ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΕΠΙΚΟΙΝΩΝΙΩΝ, ΗΛΕΚΤΡΟΝΙΚΗΣ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ

Εμπλουτισμός Δεδομένων Με Χρήση Αντιστρέψιμων Δημιουργικών Νευρωνικών Δικτύων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΑΚΑΡΕΠΗΣ ΑΝΔΡΕΑΣ

Επιβλέπων: Γεώργιος Στάμου
Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2022

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΕΠΙΚΟΙΝΩΝΙΩΝ, ΗΛΕΚΤΡΟΝΙΚΗΣ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ

Εμπλουτισμός Δεδομένων Με Χρήση Αντιστρέψιμων Δημιουργικών Νευρωνικών Δικτύων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΑΚΑΡΕΠΗΣ ΑΝΔΡΕΑΣ

Επιβλέπων: Γεώργιος Στάμου
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 11η Απριλίου 2022.

.....	.....	.....
Γεώργιος Στάμου	Θάνος Βουλοδήμος	Ανδρέας Γ. Σταφυλοπάτης
Καθηγητής Ε.Μ.Π.	Επ. Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2022

.....
Ακαρέπης Ανδρέας
Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Περίληψη

Ένα από τα μεγαλύτερα προβλήματα που αντιμετωπίζουν οι ιατρικές βάσεις δεδομένων είναι η έλλειψη και ο μικρός αριθμός δειγμάτων. Ακόμα και σε περιπτώσεις που υπάρχουν αρκετά ιατρικά δείγματα σε μια βάση, πολλά από αυτά δεν έχουν κάποια σχετική ετικέτα η οποία να κατηγοριοποιεί το δείγμα. Αυτό επιφέρει μεγάλες δυσκολίες στους ειδικούς που θέλουν να κατασκευάσουν νευρωνικά δίκτυα ταξινόμησης προκειμένου να κατηγοριοποιούν αυτόματα και με μεγαλύτερη ευκολία τα ιατρικά δεδομένα. Μια λύση σε αυτό το πρόβλημα έρχεται να δώσει πάλι η τεχνητή νοημοσύνη μέσω της αυτόματης σύνθεσης εικόνας. Μπορούμε χρησιμοποιώντας παραγωγικά νευρωνικά δίκτυα να παράξουμε νέα δεδομένα τα οποία θα εμπλουτίσουν την υπάρχουσα βάση δεδομένων η οποία πλέον θα μπορεί να χρησιμοποιηθεί πιο αποτελεσματικά για να εκπαιδεύσει ένα νευρωνικό δίκτυο ταξινόμησης. Κάτι αντίστοιχο κάναμε και στην παρούσα εργασία. Αρχικά πήραμε την ιατρική βάση δεδομένων mini-DDSM η οποία περιέχει μαστογραφίες χωρίς όγκο, με καλοήθη όγκο και με καρκίνο (για απλότητα αφαιρέσαμε τις εικόνες με τον καλοήθη όγκο) και επιχειρήσαμε να δούμε πόσο καλά μπορούμε να εκπαιδεύσουμε ένα προεκπαιδευμένο νευρωνικό δίκτυο ώστε να ταξινομεί σωστά τις εικόνες. Στη συνέχεια χρησιμοποιώντας διάφορα μοντέλα Normalizing Flow με έμφαση στο μοντέλο GLOW, παράξαμε αρκετές φωτογραφίες και εμπλουτίσαμε ένα υποσύνολο της βάσης δεδομένων που ορίσαμε ως σύνολο εκπαίδευσης. Χρησιμοποιώντας το ενισχυμένο σύνολο εκπαίδευσης, ξαναεκπαιδεύσαμε το προεκπαιδευμένο νευρωνικό δίκτυο και καταφέραμε να πάρουμε καλύτερα αποτελέσματα. Παράλληλα, επιχειρήσαμε με διάφορες τεχνικές να παρακάμψουμε κάποια προβλήματα που είχε το σλυνολο εκπαίδευσης όπως ότι κάποιες φωτογραφίες ήταν “σημαδεμένες” με άσπρο χρώμα. Κάποιες τεχνικές είχαν καλύτερα αποτελέσματα, άλλες όχι τόσο καλά.

Λέξεις Κλειδιά: Τεχνητή Νοημοσύνη, Τεχνητά Νευρωνικά Δίκτυα, Επιβλεπόμενη Μάθηση, Μη-Επιβλεπόμενη Μάθηση, Μαστογραφία, Παραγωγικά Νευρωνικά Δίκτυα, Παραγωγή Εικόνας

Abstract

One of the biggest problems medical databases are facing is the lack and small number of samples. Even in cases where there are several medical samples in a database, many of them do not have a relevant label to categorize the sample. This brings great difficulties to specialists who want to construct classification neural networks in order to categorize medical data automatically. A solution to this problem is given by artificial intelligence through automatic image synthesis. We can use generative neural networks to generate new data that will enrich the existing database which can now be used more efficiently to train a classification neural network. We did something similar in the current thesis. First, we took the mini-DDSM medical database which contains tumor-free, benign and cancer mammograms (for simplicity we removed the benign images) and tried to see how well we can train a pretrained neural network to classify correctly the images. Then, using various Normalizing Flow models with emphasis on the GLOW model, we produced several photos and enriched a subset of the database that we defined as a training set. Using the enhanced training set, we retrained the pre-trained neural network and managed to get better results. At the same time, we experimented with various techniques to bypass some problems that the training set had, such as that some photos had white markings. Some techniques worked better, others not so well.

Keywords: Artificial Intelligence, Artificial Neural Networks, Supervised Learning, Unsupervised Learning, Mammogram, Generative Neural Networks, Image synthesis

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου Γεώργιο Στάμου που με ενέπνευσε και μου έδωσε την ευκαιρία να ασχοληθώ με ένα τόσο ενδιαφέρον θέμα καθώς και τον υποψήφιο διδάκτορα Γιώργο Φιλανδριανό του οποίου η βοήθεια ήταν πολύτιμη για την εκπόνηση αυτής της εργασίας.

Θα ήθελα επίσης να ευχαριστήσω τους γονείς μου που όλα τα χρόνια των σπουδών μου έκαναν ό,τι μπορούσαν για να με στηρίξουν και να με βοηθήσουν. Θα ήθελα επίσης να ευχαριστήσω την συμφοιτήτρια και φίλη μου Ξένια Κόντη η οποία με βοήθησε πολύ να ολοκληρώσω αυτήν τη σχολή καθώς και τον συμφοιτητή και φίλο μου Μάνο Γιαννόπουλο με τον οποίο συνεργαστήκαμε πολλές φορές σε εργασίες και διαγωνισμούς.

Πίνακας Περιεχομένων

Κεφάλαιο 1: Εισαγωγή

1.1 Πρόλογος

Η παραγωγή εικόνας είναι ένα έργο το οποίο βρίσκει εφαρμογές σε ένα πολύ μεγάλο εύρος κλάδων, από τις ταινίες και τα ηλεκτρονικά παιχνίδια μέχρι και την ιατρική. Είναι συνεπώς αναμενόμενο να έχουν δημιουργηθεί πολλές και διαφορετικές τεχνικές για την παραγωγή εικόνας. Πολλές από αυτές ανήκουν στον κλάδο της τεχνητής νοημοσύνης και πιο συγκεκριμένα, στον κλάδο των νευρωνικών δικτύων. Το πρώτο νευρωνικό δίκτυο που κατασκευάστηκε και είχε καλά αποτελέσματα ήταν το λεγόμενο Variational Autoencoder (VAE)[1]. Το 2014, προτάθηκε ένα νέο είδους γεννητικού δικτύου, το Generative Adversarial Network (GAN)[2] το οποίο χρησιμοποιεί δύο επί μέρους νευρωνικά δίκτυα, έναν γεννήτορα και έναν ταξινομητή που βρίσκονται σε ανταγωνιστικό περιβάλλον. Τα αποτελέσματα αυτής της νέας αρχιτεκτονικής ήταν πολύ καλά, σε σημείο να ξεγελάνε ακόμα και το ανθρώπινο μάτι. Ωστόσο αυτή η νέα αρχιτεκτονική είχε κάποια μειονεκτήματα, που θα αναλυθούν παρακάτω, τα οποία την καθιστούσαν μη-ιδανική για πολλά είδη εφαρμογών. Αυτό οδήγησε τους ερευνητές στην εύρεση νέων νευρωνικών δικτύων που θα εξάλειπταν αυτά τα προβλήματα. Κατάφεραν να φτιάξουν ένα νέο μοντέλο, που ονόμασαν Normalizing Flow[3] το οποίο είχε εξίσου καλά αποτελέσματα, χωρίς τα προβλήματα των GANs. Με αυτήν την αρχιτεκτονική και τα διάφορα μοντέλα που έχουν προταθεί πάνω σε αυτήν, ασχολείται η παρούσα εργασία.

1.2 Σκοπός

Σκοπός της παρούσας διπλωματικής εργασίας είναι η χρήση γενετικών νευρωνικών δικτύων προκειμένου να γίνει επαύξηση της ιατρικής βάσης δεδομένων miniDDSM η οποία περιέχει μαστογραφίες με καρκίνο και χωρίς καρκίνο. Τα παραγόμενα δεδομένα δημιουργούνται με διάφορες τεχνικές όπως είναι η απλή δημιουργία από τυχαίο θόρυβο στο latent space, interpolation, random ball, latent space manipulation. Τα παραγωγικά νευρωνικά δίκτυα που χρησιμοποιούνται στην παρούσα εργασία για αυτόν τον σκοπό ονομάζονται Glow[4] και ανήκουν στην κατηγορία των Normalizing Flows. Ωστόσο, ελέγχονται και μερικά παλαιότερα μοντέλα Flow όπως το RealNVP[5]. Η αξιολόγηση των παραγόμενων εικόνων γίνεται μέσω προεκπαιδευμένων

συνελικτικών νευρωνικών δικτύων κατηγοριοποίησης στα οποία εφαρμόζουμε μεταφερόμενη μάθηση.

1.3 Είδη Εικόνας

Μια εικόνα είναι ένα τεχνούργημα που απεικονίζει κάποια οπτική αντίληψη, όπως μια φωτογραφία που δείχνει ένα θέμα (π.χ. ένα φυσικό αντικείμενο) και έτσι παρέχει στον παρατηρητή μια απεικόνισή του. Μια ψηφιακή εικόνα αποτελείται από τα στοιχεία εικόνας (picture elements), γνωστά και ως πίξελ, τα οποία είναι πεπερασμένα και έχουν διακριτές ιδιότητες ως προς την έντασή τους όταν η εικόνα είναι πολυχρωματική ή ως προς την κλίμακα του γκρι, όταν οι εικόνα ανήκει σε αυτήν την κατηγορία. Οι πολυχρωματικές εικόνες συνήθως απεικονίζονται με το μοντέλο Red Green Blue (RGB) στο οποίο κάθε πίξελ αποτελείται από αυτά τα 3 χρώματα τα οποία παίρνουν τιμές που καθορίζουν την ένταση του κάθε χρώματος. Από τον συνδυασμό αυτών των τριών χρωμάτων, προκύπτουν και τα υπόλοιπα. Τα πίξελ των εικόνων που βρίσκονται σε κλίμακα του γκρι (grayscale) ελέγχονται μόνο από μία τιμή η οποία καθορίζει το πόσο άσπρο ή μαύρο θα είναι το κάθε πίξελ. Προφανώς αυτές οι εικόνες δεν έχουν χρώματα παρά μόνο αποχρώσεις του γκρι.

Κεφάλαιο 2: Κατηγορίες της Τεχνητής Νοημοσύνης

2.1 Τεχνητή Νοημοσύνη

Ο όρος τεχνητή νοημοσύνη αναφέρεται στον κλάδο της πληροφορικής ο οποίος ασχολείται με τη σχεδίαση και την υλοποίηση υπολογιστικών συστημάτων που μιμούνται στοιχεία της ανθρώπινης συμπεριφοράς τα οποία υπονοούν έστω και στοιχειώδη ευφυΐα: μάθηση, προσαρμοστικότητα, εξαγωγή συμπερασμάτων, κατανόηση από συμφραζόμενα, επίλυση προβλημάτων κλπ. Οι πρώτες περιγραφές ενός τεχνητού νευρωνικού δικτύου έκαναν την εμφάνισή τους από τη δεκαετία του 1940. Ήδη από το 51 ξεκίνησαν να φτιάχνονται προγράμματα τα οποία ήταν ικανά να παίξουν παιχνίδια όπως σκάκι και ντάμα. Το 1956 κατασκευάστηκε το Logic Theorist, το πρώτο πρόγραμμα που στηριζόταν σε συμπερασματικούς κανόνες τυπικής λογικής και σε ευρετικούς αλγόριθμους. Σκοπός του ήταν να αποδεικνύει μαθηματικά θεωρήματα. Ένα από τα πιο διάσημα υποπεδία της τεχνητής νοημοσύνης το οποίο έχει γνωρίσει ραγδαία ανάπτυξη τα τελευταία χρόνια είναι η μηχανική μάθηση.

2.2 Μηχανική Μάθηση

Μηχανική μάθηση είναι υποπεδίο της επιστήμης των υπολογιστών, που αναπτύχθηκε από τη μελέτη της αναγνώρισης προτύπων και της υπολογιστικής θεωρίας μάθησης στην τεχνητή νοημοσύνη. Το 1959, ο Άρθουρ Σάμουελ ορίζει τη μηχανική μάθηση ως "Πεδίο μελέτης που δίνει στους υπολογιστές την ικανότητα να μαθαίνουν, χωρίς να έχουν ρητά προγραμματιστεί". Η μηχανική μάθηση διερευνά τη μελέτη και την κατασκευή αλγορίθμων που μπορούν να μάθαίνουν από τα δεδομένα και να κάνουν προβλέψεις σχετικά με αυτά. Τέτοιοι αλγόριθμοι λειτουργούν κατασκευάζοντας μοντέλα από πειραματικά δεδομένα, προκειμένου να κάνουν προβλέψεις βασιζόμενες στα δεδομένα ή να εξάγουν αποφάσεις που εκφράζονται ως το αποτέλεσμα.

