεξεργασίας της εικόνας διατηρώντας την μέγιστη ανάλυση.

Σε κάθε μία περίπτωση δόθηκε προσοχή ώστε να διατηρηθεί στην ονομασία του αρχείου ο αριθμός πεδίου που αντιστοιχεί στην θέση του στην αρχική συλλογή και η ένδειξη που αφορά σε ποια κατηγορία ανήκει η εικόνα.

Από παρατήρηση των δεδομένων προκύπτει πως σε κάθε μαστογραφία μόνο το 50%, κατά μέσο όρο, των συνολικών εικονοστοιχείων περιέχει χρήσιμη πληροφορία και το υπόλοιπο ποσοστό καλύπτεται είτε από μαύρα pixel είτε από δυνητικά επιζήμια πληροφορία (περιοχές θώρακα ασθενούς, ανάρτηση τύπου μαστογραφίας, θόρυβος εικόνας κ.ο.κ.) Η αφαίρεση αυτών των τμημάτων αυξάνει την ανάλυση των τελικών εικόνων καθώς επιτρέπει τη πυκνωση της χρήσιμης πληροφορίας. Η πυκνωση αυτή είναι επωφελής λαμβάνοντας υπόψιν τους υλικούς περιορισμούς όπως, κυρίως, τη διαθέσιμη μνήμη της κάρτας γραφικών και σε δεύτερο χρόνο την ισχύ του επεξεργαστή και τη συνολική διαθέσιμη μνήμη του συστήματος, μεγέθη τα οποία καθορίζουν τις τελικές διαστάσεις των εικόνων στις οποίες εκπαιδεύτηκαν τα δίκτυα.

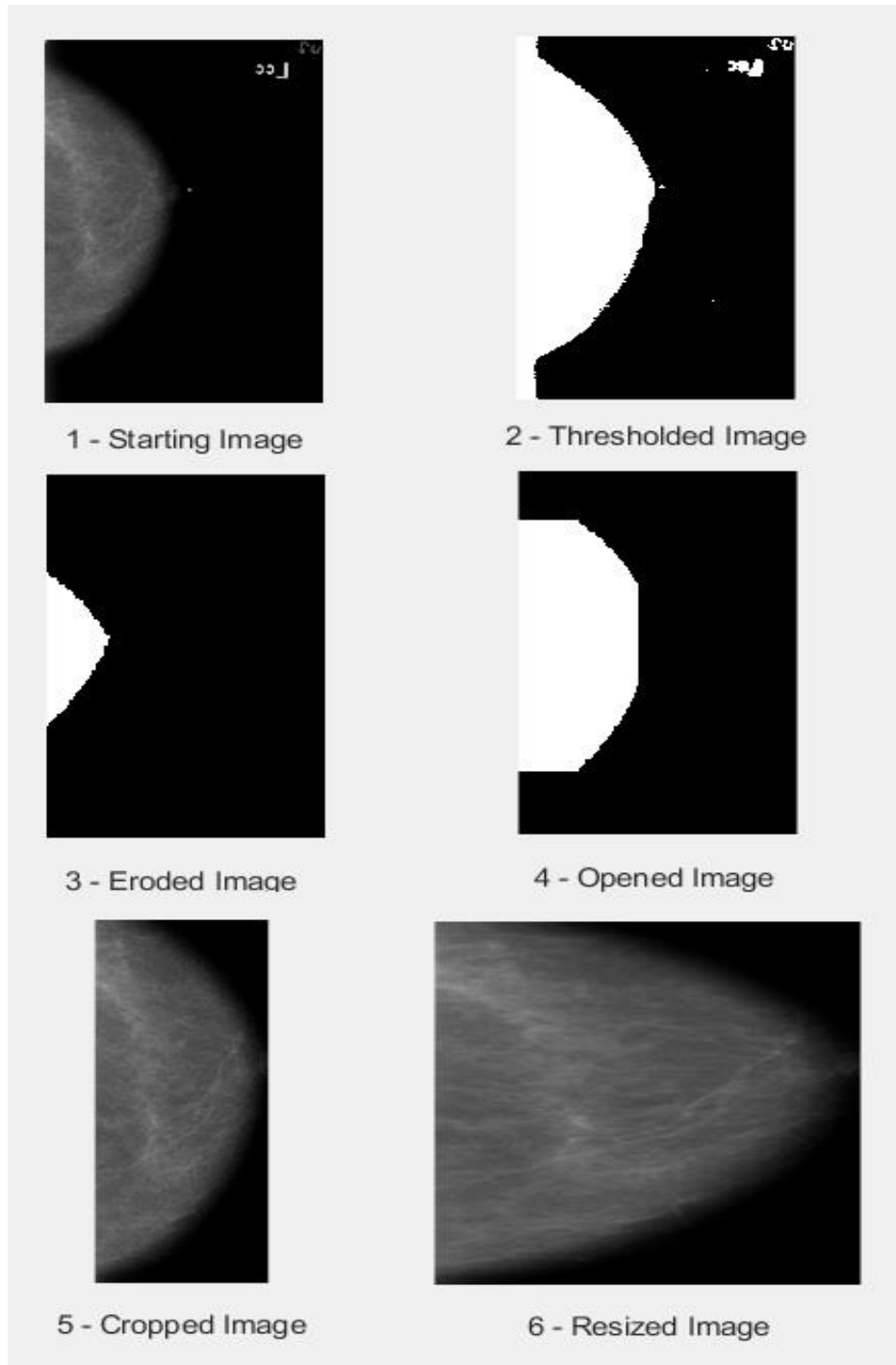
Με σκοπό την αυτοματοποίηση της επεξεργασίας των δεδομένων αξιοποιήθηκαν οι δυνατότητες επεξεργασίας και μορφοποίησης εικόνας που παρέχει το περιβάλλον MATLAB και πιο συγκεκριμένα αξιοποιήθηκαν οι εντολές `imerode` και `imdilate` για την διάνοιξη της εικόνας, με τον ίδιο, υψηλό, αριθμό επαναλήψεων και ώστε βάση του, γενικά ωοειδούς, σχήματος του μαστού να σχηματιστεί μία ενιαία περιοχή η οποία περιλαμβάνει την χρήσιμη πληροφορία. Απαραίτητος σε αυτό το σημείο ήταν ο έλεγχος της συσταλμένης εικόνας ώστε να επιβεβαιωθεί πως έχει διατηρηθεί ο πυρήνας του σχήματος που απαιτείται για την επανασύσταση του σχήματος και να χρησιμοποιηθεί μειωμένος αριθμός επαναλήψεων σε περίπτωση που η επανασύσταση δεν ήταν εφικτή. Οι εντολές `regionprops` και `imcrop` οριοθετούν και αποκόπτουν την τελική εικόνα, αφού πρώτα επιβεβαιωθεί πως στην αντίστοιχη περιοχή της εικόνας του στόχου περιλαμβάνεται το αντικείμενο ενδιαφέροντος.

Τέλος, στην αποκομμένη εικόνα, η οποία αποτελείται πλέον κατά 81% από χρήσιμη πληροφορία, εντοπίζεται με χρήση της παραμέτρου `regionprops.centroids`, το κέντρο βάρους του κεντρισμένου πλέον σχήματος. Οι περιπτώσεις στις οποίες το κέντρο βάρους εντοπίζεται στα δεξιά του μέσου της εικόνας συμπεραίνεται πως ανήκουν σε μαστογραφίες δεξιού μαστού, οι οποίες και αντικατοπτρίζονται ως προς την οριζόντια διάστασή με χρήση της εντολής `flip`. Τοιούτοτρόπως επιτυγχάνεται μεγαλύτερη ομοιογένεια και αποφεύγεται η ανάγκη εκπαίδευσης (και πιθανώς λάθους ενεργοποίησης) των

νευρώνων που αφορούν τις άνω και κάτω περιοχές της δεξιάς πλευράς των εικόνων, βεβαιώνοντας έτσι την ποιότητα της επεξεργασίας καθώς δεν θα υπάρχουν χρήσιμα δεδομένα σε περιοχές που πιθανώς να εξαιρεθούν από έλεγχο λόγω ελλειμματικής επένδυσης (padding).

2.2.2 Μετασχηματισμός

Κάθε μία από τις αρχιτεκτονικές των δικτύων που θα χρησιμοποιηθεί απαιτεί σαν είσοδο μία εικόνα συγκεκριμένων διαστάσεων. Τα δίκτυα alexnet, όπως έχουν υλοποιηθεί στο MATLAB, δέχονται εικόνες τριών χρωματικών καναλιών (rgb) και διαστάσεων 227X227 εικονοστοιχείων, άρα διαστάσεις εισόδου εικόνας [227 227 3], το GoogLeNet [224 224 3] και το YOLO [448 448 3]. Οι μορφοποιημένες εικόνες που κατασκευάστηκαν στα προηγούμενα βήματα έχουν διαστάσεις κατά μέσο όρο 5295 X 3134 pixels και, καθώς αποτελούνται από αποχρώσεις του γκρι, ένα μόνο πίνακα δεδομένων.



Εικόνα 5 – Βήματα μετασχηματισμού από αρχική σε επεξεργασμένη εικόνα

Ο αλγόριθμος trainFasterRCNNObjectDetector που θα χρησιμοποιηθεί, μορφοποιεί τις εικόνες εισόδου ώστε να έχουν τις κατάλληλες διαστάσεις κατά την εκπαίδευση, εφόσον πάντα η αρχική εικόνα έχει διαστάσεις μεγαλύτερες ή ίσες από αυτές που καθορίζονται από το image input layer αλλά δεν μπορεί να επεξεργαστεί εικόνες με διαφορετικό αριθμό καναλιών από τον απαιτούμενο. Για να ξεπεραστεί αυτόν τον προβληματισμό απαιτείται η επεξεργασία της αρχιτεκτονική του δικτύου που χρησιμοποιείται ώστε να ταιριάζει με την μορφή των δεδομένων μας. Η αντικατάσταση του στρώματος εισόδου και του πρώτου στρώματος συνέλιξης του δικτύου με τα αντίστοιχα στρώματα παραμετροποιημένα κατάλληλα είναι το πρώτο βήμα στο οποίο θα γίνει αναφορά στο Κεφάλαιο 4.2.

Για την μεγαλύτερη δυνατή ομοιογένεια και τη μείωση των απαιτήσεων μνήμης για την αποθήκευση και εκπαίδευση των δικτύων καθορίστηκαν σαν διαστάσεις εισόδου, ένα χρωματικό κανάλι με ύψος και πλάτος 224 στοιχείων. Σε αυτό το μέγεθος απαιτήθηκε και ο μετασχηματισμός των εικόνων με χρήση της εντολής `imresize`. Τα βήματα που ακολουθήθηκαν, μέχρι αυτό το σημείο, παρουσιάζονται συνοπτικά στην Εικόνα 5.

2.2.3 Ενίσχυση αντίθεσης

Οι Cheng et al. [16] αναλύουν τη σημασία της προ-επεξεργασίας των εικόνων με σκοπό την αύξηση της αντίθεσης (Contrast) μεταξύ αντικειμένου ενδιαφέροντος και υποβάθρου σε μια εικόνα ιατρικού περιεχομένου. Η μαστογραφία αποτελεί τη διςδιάστατη απεικόνιση ενός τρισδιάστατου αντικειμένου και αυτό συνεπάγεται πως σημεία της εικόνας αποτελούνται από ετερογενή στρώματα λιπώδους ιστού, γαλακτοφόρους κόλπους, αιμοφόρα αγγεία, μαστικούς αδένες και συνδέσμους, αντικείμενα τα οποία, λόγω της αθροιστικής τους πυκνότητας εμφανίζονται εντονότερα από πιθανούς όγκους. Για αυτούς τους λόγους μέθοδοι όπως η καθολική ή τοπική κατωφλίωση (thresholding) που προτείνονται δεν είναι αποτελεσματικοί για τις ανάγκες μας, σε αντίθεση με άλλους όπως η διάνοιξη του ιστογράμματος. Καθώς οι διαθέσιμες μαστογραφίες αποτελούνται από διαβαθμίσεις του γκρι, είναι χρήσιμη η επεξεργασία των δεδομένων σχετικά με το ποσοστό που καταλαμβάνει στην εικόνα κάθε διακριτή τιμή έντασης φωτεινότητας όπως απεικονίζεται και στο ιστόγραμμα της εικόνας.

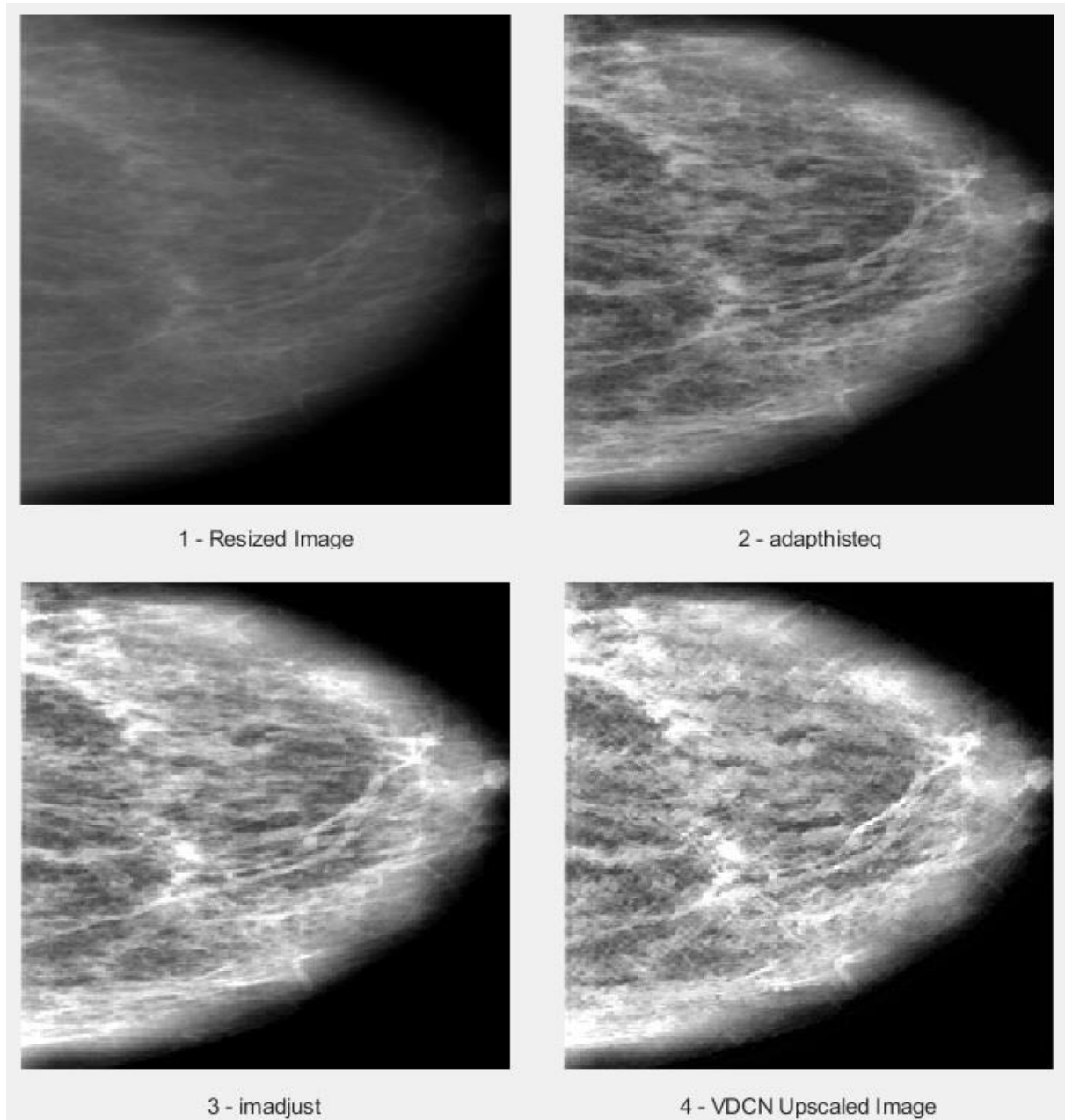
Ο εξοπλισμός της σουίτας MATLAB σχετικά με την ενίσχυση της αντίθεσης απαρτίζεται από τους παρακάτω μετασχηματισμούς:

- `Histeq`, Ο αλγόριθμος `histeq` ισοσταθμίζει το ιστόγραμμα της εικόνας, λαμβάνοντας υπόψιν του το σύνολο της εικόνας μεταβάλλει τις τιμές φωτεινότητας έτσι ώστε η στατιστική τους κατανομή να τείνει προς την κανονική κατανομή (uniform distribution).
- `Adapthisteq` σε αντίθεση με την προηγούμενη εντολή, η `adapthisteq` χωρίζει την εικόνα σε τμήματα (tiles) στα οποία εφαρμόζει εξίσωση ιστογράμματος σε κάθε ένα ξεχωριστά. Όπως παρατηρείται ο συγκεκριμένος αλγόριθμος εμφανίζει ευκρινέστερα αποτελέσματα από τον προηγούμενο ο οποίος έχει την τάση να δημιουργεί θόρυβο, ιδιαίτερα στις φωτεινές περιοχές τις εικόνας, κρύβοντας πιθανά αντικείμενα ενδιαφέροντος.
- `Imadjust`, η συγκεκριμένη εντολή αυξάνει την αντίθεση της εικόνας, μεταβάλλοντας την τιμή της φωτεινότητας κάθε εικονοστοιχείου με τέτοιο τρόπο ώστε να κορεστεί το ένα τοις εκατό των υψηλότερων και χαμηλότερων τιμών φωτεινότητας στις τιμές που αποτελούν το απόλυτα μαύρο και λευκό. Ο μετασχηματισμός αυτός δείχνει εξαιρετικά αποτελέσματα όταν ακολουθεί την ομαλοποίηση του ιστογράμματος καθώς καθιστά ευδιάκριτους ποικίλους σχηματισμούς των ιστών του μαστού όπως παρατηρείται και στην Εικόνα 6.

Στη συνέχεια γίνεται χρήση των μετασχηματισμών `adapthisteq` και `imadjust` σε σειρά για τη μεγαλύτερη δυνατή ευκρίνεια. Παρατηρείται εμφανώς σε αυτό το σημείο πως ο μετασχηματισμός των διαστάσεων της εικόνας που έλαβε χώρα στην προηγούμενη ενότητα, καθώς προκύπτει από τη μεταβολή του αρχικού παραλληλόγραμμου σχήματος σε τετραγωνισμένο, παραμορφώνει την οριζόντια διάσταση σε μεγαλύτερο βαθμό από την κάθετη. Αυτή η επέμβαση οδηγεί σε αυξημένο θόρυβο στην εικόνα, ο οποίος απαιτεί αποκατάσταση.

Οι Dong et al [32], παρατηρούν πως οι διαδρομές που προτείνονται έχουν ομοιότητες με ένα DNN και αναλύουν το SRCNN, ένα δίκτυο πλήρους συνέλιξης το οποίο μέσα από τις διαδοχικές συνελιξεις που εφαρμόζει ομαλοποιεί τις τιμές των στοιχείων. Οι Jain et al. [33] περιγράφουν ένα CNN το οποίο εμφανίζει εξαιρετικά αποτελέσματα στην απαλοιφή θορύβου Gauss. Η εργασία τους επαυξάνεται από τους Kim et al [34] οι οποίοι αντιπροτείνουν ένα βαθύτερο δίκτυο το οποίο μπορεί να αφομοιώσει την πληροφορία από μεγάλες βάσεις εικόνων όπως το ImageNet (το SRCNN εμφανίζει φθίνουσα απόδοση

κατά τη μεταφορά του από ένα δείγμα εκπαίδευσης 91 εικόνων στο ImageNet). Το δίκτυο που προτείνεται ονομάζεται Very Deep Convolutional Network (VDCN) και έχει υλοποιηθεί στο περιβάλλον του MATLAB. Έχει εκπαιδευτεί για την αύξηση των διαστάσεων μιας εικόνας διατηρώντας σταθερή την ανάλυσή της, ωστόσο, καθώς το σύστημα εξυπηρετεί στην διατήρηση αυστηρών ευθειών (σε αντίθεση με το SRCNN) για τις ανάγκες της εργασίας μπορεί να αξιοποιηθεί στην αφαίρεση της παραμόρφωσης, διατηρώντας τις διαστάσεις της εικόνας σταθερές.



Εικόνα 6 – Βήματα ενίσχυσης αντίθεσης

2.3 Συλλογή δεδομένων και καταγραφή αρχείων Ground Truth

Έχοντας φέρει σε πέρας τον κατάλληλο μετασχηματισμό των εικόνων, γίνεται χρήση του κατάλληλου αλγορίθμου ώστε να δημιουργηθούν οι πίνακες αλήθειας (ground truth), με τους οποίους θα υλοποιηθεί η εκπαίδευση και αξιολόγηση των συστημάτων. Καθώς έχει γίνει διαχωρισμός του στόχου σε δύο κατηγορίες στις οποίες θα χωριστούν τα δείγματα, δημιουργήται ένας πίνακα με τρεις στήλες. Στην πρώτη στήλη αποτυπώνεται η διεύθυνση στην οποία είναι αποθηκευμένη κάθε μια μαστογραφία που έχει υποστεί επεξεργασία ώστε να μπορεί να αξιοποιηθεί αποδοτικά από τους αλγόριθμους εκπαίδευσης.

Για κάθε μία εικόνα η οποία δεν αποτελεί μαστογραφία, επιβεβαιώνοντας πως έχει τις ίδιες ακριβώς διαστάσεις με την τελευταία μαστογραφία που υπέστη επεξεργασία, γίνεται κατανοητό πως αποτελεί την εικόνα της μαστογραφίας στην οποία είναι σημειωμένο το αντικείμενο ενδιαφέροντος-στόχος. Σε περίπτωση που η αναφερόμενη εικόνα δεν έχει τις ίδιες διαστάσεις με την τελευταία μαστογραφία και δεν αποτελεί μαστογραφία η ίδια, γίνεται αντιληπτό πως ανήκει στην δεύτερη κατηγορία εικόνων, μεγεθυμένη απεικόνιση του αντικειμένου ενδιαφέροντος ή αποτελεί λάθος καταχώρηση κατά τη σύνταξη της συλλογής και απορρίπτεται.

Οι εικόνες διαθέτουν τις ίδιες διαστάσεις με τις μαστογραφίες που ακολούθησαν, αποκόπτονται σύμφωνα με τη ωφέλιμη περιοχή που έχει προσδιοριστεί από τις προηγούμενες επαναλήψεις του αλγόριθμου, καταφλιώνονται και στη συνέχεια εξάγεται από αυτές η περιοχή στην οποία περιλαμβάνεται το αντικείμενο ενδιαφέροντος. Η περιοχή αυτή κατηγοριοποιείται σύμφωνα με τα μετα-δεδομένα της συλλογής και αποθηκεύεται στη δεύτερη ή τρίτη στήλη του πίνακα αληθείας αντίστοιχα. Με την πλήρη εκτέλεση του αλγόριθμου έχουν προκύψει δύο αρχεία πινάκων αλήθειας, αφού γίνει αφαίρεση τυχών κενών πεδίων, ένα για εκπαίδευση και ένα για την αξιολόγηση των συστημάτων.

	1 imageFilename	2 Benign	3 Malignant
1	'...\P_00001_LEFT_MLO_0.jpg'	[]	[29.5000,150.5000,22,12]
2	'...\P_00004_RIGHT_MLO_0.jpg'	[75.5000,155.5000,46,15]	[]
3	'...\P_00015_LEFT_MLO_0.jpg'	[]	[55.5000,96.5000,21,10]
4	'...\P_00021_LEFT_MLO_0.jpg'	[164.5000,127.5000,27,13]	[]
5	'...\P_00021_RIGHT_CC_0.jpg'	[153.5000,96.5000,22,11]	[]
6	'...\P_00021_RIGHT_MLO_0.jpg'	[198.5000,119.5000,26,18]	[]
7	'...\P_00026_LEFT_CC_0.jpg'	[101.5000,131.5000,19,9]	[]
8	'...\P_00026_LEFT_MLO_0.jpg'	[106.5000,144.5000,21,7]	[]
9	'...\P_00027_RIGHT_CC_0.jpg'	[59.5000,115.5000,24,17]	[]
10	'...\P_00027_RIGHT_MLO_0.jpg'	[77.5000,92.5000,30,14]	[]
11	'...\P_00034_RIGHT_CC_0.jpg'	[]	[65.5000,28.5000,33,16]
12	'...\P_00034_RIGHT_MLO_0.jpg'	[]	[143.5000,74.5000,27,14]
13	'...\P_00039_RIGHT_MLO_0.jpg'	[]	[78.5000,181.5000,45,22]
14	'...\P_00041_LEFT_CC_0.jpg'	[118.5000,39.5000,13,7]	[]
15	'...\P_00041_LEFT_MLO_0.jpg'	[125.5000,119.5000,13,8]	[]

Εικόνα 7 – Ενδεικτική απεικόνιση πίνακα αλήθειας εκπαίδευσης

Το αρχείο που θα αξιοποιείται για την εκπαίδευση των συστημάτων περιέχει 583 μαστογραφίες στις οποίες μπορούν να εντοπιστούν 633 όγκοι, εκ των οποίων 317 κακοήθεις και 316 καλοήθεις, ενώ το αρχείο αξιολόγησης αποτελείται από 359 όγκους, εκ των οποίων 143 κακοήθεις και 216 καλοήθεις. Συνολικά, οι 460 κακοήθεις όγκοι έχουν, κατά μέσο όρο, διαστάσεις Πλάτος Επί Ύψος 43.7 X 23.7 με διάμεσο 36 X 18 ενώ οι καλοήθεις 39.2 X 25.6 με διάμεσο 31 X 15.

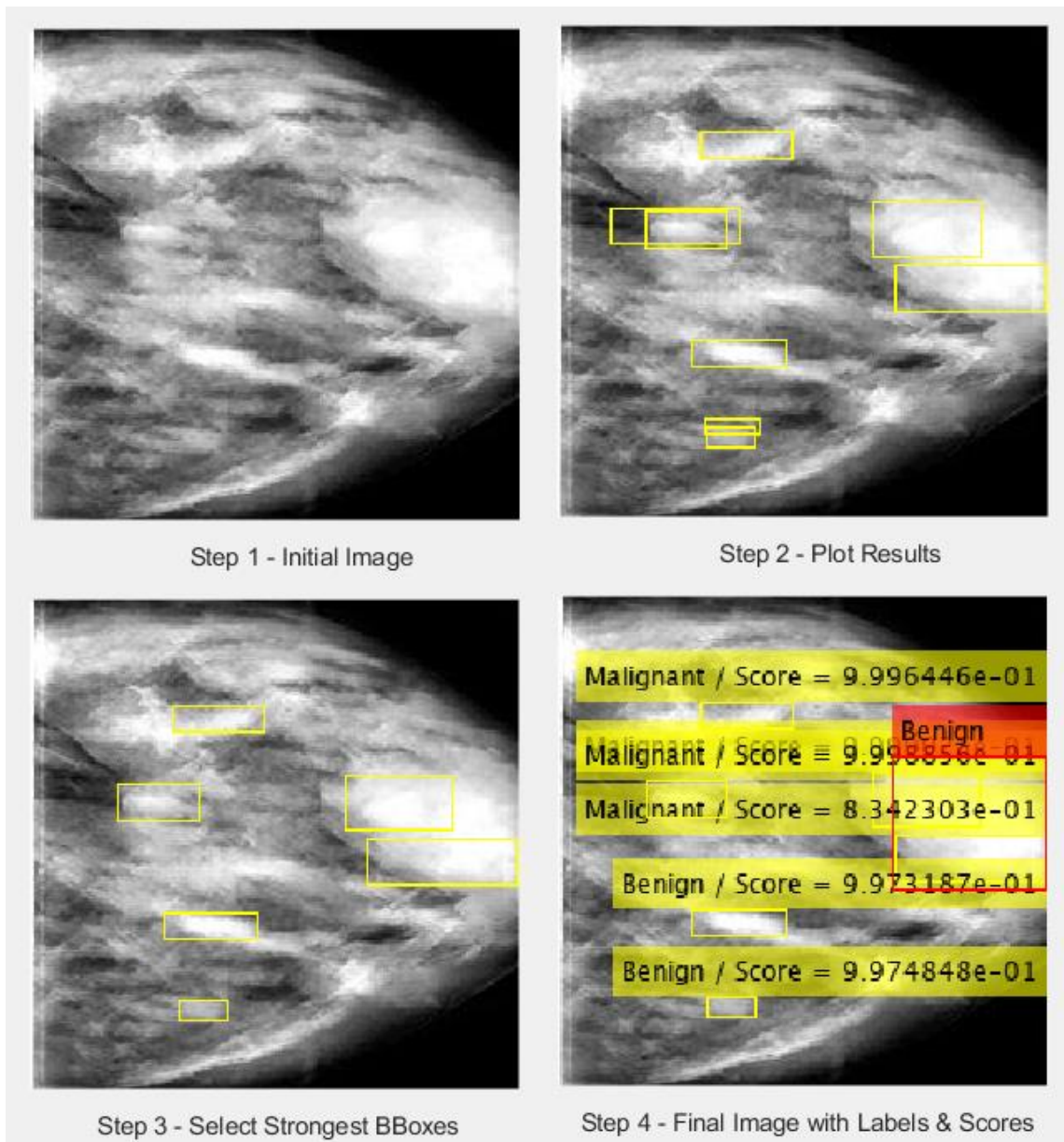
	Σετ εκπαίδευσης	Σετ ελέγχου	Σύνολο
Μαστογραφίες	583 (63%)	345 (37%)	928
Καλοήθης Όγκοι	316 (59%)	216 (41%)	532
Κακοήθης Όγκοι	317 (69%)	143 (31%)	460
Σύνολο Όγκων	633 (64%)	359 (36%)	992

Πίνακας 1 – Σύνθεση συλλογών σετ εκπαίδευσης/ελέγχου και αναλογία σε μορφολογία όγκων

3 Μέθοδοι Αξιολόγησης Ανιχνευτών

Για την αξιολόγηση των detectors που παρήχθησαν, τα δίκτυα τροφοδοτούνται με τις εικόνες του σετ ελέγχου και τα σχετικά αποτελέσματα (όρια περιοχής ενδιαφέροντος, κατηγορία και score) καταγράφονται. Τα αποτελέσματα της πρόβλεψης του δικτύου περιορίζονται από τον αλγόριθμο selectStrongestBbox ώστε σε περίπτωση που εντοπίζονται επικαλυπτόμενες περιοχές ενδιαφέροντος να διατηρηθεί μόνο αυτή με την μεγαλύτερη βεβαιότητα (score) όπως παρουσιάζεται και στην Πίνακας 1 – Σύνοψη συλλογών σετ εκπαίδευσης/ελέγχου και αναλογία σε μορφολογία όγκων. Στη συνέχεια τα αποτελέσματα συγκρίνονται με τον πίνακα αλήθειας χρησιμοποιώντας τις εντολές evaluateDetectionPrecision και evaluateDetectionMissRate οι οποίες και παράγουν έξι συγκριτικά μεγέθη που περιγράφουν την απόδοση του δικτύου. Τα μεγέθη αυτά είναι τα εξής:

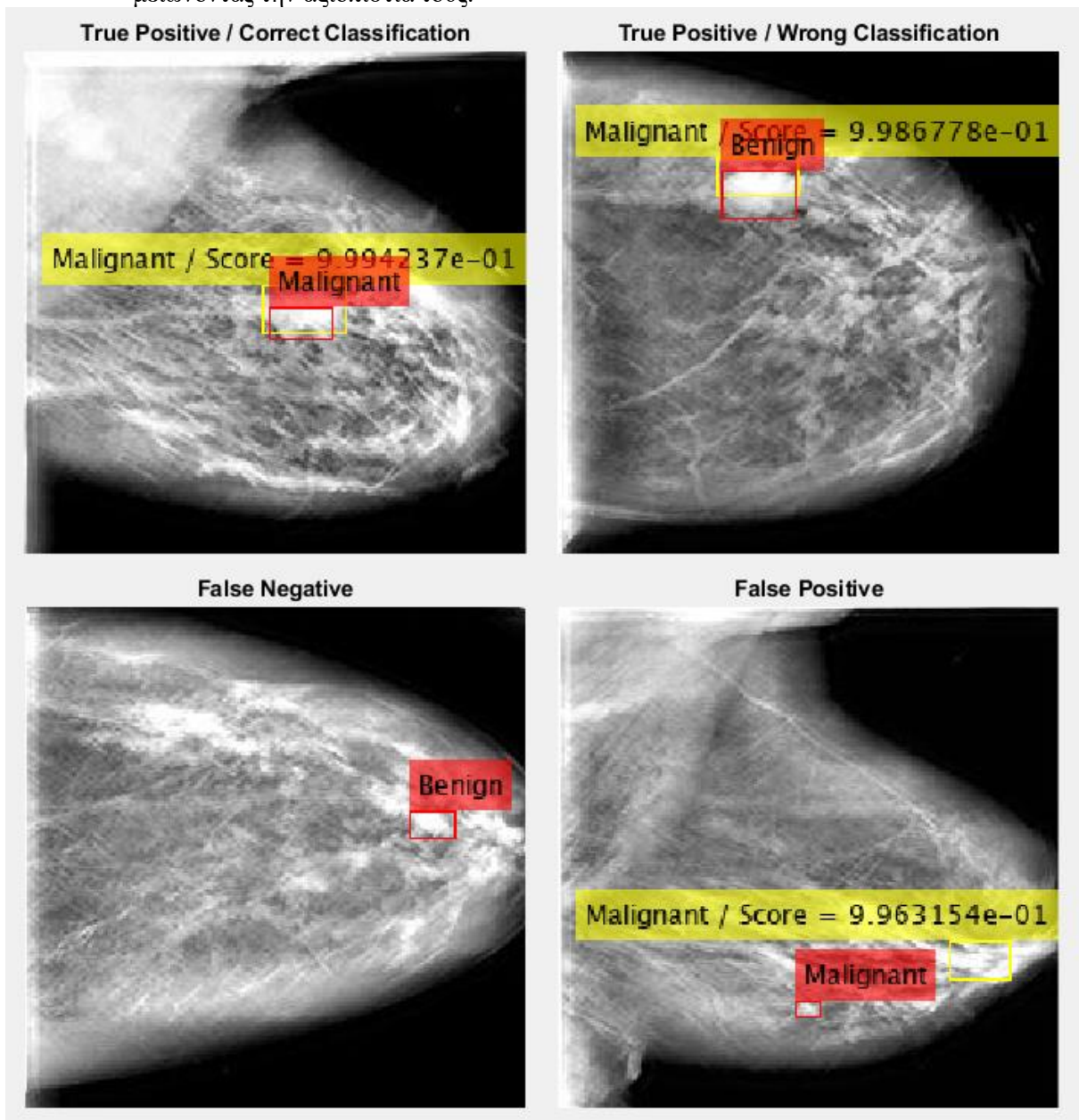
- Average Precision
- Recall
- Precision
- Average Miss Rate
- False Positives Per Image
- Miss Rate



Εικόνα 8 – Βήματα παραγωγής αποτελεσμάτων

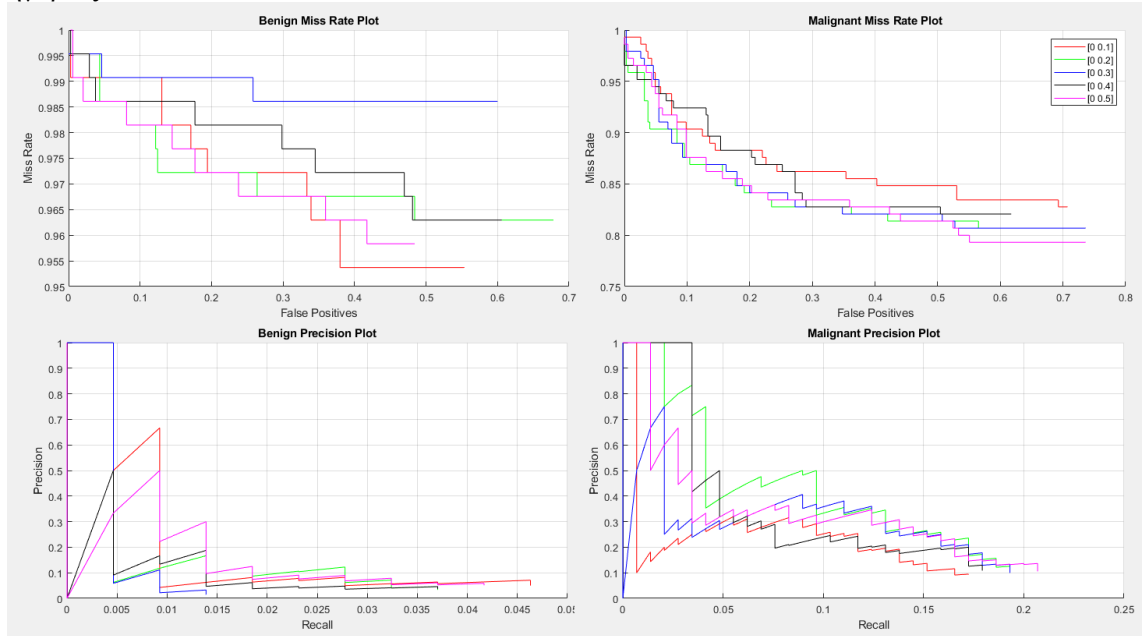
Οι βασικότερες τιμές που υποδεικνύουν σε απόλυτο βαθμό την ποιότητα των προβλέψεων είναι η μέση τιμή ακρίβειας (Average Precision) και ρυθμού λάθους (Average Miss Rate). Οι δύο αυτές τιμές προκύπτουν για καθένα από τα δύο διερευνόμενα μεγέθη, στην παρούσα εργασία, καλοήθη και κακοήθη καρκινώματα. Ο μέσος όρος των δύο αυτών τιμών μπορεί να χρησιμοποιηθεί για την κατάταξη του δικτύου. Με στόχο να γίνει εμφανέστερη η ποιοτική διαφορά μεταξύ δύο δικτύων, είναι εφικτός ο σχεδιασμός δύο καμπύλων.

