



Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Τομέας Τεχνολογίας, Πληροφορικής &
Υπολογιστών

Αναγνώριση Τεχνοτροπίας Έργων Τέχνης με Συνελικτικά Νευρωνικά Δίκτυα

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΟΡΦΕΑΣ ΜΕΝΗΣ - ΜΑΣΤΡΟΜΙΧΑΛΑΚΗΣ

Επιβλέπων : Γεώργιος Στάμου
Αν. Καθηγητής Ε.Μ.Π.

Αθήνα, Νοέμβριος 2019



Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Τομέας Τεχνολογίας, Πληροφορικής &
Υπολογιστών

Αναγνώριση Τεχνοτροπίας Έργων Τέχνης με Συνελικτικά Νευρωνικά Δίκτυα

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΟΡΦΕΑΣ ΜΕΝΗΣ - ΜΑΣΤΡΟΜΙΧΑΛΑΚΗΣ

Επιβλέπων : Γεώργιος Στάμου
Αν. Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 12η Νοεμβρίου 2019.

.....
Γεώργιος Στάμου
Αν. Καθηγητής Ε.Μ.Π.

.....
Νικόλαος Παπασπύρου
Καθηγητής Ε.Μ.Π.

.....
Ανδρέας-Γεώργιος Σταφυλοπάτης
Καθηγητής Ε.Μ.Π.

Αθήνα, Νοέμβριος 2019

.....
Ορφέας Μενής - Μαστρομιχαλάκης

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

Copyright © Ορφέας Μενής - Μαστρομιχαλάκης, 2019.
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 ψηφιοποίηση του κόσμου που λαμβάνει χώρα τις τελευταίες δεκαετίες, έχει ως αποτέλεσμα και την ψηφιοποίηση ενός τεράστιου αριθμού έργων τέχνης. Η διαχείρισή κι επεξεργασία ενός τόσο μεγάλου όγκου πληροφοριών αποτελεί μια ιδιαίτερα απαιτητική πρόκληση. Τα τελευταία χρόνια έχουν γίνει βήματα για τη χρήση της τεχνητής νοημοσύνης στην υπηρεσία της τέχνης. Τα δείγματα ήταν πολύ ελπιδοφόρα και σύντομα ξεκίνησε μια τάση στους κλάδους της μηχανικής μάθησης που αφορούσε την αναγνώριση, ταξινόμηση, επεξεργασία αλλά ακόμη και δημιουργία έργων τέχνης. Τα άλματα που έχουν γίνει την τελευταία δεκαετία στην αναγνώριση εικόνας μέσω της βαθιάς μάθησης, άνοιξαν το δρόμο για την ευρύτερη μελέτη της τέχνης μέσω του συγκεκριμένου κλάδου. Σε συνέχεια των παραπάνω, αποφασίσαμε στην παρούσα διπλωματική να ασχοληθούμε με την αναγνώριση τεχνοτροπίας έργων τέχνης με χρήση συνελκτικκών νευρικών δικτύων.

Η καλλιτεχνική τεχνοτροπία ανέκαθεν αποτελούσε το κύριο κριτήριο ομαδοποίησης και ταξινόμησης των έργων τέχνης. Μέχρι και στις μέρες μας όμως, κατά κύριο λόγο το έργο αυτό επιτελείται από ανθρώπους με χρόνια εκπαίδευσης, εμπειρίας κι εξειδίκευσης στον κλάδο, όπως κριτικοί και ιστορικοί τέχνης. Τα τελευταία χρόνια έχουν γίνει πολύ αξιόλογες προσπάθειες στην ταξινόμηση έργων τέχνης, αλλά ακόμη υπάρχουν πολλά κομμάτια που έχουν μείνει ανεξερεύνητα, αφήνοντας χώρο για σημαντικές βελτιώσεις. Η δουλειά μας επικεντρώθηκε στην κατανόηση και ανάλυση των κύριων παραμέτρων που επηρεάζουν την εκπαίδευση και την αποτελεσματικότητα ενός συνελκτικού νευρωνικού δικτύου, ενώ παράλληλα επιχειρήσαμε και επιτύχαμε να δημιουργήσουμε μοντέλα βαθιάς μάθησης τα οποία ξεπέρασαν σε επιδόσεις σχεδόν όλα όσα έχουν δημιουργηθεί μέχρι στιγμής. Πιο συγκεκριμένα συλλέξαμε, μελετήσαμε κι αξιολογήσαμε διαφορετικά σύνολα δεδομένων, εκτελέσαμε πάνω από εκατό πειράματα, εκπαιδεύοντας μοντέλα με διαφορετικές αρχιτεκτονικές και υπερπαραμέτρους, κι εφαρμόσαμε μεθόδους συνδυαστικής μάθησης στα μοντέλα που εκπαιδεύσαμε, επιτυγχάνοντας κορυφαίες επιδόσεις. Αναλύσαμε βαθύτερα τα χαρακτηριστικά κάθε τεχνοτροπίας και τον τρόπο όπου αυτά επηρεάζουν τις προβλέψεις ενός ταξινομητή και τέλος με βάση την εμπειρία και τις γνώσεις που αποκτήσαμε προτείναμε μερικές ενδιαφέρουσες μελλοντικές κατευθύνσεις της επιστημονικής μελέτης.

Λέξεις κλειδιά

Συνελκτικά Νευρωνικά Δίκτυα, Τέχνη, Καλλιτεχνική Τεχνοτροπία, Αναγνώριση Τεχνοτροπίας, Ταξινόμηση Δεδομένων, Μεταφορά Γνώσης, Συνδυαστική Μάθηση, Υπερμοντελλοποίηση

Abstract

The rapid digitization of our world that takes place the last decades, resulted in the digitization of a huge amount of artworks. The administration and processing of such a huge amount of data compound a very challenging problem. The last years, artificial intelligence has made important steps in putting science in the service of art. The results were promising and soon a tendency in the field of machine learning began, concerning recognition, classification, process and even creation of artworks. The tremendous progress that has been achieved the last decade in image recognition through deep learning, gave the opportunity for scientists to study art through that scientific field. Due to all those things stated above, we decided to study the recognition of artistic style in artworks with the use of convolutional neural networks.

Artistic style has always been the main attribute according to which art classification is performed. Even in the recent days, this task is mainly done by art experts with years of studies and experience. The last years notable results have been achieved in artwork recognition tasks, but there's still a long way to go. Our work focused on the understanding and analyzing of the main parameters that affect the training and the effectiveness of a CNN, while we achieved state of the art results. We collected, studied and evaluated different datasets, we performed more than 100 experiments using models with different architectures and hyperparameters, and we used ensemble learning which lead us to top results. We elaborated in the deeper characteristics of artistic styles and the way that they affect the prediction of a classifier. Last but not least, based on the experience and the knowledge that we gained, we suggested some interesting topics for future work.

Key words

Convolutional Neural Networks, CNN, Art, Artistic Style, Artistic Style Recognition, Data Classification, Transfer Learning, Ensemble Learning, Overfitting

Ευχαριστίες

Θα ήθελα να ευχαριστήσω ιδιαίτερα τον επιβλέποντα καθηγητή μου Δρ. Γεώργιο Στάμου, ο οποίος μου έδωσε τη δυνατότητα να ασχοληθώ και να εμβαθύνω στο αντικείμενο της Τεχνητής Νοημοσύνης μέσω αυτής της διπλωματικής αλλά και τη Δρ. Αναστασία Σοφού η οποία με καθοδήγησε καθ' όλη τη διάρκεια εκπόνησης της διπλωματικής μου εργασίας. Επίσης θα ήθελα να ευχαριστήσω το κοινωφελές ίδρυμα Κωνσταντίνου Τσαπάλα για την οικονομική στήριξη που μου παρείχε κατά τη διάρκεια των σπουδών μου. Θα ήθελα ακόμη να ευχαριστήσω ιδιαίτερα την οικογένεια μου, που έδειξε και δείχνει αμέριστη στήριξη και αγάπη. Τέλος, θα ήθελα να ευχαριστήσω τους φίλους μου και ιδιαίτερα την Ε.Μ. χωρίς τους οποίους δεν θα μπορούσα σε καμία περίπτωση να φτάσω σε αυτό το σημείο που βρίσκομαι σήμερα.

Ορφέας Μενής - Μαστρομιχαλάκης,

Αθήνα, 12η Νοεμβρίου 2019

Περιεχόμενα

Περίληψη	5
Abstract	7
Ευχαριστίες	9
Περιεχόμενα	11
Κατάλογος πινάκων	13
Κατάλογος σχημάτων	15
1. Εισαγωγή	17
Εισαγωγή	17
1.1 Το Πρόβλημα της Αναγνώρισης Τεχνοτροπίας	17
1.2 Κίνητρο	18
1.3 Συνεισφορά Διπλωματικής Εργασίας	18
1.4 Δομή Διπλωματικής Εργασίας	18
2. Θεωρητικό Υπόβαθρο	21
Θεωρητικό Υπόβαθρο	21
2.1 Αναγνώριση Τεχνοτροπίας	21
2.1.1 Αναγνώριση Τεχνοτροπίας και Ιστορική Αναδρομή	21
2.1.2 Παραλλαγές Προβλήματος	24
2.2 Μηχανική Μάθηση	28
2.2.1 Ορισμός Μηχανικής Μάθησης	28
2.2.2 Βαθιά Μάθηση	28
2.2.3 Βασικά ήδη μηχανικής μάθησης	28
2.2.4 Τεχνητά Νευρωνικά Δίκτυα	29
2.2.5 Συνελκτικά Νευρωνικά Δίκτυα	31
2.2.6 Μεταφορά Γνώσης - Transfer Learning	34
3. Σχεδιασμός και Υλοποίηση	37
Σχεδιασμός και Υλοποίηση	37
3.1 Δεδομένα	37
3.1.1 Σύνολα Δεδομένων - Datasets	37
3.1.2 Προ-επεξεργασία Δεδομένων	44
3.2 Συνελκτικό Νευρωνικό Δίκτυο	45
3.2.1 Επιλογή Δεδομένων Εισόδου	45
3.2.2 Επιλογή Αρχιτεκτονικής	47
3.2.3 Επιλογή και Βελτιστοποίηση Υπερπαραμέτρων	48
3.2.4 Αντιμέτωπιση Overfitting	49
3.2.5 Έξοδος προγράμματος και μετρικές αξιολόγησης	50

3.2.6	Αποθήκευση Μοντέλων	51
3.2.7	Συνδυαστική Μάθηση - Ensemble Learning	51
3.3	Περιβάλλον Διεξαγωγής Πειραμάτων	51
4.	Αξιολόγηση Αποτελεσμάτων και Συγκρίσεις	53
	Αξιολόγηση Αποτελεσμάτων και Συγκρίσεις	53
4.1	Συγκριτικά αποτελέσματα μοντέλων βαθιάς μάθησης	53
4.1.1	Αρχιτεκτονικές	53
4.1.2	Data Augmentation	55
4.1.3	Learning Rate	60
4.1.4	Transfer Learning	61
4.1.5	Datasets	64
4.1.6	Περικοπή και Προσαρμογή Μεγέθους	65
4.1.7	Ensemble Learning	67
4.2	Προσωπικές Παρατηρήσεις - Ποιοτική Αξιολόγηση	69
4.2.1	Μη ισορροπημένο σύνολο δεδομένων	69
4.2.2	Παρεμφερείς Κλάσεις	71
4.2.3	Ιδιαίτερες Κλάσεις και Ιδιαίτερα Χαρακτηριστικά	72
4.2.4	Παραδείγματα με μεμονωμένες προβλέψεις	75
4.3	Σύγκριση Αποτελεσμάτων	77
4.3.1	Pandora18k	77
4.3.2	Wikiart	78
5.	Επίλογος	81
	Επίλογος	81
5.1	Σύνοψη	81
5.2	Τελικά Συμπεράσματα	82
5.3	Μελλοντικές Κατευθύνσεις Επιστημονικής Μελέτης	83
	Βιβλιογραφία	85

Κατάλογος πινάκων

3.1	Το Σύνολο Δεδομένων Pandora18k	38
3.2	Το Σύνολο Δεδομένων WikiArt	42
3.3	Σύνοψη Συνόλων Δεδομένων	44
4.1	Αποτελέσματα διαφόρων αρχιτεκτονικών	53
4.2	Επίπεδα Data Augmentation	55
4.3	InceptionV3 - Data Augmentation	56
4.4	DenseNet121 - Data Augmentation	56
4.5	Με και Χωρίς Transfer Learning	62
4.6	Transfer Learning	63
4.7	Datasets Results	64
4.8	Resizing - Cropping	65
4.9	Transfer from/to crop	66
4.10	Αξιολόγηση σε δεδομένα διαφορετικής προ-επεξεργασίας	66
4.11	Συνδυαστικές Μέθοδοι	68
4.12	Προβλέψεις Μεμονωμένων Εικόνων	76
4.13	Συγκριτικά Αποτελέσματα Pandora18k	77
4.14	Συγκριτικά Αποτελέσματα Wikiart	79

Κατάλογος σχημάτων

2.1	"Impression, soleil levant", Claude Monet	22
2.2	Σχηματική απεικόνιση της μεθοδολογίας που ακολουθήθηκε από τους Rodriguez et al. (2018)	24
2.3	"The altarboy", Pablo Picasso, 1896, Realism	26
2.4	"Woman in a shawl", Pablo Picasso, 1902, Expressionism	26
2.5	"Girl with mandolin", Pablo Picasso, 1910, Cubism	26
2.6	Παραδείγματα Style Transfer από την εργασία των Gatys et al. (2015)	27
2.7	"Portrait of Edmond Belamy", ο πρώτος πίνακας, δημιούργημα τεχνητής νοημοσύνης, που πωλήθηκε ποτέ.	27
2.8	Ένα Residual Block (Rodriguez et al., 2018)	34
3.1	Κατανομή εικόνων ανά κατηγορία στο σύνολο δεδομένων Pandora18k	39
3.2	"Lady with an Ermine", Leonardo da Vinci, 1490, High Renaissance	40
3.3	"The Scream", Edvard Munch, 1893, Expressionism	40
3.4	"Still life with skull of ox", Pablo Picasso, 1942, Post-Impressionism	40
3.5	"The Persistence of Memory", Salvador Dali, 1931, Surrealism	40
3.6	"The Starry Night", Vincent van Gogh, 1889, Post-Impressionism	40
3.7	"Judith beheading Holoferne", Caravaggio, 1599, Baroque	40
3.8	"L'air", Joan Miro, 1937, Abstract Art	41
3.9	"Cow", Andy Warhol, 1966, Pop Art	41
3.10	"Self-Portrait in a Striped T-shirt", Henri Matisse, 1906, Fauvism	41
3.11	Κατανομή εικόνων ανά κατηγορία στο σύνολο δεδομένων Wikiart	43
3.12	Overfitting	49
4.1	1° Επίπεδο Data Augmentation	56
4.2	2° Επίπεδο Data Augmentation	56
4.3	3° Επίπεδο Data Augmentation	56
4.4	DenseNet121 Train and Validation Accuracy	57
4.5	InceptionV3 Train and Validation Accuracy	57
4.6	Σύγκριση μοντέλων με και χωρίς data augmentation	59
4.7	High Learning Rate Confusion Matrix	61
4.8	Low Learning Rate Confusion Matrix	61
4.9	Πίνακας σύγχυσης μη ισορροπημένων δεδομένων	70
4.10	Πίνακας σύγχυσης - Παρεμφερείς Κλάσεις	72
4.11	"Amanohashidate in Tango Province", Utagawa Hiroshige	73
4.12	"Chrysanthemums", Utagawa Hiroshige	73
4.13	"Itsukushima in Aki Province", Utagawa Hiroshige	73
4.14	"Blue Marilyn", Andy Warhol	74
4.15	"In the car", Roy Lichtenstein	74
4.16	"Red Lenin", Andy Warhol	74
4.17	Βυζαντινή Εικονογραφία 1	74
4.18	Βυζαντινή Εικονογραφία 2	74
4.19	"Guernica", Pablo Picasso, 1937, Cubism and Surrealism	75

4.20	"Self-portrait", Vincent Van Gogh, 1889 , Post-Impressionism	75
4.21	"Melting Watch", Salvador Dali , 1954, Surrealism	75
4.22	"Mona Lisa", Leonardo da Vinci, 1503, High Renaissance	76
4.23	"Man with knife", Jackson Pollock, 1940 , Abstract Expressionism	76
4.24	"Portrait of AG", Παναγιώτης Τέτσης , 1954, Post-Impressionism	76

Κεφάλαιο 1

Εισαγωγή

Τα τελευταία χρόνια η ψηφιοποίηση πολλών έργων τέχνης έχει ανοίξει το δρόμο για την εισαγωγή των υπολογιστών σε έναν τομέα που ωρίτερα δεν είχαν ιδιαίτερη ανάμειξη, την τέχνη. Εδώ και δεκαετίες αν και οι υπολογιστές είχαν μπει σε όλες σχεδόν τις πτυχές της ζωής μας, η τέχνη παρέμενε προνόμιο των ανθρώπων. Η πολυπλοκότητα καθώς και οι αφηρημένες έννοιες που συνοδεύουν την τέχνη λειτουργούσαν αποτρεπτικά στην ενασχόληση των υπολογιστών, καθαρά "λογικών" και "γραμμικών" οντοτήτων, με αυτή. Η ψηφιοποίηση του κόσμου μας όμως έφερε και την ψηφιοποίηση στην τέχνη, και τότε είδαμε τις προοπτικές που υπάρχουν μέσω αυτής. Από την αποθήκευση και επεξεργασία μέχρι την ταξινόμηση και δημιουργία, οι δυνατότητες είναι απεριόριστες. Λόγω τόσο της δυσκολίας όσο και του όγκου της πληροφορίας που συνοδεύει την τέχνη, η τεχνητή νοημοσύνη αποτελεί ένα πολύ χρήσιμο εργαλείο για τη διαχείρισή τους. Μέσω της βαθιάς μάθησης έχουμε καταφέρει να αναγνωρίζουμε εικόνες όμοια, και πολλές φορές καλύτερα από τους ανθρώπους. Στην παρούσα διπλωματική εργασία πραγματευόμαστε την αναγνώριση της τεχνοτροπίας έργων τέχνης με τη χρήσης βαθιάς μηχανικής μάθησης, και συγκεκριμένα συνελκτικών νευρωνικών δικτύων. Θα επιχειρήσουμε να μπούμε στη θέση ενός ιστορικού ή κριτικού τέχνης και να ταξινομήσουμε ένα σύνολο από έργα τέχνης ανάλογα με την τεχνοτροπία τους. Την τελευταία δεκαετία ο κόσμος της πληροφορικής έχει αρχίσει να δείχνει ολοένα και περισσότερο ενδιαφέρον στη χρήση των υπολογιστών στις υπηρεσίες της τέχνης.

1.1 Το Πρόβλημα της Αναγνώρισης Τεχνοτροπίας

Ο προσδιορισμός της τεχνοτροπίας ενός έργου τέχνης αποτελούσε για χρόνια έργο που μπορούσαν να φέρουν εις πέρας μόνο μια μικρή μερίδα ανθρώπων με εκπαίδευση κι εμπειρία στο αντικείμενο, όπως κριτικοί και ιστορικοί τέχνης. Τα τελευταία χρόνια όμως η τεχνητή νοημοσύνη έχει καταφέρει να επιδείξει αξιόλογα αποτελέσματα στο συγκεκριμένο πρόβλημα, όμως ακόμη υπάρχουν μεγάλα περιθώρια βελτίωσης. Η τεχνοτροπία ενός έργου τέχνης καθορίζεται κυρίως από την οπτική του εμφάνιση και πηγάζει από της επιλογές του δημιουργού του. Οι γραμμές ανάμεσα στις τεχνοτροπίες σε πολλές περιπτώσεις είναι θολές με πολλά έργα τέχνης να έχουν στοιχεία από διαφορετικές τεχνοτροπίες και πολλές τεχνοτροπίες να μοιράζονται κοινά χαρακτηριστικά μεταξύ τους. Η αναγνώριση της τεχνοτροπίας ενός καλλιτεχνήματος μπορεί να δώσει πολλές πληροφορίες για τις επιρροές του καλλιτέχνη καθώς και για τη σημασία του ίδιου του έργου. Αποτελεί το βασικό κριτήριο ταξινόμησης κι ομαδοποίησης των έργων τέχνης με ευρεία εφαρμογή στον κόσμο των εικαστικών τεχνών.

1.2 Κίνητρο

Η αγάπη για την τέχνη συνδυάστηκε με το ενδιαφέρον για τη βαθιά μάθηση και τα νευρωνικά δίκτυα και δημιούργησαν τη θεματολογία της διπλωματικής εργασίας. Η τεχνοτροπία έργων τέχνης αποτελεί αντικείμενο ενδιαφέροντος για εκατομμύρια ανθρώπους και βοηθάει το θεατή να κατανοήσει καλύτερα τα μηνύματα που θέλει να περάσει ο καλλιτέχνης, καθώς και να κάνει συσχετίσεις και να οδηγηθεί σε νέα συμπεράσματα. Από την άλλη ο τομέας της τεχνητής νοημοσύνης αποτελεί ένα πολύ ενδιαφέρον μέρος της επιστήμης των υπολογιστών και 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. Αυτή η περιέργεια κι ο ενθουσιασμός για το σχετικά άγνωστο κόσμο της τεχνητής νοημοσύνης συνδυάστηκαν με το ενδιαφέρον και την αγάπη για την τέχνη, και μας οδήγησαν στην απόφαση να ασχοληθούμε με τη συγκεκριμένη θεματολογία. Ένας ακόμη παράγοντας που συνέβαλε στην επιλογή μας είναι η δυσκολία του συγκεκριμένου εγχειρήματος. Η αναγνώριση εικόνας αποτελεί ένα σχετικά τετριμμένο πρόβλημα από την άποψη πως στη συντριπτική πλειοψηφία των περιπτώσεων κάτι ή είναι ή δεν είναι (π.χ. ένα αντικείμενο σε μια εικόνα ή είναι ή δεν είναι αυτοκίνητο, άλογο ή οτιδήποτε άλλο). Στο κομμάτι της τέχνης και της τεχνοτροπίας όμως αυτό δεν ισχύει. Απαιτείται βαθύτερη κατανόηση και σκέψη των οπτικών, κι όχι μόνο, ερεθισμάτων για να καταλήξεις σε κάποιο συμπέρασμα και να αξιολογήσεις αυτό που βλέπεις (ή ακούς, αγγίζεις κλπ.).

1.3 Συνεισφορά Διπλωματικής Εργασίας

Στην παρούσα διπλωματική εργασία πραγματευόμαστε την αναγνώριση τεχνοτροπίας έργων τέχνης και μέσω αυτού του προβλήματος μελετάμε κι αναλύουμε τις επιρροές διαφόρων παραμέτρων στις επιδόσεις των μοντέλων βαθιάς μάθησης. Μελετήσαμε αναλυτικά μέσω πειραμάτων την επίδραση των παραμέτρων του περιβάλλοντος αλλά και των υπερπαραμέτρων του μοντέλου στην εκπαίδευση και την αποτελεσματικότητα των δικτύων μας και παρουσιάσαμε μια πλήρη ανάλυση επί του θέματος. Παράλληλα καταφέραμε να επιτύχουμε κορυφαίες επιδόσεις στην αναγνώριση τεχνοτροπίας, καταφέροντας μάλιστα την καλύτερη ακρίβεια που έχει δημοσιευθεί μέχρι στιγμής στο ένα σύνολο δεδομένων, και τη δεύτερη καλύτερη στο άλλο. Επίσης εκτελέσαμε πειράματα με αρχιτεκτονικές οι οποίες δεν είχαν χρησιμοποιηθεί ξανά στο συγκεκριμένο πρόβλημα επιτυγχάνοντας πολύ καλά κι ελπιδοφόρα αποτελέσματα.

1.4 Δομή Διπλωματικής Εργασίας

Στο Κεφάλαιο 2 καλύπτεται το θεωρητικό υπόβαθρο όσων κρίθηκαν απαραίτητα για την κατανόηση αυτής της διπλωματικής εργασίας. Αρχικά γίνεται αναφορά στο πρόβλημα της αναγνώρισης τεχνοτροπίας αναλύοντας βασικές έννοιες, προβλήματα και ορολογίες. Στη συνέχεια γίνεται μια παρουσίαση του related work, δηλαδή των σημαντικότερων εργασιών που έχουν γίνει μέχρι σήμερα πάνω στο πρόβλημα που καλούμαστε να λύσουμε. Ακολουθεί μια αναφορά σε παραλλαγές του προβλήματός μας με την οποία κλείνει το κομμάτι θεωρίας που αφορά την τέχνη. Η δεύτερη υπο-ενότητα του κεφαλαίου αφορά την απαραίτητη θεωρία σχετικά με τον τομέα της μηχανικής μάθησης. Από τη γενική περιγραφή του τομέα, μέχρι τη βαθιά μάθηση, τα συνελκτικά

νευρωνικά δίκτυα και το transfer learning, όλα όσα θα χρειαστούν για να μπορεί ο αναγνώστης να ακολουθήσει τη ροή της εργασίας βρίσκονται εκεί.

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 και τα χαρακτηριστικά τους. Αναπτύσσονται διλήμματα που προέκυψαν στην πορεία και η λογική επίλυσή τους, και τέλος γίνεται μια αναφορά στο περιβάλλον εκτέλεσης των πειραμάτων.

Στο Κεφάλαιο 4 παρουσιάζονται τα αποτελέσματα των πειραμάτων κι εξάγονται τα συμπεράσματα. Ξεκινάμε με μια συγκριτική μελέτη των διαφορετικών μοντέλων που δημιουργήσαμε, αναλύοντας ξεχωριστά τα διάφορα χαρακτηριστικά και τον τρόπο με τον οποίο αυτά επηρεάζουν την εκπαίδευση και τα αποτελέσματα του μοντέλου. Στη συνέχεια αναλύουμε προσωπικές παρατηρήσεις του συγγραφέα και κάνουμε μια ποιοτική αξιολόγηση των αποτελεσμάτων έχοντας απομακρυνθεί από τα αριθμητικά αποτελέσματα και κοιτώντας τη γενικότερη εικόνα. Τέλος συγκρίνουμε τα αποτελέσματά μας με όσες εργασίες βρέθηκαν να έχουν δημοσιευθεί μέχρι τη στιγμή συγγραφής της διπλωματικής και είναι σχετικές και συγκρίσιμες με το πρόβλημα που καλούμαστε να λύσουμε.

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

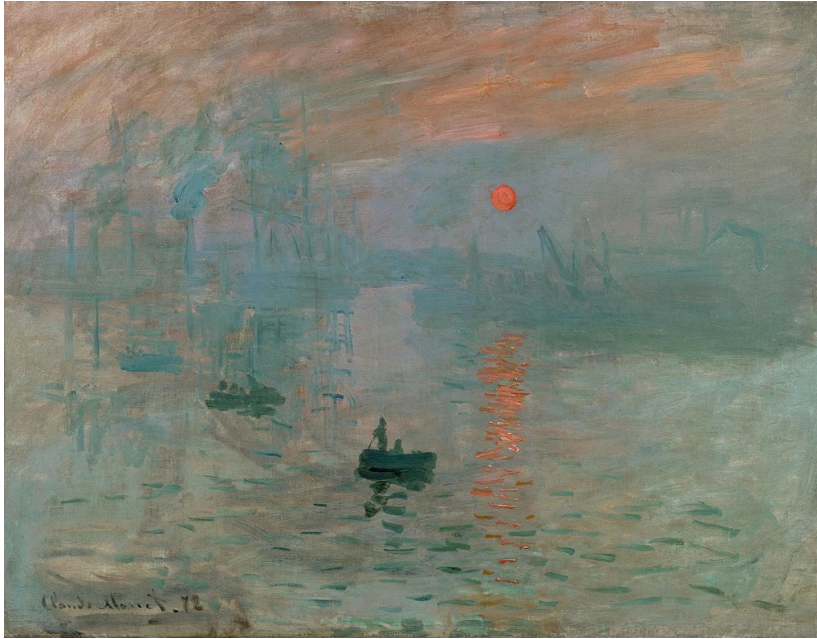
Στο παρόν κεφάλαιο θα καλυφθεί το θεωρητικό υπόβαθρο που κρίνεται απαραίτητο για την κατανόηση της διπλωματικής εργασίας. Χωρίζεται σε δύο βασικά μέρη, το πρώτο που ασχολείται με το πρόβλημα της αναγνώρισης τεχνοτροπίας αναλύοντας τις απαραίτητες έννοιες, μεθόδους και ορολογίες, και το δεύτερο που αφορά τη μηχανική μάθηση, τα εργαλεία και τις μεθόδους της. Για να είμαστε συνεπείς με την ορολογία μας πρέπει να αναφέρουμε πως μοντέλο, δίκτυο και ταξινομητής χρησιμοποιούνται ως ισοδύναμοι όροι καθώς και οι όροι κλάση και κατηγορία μεταξύ τους.

2.1 Αναγνώριση Τεχνοτροπίας

2.1.1 Αναγνώριση Τεχνοτροπίας και Ιστορική Αναδρομή

Στις εικαστικές τέχνες, τεχνοτροπία είναι ένας διακριτικός τρόπος που επιτρέπει την ομαδοποίηση των έργων σε συναφείς κατηγορίες (Fernie, 1995). Η τεχνοτροπία αφορά την οπτική εμφάνιση του έργου τέχνης η οποία το συνδέει με άλλα έργα του ίδιου καλλιτέχνη, ή/και με έργα από την ίδια εποχή ή σχολή, ίδιο τόπο ή κίνημα. Ανέκαθεν αποτελούσε τον κύριο γνώμονα ταξινόμησης έργων από τους ιστορικούς τέχνης. Συχνά η καλλιτεχνική τεχνοτροπία αναφέρεται και ως καλλιτεχνικό στιλ (ή στυλ) ή ύφος. Μια τεχνοτροπία συχνά διαιρείται σε επιμέρους στιλ/τεχνοτροπίες ανάλογα με το γενικότερο ύφος μιας περιόδου, μιας χώρας ή μιας κοινωνική ομάδας, μιας ομάδας καλλιτεχνών ή ενός καλλιτεχνικού κινήματος. Για παράδειγμα μια τεχνοτροπία συχνά διαιρείται χρονολογικά στις επί μέρους τεχνοτροπίες της "πρώιμης", "μέσης" και "όψιμης" περιόδου ("early", "middle" και "late"). Οποιοδήποτε έργο τέχνης θεωρητικά μπορεί να αναλυθεί από άποψη τεχνοτροπίας. Ούτε οι χρονικές περίοδοι ούτε οι καλλιτέχνες μπορούν να αποφύγουν να έχουν ένα στιλ. Αντίστροφα, φυσικά αντικείμενα ή θεάματα δεν μπορεί να ειπωθεί ότι έχουν ένα στιλ, καθώς το στιλ προκύπτει μόνο από τις επιλογές του δημιουργού. Οι περισσότερες στιλιστικές περίοδοι προσδιορίζονται και ορίζονται αργότερα από τους ιστορικούς τέχνης, όμως και οι καλλιτέχνες μπορούν να επιλέξουν να ορίσουν και να ονομάσουν το δικό τους στιλ.

Σημαντικός είναι ο διαχωρισμός καλλιτεχνικής τεχνοτροπίας και καλλιτεχνικού κινήματος. Σε πολλές περιπτώσεις αυτά τα δύο, λανθασμένα συγχέονται, ακόμη κι από άτομα τα οποία μελετούν προβλήματα όπως το δικό μας. Ένα καλλιτεχνικό κίνημα είναι μια καλλιτεχνική τάση ή τεχνοτροπία με μια συγκεκριμένη κοινή φιλοσοφία ή στόχο, ακολουθούμενη από μια ομάδα καλλιτεχνών σε μια περιορισμένη και συγκεκριμένη χρονική περίοδο και ορισμένες φορές συνδεδεμένη με κάποια συγκεκριμένη περιοχή. Αντίθετα, η τεχνοτροπία, όπως αναφέρεται παραπάνω, αφορά καθαρά και μόνο την οπτική εμφάνιση του έργου τέχνης. Εάν δημιουργηθεί σήμερα ένας ζωγραφικός πίνακας



Σχήμα 2.1: "Impression, soleil levant" , Claude Monet

αντίγραφο του "Impression, soleil levant" του Monet, δεν μπορούμε να πούμε πως ο συγκεκριμένος πίνακας ανήκει στο κίνημα του Ιμπρεσιονισμού. Μπορούμε όμως να αναγνωρίσουμε πως έχει την τεχνοτροπία του Ιμπρεσιονισμού. Ο Ιμπρεσιονισμός είναι ένα κίνημα που αναπτύχθηκε το 19ο αιώνα στη Γαλλία. Μόνο έργα τέχνης της συγκεκριμένης χρονικής περιόδου, από καλλιτέχνες που έζησαν, επηρεάστηκαν και διαμορφώθηκαν στα πλαίσια του συγκεκριμένου κινήματος μπορούν να αναγνωριστούν ως έργα του κινήματος του Ιμπρεσιονισμού. Το σύγχρονο αντίγραφο του ιμπρεσιονιστικού πίνακα, έχει επηρεαστεί από το συγκεκριμένο κίνημα, δεν αποτελεί όμως μέρος αυτού. Για αυτό το λόγο μελετώντας μόνο την εικόνα ενός έργου τέχνης δεν μπορούμε να αποφανθούμε για το κίνημα στο οποίο ανήκει, μπορούμε όμως να αναγνωρίσουμε την τεχνοτροπία του.

Η αναγνώριση της τεχνοτροπίας ενός έργου τέχνης δεν αποτελεί ένα τετριμμένο πρόβλημα. Μέχρι και στις μέρες μας η ταξινόμηση των έργων τέχνης επιτελείται επί των πλείστων από ιστορικούς και κριτικούς τέχνης με εμπειρία και χρόνια σπουδών και εκπαίδευσης στο συγκεκριμένο αντικείμενο, όπως γινόταν ανέκαθεν. Τα τελευταία χρόνια, με τις δυνατότητες που έχουν δημιουργηθεί μέσω της ψηφιοποίησης πολλών έργων τέχνης, αλλά και της ανάπτυξης νέων τεχνολογιών στον τομέα της πληροφορικής και της τεχνητής νοημοσύνης, έχουν γίνει αρκετές προσπάθειες να επιτελεστεί το συγκεκριμένο έργο κι από υπολογιστές. Προσπάθειες στο συγκεκριμένο τομέα έχουμε κι από το 2004 όταν οι [Jia Li and Wang \(2004\)](#) χρησιμοποίησαν δισδιάστατα μοντέλα Markov (2-dimensional Multiresolution Hidden Markov Model - MHMM) για να ταξινομήσουν τα έργα πέντε Κινέζων καλλιτεχνών, ενώ έχουμε αναφορές για προσπάθειες ταξινόμησης έργων τέχνης, όχι απαραίτητα ανά τεχνοτροπία, από το 1998 με τους [Sablatnig et al. \(1998\)](#) να μελετούν τη "δομική υπογραφή" κάθε καλλιτέχνη, με βάση τις πινελιές των έργων τέχνης. Το 2005 οι [Gunsel et al. \(2005\)](#) υλοποίησαν μερικούς απλούς αλγόριθμους μηχανικής μάθησης όπως ταξινομητές SVM, Bayesian ταξινομητές και ταξινομητές kNN για να ταξινομήσουν ένα σετ δεδομένων 107 εικόνων σε τρεις κλάσεις. Ένα χρόνο αργότερα (2006), οι [Jiang et al. \(2006\)](#) χρησιμοποίησαν ταξινομητές SVM χρησιμοποιώντας ως είσοδο χαρακτηριστικά τα οποία προέκυψαν από επεξεργασία των έργων τέχνης,

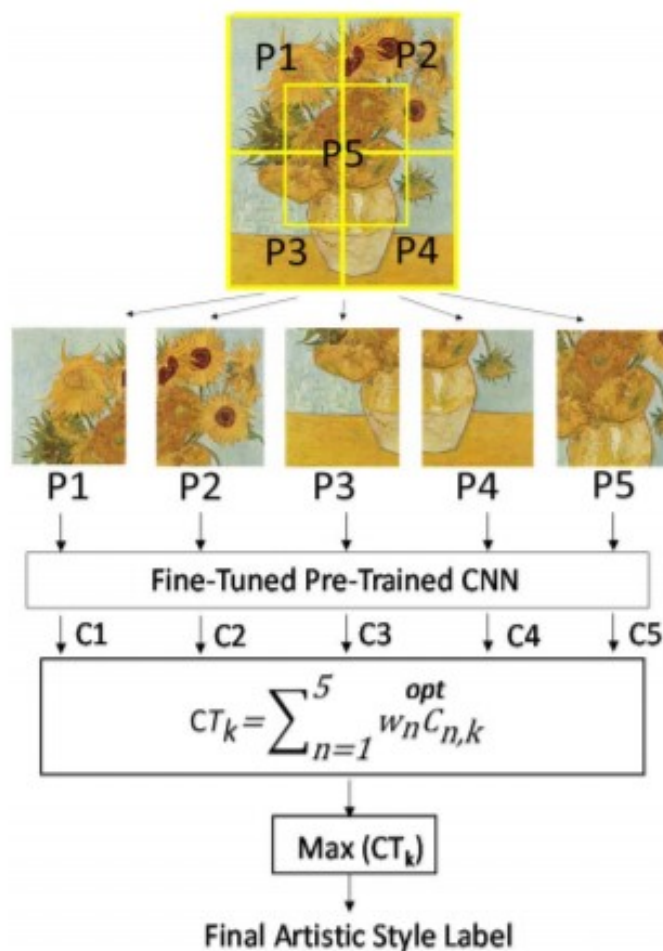
όπως ιστογράμματα χρωμάτων, διανύσματα συνοχής χρώματος κ.α., για να ταξινομήσουν έργα της παραδοσιακής κινέζικης τέχνης, και αυτοί όμως εργάστηκαν με μόνο 3 κατηγορίες. Το βασικότερο πρόβλημα των πρώτων προσπαθειών στο πεδίο της αναγνώρισης τεχνοτροπίας, ήταν η έλλειψη δεδομένων. Φυσικά δεν υπήρχαν ακόμη η γνώση και η τεχνολογία όπου έχουμε σήμερα, ώστε να φτάσουν σε συγκρίσιμα αποτελέσματα με τα state of the art, όμως αυτή η έλλειψη δεδομένων καθιστούσε το έργο τους μη γενικεύσιμο. Τα χρόνια που ακολούθησαν έγιναν αρκετές εργασίες (Arora and Elgammal, 2012; Khan et al., 2014; Shamir et al., 2010; Siddiquie et al., 2009) πάνω στο ίδιο αντικείμενο, με κοινό χαρακτηριστικό όμως την έλλειψη επαρκών δεδομένων ή/και την ταξινόμηση σε λίγες μόνο κλάσεις με αποτέλεσμα να μην προκύπτουν πολύ έγκυρα και γενικεύσιμα αποτελέσματα. Αποτέλεσαν όμως σημαντική βάση για τη δουλειά που έγινε αργότερα μόλις έγιναν διαθέσιμα σύνολα δεδομένων με πολλά έργα τέχνης και σημαντικό αριθμό κλάσεων. Οι Karayev et al. (2013) συνέλεξαν ένα εντυπωσιακά μεγάλο σύνολο δεδομένων από το WikiArt ¹, το οποίο έκτοτε χρησιμοποιείται κατά κόρον σε διάφορες παραλλαγές, σε εργασίες ταξινόμησης έργων τέχνης. Επίσης χρησιμοποίησαν συνελκτικά νευρωνικά δίκτυα για την ταξινόμηση των δεδομένων, τα οποία μέχρι και σήμερα είναι ο βασικότερος τρόπος με τον οποίο προσεγγίζεται το συγκεκριμένο πρόβλημα. Οι Bar et al. (2014) καθώς και οι Peng and Chen (2015) χρησιμοποίησαν όπως κι οι Karayev et al. (2013) συνελκτικά φίλτρα από προ-εκπαιδευμένα στο ImageNet ² βαθιά συνελκτικά δίκτυα όπως πρότειναν οι Donahue et al. (2013).

