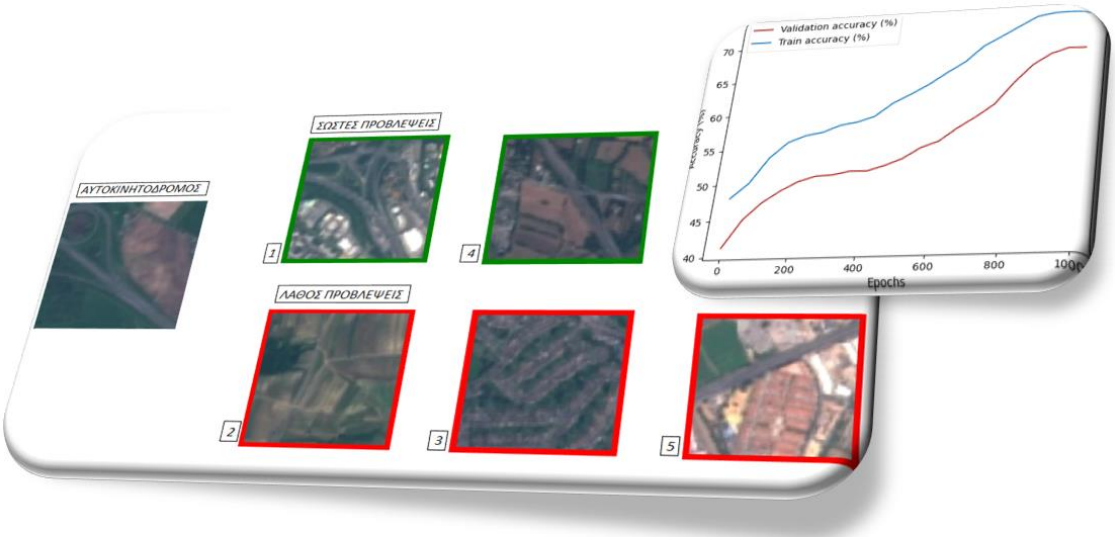
ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΑΓΡΟΝΟΜΩΝ ΤΟΠΟΓΡΑΦΩΝ ΜΗΧΑΝΙΚΩΝ – ΜΗΧΑΝΙΚΩΝ ΓΕΩΠΛΗΡΟΦΟΡΙΚΗΣ
ΕΡΓΑΣΤΗΡΙΟ ΤΗΛΕΠΙΣΚΟΠΗΣΗΣ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΜΕΛΕΤΗ ΚΑΙ ΑΞΙΟΛΟΓΗΣΗ ΜΕΘΟΔΩΝ ΑΥΤΟ-ΕΠΙΒΛΕΠΟΜΕΝΗΣ ΜΑΘΗΣΗΣ ΚΑΙ ΠΡΟΒΛΗΜΑΤΑ ΣΕ ΤΑΞΙΝΟΜΗΣΗ ΕΙΚΟΝΩΝ



ΓΩΓΟΥΣΟΥ ΙΩΑΝΝΑ

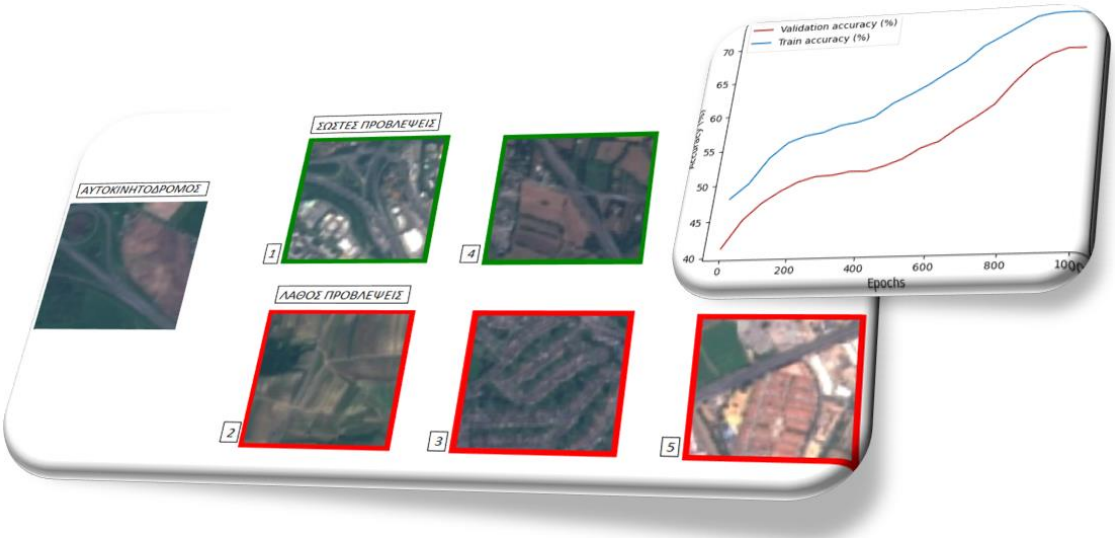
ΜΑΡΤΙΟΣ 2022

RSLab
Remote Sensing Laboratory
National Technical University of Athens

NATIONAL TECHNICAL UNIVERSITY OF ATHENS
SCHOOL OF RURAL, SURVEYING & GEOINFORMATICS ENGINEERING
REMOTE SENSING LABORATORY

DIPLOMA THESIS

STUDYING AND EVALUATING SELF-SUPERVISED LEARNING METHODS FOR IMAGE CLASSIFICATION TASKS



GOGOUSOU IOANNA

MARCH 2022

RSLab
Remote Sensing Laboratory
National Technical University of Athens

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΑΓΡΟΝΟΜΩΝ ΤΟΠΟΓΡΑΦΩΝ ΜΗΧΑΝΙΚΩΝ – ΜΗΧΑΝΙΚΩΝ ΓΕΩΠΛΗΡΟΦΟΡΙΚΗΣ
ΕΡΓΑΣΤΗΡΙΟ ΤΗΛΕΠΙΣΚΟΠΗΣΗΣ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΜΕΛΕΤΗ ΚΑΙ ΑΞΙΟΛΟΓΗΣΗ ΜΕΘΟΔΩΝ ΑΥΤΟ-ΕΠΙΒΛΕΠΟΜΕΝΗΣ ΜΑΘΗΣΗΣ ΚΑΙ ΠΡΟΒΛΗΜΑΤΑ ΣΕ ΤΑΞΙΝΟΜΗΣΗ ΕΙΚΟΝΩΝ

Τριμελής εξεταστική επιτροπή

Κ. Καράντζαλος	**Β. Καραθανάση**	**Ν. Δουλάμης**
……	……	……
Αν. Καθηγητής ΕΜΠ	**Καθηγήτρια ΕΜΠ**	**Αν. Καθηγητής ΕΜΠ**
Επιβλέπων		

Συνεπιβλέπων : Ψωμάς Βασίλειος
Υποψήφιος Διδάκτορας ΕΜΠ

ΑΘΗΝΑ, ΜΑΡΤΙΟΣ 2022

RSLab
Remote Sensing Laboratory
National Technical University of Athens

ΠΕΡΙΛΗΨΗ

Η ανάπτυξη και υλοποίηση αυτοματοποιημένων αλγορίθμων μηχανικής μάθησης για την ταξινόμηση εικόνων και αναγνώριση αντικειμένων είναι ένα πεδίο έρευνας που απασχολεί την επιστημονική κοινότητα για αρκετές δεκαετίες. Σχεδιάζονται και υλοποιούνται συνεχώς νέες μεθοδολογίες προκειμένου να βελτιωθούν οι αυτοματοποιημένες διαδικασίες κατανόησης του περιεχομένου των δεδομένων. Τα τελευταία χρόνια, προσεγγίσεις που βασίζονται σε αρχιτεκτονικές βαθιάς εκμάθησης (deep learning) έχουν ξεπεράσει σημαντικά τις ακρίβειες ανίχνευσης και ταξινόμησης τεχνικών με πιο απλές αρχιτεκτονικές. Επιπλέον η έλλειψη επισημασμένων δεδομένων (labelled data) έχει οδηγήσει τους επιστήμονες στη μελέτη και εισαγωγή νέων προσεγγίσεων οι οποίες ανακαλύπτουν μοτίβα, δομές και ποικίλα εσωτερικά χαρακτηριστικά των δεδομένων χωρίς την ανάγκη ανθρώπινης παρέμβασης.

Σε αυτό το πλαίσιο, η παρούσα διπλωματική εργασία ασχολείται με την διερεύνηση, μελέτη και αξιολόγηση μεθόδων μάθησης και κυρίως αυτο-επιβλεπόμενης μάθησης (self-supervised learning) χρησιμοποιώντας σετ δεδομένων χωρίς δεδομένα αναφοράς. Πιο συγκεκριμένα, μελετώνται και καταγράφονται σύγχρονες αρχιτεκτονικές νευρωνικών δικτύων που αφορούν την αυτό-επιβλεπόμενη μάθηση και πραγματοποιείται μια σειρά πειραμάτων πάνω σε θεμελιώδη σετ δεδομένων. Επιλέγεται μια πρόσφατη τεχνική της βιβλιογραφίας, η οποία έχει πετύχει υψηλές ακρίβειες στην ταξινόμηση εικόνας πάνω σε γνωστές συλλογές δεδομένων της Όρασης Υπολογιστών αλλά και της Τηλεπισκόπησης.

Προκειμένου να πραγματοποιηθούν τα πειράματα της ταξινόμησης εικόνας αναπτύχθηκε κώδικας σε γλώσσα προγραμματισμού Python 3 και πιο συγκεκριμένα με τη χρήση της βιβλιοθήκης PyTorch. Τέλος, τα συμπεράσματα και οι προτάσεις που παρατίθενται μπορούν να αποτελέσουν βοηθητικό υλικό για μελλοντικές ερευνητικές εργασίες.

ABSTRACT

The research and development of automated machine learning algorithms for image classification and object recognition has not stopped gaining significant attention from the scientific community for several decades. New methodologies are constantly being introduced in order to improve the automation in understanding data and their underlying structures. During the last years deep learning based approaches have surpassed the classification accuracy of simpler architectures. Moreover, the lack of labelled data has led scientists to discover new algorithms that discover hidden patterns or clusters of data without the need for human intervention.

Towards this direction, this diploma thesis focused on the study, application and evaluation of deep learning methods and mainly with self-supervised learning using unlabelled data. More specifically, modern neural network architectures related to self-supervised learning are studied and recorded and also a series of experiments are performed on benchmark datasets. A recent bibliographic technique which has achieved high accuracy in image classification task in well-known datasets for Computer Vision and Remote Sensing is selected.

In order to perform the image classification experiments, code was developed in Python 3 programming language and more specifically utilizing the PyTorch framework. Finally, the conclusions and suggestions presented can be useful material for future research work.

ΠΕΡΙΕΧΟΜΕΝΑ

ΠΕΡΙΛΗΨΗ **7**
ABSTRACT **9**
1. ΕΙΣΑΓΩΓΗ 13
1.1 Κίνητρο 15
1.2 Αντικείμενο Διπλωματικής 15
2. ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ - ΑΝΑΣΚΟΠΗΣΗ ΒΙΒΛΙΟΓΡΑΦΙΑΣ 16
2.1 Νευρωνικά δίκτυα και Ταξινόμηση εικόνας 16
2.2. Αρχιτεκτονική Νευρωνικών Δικτύων και Συνελικτικά Νευρωνικά Δίκτυα 18
2.3 Βασικές αρχιτεκτονικές Συνελικτικών Νευρωνικών Δικτύων 20
2.3.1 Περιγραφή του δικτύου AlexNet 21
2.3.2 Περιγραφή του δικτύου GoogleNet 23
2.3.3 Περιγραφή του δικτύου ResNet (Residual Network) 25
2.4 Σετ Δεδομένων και Δεδομένα Αναφοράς 28
2.4.1 Σετ Δεδομένων χωρίς δεδομένα αναφοράς, χωρίς ετικέτες (Unlabeled Data) 29
2.4.2 Σετ Δεδομένων με δεδομένα αναφοράς, με ετικέτες (Labeled Data) 29
2.5 Δομή δεδομένων αναφοράς στη μηχανική μάθηση 30
2.6 Προβλήματα Μηχανικής Μάθησης 31
2.6.1 Επιβλεπόμενη μάθηση (Supervised Learning) 31
2.6.2 Μη επιβλεπόμενη μάθηση (Unsupervised Learning) 32
2.6.3 Ενισχυτική μάθηση (Reinforcement Learning) 33
2.6.4 Ημι-επιβλεπόμενη μάθηση (Semi-supervised Learning) 33
2.7 Αυτο-επιβλεπόμενη Μάθηση (Self-supervised learning) 34
2.7.1 Pretext Tasks 36
2.7.2 «Αντιθετική» μάθηση (Contrastive learning) 40
2.8 «Απόσταξη» Γνώσης (Knowledge Distillation) 42
2.9 Σετ δεδομένων (Datasets) στην Αυτό-επιβλεπόμενη Μάθηση 45
2.10 Self-supervised Learning και Τηλεπισκόπηση 48
2.11 Self-Supervised Learning of Pretext-Invariant Representations 53
2.11.1 Επισκόπηση της μεθόδου 54
2.11.2 Πειράματα της εργασίας PIRL 54

3. ΜΕΘΟΔΟΛΟΓΙΑ ΚΑΙ ΥΛΟΠΟΙΗΣΗ 57
3.1 Μέθοδος 57

3.2 Δεδομένα 57
3.3 Υλοποίηση 60
3.4 Πειραματική διαδικασία 62
4. ΑΠΟΤΕΛΕΣΜΑΤΑ ΚΑΙ ΑΞΙΟΛΟΓΗΣΗ 64
4.1 Πρώτο πείραμα 65
4.2 Δεύτερο πείραμα 69
4.3 Τρίτο πείραμα 72
4.4 Τέταρτο πείραμα 74
4.5 Πέμπτο πείραμα 77
4.6 Έκτο πείραμα 78
4.7 Έβδομο πείραμα 79
4.8 Όγδοο πείραμα 81
4.9 Συγκεντρωτικά Αποτελέσματα 85
5. ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΠΡΟΟΠΤΙΚΕΣ 86
5.1 Γενικά Συμπεράσματα 86
5.2 Προοπτικές 87
ΠΑΡΑΡΤΗΜΑ 89
ΒΙΒΛΙΟΓΡΑΦΙΑ 93
Ξενόγλωσση βιβλιογραφία 93
Ιστοσελίδες 94

1. ΕΙΣΑΓΩΓΗ

Η όραση υπολογιστών είναι ένα πεδίο της τεχνητής νοημοσύνης (artificial intelligence-AI) το οποίο επιχειρεί να αναπαράγει, σε περιβάλλον ηλεκτρονικού υπολογιστή, τη βιολογική όραση. Επιτρέπει στους υπολογιστές και στα συστήματα να εξάγουν σημαντικές πληροφορίες από τις ψηφιακές εικόνες, τα βίντεο και άλλα ψηφιακά συστήματα, όπως σαρωτές, και με βάση αυτές τις πληροφορίες να πραγματοποιεί κάποιες ενέργειες ή να κάνει κάποιες συστάσεις. Έτσι η όραση υπολογιστών λειτουργεί παρόμοια με την ανθρώπινη όραση. Ο άνθρωπος μπορεί με χαρακτηριστική ευκολία να διακρίνει αντικείμενα, χρώματα, πρότυπα, υφή καθώς και τρισδιάστατη πληροφορία η οποία περιέχεται σε μια δισδιάστατη εικόνα. Αυτή η ευκολία όμως είναι μάλλον παραπλανητική διότι κρύβει πίσω της εκατομμύρια χρόνια εξέλιξης.

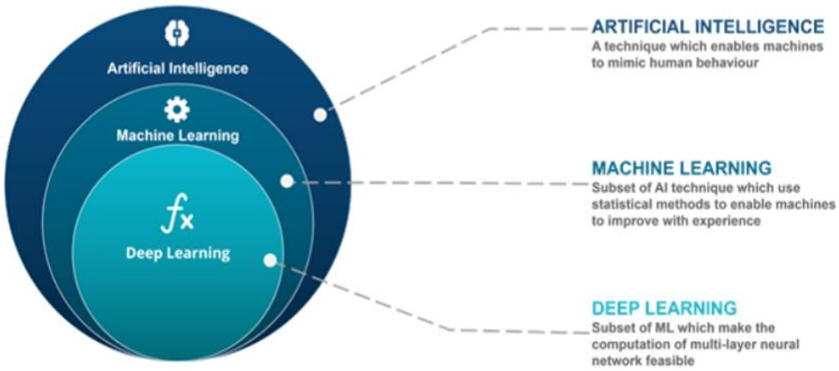
Σε κάθε ημισφαίριο του ανθρώπινου εγκεφάλου υπάρχει ένα πρωτεύον οπτικό σύστημα το οποίο αποτελείται από περίπου 140 εκατομμύρια νευρώνες, οι οποίοι συνδέονται με δισεκατομμύρια συνδέσεις μεταξύ τους. Το πολύπλοκο οπτικό αυτό σύστημα είναι ικανό να εκτελεί την επεξεργασία των εικόνων και την αναγνώριση προτύπων, που δέχεται ο εγκέφαλος μέσω του συστήματος της όρασης. Η όραση υπολογιστών εκπαιδεύει μηχανές να εκτελούν κάποιες από αυτές τις λειτουργίες της όρασης, όμως πρέπει να το κάνουν σε πολύ λιγότερο χρόνο με κάμερες, δεδομένα και αλγόριθμους και όχι με τον αμφιβληστροειδή, τα οπτικά νεύρα και τον οπτικό φλοιό όπως το ανθρώπινο σύστημα της όρασης.

Τα πρώτα πειράματα της όρασης υπολογιστή πραγματοποιήθηκαν κατά τη δεκαετία του 1950, χρησιμοποιώντας μερικά από τα πρώτα νευρωνικά δίκτυα για να ανιχνεύσουν τις ακμές ενός αντικειμένου και να ταξινομήσουν απλά αντικείμενα σε κατηγορίες όπως κύκλοι και τετράγωνα. Έπειτα την δεκαετία 1970, πραγματοποιήθηκε η πρώτη εμπορική χρήση της όρασης υπολογιστή που ερμήνευσε δακτυλογραφημένο ή χειρόγραφο κείμενο χρησιμοποιώντας οπτική αναγνώριση χαρακτήρων. Καθώς το διαδίκτυο αναπτυσσόταν την δεκαετία του 1990, μεγάλα σύνολα εικόνων ήταν διαθέσιμα για ανάλυση. Την περίοδο εκείνη τα προγράμματα αναγνώρισης προσώπων ήταν σε ανάπτυξη. Τα τελευταία χρόνια η ευκολία κτήσης των οπτικών ψηφιακών δεδομένων έχει απλοποιηθεί σε πολύ μεγάλο βαθμό. Χαρακτηριστικό παράδειγμα αποτελεί το γεγονός ότι ο κάθε άνθρωπος έχει στην κατοχή του ένα κινητό τηλέφωνο το οποίο διαθέτει μία, δύο ή και τρείς κάμερες. Ως αποτέλεσμα, υπάρχει ένας μεγάλος αριθμός ψηφιακών οπτικών δεδομένων που παράγονται καθημερινά. Έτσι είναι ευνόητο πως με την πληθώρα των διαθέσιμων οπτικών δεδομένων υπάρχει και η αντίστοιχη ανάγκη για την ανάπτυξη μεθόδων επεξεργασίας και εκμετάλλευσής τους.

Το πεδίο της όρασης υπολογιστών είναι στην πραγματικότητα ένα διεπιστημονικό πεδίο που αγγίζει πολλά τμήματα της επιστήμης και οι εφαρμογές του έχουν ένα ευρύ πεδίο δράσης. Ορισμένες μόνο εφαρμογές περιλαμβάνουν την ανίχνευση κίνησης, την αναγνώριση αντικειμένων, την τριδιάστατη ανακατασκευή μοντέλων καθώς και πλήθος ακόμα διαφορετικών εφαρμογών σε διάφορα επιστημονικά πεδία. Η αναγνώριση προτύπων-αντικειμένων αποτελεί δύσκολη διεργασία για έναν υπολογιστή, και αυτό είναι φανερό όταν επιχειρηθεί να γραφτεί ένα πρόγραμμα σε ηλεκτρονικό υπολογιστή για την υλοποίηση της εργασίας αυτής. Ενώ για τον ανθρώπινο εγκέφαλο είναι εξαιρετικά απλή η διαδικασία αντίληψης του περιεχομένου μιάς ψηφιακής εικόνας, για έναν υπολογιστή δεν είναι εύκολη υπόθεση. Οι εικόνες δεν έχουν καμία απολύτως ποιοτική σημασία για τον υπολογιστή και το μόνο που αντιλαμβάνεται είναι πίνακες χωρισμένους σε εικονοψηφία με το καθένα από αυτά να έχει μια τιμή φωτεινότητας.

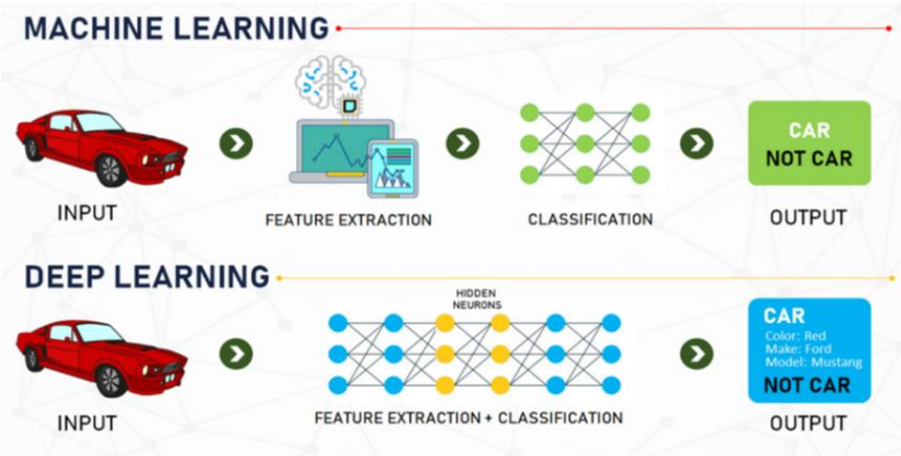
Προκειμένου λοιπόν να γίνει μια διαφορετική προσέγγιση στα προβλήματα της μηχανικής μάθησης, αξιοποιήθηκαν τα νευρωνικά δίκτυα. Η κύρια ιδέα είναι να εισάγεται στο δίκτυο ένας μεγάλος αριθμός εικόνων, από τις οποίες το σύστημα «εκπαιδεύεται» ώστε να αναγνωρίσει το περιεχόμενο τους και στη συνέχεια δοκιμάζεται σε ένα διαφορετικό σετ δεδομένων και ελέγχεται η αξιοπιστία του. Με όσες περισσότερες εικόνες τροφοδοτηθεί το σύστημα και όσο καλύτερης ανάλυσης είναι, τόσο καλύτερη η αποτελεσματικότητα του και η τελική του ακρίβεια.

Γενικότερα στην επιστημονική κοινότητα υπάρχει ένας διαχωρισμός ανάμεσα στην Τεχνητή Νοημοσύνη, τη Μηχανική Μάθηση και τη Βαθιά Μάθηση. Αρχικά, η τεχνητή νοημοσύνη είναι η τεχνική εκείνη η οποία επιτρέπει στις μηχανές να μιμηθούν την ανθρώπινη συμπεριφορά. Υποσύνολο της τεχνητής νοημοσύνης είναι η μηχανική μάθηση η οποία χρησιμοποιεί στατιστικές μεθόδους που επιτρέπουν στις μηχανές να βελτιώνονται με την εμπειρία και την πάροδο του χρόνου. Τέλος, η βαθιά μάθηση που προαναφέρθηκε, είναι επίσης υποσύνολο της μηχανικής μάθησης, και είναι ο τομέας ο οποίος πρακτικά καθιστά εφικτούς του υπολογισμούς των πολυεπίπεδων νευρωνικών δικτύων.



Εικόνα: Σχηματική απεικόνιση της Τεχνητής Νοημοσύνης, της Μηχανικής Μάθησης και της Βαθιάς Μάθησης.
(Πηγή: [18] https://dltlabs.medium.com/understanding-machine-learning-deep-learning-f5aa95264d61)

Ουσιαστικά η μηχανική μάθηση δημιουργήθηκε για να αναπτύξει έξυπνα συστήματα και με την επιστήμη του Deep Learning, εξελίχθηκε σε ένα εντελώς διαφορετικό επίπεδο. Το Machine Learning ήταν μόνο η αρχή και το Deep Learning έχει βελτιώσει τον τομέα αυτό σε σημείο που πλέον υπάρχουν αυτοκινούμενα οχήματα.



Εικόνα: Διαφορά ανάμεσα στην Μηχανική μάθηση και την βαθιά μάθηση.
(Πηγή: [19] https://dltlabs.medium.com/understanding-machine-learning-deep-learning-f5aa95264d61)

1.1 Κίνητρο

Βασικό κίνητρο για τη συγγραφή της παρούσας διπλωματικής εργασίας αποτελεί η συνεχής αναζήτηση του Μηχανικού Γεωπληροφορικής για αποτελεσματικότερες και καλύτερες μεθόδους προσδιορισμού και ανάλυσης του χώρου της Γήινης επιφάνειας. Με τη ραγδαία εξέλιξη που έχει η τεχνολογία στην σύγχρονη εποχή, μέρα με την μέρα, νέες τεχνικές έρχονται στο φως και πολλά πειράματα πραγματοποιούνται. Ως εκ τούτου και η επιστήμη της Τηλεπισκόπησης απαιτεί τη συμπόρευση του Τοπογράφου Μηχανικού με όλα τα σύγχρονα μέσα.

Τα τελευταία χρόνια λοιπόν, καθημερινά σημειώνεται μεγάλη ανάπτυξη πάνω στον τομέα της όρασης υπολογιστών και πιο συγκεκριμένα στο πεδίο της βαθιάς μάθησης. Παρουσιάζονται συνεχώς νέες μέθοδοι με δυνατότητες που πριν από μερικά χρόνια φάνταζαν αδιανόητες. Τα νευρωνικά δίκτυα πλέον είναι ευρέως χρησιμοποιούμενα και οι αρχιτεκτονικές τους γίνονται όλο και πολυπλοκότερες και η αποτελεσματικότητά τους αυξάνεται. Η βαθιά μάθηση εφαρμόζεται πλέον στην ιατρική, στη φαρμακευτική, στην παιδεία, στην ψυχαγωγία και εισχωρεί με την πάροδο του χρόνου σε όλους τους διαθέσιμους τομείς.

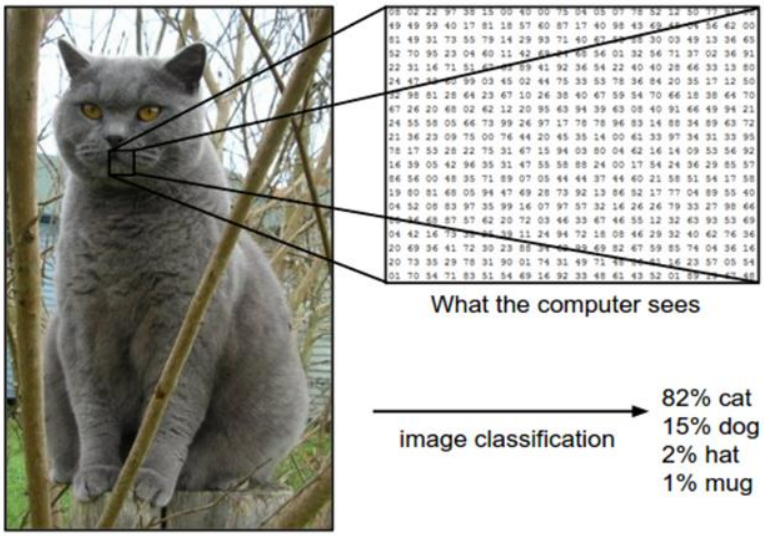
1.2 Αντικείμενο Διπλωματικής

Η παρούσα διπλωματική εργασία ασχολείται με την εκπαίδευση νευρωνικών δικτύων (Convolutional Neural Networks - CNNs) πάνω σε θεμελιώδη σετ δεδομένων. Πραγματοποιείται μια σειρά από πειράματα πάνω σε αυτο-επιβλεπόμενη μάθηση προκειμένου να συγκριθούν οι ακρίβειες που προκύπτουν σε διαφορετικά σετ δεδομένων εκπαίδευσης. Αρχικά γίνεται προ-εκπαίδευση των μοντέλων με βάση κάποιο γνωστό pretext task και έπειτα πραγματοποιείται εργασία ταξινόμησης εικόνας. Ως pretext task ορίζεται μία προσχεδιασμένη εργασία προς επίλυση από τα δίκτυα όπου χρήσιμα οπτικά χαρακτηριστικά των εικόνων μαθαίνονται μέσω αντικειμενικών συναρτήσεων της εργασιας pretext. Επίσης, η διπλωματική εργασία αυτή έχει βιβλιογραφική σημασία καθώς συγκεντρώνει και αναλύει σύγχρονες μεθόδους εκπαίδευσης και προ-εκπαίδευσης μοντέλων καθώς και τις αρχιτεκτονικές των νευρωνικων δικτύων που χρησιμοποιούνται περισσότερο σήμερα. Η εργασία δεν έχει στόχο την ανάπτυξη κάποιου αλγορίθμου αιχμής, αλλά τη μελέτη πρόσφατων αναπτυγμένων μεθόδων νευρωνικών δικτύων που αφορούν την αυτο-επιβλεπόμενη μάθηση και την ταξινόμηση εικόνας.

2. ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ - ΑΝΑΣΚΟΠΗΣΗ ΒΙΒΛΙΟΓΡΑΦΙΑΣ

2.1 Νευρωνικά δίκτυα και Ταξινόμηση εικόνας

Το πρόβλημα της ταξινόμησης εικόνας αναφέρεται στη διαδικασία χαρακτηρισμού της εικόνας από ένα καθορισμένο κατάλογο κατηγοριών-κλάσεων. Η διαδικασία αυτή αποτελεί κυρίαρχο θέμα απασχόλησης στον τομέα της όρασης υπολογιστών καθώς αποδεικνύεται ότι πολλά άλλα φαινομενικά διαφορετικά αντικείμενα, όπως η αναγνώριση αντικειμένων (image recognition) και η κατάτμηση εικόνας (segmentation), εν τέλει ανάγονται στο πρόβλημα της ταξινόμησης εικόνας.



Εικόνα 2.1.1: Ταξινόμηση εικόνας
(Πηγή: [20] https://www.kdnuggets.com/2019/11/deep-learning-image-classification-less-data.html)

Η παραπάνω εικόνα παρουσιάζει το αποτέλεσμα της ταξινόμησης μιας αρχικής εικόνας εισόδου, η οποία έχει σαν έξοδο, μια κατανομή πάνω στις καθορισμένες κλάσεις που δείχνει την εμπιστοσύνη της διαδικασίας στην κυρίαρχη κλάση που αντιπροσωπεύει το αντικείμενο. Η παραπάνω διαδικασία, η οποία φαίνεται εξαιρετικά απλή εάν εκτελεστεί από τον ανθρώπινο εγκέφαλο, αντιμετωπίζει ένα σύνολο δυσκολιών όταν εκτελείται από τον υπολογιστή.

Η προσέγγιση που ακολουθείται για την ταξινόμηση της εικόνας σε μια καθορισμένη κλάση είναι παρεμφερής με τη διαδικασία μάθησης του παιδικού ανθρώπινου εγκεφάλου όταν συναντάει για πρώτη φορά ένα άγνωστο αντικείμενο. Το σύστημα τροφοδοτείται με ένα μεγάλο αριθμό εικόνων για κάθε κλάση, και στη συνέχεια αναπτύσσονται αλγόριθμοι μάθησης με τους οποίους το μοντέλο «εκπαιδεύεται» να αναγνωρίσει μελλοντικά το κάθε άγνωστο αντικείμενο που ανήκει στις κλάσεις εκπαίδευσης. Η ολοκληρωμένη διαδικασία μπορεί να συνοψισθεί ως παρακάτω. Ως δεδομένα εισόδου χρησιμοποιούνται ένα σύνολο από N εικόνες οι οποίες συνοδεύονται από μια ταμπέλα-χαρακτηρισμό με τις K διαθέσιμες κλάσεις, όπου N >> K.

Τα δεδομένα εισόδου αναφέρονται ως δεδομένα εκπαίδευσης του δικτύου. Στη συνέχεια, μέσω κατάλληλων αλγορίθμων, το μοντέλο χρησιμοποιώντας τα δεδομένα εκπαίδευσης «μαθαίνει» τις ιδιότητες του αντικειμένου (σχήμα, μέγεθος, μορφή, πρότυπα) με βάση τα ραδιομετρικά χαρακτηριστικά του. Το στάδιο αυτό λέγεται διαδικασία εκπαίδευσης του μοντέλου. Τέλος, αξιολογείται η ποιότητα του ταξινομητή ζητώντας του να προβλέψει ετικέτες για ένα σύνολο εικόνων που δεν έχουν χρησιμοποιηθεί στη διαδικασία εκπαίδευσης.

Η υλοποίηση της παραπάνω διαδικασίας εκπαίδευσης επιτυγχάνεται χρησιμοποιώντας δύο κύριες συναρτήσεις, μια συνάρτηση βαθμολογίας (score function) και μια συνάρτηση κόστους (loss function). Η συνάρτηση βαθμολογίας συνδέει την κάθε εικόνα με μια τιμή (score) για κάθε κλάση, ενώ η συνάρτηση κόστους ποσοτικοποιεί τη διαφορά μεταξύ της προβλεπόμενης και της πραγματικής κλάσης. Ακολούθως η διαδικασία ανάγεται σε διαδικασία βελτιστοποίησης, όπου επιδιώκεται η ελαχιστοποίηση της συνάρτησης κόστους ως προς τις παραμέτρους της συνάρτησης βαθμολογίας. Η συνάρτηση κόστους χρησιμοποιείται για να μετρηθεί η αβεβαιότητα της ταξινόμησης, όσο μεγαλύτερη η τιμή του κόστους τόσο πιο «λάθος» είναι η ταξινόμηση. Υπάρχουν διάφοροι τρόποι για να καθοριστεί το κόστος. Οι πιο γνωστές μέθοδοι είναι το Multiclass Support Vector Machine (SVM) και ο ταξινομητής Softmax. Αντίστοιχα, βελτιστοποίηση είναι η διαδικασία της εύρεσης του κατάλληλου συνδυασμού παραμέτρων που ελαχιστοποιούν τη χρησιμοποιούμενη συνάρτηση κόστους. Η κύρια στρατηγική που ακολουθείται είναι να αρχικοποιηθούν οι παράμετροι με τυχαίες τιμές και στη συνέχεια μέσω μιας επαναληπτικής διαδικασίας να βελτιστοποιηθούν μέχρι να επιτευχθεί το χαμηλότερο δυνατό κόστος. Σε κάθε επανάληψη της διαδικασίας απαιτείται να βρεθεί μια μεταβολή, έστω δW, ώστε τα νέα βάρη W+δW να δίνουν μικρότερη τιμή κόστους. Η ελαχιστοποίηση επιτυγχάνεται αναλυτικά υπολογίζοντας τις μερικές παραγώγους της συνάρτησης κόστους ως προς τις μεταβλητές της συνάρτησης βαθμολογίας. Η διαδικασία υπολογισμού των μερικών παραγώγων της συνάρτησης κόστους ως προς τα βάρη του δικτύου λέγεται Gradient Descent και αποτελεί τον πυρήνα των μεθόδων που χρησιμοποιούν νευρωνικά δίκτυα. Η διαδικασία της συνεχούς αναπροσαρμογής των παραμέτρων υπολογίζοντας τις μερικές παραγώγους είναι γνωστή και ως backpropagation. Ουσιαστικά μέσω αυτής της διαδικασίας υπολογίζεται το «λάθος» της ταξινόμησης, το οποίο διοχετεύεται προς την αντίθετη κατεύθυνση του δικτύου με σκοπό να επαναληφθεί η διαδικασία. Στην συνέχεια, πραγματοποιείται μια αναπροσαρμογή στις παραμέτρους η οποία υπολογίζεται με πολλαπλασιασμό ενός μεγέθους βήματος επί τη μεταβολή των παραμέτρων. Ακολούθως η διαδικασία επαναλαμβάνεται μέχρις ότου η συνάρτηση κόστους να λάβει τόσο μικρή τιμή που η αναπροσαρμογή των βαρών να μην έχει ουσιαστική σημασία στην ταξινόμηση. Το μέγεθος του βήματος, γνωστό ως learning rate, καθορίζει το ρυθμό με τον οποίο γίνεται αναπροσαρμογή των παραμέτρων και αποτελεί αντικείμενο διερεύνησης. Εάν ο ρυθμός είναι μικρός τότε το δίκτυο θα απαιτεί αρκετό χρόνο να εκπαιδευτεί, ενώ εάν είναι μεγάλος το δίκτυο θα είναι αναξιόπιστο.

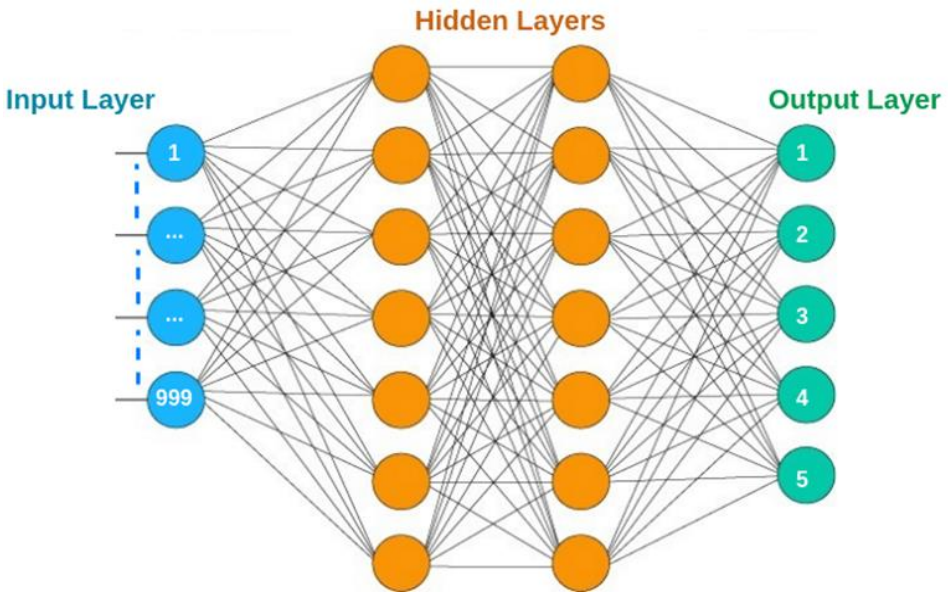
Σε εφαρμογές που τα δεδομένα εκπαίδευσης αποτελούνται από μεγάλο αριθμό εικόνων, οι παράμετροι μπορούν επίσης να φτάσουν σε μεγάλους αριθμούς. Αυτό σημαίνει πως η διαδικασία υπολογισμού των μερικών παραγώγων για όλα τα δεδομένα απαιτεί μεγάλο υπολογιστικό κόστος. Για την αντιμετώπιση αυτού του προβλήματος, συνήθως ακολουθείται μια διαδικασία διαχωρισμού των δεδομένων εκπαίδευσης σε μικρά πακέτα. Το πακέτο αυτό είναι βασικό χαρακτηριστικό του δικτύου και καλείται mini batch. Έτσι, πραγματοποιείται η εκπαίδευση ενός δικτύου με ένα προκαθορισμένο αριθμό εικόνων σύμφωνα με το mini batch και έπειτα επιλέγεται ένας νέος αριθμός δεδομένων. Μόλις ολοκληρωθεί η διαδικασία εκπαίδευσης για το σύνολο των δεδομένων, τότε έχει ολοκληρωθεί μια εποχή (epoch) εκπαίδευσης και η διαδικασία

επαναλαμβάνεται εκ νέου. Για να εξακριβωθεί η τιμή του βέλτιστου αριθμού mini batch απαιτείται πειραματισμός, συνήθως δοκιμάζοντας διαδοχικά δυνάμεις του 2 όπως 8, 16, 32, 64, 128, καθώς έτσι αποδίδει καλύτερα με ταχύτερα αποτελέσματα.