- Αφενός το σχήμα Precision-Recall το οποίο αποτυπώνει την ακρίβεια (Precision) της πρόβλεψης $\left(\frac{\text{Αληθή Θετικά}}{\text{Αληθή Θετικά} + \text{Λάθος Θετικά}} \right)$ σε σχέση με την ευαισθησία (Recall) της $\left(\frac{\text{Αληθή Θετικά}}{\text{Αληθή Θετικά} + \text{Λάθος Αρνητικά}} \right)$.
- Αφετέρου αντίστοιχα συμπεράσματα ως προς τη δυναμική του δικτύου εξάγονται από διάγραμμα που παρουσιάζει τον αριθμό λάθος προβλέψεων (Miss Rate) συναρτήσει των λάθος θετικών ανά εικόνα (False Positives Per Image / FPPI). Το συγκεκριμένο σχήμα αναδεικνύει τα δίκτυα με τάση να αυξάνουν τις θετικές προβλέψεις πιθανώς εντοπίζοντας και τα πραγματικά αντικείμενα διατηρώντας έτσι σταθερή την ακρίβειά τους αλλά μειώνοντας την αξιοπιστία τους.



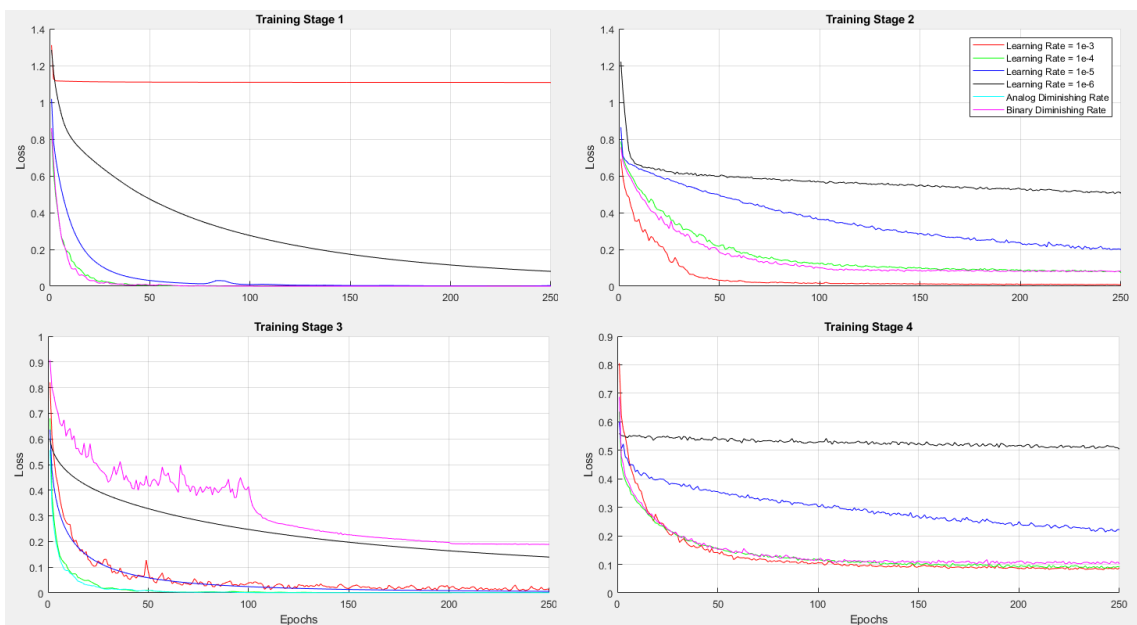
Εικόνα 9 – Κατηγοριοποίηση αποτελεσμάτων

Στην Εικόνα 9 εμφανίζονται μαστογραφίες από το σετ αξιολόγησης οι οποίες τροφοδοτήθηκαν σε εκπαιδευμένο δίκτυο και απεικονίζονται οι διαφορετικές κατηγορίες απόκρισης του συστήματος. Με κίτρινο χρώμα εμφανίζεται η πρόβλεψη του δικτύου ενώ με κόκκινο είναι σημειωμένη η θέση του πραγματικού όγκου. Προκύπτει η επιρροή που ασκεί η κατηγοριοποίηση του αντικειμένου στα μεγέθη αξιολόγησης. Ένα δίκτυο που εντοπίζει σωστά την περιοχή ενδιαφέροντος αλλά την κατηγοριοποιεί λανθασμένα αυξάνει τόσο τα Λάθος Θετικά αποτελέσματα, όσο και τα Λάθος Αρνητικά, καθώς εντόπισε λανθασμένα την λάθος κατηγορία και ταυτόχρονα δεν εντόπισε το αντικείμενο της σωστής κατηγορίας.



Εικόνα 10 – Ενδεικτικά σχήματα Miss Rate-FPPI και Precision-Recall

Τέλος συμπεράσματα σχετικά με την αποτελεσματικότητα της εκπαίδευσης του δικτύου εξάγονται και από τις καμπύλες της απώλειας κατά τα τέσσερα στάδια εκπαίδευσης. Στα σχήματα αυτά αποτυπώνονται σχέσεις όπως μη επαρκή εκπαίδευση δικτύων λόγω παραγόντων όπως μη επαρκής διάρκεια εκπαίδευσης, ή δίκτυα που έχουν αποκλίνει από το στόχο τους λόγω υψηλού ρυθμού μάθησης. Τα συμπεράσματα αυτά αξιοποιήθηκαν κυρίως στις διεργασίες που περιγράφονται στην ενότητα 5.3, από όπου προέρχονται και οι Εικόνα 10 και Εικόνα 11, ώστε να τυποποιηθούν οι παράμετροι εκπαίδευσης με σκοπό την κανονικοποίησης της απόδοσης των δικτύων.

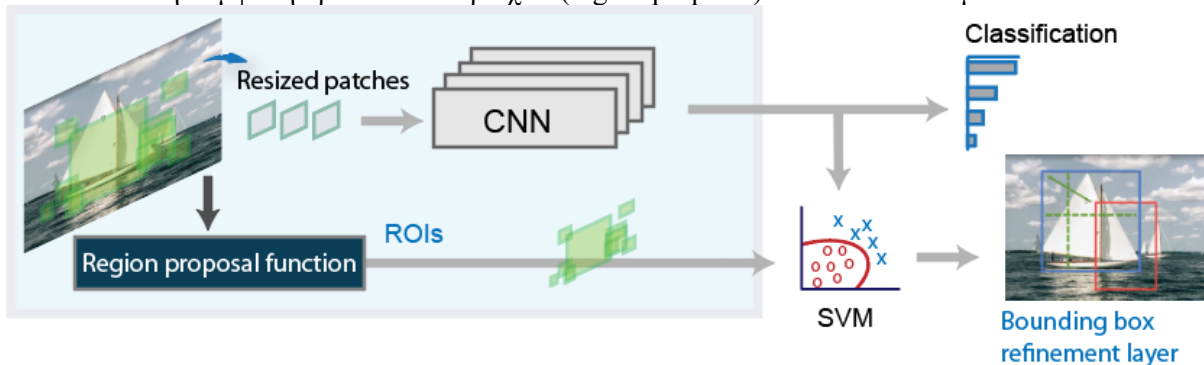


Εικόνα 11 – Ενδεικτικά σχήματα απώλειας εκπαίδευσης

4 Αρχιτεκτονική Βαθιών Νευρωνικών Δικτύων

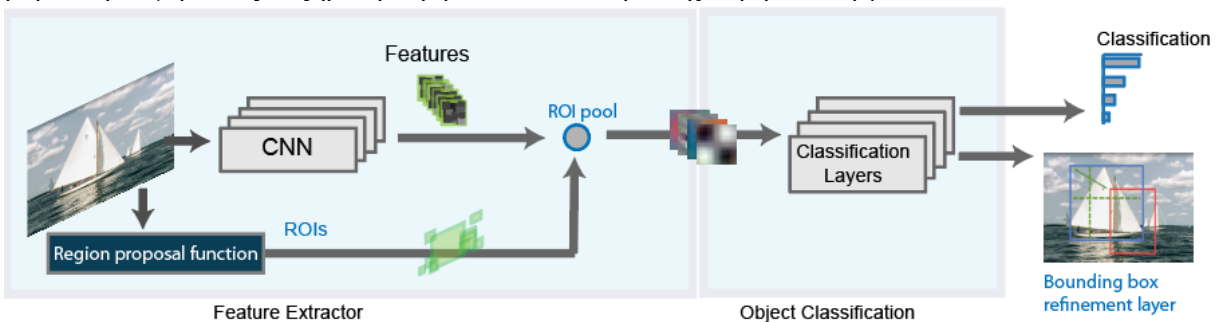
4.1 Εισαγωγή στην οικογένεια R-CNN

Το 2013 οι Girshick et al. [35] ανέλυσαν το πρόβλημα του αυτόματου εντοπισμού και κατηγοριοποίησης αντικειμένων σε εικόνα στα δύο συνθετικά του. Βασιζόμενοι στα ενθαρρυντικά αποτελέσματα του ImageNet στους πρόσφατους διαγωνισμούς κατέληξαν στο συμπέρασμα πως η έφεση των CNN στην κατηγοριοποίηση κανονικοποιημένων εικόνων [22] μπορεί να αξιοποιηθεί σε συνδυασμό με έναν αλγόριθμο ο οποίος θα αναλύει μια εικόνα σε προτάσεις περιοχών (region proposal) οι οποίες περιέχουν τα αντικείμενα ενδιαφέροντος και οι οποίες θα κατηγοριοποιούνται στη συνέχεια από το CNN δίκτυο. Προτείνουν διάφορες μεθόδους παραγωγής περιοχών προτάσεων, αγνωστικές (agnostic) ως προς την κατηγορία του αντικειμένου, με χρήση ενός μοντέλου γραμμικής οπισθοδρόμησης (linear regression model) για την ισχυροποίηση των προτάσεων. Το δίκτυο που εκπαίδευσαν έδειξε μέσο όρο ακρίβειας (mean average precision mAP) 31,4% στον διαγωνισμό ILSVRC 2013 υπερβαίνοντας το δεύτερο καλύτερο δίκτυο κατά 7 μονάδες. Καθώς το δίκτυό τους αποτελούσε τη σύμφυση προτάσεων περιοχών (region proposal) και CNN το ονόμασαν R-CNN.



Εικόνα 12 – Δομή RCNN [58]

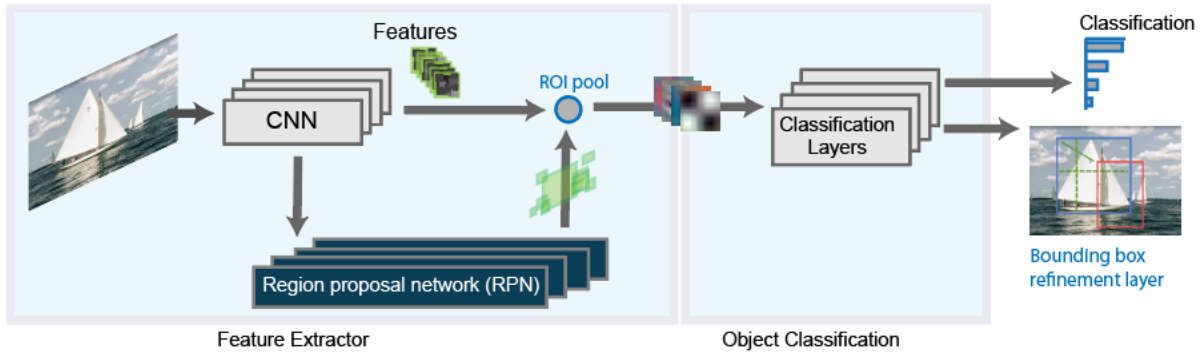
Το 2015 ο ίδιος στο [36] δημοσιεύει την επέκταση της προηγούμενης εργασίας του, όπου παρατηρεί πως το RCNN δίκτυο έχει κάποιους φυσικούς περιορισμούς. Καθώς κάθε περιοχή που προκύπτει από τον αλγόριθμο προτάσεων τροφοδοτείται μέσα από SVM και αξιολογείται ξεχωριστά από το CNN υπάρχουν απώλειες τόσο σε χρόνο εκπαίδευσης και ανίχνευσης όσο και σε ακρίβεια. Το νέο σύστημα επεξεργάζεται την αρχική εικόνα, αντί να αποκόπτει και να μεγεθύνει κάθε περιοχή πρότασης ξεχωριστά, αντίθετα αξιολογεί τις περιοχές πρότασης με τα χαρακτηριστικά που προκύπτουν από το CNN. Επιπρόσθετα οι ερευνητές αξιοποιούν τον αλγόριθμο Selective Search [37] σε αντίθεση με τον EdgeBoxes [38] που αξιοποιούσε η πρώτη υλοποίηση του RCNN. Το νέο δίκτυο Fast R-CNN συνοδεύεται από ρυθμίσεις στις μεθόδους υπολογισμού Απώλειας (Loss) και ανατροφοδότησής των βαρών προσφέροντας αυξημένη ακρίβεια και ιδιαίτερα ταχύτερη λειτουργία.



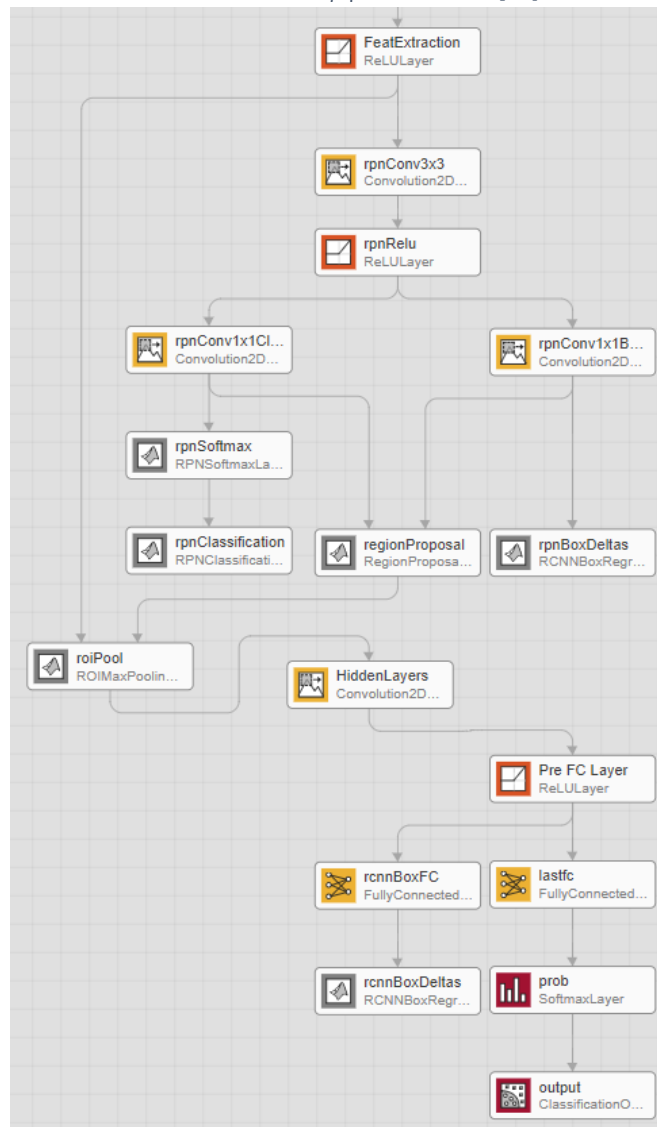
Εικόνα 13 – Δομή Fast RCNN [58]

Η επόμενη πρόοδος στην λογική της συγκεκριμένης αρχιτεκτονικής επήλθε το 2016 όπου στο [39] περιγράφεται η πλήρης αντικατάσταση του αλγόριθμου πρότασης περιοχών του Fast R-CNN δικτύου από ένα ξεχωριστό, πλήρως συνδεδεμένο, CNN. Το νέο δίκτυο πρότασης περιοχών (region proposal network - RPN) εντοπίζει τις περιοχές που περιέχουν τα αντικείμενα οι οποίες στη συνέχεια κατηγοριοποιούνται από το υπόλοιπο δίκτυο. Η εκπαίδευση του δικτύου γίνεται τμηματικά σε τέσσερα στάδια. Στην αρχή εκπαίδευεται το RPN, στο δεύτερο στάδιο με τα ευρήματά του και την περιοχή

αλήθειας που τροφοδοτείται εκπαιδεύεται το CNN. Στο τρίτο στάδιο πραγματοποιείται η πρώτη ενοποίηση των δικτύων όπου τα βάρη του CNN δικτύου χρησιμεύουν στην εκπαίδευση του RPN και τέλος, στο τέταρτο στάδιο εκπαιδεύονται τα στρώματα του CNN, ενώ παραμένουν συνδεδεμένα με το RPN. Η υλοποίηση αυτή εμφανίζει μεγαλύτερη ακρίβεια και υψηλότερη ταχύτητα λειτουργίας από τις προηγούμενες ενσαρκώσεις RCNN, ανοίγοντας τη δυνατότητα λειτουργίας ταυτόχρονης με τη πραγματικότητα (real time detection) σε αναφερόμενη συχνότητα 5 fps.



Εικόνα 14 – Δομή Faster RCNN [58]



Εικόνα 15 – Στρώματα που απαρτίζουν το Faster RCNN σε υλοποίηση Matlab

Η τελευταία προσθήκη στο γενεαλογικό δέντρο των RCNN ήρθε το 2017 με την δημοσίευση του mask RCNN [40] όπου αναλύεται η δυνατότητα προσθήκης ενός μικρού δικτύου πλήρους συνέλιξης (FCN) [41] το οποίο προβλέπει την κατάτμηση του αντικειμένου από την περιοχή ενδιαφέροντος στη

μονάδα του εικονοστοιχείου. Προσθέτοντας στον υπολογισμό της απώλειας και την διαφορά average binary cross-entropy loss καταφέρνει, χωρίς να θυσιάσει την ταχύτητα επεξεργασίας της RCNN αρχιτεκτονικής, να ξεπεράσει σε απόδοση άλλα δίκτυα προσδιορισμού και κατηγοριοποίησης όπως το FCIS [42], το οποίο επιπρόσθετα εμφανίζει αδυναμίες σε περιοχές που επικαλύπτονται από πολλαπλά αντικείμενα ενδιαφέροντος. Η συγκεκριμένη υλοποίηση δεν απασχολεί τη παρούσα έρευνα λόγω της ασαφούς φύσης των δεδομένων που επεξεργαζόμαστε σε συνάρτηση με την στόχευσή για εντοπισμό και κατηγοριοποίηση μικρών και αποστασιοποιημένων μεταξύ τους αντικειμένων και όχι συνεκτικών επιφανειών όπως η ανθρωπόμορφη ή/και συνεκτική επιφάνεια.

Στο περιβάλλον του MATLAB, από την έκδοση r2016b και εφεξής, η υλοποίηση και εκπαίδευση των RCNN δικτύων, με την εξαίρεση του mask RCNN του οποίου η υλοποίηση δεν έχει ενσωματωθεί ακόμα, υποστηρίζεται επίσημα με την χρήση των εντολών trainRCNNObjectDetector, trainFastRCNNObjectDetector και trainFasterRCNNObjectDetector. Κάθε εντολή δέχεται σαν ορίσματα τον πίνακα αληθείας των δεδομένων μας, τα στρώματα του δικτύου που θα υλοποιήσουμε και παραμέτρους εκπαίδευσης, ενώ παρουσιάζει σαν έξοδο ένα αντικείμενο detector εκπαιδευμένο με την αντίστοιχη μέθοδο και, εφόσον το επιλέξουμε ένα αρχείο με τις παραμέτρους της εκπαίδευσης. Από προεπιλογή η εκπαίδευση των δικτύων έχει τμήματα κώδικα που εκτελούνται μόνο στην CPU, όπως η εξαγωγή των προτάσεων περιοχών, ενώ το μεγαλύτερο μέρος της πραγματοποιείται στην κάρτα γραφικών, εφόσον υπάρχει διαθέσιμη μια κάρτα γραφικών με ικανότητα υπολογισμών CUDA 3.0+.

4.2 Βασικές Αρχιτεκτονικές Δομές

4.2.1 AlexNet

Οι Alex Krizhevsky et al. [22] αναλύουν την πρότασή τους για ένα βαθύ νευρωνικό δίκτυο το οποίο, ενσωματώνοντας τις τεχνολογικές προόδους της εποχής, καταφέρνει να αποσπάσει το βραβείο ImageNet Large Scale Visual Recognition Challenge (ILSVRC) του έτους 2012. Η πρότασή τους



Εικόνα 16 – Αρχική δομή AlexNet

περιείχε πέντε επίπεδα συνέλιξης (τα τρία εκ των οποίων ακολουθούμενα από στρώματα κανονικοποίησης), τρία επίπεδα συγκέντρωσης και τρία πλήρως συνδεδεμένα στρώματα που συνδέονταν με τα στρώματα κατηγοριοποίησης των αποτελεσμάτων. Το δίκτυο ονομάστηκε από το όνομα του δημιουργού του alexnet.

Εκτός από την πρακτική χρησιμότητα του δικτύου, το οποίο έχει εκπαιδευτεί σε 1,2 εκατομμύρια εικόνες της συλλογής ImageNet, η μελέτη που δημοσιεύτηκε περιέχει ιδιαίτερα ενδιαφέροντα στοιχεία, όπως την αναφορά πως η αφαίρεση ενός μόνο στρώματος συνέλιξης (που κατά τα άλλα περιέχει μόνο το ~1 τοις εκατό από τις συνολικά 6 εκατομμύρια παραμέτρους του δικτύου) μείωνε αισθητά την απόδοση. Αντίστοιχα, πραγματοποιείται τη χρήση των, πρόσφατα ανεπτυγμένων τη περίοδο εκείνη, στρωμάτων απόρριψης (dropout layer) στα δύο πρώτα πλήρως συνδεδεμένα στρώματα ώστε να αποφευχθεί η ανάπτυξη εξαιρετικά ισχυρών χαρακτηριστικών κατά την διάρκεια της εκπαίδευσης.



Εικόνα 17 – Δομή AlexNet με την προσθήκη επιπλέον Fully Connected στρώματος

Το alexnet είναι διαθέσιμο συμπληρωματικά στο MATLAB αλλά η υλοποίηση που περιλαμβάνεται διαθέτει είσοδο εικόνων με διαστάσεις [227 227 3] σε αντίθεση με τη διάσταση 224 που περιγράφεται στην δημοσίευση. Καθώς από το 102.2.2 έχει γίνει χρήση grayscale εικόνων 224 pixels σε αντίθεση με της χρωματικές των τριών καναλιών, εξυπηρετεί η χρήση των συγκεκριμένων διαστάσεων, που έχουν επιλέγει ως διαστάσεις των εικόνων της βάσης δεδομένων, ταυτόχρονα με την αφαίρεση των δύο

Πτυχιακή Εργασία, τμήμα Μηχανικών Πληροφορικής
75. Αναγνώριση καρκινώματος / Σύνταλας Δημήτριος 2051

πλεοναζόντων χρωματικών καναλιών. Με αυτό το σκοπό, πέρα από την δημιουργία του επιπέδου εισόδου με τις διαστάσεις που εξυπηρετούν, [224 224 1], είναι απαραίτητη η μεταβολή και του πρώτου στρώματος συνέλιξης του δικτύου ώστε ο αριθμός των στοιχείων στην είσοδο και την έξοδό του να ταυτίζεται με αυτόν στα γειτονικά του στρώματα.

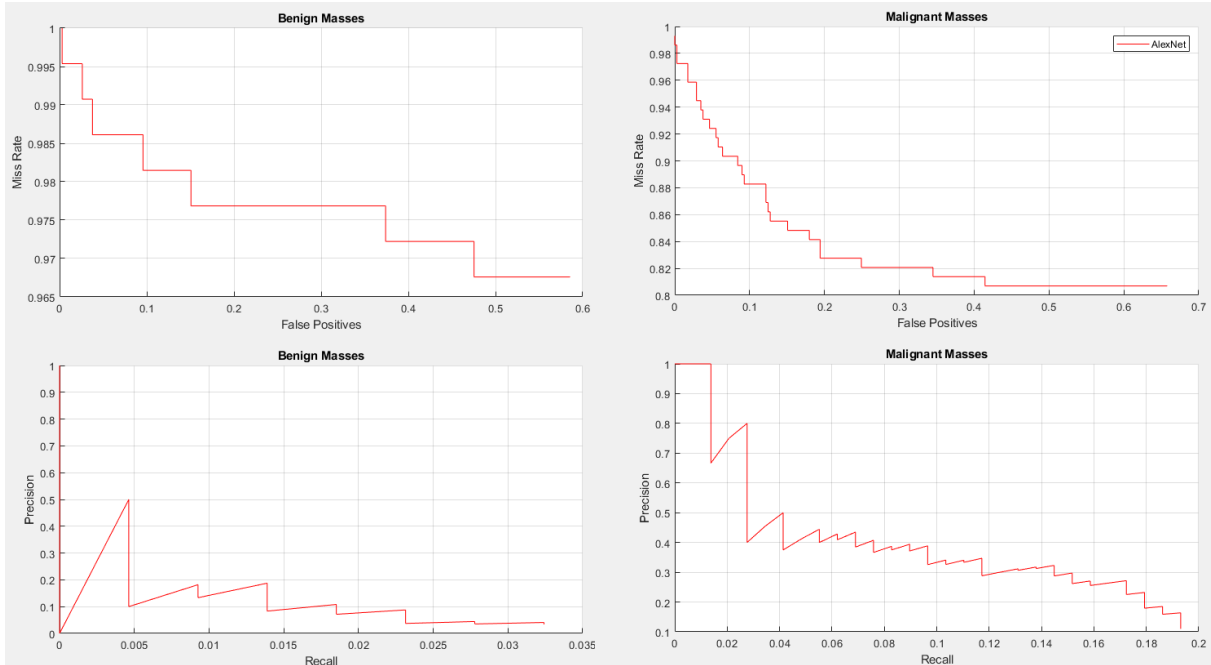
ANALYSIS RESULT				
	NAME	TYPE	ACTIVATIONS	LEARNABLES
1	imageinput 224x224x1 images with 'zero-center' normalization	Image Input	224x224x1	-
2	conv1 96 11x11x1 convolutions with stride [4 4] and padding [3 3 3 3]	Convolution	55x55x96	Weights 11x11x1x96 Bias 1x1x96
3	relu1 ReLU	ReLU	55x55x96	-
4	norm1 cross channel normalization with 5 channels per element	Cross Channel Nor...	55x55x96	-
5	pool1 3x3 max pooling with stride [2 2] and padding [0 0 0 0]	Max Pooling	27x27x96	-
6	conv2 256 5x5x48 convolutions with stride [1 1] and padding [2 2 2 2]	Convolution	27x27x256	Weights 5x5x48x256 Bias 1x1x256
7	relu2 ReLU	ReLU	27x27x256	-
8	norm2 cross channel normalization with 5 channels per element	Cross Channel Nor...	27x27x256	-
9	pool2 3x3 max pooling with stride [2 2] and padding [0 0 0 0]	Max Pooling	13x13x256	-
10	conv3 384 3x3x256 convolutions with stride [1 1] and padding [1 1 1 1]	Convolution	13x13x384	Weights 3x3x256x384 Bias 1x1x384
11	relu3 ReLU	ReLU	13x13x384	-
12	conv4 384 3x3x192 convolutions with stride [1 1] and padding [1 1 1 1]	Convolution	13x13x384	Weights 3x3x192x384 Bias 1x1x384
13	relu4 ReLU	ReLU	13x13x384	-
14	conv5 256 3x3x192 convolutions with stride [1 1] and padding [1 1 1 1]	Convolution	13x13x256	Weights 3x3x192x256 Bias 1x1x256
15	relu5 ReLU	ReLU	13x13x256	-
16	pool5 3x3 max pooling with stride [2 2] and padding [0 0 0 0]	Max Pooling	6x6x256	-
17	fc6 4096 fully connected layer	Fully Connected	1x1x4096	Weights 4096x9216 Bias 4096x1
18	relu6 ReLU	ReLU	1x1x4096	-
19	drop6 50% dropout	Dropout	1x1x4096	-
20	fc7 4096 fully connected layer	Fully Connected	1x1x4096	Weights 4096x4096 Bias 4096x1
21	relu7 ReLU	ReLU	1x1x4096	-
22	drop7 50% dropout	Dropout	1x1x4096	-
23	fc8 1000 fully connected layer	Fully Connected	1x1x1000	Weights 1000x4096 Bias 1000x1
24	relu ReLU	ReLU	1x1x1000	-
25	dropout 50% dropout	Dropout	1x1x1000	-
26	lastfc 3 fully connected layer	Fully Connected	1x1x3	Weights 3x1000 Bias 3x1
27	prob softmax	Softmax	1x1x3	-
28	output crossentropyx	Classification Output	-	-

Εικόνα 18 – Ανάλυση ροής δικτύου AlexNet που έχει υποστεί επεξεργασία

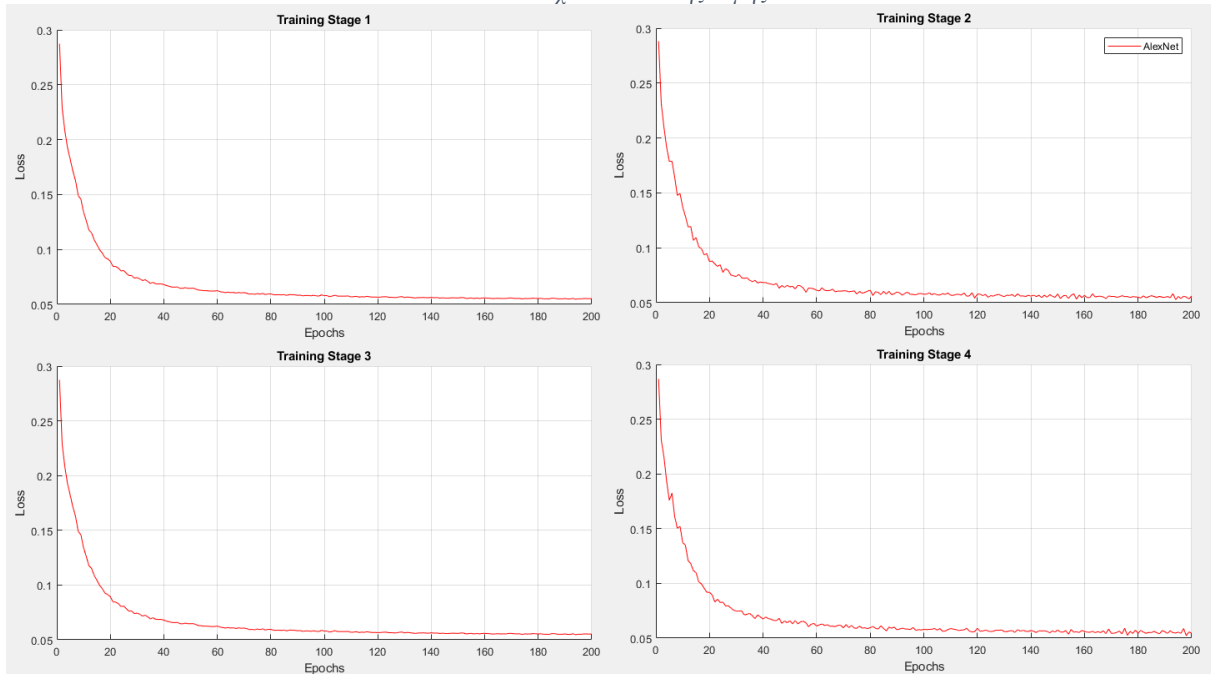
Η διαφορά που προκύπτει από τη μείωση της εικόνας εισόδου κατά τρία εικονοστοιχεία αντιμετωπίζεται εύκολα με την εισαγωγή ισόποσης επένδυσης (padding) και η σύμπτυξη τριών καναλιών επιτυγχάνεται με το άθροισμα των βαρών και των τριών καναλιών, στοιχείο προς στοιχείο, ώστε, με διαίρεση προς τον αριθμό τους, να προκύψει ο μέσος όρος ο οποίος και θα χρησιμοποιηθεί σαν βάρος του καινούριου στρώματος. Με αυτόν τον τρόπο καθίσταται εφικτή η μέγιστη μεταφορά γνώσης προς το καινούριο δίκτυο.

Πτυχιακή Εργασία, τμήμα Μηχανικών Πληροφορικής
75. Αναγνώριση καρκινώματος / Συνάπαλος Δημήτριος 2051

Τέλος, το παρόν δίκτυο, όπως και οι δομές που θα αναφερθούν στη συνέχεια, κατηγοριοποιούν τα αποτελέσματα που θα προκύψουν από τις προτάσεις περιοχών σε δύο κατηγορίες, καλοήθης ή κακοήθης, συνυπολογίζοντας μια επιπλέον κατηγορία που θα αναφέρεται στο υπόβαθρο, σε περιοχές δηλαδή που δεν θα αξιολογηθούν θετικά όσον αφορά το αν εμπεριέχουν αντικείμενο ενδιαφέροντος. Προστίθεται για αυτό τον λόγο μία επιπλέον ακολουθία από τα στρώματα Relu, Dropout και ένα επιπλέον πλήρως συνδεδεμένο στρώμα με είσοδο τους χίλιους νευρώνες του προηγούμενου στρώματος και έξοδο τις τρεις περιπτώσεις κατηγοριών.



