



Εθνικό Μετσόβιο Πολυτεχνείο

Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών

Τομέας Συστημάτων Μετάδοσης Πληροφορίας και
Τεχνολογίας Υλικών

Ανακατασκευή 3D επιφανειών με χρήση Νευρωνικών Δικτύων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΑΓΓΕΛΟΣ ΓΕΩΡΓΑΚΗΣ

Επιβλέπων : Γεώργιος Ματσόπουλος
Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούνιος 2021



Εθνικό Μετσόβιο Πολυτεχνείο

Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών

Τομέας Συστημάτων Μετάδοσης Πληροφορίας και
Τεχνολογίας Υλικών

Ανακατασκευή 3D επιφανειών με χρήση Νευρωνικών Δικτύων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΑΓΓΕΛΟΣ ΓΕΩΡΓΑΚΗΣ

Επιβλέπων : Γεώργιος Ματσόπουλος
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 25η Ιουνίου 2021.

.....
Γεώργιος Ματσόπουλος
Καθηγητής Ε.Μ.Π.

.....
Παναγιώτης Τσανάκας
Καθηγητής Ε.Μ.Π.

.....
Δημήτριος-Διονύσιος Κουτσούρης
Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούνιος 2021

.....
Άγγελος Γεωργάκης

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

Copyright © Άγγελος Γεωργάκης, 2021.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Περίληψη

Η ραγδαία ανάπτυξη των μοντέλων τεχνητής νοημοσύνης που λαμβάνει χώρα τις τελευταίες δεκαετίες, την έχει φέρει σε θέση να εισχωρήσει σε όλους τους τομείς της ανθρώπινης δραστηριότητας. Συνεπώς δεν θα μπορούσε να μην χρησιμοποιείται για την επεξεργασία και ανάλυση ιατρικών δεδομένων.

Ιδιαίτερα, στον τομέα αυτό γίνονται μεγάλα βήματα προόδου τόσο στο κομμάτι ερμηνείας δεδομένων και πρόβλεψης ασθενειών, αλλά και στο κομμάτι της απόκτησης καλύτερων και σαφέστερων δεδομένων, ώστε η ερμηνεία να γίνει σε δεύτερο χρόνο από ειδικούς επιστήμονες.

Στην παρούσα διπλωματική θα ασχοληθούμε με το κομμάτι της ανακατασκευής τρισδιάστατων κρανιακών scans. Για να το πετύχουμε αυτό θα χρησιμοποιήσουμε ίσως το πιο δυνατό εργαλείο που μας προσφέρει η τεχνητή νοημοσύνη στον τομέα της ανάλυσης και επεξεργασίας εικόνας, τα συνελκτικά νευρωνικά δίκτυα. Στην προσπάθειά μας αυτή, συναντήσαμε διάφορες δυσκολίες τις οποίες όμως αντιμετωπίσαμε κυρίως χάρη στην κατάλληλη προεπεξεργασία των δεδομένων μας αλλά και στα πολύ ισχυρά και εύχρηστα εργαλεία διαχείρισης νευρωνικών δικτύων που παρέχει η γλώσσα προγραμματισμού που χρησιμοποιήσαμε, δηλαδή η Python.

Πιο συγκεκριμένα μελετήσαμε και αξιολογήσαμε δύο διαφορετικά σύνολα δεδομένων, εκτελέσαμε πειράματα εκπαιδύοντας μοντέλα με διαφορετικές αρχιτεκτονικές και ψάξαμε τον βέλτιστο συνδυασμό υπερπαραμέτρων κάθε δικτύου, επιτυγχάνοντας εξαιρετικά καλές επιδόσεις. Τέλος χρησιμοποιήσαμε εξειδικευμένο λογισμικό για την τρισδιάστατη αναπαράσταση των εικόνων μας, και με βάση τη γνώση και εμπειρία που αποκτήσαμε, προτείνουμε μερικές ενδιαφέρουσες μελλοντικές κατευθύνσεις της επιστημονικής μελέτης.

Λέξεις κλειδιά

Συνελκτικά Νευρωνικά Δίκτυα, Αυτοκωδικοποιητής, 3D CBCT scans, Ανακατασκευή εικόνων, Αναπαράσταση 3D εικόνων

Abstract

The rapid development of Artificial Intelligence models that has been taking place the past few decades, has brought Artificial Intelligence to a point where it is a part of almost every field of human activity. So it could not be absent from processing and analyzing medical data.

Especially in the medical field, huge steps of progress are being made both in interpretation of data and precognition of diseases, but also in acquiring better and more articulate data, that will later be interpreted by specialized scientists.

In this diploma thesis we will deal with the reconstruction of 3D cranial scans. In order to achieve that we will use the most powerful tool in image processing, that Artificial Intelligence has to offer, Convolutional Neural Networks. In trying to build the most suitable Neural Network for that task, we came across many difficulties, which we overcame, mainly thanks to the right pre-processing of our data, but also thanks to the very strong and useful tools for Neural Network managing that Python(the programming language we used) has to offer.

Specifically, we studied and assessed two different sets of data, conducted experiments by training models with different architectures, and searched for the best set of hyperparameters for each one of these models, achieving great results. Finally we used some specialized software for the 3D reconstruction of our images, and according to the knowledge and experience we acquired, we suggested some possible new scientific paths in that field of research.

Key words

Convolutional Neural Networks, Autoencoder, 3D CBCT scans, Image reconstruction, 3D image representation

Ευχαριστίες

Αρχικά θα ήθελα να ευχαριστήσω τον καθηγητή κ. Γεώργιο Ματσόπουλο, για την εμπιστοσύνη που μου έδειξε, αλλά και για την ευκαιρία που μου έδωσε να ασχοληθώ με ένα τόσο ενδιαφέρον θέμα.

Επίσης θα ήθελα να ευχαριστήσω ειλικρινά τον διδάκτορα κ. Θοδωρή Οικονομόπουλο για την τυπική και άψογη επικοινωνία που είχαμε καθ' όλη τη διάρκεια της συνεργασίας μας, καθώς και για τις καίριες παρεμβάσεις του αναφορικά με την κατεύθυνση της μελέτης μας.

Ακόμη θα ήθελα να ευχαριστήσω από καρδιάς τους φίλους και τις φίλες που στάθηκαν δίπλα μου και διάνθισαν όλο αυτό το ταξίδι.

Μα πάνω απ' όλα θα ήθελα να ευχαριστήσω τη μητέρα μου για την αμέριστη στήριξη και αγάπη που έδειξε και συνεχίζει να μου δείχνει κάθε στιγμή.

Άγγελος Γεωργάκης,

Αθήνα, 25η Ιουνίου 2021

Περιεχόμενα

Περίληψη	5
Abstract	7
Ευχαριστίες	9
Περιεχόμενα	11
Κατάλογος πινάκων	13
Κατάλογος σχημάτων	15
1. Εισαγωγή	17
Εισαγωγή	17
1.1 Βασικές φυσικές αρχές	17
1.1.1 Αξονική Τομογραφία (CT)	18
1.1.2 Αξονική Τομογραφία Κωνικής Δέσμης (CBCT)	19
1.2 Κίνητρο Διπλωματικής Εργασίας	20
1.3 Συνεισφορά Διπλωματικής Εργασίας	21
1.4 Δομή Διπλωματικής Εργασίας	22
2. Θεωρητικό Υπόβαθρο	23
Θεωρητικό Υπόβαθρο	23
2.1 Μηχανική Μάθηση	23
2.1.1 Ορισμός Μηχανικής Μάθησης	23
2.1.2 Βασικά ήδη μηχανικής μάθησης	23
2.1.3 Βαθιά Μάθηση	24
2.1.4 Τεχνητά Νευρωνικά Δίκτυα	25
2.1.5 Συνελκτικά Νευρωνικά Δίκτυα	27
2.2 Autoencoder - Αυτοκωδικοποιητής	30
2.2.1 Εφαρμογές των Αυτοκωδικοποιητών	32
2.3 Βασικές έννοιες Τεχνητών Νευρωνικών Δικτύων(TNΔ)	33
2.3.1 Αλγόριθμοι Βελτιστοποίησης	34
2.3.2 Αλγόριθμος οπισθοδιάδοσης του λάθους(Backpropagation of the error algorithm)	35
2.3.3 Overfitting-Υπερμοντελοποίηση	36
2.4 Βιβλιογραφική Επισκόπηση	38
3. Σχεδιασμός, Επιλογή παραμέτρων και Υλοποίηση	41
3.1 Απόκτηση δεδομένων	41
3.2 Δεδομένα	42
3.2.1 Προεπεξεργασία Δεδομένων	42
3.2.2 Διαχωρισμός Δεδομένων	43

3.3	Περιβάλλον Διεξαγωγής Πειραμάτων	44
3.3.1	Επιλογή Δεδομένων Εισόδου	45
3.4	Επιλογή Νευρωνικού Δικτύου	46
3.4.1	Επιλογή Αρχιτεκτονικής	46
3.4.2	Επιλογή και Βελτιστοποίηση Υπερπαραμέτρων	47
3.5	Έξοδος προγράμματος και μετρικές αξιολόγησης	49
4.	Αξιολόγηση αποτελεσμάτων και συγκρίσεις	51
4.1	Μετρικές αξιολόγησης	51
4.2	Συγκριτικά αποτελέσματα μοντέλων	56
4.2.1	Παρουσίαση Μοντέλων	56
4.2.2	Subsampled δεδομένα	57
4.2.3	Fullscale δεδομένα	64
4.3	Τρισδιάστατη Αναπαράσταση	70
4.4	Προσωπικές Παρατηρήσεις	76
5.	Επίλογος	77
	Επίλογος	77
5.1	Σύνοψη	77
5.2	Τελικά Συμπεράσματα	78
5.3	Μελλοντικές Κατευθύνσεις Επιστημονικής Μελέτης	78

Κατάλογος πινάκων

4.1	Autoencoder	57
4.2	Autoencoder without BatchNormalization layers	61
4.3	Metrics according to Batch Size	64
4.4	Autoencoder	64

Κατάλογος σχημάτων

1.1	Hounsfield unit (HU) για διάφορα όργανα του ανθρώπινου σώματος	18
1.2	Πηγή ακτίνων X και ο ανιχνευτής καθώς περιστρέφονται γύρω από το σημείο ενδιαφέροντος του εκάστοτε ασθενούς.	20
2.1	Sigmoid Function	26
2.2	Tanh Function	26
2.3	ReLu Function	27
2.4	Convolutional Kernel	29
2.5	Max-Average Pooling	30
2.6	Autoencoder Architecture	32
2.7	Before/After Dropout Layers	37
3.1	Σχεδιαγραμματική απεικόνιση των βημάτων μετατροπής των δεδομένων	42
3.2	Σχεδιαγραμματική απεικόνιση των βημάτων που ακολουθήθηκαν στο σύνολο της διπλωματικής	44
4.1	Mutual Information Equation	54
4.2	Διάγραμμα Venn που δείχνει τις προσθετικές και αφαιρετικές σχέσεις πληροφορίας που σχετίζονται με τις συσχετισμένες μεταβλητές X και Y . Η περιοχή που περιέχεται και στους δύο κύκλους είναι η κοινή εντροπία $H(X,Y)$. Ο κύκλος στα αριστερά (κόκκινο και μοβ) είναι η μεμονωμένη εντροπία $H(X)$, με το κόκκινο να είναι η υπό όρους εντροπία $H(X Y)$. Ο κύκλος στα δεξιά (μπλε και μοβ) είναι η $H(Y)$, με το μπλε να είναι η $H(Y X)$. Το βιολετί είναι η αμοιβαία πληροφορία(Mutual Information) $I(X;Y)$	55
4.3	Reconstructed Images	58
4.4	a1-a2 histograms	58
4.5	Combined histogram	59
4.6	Val-Train Loss	59
4.7	Val-Train mae	60
4.8	Reconstructed Images	61
4.9	a1-a2 histograms	62
4.10	Combined histogram	62
4.11	Val-Train Loss	63
4.12	Val -Train mae	63
4.13	Reconstructed Images	65
4.14	a1-a2 histograms	65
4.15	Combined histogram	66
4.16	Val-Train Loss	66
4.17	Val-Train mae	67
4.18	Reconstructed Images	67
4.19	a1-a2 histograms	68
4.20	Combined histogram	68
4.21	Val-Train Loss	69
4.22	Val-Train mae	69

4.23	1° Πλάγια Αριστερή Όψη	70
4.24	1° Πλάγια Δεξιά Όψη	71
4.25	1° Κάτω Όψη	71
4.26	1° Μπροστά Όψη	72
4.27	1° Πίσω Όψη	72
4.28	1° Πλάγια Αριστερή Όψη	73
4.29	2° Πλάγια Δεξιά Όψη	73
4.30	2° Μπροστά Όψη	74
4.31	2° Πίσω όψη	74
4.32	2° Κάτω Όψη	75
4.33	3D απεικόνιση εικόνας 400x400x325 από Vesalius3D	75

Κεφάλαιο 1

Εισαγωγή

Η τρισδιάστατη υπολογιστική τομογραφία (3D CT), έχει αποδειχθεί ένα εξαιρετικά χρήσιμο εργαλείο για την αξιολόγηση ποικίλων παθολογιών. Οι προηγμένες τεχνικές απεικόνισης εγκάρσιας τομής επιτρέπουν στους ακτινολόγους να οπτικοποιήσουν την ασθένεια με μεγαλύτερη βεβαιότητα, αφαιρώντας την επίδραση των υπερκείμενων ιστών, επιτρέποντας έτσι την ξεχωριστή αξιολόγηση μεμονωμένων οργάνων, η οποία βοηθά στην ανίχνευση και στον χαρακτηρισμό της παθολογίας.

Στα τέλη της δεκαετίας του 1970 οι ακτινολόγοι ανακάλυψαν πως η διάγνωση που βασίζεται σε αξονικές τομογραφίες ήταν πιο εξελιγμένη απ' ό τι με την απλή ακτινογραφία, καθώς συχνά η έλλειψη τρίτης διάστασης (π.χ. οβελιαίες και στεφανιαίες διαστάσεις) δυσχέραινε το έργο τους. Βέβαια πολλοί γιατροί χωρίς βασική εκπαίδευση στην απεικόνιση διατομών εξακολουθούν να αντιμετωπίζουν δυσκολίες στην εκτίμηση της φυσιολογικής ανατομίας και της παθολογίας σε εγκάρσιες τομές CT, ώντας περισσότερο εξοικειωμένοι με την ανατομία που απεικονίζεται στο στεφανιαίο επίπεδο.

Μετά τους κάπως ακατέργαστους αλγόριθμους απόδοσης τρισδιάστατων εικόνων που αναπτύχθηκαν αρχικά στα τέλη της δεκαετίας του 1970, που επέτρεψαν το σχηματισμό τρισδιάστατων εικόνων από δεδομένα που αποκτήθηκαν στο αξονικό επίπεδο, ακολούθησε η ανάπτυξη ελικοειδούς CT μίας τομής, αλλά και πιο πρόσφατα, οι σαρωτές πολλαπλών ανιχνευτών CT (MDCT) έχουν ανοίξει νέα κεφάλαια στην τρισδιάστατη(3D) απεικόνιση. Αυτές οι προόδους έγιναν δυνατές χάρη στην ταχεία απόκτηση ογκομετρικών δεδομένων στον άξονα z, χρησιμοποιώντας λεπτές τομές και βελτιωμένους αλγόριθμους απόδοσης όγκου, οι οποίοι διευκολύνουν τις τρισδιάστατες αναπαραστάσεις, χωρίς να υποβαθμίζεται η απόδοση τους λόγω διάφορων αντανεκλαστικών κινήσεων του σώματος όπως για παράδειγμα η αναπνοή.

Στην παρούσα διπλωματική εργασία θα πραγματευτούμε την ανακατασκευή τρισδιάστατων κρανιακών scans, με τη χρήση μηχανικής μάθησης, και για την ακρίβεια νευρωνικών δικτύων. Πιο συγκεκριμένα θα δημιουργήσουμε ένα συγκεκριμένο είδος νευρωνικού δικτύου, έναν Αυτοκωδικοποιητή(Autoencoder), μέσω του οποίου θα επιχειρήσουμε να ανακατασκευάσουμε τις συγκεκριμένες τρισδιάστατες εικόνες ανακτώντας όσο το δυνατόν περισσότερη πληροφορία.

1.1 Βασικές φυσικές αρχές

Στην ενότητα αυτή θα δούμε τις βασικές φυσικές αρχές καθώς και τον τρόπο λειτουργίας της αξονικής τομογραφίας(CT) αλλά και της αξονικής τομογραφίας κωνικής δέσμης(CBCT) με την οποία και εργαστήκαμε.

1.1.1 Αξονική Τομογραφία (CT)

Οι τιμές των pixel των CT scans μετρώνται σε Hounsfield units(HU). Τα Hounsfield units πήραν το όνομά τους από τον Sir Godfrey Hounsfield, νικήτη βραβείου νόμπελ και εφευρέτη της υπολογιστικής τομογραφίας(CT).

Τα 3D CT scans αποτελούν μία μορφή απόδοσης επιφάνειας, και δημιουργούνται με τη βοήθεια εξειδικευμένου προηγμένου λογισμικού. Η διαδικασία συνίσταται στη λήψη απλών αξονικών σαρώσεων της περιοχής ενδιαφέροντος. Στη συνέχεια, στον υπολογιστή παρέχεται μια προσεκτικά επιλεγμένη τιμή εξασθένησης «κατώφλιου». Το πρόγραμμα σαρώνει κάθε τομή CT ανά γραμμή και καταγράφει τις ακριβείς συντεταγμένες κάθε pixel(εικονοστοιχείου) που δείχνει μια τιμή εξασθένησης υψηλότερη από το επιλεγμένο όριο. Για παράδειγμα, εάν επιλεγεί μια τιμή εξασθένησης +200 HU, τότε στην εικόνα 3D θα συμπεριληφθούν μόνο τα pixel με τιμή +200 HU ή περισσότερο. Αυτά τα επιλεγμένα pixel αντιπροσωπεύουν voxels(ογκοστοιχεία) τα οποία θεωρείται ότι περιέχουν ιστό πυκνότερο από το επιλεγμένο κατώφλι.

Predefined threshold value	Minimum	Maximum
Bone (CT)	226	3071
Soft Tissue (CT)	-700	225
Enamel (CT, Adult)	1553	2850
Enamel (CT, Child)	2042	3071
Compact Bone (CT, Adult)	662	1988
Compact Bone (CT, Child)	586	2198
Spongial Bone (CT, Adult)	148	661
Spongial Bone (CT, Child)	156	585
Muscle Tissue (CT, Adult)	-5	135
Muscle Tissue (CT, Child)	-25	139
Fat Tissue (CT, Adult)	-205	-51
Fat Tissue (CT, Child)	-212	-72
Skin Tissue (CT, Adult)	-718	-177
Skin Tissue (CT, Child)	-766	-202

Σχήμα 1.1: Hounsfield unit (HU) για διάφορα όργανα του ανθρώπινου σώματος

Στη συνέχεια, οι CT τομές, στοιβάζονται η μία πάνω από την άλλη μέσω του υπολογιστή και τα γειτονικά pixel του κάθετου άξονα z ενώνονται μεταξύ τους σχηματίζοντας έτσι μία ογκομετρική απόδοση της τρισδιάστατης επιφάνειας. Στη συνέχεια, σε αυτήν την εικόνα εκχωρείται σκίαση επιφάνειας από μια εικονική πηγή φωτός. Τα pixel που βρίσκονται κοντά στην πηγή φωτός φωτίζονται έντονα, ενώ τα απομακρυσμένα είναι κατάλληλα σκιασμένα. Φυσικά, εκείνα τα εικονοστοιχεία που θεωρείται ότι βρίσκονται πίσω από ένα άλλο αδιαφανές εικονοστοιχείο δεν εμφανίζονται

σε αυτήν την επιφανειακή απόδοση του όγκου. Αυτό το παιχνίδι φωτός και σκιών είναι υπεύθυνο για τη δημιουργία του τρισδιάστατου εφέ [Curry et al.](#)

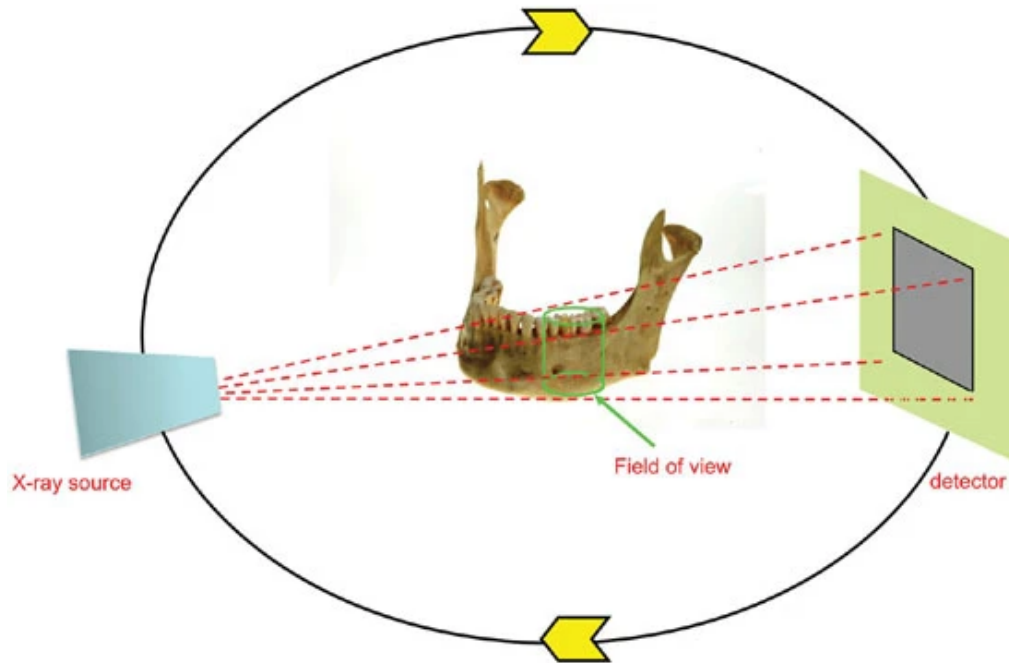
Δεδομένου ότι η επιλογή ενός pixel για την ενσωμάτωσή του στην τρισδιάστατη εικόνα είναι ένα all or nothing φαινόμενο, δεν υπάρχει καμία διαφοροποίηση γκρίζας κλίμακας μεταξύ των επιλεγμένων pixel. Αυτό σημαίνει ότι οι μεμονωμένες διαφορές ιστών δεν μπορούν να επισημανθούν με αυτήν την τεχνική [Talwar I \(1992\)](#). Ως εκ τούτου, η εφαρμογή των 3D CT scans περιορίζεται κυρίως στην απεικόνιση παθολογιών των οστών. Από την άλλη πλευρά, ένα σημαντικό συμπέρασμα αυτού είναι ότι οι αξονικές εικόνες χαμηλής αντίθεσης μπορούν να χρησιμοποιηθούν για τη δημιουργία εξαιρετικών τρισδιάστατων εικόνων. Μπορεί επίσης να επιχειρηθεί μερικές φορές η εμφάνιση ορισμένων παθολογιών εσωτερικού μαλακού ιστού, εάν η τιμή τους σε HU είναι αρκετά υψηλότερη από αυτές των γύρω περιοχών, επιλέγοντας τιμές κατωφλίου κατάλληλες για να συμπεριληφθεί η παθολογία. Αυτό είναι δυνατό στην περίπτωση αγγειακής παθολογίας, όπως δυσπλασία AV, ή αγγειακών ή ασβεστοποιημένων όγκων [Gholkar A \(1988\)](#).

1.1.2 Αξονική Τομογραφία Κωνικής Δέσμης (CBCT)

Η Αξονική Τομογραφία Κωνικής Δέσμης (CBCT) είναι μία διαφοροποιημένη εκδοχή της αξονικής τομογραφίας (CT), που χρησιμοποιείται ως επί το πλείστον στην οδοντική απεικόνιση, στην ορθοπαιδική, αλλά τελευταία βρήκε εφαρμογές και στην εξειδικευμένη απεικόνιση στήθους ανάμεσα σε άλλα. Η βασική διαφορά της με την απλή αξονική τομογραφία έγκειται στην χρήση κωνικού σχήματος ακτίνων X και ανιχνευτές δύο (2) διαστάσεων, ενώ στη συμβατική αξονική τομογραφία χρησιμοποιούνται ριπιδιόμορφου σχήματος ακτίνες X και ανιχνευτές μίας (1) διάστασης [Scarfe and Farman \(2008\)](#).

Συγκεκριμένα μία αποκλείουσα, κωνικού σχήματος πηγή ακτινοβολίας εκπέμπεται προς τον στόχο. Οι εξασθενημένες ακτίνες X ανιχνεύονται από ανιχνευτές ακτίνων X που βρίσκονται στην απεναντι πλευρά, αντιδιαμετρικά της πηγής ακπομπής των ακτίνων, με το προς σάρωση αντικείμενο να βρίσκεται ανάμεσά τους. Ο ανιχνευτής αυτός έχει πολλαπλά στοιχεία ανίχνευσης (dexels), τόσο στον X όσο και στον Y άξονα. Έτσι η απόκτηση μίας απεικόνισης όγκου μπορεί επιτευχθεί με λιγότερες περιστροφές της γέφυρας του σωλήνα ακτίνων X. Αυτό διαφοροποιεί την αξονική τομογραφία κωνικής δέσμης, σε σχέση με την απλή αξονική τομογραφία η οποία χρησιμοποιεί στοιχεία ανίχνευσης (dexels) μόνο κατά τον άξονα X.

Κατά την διάρκεια οδοντικής απεικόνισης με CBCT, ο σαρωτής περιστρέφεται γύρω από το κεφάλι του ασθενούς, παράγοντας έως και 600 διαφορετικές εικόνες δύο διαστάσεων. Για την επεμβατική ακτινολογία (interventional radiology), ο ασθενής τοποθετείται στο τραπέζι με τρόπο ώστε η περιοχή ενδιαφέροντος να βρίσκεται στο κέντρο του πεδίου όρασης της κωνικής δέσμης. Μία μοναδική περιστροφή 200 μοιρών γύρω από την περιοχή ενδιαφέροντος, οδηγεί στην απόκτηση ενός σετ εικόνων που όταν τις ανακατασκευάσει το κατάλληλο λογισμικό, δημιουργούν μία ψηφιακή εικόνα όγκου τριών διαστάσεων, την οποία μπορούμε να χειριστούμε αναλόγως.



Σχήμα 1.2: Πηγή ακτίνων Χ και ο ανιχνευτής καθώς περιστρέφονται γύρω από το σημείο ενδιαφέροντος του εκάστοτε ασθενούς.

Η τριαδιάστατη αξονική τομογραφία κωνικής δέσμης (CBCT) έχει βρει εφαρμογή σε πάρα πολλές περιπτώσεις. Παρακάτω ακολουθούν μερικά παραδείγματα :

- Εμφυτεύματα: Κατά τον προεγχειρητικό σχεδιασμό για την τοποθέτηση των εμφυτευμάτων: (α) υλοποιείται η επισημάνση και διάγνωση σημαντικών ανατομικών μορίων, όπως το κάτω φατνιακό νεύρο, (β) απεικονίζεται το σχήμα και η μορφολογία της ακρολοφίας σε τρεις διαστάσεις, (γ) ελέγχεται η ποιότητα του οστού και (δ) καθορίζεται η θέση που θα τοποθετηθούν τα εμφυτεύματα.
- Παθολογία κροταφογναθικής διάρθρωσης: Γίνεται τρισδιάστατη απεικόνιση του δίσκου, σε σχέση με την γλήνη και τον κόνδυλο.
- Έγκλειστοι κυνόδοντες, φρονιμίτες
- Κύστες, αποστήματα, νεοπλασίες
- Ακρορριζικές αλλοιώσεις, που δεν αναγνωρίζονται στην Πανοραμική.
- Ενδοδοντία
- Ορθοδοντική
- Περιοδοντολογία

1.2 Κίνητρο Διπλωματικής Εργασίας

Το ενδιαφέρον για τη βαθιά μάθηση και τα νευρωνικά δίκτυα σε συνδυασμό με την περιέργεια για πιθανές εφαρμογές τους πάνω σε ιατρικά δεδομένα και η τεράστια χρησιμότητα που αυτό μπο-

ρεί να έχει, δημιούργησαν τη θεματολογία της παρούσας διπλωματικής εργασίας.

Η ιατρική καθώς αποτελεί τον αιρχειότερο και σημαντικότερο επιστημονικό κλάδο της ανθρώπινης δραστηριότητας, καλείται να παρακολουθεί τις ραγδαίες εξελίξεις της τεχνολογίας και ταυτόχρονα να τις υιοθετεί στον βαθμό που αυτό είναι εφικτό. Από την άλλη ο τομέας της τεχνητής νοημοσύνης αποτελεί ένα πολύ ενδιαφέρον μέρος της επιστήμης των υπολογιστών που γνωρίζει μεγάλη ανάπτυξη στις μέρες μας. Αυτό έχει ως αποτέλεσμα να υπάρχουν πολλές δυνατότητες ανάπτυξης και βελτίωσης και φαίνεται να δημιουργούνται προοπτικές για σπουδαία επιτεύγματα. Ένα παράδειγμα σύνδεσης αυτών των δύο, αποτελεί η μηχανική μάθηση, η οποία θεωρείται μια από τις πιο σημαντικές ερευνητικές περιοχές. Μέσω των συστημάτων υποστήριξης κλινικών αποφάσεων, οι τεχνικές μηχανικής μάθησης μπορούν να ενσωματωθούν στις διαδικασίες της διάγνωσης, πρόβλεψης αλλά και θεραπείας ασθενειών.

Ηδη η τεχνητή νοημοσύνη, άλλες φορές ως προέκταση της ανθρώπινης και άλλες ως άξιος αντικαταστάτης της, έχει κυριαρχήσει σε πάρα πολλούς τομείς της ανθρώπινης δραστηριότητας. Ο ενθουσιασμός μας για το σχετικά ανεξερεύνητο πεδίο της τεχνητής νοημοσύνης σε συνδυασμό με την πιθανή καθιέρωσή της στην ανάλυση δεδομένων υγείας και την ανίχνευση προτύπων που θα μπορούσαν να οδηγήσουν σε νέες επιστημονικές ανακαλύψεις και να βελτιώσουν τους μεμονωμένους διαγνωστικούς ελέγχους, αλλά ακόμη και στην παροχή καθαρότερων και πλουσιότερων πληροφοριών στους ειδικούς, που θα μπορούσε να επιφέρει ευεργετικά αποτελέσματα, μας οδήγησαν στο να ασχοληθούμε με την συγκεκριμένη θεματολογία.

1.3 Συνεισφορά Διπλωματικής Εργασίας

Στην παρούσα διπλωματική εργασία πραγματευόμαστε την ανακατασκευή τρισδιάστατων CBCT scan με τη χρήση συγκεκριμένων τεχνικών μηχανικής μάθησης, και μέσω αυτού του προβλήματος μελετάμε κι αναλύουμε τις επιρροές διαφόρων παραμέτρων στις επιδόσεις των μοντέλων νευρωνικών δικτύων που κατασκευάσαμε. Αρχικά στοχεύουμε στην παρουσίαση μίας σφαιρικής εικόνας γύρω από τους διαλαφορους αλληλοσυνδεόμενους κλάδους στους οποίους ανήκει η μηχανική μάθηση, και στη συνέχεια παρουσιάζουμε τα θεωρητικά εργαλεία που απαιτούνται για την κατανόηση της λειτουργίας της.

Καθώς τα τρισδιάστατα δεδομένα που κληθήκαμε να ανακατασκευάσουμε είναι αρκετά "βαριά", τα περνάμε από ένα στάδιο προεπεξεργασίας και στη συνέχεια μελετάμε αναλυτικά μέσω πειραμάτων την επίδραση των υπερπαραμέτρων των μοντέλων στην εκπαίδευση και την αποτελεσματικότητα των δικτύων μας, ενώ δοκιμάζουμε και διάφορες αρχιτεκτονικές "χτισίματος" μοντέλων και παρουσιάζουμε μια ανάλυση επί του θέματος. Πέρα από τις διάφορες υπερπαραμέτρους με τις οποίες πειραματιστήκαμε, καθώς κάναμε δοκιμές χρησιμοποιώντας δύο διαφορετικά σετ δεδομένων, παρουσιάζουμε μία συγκριτική ανάλυση των αποτελεσμάτων ανάμεσα στα δύο αυτά σετ. Επίσης γίνεται αναφορά στα εργαλεία τρισδιάστατης απεικόνισης που εμείς βρήκαμε να είναι τα πιο εύχρηστα και αποτελεσματικά. Ενώ αξίζει να αναφερθεί πως καταφέραμε να επιτύχουμε εξαιρετικά καλές επιδόσεις στην ανακατασκευή των εικόνων, χάνοντας ελάχιστη πληροφορία.