Το 2017 οι Florea et al. (2017) δημιούργησαν το σύνολο δεδομένων Pandora18k βασισμένο πάνω στο WikiArt, το οποίο στη συνέχεια χρησιμοποιήθηκε ευρέως για την αναγνώριση τεχνοτροπίας διότι αν και μικρότερο, θεωρήθηκε εγκυρότερο από πολλά άλλα διότι αναλύθηκε τόσο από μηχανικούς όσο κι από ειδικούς σε θέματα τέχνης. Προσέγγισαν και αυτοί το πρόβλημα της αναγνώρισης τεχνοτροπίας με τη χρήση συνελκτικών νευρωνικών δικτύων χρησιμοποιώντας αρχιτεκτονικές όπως AlexNet (Krizhevsky et al., 2012). Πολλές ακόμη ομάδες ασχολήθηκαν με τη βαθιά μάθηση για να προσεγγίσουν το πρόβλημα που καλούμαστε να λύσουμε, βελτιώνοντας συνεχώς τις επιδόσεις τους, και στην ενότητα 4.3 παρουσιάζεται μια αναλυτικότερη σύγκριση μεταξύ όλων των εργασιών. Οι Lecoutre et al. (2017) το 2017, προχώρησαν σε μια βαθύτερη επανεκπαίδευση ενός δικτύου βασισμένο στο ResNet (He et al., 2015a), επιτυγχάνοντας σημαντική βελτίωση στις μέχρι τότε επιδόσεις, και αποδεικνύοντας πως η αναγνώριση εικόνας, όπως είναι η πρόκληση ImageNet, δεν έχει τόσο μεγάλη συνάφεια με την αναγνώριση τεχνοτροπίας ώστε να παίρνουμε τα δίκτυα που έχουν εκπαιδευτεί στο ImageNet και να τα χρησιμοποιούμε χωρίς επιπλέον εκπαίδευση για αναγνώριση τέχνης. Το 2018 οι Rodriguez et al. (2018) χρησιμοποίησαν μια τεχνική διαμέρισης των έργων σε τεμάχια (patches), τα οποία ταξινόμησαν ξεχωριστά και στη συνέχεια συνδύασαν τα αποτελέσματα για κάθε τεμάχιο σε ένα τελικό αποτέλεσμα για την κάθε εικόνα, επιτυγχάνοντας κορυφαία αποτελέσματα μέχρι τη στιγμή δημοσίευσης της εργασίας τους.

Ένα χρόνο αργότερα (2019), η ίδια ομάδα σε μια συνέχεια της εργασίας τους, έκανα μια δημοσίευση με τίτλο "Two-Stage Deep Learning Approach to the Classification of Fine-Art Paintings" (Sandoval et al., 2019) στην οποία πηγαίνουν την παραπάνω μεθοδολογία ένα βήμα παραπέρα. Χρησιμοποιούν το δίκτυο με την ανά τεμάχιο ταξινόμηση, που δημιούργησαν στην προηγούμενη εργασία, ως πρώτο στάδιο, και σε δεύτερη φάση χρησιμοποιούν την έξοδο του δικτύου αυτού για να εκπαιδεύσουν ένα μικρότερο δίκτυο το οποίο τοποθετούν ουσιαστικά στην κορυφή του προη-

¹ "WikiArt, Visual Art Encyclopedia", www.wikiart.org

² www.image-net.org



Σχήμα 2.2: Σχηματική απεικόνιση της μεθοδολογίας που ακολουθήθηκε από τους Rodriguez et al. (2018)

γούμενου. Αυτή η τεχνική θυμίζει τη συνδυαστική μέθοδο του boosting, και κατάφερε να δώσει μια πολύ μεγάλη αύξηση στην ακρίβεια πρόβλεψης του ταξινομητή τους. Αυτές οι επιδόσεις αποτελούσαν μέχρι και τη στιγμή συγγραφής της συγκεκριμένης εργασίας, οι κορυφαίες που έχουν δημοσιευθεί, με την παρούσα διπλωματική να επιτυγχάνει να ξεπεράσει τις επιδόσεις που αφορούν στο σύνολο δεδομένων WikiArt.

2.1.2 Παραλλαγές Προβλήματος

Τα τελευταία χρόνια έχει υπάρξει ένα έντονο ενδιαφέρον της επιστημονικής κοινότητας για την τέχνη. Για αυτό το λόγο εκτός από το μεγάλο ενδιαφέρον για την αναγνώριση τεχνοτροπίας έργων τέχνης, συναντάμε κι αρκετές παραλλαγές του προβλήματος. Υπάρχουν πολλές εργασίες οι οποίες έχουν ασχοληθεί με διάφορες μορφές τέχνης προσεγγίζοντας τις από τη σκοπιά της τεχνητής νοημοσύνης. Εκτός από τη ζωγραφική, υπάρχουν ομάδες οι οποίες έχουν ασχοληθεί και με τη μουσική (Huang and Wu, 2016; van den Oord et al., 2016). Υπάρχουν προσεγγίσεις οι οποίες αναγνωρίζουν κάποιο χαρακτηριστικό της μουσικής και συχνά ταξινομούν μουσικά κομμάτια ως

προς αυτό, όπως το μουσικό είδος, τα βρημ κ.α. , όμως υπάρχουν και προσεγγίσεις οι οποίες προ-σπαθούν, μέσω της βαθιάς μάθησης κυρίως, να συνθέσουν εντελώς νέα μουσικά κομμάτια.

Στον τομέα της ζωγραφικής και των εικαστικών τεχνών, συχνή είναι η χρήση έργων τέχνης για την ταξινόμησή τους σύμφωνα με ορισμένα κριτήρια πέρα από την τεχνοτροπία, όπως είναι ο καλλιτέχνης - δημιουργός κάθε έργου, ή το θέμα - περιεχόμενο του Tan et al. (2016). Οι Shamir et al. (2016) ασχολήθηκαν ακόμη και με το διαχωρισμό έργων αφηρημένης τέχνης, σε έργα όπου ο δημιουργός τους ήταν καλλιτέχνης από τη μία, και σε έργα όπου ο δημιουργός τους ήταν κάποιο παιδί ή ζώο, από την άλλη. Όπως παρατηρούμε από τη δουλειά που έχει γίνει στον τομέα, τόσο η αναγνώριση καλλιτέχνη όσο κι η αναγνώριση περιεχομένου αποτελούν πιθανότατα, πολύ πιο εύκολα προβλήματα για έναν υπολογιστή από την αναγνώριση τεχνοτροπίας, μιας και οι εργασίες επιτυγχάνουν πολύ μεγαλύτερα ποσοστά ακρίβειας σε αυτά τα προβλήματα. Μια πιθανή αιτία για το υψηλό ποσοστό επιτυχίας στην αναγνώριση καλλιτέχνη είναι από τη μία η έλλειψη επαρκών δεδομένων για την ασφαλή αξιολόγηση ενός μοντέλου κι από την άλλη η αναγνώριση συγκεκριμένων στοιχείων που υπάρχουν στους πίνακες ενός καλλιτέχνη όπως η μορφή των πινελιών (brush strokes) ανάλογα με το χέρι που χρησιμοποιεί, το ύψος και τη θέση του καλλιτέχνη τα οποία δεν αποτελούν κοινό χαρακτηριστικό στους πίνακες κοινής τεχνοτροπίας. Επίσης το γεγονός πως όλα τα έργα ενός καλλιτέχνη προέρχονται από μια συγκεκριμένη και περιορισμένη χρονική περίοδο έχει ως αποτέλεσμα η κατάστασή τους και η φθορά τους από το χρόνο πιθανότατα να είναι σε παρόμοια επίπεδα, με αποτέλεσμα αυτό να μπορεί να δώσει επιπλέον οπτική πληροφορία στους ταξινομητές. Μια δυσκολία όμως που υπάρχει στην αναγνώριση καλλιτέχνη είναι τα διαφορετικά στιλ τα οποία άλλαζαν πολλοί καλλιτέχνες κατά τη διάρκεια της ζωής τους, επηρεαζόμενοι από τα κινήματα και τις τάσεις της εποχής τους. Για παράδειγμα οι τρεις πίνακες που παρουσιάζονται παρακάτω (2.3,2.4,2.5) έχουν ζωγραφιστεί από τον Pablo Picasso σε διαφορετικές περιόδους της ζωής του, και έχουν δημιουργηθεί με χαρακτηριστικά τελείως διαφορετικών τεχνοτροπιών. Η πρώτη εικόνα απεικονίζει ένα έργο του μεγάλου ζωγράφου που δημιουργήθηκε κατά τα πρώτα χρόνια της καλλιτεχνικής του πορείας, και να έχει την τεχνοτροπία του ρεαλισμού. Στη δεύτερη εικόνα βλέπουμε ένα έργο που προέρχεται από την "μπλε περίοδο" του καλλιτέχνη στην οποία δημιουργούσε κατά κύριο λόγο εξπρεσιονιστικά έργα τέχνης και έργα επηρεασμένα από το κίνημα του συμβολισμού, ενώ στην τελευταία εικόνα βλέπουμε το "κορίτσι με το μαντολίνο", έναν πολύ γνωστό πίνακα του καλλιτέχνη ο οποίος ανήκει στο κίνημα του κυβισμού του οποίου μάλιστα ηγήθηκε ο Picasso, με αποτέλεσμα στις μέρες μας πολλοί να τον γνωρίζουν μόνο για αυτό το κομμάτι της δουλειάς του. Παρατηρούμε λοιπόν τρεις διαφορετικούς πίνακες με τρεις διαφορετικές τεχνοτροπίες, οι οποίοι όμως δημιουργήθηκαν από τον ίδιο καλλιτέχνη μέσα σε διάστημα 15 χρόνων. Αντιλαμβανόμαστε λοιπόν τη δυσκολία που προκύπτει στην αναγνώριση καλλιτεχνών με τόσο μεγάλο εύρος δημιουργίας.

Από την άλλη, η αναγνώριση περιεχομένου (genre) αποτελεί πρόβλημα με πολύ μεγάλη συνάφεια με την αναγνώριση εικόνας στην οποία έχουμε πετύχει ήδη καταπληκτικά αποτελέσματα, ξεπερνώντας ακόμη και τις ανθρώπινες επιδόσεις (He et al., 2015b). Πέντε ³ από τα βασικά genres στα οποία χωρίζονται τα έργα τέχνης είναι η νεκρή φύση, τα πορτραίτα, τα τοπία, το ιστορικό και θρησκευτικό περιεχόμενο, και οι σκηνές καθημερινής ζωής (genre painting). Όπως παρατηρούμε λοιπόν η αναγνώριση εικόνας είναι καθοριστική για την ταξινόμηση έργων τέχνης στις παραπάνω

³ "The 5 Painting Genres", from Visual Arts Cork, Art Encyclopedia, www.visual-arts-cork.com/painting-genres.htm



Σχήμα 2.3: "The altarboy", Pablo Picasso, 1896, Realism



Σχήμα 2.4: "Woman in a shawl", Pablo Picasso, 1902, Expressionism



Σχήμα 2.5: "Girl with mandolin", Pablo Picasso, 1910, Cubism

κατηγορίες. Η αναγνώριση ενός προσώπου μπορεί να βοηθήσει σημαντικά στην αναγνώριση ενός έργου τέχνης που αποτελεί πορτραίτο, ή η αναγνώριση φρούτων και λοιπών φαγητών στην αναγνώριση ενός πίνακα νεκρής φύσης. Αντίστοιχα η αναγνώριση ενός δέντρου ή ενός λόφου, για έναν πίνακα τοπίου κ.ο.κ. Έχουμε λοιπόν ήδη πανίσχυρα εργαλεία για την επίλυση του συγκεκριμένου προβλήματος τα οποία μας δίνουν τη δυνατότητα να φτάνουμε τα, συγκριτικά, πολύ υψηλά ποσοστά ακρίβειας που έχουν επιτευχθεί. Στα παραπάνω μπορεί να προστεθεί και το γεγονός πως το πλήθος των διαφορετικών genres είναι συνήθως πολύ μικρότερο από το αντίστοιχο των τεχνοτροπιών ή των καλλιτεχνών, με αποτέλεσμα να υπάρχουν πολύ λιγότερες κλάσεις κατά την ταξινόμηση, το οποίο ευνοεί σημαντικά την επίτευξη υψηλής ακρίβειας. Για αυτούς τους λόγους επιτυγχάνουμε εντυπωσιακά μεγαλύτερη ακρίβεια στην αναγνώριση καλλιτέχνη ή περιεχομένου σε σύγκριση με την αναγνώριση τεχνοτροπίας. Οι [Tan et al. \(2016\)](#), οι οποίοι καταπιάστηκαν και με τα τρία προβλήματα, και η ακρίβεια στην αναγνώριση τεχνοτροπίας ήταν περίπου 20 ποσοστιαίες μονάδες κάτω από τις άλλες δύο κατηγορίες αναγνώρισης. Πέρα από την ταξινόμηση των έργων τέχνης, οι [Polatkan et al. \(2009\)](#) χρησιμοποίησαν ενισχυτική μάθηση για να εντοπίσουν πλαστά έργα τέχνης, με πολύ ενθαρρυντικά αποτελέσματά. Η αναγνώριση ενός πλαστού έργου με τη χρήση των υπολογιστών μπορεί να γλιτώσει πολλά έργα τέχνης από τις φθορές που προκαλούνται από την προσπάθεια ειδικών να αποφανθούν για την αυθεντικότητά τους. Ένα ακόμη παρεμφερές κομμάτι με το οποίο έχουν καταπιαστεί πολλοί κατά καιρούς, είναι η μεταφορά του στιλ μιας φωτογραφίας ή ενός έργου τέχνης σε μια εικόνα. Υπάρχουν ομάδες οι οποίες κατάφεραν να διαχωρίσουν το στιλ από το περιεχόμενο ενός πίνακα ζωγραφικής, κι επιτυχώς να μεταφέρουν το στιλ σε μια απλή φωτογραφία ([Gatys et al., 2015](#); [Johnson et al., 2016](#)). Ένα βήμα παραπέρα από την απλή μεταφορά του στιλ ενός ζωγραφικού πίνακα ή μιας καλλιτεχνικής φωτογραφίας, αποτελεί η δημιουργία ενός εντελώς νέου πίνακα ζωγραφικής, σε αντιστοιχία με τη σύνθεση ενός νέου μουσικού κομματιού που αναφέρθηκε παραπάνω. Υπάρχουν πολλοί που έχουν αμφισβητήσει την ικανότητα υπολογιστών να δημιουργήσουν τέχνη με το επιχείρημα πως ένας υπολογιστής δεν μπορεί να δημιουργήσει απλώς να αντιγράψει και να αναπαράγει. Αυτό αποτελεί ένα πολύ ενδιαφέρον



Σχήμα 2.6: Παραδείγματα Style Transfer από την εργασία των Gatys et al. (2015)

θέμα συζήτησης το οποίο μπλέκει θεωρητικά κομμάτια της τέχνης όπως η παρθενογένεση στην τέχνη, καθώς και την ικανότητα ενός ανθρώπινου δημιουργήματος όπως ο υπολογιστής, να παράγει γνώση στην οποία δεν έχει εκτεθεί, με το οποίο όμως δεν θα ασχοληθούμε στη συγκεκριμένη διπλωματική εργασία. Παρ' όλη την αμφισβήτηση που έχει δεχθεί ο συγκεκριμένος τομέας έρευνας όμως, το 2018 έγινε η πρώτη δημοπρασία έργου τέχνης το οποίο δημιουργήθηκε εξ ολοκλήρου από ηλεκτρονικό υπολογιστή το οποίο ξεπερνώντας κάθε εκτίμηση, πουλήθηκε για το ποσό των 432.500\$!



Σχήμα 2.7: "Portrait of Edmond Belamy", ο πρώτος πίνακας, δημιούργημα τεχνητής νοημοσύνης, που πωλήθηκε ποτέ.

2.2 Μηχανική Μάθηση

2.2.1 Ορισμός Μηχανικής Μάθησης

Μηχανική μάθηση είναι το σύνολο των μεθόδων που μπορούν να ανιχνεύουν αυτόματα τα μοτίβα στα δεδομένα, και στη συνέχεια χρησιμοποιούν αυτά τα μοτίβα για να προβλέψουν μελλοντικά δεδομένα, ή να εκτελέσουν άλλα είδη αποφάσεων υπό αβεβαιότητα. (Murphy, 2012)

Η μηχανική μάθηση αποτελεί μέρος του κλάδου της πληροφορικής που συνήθως αναφέρεται με τον όρο "Τεχνητή Νοημοσύνη". Πολύ συχνά οι δύο όροι λανθασμένα συγχέονται. Η τεχνητή νοημοσύνη, όπως την όρισε ο John McCarthy, είναι η επιστήμη και η τεχνολογία κατασκευής ευφυών μηχανών (McCarthy, 2004). Πρόκειται λοιπόν για τον γενικότερο κλάδο της πληροφορικής που ασχολείται με την ανάπτυξη ευφυών μηχανών, μιας προσπάθειας δηλαδή να κάνουμε τους υπολογιστές να μιμηθούν την ευφυή συμπεριφορά του ανθρώπου ώστε να εκτελέσουμε εργασίες οι οποίες απαιτούν ανθρωπινή ευφυΐα όπως αναγνώριση λόγου, οπτική αντίληψη και λήψη αποφάσεων. Η βασική διαφορά της μηχανικής μάθησης συγκριτικά με τους γράφους γνώσης και άλλα πεδία της τεχνητής νοημοσύνης, είναι η ικανότητα της να τροποποιείται και να προσαρμόζεται στα νέα δεδομένα που δέχεται. Μέσω αυτής της ικανότητας της μηχανικής μάθησης, οι επιστήμονες μπορούν πλέον τροφοδοτώντας με δεδομένα να οδηγήσουν τους υπολογιστές σε νέα γνώση την οποία ούτε οι ίδιοι κατέχουν. Ένα από τα πρώτα παραδείγματα των παραπάνω αποτελεί η εργασία του Arthur Samuel, με τίτλο "Some Studies in Machine Learning Using the game of checkers" (Samuel, 1959), όπου ο συγγραφέας κατάφερε να φτιάξει ένα πρόγραμμα υπολογιστή με μηχανική μάθηση το οποίο έγινε καλύτερο από τον ίδιο στο παιχνίδι "Ντάμα" (Checkers). Στις μέρες μας όπου πληθώρα δεδομένων είναι διαθέσιμη και υπάρχει ολοένα και μεγαλύτερη ψηφιοποίηση αυτών των δεδομένων, η μηχανική μάθηση αποτελεί ένα πολύ ισχυρό εργαλείο για την επεξεργασία και αξιοποίηση αυτού του όγκου πληροφορίας, καθώς και για τη δημιουργία νέας γνώσης στην οποία δεν είχαμε πρόσβαση πρωτότερα.

2.2.2 Βαθιά Μάθηση

Η βαθιά μάθηση είναι ένα ιδιαίτερο είδος μηχανικής μάθησης, που επιτυγχάνει μεγάλη δύναμη και ευελιξία, μαθαίνοντας να αναπαριστά τον κόσμο ως μια εμφολευμένη ιεραρχία των εννοιών, με κάθε έννοια να ορίζεται σε σχέση με απλούστερες έννοιες και πιο αφηρημένες αντιλήψεις υπολογιζόμενες με όρους λιγότερο αφηρημένους (Goodfellow et al., 2016). Η βαθιά μάθηση χρησιμοποιεί πολλαπλά επίπεδα ώστε σταδιακά να εξάγει χαρακτηριστικά υψηλότερου επιπέδου από ανεπεξέργαστα δεδομένα. Για παράδειγμα στην επεξεργασία εικόνας, χαμηλότερα επίπεδα μπορεί να αναγνωρίζουν ακμές ή χρώματα ενώ υψηλότερα επίπεδα να αναγνωρίζουν έννοιες σχετικές με την ανθρώπινη εμπειρία όπως γράμματα, ψηφία ή πρόσωπα. Η βαθιά μάθηση βρίσκει εφαρμογή σε πληθώρα προβλημάτων όπως την όραση υπολογιστών, την αναγνώριση λόγου, την επεξεργασία φυσικής γλώσσας και την ανάλυση ιατρικών εικόνων.

2.2.3 Βασικά ήδη μηχανικής μάθησης

Τα τρία βασικότερα είδη μηχανικής μάθησης αποτελούν η Επιβλεπόμενη, ή Επιτηρούμενη Μάθηση (Supervised Learning), η Μη Επιβλεπόμενη Μάθηση (Unsupervised Learning) και η Ενισχυτική Μάθηση (Reinforced Learning).

2.2.3.1 Επιβλεπόμενη Μάθηση

Στην πρώτη περίπτωση συντίθεται ένα μαθηματικό μοντέλο από ένα σύνολο δεδομένων, τα δεδομένα εκπαίδευσης, που διαθέτει τόσο τις εισόδους όσο και τις επιθυμητές εξόδους. Ο αλγόριθμος εκπαιδεύεται με βάση τα δεδομένα και προσπαθεί να προσεγγίσει τη βέλτιστη συνάρτηση που μπορεί να τα περιγράψει, ώστε στη συνέχεια, δεδομένης μιας εισόδου να είναι σε θέση να προβλέψει την επιθυμητή έξοδο. Ο αλγόριθμος δηλαδή προσπαθεί να μοντελοποιήσει τις σχέσεις ανάμεσα στα χαρακτηριστικά των εισόδων και τις επιθυμητές εξόδους ώστε 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.

2.2.3.2 Μη Επιβλεπόμενη Μάθηση

Στη δεύτερη περίπτωση, τα δεδομένα δεν διαθέτουν επιθυμητή έξοδο. Αντίθετα, ο αλγόριθμος προσπαθεί να εντοπίσει μοτίβα, να βρει κανόνες, να ομαδοποιήσει και να εξάγει συμπεράσματα από τα δεδομένα. Είναι η μέθοδος που επιδιώκει να περιγράψει καλύτερα τα δεδομένα στο χρήστη. Συνήθως επιλέγεται για ομαδοποίηση και για συσχέτιση.

2.2.3.3 Ενισχυτική Μάθηση

Στην τρίτη περίπτωση, το μοντέλο παρατηρεί τις αλληλεπιδράσεις του με το περιβάλλον, λαμβάνει δράση με βάση αυτές και προσπαθεί να αυξήσει την ανταμοιβή ή να ελαχιστοποιήσει τον κίνδυνο. Εδώ, δηλαδή, ο αλγόριθμος επαναληπτικά 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. (Μερτζάνη, 2019)

Στη συγκεκριμένη διπλωματική θα ασχοληθούμε με την επιβλεπόμενη μάθηση και πιο συγκεκριμένα με τα τεχνητά νευρωνικά δίκτυα.

2.2.4 Τεχνητά Νευρωνικά Δίκτυα

Ένα νευρωνικό δίκτυο είναι ένας τεράστιος παράλληλος επεξεργαστής με κατανομημένη αρχιτεκτονική, ο οποίος αποτελείται από απλές μονάδες επεξεργασίας και έχει από τη φύση του τη δυνατότητα να αποθηκεύει εμπειρική γνώση και να την καθιστά διαθέσιμη για χρήση. Μοιάζει με τον ανθρώπινο εγκέφαλο σε δύο σημεία:

1. Το δίκτυο προσλαμβάνει τη γνώση από το περιβάλλον του, μέσω μιας διαδικασίας μάθησης.
2. Η ισχύς των συνδέσεων μεταξύ των νευρώνων, που αποκαλείται συναπτικό βάρος, χρησιμοποιείται για την αποθήκευση της γνώσης που αποκτιέται.

(Haykin, 2009) Τα τεχνητά νευρωνικά δίκτυα (Artificial Neural Networks - ANN) είναι μια υπολογιστική δομή εμπνευσμένη από ένα βιολογικό νευρικό σύστημα. Οι απλές επεξεργαστικές δομικές μονάδες ενός ΤΝΔ που αναφέρονται παραπάνω, ονομάζονται νευρώνες. Οι νευρώνες συνδέονται μεταξύ τους μέσω συνδέσμων με βάρη πάνω στους οποίους μπορούν να περάσουν τα σήματα. Τα τεχνητά νευρωνικά δίκτυα χρησιμοποιούνται σε πολλά διαφορετικά πεδία για την επεξεργασία

μεγάλων συνόλων δεδομένων, συχνά παρέχοντας χρήσιμες αναλύσεις που επιτρέπουν την πρόβλεψη και αναγνώριση νέων δεδομένων. Υπολογίζουν τα δομικά δεδομένα μέσω μιας διαδικασίας μάθησης και εκπαίδευσης. Τα δεδομένα που χρησιμοποιούνται συνήθως από αυτές τις δομές έχουν μη γραμμικές σχέσεις μεταξύ εισόδων και εξόδων. Χρησιμοποιούνται σε εφαρμογές όπως αναγνώριση ομιλίας, απεικόνιση, έλεγχος, εκτίμηση, βελτιστοποίηση κ.α. Βρίσκουν επίσης πρακτική εφαρμογή στους τομείς της χρηματοδότησης, της ιατρικής, των επιχειρήσεων, των μεταλλείων κ.λπ. (El-Shahat, 2018).

Αναφορικά με τη λειτουργία των νευρωνικών δικτύων, κάθε δομική μονάδα (νευρώνας) 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. Η έξοδος του νευρώνα στις περισσότερες περιπτώσεις μπορεί να είναι είτε 0 (OFF) είτε 1 (ON), λειτουργεί δηλαδή σαν μια μορφή διακόπτη ο οποίος είτε είναι ενεργός είτε ανενεργός. Εξάιρεση αποτελεί το τελευταίο επίπεδο το οποίο δίνει την έξοδο στο περιβάλλον όπου τότε έχει διαφορετική σημασία η έξοδος κάθε νευρώνα. Η τιμή της εξόδου ενός νευρώνα εξαρτάται από τη συνάρτηση ενεργοποίησής του. 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. Τρεις συναρτήσεις ενεργοποίησης οι οποίες χρησιμοποιούνται επί των πλείστων είναι οι εξής:

- Sigmoid

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

Το σύνολο τιμών για τη σιγμοειδή συνάρτηση είναι το διάστημα (0,1). Για αυτό το λόγο χρησιμοποιείται ευρέως σε περιπτώσεις όπου ο νευρώνας θέλουμε να δώσει κάποια πιθανότητα, μιας και οι πιθανότητες είναι πάντα στο διάστημα [0,1]. Η συνηθέστερη τέτοια περίπτωση είναι η ταξινόμηση των δεδομένων σε δύο κλάσεις - κατηγορίες, έστω A και B, όπου η σιγμοειδής συνάρτηση χρησιμοποιείται στο νευρώνα ο οποίος δίνει την έξοδο στο περιβάλλον και η τιμή της εξόδου του σημαίνει την πιθανότητα η είσοδος να ανήκει στη μία από τις δύο κλάσεις, έστω την A. Έτσι εισοδοί για τις οποίες η τελική έξοδος είναι μεγαλύτερο από 0,5 ταξινομούνται στην κατηγορία A, ενώ εισοδοί όπου η έξοδος του δικτύου δίνει τιμή μικρότερη του 0,5 ταξινομούνται στην κατηγορία B.

- 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.

- ReLU

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

Η συνάρτηση ReLU (Rectified Linear Unit) είναι από τις πιο ευρέως χρησιμοποιούμενες συναρτήσεις ενεργοποίησης αυτή τη στιγμή, μιας και χρησιμοποιείται σε όλα σχεδόν τα συνελκτικά νευρωνικά δίκτυα και τη βαθιά μάθηση. Το σύνολο τιμών της είναι το $[0, \infty)$. Το γεγονός που την κάνει τόσο διαδεδομένη είναι η ευκολία υπολογισμού της, πρακτικά χρειάζεται μόνο μια σύγκριση της εισόδου με το 0 και επιλέγεται το μέγιστο μεταξύ των δύο. Αυτό απλοποιεί πολύ τους υπολογισμούς που απαιτούνται κατά την εκπαίδευση βελτιώνοντας σημαντικά το χρόνο και τους πόρους που απαιτούνται.

2.2.5 Συνελκτικά Νευρωνικά Δίκτυα

2.2.5.1 Ορισμός

Τα συνελκτικά νευρωνικά δίκτυα (Convolutional Neural Networks - CNN) αποτελούν μια ειδική κατηγορία νευρωνικών δικτύων, που χρησιμοποιείται επί των πλείστων για την επεξεργασία δεδομένων τα οποία έχουν μια γνωστή πλεγματοειδή τοπολογία. Παραδείγματα περιλαμβάνουν δεδομένα χρονοσειρών, τα οποία μπορούν να θεωρηθούν ως μονοδιάστατο πλέγμα λαμβάνοντας δείγματα σε τακτά χρονικά διαστήματα, και δεδομένα εικόνας, που μπορούν να θεωρηθούν ως ένα δισδιάστατο πλέγμα από pixel. Το όνομα "Συνελκτικό Νευρωνικό Δίκτυο" υποδεικνύει ότι το δίκτυο χρησιμοποιεί μια μαθηματική πράξη που ονομάζεται συνέλιξη. Τα συνελκτικά νευρωνικά δίκτυα είναι απλώς νευρωνικά δίκτυα που χρησιμοποιούν συνέλιξη στη θέση του πολλαπλασιασμού γενικής μήτρας σε τουλάχιστον ένα από τα επίπεδά τους. (Goodfellow et al., 2016)

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

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

$$s(t) = x(t) \otimes w(t)$$

Η ανάγκη ανάπτυξης και κατ' επέκταση η χρήση συνελκτικών νευρωνικών δικτύων έγκειται στο γεγονός πως τα απλά νευρωνικά δίκτυα δεν κλιμακώνουν καλά όταν δέχονται ως είσοδο εικόνες. Για παράδειγμα για μια εικόνα μεγέθους $200 \times 200 \times 3$ pixel, ο κάθε νευρώνας του πρώτου επιπέδου θα πρέπει να έχει 120000 παραμέτρους, άρα για ένα και μόνο πυκνό επίπεδο θα χρειαζόμασταν εκατομμύρια παραμέτρους. Φυσικά θα χρειαζόταν πάνω από ένα πυκνά επίπεδα για να αναπαρασταθούν πιο περίπλοκες συναρτήσεις, οδηγώντας στη δημιουργία ενός τεράστιου δικτύου με αποτέλεσμα να είναι πολύ δύσκολη και χρονοβόρα η εκπαίδευσή του και μάλιστα χωρίς να έχει τη δυνατότητα να αποδώσει καλά λόγω του overfitting. Τα συνελκτικά 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 pixel 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.

2.2.5.2 Αρχιτεκτονική Συνελκτικών Νευρωνικών Δικτύων

Επίπεδα - Στρώματα Υπάρχουν τρία βασικά είδη επιπέδων για τη δημιουργία ενός συνελκτικού νευρωνικού δικτύου. Κυριότερο όλων από το οποίο πήραν και το όνομα τα δίκτυα, είναι το επίπεδο συνέλιξης. Στο συγκεκριμένο επίπεδο υπάρχουν πυρήνες-φίλτρα, οι παράμετροι των οποίων πρέπει να προσδιορισθούν μέσω της εκπαίδευσης του δικτύου. Αυτά τα φίλτρα έχουν προκαθορισμένο μέγεθος, μικρότερο της εισόδου, και κατά την εκπαίδευση επιτελείται η πράξη της συνέλιξης του φίλτρου διαδοχικά με ολόκληρη την είσοδο. Η σειρά κι ο τρόπος με τον οποίο γίνονται οι διάφορες συνέλιξεις (stride και padding) αφορούν σχεδιαστικές επιλογές του αρχιτέκτονα με τις οποίες δεν θα ασχοληθούμε στη συγκεκριμένη εργασία. Τα αποτελέσματα αυτής της συνέλιξης εξάγονται και διαμορφώνουν την είσοδο προς το επόμενο επίπεδο του δικτύου.

Το δεύτερο βασικό είδος επιπέδων που χρησιμοποιείται στα συνελκτικά νευρωνικά δίκτυα είναι τα επίπεδα Pooling. Σε αυτά τα επίπεδα συνοψίζονται οι έξοδοι γειτονικών γκρουπ νευρώνων εντός ενός παραθύρου με μια αντιπροσωπευτική τιμή. Πρόκειται ουσιαστικά για μια διαδικασία υπο-δειγματοληψίας. Οι επικρατέστερες κατηγορίες του pooling είναι οι max, sum και average pooling, ενώ μπορεί τα παράθυρα που χρησιμοποιούνται να επικαλύπτονται ή και όχι ανάλογα με τις ανάγκες του προβλήματος. Η διαδικασία του pooling, εκτός απ τη μείωση του μεγέθους των δεδομένων, μας δίνει τη δυνατότητα προσθήκης περισσότερης πληροφορίας στην αρχική εικόνα μέσω των αρχικών διαστάσεων ενώ είναι ανεξάρτητο μικρών μετασχηματισμών κάνοντας το μοντέλο μας πιο ανθεκτικό και στιβαρό απέναντι σε μικρο-αλλαγές. (Παπαδόπουλος, 2016)

Τέλος η τρίτη βασική κατηγορία επιπέδων που χρησιμοποιούνται στα CNN είναι τα πυκνά ή πλήρως συνδεδεμένα επίπεδα. Κάθε κόμβος ενός πυκνού επιπέδου είναι συνδεδεμένος με τις εξόδους όλων των κόμβων του προηγούμενου επιπέδου. Στα περισσότερα μοντέλα, το επίπεδο που δίνει την έξοδο προς το περιβάλλον είναι ένα πλήρως συνδεδεμένο επίπεδο.

Χαρακτηριστικές Αρχιτεκτονικές και Ιστορική Αναδρομή

- AlexNet

Είναι η αρχιτεκτονική που άλλαξε τον τρόπο με τον οποίο προσεγγίζεται η βαθιά μάθηση κι έφερε στο προσκήνιο τα βαθιά συνελκτικά νευρωνικά δίκτυα. Πρόκειται για μια αρχιτεκτονική που δημιουργήθηκε από τους Alex Krizhevsky, Ilya Sutskever και Geoffrey Hinton το 2012 (Krizhevsky et al., 2012) και κατάφερε να ξεπεράσει το δεύτερο σε επιτυχία δίκτυο στην πρόκληση ImageNet κατά 10.8%, το οποίο αποτελούσε πολύ εντυπωσιακό και σημαντικό επίτευγμα. Έτσι ξεκίνησε μια έκρηξη στη μελέτη της βαθιάς μάθησης και συγκεκριμένα των συνελκτικών νευρωνικών δικτύων. Από τότε και μέχρι σήμερα κάθε χρόνο ο νικητής της συγκεκριμένης πρόκλησης είναι ένα συνελκτικό νευρωνικό δίκτυο το οποίο δείχνει την υπεροχή των CNN έναντι των υπολοίπων στην επεξεργασία και αναγνώριση εικόνων. Η καινοτομία της συγκεκριμένη έρευνας δεν σταμάτησε στην κατά βάθος ανάπτυξη της αρχιτεκτονικής αλλά και στο γεγονός πως για να καταφέρουν να εκπαιδεύσουν το δίκτυο που δημιούργησαν χρησιμοποίησαν GPUs. Αν και δεν ήταν οι πρώτοι όπου έκαναν κάτι ανάλογο, από τότε ξεκίνησε να εφαρμόζεται πολύ ευρέως η συγκεκριμένη πρακτική. Η αρχιτεκτονική AlexNet αποτελείται από 8 επίπεδα. Τα πρώτα πέντε είναι συνελκτικά επίπεδα, ορισμένα από τα οποία ακολουθούνται από επίπεδα pooling, και τα τελευταία τρία επίπεδα είναι πλή-

πως συνδεδεμένα επίπεδα. Η συνάρτηση ενεργοποίησης που χρησιμοποιήθηκε στα κρυφά επίπεδα ήταν η ReLU.