2.2. Αρχιτεκτονική Νευρωνικών Δικτύων και Συνελικτικά Νευρωνικά Δίκτυα

Όσον αφορά την αρχιτεκτονική των νευρωνικών δικτύων, αυτά εμπνεύστηκαν από το ανθρώπινο νευρικό σύστημα, και μοντελοποιούνται μαθηματικά ως συλλογή από νευρώνες συνδεδεμένους μεταξύ τους σε ένα μη κυκλικό γράφημα (Εικόνα 2.2.1). Μη κυκλικό διότι, η έξοδος από έναν ή κάποιους νευρώνες αποτελεί είσοδο για το επόμενο επίπεδο νευρώνων, χωρίς να γίνεται κάποιου είδους κυκλική διαδικασία η οποία θα οδηγούσε σε ατέρμονο βρόχο. Ο πιο συνηθισμένος τύπος είναι το fully-connected, όπου οι νευρώνες ενός προηγούμενου επιπέδου συνδέονται με όλους τους νευρώνες του επόμενου επιπέδου.

Στην παρούσα εργασία θα χρησιμοποιηθούν συνελικτικά νευρωνικά δίκτυα (CNNs) τα οποία αποτελούν τον πυρήνα στον τομέα της βαθιάς μάθησης (deep learning). Παρόλο που η χρήση τους είναι γνωστή από τη δεκαετία του 90 (Le Cun et al., 1997), διαδόθηκαν αρκετά και έγιναν ευρέως γνωστά με τη χρησιμοποίηση τους για σκοπούς ταξινόμησης εικόνας κατά το διαγωνισμό ImageNet (Krizhevsky et al., 2012).



Εικόνα 2.2.1 : Πλήρως συνδεδεμένο νευρωνικό δίκτυο (Fully connected NN)
(Πηγή : [21] https://www.researchgate.net/figure/General-Architecure-of-the-Fully-Connected-NN_fig7_325022519)

Στην εικόνα αυτή απεικονίζεται μια τυπική μορφή απλού νευρωνικού δικτύου το οποίο αποτελείται από 3 συνολικά επίπεδα (layer), ένα layer εισόδου (δεν προσμετράται στο συνολικό αριθμό), 2 κρυμμένα layer και ένα layer εξόδου. Το δίκτυο αποτελείται από συνολικά από [7+7+5]=19 νευρώνες με τον καθένα να περιέχει μια τιμή κατωφλίου και [4x7 + 7x7 + 7x5] = 112

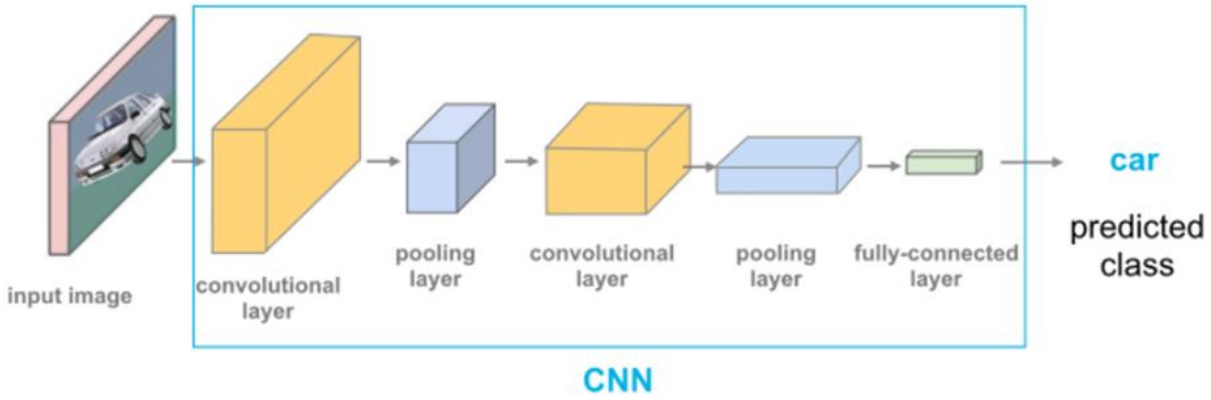
βάρη τα οποία περιέχονται σε κάθε σύνδεση. Δηλαδή ο συνολικός αριθμός εκπαιδευόμενων παραμέτρων είναι 131. Για σύγκριση αναφέρεται ότι τα σύγχρονα συνελικτικά νευρωνικά δίκτυα αποτελούνται από περίπου 100 εκατομμύρια παραμέτρους και διαθέτουν 10-20 επίπεδα (deep neural networks). Ο λόγος για τον οποίο τα νευρωνικά δίκτυα αναπαρίστανται με τον παραπάνω τρόπο είναι διότι η δομή αυτή καθιστά εύκολο το χειρισμό τους με πολλαπλασιασμούς πινάκων. Η είσοδος του δικτύου είναι ένα διάνυσμα και όλες οι παράμετροι του κάθε layer μπορούν να αναπαρασταθούν με έναν πίνακα. Η διαδικασία μετάδοσης της πληροφορίας διαμέσου των νευρώνων, από την είσοδο της μεταμορφωμένης εικόνας-στήλη έως την τελική έξοδο της στο τελευταίο layer, καλείται forward pass. Κατά τη διαδικασία αυτή, σε κάθε νευρώνα του κάθε επιπέδου εκτελείται μια προκαθορισμένη πράξη. Το αποτέλεσμα αυτής εισάγεται σαν τιμή εισόδου σε μια συνάρτηση ενεργοποίησης (activation function), χαρακτηριστική για ολόκληρο το δίκτυο. Ο κάθε νευρώνας δέχεται τιμές από όλους (ή μερικούς) νευρώνες του προηγούμενου επιπέδου και εκτελεί ένα σταθμισμένο άθροισμα και μία πρόσθεση της τιμής που περιέχει. Στη συνέχεια η τιμή αυτή διοχετεύεται σε μια συνάρτηση ενεργοποίησης και η διαδικασία επαναλαμβάνεται στο επόμενο επίπεδο. Οι πιο συνήθεις συναρτήσεις ενεργοποίησης είναι: η σιγμοειδής συνάρτηση, Tanh και ReLu (Rectified Linear Unit).

Τα συνελικτικά νευρωνικά δίκτυα λοιπόν, είναι όμοια με τα απλά νευρωνικά δίκτυα. Τα δομικά τους στοιχεία είναι οι νευρώνες, στους οποίους αντιστοιχούν κάποια βάρη. Κατά τη διαδικασία εκπαίδευσης αναζητούνται οι βέλτιστες τιμές για τις τιμές των βαρών. Το δίκτυο αναπαρίσταται από μια διαφορίσιμη συνάρτηση βαθμολογίας, η οποία λαμβάνει ως είσοδο τα pixel μιας εικόνας και εξάγει βαθμολογίες για κάθε κλάση. Χρησιμοποιείται ομοίως μια συνάρτηση κόστους η οποία ποσοτικοποιεί τη διαφορά της προβλεπόμενης από την αληθή κατηγορία. Η κύρια διαφορά τους σε σχέση με τα απλά δίκτυα είναι ότι δέχονται ως είσοδο ολόκληρη την εικόνα, αντί για ένα διάνυσμα μιας κατεύθυνσης, γεγονός που επιτρέπει την εφαρμογή ορισμένων χαρακτηριστικών στην αρχιτεκτονική του δικτύου.

Τα απλά νευρωνικά δίκτυα δέχονται ως είσοδο μια εικόνα η οποία μετατρέπεται σε διάνυσμα μιας κατεύθυνσης, και στη συνέχεια μετασχηματίζεται μέσα από ένα διαδοχικό σύνολο από κρυμμένα Layer. Κάθε κρυμμένο layer αποτελείται από νευρώνες και κάθε νευρώνας είναι πλήρως συνδεδεμένος με τους νευρώνες του προηγούμενου επιπέδου, ενώ οι νευρώνες του ίδιου επιπέδου δεν μοιράζονται μεταξύ τους κάποια σύνδεση. Η σύνδεση αυτή οδηγεί σε ένα τεράστιο αριθμό παραμέτρων του δικτύου, με επιπτώσεις τόσο στη διαδικασία μάθησης, όσο και στην αποτελεσματικότητα του δικτύου να αναγνωρίζει τα νέα άγνωστα αντικείμενα (Overfitting). Τα συνελικτικά νευρωνικά δίκτυα δέχονται ως είσοδο ολόκληρη την τρισδιάστατη εικόνα και οργανώνουν τους νευρώνες σε κάθε επόμενο Layer σε τρεις διαστάσεις. Κάθε layer μεταμορφώνει τον όγκο εισόδου σε έναν τρισδιάστατο όγκο εξόδου από ενεργοποιημένους νευρώνες μέσω μιας παραγωγίσιμης συνάρτησης ενεργοποίησης. Το τελευταίο Layer εξόδου θα έχει διαστάσεις 1x1xN, διότι ως το τέλος της διαδικασίας η αρχική εικόνα θα έχει μειωθεί σε ένα τρισδιάστατο διάνυσμα με τις N βαθμολογίες της κάθε κλάσης, διευθετημένο στη διεύθυνση του βάθους

Γενικά χρησιμοποιούνται τρεις τύποι επιπέδων για τη δημιουργία της αρχιτεκτονικής του συνελικτικού νευρωνικού δικτύου. Οι τύποι αυτοί είναι τα Convolutional layers, τα Pooling Layers και τα Fully-connected layers (FC). Κάποια από τα επίπεδα αυτά έχουν παραμέτρους ή υπερπαραμέτρους ενώ κάποια άλλα δεν έχουν. Η διαδοχή των τριών αυτών τύπων στο δίκτυο μετασχηματίζει τον αρχικό όγκο της εικόνας σε έναν τελικό όγκο με τις πιθανότητες των κλάσεων.

Το συνελικτικό επίπεδο (convolutional layer) εκτελεί το μεγαλύτερο όγκο της υπολογιστικής διαδικασίας της εκπαίδευσης του δικτύου. Οι παράμετροι του επιπέδου αυτού αποτελούν τα βάρη του δικτύου, των οποίων αναζητείται η βέλτιστη τιμή. Τα βάρη αυτά είναι οι τιμές διαφόρων χωρικών φίλτρων που χρησιμοποιούνται σε κάθε επίπεδο. Το μέγεθος του φίλτρου είναι συνήθως μικρό και εκτείνεται σε μια μικρή μόνο περιοχή της εικόνας αλλά σε όλο το βάθος της. Τα pooling layer τοποθετούνται συνήθως ανάμεσα στα συνελικτικά επίπεδα. Χρησιμοποιούνται για να μειώσουν σταδιακά τις διαστάσεις της αρχικής εικόνας με συνεπακόλουθη μείωση του αριθμού των παραμέτρων. Λειτουργούν ξεχωριστά σε κάθε επίπεδο και μειώνουν τις διαστάσεις χρησιμοποιώντας τη συνάρτηση μέγιστης επιλογής (max operation). Οι νευρώνες σε ένα Fully Connected επίπεδο i έχουν συνδέσεις, επομένως και βάρη, με όλους τους νευρώνες του προηγούμενου επιπέδου i-1, όπως ακριβώς συμβαίνει στα απλά νευρωνικά δίκτυα. Η διαφορά τους με τα συνελικτικά επίπεδα είναι ότι τα τελευταία συνδέονται μόνο με μια μικρή περιοχή του προηγούμενου επιπέδου. Η πιο συνηθισμένη αρχιτεκτονική νευρωνικών δικτύων αποτελείται από διαδοχικά συνελικτικά – RELU επίπεδα ακολουθούμενα από Max Pooling επίπεδα. Το πρότυπο αυτό επαναλαμβάνεται μέχρι ο τελικός όγκος να έχει μικρύνει αρκετά σε διαστάσεις. Στα τελικά επίπεδα είναι σύνηθες ο τελικός όγκος να διοχετεύεται σε fully-connected επίπεδα, τα οποία υπολογίζουν τις βαθμολογίες για κάθε κλάση.



Εικόνα 2.2.2: Μια τυπική αρχιτεκτονική ενός συνελικτικού νευρωνικού δικτύου.
(Πηγή: [22] https://cezannec.github.io/Convolutional Neural Networks/)

2.3 Βασικές αρχιτεκτονικές Συνελικτικών Νευρωνικών Δικτύων

Υπάρχουν αρκετές αρχιτεκτονικές συνελικτικών νευρωνικών δικτύων που έχουν αναπτυχθεί τα τελευταία χρόνια για σκοπούς ταξινόμησης εικόνας με συγκεκριμένο όνομα. Τα πιο γνωστά από αυτά παρουσιάζονται συνοπτικά παρακάτω:

- **LeNet**

Αποτελεί την πρώτη επιτυχημένη προσπάθεια η οποία επιτεύχθηκε τη δεκαετία του 90 από τον Yann LeCun. Χρησιμοποιήθηκε για σκοπούς αναγνώρισης ψηφίων.

- **Alexnet**

Αποτελεί το πρώτο εγχείρημα που έκανε γνωστά τα συνελικτικά νευρωνικά δίκτυα για σκοπούς όρασης υπολογιστών. Αναπτύχθηκε από τους Alex Krizhevsky, Ilya Sutskever και Geoff Hinton το 2012 όπου και κατέλαβε την πρώτη θέση σε αντίστοιχο ετήσιο διαγωνισμό. Το δίκτυο είχε αρκετές ομοιότητες με το LeNet, αλλά ήταν βαθύτερο, μεγαλύτερο και με εκτεταμένη χρήση συνεχομένων συνελικτικών νευρωνικών δικτύων.

- **ZF Net**

Δημιουργήθηκε από τους Matthew Zeiler και Rob Fergus το 2013, οι οποίοι τροποποιώντας κάποιες από τις υπερπαραμέτρους του Alexnet, παρουσίασαν μια βελτιωμένη έκδοση του δικτύου.

- **GoogleNet**

Δημιουργήθηκε από τους Szegedy et al. για λογαριασμό της Google. Η κύρια συνεισφορά του ήταν η ανάπτυξη ενός Inception Module το οποίο μείωσε δραστικά τον αριθμό των παραμέτρων σε 4.000.000 σε σύγκριση με τα 60.000.000 παραμέτρους του Alexnet.

- **VGGnet**

Δημιουργήθηκε από τους Karen Simonyan και Andrew Zisserman το 2014, οι οποίοι απέδειξαν ότι το βάθος του δικτύου έχει σημαντική επίδραση στην αποτελεσματικότητά του. Το δίκτυο έχει εξαιρετικά ομοιογενή αρχιτεκτονική αποτελούμενο από 16 συνελικτικά και FC επίπεδα, φίλτρα διαστάσεων 3x3 και χρησιμοποιεί Max Pooling 2x2.

- **ResNet**

Το δίκτυο αυτό ήταν ο νικητής του διαγωνισμού ILSVRC το 2015. Δημιουργήθηκε από τους Kaiming He et al. Το δίκτυο αυτό αποτελεί state of the art αρχιτεκτονικής νευρωνικού δικτύου και είναι το πιο διαδεδομένο στην επίλυση προβλημάτων ταξινόμησης εικόνας επι του παρόντος.

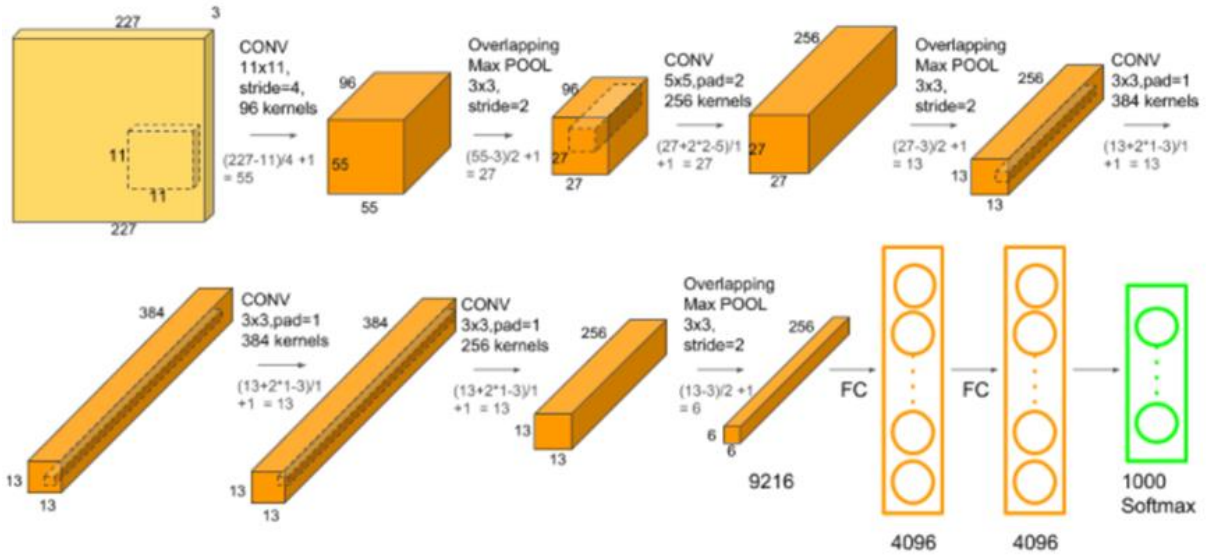
Τέλος, αναφέρεται ότι τα παραπάνω δίκτυα είναι δοκιμασμένα και λειτουργούν ικανοποιητικά για σκοπούς ταξινόμησης εικόνας. Αναλόγως του προβλήματος, μπορούν να χρησιμοποιηθούν χωρίς να απαιτείται να στηθούν από την αρχή. Ενδεχομένως να χρειάζεται να βελτιστοποιηθούν οι παράμετροι των δικτύων (fine tuning), για να προσαρμοστούν στις απαιτήσεις του προβλήματος κάθε φορά. Ως βελτιστοποίηση παραμέτρων ορίζεται η εκ νέου εκπαίδευση του ήδη εκπαιδευμένου δικτύου, αλλά με διαφορετικές εικόνες τις οποίες θα πρέπει να συλλέξει ο χρήστης. Ενδεικτικά παρακάτω γίνεται η αναλυτική περιγραφή του AlexNet, του GoogleNet και του ResNet καθώς η πλήρης περιγραφή όλων των διαθέσιμων αρχιτεκτονικών είναι πέραν του σκοπού της διπλωματικής.

2.3.1 Περιγραφή του δικτύου AlexNet

Το δίκτυο AlexNet αποτέλεσε την απαρχή για την εισαγωγή των συνελικτικών νευρωνικών δικτύων για σκοπούς ταξινόμησης και ανίχνευσης αντικειμένων σε εικόνες. Όταν προτάθηκε, πέτυχε να αναγνωρίσει αντικείμενα με ποσοστό λάθους περίπου 17% στο top-5 και 32% στο top-1 (ποσοστό μη εύρεσης της αληθούς κλάσης ανάμεσα στις πέντε πρώτες και στην πρώτη αντίστοιχα προβλέψεις), ποσοστά τα οποία ήταν πολύ καλύτερα από ό,τι είχε επιτευχθεί μέχρι τότε.

Το δίκτυο δέχεται RGB εικόνες διαστάσεων 227x227x3 και εξάγει ένα διάνυσμα διαστάσεων 1x1x1000 με τις πιθανότητες όλων των κλάσεων της βάσης Imagenet, που πιθανόν να ανήκει η εικόνα, με συνολικό άθροισμα τη μονάδα. Εάν οι διαστάσεις της εικόνας είναι διαφορετικές, τότε αυτή γίνεται resize στην επιθυμητή διάσταση. Το δίκτυο περιέχει συνολικά 650.000 νευρώνες σε όλα τα επίπεδα του και 60.000.000 εκατομμύρια παραμέτρους.

Η αρχιτεκτονική του δικτύου όπως φαίνεται στην παρακάτω εικόνα 2.18, αποτελείται από πέντε συνελικτικά νευρωνικά επίπεδα και τρία πλήρως συνδεδεμένα. Τα πρώτα δύο επίπεδα ακολουθούνται από Overlapping Max Pooling επίπεδα, ενώ το τρίτο, τέταρτο και πέμπτο είναι απευθείας συνδεδεμένα. Το πέμπτο συνελικτικό επίπεδο ακολουθείται από ένα Overlapping Max Pooling επίπεδο και το αποτέλεσμα διοχετεύεται σε δύο πλήρως συνδεδεμένα επίπεδα και τελικά σε έναν ταξινομητή Softmax. Ως συνάρτηση ενεργοποίησης χρησιμοποιείται η ReLu αντί για τη σιγμοειδή ή την tanh, ανάμεσα στα επίπεδα. Η χρήση της απέδειξε ότι η διαδικασία εκπαίδευσης επιταχύνεται σε σχέση με τις άλλες δύο κατά περίπου 6 φορές περισσότερο.



Εικόνα 2.3.1.1: Αρχιτεκτονική AlexNet (Πηγή : [23] https://neurohive.io/en/popular-networks/alexnet-imagenet-classification-with-deep-convolutional-neural-networks/)

Το επίπεδο Overlapping Max Pooling εκτελεί την ίδια διεργασία με το Max Pooling, δηλαδή εκτελεί δειγματοληψία στις τιμές των pixel επιλέγοντας τη μέγιστη τιμή σε ένα προκαθορισμένο τετράγωνο, με τη διαφορά ότι τα παράθυρα δειγματοληψίας επικαλύπτονται μεταξύ τους. Οι συγγραφείς χρησιμοποίησαν παράθυρα 3x3 με βήμα 2 γεγονός που αποδείχθηκε ότι βελτίωσε, κατά ένα μικρό ποσοστό, το error rate του top- 1 και top-3 του δικτύου, σε σχέση με τα ευρέως χρησιμοποιούμενα 2x2 παράθυρα με βήμα 2.

Η κυριότερη μέθοδος, μεταξύ άλλων, που χρησιμοποιήθηκε για να μειωθεί το overfitting του δικτύου πρωτοπαρουσιάστηκε το 2012 και λέγεται dropout [16]. Overfitting υπάρχει όταν, ενώ το δίκτυο έχει εκπαιδευτεί σωστά δηλαδή έχει επιτευχθεί η σύγκλιση της συνάρτησης κόστους, και αναγνωρίζονται τα αντικείμενα στις εικόνες εκπαίδευσης, αποτυγχάνει ωστόσο να λειτουργήσει ικανοποιητικά σε νέα άγνωστα δεδομένα.

Κυριότερη αιτία είναι ο μεγάλος αριθμός των εκπαιδευόμενων παραμέτρων, συγκεκριμένα το Alexnet περιέχει 60.000.000, και η επακόλουθη δυσκολία εκπαίδευσης του μεγάλου αυτού

αριθμού. Κατά τη διαδικασία dropout ένας νευρώνας αφαιρείται από το δίκτυο με πιθανότητα 0.5 και έτσι δεν συμμετέχει στην εκπαίδευση του δικτύου. Με αυτόν τον τρόπο, μειώνεται ο αριθμός των νευρώνων και οι ενεργοποιήσεις τους ακολουθούν διαφορετική διαδρομή. Αυτό έχει ως αποτέλεσμα οι παράμετροι να αποκτούν διαφορετικές τιμές και συμβάλλει στην ανθεκτικότητα του δικτύου κατά του overfitting. Κατά τη διαδικασία ελέγχου, το δίκτυο χρησιμοποιείται ολόκληρο, χωρίς τις τυχαίες αφαιρέσεις των νευρώνων, αλλά με πολλαπλασιασμό των αποτελεσμάτων κατά έναν παράγοντα επι 2, για να αντισταθμιστούν οι χαμένοι νευρώνες. Η χρήση του dropout αυξάνει μεν τις επαναλήψεις του δικτύου για να επιτευχθεί η σύγκλιση, αλλά μειώνει δραστικά το overfitting.

2.3.2 Περιγραφή του δικτύου GoogleNet

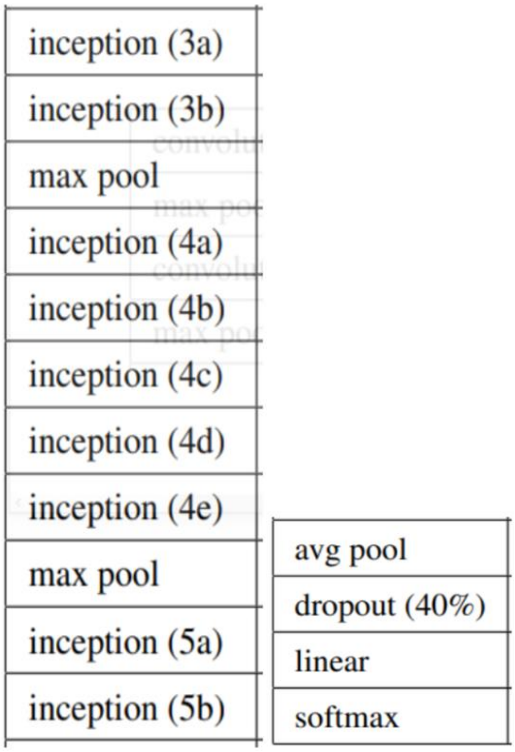
Το GoogleNet είναι ένα βαθύ συνελικτικό νευρωνικό δίκτυο με 22 επίπεδα που αποτελεί παραλλαγή του Inception Network, ενός Deep Convolutional Neural Network που αναπτύχθηκε από ερευνητές της Google. Η αρχιτεκτονική GoogLeNet που παρουσιάστηκε πρώτη φορά στο ImageNet Large-Scale Visual Recognition Challenge 2014 (ILSVRC 14) έλυσε εργασίες όρασης υπολογιστών, όπως ταξινόμηση εικόνας και ανίχνευση αντικειμένων. Σήμερα το GoogLeNet χρησιμοποιείται για πολλές εργασίες όρασης υπολογιστών, όπως ανίχνευση και αναγνώριση προσώπου, εκπαίδευση αντιπάλου κ.λπ.

Το Inception module είναι μια αρχιτεκτονική νευρωνικών δικτύων που αξιοποιεί την ανίχνευση χαρακτηριστικών σε διαφορετικές κλίμακες μέσω συνελίξεων με διαφορετικά φίλτρα και μειώνει το υπολογιστικό κόστος της εκπαίδευσης ενός εκτεταμένου δικτύου μέσω της μείωσης διαστάσεων.

Κατά τη δημιουργία του, η αρχιτεκτονική του GoogleNet σχεδιάστηκε για να είναι μια «μηχανή» με αυξημένη υπολογιστική απόδοση σε σχέση με μερικά από τα προηγούμενα ή παρόμοια δίκτυα της εποχής του. Μια μέθοδος που επιτυγχάνει η αποτελεσματικότητα του GoogLeNet είναι η μείωση της διάστασης της εικόνας εισόδου, διατηρώντας ταυτόχρονα σημαντικές χωρικές πληροφορίες για αυτή.

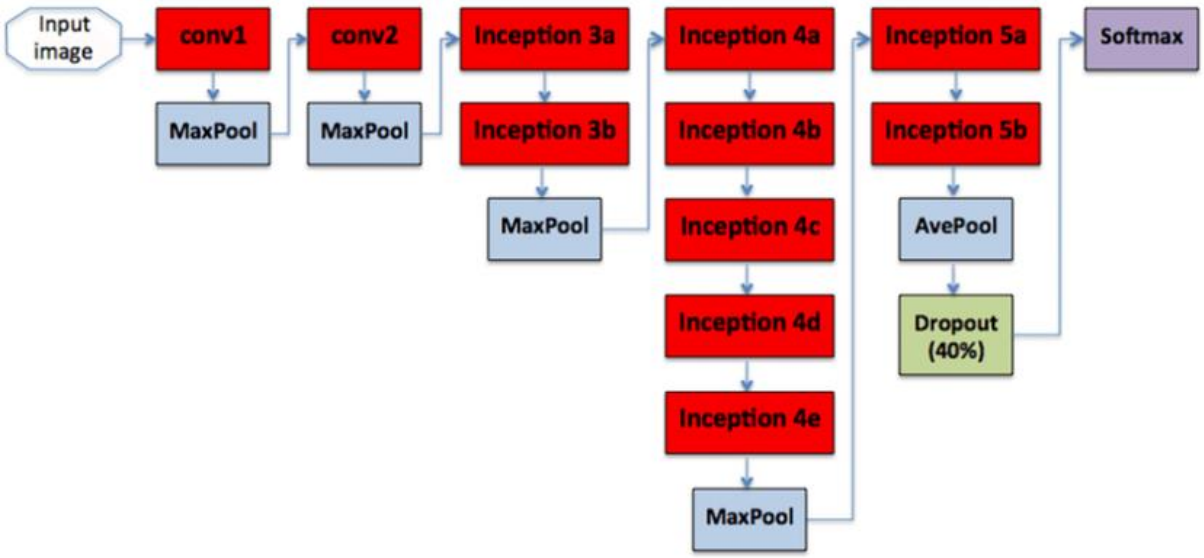
Το πρώτο συνελικτικό επίπεδο χρησιμοποιεί φίλτρο διάστασης 7x7, το οποίο είναι σχετικά μεγάλο σε σύγκριση με τα υπόλοιπα φίλτρα που βρίσκονται μέσα στο δίκτυο. Ο πρωταρχικός σκοπός αυτού του επιπέδου είναι να μειώσει αμέσως την διάσταση της εικόνας εισόδου, αλλά να μην χάσει τις χωρικές πληροφορίες της χρησιμοποιώντας μεγάλα μεγέθη φίλτρων.Το μέγεθος της εικόνας εισόδου μειώνεται με βάση έναν συντελεστή στο δεύτερο επίπεδο και ακόμη έναν συντελεστή πριν φτάσει στο επίπεδο έναρξης (first inception module). Ωστόσο καταγράφονται πολλά χαρακτηριστικά μέσω αυτής της διαδικασίας (δημιουργείται μεγάλος αριθμός feature maps).

Η αρχιτεκτονική του GoogleNet αποτελείται από 9 μοντέλα έναρξης όπως φαίνεται στην παρακάτω εικόνα. Υπάρχουν ακόμη, δύο επίπεδα επίπεδα max-pooling τα οποία χρησιμεύουν για να υποδεικνύουν την είσοδο στα δεδομένα καθώς αυτά προωθούνται μέσα στο δίκτυο. Αυτό επιτυγχάνεται μέσω της μείωσης της διαστασιμότητας των δεδομένων εισόδου. Η μείωση του μεγέθους εισόδου στην έναρξη είναι μια άλλη αποτελεσματική μέθοδος μείωσης του υπολογιστικού φορτίου του δικτύου.



Εικόνα 2.3.2.1 : Ο διαχωρισμών των επιπέδων της αρχιτεκτονικής GoogleNet.
(Πηγή: [1] https://arxiv.org/pdf/1409.4842v1.pdf)

Το average pooling επίπεδο παίρνει ένα μέσο όρο μεταξύ όλων των χαρακτηριστικών που δημιουργήθηκαν από το τελευταίο επίπεδο της έναρξης(inception 5b) και μειώνει το ύψος και το πλάτος των δεδομένων που εισάγονται σε 1x1.
Ένα dropout layer(40%) χρησιμοποιείται πριν από το γραμμικό επίπεδο (linear layer) . Το επίπεδο αυτό είναι μια τεχνική κανονικοποίησης που χρησιμοποιείται κατά τη διάρκεια της εκπαίδευσης για να αποφευχθεί το overfitting του δικτύου. Γενικότερα, η τεχνική Dropout λειτουργεί μειώνοντας τυχαία τον αριθμό των αλληλοσυνδεόμενων νευρώνων μέσα σε ένα νευρωνικό δίκτυο. Σε κάθε βήμα της εκπαίδευσης, κάθε νευρώνας έχει κάποια πιθανότητα να μείνει εκτός και να «αποσυνδεθεί» από τους υπόλοιπους νευρώνες. Το γραμμικό επίπεδο αποτελείται από 1000 κρυφές μονάδες (units), οι οποίες αντιστοιχούν στις 1000 κλάσεις που υπάρχουν στο σύνολο δεδομένων του Imagenet.
Το τελικό στρώμα είναι το στρώμα softmax. Αυτό το επίπεδο χρησιμοποιεί τη συνάρτηση ενεργοποίησης softmax, η οποία χρησιμοποιείται για να αντλήσει την κατανομή πιθανότητας ενός συνόλου αριθμών μέσα σε ένα διάνυσμα εισόδου. Το αποτέλεσμα αυτής της συνάρτησης είναι ένα διάνυσμα στο οποίο το σύνολο των τιμών του αντιπροσωπεύει την πιθανότητα εμφάνισης μιας κλάσης.



Εικόνα 2.3.2.2 : Η αρχιτεκτονική του GoogleNet
(Πηγή: [24] https://www.researchgate.net/figure/The-illustration-of-the-GoogleNet-architecture-Szegedy-et-al-2015-All-convolutional_fig1_314119821)

Υπάρχει ακόμα ένα στοιχείο που υλοποιήθηκε από τους δημιουργούς του δικτύου για την κανονικοποίηση και την αποφυγή του Overfitting. Αυτό το πρόσθετο στοιχείο είναι γνωστό ως Auxilary Classifier, δηλαδή βοηθητικός ταξινομητής.

Ένα πρόβλημα που προκύπτει συχνά στα εκτεταμένα δίκτυα είναι το "Vanishing gradient descent". Αυτό προκύπτει όταν η ενημέρωση που γίνεται στα βάρη του δικτύου από το backpropagation, είναι αμελητέα στα κατώτερα στρώματα καθώς η κλίση είναι πολύ μικρή. Ως αποτέλεσμα το δίκτυο σταματά να μαθαίνει. Οι Auxilary Classifiers προστίθενται στα ενδιάμεσα επίπεδα του δικτύου, και συγκεκριμένα στο 3ο και το 6ο (Inception 4a και Inception4d αντίστοιχα). Οι Auxilary Classifiers χρησιμοποιούνται μόνο κατά τη διάρκεια της εκπαίδευσης και αφαιρούνται κατά τη διάρκεια των συμπερασμάτων. Σκοπός τους είναι να εκτελέσει μια ταξινόμηση με βάση τις εισόδους στο μεσαίο τμήμα του δικτύου και να προσθέσει την απώλεια που υπολογίζεται κατά τη διάρκεια της εκπαίδευσης πίσω στη συνολική απώλεια του δικτύου. Ένας Auxilary Classifier αποτελείται από ένα average pool layer, ένα conv layer, δύο fully connected layers, ένα dropout layer(70%), και τέλος ένα linear layer με μια συνάρτηση ενεργοποίησης softmax. [25]

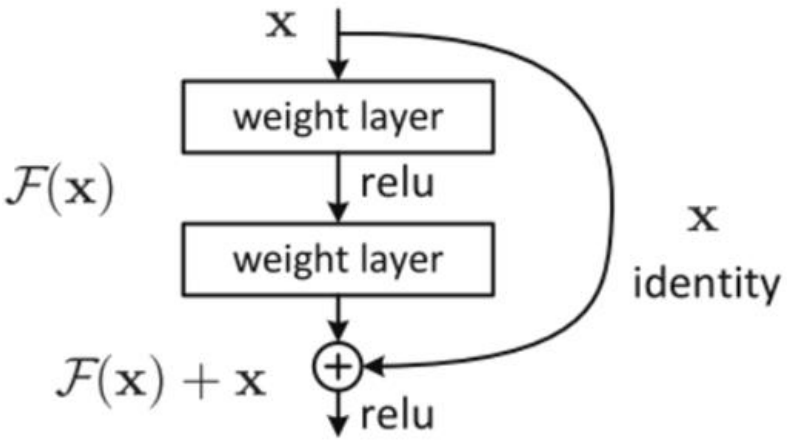
2.3.3 Περιγραφή του δικτύου ResNet (Residual Network)

Το μοντέλο του ResNet ήταν εξαιρετικά επιτυχημένο, αφού κέρδισε την πρώτη θέση στον διαγωνισμό ταξινόμησης ILSVRC 2015 με σφάλμα μόλι 3,57%. Επίσης συμμετείχε για πρώτη φορά στον διαγωνισμό του ImageNet detection, ImageNet localization, COCO detection, και COCO segmentation στους διαγωνισμούς ILSVRC & COCO το 2015.

Γενικότερα τα τελευταία χρόνια, υπήρχε μια κοινή τάση στην ερευνητική κοινότητα, πως τα νευρωνικά δίκτυα θα πρέπει να γίνουν βαθύτερα. Από το AlexNet, που η αρχιτεκτονική του είχε μόνο 5 επίπεδα, το δίκτυο VGG και το GoogleNet έχουν 19 και 22 επίπεδα αντίστοιχα. Ωστόσο, η αύξηση του βάθους του δικτύου δεν λειτουργεί απλά στοιβάζοντας επίπεδα μεταξύ τους. Τα βαθιά

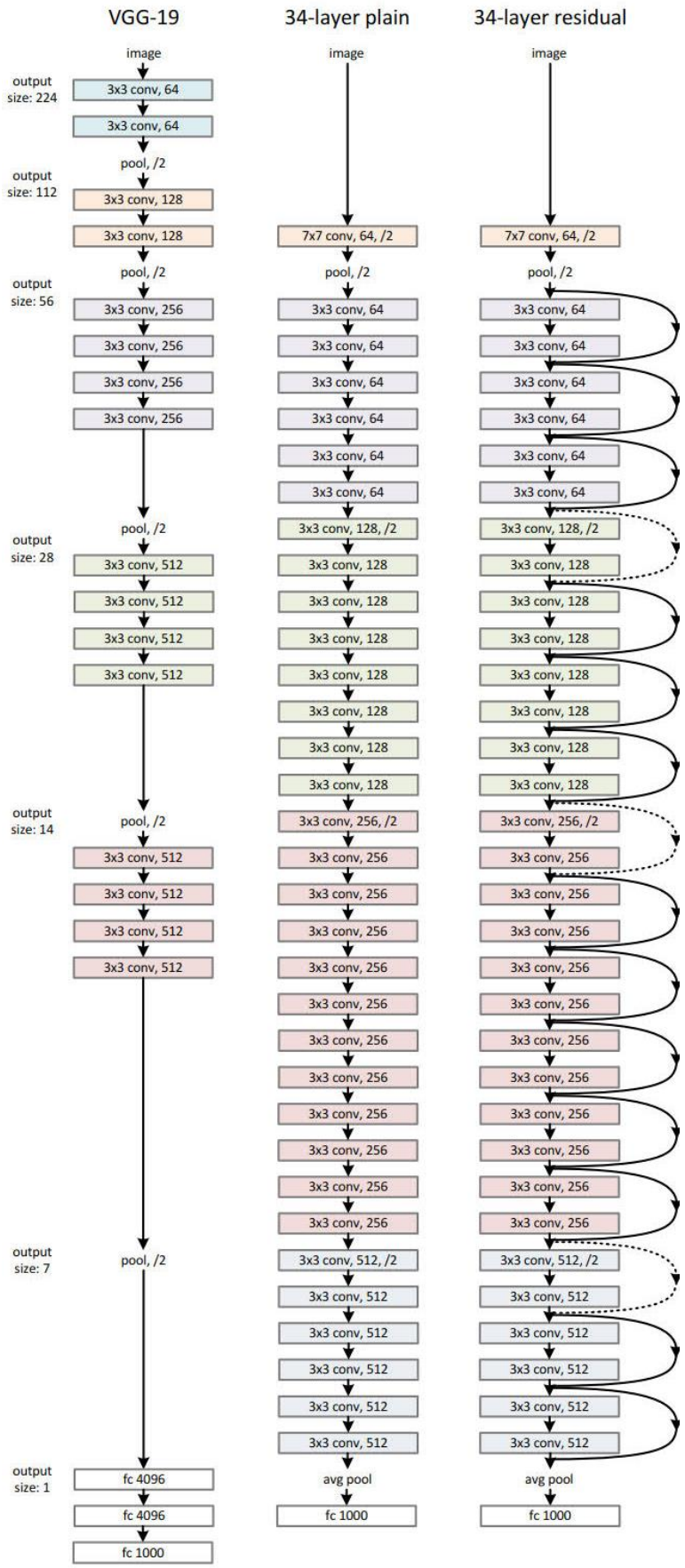
δίκτυα είναι δύσκολο να εκπαιδευτούν λόγω του διαβόητου προβλήματος vanishing gradient, που προαναφέρθηκε.