1.4 Δομή Διπλωματικής Εργασίας

Στο Κεφάλαιο 2 καλύπτεται το θεωρητικό υπόβαθρο όσων κρίθηκαν απαραίτητα για την κατανόηση της διπλωματικής εργασίας. Αρχικά γίνεται μία επεξήγηση βασικών εννοιών της τεχνητής νοημοσύνης και των νευρωνικών δικτύων, και στη συνέχεια αναλύεται σε μεγαλύτερο βάθος το πρόβλημα της ανακατασκευής εικόνας με τη χρήση αυτοκωδικοποιητών. Στη συνέχεια γίνεται μια παρουσίαση του related work, δηλαδή των σημαντικότερων εργασιών που έχουν γίνει μέχρι σήμερα πάνω στο πρόβλημα που καλούμαστε να λύσουμε. Συνολικά σε αυτό το κεφάλαιο καλύπτονται όλες οι προαπαιτούμενες γνώσεις που θα πρέπει να έχει ένας αναγνώστης ώστε να μπορέσει να ακολουθήσει τη ροή της σκέψης και της δουλειάς μας στα επόμενα κεφάλαια.

Στο Κεφάλαιο 3 αναλύονται οι σχεδιαστικές επιλογές που κάναμε κατά την ανάπτυξη των μοντέλων μας και τα βήματα υλοποίησης που ακολουθήσαμε. Παρουσιάζονται τα δεδομένα με τα οποία εργαστήκαμε καθώς και τα βήματα προεπεξεργασίας αυτών ώστε να έχουμε καλύτερα αποτελέσματα στη συνέχεια, καθώς γίνεται και μια αναφορά στο περιβάλλον εκτέλεσης των πειραμάτων.

Στο Κεφάλαιο 4 παρουσιάζονται τα αποτελέσματα των πειραμάτων κι εξάγονται συμπεράσματα. Συγκεκριμένα κάνουμε μια συγκριτική παρουσίαση των αποτελεσμάτων που πήραμε για διάφορους συνδυασμούς υπερπαραμέτρων, αλλά και δεδομένων εκπαίδευσης, και παραθέτουμε προσωπικές παρατηρήσεις του συγγραφέα.

Στο Κεφάλαιο 5 γίνεται μια σύνοψη της διπλωματικής και στη συνέχεια εξάγονται και συγκεντρώνονται τα τελικά συμπεράσματα. Τέλος αναφέρονται πιθανές μελλοντικές κατευθύνσεις της επιστημονικής μελέτης, περισσότερες από τις οποίες προκύπτουν ως λογική συνέχεια της δουλειάς που κάναμε για αυτή την εργασία.

Κεφάλαιο 2

Θεωρητικό Υπόβαθρο

Στο παρών κεφάλαιο θα καλυφθεί το θεωρητικό υπόβαθρο, που αφορά τη μηχανική μάθηση, τα εργαλεία και τις μεθόδους της, καθώς και πιο συγκεκριμένα λεπτομέρειες για τη δομή και τη λειτουργία των νευρωνικών δικτύων που χρησιμοποιήσαμε. Και στο τέλος θα δούμε αντίστοιχες προσπάθειες που έχουν γίνει στη σχετική βιβλιογραφία.

2.1 Μηχανική Μάθηση

2.1.1 Ορισμός Μηχανικής Μάθησης

Το 1959, ο Άρθουρ Σάμουελ ορίζει τη μηχανική μάθηση ως "Πεδίο μελέτης που δίνει στους υπολογιστές την ικανότητα να μάθαινουν, χωρίς να έχουν ρητά προγραμματιστεί" [Simon \(December 2015\)](#). Η μηχανική μάθηση διερευνά τη μελέτη και την κατασκευή αλγορίθμων που μπορούν να μάθαινουν 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.

Η μηχανική μάθηση αποτελεί κομμάτι του κλάδου της πληροφορικής που αποκαλείται ως "Τεχνητή Νοημοσύνη". Πολύ συχνά αυτοί οι όροι συγχέονται. Η τεχνητή νοημοσύνη, όπως ορίστηκε από ένας από τους ιδρυτές του κλάδου και εφευρέτη του όρου, John McCarthy, είναι η επιστήμη και η τεχνολογία κατασκευής ευφύων μηχανών. ([McCarthy, 2004](#)) Άρα η τεχνητή νοημοσύνη σαν όρος αποτελεί υπερσύνολο της μηχανικής μάθησης, και αποτελεί την γενικότερη προσπάθεια να κάνουμε τους υπολογιστές να μιμηθούν την ευφυή συμπεριφορά του ανθρώπου και περιλαμβάνει και άλλα πεδία όπως είναι η αναγνώριση λόγου και εικόνας και η λήψη αποφάσεων.

Ένα χαρακτηριστικό που διαχωρίζει την μηχανική μάθηση από άλλες μορφές τεχνητής νοημοσύνης είναι η ικανότητά της να μορφοποιεί κατάλληλα τον εαυτό της όταν έρχεται σε επαφή με νέα δεδομένα. Αυτό σημαίνει πως η μηχανική μάθηση είναι δυναμική και δεν χρειάζεται ανθρώπινη παρέμβαση για να πραγματοποιήσει αλλαγές στον εαυτό της. Μέσω αυτής της ικανότητας της μηχανικής μάθησης, οι επιστήμονες μπορούν πλέον να τροφοδοτήσουν με δεδομένα και να οδηγήσουν τους υπολογιστές σε νέα γνώση που ούτε οι ίδιοι κατέχουν.

2.1.2 Βασικά ήδη μηχανικής μάθησης

Οι εργασίες μηχανικής μάθησης συνήθως ταξινομούνται σε τρεις μεγάλες κατηγορίες, ανάλογα με τη φύση του εκπαιδευτικού «σήματος» ή την «ανατροφοδότηση» που είναι διαθέσιμα σε ένα σύστημα εκμάθησης. Αυτές είναι: η Επιβλεπόμενη(ή επιτηρούμενη) Μάθηση(Supervised Learn-

ing), η Μη Επιβλεπόμενη (ή επιτηρούμενη) Μάθηση (Unsupervised Learning) και η Ενισχυτική Μάθηση (Reinforced Learning) .

2.1.2.1 Επιβλεπόμενη Μάθηση

Το υπολογιστικό πρόγραμμα δέχεται τις παραδειγματικές εισόδους καθώς και τα επιθυμητά αποτελέσματα από έναν «δάσκαλο», και ο στόχος είναι να μάθει έναν γενικό κανόνα προκειμένου να αντιστοιχίσει τις εισόδους με τα αποτελέσματα. Ο αλγόριθμος δηλαδή προσπαθεί να μοντελοποιήσει τις σχέσεις ανάμεσα στα χαρακτηριστικά των εισόδων και τις επιθυμητές εξόδους ώστε να μπορέσει να βρει αντίστοιχα μοτίβα σε άγνωστα δεδομένα και να προβλέψει την αναμενόμενη έκβαση τους. Αυτή η μέθοδος επιλέγεται κυρίως σε προβλήματα ταξινόμησης και αναδρομής.

2.1.2.2 Μη Επιβλεπόμενη Μάθηση

Σε αυτή τη περίπτωση, χωρίς να παρέχεται κάποια εμπειρία στον αλγόριθμο μάθησης, δηλαδή χωρίς να γνωρίζει εκ των προτέρων ποιά είναι η επιθυμητή έξοδος, πρέπει αυτός να βρει την δομή των δεδομένων εισόδου. Η μη επιτηρούμενη μάθηση μπορεί να είναι αυτοσκοπός (ανακαλύπτοντας κρυμμένα μοτίβα σε δεδομένα) ή μέσο για ένα τέλος (χαρακτηριστικό της μάθησης). Στη συγκεκριμένη διπλωματική θα ασχοληθούμε με μία ιδιαίτερη μορφή μη επιβλεπόμενης μάθησης και πιο συγκεκριμένα με έναν αυτοκωδικοποιητή (Autoencoder).

2.1.2.3 Ενισχυτική Μάθηση

Στην ενισχυτική μάθηση ένα πρόγραμμα υπολογιστή αλληλεπιδρά με ένα δυναμικό περιβάλλον στο οποίο πρέπει να επιτευχθεί ένας συγκεκριμένος στόχος (όπως η οδήγηση ενός οχήματος), χωρίς κάποιος δάσκαλος να του λέει ρητά αν έχει φτάσει κοντά στο στόχο του ή όχι. Ένα άλλο παράδειγμα είναι να μάθει να παίζει ένα παιχνίδι εναντίον κάποιου αντιπάλου. Ουσιαστικά ο αλγόριθμος επιλέγει μια δράση, λαμβάνει μια ανατροφοδότηση από το περιβάλλον και μεταβαίνει σε μία κατάσταση. Από το ερέθισμα που του δίνεται κάθε φορά προσπαθεί να καταλάβει αν η επιλογή του ήταν καλή. Στην πάροδο του χρόνου μαθαίνει ποια είναι η επιθυμητή συμπεριφορά για να επιτύχει τη μέγιστη ανταμοιβή.

2.1.3 Βαθιά Μάθηση

Η πτώση των τιμών του υλικού των τελευταίων ετών καθώς και η ανάπτυξη των GPU για προσωπική χρήση, οδήγησε στην ανάπτυξη της ιδέας της Βαθιάς Μάθησης. Η βαθιά μάθηση προσπαθεί να μοντελοποιήσει τον τρόπο που ο ανθρώπινος εγκέφαλος επεξεργάζεται το φως και τον ήχο και τα μετατρέπει σε όραση και ακοή. Πρόκειται για ένα ιδιαίτερο είδος μηχανικής μάθησης, που αναπαριστά τον κόσμο ως μια εμφολευμένη ιεραρχία εννοιών, με κάθε έννοια να ορίζεται σε σχέση με άλλες απλούστερες. Ορισμένες επιτυχείς εφαρμογές της Βαθιάς μάθησης είναι η μηχανική όραση και η αναγνώριση ομιλίας.

2.1.4 Τεχνητά Νευρωνικά Δίκτυα

Το νευρωνικό δίκτυο είναι ένα δίκτυο από απλούς υπολογιστικούς κόμβους (νευρώνες), διασυνδεδεμένους μεταξύ τους. Είναι εμπνευσμένο από το Κεντρικό Νευρικό Σύστημα (ΚΝΣ) του ανθρώπου, το οποίο και προσπαθεί να προσομοιώσει. Οι νευρώνες είναι τα δομικά στοιχεία του δικτύου. Κάθε τέτοιος κόμβος δέχεται ένα σύνολο αριθμητικών εισόδων από διαφορετικές πηγές (είτε από άλλους νευρώνες, είτε από το περιβάλλον), επιτελεί έναν υπολογισμό με βάση αυτές τις εισόδους και παράγει μία έξοδο. Η εν λόγω έξοδος είτε κατευθύνεται στο περιβάλλον, είτε 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 επιτελούν κανέναν υπολογισμό, μεσολαβούν απλώς ανάμεσα στις περιβαλλοντικές εισόδους του δικτύου και στους υπολογιστικούς νευρώνες. Οι νευρώνες εξόδου διοχετεύουν στο περιβάλλον τις τελικές αριθμητικές εξόδους του δικτύου. Οι υπολογιστικοί νευρώνες πολλαπλασιάζουν κάθε είσοδό τους με το αντίστοιχο συναπτικό βάρος, το οποίο αποτελεί την ισχύ των συνδέσεων μεταξύ των νευρώνων, 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. Μαθηματικά όλο αυτό εκφράζεται ως εξής:

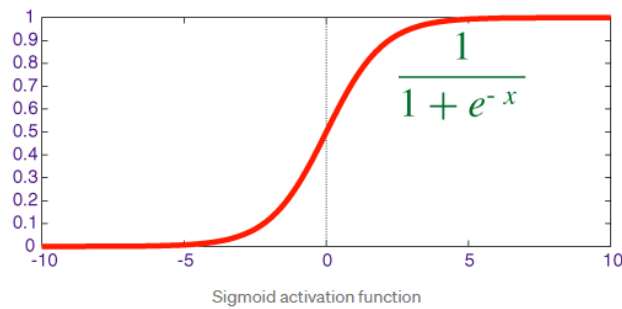
Εάν x_{ki} είναι η i -οστή είσοδος του k νευρώνα, w_{ki} : το i -οστό συναπτικό βάρος του k νευρώνα και $\phi(\cdot)$ η συνάρτηση ενεργοποίησης του νευρωνικού δικτύου, τότε η έξοδος y_k του k νευρώνα δίνεται από την εξίσωση:
$$y_k = \phi\left(\sum_{i=0}^N x_{ki}w_{ki}\right)$$

Το κύριο χαρακτηριστικό των νευρωνικών δικτύων είναι η εγγενής ικανότητα μάθησης. Ως μάθηση μπορεί να οριστεί η σταδιακή βελτίωση της ικανότητας του δικτύου να επιλύει κάποιο πρόβλημα (π.χ. η σταδιακή προσέγγιση μίας συνάρτησης). Η μάθηση επιτυγχάνεται μέσω της εκπαίδευσης, μίας επαναληπτικής διαδικασίας σταδιακής προσαρμογής των παραμέτρων του δικτύου (συνήθως των βαρών) σε τιμές κατάλληλες ώστε να επιλύεται με επαρκή επιτυχία το προς εξέταση πρόβλημα. Αφού ένα δίκτυο εκπαιδευτεί, οι παράμετροί του συνήθως «παγώνουν» στις κατάλληλες τιμές και από εκεί κι έπειτα είναι σε λειτουργική κατάσταση. Το ζητούμενο είναι το λειτουργικό δίκτυο να χαρακτηρίζεται από μία ικανότητα γενίκευσης: αυτό σημαίνει πως δίνει ορθές εξόδους για εισόδους καινοφανείς και διαφορετικές από αυτές με τις οποίες εκπαιδεύτηκε. Συνήθως τα δεδομένα που χρησιμοποιούνται από τα νευρωνικά δίκτυα έχουν μη γραμμικές σχέσεις μεταξύ εισόδων και εξόδων. Γι'αυτον ακριβώς το λόγο χρησιμοποιούνται συνήθως μη γραμμικές συναρτήσεις ενεργοποίησης. Η συνάρτηση ενεργοποίησης είναι μια συνάρτηση η οποία δέχεται ως είσοδο όλες τις εισόδους του νευρώνα και βγάζει μια έξοδο η οποία στις περισσότερες περιπτώσεις ερμηνεύεται ως 0 ή 1 αντίστοιχα, συνήθως θετικές τιμές ερμηνεύονται ως 1 και αρνητικές ως 0. Τρεις συναρτήσεις ενεργοποίησης οι οποίες χρησιμοποιούνται ως επί των πλείστων είναι οι εξής:

- Sigmoid

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

Η σιγμοειδής συνάρτηση παίρνει τιμές στο διάστημα (0,1). Για αυτόν ακριβώς το λόγο χρησιμοποιείται πολύ συχνά όταν θέλουμε ο νευρώνας μας να βγάλει ως έξοδο την πρόβλεψη της τιμής μίας πιθανότητας, αφού οι πιθανότητες έχουν πάντα πεδίο τιμών το [0,1]. Ένα χαρακτηριστικό παράδειγμα είναι χρήση ενός νευρωνικού δικτύου για την ταξινόμηση των δεδομένων σε δύο πιθανές κλάσεις. Σε αυτή την περίπτωση, η συνάρτηση ενεργοποίησης του επιπέδου εξόδου του νευρωνικού δικτύου θα είναι η σιγμοειδής και η τιμή της συμβολίζει την πιθανότητα η είσοδος να ανήκει σε μία από τις δύο κλάσεις.

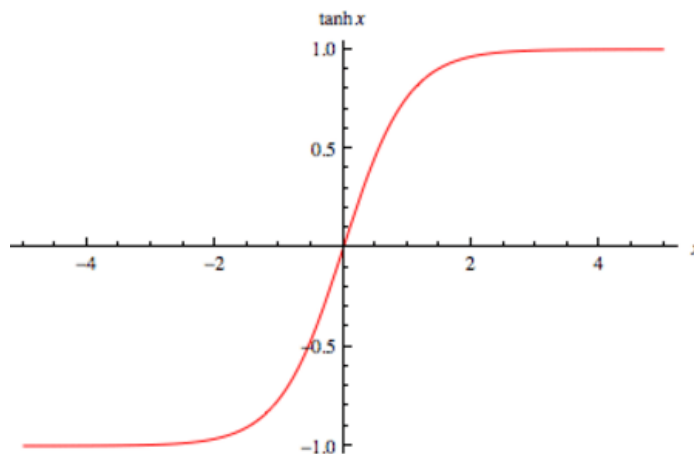


Σχήμα 2.1: Sigmoid Function

- Tanh

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

Η υπερβολική εφαπτομένη έχει ίδια μορφή με τη σιγμοειδή όμως το σύνολο τιμών της είναι το (-1,1). Βασικό προτέρημά της είναι πως αρνητικές τιμές αντιστοιχίζονται σε καθαρά αρνητικές τιμές της συνάρτησης, κι αντίστοιχα για τις θετικές, ενώ οι τιμές κοντά στο μηδέν θα βρίσκονται κοντά στο μηδέν και στην έξοδο της συνάρτησης.



Σχήμα 2.2: Tanh Function

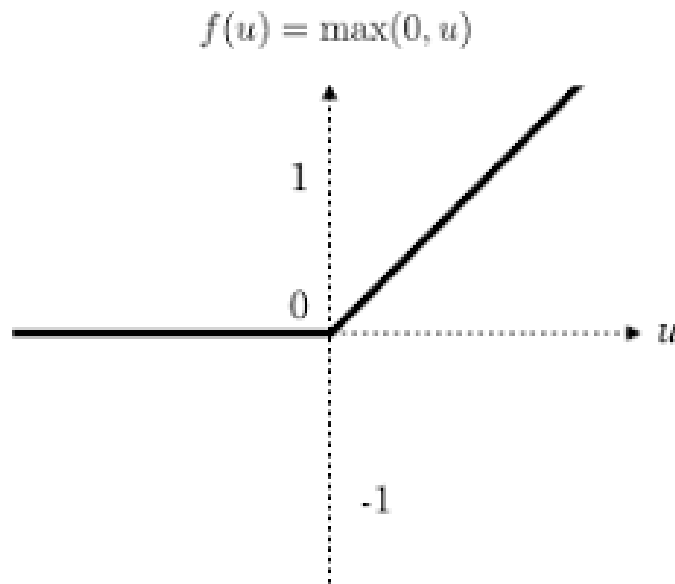
- ReLU

$$f(x) = \begin{cases} 0, & \text{if } x \leq 0 \\ x, & \text{if } x > 0 \end{cases}$$

Η συνάρτηση ReLU που το όνομά της προκύπτει απ'τό Rectified Linear Unit είναι η πιο ευρέως χρησιμοποιούμενη συνάρτηση ενεργοποίησης αυτή τη στιγμή μιας και χρησιμοποιείται σε όλα σχεδόν τα συνελκτικά νευρωνικά δίκτυα και τη βαθιά μάθηση. Το σύνολο τιμών της είναι το $[0, \infty)$. Ουσιαστικά υπολογίζει μία σύγκριση μεταξύ της εισόδου x και του μηδέν, και επιλέγει το μέγιστο μεταξύ των δύο.

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

Αυτό απλοποιεί πολύ τους υπολογισμούς που απαιτούνται κατά την εκπαίδευση βελτιώνοντας σημαντικά το χρόνο και τους πόρους που απαιτούνται για την εκπαίδευση.



Σχήμα 2.3: ReLu Function

2.1.5 Συνελκτικά Νευρωνικά Δίκτυα

2.1.5.1 Ορισμός

Τα συνελκτικά νευρωνικά δίκτυα (CNN) είναι ένα ιδιαίτερο είδος νευρωνικών δικτύων που χρησιμοποιούνται κυρίως στην ταξινόμηση εικόνων αλλά και στην αναγνώριση αντικειμένων. Για παράδειγμα συνελκτικά νευρωνικά δίκτυα (CNN) χρησιμοποιούνται σε εφαρμογές αναγνώρισης προσώπου, αντικειμένων, σημάτων οδήγησης, καρκινικών όγκων και διαφόρων άλλων.

Το όνομα "συνελκτικό νευρικό δίκτυο" υποδεικνύει ότι το δίκτυο χρησιμοποιεί μια μαθηματική λειτουργία που ονομάζεται συνέλιξη. Τα συνελκτικά νευρωνικά δίκτυα είναι απλώς νευρωνικά δίκτυα που χρησιμοποιούν συνέλιξη στη θέση γενικής μήτρας πολλαπλασιασμού σε τουλάχιστον ένα από τα στρώματά τους. (Goodfellow et al., 2016)

$$s(t) = \int x(a)w(t - a)da$$

Αυτή είναι πράξη της συνέλιξης, η οποία συνήθως συμβολίζεται με έναν αστερίσκο.

$$s(t) = x(t) * w(t)$$

Η λέξη συνέλιξη προκύπτει από το λατινογενές "convolvere", και η μαθηματική της σημασία είναι η ακέραια μέτρηση του πόσο δύο συναρτήσεις αλληλοεπικαλύπτονται καθώς η μία περνάει πάνω από την άλλη.

Η αποτελεσματικότητα των συνελικτικών νευρωνικών δικτύων στην αναγνώριση και επεξεργασία εικόνας είναι ένας από τους κύριους παράγοντες που έφεραν στο προσκύνιο την βαθιά μάθηση. Συγκεκριμένα η επιτυχία μίας αρχιτεκτονικής βαθιάς μάθησης, ενός συνελικτικού νευρωνικού δικτύου ονομαζόμενη AlexNet, το 2012 στον διαγωνισμό ImageNet, έκαναν τον κόσμο να ασχοληθεί αποφασιστικά μαζί της. Η αρχιτεκτονική AlexNet αποτελείται από 8 επίπεδα. Τα πρώτα πέντε είναι συνελικτικά επίπεδα, ορισμένα από τα οποία ακολουθούνται από επίπεδα συγκέντρωσης και τα τελευταία τρία επίπεδα είναι πυκνά - πλήρως συνδεδεμένα επίπεδα. Η συνάρτηση ενεργοποίησης που χρησιμοποιήθηκε στα κρυφά επίπεδα ήταν η ReLU.

Απο τότε τα CNN πρωτοστατούν σε διάφορους τομείς όπως η οπτική αναγνώριση χαρακτηριστικών, η όραση υπολογιστών με εφαρμογές, στην ρομποτική, σε αυτο-οδηγούμενα αυτοκίνητα, σε drone, σε ιατρικές διαγνώσεις και πολλά άλλα.

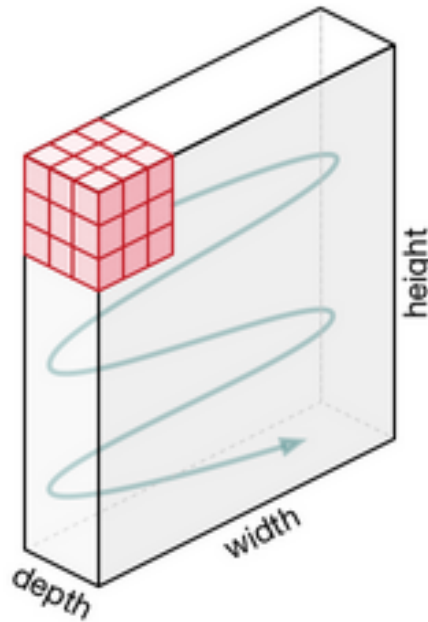
Για να καταλάβουμε γιατί τα CNN αποδίδουν τόσο καλύτερα σε σχέση με τα απλά νευρωνικά δίκτυα πρέπει να διευκρινήσουμε κάποια βασικά πράγματα για την λειτουργία τους. Ας διευκρινίσουμε αρχικά πως κάθε RGB εικόνα δύο διαστάσεων, στην γλώσσα των υπολογιστών αναπαρίσταται ως ένας πίνακας τριών διαστάσεων με την τρίτη διάσταση να αποτελεί το κανάλι χρώματος. Ας πάρουμε για παράδειγμα μία εικόνα μεγέθους $200 \times 200 \times 3$ εικονοστοιχείων (pixel), ο κάθε νευρώνας του πρώτου επιπέδου θα πρέπει να έχει $200 \times 200 \times 3 = 120000$ παραμέτρους, άρα για ένα μόνο πυκνό επίπεδο θα χρειαζόμασταν μερικά εκατομμύρια παραμέτρους. Ο αριθμός αυτός φυσικά θα εκτοξευόταν αν είχαμε εικόνες υψηλότερης ανάλυσης για παράδειγμα εικόνες $8K(7680 \times 4320)$. Επίσης προφανώς θα χρειαζόμασταν περισσότερα του ενός πυκνά επίπεδα για να αναπαρασταθούν πιο περίπλοκες συναρτήσεις, καταλήγοντας σε ένα τεράστιο δίκτυο με αποτέλεσμα να είναι πολύ δύσκολη και χρονοβόρα η εκπαίδευσή του και πιθανότατα χωρίς να έχει τη δυνατότητα να αποδώσει καλά λόγω της υπερμοντελοποίησης (Overfitting). Τα συνελικτικά δίκτυα από την άλλη εκμεταλλεύονται τη χωρική συσχέτιση των δεδομένων εισόδου. Αντί να αντιμετωπίζεται κάθε pixel μεμονωμένα ως μια απλή τιμή εισόδου, μέσω της συνέλιξης με τους πυρήνες (kernels) εξάγονται χαρακτηριστικά από συστάδες εικονοστοιχείων με τα οποία γίνεται η συνέλιξη κάθε φορά. Αυτό δίνει τη δυνατότητα στα CNN να επιτυγχάνουν πολύ καλύτερα αποτελέσματα από τα απλά νευρωνικά δίκτυα, με πολύ μικρότερο αριθμό παραμέτρων. [O'Shea and Nash \(2015\)](#)

2.1.5.2 Αρχιτεκτονική Συνελικτικών Νευρωνικών Δικτύων

Βασικά Επίπεδα Υπάρχουν τρία βασικά επίπεδα τα οποία χρειάζονται για να δομήσουμε ένα συνελικτικό νευρωνικό δίκτυο.

- Το βασικότερο όλων των επιπέδων και από το οποίο πήραν και το όνομα τους τα δίκτυα, είναι το συνελικτικό επίπεδο (Convolutional Layer). Στο συγκεκριμένο επίπεδο υπάρχουν

ορισμένοι πυρήνες-φίλτρα των οποίων τις διαστάσεις καλούμαστε να προσδιορίσουμε, πρέπει όμως σίγουρα να είναι μικρότερες αυτών της εικόνας εισόδου, καθώς και το κατά πόσα pixel μετακινείται το φίλτρο μας κάθε φορά που εκτελείται η πράξη της συνέλιξης, μεταξύ του φίλτρου και του αντίστοιχου κομματιού της εικόνας πάνω από το οποίο βρίσκεται το φίλτρο την εκάστοτε στιγμή, το οποίο και συμβαίνει μέχρι να γίνει ένα πέρασμα από ολόκληρη την εικόνα. Τα αποτελέσματα αυτής της συνέλιξης εξάγονται και διαμορφώνουν την είσοδο προς το επόμενο επίπεδο ή σπανίως την έξοδο προς το περιβάλλον.

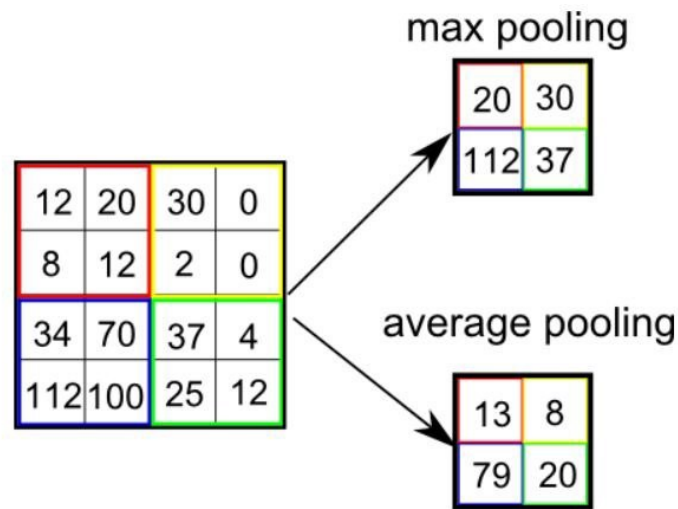


Movement of the Kernel

Σχήμα 2.4: Convolutional Kernel

- Το δεύτερο βασικό είδος επιπέδων που χρησιμοποιείται στα συνελκτικά νευρωνικά δίκτυα είναι τα επίπεδα συγκέντρωσης ή υπο-δειγματοληψίας (Pooling Layer). Ουσιαστικά επιτελούν μια διαδικασία υπο-δειγματοληψίας των δεδομένων με στόχο την μείωση των παραμέτρων του δικτύου και την κατανόηση των κυρίαρχων χαρακτηριστικών της εικόνας. Συγκεκριμένα μειώνεται ο αριθμός των διαστάσεων του αποτελέσματος της συνέλιξης που έχει προηγηθεί στο προηγούμενο επίπεδο. Υπάρχουν δύο βασικοί τρόποι ώστε να συμβεί αυτό. Το Max Pooling και το Average Pooling. Το Max Pooling επιστρέφει την μέγιστη τιμή του κομματιού της εικόνας που συνελίχθηκε με τον πυρήνα, αφού έλαβε χώρα η συνέλιξη, και το Average Pooling επιστρέφει την μέση τιμή όλων των τιμών που προέκυψαν μετά την συνέλιξη για το εκάστοτε τμήμα της εικόνας.

Ένα συνελκτικό επίπεδο σε συνδυασμό με ένα επίπεδο συγκέντρωσης αποτελούν μαζί το n -οστό επίπεδο ενός συνελκτικού δικτύου. Ο αριθμός των επιπέδων αυτών μπορεί να αυξηθεί



Σχήμα 2.5: Max-Average Pooling

αναλόγως με το βάθος των λεπτομερειών της εικόνας που θέλουμε το δίκτυό μας να κατανοήσει.

- Με αυτά τα δύο επίπεδα έχουμε καλύψει ό,τι χρειάζεται ένα συνελκτικό δίκτυο για την κατανόηση χαρακτηριστικών. Το τρίτο επίπεδο που αναφέρεται στην βιβλιογραφία είναι το πλήρως συνδεδεμένο ή πυκνό επίπεδο(Fully Connected Layer). Κάθε κόμβος αυτού του επιπέδου είναι συνδεδεμένος με τις εξόδους όλων των κόμβων του προηγούμενου επιπέδου. Τα πυκνά επίπεδα όμως χρησιμοποιούνται συνήθως ως επίπεδα εξόδου σε εργασίες κατηγοριοποίησης(Classification) και εμείς δεν τα έχουμε συμπεριλάβει στην υλοποίησή μας. [O'Shea and Nash \(2015\)](#)

2.2 Autoencoder - Αυτοκωδικοποιητής

Ο Αυτοκωδικοποιητής αποτελεί ένα είδος τεχνητού νευρωνικού δικτύου (ΤΝΔ), μη επιβλεπόμενης μάθησης, αν και στην πράξη εκπαιδεύεται μέσω τεχνικών επιβλεπόμενης μάθησης, που ονομάζονται αυτοεπιβλεπόμενη μάθηση. Ένας αυτοκωδικοποιητής αποτελείται από δύο κύρια δομικά στοιχεία : τον κωδικοποιητή(encoder) και τον αποκωδικοποιητή(decoder). Αρχικά, μέσω του κωδικοποιητή μαθαίνει πως να κωδικοποιεί και να συμπιέζει δεδομένα με αποτελεσματικό τρόπο, συνδυάζοντας τα σημαντικότερα χαρακτηριστικά τους, αγνοώντας τον θόρυβο, και μεταφέροντας τα σε έναν χώρο λιγότερων διαστάσεων. Οι μεταβλητές αυτές που προκύπτουν στον χώρο των μειωμένων διαστάσεων, συνήθως απακαλούνται ως λανθάνουσες μεταβλητές (Latent variables). Στη συνέχεια, μέσω του decoder, ο αυτοκωδικοποιητής αποκωδικοποιεί τα συμπιεσμένα αυτά δεδομένα και ανακατασκευάζει μία όσο το δυνατόν πιο πιστή αναπαράσταση των δεδομένων εισόδου.(Latent variables).

