



ΔΗΜΟΚΡΙΤΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΡΑΚΗΣ
ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ
ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΗΛΕΚΤΡΟΝΙΚΗΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ
ΣΥΣΤΗΜΑΤΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ

Αξιολόγηση της γενίκευσης των Συνελικτικών Νευρωνικών
Δικτύων που έχουν εκπαιδευθεί με έμμεσες διεργασίες

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Λάζαρος Κανδυλιώτης, ΑΕΜ: 58172

Επιβλέπων Καθηγητής: Ιωάννης, Μπούταλης, Καθηγητής,
Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών, Δ.Π.Θ.

Ξάνθη, 2024



**ΔΗΜΟΚΡΙΤΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΡΑΚΗΣ
ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ**

**ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ
ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ**

**ΤΟΜΕΑΣ ΗΛΕΚΤΡΟΝΙΚΗΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ
ΣΥΣΤΗΜΑΤΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ**

**Αξιολόγηση της γενίκευσης των Συνελικτικών Νευρωνικών
Δικτύων που έχουν εκπαιδευθεί με έμμεσες διεργασίες**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Λάζαρος Κανδυλιώτης, ΑΕΜ: 58172

ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ

Επιβλέπων Καθηγητής: Ιωάννης, Μπούταλης, Καθηγητής, Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών, Δ.Π.Θ.

2ο Μέλος: Νικόλαος, Μητιανούδης, Καθηγητής, Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών, Δ.Π.Θ.

3ο Μέλος: Αυγερινός, Αραμπατζής, Καθηγητής, Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών, Δ.Π.Θ.

Ξάνθη, 2024



DEMOCRITUS UNIVERSITY OF THRACE
SCHOOL OF ENGINEERING
DEPARTMENT OF ELECTRICAL AND COMPUTER
ENGINEERING
SECTOR OF ELECTRONICS AND INFORMATION
TECHNOLOGY SYSTEMS

Evaluation of generalization of Convolutional Neural Networks trained with proxy tasks

DIPLOMA THESIS

Lazaros Kandyliotis, Registration Number 58172

COMMITTEE OF EXAMINERS

Supervisor: Yiannis, Boutalis, Professor, Electrical and Computer Engineering, Democritus University of Thrace

Member 2: Nikolaos, Mitianoudis, Professor, Electrical and Computer Engineering, Democritus University of Thrace

Member 3: Avgerinos, Arampatzis, Professor, Electrical and Computer Engineering, Democritus University of Thrace

Xanthi, 2024

Ευχαριστίες

Με την ολοκλήρωση της παρούσας διπλωματικής εργασίας, θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα αυτής, Καθηγητή κ. Ιωάννη Μπούταλη, για την πολύτιμη βοήθεια, την εμπιστοσύνη και τη στήριξή του κατά τη διάρκεια αυτής της προσπάθειας. Η καθοδήγησή του υπήρξε καθοριστική για την ολοκλήρωση του έργου μου. Ιδιαίτερες ευχαριστίες απευθύνω στον διδάκτορα κ. Σωκράτη Γκέλιο, ο οποίος υπήρξε πολύτιμος μέντορας στην ακαδημαϊκή μου πορεία, παρέχοντάς μου διαρκώς τη βοήθειά του και τις πολύτιμες συμβουλές του σε κάθε στάδιο αυτής της διαδρομής.

Επιπλέον, θα ήθελα να εκφράσω την βαθιά μου ευγνωμοσύνη στους γονείς μου, Μάρθα και Αλέξανδρο, καθώς και στον αδερφό μου, Νίκο, για την αμέριστη και απεριόριστη στήριξη που μου προσέφεραν καθ' όλη τη διάρκεια της ζωής μου. Η αδιάκοπη αγάπη, ηθική υποστήριξη και καθοδήγησή τους αποτελούν ακρογωνιαίο λίθο των προσπαθειών και των επιτυχιών μου. Επίσης, δεν θα μπορούσα να μην αναφέρω την υποστήριξη που έλαβα από τους φίλους μου, που στάθηκαν δίπλα μου σε κάθε βήμα αυτής της διαδικασίας και με ενθάρρυναν σε όλες τις προκλήσεις.

«Μη, εί τι αυτώ σοί δυσκαταπόνητον, τούτο ανθρώπῳ αδύνατον υπολαμβάνειν, αλλ' εί τι ανθρώπῳ δυνατόν και οικείον, τούτο και σεαυτώ εφικτόν νομίζειν.»

-Μάρκος Αιρήλιος

Περίληψη

Τα τελευταία χρόνια, η Βαθιά Μάθηση (*Deep Learning - DL*) και οι εφαρμογές της στην Υπολογιστική Όραση (*Computer Vision - CV*) έχουν σημειώσει ραγδαία εξέλιξη. Ωστόσο, πολλά σύνολα δεδομένων (*datasets*) παραμένουν ανεκμετάλλευτα, καθώς δεν είναι επισημασμένα (*unlabeled*). Η επισήμανση αυτών των δεδομένων είναι μια χρονοβόρα διαδικασία και απαιτεί συχνά την παρέμβαση εξειδικευμένων ατόμων, γεγονός που καθιστά δύσκολη τη χρήση τους. Σε αυτό το σημείο, η λύση έρχεται μέσω της Αυτο-εποπτευόμενης Μάθησης (*Self-Supervised Learning - SSL*), η οποία χρησιμοποιεί έμμεσες διεργασίες (*proxy/pretext tasks*) για να εκπαιδεύσει μοντέλα χωρίς την ανάγκη επισημάνσεων. Σε αυτές τις έμμεσες διεργασίες, το μοντέλο δημιουργεί τις δικές του ετικέτες (*labels*) παραμορφώνοντας ή μετασχηματίζοντας τα δεδομένα με σκοπό να μάθει χρήσιμα χαρακτηριστικά, όπως ο προσανατολισμός των αντικειμένων, τα χρώματα και η συνοχή των εικόνων, τα οποία μπορεί να αξιοποιήσει σε άλλες εργασίες όπως η Ταξινόμηση Εικόνων (*Image Classification*) ή η Ανίχνευση Αντικειμένων (*Object Detection*).

Η παρούσα διπλωματική εργασία έχει ως στόχο τη σύγκριση τριών διαφορετικών έμμεσων διεργασιών: Πρόβλεψη Περιστροφής Εικόνας (*Image Rotation Prediction*), Χρωματοποίηση Εικόνας (*Image Colorization*) και Επιδιόρθωση/Συμπλήρωση Εικόνας (*Image Inpainting*). Σκοπός είναι να δημιουργηθεί ένα περιβάλλον δίκαιης σύγκρισης, προκειμένου να αξιολογηθεί η αποτελεσματικότητα αυτών των διεργασιών σε μία διαφορετικής φύσης εργασία (*downstream task*) ταξινόμησης εικόνων, η οποία χρειάζεται επισημασμένα δεδομένα για την εκπαίδευση της. Η διασφάλιση της δικαιοσύνης στη σύγκριση επιτυγχάνεται μέσω της χρήσης ίδιας αρχιτεκτονικής, υπερπαραμέτρων και άλλων ρυθμίσεων σε όλα τα proxy tasks καθώς και στο *downstream task*.

Για την υλοποίηση των τριών έμμεσων διεργασιών, χρησιμοποιήθηκε ως βάση το μοντέλο *ResNet-50* [4], ένα από τα πιο αποδοτικά και δημοφιλή μοντέλα στα Συνελικτικά Νευρωνικά Δίκτυα (*Convolutional Neural Networks - CNNs*), το οποίο προσαρμόστηκε ανάλογα με τη φύση κάθε διεργασίας. Η βασική αρχή της αρχιτεκτονικής του *ResNet-50* είναι τα "*skip connections*", τα οποία επιτρέπουν τη μετάδοση πληροφοριών από προηγούμενα επίπεδα του μοντέλου στα επόμενα, αποφεύγοντας το φαινόμενο της εξαφάνισης των βαθμίδων (*vanishing gradient*). Τα προεκπαιδευμένα βάρη που αποκτήθηκαν από τις έμμεσες διεργασίες μεταφέρθηκαν μέσω της τεχνικής της Μεταφοράς Μάθησης (*Transfer Learning*) και τεστάρονται στο *downstream task* της ταξινόμησης εικόνων.

Για την εκπαίδευση των proxy tasks, χρησιμοποιήθηκε το σύνολο δεδομένων *CIFAR-100* [14] ως ένα *unlabeled dataset*, δηλαδή χωρίς την χρήση ετικετών. Για το *downstream task* της ταξινόμησης εικόνων χρησιμοποιήθηκαν τα σύνολα δεδομένων *CIFAR-10* [14] και *CIFAR-100*. Στο *downstream task*, πραγματοποιήθηκε fine-tuning με δύο διαφορετικούς τρόπους: μία περίπτωση περιλάμβανε fine-tuning μόνο στο τελευταίο επίπεδο (*fully connected layer*) του μοντέλου, προκειμένου να συγκριθεί η πληροφορία των προεκπαιδευμένων βαρών μεταξύ τους και με τα τυχαία βάρη. Στη δεύτερη περίπτωση, πραγματοποιήθηκε fine-tuning σε όλα τα επίπεδα του μοντέλου, προκειμένου να αξιολογηθεί αν η πλήρης προσαρμογή των προεκπαιδευμένων βαρών αποδίδει καλύτερα από τα τυχαία βάρη.

Στο πειραματικό μέρος της εργασίας, τα αποτελέσματα και οι μετρήσεις έδειξαν ότι η αυτο-εποπτευόμενη μάθηση μέσω έμμεσων διεργασιών προσφέρει σημαντικά καλύτερες επιδόσεις σε σχέση με την εκπαίδευση με τυχαία αρχικοποιημένα βάρη. Επιπλέον, το μοντέλο καταφέρνει να πετύχει μια πολύ ικανοποιητική απόδοση, κοντά στη μέγιστη, με λιγότερες εποχές εκπαίδευσης, καθιστώντας τη διαδικασία πιο γρήγορη και οικονομική. Το πιο αποδοτικό proxy task αναδείχθηκε το *Image Inpainting*, ενώ πολύ κοντά σε απόδοση ήταν και το *Image Colorization*. Το *Image Rotation Prediction* είχε τις χαμηλότερες επιδόσεις, αλλά συνέβαλε και αυτό στην ενίσχυση της απόδοσης του μοντέλου σε σύγκριση με τα τυχαία βάρη.

Λέξεις Κλειδιά:

1. Συνελικτικά Νευρωνικά Δίκτυα
2. Αυτό-Εποπτευόμενη Μάθηση
3. Μεταφορά Μάθησης
4. Έμμεσες Διεργασίες
5. Πρόβλεψη Περιστροφής Εικόνας
6. Χρωματοποίηση Εικόνας
7. Επιδιόρθωση/Συμπλήρωση Εικόνας
8. Ταξινόμηση Εικόνων

Abstract

In recent years, *Deep Learning (DL)* and its applications in *Computer Vision (CV)* have seen rapid advancements. Despite this progress, many datasets remain underutilized due to the lack of annotations (unlabeled data). Labeling these datasets is a time-intensive process that often requires specialized human intervention, which makes them difficult to leverage effectively. Self-Supervised Learning (*SSL*) provides a promising solution to this problem by utilizing *proxy/pretext tasks* that train models without requiring labeled data. In these tasks, the model generates its own labels by manipulating the data, enabling it to learn meaningful features like object orientation, color, and image coherence, which can then be applied to tasks such as *Image Classification* or *Object Detection*.

This thesis aims to compare three different proxy tasks: *Image Rotation Prediction*, *Image Colorization*, and *Image Inpainting*. The objective is to establish a fair comparison environment to evaluate the effectiveness of these tasks in the downstream task of image classification, which is inherently different as it requires labeled data for training. To ensure a fair comparison, the same architecture, hyperparameters, and settings were consistently applied across all proxy tasks and the downstream task.

For the implementation of the three proxy tasks, the *ResNet-50* [4] model was used as the base architecture, one of the most efficient and popular models in *Convolutional Neural Networks (CNNs)*. The model was adapted to suit the specific requirements of each task. The key principle of the *ResNet-50*'s architecture is the use of "*skip connections*," which allow the transmission of information from earlier layers to subsequent layers, helping to prevent the vanishing gradient phenomenon. The pretrained weights obtained from these proxy tasks were transferred via Transfer Learning and were then evaluated on the downstream task of image classification.

For the training of the proxy tasks, the *CIFAR-100* [14] dataset was used as an unlabeled dataset, meaning only the images without their labels were utilized. For the downstream task of image classification, the *CIFAR-10* [14] and *CIFAR-100* datasets were used. In the downstream task, fine-tuning was performed in two different ways: one involved fine-tuning only the last fully connected layer of the model to compare the information learned by the pretrained weights against random weights. In the second case, fine-tuning was done on all layers of the model to assess whether fully adapting the pretrained weights yields better performance than random weights.

In the experimental part of the thesis, the experimental results and metrics showed that Self-Supervised Learning through proxy tasks offers significantly better performance compared to training with randomly initialized weights. Additionally, the model was able to achieve a high level of performance, close to its peak performance, which would typically be reached after many more epochs, making the process faster and more efficient. The most effective proxy task was Image Inpainting, followed closely by Image Colorization. Image Rotation Prediction had the lowest performance, but still contributed to improving the model's performance compared to random weights.

Key Words:

1. Convolutional Neural Networks
2. Self-Supervised Learning
3. Transfer Learning
4. Proxy/Pretext Tasks
5. Image Rotation Prediction
6. Image Colorization
7. Image Inpainting
8. Image Classification

Περιεχόμενα

Ευχαριστίες	1
Περίληψη	3
Περιεχόμενα.....	7
Εισαγωγή.....	11
0.1 Υπόβαθρο και Κίνητρα	11
0.1.1 Ανάγκη για Αυτο-Εποπτευόμενη Μάθηση και Χρήση των Έμμεσων Διεργασιών	11
0.1.2 Ρόλος και Χρησιμότητα στην Τεχνητή Νοημοσύνη	11
0.2 Σχετική Έρευνα	13
0.2.1 Ιστορία Συνελικτικών Νευρωνικών Δικτύων.....	13
0.2.2 Αυτο-Εποπτευόμενη Μάθηση και Έμμεσες Διεργασίες	14
0.2.3 Μεταφορά Μάθησης	19
0.3 Στόχοι και Συνεισφορές.....	21
0.3.1 Ερευνητικά Ερωτήματα, Στόχοι και Δομή της Εργασίας.....	21
0.3.2 Προοπτικές Μελλοντικής Έρευνας και Εφαρμογές	22
I. Θεωρητικό Μέρος.....	25
Κεφάλαιο 1°	
1. Νευρωνικά Δίκτυα	27
1.1 Εισαγωγή στα Τεχνητά Νευρωνικά Δίκτυα (ANNs)	27
1.1.1 Ορισμός και Βασικά Χαρακτηριστικά των ANNs.....	27
1.1.2 Λειτουργία, Τύποι Αρχιτεκτονικών και Εφαρμογές των ANNs	28
1.2 Είδη Μάθησης.....	35
1.2.1 Εποπτευόμενη Μάθηση	35
1.2.2 Μη Εποπτευόμενη Μάθηση	36
1.2.3 Ημι-Εποπτευόμενη Μάθηση	36
1.2.4 Αυτο-εποπτευόμενη Μάθηση.....	37
1.2.5 Ενισχυτική Μάθηση.....	38
1.3 Συνελικτικά Νευρωνικά Δίκτυα (CNNs)	41
1.3.1 Εισαγωγή στα CNNs.....	41

1.3.2 Δομή και Λειτουργία Επιπέδων στα CNNs.....	41
1.3.3 Τεχνικές Κανονικοποίησης	44
1.3.4 Σύγχρονα Μοντέλα CNNs	47
1.3.5 Αυτο-κωδικοποιητές	50
1.3.6 Εφαρμογές των CNNs.....	53
Κεφάλαιο 2^ο	
2. Αυτο-εποπτευόμενη Μάθηση και Έμμεσες Διεργασίες	55
2.1 Η Θεωρία της Αυτο-εποπτευόμενης Μάθησης	55
2.1.1 Ορισμός και Θεμελιώδεις Αρχές	55
2.1.2 Ο Ρόλος των Έμμεσων Διεργασιών στην Αυτο-εποπτευόμενη Μάθηση	56
2.1.3 Κατανόηση Σχέσεων με Άλλες Μορφές Μάθησης	56
2.1.4 Τεχνικές Υλοποίησης	57
2.2 Έμμεσες Διεργασίες	59
2.2.1 Διαφορετικοί Τύποι Έμμεσων Διεργασιών	59
2.2.2 Πρόβλεψη Περιστροφής Εικόνας.....	60
2.2.3 Χρωματοποίηση Εικόνας.....	61
2.2.4 Επιδιόρθωση/Συμπλήρωση Εικόνας.....	62
2.3 Περιορισμοί και Προκλήσεις.....	65
2.3.1 Περιορισμένη Γενίκευση των Έμμεσων Διεργασιών	65
2.3.2 Υπολογιστικό Κόστος, Πόροι και Χρόνος Εκπαίδευσης	65
2.3.3 Προβλήματα Υπερπροσαρμογής	66
2.3.4 Δυσκολίες στη Μεταφορά Χαρακτηριστικών και Ασυμβατότητα Εργασιών	66
2.4 Εφαρμογές.....	67
2.4.1 Επεξεργασία Εικόνας.....	67
2.4.2 Ανάλυση Βίντεο	67
2.4.3 Επεξεργασία Φυσικής Γλώσσας.....	68
2.4.4 Ρομποτική.....	69
2.4.5 Άλλες Εφαρμογές	70
II. Πειραματικό Μέρος.....	73
Κεφάλαιο 3^ο	
3. Διεξαγωγή Πειραμάτων	75

3.1 Μεθοδολογία Πειραμάτων	75
3.1.1 Μεθοδολογία και Λογική Συγκρίσεων	75
3.1.2 Ερωτήματα και Προσδοκίες	76
3.2 Υλοποιηση Έμμεσων Διεργασιών	79
3.2.1 Υλοποίηση Διεργασίας Πρόβλεψης Περιστροφής Εικόνας	79
3.2.2 Υλοποίηση Διεργασίας Χρωματοποίησης Εικόνας	80
3.2.3 Υλοποίηση Διεργασίας Επιδιόρθωσης/Συμπλήρωσης Εικόνας	82
3.3 Downstream Tasks και Σύνολα Δεδομένων	85
3.3.1 Ταξινόμηση Εικόνων ως Downstream Task	85
3.3.2 Σύνολα Δεδομένων που Χρησιμοποιήθηκαν	86
3.4 Μεταφορά Μάθησης και Fine-Tuning	89
3.4.1 Μεταφορά Μάθησης από τα Proxy Tasks στο Downstream Task	89
3.4.2 Fine-Tuning στο Downstream Task	89
3.5 Αποτελέσματα	91
3.5.1 Αποτελέσματα 1ου Πειράματος	91
3.5.2 Αποτελέσματα 2ου Πειράματος	94
3.5.3 Αποτελέσματα 3ου Πειράματος	97
3.5.4 Αποτελέσματα 4ου Πειράματος	102
3.5.5 Συγκεντρωτική Απεικόνιση Αποτελεσμάτων	107
3.6 Σχολιασμός Αποτελεσμάτων - Συμπεράσματα	111
3.7 Επίλογος και Μελλοντικές Επεκτάσεις	115
Βιβλιογραφία	117
Βιβλία	117
Αναφορές	117

Εισαγωγή

0.1 Υπόβαθρο και Κίνητρα

0.1.1 Ανάγκη για Αυτο-Εποπτευόμενη Μάθηση και Χρήση των Έμμεσων Διεργασιών

Η Αυτο-εποπτευόμενη Μάθηση έχει αναδειχθεί ως μια σημαντική και καινοτόμος προσέγγιση στο χώρο της Μηχανικής Μάθησης και της Τεχνητής Νοημοσύνης, απαντώντας σε κρίσιμες προκλήσεις που αντιμετωπίζουν οι παραδοσιακές μέθοδοι μάθησης. Ενώ οι εποπτευόμενες μέθοδοι απαιτούν επισημασμένα δεδομένα για να επιτύχουν υψηλά επίπεδα απόδοσης, η απόκτηση τέτοιων δεδομένων είναι συχνά δαπανηρή και χρονοβόρα. Παράλληλα, υπάρχουν ήδη τεράστια και εύκολα προσβάσιμα σύνολα δεδομένων χωρίς ετικέτες, τα οποία παραμένουν σε μεγάλο βαθμό ανεκμετάλλευτα λόγω της έλλειψης επισήμανσης. Σε πολλές περιπτώσεις, η επισήμανση δεδομένων απαιτεί την παρέμβαση εξειδικευμένου ανθρώπινου δυναμικού, γεγονός που καθιστά τη διαδικασία αυτή μη βιώσιμη σε μεγάλη κλίμακα. Αυτή η ανάγκη για επισημασμένα δεδομένα δημιουργεί περιορισμούς στην εφαρμογή και την κλιμάκωση των εποπτευόμενων (Supervised) μεθόδων, ειδικά σε τομείς όπου τα δεδομένα είναι άφθονα αλλά ανεπιτήρητα.

Σε αυτό το πλαίσιο, το συγκεκριμένο είδος μάθησης αναδύεται ως μια εναλλακτική προσέγγιση που επιτρέπει στα μοντέλα να εκμεταλλευτούν τα πλούσια σε πληροφορίες ανεπιτήρητα δεδομένα. Χρησιμοποιώντας μόνο τα διαθέσιμα δεδομένα, τα μοντέλα μπορούν να δημιουργήσουν τις δικές τους εποπτευόμενες εργασίες, γνωστές ως έμμεσες διεργασίες, όπως η πρόβλεψη περιστροφής εικόνας ή χρωματοποίηση εικόνας η συμπλήρωση ελλειπόντων τμημάτων μιας εικόνας. Επομένως, το κίνητρο πίσω από την αυτο-εποπτευόμενη μάθηση είναι αυτές οι εργασίες να συμβάλλουν στην εκμάθηση χρησιμών αναπαραστάσεων των δεδομένων από μη επισημασμένα δεδομένα χρησιμοποιώντας την έννοια της αυτο-εποπτείας και στη συνέχεια να βελτιώσουν αυτές τις αναπαραστάσεις με τη χρήση κάποιων επισημάνσεων για την επιτηρούμενη εργασία που ακολουθεί (downstream task). Αυτές οι εργασίες που ακολουθούν μπορεί να κυμαίνονται από απλές έως σύνθετες, όπως η ταξινόμηση εικόνων (image classification), η σημασιολογική τμηματοποίηση (semantic segmentation) και η ανίχνευση αντικειμένων (object detection).

Συνολικά, η αυτο-εποπτευόμενη μάθηση και η χρήση των έμμεσων διεργασιών αντιπροσωπεύουν μια σημαντική πρόοδο στην κατανόηση και αξιοποίηση των δεδομένων. Με αυτές τις μεθόδους, τα μοντέλα γίνονται πιο ανεξάρτητα και αποτελεσματικά στην εκμάθηση, χωρίς να απαιτείται εξωτερική εποπτεία. Αυτή η προσέγγιση δεν αντιμετωπίζει μόνο τις προκλήσεις που υπάρχουν στη μηχανική μάθηση, αλλά και ανοίγει νέες δυνατότητες για την ανάπτυξη πιο ισχυρών και προσαρμοστικών συστημάτων τεχνητής νοημοσύνης, τα οποία μπορούν να ανταποκριθούν αποτελεσματικά σε διάφορες εφαρμογές και πραγματικές συνθήκες.

0.1.2 Ρόλος και Χρησιμότητα στην Τεχνητή Νοημοσύνη

- Μείωση Υπολογιστικού Κόστους και Ανάγκης Ανθρώπινου Δυναμικού:

Η μείωση του υπολογιστικού κόστους αποτελεί έναν από τους βασικούς λόγους για τη χρήση της αυτο-εποπτευόμενης μάθησης. Με τις παραδοσιακές μεθόδους, η ανάγκη για μεγάλα επισημασμένα σύνολα δεδομένων οδηγεί σε αυξημένα κόστη τόσο σε υπολογιστικούς πόρους όσο και σε χρόνο. Το

πλεονέκτημα της αυτο-εποπτευόμενης μάθησης είναι η ικανότητα εκμάθησης από μη επισημασμένα δεδομένα, μειώνοντας έτσι την ανάγκη για εντατική ανθρώπινη παρέμβαση και επισήμανση. Επιπλέον, πολλές φορές αυτή η προσέγγιση καταφέρνει να επιτύχει επιδόσεις ανάλογες με αυτές των παραδοσιακών μεθόδων σε μικρότερο χρονικό διάστημα, κάτι που συμβάλλει περαιτέρω στη μείωση του υπολογιστικού κόστους. Αυτό, με τη σειρά του, επιταχύνει τη διαδικασία εκπαίδευσης και μειώνει σημαντικά το συνολικό κόστος, καθιστώντας τα μοντέλα όχι μόνο πιο αποδοτικά, αλλά και πιο οικονομικά βιώσιμα.

- **Χρήση Δεδομένων με Περιορισμούς ή Μη Επισημασμένων Δεδομένων:**

Σε ορισμένες περιπτώσεις, η πρόσβαση σε επισημασμένα δεδομένα είναι είτε δύσκολη είτε αδύνατη λόγω ηθικών, νομικών ή πρακτικών περιορισμών. Για παράδειγμα, στην ιατρική έρευνα, η επισήμανση δεδομένων μπορεί να απαιτεί την αποκάλυψη ευαίσθητων πληροφοριών, κάτι που μπορεί να εγείρει ζητήματα ιδιωτικότητας και ηθικής. Επιπλέον, σε πολλές περιπτώσεις, τα δεδομένα είναι απλώς μη επισημασμένα και η επισήμανσή τους είναι είτε μη πρακτική είτε αδύνατη λόγω του όγκου τους. Μέσω της χρήσης τεχνικών που επιτρέπουν τη μάθηση από τέτοιου είδους δεδομένα, τα μοντέλα τεχνητής νοημοσύνης μπορούν να συνεχίσουν να μαθαίνουν και να εξελίσσονται, παρακάμπτοντας αυτούς τους περιορισμούς και διασφαλίζοντας ότι η έλλειψη επισημασμένων δεδομένων δεν αποτελεί εμπόδιο στην πρόοδο της έρευνας και της ανάπτυξης.

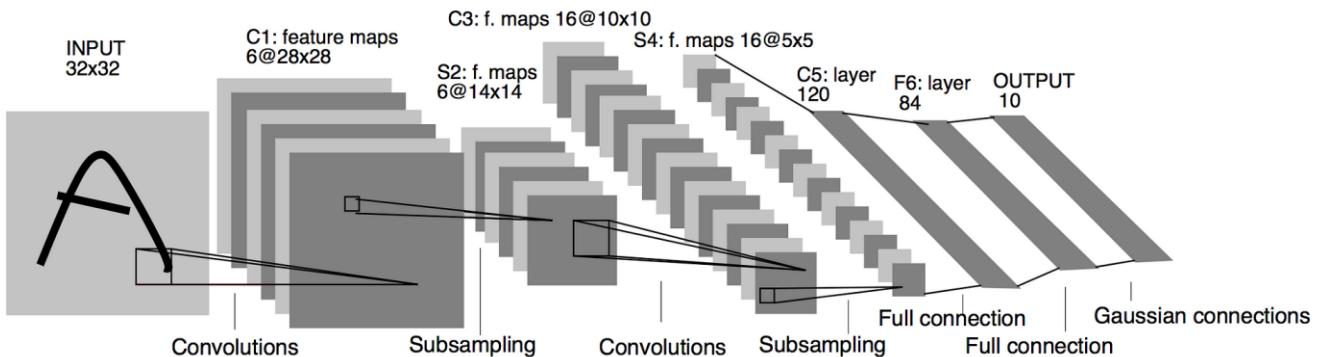
0.2 Σχετική Έρευνα

0.2.1 Ιστορία Συνελικτικών Νευρωνικών Δικτύων

Η ιστορία των Συνελικτικών Νευρωνικών Δικτύων (CNNs) ξεκίνησε με την εργασία του Yann LeCun κ.α.[1] το 1989, όταν αυτός και οι συνεργάτες του παρουσίασαν για πρώτη φορά ένα CNN που ονομάστηκε LeNet-5, για την αναγνώριση χειρόγραφων ψηφίων στο σύνολο δεδομένων MNIST. Το LeNet-5 βασίστηκε στην ικανότητα των συνελικτικών επιπέδων να ανιχνεύουν χωρικά πρότυπα και χαρακτηριστικά μέσα στις εικόνες, όπως γραμμές ή σχήματα, ενώ οι κρυφές στρώσεις μπορούσαν να συνδυάσουν αυτά τα χαρακτηριστικά για την ταξινόμηση των εικόνων.

Αυτή η προσέγγιση ήταν επαναστατική για την εποχή, διότι πριν από τα CNNs, οι περισσότερες μέθοδοι αναγνώρισης εικόνας βασίζονταν σε χειροκίνητα εξαγόμενα χαρακτηριστικά. Αντίθετα, τα CNNs επέτρεψαν στα μοντέλα να μάθουν αυτόματα ποια χαρακτηριστικά ήταν πιο σημαντικά απευθείας από τα δεδομένα, μειώνοντας την ανάγκη για χειροκίνητη παρέμβαση. Οι βασικές αρχές που εισήγαγε ο LeCun με το LeNet-5, όπως τα συνελικτικά επίπεδα για την ανίχνευση χαρακτηριστικών και τα pooling layers για μείωση της χωρικής ανάλυσης, αποτέλεσαν τη βάση για μελλοντικά, βαθύτερα δίκτυα.

Αυτή η καινοτομία άνοιξε το δρόμο για τη ραγδαία ανάπτυξη των CNNs, τα οποία σύντομα εξελίχθηκαν σε μια από τις βασικές τεχνολογίες της επεξεργασίας εικόνας και της βαθιάς μάθησης.



Εικόνα 0.2.1 - 1: LeNet Architecture.

(Πηγή: LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2222-2234.)

Στη συνέχεια, η έρευνα στα CNNs αναπτύχθηκε με την εισαγωγή των βαθύτερων νευρωνικών δικτύων, όπως το AlexNet το 2012 από τον Alex Krizhevsky [2] και τους συνεργάτες του. Το AlexNet, που κέρδισε τον διαγωνισμό ImageNet, ήταν καθοριστικό για τη μετάβαση από πιο ρηχά μοντέλα σε βαθύτερα δίκτυα που χρησιμοποιούνταν για τη βελτίωση της απόδοσης και της γενίκευσης. Η νίκη του AlexNet στον διαγωνισμό αυτό σηματοδότησε την αρχή της "επανάστασης" των CNNs στην επεξεργασία εικόνας.

Αργότερα, μοντέλα όπως το VGGNet [3], το ResNet [4], το GoogleNet [5] και το EfficientNet [6] συνέβαλαν περαιτέρω στην εξέλιξη των CNNs, αντιμετωπίζοντας κρίσιμα προβλήματα όπως το πρόβλημα του εξαφανισμένου gradient (vanishing gradient) και των περιορισμών στον αριθμό των παραμέτρων (στο κεφάλαιο των Νευρωνικών Δικτύων θα αναλύσουμε τον τρόπο λειτουργίας των παραπάνω μοντέλων). Το ResNet, ειδικότερα, εισήγαγε την υπολειμματική μάθηση (residual learning), που επέτρεψε την εκπαίδευση ακόμα βαθύτερων δικτύων χωρίς απώλεια της απόδοσης. Αυτή η καινοτομία κατέστησε το ResNet ένα από τα πιο επιτυχημένα μοντέλα CNN και χρησιμοποιείται συχνά σε πειραματικές διαδικασίες, συμπεριλαμβανομένης της παρούσας μελέτης.

0.2.2 Αυτο-Εποπτεύομενη Μάθηση και Έμμεσες Διεργασίες

Η Αυτο-Εποπτεύομενη Μάθηση αποτελεί ένα από τα πιο γρήγορα εξελισσόμενα πεδία στη μηχανική μάθηση, με ρίζες σε προηγούμενες προσεγγίσεις μη εποπτεύομενης και εποπτεύομενης μάθησης. Η έρευνα για την SSL κέρδισε ιδιαίτερο ενδιαφέρον με το έργο του DeepMind και της Facebook AI Research (FAIR), που χρησιμοποίησαν τεχνικές contrastive learning για την εκμάθηση αναπαραστάσεων χωρίς ετικέτες. Το μοντέλο MoCo (Momentum Contrast) (He et al., 2020) [7] αποτέλεσε μια σημαντική συμβολή, επιτρέποντας τη δημιουργία ευέλικτων και ισχυρών αναπαραστάσεων από τεράστια σύνολα δεδομένων χωρίς την ανάγκη ανθρώπινης επισημείωσης. Η μεθοδολογία αυτή χρησιμοποιεί τη μεγιστοποίηση της αντιπαράθεσης (contrastive loss), μαθαίνοντας να συγκρίνει τις παραστάσεις δεδομένων που θεωρούνται παρόμοια ή διαφορετικά.

Το βασικότερο στοιχείο στην Αυτο-εποπτεύομενη Μάθηση είναι οι έμμεσες διεργασίες. Στη βιβλιογραφία υπάρχουν διαφορετικά ήδη και υλοποιήσεις των έμμεσων διεργασιών. Παρακάτω παρουσιάζονται ορισμένα από τα πιο διαδεδομένα proxy tasks που έχουν δημοσιευθεί σε επιστημονικά papers. Αυτά τα tasks έχουν χρησιμοποιηθεί εκτενώς στην αυτο-εποπτεύομενη μάθηση για την εκπαίδευση μοντέλων χωρίς τη χρήση επισημασμένων δεδομένων. Κάθε ένα από αυτά τα tasks βελτιώνει την ικανότητα του μοντέλου να μαθαίνει ουσιαστικά χαρακτηριστικά των δεδομένων, επιτρέποντας τη μεταφορά αυτών των γνώσεων σε άλλες, πιο εξειδικευμένες εφαρμογές, όπως η ταξινόμηση εικόνων και η ανίχνευση αντικειμένων:

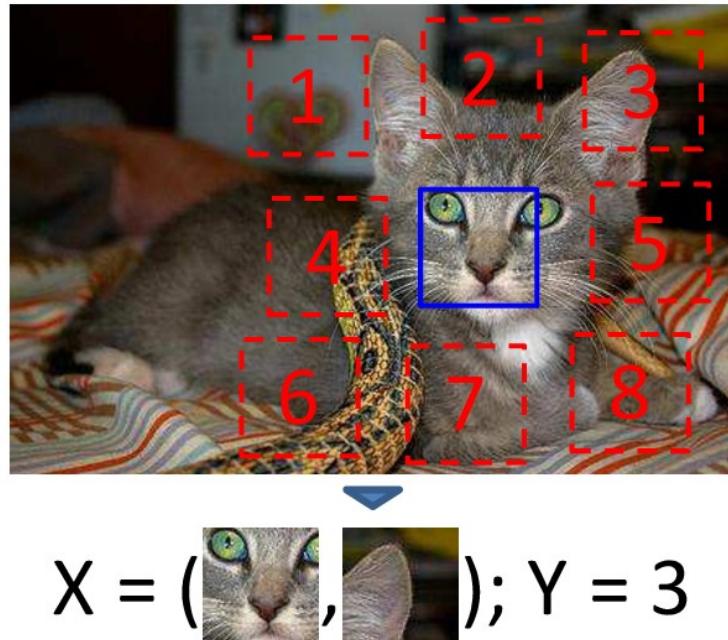
Context Prediction Task: Η διεργασία Context Prediction (Πρόβλεψη Περιεχομένου), που προτάθηκε από τους Doersch, Gupta, και Efros το 2015 [8], είχε ως στόχο την πρόβλεψη της χωρικής διάταξης δύο τμημάτων (patches) μιας εικόνας. Στο άρθρο τους με τίτλο "Unsupervised Visual Representation Learning by Context Prediction," παρουσίασαν ένα πρωτότυπο proxy task, όπου το μοντέλο εκπαιδεύεται να προβλέψει τη σχετική θέση ενός τμήματος εικόνας σε σχέση με ένα άλλο.

Πιο αναλυτικά, η εικόνα χωρίζεται σε πολλά μικρά τμήματα και δύο από αυτά επιλέγονται τυχαία. Το μοντέλο καλείται στη συνέχεια να προβλέψει πού βρίσκεται το ένα τμήμα σε σχέση με το άλλο, για παράδειγμα αν είναι αριστερά, δεξιά, πάνω ή κάτω. Καθώς το μοντέλο προσπαθεί να προβλέψει τη σχετική τους θέση, αναγκάζεται να κατανοήσει τη συνολική δομή της εικόνας, όπως το σχήμα των αντικειμένων και το πώς τα τμήματα συνδέονται μεταξύ τους.

Το κίνητρο πίσω από αυτό το proxy task είναι η ιδέα ότι, αν το μοντέλο μπορεί να κατανοήσει τη σχετική θέση των τμημάτων μέσα σε μια εικόνα, τότε θα έχει μάθει τη συνολική χωρική διάταξη και τη σημασία της εικόνας. Αυτό είναι απαραίτητο για πιο σύνθετες εργασίες, όπως η αναγνώριση αντικειμένων και η τμηματοποίηση εικόνας.

Ο Doersch και οι συνάδελφοί του έδειξαν ότι το μοντέλο τους μπορούσε να μάθει χρήσιμες αναπαραστάσεις χωρίς επισημασμένα δεδομένα. Όταν το μοντέλο δοκιμάστηκε σε datasets όπως το ImageNet, οι αναπαραστάσεις που έμαθε αποδείχθηκαν αποτελεσματικές για εργασίες μεταφοράς μάθησης, αποδεικνύοντας ανταγωνιστική απόδοση σε σχέση με τις πλήρως εποπτεύομενες προσεγγίσεις.

Αυτό το paper είχε σημαντική επιρροή στην έρευνα της αυτο-εποπτεύομενης μάθησης, καθώς εισήγαγε ένα ισχυρό proxy task που βοήθησε στη βελτίωση των μοντέλων μηχανικής μάθησης με χρήση μη επισημασμένων δεδομένων.



Εικόνα 0.2.2 - 1: Context Prediction Task: Το μοντέλο πρέπει να μάθει τη σχετική θέση μεταξύ δύο κομματιών εικόνας που έχουν αποκοπεί από την ίδια αρχική εικόνα. Στη συγκεκριμένη περίπτωση, πρέπει να προβλέψει ότι το δεξί κομμάτι είναι στη θέση 3.

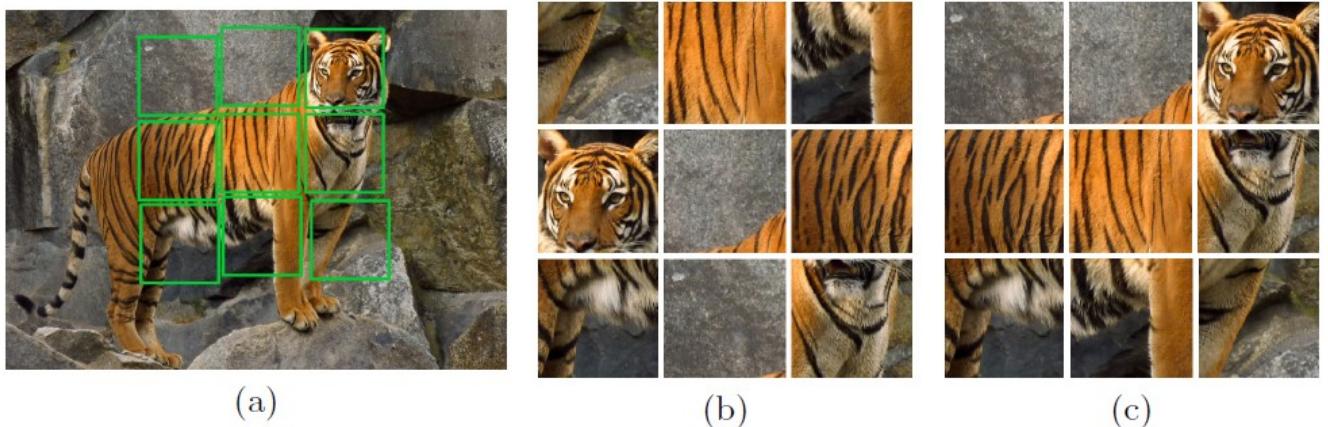
(Πηγή: Doersch, C., Gupta, A., & Efros, A. A. (2015). Unsupervised visual representation learning by context prediction. In *Proceedings of the IEEE international conference on computer vision* (pp. 1422-1430).)

Jigsaw Puzzle Solving Task: Το *Jigsaw Puzzle Solving Task* (Διεργασία Επίλυσης παζλ) είναι ένα από τα πιο γνωστά proxy tasks για την αυτο-εποπτεύομενη μάθηση, το οποίο παρουσιάστηκε το 2016 από τους Noroozi και Favaro [9] στο paper τους με τίτλο "*Unsupervised Learning of Visual Representations by Solving Jigsaw Puzzles*." Η ιδέα πίσω από αυτό το task είναι ότι το μοντέλο πρέπει να μάθει να ανασυνθέτει μια "ανακατεμένη" εικόνα, σαν να λύνει ένα παζλ, κάτι που το αναγκάζει να κατανοήσει τις σχέσεις μεταξύ των διαφορετικών τμημάτων της εικόνας.

Η διαδικασία είναι η εξής: Μια εικόνα χωρίζεται σε 9 τετράγωνα κομμάτια, τα οποία ανακατεύονται τυχαία. Το μοντέλο καλείται να μάθει να προβλέπει τη σωστή διάταξη των κομματιών, επιστρέφοντας τα στη σωστή τους θέση. Το task αυτό απαιτεί από το μοντέλο να μάθει τη χωρική σχέση και τη συνοχή μεταξύ των τμημάτων μιας εικόνας, κάτι που οδηγεί σε εκμάθηση σημαντικών χαρακτηριστικών και δομών της εικόνας, χωρίς την ανάγκη για επισημασμένα δεδομένα.

Οι Noroozi και Favaro αξιολόγησαν το μοντέλο τους σε γνωστά datasets, όπως το ImageNet [10], και διαπίστωσαν ότι το *Jigsaw Puzzle Task* βοηθά το μοντέλο να μάθει ποιοτικές αναπαραστάσεις των εικόνων, οι οποίες είναι χρήσιμες για downstream tasks όπως η ταξινόμηση. Η μεθοδολογία τους επέδειξε καλές επιδόσεις, καθώς το μοντέλο κατάφερε να εντοπίσει τα βασικά χαρακτηριστικά των εικόνων, όπως σχήματα και υφές, οδηγώντας σε καλύτερη γενίκευση των αποτελεσμάτων.

Αυτό το paper άνοιξε τον δρόμο για τη χρήση proxy tasks στην αυτο-εποπτευόμενη μάθηση, αποδεικνύοντας ότι οι εικόνες μπορούν να γίνουν αντιληπτές χωρίς χειροκίνητη επισήμανση μέσω μεθόδων που αναγκάζουν το μοντέλο να κατανοήσει τη χωρική δομή. Το *Jigsaw Puzzle Task* συνεχίζει να χρησιμοποιείται ως σημείο αναφοράς για πολλές μελλοντικές έρευνες, βοηθώντας τα μοντέλα να μάθουν αναπαραστάσεις με αυτο-εποπτευόμενο τρόπο.



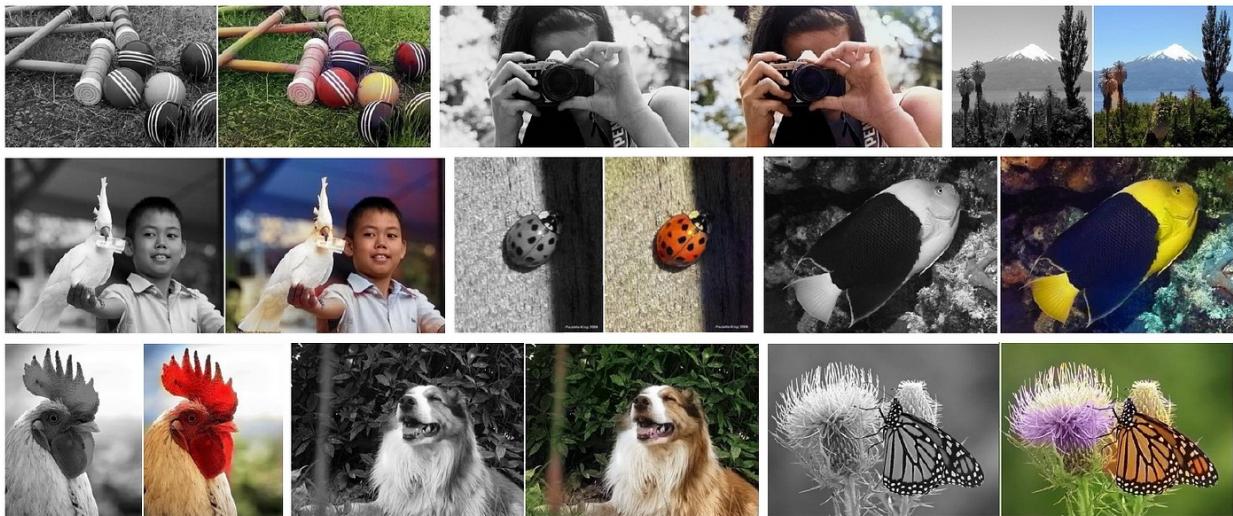
Εικόνα 0.2.2 - 2: Jigsaw Puzzle Solving Task: (a) φαίνεται η αρχική εικόνα με τα τυχαία επιλεγμένα κομμάτια της εικόνας που πρόκειται να ανακατευτούν, (b) είναι τα ανακατεμένα κομμάτια της εικόνας, (c) το μοντέλο προσπαθεί να επαναποθετήσει τα κομμάτια στην αρχική τους θέση, μαθαίνοντας τις χωρικές σχέσεις μεταξύ των χαρακτηριστικών. (Πηγή: Noroozi, M., & Favaro, P. (2016, September). Unsupervised learning of visual representations by solving jigsaw puzzles. In European conference on computer vision (pp. 69-84). Cham: Springer International Publishing.)

Image Colorization Task: Το paper με τίτλο "*Colorful Image Colorization*" από τους *Zhang, R., Isola, P., & Efros, A. A. (2016)* [11] εισήγαγε το image colorization ως ένα proxy task για την αυτο-εποπτευόμενη μάθηση. Στο έργο αυτό, το μοντέλο εκπαιδεύεται να χρωματίζει εικόνες από την grayscale εκδοχή τους, αναγκάζοντάς το να κατανοήσει τις δομές, τα αντικείμενα και τα χαρακτηριστικά που περιλαμβάνει η εικόνα για να εφαρμόσει τα κατάλληλα χρώματα.

Η διαδικασία του image colorization λειτουργεί ως εξής: η εικόνα μετατρέπεται σε grayscale και στη συνέχεια το δίκτυο καλείται να προβλέψει τα χρώματα που ταιριάζουν σε κάθε περιοχή της εικόνας, προσπαθώντας να επαναφέρει την αρχική έγχρωμη εκδοχή. Αυτό το proxy task αναγκάζει το μοντέλο να κατανοήσει το περιεχόμενο της εικόνας, όπως αντικείμενα, σχήματα και υφές, προκειμένου να προβλέψει σωστά τα χρώματα.

Μια σημαντική πτυχή αυτής της έρευνας είναι ότι τα χαρακτηριστικά που μαθαίνονται από το δίκτυο μέσω του colorization μπορούν να μεταφερθούν σε άλλα downstream tasks, όπως η ταξινόμηση εικόνων και η ανίχνευση αντικειμένων (object detection), μέσω transfer learning. Ο αλγόριθμος που προτείνεται στο paper δοκιμάστηκε σε datasets όπως το ImageNet [10] και κατάφερε να παρουσιάσει ανταγωνιστικές επιδόσεις σε άλλες εργασίες εκτός του colorization, αποδεικνύοντας την αποτελεσματικότητα της μεθόδου.

Η έρευνα αυτή έχει σημαντική συνεισφορά στον τομέα της αυτο-εποπτευόμενης μάθησης, καθώς δείχνει πως ένα proxy task όπως το colorization μπορεί να βοηθήσει το μοντέλο να κατανοήσει βαθύτερα τα δεδομένα και να χρησιμοποιήσει τις γνώσεις αυτές σε άλλες, πιο πολύπλοκες εργασίες.



Εικόνα 0.2.2 - 3: Image Colorization Task: Αφού πρώτα έχουμε μετατρέψει την αρχική εικόνα σε ασπρόμαυρη, η ασπρόμαυρη εικόνα μετατρέπεται σε χρωματισμένη. Στην αριστερή στήλη φαίνεται η ασπρόμαυρη εικόνα, ενώ στη δεξιά στήλη εμφανίζεται η εικόνα με τα προβλεπόμενα χρώματα από το μοντέλο.

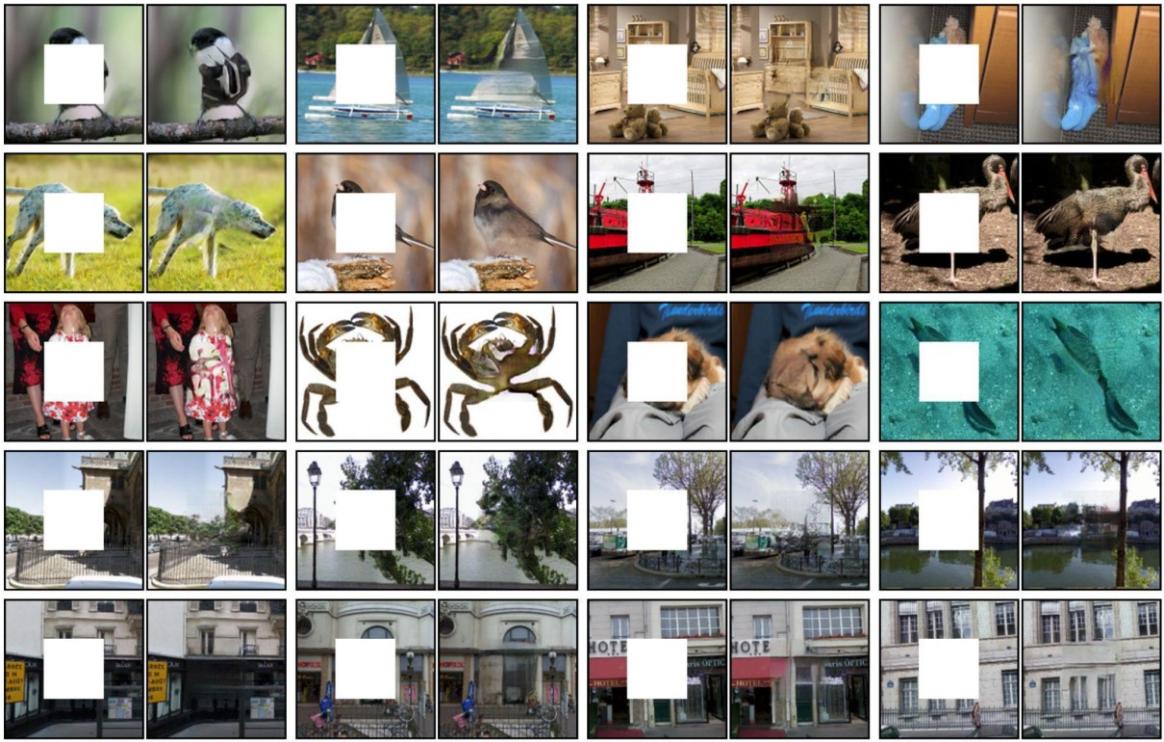
(Πηγή: Zhang, R., Isola, P., & Efros, A. A. (2016). Colorful image colorization. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part III* 14 (pp. 649–666). Springer International Publishing.)

Image inpainting Task: To paper "*Context Encoders: Feature Learning by Inpainting*" από τον Pathak κ.α. (CVPR 2016) [12] παρουσιάζει μια νέα προσέγγιση για την εκμάθηση χαρακτηριστικών μέσω του proxy task της επιδιόρθωσης/συμπλήρωσης εικόνας. Οι συγγραφείς προτείνουν ένα συνελικτικό νευρωνικό δίκτυο (CNN), το οποίο ονομάζεται *Context Encoders*, για να προβλέψει το περιεχόμενο των χαμένων τμημάτων μιας εικόνας, με βάση τα συμφραζόμενα στοιχεία που απομένουν.

Στη διαδικασία του image inpainting που περιγράφεται στο paper, η εικόνα αρχικά αλλοιώνεται με την εφαρμογή μιας μάσκας που καλύπτει τμήματα της εικόνας, συνήθως με τυχαίο τρόπο. Αυτή η μάσκα δημιουργεί κενά στην εικόνα, τα οποία το μοντέλο καλείται να "ανακατασκευάσει" ή να συμπληρώσει. Ο στόχος του μοντέλου είναι να προβλέψει τα περιεχόμενα των καλυμμένων περιοχών με βάση τα συμφραζόμενα στοιχεία της υπόλοιπης εικόνας. Το μοντέλο, λοιπόν, μαθαίνει να αναγνωρίζει τα μοτίβα και τις υφές στις γύρω περιοχές, ώστε να συμπληρώσει τις κενές περιοχές με τρόπο φυσικό και συνεκτικό. Κατά την εκπαίδευση, το μοντέλο χρησιμοποιεί μια συνάρτηση απώλειας που συγκρίνει την ανακατασκευασμένη περιοχή με την αρχική, μη αλλοιωμένη εικόνα, ώστε να βελτιώσει την ικανότητά του να προβλέπει με ακρίβεια τα χαμένα τμήματα.

Το μοντέλο εκπαιδεύεται στην επίλυση της συγκεκριμένης πρόκλησης, συνδυάζοντας δύο βασικούς όρους απώλειας: την reconstruction loss, που μετρά την απόκλιση μεταξύ της προβλεπόμενης και της πραγματικής περιοχής της εικόνας, και την adversarial loss, η οποία προσθέτει ένα στοιχείο ανταγωνιστικότητας για τη βελτίωση της ποιότητας των εικόνων. Η συνδυαστική αυτή προσέγγιση οδηγεί σε πιο ρεαλιστικές και υψηλής ποιότητας προβλέψεις για την αποκατάσταση των χαμένων περιοχών.

Το σημαντικότερο στοιχείο της εργασίας είναι ότι τα χαρακτηριστικά που μαθαίνονται από το task του inpainting, μπορούν να χρησιμοποιηθούν σε άλλα, πιο σύνθετα tasks, όπως η ταξινόμηση εικόνων, η ανίχνευση αντικειμένων, και η τμηματοποίηση εικόνας. Αυτό αποδεικνύει ότι η αυτο-εποπτευόμενη μάθηση μέσω inpainting μπορεί να οδηγήσει σε μοντέλα που γενικεύονται σε διάφορες εργασίες χωρίς την ανάγκη εκ των προτέρων επισημασμένων δεδομένων.



