Εθνικο Μετσοβιο Πολυτεχνειο
Σχολη Ηλεκτρολογων Μηχανικων και Μηχανικων Υπολογιστων
Τομεας Τεχνολογιας, Πληροφορικης και Υπολογιστων

Συσχέτιση Φωτογραφιών και Έργων Τέχνης με Συνελικτικά Νευρωνικά Δίκτυα

Διπλωματικη Εργασια

της

Στελίνας Ταράση

Επιβλέπων: Στέφανος Κόλλιας
Καθηγητής Ε.Μ.Π

Συνεπιβλέπουσα: Παρασκευή Τζούβελη
Μέλος ΕΔΙΠ

Αθήνα, Μάρτιος 2022

Εθνικο Μετσοβιο Πολυτεχνειο
Σχολη Ηλεκτρολογων Μηχανικων και Μηχανικων Υπολογιστων
Τομεας Τεχνολογιας, Πληροφορικης και Υπολογιστων

Συσχέτιση Φωτογραφιών και Έργων Τέχνης με Συνελικτικά Νευρωνικά Δίκτυα

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

Στελίνας Ταράση

Επιβλέπων: Στέφανος Κόλλιας
Καθηγητής Ε.Μ.Π

Συνεπιβλέπουσα: Παρασκευή Τζούβελη
Μέλος ΕΔΙΠ

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 1η Απριλίου 2022.

(Υπογραφή)	*(Υπογραφή)*	*(Υπογραφή)*
.....	.....	.....
Στέφανος Κόλλιας	Ανδρέας-Γεώργιος Σταφυλοπάτης	Γεώργιος Στάμου
Καθηγητής Ε.Μ.Π	Καθηγητής Ε.Μ.Π	Αν.Καθηγητής Ε.Μ.Π

Αθήνα, Μάρτιος 2022

Εθνικο Μετσοβιο Πολυτεχνειο
Σχολη Ηλεκτρολογων Μηχανικων και Μηχανικων Υπολογιστων
Τομεας Τεχνολογιας, Πληροφορικης και Υπολογιστων

ΔΗΛΩΣΗ ΜΗ ΛΟΓΟΚΛΟΠΗΣ ΚΑΙ ΑΝΑΛΗΨΗΣ ΠΡΟΣΩΠΙΚΗΣ ΕΥΘΥΝΗΣ

Με πλήρη επίγνωση των συνεπειών του νόμου περί πνευματικών δικαιωμάτων, δηλώνω ενυπογράφως ότι είμαι αποκλειστικός συγγραφέας της παρούσας Πτυχιακής Εργασίας, για την ολοκλήρωση της οποίας κάθε βοήθεια είναι πλήρως αναγνωρισμένη και αναφέρεται λεπτομερώς στην εργασία αυτή. Έχω αναφέρει πλήρως και με σαφείς αναφορές, όλες τις πηγές χρήσης δεδομένων, απόψεων, θέσεων και προτάσεων, ιδεών και λεκτικών αναφορών, είτε κατά κυριολεξία είτε βάσει επιστημονικής παράφρασης. Αναλαμβάνω την προσωπική και ατομική ευθύνη ότι σε περίπτωση αποτυχίας στην υλοποίηση των ανωτέρω δηλωθέντων στοιχείων, είμαι υπόλογος έναντι λογοκλοπής, γεγονός που σημαίνει αποτυχία στην Πτυχιακή μου Εργασία και κατά συνέπεια αποτυχία απόκτησης του Τίτλου Σπουδών, πέραν των λοιπών συνεπειών του νόμου περί πνευματικών δικαιωμάτων. Δηλώνω, συνεπώς, ότι αυτή η Πτυχιακή Εργασία προετοιμάστηκε και ολοκληρώθηκε από εμένα προσωπικά και αποκλειστικά και ότι, αναλαμβάνω πλήρως όλες τις συνέπειες του νόμου στην περίπτωση κατά την οποία αποδειχθεί, διαχρονικά, ότι η εργασία αυτή ή τμήμα της δεν μου ανήκει διότι είναι προϊόν λογοκλοπής άλλης πνευματικής ιδιοκτησίας.

(Υπογραφή)

.....
Στελίνα Ταράση

20 Μαρτίου 2022

Περίληψη

Οι εικαστικές τέχνες καθρεφτίζουν τον τρόπο με τον οποίο οι άνθρωποι αντιλαμβάνονται τον κόσμο. Αποτελούν καίριο παράγοντα για την κατανόηση της κουλτούρας και της ιστορίας κάθε πολιτισμού. Στη σύγχρονη εποχή, η άμεση αποτύπωση της πραγματικότητας γίνεται μέσω του φωτορεπορτάζ. Παρατηρώντας φωτογραφίες του 20ου και 21ου αιώνα, μπορούμε να ανακαλύψουμε ομοιότητες με παλαιότερα έργα τέχνης, εντοπίζοντας ορισμένες κοινές αντιλήψεις των ανθρώπων για την πραγματικότητα.

Η συνεχής προσπάθεια των τελευταίων χρόνων για ψηφιοποίηση των έργων τέχνης σε συνδυασμό με τις ραγδαίες εξελίξεις στο χώρο της τεχνητής νοημοσύνης, μας προσφέρουν τη δυνατότητα να μελετήσουμε τις συνδέσεις που υπάρχουν ανάμεσα στα έργα τέχνης σε μικρό χρόνο χρησιμοποιώντας πολύ μεγάλα σύνολα δεδομένων.

Στη παρούσα διπλωματική εργασία, θελήσαμε να μελετήσουμε τον τρόπο που οι υπολογιστές αντιλαμβάνονται τα έργα τέχνης μέσα από ένα σύνολο άνω των 470.000 δεδομένων, καθώς και την αποδοτικότητα τους στο να εντοπίζουν συσχετίσεις ανάμεσα σε φωτογραφίες και πίνακες ζωγραφικής. Για τους λόγους αυτούς, μέσω της μεταφοράς γνώσης, χρησιμοποιήσαμε πέντε διαφορετικά μοντέλα συνελικτικών νευρωνικών δικτύων για την εξαγωγή χαρακτηριστικών από τα δεδομένα. Στη συνέχεια, εφαρμόσαμε μεθόδους για τη μείωση των διαστάσεων των εξαγόμενων χαρακτηριστικών και τη βελτίωση της απόδοσης του συστήματος. Τέλος, για κάθε φωτογραφία που θέσαμε σαν είσοδο, υλοποιήσαμε μη επιβλεπόμενη αναζήτηση κοντινότερων γειτόνων για τον εντοπισμό των πλησιέστερων, ως προς εκείνη, έργων τέχνης.

Λέξεις Κλειδιά

Συνελικτικά Νευρωνικά Δίκτυα, Φωτογραφία, Ζωγραφική, Μεταφορά Γνώσης, Αναζήτηση Οπτικών Συνδέσμων, Μη επιβλεπόμενη μάθηση

Abstract

Visual arts reflect the way people perceive the world. They are a key factor in understanding the culture and history of each civilization. In modern times, the direct capture of reality is achieved through photojournalism. By looking at photographs of the 20th and 21st centuries, we can discover similarities with older works of art, identifying some common perceptions that people have about reality.

The increased effort of recent years to digitize works of art in combination with the constant improvements in the field of artificial intelligence, offer us the opportunity to study the links that exist between works of art in a short time using very large data sets.

In this diploma thesis, we study how computers perceive works of art through a data set of over 470,000 data, as well as their efficiency in detecting visual links between photographs and paintings. For these reasons, through transfer learning, we use five different Convolutional Neural Networks to extract features from the data. After that, we apply methods to reduce the dimensions of the exported features and improve the performance of the system. Finally, for each photograph we set as an input, we carry out an unsupervised search of nearest neighbors to find the closest works of art to it.

Keywords

Convolutional Neural Networks, Photography, Painting, Visual link retrieval, Unsupervised learning

στους γονείς μου

Ευχαριστίες

Θα ήθελα αρχικά να ευχαριστήσω θερμά τον επιβλέποντα καθηγητή μου κ. Στέφανο Κόλλια για την εμπιστοσύνη που με έδειξε, δίνοντας μου την ευκαιρία να ασχοληθώ με ένα τόσο ενδιαφέρον θέμα.

Ακόμα, θα ήθελα να ευχαριστήσω μέσα από την καρδιά μου τη Διδάκτορα Παρασκευή Τζούβελη, η οποία με στήριξε σε κάθε βήμα της εργασίας προσφέροντας μου ουσιαστική βοήθεια οποτεδήποτε το χρειαζόμουν. Δε θα είχε βγει η διπλωματική αυτή εργασία χωρίς την πολύτιμη βοήθεια της.

Ένα μεγάλο ευχαριστώ, στο φωτογράφο Νίκο Παλαιολόγο που χωρίς να το ξέρει μου έδωσε την ιδέα της παρούσας διπλωματικής εργασίας. Όσο υπέροχος φωτογράφος είναι άλλο τόσο είναι και υπέροχος άνθρωπος.

Στους γονείς μου, Κούλα και Γιάννη, οφείλω πολλά παραπάνω από ένα ευχαριστώ, για την αμέτρητη αγάπη που μου έδωσαν και τη στήριξη τους σε κάθε ιδέα που είχα, όσο περίεργη και αν ήταν, όλα αυτά τα χρόνια. Νιώθω ο πιο τυχερός άνθρωπος που τους έχω δίπλα μου.

Τέλος, θα ήθελα να ευχαριστήσω όλους τους φίλους μου και κυρίως τον Ανδρέα, τη Δήμητρα, την Ευγενία, τον Ιωσήφ, την Ευδοκία, το Σπύρο και την Πέμη, που ήταν δίπλα μου σε κάθε όμορφη και κάθε δύσκολη στιγμή. Χωρίς εκείνους δε θα είχε νόημα οποιαδήποτε επιτυχία.

Αθήνα, Μάρτιος 2022

Στελίνα Ταράση

Περιεχόμενα

Κατάλογος Σχημάτων

Κατάλογος Πινάκων

Κεφάλαιο 1

Εισαγωγή

Όπως η ιστορία επαναλαμβάνεται, το ίδιο ισχύει και για την ιστορία της τέχνης γεγονός που μας φανερώνεται πλέον με το φωτορεπορτάζ. Βλέπουμε εικόνες από όλες τις πλευρές του πλανήτη να μοιάζουν με πίνακες ζωγραφικής που έχουν δημιουργηθεί εκατοντάδες χρόνια πριν τονίζοντας τη διαχρονικότητα των γεγονότων και τη σύνδεση των δύο αυτών τεχνών.

Στην ιστορία της τέχνης, η σύγκριση και η αναζήτηση ομοιοτήτων ανάμεσα σε έργα τέχνης έχει αποτελέσει πολλές φορές επίκεντρο ερευνών. Η διαδικασία αυτή, φυσικά, απαιτούσε πολύωρες προσπάθειες για τη συλλογή των απαιτούμενων δεδομένων από διάφορες βιβλιοθήκες. Οι μελετητές χρειαζόταν να χρησιμοποιούν τεράστια σύνολα από εικόνες που τους βοηθούσαν να κάνουν τις κατάλληλες κατηγοριοποιήσεις ώστε να καταλήξουν σε συλλογές που μοιράζονται κοινά χαρακτηριστικά, έχοντας στη διάθεση τους λιγοστά μεταδεδομένα (metadata) για να τους κατευθύνουν[1].

Χάρη στις τεχνολογικές εξελίξεις και τη δραματική μείωση του κόστους, τα τελευταία χρόνια υπήρξαν αυξανόμενες προσπάθειες για ψηφιοποίηση των έργων τέχνης. Πλέον μας δίνεται η δυνατότητα να έχουμε πρόσβαση σε βάσεις δεδομένων με εκατοντάδες χιλιάδες εικόνες από διαφορετικά ινστιτούτα, όπως το WikiArt και το Met collection. Το γεγονός αυτό, σε συνδυασμό με τις συνεχείς εξελίξεις στον τομέα της Όρασης Υπολογιστών και της Αναγνώρισης Προτύπων (Pattern Recognition), έχει ανοίξει το δρόμο για καινούργιες εφαρμογές προσφέροντας νέες δυνατότητες στην καλλιτεχνική κοινότητα. Ερευνητές, από το χώρο της επιστήμης υπολογιστών, έχουν επικεντρωθεί στην ανάπτυξη εργαλείων που βοηθούν στην ανάλυση και την βαθύτερη κατανόηση των εικαστικών τεχνών δίνοντας λύσεις σε προβλήματα που ως τώρα παρέμεναν δυσεπίλυτα.

Η κατανόηση του περιεχομένου ενός πίνακα ή μίας φωτογραφίας καθώς και των νοημάτων που αποδίδει, αποτελεί αποτέλεσμα της οπτικής αντίληψης του ανθρώπου. Στην πραγματικότητα, αυτό οφείλεται στην αναγνώριση σημαντικών στοιχείων, όπως ο συνδυασμός σχημάτων, η υφή, οι αποχρώσεις και τελικά η σύνθεση του έργου[15]. Στον τομέα της Όραση Υπολογιστών, η ανάπτυξη των Συνελικτικών Νευρωνικών Δικτύων (Convolutional Neural Networks, σε συντομογραφία CNN) έχει προσφέρει λύση στο πρόβλημα της εξαγωγής και κατανόησης σημαντικών χαρακτηριστικών από εικόνες, ξεκινώντας από την αναγνώριση χρωμάτων και υφών μέχρι την εύρεση σχημάτων και αντικειμένων.

Οι νέες μέθοδοι που έχουν αναπτυχθεί, εισάγοντας την Όραση Υπολογιστών στο χώρο της Τέχνης, έχουν ένα ευρύ φάσμα εφαρμογών που αναπτύσσεται συνεχώς. Γνωστές έρευνες αφο-

ρούν την αναγνώριση τεχνοτροπίας([16],[17]), την αυτόματη αναγνώριση του καλλιτέχνη([18],[19]) και την κατηγοριοποίηση με βάση το είδος[20]. Στο πλαίσιο της παρούσας εργασίας, θα επικεντρωθούμε στην Αναζήτηση Οπτικών Συνέδεων (Visual Link Retrieval). Οι μελέτες που έχουν πραγματοποιηθεί πάνω σε αυτό τον τομέα αφορούν την εύρεση ομοιοτήτων σε πίνακες ζωγραφικής (ή και γενικότερα σε έργα τέχνης όπως αγάλματα ή αρχιτεκτονικά δημιουργήματα). Εμείς θελήσαμε να εξελίξουμε τις ήδη υπάρχουσες μεθόδους, αναζητώντας patterns, δηλαδή τα χαρακτηριστικά των εικόνων, που υπάρχουν και συνδέουν τη τέχνη της ζωγραφικής με τη φωτογραφία και πιο συγκεκριμένα, το φωτορεπορτάζ.

1.1 Σκοπός Εργασίας

Αρχικά, σκοπός της εργασίας είναι η εφαρμογή μεθόδων της Όρασης Υπολογιστών στο χώρο της τέχνης, επικεντρωνόμενοι στο visual link retrieval. Μελετήσαμε πώς οι υπολογιστές διαβάζουν πίνακες ζωγραφικής και φωτογραφίες εξάγοντας χαρακτηριστικά και τελικά πώς ταιριάζουν αυτά τα δύο είδη. Ο τομέας της Όρασης Υπολογιστών αναπτύσσεται συνεχώς, προσφέροντας μας νέες δυνατότητες για την ανάλυση εικόνων που ξεφεύγουν από τους συμβατικούς κανόνες, όπως οι πίνακες αφηρημένης τέχνης. Με αυτό τον τρόπο, έχουμε στα χέρια μας νέα εργαλεία με τα οποία μπορούμε να κατανοήσουμε την τέχνη αλλά και να τη φέρουμε πιο κοντά στο ευρύ κοινό.

Πολλές φορές μπορεί να πούμε πως αυτή η φωτογραφία είναι σαν πίνακας, εντοπίζοντας ομοιότητες από εικόνες που έχουμε στη μνήμη μας. Η ανάπτυξη της Μηχανικής Μάθησης και πιο συγκεκριμένα των συνελικτικών νευρωνικών δικτύων, μας δίνει τη δυνατότητα να αναζητούμε αυτές τις ομοιότητες μέσα από τεράστια σύνολα εικόνων χωρίς να χρειάζεται η ανθρώπινη παρέμβαση.

Ο λόγος που επικεντρωθήκαμε στη σύνδεση του φωτορεπορτάζ με τη ζωγραφική είναι η πολύ ενδιαφέρουσα σχέση τους, που παραμένει ακόμα υπό μελέτη. Βλέπουμε το παρελθόν να επαναλαμβάνεται. Σκηνές που παλαιότερα αποτυπώνονταν με πίνακες ζωγραφικής πλέον ταξιδεύουν τον κόσμο μέσω του φωτορεπορτάζ. Παρατηρώντας τις συσχετίσεις που παρουσιάζονται, βλέπουμε πώς ορισμένα πρότυπα ομορφιάς παραμένουν αναλλοίωτα, τον τρόπο που οι καλλιτέχνες αποτυπώνουν διαχρονικά τα τοπία και πώς γεγονότα του παρελθόντος αναβιώνουν σήμερα.

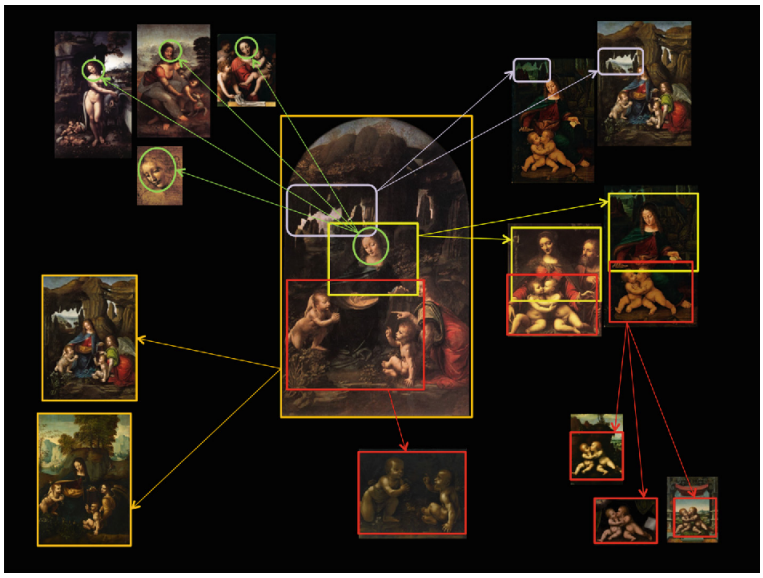
Η προτεινόμενη υλοποίηση έγινε με χρήση μη επιβλεπόμενης μάθησης. Με αυτόν τον τρόπο, δεν απαιτείται η ύπαρξη ετικετών (labels) για κάθε εικόνα, καθιστώντας τη διαδικασία πιο εύκολα υλοποιήσιμη. Επιπλέον, χρησιμοποιήσαμε πέντε διαφορετικά προεκπαιδευμένα συνελικτικά νευρωνικά δίκτυα, με στόχο την αξιολόγηση της απόδοσης τους, παρατηρώντας ποιό μας δίνει πιο ακριβή αποτελέσματα. Τέλος, βελτιώσαμε την ήδη υπάρχουσα μέθοδο, προσθέτοντας ένα ακόμη επίπεδο global average pooling στα προ-εκπαιδευμένα δίκτυα, καταλήγοντας πως τα μοντέλα παρέμεναν το ίδιο αποδοτικά απαιτώντας πολύ λιγότερο χρόνο και μνήμη.

1.2 Κίνητρο

Ο λόγος που προσωπικά επέλεξα το συγκεκριμένο θέμα, είναι αρχικά η μεγάλη αγάπη μου για τη φωτογραφία και ο θαυμασμός μου για το φωτορεπορτάζ. Παρακολουθώντας μαθήματα φωτορεπορτάζ, ήταν συχνό φαινόμενο ο καθηγητής μας να αναφέρει πίνακες ζωγραφικής συγκρίνοντας τους με σημερινές φωτογραφίες. Το γεγονός αυτό μου κίνησε την περιέργεια για τη σχέση που υπάρχει ανάμεσα σε αυτά τα δύο είδη τέχνης και πώς μπορεί η τεχνητή νοημοσύνη, η οποία με γοήτευσε βαθιά στα φοιτητικά μου χρόνια, να βοηθήσει στη μελέτη αυτή. Όπως ο ανθρώπινος εγκέφαλος μπορεί να αποθηκεύσει ένα αριθμό εικόνων, να τις συγκρίνει και να αναδείξει ομοιότητες, είναι φανερό πως η τεχνητή νοημοσύνη μπορεί να κάνει την ίδια διαδικασία με ένα πολύ μεγαλύτερο σύνολο δεδομένων. Συνδυάζοντας, λοιπόν, τους δύο αυτούς τομείς που με ενδιαφέρουν, ανοίχθηκε ένας νέος δρόμος για την παράλληλη εξερεύνηση και των δύο.