Η συνάρτηση οπισθοδιάδωσης του λάθους, σε έναν αυτοκωδικοποιητή, ονομάζεται απώλεια ανακατασκευής (reconstruction loss), διότι τιμωρεί το μοντέλο όταν οι έξοδοί του είναι πολύ διαφορετικές από τις εισόδους. Στην πιο απλή του μορφή, ένας αυτοκωδικοποιητής(autoencoder) μοιάζει στη δομή του με ένα μικρό multi-layer perceptron(MLP), με τον κωδικοποιητή(encoder)

και τον αποκωδικοποιητή(decoder) να αποτελούνται από ένα πλήρως συνδεδεμένο επίπεδο έκαστος. Με τον περιορισμό πάντα, ότι ο αριθμός των νευρώνων του επιπέδου εξόδου πρέπει να είναι ίδιος με αυτόν του επιπέδου εισόδου. Ένας αυτοκωδικοποιητής(autoencoder) του οποίου η εσωτερική αναπαράσταση έχει μικρότερο αριθμό διαστάσεων απ'ότι οι εισοδοί του χαρακτηρίζεται ως *undercomplete*. Ένας *undercomplete* αυτοκωδικοποιητής είναι ανίκανος απλά να αντιγράψει τυφλά τις εισόδους του στις κωδικοποιήσεις τους, και αντ'αυτού αναγκάζεται να μάθει τα πιο σημαντικά χαρακτηριστικά των δεδομένων, αγνοώντας τα υπόλοιπα. Εδώ πρέπει να αναφερθεί ότι σε περίπτωση που ο αυτοκωδικοποιητής χρησιμοποιεί μόνο γραμμικές συναρτήσεις ενεργοποίησης και ως απώλεια ανακατασκευής έχει οριστεί η *mean squared error (MSE)*, τότε το κομμάτι του κωδικοποιητή(encoder) του μοντέλου ουσιαστικά εφαρμόζει *principal component analysis (PCA)*.

Η ιδέα των αυτοκωδικοποιητών έγινε δημοφιλής στο πεδίο των νευρωνικών δικτύων εδώ και δεκαετίες, με τις πρώτες εφαρμογές τους ήδη από την δεκαετία του 1980. Οι πιο χαρακτηριστικές εφαρμογές τους είναι η μείωση διαστατικότητας (*Dimensionality reduction*), που μεταξύ άλλων μειώνει σημαντικά τον χρόνο εκπαίδευσης ενός δικτύου, και η εκμάθηση χαρακτηριστικών (*Feature learning*), καθώς οι πληροφορίες των δεδομένων αναγκάζονται να συμπυκνωθούν σε έναν πολύ περιορισμένο χώρο, αλλά οι αυτοκωδικοποιητές έχουν γίνει ευρέως διαδεδομένοι για την εκμάθηση γεννητικών μοντέλων δεδομένων, όπου ως γεννητικά μοντέλα, είναι ικανοί να δημιουργούν με τυχαίο τρόπο, νέα, τεχνητά δεδομένα, που μοιάζουν με αυτά στα οποία έχουν εκπαιδευτεί [Dong et al. \(2018\)](#). Κάποια από τα πιο ισχυρά μοντέλα τεχνητής νοημοσύνης της τελευταίας δεκαετίας περιέχουν αυτοκωδικοποιητές.

Βασική αρχιτεκτονική

- Όπως είδαμε και προηγουμένως ένας αυτοκωδικοποιητής αποτελείται από δύο κύρια δομικά μέρη. Έναν κωδικοποιητή και έναν αποκωδικοποιητή, που για το παρόν παράδειγμα ας τους ορίσουμε ως τις συναρτήσεις ϕ και ψ αντίστοιχα, έτσι ώστε:

$$\phi : X \rightarrow F$$

$$\psi : F \rightarrow X$$

$$\phi, \psi = \arg \min_{\phi, \psi} \|X - (\psi \circ \phi)X\|^2$$

Στην απλούστερη περίπτωση ενός μόνο κρυφού επιπέδου(hidden layer), ο κωδικοποιητής παίρνει την είσοδο $x \in R^d = \mathcal{X}$

και την χαρτογραφεί στον χώρο $h \in R^p = \mathcal{F}$:

$$h = \sigma (W x + b)$$

Ο χώρος h συνήθως αναφέρεται ως κώδικας, χώρος λανθανουσών μεταβλητών, ή ως λανθάνουσα αναπαράσταση. Εδώ, το σ αποτελεί μία συνάρτηση ενεργοποίησης, όπως για παράδειγμα η σιγμοειδής συνάρτηση. Το W είναι ο πίνακας βαρών του δικτύου και το b είναι το διάνυσμα πόλωσης(bias vector). Συνήθως οι τιμές των βαρών και των πολώσεων αρχικοποιούνται τυχαία, και στη συνέχεια ενημερώνονται επαναληπτικά κατά την εκπαίδευση του δικτύου μέσω του μηχανισμού της ανάστροφης αναδιάδοσης (*backpropagation*). Έπειτα

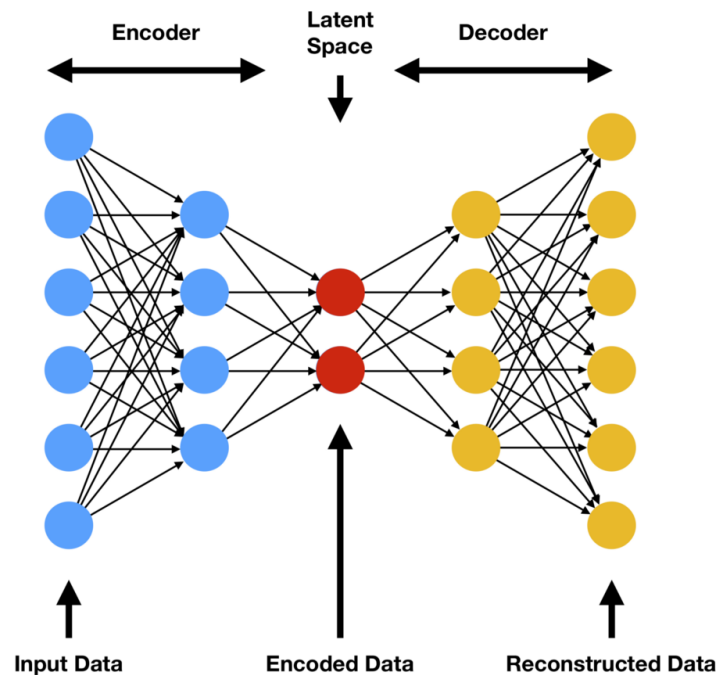
ο αποκωδικοποιητής αντιστοιχίζει τον χώρο h ώστε να ανασκευάσει το x' που θα είναι ίδιων διαστάσεων με το x :

$$x' = \sigma'(W' h + b')$$

όπου τα σ' , W' , and b' του αποκωδικοποιητή μπορεί να μην σχετίζονται με τα αντίστοιχα σ , W , and b του κωδικοποιητή.

Οι αυτοκωδικοποιητές σχεδιάζονται έτσι ώστε να ελαχιστοποιούν το λάθος ανακατασκευής (reconstruction error), πολύ συχνά χρησιμοποιούνται συναρτήσεις τετραγωνικών λαθών (mean squared error):

$$\mathcal{L}(x, x') = \|x - x'\|^2 = \|x - \sigma'(W'(\sigma(Wx + b)) + b')\|^2$$



Σχήμα 2.6: Autoencoder Architecture

2.2.1 Εφαρμογές των Αυτοκωδικοποιητών

Οι autoencoders έχουν χρησιμοποιηθεί με επιτυχία σε εφαρμογές μείωσης διαστατικότητας (Dimensionality reduction) και ανάκτησης πληροφορίας (information retrieval). Η μείωση διαστατικότητας ήταν μία από τις πρώτες εφαρμογές της βαθιάς μάθησης, και από τα βασικά κίνητρα για την μελέτη των αυτοκωδικοποιητών. Για παράδειγμα, οι Hinton και Salakhutdinov (2006) [Hinton and Salakhutdinov \(2006\)](#) εκπαίδευσαν μία στοίβα από RBMs (Restricted Boltzmann machines) και χρησιμοποίησαν στη συνέχεια τα βάρη τους για να αρχικοποιήσουν τις τιμές ενός deep autoencoder, με σταδιακά μικρότερα κρυφα επίπεδα, καταλήγοντας σε ένα bottleneck επίπεδο των 30 μονάδων. Ο αλγόριθμος που προέκυψε, επέφερε μικρότερο λάθος ανακατασκευής σε σχέση με το PCA σε 30 διαστάσεις, και η προκύπτουσα αναπράσταση ήταν ποιοτικά ευκολότερη να ερμηνευθεί και να συσχετιστεί με τις υποκείμενες κατηγορίες, οι οποίες παρουσιάζονταν ως σαφώς διαχωρισμένες συστάδες. Οι αναπαραστάσεις χαμηλότερων διαστάσεων μπορούν να αυξήσουν την απόδοση πολλών

εργασιών, όπως για παράδειγμα η κατηγοριοποίηση(classification). Τα μοντέλα αυτά απαιτούν λιγότερη μνήμη και μικρότερο χρόνο εκτέλεσης(runtime). Πολλές μορφές μείωσης διαστατικότητας τοποθετούν σημασιολογικά συναφή παραδείγματα το ένα κοντά στο άλλο [Salakhutdinov and Hinton \(2007\)](#), [Torralba et al. \(2008\)](#). Μία ακόμη διαδικασία που επωφελείται πολύ απο την μείωση διαστατικότητας είναι η ανάκτηση πληροφορίας(information retrieval), η διαδικασία εύρεσης εισόδων σε μία βάση δεδομένων που προσομοιάζουν τις ζητούμενες εισόδους. Αυτή η διαδικασία έχει τα συνήθη οφέλη απο τη μείωση διαστατικότητας, αλλά επιπλέον εποφελείται από το γεγονός ότι η αναζήτηση μπορεί να γίνει εξαιρετικά αποδοτική σε συγκεκριμένα είδη χώρων χαμηλών διαστάσεων. Ειδικά αν εκπαιδεύσουμε τον αλγόριθμο μείωσης διαστάσεων να παράγει κώδικα που είναι χαμηλής διαστατικότητας και δυαδικής μορφής, τότε μπορούμε να αποθηκεύσουμε όλες τις εισόδους της βάσης δεδομένων σε έναν πίνακα κατακερματισμού, που χαρτογραφεί τα διανύσματα διαδικού κώδικα σε εισόδους. Αυτός ο πίνακας μας επιτρέπει να ανακτήσουμε πληροφορίες απλώς επιστρέφοντας όλες τις εισόδους της βάσης δεδομένων που έχουν τον ίδιο διαδικό κώδικα με την αντίστοιχη ζητούμενη είσοδο. Αυτή η προσέγγιση της ανάκτησης πληροφορίας μέσω της μείωσης διαστατικότητας και της διαδικοποίησης ονομάστηκε σημασιολογικός κατακερματισμός(semantic hashing) και έχει εφαρμοστεί τόσο σε είσοδο κειμένου [Salakhutdinov and Hinton \(2007\)](#), [Salakhutdinov and Hinton \(2009\)](#) όσο και σε εικόνες [Torralba et al. \(2008\)](#), [Hinton et al. \(2011\)](#).

Για την παραγωγή τέτοιου δυαδικού κώδικα, χρειάζεται η χρήση της σιγμοειδούς συνάρτησης ενεργοποίησης στο τελευταίο επίπεδο, οι έξοδοι της οποίας πρέπει να εκπαιδευτούν ώστε να παραμένουν κοντά στο 0 ή στο 1, ανεξαρτήτως της τιμής της εισόδου. Αυτό μπορεί να επιτευχθεί εισάγοντας επιπλέον θόρυβο ακριβώς πριν τη μη γραμμικότητα της σιγμοειδούς συνάρτησης κατά τη διάρκεια της εκπαίδευσης. Το μέγεθος του θορύβου θα πρέπει να αυξάνεται με την πάροδο του χρόνου, και για να αντισταθμιστεί ο θόρυβος και να διατηρήσουμε όσο το δυνατόν περισσότερη πληροφορία, το δίκτυό μας θα πρέπει να αυξάνει τις εισόδους της σιγμοειδούς συνάρτησης, έως ότου επέλθει κορεσμός. Η ιδέα της εκμάθησης μιας συνάρτησης κατακερματισμού έχει εξερευνηθεί περαιτέρω σε διάφορες κατευθύνσεις, συμπεριλαμβανομένης και της ιδέας της εκπαίδευσης των αναπαραστάσεων για τη βελτιστοποίηση του λάθους ανακατασκευής, συνδεδεμένου με τη διαδικασία εύρεσης γειτονικών παραδειγμάτων στον πίνακα κατακερματισμού. [Norouzi and Fleet \(2011\)](#)

Μία ακόμη πολύ σημαντική εφαρμογή είναι η χρήση ενός συγκεκριμένου είδους αυτοκωδικοποιητών, των Denoising Autoencoders, που ειδικεύονται στην ανακατασκευή διεφθαρμένων δεδομένων, αγνοώντας τα δεδομένα θορύβου, για την καλύτερη εκπαίδευση βαθιών νευρωνικών δικτύων σε προβλήματα κατηγοριοποίησης, δίνοντας εξαιρετικά καλά αποτελέσματα. [Vincent et al. \(2010\)](#), [Vincent et al. \(2008\)](#)

2.3 Βασικές έννοιες Τεχνιτών Νευρωνικών Δικτύων(TND)

Στην παρούσα υποενότητα θα αναφερθούμε σε μερικά από τα βασικά δομικά στοιχεία ενός τεχνιτού νευρωνικού δικτύου, με τα οποία πειραματιστήκαμε στη συνέχεια.

2.3.1 Αλγόριθμοι Βελτιστοποίησης

- **Gradient Descent.** Το gradient descent είναι ένας επαναληπτικός αλγόριθμος βελτιστοποίησης για την εύρεση ελαχίστων μίας συνάρτησης. Στην περίπτωση μας θα χρησιμοποιηθεί για την εύρεση των ελαχίστων της συνάρτησης απώλειας (Loss Function) κατά το τρέξιμο του αλγορίθμου οπισθοδιάδωσης του λάθους (Backpropagation algorithm), όπως θα δούμε παρακάτω. Είναι μέθοδος που λαμβάνει υπόψη μόνο τις μερικές παραγώγους πρώτης τάξης της συνάρτησης, και εκμεταλλεύεται το γεγονός ότι το διάνυσμά τους δείχνει την κατεύθυνση στον n -διάστατο χώρο προς την οποία θα αυξηθεί περισσότερο η τιμή της. Γι' αυτό και για την εύρεση των ελαχίστων σημείων ο αλγόριθμος κάνει βήματα προς την αντίθετη κατεύθυνση. Αρχικά οι μερικές παράγωγοι πολλαπλασιάζονται με έναν αρκετά μικρό αριθμό που ονομάζεται ρυθμός μάθησης (learning rate), ο οποίος καθορίζει το μέγεθος των αλλαγών στην κατεύθυνση των παραγώγων, και είναι τόσο μικρός για λόγους σταθερότητας της εκπαίδευσης. Υπάρχουν τρεις παραλλαγές του gradient descent που χρησιμοποιούνται συχνά και διαφέρουν στην ποσότητα των δεδομένων που λαμβάνονται υπόψη κάθε φορά πριν τον υπολογισμό των μερικών παραγώγων.

1. **Batch Gradient Descent.** Στην αρχική υλοποίηση του gradient descent, που ονομάστηκε batch gradient descent, οι μερικές παράγωγοι της συνάρτησης απώλειας υπολογίζονταν κάθε φορά αφού είχαν ληφθεί υπόψη όλα τα δεδομένα εκπαίδευσης. Το γεγονός πως για κάθε βήμα εκπαίδευσης απαιτούταν ο υπολογισμός των παραγώγων για όλα τα δεδομένα εκπαίδευσης είχε ως αποτέλεσμα το batch gradient descent να είναι εξαιρετικά αργό, και πρακτικά μη υλοποιήσιμο από άποψης χρόνου για μεγάλα σετ δεδομένων.
2. **Stochastic Gradient Descent.** Το stochastic gradient descent είναι το ακριβώς αντίθετο του batch gradient descent, και ενημερώνει τις παραμέτρους για κάθε ένα δεδομένο εκπαίδευσης. Η λογική πίσω από το stochastic gradient descent είναι ότι σε μεγάλα σετ δεδομένων το batch gradient descent πραγματοποιεί παραπάνω υπολογισμούς απ' όσους χρειάζονται, αφού υπολογίζει ξανά και ξανά σχεδόν ίδιες παραγώγους για παρόμοια δεδομένα. Το stochastic gradient descent είναι φυσικά πολύ πιο γρήγορο, όμως οδηγεί σε σχετική αστάθεια στην εκπαίδευση, αφού η έλλειψη πληροφοριών για την στατιστική κατανομή των δεδομένων καθιστά τον αλγόριθμο ευάλωτο σε ακραία, μη αντιπροσωπευτικά δεδομένα.
3. **Mini-batch Gradient Descent.** Το mini-batch gradient descent συνδυάζει τα πλεονεκτήματα των δύο παραπάνω προσεγγίσεων. Αυτό το πετυχαίνει ενημερώνοντας τις παραμέτρους για κάθε mini-batch αριθμό δεδομένων. Έτσι, και μειώνεται η έντονη διακύμανση των συνεχών ενημερώσεων του stochastic gradient descent, και δεν είναι τόσο απαγορευτικό από άποψη χρόνου και μνήμης όσο το batch gradient descent, αφού μπορούμε να ορίσουμε το μέγεθος του mini-batch όπως εμείς επιθυμούμε. Παράλληλα, λόγω της πολύ συχνής χρήσης του τα τελευταία χρόνια, έχουν αναπτυχθεί βιβλιοθήκες που χρησιμοποιούν βελτιστοποιήσεις για εξαιρετικά γρήγορες πράξεις μεταξύ πινάκων. Πολλές φορές επικρατεί στη βιβλιογραφία η χρήση του ονόματος stochastic gradient descent (SGD) ακόμα και όταν εφαρμόζεται mini-batch gradient descent. Ένα πρόβλημα

που εμφανίζεται συχνά με τη χρήση του mini-batch gradient descent είναι η επιλογή του κατάλληλου learning rate: Αν είναι πολύ μικρό, η σύγκλιση θα είναι απαγορευτικά αργή, ενώ αν είναι πολύ μεγάλο, είναι πολύ πιθανό προς το τέλος της εκπαίδευσης η συνάρτηση απώλειας να ταλαντώνεται γύρω από κάποιο ελάχιστο σημείο χωρίς ποτέ να φτάσει σε αυτό. Για το λόγο αυτό έχουν αναπτυχθεί αλγόριθμοι βελτιστοποίησης που ρυθμίζουν το learning rate αναλόγως κατά τη διάρκεια της εκπαίδευσης. Δύο από τους πιο σημαντικούς και με τους οποίους πειραματιστήκαμε και εμείς κατά κύριο λόγο είναι ο Adam και ο RMSProp.

- RMSProp. Η λειτουργία του RMSProp έγκειται στο ότι δεν αποθηκεύονται οι παράγωγοι σε όλα τα βήματα από την αρχή της εκπαίδευσης, αντίθετα αποθηκεύονται μόνο οι τελευταίοι παράγωγοι, των οποίων η τιμή είναι και πιο έμπιστη. Έτσι, τυχόν μεγάλες ανανεώσεις στα πρώτα βήματα της εκπαίδευσης, όπου τα βάρη είναι στην ουσία τυχαία, σταματούν να έχουν σημασία καθώς το δίκτυο συγκλίνει, ενώ ως επιπλέον πλεονέκτημα του αλγορίθμου οι μικρές ανανεώσεις στα επόμενα στάδια της εκπαίδευσης τον καθιστούν πολύ πιο κατάλληλο για “τελειοποίηση” (fine tuning) των παραμέτρων.
- Adam (Adaptive Moment Estimation). Ο Adam είναι ένας ακόμη αλγόριθμος που υπολογίζει και προσαρμόζει τις τιμές των learning rates για κάθε παράμετρο. Πέρα από το να αποθηκεύει μία εκθετικά φθίνουσα μέση τιμή των τετραγώνων προηγούμενων παραγώγων, όπως κάνει ο RMSProp, ο Adam αποθηκεύει επίσης και μία εκθετικά φθίνουσα τιμή των προηγούμενων παραγώγων σε εντιστοιχία με το πως λειτουργεί το momentum. [Kingma and Ba \(2017\)](#)

2.3.2 Αλγόριθμος οπισθοδιάδοσης του λάθους (Backpropagation of the error algorithm)

Ο αλγόριθμος αυτός ανήκει στην κατηγορία της επιβλεπόμενης μάθησης, αλλά χρησιμοποιείται και σε εφαρμογές ήμι-επιβλεπόμενης μάθησης όπως είναι οι αυτοκωδικοποιητές. Ο λόγος που χρειαζόμαστε την εκ των προτέρων γνώση της επιθυμητής εξόδου για κάθε αντίστοιχη είσοδο είναι πως ο αλγόριθμος αυτός προσπαθεί να υπολογίσει τα συναπτικά βάρη του δικτύου που ελαχιστοποιούν τη συνάρτηση κόστους. Η συνάρτηση κόστους συνήθως είναι το τετραγωνικό σφάλμα της τρέχουσας εξόδου του νευρωνικού δικτύου σε σχέση με την επιθυμητή έξοδο. Η λειτουργία του αλγορίθμου χωρίζεται σε τέσσερα διακριτά μέρη :

1. Διάδοση : Αρχικά υπολογίζεται η έξοδος του κάθε νευρώνα σε κάθε επίπεδο του δικτύου, με σκοπό τον υπολογισμό των εξόδων του τελευταίου επιπέδου. Στην περίπτωση που πρόκειται για νευρώνα κρυφού επιπέδου, η κάθε είσοδος αποτελεί έξοδο νευρώνα προηγούμενου επιπέδου ενώ αν πρόκειται για νευρώνα του επιπέδου εισόδου, η έξοδος ισούται με την είσοδό του. Στη συνέχεια εξάγονται οι διαφορές μεταξύ επιθυμητών και πραγματικών τιμών εξόδου στους νευρώνες του τελευταίου επιπέδου. Έτσι, προκύπτει η διαφορά : $d_b = target - output$
2. Ανανέωση Βαρών Εξόδου : Τα βάρη εξόδου ανανεώνονται σύμφωνα με την εξίσωση $W'_{AB} = W_{AB} + d_b * output_A$ όπου W'_{AB} το νέο βάρος και W_{AB} το παλιό.

3. Υπολογισμός βαρών κρυμμένων επιπέδων : Αντίθετα με το επίπεδο εξόδου, στα κρυμμένα επίπεδα δεν υπάρχει κάποια γνωστή επιθυμητή έξοδος με την οποία μπορούμε να συγκρίνουμε την πραγματική τιμή εξόδου που έχουμε και να καταλήξουμε άμεσα σε ένα συμπέρασμα. Έτσι πραγματοποιείται οπισθοδιάδοση της τιμής δ που υπολογίσαμε προηγουμένως, από το επίπεδο εξόδου προς τα κρυφά επίπεδα του δικτύου. Οι διαφορές δ των νευρώνων εξόδου εισάγονται στους νευρώνες των κρυφών επιπέδων και από αυτούς λαμβάνονται οι αντίστοιχες διαφορές.
4. Προσαρμογή βαρών κρυφών επιπέδων : Αντίστοιχα με πριν ανανεώνονται οι τιμές των συναπτικών βαρών των νευρώνων των κρυφών επιπέδων.

Η διαδικασία εκπαίδευσης του δικτύου γίνεται σε κύκλους εκπαίδευσης, όπου σε κάθε κύκλο πραγματοποιούνται ακριβώς οι ίδιες ενέργειες. Η διαδικασία εκπαίδευσης σταματά μέχρι να ικανοποιηθεί κάποιο κριτήριο τερματισμού, το οποίο συνήθως είναι είτε η ολοκλήρωση ενός αριθμού κύκλων είτε ότι το συνολικό τετραγωνικό σφάλμα των εξόδων να μειωθεί κάτω από ένα συγκεκριμένο κατώφλι ϵ .

2.3.3 Overfitting-Υπερμοντελοποίηση

Υπερμοντελοποίηση (overfitting). Το φαινόμενο της υπερμοντελοποίησης (overfitting) αποτελεί μία από τις μεγαλύτερες προκλήσεις στην εκπαίδευση ενός μοντέλου μηχανικής μάθησης. Συμβαίνει όταν το μοντέλο μας, είτε λόγω της αρχιτεκτονικής του, είτε ελλείψη αρκετού πλήθους προς εκπαίδευση δεδομένων, αντί να μαθαίνει χρήσιμα χαρακτηριστικά και σχέσεις μεταξύ των δεδομένων εκπαίδευσης, μαθαίνει να “αποστηθίζει” τυφλά τα δεδομένα αυτά. Το overfitting είναι λογικό να συμβαίνει, αφού το gradient descent, αν εντοπίσει ότι στο μοντέλο υπάρχουν παράμετροι που δε χρησιμοποιούνται με αποδοτικό τρόπο καθώς δεν υπάρχουν αρκετά δεδομένα για να ανακαλυφθούν καινούρια μοτίβα, θα το οδηγήσει να μάθει “απ’έξω” τα δεδομένα εκπαίδευσής του, αφού έτσι ελαχιστοποιείται η απώλειά του.

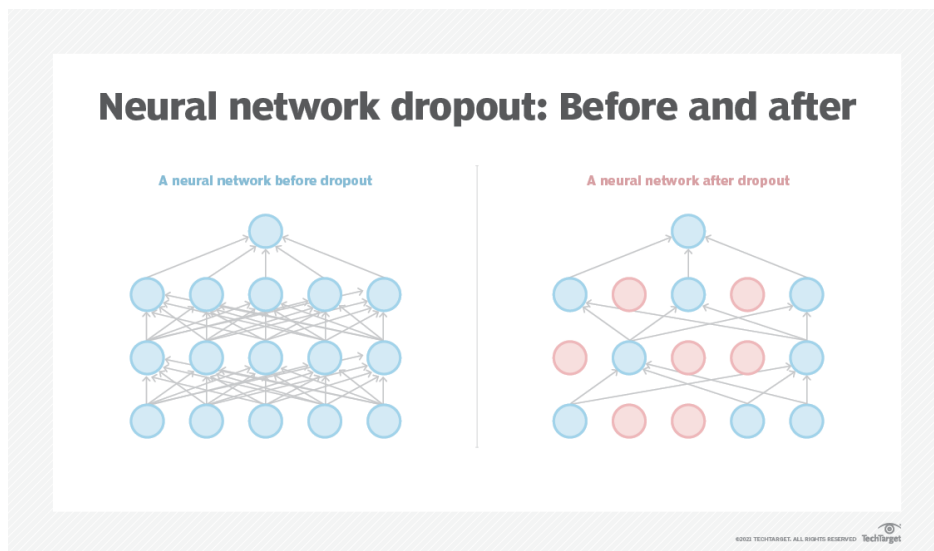
Για να εντοπίσουμε αν το μοντέλο μας πάσχει από υπερμοντελοποίηση, θέτουμε ένα ποσοστό των δεδομένων μας, συνήθως κοντά στο 20 τις εκατό, ως το Validation set (σετ επιβεβαίωσης), το οποίο παραμένει κρυφό από το δίκτυό μας και μόνο στο τέλος κάθε εποχής το μοντέλο μας κάνει προβλέψεις πάνω στο σύνολο αυτό. Overfitting υπάρχει όταν η μέση απώλεια επιβεβαίωσης (validation loss) είναι αισθητά μεγαλύτερη από την απώλεια εκπαίδευσης (training loss), ή παρόμοια όταν η επίδοση επιβεβαίωσης του μοντέλου υπό κάποια μετρική (accuracy, precision κ.λ.π.) είναι αισθητά χειρότερη από την επίδοση εκπαίδευσης. Αυτό σημαίνει ότι ενώ το δίκτυό μας μπορεί να αποδώσει καλά στα δεδομένα πάνω στα οποία έχει εκπαιδευτεί, όταν το τροφοδοτούμε με καινούρια δεδομένα δεν μπορεί να γενικεύσει πάνω σε αυτά τη γνώση που έχει αποκτήσει.

Οι τεχνικές που έχουν αναπτυχθεί για την αντιμετώπιση του overfitting εντάσσονται σε δύο γενικές κατηγορίες:

- Στην πρώτη ανήκουν αυτές που προσπαθούν να περιορίσουν την ισχύ του εκάστοτε μοντέλου, με μία προφανή ιδέα να είναι το σταμάτημα της εκπαίδευσης όταν παρατηρηθεί άυξηση της διαφοράς των δύο απωλειών ή σταθεροποίηση του validation loss (early stopping). Άλλες γνωστές τεχνικές είναι η $L1$ και $L2$ ομαλοποιήσεις (regularization), που εισάγουν στη συνάρτηση απώλειας επιπλέον όρους, και συγκεκριμένα τις $L1$ και $L2$ νόρμες των παραμέτρων του

μοντέλου. Έτσι, στο μοντέλο δίνονται λιγότεροι βαθμοί ελευθερίας, και άρα είναι δυσκολότερο να υπερπροσαρμοστεί στα δεδομένα του. Επίσης στην κατηγορία αυτή εντάσσεται η τεχνική της εισαγωγής επιπέδων Dropout και Batchnormalization στην αρχιτεκτονική του μοντέλου μας.

- Dropout. Το επίπεδο Dropout σε κάθε φάση της διαδικασίας εκπαίδευσης αγνοεί κάποιους τυχαίους κόμβους(νευρώνες) και τους βγάζει εκτός δικτύου με πιθανότητα $1-p$, ή τους συμπεριλαμβάνει στο δίκτυο με πιθανότητα p , με αποτέλεσμα να προκύπτει ένα δίκτυο με αρκετά λιγότερους νευρώνες. Με αυτόν τον τρόπο χάνουμε πληροφορία και ο αλγόριθμός μας αργεί αισθητά περισσότερο να φτάσει σε σύγκλιση, αλλά το τυχαία "σβήσιμο" κάποιων νευρώνων κάνει το δίκτυο μας να μην μπορεί απλά να μάθει "απ'έξω" τη συσχέτιση μεταξύ νευρώνων και έτσι να μειώνεται πολύ η πιθανότητα υπερμοντελοποίησης.



Σχήμα 2.7: Before/After Dropout Layers

- BatchNormalization. Όταν αναφερόμαστε σε κανονικοποίηση συνελκτικών νευρωνικών δικτύων φαίνεται η εισαγωγή BtatchNormalization επιπέδων στο μοντέλο μας να αποδίδει αισθητά καλύτερα από ότι αυτή Dropout επιπέδων. Ένα BatchNormalization επίπεδο, είναι ένα επίπεδο που κανονικοποιεί τις εισόδους του, και διατηρεί την μέση τιμή της εξόδου κοντά στο 0 και την στάνταρ διακύμανση της κοντά στο 1. [Ioffe and Szegedy \(2015\)](#)
- Παράλληλα, υπάρχουν τεχνικές που προσπαθούν να δώσουν στο μοντέλο την εντύπωση ότι εκπαιδύεται με περισσότερα δεδομένα από αυτά που πράγματι έχουμε στη διάθεσή μας, και αποτελούν τη δεύτερη κατηγορία, τον εμπλουτισμό δεδομένων (data augmentation). Μερικές από τις πιο γνωστές τεχνικές είναι η τυχαία αλλαγή των χρωμάτων, η περιστρο-φή της εικόνας κατά μερικές μοίρες (rotation), ο "καθρεφτισμός" της εικόνας κατά τον οριζόντιο ή τον κάθετο άξονα (horizontal και vertical flipping), η εστίαση (zooming) κατά ένα μικρό ποσοστό, η οριζόντια ή κάθετη μετακίνηση της εικόνας κατά μερικά pixels και η αλλαγή της αντίθεσης. Παρά τα τεράστια πιθανά τους οφέλη, οι τεχνικές αυτές έχουν έναν πολύ σημαντικό

περιορισμό: Το εύρος των τιμών των διάφορων τεχνικών, αλλά πολλές φορές ακόμα και η ίδια η δυνατότητα χρήσης τους καθορίζεται από τη φύση των δεδομένων εκπαίδευσης. Για παράδειγμα, αν ένα συνελκτικό δίκτυο εκπαιδεύεται να αναγνωρίζει ανθρώπινα πρόσωπα, το κάθετο flipping στην πραγματικότητα είναι πολύ πιθανό να χειροτερέψει την επίδοση, αφού το μοντέλο θα έχει εκπαιδευτεί πάνω σε δεδομένα που δεν έχουν φυσική σημασία, και άρα θα έχει σπαταλήσει πολύτιμο “χώρο” για την αποθήκευσή των μοτίβων τους.

2.4 Βιβλιογραφική Επισκόπηση

Υπάρχει αρκετό υλικό στην σχετική επιστημονική βιβλιογραφία, με ερευνητές που εξετάζουν την ανακατασκευή 3D σχημάτων, με χρήση διαφόρων μορφών νευρωνικών δικτύων. Εδώ [Tan et al. \(2017\)](#), εξετάζεται η χρήση ενός συγκεκριμένου είδους autoencoder, του variational autoencoder με σκοπό την ανακατασκευή πλεγμάτων με ακανόνιστη τοπολογία, αλλά και τη δημιουργία νέων.