Εικόνα 0.2.2 - 4: Image Inpainting Task: Στην αριστερή στήλη φαίνεται η εικόνα με εφαρμοσμένη μια μάσκα που κρύβει μέρος της εικόνας, ενώ στην δεξιά στήλη φαίνεται η συμπληρωμένη εικόνα που προέβλεψε το μοντέλο.

(Πηγή: Pathak, D., Krahenbuhl, P., Donoseahue, J., Darrell, T., & Efros, A. A. (2016). Context encoders: Feature learning by inpainting. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2536-2544).)

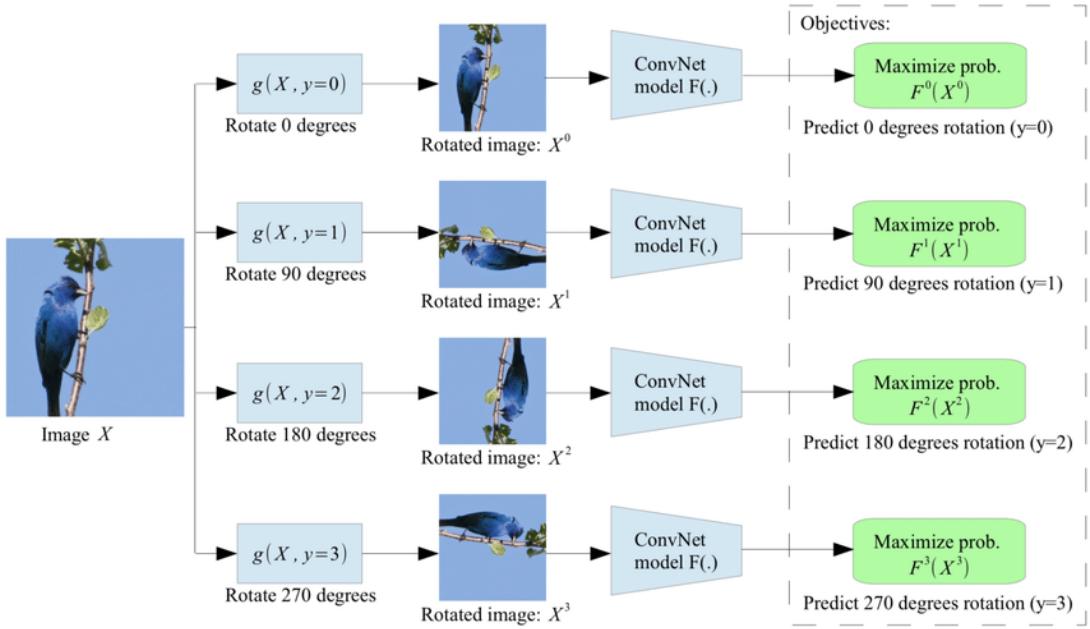
Image Rotation Prediction Task: Το *Image Rotation Task* εισήχθη από τους Gidaris, Singh και Komodakis [13] στο paper τους με τίτλο "Unsupervised Representation Learning by Predicting Image Rotations" (2018). Σε αυτή την εργασία, προτείνεται ένα καινοτόμο proxy task για την εκπαίδευση μοντέλων χωρίς επισημασμένα δεδομένα, το οποίο βασίζεται στην πρόβλεψη της γωνίας περιστροφής μιας εικόνας.

Η βασική ιδέα του task είναι η εξής: μια εικόνα περιστρέφεται σε μία από τέσσερις προκαθορισμένες γωνίες (0° , 90° , 180° , 270°), και το μοντέλο πρέπει να προβλέψει τη γωνία περιστροφής. Αυτή η προσέγγιση αναγκάζει το μοντέλο να μάθει ουσιαστικά χαρακτηριστικά της εικόνας, όπως το σχήμα και τον προσανατολισμό των αντικειμένων, για να είναι σε θέση να προβλέψει σωστά τη γωνία. Η πρόβλεψη της γωνίας λειτουργεί ως εποπτικό σήμα, χωρίς την ανάγκη για χειροκίνητα επισημασμένα δεδομένα.

Οι ερευνητές δοκίμασαν αυτή τη μέθοδο σε διάφορα γνωστά datasets, όπως το CIFAR-10 και το ImageNet, και διαπίστωσαν ότι η προσέγγιση αυτή βελτίωσε την ποιότητα των αναπαραστάσεων που μαθαίνει το μοντέλο. Το *Image Rotation* task πέτυχε ιδιαίτερα καλές επιδόσεις στα downstream tasks, όπως η ταξινόμηση και η ανίχνευση αντικειμένων, καθώς το μοντέλο είχε μάθει να κατανοεί ουσιαστικά χαρακτηριστικά της εικόνας. Επιπλέον, αυτό το task έδειξε ότι μπορεί να ενσωματωθεί με επιτυχία σε άλλες τεχνικές αυτο-εποπτευόμενης μάθησης, ενισχύοντας τη γενίκευση του μοντέλου.

Η συγκεκριμένη εργασία αποτελεί ορόσημο για την αυτο-εποπτευόμενη μάθηση, καθώς απέδειξε ότι απλά tasks όπως η πρόβλεψη γωνίας μπορούν να οδηγήσουν σε πολύ ισχυρές αναπαραστάσεις χωρίς τη χρήση

επισημασμένων δεδομένων. Το *Image Rotation* task εξακολουθεί να χρησιμοποιείται σε πολλές έρευνες που ασχολούνται με την αυτο-εποπτεύόμενη μάθηση.



Εικόνα 0.2.2 - 5: Image Rotation Prediction Task: Η εικόνα περιστρέφεται σε τέσσερις διαφορετικές γωνίες ($0^\circ, 90^\circ, 180^\circ, 270^\circ$), και το μοντέλο εκπαιδεύεται ώστε να προβλέψει τη γωνία περιστροφής της εικόνας.
(Πηγή: Gidaris, S., Singh, P., & Komodakis, N. (2018). Unsupervised representation learning by predicting image rotations. *arXiv preprint arXiv:1803.07728*.)

0.2.3 Μεταφορά Μάθησης

Η μεταφορά μάθησης είναι μια καθοριστική τεχνική στη μηχανική μάθηση, που προσφέρει λύσεις όταν δεν υπάρχουν αρκετά δεδομένα για την πλήρη εκπαίδευση ενός μοντέλου από την αρχή. Αυτή η προσέγγιση επιτρέπει σε μοντέλα που έχουν εκπαιδευτεί σε μεγάλα σύνολα δεδομένων να προσαρμόζονται σε νέα, μικρότερα σύνολα δεδομένων με ελάχιστη εκπαίδευση. Με τη μεταφορά γνώσης από παρόμοιες διεργασίες, βελτιώνονται οι επιδόσεις του μοντέλου σε νέες διεργασίες, μειώνοντας τον χρόνο και τους πόρους που απαιτούνται για την εκπαίδευση.

Το πρώτο σημαντικό άρθρο που αναφέρθηκε στη μεταφορά μάθησης είναι το "A Survey on Transfer Learning" των Pan και Yang (2009) [15], το οποίο κατηγοριοποίησε τις διάφορες μορφές μεταφοράς μάθησης, όπως την επαγωγική, την μη-επαγωγική και την ενδοτομειακή μεταφορά. Σημείωσε πώς τα μοντέλα μπορούν να εκπαιδευτούν σε ένα πεδίο με άφθονα δεδομένα και να προσαρμοστούν σε άλλα πεδία με διαφορετικά χαρακτηριστικά, επιτρέποντας τη γενίκευση των χαρακτηριστικών τους.

Ένα σημαντικό επόμενο βήμα ήταν το άρθρο "Learning Transferable Features with Deep Adaptation Networks" από τον Long κ.α. (2015) [16], το οποίο πρότεινε τα Deep Adaptation Networks (DANs). Το DAN προσφέρει μια μέθοδο για τη μείωση της απόκλισης μεταξύ δεδομένων πηγής και στόχου, προσαρμόζοντας το μοντέλο σε κοινά χαρακτηριστικά που είναι "μεταφερόμενα" μεταξύ των δύο τομέων. Αυτό επέτρεψε την εφαρμογή σε διάφορες εργασίες ταξινόμησης, επιτρέποντας στο μοντέλο να επιτύχει υψηλή απόδοση σε τομείς όπου τα δεδομένα του στόχου δεν είχαν τη μορφή που απαιτούσε η αρχική εκπαίδευση.

Στη συνέχεια, το άρθρο "*Attention Is All You Need*" των Vaswani *et al.* (2017) [17] παρουσίασε τους Transformers, ένα από τα πιο ισχυρά μοντέλα για τη μεταφορά μάθησης. Με τους Transformers, η ιδέα της προεκπαίδευσης σε τεράστια σύνολα δεδομένων βελτιώθηκε, και μετά έγινε fine-tuning για συγκεκριμένες διεργασίες. Το fine-tuning αναφέρεται στη διαδικασία όπου οι παράμετροι του μοντέλου προσαρμόζονται περαιτέρω χρησιμοποιώντας επισημασμένα δεδομένα για μια νέα εργασία. Αυτή η προσέγγιση έφερε επαναστατικές επιδόσεις στη φυσική γλώσσα, στη μετάφραση και σε άλλες γλωσσικές εργασίες, δείχνοντας πώς η μεταφορά μάθησης μέσω του fine-tuning μπορεί να βελτιώσει την απόδοση σε σύνθετα προβλήματα.

0.3 Στόχοι και Συνεισφορές

Η παρούσα διπλωματική εργασία επικεντρώνεται στην εφαρμογή της μεθόδου της Αυτο-Εποπτευόμενης Μάθησης με τη χρήση έμμεσων διεργασιών για την ενίσχυση της απόδοσης σε προβλήματα ταξινόμησης εικόνων. Η κεντρική ιδέα βασίζεται στην προεκπαίδευση μοντέλων σε διεργασίες που δεν απαιτούν επισημασμένα δεδομένα, όπως η πρόβλεψη περιστροφής εικόνων, η χρωματοποίηση εικόνας και η επιδιόρθωση/συμπλήρωση εικόνας. Τα προεκπαιδευμένα αυτά βάρη στη συνέχεια μεταφέρονται σε ένα downstream task, όπου το μοντέλο αξιολογείται σε προβλήματα ταξινόμησης εικόνων. Σε αυτή την εργασία, τα τρία proxy tasks εκπαιδεύτηκαν στο CIFAR-100 χωρίς την χρήση των ετικετών των εικόνων, δηλαδή ως ένα unlabeled dataset, και τα προεκπαιδευμένα βάρη τους δοκιμάστηκαν σε δύο datasets ταξινόμησης εικόνων, το CIFAR-10 και το CIFAR-100, τόσο με πλήρες fine-tuning ολόκληρου του μοντέλου όσο και με fine-tuning μόνο του τελευταίου επιπέδου. Η εργασία αυτή επιδιώκει τη δίκαιη σύγκριση των τριών proxy tasks, χρησιμοποιώντας ως βάση την ίδια αρχιτεκτονική μοντέλου (ResNet-50) για τα proxy tasks και τις ίδιες υπερπαραμέτρους για την εκπαίδευση στο downstream task.

0.3.1 Ερευνητικά Ερωτήματα, Στόχοι και Δομή της Εργασίας

Τα ερευνητικά ερωτήματα που τέθηκαν σε αυτή την εργασία επικεντρώνονται στη μελέτη της αποτελεσματικότητας των proxy tasks και της εφαρμογής τους στο downstream task της ταξινόμησης εικόνων, καθώς και στη συμβατότητα των συνόλων δεδομένων που χρησιμοποιήθηκαν. Συγκεκριμένα, η εργασία εξετάζει τα παρακάτω ερωτήματα:

1. Ποιο από τα τρία proxy tasks (Image Rotation Prediction, Image Colorization, Image Inpainting) παράγει τα πιο αποτελεσματικά προεκπαιδευμένα βάρη για τη βελτίωση της απόδοσης ενός μοντέλου ταξινόμησης εικόνων;
2. Πώς συγκρίνεται η απόδοση των proxy tasks σε διαφορετικά σύνολα δεδομένων ταξινόμησης, όπως το CIFAR-10 και το CIFAR-100;
3. Ποια είναι η επίδραση της εκπαίδευσης ολόκληρου του μοντέλου (fine-tuning all layers) σε σχέση με την εκπαίδευση μόνο του τελευταίου επιπέδου (fine-tuning last layer) στην απόδοση του μοντέλου, τόσο στο CIFAR-10 όσο και στο CIFAR-100;
4. Σε ποιο βαθμό τα proxy tasks είναι συμβατά με το downstream task της ταξινόμησης εικόνων και ποιο task προσφέρει την καλύτερη γενίκευση των χαρακτηριστικών;
5. Πώς επηρεάζει η συμβατότητα μεταξύ των συνόλων δεδομένων (CIFAR-100 ως σύνολο προεκπαίδευσης και CIFAR-10/CIFAR-100 ως σύνολα ταξινόμησης) την απόδοση της μεταφοράς μάθησης, και κατά πόσο επηρεάζεται η απόδοση όταν το μοντέλο μεταφέρεται από το ίδιο σύνολο δεδομένων (CIFAR-100) σε διαφορετικό (CIFAR-10);

Ο στόχος της παρούσας μελέτης είναι να πραγματοποιηθεί μια δίκαιη σύγκριση των τριών proxy tasks σε ένα ελεγχόμενο περιβάλλον, χρησιμοποιώντας ως βάση την ίδια αρχιτεκτονική (ResNet-50) και τις ίδιες υπερπαραμέτρους κατά την εκπαίδευση του downstream task. Μέσω αυτής της σύγκρισης, θα αναδειχθεί ποιο proxy task είναι πιο αποδοτικό στη μεταφορά μάθησης για το downstream task της ταξινόμησης εικόνων, ενώ θα εξεταστεί επίσης η επίδραση της συμβατότητας των συνόλων δεδομένων και η συνάφεια των χαρακτηριστικών που έχουν μάθει τα proxy tasks, τόσο σε παρόμοια όσο και σε διαφορετικά datasets.

Η εργασία αυτή χωρίζεται σε δύο κύρια μέρη: το Θεωρητικό και το Πειραματικό Μέρος. Στο Θεωρητικό Μέρος (Κεφάλαια 1 και 2), παρουσιάζονται βασικές έννοιες που σχετίζονται με τα Νευρωνικά Δίκτυα και

την Αυτο-Εποπτευόμενη Μάθηση. Στο πρώτο κεφάλαιο αναλύονται οι θεμελιώδεις αρχές των Τεχνητών Νευρωνικών Δικτύων, η λειτουργία τους και οι διάφοροι τύποι τους, ενώ γίνεται ειδική αναφορά στα Συνελικτικά Νευρωνικά Δίκτυα. Στο δεύτερο κεφάλαιο εξετάζεται η θεωρία της Αυτο-Εποπτευόμενης Μάθησης και τα είδη των έμμεσων διεργασιών που χρησιμοποιούνται για την εκπαίδευση μοντέλων χωρίς επισημασμένα δεδομένα. Παρουσιάζονται επίσης οι προκλήσεις και οι περιορισμοί της μεθόδου, καθώς και οι εφαρμογές της σε διάφορους τομείς της Υπολογιστικής Όρασης.

Το Πειραματικό Μέρος (Κεφάλαιο 3) εστιάζει στην περιγραφή της πειραματικής διαδικασίας και την αξιολόγηση των αποτελεσμάτων. Στην αρχή του κεφαλαίου γίνεται ανάλυση της μεθοδολογίας που ακολουθήθηκε, καθώς και των ερευνητικών ερωτημάτων και προσδοκιών της μελέτης. Στη συνέχεια παρουσιάζεται ο τρόπος υλοποίησης των τριών έμμεσων διεργασιών (Image Rotation Prediction, Image Colorization, και Image Inpainting), η διαδικασία της μεταφοράς μάθησης και του fine-tuning των μοντέλων και τα σύνολα δεδομένων που χρησιμοποιήθηκαν. Τα αποτελέσματα από τα πειράματα παρουσιάζονται αναλυτικά και σχολιάζονται, με ιδιαίτερη έμφαση στην αποτελεσματικότητα των προεκπαίδευμένων βαρών στις έμμεσες διεργασίες, σε σχέση με τα τυχαία βάρη, στο task της ταξινόμησης εικόνων. Τέλος, ο σχολιασμός των αποτελεσμάτων και τα συμπεράσματα της εργασίας συνοψίζονται στο τελευταίο τμήμα του κεφαλαίου.

0.3.2 Προοπτικές Μελλοντικής Έρευνας και Εφαρμογές

Προοπτικές Μελλοντικής Έρευνας:

Οι προοπτικές για μελλοντική έρευνα στον τομέα της Αυτο-Εποπτευόμενης Μάθησης με τη χρήση έμμεσων διεργασιών είναι ιδιαίτερα σημαντικές και προσφέρουν πολλά υποσχόμενες κατευθύνσεις για περαιτέρω μελέτη και ανάπτυξη. Μία από τις βασικές προοπτικές είναι η διερεύνηση της χρήσης πιο σύνθετων και βαθιών αρχιτεκτονικών νευρωνικών δικτύων, όπως τα μεγαλύτερα μοντέλα, τα οποία έχουν τη δυνατότητα να καταγράφουν ακόμα πιο λεπτομερή χαρακτηριστικά από τα δεδομένα. Οι σύγχρονες προσεγγίσεις, που περιλαμβάνουν μηχανισμούς προσοχής (*attention mechanisms*), όπως τα Transformer-based μοντέλα [14], θα μπορούσαν να ενσωματωθούν για να επιτρέψουν στο μοντέλο να εστιάζει σε πιο σημαντικές περιοχές των εικόνων και να αποδίδει καλύτερα σε πιο σύνθετα tasks.

Μια άλλη σημαντική κατεύθυνση είναι η επέκταση της μελέτης σε μεγαλύτερα και πιο ετερογενή σύνολα δεδομένων, τα οποία περιλαμβάνουν πολυπλοκότερες και πιο ποικίλες κατηγορίες, όπως το ImageNet ή και δεδομένα από τον πραγματικό κόσμο που δεν είναι ιδανικά κατηγοριοποιημένα. Αυτό θα επιτρέψει την αξιολόγηση των proxy tasks σε συνθήκες που προσομοιώνουν καλύτερα τα πραγματικά προβλήματα και θα εξεταστεί κατά πόσο τα χαρακτηριστικά που μαθαίνονται από τα proxy tasks είναι πραγματικά γενικεύσιμα σε πιο ποικίλα περιβάλλοντα.

Μια ακόμα ενδιαφέρουσα προοπτική θα μπορούσε να είναι η καλύτερη προσαρμογή των proxy tasks για πιο απαιτητικά και σύνθετα tasks, όπως η πολυδιάστατη ανάλυση δεδομένων, η επεξεργασία βίντεο ή δεδομένων από αισθητήρες. Για παράδειγμα, η ενσωμάτωση χωροχρονικών πληροφοριών από βίντεο, μέσω proxy tasks που επιτρέπουν στο μοντέλο να μαθαίνει όχι μόνο από τις χωρικές αλλά και από τις χρονικές σχέσεις των δεδομένων, θα μπορούσε να βελτιώσει την απόδοση των μοντέλων σε εφαρμογές που απαιτούν ανάλυση σειρών εικόνων ή συνεχούς ροής δεδομένων. Η ενοποίηση τέτοιων μεθόδων θα προετοιμάσει τα μοντέλα για την αντιμετώπιση πολυδιάστατων και πιο απαιτητικών περιβαλλόντων.

Επιπροσθέτως, μελλοντικές έρευνες θα μπορούσαν να επικεντρωθούν στη βελτιστοποίηση της διαδικασίας fine-tuning. Οι τρέχουσες μέθοδοι συχνά επικεντρώνονται είτε στην προσαρμογή ολόκληρου του μοντέλου είτε μόνο του τελευταίου επιπέδου, αλλά θα μπορούσε να διερευνηθεί η χρήση μεθόδων "progressive fine-tuning", όπου σταδιακά περισσότερα επίπεδα του μοντέλου θα ξεπαγώνονται κατά τη διαδικασία εκπαίδευσης, επιτρέποντας τη βελτιστοποίηση του μοντέλου με πιο στοχευμένο τρόπο. Αυτό μπορεί να οδηγήσει σε μεγαλύτερη αποδοτικότητα και εξοικονόμηση πόρων κατά την εκπαίδευση, ιδίως όταν τα δεδομένα του downstream task είναι περιορισμένα.

Εφαρμογές:

Η Αυτο-Εποπτευόμενη Μάθηση με τη χρήση έμμεσων διεργασιών έχουν ήδη αποδείξει την αξία τους σε πολλές εφαρμογές της υπολογιστικής όρασης, ιδίως σε περιπτώσεις όπου τα επισημασμένα δεδομένα είναι περιορισμένα ή δύσκολα να αποκτηθούν. Ένα από τα πιο σημαντικά πεδία εφαρμογής είναι η ανάλυση ιατρικών εικόνων, όπου οι αυτοεπιβλεπόμενες τεχνικές μπορούν να υποστηρίζουν την αυτόματη ανάλυση και διάγνωση, βοηθώντας στην αναγνώριση ασθενειών ή ανωμαλιών με ελάχιστη ανθρώπινη παρέμβαση. Συστήματα ανίχνευσης αντικειμένων σε αυτόνομα οχήματα αποτελούν ένα άλλο σημαντικό πεδίο εφαρμογής, όπου η ικανότητα των μοντέλων να μαθαίνουν από μεγάλες ποσότητες μη επισημασμένων δεδομένων μπορεί να ενισχύσει την ακρίβεια και την ασφάλεια των συστημάτων.

Επιπλέον, τα proxy tasks μπορούν να εφαρμοστούν σε εφαρμογές που αφορούν την ασφάλεια, όπως η αναγνώριση προσώπου σε διάφορες συνθήκες φωτισμού ή η παρακολούθηση ατόμων σε πολυπληθείς δημόσιους χώρους. Η ικανότητα των μοντέλων να μαθαίνουν από σύνθετα και πολυδιάστατα δεδομένα τα καθιστά ιδιαίτερα χρήσιμα σε περιπτώσεις όπου η ανάγκη για γρήγορη και αξιόπιστη ανάλυση είναι κρίσιμη.

Με την πρόοδο της τεχνολογίας και την αύξηση των διαθέσιμων δεδομένων, η Αυτο-Εποπτευόμενη Μάθηση μέσω proxy tasks αναμένεται να βρει ακόμα περισσότερες εφαρμογές σε διάφορους τομείς, από την επιστήμη των δεδομένων και την τεχνητή νοημοσύνη μέχρι τη βιομηχανία και την έρευνα.

I

Θεωρητικό Μέρος

Κεφάλαιο 1^ο

Νευρωνικά Δίκτυα

1.1 Εισαγωγή στα Τεχνητά Νευρωνικά Δίκτυα (ANNs)

Η ιδέα των Τεχνητών Νευρωνικών Δικτύων (ANNs) προτάθηκε αρχικά από τους Warren McCulloch και Walter Pitts το 1943 [18]. Στο θεμελιώδες έργο τους "A Logical Calculus of the Ideas Immanent in Nervous Activity", περιέγραψαν ένα υπολογιστικό μοντέλο που αποτελούνταν από νευρώνες, οι οποίοι μπορούσαν να εκτελούν λογικές πράξεις μέσω συνδέσεων μεταξύ τους. Αυτή η προσέγγιση έθεσε τις βάσεις για την ανάπτυξη των τεχνητών νευρωνικών δικτύων όπως τα γνωρίζουμε σήμερα.

Στη συνέχεια, ένας από τους πρώτους επίσημους και ευρέως αποδεκτούς ορισμούς των Τεχνητών Νευρωνικών Δικτύων δόθηκε από τον Hecht-Nielsen το 1989:

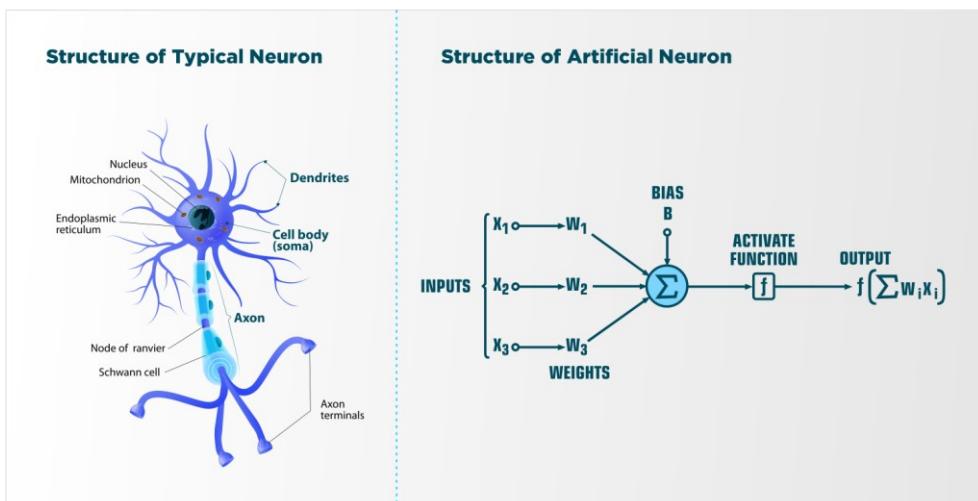
"Ενα τεχνητό νευρωνικό δίκτυο είναι μια παράλληλη διάταξη, που επεξεργάζεται διανεμημένες πληροφορίες, αποτελείται από μονάδες επεξεργασίας (οι οποίες μπορούν να κατέχουν μια τοπική μνήμη και να διεκπεραιώνουν λειτουργίες τοπικής επεξεργασίας πληροφορίας) και είναι διασυνδεδεμένο μέσω καναλιών πολλών διευθύνσεων, τα οποία καλούνται και συνδέσεις. Κάθε μονάδα επεξεργασίας έχει μια μοναδική έξοδο που κατευθύνεται σε όσες συνδέσεις είναι επιθυμητό. Κάθε σύνδεση μεταφέρει το ίδιο σήμα - το σήμα εξόδου της μονάδας επεξεργασίας. Το σήμα αυτό μπορεί να είναι οποιουδήποτε επιθυμητού μαθηματικού τύπου. Η επεξεργασία πληροφορίας η οποία φτάνει σε κάθε μονάδα επεξεργασίας μπορεί να οριστεί αυθαίρετα με τον περιορισμό όμως ότι αυτή θα είναι εντελώς τοπική. Αυτό σημαίνει ότι πρέπει να εξαρτάται από τις τρέχουσες τιμές των σημάτων εισόδου που φτάνουν στην μονάδα επεξεργασίας μέσω των συνδέσεων και με τιμές που αποθηκεύονται στην τοπική μνήμη της μονάδας."

Τα Τεχνητά Νευρωνικά Δίκτυα (ANNs) αποτελούν μια από τις πιο σημαντικές τεχνολογικές εξελίξεις στον τομέα της τεχνητής νοημοσύνης και της μηχανικής μάθησης. Για να κατανοήσουμε τη λειτουργία και τις εφαρμογές τους, είναι απαραίτητο να εξετάσουμε τα βασικά χαρακτηριστικά που τα διακρίνουν. Αυτά τα χαρακτηριστικά καθορίζουν την ικανότητα των ANNs να μαθαίνουν από δεδομένα, να αναγνωρίζουν πρότυπα και να προσαρμόζονται σε νέες πληροφορίες, καθιστώντας τα απαραίτητα εργαλεία για την επίλυση σύνθετων προβλημάτων σε πολλούς τομείς της επιστήμης και της τεχνολογίας. Στην συνέχεια, θα παρουσιαστούν τα κύρια χαρακτηριστικά που καθορίζουν την λειτουργία και την αποδοτικότητα των Τεχνητών Νευρωνικών Δικτύων:

- Δομή Δικτύου (Network Structure):** Τα Τεχνητά Νευρωνικά Δίκτυα αποτελούνται από πολλαπλά επίπεδα νευρώνων (neurons), τα οποία είναι διασυνδεδεμένα. Η βασική δομή ενός ANN περιλαμβάνει τρία κύρια επίπεδα: το επίπεδο εισόδου (input layer), το (ή τα) κρυφό(ά) επίπεδο(α) (hidden layer(s)) και το επίπεδο εξόδου (output layer). Κάθε νευρώνας σε ένα επίπεδο είναι συνδεδεμένος με κάθε νευρώνα στο επόμενο επίπεδο μέσω συνδέσεων που έχουν βάρη (weights).
- Νευρώνες και Συνδέσεις (Neurons and Connections):** Οι νευρώνες είναι οι θεμελιώδεις μονάδες επεξεργασίας (processing units) στα ANNs. Κάθε νευρώνας λαμβάνει σήματα εισόδου (input signals) από άλλους νευρώνες, πολλαπλασιάζει τα εισερχόμενα σήματα με τα αντίστοιχα βάρη, και στη συνέχεια εφαρμόζει μια συνάρτηση ενεργοποίησης (activation function) στο άθροισμα αυτών των σημάτων για να παράγει ένα σήμα εξόδου (output signal). Η σύνδεση (connection) μεταξύ των

νευρώνων καθορίζεται από τα βάρη, τα οποία μαθαίνονται και προσαρμόζονται κατά τη διαδικασία εκπαίδευσης (training) του δικτύου.

3. **Συνάρτηση Ενεργοποίησης (Activation Function):** Οι συναρτήσεις ενεργοποίησης καθορίζουν πώς θα μετατραπεί το αθροισμένο σήμα εισόδου σε έξοδο από έναν νευρώνα. Συχνά χρησιμοποιούμενες συναρτήσεις ενεργοποίησης είναι η sigmoid, η ReLU (Rectified Linear Unit), και η tanh (tangent hyperbolic). Αυτές οι συναρτήσεις επιτρέπουν στο δίκτυο να εισάγει μη γραμμικότητα (non-linearity) στο μοντέλο, το οποίο είναι κρίσιμο για την εκμάθηση πολύπλοκων σχέσεων στα δεδομένα.
4. **Μάθηση και Προσαρμοστικότητα (Learning and Adaptability):** Τα ANNs μαθαίνουν προσαρμόζοντας τα βάρη των συνδέσεων μεταξύ των νευρώνων. Κατά τη διαδικασία μάθησης (learning process), το δίκτυο συγκρίνει την προβλεπόμενη έξοδο (predicted output) με την πραγματική έξοδο (actual output) και υπολογίζει το σφάλμα (error) μέσω μιας συνάρτησης απώλειας (loss function). Στη συνέχεια, το δίκτυο χρησιμοποιεί αλγορίθμους όπως η οπισθοδιάδοση (backpropagation) για να προσαρμόσει τα βάρη έτσι ώστε να μειωθεί το σφάλμα σε μελλοντικές προβλέψεις.
5. **Προσαρμοστικότητα σε Νέα Δεδομένα (Adaptability to New Data):** Τα ANNs είναι δυναμικά μοντέλα που μπορούν να προσαρμόζονται σε νέα δεδομένα. Με κάθε νέο παράδειγμα που παρουσιάζεται στο δίκτυο, τα βάρη των συνδέσεων ενημερώνονται, επιτρέποντας στο δίκτυο να βελτιώνει την ικανότητά του να κάνει ακριβείς προβλέψεις (accurate predictions). Αυτό καθιστά τα ANNs ιδανικά για προβλήματα όπου η φύση των δεδομένων αλλάζει με την πάροδο του χρόνου.
6. **Γενικευτική Ικανότητα (Generalization Ability):** Ένα καλά εκπαιδευμένο ANN έχει τη δυνατότητα να γενικεύει (generalize), δηλαδή να αποδίδει καλά όχι μόνο στα δεδομένα εκπαίδευσης (training data) αλλά και σε νέα, αόρατα δεδομένα (unseen data). Αυτό το χαρακτηριστικό είναι κρίσιμο για την επιτυχία των ANNs σε πραγματικές εφαρμογές.
7. **Παράλληλη Επεξεργασία (Parallel Processing):** Τα ANNs λειτουργούν με παράλληλη επεξεργασία πληροφοριών, πράγμα που σημαίνει ότι πολλοί νευρώνες μπορούν να εκτελούν υπολογισμούς ταυτόχρονα. Αυτή η παράλληλη επεξεργασία συμβάλλει στη μεγάλη υπολογιστική ισχύ (computational power) και ταχύτητα (speed) των ANNs, κάνοντάς τα ιδανικά για εφαρμογές που απαιτούν υψηλές αποδόσεις.



Εικόνα 1.1.1 - 1: Σύγκριση της δομής ενός τυπικού βιολογικού νευρώνα με έναν τεχνητό νευρώνα.

(Πηγή: NIIT BIRGUNJ)

1.1.2 Λειτουργία, Τύποι Αρχιτεκτονικών και Εφαρμογές των ANNs

Αφού ορίσαμε την έννοια των Τεχνητών Νευρωνικών Δικτύων, ας εξετάσουμε την αρχιτεκτονική και τη λειτουργία τους. Τα Τεχνητά Νευρωνικά Δίκτυα αποτελούνται από διάφορα είδη δομών, οι οποίες είναι σχεδιασμένες για να ανταποκρίνονται σε συγκεκριμένες απαιτήσεις και εφαρμογές. Ένα βασικό στοιχείο

της αρχιτεκτονικής των ANNs είναι το πώς οι νευρώνες συνδέονται και συνεργάζονται για να επεξεργαστούν τα δεδομένα και να εκτελέσουν λειτουργίες που μπορούν να οδηγήσουν σε χρήσιμα αποτελέσματα.

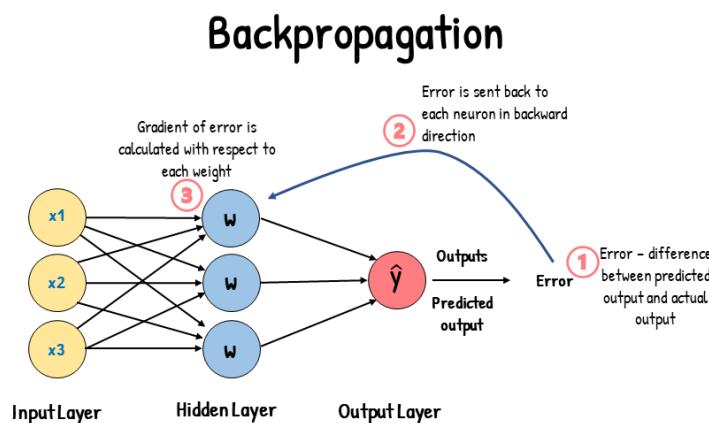
Βάρη (Weights): Τα βάρη στα νευρωνικά δίκτυα είναι κρίσιμοι παράγοντες που καθορίζουν πώς οι εισερχόμενες πληροφορίες επηρεάζουν την τελική έξοδο του δικτύου. Κάθε σύνδεση μεταξύ νευρώνων έχει ένα συνδεδεμένο βάρος, το οποίο προσαρμόζει το εισερχόμενο σήμα από τον προηγούμενο νευρώνα. Ουσιαστικά, τα βάρη πολλαπλασάζονται με τα σήματα που λαμβάνουν οι νευρώνες, ενισχύοντας ή αποδυναμώνοντας αυτά τα σήματα πριν τα περάσουν στον επόμενο νευρώνα. Σταδιακά, μέσω της διαδικασίας εκπαίδευσης, τα βάρη προσαρμόζονται ώστε το δίκτυο να μάθει να προβλέπει σωστά ή να ταξινομεί τα δεδομένα. Η προσαρμογή των βαρών γίνεται με στόχο τη μείωση της συνάρτησης απώλειας, δηλαδή της απόκλισης μεταξύ των προβλέψεων του δικτύου και των πραγματικών αποτελεσμάτων.

Backpropagation: Το Backpropagation (αναδρομική διάδοση σφάλματος) είναι η κύρια μέθοδος που χρησιμοποιείται για την εκπαίδευση νευρωνικών δικτύων. Αυτή η μέθοδος επιτρέπει την αποτελεσματική προσαρμογή των βαρών του δικτύου με βάση τα σφάλματα που προκύπτουν από τις προβλέψεις του.

Η διαδικασία λειτουργεί ως εξής:

1. **Προώθηση (Forward Pass):** Το δίκτυο λαμβάνει ένα σύνολο εισόδων, και αυτές οι είσοδοι πολλαπλασάζονται με τα βάρη των συνδέσεων και περνούν μέσω των νευρώνων του δικτύου μέχρι να φτάσουν στο επίπεδο εξόδου, όπου παράγεται η τελική πρόβλεψη.
2. **Υπολογισμός Σφάλματος (Error Calculation):** Η συνάρτηση απώλειας υπολογίζει τη διαφορά μεταξύ της προβλεπόμενης εξόδου και της πραγματικής τιμής. Αυτό το σφάλμα είναι αυτό που το δίκτυο καλείται να ελαχιστοποιήσει.
3. **Αναδρομική Διάδοση (Backward Pass):** Το σφάλμα στη συνέχεια διαδίδεται προς τα πίσω μέσω του δικτύου, από το επίπεδο εξόδου προς τα προηγούμενα επίπεδα. Κατά τη διάρκεια αυτής της διαδικασίας, το σφάλμα κατανεμείται σε κάθε νευρώνα και κάθε βάρος, καθορίζοντας πόσο κάθε βάρος συνέβαλε στο συνολικό σφάλμα. Αυτό επιτυγχάνεται μέσω της χρήσης μερικών παραγώγων, που υπολογίζουν την επίδραση κάθε βάρους στη συνάρτηση απώλειας, επιτρέποντας την προσαρμογή των βαρών με ακρίβεια.
4. **Ενημέρωση Βαρών (Weight Update):** Τα βάρη ενημερώνονται με βάση το σφάλμα που υπολογίστηκε κατά το στάδιο της αναδρομικής διάδοσης. Αυτή η ενημέρωση γίνεται με τη βοήθεια του αλγορίθμου Gradient Descent, που προσαρμόζει τα βάρη ώστε να μειώσει το σφάλμα στις επόμενες προβλέψεις του δικτύου.

Η διαδικασία αυτή επαναλαμβάνεται πολλές φορές κατά τη διάρκεια της εκπαίδευσης, μέχρι το δίκτυο να πετύχει το επιθυμητό επίπεδο ακρίβειας.

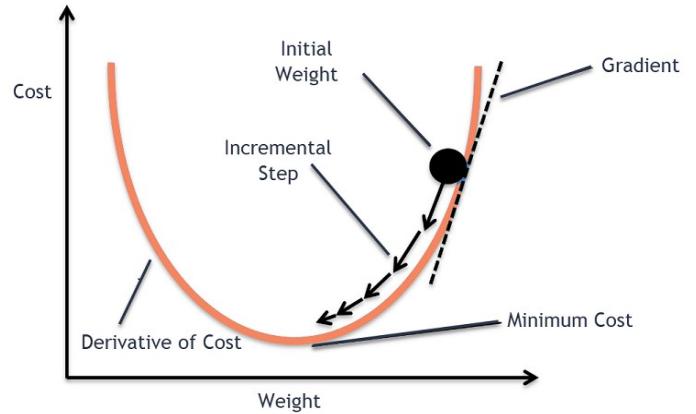


Εικόνα 1.1.2 - 1: Η διαδικασία της Αναδρομικής Διάδοσης (Backpropagation) σε ένα νευρωνικό

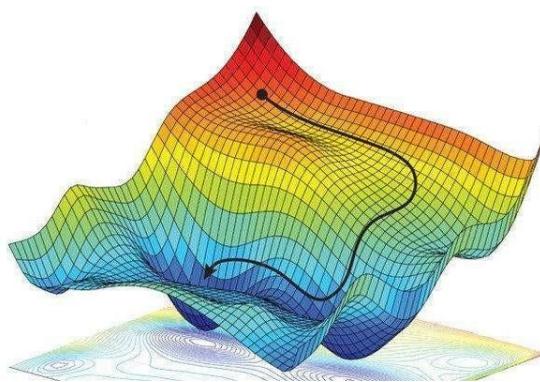
δίκτυο, όπου το σφάλμα υπολογίζεται από τη διαφορά μεταξύ της προβλεπόμενης εξόδου και της πραγματικής τιμής. Στη συνέχεια, το σφάλμα αυτό διαδίδεται προς τα πίσω μέσα από το δίκτυο, ενημερώνοντας τα βάρη κάθε συνδέσμου με βάση την κλίση του σφάλματος. (Πηγή: ResearchGate)

Αλγόριθμοι Βελτιστοποίησης (Optimization Algorithms): Για την εκπαίδευση αυτών των δικτύων, χρησιμοποιούνται διάφοροι αλγόριθμοι, με κυριότερο τον Gradient Descent και την παραλλαγή του, τον Stochastic Gradient Descent (SGD). Πλέον χρησιμοποιούνται και πιο προηγμένοι αλγόριθμοι όπως ο Adam(Adaptive Moment Estimation) [19]. Αυτοί οι αλγόριθμοι επιτρέπουν την ενημέρωση των βαρών των συνδέσεων του δικτύου προς την κατεύθυνση του ελάχιστου της συνάρτησης απώλειας, βελτιστοποιώντας έτσι το μοντέλο.

- Ο αλγόριθμος Gradient Descent είναι μια μέθοδος βελτιστοποίησης που χρησιμοποιείται για την ελάχιστοποίηση της συνάρτησης απώλειας σε ένα νευρωνικό δίκτυο. Λειτουργεί υπολογίζοντας το gradient της συνάρτησης απώλειας ως προς τα βάρη του δικτύου και στη συνέχεια ενημερώνοντας τα βάρη προς την αντίθετη κατεύθυνση του gradient, με στόχο τη μείωση της απώλειας. Αυτή η διαδικασία επαναλαμβάνεται έως ότου το δίκτυο συγκλίνει σε ένα τοπικό ή παγκόσμιο ελάχιστο.
- Ο Stochastic Gradient Descent είναι μια παραλλαγή του Gradient Descent, όπου αντί να χρησιμοποιείται ολόκληρο το σύνολο δεδομένων για τον υπολογισμό του gradient, κάθε ενημέρωση των βαρών γίνεται χρησιμοποιώντας μόνο ένα δείγμα δεδομένων ή ένα μικρό υποσύνολο (mini-batch). Αυτό κάνει τον SGD ταχύτερο και πιο αποδοτικό, ειδικά σε μεγάλα σύνολα δεδομένων, αν και η πορεία προς τη σύγκλιση μπορεί να είναι πιο θορυβώδης λόγω της τυχαιότητας που εισάγεται από τη χρήση μεμονωμένων δειγμάτων.
- Ο αλγόριθμος Adam (Adaptive Moment Estimation) είναι ένας προηγμένος αλγόριθμος βελτιστοποίησης που προσαρμόζει δυναμικά τον ρυθμό μάθησης για κάθε παράμετρο του νευρωνικού δικτύου. Συνδυάζει τα πλεονεκτήματα του SGD και του RMSprop, χρησιμοποιώντας τόσο τη μέση τιμή όσο και τη διακύμανση των παραγώγων. Αυτό τον καθιστά ιδιαίτερα αποτελεσματικό για εκπαίδευση σε μεγάλα και θορυβώδη δεδομένα, προσφέροντας ταχύτερη και πιο σταθερή σύγκλιση σε σχέση με άλλους παραδοσιακούς αλγόριθμους βελτιστοποίησης.



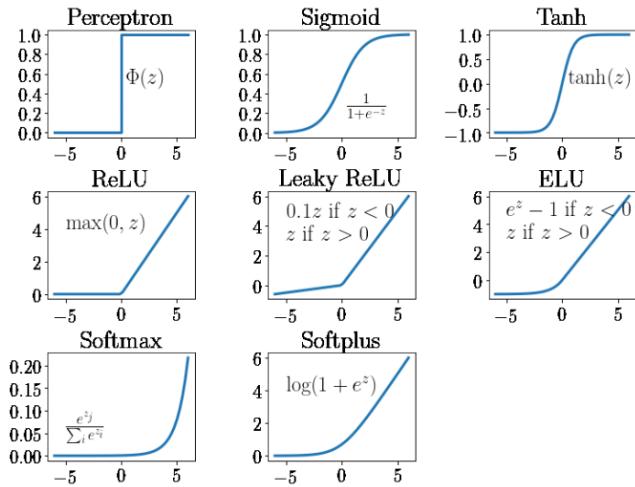
Εικόνα 1.1.2 - 2: Διάγραμμα που απεικονίζει τη διαδικασία εύρεσης του ελάχιστου κόστους μέσω της βελτιστοποίησης των βαρών, χρησιμοποιώντας τη μέθοδο του Gradient Descent.
(Πηγή: Analytics Vidhya)



Εικόνα 1.1.2 - 3: Οπτικοποίηση της Διαδικασίας Gradient Descent. Πορεία ενός αλγορίθμου Gradient Descent που προσπαθεί να βρει το ελάχιστο μιας συνάρτησης απώλειας, μεταβαίνοντας από υψηλότερα επίπεδα της συνάρτησης προς το χαμηλότερο σημείο της, όπου βρίσκεται το καθολικό ελάχιστο.(Πηγή: ResearchGate)

Συναρτήσεις Ενεργοποίησης και Συναρτήσεις Εξόδου (Activation Functions and Output Functions): Στα νευρωνικά δίκτυα, οι συναρτήσεις ενεργοποίησης και οι συναρτήσεις εξόδου διαδραματίζουν κρίσιμο ρόλο στη διαδικασία της μάθησης και της παραγωγής προβλέψεων. Οι συναρτήσεις ενεργοποίησης εφαρμόζονται στα κρυφά στρώματα του δικτύου, εισάγοντας μη γραμμικότητα και επιτρέποντας στο δίκτυο να μάθει σύνθετα μοτίβα και σχέσεις στα δεδομένα. Από την άλλη πλευρά, οι συναρτήσεις εξόδου εφαρμόζονται στο τελικό επίτεδο του δικτύου, μετατρέποντας τις τελικές ενεργοποιήσεις σε προβλέψεις ή κατηγορίες, ανάλογα με το εκάστοτε πρόβλημα. Θα εξετάσουμε, παρακάτω, τις κυριότερες συναρτήσεις ενεργοποίησης και εξόδου που χρησιμοποιούνται συνήθως στα τεχνητά νευρωνικά δίκτυα, αναλύοντας τη λειτουργία και τη σημασία τους.

- **Linear:** Χρησιμοποιείται κυρίως σε προβλήματα παλινδρόμησης, όπου η έξοδος μπορεί να είναι οποιαδήποτε πραγματική τιμή. Η Linear συνάρτηση εξόδου δεν τροποποιεί το σήμα που εισέρχεται, επιτρέποντας τη δημιουργία συνεχών προβλέψεων.
- **Sigmoid:** Είναι μία ομαλή καμπύλη που συμπιέζει την είσοδο σε μια τιμή μεταξύ 0 και 1. Χρησιμοποιείται τόσο ως συνάρτηση ενεργοποίησης όσο και ως συνάρτηση εξόδου, ιδιαίτερα σε δυαδικά προβλήματα ταξινόμησης, όπου επιστρέφει την πιθανότητα το δείγμα να ανήκει σε μια συγκεκριμένη κατηγορία.
- **Tanh (Hyperbolic Tangent):** Η Tanh είναι παρόμοια με τη Sigmoid, αλλά συμπιέζει την είσοδο σε μια τιμή μεταξύ -1 και 1. Αυτή η συνάρτηση είναι συχνά προτιμότερη από τη Sigmoid στα κρυφά στρώματα, επειδή οι έξοδοι της είναι κεντραρισμένες γύρω από το μηδέν, επιτρέποντας καλύτερη σύγκλιση κατά την εκπαίδευση.
- **ReLU (Rectified Linear Unit):** Αποτελεί μια από τις πιο δημοφιλείς συναρτήσεις ενεργοποίησης λόγω της απλότητάς της και της αποδοτικότητάς της. Επιστρέφει το μέγιστο μεταξύ μηδενός και της εισόδου, δηλαδή ενεργοποιεί μόνο τις θετικές τιμές και μετατρέπει όλες τις αρνητικές τιμές σε μηδέν. Χρησιμοποιείται συχνά στα κρυφά στρώματα για να εισάγει μη γραμμικότητα στο δίκτυο.
- **Leaky ReLU:** Η Leaky ReLU είναι μια παραλλαγή της ReLU που επιτρέπει μια μικρή κλίση για τις αρνητικές τιμές αντί να τις θέτει απλώς στο μηδέν. Αυτό σημαίνει ότι, ενώ η ReLU επιστρέφει μηδέν για όλες τις αρνητικές εισόδους, η Leaky ReLU επιστρέφει μια μικρή, αρνητική τιμή (συνήθως ένα κλάσμα της εισόδου). Αυτό βοηθά στην αντιμετώπιση του προβλήματος της «εξαφάνισης των νευρώνων» (dying ReLU problem), όπου κάποιοι νευρώνες μπορεί να μην ενεργοποιούνται ποτέ.
- **Softmax:** Είναι μια συνάρτηση εξόδου που χρησιμοποιείται κυρίως σε προβλήματα ταξινόμησης πολλαπλών κατηγοριών. Μετατρέπει τις ενεργοποιήσεις του τελευταίου επιπέδου σε πιθανότητες, κανονικοποιώντας τις ώστε να αθροίζουν στο 100%. Κάθε κατηγορία λαμβάνει μια πιθανότητα, δείχνοντας πόσο πιθανό είναι το δείγμα να ανήκει σε αυτήν την κατηγορία.



Εικόνα 1.1.2 - 4: Συναρτήσεις Ενεργοποίησης (Activation Functions) και Συναρτήσεις Εξόδου (Output Functions) που χρησιμοποιούνται στα Τεχνητά Νευρωνικά Δίκτυα.
(Πηγή: ResearchGate)

Τεχνικές Κανονικοποίησης (Regularization Techniques): Οι τεχνικές κανονικοποίησης παίζουν σημαντικό ρόλο στη σταθερότητα και τη γενίκευση των δικτύων. Μερικές από αυτές είναι η Batch Normalization [20], που κανονικοποιεί την έξοδο κάθε επιπέδου, και το Dropout [21], που τυχαία απενεργοποιεί ορισμένους νευρώνες κατά την εκπαίδευση, καθώς και οι τεχνικές L1 και L2 regularization (Weight Decay). Οι L1 και L2 regularization [28,29] τεχνικές προσθέτουν έναν όρο ποινής στην συνάρτηση απώλειας ο οποίος λειτουργεί ως αποτρεπτικός παράγοντας ώστε να αποφευχθούν οι αναθέσεις μεγάλων τιμών στα βάρη. Αυτές οι τεχνικές συμβάλλουν στη μείωση της πολυπλοκότητας του μοντέλου και στην αποφυγή της υπερπροσαρμογής (overfitting), διασφαλίζοντας ότι το δίκτυο δεν προσαρμόζεται υπερβολικά στα

δεδομένα εκπαίδευσης.

Συνάρτηση απώλειας (Loss Function): Η συνάρτηση απώλειας είναι ένα κρίσιμο στοιχείο στη διαδικασία εκπαίδευσης των νευρωνικών δικτύων, καθώς μετρά πόσο καλά ή κακά αποδίδει το δίκτυο στις προβλέψεις του. Συγκρίνει την πρόβλεψη του δικτύου με την πραγματική τιμή (label) και υπολογίζει το σφάλμα που προκύπτει. Το σφάλμα αυτό είναι που χρησιμοποιείται στη συνέχεια για την προσαρμογή των βαρών του δικτύου μέσω της διαδικασίας της αναδρομικής διάδοσης (backpropagation).

Υπάρχουν διάφοροι τύποι συναρτήσεων απώλειας, οι οποίες επιλέγονται ανάλογα με το είδος του προβλήματος:

- **Mean Squared Error (MSE):** Χρησιμοποιείται συνήθως σε προβλήματα παλινδρόμησης, όπου η συνάρτηση αυτή μετρά τη μέση τετραγωνική απόκλιση μεταξύ των πραγματικών τιμών και των προβλέψεων του δικτύου.

Μαθηματική έκφραση του Mean Squared Error:

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 N$$

Όπου y_i η πραγματική τιμή, \hat{y}_i η προβλεπόμενη τιμή και N το πλήθος των δειγμάτων.

- **Cross-Entropy Loss:** Χρησιμοποιείται συχνά σε προβλήματα ταξινόμησης, ιδίως όταν υπάρχουν πολλές κατηγορίες. Η συνάρτηση αυτή υπολογίζει την απόκλιση μεταξύ της προβλεπόμενης κατανομής πιθανοτήτων και της πραγματικής κατανομής.

Μαθηματική έκφραση του Cross-Entropy Loss:

$$\text{Cross-Entropy Loss} = - \sum_{i=1}^n \sum_{c=1}^C y_{i,c} \log(\hat{y}_{i,c})$$

Όπου $y_{i,c}$ η ένδειξη αν το δείγμα i ανήκει στην κατηγορία c (1 αν ανήκει, 0 αν δεν ανήκει), $\hat{y}_{i,c}$ η προβλεπόμενη πιθανότητα ότι το δείγμα i ανήκει στην κατηγορία c και C ο αριθμός των κατηγοριών.

- **Binary Cross-Entropy:** Ειδική περίπτωση της cross-entropy loss που χρησιμοποιείται σε δυαδικά προβλήματα ταξινόμησης, όπου οι προβλέψεις είναι μεταξύ δύο κατηγοριών.

$$\text{Binary Cross-Entropy Loss} = - \frac{1}{n} \sum_{i=1}^n [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]$$

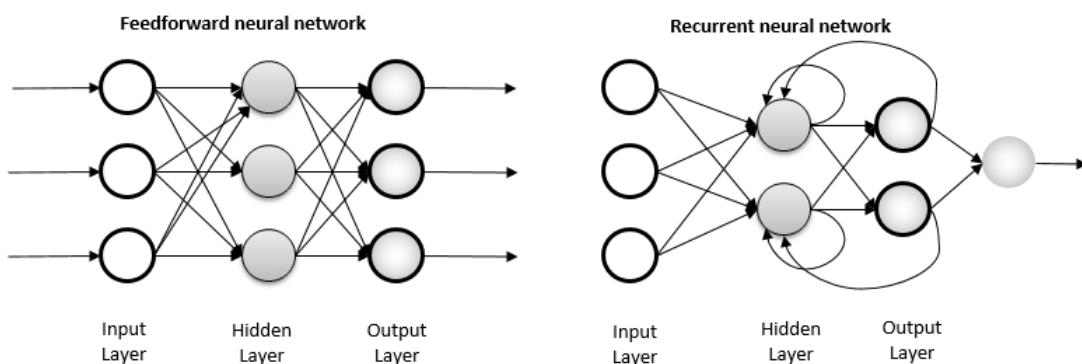
Όπου y_i η πραγματική ετικέτα (0 ή 1), \hat{y}_i η προβλεπόμενη πιθανότητα ότι η ετικέτα έιναι 1 και n το πλήθος των δειγμάτων.