Ο Tom M. Mitchell πρότεινε έναν επίσημο ορισμό που χρησιμοποιείται ευρέως[6]: «Ένα πρόγραμμα υπολογιστή λέγεται ότι μαθαίνει από εμπειρία E

ως προς μια κλάση εργασιών T και ένα μέτρο επίδοσης P, αν η επίδοσή του σε εργασίες της κλάσης T, όπως αποτιμάται από το μέτρο P, βελτιώνεται με την εμπειρία E».

Η μηχανική μάθηση συνήθως χωρίζεται σε 3 βασικές κατηγορίες ανάλογα με το εκπαιδευτικό σύστημα που χρησιμοποιείται:

- Επιβλεπόμενη μάθηση (supervised learning)[7]: Το υπολογιστικό πρόγραμμα δέχεται τις παραδειγματικές εισόδους καθώς και τα επιθυμητά αποτελέσματα από έναν επιβλέπων, και ο στόχος είναι να μάθει έναν γενικό κανόνα προκειμένου να αντιστοιχίσει τις εισόδους με τα αποτελέσματα.
- Μη επιβλεπόμενη μάθηση (unsupervised learning)[8]: Χωρίς κάποια εμπειρία στον αλγόριθμο μάθησης, πρέπει να βρεθεί η δομή των δεδομένων εισόδου.
- Ενισχυτική μάθηση (reinforced learning)[9]: Ένα πρόγραμμα υπολογιστή αλληλεπιδρά με ένα δυναμικό περιβάλλον στο οποίο πρέπει να επιτευχθεί ένας συγκεκριμένος στόχος, χωρίς κάποιος επιβλέπων να του λέει ρητά αν έχει φτάσει στον στόχο του. Συνήθως περιλαμβάνει ένα σύστημα ανταμοιβής.

Ένας άλλος τρόπος κατηγοριοποίησης των προβλημάτων μηχανικής μάθησης αφορά το επιθυμητό αποτέλεσμα μετά την μάθηση. Η κατηγοριοποίηση μπορεί να γίνει σε:

- Στην ταξινόμηση (classification), τα δεδομένα εισόδου χωρίζονται σε δύο ή περισσότερες κλάσεις, και η μηχανή πρέπει να κατασκευάσει ένα μοντέλο, το οποίο θα αντιστοιχίζει τα δεδομένα σε μία ή περισσότερες κλάσεις. Αυτό συνήθως εμπίπτει στην επιβλεπόμενη μάθηση.
- Στην παλινδρόμηση (regression), η οποία είναι και αυτή πρόβλημα επιβλεπόμενης μάθησης, όπου τα αποτελέσματα αυτήν τη φορά είναι συνεχή και όχι διακριτά.
- Στην συσταδοποίηση (clustering), όπου ένα σύνολο εισόδων πρόκειται να χωριστεί σε ομάδες. Αντίθετα με την ταξινόμηση, οι ομάδες δεν είναι γνωστές εκ των προτέρων, καθιστώντας αυτόν τον διαχωρισμό τυπική εργασία μη επιβλεπόμενης μάθησης.

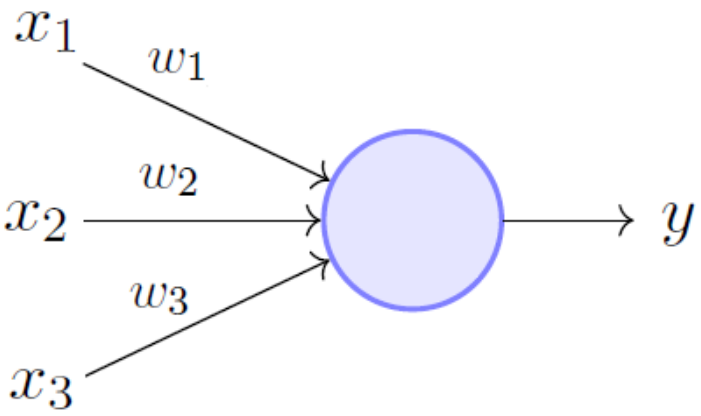
2.3 Νευρωνικά Δίκτυα

Νευρωνικό δίκτυο ονομάζεται ένα κύκλωμα διασυνδεδεμένων τεχνητών νευρώνων[10]. Πρόκειται για ένα αφηρημένο αλγοριθμικό κατασκεύασμα το οποίο εμπίπτει στον τομέα της υπολογιστικής νοημοσύνης και έχει ως στόχο την επίλυση κάποιου υπολογιστικού προβλήματος. Ένας αλγόριθμος εκμάθησης νευρωνικού δικτύου, είναι ένας αλγόριθμος μάθησης, που εμπνέεται από τη δομή και τις λειτουργικές πτυχές των βιολογικών νευρωνικών δικτύων. Η δομή των υπολογισμών βασίζεται σε μια ομάδα εσωτερικά διασυνδεδεμένων τεχνητών νευρώνων, οι οποίοι επεξεργάζονται την πληροφορία και εκτελούν υπολογισμούς επικοινωνώντας μεταξύ τους.

Οι νευρώνες είναι τα δομικά στοιχεία του δικτύου. Κάθε τέτοιος κόμβος δέχεται ένα σύνολο αριθμητικών εισόδων από διαφορετικές πηγές, κάνει έναν υπολογισμό με βάση αυτές τις εισόδους και παράγει μία έξοδο. Αυτή η έξοδος είτε κατευθύνεται στο περιβάλλον ή τροφοδοτείται ως είσοδος σε άλλους νευρώνες του δικτύου. Υπάρχουν τρεις τύποι νευρώνων: οι νευρώνες εισόδου, οι νευρώνες εξόδου και οι υπολογιστικοί νευρώνες ή αλλιώς κρυμμένοι νευρώνες. Οι νευρώνες εισόδου δεν κάνουν κανέναν υπολογισμό, μεσολαβούν απλώς ανάμεσα στις περιβαλλοντικές εισόδους του δικτύου και στους υπολογιστικούς νευρώνες. Οι νευρώνες εξόδου εξάγουν τις τελικές αριθμητικές εξόδους του δικτύου. Οι υπολογιστικοί νευρώνες πολλαπλασιάζουν κάθε είσοδό τους με το αντίστοιχο συναπτικό βάρος και υπολογίζουν το ολικό άθροισμα των γινομένων. Το άθροισμα αυτό τροφοδοτείται ως όρισμα στη συνάρτηση ενεργοποίησης, την οποία υλοποιεί εσωτερικά κάθε κόμβος. Η τιμή που λαμβάνει η συνάρτηση για το εν λόγω όρισμα είναι και η έξοδος του νευρώνα για τις τρέχουσες εισόδους και βάρη.

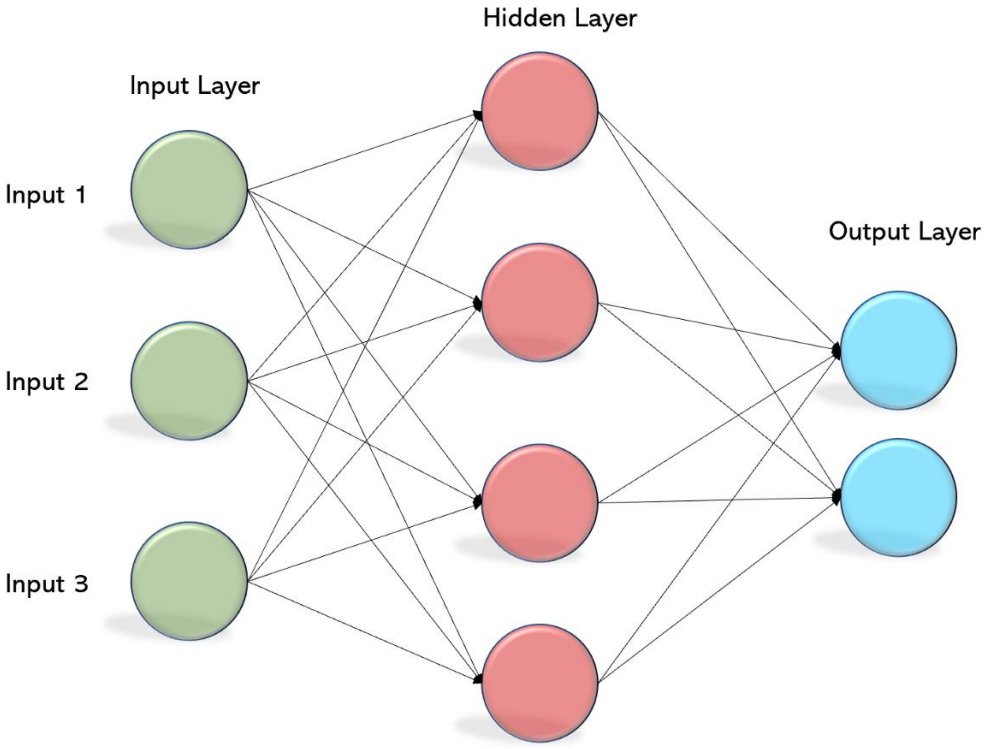
Έστω X_{ki} η i-οστή είσοδος του νευρώνα k, w_{ki} το i-οστό βάρος του νευρώνα και Φ η συνάρτηση ενεργοποίησης του νευρωνικού δικτύου, τότε η έξοδος y_k του νευρώνα δίνεται από την εξίσωση:

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



Perceptron Model (Minsky-Papert in 1969)

Πολλοί τέτοιοι νευρώνες συνδυάζονται για να κατασκευάσουν ένα σύνθετο δίκτυο με μεγαλύτερες ικανότητες πρόβλεψης. Ένα κλασικό παράδειγμα είναι τα Multilayer Perceptrons τα οποία χρησιμοποιούν “κρυφά” επίπεδα ανάμεσα στα επίπεδα εισόδου και εξόδου.



Τα νευρωνικά δίκτυα είναι εφαρμόσιμα σχεδόν σε κάθε κατάσταση στην οποία ισχύει μια σχέση μεταξύ μεταβλητών πρόβλεψης και προβλεπόμενων μεταβλητών, ακόμα και όταν αυτή η σχέση είναι πολύ περίπλοκη για να

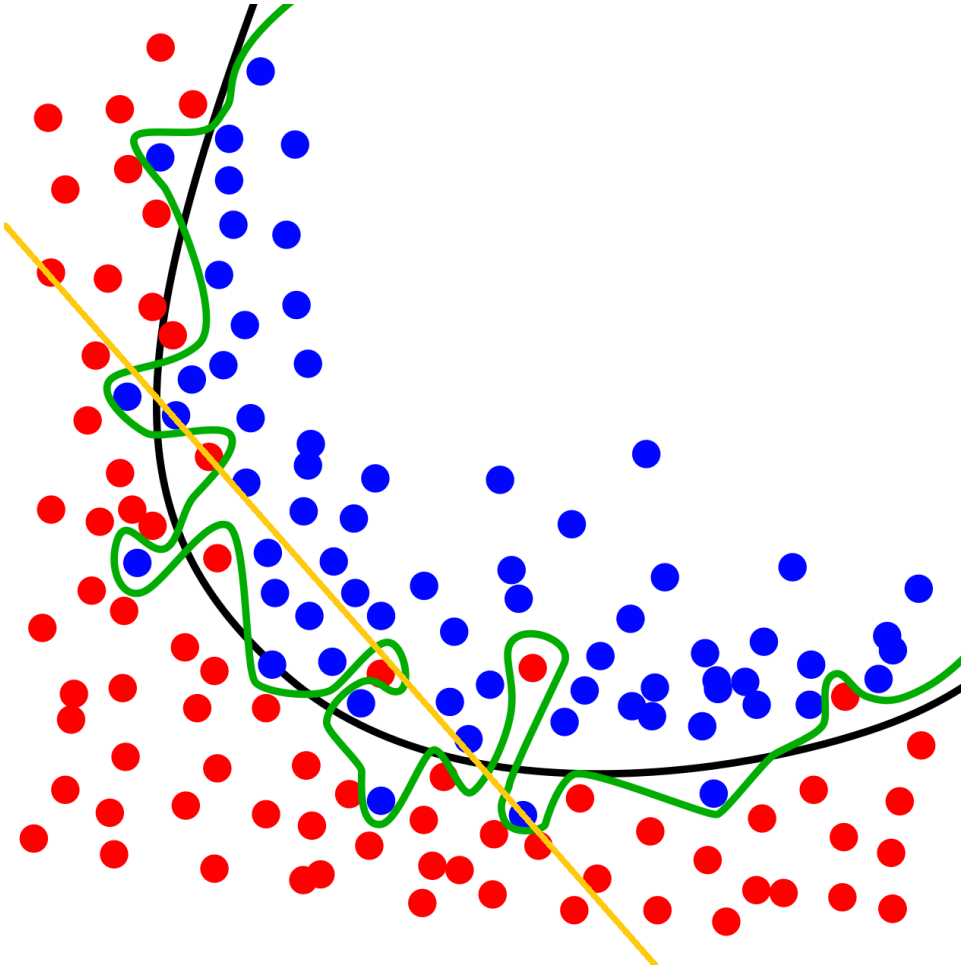
αποδοθεί με τους συνηθισμένους όρους της συσχέτισης ή των διαφόρων ομάδων.

Μια μεγάλη υποκατηγορία των νευρωνικών δικτύων είναι τα γενετικά νευρωνικά δίκτυα(generative neural networks). Σε αυτήν ανήκουν και τα FLOW τα οποία είναι τα νευρωνικά δίκτυα που μελετώνται στην παρούσα εργασία.