Η βασική ιδέα του ResNet είναι η εισαγωγή μιας λεγόμενης "σύνδεσης συντόμευσης ταυτότητας" (identity shortcut connection) που παραλείπει ένα ή περισσότερα επίπεδα, όπως φαίνεται στο παρακάτω σχήμα:



Εικόνα 2.3.3.1 : Ένα residual μπλοκ
(Πηγή : [2] https://arxiv.org/pdf/1512.03385.pdf)

Στην αρχιτεκτονική του ResNet υπάρχει ένα απλό δίκτυο 34αρων επιπέδων που είναι εμπνευσμένο από το VGG-19 στο οποίο προστίθενται η συντόμευση ή οι παραλείψεις. Αυτές οι παραλείψεις συνδέσεων ή τα residual μπλοκ μετατρέπουν την αρχιτεκτονική στο υπόλοιπο δίκτυο όπως φαίνεται παρακάτω :



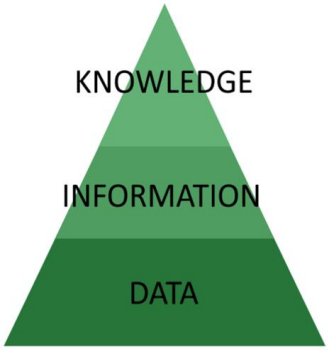
Εικόνα 2.3.3.2 : Παράδειγμα αρχιτεκτονική δικτύου για το ImageNet. Το VGG-19, ένα απλό δίκτυο με 34 επίπεδα και ένα residual δίκτυο με 34 επίπεδα. (Πηγή : [2] https://arxiv.org/pdf/1512.03385.pdf)

Η πρώτη αρχιτεκτονική ResNet ήταν η Resnet-34 η οποία περιελάμβανε την εισαγωγή shortcut συνδέσεων για τη μετατροπή ενός απλού δικτύου στο αντίστοιχο του residual-υπολειπόμενου δικτύου. Έτσι το απλό δίκτυο εμπνεύστηκε από τα νευρωνικά δίκτυα VGG, με τα συνελικτικά δίκτυα να έχουν φίλτρα διάστασης [3×3]. Ωστόσο, σε σύγκριση με τα δίκτυα VGG, το ResNets έχει λιγότερα φίλτρα και χαμηλότερη πολυπλοκότητα.

Ένα ακόμη μοντέλο που χρησιμοποιείται πολύ στην σημερινή εποχή είναι το ResNet50. Ενώ η αρχιτεκτονική Resnet50 βασίζεται στο παραπάνω μοντέλο, υπάρχει μια σημαντική διαφορά. Σε αυτή την περίπτωση, το "δομικό" στοιχείο μετατράπηκε σε σχέδιο bottleneck-συμφόρησης λόγω ανησυχιών για τον χρόνο που χρειάζεται για την εκπαίδευση των επιπέδων. Έτσι χρησιμοποιήθηκε μια στοίβα από 3 επίπεδα έναντι των 2 που ήταν στο προηγούμενο μοντέλο. Επομένως, καθένα από τα μπλοκ 2 επιπέδων στο Resnet 34 αντικαταστάθηκε με ένα μπλοκ 3ων επιπέδων, σχηματίζοντας την αρχιτεκτονική Resnet 50 και πετυχαίνοντας πολύ μεγαλύτερη ακρίβεια.

2.4 Σετ Δεδομένων και Δεδομένα Αναφοράς

Η παραδοσιακή αρχιτεκτονική για την ανάπτυξη συστημάτων τεχνητής νοημοσύνης περιλαμβάνει τη διάκριση μεταξύ δεδομένων, πληροφοριών, γνώσης και την διάταξή τους σε μια ιεραρχική δομή:



Εικόνα 2.4.1.1 : Αρχιτεκτονική για την ανάπτυξη συστημάτων τεχνητής νοημοσύνης.
(Πηγή : [26] https://www.baeldung.com/cs/ml-labeled-vs-unlabeled-data **)**

Σύμφωνα με αυτό το μοντέλο, ένα σύστημα τεχνητής νοημοσύνης κατανοεί τον κόσμο συγκεντρώνοντας δεδομένα σε πληροφορίες, στη συνέχεια επεξεργάζεται τις πληροφορίες για να αποσπάσει τη γνώση και στη συνέχεια χρησιμοποιεί αυτή τη γνώση για να κατευθύνει τη μεταγενέστερη συλλογή δεδομένων.

Με τον όρο δεδομένα θεωρείται κάθε ανεπεξέργαστο γεγονός, αξία, κείμενο, ήχος ή εικόνα που δεν ερμηνεύεται και δεν αναλύεται. Τα δεδομένα είναι το πιο σημαντικό μέρος όλων των εφαρμογών στην ανάλυση δεδομένων (Data Analytics), στην μηχανική μάθηση (Machine Learning) και την τεχνητή νοημοσύνη (Artificial Intelligence). Χωρίς δεδομένα, δεν μπορεί να εκπαιδευτεί κανένα μοντέλο και όλες οι σύγχρονες έρευνες και οι αυτοματισμοί θα ήταν μάταιοι. Οι μεγάλες

επιχειρήσεις ξοδεύουν πολλά χρήματα μόνο και μόνο για να συγκεντρώσουν όσο το δυνατόν περισσότερα δεδομένα.

Οι πληροφορίες στην συνέχεια, είναι δεδομένα που έχουν ερμηνευτεί και επεξεργαστεί και έχουν πλέον κάποιο ουσιαστικό συμπέρασμα για τους χρήστες. Οι πληροφορίες έχουν καλύτερη ικανότητα να συνοψίζουν την πολυπλοκότητα της εξωτερικής πραγματικότητας από τα ακατέργαστα δεδομένα και αυτός είναι ο λόγος που τοποθετούνται σε υψηλότερο επίπεδο στην πυραμίδα.

Η γνώση είναι συνδυασμός συμπερασματικών πληροφοριών, εμπειριών, μάθησης και ιδεών. Μόλις εξαχθούν μοτίβα σε ένα σύνολο δεδομένων, μπορούν να χρησιμοποιηθούν για να προβλέψουν τη μελλοντική κατάσταση του κόσμου, που προκύπτει από τις ενέργειες του συστήματος. Έτσι, μπορεί να γενικευθεί η γνώση που αποκτήθηκε για την εξαγωγή αυτού του συγκεκριμένου μοτίβου σε μελλοντικές, αθέατες καταστάσεις.

Τα δεδομένα λοιπόν χωρίζονται σε δύο βασικές κατηγορίες : τα επισημασμένα δεδομένα ή τα δεδομένα με ετικέτες, και τα μη επισημασμένα ή δεδομένα χωρίς ετικέτες.

2.4.1 Σετ Δεδομένων χωρίς δεδομένα αναφοράς, χωρίς ετικέτες (Unlabeled Data)

Εάν ενεργοποιηθεί ένας αισθητήρας ή αν γίνουν παρατηρήσεις χωρίς να είναι γνωστό τίποτα για το περιβάλλον ή τον τρόπο με τον οποίο λειτουργεί ο κόσμος, τότε συλλέγονται δεδομένα χωρίς ετικέτα. Μπορεί ακόμη να οριστούν και ως κομμάτια δεδομένων που δεν έχουν επισημανθεί με ετικέτες που προσδιορίζουν χαρακτηριστικά, ιδιότητες ή ταξινομήσεις. Τα δεδομένα χωρίς ετικέτα δεν έχουν ετικέτες ή στόχους για πρόβλεψη, αλλά μόνο χαρακτηριστικά που τα αντιπροσωπεύουν. Έτσι, τα μη επισημασμένα δεδομένα είναι βασικά ακατέργαστα δεδομένα που δεν έχουν «σημειωθεί» από ειδικούς.

2.4.2 Σετ Δεδομένων με δεδομένα αναφοράς, με ετικέτες (Labeled Data)

Τα επισημασμένα δεδομένα είναι δεδομένα που υπόκεινται σε προηγούμενη κατανόηση του τρόπου με τον οποίο λειτουργεί ο κόσμος. Ένας άνθρωπος ή αυτόματο μηχάνημα, πρέπει να χρησιμοποιήσει τις γνώσεις του για να επιβάλλει πρόσθετες πληροφορίες στα δεδομένα. Ωστόσο, αυτή η γνώση δεν υπάρχει στις μετρήσεις που πραγματοποιούνται. Χαρακτηριστικά παραδείγματα δεδομένων με ετικέτα είναι: μια εικόνα γάτας ή σκύλου, με σχετική ετικέτα «γάτα» ή «σκύλος», περιγραφή κειμένου για τον έλεγχο ενός προϊόντος και η βαθμολογία του από έναν χρήστη για αυτό το προϊόν, τα χαρακτηριστικά ενός σπιτιού προς πώληση και η τιμή πώλησής του.

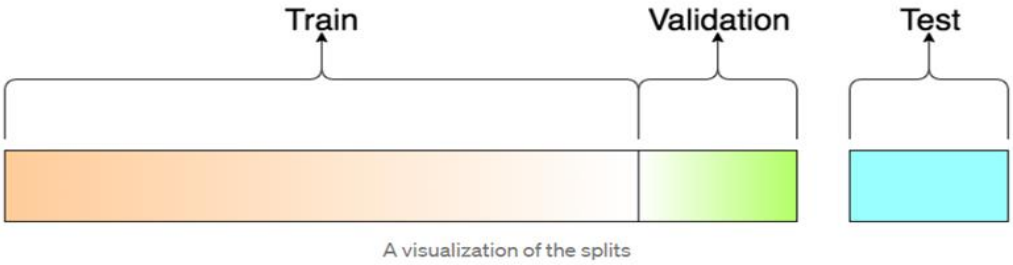
Εικόνα 2.4.3.1: Δεδομένα με ετικέτες και χωρίς ετικέτες.
(Πηγή : [27]https://labelyourdata.com/articles/unlabeled-data-in-machine-learning **)**

2.5 Δομή δεδομένων αναφοράς στη μηχανική μάθηση

Έκτος από τον διαχωρισμό των δεδομένων σε αυτά με ετικέτες και αυτά χωρίς ετικέτες, τα δεδομένα που χρησιμοποιούνται για τους αλγορίθμους της μηχανικής μάθησης χωρίζονται σε ακόμα 3 κατηγορίες προκειμένου να γίνουν ακριβείς προβλέψεις . Υπάρχουν τα δεδομένα εκπαίδευσης, τα δεδομένα δοκιμών και τα δεδομένα επικύρωσης, ή αλλιώς, train, test and validation data.

- **Training data** : αυτή η κατηγορία δεδομένων δημιουργεί τον αλγόριθμο της μηχανικής μάθησης. Τα δεδομένα αυτά δίνονται στον αλγόριθμο σαν είσοδος και αυτά αντιστοιχούν σε μια αναμενόμενη έξοδο. Το μοντέλο αξιολογεί επανειλημμένα τα δεδομένα για να μάθει περισσότερα για τη συμπεριφορά τους και στη συνέχεια προσαρμόζεται αναλόγως προκειμένου να εξυπηρετήσει τον επιδιωκόμενο σκοπό του.
- **Validation data** : κατά τη διάρκεια της εκπαίδευσης του μοντέλου, τα δεδομένα επικύρωσης είναι νέα δεδομένα που εισάγονται στο μοντέλο και δεν έχουν αξιολογηθεί στο παρελθόν. Τα δεδομένα αυτά αποτελούν την πρώτη δοκιμή σε νέα-«ξένα» δεδομένα και έτσι επιτρέπουν στους επιστήμονες να αξιολογήσουν πόσο καλά το μοντέλο κάνει προβλέψεις με βάση νέα δεδομένα. Δεν χρησιμοποιούν όλοι οι επιστήμονες δεδομένα επικύρωσης, όμως αυτά μπορούν να παρέχουν ορισμένες χρήσιμες πληροφορίες για τη βελτιστοποίηση των υπερπαραμέτρων, οι οποίες επηρεάζουν τον τρόπο με τον οποίο το μοντέλο αξιολογεί τα δεδομένα.
- **Testing data** : αφού κατασκευαστεί το μοντέλο, τα δεδομένα δοκιμών επιβεβαιώνουν για άλλη μια φορά ότι μπορούν να γίνουν ακριβείς προβλέψεις. Εάν τα δεδομένα εκπαίδευσης και επικύρωσης έχουν ετικέτες για την παρακολούθηση της απόδοσης του μοντέλου, τα δεδομένα δοκιμών θα πρέπει να μην φέρουν ετικέτα. Τα δεδομένα δοκιμής παρέχουν έναν τελικό, πραγματικό κόσμο ενός αόρατου-«ξένου» συνόλου δεδομένων για να επιβεβαιώσουν ότι ο αλγόριθμος μηχανικής μάθησης εκπαιδεύτηκε αποτελεσματικά.

Οι αλγόριθμοι μηχανικής μάθησης λοιπόν, απαιτούν δεδομένα εκπαίδευσης για την επίτευξη ενός στόχου. Ο αλγόριθμος θα αναλύσει αυτό το σύνολο δεδομένων, θα ταξινομήσει τις εισόδους και τις εξόδους και στη συνέχεια θα το αναλύσει ξανά. Εκπαιδευμένος αρκετά, ένας αλγόριθμος θα απομνημονεύσει ουσιαστικά όλες τις εισόδους και τις εξόδους σε ένα σύνολο δεδομένων εκπαίδευσης. Ωστόσο, αυτό δημιουργεί πρόβλημα όταν χρειάζεται να λάβει υπόψη του δεδομένα από άλλες πηγές. Εδώ δίνουν την λυση τα δεδομένα επικύρωσης. Τα validation data παρέχουν έναν αρχικό έλεγχο ότι το μοντέλο μπορεί να πραγματοποιήσει χρήσιμες προβλέψεις σε πραγματικό περιβάλλον, δηλαδή δεδομένα που δεν έχει ξαναδεί το μοντέλο, κάτι που δεν μπορούν να κάνουν τα δεδομένα εκπαίδευσης. Ο αλγόριθμος τότε μπορεί να αξιολογήσει δεδομένα εκπαίδευσης και δεδομένα επικύρωσης ταυτόχρονα.[29]



Εικόνα 2.5.1 : Διαχωρισμός των δεδομένων για αλγορίθμους μηχανικής μάθησης
(Πηγή : [28] https://towardsdatascience.com/train-validation-and-test-sets-72cb40cba9e7)

2.6 Προβλήματα Μηχανικής Μάθησης

Η μηχανική μάθηση είναι ένα μεγάλο πεδίο μελέτης που επικαλύπτεται και κληρονομεί ιδέες από πολλούς συναφείς τομείς όπως η τεχνητή νοημοσύνη. Το επίκεντρο του πεδίου είναι η μάθηση, δηλαδή η απόκτηση δεξιοτήτων ή γνώσεων με βάση την εμπειρία. Συνήθως, αυτό σημαίνει σύνθεση χρήσιμων εννοιών από δεδομένα. Ως εκ τούτου, υπάρχουν πολλοί διαφορετικοί τύποι μάθησης που μπορεί κανείς να συναντήσει. Οι τρεις βασικοί τύποι είναι : η επιβλεπόμενη μάθηση, η μη επιβλεπόμενη μάθηση και η ενισχυτική μάθηση. Ωστόσο υπάρχουν και άλλες ευρείες τεχνικές που χρησιμοποιούνται όλο και περισσότερο.

2.6.1 Επιβλεπόμενη μάθηση (Supervised Learning)

Η επιβλεπόμενη μάθηση, γνωστή και ως επιβλεπόμενη μηχανική μάθηση, είναι υποκατηγορία της μηχανικης μάθησης και της τεχνητής νοημοσύνης. Ορίζεται από την χρήση επισημασμένων δεδομένων με ετικέτες (labeled datasets) για να εκπαιδεύσουν αλγορίθμους να ταξινομούν δεδομένα ή να προβλέπουν αποτελέσματα με ακρίβεια. Καθώς τα δεδομένα εισόδου τροφοδοτούνται στο μοντέλο, προσαρμόζονται τα βάρη του έως ότου το μοντέλο να προσαρμοστεί κατάλληλα, το οποίο είναι μέρος της διαδικασίας επικύρωσης (validation process). Η επιβλεπόμενη μάθηση χρησιμοποιεί ένα εκπαιδευτικό σετ-σύνολο δεδομένων για να διδάξει τα μοντέλα να αποδώσουν το επιθυμητό αποτέλεσμα. Αυτό το σύνολο δεδομένων περιλαμβάνει δεδομένα εισόδου και σωστά αποτελέσματα εξόδου, το οποίο επιτρέπει στο μοντέλο να μαθαίνει με την πάροδο του χρόνου. Ο αλγόριθμος μετρά την ακρίβεια του μοντέλου μέσω της συνάρτησης απώλειας (loss function), προσαρμόζοντάς την μέχρι να ελαχιστοποιηθεί επαρκώς το σφάλμα.

Η επιβλεπόμενη μάθηση μπορεί να χωριστεί σε 2 τύπους προβλημάτων, την ταξινόμηση και την παλινδρόμηση :

- Ταξινόμηση: χρησιμοποιεί έναν αλγόριθμο για να αντιστοιχίζει με ακρίβεια τα τεστ δεδομένα σε συγκεκριμένες κατηγορίες. Αναγνωρίζει συγκεκριμένες οντότητες μέσα στο σύνολο δεδομένων και προσπαθεί να βγάλει κάποια συμπεράσματα για το πώς αυτές οι οντότητες πρέπει να επισημανθούν ή να οριστούν. Οι συνήθεις αλγόριθμοι ταξινόμησης είναι οι γραμμικοί ταξινομητές, οι μηχανές υποστήριξης διανυσμάτων (SVM), τα «δέντρα αποφάσεων» (decision trees), ο πλησιέστερος γείτονας (k-nearest neighbor) και το «τυχαίο δάσος» (random forest).
- Παλινδρόμηση : χρησιμοποιείται για την κατανόηση της σχέσης μεταξύ εξαρτημένων και ανεξάρτητων μεταβλητών. Συνήθως χρησιμοποιείται για την πραγματοποίηση προβολών, όπως για έσοδα από πωλήσεις για μια συγκεκριμένη επιχείρηση. Η γραμμική παλινδρόμηση, η υλικοτεχνική παλινδρόμηση και η πολυωνυμική παλινδρόμηση είναι οι πιο δημοφιλείς αλγόριθμοι παλινδρόμησης. [30]

2.6.2 Μη επιβλεπόμενη μάθηση (Unsupervised Learning)

Η μη επιβλεπόμενη μάθηση, γνωστή και ως μη εποπτευόμενη μηχανική μάθηση, χρησιμοποιεί αλγόριθμους μηχανικής μάθησης για να αναλύσει και να ομαδοποιήσει σύνολα δεδομένων χωρίς ετικέτα (Unlabeled datasets). Αυτοί οι αλγόριθμοι ανακαλύπτουν κρυφά μοτίβα ή ομαδοποιήσεις δεδομένων χωρίς την ανάγκη ανθρώπινης παρέμβασης. Η ικανότητά του να ανακαλύπτει ομοιότητες και διαφορές στις πληροφορίες που του δίνονται, το καθιστά την ιδανική λύση για διερευνητικές αναλύσεις δεδομένων, στρατηγικές διασταυρωμένων πωλήσεων (cross-selling strategies), τμηματοποίηση πελατών (customer segmentation) και αναγνώριση εικόνας. Τα μοντέλα μάθησης χωρίς επίβλεψη χρησιμοποιούνται για τρεις κύριες εργασίες: ομαδοποίηση, συσχέτιση και μείωση διαστάσεων.

- Ομαδοποίηση (Clustering) : είναι μια τεχνική η οποία ομαδοποιεί δεδομένα χωρίς ετικέτα με βάση τις ομοιότητες ή τις διαφορές τους. Οι αλγόριθμοι ομαδοποίησης χρησιμοποιούνται για την επεξεργασία ακατέργαστων, μη ταξινομημένων δεδομένων σε ομάδες που αντιπροσωπεύονται από δομές ή μοτίβα στις πληροφορίες. Οι αλγόριθμοι ομαδοποίησης κατηγοριοποιούνται σε: αποκλειστικούς, επικαλυπτομενος, ιεραρχικούς και πιθανολογικούς.
- Συσχέτιση (Association) : ένας κανόνας συσχέτισης είναι μια μέθοδος βασισμένη σε κανόνες για την εύρεση σχέσεων μεταξύ μεταβλητών σε ένα δοσμένο σύνολο δεδομένων. Αυτές οι μέθοδοι χρησιμοποιούνται συχνά για την ανάλυση του καλαθιού στις αγορές, επιτρέποντας στις εταιρείες να κατανοήσουν καλύτερα τις σχέσεις μεταξύ διαφορετικών προϊόντων. Η κατανόηση των καταναλωτικών συνηθειών των πελατών επιτρέπει στις επιχειρήσεις να αναπτύξουν καλύτερες στρατηγικές διασταύρωσης και μηχανές συστάσεων για νέα προϊόντα.
- Μείωση διαστάσεων (Dimensionality reduction) : περισσότερα δεδομένα αποδίδουν πιο ακριβή αποτελέσματα, επηρεάζουν όμως την απόδοση των αλγορίθμων (π.χ. overfitting-υπερφόρτωση) και μπορεί επίσης να δυσχεράνει την απεικόνιση συνόλων δεδομένων. Η μείωση διαστάσεων είναι μια τεχνική που χρησιμοποιείται όταν ο αριθμός των

χαρακτηριστικών ή των διαστάσεων σε ένα δοσμένο σύνολο δεδομένων είναι πολύ υψηλός. Μειώνει τον αριθμό των δεδομένων που εισάγονται σε ένα διαχειρίσιμο μέγεθος διατηρώντας παράλληλα την ακεραιότητα του συνόλου δεδομένων όσο το δυνατόν περισσότερο. [31]

2.6.3 Ενισχυτική μάθηση (Reinforcement Learning)

Η ενισχυτική μάθηση περιγράφει μια κατηγορία προβλημάτων όπου ένας «πράκτορας» λειτουργεί σε ένα περιβάλλον και πρέπει να μάθει να λειτουργεί χρησιμοποιώντας τις συνεχείς πληροφορίες που του δίνονται. Η ενισχυτική μάθηση λοιπόν είναι η διαδικασία για να αντιστοιχηθούν οι καταστάσεις σε ενέργειες, έτσι ώστε να μεγιστοποιηθεί ένα αριθμητικό σήμα ανταμοιβής. Ο μαθητής δεν ενημερώνεται ποιες ενέργειες πρέπει να κάνει, αλλά αντίθετα πρέπει να ανακαλύψει ποιες ενέργειες αποφέρουν την μεγαλύτερη ανταμοιβή δοκιμαζοντας τις.

Όπως και στην επιβλεπόμενη μάθηση, το μοντέλο έχει κάποια απάντηση από την οποία μπορεί να μάθει, αν και η ανατροφοδότηση μπορεί να καθυστερεί και να είναι στατιστικά θορυβώδης, οπότε είναι δύσκολο για το μοντέλο να συνδέσει αιτία και αποτέλεσμα.

2.6.4 Ημι-επιβλεπόμενη μάθηση (Semi-supervised Learning)

Τα όρια μεταξύ της μη επιβλεπόμενης και της επιβλεπόμενης μάθησης είναι θολά και υπάρχουν πολλές «υβριδικές» προσεγγίσεις που εξάγονται από κάθε πεδίο μελέτης. Η ημι-επιβλεπόμενη μάθηση είναι ουσιαστικά επιβλεπόμενη μάθηση όπου τα δεδομένα εκπαίδευσης περιέχουν πολύ λίγα επισημασμένα δεδομένα (labeled data) και πολλά δεδομένα χωρίς ετικέτες (unlabeled data). Ο στόχος ενός ημι-επιβλεπόμενου μοντέλου μάθησης είναι να κάνει αποτελεσματική χρήση όλων των διαθέσιμων δεδομένων, όχι μόνο των επισημασμένων δεδομένων όπως συμβαίνει στην επιβλεπόμενη μάθηση.

Η αποτελεσματική χρήση μη επισημασμένων δεδομένων μπορεί να απαιτεί τη χρήση μεθόδων χωρίς επίβλεψη, όπως ομαδοποίηση και εκτίμηση πυκνότητας. Εφόσον ανακαλυφθούν ομάδες ή μοτίβα, επιβλεπόμενες μέθοδοι ή ιδέες από την επιβλεπόμενη μάθηση μπορούν να χρησιμοποιηθούν για να επισημανθούν τα δεδομένα χωρίς ετικέτα και στην συνέχεια αυτά να χρησιμοποιηθούν για προβλέψεις. [32]

Εικόνα 2.6.4.1 : Περιγραφή των διαφορετικών προσεγγίσεων μηχανικής μάθησης
(Πηγή : [33] https://business.blogthinkbig.com/semi-supervised-learning-the-great-unknown/ **)**

2.7 Αυτο-επιβλεπόμενη Μάθηση (Self-supervised learning)

Η αυτο-επιβλεπόμενη μάθηση αναφέρεται σε ένα πρόβλημα μάθησης χωρίς επίβλεψη που διαμορφώνεται ως πρόβλημα επιβλεπόμενη μάθησης προκειμένου να εφαρμοστούν αντίστοιχοι αλγόριθμοι για την επίλυσή του. Επιβλεπόμενοι αλγόριθμοι μάθησης (supervised learning algorithms) χρησιμοποιούνται για την επίλυση μιας εναλλακτικής εργασίας ή προπαρασκευαστικής εργασίας (pretext task), το αποτέλεσμα της οποίας είναι ένα μοντέλο που μπορεί να χρησιμοποιηθεί στην επίλυση του αρχικού (πραγματικού) προβλήματος μοντελοποίησης.

Στην αυτο-επιβλεπόμενη μάθηση, τα μοντέλα αν και εν τέλει πρόκειται να επιλύσουν ένα πρόβλημα κλασικής επιβλεπόμενης ταξινόμησης, αντί να χρησιμοποιούν αυτές καθ' αυτές τις ετικέτες αυτού του προβλήματος, χρησιμοποιούν ετικέτες που έχουν δημιουργηθεί αυτόματα και αναφέρονται σε κάποια υποκατάστατη εργασία. Η υποκατάστατη εργασία έχει σχεδιαστεί για να αποδίδει γενικά χαρακτηριστικά που είναι περιγραφικά και εύρωστα σε τυπικές διακυμάνσεις που γίνονται στα δεδομένα. Έτσι κάθε εικόνα αποτελεί την δική της κλάση και άλλα δείγματα δημιουργούνται με εφαρμογή διαφορετικών augmentation στην αρχική εικόνα.

Ένας άλλος τρόπος δημιουργίας ετικετών είναι η χρήση του αλγορίθμου k-means. Σε αυτήν την περίπτωση οι κλάσεις καθορίζονται με την εφαρμογή της ομαδοποίησης k-means στις αναπαραστάσεις που εξάγονται από το αυτο-επιβλεπόμενη μοντέλο. Σε κάθε epoch καθορίζονται νέες κλάσεις, ενώ σε κάθε εικόνα εκχωρείται αυτόματα μια ετικέτα που αντιστοιχεί στην κοντινότερη συστάδα. [9]

Η απόδοση από τα βαθιά συνελικτικά νευρωνικά δίκτυα (CNNs) εξαρτάται σημαντικά από την ικανότητα και το ποσοστό των δεδομένων εκπαίδευσης. Διαφορετικά είδη από αρχιτεκτονικές δικτύων έχουν εφευρεθεί για να αυξήσουν την χωρητικότητα των μοντέλων των δικτύων, και όλο

και μεγαλύτερα σετ δεδομένων συλλέγονται στην σύγχρονη εποχή. Ποικίλα δίκτυα συμπεριλαμβανομένων και των AlexNet, VGG, GoogLeNet, ResNet, και DenseNet και μεγάλης κλίμακας σετ δεδομένων όπως το ImageNet και το OpenImage έχουν προταθεί. Με εξεζητημένη αρχιτεκτονική και μεγάλης κλίμακας δεδομένα, η απόδοση των ConvNets εξακολουθεί να είναι η τελευταία λέξη της τεχνολογίας για πολλές δοκιμασίες στην όραση των υπολογιστών. Παρόλα αυτά, η συλλογή και η επισήμανση αυτών των μεγάλης κλίμακας δεδομένων είναι πολύ χρονοβόρα και κοστοβόρα όπως προαναφέρθηκε. Για εξοικονόμηση χρόνου και πόρων λοιπόν, πολλές αυτό-επιβλεπόμενες μέθοδοι έχουν προταθεί για να εκπαιδευτούν σε οπτικά χαρακτηριστικά από μεγάλα σετ μη επισημασμένων εικόνων (unlabeled) χωρίς την ανθρώπινη παρουσία. Προκειμένου να «μάθουν» τα δίκτυα οπτικά χαρακτηριστικά από μη επισημασμένα δεδομένα, μια δημοφιλής λύση είναι η επίλυση προσχημάτων-pretext tasks. Υπάρχουν και άλλοι μέθοδοι που χρησιμοποιούνται στην αυτό-επιβλεπόμενη μάθηση, οι οποίες χρησιμοποιούν pretext tasks και θα αναλυθούν παρακάτω περαιτέρω.

Το γενικότερο σχήμα της αυτό-επιβλεπόμενη μάθησης λοιπον, περιγράφεται στο παρακάτω σχήμα. Κατά τη διάρκεια της αυτό-επιβλεπόμενης εκπαιδευτικής φάσης, ένα προκαθορισμένο pretext task σχεδιάζεται για να λυθεί από το συνελικτικό δίκτυο, και οι ψευδο-ετικέτες για το pretext task γεννιούνται αυτόματα βασισμένα σε κάποια χαρακτηριστικά των δεδομένων. Τότε το δίκτυο εκπαιδεύεται για να μάθει αντικειμενικές λειτουργίες του pretext task. Μετά το τέλος της αυτό-επιβλεπόμενης εκπαίδευσης, τα γνωστά οπτικά χαρακτηριστικά μπορούν να μεταφερθούν σε μεταγενέστερες εργασίες (downstream tasks) σαν προ-εκπαιδευμένα μοντέλα για να βελτιωθεί η απόδοση και να αποφευχθεί το overfitting.

Εικόνα 2.7.1 : Η γενική ιδέα της αυτό-επιβλεπόμενης μάθησης.
(Πηγή: [10] https://arxiv.org/abs/1902.06162)

Όπως φαίνεται στην εικόνα, τα οπτικά χαρακτηριστικά μαθαίνονται μέσω της διαδικασίας εκπαίδευσης του ConvNet για την επίλυση του pretext task. Αφού τελειώσει η εκπαίδευση οι παράμετροι μεταφέρονται σε άλλα μεταγενέστερα tasks και χρησιμοποιούνται ως προ-εκπαιδευμένα μοντέλα. Η απόδοση σε αυτά τα tasks χρησιμοποιείται για να αξιολογηθεί η ποιότητα της μάθησης των χαρακτηριστικών.

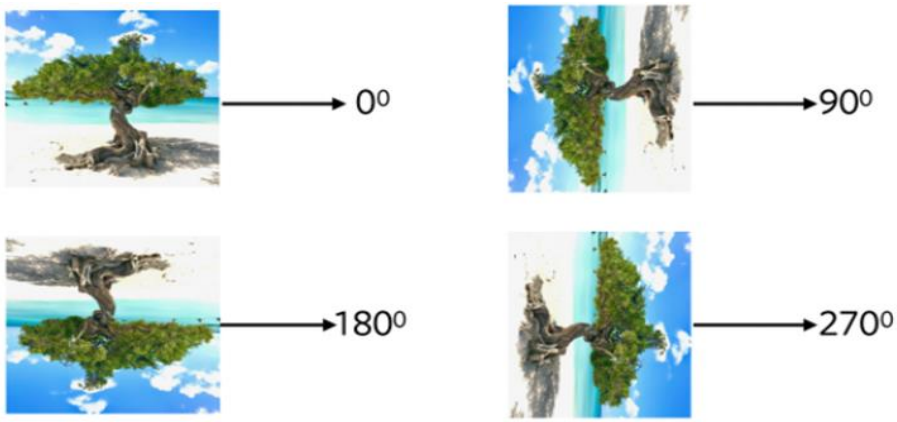
2.7.1 Pretext Tasks

Τα pretext tasks χρησιμοποιούνται στην αυτό-επιβλεπόμενη μάθηση προκειμένου να δημιουργηθούν «χρήσιμες» αναπαραστάσεις χαρακτηριστικών. Με τον όρο «χρήσιμες» νοούνται οι αναπαραστάσεις που θα πρέπει να μπορούν εύκολα να προσαρμοστούν σε άλλες εργασίες-tasks, που είναι άγνωστες κατά τη διάρκεια της εκπαίδευσης. Τα pretext tasks έχουν δύο κοινά: (1) τα οπτικά χαρακτηριστικά των εικόνων πρέπει να είναι αντιληπτά από τα συνελικτικά δίκτυα για να επιλυθούν αυτά τα tasks και (2) οι ψευδο-ετικέτες των pretext tasks μπορούν να δημιουργηθούν αυτόματα βασισμένα στα χαρακτηριστικά των εικόνων. Τα tasks αυτά, παρέχουν είτε τα χαρακτηριστικά που έμαθαν (learned feature representations) είτε τα βάρη του μοντέλου, για να χρησιμοποιηθούν στην συνέχεια στα λεγόμενα downstream tasks. Τα downstream tasks είναι εφαρμογές όρασης υπολογιστή που χρησιμοποιούνται για την αξιολόγηση της ποιότητας των χαρακτηριστικών που μαθαίνει η αυτο-επιβλεπόμενη μάθηση.

Κάποια από τα πιο διάσημα Pretext tasks είναι :

- **Image Rotation**

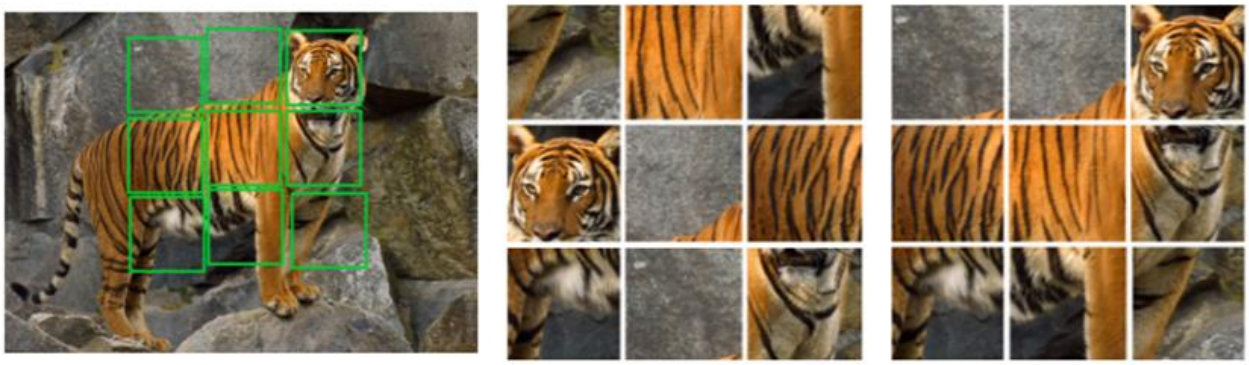
Σε αυτό το task η εικόνα υπόκειται σε έναν γεωμετρικό μετασχηματισμό. Πιο συγκεκριμένα η εικόνα περιστρέφεται κατά {0^o, 90°,180°, 270°}. Ένα συνελικτικό δίκτυο εκπαιδεύεται προκειμένου να μπορέσει να αναγνωρίσει την περιστροφή που έχει δεχτεί η εικόνα. Σαν αποτέλεσμα δίνεται μια πιθανότητα περιστροφής που προβλέπεται από το μοντέλο. Η βασική ιδέα είναι πως αν κάποιος δεν γνωρίζει τα αντικείμενα-χαρακτηριστικά που απεικονίζονται στην εικόνα, τότε δεν θα μπορέσει να αναγνωρίσει ούτε την περιστροφή που αυτή έχει δεχτεί [3].



Εικόνα 2.7.1.1: Περιστροφή εικόνας κατά {0^o, 90°,180°, 270°}
(Πηγή : [34] https://atcold.github.io/pytorch-Deep-Learning/en/week10/10-1/)

- **Jigsaw Puzzle**

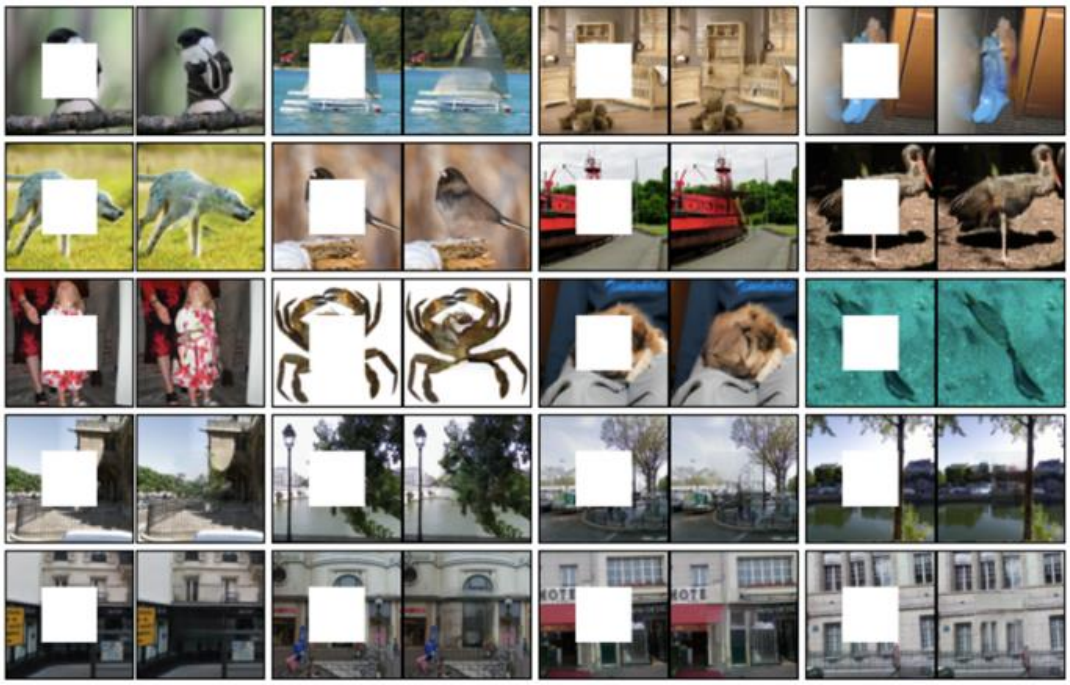
Σύμφωνα με αυτό το task, ένα κομμάτι της εικόνας ή ολόκληρη η εικόνα χωρίζεται σε 9 κομμάτια και στην συνέχεια αυτά τα κομμάτια ανακατεύονται μεταξυ τους δημιουργώντας ένα παζλ που πρέπει να λυθεί. Ορισμένα κομμάτια μπορεί να είναι άμεσα αναγνωρίσιμα σαν μέρη του αντικειμένου, όμως άλλα είναι διφορούμενα (π.χ. έχουν παρόμοια σχέδια) και η αναγνώρισή τους είναι πολύ πιο αξιόπιστη όταν όλα τα κομμάτια αξιολογούνται από κοινού [4].



Εικόνα 2.7.1.2 : Εκμάθηση χαρακτηριστικών εικόνας επιλύοντας Jigsaw Puzzle
(Πηγή : [4] https://arxiv.org/pdf/1603.09246.pdf)

- **Image Inpainting**

Το Image Inpainting αναφέρεται στη δημιουργία/αναδημιουργία της περιοχής μιας εικόνας που έχει αφαιρεθεί. Από μια εικόνα, αφαιρείται ένα μέρος της και το μοντέλο εκπαιδεύεται για να δημιουργήσει το κομμάτι της εικόνας που λείπει. Αυτή η τεχνική είναι μια πρωτοπόρα μέθοδος, που σημαίνει ότι το μοντέλο εκπαιδεύεται για να δημιουργεί πληροφορίες προκειμένου να μάθει τα χαρακτηριστικά της εικόνας. Είναι σημαντικό ο κωδικοποιητής να καταλάβει όλο το περιεχόμενο της εικόνας, καθώς και να παράγει μια πιθανή υπόθεση για το κομμάτι που λείπει. [35]



Εικόνα 2.7.1.3 : Image Inpainting Task
(Πηγή : [5] https://arxiv.org/pdf/1604.07379.pdf)

- **Colorization of grayscale images**

Ο χρωματισμός των εικόνων που έχουν μόνο τόνους του γκρι, είναι μια εύκολη διαδικασία για την ανθρώπινη φαντασία. Ωστόσο, απαιτείται υψηλό επίπεδο κατανόησης για αυτή την διαδικασία και γι αυτό ακριβώς τον λόγο, η ανάπτυξη των πλήρως αυτοματοποιημένων αλγορίθμων χρωματισμού αποτελούν μέχρι και σήμερα μια δοκιμασία για τους ερευνητές. Έχουν πραγματοποιηθεί πολλές έρευνες στο συγκεκριμένο πεδίο και υπάρχουν διαφορετικές προσεγγίσεις που χρησιμοποιούνται και σαν pretext tasks σε μεθόδους αυτό-επιβλεπόμενης μάθησης.