- VGG

Η αρχιτεκτονική VGG αναπτύχθηκε από την ομάδα οπτικής γεωμετρίας του πανεπιστημίου της Οξφόρδης (Simonyan and Zisserman, 2014). Βασίστηκε πάνω στην αρχιτεκτονική AlexNet με κύρια ιδέα τη μεγαλύτερη εμβάθυνση του δικτύου. Ενώ άλλες προσεγγίσεις είχαν επικεντρωθεί στο μέγεθος των παραθύρων των φίλτρων, με μια τάση προς τα μικρότερα μεγέθη, και στη διαφορετική προσπέλαση των εικόνων με τα φίλτρα (stride), οι αρχιτέκτονες του VGG επέλεξαν να προσθέσουν περισσότερα επίπεδα, το οποίο τους έκανε να διαφοροποιηθούν από τις άλλες ομάδες. Ενώ το AlexNet αποτελούνταν από 8 μόνο επίπεδα, στα μοντέλα VGG ο αριθμός των επιπέδων υπερδιπλασιάστηκε, φτάνοντας τα 16 επίπεδα για την αρχιτεκτονική VGG16 και τα 19 για τη VGG19 όπως υποδεικνύουν και τα ονόματά τους. Επίσης στις VGG αρχιτεκτονικές χρησιμοποιήσαν αρκετά μικρότερο μέγεθος στα φίλτρα των επιπέδων συνέλιξης συγκριτικά με το AlexNet (3×3 αντί για 11×11), και μικρότερο stride (το οποίο όμως συνδέεται άμεσα και με το μέγεθος του φίλτρου (1 αντί 3))⁴.

- Inception

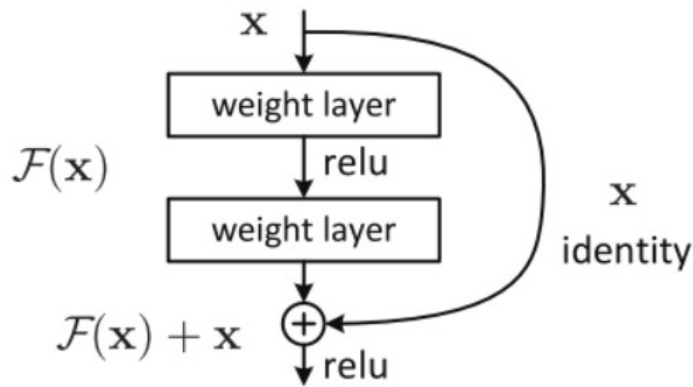
Η αρχιτεκτονική Inception αναπτύχθηκε από την ομάδα των 729 στα πλαίσια της έρευνας από τη Google, εξού και το δεύτερο όνομα με το οποίο έγινε γνωστό το πρώτο δίκτυο Inception, InceptionV1 ή αλλιώς GoogleNet. Η καινοτομία των σχεδιαστών του δικτύου ήταν να κάνουν το δίκτυο πιο "πλατύ" εκτός από απλώς πιο βαθύ όπου ήταν η τάση της εποχής. Αυτό το πέτυχαν χρησιμοποιώντας φίλτρα διαφορετικού μεγέθους στο ίδιο επίπεδο, με βάση τη λογική πως σε μια εικόνα μπορεί το αντικείμενο να βρίσκεται σε πρώτο πλάνο και να καταλαμβάνει μεγάλο μέρος της, αλλά μπορεί να βρίσκεται και σε μια μονάχα γωνία της, έχοντας πολύ μικρότερο μέγεθος. Ένα επίπεδο θα έπρεπε λοιπόν να είναι ικανό να εντοπίσει το αντικείμενο σε όλες τις περιπτώσεις. Στα επόμενα χρόνια δημιουργήθηκαν από τους σχεδιαστές του αρχικού InceptionV1, και επόμενα μοντέλα, όπου το καθένα αποτελούσε βελτίωσή του προηγούμενου. Τα InceptionV2 και InceptionV3 παρουσιάστηκαν μαζί (szegedy) ενώ μερικούς μήνες αργότερα παρουσιάστηκαν τα InceptionV4 και Inception-ResNet (εμπνευσμένο κι από την αρχιτεκτονική ResNet όπως υποδυκνύει και το όνομα) (Szegedy et al., 2016)⁵.

- ResNet

Τα ResNets κι οι αντίστοιχες αρχιτεκτονικές αναπτύχθηκαν το 2015 από την ομάδα των He et al. (2015a) ως μέρος έρευνας για τη Microsoft. Μέχρι τότε εάν και είχε ξεκινήσει η τάση των βαθιών συνελκτικών νευρωνικών δικτύων, οι αρχιτέκτονες είχαν πέσει πάνω στο πρόβλημα του vanishing gradient, το γεγονός δηλαδή πως κατά το backpropagation η παράγωγος λόγω της τιμής των βαρών γίνεται τόσο μικρή με αποτέλεσμα να αδυνατεί να επηρεάσει τα υπόλοιπα βάρη και πρακτικά να επιβραδύνεται η εκπαίδευση του δικτύου ή ακόμη και να σταματάει εντελώς. Δεν μπορούσαν λοιπόν να στοιβάζουν απλώς επίπεδα ελπίζοντας να πάρουν καλύτερα αποτελέσματα, διότι λόγω αυτού του προβλήματος, όταν ανέβαινε ο

⁴ "VGG Neural Networks: The Next Step After AlexNet", Jerry Wei, towardsdatascience.com/vgg-neural-networks-the-next-step-after-alexnet-3f91fa9ffe2c

⁵ "A Simple Guide to the Versions of the Inception Network", Bharath Raj, towardsdatascience.com/a-simple-guide-to-the-versions-of-the-inception-network-7fc52b863202



Σχήμα 2.8: Ένα Residual Block (Rodriguez et al., 2018)

αριθμός των επιπέδων τα αποτελέσματα που προέκυπταν ήταν χειρότερα. Την απάντηση στο πρόβλημα βρήκαν οι He et al. (2015a) μέσω των Residual Networks, ή ResNets. Το βασικό προτέρημα των ResNets είναι τα residual blocks τα οποία πρακτικά αποτελούν συνδέσεις οι οποίες επιτρέπουν την παρακάμψη ενός ή περισσοτέρων επιπέδων .

Τα ResNets δεν αποτελούν την πρώτη εφαρμογή τέτοιων παρακάμψεων. Στην εργασία των sri για τα Highway Networks, εισάγονται ορισμένες παρακάμψεις με έλεγχο στο πλήθος των δεδομένων που μπορούν να μεταφερθούν μέσω αυτών. Τα ResNets μπορούν λοιπόν να θεωρηθούν μια ειδικότερη περίπτωση των Highway Networks. Η χρήση αυτών των παρακάμψεων επέτρεψε στα ResNets να έχουν πολλά περισσότερα επίπεδα συγκριτικά με τα προηγούμενα μοντέλα. Για παράδειγμα το VGG Simonyan and Zisserman (2014) είχε μέχρι 19 επίπεδα και το GoogleNet (ή αλλιώς InceptionV1) 729 μέχρι 22. Αντίθετα τα ResNets κατάφεραν να φτάσουν τα 50, 101 και 152 με τις αρχιτεκτονικές ResNet50, ResNet101 και ResNet152 αντίστοιχα. Μάλιστα οι δημιουργοί των ResNets εκπαιδύσαν και τεράστια δίκτυα με βάθος ακόμη και πάνω των χιλίων επιπέδων. Υπάρχουν και μικρότερες εκδοχές των δικτύων όπως τα ResNet18 και ResNet34 τα οποία χρησιμοποιούν residual blocks δύο επιπέδων σε αντίθεση με τα μεγαλύτερα, τα οποία χρησιμοποιούν residual blocks τριών επιπέδων ⁶, και χρησιμοποιούνται συνήθως σε περιπτώσεις που τα ResNets με 50 επίπεδα και πάνω είναι πολύ μεγάλα ή/και "βαριά" για κάποιο υπολογιστικό σύστημα ⁷.

2.2.6 Μεταφορά Γνώσης - Transfer Learning

Δεδομένου ενός πεδίου D_s , ενός εκπαιδευτικού έργου T_s , ενός πεδίου D_t κι ενός εκπαιδευτικού έργου T_t , η μεταφορά γνώσης στοχεύει να συμβάλει στη βελτίωση της μάθησης της συνάρτησης πρόγνωσης $f(\cdot)$ στο D_t χρησιμοποιώντας τις γνώσεις από τα D_s και T_s , όπου $D_s \neq D_t$, ή $T_s \neq T_t$. (Lin and Jung, 2017) Σκοπός της μεταφοράς γνώσης λοιπόν είναι να χρησιμοποιηθεί γνώση η οποία έχει προέλθει από την επίλυση ενός προβλήματος A για την επίλυση ενός διαφορετικού προβλήματος B το οποίο μπορεί να αφορά διαφορετικά δεδομένα ή/και διαφορετικό πλαίσιο, πρέπει όμως

⁶ "ResNet (34, 50, 101): Residual CNNs for Image Classification Tasks", Muneeb ul Hassan, neurohive.io/en/popular-networks/resnet/

⁷ "An Overview of ResNet and its Variants", Vincent Fung, towardsdatascience.com/an-overview-of-resnet-and-its-variants-5281e2f56035

να έχει κάποια συνάφεια με το αρχικό πρόβλημα A, ώστε να μπορεί να χρησιμοποιηθεί με κάποιον τρόπο η προϋπάρχουσα γνώση. Για παράδειγμα εάν έχουμε ένα νευρωνικό δίκτυο το οποίο έχει εκπαιδευτεί στο να αναγνωρίζει πρόσωπα, μπορούμε να χρησιμοποιήσουμε την ιδιότητά του αυτή για να αναγνωρίσουμε ανθρώπους σε μια εικόνα. Η εφαρμογή της μεταφοράς γνώσης γίνεται με τον εξής τρόπο: αρχικά εκπαιδεύουμε ένα νευρωνικό δίκτυο να επιτελέσει ένα συγκεκριμένο έργο, πιο απλοϊκό συνήθως από το τελικό πρόβλημα το οποίο θέλουμε να επιλύσουμε, για το οποίο όμως έχουμε περισσότερα δεδομένα. Η συνήθης πρακτική είναι αφού εκπαιδεύσουμε το νευρωνικό δίκτυο να πάρουμε τη δομή του και τα βέλτιστα βάρη που προέκυψαν από τον εκπαίδευση και αλλάζοντας τα τελευταία επίπεδα αλλά κρατώντας σταθερά τα προηγούμενα να εκπαιδεύσουμε ξανά τα κορυφαία επίπεδα ώστε να προσαρμοστούν στο τελικό πρόβλημα το οποίο θέλουμε να επιλύσουμε.

Υπάρχει η δυνατότητα να "παγώσουν" ορισμένα επίπεδα του νευρωνικού δικτύου, με αποτέλεσμα τα βάρη τους να μην επηρεάζονται κατά την εκπαίδευση, το οποίο αποτελεί και μια τεχνική με ευρεία χρήση στο transfer learning. Αυτό συνήθως το επιλέγουμε για να αποφύγουμε το overfitting και να μην χαλάσουμε τη γνώση που έχει ήδη αποκτήσει το δίκτυό μας. Ανάλογα με το πλήθος των δεδομένων αλλά και τη συνάφεια του προβλήματος και των δεδομένων με το αρχικό πρόβλημα, αποφασίζεται το βάθος της επανεκπαίδευσης το δικτύου. Σε περίπτωση όπου έχουμε αρκετά μεγάλο σύνολο δεδομένων μπορούμε να εκπαιδεύσουμε μεγάλο μέρος του δικτύου ή ακόμη κι ολόκληρο το δίκτυο, και να χρησιμοποιήσουμε την υπάρχουσα γνώση ως αρχικοποίηση. Στην περίπτωση όμως που δεν έχουμε ικανοποιητικό πλήθος δεδομένων, μια βαθιά επανεκπαίδευση θα μπορούσε να οδηγήσει σε overfitting όπως αναφέρθηκε παραπάνω, και να καταστρέψει τη γνώση που έχει μεταφερθεί, αλλάζοντας προς λάθος κατευθύνσεις τα βάρη των επιπέδων. Επίσης στην περίπτωση όπου το πρόβλημα που καλούμαστε να επιλύσουμε δεν έχει μεγάλη συνάφεια με το αρχικό πρόβλημα από το οποίο αντλήσαμε τη "γνώση", τότε κρίνεται καλύτερη η εκπαίδευση περισσότερων επιπέδων εκτός του επιπέδου ταξινόμησης που βρίσκεται στην κορυφή. Το βάθος της εκπαίδευσης από εκεί και πέρα καθορίζεται κι από το πλήθος των δεδομένων όπως αναφέρθηκε παραπάνω, υπάρχει όμως μια προτίμηση προς βαθύτερες εκπαιδεύσεις. Αντίθετα, σε περιπτώσεις όπου το πρόβλημα είναι πολύ παρεμφερές με το αρχικό, επιλέγουμε εκπαίδευση μόνο του κορυφαίου επιπέδου ή και πολύ λίγων ακόμη επιπέδων, διότι δεν κρίνεται απαραίτητη η περαιτέρω εκπαίδευση μιας κι όλη η γνώση που απαιτείται από τα χαμηλότερα επίπεδα είναι παρούσα ήδη. Για παράδειγμα η χρήση ενός δικτύου όπου μπορεί να αναγνωρίσει ζώα για τη δημιουργία ενός μοντέλου το οποίο να αναγνωρίζει γάτες και σκύλους δεν απαιτεί σημαντική εκπαίδευση, ειδικά στην περίπτωση όπου στα δεδομένα του αρχικού μοντέλου υπήρχαν και γάτες και σκύλοι, διότι η γνώση για την αναγνώριση ενός σκύλου και μιας γάτας υπάρχει ήδη στο δίκτυο, το μόνο που χρειάζεται είναι να προσαρμόσουμε το πλήθος των κλάσεων στις οποίες χωρίζει τα δεδομένα ο ταξινομητής μας και να εκπαιδεύσουμε το συγκεκριμένο επίπεδο που επιτελεί αυτή τη δουλειά. Αντίθετα το πρόβλημά μας δεν είναι τόσο κοντά στην αναγνώριση εικόνας, για αυτό το λόγο όπως θα αναλυθεί και παρακάτω επιλέξαμε κατά κύριο λόγο μια βαθύτερη εκπαίδευση των μοντέλων μας.

Η βασική λογική της μεταφοράς γνώσης είναι πως για να μάθει κανείς να επιτελεί ένα πολύπλοκο έργο, πρέπει να έχει μάθει πρώτα να επιλύει τα πιο απλά προβλήματα από τα οποία απαρτίζεται το τελικό έργο. Στην περίπτωση της εικόνας που έχουμε εμείς, για να καταφέρει ένα δίκτυο

να αναγνωρίσει αντικείμενα, πρέπει να μάθει πρώτα να αναγνωρίζει βασικά χαρακτηριστικά αυτών όπως είναι οι ακμές και τα χρώματα. Τα πρώτα επίπεδα ενός νευρωνικού δικτύου το οποίο έχει εκπαιδευτεί για αναγνώριση εικόνας λοιπόν, έχουν εκπαιδευτεί να αναγνωρίζουν βασικά τέτοια χαρακτηριστικά τα οποία είναι χρήσιμα σχεδόν σε οποιοδήποτε πρόβλημα αναγνώρισης εικόνων, όπως είναι και το δικό μας. Σαν παρεμφερές πρόβλημα από το οποίο θα αντλήσουμε τη γνώση επιλέξαμε την αναγνώριση εικόνας για το λόγο ότι αποτελεί ένα πολύ γενικότερο πρόβλημα από το οποίο θα μπορούσαμε να μάθουμε τα βασικά χαρακτηριστικά που απαιτούνται και για την αναγνώριση τεχνοτροπίας αλλά και για τον κυριότερο λόγο ότι αποτελεί ένα πρόβλημα με τεράστιο όγκο δεδομένων, με αποτέλεσμα να μπορούν να εκπαιδευτούν πολύ πιο αποτελεσματικά μοντέλα. Συγκεκριμένα επιλέξαμε να χρησιμοποιήσουμε τα βάρη από δίκτυα τα οποία εκπαιδεύτηκαν στο ImageNet. Το ImageNet αποτελεί μια βάση δεδομένων με εκατομμύρια εικόνες χωρισμένες σε χιλιάδες κατηγορίες. Με βάση αυτό το σύνολο δεδομένων έχει δημιουργηθεί ο διαγωνισμός ImageNet Large-Scale Visual Recognition Challenge (ILSVRC)⁸, στον οποίο επιλέγονται από το σύνολο δεδομένων ImageNet 1.000 κατηγορίες με πάνω από 1.000.000 εικόνες για εκπαίδευση, 50.000 για validation και 150.000 για testing, πάνω στις οποίες εκπαιδεύεται κι αξιολογείται κάθε δίκτυο^{9,10}.

Τα τελευταία χρόνια φτάσαμε στο εντυπωσιακό σημείο να ξεπεράσουμε με τη χρήση CNN τις ανθρώπινες επιδόσεις στην αναγνώριση εικόνας. Θα χρησιμοποιήσουμε λοιπόν τα βάρη δικτύων τα οποία έχουν εκπαιδευτεί στο ImageNet ως βάση για την ανάπτυξη μοντέλων αναγνώρισης τεχνοτροπίας. Το Keras¹¹ μάλιστα παρέχει έτοιμα αυτά τα βάρη για πολλές αρχιτεκτονικές CNN κάνοντας το έργο μας πολύ ευκολότερο.

⁸ www.image-net.org/challenges/LSVRC/

⁹ "VGG Neural Networks: The Next Step After AlexNet", Jerry Wei, towardsdatascience.com/vgg-neural-networks-the-next-step-after-alexnet-3f91fa9ffe2c

¹⁰ "A Gentle Introduction to the ImageNet Challenge (ILSVRC)", Jason Brownlee, machinelearningmastery.com/introduction-to-the-imagenet-large-scale-visual-recognition-challenge-ilsvrc/targetText=The%20ImageNet%20Large%20Scale%20Visual%20Recognition%20Challenge%2C%20or%20ILSVRC%2C%20is,of%2Dthe%2Dart%20of%20the%20challenge

¹¹ "Keras: The Python Deep Learning library", keras.io/

Κεφάλαιο 3

Σχεδιασμός και Υλοποίηση

Στο παρόν κεφάλαιο αναλύονται οι σχεδιαστικές επιλογές και τα βήματα υλοποίησης που εκτελέστηκαν κατά την ανάπτυξη της διπλωματικής εργασίας. Ο σχεδιασμός κι η υλοποίηση της διπλωματικής εργασίας εξελίχθηκε σε 4 βασικά στάδια: α) την εύρεση κι επεξεργασία δεδομένων, β) τη σχεδίαση του προγράμματος, γ) τη δημιουργία ενός περιβάλλοντος για τη διεξαγωγή των πειραμάτων και δ) την ανάπτυξη και υλοποίηση του προγράμματος. Παρακάτω παρουσιάζονται αναλυτικότερα τα στάδια αυτά παραθέτοντας μαζί τα στάδια β) και δ) τα οποία αφορούν το σχεδιασμό και την υλοποίηση των πειραμάτων.

3.1 Δεδομένα

Η επιλογή των κατάλληλων δεδομένων είναι από τους σημαντικότερους παράγοντες, αν όχι ο πιο σημαντικός, 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. Για ένα πρόγραμμα επιτηρούμενης μάθησης, όπως αυτό που θα αναπτύξουμε, τα δεδομένα μας θέλουμε να είναι πρωτίστως έγκυρα και ορθά κατηγοριοποιημένα. Ένα αποτέλεσμα δεν έχει καμία αξία εάν έχει βασιστεί σε ψευδή ή λάθος δεδομένα. Σημαντικό για το πρόγραμμά μας είναι και το πλήθος των δεδομένων, μιας και όσο μεγαλύτερο είναι αυτό, τόσο πιο έμπιστο και στιβαρό γίνεται το μοντέλο μας, και τόσο καλύτερα μπορούμε να γενικεύσουμε τα αποτελέσματά μας σε νέα δεδομένα, χωρίς βέβαια αυτό να αποτελεί το μόνο παράγοντα. Είναι χρήσιμο ένα σύνολο δεδομένων να είναι ισορροπημένο, να υπάρχει δηλαδή ισοκατανομή των στοιχείων στις κλάσεις του. Σε περίπτωση βέβαια που δεν υπάρχει η εν λόγω ισορροπία έχουν αναπτυχθεί εργαλεία και τεχνικές για την αντιμετώπιση του προβλήματος, μερικές από τις οποίες αναφέρονται παρακάτω. Φυσικά κανένα από τα παραπάνω χαρακτηριστικά δεν αποτελεί γνώμονα από μόνο του. Χρειάζεται ένας συνδυασμός τους για να θεωρηθούν τα δεδομένα επαρκή για την ανάπτυξη ενός αξιόπιστου μοντέλου μηχανικής μάθησης.

3.1.1 Σύνολα Δεδομένων - Datasets

Η εργασία μας πραγματεύεται την ταξινόμηση έργων τέχνης σύμφωνα με την τεχνοτροπία τους. Για να επιτελέσουμε αυτό το έργο στα πλαίσια της επιβλεπόμενης μάθησης, χρειαζόμαστε ένα σύνολο δεδομένων με έργα τέχνης, ή χαρακτηριστικά αυτών, χωρισμένα σε κατηγορίες ανάλογα με την τεχνοτροπία. Υπενθυμίζουμε πως στην επιβλεπόμενη μάθηση πρέπει όλα τα δεδομένα να είναι ήδη κατηγοριοποιημένα ώστε να μπορεί να εκπαιδευτεί το μοντέλο μας και στη συνέχεια να εξαχθούν συμπεράσματα για τη λειτουργικότητά του και την εγκυρότητα των αποτελεσμά-

Πίνακας 3.1: Το Σύνολο Δεδομένων Pandora18k

Κίνημα	Πλήθος Πινάκων	Ημερομηνίες Κλειδί
Byzantin Iconography	847	500-1400
Early Renaissance	751	1280-1450
Northern Renaissance	821	1497-1550
High Renaissance	832	1490-1527
Baroque	990	1590-1725
Rococo	832	1650-1850
Romanticism	895	1770-1880
Realism	1199	1840-1880
Impressionism	1257	1860-1950
Post Impressionism	1276	1860-1925
Expressionism	1027	1905-1925
Symbolism	1057	1850-1900
Fauvism	719	1905-1908
Cubism	1227	1907-1920
Surrealism	1072	1920-1940
Abstract Art	1063	1910-σήμερα
Naive Art	1053	1890-1950
Pop Art	1120	1950-1969

των. Επιλέξαμε να χρησιμοποιήσουμε δύο διαφορετικά σύνολα δεδομένων, το Pandora18k¹ και το wikiart² τα οποία περιγράφονται παρακάτω. Μεταξύ των δύο συνόλων δεδομένων υπάρχουν επικαλύψεις μιας και μεγάλο μέρος του Pandora18k έχει αντληθεί από το wikiart, όμως παρακάτω γίνεται σαφές γιατί επιλέξαμε να δουλέψουμε και με τα δύο.

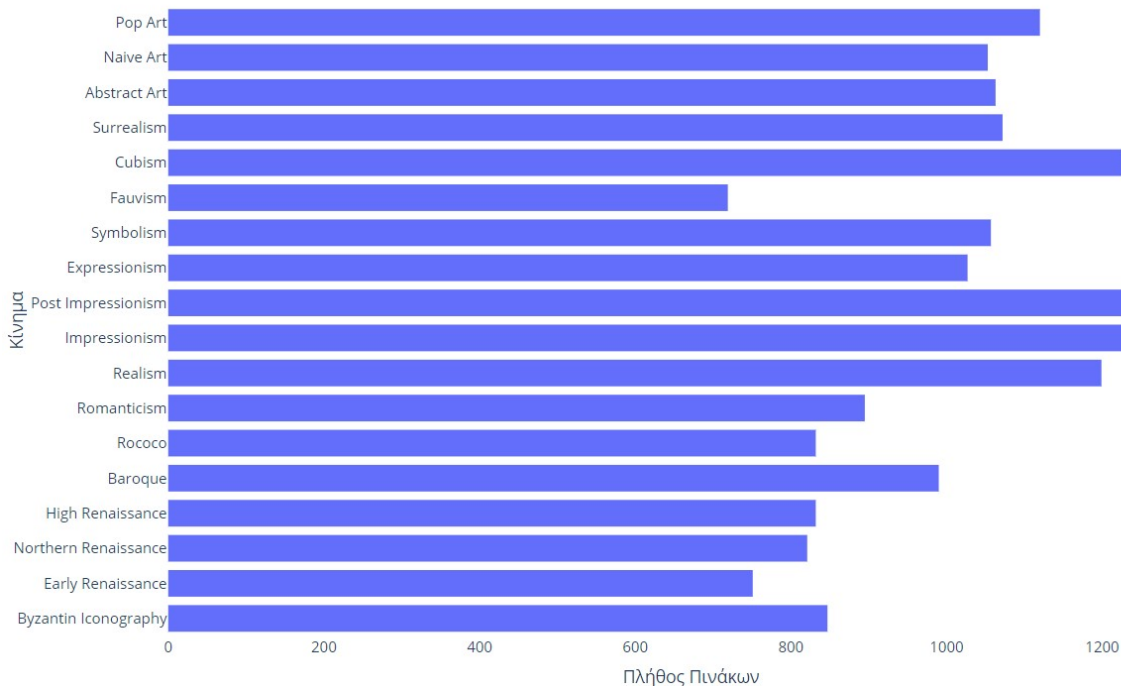
3.1.1.1 Pandora18k

Το σύνολο δεδομένων Pandora18k είναι ένα σύνολο από έργα τέχνης το οποίο συλλέχτηκε και κατηγοριοποιήθηκε από μια ομάδα του εργαστηρίου επεξεργασίας κι ανάλυσης εικόνας του πολυτεχνείου του Βουκουρεστίου(Florea et al., 2017).

Το συγκεκριμένο σύνολο δεδομένων έχει ελεγχθεί από μηχανικούς για να επιβεβαιωθεί πως μόνο το συναφές κομμάτι του πίνακα εμφανίζεται στις εικόνες (κι όχι κορνίζες, τοίχος ή οτιδήποτε άλλο δεν αποτελεί μέρος του έργου τέχνης) κι επίσης έχει ελεγχθεί από ειδικούς σε θέματα τέχνης για να διασφαλιστεί πως η κατηγοριοποίηση των πινάκων είναι ορθή. Αποτελείται από 18038 πίνακες χωρισμένους σε δεκαοκτώ διαφορετικές κατηγορίες ανάλογα με το καλλιτεχνικό κίνημα στο οποίο ανήκουν. Οι κατηγορίες είναι οι εξής: Byzantin Iconography, Early Renaissance, Northern Renaissance, High Renaissance, Baroque, Rococo, Romanticism, Realism, Impressionism, Post Impressionism, Expressionism, Symbolism, Fauvism, Cubism, Surrealism, Abstract Art, Naive Art και Pop Art. Το συνολικό μέγεθος των αρχείων είναι 1,34 GB. Το πλήθος των έργων τέχνης που αντιστοιχούν σε κάθε κίνημα, καθώς και μια προσέγγιση για την εποχή στην οποία αναπτύχθηκε και έζησε το κάθε κίνημα(Little, 2004), φαίνονται στον παρακάτω πίνακα.

¹ "Pandora Database", imag.pub.ro/pandora/pandora_download.html

² WikiArt, Visual Art Encyclopedia, www.wikiart.org



Σχήμα 3.1: Κατανομή εικόνων ανά κατηγορία στο σύνολο δεδομένων Pandora18k

Αξίζει να σημειωθεί ότι η σειρά με την οποία παρουσιάζονται τα διάφορα καλλιτεχνικά κινήματα είναι χρονολογική, εάν λάβουμε και υπόψιν το χρονικό περιθώριο λάθους που έχουν οι παραπάνω εκτιμήσεις. Επειδή πολλές φορές καλλιτεχνικά κινήματα που συνυπήρχαν χρονικά ή/και χωρικά αλλά και κινήματα όπου γεννήθηκαν κι ακολούθησαν χρονικά άλλα, έχουν πολλά κοινά χαρακτηριστικά τόσο στην όψη όσο κι στη θεματολογία, αναμένουμε να είναι πιο δύσκολος ο διαχωρισμός τους. Το γεγονός πως τα κινήματα είναι με αυτή τη σειρά λοιπόν θα φανεί χρήσιμο στην οπτικοποίηση των αποτελεσμάτων παρακάτω και πιο συγκεκριμένα στους πίνακες σύγχυσης (Confusion Matrices).

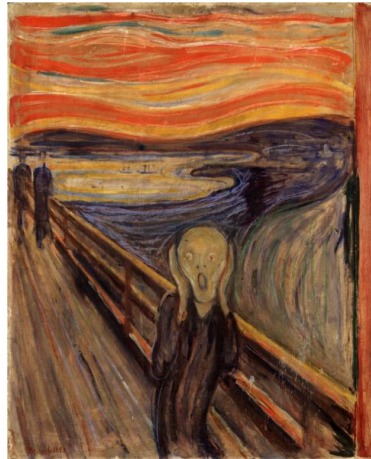
Όπως παρατηρούμε, κι από το σχετικό ραβδόγραμμα στο σχήμα 3.1, αν και υπάρχουν οριακές κατηγορίες οι οποίες έχουν ορατή διαφορά στο πλήθος των εικόνων που τους αντιστοιχούν, σε γενικές γραμμές υπάρχει ισοκατανομή των έργων τέχνης, έχουμε λοιπόν ένα ισορροπημένο σύνολο δεδομένων.

Επιλέξαμε να εργαστούμε με το συγκεκριμένο σύνολο δεδομένων για τους τρεις ακόλουθους βασικούς λόγους: α) διότι έχει ελεγχθεί τόσο ως προς την εγκυρότητα, όσο και ως προς την ποιότητα των δεδομένων του, όπως ήδη αναφέρθηκε, β) διότι έχουν υπάρξει πολλές εργασίες οι οποίες επιχειρήσαν αναγνώριση τεχνοτροπίας πάνω στο ίδιο σύνολο δεδομένων άρα έχουμε ένα μέτρο σύγκρισης για τα αποτελέσματά μας, και γ) επειδή αποτελεί ένα μικρότερο σύνολο δεδομένων από την δεύτερη επιλογή μας, καταλαμβάνοντας σημαντικά μικρότερο μέρος μνήμης, καθιστώντας την εκπαίδευσή του νευρωνικού δικτύου σημαντικά γρηγορότερη κι επιτρέποντας τη χρήση υπολογιστικών περιβαλλόντων με λιγότερη μνήμη.

Στη συνέχεια παρουσιάζονται μερικά έργα τέχνης από το σύνολο δεδομένων.



Σχήμα 3.2: "Lady with an Ermine", Leonardo da Vinci, 1490, High Renaissance



Σχήμα 3.3: "The Scream", Edvard Munch, 1893, Expressionism



Σχήμα 3.4: "Still life with skull of ox", Pablo Picasso, 1942, Post-Impressionism



Σχήμα 3.5: "The Persistence of Memory", Salvador Dali, 1931, Surrealism



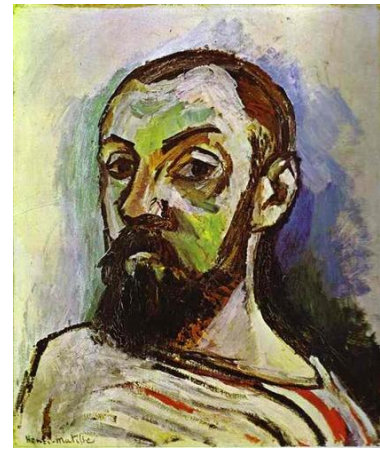
Σχήμα 3.6: "The Starry Night", Vincent van Gogh, 1889, Post-Impressionism



Σχήμα 3.7: "Judith beheading Holoferne", Caravaggio, 1599, Baroque

3.1.1.2 Wikiart

Το σύνολο δεδομένων wikiart (Karayev et al., 2013; Tan et al., 2018) αποτελείται από έργα τέχνης τα οποία έχουν αντληθεί από την ιστοσελίδα "wikiart.org", όπου αποτελεί μια διαδικτυακή ψηφιακή εγκυκλοπαίδεια έργων τέχνης από μουσεία, πανεπιστήμια, δημαρχεία και άλλα δημόσια κτίρια. Στο διαδίκτυο κυκλοφορούν αρκετά σύνολα δεδομένων όπου προέρχονται από τη συγκεκριμένη ιστοσελίδα. Το σύνολο που χρησιμοποιήσαμε εμείς προέρχεται από το αποθετήριο των Tan et al. (2018) που συνοδεύει την εργασία τους "Improved ArtGAN for Conditional Synthesis of Natural Image and Artwork". Περιέχει πάνω από ογδόντα χιλιάδες εικόνες από ζωγραφικούς πίνακες, ταξινομημένες σε είκοσι επτά κατηγορίες, σύμφωνα με την τεχνοτροπία τους. Το συνολικό μέγεθος των αρχείων είναι 25,7 GB. Οι κατηγορίες στις οποίες είναι χωρισμένες οι εικόνες είναι οι εξής: Abstract Expressionism, Action Painting, Analytical Cubism, Art Nouveau (Modern), Baroque, Color Field Painting, Contemporary Realism, Cubism, Early Renaissance, Expressionism, Fauvism, High Renaissance, Impressionism, Mannerism (Late Renaissance), Minimalism, Naive Art (Primitivism), New Realism, Northern Renaissance, Pointillism, Pop Art, Post Impressionism, Realism, Rococo, Romanticism, Symbolism, Synthetic Cubism και Ukiyo-e. Στον πίνακα 3.2 παρουσιάζεται



Σχήμα 3.8: "L'air", Joan Miro, 1937, Abstract Art

Σχήμα 3.9: "Cow", Andy Warhol, 1966, Pop Art

Σχήμα 3.10: "Self-Portrait in a Striped T-shirt", Henri Matisse, 1906, Fauvism

η κατανομή των εικόνων ανά κατηγορία. Όπως παρατηρούμε κι από το παρακάτω ραβδόγραμμα του σχήματος 3.11, σε αντίθεση με το προηγούμενο σύνολο δεδομένων, το wikiart είναι εκτός ισορροπίας. Έχει κλάσεις με πλήθος στοιχείων που ξεκινάει κάτω από τα εκατό (Action Painting) και φτάνει μέχρι και πάνω από δεκατρείς χιλιάδες (Impressionism).

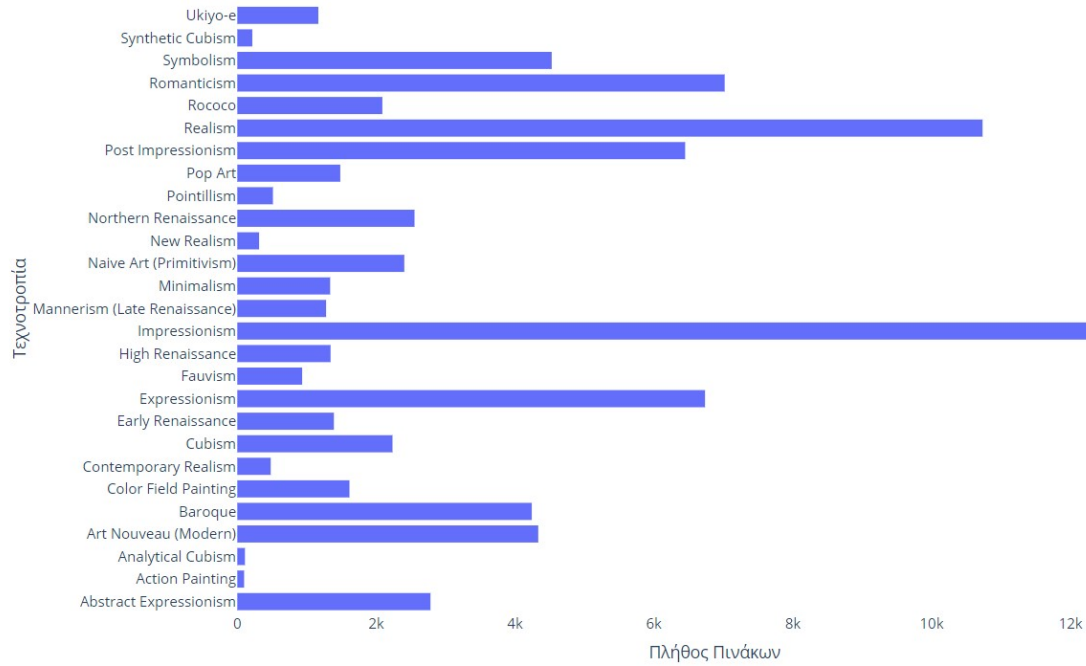
Το πρόβλημα με τα μη ισορροπημένα σύνολα δεδομένων είναι πως ο ταξινομητής έχει μια τάση να προβλέπει πολύ συχνότερα τις κλάσεις από τις οποίες έχει περισσότερα δεδομένα μιας και με αυτό τον τρόπο έχει μεγαλύτερη ακρίβεια (accuracy). Για παράδειγμα σε ένα πρόβλημα δυαδικής ταξινόμησης (binary classification) όπου το σύνολο δεδομένων αποτελείται κατά 99% από την κλάση A και το υπόλοιπο 1% ανήκει στην κλάση B, τότε ένας ταξινομητής όπου προβλέπει πάντα την κλάση A θα έχει ακρίβεια 99%. Στην ταξινόμηση ενός τέτοιου συνόλου δεδομένων είναι πολύ συχνό φαινόμενο, ο ταξινομητής να ταξινομεί δεδομένα της κλάσης B ως κλάση A διότι αυτό δεν του κοστίζει πολύ σε ακρίβεια, την οποία συνήθως προσπαθεί να βελτιστοποιήσει. Να επισημάνουμε πως το φαινόμενο όπου ένας ταξινομητής μαθαίνει πολύ καλά τα χαρακτηριστικά μιας κλάσης αλλά όχι μιας άλλης επειδή η δεύτερη έχει πολύ λιγότερα δεδομένα, με αποτέλεσμα να μην γενικεύει σωστά σε νέα δεδομένα, δεν αποτελεί πρόβλημα μη ισορροπημένου συνόλου δεδομένων αλλά ελλιπών δεδομένων. Για παράδειγμα εάν έχω ένα πρόβλημα ταξινόμησης εικόνων σε ξυλοκόπους και ναυτικούς, και έχω πολλές εικόνες ξυλοκόπων αλλά πολύ λίγες ναυτικών, όπου στις περισσότερες εικόνες οι ναυτικοί καπνίζουν μια πίπα, τότε μπορεί ο ταξινομητής μου να θεωρήσει πως όποιος καπνίζει πίπα είναι ναυτικός με αποτέλεσμα να οδηγηθεί σε μια τελείως λάθος κατεύθυνση ως προς το χαρακτηριστικό ταξινόμησης. Αυτό δεν αποτελεί παράδειγμα του προβλήματος μη ισορροπημένων κλάσεων αλλά τα περιορισμένα δεδομένα που του είναι διαθέσιμα για κάποια κλάση δεν επιτρέπουν την εκμάθηση των σωστών χαρακτηριστικών.