Κεφάλαιο 3: Χαρακτηριστικά των Νευρωνικών Δικτύων

3.1 Overfitting/Underfitting

Όταν εκπαιδεύουμε ένα νευρωνικό δίκτυο σε ένα σύνολο εκπαίδευσης, ο στόχος μας είναι να το εκπαιδεύσουμε τόσο ώστε να μπορεί να γενικεύσει αρκετά καλά στο σύνολο των δειγμάτων. Το σύνολο της εκπαίδευσης είναι ένα (συνήθως πολύ) μικρό υποσύνολο του συνόλου των δειγμάτων, οπότε δημιουργούνται κάποια προβλήματα. Ένα από αυτά είναι το πρόβλημα της υπερπροσαρμογής (overfitting)[11]. Αν φτιάξουμε ένα αρκετά περίπλοκο και βαθύ μοντέλο και το εκπαιδεύσουμε υπερβολικά στο σύνολο εκπαίδευσης, είναι πιθανό να εκπαιδευτεί τέλεια ή σχεδόν τέλεια σε αυτό το σύνολο. Όταν όμως το δοκιμάσουμε στο σύνολο της δοκιμής, μάλλον θα δούμε αποτελέσματα πολύ χειρότερα. Αυτό ονομάζεται υπερπροσαρμογή, δηλαδή η ικανότητα ενός μοντέλου να αναλύει καλά ένα πολύ συγκεκριμένο υποσύνολο του συνόλου δειγμάτων αλλά να έχει κακή ικανότητα να γενικεύσει στο ευρύτερο σύνολο. Το αντίθετο, δηλαδή η ανικανότητα ενός μοντέλου να μάθει να αναλύει καλά το σύνολο εκπαίδευσης, ονομάζεται underfitting. Αυτό μπορεί να οφείλεται σε ανεπαρκές/κακοφτιαγμένο σύνολο εκπαίδευσης, μοντέλο που δεν είναι αρκετά σύνθετο για να πιάσει τις σχέσεις και τα μοτίβα στα δεδομένα εκπαίδευσης ή σε μικρό χρόνο εκπαίδευσης.



Τα μπλε και τα κόκκινα είναι τα δείγματα εκπαίδευσης 2 κλάσεων. Η πράσινη γραμμή είναι η διαχωριστική γραμμή ενός μοντέλου που έχει υποστεί overfitting. Ενώ διαχωρίζει τέλεια τα δεδομένα εκπαίδευσης, μάλλον θα έχει πρόβλημα γενίκευσης. Η πορτοκαλή γραμμή είναι από ένα μοντέλο που έχει κάνει underfitting. Δεν μπορεί να διαχωρίσει καλά τα δεδομένα. Η μαύρη γραμμή είναι από ένα μοντέλο που έχει μάθει σε καλό βαθμό το σύνολο εκπαίδευσης και φαίνεται ότι θα μπορεί να γενικεύσει αρκετά καλά.

3.2 Bias

Το νευρωνικό δίκτυο προσπαθεί να εντοπίσει μοτίβα στο σύνολο εκπαίδευσης προκειμένου να γενικεύσει και να μπορεί να κάνει σωστές εκτιμήσεις σε άγνωστα δείγματα. Όσο μεγάλο όμως και να είναι το σύνολο εκπαίδευσης, δεν μπορεί να συμπεριλαμβάνει όλες τις περιπτώσεις της κατανομής του δείγματος. Αυτό σημαίνει ότι κάποια μοτίβα δε θα τα εντοπίσει, ενώ μπορεί να εντοπίσει κάποια μοτίβα που ανταποκρίνονται στα δεδομένα της εκπαίδευσης αλλά όχι στην πραγματικότητα.

Ένα τυχαίο παράδειγμα είναι ένα νευρωνικό δίκτυο που ξεχωρίζει ανθρώπους από ζώα. Στο σύνολο εκπαίδευσης για την κατηγορία των ανθρώπων, έχουμε δώσει στο νευρωνικό δίκτυο ένα σύνολο από φωτογραφίες από μερικά κέντρα φωτογράφισης ταυτότητας ενώ για τα ζώα έχουν παρθεί τυχαίες φωτογραφίες από το ίντερνετ. Όλες οι ανθρώπινες φωτογραφίες λόγω της επιλογής που κάναμε, θα έχουν λευκό background οπότε το νευρωνικό μας δίκτυο θα μάθει ότι το λευκό background ισοδυναμεί με άνθρωπο. Αυτό ενώ μπορεί να είναι αληθές για τα δεδομένα εκπαίδευσης, προφανώς δεν ανταποκρίνεται στην πραγματικότητα και το νευρωνικό μας δίκτυο θα έχει πρόβλημα να γενικεύσει στο σύνολο της κατανομής.

Όταν γίνεται κάτι τέτοιο, λέμε ότι το σύνολο εκπαίδευσης που έχουμε έχει κάποιο bias και πρέπει κάπως να το αντιμετωπίσουμε αν θέλουμε να εκπαιδευτεί σωστά το μοντέλο μας. Υπάρχουν διάφορες τεχνικές regularization προκειμένου να βοηθήσουμε το μοντέλο μας με το bias και το overfitting. Η πιο διαδεδομένη είναι το dropout κατά την οποία αποφεύγονται πολύπλοκες υπερπροσαρμογές βασιζόμενες σε σύνθετα μοτίβα, απορρίπτοντας τυχαία κάποια βάρη ανά στρώμα κατά τη διαδικασία εκπαίδευσης κάθε στοιχείου.

3.3 Image Augmentation

Η επαύξηση των εικόνων[12] είναι μια τεχνική που χρησιμοποιείται για την εκπαίδευση νευρωνικών δικτύων όταν το σύνολο δεδομένων που έχουμε στη διάθεσή μας αποτελείται από εικόνες. Αυξάνοντας το σύνολο των διαθέσιμων εικόνων παράγοντας νέες εικόνες παραπλήσιες με τις ήδη υπάρχουσες, μπορούμε να εκπαιδεύσουμε καλύτερα το μοντέλο μας και να το βοηθήσουμε να γενικεύσει με μεγαλύτερη επιτυχία. Οι πιο απλές τεχνικές που εφαρμόζονται είναι η τυχαία περιστροφή των εικόνων κατά κάποια μοίρες προκειμένου να φαίνεται η εικόνα και από άλλες οπτικές. Μια άλλη τεχνική είναι το αναποδογύρισμα (flipping) το οποίο μπορεί να γίνει ως προς τον άξονα Χ, ως προς τον άξονα Υ ή ως προς την αρχή των αξόνων. Μια άλλη τεχνική είναι η κλιμάκωση (scaling) κατά την οποία η εικόνα μεγεθύνεται και κόβεται περιμετρικά προκειμένου το μέγεθος να παραμείνει ίδιο ή συρρικνώνεται και τα εξωτερικά πίξελ γεμίζονται κατά προσέγγιση. Είναι προφανές ότι για να μεγεθύνουμε μια εικόνα πρέπει το τμήμα της το οποίο μας ενδιαφέρει να μη βρίσκεται στις άκρες της εικόνας αλλιώς θα χάσουμε πληροφορία. Αντίστοιχα για να γίνει η συρρίκνωση, πρέπει η εικόνα

περιμετρικά να είναι εύκολο να γεμίσει με κάτι που να μην αποκλίνει από την αρχική εικόνα, π.χ. η εικόνα να είναι μονοχρωματική περιμετρικά. Μια ακόμα γνωστή τεχνική είναι η Translation κατά την οποία η εικόνα μετακινείται κατά μήκος του άξονα X ή Y και τα νέα τμήματα γεμίζονται πάλι προσεγγιστικά. Όλες αυτές οι τεχνικές μπορούν να εφαρμοστούν σε ένα σύνολο δεδομένων προκειμένου να το κάνουν πιο πλούσιο και έτσι να βοηθήσουν το νευρωνικό δίκτυο να μάθει καλύτερα. Αυτές οι τεχνικές όμως, δεν εφαρμόζονται σε όλες τις περιπτώσεις οπότε πρέπει να αξιολογούμε το dataset που έχουμε και να εκτιμούμε ποιες θα μας βοηθήσουν και ποιες θα έχουν ουδέτερο ή ακόμα και αρνητικό αποτέλεσμα στην εκπαίδευση του δικτύου μας.

Horizontal και vertical flip μιας εικόνας ενός αυτοκινήτου

Κεφάλαιο 4: Παραγωγικά Νευρωνικά Δίκτυα

4.1 Παραγωγικά Ανταγωνιστικά Δίκτυα

Την τελευταία δεκαετία έχει γίνει ραγδαία ανάπτυξη στον κλάδο των νευρωνικών δικτύων. Μεγάλη έμφαση έχει δοθεί στον κλάδο των παραγωγικών νευρωνικών δικτύων όπου έχουν εισαχθεί πολλά νέα μοντέλα. Ένα από τα πιο σημαντικά είναι α Παραγωγικά Ανταγωνιστικά Δίκτυα (Generative Adversarial Networks - GAN). Τα GAN είναι μια κατηγορία συστημάτων μηχανικής μάθησης που εφευρέθηκε από τον Ian Goodfellow και τους συναδέλφους του το 2014. Βασίζονται στην λογική της αντιπαλικής μάθησης. Δύο νευρωνικά δίκτυα διαγωνίζονται σε ένα παίγνιο. Δοθέντος ενός συνόλου εκπαίδευσης, αυτή η τεχνική μαθαίνει να δημιουργεί νέα δεδομένα με τα ίδια στατιστικά στοιχεία. Για παράδειγμα, ένα GAN εκπαιδευμένο σε φωτογραφίες, μπορεί να δημιουργήσει νέες φωτογραφίες που φαίνονται αυθεντικές στους ανθρώπινους παρατηρητές, έχοντας πολλά ρεαλιστικά χαρακτηριστικά. Αν και αρχικά προτάθηκαν αμιγώς ως μορφή παραγωγικού μοντέλου για εφαρμογές μη επιβλεπόμενης μάθησης, τα GAN έχουν επίσης αποδειχθεί χρήσιμα για την ημιεπιβλεπόμενη μάθηση, την επιβλεπόμενη μάθηση και την ενισχυτική μάθηση.

Το δίκτυο αυτό αποτελείται από δύο υποδίκτυα, το παραγωγικό (generator) και το διαχωριστικό (classifier). Το παραγωγικό δίκτυο δημιουργεί υποψηφίους (candidates) ενώ το διαχωριστικό δίκτυο τους αξιολογεί. Το παραγωγικό δίκτυο μαθαίνει να προβάλλει από έναν latent space σε μια επιθυμητή κατανομή δεδομένων, ενώ το διαχωριστικό δίκτυο διακρίνει τους candidates από την πραγματική κατανομή. Ο στόχος εκπαίδευσης του παραγωγικού δικτύου είναι η αύξηση του ποσοστού σφάλματος του διακριτικού δικτύου δηλαδή η παραγωγή τόσο καλών candidates ώστε το διακριτό δίκτυο να μην μπορεί εύκολα να ξεχωρίσει αν ανήκουν στην πραγματική κατανομή ή όχι.

Τα GANs χωρίζονται σε διάφορες υποκατηγορίες. Αρχικά, υπάρχει το κλασικό μοντέλο το οποίο είναι αυτό που περιγράψαμε. Μια άλλη αρκετά δημοφιλής αρχιτεκτονική είναι το CycleGAN[13] το οποίο μαθαίνει μετασχηματισμούς ανάμεσα σε εικόνες όπως για παράδειγμα χειμωνιάτικες φωτογραφίες σε καλοκαιρινές φωτογραφίες. Την ονομασία του την έχει πάρει από το γεγονός ότι προσπαθεί να μάθει και το αντίστροφο ταίριασμα, από την παραγόμενη

εικόνα στην αρχική. Συνεπώς έχει δύο γεννήτορες. Για να εκπαιδεύσουμε ένα CycleGAN είναι εμφανές ότι χρειαζόμαστε 2 κατηγορίες δεδομένων, άρα θέλουμε labeled data.

Τα GAN είναι implicit generative models, που σημαίνει ότι δεν μοντελοποιούν ρητά τη συνάρτηση πιθανότητας ούτε δίνουν τρόπους για την εύρεση της latent variable που αντιστοιχεί σε ένα δεδομένο δείγμα, σε αντίθεση με εναλλακτικές λύσεις όπως το γενετικό μοντέλο FLOW. Με άλλα λόγια, δεν μπορούμε από την παραγόμενη εικόνα να επιστρέψουμε με ακρίβεια στο latent code, μόνο κατά προσέγγιση.

4.2 Latent Space

Ένας latent space, είναι μια ενσωμάτωση ενός συνόλου στοιχείων μέσα σε ένα manifold στο οποίο αντικείμενα που μοιάζουν περισσότερο μεταξύ τους τοποθετούνται πιο κοντά. Η θέση μέσα στον latent space μπορεί να θεωρηθεί ότι ορίζεται από ένα σύνολο latent variables που προκύπτουν από τις ομοιότητες των αντικειμένων. Στις περισσότερες περιπτώσεις, η διάσταση του latent space επιλέγεται να είναι μικρότερη από τη διάσταση του χώρου των χαρακτηριστικών από τον οποίο αντλούνται τα σημεία των δεδομένων. Υπάρχει ένας αριθμός αλγορίθμων για τη δημιουργία ενσωματώσεων latent space, δεδομένου ενός συνόλου στοιχείων δεδομένων και μιας συνάρτησης ομοιότητας. Στα GAN που παράγουν εικόνες, ο latent space είναι διάνυσμα ενώ οι εικόνες έχουν 2 διαστάσεις (grayscale). Στα Flows, η διάσταση του latent space είναι ίδια με τη διάσταση των εικόνων που παράγει το δίκτυο.