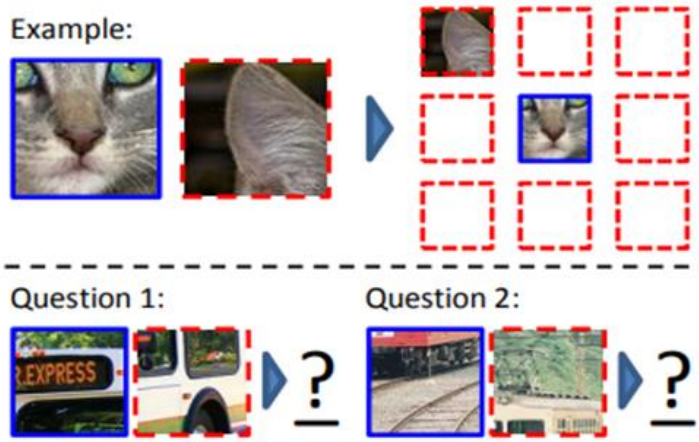
Μία από αυτές αναφέρεται στον προσδιορισμό χρωματικού ιστογράμματος για κάθε pixel της εικόνας. Στην μέθοδο αυτή αρχικά προσδιορίζονται σημαντικές πληροφορίες για την εικόνα, δηλαδή το τι αυτή περιέχει για παράδειγμα πρόσωπο, αυτοκίνητο, φυτά και άλλα και έπειτα προσδιορίζεται η τοπική τους θέση στην εικόνα. Τα νευρωνικά δίκτυα (CNNs) αποτελούν σημαντικό εργαλείο προκειμένου στην συνέχεια να παραχθεί και το επιθυμητό χρωματικό ιστόγραμμα για την αρχική γκρίζα εικόνα.

Ένα άλλο ενδιαφέρον πείραμα για τον χρωματισμό εικόνας, προτείνει μια πλήρη τυποποιημένη μέθοδο ρεαλιστικού και αληθοφανους χρωματισμού. Οι ερευνητές ενθαρρύνουν την αβεβαιότητα του προβλήματος με το να το παρουσιάζουν σαν task ταξινόμησης και χρησιμοποιούν μια επανεξισορρόπηση κλάσεων στο χρόνο εκπαίδευσης για να αυξήσουν την ποικιλομορφία των χρωμάτων στο αποτέλεσμα. [6],[7].

Εικόνα 2.7.1.4 : Ασπρόμαυρες φωτογραφίες που αποκτούν χρώματα με την μέθοδο του colorization.
(Πηγή: [6] https://arxiv.org/pdf/1603.08511.pdf)

- **Context Prediction**

Η εργασία αυτή έχει ομοιότητες με το παραπάνω jigsaw task ωστόσο αποτελεί ένα διαφορετικό έργο. Αρχικά γίνεται εξαγωγή τυχαίων κομματιών μιας εικόνας, και έπειτα μέσα από την εκπαίδευση ενός νευρωνικού δικτύου γίνεται η προσπάθεια να προβλεφθεί η σωστή θέση ενός κομματιού της εικόνας, σε σχέση με ένα αρχικό. Προκειμένου να λειτουργήσει σωστά ένα τέτοιο task απαιτείται από το μοντέλο να μάθει να αναγνωρίζει αντικείμενα και κομμάτια από αυτά.



Εικόνα 2.7.1.5 : Το task του Context Prediction
(Πηγή : [8] 1505.05192.pdf (arxiv.org))

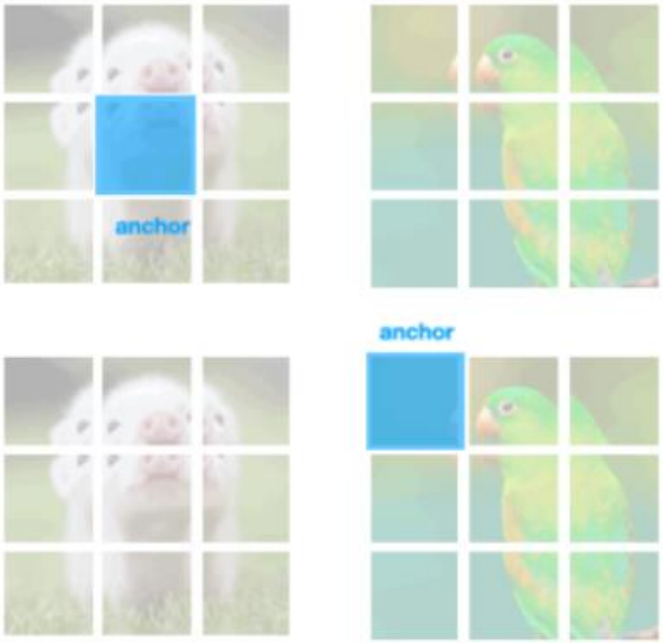
2.7.2 «Αντιθετική» μάθηση (Contrastive learning)

Στην περίπτωση του contrastive learning, το pretext task ορίζεται με τρόπο που οι παρόμοιες εικόνες θα πρέπει να βρίσκονται κοντά στον χώρο της αναπαράστασης, ενώ οι ανόμοιες εικόνες θα πρέπει να είναι όσο το δυνατόν πιο μακριά. Οι όμοιες εικόνες αποτελούν τα θετικά δείγματα, ενώ οι ανόμοιες τα αρνητικά. Έτσι το κύριο έργο, είναι πλέον το πως θα βρεθούν όμοιες και ανόμοιες εικόνες.

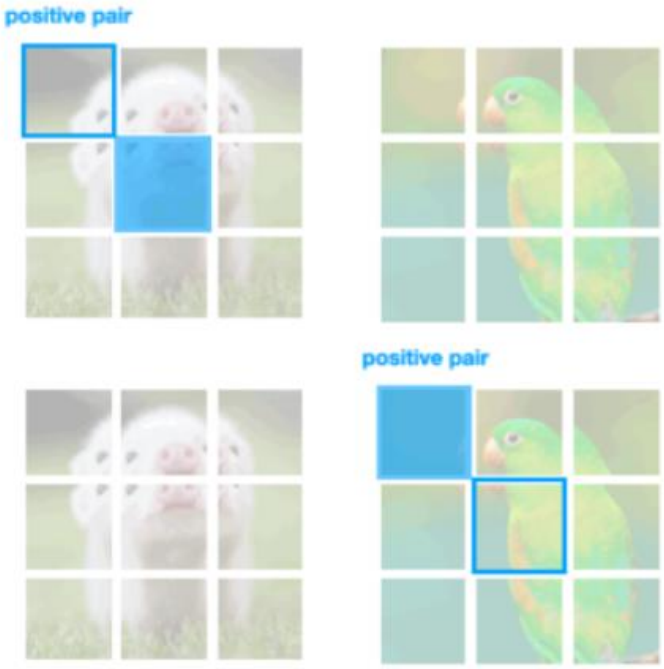
Εικόνα 2.7.2.1 : Παράδειγμα θετικου και αρνητικου ζευγαριού εικόνων. (Πηγή : [42] https://towardsdatascience.com/a-framework-for-contrastive-self-supervised-learning-and-designing-a-new-approach-3caab5d29619)

Παρακάτω παρουσιάζονται πειράματα που έχουν πραγματοποιηθεί προκειμένου να προκύψουν θετικά και αρνητικά ζεύγη [9]:

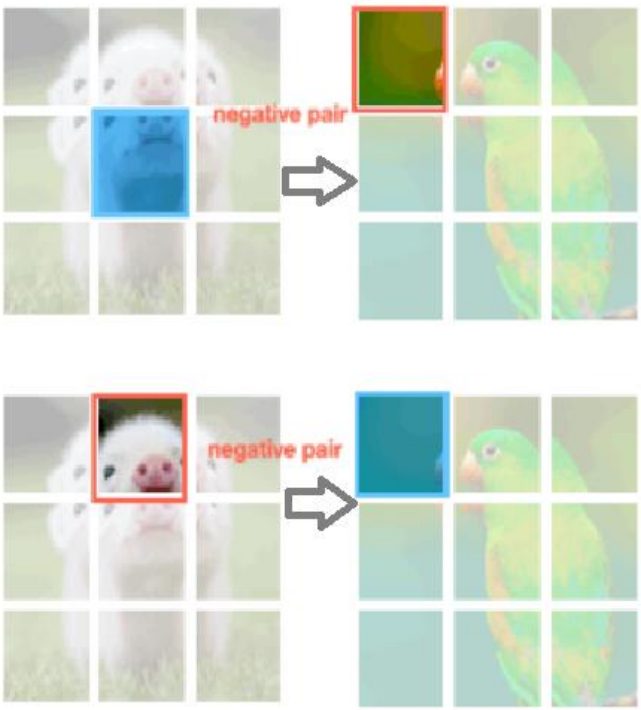
- Σε ένα από τα πειράματα που έχουν πραγματοποιηθεί στον συγκεκριμένο τομέα, οι ερευνητές χρησιμοποιούν κομμάτια απο μια αρχική εικόνα ως θετικά ζεύγη και κομμάτια που εξάγονται από διαφορετικές εικόνες, ως αρνητικά ζεύγη.



Εικόνα 2.7.2.2 : Στο 1ο στάδιο ένα κομμάτι της εικόνας θεωρείται ως «άγκυρα»-θετικό ζευγάρι.

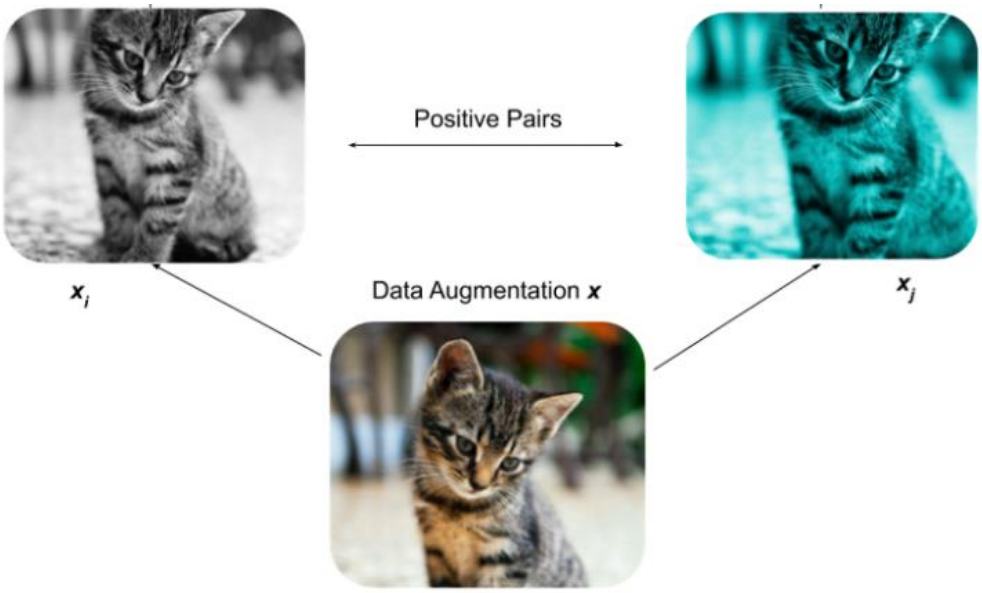


Εικόνα 2.7.2.3 : Στο 2ο στάδιο βρίσκονται τα θετικά ζευγάρια από την ίδια εικόνα.



Εικόνα 2.7.2.4 : Τέλος, στο 3ο στάδιο βρίσκονται τα αρνητικά ζεύγη από κομμάτια διαφορετικών εικόνων. (Πηγή: [43])

- Μία άλλη μέθοδος για δημιουργία θετικών ζευγών, βασίζεται στην ιδέα της εφαρμογής augmentation σε μια εικόνα (τυχαίες στροφές, jittering, αλλαγές στα χρώματα, κλπ.). Θετικά ζεύγη θεωρούνται οι augmented εικόνες που έχουν προκύψει από την ίδια εικόνα, ενώ αρνητικά ζεύγη όλα τα υπόλοιπα.

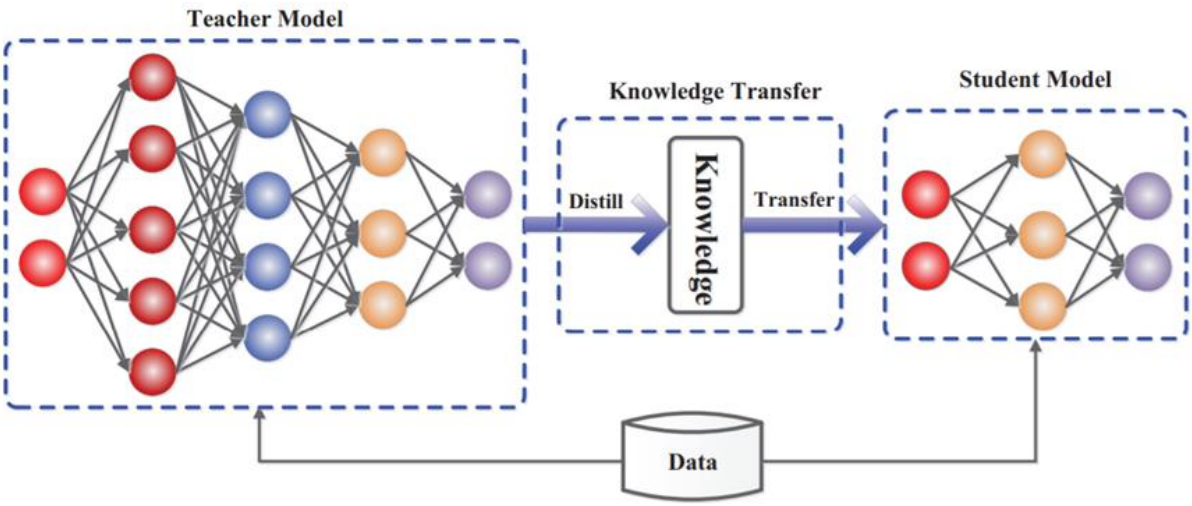


Εικόνα 2.7.2.5 :Παράδειγμα θετικού ζεύγος από εικόνα που έχει υποστεί κάποιο augmentation. (Πηγή:[43])

- Έχουν ακόμα προταθεί και άλλες μέθοδοι που βασίζονται στην χρήση αναπαραστάσεων εικόνων από διαφορετικά επίπεδα του νευρωνικού δικτύου ως θετικά δείγματα.

2.8 «Απόσταξη» Γνώσης (Knowledge Distillation)

Η μέθοδος της «απόσταξης» γνώσης αναφέρεται στην ιδέα της διδασκαλίας ενός μικρού δικτύου «μαθητή», βήμα προς βήμα, κατευθυνοντας το με βάση ένα μεγαλύτερο ήδη εκπαιδευμένο δίκτυο «δάσκαλο». Στην διαδικασία αυτή, υπάρχουν «μαλακές» ετικέτες (soft labels, δηλαδή οι εικόνες έχουν προσκολλημένη πάνω τους ετικέτα με πιθανότητα να ανήκουν σε κάποια ή κάποιες κλάσεις). Οι ετικέτες αυτές λοιπόν, αναφέρονται στους χάρτες χαρακτηριστικών εξόδου από το μεγαλύτερο δίκτυο μετά από κάθε επίπεδο συνέλιξης. Το μικρότερο δίκτυο στη συνέχεια εκπαιδεύεται για να μάθει την ακριβή συμπεριφορά του μεγαλύτερου δικτύου προσπαθώντας να αναπαράγει τις εξόδους του σε κάθε επίπεδο (όχι μόνο την τελική απώλεια/loss). Η μέθοδος αυτή επινοήθηκε καθώς σε πολλούς τομείς της βαθιάς μάθησης τα μοντέλα είναι πολύ ακριβά υπολογιστικά προκειμένου να λειτουργήσουν σε συσκευές όπως τηλέφωνα ή ενσωματωμένες συσκευές.



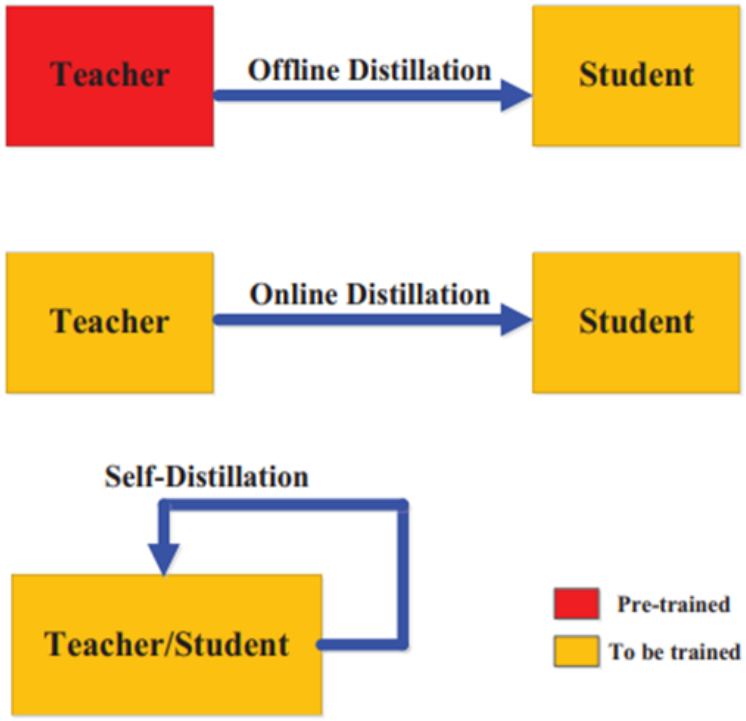
Εικόνα 2.8.1 : Το γενικότερο πλαίσιο του «δασκάλου-μαθητή» από την απόσταξη γνώσης .
(Πηγή: [11] https://arxiv.org/pdf/2006.05525.pdf)

Στην απόσταξη γνώσης λοιπόν, ο τύπος της γνώσης, οι στρατηγικές απόσταξης και οι αρχιτεκτονικές δασκάλου-μαθητή, παίζουν σημαντικό καθοριστικό ρόλο στην μάθηση του μικρού–δικτύου μαθητή. Ωστόσο, η μάθηση που μεταφέρεται από το μεγάλο- δίκτυο δάσκαλο στο μικρό-δίκτυο μαθητή χωρίζεται σε 3 κατηγορίες συμφωνα με τους επιστήμονες. Υπάρχει η γνώση με βάση την απάντηση (response-based knowledge), η γνώση βασισμένη σε χαρακτηριστικά (feature-based knowledge) και η βασισμένη σε σχέσεις γνώση (relation-based knowledge).

- Η response-based knowledge είναι η γνώση που αναφέρεται συνήθως στην νευρική απόκριση του τελευταίου επιπέδου εξόδου του μοντέλου δασκάλου. Η κύρια ιδέα είναι να μιμηθείτε απευθείας την τελική πρόβλεψη του μοντέλου δασκάλου. Η απόσταξη της γνώσης που βασίζεται στην απάντηση είναι απλή αλλά αποτελεσματική για τη συμπίεση μοντέλων και έχει χρησιμοποιηθεί ευρέως σε διάφορες εργασίες και εφαρμογές.
- Τα βαθιά νευρωνικά δίκτυα είναι καλά στην εκμάθηση πολλαπλών επιπέδων αναπαράστασης χαρακτηριστικών με την μέθοδο αυξητικής αφαίρεσης. Αυτό είναι γνωστό ως representation learning. Επομένως, τόσο η έξοδος του τελευταίου στρώματος και η έξοδος των ενδιάμεσων στρωμάτων, δηλαδή χάρτες χαρακτηριστικών, μπορούν να χρησιμοποιηθούν ως γνώση για την επίβλεψη της εκπαίδευσης του μοντέλου μαθητή. Συγκεκριμένα, η feature-based knowledge από τα ενδιάμεσα επίπεδα είναι καλή επέκταση της γνώσης που βασίζεται στην απάντηση, ειδικά για την εκπαίδευση λεπτότερων και βαθύτερων δικτύων.
- Οι 2 παραπάνω κατηγορίες γνώσης, χρησιμοποιούν ένα συγκεκριμένο επίπεδο στο μοντέλο του δασκάλου. Η relation-based knowledge διερευνά περαιτέρω τις σχέσεις μεταξύ διαφορετικών επιπέδων ή δειγμάτων δεδομένων.

Όπως συμβαίνει και με τις παραπάνω κατηγορίες γνώσης, υπάρχουν και διαφορετικοί τύποι απόσταξης γνώσης. Σύμφωνα με τους ερευνητες, ανάλογα με το αν το μοντέλο «δάσκαλος» εκπαιδεύεται ταυτόχρονα με το μοντέλο «μαθητή» ή όχι, τα σχήματα της απόσταξης γνώσης χωρίζονται σε 3 κατηγορίες : offline distillation, online distillation and self-distillation. Πιο αναλυτικά :

- Στην **offline distillation** η γνώση μεταφέρεται από ένα προεκπαιδευμένο μοντέλο δάσκαλο σε ένα δίκτυο μαθητή. Έτσι, η όλη διαδικασία της μάθησης έχει δύο στάδια: 1) το μεγάλο μοντέλο-δάσκαλος εκπαιδεύεται πρώτα σε ένα σύνολο από δεδομένα εκπαίδευσης πριν την απόσταξη και 2) το μοντέλο του δασκάλου χρησιμοποιείται είτε για την εξαγωγή της γνώσης με τη μορφή logits είτε για τα ενδιάμεσα χαρακτηριστικά, τα οποία στη συνέχεια χρησιμοποιούνται για καθοδήγηση στην εκπαίδευση του μοντέλου μαθητή κατά την απόσταξη.
- Αν και οι μέθοδοι απόσταξης εκτός σύνδεσης είναι απλές και αποτελεσματικές, ορισμένα ζητήματα στην απόσταξη εκτός σύνδεσης έχουν προσελκύσει την προσοχή της ερευνητικής κοινότητας. Για να ξεπεραστεί ο περιορισμός της απόσταξης εκτός σύνδεσης, προτάθηκε η **online distillation** για περαιτέρω βελτίωση της απόδοσης του μοντέλου μαθητή, ειδικά όταν δεν είναι διαθέσιμο ένα μεγάλο μοντέλο-δάσκαλος υψηλής απόδοσης μεγάλης χωρητικότητας. Στην online απόσταξη, και τα 2 μοντέλα δασκάλου και μαθητή ενημερώνονται ταυτόχρονα και όλο το πλαίσιο της απόσταξης γνώσης είναι εκπαιδεύσιμο από άκρο σε άκρο (end-to-end).
- Τέλος, υπάρχει και η τεχνική του **self-distillation** όπου ίδιο δίκτυο χρησιμοποιείται και σαν δάσκαλος και σαν μαθητής. Αυτό μπορεί να θεωρηθεί ένα ειδικό είδος της online απόσταξης. Πιο συγκεκριμένα, προτάθηκε αυτή η μέθοδος της αυτό-απόσταξης, στην οποία η γνώση από τα βαθύτερα τμηματα του δικτύου, μεταφέρονται στα πιο «ρηχά» τμήματα.



Εικόνα 2.8.2 : Τα διαφορετικά είδη της απόσταξης γνώσης. Με κόκκινο χρώμα είναι τα προεκπαιδευμένα μοντέλα που έχουν εκπαιδευτεί πριν την απόσταξη, και με κίτρινο χρώμα είναι τα μοντέλα που εκπαιδεύονται κατά τη διάρκεια της απόσταξης. (Πηγή: [11] https://arxiv.org/pdf/2006.05525.pdf)

2.9 Σετ δεδομένων (Datasets) στην Αυτό-επιβλεπόμενη Μάθηση

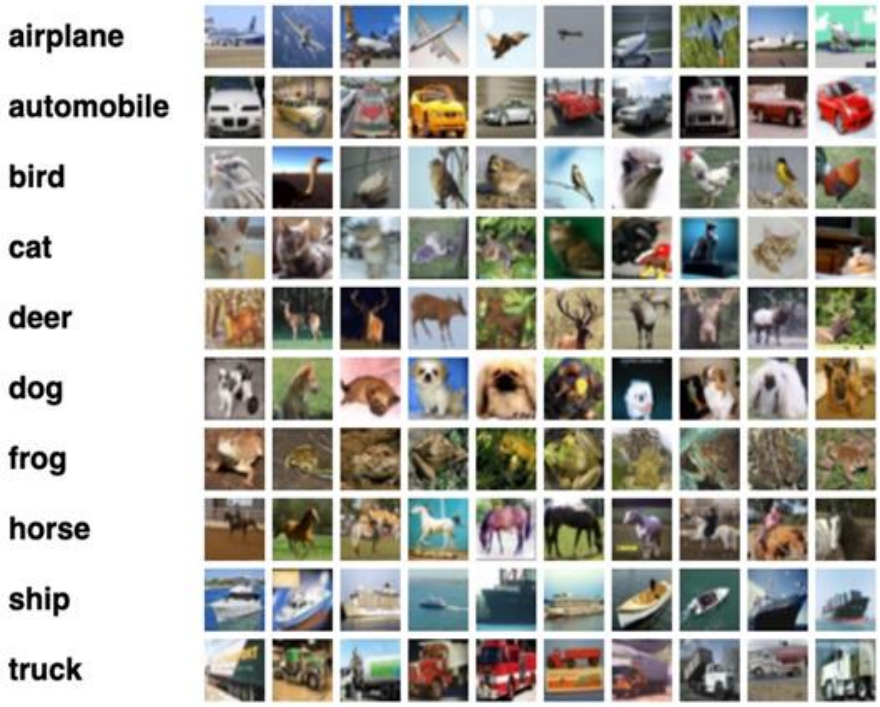
Τα νευρωνικά δίκτυα είναι απαραίτητο να εκπαιδευτούν -βελτιστοποιηθούν με ένα μεγάλο αριθμό από δεδομένα. Όπως προαναφέρθηκε η διεργασία της προσθήκης ετικέτας στα δεδομένα απαιτεί πολλές ώρες απασχόλησης για έναν άνθρωπο και έτσι το θέμα αυτό απασχολεί αρκετά τη διεθνή επιστημονική κοινότητα προκειμένου να αυτοματοποιηθεί η διαδικασία. Οι εικόνες εκπαίδευσης μπορούν να αντληθούν από συλλογές δεδομένων εικόνων κατάλληλες για τη διαδικασία της αναγνώρισης-ανίχνευσης. Η χρήση των παραπάνω συλλογών επιτρέπουν στο χρήστη να περιορίσει το χρόνο συλλογής, επεξεργασίας και μορφοποίησης των δεδομένων και να ασχοληθεί αποκλειστικά με την ανάπτυξη των μεθόδων και των τεχνικών. Παρακάτω περιγράφονται κάποιες από τις βασικές συλλογές δεδομένων που χρησιμοποιούνται στην όραση υπολογιστών και κάποιες συλλογές που θα χρησιμοποιηθούν στην παρούσα διπλωματική εργασία.

- **CIFAR-10**

Το σύνολο δεδομένων CIFAR-10 περιέχει 60.000 έγχρωμες εικόνες, διαστάσεων 32x32 σε 10 διαφορετικές κατηγορίες-κλάσεις. Οι 10 διαφορετικές κατηγορίες αντιπροσωπεύουν: αεροπλάνα, αυτοκίνητα, πουλιά, γάτες, ελάφια, σκύλους, βατράχους, άλογα, πλοία και φορτηγά. Υπάρχουν 6.000 εικόνες σε κάθε τάξη. Λόγω της χαμηλής του ανάλυσης, το dataset Cifar-10 χρησιμοποιείται συχνά σε εργασίες αναγνώρισης αντικειμένων.

- **CIFAR-100**

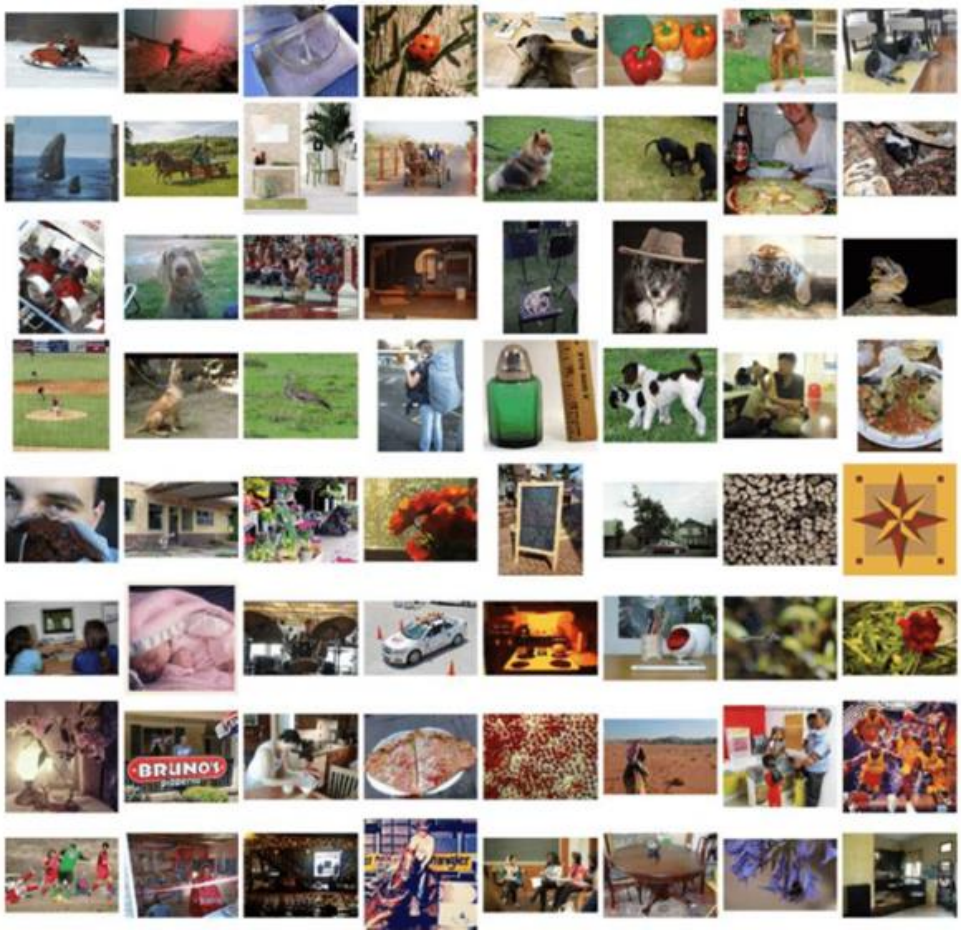
Το σύνολο δεδομένων CIFAR-10 περιέχει 60.000 έγχρωμες εικόνες, διαστάσεων 32x32 σε 10 διαφορετικές κατηγορίες-κλάσεις. Οι 10 διαφορετικές κατηγορίες αντιπροσωπεύουν: αεροπλάνα, αυτοκίνητα, πουλιά, γάτες, ελάφια, σκύλους, βατράχους, άλογα, πλοία και φορτηγά. Υπάρχουν 6.000 εικόνες σε κάθε τάξη. Λόγω της χαμηλής του ανάλυσης, το dataset Cifar-10 χρησιμοποιείται συχνά σε εργασίες αναγνώρισης αντικειμένων.



Εικόνα 2.9.1 : 10 τυχαίες εικόνες από κάθε κλάση του σετ δεδομένων CIFAR-10
(Πηγή : [36] https://www.cs.toronto.edu/~kriz/cifar.html)

- **ImageNet Database**

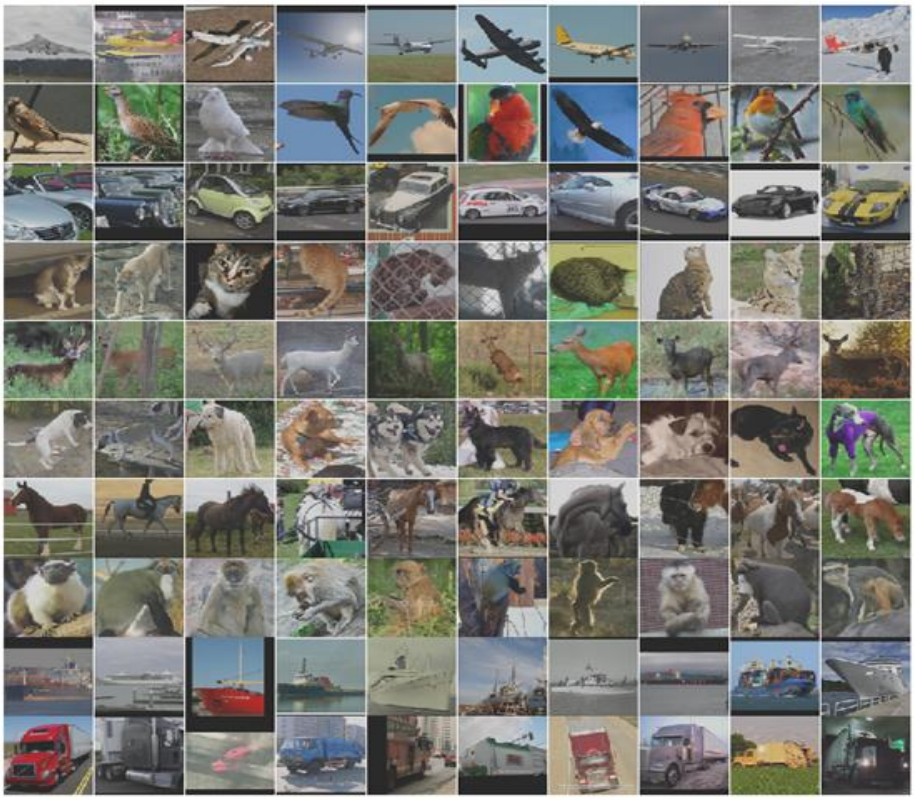
Η συλλογή αυτή είναι από τις πιο γνωστές και πιο χρησιμοποιημένες. Η βάση περιέχει πάνω από 14 εκατομμύρια εικόνες με τις αντίστοιχες περιγραφές τους (labels) και για 1 εκατομμύριο από αυτές είναι επιπλέον διαθέσιμα περιγεγραμμένα τετράγωνα (bounding boxes) των αντικειμένων τους. Οι συνολικές κατηγορίες των αντικειμένων είναι περίπου 20.000 με εκατοντάδες εικόνες για κάθε μία. Από το 2010 λαμβάνει χώρα αντίστοιχος ετήσιος διαγωνισμός γνωστός ως ImageNet Large Scale Visual Recognition Challenge (ILSVRC) όπου ερευνητές ανταγωνίζονται σε μεθόδους αναγνώρισης και κλασικοποίησης αντικειμένων από εικόνες. Ο συγκεκριμένος διαγωνισμός αποτέλεσε την αιτία της έξαρσης των νέων επιτυχημένων μεθόδων μέσα σε μικρό χρονικό διάστημα (2012 -2017).[37]



Εικόνα: Παραδείγματα εικόνων από την βάση δεδομένων ImageNet
(Πηγή : [38] https://www.researchgate.net/figure/Examples-in-the-ImageNet-dataset_fig7_314646236)

- **STL-10**

Το σύνολο δεδομένων STL-10 είναι ένα σύνολο δεδομένων αναγνώρισης εικόνας για την ανάπτυξη αλγορίθμων μάθησης χωρίς επίβλεψη (unsupervised feature learning), βαθιάς μάθησης και αλγορίθμους αυτοδιδασκαλίας. Είναι εμπνευσμένο από το σύνολο δεδομένων CIFAR-10 που εξηγήθηκε παραπάνω αλλά με ορισμένες τροποποιήσεις. Συγκεκριμένα, κάθε τάξη έχει λιγότερα παραδείγματα εκπαίδευσης με ετικέτα από ό,τι στο CIFAR-10, αλλά παρέχεται ένα πολύ μεγάλο σύνολο δειγμάτων χωρίς ετικέτα για την εκμάθηση των μοντέλων των εικόνων πριν από την επιβλεπόμενη εκπαίδευση. Η κύρια πρόκληση είναι να χρησιμοποιηθούν τα δεδομένα χωρίς ετικέτα (τα οποία προέρχονται από παρόμοια αλλά διαφορετική κατανομή από τα δεδομένα με ετικέτα) για να δημιουργηθεί ένα χρήσιμο υπόβαθρο. Το σύνολο αυτό περιέχει 10 κλάσεις : αεροπλάνο, πουλί, αυτοκίνητο, γάτα, σκύλος, ελάφι, άλογο, μαϊμού, πλοίο και φορτηγό. Οι εικόνες είναι έγχρωμες και διαστάσεων (96×96) pixels. Οι εικόνες χωρίζονται σε 500 εικόνες για εκπαίδευση, 800 εικόνες για τεστάρισμα και 100000 μη επισημασμένες εικόνες για μη-επιβλεπόμενη μάθηση (αυτά τα παραδείγματα εξάγονται από παρόμοια αλλά ευρύτερη κατανομή εικόνων. Για παράδειγμα, περιέχει άλλους τύπους ζώων (αρκούδες, κουνέλια κ.λπ.) και οχήματα (τρένα, λεωφορεία κ.λπ.) εκτός από αυτά του σετ με ετικέτα.) Οι εικόνες αυτές πάρθηκαν από παραδείγματα με ετικέτα του ImageNet. [39]



Εικόνα 2.9.2 : Δείγματα των εικόνων του dataset STL-10
(Πηγή: [39] https://ai.stanford.edu/~acoates/stl10/)

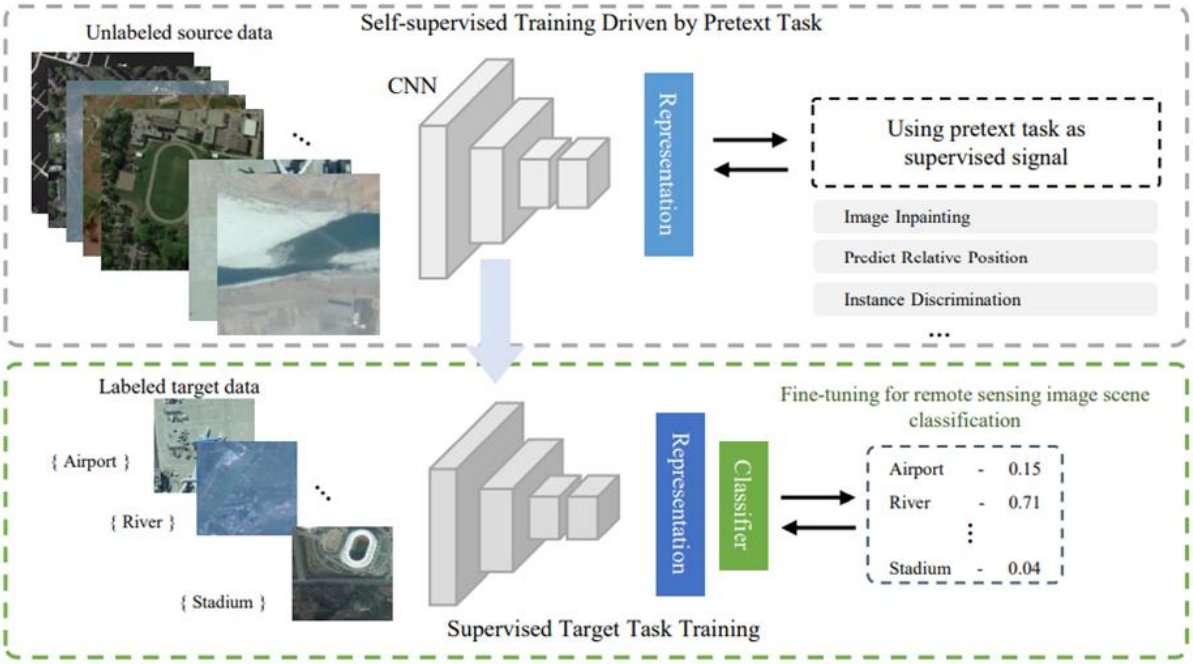
2.10 Self-supervised Learning και Τηλεπισκόπηση

Στο κεφάλαιο αυτό θα αναφερθούν κάποιες από τις πιο σύγχρονες μελέτες και έρευνες που έχουν πραγματοποιηθεί στον τομέα της τηλεπισκόπησης σε συνδυασμό με την αυτό-επιβλεπόμενη μάθηση.