Επειδή όμως σε πραγματικά προβλήματα η ανισορροπία των δεδομένων είναι ένα πολύ συχνό φαινόμενο, δημιουργήθηκαν κάποιοι τρόποι αντιμετώπισης του προβλήματος. Αρχικά, μιας

Πίνακας 3.2: Το Σύνολο Δεδομένων WikiArt

Τεχνοτροπία	Πλήθος Πινάκων
Abstract Expressionism	2782
Action Painting	98
Analytical Cubism	110
Art Nouveau (Modern)	4334
Baroque	4241
Color Field Painting	1615
Contemporary Realism	481
Cubism	2235
Early Renaissance	1391
Expressionism	6736
Fauvism	934
High Renaissance	1343
Impressionism	13060
Mannerism (Late Renaissance)	1279
Minimalism	1337
Naive Art (Primitivism)	2405
New Realism	314
Northern Renaissance	2552
Pointillism	513
Pop Art	1483
Post Impressionism	6450
Realism	10733
Rococo	2089
Romanticism	7019
Symbolism	4528
Synthetic Cubism	216
Ukiyo-e	1167

και όπως είδαμε σε αυτές τις περιπτώσεις το accuracy δεν αποτελεί τον κατάλληλο γνώμονα για την επίδοση ενός μοντέλου, η χρήση άλλων μετρικών, όπως precision και f1, για τη βελτιστοποίηση και αξιολόγηση του μοντέλου μπορεί να οδηγήσει σε καλύτερα αποτελέσματα, όπου το καλύτερα βέβαια πάντα εξαρτάται από το πρόβλημα. Μια άλλη κατεύθυνση αντιμετώπισης του προβλήματος είναι η προσπάθεια εξισορρόπησης των κλάσεων. Αυτό μπορεί να επιτευχθεί τόσο με υπο-δειγματοληψία (under-sampling) όσο και με υπερ-δειγματοληψία (over-sampling). Πρόκειται για τις δύο όψεις του ίδιου νομίσματος. Κατά την υπο-δειγματοληψία επιλέγονται τυχαία ορισμένα δεδομένα από τις κατηγορίες όπου έχουν περισσότερα δεδομένα ώστε να χρησιμοποιηθούν στην εκπαίδευση. Με αυτό τον τρόπο μειώνεται το πλήθος των δεδομένων τους, οδηγώντας το σύνολο σε μια πιο ισορροπημένη κατάσταση. Αυτή η τεχνική επιλέγεται συνήθως στις περιπτώσεις όπου και οι κλάσεις με τα λιγότερα δεδομένα, έχουν αρκετά για την εκπαίδευση, κι απλώς κάποιες κλάσεις έχουν υπερβολικά πολλά. Κατά την υπερ-δειγματοληψία από την άλλη, δημιουργούμε αντίγραφα των δεδομένων των κλάσεων όπου έχουν λιγότερα δεδομένα. Μερικές φορές κάνουμε και κάποια παραμόρφωση ή τροποποίηση η οποία να μην επηρεάζει την κατηγοριοποίησή τους, ώστε να μην βλέπει ο ταξινομητής πολλές φορές τα ίδια δεδομένα. Αυτή η τεχνική, ενέχει τον κίνδυνο του overfitting διότι ο ταξινομητής μπορεί να προσαρμοστεί πέραν του επιθυμητού στα δεδομένα



Σχήμα 3.11: Κατανομή εικόνων ανά κατηγορία στο σύνολο δεδομένων Wikiart

εκπαίδευσης λόγω του ότι θα βλέπει τα ίδια δεδομένα, έστω και με μικρές αλλαγές, προσαρμόζοντας τα βάρη του σε συγκεκριμένα τους χαρακτηριστικά. Αυτή η μέθοδος επιλέγεται συνήθως όταν οι κατηγορίες με λιγότερα δεδομένα έχουν πολύ μικρό πλήθος στοιχείων, όπου σε αυτή την περίπτωση η μείωση του μεγέθους των κλάσεων που υπερέχουν σε πλήθος, ενέχει τον κίνδυνο να προκύψουν προβλήματα που σχετίζονται με τη γενικότερη έλλειψη δεδομένων.

Στη συγκεκριμένη εργασία αν και όπως είδαμε το σύνολο δεδομένων wikiart έχει εμφανή ανισορροπία, δεν προχωρήσαμε στην εφαρμογή υπερ ή υπο-δειγματοληψίας. Χρησιμοποιήσαμε και σε αυτή την περίπτωση τη μετρική accuracy ως βασικό μέτρο αξιολόγησης διότι κι όλες οι προηγούμενες εργασίες είχαν πράξει το ίδιο, άρα έπρεπε να έχουμε κάποιο μέτρο σύγκρισης μαζί τους. Δώσαμε όμως περισσότερη σημασία στις άλλες μετρικές που αναφέρθηκαν παραπάνω και παρατηρήσαμε πως αν και μη ισορροπημένο το σύνολο, μας έδινε καλές, ή τουλάχιστον συγκρίσιμες μεταξύ τους, τιμές σε όλες τις μετρικές, κι έτσι δεν κρίθηκε αναγκαία η εφαρμογή κάποιας από τις τεχνικές εξισορρόπησης του δείγματος. Παρατηρώντας όμως το σύνολο δεδομένων παρατηρούμε πως ορισμένες κλάσεις έχουν ιδιαίτερη συνάφεια, όπως Cubism, Analytical Cubism και Synthetic Cubism. Οι δύο τελευταίες κλάσεις αποτελούν μέρος του γενικότερου κινήματος του κυβισμού. Συγκεκριμένα ο αναλυτικός κυβισμός αποτελεί τη δεύτερη περίοδο του κινήματος του κυβισμού όπου διήρκεσε από το 1910 μέχρι και το 1912 της οποίας ηγήθηκαν οι Pablo Picasso και Georges Braque³, και ο συνθετικός κυβισμός αφορά μια περίοδο του κινήματος του κυβισμού όπου διήρκεσε από το 1912 έως και το 1914, γεννήθηκε από τον αναλυτικό κυβισμό με τον οποίο είχε τους ίδιους εμπνευστές και καθοδηγητές⁴. Παρατηρούμε λοιπόν πως και οι τρεις κλάσεις αφορούν το ίδιο καλλιτεχνικό κίνημα. Για αυτό το λόγο αποφασίσαμε αφού εκτελέσουμε ορισμένα πειράματα με το σύνολο

³ "What Is Analytic Cubism in Art?", Beth S. Gersh-Nesic, www.thoughtco.com/analytical-cubism-183189

⁴ "What Is the Definition of Synthetic Cubism?", Beth S. Gersh-Nesic, www.thoughtco.com/synthetic-cubism-definition-183242

δεδομένων ως έχει, να ενώσουμε τις 3 αυτές κατηγορίες κάτω από την κλάση Cubism. Μια ακόμη κλάση με πολύ λίγα δεδομένα είναι η Action Painting. Η τεχνοτροπία του action painting όμως είναι στενά συνδεδεμένη με το κίνημα του αφηρημένου εξπρεσιονισμού, τόσο μάλιστα όπου ορισμένοι κριτικοί τέχνης χρησιμοποιούν τους δύο όρους εναλλάξιμα⁵. Επειδή λοιπόν η συγκεκριμένη τεχνοτροπία ανήκει σε μια γκρίζα ζώνη και έχουμε πολύ λίγα δεδομένα συγκριτικά με το σύνολο, επιλέξαμε να την απαλείψουμε από το σύνολο δεδομένων μας, διαγράφοντας τις εικόνες της. Θα μπορούσαμε να εντάξουμε τους πίνακες της στην κλάση του αφηρημένου εξπρεσιονισμού, όμως επειδή δεν υπάρχει κάποια ξεκάθαρη άποψη στους επιστημονικού κύκλους για το κατά πόσο είναι η ίδια τεχνοτροπία και δεν πρόκειται για μεγάλο μέρος των δεδομένων, επιλέξαμε τη διαγραφή τους για να μην διακινδυνεύσουμε την εγκυρότητα των δεδομένων η οποία είναι ήδη αμφισβητήσιμη από ορισμένους (Sandoval et al., 2019) λόγω του ότι δεν έχουν ελεγχθεί από ειδικούς. Σε αντίστοιχη λογική και για να εξισορροπήσουμε το δείγμα μας αλλά και να δημιουργήσουμε ένα σύνολο δεδομένων με το οποίο θα μπορούμε να συγκρίνουμε τα αποτελέσματα με τη δουλειά άλλων, διαγράψαμε και τις κλάσεις New Realism, Contemporary Realism και Pointillism σε αντιστοιχία με τη δουλειά των Sandoval et al. (2019). Έτσι δημιουργήθηκε ένα νέο σύνολο δεδομένων βασισμένο και επί των πλείστων ίδιο με το wikiart, το οποίο για λόγους ευκολίας θα ονομάσουμε wikiart2.

Επιλέξαμε να εργαστούμε με το συγκεκριμένο σύνολο δεδομένων για τους ακόλουθους λόγους: α) διότι αποτελεί το μεγαλύτερο σύνολο δεδομένων με έργα τέχνης σε ψηφιακή μορφή αυτή τη στιγμή διαθέσιμο στο ευρύ κοινό, και β) διότι σχεδόν όλες οι εργασίες όπου έχουν γίνει σχετικά με το πρόβλημα όπου καλούμαστε να επιλύσουμε, κι έχουν δημοσιευθεί μετά τη δημιουργία του συνόλου, έχουν κάνει χρήση του wikiart τουλάχιστον σε ένα κομμάτι της εργασίας, δίνοντάς μας έτσι τη δυνατότητα να έχουμε πολλές διαφορετικές εργασίες για να συγκρίνουμε τα αποτελέσματά μας.

Πίνακας 3.3: Σύνοψη Συνόλων Δεδομένων

Όνομα	Πλήθος Εικόνων	Κλάσεις	Συνολική Μνήμη
Pandora18k	18038	18	1,34 GB
Wikiart	81445	27	25,7 GB
Wikiart2	80039	21	25,2 GB

3.1.2 Προ-επεξεργασία Δεδομένων

Τα σύνολα δεδομένων αποτελούνται από φακέλους στους οποίους είναι χωρισμένες οι εικόνες ανά κατηγορία. Κάθε κλάση - τεχνοτροπία αντιστοιχεί σε ένα φάκελο. Τα εργαλεία που μας δίνονται από τη βιβλιοθήκη Keras⁶ (Chollet et al., 2015) μας επιτρέπουν να αντλήσουμε την κατηγορία κάθε εικόνας από τον φάκελο στον οποίο βρίσκεται, χωρίς να χρειάζεται κάποιο επιπλέον αρχείο που να ορίζει ρητά την κατηγορία κάθε εικόνας. 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

⁵ "Art History Definition: Action Painting", Beth S. Gersh-Nesic, www.thoughtco.com/art-history-definition-action-painting-183188

⁶ "Keras: The Python Deep Learning library", keras.io/

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 αυτό το διαχωρισμό δεδομένων με τη χρήση του πακέτου `split-folders`⁷, μέσω της ομώνυμης εντολής. Η συγκεκριμένη εντολή παίρνει ως είσοδο έναν φάκελο με δεδομένα και ορισμένες παραμέτρους και δημιουργεί έναν καινούριο φάκελο ο οποίος περιέχει τρεις υποφακέλους, `train`, `val` και `test` με τα αντίστοιχα δεδομένα να περιέχονται στον καθένα. Η κατανομή των εικόνων στις διάφορες κατηγορίες παραμένει σταθερή στα τρία σύνολα κι εμείς επιλέξαμε να χρησιμοποιήσουμε το 80% των δεδομένων για εκπαίδευση, το 10% για validation και το υπόλοιπο 10% για Testing.

Αφού γίνει η συλλογή των δεδομένων, οι δημιουργοί του προγράμματος καλούνται να επιλέξουν τη μορφή με την οποία θα εισάγουν τα δεδομένα στο πρόγραμμα και να τα προ-επεξεργαστούν κατάλληλα. Οι σχεδιαστικές επιλογές που πάρθηκαν αναλύονται στην υπο-ενότητα 3.2.1. Στο συγκεκριμένο σημείο αναλύουμε την προ-επεξεργασία την οποία κάναμε στα δεδομένα σύμφωνα με αυτές τις επιλογές. Λόγω της απαίτησης ενός CNN να τροφοδοτείται με εικόνες σταθερών και προκαθορισμένων διαστάσεων, δημιουργήσαμε δύο αντίγραφα του Pandora18k τα οποία αποτελούνταν από τμήματα των αρχικών εικόνων μετά από περικοπή, ή προσαρμογή μεγέθους εάν δεν επιτρέπουν οι αρχικές διαστάσεις την περικοπή, στις διαστάσεις 299×299 , ώστε να μπορούν να εισαχθούν ακέραια σε ένα CNN που χρησιμοποιεί αρχιτεκτονικές βασισμένες στην Inception, οι οποίες δέχονται εικόνες στο συγκεκριμένο μέγεθος. Το πρώτο αντίγραφο του Pandora18k περιέχει ένα τυχαία επιλεγμένο τμήμα για κάθε εικόνα, ενώ το δεύτερο αντίγραφο περιέχει το κεντρικό τμήμα κάθε εικόνας.

3.2 Συνελικτικό Νευρωνικό Δίκτυο

3.2.1 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 εργαστούμε. 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 είσοδος 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. Η επιλογή λίγων χαρακτηριστικών μπορεί να στερούσε τη δυνατότητα από το μοντέλο μας να κατηγοριοποιήσει σωστά τις εικόνες λόγω λειψών δεδομένων εισόδου, αλλά η επιλογή πολλών χαρακτηρι-

⁷ pypi.org/project/split-folders/

κών μπορούσε να αποβεί αρνητική στο αποτέλεσμα λόγω της κατάρας της διαστατικότητας (Curse of dimensionality) (Bellman and Collection, 1961). Ο δεύτερος και κύριος 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 κι άλλων προ-υπολογισμένων 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.

Δεδομένης πλέον της χρήσης εικόνων ως είσοδο στο πρόγραμμα επιλέξαμε τα συνελκτικα νευρωνικά δίκτυα για την εκτέλεση της εργασίας καθώς είναι το ισχυρότερο εργαλείο που έχουμε στα χέρια μας για την επεξεργασία και αναγνώριση εικόνων. Στη συνέχεια θα αναλύσουμε τις σχεδιαστικές επιλογές που κάναμε έχοντας ως δεδομένη τη χρήση των CNN. Οι εικόνες στα σύνολα δεδομένων έχουν ποικίλα μεγέθη. Ένα CNN απαιτεί οι εικόνες που του δίνονται ως είσοδος να είναι όλες ίδιων διαστάσεων. Οι επιλογές που θα μπορούσαν να λύσουν ο συγκεκριμένο πρόβλημα κι έχουν εφαρμοσθεί και στο παρελθόν είναι κατά κύριο λόγο τρεις:

1. Περικοπή των εικόνων στο επιθυμητό μέγεθος.
2. Προσαρμογή του μεγέθους των εικόνων στις επιθυμητές διαστάσεις (Resizing).
3. Αρχικά προσαρμογή του μεγέθους των εικόνων σε κάποιο προκαθορισμένο μέγεθος και στη συνέχεια περικοπή στο επιθυμητό μέγεθος που ορίζει το νευρωνικό δίκτυο.

Η βιβλιοθήκη Keras παρέχει έναν πολύ εύκολο και απλό τρόπο για εισαγωγή εικόνων στο πρόγραμμα μέσω του ImageDataGenerator ^{*}. Πρόκειται για ένα πολύ χρήσιμο εργαλείο για την εισαγωγή και επεξεργασία εικόνων. Συγκεκριμένα η μέθοδος `flow_from_directory` του ImageDataGenerator επιτρέπει την εισαγωγή εικόνων από έναν φάκελο και την επεξεργασία κατά την εισαγωγή, διευκολύνοντας πολύ τη διαδικασία προ-επεξεργασίας των δεδομένων. Έχει όμως έναν περιορισμό, υποχρεωτικά πρέπει να προσαρμόζεται το μέγεθος κάθε εικόνας που εισάγεται σε κάποιες διαστάσεις οι οποίες ορίζονται από τον προγραμματιστή, προτού γίνει οποιαδήποτε άλλη επεξεργασία της εικόνας στα πλαίσια του προγράμματος. Αυτό δεν μπορεί να παρακαμφθεί διότι το πρόγραμμα πρέπει να γνωρίζει και να δεσμεύσει συγκεκριμένο χώρο μνήμης για κάθε εικόνα αρχικοποιώντας συγκεκριμένους πίνακες, γεγονός που δεν θα μπορούσε εάν κάθε εικόνα είχε διαφορετικό μέγεθος. Γι' αυτό το λόγο, το ίδιο το εργαλείο αλλάζει τις διαστάσεις των εικόνων που εισάγονται στις επιθυμητές τιμές. Αυτό το χαρακτηριστικό καθιστά πολύ εύκολη τη δεύτερη επιλογή που αναφέρεται παραπάνω, μιας και πρόκειται για μια διαδικασία η οποία γίνεται με μια και μόνο εντολή. Επίσης με ορισμένες τροποποιήσεις επιτρέπει και την τρίτη επιλογή μιας και ο προγραμματιστής μπορεί να ορίσει κάποιο μεγαλύτερο μέγεθος εικόνας στη συνάρτηση `flow_from_directory` και στη συνέχεια μέσω μιας συνάρτησης να περικόψει την εικόνα που έχει ήδη εισαχθεί στο πρόγραμμα, στις διαστάσεις που επιθυμεί. Είναι όμως αδύνατο να υλοποιηθεί η πρώτη επιλογή με χρήση του συγκεκριμένου εργαλείου. Ο μόνος τρόπος να υλοποιηθεί περικοπή της εικόνας χωρίς προσαρμογής μεγέθους κατά την εκτέλεση είναι ο προγραμματιστής να δημιουργήσει εξολοκλήρου έναν δικό του Data Generator στον οποίο να προσθέσει το συγκεκριμένο χαρακτηριστικό στη μέθοδο που εισάγει τα δεδομένα.

^{*} keras.io/preprocessing/image/

Η επιλογή να εργαστεί κανείς με τμήματα εικόνων βασίζεται στην υπόθεση πως η τεχνοτροπία ενός έργου τέχνης είναι εμφανής σε κάθε σημείο του κι όχι μόνο σε συγκεκριμένα, άρα ένα τμήμα ενός έργου περιέχει επαρκείς πληροφορίες ώστε ένα CNN να μπορεί να προσδιορίσει την τεχνοτροπία του. Με την αποφυγή της αλλαγής του μεγέθους του πίνακα παραμένουν άθικτα χαρακτηριστικά τα οποία μπορεί να χαθούν με τη σμίκρυνση ενός έργου τέχνης και διατηρούνται οι λεπτομέρειες σε επίπεδο pixel (Viswanathan, 2017). Από την άλλη, η επιλογή ολόκληρης της εικόνας ως είσοδο βασίζεται στην υπόθεση πως ακόμη και με την αλλαγή μεγέθους μιας εικόνας, η τεχνοτροπία της δεν αλλάζει. Μάλιστα ορισμένες καλλιτεχνικές τεχνοτροπίες έχουν και συγκεκριμένη θεματολογία με την οποία ασχολούνται, δίνοντας ένα επιπλέον χαρακτηριστικό για την αναγνώριση τους, το οποίο στις περισσότερες περιπτώσεις χάνεται με την περικοπή ενός μικρού μόνο μέρους του πίνακα.

Στην παρούσα διπλωματική επιλέξαμε να εργαστούμε τόσο με ολόκληρες εικόνες προσαρμόζοντας το μέγεθός τους κατά την εισαγωγή στο πρόγραμμα, όσο και με τμήματα εικόνων τα οποία προήλθαν μετά από περικοπή συγκεκριμένων διαστάσεων των αρχικών δεδομένων.

3.2.2 Επιλογή Αρχιτεκτονικής

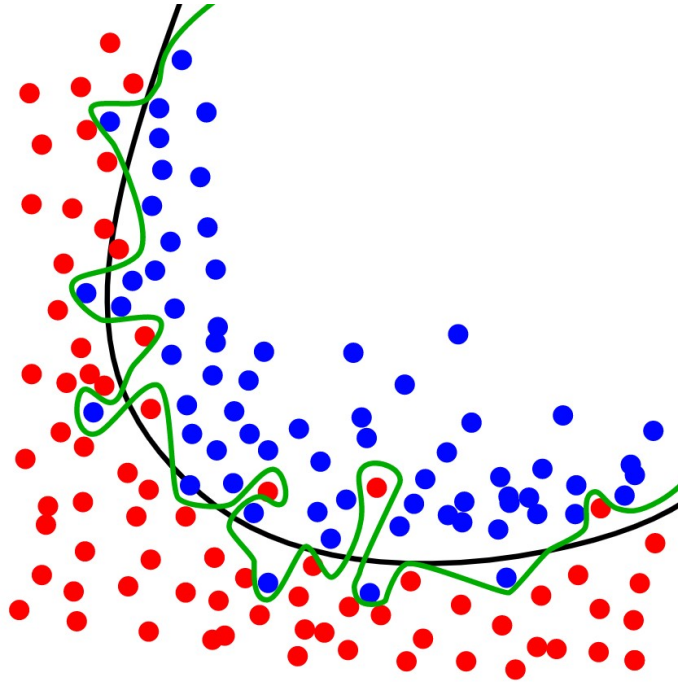
Η δεύτερη βασική επιλογή που είχαμε να κάνουμε αφορούσε την αρχιτεκτονική του συνελκτικού νευρωνικού δικτύου. Είχαμε να αποφασίσουμε αρχικά εάν θα δημιουργήσουμε κάποια δική μας αρχιτεκτονική ή θα χρησιμοποιήσουμε κάποια προϋπάρχουσα καθώς και εάν χρησιμοποιήσουμε κάποια "έτοιμη" αρχιτεκτονική ποία ή ποιές θα είναι αυτές. Στο πρώτο δίλημμα, έτοιμη ή νέα αρχιτεκτονική, η απόφαση ήταν σχετικά εύκολη, και ήταν να χρησιμοποιήσουμε αρχιτεκτονικές που υπάρχουν ήδη, διότι αυτές οι αρχιτεκτονικές έχουν σχεδιαστεί ειδικά για την αναγνώριση εικόνων από ομάδες ατόμων με εξειδίκευση στον τομέα και μετά από μήνες ή και χρόνια έρευνας. Δεν θεωρήσαμε λοιπόν πως οι γνώσεις μας είναι επαρκείς ώστε να δημιουργήσουμε από το μηδέν κάτι καλύτερο από αυτά που μας παρέχονται ήδη. Μπορεί το πρόβλημα που επιλύουμε να μην είναι ακριβώς αυτό για το οποίο σχεδιάστηκαν οι περισσότερες αρχιτεκτονικές, και πάλι όμως αφορά ανάλυση εικόνων και δεν έχουμε κάποια ένδειξη που να μας κατευθύνει προς τη δημιουργία κάποιας καλύτερης αρχιτεκτονικής για το δικό μας πρόβλημα. Μάλιστα υπάρχουν ομάδες (Yu et al., 2017) που επιχείρησαν να δημιουργήσουν νέες αρχιτεκτονικές εμπνευσμένες από τις ήδη υπάρχουσες, κατέληξαν σε καλύτερα αποτελέσματα από μερικές απλές έτοιμες μεν αλλά πολύ χειρότερα από τα βέλτιστα τα οποία επιτύχαμε στη συγκεκριμένη εργασία, επιβεβαιώνοντας τις σκέψεις μας για χρήση καθαρά και μόνο έτοιμων αρχιτεκτονικών. Τέλος, η χρήση μιας δικής μας αρχιτεκτονικής δεν θα μας έδινε τη δυνατότητα να χρησιμοποιήσουμε Transfer Learning, τεχνικής η οποία δίνει σταθερά τα βέλτιστα αποτελέσματα σχεδόν σε όλες τις εργασίες που έχει εφαρμοστεί. Αφού επιλέξαμε λοιπόν να χρησιμοποιήσουμε έτοιμες αρχιτεκτονικές, έπρεπε να επιλέξουμε ποιές θα χρησιμοποιήσουμε μιας και υπάρχει μια πολύ μεγάλη γκάμα. Επιλέξαμε κατά κύριο λόγο σύμφωνα με τις επιδόσεις τους στο αντικείμενο για το οποίο σχεδιάστηκαν, την αναγνώριση εικόνων από το σύνολο δεδομένων ImageNet.

Δοκιμάστηκαν αρκετές αρχιτεκτονικές όμως καταλήξαμε σε 4 διαφορετικές αρχιτεκτονικές στις οποίες βασίσαμε την εργασία: την αρχιτεκτονική VGG Simonyan and Zisserman (2014) και συγκεκριμένα τις VGG16 και VGG19, την αρχιτεκτονική ResNet He et al. (2015a) και συγκεκριμένα τις ResNet50, ResNet152 και ResNetV2, την αρχιτεκτονική Inception Szegedy et al. (2014) και πιο

συγκεκριμένα τις InceptionV3 (sze) και Inception-ResNetV2 (Szegedy et al., 2016) και την αρχιτεκτονική DenseNet(Huang et al., 2016) και πιο συγκεκριμένα τη DenseNet121. Κάθε μία από αυτές έχει συγκεκριμένα χαρακτηριστικά τα οποία τη διακρίνουν από όλες τις υπόλοιπες όπως για παράδειγμα τα δίκτυα ResNet είναι σαν να δημιουργούνται από πολλά επιμέρους δίκτυα και έχουν skip-connections ή short-cuts για να παρακάμπτονται επίπεδα σε ορισμένα σημεία. Και οι τρεις αρχιτεκτονικές δημιουργούν πολύ ισχυρά μοντέλα για την αναγνώριση εικόνων και παρέχονται από τη βιβλιοθήκη Keras. Μάλιστα μέσω του Keras παρέχονται και τα βάρη για τα παραπάνω δίκτυα, από την εκπαίδευσή τους στο ImageNet. Αυτό καθιστά τη χρήση της τεχνικής Transfer Learning πολύ πιο εύκολη, όμως θα αναφερθούμε αναλυτικότερα σε αυτό παρακάτω. Δεν εργαστήκαμε μόνο με τα καλύτερα διαθέσιμα δίκτυα, αλλά δοκιμάσαμε και δίκτυα τα οποία δεν είχαν τόσο καλές επιδόσεις στην αναγνώριση εικόνας για δύο βασικούς λόγους: Πρώτον, όπως έχει αναφερθεί και σε άλλες εργασίες όπου έχουν ασχοληθεί με την αναγνώριση τέχνης, οι κορυφαίες αρχιτεκτονικές στην αναγνώριση εικόνας, δεν είναι πάντοτε και κορυφαίες στην αναγνώριση τέχνης (Lecoutre et al., 2017). Δεύτερον, θέλαμε να έχουμε και μια σύγκριση μεταξύ των διαφορετικών αρχιτεκτονικών, αλλά και να ξεκινήσουμε εκπαιδεύοντας κάποια πιο απλά και κατ' επέκταση λιγότερο χρονοβόρα δίκτυα, και στη συνέχεια να προχωρήσουμε στην εκπαίδευση μεγαλύτερων CNN. Στο μεγαλύτερο μέρος των πειραμάτων έγινε χρήση της αρχιτεκτονικής InceptionV3 διότι αυτή ήταν που είχε δώσει τα καλύτερα αποτελέσματα τη δεδομένη στιγμή και έπρεπε να επιλέξουμε μια αρχιτεκτονική που θα κρατήσουμε σταθερή ώστε να ελέγξουμε την επιρροή και των υπόλοιπων υπερπαραμέτρων στην ακρίβεια προβλέψεων του μοντέλου.

3.2.3 Επιλογή και Βελτιστοποίηση Υπερπαραμέτρων

Μετά την επιλογή της αρχιτεκτονικής του CNN, έπρεπε να επιλέξουμε τις υπερπαραμέτρους με τις οποίες θα εργαστούμε, οι οποίες επίσης καθορίζουν τόσο τη δομή του νευρωνικού δικτύου, όσο και τον τρόπο εκπαίδευσής του. Αρχικά έπρεπε να επιλέξουμε τα επίπεδα που θα βάλουμε στην κορυφή του CNN, μιας και τα μοντέλα που παρέχονται από το Keras είναι φτιαγμένα να κατηγοριοποιούν τις δεδομένα σε 1000 κατηγορίες σύμφωνα με το ImageNet, άρα εμείς έπρεπε να ορίσουμε δικά μας επίπεδα στην κορυφή του CNN παραλείποντας τα αντίστοιχα έτοιμα, ώστε να κατηγοριοποιήσουμε τα δεδομένα στο επιθυμητό πλήθος κλάσεων. Έγιναν πειράματα με τρεις βασικές περιπτώσεις: α) η χρήση δύο πλήρως συνδεδεμένων επιπέδων στην κορυφή του CNN κι ενός επιπλέον ως επίπεδο ταξινόμησης, β) η χρήση ενός πυκνού επιπέδου πριν το τελικό επίπεδο ταξινόμησης, και γ) η χρήση μόνο του επιπέδου ταξινόμησης, όπου είναι και η αντίστοιχη δομή που χρησιμοποιείται στην πλήρη αρχιτεκτονική που παρέχεται από το Keras. Μετά από τα πρώτα πειράματα όμως φάνηκε να μην βελτιώνει τα αποτελέσματα η προσθήκη επιπέδων κι έτσι εγκαταλείφθηκαν γρήγορα οι επιλογές α) και β) κι από εκεί και πέρα εργαστήκαμε αλλάζοντας μόνο το επίπεδο ταξινόμησης σε σύγκριση με τα δίκτυα που παρέχονται. Επιλέξαμε να χρησιμοποιήσουμε τον Adam optimizer, όπου αποτελεί έναν από τους πιο ευρέως χρησιμοποιούμενους optimizers στο πεδίο της αναγνώρισης εικόνας. Η υπερπαραμέτρος η οποία όμως 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. Τέλος το πλήθος των εποχών κυμάνθηκε από τις 10 μέχρι και τις 30 στις περισσότερες περιπτώσεις, με πολύ λίγα παραδείγματα να φτάνουν μέχρι και τις 60 με



Σχήμα 3.12: Overfitting

70 εποχές. Πολλές φορές η εκπαίδευση έγινε σε στάδια. Όταν υπήρχε η εκτίμηση πως περισσότερες εποχές μπορούν να συνεισφέρουν στη βελτίωση του ταξινομητή, επαναφέραμε την κατάσταση του ταξινομητή στην οποία είχε μείνει και τον εκπαιδεύαμε για επιπλέον εποχές. Φυσικά ο προσδιορισμός των βέλτιστων τιμών των υπερπαραμέτρων έγινε μέσω της εκπαίδευσης των CNN και την προσαρμογή των υπερπαραμέτρων στα αποτελέσματα που λαμβάναμε.

3.2.4 Αντιμετώπιση Overfitting

Ένα από τα μεγαλύτερα προβλήματα κατά την εκπαίδευση ενός νευρωνικού δικτύου είναι το overfitting, η υπερβολική προσαρμογή δηλαδή του δικτύου στα δεδομένα εκπαίδευσης η οποία οδηγεί σε χειρότερη γενίκευση σε νέα δεδομένα. Στο σχήμα 3.12 παρουσιάζεται μια αναπαράσταση ενός μοντέλου με overfitting κι ενός βέλτιστου - κανονικοποιημένου μοντέλου.

Σε γενικές γραμμές μεγάλο πλήθος δεδομένων βοηθάει στην αποφυγή του overfitting όμως και αυτό δεν αρκεί από μόνο του. Έχουν αναπτυχθεί λοιπόν αρκετοί τρόποι για την αντιμετώπιση του συγκεκριμένου φαινομένου. Θα αναφερθούμε σε μερικούς από αυτούς, τους οποίους επί των πλείστον χρησιμοποιήσαμε και στην εργασία μας. Η προσθήκη επιπέδων Dropout στο μοντέλο είναι ένας από τους τρόπους που επιλέξαμε για την αντιμετώπιση του overfitting στα μοντέλα όπου προσθέσαμε επιπλέον πυκνά επίπεδα. Πρόκειται για μια πατέντα της Google, και αυτό που κάνει είναι να "απενεργοποιεί" κόμβους του νευρωνικού δικτύου κατά την εκπαίδευση με αποτέλεσμα σε κάθε εποχή να χρησιμοποιούνται διαφορετικοί συνδυασμοί κόμβων έτσι ώστε να μην προσαρμόζεται το σύνολο των βαρών στα δεδομένα εκπαίδευσης. Με τη χρήση των επιπέδων Dropout πετύχαμε σημαντική βελτίωση σε ορισμένα μοντέλα τα οποία είχαν εμφανές πρόβλημα overfitting. Μια ακόμη τεχνική αντιμετώπισης του overfitting αποτελεί η επιβολή "ποινών" σε πολύπλοκα μοντέλα. Όσο πιο πολύπλοκο είναι ένα μοντέλο, τόσο πιο πιθανό είναι να προσαρμοστεί πέραν του επιθυμητού

στα δεδομένα εκπαίδευσης. Για αυτό το λόγο ωθούμε το μοντέλο να επιλέξει την πιο απλή από τις λύσεις που έχει, επιβάλλοντας μια μορφή ποινών σε βάρη με πολύ μεγάλες τιμές μέσω της κανονικοποίησης L1 (Lasso Regression) και L2 (Ridge Regression). Στα μοντέλα μας αν και κάναμε μερικά πειράματα με τη χρήση αυτής της τεχνικής, δεν τη χρησιμοποιήσαμε στο μεγαλύτερο μέρος των δοκιμών μας. Μια άλλη τεχνική αντιμετώπισης του overfitting η οποία υποτιμήθηκε από άλλες εργασίες, είναι το data augmentation. Ο τρόπος που υλοποιείται μέσω των εργαλείων που παρέχει η βιβλιοθήκη Keras δεν αφορά, όπως κανείς περιμένει από το όνομα (augmentation = αύξηση), την αύξηση των δεδομένων, αλλά την τροποποίηση αυτών. Η τεχνική εφαρμόζεται ευρέως στην αναγνώριση αντικειμένων και βασίζεται στη λογική πως ένα αντικείμενο παραμένει το ίδιο αντικείμενο από όποια οπτική γωνία κι αν το δούμε, σε όποιο κομμάτι του κι αν κοιτάξουμε. Με βάση αυτό λοιπόν, κατά τη διάρκεια της εκπαίδευσης του CNN οι εικόνες καθώς τροφοδοτούνται στο δίκτυο παραμορφώνονται με κάποιους τρόπους που θα αναφερθούν στη συνέχεια, εισάγοντας κάθε φορά μια ελαφρώς διαφορετική οπτική της ίδιας εικόνας με αποτέλεσμα να μην επιτρέπεται στον ταξινομητή να μάθει ακριβώς τα χαρακτηριστικά της συγκεκριμένης εικόνας, αφήνοντας όμως το περιεχόμενο κι αυτό που την κατηγοριοποιεί εμφανές. Οι παραμορφώσεις που εφαρμόζονται συνήθως αφορούν τυχαία περιστροφή, καθρέφτισμα ή ζουμ της εικόνας. Σε ορισμένες εργασίες που έχουν γίνει πάνω στην αναγνώριση τέχνης, οι συγγραφείς αναφέρουν πως δεν εφαρμόζουν data augmentation διότι θεωρούν πως ένα έργο τέχνης καθορίζεται και μπορεί να κριθεί μόνο ως μια ολότητα χωρίς παραμορφώσεις, και πως η αλλοίωση κάποιων χαρακτηριστικών θα οδηγήσει σε χειρότερα αποτελέσματα. Αυτά μπορεί να ισχύουν όσον αφορά την εκτίμηση ενός ανθρώπου για την τεχνοτροπία ενός πίνακα, όμως στην περίπτωση του υπολογιστή εμείς είδαμε σημαντική βελτίωση με τη χρήση της συγκεκριμένης τεχνικής. Αυτό μας δείχνει και πως πολλά δεδομένα που ισχύουν για την εκτέλεση μιας εργασίας από τον άνθρωπο δεν ισχύουν απαραίτητα και για την εκτέλεση της ίδιας εργασίας από τον υπολογιστή. Μπορεί η ανθρώπινη εμπειρία να αποτελεί βάση κι έμπνευση για της μεθόδους τεχνητής νοημοσύνης, αλλά δεν αποτελεί και απόλυτο γνώμονα. Τέλος δύο ακόμη τεχνικές που χρησιμοποιήσαμε είναι ο πρόωρος τερματισμός (early stopping) και η μείωση του ρυθμού εκπαίδευσης (learning rate reduction). Η πρώτη σταματάει την εκπαίδευση όταν για κάποιες εποχές δεν έχει υπάρξει βελτίωση σε μια μετρική που ορίζεται από τον προγραμματιστή ώστε να σταματήσει την υπερμοντελοποίηση του δικτύου το οποίο όπως φαίνεται οδηγείται προς τα εκεί, και η δεύτερη μειώνει το ρυθμό εκπαίδευσης (learning rate) σε μια αντίστοιχη περίπτωση ώστε να μην επιτρέψει στο μοντέλο να αλλάξει κατά πολύ τα βάρη του και να προσαρμοστεί υπερβολικά στα δεδομένα εκπαίδευσης.