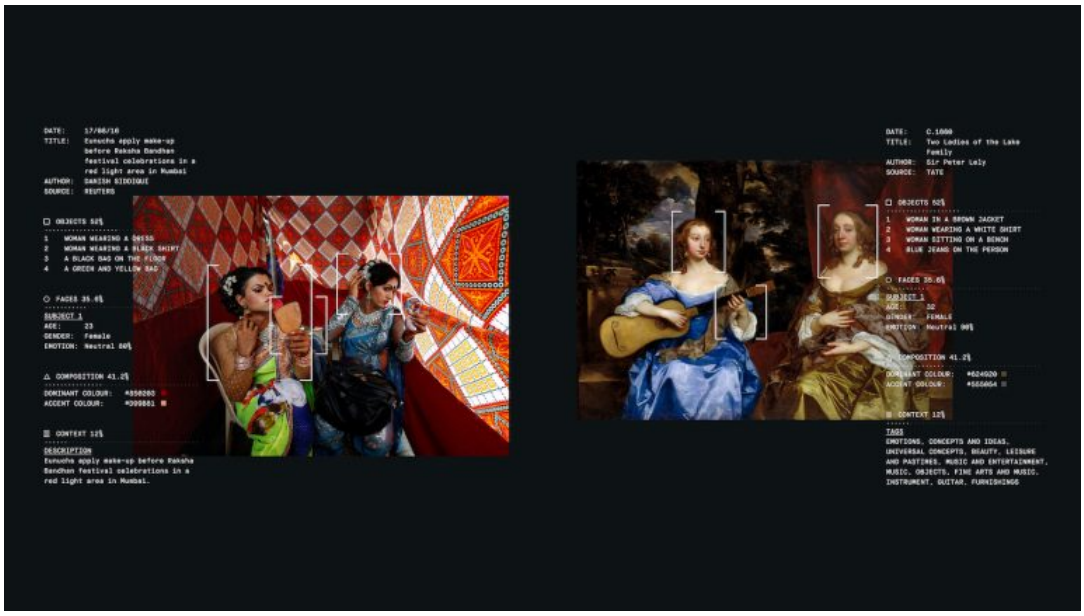
1.3 Προηγούμενες Μελέτες

Η υλοποίηση του visual link retrieval για την αναζήτηση patterns σε έργα εικαστικών τεχνών χρησιμοποιώντας βαθιά μάθηση, είναι μία μέθοδος που άρχισε να αναπτύσσεται τα τελευταία χρόνια. Το 2016, ο Benoit Seguin δημοσίευσε με την ομάδα του το paper “Visual Link Retrieval in a Database of Paintings”[1]. Στην έρευνα τους, συνέκριναν την κλασική μέθοδο bag-of-words με τη χρήση ενός προ-εκπαιδευμένου CNN μοντέλου, με στόχο την εύρεση πινάκων που σχετίζονται μεταξύ τους σύμφωνα με τις υποδείξεις ειδικού. Η μέθοδος αυτή έδειξε πως το CNN παρείχε τελικά καλύτερα αποτελέσματα. Παρόλα αυτά, η παραπάνω υλοποίηση βασιζόταν σε επιβλεπόμενη μάθηση που απαιτούσε την εισαγωγή ετικετών από τους ειδικούς, διαδικασία αρκετά χρονοβόρα που υπόκειται στην υποκειμενική κρίση των ανθρώπων. Αντίθετα, η χρήση μη επιβλεπόμενης μάθησης θα απέφευγε τα συγκεκριμένα προβλήματα. Στο paper “Discovering Visual Patterns in Art Collections with Spatially-consistent Feature Learning”[21], οι Shen et al. χρησιμοποίησαν ένα βαθύ νευρωνικό δίκτυο για να αναγνωρίσουν near-duplicate patterns σε ένα σύνολο έργων από το ζωγράφο Juan Brueghel. Η πρότασή τους βασιζόταν στην εισαγωγή ενός συγκεκριμένου χαρακτηριστικού στη διαδικασία, βελτιστοποιώντας το στη συγκεκριμένη συλλογή δεδομένων χρησιμοποιώντας self-supervised learning. Με το συγκεκριμένο τρόπο εκπαίδευσης, τα training labels αποκτούνταν από τα δεδομένα εισόδου. Το 2020, ο Gennaro Vessio και οι συνεργάτες του δημοσίευσαν το “Visual link retrieval and knowledge discovery in painting datasets”[15] αναζητώντας ομοιότητες μεταξύ πινάκων ζωγραφικής, καταλήγοντας στη δημιουργία ενός γράφου που παρουσιάζει πως κάθε ζωγράφος έχει επηρεαστεί από άλλους. Η μέθοδος που χρησιμοποίησε είναι και εκείνη στην οποία στηριχθήκαμε. Σε σχέση με τη συγκεκριμένη υλοποίηση, επιλέξαμε αρχικά να χρησιμοποιήσουμε ένα πολύ μεγαλύτερο σύνολο δεδομένων (dataset) από έργα τέχνης, για να έχουμε ακόμη πιο σαφή αποτελέσματα, στο οποίο προσθέσαμε ένα σύνολο φωτογραφιών. Επιπλέον, χρησιμοποιήσαμε πέντε διαφορετικά προ-εκπαιδευμένα μοντέλα στα οποία προσθέσαμε ένα νέο επίπεδο, το Global Average Pooling, εξάγωντας μικρότερο αριθμό χαρακτηριστικών και βελτιώνοντας έτσι τον χρόνο και την απαιτούμενη μνήμη, παίρνοντας πολύ ικανοποιητικά αποτελέσματα.



Σχήμα 1.1: *Παράδειγμα από το Visual link retrieval in a Database of paintings[1]*

Σχετικά με το θέμα που επιλέξαμε, δηλαδή την αναζήτηση συσχετίσεων όχι μόνο ανάμεσα σε πίνακες ζωγραφικής αλλά και με φωτογραφίες, βρήκαμε μία πολύ ενδιαφέρουσα εφαρμογή που πραγματοποιήθηκε στο πλαίσιο του IK Prize 2016, το Recognition. Ο διαγωνισμός καλούσε τους συμμετέχοντες να χρησιμοποιήσουν την τεχνητή νοημοσύνη για να εξερευνήσουν τη συλλογή British art του Tate Britain. Η νικήτρια ομάδα, που αποτελείται από τους Angelo Semeraro, Coralie Gourguechon και Monica Lanaro, υλοποίησε ένα σύστημα στο οποίο εισήγαγε 30.000 έργα τέχνης αντιστοιχώντας τα με εικόνες από το πρακτορείο Reuters. Η μέθοδος τους αποτελείται από 4 διαφορετικούς αλγόριθμους (Object recognition, Facial recognition,Composition recognition και Context recognition) για την ανάλυση των εικόνων. Τα αποτελέσματα τους είχαν μεγάλο ενδιαφέρον και το μουσείο χρησιμοποίησε την εφαρμογή τους τόσο σαν ιστοσελίδα που μπορεί το κοινό να πειραματιστεί με δικές του εικόνες, όσο και σαν installation στο χώρο του μουσείου.



Σχήμα 1.2: *Παράδειγμα από τα αποτελέσματα του Recognition.Η εικόνα ανήκει στο Tate Britain.*

Η επιτυχής αναγνώριση του περιεχομένου μίας εικόνας συνδέεται άμεσα με το κύριο πρόβλημα που προκύπτει στην ανάλυση εικόνων, την εύρεση υψηλού επιπέδου εννοιών (high-level concept detection). Στο πλαίσιο αυτό, έχει παρουσιαστεί το paper "Using Visual Context and Region Semantics for High-Level Concept Detection" [22], παρουσιάζοντας νέους τρόπους προσέγγυσης του προβλήματος.

Η ικανότητα των νευρωνικών δικτύων να αξιοποιούν χαρακτηριστικά που έχουν εξαγάγει από σύνολα δεδομένων έχει οδηγήσει στην ένταξη τους σε εφαρμογές από διαφορετικούς τομείς. Παραδείγματα είναι η χρήση τους για προβλέψεις σε Ιατρικές Εικόνες [23],[24], [25] και στην Επεξεργασία Σήματος[26]. Ακόμη, έχουν χρησιμοποιηθεί στην ανάλυση εικόνων και βίντεο για την αναγνώριση προσώπων [27], [28] αλλά στην προσπάθεια σύνθεσης ανθρώπινων εκφράσεων και συναισθημάτων [29].

1.4 Δομή Εργασίας

Στο Κεφάλαιο 2,παρουσιάζεται το θεωρητικό υπόβαθρο των απαραίτητων εννοιών, για την καλύτερη κατανόηση της εργασίας. Αρχικά, γίνεται αναφορά στη σχέση της φωτογραφίας με τη ζωγραφική ιστορικά, καθώς και στη γέννηση του φωτορεπορτάζ. Στη δεύτερη υποενότητα του κεφαλαίου αναλύουμε βασικές έννοιες του χώρου της Βαθιάς Μάθησης και των Νευρωνικών Δικτύων, δίνοντας έμφαση στα Συνελικτικά Νευρωνικά Δίκτυα. Ακόμη, περιγράφεται η αρχιτεκτονική Σύγχρονων Συνελικτικών Δικτύων συμπεριλαμβανομένων εκείνων που χρησιμοποιήσαμε για τη διεξαγωγή των πειραμάτων.

Στο κεφάλαιο 3, γίνεται αρχικά παρουσίαση των συνόλων δεδομένων που επιλέχθηκαν για τα πειράματα μας. Στη συνέχεια, περιγράφονται οι σχεδιαστικές επιλογές των προγράμματων και τα βήματα που ακολουθήσαμε για την υλοποίηση τους. Τέλος, γίνεται αναφορά στο περιβάλλον εκτέλεσης των πειραμάτων.

Στο 4ο κεφάλαιο, παρουσιάζονται τα αποτελέσματα των πειραμάτων. Πιο συγκεκριμένα, παρατίθενται τα αποτελέσματα που εξήγαγε κάθε πείραμα για διαφορετικές φωτογραφίες. Στη συνέχεια, ακολουθεί σχολιασμός επί των αποτελεσμάτων και αξιολόγηση της επίδοσης των πέντε διαφορετικών μοντέλων που χρησιμοποιήθηκαν.

Στο 5ο και τελευταίο κεφάλαιο, πραγματοποιείται μία σύνοψη της παρούσας διπλωματικής εργασίας και των συμπερασμάτων που εξήχθησαν. Τέλος, παρουσιάζονται μελλοντικές επεκτάσεις και πιθανοί τρόποι αξιοποίησης της προτεινόμενης μεθόδου.

Μέρος Ι

Θεωρητικό Μέρος

Κεφάλαιο 2

Θεωρητικό υπόβαθρο

.

2.1 Η Σχέση της Φωτογραφίας με τη Ζωγραφική

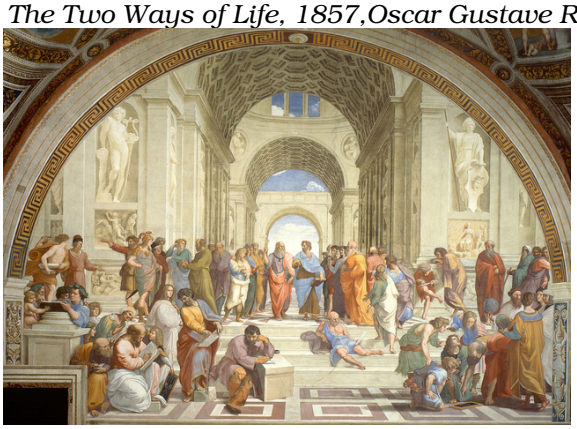
Ο 19ος αιώνας έφερε τη γέννηση μίας νέας μεθόδου αποτύπωσης της πραγματικότητας: τη Φωτογραφία. Το γεγονός αυτό πυροδότησε μία σειρά ερωτημάτων όπως ¨Είναι η Φωτογραφία τέχνη ή τεχνική¨ και ¨Ποια πρέπει να είναι η σχέση της με Ζωγραφική¨, τα οποία παραμένουν μέχρι και σήμερα θέματα συζήτησης και έρευνας. Για να κατανοήσουμε τη σχέση των δύο αυτών τεχνών, θα προχωρήσουμε σε μία ιστορική αναδρομή, παρουσιάζοντας πως επηρέασε η μία την άλλη στο πέρασμα του χρόνου.

Όταν τα χημικά φαινόμενα συνδυάστηκαν με τα φυσικά, τότε οδηγηθήκαμε στη γέννηση της φωτογραφίας. Ο πρώτος που πέτυχε την ένωση αυτή ήταν ο γάλλος Joseph Nicephore Niepce (1765-1833). Το 1816 χρησιμοποιώντας μία μηχανή με φακό κατόρθωσε να δημιουργήσει τα πρώτα φωτογραφικά αρνητικά πάνω σε χαρτί, το οποίο είχε επιστρωθεί με χλωριούχο άργυρο. Το 1829 ο Niepce, που στο μεταξύ είχε βελτιώσει πολύ τη μέθοδο του την οποία αποκαλούσε Ηλιογραφία, συνεργάστηκε με τον γάλλο ζωγράφο Louis Jacques Daguerre (1787-1851). Το 1833 μετά το θάνατο του Niepce, ο Daguerre θα συνεχίσει μόνος του τα πειράματα και θα καταλήξει στην ομώνυμη μέθοδο αποτύπωσης, τη δαγκεροτυπία.

Σχήμα 2.1: *L'Atelier de l'artiste. Πρώιμη δαγκεροτυπία του Daguerre κατασκευασμένη το 1837.*

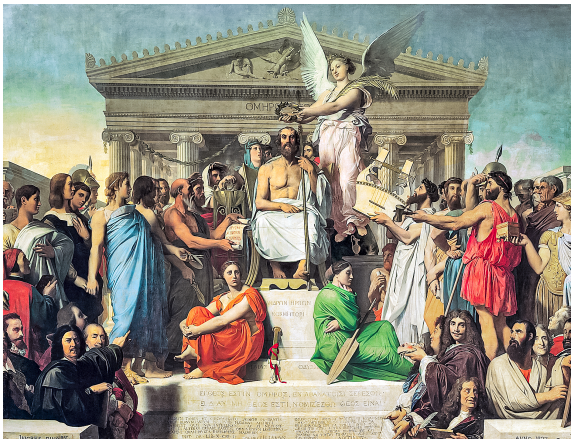
Η εφεύρεση της φωτογραφίας προήλθε από την αναγκαιότητα δημιουργίας και ενασχόλησης με μία νέα μορφή τέχνης που δημιουργήθηκε ύστερα από τη Βιομηχανική Επανάσταση, τις κοινωνικές και πολιτικές ανακατατάξεις στην Ευρώπη, στις αρχές του 19ου αιώνα καθώς και με τους προβληματισμούς στον τομέα της τέχνης. Η ανακάλυψη της φωτογραφίας όμως αποτέλεσε εκείνη τη χρονική στιγμή απειλή και θεωρήθηκε επικίνδυνη για την ιεραρχία της τέχνης. Η τέχνη του 19ου αιώνα ήταν μία πυραμίδα όπου όλοι προσπαθούσαν να πάρουν μία θέση. Την κορυφή κατείχαν διάσημοι ζωγράφοι, ακολουθούσαν εκείνοι που περιορίζονταν στη δημιουργία πορτραίτων για τους αστούς, ενώ στη βάση βρίσκονταν σχεδιαστές και εικονογράφοι βιβλίων. Το πνεύμα της εποχής υποστήριζε την ύπαρξη αυτής της ιεραρχίας και δεχόταν τον διαχωρισμό της τέχνης σε υψηλή και χαμηλή. Με την ανακάλυψη της φωτογραφίας ωστόσο η ικανότητα που ξεχώριζε τον καλλιτέχνη από τον κοινό άνθρωπο, έγινε κτήμα ενός μηχανικού μέσου. Αναμενόμενο ήταν λοιπόν, η φωτογραφία να προξενήσει την περιέργεια, τον φθόνο, τον πανικό ή τον ενθουσιασμό. Μέσα από τη φωτογραφία, οι άνθρωποι βλέπουν ένα θαυμαστό τρόπο να δημιουργούν καθετί που μπορούσε να κάνει ο ζωγράφος και να το κάνουν με μεγαλύτερη ακρίβεια, γρηγορότερα και φθηνότερα. Στις αρχές του 19ου αιώνα η ζωγραφική αλλάζει χαρακτήρα και εγκαταλείπει την ανατομική ακρίβεια στην απεικόνιση με στόχο μία αναπαράσταση πιο πνευματική από την ίδια την πραγματικότητα. Η φωτογραφία γεννιέται μέσα σε αυτή τη μεταβατική περίοδο της ζωγραφικής αντίληψης. Στην πρώτη δεκαετία της, που συμπίπτει με την περίοδο της δαγκεροτυπίας, η φωτογραφία και η νέα σχολή του ρομαντισμού στη ζωγραφική, εξελίσσονται παράλληλα. Το 1950 η φωτογραφία δεχόταν την κριτική των ζωγράφων, με την κατηγορία της ψυχρής αντιγραφής της φύσης. Το γεγονός αυτό ώθησε τους φωτογράφους να δημιουργήσουν τα περήφημα ταμπλώ-βιβαν, που μιμούνται γνωστούς ζωγραφικούς πίνακες με σκοπό την απομάκρυνση της φωτογραφίας από την αληθοφάνεια και την ακρίβεια, που τη θεωρούσαν μηχανική και απρόσωπη. Υποταγμένη στη ζωγραφική, η φωτογραφία είχε τις περισσότερες φορές σαν κύριο σκοπό να της μοιάσει. Ιδιαίτερα στη βικτωριανή Αγγλία, μία ομάδα φωτογράφων που ονομάστηκαν “Pictorialists”, δηλαδή εικονογράφοι, προσπαθούν με βουκολικά θέματα, ταμπλώ-βιβαν και αλληγορικές σκηνές να αντιγράψουν τη ζωγραφική. Κύριοι εκπρόσωποι της σχολής αυτής είναι ο Oscar Rejlander (1813-1875) και ο Henry Peach Robinson (1830-1901)[30].

Σχήμα 2.2: *The Two Ways of Life, 1857,Oscar Gustave Rejlander [2]*



Σχήμα 2.3: *Raphael, School of Athens, 1509-1511, fresco (Stanza della Segnatura, Palazzi Pontifici, Vatican [3]*

Σχήμα 2.4: *Apotheosis de Degas (After Ingres' L'apotheose d'Homere) Walter Barnes, Edgar Degas1885 [4]*



Σχήμα 2.5: *Jean Auguste Dominique Ingres, Apotheosis of Homer, 1827 [5]*

Το 1900 πρωτοεμφανίζεται ο φωτογράφος Weston δημιουργώντας φωτογραφίες σε στούντιο. Τα γυμνά που δημιουργεί καθιερώνουν το γυμνό, για πρώτη φορά, ως δημιουργικό θέμα στη φωτογραφία. Πρόκειται για ένα είδος Αφηρημένου Εξπρεσιονισμού στη φωτογραφία.

Το 1940 το πολεμικό φωτορεπορτάζ θέτει τις βάσεις της σύγχρονης αντίληψης για τη φωτογραφία-ντοκουμέντο. Πλέον, η φωτογραφία δεν είχε ως κύριο σκοπό εδώ τον καλλιτεχνικό πειραματισμό και τη δημιουργία. Χρησιμοποιεί όμως τις τεχνικές και εκφραστικές της δυνατότητες για να καταγράψει και να αναδείξει γεγονότα. Αισθητικό χαρακτηριστικό,των περισσότερων πολεμικών φωτογραφιών θα μπορούσε να θεωρηθεί η απευλευθερωμένη σύλληψη του ντοκουμέντου και της στιγμής, όπως πραγματοποιείται. Κατά τη συγκεκριμένη περίοδο, η ανάγκη της φωτογραφίας να αναδείξει την αλήθεια καθιερώνει μία σημαντική έως σήμερα φωτογραφική αντίληψη: τον απόλυτο ρεαλισμό [31].

2.2 Μηχανική Μάθηση

2.2.1 Ορισμός Μηχανικής Μάθησης

Η Μηχανική Μάθηση (machine learning) αποτελεί κλάδο της Τεχνητής Νοημοσύνης και έχει σημαντική επίδραση στην τεχνολογική ανάπτυξη του 21ου αιώνα. Στο πεδίο αυτό μελετούνται και αναπτύσσονται αλγόριθμοι ικανοί να πραγματοποιούν προβλέψεις ή να εξάγουν αποφάσεις βασιζόμενοι σε σύνολα δεδομένων, μιμούμενοι τον ανθρώπινο τρόπο σκέψης.

Τα μικρά παιδιά είναι ευκολότερο να μάθουν τη διαφορά ανάμεσα σε δύο φρούτα συνδέοντας τα με τις εικόνες τους παρά να δεχθούν και να αποθηκεύσουν ένα σύνολο ορισμών που τα προσδιορίζει και τα διαφοροποιεί. Με την ίδια λογική λειτουργεί η μηχανική μάθηση. Αντί να κωδικοποιεί τις πληροφορίες που δέχεται το σύστημα, αναζητά σημαντικές συνδέσεις και μοτίβα από παραδείγματα και παρατηρήσεις [32].

Τον όρο ¨Μηχανική Μάθηση" εισήγαγε πρώτος ο Arthur Samuel το 1959 στην εργασία του με τίτλο “Some Studies in Machine Learning Using the game of checkers”, παρουσιάζοντας το πρώτο πρόγραμμα μηχανικής μάθησης για το παιχνίδι ¨Ντάμα¨ [33]. Ο ίδιος όρισε ως μηχανική μάθηση το πεδίο μελέτης που δίνει στους υπολογιστές την ικανότητα να μαθαίνουν χωρίς να έχουν προγραμματιστεί ρητά. Ένας πιο σύγχρονος ορισμός και ευρέως διαδεδομένος δόθηκε από τον Tom Mitchell ως ¨Ένα πρόγραμμα υπολογιστή λέγεται ότι μαθαίνει από εμπειρία Ε ως προς μια κλάση εργασιών Τ και ένα μέτρο επίδοσης Π, αν η επίδοσή του σε εργασίες της κλάσης Τ, όπως αποτιμάται από το μέτρο Ρ, βελτιώνεται με την εμπειρία Ε» [34].

2.2.2 Βασικά Είδη Μηχανικής Μάθησης

Τα τρία είδη Μηχανικής Μάθησης είναι η Επιβλεπόμενη ή Επιτηρούμενη (Supervised Learning), η Μη Επιβλεπόμενη (Unsupervised Learning) και η Ενισχυτική Μάθηση (Reinforced Learning).

1. Επιβλεπόμενη Μάθηση

Στην Επιβλεπόμενη Μάθηση ο αλγόριθμος που χρησιμοποιείται εκπαιδεύεται με παραδείγματα. Τα δεδομένα που χρησιμοποιούνται ονομάζονται δεδομένα εκπαίδευσης και αποτελούνται από τις εισόδους και τις αντίστοιχες εξόδους. Κατά την εκπαίδευση ο αλγόριθμος αναλύει τα δεδομένα αναζητώντας μοτίβα που συνδέουν τις εισόδους με τις επιθυμητές εξόδους. Στόχος είναι, μετά την εκπαίδευση ο αλγόριθμος να μπορεί να αντιστοιχίσει άγνωστα δεδομένα ως εισόδους με τις κατάλληλες εξόδους, βασιζόμενος στις πληροφορίες που έχει συλλέξει. Η μέθοδος αυτή χρησιμοποιείτα κυρίως για προβλήματα ταξινόμησης και πρόγνωσης. Για παράδειγμα θα χρησιμοποιήσουμε ένα πολύ γνωστό σύνολο δεδομένων, το Iris data set από το machine-learning repository του πανεπιστημίου της Καλιφόρνια [35]. Στο συγκεκριμένο σύνολο δεδομένων περιέχονται 150 φυτά, όπου για κάθε ένα δίνονται πληροφορίες για το μέγεθος και το πλάτος των διαφόρων φύλλων του. Κάθε είσοδος του συνόλου περιλαμβάνει και μία ετικέτα (label) που επισημαίνει σε ποιά κατηγορία ανήκει το συγκεκριμένο φυτό (Iris Setosa, Iris Versicolor, Iris Virginica). Με αυτό τον τρόπο, το δίκτυο μαθαίνει να προσαρμόζει τα βάρη και τα thresholds του ώστε να μπορεί να αναγνωρίζει σε ποιά κατηγορία ανήκει κάθε φυτό. Η εκπαίδευση γίνεται με το training data set και η επαλήθευση με το test data set [32].