Η ανακατασκευή τρισδιάστατων αντικειμένων είναι ένα σύνθετο επιστημονικό πρόβλημα που συναντάται σε πολλούς τεχνολογικούς τομείς, όπως τα computer graphics, το computer animation, η όραση υπολογιστών, αλλά και η εμφάνιση ιατρικών εικόνων. Πρόσφατα, δίνεται όλο και περισσότερη προσοχή στο συγκεκριμένο επιστημονικό πεδίο, κυρίως λόγω της επιτυχίας του 3d printing [Walker et al. \(2018\)](#) και των τρισδιάστατων ιατρικών εικόνων [Heinrich et al. \(2018\)](#). Επίσης η εμφάνιση μεγάλης κλίμακας βάσεων δεδομένων με 3d μοντέλα όπως το Shapenet, σίγουρα έχει βοηθήσει σημαντικά. Στην βιβλιογραφία συνήθως ζητείται η ανακατασκευή των τρισδιάστατων επιφανειών, από μία ή περισσότερες εικόνες δύο διαστάσεων. Αυτό το πρόβλημα ερευνάται εδώ και δεκαετίες. Οι πρώτες μέθοδοι που δοκίμασαν να το επιλύσουν, εστίασαν στη γεωμετρική προσέγγιση και την μαθηματική κατανόηση του προβλήματος της προβολής των 3 διαστάσεων σε 2, έτσι ώστε να προσεγγίσουν αλγοριθμικές λύσεις που να επιλύουν το ανάστροφο. Οι περισσότερες από αυτές τις μεθόδους [Hartley and Zisserman \(2004\)](#), [Laurentini \(1994\)](#), αν και σε μερικές περιπτώσεις κατάφεραν να επιτύχουν αξιοπρεπή αποτελέσματα, απαιτούσαν πολλές, πρεσεκτικά ζυγισμένες λήψεις, από συγκεκριμένες γωνίες, του προς ανακατασκευή αντικειμένου, πράγμα που σε πολλές περιπτώσεις δεν καθίσταται εφικτό. Αντιθέτως οι άνθρωποι φαίνεται να είναι ενστικτωδώς πολύ καλοί στη διαδικασία της ανακατασκευής τρισδιάστατων σχημάτων, ακόμα και στο να μαντέψουν πως θα μοιάζει από πλευρές που δεν έχουν οπτική επαφή. Αυτό το καταφέρνουμε κυρίως χρησιμοποιώντας τη μνήμη μας, έχοντας δει αντίστοιχες εικόνες στο παρελθόν.

Η δεύτερη γενιά μεθόδων τρισδιάστατης ανακατασκευής, χρησιμοποίησε τεχνικές βαθιάς μάθησης, κάτι που κατέστη εφικτό και χάρη στον ολοένα αυξανόμενο όγκο δεδομένων, έτοιμων για αξιοποίηση ως προς την εκπαίδευση των δικτύων, και έχει ήδη πετύχει εξαιρετικά και πολύ φιλόδοξα αποτελέσματα. Υπάρχουν πολλές διαφορετικές προσεγγίσεις που μπορεί να ακολουθήσει κανείς, τόσο σε ό,τι αφορά την αρχιτεκτονική του δικτύου, αλλά και στους ίδιους τους σκοπούς της ανακατασκευής (Volumetric representations, Surface-based representations, Intermediation). Αυτό όμως που φαίνεται να παραμένει σταθερό, παρά τις επιμέρους διαφορετικές στοχοθετήσεις, είναι η ραχοκοκκαλιά του δικτύου, η οποία αποτελείται από έναν κωδικοποιητή h , ακολουθούμενο από έναν αποκωδικοποιητή g , i.e., $f=g \circ h$. Ο κωδικοποιητής h μεταφέρει τα δεδομένα εισόδου στον χώρο λανθάνουσών μεταβλητών, όπως αναλύθηκε και προηγουμένως, ο κωδικοποιητής αποτελείται από μια αλληλουχία συνελκτικών (convolution layer) επιπέδων, και επιπέδων ομαδοποίησης (pooling layer), ακολουθούμενα από πλήρως συνδεδεμένα επίπεδα νευρώνων. Ο αποκω-

δικοποιητής χρησιμοποιεί είτε πλήρως συνδεδεμένα επίπεδα νευρώνων, είτε ένα δίκτυο αποσυγκέντρωσης(deconvolution network), δηλαδή μία σειρά από συνελκτικά(convolution) και υπερδειγματοληπτικά(upsampling) επίπεδα. Πάνω στην αρχιτεκτονική που μόλις περιγράφηκε έχουν γίνει πολλές επεκτάσεις δημιουργώντας κάποια από τα πλέον γνωστά δίκτυα σήμερα(ResNet, Convolutional Neural Networks (CNN), Generative Adversarial Networks(GAN), Variational Auto-Encoders). Ο χώρος λανθανουσών μεταβλητών, μετά την κωδικοποίηση, μπορεί να είναι είτε συνεχής είτε διακριτός. Οι Wu et al. στην εργασία τους Wu et al. (2015) εισήγαγαν το 3D ShapeNet, ένα δίκτυο κωδικοποίησης, που κωδικοποιεί ένα τρισδιάστατο σχήμα, το οποίο αναπαρίσταται ως ένα πλέγμα μεγέθους 303 σε μία λανθάνουσα αναπαράσταση μεγέθους 4000×1. Ο πυρήνας του δικτύου αυτού αποτελείται από 3 συνελκτικά επίπεδα(convolutional layers) κάθε ένα από τα οποία χρησιμοποιεί τρισδιάστατα φίλτρα, ακολουθούμενα από 3 επίπεδα πλήρως συνδεδεμένων νευρώνων. Αυτό το πρότυπο αρχιτεκτονικής έχει χρησιμοποιηθεί σε εργασίες κατηγοριοποίησης(classification) και ανάκτησης Wu et al. (2015), αλλά και σε ανακατασκευή τρισδιάστατων σχημάτων Wu et al. (2015). Για δυδιάστατες κωδικοποιήσεις σε λανθάνων χώρο χρησιμοποιείται η ίδια αρχιτεκτονική με το 3D ShapeNet Wu et al. (2015), μόνο που χρησιμοποιούνται συνελκτικά επίπεδα δύο διαστάσεων Yan et al. (2016), Grant et al. (2016), Wu et al. (2017), Choy et al. (2016), Tulsiani et al. (2017), Sun et al. (2018), Wiles and Zisserman (2017) , Tulsiani et al. (2018) . Πολλές μελέτες διαφέρουν μεταξύ τους στον τύπο και τον αριθμό των επιπέδων που χρησιμοποιούν. Για παράδειγμα, οι Yan et al. Yan et al. (2016) χρησιμοποιούν 3 συνελκτικά επίπεδα(convolutional layers) με 64,128 και 256 κανάλια αντίστοιχα, 3 πλήρως συνδεδεμένα επίπεδα νευρώνων με 1024, 1024, και 512 νευρώνες αντίστοιχα. Οι Wiles και Zisserman Wiles and Zisserman (2017)[10] χρησιμοποιούν 6 συνελκτικά επίπεδα (convolutional layers) από 3, 64, 128, 256, 128, 160 κανάλια αντίστοιχα. Σε άλλες εργασίες βλέπουμε να προσθέτονται επίπεδα συγκέντρωσης(pooling layers)Choy et al. (2016), Johnston et al. (2017), και η Leaky Rectified Linear Unit (ReLU) ως συνάρτηση ενεργοποίησης Choy et al. (2016), Johnston et al. (2017), Peng et al. (2021). Για παράδειγμα, οι Wiles and Zisserman Wiles and Zisserman (2017) χρησιμοποιούν επίπεδα μέγιστης συγκέντρωσης(maxpooling layers) ανάμεσα σε κάθε ζευγάρι συνελκτικών επιπέδων, εκτός από το πρώτο επίπεδο και πριν από το τελευταίο. Η ReLU ως συνάρτηση ενεργοποίησης βελτιώνει την εκπαίδευση του αλγορίθμου, καθώς η παράγωγος κατά την οπισθοδιάδοση(backpropagation) του λάθους δεν μηδενίζεται ποτέ.

Σε ότι αφορά τον συνεχή λανθάνων χώρο, οι Variational Autoencoders(VAE) είναι το πιο χαρακτηριστικό παράδειγμα. Χάρη στην ιδιότητά τους αυτή, έχουν χρησιμοποιηθεί σε ογκομετρική[17] Wu et al. (2016), Liu et al. (2018) , σε βασισμένη στο βάθος Arsalan Soltani et al. (2017)[19], βασισμένη στην επιφάνεια Henderson and Ferrari (2018)[20], και σε βασισμένη σε σημεία Mandikal et al. (2019)[21], Gadelha et al. (2018)[22] τρισδιάστατη ανακατασκευή. Στην εργασία των Wu et al. (2016), για παράδειγμα, ο κωδικοποιητής δέχεται μία 256×256 RGB εικόνα και παράγει δύο 200-διάστατα διανύσματα που αναπαριστούν αντίστοιχα τη μέση και τη κανονική απόκλιση μιας Γκαουσιανής κατανομής στον 200διάστατο χώρο. Συγκριτικά με τους κανονικούς κωδικοποιητές, οι 3D-VAE μπορούν να χρησιμοποιηθούν ώστε δειγματοληπώντας τυχαία από τον λανθάνων χώρο, να παράξουν παραλλαγές της εισόδου, και να ανακατασκευάσουν πολλαπλά πιθανά τρισδιάστατα σχήματα, δοθείσης της εικόνας εισόδου Mandikal et al. (2019), Gadelha et al. (2018). Επίσης έχει καλή ικανότητα γενίκευσης σε εικόνες που δεν έχει δει ποτέ κατά την διαδικασία της εκπαίδευσης. Οι Wang et al. (2017) αντιμετωπίζουν την προς ανακατασκευή τρισδιάστατη εικόνα, ως

ένα σύνολο απο εικόνες δύο διαστάσεων, τις οποίες ανακατασκευάζουν μία προς μία με μεγάλη ακρίβεια. Ο κίνδυνος που εμπεριέχει αυτή η μέθοδος, είναι η λάθος τοποθέτηση των 2D εικόνων που αποτελούν την 3D εικόνα. Για να μην υπάρχουν ασυνέχειες ανάμεσα στις 2D εικόνες, οι [Wang et al. \(2017\)](#) χρησιμοποίησαν ένα μακροπρόθεσμο επαναλαμβανόμενο νευρωνικό δίκτυο(Long term Recurrent Convolutional Network (LRCN)) [Donahue et al. \(2016\)](#) αποτελούμενο από έναν 3D κωδικοποιητή, μία LSTM μονάδα, και έναν 2D αποκωδικοποιητή. Σε κάθε επανάληψη ο κωδικοποιητής επεξεργάζεται 5 συνεχόμενες 2D εικόνες, παράγοντας μία στάνταρ μεγέθους αναπαράσταση διανύσματος, την οποία και προωθεί ως είσοδο στην LSTM. Η έξοδος της LSTM περνάει από τον 2D αποκωδικοποιητή, ο οποίος εν τέλη παράγει μίας υψηλής ανάλυσης 2D εικόνα. Η συγκέντρωση όλων των 2D εικόνων, μας δίνουν εν τέλη, μία υψηλής ανάλυσης 3D εικόνα. Αυτής της μορφής η υλοποίηση είναι πολύ κοντά στην διαδικασία που ακολουθήσαμε και εμείς στην παρούσα διπλωματική.

Κεφάλαιο 3

Σχεδιασμός, Επιλογή παραμέτρων και Υλοποίηση

Στο παρών κεφάλαιο αναλύονται οι σχεδιαστικές επιλογές καθώς και τα βήματα υλοποίησης και επιλογής των παραμέτρων του δικτύου μας. Ο σχεδιασμός κι η υλοποίηση της διπλωματικής εργασίας εξελίχθηκε σε 5 βασικά στάδια: α) την προεπεξεργασία των δεδομένων, β) τη σχεδίαση του προγράμματος, γ) την επιλογή των υπερπαραμέτρων του δικτύου δ) τη δημιουργία ενός περιβάλλοντος για τη διεξαγωγή των πειραμάτων και ε) την ανάπτυξη και υλοποίηση του προγράμματος.

3.1 Απόκτηση δεδομένων

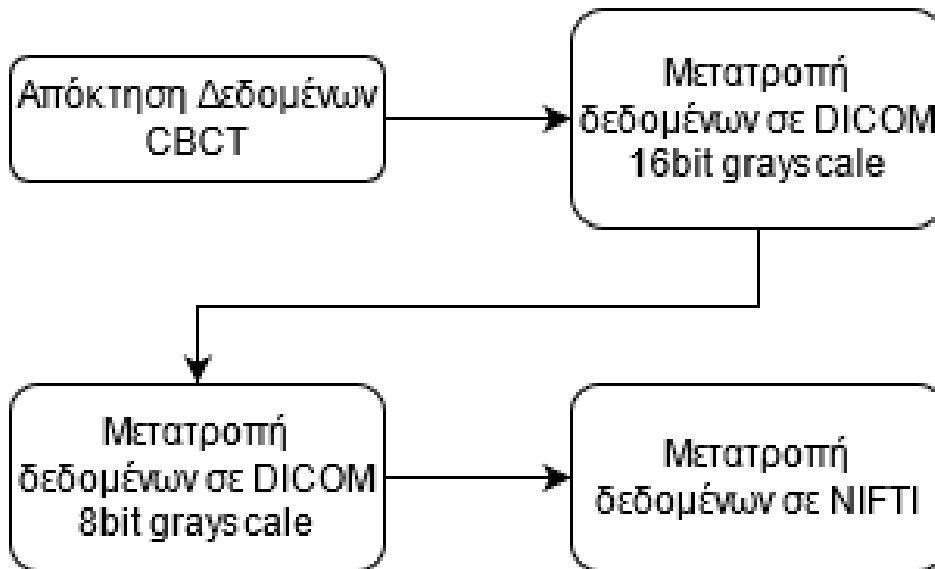
Τα δεδομένα που χρησιμοποιήσαμε στην παρούσα διπλωματική και με τα οποία εκπαιδεύσαμε το νευρωνικό δίκτυο, προέρχονται απο περισσότερους από 30 ασθενείς οι οποίοι χρειάζονταν μεταμόσχευση καθώς παρουσίαζαν οδοντοφατνιακά ελαττώματα. Παρήχθη 1 σετ δεδομένων για κάθε ασθενή. Οι χειρουργικές επεμβάσεις πραγματοποιήθηκαν και τα σετ δεδομένων παρήχθησαν από το τμήμα Περιοδοντολογίας του Semmelweis University, Budapest, Hungary, του οποίου η τοπική και ιδρυματική επιτροπή επιστήμης και ερευνητικής ηθικής (reference number: 77/2011), ενέκρινε την χρήση τους. Κάθε σετ αποτελείται από δύο όγκους CBCT εικόνων, έναν για την προεγχειρητική και έναν για τη μετεγχειρητική αναπαράσταση κάθε ασθενούς. Ο συνδυασμός των CT τομών σε κάθε περίπτωση αποτελεί μια πλήρη τρισδιάστατη αναπαράσταση κάθε ασθενή, δίνοντας έμφαση στις οδοντοφατνιακές αλλαγές, πριν, και μετά το πέρας 9 έως 12 μηνών μετά τη χειρουργική επέμβαση. Όλες οι σαρώσεις CBCT λήφθηκαν από τον i-CAT (Imaging Sciences International, Hatfield, PA) εξοπλισμό ακτίνων X. Τα σύνολα δεδομένων λήφθηκαν κατα τη διάρκεια σύνθετων επεμβάσεων μεταμόσχευσης και προσθετικής στους ασθενείς, λαμβάνοντας υπόψη ότι η δοσολογία ακτινοβολίας μιας σάρωσης i-CAT δεν είναι σημαντικά υψηλότερη σε σχέση με αυτή ενός CBCT scan μικρότερου όγκου, συνεπώς το σύστημα i-CAT προτιμήθηκε αφού σε όλες τις περιπτώσεις υπήρχαν επιπλέον ενδοδοντικές θεραπείες που έπρεπε να αξιολογηθούν.

Όλα τα σετ δεδομένων λήφθηκαν, χρησιμοποιώντας ένα κοινό σετ παραμέτρων για τη λειτουργία του i-CAT, όπως υψηλές συχνότητες, 3–8mA (pulse mode), ο ανιχνευτής εικόνας ήταν μία πλάκα από σιλικόνη διαστάσεων 20 x 25 cm, και η λήψη της εικόνας πραγματοποιήθηκε ώντας σε καθιστή θέση, από μία μόνο περιστροφή 360 μοιρών, με χρονικό διάστημα σάρωσης σταθερά τα 40 δευτερόλεπτα.

Όλα τα διαθέσιμα σετ δεδομένων μετατράπηκαν σε μορφή DICOM(digital imaging and communications in medicine) ως σειρά 16-bit εικόνων γκριζας κλίμακας. Το μέγεθος κάθε εικόνας ήταν 400 x 400 pixel, με μέγεθος 0,4 x 0,4 mm ανά pixel. Η απόσταση μεταξύ των τομών στον z άξονα,

διατηρήθηκε σταθερή για όλα τα σύνολα επίσης στα 0,4 mm. Ο αριθμός των τομών (CT slices) σε κάθε σετ ήταν 322–327, ανάλογα με τον ασθενή. Για την απλοποίηση της επεξεργασίας των δεδομένων, όλα τα σύνολα DICOM μετατράπηκαν σε 8-bit RAW gray scale δεδομένα, χρησιμοποιώντας το λογισμικό ανοιχτού κώδικα XMedCon. [Economopoulos TL \(2012\)](#)

Παρακάτω ακολουθεί σχεδιάγραμμα στο οποίο βλέπουμε τα βήματα που ακολουθήθηκαν από την απόκτηση των δεδομένων μέχρι να τα φέρουμε στην επιθυμητή για εμάς μορφή, ώστε να μας είναι ευκολότερο να τα επεξεργαστούμε.



Σχήμα 3.1: Σχεδιαγραμματική απεικόνιση των βημάτων μετατροπής των δεδομένων

3.2 Δεδομένα

Η επιλογή των κατάλληλων δεδομένων είναι από τους σημαντικότερους παράγοντες που καθορίζουν την επιτυχία ενός προγράμματος μηχανικής μάθησης. Η εύρεση όμως των κατάλληλων δεδομένων αποτελεί ένα πολύ μεγάλο πρόβλημα στο χώρο της μηχανικής μάθησης και ειδικότερα της βαθιάς μάθησης. Σημαντικό για το πρόγραμμά μας είναι και το πλήθος των δεδομένων μιας και όσο μεγαλύτερο είναι αυτό, τόσο πιο έμπιστο γίνεται το μοντέλο μας, και τόσο καλύτερα μπορούμε να γενικεύσουμε τα αποτελέσματά μας σε νέα δεδομένα, χωρίς βέβαια αυτό να αποτελεί το μόνο παράγοντα. Φυσικά κανένα από τα παραπάνω χαρακτηριστικά δεν αποτελεί γνώμονα από μόνο του, αλλά χρειάζεται ένας συνδυασμός τους για να θεωρηθούν τα δεδομένα επαρκή για την ανάπτυξη ενός αξιόπιστου μοντέλου μηχανικής μάθησης. Ο τρόπος απόκτησης των δεδομένων που χρησιμοποιήσαμε σχολιάστηκε αναλυτικά στο κεφάλαιο 2.

3.2.1 Προεπεξεργασία Δεδομένων

Όπως αναφέρθηκε και σε προηγούμενη ενότητα, εμείς κληθήκαμε να δουλέψουμε με CT scans, τα οποία λάβαμε σε μορφή dicom. Ένα από τα πρώτα βήματα της προεπεξεργασίας των δεδομένων, ήταν να μετατρέψουμε τα δεδομένα dicom σε μορφή nifti (nii.gz). Ο τύπος δεδομένων nifti είναι αρκετά απλούστερος στην επεξεργασία από τον dicom και καταναλώνει αισθητά λιγότερη μνήμη. Ένας ακόμη λόγος που κάναμε αυτή την μετατροπή, ήταν πως πολλά από τα metadata που

κρατώνται στα dicom αρχεία δεν μας χρειάζονταν στην παρούσα διπλωματική. Αξίζει να τονιστεί πως το γεγονός, πως το μέγεθος κάθε εικόνας ήταν σταθερά στα 400 x 400 pixel, με μέγεθος 0,4 x 0,4 mm ανά pixel και η απόσταση μεταξύ των τομών στον z άξονα, διατηρήθηκε σταθερή για όλα τα σύνολα στα 0,4 mm, αλλά και πως ο αριθμός των τομών (CT slices) σε κάθε σετ ήταν περίπου σταθερός (322–327 ανάλογα με τον ασθενή), μας βοήθησε πολύ να απαλλαγούμε από περιττά metadata και από επιπλέον βήματα προεπεξεργασίας, αλλά ακόμη περισσότερο βοήθησε την διαδικασία εκπαίδευσης του δικτύου μας, όπως θα δουμε αργότερα.

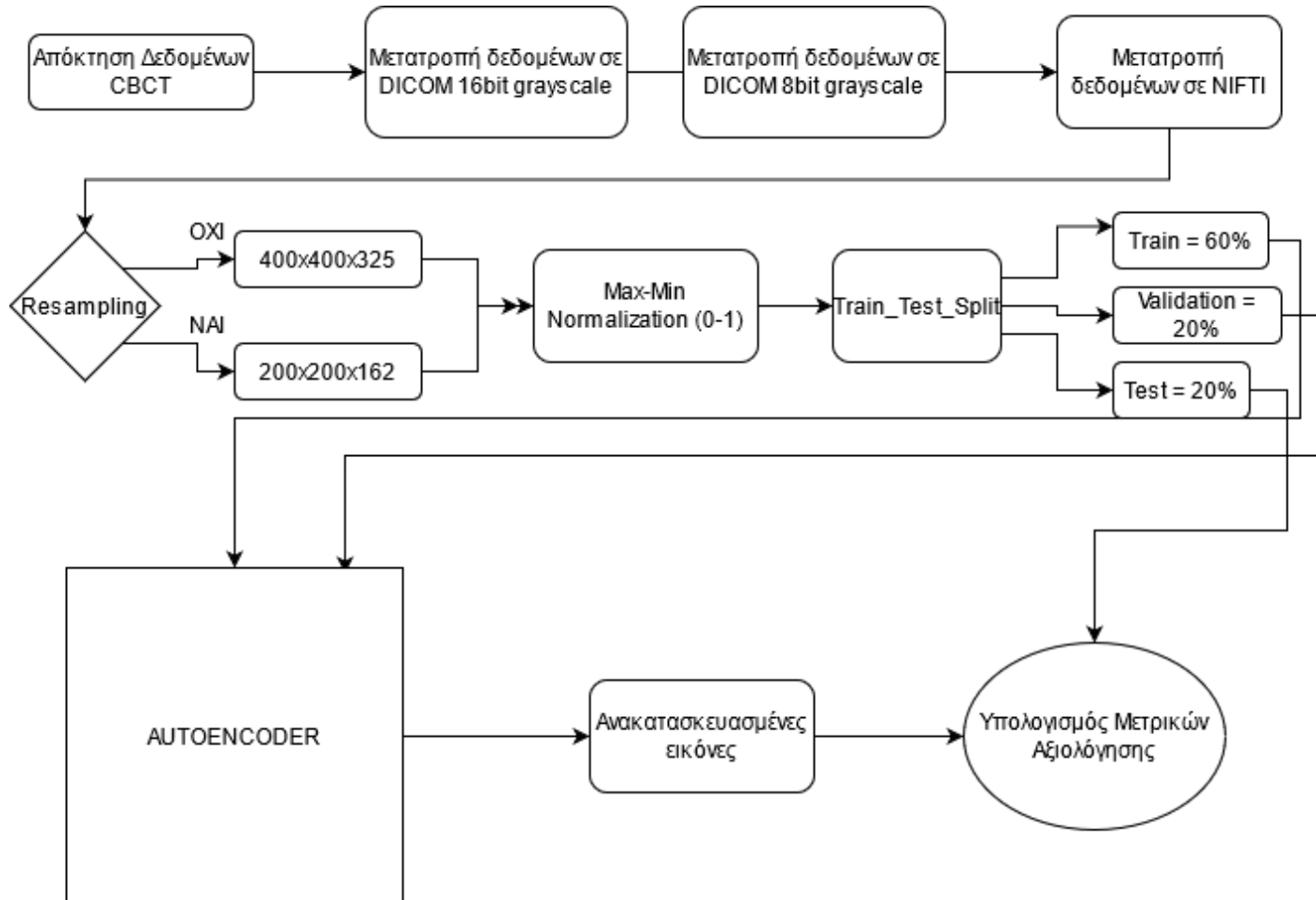
Η αρχική μας υλοποίηση που έλαβε χώρα στο υπολογιστικό περιβάλλον του Google Colab, απαιτούσε το resampling των δεδομένων μας λόγω έλλειψης μνήμης όπως θα δούμε παρακάτω, ενώ για τα πειράματα που τρέξαμε τοπικά σε προσωπικό υπολογιστή, χρησιμοποιήσαμε τα full-scale δεδομένα χωρίς να προβούμε σε κανένα resampling. Τα υπόλοιπα βήματα προεπεξεργασίας ισχύουν κανονικά και για τις δύο υλοποιήσεις. Το πρώτο βήμα προεπεξεργασίας που εφαρμόσαμε λοιπόν ήταν το resampling των δεδομένων μας. Συγκεκριμένα μετατρέψαμε κάθε τρισδιάστατο CT scan, αρχικού μεγέθους 400x400x322-327, σε μέγεθος 200x200x162, καταλαμβάνοντας σημαντικά μικρότερο μέρος μνήμης, καθιστώντας την εκπαίδευσή του νευρωνικού δικτύου σημαντικά γρηγορότερη κι επιτρέποντας τη χρήση υπολογιστικών περιβαλλόντων με λιγότερη μνήμη. Στη συνέχεια χρησιμοποιώντας max-min normalization, κανονικοποιήσαμε τις τιμές έντασης των scans στο διάστημα 0-1, εξοικονομώντας έτσι μνήμη και υπολογιστική ισχύ. Επίσης κάναμε σίγουρο πως οι διαστάσεις του κάθε slice (εδώ 200x200) αποτελούνται από ζυγούς αριθμούς, έτσι ώστε να μας είναι πιο εύκολο στη συνέχεια το downsampling περνώντας τες από το μοντέλο.

3.2.2 Διαχωρισμός Δεδομένων

Τα σύνολα δεδομένων αποτελούνται από φακέλους στους οποίους είναι χωρισμένες οι εικόνες σε τρισδιάστατα scan. Κάθε φάκελος αποτελείται από τις dicom εικόνες που στο σύνολο τους συνιστούν το τρισδιάστατο scan. Αυτό που χρειάζεται να κάνουμε εμείς όμως είναι να χωρίσουμε τα δεδομένα σε Train, Validation και Test. Χρειαζόμαστε ένα σύνολο δεδομένων πάνω στο οποίο θα εκπαιδεύσουμε το CNN (Train Set), ένα διαφορετικό σύνολο δεδομένων η χρήση του οποίου βοηθάει στην αξιολόγηση του νευρωνικού μας δικτύου κατά την εκπαίδευση και την προσαρμογή του ανάλογα με τις επιδόσεις στο συγκεκριμένο σύνολο (Validation Set) και τέλος ένα σύνολο δεδομένων το οποίο δεν έχει χρησιμοποιηθεί σε καμία φάση της εκπαίδευσης, πάνω στο οποίο θα γίνει η αξιολόγηση του μοντέλου (Test Set). Είναι πολύ σημαντικό τα τρία αυτά σύνολα να μην έχουν καμία επικάλυψη. Υλοποιήσαμε αυτόν τον διαχωρισμό δεδομένων χρησιμοποιώντας δύο διαδοχικές φορές την συνάρτηση *train – test – split* της βιβλιοθήκης *sklearn*. Την πρώτη φορά της περάσαμε ως παραμέτρους 2 φορές τον φάκελο με το πλήρες dataset εικόνων που θέλουμε να χωρίσουμε, μία για την είσοδο και μία για την επιθυμητή έξοδο, που στην περίπτωσή μας ταυτίζονται, την μεταβλητή *random – state* να ισούται με έναν ακέραιο αριθμό, ώστε να χωρίζεται με τον ίδιο τρόπο κάθε φορά το dataset και ο διαχωρισμός να είναι αναπαράξιμος σε κάθε μας δοκιμή και την μεταβλητή *test – size = 0.2*, που δηλώνει πως θα κρατήσουμε το 20 εκατό του dataset για το validation set μας, και από το υπόλοιπο 80 τις εκατό θα προκύψουν το train και το test set. Στην δεύτερη κλήση της *train – test – split* περάσαμε ως παραμέτρους 2 φορές το 80 τις εκατό του dataset που μας επέστρεψε η προηγούμενη κλήση, πάλι καθώς η είσοδος μας ταυτίζεται με την επιθυμητή έξοδο, την μεταβλητή *random state* ίση με έναν ακέραιο αριθμό όπως και προηγουμένως, και την

μεταβλητή $test-size = 0.25$. Έτσι προκύπτει το train set ίσο με $0.8 * 0.75 = 0.6$ του αρχικού dataset, και το test set ίσο με $0.8 * 0.25 = 0.2$ του αρχικού dataset.

Ακολουθεί σχεδιάγραμμα, όπου αναπαρίστανται τα κύρια βήματα που ακολουθήθηκαν σε όλη την διπλωματική, από την απόκτηση των δεδομένων έως και την μέτρηση της ποιότητας ανακατασκευής των εικόνων του δικτύου μας.



Σχήμα 3.2: Σχεδιαγραμματική απεικόνιση των βημάτων που ακολουθήθηκαν στο σύνολο της διπλωματικής

3.3 Περιβάλλον Διεξαγωγής Πειραμάτων

Για την ανάπτυξη των προγραμμάτων χρησιμοποιήσαμε το Keras [Chollet et al. \(2015\)](#) το οποίο χρησιμοποιεί το Tensorflow [Abadi et al. \(2015\)](#) ως backend για να τρέξει. Το Keras είναι μια βιβλιοθήκη υψηλού επιπέδου και χρειάζεται κάποια βιβλιοθήκη λογισμικού με το κατάλληλο API για να τρέξει "πάνω" σε αυτή. Το Tensorflow είναι μια open-source βιβλιοθήκη λογισμικού η οποία χρησιμοποιείται ευρέως, ειδικά σε κομμάτια που αφορούν τη βαθιά μάθηση. Πρόκειται για δύο βιβλιοθήκες οι οποίες παρέχουν πάρα πολλά και εύχρηστα εργαλεία για την ανάπτυξη προγραμ-

μάτων σχετικά με τη βαθιά μάθηση και ειδικά για τα συνελκτικά νευρωνικά δίκτυα, τις οποίες και χρησιμοποιήσαμε.

Το υπολογιστικό περιβάλλον όπου χρησιμοποιήσαμε για τα πρώτα πειράματα είναι το Google Colab [Bisong \(2019\)](#). Πρόκειται για ένα εύχρηστο περιβάλλον το οποίο δίνει τη δυνατότητα στο χρήστη να τρέχει jupyter notebooks online. Οι βασικοί λόγοι για τη συγκεκριμένη επιλογή ήταν οι εξής: α) Παρέχει δωρεάν τη δυνατότητα επιτάχυνσης της εκπαίδευσης με χρήση GPU (μία εκ των Nvidia K80s, T4s, P4s and P100s, 12GB GDDR5 VRAM) , χαρακτηριστικό χωρίς το οποίο θα ήταν αδύνατη η εκπαίδευση των νευρωνικών δικτύων που δημιουργήσαμε μιας και ο χρόνος εκπαίδευσης θα ήταν απαγορευτικός. β) Παρέχει τη δυνατότητα σύνδεσης με το Google Drive, καθιστώντας πολύ εύκολη αν και χρονοβόρα την εισαγωγή δεδομένων στο πρόγραμμα.

Τα πρώτα μας πειράματα έγιναν στο google Colab, αλλά ακόμη και τα 12GB RAM που αυτό μας παρέχει ήταν πολύ λίγα για να εκπαιδύσουμε το δίκτυό μας χρησιμοποιώντας τα full scale δεδομένα που είχαμε στη διάθεσή μας. Γι'αυτό προβήκαμε στο resampling των δεδομένων μας, όπως αναλύθηκε και προηγουμένως. Μια μέθοδος που εφαρμόζεται συχνά κατά την εκπαίδευση νευρωνικών δικτύων όταν το προς εκπαίδευση σύνολο δεδομένων δεν χωράει στη μνήμη RAM, όπως στην περίπτωσή μας, είναι η χρήση generators. Οι generators είναι κλάσεις της Python, που παίρνουν επαναληπτικά κομμάτια (batches) των δεδομένων από τον σκληρό δίσκο και τα φέρνουν στην μνήμη ram για να εκπαιδευτεί με βάση αυτά το δίκτυο. Η διαδικασία αυτή επαναλαμβάνεται μέχρι να εξαντληθούν τα δεδομένα και για όσες εποχές έχουμε ορίσει. Βέβαια η συνεχής αυτή επικοινωνία με τον σκληρό δίσκο του συστήματός μας, κάνει την όλη διαδικασία εκπαίδευσης πολύ πιο αργή, και ειδικά στο περιβάλλον του google colab, όπου η σύνδεση με το google Drive είναι ιδιαίτερος χρονοβόρα, η χρήση generators για την εκπαίδευση των μοντέλων μας ήταν απαγορευτική, καθώς για μεγάλο αριθμό εποχών μπορεί να χρειαζόταν από μέρες έως και βδομάδες για την εκπαίδευση του δικτύου μας.