Η συνάρτηση απώλειας διαδραματίζει βασικό ρόλο στη διαδικασία εκπαίδευσης του νευρωνικού δικτύου, καθώς είναι η βάση για την ενημέρωση των βαρών μέσω των αλγορίθμων βελτιστοποίησης όπως ο Gradient Descent ή ο Adam. Για παράδειγμα, στο πλαίσιο των Συνελικτικών Νευρωνικών Δικτύων (CNNs), η συνάρτηση απώλειας παραμένει κρίσιμη, ιδιαίτερα σε προβλήματα όπως η αναγνώριση εικόνας ή η ανίχνευση αντικειμένων, όπου οι σωστές προβλέψεις είναι ζωτικής σημασίας. Η επιλογή της κατάλληλης συνάρτησης απώλειας είναι καίρια για την επιτυχή εκπαίδευση του μοντέλου και την επίτευξη υψηλής απόδοσης.

Τύποι Αρχιτεκτονικών: Τα τεχνητά νευρωνικά δίκτυα ξεκίνησαν από μια απλή ιδέα που εισήχθη με το Περσέπτρον (Perceptron), το οποίο αποτέλεσε το θεμέλιο για την ανάπτυξη και εξέλιξη των νευρωνικών δικτύων. Το Περσέπτρον, που αναπτύχθηκε από τον Frank Rosenblatt το 1958 [22], ήταν η πρώτη προσπάθεια δημιουργίας ενός τεχνητού νευρώνα που θα μπορούσε να πραγματοποιήσει βασικές ταξινομήσεις με βάση γραμμικούς διαχωρισμούς των δεδομένων. Αυτή η πρωτόλεια μορφή τεχνητού νευρώνα αποτέλεσε τη βάση πάνω στην οποία αναπτύχθηκαν πιο πολύπλοκα και ισχυρά νευρωνικά δίκτυα.

Με την πάροδο του χρόνου, η αρχική ιδέα του Περσέπτρον εξελίχθηκε και διαμορφώθηκε σε διάφορες μορφές νευρωνικών δικτύων, τα οποία είναι ικανά να αντιμετωπίζουν πιο σύνθετα προβλήματα. Από τις πρώτες αυτές εξελίξεις προέκυψαν αρκετοί διαφορετικοί τύποι αρχιτεκτονικών των τεχνητών νευρωνικών δικτύων, με πέντε από αυτούς να αποτελούν τις πιο βασικές και σημαντικές κατηγορίες σήμερα:

- **Προσαγωγικά ή Προστροφοδοτούμενα Νευρωνικά Δίκτυα (Feedforward Neural Networks):** Αυτά τα νευρωνικά δίκτυα είναι από τις πιο απλές μορφές Τεχνητών Νευρωνικών Δικτύων, όπου τα δεδομένα τροφοδοτούνται στο δίκτυο και ταξιδεύουν μόνο προς μία κατεύθυνση. Τα δεδομένα περνούν μέσα από τους κόμβους εισόδου και εξέρχονται στους κόμβους εξόδου. Αυτό το νευρωνικό δίκτυο μπορεί να έχει ή να μην έχει κρυφά επίπεδα, γι' αυτό και ονομάζεται προσαγωγικό δίκτυο.
- **Νευρωνικά Δίκτυα με Ακτινικές Συναρτήσεις Βάσης (Radial Basis Function Neural Networks):** Οι ακτινικές συναρτήσεις είναι ένα δημοφιλές σύνολο συναρτήσεων που χρησιμοποιούνται στους υπολογισμούς αποστάσεων, λαμβάνοντας υπόψη την απόσταση ενός σημείου σε σχέση με το κέντρο. Τα RBF νευρωνικά δίκτυα [23] χρησιμοποιούν αυτές τις συναρτήσεις ως ενεργοποιήσεις. Η έξοδος του δικτύου είναι ένας γραμμικός συνδυασμός των συναρτήσεων εξόδων των ακτινικών συναρτήσεων και των βαρών των νευρώνων, όπου οι ακτινικές συναρτήσεις έχουν ως είσοδο τις εισόδους του δικτύου.
- **Αναδρομικά Νευρωνικά Δίκτυα (Recurrent Neural Networks):** Ένα αναδρομικό νευρωνικό δίκτυο (RNN) [24] είναι ένας τύπος τεχνητού νευρωνικού δικτύου που χρησιμοποιείται συνήθως στην αναγνώριση ομιλίας και στην επεξεργασία φυσικής γλώσσας (NLP). Τα RNN είναι σχεδιασμένα για να αναγνωρίζουν τα διαδοχικά χαρακτηριστικά των δεδομένων και να χρησιμοποιούν μοτίβα για την πρόβλεψη του πιο πιθανού επόμενου σεναρίου.
- **Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks):** Τα συνελικτικά νευρωνικά δίκτυα είναι παρόμοια με τα προσαγωγικά νευρωνικά δίκτυα, όπου οι νευρώνες έχουν βαρύτητες (weights) και μεροληψίες (biases) που μπορούν να μαθευτούν. Τα CNNs έχουν πολλές εφαρμογές σε πολλούς τομείς, όπως η αναγνώριση εικόνας, η ανίχνευση αντικειμένων, η αναγνώριση προσώπων, η ανάλυση εικόνας και άλλες. (Σε επόμενη ενότητα θα αναλύσουμε σε βάθος τα Συνελικτικά Νευρωνικά Δίκτυα)
- **Αρθρωτά Νευρωνικά Δίκτυα (Modular Neural Networks):** Τα Αρθρωτά Νευρωνικά Δίκτυα αποτελούνται από μια συλλογή διαφορετικών δικτύων που λειτουργούν ανεξάρτητα και συμβάλλουν στο τελικό αποτέλεσμα. Κάθε νευρωνικό δίκτυο έχει ένα σύνολο εισόδων που είναι μοναδικές σε σύγκριση με άλλα δίκτυα, κατασκευάζοντας και εκτελώντας υπο-εργασίες.



Εικόνα 1.1.2 - 5: Σύγκριση Feedforward Neural Network(FFNN) με Recurrent Neural Network (RNN).
(Πηγή: ResearchGate)

Εφαρμογές και Παραδείγματα Χρήσης: Τα Τεχνητά Νευρωνικά Δίκτυα έχουν βρει εφαρμογή σε πολλούς και διαφορετικούς τομείς, χάρη στην ικανότητά τους να μαθαίνουν από δεδομένα και να κάνουν προβλέψεις ή κατηγοριοποιήσεις με ακρίβεια. Στην αναγνώριση εικόνων, τα ANNs χρησιμοποιούνται για την κατηγοριοποίηση και την ανάλυση οπτικών δεδομένων, όπως είναι η ανίχνευση αντικειμένων, προσώπων και χειρόγραφων κειμένων. Αυτές οι τεχνικές χρησιμοποιούνται σε εφαρμογές όπως η αυτόματη αναγνώριση πινακίδων οχημάτων (ANPR)[54], η ιατρική διάγνωση μέσω ανάλυσης ιατρικών εικόνων, και τα συστήματα ασφαλείας με βάση την αναγνώριση προσώπων.

Στον τομέα της επεξεργασίας φυσικής γλώσσας (NLP), τα ANNs βοηθούν στην κατανόηση, μετάφραση και παραγωγή κειμένου, βελτιώνοντας τις επιδόσεις των μηχανών αναζήτησης, των εικονικών βιοηθών (όπως είναι η Siri και η Alexa) και των συστημάτων αυτόματης μετάφρασης. Συστήματα συστάσεων, όπως αυτά που χρησιμοποιούνται από το Netflix, το YouTube και το Amazon, βασίζονται επίσης σε ANNs για να προτείνουν περιεχόμενο στους χρήστες με βάση τις προτιμήσεις και τις συνήθειές τους.

Στη ρομποτική, τα ANNs ενισχύουν την αυτονομία και τη λήψη αποφάσεων των ρομπότ, επιτρέποντάς τους να αλληλεπιδρούν με το περιβάλλον τους με μεγαλύτερη ακρίβεια, να αποφεύγουν εμπόδια και να εκτελούν πολύπλοκες εργασίες. Επιπλέον, στην ανάλυση μεγάλων δεδομένων (Big Data), τα ANNs χρησιμοποιούνται για την ανίχνευση προτύπων, την πρόβλεψη τάσεων και την αναγνώριση ανωμαλιών, όπως στην ανίχνευση απάτης στις τραπεζικές συναλλαγές. Η δυνατότητα των ANNs να προσαρμόζονται και να μαθαίνουν από νέα δεδομένα τα καθιστά πολύτιμο εργαλείο για την επίλυση προβλημάτων σε πολλούς τομείς της σύγχρονης τεχνολογίας και επιστήμης.

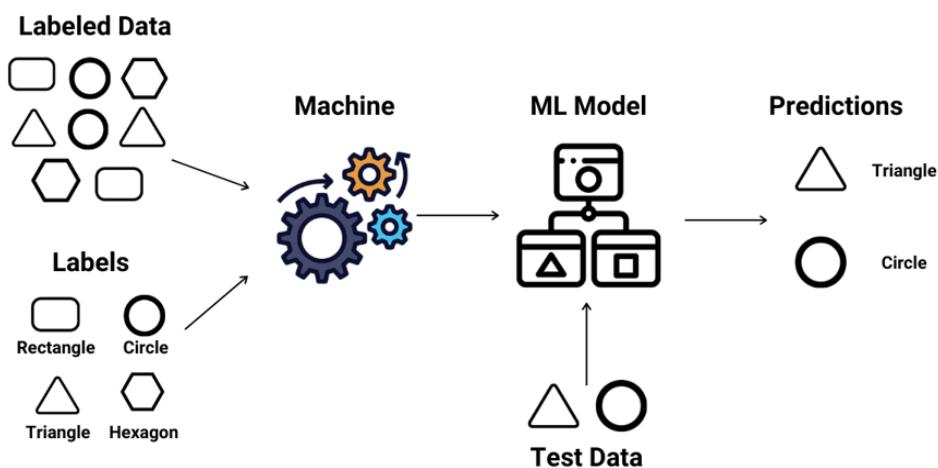
1.2 Είδη Μάθησης

1.2.1 Εποπτευόμενη Μάθηση (Supervised Learning)

Η Εποπτευόμενη Μάθηση (Supervised Learning) είναι μια από τις πιο διαδεδομένες και θεμελιώδεις τεχνικές στον τομέα της μηχανικής μάθησης. Σε αυτήν τη μορφή μάθησης, το μοντέλο εκπαιδεύεται χρησιμοποιώντας ένα σύνολο δεδομένων που είναι ήδη επισημασμένο με τις σωστές απαντήσεις. Κάθε δείγμα δεδομένων συνοδεύεται από μια ετικέτα ή μια επιθυμητή έξιδο, η οποία χρησιμεύει ως οδηγός για το μοντέλο κατά τη διαδικασία της μάθησης. Ο στόχος είναι να μάθει το μοντέλο να αντιστοιχίζει σωστά τα δεδομένα εισόδου με τις αντίστοιχες εξόδους, ώστε να μπορεί να κάνει προβλέψεις ή ταξινομήσεις σε νέα, αταξινόμητα δεδομένα.

Η διαδικασία της αυτού του είδους μάθησης περιλαμβάνει τη διάσπαση του συνόλου των δεδομένων σε τρία βασικά υποσύνολα: το σύνολο εκπαίδευσης (training set), το σύνολο δοκιμής (testing set) και το σύνολο επικύρωσης (validation set). Το σύνολο εκπαίδευσης χρησιμοποιείται για την εκμάθηση των παραμέτρων του μοντέλου, ενώ το σύνολο δοκιμής χρησιμοποιείται για την αξιολόγηση της απόδοσης του μοντέλου σε άγνωστα δεδομένα. Το σύνολο επικύρωσης βοηθά στη ρύθμιση των υπερπαραμέτρων του μοντέλου για να διασφαλιστεί ότι δεν υπάρχει υπερπροσαρμογή στα δεδομένα εκπαίδευσης.

Παρόλο που η Εποπτευόμενη Μάθηση είναι ισχυρή και αποτελεσματική, έχει και ορισμένα μειονεκτήματα. Το μεγαλύτερο από αυτά είναι η ανάγκη για μεγάλο όγκο επισημασμένων δεδομένων, τα οποία συχνά απαιτούν σημαντικό χρόνο και κόπο για να συλλεχθούν και να επισημανθούν από ανθρώπους. Επιπλέον, η ποιότητα των ετικετών είναι κρίσιμη, καθώς τυχόν λάθη στις επισημάνσεις μπορούν να οδηγήσουν σε ανακριβείς προβλέψεις από το μοντέλο. Παρά τις προκλήσεις, η Εποπτευόμενη Μάθηση παραμένει ένα από τα πιο ευρέως χρησιμοποιούμενα εργαλεία στη μηχανική μάθηση, ιδίως σε εφαρμογές όπως η αναγνώριση εικόνας, η ανάλυση κειμένου και η πρόβλεψη τάσεων.



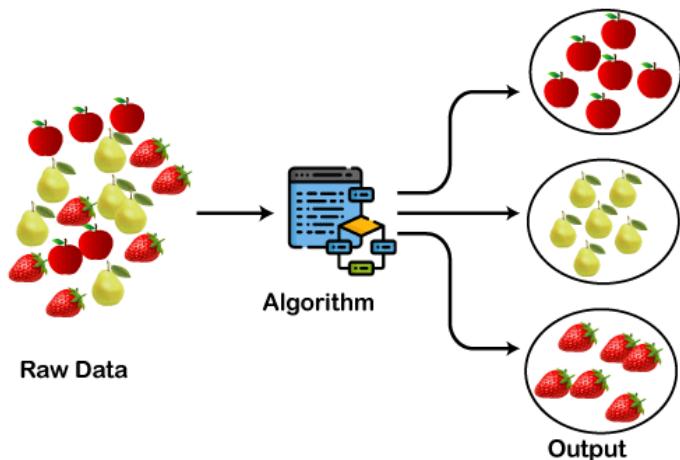
Εικόνα 1.2.1 - 1: Χρήση Εποπτευόμενης Μάθησης για ταξινόμηση επισημασμένων δεδομένων (labeled data). (Πηγή: Enjoy Algorithms)

1.2.2 Μη Εποπτευόμενη Μάθηση (Unsupervised Learning)

Η Μη Εποπτευόμενη Μάθηση (Unsupervised Learning) είναι μια κατηγορία μηχανικής μάθησης όπου τα μοντέλα εκπαιδεύονται χρησιμοποιώντας δεδομένα που δεν είναι επισημασμένα, δηλαδή δεν έχουν προκαθορισμένες ετικέτες ή κατηγορίες. Σε αντίθεση με την Εποπτευόμενη Μάθηση, όπου το σύστημα μαθαίνει να προβλέπει αποτελέσματα βάσει επισημασμένων δεδομένων, η Μη Εποπτευόμενη Μάθηση στοχεύει στην ανακάλυψη κρυφών μοτίβων, δομών ή σχέσεων μέσα στα δεδομένα.

Ένα από τα κύρια παραδείγματα εφαρμογής της μη εποπτευόμενης μάθησης είναι η ομαδοποίηση (clustering), όπου τα δεδομένα χωρίζονται σε ομάδες βάσει των ομοιοτήτων τους, χωρίς να υπάρχουν προκαθορισμένες κατηγορίες. Άλλη εφαρμογή είναι η μείωση διαστάσεων (dimensionality reduction), η οποία χρησιμοποιείται για να συμπυκνώσει δεδομένα υψηλής διάστασης σε λιγότερες διαστάσεις, διατηρώντας τις πιο σημαντικές πληροφορίες.

Η Μη Εποπτευόμενη Μάθηση είναι ιδιαίτερα χρήσιμη σε περιπτώσεις όπου η επισήμανση των δεδομένων είναι δύσκολη ή αδύνατη, και επιτρέπει στα μοντέλα να ανακαλύπτουν μοτίβα και σχέσεις που δεν είναι άμεσα ορατές, προσφέροντας πολύτιμες γνώσεις και πληροφορίες χωρίς την ανάγκη ανθρώπινης παρέμβασης.



Εικόνα 1.2.2 - 1: Χρήση Μη Εποπτευόμενης Μάθησης μέσω Clustering για ομαδοποίηση δεδομένων χωρίς επισημάνσεις(unlabeled data).
(Πηγή: DataCamp)

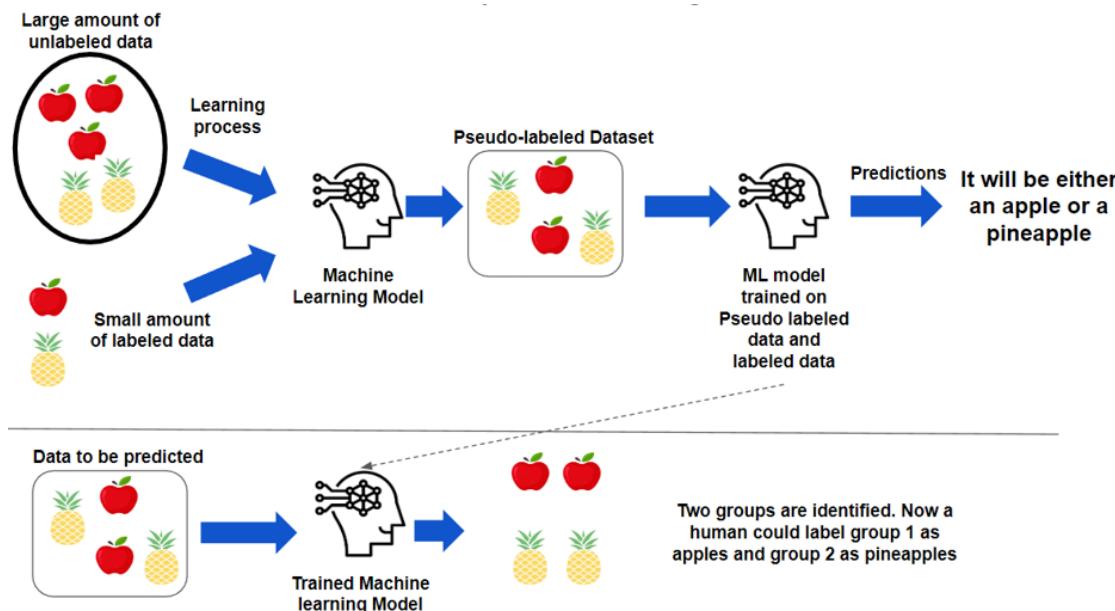
1.2.3 Ημι-Εποπτευόμενη Μάθηση (Semi-Supervised Learning)

Η ημι-εποπτευόμενη μάθηση (Semi-Supervised Learning) είναι ένα υβριδικό είδος μηχανικής μάθησης που συνδυάζει στοιχεία από την εποπτευόμενη και τη μη εποπτευόμενη μάθηση. Σε αυτή την προσέγγιση, το μοντέλο εκπαιδεύεται με ένα μικρό σύνολο επισημασμένων δεδομένων και ένα πολύ μεγαλύτερο σύνολο μη επισημασμένων δεδομένων. Αυτό επιτρέπει στο μοντέλο να αξιοποιήσει τη δομή των μη επισημασμένων δεδομένων για να βελτιώσει την απόδοσή του χωρίς την ανάγκη πλήρους συνόλου επισημάνσεων.

Η ημι-εποπτευόμενη μάθηση είναι ιδιαίτερα χρήσιμη σε περιπτώσεις όπου η συλλογή και η επισήμανση μεγάλων συνόλων δεδομένων είναι δαπανηρή ή χρονοβόρα. Το μικρό ποσοστό επισημασμένων δεδομένων

που χρησιμοποιείται βοηθά στη σωστή καθοδήγηση του μοντέλου, ενώ τα μη επισημασμένα δεδομένα συμβάλλουν στην εκμάθηση σημαντικών προτύπων και δομών που υπάρχουν στα δεδομένα.

Αυτή η προσέγγιση χρησιμοποιείται σε διάφορες εφαρμογές, όπως στην αναγνώριση εικόνας και στην επεξεργασία φυσικής γλώσσας, όπου η επισήμανση δεδομένων μπορεί να είναι δύσκολη ή ακριβή. Με τη χρήση ημι-εποπτευόμενης μάθησης, τα μοντέλα μπορούν να επιτύχουν βελτιωμένες προβλέψεις και μεγαλύτερη ακρίβεια χωρίς την ανάγκη πλήρους εξάρτησης από τα επισημασμένα δεδομένα.



Εικόνα 1.2.3 - 1: Παράδειγμα Ημι-Εποπτευόμενης μάθησης, όπου ένα μοντέλο μαθαίνει από ένα μικρό σύνολο επισημασμένων δεδομένων και ένα μεγαλύτερο σύνολο μη επισημασμένων δεδομένων, δημιουργώντας ένα ψευδο-επισημασμένο σύνολο δεδομένων για βελτιωμένες προβλέψεις.

(Πηγή: DataDrivenInvsetor)

1.2.4 Αυτο-εποπτευόμενη Μάθηση (Self-Supervised Learning)

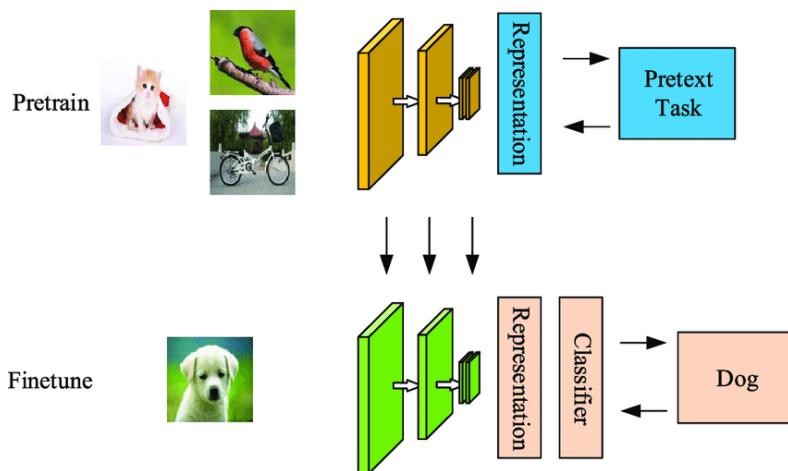
Η Αυτο-Εποπτευόμενη Μάθηση (*Self-Supervised Learning - SSL*) είναι μια καινοτόμος προσέγγιση στον τομέα της μηχανικής μάθησης, που γεφυρώνει το κενό μεταξύ εποπτευόμενης και μη εποπτευόμενης μάθησης. Σε αυτήν τη μέθοδο, τα μοντέλα μαθαίνουν να εξάγουν πληροφορίες από δεδομένα χωρίς να απαιτούνται επισημάνσεις που έχουν προετοιμαστεί από ανθρώπους. Αντί να βασίζονται σε εξωτερικές επισημάνσεις, τα μοντέλα δημιουργούν τις δικές τους εποπτευόμενες εργασίες, γνωστές ως έμμεσες διεργασίες, αξιοποιώντας τα διαθέσιμα δεδομένα. Έτσι, μέσω έμμεσων διεργασιών, επιτυγχάνεται η εκμάθηση κάποιων γενικευμένων χαρακτηριστικών που μπορούν να φανούν χρήσιμα ώστε να βελτιώσουν την απόδοση στις διαδικασίες που ακολουθούν, οι οποίες πολλές φορές χρειάζονται επισημασμένα σύνολα δεδομένων.

Ένα χαρακτηριστικό παράδειγμα μιας τέτοιας έμμεσης διεργασίας είναι το *Image Inpainting* (η τεχνική αυτή θα υλοποιηθεί και θα χρησιμοποιηθεί στο Πειραματικό Μέρος), όπου το μοντέλο καλείται να συμπληρώσει τα ελλείποντα τμήματα μιας εικόνας. Μέσω αυτής της διαδικασίας, το μοντέλο μαθαίνει να κατανοεί τη

δομή και τα μοτίβα της εικόνας, χωρίς να χρειάζεται προκαθορισμένες επισημάνσεις. Έτσι, το μοντέλο εκπαιδεύεται να αναγνωρίζει τις βασικές ιδιότητες των δεδομένων, κάτι που μπορεί να εφαρμοστεί στα downstream tasks.

Η Αυτο-Εποπτευόμενη Μάθηση προσφέρει σημαντικά πλεονεκτήματα, καθώς επιτρέπει την εκμετάλλευση μεγάλων ποσοτήτων μη επισημασμένων δεδομένων, μειώνοντας την εξάρτηση από τη χρονοβόρα και δαπανηρή διαδικασία επισήμανσης. Αυτή η προσέγγιση ανοίγει νέες δυνατότητες στη μηχανική μάθηση, καθιστώντας τα μοντέλα πιο αποδοτικά και ικανά να γενικεύουν σε ευρύτερες εφαρμογές.

Σε αυτό το σημείο είναι σημαντικό να αναφέρουμε την τεχνική της Μεταφοράς Μάθησης (Transfer Learning) και του fine-tuning. Η μεταφορά μάθησης επιτρέπει τη χρήση ενός προ-εκπαιδευμένου μοντέλου (π.χ. μιας έμμεσης διεργασίας), το οποίο έχει μάθει να εξάγει χρήσιμα χαρακτηριστικά από ένα μεγάλο σύνολο δεδομένων, σε μια νέα εργασία με περιορισμένα δεδομένα. Το fine-tuning αναφέρεται στην προσαρμογή αυτού του προ-εκπαιδευμένου μοντέλου στα νέα δεδομένα, επιτρέποντας την εξειδίκευση του μοντέλου για την επίτευξη καλύτερων αποτελεσμάτων στη νέα εργασία. Αυτή η διαδικασία είναι ιδιαίτερα αποτελεσματική όταν εφαρμόζεται σε συνδυασμό με την αυτο-εποπτευόμενη μάθηση, βελτιώνοντας την απόδοση του μοντέλου στα downstream tasks.



Εικόνα 1.2.4 - 1: Διάγραμμα που απεικονίζει τη διαδικασία της Αυτο-Εποπτευόμενης Μάθησης, όπου το μοντέλο αρχικά εκπαιδεύεται σε μια έμμεση διεργασία για να μάθει χρήσιμες αναπαραστάσεις από μη επισημασμένα δεδομένα και στη συνέχεια γίνεται Μεταφορά Μάθησης και βελτιστοποίηση για την εκτέλεση μιας συγκεκριμένης εποπτευόμενης εργασίας, όπως η ταξινόμηση εικόνας.

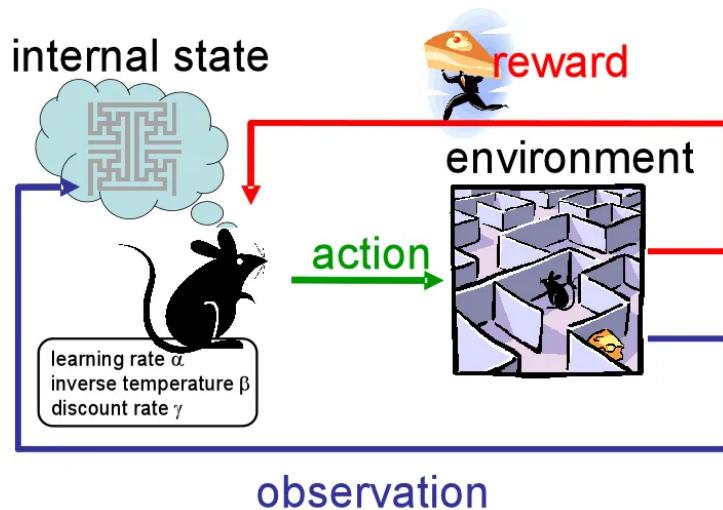
(Πηγή: V7 Labs)

1.2.5 Ενισχυτική Μάθηση (Reinforcement Learning)

To Reinforcement Learning (RL) είναι μια υποκατηγορία της μηχανικής μάθησης όπου ένας πράκτορας (agent) μαθαίνει να λαμβάνει αποφάσεις αλληλεπιδρώντας με το περιβάλλον του. Ο πράκτορας επιδιώκει να μεγιστοποιήσει τις συνολικές ανταμοιβές (rewards) που λαμβάνει μέσω των ενεργειών του. Η βασική ιδέα είναι ότι ο πράκτορας λαμβάνει ανατροφοδότηση από το περιβάλλον του με τη μορφή επιβραβεύσεων ή ποινών ανάλογα με τις αποφάσεις του. Αυτή η ανατροφοδότηση χρησιμοποιείται για την προσαρμογή της συμπεριφοράς του πράκτορα έτσι ώστε να επιτυγχάνει καλύτερες επιδόσεις στο μέλλον.

Ένα σημαντικό στοιχείο του RL είναι η μάθηση μιας πολιτικής (policy), δηλαδή μιας στρατηγικής που καθορίζει ποιες ενέργειες πρέπει να λαμβάνονται σε κάθε δεδομένη κατάσταση. Οι αλγόριθμοι Reinforcement Learning, όπως ο Q-learning [25] και το SARSA [26], επιτρέπουν στον πράκτορα να ανακαλύπτει ποια ενέργεια είναι η καλύτερη για τη μεγιστοποίηση της συνολικής ανταμοιβής. Επιπλέον, με την πρόοδο της τεχνολογίας, έχουν αναπτυχθεί πιο προχωρημένες τεχνικές, όπως τα Deep Q-Networks (DQN) [27], που χρησιμοποιούν νευρωνικά δίκτυα για την επίλυση πιο σύνθετων προβλημάτων.

Το Reinforcement Learning έχει βρει εφαρμογή σε πολλούς τομείς, από τη ρομποτική και τα συστήματα συστάσεων μέχρι τον τομέα των βιντεοπαιχνιδιών. Στη ρομποτική, οι πράκτορες μαθαίνουν να εκτελούν περίπλοκες εργασίες σε πραγματικά περιβάλλοντα, όπως η πλοήγηση σε αχαρτογράφητους χώρους. Στα συστήματα συστάσεων, το RL βοηθά στη βελτίωση της ακρίβειας των προτάσεων προς τους χρήστες, ενώ στα βιντεοπαιχνίδια, οι πράκτορες έχουν επιτύχει εξαιρετικές επιδόσεις, καταφέρνοντας να νικούν ακόμα και τους καλύτερους ανθρώπινους παίκτες.



Εικόνα 1.2.5 - 1: Ένα απλό παράδειγμα Ενισχυτικής Μάθησης όπου ένας πράκτορας (το ποντίκι) αλληλεπιδρά με το περιβάλλον του, παρατηρεί τις αλλαγές, λαμβάνει ενέργειες και μαθαίνει μέσω ανταμοιβών (το τυρί) για να βελτιστοποιήσει τη συμπεριφορά του.
(Πηγή: Medium)

1.3 Συνελικτικά Νευρωνικά Δίκτυα (CNNs)

1.3.1 Εισαγωγή στα CNNs

Τα Συνελικτικά Νευρωνικά Δίκτυα (*Convolutional Neural Networks - CNNs*) αποτελούν ένα από τα πιο σημαντικά επιτεύγματα στον τομέα της Μηχανικής Μάθησης και της Τεχνητής Νοημοσύνης, ιδιαίτερα στον τομέα της Επεξεργασίας Εικόνας και Αναγνώρισης Προτύπων. Τα CNNs παρουσιάστηκαν για πρώτη φορά από τον Yann LeCun το 1989, με την εισαγωγή του LeNet-5, ενός δικτύου που σχεδιάστηκε για την αναγνώριση χειρόγραφων ψηφίων στο σύστημα ταχυδρομικών κωδικών. Αυτή η πρωτοποριακή εργασία αποτέλεσε τη βάση για την εξέλιξη των CNNs, τα οποία σήμερα χρησιμοποιούνται ευρέως σε ποικίλες εφαρμογές, από την ανάλυση εικόνας και βίντεο μέχρι την αναγνώριση ομιλίας και την επεξεργασία φυσικής γλώσσας.

Με την πάροδο του χρόνου, τα CNNs έχουν υποστεί σημαντικές βελτιώσεις και τροποποιήσεις, με την εισαγωγή νέων αρχιτεκτονικών, όπως το AlexNet, το VGGNet, και το ResNet, οι οποίες έχουν συμβάλει στην επίτευξη καλύτερων αποτελεσμάτων και στη γενίκευση των μοντέλων σε διάφορα προβλήματα. Αυτές οι εξελιγμένες αρχιτεκτονικές, σε συνδυασμό με προχωρημένες τεχνικές εκπαίδευσης και βελτιστοποίησης, έχουν καταστήσει τα CNNs ένα από τα πιο ισχυρά εργαλεία στη σύγχρονη Τεχνητή Νοημοσύνη.

Τα CNNs διαφοροποιούνται από τα παραδοσιακά Τεχνητά Νευρωνικά Δίκτυα λόγω της ειδικής τους δομής, η οποία είναι προσαρμοσμένη για την επεξεργασία δεδομένων με δομή, όπως οι εικόνες. Η βασική αρχή λειτουργίας τους στηρίζεται στη χρήση συνελικτικών επιπέδων (*convolutional layers*), τα οποία έχουν τη δυνατότητα να ανιχνεύουν και να εξάγουν τοπικά χαρακτηριστικά από τα δεδομένα εισόδου. Αυτή η ικανότητα καθιστά τα CNNs ιδανικά για εφαρμογές όπως η αναγνώριση αντικειμένων και η κατηγοριοποίηση εικόνων, όπου η τοπική πληροφορία παίζει κρίσιμο ρόλο.

1.3.2 Δομή και Λειτουργία Επιπέδων στα CNNs:

Τα Συνελικτικά Νευρωνικά Δίκτυα αποτελούνται από διάφορα επίπεδα, καθένα από τα οποία παίζει έναν ειδικό ρόλο στην ανάλυση των δεδομένων εισόδου και την παραγωγή της τελικής εξόδου. Σε αυτό το κεφάλαιο, θα εξετάσουμε λεπτομερώς τα κύρια είδη επιπέδων που συνθέτουν την αρχιτεκτονική των CNNs, καθώς και τις σημαντικές αρχιτεκτονικές που έχουν αναπτυχθεί για να βελτιώσουν την απόδοση των δικτύων σε ποικίλα προβλήματα.

Συνελικτικά Επίπεδα (*Convolutional Layers*): Κατά τη διάρκεια της εμπρός διέλευσης (*forward pass*), ο πυρήνας/φίλτρο (*kernel*) διασχίζει την εικόνα κατά μήκος του ύψους και του πλάτους, δημιουργώντας μια αναπαράσταση της περιοχής υποδοχής που καλύπτει. Αυτό έχει ως αποτέλεσμα τη δημιουργία ενός δισδιάστατου χάρτη ενεργοποίησης (*activation map*), που απεικονίζει την απόκριση του πυρήνα σε κάθε χωρική θέση της εικόνας. Το μέγεθος του βήματος (*stride*) καθορίζει πόσο μακριά μετακινείται ο πυρήνας σε κάθε βήμα.

Αν, για παράδειγμα, έχουμε μια είσοδο μεγέθους $W \times W \times D$ και χρησιμοποιούμε Dout πυρήνες με χωρικό μέγεθος F , βήμα S και ποσότητα επένδυσης (*padding*) P (Η επένδυση (*padding*) είναι η διαδικασία πρόσθεσης επιπλέον μηδενικών pixel γύρω από την αρχική εικόνα ώστε να μην χαθεί πληροφορία λόγω της σμίκρυνσης της εικόνας εισόδου κατά τις αλλαγές από το φίλτρο), τότε το μέγεθος του όγκου εξόδου μπορεί να καθοριστεί

με τον παρακάτω μαθηματικό τύπο:

$$W_{out} = \frac{W - F + 2P}{S} + 1$$

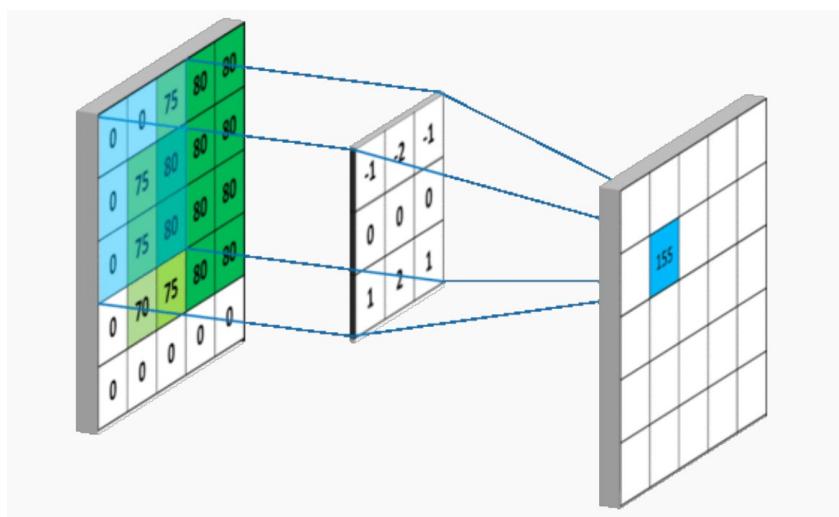
Αυτό θα οδηγήσει σε έναν όγκο εξόδου μεγέθους $W_{out} \times W_{out} \times D_{out}$.

Κίνητρα πίσω από την Συνέλιξη (Convolution): Η Συνέλιξη βασίζεται σε τρεις κύριες ιδέες που έχουν επηρεάσει σημαντικά την επεξεργασία εικόνας: την αραιή αλληλεπίδραση (sparse interaction), την κοινή χρήση παραμέτρων (parameter sharing) και την ισοβαρή αναπαράσταση (equivariant representation).

Στα παραδοσιακά Νευρωνικά Δίκτυα, κάθε μονάδα εξόδου συνδέεται με κάθε μονάδα εισόδου, απαιτώντας μεγάλο αριθμό παραμέτρων. Αντίθετα, στα Συνελικτικά Νευρωνικά Δίκτυα, χρησιμοποιούνται μικρότεροι πυρήνες/φίλτρα (kernels) που αλληλεπιδρούν με ένα μικρότερο μέρος της εισόδου. Αυτό μειώνει τις απαιτήσεις μνήμης και αυξάνει την αποδοτικότητα του μοντέλου.

Επιπλέον, τα Συνελικτικά Νευρωνικά Δίκτυα εφαρμόζουν τα ίδια βάρη σε διαφορετικές θέσεις της εικόνας, κάνοντας χρήση της κοινής χρήσης παραμέτρων. Αυτή η πρακτική όχι μόνο απλοποιεί τον υπολογισμό, αλλά διασφαλίζει ότι το δίκτυο μπορεί να αναγνωρίζει χαρακτηριστικά ανεξάρτητα από τη θέση τους στην εικόνα, καθιστώντας το δίκτυο ισοβαρές σε μεταθέσεις. Αυτό σημαίνει ότι οι αλλαγές στην είσοδο αντικατοπτρίζονται με αντίστοιχο τρόπο στην έξοδο.

Τα CNNs χρησιμοποιούν την ίδια ομάδα βαρών για διαφορετικά σημεία της εικόνας, κάτι που ονομάζεται κοινή χρήση παραμέτρων. Αυτό κάνει τους υπολογισμούς πιο απλούς και επιτρέπει στο δίκτυο να αναγνωρίζει χαρακτηριστικά ανεξάρτητα από το πού βρίσκονται στην εικόνα. Με άλλα λόγια, αν η είσοδος αλλάξει, η έξοδος θα αλλάξει με τον ίδιο τρόπο, κάνοντας το δίκτυο ικανό να διαχειρίζεται μετατοπίσεις στην εικόνα χωρίς να χάνει την ακρίβειά του.



Εικόνα 1.3.2 - 1: Εφαρμογή ενός πυρήνα/φίλτρου (kernel) σε ένα τμήμα της εικόνας κατά τη διάρκεια της Συνέλιξης (Convolution) για την εξαγωγή χαρακτηριστικών, παράγοντας έναν Χάρτη Ενεργοποίησης (Activation Map). (Πηγή: TowardsDataScience)

Επίπεδα Υποδειγματοληψίας (Pooling Layers): Η λειτουργία τους στο Νευρωνικό Δίκτυο συνίσταται στην αντικατάσταση της εξόδου του δικτύου σε ορισμένα σημεία, υπολογίζοντας μια συνοπτική στατιστική των γειτονικών εξόδων. Αυτό έχει ως αποτέλεσμα τη μείωση του χωρικού μεγέθους της αναπαράστασης, πράγμα που μειώνει τις απαιτήσεις υπολογιστικής ισχύος και των βαρών που χρειάζονται για την επεξεργασία. Η διαδικασία αυτή εφαρμόζεται ξεχωριστά σε κάθε τμήμα της αναπαράστασης.

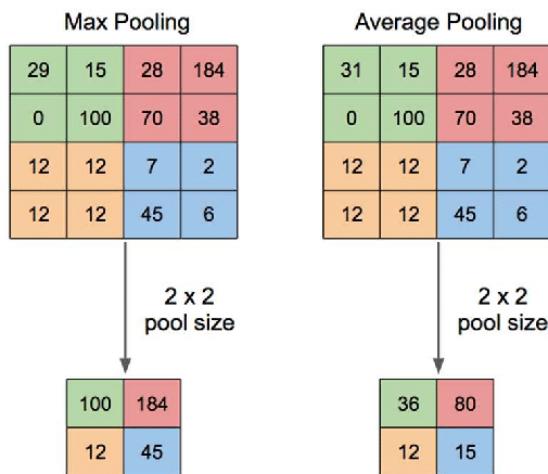
Υπάρχουν διάφορες μέθοδοι pooling, όπως η λήψη του μέσου όρου των τιμών μιας ορθογώνιας περιοχής, η εφαρμογή του L2 κανόνα (L2 norm) σε αυτή την περιοχή, ή ένας σταθμισμένος μέσος όρος βάσει της απόστασης από το κεντρικό pixel. Ωστόσο, η πιο συνηθισμένη μέθοδος είναι το max pooling, όπου επιλέγεται η μέγιστη τιμή από τη γειτονιά (Βλέπε Εικόνα 1.2.1 - 4).

Αν υποθέσουμε ότι έχουμε έναν χάρτη ενεργοποίησης (activation map) με μέγεθος $W \times W \times D$, έναν πυρήνα/φίλτρο (kernel) υποδειγματοληψίας με χωρικό μέγεθος F , και βήμα (stride) S , τότε το μέγεθος του εξόδου μπορεί να υπολογιστεί από τον παρακάτω μαθηματικό τύπο:

$$W_{out} = \frac{W - F}{S} + 1$$

που δίνει ως αποτέλεσμα έναν όγκο εξόδου με μέγεθος $W_{out} \times W_{out} \times D$.

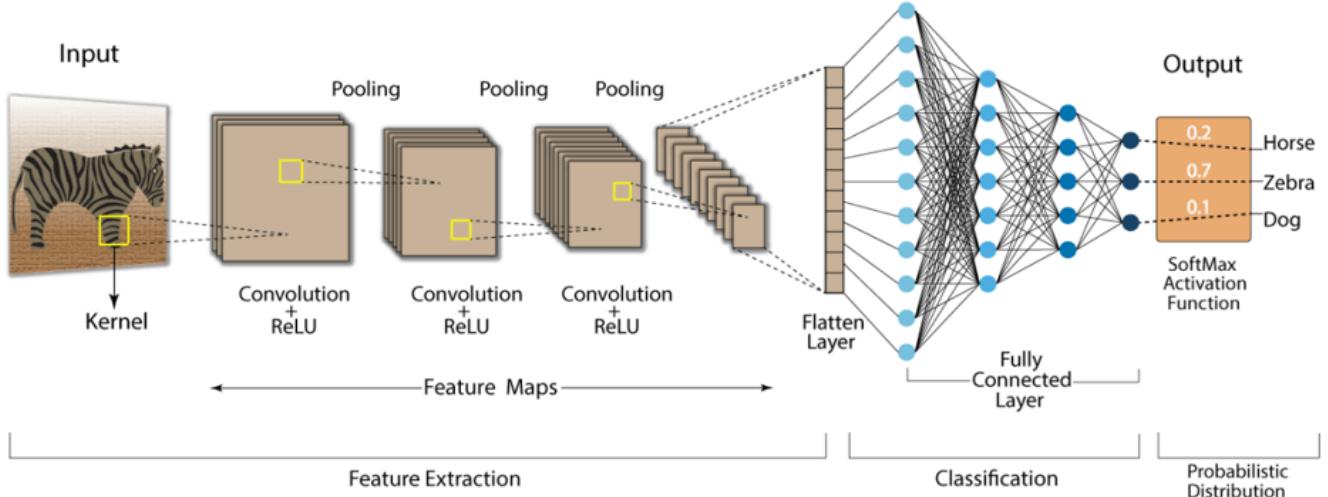
Το pooling προσφέρει επίσης κάποια μετάθεση ανεξαρτησίας (translation invariance), που σημαίνει ότι ένα αντικείμενο μπορεί να αναγνωριστεί ανεξάρτητα από το πού εμφανίζεται στο πλαίσιο.



Εικόνα 1.3.2 - 2: Διαδικασίες Max Pooling και Average Pooling σε ένα χάρτη ενεργοποίησης με χρήση kernel μεγέθους 2x2. (Πηγή: ResearchGate)

Πλήρως Συνδεδεμένο Επίπεδο (Fully Connected Layer): Στο Πλήρως Συνδεδεμένο Επίπεδο (συχνά αναφέρεται και σαν Dense Layer), κάθε νευρώνας συνδέεται πλήρως με όλους τους νευρώνες τόσο στο προηγούμενο όσο και στο επόμενο επίπεδο, όπως συμβαίνει και στα παραδοσιακά νευρωνικά δίκτυα. Αυτή η δομή επιτρέπει την υπολογιστική διαδικασία με πολλαπλασιασμό μητρώων, ακολουθούμενο από την προσθήκη ενός όρου μεροληψίας (bias). Το πλήρως συνδεδεμένο επίπεδο είναι κρίσιμο για τη χαρτογράφηση της αναπαράστασης των δεδομένων από την είσοδο προς την έξοδο του δικτύου.

Επίπεδα Μη-Γραμμικότητας (Non-Linearity Layers): Επειδή η συνέλιξη είναι μια γραμμική διαδικασία, και οι εικόνες δεν είναι γραμμικές, τα επίπεδα μη-γραμμικότητας τοποθετούνται συχνά αμέσως μετά από το συνελικτικό επίπεδο. Αυτά τα επίπεδα προσθέτουν μη-γραμμικότητα στον χάρτη ενεργοποίησης, κάτι που επιτρέπει στο δίκτυο να διαχειρίζεται και να αναγνωρίζει πιο σύνθετες δομές και σχέδια στα δεδομένα.



Εικόνα 1.3.2 - 3: Διάγραμμα ενός Συνελικτικού Νευρωνικού Δικτύου (CNN), που απεικονίζει τη διαδικασία εξαγωγής χαρακτηριστικών (feature extraction) μέσω των επιπέδων συνέλιξης (convolution layers), ReLU και pooling, και την τελική ταξινόμηση με χρήση των Flatten και Fully Connected επιπέδων, οδηγώντας σε μια πιθανολογική κατανομή κατηγοριών μέσω της συνάρτησης SoftMax. (Πηγή: Analytics Vidhya)

1.3.3 Τεχνικές Κανονικοποίησης

Στην ενότητα "Εισαγωγή στα Τεχνητά Νευρωνικά Δίκτυα (ANNs)", αναφέρθηκαν επιγραμματικά οι τεχνικές κανονικοποίησης (Normalization Techniques) που χρησιμοποιούνται για τη βελτίωση της σταθερότητας και της γενίκευσης των νευρωνικών δικτύων. Τώρα, θα προχωρήσουμε σε μια πιο λεπτομερή ανάλυση αυτών των τεχνικών, εξηγώντας τον ρόλο τους στη λειτουργία των Συνελικτικών Νευρωνικών Δικτύων. Αυτές οι τεχνικές είναι ζωτικής σημασίας για την αποτροπή του overfitting, τη σταθεροποίηση της διαδικασίας εκπαίδευσης και την επιτάχυνση της σύγκλισης του μοντέλου, καθιστώντας τα CNNs πιο αποδοτικά και αποτελεσματικά. Στις επόμενες παραγράφους, θα αναλύσουμε τις πιο κοινές τεχνικές κανονικοποίησης, όπως η Batch Normalization, το Dropout, καθώς και οι τεχνικές L1 και L2 regularization.

Η κανονικοποίηση των δεδομένων εισόδου (Input Data Normalization) είναι ένα κρίσιμο βήμα στη διαδικασία εκπαίδευσης των νευρωνικών δικτύων, ιδιαίτερα στα Συνελικτικά Νευρωνικά Δίκτυα. Η κανονικοποίηση εξασφαλίζει ότι όλα τα χαρακτηριστικά των δεδομένων εισόδου έχουν παρόμοια κλίμακα, συνήθως μετατρέποντάς τα σε μια κατανομή με μέση τιμή 0 και τυπική απόκλιση 1. Αυτό βοηθά στην αποφυγή του φαινομένου όπου συγκεκριμένες είσοδοι κυριαρχούν λόγω της μεγάλης κλίμακας τους, γεγονός που μπορεί να δυσκολέψει την εκπαίδευση και να οδηγήσει σε βραδύτερη σύγκλιση. Η κανονικοποίηση των δεδομένων εισόδου επιτρέπει στο μοντέλο να μαθαίνει πιο αποτελεσματικά, καθώς βελτιστοποιεί τη διαδικασία του Gradient Descent, καθιστώντας τον υπολογισμό των βαρών πιο σταθερό και γρήγορο.

Batch Normalization: Η κανονικοποίηση μέσω batch (Batch Normalization) [20] είναι μια τεχνική που προτάθηκε για την αντιμετώπιση των προκλήσεων που προκύπτουν κατά την εκπαίδευση βαθιών νευρωνικών δικτύων. Η ιδέα πίσω από την κανονικοποίηση μέσω batch είναι να μειώσει τη μεταβολή της κατανομής των εισόδων σε κάθε επίπεδο του δικτύου, ένα πρόβλημα που είναι γνωστό ως εσωτερική διακύμανση της ενεργοποίησης (internal covariate shift). Αυτό επιτυγχάνεται με την κανονικοποίηση της εξόδου κάθε επιπέδου μέσω της αφαίρεσης της μέσης τιμής και της διαίρεσης με την τυπική απόκλιση, για το συγκεκριμένο batch εισόδων. Στη συνέχεια, εισάγονται δύο επιπλέον παραμετροποιήσιμες τιμές, η κλίση (scale) και η μετατόπιση (shift), που επιτρέπουν στο δίκτυο να προσαρμόσει τη διαδικασία κανονικοποίησης ανάλογα με τις απαιτήσεις του μοντέλου. Αυτό βοηθά στη σταθεροποίηση της εκπαίδευσης και επιτρέπει τη χρήση υψηλότερων μαθησιακών ρυθμών, βελτιώνοντας την απόδοση και επιταχύνοντας τη διαδικασία σύγκλισης.

L1 Regularization: Η κανονικοποίηση L1 (L1 Regularization) [28] είναι μια τεχνική που χρησιμοποιείται για την αποτροπή του overfitting ενός νευρωνικού δικτύου. Η κανονικοποίηση αυτή επιβάλλει έναν όρο ποινής στο μέγεθος των βαρών του δικτύου, προσθέτοντας το απόλυτο άθροισμα των τιμών των βαρών στη συνάρτηση κόστους. Αυτό ενθαρρύνει το μοντέλο να παράγει αραιές αναπαραστάσεις (sparse representations), όπου πολλά από τα βάρη μπορεί να είναι μηδενικά, οδηγώντας σε ένα απλούστερο μοντέλο που είναι λιγότερο πιθανό να υπερπροσαρμοστεί στα δεδομένα εκπαίδευσης. Η L1 κανονικοποίηση είναι ιδιαίτερα χρήσιμη σε περιπτώσεις όπου η επιλογή χαρακτηριστικών είναι σημαντική, καθώς τείνει να μηδενίζει τα λιγότερο σημαντικά χαρακτηριστικά.

Μαθηματική Έκφραση για την L1 Regularization:

$$L1 \text{ Regularization} = \lambda \sum_{j=0}^M |w_j|$$

Η οποία προστίθεται στη συνάρτηση κόστους και, επομένως, η συνάρτηση κόστους διαμορφώνεται ως εξής:

$$Cost = \sum_{i=0}^N (y_i - \sum_{j=0}^M x_{ij} W_j)^2 + \lambda \sum_{j=0}^M |w_j|$$

Όπου ο πρώτος όρος, $\sum_{i=0}^N (y_i - \sum_{j=0}^M x_{ij} W_j)^2$, είναι η συνάρτηση απώλειας. Με y_i την πραγματική ετικέτα, x_{ij} την τιμή χαρακτηριστικού του j -οστού χαραλτηριστικού για το i -οστό παράδειγμα, W_j το βάρος που σχετίζεται με το j -οστό χαρακτηριστικό του μοντέλου, N τον αριθμό δειγμάτων στο σύνολο δεδομένων, M τον αριθμό χαρακτηριστικών στο σύνολο δεδομένων και λ την υπερπαράμετρο κανονικοποίησης που ελέγχει τη δύναμη της κανονικοποίησης. Μεγαλύτερες τιμές αυξάνουν την ποινή για μεγαλύτερα βάρη.

Ο όρος L1 regularization αθροίζει τις απόλυτες τιμές των βαρών, προάγοντας την αραιότητα (πολλά βάρη γίνονται μηδέν).

L2 Regularization: Η κανονικοποίηση L2 (L2 Regularization) [29], γνωστή και ως Ridge Regularization, είναι μια άλλη τεχνική κανονικοποίησης που επιβάλλει έναν ποινικό όρο στη συνάρτηση κόστους, αλλά αντί για το απόλυτο άθροισμα, προσθέτει το τετραγωνισμένο άθροισμα των τιμών των βαρών. Αυτός ο όρος ποινής ενθαρρύνει το μοντέλο να παράγει μικρά, ομαλά βάρη, αποτέλοντας έτοι την υπερπροσαρμογή. Σε

αντίθεση με την L1 κανονικοποίηση, η L2 κανονικοποίηση τείνει να διατηρεί όλα τα χαρακτηριστικά, αλλά μειώνει τη σημασία των λιγότερο σημαντικών. Η L2 κανονικοποίηση είναι ιδανική για προβλήματα όπου η υπερπροσαρμογή αποτελεί ανησυχία, αλλά ταυτόχρονα είναι σημαντικό να διατηρηθούν όλα τα χαρακτηριστικά, μειώνοντας απλώς τη βαρύτητά τους αντί να τα εξαλείψει πλήρως.