2. Μη Επιβλεπόμενη Μάθηση

Στη Μη Επιβλεπόμενη Μάθηση τα δεδομένα που χρησιμοποιούνται στην εκπαίδευση δεν έχουν επιθυμητή έξοδο. Για το λόγο αυτό, ο αλγόριθμος λειτουργεί χωρίς καθοδήγηση. Στόχος είναι η κατάλληλη ομαδοποίηση των δεδομένων βάση ομοιοτήτων, μοτίβων και διαφορών. Χρησιμοποείται κυρίως για προβήματα Ομαδοποίησης (Clustering) και Συσχέτισης (Association Rule). Η ομαδοποίηση αποτελεί τη βασική μέθοδο αξιοποίησης της μη επιβλεπόμενης μάθησης. Όπως κάθε άλλο πρόβλημα αυτού του είδους, προσπαθεί να βρει τη δομή των δεδομένων εισόδου ώστε να ομαδοποιήσει εκείνα με κοινά χαρακτηριστικά, δημιουργώντας clusters. Η μέθοδος συσχετίσεων επικεντρώνεται από την άλλη στην εύρεση συσχετισμών (σχέσεις, εξαρτήσεις) σε ένα μεγάλο σύνολο δεδομένων [34].

3. Ενισχυτική Μάθηση

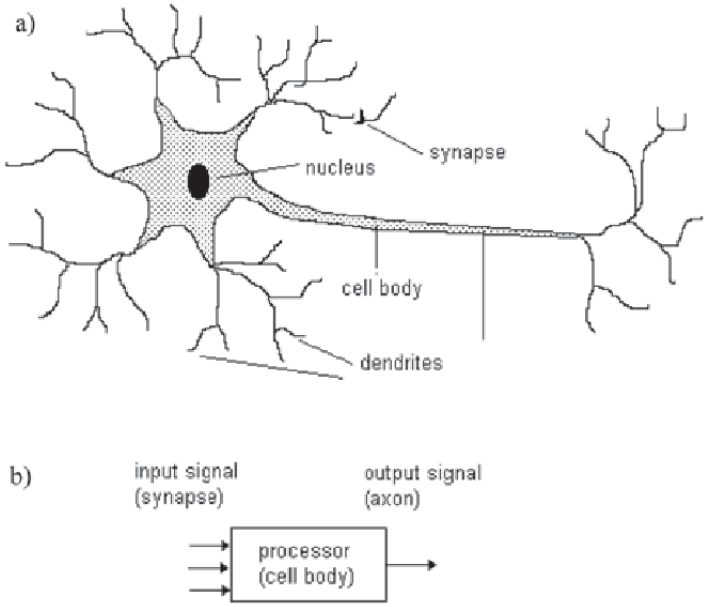
Στην τρίτη μέθοδο εκπαίδευσης, ο αλγόριθμος μαθαίνει μέσα από αλληλεπίδραση με το περιβάλλον. Ο πράκτορας λαμβάνει μία αναπαράσταση της τρέχουσας κατάστασης του περιβάλλοντος και ενεργεί σύμφωνα με μία πολιτική. Το περιβάλλον του παρέχει αριθμητικές ανταμοιβές ανταποκρινόμενο στις ενέργειες του. Με αυτόν τον τρόπο μαθαίνει ποιά είναι η επιθυμητή συμπεριφορά ώστε να πετύχει τη μέγιστη ανταμοιβή.

2.2.3 Νευρωνικά Δίκτυα

Ο όρος νευρωνικά δίκτυα ιστορικά αναφέρεται σε δίκτυα νευρώνων στον ανθρώπινο εγκέφαλο. Διαφορετικές περιοχές του εγκεφάλου πραγματοποιούν και διαφορετικές διεργασίες. Ο εγκεφαλικός φλοιός αποτελεί το εξωτερικό μέρος του ανθρώπινου εγκεφάλου και είναι ένα από τις πιο μεγάλα και ανεπτυγμένα τμήματα, έχοντας κεντρικό ρόλο σε όλες τις

ανώτερες εγκεφαλικές λειτουργίες όπως η μνήμη, η προσοχή, η αντίληψη, η σκέψη, η γλώσσα και η συνείδηση . Έχει πάχος 2-5mm και σχηματίζει μία πολυεπίπεδη δομή στην οποία υπάρχει μεγάλος αριθμός νευρικών κυττάρων, οι νευρώνες. Υπολογίζεται πως ο ανθρώπινος εγκεφαλικός φλοιός έχει 10^{10} νευρώνες περίπου [36].

Κάθε νευρώνας έχει τη δυνατότητα μετάδοσης ενός ηλεκτροχημικού σήματος. Αποτελείται από μία διακλαδωτική διάρθρωση εισροών (δενδρίτες), το κυτταρικό σώμα και μία διακλαδωτική δομή εκροών (τον άξονα). Οι άξονες του ενός κυττάρου συνδέονται με τους δενδρίτες του άλλου μέσω μίας σύναψης. Καθώς ένας άξονας ενεργοποιείται, πυροδοτείται ένα ηλεκτροχημικό σήμα κατά μήκος του. Κάθε σύναψη περιέχει νευροδιαβιβαστές χημικών ουσιών για τη μετάδοση του μηνύματος και η ισχύς του σήματος που λαμβάνεται από ένα νευρώνα εξαρτάται από την αποτελεσματικότητα των συνάψεων αυτών.



Σχήμα 2.6: *Δομή ενός νευρικού κυττάρου [6]*

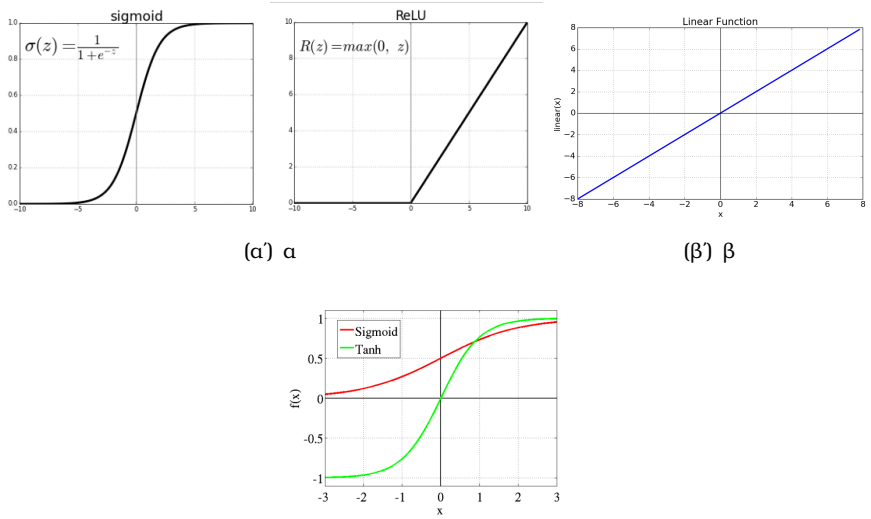
2.2.4 Τεχνητά Νευρωνικά Δίκτυα

Ο όρος τεχνητά νευρωνικά δίκτυα (ANNs) παρουσιάστηκε πρώτη φορά το 1943 από τους Warren McCullough και Walter Pitss [37], ερευνητές στο πανεπιστήμιο του Σικάγο και έγιναν ευρέως γνωστά στη δεκαετία του 80 με τη δημιουργία αλγορίθμων που μπορούν να μαθαίνουν βασιζόμενοι σε δεδομένα εκπαίδευσης. Λόγω της έλλειψης μεγάλων dataset και υπολογιστικής ισχύς έμειναν για πολλά χρόνια στην αφάνεια μέχρι και τις αρχές του 2000. Πλέον με την ύπαρξη τεράστιων συνόλων δεδομένων, όπως για παράδειγμα το ImageNet [38], και την ανάπτυξη συστημάτων παράλληλης επεξεργασίας, όπως GPUs, έχουν καταφέρει να δώσουν λύση σε πολύ δύσκολες διεργασίες όπως image classification και natural language processing.

Τα τεχνητά νευρωνικά δίκτυα συγκεντρώνουν το ενδιαφέρον της επιστημονικής κοινότητας λόγω της ευέλικτης δομής τους που τους επιτρέπει να χρησιμοποιούνται σε πολλαπλές εφαρμογές και στους τρεις τύπους της μηχανικής μάθησης. Εμπνευσμένα από τις βασικές αρχές των βιολογικών συστημάτων, αποτελούνται από χιλιάδες ή ακόμα και εκατομμύρια απλές, ισχυρά συνδεδεμένες μονάδες επεξεργασίας που ονομάζονται νευρώνες. Τα περισσότε-

ρες νευρωνικά δίκτυα είναι οργανωμένα σε επίπεδα (layers) και ορίζονται ως “feed-forward”, δηλαδή τα δεδομένα κινούνται προς μία σταθερή κατεύθυνση. Ένας νευρώνας μπορεί να είναι συνδεδεμένος με πολλούς άλλους από το προηγούμενο επίπεδο από το οποίο δέχεται δεδομένα καθώς και με πολλούς στο επόμενο όπου στέλνει τα αποτελέσματα που προκύπτουν ύστερα από την επεξεργασία. Τα βασικά επίπεδα νευρώνων είναι το επίπεδο εισόδου (input layer), τα κρυφά επίπεδα (hidden layers) και το επίπεδο εξόδου (output layer).

Όπως οι συνάψεις στον εγκέφαλο, κάθε σύνδεση μεταξύ των νευρώνων μεταφέρει σήματα τα οποία καθορίζονται από βάρη που μεταβάλλονται συνεχώς κατά τη διάρκεια της εκπαίδευσης. Κάθε φορά που το δίκτυο ενεργοποιείται, ο νευρώνας δέχεται δεδομένα από κάθε σύνδεση και τα πολλαπλασιάζει με το αντίστοιχο βάρος. Ύστερα από τους κατάλληλους υπολογισμούς εξάγει το τελικό αποτέλεσμα. Αν αυτό υπερβαίνει ένα κατώφλι (threshold), ο νευρώνας περνά τα δεδομένα στο επόμενο επίπεδο. Η συνάρτηση που καθορίζει την έξοδο του νευρώνα ονομάζεται συνάρτηση ενεργοποίησης και στις περισσότερες περιπτώσεις (εκτός του τελικού επιπέδου) είναι 0 (OFF) ή 1 (ON). Οι πιο γνωστές συναρτήσεις ενεργοποίησης είναι η σιγμοειδής (sigmoid), η ανορθωμένη γραμμική μονάδα (ReLU), η γραμμική (linear) και η υπερβολική εφαπτομένη (Tanh).



(α΄) α (β΄) β

(γ΄) γ

Σχήμα 2.7: *(α) Sigmoid (αριστερά), ReLU (δεξιά) (β) Linear (γ) Tanh [7]*

Στα αρχικά νευρωνικά δίκτυα από τους McCullough και Pitts υπήρχαν βάρη και thresholds αλλά δεν είχε υλοποιηθεί κάποιος μηχανισμός εκπαίδευσης. Αυτό που είχαν ως στόχο να δείξουν ήταν ότι ένα νευρωνικό δίκτυο μπορούσε να υλοποιήσει οποιαδήποτε συνάρτηση όπως ένας υπολογιστής, και άρα ο ανθρώπινος εγκέφαλος μπορεί να θεωρηθεί μία υπολογιστική μηχανή.

Το πρώτο εκπαιδευόμενο νευρωνικό δίκτυο, το Perceptron, παρουσιάστηκε το 1957 από τον ψυχολόγο F. Rosenblatt από το πανεπιστήμιο του Cornell [39]. Ο σχεδιασμός του ήταν όπως τα σημερινά δίκτυα με τη διαφορά ότι είχε μόνο ένα επίπεδο με μεταβαλλόμενα βάρη και thresholds, και λειτουργούσε ως γραμμικός ταξινομητής.

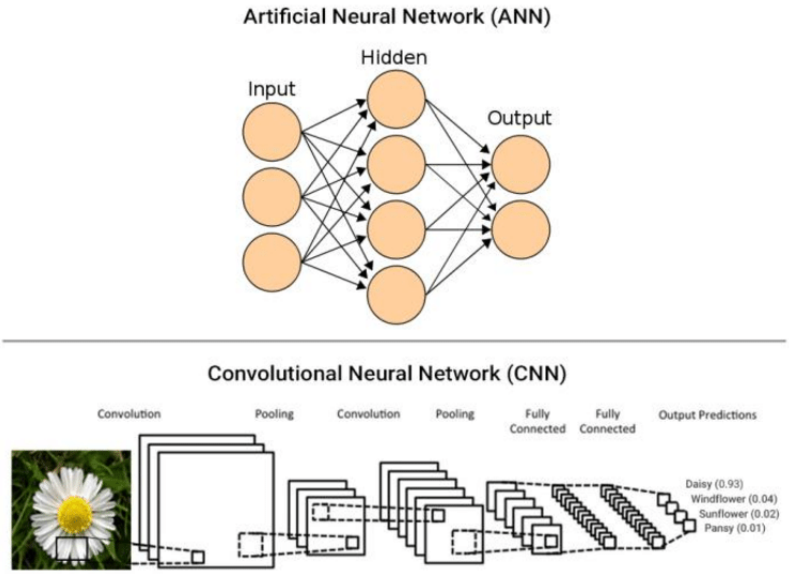
2.2.5 Συνελικτικά Νευρωνικά Δίκτυα

Τα Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks) αποτελούν το πιο διαδεδομένο είδος τεχνητών νευρωνικών δικτύων στον τομέα της Όρασης Υπολογιστών (Computer Vision) [40]. Είναι feedforward δίκτυα ικανά να εξάγουν χαρακτηριστικά από ένα σύνολο δεδομένων.

Η διαφορά τους με τα απλά τεχνητά νευρωνικά δίκτυα (ANNs) έγκειται σε τρεις παραμέτρους. Αρχικά, κάθε νευρώνας δε συνδέεται με όλους εκείνους από το προηγούμενο επίπεδο αλλά με ένα μικρό μέρος αυτών, γεγονός που μειώνει τον αριθμό των παραμέτρων και αυξάνει την ταχύτητα υπολογισμών. Επιπλέον παρέχεται η δυνατότητα σε μία ομάδα συνδέσεων να μοιράζεται τα ίδια βάρη, μειώνοντας περαιτέρω τις παραμέτρους. Τέλος, βασίζεται στη μείωση διαστάσεων των εικόνων με δειγματοληψία διατηρώντας παράλληλα τις σημαντικές πληροφορίες. Ο όρος ¨συνελικτικά¨ προέρχεται από τη χρήση της μαθηματικής πράξης της συνέλιξης που πραγματοποιείται ανάμεσα σε ένα δισδιάστατο πίνακα από βάρη που ονομάζεται πυρήνας (kernel) ή φίλτρο (filter) και την είσοδο ως διάνυσμα. Με συνάρτηση εισόδου x και συνάρτηση βαρών w, τότε η συνέλιξη $x*w$ ορίζεται ως:

$$s(t) = x(t) * w(t) = \int x(h)w(t-h)\,dh$$

Η ανάγκη δημιουργίας τους βασίστηκε στην αδυναμία των απλών νευρωνικών δικτύων να επεξεργαστούν μεγάλες εικόνες. Πιο συγκεκριμένα, έχοντας για παράδειγμα ως είσοδο εικόνες μεγέθους 32x32x3 ένας νευρώνας πλήρως συνδεδεμένος στο πρώτο κρυφό επίπεδο αντιστοιχεί σε 3072 βάρη. Βλέπουμε λοιπόν πως για τη δημιουργία ενός δικτύου πολλών επιπέδων θα χρειαζόμασταν έναν τεράστιο αριθμό παραμέτρων κάνοντας τη διαδικασία της εκπαίδευσης ιδιαίτερα δύσκολη και χρονοβόρα. Αντίθετα, τα συνελικτικά δίκτυα εκμεταλλεύονται τη χωρική συσχέτιση των δεδομένων εισόδου υλοποιώντας συνέλιξη μεταξύ του πυρήνα και μίας συστάδας από pixels. Καταφέρνουμε λοιπόν να έχουμε πολύ καλύτερα και γρηγορότερα αποτελέσματα με πολύ μικρότερο αριθμό παραμέτρων.



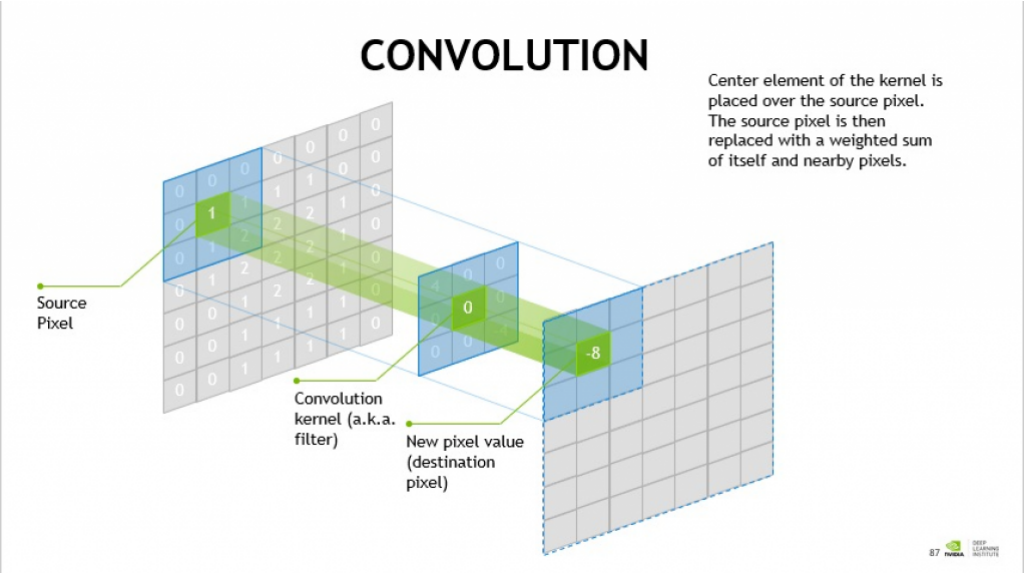
Σχήμα 2.8: *Σύγκριση αρχιτεκτονικής ενός ANN και ενός CNN [8]*

2.2.6 Αρχιτεκτονική Συνελικτικών Νευρωνικών Δικτύων

Η αρχιτεκτονική των συνεκτικών νευρωνικών δικτύων αποτελείται από τρεις τύπους επιπέδων: τα συνελικτικά επίπεδα, τα επίπεδα δειγματοληψίας και τα πυκνά ή πλήρως συνδεδεμένα επίπεδα.

Συνελικτικό επίπεδο (Convolutional Layer)

Το συνελικτικό επίπεδο είναι το κύριο στοιχείο των ομώνυμων δικτύων και πάντα αποτελεί τουλάχιστον το πρώτο επίπεδο της αρχιτεκτονικής. Η λειτουργία τους βασίζεται στη χρήση των πυρήνων. Οι πυρήνες-φίλτρα είναι συνήθως μικροί σε διαστάσεις αλλά εφαρμόζονται σε ολόκληρη την εικόνα διαδοχικά. Κατά τη διάρκεια της εκπαίδευσης πραγματοποιείται συνέλιξη κάθε τμήματος της εικόνας εισόδου με το φίλτρο παράγοντας έναν χάρτη ενεργοποίησης ή αλλιώς χάρτη χαρακτηριστικών (activation map ή feature map) δύο διαστάσεων. Με αυτό τον τρόπο το δίκτυο μαθαίνει ποιά φίλτρα ενεργοποιούνται όταν εντοπίζουν συγκεκριμένα χαρακτηριστικά σε ένα δοσμένο σημείο της εισόδου. Σε κάθε επίπεδο υπάρχουν πολλά φίλτρα τα οποία έχουν ένα αντίστοιχο χάρτη ενεργοποίησης. Κάθε τέτοιος χάρτης "στοιβάζεται" με τους υπόλοιπους παράγοντας το τελικό αποτέλεσμα του συνελικτικού επιπέδου. Με τη δόμηση ενός μοντέλου από συνεχόμενα συνελικτικά επίπεδα επιτρέπεται στα αρχικά επίπεδα να μαθαίνουν βασικά χαρακτηριστικά της εικόνας όπως ακμές και όσο προχωρούν εις βάθος να εντοπίζουν σχηματα και τελικά συγκεκριμένα αντικείμενα.



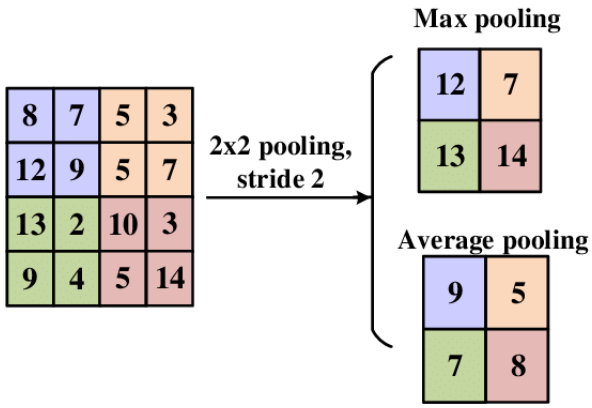
Σχήμα 2.9: *Παράδειγμα συνέλιξης της εισόδου με ένα πυρήνα 3x3 [9]*

Όπως αναφέραμε παραπάνω, τα συνελικτικά δίκτυα υπερτερούν έναντι των απλών δικτύων λόγω της ιδιότητας του κάθε νευρώνα στο συνελικτικό επίπεδο να συνδέεται μόνο με μία μικρή περιοχή της εισόδου. Επιπλέον μπορούν να μειώσουν σημαντικά την πολυπλοκότητα του μοντέλου βελτιστοποιώντας τις εξόδους τους. Αυτό επιτυγχάνεται με τρεις υπερπαραμέτρους : το depth, το stride και θέτοντας zero-padding τα οποία καθορίζονται από τον δημιουργό [41].