3.2.5 Έξοδος προγράμματος και μετρικές αξιολόγησης

Μετά από το βασικό σχεδιασμό του συνελκτικού νευρωνικού δικτύου έπρεπε να αποφασίσουμε σε ποιές μετρικές θα βασίσουμε την εκπαίδευσή του αλλά και τι έξοδο θα πάρουμε από το πρόγραμμα. Εφόσον η μετρική η οποία αποτελεί μέτρο σύγκρισης με υπόλοιπες εργασίες είναι το accuracy, επιλέξαμε να εκπαιδεύσουμε τα CNN με στόχο της βελτιστοποίησης της συγκεκριμένης μετρικής. Η εκπαίδευση με βάση το accuracy είναι μια πρακτική που εφαρμόζεται στις περισσότερες περιπτώσεις εκπαίδευσης νευρωνικών δικτύων. Εκτός από το accuracy όμως ελέγχουμε και την τιμή της συνάρτησης απώλειας (loss function) η οποία δείχνει πόσο "μακριά" είναι η πρόβλεψη από την πραγματικότητα. Δηλαδή μπορεί κάποια πρόβλεψη να είναι λανθασμένη αλλά να είναι

πολύ πιο κοντά στη σωστή από κάποια άλλη λανθασμένη πρόβλεψη. Σύμφωνα με τις τιμές της `loss function`, και σε ορισμένες περιπτώσεις και των τιμών της ακρίβειας (`accuracy`) λειτουργούν και οι δύο τελευταίες τεχνικές αντιμετώπισης του `overfitting` που αναφέρθηκαν παραπάνω. Στο τέλος της εκπαίδευσης του νευρωνικού εξάγεται η ακρίβεια τόσο των δεδομένων εκπαίδευσης, όσο και των δεδομένων `validation` και φυσικά `testing`. Η ακρίβεια στα τελευταία δεδομένα είναι και η αντικειμενική αξιολόγηση του μοντέλου μας. Στο τέλος επίσης εξάγονται οι τιμές των μετρικών `precision`, `recall` και `f1` για κάθε κατηγορία αλλά και οι συνολικές. Επιλέξαμε να παίρνουμε από κάθε μοντέλο μια γραφική παράσταση η οποία παρουσιάζει τις τιμές του `accuracy` και της `loss function` πάνω στα δεδομένα εκπαίδευσης και επαλήθευσης (`validation`) στην πορεία της εκπαίδευσης.

3.2.6 Αποθήκευση Μοντέλων

Μια πολύ καλή πρακτική κατά την εκπαίδευση νευρωνικών δικτύων είναι να αποθηκεύουμε τα μοντέλα που εκπαιδεύουμε, ή τα βάρη αυτών. Στην περίπτωση μας όπου δουλέυαμε σε έναν απομακρυσμένο υπολογιστή ήταν πολύ σημαντικό να αποθηκεύονται στιγμιότυπα του μοντέλου ακόμη και κατά την εκπαίδευση διότι ακόμη κι η παραμικρή διακοπή της σύνδεσης στο διαδίκτυο μπορούσε να οδηγήσει σε τερματισμό της σύνδεσης SSH με αποτέλεσμα εάν εκπαιδεύεται ένα νευρωνικό εκείνη την ώρα, να διακοπεί η εκπαίδευση και να χαθούν ακόμη και ώρες εκπαίδευσης. Για αυτό το λόγο χρησιμοποιήσαμε `checkpoints`, το οποίο σημαίνει ότι ανά κάποιες εποχές, που ορίζουμε εμείς, ή όταν υπάρχει βελτίωση σε κάποια μετρική, το οποίο και πάλι είναι στην ευχέρεια του προγραμματιστή, αποθηκεύονται τα βάρη ή ένα συνολικό στιγμιότυπο του μοντέλου, κι έτσι εάν για ον οποιοδήποτε λόγο θέλουμε, μπορούμε να επαναφέρουμε την εκπαίδευση από το συγκεκριμένο στάδιο. Εκτός από `checkpoints` όμως αποθηκεύαμε τα βάρη του CNN στο τέλος της εκπαίδευσης έτσι ώστε να μπορούμε να συνεχίσουμε την εκπαίδευσή του εάν το επιθυμούμε ή και να χρησιμοποιήσουμε τη γνώση που έχει αποκτήσει το δίκτυο για να τροφοδοτήσουμε κάποιο άλλο όπως στο `transfer learning`.

3.2.7 Συνδυαστική Μάθηση - Ensemble Learning

Μετά την εκπαίδευση των μοντέλων μας κάναμε μερικά πειράματα με συνδυαστική μάθηση. Αυτό που κάναμε πρακτικά είναι να πάρουμε μοντέλα εκπαιδευμένα να κάνουν την εργασία που επιθυμούμε και να δημιουργήσουμε έναν υπερ-ταξινομητή η απόφαση του οποίου προκύπτει από το συνδυασμό των εξόδων των εκπαιδευμένων μοντέλων που χρησιμοποιούμε. Δηλαδή κατά την ταξινόμηση μια εικόνα μπαίνει ως είσοδος στους επί μέρους ταξινομητές, αυτοί βγάζουν έναν πίνακα με τις πιθανότητες αυτή η εικόνα να ανήκει στην κάθε κατηγορία και ο υπερ-ταξινομητής συνδυάζει με κάποιο τρόπο αυτές τις εξόδους, συνήθως χρησιμοποιώντας τη μέση, μέγιστη ή ελάχιστη τιμή ανά κατηγορία κι έτσι προκύπτει η τελική έξοδος. Πρόκειται για τη μέθοδο που μας οδήγησε στα κορυφαία αποτελέσματα που πετύχαμε σε αυτή την εργασία.

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 προγραμμάτων σχετικά με τη βαθιά μάθηση και ειδικά για τα συνελκτικά νευρωνικά δίκτυα που τις χρησιμοποιήσαμε εμείς.

Το υπολογιστικό περιβάλλον όπου χρησιμοποιήσαμε για τα πρώτα πειράματα είναι το Google Colab ⁹. 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, 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 την εισαγωγή δεδομένων στο πρόγραμμα.

Τα πρώτα μας πειράματα έγιναν πάνω στο σύνολο δεδομένων Pandora18k διότι τα άλλα σύνολα δεδομένων δεν χωρούσαν στο χώρο που παρέχεται δωρεάν από το Google Drive (παρέχονται δωρεάν 15 GB ενώ τα wikiart και wikiart2 ξεπερνούν τα 25 GB), αλλά και διότι ακόμη και με την επιτάχυνση της GPU που παρέχει το Google Colab, η εκπαίδευση στα μεγάλα σύνολα δεδομένων θα διαρκούσε ημέρες. Μετά την εξοικείωσή μας με τα εργαλεία και τον κώδικα ξεκινήσαμε να τρέχουμε το μεγαλύτερο μέρος των πειραμάτων μας στο σέρβερ pinkfloyd του εργαστηρίου ευφυών συστημάτων του Εθνικού Μετσόβιου Πολυτεχνείου. Ο συγκεκριμένος υπολογιστής είναι εξοπλισμένος με δύο GPUs της NVidia, μία GeForce GTX 1080 Ti και μία TITAN Xp. Τα πειράματα "τρέχανε" σε μία από τις δύο κάρτες γραφικών κάθε φορά μιας και πρόκειται για έναν υπολογιστή ο οποίος μοιράζεται και με άλλα μέλη της ακαδημαϊκής κοινότητας. Η σύνδεση στον υπολογιστή γινόταν εξ αποστάσεως μέσω του πρωτοκόλλου SSH με τη βοήθεια του PuTTY και η μεταφορά αρχείων μέσω του πρωτοκόλλου SFTP.

⁹ "Google Colab", colab.research.google.com

Κεφάλαιο 4

Αξιολόγηση Αποτελεσμάτων και Συγκρίσεις

Στο παρόν κεφάλαιο παρουσιάζονται τα αποτελέσματα των πειραμάτων μαζί με τις παρατηρήσεις και τα συμπεράσματα που προκύπτουν από αυτά. Αρχικά γίνεται μια συγκριτική παρουσίαση των πειραμάτων που εκτελέστηκαν αναλύοντας την επιρροή διαφόρων παραμέτρων, ενώ στη συνέχεια γίνεται μια ποιοτική αξιολόγηση των αποτελεσμάτων η οποία συνοδεύεται από προσωπικές παρατηρήσεις του συγγραφέα. Τέλος συγκρίνονται τα αποτελέσματα της διπλωματικής με άλλες εργασίες που έχουν πραγματοποιηθεί το ίδιο πρόβλημα.

4.1 Συγκριτικά αποτελέσματα μοντέλων βαθιάς μάθησης

Στην παρούσα ενότητα παρουσιάζονται τα αποτελέσματα των πειραμάτων που εκτελέσαμε, ομαδοποιημένα σε συγκεκριμένους θεματικούς κύκλους, στους οποίους αναφέρεται κάθε φορά το χαρακτηριστικό βάση του οποίου διαφοροποιούνται τα μοντέλα που παρουσιάζονται, και ακολουθεί η ανάλυση της επιρροής του συγκεκριμένου χαρακτηριστικού στα αποτελέσματα.

4.1.1 Αρχιτεκτονικές

Αρχικά παρουσιάζουμε τη βέλτιστη επίδοση που επιτεύχθηκε στο test set με κάθε αρχιτεκτονική που χρησιμοποιήσαμε, ανεξάρτητα των τιμών των υπερπαραμέτρων. Στον πίνακα 4.1 βλέπουμε τις τιμές της μετρικής accuracy στα τρία υποσύνολα του συνόλου δεδομένων μας (train, validation και test set) από τα μοντέλα που εκπαιδεύτηκαν στο σύνολο δεδομένων Pandora18k καθώς και το πλήθος των παραμέτρων κάθε αρχιτεκτονικής, επιλέγοντας για κάθε αρχιτεκτονική το μοντέλο που έδωσε το υψηλότερο accuracy στο test set.

Πίνακας 4.1: Αποτελέσματα διαφόρων αρχιτεκτονικών

Αρχιτεκτονική	Accuracy			Πλήθος Παραμέτρων
	Test Set	Validation Set	Train Set	
DenseNet121	64,2%	65,1%	74,0%	7,055,954
VGG16	56,1%	55,8%	81,1%	14,723,922
VGG19	53,0%	53,3%	85,7%	20,033,618
ResNet50	52,5%	54,3%	68,1%	23,624,594
ResNet152	59,1%	59,9%	70,8%	58,370,944
ResNet50 V2	55,4%	57,1%	95,2%	23,601,682
Inception V3	67,6%	68,3%	76,6%	21,839,666
Inception - ResNet V2	66,6%	65,7%	71,2%	54,364,402

Πριν αρχίσουμε την ανάλυση των αποτελεσμάτων πρέπει να αναφέρουμε πως κατά την εκτέλεση πειραμάτων για τη βελτιστοποίηση των υπερπαραμέτρων και την παρατήρηση της συμπεριφοράς του μοντέλου, χρησιμοποιήθηκε ως βασική αρχιτεκτονική η InceptionV3. Για αυτό το λόγο οι υπερπαραμέτροι που χρησιμοποιήσαμε ως βάση στη συνέχεια, πιθανόν να προσαρμόστηκαν αρκετά στη συγκεκριμένη αρχιτεκτονική το οποίο μαζί με το γεγονός ότι δοκιμάστηκαν πολλοί περισσότεροι συνδυασμοί υπερπαραμέτρων συγκριτικά με τις υπόλοιπες, μπορεί να εξηγήει το λόγο για τον οποίο πετύχαμε βέλτιστα αποτελέσματα με τη συγκεκριμένη αρχιτεκτονική. Αυτά βέβαια χωρίς να ξεχνάμε ότι η InceptionV3 αποτελεί μια πολύ ισχυρή και όπως φαίνεται "ταιριαστή" αρχιτεκτονική για το πρόβλημά μας. Η InceptionV3 δεν επιλέχθηκε αυθαίρετα, αλλά μετά από ένα μέρος αρχικών πειραμάτων με μερικές βασικές αρχιτεκτονικές, παρατηρήσαμε πως η συγκεκριμένη μας έδινε πολύ καλύτερα αποτελέσματα από τις υπόλοιπες, για αυτό και επιλέξαμε να βελτιστοποιήσουμε το μοντέλο μας με βάση αυτή. Σε γενικές γραμμές προσπαθήσαμε να βελτιστοποιήσουμε τις αποδόσεις των μοντέλων κρατώντας όμως το μεγαλύτερο κομμάτι των υπερπαραμέτρων και της επεξεργασίας των δεδομένων σταθερό ώστε να μπορεί να μελετηθεί η συμπεριφορά των μοντέλων κάτω από παρόμοιες συνθήκες. Πρέπει να αναφέρουμε πως στο σύνολο δεδομένων wikiart2 το καλύτερο αποτέλεσμα δεν έδωσε ένα μοντέλο βασισμένο στην αρχιτεκτονική InceptionV3 όπως στο Pandora18k, αλλά μοντέλο με βάση την αρχιτεκτονική Inception - ResNetV2 το οποίο ξεπέρασε κατά λίγο το αντίστοιχο μοντέλο του InceptionV3. Βλέπουμε λοιπόν πως μπορεί σε διαφορετικά σύνολα δεδομένων να παίρνουμε βέλτιστα αποτελέσματα από διαφορετικής δομής μοντέλα. Βέβαια ανεξάρτητα του συνόλου δεδομένων που χρησιμοποιήθηκε τα μοντέλα με ίδιες αρχιτεκτονικές έδιναν αντίστοιχα καλή ή κακή ακρίβεια, όμως σε περιπτώσεις όπου οι επιδόσεις ήταν πολύ κοντά όπως έγινε με τις αρχιτεκτονικές InceptionV3 και Inception - ResNetV2, το κάθε σύνολο δεδομένων έδωσε ελαφρώς διαφορετικά αποτελέσματα με αποτέλεσμα στην περίπτωση του wikiart2 να ανέβει στην κορυφή το Inception - ResNetV2.

Το ότι περισσότερες παράμετροι σε ένα δίκτυο οδηγούν σε καλύτερα αποτελέσματα αποτελεί μια κοινή παρανόηση για όσους κάνουν τα πρώτα τους βήματα στο κομμάτι της βαθιάς μάθησης. Ο παραπάνω πίνακας αποτελεί τρανό παράδειγμα πως κάτι τέτοιο δεν ισχύει μιας και την τρίτη καλύτερη επίδοση, την πετύχαμε με την αρχιτεκτονική DenseNet121 η οποία αποτελεί την αρχιτεκτονική με τις λιγότερες παραμέτρους, λιγότερες μάλιστα από τις μισές σε σύγκριση με τη VGG16 όπου είναι η αμέσως "μεγαλύτερη". Με την εξέλιξη των αρχιτεκτονικών, δημιουργούνται ολοένα και αποδοτικότερα μοντέλα άλλοτε με μεγαλύτερο κι άλλοτε με μικρότερο πλήθος παραμέτρων. Η αύξηση των παραμέτρων ενός μοντέλου όμως έχει μεγάλο υπολογιστικό κόστος. Να επισημάνουμε εδώ πως το πλήθος των (εκπαιδευσιμων) παραμέτρων επηρεάζει καθοριστικά το χρόνο εκπαίδευσης ενός μοντέλου το οποίο όταν ασχολούμαστε με δίκτυα εκατομμυρίων παραμέτρων και σύνολα δεδομένων χιλιάδων ή και εκατομμυρίων στοιχείων τότε μια μείωση λίγων ποσοστιαίων μονάδων στο χρόνο, όπως της τάξης του 5%, μπορεί να σημαίνει μείωση του χρόνου εκπαίδευσης του δικτύου κατά πολλές ώρες ή και ημέρες. 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 λαμβάνουν υπόψιν πάντα και τους πόρους που δαπανήθηκαν για να επιτευχθεί ένα αποτέλεσμα.

Πίνακας 4.2: Επίπεδα Data Augmentation

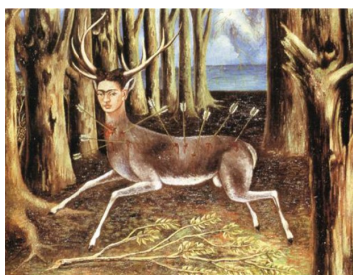
	1° Επίπεδο	2° Επίπεδο	3° Επίπεδο
rotation	x	x	✓
width shift	x	x	✓
height shift	x	x	✓
zoom	x	x	✓
shear	x	✓	✓
horizontal flip	x	✓	✓
vertical flip	x	✓	✓

Παρατηρώντας τη διαφορά της τιμής του accuracy στο training set και στο test set μπορούμε να δούμε πόσο ισχυρό είναι το overfitting σε κάθε μοντέλο, πόσο έχει προσαρμοστεί δηλαδή στα δεδομένα εκπαίδευσης. Φυσικά είναι πολλοί παράμετροι που καθορίζουν το πόσο προσαρμόζεται ένα μοντέλο σε ένα σύνολο των δεδομένων όπως έχει αναφερθεί και σε παραπάνω κεφάλαιο, όμως το πλήθος των παραμέτρων είναι ένα χαρακτηριστικό που συχνά δείχνει τη δυνατότητα ενός μοντέλου να μάθει πιο περίπλοκες συναρτήσεις, με αποτέλεσμα να κινδυνεύει από το overfitting. Όπως παρατηρούμε από τον παραπάνω πίνακα όμως, δεν ισχύει πως περισσότερες παράμετροι οδηγούν οπωσδήποτε σε μεγαλύτερο overfitting. Θα μελετήσουμε εκτενέστερα τη συμπεριφορά πολλών μοντέλων σχετικά με το overfitting στη συνέχεια, όμως τα παραπάνω αποτελούν μια βασική παρατήρηση για την κατανόηση και αντιμετώπιση του συγκεκριμένου προβλήματος.

4.1.2 Data Augmentation

Μία από τις βασικότερες τεχνικές που χρησιμοποιήσαμε για την αντιμετώπιση του overfitting ήταν το data augmentation. Όπως περιγράφηκε και σε παραπάνω κεφάλαιο, η τεχνική του data augmentation λειτουργεί με τον ακόλουθο τρόπο: σε κάθε εποχή, όταν οι εικόνες εισάγονται στο δίκτυο ώστε να εκπαιδευτεί, γίνονται ορισμένες μετατροπές στις εικόνες ώστε να μην είναι ακριβώς ίδιες με αυτές που είδε το μοντέλο στις προηγούμενες εποχές. Οι τροποποιήσεις που γίνονται στις εικόνες περιέχουν κάποια τυχαιότητα, δηλαδή πολλές εφαρμόζονται με κάποια συγκεκριμένη πιθανότητα ή ο βαθμός επιρροής τους καθορίζεται από κάποια τυχαία μεταβλητή, ούτως ώστε να μην επαναλαμβάνονται οι τροποποιημένες εικόνες. Μελετήσαμε την επιρροή του data augmentation τόσο στο συνολικό accuracy ενός μοντέλου, όσο και στην αποφυγή του overfitting, εκπαιδεύοντας δύο βασικά διαφορετικά μοντέλα στα οποία αφήσαμε όλα τα χαρακτηριστικά σταθερά, και δοκιμάσαμε τρία διαφορετικά "επίπεδα" data augmentation στο καθένα. Το πρώτο μοντέλο που δοκιμάστηκε είχε ως βάση την αρχιτεκτονική Inception V3 ενώ το δεύτερο, τη DenseNet121. Τα τρία "επίπεδα" data augmentation που δοκιμάστηκαν όπως προαναφέρθηκε, παρουσιάζονται στον πίνακα 4.2.

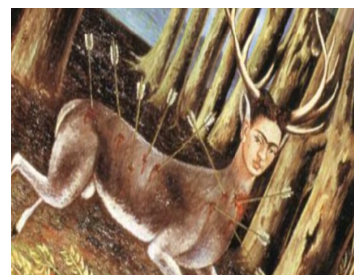
Παρατηρούμε λοιπόν πως στο πρώτο επίπεδο δεν επιτελείται καμία λειτουργία του data augmentation, δεν έχουμε λοιπόν καμία παραμόρφωση της εικόνας. Στο δεύτερο επίπεδο έχουμε ορισμένους τυχαίους αντικατοπτρισμούς της εικόνας και shear mapping αυτής. Στο τρίτο επίπεδο εφαρμόζουμε επτά διαφορετικούς τρόπους παραμόρφωσης της εικόνας όπως αντικατοπτρισμούς, περιστροφές, zoom κ.α. Στις εικόνες 4.1, 4.2 και 4.3 παρουσιάζονται τα τρία επίπεδα data augmentation όταν εφαρμοστούν σε ένα έργο τέχνης, όπως το "The wounded deer" της Frieda Kahlo. Στους πίνακες 4.3 και 4.4 παρουσιάζονται οι τρεις τιμές του accuracy, μία για κάθε ένα από τα τρία μέρη του



Σχήμα 4.1: 1^ο Επίπεδο Data Augmentation



Σχήμα 4.2: 2^ο Επίπεδο Data Augmentation



Σχήμα 4.3: 3^ο Επίπεδο Data Augmentation

Πίνακας 4.3: InceptionV3 - Data Augmentation

Επίπεδο	Accuracy		
	Test Set	Validation Set	Train Set
1 ^ο	58,3%	59,7%	99,4%
2 ^ο	61,3%	59,5%	82,9%
3 ^ο - A	52,5%	53,6%	52,5%
3 ^ο - B	65,0%	66,5%	70,4%

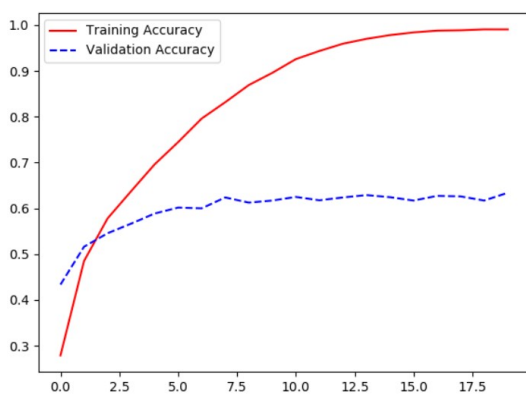
Πίνακας 4.4: DenseNet121 - Data Augmentation

Επίπεδο	Accuracy		
	Test Set	Validation Set	Train Set
1 ^ο	61,1%	62,7%	99,7%
2 ^ο	63,1%	65,0%	91,1%
3 ^ο - A	56,7%	56,4%	70,2%
3 ^ο - B	64,2%	65,1%	74,0%

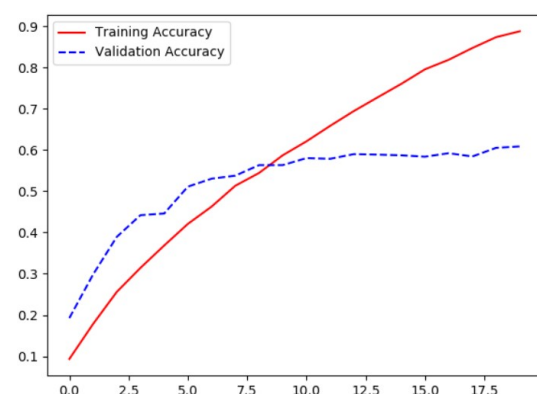
συνόλου δεδομένων (train, validation και test) στα δύο μοντέλα που δημιουργήσαμε, μαζί με το αντίστοιχο επίπεδο data augmentation. Στον πίνακα προσθέτουμε δύο καταχωρήσεις για την περίπτωση του επιπέδου 3. Η πρώτη καταχώρηση (A) αφορά τις τιμές των μετρικών για εκπαίδευση του δικτύου για ίδιο πλήθος εποχών με τις άλλες δύο περιπτώσεις (Επίπεδο 1 και Επίπεδο 2), ενώ η δεύτερη καταχώρηση (B) αφορά τις καλύτερες τιμές των μετρικών που επιτεύχθηκαν με μεγάλο data augmentation, σε περισσότερες όμως εποχές. Συγκεκριμένα για την αρχιτεκτονική Inception V3 για το επίπεδο 3, επιτύχαμε βέλτιστες μετρικές μετά από 66 εποχές, ενώ για τη DenseNet121 μετά από 55.

Τα συμπεράσματα που προκύπτουν είναι λίγο - πολύ τα αναμενόμενα. Αρχικά αξίζει να σημειωθεί πως και στις δύο περιπτώσεις παρουσιάζεται παρόμοια συμπεριφορά. Παρατηρούμε πως στην περίπτωση που δεν χρησιμοποιηθεί καθόλου data augmentation, οι ταξινομητές προσαρμόζονται υπερβολικά πολύ στο σύνολο δεδομένων πάνω στο οποίο εκπαιδεύονται, για αυτό το λόγο και παρατηρούμε τα τόσο υψηλά ποσοστά ακρίβειας στο train set, χωρίς όμως αυτό να συνοδεύεται με αντίστοιχα υψηλή ακρίβεια και στα άλλα δύο σετ. Όπως περιμέναμε λοιπόν σε αυτές τις περιπτώσεις το φαινόμενο του overfitting είναι πολύ έντονο, όπως φαίνεται και στις γραφικές παραστάσεις των σχημάτων 4.4 και 4.5 όπου και στις δύο περιπτώσεις είναι πολύ εμφανής ο διαχωρισμός της γραμμής που αντιστοιχεί στο training accuracy, και αυτής που αντιστοιχεί στο validation accuracy. Από κάποια εποχή και μετά λοιπόν η γνώση που παίρνει ο ταξινομητής αφορά συγκεκριμένα χα-

ρακτηριστικά των εικόνων του training set, κι όχι γενικά χαρακτηριστικά της κάθε κλάσης που θα οδηγούσαν σε καλύτερη γενίκευση. Ο ταξινομητής αδυνατεί να ξεχωρίσει ποία χαρακτηριστικά καθορίζουν μια κλάση και ποία άλλα υπάρχουν μόνο στις συγκεκριμένες εικόνες, για αυτό και παρατηρούμε μια αλλαγή στην κλήση του validation accuracy το οποίο αρχικά έχει ανοδική πορεία ενώ μετά από κάποιο σημείο παραμένει περίπου σταθερό, ενώ το training accuracy μέχρι και το τέλος συνεχίζει την ανοδική του πορεία. Αξίζει να μείνουμε λίγο παραπάνω στην εικόνα με τις γραφικές παραστάσεις και να αναφερθούμε και στις διαφορές μεταξύ των δύο μοντέλων, αφού μέχρι στιγμής έχουμε αναφερθεί μόνο στα κοινά τους χαρακτηριστικά. Η βασικότερη και μόνη ουσιαστική διαφορά των γραφικών παραστάσεων των δύο μοντέλων είναι το από ποίο σημείο και μετά ξεκινά η υπερμοντελοποίηση και σταματά η βελτίωση του validation accuracy. Στην περίπτωση του DenseNet121, από τις πρώτες 2-3 εποχές ήδη παρατηρούμε πως από τη μία το validation accuracy έχει φτάσει σε ικανοποιητικά υψηλά επίπεδα, κι από την άλλη έχει ξεκινήσει να ανοίγει η διαφορά μεταξύ του train και του validation accuracy. Στην εκπαίδευση του Inception V3 όμως αυτή η συμπεριφορά παρατηρείται μετά από 8-9 εποχές. Εδώ λοιπόν πρέπει να αναφερθεί πως το μοντέλο που χρησιμοποιήθηκε στην περίπτωση του DenseNet121 αποτελείται από τη βασική αρχιτεκτονική, με ένα πλήρως συνδεδεμένο επίπεδο ταξινόμησης με 18 κόμβους στην κορυφή. Στο μοντέλο που χρησιμοποιήθηκε ως βάση το Inception V3, πέραν του επιπέδου ταξινόμησης μεγέθους 18 κόμβων, προστέθηκε κι ακόμη ένα πυκνό επίπεδο 650 κόμβων ανάμεσα στη βασική αρχιτεκτονική και το τελικό πυκνό επίπεδο. Αυτό έχει ως αποτέλεσμα το συνολικό μέγεθος του δικτύου να είναι πάνω από 15 φορές μεγαλύτερο από το αντίστοιχο του μοντέλου του DenseNet. Παρατηρούμε λοιπόν πως το μικρότερο μοντέλο προσαρμόστηκε πολύ πιο γρήγορα στα δεδομένα τόσο στο σημείο να έχει μια καλή γενίκευση και στο test set, αλλά και στην υπερμοντελοποίηση στα δεδομένα του train set, ενώ το τεράστιο δίκτυο που δημιουργήθηκε στην άλλη περίπτωση χρειάστηκε αρκετό παραπάνω χρόνο για να προσαρμόσει τα βάρη του στα δεδομένα που του δόθηκαν. Για αυτό το λόγο έφτασε αργότερα κοντά στα μέγιστα του ποσοστά validation accuracy αλλά και άργησε να εμφανιστεί το φαινόμενο του overfitting.



Σχήμα 4.4: DenseNet121 Train and Validation Accuracy



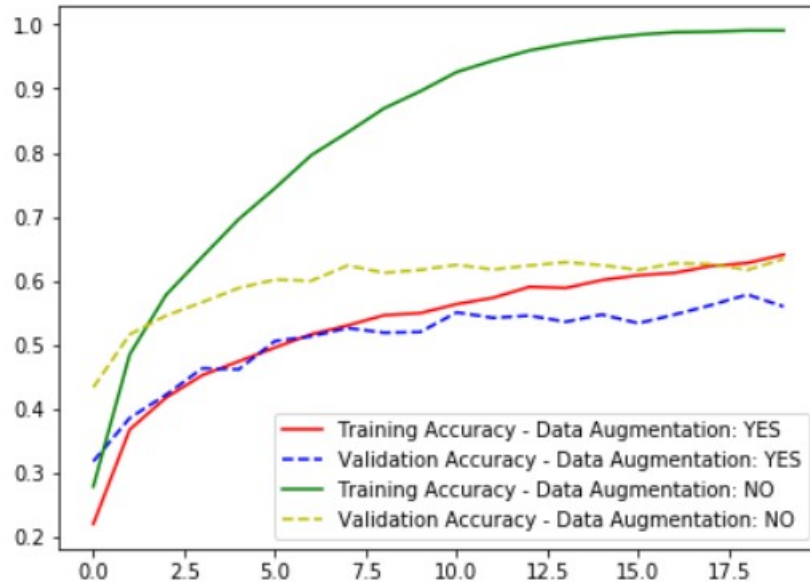
Σχήμα 4.5: InceptionV3 Train and Validation Accuracy

Μια ακόμη σημαντική παρατήρηση από τους παραπάνω πίνακες είναι το γεγονός πως και στις δύο περιπτώσεις όσο περισσότερο data augmentation χρησιμοποιούμε, τόσο μικρότερο overfitting έχουμε, κι όσο μικρότερο overfitting έχουμε, τόσο καλύτερα αποτελέσματα παίρνουμε και στη μετρική accuracy για το test set. Η πρώτη συνεπαγωγή είναι προφανής κι αναμενόμενη, και είναι κι ο

κύριος λόγος που χρησιμοποιούμε το data augmentation στα μοντέλα μας. Η δεύτερη συνεπαγωγή όμως αν και αναμενόμενη δεν είναι τετριμμένη. Αυτό αποτελεί ένα σημαντικό λάθος στο οποίο μπορεί να υποπέσουν πολλοί οδηγούμενοι από τα καλύτερα αποτελέσματα που δίνει μέχρι ενός σημείου η μείωση του overfitting. Η χρήση του data augmentation, τουλάχιστον στη μορφή που παρέχεται από τα εργαλεία του keras, έχει σκοπό να τροποποιεί σε κάθε εποχή τις εικόνες όπως προαναφέρθηκε, με σκοπό να μην εμμένει σε συγκεκριμένα χαρακτηριστικά ο ταξινομητής αλλά να επικεντρώνεται στα γενικότερα χαρακτηριστικά τα οποία είναι παρόντα και μετά τις τροποποιήσεις. Στο τελευταίο κομμάτι των παραπάνω βρίσκεται ο κίνδυνος του λάθους. Η υπερβολική τροποποίηση των δεδομένων ενέχει τον κίνδυνο να αλλοιωθούν και χαρακτηριστικά σημαντικά για την ταξινόμηση τους στη σωστή κλάση. Αποτέλεσμα αυτού είναι ο ταξινομητής να μην έχει τη δυνατότητα να αναγνωρίσει τα στοιχεία της εισόδου που καθορίζουν την κλάση κι ως αποτέλεσμα να αδυνατεί να ταξινομήσει ορθά ένα μεγάλο μέρος των εισόδων, διότι απλώς δεν έχει τα απαραίτητα δεδομένα. Αυτό οδηγεί σε ταξινομητές με πολύ μικρό overfitting και πάλι, αλλά με χειρότερες τιμές στις μετρικές αξιολόγησης, όπως το test set accuracy.

Όπως αναφέρθηκε νωρίτερα ένα μοντέλο στο οποίο τροφοδοτούμε δεδομένα τα οποία τροποποιούμε με την τεχνική του data augmentation, χρειάζεται περισσότερες εποχές για να συγκλίνει στις βέλτιστες τιμές του. Παρατηρείται λοιπόν μια μικρότερη κλίση στην καμπύλη που ορίζεται από τις τιμές τους accuracy σε συνάρτηση με τις εποχές. Στη γραφική παράσταση του σχήματος 4.6 παρουσιάζονται οι καμπύλες που ορίζουν οι τιμές του accuracy για το train (συνεχείς γραμμές) και validation (διακεκομμένες γραμμές) set σε συνάρτηση με τις εποχές, για το μοντέλο με την αρχιτεκτονική DenseNet121 έχοντας χρησιμοποιήσει data augmentation επιπέδου 1 (Πράσινο - Κίτρινο) και επιπέδου 3 (Κόκκινο - Μπλε), συγκεκριμένα παρουσιάζεται το Επίπεδο 3 - A, διότι κρατήσαμε μόνο τις πρώτες 20 εποχές της εκπαίδευσης με εφαρμογή πολλών παραμορφώσεων. Πρόκειται λοιπόν για μια σύγκριση μοντέλων με και χωρίς data augmentation. Παρατηρούμε κι εδώ πως το μοντέλο χωρίς τη χρήση του data augmentation φτάνει πολύ υψηλές τιμές training accuracy, λόγω του overfitting. Από τη συγκεκριμένη γραφική θα κρατήσουμε όμως τη διαφορά στις κλίσεις των συναρτήσεων, μιας και όπως παρατηρούμε οι κλίσεις των συναρτήσεων του μοντέλου χωρίς data augmentation είναι εμφανώς μεγαλύτερες σε σύγκριση με τις αντίστοιχες συναρτήσεις του μοντέλου με data augmentation. Ειδικά στο κομμάτι των δεδομένων εκπαίδευσης η διαφορά είναι πολύ μεγάλη, όμως και στο κομμάτι των δεδομένων validation παρατηρούμε πως η μπλε διακεκομμένη γραμμή συνεχίζει κι αυξάνεται μέχρι και το τέλος των 20 εποχών, ενώ η κίτρινη από την πέμπτη-έκτη περίπου εποχή και μετά παραμένει σχετικά σταθερή έχοντας ήδη φτάσει κοντά στις μέγιστες τιμές της. Επίσης παρατηρούμε πως στην περίπτωση χρήσης του data augmentation οι γραμμές του train και validation accuracy πορεύονται περίπου μαζί, για τις πρώτες 6-7 εποχές μάλιστα μπλέκονται μεταξύ τους, ενώ στην άλλη περίπτωση από τις πρώτες κιόλας εποχές όπως αναφέρθηκε και παραπάνω, οι δυο γραμμές χωρίζονται, όπου αποτελεί ένα χαρακτηριστικό του overfitting. Ο λόγος της διαφοράς αυτής είναι κι ο λόγος όπου λειτουργεί η τεχνική του data augmentation. Αυτός είναι πως με τη συνεχή παραμόρφωση των εικόνων ο ταξινομητής δεν μπορεί να αναγνωρίσει εύκολα τα χαρακτηριστικά που καθορίζουν την κάθε κλάση, κι έτσι εποχή με εποχή προσαρμόζει λίγο τα βάρη του ώστε να φτάσει σε μια γενικευμένη εικόνα της κάθε κλάσης, ενώ αντίθετα όταν τα δεδομένα παραμένουν αναλλοίωτα, μετά από λίγες μόνο εποχές το μοντέλο έχει μάθει την αναπαράσταση κάθε κλάσης όπως θεωρεί αυτό πως είναι, και συνεχίζει να προσαρμόζεται στα

επακριβή δεδομένα που του παρέχουμε κατά την εκπαίδευση. Αυτή η δυσκολία που βάζουμε στον ταξινομητή είναι αυτή που τον κάνει να χρειάζεται περισσότερο χρόνο ώστε να συγκλίνει σε καλύτερη ακρίβεια ταξινόμησης, είναι όμως κι αυτή που του επιτρέπει να προσαρμόσει καταλληλότερα τα βάρη του ώστε να δώσει καλύτερα συνολικά αποτελέσματα.