Μαθηματική Έκφραση για την L1 Regularization:

$$L2 \text{ Regularization} = \lambda \sum_{j=0}^M W_j^2$$

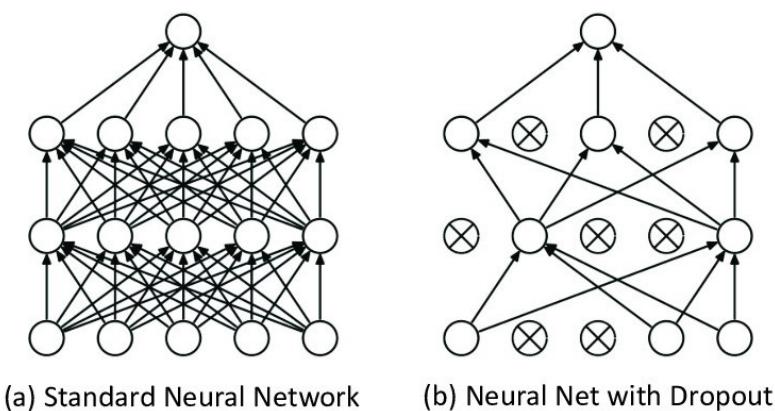
Η οποία προστίθεται στη συνάρτηση κόστους και, επομένως, η συνάρτηση κόστους διαμορφώνεται ως εξής:

$$Cost = \sum_{i=0}^N (y_i - \sum_{j=0}^M x_{ij} W_j)^2 + \lambda \sum_{j=0}^M W_j^2$$

Όπου ο πρώτος όρος, $\sum_{i=0}^N (y_i - \sum_{j=0}^M x_{ij} W_j)^2$, είναι η συνάρτηση απώλειας. Με y_i την πραγματική ετικέτα, x_{ij} την τιμή χαρακτηριστικού του j -οστού χαραλτηριστικού για το i -οστό παράδειγμα, W_j το βάρος που σχετίζεται με το j -οστό χαρακτηριστικό του μοντέλου, N τον αριθμό δειγμάτων στο σύνολο δεδομένων, M τον αριθμό χαρακτηριστικών στο σύνολο δεδομένων και λ την υπερπαράμετρο κανονικοποίησης που καθορίζει την ποσότητα της ποινής που προστίθεται στη συνάρτηση κόστους με βάση τα βάρη.

Ο όρος L2 regularization αθροίζει τα τετράγωνα των βαρών, ποινικοποιώντας τα μεγαλύτερα βάρη, αλλά δεν τα αναγκάζει να γίνουν μηδέν.

Dropout: Η τεχνική Dropout [21] είναι μια ακόμα μέθοδος κανονικοποίησης που χρησιμοποιείται για την αποφυγή υπερπροσαρμογής στα νευρωνικά δίκτυα. Κατά τη διάρκεια της εκπαίδευσης, τυχαία "απενεργοποιούνται" ορισμένοι νευρώνες σε κάθε επίπεδο, με την πιθανότητα που καθορίζεται από μία υπερπαράμετρο που ονομάζεται ρυθμός dropout (dropout rate). Με αυτόν τον τρόπο, το δίκτυο δεν βασίζεται υπερβολικά σε συγκεκριμένους νευρώνες, αλλά μαθαίνει να διαχέει την πληροφορία σε όλους τους νευρώνες του δικτύου. Το αποτέλεσμα είναι ένα πιο γενικευμένο μοντέλο που είναι λιγότερο επιρρεπές στην υπερεκπαίδευση και πιο ικανό να χειρίζεται άγνωστα δεδομένα.

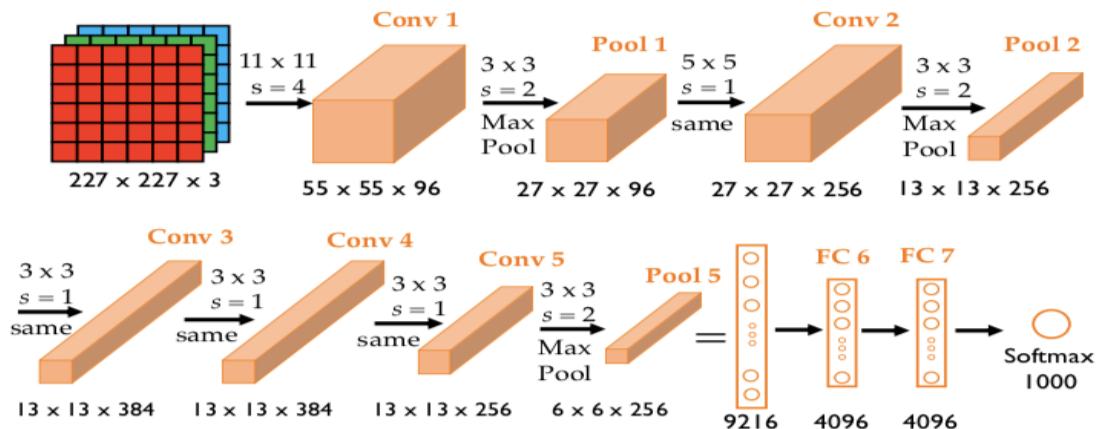


Εικόνα 1.3.3 - 1: Παράδειγμα Δικτύου με και χωρίς Dropout. Η εικόνα δείχνει ένα αρχικό νευρωνικό δίκτυο (αριστερά) και το ίδιο δίκτυο αφού έχουν εφαρμοστεί τεχνικές Dropout, με μερικούς νευρώνες να έχουν απενεργοποιηθεί τυχαία (δεξιά). (Πηγή: Medium)

1.3.4 Σύγχρονα Μοντέλα CNNs

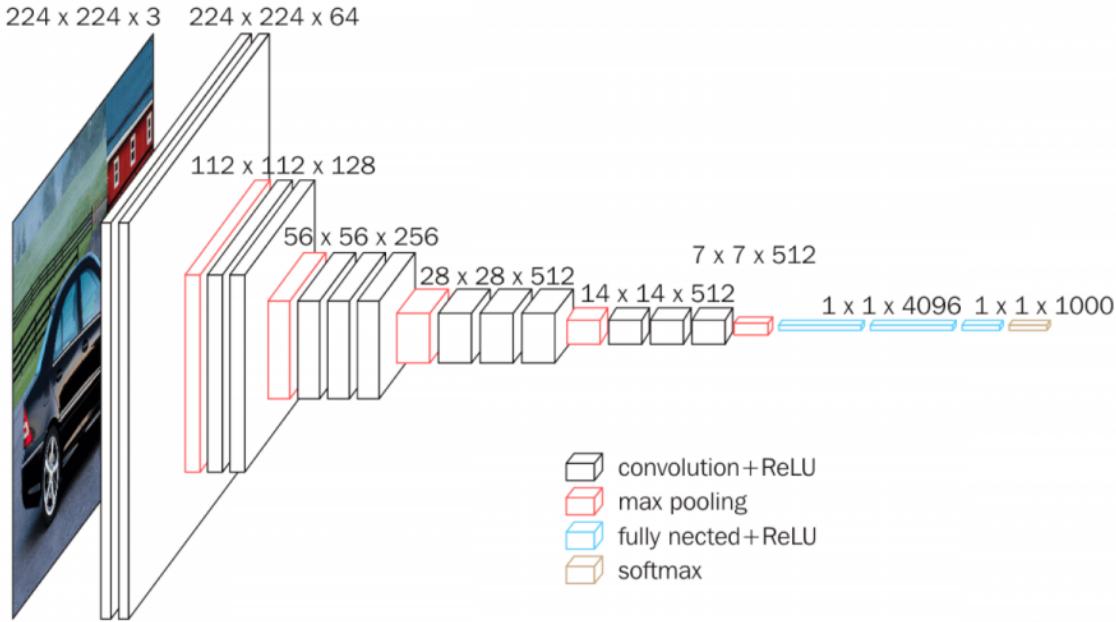
Τα CNNs έχουν εξελιχθεί σημαντικά τα τελευταία χρόνια, με την ανάπτυξη νέων και πιο αποδοτικών αρχιτεκτονικών. Παρακάτω παρουσιάζονται μερικά από τα πιο σημαντικά μοντέλα CNNs, με έμφαση στο ResNet, που έχει επηρεάσει σημαντικά τον τομέα της επεξεργασίας εικόνας (το ResNet θα χρησιμοποιηθεί στο Πειραματικό Μέρος).

AlexNet: Το AlexNet [2] είναι ένα από τα πρώτα μοντέλα CNN που ανέδειξαν τη δύναμη των Συνελικτικών Νευρωνικών Δικτύων στη μηχανική μάθηση και την επεξεργασία εικόνας. Δημιουργήθηκε από τον Alex Krizhevsky και την ομάδα του το 2012 και κέρδισε τον διαγωνισμό ImageNet, σημειώνοντας τεράστια επιτυχία. Το AlexNet αποτελείται από 8 επίπεδα, εκ των οποίων τα 5 είναι συνελικτικά και τα 3 πλήρως συνδεδεμένα (fully connected). Χρησιμοποιεί επίσης τεχνικές όπως η κανονικοποίηση με χρήση του ReLU (Rectified Linear Unit) και το Dropout για τη μείωση του overfitting. Αυτό το μοντέλο έθεσε τις βάσεις για την ανάπτυξη πιο προηγμένων CNNs.



Εικόνα 1.3.4 - 1: Αρχιτεκτονική AlexNet
(Πηγή: DevGenius)

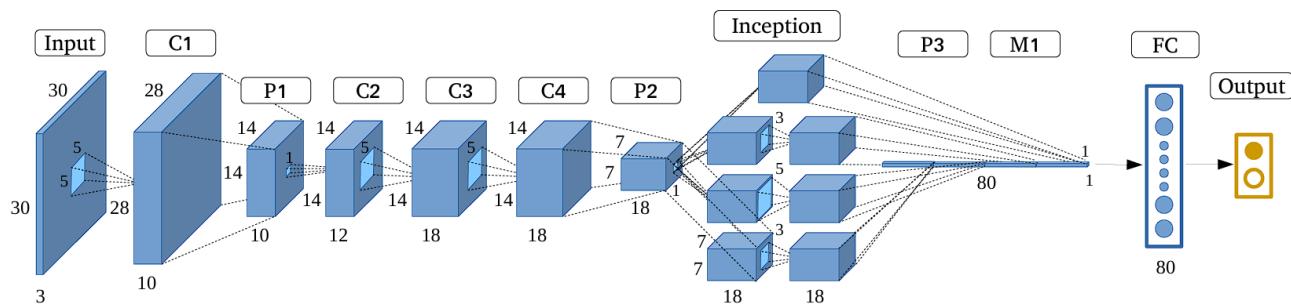
VGGNet: Το VGGNet [3] είναι ένα άλλο σημαντικό μοντέλο CNN που αναπτύχθηκε από την ομάδα του Visual Geometry Group (VGG) στο Πανεπιστήμιο της Οξφόρδης. Το VGGNet εισήγαγε την ιδέα της χρήσης μικρών φίλτρων 3x3 σε όλο το δίκτυο, σε συνδυασμό με την αύξηση του βάθους του δικτύου. Η απλότητα της αρχιτεκτονικής του, με τη διαδοχική τοποθέτηση συνελικτικών επιπέδων, το έκανε ιδιαίτερα επιτυχημένο σε διάφορες εφαρμογές επεξεργασίας εικόνας. Το VGGNet βοήθησε στην κατανόηση του ρόλου του βάθους των δικτύων στην επίτευξη υψηλής απόδοσης.



Εικόνα 1.3.4 - 2: Αρχιτεκτονική VGGNet
(Πηγή: ResearchGate)

GoogLeNet (Inception): Το GoogLeNet, γνωστό και ως Inception [5], είναι ένα άλλο σημαντικό μοντέλο CNN που αναπτύχθηκε από την ομάδα της Google. Η αρχιτεκτονική Inception χρησιμοποιεί πολλαπλά φίλτρα διαφορετικών μεγεθών στο ίδιο επίπεδο, επιτρέποντας την εξαγωγή χαρακτηριστικών σε διαφορετικές κλίμακες.

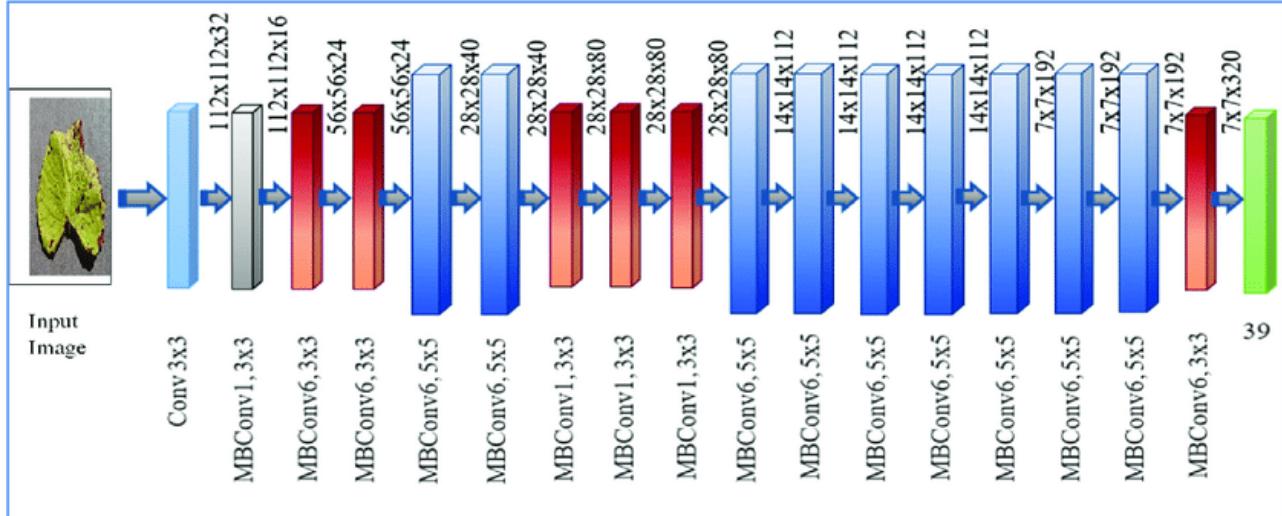
Αυτή η προσέγγιση επιτρέπει στο δίκτυο να μαθαίνει πλούσιες και ποικιλόμορφες αναπαραστάσεις των δεδομένων, χωρίς να αυξάνει δραματικά τον αριθμό των παραμέτρων. Το GoogLeNet κέρδισε τον διαγωνισμό ImageNet το 2014 (Το ImageNet Competition είναι ένας ετήσιος διαγωνισμός αναγνώρισης εικόνων που συμβάλλει στην εξέλιξη των τεχνικών τεχνητής νοημοσύνης), δείχνοντας ότι η προσέγγιση της Inception είναι αποτελεσματική για πολύπλοκες εργασίες αναγνώρισης εικόνας.



Εικόνα 1.3.4 - 3: Αρχιτεκτονική GoogleNet (Inception)
(Πηγή: ResearchGate)

EfficientNet: Το EfficientNet, γνωστό και ως EffNet [6], είναι ένα νεότερο μοντέλο CNN που αναπτύχθηκε με στόχο τη βελτιστοποίηση της απόδοσης σε σχέση με τον αριθμό των παραμέτρων και την υπολογιστική ισχύ που απαιτείται. Το EfficientNet εισάγει την ιδέα του compound scaling, όπου το βάθος, το πλάτος και η ανάλυση εισόδου του δικτύου αυξάνονται με έναν συντονισμένο τρόπο, διατηρώντας την ισορροπία

μεταξύ της απόδοσης και της υπολογιστικής αποδοτικότητας. Το EfficientNet έχει δείξει εξαιρετικά αποτελέσματα σε διάφορα benchmarks και έχει χρησιμοποιηθεί ευρέως σε εφαρμογές όπου η αποδοτικότητα είναι κρίσμη.

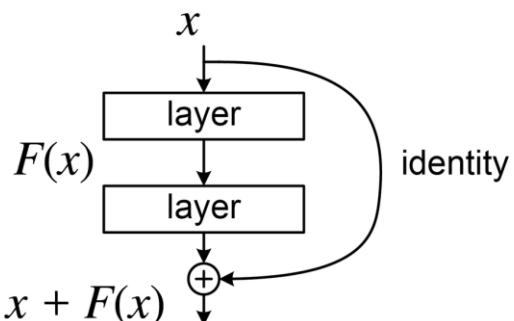


Εικόνα 1.3.4 - 4: Αρχιτεκτονική EfficientNet (Συγκεκριμένα EfficientNet B7)
(Πηγή: ResearchGate)

ResNet (Residual Networks): Το ResNet (Residual Network) [4] είναι μια από τις πιο σημαντικές και επιδραστικές αρχιτεκτονικές Συνελικτικών Νευρωνικών Δικτύων, το οποίο εισήχθη από τους Kaiming He, Xiangyu Zhang, Shaoqing Ren, και Jian Sun το 2015. Το ResNet κέρδισε την πρώτη θέση στο διαγωνισμό ImageNet το 2015 και είναι γνωστό για την ικανότητά του να εκπαιδεύει πολύ βαθιά δίκτυα, κάτι που προηγουμένως θεωρούνταν εξαιρετικά δύσκολο λόγω του φαινομένου της εξαφάνισης του gradient (vanishing gradient problem).

Residual Learning (Η βασική ιδέα του ResNet): Το κύριο καινοτόμο στοιχείο του ResNet είναι η εισαγωγή του "residual learning", μια τεχνική που διευκολύνει την εκπαίδευση πολύ βαθιών δικτύων. Η βασική ιδέα πίσω από το residual learning είναι η χρήση "skip connections" (ή "shortcut connections") που παρακάμπτουν ένα ή περισσότερα επίπεδα του δικτύου. Με αυτόν τον τρόπο, το δίκτυο μαθαίνει ένα "residual" ή υπόλοιπο, δηλαδή τη διαφορά μεταξύ της επιθυμητής εξόδου και της αρχικής εισόδου.

Skip Connections: Οι skip connections επιτρέπουν στο gradient να ρέει πίσω, μέσα στο δίκτυο, χωρίς να περνάει από όλα τα επίπεδα, κάτι που μειώνει τον κίνδυνο εξαφάνισης (Vanishing) ή έκρηξης (Exploding) του gradient. Αυτό επιτρέπει την αποτελεσματική εκπαίδευση πολύ βαθιών δικτύων, με εκατοντάδες ή ακόμα και χιλιάδες επίπεδα, κάτι που προηγουμένως ήταν ανέφικτο.

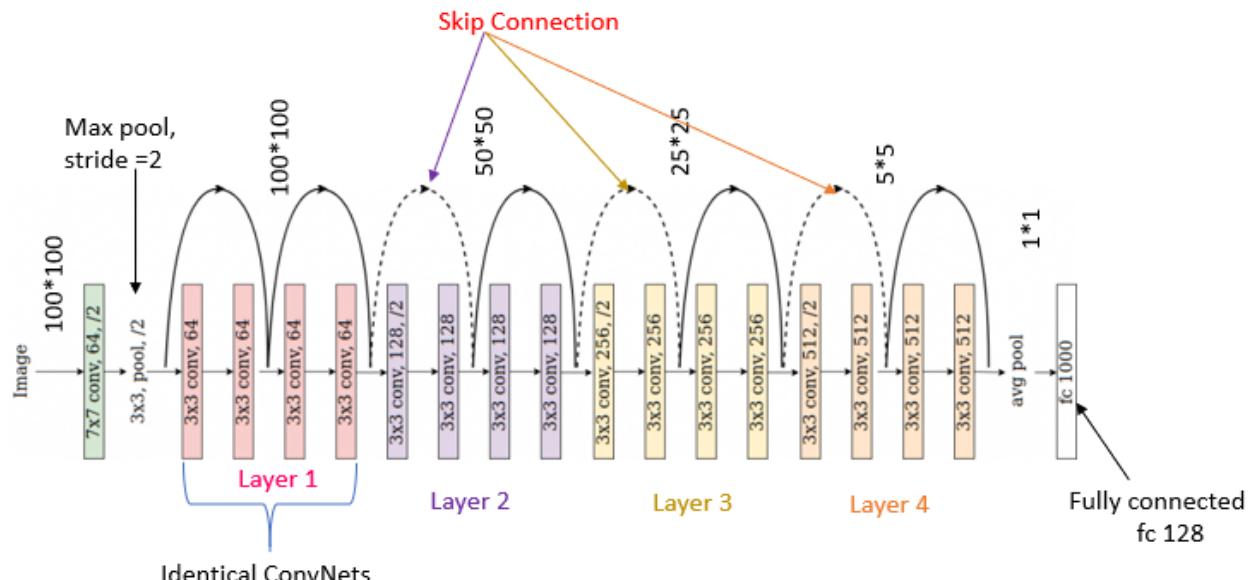


Εικόνα 1.3.4 - 5: Residual learning: a building block
(Πηγή: He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).)

Αρχιτεκτονική του ResNet: Η αρχιτεκτονική του ResNet αποτελείται από πολλά residual blocks, το καθένα από τα οποία περιλαμβάνει δύο ή τρία συνελικτικά επίπεδα, τα οποία ακολουθούνται από μια προσθετική συνάρτηση που ενσωματώνει την είσοδο του block με την έξοδο των συνελικτικών επιπέδων. Αυτά τα blocks διατάσσονται σε βάθος για να δημιουργήσουν το πλήρες δίκτυο.

Παραλλαγές του ResNet: Το αρχικό ResNet έχει πολλές παραλλαγές, ανάλογα με τον αριθμό των επιπέδων:

- ResNet-18: Έχει 18 επίπεδα.
- ResNet-34: Έχει 34 επίπεδα.
- ResNet-50: Χρησιμοποιεί ένα πιο βαθύ δίκτυο με 50 επίπεδα, εισάγοντας bottleneck blocks για τη μείωση της πολυπλοκότητας.
- ResNet-101 και ResNet-152: Έχουν 101 και 152 επίπεδα αντίστοιχα, επιτρέποντας ακόμα μεγαλύτερο βάθος στο δίκτυο.



Εικόνα 1.3.4 - 6: Παράδειγμα ResNet-18

(Πηγή: ResearchGate)

Πλεονεκτήματα και Χρήσεις του ResNet: Το ResNet επέτρεψε την εκπαίδευση δικτύων με πρωτοφανή βάθη, τα οποία έχουν αποδειχθεί εξαιρετικά αποτελεσματικά σε μια ευρεία γκάμα εφαρμογών, όπως η αναγνώριση εικόνων, η ανίχνευση αντικειμένων, και η επεξεργασία φυσικής γλώσσας. Το ResNet έχει επίσης αποτελέσει τη βάση για πολλές άλλες σύγχρονες αρχιτεκτονικές CNNs, και παραμένει ένα πρότυπο για τη σχεδίαση και την ανάπτυξη νέων, ακόμα πιο προηγμένων μοντέλων.

1.3.5 Αυτο-κωδικοποιητές

Οι Αυτο-κωδικοποιητές (Auto-encoders) [30] αποτελούν ένα είδος νευρωνικών δικτύων που χρησιμοποιούνται κυρίως για τη μάθηση αποδοτικών αναπαραστάσεων των δεδομένων, συχνά για σκοπούς μείωσης διαστάσεων ή συμπίεσης δεδομένων. Ο βασικός στόχος ενός autoencoder είναι να μάθει να αναδημιουργεί την είσοδό του στην έξοδο, μέσω ενός περιορισμένου χώρου αναπαράστασης, γνωστού ως "latent space" (Οι autoencoders θα χρησιμοποιηθούν σε κάποιες από τις έμμεσες διεργασίες στο Πειραματικό Μέρος).

Βασική Αρχιτεκτονική: Η αρχιτεκτονική ενός autoencoder αποτελείται από τρία κύρια μέρη:

- Κωδικοποιητής (Encoder):** Ο κωδικοποιητής είναι το τμήμα του δικτύου που μετατρέπει την είσοδο σε μια συμπιεσμένη αναπαράσταση, περιορισμένων διαστάσεων. Αυτό επιτυγχάνεται μέσω ενός ή περισσότερων επιπέδων νευρώνων, τα οποία εκτελούν συναρτήσεις όπως συνελίξεις (στην περίπτωση συνελικτικών autoencoders) ή πλήρως συνδεδεμένων επιπέδων. Ο στόχος του κωδικοποιητή είναι να διατηρήσει τα πιο σημαντικά χαρακτηριστικά των δεδομένων στην συμπιεσμένη μορφή τους.
- Latent Space:** Το latent space ή κρυφό επίπεδο (hidden layer) είναι ο χώρος όπου αποθηκεύεται η συμπιεσμένη αναπαράσταση των δεδομένων. Αυτός ο χώρος έχει χαμηλότερες διαστάσεις από τα αρχικά δεδομένα και περιέχει την "ουσία" των δεδομένων, διατηρώντας τα βασικά τους χαρακτηριστικά. Το μέγεθος του latent space καθορίζει την ικανότητα του δικτύου να συμπιέζει τα δεδομένα χωρίς απώλειες.
- Αποκωδικοποιητής (Decoder):** Ο αποκωδικοποιητής είναι το τμήμα του δικτύου που αναλαμβάνει να αναδημιουργήσει την αρχική είσοδο από την συμπιεσμένη αναπαράσταση. Ο αποκωδικοποιητής αντιστρέφει τη διαδικασία του κωδικοποιητή, αυξάνοντας σταδιακά τις διαστάσεις της αναπαράστασης μέσω των επιπέδων νευρώνων. Ο στόχος του αποκωδικοποιητή είναι να παράγει μια έξοδο όσο το δυνατόν πιο πιστή στην αρχική είσοδο.

Εκπαίδευση: Οι autoencoders εκπαιδεύονται χρησιμοποιώντας μια συνάρτηση απώλειας που συγκρίνει την έξοδο του αποκωδικοποιητή με την αρχική είσοδο. Η πιο κοινή συνάρτηση απώλειας είναι το Μέσο Τετραγωνικό Σφάλμα (Mean Squared Error - MSE, ο ορισμός του υπάρχει στην ενότητα 1.1.2), το οποίο μετρά την απόκλιση μεταξύ της αναδημιουργημένης εξόδου και της αρχικής εισόδου. Ο στόχος της εκπαίδευσης είναι η ελαχιστοποίηση αυτής της απώλειας, διασφαλίζοντας ότι το δίκτυο μπορεί να αναδημιουργήσει την είσοδο όσο το δυνατόν πιο πιστά.

Είδη Αυτο-κωδικοποιητών:

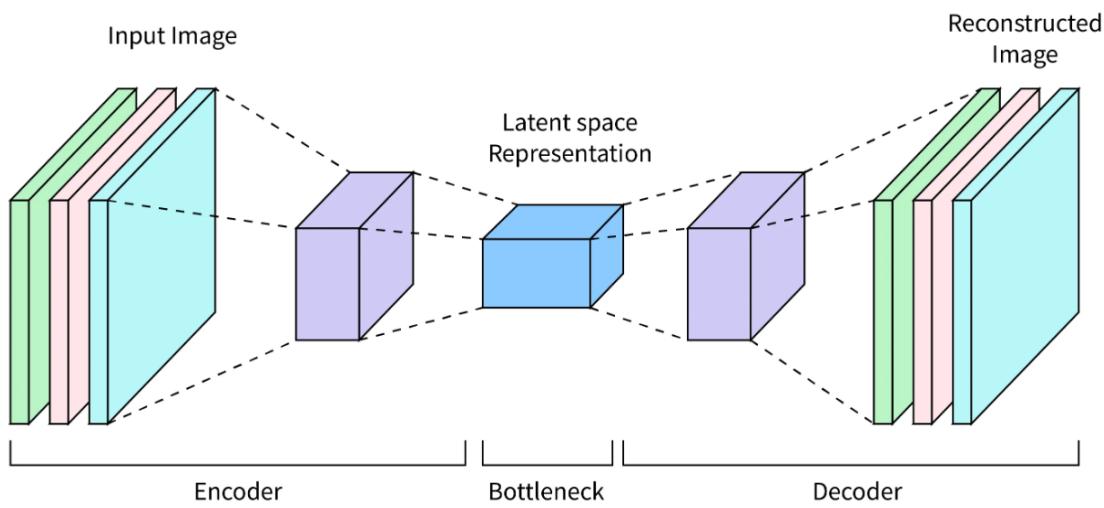
Βαθιοί Αυτο-κωδικοποιητές (Deep Autoencoders): Οι βαθιοί autoencoders [31] είναι μια εξέλιξη των βασικών autoencoders, διαθέτοντας πολλαπλά κρυφά επίπεδα τόσο στον κωδικοποιητή όσο και στον αποκωδικοποιητή. Αυτό τους επιτρέπει να μάθουν πιο σύνθετες και υψηλότερης διάστασης αναπαραστάσεις των δεδομένων. Η αυξημένη πολυπλοκότητα των μοντέλων αυτών τους δίνει τη δυνατότητα να διαχειρίζονται καλύτερα δεδομένα με περίπλοκες δομές, όπως εικόνες υψηλής ανάλυσης ή πολυδιάστατα δεδομένα.

Συνελικτικοί Αυτο-κωδικοποιητές (Convolutional Autoencoders): Οι συνελικτικοί autoencoders [32] ενσωματώνουν συνελικτικά επίπεδα τόσο στον κωδικοποιητή όσο και στον αποκωδικοποιητή τους. Αυτή η αρχιτεκτονική είναι ιδιαίτερα κατάλληλη για την επεξεργασία δεδομένων εικόνας, καθώς οι συνελικτικοί πυρήνες επιτρέπουν την εξαγωγή τοπικών χαρακτηριστικών και τη διατήρηση της χωρικής δομής των δεδομένων. Αυτό καθιστά τους συνελικτικούς autoencoders εξαιρετικά αποτελεσματικούς για εφαρμογές όπως η ανάλυση εικόνων, η αφαίρεση θορύβου και η συμπίεση εικόνας.

Αραιοί Αυτο-κωδικοποιητές (Sparse Autoencoders): Οι αραιοί autoencoders [33] εισάγουν έναν επιπλέον όρο στη συνάρτηση απώλειας που επιδιώκει την αραιώση στο latent space, δηλαδή, διασφαλίζουν ότι οι περισσότεροι νευρώνες θα έχουν τιμές κοντά στο μηδέν. Αυτή η αραιώση προάγει τη γενίκευση του μοντέλου και μειώνει την πιθανότητα overfitting. Οι αραιοί autoencoders είναι ιδιαίτερα χρήσιμοι όταν τα δεδομένα έχουν πολλαπλές διαστάσεις και η εξαγωγή χαρακτηριστικών με μικρότερη σημασία είναι επιθυμητή.

Θορυβώδεις Αυτο-κωδικοποιητές (Denoising Autoencoders): Οι θορυβώδεις autoencoders [34] εκπαιδεύονται για να αναδημιουργούν την αρχική, καθαρή είσοδο από μια θορυβώδη εκδοχή της. Κατά τη διάρκεια της εκπαίδευσης, εισάγεται σκόπιμα θόρυβος στα δεδομένα εισόδου, και το μοντέλο μαθαίνει να αφαιρεί αυτόν τον θόρυβο, αναδημιουργώντας την αρχική εικόνα ή τα δεδομένα. Αυτή η τεχνική είναι εξαιρετικά χρήσιμη για την αφαίρεση θορύβου από δεδομένα εικόνας ή ήχου και για τη βελτίωση της ποιότητας των δεδομένων.

Αυτο-κωδικοποιητές Μεταβλητών (Variational Autoencoders - VAEs): Οι VAEs [35] είναι μια προχωρημένη μορφή autoencoder, η οποία εισάγει στατιστικές ιδιότητες στον latent space. Συγκεκριμένα, οι VAEs επιδιώκουν να μετατρέψουν τον latent space σε μια στατιστική κατανομή (π.χ., κανονική κατανομή), επιτρέποντας τη δειγματοληψία νέων, παρόμοιων δεδομένων που μιμούνται τα δεδομένα εκπαίδευσης. Αυτό τους καθιστά χρήσιμους για τη δημιουργία συνθετικών δεδομένων, την παραγωγή εικόνων και την κατανόηση των εσωτερικών δομών των δεδομένων.



Εικόνα 1.3.5 - 1: Διάγραμμα ενός Autoencoder, που απεικονίζει τη διαδικασία κωδικοποίησης της αρχικής εικόνας (Input Image) σε έναν συμπιεσμένο χωρικό χώρο αναπαράστασης (Latent Space) και στη συνέχεια την αναδημιουργία της εικόνας (Reconstructed Image) μέσω του αποκωδικοποιητή (Decoder). (Πηγή: Scaler)

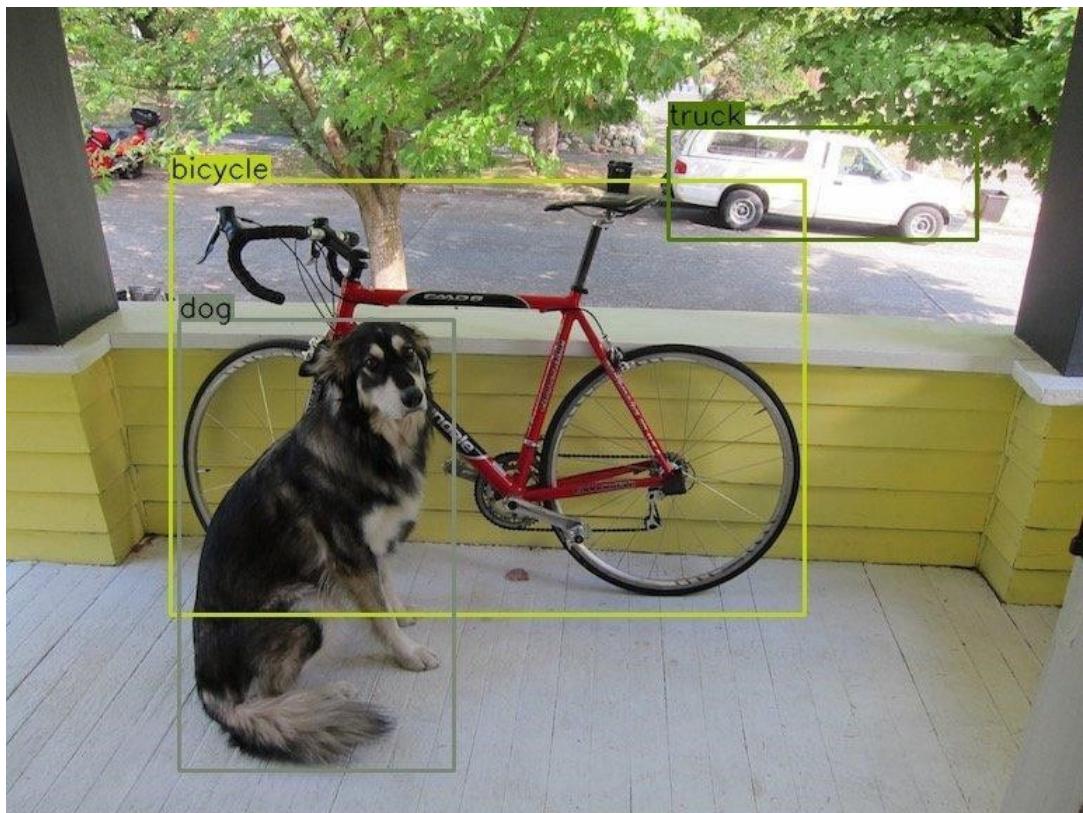
Εφαρμογές: Οι autoencoders έχουν ευρεία χρήση σε διάφορες εφαρμογές:

- **Μείωση Διαστάσεων (Dimensionality Reduction):** Χρησιμοποιούνται για να συμπιέσουν δεδομένα σε χαμηλότερων διαστάσεων αναπαραστάσεις, διατηρώντας όσο το δυνατόν περισσότερη πληροφορία, γεγονός που τους καθιστά χρήσιμους για την οπτικοποίηση δεδομένων και την προεπεξεργασία πριν από άλλους αλγορίθμους μηχανικής μάθησης [30].
- **Ανίχνευση Ανωμαλιών (Anomaly Detection):** Χρησιμοποιούνται για την ανίχνευση ανωμαλιών σε δεδομένα, όπως στην ασφάλεια ή στη βιομηχανική συντήρηση, όπου το δίκτυο μαθαίνει την κανονική συμπεριφορά και μπορεί να αναγνωρίσει αποκλίσεις [30,36].
- **Αφαίρεση Θορύβου (Denoising):** Χρησιμοποιούνται για την αφαίρεση θορύβου από δεδομένα εικόνας, βίντεο ή ήχου, βελτιώνοντας την ποιότητα των δεδομένων [30,34].
- **Γενετική Μοντελοποίηση (Generative Modeling):** Οι κωδικοποιητές μεταβλητών (VAEs) χρησιμοποιούνται για τη δημιουργία νέων, συνθετικών δεδομένων, όπως εικόνες, που είναι στατιστικά παρόμοια με τα δεδομένα εκπαίδευσης [35].

1.3.6 Εφαρμογές των CNNs

Τα Συνελικτικά Νευρωνικά Δίκτυα έχουν φέρει επανάσταση σε πολλούς τομείς της τεχνολογίας και της επιστήμης, χάρη στην ικανότητά τους να εξάγουν και να αναλύουν χαρακτηριστικά από δεδομένα με τρόπο που προσομοιάζει την ανθρώπινη αντίληψη. Ειδικότερα, τα CNNs έχουν γίνει η κυρίαρχη τεχνολογία για την επεξεργασία και την κατανόηση εικόνων και βίντεο, ενώ βρίσκουν εφαρμογή και σε τομείς όπως η επεξεργασία φυσικής γλώσσας, η ιατρική απεικόνιση και τα αυτόνομα συστήματα. Στην παρούσα ενότητα, θα εξετάσουμε τις βασικές εφαρμογές των CNNs, αναδεικνύοντας την ευελιξία και την απόδοσή τους σε ένα ευρύ φάσμα πραγματικών προβλημάτων.

- **Αναγνώριση Εικόνας (Image Recognition):** Η αναγνώριση εικόνας είναι μία από τις πιο κοινές εφαρμογές των CNNs. Σε αυτή την περίπτωση, τα CNNs χρησιμοποιούνται για την αναγνώριση και την κατηγοριοποίηση αντικειμένων μέσα σε εικόνες. Ένα από τα πιο γνωστά παραδείγματα είναι το ImageNet, ένας διαγωνισμός στον οποίο μοντέλα CNNs έχουν επιδείξει εξαιρετική ακρίβεια στην αναγνώριση χιλιάδων διαφορετικών κατηγοριών αντικειμένων [37].
- **Ανίχνευση Αντικειμένων (Object Detection):** Η ανίχνευση αντικειμένων επεκτείνει την αναγνώριση εικόνας, προσθέτοντας τη δυνατότητα όχι μόνο να εντοπίζονται αλλά και να τοποθετούνται τα αντικείμενα σε συγκεκριμένες θέσεις μέσα σε μια εικόνα. Τεχνικές όπως το YOLO (You Only Look Once) και το Faster R-CNN [38,39] είναι κλασικά παραδείγματα εφαρμογών των CNNs στον τομέα αυτό, επιτρέποντας την ταυτόχρονη ανίχνευση πολλαπλών αντικειμένων σε πραγματικό χρόνο.
- **Επεξεργασία Φυσικής Γλώσσας (Natural Language Processing - NLP):** Αν και τα CNNs είναι περισσότερο γνωστά για την επεξεργασία εικόνας, χρησιμοποιούνται επίσης και στην επεξεργασία φυσικής γλώσσας [40]. Για παράδειγμα, τα CNNs μπορούν να χρησιμοποιηθούν για την ανάλυση συναισθήματος, την αναγνώριση οντοτήτων, και την κατηγοριοποίηση κειμένων. Σε αυτό το πλαίσιο, τα CNNs μαθαίνουν από τη δομή των προτάσεων και των λέξεων για να εξάγουν χαρακτηριστικά που είναι κρίσιμα για τη σωστή κατηγοριοποίηση ή ανάλυση του κειμένου.
- **Ιατρική Απεικόνιση (Medical Imaging):** Στον τομέα της ιατρικής, τα CNNs έχουν βρει ευρεία εφαρμογή στην ανάλυση και την ερμηνεία ιατρικών εικόνων, όπως οι ακτινογραφίες, οι μαγνητικές τομογραφίες και οι υπέρηχοι [41]. Τα CNNs μπορούν να βοηθήσουν στον εντοπισμό ανωμαλιών, όπως όγκων ή άλλων παθολογικών καταστάσεων, με ακρίβεια που συχνά υπερβαίνει αυτή των ανθρώπινων ειδικών.
- **Αυτόνομα Οχήματα (Autonomous Vehicles):** Τα αυτόνομα οχήματα χρησιμοποιούν CNNs για την ανάλυση δεδομένων από κάμερες και άλλους αισθητήρες, επιτρέποντας την αναγνώριση και την ερμηνεία των συνθηκών του περιβάλλοντος, όπως η αναγνώριση πινακίδων οδικής σήμανσης, η ανίχνευση πεζών και άλλων οχημάτων, και η λήψη αποφάσεων σε πραγματικό χρόνο [42].
- **Δημιουργία Τέχνης (Art Generation):** Τα CNNs μπορούν να χρησιμοποιηθούν στη δημιουργία τέχνης, όπως η δημιουργία νέων εικόνων ή η μετατροπή του στυλ μιας εικόνας (style transfer). Μέσω τεχνικών όπως οι Generative Adversarial Networks (GANs) [43], τα CNNs έχουν χρησιμοποιηθεί για τη δημιουργία έργων τέχνης που μοιάζουν να έχουν δημιουργηθεί από ανθρώπους καλλιτέχνες.
- **Ανάλυση Βίντεο (Video Analysis):** Στην ανάλυση βίντεο, τα CNNs χρησιμοποιούνται για την αναγνώριση και την κατηγοριοποίηση αντικειμένων, την ανίχνευση κινήσεων και την ερμηνεία σκηνών. Οι εφαρμογές αυτές είναι σημαντικές για την παρακολούθηση ασφαλείας, την ανάλυση αθλητικών αγώνων, και άλλες εφαρμογές όπου η κατανόηση του περιεχομένου βίντεο είναι κρίσιμη [44].



Εικόνα 1.3.6 - 1: Χρήση Συνελικτικών Νευρωνικών Δικτύων για Ανίχνευση Αντικεμένων.
(Πηγή: TowardsDataScience)

Κεφάλαιο 2^ο

Αυτο-Εποπτευόμενη Μάθηση και Έμμεσες Διεργασίες

2.1 Η Θεωρία της Αυτο-εποπτευόμενης Μάθησης

2.1.1 Ορισμός και Θεμελιώδεις Αρχές

Η Αυτο-εποπτευόμενη Μάθηση αποτελεί μια σύγχρονη προσέγγιση στο πεδίο της μηχανικής μάθησης, που έχει κερδίσει σημαντικό ύδαφος τα τελευταία χρόνια λόγω της ικανότητάς της να εκπαιδεύει μοντέλα χωρίς την ανάγκη για μεγάλα σύνολα επισημασμένων δεδομένων. Σε αυτή την προσέγγιση, τα μοντέλα χρησιμοποιούν τα ίδια τα δεδομένα για να δημιουργήσουν σήματα εποπτείας, αντικαθιστώντας την παραδοσιακή ανάγκη για εξωτερικές ετικέτες που παρέχονται από ανθρώπους. Με αυτό τον τρόπο, η *SSL* γεφυρώνει το χάσμα μεταξύ της επιβλεπόμενης και της μη επιβλεπόμενης μάθησης, προσφέροντας μια λύση που επιτρέπει στα μοντέλα να μαθαίνουν χρήσιμες αναπαραστάσεις από μη επισημασμένα δεδομένα.

Βασική ιδέα της Αυτο-εποπτευόμενης Μάθησης είναι η εκμετάλλευση των εγγενών δομών και σχέσεων που υπάρχουν στα δεδομένα εισόδου. Αυτές οι σχέσεις μπορεί να περιλαμβάνουν την τάξη των λέξεων σε μια πρόταση, τις γωνίες και τις σκιές σε μια εικόνα, ή τη συνέχεια των ήχων σε ένα ηχητικό σήμα. Αντί να βασίζεται σε προκαθορισμένα *labels*, το μοντέλο δημιουργεί τα δικά του εποπτευόμενα σήματα (*supervisory signals*) μέσω της εκμετάλλευσης αυτών των εγγενών χαρακτηριστικών. Αυτό το πλαίσιο μάθησης επιτρέπει στα μοντέλα να αποκτούν βαθιά κατανόηση των δεδομένων, που μπορεί να μεταφερθεί σε άλλες, πιο εξειδικευμένες εργασίες.

Στην πρακτική εφαρμογή της Αυτο-εποπτευόμενης Μάθησης, τα δεδομένα τροποποιούνται σκόπιμα με την εισαγωγή διαφόρων μετασχηματισμών, όπως προσθήκη θορύβου, αποκοπή, περιστροφή, ή εφαρμογή δυαδικών μασκών που αποκρύπτουν τμήματα των δεδομένων. Για παράδειγμα, σε μια εικόνα μπορεί να προστεθεί θόρυβος ή να αποκρυφθούν τμήματά της, δημιουργώντας ένα "κατεστραμμένο" δεδομένο. Το τροποποιημένο αυτό δεδομένο στη συνέχεια χρησιμοποιείται ως είσοδος στο μοντέλο, ενώ η αρχική, μη αλλοιωμένη έκδοση της εικόνας χρησιμοποιείται ως ετικέτα. Ο στόχος του μοντέλου είναι να μάθει να ανακατασκευάζει ή να προβλέπει το αρχικό, μη αλλοιωμένο δεδομένο από την αλλοιωμένη έκδοση. Μέσα από αυτή τη διαδικασία, το μοντέλο αναγκάζεται να κατανοήσει και να απομονώσει τις πιο σημαντικές ιδιότητες και χαρακτηριστικά των δεδομένων, κάτι που το βοηθά να γενικεύσει καλύτερα και να επιτυγχάνει υψηλότερες επιδόσεις όταν εφαρμόζεται σε νέα, μη επισημασμένα δεδομένα.

Ένας σημαντικός στόχος της Αυτο-εποπτευόμενης Μάθησης είναι να ανακαλύπτει και να μαθαίνει αναπαραστάσεις που είναι τόσο πλούσιες και γενικεύσιμες, ώστε να μπορούν να χρησιμοποιηθούν σε άλλα προβλήματα ή *tasks*, γνωστά ως *downstream tasks*, που συχνά απαιτούν επισημασμένα δεδομένα για την επίτευξη υψηλής απόδοσης. Για παράδειγμα, αφού εκπαιδευτεί το μοντέλο με *SSL*, μπορεί να

χρησιμοποιηθεί σε προβλήματα όπως η ταξινόμηση εικόνων, η αναγνώριση αντικειμένων, ή η ανάλυση κειμένου, με μικρή ή καθόλου επιπλέον εκπαίδευση.

Αυτό το είδος μάθησης έχει αποδείξει την αποτελεσματικότητά του σε μια ποικιλία εφαρμογών, από την επεξεργασία εικόνας και ήχου μέχρι τη φυσική γλώσσα, και έχει βρει ευρεία χρήση σε πραγματικές εφαρμογές όπου η πρόσβαση σε επισημασμένα δεδομένα είναι περιορισμένη ή ακριβή. Η Αυτο-εποπτευόμενη Μάθηση αποτελεί μια κανονόμο προσέγγιση που προσφέρει ισχυρά εργαλεία για την επεξεργασία και κατανόηση δεδομένων, χωρίς να απαιτείται ανθρώπινη εποπτεία ή επισημασμένα δεδομένα. Η ικανότητα των μοντέλων να μαθαίνουν χρήσιμες αναπαραστάσεις από μη επισημασμένα δεδομένα τα καθιστά ιδιαίτερα πολύτιμα για μια σειρά από εφαρμογές, καθιστώντας τα μοντέλα πιο ευέλικτα και ικανά να προσαρμόζονται σε νέες, πιο σύνθετες εργασίες.

2.1.2 Ο Ρόλος των Έμμεσων Διεργασιών στην Αυτό-εποπτευόμενη Μάθηση

Οι έμμεσες διεργασίες διαδραματίζουν καθοριστικό ρόλο στην Αυτο-Εποπτευόμενη Μάθηση, καθώς υποχρεώνουν το μοντέλο να εξαγάγει ουσιαστικά χαρακτηριστικά από τα δεδομένα. Καθώς το μοντέλο προσπαθεί να λύσει μια έμμεση διεργασία, όπως η ανακατασκευή ενός τμήματος εικόνας ή η πρόβλεψη του επόμενου καρέ ενός βίντεο, αναγκάζεται να αναπτύξει μια ενδελεχή κατανόηση των βασικών δομών και συσχετισμών που υπάρχουν στα δεδομένα. Αυτή η βαθύτερη κατανόηση είναι θεμελιώδης για την επιτυχημένη μεταφορά της μάθησης σε downstream tasks, όπου συχνά απαιτούνται επισημασμένα δεδομένα για την επίτευξη υψηλών επιδόσεων.

Η χρήση έμμεσων διεργασιών επιτρέπει στο μοντέλο να γενικεύει καλύτερα, δηλαδή να αποδίδει καλά όχι μόνο στα δεδομένα εκπαίδευσης αλλά και σε νέα, άγνωστα δεδομένα. Αυτό επιτυγχάνεται επειδή οι έμμεσες διεργασίες οδηγούν το μοντέλο να μάθει γενικά πρότυπα και δομές που είναι κοινά στα δεδομένα, παρά να απομνημονεύσει συγκεκριμένες λεπτομέρειες. Έτσι, όταν το μοντέλο εφαρμόζεται σε διαφορετικές, πραγματικές εφαρμογές, είναι σε θέση να αποδώσει καλύτερα, ακόμα και αν τα δεδομένα της νέας εργασίας διαφέρουν από τα δεδομένα εκπαίδευσης.

Ένα από τα πιο σημαντικά πλεονεκτήματα των έμμεσων διεργασιών είναι η μείωση της εξάρτησης από μεγάλα σύνολα επισημασμένων δεδομένων. Σε πολλές εφαρμογές, η συλλογή και επισήμανση δεδομένων είναι δαπανηρή και χρονοβόρα. Οι έμμεσες διεργασίες επιτρέπουν στα μοντέλα να χρησιμοποιούν ανεπιτήρητα δεδομένα για την εκμάθηση, μειώνοντας την ανάγκη για επισημασμένα δεδομένα και καθιστώντας την όλη διαδικασία εκπαίδευσης πιο αποδοτική και οικονομικά βιώσιμη.

Τέλος, παρόλο που οι έμμεσες διεργασίες προσφέρουν σημαντικά οφέλη, η επιλογή της σωστής διεργασίας αποτελεί πρόκληση. Η αποτελεσματικότητα μιας έμμεσης διεργασίας εξαρτάται από το πόσο καλά τα χαρακτηριστικά που μαθαίνονται μπορούν να μεταφερθούν σε άλλες εργασίες. Διαφορετικές διεργασίες μπορεί να οδηγήσουν σε διαφορετικά χαρακτηριστικά, και όχι όλα τα χαρακτηριστικά είναι εξίσου χρήσιμα για όλες τις downstream tasks. Παρά τις προκλήσεις, η έρευνα στον τομέα αυτό συνεχίζει να βελτιώνει τις τεχνικές επιλογής και σχεδιασμού έμμεσων διεργασιών, ανοίγοντας το δρόμο για μοντέλα που μπορούν να μάθουν πιο αποδοτικά και να εφαρμόζουν τις γνώσεις τους σε ένα ευρύ φάσμα εφαρμογών.

2.1.3 Κατανόηση Σχέσεων με Άλλες Μορφές Μάθησης

Η Αυτο-εποπτευόμενη μάθηση συνδυάζει στοιχεία από την εποπτευόμενη και την μη εποπτευόμενη μάθηση, δημιουργώντας μια ενδιάμεση κατηγορία που αξιοποιεί τα πλεονεκτήματα και των δύο προσεγγίσεων. Αντί να απαιτεί επισημασμένα δεδομένα, όπως στην εποπτευόμενη μάθηση, η SSL χρησιμοποιεί τα μη επισημασμένα δεδομένα για να παράγει δικά της σήματα εποπτείας μέσω έμμεσων διεργασιών.

Μια σημαντική διαφορά με τη μη εποπτευόμενη μάθηση είναι ότι, ενώ και οι δύο μέθοδοι δουλεύουν με

μη επισημασμένα δεδομένα, η αυτο-εποπτευόμενη μάθηση δημιουργεί εποπτευόμενα σήματα για να κατευθύνει τη διαδικασία μάθησης. Αντίθετα, η μη εποπτευόμενη μάθηση συχνά στοχεύει στην ανακάλυψη κρυφών δομών ή ομάδων μέσα στα δεδομένα χωρίς τη χρήση οποιουδήποτε εποπτευτικού σήματος.

Σε σύγκριση με την ημι-εποπτευόμενη μάθηση, όπου ένα μέρος των δεδομένων είναι επισημασμένο και το υπόλοιπο όχι, η αυτο-εποπτευόμενη μάθηση διαφέρει επειδή δεν βασίζεται σε προϋπάρχοντα επισημασμένα δεδομένα, αλλά παράγει δικά της μέσω έμμεσων διεργασιών. Αυτή η προσέγγιση καθιστά την *SSL* ιδιαίτερα χρήσιμη όταν τα επισημασμένα δεδομένα είναι σπάνια ή δύσκολο να αποκτηθούν, δίνοντας τη δυνατότητα για εκμάθηση σε κλίμακα χωρίς την ανάγκη χειροκίνητης επισήμανσης.

2.1.4 Τεχνικές Υλοποίησης

Η υλοποίηση της *SSL* απαιτεί μια συνδυαστική προσέγγιση που περιλαμβάνει τη σωστή επιλογή δεδομένων, την επιλογή των κατάλληλων μετασχηματισμών και τη χρήση εξειδικευμένων αλγορίθμων και βιβλιοθηκών. Η διαδικασία υλοποίησης περιλαμβάνει αρχικά τον καθορισμό των στόχων της μάθησης και των εποπτευόμενων σημάτων που θα παραχθούν από τα δεδομένα. Στη συνέχεια, το μοντέλο εκπαιδεύεται με μια σειρά από προσεκτικά σχεδιασμένα βήματα που επιτρέπουν την εξαγωγή χρήσιμων αναπαραστάσεων, οι οποίες στη συνέχεια μπορούν να χρησιμοποιηθούν σε downstream tasks, όπως η ταξινόμηση ή η ανάλυση δεδομένων.