Επίπεδο Δειγματοληψίας (Pooling Layer)

Το επίπεδο δειγματοληψία συνήθως τοποθετείται μετά τα συνελικτικά επίπεδα και έχει ως στόχο της σταδιακή μείωση των διαστάσεων των δεδομένων και συνεπώς τη μείωση των παραμέτρων και της πολυπλοκότητας του μοντέλου.

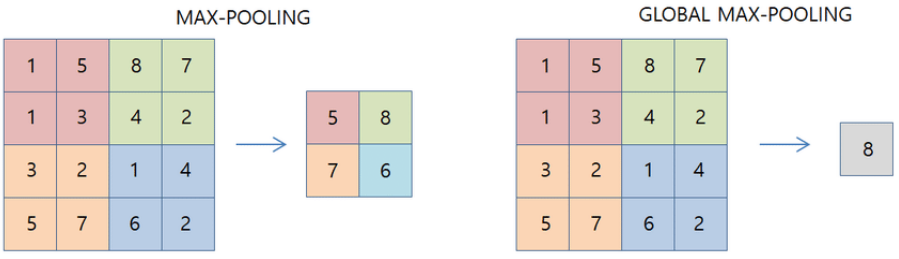
Ο τρόπος λειτουργίας του βασίζεται στην εφαρμογή μίας διαδικασίας σε κάθε χάρτη ενεργοποίησης, όπως θα λειτουργούσε ένα φίλτρο. Πιο γνωστές μέθοδοι είναι εκείνες της μέγιστης τιμής υλοποιώντας το max-pooling ή μέσης τιμής με average-pooling. Συνήθως τα επίπεδα αυτά έχουν πυρήνες διαστάσεων 2x2 με stride 2 (μέγεθος μετατόπισης παραθύρου στην είσοδο) ή 3x3 και stride 2 πραγματοποιώντας επικάλυψη παραθύρων (overlapping pooling). Λόγω της αρχιτεκτονικής του επιπέδου δειγματοληψίας, ένας πυρήνας μεγαλύτερος του 3x3 θα προκαλούσε μείωση της αποδοτικότητας του μοντέλου.



Σχήμα 2.10: *Παράδειγμα χρήσης max-pooling και average-pooling [10]*

Ένα ακόμα πλεονέκτημα που προσφέρει η προσθήκη των συγκεκριμένων επιπέδων είναι ότι το μοντέλο γίνεται πιο σταθερό όταν συμβαίνουν μικρές αλλαγές στις αρχικές εικόνες, αφού οι περισσότεροι χάρτες ενεργοποίησης μετά τη δειγματοληψία δε θα μεταβάλλονται όταν υπάρχουν μικρές διαφοροποιήσεις στις εισόδους [42].

Ένα ακόμη είδος pooling layer είναι εκείνο του Global Average Pooling ή Global Max Pooling Layer. Η διαφορά του με τα προηγούμενα που αναφέρονται είνα πως εδώ το pool size τείθεται ίσο με το μέγεθος της εισόδου. Μπορεί να χρησιμοποιηθεί σε διάφορες περιπτώσεις, όπως στη μείωση των διαστάσεων του εξαγόμενου, από προηγούμενα συνελικτικά επίπεδα, feature map, αντικαθιστώντας ένα αναγκαίο Flattening ή και σε ορισμένες περιπτώσεις τα Dense layers.



Σχήμα 2.11: *Η διαφορά ανάμεσα σε max-pooling layer και global-max-pooling layer [6]*

Πλήρως Συνδεδεμένο Επίπεδο (Fully Connected Layer)

Η τρίτη βασική κατηγορία επιπέδων στα συνελικτικά δίκτυα είναι τα πυκνά ή πλήρως συνδεδεμένα επίπεδα. Στα επίπεδα αυτά κάθε νευρώνας είναι συνδεδεμένος με τους όλους τους νευρώνες από το προηγούμενο επίπεδο, παρόμοια με τη λειτουργία των απλών νευρωνικών δικτύων. Συνήθως αποτελεί το τελευταίο επίπεδο του μοντέλου και δίνει σαν έξοδο τη πιθανότητα που αντιστοιχεί σε κάθε κλάση.

2.2.7 Σύγχρονες Αρχιτεκτονικές Συνελικτικών Νευρωνικών Δικτύων

Τα τελευταία χρόνια οι αρχιτεκτονικές των συνελικτικών δικτύων βελτιώνονται συνεχώς πετυχαίνοντας όλο και μεγαλύτερη ακρίβεια. Η πρώτη αποτελεσματική αρχιτεκτονική ήταν το LeNet που παρουσιάστηκε από τον Yann Lecun το 1990. Τα μετέπειτα χρόνια, οι αρχιτεκτονικές των CNN εξελίσσονταν συνεχώς δημιουργώντας νέα μοντέλα. Παρακάτω παρουσιάζονται οι πλέον διαδεδομένες αρχιτεκτονικές στον τομέα της Όρασης Υπολογιστών.

AlexNet

Το 2012 πραγματοποιήθηκε ο διαγωνισμός ταξινόμησης εικόνων στο σύνολο δεδομένων ImageNet [38] με νικητές τους Alex Krizhevsky, Ilya Sutskever και Geoff Hinton, δημιουργούς του AlexNet, επιτυγχάνοντας εξαιρετικά αποτελέσματα. Το μοντέλο είχε εκπαιδευτεί χρησιμοποιώντας GPUs, διευκολύνοντας το να τρέξει με μεγαλύτερες ταχύτητες, κερδίζοντας το ενδιαφέρον της επιστημονικής κοινότητας. Αποτελείται από συνολικά 8 επίπεδα, 5 συνελικτικά και 3 πλήρως συνδεδεμένα. Σημαντικό στοιχείο για την υψηλή του επίδοση ήταν η χρήση μονάδων ReLU ως συναρτήσεις ενεργοποίησης έναντι της tanh μειώνοντας σημαντικά τον χρόνο εκπαίδευσης. Επιπλέον, παρά την παραδοσιακή μέθοδο ως τότε, όπου οι έξοδοι του pooling layer δεν επικαλύπτονταν, οι δημιουργοί παρατήρησαν μείωση του ποσοστού λάθους κατά 0.5% χρησιμοποιώντας overlapping pooling ενώ ήταν πιο δύσκολο να παρουσιαστεί overfitting, ένα από τα βασικότερα προβλήματα λόγω του μεγάλου αριθμού παραμέτρων.

Hidden Layer		Design
Convolution	1	96 filters in size 11x11 with max pooling in size 3x3
	2	256 filters in size 5x5 with max pooling in size 3x3
	3	384 filters in size 3x3 without pooling in size 3x3
	4	384 filters in size 3x3 without pooling
	5	256 filters in size 3x3 with max pooling in size 3x3
Fully Connected	1	4096 nodes with LeakyRelu activation function
	2	4096 nodes with LeakyRelu activation function
	3	100 nodes with LeakyRelu activation function

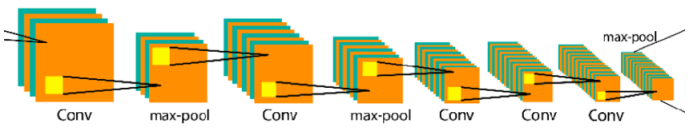


FIGURE 3: AlexNet architecture

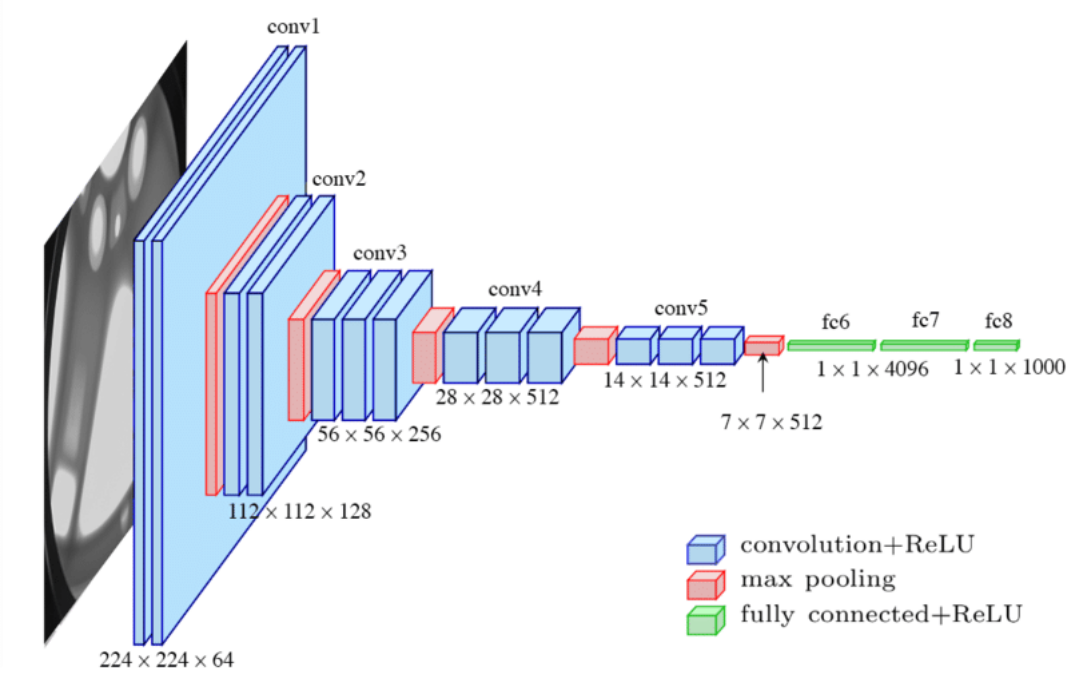
Σχήμα 2.12: *Αρχιτεκτονική AlexNet [11]*

VGG16-VGG19

Το VGG16 είναι ένα συνελικτικό νευρωνικό δίκτυο που δημιουργήθηκε από τους K. Simonyan και A. Zisserman από το πανεπιστήμιο της Οξφόρδης και δημοσιεύτηκε πρώτη φορά στο paper "Very Deep Convolutional Networks for Large-Scale Image Recognition" [43]. Το μοντέλο πετυχαίνει ακρίβεια 92.7% στο σύνολο δεδομένων ImageNet και ήταν ένα από τα πιο διάσημα μοντέλα που κατατέθηκαν στο ILSVRC-2014. Ο αριθμός 16 στο όνομα του οφείλεται στην αρχιτεκτονική του, η οποία αποτελείται από 13 επίπεδα συνέλιξης και 3 πλήρως συνδεδεμένα επίπεδα. Το δίκτυο έχει αρκετά μεγάλο μέγεθος με περίπου 138.000.000 παραμέτρους.

Στο ImageNet οι εικόνες έχουν σταθερό μέγεθος 224x224 RGB, για το λόγο αυτό η είσοδος στο πρώτο convolutional layer έχει μέγεθος 224x224x3. Η εικόνα στη συνέχεια περνάει από μία σειρά από convolutional layers στα οποία τα φίλτρα που χρησιμοποιούνται έχουν μέγεθος 3x3. Το convolutional stride είναι σταθερό και ίσο με 1 όπως και το spatial padding για τα φίλτρα μεγέθους 3x3. Ακόμη, υπάρχουν 5 max pooling layers τα οποία ακολουθούν ορισμένα από τα convolutional layers. Στο τέλος βρίσκονται 3 Fully-Connected (FC) layers, εκ των οποίων τα πρώτα δύο έχουν 4096 κανάλια το καθένα ενώ το τελευταίο πραγματοποιεί την κατηγοριοποίηση σε 1000 κατηγορίες και συνεπώς αποτελείται από 1000 κανάλια. Το τελευταίο επίπεδο είναι ένα soft-max layer που χρησιμοποιείται ως συνάρτηση ενεργοποίησης για την κατάλληλη κατηγοριοποίηση.

Το VGG19 είναι ένα δίκτυο παρόμοιας αρχιτεκτονικής με το VGG16 με τη διαφορά ότι έχει 3 παραπάνω συνελικτικά επίπεδα, πετυχαίνοντας ακρίβεια 92% στο σύνολο δεδομένων ImageNet.

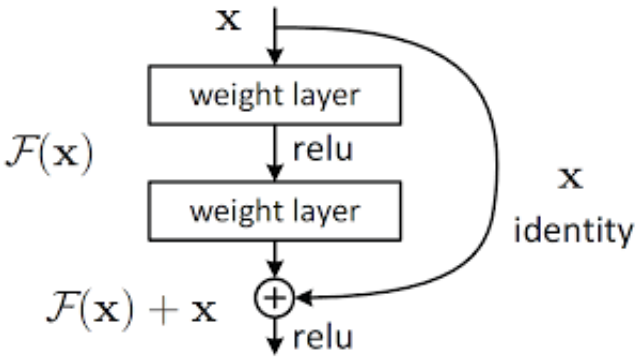


Σχήμα 2.13: *Αρχιτεκτονική VGG16 [12]*

ResNet

Το μοντέλο ResNet δημοσιεύθηκε πρώτη φορά το 2015 από τους Kaiming He, Xiangyu Zhang, Shaoqing Ren και Jian Sun στο paper “Deep Residual Learning for Image Recognition” [44]. Την ίδια χρονιά πήρε την πρώτη θέση στον διαγωνισμό ILSVRC 2015 με ποσοστό top-5 error 3.57%.

Μέχρι τη δημιουργία του, παρά το έντονο ενδιαφέρον των ερευνητών για τα βαθιά συνελικτικά δίκτυα και τη συνεχή βελτίωση τους, είχε παρατηρηθεί πως με την αύξηση του βάθους του δικτύου η επίδοση τους από ένα σημείο και μετά μειωνόταν δραματικά λόγω του vanishing gradient.Το πρόβλημα αυτό προκύπτει κατά το backpropagation όπου οι τιμές των gradients του δικτύου παίρνουν πολύ μικρές τιμές, με αποτέλεσμα τα βάρη να μην επηρεάζονται και η εκπαίδευση να επιβραδύνεται.



Σχήμα 2.14: *Residual Block [13]*

Αυτό μπόρεσε να επιλυθεί χρησιμοποιώντας residual blocks, τα οποία προσφέρουν τη δυνατότητα στις συνδέσεις να παρακάμψουν ένα ή περισσότερα επίπεδα υλοποιώντας ένα residual mapping. Πιο συγκεκριμένα, με τη σύνδεση αυτή που ονομάζεται ‘skip connection’, το τελικό αποτέλεσμα δεν είναι ίδιο με πριν. Χωρίς το skip connection, η είσοδος X πολλαπλασιάζεται με τα βάρη του επιπέδου προσθέτοντας το bias, στη συνέχεια περνά από τη συνάρτηση ενεργοποίησης και παρέχεται η έξοδος H(x).

$$H(x) = f(wx + b)$$

ή

$$H(x) = f(x)$$

Με τη προσθήκη της συντόμευσης το αποτέλεσμα H(x) γίνεται ίσο με H(x)=f(x)+x.

Στις περιπτώσεις που οι διαστάσεις της εισόδου και εξόδου είναι ίσες, οι συνδέσεις συντόμευσης υλοποιούνται ανά δύο συνελικτικά επίπεδα, πολλαπλασιάζοντας την είσοδο με τον μοναδιαίο πίνακα.Στην αντίθετη περίπτωση πραγματοποιείται zero-padding για την αύξηση των διαστάσεων με τη προσθήκη μηδενικών στις άκρες ή την εφαρμογή ενός 1x1 συνελικτικού δικτύου που αντιστοιχεί στις διαστάσεις της εισόδου με το αποτέλεσμα να διαμορφώνεται από τη σχέση

$$H(x) = f(x) + w1$$

Με αυτό τον τρόπο το ResNet λύνει το πρόβλημα το vanishing gradient προσφέροντας

εναλλακτικά μονοπάτια ενώ όταν ένα επίπεδο βλάπτει την απόδοση μπορεί να παρακαμφθεί. Δημιουργήθηκαν, συνεπώς, δίκτυα με πολύ μεγαλύτερο βάθος που μπορούσαν εύκολα να βελτιστοποιηθούν, χωρίς να αυξάνεται το σφάλμα εκπαίδευσης με την αύξηση των επιπέδων, παράγοντας καλύτερα αποτελέσματα από τα νευρωνικά δίκτυα που υπήρχαν έως τότε.

Η αρχιτεκτονική τους αποτελείται από συνελιτικά επίπεδα των οποίων οι πυρήνες αυξάνονται με την αύξηση των επιπέδων. Υπάρχει μόνο ένα max-pooling layer με pooling size 3x3 και stride 2, ύστερα από το πρώτο επίπεδο. Στη συνέχεια δεν πραγματοποιείται ιδιαίτερη μείωση των διαστάσεων κατά τη διάρκεια της εκπαίδευσης. Ακόμη, χρησιμοποιούνται average-pooling layers στη θέση των fully connected, μειώντας την πολυπλοκότητα του μοντέλου και βελτιώνοντας την ικανότητα του να συνδέει τους χάρτες ενεργοποίησης(feature maps) με τις κατηγορίες εξόδου. Τέλος, το output layer έχει 1000 νευρώνες, ίσο τον αριθμό των κατηγοριών στο ImageNet, ενώ εφαρμόζεται softmax συνάρτηση ενεργοποίησης για την παραγωγή των αντίστοιχων πιθανοτήτων για κάθε κατηγορία.

Inception

Η αρχιτεκτονική του Inception v1 παρουσιάστηκε πρώτη φορά στον διαγωνισμό ILSVRC 2014 κερδίζοντας την πρώτη θέση με την ονομασία GoogLeNet [45]. Στην περίοδο όπου οι επιστημονική κοινότητα επικεντρωνόταν στην αύξηση των επιπέδων για μεγαλύτερη ακρίβεια, κάνοντας ένα δίκτυο πιο βαθύ, ερευνητές της Google παρατήρησαν πως ένα δίκτυο μπορούσε να βελτιώσει την απόδοση του με το να γίνει πιο "πλατύ".

Ένα από τα προβλήματα που κλήθηκε να λύσει ήταν η επιλογή του σωστού μεγέθους πυρήνα. Είναι σύνηθες, όπως θα δούμε και στο σύνολο δεδομένων μας παρακάτω, η θέση ενός αντικειμένου στην εικόνα να διαφέρει σημαντικά ανάμεσα στα δεδομένα και η επιλογή φίλτρου να είναι δύσκολη. Ακόμη, τα βαθιά νευρωνικά δίκτυα αντιμετώπιζαν το πρόβλημα της υπερεκπαίδευσης ενώ ήταν ιδιαίτερα ακριβά υπολογιστικά.

Για τους λόγους αυτούς, οι αρχιτέκτονες του Inception πρότειναν τη χρήση πολλαπλών φίλτρων, διαφορετικού μεγέθους, στο ίδιο επίπεδο. Πιο συγκεκριμένα, σε ένα μόνο επίπεδο το Inception μπορεί να εφαρμόζει μία συνέλιξη 5x5, μία συνέλιξη 3x3, μία συνέλιξη 1x1 και ένα max-pooling, αφήνοντας το μοντέλο να επιλέξει πώς θα χρησιμοποιήσει κάθε πληροφορία που παράγεται. Φυσικά, η αύξηση της πυκνότητας ενός μοντέλου συνεπάγεται και την αύξηση του υπολογιστικού κόστους, αφού τα συνελικτικά φίλτρα μεγέθους 5x5 είναι ακριβά υπολογιστικά και τα εξαγόμενα feature maps για κάθε επίπεδο είναι σαφώς περισσότερα.

Για την επίλυση του προβλήματος αυτού, οι ερευνητές εφάρμοσαν πριν από κάθε συνέλιξη 3x3 και 5x5 μία επιπρόσθετη συνέλιξη διαστάσεων 1x1. Η συνέλιξη μεγέθους 1x1 κοιτάει μία τιμή κάθε φορά, εφαρμοζόμενη όμως στα διαφορετικά κανάλια εξάγει χωρικές πληροφορίες που υπάρχουν μεταξύ τους, μειώνοντας τις τελικές διαστάσεις. Με τη μείωση των διαστάσεων των πινάκων που δέχεται σαν είσοδο κάθε επίπεδο, στο Inception μπορούσαν να εφαρμοστούν παράλληλοι μετασχηματισμοί για κάθε επίπεδο, με αποτέλεσμα να δημιουργούνται δίκτυα που ήταν, αρχικά, βαθιά (πολλά επίπεδα) αλλά και πλατιά (πολλές παράλληλες διεργασίες).

Η ιδέα στην οποία βασίζεται το Inception οδήγησε στη δημιουργία των Inception v2 και Inception v3 τα οποία ήταν μία βελτιστοποιημένη έκδοση του αρχικού, κυρίως λόγω της μείωσης των διαστάσεων των πυρήνων. Για παράδειγμα, στο Inception V3 η συνέλιξη

5x5 αντικαταστάθηκε με δύο συνεχόμενες συνελίξεις 3x3. Στη συνέχεια δημιουργήθηκε το Inception v4 που συνδυαζόμενο με την ιδέα των υπολειπόμενων συνδέσεων (residual connections) του ResNet οδήγησε στην ανάπτυξη του Inception-Resnet.

Inception - Resnet v1 και v2

Βασιζόμενοι στην κεντρική ιδέα του Resnet σε συνδυασμό με την αρχιτεκτονική του Inception, το 2016 παρουσιάστηκαν από την Google δύο νέα μοντέλα, τα Inception - Resnet v1 και v2. Το Inception - Resnet v1 είχε παρόμοιο υπολογιστικό κόστος με το Inception v3 και το Inception - Resnet v2 με το Inception v4, έχοντας διαφορετικές υπερπαραμέτρους μεταξύ τους. Στόχος ήταν η προσαρμογή των residual connections στην έξοδο που παράγουν τα Inception blocks. Για να επιτευχθεί αυτό, χρειαζόταν η είσοδος και η έξοδος πριν και μετά τις διεργασίες που πραγματοποιεί το Inception σε κάθε block να έχουν το ίδιο μέγεθος. Για αυτό, προστέθηκε μία συνέλιξη 1x1 μετά από κάθε άλλη συνέλιξη, ενώ τα pooling layers αφαιρέθηκαν (παρόλα αυτά συνέχισαν να υπάρχουν μέσα στα reduction blocks) [46].