- **Remote Sensing Image Scene Classification with Self-Supervised Paradigm under Limited Labeled Samples (2020)**

Ο Chao Tao που ανήκει στην Σχολή Γεωεπιστημών και Πληροφορικής του πανεπιστημίου Central South University της Κίνας, δημοσίευσε την έρευνα αυτή μαζί με τους Ji Qi, Weipeng Lu, Hao Wang and Haifeng Li. Αυτή η δημοσίευση ήταν και η πρώτη φορά που χρησιμοποιήθηκε ο μηχανισμός του self-supervised learning στην ταξινόμηση των εικόνων τηλεπισκόπηση. Με την συγκεκριμένη έρευνα αποδεικνύεται ότι η αυτο-επιβλεπόμενη μάθηση είναι ένα εντελώς νέο παράδειγμα που μαθαίνει χαρακτηριστικά από ογκώδεις εικόνες χωρίς ετικέτα για την κατανόηση εικόνων τηλεπισκόπηση. Αυτό το παράδειγμα είναι εξαιρετικά κατάλληλο για tasks κατανόησης τηλεπισκοπικών εικόνων γιατί υπάρχει πολύ εύκολα πρόσβαση σε μεγάλο αριθμό εικόνων από διαφορετικές περιοχές και σε διαφορετικές εποχές. Επίσης πειράματα που έγιναν σε 3 σύνολα δεδομένων ταξινόμησης σκηνής τηλεπισκόπησης έδειξαν πως η προτεινόμενη μέθοδος υπερβαίνει την παραδοσιακή προσέγγιση προεκπαίδευσης του συνόλου δεδομένων ImageNet όταν τα δεδομένα με ετικέτα είναι ανεπαρκή.

Όπως ακριβώς παρουσιάζεται στην παρακάτω εικόνα, ο γενικός «αγωγός» του παραδείγματος αυτό-επιβλεπόμενη μάθησης αποτελείται από 2 μέρη. Στη φάση της αυτό-επιβλεπόμενη εκπαίδευσης, ένα βαθύ συνελικτικό νευρωνικό δίκτυο (DCNN) εκπαιδεύεται στην επίλυση προκαθορισμένης εργασίας-pretext task, για εκμάθηση πιθανών χρήσιμων αναπαραστάσεων σε πηγές μεγάλων σετ δεδομένων χωρίς ετικέτα. Οι αναπαραστάσεις που μαθαίνει αποθηκεύονται ως παράμετροι του κωδικοποιητή του DCNN. Σε δεύτερη φάση λοιπόν, οι πλέον γνωστές αναπαραστάσεις μεταφέρονται σε άλλες εργασίες-στόχους ως προεκπαιδευμένα μοντέλα. Σε σύγκριση με την εκπαίδευση από την αρχή, η «τελειοποίηση» ή αλλιώς fine-tuning, του προεκπαιδευμένου μοντέλου με κάλες αναπαραστάσεις μπορεί να ξεπεράσει το overfitting και να πετύχει υψηλότερη απόδοση στην εργασία στόχο, ειδικά όταν τα δείγματα με ετικέτα είναι ανεπαρκή. [12]

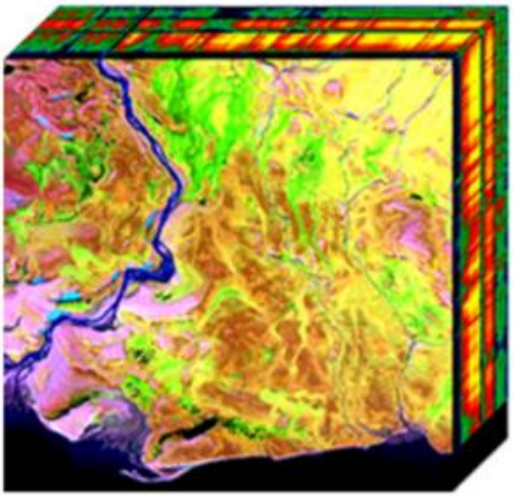


Εικόνα 2.10.1.1 : Διάγραμμα ροής του παραδείγματος της αυτό-επιβλεπόμενης μάθησης για ταξινόμηση σκηνών εικόνων τηλεπισκόπηση. (Πηγή: [12] 2010.00882.pdf (arxiv.org))

- **Self-Supervised Learning with Adaptive Distillation for Hyperspectral Image Classification (2021)**

Η δημοσίευση αυτή πραγματοποιήθηκε από τους Jun Yue, Leyuan Fang και Pedram Ghamisi. Η συγκεκριμένη έρευνα χρησιμοποιεί υπερφασματικά δεδομένα, δηλαδή εικόνες που συλλέγουν πληροφορία από όλο το φάσμα. Πιο συγκεκριμένα, οι υπερφασματικές εικόνες αποτελούνται από διαδοχικές εικόνες της ίδιας περιοχής αναφοράς τοποθετημένες η μία πάνω στην άλλη ώστε να προκύψει ένα τρισδιάστατο διάνυσμα του οποίου η τρίτη διάσταση αναφέρεται στο ηλεκτρομαγνητικό φάσμα. Το διάνυσμα αυτό ή αλλιώς ο λεγόμενος φασματικός κύβος, περιέχει το σύνολο των τιμών της κάθε φασματικής κατηγορίας που έχει συλλέξει ο δέκτης, για κάθε εικονοστοιχείο της φυσικής επιφάνειας στην οποία αναφέρεται η εικόνα. Η υπερφασματική απεικόνιση είναι ουσιαστικά η καταγραφή μιας «σκηνής» σε πολλαπλά φάσματα στο εύρος του

ορατού και του υπερύθρου. Σε αντίθεση με την πολυφασματική απεικόνιση, όπου η καταγραφή γίνεται σε λιγότερες φασματικές ζώνες (περίπου 10), στην υπερφασματική απεικόνιση η καταγραφή είναι συνεχής και καλύπτει εκατοντάδες ζώνες, κάθε μια σε ένα πολύ στενό εύρος του μήκους κύματος.



Εικόνα 2.10.2.1 : Υπερφασματικός κύβος
(Πηγή : [40] https://en.wikipedia.org/wiki/File:HyperspectralCube.jpg)

Η ταξινόμηση των υπερφασματικών εικόνων έχει απασχολήσει αρκετά την επιστημονική κοινότητα και υπάρχει έλλειψη στα δεδομένα αυτά, και έτσι οι μέθοδοι βαθιάς μάθησης δεν έχουν εξελιχθεί ιδιαίτερα στον τομέα αυτό. Με την δημοσίευση αυτή, η μέθοδος της αυτό-επιβλεπόμενης μάθησης (self-supervised learning) με προσαρμοστική «απόσταξη» (adaptive distillation) προτάθηκε για να εκπαιδεύσει νευρωνικά δίκτυα με εκτεταμένα δείγματα χωρίς ετικέτες (unlabeled data). Η προτεινόμενη μέθοδος αποτελείται από 2 μοντέλα: την προσαρμοστική απόσταξη γνώσης με χωρικο-φασματική ομοιότητα και τον τρισδιάστατο μετασχηματισμό του υπερφασματικού κύβου. [13]

Τα βασικά συμπεράσματα που προκύπτουν από την εργασία αυτή είναι :

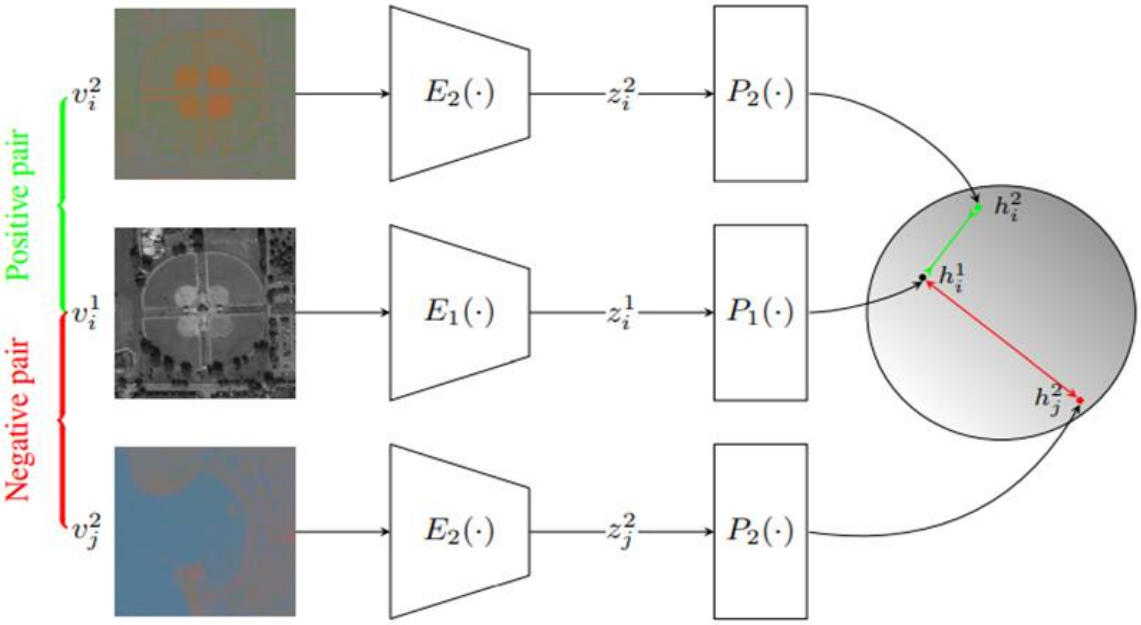
- Η προτεινόμενη μέθοδος υπερέχει της υπάρχουσας δημοφιλής μεθόδου ταξινόμησης υπερφασματικών σκηνών, παρουσία μικρών επισημασμένων δειγμάτων εκπαίδευσης. Η συγκριτική ανάλυση των αποτελεσμάτων υποδηλώνει ότι η προτεινόμενη μέθοδος επιτυγχάνει τα πιο σύγχρονα αποτελέσματα για την ταξινόμηση υπερφασματικών εικόνων με λίγες λήψεις.
- Η μελέτη κατάλυσης που πραγματοποιείται δείχνει ότι και οι δύο ενότητες (SSL με προσαρμοστική απόσταξη γνώσης και SSL με 3-D μετασχηματισμό) είναι αποτελεσματικές για τη βελτίωση της ακρίβειας.
- Η στρατηγική σύντηξης (fusion strategies) βελτιώνει την απόδοση της ταξινόμησης σε επίπεδο pixel με περιορισμένα επισημασμένα δεδομένα εκπαίδευσης.

- **Self-Supervised Learning of Remote Sensing Scene Representations Using Contrastive Multiview Coding (2021)**

Σε αυτήν την εργασία, που πραγματοποιήθηκε από επιστήμονες του Πανεπιστημίου της Banja Luka, Bosnia Herzegovina, διεξάγεται μια εκτενής ανάλυση της δυνατότητας εφαρμογής των αυτοεποτευόμενων μεθόδων μάθησης στον τομέα της τηλεπισκόπησης. Αναλύεται η επιρροή του αριθμού και του τομέα των εικόνων που χρησιμοποιούνται για προεκπαίδευση αυτό-επιβλεπόμενη μάθησης στην απόδοση των downstream εργασιών. Αποδεικνύεται ότι για το downstream task της ταξινόμησης εικόνων τηλεπισκόπηση, χρησιμοποιώντας αυτό-επιβλεπόμενη προεκπαίδευση σε εικόνες τηλεπισκόπησης, μπορούν να δοθούν καλύτερα αποτελέσματα από το να προεκπαιδεύονταν εικόνες φυσικών σκηνών. Επιπλέον, αποδεικνύεται επίσης ότι η αυτο-επιβλεπόμενη προ-εκπαίδευση μπορεί εύκολα να επεκταθεί σε πολυφασματικές απεικονίσεις που παράγουν ακόμα καλύτερα αποτελέσματα στις downstream εργασίες.

Η μέθοδος που χρησιμοποιείται σε αυτήν την δουλειά αποτελείται από 2 βήματα: αυτό-επιβλεπόμενη προεκπαίδευση (self-supervised pre-training) και μεταφορά της γνώσης και αξιολόγηση των αναπαραστάσεων που μαθαίνονται.

Η εργασία χρησιμοποιεί τη μέθοδο του contrastive learning και έτσι το pretext task ορίζεται με τρόπο που οι παρόμοιες εικόνες θα πρέπει να βρίσκονται κοντά στον χώρο της αναπαράστασης, ενώ οι ανόμοιες εικόνες θα πρέπει να είναι όσο το δυνατόν πιο μακριά. Αναζητούνται «θετικά» και «αρνητικά» ζευγάρια. Έτσι για την αυτό-επιβλεπόμενη μάθηση χρησιμοποιείται το Contrastive Multiview Coding – CMC.

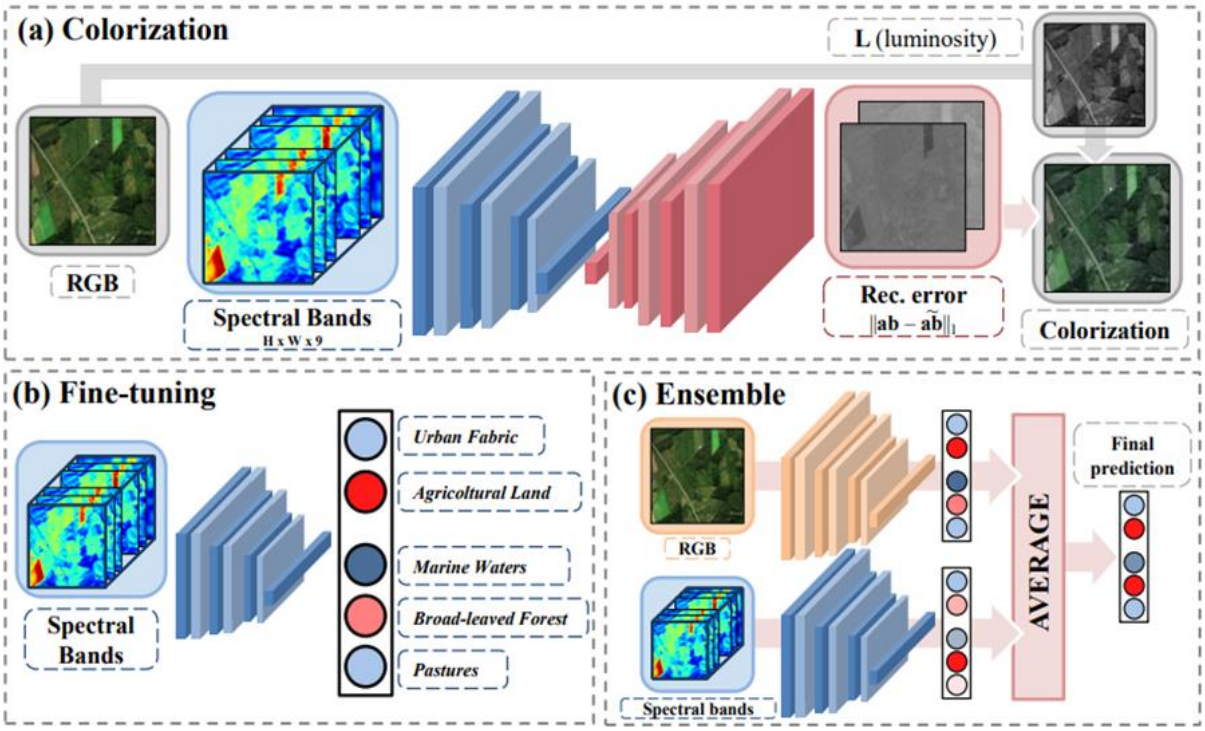


Εικόνα 2.10.3.1 : Σχηματική απεικόνιση του αλγορίθμου CMC, όπου τα θετικά ζεύγη εικόνων έχουν μικρότερη απόσταση από τα αρνητικά ζεύγη. (Πηγή: [14] 2104.07070.pdf (arxiv.org))

- **The color out of space: learning self-supervised representations for Earth Observation imagery (2020)**

Πολλοί επιστήμονες από 3 Πανεπιστήμια και Ιστινστιτούτα της Ιταλίας, δημοσίευσαν αυτήν την την εργασία αναφορικά με την αυτό-επιβλεπόμενη μάθηση και τις δορυφορικές απεικονίσεις. Πιο συγκεκριμένα, προτείνεται να παρθούν σημαντικές και ουσιαστικές αναπαραστάσεις από τις δορυφορικές εικόνες, αξιοποιώντας τις φασματικές ζώνες με πολλές διαστάσεις για να ανασκευαστεί το ορατό χρώμα. Πραγματοποιούνται πειράματα για την ταξινόμηση της κάλυψης Γης (BigEarthNet) και πειράματα ανίχνευσης του ιού του Δυτικού Νείλου, που δείχνει ότι ο χρωματισμός είναι ένα σταθερό pretext task για την εκπαίδευση ενός εξαγωγέα χαρακτηριστικών.

Πιο αναλυτικά, προτείνετε μια νέα διαδικασία εκμάθησης αναπαραστάσεων για δορυφορικές εικόνες, η οποία επινοεί έναν αυτοεπιβλεπόμενο αλγόριθμο. Πιο συγκεκριμένα, απαιτεί από το δίκτυο να ανακτήσει πληροφορίες RGB μόνο μέσω άλλων φασματικών ζωνών (δηλαδή ένα υποσύνολο ζωνών χωρίς την RBG πληροφορία). Το task του χρωματισμού θεωρείται ότι είναι ιδιαίτερα χρήσιμο για τις δορυφορικές εικόνες, καθώς η σύνδεση μεταξύ των χρωμάτων και η σημασιολογία είναι ισχυρή: για παράδειγμα, τα θαλάσσια ύδατα χαρακτηρίζονται με το μπλε χρώμα, περιοχές βλάστησης με το πράσινο ή στις καλλιεργήσιμες εκτάσεις προτιμούνται ζεστοί τόνοι. Μια τέτοια προηγούμενη γνώση εισάγεται μέσω μιας αρχιτεκτονικής κωδικοποιητή-αποκωδικοποιητή που εκμεταλλεύεται τις φασματικές ζώνες. Μόλις το μοντέλο φτάσει σε καλές δυνατότητες στον χρωματισμό πλακιδίων, χρησιμοποιείται ο κωδικοποιητής σαν εξαγωγέας χαρακτηριστικών για το επόμενο βήμα, δηλαδή μια εργασία τηλεπισκόπησης. Οι αναπαραστάσεις που εκπαιδεύτηκαν στον χρωματισμό οδήγησαν σε αξιοσημείωτα αποτελέσματα και αποκλίνουν σημασιολογικά από αυτά που υπολογίστηκαν πάνω σε κανάλια RGB. [15]



Εικόνα 2.10.4.1 : Μια επισκόπηση του γενικού «αγωγού» για εκμάθηση χαρακτηριστικών σε δορυφορικές εικόνες
(Πηγή: [15] 2006.12119.pdf (arxiv.org))

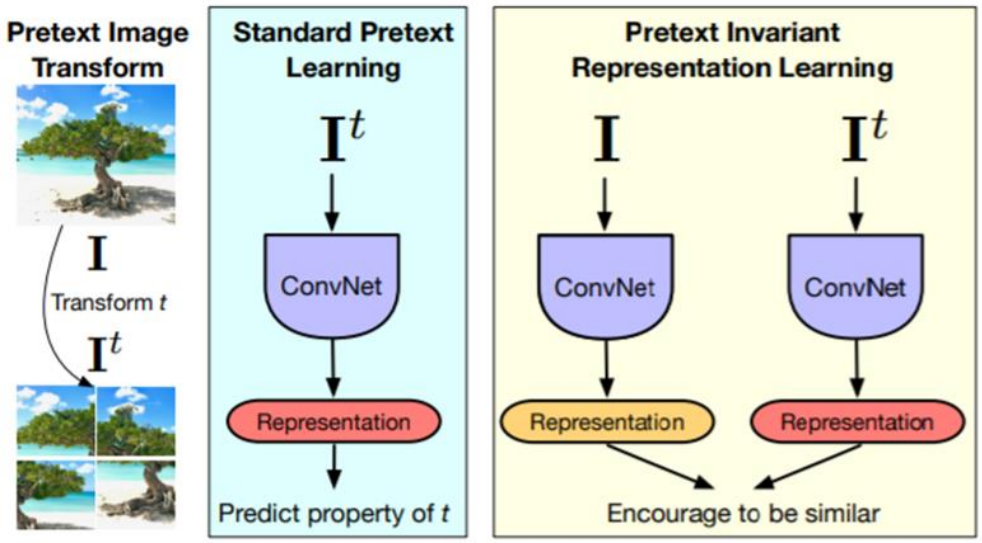
2.11 Self-Supervised Learning of Pretext-Invariant Representations

Το 2019 η ερευνητική ομάδα της Facebook, έδωσε στην δημοσιότητα μια εργασία με τίτλο «Self-Supervised Learning of Pretext-Invariant Representations». Η παρούσα διπλωματική εργασία βασίζεται στην παραπάνω ιδέα που θα αναλυθεί στην συνέχεια.

Ξεκινώντας, όπως έχει αναλυθεί προηγουμένως, ο στόχος της αυτό-επιβλεπόμενη μάθησης από εικόνες, είναι η κατασκευή αναπαραστάσεων μέσω pretext tasks που έχουν σημασιολογικό νόημα και δεν απαιτούν ετικέτες-σχολιασμούς για ένα μεγάλο σύνολο δεδομένων εκπαίδευσης. Πολλά pretext tasks οδηγούν σε αναπαραστάσεις που είναι διαφορετικές σε σχέση με τον μετασχηματισμό που έχει υποστεί η αρχική εικόνα. Αντίθετα, οι ερευνητές υποστηρίζουν πως η σημασιολογική αναπαράσταση μιας εικόνας πρέπει να είναι αμετάβλητη ακόμα και αν υποστεί μετασχηματισμό η αρχική εικόνα. Πιο συγκεκριμένα, αναπτύξαν το Pretext-Invariant Representation Learning ή αλλιώς PIRL το οποίο μαθαίνει αμετάβλητες αναπαραστάσεις εικόνων που βασίζονται στα pretext tasks. Χρησιμοποιούν το PIRL με ένα πολυχρησιμοποιημένο pretext task, την επίλυση του jigsaw puzzle. Οι ερευνητές αποδεικνύουν πως το PIRL ουσιαστικά βελτιώνει την ποιότητα των αναπαραστάσεων που μαθαίνονται. Ακόμη, το PIRL, παρά το γεγονός ότι γίνεται χωρίς επίβλεψη, ξεπερνά τις αποδόσεις της επιβλεπόμενης προεκπαίδευσης στην εκμάθηση αναπαραστάσεων εικόνων για εργασίες ανίχνευσης αντικειμένων.

Γενικότερα η αυτό-επιβλεπόμενη μάθηση προσπαθεί να μάθει αναπαραστάσεις από τα ίδια τα pixels χωρίς όμως να βασίζεται σε προκαθορισμένες επισημάνσεις. Αυτό, συχνά επιτυγχάνεται μέσω των pretext tasks, τα οποία εφαρμόζουν μετασχηματισμούς στην εικόνα εισόδου και απαιτούν από τον εκπαιδευτή να προβλέψει τις ιδιότητες-χαρακτηριστικά του μετασχηματισμού για την μετασχηματισμενη εικονα.

Με βάση όλα τα παραπάνω, προτείνετε μια μέθοδος που μαθαίνει αμετάβλητες αναπαραστάσεις των αρχικών εικόνων αντί για μεταβλητές. Το Pretext-Invariant Representation Learning (PIRL) αντί να προβλέψει τις ιδιότητες του μετασχηματισμού, κατασκευάζει αναπαραστάσεις εικόνων που είναι παρόμοιες με την αναπαράσταση μετασχηματισμένων εκδόσεων της ίδιας εικόνας και διαφορετική από τις αναπαραστάσεις άλλων εικόνων. Ουσιαστικά ξεπερνά κάθε προηγούμενη τεχνική στην αυτο-επιβλεπόμενη μάθηση από το ImageNet και από δεδομένα εικόνων που δεν είναι προσεκτικά διαλεγμένα. Είναι ενδιαφέρον ότι το PIRL ξεπερνά ακόμη και την επιβλεπόμενη προεκπαίδευση στην εκμάθηση αναπαραστάσεων εικόνων κατάλληλων για ανίχνευση αντικειμένων.



Εικόνα 2.11.1.1 : Pretext-Invariant Representation Learning (PIRL). Πολλά pretext tasks για την αυτό-επιβλεπόμενη μάθηση, συμπεριλαμβάνουν τον μετασχηματισμό της εικόνας I, υπολογίζοντας την αναπαράσταση της τροποποιημένης εικόνας και προβλέποντας τις ιδιότητες του μετασχηματισμού t από την αναπαράσταση. Σαν αποτέλεσμα, η αναπαράσταση πρέπει να διαφέρει απο τον μετασχηματισμό t και πιθανόν δεν πρέπει να περιέχει πολλές σημαντικές πληροφορίες. Αντίθετα, το PIRL μαθαίνει αναπαραστάσεις που είναι αμετάβλητες στον μετασχηματισμό t και διατηρούν σημαντικές πληροφορίες.
(Πηγή : [16] https://arxiv.org/pdf/1912.01991.pdf)

2.11.1 Επισκόπηση της μεθόδου

Η εργασία επικεντρώνεται σε pretext tasks για αυτό-επιβλεπόμενη μάθηση, στην οποία ένας γνωστός μετασχηματισμός εικόνας εφαρμόζεται στην εικόνα εισόδου. Για παράδειγμα, το «Jigsaw puzzle» χωρίζει την εικόνα σε εννέα κομμάτια και διαταράσσει την εικόνα μεταθέτοντας τυχαία τα κομμάτια αυτά. Προηγούμενες εργασίες που έχουν πραγματοποιηθεί, χρησιμοποιούσαν το Jigsaw puzzle ως task για πρόβλεψη της μετάθεσης από την διαταραγμένη εικόνα εισόδου. Αυτό απαιτεί από τον εκπαιδευτή να κατασκευάσει μια αναπαράσταση που είναι αμετάβλητη της αρχικής διαταραχής. Ακριβώς το ίδιο συμβαίνει και με άλλα pretext tasks που έχουν μελετηθεί τελευταία. Στην εργασία αυτή υιοθετείται το υπάρχον pretext task Jigsaw puzzle με τρόπο που να προτρέπει την αναπαράσταση της εικόνας να είναι αμετάβλητη στην διαταραχή των κομματιών της εικόνας. Σύμφωνα με τους ερευνητές η προσέγγιση αυτή είναι εύκολα εφαρμόσιμη και σε κάθε άλλο pretext task που περιλαμβάνει μετασχηματισμούς εικόνων, ωστόσο η διπλωματική εργασία αυτή εστιάζει στο Jigsaw puzzle.

2.11.2 Πειράματα της εργασίας PIRL

Ακολουθώντας την κοινή πρακτική στην αυτο-επιβλεπόμενη μάθηση, αξιολογήθηκε η απόδοση του PIRL σε πειράματα μεταφοράς μάθησης (transfer-learning). Πραγματοποιήθηκαν πειράματα

σε ποικίλα σύνολα δεδομένων, με επίκεντρο τον εντοπισμό αντικειμένων και την εργασία ταξινόμησης εικόνας. Οι εμπειρικές αξιολογήσεις των ερευνητών καλύπτουν :

1. Μια ρύθμιση της εκμάθησης στην οποία οι παράμετροι του συνελικτικού δικτύου ρυθμίζονται με ακρίβεια κατά την μεταφορά, αξιολογώντας έτσι την «αρχικοποίηση» του δικτύου που πραγματοποιήθηκε με την χρήση της αυτό-επιβλεπόμενης μάθησης

2. Ένα περιβάλλον μάθησης στο οποίο οι παράμετροι του δικτύου σταθεροποιούνται κατά την διαδικασία της μεταφορά της εκμάθησης, χρησιμοποιώντας έτσι το δίκτυο ως εξαγωγέα χαρακτηριστικών.

Προκειμένου να διευκολυνθεί η σύγκριση με προηγούμενες έρευνες που έχουν πραγματοποιηθεί, οι ερευνητές της Facebook χρησιμοποίησαν το σετ δεδομένων του ImageNet με 1,28 εκατομμύρια εικόνες για την προ-εκπαίδευση του μοντέλου. Ακόμη πρέπει να αναφερθεί πως χρησιμοποιούν κυρίως το μοντέλο ResNet.

Χρησιμοποιείται μοντέλο ResNet50, για τον υπολογισμό των αναπαραστάσεων των εικόνων (και των αρχικών και αυτών που έχουν υποστεί μετασχηματισμό). Τα μοντέλα εκπαιδεύονται χρησιμοποιώντας mini-batch SGD χρησιμοποιώντας το σχήμα διάσπασης ρυθμού μάθησης συνημιτόνου με αρχικό ρυθμό μάθησης (learning rate) $1{,}2\times10^{-1}$ και ένα τελικό ποσοστό μάθησης του $1{,}2 \times 10^{-4}$. Τα μοντέλα εκπαιδεύονται για 800 εποχές (epochs), χρησιμοποιώντας batch size 1,024 εικόνες ανα παρτίδα. Επίσης, δεν χρησιμοποιούνται προσεγγίσεις επαύξησης των δεδομένων επειδή αυτό είναι αποτέλεσμα από προσεγγίσεις επιβλεπόμενης μάθησης. Τέλος, οι υπερπαράμετροι που χρησιμοποιούνται στην μεταφορά μάθησης, παίζουν σημαντικό ρόλο σύμφωνα με προηγούμενα πειράματα που έχουν πραγματοποιηθεί.

Συνοπτικά τα βήματα που ακολουθούνται είναι : αρχικά εξάγονται 9 κομμάτια της εικόνας, υπολογίζεται μια αναπαράσταση για κάθε κομμάτι ξεχωριστά εξάγοντας τις ενεργοποιήσεις από το επίπεδο res5 του ResNet-50 και υπολογίζοντας τη μέση τιμή της ενεργοποιήσεις. Έπειτα εφαρμόζεται μια γραμμική προβολή για να ληφθεί ένα διάνυσμα-κομμάτι 128-διαστάσεων και στην συνέχεια συνενώνονται οι αναπαραστάσεις των κομματιών σε τυχαία σειρά και εφαρμόζεται μια δεύτερη γραμμική προβολή στο αποτέλεσμα για να ληφθεί το τελικό διάνυσμα (128-διαστάσεων) αναπαράσταση εικόνας.

Η παρούσα διπλωματική εργασία ενδιαφέρεται περισσότερο για την εργασία της ταξινόμησης εικόνων παρά για την ανίχνευση αντικειμένων. Έτσι παρακάτω παρουσιάζεται ένας πίνακας με τα αποτελέσματα της εργασίας σε διαφορετικά σετ δεδομένων. Πραγματοποιείται ταξινόμηση εικόνων με γραμμικά μοντέλα σε 4 σετ δεδομένων χρησιμοποιώντας κάποιες ρυθμίσεις από προηγούμενη έρευνα της ομάδας της Facebook [17]. Γραμμικοί ταξινομητές εκπαιδεύονται σε αναπαραστάσεις εικόνων που λαμβάνονται από αυτό-επιβλεπόμενους «μαθητές» που ήταν προεκπαιδευμένοι στο ImageNet (χωρίς ετικέτες). Αναφέρεται η απόδοση για το επίπεδο με την καλύτερη επίδοση για κάθε μέθοδο.

Method	Parameters	Transfer Dataset			
		ImageNet	VOC07	Places205	iNat.
ResNet-50 using evaluation setup of [19]					
Supervised	25.6M	75.9	87.5	51.5	45.4
Colorization [19]	25.6M	39.6	55.6	37.5	–
Rotation [18]	25.6M	48.9	63.9	41.4	23.0
NPID++ [72]	25.6M	59.0	76.6	46.4	32.4
MoCo [24]	25.6M	60.6	–	–	–
Jigsaw [19]	25.6M	45.7	64.5	41.2	21.3
PIRL (**ours**)	25.6M	**63.6**	**81.1**	**49.8**	**34.1**
Different architecture or evaluation setup					
NPID [72]	25.6M	54.0	–	45.5	–
BigBiGAN [12]	25.6M	56.6	–	–	–
AET [76]	61M	40.6	–	37.1	–
DeepCluster [6]	61M	39.8	–	37.5	–
Rot. [33]	61M	54.0	–	45.5	–
LA [80]	25.6M	60.2†	–	50.2†	–
CMC [64]	51M	64.1	–	–	–
CPC [51]	44.5M	48.7	–	–	–
CPC-Huge [26]	305M	61.0	–	–	–
BigBiGAN-Big [12]	86M	61.3	–	–	–
AMDIM [4]	670M	68.1	–	55.1	–

Εικόνα 2.11.3.1 : Ταξινόμηση εικόνας με γραμμικούς ταξινομητές. Με έντονη γραφή είναι οι καλύτερες επιδόσεις στην αυτό-επιβλεπόμενη εκπαίδευση σε κάθε σύνολο δεδομένων.
(Πηγή:[16] https://arxiv.org/pdf/1912.01991.pdf)

3. ΜΕΘΟΔΟΛΟΓΙΑ ΚΑΙ ΥΛΟΠΟΙΗΣΗ

3.1 Μέθοδος

Η παραπάνω ιδέα (Pretext-Invariant Representation Learning) αποτέλεσε και το κίνητρο για μελέτη και πειραματισμούς στην παρούσα διπλωματική εργασία. Συγκεκριμένα, η πειραματική διαδικασία της διπλωματικής εργασίας ακολουθείται επίσης και στις πιο πρόσφατες δημοσιεύσεις με μεγαλύτερες αρχιτεκτονικές που δεν μπορούσαν να υλοποιηθούν εδώ λόγω έλλειψης πόρων.

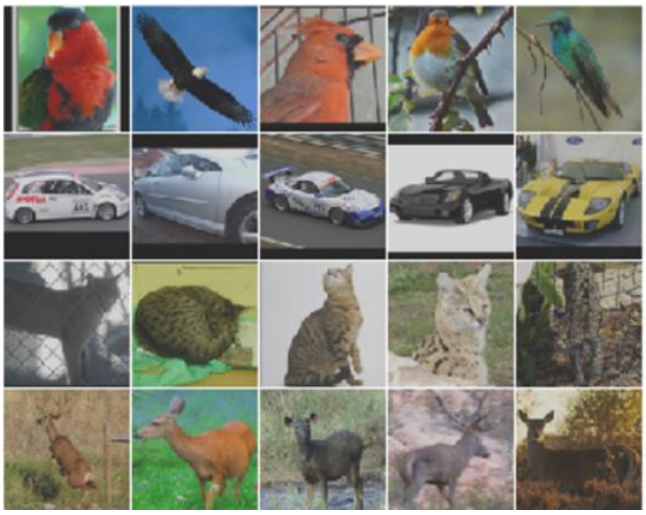
Συνοπτικά, η βασική ιδέα είναι να μείνουν ανεπηρέαστα (invariant) τα δεδομένα μετά τον μετασχηματισμό που υπόκεινται. Δηλαδή τα δεδομένα δεν πρέπει να μάθουν τον εκάστοτε μετασχηματισμό (π.χ. puzzle, rotation), αλλά να είναι αμετάβλητα ως προς αυτόν και η τελική αναπαράσταση να είναι η ίδια αρχική εικόνα. Έτσι, πραγματοποιείται μια σειρά από πειράματα.

Πρώτα γίνεται μια προ-εκπαίδευση των δικτύων σε διάφορα σετ δεδομένων πάνω στο pretext task του Jigsaw puzzle που αναλύθηκε. Έπειτα, μετά την προεκπαίδευση αυτή, επιλέγεται το μοντέλο που έδωσε τα καλύτερα αποτελέσματα και εφαρμόζεται στην συνέχεια στο ζητούμενο downstream task που είναι η ταξινόμηση εικόνας στην προκειμένη περίπτωση. Επιλέχθηκαν 4 διαφορετικά σετ δεδομένων και πραγματοποιήθηκαν σχεδόν όλοι οι δυνατοί συνδυασμοί. Φυσικά, το μοντέλο είναι εφικτό να προ-εκπαιδευτεί και πάνω σε άλλα pretext tasks που προαναφέρθηκαν. Τέλος, υπολογίζεται η ακρίβεια της ταξινόμησης (accuracy), που αποδεικνύει πόσο καλά έχει εκπαιδευτεί το μοντέλο και κατά πόσο μπορεί το δίκτυο να αναγνωρίσει χαρακτηριστικά από εικόνες που δεν έχει ξανασυναντήσει.

3.2 Δεδομένα

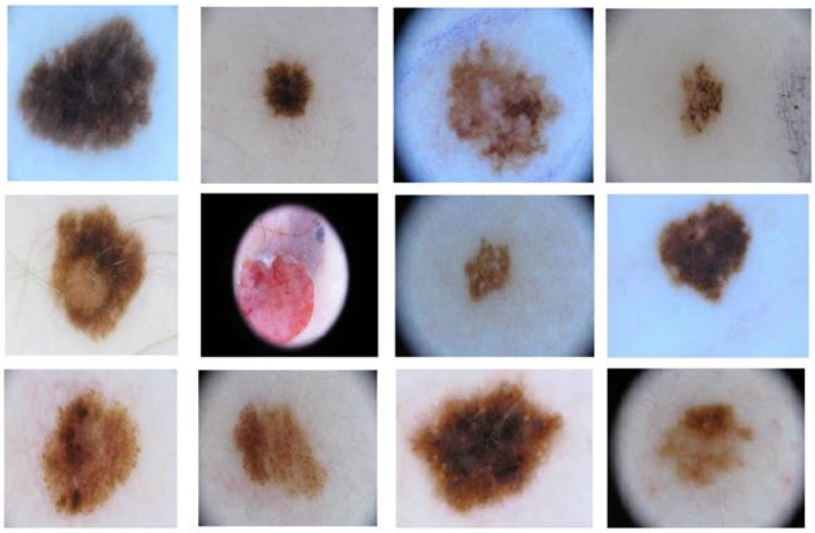
Αρχικά, χρησιμοποιήθηκαν τα σετ δεδομένων του STL10 και Cifar10 τα οποία παρουσιάζουν πολλές ομοιότητες μεταξύ τους. Για λόγους λειτουργικότητας , το σετ δεδομένων Cifar10 υπέστη κάποιες τροποποιήσεις. Οι εικόνες του απο (32×32) pixels έγιναν στην διάσταση (96×96)pixels. Αυτό φυσικά προσθέτει εξτρα θόρυβο στην εικόνα που επηρεάζει και την τελική ακρίβεια .

Εικόνα 3.2.1 : Παράδειγμα αρχικής και μετασχηματιζόμενης εικόνας από το σετ δεδομένων Cifar10 .



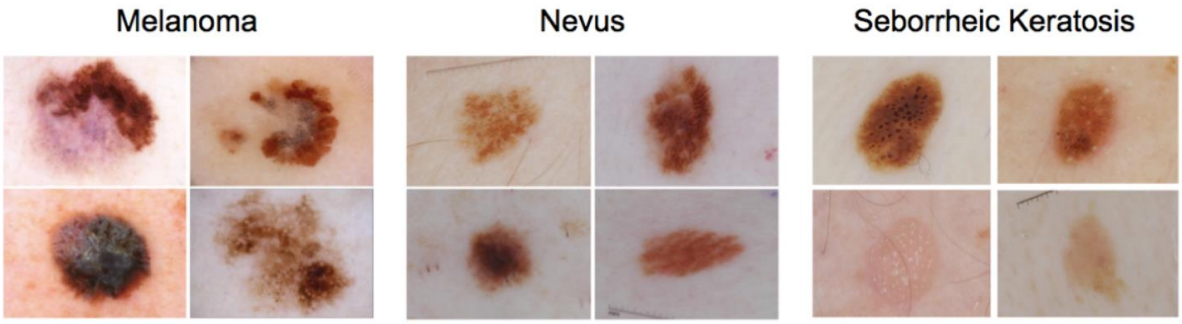
Εικόνα 3.2.2 : Εικόνες από το σετ δεδομένων STL10
(Πηγή : [39] https://cs.stanford.edu/~acoates/stl10/)

Επίσης, χρησιμοποιήθηκε ένα σετ δεδομένων που αφορά δερματικά νοσήματα. Πιο συγκεκριμένα, το σετ αυτό χρησιμοποιήθηκε στον διαγωνισμό ISIC CHALLENGE το 2017. Στόχος του διαγωνισμού είναι να βοηθήσει τους συμμετέχοντες να αναπτύξουν τα εργαλεία ανάλυσης εικόνας για να καταστεί δυνατή η αυτοματοποιημένη διάγνωση του μελανώματος από δερματοσκοπικές εικόνες. Ουσιαστικά πρόκειται για μια εφαρμογή ανάλυσης και ταξινόμησης εικόνας. Η ανάλυση των δερματικών βλαβών αποτελείται από τρια μέρη: τμηματοποίηση των αλλοιώσεων, ανίχνευση και εντοπισμός οπτικών δερμοσκοπικών χαρακτηριστικών-μοτίβων και τέλος ταξινόμηση των ασθενειών. Το σετ περιλαμβάνει 2000 εικόνες για εκπαίδευση. Χρησιμοποιήθηκε το συγκεκριμένο σετ δεδομένων χάρη στον μικρό όγκο του.