Η επιλογή των κατάλληλου περιβάλλοντος, των κατάλληλων εργαλείων και βιβλιοθηκών για την υλοποίηση της *SSL* είναι επίσης κρίσιμη, καθώς αυτά καθορίζουν την αποτελεσματικότητα και την αποδοτικότητα της διαδικασίας. Παρακάτω, θα παρουσιαστούν τα βασικά βήματα για την υλοποίηση της Αυτο-εποπτευόμενης Μάθησης, καθώς και οι πιο διαδεδομένες βιβλιοθήκες και εργαλεία που μπορούν να χρησιμοποιηθούν για αυτόν τον σκοπό.

1. Επιλογή Δεδομένων (Data Selection): Επιλογή ενός μεγάλου συνόλου μη επισημασμένων δεδομένων που θα χρησιμοποιηθούν για την εκπαίδευση του μοντέλου.
2. Επιλογή Ιδιότητας προς Πρόβλεψη (Select a Property to Predict): Καθορισμός μιας ιδιότητας των δεδομένων που το μοντέλο θα προσπαθήσει να προβλέψει. Αυτή η ιδιότητα μπορεί να είναι, για παράδειγμα, η επόμενη λέξη σε μια ακολουθία κειμένου, η κατεύθυνση ενός αντικειμένου σε μια εικόνα, ή η συμπλήρωση μιας κενής περιοχής σε μια εικόνα.
3. Δημιουργία Εποπτικών Σημάτων (Create Supervisory Signals): Μετασχηματισμός των δεδομένων ώστε να δημιουργηθούν παραδείγματα για το μοντέλο. Ένα παράδειγμα είναι η πρόσθεση θορύβου σε μια εικόνα με σκοπό να ζητηθεί από το μοντέλο να προβλέψει την καθαρή έκδοση της εικόνας. Τα μετασχηματισμένα δεδομένα θα χρησιμοποιηθούν ως είσοδος, ενώ τα αρχικά δεδομένα θα χρησιμεύσουν ως "ετικέτες".
4. Ορισμός Συνάρτησης Απώλειας (Define a Loss Function): Ορισμός μιας συνάρτησης απώλειας που θα αξιολογήσει την απόδοση του μοντέλου στην πρόβλεψη της ιδιότητας των δεδομένων. Αυτή η συνάρτηση θα καθοδηγήσει το μοντέλο στο να μάθει τις ουσιαστικές ιδιότητες των δεδομένων.
5. Εκπαίδευση του Μοντέλου (Model Training): Εκπαίδευση του μοντέλου χρησιμοποιώντας τα μετασχηματισμένα δεδομένα και τις "ετικέτες" που δημιουργήθηκαν. Χρήση ενός αλγορίθμου βελτιστοποίησης, όπως ο Stochastic Gradient Descent - SGD ή Adam, για την ελαχιστοποίηση της συνάρτηση απώλειας.
6. Επαλήθευση και Αξιολόγηση (Validate and Evaluate/Test): Αφού ολοκληρωθεί η εκπαίδευση, το μοντέλο μπορεί να επαληθευτεί χρησιμοποιώντας ένα ξεχωριστό σύνολο δεδομένων. Αυτή η

διαδικασία διασφαλίζει ότι το μοντέλο γενικεύει καλά και δεν είναι υπερπροσαρμοσμένο στα δεδομένα εκπαίδευσης, επιτρέποντας την αξιολόγηση της απόδοσής του σε νέα, μη επισημασμένα δεδομένα.

7. Fine-Tuning για Ειδικές Εργασίες (Fine-Tune for Specific Tasks): Το μοντέλο μπορεί να προσαρμοστεί σε συγκεκριμένες εργασίες με την προσθήκη labeled data και την προσαρμογή του μέσω μιας διαδικασίας fine-tuning. Αυτό επιτρέπει στο μοντέλο να μάθει εξειδικευμένα χαρακτηριστικά που σχετίζονται με τη συγκεκριμένη εργασία, βελτιώνοντας την απόδοσή του σε συγκεκριμένα tasks.
8. Εφαρμογή σε Πραγματικά Δεδομένα (Deploy on Real Data): Μετά την ολοκλήρωση του fine-tuning, το μοντέλο μπορεί να επαληθευτεί χρησιμοποιώντας το σύνολο δεδομένων αξιολόγησης (Test data) του downstream task ώστε να αξιολογηθεί η γενίκευση του σε νέα/ άγνωστα δεδομένα της προοριζόμενης εργασίας .

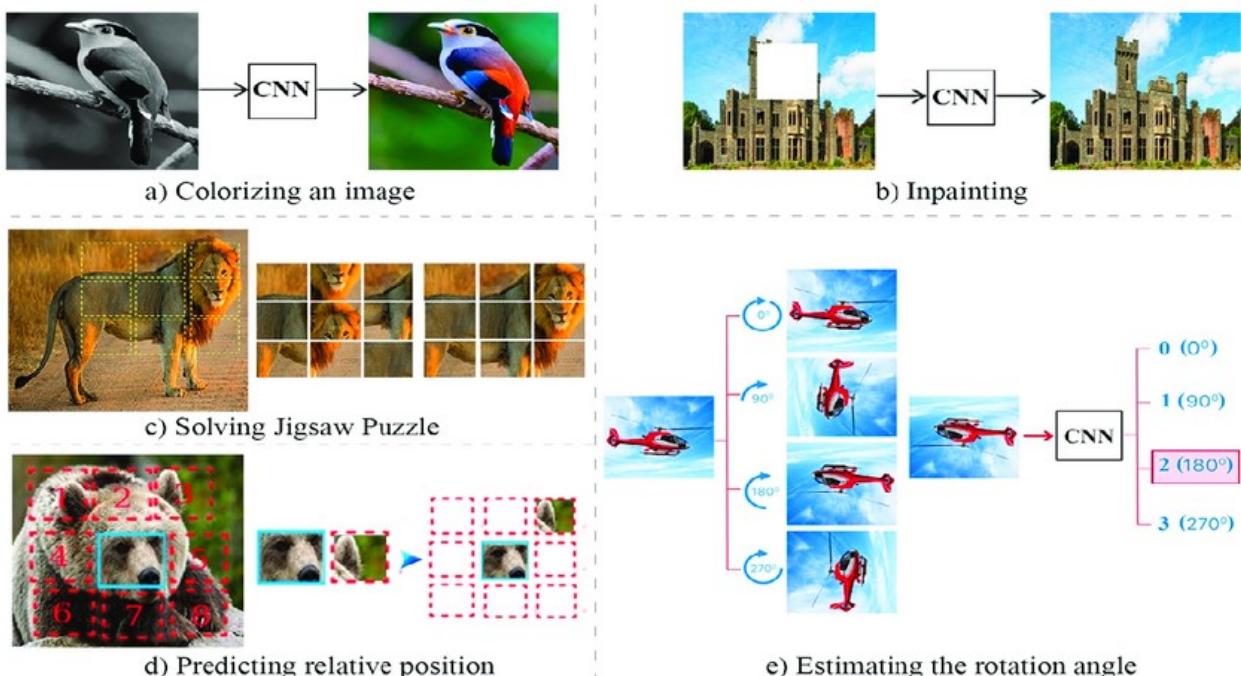
Διαδεδομένες Βιβλιοθήκες και Εργαλεία για την Υλοποίηση της *SSL* (και Νευρωνικών Δικτύων γενικότερα):

- PyTorch: Μία από τις πιο διαδεδομένες βιβλιοθήκες για την ανάπτυξη και εκπαίδευση νευρωνικών δικτύων, η *PyTorch* [45] προσφέρει ενσωματωμένες δυνατότητες για την υλοποίηση αυτο-εποπτευόμενης μάθησης μέσω της χρήσης μετασχηματισμών δεδομένων και των autograd λειτουργιών του(η βιβλιοθήκη *PyTorch* είναι το κύριο εργαλείο που χρησιμοποιήθηκε στο Πειραματικό Μέρος).
- TensorFlow: Άλλη μια δημοφιλής βιβλιοθήκη, το *TensorFlow* [46] παρέχει εργαλεία για την ανάπτυξη και εκπαίδευση μοντέλων αυτο-εποπτευόμενης μάθησης. Η δυνατότητα χρήσης των Keras APIs μέσα από το *TensorFlow* καθιστά την υλοποίηση πιο προσιτή και ευέλικτη.
- Scikit-learn: Παρόλο που είναι γνωστό για την απλότητά του, το *Scikit-learn* [47] μπορεί να χρησιμοποιηθεί σε συνδυασμό με άλλες βιβλιοθήκες για την προεπεξεργασία δεδομένων και την υλοποίηση βασικών μοντέλων αυτο-εποπτευόμενης μάθησης.
- Hugging Face Transformers: Για την επεξεργασία φυσικής γλώσσας και άλλα σχετικά tasks, η βιβλιοθήκη *Hugging Face* [48] παρέχει προκαθορισμένα μοντέλα και εργαλεία για την εφαρμογή αυτο-εποπτευόμενης μάθησης σε ευρεία κλίμακα.
- FastAI: Μια βιβλιοθήκη που βασίζεται στο *PyTorch* και παρέχει εύχρηστα εργαλεία και αλγορίθμους για την ανάπτυξη μοντέλων αυτο-εποπτευόμενης μάθησης με ελάχιστη προσπάθεια [45,49].

2.2 Έμμεσες Διεργασίες

2.2.1 Διαφορετικοί Τύποι Έμμεσων Διεργασιών

Στην Αυτο-εποπτεύομενη Μάθηση, οι έμμεσες διεργασίες παίζουν κεντρικό ρόλο στην εκπαίδευση των μοντέλων χωρίς την ανάγκη για επισημασμένα δεδομένα. Μέσω αυτών των διεργασιών, το μοντέλο καλείται να επιλύσει τεχνητά δημιουργημένα προβλήματα, τα οποία σχεδιάζονται με τέτοιο τρόπο ώστε να αναγκάζουν το μοντέλο να μάθει ουσιαστικές ιδιότητες και χαρακτηριστικά των δεδομένων. Κάθε έμμεση διεργασία αντιπροσωπεύει μια μοναδική προσέγγιση για την εξαγωγή γνώσης από μη επισημασμένα δεδομένα, και μπορεί να εφαρμοστεί σε ποικίλους τομείς, από την επεξεργασία εικόνας και ήχου έως την αναγνώριση προτύπων. Παρακάτω, θα εξετάσουμε τους διαφορετικούς τύπους αυτών των έμμεσων διεργασιών και πώς αυτοί συμβάλλουν στην ανάπτυξη αποτελεσματικών και γενικευμένων μοντέλων.



Εικόνα 2.2.1 - 1: Παραδείγματα Έμμεσων Διεργασιών
(Πηγή: ResearchGate)

1. Συμπλήρωση ή Αποκατάσταση Ελλιπών Δεδομένων (Completion or Restoration of Missing Data):

- Αυτή η κατηγορία περιλαμβάνει tasks όπου λείπουν τμήματα της εικόνας ή των δεδομένων, και το μοντέλο καλείται να τα ανακατασκευάσει ή να τα συμπληρώσει. Παραδείγματα περιλαμβάνουν:
 - Image Inpainting: Συμπλήρωση χαμένων ή κρυφών τμημάτων μιας εικόνας.
 - Text Infilling: Συμπλήρωση χαμένων λέξεων ή φράσεων σε ένα κείμενο [50].

2. Ανακατασκευή Αρχικών Δεδομένων από Μετασχηματισμένα Δεδομένα (Reconstruction from Transformed Data):

- Σε αυτήν την κατηγορία, το μοντέλο παίρνει ως είσοδο μια τροποποιημένη ή αλλοιωμένη εκδοχή των δεδομένων και προσπαθεί να ανακτήσει την αρχική τους μορφή. Παραδείγματα περιλαμβάνουν:
 - Image Colorization: Χρωματοποίηση ασπρόμαυρων εικόνων.
 - Denoising: Αφαίρεση θορύβου από εικόνες ή ήχο [55].

3. Προσανατολισμός και Μετασχηματισμοί Δεδομένων (Orientation and Transformation Tasks):

- Αυτή η κατηγορία περιλαμβάνει tasks που σχετίζονται με την αναγνώριση ή τη διόρθωση του προσανατολισμού και άλλων γεωμετρικών μετασχηματισμών των δεδομένων. Παραδείγματα περιλαμβάνουν:
 - Rotation Prediction: Αναγνώριση του σωστού προσανατολισμού μιας εικόνας μετά από τυχαία περιστροφή.
 - Flip Prediction: Αναγνώριση αν μια εικόνα έχει αναστραφεί (flipped).

4. Μάθηση μέσω Αντίθεσης (Contrastive Learning):

- Σε αυτή την κατηγορία, το μοντέλο εκπαιδεύεται να αναγνωρίζει τη διαφορά ή την ομοιότητα μεταξύ δύο παραλλαγών του ίδιου δεδομένου. Αυτή η προσέγγιση ενισχύει τη μάθηση των χαρακτηριστικών που είναι ανθεκτικά σε αλλαγές ή θόρυβο. Παραδείγματα περιλαμβάνουν:
 - SimCLR (Simple Framework for Contrastive Learning of Visual Representations): Το μοντέλο προσπαθεί να διαφοροποιήσει μεταξύ ζευγών εικόνων που είναι παρόμοιες ή διαφορετικές [56].
 - Instance Discrimination: Κατηγοριοποίηση δεδομένων με βάση τη διακριτότητά τους [57].

5. Προβλέψεις Χρονικών Ακολουθιών (Temporal Sequence Prediction):

- Εδώ το μοντέλο καλείται να προβλέψει την επόμενη κατάσταση ή την επόμενη χρονική στιγμή σε μια σειρά δεδομένων. Παραδείγματα περιλαμβάνουν:
 - Next Word Prediction: Προβλέψεις της επόμενης λέξης σε ένα κείμενο [58].
 - Future Frame Prediction: Προβλέψεις της επόμενης εικόνας σε μια ακολουθία βίντεο [59].

Στις επόμενες ενότητες θα αναλύσουμε κάποιες από αυτές τις Έμμεσες Διεργασίες οι οποίες θα χρησιμοποιηθούν στο Πειραματικό Μέρος.

2.2.2 Πρόβλεψη Περιστροφής Εικόνας

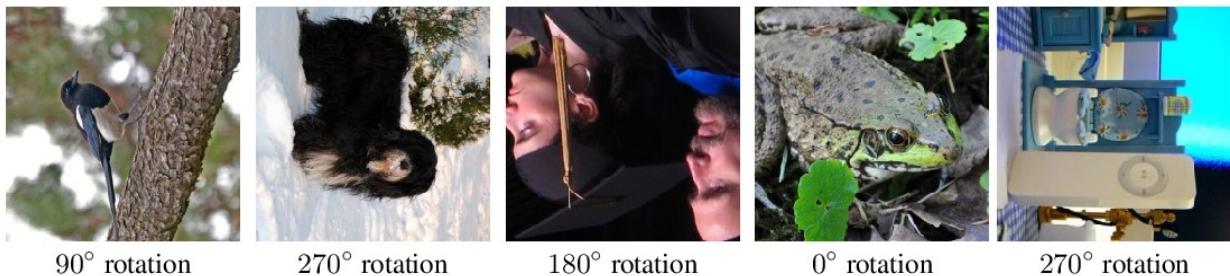
Η Πρόβλεψη Περιστροφής Εικόνας (Image Rotation Prediction) είναι ένα από τα πιο κλασικά και διαδεδομένα proxy tasks που χρησιμοποιούνται στην αυτο-εποπτευόμενη μάθηση. Στόχος αυτού του task είναι η πρόβλεψη του βαθμού περιστροφής μιας εικόνας από το μοντέλο. Κατά τη διαδικασία αυτή, μια εικόνα περιστρέφεται τυχαία κατά 0° , 90° , 180° , ή 270° , και το μοντέλο πρέπει να μάθει να προβλέπει τη σωστή γωνία περιστροφής.

Η διαδικασία ξεκινά με την εισαγωγή των περιστρεφόμενων εκδόσεων της εικόνας ως είσοδο στο μοντέλο, ενώ η πραγματική γωνία περιστροφής χρησιμοποιείται ως label για την εκπαίδευση. Αυτό αναγκάζει το μοντέλο να μάθει χαρακτηριστικά που σχετίζονται με την υφή, τα περιγράμματα και τις δομές μέσα στην εικόνα, ώστε να μπορεί να εντοπίσει τη σωστή κατεύθυνση και γωνία της περιστροφής.

Κατά την εκπαίδευση, χρησιμοποιείται μια συνάρτηση απώλειας που μετρά την ακρίβεια των προβλέψεων του μοντέλου σχετικά με τη γωνία περιστροφής. Στη συνέχεια, εφαρμόζεται ένας αλγόριθμος βελτιστοποίησης για να προσαρμόσει τα βάρη του μοντέλου, με στόχο τη βελτίωση της ικανότητάς του να διακρίνει με ακρίβεια τη σωστή γωνία περιστροφής σε κάθε εικόνα. Το μοντέλο μαθαίνει έτσι να εξάγει πληροφορίες σχετικές με τη δομή και την κατεύθυνση των αντικειμένων μέσα στις εικόνες, κάτι που το

βοηθά να γενικεύσει και σε άλλες, πιο σύνθετες εργασίες, όπως η αναγνώριση αντικειμένων ή η κατηγοριοποίηση εικόνων.

Η πρόβλεψη περιστροφής είναι ιδιαίτερα χρήσιμη καθώς αναγκάζει το μοντέλο να κατανοήσει τα γεωμετρικά χαρακτηριστικά των εικόνων, βελτιώνοντας τη συνολική του απόδοση σε μια σειρά από downstream tasks.



Εικόνα 2.2.2 - 1: Εικόνα που απεικονίζει το έργο της Πρόβλεψης Περιστροφής Εικόνας (Image Rotation Prediction Task). Εικόνες που έχουν περιστραφεί κατά 0, 90, 180, 270 μοίρες.

(Πηγή: Gidaris, S., Singh, P., & Komodakis, N. (2018). Unsupervised representation learning by predicting image rotations. *arXiv preprint arXiv:1803.07728.*)

2.2.3 Χρωματοποίηση Εικόνας

Η Χρωματοποίηση Εικόνας (Image Colorization) είναι μία από τις πιο ευρέως χρησιμοποιούμενες έμμεσες διεργασίες στην Αυτο-εποπτευόμενη Μάθηση, και έχει ως στόχο την εκμάθηση πλούσιων και γενικευμένων αναπαραστάσεων των δεδομένων. Η διαδικασία αυτή ξεκινά με την απόκτηση αρχικών έγχρωμων εικόνων, οι οποίες στη συνέχεια μετατρέπονται σε ασπρόμαυρες (grayscale). Αυτή η μετατροπή γίνεται με την αφαίρεση των χρωματικών πληροφοριών, αφήνοντας μόνο τις αποχρώσεις του γκρι που αναπαριστούν την ένταση του φωτός σε κάθε σημείο της εικόνας.

Στο επόμενο βήμα, το μοντέλο λαμβάνει αυτές τις ασπρόμαυρες εικόνες ως είσοδο και προσπαθεί να ανακατασκευάσει τα αρχικά χρώματα, δηλαδή να "χρωματίσει" την εικόνα. Ο στόχος είναι να προβλεφθούν οι ακριβείς χρωματικές τιμές για κάθε pixel της εικόνας. Αυτή η διαδικασία απαιτεί από το μοντέλο να κατανοήσει τις χωρικές σχέσεις και το περιεχόμενο της εικόνας, όπως τα όρια των αντικειμένων, τις σκιάσεις, καθώς και τις σχέσεις μεταξύ των διαφορετικών αντικειμένων που απεικονίζονται.

Κατά την εκπαίδευση του μοντέλου, οι προβλέψεις των χρωμάτων συγκρίνονται με τα πραγματικά χρώματα της αρχικής έγχρωμης εικόνας, η οποία χρησιμοποιείται ως το "ground truth". Μέσω μιας συνάρτησης απώλειας, που μετρά τη διαφορά μεταξύ των προβλεπόμενων από το μοντέλο και των πραγματικών χρωμάτων, και με τη χρήση ενός αλγορίθμου βελτιστοποίησης (optimizer), το μοντέλο προσαρμόζει τα βάρη του για να βελτιώσει την ακρίβεια των προβλέψεών του. Αυτή η διαδικασία επαναλαμβάνεται πολλές φορές κατά τη διάρκεια της εκπαίδευσης, οδηγώντας το μοντέλο να αναπτύξει βαθύτερη κατανόηση της δομής, του περιεχομένου, και των χρωματικών σχέσεων που υπάρχουν στις εικόνες.

Η χρωματοποίηση εικόνας ως έμμεση διεργασία είναι πολύτιμη, καθώς επιτρέπει στο μοντέλο να μάθει αναπαραστάσεις που μπορούν να μεταφερθούν σε άλλα, πιο πολύπλοκα downstream tasks, όπως η αναγνώριση αντικειμένων ή η ταξινόμηση εικόνων. Μέσα από αυτή τη διαδικασία, το μοντέλο αναπτύσσει την ικανότητα να κατανοεί γενικά πρότυπα και δομές που είναι κοινά στα δεδομένα, γεγονός που το καθιστά πιο αποτελεσματικό και ευέλικτο σε μια ποικιλία εφαρμογών.



Εικόνα 2.2.3 - 1: Εικόνα που απεικονίζει το έργο της Χρωματοποίησης Εικόνας (image colorization task). Η αριστερή εικόνα είναι η grayscale είσοδος (input) στο μοντέλο, η μεσαία εικόνα είναι η πρόβλεψη του μοντέλου (output/prediction), ενώ η δεξιά εικόνα είναι η αρχική με τα πραγματικά χρώματα (ground truth). (Πηγή: AiProjects)

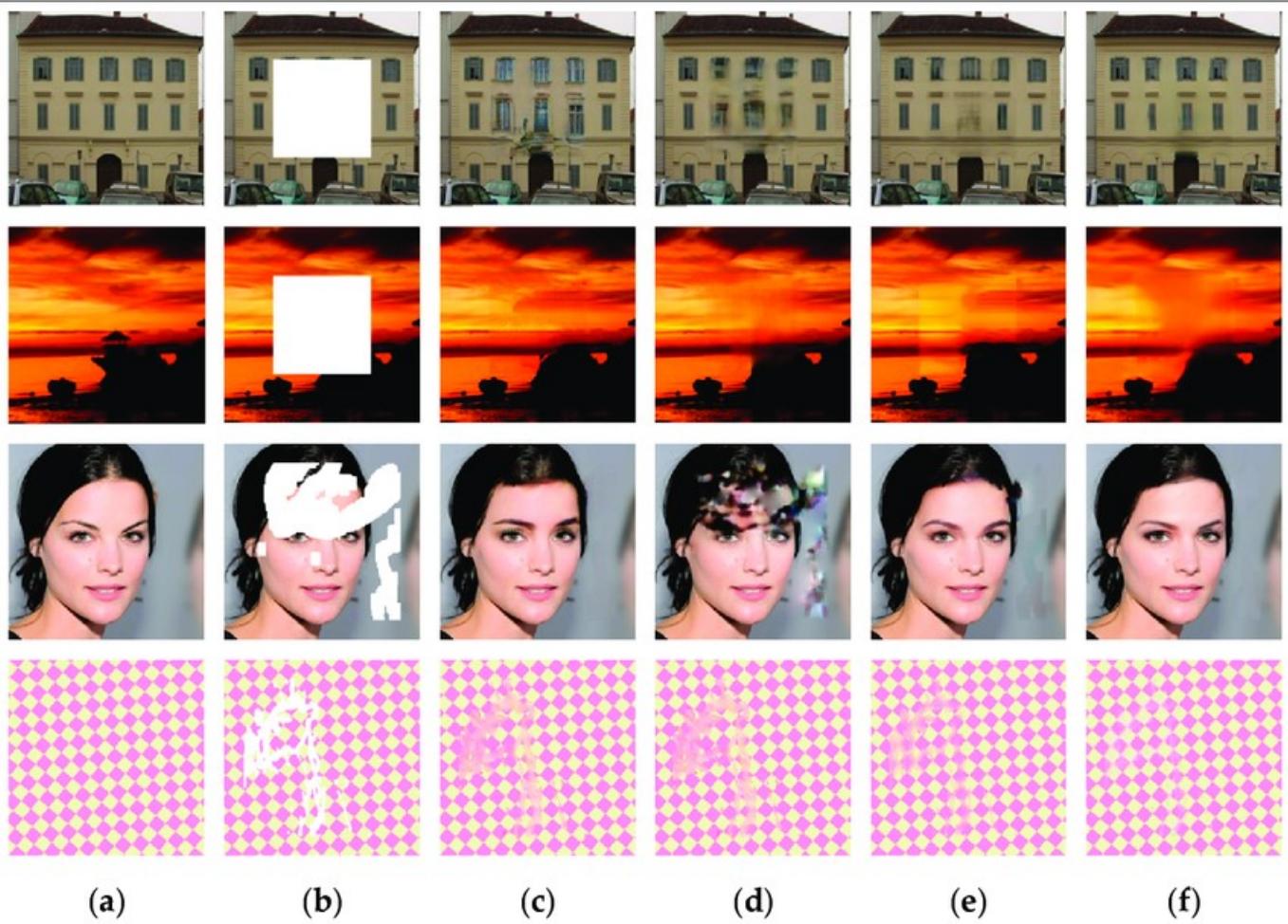
2.2.4 Επιδιόρθωση/Συμπλήρωση Εικόνας

Η Επιδιόρθωση/Συμπλήρωση Εικόνας (Image Inpainting) είναι ένα από τα πιο διαδεδομένα proxy tasks στην αυτο-εποπτεύομένη μάθηση, το οποίο έχει ως στόχο την αποκατάσταση των κενών ή κατεστραμμένων περιοχών μιας εικόνας με τρόπο που να μοιάζει φυσικός και να διατηρεί τη συνοχή της υπόλοιπης εικόνας.

Η διαδικασία ξεκινά με την εισαγωγή ενός είδους αλλοίωσης στην αρχική εικόνα. Συγκεκριμένα, τμήματα της εικόνας αφαιρούνται ή καλύπτονται με μια μάσκα, δημιουργώντας κενά που το μοντέλο πρέπει να ανακατασκευάσει. Αυτά τα καλυμμένα τμήματα της εικόνας λειτουργούν ως είσοδος στο μοντέλο, ενώ η πλήρης, μη αλλοιωμένη εικόνα χρησιμεύει ως το "ground truth" (δηλαδή, τα πραγματικά δεδομένα με τα οποία θα συγκριθούν οι προβλέψεις του μοντέλου).

Κατά την εκπαίδευση του μοντέλου, χρησιμοποιείται μια συνάρτηση απώλειας που μετρά τη διαφορά μεταξύ της ανακατασκευασμένης περιοχής (που προβλέπει το μοντέλο) και της πραγματικής περιοχής στην πλήρη εικόνα. Στη συνέχεια, ένας αλγόριθμος βελτιστοποίησης χρησιμοποιείται για να προσαρμόσει τα βάρη του μοντέλου, ελαχιστοποιώντας το σφάλμα στην πρόβλεψη και βελτιώνοντας έτσι την ικανότητα του μοντέλου να ανακατασκευάζει με ακρίβεια τις καλυμμένες περιοχές.

Μέσω αυτής της διαδικασίας, το μοντέλο μαθαίνει να αναγνωρίζει και να κατανοεί τις δομές και τα πρότυπα στις εικόνες, ώστε να μπορεί να αναπληρώνει τα χαμένα τμήματα με τρόπο που διατηρεί τη συνοχή και τη φυσικότητα της εικόνας.



Εικόνα 2.2.4 - 1: Εικόνα που απεικονίζει το έργο της Επιδιόρθωσης/ Συμπλήρωσης Εικόνας (Image Inpainting task). Η στήλη (a) δείχνει την αρχική εικόνα (ground truth), η στήλη (b) παρουσιάζει την εικόνα με την καλυμμένη περιοχή (masked εικόνα), ενώ οι στήλες (c), (d), (e) και (f) δείχνουν τις προβλέψεις από τέσσερα διαφορετικά μοντέλα. (Πηγή: ResearchGate)

2.3 Περιορισμοί και Προκλήσεις

2.3.1 Περιορισμένη Γενίκευση Έμμεσων Διεργασιών

Στην Αυτο-Εποπτευόμενη Μάθηση υπάρχει το ζήτημα της περιορισμένης γενίκευσης των έμμεσων διεργασιών. Ένα σημαντικό πρόβλημα που συναντάται είναι ότι οι διεργασίες που χρησιμοποιούνται για να δημιουργηθούν εποπτευόμενα σήματα από μη επισημασμένα δεδομένα ενδέχεται να μην είναι πάντα κατάλληλες ή να μην αποδίδουν εξίσου καλά σε όλα τα είδη δεδομένων, και το σημαντικότερο, να μην συνάδουν με τις απαιτήσεις των downstream tasks.

Ένα βασικό πρόβλημα είναι ότι το μοντέλο μπορεί να εκπαιδευτεί να αναγνωρίζει και να χρησιμοποιεί χαρακτηριστικά που είναι πολύ συγκεκριμένα για την έμμεση διεργασία που επιλέχθηκε. Για παράδειγμα, ένα μοντέλο που έχει εκπαιδευτεί να χρωματίζει εικόνες μπορεί να μάθει να εστιάζει σε συγκεκριμένα χρωματικά μοτίβα ή υφές που είναι χρήσιμα για την επιτυχία της συγκεκριμένης διεργασίας. Ωστόσο, αυτά τα χαρακτηριστικά μπορεί να μην είναι εξίσου χρήσιμα ή ακόμα και άσχετα όταν το μοντέλο καλείται να επιλύσει μια διαφορετική εργασία, όπως η αναγνώριση αντικειμένων, η ανίχνευση ανωμαλιών ή η κατηγοριοποίηση εικόνων.

Στο πρακτικό μέρος της μελέτης μας, θα εξετάσουμε τη δυνατότητα μεταφοράς των γνώσεων που έχει αποκτήσει το μοντέλο από τις έμμεσες διεργασίες σε άλλες, πιο συγκεκριμένες εφαρμογές. Ένας από τους στόχους μας είναι να διαπιστώσουμε κατά πόσο οι έμμεσες διεργασίες μπορούν να συμβάλλουν αποτελεσματικά στη βελτίωση της απόδοσης των downstream tasks ή αν τελικά περιορίζουν τη γενίκευση του μοντέλου.

Αυτό το ζήτημα έχει ανοίξει έναν ευρύ επιστημονικό διάλογο, καθώς οι ερευνητές αναζητούν τρόπους να βελτιώσουν τις μεθόδους αυτο-εποπτευόμενης μάθησης ώστε να προσφέρουν καλύτερη απόδοση σε πρακτικές εφαρμογές. Υπάρχει έντονη συζήτηση για το κατά πόσο οι έμμεσες διεργασίες μπορούν να οδηγήσουν σε μοντέλα που γενικεύουν αποτελεσματικά ή αν περιορίζουν την απόδοση σε νέες εργασίες.

2.3.2 Υπολογιστικό Κόστος, Πόροι και Χρόνος Εκπαίδευσης

Η Αυτο-Εποπτευόμενη Μάθηση, όπως και άλλες προηγμένες μέθοδοι μηχανικής μάθησης, απαιτεί σημαντικούς υπολογιστικούς πόρους, τόσο σε επίπεδο υλικού όσο και σε χρόνο εκπαίδευσης. Εξαιτίας της φύσης της, όπου μεγάλες ποσότητες δεδομένων πρέπει να επεξεργαστούν χωρίς άμεση επίβλεψη, οι απαιτήσεις σε υπολογιστική ισχύ, μνήμη, και αποθηκευτικό χώρο είναι αυξημένες.

Το υπολογιστικό κόστος αυξάνεται ιδιαίτερα όταν πρόκειται για μεγάλα μοντέλα με εκατομμύρια ή δισεκατομμύρια παραμέτρους. Αυτά τα μοντέλα απαιτούν εξειδικευμένο υλικό, όπως GPU ή TPU, για να επιταχυνθεί η διαδικασία εκπαίδευσης. Η αγορά και συντήρηση τέτοιου εξοπλισμού συνεπάγεται σημαντικές δαπάνες, που μπορεί να είναι απαγορευτικές για μικρότερες επιχειρήσεις ή ερευνητικά κέντρα με περιορισμένο προϋπολογισμό.

Επιπλέον, ο χρόνος εκπαίδευσης αυτών των μοντέλων μπορεί να είναι ιδιαίτερα μεγάλος. Καθώς τα δεδομένα είναι συχνά ακατέργαστα και απρόβλεπτα, η διαδικασία εκμάθησης των ουσιαστικών χαρακτηριστικών απαιτεί πολλές επαναλήψεις (epochs) εκπαίδευσης, γεγονός που επιμηκύνει τον συνολικό χρόνο που χρειάζεται μέχρι το μοντέλο να γίνει αποδοτικό και αξιόπιστο. Αυτή η χρονική καθυστέρηση μπορεί να επηρεάσει τη δυνατότητα γρήγορης υλοποίησης λύσεων και να προκαλέσει προβλήματα στην ευελιξία και την προσαρμοστικότητα μιας επιχείρησης.

Τέλος, οι αυξημένες απαιτήσεις σε υπολογιστικούς πόρους και χρόνο εκπαίδευσης δημιουργούν και μια ακόμα πρόκληση: την ανάγκη για συνεχή παρακολούθηση και προσαρμογή των μοντέλων, ώστε να διασφαλιστεί η αποδοτικότητά τους χωρίς να ξεφεύγουν οι απαιτήσεις σε πόρους πέρα από τα διαθέσιμα

όρια. Ως εκ τούτου, το υπολογιστικό κόστος, οι πόροι και ο χρόνος εκπαίδευσης αποτελούν κρίσιμες προκλήσεις στην αυτο-εποπτευόμενη μάθηση, περιορίζοντας συχνά την εφαρμοσιμότητά της σε ευρύτερες κλίμακες.

2.3.3 Προβλήματα Υπερπροσαρμογής

Η υπερπροσαρμογή (overfitting) αποτελεί ένα από τα πιο σοβαρά προβλήματα στην εκπαίδευση μοντέλων Μηχανικής Μάθησης, και η Αυτο-Εποπτευόμενη μάθηση δεν αποτελεί εξαίρεση. Η υπερπροσαρμογή συμβαίνει όταν ένα μοντέλο μαθαίνει να απομνημονεύει τα δεδομένα εκπαίδευσης τόσο καλά που χάνει την ικανότητά του να γενικεύει σε νέα, αόρατα δεδομένα. Αυτό μπορεί να οδηγήσει σε υψηλή ακρίβεια κατά τη διάρκεια της εκπαίδευσης, αλλά σε χαμηλή απόδοση όταν το μοντέλο χρησιμοποιείται σε πραγματικές εφαρμογές ή σε διαφορετικά σύνολα δεδομένων.

Στην Αυτο-εποπτευόμενη Μάθηση, ο κίνδυνος υπερπροσαρμογής είναι αυξημένος λόγω της φύσης των έμμεσων διεργασιών που χρησιμοποιούνται για την εκπαίδευση του μοντέλου. Εάν το μοντέλο επικεντρωθεί υπερβολικά στην επίλυση του συγκεκριμένου proxy task, μπορεί να μάθει να εκμεταλλεύεται τα συγκεκριμένα χαρακτηριστικά που σχετίζονται με αυτό το task, χωρίς να αναπτύσσει την ικανότητα να γενικεύει σε άλλα καθήκοντα. Αυτό μπορεί να σημαίνει ότι το μοντέλο αποδίδει εξαιρετικά καλά στο proxy task, αλλά αποτυγχάνει όταν προσπαθεί να αντιμετωπίσει το πραγματικό downstream task για το οποίο προορίζοταν.

Η αντιμετώπιση της υπερπροσαρμογής στην αυτο-εποπτευόμενη μάθηση απαιτεί την εφαρμογή τεχνικών κανονικοποίησης, όπως το dropout, το batch normalization, και άλλες, καθώς και την χρήση προσεκτικά επιλεγμένων proxy tasks που προάγουν τη γενίκευση. Η επιτυχής αντιμετώπιση της υπερπροσαρμογής αποτελεί κρίσιμο παράγοντα για την αποτελεσματική εφαρμογή της Αυτο-Εποπτευόμενης μάθησης, καθώς και για την αξιοπιστία των μοντέλων που παράγονται από αυτήν την προσέγγιση.

2.3.4 Δυσκολίες στη Μεταφορά Χαρακτηριστικών και Ασυμβατότητα Εργασιών

Η μεταφορά χαρακτηριστικών μέσω του transfer learning αποτελεί έναν από τους βασικούς στόχους της Αυτο-Εποπτευόμενης μάθησης. Ωστόσο, η διαδικασία αυτή δεν είναι πάντα χωρίς προκλήσεις. Συχνά υπάρχει ασυμβατότητα ανάμεσα σε proxy και downstream εργασίες. Η ασυμβατότητα αυτή μπορεί να οφείλεται σε διάφορους παράγοντες, όπως η διαφορετική φύση των δεδομένων μεταξύ του proxy task και του downstream task ή οι διαφορετικές απαιτήσεις που έχει κάθε εργασία από το μοντέλο. Παρά τις βελτιώσεις που έχουν σημειωθεί στον τομέα του transfer learning, η βέλτιστη μεταφορά γνώσεων παραμένει ένα ανοιχτό ζήτημα στον επιστημονικό χώρο, με τους ερευνητές να εξετάζουν συνεχώς νέες τεχνικές για τη βελτίωση της απόδοσης και της γενίκευσης των μοντέλων.

Επιπλέον, η διαδικασία του fine-tuning, που συνήθως ακολουθεί το transfer learning, απαιτεί λεπτή ισορροπία. Αν δεν γίνει σωστά, υπάρχει ο κίνδυνος το μοντέλο είτε να μην εκμεταλλεύεται επαρκώς τα ήδη μαθημένα χαρακτηριστικά, είτε να προσαρμοστεί υπερβολικά στα νέα δεδομένα, χάνοντας τη γενικότητα που είχε αποκτήσει κατά την εκπαίδευση στο proxy task.

2.4 Εφαρμογές

2.4.1 Επεξεργασία Εικόνας

Η Αυτο-εποπτεύμενη Μάθηση έχει αρχίσει να διαδραματίζει κρίσιμο ρόλο στην επεξεργασία εικόνας, ανοίγοντας νέους ορίζοντες για την ανάλυση και την κατανόηση οπτικών δεδομένων χωρίς την ανάγκη μεγάλων συνόλων επισημασμένων δεδομένων. Στην επεξεργασία εικόνας, η *SSL* επιτρέπει στα μοντέλα να μάθουν αναπαραστάσεις υψηλού επιπέδου που μπορούν να χρησιμοποιηθούν για μια ποικιλία εφαρμογών, όπως η βελτίωση της ανάλυσης και η αποκατάσταση των εικόνων.

Μέσω της αυτο-εποπτεύμενης μάθησης, τα μοντέλα μπορούν να μάθουν να αναγνωρίζουν τα βασικά χαρακτηριστικά των εικόνων, όπως τα περιγράμματα, οι σκιές, οι υφές, και οι χρωματικές σχέσεις, με ελάχιστη ανθρώπινη παρέμβαση. Αυτή η ικανότητα επιτρέπει στα συστήματα να βελτιώνουν την ανάλυση των εικόνων, να ανακατασκευάζουν χαμένες ή κατεστραμμένες περιοχές και να προσαρμόζουν την απόδοση σε διαφορετικούς τύπους οπτικών δεδομένων.

Επιπλέον, η *SSL* έχει αποδειχθεί πολύτιμη για την αυτοματοποίηση της κατηγοριοποίησης και της ανάλυσης μεγάλων συνόλων εικόνων, προσφέροντας εργαλεία για τη γρήγορη και ακριβή αναγνώριση αντικειμένων και σκηνών. Με τη χρήση μεθόδων αυτο-εποπτεύμενης μάθησης, οι εφαρμογές επεξεργασίας εικόνας μπορούν να επεκταθούν σε νέους τομείς, όπου τα δεδομένα είναι πλούσια αλλά μη επισημασμένα, οδηγώντας σε καινοτόμες λύσεις για την κατανόηση και την αξιοποίηση οπτικών πληροφοριών.

2.4.2 Ανάλυση Βίντεο

Η Αυτο-εποπτεύμενη Μάθηση έχει αρχίσει να διαδραματίζει σημαντικό ρόλο στην ανάλυση βίντεο, προσφέροντας νέες μεθόδους για την επεξεργασία και κατανόηση μεγάλων συνόλων βιντεοσκοπημένου υλικού. Στην ανάλυση βίντεο, η *SSL* επιτρέπει στα μοντέλα να μάθουν χρήσιμες αναπαραστάσεις και να εξάγουν ουσιαστικές πληροφορίες χωρίς την ανάγκη επισημασμένων δεδομένων, τα οποία συχνά είναι δύσκολο και δαπανηρό να αποκτηθούν.

Μέσα από τη διαδικασία της αυτο-εποπτεύμενης μάθησης, τα μοντέλα μπορούν να αναγνωρίζουν μοτίβα και κινήσεις σε βίντεο, όπως την αναγνώριση δραστηριοτήτων, την παρακολούθηση αντικειμένων και τη διάγνωση ανωμαλιών. Η ικανότητα των *SSL* μοντέλων να μαθαίνουν από ακατέργαστα δεδομένα καθιστά δυνατό τον εντοπισμό και την κατανόηση πολύπλοκων αλληλουχιών γεγονότων σε βίντεο, βελτιώνοντας την ακρίβεια και την απόδοση των εφαρμογών.

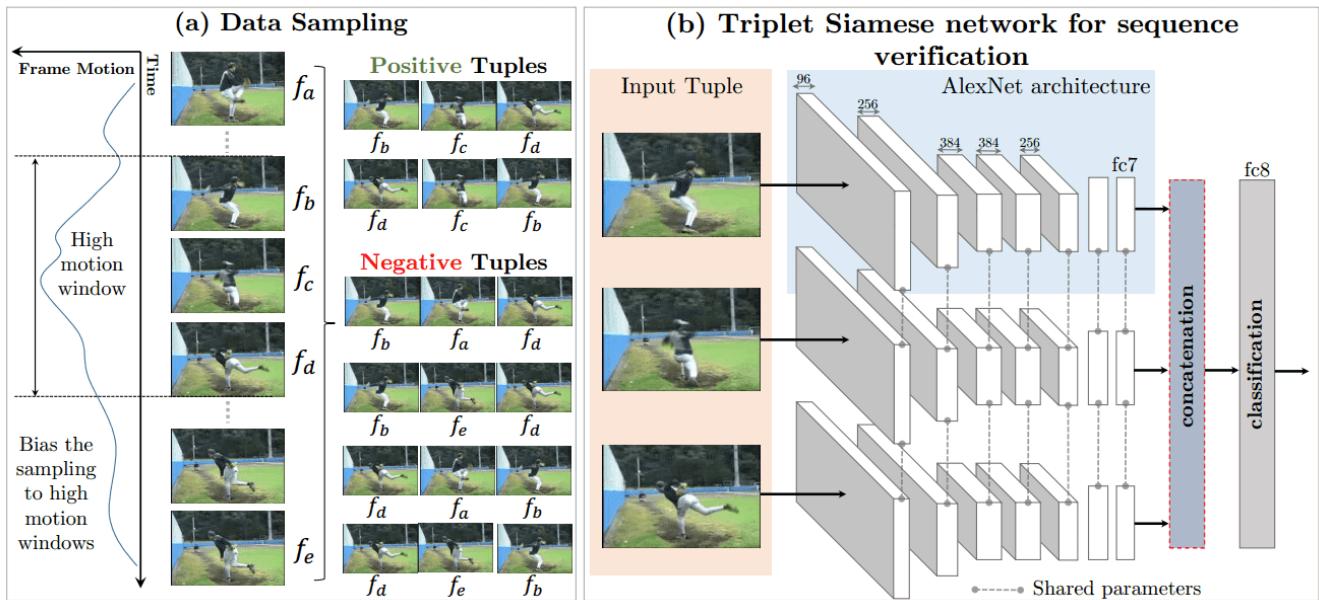
Επιπλέον, η *SSL* μπορεί να εφαρμοστεί για τη βελτίωση της ανάλυσης βίντεο σε πραγματικό χρόνο, επιτρέποντας την αυτοματοποίηση της παρακολούθησης και της ανάλυσης βίντεο από κάμερες ασφαλείας, αθλητικές μεταδόσεις, και άλλες πηγές βίντεο. Αυτή η προσέγγιση προσφέρει την ικανότητα να εξάγονται χρήσιμα συμπεράσματα από τεράστιους όγκους βιντεοσκοπημένου υλικού με ελάχιστη ανθρώπινη παρέμβαση, καθιστώντας την ανάλυση βίντεο πιο αποτελεσματική και ακριβή.

Παράδειγμα:

Ένα παράδειγμα εφαρμογής της *SSL* στην ανάλυση βίντεο είναι η αναγνώριση και η κατηγοριοποίηση ανθρώπινων δραστηριοτήτων σε βίντεο επιτήρησης. Σε αυτό το σενάριο, το μοντέλο εκπαιδεύεται χρησιμοποιώντας βίντεο χωρίς ετικέτες. Αρχικά, το μοντέλο μπορεί να μάθει να αναγνωρίζει βασικά μοτίβα κίνησης, όπως το περπάτημα, το τρέξιμο ή το σταμάτημα, μέσω έμμεσων διεργασιών proxy tasks όπως η πρόβλεψη του επόμενου καρέ (frame prediction) ή η ανίχνευση της αλλαγής κατεύθυνσης κίνησης.

Μόλις το μοντέλο έχει μάθει να αναγνωρίζει αυτά τα βασικά μοτίβα, μπορεί να προσαρμοστεί (fine-tuned) για πιο εξειδικευμένες εργασίες, όπως η αναγνώριση υπόπτων συμπεριφορών ή η ανίχνευση περιστατικών πτώσης. Αυτή η διαδικασία εξαλείφει την ανάγκη για μεγάλα σύνολα δεδομένων με ετικέτες, καθιστώντας

την ανάλυση βίντεο πιο αποδοτική και προσιτή σε πραγματικό χρόνο.



Εικόνα 2.4.2 - 1: Χρήση Αυτο-Εποπτεύμενης Μάθησης για ανάλυση βίντεο. (Πηγή: AiSummer)

2.4.3 Επεξεργασία Φυσικής Γλώσσας

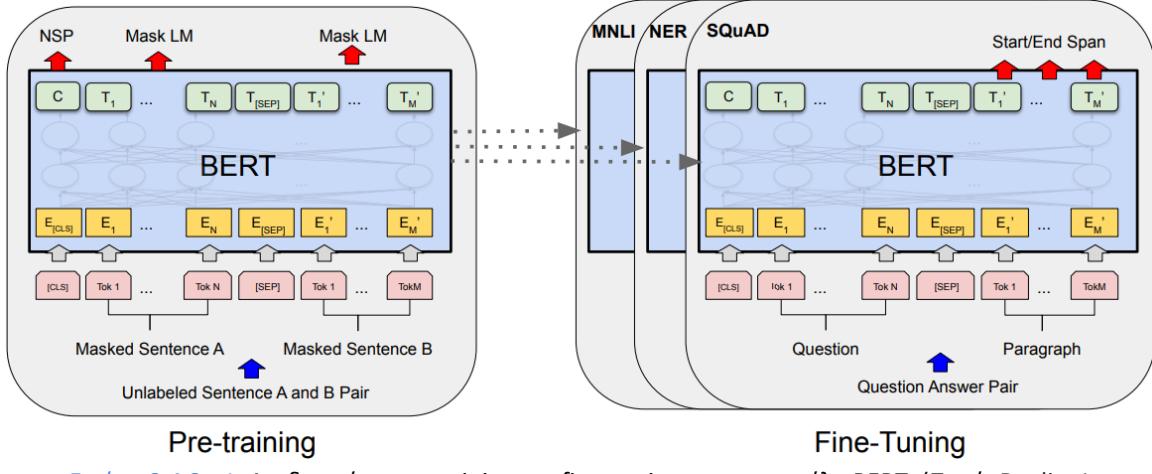
Η SSL έχει αρχίσει να διαδραματίζει σημαντικό ρόλο στην Επεξεργασία Φυσικής Γλώσσας (Natural Language Processing - NLP), επιτρέποντας την εκπαίδευση μοντέλων που κατανοούν και επεξεργάζονται κείμενο με λιγότερα δεδομένα με ετικέτες. Ένα από τα βασικά πλεονεκτήματα της SSL στην επεξεργασία φυσικής γλώσσας είναι η ικανότητά της να μαθαίνει αναπαραστάσεις κειμένου (text embeddings) από ακατέργαστο κείμενο, χωρίς την ανάγκη για επισημασμένα δεδομένα.

Μια από τις πιο γνωστές εφαρμογές της SSL στην επεξεργασία φυσικής γλώσσας είναι η εκπαίδευση γλωσσικών μοντέλων, όπως το BERT (Bidirectional Encoder Representations from Transformers) [50]. Στην αρχική φάση εκπαίδευσης, το BERT χρησιμοποιεί μια έμμεση διεργασία που ονομάζεται "Masked Language Modeling", όπου κάποιες λέξεις μέσα σε προτάσεις καλύπτονται τυχαία και το μοντέλο μαθαίνει να τις προβλέπει. Μέσω αυτής της διαδικασίας, το μοντέλο καταφέρνει να κατανοήσει τις σχέσεις μεταξύ των λέξεων και των εννοιών στο κείμενο.

Αυτή η τεχνική έχει αποδειχθεί εξαιρετικά αποτελεσματική για πολλές εφαρμογές NLP, όπως η ανάλυση συναίσθημάτος, η εξαγωγή οντοτήτων, και η μετάφραση κειμένου. Η SSL επιτρέπει στα μοντέλα να κατανοούν βαθύτερες γλωσσικές δομές και να αποδίδουν καλύτερα σε διάφορα γλωσσικά tasks, ακόμη και με περιορισμένα δεδομένα με ετικέτες.

Παράδειγμα:

Ένα χαρακτηριστικό παράδειγμα εφαρμογής της Αυτο-εποπτεύμενης Μάθησης στην Επεξεργασία Φυσικής Γλώσσας είναι η εκπαίδευση του γλωσσικού μοντέλου GPT (Generative Pre-trained Transformer) [51], το οποίο χρησιμοποιείται για τη δημιουργία κειμένου. Στο στάδιο της προεκπαίδευσης, το GPT χρησιμοποιεί μια έμμεση διεργασία που ονομάζεται "Language Modeling", όπου το μοντέλο μαθαίνει να προβλέπει την επόμενη λέξη σε μια ακολουθία λέξεων. Για παράδειγμα, αν το κείμενο εισόδου είναι "Το καλοκαίρι ο ουρανός είναι", το μοντέλο μαθαίνει να προβλέπει τη λέξη "γαλάζιος". Αυτή η διαδικασία επιτρέπει στο GPT να κατανοεί τη δομή και το περιεχόμενο του κειμένου, καθιστώντας το ικανό να δημιουργεί συνεκτικό και φυσικό κείμενο σε ποικιλία εφαρμογών, όπως αυτόματη παραγωγή περιεχομένου, μετάφραση γλώσσας, και διάλογος με χρήστες.



Εικόνα 2.4.3 - 1: Διαδικασίες pre-training και fine-tuning για το μοντέλο BERT. (Πηγή: Devlin, J. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805.*)

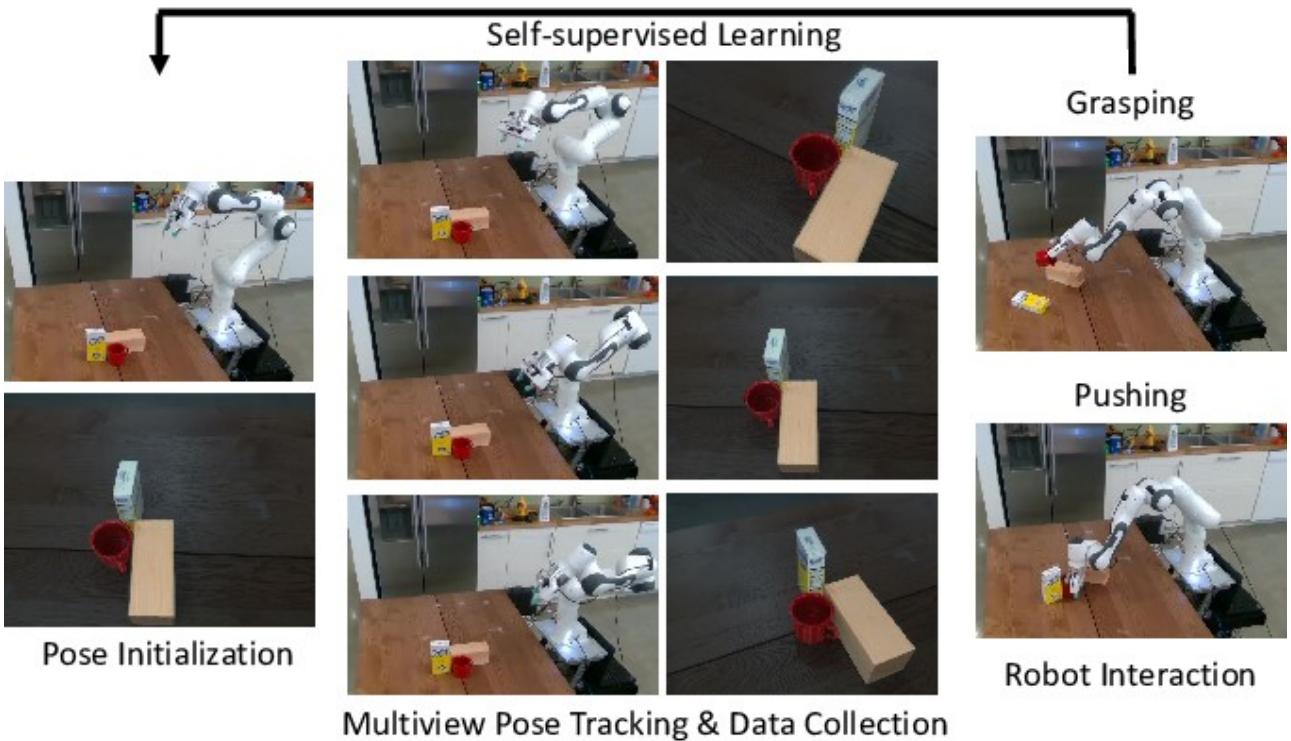
2.4.4 Ρομποτική

Η SSL έχει βρει πολλές εφαρμογές στον τομέα της ρομποτικής, καθώς επιτρέπει στα ρομπότ να μαθαίνουν από τα δεδομένα τους χωρίς την ανάγκη για μεγάλες ποσότητες επισημασμένων δεδομένων, κάτι που είναι συχνά δύσκολο ή ακριβό να αποκτηθεί. Μερικές από τις βασικές εφαρμογές της SSL στη ρομποτική περιλαμβάνουν:

- **Αντίληψη του Περιβάλλοντος:** Τα ρομπότ χρησιμοποιούν αυτόν τον τύπο μάθησης για να βελτιώσουν την ικανότητά τους να κατανοούν και να αντιλαμβάνονται το περιβάλλον γύρω τους. Για παράδειγμα, η SSL μπορεί να χρησιμοποιηθεί για την εκμάθηση της αναγνώρισης αντικειμένων σε περιβάλλοντα όπου δεν υπάρχουν επισημασμένα δεδομένα, επιτρέποντας στα ρομπότ να αναγνωρίζουν και να αλληλεπιδρούν με αντικείμενα με μεγαλύτερη ακρίβεια.
- **Αυτόνομη Πλοήγηση:** Τα ρομπότ μπορούν να χρησιμοποιήσουν SSL για να μάθουν πώς να πλοιηγούνται σε άγνωστα περιβάλλοντα, αναγνωρίζοντας εμπόδια και σχεδιάζοντας ασφαλείς διαδρομές. Αυτό είναι ιδιαίτερα χρήσιμο σε περιπτώσεις όπου τα ρομπότ πρέπει να πλοιηγηθούν σε περιβάλλοντα που δεν έχουν χαρτογραφηθεί εκ των προτέρων.
- **Χειρισμός Αντικειμένων:** Με τη χρήση SSL, τα ρομπότ μπορούν να μάθουν να χειρίζονται διάφορα αντικείμενα, αναπτύσσοντας δεξιότητες όπως η αρπαγή, η μεταφορά και η τοποθέτηση αντικειμένων. Αυτή η διαδικασία μπορεί να περιλαμβάνει την εκμάθηση των φυσικών ιδιοτήτων των αντικειμένων, όπως το βάρος, το σχήμα και η υφή, ώστε να βελτιωθεί η ακρίβεια και η ευαισθησία κατά τον χειρισμό.