Xception

Το Xception πήρε το όνομά του από το Extreme Inception [47], αποτελεί δηλαδή μία extreme έκδοση του μοντέλου που αναφέρουμε παραπάνω. Η υπερβολική εκδοχή του Inception βασίζεται στην ιδέα πως πρώτα θα χρησιμοποιείται μία συνέλιξη μεγέθους 1x1 για την εύρεση των διακαναλικών συσχετίσεων και στη συνέχεια θα υπολογίζονται οι υπόλοιπες nxn συνελίξεις.

Στο σημείο αυτό, θα εξηγήσουμε έναν νέο όρο, τη βαθιά διαχωρίσιμη συνέλιξη (depthwise separable convolution), μία λειτουργία που από το 2014 έχει αποκτήσει ευρεία χρήση στα νευρωνικά δίκτυα. Η βαθιά διαχωρίσιμη συνέλιξη αποτελείται από μία βαθιά συνέλιξη (μία συνέλιξη που πραγματοποιείται ανεξάρτητα σε κάθε κανάλι της εισόδου) και ακολουθείται από μία σημείο-προς-σημείο (pointwise) συνέλιξη (συνέλιξη 1x1) για την τροποποίηση των διαστάσεων. Η διαφορά του Xception με τη βαθιά διαχωρίσιμη συνέλιξη είναι πως το pointwise convolution πραγματοποιείται πρώτο, ακολουθώντας την υλοποίηση του Inception v3.

Το Xception, λοιπόν, αποτελεί μία αρχιτεκτονική δομημένη από βαθιά διαχωρίσιμα συνελικτικά blocks προσθέτοντας max pooling layers. Τα στοιχεία αυτά, είναι συνδεδεμένα με υπολειπόμενες συνδέσεις όπως είδαμε και στο ResNet.

EfficientNet

Το 2019, στο paper "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks" [48] παρουσιάστηκε μία καινούργια μέθοδος που χρησιμοποιούσε σύνθετους συντελεστες (compound coefficient) για να αυξήσει τις διαστάσεις της δομής ενός CNN. Αντίθετα με τα προηγούμενα μοντέλα, που χρησιμοποιούν αυθαίρετους συντελεστές για την τροποποίηση του πλάτους, βάθους και ανάλυσης της εικόνας, η μέθοδος αυτή διαμόρφωνε τα μεγέθη ομοιόμορφα χρησιμοποιώντας προκαθορισμένους συντελεστές. Για παράδειγμα, εάν θέλουμε να χρησιμοποιήσουμε 2^N περισσότερους υπολογιστικούς πόρους, τότε μπορούμε απλά να αυξήσουμε το βάθος του δικτύου επί α^N, το πλάτος επί β^N και το μέγεθος της εικόνας επί γ^N, όπου τα α,β,γ είναι σταθεροί συντελεστές καθορισμένοι από μία αναζήτηση

πλέγματος (grid search) στο αρχικό μικρότερο μοντέλο. Το Efficientnet χρησιμοποιεί ένα σύνθετο συντελεστή φ για να διαμορφώσει κατάλληλα και ομοιόμορφα τις διαστάσεις του δικτύου. Η λογική πίσω από αυτή τη μέθοδο, βασίζεται στην ιδέα πως εάν θέλουμε να εισάγουμε στο σύστημα μία μεγαλύτερη εικόνα, τότε το δίκτυο χρειάζεται περισσότερα επίπεδα και περισσότερα κανάλια για να εντοπίσει λεπτομέρειες.

Τα μοντέλα EfficientNet, αυτή τη στιγμή, πετυχαίνουν μεγαλύτερη απόδοση στο ImageNet από οποιοδήποτε άλλο δίκτυο. Χαρακτηριστικά, το EfficientNet-B7 πετυχαίνει state-of-the-art 84.4% top-1 / 97.1% top-5 accuracy, ενώ παραμένει σχετικά μικρό σε μέγεθος και γρήγορο.

2.2.8 Μεταφορά Γνώσης (Transnfer Learning)

Δεδομένου ενός πεδίου πηγής Ds με αντίστοιχη διεργασία Ts, και το πεδίο προορισμού Dt με αντίστοιχη διεργασία στόχο Tt, η μεταφορά γνώσης έχει ως σκοπό να μεταφέρει τη σχετική γνώση που περιέχεται στο Ts και Ds για την ενίσχυση της απόδοσης της συνάρτησης πρόγνωσης fT(.) στη διεργασία στόχο Tt και πεδίο στόχο Dt, όπου Ds ≠ *Dt* ή Ts ≠ *Tt*[49].

Ο παραπάνω ορισμός πρόκειται για μεταφορά γνώσης από μία πηγή, το οποίο και αποτελεί συνηθέστερη περίπτωση στη μελέτη της μεταφοράς γνώσης.

Στόχος, λοιπόν, της μεταφοράς γνώσης είναι η βελτίωση της απόδοσης και η μείωση των αναγκαίων ετικετών (labels) για τα δεδομένα που απαιτούνται σε ένα πεδίο στόχο, χρησιμοποιώντας ήδη υπάρχουσα γνώση από ένα παρόμοιο πεδίο. Στη περιπτωση που το πεδίο της πηγής και του στόχου δεν σχετίζονται, η μεταφορά γνώσης μπορεί να μην είναι τελικά επιτυχημένη [50].

Πιο συγκεκριμένα, στον τομέα της κατηγοριοποίησης εικόνων (image classification) η μεταφορά γνώσης βασίζεται στη θεωρία πως εάν ένα μοντέλο έχει ήδη εκπαιδευτεί σε ένα αρκετά μεγάλο και γενικό σύνολο δεδομένων, αυτό μπορεί να χρησιμοποιηθεί ως ένα γενικό μοντέλο του οπτικού κόσμου. Με αυτό τον τρόπο, αξιοποιούνται οι ήδη υπάρχοντες χάρτες χαρακτηριστικών (feature maps) χωρίς να απαιτείται η εκπαίδευση ενός μοντέλου από την αρχή σε ένα πολύ μεγάλο σύνολο δεδομένων.

Σε περίπτωση που έχουμε μικρά σύνολα δεδομένων, η μεταφορά γνώσης μας βοηθά να αποφύγουμε το ενδεχόμενο της υπερεκπαίδευσης (overfitting), που προκαλεί μείωση της απόδοσης του συστήματος. Με τη μεταφορά γνώσης, μπορούμε να εκπαιδεύσουμε ορισμένα επίπεδα του μοντέλου (ή ακόμη και ολόκληρο το δίκτυο αν απαιτείται) μέσω fine tuning χρησιμοποιώντας σύνολα δεδομένων που δεν έχουν αρκετές εικόνες[51].

Πιο συγκεριμένα, η χρήση των προ-εκπαιδευμένων μοντέλων υλοποιείται με τις εξής μεθόδους :

- Εξαγωγή Χαρακτηριστικών (Feature extraction): Χρησιμοποιώντας την έξοδο ενός ενδιάμεσου επιπέδου του pre-trained CNN μπορούμε να εξάγουμε χαρακτηριστικά για μία εικόνα. Ο χάρτης χαρακτηριστικών περιλαμβάνει τα χαρακτηριστικά της εισόδου για τις διάφορες περιοχές της εικόνας. Το τελικό μέρος του pre-trained μοντέλου είναι εξειδικευμένο στην κατηγοριοποίηση για την οποία χρησιμοποιήθηκε. Για τον λόγο αυτό, αφαιρούμε το επίπεδο εξόδου του CNN και χρησιμοποιούμε το υπόλοιπο ως εξαγωγέα χαρακτηριστικών.

- Fine tuning : Στις περιπτώσεις που το σύνολο δεδομένων του προβλήματος και του αρχικού pre-trained μοντέλου διαφέρουν, μπορούμε να βελτιώσουμε την απόδοση του μοντέλου με τη χρήση του Fine Tuning. Στη συγκεκριμένη μέθοδο, χρησιμοποιούμε τα βάρη από την προ-εκπαίδευση και ¨ξεπαγώνουμε¨ ορισμένα από τα top layers ώστε να εκπαιδευτούν μαζί με το υπόλοιπο μοντέλο μας. Με αυτό τον τρόπο, δίνεται η δυνατότητα να διαφοροποιηθούν τα χαρακτηριστικά που παράγονται στα top layers ώστε να είναι πιο σχετικά με το πρόβλημα μας.

Μέρος II

Πρακτικό Μέρος

Κεφάλαιο 3

Ανάλυση και Σχεδίαση

Στο κεφάλαιο αυτό παρουσιάζονται αναλυτικά τα βήματα που ακολουθήθηκαν για την υλοποίηση του συστήματος. Αρχικά περιγράφεται η επιλογή του συνόλου δεδομένων και τα προγραμματιστικά εργαλεία που χρησιμοποιήθηκαν. Στη συνέχεια αναλύεται η υλοποίηση και οι βασικοί αλγόριθμοι του συστήματος καθώς και η δομή του κώδικα. Τέλος, περιγράφονται τα υπολογιστικά συστήματα που χρησιμοποιήθηκαν για την διεξαγωγή των πειραμάτων.

3.1 Δεδομένα

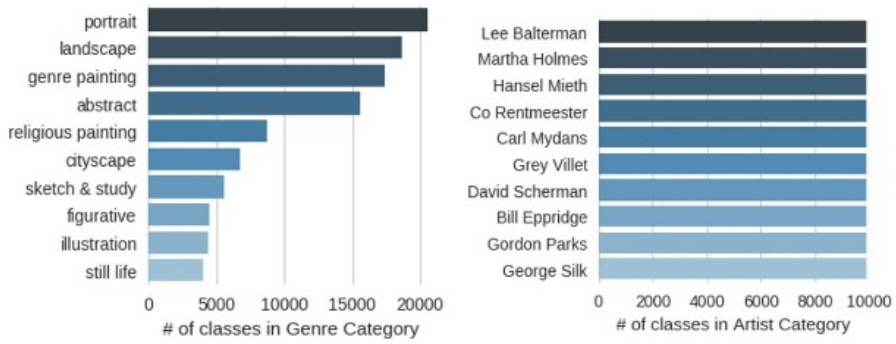
Η επιλογή του συνόλου δεδομένων (dataset) είναι κύριο στοιχείο της διεξαγωγής ενός πειράματος στον τομέα των Νευρωνικών Δικτύων. Η εργασία μας απαιτεί δύο είδη dataset, ένα που αποτελείται από έργα τέχνης στον χώρο, κυρίως, της ζωγραφικής και ένα στον τομέα της φωτογραφίας. Λόγω της αυξανόμενης ψηφιοποίησης των έργων τέχνης, ήταν δυνατή η πρόσβαση σε μεγάλα και ολοκληρωμένα σύνολα δεδομένων που έχουν συλλεχθεί από online μουσεία και art gallery βάσεις δεδομένων.

3.1.1 Σύνολο δεδομένων έργων τέχνης

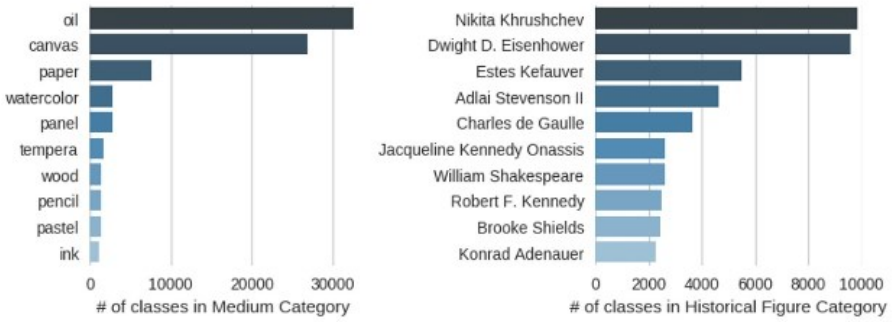
Για την πρόσβαση σε πίνακες ζωγραφικής, χρησιμοποιήθηκε το dataset ART5OOK [14], ένα σύνολο δεδομένων με πάνω από 500.000 έργα τέχνης. Δημιουργήθηκε το 2017 από τους Hui Mao, Ming Cheung και James She για την υλοποίηση του project τους “DeepArt” [52]. Το dataset περιλαμβάνει έργα τέχνης που έχουν συλλεχθεί από τέσσερις κυρίως συλλογές στο διαδίκτυο : Rijksmuseum, Google Arts and Culture, WikiArt.org - Visual Art Encyclopedia και Web Gallery of Art. Τα πλεονεκτήματα του συγκεκριμένου συνόλου δεδομένων είναι τρία. Αρχικά, αποτελεί μία τεράστια συλλογή έργων τέχνης, πολύ μεγαλύτερη από τα περισσότερα αντίστοιχα σύνολα. Αυτός ήταν και ο λόγος για τον οποίο επιλέχθηκε για τα πειράματα μας. Ο μεγάλος όγκος δεδομένων μας δίνει τη δυνατότητα να έχουμε πιο σαφή αποτελέσματα και πιο ολοκληρωμένη έρευνα πάνω στη σχέση της φωτογραφίας με τη ζωγραφική, έχοντας πρόσβαση σε ένα πολύ σημαντικό αριθμό έργων τέχνης. Επιπλέον, περιέχει πληροφορίες για κάθε έργο, προσδιορίζοντας τον καλλιτέχνη, το είδος, την χρονιά κ.ο.κ. Τέλος, το dataset είναι οργανωμένο σε φακέλους βοηθώντας τους χρήστες του να το αξιοποιήσουν όπως εκείνοι επιθυμούν, ενώ είναι δημόσιο και συνεπώς προσβάσιμο σε κάθε πρόσωπο που επιθυμεί να

το χρησιμοποιήσει για την έρευνα του.

Σχήμα 3.1: *Παραδείγματα του συνόλου Art500k [14]*



(α΄) α



(β΄) β

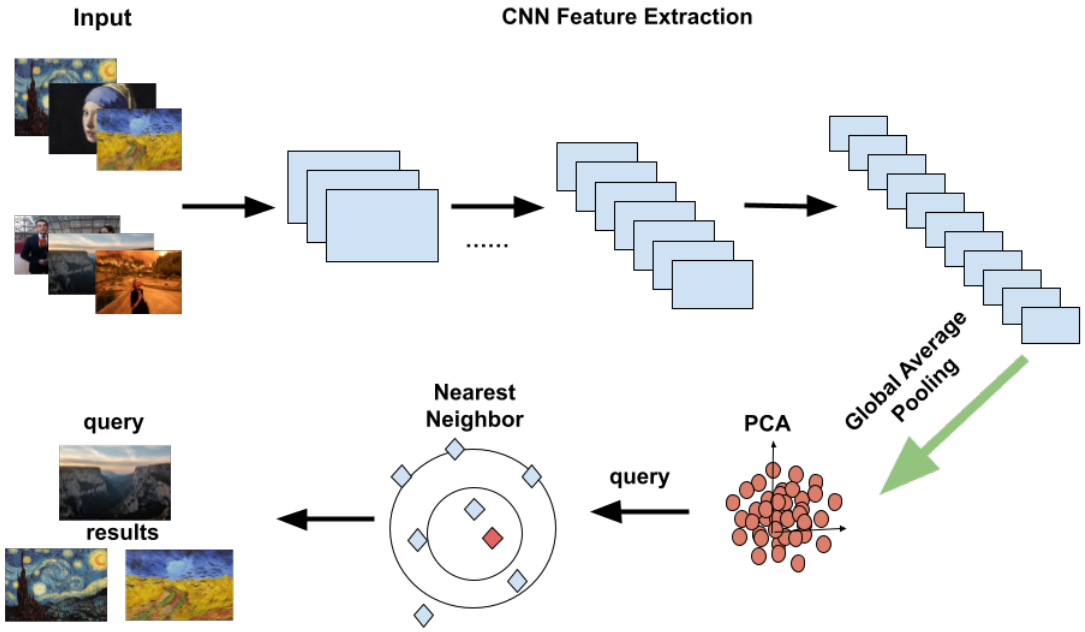
Σχήμα 3.2: *(α) Στατιστικά για τις κατηγορίες Είδους (αριστερά) και Καλλιτέχνη (δεξιά) (β) Στατιστικά για τις κατηγορίες Μέσου (αριστερά) και Ιστορικής Φιγούρας (δεξιά) [14].*

3.1.2 Σύνολο Φωτογραφιών

Για το δεύτερο μέρος, χρειαστήκαμε ένα σύνολο δεδομένων από φωτογραφίες. Θεωρήσαμε πως θα ήταν περισσότερο ενδιαφέρον, η σύγκριση έργων τέχνης με φωτογραφίες που προκύπτουν από σκηνές τις καθημερινότητας και δεν αποτελούν επιτηδευμένη αναπαράσταση ενός προγενέστερου έργου αφού τότε η εύρεση ομοιοτήτων θα ήταν ευκολότερη. Όπως αναλύσαμε και στο κεφάλαιο 1, έχουμε δει και προγενέστερες μελέτες που πραγματοποιούν μία προσπάθεια σύνδεσης των δύο αυτών συνόλων καταλήγοντας σε πολύ ενδιαφέροντα συμπεράσματα. Στη δική μας περίπτωση, χρησιμοποιήθηκαν φωτογραφίες από τη βάση δεδομένων του φωτογραφικού πρακτορείου SOOC. Το σύνολο των φωτογραφιών αποτελείται κυρίως από εικόνες της καθημερινότητας καθώς και της επικαιρότητας των τελευταίων χρόνων. Οι φωτογραφίες αυτές, αν και έχουν ως στόχο την αποτύπωση των γεγονότων στο πλαίσιο του ρεπορτάζ, έχουν διαμορφωθεί από τις βασικές αρχές της καλλιτεχνικής φωτογραφίας. Το SOOC καλύπτει ένα ευρύ φάσμα γεγονότων, δίνοντας μας τη δυνατότητα να εισάγουμε στα πειράματα μας φωτογραφίες από πολλαπλά είδη φωτογραφίας όπως πορτραίτα, τόπια, αρχιτεκτονική φωτογραφία κ.α. Το γεγονός αυτό είναι ιδιαίτερα χρήσιμο, αφού μπορούμε να ελέγξουμε την αποτελεσματικότητα των μοντέλων στα διάφορα χαρακτηριστικά που αναγνωρίζουν και την ακρίβεια στην εύρεση συνδέσεων σε διαφορετικές θεματολογίες.

Σχήμα 3.3: *Παραδείγματα Φωτογραφιών από SOOC*

3.2 Σχεδιασμός Υλοποίησης



Σχήμα 3.4: *Σχηματική Αναπαράσταση της Υλοποίησης. Σημ. Τα παραδείγματα που φαίνονται στην εικόνα δεν είναι τα πραγματικά αποτελέσματα, χρησιμοποιούνται μόνο για την καλύτερη κατανόηση.*

3.2.1 Προ-επεξεργασία Δεδομένων

Για την επεξεργασία των δεδομένων μας χρησιμοποιήσαμε εργαλεία από τη βιβλιοθήκη Keras [53]. Το Keras είναι μία βιβλιοθήκη ανοικτού κώδικα που παρέχει στη γλώσσα προγραμματισμού Python τη δυνατότητα να χρησιμοποιεί τεχνητά νευρωνικά δίκτυα, ενώ λειτουργεί σαν interface για τη βιβλιοθήκη Tensorflow.

Αρχικά, τα δεδομένα μας είναι χωρισμένα στους φακέλους Artists1, Artists2, Art Movement. Κάθε φάκελος από τους δύο πρώτους περιέχει υπο-φακέλους με το όνομα του Καλλιτέχνη όπου περιέχει τα έργα του. Στον φάκελο Art Movement έχουμε είδη ζωγραφικής με κάθε φάκελο να περιέχει ορισμένα παραδείγματα. Ο φάκελος αυτος δε θα χρησιμοποιηθεί στα πειράματα μας, αφού περιέχει δεδομένα που προυπάρχουν στους άλλους δύο. Στην ανάκτηση των αρχείων, η κατηγοριοποίηση σε φακέλους δε θα παίξει κάποιο ρόλο, αφού δεν μας ενδιαφέρουν στοιχεία που λειτουργούν σαν ετικέτες. Με τη χρήση του module pathlib, μπορούμε να χρησιμοποιούμε τα paths για κάθε αρχείο στον υπολογιστή μας, ώστε να έχουμε πρόσβαση σε αυτό.

Ένα σημαντικό πρόβλημα που έπρεπε να επιλυθεί ήταν η εύρεση των εικόνων που ήταν ¨σπασμένες¨, δηλαδή δεν ήταν σωστά αποθηκευμένες στο dataset. Για τον λόγο αυτό χρειάστηκε να διατρέξουμε όλες τις εικόνες και να εντοπίσουμε όσες χρειαζόταν να αφαιρέσουμε.

Ευτυχώς, ο αριθμός τους ήταν αρκετά μικρός και το σύνολο δεδομένων μειώθηκε σε αμελητέο βαθμό.

Στη συνέχεια πραγματοποιούμε την προ-επεξεργασία στα δεδομένα. Πρώτο βήμα είναι η μετατροπή των εικόνων σε πίνακες στη δομή Ύψος, Πλάτος, Κανάλι για να είναι πιο εύκολη η επεξεργασία τους. Αμέσως μετά γίνεται διαίρεση κάθε στοιχείου του πίνακα με το 255, ώστε οι τιμές να κυμαίνονται μεταξύ 0 και 1 και οι υπολογισμοί να είναι πιο εύκολοι και γρήγοροι. Τέλος, ένα CNN χρειάζεται να δέχεται ως είσοδο εικόνες με σταθερό και προκαθορισμένο μέγεθος. Για τον λόγο αυτό ορίσαμε τις διαστάσεις των εικόνων σύμφωνα με τις απαιτήσεις του κάθε μοντέλου.