Σχήμα 4.6: Σύγκριση μοντέλων με και χωρίς data augmentation

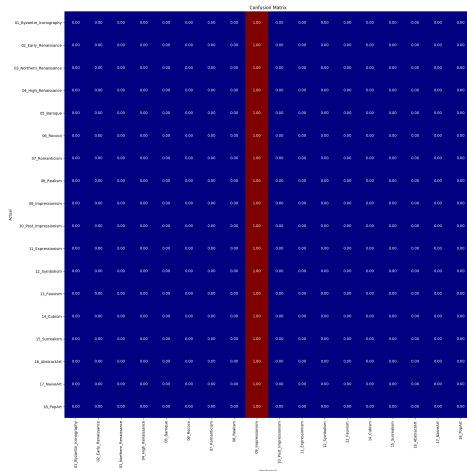
Όπως αναφέρθηκε και παραπάνω, μια άποψη εναντίον της χρήσης του data augmentation, παραλλαγή της οποίας συναντήσαμε και σε εργασίες φοιτητών που πραγματεύονται το πρόβλημά μας, είναι πως ένα έργο τέχνης αποτελεί μια ολότητα (το οποίο αποτελεί κι επιχείρημα κατά της περικοπής των εικόνων), και μόνο ως ένα ακέραιο κι ανεπηρέαστο κομμάτι μπορεί να αναγνωριστεί. Οποιοσδήποτε αλλαγές στην όψη του έργου αλλοιώνουν το ίδιο το έργο τέχνης κι έτσι πλέον χαρακτηρίζουμε κάτι διαφορετικό κι όχι το έργο τέχνης από το οποίο ξεκινήσαμε. Σε αυτό το επιχείρημα υπάρχει η προσέγγιση από την πλευρά της θεωρίας της τέχνης, με την οποία όμως όπως έχουμε αναφέρει ήδη δεν θα ασχοληθούμε στη συγκεκριμένη διπλωματική εργασία, υπάρχει όμως και η προσέγγιση από την πλευρά της τεχνητής νοημοσύνης. Οι αλλαγές που επιτελούνται στην είσοδο χρησιμοποιούν ως βάση μια συγκεκριμένη εικόνα και η τροποποίηση αυτής αφορά τον τρόπο με τον οποία μαθαίνει ένα νευρωνικό δίκτυο. Δεν αναγνωρίζουμε μια τροποποιημένη έκδοση της εικόνας μιας και το data augmentation εφαρμόζεται μόνο πάνω στο train set, κι όχι στα validation και test. Στην τελική αξιολόγηση λοιπόν αναγνωρίζουμε ένα έργο τέχνης χωρίς παραμορφώσεις. Ακόμη κι ένας άνθρωπος όταν επιτελεί πολλές εργασίες τροποποιεί τα δεδομένα του, απλώς συχνά το κάνει με έναν τρόπο που δεν είναι εμφανής στους παρατηρητές γι' αυτό κι αυτή η πρακτική "ξενίζει" πολλούς. Η ανθρώπινη εμπειρία μπορεί να χρησιμοποιηθεί ως έμπνευση για τον τομέα της τεχνητής νοημοσύνης, αποτελεί μάλιστα έμπνευση αυτής, δεν θα πρέπει όμως να συγχέουμε και να ταυτίζουμε της ανθρώπινες τακτικές με τις τακτικές που χρησιμοποιεί ένας υπολογιστής για να επιτελέσει ένα έργο. Το γεγονός πως μια παραμόρφωση της εικόνας μπορεί να δυσχεραίνει το έργο ενός ανθρώπου να αναγνωρίσει την τεχνοτροπία της, δεν σημαίνει απαραίτητα πως το ίδιο συμβαίνει και στην περίπτωση του υπολογιστή. Όπως παρατηρούμε μάλιστα ισχύει ακριβώς το αντίθετο. Η χρήση του data augmentation μπορεί να συμβάλει σημαντικά στη βελτίωση ενός μοντέλου, ειδικά σε περιπτώσεις όπου δεν είναι διαθέσιμο πολύ μεγάλο πλήθος δεδομένων άρα

είναι εύκολο για το μοντέλο να χρησιμοποιήσει λάθος χαρακτηριστικά για την ταξινόμηση των δεδομένων.

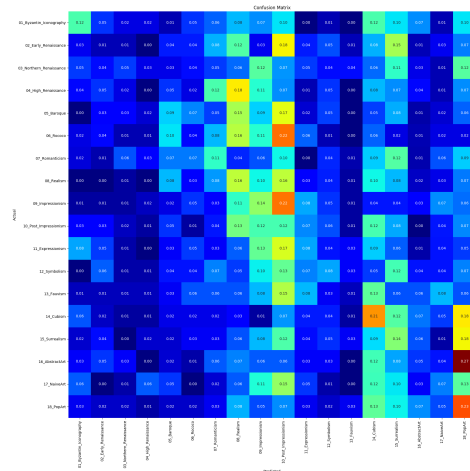
4.1.3 Learning Rate

Μια από τις βασικότερες υπερπαραμέτρους για την επιτυχία ενός μοντέλου, είναι ο ρυθμός μάθησης (learning rate). Στη συγκεκριμένη υποενότητα θα ασχοληθούμε με την επιρροή αυτής της υπερπαραμέτρου στην εκπαίδευση και την απόδοση ενός μοντέλου. Ο ρυθμός μάθησης καθορίζει το πόσο μεγάλες αλλαγές γίνονται στα βάρη του δικτύου κατά την εκπαίδευση. Αποτελεί πρακτικά έναν συντελεστή που ορίζει το μέγεθος του βήματος που θα κάνει το δίκτυο προς την ελαχιστοποίηση της loss function. Εκτελέσαμε πειράματα με διαφορετικές τιμές του learning rate για να παρακολουθήσουμε τη συμπεριφορά των μοντέλων κατά την εκπαίδευση. Κατά τα γνωστά, τα υπόλοιπα χαρακτηριστικά του μοντέλου έμειναν σχεδόν σταθερά ώστε να συγκρίνουμε την επιρροή καθαρά και μόνο αυτής της υπερπαραμέτρου. Αν και για την εύρεση της βέλτιστης τιμής γίνονται συνήθως μικρές αλλαγές στην τιμή μιας υπερπαραμέτρου γύρω από μια καλή συγκριτικά τιμή, εμείς επιλέξαμε να μην παρουσιάσουμε αυτό το κομμάτι αλλά να κάνουμε τρία πειράματα με μεγάλη διαφορά στην τιμή του learning rate μεταξύ τους, μιας και θεωρήσαμε πως αυτά θα δείξουν τον τρόπο με τον οποίο επηρεάζει ο ρυθμός μάθησης την εκπαίδευση ενός μοντέλου. Χρησιμοποιήσαμε ως βάση μια τιμή του learning rate που είδαμε πως δίνει καλά αποτελέσματα και στη συνέχεια εκπαιδεύσαμε δύο επιπλέον μοντέλα, ένα με learning rate 100 φορές μεγαλύτερο, κι ένα με 100 φορές μικρότερο. Καταλήξαμε λοιπόν με τρία μοντέλα, ένα με πολύ μικρό learning rate, ένα με ένα μέσο και καλό για το συγκεκριμένο πρόβλημα learning rate, κι ένα με πολύ υψηλό learning rate. Σε αυτό το σημείο να επισημάνουμε πως οι βέλτιστες τιμές του ρυθμού μάθησης καθορίζονται από το περιβάλλον του προβλήματος, όπως τη μορφή των δεδομένων και των δικτύων που χρησιμοποιούνται στο μοντέλο. Μπορεί λοιπόν μια βέλτιστη τιμή του learning rate για ένα πρόβλημα να είναι υπερβολικά μεγάλη ή μικρή για ένα άλλο πρόβλημα. Δεν παίζει ρόλο λοιπόν η αναφορά απόλυτων τιμών του ρυθμού μάθησης αλλά μας ενδιαφέρει η συσχέτιση μεταξύ των διαφορετικών τιμών. Η συγκεκριμένη σειρά πειραμάτων μας έδωσε ορισμένα ενδιαφέροντα αποτελέσματα. Και στις δύο ακραίες περιπτώσεις του πολύ υψηλού κι αντίστοιχα πολύ χαμηλού learning rate η ακρίβεια του μοντέλου ήταν κάκιστη, και συγκεκριμένα κάτω του 10% στο test set, κι αντίστοιχες τιμές περίπου στα train και validation σετ. Το ενδιαφέρον όμως έγκειται στους λόγους που οδήγησαν σε αυτές τις πολύ κακές επιδόσεις. Οι confusion Matrices των δύο μοντέλων που παρουσιάζονται στα σχήματα 4.7 και 4.8 βοηθούν πολύ στην κατανόηση της συμπεριφοράς των μοντέλων. Εδώ να αναφερθεί επιγραμματικά τι είναι ένας Confusion Matrix: πρόκειται για έναν πίνακα όπου στη γραμμή είναι η πραγματική κλάση μιας εικόνας και στη στήλη η κλάση όπου προέβλεψε ο ταξινομητής. Κάθε κουτάκι του πίνακα λοιπόν με συντεταγμένες (X,Y) δείχνει πόσες εικόνες της κλάσης X ταξινομήθηκαν από το μοντέλο ως κλάση Y. Το μπλε χρώμα δείχνει χαμηλές τιμές, ενώ όσο πιο κόκκινο το χρώμα τόσο πιο μεγάλες τιμές, δηλαδή τόσο περισσότερες εικόνες αντιστοιχούν σε αυτό το κουτάκι. Ενδιάμεσες τιμές είναι η χρωματική παλέτα από σκούρο μπλε, κίτρινο, πορτοκαλί μέχρι το σκούρο κόκκινο. Σωστή ταξινόμηση έχουν τα στοιχεία της κύρια διαγωνίου (X,X).

Δεν μας ενδιαφέρουν οι συγκεκριμένες τιμές του κάθε στοιχείου του πίνακα αλλά η γενική μορφή του. Από το σχήμα 4.8 παρατηρούμε πως ο ταξινομητής βρίσκεται σε μια σύγχυση. Δεν μπορεί να ξεχωρίσει τις κλάσεις και ταξινομεί γύρω από μερικές κλάσεις σχεδόν τυχαία, βάση



Σχήμα 4.7: High Learning Rate Confusion Matrix



Σχήμα 4.8: Low Learning Rate Confusion Matrix

μάλλον της μορφής που είχαν τα βάρη του δικτύου του όταν ξεκίνησε η εκπαίδευση. Η πολύ μικρή τιμή του ρυθμού μάθησης λοιπόν δεν επιτρέπει στο μοντέλο να αλλάξει αρκετά τις τιμές των βαρών του ώστε να ταξινομήσει σωστά τις εικόνες. Για αυτό το λόγο ακόμη και το training accuracy αδυνατεί να πάρει μεγάλες τιμές. Καθ όλη τη διάρκεια της εκπαίδευσής του οι μετρικές του δικτύου απλώς κυμαίνονται γύρω από τις χαμηλές αυτές τιμές αδυνατώντας να μάθει κάτι περισσότερο από αυτά που του δόθηκαν πρακτικά έτοιμα. Από την άλλη όπως παρατηρούμε από το σχήμα 4.7 ο ταξινομητής ο οποίος εκπαιδεύτηκε με πολύ μεγάλο learning rate, στο τέλος προβλέπει μόνο μία κλάση. Ο πιθανότερος λόγος για αυτό είναι πως ο ρυθμός μάθησης ήταν τόσο μεγάλος ώστε από τις πρώτες εικόνες μετέβαλε τόσο πολύ τα βάρη ώστε τα έριξε σε κάποιο τοπικό ελάχιστο όπου ήταν η πρόβλεψη μιας και μόνο κατηγορίας και στη συνέχεια τα βάρη προσαρμόστηκαν τόσο γρήγορα σε αυτό ώστε στη συνέχεια αδυνατούσαν να φύγουν από αυτό. Παρατηρούμε λοιπόν πόσο σημαντική είναι η κατάλληλη τιμή του learning rate για ένα δίκτυο μιας και μια πολύ χαμηλή τιμή μπορεί να στερήσει τη δυνατότητα από το δίκτυο να μάθει ορισμένα (ή όπως στην περίπτωσή μας οποιαδήποτε) χαρακτηριστικά, αλλά μια μεγάλη τιμή του ρυθμού μάθησης μπορεί είτε να κάνει τα βάρη να μεταβληθούν υπερβολικά πολύ και να φτάσουν σε μια κατάσταση που δεν μπορεί πια να αλλάξει όπως στο παράδειγμά μας, είτε να τα κάνει να πηδάνε από εδώ κι από εκεί χωρίς να κλίνουν προς κάποιο τοπικό ή ολικό ελάχιστο, μιας και η παραμικρή προσπάθεια μεταβολής των βαρών προς μια κατεύθυνση μετατρέπεται σε πολύ μεγάλες αλλαγές στις τιμές αυτών με αποτέλεσμα να μην μπορεί να κατευθυνθεί το δίκτυο προς τα επιθυμητά σημεία.

4.1.4 Transfer Learning

Σε παραπάνω κεφάλαιο έχει αναλυθεί τι είναι το transfer learning. Πρακτικά, όπως υποδεικνύει και το όνομα, αφορά τη μεταφορά γνώσης που έχει αποκτήσει ένα μοντέλο από ένα πρόβλημα, σε ένα δεύτερο διαφορετικό πρόβλημα. Στην περίπτωσή μας εφαρμόσαμε transfer learning χρησιμοποιώντας ως αφετηρία τα βάρη των δικτύων τα οποία είχαν εκπαιδευτεί για αναγνώριση εικόνας, στο σύνολο δεδομένων ImageNet. Μελετήσαμε τη σημασία και τις δυνατότητες που παρέχει η συγκεκριμένη πρακτική. Εκτελέσαμε πειράματα τόσο με και χωρίς τη χρήση της μεταφοράς γνώσης, όσο και με διαφορετικές παραλλαγές αυτής που θα αναλύσουμε παρακάτω. Πιο συγκεκρι-

Πίνακας 4.5: Με και Χωρίς Transfer Learning

Transfer Learning	Accuracy		
	Test Set	Validation Set	Train Set
NAI	63,1%	65,0%	91,1%
OXI	42,2%	43,4%	62,0%

κρίμενα εκπαιδεύσαμε παρόμοια μοντέλα με δύο διαφορετικές αρχικοποιήσεις στα βάρη τους. Η μία ήταν η τυχαία αρχικοποίηση (χωρίς transfer learning) και η άλλη ήταν με τις τιμές από την εκπαίδευση στο ImageNet όπως παρέχονται από τη βιβλιοθήκη keras (με transfer learning). Στην αρχή εκπαιδεύσαμε ολόκληρα τα δύο δίκτυα. Στη συνέχεια στο μοντέλο με τη μεταφορά γνώσης, δοκιμάσαμε να παγώσουμε ορισμένα επίπεδα του δικτύου, πρακτική η οποία έχει χρησιμοποιηθεί κατά κόρον σε εργασίες που πραγματεύονται το πρόβλημά μας. Η λογική πίσω από αυτό είναι πως έχουμε έτοιμα ορισμένα βάρη από ένα παρεμφερές πρόβλημα και δεν θέλουμε να τα "χαλάσουμε" με την εκπαίδευσή τους μιας και υπάρχει ο κίνδυνος του overfitting. Έτσι μπορούμε να "παγώσουμε" ορισμένα επίπεδα, ειδικά τα κατώτερα τα οποία θεωρητικά αναγνωρίζουν βασικά στοιχεία μιας εικόνας όπως είναι οι ακμές και τα χρώματα. Με αυτό τον τρόπο μειώνουμε και σε σημαντικό βαθμό το πλήθος των παραμέτρων που εκπαιδεύονται με αποτέλεσμα να παρατηρείται σημαντική μείωση και στο χρόνο εκπαίδευσης των δικτύων. Αρχικά όμως θα ασχοληθούμε με τη σύγκριση της χρήσης transfer learning ή μη και στη συνέχεια θα ασχοληθούμε και με τις διάφορες παραλλαγές του. Στον πίνακα 4.5 παρουσιάζονται οι τιμές της μετρικής accuracy στα τρία υποσύνολα του συνόλου δεδομένων μας (train, validation και test) για τις δύο περιπτώσεις, με και χωρίς μεταφορά γνώσης.

Τα συγκεκριμένα αποτελέσματα προέκυψαν από την εκπαίδευση των μοντέλων για 20 εποχές. Εκπαιδεύσαμε το μοντέλο χωρίς transfer learning και για περισσότερες εποχές όμως δεν έδωσε καλύτερα αποτελέσματα. Μια αρχική παρατήρηση είναι πως η μεταφορά γνώσης έδωσε καλύτερα αποτελέσματα, όμως στον ίδιο αριθμό εποχών αυτό είναι ίσως αναμενόμενο μιας και έχει ένα προβάδισμα. Το πιο ενδιαφέρον συμπέρασμα όμως προκύπτει από αυτά που δεν φαίνονται στον πίνακα, πως δηλαδή ακόμη και με περισσότερες εποχές εκπαίδευσης και με αλλαγή των υπερπαραμέτρων το μοντέλο χωρίς μεταφορά γνώσης αδυνατεί ακόμη και να πλησιάσει τα αποτελέσματα του άλλου. Σε αυτό το σημείο βρίσκεται η σπουδαιότητα του transfer learning. Δεν αποτελεί απλώς ένα καλύτερο σημείο εκκίνησης. Δίνει στο μοντέλο τη δυνατότητα να φτάσει σε κομμάτι γνώσης στο οποίο δεν θα είχε τη δυνατότητα από μόνο του. Η αδυναμία του μοντέλου να φτάσει από μόνο του σε αντίστοιχα επίπεδα οφείλεται από τη μία στο overfitting, μιας και δεν υπάρχουν αρκετά δεδομένα στα οποία μπορεί να εκπαιδευτεί ένα στιβαρό μοντέλο, αλλά και στο μέγεθος του χώρου αναζήτησης βέλτιστων λύσεων, το οποίο έχει ως αποτέλεσμα το μοντέλο να μην μπορεί να υπερπηδήσει κάποια εμπόδια για να φτάσει σε βέλτιστα αποτελέσματα. Εκεί έρχεται η μεταφορά γνώσης να μας λύσει τα χέρια δίνοντάς μας τη δυνατότητα να χρησιμοποιήσουμε γνώση που έχει προέλθει από δεδομένα εκατομμυρίων εικόνων και ώρες ή και ημέρες εκπαίδευσης. Εάν φανταστούμε το χώρο αναζήτησης του μοντέλου μας σαν ένα χάρτη με χιλιάδες νησιά όπου αναζητούμε το ελάχιστο υψόμετρο, αυτό που μας κάνει η μεταφορά γνώσης είναι να μας τοποθετήσει σε ένα σύμπλεγμα νησιών το οποίο θα μας δώσει ικανοποιητικά αποτελέσματα και πιθανότατα, αλλά όχι εγγυημένα, εκεί γύρω βρίσκεται και η βέλτιστη λύση. Το πρόβλημα είναι πως ο χώρος αναζήτησης είναι τόσο

Πίνακας 4.6: Transfer Learning

Μοντέλο	Accuracy			Χρόνος ανά εποχή
	Test Set	Validation Set	Train Set	
Περίπτωση 1	28,3%	28,6%	29,5%	45 sec
Περίπτωση 2	38,2%	40,7%	42,1%	58 sec
Περίπτωση 3	42,4%	44,0%	47,6%	75 sec
Περίπτωση 4	63,1%	65,0%	91,1%	153 sec

μεγάλος που χωρίς αυτή τη βοήθεια δεν θα μπορούσαμε να φτάσουμε ποτέ σε αυτό το σύμπλεγμα των νησιών με αποτέλεσμα να βρίσκουμε άλλα τοπικά ελάχιστα και να μένουμε σε αυτά χωρίς τη δυνατότητα να φτάσουμε ποτέ στα βέλτιστα αποτελέσματα. Για αυτό ακριβώς το λόγο και σχεδόν όλες οι εργασίες που έχουν γίνει τα τελευταία χρόνια πάνω στην αναγνώριση τεχνοτροπίας έχουν χρησιμοποιήσει transfer learning, και μάλιστα μέσω αυτού έχουν επιτύχει βέλτιστα αποτελέσματα. Συμπεραίνουμε λοιπόν πως η συγκεκριμένη τεχνική εκπαίδευσης είναι απαραίτητη για να πάρουμε ικανοποιητικά αποτελέσματα με τα δεδομένα που έχουμε. Για αυτό το λόγο κι όλα τα άλλα πειράματα που έγιναν έκαναν χρήση των προ-εκπαιδευμένων δικτύων που παρέχονται από το keras. Στη συνέχεια προχωρήσαμε ένα βήμα παραπέρα κι όπως προαναφέρθηκε, επιχειρήσαμε να "παγώσουμε" ορισμένα επίπεδα του δικτύου ώστε να αποφύγουμε το overfitting και να αφήσουμε τα ανώτερα επίπεδα να προσαρμοστούν καλύτερα στα χαρακτηριστικά υψηλότερου επιπέδου όπως είναι κι εν τέλη η τεχνοτροπία της εικόνας. Πρόκειται για μια πρακτική που χρησιμοποιείται συνήθως όταν το σύνολο δεδομένων δεν είναι αρκετά μεγάλο, και χρησιμοποιήθηκε από πολλές εργασίες που ασχολήθηκαν με το πρόβλημα της αναγνώρισης τεχνοτροπίας. Δοκιμάσαμε αρχικά να παγώσουμε όλα τα επίπεδα και να επιτρέψουμε την εκπαίδευση μόνο του τελευταίου πυκνού επιπέδου ταξινόμησης (περίπτωση 1), μια τακτική που ακολουθείται συνήθως όταν το πρόβλημα από το οποίο αντλούμε τη γνώση είναι πολύ παρεμφερές με το πρόβλημα που επιλύουμε. Αναμενόμενα τα αποτελέσματα δεν ήταν πολύ ικανοποιητικά. Επιτεύχθηκε ακρίβεια (accuracy) 28,3% στο test set μετά από 40 εποχές, χωρίς όμως να παρατηρείται σχεδόν καθόλου overfitting. Ο κυριότερος λόγος της χαμηλής ακρίβειας είναι ότι η αναγνώριση εικόνας δεν αποτελεί πολύ παρεμφερές πρόβλημα με την αναγνώριση τεχνοτροπίας άρα τα επίπεδα τα οποία παγώσαμε δεν συμβάλουν ικανοποιητικά στην ταξινόμηση. Στη συνέχεια επιχειρήσαμε να παγώσουμε ένα μικρότερο τμήμα του δικτύου, επιτρέποντας να εκπαιδευτούν τα επίπεδα που ανήκουν στο 20% των υψηλότερων επιπέδων του δικτύου (περίπτωση 2). Τα αποτελέσματα ήταν καλύτερα, και πάλι όμως δεν ήταν συγκρίσιμα με την πλήρη εκπαίδευση του δικτύου. Καταφέραμε να φτάσουμε 38,2% ακρίβεια στο test set. Επαναλάβαμε το πείραμα επιτρέποντας να εκπαιδευτεί το 40% του δικτύου (περίπτωση 3) και τέλος φυσικά έχουμε την εκπαίδευση του πλήρους δικτύου κι από το προηγούμενο μέρος αυτής της υπο-ενότητας (περίπτωση 4). Στον πίνακα 4.6 παρουσιάζονται συνοπτικά τα αποτελέσματα των τεσσάρων περιπτώσεων που αναφέρθηκαν παραπάνω μαζί με τους χρόνους εκπαίδευσης του κάθε μοντέλου ανά εποχή.

Παρατηρούμε πως όσο βαθύτερη είναι η εκπαίδευση του δικτύου τόσο καλύτερα αποτελέσματα παίρνουμε, τόσο μεγαλύτερο είναι το overfitting και φυσικά τόσο μεγαλύτερη η διάρκεια εκπαίδευσης. Το τελευταίο κομμάτι είναι προφανές μιας και παγώνοντας ορισμένα επίπεδα, μειώνεται ο αριθμός των παραμέτρων προς εκπαίδευση, άρα μειώνεται κι ο χρόνος εκπαίδευσης ολόκληρου του δικτύου. Όσον αφορά το κομμάτι του overfitting, αυξάνοντας το βάθος εκπαίδευσης,

Πίνακας 4.7: Datasets Results

Σύνολο Δεδομένων	Accuracy			Χρόνος ανά εποχή
	Test Set	Validation Set	Train Set	
Pandora18k	67,6%	68,3%	76,6%	169 sec
Wikiart	63,5%	64,8%	77,1%	1912 sec
Wikiart2	64,8%	64,3%	79,0%	1843 sec

αυξάνεται η "χωρητικότητα" του δικτύου, με αποτέλεσμα να μπορεί να μάθει πιο περίπλοκες συναρτήσεις και να προσαρμοστεί καλύτερα στα δεδομένα. Αυτό όμως έχει κι ως αποτέλεσμα δυστυχώς την υπερβολική προσαρμογή των παραμέτρων στα δεδομένα εισόδου δίνοντας το αποτέλεσμα του overfitting που παρατηρούμε. Στο βασικότερο κομμάτι όπου είναι αυτό που μας ενδιαφέρει πάντοτε, η καλή επίδοση του μοντέλου μας, η βαθύτερη εκπαίδευση φαίνεται να είναι πιο ευεργετική για το μοντέλο. Αυτό πιθανότατα οφείλεται στο γεγονός πως από τη μία τα δεδομένα μας δεν είναι τόσο λίγα ώστε μια πλήρης εκπαίδευση να δημιουργεί σημαντικό πρόβλημα, αν και όπως έχει ειπωθεί πολλές φορές το πρόβλημα του overfitting είναι πάντοτε παρόν, και από την άλλη το πρόβλημα της αναγνώρισης τεχνοτροπίας είναι αρκετά κοντά στην αναγνώριση εικόνας ώστε να μπορεί να επωφεληθεί από τη γνώση που του παρέχεται από το άλλο πρόβλημα, όμως δεν είναι τόσο παρεμφερές ώστε να χρησιμοποιηθεί κάποιο τμήμα του δικτύου αυτούσιο.

4.1.5 Datasets

Σε αυτή τη διπλωματική χρησιμοποιήσαμε δύο βασικά σύνολα δεδομένων, και μια παραλλαγή του δεύτερου, όπως αναλύσαμε και σε παραπάνω κεφάλαιο. Αξίζει λοιπόν να γίνει μια αναφορά και σύγκριση των αποτελεσμάτων που πήραμε και των παρατηρήσεων που κάναμε κατά την εκπαίδευση αυτών. Η περιγραφή του κάθε συνόλου δεδομένων έχει γίνει ήδη για αυτό στο συγκεκριμένο σημείο θα αναφερθούμε επιγραμματικά στα χαρακτηριστικά τα οποία θεωρούμε πως επηρέασαν τη διαδικασία της εκπαίδευσης και τα αποτελέσματα. Η βασική διαφορά των δύο αρχικών συνόλων δεδομένων είναι το μέγεθος. Το Pandora18k αποτελείται από περίπου 18000 εικόνες χωρισμένες σε 18 κατηγορίες, ενώ το wikiart έχει πάνω από 80000 εικόνες χωρισμένες σε 27 και 22 αντίστοιχα κατηγορίες για το βασικό dataset κι αυτό που δημιουργήσαμε εμείς με βάση το αρχικό. Το Pandora18k είναι ένα ισορροπημένο σύνολο δεδομένων, ενώ τα wikiart έχουν πολύ μεγάλες διαφορές στο πλήθος των στοιχείων από κλάση σε κλάση. Ένα ακόμη στοιχείο το οποίο δεν μπορεί να αμεληθεί είναι πως το Pandora18k έχει ελεγχθεί ως προς την εγκυρότητά του, ενώ στο wikiart αποτελεί μια μαζική συλλογή έργων τέχνης στην οποία όμως βρήκαμε μέχρι και φωτογραφίες από γλυπτά. Στον πίνακα 4.7 παρουσιάζονται τα βέλτιστα αποτελέσματα που επιτεύχθηκαν σε κάθε σύνολο δεδομένων μαζί με ένα μέσο χρόνο εκπαίδευσης ανά εποχή.

Να σημειωθεί βέβαια πως λόγω της πολύ χρονοβόρας εκπαίδευσης των wikiart αλλά και της εγκυρότητας που μας παρείχε το Pandora18k εκτελέσαμε την πλειοψηφία των πειραμάτων στο Pandora18k άρα είχαμε τη δυνατότητα να προσαρμόσουμε καλύτερα τις υπερπαραμέτρους των μοντέλων στις απαιτήσεις του συγκεκριμένου συνόλου δεδομένων. Αυτός είναι ο ένας λόγος που πιθανόν να οφείλεται για τα ελαφρώς καλύτερα αποτελέσματα που έχουμε στο συγκεκριμένο σύνολο. Σε γενικές γραμμές παρατηρούμε πως τα ποσοστά ακρίβειας και στα τρία σύνολα κυμαίνονται περίπου στα ίδια. Τόσο το βέλτιστο accuracy στο test set, όσο και η σύγκριση με το train

Πίνακας 4.8: Resizing - Cropping

Προ-επεξεργασία	Accuracy		
	Test Set	Validation Set	Train Set
Resizing	65,4%	64,7%	86,6%
Cropping	62,2%	64,5%	87,1%

accuracy δείχνουν πως σε όλα τα μοντέλα καταφέραμε να επιτύχουμε αξιόλογα αποτελέσματα αλλά φυσικά με το overfitting να είναι εμφανές. Να σημειώσουμε πως αυτά δεν αποτελούν τα βέλτιστα αποτελέσματα που επιτύχαμε, αλλά αποτελέσματα πειραμάτων που εκτελέστηκαν για τη σύγκριση των συνόλων δεδομένων. Αυτό που έχει ενδιαφέρον είναι πως το τεράστιο μέγεθος (μεγαλύτερο του τετραπλάσιου σε σύγκριση με το Pandora18k) των wikiart δεν οδήγησε σε σημαντικά μικρότερο overfitting. Αυτό οφείλεται από τη μία κυρίως στο γεγονός πως αντιμετωπίσαμε αρκετά καλά το overfitting στο κομμάτι του μοντέλου, άρα το μεγάλο μέγεθος δεδομένων δεν ήταν απαραίτητο, κι από την άλλη στο γεγονός πως και το Pandora18k είχε ένα επαρκώς μεγάλο αριθμό δεδομένων ώστε να μπορέσουμε να εκπαιδεύσουμε τα CNN πάνω σε αυτά. Το γεγονός πως κάποιο σύνολο δεδομένων είναι εκτός ισορροπίας δεν θα έπρεπε να συμβάλει στη μείωση του συνολικού accuracy, τουλάχιστον σε σημαντικό βαθμό, διότι αν το μοντέλο έχει μια τάση προς τις κατηγορίες με περισσότερα δεδομένα τότε στην περίπτωση που και το test set έχει την ίδια κατανομή, περιμένουμε αυτό να μην επηρεάσει αρνητικά το τελικό accuracy. Σημαντική είναι η διαφορά στο χρόνο εκπαίδευσης των δικτύων ανάλογα με τη χρήση του κάθε συνόλου δεδομένων. Συγκεκριμένα ένα δίκτυο που εκπαιδεύεται πάνω σε ένα από τα σύνολα δεδομένων wikiart, χρειάζεται πάνω από το δεκαπλάσιο χρόνο σε σύγκριση με το ίδιο δίκτυο εάν εκπαιδευτεί στο Pandora18k.

4.1.6 Περικοπή και Προσαρμογή Μεγέθους

Εκτός από τα διαφορετικά σύνολα δεδομένων που χρησιμοποιήσαμε, στα πειράματά μας δοκιμάσαμε και διαφορετική προσέγγιση στο μορφή που εισάγονται οι εικόνες στο δίκτυο. Συγκεκριμένα όπως αναφέρθηκε και στο κομμάτι της προ-επεξεργασίας δεδομένων, εκτελέσαμε πειράματα στα οποία εισάγαμε τις εικόνες τους συνόλου δεδομένων με δύο βασικούς διαφορετικούς τρόπους:

- προσαρμογή του μεγέθους τους (resizing),
- και περικοπή της εικόνας (cropping).

Στη δεύτερη περίπτωση δοκιμάσαμε τυχαία περικοπή της εικόνας, αλλά και περικοπή του κεντρικού τμήματος. Τα αποτελέσματα μεταξύ των δύο αυτών υποπεριπτώσεων δεν έδωσαν ιδιαίτερα διαφορετικά αποτελέσματα. Το ενδιαφέρον έγκειται στη διαφορά μεταξύ της χρήσης ολόκληρης εικόνας και τμήματος αυτής, και κυρίως στην ανάμιξη αυτών των δύο. Το τελευταίο θα γίνει πιο κατανοητό στη συνέχεια. Στον πίνακα 4.8 παρουσιάζονται τα αποτελέσματα της εκπαίδευσης δύο δικτύων με διαφορετική προ-επεξεργασία των δεδομένων εισόδου τους, ενός με resizing (μοντέλο A) και ενός με cropping (μοντέλο B) αντίστοιχα.

Παρατηρούμε πως τα μοντέλα έδωσαν πολύ παρόμοια ποσοστά σε όλους τους τομείς με την περίπτωση της χρήσης ολόκληρων εικόνων να γενικεύει κάπως καλύτερα στα δεδομένα του test set. Συγκρίνοντας τα αναλυτικότερα αποτελέσματα παρατηρούμε πως κάθε ταξινόμητης φαίνεται

Πίνακας 4.9: Transfer from/to crop

Μοντέλο	Test Accuracy	
	1	2
A	65,4%	66,4%
B	61,7%	62,9%

Πίνακας 4.10: Αξιολόγηση σε δεδομένα διαφορετικής προ-επεξεργασίας

Μοντέλο	Test Accuracy		
	resizing	random cropping	center cropping
A	63,3%	43,1%	43,3%
B	61,9%	60,0%	61,7%

να προβλέπει καλύτερα ορισμένες κατηγορίες στις οποίες ο άλλος ταξινομητής είναι πιο αδύναμος κι αντίστροφα. Αυτό φυσικά μπορεί να έχει να κάνει με την πορεία της εκπαίδευσης, θα μπορούσε όμως να ερμηνευθεί και από το γεγονός πως ορισμένες τεχνοτροπίες μπορεί να έχουν πολύ ισχυρή σύνδεση με το περιεχόμενο άρα η χρήση ολόκληρων εικόνων να βοηθάει ενώ αντίστοιχα σε άλλες να παίζουν σημαντικότερο ρόλο η λεπτομέρειες οι οποίες φαίνονται σε κάθε κομμάτι του πίνακα αλλά χάνονται όταν αλλάζουμε το μέγεθος της εικόνας. Για να αποφανθούμε όμως για αυτό το κομμάτι θα πρέπει να συμβουλευτούμε ιστορικούς τέχνης ώστε να κατανοήσουμε το θεωρητικό υπόβαθρο πίσω από κάθε τεχνοτροπία και να εκτελέσουμε περισσότερα πειράματα, το οποίο αφήνεται ως μελλοντική δουλειά.

Στη συνέχεια επιχειρήσαμε να πάρουμε τη γνώση που έχει αποκτήσει το δίκτυο από την εκπαίδευση σε ολόκληρες εικόνες και να τη μεταφέρουμε στο δίκτυο το οποίο τροφοδοτείται με τμήματα εικόνας, και αντίστροφα. Στον πίνακα 4.9 βλέπουμε τις τιμές της μετρικής accuracy για το test set πριν και μετά τη χρήσης του transfer learning από το μοντέλο A (resizing) στο B(croppig) κι αντίστροφα. Η στήλη 1 δείχνει το ποσοστό accuracy για το μοντέλο A χωρίς τη μεταφορά γνώση από το B, ενώ η στήλη 2 με τη μεταφορά γνώσης από το B. Προφανώς το αντίστροφο ισχύει για τη γραμμή του μοντέλου B.

Παρατηρούμε λοιπόν πως και στις δύο περιπτώσεις έχουμε βελτίωση των αποτελεσμάτων του μοντέλου το οποίο οφείλεται στο γεγονός πως χρησιμοποιήσαμε ως βάση ένα μοντέλο το οποίο έχει μάθει ικανοποιητικά καλά το πρόβλημα που καλούμαστε να λύσουμε, απλώς με μια μικρή τροποποίηση στα δεδομένα εισόδου. Η μεταφορά γνώσης που χρησιμοποιούμε σε αυτό το σημείο δεν παραβιάζει την εγκυρότητα του μοντέλου μιας και τα σύνολα δεδομένων που χρησιμοποιήσαμε και στις δύο φορές αποτελούνται από τις ίδιες εικόνες στα train, validation και test set, δεν υπάρχει δηλαδή η περίπτωση κάποιο μοντέλο να εκπαιδεύτηκε στο τμήμα μιας εικόνας την οποία θα δει ολόκληρη στην αξιολόγησή του στο test set ή αντίστροφα.