Παράδειγμα:

Ένα παράδειγμα εφαρμογής της SSL στη ρομποτική είναι η εκπαίδευση ρομποτικών βραχιόνων για τη συναρμολόγηση εξαρτημάτων χωρίς την ανάγκη ανθρώπινης επίβλεψης. Με τη χρήση SSL, ο ρομποτικός βραχίονας μπορεί να μάθει να συναρμολογεί αντικείμενα αναλύοντας τα δεδομένα κίνησης και αλληλεπίδρασης με τα εξαρτήματα, χωρίς να χρειάζεται να του δοθούν συγκεκριμένες οδηγίες για κάθε βήμα. Αυτή η διαδικασία καθιστά τα ρομπότ πιο ευπροσάρμοστα και ικανά να εκτελούν σύνθετες εργασίες σε πραγματικά περιβάλλοντα παραγωγής.



Multiview Pose Tracking & Data Collection

[Εικόνα 2.4.4 - 1](#): Χρήση Αυτο-εποπτευόμενης Μάθησης στον τομέα της Ρομποτικής. (Πηγή: ResearchGate)

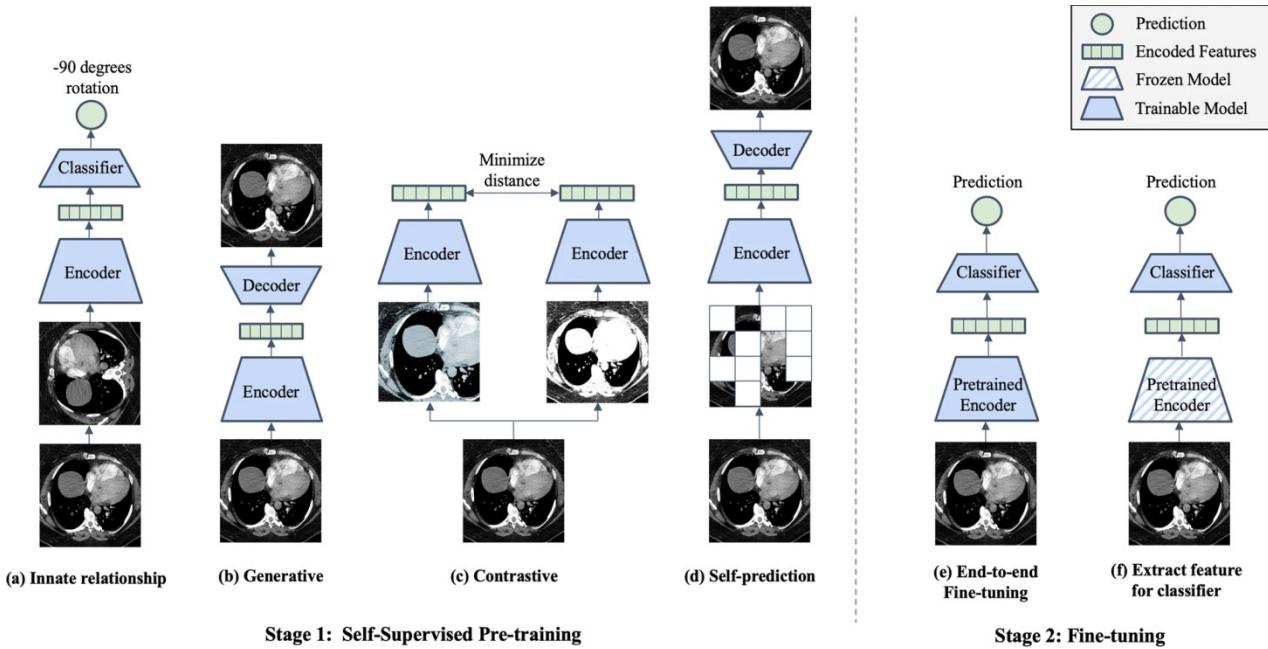
2.4.5 Άλλες Εφαρμογές

Ιατρική: Η Αυτο-εποπτευόμενη Μάθηση έχει αρχίσει να παίζει σημαντικό ρόλο στην ιατρική, ιδίως σε τομείς όπως η ιατρική απεικόνιση και η διάγνωση. Οι ιατρικές εικόνες, όπως οι μαγνητικές τομογραφίες (MRI), οι αξονικές τομογραφίες (CT scans), και οι ακτινογραφίες, παράγουν τεράστιες ποσότητες δεδομένων, ωστόσο η επισήμανση αυτών των δεδομένων είναι χρονοβόρα και απαιτεί εξειδικευμένη γνώση από ιατρούς. Η SSL επιτρέπει στα μοντέλα να μάθουν από αυτά τα μη επισημασμένα δεδομένα, προσφέροντας τις εξής δυνατότητες:

- **Βελτίωση της Διάγνωσης:** Η SSL μπορεί να χρησιμοποιηθεί για να εκπαιδευτούν μοντέλα που εντοπίζουν ανωμαλίες σε ιατρικές εικόνες, ακόμα και όταν δεν υπάρχουν επισημασμένα δεδομένα για όλες τις περιπτώσεις. Για παράδειγμα, η SSL μπορεί να βοηθήσει στην ανίχνευση όγκων, κρυφών βλαβών, ή άλλων ανωμαλιών με μεγαλύτερη ακρίβεια.
- **Αναγνώριση Παθολογιών:** Η SSL μπορεί επίσης να χρησιμοποιηθεί για την αναγνώριση και ταξινόμηση διαφόρων παθολογιών από ιατρικές εικόνες. Η ικανότητα της SSL να μαθαίνει από μη επισημασμένα δεδομένα μειώνει την εξάρτηση από τα επισημασμένα σύνολα δεδομένων, επιτρέποντας την εκπαίδευση μοντέλων που μπορούν να αναγνωρίσουν σπάνιες ή δύσκολες παθήσεις.

Παράδειγμα:

Ένα παράδειγμα εφαρμογής της SSL στην ιατρική είναι η ανάπτυξη μοντέλων για την αυτόματη ανάλυση ακτινογραφιών θώρακα. Χρησιμοποιώντας SSL, τα μοντέλα μπορούν να μάθουν να αναγνωρίζουν πνευμονικές βλάβες και άλλες παθολογικές καταστάσεις χωρίς την ανάγκη για εκτενή επισημασμένα δεδομένα, βελτιώνοντας την ταχύτητα και την ακρίβεια της διάγνωσης.



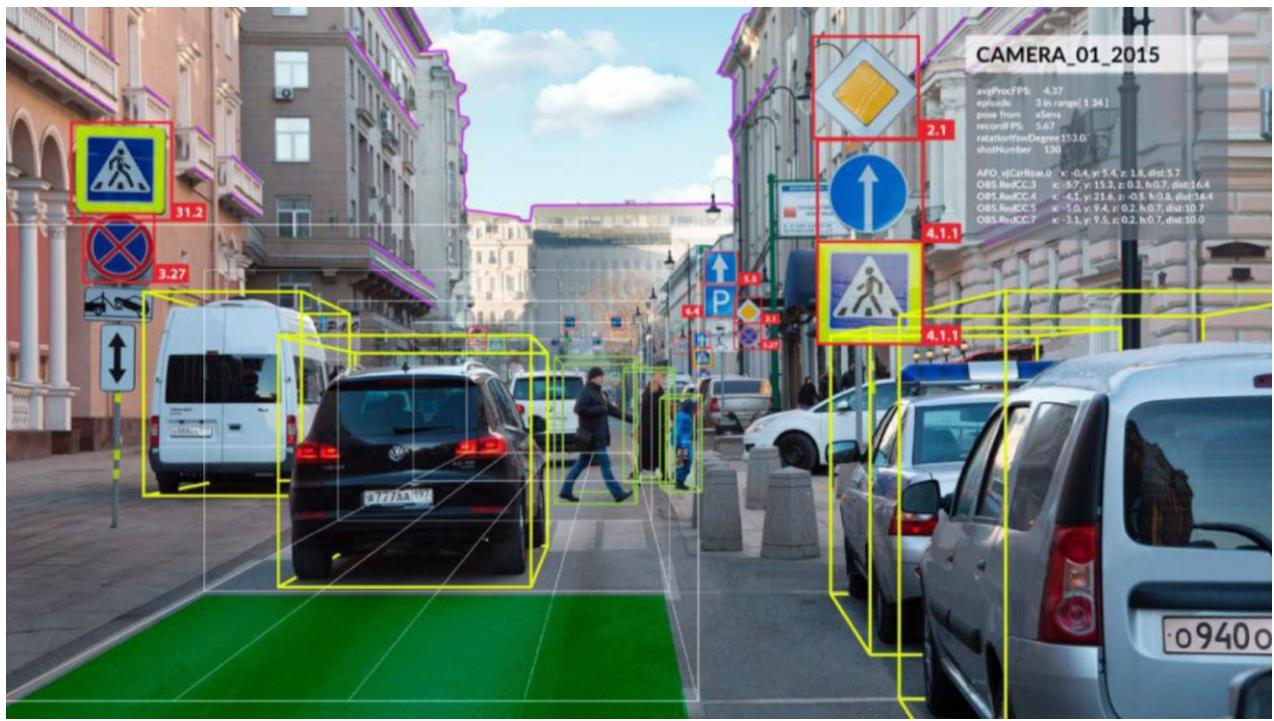
Εικόνα 2.4.5 - 1: Χρήση Αυτο-εποπτευόμενης Μάθησης στον τομέα της Ιατρικής. (Πηγή: TowardsDataScience)

Αυτόνομη Οδήγηση: Στον τομέα της αυτόνομης οδήγησης, η *SSL* συμβάλλει στην ανάπτυξη αυτοκινούμενων οχημάτων που μπορούν να πλοηγούνται με ασφάλεια και ακρίβεια σε διάφορα περιβάλλοντα, χρησιμοποιώντας μη επισημασμένα δεδομένα για την εκπαίδευσή τους. Η *SSL* μπορεί να ενισχύσει την απόδοση των μοντέλων που χρησιμοποιούνται για:

- **Αντίληψη του Περιβάλλοντος:** Τα αυτόνομα οχήματα πρέπει να κατανοούν το περιβάλλον γύρω τους, να αναγνωρίζουν εμπόδια, σήματα κυκλοφορίας, πεζούς, και άλλα οχήματα. Η *SSL* βοηθά στην εκπαίδευση αυτών των μοντέλων με τη χρήση μεγάλων ποσοτήτων μη επισημασμένων δεδομένων, όπως εικόνες και βίντεο από κάμερες, lidar και άλλους αισθητήρες.
- **Λήψη Αποφάσεων:** Η *SSL* μπορεί επίσης να χρησιμοποιηθεί για την ανάπτυξη μοντέλων που λαμβάνουν αποφάσεις σε πραγματικό χρόνο, όπως η αλλαγή λωρίδας, η επιβράδυνση ή η επιτάχυνση, και η πλοήγηση σε πολύπλοκα σενάρια κυκλοφορίας.

Παράδειγμα:

Ένα παράδειγμα εφαρμογής της *SSL* στην αυτόνομη οδήγηση είναι η εκπαίδευση ενός αυτοκινήτου να αναγνωρίζει και να αποφεύγει εμπόδια χρησιμοποιώντας βίντεο από κάμερες που έχουν καταγράψει διάφορες διαδρομές. Το μοντέλο μπορεί να μάθει να προβλέπει την κίνηση των αντικειμένων και να αντιδρά κατάλληλα, διασφαλίζοντας την ασφάλεια των επιβατών και των άλλων χρηστών του δρόμου.



Εικόνα 2.4.5 - 2: Χρήση Αυτο-Εποπτευόμενης Μάθησης στην Αυτόνομη Οδήγηση (Πηγή: ResearchGate)

II

Πειραματικό Μέρος

Κεφάλαιο 3^ο

Διεξαγωγή Πειραμάτων

3.1 Μεθοδολογία Πειραμάτων

Σε αυτή την ενότητα θα αναλυθεί η μεθοδολογία που ακολουθήθηκε κατά τη διάρκεια των πειραμάτων και θα παρουσιαστεί η λογική πίσω από τις συγκρίσεις που πραγματοποιήθηκαν. Η διεξαγωγή των πειραμάτων είχε ως στόχο να αξιολογηθεί η απόδοση των τριών διαφορετικών proxy tasks (*Image Rotation Prediction*, *Image Colorization*, *Image Inpainting*) μέσω της μεταφοράς μάθησης στο downstream task της ταξινόμησης εικόνων. Επιπλέον, θα εξηγήσουμε πώς οργανώθηκαν οι δοκιμές και η διαδικασία fine-tuning σε δύο διαφορετικά σύνολα δεδομένων (CIFAR-10 και CIFAR-100), και θα συζητηθεί η λογική των συγκρίσεων που πραγματοποιήθηκαν μεταξύ των διαφορετικών μεθόδων και παραμέτρων εκπαίδευσης, προκειμένου να διασφαλιστεί η δίκαιη και ακριβής αξιολόγηση των αποτελεσμάτων.

3.1.1 Μεθοδολογία και Λογική Συγκρίσεων

Για να διασφαλίσουμε τη δίκαιη σύγκριση μεταξύ των τριών proxy tasks, επιλέξαμε το *Image Rotation Prediction*, το *Image Colorization* και το *Image Inpainting*. Αυτά τα τρία tasks είναι διαφορετικής φύσης: το πρώτο έχει να κάνει με τον προσανατολισμό των αντικειμένων μέσα στις εικόνες (rotation prediction), το δεύτερο αφορά τα χρώματα των εικόνων (colorization), ενώ το τρίτο εστιάζει στο περιεχόμενο, το πλαίσιο και τη συνοχή των εικόνων (inpainting). Χρησιμοποιήσαμε ως βάση το ResNet-50 για όλα τα tasks. Η ίδια αρχιτεκτονική εφαρμόστηκε σε κάθε task, προσαρμόζοντάς τη κατάλληλα ανάλογα με τις απαιτήσεις του κάθε proxy task. Σε επόμενη ενότητα θα αναλύσουμε ακριβώς πώς υλοποιήθηκε η προσαρμογή του ResNet-50 για κάθε ένα από αυτά τα tasks. Αυτή η ομοιομορφία στην αρχιτεκτονική επιτρέπει την ακριβή σύγκριση των επιδόσεων, χωρίς να επηρεάζονται τα αποτελέσματα από διαφοροποιήσεις στην αρχιτεκτονική του δικτύου.

Η λογική των συγκρίσεων εστιάζει σε δύο βασικές στρατηγικές fine-tuning:

1. Fine-tuning μόνο του τελευταίου επιπέδου: Σε αυτή την περίπτωση, επανεκπαιδεύσαμε μόνο το τελευταίο πλήρως συνδεδεμένο επίπεδο του ResNet-50 στο downstream task της ταξινόμησης εικόνων. Ο στόχος ήταν να δούμε πώς συγκρίνονται τα βάρη που αποκτήθηκαν από τα proxy tasks σε σχέση με τυχαία βάρη (*random weights*), επιτρέποντας μας να αξιολογήσουμε τι πληροφορία έχουν μάθει τα proxy tasks και αν αυτή η πληροφορία είναι χρήσιμη στο downstream task. Η σύγκριση με τα τυχαία βάρη μας δίνει τη δυνατότητα να αξιολογήσουμε κατά πόσο τα βάρη που έμαθε το δίκτυο μέσω του self-supervised learning είναι συμβατά και ωφέλιμα για την ταξινόμηση εικόνων.
2. Fine-tuning ολόκληρου του μοντέλου: Σε αυτή την περίπτωση, επανεκπαιδεύσαμε όλα τα επίπεδα του μοντέλου στο downstream task. Ο στόχος εδώ είναι να δούμε αν η εκπαίδευση των proxy tasks πραγματικά αξίζει να γίνει, βοηθώντας το μοντέλο να βελτιώσει την απόδοσή του σε σχέση με την αρχικοποίηση με τυχαία βάρη. Σημειώνουμε ότι χρησιμοποιήθηκαν οι ίδιες υπερπαράμετροι σε όλες τις περιπτώσεις για να διασφαλιστεί ότι η μόνη διαφορά ανάμεσα στα πειράματα ήταν η αρχική κατάσταση των βαρών.

Δεν χρησιμοποιήσαμε καθόλου *data augmentation* στα πειράματά μας, καθώς αυτό θα έθετε σε κίνδυνο τη δίκαιη σύγκριση. Κάθε proxy task έχει διαφορετικές απαιτήσεις σχετικά με τους μετασχηματισμούς (*transforms*), γεγονός που θα μπορούσε να προκαλέσει προβλήματα κατά την εκπαίδευση. Για παράδειγμα, στο *image rotation prediction task*, εάν εφαρμοστεί ένα *random rotation* 180° σε μια εικόνα κατά την

προεπεξεργασία δεδομένων και της αποδοθεί η κλάση 2 (180° περιστροφή), το ίδιο *random rotation* από τους μετασχηματισμούς μπορεί να την επαναφέρει στην αρχική της θέση. Αυτό θα την έκανε να ανήκει ξανά στην κλάση 0, όμως εμείς θα της είχαμε δώσει λανθασμένα την κλάση 2. Ακόμη και αν το μοντέλο κάνει τη σωστή πρόβλεψη για την κλάση 0, το λάθος στις ετικέτες θα το καταγράψει ως λάθος πρόβλεψη, γεγονός που θα καταστρέψει την εκπαίδευση. Αντίστοιχα προβλήματα προκύπτουν και σε tasks όπως το *colorization*, όπου ο μετασχηματισμός *color jitter* θα μπορούσαν να αλλοιώσουν τα χρώματα, καθιστώντας δύσκολη την ακριβή εκπαίδευση. Σε κανονικές συνθήκες, μπορούμε να χρησιμοποιήσαμε κανονικά τους μετασχηματισμούς (εφόσον δεν καταστρέφουν την εκπαίδευση της έμμεσης διεργασίας) για να επιτύχουμε υψηλότερη ακρίβεια ωστόσο όπως αναφέραμε σκοπός ήταν η δίκαιη σύγκριση και όχι η υψηλότερη δυνατή επίδοση. Για αυτούς τους λόγους, περιοριστήκαμε σε βασικά *transforms*, όπως *resize*, *normalize* και *totensor*. Σημαντικό είναι να τονίσουμε ότι επειδή ο στόχος μας δεν ήταν να πετύχουμε την υψηλότερη δυνατή επίδοση, αλλά να εξασφαλίσουμε δίκαιη σύγκριση των μεθόδων, οι επιδόσεις που θα καταγραφούν θα είναι πιθανώς ελάχιστα χαμηλότερες από τα αναμενόμενα standards.

Τα proxy tasks εκπαιδεύτηκαν στο CIFAR-100 χωρίς τη χρήση των labels του dataset, καθώς το χρησιμοποιήσαμε ως *unlabeled* σύνολο δεδομένων. Στη συνέχεια, τα αποθηκευμένα βάρη από τα proxy tasks μεταφέρθηκαν και τεσταρίστηκαν τόσο στο CIFAR-10 όσο και στο CIFAR-100 για το downstream task του *image classification*. Η χρήση και των δύο datasets ήταν σκόπιμη, καθώς θέλαμε να αξιολογήσουμε την απόδοση τόσο στο ίδιο dataset που χρησιμοποιήθηκε για την εκπαίδευση των proxy tasks (CIFAR-100) όσο και σε ένα διαφορετικό αλλά παρόμοιο dataset (CIFAR-10). Τα δύο σύνολα δεδομένων επιλέχθηκαν λόγω της ομοιότητάς τους, καθώς έχουν το ίδιο μέγεθος και περιέχουν εικόνες παρόμοιας φύσης, γεγονός που επιτρέπει την πιο ομαλή σύγκριση των αποτελεσμάτων.

Εκτός από τη σύγκριση των τελικών επιδόσεων των proxy tasks, εξετάσαμε επίσης το πόσο γρήγορα κάθε proxy task επιτυγχάνει υψηλή απόδοση. Δηλαδή, μας ενδιαφέρει να δούμε όχι μόνο ποιο task αποδίδει καλύτερα, αλλά και ποιο μπορεί να φτάσει σε ένα ικανοποιητικό επίπεδο ακρίβειας σε λιγότερες εποχές εκπαίδευσης. Για παράδειγμα, εάν ένα μοντέλο που έχει εκπαιδευτεί με SSL μπορεί να φτάσει σε ικανοποιητικό επίπεδο ακρίβειας με λιγότερες εποχές σε σχέση με τυχαία βάρη, τότε αυτό θα ήταν οικονομικά και χρονικά πιο αποδοτικό. Τα διαγράμματα που απεικονίζουν τη διαφορά στην απόδοση μεταξύ των proxy task weights και των random weights μας βοηθούν να αξιολογήσουμε αυτήν τη διάσταση.

Τέλος, πρέπει να σημειωθεί ότι δεν πραγματοποιήθηκε εκτενές *hyperparameter tuning* στο downstream task ανάλογα με την χρήση των διαφορετικών βαρών από τα διαφορετικά tasks. Η λογική μας ήταν να διατηρήσουμε τις υπερπαραμέτρους σταθερές σε όλα τα πειράματα για να διασφαλίσουμε τη δίκαιη σύγκριση. Αυτό σημαίνει ότι δεν επιδιώξαμε να πιάσουμε τις καλύτερες δυνατές επιδόσεις, αλλά να δημιουργήσουμε ένα ισοδύναμο πειραματικό περιβάλλον για τη σύγκριση των μεθόδων. Επομένως, οι επιδόσεις των μοντέλων δεν ανταγωνίζονται τα βέλτιστα πρότυπα απόδοσης, αλλά αποσκοπούν στην απόδειξη της αποτελεσματικότητας των proxy tasks σε σχέση με τα random weights και στην ανάδειξη της πιο αποδοτικής διεργασίας μεταξύ των proxy tasks ως προ-εκπαιδευτική μέθοδος των βαρών για το downstream task της ταξινόμησης εικόνων.

3.1.2 Ερωτήματα και Προσδοκίες

Με βάση τη μεθοδολογία που αναλύθηκε στην προηγούμενη ενότητα, τα κύρια ερευνητικά ερωτήματα και προσδοκίες των πειραμάτων μας συνοψίζονται ως εξής:

Επιδόσεις με τη χρήση Αυτο-Εποπτευόμενης Μάθησης σε σχέση με τα τυχαία βάρη

Αρχικό και βασικό μας ερώτημα είναι εάν η χρήση των proxy tasks μέσω της αυτο-εποπτευόμενης μάθησης (*self-supervised learning*) μπορεί να προσφέρει καλύτερα αποτελέσματα από τα τυχαία βάρη. Αναμένουμε ότι σε σχεδόν όλες τις μετρήσεις, οι επιδόσεις θα είναι καλύτερες με τη χρήση των προεκπαιδευμένων βαρών, καθώς αυτά θα έχουν μάθει κάποια βασικά χαρακτηριστικά από τα proxy tasks, σε αντίθεση με τα

τυχαία βάρη που δεν φέρουν καμία πληροφορία.

Επιδόσεις στα διαφορετικά datasets (CIFAR-10 vs CIFAR-100)

Αναμένουμε ότι οι επιδόσεις στο CIFAR-100 θα είναι χαμηλότερες σε σύγκριση με το CIFAR-10, καθώς το CIFAR-100 έχει 100 κλάσεις, ενώ το CIFAR-10 έχει μόνο 10. Αυτό σημαίνει ότι το task ταξινόμησης είναι πιο πολύπλοκο στο CIFAR-100. Ωστόσο, περιμένουμε ότι τα pretrained weights που εκπαιδεύτηκαν στο CIFAR-100 θα λειτουργήσουν εξίσου καλά και στο CIFAR-10, καθώς τα δύο datasets περιέχουν παρόμοιες εικόνες και χαρακτηριστικά. Επομένως, αναμένουμε ότι τα proxy tasks θα είναι ικανά να μεταφέρουν τις γνώσεις τους στο νέο dataset και να διατηρήσουν καλές αποδόσεις.

Fine-tuning μόνο στο τελευταίο επίπεδο

Σε αυτή την περίπτωση, το ερώτημα είναι κατά πόσο το fine-tuning μόνο του τελευταίου επιπέδου του δικτύου θα αποφέρει καλύτερα αποτελέσματα με τα proxy tasks σε σχέση με τα τυχαία βάρη. Αναμένουμε ότι οι επιδόσεις των proxy tasks θα είναι σημαντικά καλύτερες, καθώς τα βάρη του δικτύου έχουν ήδη εκπαιδευτεί σε ένα συναφές task. Εδώ, τα proxy tasks παρέχουν προεκπαιδευμένα βάρη που περιέχουν χρήσιμη πληροφορία, ενώ τα τυχαία βάρη απαιτούν να μάθουν τα πάντα από το μηδέν. Επομένως, αφού εκπαιδεύουμε μόνο το τελευταίο επίπεδο, και δεδομένου ότι τα τυχαία βάρη δεν έχουν υποστεί εκπαίδευση εκτός από το τελευταίο επίπεδο, ενώ τα προεκπαιδευμένα βάρη περιέχουν ήδη κάποια σχετική πληροφορία στο μεγαλύτερο ποσοστό του μοντέλου, αναμένουμε σημαντικά καλύτερες επιδόσεις από τα proxy tasks.

Fine-tuning σε ολόκληρο το μοντέλο

Το ερώτημα εδώ είναι εάν το fine-tuning ολόκληρου του μοντέλου με τα προεκπαιδευμένα βάρη θα είναι καλύτερο από το fine-tuning ολόκληρου του μοντέλου με τυχαία βάρη. Προφανώς, περιμένουμε να πετύχουμε πολύ καλύτερες επιδόσεις σε σύγκριση με το fine-tuning μόνο του τελευταίου επιπέδου, καθώς σε αυτή την περίπτωση επανεκπαιδεύονται όλα τα επίπεδα του δικτύου και όχι μόνο το τελευταίο. Όσον αφορά το fine-tuning ολόκληρου του μοντέλου, αναμένουμε ότι οι τελικές επιδόσεις με τα pre-trained weights θα είναι ελάχιστα καλύτερες από αυτές με τα random weights, καθώς τα βάρη είναι μόνο τα αρχικά - είτε προεκπαιδευμένα είτε τυχαία - και στη συνέχεια όλα τροποποιούνται κατά τη διάρκεια της εκπαίδευσης. Επομένως, περιμένουμε μία ελαφρώς καλύτερη σύγκλιση (convergence) στο τέλος με τα pre-trained weights, αλλά η διαφορά δεν θα είναι δραματική.

Ταχύτητα επίτευξης ικανοποιητικής επίδοσης με fine-tuning όλων των επιπέδων του μοντέλου

Εκτός από τις τελικές επιδόσεις, μας ενδιαφέρει και η ταχύτητα με την οποία τα proxy tasks θα βοηθήσουν το μοντέλο να φτάσει σε ένα ικανοποιητικό επίπεδο απόδοσης. Αναμένουμε ότι τα proxy tasks θα επιτρέψουν στο μοντέλο να πετύχει γρήγορα υψηλά επίπεδα ακρίβειας. Πιστεύουμε ότι σε πολύ λιγότερες εποχές από ότι με τα τυχαία βάρη, το μοντέλο θα φτάσει σε υψηλή ακρίβεια.

Έμμεση διεργασία με τις καλύτερες επιδόσεις

Ένα βασικό ερώτημα που θέλουμε να απαντήσουμε είναι ποιο από τα τρία proxy tasks (Image Rotation Prediction, Image Colorization, Image Inpainting) θα αποδειχθεί πιο αποτελεσματικό στην ενίσχυση της απόδοσης του downstream task. Περιμένουμε ότι το *colorization* ή το *inpainting* θα αποδειχθούν τα πιο αποδοτικά, καθώς είναι πιο σύνθετες διεργασίες σε σχέση με την πρόβλεψη περιστροφής, και επομένως προσφέρουν περισσότερη πληροφορία στο μοντέλο.

3.2 Υλοποίηση Έμμεσων Διεργασιών

Στην ενότητα αυτή, θα παρουσιαστεί η υλοποίηση τριών έμμεσων διεργασιών που χρησιμοποιήθηκαν για την εκπαίδευση του μοντέλου μέσω αυτο-εποπτεύμενης μάθησης: η πρόβλεψη περιστροφής εικόνας, η χρωματοποίηση εικόνας, και η επιδιόρθωση/συμπλήρωση εικόνας (image inpainting). Θα αναλύσουμε τον τρόπο με τον οποίο υλοποιήθηκαν αυτά τα proxy/pretext tasks, τα εργαλεία που χρησιμοποιήθηκαν για την κατασκευή τους, καθώς και τις υπερπαραμέτρους που επιλέχθηκαν για την εκπαίδευση των μοντέλων σε αυτά τα tasks. Επιπλέον θα παρουσιαστούν εικόνες που δείχνουν την επεξεργασία των αρχικών εικόνων και τις προβλέψεις του εκάστοτε μοντέλου.

Οι φωτογραφίες που χρησιμοποιήθηκαν για την απεικόνιση της απόδοσης των μοντέλων δεν προέρχονται από το σύνολο δεδομένων που χρησιμοποιήθηκε στα πειράματα, το οποίο θα περιγραφεί σε επόμενη ενότητα. Αυτό συμβαίνει επειδή οι εικόνες του συγκεκριμένου συνόλου δεδομένων είναι πολύ μικρού μεγέθους, γεγονός που δυσκολεύει την οπτική αντίληψη των διαφορών από το ανθρώπινο μάτι. Αντ' αυτού, χρησιμοποιήθηκαν εικόνες από το σύνολο δεδομένων Caltech-256 [52], το οποίο περιλαμβάνει εικόνες με επαρκείς διαστάσεις, επιτρέποντας έτσι την καλύτερη παρατήρηση και κατανόηση των διαφορών μεταξύ των εικόνων.

3.2.1 Υλοποίηση Διεργασίας Πρόβλεψης Περιστροφής Εικόνας

Προεπεξεργασία Δεδομένων: Η διαδικασία προεπεξεργασίας δεδομένων στην διεργασία πρόβλεψης περιστροφής εικόνας περιλαμβάνει αρχικά την αναπροσαρμογή των εικόνων σε σταθερό μέγεθος 224x224 pixels (μετασχηματισμός “*Resize()*”), ώστε να διασφαλιστεί η ομοιομορφία των εισόδων στο μοντέλο. Επίσης οι εικόνες υφίστανται μετασχηματισμούς, όπως η μετατροπή τους σε tensors (μετασχηματισμός “*ToTensor*” - οι τιμές των pixels μετατρέπονται από το εύρος [0, 255] σε [0,1], μας επιστρέφεται ο τένσορας με διαστασεis [Κανάλια, Ύψος, Πλάτος]), καθώς και κανονικοποίηση (μετασχηματισμός “*Normalize(mean, std)*” - κανονικοποιεί τα δεδομένα εισόδου ώστε η εικόνα να έχει συγκεκριμένη μέση τιμή(μean) και τυπική απόκλιση (std)) κάτι που επιτρέπει την αποτελεσματική επεξεργασία τους από τα επίπεδα του νευρωνικού δικτύου.

Αυτή η διαδικασία προετοιμάζει τα δεδομένα ώστε να είναι έτοιμα για εκπαίδευση, ενώ η τυχαία περιστροφή σε κάθε εποχή διασφαλίζει ότι το μοντέλο εκτίθεται σε διαφορετικές γωνίες της ίδιας εικόνας, ενισχύοντας έτσι την ικανότητά του να μάθει τις χωρικές δομές των εικόνων. Στις εικόνες που ανακτώνται από τον φάκελο, εφαρμόζεται τυχαία περιστροφή σε γωνίες 0°, 90°, 180° ή 270°.

Ο κύριος στόχος είναι η ταξινόμηση των περιστροφών, όπου κάθε εικόνα ανήκει σε μία από τις τέσσερις κατηγορίες, ανάλογα με τη γωνία περιστροφής της. Η τιμή της ετικέτας καθορίζεται από τον τύπο: *ετικέτα = γωνία περιστροφής / 90*, δίνοντας τιμές ετικέτας 0, 1, 2 ή 3 για τις αντίστοιχες περιστροφές 0°, 90°, 180° και 270°. Έτσι, οι 4 κατηγορίες προκύπτουν από τις διαφορετικές γωνίες, και το μοντέλο πρέπει να μάθει να αναγνωρίζει σωστά την περιστροφή κάθε εικόνας. Θα μπορούσαμε να πούμε ότι αναγάγουμε τη συγκεκριμένη διεργασία σε μια διεργασία ταξινόμησης εικόνας, αλλά με ετικέτες που υποδηλώνουν τη γωνία περιστροφής αντί για το περιεχόμενο της εικόνας όπως στην κλασική διεργασία ταξινόμησης εικόνων.

Αρχιτεκτονική: Η αρχιτεκτονική που χρησιμοποιήθηκε είναι το μοντέλο *ResNet50* που αποτελείται από 50 στρώσεις (layers), οι οποίες περιλαμβάνουν συνελικτικά στρώματα (convolutional layers), στρώματα pooling, και πλήρως συνδεδεμένα στρώματα. Οι τελευταίες στρώσεις είναι πλήρως συνδεδεμένες, με την τελική να έχει 1000 εξόδους για προβλήματα ταξινόμησης στο *ImageNet*.

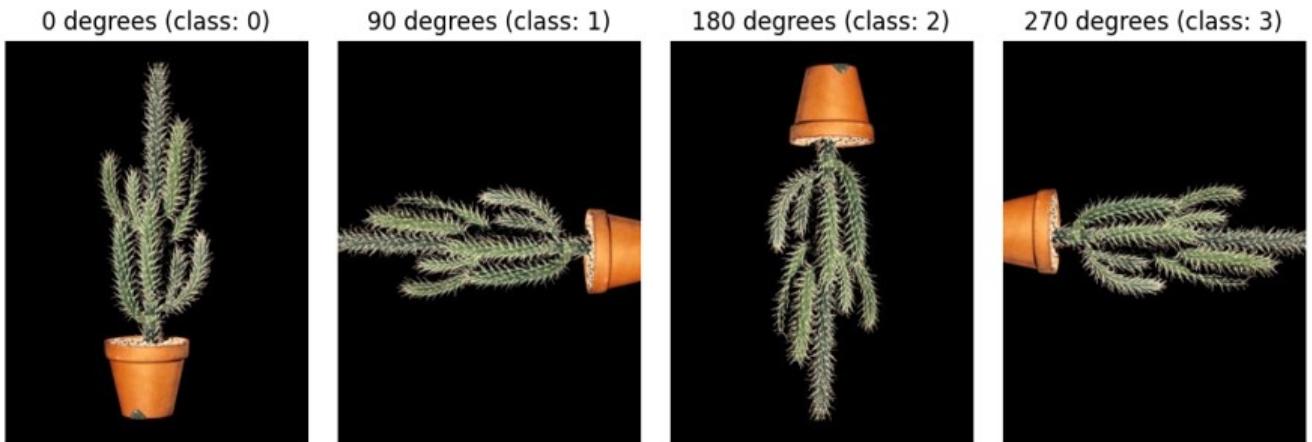
Στο συγκεκριμένο task της πρόβλεψης περιστροφής εικόνας, αυτή η τελική στρώση τροποποιήθηκε ώστε να έχει 4 εξόδους, μία για κάθε πιθανή γωνία περιστροφής (0°, 90°, 180°, 270°). Το υπόλοιπο του δικτύου

παραμένει το ίδιο και προσαρμόζεται ώστε να εξάγει χαρακτηριστικά που βοηθούν στην πρόβλεψη της σωστής γωνίας.

Εκπαίδευση: Η εκπαίδευση του μοντέλου για το task της πρόβλεψης περιστροφής εικόνας υλοποιήθηκε με τη χρήση του *CrossEntropyLoss*, το οποίο είναι κατάλληλο για προβλήματα ταξινόμησης όπως αυτό. Το *CrossEntropyLoss* μετρά τη διαφορά μεταξύ της προβλεπόμενης κατανομής πιθανοτήτων του μοντέλου και των πραγματικών ετικετών. Ειδικά για το συγκεκριμένο task, οι τέσσερις γωνίες περιστροφής (0° , 90° , 180° , 270°) αντιστοιχούν σε τέσσερις κατηγορίες, και το *CrossEntropyLoss* είναι ιδανικό για ταξινομήσεις πολλαπλών κατηγοριών.

Η επιλογή του Adam optimizer έγινε γιατί παρέχει γρήγορη σύγκλιση και δυναμική προσαρμογή του learning rate, βελτιστοποιώντας το μοντέλο με τη χρήση των παραγώγων της συνάρτησης απώλειας για την ενημέρωση των βαρών. Το learning rate έχει οριστεί στο 0.001 ώστε να διατηρεί τη σταθερότητα του μοντέλου, ενώ το weight decay στο 0.0001 βοηθάει στην αποτροπή της υπερπροσαρμογής μειώνοντας την πολυπλοκότητα του μοντέλου.

Το μοντέλο εκπαιδεύεται για 50 εποχές με 64 παρτίδες (batches), λαμβάνοντας δεδομένα που περιστρέφονται σε τυχαίες γωνίες και προσπαθώντας να προβλέψει σωστά την γωνία κάθε φορά.



Εικόνα 3.2.1 - 1: Παραδείγματα περιστροφής εικόνας για το task του Image Rotation Prediction (Caltech-256 Dataset): Η αρχική εικόνα περιστρέφεται κατά 0° , 90° , 180° , και 270° , με κάθε γωνία να αντιστοιχεί σε μια κλάση (0, 1, 2, 3).

3.2.2 Υλοποίηση Διεργασίας Χρωματοποίησης Εικόνας

Προεπεξεργασία Δεδομένων: Αρχικά χρησιμοποιούνται τα ίδια transforms όπως και στα υπόλοιπα tasks. Αυτά περιλαμβάνουν τη μετατροπή των μεγέθους των εικόνων σε σταθερές διαστάσεις με το *Resize*, την κανονικοποίηση των τιμών των pixels με το *Normalize*, και τη μετατροπή της εικόνας σε μορφή tensor μέσω του *ToTensor*, ώστε να είναι κατάλληλη για εισαγωγή στο νευρωνικό δίκτυο.

Στη συνέχεια, για να δημιουργηθούν τα κατάλληλα δεδομένα εισόδου για το μοντέλο, οι εικόνες μετατρέπονται από έγχρωμες (RGB) σε ασπρόμαυρες (grayscale). Η μετατροπή αυτή είναι απαραίτητη καθώς το task του *image colorization* στοχεύει στο να εκπαιδεύσει το μοντέλο να "χρωματίσει" μια ασπρόμαυρη εικόνα. Για να μπορέσει όμως αυτή η ασπρόμαυρη εικόνα να εισαχθεί στο δίκτυο, χρειάζεται να ταιριάξει τη μορφή της με αυτή των έγχρωμων εικόνων, δηλαδή να έχει τρία κανάλια. Γι' αυτό επαναλαμβάνουμε την *grayscale* εικόνα τρεις φορές, δημιουργώντας ένα *fake RGB* format. Κάθε ένα από τα τρία κανάλια της νέας εικόνας έχει την ίδια πληροφορία με την *grayscale* εικόνα, έτσι ώστε η είσοδος να έχει τη σωστή διαστασιολόγηση για το νευρωνικό δίκτυο, που είναι σχεδιασμένο να δουλεύει με εικόνες

τριών καναλιών.

Τέλος, κάθε δείγμα δεδομένων που επιστρέφεται περιέχει δύο στοιχεία: την ασπρόμαυρη εικόνα που χρησιμοποιείται ως είσοδος στο μοντέλο, και την έγχρωμη εικόνα ως στόχο, την οποία το μοντέλο καλείται να ανακατασκευάσει με την κατάλληλη χρωματική πληροφορία.

Αρχιτεκτονική: Η αρχιτεκτονική του δικτύου που χρησιμοποιείται για το task του *image colorization*, είναι μια μορφή autoencoder με βάση το ResNet50. Αρχικά, αφαιρούνται τα τελικά πλήρως συνδεδεμένα επίπεδα (fully connected layers) του ResNet, καθώς αυτά είναι σχεδιασμένα για ταξινόμηση και δεν είναι κατάλληλα για το task της χρωματοποίησης εικόνας. Αντικαθίστανται με διαδοχικά επίπεδα αποκωδικοποίησης (decoder), που βασίζονται σε transposed convolutions.

Το δίκτυο λειτουργεί με τον εξής τρόπο: Υπάρχουν 5 επίπεδα του encoder και 5 επίπεδα του decoder. Κατά την κωδικοποίηση (encoder), τα χαρακτηριστικά της εικόνας περνούν μέσα από τα συνελικτικά επίπεδα του ResNet, που εξάγουν πληροφορίες υψηλού επιπέδου από την εικόνα. Στη συνέχεια, κατά την αποκωδικοποίηση (decoder), το μοντέλο χρησιμοποιεί transposed convolutions για να επαναφέρει το μέγεθος της εικόνας σε κανονικό επίπεδο. Πιο συγκεκριμένα, τα transposed convolutions, γνωστά και ως deconvolutions ή upsampling convolutions, είναι μια διαδικασία που αντιστρέφει τη λειτουργία των κανονικών συνελικτικών επιπέδων. Αντί να μειώνουν το μέγεθος των χαρακτηριστικών της εικόνας, όπως κάνουν τα κανονικά convolutions, τα transposed convolutions αυξάνουν το μέγεθος, επαναφέροντας τη διαστασιολογία της εικόνας στα αρχικά της επίπεδα. Αυτό τα καθιστά ιδιαίτερα χρήσιμα στα decoders, όπως στα autoencoders, όπου το μοντέλο χρειάζεται να ανακατασκευάσει εικόνες από μειωμένες αναπαραστάσεις. Σε κάθε επίπεδο του decoder εφαρμόζονται λειτουργίες BatchNorm2D και ReLU για κανονικοποίηση και εισαγωγή μη-γραμμικότητας, που συμβάλλουν στη σταθεροποίηση της εκπαίδευσης και τη βελτίωση των προβλέψεων.

Σημαντικό ρόλο παίζουν οι συνδέσεις παράκαμψης (skip connections) μεταξύ των επιπέδων του encoder και του decoder. Αυτές οι συνδέσεις διασφαλίζουν ότι οι πληροφορίες από τα αρχικά στάδια της επεξεργασίας της εικόνας δεν χάνονται κατά την αποκωδικοποίηση. Έτσι, το μοντέλο καταφέρνει να αναπαραγάγει λεπτομέρειες και υφές με μεγαλύτερη ακρίβεια.

Το τελικό επίπεδο του μοντέλου είναι υπεύθυνο για την παραγωγή της ανακατασκευασμένης εικόνας, διασφαλίζοντας ότι το αποτέλεσμα έχει το ίδιο μέγεθος και τα ίδια χαρακτηριστικά με την αρχική εικόνα.

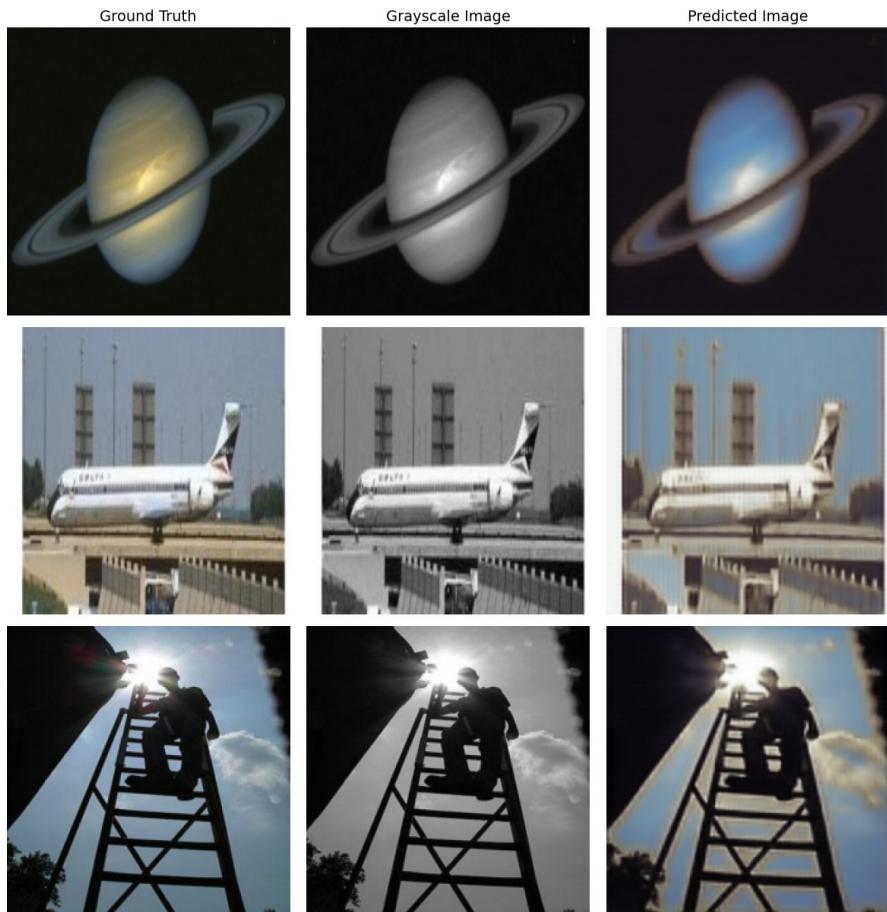
Εκπαίδευση: Για το task του *image colorization*, η διαδικασία εκπαίδευσης επικεντρώνεται στην εκμάθηση του μοντέλου να ανακατασκευάζει ολόκληρη την εικόνα, προσθέτοντας τη σωστή χρωματική πληροφορία στα ασπρόμαυρα δεδομένα που του παρέχονται ως είσοδος. Κατά τη διάρκεια της εκπαίδευσης, γίνεται χρήση της συνάρτησης απώλειας MSE Loss (Mean Squared Error), η οποία μετράει το τετραγωνικό σφάλμα μεταξύ των προβλεπόμενων χρωμάτων και των πραγματικών τιμών των pixels σε όλη την εικόνα.

Σε κάθε βήμα της εκπαίδευσης, το μοντέλο λαμβάνει ως είσοδο μια ασπρόμαυρη εικόνα και προσπαθεί να παράγει την αντίστοιχη έγχρωμη εκδοχή της. Η συνάρτηση MSE Loss υπολογίζει τη διαφορά ανάμεσα στις προβλεπόμενες τιμές των pixels και στις πραγματικές τιμές των χρωμάτων (RGB) σε ολόκληρη την εικόνα. Σε αντίθεση με το *image inpainting* που θα δούμε παρακάτω, όπου η απώλεια (loss) υπολογίζεται μόνο για τα καλυμμένα τμήματα, εδώ η MSE Loss εφαρμόζεται σε κάθε pixel της εικόνας, καθώς το μοντέλο πρέπει να χρωματίσει πλήρως όλες τις περιοχές της εικόνας.

Η επιλογή της MSE Loss είναι ιδανική για το *image colorization* επειδή τα δεδομένα της εικόνας είναι συνεχείς τιμές (τιμές pixels), και η MSE είναι κατάλληλη για την ελαχιστοποίηση των διαφορών μεταξύ της προβλεψης και της πραγματικής εικόνας. Στοχεύει στο να προσαρμόζει το χρώμα σε κάθε pixel της εικόνας με τέτοιο τρόπο ώστε οι προβλεπόμενες τιμές να πλησιάζουν όσο το δυνατόν περισσότερο τις πραγματικές τιμές χρώματος.

Κατά την εκπαίδευση, το μοντέλο περνά από πολλαπλά βήματα, γνωστά ως εποχές, στα οποία επεξεργάζεται εικόνες από το *training set*. Το σύνολο δεδομένων χωρίζεται σε *batches* με μέγεθος 64, και το μοντέλο εκτελεί επαναλαμβανόμενες ενημερώσεις των παραμέτρων του χρησιμοποιώντας τον βελτιστοποιητή Adam, με ρυθμό μάθησης 0.001. Κατά τη διάρκεια κάθε epoch (50 συνολικά), το μοντέλο εκπαιδεύεται και η απώλεια (loss) υπολογίζεται και συσσωρεύεται για να παρακολουθείται η συνολική απόδοση.

Παρακάτω παρατίθενται ορισμένες προβλέψεις του μοντέλου σε εικόνες του Caltech-256 Dataset (για την καλύτερη κατανόηση των διαφορών των εικόνων), οι οποίες απεικονίζουν τη διαδικασία χρωματοποίησης των ασπρόμαυρων εικόνων (ορισμένες από τις προβλέψεις του μοντέλου είναι επιτυχημένες, ενώ άλλες δεν αποδίδουν το χρωματικό αποτέλεσμα με την ίδια ακρίβεια):



Εικόνα 3.2.2 - 1: Παραδείγματα από το task της χρωματοποίησης εικόνων (image colorization). Στην πρώτη στήλη εμφανίζονται οι αρχικές έγχρωμες εικόνες (Ground Truth), στη δεύτερη στήλη οι ασπρόμαυρες εκδοχές των εικόνων (Grayscale Image) και στην τρίτη στήλη οι προβλέψεις του μοντέλου για την επαναφορά των χρωμάτων (Predicted Image).

3.2.3 Υλοποίηση Διεργασίας Επιδιόρθωσης/Συμπλήρωσης Εικόνας

Προεπεξεργασία Δεδομένων: Στην υλοποίηση της διαδικασίας για το image inpainting task, αρχικά κάθε εικόνα προετοιμάζεται με τη μετατροπή της σε ομοιόμορφο μέγεθος και μορφή που μπορεί να επεξεργαστεί το μοντέλο μετασχηματίζοντας τις διαστάσεις της εικόνας σε 224x224 (Resize), κανονικοποιώντας την (Normalize) και μετατρέποντας την σε τένσορα (ToTensor), ακριβώς όπως και στα

προηγούμενα tasks. Η επεξεργασία περιλαμβάνει την εφαρμογή μιας δυαδικής μάσκας στην εικόνα, η οποία αποτελείται από τιμές 0 και 1, όπου τα 0 αναπαριστούν τις περιοχές που θέλουμε να καλύψουμε και να συμπληρώσει το μοντέλο, και τα 1 αναπαριστούν τις περιοχές που παραμένουν ακέραιες.

Η μάσκα δημιουργείται τυχαία με τη χρήση γεωμετρικών σχημάτων, όπως γραμμές και κύκλοι, που τοποθετούνται σε τυχαία σημεία στην εικόνα. Αυτό έχει ως αποτέλεσμα την κάλυψη ορισμένων περιοχών της εικόνας, τις οποίες το μοντέλο καλείται να ανακατασκευάσει. Το επόμενο βήμα είναι ο πολλαπλασιασμός της εικόνας με τη μάσκα, δημιουργώντας την "masked" εικόνα.

Σε κάθε εποχή εκπαίδευσης, η μάσκα εφαρμόζεται τυχαία, δηλαδή κάθε εικόνα μπορεί να λάβει διαφορετική μάσκα σε κάθε επανάληψη. Αυτό βοηθά το μοντέλο να εκπαιδευτεί πάνω σε πολλές παραλαγές της ίδιας εικόνας, βελτιώνοντας την ικανότητά του να γενικεύει και να συμπληρώνει σωστά τα κενά.

Αρχιτεκτονική: Για το task του image inpainting, χρησιμοποιείται η ίδια αρχιτεκτονική με αυτή που περιγράφηκε για το image colorization, δηλαδή ένα autoencoder βασισμένο στο ResNet50. Η κύρια διαφορά είναι ότι το μοντέλο εστιάζει στην επιδιόρθωση των κενών περιοχών της εικόνας όπως θα περιγράψουμε και παρακάτω στην εκπαίδευση, ενώ για την ανακατασκευή χρησιμοποιούνται τα ίδια επίπεδα αποκωδικοποίησης με transposed convolutions και οι συνδέσεις παράκαμψης (skip connections) για διατήρηση των λεπτομερειών.

Εκπαίδευση: Κατά την εκπαίδευση του image inpainting, έγινε χρήση της συνάρτησης MSE Loss (Mean Squared Error), η οποία είναι ιδανική για τη συγκεκριμένη περίπτωση, καθώς στοχεύει στην ελαχιστοποίηση της διαφοράς ανάμεσα στις προβλεπόμενες και πραγματικές τιμές των pixels μόνο στα καλυμμένα μέρη της εικόνας (masked regions). Η MSE Loss υπολογίζει το τετραγωνικό σφάλμα μεταξύ των προβλεπόμενων και πραγματικών τιμών των pixels. Σε αυτό το task, όμως, η απώλεια (loss) υπολογίζεται μόνο για τα καλυμμένα (masked) μέρη της εικόνας, δηλαδή τα τμήματα που έχουν αλλοιωθεί σκόπιμα.