3.2.2 Εξαγωγή χαρακτηριστικών

Το επόμενο βήμα είναι η εξαγωγή σημαντικών χαρακτηριστικών από τις εικόνες εισόδου. Όπως αναλύσαμε και στο κεφάλαιο 2, η εξαγωγή χαρακτηριστικών είναι μία διαδικασία που υλοποιείται με transfer learning. Για την υλοποίηση της, κατά τη διάρκεια του transfer learning αφαιρούμε τα fully-connected layers στο τέλος του μοντέλου και εξάγουμε τα χαρακτηριστικά στο προηγούμενο επίπεδο. Αυτή είναι μία συνήθης μέθοδος για την εξαγωγή χαρακτηριστικών, αφού θέλουμε να αποφύγουμε την κατηγοριοποίηση των εικόνων στο τέλος του μοντέλου. Η διαδικασία αυτή γίνεται ιδιαίτερα εύκολη θέτοντας το include top = False κατά τη μεταφορά του εκάστοτε μοντέλου. Τα convolutional layers στην υπόλοιπη αρχιτεκτονική, χρησιμοποιούνται ως εξαγωγείς χαρακτηριστικών. Στην αρχή ανιχνεύουν ακμές και σχήματα και όσο προχωράει η εκπαίδευση αναγνωρίζουν αντικείμενα και πιο περίπλοκες μορφές. Στο τέλος, έχουμε ένα πολύ μεγάλο αριθμό από βασικά χαρακτηριστικά, που περιλαμβάνει κάθε εικόνα εισόδου, σε αριθμητική μορφή.

Στο συγκεκριμένο σημείο, επιλέξαμε να υλοποιήσουμε το σύστημα μας με διαφορετικά μοντέλα θέλοντας να μελετήσουμε τον τρόπο που κάθε μοντέλο εξάγει χαρακτηριστικά και σε ποια σημεία εστιάζει. Τα μοντέλα αυτά είναι το VGG16, VGG19, Xception, Inception-Resnet και το EfficientNetb7. Για κάθε ένα από τα μοντέλα, εφαρμόσαμε predict(X), όπου X η εικόνα εισόδου, για ολόκληρο το σύνολο των δεδομένων μας.

3.2.3 Global Average Pooling

Τα συνήθη συνελικτικά νευρωνικά δίκτυα πραγματοποιούν τις συνελίξεις στα χαμηλότερα επίπεδα του δικτύου. Για να γίνει η κατηγοριοποίηση, τα feature maps από το τελευταίο συνελικτικό δίκτυο μετατρέπονται σε διανύσματα και εισάγονται σε fully connected layers ακολουθούμενα από ένα softmax logistic regression layer. Με αυτό τον τρόπο τα συνελικτικά επίπεδα λειτουργούν ως feature extractors και τα τελικά χαρακτηριστικά κατηγοριοποιούνται κατά τις συνήθεις μεθόδους.

Παρόλα αυτά, τα fully connected layers τείνουν να ενισχύουν την πιθανότητα για overfitting και να παρεμποδίζουν την ικανότητα του δικτύου για γενίκευση. Σε αυτή την περίπτωση προτείνεται συνήθως η χρήση του Dropout out layer το οποίο θέτει τυχαία ορισμένες συναρτήσεις ενεργοποίησης των fully connected layers ίσες με το 0.

Για την αποφυγή χρήσης των fully connected layers μετά την εξαγωγή χαρακτηριστικών, αποφασίσαμε να χρησιμοποιήσουμε το global average pooling. Έστω πως έχουμε feature

maps διαστάσεων HxWxD. Εφαρμόζοντας το Global Average Pooling παίρνουμε για κάθε feature map το μέσο όρων των τιμών του. Με αυτόν τον τρόπο οι τελικές διαστάσεις μειώνονται σε 1x1xD. Συνεπώς, το τελικό μέγεθος των χαρακτηριστικών που εξάγονται και αποτελούν την είσοδο του PCA είναι σαφώς μικρότερο.

Το πλεονέκτημα του average global pooling έναντι των fully connected layers είναι πως τα χαρακτηριστικά είναι πιο πιστά ως προς τα αποτελέσματα που έχουν εξάγει τα προηγούμενα συνελικτικά επίπεδα και δεν απαιτείται η συσχέτιση τους με τις τελικές κατηγορίες του δικτύου. Ένα ακόμη πλεονέκτημα, είναι πως δεν υπάρχει κάποια παράμετρος για βελτιστοποίηση στο επίπεδο αυτό και συνεπώς αποφεύγεται το ενδεχόμενο του overfitting. Τέλος, το global average pooling αθροίζει τις χωρικές πληροφορίες, με αποτέλεσμα τα εξαγόμενα χαρακτηριστικά να μένουν πιο πιστά στις αρχικές χωρικές πληροφορίες των δεδομένων εισόδου.

3.2.4 Ανάλυση Κύριων Συνιστωσών (Principal Component Analysis)

Στο σημείο αυτό, προχωρήσαμε στην περαιτέρω μείωση των διαστάσεων των feature maps χρησιμοποιήσαμε μία πολύ σημαντική μέθοδο μείωσης διαστάσεων το PCA (Principal Component Analysis) ή αλλιώς ανάλυση κύριων συνιστωσών.

Φυσικά, δε γίνεται να μειώσουμε τις διαστάσεις των διανυσμάτων ελαττώνοντας απλά το μέγεθος τους καθώς είναι πολύ πιθανόν να χαθούν σημαντικές πληροφορίες. Το PCA επιτρέπει να μειώσουμε τον αριθμό των παραμέτρων διατηρώντας όσες περισσότερες σημαντικές πληροφορίες είναι δυνατό. Για να μην έχουμε πολύ μεγάλες και περίπλοκες υπολογιστικές απαιτήσεις καθώς και μεγάλο χρόνο εκτέλεσης του προγράμματος, κληθήκαμε να επιλέξουμε έναν αριθμό χαρακτηριστικών (principal components) που να είναι αντιπροσωπευτικός και να μη χάνει σημαντικά δεδομένα.

Στη μέθοδο αυτή, αναπαριστούμε ένα σύνολο αρχικών μεταβλητών ως ένα νέο σύνολο που προκύπτει από το γραμμικό συνδυασμό τους. Κάθε συνιστώσα που προκύπτει δεν έχει την ίδια βαρύτητα. Η πρώτη περιέχει τη μεγαλύτερη διακύμανση, δηλαδή αντιπροσωπεύει τη μεγαλύτερη δυνατή μεταβλητότητα των δεδομένων. Όσο μεγαλύτερη είναι η διακύμανση, τόσο περισσότερη πληροφορία είναι καταχωρημένη από την πρώτη συνιστώσα. Με την ίδια λογική, ακολουθούν και οι υπόλοιπες συνιστώσες που προκύπτουν ύστερα από την εφαρμογή του.

Στην πράξη, αφού είχαμε όλα τα χαρακτηριστικά από τους πίνακες ζωγραφικής καθώς και από τις φωτογραφίες που λειτουργούν ως queries, υλοποιήσαμε το PCA μέσω του sklearn. Ο αριθμός των συνιστωσών (components) σε κάθε περίπτωση ήταν διαφορετικός, καθώς τα μοντέλα εξήγαγαν διαφορετικό αριθμό χαρακτηριστικών. Έτσι, μοντέλα με μεγαλύτερους εξαγόμενους πίνακες χρειάζονταν και μεγαλύτερο αριθμό συνιστωσών.

3.2.5 Κοντινότεροι Γείτονες (Nearest Neighbor)

Στη συνέχεια, με βάση τα χαρακτηριστικά που είχαμε πλέον στις κατάλληλες διαστάσεις, προχωρήσαμε στην αναζήτηση των πλησιέστερων έργων τέχνης ως προς τις φωτογραφίες εισόδου. Για την πραγματοποίηση του, χρησιμοποιήσαμε το μέθοδο του Nearest Neighbor. Η

μη επιβλεπόμενη υλοποίηση του Nearest Neighbor αποτελεί βάση για πολλές άλλες σημαντικές μεθόδους που αφορούν κυρίως τη μείωση διαστάσεων και την ομαδοποίηση.

Στη μέθοδο αυτή, για κάθε query που δίνουμε στο σύστημα μας επιστρέφονται τα k ¨κοντινότερα¨ σε αυτό δεδομένα. Η τιμή του k καθορίζεται από εμάς, και στην περίπτωση μας είναι ίσο με 10. Η απόσταση μεταξύ των σημείων μπορεί να υπολογιστεί με πολλούς τρόπους, παρόλα αυτά η πιο ευρέως διαδεδομένη, και αυτή που χρησιμοποιήσαμε, είναι η Ευκλείδεια απόσταση, που υπολογίζεται ως: l(p,q) = $\sqrt{(p_i - q_i)^2}$

Ο χρόνος και η μνήμη που απαιτείται για την υλοποίηση του Nearest Neighbor εξαρτάται από τον τρόπο υλοποίησης του. Ο πιο απλός τρόπος είναι με brute force, δηλαδή ο υπολογισμός της απόστασης για κάθε ζευγάρι δεδομένων. Έτσι, έχοντας N εισόδους και D διαστάσεις, η χρονική πολυπλοκότητα ορίζεται ως O(DN). Με την αύξηση, συνεπώς, των δεδομένων ο αλγόριθμος αρχίζει να μην είναι χρονικά αποδοτικός. Για τον λόγο αυτό δημιουργήθηκαν δύο νέοι τρόποι υλοποίησης, το K-D Tree και το Ball Tree.

Το K-D Tree στηρίζεται στη δημιουργία δομών με βάση τα δέντρα στοχεύοντας στη μείωση των απαιτούμενων υπολογισμών για την εύρεση αποστάσεων.Η βασική ιδέα είναι πως αν ένα σημείο A είναι πολύ μακριά από ένα σημείο B, τότε το B θα είναι πολύ κοντα στο σημείο C, για το οποίο γνωρίζουμε πως είναι πολύ μακριά από το A, χωρίς να απαιτείται ο ακριβής υπολογισμός της απόστασής τους. Σε αυτή την περίπτωση το υπολογιστικό κόστος μπορεί να μειωθεί σε O[Dlog(N)] για κάθε query. Όπως βλέπουμε, το K-D Tree μπορεί να είναι πολύ αποτελεσματικό κάτι που όμως απαιτεί μικρό αριθμό διαστάσεων (D < 20).

Για την επίλυση των προβλημάτων που προκύπτουν από το K-D Tree για πολύ μεγάλα D, υλοποιήθηκε η μέθοδος του Ball Tree. Σε αυτή την περίπτωση το κόστος υλοποίησης του δέντρου αυξάνεται, παρόλα αυτά ο αλγόριθμος μπορεί να είναι αποδοτικός ακόμη και σε πολύ μεγάλες διαστάσεις. Στη μέθοδο αυτή, τα δεδομένα χωρίζονται επαναληπτικά σε κόμβους που εκπροσωπούνται από ένα κέντρο C και μία ακτίνα r, ώστε κάθε στοιχείο του κόμβου να καθορίζεται από αυτά. Ο αριθμός, συνεπώς, των δεδομένων που ερευνώνται ως πιθανοί κοντινότεροι γείτονες μειώνεται, χρησιμοποιώντας τη τριγωνική ανισότητα επιτυγχάνοντας χρόνο O(Dlog(N)). Με βάση αυτή τη λογική, το Ball tree μπορεί να ξεπεράσει σε απόδοση το K-D tree για πολύ μεγάλες διαστάσεις, αν και ο βαθμός απόδοσής του εξαρτάται σημαντικά από τη δομή των δεδομένων.

Η βέλτιστη επιλογή της μεθόδου υλοποίησης για το Nearest Neighbor εξαρτάται όπους βλέπουμε από διάφορους παράγοντες,όπως το μέγεθος του dataset,τη δομή των δεδομένων, τον αριθμό k των κοντινότερων γειτόνων καθώς και τον αριθμό των queries. Στην υλοποίηση μας επιτρέψαμε στον αλγόριθμο να επιλέξει μόνος του την καταλληλότερη μέθοδο, βασιζόμενος στους παραπάνω παράγοντες.

3.3 Υπολογιστικά Συστήματα

Αρχικά, η υλοποίηση του κώδικα και η διεξαγωγή πειραμάτων πραγματοποιήθηκαν στο Google Colaboratory. Το υπολογιστικό αυτό περιβάλλον χρησιμοποιείται ευρέως καθώς διατίθεται δωρεάν στους χρήστες και τους δίνει τη δυνατότητα να τρέξουν προγράμματα σε γλώσσα Python παρέχοντας πρόσβαση σε GPUs. Για τους λόγους αυτούς επιλέχθηκε και από εμάς για το αρχικό στάδιο. Το μειονέκτημα του συγκεκριμένου περιβάλλοντος είναι η

περιορισμένη μνήμη που μπορούμε να χρησιμοποιήσουμε (15GB) καθώς και η μη επαρκής παροχή GPUs. Με τους παραπάνω περιορισμούς, τα αρχικά πειράματα πραγματοποιήθηκαν σε ένα μικρό αριθμό δεδομένων, χρησιμοποιώντας τμήμα του Toy Artwork Dataset που παρέχεται από το ART500K.

Στη συνέχεια τα πειράματα πραγματοποιήθηκαν στο Εθνικό Υπερυπολογιστικό Σύστημα ARIS (Advanced Research Information System) του Εθνικού Δικτύου Υποδομών Τεχνολογίας και Έρευνας (ΕΔΥΤΕ). Η σύνδεση στον υπερυπολογιστή γινόταν εξ αποστάσεως μέσω του πρωτοκόλλου SSH. Στη διεξαγωγή των πειραμάτων μας χρησιμοποιήσαμε κόμβους GPU, που ο καθένας διαθέτει 2 GPU Nvidia Tesla K40 και μνήμη 64GB. Η βοήθεια αυτή ήταν πολύτιμη, αφού μπορούσαμε να επεξεργαστούμε ολόκληρο το σύνολο δεδομένων παρά το μεγάλο μέγεθος και να τρέξουμε τα προγράμματα σε πολύ αποδοτικό χρόνο.

Κεφάλαιο 4

Αξιολόγηση Αποτελεσμάτων

Στο κεφάλαιο αυτό παρουσιάζονται τα αποτελέσματα των πειραμάτων και αναλύονται τα συμπεράσματα που εξαγάγαμε από αυτά.

4.1 Υλοποήσεις με διαφορετικά CNN μοντέλα

Για τα πειράματα μας, χρησιμοποιήσαμε 5 διαφορετικά προ-εκπαιδευμένα μοντέλα με διαφορετικό accuracy στο ImageNet. Στο τέλος κάθε μοντέλου, προσθέσαμε ένα global average pooling layer για τη μείωση των διαστάσεων του εξαγώμενου πίνακα χαρακτηριστικών. Με αυτόν τον τρόπο, θελήσαμε να παρατηρήσουμε πού εστιάζει κάθε μοντέλο κατά την εξαγωγή χαρακτηριστικών, πόσο αποδοτικά είναι στη συσχέτιση έργων τέχνης με μη επιβλεπόμενη μάθηση και τελικά ποιο θα ήταν προτιμότερο για τη συγκεκριμένη διεργασία.

Για κάθε μοντέλο, μετά την εφαρμογή του prediction σε κάθε εικόνα, παράγεται ένας πίνακας ίσος με 471.820 επί τον αριθμό των χαρακτηριστικών που εξάγονται κάθε φορά. Ο πίνακας αυτός αποτελεί την είσοδο του PCA, το οποίο μειώνει τις διαστάσεις του σε 471.820 επί τον αριθμό των components.

Στη συνέχεια, παρουσιάζουμε τα τελικά επίπεδα για κάθε αρχιτεκτονική μαζί με τον αριθμό των εξαγόμενων χαρακτηριστικών. Επίσης περιλαμβάνονται τα διαγράμματα του cumulative explained variance μετά την εφαρμογή του PCA, τα οποία δείχνουν ποιο ποσοστό των αρχικών χαρακτηριστικών παραμένει μετά τη μείωση των διαστάσεων σε σχέση με τον αριθμό των components.

Model	Top-1 Accuracy	Top-5 Accuracy
VGG16	71.3%	90.1%
VGG19	71.3%	90.0%
Xception	79.0%	94.5%
InceptionResNetV2	80.3%	95.3%
EfficientNetB7	84.3%	97.0%

Πίνακας 4.1: *Accuracy on ImageNet*

4.1.1 VGG16

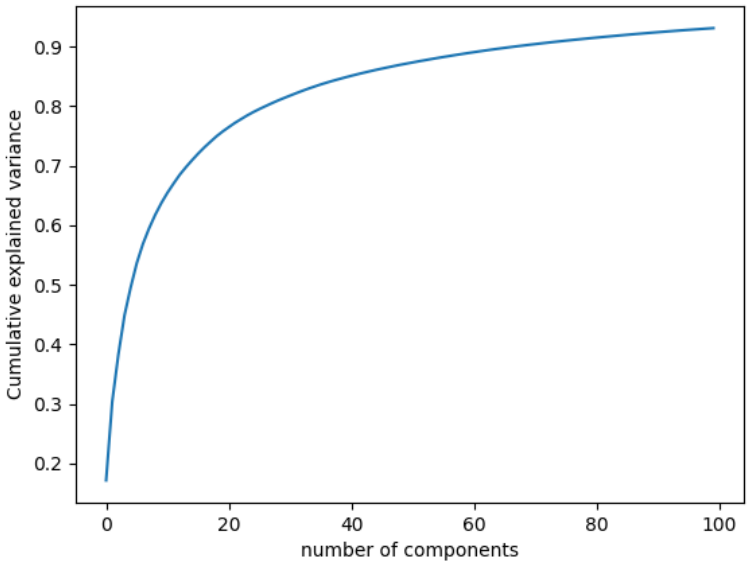
Ξεκινήσαμε τα πειράματα μας με ένα πολύ συνηθισμένο μοντέλο για εξαγωγή χαρακτηριστικών, το VGG16. Αποτελεί ένα από τα πιο βασικά και αποτελεσματικά μοντέλα, λόγω

του μικρού αριθμού υπερπαραμέτρων και της μεγάλης του αποδοτικότητας. Συνοπτικά, διαθέτει 16 επίπεδα (13 convolutional layers, 3 fully connected, 3 max pooling και 1 softmax layer). Στην υλοποίηση μας αφαιρέσαμε τα τελευταία 3 Dense layers και προσθέσαμε το global average pooling layer όπως φαίνεται παρακάτω. Ενώ ο αρχικός αριθμός των εξαγώμενων χαρακτηριστικών για κάθε εικόνα θα ήταν (8x8x512) = 32.768, με τη πρόσθεση του global average pooling μειώνονται στα 512.

block5_conv1 (Conv2D)	(None, 16, 16, 512)	2359808
block5_conv2 (Conv2D)	(None, 16, 16, 512)	2359808
block5_conv3 (Conv2D)	(None, 16, 16, 512)	2359808
block5_pool (MaxPooling2D)	(None, 8, 8, 512)	0
global_average_pooling2d (Gl	(None, 512)	0

Σχήμα 4.1: *VGG16 Final Architecture (top layers)*

Στη συνέχεια, ο πίνακας με τα χαρακτηριστικά για κάθε εικόνα αποτέλεσε την είσοδο του PCA. Δοκιμάσαμε διαφορετικούς αριθμούς για το πλήθος των components, καταλήγοντας στον αριθμό των 100 όπως βλέπουμε και στο παρακάτω διάγραμμα. Στόχος μας ήταν να σταθεροποιηθεί το cumulative explained variance πλησιέστερα στο 1, με το ελάχιστο αριθμό components. Θέλουμε ο αριθμός αυτός να είναι όσο των δυνατών μικρότερος αρχικά για την εξοικονόμηση μνήμης και έπειτα για τον ευκολότερο υπολογισμό των αποστάσεων μεταξύ των εικόνων στο επόμενο βήμα, το Nearest Neighbors.



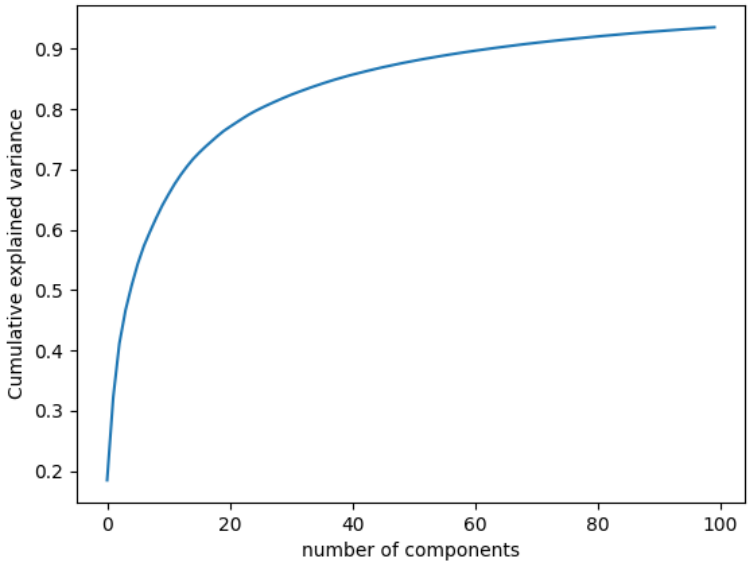
Σχήμα 4.2: *cumulative explained variance VGG16*

4.1.2 VGG19

Το VGG19 έχει παρόμοια υλοποίηση με το προηγούμενο μοντέλο μας, με τη διαφορά πως διαθέτει 19 επίπεδα (16 convolution layers, 3 Fully connected layer, 5 MaxPool layers και 1 SoftMax layer). Για την προσαρμογή του στην υλοποίηση μας ακολουθήσαμε τις ίδιες μεθόδους με παραπάνω, αφού ο αριθμός των εξαγώμενων χαρακτηριστικών ήταν κοινός.

block4_conv4 (Conv2D)	(None, 32, 32, 512)	2359808
block4_pool (MaxPooling2D)	(None, 16, 16, 512)	0
block5_conv1 (Conv2D)	(None, 16, 16, 512)	2359808
block5_conv2 (Conv2D)	(None, 16, 16, 512)	2359808
block5_conv3 (Conv2D)	(None, 16, 16, 512)	2359808
block5_conv4 (Conv2D)	(None, 16, 16, 512)	2359808
block5_pool (MaxPooling2D)	(None, 8, 8, 512)	0
global_average_pooling2d (Gl	(None, 512)	0

Σχήμα 4.3: *VGG19 Final Architecture (top layers)*



Σχήμα 4.4: *cumulative explained variance VGG19*

4.1.3 Xception

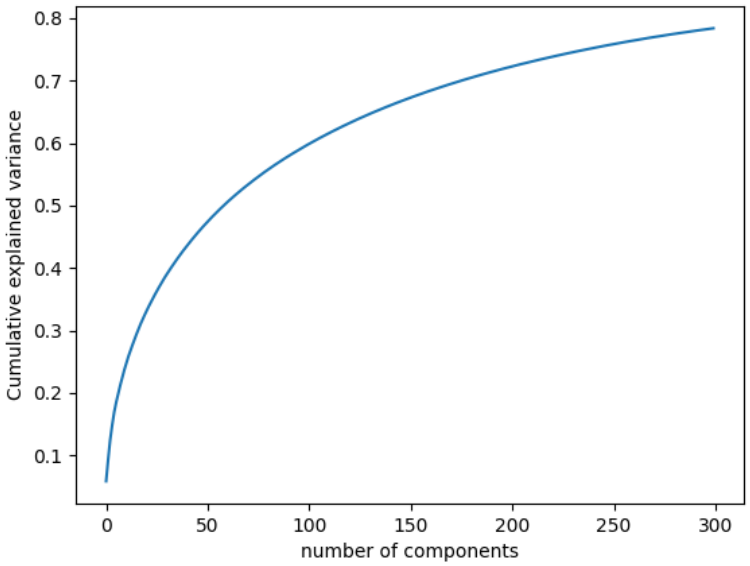
Το Xception βασίζεται αποκλειστικά στα βαθιά διαχωρίσιμα συνελικτικά επίπεδα. Η βαθιά διαχωριστική συνέλιξη προσφέρει τη δυνατότητα να διαμοιραστεί μία συνέλιξη σε δύο ή περισσότερες συνελίξεις παράγοντας το ίδιο αποτέλεσμα, κρατώντας το βάθος της εικόνας σταθερό. Για παράδειγμα, θεωρώντας μία εικόνα εισόδου με μέγεθος 7x7x3, μπορούμε να εφαρμόσουμε συνέλιξη με τρεις πυρήνες διαστάσεων 3x3x1. Κάθε πυρήνας εφαρμόζεται σε ένα κανάλι μόνο της εικόνας εξάγοντας ένα ενδιάμεσο αποτέλεσμα ίσο με 5x5x1, τα οποία στο τέλος συνθέτουν το τελικό αποτέλεσμα ίσο με 5x5x3. Η μέθοδος αυτή είναι ιδιαίτερα χρήσιμη, αφού ιδανικά μειώνει τον χρόνο υπολογισμών. Η αρχιτεκτονική του Xception περιλαμβάνει 36 συνελικτικά επίπεδα σχηματίζοντας τη βάση εξαγωγής χαρακτηριστικών του δικτύου ενώ χρησιμοποιεί γραμμικές υπολειπόμενες συνδέσεις (Residual Connections), όπως είδαμε και στο κεφάλαιο 2.