Τέλος, επιχειρήσαμε να αξιολογήσουμε το κάθε μοντέλο (A και B) στα δεδομένα του άλλου μοντέλου. Δηλαδή αξιολογήσαμε το μοντέλο που εκπαιδεύτηκε σε ολόκληρες εικόνες (μοντέλο A), με τμήματα εικόνων (δεδομένα του μοντέλου B) κι αντίστροφα. Εδώ προέκυψαν μερικά ενδιαφέροντα αποτελέσματα τα οποία παρουσιάζονται στον πίνακα 4.10. Οι στήλες του πίνακα δείχνουν το test accuracy του μοντέλου για τις εξής περιπτώσεις προ-επεξεργασίας των εικόνων: προσαρμογή μεγέθους, τυχαία περικοπή τμήματος και περικοπή του κεντρικού τμήματος.

Να σημειώσουμε πως σε αυτό το κομμάτι των πειραμάτων χρησιμοποιήθηκαν ελαφρώς διαφορετικά μοντέλα από τα παραπάνω, για αυτό και παρατηρείται μια απόκλιση στις επιδόσεις του κάθε μοντέλου στα "δικά του" δεδομένα. Αυτό όμως που μας ενδιαφέρει σε αυτό το σημείο είναι η σύγκριση των αποτελεσμάτων μεταξύ τους. Αρχικά παρατηρούμε πως και για τα δύο μοντέλα δεν παίζει ιδιαίτερο ρόλο εάν τα τμήματα της εικόνας έχουν περικοπή από το κεντρικό κομμάτι της ή τυχαία, με τα κεντρικά τμήματα όμως να δίνουν ελαφρώς καλύτερα αποτελέσματα. Αυτό μπορεί να οφείλεται πως συνήθως στο κέντρο ενός έργου τέχνης υπάρχει περισσότερη πληροφορία από τις άκρες οι οποίες μπορεί να χρησιμοποιηθούν από μια τυχαία περικοπή. Ιδιαίτερο ενδιαφέρον παρουσιάζει όμως η επίδοση του κάθε μοντέλου στην προ-επεξεργασία δεδομένων πάνω στα οποία δεν εκπαιδεύτηκε. Παρατηρούμε πως το μοντέλο που εκπαιδεύτηκε σε τμήματα εικόνας μπορεί και γενικεύει εντυπωσιακά καλά στις ολόκληρες εικόνες επιτυγχάνοντας μάλιστα καλύτερη τιμή στο test accuracy από την περίπτωση των τμημάτων της εικόνας. Αντιλαμβανόμαστε λοιπόν πως ο ταξινομητής πιθανότατα μαθαίνει όντως την τεχνοτροπία και μόνο αυτή, και μπορεί κατ'επέκταση να αντεπεξέλθει και σε νέα δεδομένα τα οποία δεν έχουν ακριβώς τη μορφή που έχει μάθει. Αντίθετα παρατηρούμε πως το αντίστροφο δεν ισχύει. Η πρόβλεψη του μοντέλου A που εκπαιδεύτηκε σε ολόκληρες εικόνες, πάνω στο σύνολο δεδομένων από τμήματα των εικόνων, έχουν κατά περίπου 20% χαμηλότερη ακρίβεια σε σύγκριση με τη "δική του" μορφή δεδομένων. Αυτή η παρατήρηση θα μπορούσε να μας οδηγήσει σε δύο χωριστά συμπεράσματα. Το ένα θα ήταν πως η τεχνοτροπία τελικά δεν είναι παρούσα σε όλα τα μέρη του πίνακα άρα περικόπτοντας ένα μόνο μέρος μπορούν να χαθούν πληροφορίες που αφορούν την τεχνοτροπία του έργου. Αυτό όμως αναιρείται από το γεγονός πως ο ταξινομητής που εκπαιδεύτηκε σε τμήματα εικόνων κατάφερε κι έφτασε πολύ ικανοποιητικά επίπεδα ακρίβειας ακόμη και όταν αξιολογήθηκε σε ολόκληρες εικόνες, άρα τα χαρακτηριστικά που βοηθούν στην ταξινόμηση των εικόνων είναι παρόντα. Αυτό μας αφήνει με το δεύτερο συμπέρασμα το οποίο είχαμε αναφέρει και ως αρνητικό της περικοπής των έργων τέχνης, πως πιθανότατα οι ταξινομητές αντλούν πληροφορία κι από το περιεχόμενο των πινάκων το οποίο όμως σε πολλές περιπτώσεις χάνεται όταν γίνει η περικοπή. Ο ταξινομητής λοιπόν αντλεί χαρακτηριστικά κι από το περιεχόμενο των έργων τέχνης, τα οποία όμως δεν είναι παρόντα στα τμήματα που τον βάζουμε να ταξινομήσει για αυτό το λόγο παρατηρούμε τη μεγάλη αυτή διαφορά στις επιδόσεις του μοντέλου A.

4.1.7 Ensemble Learning

Όπως αναφέρθηκε στο κομμάτι του σχεδιασμού και υλοποίησης, μετά την ανάλυση των χαρακτηριστικών των μοντέλων δοκιμάσαμε τη χρήση μερικών μεθόδων συνδυαστικής μάθησης. Η τεχνική που ακολουθήσαμε περιγράφεται στην ενότητα 3.2.7. Πρακτικά δημιουργήσαμε έναν υπερταξινομητή ο οποίος βασιζόμενος στις εξόδους των επί μέρους ταξινομητών δημιουργεί μια δική του συνδυαστική έξοδο. Ο λόγος για τον οποίο αυτή η τεχνική μπορεί να δώσει καλύτερα αποτελέσματα είναι πως ο ένας ταξινομητής μπορεί να καλύπτει τις αδυναμίες του άλλου κι αντίστροφα. Υπάρχουν πολλές περιπτώσεις στις οποίες οι συνδυαστικές μέθοδοι δίνουν καλύτερα αποτελέσματα κι αυτός είναι ο λόγος για τον οποίο οι περισσότερες ομάδες οι οποίες ασχολούνται με την αναγνώριση εικόνας εφαρμόζουν στο τέλος κάποιο συνδυασμό των ταξινομητών τους, με τον οποίο επιτυγχάνουν συνήθως και τα βέλτιστα αποτελέσματα. Εμείς μελετήσαμε τρεις βασικούς

Πίνακας 4.11: Συνδυαστικές Μέθοδοι

Μοντέλα	Test Accuracy		
	Average	Maximum	Minimum
A-B	61,3%	61,3%	61,3%
A-Γ	70,2%	70,3%	70,0%
B-Γ	68,3%	67,4%	68,6%
A-B-Γ	69,2%	70,2%	69,8%
A	67,6%		
B	61,7%		
Γ	66,6%		

τρόπους συνδυασμού των εξόδων των επί μέρους ταξινομητών για τη σύνθεση της τελικής εξόδου, οι οποίοι είναι οι εξής:

- Average
- Maximum
- Minimum

Σε κάθε περίπτωση εκτελείται η πράξη που ορίσαμε ανά κλάση στην έξοδο των ταξινομητών και βάση αυτή προκύπτει η ανά κλάση έξοδος του υπερ-ταξινομητή. Στη συνέχεια κατά τα γνωστά επιλέγεται ως κλάση της εικόνας αυτή με τη μεγαλύτερη τιμή στην αντίστοιχη έξοδο. Εκτελέσαμε τα περισσότερα πειράματα και σε αυτή την περίπτωση με τα μοντέλα που δέχονται ως είσοδο το Pandora18k. Χρησιμοποιήσαμε τρία βασικά μοντέλα στα οποία εφαρμόσαμε τη συνδυαστική μέθοδο που περιγράψαμε τόσο ανά δύο καθώς και με τα τρία μαζί. Τα μοντέλα που χρησιμοποιήσαμε ήταν ένα μοντέλο με βάση την αρχιτεκτονική InceptionV3 το οποίο εκπαιδεύτηκε σε ολόκληρες εικόνες (εφεξής μοντέλο A), ένα μοντέλο με βάση την ίδια αρχιτεκτονική το οποίο όμως εκπαιδεύτηκε σε τμήματα εικόνων (εφεξής μοντέλο B), και τέλος ένα μοντέλο με βάση την αρχιτεκτονική Inception - ResNetV2 το οποίο εκπαιδεύτηκε όπως και το πρώτο σε ολόκληρες εικόνες (εφεξής μοντέλο Γ). Η αξιολόγηση έγινε πάνω σε ολόκληρες εικόνες επειδή όπως δείξαμε και παραπάνω οι ταξινομητές που εκπαιδεύονται σε τμήματα εικόνας αποδίδουν εξίσου καλά και στις ολόκληρες ενώ το αντίστροφο δεν ισχύει. Στον πίνακα 4.11 παρουσιάζονται τα αποτελέσματα του test accuracy που επιτύχαμε με τη συγκεκριμένη μέθοδο με τους διάφορους συνδυασμούς των παραπάνω μοντέλων καθώς και με τους τρεις διαφορετικούς τρόπους συνδυασμού των εξόδων τους που αναφέρθηκαν παραπάνω (average, maximum και minimum). Θεωρήσαμε χρήσιμο να παρουσιαστούν μαζί και τα ποσοστά του κάθε μοντέλου που χρησιμοποιήσαμε ξεχωριστά ώστε να είναι ευκολότερη η σύγκριση των αποτελεσμάτων στη συνέχεια.

Αρχικά παρατηρούμε πως η επιλογή της πράξης με την οποία θα συνδυαστούν οι έξοδοι των επί μέρους ταξινομητών δεν επηρεάζει ιδιαίτερα το αποτέλεσμα μιας και σε όλες τις περιπτώσεις η διαφορά του test accuracy μεταξύ των επιλογών average, maximum και minimum ήταν το πολύ 1,2%. Στη συνέχεια αξίζει να σημειωθεί πως αν και δείχνει η συγκεκριμένη μέθοδος να δίνει πολύ καλύτερα αποτελέσματα, αυτό δεν ισχύει πάντοτε. Συγκεκριμένα στο συνδυασμό των μοντέλων A και B βλέπουμε πως το τελικό αποτέλεσμα είναι χειρότερο από τα αποτελέσματα και των δύο ταξινομητών. Στις υπόλοιπες περιπτώσεις όμως η μέθοδος της συνδυαστικής μάθησης οδήγησε σε

σημαντική αύξηση της ακρίβειας και μας επέτρεψε να ξεπεράσουμε το φράγμα του 70% το οποίο δεν είχαμε καταφέρει να πετύχουμε με τη χρήση ενός και μόνο δικτύου. Να αναφέρουμε επίσης πως μέσω του ensembled learning επιτύχαμε και τα βέλτιστα αποτελέσματα στο σύνολο δεδομένων wikiart2 με το test accuracy να αγγίζει το 67,9% ξεπερνώντας όλα τα ποσοστά που έχουν δημοσιευθεί μέχρι στιγμής σχετικά με την αναγνώριση τεχνοτροπίας στο wikiart. Εάν επικεντρωθούμε περισσότερο στη συγκεκριμένη μέθοδο και αναλύσουμε τα αποτελέσματα των ταξινομητών που έχουμε στη διάθεσή μας, πιθανόν να μπορούσαμε να προσδιορίσουμε τον καλύτερο γραμμικό ή και μη γραμμικό συνδυασμό για τις εξόδους των δικτύων που θα μπορούσε να οδηγήσει σε ακόμη καλύτερα αποτελέσματα. Θα μπορούσαμε επίσης να επιχειρήσουμε να εκπαιδύσουμε μοντέλα τα οποία αναγνωρίζουν με πολύ μεγάλη σιγουριά ορισμένες κατηγορίες και στο τέλος να τα συνδυάσουμε μέσω της πράξης maximum επιλέγοντας κάθε φορά τη μέγιστη τιμή μιας κλάσης εφόσον θα γνωρίζουμε πως κάθε μοντέλο προβλέπει με μεγάλα ποσοστά μόνο τις εικόνες για τις οποίες γνωρίζει σίγουρα την κλάση. Υπάρχει πολλή δουλειά που θα μπορούσε να γίνει λοιπόν στο συγκεκριμένο κομμάτι χωρίς όμως να υπάρχει κάποια εγγύηση πως θα δώσει καλύτερα αποτελέσματα, όσα έχουμε δει μέχρι στιγμής όμως είναι ενθαρρυντικά. Τα παραπάνω αφήνονται ως ιδέες για μελλοντικές κατευθύνσεις μελέτης.

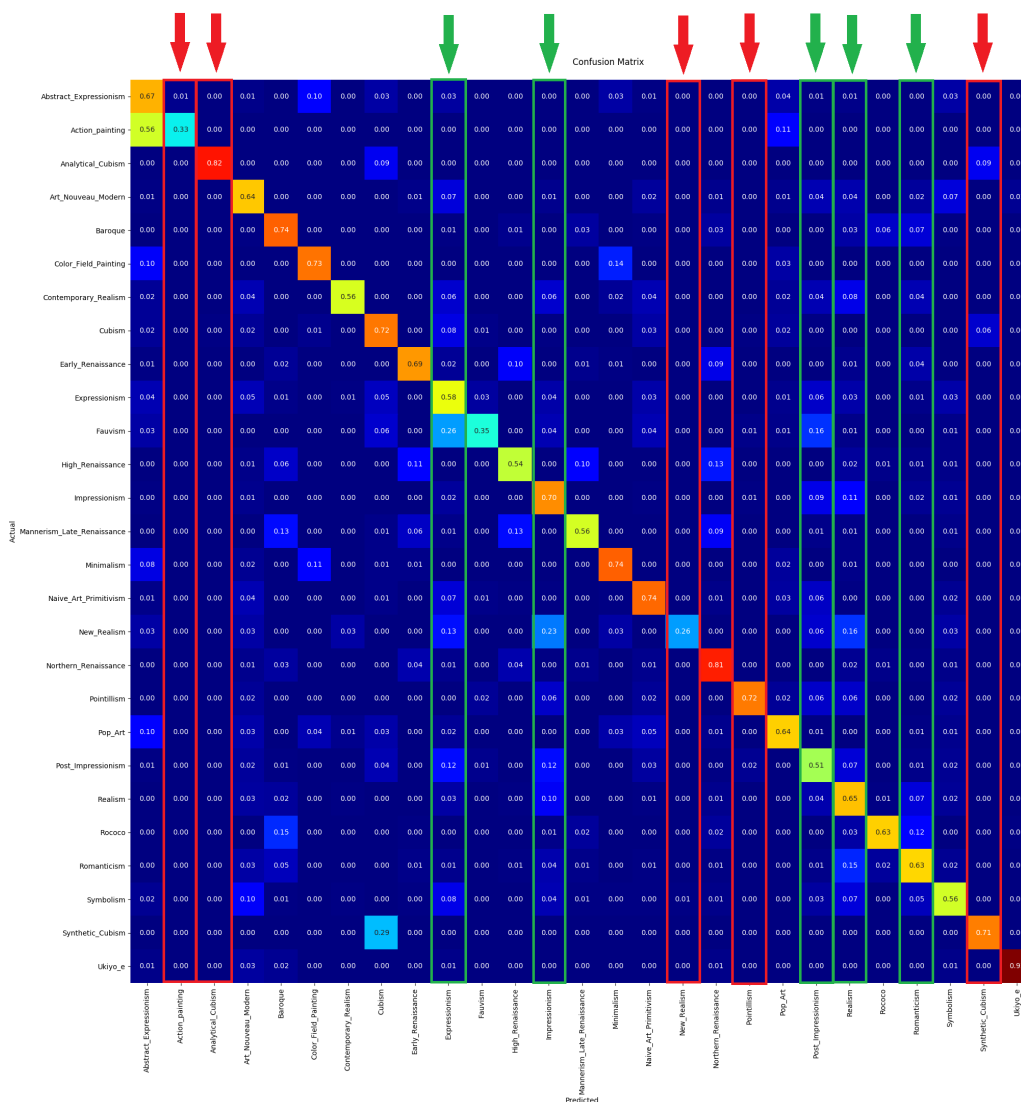
4.2 Προσωπικές Παρατηρήσεις - Ποιοτική Αξιολόγηση

Στη συγκεκριμένη ενότητα θα παραθέσουμε ορισμένες προσωπικές παρατηρήσεις του συγγραφέα σε συνδυασμό με μια ποιοτική αξιολόγηση των αποτελεσμάτων.

4.2.1 Μη ισορροπημένο σύνολο δεδομένων

Η πρώτη παρατήρηση αφορά τα χαρακτηριστικά ενός μη ισορροπημένου δείγματος. Συγκεκριμένα το σύνολο δεδομένων wikiart το οποίο χρησιμοποιήσαμε είχε ορισμένες κατηγορίες οι οποίες είχαν σημαντικά λιγότερα δεδομένα από τις υπόλοιπες, καθώς κι άλλες οι οποίες είχαν πολύ περισσότερα από το μέσο όρο. Παρατηρώντας τον πίνακα σύγχυσης (Confusion Matrix) ενός μοντέλου το οποίο εκπαιδεύτηκε στο wikiart μπορούμε να παρατηρήσουμε την επιρροή της ανισορροπίας του συνόλου δεδομένων στις προβλέψεις του μοντέλου. Στην εικόνα 4.9 παρουσιάζεται ο εν λόγω πίνακας σύγχυσης. Με κόκκινο χρώμα έχουμε σημειωθεί οι 5 κατηγορίες με το μικρότερο πλήθος δεδομένων (Synthetic Cubism, Pointillism, New Realism, Analytical Cubism και Action Painting), ενώ με πράσινο χρώμα έχουν σημειωθεί οι 5 μεγαλύτερες σε πληθικότητα κατηγορίες (Romanticism, Realism, Post Impressionism, Impressionism και Expressionism).

Υπενθυμίζουμε πως το πρόβλημα που δημιουργούν τα μη ισορροπημένα σύνολα δεδομένων είναι η τάση του ταξινομητή να επιλέγει κλάσεις οι οποίες έχουν μεγάλο πλήθος δεδομένων έναντι κλάσεων οι οποίες έχουν λιγότερα στοιχεία διότι απλώς αυτό του δίνει μεγαλύτερη ακρίβεια ή πιο συγκεκριμένα του κοστίζει λιγότερο σε ακρίβεια. Για να μιλήσουμε και με όρους μηχανικής μάθησης, οι κλάσεις με πολύ μεγαλύτερο πλήθος δεδομένων έχουν και πολλά περισσότερα false positives, ενώ σε αντίθεση η κλάσεις με πολύ μικρό πλήθος δεδομένων δεν έχουν σχεδόν καθόλου false positives. Αυτό ακριβώς παρατηρούμε και από τη συγκεκριμένη εικόνα. Οι στήλες αντιστοιχούν στις προβλέψεις του μοντέλου ενώ οι γραμμές ορίζουν την πραγματική κατηγορία. Παρατηρούμε λοιπόν πως οι στήλες που έχουν επισημανθεί με κόκκινο, εκτός της σωστής πρόβλεψης, είναι



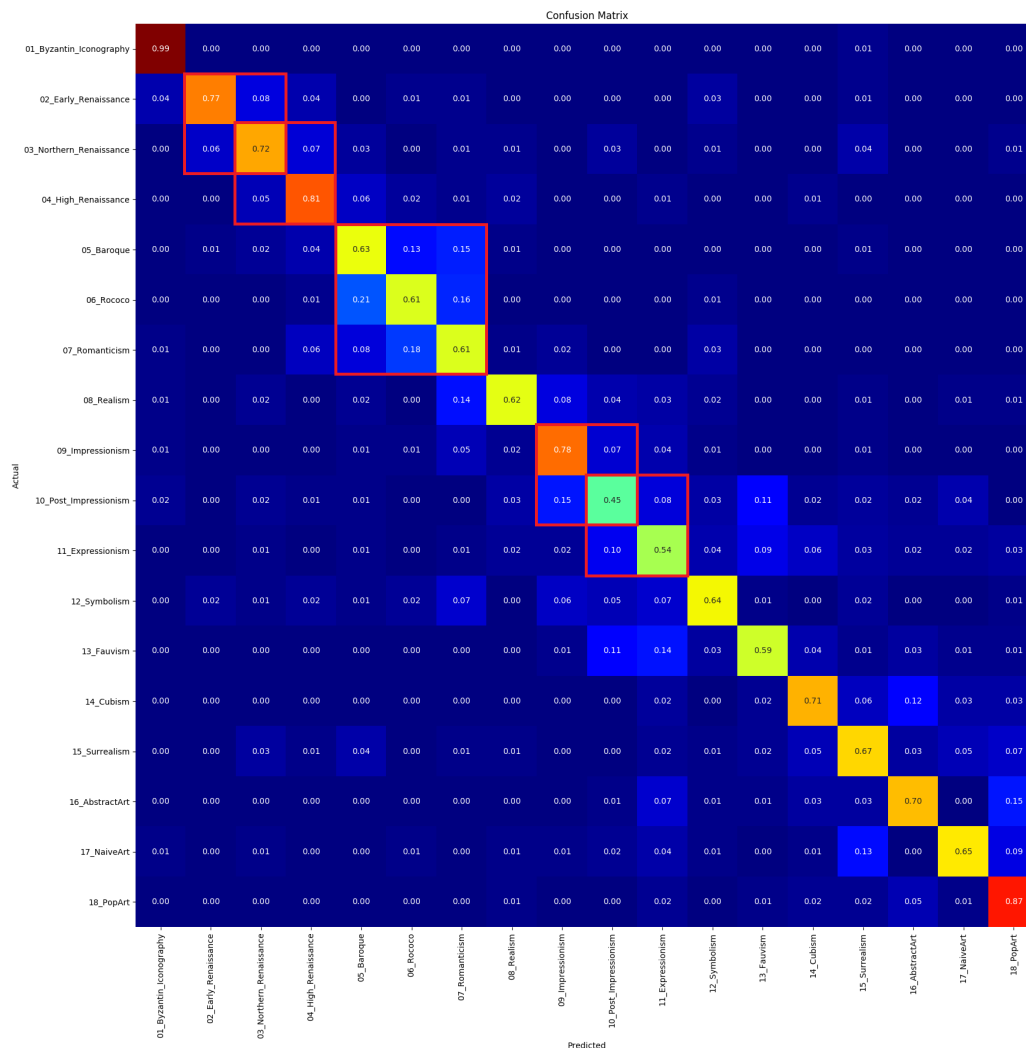
Σχήμα 4.9: Πίνακας σύγκρισης μη ισορροπημένων δεδομένων

όλες σκούρο μπλε με εξαίρεση μόνο 2-3 στοιχεία τα οποία όμως έχουν ξεχωριστό λόγο ύπαρξης ο οποίος θα αναλυθεί στην επόμενη ακριβώς ενότητα. Παρατηρούμε δηλαδή πως ο ταξινομητής δεν ταξινομεί (σχεδόν) καμία εικόνα λανθασμένα στις συγκεκριμένες κατηγορίες. Αντίθετα στις στήλες που έχουν επισημανθεί με πράσινο χρώμα παρατηρούμε πολλές θέσεις του πίνακα εκτός της κύριας διαγωνίου που δεν είναι σκούρο μπλε, άρα αντιστοιχούν σε λανθασμένες προβλέψεις. Βλέπουμε αυτή την τάση του ταξινομητή να επιλέγει τις κλάσεις που έχουν πολλά δεδομένα διότι και βάση πιθανοτήτων, είναι πιθανότερο μια εικόνα να ανήκει στις συγκεκριμένες κλάσεις. Αυτό το φαινόμενο βέβαια όπως έχει αναφερθεί ξανά, δεν επηρεάζει σημαντικά τη συνολική ακρίβεια του μοντέλου το οποίο μας απασχολεί στο συγκεκριμένο πρόβλημα, επηρεάζει όμως άλλες μετρικές οι οποίες θα μπορούσαν να είναι ζωτικής σημασίας σε άλλα προβλήματα.

4.2.2 Παρεμφερείς Κλάσεις

Φυσικά το μη ισορροπημένο δείγμα δεν είναι ο μόνος λόγος που τα μοντέλα μας ταξινομούν λάθος τα δεδομένα. Μεγάλο ενδιαφέρον, ειδικά από τη σκοπιά της τέχνης, έχει ο τρόπος με τον οποίο οι ταξινομητές μας προβλέπουν τις λάθος κλάσεις. Η προηγούμενη ενότητα αφορούσε την επιρροή της βαθιάς μάθησης στις λάθος προβλέψεις. Σε αυτή την ενότητα θα ασχοληθούμε με την επιρροή της ίδιας της τέχνης στις λάθος προβλέψεις. Αυτό που εννοούμε είναι πως πολλές φορές δεν ευθύνεται ο ίδιος ο ταξινομητής, ή τουλάχιστον όχι εξολοκλήρου, για τις λάθος προβλέψεις. Υπάρχουν πολλές τεχνοτροπίες οι οποίες έχουν επηρεαστεί και μοιάζουν σε μεγάλο βαθμό οπτικά με άλλες. Επίσης υπάρχουν έργα τα οποία έχουν στοιχεία από πολλές τεχνοτροπίες όμως ο ταξινομητής μας καλείται να επιλέξει μόνο μία. Παρατηρώντας τους πίνακες σύγχυσης μπορούμε να εντοπίσουμε κλάσεις οι οποίες συχνά περιπλέκονται στις προβλέψεις μιας και το μοντέλο προβλέπει πολλά δεδομένα της μιας κλάσης ως στοιχεία της άλλης κι αντίστροφα. Στην εικόνα [4.10](#) παρουσιάζεται ο Confusion Matrix ενός μοντέλου που εκπαιδεύσαμε στο σύνολο δεδομένων Pandora18k.

Στον πίνακα έχουν σημειωθεί και με κόκκινο τετράγωνο ορισμένες ομάδες στοιχείων του πίνακα που μας δείχνουν τη συσχέτιση μεταξύ κλάσεων. Η πιο εμφανής από όλες είναι η συσχέτιση μεταξύ Baroque, Rococo και Romanticism. Πρόκειται για το τετράγωνο 3×3 που έχει σημειωθεί στον πίνακα και βλέπουμε πως ο ταξινομητής δυσκολεύεται ιδιαίτερα να ξεχωρίσει τις τρεις τεχνοτροπίες. Ταξινομεί μεγάλο μέρος των Baroque έργων τέχνης ως Rococo ή Romanticism κι αντίστροφα. Αυτό όμως οφείλεται κυρίως στο γεγονός πως οι συγκεκριμένες τεχνοτροπίες είναι πολύ κοντά τόσο αισθητικά όσο και χρονολογικά. Το Rococo συχνά αναφέρεται ως ύστερο Baroque (late Baroque) και το κίνημα του Ρομαντισμού επηρεάστηκε σημαντικά από καλλιτέχνες του κινήματος του ύστερου Μπαρόκ. Αντίστοιχα παρατηρούμε συσχετίσεις μεταξύ Ιμπρεσιονισμού και μετα-ιμπρεσιονισμού, Early, Northern και High Renaissance κ.α. Επίσης στο σύνολο δεδομένων wikiart και συγκεκριμένα στην αρχική του μορφή, παρατηρούμε συχνά να συγχέονται οι κλάσεις Cubism, Analytical Cubism και Synthetic Cubism. Όπως υποδεικνύουν και τα ονόματά τους πρόκειται για τρεις κλάσεις τα έργα των οποίων έχουν πάρα πολλά κοινά χαρακτηριστικά και μάλιστα από τους περισσότερους θεωρούνται ότι όλες ανήκουν στην τεχνοτροπία του κυβισμού (για αυτό το λόγο και τις συγχωνεύσαμε στο wikiart2). Παρατηρούμε λοιπόν γενικά πως ένα νευρωνικό δίκτυο είναι σε θέση να μας δώσει περισσότερη πληροφορία πέρα από την ταξινόμηση η οποία αποτελεί τον άμεσο στόχο. Ο πίνακας της εικόνας [4.10](#), αποτελεί πολύ καλό εργαλείο για τον εντοπισμό τέτοιων συστάδων στοιχείων μιας και οι τεχνοτροπίες είναι με (περίπου) χρονολογική σειρά. Είναι πολύ πιθανότερο δυο τεχνοτροπίες οι οποίες συνυπήρξαν ή η μια διαδέχθηκε την άλλη να έχουν πολλά κοινά στοιχεία τα οποία είναι εμφανή και στις προβλέψεις. Χάρη στη χρονολογική σειρά παρατηρούμε πως το μεγαλύτερο μέρος των λάθος προβλέψεων βρίσκεται γύρω από την κύρια διαγώνιο για αυτόν ακριβώς το λόγο. Μπορούμε λοιπόν να αντλήσουμε πληροφορίες σχετικά με την ομοιότητα ή μη δύο τεχνοτροπιών κι ακόμη και να εντοπίσουμε συσχετίσεις τις οποίες δεν είχαμε εντοπίσει νωρίτερα απλώς παρατηρώντας τις λάθος προβλέψεις ενός ταξινομητή. Αυτό μπορεί να μας βοηθήσει να κατανοήσουμε βαθύτερα τις καλλιτεχνικές τεχνοτροπίες και τα καλλιτεχνικά κινήματα καθώς και τα στοιχεία που τα συνθέτουν. Αυτό που είναι πολύ ενδιαφέρον σχετικά με το συγκεκριμένο κομμάτι είναι πως και οι υπολογιστές αντιλαμβάνονται τις ίδιες συσχετίσεις που αντιλαμβάνονται και οι άνθρωποι. Εντοπίζουν ομοιότητες μεταξύ τεχνοτροπιών μέσω των λαθών



Σχήμα 4.10: Πίνακας σύγχυσης - Παρεμφερείς Κλάσεις

που κάνουν. Μπορούμε λοιπόν ερμηνεύοντας κατάλληλα τα αποτελέσματα, να αντλήσουμε πληροφορία ακόμη κι από τα λάθη που κάνει ένας ταξινομητής, και πολλές φορές αυτή η πληροφορία μπορεί να έχει και πολύ μεγαλύτερο βάθος και ουσία.

4.2.3 Ιδιαίτερες Κλάσεις και Ιδιαίτερα Χαρακτηριστικά

Σε αντίθετη τροχιά με τα προηγούμενα, κατά την εκτέλεση των πειραμάτων παρατηρήσαμε πως υπήρχαν ορισμένες κλάσεις οι οποίες σταθερά έδιναν πολύ καλύτερα αποτελέσματα από τις υπόλοιπες. Στο σύνολο δεδομένων Pandora18k οι κλάσεις Byzantine Iconography και Pop Art έδωσαν σχεδόν σε όλα τα πειράματα ιδιαίτερα υψηλά ποσοστά ακρίβειας ακόμη και σε περιπτώσεις όπου η συνολική ακρίβεια του μοντέλου δεν ήταν υψηλή. Αντίστοιχα στα σύνολα δεδομένων βασισμένα στο wikiart η κλάση Ukiyo-e που αντιστοιχεί στην ομώνυμη ιαπωνική τεχνοτροπία έδωσε σταθερά υψηλά ποσοστά ακρίβειας. Ο βασικός λόγος για αυτό είναι η μεγάλη οπτική δια-

φορά όπου έχουν οι συγκεκριμένες κλάσεις σε σύγκριση με τις υπόλοιπες. Η Ιαπωνική τεχνοτροπία Ukiyo-e έχει ένα ιδιαίτερο στιλ το οποίο δεν μοιάζει με κανένα άλλο από αυτά που μελετάμε. Αυτό κάνει τα έργα τέχνης της συγκεκριμένης τεχνοτροπίας να ταξινομούνται στο μεγαλύτερο βαθμό ορθά. Ρίχνοντας μια κοντινότερη ματιά όμως στο σύνολο δεδομένων εντοπίζουμε ένα ακόμη χαρακτηριστικό το οποίο μπορεί να βοηθάει τον ταξινομητή να αναγνωρίσει τη συγκεκριμένη κλάση το οποίο όμως δεν σχετίζεται απαραίτητα με την τεχνοτροπία καθαυτή, τουλάχιστον όχι μόνο με το ζωγραφικό κομμάτι αυτής. Στις εικόνες 4.11, 4.12 και 4.13 παρουσιάζονται μερικά έργα τέχνης της συγκεκριμένης τεχνοτροπίας τα οποία όπως παρατηρούμε ένα ιδιαίτερο σχήμα. Ο λόγος για αυτό το σχήμα είναι πως τα συγκεκριμένα έργα τέχνης είναι βεντάλιες. Ο καλλιτέχνης Utagawa Hiroshige, ο οποίος είναι και δημιουργός των συγκεκριμένων, δημιούργησε πολλά έργα τέχνης πάνω σε ξύλινες βεντάλιες χρησιμοποιώντας ορισμένες φορές και κομμάτια χαρτιού. Παρατηρούμε λοιπόν πως στην κλάση Ukiyo-e υπάρχουν έργα τέχνης με ένα ιδιαίτερο σχήμα το οποίο δεν υπάρχει σε καμία άλλη κλάση, άρα πιθανότατα ο ταξινομητής αντιλαμβάνεται κι αυτό ως ένα χαρακτηριστικό της συγκεκριμένης κλάσης. Φυσικά η συγκεκριμένη κλάση δεν αποτελείται μόνο από βεντάλιες, μάλιστα οι βεντάλιες είναι μόνο ένα μικρό κομμάτι αυτής, άρα ο ταξινομητής πρέπει να μάθει και άλλα χαρακτηριστικά τα οποία ορίζονται από το ζωγραφικό στιλ της τεχνοτροπίας, όμως αυτό δεν σημαίνει πως και το ιδιαίτερο σχήμα δεν βοηθάει το μοντέλο να ταξινομήσει καλύτερα ορισμένα έργα τέχνης.



Σχήμα 4.11: "Amanohashidate in Tango Province", Utagawa Hiroshige



Σχήμα 4.12: "Chrysanthemums", Utagawa Hiroshige



Σχήμα 4.13: "Itsukushima in Aki Province", Utagawa Hiroshige

Η Pop Art αντίστοιχα διαφέρει αρκετά από τις υπόλοιπες τεχνοτροπίες χάρη στη χρήση έντονων και ζωντανών χρωμάτων τα οποία είναι χαρακτηριστικά της συγκεκριμένης τεχνοτροπίας. Επίσης συχνή είναι η απότομη αλλαγή χρωμάτων στο οποίο συμβάλει και το γεγονός πως συχνά έργα pop τέχνης είχαν στοιχεία από κολάζ, με αποτέλεσμα στα σημεία ένωσης των κομματιών να έχουμε αυστηρές γραμμές κι απότομη εναλλαγή πολύ διαφορετικών χρωμάτων το οποίο δεν είναι πολύ συνηθισμένο, ιδιαίτερα στις τεχνοτροπίες οι οποίες αναπτύχθηκαν πριν τον εικοστό αιώνα. Αυτός είναι κι ο βασικός λόγος για τον οποίο παρατηρούμε τα πολύ υψηλά ποσοστά ακρίβειας. Στις εικόνες 4.14, 4.15 και 4.16 παρουσιάζονται τρεις Pop Art πίνακες, από δύο από τους πιο αναγνωρίσιμους καλλιτέχνες του pop κινήματος, τους Andy Warhol και Roy Lichtenstein, στις οποίες είναι εμφανή τα χαρακτηριστικά που αναφέρθηκαν παραπάνω.

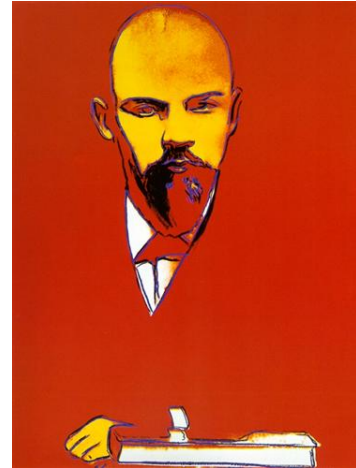
Αντίστοιχα με την τεχνοτροπία Ukiyo-e και η βυζαντινή εικονογραφία έχει το δικό της μοναδικό καλλιτεχνικό στιλ. Είναι αρκετά δύσκολο να ταξινομηθεί λάθος ένας πίνακας αυτής της κλάσης μιας και τα χαρακτηριστικά της δεν θυμίζουν καμία από τις υπόλοιπες κλάσεις. Για αυτό το λόγο παρατηρούμε ότι η συγκεκριμένη κλάση έχει ακρίβεια που πλησιάζει συχνά το 100% στα



Σχήμα 4.14: "Blue Marilyn", Andy Warhol



Σχήμα 4.15: "In the car", Roy Lichtenstein



Σχήμα 4.16: "Red Lenin", Andy Warhol

περισσότερα μοντέλα που εκπαιδεύσαμε. Μελετώντας όμως περισσότερο το σύνολο δεδομένων μας εντοπίζουμε κι εδώ ένα ιδιαίτερο χαρακτηριστικό το οποίο δεν αφορά την τεχνοτροπία, είναι όμως παρόν σε πολλές από τις εικόνες της βυζαντινής εικονογραφίας και ο ταξινομητής μπορεί να το χρησιμοποιεί για την καλύτερη ταξινόμηση των δεδομένων. Το χαρακτηριστικό αυτό είναι η φθορά. Το γεγονός πως η βυζαντινή εικονογραφία είναι η παλαιότερη τεχνοτροπία με μεγάλη διαφορά από τις υπόλοιπες στο σύνολο δεδομένων, καθώς και το γεγονός ότι πολλά έργα τέχνης είναι τοιχογραφίες οι οποίες φθείρονται πιο εύκολα, έχουν ως αποτέλεσμα η φθορά να είναι εμφανής σε πολλές από τις εικόνες των έργων τέχνης που χρησιμοποιούμε. Στα σχήματα 4.17 και 4.18 παρουσιάζονται μερικά παραδείγματα εικόνων από το σύνολο δεδομένων όπου η φθορά είναι εμφανής.