Η υπόλοιπη εικόνα, που παραμένει άθικτη, δεν επηρεάζεται από το μοντέλο. Ο σκοπός είναι το μοντέλο να μάθει να ανακατασκευάζει μόνο τα χαμένα τμήματα, διατηρώντας τις δομές και τις λεπτομέρειες της αρχικής εικόνας. Έτσι, με τη χρήση της MSE Loss για τα masked κομμάτια, το μοντέλο μαθαίνει να παράγει αποτελέσματα που προσεγγίζουν όσο το δυνατόν περισσότερο τα πραγματικά δεδομένα στα καλυμμένα σημεία, χωρίς να επηρεάζει τις υπόλοιπες περιοχές.

Η επιλογή της MSE είναι κατάλληλη, καθώς τα δεδομένα της εικόνας είναι συνεχής πληροφορία, και η MSE εξασφαλίζει την ελαχιστοποίηση της διαφοράς μεταξύ της πρόβλεψης και της πραγματικής εικόνας με τρόπο που λαμβάνει υπόψη τις τιμές των pixels και τις μεταβολές τους.

Συνοψίζοντας, η διαδικασία εκπαίδευσης για το image inpainting ακολουθεί τα εξής βήματα: Κάθε εικόνα αφού πολλαπλασιαστεί με μια τυχαία παραγόμενη μάσκα, η οποία καλύπτει ορισμένα τμήματα της εικόνας έχει μία masked μορφή. Η masked εικόνα χρησιμοποιείται ως είσοδος στο μοντέλο, το οποίο προσπαθεί να προβλέψει τα καλυμμένα σημεία. Η πρόβλεψη συγκρίνεται μόνο με τα masked μέρη της αρχικής εικόνας, ενώ η υπόλοιπη εικόνα παραμένει αμετάβλητη. Η διαδικασία αυτή εκτελείται για 50 epochs με batch size 64, εξασφαλίζοντας σταδιακή βελτίωση της ικανότητας του μοντέλου να ανακατασκευάζει τα χαμένα τμήματα. Για την βελτιστοποίηση χρησιμοποιήθηκε ο Adam Optimizer με learning rate ίσο με 0.0001. Σε κάθε epoch, μια διαφορετική τυχαία μάσκα εφαρμόζεται σε κάθε εικόνα, γεγονός που επιτρέπει στο μοντέλο να βλέπει την ίδια εικόνα με πολλές διαφορετικές απώλειες δεδομένων (μάσκες). Αυτό ενισχύει την ικανότητά του να γενικεύει και να προσαρμόζεται σε διαφορετικά σενάρια όπου τμήματα της εικόνας λείπουν.

Στη συνέχεια, παρουσιάζονται κάποιες ενδεικτικές προβλέψεις του μοντέλου σε εικόνες του Caltech-256 Dataset (για την καλύτερη αντίληψη των διαφορών των εικόνων) όπου φαίνεται η επιδιόρθωση/συμπλήρωση των τυχαίων masked περιοχών των εικόνων (μερικές από τις προβλέψεις του μοντέλου είναι

πετυχημένες, ενώ σε άλλες περιπτώσεις η αποκατάσταση των κενών είναι λιγότερο ακριβής):



Εικόνα 3.2.3 - 1: Παραδείγματα εικόνων από το task της επιδιόρθωσης/συμπλήρωσης εικόνας (image inpainting). Στην πρώτη στήλη απεικονίζεται η αρχική εικόνα (Ground Truth), στη δεύτερη στήλη παρουσιάζεται η εικόνα με τα καλυμμένα τμήματα (Masked Image), και στην τρίτη στήλη η εικόνα που προβλέφθηκε από το μοντέλο (Predicted Image).

3.3 Downstream Tasks και Σύνολα Δεδομένων

Τα downstream tasks αποτελούν το επόμενο στάδιο όπου αξιολογείται η απόδοση των προεκπαίδευμένων μοντέλων σε πραγματικές εφαρμογές, όπως η ταξινόμηση εικόνων. Στην περίπτωση αυτή, το μοντέλο που εκπαιδεύτηκε σε proxy tasks, όπως η πρόβλεψη περιστροφής, η χρωματοποίηση ή η συμπλήρωση εικόνας, δοκιμάζεται στο downstream task της ταξινόμησης εικόνων. Αυτό το στάδιο είναι κρίσιμο, καθώς αναδεικνύει κατά πόσο τα χαρακτηριστικά που έμαθε το μοντέλο από τα proxy tasks μπορούν να γενικευτούν σε πιο πρακτικές και απαιτητικές εφαρμογές. Επιπλέον, η επιλογή των κατάλληλων συνόλων δεδομένων είναι εξίσου σημαντική, καθώς αυτά επηρεάζουν άμεσα την απόδοση και τη δυνατότητα γενίκευσης του μοντέλου. Στην ενότητα αυτή θα εξηγηθούν τόσο το task της ταξινόμησης εικόνων όσο και τα σύνολα δεδομένων που χρησιμοποιήθηκαν.

3.3.1 Ταξινόμηση Εικόνων ως Downstream Task

Η Ταξινόμηση Εικόνων (*Image Classification*) είναι ένα κλασικό πρόβλημα στην υπολογιστική όραση, το οποίο στοχεύει στην ανάθεση μιας εικόνας σε μία από πολλές προκαθορισμένες κατηγορίες. Το task αυτό απαιτεί την ικανότητα του μοντέλου να εξάγει αποδοτικά χαρακτηριστικά από την εικόνα και να τα συσχετίζει με τις αντίστοιχες κλάσεις. Η ταξινόμηση εικόνων λειτουργεί ως downstream task σε πολλά ερευνητικά πειράματα, επιτρέποντας την αξιολόγηση της γενίκευσης των χαρακτηριστικών που έχουν μάθει τα μοντέλα από προγενέστερα proxy tasks.

Στο πλαίσιο αυτό, το μοντέλο εκπαιδεύεται με στόχο να αναγνωρίσει υψηλού επιπέδου χαρακτηριστικά, όπως σχήματα, υφές και χρώματα, τα οποία είναι κρίσιμα για την ορθή κατηγοριοποίηση των εικόνων. Η προσέγγιση αυτή στηρίζεται στις αρχές της πολυεπίπεδης εκμάθησης (*hierarchical learning*), όπου τα πρώτα επίπεδα του νευρωνικού δικτύου εξάγουν απλά χαρακτηριστικά, όπως ακμές, ενώ τα επόμενα επίπεδα συνδυάζουν αυτά τα χαρακτηριστικά για να σχηματίσουν πιο σύνθετες αναπαραστάσεις.

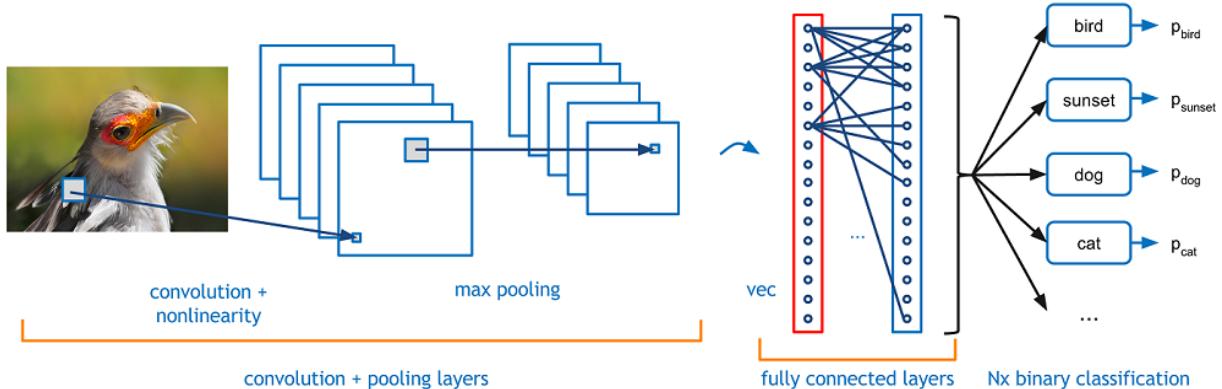
Η διαδικασία της εκπαίδευσης περιλαμβάνει 50 εποχές και το μέγεθος παρτίδας είναι 64. Επιπλέον, ως συνάρτηση απώλειας χρησιμοποιήθηκε η Cross-Entropy Loss και για τη βελτιστοποίηση χρησιμοποιήσαμε τον Adam Optimizer με learning rate ίσο με 0.001 και με weight decay ίσο με 0.0001 ως μία τεχνική κανονικοποίησης. Τέλος, χρησιμοποιήθηκε η τεχνική Dropout όπου κρίθηκε απαραίτητο για την αποφυγή της υπερπροσαρμογής.

Η απόδοση του μοντέλου στο task της ταξινόμησης εικόνων αξιολογείται μέσω μετρικών όπως η ακρίβεια (*accuracy*), η οποία αντικατοπτρίζει το ποσοστό των σωστών προβλέψεων σε σχέση με το σύνολο των εικόνων. Μια υψηλή ακρίβεια υποδηλώνει ότι το μοντέλο έχει μάθει να αναγνωρίζει σωστά τις κλάσεις, ακόμα και όταν οι εικόνες είναι οπτικά περίπλοκες ή περιέχουν θόρυβο. Στο συγκεκριμένο downstream task, η δυνατότητα του μοντέλου να εκμεταλλευτεί τα χαρακτηριστικά που έμαθε κατά την προεκπαίδευση μέσω άλλων έμμεσων διεργασιών είναι κρίσιμη για την τελική του απόδοση.

Τα νευρωνικά δίκτυα που χρησιμοποιούνται για ταξινόμηση εικόνων, όπως το ResNet, εφαρμόζουν συνήθως διαδοχικές συνελικτικές (*convolutional*) και πλήρως συνδεδεμένες (*fully connected*) στρώσεις. Τα τελικά επίπεδα του δικτύου εξάγουν μια αναπαράσταση της εικόνας σε έναν χώρο χαρακτηριστικών υψηλής διάστασης και εφαρμόζεται μια συνάρτηση ενεργοποίησης (π.χ. *softmax*) για την εκτίμηση της πιθανότητας ότι η εικόνα ανήκει σε κάθε κατηγορία. Η κατηγορία με την υψηλότερη πιθανότητα επιλέγεται ως η τελική πρόβλεψη του μοντέλου.

Η ταξινόμηση εικόνων ως downstream task επιτρέπει την αξιολόγηση του πώς τα προεκπαίδευμένα βάρη και τα χαρακτηριστικά που μαθαίνει το μοντέλο από proxy tasks μπορούν να συμβάλλουν στη γενίκευση και εφαρμογή αυτών των χαρακτηριστικών σε πιο συγκεκριμένες και απαιτητικές εργασίες υπολογιστικής

όρασης.



Εικόνα 3.3.1 - 1: Διαδικασία του task της ταξινόμησης εικόνων, όπου ένα συνελικτικό νευρωνικό δίκτυο (CNN) επεξεργάζεται μια εικόνα μέσω συνελικτικών και max pooling επιπέδων, για να εξάγει χαρακτηριστικά και να πραγματοποιήσει ταξινόμηση σε προκαθορισμένες κατηγορίες σύμφωνα με τις παραγόμενες πιθανότητες P .
(Πηγή: TowardsDataScience)

3.3.2 Σύνολα Δεδομένων που Χρησιμοποιήθηκαν

Στα πειράματα χρησιμοποιήθηκαν τα σύνολα δεδομένων CIFAR-10 και CIFAR-100, τα οποία είναι από τα πιο ευρέως χρησιμοποιούμενα datasets στον τομέα της υπολογιστικής οράσης. Αυτά τα σύνολα δεδομένων περιέχουν έγχρωμες εικόνες μικρών διαστάσεων (32x32 pixels), και έχουν ως στόχο την ταξινόμηση εικόνων σε πολλαπλές κατηγορίες.

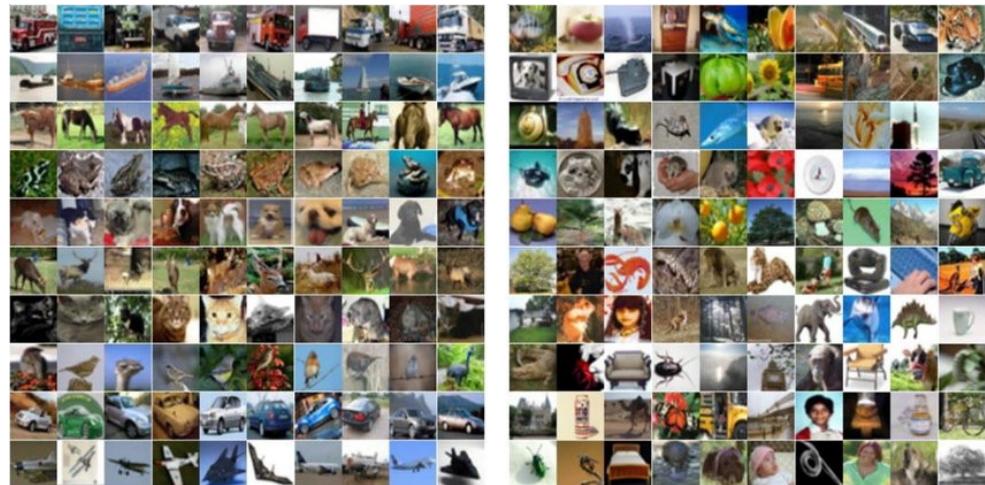
Το CIFAR-10 περιέχει συνολικά 60.000 εικόνες, κατανεμημένες σε 10 διαφορετικές κλάσεις. Κάθε κλάση περιέχει 6.000 εικόνες, εκ των οποίων 5.000 χρησιμοποιούνται για την εκπαίδευση και 1.000 για την αξιολόγηση του μοντέλου. Οι κλάσεις περιλαμβάνουν διάφορα αντικείμενα όπως αυτοκίνητα, γάτες, αεροπλάνα, σκύλους, και άλλα καθημερινά αντικείμενα και ζώα. Το CIFAR-10 χρησιμοποιείται συχνά σε tasks ταξινόμησης εικόνων, καθώς οι εικόνες του είναι απλές και ευρέως κατανοητές, ενώ η μικρή τους διάσταση επιτρέπει ταχύτερη εκπαίδευση.

Το CIFAR-100 είναι παρόμοιο με το CIFAR-10, αλλά περιέχει 100 κατηγορίες αντί για 10, καθιστώντας το task της ταξινόμησης πολύ πιο απαιτητικό. Κάθε κλάση περιλαμβάνει 600 εικόνες, με 500 για την εκπαίδευση και 100 για την αξιολόγηση. Οι κατηγορίες του CIFAR-100 είναι πιο εξειδικευμένες και περιλαμβάνουν υποκατηγορίες όπως έντομα, φάρια, έπιπλα, και άλλα καθημερινά αντικείμενα. Παρά τη μεγαλύτερη ποικιλία κατηγοριών, οι εικόνες του CIFAR-100 έχουν τις ίδιες διαστάσεις (32x32 pixels) και παρόμοια πολυπλοκότητα με το CIFAR-10.

Για το training των proxy tasks, χρησιμοποιήθηκε το CIFAR-100. Τα χαρακτηριστικά που έμαθε το μοντέλο κατά τη διάρκεια της εκπαίδευσης στα proxy tasks με αυτό το dataset μεταφέρθηκαν στο downstream task της ταξινόμησης εικόνων. Για την ταξινόμηση, τα pre-trained βάρη από το CIFAR-100 δοκιμάστηκαν τόσο στο ίδιο dataset (CIFAR-100) όσο και σε ένα διαφορετικό dataset, το CIFAR-10, προκειμένου να αξιολογηθεί η ικανότητα του μοντέλου να γενικεύει σε διαφορετικά σύνολα δεδομένων.

Τα CIFAR-10 και CIFAR-100 είναι συμβατά μεταξύ τους, καθώς έχουν παρόμοια χαρακτηριστικά. Και τα δύο datasets περιέχουν εικόνες με την ίδια ανάλυση (32x32) και έχουν ένα ευρύ φάσμα κατηγοριών αντικειμένων. Αυτή η συμβατότητα μεταξύ των συνόλων δεδομένων επιτρέπει τη δοκιμή των προεκπαιδευμένων βαρών από το CIFAR-100 στο CIFAR-10, παρέχοντας χρήσιμες πληροφορίες σχετικά με

το κατά πόσο τα χαρακτηριστικά που έμαθε το μοντέλο μπορούν να γενικευτούν στο task ταξινόμησης.



(a)

(b)

Εικόνα 3.3.2 - 1: (a) Εικόνες από το CIFAR-10 και (b) εικόνες από το CIFAR-100, παρουσιάζοντας παραδείγματα κατηγοριών που περιέχονται στα datasets. (Πηγή: ResearchGate)

3.4 Μεταφορά Μάθησης και Fine-Tuning

Η Μεταφορά Μάθησης (*Transfer learning*) και το Fine-Tuning αποτελούν κρίσιμες τεχνικές στη σύγχρονη μηχανική μάθηση, ιδιαίτερα σε περιπτώσεις όπου το σύνολο δεδομένων για το downstream task είναι περιορισμένο. Η βασική ιδέα της μεταφοράς μάθησης είναι η χρήση ενός προεκπαιδευμένου μοντέλου, το οποίο έχει μάθει χρήσιμα χαρακτηριστικά από ένα διαφορετικό task, με στόχο να βελτιώσει την απόδοση σε ένα νέο, αλλά συναφές task. Το fine-tuning συνίσταται στην περαιτέρω προσαρμογή των προεκπαιδευμένων βαρών του μοντέλου στο νέο task, είτε προσαρμόζοντας μόνο τα τελευταία επίπεδα του δικτύου είτε επανεκπαιδεύοντας ολόκληρο το μοντέλο. Στην παρούσα ενότητα, θα εξεταστούν οι στρατηγικές μεταφοράς μάθησης από τα proxy tasks στο downstream task της ταξινόμησης εικόνων, καθώς και οι τεχνικές fine-tuning που εφαρμόστηκαν για τη βελτιστοποίηση της απόδοσης του μοντέλου.

3.4.1 Μεταφορά Μάθησης από τα Proxy Tasks στο Downstream Task

Η διαδικασία της μεταφοράς μάθησης από τα proxy tasks στο downstream task βασίζεται στη χρήση των βέλτιστων pretrained weights που αποκτήθηκαν κατά την εκπαίδευση στις τρεις έμμεσες διεργασίες (proxy tasks): πρόβλεψη περιστροφής, χρωματοποίηση και επιδιόρθωση/συμπλήρωση εικόνας (inpainting), οι οποίες εκπαιδεύτηκαν στο σύνολο δεδομένων CIFAR-100. Είναι σημαντικό να σημειωθεί ότι το CIFAR-100 χρησιμοποιήθηκε ως ένα unlabeled dataset για την εκπαίδευση των proxy tasks, χωρίς να χρησιμοποιήθουν τα labels των εικόνων που υποδηλώνουν τις κατηγορίες στις οποίες ανήκουν. Κατά τη διάρκεια της εκπαίδευσης στα proxy tasks, παρακολουθήσαμε τη βελτίωση του validation accuracy και, κάθε φορά που το μοντέλο σημείωνε την υψηλότερη απόδοση στο validation set, αποθηκεύαμε το state του μοντέλου. Αυτό εξασφάλισε ότι διατηρήσαμε τα πιο αποδοτικά και γενικεύσιμα βάρη, απαραίτητα για τη μεταφορά στο downstream task.

Αφού ολοκληρώθηκε η εκπαίδευση στα proxy tasks, κατασκευάσαμε ένα πανομοιότυπο μοντέλο για το downstream task της ταξινόμησης εικόνων. Για να μπορέσουμε να χρησιμοποιήσουμε τα αποθηκευμένα βάρη, φορτώσαμε τα pretrained weights στο νέο μοντέλο και στη συνέχεια τροποποιήσαμε το τελευταίο πλήρως συνδεδεμένο επίπεδο (fully connected layer), έτσι ώστε να προσαρμοστεί στον αριθμό των κλάσεων που απαιτούσε το εκάστοτε dataset. Συγκεκριμένα, για το σύνολο δεδομένων CIFAR-10, το τελικό επίπεδο αναδιαμορφώθηκε ώστε να παράγει 10 εξόδους, όσες είναι οι κατηγορίες στο CIFAR-10, ενώ για το CIFAR-100 το τελικό επίπεδο τροποποιήθηκε ώστε να παράγει 100 εξόδους.

Η προσαρμογή αυτή του τελευταίου επιπέδου είναι σημαντική, καθώς επιτρέπει τη σωστή αντιστοίχιση του χώρου χαρακτηριστικών που έχει μάθει το μοντέλο με τον αριθμό των κλάσεων που απαιτούνται από το εκάστοτε dataset. Με τη διαδικασία αυτή, μπορέσαμε να αξιολογήσουμε την αποτελεσματικότητα των προεκπαιδευμένων βαρών που αποκτήθηκαν από το CIFAR-100 όχι μόνο στο ίδιο dataset αλλά και σε ένα διαφορετικό dataset, το CIFAR-10. Αυτό μας επιτρέπει να εξετάσουμε κατά πόσο τα χαρακτηριστικά που έμαθε το μοντέλο σε ένα σύνολο δεδομένων μπορούν να γενικευτούν και να βελτιώσουν την απόδοση σε ένα διαφορετικό σύνολο δεδομένων.

3.4.2 Fine-Tuning στο Downstream Task

Στη διαδικασία του fine-tuning στο downstream task της ταξινόμησης εικόνων, πραγματοποιήσαμε δύο διαφορετικές προσεγγίσεις για την εκπαίδευση του μοντέλου. Συγκεκριμένα, εκπαιδεύσαμε το μοντέλο σε κάθε σύνολο δεδομένων (CIFAR-10 και CIFAR-100) με δύο τρόπους:

1. Εκπαίδευση ολόκληρου του μοντέλου, προσαρμόζοντας τα βάρη όλων των επιπέδων (fine-tuning all layers).
2. Εκπαίδευση μόνο του τελευταίου επιπέδου του δικτύου (fine-tuning only the last layer), ενώ διατηρήσαμε τα υπόλοιπα επίπεδα "παγωμένα" (frozen).

Στην πρώτη περίπτωση, πραγματοποιήθηκε πλήρης εκπαίδευση του δικτύου σε κάθε task, επιτρέποντας στα βάρη όλων των επιπέδων να ενημερώνονται κατά τη διαδικασία της εκπαίδευσης. Στη δεύτερη περίπτωση, για να επανεκπαιδευτεί μόνο το τελευταίο επίπεδο, παγώσαμε (freeze) όλα τα επίπεδα του δικτύου και “ξεπαγώσαμε” (unfreeze) το τελευταίο επίπεδο, αφήνοντας ελεύθερο προς εκπαίδευση μόνο το τελευταίο πλήρως συνδεδεμένο επίπεδο. Με αυτόν τον τρόπο, μπορέσαμε να αξιολογήσουμε πώς τα προεκπαιδευμένα βάρη επηρεάζουν την απόδοση του μοντέλου, όταν γίνεται fine-tuning μόνο στο τελευταίο επίπεδο.

Αυτή η διαδικασία εφαρμόστηκε σε όλα τα proxy tasks (πρόβλεψη περιστροφής, χρωματοποίηση και επιδιόρθωση/συμπλήρωση εικόνας). Ο στόχος ήταν να αξιολογήσουμε την απόδοση του μοντέλου σε διαφορετικά σύνολα δεδομένων και με διαφορετικές στρατηγικές fine-tuning.

Για την προεπεξεργασία των δεδομένων, χρησιμοποιήσαμε τους εξής μετασχηματισμούς: αλλαγή μεγέθους των εικόνων σε 224x224 pixels, μετατροπή τους σε tensors και κανονικοποίηση (normalize) των pixel values με τις τιμές μέσου όρου και τυπικής απόκλισης που αντιστοιχούν στο CIFAR-10 και CIFAR-100. Η χρήση της κανονικοποίησης εξασφαλίζει ότι τα δεδομένα εισόδου έχουν σταθεροποιημένες τιμές, κάτι που βελτιώνει την εκπαίδευση του μοντέλου.

Οι υπερπαράμετροι της εκπαίδευσης ήταν κοινές για όλα τα tasks, τα μοντέλα των οποίων εκπαιδεύτηκαν για 50 εποχές με το batch size να ορίζεται σε 64. Επίσης, ο αλγόριθμος βελτιστοποίησης ήταν ο Adam, με learning rate 0.001 και weight decay 1e-4. Η εκπαίδευση πραγματοποιήθηκε με χρήση της συνάρτησης απώλειας CrossEntropyLoss, καθώς το downstream task αφορά ταξινόμηση πολλών κατηγοριών.

Με τη στρατηγική αυτή, μπορέσαμε να αξιολογήσουμε την απόδοση των προεκπαιδευμένων βαρών σε διαφορετικά σύνολα δεδομένων, τόσο με πλήρες fine-tuning όσο και με fine-tuning μόνο του τελευταίου επιπέδου. Η διαδικασία που περιγράφηκε, εφαρμόστηκε ακριβώς με τον ίδιο τρόπο για τα βάρη όλων των proxy tasks ώστε να διασφαλιστεί η δίκαιη σύγκριση μεταξύ των έμμεσων διεργασιών.

3.5 Αποτελέσματα

Σε αυτή την ενότητα θα παρουσιαστούν τα αποτελέσματα των πειραμάτων που πραγματοποιήθηκαν με σκοπό την αξιολόγηση των αποδόσεων των προεκπαιδευμένων βαρών στο task της ταξινόμησης εικόνων.

3.5.1 Αποτελέσματα 1ου πειράματος

Task: Image Classification

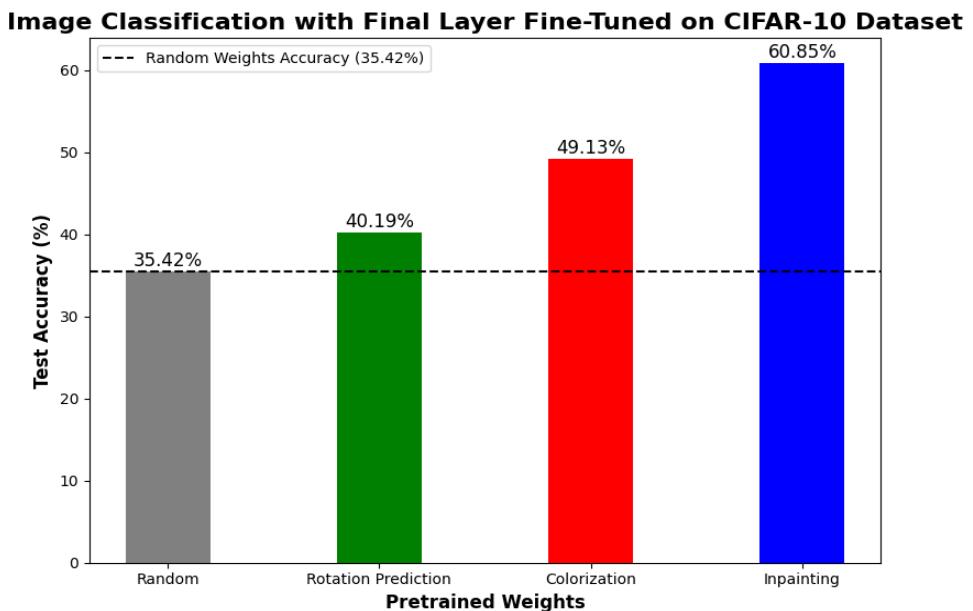
Fine-Tuning: Final Layer

Weights: Randomly Initialized, Image Rotation Prediction Pretraining, Image Colorization Pretraining, Image Inpainting Pretraining

Dataset: CIFAR-10

Accuracy Type: Test Accuracy

Weights	Test Accuracy
Randomly Initialized	35.42%
Image Rotation Prediction Pretraining	40.19%
Image Colorization Pretraining	49.13%
Image Inpainting Pretraining	60.85%



Πίνακας-Διάγραμμα 3.5.1 - 1: Αποτελέσματα ακρίβειας (Test Accuracy) στην ταξινόμηση εικόνων, επανεκπαιδεύοντας μόνο το τελευταίο στρώμα του μοντέλου στη συγκεκριμένη διεργασία, χρησιμοποιώντας τυχαία και προεκπαιδευμένα βάρη από τις τρεις έμμεσες διεργασίες (Image Rotation Prediction, Image Colorization, Image Inpainting) στο σύνολο δεδομένων CIFAR-10.

Αφού παρατέθηκαν οι τελικές τιμές Test Accuracy για το πρώτο πείραμα, στη συνέχεια θα παρουσιαστούν πιο συγκεκριμένες μετρήσεις για το ίδιο πείραμα, αναλύοντας τις τιμές του Validation Accuracy ανά εποχή. Αυτές οι μετρήσεις θα μας δώσουν μια πιο λεπτομερή εικόνα για την απόδοση του μοντέλου κατά τη διάρκεια της εκπαίδευσης:

Image Classification with Final Layer Fine-Tuned on CIFAR-10 Dataset

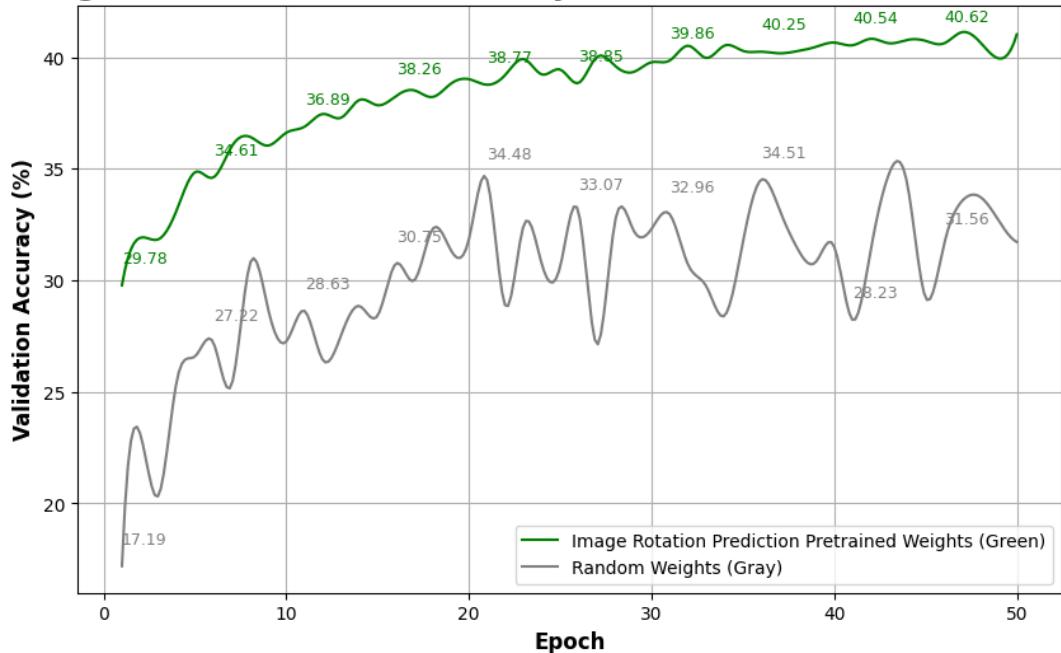


Image Classification with Final Layer Fine-Tuned on CIFAR-10 Dataset

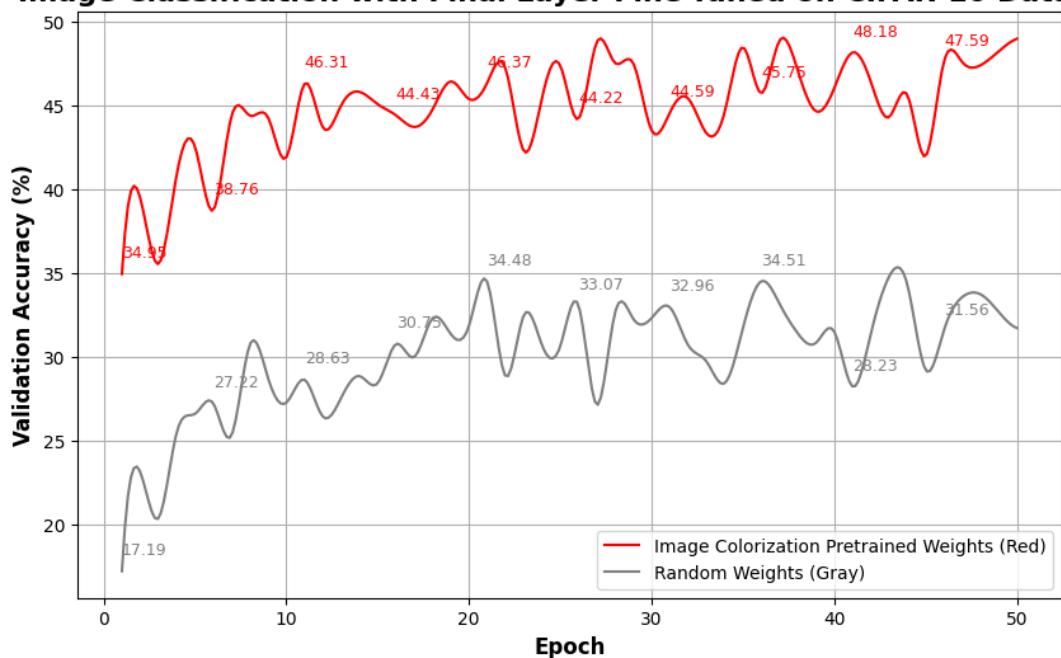
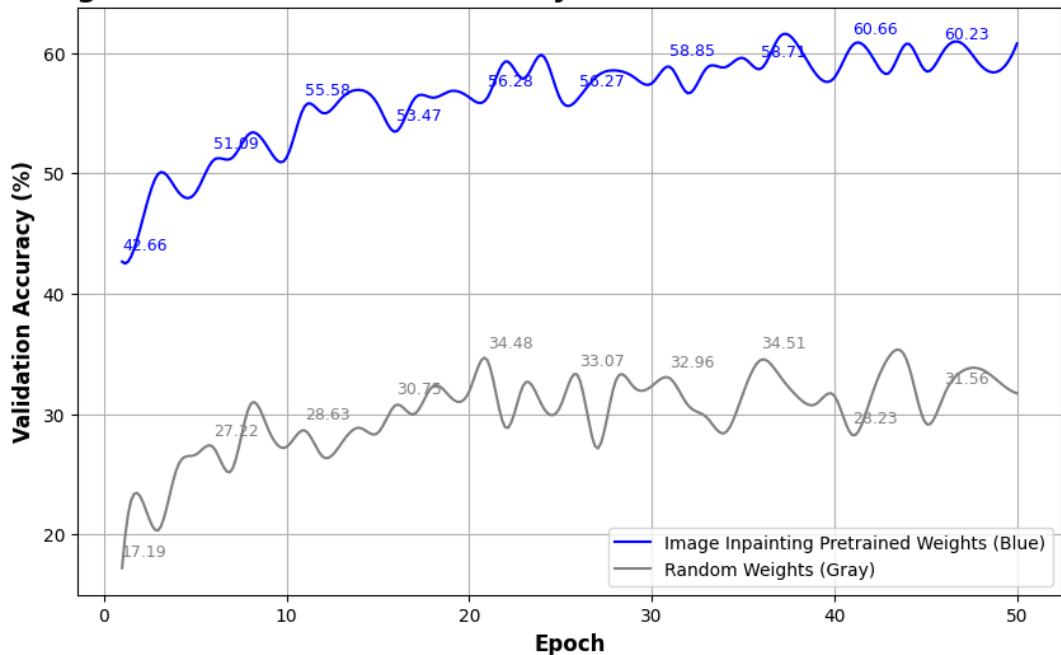


Image Classification with Final Layer Fine-Tuned on CIFAR-10 Dataset



Διαγράμματα 3.5.1 - 2,3,4: Αποτελέσματα ακρίβειας (στο validation set) ανά εποχή στην ταξινόμηση εικόνων, επανεκπαιδεύοντας μόνο το τελευταίο στρώμα του μοντέλου στη συγκεκριμένη διεργασία, χρησιμοποιώντας τυχαία βάρη ως σημείο αναφοράς και τα προεκπαιδευμένα βάρη από τις τρεις έμμεσες διεργασίες: Image Rotation Prediction, Image Colorization, Image Inpainting αντίστοιχα, στο σύνολο δεδομένων CIFAR-10.

3.5.2 Αποτελέσματα 2ου πειράματος

Task: Image Classification

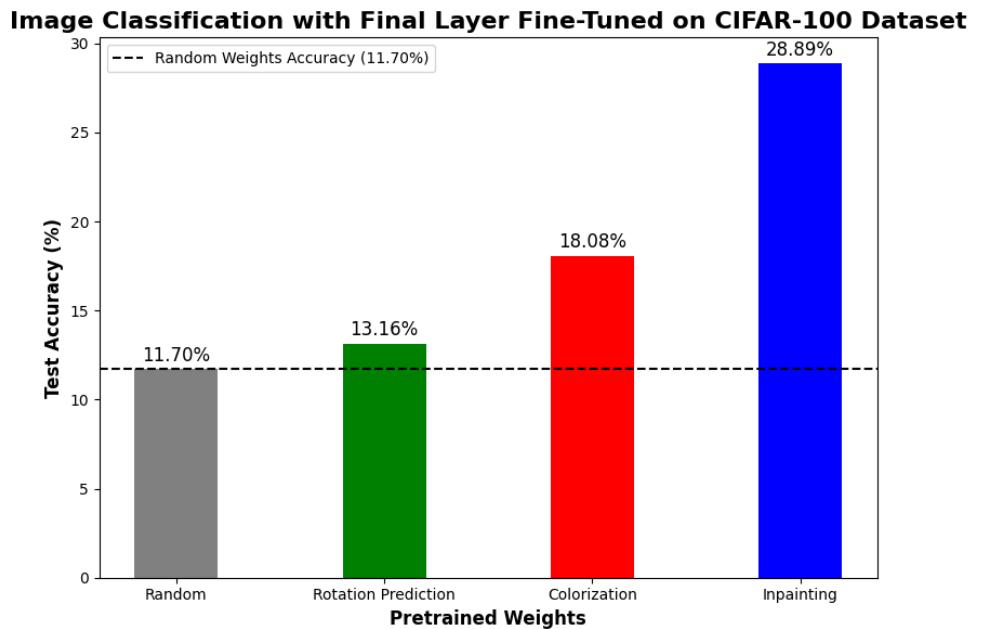
Fine-Tuning: Final Layer

Weights: Randomly Initialized, Image Rotation Prediction Pretraining, Image Colorization Pretraining, Image Inpainting Pretraining

Dataset: CIFAR-100

Accuracy Type: Test Accuracy

Weights	Test Accuracy
Randomly Initialized	11.70%
Image Rotation Prediction Pretraining	13.16%
Image Colorization Pretraining	18.08%
Image Inpainting Pretraining	28.89%



Πίνακας-Διάγραμμα 3.5.2 - 1: Αποτελέσματα ακρίβειας (Test Accuracy) στην ταξινόμηση εικόνων, επανεκπαιδεύοντας μόνο το τελευταίο στρώμα του μοντέλου στη συγκεκριμένη διεργασία, χρησιμοποιώντας τυχαία και προεκπαιδευμένα βάρη από τις τρεις έμμεσες διεργασίες (Image Rotation Prediction, Image Colorization, Image Inpainting) στο σύνολο δεδομένων CIFAR-100.

Αφού παρατέθηκαν οι τελικές τιμές Test Accuracy για το δεύτερο πείραμα, στη συνέχεια θα παρουσιαστούν πιο συγκεκριμένες μετρήσεις για το ίδιο πείραμα, αναλύοντας τις τιμές του Validation Accuracy ανά εποχή. Αυτές οι μετρήσεις θα μας δώσουν μια πιο λεπτομερή εικόνα για την απόδοση του μοντέλου κατά τη διάρκεια της εκπαίδευσης:

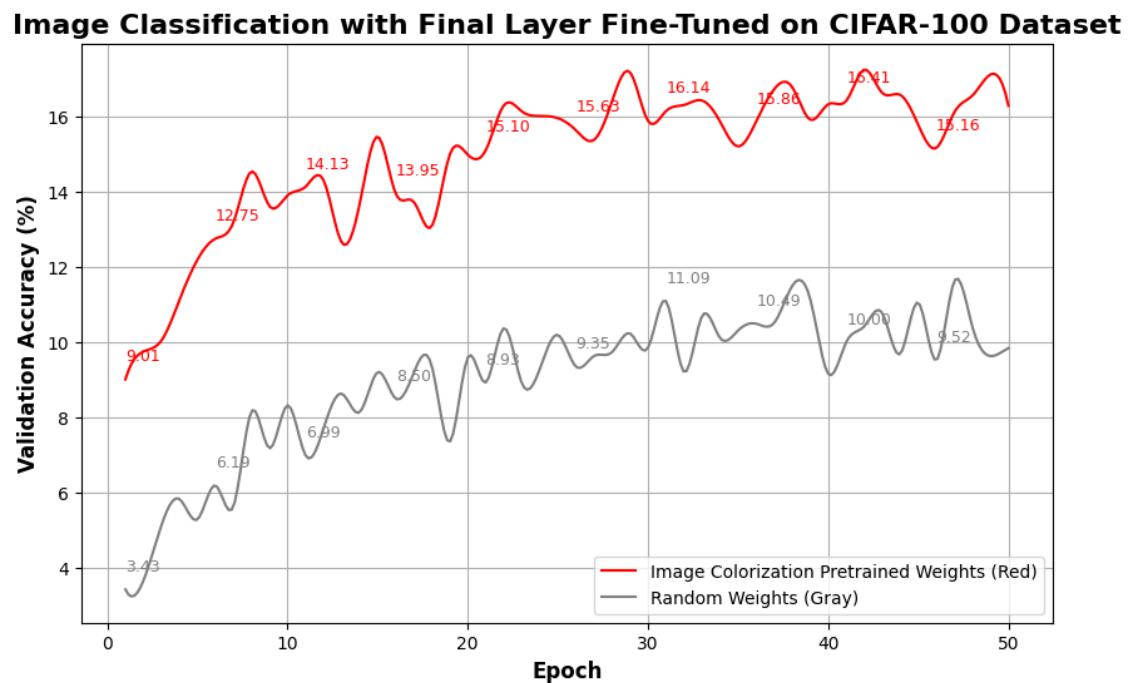
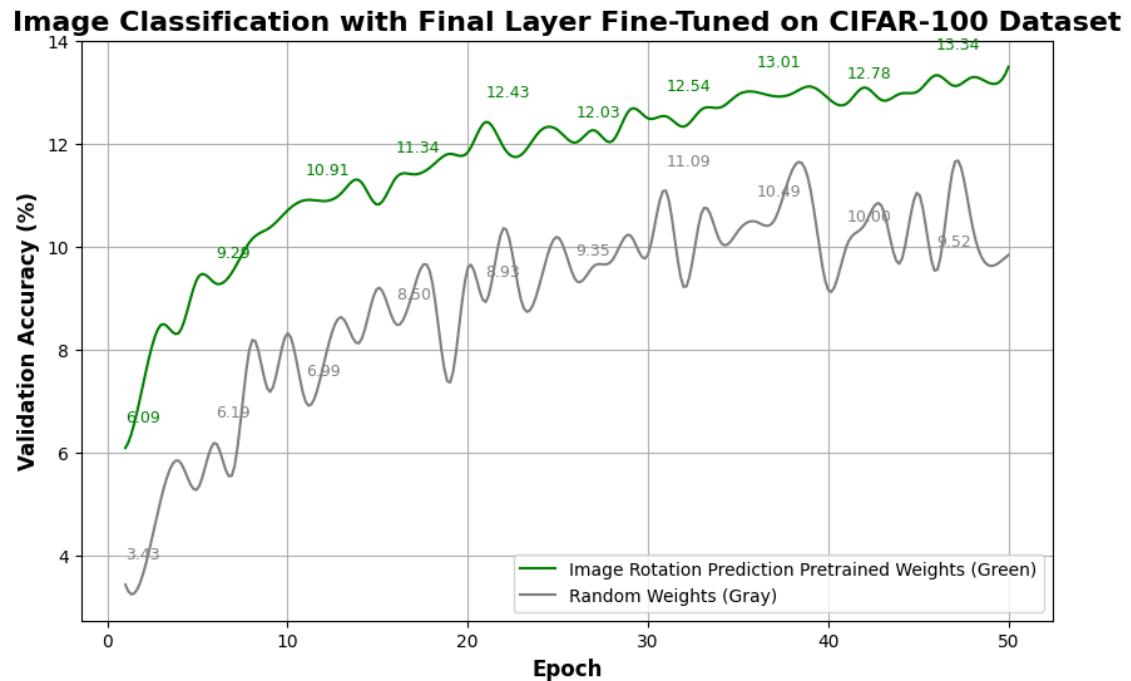
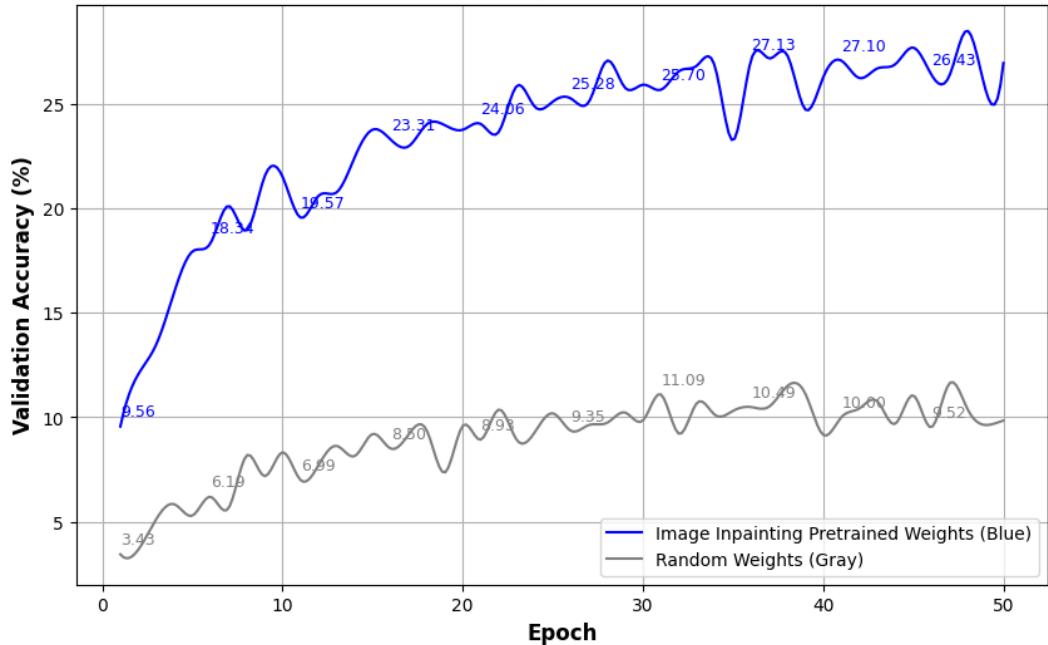


Image Classification with Final Layer Fine-Tuned on CIFAR-100 Dataset



Διαγράμματα 3.5.2 - 2,3,4: Αποτελέσματα ακρίβειας (στο validation set) ανά εποχή στην ταξινόμηση εικόνων, επανεκπαιδεύοντας μόνο το τελευταίο στρώμα του μοντέλου στη συγκεκριμένη διεργασία, χρησιμοποιώντας τυχαία βάρη ως σημείο αναφοράς και τα προεκπαιδευμένα βάρη από τις τρεις έμμεσες διεργασίες: Image Rotation Prediction, Image Colorization, Image Inpainting αντίστοιχα, στο σύνολο δεδομένων CIFAR-100.

3.5.3 Αποτελέσματα 3ου πειράματος

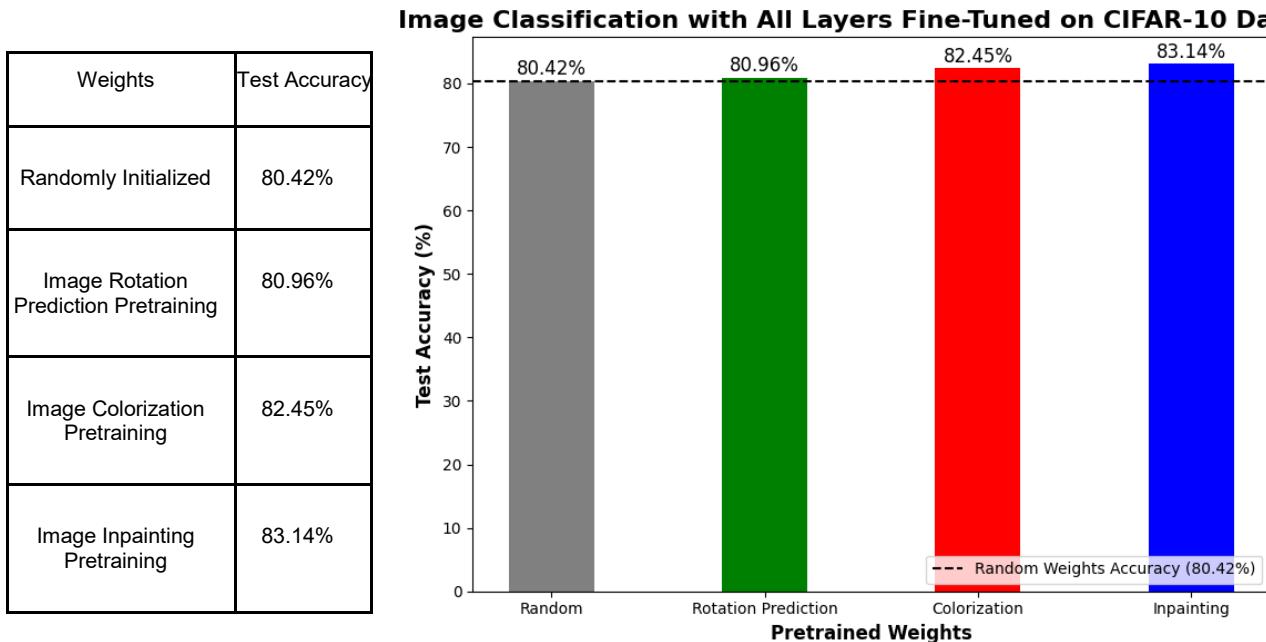
Task: Image Classification

Fine-Tuning: All Layers

Weights: Randomly Initialized, Image Rotation Prediction Pretraining, Image Colorization Pretraining, Image Inpainting Pretraining

Dataset: CIFAR-10

Accuracy Type: Test Accuracy



Πίνακας-Διάγραμμα 3.5.3 - 1: Αποτελέσματα ακρίβειας (Test Accuracy) στην ταξινόμηση εικόνων, επανεκπαιδεύοντας όλα τα στρώματα του μοντέλου στη συγκεκριμένη διεργασία, χρησιμοποιώντας τυχαία και προεκπαιδευμένα βάρη από τις τρεις έμμεσες διεργασίες (Image Rotation Prediction, Image Colorization, Image Inpainting) στο σύνολο δεδομένων CIFAR-10.