3.3.1 Επιλογή Δεδομένων Εισόδου

Η πρώτη και βασικότερη επιλογή που είχαμε να κάνουμε στην αρχή της δημιουργίας του προγράμματος ήταν ο τύπος δεδομένων με τα οποία θα δουλέψουμε. Αυτή η επιλογή καθορίζει σε μεγάλο βαθμό τη μορφή ολόκληρου του υπόλοιπου προγράμματος μιας και χρησιμοποιούνται εντελώς διαφορετικά δίκτυα για την επεξεργασία εικόνων σε σχέση με την επεξεργασία πινάκων χαρακτηριστικών. Μπορούσαμε να κινηθούμε σε δύο άξονες: α) να χρησιμοποιήσουμε τις εικόνες αυτούσιες ως είσοδο στο πρόγραμμα, και β) να εξάγουμε χαρακτηριστικά από τις εικόνες, όπως ιστογράμματα χρωμάτων κλπ., τα οποία θα χρησιμοποιηθούν ως είσοδο στο πρόγραμμα. Επιλέξαμε να χρησιμοποιήσουμε τις εικόνες ως είσοδο, κι όχι χαρακτηριστικά αυτών για δύο βασικούς λόγους. Πρώτον αυτή η επιλογή δεν απαιτούσε επιπλέον προ-επεξεργασία στα δεδομένα και δεν απαιτούσε επιπλέον σχεδιαστικές επιλογές όπως ποια και πόσα χαρακτηριστικά να εξαχθούν από τις εικόνες, σε τι μορφή να εισαχθούν στο πρόγραμμα και πολλές άλλες, η λήψη των οποίων αποφάσεων θα ήταν και χρονοβόρα και επικίνδυνη να φέρει μεγαλύτερα περιθώρια λάθους. Ο δεύτερος και κύριο λόγος όμως για τον οποίο επιλέξαμε να εργαστούμε με τις εικόνες κι όχι με χαρακτηριστικά αυτών είναι διότι όπως έχει δείξει προηγούμενη δουλειά πάνω στο συγκεκριμένο πεδίο η χρήση των εικόνων δίνει καλύτερα αποτελέσματα σε σύγκριση με χαρακτηριστικά που εξάγονται από αυτές [Lecoutre et al. \(2017\)](#).

Η χρήση ιστογραμμάτων χρωμάτων, χωρικής συσχέτισης και άλλων προυπολογισμένων χαρακτηριστικών από εικόνες μπορεί να δώσει μια βαθύτερη ματιά στη συμπεριφορά του μοντέλου και να βοηθήσει στην κατανόηση αυτού, δεν φαίνεται όμως να δίνει και καλύτερα συνολικά αποτελέσματα.

3.4 Επιλογή Νευρωνικού Δικτύου

Μία ακόμη καθοριστική επιλογή που κληθήκαμε να πάρουμε, ήταν το είδος του Νευρωνικού δικτύου που θα φτιάχναμε ώστε να ανακατασκευάσουμε τις εικόνες του dataset. Με μία ματιά στην σχετική βιβλιογραφία [Liu and Lam \(2018\)](#) βλέπουμε πως στην συντηρητική πλειοψηφία των περιπτώσεων που έχουμε να κάνουμε με ανακατασκευή εικόνων, χρησιμοποιούνται διαφόρων ειδών Autoencoders. Και όχι τυχαία, καθώς πετυχαίνουν πολύ καλά αποτελέσματα σε εργασίες ανακατασκευής, ανάκτησης πληροφορίας και ελαχιστοποίησης θορύβου. Μετά από αυτή την επιλογή, με τη μία πάρθηκε η απόφαση να χρησιμοποιηθούν συνελκτικά νευρωνικά δίκτυα για την εκτέλεση της εργασίας καθώς είναι το ισχυρότερο εργαλείο που έχουμε στα χέρια μας για την επεξεργασία και αναγνώριση εικόνων. Άρα η χρήση των CNN αποτελούσε μονόδρομο για εμάς από τη στιγμή που αποφασίσαμε να εργαστούμε με εικόνες κι όχι με προϋπολογισμένα χαρακτηριστικά όπως αναφέρθηκε παραπάνω. Στη συνέχεια θα αναλύσουμε τις σχεδιαστικές επιλογές που κάναμε έχοντας ως δεδομένη τη χρήση των ενός convolutional autoencoder.

3.4.1 Επιλογή Αρχιτεκτονικής

Έχοντας καταλήξει στην κατασκευή ενός Συνελκτικού Αυτοκωδικοποιητή (Convolutional Autoencoder), η επόμενη σχεδιαστική επιλογή που έπρεπε να πάρουμε, αφορούσε την αρχιτεκτονική του ίδιου του συνελκτικού αυτοκωδικοποιητή. Πιο συγκεκριμένα στο αν θα τον κατασκευάζαμε χρησιμοποιώντας συνελκτικά επίπεδα 2 ή 3 διαστάσεων, με κάθε επιλογή να έχει τα πλεονεκτήματα και τα μειονεκτήματά της. Έχουν γίνει αρκετές μελέτες που εξετάζουν αν ο επιπλέον υπολογιστικός φόρτος των 3D συνελκτικών επιπέδων προσφέρει αντίστοιχα καλύτερα αποτελέσματα, το οποίο εξαρτάται και από το είδος της εργασίας που έχουν να επιτελέσουν. Εδώ [McCarthy \(2004\)](#) φαίνεται πως σε εργασίες κατηγοριοποίησης(classification) CT εικόνων τα δισδιάστατα δίκτυα αποδίδουν καλύτερα, ενώ υπάρχουν και μελέτες που δοκιμάζουν τη χρήση μεικτών 2D-3D δικτύων για βέλτιστη απόδοση. [Roy et al. \(2020\)](#). Για εργασίες ανακατασκευής από 3D σε 3D επιφάνειες, όπως στην δική μας περίπτωση δεν βρήκαμε να υπάρχει κάτι αποφασιστικά καταληκτικό στην βιβλιογραφία, αλλά ο μειωμένος υπολογιστικός φόρτος των 2D δικτύων σε συνδυασμό με τους υπολογιστικούς πόρους που είχαμε στη διάθεσή μας, έπαιξε καθοριστικό ρόλο στην απόφασή μας να εργαστούμε με δίκτυα δύο διαστάσεων.

Συνελκτικά επίπεδα τριών(3) διαστάσεων χρησιμοποιούνται συνήθως για δεδομένα video δύο διαστάσεων με τον χρόνο να αποτελεί την τρίτη διάσταση, ή για εικόνες τριών (3) διαστάσεων όπως στην περίπτωσή μας.

Η βασική διαφορά μεταξύ τους, όπως δηλώνει και το όνομά τους, ανάγεται στον αριθμό των διαστάσεων που δέχεται σαν είσοδο το κάθε επίπεδο. Όπου ένα συνελκτικό επίπεδο 2 διαστάσεων(

conv2D) θα πάρει ως είσοδο και θα επεξεργαστεί μία εικόνα 2 διαστάσεων, η οποία θα αναπαρσταθεί σε έναν πίνακα 3 διαστάσεων, η επιπλέον διάσταση αναπαριστά το κανάλι χρώματος (colour channel). Αντίστοιχα, ένα συνελκτικό επίπεδο 3 διαστάσεων θα πάρει ως είσοδο μία εικόνα 3 διαστάσεων, σε μορφή πίνακα τεσσάρων(4) διαστάσεων. Η επιπλέον αυτή διάσταση, μετά από δοκιμές που πραγματοποιήσαμε είδαμε πως προκαλεί στην καλύτερη περίπτωση τον τριπλασιασμό των παραμέτρων του δικτύου μας κάτι που αυξάνει την πιθανότητα υπερμοντελοποίησης και σε συνδυασμό με τα ήδη πολύ βαριά τρισδιάστατα(3D) δεδομένα με τα οποία θα έπρεπε να το εκπαιδύσουμε αυξάνουν κατά πολύ το υπολογιστικό κόστος και από πλευράς χρόνου και μνήμης. Αυτό σε συνδυασμό με το γεγονός πως δεν είχαμε αρκετά 3D δεδομένα, και θα έπρεπε να προβούμε σε επιπλέον βήματα προεπεξεργασίας για να μεγενθύνουμε το dataset (data augmentation) ώστε να εκπαιδύσουμε το δίκτυο, κάτι που θα σήμαινε και μεγαλύτερα περιθώρια λάθους, μας οδήγησαν στην απόφαση να δουλέψουμε με δισδιάστατα(2D) συνελκτικά επίπεδα. Χρησιμοποιώντας κάθε φέτα (slice) της τρισδιάστατης αρχικής εικόνας, ως μία ξεχωριστή είσοδο.

Παρ'ότι προβήκαμε σε πολλά πειράματα μέχρι να καταλήξουμε στην αρχιτεκτονική που κρίναμε πως αποδίδει βέλτιστα, οι αρχιτεκτονικές επιλογές σχεδιασμού του δικτύου που δοκιμάσαμε ήταν σε μεγάλο βαθμό επηρεασμένες από το αρχιτεκτονικό μοτίβο που είδαμε πως συναντάται πιο συχνά στην βιβλιογραφία για εργασίες αντίστοιχες με τη δική μας.

Βασιστήκαμε λοιπόν σε μία αλληλουχία από συνελκτικά επίπεδα δύο(2) διαστάσεων, ακολουθούμενα από επίπεδα MaxPooling, τα οποία είναι απαραίτητα για το downsampling της πληροφορίας στον κωδικοποιητή, και αντίστοιχα από συνελκτικά επίπεδα δύο διαστάσεων ακολουθούμενα από επίπεδα UpSampling στον αποκωδικοποιητή. Πειραματιστήκαμε αρκετά με τον αριθμό των συνελκτικών επιπέδων που θα χρησιμοποιούσαμε αλλά και με το πόσο θα κωδικοποιήσουμε την πληροφορία της εισόδου, δηλαδή πόσα επίπεδα MaxPooling θα εισάγουμε στο μοντέλο, ώστε η είσοδος να είναι ικανοποιητικά αναπαράξιμη στην έξοδο. Επίσης δοκιμάσαμε την χρήση Dropout αλλά και BatchNormalization επιπέδων για την αποφυγή υπερμοντελοποίησης(overfitting) του δικτύου μας, κάτι στο οποίο θα αναφερθούμε πιο αναλυτικά παρακάτω.

3.4.2 Επιλογή και Βελτιστοποίηση Υπερπαραμέτρων

Οι υπερπαραμέτροι είναι μεταβλητές που ορίζουν τη δομή του δικτύου αλλά και το πως αυτό εκπαιδεύεται. Επίσης καθορίζονται πριν αρχίσει η εκπαίδευση, δηλαδή πριν αρχίσει η διαδικασία βελτιστοποίησης των βαρών του δικτύου. Οι βασικές υπερπαραμέτροι που σχετίζονται με την δομή του δικτύου είναι :

- **Ο αριθμός των κρυφών επιπέδων**, όπου κρυφά επίπεδα αποκαλούμε τα επίπεδα ανάμεσα στο επίπεδο εισόδου και το επίπεδο εξόδου.
 - Εδώ πειραματιστήκαμε κυρίως χρησιμοποιώντας από 5 έως 10 κρυφά επίπεδα.
- **Η συνάρτηση ενεργοποίησης**. Η χρήση της είναι απαραίτητη για να εισάγει το στοιχείο της μη γραμμικότητας στο μοντέλο, κάτι που παρέχει τη δυνατότητα στα νευρωνικά δίκτυα να εκπαιδεύονται προς μη γραμμικά όρια προβλέψεων. Η πιο δημοφιλής είναι η συνάρτηση **γραμμικής ανόρθωσης(ReLU)** η οποία θα βγάλει στην έξοδο κατευθείαν την είσοδο αν αυτή είναι θετική, ενώ θα βγάλει μηδέν(0) αν είναι αρνητική. Επίσης αν θέλουμε το μοντέλο μας

να κάνει δυαδικές προβλέψεις, τότε στο επίπεδο εξόδου χρησιμοποιείται η **σιγμοειδής συνάρτηση**, ενώ αν κάνει προβλέψεις μεταξύ πολλαπλών κλάσεων, συχνά χρησιμοποιείται η **Softmax**.

– Εδώ στο επίπεδο εισόδου και στα κρυφά επίπεδα χρησιμοποιήσαμε την ReLu και στο επίπεδο εξόδου την σιγμοειδή συνάρτηση.

- **Dropout.** Το Dropout είναι μια τεχνική κανονικοποίησης ώστε να **αποφύγουμε την υπερμοντελοποίηση** [Srivastava et al. \(2014\)](#), δηλαδή να αυξάνεται η ικανότητα γενίκευσης του μοντέλου μας, και εκτός των δειγμάτων με τα οποία το είχαμε εκπαιδεύσει.

– Αν και δοκιμάσαμε υλοποιήσεις με την εισαγωγή dropout επιπέδων είδαμε πως το μοντέλο μας αποδίδει καλύτερα χωρίς αυτά.

- **Ο ρυθμός μάθησης (learning rate).** Ο ρυθμός μάθησης καθορίζει το πόσο γρήγορα ενημερώνονται οι παράμετροι του δικτύου. Χαμηλοί ρυθμοί μάθησης, καθυστερούν την εκπαίδευση αλλά συγκλίνουν ομαλά, ενώ υψηλοί ρυθμοί μάθησης επιταχύνουν την διαδικασία μάθησης αλλά ίσως και να μην συγκλίνουν.

– Η επιρροή του learning rate διαφέρει ανάλογα με τον optimizer που θα χρησιμοποιήσουμε. Είδαμε πως επιλέγοντας τον Adam ως optimizer έχουμε καλύτερα αποτελέσματα για την default τιμή του ρυθμού μάθησης, δηλαδή για $lr = 0.001$.

- **Ο αριθμός των εποχών,** δηλαδή ο αριθμός των φορών που το πλήρες dataset περνάει από το δίκτυο για εκπαίδευση.

– Δοκιμάσαμε να εκπαιδεύσουμε το δίκτυό μας από 10 μέχρι και 30 εποχές, ενώ σε πολύ λίγες περιπτώσεις φτάσαμε και τις 40. Ο πολύ μεγάλος χρόνος εκπαίδευσης που απαιτεί ένας μεγάλος αριθμός εποχών, δεν μας άφησε να κάνουμε πολλά τέτοια πειράματα, αλλά φάνηκε πως για 20 έως 30 εποχές παίρνουμε τα καλύτερα αποτελέσματα.

- **Batch size.** Με τον όρο batch size, αποκαλούμε τον αριθμό των δειγμάτων που χρησιμοποιεί το δίκτυο μας κάθε φορά γίνει η ανανέωση των παραμέτρων του.

– Και εδώ πειραματιστήκαμε αρκετά, αλλά φάνηκε πως για batch size από 4 έως 32 πήραμε τα καλύτερα αποτελέσματα, για τους διάφορους συνδυασμούς των υπόλοιπων υπερπαραμέτρων.

Δύο ακόμη υπερπαραμέτροι που παίζουν καθοριστικό ρόλο στην συμπεριφορά του δικτύου μας είναι:

- **Ο αλγόριθμος βελτιστοποίησης (optimizer) που επιλέγουμε.** Συγκεκριμένα πειραματιστήκαμε κυρίως ανάμεσα σε δύο optimizers, τον RMSprop, και τον Adam, ο οποίος αποτελεί έναν από τους πιο ευρέως χρησιμοποιούμενους optimizers στο πεδίο της αναγνώρισης εικόνας, και που όντως φάνηκε να αποδίδει αισθητά καλύτερα.

- **Η συνάρτηση οπισθολογισμού του λάθους που επιλέγουμε (loss function).** Εδώ πειραματιστήκαμε κυρίως ανάμεσα στις :

- **Binary crossentropy**
- Mean Squared Error
- Mean Absolute Error

Η υπερπαραμέτρος η οποία έδειξε να παίζει καθοριστικό ρόλο στην επιτυχία του μοντέλου ήταν ο ρυθμός μάθησης (learning rate). Έγιναν πειράματα με διάφορους ρυθμούς μάθησης μέσω των οποίων προσδιορίστηκε ένας βέλτιστος ο οποίος χρησιμοποιήθηκε και στα επόμενα πειράματα. Φυσικά ο προσδιορισμός των βέλτιστων τιμών των υπερπαραμέτρων έγινε μέσω της εκπαίδευσης των CNN και την προσαρμογή των υπερπαραμέτρων στα αποτελέσματα που λαμβάνουμε.

3.5 Έξοδος προγράμματος και μετρικές αξιολόγησης

Μετά από το βασικό σχεδιασμό του αυτοκωδικοποιητή έπρεπε να αποφασίσουμε σε ποιά μετρική θα βασίσουμε την εκπαίδευσή του αλλά και τι έξοδο θα πάρουμε από το πρόγραμμα. Η μετρική η οποία χρησιμοποιείται και στις περισσότερες εργασίες ανακατασκευής είναι το Mean absolute error(mae), επιλέξαμε να εκπαιδεύσουμε τα CNN με στόχο την ελαχιστοποίηση της συγκεκριμένης μετρικής. Στο τέλος της εκπαίδευσης του νευρωνικού εξάγεται το μέσο απόλυτο λάθος ανακατασκευής(mae) τόσο των δεδομένων εκπαίδευσης, όσο και των δεδομένων validation και test. Έτσι από κάθε μοντέλο παίρνουμε μια γραφική παράσταση που παρουσιάζει τις τιμές του mae και της loss function πάνω στα δεδομένα εκπαίδευσης και επαλήθευσης (validation) με πορεία της εκπαίδευσης. Αυτές οι γραφικές παραστάσεις παρουσιάζονται στο επόμενο κεφάλαιο.

Κεφάλαιο 4

Αξιολόγηση αποτελεσμάτων και συγκρίσεις

Στο παρών κεφάλαιο παρουσιάζονται τα αποτελέσματα των πειραμάτων και αναλύονται τόσο ποιοτικά όσο και ποσοτικά. Αρχικά γίνεται μια συγκριτική παρουσίαση μεταξύ των πειραμάτων που εκτελέστηκαν αναλύοντας την επιρροή διαφόρων παραμέτρων, ενώ στη συνέχεια γίνεται μια ποιοτική αξιολόγηση των αποτελεσμάτων η οποία συνοδεύεται από προσωπικές παρατηρήσεις.

Τα αποτελέσματα που θα παρουσιαστούν, χωρίζονται σε δύο βασικές υποενότητες, ανάλογα με τον υπολογιστικό τόπο διεξαγωγής τους, και κατά συνέπεια με το μέγεθος και είδος του σετ δεδομένων που χρησιμοποιήθηκε για την εκπαίδευση των μοντέλων.

4.1 Μετρικές αξιολόγησης

Καθώς ο στόχος της παρούσας διπλωματικής είναι η ανακατασκευή εικόνων, και καθώς η ανθρώπινη αντίληψη είναι υποκειμενική και πολύ ευκολότερο να εξαπατηθεί, χρειαζόμαστε κάποιες μετρικές αξιολόγησης που θα μας λένε με μεγαλύτερη αξιοπιστία το πόσο καλά έχει γίνει η ανακατασκευή των εικόνων σε κάθε μοντέλο.

Οι κύριες μετρικές αξιολόγησης που χρησιμοποιούνται για σύγκριση εικόνων και που θα χρησιμοποιήσουμε και εμείς στην εργασία μας είναι οι παρακάτω :

- **Peak Signal To Noise Ratio (PSNR)** Ο λόγος σήματος προς θόρυβο (SNR ή S/N) είναι ένα μέτρο που χρησιμοποιείται στην επιστήμη και τη μηχανική που συγκρίνει το επίπεδο ενός επιθυμητού σήματος με το επίπεδο του θορύβου περιβάλλοντος. Το SNR ορίζεται ως ο λόγος ισχύος σήματος προς την ισχύ θορύβου, που συχνά εκφράζεται σε ντεσιμπέλ (dB). Μια αναλογία υψηλότερη από 1:1 (μεγαλύτερη από 0 dB) υποδεικνύει περισσότερο σήμα από θόρυβο.

$$\text{SNR} = \frac{P_{\text{signal}}}{P_{\text{noise}}}$$

όπου το P είναι η μέση ισχύς. Τόσο το σήμα όσο και η ισχύς θορύβου πρέπει να μετρώνται στα ίδια ή ισοδύναμα σημεία ενός συστήματος και εντός του ίδιου εύρους ζώνης.

Επειδή πολλά σήματα έχουν πολύ μεγάλο δυναμικό εύρος τιμών, τα σήματα εκφράζονται συχνά χρησιμοποιώντας τη λογαριθμική κλίμακα ντεσιμπέλ. Με βάση τον ορισμό του ντεσιμπέλ, το σήμα και ο θόρυβος μπορεί να εκφράζονται σε ντεσιμπέλ (dB) ως

$$P_{\text{signal,dB}} = 10 \log_{10} (P_{\text{signal}}) \quad (4.1)$$

$$P_{\text{noise,dB}} = 10 \log_{10} (P_{\text{noise}}) \quad (4.2)$$

Και εφαρμόζοντας βασικές ιδιότητες λογαρίθμων προκύπτει:

$$\text{SNR}_{\text{dB}} = P_{\text{signal,dB}} - P_{\text{noise,dB}}. \quad (4.3)$$

Η μετρική που χρησιμοποιούμε εμείς το PSNR είναι ένας όρος για την αναλογία μεταξύ της μέγιστης δυνατής ισχύος ενός σήματος και της ισχύος του θορύβου που επηρεάζει την πιστότητα της ανακατασκευής του σήματος. Το PSNR επίσης μετράται σε ντεσιμπέλ(dB) και χρησιμοποιείται πολύ συχνά για τον ποσοτικό προσδιορισμό της ποιότητας ανακατασκευής σε εικόνες και βίντεο που υπόκεινται σε κάποια απώλεια συμπίεσης. Όσο μεγαλύτερο είναι το PSNR, τόσο καλύτερη είναι και η ανακατασκευή της συμπιεσθείσας εικόνας.

Το μέσο τετραγωνικό λάθος(mean-square error(MSE)) και ο λόγος μέγιστου σήματος προς θόρυβο(PSNR) χρησιμοποιούνται για να συγκρίνουν την ποιότητα συμπίεσης εικόνων. Το μέσο τετραγωνικό λάθος (MSE) αναπαριστά το συσσωρευτικό τετραγωνικό λάθος ανάμεσα στην συμπιεσμένη και την αρχική εικόνα, όσο χαμηλότερη η τιμή του MSE, τόσο χαμηλότερο και το λάθος.

Για τον υπολογισμό του PSNR, χρειάζεται πρώτα ο υπολογισμός του MSE με τον τρόπο που ακολουθεί:

Δοθείσης μίας μονοχρωματικής χωρίς θόρυβο εικόνας I, διαστάσεων m x n, και της θορυβώδους αναπαράστασής της K, τότε το μέσο τετραγωνικό λάθος(MSE) ορίζεται ως :

$$\text{MSE} = \frac{1}{m n} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i, j) - K(i, j)]^2$$

και το PSNR ως :

$$\begin{aligned} \text{PSNR} &= 10 \cdot \log_{10} \left(\frac{\text{MAX}_I^2}{\text{MSE}} \right) \\ &= 20 \cdot \log_{10} \left(\frac{\text{MAX}_I}{\sqrt{\text{MSE}}} \right) \\ &= 20 \cdot \log_{10} (\text{MAX}_I) - 10 \cdot \log_{10} (\text{MSE}) \end{aligned}$$

Όπου στις παραπάνω εξισώσεις το MAX_I είναι η μέγιστη δυνατή τιμή εικονοστοιχείου (pixel) της εικόνας.

Γενικά έχει δεχθεί πως το PSNR δεν αποδίδει πάντα βέλτιστα και σε συμφνία με την ανθρώπινη αντίληψη, [Huynh-Thu and Ghanbari \(2012\)](#), [Huynh-Thu \(2008\)](#) γι' αυτό θα χρησιμοποιήσουμε και άλλες μετρικές, για να είμαστε σίγουροι για το αποτέλεσμα της ποιότητας ανακατασκευής των εικόνων μας.

- **Συντελεστής συσχέτισης(correlation coefficient)** Ο συντελεστής συσχέτισης είναι ένα αριθμητικό μέτρο συσχέτισης, που δηλώνει μια στατιστική σχέση μεταξύ δύο μεταβλητών. Οι

μεταβλητές μπορεί να είναι δύο στήλες ενός συνόλου δεδομένων. Έτσι είναι πολύ εύκολο να μετατρέψουμε εικόνες σε μονοδιάστατους πίνακες, ειδικά στην γλώσσα προγραμματισμού που χρησιμοποιήσαμε, δηλαδή στην *python*, και στη συνέχεια να μετρήσουμε τον συντελεστή συσχέτισής τους.

Υπάρχουν διάφοροι τύποι συντελεστή συσχέτισης, καθένας με τον δικό του ορισμό και το δικό του εύρος χρησιμότητας και χαρακτηριστικών. Όλοι υποθέτουν τιμές στο εύρος από -1 έως $+1$, όπου ± 1 δείχνει την ισχυρότερη δυνατή συμφωνία και 0 την ισχυρότερη δυνατή διαφωνία. Ως εργαλεία ανάλυσης, οι συντελεστές συσχέτισης παρουσιάζουν ορισμένα προβλήματα, συμπεριλαμβανομένης της τάσης ορισμένων τύπων που παραμορφώνονται από τα ακραίες τιμές και τη δυνατότητα εσφαλμένης χρήσης για να συναχθεί μια αιτιώδης σχέση μεταξύ των μεταβλητών.

Συγκεκριμένα εμείς χρησιμοποιήσαμε τον συντελεστή συσχέτισης Pearson, του οποίου η χρήση είναι αρκετά διαδεδομένη στη σύγκριση εικόνων. Ο μαθηματικός του τύπος είναι ο εξής :

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

Όπου :

n είναι το μέγεθος του δείγματος

x_i, y_i είναι οι τιμές έντασης των pixel των εικόνων 1 και 2 αντίστοιχα

$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ είναι η μέση τιμή του δείγματος και αντίστοιχα ισχύει για το \bar{y}

Ο συντελεστής συσχέτισης έχει την τιμή $r = 1$ εάν οι δύο εικόνες είναι απολύτως πανομοιότυπες, $r = 0$ εάν είναι εντελώς μη συσχετισμένες, και $r = -1$ εάν είναι εντελώς αντισυσχετιζόμενες, για παράδειγμα, εάν η μία εικόνα είναι το αρνητικό της άλλης.

Ένα από τα προφανή πλεονεκτήματα του συντελεστή συσχέτισης Pearson είναι ότι συμπυκνώνει τη σύγκριση των 2 (συχνά μεγάλων) διδιάστατων εικόνων σε έναν μόνο αριθμό, r . Επιπλέον, ο συντελεστής συσχέτισης είναι εντελώς αναλλοίωτος σε γραμμικούς μετασχηματισμούς x ή y . Ως αποτέλεσμα, το r δεν επηρεάζεται ιδιαίτερα από ομοιόμορφες παραλλαγές φωτεινότητας ή αντίθεσης σε μια εικόνα. Τέτοιες χωρικά ομοιόμορφες παραλλαγές μπορεί να προκληθούν, για παράδειγμα, από διαφορές στη φωτεινότητα της πηγής φωτός με την πάροδο του χρόνου, αλλάζοντας τα επίπεδα φωτισμού φόντου ή από παραλλαγές στο κέρδος του συστήματος απεικόνισης.

Παρά τα πλεονεκτήματά του, ο συντελεστής συσχέτισης έχει και αυτός προβλήματα και περιορισμούς [Yen and Roger G. Johnston](#). Το κύριο μειονέκτημα του είναι ότι είναι υπολογιστικά δαπανηρό. Αυτό περιορίζει συχνά τη χρησιμότητά του για την σύγκριση εικόνων. Ο συντελεστής συσχέτισης είναι επίσης εξαιρετικά ευαίσθητος σε διαφόρων ειδών παραμορφώσεις της εικόνας, όπως το *image skewing* (μια εικόνα κλίνει προς τη μία κατεύθυνση), το *βινιετάρισμα* της εικόνας(μείωση της έντασης της εικόνας κοντά στις άκρες), καθώς και το *image pincushioning* (οι άκρες της εικόνας είναι κοίλες). Αυτές οι παραμορφώσεις αναπόφευκτα εμφανίζονται συχνά στα συστήματα απεικόνισης. Τέτοιες παραμορφώσεις είναι ιδιαίτερα διαδεδομένες σε εικόνες ηλεκτρονικού μικροσκοπίου λόγω των μη γραμμικοτήτων

και της πολυπλοκότητας της ηλεκτρονικής οπτικής.

Ένα άλλο πρόβλημα που συχνά παραβλέπεται σε πρακτικές εφαρμογές είναι ότι το r είναι απροσδιόριστο - λόγω μηδενικής διαίρεσης - εάν μία από τις δοκιμαστικές εικόνες έχει σταθερή, ομοιόμορφη ένταση. Άλλα προβλήματα με τον συντελεστή συσχέτισης περιλαμβάνουν πιθανή προκατάληψη Good and Smith (1984), την πολυπλοκότητα ερμηνείας του ίδιου του συντελεστή συσχέτισης Rodgers and Nicewander (1988), υπερβολική ευαισθησία σε θόρυβο pixel και παραλλαγές κέρδους Jenkin et al. (1991), Barnea and Silverman (1972), CROMBIE (1983), δυσκολίες στην διαχείριση της προοπτικής καθώς και όταν έχουμε κινούμενες πηγές φωτισμού Jenkin et al. (1991), CROMBIE (1983), CRACKNELL (1985). Επίσης υπάρχει ένα επιπλέον πρόβλημα με τον συντελεστή συσχέτισης που δεν φαίνεται να έχει συζητηθεί ή να αποδειχθεί λεπτομερώς. Μια εικόνα μπορεί να τροποποιηθεί σε μεγάλο βαθμό, χωρίς αυτό να ανιχνευθεί απ' τον συντελεστή συσχέτισης, αρκεί ο τοπικός μέσος όρος και / ή το ιστόγραμμα των εντάσεων των εικονοστοιχείων να είναι σχετικά αμετάβλητοι.

- **Mutual Information(MI)** Η αμοιβαία πληροφορία ορίστηκε και αναλύθηκε από τον Claude Shannon στο έργο του "Mathematical Theory of Communication", αν και αρχικά δεν την ονόμασε «αμοιβαία πληροφορία». Αυτός ο όρος επινοήθηκε αργότερα από τον Robert Fano. Η αμοιβαία πληροφορία είναι επίσης γνωστή ως κέρδος πληροφορίας (information gain). Στη θεωρία πιθανοτήτων και στη θεωρία πληροφορίας, η αμοιβαία πληροφορία (Mutual Information) δύο τυχαίων μεταβλητών είναι ένα μέτρο της αμοιβαίας εξάρτησής τους. Πιο συγκεκριμένα, ποσοτικοποιεί την "ποσότητα πληροφορίας" (σε μονάδες όπως shannons (bits), nats ή hartleys) που λαμβάνεται για μια τυχαία μεταβλητή από την παρατήρηση μίας άλλης τυχαίας μεταβλητής. Η έννοια της αμοιβαίας πληροφορίας συνδέεται στενά με αυτήν της εντροπίας μιας τυχαίας μεταβλητής, μια θεμελιώδη έννοια στη θεωρία της πληροφορίας που ποσοτικοποιεί την αναμενόμενη «ποσότητα πληροφορίας» που βρίσκεται σε μια τυχαία μεταβλητή.

Η αμοιβαία πληροφορία δεν περιορίζεται σε τυχαίες μεταβλητές πραγματικής τιμής και ούτε από την γραμμική εξάρτηση όπως ο συντελεστής συσχέτισης, η MI είναι πιο γενική και καθορίζει πόσο διαφορετική είναι η κοινή κατανομή του ζεύγους (X, Y) στο προϊόν των οριακών κατανομών των X και Y . Η MI είναι η αναμενόμενη τιμή της αμοιβαίας πληροφορίας σημείου (Point Mutual Information(PMI)).