Εικόνα 3.2.3 : Εικόνες από το σετ δεδομένων του ISIC Challenge 2017.
(Πηγή: [41] https://challenge.isic-archive.com/data/)

Παρακάτω παρουσιάζονται οι 3 κλάσεις της ταξινόμησης που πραγματοποιείται στον διαγωνισμό. Πρώτη κατηγορία αποτελεί το «Μελάνωμα», και στην συνέχεια υπάρχει η κατηγορία «Σπίλος» και η «Σμηγματορροϊκή κεράτωση».



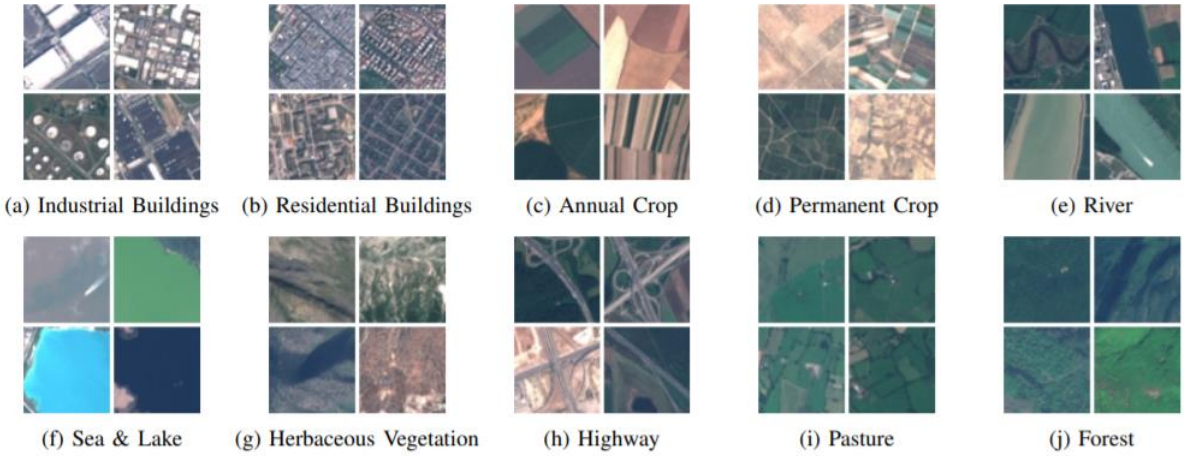
Εικόνα 3.2.4 : Κατηγορίες ταξινόμησης από το σετ δεδομένων του ISIC Challenge 2017.
(Πηγή: [41] https://challenge.isic-archive.com/data/)

Τέλος, χρησιμοποιήθηκε το σετ δεδομένων της EuroSAT. Στην σημερινή εποχή, κυβερνητικά προγράμματα όπως το Copernicus της ESA και το Landsat της NASA καταβάλλουν σημαντικές προσπάθειες για να παρέχουν δημόσια και συνεχής πρόσβαση σε δεδομένα δορυφορικών εικόνων για τη Γη. Τα ελεύθερα και διαθέσιμα δεδομένα αυτά χρησιμοποιούνται για εμπορικούς και μη εμπορικούς σκοπούς με στόχο να τροφοδοτήσουν την καινοτομία και την επιχειρηματικότητα. Το συγκεκριμένο σετ δεδομένων της EuroSAT προέρχεται από τον δορυφόρο Sentinel-2 και χρησιμοποιείται για το έργο της ταξινόμησης χρήσεων και κάλυψης γης. Αποτελείται από 27.000 εικόνες με ετικέτα με 10 διαφορετικές κατηγορίες χρήσεων και κάλυψης γης. Το σετ δεδομένων της EuroSAT περιέχει ακόμη και σετ με πολυφασματικές εικόνες που καλύπτουν 13 φασματικές ζώνες που όμως δεν θα χρησιμοποιηθούν στο συγκεκριμένο πείραμα. Επιπλέον, τα δεδομένα είναι γεωαναφερμένα και βασίζονται σε ανοιχτά και ελεύθερα προσβάσιμα δεδομένα για χρήση και κάλυψη γής που επιτρέπουν μεγάλη γκάμα εφαρμογών.

Εικόνα 3.2.5 : Εικόνες από το σετ δεδομένων της EuroSAT.
(Πηγή: [44] https://github.com/phelber/eurosat)

Πιο συγκεκριμένα, για το πείραμα που πραγματοποιήθηκε στην παρούσα εργασία χρησιμοποιήθηκε το σετ δεδομένων που περιέχει εικόνες με 3 μόνο κανάλια (RGB) και είναι σε διάσταση (64×64) pixels. Στην εικόνα 2.3.5 και 2.3.6 φαίνονται ενδεικτικές εικόνες από κάθε μια κλάση του σετ. Οι κλάσεις αναλυτικά είναι: Βιομηχανικά κτίρια, Κτίρια κατοικιών, Ετήσια

καλλιέργεια, Μόνιμη καλλιέργεια, Ποτάμι, Θάλασσα και λίμνη, Ποώδης Βλάστηση, Αυτοκινητόδρομος, Βοσκότοπος και Δάσος. Κάθε κλάση περιέχει απο 2000 έως και 3000 εικόνες, με τον συνολικό αριθμό εικόνων να είναι στις 27000 εικόνες με γεωαναφορά. [18]

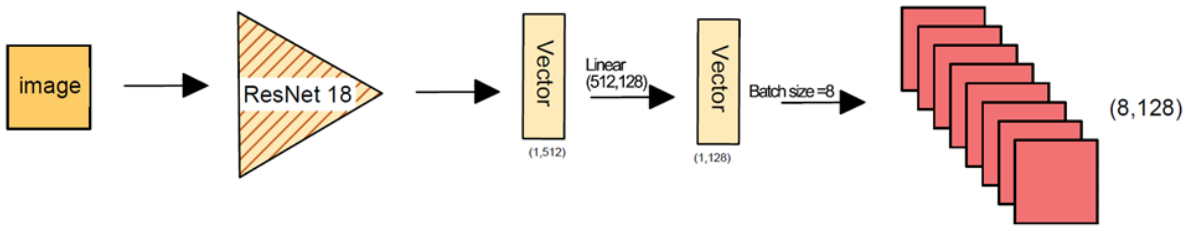


Εικόνα 3.2.6 : Οι 10 κλάσεις του σετ δεδομένων της EuroSAT.
(Πηγή : [18] https://arxiv.org/pdf/1709.00029.pdf)

Όπως αναφέρθηκε και προηγουμένως για το σετ δεδομένων του Cifar10, και σε αυτήν την περίπτωση τα δεδομένα μεγεθύνονται για λόγους λειτουργικότητας και γίνονται στην διάσταση (96×96) pixels. Η ενέργεια αυτή όπως και προηγουμένως, προσθέτει επιπλέον θόρυβο στις εικόνες.

3.3 Υλοποίηση

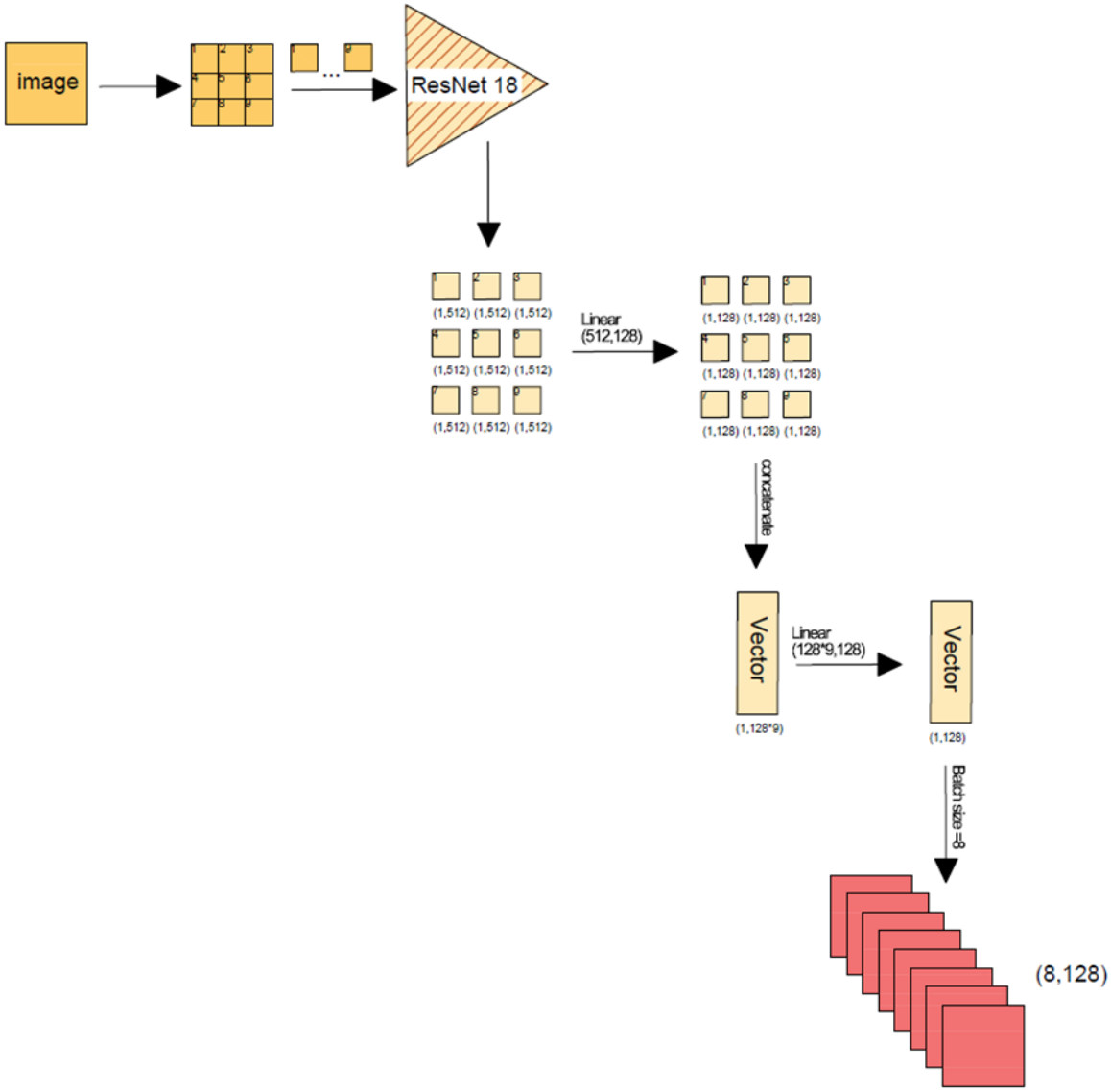
Σε ότι αφορά την υλοποίηση των πειραμάτων, σε θεωρητικό επίπεδο, η εικόνα καθώς εισάγεται στο δίκτυο ακολουθεί μια συγκεκριμένη «διαδρομή». Το ίδιο συμβαίνει και όταν η εικόνα υποστεί κάποιο μετασχηματισμό, όπως ο διαμερισμός της σε 9 επιμέρους κομμάτια ή η περιστροφή της. Η παρακάτω εικόνα απεικονίζει σχηματικά την «διαδρομή» αυτή και δείχνει την γενική ιδέα που ακολουθείται και στην υλοποίηση της παρούσας εργασίας. Για λόγους ευκολίας επιδεικνύεται σαν παράδειγμα η αρχιτεκτονική του μοντέλου ResNet18. Στην παρούσα εργασία χρησιμοποιείται η αρχιτεκτονική του ResNet50 και προκύπτουν τα αντίστοιχα νούμερα και αποτελέσματα.



Εικόνα 3.3.1 : Εισαγωγή και διαδρομή που ακολουθεί η αρχική εικόνα στο δίκτυο.

Όπως φαίνεται και παραπάνω, η εικόνα χωρίς να υποστεί κάποιο μετασχηματισμό, διέρχεται μέσα από το μοντέλο του ResNet18. Μέσω της αρχιτεκτονικής του μοντέλου, προκύπτει ένα διάνυσμα διαστάσεων (1,512). Στην συνέχεια το διάνυσμα αυτό διέρχεται από ένα Fully Connected Layer

διάστασης (512,128) και ως αποτέλεσμα δίνει ένα ακόμη διάνυσμα διαστάσεων (1,128) αφού γίνεται πολλαπλασιασμός των πινάκων. Επειδή το mini batch που έχει οριστεί είναι 8, στο τέλος προκύπτει ένα σύνολο από εικόνες-διανύσματα διάστασης (8,128).



Εικόνα 3.3.2 : Εισαγωγή και διαδρομή που ακολουθεί η αρχική εικόνα στο δίκτυο μετά τον μετασχηματισμό της.

Στην προκειμένη περίπτωση, επιδεικνύεται σχηματικά η διαδρομή που ακολουθεί η μετασχηματισμένη εικόνα. Αρχικά, γίνεται ο διαμελισμός της σε 9 επιμέρους κομμάτια, τα οποία εισάγονται ένα ένα στο μοντέλου του ResNet18. Όπως και προηγουμένως, προκύπτουν διανύσματα διαστάσεων (1,512), όμως αυτήν τη φορά είναι 9 μικρά διανύσματα. Έπειτα, το καθένα από αυτά διέρχεται από ένα Linear Fully Connected επίπεδο διάστασης (512,128) και προκύπτουν 9 διανύσματα διάστασης (1,128). Η εντολή concatenate που φαίνεται λαμβάνει ως είσοδο μια λίστα τανυστών-tensors, που όλοι έχουν το ίδιο σχήμα εκτός από τον άξονα συνένωσης, και επιστρέφει έναν μόνο τανυστή που είναι η συνένωση όλων των εισόδων. Έτσι η χρήση του concatenate δίνει ένα διάνυσμα διάστασης (1,128×9)=(1,1152). Ξανά το vector αυτό περνά μέσω ενός Linear επιπέδου διάστασης (128×9,128) και το αποτέλεσμα που δίνει είναι ένα διάνυσμα

διάστασης (1,128) όπως και προηγουμένως. Το mini batch είναι και πάλι 8 και έτσι προκύπτει δια δέσμη από εικόνες-διανύσματα συνολικής διάστασης (8,128).

Η γενική ιδέα είναι πως πρέπει αυτά τα 2 σύνολα διανυσμάτων, οι αναπαραστάσεις αυτές πρέπει να είναι ίδιες. Δηλαδή, ανεξαρτήτως του μετασχησμού που υπέστη η εικόνα, οι τελικές αναπαραστάσεις πρέπει να έχουν τα ίδια ακριβώς χαρακτηριστικά όταν περάσουν μέσα από την αρχιτεκτονική που περιγράφηκε παραπάνω.

Στην παρούσα εργασία, όπως προαναφέρθηκε υλοποιείται αρχικά το pretext task του Jigsaw puzzle και προεκπαιδεύεται το δίκτυο. Έπειτα επιλέγεται το καλύτερο μοντέλο από την προεκπαιδευση και εφαρμόζεται σε εργασία ταξινόμησης εικόνας. Ωστόσο, σε αυτό το στάδιο πραγματοποιείται εκπαίδευση μόνο σε έναν linear classifier. Ουσιαστικά σε όλα τα πειράματα γίνεται εκπαίδευση-train στο backbone (εκμάθηση των representations) σε ένα dataset με το pretext invariance task και μετά πραγματοποιείται train **MONO** σε έναν classifier στο downstream dataset.

3.4 Πειραματική διαδικασία

Όπως προαναφέρθηκε χρησιμοποιούνται 4 σετ δεδομένων με τους περισσότερους δυνατούς συνδυασμούς σε pretext και downstream task.

- ✓ Αρχικά, στο **1ο πείραμα** πραγματοποιείται προεκπαίδευση με Jigsaw puzzle στο σετ δεδομενων STL10, και έπειτα το ίδιο σετ δοκιμάζεται και στο downstream task.
- ✓ Το ίδιο ακριβώς συμβαίνει και στο **2ο πείραμα** με το μετασχηματισμένο σετ δεδομένων Cifar10.
- ✓ Στο **3ο πείραμα** πραγματοποιείται προεκπαίδευση με Jigsaw puzzle στο σετ δεδομενων STL10 και downstream task στο Cifar10.
- ✓ Στο **4ο πείραμα** πραγματοποιεί ακριβώς η ανάποδη διαδικασία του 3ου πειράματος, τεστάροντας το καλύτερο μοντέλο του pretext task στο σετ Cifar10 σε εργασία ταξινόμησης εικόνων του σετ δεδομένων STL10.
- ✓ Για **5ο πείραμα** χρησιμοποιείται στο pretext task το σετ με τα δερματικά νοσήματα (ISIC 2017) και στην εργασία ταξινόμησης δοκιμάζεται το σετ δεδομενων STL10.
- ✓ Ομοίως, στο **6ο πείραμα** γίνεται προεκπαίδευση στο σετ δεδομενων STL10 και για την εργασία ταξινόμησης δίνεται το σετ με τα δερματικά νοσήματα.
- ✓ Στο **7ο πείραμα** και το pretext και το downstream task γίνονται στο σετ δεδομένων του ISIC 2017.
- ✓ Τέλος στο **8ο πείραμα** πραγματοποιείται και pretext και downstream task στο σετ δεδομένων της EuroSAT.

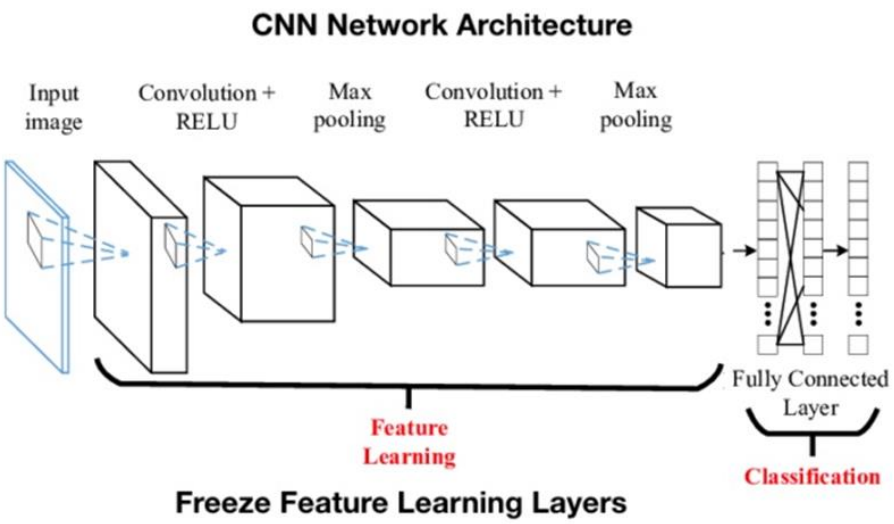
Πιο συνοπτικά τα πειράματα παρουσιάζονται στον παρακάτω πίνακα:

	Pretext task (Jigsaw puzzle)	Downstream task (Image Classification)
Dataset	STL10	STL10
	Cifar10	Cifar10
	STL10	Cifar10
	Cifar10	STL10
	ISIC 2017	STL10
	STL10	ISIC 2017
	ISIC 2017	ISIC 2017
	EuroSAT	EuroSAT

Πίνακας 3.4.1 : Πειράματα και σετ δεδομένων.

Το Batch size που χρησιμοποιείται είναι μικρό, και συγκεκριμένα 8, καθώς όσο πιο μεγάλος είναι ο αριθμός τόσο περισσότερος υπολογιστικός χώρος δεσμεύεται στην μνήμη του υπολογιστή. Όλα τα πειράματα πραγματοποιήθηκαν για 1000 εποχές.

Στην διαδικασία του downstream task καθώς η εικόνα εισάγεται στο δίκτυο περνάει από το κομμάτι που γίνεται η εξαγωγή των χαρακτηριστικών (feature extractor). Στην προκειμένη περίπτωση γίνεται «πάγωμα»-freeze των Convolution επιπέδων και η εικόνα περνά απευθείας στον classifier που είναι το τελευταίο επίπεδο. Ουσιαστικά με το πάγωμα αυτό εκπαιδεύεται μόνο ο classifier, ο οποίος δίνει σαν αποτέλεσμα ένα διάνυσμα διαστάσεων όσες και οι κατηγορίες των κλάσεων.



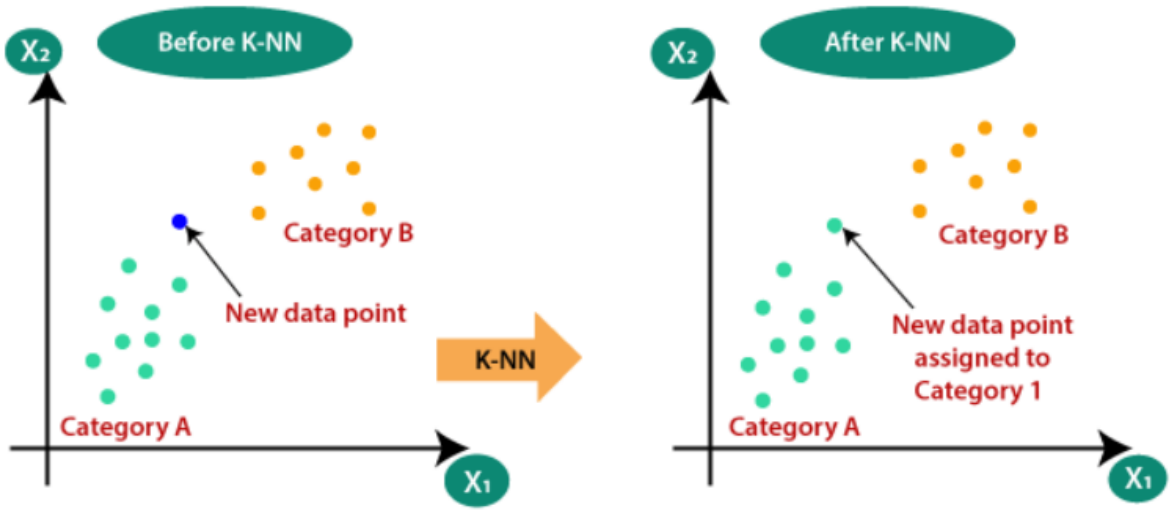
Εικόνα 3.4.1 : Παράδειγμα σχήματος για την απεικόνιση των παγωμένων επιπέδων.
(Πηγή: [42] https://spr.com/can-deep-transfer-learning-help-detect-covid-19/)

4. ΑΠΟΤΕΛΕΣΜΑΤΑ ΚΑΙ ΑΞΙΟΛΟΓΗΣΗ

Στην ενότητα αυτή παρατίθενται τα αποτελέσματα της ακρίβειας του downstream task, δηλαδή της εργασίας ταξινόμησης εικόνας. Από το πρώτο κομμάτι των πειραμάτων, πραγματοποιείται η εργασία pretext-task και παρατηρείται ότι η συνάρτηση loss μειώνεται όσο περνάνε οι εποχές, όμως αυτό το κεφάλαιο αφορά μόνο την δεύτερη φάση των πειραμάτων. Οι ακρίβειες που παρουσιάζονται είναι η ακρίβεια της εκπαίδευσης-Train accuracy και η Validation accuracy που δείχνει δηλαδή την ικανότητα γενίκευσης του μοντέλου. Ουσιαστικά η Validation ακρίβεια είναι το μέτρο για το εάν το δίκτυο μπορεί να αναγνωρίσει εικόνες που δεν έχει ξανασυναντήσει.

Επίσης για την οπτικοποίηση του αποτελέσματος της ακρίβειας, παρατίθενται και τα διαγράμματα ακρίβειας(%)-εποχών που δημιουργήθηκαν χρησιμοποιώντας Matplotlib, μια βιβλιοθήκη σχεδίασης για τη γλώσσα προγραμματισμού Python. Επιπλέον, γίνεται μια ποιοτική απεικόνιση των αποτελεσμάτων χρησιμοποιώντας τον αλγόριθμο k-NN. Γενικότερα ο αλγόριθμος k-NN κατατάσει στις ήδη υπάρχουσες κλάσεις τα νέα δείγματα. Η λειτουργία του αλγορίθμου λειτουργεί ως εξής: επιλέγεται ο αριθμός k των γειτόνων, υπολογίζεται η ευκλείδεια απόσταση των k δειγμάτων των γειτόνων, στην συνέχεια επιλέγονται οι k πλησιέστεροι γείτονες σύμφωνα με την υπολογισμένη ευκλείδεια απόσταση, έπειτα μετριέται ο αριθμός των σημείων δεδομένων κάθε κατηγορίας και τέλος τα νέα σημεία δεδομένων κατατάσσονται σε εκείνη την κατηγορία για την οποία ο αριθμός του γείτονα είναι ο μέγιστος. [45]

Στην παρακάτω εικόνα παρουσιάζεται η λειτουργία του αλγορίθμου k-NN:



Εικόνα 4.1 : Λειτουργία του αλγορίθμου k-NN για ταξινόμηση νέων δεδομένων.
(Πηγή : [45] https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine-learning)

Όπως και στην παραπάνω επεξήγηση με τα σημεία, έτσι και στην περίπτωση των εικόνων, αυτές κατατάσσονται σε κάποια από τις υπαρχουσες κλάσεις. Επιλέγονται 5 τυχαία δείγματα και με βάση τα χαρακτηριστικά των εικόνων, αυτά κατατάσσονται σε κατηγορίες. Έτσι εάν κάποια από

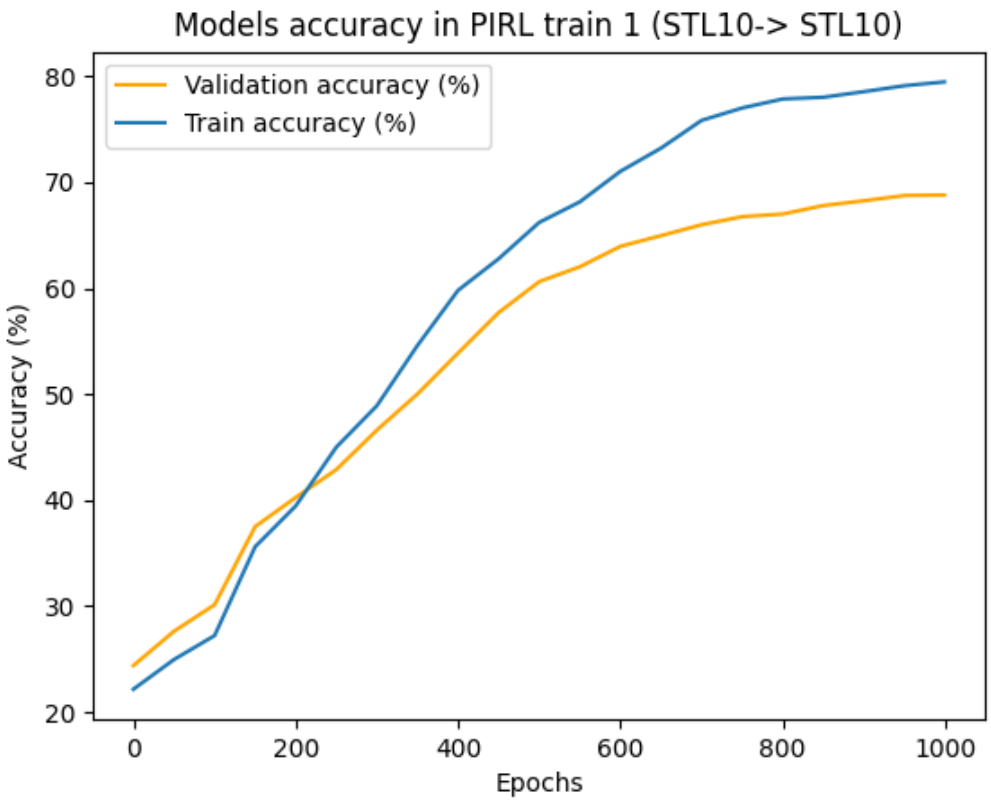
τα 5 δείγματα ανήκουν στην ίδια κλάση- έχουν κοινά χαρακτηριστικά με την αρχική εικόνα, σηματοδοτούνται με πράσινο χρώμα στο περίγραμμά τους. Διαφορετικά, το σύστημα αναγνωρίζει πως ανήκουν σε διαφορετικές κλάσεις και περιβάλλονται από ένα κόκκινο περίγραμμα.

4.1 Πρώτο πείραμα

Για το πείραμα αυτό τα αποτελέσματα της τελικής ακρίβειας, δηλαδή της εργασίας ταξινόμησης εικόνας, φαίνονται στο παρακάτω διάγραμμα καθώς και στον Πίνακα 1 του παραρτήματος. Χρησιμοποιείται το σετ δεδομένων STL10 και στην περίπτωση της προ-εκπαίδευσης και στην περίπτωση της ταξινόμησης.

Η μεγαλύτερη επί τις εκατό ακρίβεια φαίνεται να είναι στην τελευταία εποχή με τιμή 68,79%. Απο τις τελευταίες εποχές φαίνεται πως η Validation ακρίβεια σταθεροποιείται κοντά στην τιμή 68%.

Περίπου στη εποχή 200 η ακρίβεια της εκπαίδευσης και η Validation ακρίβεια, ταυτίζονται με τιμή κοντά στο 40%. Όσο το διάγραμμα προχωράει προς τα αριστερά υπάρχει σταθερή διαφορά ανάμεσα στις δύο ακρίβειες περίπου στο 10% και δεν φαίνεται να υπάρχει overfit στο μοντέλο.



Εικόνα 4.1.1 : Επί της εκατό ακρίβεια με την πάροδο των εποχών για το 1ο πείραμα.

Παρακάτω στην εικόνα 4.1.2 γίνεται μια ποιοτική παρουσίαση των αποτελεσμάτων. Αρχικά επιλέγεται μια τυχαία εικόνα απο μια κλάση. Στην συγκεκριμένη περίπτωση είναι το αυτοκίνητο. Έπειτα αφού έχει εκπαιδευτεί το μοντέλο για την ταξινόμηση , χρησιμοποιείται η μέθοδος του εγγύτερου γείτονα (k-NN) που αναλύθηκε παραπάνω. Στο κάτω αριστερά μέρος τις κάθε εικόνας

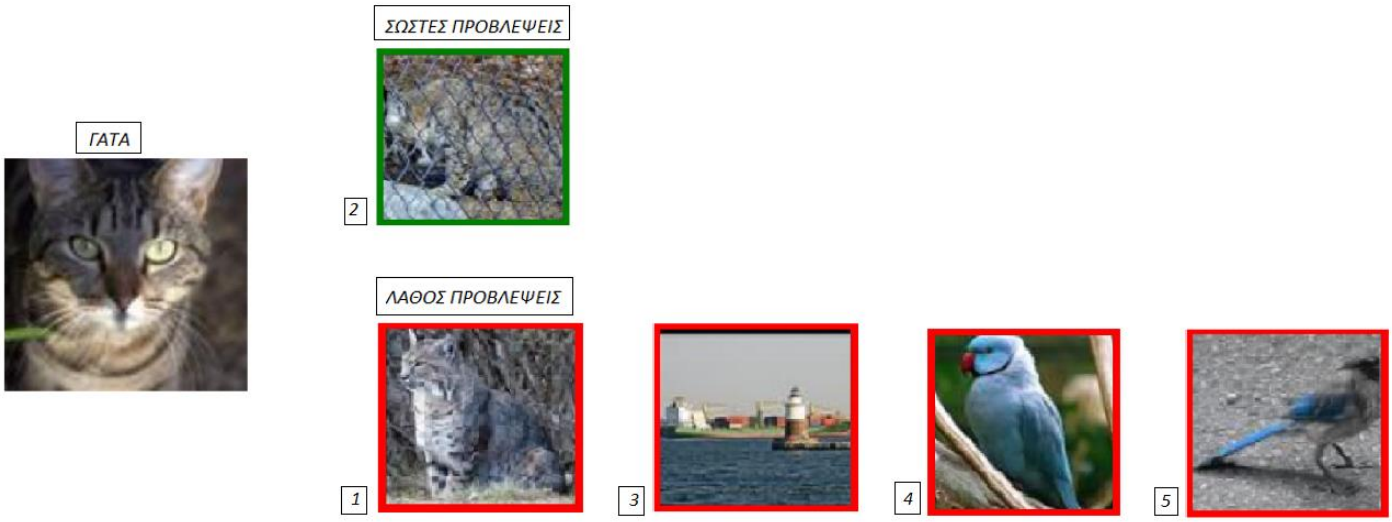
φαίνεται ο αριθμός του κάθε «γείτονα». Ο πιο κοντινός γείτονας είναι με το νούμερο 1 και ο πιο μακρινός με το νούμερο 5.

Στην προκειμένη περίπτωση, το σύστημα αναγνωρίζει σωστά τα 3 αυτοκίνητα από τις εικόνες των δειγμάτων και φαίνεται πως υπάρχει και διαχωρισμός με το φορτηγό που φαίνεται στην τελευταία εικόνα από τις λάθος προβλέψεις.

Εικόνα 4.1.2. : Παράδειγμα σωστών και λανθασμένων προβλέψεων για την κατηγορία «Αυτοκίνητο».

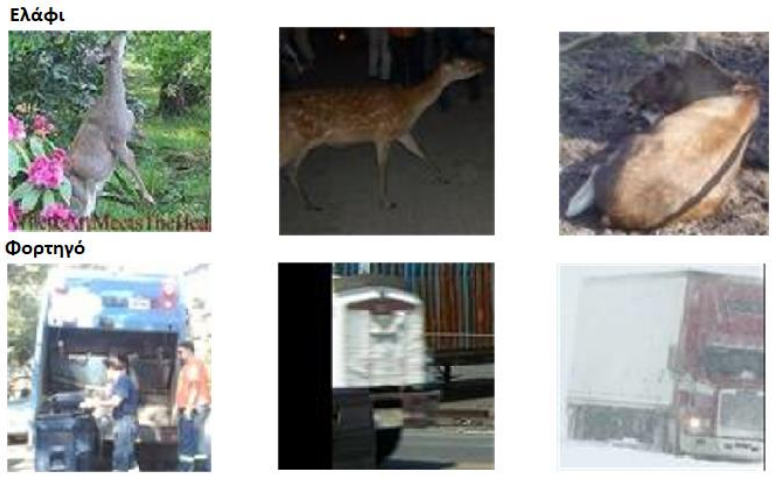
Ωστόσο η ακρίβεια 68% δείχνει πως πιθανον υπάρχουν κάποιες κατηγορίες που περιέχουν εικόνες που δεν αναγνωρίζονται εύκολα από το δίκτυο. Στην περίπτωση του αυτοκινήτου, η αναγνώριση έγινε πολύ σωστά παρόλο που οι εικόνες των δειγμάτων των αυτοκινήτων αποτελούνται και από παλαιότερα αυτοκίνητα και φαίνεται μόνο ένα μέρος τους.

Στην εικόνα 4.1.3. όπως φαίνεται στις σωστές προβλέψεις το δίκτυο έχει αναγνωρίσει ορθά μια γάτα. Ωστόσο, η πρώτη εικόνα των λανθασμένων προβλέψεων ανήκει επίσης στην κατηγορία της γάτας και μάλιστα είναι ο «κοντινότερος γείτονας». Αυτή είναι μια από τις περιπτώσεις που το σύστημα δεν αναγνωρίζει σωστά την κλάση της εικόνας. Αυτό μπορεί να οφείλεται στο γεγονός ότι το υπόβαθρο της φωτογραφίας είναι πανομοιότυπο με το τρίχωμα της γάτας και έτσι ίσως είναι δύσκολο να γίνει ο διαχωρισμός. Παρόλα αυτά το δίκτυο αναγνωρίζει σωστά τα υπόλοιπα δείγματα και δεν τα συγχέει με την κατηγορία της γάτας.



Εικόνα 4.1.3. : Παράδειγμα σωστών και λανθασμένων προβλέψεων για την κατηγορία «Γάτα» του σετ STL10.

Τέλος στην παρακάτω εικόνα παρουσιάζονται κάποιες εικόνες που δημιουργούν πρόβλημα στην σωστή ταξινόμηση λόγω των ιδιαίτερων συνθηκών που έχουν παρθεί οι φωτογραφίες. Για παράδειγμα η δεύτερη εικόνα της κλάσης «Ελάφι» είναι αρκετά σκοτεινή. Επίσης η τρίτη εικόνα της κλάσης «Φορτηγό» λόγω καιρικών συνθηκών, έχει πολύ θόρυβο και πιθανόν αυτό να δημιουργεί πρόβλημα στην σωστή κατάταξή τους.



Εικόνα 4.1.4. : Παραδείγματα εικόνων του σετ STL10 που δεν ταξινομήθηκαν σωστά.

Στην εικόνα 4.1.5 φαίνεται ακόμη ένα παράδειγμα σωστής ταξινόμησης. Όλα τα δείγματα κατατάχθηκαν σωστά και δεν υπήρχε κάποια σύγχυση στο δίκτυο. Παρόλο που το πτηνό της εικόνας αναφοράς διαφέρει αρκετά με τα δύο πτηνά που καλείται το δίκτυο να αναγνωρίσει, αυτό τα καταφέρνει με μεγάλη επιτυχία.

Εικόνα 4.1.5. : Παράδειγμα σωστών και λανθασμένων προβλέψεων για την κατηγορία «Πτηνό» του σετ STL10.

Τέλος παρουσιάζεται ακόμη ένα παράδειγμα για το συγκεκριμένο πείραμα. Η κλάση αυτή τη φορά είναι: «Άλογο». Λόγω της καλής ακρίβειας της ταξινόμησης που φτάνει σχεδόν το 70%, παρατηρείται πως στο συγκεκριμένο παράδειγμα έχει γίνει μόνο ένα λάθος. Η εικόνα με αριθμό 4, δεν έχει μπεί στις σωστές προβλέψεις πιθανόν λόγω της παρουσίας ανθρώπου που καταλαμβάνει μεγάλο όγκο στην εικόνα. Ωστόσο το δίκτυο έχει ξεχωρίσει τα δύο ελάφια και το σκύλο που ήταν πιθανό να μπερδευτούν με το άλογο καθώς έχουν τους ίδιου τόνους χρωματικά και παρόμοια εμφάνιση.

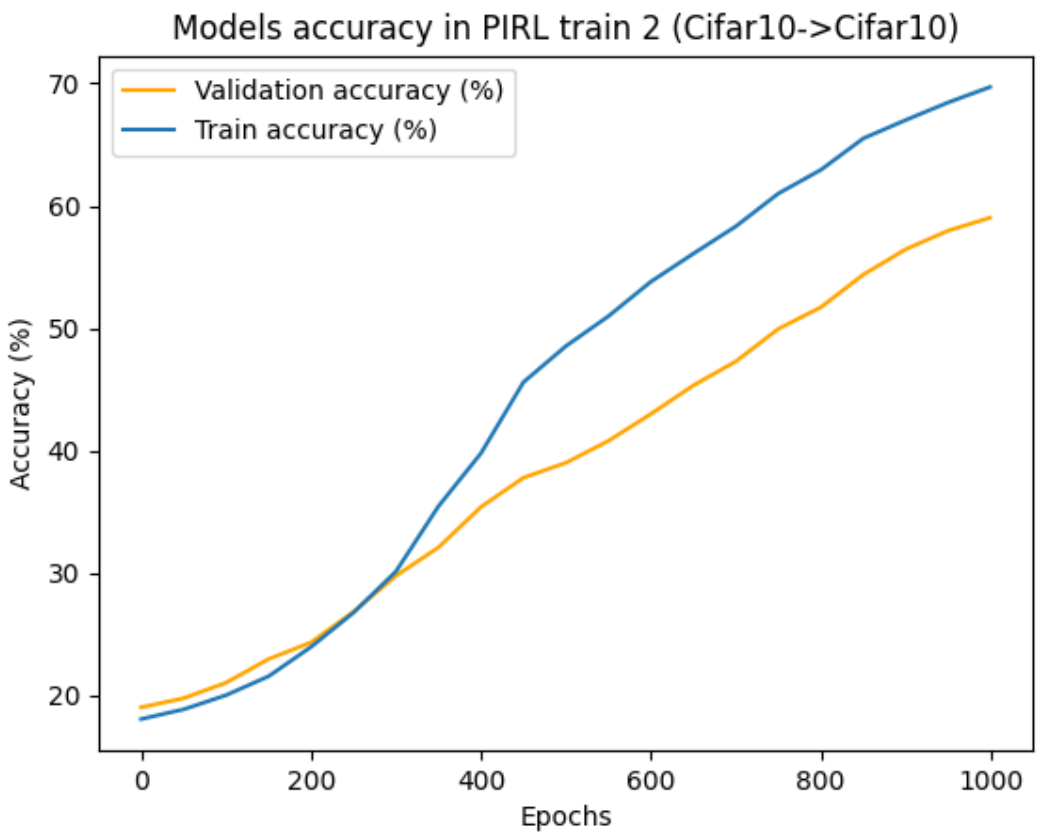
Εικόνα 4.1.6. : Παράδειγμα σωστών και λανθασμένων προβλέψεων για την κατηγορία «Άλογο» του σετ STL10.