Εικόνα 19 – Σχέδιο απόδοσης δομής AlexNet



Εικόνα 20 – Απόδοση εκπαίδευσης δομής AlexNet

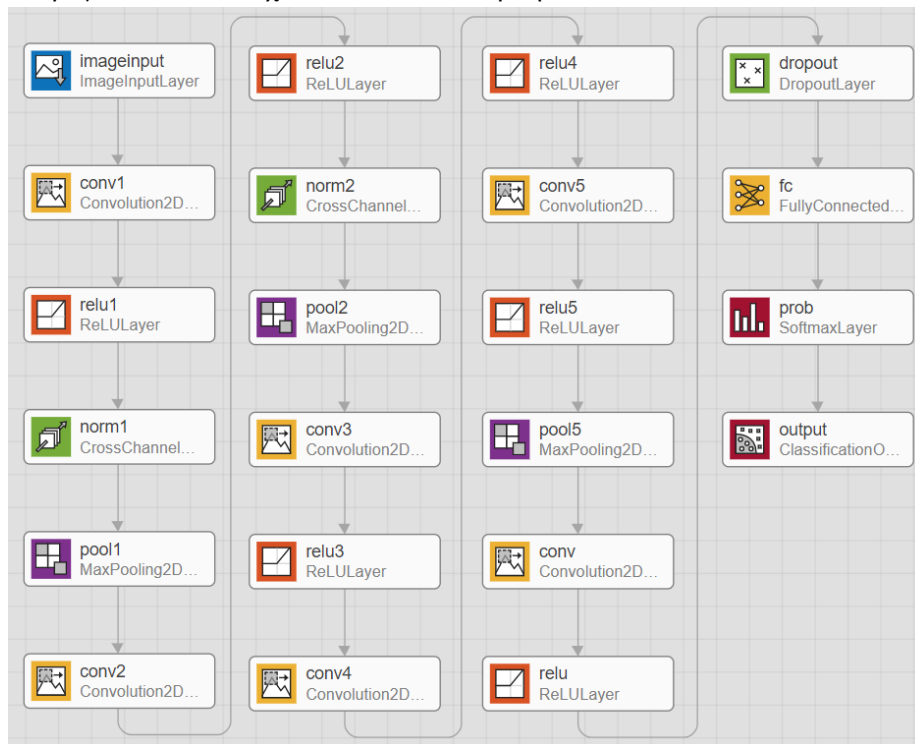
Network Base	Benign Average Precision	Malignant Average Precision	Benign Average Miss Rate	Malignant Average Miss Rate	Ποσοστό εικόνων με εύρεση ROI
AlexNet	0.0053	0.0811	0.9819	0.8821	60.29%

Πίνακας 2 – Απόδοση δομής AlexNet

4.2.2 R-FCN

Οι He et al. [43] εμβαθύνουν στον διαχωρισμό της οικογένειας των RCNN στα συνθετικά τους δίκτυα. Παρατηρούν πως στην περίπτωση του Faster RCNN το RPN τμήμα του δικτύου είναι ένα αυτόνομο δίκτυο που αποτελείται εξ ολοκλήρου από συνελίξεις και συνδέεται μέσω ενός στρώματος συγκέντρωσης με ένα δεύτερο πλήρως συνδεδεμένο δίκτυο το οποίο κατηγοριοποιεί τις προτάσεις που προκύπτουν από το RPN .

Η πρότασή τους δεν προκύπτει εκ του μηδενός. Ήδη το 2015 οι Long et al. [41] έχουν εντοπίσει την αδυναμία των πλήρως συνδεδεμένων δικτύων να γενικεύσουν την πληροφορία που λαμβάνουν με αποτέλεσμα η εκπαίδευσή τους να γίνεται σε περιορισμένο χωρικά επίπεδο. Προτείνουν ένα δίκτυο πλήρους συνέλιξης για την απόδοση χαρακτηριστικών στοιχείο προς στοιχείο (εργασία αντίστοιχη με αυτή που πραγματοποιεί το Mask RCNN που έχουμε ήδη αναλύσει) αξιοποιώντας τη μέθοδο ταχείας ανάγνωσης (Fast Scanning) των Giusti et al. [44] ώστε να αντιστοιχηθεί η τιμή των αρχικών (raw) πινάκων που παράγονται στα στοιχεία από τα οποία προήλθαν.



Εικόνα 21 – Δομή FCN AlexNet

Εντοπίζουν ότι τα πλήρως συνδεδεμένα στρώματα εγκλωβίζουν χωρικά την γνώση που προκύπτει και οδηγούν σε μείωση ακρίβειας και ανάγκη χρήσης τεχνασμάτων για την βελτίωσή της. Γίνεται αναφορά σε δίκτυα όπως το GoogLeNet τα οποία χρησιμοποιούν, από κατασκευής, δομή πλήρους συνέλιξης, αξιοποιώντας πλήρως συνδεδεμένα στρώματα μόνο στα τελευταία επίπεδα της κατηγοριοποίησης, σε αντίθεση με τα RCNN, τα οποία αποκαλούν Δίκτυα Μερικής Συνέλιξης (Semi Convolutional Network).

Το δίκτυο που προτείνουν συνδυάζει το RPN που προαναφέρθηκε με ένα δίκτυο πλήρους συνέλιξης το οποίο αναλύει τις περιοχές που προτείνονται και τις αθροίζει σε ένα τελικό στρώμα συγκέντρωσης το οποίο και αποφαινεται για την κατηγορία και τη βαθμολογία. Το δίκτυό τους, υλοποιημένο σε σκελετό του δικτύου ResNet [45] υπερτερεί απέναντι σε ένα Faster RCNN βασισμένο στην ίδια αρχιτεκτονική και ονομάζεται R-FCN. Στα δίκτυα ResNet γίνεται αναφορά στη συνέχεια του κεφαλαίου και η δομή τους χρησιμοποιείται σαν σκελετός για την υλοποίηση του detector.

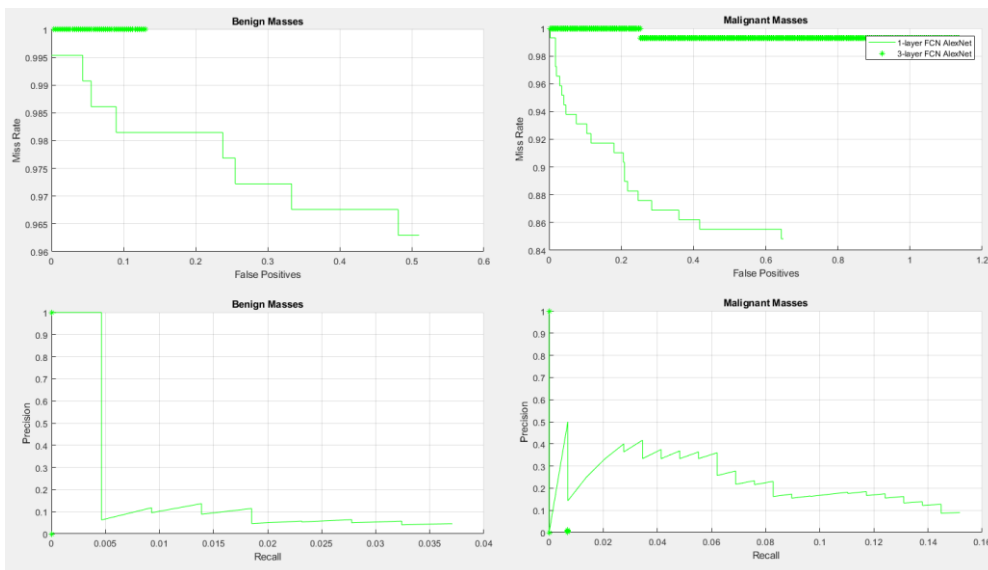
Σε αυτά τα πλαίσια αντικαθίστανται τα στρώματα 17-23, του fully connected τμήματος του βασικού δικτύου AlexNet με ένα στρώμα συνέλιξης 192 φίλτρων 3X3 με stride [1, 1] και padding [1, 1, 1, 1] ώστε να προκύψει ένα δίκτυο πλήρους συνέλιξης κατά τα πρότυπα του R-FCN. Παρατηρώντας τα αποτελέσματά του στις Εικόνα 22 και Εικόνα 23 και Πίνακας 3 είναι φανερό πως η αφαίρεση των πλήρως συνδεδεμένων στρωμάτων δεν βελτιώνει τα αποτελέσματά του δικτύου.

Πτυχιακή Εργασία, τμήμα Μηχανικών Πληροφορικής
75. Αναγνώριση καρκινώματος / Συνάπαλος Δημήτριος 2051

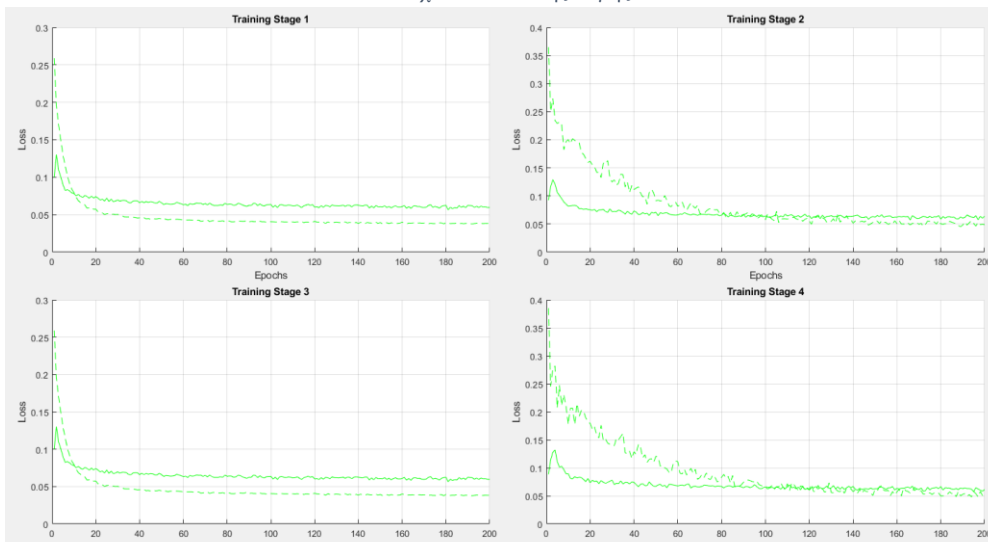
Χρησιμοποιώντας 3 στρώματα συνέλιξης (με κλιμακούμενο αριθμό 64, 32 και 16 φίλτρων 3X3) και τα συνοδευτικά τους στρώματα τα αποτελέσματα που παρουσιάζονται είναι απογοητευτικά. Καταλήγοντας, τα fully connected layers στην απόληξη της δομής είναι βοηθητικά και χρήσιμα για τις ανάγκες του δικτύου και η ουσία της πρότασης του R-FCN βρίσκεται στην αυτούσια υλοποίησή του και όχι σε μεταφορά του σε faster RCNN πρότυπα.

Network Base	Benign Average Precision	Malignant Average Precision	Benign Average Miss Rate	Malignant Average Miss Rate	Ποσοστό εικόνων με εύρεση ROI
1-layer FCN AlexNet	0.0074	0.0392	0.9814	0.9190	43.19%
3-layer FCN AlexNet	0.0000	0.0001	1.0000	0.9977	7.83%

Πίνακας 3 – Απόδοση δομής FCN AlexNet



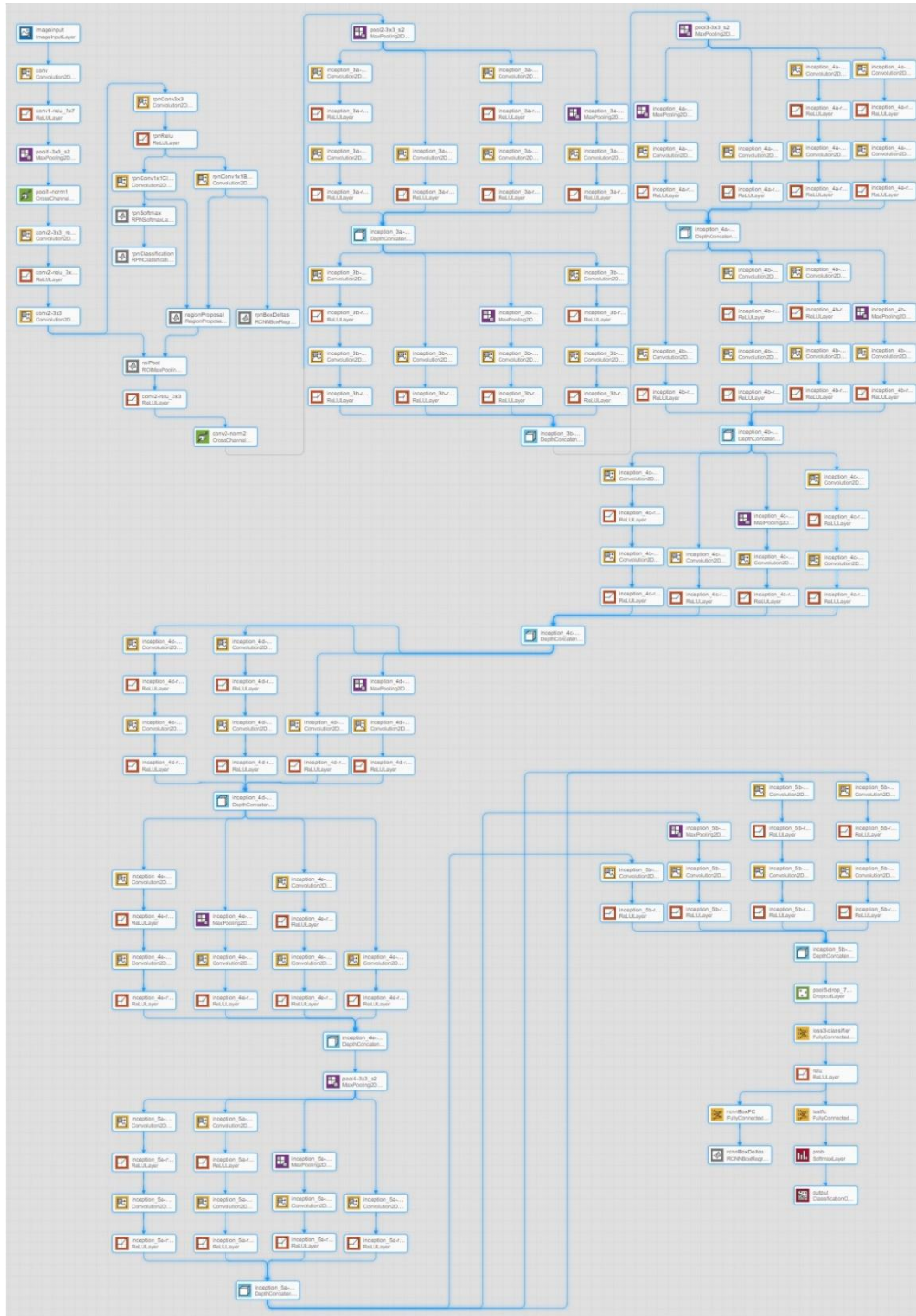
Εικόνα 22 – Σχέδιο απόδοσης δομής FCN AlexNet



Εικόνα 23 – Απόλεια εκπαίδευσης δομής FCN AlexNet

4.2.3 GoogLeNet

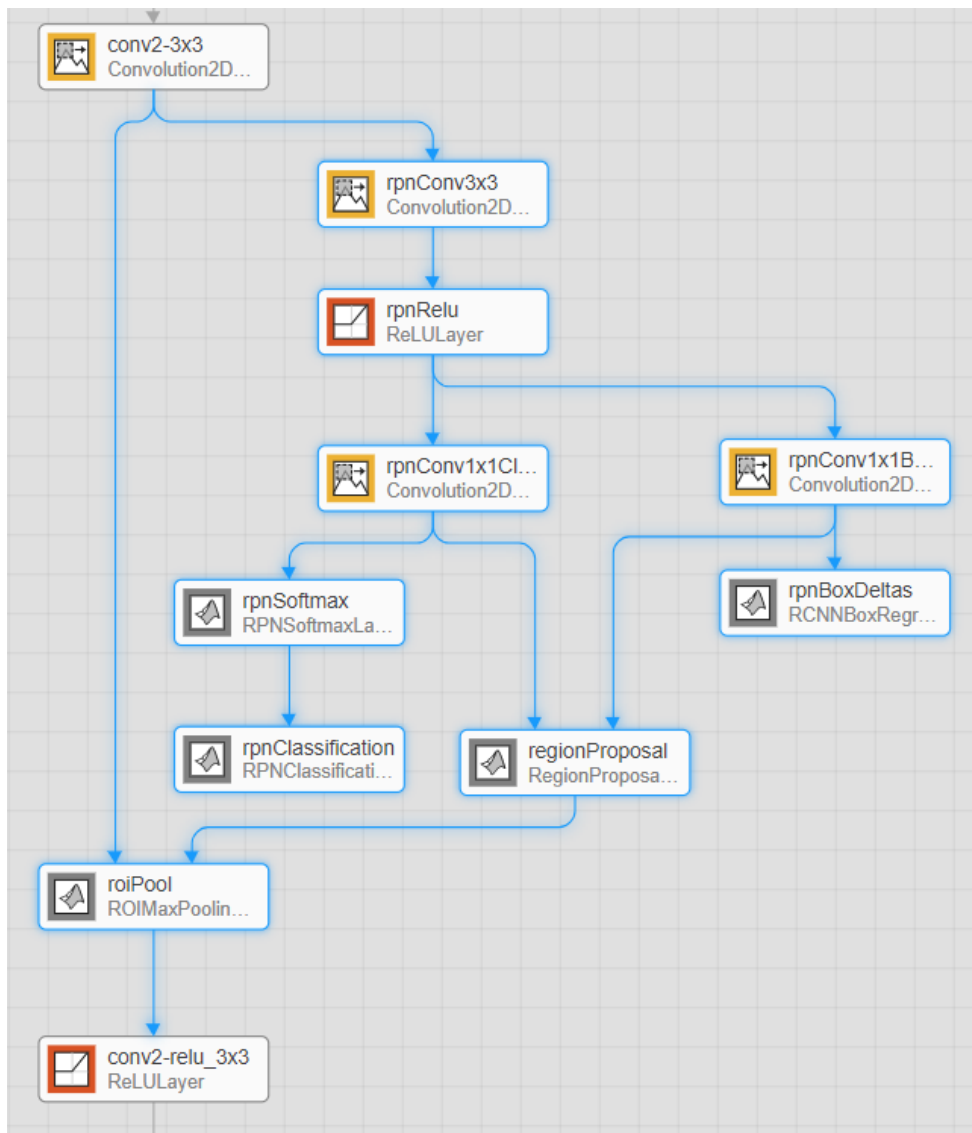
Το GoogLeNet αναπτύχθηκε από τους Szegedy et al. [46] στα πλαίσια του διαγωνισμού ILSVRC 2014 στον οποίο και κατέλαβε την πρώτη θέση. Η ομάδα της εργασίας ανέπτυξε τη συγκεκριμένη υλοποίηση στην προσπάθειά της να δημιουργήσει ένα βαθύτερο δίκτυο που θα έχει τη δυνατότητα να εκπαιδευτεί σε ακόμη μεγαλύτερες βάσεις δεδομένων χωρίς να υπερκαλύπτεται (overfitting). Σε αντίθεση με τα σύγχρονά του δίκτυα, τα οποία στοιβάζουν στρώματα συνέλιξης, ενεργοποίησης και συγκέντρωσης με διάταξη σε σειρά, το GoogLeNet δημιουργεί δυσδιάστατες δομές οι οποίες επεξεργάζονται ταυτόχρονα συνελίξεις οι οποίες αθροίζονται στην συνέχεια σε μία έξοδο η οποία τροφοδοτεί το επόμενο στρώμα.



Εικόνα 24 – Δομή GoogLeNet

Η τεχνική αυτή βασίστηκε στις παρατηρήσεις των Agora et al. [47] πως, υποθετικά, μία αραιά συνδεδεμένη δομή δικτύου (μια δομή που τείνει να μιμηθεί την φυσική δομή του ανθρώπινου εγκεφάλου) θα υπερβαίνει τις πυκνά και πλήρως συνδεδεμένες δομές που αξιοποιούνταν κατά κόρον. Αξίζει να αναφερθούμε στο γεγονός πως η υλοποίηση που πρότειναν για τον διαγωνισμό πλησιάζει τα δίκτυα πλήρους συνέλιξης (Fully Convolutional Network - FCN) καθώς διαθέτει πλήρως συνδεδεμένο δίκτυο μόνο στην έξοδό του ώστε να εξυπηρετήσει την κατηγοριοποίηση των αποτελεσμάτων.

Η τεχνική που χρησιμοποίησαν για την εμβάθυνση του δικτύου ονομάστηκε Inception (σημείο εκκίνησης) και αποτελείται από παράλληλα στρώματα μικρών φίλτρων συνέλιξεων (1X1) με σκοπό τη μείωση των διαστάσεων, διατηρώντας χρήσιμη πληροφορία η οποία τροφοδοτείται σε, υπολογιστικά ασύμφορα, 3X3 και 5X5 φίλτρα. Βασιζόμενοι στη θέση πως τα, σε σειρά στοιβαγμένα, δυσδιάστατα ανώτερα στρώματα θα αποκτούν γνώση χαρακτηριστικών όλο και υψηλότερης αφαίρεσης διατηρούν τα χαμηλότερα στρώματα στο συνηθισμένο μοτίβο συνέλιξης χρησιμοποιώντας περιστασιακά στρώματα συγκέντρωσης ώστε να υποδιπλασιάσουν το κλιμακούμενο όγκο του δικτύου.



Εικόνα 25 – Το RPN τμήμα της υλοποίησης GoogLeNet

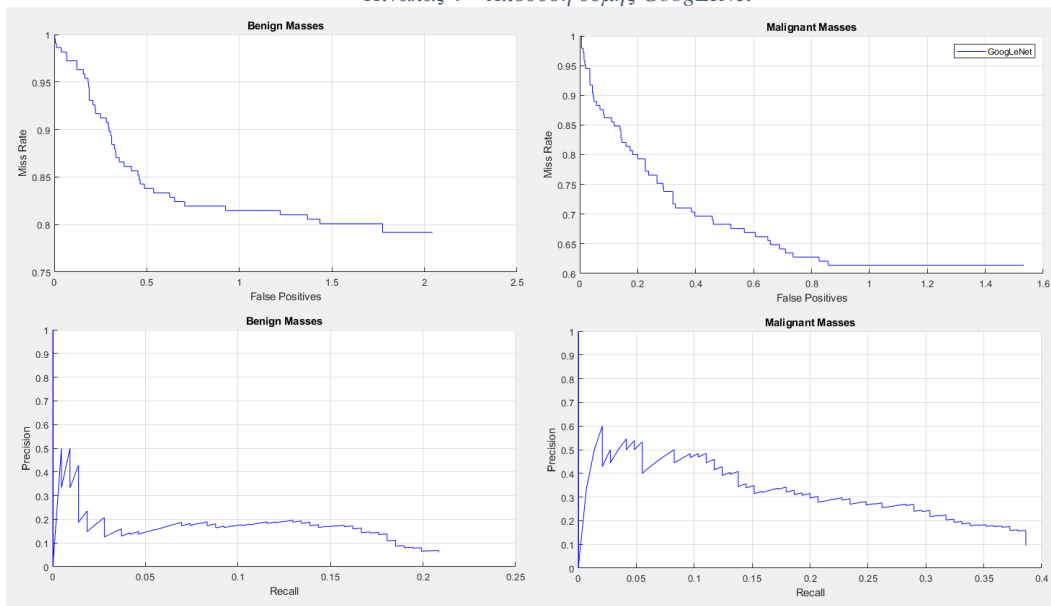
Εισάγοντας την υλοποίηση του GoogLeNet η επεξεργασία ξεκινάει με τις ίδιες κινήσεις που χρησιμοποιήθηκαν και για το Alexnet. Αφού δημιουργηθεί ένα νέο στρώμα εισόδου στις διαστάσεις που εξυπηρετεί, το πρώτο στρώμα συνέλιξης μορφοποιείται σε ένα χρωματικό κανάλι χρησιμοποιώντας το μέσο όρο των τριών αρχικών καναλιών σαν βάρος και κλίση για κάθε στοιχείο και εισάγεται ένα νέο πλήρως συνδεδεμένο στρώμα με είσοδο τους 1000 νευρώνες της εξόδου του προηγούμενου στρώματος και έξοδο τις τρεις κατηγορίες που εξυπηρετούν τους σκοπούς της εργασίας (καλοήθης, κακοήθης, υπόβαθρο).

Λόγω της περιπλοκότητας του δικτύου είναι απαραίτητο να οριστεί χειροκίνητα το επίπεδο πρόσδεσης του RPN δικτύου καθώς η εντολή trainFasterRCNN δεν μπορεί να ανταποκριθεί στις απαιτήσεις του δυσδιάστατου βάθους μοντέλου που χρησιμοποιούμε. Επιλέγεται σαν σημείο εξαγωγής χαρακτηριστικών (feature extraction) το τρίτο επίπεδο συνέλιξης του δικτύου καθώς έχει το απαιτούμενο ύψος για να εμφανίσει ικανοποιητικό βαθμό εντύπωσης των χαρακτηριστικών χωρίς να κινδυνεύει όσο τα επόμενα του στρώματα από υπερκάλυψη.

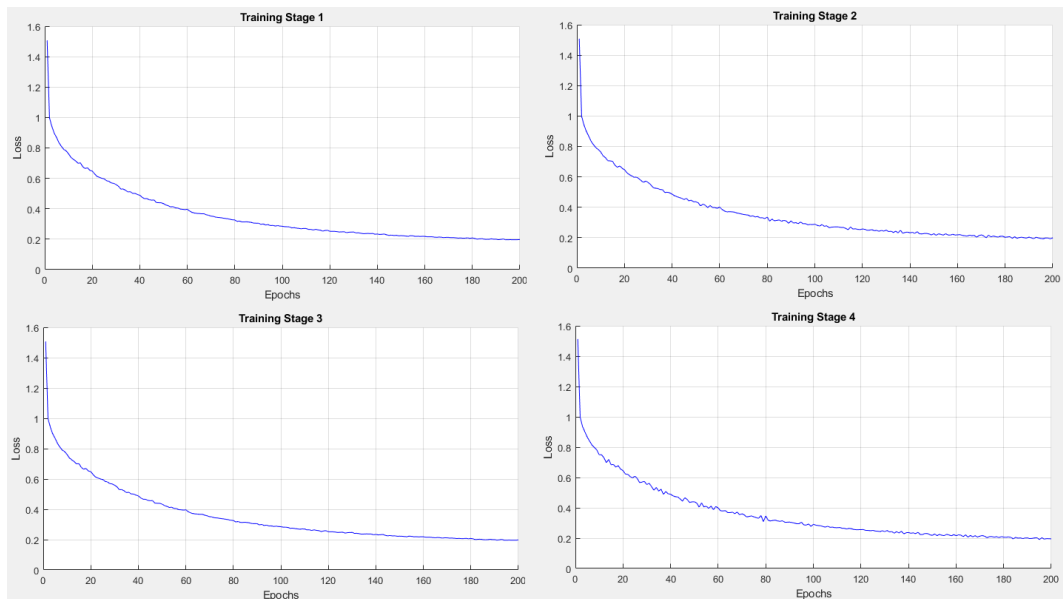
Εκπαιδευόντας το δίκτυο που προκύπτει και αξιολογώντας το με χρήση των αντίστοιχων σετ εντοπίζει αντικείμενα ενδιαφέροντος σε 334 εκ των 345 διαθέσιμων μαστογραφιών. Τα στατιστικά αποτελέσματα εμφανίζονται παρακάτω και παρουσιάζονται βελτιωμένα σε σχέση με τις δομές AlexNet που παρουσιάστηκαν έως τώρα.

Network Base	Benign Average Precision	Malignant Average Precision	Benign Average Miss Rate	Malignant Average Miss Rate	Ποσοστό εικόνων με εύρεση ROI
GoogLeNet	0.0380	0.1304	0.9312	0.8196	96.81%

Πίνακας 4 – Απόδοση δομής GoogLeNet



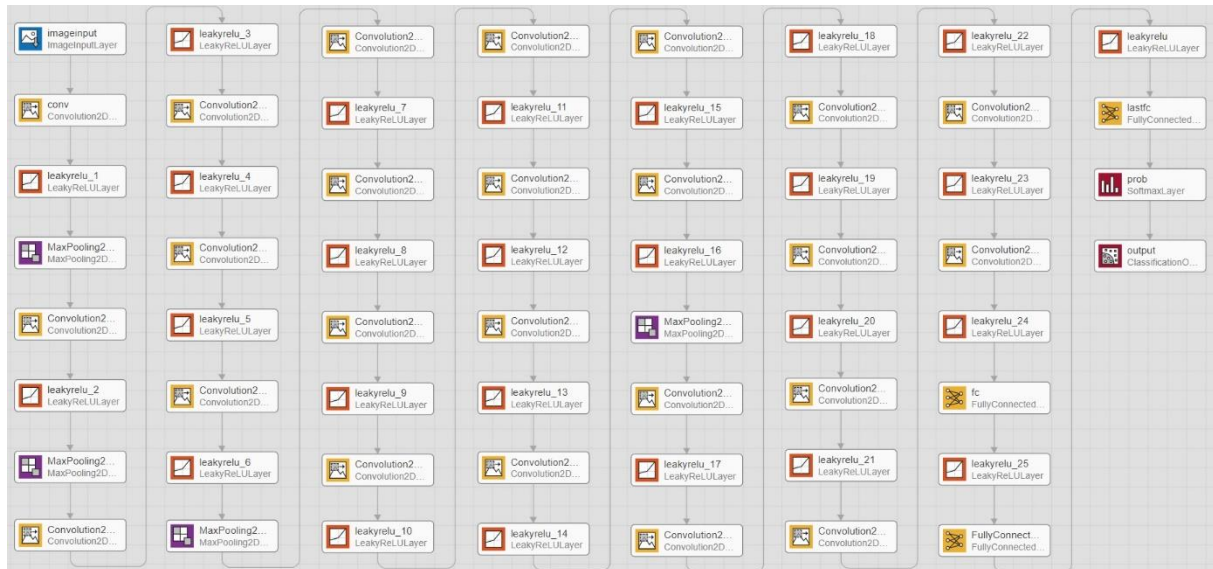
Εικόνα 27 – Σχέδιο απόδοσης δομής GoogLeNet



Εικόνα 26 – Απόλεια εκπαίδευσης δομής GoogLeNet

4.2.4 YOLO

Όπως έχει ήδη αναφερθεί τα RCNN δίκτυα χρησιμοποιούν ποικίλες μεθόδους παραγωγής προτάσεων περιοχών οι οποίες στη συνέχεια κατηγοριοποιούνται, η κάθε μια ξεχωριστά, αποκομμένη από τις υπόλοιπες περιοχές και το σύνολο της εικόνας. Μετά την κατηγοριοποίηση εφαρμόζονται διορθωτικοί αλγόριθμοι για τη μεταβολή της περιοχής που περιλαμβάνει το αντικείμενο ενδιαφέροντος σε αυτή που θα το περιβάλλει στον μεγαλύτερο βαθμό και με τη μεγαλύτερη ακρίβεια.



Εικόνα 28 – Δομή YOLO

Σύμφωνα με τους Redmon et al. στο [48] αυτή η παλινδρόμηση ανάμεσα στην περιοχή που προτείνεται, τη κατηγοριοποίηση της, τον επαναπροσδιορισμό της περιοχής και την επανακατηγοριοποίηση προκαλεί καθυστερήσεις στην εκτέλεση του εντοπισμού και δημιουργεί ένα περιβάλλον δύσκολο ρυθμιζόμενο κατά την εκπαίδευση. Η αντιπρόταση της έρευνας στον έλεγχο κατακερματισμένων περιοχών και τις καθυστερήσεις που αυτός συνεπάγεται είναι ο μοναδικός έλεγχος ολόκληρης της εικόνας. Το δίκτυο που προτείνεται (you only look once - YOLO) ελέγχει ολόκληρη την εικόνα εισόδου μία φορά και κατηγοριοποιεί και οριοθετεί τα αντικείμενα που εντοπίζει επιτόπου.

Ενώ το YOLO είναι γρήγορο και αποτελεσματικό για την κατηγοριοποίηση περιοχών, όπως η αναγνώριση του ανθρώπινου σχήματος ή αντικειμένων με μεγάλη επιφάνεια εικόνας όπως οχήματα, είναι αδύναμο για την ανίχνευση και την κατηγοριοποίηση μικρών αντικειμένων, ιδιαίτερα σε ομάδες όπως σμήνη πουλιών που αναφέρονται χαρακτηριστικά στη δημοσίευση. Αυτό προκύπτει από τον τρόπο με τον οποίο εξάγει τις περιοχές που κατηγοριοποιούνται και υποδεικνύει πως δεν είναι η κατάλληλη αρχιτεκτονική για τις ανάγκες της έρευνας, λόγω του θολού και συγκεχυμένου χαρακτήρα των απεικονίσεων που χρησιμοποιούνται και των αντικειμένων που περιέχουν.

Το βάθος του δικτύου φαίνεται εκ πρώτης όψεως ενθαρρυντικό και το δίκτυο μπορεί εύκολα να χρησιμοποιηθεί σαν σκελετός για την εκπαίδευση ενός Faster RCNN. Απαιτείται η επεξεργασία των στρωμάτων του δικτύου με τρόπο αντίστοιχο με τα προηγούμενα δίκτυα που αναφέρθηκαν, αντικαθιστώντας τα στρώματα εισόδου και το πρώτο δίκτυο συνέλιξης κατάλληλα ρυθμισμένα για ένα κανάλι δεδομένων και τέλος την προσθήκη του πλήρως συνδεδεμένου δικτύου και του στρώματος ταξινόμησης στην απόληξη, ρυθμισμένα αντίστοιχα για τους τρεις τύπους δεδομένων.

Παρόλο που δεν αξιοποιείται το συγκεκριμένο δίκτυο στην ενδεδειγμένη του υλοποίηση σαν ένα αυτόνομο δίκτυο αλλά σαν σκελετό για την κατασκευή ενός Faster RCNN θα είναι εξαιρετικά χρήσιμο για την εξαγωγή συμπερασμάτων σε σχέση με το βάθος του δικτύου και τη δομή του. Είναι εφικτό να παρατηρηθεί πως το YOLO αποτελεί την επανάληψη στρωμάτων συνέλιξης και relu με μερικά pooling layers ώστε να περιορίζουν την εκρηκτική αύξηση του αριθμού μεταβλητών. Επιπλέον παρατηρείται πως περιέχει τρία πλήρως συνδεδεμένα στρώματα και τελικά η δομή του είναι παρόμοια με ενός βαθύτερου AlexNet από το οποίο έχουν αφαιρεθεί τα dropout layers, γεγονός το οποίο πιθανός θα μειώνει την ακρίβεια του δικτύου σε σχέση με την αναμενόμενη βάση των μέχρι τώρα παρατηρήσεων.