Έχει γίνει μεγάλη απόπειρα προκειμένου να βρεθεί τρόπος να μπορούμε να αντιστρέφουμε την παραγόμενη εικόνα στον latent space και μετά να την ξανακατασκευάζουμε χωρίς απώλειες[14]. Υπάρχουν πολλοί λόγοι για τους οποίους μπορεί να θέλουμε κάτι τέτοιο. Ένας από αυτούς είναι ότι αν η διάσταση του latent space είναι μικρότερη από της εικόνας, μπορούμε έτσι να πετύχουμε τεράστια συμπίεση της εικόνας με αυτόν τον τρόπο. Ένας άλλος είναι ότι μπορούμε από το latent space να τροποποιήσουμε γρήγορα την εικόνα. Κοιτώντας, για παράδειγμα, από ένα GAN το latent code εικόνων ανθρώπων που χαμογελάνε, μπορούμε να δούμε κάποια μοτίβα, να πάρουμε την εικόνα ενός ανθρώπου που δε χαμογελάει, να την αντιστρέψουμε και τροποποιώντας κατάλληλα το latent code της εικόνας, να ξανακατασκευάσουμε την εικόνα και πλέον ο άνθρωπος αυτός να χαμογελάει

χωρίς να έχουν αλλοιωθεί άλλα χαρακτηριστικά του. Έχουν γίνει έρευνες προκειμένου να βρεθούν κατευθύνσεις και ερμηνεύσιμοι έλεγχοι που να τροποποιούν μια εικόνα χρησιμοποιώντας GAN[15]. Όπως αναφέρθηκε όμως, τα GAN δεν μπορούν από την παραγόμενη εικόνα να επιστρέψουν στο latent space με 100% ακρίβεια. Μάλιστα, δυσκολεύονται πολύ να κάνουν κάτι τέτοιο για εικόνες που δεν έχουν συντεθεί από το ίδιο το GAN. Οι προσεγγίσεις μπορεί να είναι αρκετά ικανοποιητικές όταν ενδιαφερόμαστε για τροποποίηση της εικόνας κάποιου ανθρώπου. Όταν όμως επεξεργαζόμαστε ιατρικά δεδομένα, μια απλή προσέγγιση δεν αρκεί. Αυτό το πρόβλημα έρχονται να λύσουν τα Flows τα οποία εξ ορισμού μοντελοποιούν ρητά τη συνάρτηση πιθανότητας και δίνουν μια πλήρη αντιστοιχία ανάμεσα στο latent space και στον χώρο των δειγμάτων,

4.3 Flow-Based Generative Model

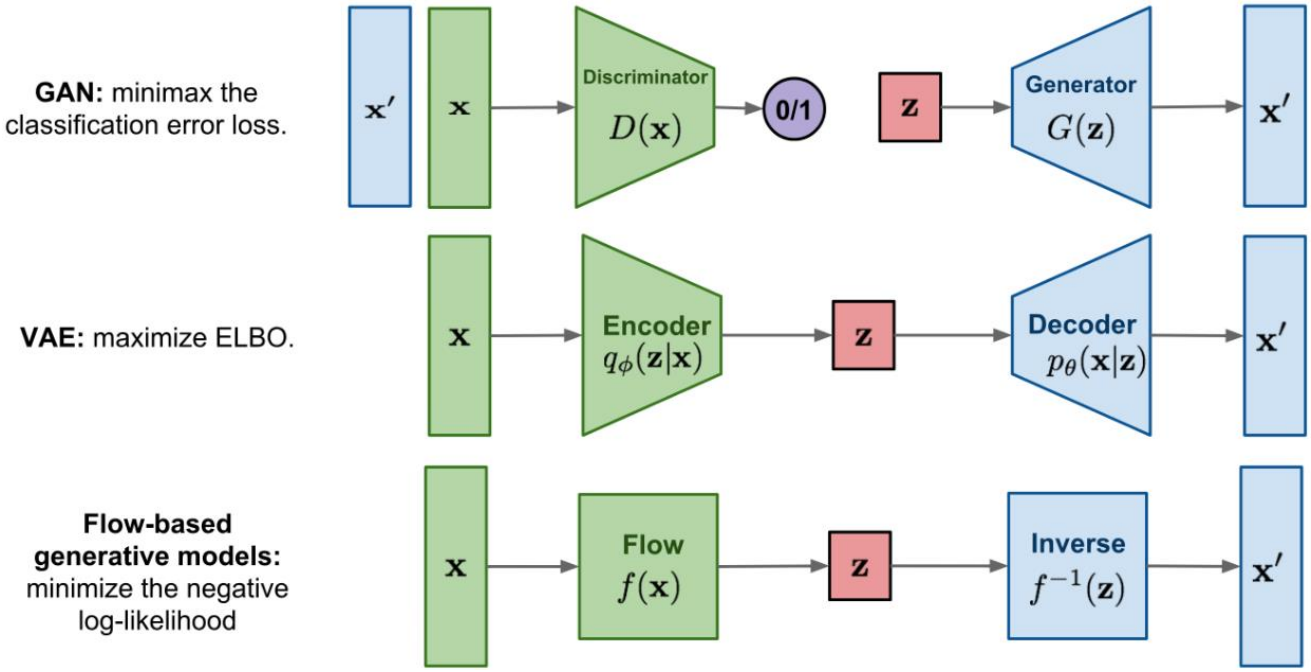
Ένα Flow είναι ένα παραγωγικό νευρωνικό δίκτυο το οποίο μοντελοποιεί μια συνάρτηση κατανομής χρησιμοποιώντας μια ροή κανονικοποίησης (normalizing flow), η οποία είναι μια στατιστική μέθοδος που χρησιμοποιεί τον κανόνα της αλλαγής μεταβλητής προκειμένου να μετατρέψει μια απλή κατανομή σε μια σύνθετη. Καινούρια δείγματα μπορούν να παραχθούν παίρνοντας τυχαία δείγματα από την αρχική κατανομή και περνώντας τα από το Flow. Το βασίκο πλεονέκτημα που έχει αυτό το μοντέλο είναι ότι μαθαίνει ακριβώς τη συνάρτηση πιθανοφάνειας, αντίθετα με τα GAN και τα Variational autoencoder (VAE).

Έστω z_0 μια τυχαία (ενδεχομένως πολυδιάστατη) μεταβλητή με κατανομή $p_0(z_0)$. Έστω μια ακολουθία $z_1,z_2,...,z_k$ η οποία έχει προέλθει με διαδοχικούς μετασχηματισμούς του z_0 μέσω των συναρτήσεων $f_1,f_2,...,f_k$ δηλαδή $f_i(z_{i-1})=z_i$. Οι συναρτήσεις f_i πρέπει να είναι αντιστρέψιμες. Η τελική έξοδος του μοντέλου είναι το z_k το οποίο μοντελοποιεί την κατανομή-στόχος. Η λογαριθμική πιθανοφάνεια του z_k είναι:

$$\log p_K(z_K) = \log p_0(z_0) - \sum_{i=1}^{K} \log \left| \det \frac{df_i(z_{i-1})}{dz_{i-1}} \right|$$

Για να υπολογίζεται εύκολα, θέλουμε η συνάρτηση να είναι εύκολα αντιστρέψιμη και να υπολογίζεται εύκολα η ορίζουσα του Ιακωβιανού πίνακα. Συνήθως οι συναρτήσεις f μοντελοποιούνται χρησιμοποιώντας βαθιά

νευρωνικά δίκτυα και εκπαιδεύονται προκειμένου να ελαχιστοποιούν την αρνητική λογαριθμική πιθανοφάνεια των δειγμάτων της αρχικής κατανομής από την κατανομή-στόχος. Αυτά τα μοντέλα συνήθως σχεδιάζονται με αρχιτεκτονικές που επιτρέπουν τον υπολογισμό της Ιακωβιανής ορίζουσας και της αντίστροφης συνάρτησης μόνο με πέρασμα προς τα εμπρός (forward pass). Μερικά παραδείγματα τέτοιων αρχιτεκτονικών είναι τα NICE[16], RealNVP και Glow. Στην εργασία αυτήν ασχοληθήκαμε περισσότερο με το τελευταίο.



Κεφάλαιο 5: Glow

5.1 Glow: Generative Flow with Invertible 1×1 Convolutions

Η αρχιτεκτονική του μοντέλου Glow βασίζεται στην αρχιτεκτονική των προηγούμενων γνωστών μοντέλων NICE και RealNVP. Το glow εφαρμόζει 3 βήματα:

1)Activation Normalization

Εφαρμόζει έναν συγγενή μετασχηματισμό χρησιμοποιώντας κλίμακα και bias για κάθε κανάλι. Οι παράμετροι είναι εκπαιδεύσιμες.

2)Αντιστρέψιμη 1x1 Συνέλιξη

Ανάμεσα στα επίπεδα, η σειρά των καναλιών αντιστρέφεται ώστε να έχουν την ευκαιρία όλα τα διαστατικά δεδομένα να αλλαχθούν. Μια συνέλιξη 1χ1 με ίσο αριθμό καναλιών εισόδου και εξόδου, κάνει ακριβώς αυτό, είναι η γενίκευση κάθε μετάθεσης της σειράς των καναλιών που είχε προταθεί αρχικά. Η λογαριθμική ορίζουσα μιας αντιστρέψιμης 1 × 1 συνέλιξης ενός h × w × c τανυστή **h** με c × c πίνακα βαρών W είναι απλή στον υπολογισμό:

$$\log \left| \det \left(\frac{d\,\mathrm{conv2D}(\mathbf{h}; \mathbf{W})}{d\,\mathbf{h}} \right) \right| = h \cdot w \cdot \log |\det(\mathbf{W})|$$

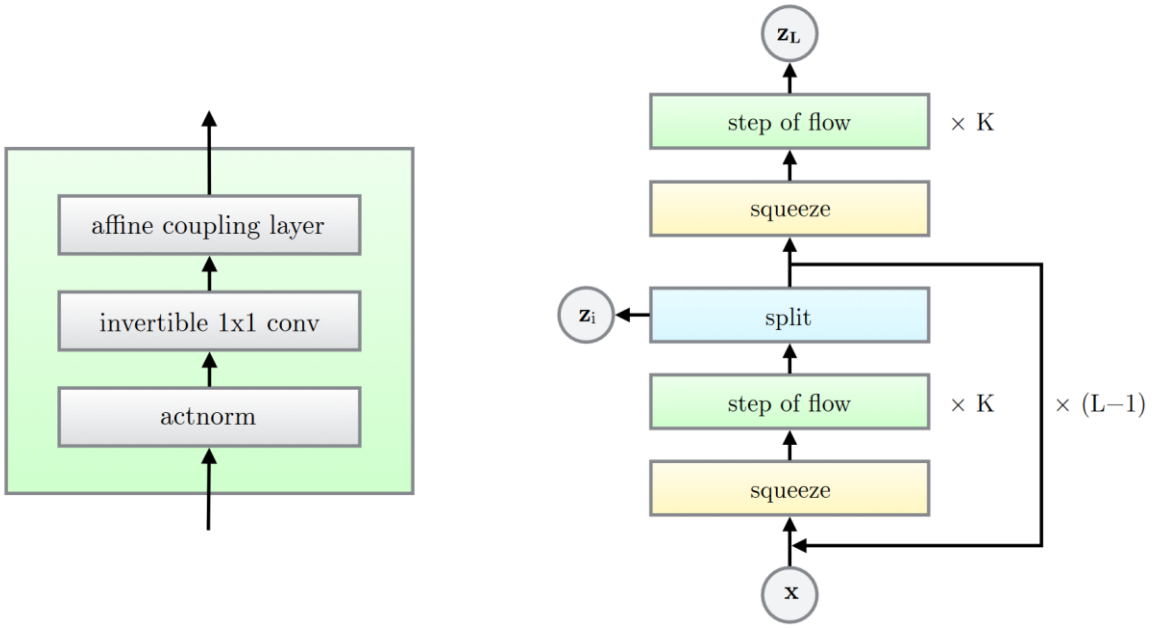
3)Affine Coupling Layer

Για να φτιάξουμε ένα normalizing flow, στοιβάζουμε μια ακολουθία από αντιστρέψιμες συναρτήσεις μετασχηματισμού. Σε κάθε συνάρτηση, η οποία είναι γνωστή και ως συγγενικό στρώμα ζεύξης (affine coupling layer), οι διαστάσεις εισόδου χωρίζονται σε 2 κομμάτια: Οι πρώτες d διαστάσεις μένουν ως έχουν. Οι επόμενες d+1 έως D διαστάσεις περνάνε από έναν συγγενικό μετασχηματισμό (affine transformation) ο οποίος περιλαμβάνει κλιμάκωση και μετατόπιση. Οι παράμετροι της κλιμάκωσης και της μετατόπισης είναι συναρτήσεις των d πρώτων διαστάσεων.

Για το μοντέλο RealNVP, που δεν έχει τη συνέλιξη, ισχύει ότι σε ένα coupling layer ορισμένες διαστάσεις (channels) παραμένουν αμετάβλητες. Για να βεβαιωθούμε ότι όλες οι είσοδοι έχουν την ευκαιρία να αλλάξουν, το μοντέλο

αντιστρέφει τη σειρά σε κάθε επίπεδο έτσι ώστε διαφορετικά στοιχεία να μένουν αμετάβλητα κάθε φορά.

Αυτά τα βήματα (αριστερά) συνδυάζονται με μια αρχιτεκτονική πολλαπλής κλίμακας (δεξιά) προκειμένου να προκύψει το τελικό μοντέλο. Στα πειράματά μας έγιναν δοκιμές με διάφορα μεγέθη K και L προκειμένου να εξαχθούν τα καλύτερα αποτελέσματα σε λογικούς χρόνους εκπαίδευσης.



5.2 Multi-scale Architecture

Ένα μειονέκτημα που έχουν τα flows είναι ότι λειτουργούν με τις ίδιες διαστάσεις που έχει και η είσοδος. Αν η είσοδος έχει πολλές διαστάσεις τότε το ίδιο θα ισχύει και για το latent space άρα θα απαιτείται μεγάλο υπολογιστικό κόστος για να εκπαιδευτεί το συγκεκριμένο δίκτυο. Εκμεταλλευόμενοι όμως το γεγονός ότι πολλά πίξελ είναι μικρής σημασίας για την εκμάθηση, μπορούμε να τα αφαιρέσουμε και να μειώσουμε το μέγεθος της εικόνας. Μετά από K μετασχηματισμούς, χωρίζουμε τις μισές διαστάσεις και τις αξιολογούμε. Τις υπόλοιπες μισές τις περνάμε από K ακόμα μετασχηματισμούς. Η διαδικασία αυτή επαναλαμβάνεται L φορές.