Τα αρχικά εξαγόμενα χαρακτηριστικά είχαν μέγεθος (8x8x2048) = 131.072 τα οποία μειώθηκαν σε 2048.

block14_sepconv1 (SeparableConv (None, 8, 8, 1536)	1582080	add_11[0][0]
block14_sepconv1_bn (BatchNorma (None, 8, 8, 1536)	6144	block14_sepconv1[0][0]
block14_sepconv1_act (Activatio (None, 8, 8, 1536)	0	block14_sepconv1_bn[0][0]
block14_sepconv2 (SeparableConv (None, 8, 8, 2048)	3159552	block14_sepconv1_act[0][0]
block14_sepconv2_bn (BatchNorma (None, 8, 8, 2048)	8192	block14_sepconv2[0][0]
block14_sepconv2_act (Activatio (None, 8, 8, 2048)	0	block14_sepconv2_bn[0][0]
global_average_pooling2d (Globa (None, 2048)	0	block14_sepconv2_act[0][0]

Σχήμα 4.5: *Xception Final Architecture*

Λόγο του σχετικά μεγάλου αριθμού χαρακτηριστικών, ο αριθμός των components για το PCA ήταν μεγαλύτερος από ότι στα προηγούμενα και ίσος με 300.



Σχήμα 4.6: *cumulative explained variance Xception*

4.1.4 Inception-ResNet-v2

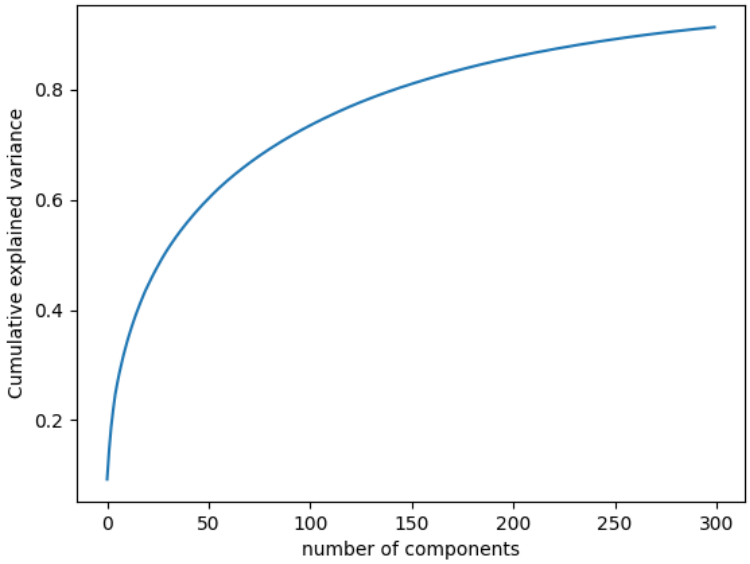
Το Inception-ResNet-v2 είναι μία αρχιτεκτονική συνελικτικών νευρωνικών δικτύων που βασίζεται σε αυτή του Inception περιλαμβάνοντας, επίσης, υπολειπόμενες συνδέσεις. Διαθέτει 164 επίπεδα και έχει εκπαιδευτεί σε πάνω από 1 εκατομμύριο εικόνες του Imagenet, με αποτέλεσμα να εξάγει χαρακτηριστικά για ένα ευρύ φάσμα διαφορετικών εικόνων.

Όπως βλέπουμε στο σχήμα παρακάτω, ο αριθμός των χαρακτηριστικών που εξάγονται τελικά μειώνεται από 55.296 σε 1.536.

block8_10_mixed (Concatenate)	(None, 6, 6, 448)	0	activation_199[0][0] activation_202[0][0]
block8_10_conv (Conv2D)	(None, 6, 6, 2080)	933920	block8_10_mixed[0][0]
block8_10 (Lambda)	(None, 6, 6, 2080)	0	block8_9_ac[0][0] block8_10_conv[0][0]
conv_7b (Conv2D)	(None, 6, 6, 1536)	3194880	block8_10[0][0]
conv_7b_bn (BatchNormalization)	(None, 6, 6, 1536)	4608	conv_7b[0][0]
conv_7b_ac (Activation)	(None, 6, 6, 1536)	0	conv_7b_bn[0][0]
global_average_pooling2d (Globa	(None, 1536)	0	conv_7b_ac[0][0]

Σχήμα 4.7: *InceptionResNetV2 Final Architecture*

Για το PCA χρειάστηκε αντίστοιχα ένας μεγάλος αριθμός components ίσος με 300.



Σχήμα 4.8: *cumulative explained variance InceptionResNetV2*

4.1.5 EfficientNet B7

Τα EfficientNets είναι μία οικογένεια μοντέλων που πετυχαίνει καλύτερα αποτελέσματα σε accuracy και efficiency από ότι οποιοδήποτε άλλο προηγούμε συνελικτικό δίκτυο. Συγκεκριμένα, το EfficientNet B7 πετυχαίνει τα υψηλότερα ποσοστά ενώ είναι 8.2 φορές μικρότερο και 6.1 φορές ταχύτερο σε σχέση με το προηγούμενο σε αποτελεσματικότητα συνελικτικό δίκτυο [48]. Η υλοποίηση του βασίζεται στην αλλαγή των διαστάσεων του βάθους/πλάτους/ανάλυσης των δικτύων. Το EfficientNet B7 δέχεται ιδανικά σαν μέγεθος εικόνας 600x600, νούμερο αρκετά μεγαλύτερο από ότι απαιτούν όλα τα προηγούμενα. Ακόμη, πρέπει να σημειώσουμε πως για το συγκεκριμένο μοντέλο δεν έχει προηγηθεί κανονικοποίηση των τιμών (διαιρώντας τις τιμές των pixels με το 255) των εισαγόμενων εικόνων αφού είναι μία διαδικασία που πραγματοποιεί το ίδιο.

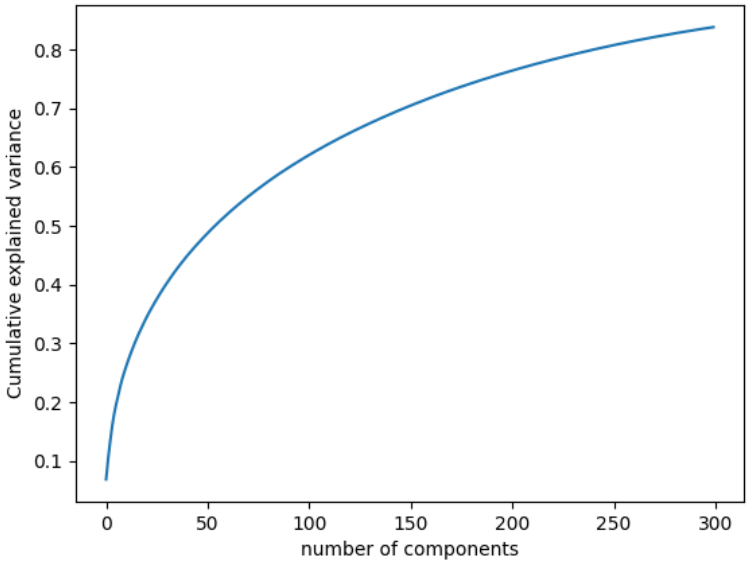
Βλέπουμε πως ο αριθμός των χαρακτηριστικών κάθε εικόνας χωρίς την εφαρμογή του

global average pooling θα ήταν ίσος με 163.840, νούμερο ιδιαίτερα μεγάλο. Με την εφαρμογή του επιπρόθετου επιπέδου, ο αριθμός μειώνεται σε 2.560.

block7d_add (Add)	(None, 8, 8, 640)	0	block7d_drop[0][0] block7c_add[0][0]
top_conv (Conv2D)	(None, 8, 8, 2560)	1638400	block7d_add[0][0]
top_bn (BatchNormalization)	(None, 8, 8, 2560)	10240	top_conv[0][0]
top_activation (Activation)	(None, 8, 8, 2560)	0	top_bn[0][0]
avg_pool (GlobalAveragePooling2)	(None, 2560)	0	top_activation[0][0]

Σχήμα 4.9: *EfficientNetB7 Final Architecture*

Για την εφαρμογή του PCA, παρόμοια με τα προηγούμενα ο αριθμός των components είναι ίσος με 300 φτάνοντας το cumulative explained variance ίσο με 0.8.

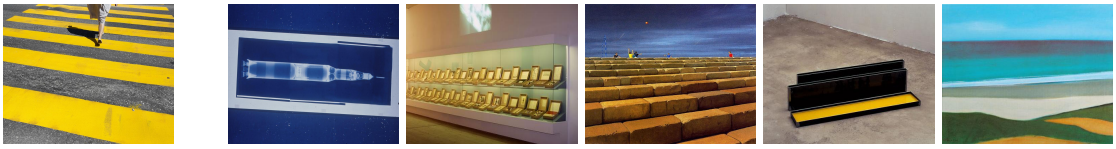


Σχήμα 4.10: *cumulative explained variance EfficientNetB7*

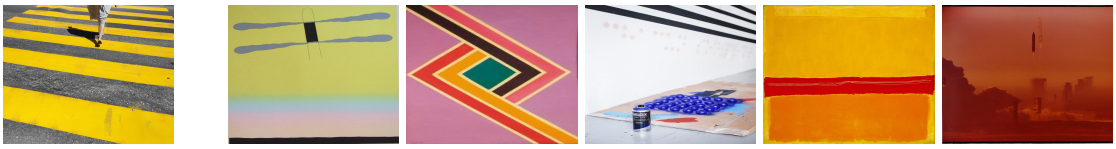
4.2 Σύκριση Αποτελεσμάτων

Μία δυσκολία που αντιμετωπίζει η μέθοδος μας είναι πως δεν υπάρχει κάποιο αντικειμενικό κριτήριο για τη σύγκριση των αποτελεσμάτων που δίνουν τα διαφορετικά μοντέλα λόγω της απουσίας ετικετών. Στο σημείο αυτό, προηγούμενες μελέτες που ακολουθούν την ίδια μέθοδο [15], ζήτησαν τη γνώμη των ειδικών για την απόδοση της αποτελεσματικότητας του συστήματος. Στην εργασία μας, παρουσιάζουμε για διαφορετικές εικόνες τα 5 πρώτα αποτελέσματα που έδωσε κάθε μοντέλο. Οι εικόνες αυτές ανήκουν σε διαφορετικές θεματολογίες(πορταίτα, φύση, γεγονότα, γεωμετρικά σχέδια). Με αυτόν τον τρόπο, ο αναγνώστης μπορεί να παρατηρήσει ποιό μοντέλο δίνει καλύτερα αποτελέσματα στις διαφορετικές ενότητες.

Query Results



Σχήμα 4.11: *VGG16*



Σχήμα 4.12: *VGG19*

Σχήμα 4.13: *Xception*

Σχήμα 4.14: *InceptionResNetV2*

Σχήμα 4.15: *EfficientNetB7*

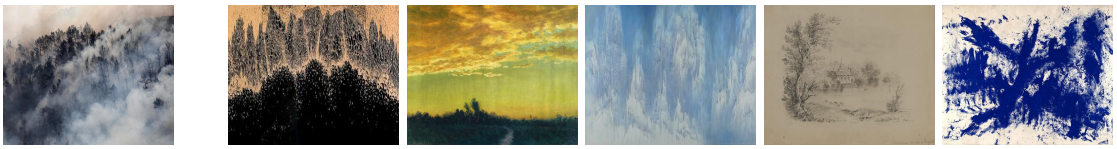
Query Results

Σχήμα 4.16: *VGG16*

Σχήμα 4.17: *VGG19*

Σχήμα 4.18: *Xception*

Σχήμα 4.19: *InceptionResNetV2*



Σχήμα 4.20: *EfficientNetB7*

Query Results

Σχήμα 4.21: *VGG16*

Σχήμα 4.22: *VGG19*

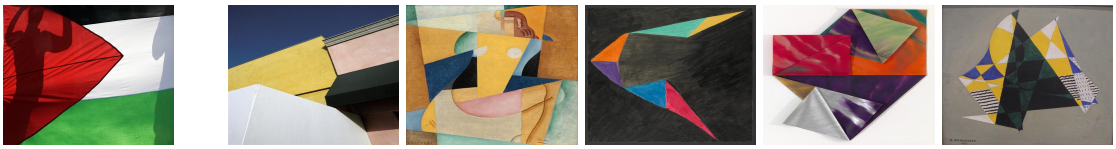
Σχήμα 4.23: *Xception*

Σχήμα 4.24: *InceptionResNetV2*

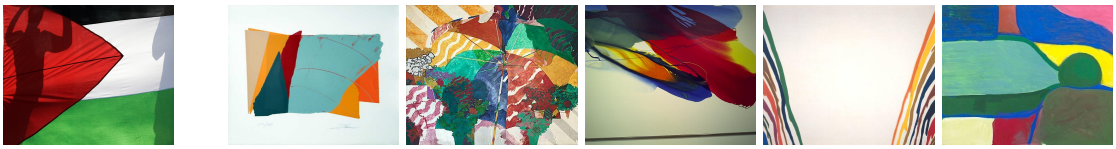
Σχήμα 4.25: *EfficientNetB7*

Query Results

Σχήμα 4.26: *VGG16*



Σχήμα 4.27: *VGG19*



Σχήμα 4.28: *Xception*

Σχήμα 4.29: *InceptionResNetV2*

Σχήμα 4.30: *EfficientNetB7*

Query Results

Σχήμα 4.31: *VGG16*

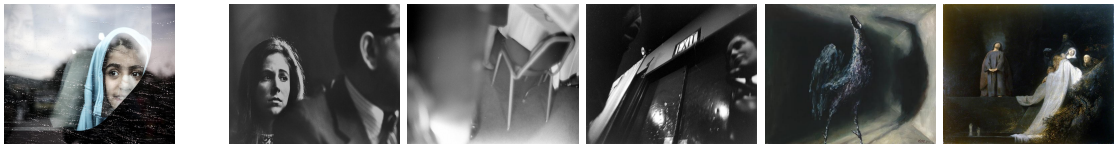
Σχήμα 4.32: *VGG19*

Σχήμα 4.33: *Xception*

Σχήμα 4.34: *InceptionResNetV2*

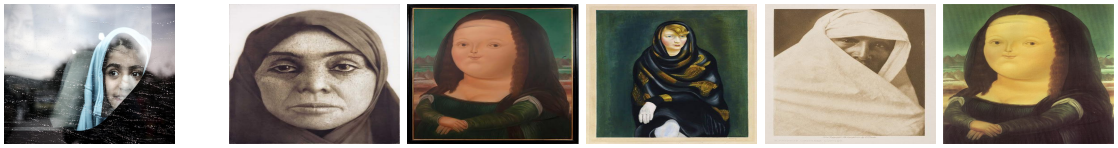
Σχήμα 4.35: *EfficientNetB7*

Query Results



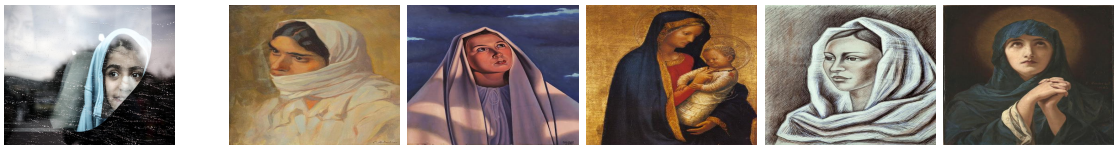
Σχήμα 4.36: *VGG16*

Σχήμα 4.37: *VGG19*



Σχήμα 4.38: *Xception*

Σχήμα 4.39: *InceptionResNetV2*



Σχήμα 4.40: *EfficientNetb07*

Query Results

Σχήμα 4.41: *VGG16*

Σχήμα 4.42: *VGG19*

Σχήμα 4.43: *Xception*

Σχήμα 4.44: *Resnet-Inception*

Σχήμα 4.45: *EfficientNetb07*

4.3 Σχολιασμός Αποτελεσμάτων

Σύμφωνα με τα παραπάνω αποτελέσματα, είμαστε σε θέση να εξάγουμε κάποια βασικά συμπεράσματα για την απόδοση των μοντέλων. Αρχικά, όλα τα μοντέλα δώσανε πολύ καλά αποτελέσματα και κάθε ένα από αυτά θα μπορούσε να χρησιμοποιηθεί για τη συγκεκριμένη διαδικασία. Αν μας ενδιέφερε κυρίως η ταχύτητα του προγράμματος θα επιλέγαμε κάποιο ανάμεσα στα VGG16 και VGG19 αφού ήταν περίπου 6 φορές ταχύτερα από τα υπόλοιπα. Παρόλα αυτά, η απόδοση τους δεν ήταν αντίστοιχη των υπολοίπων. Παρατηρούμε πως στον εντοπισμό προσώπων και την αντιστοίχιση τους με πορτραίτα, τα Xception, InceptionResNetv2 και EfficientNetB7 δίνουν σαφώς καλύτερα αποτελέσματα, αφού είναι σε θέση να αναγνωρίσουν ακόμη και λεπτομέρειες που απαρτίζουν τις εικόνες. Ακόμη, βλέπουμε πως τα συγκεκριμένα μοντέλα είναι σε θέση να εντοπίσουν πιο δύσκολα στοιχεία όπως ζώα και να κατανοήσουν καλύτερα το περιεχόμενο των εικόνων όταν αφορά γεγονότα. Στον εντοπισμό σχημάτων και χρωματικών αποχρώσεων όλα τα μοντέλα ανταποκρίνονται εξίσου ικανοποιητικά. Το EfficientNetB7 φαίνεται να είναι εκείνο που μπορεί να αντιστοιχίσει καλύτερα το περιεχόμενο της εικόνας, γεγονός αναμενόμενο λόγω της υψηλής απόδοσης του. Εν κατακλείδι, η επιλογή του μοντέλου για την υλοποίηση της συγκεκριμένης μεθόδου έγκειται στα δεδομένα που περιλαμβάνει κάθε σύνολο δεδομένων και στις ανάγκες που παρουσιάζει κάθε υλοποίηση.

4.4 Μεμονομένες Περιπτώσεις

Στο τελευταίο κομμάτι της εργασίας, παραθέτουμε κάποια από τα αποτελέσματα που θεωρούμε άξια αναφοράς λόγω της ομοιότητας τους. Για κάθε εικόνα υπάρχει δίπλα ένας πίνακας από τους 10 που επιστρέφει το πρόγραμμα μας για κάθε query (όχι αναγκαστικά το 1ο αποτέλεσμα).