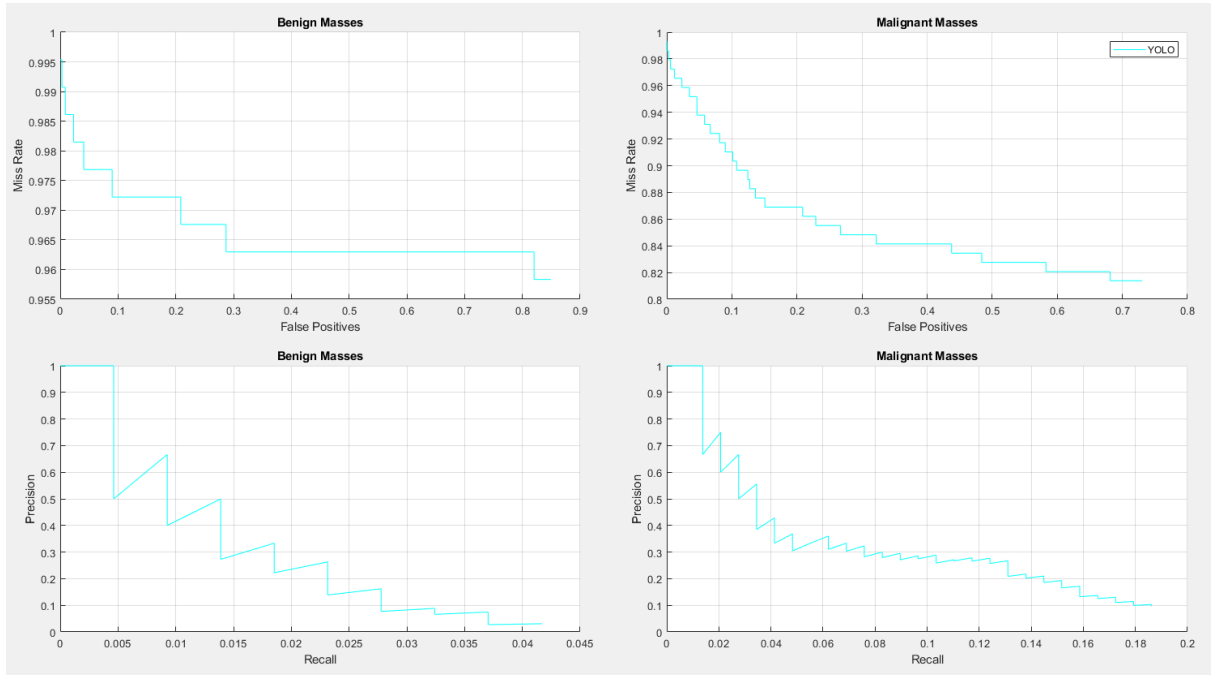
Το δίκτυο λόγω βάθους (130 εκατομμύρια μεταβλητές) απαιτεί μεγάλο διάστημα εκπαίδευσης, (απαιτήθηκαν 134 ώρες εκπαίδευσης για τη διέλευση 200 εποχών). Εν τέλει το δίκτυο εντόπισε 244 στις

Πτυχιακή Εργασία, τμήμα Μηχανικών Πληροφορικής
75. Αναγνώριση καρκινώματος / Συνάπαλος Δημήτριος 2051

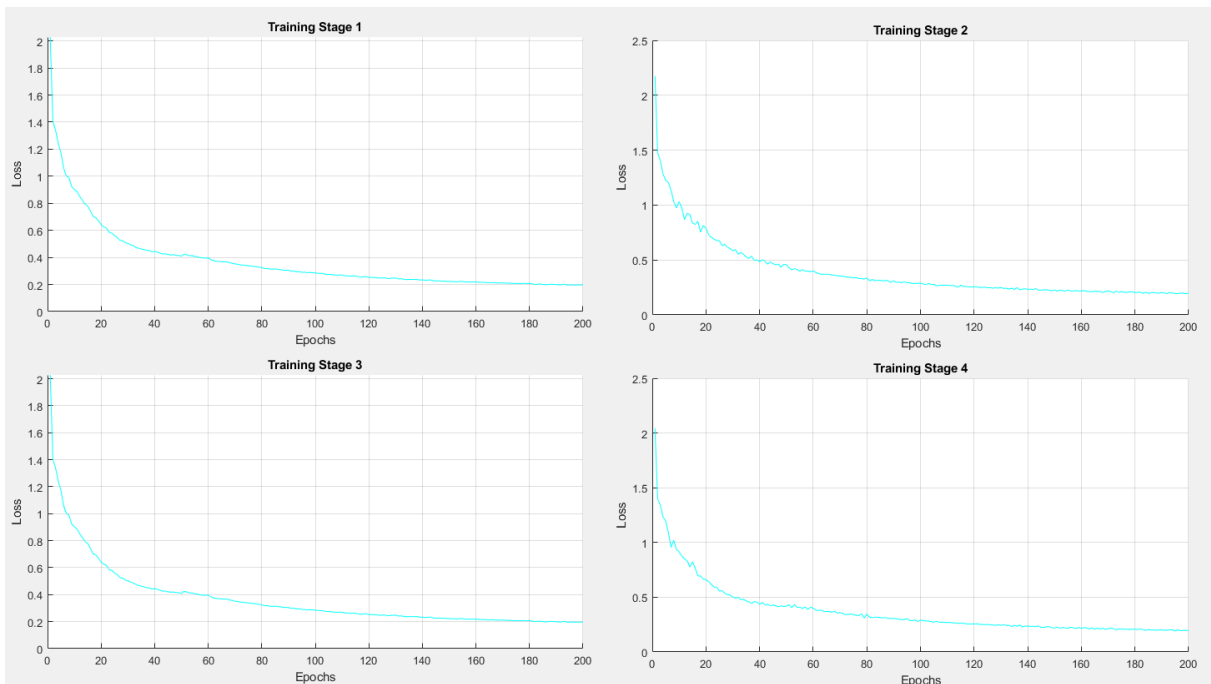
345 εικόνες, υψηλότερη απόδοση από τα AlexNet αλλά μικρότερη από την αντίστοιχη του GoogLeNet το οποίο απαιτήσε και μικρότερο χρόνο εκπαίδευσης. Ενδέχεται το συγκεκριμένο δίκτυο να απαιτεί περαιτέρω διάρκεια εκπαίδευσης από τις 200 εποχές που έχουν τυποποιηθεί και αναφέρονται στο Κεφάλαιο 5.3, ωστόσο η αύξηση των εποχών θα επιφέρει την ανάλογη αύξηση στις ήδη υψηλές χρονικές απαιτήσεις του δικτύου.

Network Base	Benign Average Precision	Malignant Average Precision	Benign Average Miss Rate	Malignant Average Miss Rate	Ποσοστό εικόνων με εύρεση ROI
YOLO	0.0144	0.0666	0.9732	0.8985	70.72%

Πίνακας 5 – Απόδοση δομής YOLO



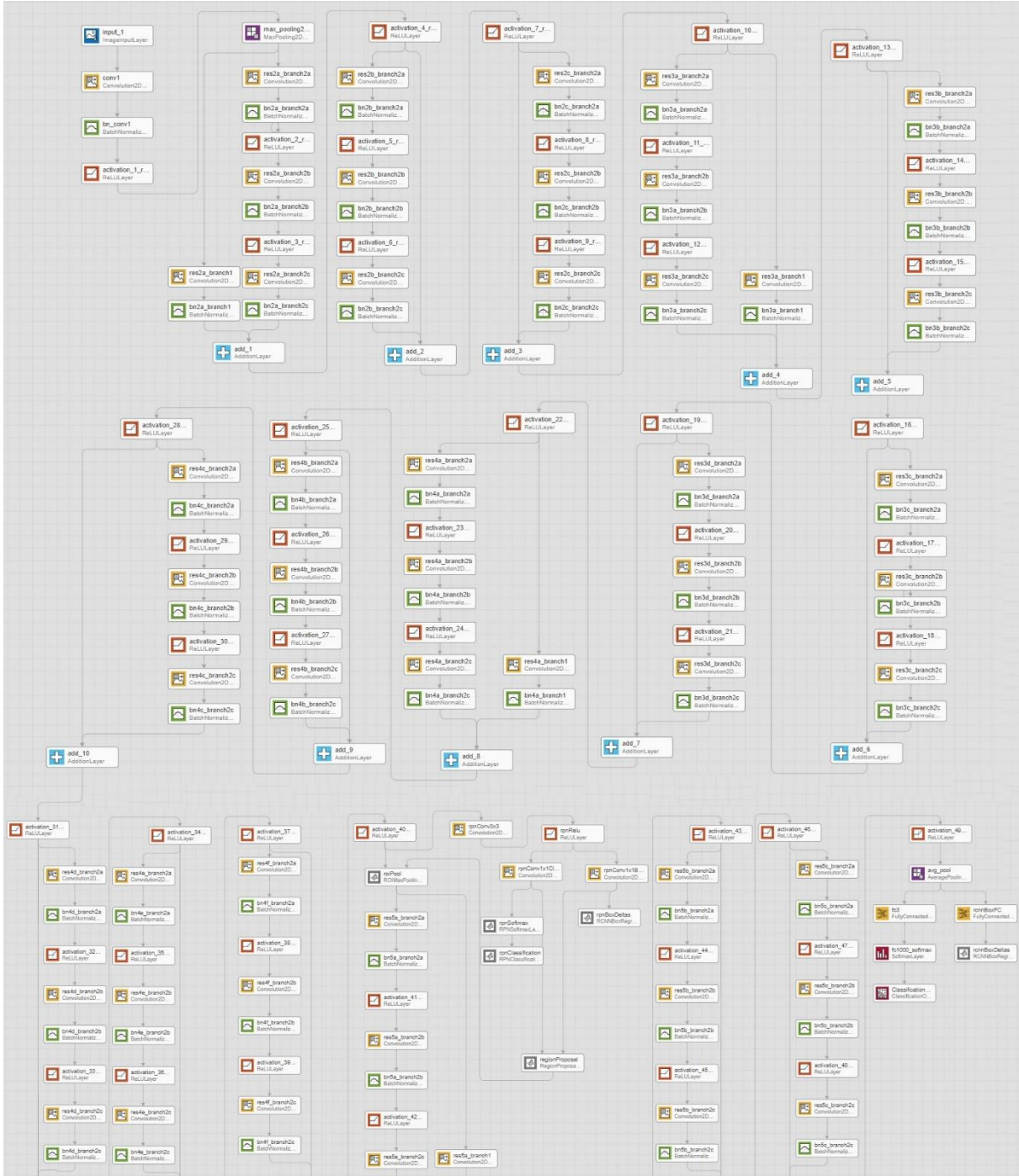
Εικόνα 30 – Σχέδιο απόδοσης δομής YOLO



Εικόνα 29 – Απόλεια εκπαίδευσης δομής YOLO

4.2.5 ResNet

Οι He et al. [45] αναφέρουν πως η αύξηση του βάθους ενός δικτύου δεν συνδέεται γραμμικά με την αύξηση της ακριβείας του. Παρατηρούν πως η αύξηση των στρωμάτων αυξάνει το ποσοστό λάθους τόσο κατά την εκπαίδευση όσο και κατά την αξιολόγηση, αυξάνοντας ταυτόχρονα την πολυπλοκότητα και τον απαιτούμενο χρόνο εκπαίδευσης.



Εικόνα 31 – Δομή ResNet50 με σύνδεση RPN στη ReLU Activation 40

Υποθέτουν πως συχνά εμφανιζόμενα προβλήματα σχετικά με την κλίση της εκπαίδευσης (gradient) και απότομες μεταβολές της έχουν λυθεί σε ικανοποιητικό βαθμό με την χρήση αρχικοποιημένων βαρών και την εισαγωγή στρωμάτων κανονικοποίησης στα δίκτυα, αντίθετα παρατηρούν εκφυλισμό στην ακρίβεια των δικτύων ο οποίος δεν εξηγείται από υπέρ/υπό κάλυψη του δικτύου καθώς φαίνεται μόνο να εντείνεται από την εμπάθυσή του.

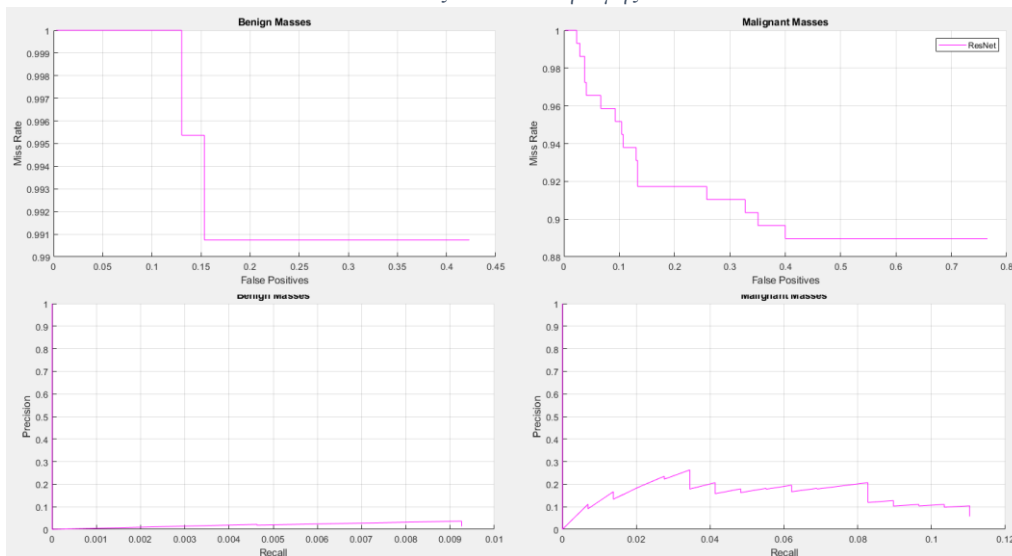
Η πρότασή τους σε απάντηση του ανωτέρω προβλήματος εστιάζεται στη χρήση συνδέσεων ανάμεσα στα επίπεδα του δικτύου, προσπαθώντας να αξιοποιήσουν την ταυτοτική μνήμη (identity mapping) των προηγούμενων στρωμάτων, προσθέτοντας το αποτέλεσμά τους με αυτό των επόμενων στρωμάτων. Το προτεινόμενο δίκτυο ονομάστηκε υπολειμματικό (Residual Network - ResNet) από την ιδιότητά του να διατηρεί και να αναπαράγει την γνώση των προηγούμενων στρωμάτων στα επόμενα και παρουσιάζει την ικανότητα να αυξάνει την ακρίβειά του όσο αυξάνεται το βάθος του.

Τα δίκτυα ResNet αποτελούν, κατά τα άλλα, βαθιά δίκτυα πλήρους συνέλιξης και μπορούν να ενσωματωθούν εύκολα σαν σκελετός για την εκπαίδευση ενός Faster RCNN δικτύου επαρκούς βάθους για τις ανάγκες της εργασίας. Η χρήση της υλοποίησης του ResNet-50 επιλέχθηκε καθώς, παρότι το βαθύτερο ResNet-100 θα εμφανίζει θεωρητικά μεγαλύτερη ακρίβεια, το βάθος του οδηγεί σε κορεσμό της διαθέσιμης μνήμης της κάρτας γραφικών λόγω του όγκου των μεταβλητών που καλείται να διαχειριστεί.

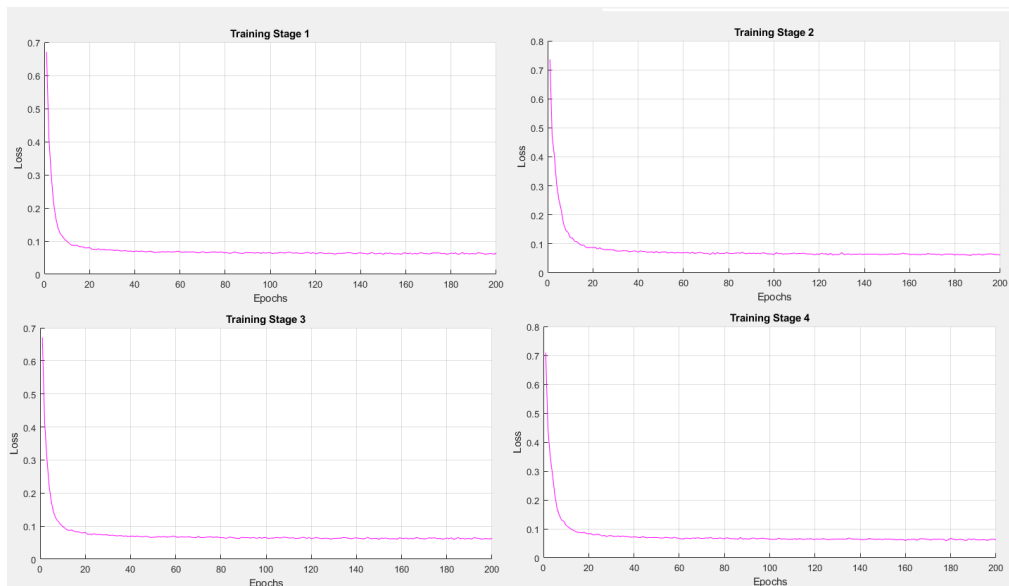
Η εκπαίδευση του δικτύου ResNet διήρκεσε περί τις 63 ώρες με τα αποτελέσματά του να μην παρουσιάζουν την ανάλογη απόδοση όπως εμφανίζεται παρακάτω.

Network Base	Benign Average Precision	Malignant Average Precision	Benign Average Miss Rate	Malignant Average Miss Rate	Ποσοστό εικόνων με εύρεση ROI
ResNet	0.0003	0.0190	0.9959	0.9446	40.00%

Πίνακας 6 – Απόδοση δομής ResNet



Εικόνα 33 – Σχέδιο απόδοσης δομής ResNet



Εικόνα 32 – Απόλεια εκπαίδευσης δομής ResNet

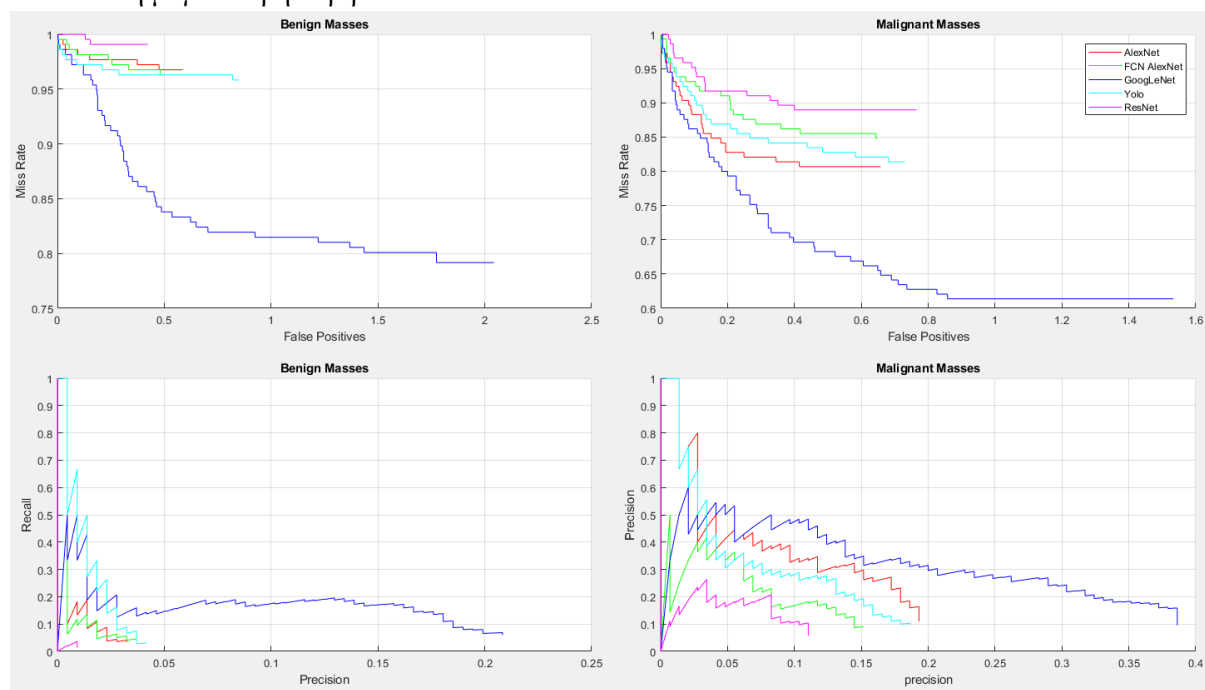
4.3 Ανασκόπηση αποτελεσμάτων βασικών δομών

Έως τώρα έχει γίνει αναφορά στη μορφοποίηση, υλοποίηση και εκπαίδευση πέντε δημοφιλών δομών CNN οι οποίες αξιοποιήθηκαν για την υλοποίηση πέντε διακριτών δικτύων Faster RCNN. Τα δίκτυα εκπαιδεύτηκαν χρησιμοποιώντας τους πίνακες αλήθειας που κατασκευάστηκαν στο Κεφάλαιο 2 και με παραμέτρους εκπαίδευσης στις οποίες θα γίνει αναφορά στο Κεφάλαιο 4 της εργασίας. Τέλος η απόδοσή τους αξιολογήθηκε με μεθόδους που παρουσιάζονται εκτενέστερα στο Κεφάλαιο 5, από τις οποίες προέκυψαν οι πίνακες και τα σχέδια που έχουν παρατεθεί μεμονωμένα με κάθε αρχιτεκτονική και παρουσιάζονται εδώ συγκεντρωτικά:

Network Base	Benign Average Precision	Malignant Average Precision	Benign Average Miss Rate	Malignant Average Miss Rate	Ποσοστό εικόνων με εύρεση ROI
AlexNet	0.0053	0.0811	0.9819	0.8821	60.29%
FCN AlexNet	0.0074	0.0392	0.9814	0.9190	43.18%
GoogLeNet	0.0380	0.1304	0.9312	0.8196	96.81%
YOLO	0.0144	0.0666	0.9732	0.8985	70.72%
ResNet	0.0003	0.0190	0.9959	0.9446	40.00%

Πίνακας 7 – Πίνακας απόδοσης Βασικών Δομών

Από την μέχρι τώρα απόδοση των δικτύων είναι εφικτή η εξαγωγή κάποιων συμπεράσματος. Παρατηρείται πως το δίκτυο AlexNet έχει μια εξαιρετικά ικανοποιητική απόδοση σε σχέση με το συνολικό του μέγεθος, τον αριθμό των μεταβλητών του και τον χρόνο εκπαίδευσης, καθώς κατατάσσεται τρίτο σε σειρά απόδοσης ωστόσο η ευκολία παραμετροποίησής του επιβάλλει περαιτέρω επεξεργασία με στόχο τη βελτίωση της απόδοσης. Από την απόδοση του FCN AlexNet συμπεραίνεται πως είναι βοηθητική για τα δίκτυα ή προσθήκη ενός ρηχού πλήρως συνδεδεμένου δικτύου στην απόληξη των δικτύων συνέλιξης. Στο Κεφάλαιο 6 θα παρατεθούν οι υλοποιήσεις περαιτέρω πειραμάτων ώστε να προσδιοριστεί το βάθος και το εύρος των πλήρως συνδεδεμένων στρωμάτων που αποδίδει τη μεγαλύτερη ακρίβεια.

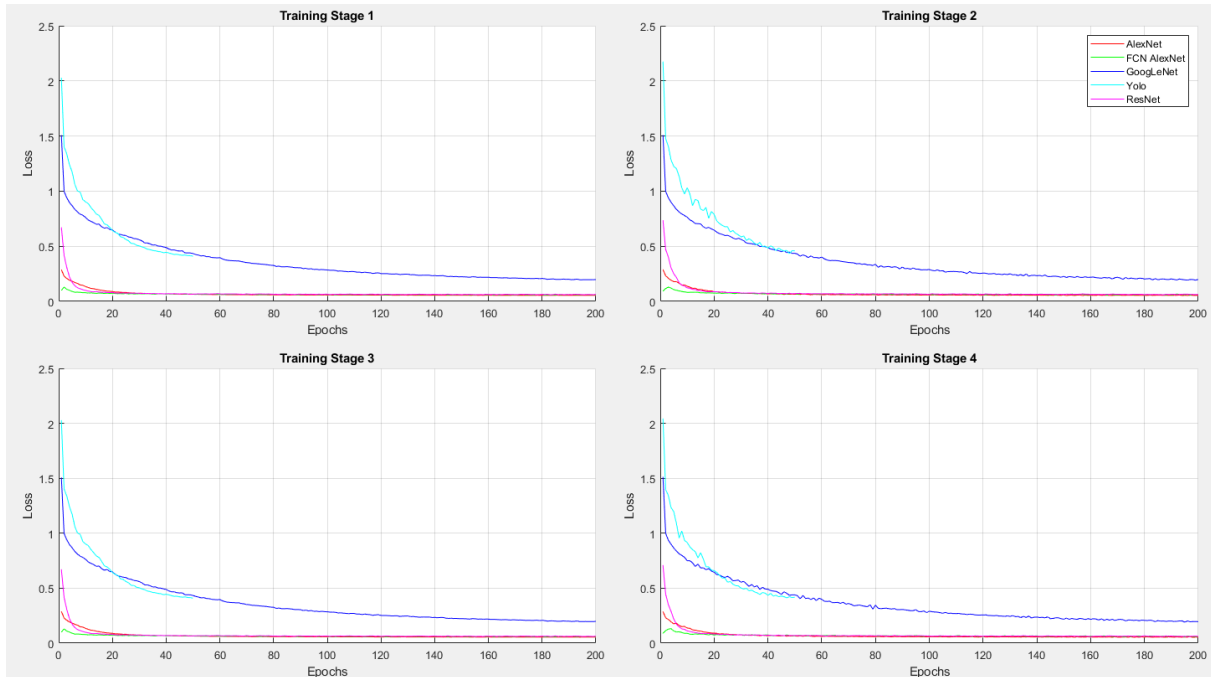


Εικόνα 34 – Σχέδιο απόδοσης Βασικών Δομών

Το δίκτυο GoogLeNet εμφανίζει τη μεγαλύτερη απόδοση που επετεύχθη, εντοπίζοντας αντικείμενα ενδιαφέροντος σχεδόν στο σύνολο του σετ ελέγχου. Η ερμηνεία που δίνεται είναι πως σε αυτό εξυπηρετεί η χαρακτηριστική πρωτοτυπία των δυοδιάστατων δομών Inception που υλοποιεί το συγκεκριμένο δίκτυο και θα παρατεθούν αποτελέσματα πειραματισμών ώστε να βελτιωθεί, αν είναι εφικτό, η απόδοσή του, όσο και προσπάθειες να μεταγισθούν τμήματα των στρωμάτων Inception σε δομές άλλων δικτύων.

Όσον αφορά τα δίκτυα YOLO και ResNet το βάθος τους αποτελεί ανασταλτικό παράγοντα υπό τις συγκεκριμένες παραμέτρους και καθώς κάθε πειραματισμός είναι εξαιρετικά χρονοβόρος η πληθώρα των στρωμάτων τους δεν είναι φιλική προς επέμβαση. Ενώ το YOLO πέτυχε ικανοποιητική απόδοση (στα πλαίσια πάντα των έως τώρα πειραμάτων) εμφανίζεται χρησιμότερη η επένδυση σε μικρότερα δίκτυα όπως το AlexNet κυρίως λόγω του φόρτου εργασίας που εναποθέτει στον εξοπλισμό η εκπαίδευση δικτύων αντίστοιχου μεγέθους. Το δίκτυο ResNet δεν απέδωσε αποτελέσματα αντίστοιχα των απαιτήσεων εκπαίδευσης του και αξιοποιείται σαν στατιστικό μέγεθος.

Παρουσιάζεται εδώ η υπόνοια πως η έλλειψη απόδοσης σε σχέση με το μέγεθος των YOLO και ResNet αφορά τη διαφορά της υλοποίησης, στα πλαίσια δομής Faster RCNN, από την αρχική τους δομή ως αυτόνομα και αυτόρρηκτα δίκτυα και υπό των περιορισμών που τίθενται από τις παραμέτρους στις οποίες γίνεται εκτενής αναφορά στο κεφάλαιο που ακολουθεί.



Εικόνα 35 – Απόλεια εκπαίδευσης Βασικών Δομών

5 Παράμετροι Εκπαίδευσης Πειραμάτων

5.1 Αναφορά στην λειτουργία της Μηχανικής Μάθησης

Για να αναπτυχθεί ο τρόπος με τον οποίο ένα ηλεκτρονικό σύστημα νευρώνων εκπαιδεύεται και να αξιολογηθούν οι αλγόριθμοι εκπαίδευσης, ώστε να επιλεγεί ο κατάλληλος για τις ανάγκες της εργασίας, είναι βοηθητικό να γίνει μια αναφορά στον τρόπο λειτουργίας τους και τον λόγο για τον οποίο αναπτύχθηκαν.

Ήδη από τα πρώτα νευρωνικά δίκτυα, όπως το Perceptron [8], χρειάστηκε να αναπτυχθεί μια μέθοδος με την οποία θα παραμετροποιούνται οι μεταβλητές τιμές των βαρών και των κλίσεων των νευρώνων. Για να γίνει αυτό έπρεπε πρώτα να οριστεί η μέθοδος αξιολόγησης του δικτύου, ένας τρόπος για να κβαντιστεί η απόδοσή του και η ανάγκη αυτή υλοποιήθηκε με την εισαγωγή του μετρικού που ονομάζουμε κόστος (Cost) ή απώλεια (Loss). Στην περίπτωση ενός δικτύου με δύο μόνο πλήρως συνδεδεμένα στρώματα (m εισόδους και n εξόδους), η τιμή κάθε νευρώνας εξόδου (L) αποτελεί το άθροισμα κάθε τιμής εισόδου (a) πολλαπλασιασμένης με ένα βάρος και με έναν επιπλέον αριθμό που λειτουργεί ως προκατάληψη, ενώ συχνά το αποτέλεσμα αυτό διοχετεύεται σε έναν αλγόριθμο ενεργοποίησης όπως μια σιγμοειδή πράξη ή ένα Rectified Linear Unit - Relu [49].

$$o_j^{(L)} = \sigma \left(w_{1j}^{(L)} o_{1j}^{(L-1)} + w_{2j}^{(L)} o_{2j}^{(L-1)} + \dots + w_{ij}^{(L)} o_{ij}^{(L-1)} + b_j^{(L)} \right)$$

Εξίσωση 1 – Αποτέλεσμα νευρώνας με σιγμοειδή λειτουργία ενεργοποίησης

Το αποτέλεσμα του δικτύου εξαρτάται από τον βαθμό της ενεργοποίησης των νευρώνων εξόδου, οπότε μπορούμε να συγκρίνουμε το αποτέλεσμα της ενεργοποίησης του δικτύου με το αποτέλεσμα που είναι γνωστό σε εμάς. Τετραγωνίζοντας τη διαφορά της ενεργοποίησης του νευρώνα από το επιθυμητό αποτέλεσμα και αθροίζοντας το αποτέλεσμα για όλους τους n νευρώνες εξόδου εξάγουμε την τιμή του κόστους για την συγκεκριμένη περίπτωση, δηλαδή ποσοτικοποιούμε την ικανότητα ορθής κατηγοριοποίησης του δικτύου μας. Επαναλαμβάνοντας τον υπολογισμό της απώλειας για όλα τα k δείγματα εισόδων μπορούμε να έχουμε μια εικόνα για την συνολική απόδοση του δικτύου.

$$C = \frac{1}{k} \sum_{p=1}^k \sum_{j=1}^n \left(o_j^{(L)} - Target_j \right)_p^2$$

Εξίσωση 2 – Υπολογισμός μέσου κόστους συστήματος

Από την στιγμή που έχουμε μια τιμή η οποία αξιολογεί την απόδοση του δικτύου μας μπορούμε να την συγκρίνουμε με το αποτέλεσμα του δικτύου με διαφορετικές τιμές βαρών. Σε αυτό το σημείο μπορούμε να διορθώσουμε τις τιμές των μεταβλητών μας ώστε η νέα ενεργοποίηση του δικτύου να μας δίνει μια τιμή απώλειας μικρότερη από την προηγούμενη. Για τον υπολογισμό του ποσού μεταβολής των μεταβλητών αξιοποιείται η τεχνική της ανάστροφης διάδοσης (Backpropagation) [7]. Ο συγκεκριμένος αλγόριθμος υπολογίζει την απόσταση ανάμεσα στο αποτέλεσμα της ενεργοποίησης και της επιθυμητής κατάστασης και προσπαθεί να μεταβάλλει κατά V τα βάρη των προηγούμενων στρωμάτων κατά $\partial w^{(L)}$ με τέτοιο τρόπο ώστε να δημιουργηθούν οι απαραίτητες συνθήκες, υπό τις οποίες οι επόμενοι νευρώνες θα έχουν τιμές που θα πλησιάζουν έστω και ελάχιστα περισσότερο τις επιθυμητές ∂C .

$$\gamma \text{ για } C = \left(o^{(L)} - Target \right)^2, Z^{(L)} = w^{(L)} o^{(L-1)} + b^{(L)}$$

$$\frac{\partial C}{\partial w^{(L)}} = \frac{\partial Z^{(L)}}{\partial w^{(L)}} \frac{\partial o^{(L)}}{\partial Z^{(L)}} \frac{\partial C}{\partial o^{(L)}}$$

Εξίσωση 3 – Λόγος μεταβολής κόστους/βάρους

Μετά τον επανυπολογισμό των βαρών η καινούρια πλήρης ενεργοποίηση του δικτύου οδηγεί στην επόμενη μεταβολή των βαρών έως ότου εντοπιστούν οι συντεταγμένες στις οποίες η απώλεια έχει τη μικρότερη δυνατή τιμή. Ένα χρήσιμο μέτρο ώστε να υπολογιστεί το ποσό της μεταβολής της τιμής κάθε μεταβλητής που θα έχει ως αποτέλεσμα τη μεγαλύτερη δυνατή μείωση του κόστους είναι η παράγωγος του μεταξύ τους λόγου.

$$\delta^{(L)} = \frac{\partial C}{\partial w^{(L)}} = \frac{\partial Z^{(L)}}{\partial w^{(L)}} \frac{\partial o^{(L)}}{\partial Z^{(L)}} \frac{\partial C}{\partial o^{(L)}} = o^{(L-1)} \sigma'(Z^{(L)}) 2(o^{(L)} - Target)$$

Εξίσωση 4 – Παράγωγος μεταβολής

Με αυτό τον τρόπο μπορούμε να υπολογίσουμε την μεταβολή της τιμής της κάθε μεταβλητής που θα έχει σαν αποτέλεσμα την μεγαλύτερη αύξηση του μέσου κόστους για μία επανάληψη όλων των

δειγμάτων μας. Οι μεταβολές αυτές ονομάζονται συλλογικά κλίση (Gradient - ΔC) και εφαρμόζοντας το αρνητικό πρόσημο μπορούμε να εντοπίσουμε την κλίση που θα μειώσει το κόστος μας με τον μεγαλύτερο βαθμό

$$\nabla C = \begin{bmatrix} \frac{\partial C}{\partial w^{(1)}} \\ \frac{\partial C}{\partial b^{(1)}} \\ \vdots \\ \frac{\partial C}{\partial w^{(L)}} \\ \frac{\partial C}{\partial b^{(L)}} \end{bmatrix} = (-\Delta C)$$

Εξίσωση 5 – Κλίση κατάβασης

Ο αλγόριθμος που υπολογίζει αντίστοιχα την μεταβολή του κάθε βάρους που θα επιφέρει τη μεγαλύτερη δυνατή μεταβολή στα επερχόμενα στρώματα και με κατάληξη τη μεγαλύτερη δυνατή μείωση του μέσου κόστους ονομάζεται στοχαστική κλίση κατάβασης (Stochastic Gradient Descent – SGD). Ο αλγόριθμος αυτός δεν αποτελεί πανάκια καθώς, παρότι είναι εξαιρετικά χρήσιμος για τον εντοπισμό τοπικών ελαχίστων κόστους, μπορεί, λόγω μικρού βηματισμού, να παγιδευτεί σε τοπικό επίπεδο και να μην εντοπίσει το απόλυτα χαμηλότερο κόστος, ενώ αντίθετα μπορεί από μεγάλο βήμα να υπερπηδήσει το ελάχιστο. Ένα ακόμα πρόβλημα προκύπτει από την τάση της κλίσης $\delta^{(L)}$ είτε να μειώνεται έως και μηδενίζεται, στην οποία περίπτωση τα βάρη σταματάν αν αλλάζουν (vanishing Gradient), είτε να εκτινάσσεται σε υψηλές τιμές οδηγώντας σε τυχαίες οδούς τη πορεία μείωσης (exploding Gradient).