Η εικόνα πρώτα πρώτού χωριστεί (split), συμπιέζεται (squeeze) ως εξής: Η εικόνα χωρίζεται σε τετραγωνάκια μεγέθους 2x2xC και μετασχηματίζονται σε μπλοκ 1x1x4C. Συνεπώς, μειώνεται το μήκος και το πλάτος της εικόνας στη μέση και τα κανάλια τετραπλασιάζονται. Μετά είναι εύκολο να χωρίσουμε την εικόνα χωρίς να χρειαστεί να αναδιατάξουμε τα πίξελ.

5.3 Variational Dequantization

Τα normalizing flows βασίζονται στον κανόνα της αλλαγής των μεταβλητών, ο οποίος ορίζεται φυσικά στον συνεχή χώρο. Η εφαρμογή flow απευθείας σε διακριτά δεδομένα οδηγεί σε μοντέλα ανεπιθύμητης πυκνότητας όπου τοποθετείται αυθαίρετα υψηλή πιθανότητα σε λίγες, συγκεκριμένες τιμές. Για την αποφυγή τέτοιων εκφυλισμένων λύσεων, μια κοινή λύση είναι η προσθήκη μικρής ποσότητας θορύβου σε κάθε διακριτή τιμή, η οποία αναφέρεται επίσης ως αποκβαντικοποίηση (dequantization). Θεωρώντας το χ ακέραιο η αποκβαντικοποιημένη αναπαράσταση u μπορεί να γραφτεί ως u = x + v όπου u ∈[0,1)^D. Ο στόχος μας της μοντελοποίησης του p(x) πλέον γίνεται:

$$p(x) = \int p(x+u)du = \int \frac{q(u|x)}{q(u|x)} p(x+u)du = \mathbb{E}_{u\sim q(u|x)}\left[\frac{p(x+u)}{q(u|x)}\right]$$

Όπου q(u|x) είναι η κατανομή του θορύβου.

Αφού προσθέσουμε τον θόρυβο στις διακριτές τιμές, μετατρέπουμε επιπλέον τον όγκο σε Γκαουσιανό σχήμα. Αυτό γίνεται προσαρμόζοντας την κλίμακα στο x + u μεταξύ του 0 και 1 και εφαρμόζοντας την αντιστροφή της σιγμοειδούς συνάρτησης:

$$\sigma(z)^{-1} = \log z - \log 1 - z.$$

Η αποκβαντικοποίηση χρησιμοποιεί μια ομοιόμορφη κατανομή του θορύβου που οδηγεί αποτελεσματικά στην αναπαράσταση των εικόνων ως υπερκύβων με ευκρινή περιγράμματα. Ωστόσο, η μοντελοποίηση τέτοιων απότομων περιγραμμάτων δεν είναι εύκολη για ένα flow, καθώς χρησιμοποιεί ομαλούς μετασχηματισμούς για να τη μετατρέψει σε Γκαουσιανή κατανομή. Η αποκβαντοποίηση έχει επεκταθεί επομένως σε πιο εξελιγμένες κατανομές που μπορούν να μαθευτούν, πέρα από την ομοιόμορφη, σε ένα μεταβλητό πλαίσιο. Ειδικότερα, αν θυμηθούμε τον στόχο της μάθησης

$$\log p(x) = \log \mathbb{E}_u\left[\frac{p(x+u)}{q(u|x)}\right]$$

η ομοιόμορφη κατανομή μπορεί να αντικατασταθεί από μια μαθημένη κατανομή $q_\theta(u|x)$ με υποστήριξη στο u ∈ [0,1)^D. Αυτή η προσέγγιση ονομάζεται μεταβλητή αποκβαντικοποίηση (Variational Dequantization)[17]. Για να μάθουμε μια τέτοια κατανομή, μπορούμε να χρησιμοποιήσουμε ένα

δεύερο normalizing flow που παίρνει το x ως εξωτερική είσοδο και μαθαίνει μια ευέλικτη κατανομή πάνω στο u. Για να εξασφαλίσουμε υποστήριξη στο [0,1)^D, μπορούμε να εφαρμόσουμε μια σιγμοειδή συνάρτηση ενεργοποίησης ως τελικό μετασχηματισμό flow.

Κεφάλαιο 6: Σχετικά Εγχειρήματα

6.1 PGGAN για Εγκεφαλικές Τομογραφίες

Χρησιμοποιώντας ένα δίκτυο PGGAN[18] (Progressive Growing of GAN), οι ερευνητές κατάφεραν να παράξουν συνθετικά δεδομένα από αξονικές και μαγνητικές τομογραφίες εγκεφάλου[19]. Η σχετική έρευνα είχε δείξει ότι το μοντέλο GAN μάλλον δεν επηρέαζε σημαντικά τα παραγόμενα αποτελέσματα. Οι αξονικές τομογραφίες είχαν χωριστεί σε 3 κλάσεις: cortical CSF, brain stem CSF και ventricular CSF. Αντίθετα οι μαγνητικές αποτελούταν μόνο από μία κλάση. Εκπαίδευσαν τον ταξινομητή με διάφορους τρόπους και με διάφορα ποσοστά των δεδομένων, χρησιμοποιώντας τις κλασικές τεχνικές ενίσχυσης. Τα αποτελέσματα που έλαβαν ήταν αρκετά αληθοφανή και όντως κατάφεραν να ενισχύσουν το σύνολο εκπαίδευσης, σε ορισμένες περιπτώσεις με σημαντική αύξηση. Τα μοντέλο κατηγοριοποίησης που χρησιμοποίησαν ήταν τα Unet, UResNet, Deep Medic τα οποία είναι μοντέλα εξειδικευμένα για ιατρικά δεδομένα όπου το σύνολο εκπαίδευσης είναι μικρό. Το μοντέλο GAN εκπαιδεύτηκε λαμβάνοντας κάθε φορά μια εικόνα με το label που της αντιστοιχούσε έτσι μάθαινε τη συσχέτιση ανάμεσα στην εικόνα και στον τύπο της αξονικής. Τα δεδομένα στα οποία εκπαιδεύτηκε το GAN ήταν και αυτά ενισχυμένα με τους συμβατικούς τρόπους.

6.2 Gan για Ηπατικά Τραύματα

Ένα άλλο σχετικό εγχείρημα που έχει γίνει είναι η παραγωγή ιατρικών δεδομένων από βάση δεδομένων ηπατικών τραυμάτων χρησιμοποιώντας GAN[20]. Οι εικόνες σε αυτήν την έρευνα ήταν αρκετά περιορισμένες, συνολικά 182 και διακρίνονταν σε 3 κλάσεις: Κύστη, Μετάσταση και Αιμαγγείωμα. Προσθέτοντας τα συνθετικά δεδομένα που παρήχθησαν από το μοντέλο τους, κατάφεραν να πετύχουν σημαντική βελτίωση στην ακρίβεια του ταξινομητή που χρησιμοποίησαν. Αυτό δείχνει ότι οι κλασικοί μέθοδοι image augmentation όταν εφαρμοστούν σε συνδυασμό με ενίσχυση του συνόλου δεδομένου μέσω παραγόμενων εικόνων, μπορεί να μας δώσει σπουδαία αύξηση όταν το σύνολο δεδομένων είναι πολύ μικρό.

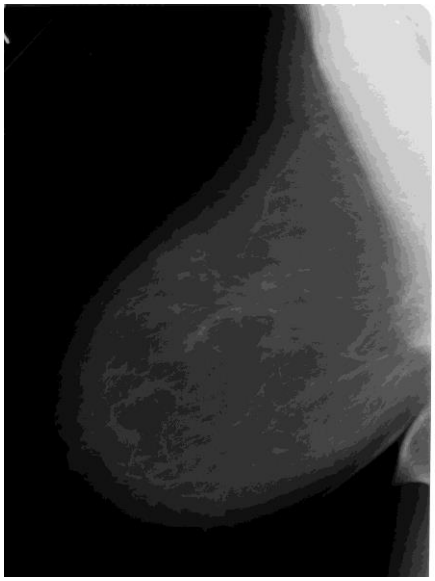
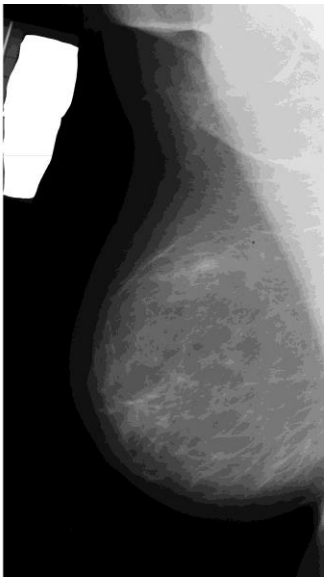
6.3 Κατηγοριοποίηση της Mini-DDSM

Έχει γίνει ήδη μια απόπειρα ταξινόμησης της ιατρικής βάσης δεδομένων mini-DDSM χρησιμοποιώντας προεκπαιδευμένα νευρωνικά δίκτυα[21] όμως με περιορισμένη επιτυχία. Σε αυτήν την έρευνα χρησιμοποιήθηκαν πιο στιβαρά μοντέλα, βασικό image augmentation αλλά χρησιμοποιήθηκε και η κλάση benign. Το καλύτερο accuracy που πέτυχαν ήταν σχεδόν 66%.

Κεφάλαιο 7: Δεδομένα

7.1 Δεδομένα και Βασικές Τροποποιήσεις

Για την εργασία, χρησιμοποιήθηκε η ιατρική βάση Mini-DDSM[22] η οποία περιέχει grayscale φωτογραφίες από μαστογραφίες. Η βάση αυτή περιέχει τριών ειδών μαστογραφίες: Φυσιολογικές (normal), με καλοήθη όγκο (benign) και με καρκίνο (cancer). Στην εργασία μας δεν ασχοληθήκαμε με τις benign φωτογραφίες. Οι περισσότερες μαστογραφίες ήταν αρκετά μεγάλες, το μήκος ήταν 1000-2000 πίξελ ενώ το ύψος 2000-3000 στις περισσότερες. Αυτά τα μεγέθη είναι απαγορευτικά για εκπαίδευση νευρωνικού δικτύου οπότε τις κάναμε downsize και rescale σε 128x128. Οι φωτογραφίες είναι τεσσάρων ειδών: Είτε κατακόρυφες (craniocaudal, CC) είτε πλαϊνές (mediolateral oblique, MLO) για τον αριστερό και τον δεξιό μαστό. Στη βάση μας έχουμε 2716 φωτογραφίες με καρκίνο και 2408 φυσιολογικές. Για κάθε είδος φωτογραφίας, εκπαιδεύσαμε διαφορετικό μοντέλο οπότε οι φωτογραφίες αυτές δεν αναμείχθηκαν. Ένα πρόβλημα που είχαν τα δεδομένα είναι ότι πάνω σε κάποιες μαστογραφίες υπήρχαν τα στοιχεία του εξεταζόμενου τα οποία είχαν σβηστεί. Στην περίπτωση των φωτογραφιών με καρκίνο, τα στοιχεία ήταν σβησμένα με μαύρο χρώμα το οποίο αναμειγνυόταν με το background οπότε δεν υπήρχε ουσιαστικό πρόβλημα. Στην περίπτωση όμως των φυσιολογικών φωτογραφιών, τα στοιχεία είχαν σβηστεί με άσπρο χρώμα (περίπου το 70% των δεδομένων). Αυτό εισήγαγε ένα bias και στο Glow αλλά και στον ταξινομητή που χρησιμοποιήθηκε αργότερα για την αξιολόγηση των παραγόμενων αποτελεσμάτων. Ενδεικτικά 2 δείγματα, αριστερά από το σύνολο του καρκίνου και δεξιά από το φυσιολογικό.

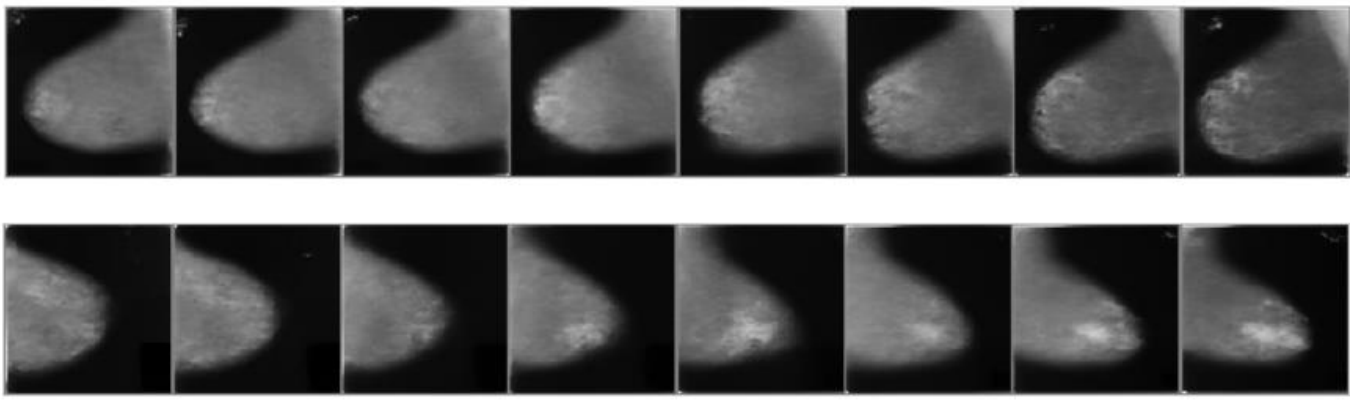



Είναι προφανές ότι ένα τόσο μεγάλο bias που υπάρχει σε ένα τόσο μεγάλο ποσοστό του συνόλου εκπαίδευσης, θα δυσκολέψει πολύ την εκπαίδευση του δικτύου άρα πρέπει με κάποιον τρόπο να το εξαλείψουμε. Οι κλασικές τεχνικές όμως όπως το dropout είναι εμφανές ότι δε θα δουλέψουν μιας και είναι ένα πολύ προφανές μοτίβο.