Σχήμα 4.17: Βυζαντινή Εικονογραφία 1



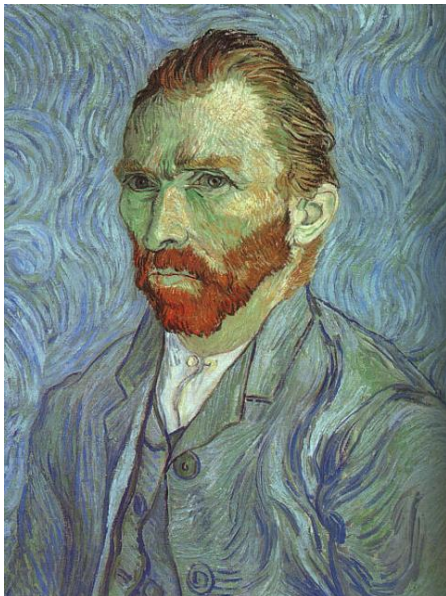
Σχήμα 4.18: Βυζαντινή Εικονογραφία 2

4.2.4 Παραδείγματα με μεμονωμένες προβλέψεις

Για να αποδείξουμε τη λειτουργικότητα των μοντέλων μας αλλά και να βοηθήσουμε τον αναγνώστη να κατανοήσει καλύτερα τον τρόπο με τον οποίο εξάγονται τα αποτελέσματα ενός CNN, βάλαμε ένα από τα μοντέλα που δημιουργήσαμε να προβλέψει την τεχνοτροπία σε μερικές μεμονωμένες εικόνες. Χρησιμοποιήσαμε πασίγνωστα έργα τέχνης από το train set καθώς κι άλλα από το test set, ακόμη κι εκτός του συνόλου δεδομένων. Στον πίνακα 4.12 παρουσιάζονται τα αποτελέσματα των προβλέψεων.



Σχήμα 4.19: "Guernica", Pablo Picasso, 1937, Cubism and Surrealism



Σχήμα 4.20: "Self-portrait", Vincent Van Gogh, 1889, Post-Impressionism



Σχήμα 4.21: "Melting Watch", Salvador Dali, 1954, Surrealism

Αναμενόμενα παρατηρούμε ο ταξινομητής να προβλέπει με πολύ μεγάλη ακρίβεια την τεχνοτροπία του έργου "Guernica" του Pablo Picasso και του έργου "Mona Lisa" του Leonardo da Vinci, τα οποία αποτελούν μέρος του train set του Pandora18k. Το μοντέλο έχει εκπαιδευτεί πάνω στα συγκεκριμένα δεδομένα άρα έχει προσαρμόσει και τα βάρη του ώστε να τα προβλέπει με ακρίβεια.

Πίνακας 4.12: Προβλέψεις Μεμονωμένων Εικόνων

Piece of Art	Prediction	Actual	Part of Dataset
"Guernica"(4.19)	1.Surrealism 99,93%	Surrealism	YES - train set
"Mona Lisa" (4.22)	1.High Renaissance 99,99%	High Renaissance	YES - train set
"Portrait of AG" (4.24)	1.Expressionism 83,35% 2.Post-Impressionism 4,95%	Post-Impressionism	YES - test set
"Man with knife" (4.23)	1.Abstract Art 86,30% 2.Expressionism 12,39%	Abstract Art	YES - test set
"Self-portrait" (4.20)	1.Post-Impressionism 95,35%	Post-Impressionism	YES - test set
"Melting Watch" (4.21)	1.Surrealism 99,95%	Surrealism	NO

Περισσότερο ενδιαφέρον έχουν οι προβλέψεις των υπόλοιπων έργω τέχνης. Από τα πειράματα που εκτελέσαμε, σε μόνο ένα ο ταξινομητής μας προέβλεψε λάθος τεχνοτροπία, κι αυτό ήταν το "Portrait of AG" του Παναγιώτη Τέτση. Το μοντέλο μας κατέταξε τον συγκεκριμένο πίνακα στην κλάση του Εξπρεσιονισμού ενώ στην πραγματικότητα ταιριάζει περισσότερο στην τεχνοτροπία του μετα-Ιμπρεσιονισμού (Post-Impressionism). Αυτό μας δείχνει όσα αναφέρθηκαν και παραπάνω για παρεμφερείς τεχνοτροπίες μιας και τα δύο αυτά κινήματα μοιράζονται πολλά κοινά χαρακτηριστικά. Αξίζει να σημειωθεί πως η δεύτερη σε σειρά τεχνοτροπία που προέβλεψε το μοντέλο ήταν η ορθή, με μεγάλη όμως ποσοστιαία διαφορά από την πρώτη όπως παρατηρούμε κι από τον πίνακα. Τους υπόλοιπους πίνακες ζωγραφικής τους οποίους δεν είχε ξαναδεί ο ταξινομητής τους κατέταξε σωστά και με πολύ υψηλά ποσοστά πιθανοτήτων τους περισσότερους. Μάλιστα το μοντέλο ανταποκρίθηκε εντυπωσιακά στην είσοδο ενός έργου τέχνης το οποίο δεν υπάρχει καν στο σύνολο δεδομένων, το "Melting Watch" του Salvador Dali, προβλέποντας τη σωστή τεχνοτροπία με πιθανότητα 99,95%.



Σχήμα 4.22: "Mona Lisa", Leonardo da Vinci, 1503, High Renaissance



Σχήμα 4.23: "Man with knife", Jackson Pollock, 1940, Abstract Expressionism



Σχήμα 4.24: "Portrait of AG", Παναγιώτης Τέτσης, 1954, Post-Impressionism

Πίνακας 4.13: Συγκριτικά Αποτελέσματα Pandora18k

Εργασία	Test Accuracy
Florea et al. (2017)	50,1%
Yu et al. (2017)	56,6%
Florea and Gieseke (2018)	63,5%
Rodriguez et al. (2018)	68,1% (*70,2%)
Sandoval et al. (2019)	76,0% (*77,5%)
This work	70,3%

4.3 Σύγκριση Αποτελεσμάτων

Στη συγκεκριμένη ενότητα θα συγκρίνουμε τα αποτελέσματά μας με τα αποτελέσματα άλλων εργασιών που έχουν γίνει πάνω στην αναγνώριση τεχνοτροπίας. Θεωρούμε πως δεν έχει αξία να συγκριθούμε με εργασίες που έγιναν πολύ παλαιότερα με πολύ μικρότερα σύνολα δεδομένων διότι δεν έχουμε κάποιο κοινό σημείο αναφοράς. Αποφασίσαμε λοιπόν να παρουσιάσουμε μια συγκριτική μελέτη ανάμεσα σε εργασίες οι οποίες έχουν χρησιμοποιήσει τα σύνολα δεδομένων που χρησιμοποιήσαμε κι εμείς ή παρεμφερή. Θα εξετάσουμε πρώτα τα αποτελέσματα που επιτύχαμε στο σύνολο δεδομένων Pandora18k και στη συνέχεια στα σύνολα δεδομένων wikiart και wikiart2 μαζί μιας και το τελευταίο αποτελεί απλώς μια παραλλαγή του wikiart.

4.3.1 Pandora18k

Το σύνολο δεδομένων Pandora18k δημιουργήθηκε το 2017 από τους [Florea et al. \(2017\)](#) όπως έχει αναφερθεί ήδη. Πρόκειται λοιπόν για ένα σχετικά καινούριο σύνολο δεδομένων, χάρη όμως στην ισορροπία και την εγκυρότητά του, χρησιμοποιείται πολύ συχνά στις μελέτες της τεχνοτροπίας που έχουν γίνει έκτοτε. Στον πίνακα 4.13 παρουσιάζονται συνοπτικά τα αποτελέσματα όσων εργασιών έχουν γίνει μέχρι σήμερα πάνω στην αναγνώριση τεχνοτροπίας κι έχουν χρησιμοποιήσει το σύνολο δεδομένων Pandora18k.

Αρχικά να εξηγήσουμε τη σημασία των αστερίσκων σε δύο από τις καταχωρήσεις. Πρόκειται για την ίδια ομάδα, αποτελούμενη επί των πλείστων από άτομα του Πανεπιστημίου RMIT της Μελβούρνης, η οποία πραγματεύτηκε το πρόβλημα της αναγνώρισης τεχνοτροπίας. Η συγκεκριμένη ομάδα χρησιμοποίησε το σύνολο δεδομένων Pandora18k στο οποίο όμως προσέθεσε και μια κατηγορία με Αυστραλιανή τέχνη και τέχνη των Αβορίγιων. Η συγκεκριμένη κατηγορία όμως λόγω του ότι είναι πολύ διαφορετική από τις υπόλοιπες έδωσε πολύ υψηλά ποσοστά ακρίβειας, ανεβάζοντας τη συνολική ακρίβεια του μοντέλου τους. Σύμφωνα με τα δεδομένα που έδωσαν στις εργασίες τους κάναμε αναγωγή των ποσοστών στο σύνολο δεδομένων Pandora18k όπου είναι και τα ποσοστά που παρουσιάζονται στον πίνακα. Στην παρένθεση βρίσκονται τα συνολικά ποσοστά του μοντέλου τους, περιλαμβανομένης και της επιπλέον κλάσης. Η δουλειά που έγινε από τις περισσότερες ομάδες εκτός αυτής του RMIT, επικεντρώθηκε στην εκπαίδευση ενός CNN και τη βελτιστοποίηση αυτού. Και στις εργασίες των [Sandoval et al. \(2019\)](#) και [Rodriguez et al. \(2018\)](#) φυσικά ξεκινάμε με βάση ένα απλό μοντέλο ταξινόμησης. Αξίζει να σημειωθεί πως με βάση το test accuracy στα βασικά μοντέλα όλων των εργασιών, το δικό μας πέτυχε τα υψηλότερα ποσοστά σε όλα τα σύνολα δεδομένων. Αυτό που επέτρεψε όμως στην ομάδα του RMIT να πάει ένα βήμα παραπέρα είναι αρχικά η χρήση ενός συστήματος συνδυασμού patches για την πρόβλεψη ([Rodriguez](#)

et al., 2018) αλλά αυτό που έκανε το άλμα στις επιδόσεις ήταν η χρήση μιας μεθόδου boosting στην τελευταία εργασία τους (Sandoval et al., 2019) όπου εκπαιδεύουν ένα μικρό νευρωνικό δίκτυο στην κορυφή του CNN που είχαν δημιουργήσει στην προηγούμενη εργασία. Αναλυτικότερα η μεθοδολογία που ακολούθησαν έχει παρουσιαστεί στην ενότητα 2.1.1. Αυτό που μπορούμε λοιπόν να κρατήσουμε αναφορικά με την επιτυχία της δουλειάς που κάναμε εμείς είναι πως πετύχαμε το υψηλότερο test accuracy με τη χρήση ενός και μόνο CNN ενώ η μόνη εργασία η οποία κατάφερε να ξεπεράσει τα ποσοστά μας, χρησιμοποίησε δύο τεχνικές οι οποίες δεν είχαν εφαρμοστεί στο συγκεκριμένο πρόβλημα μέχρι τη δεδομένη στιγμή. Ακόμη κι η πρώτη εργασία της συγκεκριμένης ομάδας η οποία εισήγαγε τη διαμέριση της εικόνας σε τμήματα και το συνδυασμό αυτών στο τέλος, μια πρακτική που έδωσε σημαντική αύξηση της ακρίβειας στα μοντέλα τους, δεν κατάφερε να φτάσει μεγαλύτερα ποσοστά από τα δικά μας μοντέλα. Αντιλαμβανόμαστε λοιπόν πως αν και υπάρχουν ακόμη περιθώρια βελτίωσης ενός και μόνο CNN, και το αποδείξαμε με την παρούσα εργασία, για να καταφέρουμε να φτάσουμε ένα σκαλοπάτι παραπάνω πρέπει να ξεφύγουμε από την πεπατημένη και να αντλήσουμε έμπνευση από τις τεχνικές που χρησιμοποιούνται σε άλλα προβλήματα βαθιάς μάθησης. Με αυτό τον τρόπο πετύχαμε τα κορυφαία αποτελέσματα για την εργασία αυτή, συνδυάζοντας πάνω από ένα δίκτυα μέσω μιας μορφής της μεθόδου bagging, όπως αναφέρθηκε σε παραπάνω ενότητα.

4.3.2 Wikiart

Το σύνολο δεδομένων wikiart προέρχεται από την ομώνυμη ψηφιακή εγκυκλοπαίδεια εικαστικών τεχνών και χρησιμοποιήθηκε για πρώτη φορά από τους Karayev et al. (2013). Έκτοτε αποτέλεσε τη βάση όλων σχεδόν των συνόλων δεδομένων που χρησιμοποιήθηκαν για επεξεργασία κι αναγνώριση τέχνης με βαθιά μάθηση. Αναλυτική περιγραφή και ανάλυση των προτερημάτων και των αδυναμιών του συγκεκριμένου συνόλου δεδομένων έχει γίνει σε παραπάνω κεφάλαιο. Στη συγκεκριμένη ενότητα θα ασχοληθούμε μόνο με τα αποτελέσματα που προέκυψαν από διάφορες εργασίες που εργάστηκαν με αυτό. Το πρόβλημα στη σύγκριση των αποτελεσμάτων με το συγκεκριμένο σύνολο δεδομένων είναι πως κάθε ομάδα έχει εργαστεί με μια ελαφρώς διαφοροποιημένη έκδοση του αρχικού dataset. Άλλες ομάδες προσέθεσαν δεδομένα, άλλες αφαίρεσαν δεδομένα ή/και κλάσεις, άλλες ομαδοποίησαν κλάσεις κλπ. Αντίστοιχα κι εμείς εφαρμόσαμε ορισμένες τροποποιήσεις και δημιουργήσαμε το wikiart2 όπως αναφέρεται παραπάνω. Για αυτό το λόγο λοιπόν μαζί με τα ποσοστά test accuracy θα παρουσιάζονται με μερικά επιπλέον χαρακτηριστικά του συνόλου δεδομένων που χρησιμοποιήθηκε σε κάθε περίπτωση, όπως το πλήθος των δεδομένων και ο αριθμός των κλάσεων. Όμοια με παραπάνω, στον πίνακα 4.14 παρουσιάζονται τα αποτελέσματα.

Από τα πρώτα πράγματα που παρατηρούμε είναι πως σχεδόν καμία εργασία δεν έχει δουλέψει με ακριβώς το ίδιο σύνολο δεδομένων με κάποια άλλη. Αυτό κάνει τη σύγκριση κάπως άνιση σε ορισμένες περιπτώσεις μιας και για παράδειγμα οι Sandoval et al. (2019) δούλεψαν με ένα dataset το οποίο προήλθε από το wikiart, το επεξεργάστηκαν όμως έτσι ώστε να δημιουργήσουν ένα πλήρως ισορροπημένο σύνολο δεδομένων προσθέτοντας και πάλι μια επιπλέον κλάση με Αυστραλιανή τέχνη και τέχνη των Αβορίγιων. Σε κάθε περίπτωση όμως οι επιδόσεις των μοντέλων μας είναι οι κορυφαίες. Υπήρξαν ομάδες οι οποίες δούλεψαν με χαρακτηριστικά που εξήγαγαν από τις εικόνες, ενώ άλλες με αυτούσιες τις εικόνες όπως κάναμε κι εμείς. Υπήρξαν πολλές διαφορετικές προσεγγίσεις, με τα CNNs να έχουν κυρίαρχη θέση στις περισσότερες. Σε γενικές γραμμές οι επιδόσεις

Πίνακας 4.14: Συγκριτικά Αποτελέσματα Wikiart

Εργασία	Test Accuracy	Κλάσεις	Πλήθος Δεδομένων
Karayev et al. (2013)	44,1%	25	85000
Bar et al. (2014)	43,0%	27	40724
Saleh and Elgammal (2015)	46,0%	27	81449
Tan et al. (2016)	54,5%	27	80000
Lecoutre et al. (2017)	62,8%	25	80000
Florea and Gieseke (2018)	46,2%	25	85000
Cetinic et al. (2018)	56,4%	27	96014
Zhong et al. (2019)	59,0%	25	30825
Sandoval et al. (2019)	66,7%	22	26400
This work - wikiart	63,5%	27	81445
This work - wikiart2	67,9%	21	80039

των μοντέλων μας στα συγκεκριμένα σύνολα δεδομένων είναι πολύ ικανοποιητικές, χρήζουν όμως φυσικά μεγάλης βελτίωσης, μιας και υπάρχει περιθώριο αύξησης της ακρίβειας του με καλύτερη προσαρμογή των υπερπαραμέτρων το οποίο όπως δείξαμε στη συγκεκριμένη εργασία παίζει πολύ σημαντικό ρόλο, αλλά φυσικά και με την εφαρμογή μεθόδων συνδυαστικής μάθησης όπως boosting και bagging τα οποία όπως είδαμε συμβάλουν σημαντικά στη βελτίωση της ακρίβειας των ταξινομητών.

Κεφάλαιο 5

Επίλογος

Στο παρόν 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 πραγματευθήκαμε το πρόβλημα της αναγνώρισης τεχνοτροπίας έργων τέχνης. Πρόκειται για ένα δύσκολο κι απαιτητικό πρόβλημα ακόμη και για τον άνθρωπο. Σκοπός μας ήταν η μελέτη κι ανάλυση της λειτουργίας των συνελκτικών νευρωνικών δικτύων, και του τρόπου που επηρεάζουν την εκπαίδευσή τους οι διάφορες παράμετροι τόσο του περιβάλλοντος όσο και του ίδιου του νευρωνικού δικτύου. Αναλυτικότερα μελετήσαμε την επιρροή των δεδομένων στην εκπαίδευση και την ακρίβεια ενός μοντέλου, δείχνοντας μέσω πειραμάτων την επίδραση ενός μη ισορροπημένου dataset, και τη σημασία της προ-επεξεργασίας των δεδομένων. Εργαστήκαμε με δύο βασικά σύνολα δεδομένων με αρκετές διαφορές μεταξύ τους και εντοπίσαμε τις επιρροές του καθενός στο δίκτυο. Αποδείξαμε τη χρησιμότητα του data augmentation και τη συμβολή του στην αντιμετώπιση του overfitting. Μελετήσαμε την καθοριστική σημασία που έχει το learning rate για ένα μοντέλο και τις καταστάσεις στις οποίες οδηγεί ένα δίκτυο όταν παίρνει πολύ υψηλές ή χαμηλές τιμές. Εργαστήκαμε χρησιμοποιώντας την τεχνική του Transfer Learning και δείξαμε μέσω πειραμάτων τις τρομερές δυνατότητες που έχει η συγκεκριμένη μέθοδος. Εξετάσαμε τα χαρακτηριστικά πολλών αρχιτεκτονικών συνελκτικών νευρωνικών δικτύων και αναλύσαμε τα προτερήματα και τα μειονεκτήματα της καθέμιας. Εκτελέσαμε συγκριτικά πειράματα ανάμεσα σε οκτώ διαφορετικές αρχιτεκτονικές, περισσότερες από οποιαδήποτε άλλη εργασία έχει πραγματευθεί το πρόβλημα της αναγνώρισης τεχνοτροπίας. Μάλιστα χρησιμοποιήσαμε ορισμένες αρχιτεκτονικές όπως η DenseNet121 και η Inception - ResNetV2 οι οποίες έδωσαν ιδιαίτερα ελπιδοφόρα αποτελέσματα και δεν έχουν ξαναχρησιμοποιηθεί σε αντίστοιχο πρόβλημα. Εφαρμόσαμε μια μέθοδο συνδυαστικής μάθησης (ensemble learning) η οποία αποτελεί μια μορφή bagging κατά την οποία συνδυάζονται οι προβλέψεις δύο ή περισσότερων ταξινομητών και μέσω ενός γραμμικού συνδυασμού των αποτελεσμάτων προκύπτει η έξοδος ενός νέου υπερ-ταξινομητή. Με τη συγκεκριμένη μέθοδο πετύχαμε σημαντική αύξηση της ακρίβειας των μοντέλων μας, ξεπερνώντας το φράγμα του 70% στην περίπτωση του συνόλου δεδομένων Pandora18k και διαμορφώνοντας το βέλτιστο αποτέλεσμα που επιτύχαμε το οποίο έφτασε το 70,2%. Αντίστοιχα στο σύνολο δεδομένων wikiart πετύχαμε την κορυφαία ακρίβεια που έχει δημοσιευθεί μέχρι σήμερα κι έφτασε το 67,9%. Κα-

τανοήσαμε βαθύτερα κομμάτια της τέχνης και συνδυάζοντας τη νέα αυτή γνώση με τη γνώση μας σχετικά με τη μηχανική μάθηση οδηγηθήκαμε σε ορισμένα ενδιαφέροντα ποιοτικά συμπεράσματα σχετικά με τη συμπεριφορά των μοντέλων. Παρατηρήσαμε τα "ανθρώπινα" λάθη που κάνουν οι ταξινομητές ανάμεσα σε τεχνοτροπίες οι οποίες μοιράζονται κοινά χαρακτηριστικά. Αναλύσαμε τις θεωρητικές και πρακτικές δυσκολίες της αναγνώρισης της τεχνοτροπίας ενός έργου τέχνης, πολλές από τις οποίες εμφανίστηκαν και στα πειράματά μας. Μέσα από τη μελέτη κι εφαρμογή όσων διαβάσαμε, εξοικειωθήκαμε με τα εργαλεία και τις μεθόδους της βαθιάς μάθησης και κατανοήσαμε σε βάθος τις αρχές και τη λειτουργία των συνελκτικών νευρωνικών δικτύων. Αυτή η διπλωματική εργασία λοιπόν αποτελεί μια συνολική παρουσίαση των CNN και των παραμέτρων που τα επηρεάζουν, καθώς και μια συνολική συγκριτική διαδικασία ανάμεσα στις εργασίες που έχουν δημοσιευθεί μέχρι στιγμής σχετικά με την αναγνώριση τεχνοτροπίας. Επιτυγχάνει αξιοσημείωτη ακρίβεια με πολύ λίγους πόρους σε ορισμένα κομμάτια και τα μοντέλα που αναπτύχθηκαν στο πλαίσιο αυτής ξεπερνούν τις επιδόσεις των περισσότερων state of the art μοντέλων.

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. Επιγραμματικά αυτά είναι:

- Τα συνελκτικά νευρωνικά δίκτυα δείχνουν να υπερέχουν των υπόλοιπων νευρωνικών δικτύων στον τομέα της επεξεργασίας κι αναγνώρισης εικόνων.
- Τα δεδομένα αποτελούν αναπόσπαστο κι από τα σπουδαιότερα κομμάτια ενός μοντέλου μηχανικής μάθησης, γι' αυτό και πρέπει πάντα να επιλέγονται τα κατάλληλα δεδομένα, να μελετούνται καλά και να τα προ-επεξεργαζόμαστε ανάλογα. Ένα μικρό ή/και μη έγκυρο σύνολο δεδομένων δεν μπορεί να δώσει ιδιαίτερα αξιόπιστα αποτελέσματα, αλλά το ικανοποιητικό μέγεθος για ένα σύνολο δεδομένων καθορίζεται από το ίδιο το πρόβλημα.
- Το overfitting αποτελεί από τα μεγαλύτερα προβλήματα κατά την εκπαίδευση μοντέλων βαθιάς μάθησης, η αντιμετώπιση του οποίου είναι καίριας σημασίας για την επιτυχία ενός μοντέλου. Η μέθοδος του data augmentation συμβάλει τα μέγιστα προς αυτή την κατεύθυνση.
- Το πλήθος των παραμέτρων μιας αρχιτεκτονικής δεν καθορίζει απαραίτητα την ικανότητά της και καλό είναι να δοκιμάζονται και νέες αρχιτεκτονικές οι οποίες μπορεί να μην έχουν ξαναεφαρμοστεί στο εκάστοτε πρόβλημα.
- Το learning rate αποτελεί μια βασικότερη υπερπαραμέτρο η οποία θέλει ιδιαίτερη προσοχή στον προσδιορισμό των τιμών της.

- Η χρήση της μεθόδου transfer learning είναι ζωτικής σημασίας για την επιτυχία ενός μοντέλου και μας επιτρέπει να φτάσουμε γρήγορα σε επίπεδα ακρίβειας χωρίς της οποίας θα ήταν σχεδόν αδύνατο.
- Η χρήση μεθόδων ensemble learning μπορεί να οδηγήσει σε σημαντική βελτίωση των αποτελεσμάτων, πρέπει όμως να μελετηθεί και να επιλεγεί η κατάλληλη.
- Η περικοπή έργων τέχνης και η μελέτη τους μόνο κατά τμήματα προσφέρει πληροφορία όσο και αφαιρεί. Η χρήση τμημάτων ή και ολόκληρων εικόνων δεν δείχνει να παίζει σημαντικό ρόλο στην απόδοση του μοντέλου μας, ο συνδυασμός τους όμως δείχνει να έχει ενδιαφέρον μελέτης.
- Τα ιδιαίτερα χαρακτηριστικά κάθε κλάσης μπορούν να συμβάλουν στη βελτίωση ή και τη μείωση της ακρίβειας ενός μοντέλου. Οι υπολογιστές φαίνονται συχνά να υποπίπτουν σε ίδια λάθη με τους ανθρώπους και να εντοπίζουν τις ομοιότητες μεταξύ κλάσεων τις οποίες εντοπίζουν κι οι άνθρωποι.
- Μπορούμε να αντλήσουμε σημαντική πληροφορία από τα λάθη ενός ταξινομητή, ειδικά όσον αφορά τη συσχέτιση δύο κλάσεων.
- Πρέπει να ξεφεύγουμε από την πεπατημένη και να αντλούμε έμπνευση από άλλα κομμάτια της βαθιάς μάθησης κι όχι μόνο, για να επιτύχουμε σημαντικές βελτιώσεις στα μοντέλα μας.
- Πάντοτε υπάρχουν περιθώρια βελτίωσης της υπάρχουσας δουλειάς, ειδικά στο κομμάτι της βαθιάς μάθησης όπου ο χώρος αναζήτησης των βέλτιστων παραμέτρων είναι τεράστιος.
- Η επιστήμη και η τέχνη μπορούν να συνδυαστούν αρμονικά δίνοντας πολύ ενδιαφέροντα αποτελέσματα.

5.3 Μελλοντικές Κατευθύνσεις Επιστημονικής Μελέτης

Δεδομένων των καλών ποσοστών ακρίβειας των μοντέλων μας, υπάρχει η προοπτική να αποτελέσουν μια ισχυρή βάση για περαιτέρω ανάπτυξη μοντέλων για αναγνώριση τεχνοτροπίας. Εκτενέστερη μελέτη των υπερπαραμέτρων που χρησιμοποιήθηκαν καθώς και συλλογή ενός μεγαλύτερου και πιο αξιόπιστου συνόλου δεδομένων θα μπορούσαν να οδηγήσουν σε σημαντική αύξηση της απόδοσης του μοντέλου. Περικοπή των εικόνων κατά την εισαγωγή με τυχαία επιλογή του μέρους περικοπής σε κάθε εποχή είναι μια ιδέα η οποία θα μπορούσε να συμβάλει σημαντικά στη μείωση του overfitting και να δώσει επίσης καλύτερα αποτελέσματα. Φυσικά εφαρμογή κι επέκταση των μεθόδων που χρησιμοποιήθηκαν στο state of the art δίκτυο από την ομάδα του RMIT University οι οποίες όπως φαίνεται συμβάλουν σε μεγάλο βαθμό στην επιτυχία ενός μοντέλου για τη μελέτη έργων τέχνης. Τα παραπάνω αφορούν μελλοντική μελέτη στον άξονα στον οποίο κινηθήκαμε στα πλαίσια της διπλωματικής εργασίας. Ξεφεύγοντας από τα αυστηρά όρια της αναγνώρισης τεχνοτροπίας μεγάλο ενδιαφέρον παρουσιάζει η δημιουργία έργων τέχνης με χρήση συνελκτικών νευρωνικών δικτύων η οποία έχει δώσει εντυπωσιακά αποτελέσματα μέχρι στιγμής. Στις εργασίες που έχουν πραγματοποιηθεί τη συγκεκριμένη πρόκληση χρησιμοποιούνται επί των πλείστων CNNs εκπαιδευμένα στο σύνολο δεδομένων ImageNet. Θεωρούμε πως η χρήση ενός

CNN το οποίο έχει εκπαιδευτεί στην αναγνώριση τεχνοτροπίας, όπως τα δικά μας, μπορεί να δώσει ακόμη καλύτερα ή τουλάχιστον πιο ενδιαφέροντα αποτελέσματα. Τέλος ενδιαφέρουσα είναι η προσέγγιση της αναγνώρισης κινήματος κι όχι απλώς τεχνοτροπίας. Αυτό θα μπορούσε να επιτευχθεί με έναν συνδυασμό των δεδομένων που χρησιμοποιήσαμε για την αναγνώριση τεχνοτροπίας και μετα-δεδομένων σχετικά με το δημιουργό, την περίοδο, το μέρος ή ακόμη και το υλικό κάθε έργου τέχνης.

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.
- Ravneet Singh Arora and Ahmed M. Elgammal. Towards automated classification of fine-art painting style: A comparative study. *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, pages 3541–3544, 2012.
- Yaniv Bar, Noga Levy, and Lior Wolf. Classification of artistic styles using binarized features derived from a deep neural network. In *ECCV Workshops*, 2014.
- R. Bellman and Karreman Mathematics Research Collection. *Adaptive Control Processes: A Guided Tour*. Princeton Legacy Library. Princeton University Press, 1961. URL <https://books.google.gr/books?id=POAmAAAAMAAJ>.
- Eva Cetinic, Tomislav Lipic, and Sonja Grgic. Fine-tuning convolutional neural networks for fine art classification. *Expert Systems with Applications*, 114:107 – 118, 2018. ISSN 0957-4174. doi: <https://doi.org/10.1016/j.eswa.2018.07.026>. URL <http://www.sciencedirect.com/science/article/pii/S0957417418304421>.
- François Chollet et al. Keras. <https://github.com/fchollet/keras>, 2015.
- Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, and Trevor Darrell. Decaf: A deep convolutional activation feature for generic visual recognition, 2013.
- Adel El-Shahat. *Advanced Applications for Artificial Neural Networks*. 2018. ISBN 978-953-51-3781-8. URL <https://www.intechopen.com/books/advanced-applications-for-artificial-neural-networks>.

- Eric Campbell Fernie. *Art History and its Methods: A critical anthology*. 1995. ISBN 978-0-7148-2991-3.
- C. Florea, C. Toca, and F. Gieseke. Artistic movement recognition by boosted fusion of color structure and topographic description. In *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 569–577, March 2017. doi: 10.1109/WACV.2017.69.
- Corneliu Florea and Fabian Gieseke. Artistic movement recognition by consensus of boosted svm based experts. *Journal of Visual Communication and Image Representation*, 56:220 – 233, 2018. ISSN 1047-3203. doi: <https://doi.org/10.1016/j.jvcir.2018.09.015>. URL <http://www.sciencedirect.com/science/article/pii/S104732031830230X>.
- Corneliu Florea, Razvan Condorovici, Constantin Vertan, Raluca Boia, Laura Florea, and Ruxandra Vranceanu. Pandora: Description of a painting database for art movement recognition with baselines and perspectives. 02 2016.
- Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. A neural algorithm of artistic style, 2015.
- Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.
- B. Günsel, Sanem Sariel, and Oguz Icoğlu. Content-based access to art paintings. volume 2, pages II – 558, 10 2005. ISBN 0-7803-9134-9. doi: 10.1109/ICIP.2005.1530116.
- Simon S. Haykin. *Neural networks and learning machines*. Pearson Education, Upper Saddle River, NJ, third edition, 2009.
- Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015a.
- Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification, 2015b.
- Allen Huang and Raymond Wu. Deep learning for music, 2016.
- Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. Densely connected convolutional networks, 2016.
- Jia Li and J. Z. Wang. Studying digital imagery of ancient paintings by mixtures of stochastic models. *IEEE Transactions on Image Processing*, 13(3):340–353, March 2004. doi: 10.1109/TIP.2003.821349.
- Shuqiang Jiang, Qixiang Ye, and Wen Gao. An effective method to detect and categorize digitized traditional chinese paintings. *Pattern Recognition Letters*, 27:734–746, 05 2006. doi: 10.1016/j.patrec.2005.10.017.
- Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution, 2016.
- Sergey Karayev, Matthew Trentacoste, Helen Han, Aseem Agarwala, Trevor Darrell, Aaron Hertzmann, and Holger Winnemoeller. Recognizing image style, 2013.

- Fahad Shahbaz Khan, Shida Beigpour, Joost van de Weijer, and Michael Felsberg. Painting-91: a large scale database for computational painting categorization. *Machine Vision and Applications*, 25(6): 1385–1397, Aug 2014. ISSN 1432-1769. doi: 10.1007/s00138-014-0621-6. URL <https://doi.org/10.1007/s00138-014-0621-6>.
- Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton. Imagenet classification with deep convolutional neural networks. *Neural Information Processing Systems*, 25, 01 2012. doi: 10.1145/3065386.
- Adrian Lecoutre, Benjamin Négrevergne, and Florian Yger. Recognizing art style automatically in painting with deep learning. In *ACML*, 2017.
- Yuan-Pin Lin and T. Jung. Improving eeg-based emotion classification using conditional transfer learning. In *Front. Hum. Neurosci.*, 2017.
- Stephen Little. *Isms: Understanding Art*. Universe, 2004. ISBN 978-0789312099.
- John McCarthy. What is artificial intelligence? 01 2004.
- Kevin P. Murphy. *Machine Learning: A Probabilistic Perspective*. The MIT Press, 2012. ISBN 0262018020, 9780262018029.
- K. Peng and T. Chen. Cross-layer features in convolutional neural networks for generic classification tasks. In *2015 IEEE International Conference on Image Processing (ICIP)*, pages 3057–3061, Sep. 2015. doi: 10.1109/ICIP.2015.7351365.
- G. Polatkan, S. Jafarpour, A. Brasoveanu, S. Hughes, and I. Daubechies. Detection of forgery in paintings using supervised learning. In *2009 16th IEEE International Conference on Image Processing (ICIP)*, pages 2921–2924, Nov 2009. doi: 10.1109/ICIP.2009.5413338.
- C. S. Rodriguez, M. Lech, and E. Pirogova. Classification of style in fine-art paintings using transfer learning and weighted image patches. In *2018 12th International Conference on Signal Processing and Communication Systems (ICSPCS)*, pages 1–7, Dec 2018. doi: 10.1109/ICSPCS.2018.8631731.
- R. Sablatnig, P. Kammerer, and E. Zolda. Hierarchical classification of paintings using face- and brush stroke models. In *Proceedings. Fourteenth International Conference on Pattern Recognition (Cat. No.98EX170)*, volume 1, pages 172–174 vol.1, Aug 1998. doi: 10.1109/ICPR.1998.711107.
- Babak Saleh and Ahmed Elgammal. Large-scale classification of fine-art paintings: Learning the right metric on the right feature, 2015.
- A. L. Samuel. Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 3(3):210–229, July 1959. doi: 10.1147/rd.33.0210.
- C. Sandoval, E. Pirogova, and M. Lech. Two-stage deep learning approach to the classification of fine-art paintings. *IEEE Access*, 7:41770–41781, 2019. doi: 10.1109/ACCESS.2019.2907986.
- Lior Shamir, Tomasz Macura, Nikita Orlov, D. Mark Eckley, and Ilya G. Goldberg. Impressionism, expressionism, surrealism: Automated recognition of painters and schools of art. *ACM Trans.*

- Appl. Percept.*, 7(2):8:1–8:17, February 2010. ISSN 1544-3558. doi: 10.1145/1670671.1670672. URL <http://doi.acm.org/10.1145/1670671.1670672>.
- Lior Shamir, Jenny Nissel, and Ellen Winner. Distinguishing between abstract art by artists vs. children and animals: Comparison between human and machine perception. *ACM Trans. Appl. Percept.*, 13(3):17:1–17:17, May 2016. ISSN 1544-3558. doi: 10.1145/2912125. URL <http://doi.acm.org/10.1145/2912125>.
- B. Siddiquie, S. N. Vitaladevuni, and L. S. Davis. Combining multiple kernels for efficient image classification. In *2009 Workshop on Applications of Computer Vision (WACV)*, pages 1–8, Dec 2009. doi: 10.1109/WACV.2009.5403040.
- Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition, 2014.
- Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions, 2014.
- Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alex Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning, 2016.
- W. R. Tan, C. S. Chan, H. E. Aguirre, and K. Tanaka. Ceci n’est pas une pipe: A deep convolutional network for fine-art paintings classification. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 3703–3707, Sep. 2016. doi: 10.1109/ICIP.2016.7533051.
- Wei Ren Tan, Chee Seng Chan, Hernan Aguirre, and Kiyoshi Tanaka. Improved artgan for conditional synthesis of natural image and artwork. *IEEE Transactions on Image Processing*, PP: 1–1, 08 2018. doi: 10.1109/TIP.2018.2866698.
- Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. Wavenet: A generative model for raw audio, 2016.
- Nitin Viswanathan. Artist identification with convolutional neural networks. 2017.
- Yangyang Yu, Olivier Jin, and Daniel Hsu. Artistic movement recognition using deep cnns. 2017.
- Sheng-hua Zhong, Xingsheng Huang, and Zhijiao Xiao. Fine-art painting classification via two-channel dual path networks. *International Journal of Machine Learning and Cybernetics*, May 2019. ISSN 1868-808X. doi: 10.1007/s13042-019-00963-0. URL <https://doi.org/10.1007/s13042-019-00963-0>.
- Ασημίνα Μερτζάνη. Τεχνικές Μηχανικής Μάθησης σε Προβλήματα Δρομολόγησης και Ανάθεσης Πόρων. Master’s thesis, Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών, Εθνικό Μετσόβιο Πολυτεχνείο, 9 2019.
- Αθανάσιος Παπαδόπουλος. ΣΥΝΕΛΙΚΤΙΚΑ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ ΣΤΗΝ ΥΠΟΛΟΓΙΣΤΙΚΗ ΎΠΑΡΞΗ. Master’s thesis, 2016. URL <https://nemertes.lis.upatras.gr/jspui/bitstream/10889/9623/3/PapadopoulosAth%28phys%29.pdf>.