5.2 Επεξήγηση αλγορίθμων κατάβασης απώλειας

Για την εξάλειψη των φαινομένων που επισημάνθηκαν έχουν αναπτυχθεί οι παρακάτω τεχνολογίες.

5.2.1 SGDM

Στο [9] διερευνάται ο κατάλληλος ρυθμός μάθησης για την εκπαίδευση του προτεινόμενου δικτύου. Οι Rumelhart et al. παρατηρούν πως η στοχαστική μέθοδος για να λειτουργήσει σωστά απαιτεί μικρό βηματισμό και μια τάση αδράνειας η οποία θα αποτρέπει την παλινδρόμηση στο ίδιο σημείο. Ο κανόνας που περιέγραψαν, όπου $\Delta w_{ji}(n+1)$ η κλίση του βάρους w_{ji} για την παρούσα περίοδο, $\Delta w_{ji}(n)$ η κλίση του ίδιου βάρους στην προηγούμενη περίοδο, η ο ρυθμός μάθησης και α η σταθερά απομείωσης, είναι ο εξής:

$$\Delta w_{ij}^{(L+1)} = \eta (\delta_{ij}^{(L)}) + \alpha \Delta w_{ij}^{(L)}$$

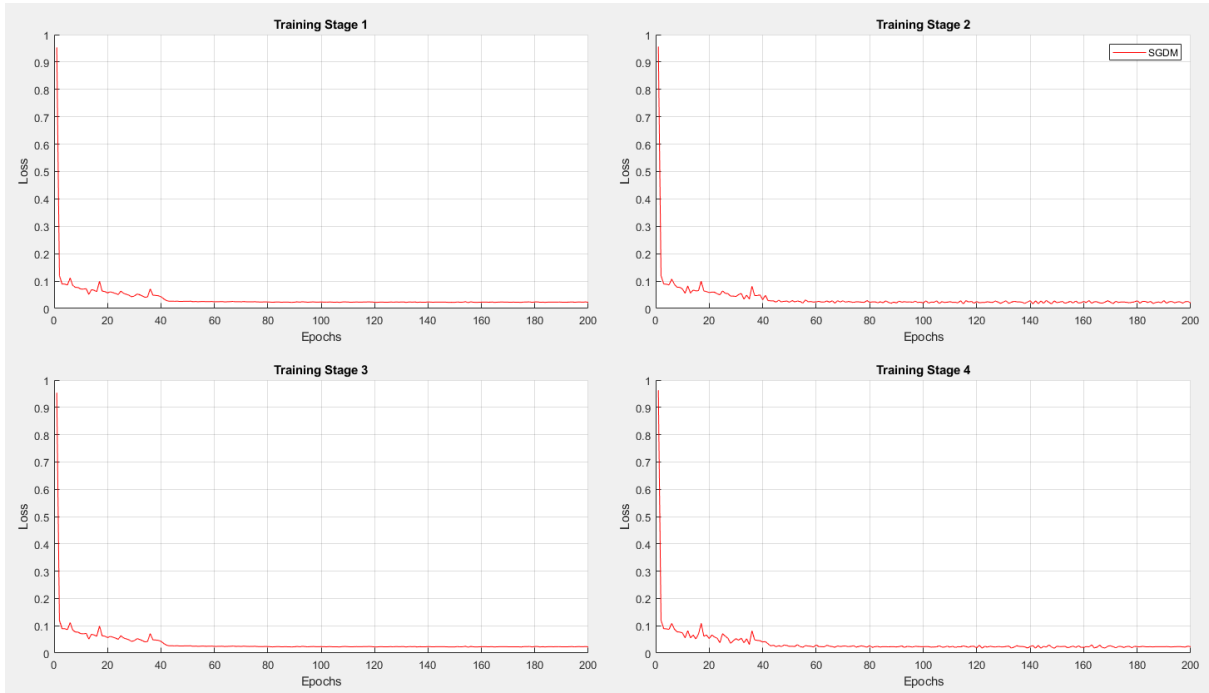
Εξίσωση 6 – Στοχαστική Κλίση Κατάβασης με Ορμή

Η ανωτέρω πρόταση, αξιοποιείται συχνά σε βαθιά δίκτυα και αποτελεί τον συνδυασμό του στοχαστικού αλγορίθμου με την εφαρμογή ορμής (momentum), όπως την περιέγραψε το 1964 ο σοβιετικός μαθηματικός Boris Polyak [50]. Η φυσική αναλογία της χρήσης της μνήμης του δικτύου για την προηγούμενη περίοδο είναι η αξιοποίηση της ορμής σαν αδράνεια, την τάση να κινείται προς μια κατεύθυνση, αποφεύγοντας έτσι την παλινδρόμηση γύρω από το ίδιο σημείο. Οι τιμές η και α καθορίζουν τον βαθμό στον οποίο επιδρούν στο τελικό αποτέλεσμα η κλίση κατάβασης και η ορμή.

Εάν αντιληφθούμε τη τιμή της κλίσης σαν το ύψος μιας δυσδιάστατης επιφάνειας, με διαστάσεις τις τιμές των βαρών, η επιτυχία της συγκεκριμένης μεθόδου, όπως αναφέρει και ο Matthew Zeiler [51], έγκειται στην ικανότητα της να επιταχύνει τη διάβαση της πεδιάδας και να διατηρεί την κατεύθυνση της διαδρομής όπου υπάρχει η τάση για στασιμότητα. Στο ίδιο έργο συγκρίνει τις τιμές λάθους για διαφορετικές υπερπαραμέτρους και αλγόριθμους κλίσης για ένα δίκτυο που έχει εκπαιδευτεί για 6 εποχές στην αναγνώριση των αριθμών MNIST και παρατηρεί βελτίωση των αποτελεσμάτων του δικτύου που εκπαιδεύτηκε με τον αλγόριθμο SGDM σε σχέση με την απλή στοχαστική κλίση.

Έκτοτε έχουν προταθεί επιπλέον αλγόριθμοι που επαυξάνουν ή διορθώνουν τη βασική λογική των στοχαστικών αλγορίθμων όπως ο EASGD των LeCun et al. [52], ο R-SGD των Huang et al. [53], ο ADAGRAD των Duchi et al [54] και ο ADADELTA των Zeiller et al [51]. Κάθε μία πρόταση προσφέρει μια καινούρια οπτική που μπορεί να ακολουθηθεί και θα απαιτηθεί βαθύτερη έρευνα, πέρα από τα

πλαίσια της παρούσας εργασίας για να εντοπιστούν τα όρια των αλγορίθμων, οι περιπτώσεις και οι παράμετροι με τις οποίες αποδίδουν.



Εικόνα 36 – Κατάβαση απώλειας SGDM

5.2.2 RSMPROP

Στο [55] οι Riedmiller et al. περιγράφουν την πρότασή τους για τη βελτίωση του αλγορίθμου SGD. Παρατηρούν πως η χρήση της ορμής επιτρέπει σε ένα δίκτυο να ξεφεύγει από τοπικά ακρότατα χωρίς να απαιτείται πολύ μεγάλο βήμα εκπαίδευσης, κάτι που μπορεί να μειώσει την ακρίβεια του δικτύου και συνοδεύεται με επιπλέον προβληματισμούς. Εντοπίζουν όμως πως η τάξη μεγέθους (magnitude) της ορμής μπορεί να είναι διαφορετική για διαφορετικά βάρη και αντίστοιχα είναι προβληματική η επιλογή μίας καθολικής τιμής που θα γενικευτεί στο σύνολο των περιπτώσεων. Προτείνουν τη μέθοδο rprop η οποία αξιοποιεί μόνο το πρόσημο της κλίσης και όχι την τιμή της, διαιρώντας πρακτικά την κλίση με το μέτρο της.

$$\Delta_{ij}^{(L)} = \begin{cases} \eta^- \Delta_{ij}^{(L-1)}, & \text{if } \delta_{ij}^{(L-1)} * \delta_{ij}^{(L)} > 0 \\ \eta^+ \Delta_{ij}^{(L-1)}, & \text{if } \delta_{ij}^{(L-1)} * \delta_{ij}^{(L)} < 0 \\ \Delta_{ij}^{(L-1)}, & \text{else} \end{cases}$$

$$\Delta w_{ij}^{(L)} = \begin{cases} -\Delta_{ij}^{(L)}, & \text{if } \delta_{ij}^{(L)} > 0 \\ +\Delta_{ij}^{(L)}, & \text{if } \delta_{ij}^{(L)} < 0 \\ 0, & \text{else} \end{cases}$$

$$w_{ij}^{(L+1)} = w_{ij}^{(L)} + \Delta w_{ij}^{(L)}$$

Εξίσωση 7 – Απόφαση κλίσης RPROP

Ουσιαστικά ελέγχει την συνέχεια του πρόσημου της κλίσης και κλιμακώνει πολλαπλασιαστικά ανάλογα (με αριθμό μεγαλύτερο του 1 εφόσον υπάρχει συνέχεια, με αριθμό μικρότερο του 1 και μεγαλύτερο του 0, εάν το πρόσημο αντιστρέφεται). Αυτό οδηγεί σε συμπεριφορά αντίστοιχη με τη χρήση ορμής για το ξεπέρασμα των τοπικών ελαχίστων αλλά με μία γενικευμένη τάξη μεγέθους για όλες τις αλλαγές βαρών της περιόδου.

Ο αλγόριθμος rprop λειτουργεί σωστά μόνο σε δέσμες (mini-batches) που περιλαμβάνουν το σύνολο των διαθέσιμων δειγμάτων καθώς η κλίση αλλάζει σε κάθε επανάληψη και όταν αντιμετωπίζει δέσμες μικρότερες του συνόλου σε κάθε δέσμη χρησιμοποιείται διαφορετική κλίση αντί μιας για όλες τις δέσμες. Κάτι τέτοιο αντιμετωπίζεται από τους Tieleman και Hinton [56] με την υλοποίηση του

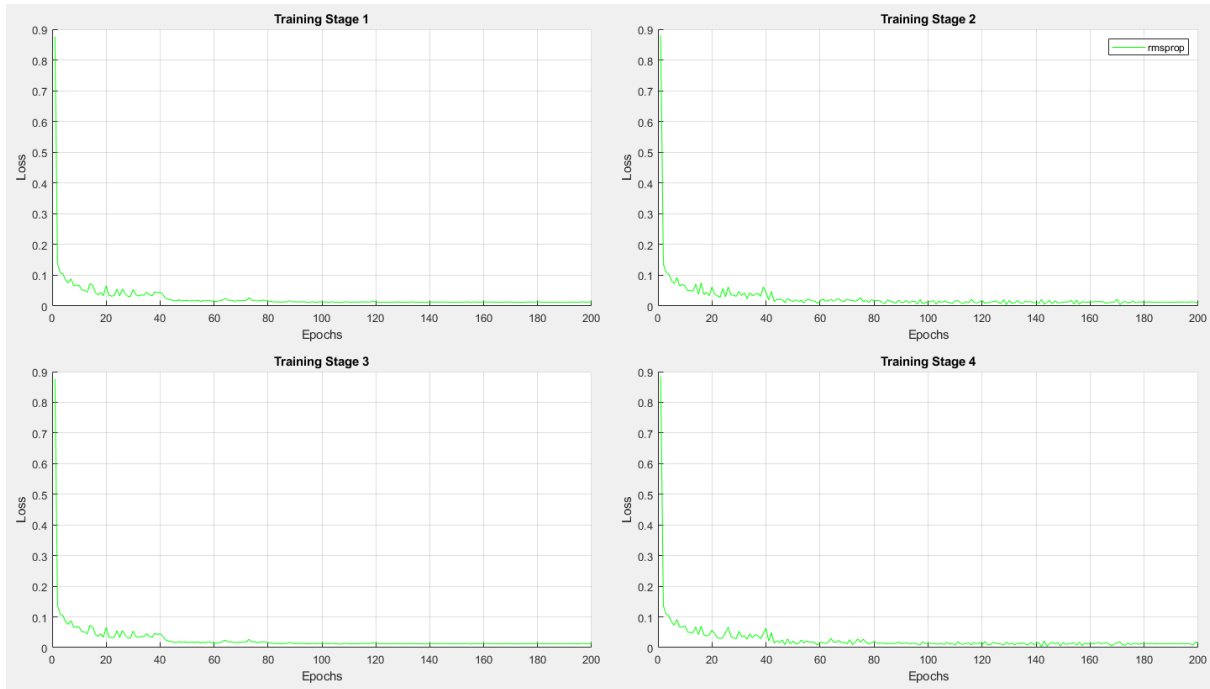
αλγόριθμου rmsprop (root mean square propagation) ο οποίος χρησιμοποιεί τον μέσο όρο από τα τετράγωνα των κλήσεων από κάθε κοντινή περίοδο ώστε να κανονικοποιήσει τον παρονομαστή της διαίρεσης για όλες τις δέσμες.

$$m_w^{(L)} = \frac{\eta m_w^{(L-1)}}{10} + \frac{(\delta^{(L)})^2}{10}$$

$$\Delta w_{ij}^{(L)} = \frac{\eta (\delta_{ij}^{(L)}) + a \Delta w_{ij}^{(L)}}{\sqrt{m_w^{(L)}}}$$

$$w_{ij}^{(L+1)} = w_{ij}^{(L)} + \Delta w_{ij}^{(L)}$$

Εξίσωση 8 – Απόφαση κλίσης RMSPROP



Εικόνα 37 – Κατάβαση απώλειας RMSPROP

5.2.3 ADAM

Ο αλγόριθμος ADAM (Adaptive Moment Estimation) [57] προέκυψε από την ανάγκη να γενικευτούν οι εφαρμογές των αλγορίθμων που επαυξάνουν την λειτουργικότητα της κλίσης κατάβασης. Όπως είδαμε οι τεχνικές που περιγράψαμε έχουν εξειδικευμένες δυνατότητες και χρειάζεται ιδιαίτερος πειραματισμός ώστε να εντοπιστεί η κατάλληλη μέθοδος για την κάθε ανάγκη. Οι Kingma και Ba σχημάτισαν τον ADAM βασίζόμενοι στους υπάρχοντες αλγόριθμους RMSPROP και ADAGRAD, επιδιώκοντας να συσχετίσουν την αντοχή του RMSPROP σε σειριακά και δυναμικά δεδομένα, με την ικανότητα του ADAGRAD να διαχειρίζεται αραιές κλίσεις. Πρακτικά, η χρήση των κινούμενων μέσων όρων του RMSPROP ενισχύεται από έναν αλγόριθμο πρόβλεψης κλίσεων.

$$E[m^{(L)}] = E[\Delta^{(L)}]$$

$$E[v^{(L)}] = E[(\Delta^{(L)})^2]$$

Εξίσωση 9 – Πρόβλεψη δύο στιγμιότυπων (moments)

Οι προβλέψεις επιτρέπουν τη χρήση των μέσων όρων από τις πρώτες επαναλήψεις. Για τον υπολογισμό των μέσων όρων m για την πρώτη πρόβλεψη και v για την δεύτερη, χρησιμοποιούνται δύο νέες υπερπαραμέτροι οι β_1, β_2 με τιμές 0.9 και 0.999 αντίστοιχα. Οι τιμές αυτές σπάνιος διαφέρουν καθώς έχει φανεί να λειτουργούν ικανοποιητικά για τις ανάγκες των περισσότερων πειραμάτων. Από την μορφή των εξισώσεων παρατηρούμε την αναδρομικότητά τους οπότε μπορούμε να γενικεύσουμε για τις επόμενες περιόδους.

$$m^{(L)} = \beta_1 m^{(L-1)} + (1 - \beta_1) \Delta^{(L)}$$

$$v^{(L)} = \beta_2 v^{(L-1)} + (1 - \beta_2) (\Delta^{(L)})^2$$

Εξίσωση 10 – Υπολογισμός μέσων όρων προβλέψεων

$$m^{(L)} = (1 - \beta_1) \sum_{i=0}^L \beta_1^{(L-i)} \Delta^{(i)}$$

$$v^{(L)} = (1 - \beta_2) \sum_{i=0}^L \beta_2^{(L-i)} (\Delta^{(i)})^2$$

Εξίσωση 11 – Γενικευμένος μέσος όρος προβλέψεων

Σε αυτό το σημείο ο αλγόριθμος λαμβάνει υπόψιν του την προκατάληψη που έχουμε ορίσει και επανυπολογίζει τις προβλέψεις του χρησιμοποιώντας τα καινούρια δεδομένα. Η κίνηση αυτή έχει σκοπό τη διόρθωση του βήματος που έχει προβλεφθεί καθώς είναι ουσιαστικότερο να διορθώσεις ένα λάθος που έχει γίνει παρά να προβλέψεις εξ αρχής τον σωστό βηματισμό.

$$E[m^{(L)}] = E \left[(1 - \beta_1) \sum_{i=1}^L \beta_1^{(L-i)} \Delta^{(i)} \right]$$

$$= E[\Delta^{(i)}] (1 - \beta_1) \sum_{i=1}^L \beta_1^{(L-i)} + b^{(L)}$$

$$= E[\Delta^{(i)}] (1 - \beta_1) + b^{(L)}$$

$$E[v^{(L)}] = E \left[(1 - \beta_2) \sum_{i=1}^L \beta_2^{(L-i)} \Delta^{(i)} \right]$$

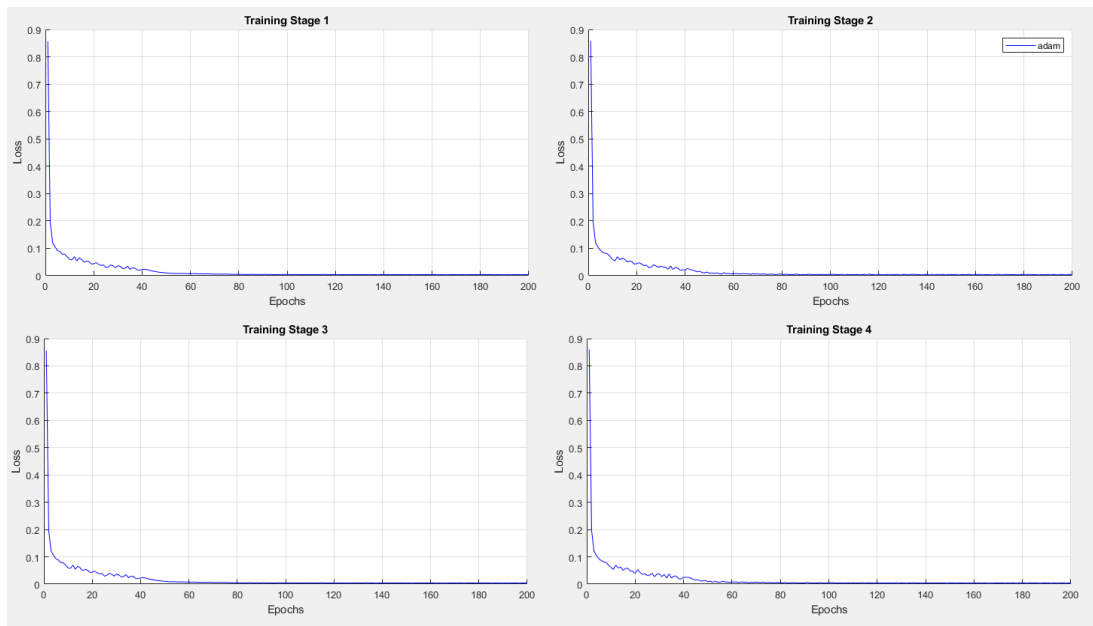
$$= E[\Delta^{(i)}] (1 - \beta_2) \sum_{i=1}^L \beta_2^{(L-i)} + b^{(L)}$$

$$= E[\Delta^{(i)}] (1 - \beta_2) + b^{(L)}$$

$$\hat{m}^{(L)} = \frac{m^{(L)}}{1 - \beta_1}$$

$$\hat{v}^{(L)} = \frac{v^{(L)}}{1 - \beta_2}$$

Εξίσωση 12 – Υπολογισμός προβλέψεων λαμβάνοντας υπόψιν την προκατάληψη



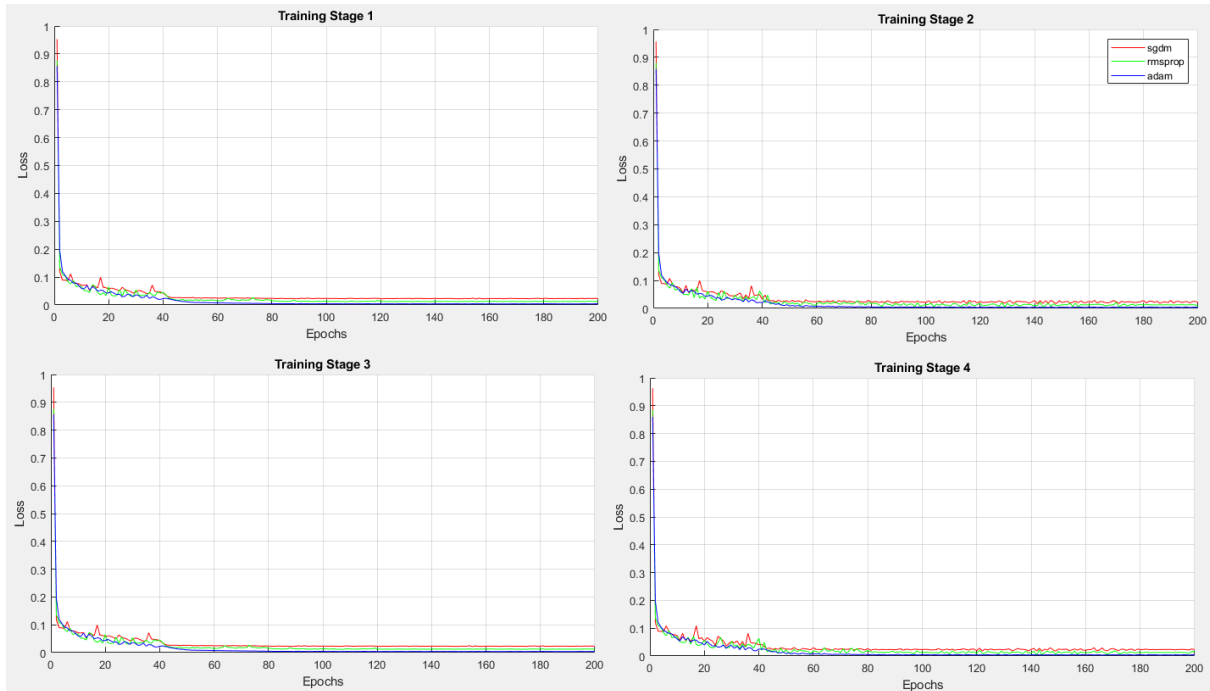
Εικόνα 38 – Κατάβαση απώλειας ADAM

Τέλος οι καινούριες προβλέψεις εξυπηρετούν στον υπολογισμό των βαρών. Σε αυτό το σημείο αξιοποιούμε τις παραμέτρους ήτα ως το βηματισμό της μάθησης και την τιμή έψιλον, μια μικρή τιμή που σαν σκοπό έχει να αποφεύγει τον πιθανό μηδενισμό του παρονομαστή του κλάσματος.

$$w^{(L+1)} = w^{(L)} - \eta \frac{\hat{m}^{(L)}}{\sqrt{\hat{v}^{(L)} + \epsilon}}$$

Εξίσωση 13 – Υπολογισμός βαρών ADAM

Από τα σχήματα της Εικόνα 37 και Εικόνα 38 είναι εμφανές πως και οι τρεις μέθοδοι κατάβασης παρουσιάζουν αντίστοιχη κατάβαση απώλειας και συγκλίνουν περίπου στις 80 εποχές για ένα AlexNet δίκτυο. Λόγω εξοικείωσης και παραμετροποιησιμότητας επιλέγεται ο SGDM για το σύνολο της εργασίας.



Εικόνα 39 – Συγκριτικό σχήμα κατάβασης απώλειας

5.3 Τυποποίηση παραμέτρων εκπαίδευσης

Κατά τη διάρκεια της εκπαίδευσης μπορούμε να ορίζονται παράμετροι, κατάλληλες προς τα ιδιαίτερα πειραματικά δεδομένα ώστε να επιτευχθεί βελτίωση της ακρίβειας εντοπισμού και μείωση του χρόνου εκπαίδευσης. Το περιβάλλον Matlab επιτρέπει την επιλεκτική ρύθμιση δεκάδων παραγόντων, ωστόσο, η παρούσα εργασία επικεντρώνεται μόνο στους βασικότερους εξ αυτών. Για τους πειραματισμούς χρησιμοποιούνται οι default τιμές των παραμέτρων, προσδιορίζοντας συγκεκριμένα μόνο τις παραμέτρους που ελέγχονται. Για την υλοποίηση του δικτύου αξιοποιείται η αρχιτεκτονική AlexNet όπως παρουσιάζεται στην Εικόνα 17.

5.3.1 Διάρκεια Εκπαίδευσης

Σημαντικός ρυθμιστικός παράγοντας της ποιότητας εκπαίδευσης είναι η συνολική διάρκεια της, η οποία καθορίζεται από τον αριθμό των εποχών για τον οποίο θα πραγματοποιηθεί το κάθε στάδιο της εκπαίδευσης. Ως μία εποχή ορίζεται η ενεργοποίηση του δικτύου με όλες τις εικόνες εκπαίδευσης και στο τέλος της υπολογίζονται σωρευτικά η απώλεια και οι αλλαγές των βαρών που είναι απαραίτητες για τη μείωσή της.

Θεωρητικά μεγαλύτερη διάρκεια εκπαίδευσης οδηγεί σε μεγαλύτερη ακρίβεια του εκπαιδευόμενου συστήματος, ωστόσο, λαμβάνοντας υπόψιν πως ο χρόνος εκπαίδευσης είναι περατό μέγεθος και κύριος παράγοντας του κόστους εκπαίδευσης ενός δικτύου, απαιτείται ο εντοπισμός της απολύτως απαραίτητης διάρκειας εκπαίδευσης, όπως και ο ορίσμός του σημείου από το οποίο και στη συνέχεια αύξηση του χρόνου εκπαίδευσης δεν οδηγεί σε αντίστοιχη αύξηση της ακρίβειας του εκπαιδευόμενου δικτύου. Έχει υποδειχθεί από το 4.2 πως βαθιά δίκτυα όπως το YOLO έχουν μεγάλες χρονικές

απαιτήσεις για την εκπαίδευσή τους, άμεσο απότοκο του εξαιρετικού όγκου των μεταβλητών τους. Το γεγονός αυτό εξαναγκάζει στο συμβιβασμο με δίκτυα που δεν θα έχουν εκπαιδευτεί ολοκληρωμένα καθώς περαιτέρω επένδυση χρόνο θα στερήσει τη δυνατότητα εμβάθυνσης πειραματικά σε ρηχότερες δομές οι οποίες εκπαιδεύονται γρηγορότερα σε εποχές εκπαίδευσης και συνεπώς σε συνολικό απαιτούμενο χρόνο, όπως και αντικατοπτρίζουν καλύτερα τα αποτελέσματα των επεμβάσεων στις δομές τους. Απαιτείται έτσι ο προσδιορισμός ενός ενιαίου πλαισίου εποχών το οποίο θα επιτρέψει τη μεγαλύτερη δυνατή αναλογία μείωσης απώλειας προς χρόνο εκπαίδευσης.

Από την Εικόνα 40 είναι εμφανές πως μετά την εποχή 150 η μείωση της απώλειας έχει ολοκληρωθεί σε μεγάλο βαθμό και περαιτέρω εκπαίδευση ενδεχόμενος να μην είναι συμφέρουσα ως προς την απόδοσή της. Παρουσιάζονται μαζί με τα στοιχεία του δικτύου AlexNet και οι αντίστοιχες τιμές του GoogLeNet ώστε να γίνει αντιληπτός ο ρόλος που διαδραματίζει στην απαραίτητη διάρκεια εκπαίδευσης το βάθος του δικτύου. Προκύπτει πως το GoogLeNet συγκλίνει γρηγορότερα από το AlexNet και αντίστοιχα μπορούμε να υποθέσουμε πως δεν απαιτείται σημαντικά μεγαλύτερος χρόνος εκπαίδευσης για βαθύτερα δίκτυα, διατηρώντας πάντα την επιφύλαξη που προκύπτει από το γεγονός πως η συμπεριφορά του δικτύου GoogLeNet καθορίζεται από την πρωτότυπη αρχιτεκτονική του.

Είναι σημαντικό εδώ σημειωθεί το γεγονός πως το πρώτο και το τρίτο στάδιο εκπαίδευσης αφορούν μόνο το RPN τμήμα του δικτύου οπότε η ουσιαστική εκπαίδευση των βαρών του σκελετού του δικτύου που αξιοποιείται πραγματοποιείται στο δεύτερο και, κυρίως, στο τέταρτο στάδιο εκπαίδευσης. Με σκοπό να διευκρινιστεί η βέλτιστη περίοδος εκπαίδευσης παρουσιάζονται τα βήματα που εμφανίζει η μείωση της απώλειας όπως παρακάτω:

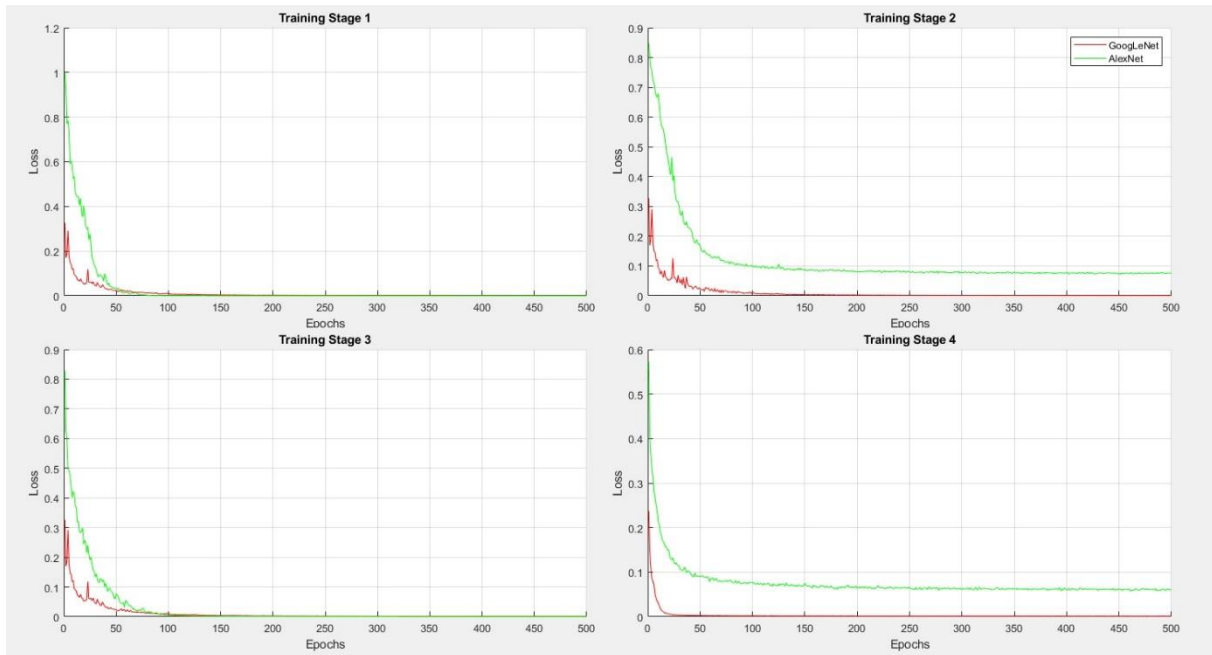
Epoch	AlexNet				GoogleNet			
	Stage 1	Stage 2	Stage 3	Stage 4	Stage 1	Stage 2	Stage 3	Stage 4
1	1.0002	0.8476	0.8282	0.5728	0.3265	0.3275	0.3265	0.2371
50	0.0333	0.1663	0.0777	0.0905	0.0232	0.0225	0.0232	0.0027
100	0.0005	0.0997	0.0056	0.076	0.0094	0.0119	0.0094	0.0018
150	0.0002	0.089	0.0018	0.0751	0.0035	0.0038	0.0035	0.0015
200	0.0001	0.0808	0.0009	0.0635	0.0026	0.0027	0.0026	0.0015
250	0.0001	0.0812	0.0007	0.0635	0.0018	0.0018	0.0018	0.0014
300	0	0.079	0.0005	0.061	0.0015	0.0015	0.0015	0.0014
350	0	0.0808	0.0005	0.0622	0.0013	0.001	0.0013	0.0014
400	0	0.0796	0.0004	0.0599	0.0012	0.0013	0.0012	0.0013
450	0	0.0751	0.0004	0.0625	0.001	0.0009	0.001	0.0012
500	0	0.0765	0.0004	0.0605	0.0009	0.0009	0.0009	0.0012
Epoch	Stage 1		Stage 2		Stage 3		Stage 4	
	AlexNet	GoogleNet	AlexNet	GoogleNet	AlexNet	GoogleNet	AlexNet	GoogleNet
1	1.0002	0.3265	0.8476	0.3265	0.8282	0.3265	0.5728	0.2371
50	0.0333	0.0232	0.1663	0.0232	0.0777	0.0232	0.0905	0.0027
100	0.0005	0.0094	0.0997	0.0094	0.0056	0.0094	0.076	0.0018
150	0.0002	0.0035	0.089	0.0035	0.0018	0.0035	0.0751	0.0015
200	0.0001	0.0026	0.0808	0.0026	0.0009	0.0026	0.0635	0.0015
250	0.0001	0.0018	0.0812	0.0018	0.0007	0.0018	0.0635	0.0014
300	0	0.0015	0.079	0.0015	0.0005	0.0015	0.061	0.0014
350	0	0.0013	0.0808	0.0013	0.0005	0.0013	0.0622	0.0014
400	0	0.0012	0.0796	0.0012	0.0004	0.0012	0.0599	0.0013
450	0	0.001	0.0751	0.001	0.0004	0.001	0.0625	0.0012
500	0	0.0009	0.0765	0.0009	0.0004	0.0009	0.0605	0.0012

Πίνακας 8 – Πίνακας Απώλειας ανά Εποχή

Από τον Πίνακα 2, εστιάζοντας στο τέταρτο στάδιο της εκπαίδευσης, παρατηρούμε πως έως την εποχή 200 η σύγκλιση έχει ολοκληρωθεί στον μεγαλύτερο βαθμό ενώ ένα μικρό βήμα ακόμα γίνεται μέχρι και την 250 σε δίκτυα βαθύτερα του AlexNet. Περαιτέρω εκπαίδευση φαίνεται να παρουσιάζει μειούμενη απόδοση όπως παρατηρούμε από τη μείωση της απώλειας για το δίκτυο GoogLeNet, στο στάδιο 4, κατά 0.0001 σε περίοδο 150 εποχών (εποχές 250-400). Από τα παραπάνω επιλέγουμε την

Πτυχιακή Εργασία, τμήμα Μηχανικών Πληροφορικής
75. Αναγνώριση καρκινώματος / Συνάπαλος Δημήτριος 2051

εκπαίδευση των δικτύων μας για 200 εποχές καθώς θεωρούμε πως η αύξηση κατά 25% του χρόνου εκπαίδευσης δεν συμφέρει όταν αποδίδει μείωση της απώλειας μόλις 6% για το GoogLeNet (σε αντίθεση η μάθηση κατά τις εποχές 150-200 μείωσε την απώλεια κατά 30% για το δίκτυο AlexNet).