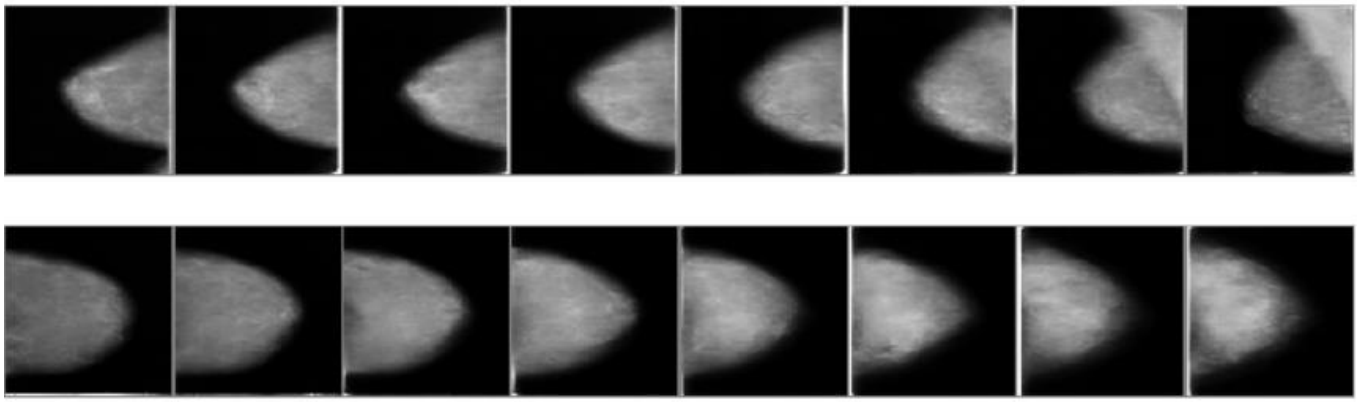
Κεφάλαιο 8: Αρχικά Πειράματα

8.1 Πρώτα Μοντέλα

Στα πρώτα πειράματα που έγιναν χρησιμοποιήθηκαν ρηχά μοντέλα RealNVP. Τα παραγόμενα αποτελέσματα δεν έφεραν ιδιαίτερη ομοιότητα στα δείγματα εκπαίδευσης, ένας παρατηρητής δε θα μπορούσε να καταλάβει ότι πρόκειται για μαστογραφίες. Αυτό ήταν αναμενόμενο σύμφωνα με την υπάρχουσα έρευνα και το αντίστοιχο δημοσίευμα μιας και τα μικρό K και L δεν επέτρεψαν στο μοντέλο να μάθει να παράγει μια τέτοια σύνθετη εικόνα. Στη συνέχεια έγινε μια απόπειρα να παραχθούν αποτελέσματα χρησιμοποιώντας την τεχνική interpolation. Οι εικόνες που παράχθηκαν έδειξαν να έχουν προοπτικές αλλά προφανώς έμοιαζαν αρκετά με τις προϋπάρχουσες.



Interpolated εικόνες από το σύνολο των κανονικών μαστογραφιών

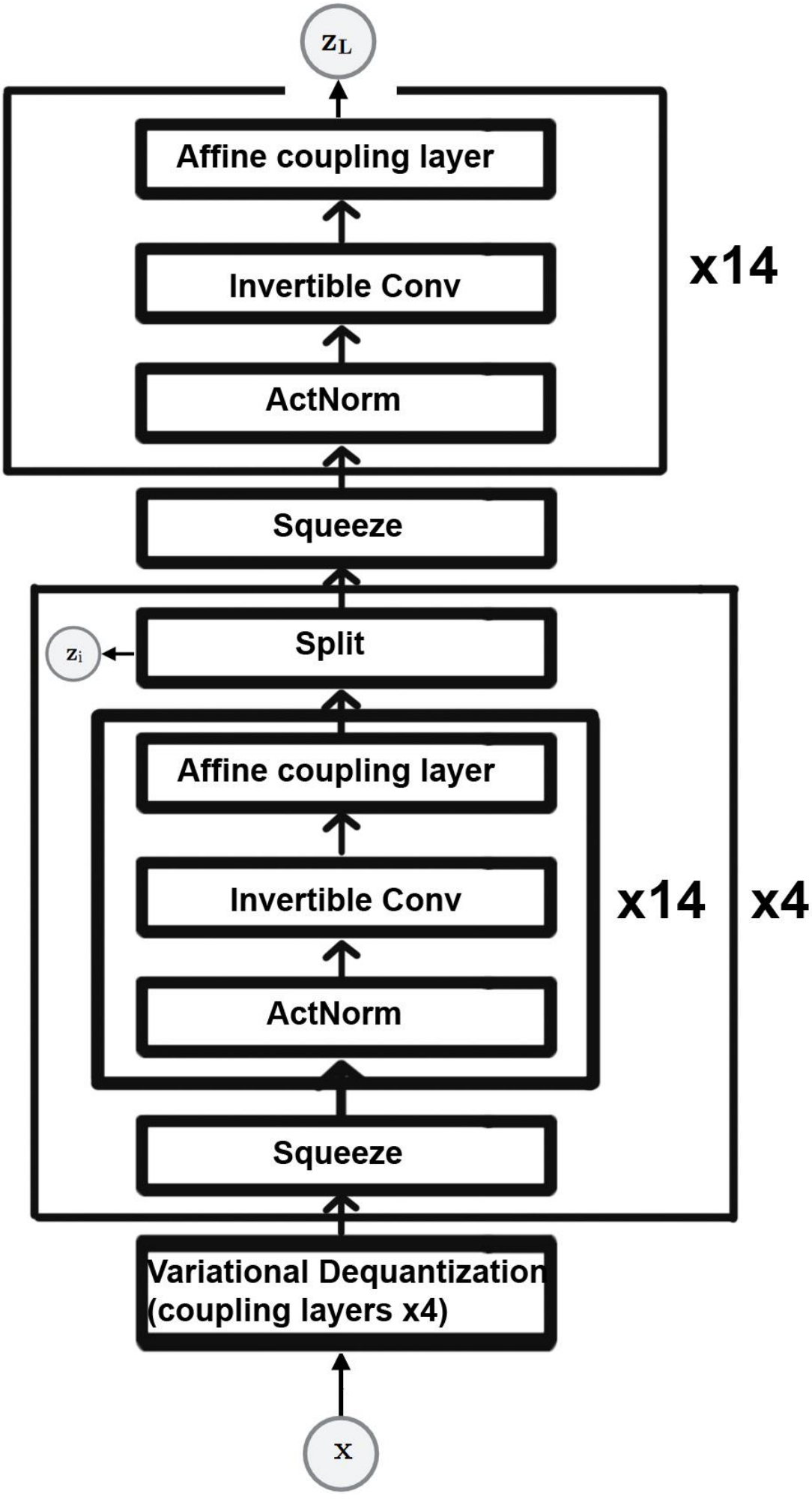


Interpolated εικόνες από το σύνολο των μαστογραφιών με καρκίνο

8.2 Το Βασικό μοντέλο

Στη συνέχεια υλοποιήθηκαν βαθύτερα μοντέλα τύπου GLOW τα οποία εμφάνισαν αμέσως καλύτερα αποτελέσματα. Δοκιμάστηκαν μοντέλα με L = 4, 5 και K με διάφορες τιμές από 4 μέχρι 24. Τα καλύτερα αποτελέσματα παρατηρήθηκαν για L=5 και K=14. Εάν είχαμε στη διάθεσή μας πιο ισχυρές κάρτες γραφικών, θα μπορούσαμε ενδεχομένως να δοκιμάζαμε και μεγαλύτερα L. Ο λόγος που δεν έγινε αυτό είναι επειδή για μεγαλύτερα L, μια εικόνα μεγέθους 128χ128 δε θα παρουσίαζε καλά αποτελέσματα λόγω του split and squeeze, θα έπρεπε να είχαμε μεγαλύτερες εικόνες.

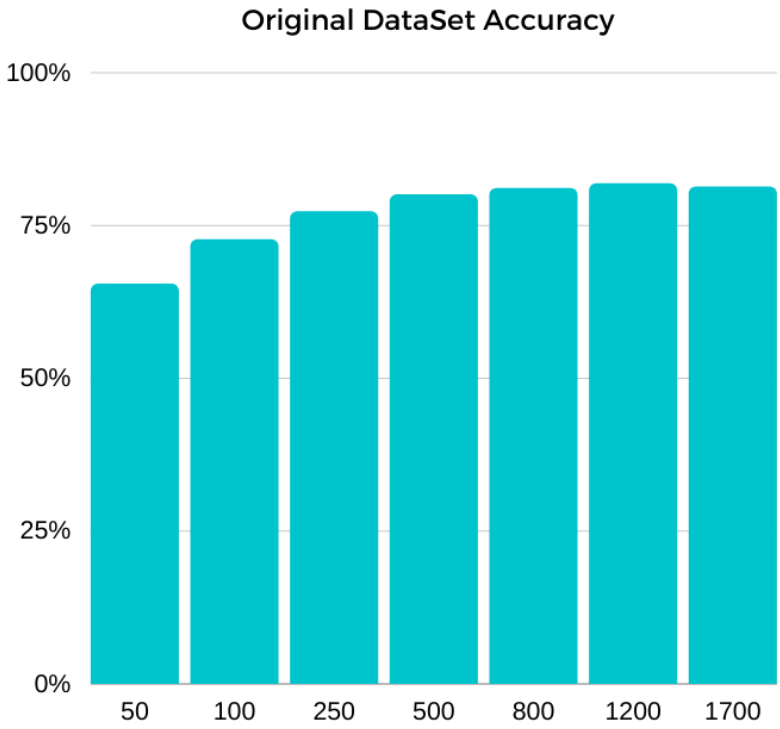
Το τελικό μοντέλο μας φαίνεται ακολούθως:


z_L
Affine coupling layer
Invertible Conv
ActNorm
x14
Squeeze
z_i
Split
Affine coupling layer
Invertible Conv
ActNorm
x14
x4
Squeeze
Variational Dequantization
(coupling layers x4)
x

8.3 Το πλήρες Dataset

Αρχικά, πήραμε το σύνολο των δεδομένων, το χωρίσαμε σε training set και test set και χρησιμοποιήσαμε ένα προεκπαιδευμένο μοντέλο, το MobileNetV2[23] χωρίς το κορυφαίο επίπεδο, με το οποίο προσπαθήσουμε να ταξινομήσουμε τα δεδομένα στις 2 κλάσεις: με καρκίνο και χωρίς καρκίνο. Παρακάτω φαίνεται ένας πίνακας με το κάθε πείραμα που έγινε και τα αποτελέσματα που πήραμε:

Training Set Size	Score
50	65.38%
100	72.68%
250	77.25%
500	80.04%
800	81.07%
1200	81.82%
1700	81.27%

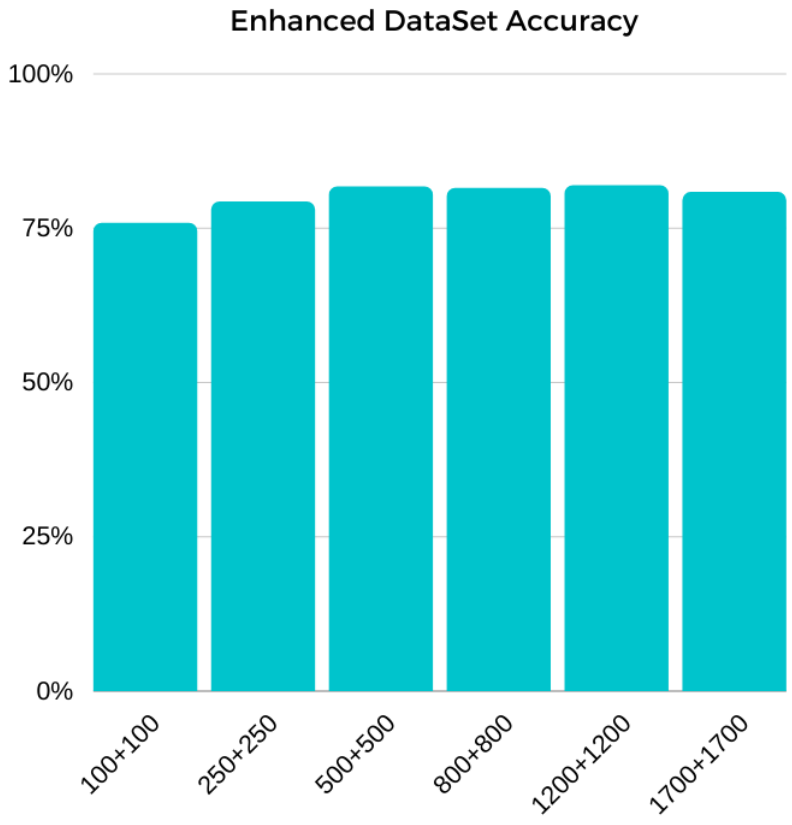


Η ακρίβεια του αρχικού συνόλου εκπαίδευσης με 50 έως και 1700 μαστογραφίες ανά κατηγορία

Το validation set ήταν 700 εικόνες από κάθε κατηγορία.