4.4.1 VGG16

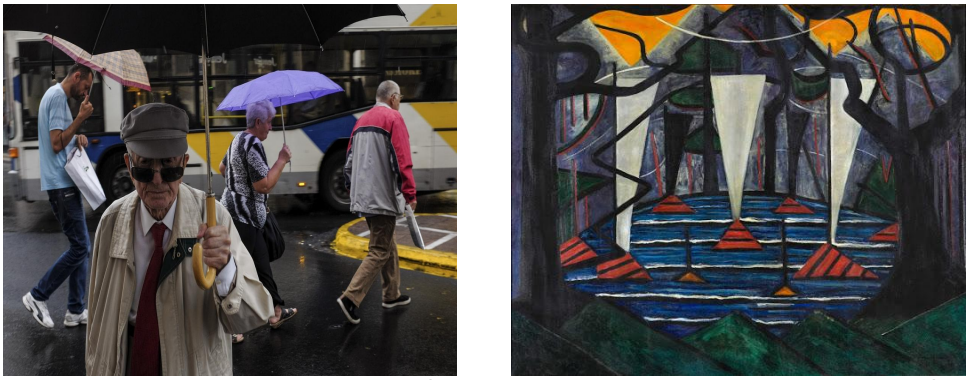
Σχήμα 4.46: *Αριστερά: Nick Paleologos/SOOC, Δεξιά: Jackson Pollock/Mural On Indian Red Ground 1950*

Σχήμα 4.47: *Αριστερά: Odysseas Chloridis/SOOC, Δεξιά: Edvard Munch/Friedrich Nietzsche 1906*

4.4.2 VGG19

Σχήμα 4.48: *Αριστερά: Alexandros Michailidis/SOOC, Δεξιά: David Wilkie/The First Council Of Queen Victoria 1838*

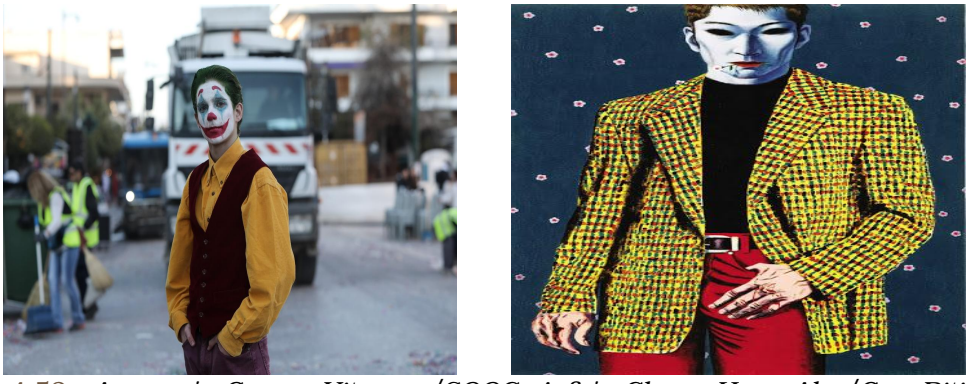
Σχήμα 4.49: *Αριστερά: Alexandros Michailidis/SOOC, Δεξιά: Claude Monet/Water Lilies 1899 2*



Σχήμα 4.50: *Αριστερά: Nick Paleologos/SOOC, Δεξιά: Jacoba van Heemskerck/Composition No 23*

Σχήμα 4.51: *Αριστερά: Nick Paleologos/SOOC, Δεξιά: William Turner/Stonehenge Twiligh*

4.4.3 Xception



Σχήμα 4.52: *Αριστερά: George Vitsaras/SOOC, Δεξιά: Chang Hong Ahn/Guy Biting The Flower*


Σχήμα 4.53: *Αριστερά: George Vitsaras/SOOC, Δεξιά: Joaquin Sorolla/On The Sand Valencia Beach 1908*

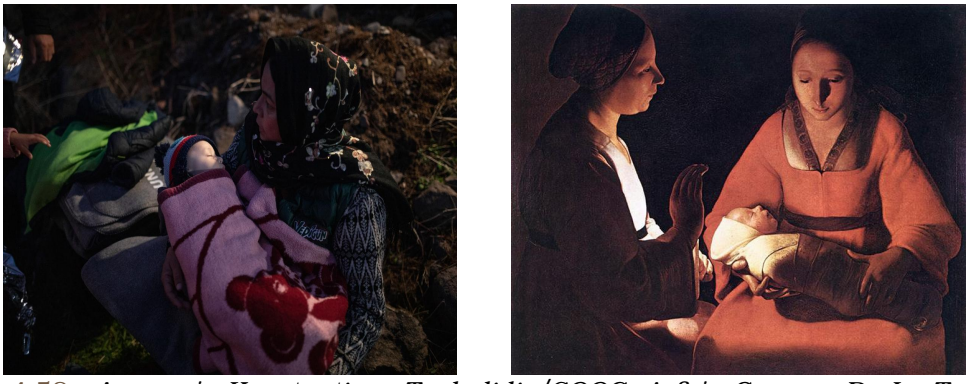

Σχήμα 4.54: *Αριστερά: Nikos Libertas/SOOC, Δεξιά: Clyfford Still/Ph 118*


Σχήμα 4.55: *Αριστερά: Nick Paleologos/SOOC, Δεξιά: Henry William Banks Davis/A Gleamy Day In Picardy 1900*

4.4.4 InceptionResnetV2


Σχήμα 4.56: *Αριστερά: Konstantinos Tsakalidis/SOOC, Δεξιά: Thomas Gainsborough/Edward 2Nd Viscount Ligonier 1770*

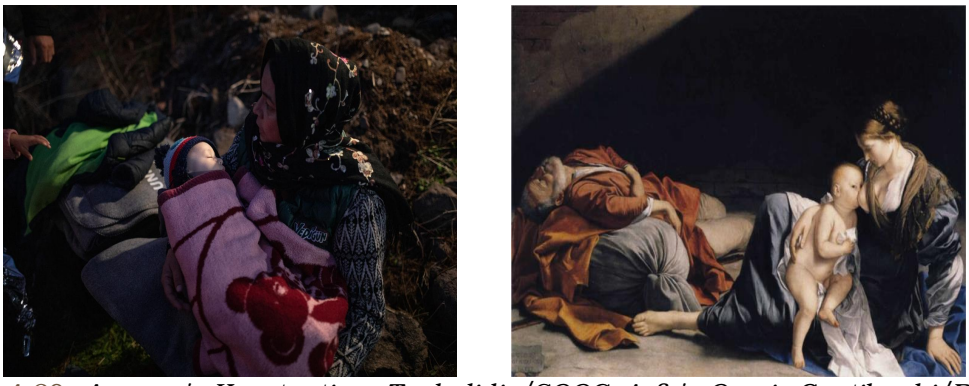
Σχήμα 4.57: *Αριστερά: Alexandros Michailidis/SOOC, Δεξιά: Felice Giani/Triumphal Arch In Ponte Sant Angelo Built For The Liberation Celebration*



Σχήμα 4.58: *Αριστερά: Konstantinos Tsakalidis/SOOC, Δεξιά: Georges De La Tour/The Newborn*

4.4.5 EfficientNetB7

Σχήμα 4.59: *Αριστερά: George Vitsaras/SOOC, Δεξιά: Mykola Pymonenko/A Girl*



Σχήμα 4.60: *Αριστερά: Konstantinos Tsakalidis/SOOC, Δεξιά: Orazio Gentileschi/Rest On The Flight To Egypt 1628*

Μέρος III

Επίλογος

Κεφάλαιο 5

Επίλογος

5.1 Σύνοψη Εργασίας

Στην παρούσα διπλωματική εργασία παρουσιάσαμε τη μέθοδο που ακολουθήσαμε για την αναζήτηση συσχετίσεων ανάμεσα σε φωτογραφίες και πίνακες ζωγραφικής. Για την υλοποίηση της συγκεκριμένης μεθόδου, αρχικά πραγματοποιήσαμε μία εκτενή μελέτη πάνω στα συνελικτικά νευρωνικά δίκτυα και τον τρόπο εξαγωγής χαρακτηριστικών από εικόνες. Πιο συγκεκριμένα, χρησιμοποιήσαμε, μέσω μεταφορά γνώσης, πέντε διαφορετικά προεκπαιδευμένα μοντέλα συνελικτικών νευρωνικών δικτύων : τα VGG16, VGG19, Xception, Inception-ResNet v2 και το EfficientNet B7. Στα μοντέλα, εισάγαμε σαν είσοδο τις εικόνες από τα δύο σύνολα δεδομένων που χρησιμοποιήσαμε, το ART500k για έργα τέχνης και τις φωτογραφίες από το πρακτορείο SOOC, ύστερα από την κατάλληλη προ-επεξεργασία. Σε κάθε μοντέλο, αφαιρέσαμε τα τελικά πλήρως συνδεδεμένα επίπεδα και προσθέσαμε ένα επίπεδο Global Average Pooling Layer για τη μείωση των διαστάσεων των τελικών εξαγόμενων χαρακτηριστικών για κάθε εικόνα. Στη συνέχεια, χρησιμοποιήσαμε τη μέθοδο PCA για την περαιτέρω μείωση των διαστάσεων και την επιτάχυνση του τελικού προγράμματος. Οι τελικοί πίνακες χαρακτηριστικών (feature maps) αποτέλεσαν την είσοδο της μη επιβλεπόμενης αναζήτησης κοντινότερων γειτόνων (Nearest Neighbor). Με αυτόν τον τρόπο, για κάθε φωτογραφία που θέσαμε σαν είσοδο, λάβαμε τους πέντε πλησιέστερους πίνακες μέσα από ένα σύνολο δεδομένων με πάνω από 471.000 έργα τέχνης. Στη συνέχεια παρουσιάσαμε τα αποτελέσματα που πήραμε από κάθε μοντέλο, συγκρίνοντας την απόδοση τους για εικόνες με διαφορετικές θεματολογίες. Με αυτόν τον τρόπο, μπορέσαμε να αξιολογήσουμε ποιό μοντέλο είναι πιο αποδοτικό σε κάθε περίπτωση μελετώντας τόσο την ακρίβεια των αποτελεσμάτων όσο και τον χρόνο που απαιτούσε κάθε περίπτωση.

5.2 Τελικά Συμπεράσματα και Μελλοντικές Επεκτάσεις

Μέχρι πρότινος, η διαδικασία σύγκρισης και εύρεσης συσχετίσεων σε έργα τέχνης ήταν έργο ειδικά καταρτισμένων ειδικών που απαιτούσε πολύωρες μελέτες και χρονοβόρες προσπάθειες για τη συλλογή δεδομένων. Στη μέθοδο που προτείνουμε, η διαδικασία αυτή μπορεί να γίνει γρήγορα και χωρίς την παρέμβαση του ανθρώπινου παράγοντα χρησιμοποιώντας σύνολα δεδομένων από ψηφιοποιημένα έργα τέχνης και εφαρμόζοντας μεθόδους από τον χώρο της μηχανικής μάθησης. Φυσικά, τα αποτελέσματα που εξάγει η τεχνητή νοημο-

σύνη δεν είναι πάντα ακριβή, ειδικά σε περιπτώσεις όπως η τέχνη που το περιεχόμενο των δεδομένων είναι δύσκολο να γίνει κατανοητό ακόμα και από τους ανθρώπους. Παρόλα αυτά, η προτεινόμενη υλοποίηση μπορεί να σταθεί αρωγός στη μελέτη των ειδικών και να φέρει την τέχνη ακόμη πιο κοντά κοντά στο ευρύ κοινό.

Πιο συγκεκριμένα, η μέθοδος που ακολουθήσαμε μπορεί να εφαρμοστεί αρχικά σε ιστοσελίδες μουσείων ή ακόμα και στους φυσικούς χώρους, καλώντας τους επισκέπτες να εισάγουν δικές τους φωτογραφίες. Ακόμα, μπορεί να παροτρύνει νέους φοιτητές να ασχοληθούν με την εφαρμογή της τεχνητής νοημοσύνης στην τέχνη, βρίσκοντας νέες μεθόδους και εξελίσσοντας τις ήδη υπάρχουσες.

Επιπλέον, μπορεί να βοηθήσει στη μελέτη και κατανόηση των συνδέσεων που υπάρχουν ανάμεσα στην τέχνη της φωτογραφίας και της ζωγραφικής από κριτικούς τέχνης. Αν και το σύνολο φωτογραφιών που θέσαμε σαν είσοδο είναι σχετικά μικρό, για τις περισσότερες φωτογραφίες τα μοντέλα παρουσίασαν πίνακες με φανερές ομοιότητες. Αυτό μας δείχνει, πως παρά το πέρασμα των αιώνων, κάποια στοιχεία στις εικαστικές τέχνες παραμένουν αναλλοίωτα. Είναι άξιο παρατήρησης ο τρόπος με τον οποίο ο άνθρωπος αποτυπώνει τον πόνο, τη χαρά, τον πόλεμο, τη μητρότητα, τη φύση και πολλά ακόμη θέματα, διαχρονικά, ανεξάρτητα του μέσου που επιλέγει.

Πέρα από τη σύνδεση φωτογραφιών με πίνακες, η μέθοδος που προτείνεται μπορεί να χρησιμοποιηθεί ακόμα και σε περιπτώσεις εύρεσης αντιγραφής ενός έργου. Θέτοντας έναν πίνακα ή μία φωτογραφία ως query, με τα κατάλληλα σύνολα δεδομένων, το πρόγραμμα μπορεί να εμφανίσει προγενέστερα έργα από τον ίδιο χώρο, με έντονες ομοιότητες, οι οποίες να αποτελούν ένδειξη αντιγραφής.

Κάποιες βελτιώσεις που θα μπορούσαν να γίνουν σε μετέπειτα μελέτες, θα ήταν αρχικά η χρήση ενός μεγαλύτερου συνόλου φωτογραφιών για τη την εξαγωγή περισσότερων αποτελεσμάτων. Επιπλέον, θα μπορούσαν να αφαιρεθούν οι εικόνες που δεν αφορούν πίνακες ζωγραφικής από το σύνολο δεδομένων ART500k ώστε τα αποτελέσματα να περιορίζονται μόνο σε αυτόν τον τομέα. Τέλος, η σημαντικότερη κατά τη γνώμη μας βελτίωση, θα ήταν να βρεθεί ένας τρόπος για την αντικειμενική αξιολόγηση της απόδοσης των πειραμάτων.

Bibliography

[1] Seguin, Benoit and Striolo, Carlotta and diLenardo, Isabella and Kaplan, Frederic, "Visual Link Retrieval in a Database of Paintings," in *Computer Vision - ECCV 2016 Workshops*, Hua, Gang and Jégou, Hervé, Ed. Cham: Springer International Publishing, 2016, pp. 753–767.

[2] The Met Museums, "Two ways of life 1857, printed 1920s," 1857, printed 1920s, [Online; accessed March 21, 2022]. [Online]. Available: https://www.metmuseum.org/art/collection/search/294822

[3] Khacademy, "Raphael, school of athens," [Online; accessed March 21, 2022]. [Online]. Available: https://www.khanacademy.org/humanities/ap-art-history/early-europe-and-colonial-americas/renaissance-art-europe-ap/a/raphael-school-of-athens

[4] International Visual Art, "The apotheosis of degas - photo," [On flickr].

[5] Wikipedia, the free encyclopedia, "Jean auguste dominique ingres, apotheosis of homer," 1827, [Online; accessed March 20, 2022]. [Online]. Available: https://commons.wikimedia.org/wiki/File:Jean_Auguste_Dominique_Ingres,_Apotheosis_of_Homer,_1827.jpg

[6] J. Hoła and K. Schabowicz, "Application of artificial neural networks to determine concrete compressive strength based on non-destructive tests," *JOURNAL OF CIVIL ENGINEERING AND MANAGEMENT*, vol. 11, pp. 23–32, 03 2005.

[7] Towards Data Science, "Activation functions in neural networks," 2017, [Online; accessed March 21, 2022]. [Online]. Available: https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6

[8] I. Gogul and S. Kumar, "Flower species recognition system using convolution neural networks and transfer learning," 03 2017, pp. 1–6.

[9] edge ai + vision alliance, "What's the difference between a cnn and an rnn?" 2013, [Online; accessed March 20, 2022]. [Online]. Available: https://www.edge-ai-vision.com/2018/09/whats-the-difference-between-a-cnn-and-an-rnn/

[10] H. Yingge, I. Ali, and K.-Y. Lee, "Deep neural networks on chip - a survey," 02 2020, pp. 589–592.

[11] K. Prilianti, T. Brotosudarmo, S. Anam, and A. Suryanto, "Performance comparison of the convolutional neural network optimizer for photosynthetic pigments prediction on plant digital image," vol. 2084, 03 2019.

[12] M. Ferguson, R. ak, Y.-T. Lee, and K. Law, "Automatic localization of casting defects with convolutional neural networks," 12 2017, pp. 1726–1735.

[13] Neurohive, "Residual cnns for image classification tasks," 2019, [Online; accessed March 20, 2022]. [Online]. Available: https://neurohive.io/en/popular-networks/resnet/

[14] H. Mao, J. She, and M. Cheung, "Visual arts search on mobile devices," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 15, no. 2s, p. 60, 2019.

[15] G. Castellano, E. Lella, and G. Vessio, "Visual link retrieval and knowledge discovery in painting datasets," *Multimedia Tools and Applications*, vol. 80, no. 5, p. 6599–6616, Oct 2020. [Online]. Available: http://dx.doi.org/10.1007/s11042-020-09995-z

[16] L. Shamir, T. J. Macura, N. V. Orlov, D. M. Eckley, and I. G. Goldberg, "Impressionism, expressionism, surrealism: Automated recognition of painters and schools of art," *ACM Trans. Appl. Percept.*, vol. 7, pp. 8:1–8:17, 2010.

[17] C. Sandoval, E. Pirogova, and M. Lech, "Two-stage deep learning approach to the classification of fine-art paintings," *IEEE Access*, vol. 7, pp. 41 770–41 781, 2019.

[18] D. Keren, "Painter identification using local features and naive bayes," *Object recognition supported by user interaction for service robots*, vol. 2, pp. 474–477 vol.2, 2002.

[19] E. Cetinic and S. Grgic, "Automated painter recognition based on image feature extraction," *Proceedings ELMAR-2013*, pp. 19–22, 2013.

[20] S. Agarwal, H. Karnick, N. Pant, and U. Patel, "Genre and style based painting classification," *2015 IEEE Winter Conference on Applications of Computer Vision*, pp. 588–594, 2015.

[21] X. Shen, A. Efros, and A. Mathieu, "Discovering visual patterns in art collections with spatially-consistent feature learning," 03 2019.

[22] P. Mylonas, E. Spyrou, Y. Avrithis, and S. Kollias, "Using visual context and region semantics for high-level concept detection," *IEEE Transactions on Multimedia*, vol. 11, no. 2, pp. 229–243, 2009.

[23] D. Kollias, Y. Vlaxos, M. Seferis, I. Kollia, L. Sukissian, J. Wingate, and S. Kollias, "Transparent adaptation in deep medical image diagnosis," in *International Workshop on the Foundations of Trustworthy AI Integrating Learning, Optimization and Reasoning*. Springer, 2020, pp. 251–267.

[24] D. Kollias, A. Arsenos, L. Soukissian, and S. Kollias, "Mia-cov19d: Covid-19 detection through 3-d chest ct image analysis," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 537–544.

[25] I. Kollia, A.-G. Stafylopatis, and S. Kollias, "Predicting parkinson's disease using latent information extracted from deep neural networks," in *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2019, pp. 1–8.

[26] F. Caliva, F. S. De Ribeiro, A. Mylonakis, C. Demazi'ere, P. Vinai, G. Leontidis, and S. Kollias, "A deep learning approach to anomaly detection in nuclear reactors," in *2018 International joint conference on neural networks (IJCNN)*. IEEE, 2018, pp. 1–8.

[27] Y. Avrithis, N. Tsapatsoulis, and S. Kollias, "Broadcast news parsing using visual cues: A robust face detection approach," in *In: Proceedings of IEEE International Conference on Multimedia and Expo (ICME 2000*, 2000, pp. 1469–1472.

[28] N. Tsapatsoulis and S. D. Kollias, "Face detection in color images and video sequences," 2000.

[29] G. Caridakis, A. Raouzaiou, K. Karpouzis, and S. Kollias, "Synthesizing gesture expressivity based on real sequences," in *Workshop Programme*, vol. 10, p. 19.

[30] ΞΑΝΘΑΚΗΣ ΑΛΚΗΣ, *ΙΣΤΟΡΙΑ ΤΗΣ ΦΩΤΟΓΡΑΦΙΚΗΣ ΑΙΣΘΗΤΙΚΗΣ 1839-1975*. ΑΙΓΟΚΕΡΩΣ, 1999.

[31] ΣΑΜΠΑΝΙΚΟΥ ΕΥΗ, *ΦΩΤΟΓΡΑΦΙΑ ΚΑΙ ΖΩΓΡΑΦΙΚΗ 19ος - 20ος ΑΙΩΝΑΣ*. ΤΥΠΩΘΗΤΩ / ΔΑΡΔΑΝΟΣ, 2003.

[32] C. Janiesch, P. Zschech, and K. Heinrich, "Machine learning and deep learning," *Electronic Markets*, vol. 31, no. 3, p. 685–695, Apr 2021. [Online]. Available: http://dx.doi.org/10.1007/s12525-021-00475-2

[33] A. L. Samuel, "Some studies in machine learning using the game of checkers," *IBM Journal of Research and Development*, vol. 3, no. 3, pp. 210–229, 1959.

[34] Cios K.J., Swiniarski R.W., Pedrycz W., Kurgan L.A. , *Unsupervised Learning: Association Rules. In: Data Mining*. Springer, Boston, MA, 2007.

[35] D. Dua and C. Graff, "UCI machine learning repository," 2017. [Online]. Available: http://archive.ics.uci.edu/ml

[36] B. Mehlig, "Machine learning with neural networks," Oct 2021. [Online]. Available: http://dx.doi.org/10.1017/9781108860604

[37] W. S. Mcculloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Journal of Symbolic Logic*, vol. 9, no. 2, pp. 49–50, 1943.

[38] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale

Visual Recognition Challenge," *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211–252, 2015.

[39] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain." *Psychological Review*, vol. 65, no. 6, pp. 386–408, 1958. [Online]. Available: http://dx.doi.org/10.1037/h0042519

[40] Z. Li, W. Yang, S. Peng, and F. Liu, "A survey of convolutional neural networks: Analysis, applications, and prospects," 2020.

[41] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," 2015.

[42] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. [Online]. Available: http://www.deeplearningbook.org

[43] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015.

[44] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015.

[45] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," 2014. [Online]. Available: https://arxiv.org/abs/1409.4842

[46] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," 2016. [Online]. Available: https://arxiv.org/abs/1602.07261

[47] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," 2016. [Online]. Available: https://arxiv.org/abs/1610.02357

[48] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," 2019. [Online]. Available: https://arxiv.org/abs/1905.11946

[49] A. Farahani, B. Pourshojae, K. Rasheed, and H. R. Arabnia, "A concise review of transfer learning," 2021.

[50] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A comprehensive survey on transfer learning," 2020.

[51] D. Kollias, A. Tagaris, A. Stafylopatis, S. Kollias, and G. Tagaris, "Deep neural architectures for prediction in healthcare," *Complex & Intelligent Systems*, vol. 4, no. 2, pp. 119–131, 2018.

[52] H. Mao, M. Cheung, and J. She, "Deepart: Learning joint representations of visual arts," in *Proceedings of the 25th ACM international conference on Multimedia*. ACM, 2017, pp. 1183–1191.

[53] F. Chollet *et al.* (2015) Keras. [Online]. Available: https://github.com/fchollet/keras