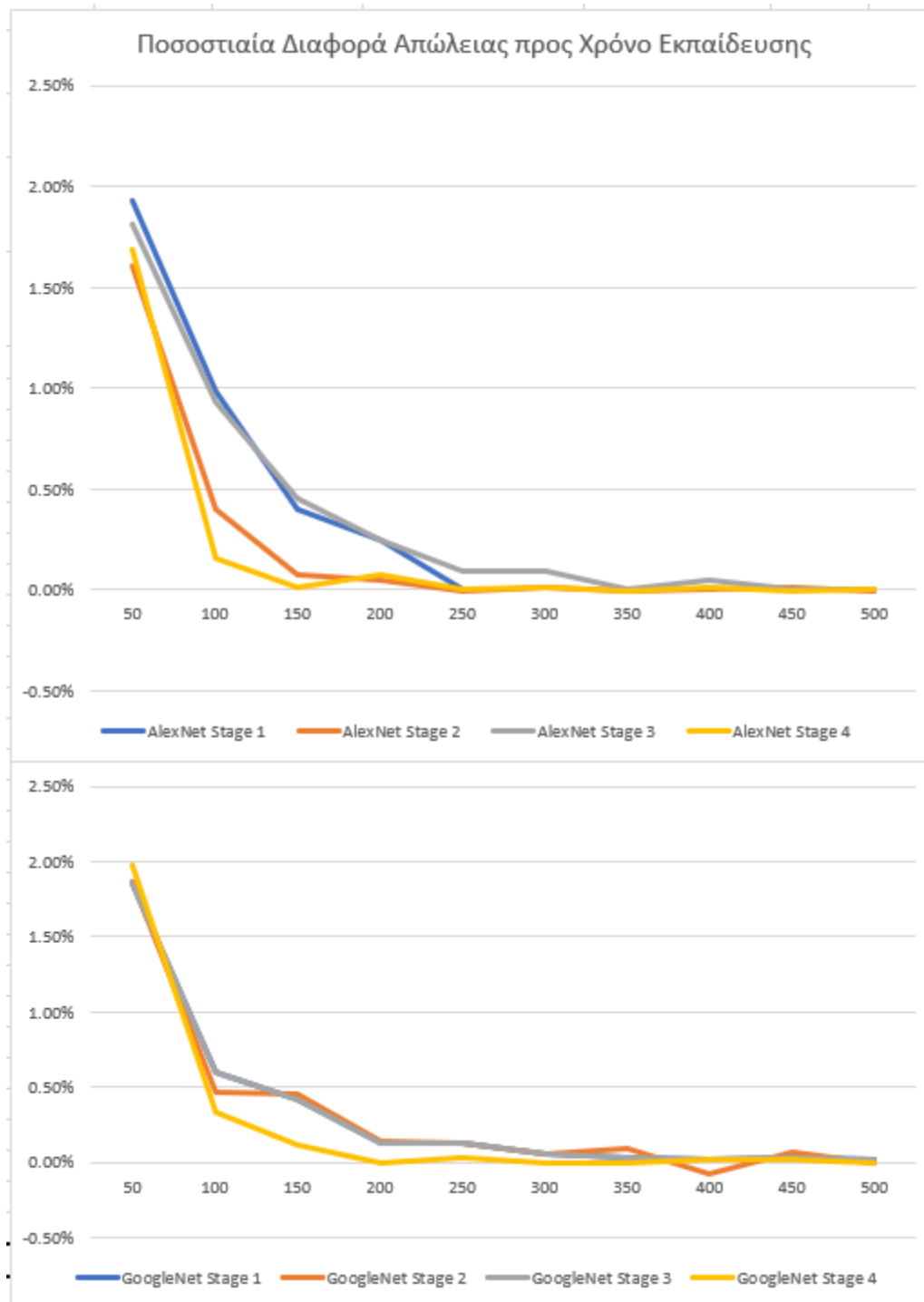
Εικόνα 40 - Κατάβαση Απώλειας σε 500 εποχές

Στον Πίνακα 9 παρουσιάζεται ο λόγος μείωσης της απώλειας προς τη διάρκεια εκπαίδευσης ώστε να εντοπιστεί το σημείο στο οποίο περεταίρω εκπαίδευση είναι ασύμφορη.

Training Duration	AlexNet				GoogLeNet			
	Stage 1	Stage 2	Stage 3	Stage 4	Stage 1	Stage 2	Stage 3	Stage 4
50	1.93%	1.61%	1.81%	1.68%	1.86%	1.86%	1.86%	1.98%
100	0.98%	0.40%	0.93%	0.16%	0.59%	0.47%	0.59%	0.33%
150	0.40%	0.07%	0.45%	0.01%	0.42%	0.45%	0.42%	0.11%
200	0.25%	0.05%	0.25%	0.08%	0.13%	0.14%	0.13%	0.00%
250	0.00%	0.00%	0.09%	0.00%	0.12%	0.13%	0.12%	0.03%
300	0.00%	0.01%	0.10%	0.01%	0.06%	0.06%	0.06%	0.00%
350	0.00%	-0.01%	0.00%	-0.01%	0.04%	0.10%	0.04%	0.00%
400	0.00%	0.00%	0.05%	0.01%	0.02%	-0.08%	0.02%	0.02%
450	0.00%	0.01%	0.00%	-0.01%	0.04%	0.07%	0.04%	0.02%
Training Duration	Stage 1		Stage 2		Stage 3		Stage 4	
	AlexNet	GoogLeNet	AlexNet	GoogLeNet	AlexNet	GoogLeNet	AlexNet	GoogLeNet
50	1.93%	1.86%	1.61%	1.86%	1.81%	1.86%	1.68%	1.98%
100	0.98%	0.59%	0.40%	0.59%	0.93%	0.59%	0.16%	0.33%
150	0.40%	0.42%	0.07%	0.42%	0.45%	0.42%	0.01%	0.11%
200	0.25%	0.13%	0.05%	0.13%	0.25%	0.13%	0.08%	0.00%
250	0.00%	0.12%	0.00%	0.12%	0.09%	0.12%	0.00%	0.03%
300	0.00%	0.06%	0.01%	0.06%	0.10%	0.06%	0.01%	0.00%
350	0.00%	0.04%	-0.01%	0.04%	0.00%	0.04%	-0.01%	0.00%
400	0.00%	0.02%	0.00%	0.02%	0.05%	0.02%	0.01%	0.02%
450	0.00%	0.04%	0.01%	0.04%	0.00%	0.04%	-0.01%	0.02%

Πίνακας 9 – Πίνακας ποσοστιαίας διαφοράς Απώλειας προς Χρόνο Εκπαίδευσης

Όπως προκύπτει από τα παραπάνω και εστιάζοντας στην κάλυψη της ανάγκης για εκπαίδευση πυκνής απόδοσης λόγω του αριθμού των πειραμάτων που ήρθαν σε πέρας, επιλέγεται προς χρήση ο αριθμός των 200 εποχών ως διάρκεια εκπαίδευσης των δικτύων, καθώς παρουσιάζει την μεγαλύτερη απόδοση στον αντίστοιχο πίνακα.



Εικόνα 41 – Σχήμα ποσοστιαίας διαφοράς Απώλειας προς Χρόνο Εκπαίδευσης

5.3.2 Ρυθμός Μεταβολής Μάθησης

Όπως έχει ήδη αναπτυχθεί στο παρών κεφάλαιο, ο ρυθμός μεταβολής της μάθησης (Learning Rate) επηρεάζει το μέγεθος της μεταβολής της τιμής των βαρών που λαμβάνει χώρα σε κάθε εποχή εκπαίδευσης. Μπορεί να γίνει ευκολά αντιληπτή η σημασία της σωστής ρύθμισής του, λαμβάνοντας υπόψιν τον βασικό στόχο της εκπαίδευσης του δικτύου, τη μεγαλύτερη δυνατή μείωση της απώλειας.

Αν απεικονίσουμε την τιμή της απώλειας ως προς την τιμή δυνατών βαρών και bias ενός και μόνο νευρώνας μπορούμε να αντιληφθούμε ως στόχο τις κοιλάδες που προκύπτουν, τις τιμές weight και bias για τις οποίες η απώλεια ελαχιστοποιείται. Καθώς ξεκινάμε την αναζήτηση από ένα τυχαίο σημείο αναφοράς και βηματίζουμε αναζητώντας τα τοπικά ελάχιστα, ο ρυθμός μεταβολής αποτελεί το μέγεθος του βήματος μας. Με μεγάλο βήμα θα κινούμαστε με γρήγορο ρυθμό στην επιφάνεια αλλά είναι πιθανό πως θα υπερπηδούμε αντί να εντοπίζουμε τα ελάχιστα που αναζητούμε. Αντίθετα με πολύ μικρό βήμα ο χρόνος αναζήτησης αυξάνεται και υπάρχει η πιθανότητα να εγκλωβιστούμε σε μικρές κοιλάδες αγνοώντας ικανοποιητικότερα αποτελέσματα.

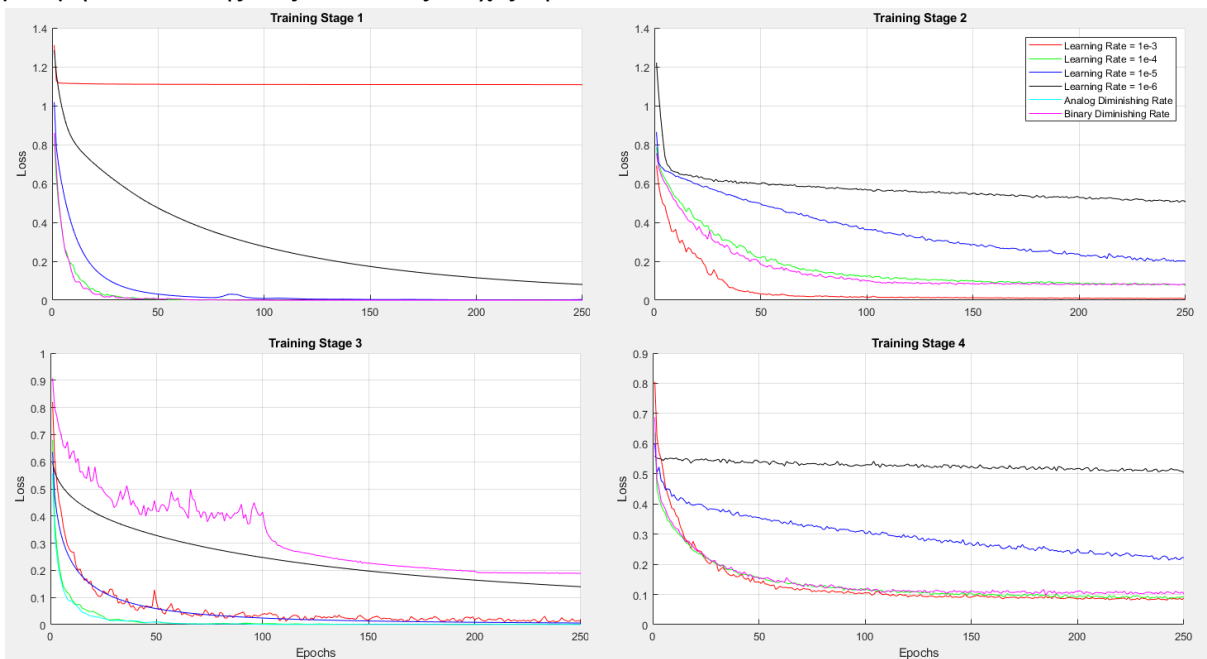
Στην Εικόνα 42 παρουσιάζεται η καμπύλη απώλειας κατά τα τέσσερα στάδια της εκπαίδευσης του AlexNet για 250 εποχές με τέσσερις διαφορετικούς ρυθμούς εκπαίδευσης που παρέμειναν σταθεροί. Κάθε σημείο αντιπροσωπεύει τη μέση τιμή απώλειας για κάθε εποχή. Από το σχήμα μπορεί να παρατηρηθεί πως:

- Το μεγαλύτερο μέρος της εκπαίδευσης έχει ολοκληρωθεί μέχρι την εποχή 100 και έκτοτε τα δίκτυα δεν συγκλίνουν ικανοποιητικά.
- Σε αυτές τις 100 εποχές μόνο οι δύο μεγαλύτεροι ρυθμοί έχουν φέρει ικανοποιητικό αποτέλεσμα με τους δύο μικρότερους να μην έχουν επαρκή αριθμό εποχών ώστε να συγκλίνουν με τους υψηλότερους ρυθμούς.
- Οι υψηλοί ρυθμοί, ενώ συγκλίνουν γρηγορότερα, έχουν την τάση, σε περιπτώσεις, να αποκλίνουν απότομα όπως παρατηρούμε στη πρώτη περίοδο άρα δεν είναι ιδανικοί.

Από τα παραπάνω προκύπτει τα ικανοποιητικότερα αποτελέσματα εμφανίζονται με την χρήση μεταβαλλόμενου ρυθμού μάθησης, αξιοποιώντας έναν υψηλό ρυθμό για μικρό διάστημα ώστε να εμφανιστεί γρήγορη σύγκλιση αλλά και τη ταυτόχρονη μείωσή του με στόχο τη μεγαλύτερη μείωση απώλειας. Για την υλοποίηση αυτού του σχήματος ελέγχονται δύο μέθοδοι:

1. Ορίζεται πως ανά τακτικά διαστήματα, έστω 100 εποχές, ο ρυθμός θα υποδεκαπλασιάζεται.
2. Εναλλακτικά ο ρυθμός μειώνεται σε μικρότερα διαστήματα ώστε να εμφανιστεί πιο διαβαθμισμένη κατάβαση.

Από την Εικόνα 42 είναι προφανές πως τα καλύτερα αποτελέσματα εμφανίζονται από τις μεθόδους που περιλαμβάνουν τις τιμές $e-3$, $e-4$ και τη σταδιακή μείωση του ρυθμού μάθησης. Με στόχο την αποφυγή του κινδύνου απόκλισης που παρουσιάζει η τιμή $e-3$ τοποθετείται η έναρξη της κατάβασης στο υποδιπλάσιό της, $5e-4$ και ο ρυθμός μείωσης ορίζεται στο 10% της τρέχουσας τιμής ανά 10 εποχές με τιμή εκπαίδευσης στις τελευταίες εποχές τη $6.75e-5$.



Εικόνα 42 – Συγκεντρωτικό σχήμα σχέσης Loss με learning rate

5.3.3 Τομή-προς-Ένωση Θετικών-Αρνητικών Δειγμάτων

Τελευταία υπερπαράμετρος που θα παρουσιαστεί στο παρών κεφάλαιο είναι η τομή ως προς την ένωση θετικών και αρνητικών δειγμάτων. Η εντολή `trainFasterRCNNObjectDetector` επιτρέπει τον ορισμό του ποσοστού επικάλυψης της πρότασης περιοχής (region proposal) με τον πίνακα αλήθειας του στόχου (ground truth) που είναι απαραίτητο για να θεωρηθεί ένα αποτέλεσμα σαν θετικό ή αρνητικό. Οι προρυθμισμένες τιμές για τη συγκεκριμένη παράμετρο είναι τιμή επικάλυψης μικρότερη του 50% όπου θεωρείται το δείγμα αρνητικό ενώ με μεγαλύτερη θετικό. Είναι σημαντικό εδώ να επισημανθεί πως, καθώς το δίκτυο δημιουργεί καταλόγους θετικών παραδειγμάτων ξεχωριστά από τους καταλόγους αρνητικών, ένα παράδειγμα μπορεί να καταγραφεί και στους δύο καταλόγους.

Από τον ορισμό συμπεραίνεται πως το μεγάλο εύρος στον ορισμό των θετικών δειγμάτων θα επιτρέψει τη μεγαλύτερη ακρίβεια στην πρόβλεψη, ενώ αντίθετα μικρό εύρος στα αρνητικά παραδείγματα θα δίνει την εικόνα περισσότερων λάθος θετικών (False Positives) διαγνώσεων. Στον Πίνακα 10 παρουσιάζονται εννέα πειραματισμοί με ενδεικτικές τιμές παραμέτρου ώστε να εντοπιστεί η σχέση ανάμεσα στο εύρος των τιμών και την ακρίβεια των δικτύων που παράγουν, χρησιμοποιώντας τα όρια των 0,1 0,5 και 0,9 σε συνδυασμό μεταξύ τους τόσο για τα θετικά όσο και για τα αρνητικά παραδείγματα.

Η εκπαίδευση πραγματοποιείται για 100 εποχές με αρχική τιμή μάθησης $5e-4$ και μειώνεται κατά 10% κάθε 10 εποχές όπως παρουσιάστηκε στην προηγούμενη ενότητα. Αφού εκπαιδευτούν τα δίκτυα, οι παραγόμενοι detectors τροφοδοτούνται με το σύνολο του σετ ελέγχου και τα αποτελέσματά τους καταγράφονται και αξιολογούνται με τη μέθοδο που θα παρουσιάστηκε εκτενέστερα στο Κεφάλαιο 3.

Negative Overlap Range	Positive Overlap Range	Benign Average Precision	Malignant Average Precision	Benign Average Miss Rate	Malignant Average Miss Rate
[0 0.1]	[0.1 1]	0.001	0.075	1.000	0.995
[0 0.1]	[0.5 1]	0.000	0.037	0.991	0.888
[0 0.1]	[0.9 1]	0.000	0.009	1.000	0.928
[0 0.5]	[0.1 1]	0.003	0.057	0.995	0.963
[0 0.5]	[0.5 1]	0.000	0.012	0.983	0.896
[0 0.5]	[0.9 1]	0.000	0.000	1.000	0.954
[0 0.9]	[0.1 1]	0.010	0.055	1.000	1.000
[0 0.9]	[0.5 1]	0.000	0.025	0.972	0.891
[0 0.9]	[0.9 1]	0.000	0.000	1.000	0.940

Πίνακας 10 – Πίνακας αποτελεσμάτων Τομής ως προς Ένωσης

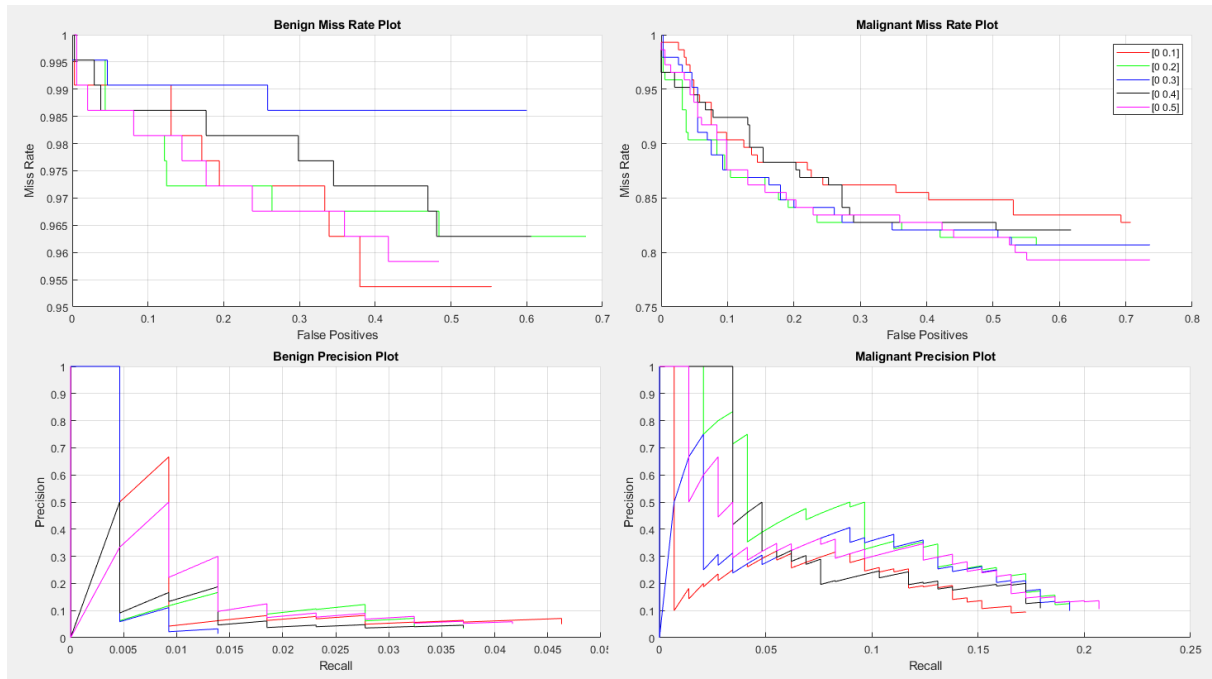
Από τον Πίνακα 10 αποκτά μορφή η σχέση της σημαντικότερης επιρροής στην ακρίβεια, η οποία είναι ο ορισμός μεγάλου εύρους ως τιμή Positive Overlap, ωστόσο προκύπτει αντίστοιχα πως οι χαμηλότερες τιμές Miss Rate προέρχονται από το εύρος [0.5 1], γεγονός που υποδεικνύει πως απαιτείται να συμβιβασμός της ανάγκη για εντοπισμό όλων των καρκινωμάτων με την τάση του συστήματος να εντοπίζει False Positives. Για τις ανάγκες της παρούσης εργασίας ο πρωτεύον στόχος είναι η παραγωγή του ασφαλέστερου δυνατού συμπεράσματος και αξιολογείται πως βασικό κριτήριο αποτελεί η ο εντοπισμός του μεγαλύτερου αριθμού αντικειμένων συνεπικουρούμενος από την ακρίβεια του δικτύου και με τριτεύουσα σημασία στο ρυθμό λάθους.

Από τα μέχρι τώρα παρουσιαζόμενα δεδομένα είναι εμφανές πως υπάρχει η τάση με την αύξηση του εύρους του Negative Overlap να αυξάνεται η ακρίβεια για τους καλοήθεις όγκους και να μειώνεται ταυτόχρονα για τους κακοήθεις, ωστόσο αυτό μπορεί να οφείλεται στον μικρό αριθμό δειγμάτων. Το συμπέρασμα προς τη βέλτιστη τιμή για το Negative Overlap προκύπτει από τη διενέργεια επιπλέον πειραμάτων τα αποτελέσματα των οποίων παρατίθενται στον Πίνακα 11. Ποιο συγκεκριμένα διενεργήθηκαν πέντε επιπλέον πειράματα, διατηρώντας τις ίδιες παραμέτρους με εξαίρεση την τιμή του Negative Overlap η οποία κυμάνθηκε σε εύρος τιμών από 0.1 έως και 0.5. Τα αποτελέσματα απορρίπτουν την υπόθεση που διατυπώθηκε περί αντίστροφης σχέσης ακρίβειας καλοηθών και κακοηθών αποτελεσμάτων.

Negative Overlap Range	Positive Overlap Range	Benign Average Precision	Malignant Average Precision	Benign Average Miss Rate	Malignant Average Miss Rate
[0 0.1]	[0.1 1]	0.0076	0.0733	0.9788	0.9120
[0 0.2]	[0.1 1]	0.0080	0.0449	0.9799	0.8815
[0 0.3]	[0.1 1]	0.0081	0.0882	0.9907	0.8886
[0 0.4]	[0.1 1]	0.0053	0.0638	0.9819	0.8976
[0 0.5]	[0.1 1]	0.0051	0.0711	0.9768	0.8839

Πίνακας 11 – Αποτελέσματα διαφορετικών τιμών Negative Overlap

Είναι εμφανές πως μια σχετικά μικρή τιμή παρουσιάζει ικανότερα αποτελέσματα από υψηλότερες με ξεκάθαρη υπεροχή να παρουσιάζει η τιμή 0.3 με τον υψηλότερο μέσο όρο ακρίβειας.



Εικόνα 43 – Σχήμα απόλειας τιμών Negative Overlap

Σε αυτό το σημείο και πρώτου γίνει μετάβαση στο επόμενο κεφάλαιο όπου παρουσιάζονται οι μέθοδοι αξιολόγησης που χρησιμοποιήθηκαν για την εξαγωγή των συμπερασμάτων που έχουν διατυπωθεί παρατίθενται λίγες γραμμές κώδικα στις οποίες συμπεκνώνονται οι παράμετροι που τυποποιήθηκαν στο παρών κεφάλαιο. Όπου layers τα στρώματα του δικτύου προς υλοποίηση και trainingData ο πίνακας αλήθειας εκπαίδευσης που έχει προκύψει από το Κεφάλαιο 2.3. Η παράμετρος MiniBatchSize ορίζεται σαν 1 από τις απαιτήσεις του αλγορίθμου εκπαίδευσης.

```
options = trainingOptions('sgdm', ...
    'LearnRateSchedule','piecewise', ...
    'InitialLearnRate', 5e-4, ...
    'LearnRateDropFactor', 0.9, ...
    'LearnRateDropPeriod', 10, ...
    'MaxEpochs', 200, ...
    'MiniBatchSize', 1);

[detector info] = trainFasterRCNNObjectDetector(trainingData, ...
    layers, options, ...
    'NegativeOverlapRange', [0 0.3], ...
    'PositiveOverlapRange', [0.1 1])
```

6 Επιπλέον Πειραματισμοί σε Δομές και Παραμέτρους

Στο παρόν κεφάλαιο παρατίθενται πειράματα μετασχηματισμών στη δομή των δικτύων AlexNet και GoogLeNet και η απόδοση των αντίστοιχων αποτελεσμάτων τους. Εμφανίζονται συγκριτικοί πίνακες που περιλαμβάνουν το σύνολο των πειραμάτων που έχουν παρουσιαστεί, από τα οποία επιλέγονται τα τρία ικανότερα ώστε να μελετηθούν περαιτέρω στο επόμενο κεφάλαιο.

6.1 Πειράματα Δομών Αρχιτεκτονικής

Αρχικά παρουσιάζονται επεμβάσεις στις δημοφιλείς δομές που περιεγράφηκαν στο Κεφάλαιο 4.2 και τα αποτελέσματα που παρήχθησαν.

6.1.1 Πειραματισμοί σε βάση δικτύου AlexNet

Στην ενότητα 4.2.2 παρατίθενται τα αποτελέσματα ενός Fully Connected δικτύου βασισμένου στη δομή του Alexnet όπως έχει ήδη παρουσιαστεί. Ένα επιπλέον πείραμα έχει να κάνει με την αφαίρεση των στρωμάτων 10-16 (τρία στρώματα συνέλιξης και τρεις relu) του δικτύου ώστε να κριθεί η απόδοση ενός ρηχότερου δικτύου.



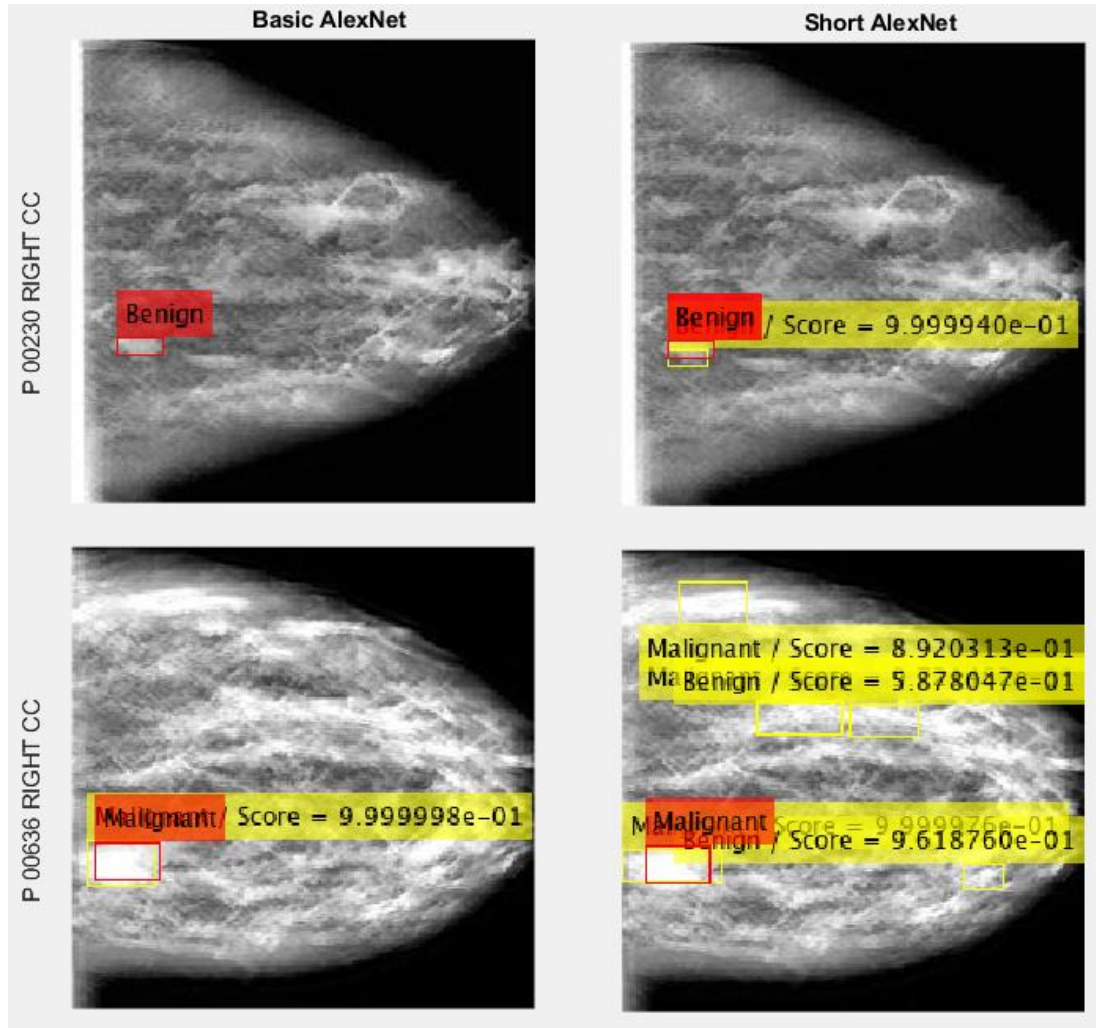
Εικόνα 44 – Δομή Short AlexNet

Το δίκτυο που προκύπτει εμφανίζει εξαιρετική απόδοση τόσο στην ακρίβεια του, ιδιαίτερα όσον αφορά κακοήθης όγκους, αλλά κυρίως μέσω της εύρεσης αντικειμένου ενδιαφέροντος σχεδόν στο σύνολο των διαθέσιμων εικόνων. Αρκεί να αναφερθεί πως το αρχικό δίκτυο Alexnet κατάφερε να εντοπίσει αντικείμενα σε 208 εικόνες, σε σύνολο 345, ενώ το περιτμημένο δίκτυο εντόπισε αντικείμενα σε 311 εικόνες. Η παρατήρηση των αποτελεσμάτων του υποδεικνύει την τάση του δικτύου να εντοπίζει αρκετά λάθος θετικά σε κάθε εικόνα, ωστόσο καταφέρνει αντίστοιχα να εντοπίσει μεγάλο αριθμό των πραγματικών αντικειμένων που απεικονίζονται στις μαστογραφίες, όπως προκύπτει και από την Εικόνα 45.

Network Base	Benign Average Precision	Malignant Average Precision	Benign Average Miss Rate	Malignant Average Miss Rate	Ποσοστό εικόνων με εύρεση ROI
Short AlexNet	0.0125	0.1618	0.9664	0.8103	90.14%

Πίνακας 12 – Αποτελέσματα δομής Short AlexNet

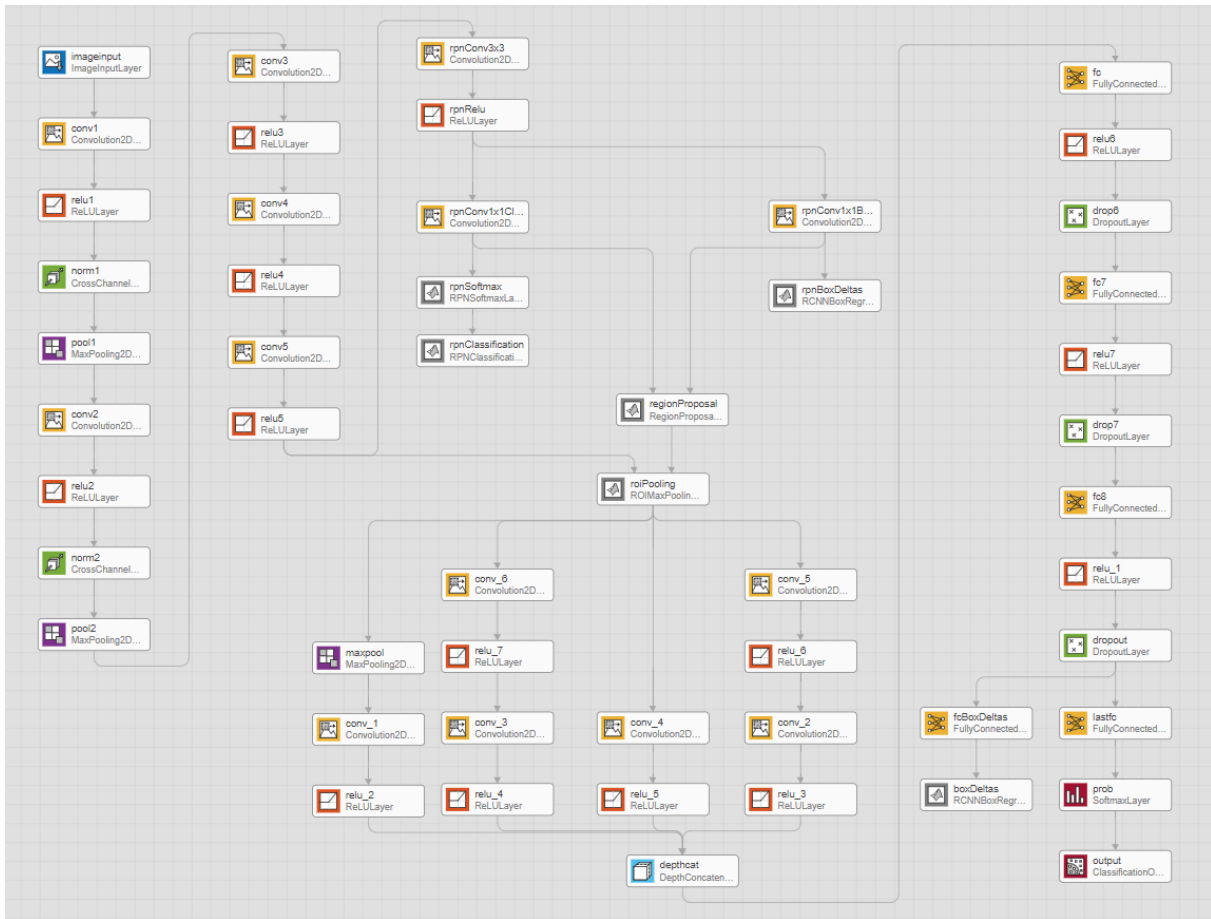
Στην συνέχεια παραθέτουμε τα αποτελέσματα μιας σειράς πειραμάτων με σκοπό τη μετάγγιση στρωμάτων Inception προερχόμενα από το δίκτυο GoogLeNet και την ενσωμάτωσή τους στα υπάρχοντα δίκτυα AlexNet και Short AlexNet. Πραγματοποιήθηκαν συνολικά τέσσερα πειράματα που περιλαμβάνουν την εισαγωγή των Inception layers 5a στις δομές AlexNet και Short AlexNet και 5a και 5b στα ίδια δίκτυα. Χάριν ευκολίας στα δίκτυα που παρήχθησαν με την εισαγωγή μόνο των στρωμάτων 5a θα αναφερόμαστε με την ονομασία Inception AlexNet και στα αντίστοιχα που περιλαμβάνουν τα στρώματα 5a και 5b Inception AlexNet+. Ενδεικτικά παρουσιάζονται οι δομές των δικτύων Inception AlexNet (Εικόνα 46) και Short Inception AlexNet+ (Εικόνα 47)



Εικόνα 45 – Διαφορά αποτελεσμάτων AlexNet / Short AlexNet

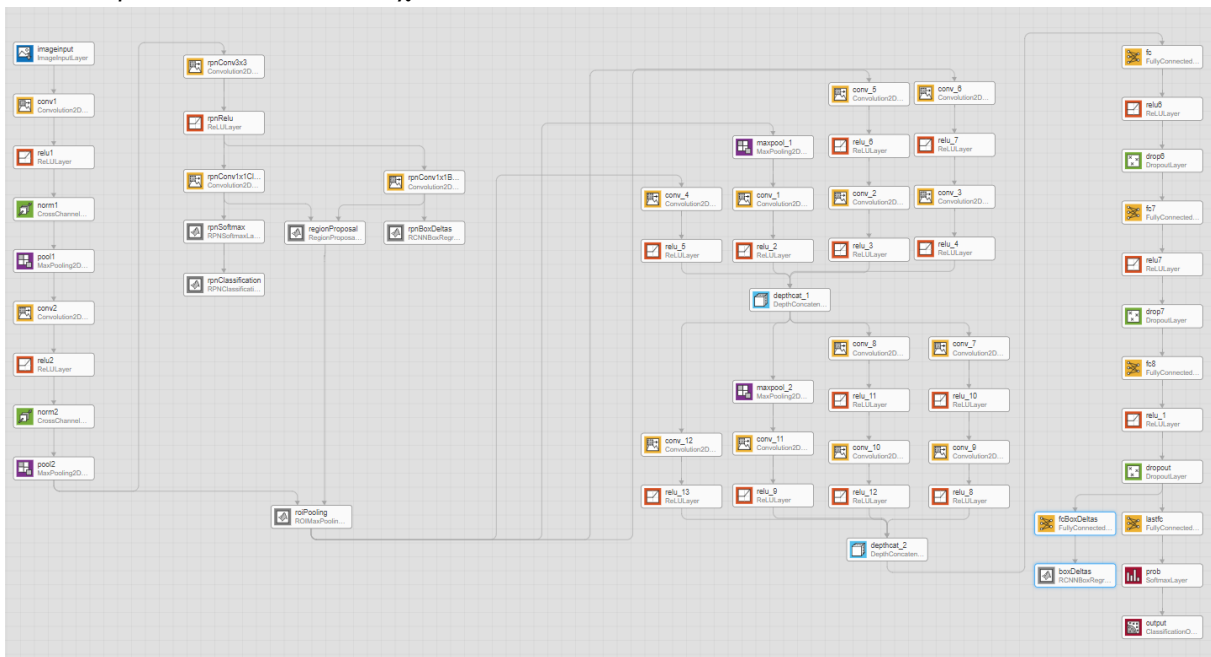
Network Base	Benign Average Precision	Malignant Average Precision	Benign Average Miss Rate	Malignant Average Miss Rate	Ποσοστό εικόνων με εύρεση ROI
Inception AlexNet	0.0011	0.0570	0.9923	0.8982	47.83%
Inception AlexNet+	0.0024	0.0541	0.9871	0.9054	68.12%
Short Inception AlexNet	0.0010	0.0331	0.9912	0.9312	80.00%
Short Inception AlexNet+	0.0010	0.0236	0.9922	0.9427	80.29%

Πίνακας 13 – Αποτελέσματα δομών Inception AlexNet



Εικόνα 46 – Δομή Inception AlexNet

Τα αποτελέσματα που προκύπτουν δεν είναι ενθαρρυντικά. Η μειωμένη απόδοση αποδίδεται στην πληθώρα των μεταβλητών που εισάγονται (κάθε αλληλουχία Inception layers αποτελείται από 177.000 μεταβλητές που απαιτούν εκπαίδευσης) συναρτήσει το μικρού χρόνου εκπαίδευσης και των περιορισμένου αριθμού δειγμάτων που διαχειρίζονται τα πειράματα που παρατίθενται. Για τις ανάγκες τις εργασίας τα πιο ενθαρρυντικά αποτελέσματα εμφανίζονται από το δίκτυο Short AlexNet στο οποίο και θα παρατεθούν επιπλέον στοιχεία.