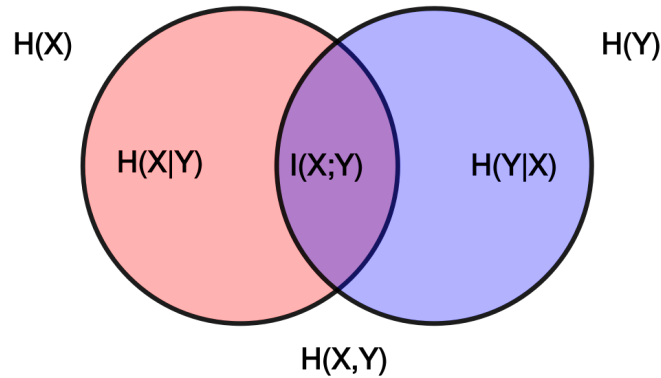
Αν θεωρήσουμε (X, Y) να είναι ένα ζεύγος τυχαίων μεταβλητών με τιμές πάνω από το διάστημα $(X \times Y)$, και εάν η κοινή κατανομή τους είναι $P(X,Y)$ και οι οριακές κατανομές τους είναι P_x και P_y , τότε η αμοιβαία πληροφορία μεταξύ τους ορίζονται ως :

$$I(X; Y) = D_{KL}(P_{(X,Y)} || P_X \otimes P_Y)$$

Σχήμα 4.1: Mutual Information Equation

όπου D_{kl} είναι η απόκλιση Kullback – Leibler. Σύμφωνα με την ιδιότητα της απόκλισης Kullback – Leibler, η αμοιβαία πληροφορία $I(X;Y)$ ισούται με το μηδέν ακριβώς όταν η κοινή

κατανομή συμπίπτει με το προϊόν των οριακών κατανομών, δηλαδή όταν τα X και Y είναι ανεξάρτητα (και επομένως παρατηρώντας το Y δεν μας δίνει καμία πληροφορία για το X).



Σχήμα 4.2: Διάγραμμα Venn που δείχνει τις προσθετικές και αφαιρετικές σχέσεις πληροφορίας που σχετίζονται με τις συσχετισμένες μεταβλητές X και Y . Η περιοχή που περιέχεται και στους δύο κύκλους είναι η κοινή εντροπία $H(X,Y)$. Ο κύκλος στα αριστερά (κόκκινο και μοβ) είναι η μεμονωμένη εντροπία $H(X)$, με το κόκκινο να είναι η υπό όρους εντροπία $H(X|Y)$. Ο κύκλος στα δεξιά (μπλε και μοβ) είναι η $H(Y)$, με το μπλε να είναι η $H(Y|X)$. Το βιολετί είναι η αμοιβαία πληροφορία (Mutual Information) $I(X;Y)$

Με πιο απλά λόγια, η αμοιβαία πληροφορία, είναι μία μετρική που δείχνει το πόσο καλά μπορούμε να προβλέψουμε το σήμα στην εικόνα Y , γνωρίζοντας τις τιμές του σήματος της εικόνας X . Το εύρος τιμών της πάει από το μηδέν(0) έως το θετικό άπειρο(+ ∞) με τις μεγαλύτερες τιμές να δηλώνουν και μεγαλύτερη αμοιβαία πληροφορία μεταξύ των εικόνων. Όταν οι προς σύγκριση εικόνες έχουν την ίδια κωδικοποίηση και αντίστοιχη ευθυγράμμιση στους άξονες, τότε το σήμα στις δύο εικόνες μπορεί να συγκριθεί με βάση τις τιμές των αντίστοιχων pixel των εικόνων. Σε αυτή την περίπτωση μία μετρική όπως ο συντελεστής συσχέτισης μας καλύπτει. Στην περίπτωση όμως που έχουν διαφορετική κωδικοποίηση, μπορεί διαφορετικές τιμές έντασης σήματος να αντιστοιχούν στο ίδιο κομμάτι ιστού της εικόνας, όπως συμβαίνει για παράδειγμα αν θέλουμε να συγκρίνουμε εικόνες T1-weighted MRI και T2-weighted MRI, κάτι που προκύπτει συχνά σε ιατρικές εφαρμογές. Σε αυτή την περίπτωση μία μετρική όπως είναι ο συντελεστής συσχέτισης δεν μας αρκεί Η αμοιβαία πληροφορία (Mutual Information) είναι μία μετρική που δεν απαιτεί τις ίδιες τιμές σήματος στις προς σύγκριση εικόνες.

Ο υπολογισμός του Mutual Information είναι αρκετά πιο περίπλοκος απ'ότι αυτός του συντελεστή συσχέτισης, καθώς χρειάστηκε πρώτα να υπολογίσουμε το ιστόγραμμα κάθε εικόνας και στη συνέχεια το κοινό τους ιστόγραμμα, και με βάση αυτό να υπολογίσουμε την αμοιβαία πληροφορία των εικόνων.

4.2 Συγκριτικά αποτελέσματα μοντέλων

Στο παρών κεφάλαιο παρουσιάζονται τα αποτελέσματα των πειραμάτων μαζί με τις παρατηρήσεις και τα συμπεράσματα που προκύπτουν από αυτά. Αρχικά γίνεται μια συγκριτική παρουσίαση μεταξύ των πειραμάτων που εκτελέστηκαν αναλύοντας την επιρροή διάφορων παραμέτρων, ενώ στη συνέχεια γίνεται μια ποιοτική αξιολόγηση των αποτελεσμάτων η οποία συνοδεύεται και από προσωπικές παρατηρήσεις.

4.2.1 Παρουσίαση Μοντέλων

Σχέδον όλα τα μοντέλα που δοκιμάσαμε ήταν αρκετά "φτηνά" υπολογιστικά, δηλαδή το σύνολο των εκπαιδευσιμων παραμέτρων τους, δεν ξεπερνούσε το μισό εκατομμύριο στις περισσότερες περιπτώσεις. Ο αριθμός των παραμέτρων ενός δικτύου είναι καθοριστικός ως προς τον χρόνο εκπαίδευσης ενός μοντέλου, το οποίο είναι καθοριστικής σημασίας όταν θέλουμε να κάνουμε δεκάδες ίσως και εκατοντάδες τρεξίματα για να μελετήσουμε τη συμπεριφορά ενός μοντέλου. Επίσης ένα μοντέλο με περισσότερες παραμέτρους απαιτεί κι ανάλογα μεγαλύτερο χώρο στη μνήμη τόσο για την εκπαίδευσή όσο και την αποθήκευσή του, το οποίο μπορεί να δημιουργήσει σημαντικό πρόβλημα εάν δεν είναι διαθέσιμο το κατάλληλο υλικό. Πρέπει λοιπόν να λαμβάνουμε υπόψιν πάντα και τους πόρους που δαπανήθηκαν για να επιτευχθεί ένα αποτέλεσμα.

Αρχικά θα γίνει μία παρουσίαση των μοντέλων εκείνων που εκπαιδεύτηκαν στο υπολογιστικό περιβάλλον του google colab με τα subsampled(200x200x162) δεδομένα και που έφεραν τα καλύτερα αποτελέσματα ανακατασκευής και θα αναλυθούν ως προς τις διάφορες παραμέτρους που τα διαφοροποιούν, και στη συνέχεια θα δούμε πως τα καλύτερα από αυτά τα μοντέλα απέδωσαν εκπαιδευόμενα με μεγαλύτερο dataset full scale(400x400x325) δεδομένων. Ο λόγος που γίνεται αυτός ο διαχωρισμός στην ανάλυση των αποτελεμάτων είναι πως στο google colab χάρη στην επιτάχυνση υλικού με gru που μας προσφέρεται αλλά και στον σχεδόν μισό όγκο δεδομένων εκπαίδευσης, μπορέσαμε να εκτελέσουμε πολύ περισσότερα πειράματα σε πολύ λιγότερο χρόνο, και μόνο τα βέλτιστα από αυτά τα δοκιμάσαμε και σε τοπικό υπολογιστή χρησιμοποιώντας τα full scale δεδομένα.

4.2.2 Subsampled δεδομένα

Μία αρχιτεκτονική που φάνηκε να αποδίδει σταθερά καλά αποτελέσματα είναι η παρακάτω :

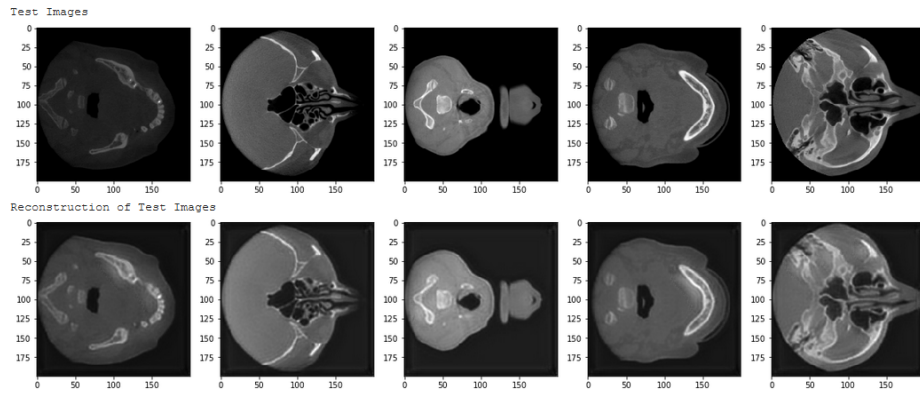
Encoder	Decoder
Conv2D	Conv2D
BatchNormalization	BatchNormalization
Conv2D	Conv2D
BatchNormalization	BatchNormalization
MaxPooling2D	UpSampling2D
Conv2D	Conv2D
BatchNormalization	BatchNormalization
Conv2D	Conv2D
BatchNormalization	BatchNormalization
MaxPooling2D	UpSampling2D
Conv2D	Conv2D
BatchNormalization	
Conv2D	
BatchNormalization	

Πίνακας 4.1: Autoencoder

Δοκιμάσαμε την παραπάνω αρχιτεκτονική, με τον συνδυασμό των υπερπαραμέτρων που φάνηκε να αποδίδουν καλύτερα, δηλαδή :

- Optimizer : Adam
- Learning Rate: 0.001
- Loss Function : Binary Crossentropy
- Batch size : 8
- Number of epochs : 20
- Activation Function : ReLu
- Activation Function of output layer : Sigmoid
- Convolution layer's filter size : 3x3

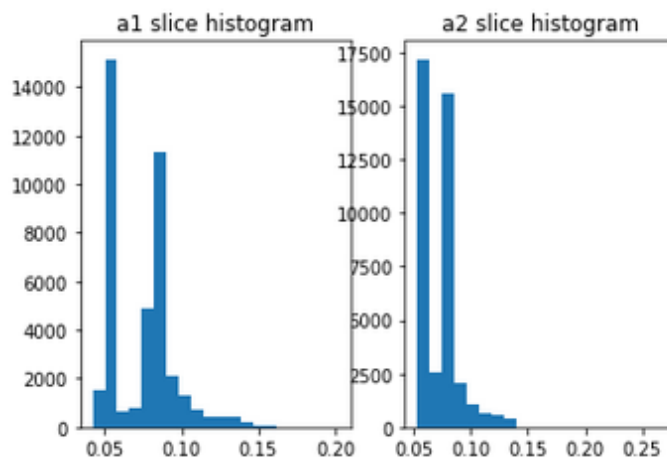
Στη συνέχεια, στο σχήμα 4.3 βλέπουμε τις εικόνες που ανακατασκευάσαμε. Με το ανθρώπινο μάτι η ανακατασκευή τους φαίνεται ικανοποιητική, κάτι που επιβεβαιώνεται βέβαια και από τις μετρικές αξιολόγησης που χρησιμοποιήσαμε όπως φαίνεται στον παρακάτω πίνακα :



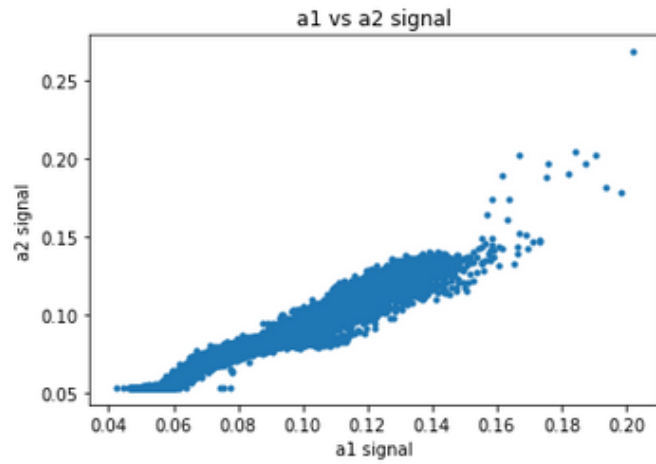
Σχήμα 4.3: Reconstructed Images

Κοιτάζοντας και τις μετρικές αξιολογήσεις βλέπουμε πως οι ανακατασκευασμένες εικόνες, σε σχέση με τις αρχικές έχουν :

- PSNR: PSNR = 45.77 dB
- Συντελεστή Συσχέτισης: Correlation Coefficient = 0.9861335151197899
- **Αμοιβαία πληροφορία(Mutual Information)** Για τον υπολογισμό της Αμοιβαίας πληροφορίας, υπολογίσαμε πρώτα τα ιστογράμματα δύο εικόνων. Μίας τυχαίας εικόνας από το αρχικό σετ εκπαίδευσης, την a1 και της αντίστοιχης ανακατασκευής της, την a2, όπως φαίνεται παρακάτω, και στη συνέχεια υπολογίσαμε το κοινό τους ιστόγραμμα όπου όσο πιο κοντά στην ευθεία $y = x$ είναι τα σημεία του κοινού ιστογράμματος, τόσο πιο παρόμοια είναι τα ιστογράμματα των δύο εικόνων :



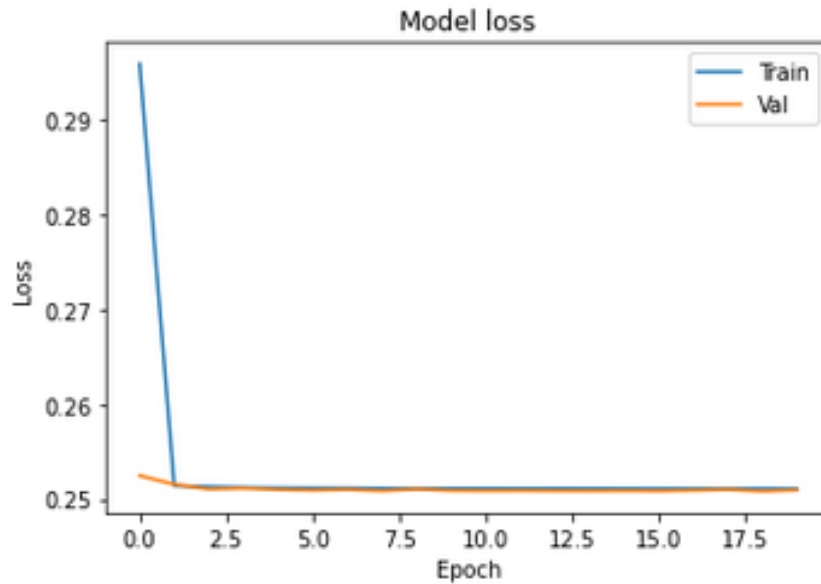
Σχήμα 4.4: a1-a2 histograms



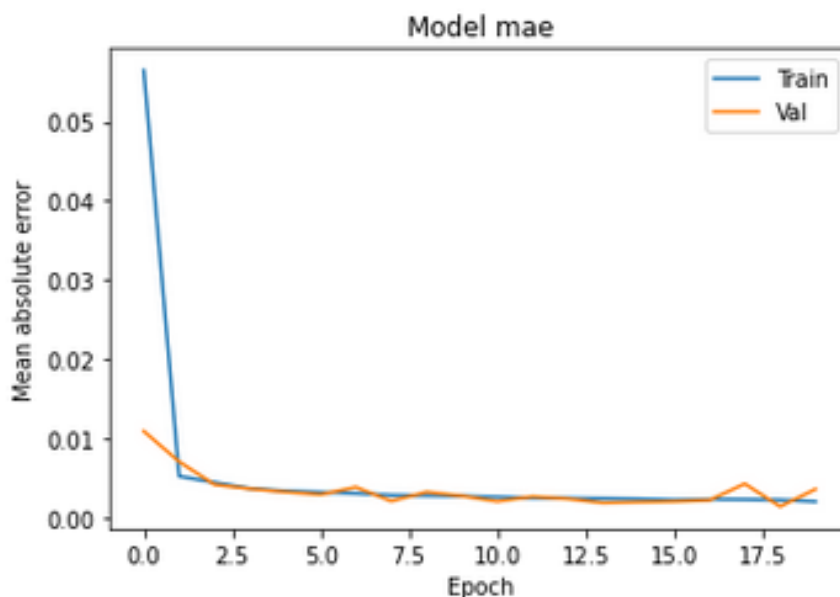
Σχήμα 4.5: Combined histogram

Και η αμοιβαία πληροφορία είναι : $Mutual\ Information = 1.059136776321159$

Βλέπουμε λοιπόν πως το συγκεκριμένο μοντέλο αποδίδει αρκετά καλά στο κομμάτι της ανακατασκευής, ας δούμε τώρα και δύο γραφήματα για τις μετρικές του ίδιου του μοντέλου :



Σχήμα 4.6: Val-Train Loss



Σχήμα 4.7: Val-Train mae

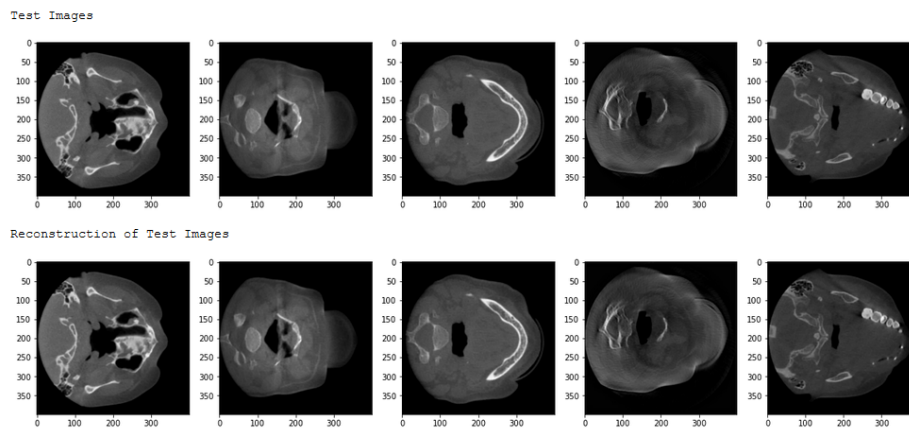
Βλέπουμε πως οι απώλειες του validation set ακολουθούν αυτές του train set, και το ίδιο με λίγες περισσότερες διακυμάνσεις συμβαίνει και για την μετρική αξιολόγησης του μοντέλου που χρησιμοποιήσαμε, το Mean Absolute Error(MAE), άρα μπορούμε να πούμε με σιγουριά ότι το μοντέλο μας δεν πάσχει από υπερμοντελοποίηση και έχει πολύ καλές δυνατότητες γενίκευσης. Αυτό φάνηκε και προηγουμένως, καθώς όλες οι εικόνες που ανακατασκευάσαμε, άνηκαν στο test set, δηλαδή ήταν εικόνες που το μοντέλο μας δεν είχε ξαναδεί ποτέ του.

Εδώ αξίζει να αναφερθεί ότι στη συγκεκριμένη αρχιτεκτονική προσθήσαμε τα Batchnormalization επίπεδα, έπειτα από κάθε συνελικτικό επίπεδο, ώστε να αποφύγουμε την πιθανή υπερμοντελοποίηση, και να μην χρειαστεί να εισάγουμε επίπεδα Dropout στο μοντέλο μας, όπως ακριβώς αναλύθηκε στο κεφάλαιο δύο(2). Παρ'όλα αυτά, δοκιμάζοντας να αφαιρέσουμε τα Batchnormalization επίπεδα, είδαμε πως το μοντέλο μας όχι μόνο δεν παρουσιάζει στοιχεία overfitting, αλλά αντιθέτως αποδίδει αισθητά καλύτερα αποτελέσματα, όπως βλέπουμε παρακάτω :

Εδώ φαίνεται η νέα αρχιτεκτονική του μοντέλου μας χωρίς τα επίπεδα BatchNormalization και ακολουθεί, όπως και πριν, ένα δείγμα 5 εικόνων του αρχικού dataset και των αντίστοιχων ανακατασκευών τους :

Encoder	Decoder
Conv2D	Conv2D
Conv2D	Conv2D
MaxPooling2D	UpSampling2D
Conv2D	Conv2D
Conv2D	Conv2D
MaxPooling2D	UpSampling2D
Conv2D	Conv2D
Conv2D	-

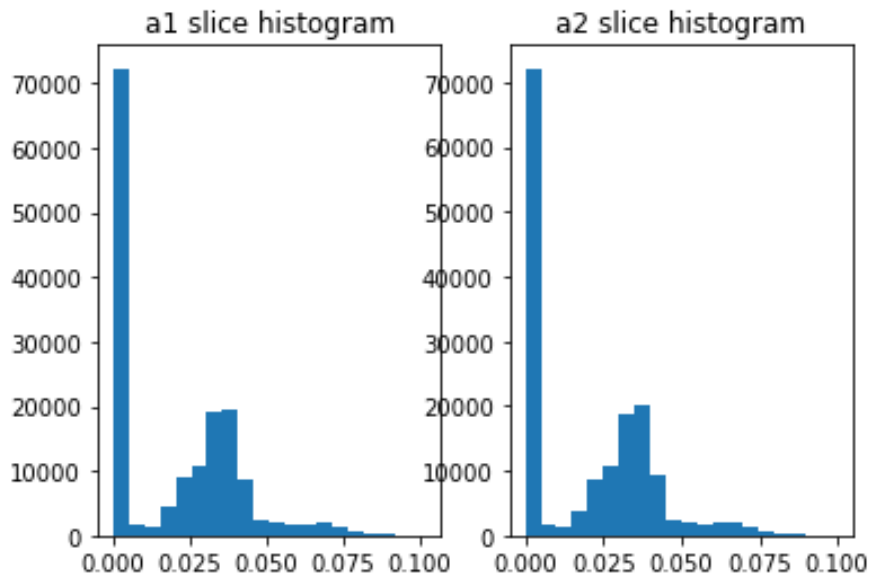
Πίνακας 4.2: Autoencoder without BatchNormalization layers



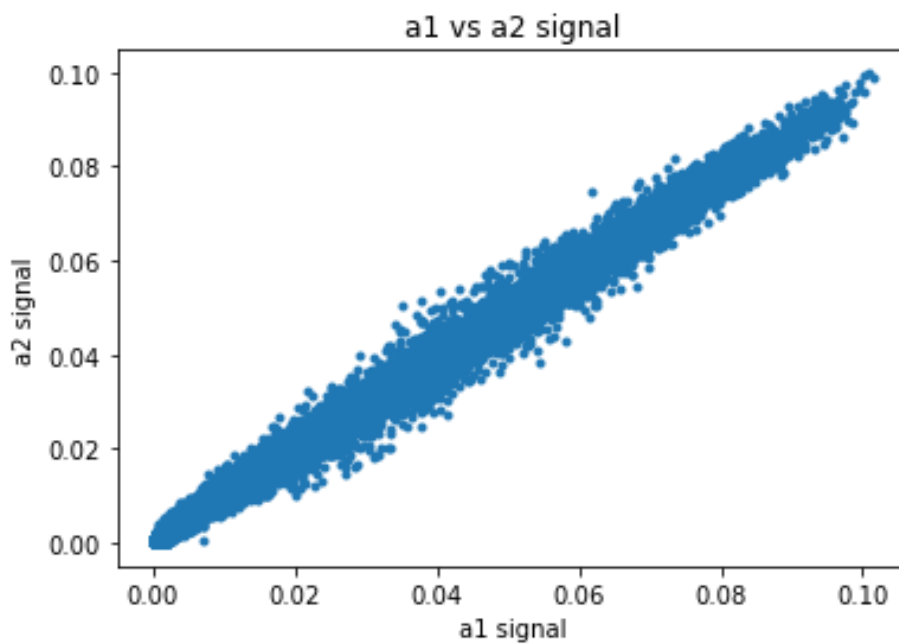
Σχήμα 4.8: Reconstructed Images

Μετρικές αξιολόγησης :

- PSNR : PSNR = 55.26dB
- Συντελεστής Συσχέτισης : Correlation coefficient = 0.9984484132883019
- Mutual Information(Αμοιβαία πληροφορία) : MU = 1.5623168914820418

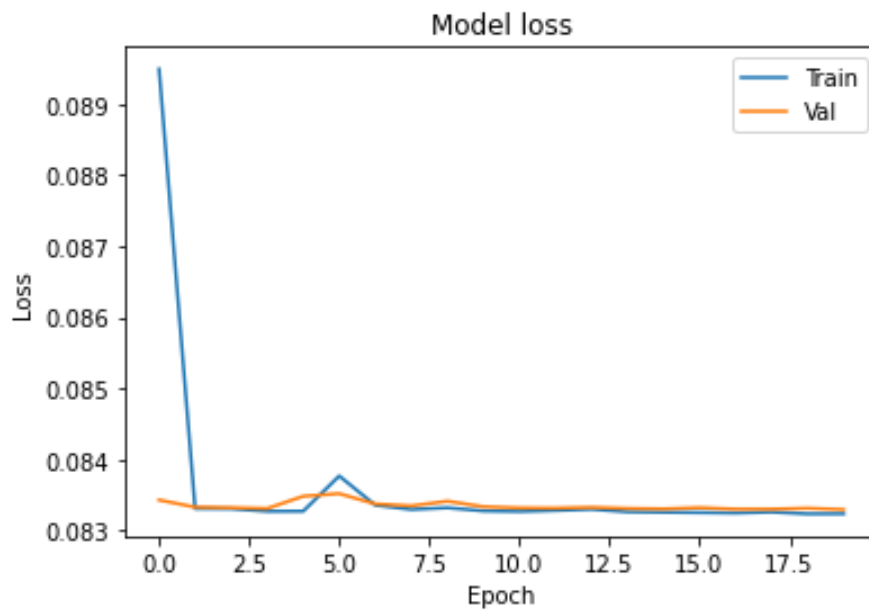


Σχήμα 4.9: a1-a2 histograms

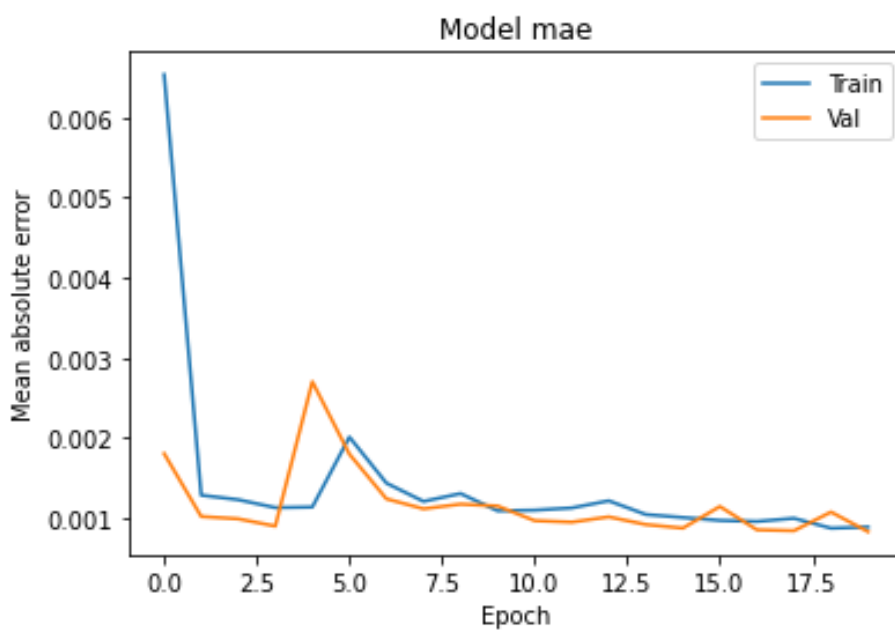


Σχήμα 4.10: Combined histogram

Βλέπουμε λοιπόν, πως και όλες οι εικόνες μας φαίνονται καλύτερες αλλά και όλες οι μετρικές αξιολόγησης συμφωνούν με την ανθρώπινη αντίληψη για την συγκεκριμένη ανακατασκευή. Παρακάτω βλέπουμε τις μετρικές απόδοσης του μοντέλου μας όπως και προηγουμένως :



Σχήμα 4.11: Val-Train Loss



Σχήμα 4.12: Val -Train mae

Όπως και πριν βλέπουμε πως οι απώλειες του validation set μας ακολουθούν αυτές του training set, πράγμα που σημαίνει ότι το μοντέλο μας μπορεί να γενικεύσει χωρίς πρόβλημα.

Στον παρακάτω πίνακα βλέπουμε τις επιδόσεις του μοντέλου μας για διάφορες τιμές του batch size, κρατώντας σταθères τις υπόλοιπες υπερπαραμέτρους στις τιμές που αρχικά αναφέραμε.

Πίνακας 4.3: Metrics according to Batch Size

Batch Size	PSNR	Correlation Coefficient	Mutual Information
4	55.2dB	0.9977341972362747	1.2709070373417723
8	55.26dB	0.9984484132883019	1.5623168914820418
16	47.39 dB	0.9948873492437569	1.2079608603045642
32	51.99dB	0.9951206209490512	1.1764364708288404

4.2.3 Fullscale δεδομένα

Μετά τις αρχικές δοκιμές που κάναμε στο υπολογιστικό περιβάλλον του Google Colab, έγιναν και πειράματα τοπικά, σε προσωπικό φορητό υπολογιστή, όπου όπως αναφέρθηκε προηγουμένως, με τη χρήση της κλάσης των Generators, μπορέσαμε να εκπαιδεύσουμε το δίκτυό μας χρησιμοποιώντας τα full-scale δεδομένα. Δηλαδή τρισδιάστατες εικόνες μεγέθους 400x400x325. Επίσης είναι πολύ σημαντικό πως έτσι μπορέσαμε να τροφοδοτήσουμε το δίκτυό μας με σχεδόν διπλάσιο αριθμό, προς εκπαίδευση, εικόνων, κάτι που σημαίνει πως μειώνονται πολύ οι πιθανότητες υπερμοντελοποίησης, και θα περιμέναμε να δούμε καλύτερα αποτελέσματα ανακατασκευής.

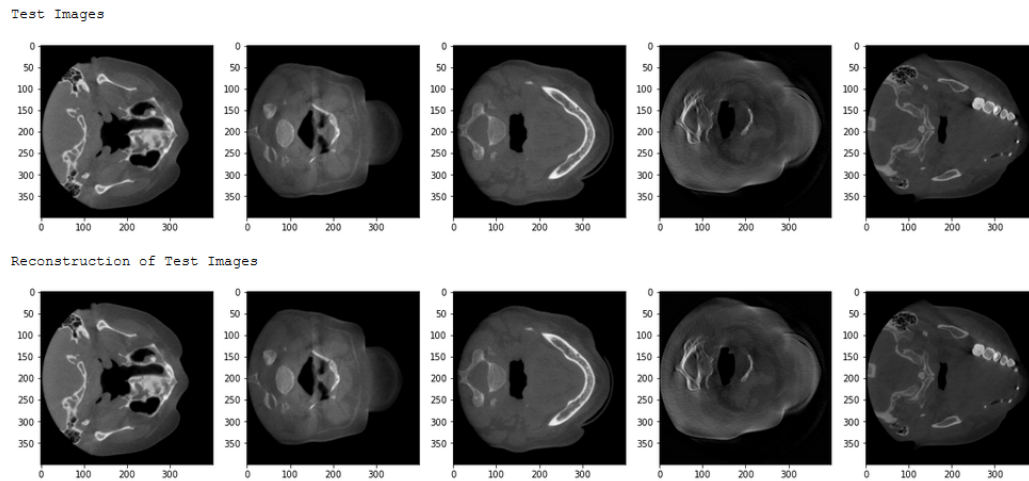
Η αρχιτεκτονική που θα δούμε είναι ίδια με προηγουμένως:

Encoder	Decoder
Conv2D	Conv2D
BatchNormalization	BatchNormalization
Conv2D	Conv2D
BatchNormalization	BatchNormalization
MaxPooling2D	UpSampling2D
Conv2D	Conv2D
BatchNormalization	BatchNormalization
Conv2D	Conv2D
BatchNormalization	BatchNormalization
MaxPooling2D	UpSampling2D
Conv2D	Conv2D
BatchNormalization	
Conv2D	
BatchNormalization	

Πίνακας 4.4: Autoencoder

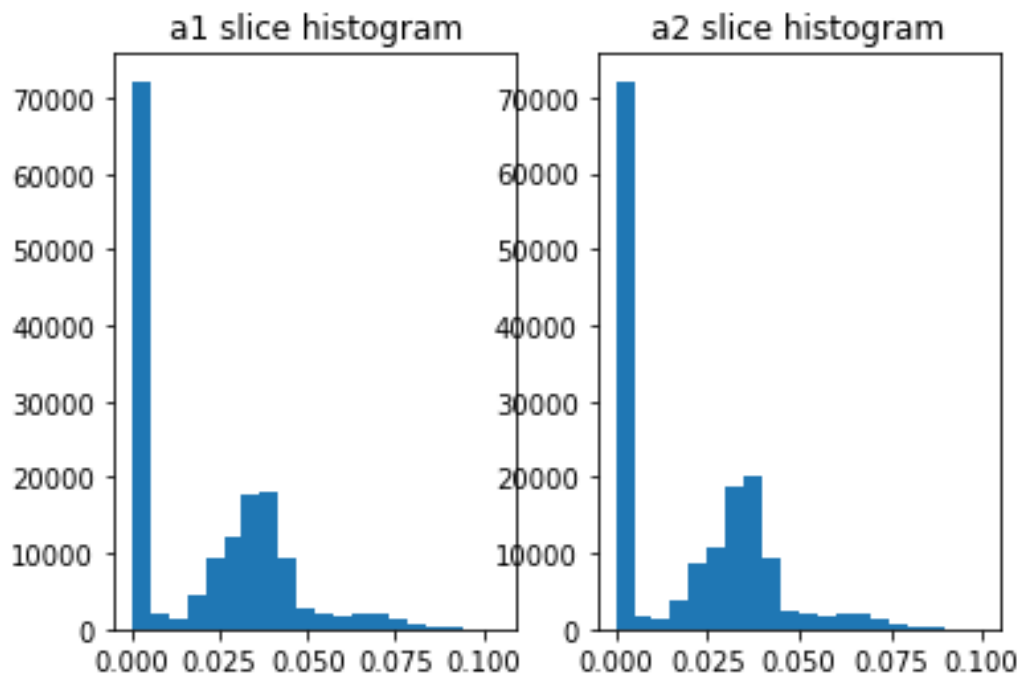
Χρησιμοποιώντας την ίδια αρχιτεκτονική με πριν και με όλες τις υπερπαραμέτρους ίδιες με προηγουμένως βλέπουμε πως όντως τα αποτελέσματά μας είναι αισθητά καλύτερα.