4.2 Δεύτερο πείραμα

Αντίστοιχα τα αποτελέσματα για το δεύτερο πείραμα παρατίθενται στον Πίνακα 2 του παραρτήματος και φαινονται σχηματικά στον παρακάτω διάγραμμα. Τα δεδομένα που χρησιμοποιούνται είναι αυτά του CIFAR-10 και στην εργασία προ-εκπαίδευσης και στην εργασία ταξινόμησης.

Σε αυτό το πείραμα η μέγιστη ακρίβεια φτάνει σχεδόν στο 60%. Παρά το γεγονός ότι τα σετ δεδομένων του Cifar 10 και του STL10 είναι όμοια, σε αυτό το πείραμα σε σχέση με το προηγούμενο η Validation accuracy είναι πεσμένη πιθανώς και λόγω της μεγέθυνσης που έχουν υποστεί οι εικόνες. Απο (32×32) pixels έγιναν στην διάσταση (96×96) pixels και αυτό πρόσθεσε επιπλέον θόρυβο στις εικόνες και έτσι η τελική ακρίβεια είναι χαμηλότερη συγκριτικά με το πρώτο πείραμα.

Σχηματικά, φαίνεται στην παρακάτω εικόνα πως κοντά στην εποχή 300 οι 2 ακρίβειες συμπίπτουν και υπάρχει σταθερή διαφορά μεταξύ τους (>10%) καθώς οι εποχές προχωρούν. Η ακρίβεια φαίνεται πως τείνει να σταθεροποιηθεί κοντά στο 60% μετά τις 900 εποχές.



Εικόνα 4.2.1 : Επί της εκατό ακρίβεια με την πάροδο των εποχών για το 2ο πείραμα.

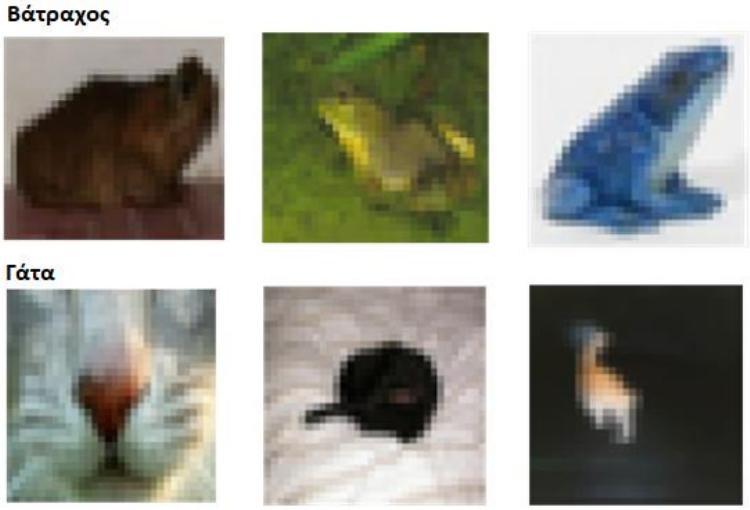
Στην παρακάτω εικόνα παρουσιάζεται ένα παράδειγμα λανθασμένης ταξινόμησης σύμφωνα με την μέθοδο του KNN με 5 τυχαία δείγματα και την εικόνα αναφοράς να ανήκει στην κλάση «Σκύλος». Όπως διακρίνεται στην εικόνα 4.2.2., το δίκτυο έχει προβλέψει σωστά 2 εικόνες σκύλων,

όμως στην δεύτερη εικόνα των λανθασμένων προβλέψεων το δίκτυο έχει κάνει λάθος. Πιθανόν λόγω της στάσης που έχει πάρει ο σκύλος στην συγκεκριμένη λήψη το σύστημα αδυνατεί να αναγνωρίσει τα χαρακτηριστικά του σκύλου παρόλο που τα χρώματα των δύο σκύλων είναι στους ίδιους τόνους. Ωστόσο, μπορεί εύστοχα να διακρίνει την διαφορά γάτας-σκύλου καθώς κατατάσει στα λανθασμένα αποτελέσματα την γάτα με τον αριθμό 4.

Εικόνα 4.2.2. : Παράδειγμα σωστών και λανθασμένων προβλέψεων για την κατηγορία «Σκύλος» του σετ Cifar10.

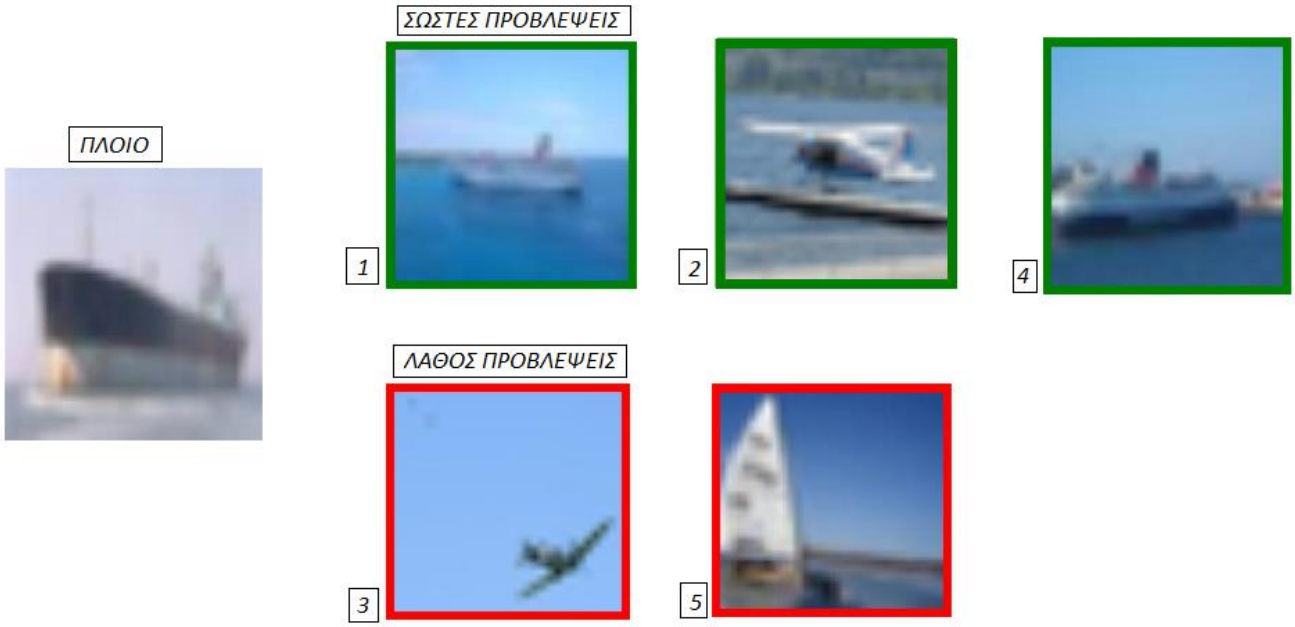
Επίσης παρουσιάζονται ενδεικτικά και 3 εικόνες (4.2.3.) από τις κλάσεις : «Βάτραχος» και «Γάτα» οι οποίες δεν κατατάχθηκαν στις σωστές κλάσεις. Αρχικά στην κλάση «Βάτραχος» παρατηρείται πως τα χρώματα του υποβάθρου αλλά και του ίδιου του βατράχου δεν βοηθούν στην σωστή ταξινόμηση. Ο βάτραχος της πρώτης εικόνας δεν έχει κάποιο ιδιαίτερο μοτίβο στο σώμα του, και η στάση του στην συγκεκριμένη εικόνα δεν είναι χαρακτηριστική και δεν παραπέμπει σε αυτήν την κατηγορία. Αντίστοιχα και ο βάτραχος την τελευταίας εικόνας έχει πιο σπάνιο χρώμα και αυτό μπερδεύει το δίκτυο. Τέλος ο δεύτερος βάτραχος έχει σχεδόν τους ίδιους τόνους πράσινου με το υπόβαθρο και δεν διακρίνεται εμφανώς.

Ομοίως και στην περίπτωση της κλάσης «Γάτα» , οι 3 εικόνες που φαίνονται παραπάνω δεν ταξινομήθηκαν σωστά. Στην πρώτη εικόνα παρουσιάζεται η μουσούδα της γάτας χωρίς να φαίνονται λοιπά χαρακτηριστικά της. Στην δεύτερη και τρίτη εικόνα φαίνονται δύο γάτες οι οποίες λόγω των συνθηκών που τραβήχτηκε η φωτογραφία δεν είναι εύκολα αναγνωρίσιμες. Στις εικόνες αυτές ο φωτισμός είναι σχετικά χαμηλός και δεν διακρίνονται εύκολα τα χαρακτηριστικά των γατων. Λόγω και την μεγέθυνσης των εικόνων που αναφέρθηκε παραπάνω, έχει προστεθεί περισσότερος θόρυβος στις ήδη σκοτεινές εικόνες και το δίκτυο αδυνατεί να διακρίνει χαρακτηριστικά της γάτας για να μπορέσει να την κατατάξει στην σωστή κατηγορία.



Εικόνα 4.2.3. : Εικόνες που δεν ταξινομήθηκαν σωστά απο τις κλάσεις «Βάτραχος» και «Γάτα» του σετ Cifar10.

Ακόμη, στην εικόνα 4.2.4. παρουσιάζεται ένα παράδειγμα για την κλάση «Πλοίο». Στην περίπτωση αυτή δεν ταξινομούνται σωστά δύο από τις πέντε εικόνες. Στις σωστές προβλέψεις η δεύτερη εικόνα φαίνεται φανερά πως είναι ένα αεροπλάνο αλλά επειδή βρίσκεται στην επιφάνεια του νερού το δίκτυο πιθανόν το έχει μπερδέψει με πλοίο. Αντίστοιχα στην δεύτερη εικόνα των λάθος προβλέψεων το πλοίο δεν αναγνωρίζεται πιθανόν γιατί έχει πανιά και δεν είναι σαν τα υπόλοιπα πλοία που έχει συναντήσει προηγουμένως το δίκτυο.

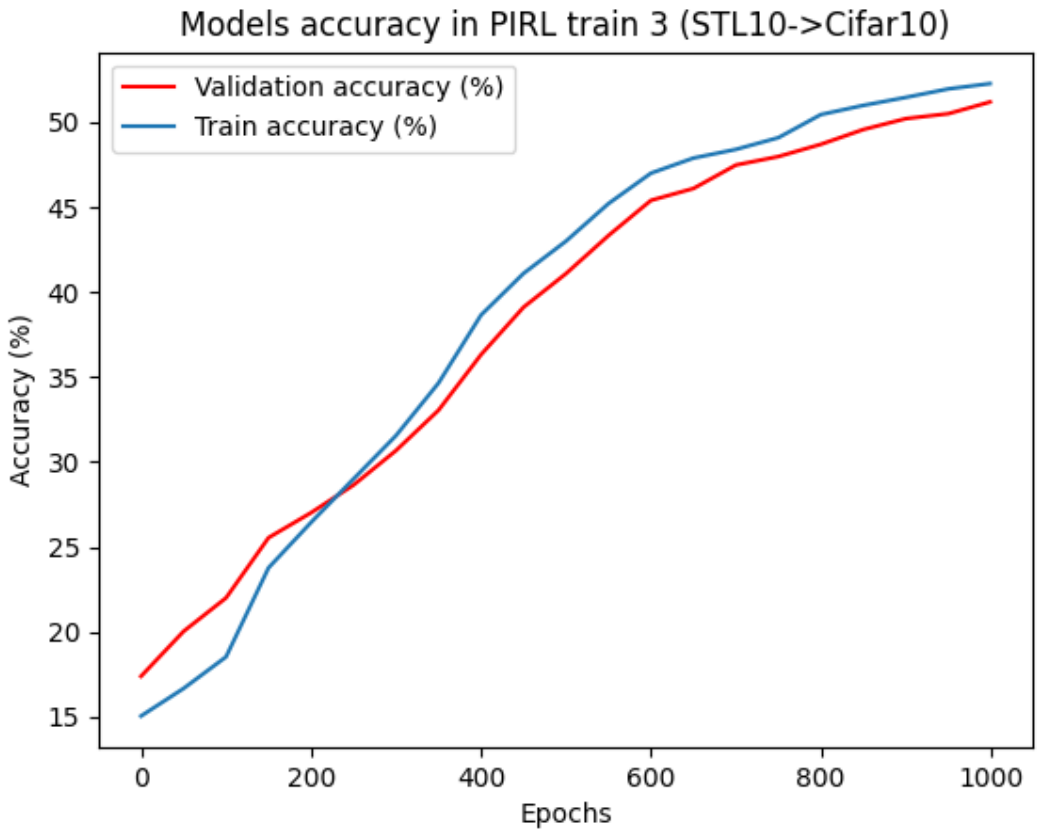


Εικόνα 4.2.4. : Παράδειγμα σωστών και λανθασμένων προβλέψεων για την κατηγορία «Πλοίο» του σετ Cifar10.

4.3 Τρίτο πείραμα

Στο πείραμα αυτό πραγματοποιήθηκε προεκπαίδευση σε Jigsaw puzzle στο σετ δεδομένων STL10 και έπειτα, το βέλτιστο μοντέλο δοκιμάστηκε σε εργασία ταξινόμησης εικόνας του σετ Cifar10. Τα αποτελέσματα φαίνονται παρακάτω στο διάγραμμα.

Στην προκειμένη περίπτωση η ακρίβεια φτάνει έως 51%. Όπως όμως φαίνεται και απο τον Πινακα 4.3 και από την εικόνα παρακάτω, η Train accuracy και η Validation accuracy, είναι πολύ κοντά μεταξύ τους. Το γεγονός αυτό σημαίνει πως το μοντέλο μπορεί να γενικεύει και να αναγνωρίζει εικόνες που δεν έχει ξανασυναντήσει. Φυσικά το γεγονός ότι τα 2 σετ δεδομένων μοιάζουν ως προς το περιεχόμενο των εικόνων βοηθάει στο συγκεκριμένο πείραμα.



Εικόνα 4.3.1: Επί της εκατό ακρίβεια με την πάροδο των εποχών για το 3ο πείραμα.

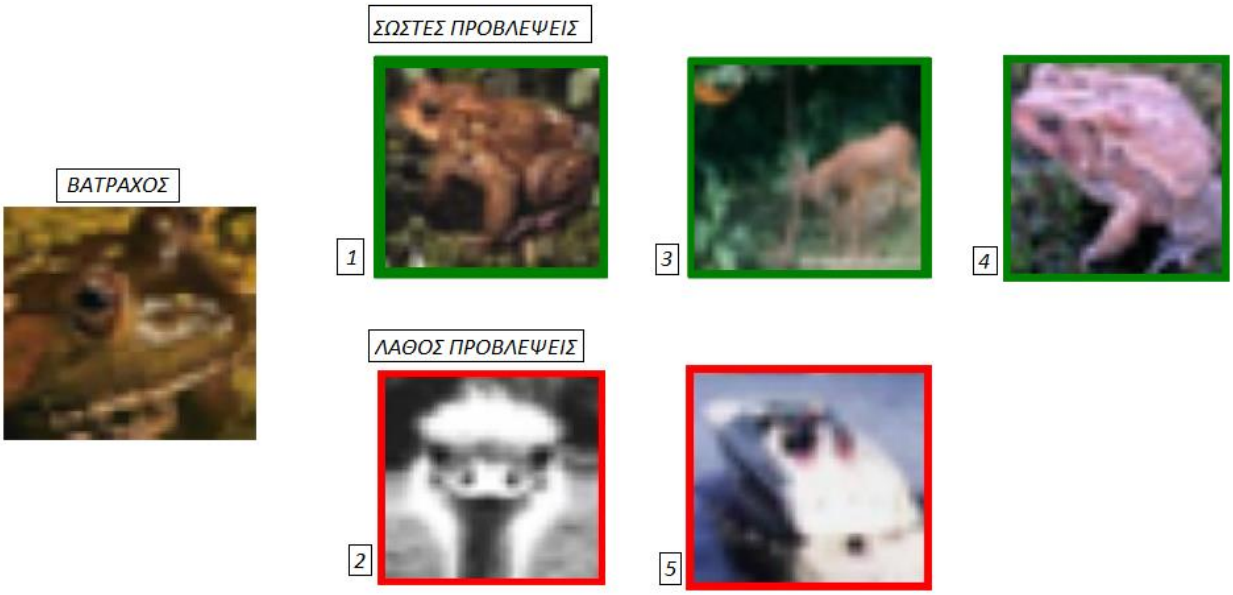
Η ακρίβεια 51% σημαίνει πως το δίκτυο δεν μπορεί να αναγνωρίσει πολλές νέες εικόνες. Χαρακτηριστικό παράδειγμα παρουσιάζεται στην εικόνα 4.3.2 όπου το δίκτυο καλείται να ταξινομήσει σωστά 5 εικόνες εκ των οποίων οι 4 ανήκουν στην κατηγορία αναφοράς. Η εικόνα αναφοράς είναι η κλάση «Αεροπλάνο». Όπως μπορεί εύκολα να παρατηρηθεί, το δίκτυο ταξινομεί σωστά τις 2 εικόνες από τις 4 εικόνες που περιέχουν αεροπλάνα. Ωστόσο, οι 2 τελευταίες εικόνες των λαθος προβλέψεων ανήκουν και αυτές στην κατηγορία «Αεροπλάνο». Το δίκτυο μπορεί να

διακρίνει την διαφορά ανάμεσα σε αεροπλάνο και πλοίο, όμως μπερδεύει την δεύτερη εικόνα με κάποιο όχημα και την τρίτη εικόνα με κάποιο πετούμενο πτηνό.

Ουσιαστικά αυτό το παράδειγμα δείχνει πως το δίκτυο αναγνωρίζει σωστά σχεδόν τις μισές από τις εικόνες που θα του δοθούν για να ταξινομήσει. Αυτό δείχνει και η ακρίβεια 51%. Όπως ήδη έχει αναφερθεί το δίκτυο εκπαιδεύτηκε σε ένα παρόμοιο σετ δεδομένων (STL10) και τώρα καλείται να αναγνωρίσει νέες παρόμοιες εικόνες, με περισσότερο θόρυβο, που δεν έχει όμως ξανασυναντήσει (Cifar10).

Εικόνα 4.3.2. : Παράδειγμα σωστών και λανθασμένων προβλέψεων για την κατηγορία «Αεροπλάνο» για το σετ Cifar10.

Ομοίως στην παρακάτω εικόνα φαίνεται ακόμη ένα παράδειγμα που το δίκτυο έχει κάνει λάθος σε μία από τις πέντε εικόνες που καλείται να αναγνωρίσει. Η εικόνα με τον αριθμό 3 του εγγύτερου γείτονα, είναι ελάφι και όμως το δίκτυο το αναγνωρίζει σαν βάτραχο. Αυτό είναι πιθανόν να συμβαίνει λόγω των χρωμάτων του συγκεκριμένου βατράχου και του ελαφιού. Ωστόσο οι υπόλοιπες εικόνες αναγνωρίζονται σωστά.



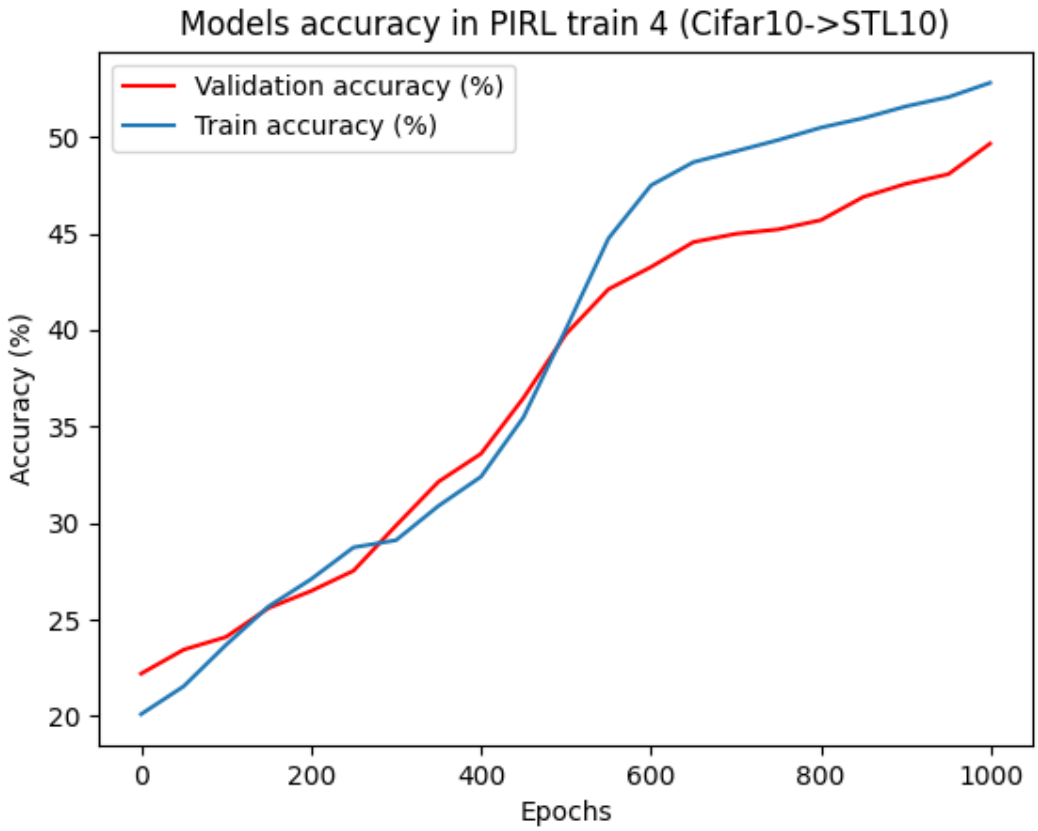
Εικόνα 4.3.3. : Παράδειγμα σωστών και λανθασμένων προβλέψεων για την κατηγορία «Βάτραχος» για το σετ Cifar10.

4.4 Τέταρτο πείραμα

Στο τέταρτο πείραμα πραγματοποιείται η ανάποδη διαδικασία του τρίτου πειράματος. Το καλύτερο μοντέλο απο το pretext task με τις εικόνες του Cifar10 χρησιμοποιείται στο downstream task του STL10.

Στο πείραμα αυτό η ακρίβεια είναι πεσμένη στο 49% . Αυτό πιστεύεται ότι οφείλεται στο γεγονός ότι τα δεδομένα του Cifar10 που χρησιμοποιήθηκαν για την προεκπαίδευση, έχουν περισσότερο θόρυβο και εν τέλη το μοντέλο έχει πρόβλημα στο να γενικευση στο downstream task.

Όπως φαίνεται και στην εικόνα παρακάτω, μετά την εποχή 500, οι δύο ακρίβειες αρχίζουν να αποκλίνουν. Το μοντέλο δεν μπορεί να γενικεύσει τόσο καλά σε εικόνες που δεν έχει ξανασυναντήσει ακόμη και αν έχουν το ίδιο περιεχόμενο.



Εικόνα 4.4.1: Επί της εκατό ακρίβεια με την πάροδο των εποχών για το 4ο πείραμα.

Η σχετικά χαμηλή ακρίβεια φαίνεται και ποιοτικά στην εικόνα 4.4.2.. Πιο αναλυτικά στην κλάση «Πίθηκος» φαίνεται πως στις σωστές προβλέψεις έχουν ταξινομηθεί 3 εικόνες εκ των οποίων μόνο δύο περιέχουν πραγματικά έναν πίθηκο, και στις λάθος προβλέψεις έχουν ταξινομηθεί 2 εικόνες εκ των οποίων μόνο μια περιέχει πίθηκο. Δηλαδή το δίκτυο έχει ταξινομήσει λάθος μια γάτα και την έχει κατατάξει στην κατηγορία του πιθήκου. Αντίστοιχα, η πρώτη εικόνες των λάθος προβλέψεων περιέχει πίθηκο και παρόλα αυτά έχει ταξινομηθεί λάθος. Το υπόβαθρο αυτής της εικόνας και το τρίχωμα του πιθήκου ίσως μπερδεύει το δίκτυο και το κατατάσει σε λανθασμένη κλάση. Συμπερασματικά, απο τις 5 εικόνες το σύστημα κατάφερε να ταξινομήσει σωστά μόνο τις 3. Γι'αυτό αναφέρεται και πως η ακρίβεια της ταξινόμησης είναι σχετικά χαμηλή.

Στην εικόνα 4.4.3 φαίνεται μια πιο σωστή ταξινόμηση της κατηγορίας «Φορτηγό». Ωστόσο και σε αυτή την περίπτωση υπάρχει λάθος ταξινομημένη εικόνα. Η εικόνα με τον αριθμό 4 του εγγύτερου γείτονα ανήκει στην κατηγορία του φορτηγού όμως φαίνεται να έχει ταξινομηθεί σε άλλη κατηγορία. Η χαμηλή ανάλυση της εικόνας και τα έντονα χρώματα του φορτηγό έχουν μπερδέψει το δίκτυο.

Εικόνα 4.4.2: Παράδειγμα σωστών και λανθασμένων προβλέψεων για την κατηγορία «Πίθηκος» του σετ STL10.

Εικόνα 4.4.3: Παράδειγμα σωστών και λανθασμένων προβλέψεων για την κατηγορία «Φορτηγό» του σετ STL10.

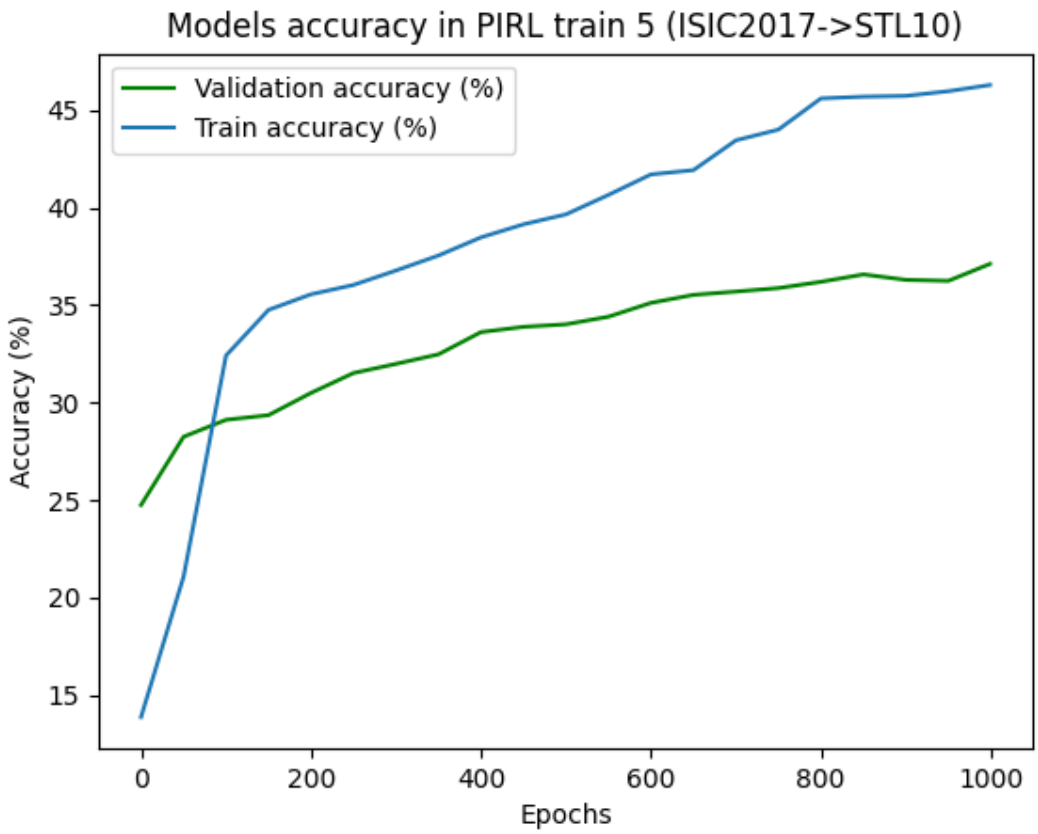
4.5 Πέμπτο πείραμα

Για το πέμπτο πείραμα χρησιμοποιήθηκε το σετ δεδομένων ISIC 2017 που αφορά δερματικά νοσήματα. Πραγματοποιήθηκε προεκπαιδευση σε αυτό το σετ δεδομένων και δοκιμάστηκε η ταξινόμηση εικόνας στο STL10.

Το πείραμα αυτό δεν έχει πρακτική σημασία καθώς είναι 2 πολύ διαφορετικά σετ δεδομένων που χρησιμοποιούνται. Το δίκτυο εκπαιδεύεται εξ αρχής στα δερματικά νοσήματα και δοκιμάζεται εάν μπορεί να αναγνωρίσει εικόνες όπως αυτοκινητο, γάτα, σκύλος και λοιπά. Η ακρίβεια φτάνει το μέγιστο 37%, το οποίο όμως είναι και αναμενόμενο.

Στην εικόνα πιο κάτω παρουσιάζεται και σχηματικά το διάγραμμα με τις 2 ακρίβειες. Οι διαφορά που έχουν η Train και η Validation accuracy φαίνεται να μεγαλώνει καθώς οι εποχές περνούν και έτσι φαίνεται να υπάρχει overfitting. Όμως κάτι τέτοιο είναι αναμενόμενο καθώς τα σετ δεδομένων δεν έχουν κοινά χαρακτηριστικά.

Στο πείραμα αυτό λόγω των πολύ διαφορετικών σετ δεδομένων που χρησιμοποιούνται και λόγω της χαμηλής ακρίβειας, δεν παρουσιάζεται ποιοτική απεικόνιση των αποτελεσμάτων με εικόνες.



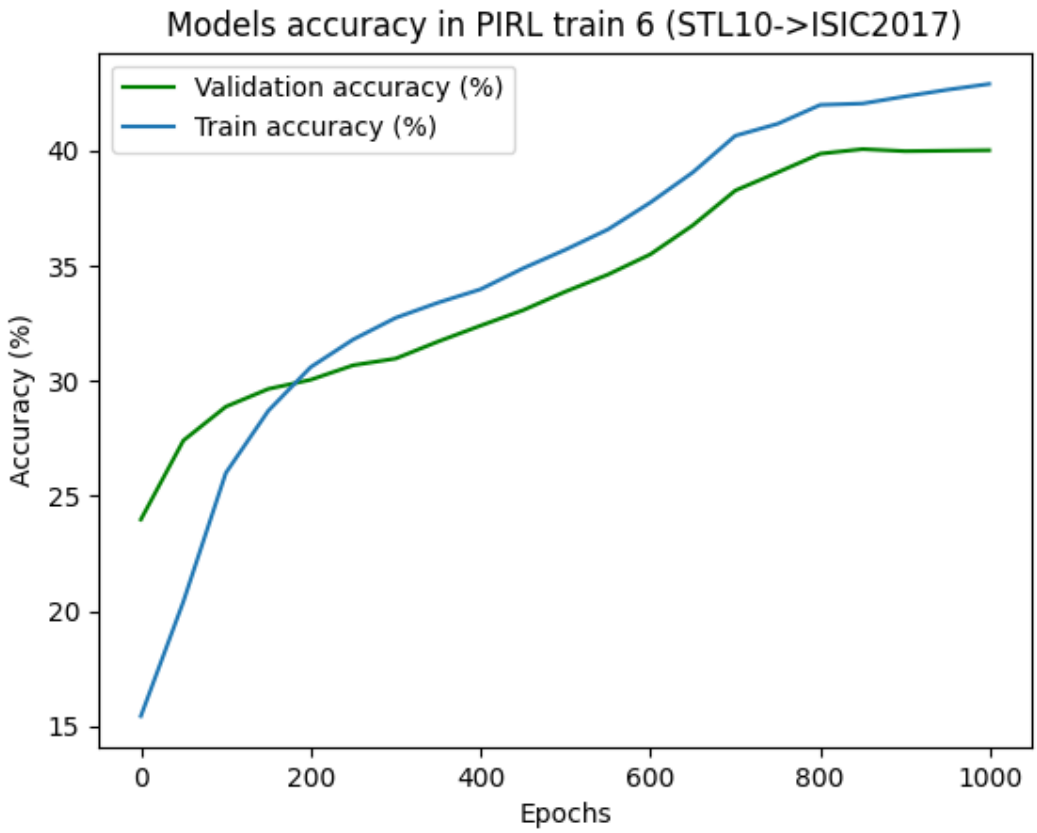
Εικόνα 4.5.1: Επί της εκατό ακρίβεια με την πάροδο των εποχών για το 5ο πείραμα.

4.6 Έκτο πείραμα

Εδώ πραγματοποιείται η ανάποδη διαδικασία. Το Jigsaw puzzle εκπαιδεύει το δίκτυο με τα δεδομένα του STL10 και το καλύτερο μοντέλο της προεκπαίδευσης καλείται να ταξινομήσει εικόνες που περιέχουν δερματικά νοσήματα από το dataset του ISIC 2017.

Και σε αυτήν την περίπτωση η ακρίβεια είναι χαμηλή και φτανει το σχεδόν 37%. Ισχύει ότι ακριβώς και στο παραπάνω πείραμα. Τα δεδομένα που χρησιμοποιούνται στα 2 task είναι πολύ διαφορετικά και η ακρίβεια είναι χαμηλή. Συγκριτικά με το διάγραμμα του πειράματος 5, όπως φαίνεται στην εικόνα οι 2 ακρίβειες φαίνεται να συγκλίνουν περισσότερο και έτσι δεν υπάρχει overfit. Ωστόσο η ακρίβεια είναι αρκετά χαμηλή.

Όπως και στο προηγούμενο πείραμα, έτσι και εδώ, δεν γίνεται ποιοτική αναπαράσταση των αποτελεσμάτων λόγω της χαμηλής ακρίβειας της ταξινόμησης και των διαφορετικών σετ δεδομένων.



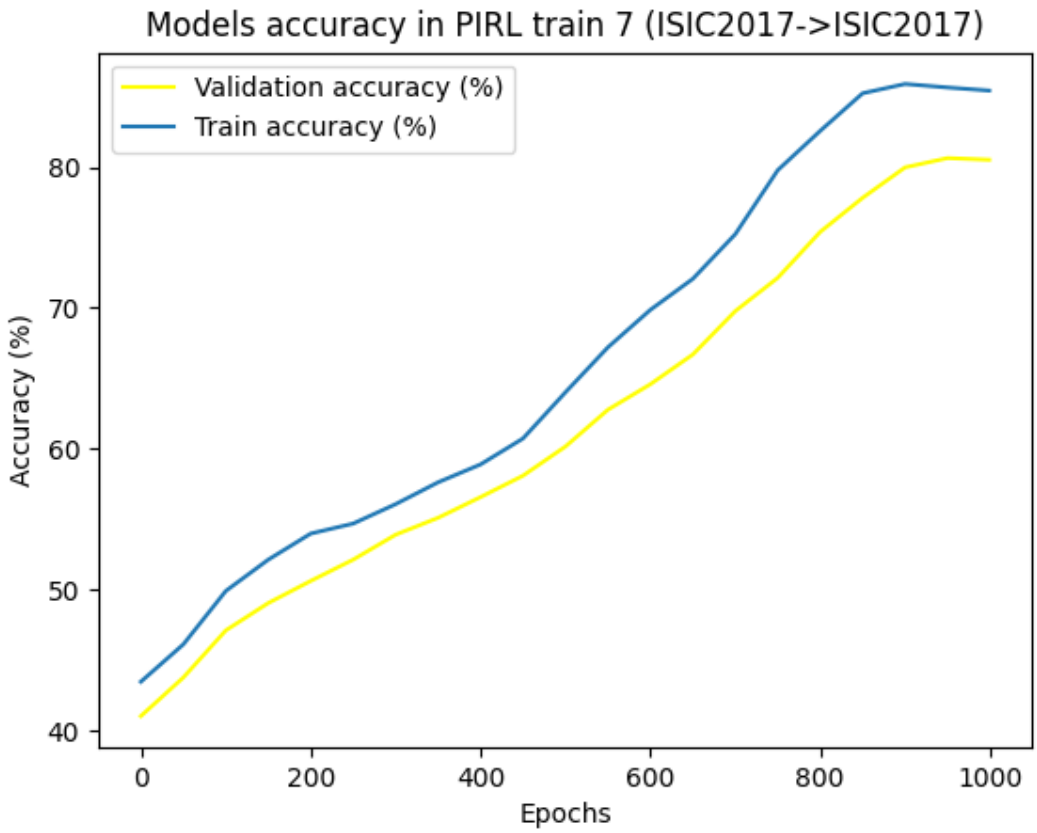
Εικόνα 4.6.1 : Επί της εκατό ακρίβεια με την πάροδο των εποχών για το 6ο πείραμα.

4.7 Έβδομο πείραμα

Στο πείραμα αυτό γίνεται προεκπαίδευση του δικτύου με το σετ δεδομένων ISIC2017. Έπειτα, το ίδιο σετ δεδομένων χρησιμοποιείται και στην εργασία ταξινόμησης εικόνας. Τα αποτελέσματα παρουσιάζονται στον Πίνακα 7 του παραρτήματος και οπτικά στο παρακάτω διάγραμμα.

Το συγκεκριμένο σετ δεδομένων είναι μικρό σε όγκο καθώς περιέχει μόνο 2000 εικόνες που είναι παρεμφερείς μεταξύ τους. Στον διαγωνισμό που χρησιμοποιήθηκαν τα δεδομένα αυτά, κύριος σκοπός ήταν η διάκριση του μελανώματος από άλλες δύο δερματοσκοπικές ασθένειες. Έτσι λόγω της ομοιότητας των δεδομένων, όπως φαίνεται η ακρίβεια της ταξινόμησης φτάνει μέχρι και το 80%.

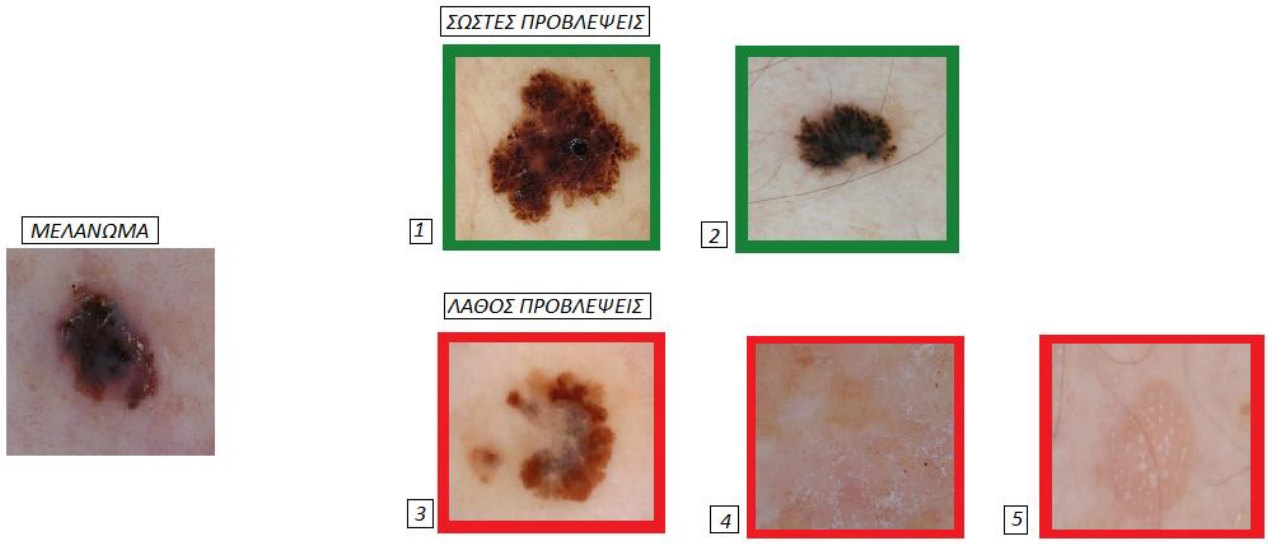
Στην σχηματική απεικόνιση φαίνεται πως οι 2 ακρίβειες είναι κοντά μεταξύ τους και δεν υπάρχει overfit.