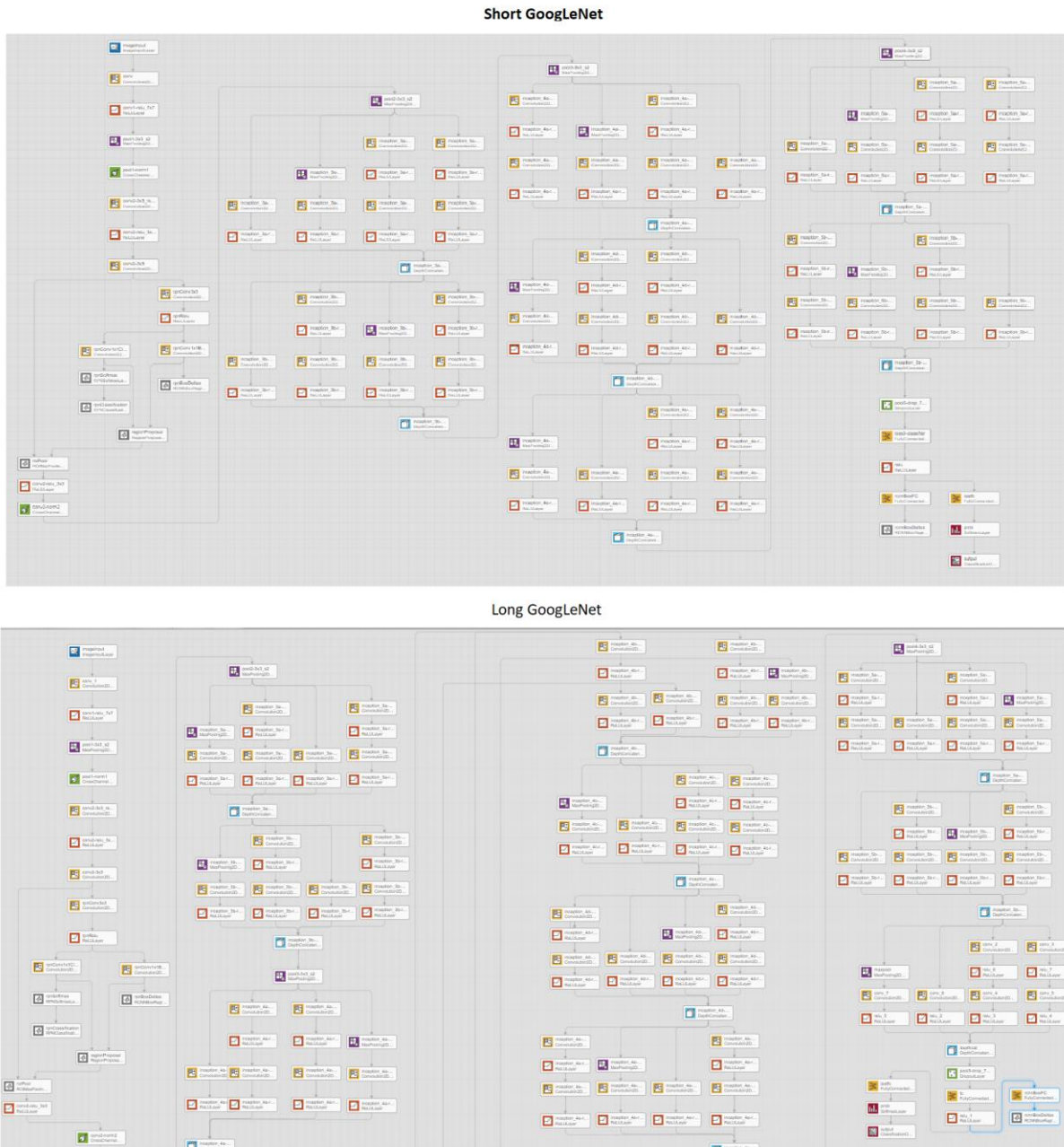


Εικόνα 47 – Δομή Short Inception AlexNet+

6.1.2 Πειραματισμοί σε βάση δικτύου GoogLeNet

Το δίκτυο GoogLeNet έχει σημειώσει την μεγαλύτερη απόδοση από τα βασικά δίκτυα. Παρατήρηση της δομής του υποδεικνύει πως είναι δύσκολη η επέμβαση στα Inception layers λόγω της ιδιαίτερης αρχιτεκτονικής τους. Τα πειράματα που παρατίθενται μπορούν να διακριθούν σε δύο κατηγορίες.

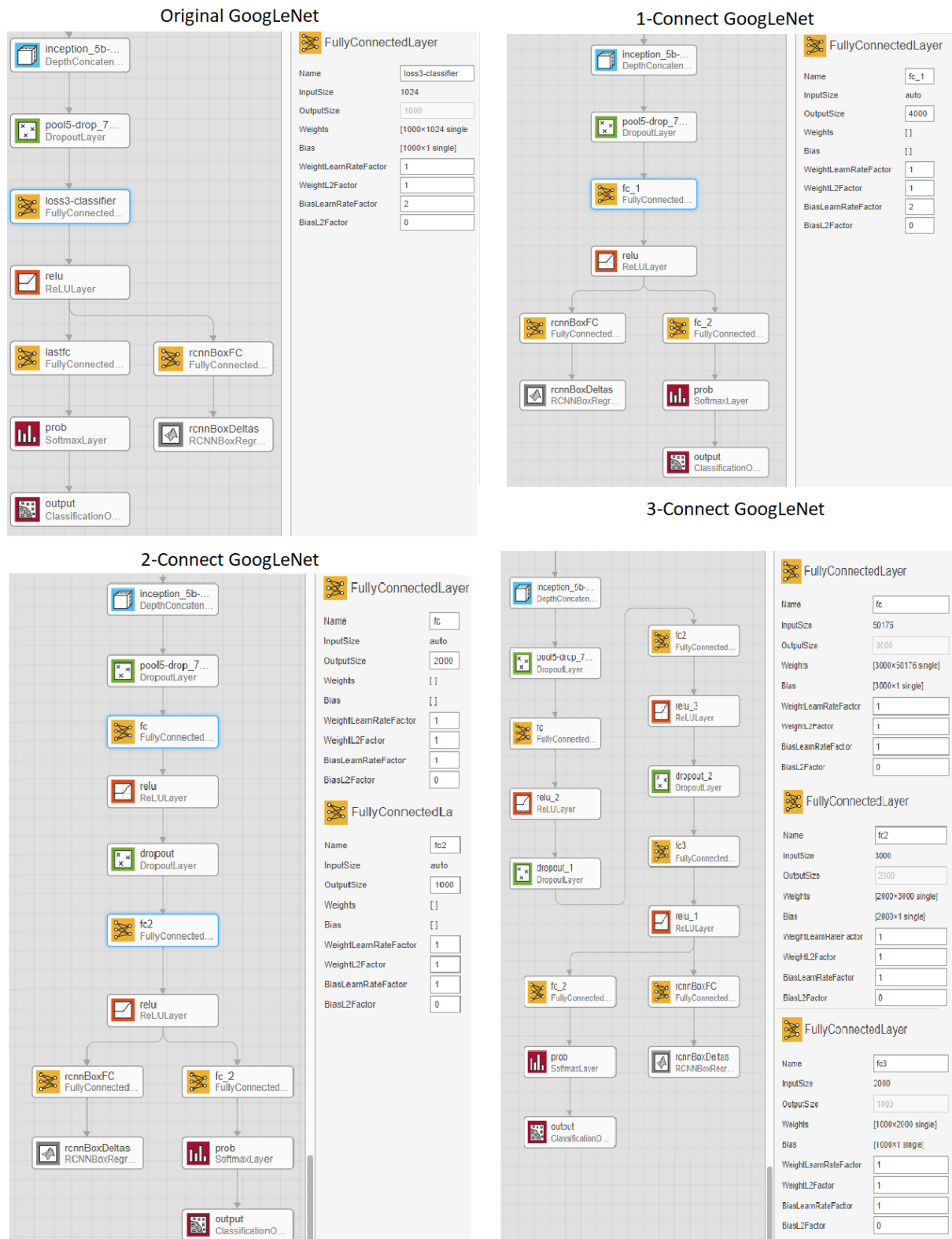
Η πρώτη σειρά πειραμάτων αφορά τη μορφολογία των στρωμάτων GoogLeNet, τα οποία περιέχουν εννέα διασδιάστατες δομές Inception. Πραγματοποιήθηκαν δύο σχετικοί πειραματισμοί όπου στον πρώτο αφαιρέθηκαν οι δύο τελευταίες δομές ενώ στον δεύτερο προστέθηκαν αντίστοιχα δύο επιπλέον δομές στο τέλος της ακολουθίας. Χάριν ευκολίας το περιτμημένο δίκτυο θα παρατεθεί ως Short GoogLeNet ενώ το επαυξημένο ως Long GoogLeNet όπως και εμφανίζονται στην Εικόνα 48.



Εικόνα 48 – Δομές Short & Long GoogLeNet

Η δεύτερη σειρά πειραμάτων αφορά την αντικατάσταση του πλήρως συνδεδεμένου στρώματος των 1000 νευρώνων από τρεις δομές πλήρως συνδεδεμένων δικτύων. Η πρώτη δομή αντικαθιστά το υπάρχον στρώμα 1000 νευρώνων από ένα μεγαλύτερο 4000 νευρώνων, η δεύτερη με δύο στρώματα από 2000 και 1000 νευρώνες και η τρίτη με τρία στρώματα 3000, 2000 και 1000 νευρώνων όπως παρουσιάζεται στην Εικόνα 49.

Πτυχιακή Εργασία, τμήμα Μηχανικών Πληροφορικής
75. Αναγνώριση καρκινώματος / Σύνταπος Δημήτριος 2051



Εικόνα 49 – Δομές X-Connect GoogLeNet

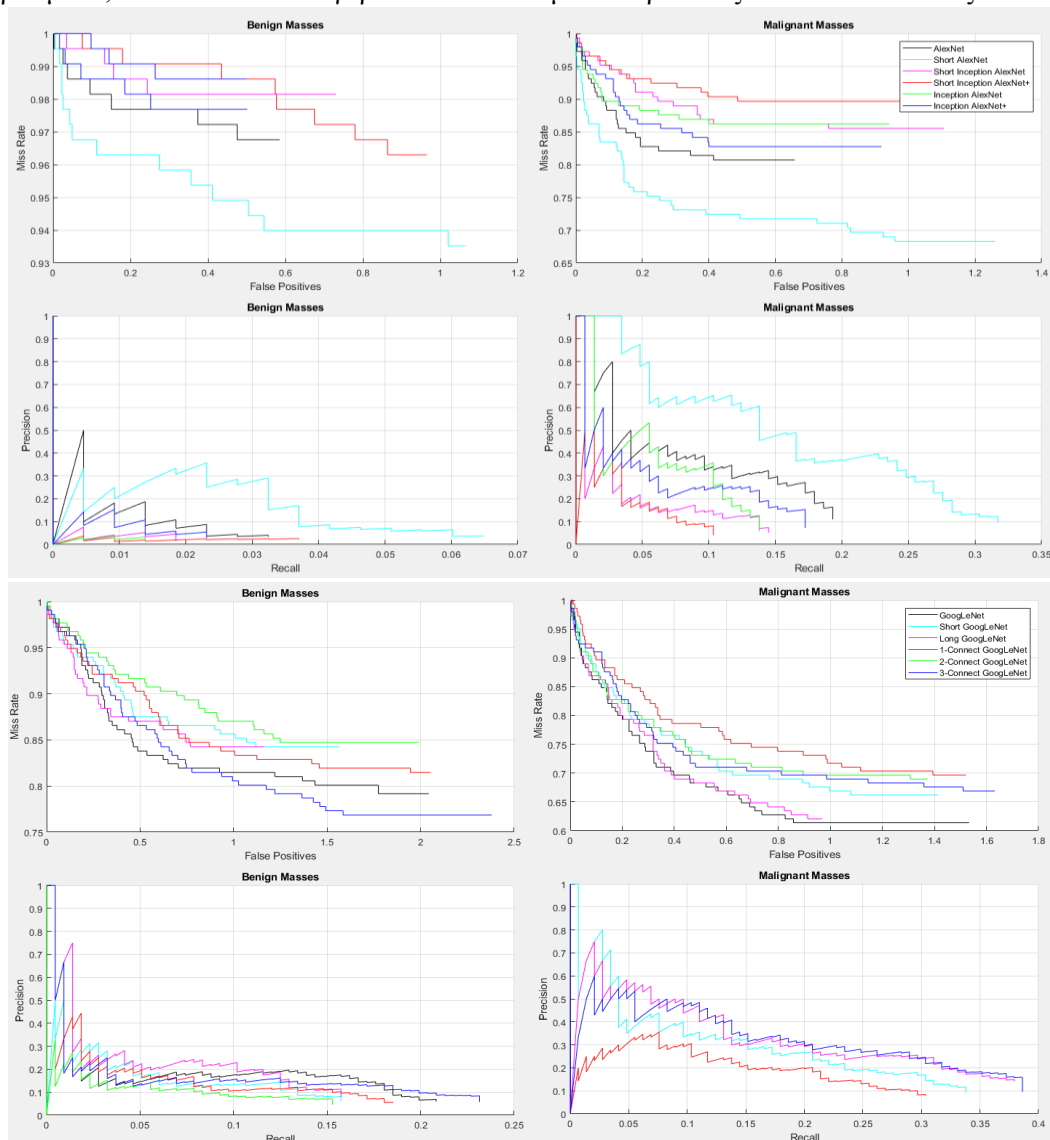
Τα δίκτυα που παρουσιάστηκαν διατηρούν κατά τα άλλα τη βασική δομή του GoogLeNet και παρατίθενται ώστε να σημειωθεί η επιρροή που παρουσιάζουν δομικές αλλαγές στην απόδοση των δικτύων και τα αποτελέσματά τους εμφανίζονται στον Πίνακα 14. Παρατηρούνται ικανοποιητικά αποτελέσματα για τη δομή Long GoogLeNet ενώ οι υπόλοιπες δομές δεν εμφανίζουν αντίστοιχη απόδοση. Εκτενέστερη αναφορά στα αποτελέσματα των πειραμάτων και παρουσίαση των σχετικών γραφημάτων πραγματοποιείται στην επόμενη ενότητα.

Network Base	Benign Average Precision	Malignant Average Precision	Benign Average Miss Rate	Malignant Average Miss Rate	Ποσοστό εικόνων με εύρεση ROI
Short GoogLeNet	0.0297	0.1108	0.9402	0.8463	95.07%
Long GoogLeNet	0.0407	0.1321	0.9311	0.8226	87.83%
1-Connect GoogLeNet	0.0302	0.0656	0.9405	0.8808	98.55%
2-Connect GoogLeNet	0.0189	0.1007	0.9564	0.8523	97.10%
3-Connect GoogLeNet	0.0400	0.1003	0.9356	0.8531	98.84%

Πίνακας 14 – Αποτελέσματα πειραμάτων δομών GoogLeNet

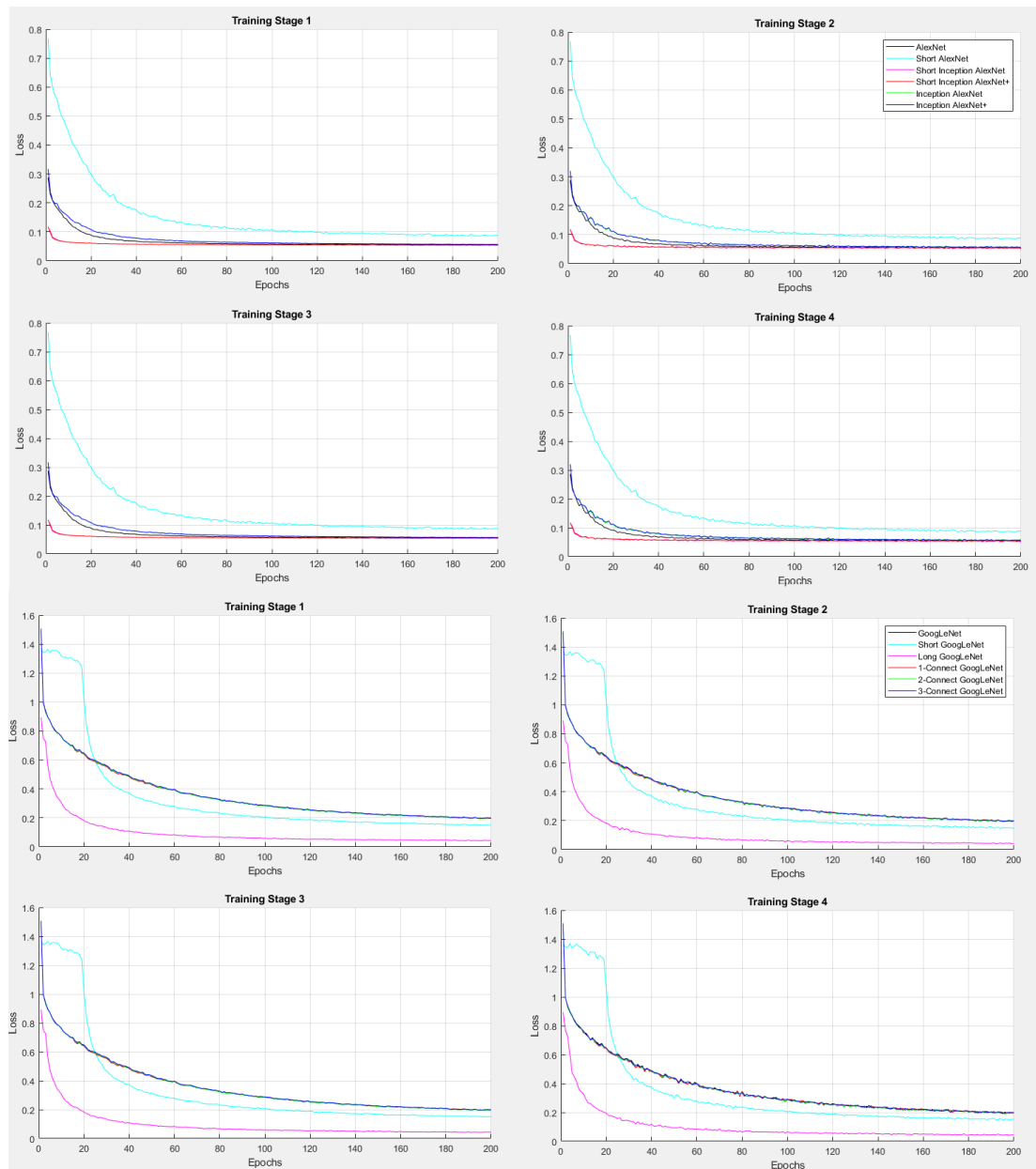
6.1.3 Σύνοψη αποτελεσμάτων πειραματικών δομών

Σε αυτή την ενότητα παρατίθενται τα συνολικά αποτελέσματα των πειραμάτων που έχουν αναλυθεί μέχρι αυτό το σημείο της εργασίας. Πιο συγκεκριμένα στην Εικόνα 50 και εμφανίζονται τα γραφικά των πειραμάτων που παρουσιάστηκαν στο παρόν κεφάλαιο από τα οποία εξάγονται ενδιαφέροντα συμπεράσματα, ενώ αναλυτικά συγκριτικά αποτελέσματα παρουσιάζονται στον Πίνακα 15.



Εικόνα 50 – Αποτελέσματα πειραματικών δομών

- Είναι εμφανής η απόδοση των αποτελεσμάτων του δικτύου Short AlexNet σε σχέση με τα συγγενή του δίκτυα, ξεπερνώντας και την απόδοση του βασικού δικτύου AlexNet. Το αποτέλεσμα αυτό έρχεται σε ρήξη με τη θέση των δημιουργών του δικτύου, πως μείωση των στρωμάτων επιφέρει ανάλογη μείωση δυνατοτήτων, όπως παρατέθηκε και στην ενότητα 4.2.1. Η απρόβλεπτη αύξηση της απόδοσης μπορεί να αποδοθεί σε δύο γεγονότα:
 - Ο μειωμένος αριθμός μεταβλητών προκάλεσε «σύγχυση» στο δίκτυο οδηγώντας το σε εντοπισμό λανθασμένων αντικειμένων όπως υποσημειώθηκε και από την Εικόνα 45, τα οποία αντισταθμίστηκαν, ωστόσο, από τον εντοπισμό σωστών αντικειμένων σε μεγαλύτερο αριθμό από τα συναφή του δίκτυα. Το συμπέρασμα αυτό υποδεικνύεται και από την Εικόνα 51 όπου εμφανίζεται αυξημένη απώλεια για την ίδια περίοδο εκπαίδευσης.
 - Επιπρόσθετα πρέπει να υπενθυμισθεί πως το δίκτυο AlexNet δεν σχεδιάστηκε για εντοπισμό αντικειμένων σε εικόνα αλλά για κατηγοριοποίηση κεντρισμένου αντικειμένου σε εικόνα. Στην παρούσα εργασία, το μικρότερο δίκτυο περιείχε φίλτρα συνέλιξης επαρκούς μεγέθους για τον εντοπισμό των αντικειμένων ενδιαφέροντος, γεγονός το οποίο σε συνδυασμό με την πρώτη θέση επέτρεψε τα αποτελέσματα αυτά.



Εικόνα 51 – Απόλεια κατάβασης πειραμάτων

- Παρατηρείται μη ικανοποιητική απόδοση από τα δίκτυα στα οποία ενσωματώθηκαν δομές Inception. Σε μεταξύ τους σύγκριση προκύπτει πως οι δομές που βασίζονται στο πλήρες δίκτυο AlexNet έχουν περιορισμένη εφαρμογή καθώς εντόπισαν αντικείμενο ενδιαφέροντος σε μικρό αριθμό εικόνων, ενώ τα δίκτυα που βασίζονται στη Short AlexNet αρχιτεκτονική εντόπισαν αντικείμενο στα τέσσερα πέμπτα των εικόνων, εντοπίζοντας βέβαια, μεγάλο αριθμό λαθών, γεγονός που προκύπτει από τη μικρή απόδοση Average Precision.
- Αναφορικά με τις δομές που βασίστηκαν στην αρχιτεκτονική GoogLeNet, εμφανίζονται δύο διακριτές προσεγγίσεις στην επεξεργασία, με τη μία οικογένεια δομών να αφορά τη Long και Short αρχιτεκτονική και τη δεύτερη τις δομές όπου αντικαταστάθηκε το πλήρως συνδεδεμένο στρώμα, στις οποίες θα αναφερθούμε με τον όρο X-Connect.
 - Η Long GoogLeNet εμφανίζει τα ικανότερα αποτελέσματα, με απόδοση που υπερβαίνει του βασικού δικτύου. Η αύξηση της ακρίβειας πηγάζει από την αύξηση του βάθους του δικτύου και οδηγεί στο συμπέρασμα πως οι δομές Inception εμφανίζουν σε άθροιση ικανότερα αποτελέσματα από τις άλλες βαθιές δομές που μελετήθηκαν όπως τη YOLO και τη ResNet. Από την Εικόνα 51 προκύπτει πως η δομή Long GoogLeNet εμφάνισε την ικανότερη κατάβαση απώλειας από τις συναφής πειραματικές δομές.
 - Από τις X-Connect αρχιτεκτονικές ικανότερα αποτελέσματα εμφάνισε η δομή που περιείχε τα τρία πλήρως συνδεδεμένα στρώματα, ενώ αναντίστοιχη απόδοση εμφάνισε η 1-Connect η οποία διατήρησε την ίδια δομή με το βασικό GoogLeNet διαφοροποιούμενη μόνο στον αριθμό των νευρώνων του σχετικού στρώματος. Η μειωμένη αποτελεσματικότητα αποδίδεται στην απώλεια της μάθησης του αρχικού δικτύου και στην ανάγκη εκπαίδευσης των τιμών πολλαπλασίων νευρώνων.

Network Base	Benign Average Precision	Malignant Average Precision	Benign Average Miss Rate	Malignant Average Miss Rate	Ποσοστό εικόνων με εύρεση ROI
AlexNet	0.0053	0.0811	0.9819	0.8821	60.29%
FCN AlexNet	0.0074	0.0392	0.9814	0.9190	43.18%
GoogLeNet	0.0380	0.1304	0.9312	0.8196	96.81%
YOLO	0.0144	0.0666	0.9732	0.8985	70.72%
ResNet	0.0003	0.0190	0.9959	0.9446	40.00%
Short AlexNet	0.0125	0.1618	0.9664	0.8103	90.14%
Inception AlexNet	0.0011	0.0570	0.9923	0.8982	47.83%
Inception AlexNet+	0.0024	0.0541	0.9871	0.9054	68.12%
Short Inception AlexNet	0.0010	0.0331	0.9912	0.9312	80.00%
Short Inception AlexNet+	0.0010	0.0236	0.9922	0.9427	80.29%
Short GoogLeNet	0.0297	0.1108	0.9402	0.8463	95.07%
Long GoogLeNet	0.0407	0.1321	0.9311	0.8226	87.83%
1-Connect GoogLeNet	0.0302	0.0656	0.9405	0.8808	98.55%
2-Connect GoogLeNet	0.0189	0.1007	0.9564	0.8523	97.10%
3-Connect GoogLeNet	0.0400	0.1003	0.9356	0.8531	98.84%

Πίνακας 15 – Αποτελέσματα πειραματικών δομών

Network Base	Total Average Precision	Total Average Miss Rate
Short AlexNet	0.0872	0.8884
Long GoogLeNet	0.0864	0.8769
GoogLeNet	0.0842	0.8754
Short GoogLeNet	0.0703	0.8933
3-Connect GoogLeNet	0.0702	0.8944
2-Connect GoogLeNet	0.0598	0.9044
1-Connect GoogLeNet	0.0479	0.9107
AlexNet	0.0432	0.9320
YOLO	0.0405	0.9359
Inception AlexNet	0.0291	0.9453
Inception AlexNet+	0.0283	0.9463
FCN AlexNet	0.0233	0.9502
Short Inception AlexNet	0.0171	0.9612
Short Inception AlexNet+	0.0123	0.9675
ResNet	0.0097	0.9703

Πίνακας 16 – Συγκριτική απόδοση πειραματικών δομών

Τελικώς στον Πίνακα 16 παρατίθενται οι μέσοι όροι ακρίβειας και λάθους όλων των δικτύων που έχουν παραχθεί στα μέχρι τώρα πλαίσια της εργασίας. Από την κατάταξή τους με βάρος στην μεγαλύτερη ακρίβεια παρουσιάζονται τα τρία ικανότερα δίκτυα, Short AlexNet, Long GoogLeNet και βασικό GoogLeNet. Τα τρία αυτά δίκτυα, στα οποία έχει γίνει αναφορά στο παρών κεφάλαιο, επιλέγονται για περαιτέρω μελέτη στην ενότητα που ακολουθεί.

6.2 Πειραματισμοί επανεκπαίδευσης δικτύων

Στην προηγούμενη ενότητα παρουσιάστηκαν τα αποτελέσματα του συνόλου των πειραματισμών που πραγματοποιήθηκαν κατά τη διάρκεια της παρούσης εργασίας. Από τις διαθέσιμες δομές, διακρίθηκαν για την απόδοσή τους οι εξής τρεις:

1. Short AlexNet
2. Long GoogLeNet
3. GoogLeNet

Με στόχο τη διεύρυνση της απόδοσής τους και εντοπισμού των εξ αυτών δικτύων με την υψηλότερη ικανότητα μάθησης σε εύρος χρόνου, ο πειραματισμός αφορά την επανεκπαίδευση του detector για μεγαλύτερο χρονικό διάστημα και την σύγκρισή του με τον αρχικό ανιχνευτή όπως και με έναν τρίτο οποίος θα εκπαιδευτεί εξ αρχής για το ίδιο αυξημένο χρονικό διάστημα.

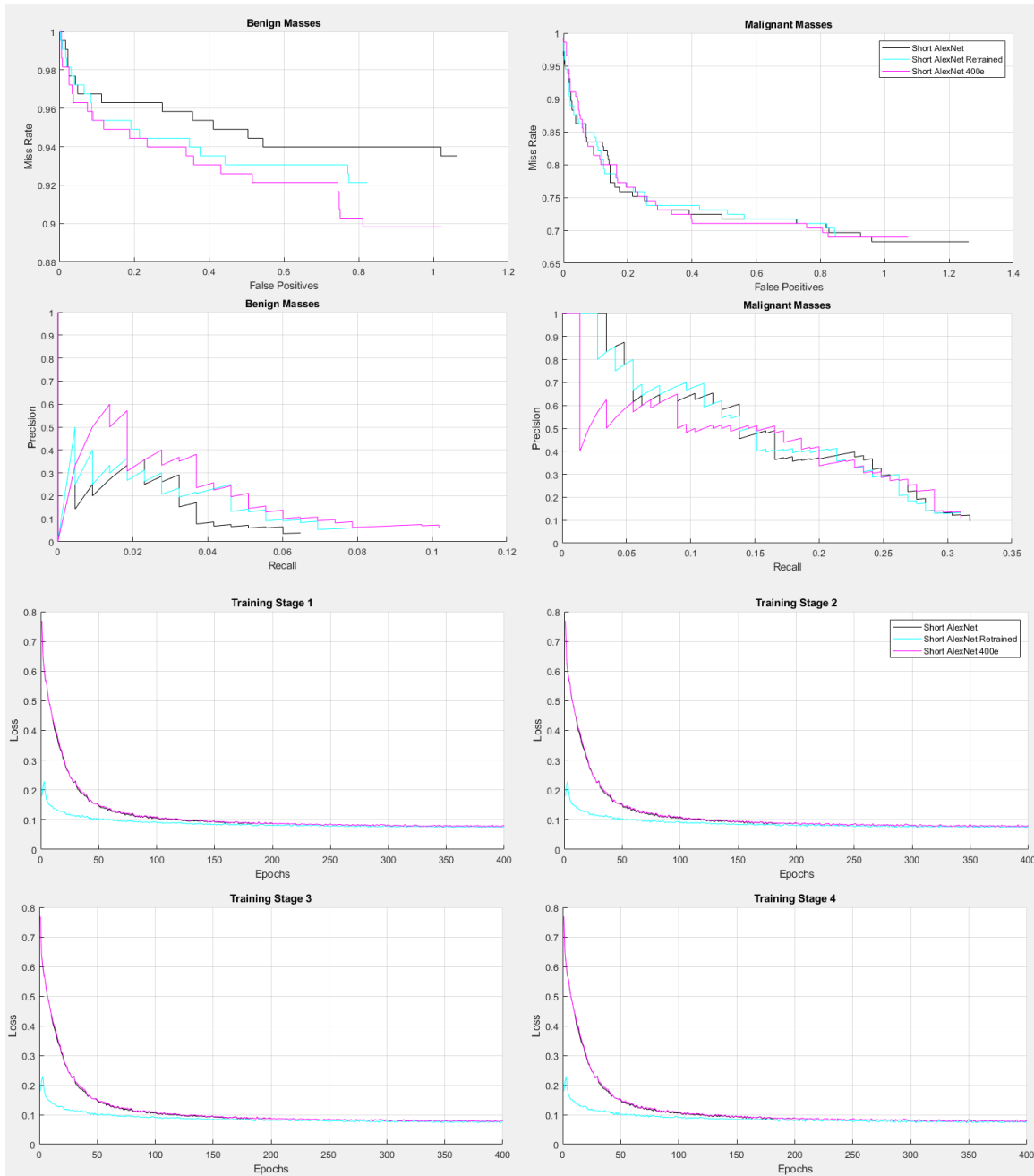
Συγκεκριμένα οι αρχικοί detectors παρήχθησαν χρησιμοποιώντας τις τυποποιημένες παραμέτρους που αφορούν 200 εποχές εκπαίδευσης με κατάβαση μάθησης 10% κάθε 10 εποχές. Για τα πειράματα της παρούσας ενότητας χρησιμοποιήθηκαν παράμετροι που αφορούν διάρκεια εκπαίδευσης 400 εποχών με κατάβαση μάθησης 10% κάθε 20 εποχές. Οι ανιχνευτές που αφορούν την επανεκπαίδευση των αρχικών detectors ονομάζονται στο εξής «Retrained» ενώ αυτοί που παρήχθησαν με μοναδική εκπαίδευση των 400 εποχών σημειώνονται με το όνομα της δομής και τα στοιχεία «_400e».