Στη συνέχεια, εμπλουτίσαμε τα training set χρησιμοποιώντας αποτελέσματα που παρήχθησαν από τα μοντέλα μας και πήραμε νέες τιμές. Αυτές φαίνονται ακολούθως:

100+100	75.80%
250+250	79.25%
500+500	81.68%
800+800	81.42%
1200+1200	81.88%
1700+1700	80.78%

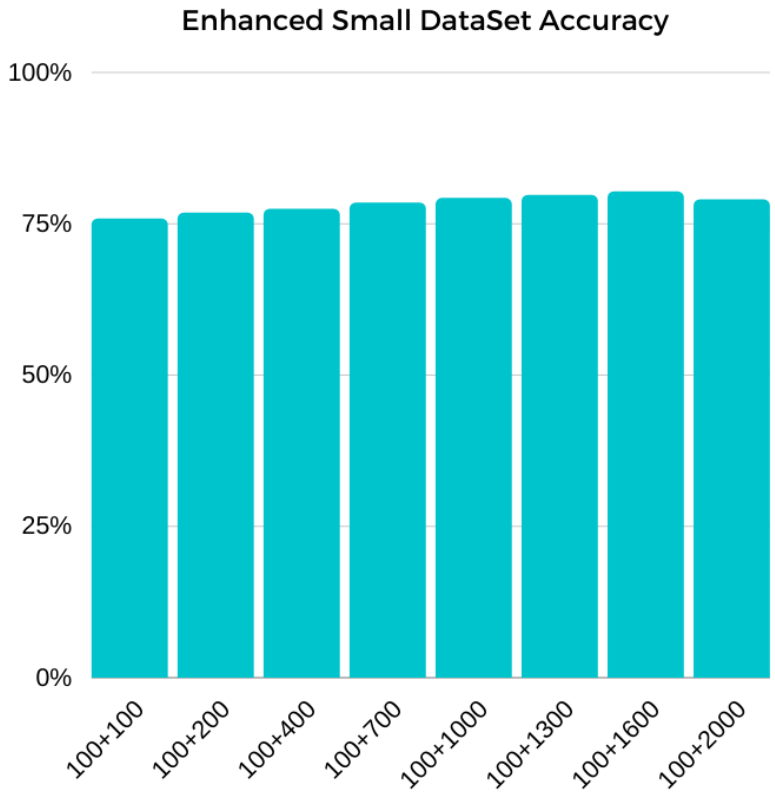


Η ακρίβεια του αρχικού training set με διάφορες ποσότητες και διάφορες ενισχύσεις

Το validation set ήταν το οι ίδιες 700 εικόνες με πριν. Είναι εμφανές ότι βλέπουμε μια σημαντική βελτίωση στις περιπτώσεις που το training set είναι πιο μικρό αλλά φαίνεται να βρίσκουμε ένα πάνω όριο όταν το training set μεγαλώνει.

100+100	75.80%
100+200	76.76%
100+400	77.41%
100+700	78.43%
100+1000	79.17%

100+1300	79.67%
100+1600	80.27%
100+2000	78.95%



100 εικόνες του αρχικού training set με ενισχύσεις από 100 εικόνες έως και 2000

Χρησιμοποιώντας ένα μικρό υποσύνολο του training set, βλέπουμε μια καλή αύξηση μέχρι τις 1000 εικόνες. Βέβαια, το μοντέλο μας έχει δει όλο το dataset.

Στη συνέχεια δοκιμάσαμε να εκπαιδεύσουμε ένα μοντέλο ταξινόμησης με 5000 παραγόμενες εικόνες χωρίς καμία εικόνα από το αρχικό training set. Το accuracy που έπιασε ήταν 79.41%.

8.4 Η επιβεβαίωση του bias

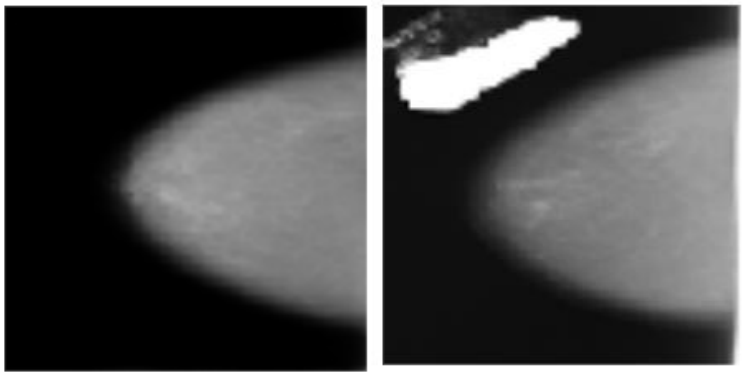
Αυτό το όριο γύρω στο 81-82% πιθανότατα οφείλεται στο άσπρο χρώμα που υπάρχει και δημιουργεί ένα bias στο νευρωνικό δίκτυο ταξινόμησης. Υπενθυμίζουμε ότι οι normal εικόνες με το άσπρο χρώμα είναι γύρω στο 75%. Για να επιβεβαιώσουμε την ύπαρξη του bias, πήραμε προσεγγιστικά το υποσύνολο των εικόνων με το λευκό χρώμα, ισάριθμες εικόνες με καρκίνο, το χωρίσαμε σε training set και validation set και εκπαιδεύσαμε το δίκτυο

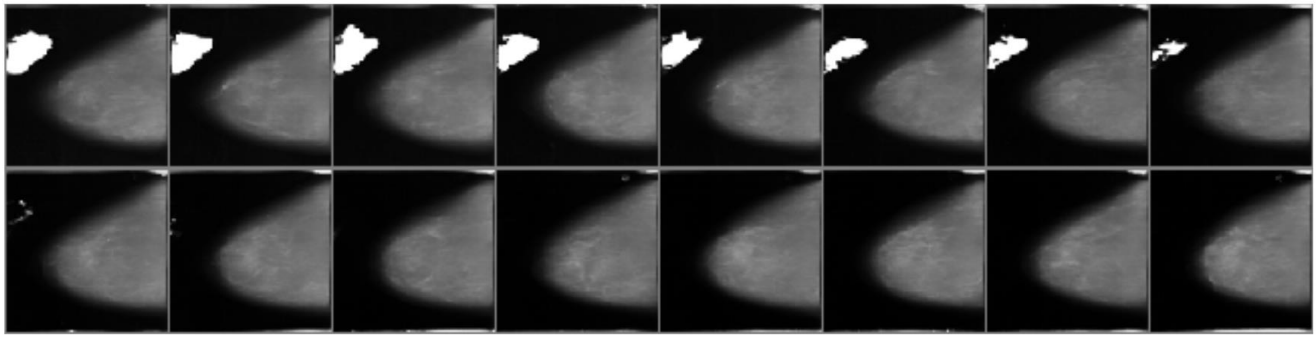
ταξινόμησης. Επέτυχε accuracy 91.15% επιβεβαιώνοντας έτσι τις υποψίες μας. Το μοντέλο μαθαίνει περισσότερο να ξεχωρίζει τις φυσιολογικές βρίσκοντας το άσπρο χρώμα παρά εντοπίζοντας τον καρκίνο.

Στη συνέχεια κάναμε το ίδιο στο εναπομείναν subset το οποίο πάλι χωρίσαμε σε training set (600 εικόνες) και validation set (112 εικόνες) και εκπαιδεύσαμε το μοντέλο ταξινόμησης. Αυτήν τη φορά το accuracy ήταν 72.47%.

8.5 Latent Space Manipulation για Εξάλειψη του Bias

Προκειμένου να αφαιρέσουμε το bias, κάναμε latent space manipulation. Πήραμε όλες τις εικόνες χωρίς καρκίνο(ανά τύπου μαστογραφίας) που είχαν το άσπρο σβήσιμο (Znegative), τις μεταφέραμε σε latent space, πήραμε τον μέσο όρο τους και μετά κάναμε το ίδιο στις μαστογραφίες χωρίς το bias (Zpositive). Στη συνέχεια χρησιμοποιήσαμε την κατεύθυνση Zpos-Zneg για να κάνουμε manipulate τις εικόνες και έτσι πήραμε τις φωτογραφίες με το άσπρο και το αφαιρέσαμε.





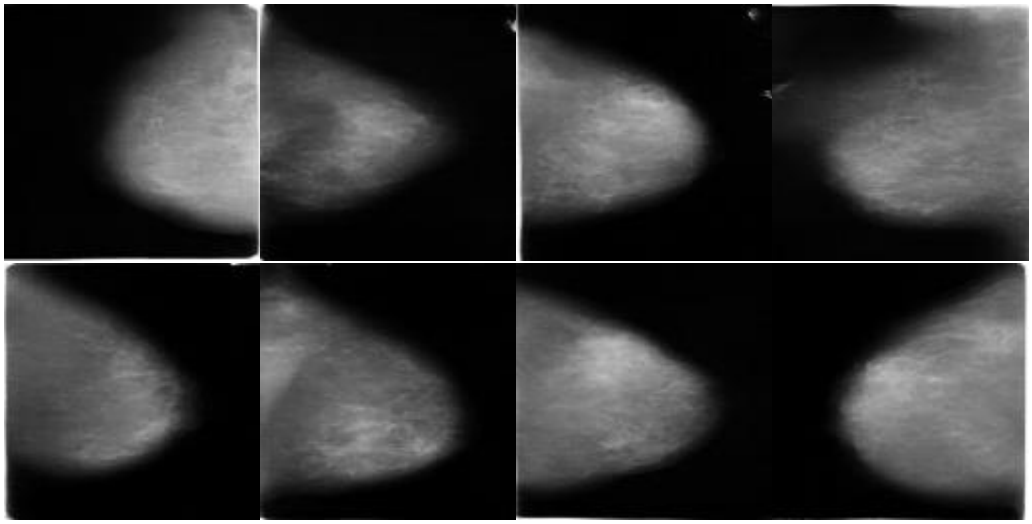
1) Η εικόνα που προκύπτει από το latent space του Zpos
2) Η εικόνα που προκύπτει από το latent space του Zneg
3) Latent Space Manipulation στην κατεύθυνση Zpos-Zneg

Ύστερα έγιναν πειράματα ταξινόμησης με το υποσύνολο που δεν είχε το bias και στη συνέχεια το ίδιο υποσύνολο ενισχύθηκε με τις manipulated εικόνες. Τα αποτελέσματα αυτήν τη φορά ήταν ελαφρώς χειρότερα. Πιθανότατα οφείλεται στο γεγονός ότι και η αρχική εικόνα επηρεάζεται ελαφρώς το οποίο κάνει την εκπαίδευση χειρότερη.

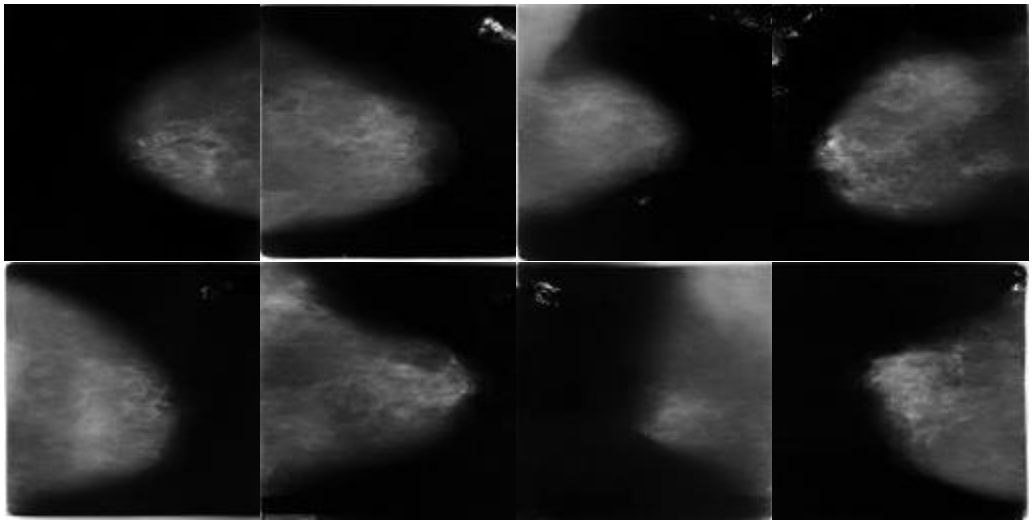
Κεφάλαιο 9: Πειράματα Χωρίς το Bias

9.1 Εκπαίδευση μόνο με τις καθαρές εικόνες

Στη συνέχεια εκπαιδεύσαμε ένα μοντέλο το οποίο είχε ως training set το υποσύνολο των normal εικόνων χωρίς το bias. Εκπαιδεύτηκε σε 600 εικόνες οι οποίες πέρασαν από ένα augmentation προκειμένου το training set του μοντέλου μας να είναι λίγο μεγαλύτερο. Αντίστοιχη εκπαίδευση είχε και ένα δεύτερο μοντέλο με εικόνες καρκίνου, αν και εδώ δεν υπήρχε το bias, πήραμε 600 εικόνες για να είναι δίκαιη η εκπαίδευση και στη συνέχεια η ενίσχυση. Το μοντέλο μας αυτή τη φορά δεν είχε δει καθόλου το test set. Ακολούθως φαίνονται τα αποτελέσματα αυτού του πειράματος:



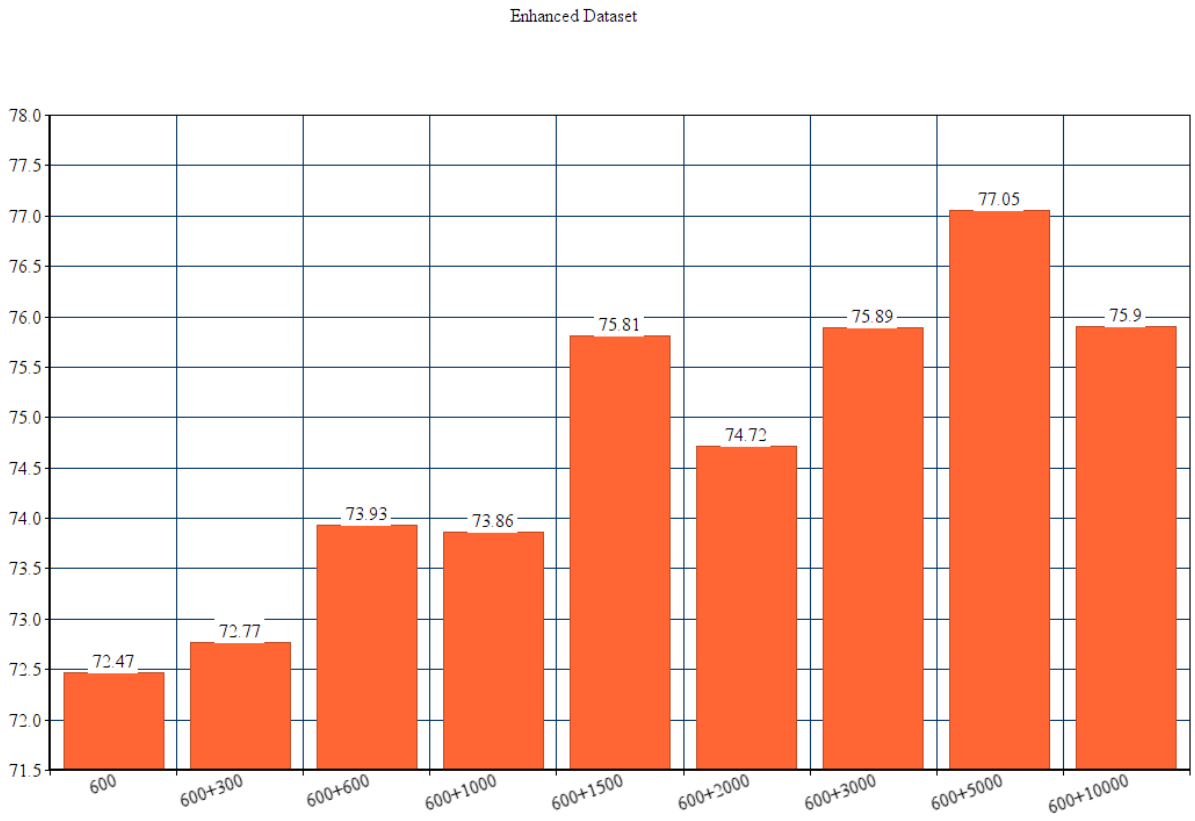
Generated εικόνες από το μοντέλο που εκπαιδεύσαμε στο καθαρό set με τις φυσιολογικές εικόνες



Generated εικόνες από το μοντέλο που εκπαιδεύσαμε στο υποσύνολο με τις μαστογραφίες με καρκίνο

Τα αποτελέσματα που πήραμε ήταν αρκετά καλά. Αξίζει σε αυτό το σημείο να τονιστεί πώς το GLOW μοντέλο μας δεν είχε δει σε αυτήν την περίπτωση ούτε το test set αλλά ούτε και το υπόλοιπο training set. Οι καινούριες εικόνες παρήχθησαν εξ ολοκλήρου από το υπάρχον training set. Το training set ενισχυμένο με augmented εικόνες όπως αυτές που εκπαιδεύτηκε το Glow, έπιασε 73.06% accuracy. Το augmented training set ενισχυμένο με 5000 εικόνες από το Glow κατάφερε να βγάλει 77.51% accuracy.

600	72.47%
600+300g	72.77%
600+600g	73.93%
600+1000g	73.86%
600+1500g	75.81%
600+2000g	74.72%
600+3000g	75.89%
600+5000g	77.05%
600+10000g	75.90%



Το καθαρό dataset ενισχυμένο με παραγώμενες εικόνες

9.2 Πειράματα με Interpolation

Δοκιμάσαμε πάλι να εφαρμόσουμε την τεχνική του interpolation και αυτήν τη φορά τα αποτελέσματα ήταν αρκετά καλά, παραπλήσια με τα αποτελέσματα των καθαρά generated εικόνων. Ο συνδυασμός generated και interpolated εικόνων πάντως δεν έδωσε κάποια αύξηση σε σχέση με την κάθε κατηγορία ξεχωριστά. Τα πειράματα έγιναν αυτήν τη φορά χρησιμοποιώντας το augmented dataset (εξού και οι 1200 εικόνες).

1200+1200i	76.48%
1200+5000g+1200i	77.03%
1200+1200g+1200i	76.90%
1200+2400i	76.43%

9.3 Latent Space Manipulation στην Άλλη Κλάση

Σε ένα άλλο πείραμα που κάναμε, εφαρμόσαμε την τεχνική του latent space manipulation. Αυτήν τη φορά, πήραμε το μοντέλο που παράγει εικόνες καρκίνου, πήραμε τον μέσο όρο του latent code ανά είδος εικόνας (αριστερό MLO, δεξί MLO, αριστερό CC, δεξί CC) για τις εικόνες του καρκίνου και για τις εικόνες χωρίς καρκίνο. Στη συνέχεια κάναμε manipulation όλες τις εικόνες που δεν είχαν καρκίνο προκειμένου να τους δώσουμε καρκίνο. Μετά κάναμε το ίδιο χρησιμοποιώντας το μοντέλο που παράγει φυσιολογικές μαστογραφίες για να αφαιρέσουμε τον καρκίνο στις μαστογραφίες με τον καρκίνο. Τα αποτελέσματα ήταν εκπληκτικά. Απλά προσθέτοντας τις νέες εικόνες ο ταξινομητής πέτυχε ποσοστό 77.16%. Το πιο σημαντικό όμως είναι ότι αυτές οι εικόνες φαίνεται να εισήγαγαν νέα πληροφορία, που παραπέμπει αρκετά σε πραγματικά δεδομένα, διαφορετική από τις απλές παραγόμενες εικόνες διότι όταν συνδυάστηκαν με τις παραγόμενες μαστογραφίες, ο ταξινομητής πέτυχε ποσοστό 80.32% που είναι τεράστια άνοδος σχετικά με το 77.5% που είχαν οι απλά παραγόμενες εικόνες και το 73.5% που είχε το απλό augmented σύνολο δεδομένων. Όταν συνδυάστηκαν και με τις interpolated, τα αποτελέσματα ήταν παρόμοια. Είναι πλέον αρκετά εμφανές ότι οι interpolated εικόνες δεν εισάγουν πληροφορία που να είναι αισθητά διαφορετική από τις generated.

1200+1200m	77.16%
1200+1200g+1200m	80.32%
1200+1200g+1200m+1200i	80.43%

9.4 Λοιπά πειράματα

Τέλος, προσπαθήσαμε να ενισχύσουμε το training set και με εικόνες που παρήχθησαν με την τεχνική του random ball αλλά τα αποτελέσματα ήταν παρόμοια με το αρχικό training set έως και ελαφρώς χειρότερα.

Για να δούμε πόσο καλά στέκουν από μόνες τους οι generated εικόνες, χρησιμοποιήσαμε 5000 από το κάθε είδος ως training set χωρίς καμία από τις αρχικές. Το αποτέλεσμα που πιάσαμε στο test set ήταν 70.63%. Ελαφρώς χειρότερο από το αρχικό training set αλλά είναι εμφανές ότι αυτές οι εικόνες

είναι αρκετά καλές ώστε να μπορούν να εκπαιδεύσουν αρκετά αποτελεσματικά ένα νευρωνικό δίκτυο ταξινόμησης από μόνες τους.

Εμπνευσμένοι από τα GAN, επιχειρήσαμε να δούμε πόσο αληθοφανείς είναι οι παραγόμενες εικόνες. Έτσι λοιπόν, φτιάξαμε ένα training dataset με 2 κλάσεις. Η πρώτη περιείχε τις 600 αρχικές εικόνες χωρίς τον καρκίνο και 600 εικόνες χωρίς καρκίνο που είχαν φτιαχτεί από το μοντέλο μας. Μετά φτιάξαμε και άλλο ένα dataset με τις εικόνες με τον καρκίνο. Στη συνέχεια φτιάξαμε αντίστοιχα και τα test sets. Χρησιμοποιήσαμε το ίδιο προεκπαιδευμένο μοντέλο. Το μοντέλο κατάφερε να αναγνωρίσει τις ψεύτικες φυσιολογικές εικόνες με ποσοστό 94% ενώ τις εικόνες με τον καρκίνο με ποσοστό 96%. Αν και ένας άπειρος παρατηρητής θα δυσκολευόταν να ανιχνεύσει εύκολα τις ψεύτικες εικόνες, ένα ισχυρό pretrained μοντέλο όπως το mobilenet_v2 κατάφερε να βρει τα μοτίβα που πρόδιδαν ότι οι εικόνες δεν ήταν αληθινές σε αρκετά μεγάλο ποσοστό.

Κεφάλαιο 10: Περαιτέρω Μελέτη

Όπως έχει ήδη αναφερθεί, οι εικόνες που χρησιμοποιήθηκαν είχαν γίνει rescale σε 128χ128. Ενδιαφέρον θα είχε να χρησιμοποιούσαμε εικόνες μεγέθους 256χ256 ή ακόμα και 512χ512 επειδή πρώτον θα ήταν πιο ευκρινείς και αναλυτικές και δεύτερον θα είχαμε τη δυνατότητα να υλοποιήσουμε πιο βαθιά μοντέλα τα οποία θα μάθαιναν καλύτερα τα χαρακτηριστικά των εικόνων. Δυστυχώς η υπολογιστική ισχύς που διαθέταμε δε μας επέτρεψε κάτι τέτοιο.

Αναφέραμε στην εργασία το πρόβλημα με το bias του άσπρου σημαδιού που υπήρχε στις φωτογραφίες και τις απόπειρες που κάναμε προκειμένου να το προσπελάσουμε. Μία προφανής λύση θα ήταν να πάρουμε όλες τις εικόνες ξεχωριστά και να περάσουμε με μαύρο χρώμα το λευκό σημάδι με κάποιο λογισμικό επεξεργασίας εικόνας (π.χ. photoshop) προκειμένου να το εξαφανίσουμε αλλά προφανώς αυτό θα απαιτούσε πολύ χρόνο και θα ήταν μια δουλειά αρκετά κουραστική. Παρ’ όλ’ αυτά, θα είχε ενδιαφέρον να δούμε τι αποτελέσματα θα μπορούσαμε να εξάγουμε από ένα πλήρως unbiased σύνολο δεδομένων.

Υλοποίηση

Όλα τα πειράματα που αναγράφονται σε αυτήν την εργασία σχετικά με την επεξεργασία των δεδομένων, την εκπαίδευση των νευρωνικών δικτύων, την παραγωγή των εικόνων και την κατηγοριοποίησή τους, έγιναν στην γλώσσα προγραμματισμού Python. Για την υλοποίηση των παραγωγικών μοντέλων χρησιμοποιήσαμε τη βιβλιοθήκη Pytorch[24] που έχει κατασκευάσει η Meta ενώ για το προεκπαιδευμένο νευρωνικό δίκτυο ταξινόμησης χρησιμοποιήσαμε τη βιβλιοθήκη Tensorflow[25] την οποία έχει αναπτύξει η Google. Χάρη σε αυτές τις βιβλιοθήκες αξιοποιήσαμε την διεπαφή CUDA της Nvidia προκειμένου να εκπαιδεύσουμε εύκολα και σχετικά γρήγορα τα νευρωνικά μας δίκτυα στην κάρτα γραφικών που διαθέταμε, η οποία είναι η Nvidia Titan Xp.

Βιβλιογραφία

[1] Kingma, Diederik P., and Max Welling. "Auto-encoding variational bayes." *arXiv preprint arXiv:1312.6114* (2013).

[2] Goodfellow, Ian, et al. "Generative adversarial nets." *Advances in neural information processing systems* 27 (2014).

[3] Rezende, Danilo, and Shakir Mohamed. "Variational inference with normalizing flows." *International conference on machine learning*. PMLR, 2015.

[4] Kingma, Durk P., and Prafulla Dhariwal. "Glow: Generative flow with invertible 1x1 convolutions." *Advances in neural information processing systems* 31 (2018).

[5] Dinh, Laurent, Jascha Sohl-Dickstein, and Samy Bengio. "Density estimation using real nvp." *arXiv preprint arXiv:1605.08803* (2016).

[6] Mitchell, Tom M., and Tom M. Mitchell. *Machine learning*. Vol. 1. No. 9. New York: McGraw-hill, 1997.

[7] https://en.wikipedia.org/wiki/Artificial_Intelligence:_A_Modern_Approach

[8] Hinton, Geoffrey, and Terrence J. Sejnowski, eds. *Unsupervised learning: foundations of neural computation*. MIT press, 1999.

[9] Kaelbling, Leslie Pack, Michael L. Littman, and Andrew W. Moore. "Reinforcement learning: A survey." *Journal of artificial intelligence research* 4 (1996): 237-285.

[10] https://news.mit.edu/2017/explained-neural-networks-deep-learning-0414

[11] Hawkins, Douglas M. "The problem of overfitting." *Journal of chemical information and computer sciences* 44.1 (2004): 1-12.

[12] Shorten, Connor, and Taghi M. Khoshgoftaar. "A survey on image data augmentation for deep learning." *Journal of big data* 6.1 (2019): 1-48.

[13] Zhu, Jun-Yan, et al. "Unpaired image-to-image translation using cycle-consistent adversarial networks." *Proceedings of the IEEE international conference on computer vision*. 2017.

[14] Xia, Weihao, et al. "Gan inversion: A survey." *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2022).

[15] Härkönen, Erik, et al. "Ganspace: Discovering interpretable gan controls." *Advances in Neural Information Processing Systems* 33 (2020): 9841-9850.

[16] Dinh, Laurent, David Krueger, and Yoshua Bengio. "Nice: Non-linear independent components estimation." *arXiv preprint arXiv:1410.8516* (2014).

[17] Ho, Jonathan, et al. "Flow++: Improving flow-based generative models with variational dequantization and architecture design." *International Conference on Machine Learning*. PMLR, 2019.

[18] Karras, Tero, et al. "Progressive growing of gans for improved quality, stability, and variation." *arXiv preprint arXiv:1710.10196* (2017).

[19] Bowles, Christopher, et al. "Gan augmentation: Augmenting training data using generative adversarial networks." *arXiv preprint arXiv:1810.10863* (2018).

[20] Frid-Adar, Maayan, et al. "Synthetic data augmentation using GAN for improved liver lesion classification." *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*. IEEE, 2018.

[21] https://www.sciencedirect.com/science/article/pii/S2666412722000162

[22] https://ardisdataset.github.io/MiniDDSM/

[23] Sandler, Mark, et al. "Mobilenetv2: Inverted residuals and linear bottlenecks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.

[24] https://pytorch.org/

[25] https://www.tensorflow.org/