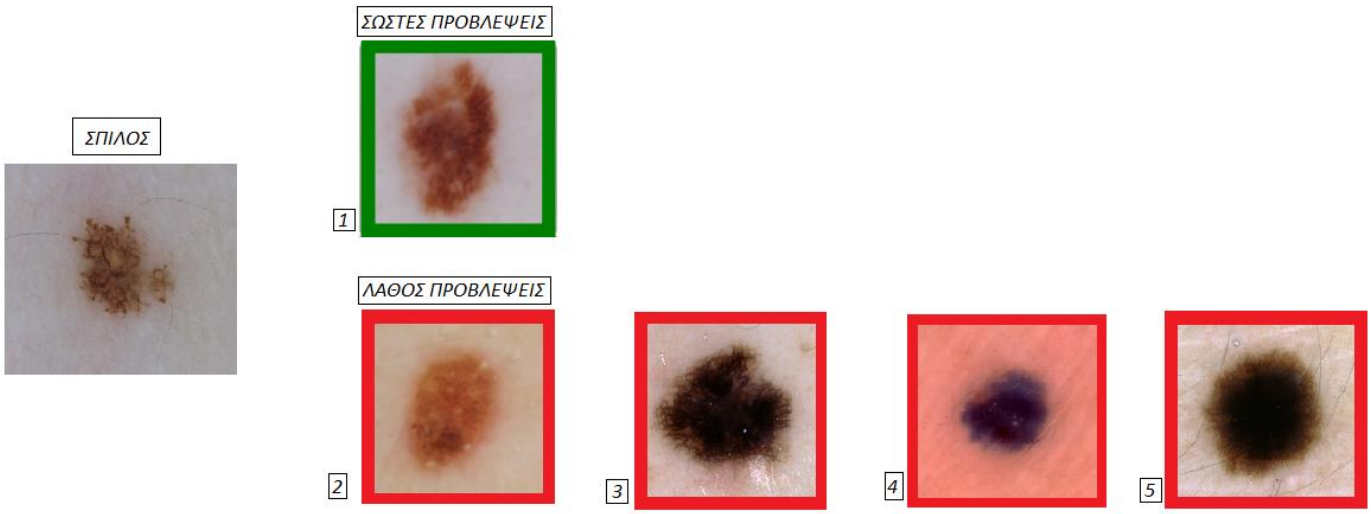
Εικόνα 4.7.1: Επί της εκατό ακρίβεια με την πάροδο των εποχών για το 7ο πείραμα.

Στην παρακάτω εικόνα φαίνεται το παράδειγμα της ταξινόμησης των εικόνων του δέρματος. Προφανώς δεν γίνεται να γνωρίζουμε με ακρίβεια εάν αυτά τα ενδεικτικά αποτελέσματα είναι σωστά καθώς το αντικείμενο των εικόνων πραγματεύεται ένα πολύ εξειδικευμένο θέμα. Ωστόσο, στην εικόνα 4.7.2. φαίνεται πως το δίκτυο έχει ταξινομήσει σωστά όλες τις εικόνες. Θα μπορούσε κανείς να έχει αμφιβολία μόνο για την πρώτη εικόνα των λανθασμένων προβλέψεων με τον αριθμό 3 καθώς η κλάση «Μελάνωμα» αποτελείται κυρίως από κηλίδες με πιο σκούρα χρώματα.

Παρόλα αυτά το δίκτυο έχει εξαιρετικά υψηλή ακρίβεια και αυτό σημαίνει πως μπορεί εύκολα να διακρίνει τις 3 κατηγορίες νοσημάτων. Κάποιο πιο εξειδικευμένο προσωπικό θα μπορούσε να καταλάβει εάν τα αποτελέσματα της ταξινόμησης είναι πράγματι τόσο καλά όσο αναμένονται να είναι. Στην προκειμένη περίπτωση, θεωρείται πως το δίκτυο αναγνωρίζει και ταξινομεί σωστά τις εικόνες αυτές.



Εικόνα 4.7.2: Παράδειγμα σωστών και λανθασμένων προβλέψεων για την κατηγορία «Μελάνωμα» του σετ ISIC 2017.



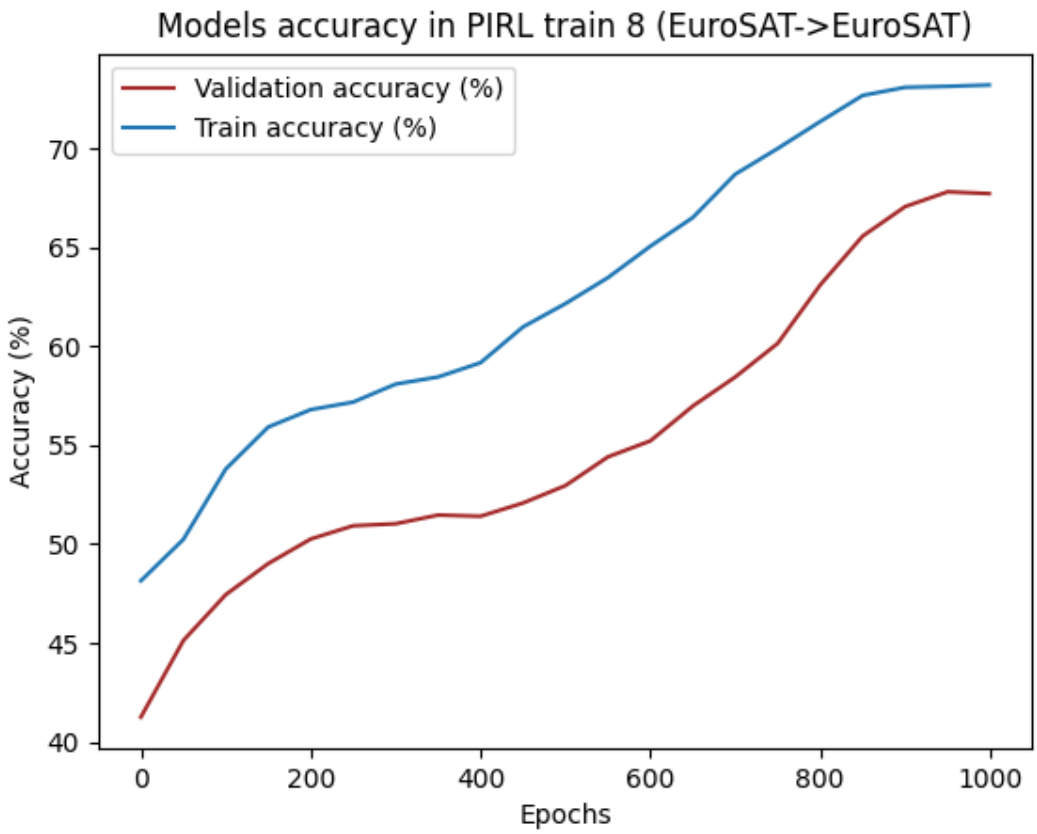
Εικόνα 4.7.3: Παράδειγμα σωστών και λανθασμένων προβλέψεων για την κατηγορία «Μελάνωμα» του σετ ISIC 2017.

Αντιστοιχα και στην παραπάνω εικόνα παρουσιάζεται το παράδειγμα για την κατηγορία «Σπίλος». Αυτή τη φορά το δίκτυο φαίνεται να έχει κατατάξει στις σωστές προβλέψεις μία εικόνα και τέσσερις εικόνες στις λάθος προβλέψεις. Στις 3 τελευταίες εικόνες των λάθος προβλέψεων η ασθένεια μοιάζει να είναι μελάνωμα λόγω του σκούρου χρώματος των κηλίδων. Ωστόσο και σε αυτήν την περίπτωση μόνο εξειδικευμένο προσωπικό μπορεί να κάνει σωστή αναγνώριση και ταξινόμηση των νοσημάτων αυτών με οπτική παρατήρηση.

4.8 Όγδοο πείραμα

Τέλος για το πείραμα αυτό χρησιμοποιήθηκε το δορυφορικό σετ δεδομένων της EuroSAT. Όπως και στα προηγούμενα πειράματα, έτσι και σε αυτό, πραγματοποιήθηκε προ-εκπαίδευση του δικτύου με το συγκεκριμένο σετ δεδομένων και έπειτα παγώνοντας όλα τα προηγούμενα επίπεδα, εκπαιδεύεται ένας classifier και υλοποιείται εργασια ταξινόμησης εικόνας στο ίδιο σετ δεδομένων. Τα αποτελέσματα παρουσιάζονται οπτικά παρακάτω, και στον Πίνακα 8 του παραρτήματος.

Η ακρίβεια σε αυτό το τελευταίο πείραμα φτάνει σχεδόν στο 68 % και όπως φαίνεται από το παρακάτω σχηματικό διάγραμμα, δεν υπάρχει overfit στο μοντέλο. Αυτό σημαίνει πως το μοντέλο μπορεί να γενικεύσει και να αναγνωρίσει με σχετικά καλή ακρίβεια εικόνες που δεν έχει ξανασυναντήσει.



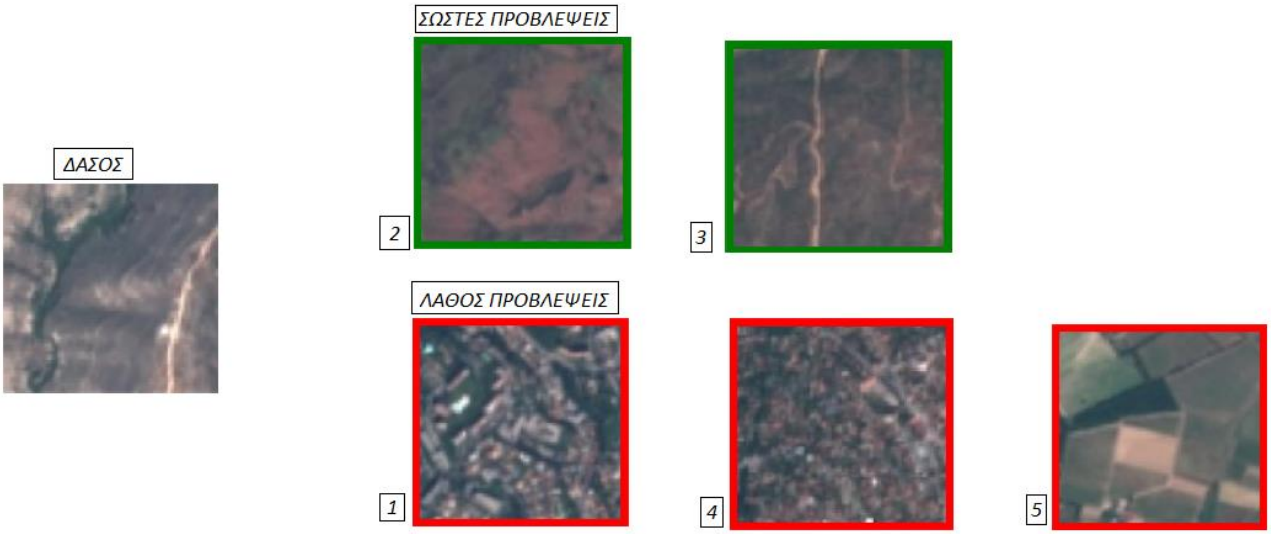
Εικόνα 4.8.1: Επί της εκατό ακρίβεια με την πάροδο των εποχών για το 8ο πείραμα.

Η ακρίβεια αυτής της ταξινόμησης θεωρείται πως είναι καλή καθώς φτάνει σχεδόν στο 70%. Για να μπορέσει κανείς να αντιληφθεί την σημασία του αριθμού αυτού, παρουσιάζονται παρακάτω μερικά αποτελέσματα εικόνων προκειμένου ποιοτικά να δειχθεί η ακρίβεια της ταξινόμησης.

Στην εικόνα 4.8.2 παρουσιάζεται ενδεικτικά η κλάση «Αυτοκινητόδρομος». Στις σωστές προβλέψεις φαίνεται πως υπάρχει οδικός άξονας φανερά και στις 2 εικόνες και είναι το στοιχείο που κυριαρχεί. Στις λανθασμένες προβλέψεις, οι 2 πρώτες εικόνες φαίνεται να έχουν κάποιο γραμμικό πρότυπο ή γεωμετρικούς σχηματισμούς και αυτό το αναγνωρίζει το δίκτυο. Ωστόσο στην τελευταία εικόνα η λήψη φαίνεται να έχει έναν μεγάλο οδικό άξονα εντός πόλης. Το δίκτυο μπερδεύεται και πιθανόν κατατάσει την εικόνα στην κατηγορία με την βιομηχανική-αστική περιοχή.

Εικόνα 4.8.2: Παράδειγμα σωστών και λανθασμένων προβλέψεων για την κατηγορία «Αυτοκινητόδρομος» του σετ EuroSAT.

Αντίστοιχα στο δεύτερο παράδειγμα που παρουσιάζεται ποιοτικά στην εικόνα 4.8.3. φαίνεται μια ενδεικτική ταξινόμηση 5 δειγμάτων της κλάσης «Δάσος». Αυτή τη φορά φαίνεται το δίκτυο να έχει ταξινομήσει σωστά όλα τα δείγματα. Στις σωστές προβλέψεις μπορεί και αναγνωρίζει πως πρόκειται για δασική έκταση παρόλο που τα μέρη δεν περιέχουν πυκνό δάσος και υπάρχουν και στοιχεία από επαρχιακό δρόμο. Αντίστοιχα οι λανθασμένες προβλέψεις ανήκουν πράγματι σε διαφορετικές κλάσεις. Οι πρώτες δύο εικόνες κατατάσσονται στην κατηγορία βιομηχανική-αστική περιοχή, καθώς περιέχουν συγκεκριμένα πρότυπα που ορίζονται από την δόμηση της περιοχής και η τρίτη και τελευταία εικόνα φαίνεται να έχει και αυτή κάποιο γεωμετρικό σχηματισμό στα χωράφια που φανερώνει πως πρόκειται για καλλιεργήσιμη έκταση ή βοσκότοπο.

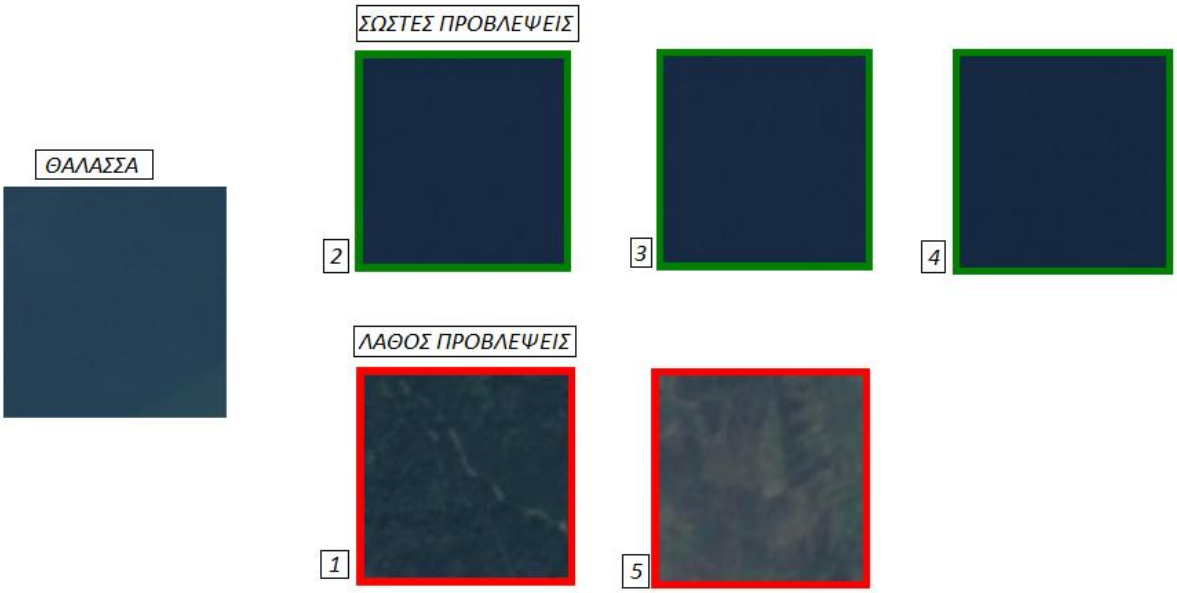


Εικόνα 4.8.3: Παράδειγμα σωστών και λανθασμένων προβλέψεων για την κατηγορία «Δάσος» του σετ EuroSAT.

Παρουσιάζεται ακόμη παράδειγμα με την κλαση «Αστική περιοχή» και την κλάση «Θάλασσα-Λίμνη». Στην περίπτωση της αστικής περιοχής το δίκτυο έχει προβλέψει μία σωστή εικόνα και τέσσερις λάθος. Όπως φαίνεται, η εικόνα που ανήκει στις σωστές προβλέψεις έχει παρόμοιο πρότυπο και δόμηση με την εικόνα να αναφοράς. Ακόμη, η τελευταία εικόνα με τον αριθμό 5 του εγγύτερου γείτονα φαίνεται να μοιάζει με την αρχική εικόνα. Παρόλα αυτά το δίκτυο την έχει κατατάξει σε άλλη κλάση. Το πιθανότερο είναι η εικόνα να έχει καταταχθεί στην κατηγορία «Ποώδης Βλάστηση». Οι υπόλοιπες εικόνες ωστόσο, φαίνεται να είναι σωστα ταξινομημένες.

Η κατηγορία «Θάλασσα-Λίμνη» φαίνεται να είναι απλό και εύκολο παράδειγμα ταξινόμησης εικόνας. Οι εικόνες που περιέχουν θάλασσα ή λίμνη έχουν διάφορους τόνους του σκούρου μπλε χρώματος και συνήθως δεν περιέχουν άλλα αντικείμενα. Η μόνη σύγχυση που πιθανόν να υπήρχε θα ήταν με κάποια εικόνα ενός μεγάλου ποταμιού. Όμως σε αυτή την περίπτωση το δίκτυο αναγνωρίζει πολύ σωστά όλες τις εικόνες.

Εικόνα 4.8.3: Παράδειγμα σωστών και λανθασμένων προβλέψεων για την κατηγορία «Αστική Περιοχή» του σετ EuroSAT.



Εικόνα 4.8.3: Παράδειγμα σωστών και λανθασμένων προβλέψεων για την κατηγορία «Θάλασσα-Λίμνη» του σετ EuroSAT.

Έτσι συμπερασματικά, για αυτά τα δορυφορικά σετ δεδομένων το δίκτυο είναι σε θέση να μαθαίνει χαρακτηριστικά των εικόνων όπως για παράδειγμα τα διάφορα γεωμετρικά πρότυπα ή χρώματα που μπορεί να υπάρχουν σε κάθε εικόνα και στην συνέχεια να αναγνωρίσει εικόνες που δεν έχει συναντήσει ξανα. Το συγκεκριμένο σετ της EuroSAT έχει 10 κλάσεις οι οποίες μπορεί εύκολα να μπερδευτούν μεταξύ τους καθώς έχουν παρόμοια χαρακτηριστικά. Ακόμη και ένας άνθρωπος θα μπορούσε να κατατάξει σε λάθος κατηγορία κάποιες από τις εικόνες. Έτσι το δίκτυο αυτό θεωρείται πως δίνει αξιόλογα αποτελέσματα σε ότι αφορά την ακρίβεια της ταξινόμησης.

4.9 Συγκεντρωτικά Αποτελέσματα

Με βάση τα αποτελέσματα των 8 πειραμάτων λοιπόν, κατασκευάστηκε ο παρακάτω συγκεντρωτικός πίνακας που αναφέρει τις τελικές επί τις εκατό Validation accuracies. Δεν είναι εφικτή η σύγκριση με τα αποτελέσματα του αρχικού πειράματος PIRL καθώς εκεί χρησιμοποιήθηκαν διαφορετικά και μεγάλα σε όγκο σετ δεδομένων που δεν μπορούσαν να χρησιμοποιηθούν στην παρούσα εργασία λόγω έλλειψης πόρων.

Πρέπει ακόμα να τονιστεί πως όλα τα πειράματα ενδεχομένως να έδιναν κάποιες καλύτερες ακρίβειες αν πραγματοποιούνταν για περισσότερες από 1000 εποχές και αν γινόταν εύρεση του σωστού συνόλου υπερ-παραμέτρων. Ωστόσο, στα περισσότερα διαγράμματα φαίνεται πως όσο φτάνουν προς το τέλος οι εποχές, η ακρίβεια αρχίζει και σταθεροποιείται.

Models accuracy (%) with different datasets		**Downstream task**			
		STL10	CIFAR10	ISIC2017	EuroSAT
Pretext task	STL10	**68.79**	**51.19**	**40.07**	
	CIFAR10	**49.65**	**59.03**		
	ISIC2017	**37.12**		**80.52**	
	EuroSAT				**67.81**

Πίνακας 4.9 : Βέλτιστη επί τις εκατό ακρίβεια για κάθε συνδυασμό δεδομένων.

5. ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΠΡΟΟΠΤΙΚΕΣ

5.1 Γενικά Συμπεράσματα

Σε γενικό πλαίσιο, η ιδέα των πειραμάτων της εργασίας PIRL είναι ένα πολύ χρήσιμο εργαλείο που βελτιώνει την σημασιολογική ποιότητα των αναπαραστάσεων των εικόνων. Το σκεπτικό πίσω από το PIRL είναι η αμεταβλητότητα στους μετασχηματισμούς εικόνας διατηρώντας τις σημασιολογικές πληροφορίες στην αναπαράσταση. Ουσιαστικά εκμεταλλεύεται και αναδεικνύει τις ιδιότητες της αυτο-επιβλεπόμενης μάθησης. Στην συγκεκριμένη διπλωματική εργασία με βάση τα πειράματα που πραγματοποιήθηκαν αποδεικνύεται πως θα μπορούσε να χρησιμοποιηθεί οποιοδήποτε pretext task για την προ-εκπαίδευση του μοντέλου καθώς η εργασία επικεντρώνεται στην εργασία ταξινόμησης εικόνας σχετικά με τα διαφορετικά σετ δεδομένων.

Πιο συγκεκριμένα, με βάση τα πειράματα που διεξήχθησαν συμπεραίνεται ότι επιτυχία του αλγορίθμου εξαρτάται από την ποιότητα των διατιθέμενων δεδομένων. Η ανάλυση των εικόνων αλλά και το περιεχόμενό τους, επηρεάζουν σημαντικά την ακρίβεια του δικτύου. Ωστόσο, και οι εποχές που πραγματοποιούνται τα πειράματα άλλα και η αναζήτηση και εφαρμογή των κατάλληλων υπερ παραμέτρων παίζουν σημαντικό ρόλο στην τελική ακρίβεια της εργασίας ταξινόμησης.

Το σετ δεδομένων του STL10 έδωσε καλά αποτελέσματα στο πρώτο πείραμα που χρησιμοποιήθηκε και για την προ-εκπαίδευση και για την εργασία ταξινόμησης (Pretext & Downstream Task). Η καλή ανάλυση των εικόνων του σετ αλλά και ο αριθμός των δεδομένων εκπαίδευσης (100000 εικόνες) είναι ένας σημαντικός παράγοντας για την καλή ακρίβεια. Εντούτοις, το περιεχόμενο των εικόνων σε πολλές περιπτώσεις μπέρδεψε το δίκτυο και κυρίως στις περιπτώσει που το μοντέλο είχε προ-εκπαιδευτεί σε άλλο σετ δεδομένων.

Το Cifar10 είναι ένα πολυχρησιμοποιημένο και βασικό σετ δεδομένων για εργασίες Μηχανικής Μάθησης. Παρόλα αυτά, στα πειράματα της διπλωματικής εργασίας αυτής η χαμηλή ανάλυση των εικόνων δεν λειτούργησε θετικά. Πρακτικοί λόγοι οδήγησαν στην μεγέθυνση των δεδομένων και επομένως προσθήκη θορύβου και εν τέλη μείωση της ακρίβειας. Στο δεύτερο πείραμα επιτυγχάνεται και η υψηλότερη ακρίβεια σε πείραμα που συμμετέχει το σετ Cifar10.

Το σετ δεδομένων του ISIC 2017, όπως ήδη έχει προαναφερθεί πρόκειται για ένα μικρό σετ δεδομένων αποτελούμενο από μόνο τρεις κατηγορίες εικόνων με δερματοσκοπικές παθήσεις. Το συγκεκριμένο σετ αποδίδει την καλύτερη ακρίβεια όλων των πειραμάτων στο εβδομο πείραμα όπου το μοντέλο προεκπαιδεύεται και δοκιμάζεται στο ίδιο σετ δεδομένων του διαγωνισμού ISIC 2017. Το σετ περιέχει εικόνες που έχουν παρθεί από μικροσκόπιο και έτσι η ανάλυση των εικόνων είναι μεγάλη. Οι κλάσεις των εικόνων αυτών είναι τρεις και ως αποτέλεσμα υπάρχει μεγαλύτερη πιθανότητα μια εικονα να καταταχθεί σωστα.

Το τελευταίο σετ δεδομένων που χρησιμοποιήθηκε είναι του EuroSAT και είναι ένα δορυφορικό σετ που περιέχει 27.000 εικόνες σχετικά καλής ανάλυσης. Η δυσκολία του συγκεκριμένου σετ δεδομένων ήταν η ομοιότητα εικόνων από διαφορετικές κλάσεις. Το δίκτυο δυσκολεύτηκε στον

διαχωρισμό εικόνων ή τμήματα εικόνων που περιέχουν παρόμοια μοτίβα και έτσι κατα κάποιο τρόπο μειωνόταν η ακρίβεια της ταξινόμησης.

Τέλος, από τα παραπάνω πειράματα είναι προφανές πως σετ δεδομένων τα οποία παρουσιάζουν οπτική ομοιότητα (π.χ. STL10 και Cifar10) παρουσιάζουν και αντίστοιχα καλύτερα αποτελέσματα στην εργασία ταξινόμησης. Αντίστοιχα, σετ δεδομένων που έχουν συγκεκριμένο περιεχόμενο (π.χ. ISIC2017), όταν συνδιάζονται με άλλα σετ δεδομένων δεν δίνουν καλα αποτελέσματα στο downstream task. Συνοψίζοντας λοιπόν, σύμφωνα με τα πειράματα που πραγματοποιήθηκαν, μπορεί κανείς να συμπεράνει πως η εργασία ταξινόμησης εικόνας χρησιμοποιώντας Self-Supervised Learning απαιτεί ένα σχετικά μεγάλο σετ δεδομένων, καλής ανάλυσης, με καθαρές εικόνες. Φυσικά η αρχιτεκτονική του μοντέλου και η εύρεση των κατάλληλων υπερ-παραμέτρων παίζουν πολύ σημαντικό ρόλο στην τελική ακρίβεια της ταξινόμησης.

5.2 Προοπτικές

Τα νευρωνικά δίκτυα τα τελευταία χρόνια κάνουν μεγάλα άλματα και είναι πλέον εδραιωμένα στην επιστήμη της όρασης υπολογιστών και όχι μόνο. Έγινε φανερό από την παρούσα εργασία η πολυπλοκότητα της αρχιτεκτονικής τους και ο μεγάλος αριθμός των εκπαιδευόμενων παραμέτρων. Οι προοπτικές που παρατηρούνται στον τομέα πλέον απευθύνονται στην εύρεση νέων μεθοδολογιών που θα απαιτούν λιγότερη υπολογιστική ισχύ και θα έχουν μεγαλύτερη αποτελεσματικότητα. Υπάρχει πληθώρα δεδομένων στην σύγχρονη εποχή, όμως απαιτούνται ισχυρά υπολογιστικά μέσα που δεν διαθέτουν όλοι. Οι αρχιτεκτονικές των δικτύων έχουν μεγαλώσει πολύ καθώς χρησιμοποιούνται νέες τεχνικές αντί των CNN που λέγοναι Transformers. Ακόμη τα pretext task εξελίσονται συνεχώς και γίνονται όλο και καλύτερα. Σε ότι αφορά το Self-Supervised Learning, η επιστημονική κοινότητα χρησιμοποιεί πολύ το σετ δεδομένων του ImageNet για προ-εκπαίδευση, το οποίο όπως αναφέρθηκε είναι ένα πολύ μεγάλο σετ δεδομένων με 1000 κλάσεις που έχει μεγάλη οπτική ποικιλία. Στο downstream task χρησιμοποιούνται πιο μικρά σετ δεδομένων.

Η συγκεκριμένη εργασία χρησιμοποίησε βασικά σετ δεδομένων Μηχανικής Μάθησης και ένα δορυφορικό σετ εικόνων RGB. Μια πρόταση είναι η πραγματοποίηση νέων πειραμάτων σε πιο εξειδικευμένα δεδομένα. Η εξέλιξη των δορυφορικών μέσων είναι τέτοια ώστε πλέον είναι άμεσα διαθέσιμα δεδομένα ακόμα και video από δορυφορικά μέσα. Έχουν γίνει ήδη προσπάθειες από την διεθνή επιστημονική κοινότητα πάνω στα υπερφασματικά δεδομένα, όμως το διαθέσιμο υλικό είναι μεγάλο, με αποτέλεσμα να γεννάται η ανάγκη διαχείρισης των δεδομένων αυτών με αντίστοιχο τρόπο.

Ακόμη σχετικά με την παρούσα εργασία, ενδιαφέρον πείραμα θα ήταν και η δοκιμή διαφόρων pretext task σε ένα σετ δεδομένων και εφαρμογή του ίδιου σετ σε downstream task ταξινόμησης προκειμένου να διαπιστωθεί η διαφορά που προκύπτει στην τελική ακρίβεια.

Τέλος, στα ίδια ή και σε άλλα πειράματα θα ήταν ενδιαφέρον να αναπτυχθεί αλγόριθμος για να παρατηρηθεί το που εστιάζει το νευρωνικό κατα την ταξινόμηση των εικόνων. Το πεδίο αυτό έχει ήδη αναπτυχθεί και ονομάζεται Attention in Machine Learning και η σκέψη πίσω από αυτό είναι ότι το δίκτυο θα πρέπει να εστιάσει περισσότερο σε κάποιο μικρό αλλά σημαντικό μέρος των δεδομένων.

ΠΑΡΑΡΤΗΜΑ

Στο παράρτημα αυτό αναγράφονται τα αποτελέσματα των πειραμάτων που πραγματοποιήθηκαν στην παρούσα εργασία. Παρουσιάζονται οι επι τοις εκατό (%) ακρίβειες Train και Validation.

PIRL train 1 (STL10->STL10)		
EPOCH	TRAIN ACCURACY (%)	VALIDATION ACCURACY (%)
0	22,15	24,38
50	24,95	27,62
100	27,21	30,12
150	35,60	37,50
200	39,45	40,25
250	45,00	42,88
300	48,91	46,61
350	54,60	50,02
400	59,80	53,88
450	62,78	57,70
500	66,21	60,62
550	68,15	62,02
600	71,03	63,95
650	73,22	64,96
700	76,84	65,99
750	77,01	66,75
800	77,86	67,00
850	78,02	67,80
900	78,55	68,25
950	79,11	68,75
999	79,47	**68,79**

Πίνακας 1: Αποτελέσματα ακρίβειας 1ου πειράματος.

PIRL train 2 (Cifar10->Cifar10)		
EPOCH	TRAIN ACCURACY (%)	VALIDATION ACCURACY (%)
0	18,09	19,05
50	18,88	19,79
100	20,04	21,06
150	21,58	22,99
200	23,98	24,35
250	26,78	26,86
300	30,15	29,79
350	35,47	32,11
400	39,80	35,41
450	45,60	37,80
500	48,55	39,02
550	50,99	40,81
600	53,81	43,01
650	56,10	45,34
700	58,33	47,30
750	61,01	49,96
800	62,94	51,72
850	65,51	54,39
900	67,01	56,46
950	68,45	57,98
999	69,70	**59,03**

Πίνακας 2: Αποτελέσματα ακρίβειας 2ου πειράματος.

PIRL train 3 (STL10 ->Cifar10)		
EPOCH	TRAIN ACCURACY (%)	VALIDATION ACCURACY (%)
0	15,06	17,40
50	16,68	20,04
100	18,54	22,01
150	23,78	25,55
200	26,45	27,02
250	29,01	28,65
300	31,56	30,68
350	34,65	33,05
400	38,65	36,32
450	41,10	39,11
500	43,00	41,09
550	45,21	43,33
600	47,00	45,40
650	47,89	46,10
700	48,40	47,48
750	49,09	47,98
800	50,45	48,70
850	50,99	49,57
900	51,46	50,21
950	51,96	50,50
999	52,27	**51,19**

Πίνακας 3: Αποτελέσματα ακρίβειας 3ου πειράματος.

PIRL train 4 (Cifar10 ->STL10)		
EPOCH	TRAIN ACCURACY (%)	VALIDATION ACCURACY (%)
0	20,10	22,20
50	21,54	23,44
100	23,69	24,10
150	25,68	25,60
200	27,09	26,48
250	27,74	27,53
300	29,11	29,87
350	30,89	32,14
400	32,40	33,59
450	35,48	36,48
500	40,05	39,79
550	44,74	42,11
600	47,49	43,26
650	48,69	44,55
700	49,26	44,98
750	49,84	45,21
800	50,48	45,69
850	50,97	46,89
900	51,58	47,57
950	52,06	48,07
999	52,80	**49,65**

Πίνακας 4: Αποτελέσματα ακρίβειας 4ου πειράματος.

PIRL train 5 (ISIC2017->STL10)		
EPOCH	TRAIN ACCURACY (%)	VALIDATION ACCURACY (%)
0	13,87	24,75
50	21,07	28,25
100	32,42	29,12
150	34,76	29,36
200	35,56	30,50
250	36,04	31,52
300	36,78	31,99
350	37,55	32,48
400	38,48	33,62

PIRL train 6 (STL10->ISIC2017)		
EPOCH	TRAIN ACCURACY (%)	VALIDATION ACCURACY (%)
0	15,44	23,98
50	20,41	27,41
100	26,00	28,88
150	28,70	29,65
200	30,60	30,05
250	31,80	30,68
300	32,74	30,97
350	33,40	31,71
400	33,98	32,40

450	39,15	33,89
500	39,66	34,02
550	40,66	34,41
600	41,71	35,12
650	41,93	35,53
700	43,46	35,70
750	44,01	35,88
800	45,61	36,20
850	45,70	36,58
900	45,74	36,30
950	45,98	36,25
999	46,30	**37,12**

Πίνακας 5: Αποτελέσματα ακρίβειας 5ου πειράματος.

450	34,89	33,07
500	35,70	33,88
550	36,58	34,62
600	37,75	35,20
650	39,06	36,76
700	40,65	38,27
750	41,17	39,06
800	41,99	39,87
850	42,05	**40,07**
900	42,36	39,98
950	42,65	40,00
999	42,90	40,02

Πίνακας 6 : Αποτελέσματα ακρίβειας 6ου πειράματος

PIRL train 7 (ISIC2017->ISIC2017)		
EPOCH	**TRAIN ACCURACY (%)**	**VALIDATION ACCURACY (%)**
0	43,48	41,04
50	46,13	43,78
100	49,92	47,12
150	52,14	49,06
200	53,99	50,63
250	54,70	52,14
300	56,08	53,91
350	57,63	55,11
400	58,90	56,58
450	60,74	58,10
500	64,02	60,17
550	67,21	62,78
600	69,87	64,59
650	72,07	66,70
700	75,24	69,78
750	79,78	72,14
800	82,57	75,40
850	85,24	77,82
900	85,90	79,98
950	85,65	80,64
999	85,43	**80,52**

Πίνακας 7: Αποτελέσματα ακρίβειας 7ου πειράματος.

PIRL train 8 (EuroSAT->EuroSAT)		
EPOCH	**TRAIN ACCURACY (%)**	**VALIDATION ACCURACY (%)**
0	48,15	41,26
50	50,24	45,13
100	53,8	47,45
150	55,92	49,02
200	56,8	50,26
250	57,18	50,93
300	58,09	51,03
350	58,45	51,47
400	59,17	51,41
450	60,98	52,08
500	62,16	52,96
550	63,47	54,41
600	65,06	55,22
650	66,51	56,98
700	68,7	58,45
750	70	60,15
800	71,36	63,1
850	72,68	65,57
900	73,09	67,06
950	73,14	**67,81**
999	73,21	67,72

Πίνακας 8: Αποτελέσματα ακρίβειας 8ου πειράματος.

ΒΙΒΛΙΟΓΡΑΦΙΑ

Ξενόγλωσση βιβλιογραφία

[1] Going deeper with convolutions, 2014. (1409.4842v1.pdf (arxiv.org))

[2] Deep Residual Learning for Image Recognition, 2015. (1512.03385.pdf (arxiv.org))

[3] SUPERVISED REPRESENTATION LEARNING BY PREDICTING IMAGE ROTATIONS,2018. (https://arxiv.org/pdf/1803.07728.pdf)

[4] Unsupervised Learning of Visual Representations by Solving Jigsaw Puzzles, 2017. (https://arxiv.org/pdf/1603.09246.pdf)

[5] Context Encoders: Feature Learning by Inpainting, 2016. (https://arxiv.org/pdf/1604.07379.pdf)

[6] Colorful Image Colorization . 2016 (https://arxiv.org/pdf/1603.08511.pdf)

[7] Learning Representations for Automatic Colorization, 2017. (https://arxiv.org/pdf/1603.06668.pdf)

[8] Unsupervised Visual Representation Learning by Context Prediction, 2016. (https://arxiv.org/pdf/1505.05192.pdf)

[9] Self-Supervised Learning of Remote Sensing Scene Representations Using Contrastive Multiview Coding, 2021. (2104.07070.pdf (arxiv.org))

[10] Self-supervised Visual Feature Learning with Deep Neural Networks: A Survey, 2019.(1902.06162.pdf (arxiv.org))

[11] Knowledge Distillation: A Survey, 2021.(https://arxiv.org/pdf/2006.05525.pdf)

[12] Remote Sensing Image Scene Classification with Self-Supervised Paradigm under Limited Labeled Samples, 2020.(2010.00882.pdf (arxiv.org))

[13] Self-Supervised Learning With Adaptive Distillation for Hyperspectral Image Classification, 2020.(Self-Supervised Learning With Adaptive Distillation for Hyperspectral Image Classification | IEEE Journals & Magazine | IEEE Xplore)

[14] Self-Supervised Learning of Remote Sensing Scene Representations Using Contrastive Multiview Coding, 2021. (2104.07070.pdf (arxiv.org))

[15] The color out of space: learning self-supervised representations for Earth Observation imagery, 2020.(2006.12119.pdf (arxiv.org))

[16] Self-Supervised Learning of Pretext-Invariant Representations, 2019. (https://arxiv.org/pdf/1912.01991.pdf)

[17] Scaling and Benchmarking Self-Supervised Visual Representation Learning, 2019.(https://arxiv.org/pdf/1905.01235.pdf)

[18] EuroSAT: A Novel Dataset and Deep Learning Benchmark for Land Use and Land Cover Classification (https://arxiv.org/pdf/1709.00029.pdf)

Ιστοσελίδες

[19] https://dltlabs.medium.com/understanding-machine-learning-deep-learning-f5aa95264d61

[20] https://www.kdnuggets.com/2019/11/deep-learning-image-classification-less-data.html

[21] https://www.researchgate.net/figure/General-Architecure-of-the-Fully-Connected-NN_fig7_325022519

[22] https://cezannec.github.io/Convolutional_Neural_Networks/

[23] https://neurohive.io/en/popular-networks/alexnet-imagenet-classification-with-deep-convolutional-neural-networks/

[24] https://www.researchgate.net/figure/The-illustration-of-the-GoogleNet-architecture-Szegedy-et-al-2015-All-convolutional_fig1_314119821

[25] https://towardsdatascience.com/deep-learning-googlenet-explained-de8861c82765

[26] https://www.baeldung.com/cs/ml-labeled-vs-unlabeled-data

[27] https://labelyourdata.com/articles/unlabeled-data-in-machine-learning

[28] https://www.applause.com/blog/training-data-validation-data-vs-test-data

[29] https://towardsdatascience.com/train-validation-and-test-sets-72cb40cba9e7

[30] https://www.ibm.com/cloud/learn/supervised-learning

[31] https://www.ibm.com/cloud/learn/unsupervised-learning

[32] https://machinelearningmastery.com/types-of-learning-in-machine-learning/

[33] https://business.blogthinkbig.com/semi-supervised-learning-the-great-unknown/

[34] https://atcold.github.io/pytorch-Deep-Learning/en/week10/10-1/

[35] https://deeplearninguniversity.com/pretext-training-tasks-for-self-supervised-learning-in-computer-vision/

[36] https://www.cs.toronto.edu/~kriz/cifar.html

[37] https://en.wikipedia.org/wiki/ImageNet

[38] https://www.researchgate.net/figure/Examples-in-the-ImageNet-dataset_fig7_314646236

[39] https://ai.stanford.edu/~acoates/stl10/

[40] https://en.wikipedia.org/wiki/File:HyperspectralCube.jpg

[41] https://challenge.isic-archive.com/data/

[42] https://spr.com/can-deep-transfer-learning-help-detect-covid-19/

[43] https://towardsdatascience.com/a-framework-for-contrastive-self-supervised-learning-and-designing-a-new-approach-3caab5d29619

[44] https://github.com/phelber/eurosat

[45] https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine-learning