Στην συνέχεια παρατίθενται τα σχήματα που προέκυψαν από την εκπαίδευση και αξιολόγηση των σχετικών ανιχνευτών ενώ στην ενότητα 6.2.4 παρουσιάζονται συνοπτικά τα αποτελέσματα που προέκυψαν.

6.2.1 Επανεκπαίδευση Short AlexNet

Ερμηνεύοντας τα δεδομένα που παρουσιάζονται στην Εικόνα 52 προκύπτουν οι περιορισμένες δυνατότητες της δομής αυτής. Από τα σχήματα απώλειας εμφανίζεται εικόνας ελάχιστης βελτίωσης από την αύξηση του χρόνου εκπαίδευσης, γεγονός που ανάγεται στον μειωμένο αριθμό μεταβλητών της δομής, ο οποίος περιορίζει και την ικανότητά της στη μάθηση. Επιπλέον εμφανίζεται ικανότερη κατάβαση για το Retrained δίκτυο, το οποίο συγκλίνει γρηγορότερα από τα δύο αρχικής εκπαίδευσης, χωρίς αυτό να συνοδεύεται από αντίστοιχη μείωση της απώλειας.

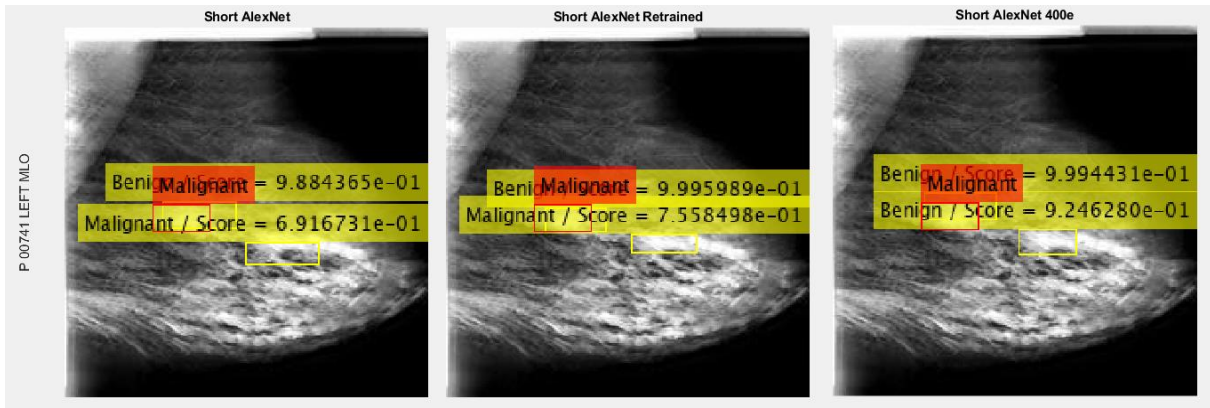
Από τα σχήματα Precision-Recall και FPPI-Miss Rate εντοπίζεται αντίστοιχη απόδοση μεταξύ των δικτύων με το αρχικό συμπέρασμα να είναι πως το συγκεκριμένο δίκτυο δεν έχει την ικανότητα να αυξήσει την απόδοσή του αναλογικά με την διάρκεια της εκπαίδευσης.



Εικόνα 52 – Σχήμα απόδοσης και απώλειας επανεκπαίδευσεων Short AlexNet

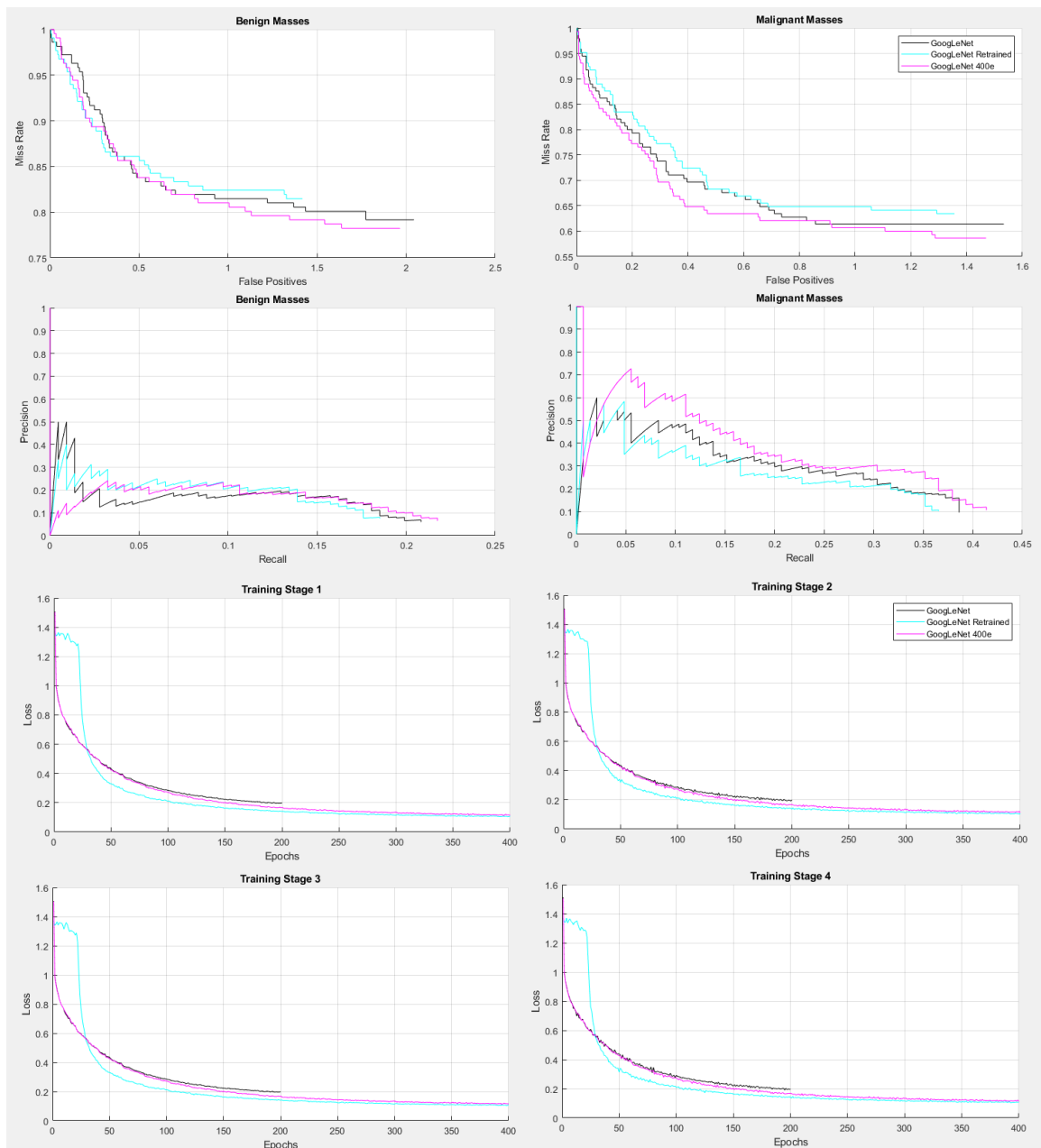
Στο σχήμα της Εικόνα 53 εμφανίζονται αποτελέσματα που προέκυψαν από την τροφοδοσία ενδεικτικής εικόνας του σετ ελέγχου, στα οποία παρατηρούμε μικρή διαφοροποίηση του αποτελέσματος. Τα τρία δίκτυα εντοπίζουν πολλαπλά αντικείμενα, ανάμεσά τους και το σωστό, το οποίο επισημαίνουν με επαρκή βεβαιότητα, με το τρίτο, ωστόσο, δίκτυο να το κατηγοριοποιεί λανθασμένα.

Το ίδιο δίκτυο εντοπίζει ένα λάθος θετικό, που εντοπίζουν και τα συναφή του δίκτυα, ωστόσο το κατηγοριοποιεί με υψηλή βεβαιότητα, ενώ τα υπόλοιπα του αναθέτουν μικρότερη βεβαιότητα. Αυτό το στοιχείο είναι σημαντικό καθώς στη περίπτωση που εφαρμοζόταν υπιερατό φίλτρο αποτελεσμάτων στα δίκτυα, το οποίο θα επέτρεπε τη χρήση μόνο αποτελεσμάτων με υψηλή βεβαιότητα, τα συναφή δίκτυα θα εμφάνιζαν μεγαλύτερη ακρίβεια, κατ' εκτίμηση του στιγμιότυπου που αναλύεται.



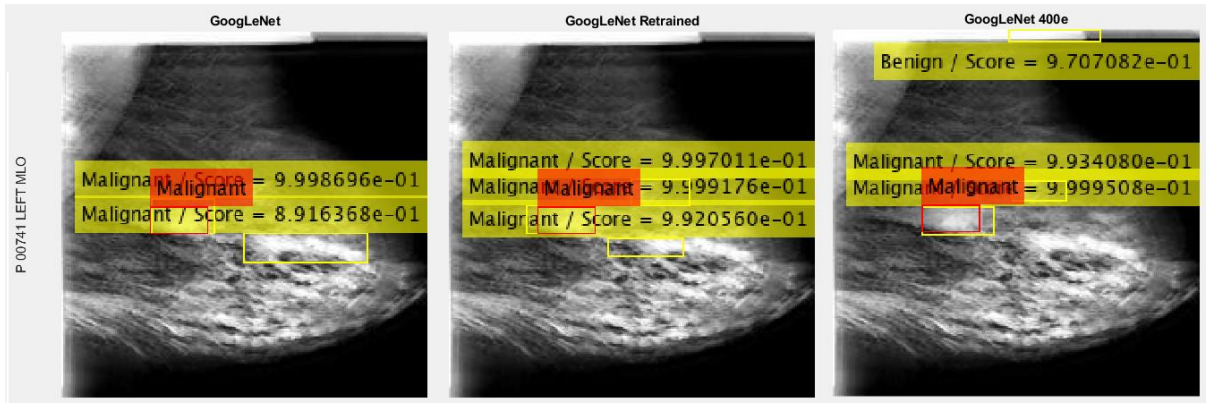
Εικόνα 53 – Ενδεικτικά αποτελέσματα δικτύων Short AlexNet

6.2.2 Επανεκπαίδευση GoogLeNet



Εικόνα 54 – Σχήμα απόδοσης και απώλειας επανεκπαιδύσεων GoogLeNet

Στη παρούσα ενότητα συνεχίζεται η ανάλυση της επιρροής που φέρει η επανεκπαίδευση δικτύων στην απόδοσή τους εστιάζοντας στο δίκτυο GoogLeNet. Το συγκεκριμένο δίκτυο έχει διακριθεί κατά τη διάρκεια της εργασίας για τις επιδόσεις του και διαθέτει τη πιο διακριτή δομή, απαρτιζόμενο από τα πρωτότυπα στρώματα Inception, τα οποία έχουν ήδη αναλυθεί εκτενώς. Αξίζει να σημειωθεί πως η δομή GoogLeNet ήταν η μοναδική που εμφάνισε επαρκώς υψηλή απόδοση ώστε να συμπεριληφθεί αυτούσια στις δομές που μελετώνται στο παρόν κεφάλαιο.



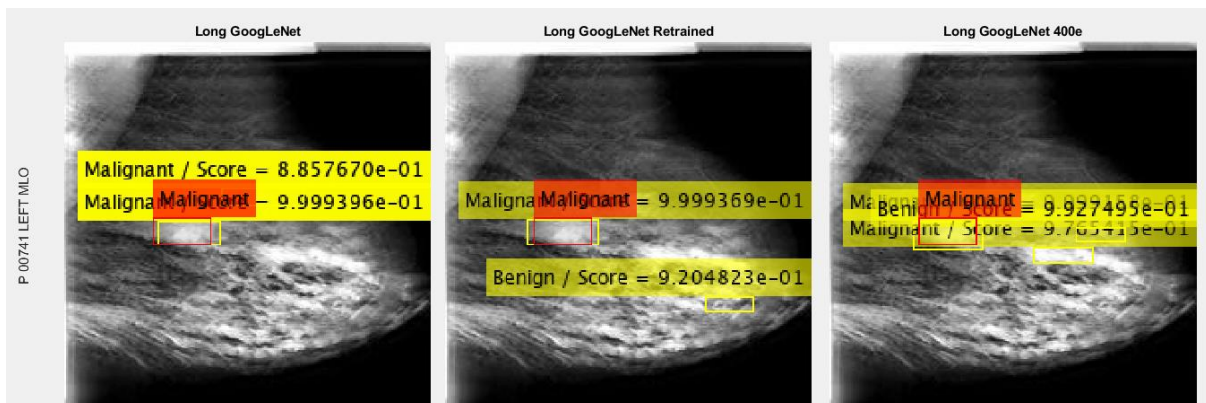
Εικόνα 55 – Ενδεικτικά αποτελέσματα δικτύων GoogLeNet

Ξεκινώντας την ανάλυση από τα σχήματα απώλειας που εμφανίζονται στην Εικόνα 54 παρατηρείται πως για τις πρώτες 25 εποχές το Retrained δίκτυο εμφάνισε δυσκολία σύγκλισης, η οποία αποκαταστάθηκε στη συνέχεια με αποτέλεσμα να μην επηρεαστεί από μη επαρκή εκπαίδευση, ωστόσο, όπως προκύπτει από τις καμπύλες ευαισθησίας το 400e δίκτυο εμφάνισε ικανοποιητικά αποτελέσματα ξεπερνώντας το βασικό δίκτυο.

Ερμηνεύοντας τα αποτελέσματα στην Εικόνα 55 είναι εμφανές πως το βασικό δίκτυο εμφάνιζε ικανότερη ακρίβεια εντοπίζοντας μόνο ένα λάθος θετικό, τη στιγμή που τα επανεκπαιδευμένα δίκτυα εντόπισαν δύο λάθος θετικά. Σημαντικό είναι πως και τα τρία δίκτυα εντόπισαν το σωστό αντικείμενο ενδιαφέροντος και την αντίστοιχα σωστή κατηγοριοποίησή του.

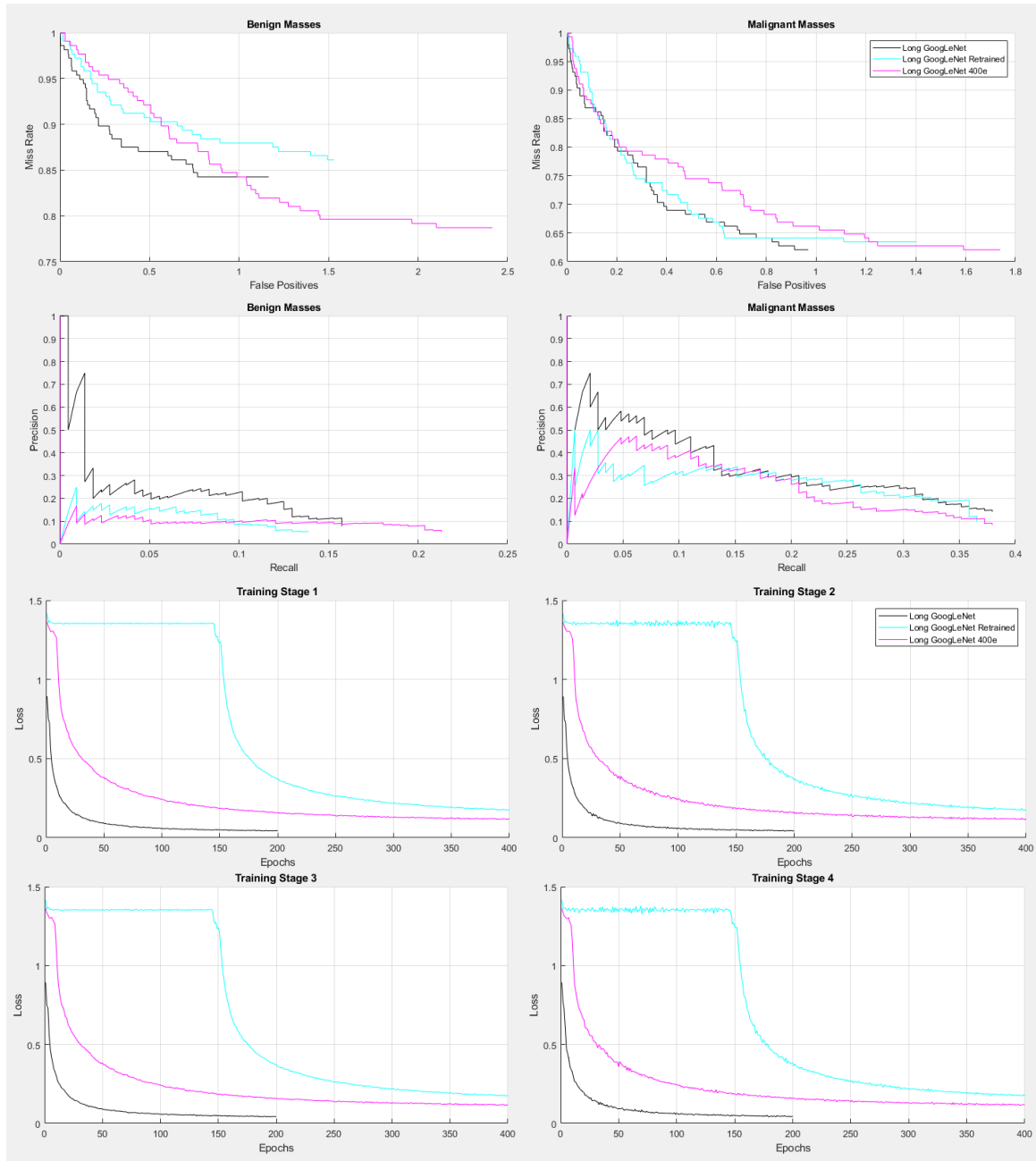
6.2.3 Επανεκπαίδευση Long GoogLeNet

Τελευταίο δίκτυο που μελετήθηκε στη παρούσα σειρά πειραματισμών είναι η δομή Long GoogLeNet. Ερμηνεύοντας τα αποτελέσματα που παρατίθενται στα σχήματα απώλειας της εικόνας 55 είναι εμφανές πως υπήρξε απόκλιση κατά την εκπαίδευση των δικτύων, τόσο του 400e αλλά κυρίως του Retrained δικτύου, το οποίο ξεκίνησε να συγκλίνει μετά την εποχή 150 και δεν κατάφερε να ανακάμψει μέχρι το τέλος της εκπαίδευσης. Εντοπίζεται εδώ και καθίσταται έντονο το αντίστοιχο φαινόμενο που είχε παρατηρηθεί και στην επανεκπαίδευση της βασικής δομής GoogLeNet όπου ένα μέρος της εκπαίδευσης διέρχεται χωρίς μείωση της απώλειας, η οποία επανέρχεται μετά από αυτό το σημείο με αποτέλεσμα την αραίωση της απόδοσης της εκπαίδευσης. Για την αντιμετώπιση αυτού του φαινομένου μπορεί να χρησιμοποιηθεί μικρότερη τιμή initial learning rate, ώστε να προσομοιωθεί η συνέχιση της εκπαίδευσης από ένα προηγούμενο σημείο.



Εικόνα 56 – Ενδεικτικά αποτελέσματα δικτύων Long GoogLeNet

Από τις καμπύλες ευαισθησίας εντοπίζεται το αρχικό δίκτυο να έχει τη μεγαλύτερη ακρίβεια, ενώ τα επανεκπαιδευμένα δίκτυα δεν προσεγγίζουν τα αποτελέσματά του. Για το 400ε δίκτυο μπορεί να παρατηρηθεί πως εμφανίζει σε σημεία άνω των δύο λάθος θετικών ανά εικόνα, ενώ τα συγγενή του δίκτυα περιορίζονται σε τιμές μεταξύ του 1 και 1.5. Προκύπτει συνεπώς από τις τιμές η μη επαρκής εκπαίδευση του δικτύου, ενώ είναι πιθανό να οφείλεται για την ανεπαρκή εκπαίδευση εγκλωβισμός των βαρών του σε τοπικό ελάχιστο που δεν ανταποκρινόταν στις πραγματικές δυνατότητες της δομής, όπως παρατηρήθηκαν στο βασικό δίκτυο.



Εικόνα 57 – Σχήμα απόδοσης και απώλειας επανεκπαιδύσεων Long GoogLeNet

Ο συσχετισμός αυτός επιβεβαιώνεται από την εικόνα Εικόνα 56 στην οποία είναι εμφανής η ανεπάρκεια του δικτύου από τη λάθος κατηγοριοποίηση του ορθού αντικειμένου ενδιαφέροντος και τον ταυτόχρονο εντοπισμό ενός επιπλέον λάθος θετικού αποτελέσματος. Από αυτή την ανεπάρκεια προκύπτει η συνολική αδυναμία του δικτύου να εμφανίσει απόδοση ανάλογη της διάρκειας εκπαίδευσης όπως παρατηρήθηκε στη βασική δομή GoogLeNet, ωστόσο προκύπτει πως η συγκεκριμένη δομή διατηρεί στοιχεία τα οποία μπορούν να της δώσουν το προβάδισμα εφόσον αντιμετωπιστούν τα προβλήματα-τροχοπέδη στα οποία έγινε αναφορά.

6.2.4 Σύνοψη αποτελεσμάτων επανεκπαιδεύσεων

Σε αυτό το σημείο παρατίθενται τα συγκριτικά αποτελέσματα που προέκυψαν από την επανεκπαίδευση των δικτύων που αναλύθηκαν στις προηγούμενες ενότητες του παρών κεφαλαίου. Όπως προβλέφθηκε από τα σχετικά σχήματα δίκτυα διακρίνονται από υψηλές τιμές εντοπισμού αντικειμένων, έχοντας όλα εντοπίσει αντικείμενο ενδιαφέροντος στο ~90% των εικόνων.

Network Base	Benign Average Precision	Malignant Average Precision	Benign Average Miss Rate	Malignant Average Miss Rate	Ποσοστό εικόνων με εύρεση ROI
Short AlexNet	0.0125	0.1618	0.9664	0.8103	90.14%
Short AlexNet Retrained	0.0177	0.1596	0.9591	0.8131	88.70%
Short AlexNet_400e	0.0244	0.1431	0.9508	0.8158	88.41%
Long GoogLeNet	0.0407	0.1321	0.9311	0.8226	87.83%
Long GoogLeNet Retrained	0.0180	0.1062	0.9538	0.8362	94.20%
Long GoogLeNet_400e	0.0211	0.1008	0.9570	0.8520	97.97%
GoogLeNet	0.0380	0.1304	0.9312	0.8196	96.81%
GoogLeNet Retrained	0.0396	0.1121	0.9248	0.8369	95.94%
GoogLeNet_400e	0.0382	0.1654	0.9304	0.7906	96.23%

Πίνακας 17 – Αποτελέσματα επανεκπαίδευσης δομών

Κατατάσσοντας τα δίκτυα σε σειρά μειούμενης απόδοσης εμφανίζεται σαν δίκτυο με την ικανότερη απόδοση η δομή του βασικού GoogLeNet που εκπαιδεύτηκε για 400 εποχές. Η δομή αυτή πετυχαίνει τη μεγαλύτερη τιμή Average Precision που έχει παρατηρηθεί στην παρούσα εργασία, διατηρώντας καλές τιμές Average Miss Rate και ποσοστού εικόνων με εύρεση ROI.

Network Base	Total Average Precision	Total Average Miss Rate
GoogLeNet_400e	0.1018	0.8765
Short AlexNet Retrained	0.0887	0.8501
Short AlexNet	0.0872	0.8559
Long GoogLeNet	0.0864	0.8505
GoogLeNet	0.0842	0.8939
Short AlexNet_400e	0.0838	0.8500
GoogLeNet Retrained	0.0759	0.8982
Long GoogLeNet Retrained	0.0621	0.8891
Long GoogLeNet_400e	0.0610	0.9159

Πίνακας 18 – Συγκριτική απόδοση επανεκπαίδευσης δομών

Συνοψίζοντας, επιβεβαιώνεται το προτέρημα των διδιάστατων δομών Inception στην εξαγωγή features από τα αντικείμενα ενδιαφέροντος και η ικανότητα της ως επαρκώς βαθιάς δομής ώστε να μπορεί να εκμεταλλευτεί μεγαλύτερα διαστήματα εκπαίδευσης αυξάνοντας ανάλογα και τις επιδόσεις της. Ολοκληρώνοντας τη παρούσα εργασία έχει εντυπωθεί το συμπέρασμα της δυναμικής που διαθέτουν τα βαθιά νευρωνικά δίκτυα στην λύση πραγματικών προβλημάτων, όπως και οι ποικίλες πλευρές που τίθενται στην πρόκληση της αυτοματοποιημένης αναγνώρισης καρκινωμάτων και της διαρκούς βελτίωσης της ποιότητας των αποτελεσμάτων.

Βιβλιογραφία

- [1] F. T. L. D. J. S. J. R. M. A. P. & H. J. C. de Dombal, "Computer-aided diagnosis of acute abdominal pain. British medical journal," *British Medical Journal*, vol. 2, no. 5804, pp. 9-13, 1972.
- [2] H. Hollerith, "The Electric Tabulating Machine," *Journal of the Royal Statistical Association*, vol. 57, no. 4, pp. 678-682, 1894.
- [3] R. & H. A. B. Doll, "Smoking and carcinoma of the lung; preliminary report.," *British medical journal*, vol. 2(4682), pp. 739-48., 1950.
- [4] H. Ghayoumi zadeh, "Diagnosing Breast Cancer with the Aid of Fuzzy Logic Based on Data Mining of a Genetic Algorithm in Infrared Images," *Middle East Journal of Cancer*, vol. 3, pp. 119-129, 2011.
- [5] B. M. S. S. K. & M. M. Asl, "Support vector machine-based arrhythmia classification using reduced features of heart rate variability signal.," *Artificial intelligence in medicine*, vol. 4, no. 1, pp. 51-64., 2008.
- [6] R. Z. M. M. B. O. M. S. P. G. & F.-S. L. Ferri, "Computer-assisted detection of nocturnal leg motor activity in patients with restless legs syndrome and periodic leg movements during sleep," *Sleep*, vol. 28, no. 8, pp. 998-1004., 2005.
- [7] Y. B. B. D. J. S. H. D. H. R. E. H. W. & J. L. D. LeCun, "Backpropagation applied to handwritten zip code recognition," *Neural computation*, vol. 1, no. 4, pp. 541-551, 1989.
- [8] F. Rosenblatt, "The perceptron: a probabilistic model for information storage and organization in the brain.," *Psychological review*, vol. 65, no. 6, p. 386, 1958.
- [9] D. E. H. G. E. & W. R. J. Rumelhart, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, p. 533, 1986.
- [10] Y. B. Y. & H. G. LeCun, "Deep learning," *nature*, vol. 521, no. 7553, p. 436, 2015.
- [11] V. & H. G. E. Nair, "Rectified linear units improve restricted boltzmann machines," *Proceedings of the 27th international conference on machine learning*, vol. 10, pp. 807-814, 2010.
- [12] M. & N. S. & Z. A. Razzak, "Deep Learning for Medical Image Processing: Overview, Challenges and Future," *arXiv:1704.06825*, 2017.
- [13] S. Z. H. F. Y. & L. W. Liu, "Prostate cancer diagnosis using deep learning with 3D multiparametric MRI.," *Medical Imaging 2017: Computer-Aided Diagnosis*, vol. 10134, p. 1013428, 2017.
- [14] F. C. K. V. R. S. S. A. G. P. J. C. T. S. E. S.-P. C. W. M. M. A. P. U. P. M. & V. G. B. Ciompi, "Towards automatic pulmonary nodule management in lung cancer screening with deep learning," *Scientific Reports*, vol. 7, no. 46479, 2017.
- [15] X. S. J. & V. V. Huang, "Lung nodule detection in CT using 3D convolutional neural networks," *IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pp. 379-383, 2017.
- [16] H. D. S. X. J. M. R. H. L. M. C. X. P. & D. H. N. Cheng, "Approaches for automated detection and classification of masses in mammograms," *Pattern recognition*, pp. 646-668, 2006.
- [17] W. & D. K. & G. M. & N. R. & S. R. Zhang, "An improved shift-invariant artificial neural network for computerized detection of clustered microcalcifications in digital mammograms," *Medical physics*, vol. 23, pp. 595-601, 1996.
- [18] F. D. L. A. Papadopoulos A, "An automatic microcalcification detection system based on a hybrid neural network classifier," *Artif Intell Med.*, 2002.
- [19] J. G. F. A. R.-P. R. O. J. L. & L. M. A. G. Arevalo, "Convolutional neural networks for mammography mass lesion classification," *37th Annual International Conference of the IEEE*, vol. Engineering in Medicine and Biology Society (EMBC), pp. 797-800, 2015.

- [20] H. & K. P. & J. A. & S. Alharbi, "Fuzzy image segmentation for mass detection in digital mammography: Recent advances and techniques. Multidisciplinary Computational Intelligence Techniques: Applications in Business, Engineering, and Medicine," *A.S.M.*, vol. 2, pp. 378-402, 2012.
- [21] T. K. M. P. T. J. S. E. O. & Y. D. A. Arodz, "Detection of clustered microcalcifications in small field digital mammography," *Computer methods and programs in biomedicine*, vol. 1, no. 1, pp. 56-65, 2006.
- [22] A. S. I. & H. G. E. Krizhevsky, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, pp. 1097-1105, 2012.
- [23] R. F. Matthew D Zeiler, "Visualizing and Understanding Convolutional Networks," *European conference on computer vision*, pp. 818-833, 2014.
- [24] O. A. A. A. T. C. C. W. G. N. K. A. P. K. S. T. W. F. A. Morgan P. McBee, "Deep Learning in Radiology," *Academic Radiology*, vol. 25, no. 11, pp. 1472-1480, 2018.
- [25] J. Y. S. S. R. G. R. T. H. C. B. K. M. B. G. J. L. N. Tajbakhsh, "Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1299-1312, 2016.
- [26] Y. & D. I. & W. L. & L. S. & K. E. & G. H. Bar, "Deep learning with non-medical training used for chest pathology identification," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 2018.
- [27] A. A. S. E. Y. S. & C. P. Huqqani, "Multicore and gpu parallelization of neural networks for face recognition.," *Procedia Computer Science*, vol. 18, pp. 349-358, 2013.
- [28] E. Ç. A. & G. Z. Cengil, "A GPU-based convolutional neural network approach for image classification.," *In Artificial Intelligence and Data Processing Symposium*, pp. 1-6, 2017.
- [29] V. B. S. K. F. J. K. J. K. P. M. S. P. S. M. D. P. M. T. L. P. F. Clark K, "The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository," *Journal of Digital Imaging*, vol. 26, no. 6, pp. 1045-1057, 2013.
- [30] F. G. A. H. K. K. M. M. G. & D. L. R. Rebecca Sawyer Lee, "A curated mammography data set for use in computer-aided detection and diagnosis research," *Scientific Data*, vol. 4, no. 170177, 2017.
- [31] F. G. A. H. ., D. R. Rebecca Sawyer Lee, "Curated Breast Imaging Subset of DDSM," *The Cancer Imaging Archive*, no. <http://dx.doi.org/10.7937/K9/TCIA.2016.7O02S9CY>, 2016.
- [32] C. L. C. C. H. K. & T. X. Dong, "Image Super-Resolution Using Deep Convolutional Networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, pp. 295-307, 2016.
- [33] V. & S. S. Jain, "Natural Image Denoising with Convolutional Networks," *Advances in Neural Information Processing Systems*, pp. 769-776, 2009.
- [34] J. K. L. J. & M. L. K. Kim, "Accurate Image Super-Resolution Using Very Deep Convolutional Networks," *In Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1646-1654, 2016.
- [35] R. & D. J. & D. T. & M. J. Girshick, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition," *CVPR*, vol. CVPR.2014, 2013.
- [36] R. Girshick, "Fast R-CNN," *CoRR*, vol. abs/1504.08083, 2015.
- [37] J. R. V. D. S. K. E. G. T. & S. A. W. Uijlings, "Selective Search for Object Recognition," *Journal of computer vision*, vol. 104, pp. 154-171, 2013.
- [38] C. L. & D. P. Zitnick, "Edge Boxes: Locating Object Proposals from Edges," *Springer, Cham*, pp. 391-405, 2014.
- [39] S. & H. K. & G. R. & S. J. Ren, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, 2015.

- [40] G. G. P. D. a. R. G. Kaiming He, "Mask r-cnn," *IEEE International Conference on Computer Vision*, p. 2980–2988, 2017.
- [41] J. S. E. & D. T. Long, "Fully Convolutional Networks for Semantic Segmentation," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431-3440, 2015.
- [42] H. Q. J. D. X. J. a. Y. W. Y. Li, "Fully Convolutional Instance-Aware Semantic Segmentation," *IEEE Conference on Computer Vision and Pattern Recognition*, vol. CVPR 2017, pp. 4438-4446, 2017.
- [43] J. L. Y. H. K. & S. J. Dai, "R-FCN: Object Detection via Region-based Fully Convolutional Networks," *Advances in neural information processing systems*, pp. 379-387, 2016.
- [44] A. C. D. C. M. J. G. L. M. & S. J. Giusti, "Fast image scanning with deep max-pooling convolutional neural networks," *IEEE International Conference on Image Processing (ICIP)*, pp. 4034-4038, 2013.
- [45] K. Z. X. R. S. & S. J. He, "Deep Residual Learning for Image Recognition," *In Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778, 2016.
- [46] C. L. W. J. Y. S. P. R. S. A. D. & R. Szegedy, "Going deeper with convolutions," *In Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1-9, 2015.
- [47] S. B. A. G. R. & M. T. Arora, "Provable Bounds for Learning Some Deep," *International Conference on Machine Learning*, pp. 584-592, 2014.
- [48] S. D. R. G. A. F. J. Redmon, "You only look once: Unified, real-time object detection," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, p. 779–788, 2016.
- [49] B. Z. Q. V. L. Prajit Ramachandran, "Searching for Activation Functions," *arxiv:1710.05941*, 2017.
- [50] B. Polyak, "Some methods of speeding up the convergence of iteration methods," *Ussr Computational Mathematics and Mathematical Physics*, vol. 4, pp. 1-17, 1964.
- [51] M. D. Zeiler, "ADADELTA: an adaptive learning rate method," 2012, no. arXiv preprint arXiv:1212.5701..
- [52] S. C. A. E. & L. Y. Zhang, "Deep learning with Elastic Averaging SGD," *Advances in Neural Information Processing Systems*, pp. 685-693, 2015.
- [53] H. & T. T. Huang, "Reinforced stochastic gradient descent for deep neural network learning," *arXiv:1701.07974*, 2017.
- [54] J. H. E. & S. Y. Duchi, "Adaptive subgradient methods for online learning and stochastic optimization," *Journal of Machine Learning Research*, vol. 12, pp. 2121-2159, 2011.
- [55] M. & B. H. Riedmiller, "A direct adaptive method for faster backpropagation learning: The RPROP algorithm," *IEEE International Conference on Neural Networks*, pp. 586-591, 1993.
- [56] T. a. H. G. Tieleman, "Lecture 6.5-rmsprop: Divide the Gradient by a Running Average of its Recent Magnitude," *COURSERA: Neural Networks for Machine Learning*, vol. 4, pp. 26-30, 2012.
- [57] D. P. & B. J. Kingma, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [58] M. Matlab, "R-CNN, Fast R-CNN, and Faster R-CNN Basics," <https://www.mathworks.com/help/vision/ug/faster-r-cnn-basics.html>.