Εδώ βλέπουμε τις ανακατασκευασμένες εικόνες και ακολούθως τις μετρικές αξιολόγησής τους.

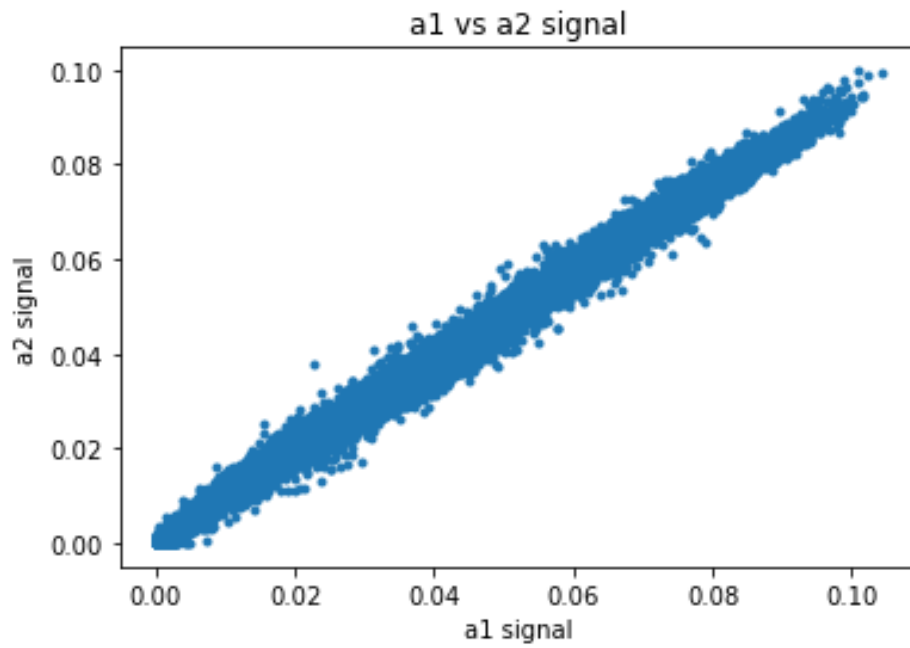


Σχήμα 4.13: Reconstructed Images

- PSNR : PSNR = 53.78dB
- Συντελεστής Συσχέτισης : Correlation coefficient = 0.9986706567944232
- Αμοιβαία Πληροφορία(Mutual Information) : MI = 1.5798891911895605

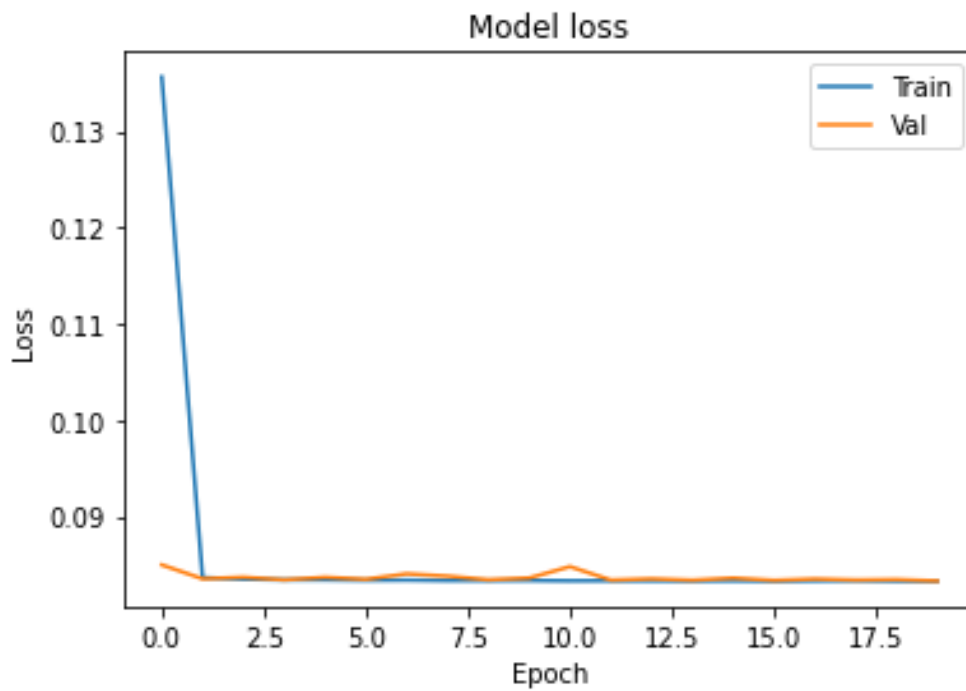


Σχήμα 4.14: a1-a2 histograms

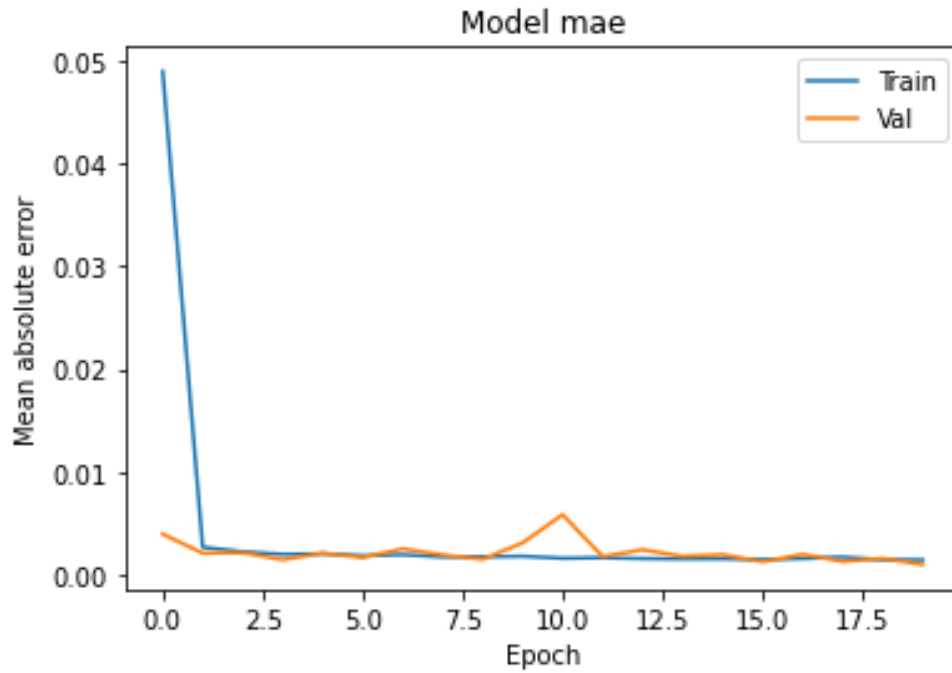


Σχήμα 4.15: Combined histogram

Ακολούθως βλέπουμε πάλι πως οι απώλειες του validation set ακολουθούν αυτές του training set μας, συνεπώς το μοντέλο μας συνεχίζει να έχει πολύ καλές ικανότητες γενίκευσης.

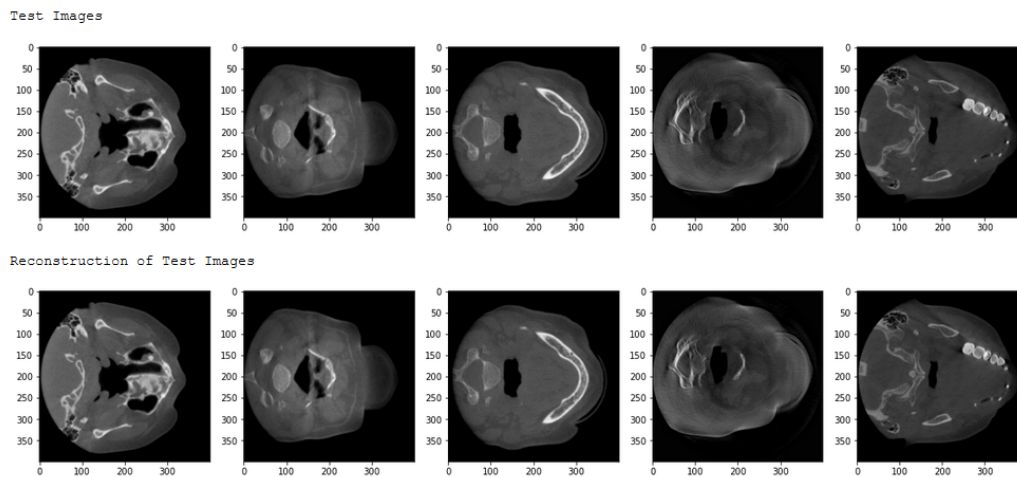


Σχήμα 4.16: Val-Train Loss



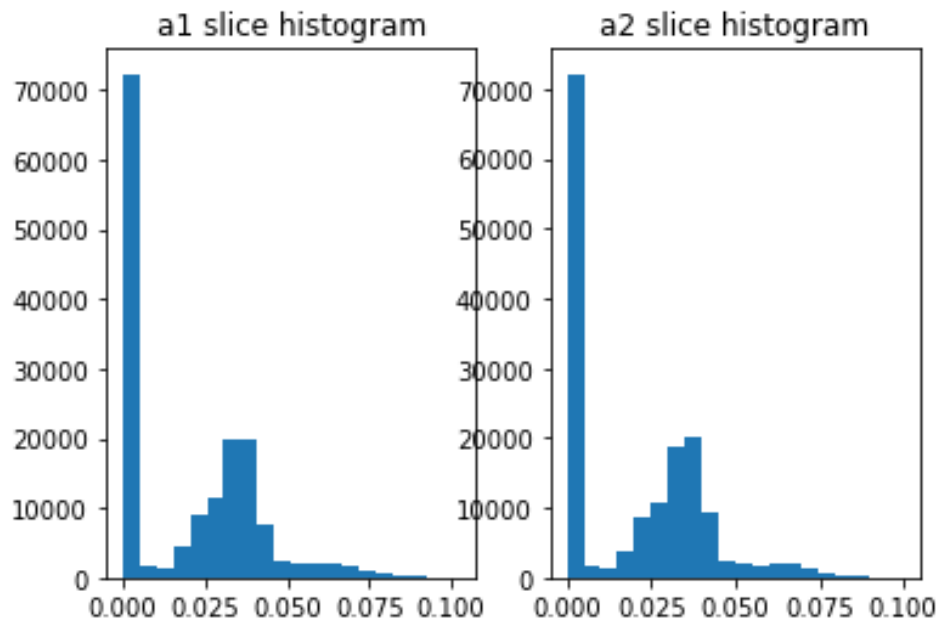
Σχήμα 4.17: Val-Train mae

Αντίστοιχα θα δούμε και τα αποτελέσματα της αρχιτεκτονικής χωρίς τα BatchNormalization επίπεδα, όταν χρησιμοποιήσαμε generators, πάλι με όλες τις υπερπαραμέτρους να παραμένουν ίδιες :

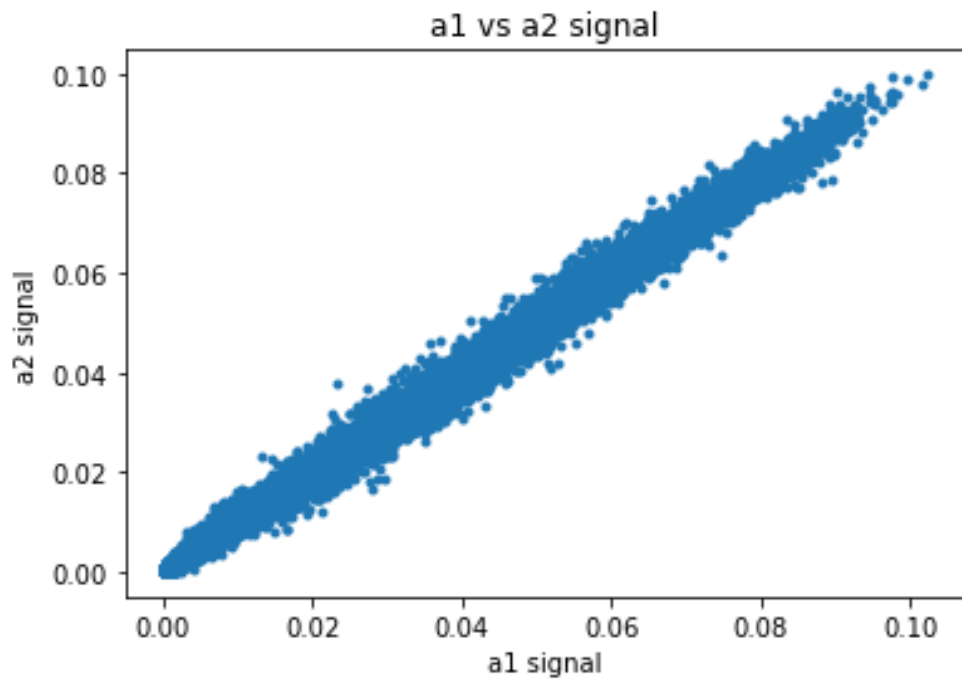


Σχήμα 4.18: Reconstructed Images

- PSNR: PSNR = 56.4dB
- Συντελεστής Συσχέτισης : Correlation Coefficient = 0.998893914767466
- Αμοιβαία Πληροφορία(Mutual Information) : MI = 1.5823484772650072
-



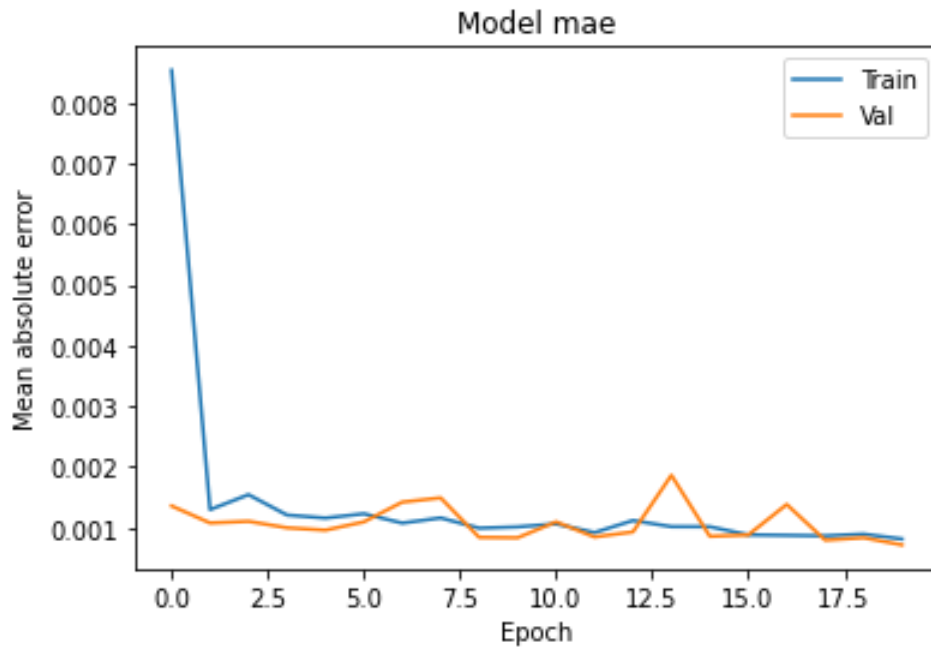
Σχήμα 4.19: a1-a2 histograms



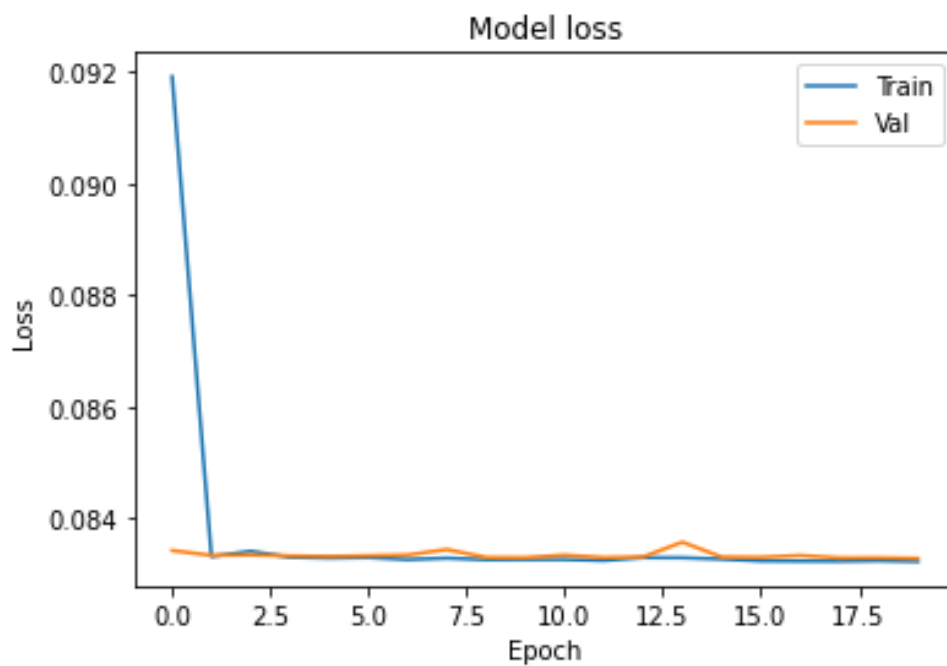
Σχήμα 4.20: Combined histogram

Στη συνέχεια βλέπουμε ξανά τα γραφήματα με τις απώλειες του μοντέλου μας. Παρατηρούμε πως έχουμε με διαφορά τις μικρότερες απέλειες για το σετ επιβεβαίωσης(validation set) που έχουμε δει ως τώρα, κάτι αναμενόμενο αφού το συγκεκριμένο μοντέλο πέτυχε και τα καλύτερα αποτελέσματα ανακατασκευής. Και αντίστοιχα με πριν βλέπουμε πως οι απώλειες του validation set ακολουθούν

αυτές του training set, αρά συμπεραίνουμε όπως ήταν αναμενόμενο πώς ούτε τώρα το μοντέλο μας πάσχει από υπερμοντελοποίηση(overfitting).



Σχήμα 4.21: Val-Train Loss

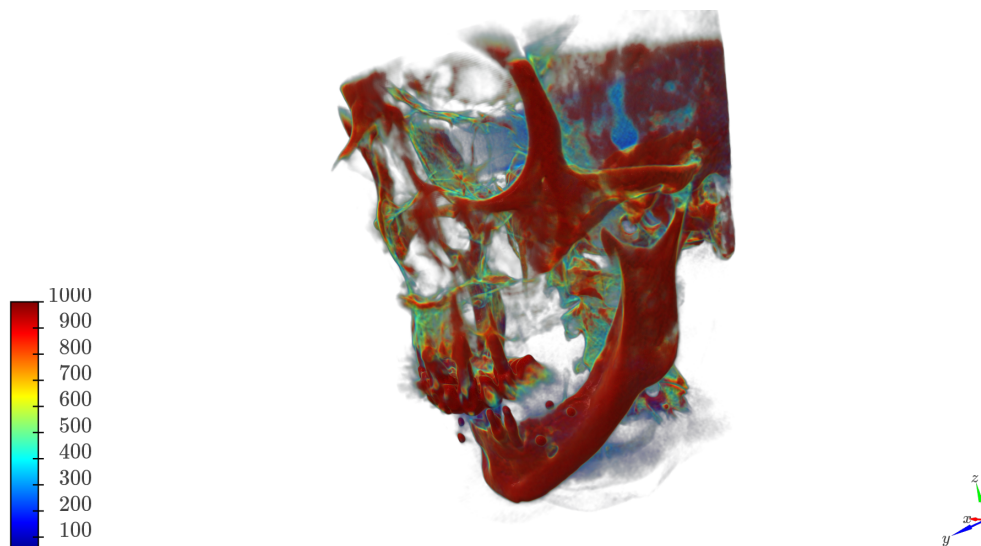


Σχήμα 4.22: Val-Train mae

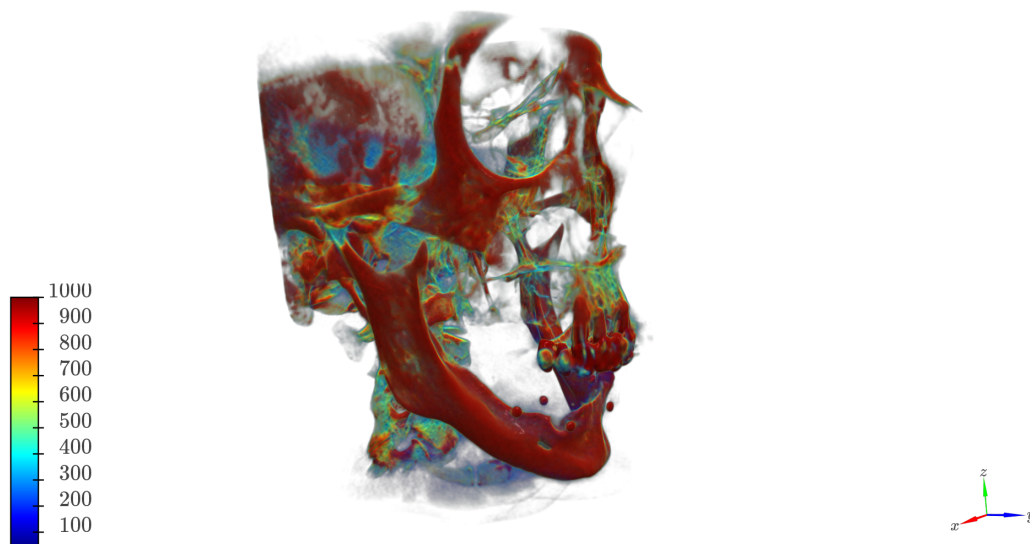
4.3 Τρισδιάστατη Αναπαράσταση

Μέχρι στιγμής ασχοληθήκαμε με μοντέλα που ανακατασκευάζουν εικόνες δύο διαστάσεων(2D), ο σκοπός όμως της διπλωματικής μας είναι η ανακατασκευή τρισδιάστατων(3D) κρανιακών scans. Για να το πετύχουμε αυτό ασχοληθήκαμε με πολλά εργαλεία που προσφέρουν απεικόνιση ογκομετρικών δεδομένων, πολλά από αυτά εμφάνιζαν απλώς στατικές όψεις της τρισδιάστατης εικόνας, χωρίς να έχουμε εμείς τη δυνατότητα να επιλέξουμε τις όψεις που θέλουμε, και να επεξεργαστούμε την εικόνα εν συνόλω. Έτσι καταλήξαμε στο K3D-jupyter το οποίο είναι ένα πακέτο τρισδιάστατης οπτικοποίησης σε Jupyter notebook. Το K3D είναι από τα πιο εύχρηστα και αποτελεσματικά εργαλεία που χρησιμοποιήσαμε για τρισδιάστατη απεικόνιση. Έχει σχεδιαστεί με εγγενή διαλειτουργικότητα με υπάρχουσες ισχυρές βιβλιοθήκες όπως η VTK, ωστόσο, δεν εξαρτάται πλήρως από αυτές. Τα γραφήματα K3D είναι ipywidget, διασφαλίζοντας έτσι την τέλεια αλληλεπίδραση ενός widget προγράμματος περιήγησης με τον πυρήνα της Python με ένα τυποποιημένο πρωτόκολλο. Παρακάτω θα δούμε διάφορες όψεις μίας τρισδιάστατης εικόνας όπως την εμφανίσαμε μέσω του K3D. Αρχικά θα δούμε την αναπαράσταση μίας εικόνας στις πραγματικές της διαστάσεις, δηλαδή 400x400x325 και στη συνέχεια θα δούμε την αναπαράσταση της ίδια εικόνας μετά από resampling όπου οι διαστάσεις της είναι 200x200x162. Resampling το οποίο και χρησιμοποιήσαμε για τις εικόνες που τροφοδοτήσαμε τα μοντέλα μας στο Google Colab.

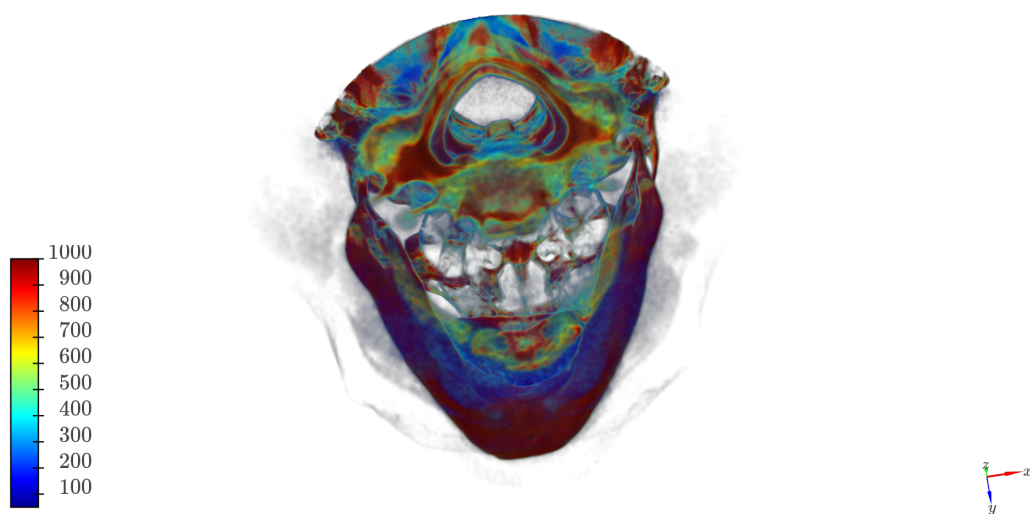
Τρισδιάστατη απεικόνιση εικόνων στις πραγματικές τους διαστάσεις :



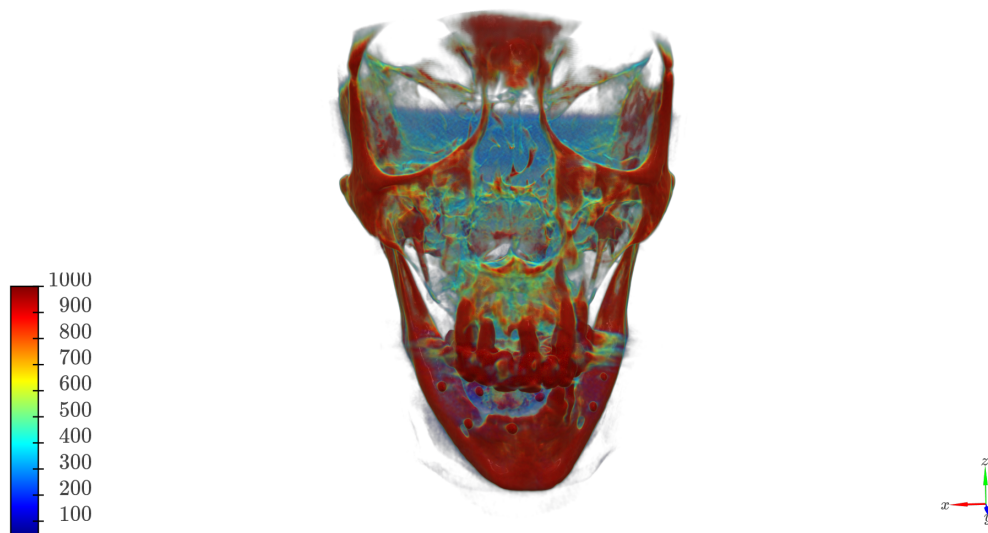
Σχήμα 4.23: 1^ο Πλάγια Αριστερή Όψη



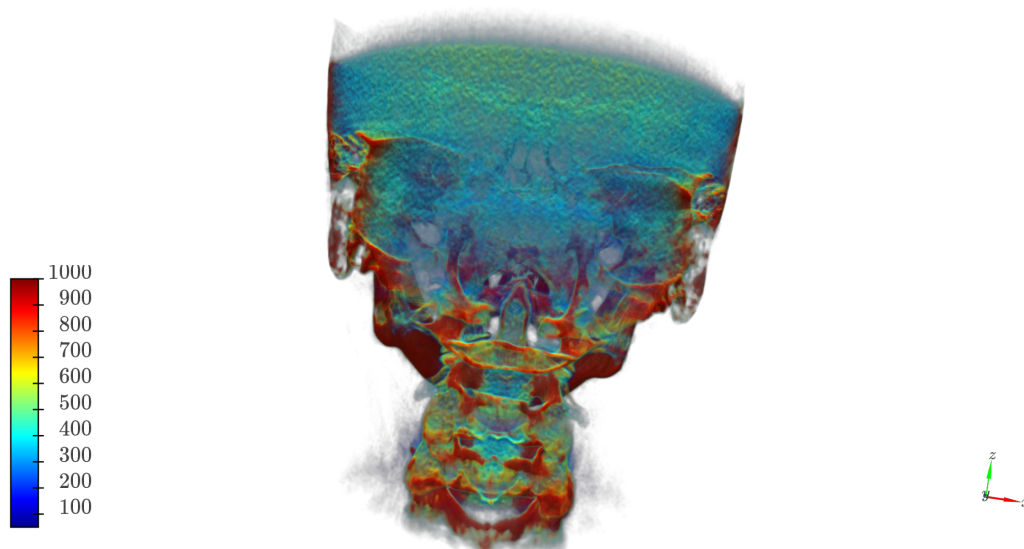
Σχήμα 4.24: 1° Πλάγια Δεξιά Όψη



Σχήμα 4.25: 1° Κάτω Όψη

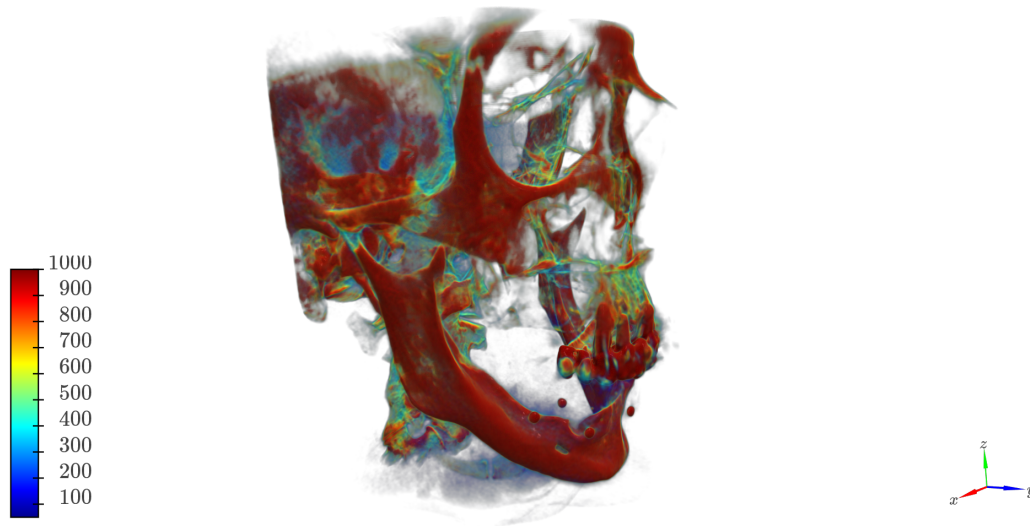


Σχήμα 4.26: 1^ο Μπροστά Όψη

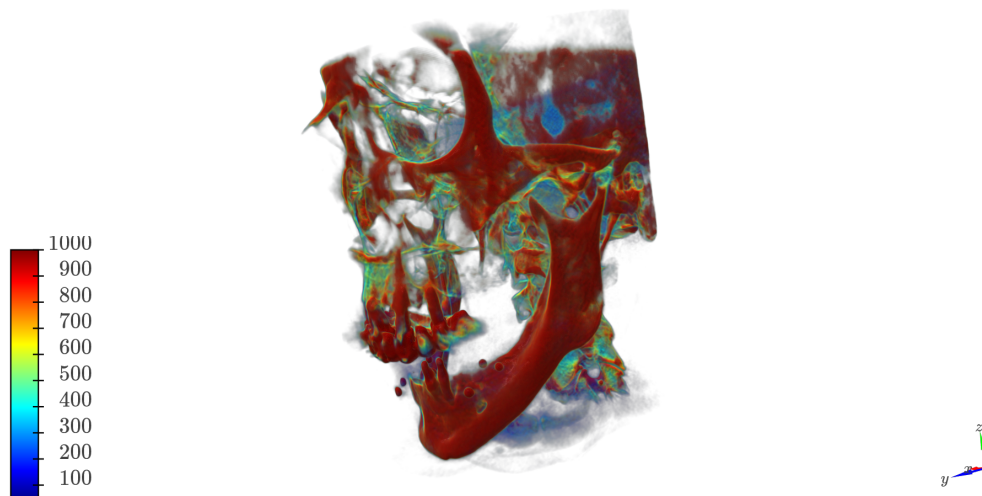


Σχήμα 4.27: 1^ο Πίσω Όψη

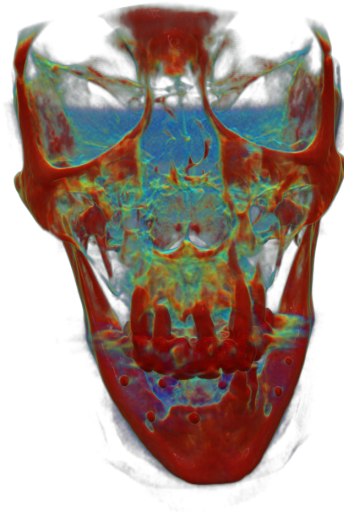
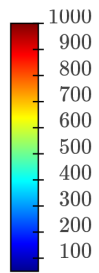
Τρισδιάστατη απεικόνιση εικόνων μετά από resampling :



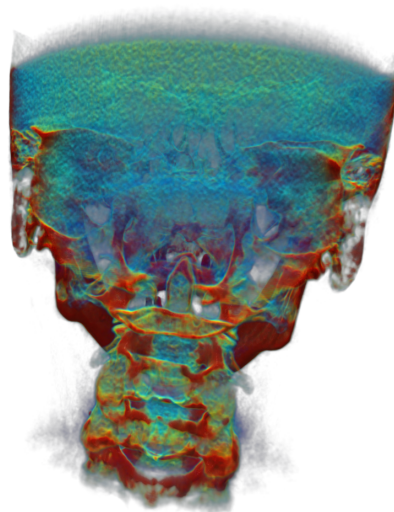
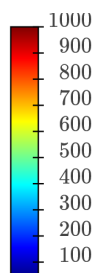
Σχήμα 4.28: 1° Πλάγια Αριστερή Όψη



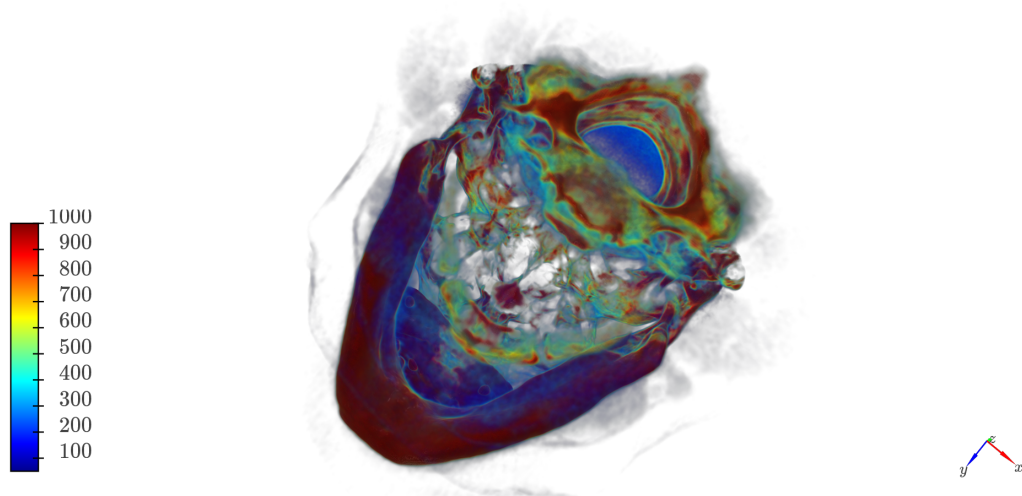
Σχήμα 4.29: 2° Πλάγια Δεξιά Όψη



Σχήμα 4.30: 2° Μπροστά Όψη

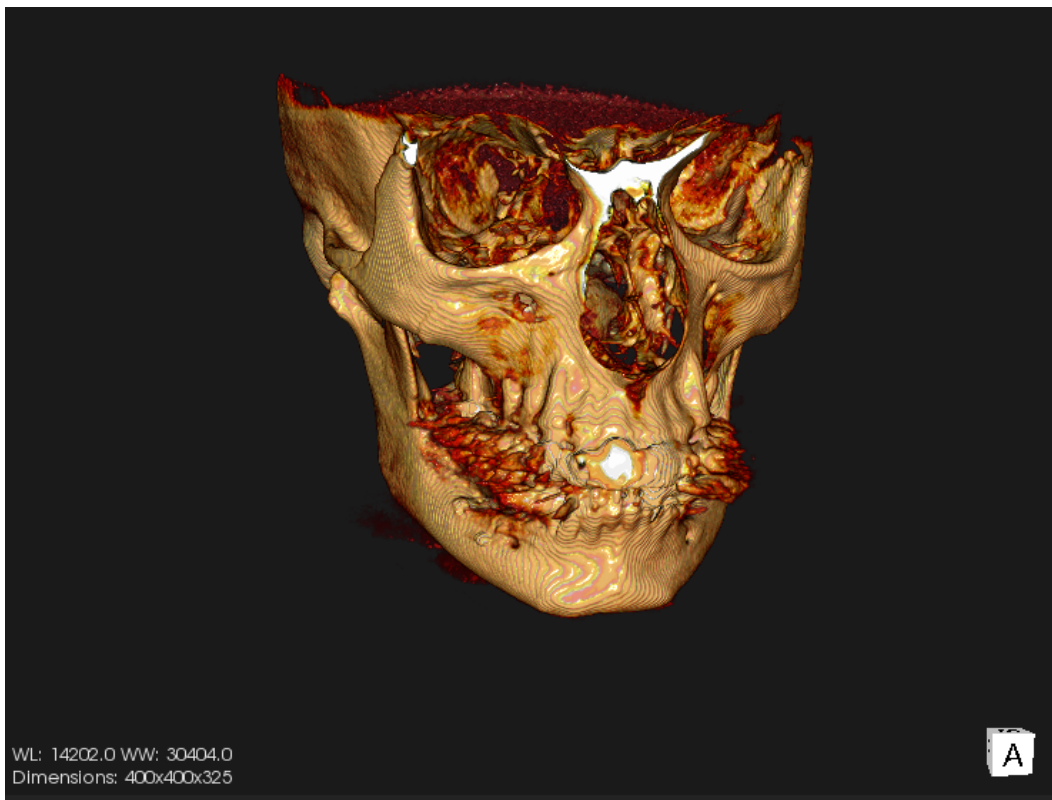


Σχήμα 4.31: 2° Πίσω όψη



Σχήμα 4.32: 2^ο Κάτω Όψη

Ένα ακόμη εργαλείο που χρησιμοποιήσαμε με ικανοποιητικά αποτελέσματα, είναι το Vesalius3D, πρόγραμμα που εγκαταστήσαμε σε τοπικό υπολογιστή, και αποκτήσαμε άδεια χρήσης του ως μέλη της πολυτεχνιακής κοινότητας, παρέχοντας τα κατάλληλα διαπιστευτήρια.



Σχήμα 4.33: 3D απεικόνιση εικόνας 400x400x325 από Vesalius3D

4.4 Προσωπικές Παρατηρήσεις

Στη διπλωματική αυτή εργαστήκαμε με δύο ελαφρώς διαφορετικά σύνολα δεδομένων. Η περιγραφή του κάθε συνόλου δεδομένων έχει γίνει ήδη σε προηγούμενα κεφάλαια. Η βασική διαφορά των δύο συνόλων δεδομένων είναι το μέγεθος, καθώς το dataset του Google Colab περιελάμβανε κοντά στις 4000 εικόνες, ενώ αντίθετα το dataset με το οποίο εργαστήκαμε τοπικά, περιελάμβανε κοντά στις 8000 εικόνες, δηλαδή είχε σχεδόν το διπλάσιο μέγεθος. Συνεπώς τα αποτελέσματα που προκύπτουν από το μεγαλύτερο dataset είναι σίγουρα πιο αξιόπιστα, καθώς το μοντέλο έχει εκπαιδευτεί "βλέποντας" μία πολύ μεγαλύτερη ποικιλία εικόνων, μειώνοντας κατά πολύ την πιθανότητα υπερμοντελοποίησης. Επίσης αξίζει να αναφερθεί πως με βάση τις μετρικές αξιολογήσεις που παρουσιάσαμε φαίνεται και από την σχετική βιβλιογραφία πως πετύχαμε εξαιρετικά αποτελέσματα στο κομμάτι της ανακατασκευής εικόνας δύο διαστάσεων. Καταληκτικά θα θέλαμε να επισημάνουμε και τη διαφορά στο χρόνο εκπαίδευσης των δικτύων ανάλογα με το σύνολο δεδομένων με το οποίο το εκπαιδεύσαμε. Βέβαια η διαφορά αυτή δεν οφείλεται μόνο στη διαφορά μεγέθους των δύο συνόλων αλλά και στον διαφορετικό τρόπο εκπαίδευσης που υλοποιήσαμε στις δύο περιπτώσεις, ο οποίος εκτόξευσε ακόμα περισσότερο τη διαφορά αυτή. Συγκεκριμένα το ίδιο ακριβώς δίκτυο απαιτούσε μέχρι και 120 φορές περισσότερο χρόνο όταν εκπαιδευόταν με το μεγαλύτερο dataset. Να τονίσουμε όμως, καθώς τα πειράματα αυτά γίνονταν τοπικά σε προσωπικό υπολογιστή, ότι τα τεχνικά χαρακτηριστικά του εκάστοτε υπολογιστή επηρεάζουν σε πολύ μεγάλο βαθμό τον απαιτούμενο χρόνο εκπαίδευσης.

Κεφάλαιο 5

Επίλογος

Στο παρών κεφάλαιο συνοψίζεται η δουλειά που έγινε στη διπλωματική και στη συνέχεια παρουσιάζονται επιγραμματικά τα συμπεράσματα που προέκυψαν από τα προηγούμενα κεφάλαια. Τέλος παρουσιάζονται μερικές προτάσεις για τις μελλοντικές κατευθύνσεις της επιστημονικής μελέτης.

5.1 Σύνοψη

Στην παρούσα διπλωματική εργασία πραγματοποιήσαμε το πρόβλημα της ανακατασκευής τρισδιάστατων εικόνων. Πρόκειται για ένα δύσκολο και απαιτητικό πρόβλημα με εφαρμογή σε μεγάλο πλήθος σύγχρονων τεχνολογιών. Για να το πετύχουμε αυτό χρησιμοποιήσαμε ένα από τα πιο ισχυρά εργαλεία της τεχνητής νοημοσύνης που έχουμε στη διάθεσή μας, τα συνελκτικά νευρωνικά δίκτυα. Έτσι με βάση τα συνελκτικά επίπεδα, δομίσουμε έναν αυτοκωδικοποιητή και προσπαθήσαμε να βρούμε την βέλτιστη αρχιτεκτονική αλλά και αλληλουχία υπερπαραμέτρων που θα μας προσέφερε τα καλύτερα αποτελέσματα ανακατασκευής. Επίσης εκτός από τις διάφορες υπερπαραμέτρους μελετήσαμε και το πως το μέγεθος του σετ δεδομένων που χρησιμοποιούμε για να εκπαιδεύσουμε το δίκτυό μας επηρεάζει και την απόδοσή του. Για να το μελετήσουμε αυτό χρησιμοποιώντας τα full-scale δεδομένα τα οποία δεν χωρούσαν στην μνήμη RAM του υπολογιστή μας, χρειάστηκε να εφαρμόσουμε μία τεχνική εκπαίδευσης νευρωνικών δικτύων που χρησιμοποιείται πολύ συχνά όταν ερχόμαστε αντιμέτωποι με τόσο μεγάλο όγκο δεδομένων. Η τεχνική αυτή κάνει χρήση της λογικής των generators, όπου σε κάθε κύκλο εκπαίδευσης, παίρνουμε batch size αριθμό εικόνων από τον σκληρό δίσκο και τις τροφοδοτούμε στο δίκτυό μας. Η διαδικασία αυτή επαναλαμβάνεται μέχρι να εξαντληθούν τα προς εκπαίδευση δεδομένα.

Στη συνέχεια πειραματιστήκαμε με το πόσο σημαντική είναι η προεπεξεργασία των δεδομένων μας και η κανονικοποίηση των τιμών των εικονοστοιχείων, καθώς και με κάποιες τεχνικές αποφυγής υπερμοντελοποίησης τις οποίες όμως τελικά δεν υιοθετήσαμε.

Για την σύγκριση των αποτελεσμάτων κάθε μοντέλου ώστε να μπορέσουμε να καταλήξουμε σε μια βέλτιστη αρχιτεκτονική, και σε έναν βέλτιστο συνδυασμό υπερπαραμέτρων χρησιμοποιήσαμε τρεις(3) διαφορετικές μετρικές αξιολόγησεις που χρησιμοποιούνται ευρέως στην σύγκριση εικόνων. Τον λόγο μέγιστου σήματος προς θόρυβο(PSNR), τον συντελεστή συσχέτισης(correlation coefficient), και την αμοιβαία πληροφορία(Mutual Information). Τέλος αφού το δίκτυό μας κατασκευάστηκε με σκοπό την ανακατασκευή εικόνων δύο διαστάσεων, χρειάστηκε η χρήση επιπλέον εξειδικευμένου λογισμικού ώστε να έχουμε μία τριαδιάστατη αναπαράσταση των εικόνων που χρησιμοποιήσαμε, το οποίο ήταν και ουσιαστικά ο σκοπός της διπλωματικής. Μέσα από τη μελέτη κι εφαρμογή όσων

διαβάσαμε, εξοικειωθήκαμε με τα εργαλεία και τις μεθόδους της βαθιάς μάθησης και κατανοήσαμε τη λειτουργία των συνελκτικών νευρωνικών δικτύων. Αυτή η διπλωματική εργασία λοιπόν αποτελεί μια παρουσίαση των συνελκτικών αυτοκωδικοποιητών και των παραμέτρων που τους επηρεάζουν. Καταληκτικά καταφέραμε να επιτύχουμε αξιοσημείωτη ακρίβεια στο κομμάτι της ανακατασκευής εικόνας.

5.2 Τελικά Συμπεράσματα

Όπως αναφέρθηκε και παραπάνω ο βασικός σκοπός της διπλωματικής εργασίας ήταν η εξοικείωση με τη βαθιά μάθηση, τα εργαλεία και της μεθόδους της καθώς και μια εκτενής μελέτη των χαρακτηριστικών που οδηγούν σε ένα καλό μοντέλο. Θεωρούμε πως όλα τα παραπάνω καλύφθηκαν κατά την εκτέλεση της εργασίας κι επιπλέον τα αποτελέσματα τα οποία επιτύχαμε κρίνονται ικανοποιητικά σε σύγκριση με άλλες εργασίες. Τα συμπεράσματα στα οποία καταλήξαμε με την εκπόνηση αυτής της διπλωματικής προκύπτουν τόσο από τα πειράματα που εκτελέσαμε όσο κι από τη σύγκριση με εργασίες άλλων ομάδων πάνω στο ίδιο πρόβλημα. Συνοπτικά - Επιγραμματικά αυτά είναι:

- Τα συνελκτικά νευρωνικά δίκτυα δείχνουν να υπερέχουν των υπόλοιπων νευρωνικών δικτύων στον τομέα της ανακατασκευής εικόνων.
- Τα δεδομένα αποτελούν από τα σπουδαιότερα κομμάτια ενός μοντέλου, γι' αυτό και πρέπει πάντα να επιλέγονται τα κατάλληλα δεδομένα, να μελετούνται καλά και να τα προεπεξεργαζόμαστε καταλλήλως. Ένα μικρό ή μη έγκυρο σύνολο δεδομένων ίσως να μην δώσει το ίδιο αξιόπιστα αποτελέσματα, αλλά το ικανοποιητικό μέγεθος για ένα σύνολο δεδομένων καθορίζεται από το ίδιο το πρόβλημα.
- Το πλήθος των παραμέτρων μιας αρχιτεκτονικής δεν καθορίζει απαραίτητα την ικανότητά της.
- Το learning rate, ο optimizer που επιλέγουμε αλλά το batch size αποτελούν μερικές από τις βασικότερες υπερπαραμέτρους ενός δικτύου και γι' αυτό χρειάζεται και ιδιαίτερη προσοχή στον προσδιορισμό τους.
- Πάντοτε υπάρχουν περιθώρια βελτίωσης της υπάρχουσας δουλειάς, ειδικά στο κομμάτι της βαθιάς μάθησης όπου ο χώρος αναζήτησης των βέλτιστων παραμέτρων είναι τεράστιος.

5.3 Μελλοντικές Κατευθύνσεις Επιστημονικής Μελέτης

Δεδομένων των καλών μετρικών ανακατασκευής του δικτύου μας, υπάρχει η προοπτική να αποτελέσει μια βάση για περαιτέρω ανάπτυξη μοντέλων για ανακατασκευή εικόνων. Επίσης η ύπαρξη ενός ισχυρότερου υπολογιστικού περιβάλλοντος που θα μείωνε τους χρόνους εκπαίδευσης του μοντέλου μας, θα μας επέτρεπε να το εκπαιδεύσουμε χρησιμοποιώντας ένα ακόμη μεγαλύτερο σετ δεδομένων και σε συνδυασμό με μια εκτενέστερη μελέτη των προς χρήση υπερπαραμέτρων θα μπορούσαν να οδηγήσουν σε σημαντική αύξηση της απόδοσης του μοντέλου. Τα παραπάνω αφορούν μελλοντική μελέτη στον άξονα στον οποίο κινηθήκαμε στα πλαίσια της διπλωματικής

εργασίας. Ξεφεύγοντας από τα αυστηρά όρια της ανακατασκευής εικόνας μεγάλο ερευνητικό ενδιαφέρον παρουσιάζει η δυνατότητα δημιουργίας νέων εικόνων, παρόμοιων με αυτές με τις οποίες εκπαιδεύτηκαν, αλλά βελτιωμένες ως προς κάποιο συγκεκριμένο χαρακτηριστικό. Για παράδειγμα στο συγκεκριμένο σετ δεδομένων που εργαστήκαμε, τα συνθετικά προσθετικά μέρη που είχαν πολλοί ασθενείς στα δόντια τους, δημιουργούσαν αντανάκλασεις των ακτινών με αποτέλεσμα οι προκύπτουσες ακτινογραφίες να είναι θορυβώδεις στις συγκεκριμένες περιοχές. Η δημιουργία ενός νευρωνικού δικτύου που θα μπορούσε να ανακατασκευάζει τις συγκεκριμένες εικόνες, μαθαίνοντας όμως στα συγκεκριμένα σημεία των ανακλάσεων να δημιουργεί τον ιστό που κανονικά θα φαινόταν, θα είχε ιδιαίτερο ερευνητικό ενδιαφέρον, και πιθανότατα θα μπορούσε να υλοποιηθεί με τη χρήση ενός Denoising Autoencoder.

Βιβλιογραφία

- Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL <http://tensorflow.org/>. Software available from tensorflow.org.
- Amir Arsalan Soltani, Haibin Huang, Jiajun Wu, Tejas D. Kulkarni, and Joshua B. Tenenbaum. Synthesizing 3d shapes via modeling multi-view depth maps and silhouettes with deep generative networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.
- Daniel I. Barnea and Harvey F. Silverman. A class of algorithms for fast digital image registration. *IEEE Transactions on Computers*, C-21(2):179–186, 1972. doi: 10.1109/TC.1972.5008923.
- Ekaba Bisong. *Google Colaboratory*, pages 59–64. Apress, Berkeley, CA, 2019. ISBN 978-1-4842-4470-8. doi: 10.1007/978-1-4842-4470-8_7. URL https://doi.org/10.1007/978-1-4842-4470-8_7.
- François Chollet et al. Keras. <https://github.com/fchollet/keras>, 2015.
- Christopher B. Choy, Danfei Xu, JunYoung Gwak, Kevin Chen, and Silvio Savarese. 3d-r2n2: A unified approach for single and multi-view 3d object reconstruction, 2016.
- A. P. CRACKNELL. A review of: “techniques for image processing and classification in remote sensing”. by r. a. schowengerdt. (new york: Academic press, 1983.) [pp. 249.] price \$25·00. *International Journal of Remote Sensing*, 6(2):3678–3678, 1985. doi: 10.1080/01431168508948454. URL <https://doi.org/10.1080/01431168508948454>.
- MICHAEL A. CROMBIE. Coordination of stereo image registration and pixel classification. *American Society of Photogrammetry*, 49:529–532, 1983.

T S Curry, J E Dowdey, and R C Murray, Jr. Christensen's physics of diagnostic radiology. 4th edition. URL <https://www.osti.gov/biblio/5443052>.

Jeff Donahue, Lisa Anne Hendricks, Marcus Rohrbach, Subhashini Venugopalan, Sergio Guadarrama, Kate Saenko, and Trevor Darrell. Long-term recurrent convolutional networks for visual recognition and description, 2016.

Ganggang Dong, Guisheng Liao, Hongwei Liu, and Gangyao Kuang. A review of the autoencoder and its variants: A comparative perspective from target recognition in synthetic-aperture radar images. *IEEE Geoscience and Remote Sensing Magazine*, 6:44–68, 09 2018. doi: 10.1109/MGRS.2018.2853555.

Matsopoulos GK Molnár B Windisch P. Economopoulos TL, Asvestas PA. Volumetric difference evaluation of registered three-dimensional pre-operative and post-operative ct dental data. pages 328–339, 2012. doi: 10.1259/dmfr/94040044.

Matheus Gadelha, Rui Wang, and Subhransu Maji. Multiresolution tree networks for 3d point cloud processing, 2018.

Isherwood I. Gholkar A. Three-dimensional computed tomographic reformations of intracranial vascular lesions. pages 258–261, 03 1988. doi: 10.1259/0007-1285-61-723-258.

I. J. Good and Eric P. Smith. C184.the possible bias of the pearson chi-squared test in non-equiprobable cases. *Journal of Statistical Computation and Simulation*, 19(1):79–82, 1984. doi: 10.1080/00949658408810714. URL <https://doi.org/10.1080/00949658408810714>.

Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.

Edward Grant, Pushmeet Kohli, and Marcel van Gerven. Deep disentangled representations for volumetric reconstruction, 2016.

Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2 edition, 2004. doi: 10.1017/CBO9780511811685.

Mattias P. Heinrich, Max Blendowski, and Ozan Oktay. Ternarynet: faster deep model inference without gpu for medical 3d segmentation using sparse and binary convolutions. *International Journal of Computer Assisted Radiology and Surgery*, 13(9):1311–1320, May 2018. ISSN 1861-6429. doi: 10.1007/s11548-018-1797-4. URL <http://dx.doi.org/10.1007/s11548-018-1797-4>.

Paul Henderson and Vittorio Ferrari. Learning to generate and reconstruct 3d meshes with only 2d supervision, 2018.

G. E. Hinton and R. R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, 2006. ISSN 0036-8075. doi: 10.1126/science.1127647. URL <https://science.sciencemag.org/content/313/5786/504>.

Geoffrey E. Hinton, Alex Krizhevsky, and Sida D. Wang. Transforming auto-encoders. In Timo Honkela, Włodzisław Duch, Mark Girolami, and Samuel Kaski, editors, *Artificial Neural Networks*

- and Machine Learning – ICANN 2011*, pages 44–51, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg. ISBN 978-3-642-21735-7.
- Q. Huynh-Thu. Scope of validity of psnr in image/video quality assessment. *Electronics Letters*, 44:800–801(1), June 2008. ISSN 0013-5194. URL https://digital-library.theiet.org/content/journals/10.1049/el_20080522.
- Q. Huynh-Thu and M. Ghanbari. The accuracy of psnr in predicting video quality for different video scenes and frame rates. *Telecommunication Systems*, 49:35–48, 2012.
- Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37, ICML'15*, page 448–456. JMLR.org, 2015.
- Michael R.M. Jenkin, Allen D. Jepson, and John K. Tsotsos. Techniques for disparity measurement. *CVGIP: Image Understanding*, 53(1):14–30, 1991. ISSN 1049-9660. doi: [https://doi.org/10.1016/1049-9660\(91\)90002-7](https://doi.org/10.1016/1049-9660(91)90002-7). URL <https://www.sciencedirect.com/science/article/pii/1049966091900027>.
- Adrian Johnston, Ravi Garg, Gustavo Carneiro, Ian Reid, and Anton van den Hengel. Scaling cnns for high resolution volumetric reconstruction from a single image. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV) Workshops*, Oct 2017.
- Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017.
- A. Laurentini. The visual hull concept for silhouette-based image understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(2):150–162, 1994. doi: 10.1109/34.273735.
- Adrian Lecoutre, Benjamin Négrevergne, and Florian Yger. Recognizing art style automatically in painting with deep learning. In *ACML*, 2017.
- Po-Yu Liu and Edmund Y. Lam. Image reconstruction using deep learning, 2018.
- Shikun Liu, C. Lee Giles, and Alexander G. Ororbia II au2. Learning a hierarchical latent-variable model of 3d shapes, 2018.
- Priyanka Mandikal, K L Navaneet, Mayank Agarwal, and R. Venkatesh Babu. 3d-lmnet: Latent embedding matching for accurate and diverse 3d point cloud reconstruction from a single image, 2019.
- John McCarthy. What is artificial intelligence? 01 2004.
- Michael A. Nielsen. *Neural Networks and Deep Learning*. 2015. URL <http://neuralnetworksanddeeplearning.com/>.
- Mohammad Norouzi and David J. Fleet. Minimal loss hashing for compact binary codes. In *Proceedings of the 28th International Conference on International Conference on Machine Learning, ICML'11*, page 353–360, Madison, WI, USA, 2011. Omnipress. ISBN 9781450306195.

- K. O’Shea and Ryan Nash. An introduction to convolutional neural networks. *ArXiv*, abs/1511.08458, 2015.
- Bo Peng, Wei Wang, Jing Dong, and Tieniu Tan. Learning pose-invariant 3d object reconstruction from single-view images. *Neurocomputing*, 423:407–418, 2021. ISSN 0925-2312. doi: <https://doi.org/10.1016/j.neucom.2020.10.089>. URL <https://www.sciencedirect.com/science/article/pii/S0925231220317021>.
- Joseph Lee Rodgers and W. Alan Nicewander. Thirteen ways to look at the correlation coefficient. *The American Statistician*, 42(1):59–66, 1988. ISSN 00031305. URL <http://www.jstor.org/stable/2685263>.
- Swalpa Kumar Roy, Gopal Krishna, Shiv Ram Dubey, and Bidyut B. Chaudhuri. Hybridsn: Exploring 3-d-2-d cnn feature hierarchy for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 17(2):277–281, 2020. doi: 10.1109/LGRS.2019.2918719.
- D. E. Rumelhart, G. E. Hinton, and R. J. Williams. *Learning Internal Representations by Error Propagation*, page 318–362. MIT Press, Cambridge, MA, USA, 1986. ISBN 026268053X.
- Ruslan Salakhutdinov and Geoff Hinton. Learning a nonlinear embedding by preserving class neighbourhood structure. In Marina Meila and Xiaotong Shen, editors, *Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics*, volume 2 of *Proceedings of Machine Learning Research*, pages 412–419, San Juan, Puerto Rico, 21–24 Mar 2007. PMLR. URL <http://proceedings.mlr.press/v2/salakhutdinov07a.html>.
- Ruslan Salakhutdinov and Geoffrey Hinton. Deep boltzmann machines. In David van Dyk and Max Welling, editors, *Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics*, volume 5 of *Proceedings of Machine Learning Research*, pages 448–455, Hilton Clearwater Beach Resort, Clearwater Beach, Florida USA, 16–18 Apr 2009. PMLR. URL <http://proceedings.mlr.press/v5/salakhutdinov09a.html>.
- William C. Scarfe and Allan G. Farman. What is cone-beam ct and how does it work? *Dental Clinics of North America*, 52(4):707–730, 2008. ISSN 0011-8532. doi: <https://doi.org/10.1016/j.cden.2008.05.005>. URL <https://www.sciencedirect.com/science/article/pii/S001185320800044X>. Contemporary Dental and Maxillofacial Imaging.
- Phil Simon. *Too Big to Ignore: The Business Case for Big Data*. December 2015. ISBN 978-1-119-21784-8. URL https://el.wikipedia.org/wiki/%CE%9C%CE%B7%CF%87%CE%B1%CE%BD%CE%B9%CE%BA%CE%AE_%CE%BC%CE%AC%CE%B8%CE%B7%CF%83%CE%B7.
- Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(56):1929–1958, 2014. URL <http://jmlr.org/papers/v15/srivastava14a.html>.
- Xingyuan Sun, Jiajun Wu, Xiuming Zhang, Zhoutong Zhang, Chengkai Zhang, Tianfan Xue, Joshua B. Tenenbaum, and William T. Freeman. Pix3d: Dataset and methods for single-image 3d shape modeling, 2018.

- Chaudhari M, Talwar I, Jankharia B, Hemrajani A. Three-dimensional ct- principles and techniques. pages 85–96, 1992.
- Qingyang Tan, Lin Gao, Yu-Kun Lai, Jie Yang, and Shihong Xia. Mesh-based autoencoders for localized deformation component analysis, 2017.
- Antonio Torralba, Rob Fergus, and William T. Freeman. 80 million tiny images: A large data set for nonparametric object and scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(11):1958–1970, 2008. doi: 10.1109/TPAMI.2008.128.
- Shubham Tulsiani, Tinghui Zhou, Alexei A. Efros, and Jitendra Malik. Multi-view supervision for single-view reconstruction via differentiable ray consistency, 2017.
- Shubham Tulsiani, Alexei A. Efros, and Jitendra Malik. Multi-view consistency as supervisory signal for learning shape and pose prediction, 2018.
- Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th International Conference on Machine Learning, ICML '08*, page 1096–1103, New York, NY, USA, 2008. Association for Computing Machinery. ISBN 9781605582054. doi: 10.1145/1390156.1390294. URL <https://doi.org/10.1145/1390156.1390294>.
- Pascal Vincent, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, and Pierre-Antoine Manzagol. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *J. Mach. Learn. Res.*, 11:3371–3408, December 2010. ISSN 1532-4435.
- Jason M. Walker, E. Harris, Charles Lynagh, Andrea Beck, Rich Lonardo, Brian Vuksanovich, J. Thiel, Kirk Rogers, B. Conner, and E. MacDonald. 3d printed smart molds for sand casting. *International Journal of Metalcasting*, 12:785–796, 2018.
- Weiyue Wang, Qianguai Huang, Suya You, Chao Yang, and Ulrich Neumann. Shape inpainting using 3d generative adversarial network and recurrent convolutional networks, 2017.
- Olivia Wiles and Andrew Zisserman. Silnet : Single- and multi-view reconstruction by learning from silhouettes, 2017.
- Jiajun Wu, Chengkai Zhang, Tianfan Xue, William T. Freeman, and Joshua B. Tenenbaum. Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling. In *Proceedings of the 30th International Conference on Neural Information Processing Systems, NIPS'16*, page 82–90, Red Hook, NY, USA, 2016. Curran Associates Inc. ISBN 9781510838819.
- Jiajun Wu, Yifan Wang, Tianfan Xue, Xingyuan Sun, William T. Freeman, and Joshua B. Tenenbaum. Marrnet: 3d shape reconstruction via 2.5d sketches. In *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17*, page 540–550, Red Hook, NY, USA, 2017. Curran Associates Inc. ISBN 9781510860964.
- Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets: A deep representation for volumetric shapes, 2015.

Xinchen Yan, Jimei Yang, Ersin Yumer, Yijie Guo, and Honglak Lee. Perspective transformer nets: Learning single-view 3d object reconstruction without 3d supervision. In *NIPS*, 2016.

Eugene K. Yen and CPP Roger G. Johnston, Ph.D. The ineffectiveness of the correlation coefficient for image comparisons. URL <http://lib-www.lanl.gov/la-pubs/00418797.pdf>.