Αφού παρατέθηκαν οι τελικές τιμές Test Accuracy για το τρίτο πείραμα, στη συνέχεια θα παρουσιαστούν πιο συγκεκριμένες μετρήσεις για το ίδιο πείραμα, αναλύοντας τις τιμές του Validation Accuracy ανά εποχή. Αυτές οι μετρήσεις θα μας δώσουν μια πιο λεπτομερή εικόνα για την απόδοση του μοντέλου κατά τη διάρκεια της εκπαίδευσης:

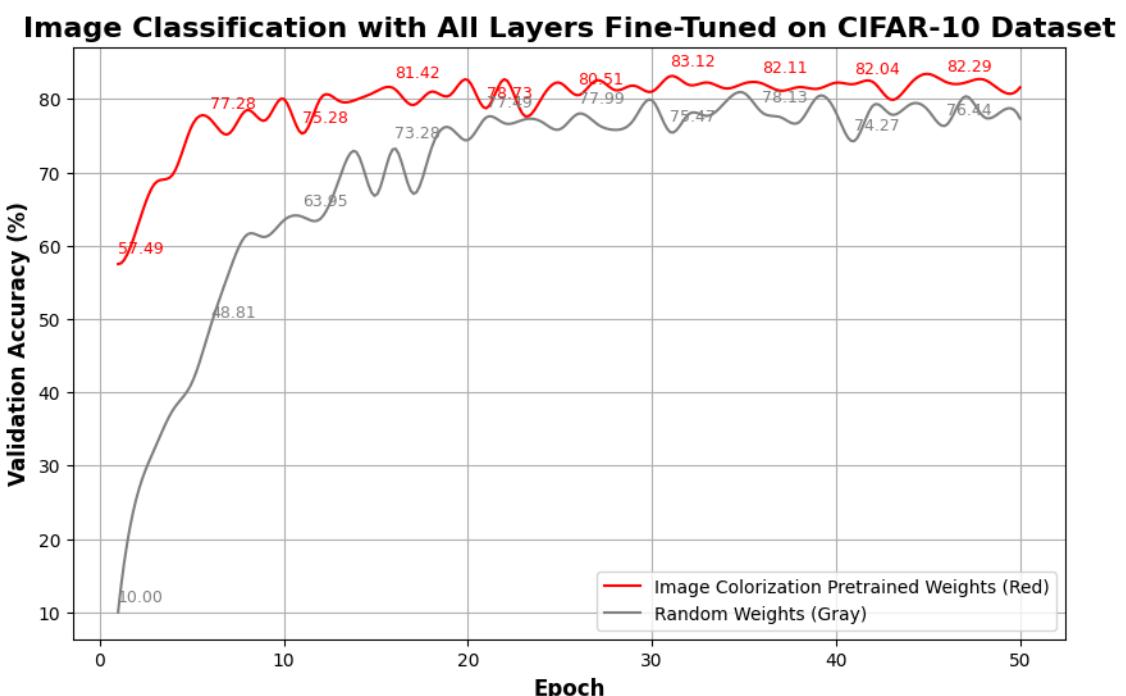
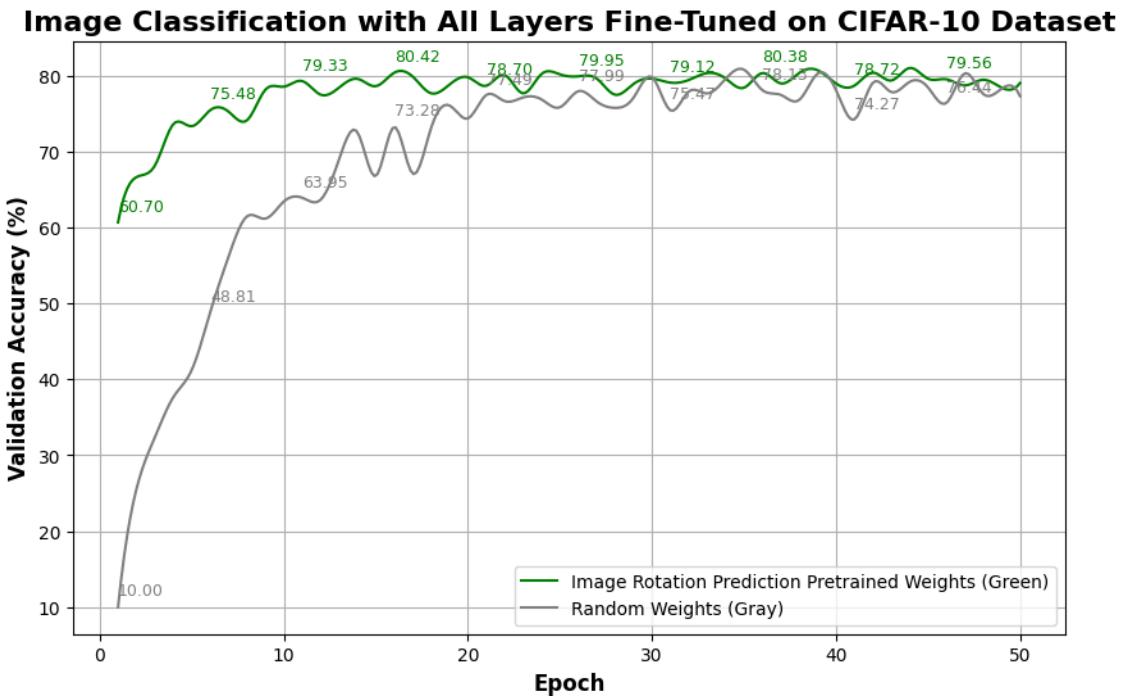
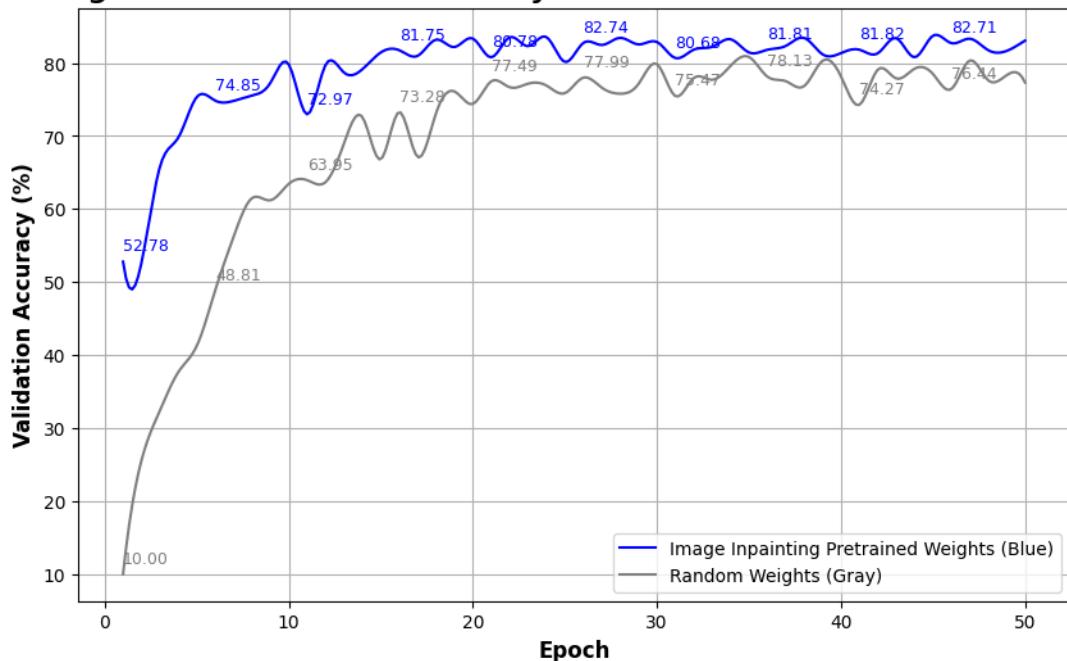


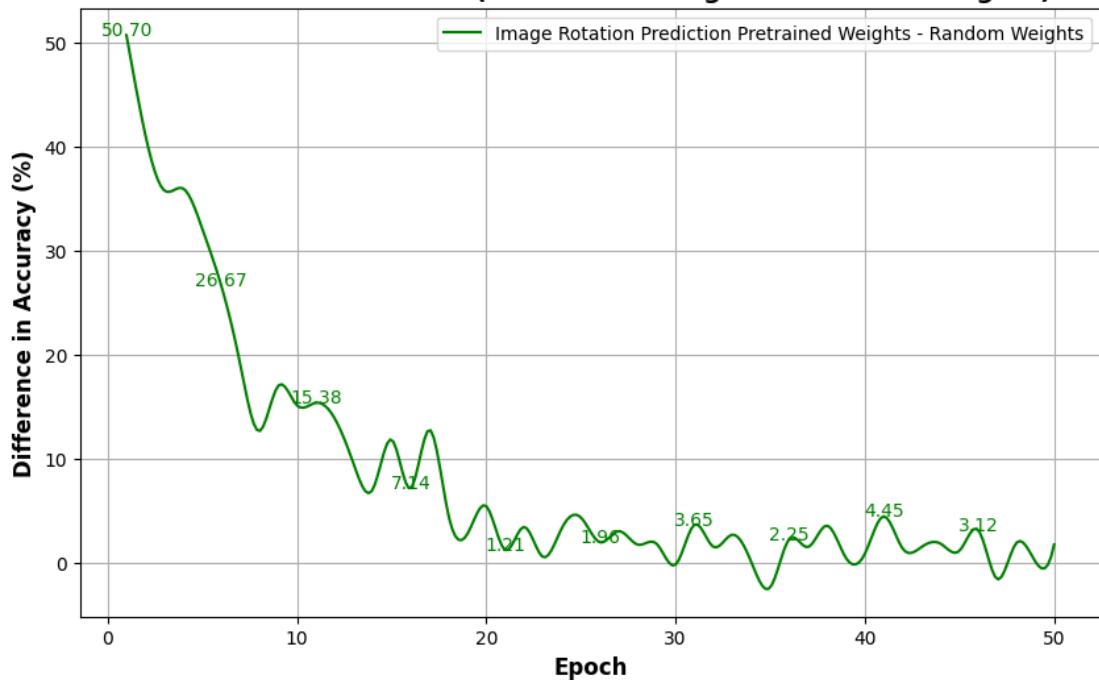
Image Classification with All Layers Fine-Tuned on CIFAR-10 Dataset



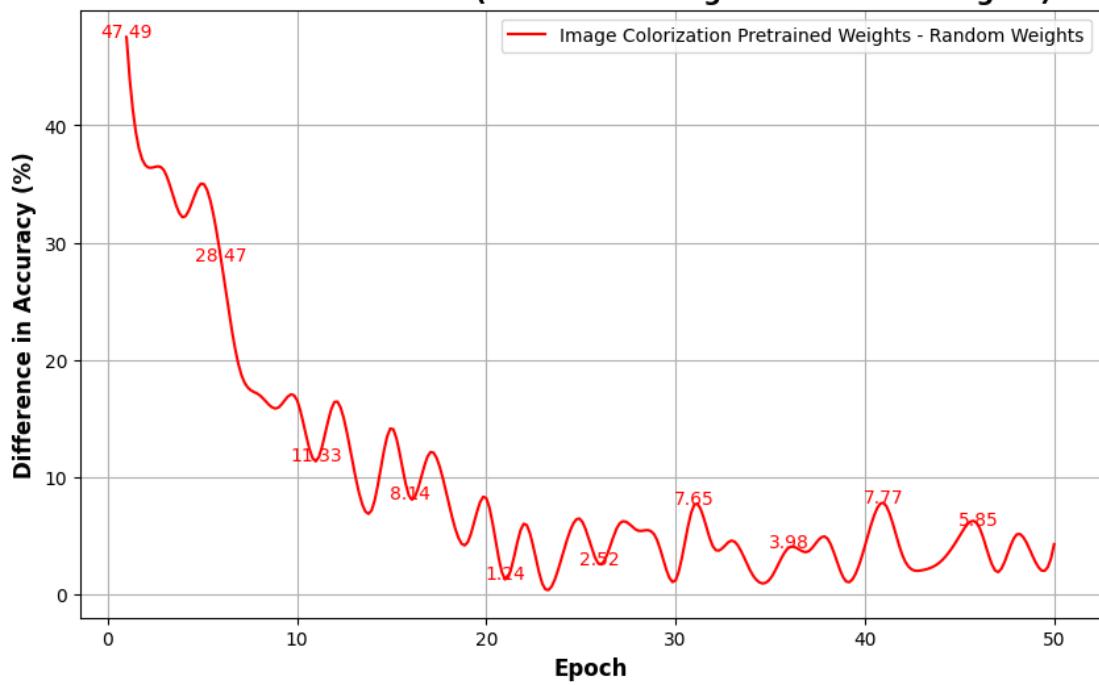
Διαγράμματα 3.5.3 - 2,3,4: Αποτελέσματα ακρίβειας (στο validation set) ανά εποχή στην ταξινόμηση εικόνων, επανεκπαιδεύοντας όλα τα στρώματα του μοντέλου στη συγκεκριμένη διεργασία, χρησιμοποιώντας τυχαία βάρη ως σημείο αναφοράς και τα προεκπαιδευμένα βάρη από τις τρεις έμμεσες διεργασίες: Image Rotation Prediction, Image Colorization, Image Inpainting αντίστοιχα ως αρχικά βάρη, στο σύνολο δεδομένων CIFAR-10.

Επιπλέον, στο συγκεκριμένο όπως και στο επόμενο πείραμα παρουσιάζονται τα διαγράμματα που δείχνουν τη διαφορά της ακρίβειας ανάμεσα στις τιμές που μετρήθηκαν για το validation accuracy του μοντέλου με τα προεκπαιδευμένα βάρη από την κάθε έμμεση διεργασία και τις τιμές που μετρήθηκαν για το validation accuracy του μοντέλου με random weights. Αυτή η σύγκριση θα μας βοηθήσει να κατανοήσουμε καλύτερα την επίδραση των προεκπαιδευμένων βαρών σε σχέση με τα τυχαία βάρη και να εξάγουμε πιο ολοκληρωμένα συμπεράσματα για την απόδοση του μοντέλου ανά τις εποχές, δηλαδή κατά τη διάρκεια της εκπαίδευσης:

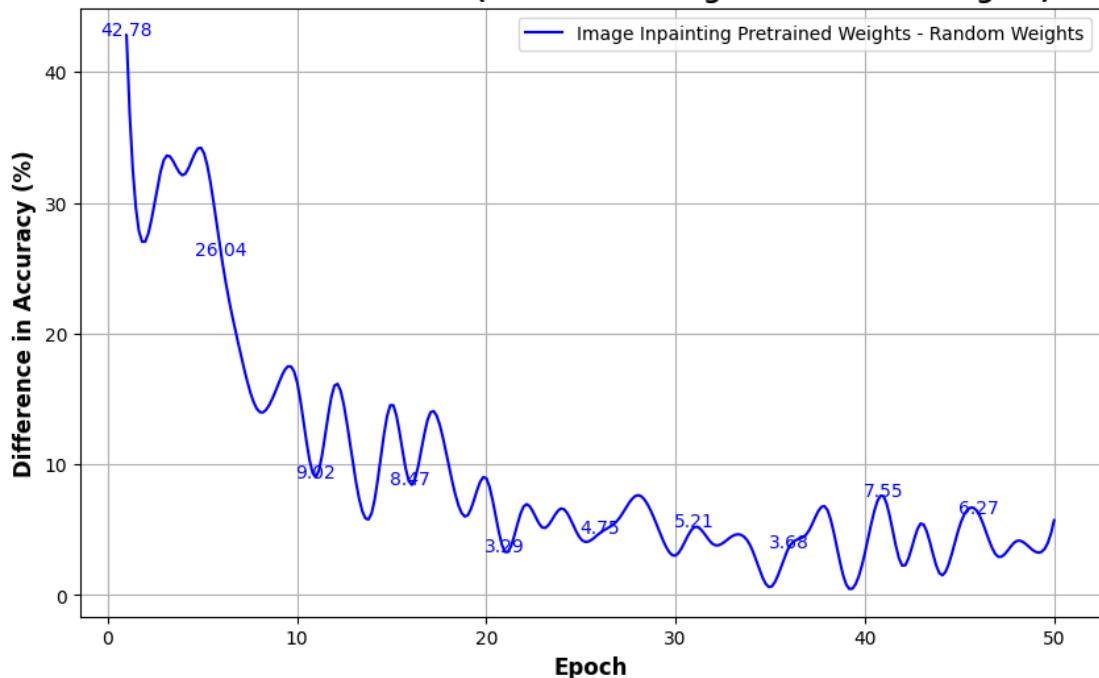
Difference in Validation Accuracy of Image Classification with All Layers Fine-Tuned on CIFAR-10 (Pretrained Weights - Random Weights)



Difference in Validation Accuracy of Image Classification with All Layers Fine-Tuned on CIFAR-10 (Pretrained Weights - Random Weights)



Difference in Validation Accuracy of Image Classification with All Layers Fine-Tuned on CIFAR-10 (Pretrained Weights - Random Weights)



Διαγράμματα 3.5.3 - 5,6,7: Διαφορά του Validation Accuracy στην ταξινόμηση εικόνων, επανεκπαιδεύοντας όλα τα στρώματα στο συγκεκριμένο task, ανάμεσα στο μοντέλο αρχικοποιημένο με τα προεκπαιδευμένα βάρη και στο μοντέλο αρχικοποιημένο με τυχαία βάρη για κάθε έμμεση διεργασία (Image Rotation Prediction, Image Colorization, Image Inpainting) αντίστοιχα στο σύνολο δεδομένων CIFAR-10.

3.5.4 Αποτελέσματα 4ου πειράματος

Task: Image Classification

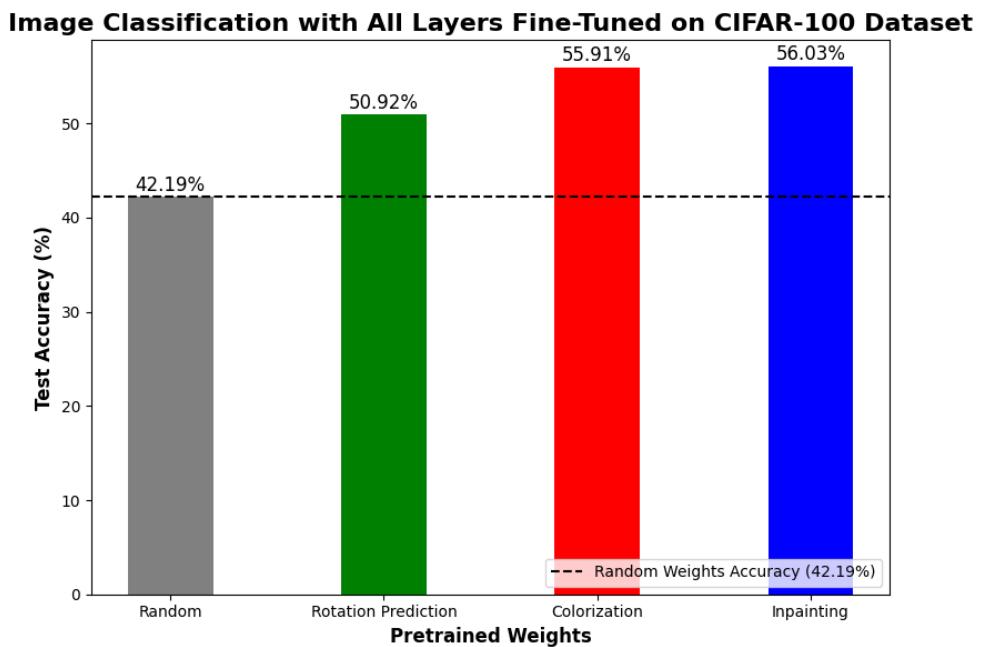
Fine-Tuning: All Layers

Weights: Randomly Initialized, Image Rotation Prediction Pretraining, Image Colorization Pretraining, Image Inpainting Pretraining

Dataset: CIFAR-100

Accuracy Type: Test Accuracy

Weights	Test Accuracy
Randomly Initialized	42.19%
Image Rotation Prediction Pretraining	50.92%
Image Colorization Pretraining	55.91%
Image Inpainting Pretraining	56.03%



Πίνακας-Διάγραμμα 3.5.4 - 1: Αποτελέσματα ακρίβειας (στο test set) στην ταξινόμηση εικόνων, επανεκπαιδεύοντας όλα τα στρώματα του μοντέλου στη συγκεκριμένη διεργασία, χρησιμοποιώντας τυχαία και προεκπαιδευμένα βάρη από τις τρεις έμμεσες διεργασίες (Image Rotation Prediction, Image Colorization, Image Inpainting) στο σύνολο δεδομένων CIFAR-100.

Αφού παρατέθηκαν οι τελικές τιμές Test Accuracy για το τέταρτο πείραμα, στη συνέχεια θα παρουσιαστούν πιο συγκεκριμένες μετρήσεις για το ίδιο πείραμα, αναλύοντας τις τιμές του Validation Accuracy ανά εποχή. Αυτές οι μετρήσεις θα μας δώσουν μια πιο λεπτομερή εικόνα για την απόδοση του μοντέλου κατά τη διάρκεια της εκπαίδευσης:

Image Classification with All Layers Fine-Tuned on CIFAR-100 Dataset

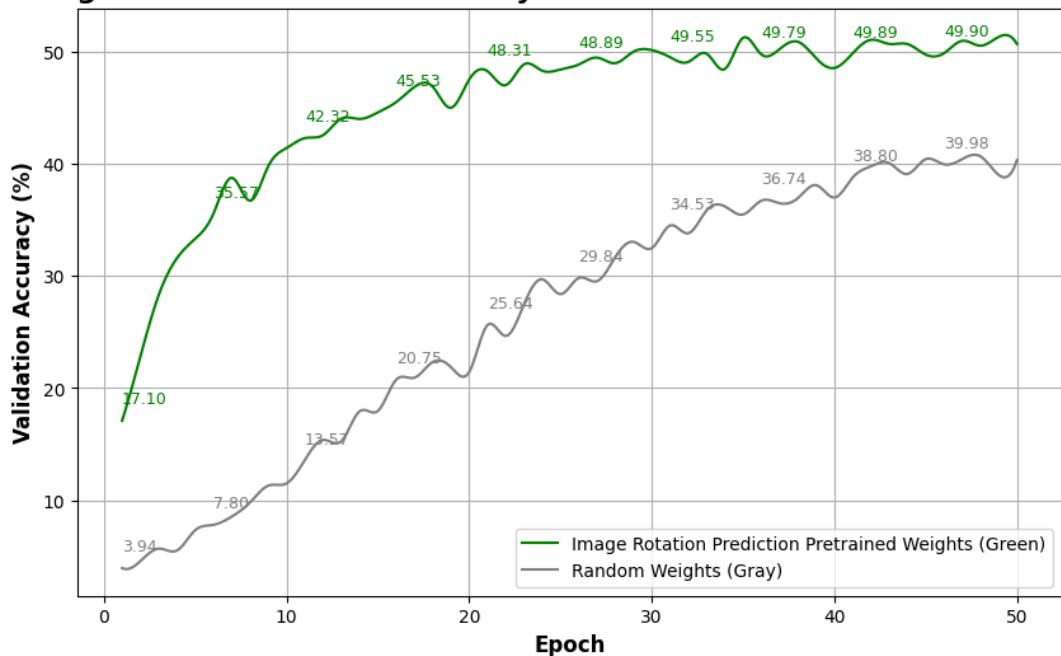


Image Classification with All Layers Fine-Tuned on CIFAR-100 Dataset

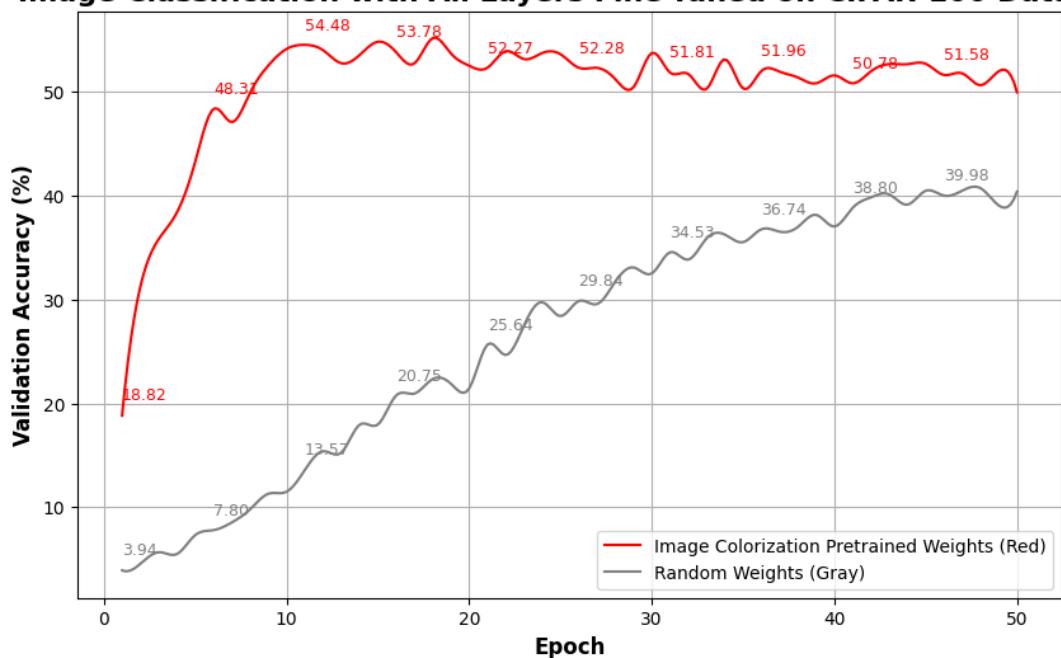
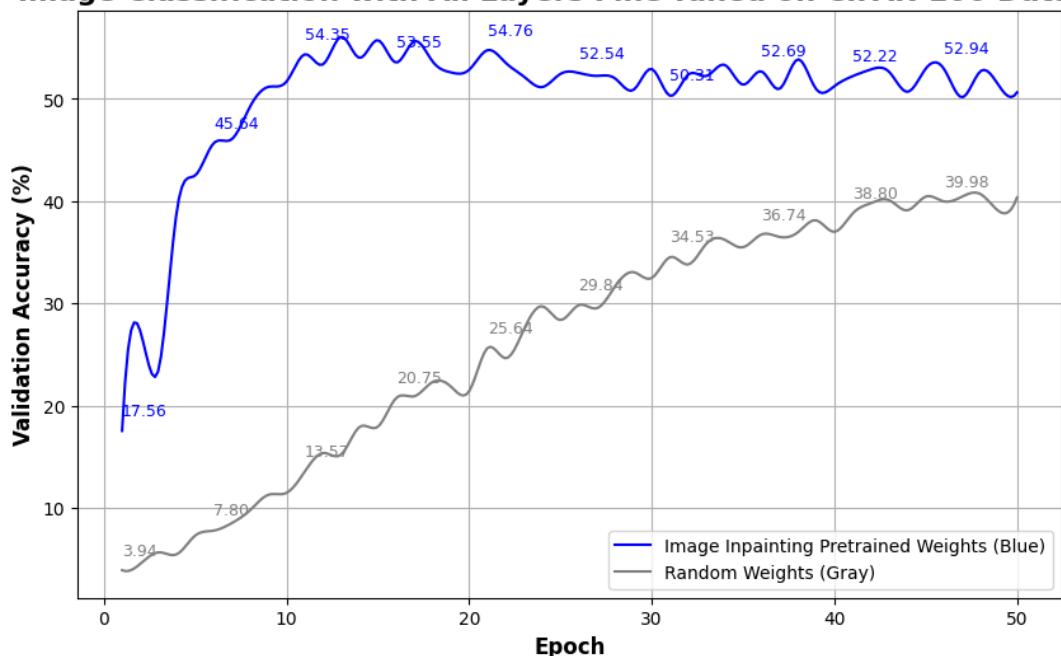


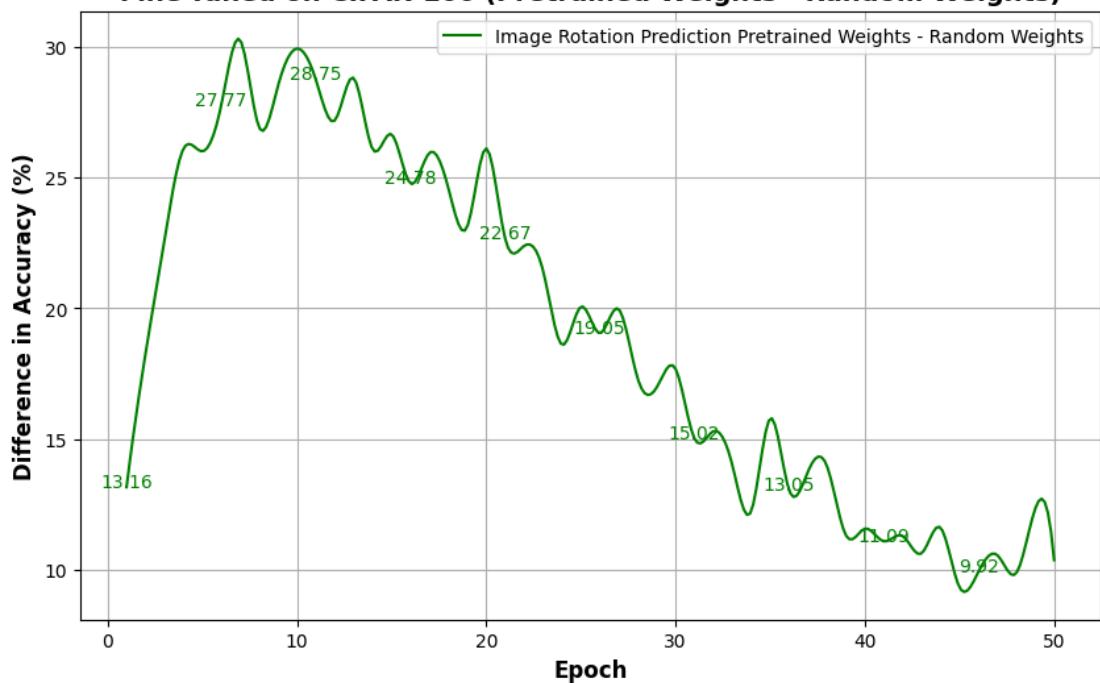
Image Classification with All Layers Fine-Tuned on CIFAR-100 Dataset



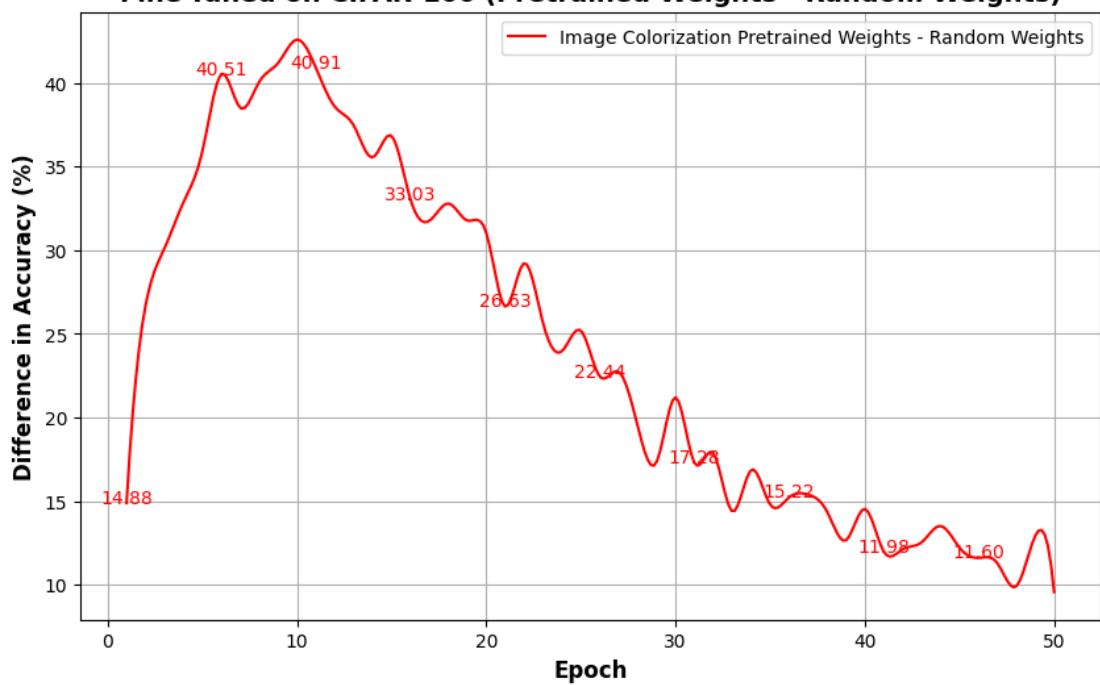
Διαγράμματα 3.5.4 - 2,3,4: Αποτελέσματα ακρίβειας (στο validation set) ανά εποχή στην ταξινόμηση εικόνων, επανεκπαιδεύοντας όλα τα στρώματα του μοντέλου στη συγκεκριμένη διεργασία, χρησιμοποιώντας τυχαία βάρη ως σημείο αναφοράς και τα προεκπαιδευμένα βάρη από τις τρεις έμμεσες διεργασίες: Image Rotation Prediction, Image Colorization, Image Inpainting αντίστοιχα ως αρχικά βάρη, στο σύνολο δεδομένων CIFAR-100.

Επιπλέον, παρουσιάζονται τα διαγράμματα που δείχνουν τη διαφορά της ακρίβειας ανάμεσα στις τιμές που μετρήθηκαν για το validation accuracy του μοντέλου με τα προεκπαιδευμένα βάρη από την κάθε έμμεση διεργασία και τις τιμές που μετρήθηκαν για το validation accuracy του μοντέλου με random weights. Αυτή η σύγκριση θα μας βοηθήσει να κατανοήσουμε καλύτερα την επίδραση των προεκπαιδευμένων βαρών σε σχέση με τα τυχαία βάρη και να εξάγουμε πιο ολοκληρωμένα συμπεράσματα για την απόδοση του μοντέλου ανά τις εποχές, δηλαδή κατά τη διάρκεια της εκπαίδευσης:

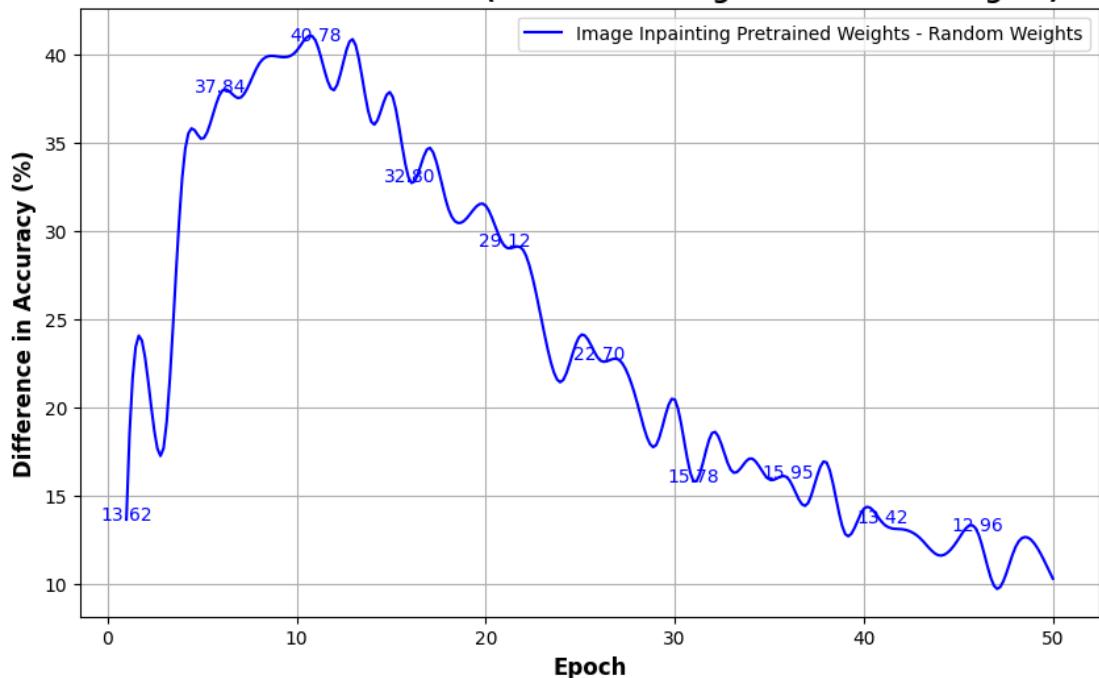
Difference in Validation Accuracy of Image Classification with All Layers Fine-Tuned on CIFAR-100 (Pretrained Weights - Random Weights)



Difference in Validation Accuracy of Image Classification with All Layers Fine-Tuned on CIFAR-100 (Pretrained Weights - Random Weights)



Difference in Validation Accuracy of Image Classification with All Layers Fine-Tuned on CIFAR-100 (Pretrained Weights - Random Weights)

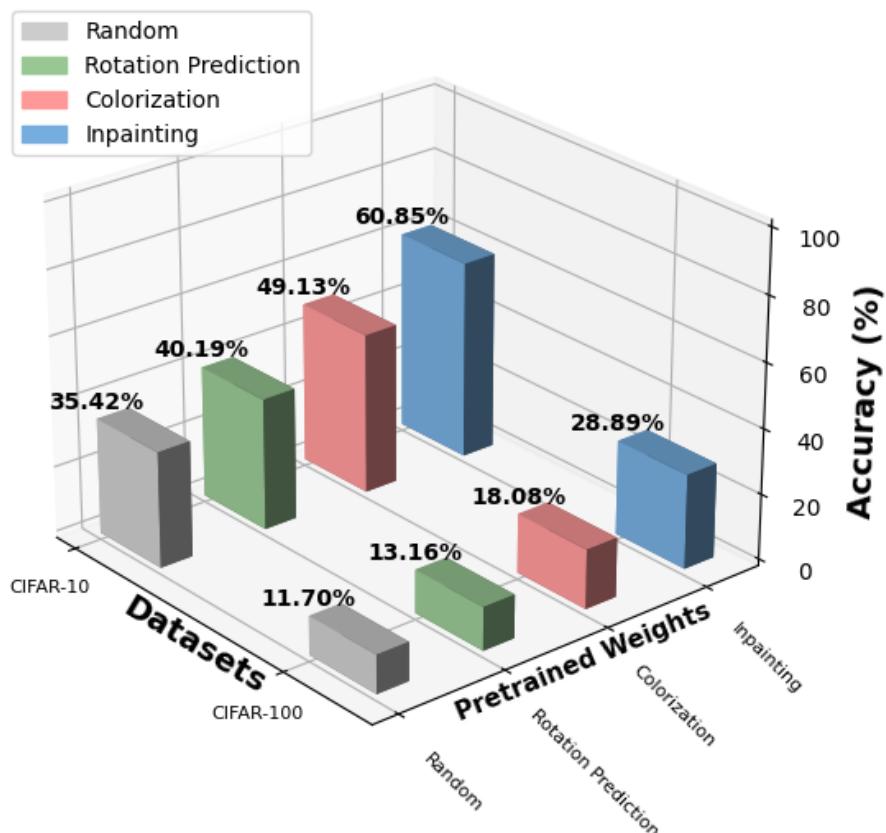


Διαγράμματα 3.5.4 - 5,6,7: Διαφορά του Validation Accuracy στην ταξινόμηση εικόνων, επανεκπαιδεύοντας όλα τα στρώματα στο συγκεκριμένο task, ανάμεσα στο μοντέλο αρχικοποιημένο με τα προεκπαιδευμένα βάρη και στο μοντέλο αρχικοποιημένο με τυχαία βάρη για κάθε έμμεση διεργασία (Image Rotation Prediction, Image Colorization, Image Inpainting) αντίστοιχα στο σύνολο δεδομένων CIFAR-100.

3.5.5 Συγκεντρωτική Απεικόνιση Αποτελεσμάτων

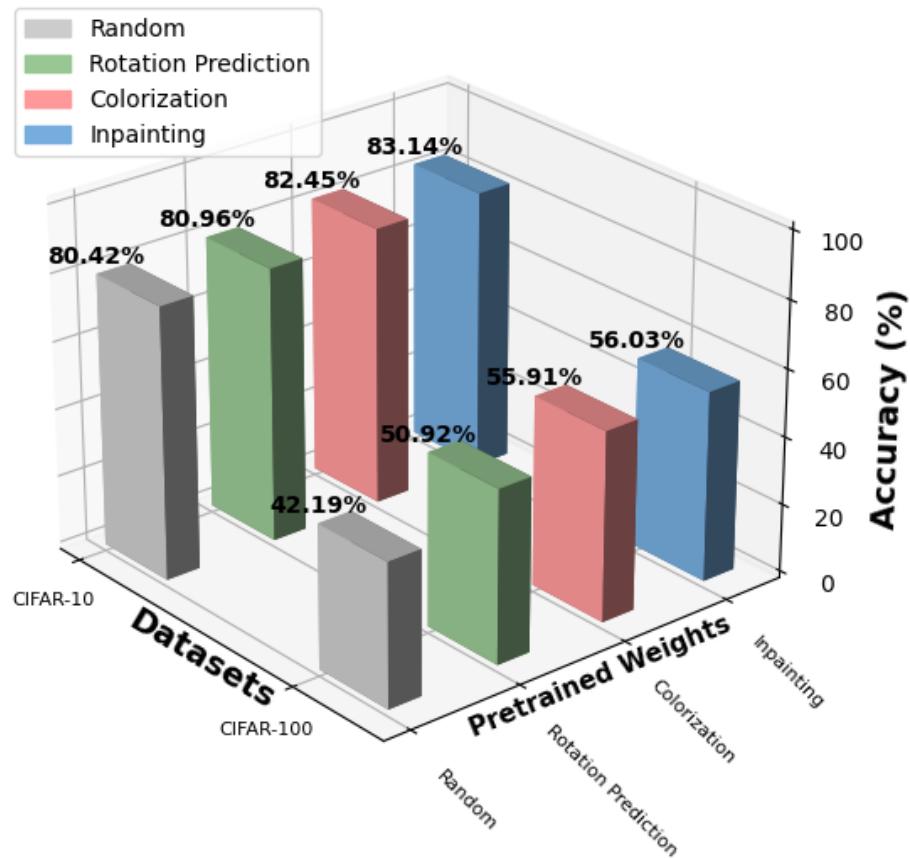
Τέλος, θα συνοψίσουμε όλες τις μετρήσεις στα παρακάτω διαγράμματα, ώστε να διευκολύνουμε την κατανόηση των διαφορών ανάμεσα στα tasks και να παρέχουμε μια πιο ολοκληρωμένη εικόνα της απόδοσης του μοντέλου. Με τη συγκεντρωτική απεικόνιση των αποτελεσμάτων, θα καταστεί ευκολότερη η εξαγωγή συμπερασμάτων σχετικά με την επίδραση των προεκπαϊδευμένων βαρών της κάθε έμμεσης διεργασίας, καθώς και η σύγκρισή τους με τα τυχαία βάρη.

Image Classification with Final Layer Fine-Tuned



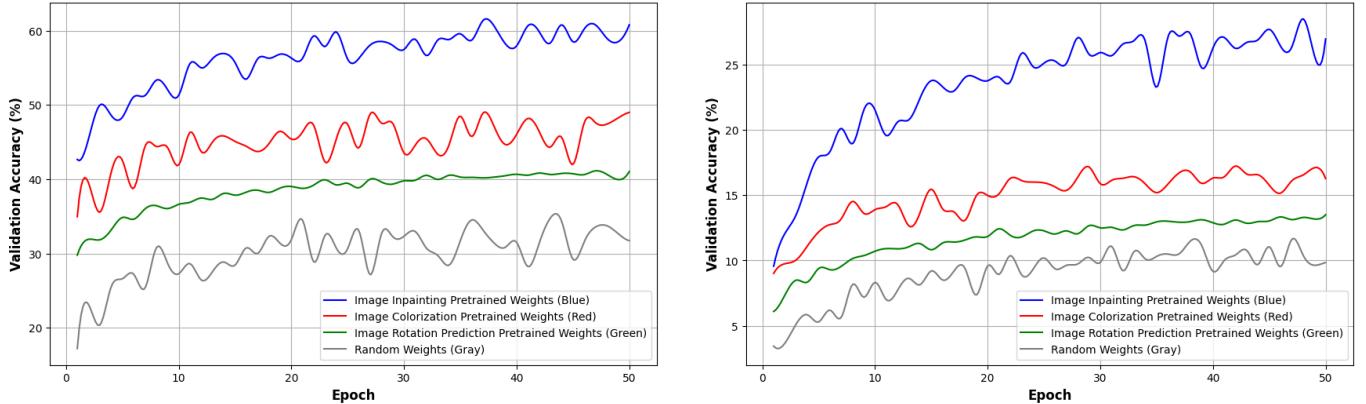
Διαγράμματα 3.5.5 - 1: Αποτελέσματα ακρίβειας (Test Accuracy) στην ταξινόμηση εικόνων, επανεκπαϊδεύοντας μόνο το τελευταίο στρώμα του μοντέλου στη συγκεκριμένη διεργασία, χρησιμοποιώντας τυχαία και προεκπαϊδευμένα βάρη από τις τρεις έμμεσες διεργασίες (Image Rotation Prediction, Image Colorization, Image Inpainting) στα σύνολα δεδομένων CIFAR-10 και CIFAR-100.

Image Classification with All Layers Fine-Tuned



Διάγραμμα 3.5.5 - 2: Αποτελέσματα ακρίβειας (Test Accuracy) στην ταξινόμηση εικόνων, επανεκπαιδεύοντας όλα τα στρώματα του μοντέλου στη συγκεκριμένη διεργασία, χρησιμοποιώντας τυχαία και προεκπαίδευμένα βάρη από τις τρεις έμμεσες διεργασίες (Image Rotation Prediction, Image Colorization, Image Inpainting) στα σύνολα δεδομένων CIFAR-10 και CIFAR-100.

Image Classification with Final Layer Fine-Tuned on CIFAR-10 Dataset **Image Classification with Final Layer Fine-Tuned on CIFAR-100 Dataset**



Διαγράμματα 3.5.5 – 3.4: Αποτελέσματα ακρίβειας (στο validation set) ανά εποχή στην ταξινόμηση εικόνων, επανεκπαιδεύοντας μόνο το τελευταίο στρώμα του μοντέλου στη συγκεκριμένη διεργασία, χρησιμοποιώντας τυχαία βάρη ως σημείο αναφοράς και τα προεκπαιδευμένα βάρη από τις τρεις έμμεσες διεργασίες: Image Rotation Prediction, Image Colorization, Image Inpainting ως αρχικά βάρη, στα σύνολα δεδομένων CIFAR-10 και CIFAR-100.

Image Classification with All Layers Fine-Tuned on CIFAR-10 Dataset

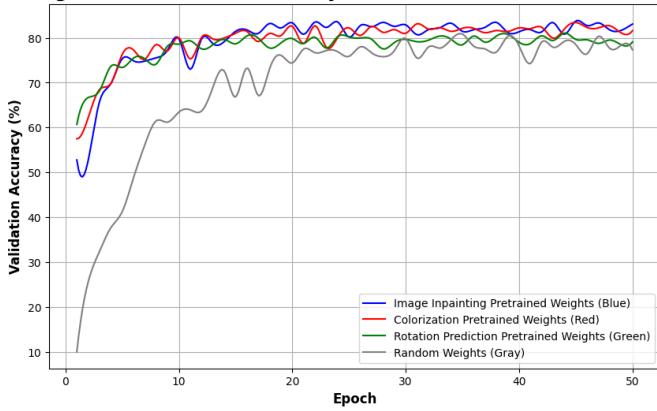
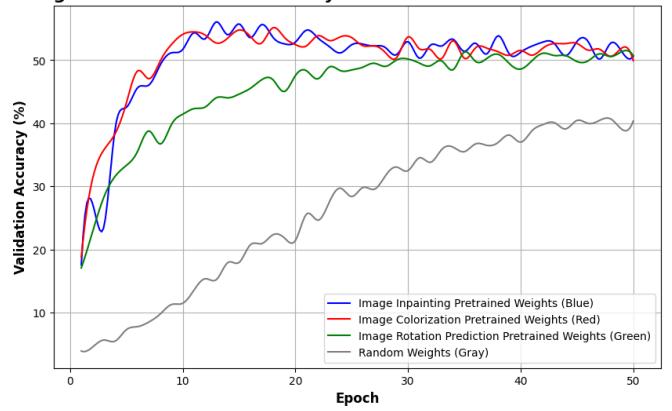


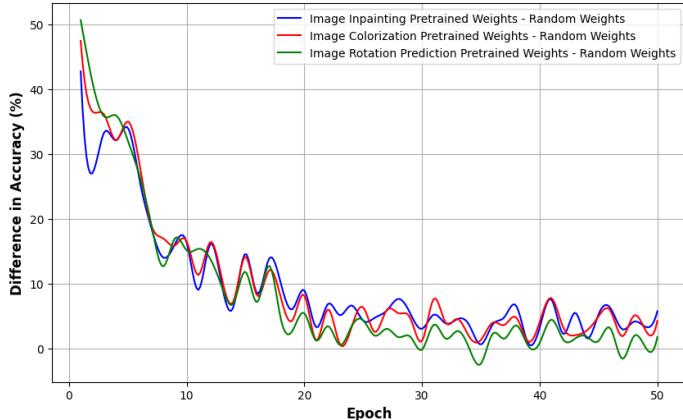
Image Classification with All Layers Fine-Tuned on CIFAR-100 Dataset



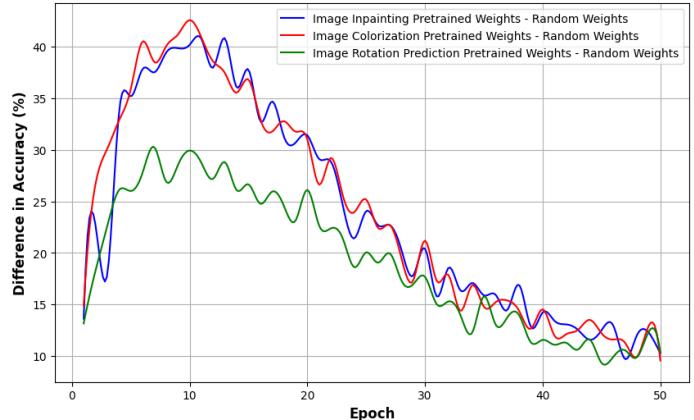
Δ

Διαγράμματα 3.5.5 – 5.6: Αποτελέσματα ακρίβειας (στο validation set) ανά εποχή στην ταξινόμηση εικόνων, επανεκπαιδεύοντας όλα τα στρώματα του μοντέλου στη συγκεκριμένη διεργασία, χρησιμοποιώντας τυχαία βάρη ως σημείο αναφοράς και τα προεκπαιδευμένα βάρη από τις τρεις έμμεσες διεργασίες: Image Rotation Prediction, Image Colorization, Image Inpainting ως αρχικά βάρη, στα σύνολα δεδομένων CIFAR-10 και CIFAR-100.

Difference in Validation Accuracy of Image Classification with All Layers Fine-Tuned on CIFAR-10 (Pretrained Weights - Random Weights)



Difference in Validation Accuracy of Image Classification with All Layers Fine-Tuned on CIFAR-100 (Pretrained Weights - Random Weights)



Διαγράμματα 3.5.5 - 7,8: Διαφορά του Validation Accuracy στην ταξινόμηση εικόνων, επανεκπαιδεύοντας όλα τα στρώματα στο συγκεκριμένο task, ανάμεσα στο μοντέλο αρχικοποιημένο με τα προεκπαιδευμένα βάρη και στο μοντέλο αρχικοποιημένο με τυχαία βάρη για τις τρεις έμμεσες διεργασίες (Image Rotation Prediction, Image Colorization, Image Inpainting) στα σύνολα δεδομένων CIFAR-10 και CIFAR-100.

3.6 Σχολιασμός Αποτελεσμάτων - Συμπεράσματα

Επιδόσεις με τη χρήση Έμμεσων Διεργασιών για την προ-εκπαίδευση των βαρών σε σχέση με τα τυχαία βάρη

Σύμφωνα με τις μετρήσεις και όπως φαίνεται στα διαγράμματα που παρατίθενται στην προηγούμενη ενότητα, οι επιδόσεις όλων των proxy tasks ήταν σταθερά (σε όλα τα πειράματα) καλύτερες από εκείνες που επιτεύχθηκαν με τη χρήση random weights. Αυτό επιβεβαιώνει την αρχική μας προσδοκία ότι η SSL προσφέρει σημαντικό πλεονέκτημα, καθώς τα προεκπαιδευμένα βάρη έχουν ήδη μάθει βασικά χαρακτηριστικά από τα proxy tasks. Σε αντίθεση, τα τυχαία βάρη δεν φέρουν καμία αρχική πληροφορία και απαιτούν περισσότερο χρόνο και δεδομένα για να μάθουν ουσιαστικά χαρακτηριστικά. Επομένως, τα proxy tasks προσέφεραν σημαντική βελτίωση στις επιδόσεις σε όλες τις μετρήσεις, επιβεβαιώνοντας τον σκοπό της αυτο-εποπτευόμενης μάθησης.

Επιδόσεις στα Διαφορετικά Datasets (CIFAR-10, CIFAR-100)

Σύμφωνα με τα αποτελέσματα, οι επιδόσεις στο CIFAR-100 ήταν χαμηλότερες σε σύγκριση με το CIFAR-10, κάτι που επιβεβαιώνει την αρχική μας υπόθεση ότι το task ταξινόμησης στο CIFAR-100 είναι πιο πολύπλοκο λόγω του μεγαλύτερου αριθμού κλάσεων (100 έναντι 10). Αυτό είναι εμφανές τόσο στην περίπτωση του fine-tuning μόνο του τελευταίου επιπέδου όσο και στην περίπτωση του fine-tuning ολόκληρου του μοντέλου, όπου οι επιδόσεις στο CIFAR-100 παρέμειναν χαμηλότερες.

Ωστόσο, τα pretrained weights που εκπαιδεύτηκαν στο CIFAR-100 φαίνεται να μεταφέρουν ικανοποιητικά τις γνώσεις τους στο CIFAR-10. Παρά την διαφορετικότητα των συγκεκριμένων συνόλων δεδομένων, τα προεκπαιδευμένα βάρη που προέκυψαν από το CIFAR-100 συνέβαλαν στην επίτευξη καλών επιδόσεων στο CIFAR-10, όπως αναμενόταν. Παρά το γεγονός ότι τα proxy tasks εκπαιδεύτηκαν σε 90 επιπλέον κλάσεις που δεν υπάρχουν στο CIFAR-10 και θα μπορούσαν να θεωρηθούν άχρηστες, το μοντέλο κατάφερε να μάθει σημαντικές πληροφορίες που ήταν χρήσιμες και στο CIFAR-10, επιτυγχάνοντας ικανοποιητικές αποδόσεις.

Επιπλέον, αυτό μας δείχνει ότι τα proxy tasks μαθαίνουν βασικά χαρακτηριστικά που σχετίζονται με τον προσανατολισμό, το χρώμα και τη συνοχή της εικόνας (για τα τρία proxy tasks αντίστοιχα) και όχι με βάση την κλάση των εικόνων. Αυτό επιτρέπει στα μοντέλα να μεταφέρουν χρήσιμες πληροφορίες ανεξαρτήτως των κλάσεων στις οποίες ανήκουν οι εικόνες. Αυτό συμβαίνει και επειδή χρησιμοποιήσαμε το CIFAR-100 ως unlabeled dataset για την εκπαίδευση στις έμμεσες διεργασίες. Επομένως, ο ισχυρισμός ότι το μοντέλο έμαθε από τα labels, ακόμα και αν δεν τα χρησιμοποιήσαμε άμεσα, δεν ευσταθεί διότι τα αποτελέσματα επιβεβαιώνουν ότι το μοντέλο μαθαίνει βάσει χαρακτηριστικών όπως ο προσανατολισμός, το χρώμα και η συνοχή της εικόνας, και όχι βάσει των κλάσεων των εικόνων.

Συμπερασματικά, οι επιδόσεις ήταν χαμηλότερες στο CIFAR-100 λόγω της πολυπλοκότητας του task, αλλά τα proxy tasks απέδειξαν ότι μπορούν να μεταφέρουν χρήσιμη πληροφορία στο CIFAR-10, διατηρώντας ικανοποιητικές αποδόσεις.

Σύγκριση πληροφορίας ανάμεσα στα προ-εκπαίδευμένα βάρη από τις έμμεσες διεργασίες και τα τυχαία βάρη, με fine-tuning μόνο στο τελευταίο επίπεδο του μοντέλου

Στο CIFAR-10, παρατηρούμε μια αύξηση απόδοσης σε σχέση με την αρχική επίδοση των τυχαίων βαρών:

- 13.47% για το rotation (από 35.42% σε 40.19%),
- 38.71% για το colorization (από 35.42% σε 49.13%),
- 71.80% για το inpainting (από 35.42% σε 60.85%).

Αντίστοιχα, στο CIFAR-100 παρατηρούμε μια αύξηση απόδοσης σε σχέση με τα τυχαία βάρη:

- 12.48% για το rotation (από 11.70% σε 13.16%),
- 54.53% για το colorization (από 11.70% σε 18.08%),
- 146.92% για το inpainting (από 11.70% σε 28.89%).

Επομένως, με αυτήν την άμεση σύγκριση των επιδόσεων, μπορούμε εύκολα να αντιληφθούμε ότι η πληροφορία που έμαθαν τα βάρη στα proxy tasks είναι ιδιαίτερα χρήσιμη και βοηθά σημαντικά στο task της ταξινόμησης εικόνων.

Σύγκριση τελικών επιδόσεων ανάμεσα στα προ-εκπαιδευμένα βάρη από τις έμμεσες διεργασίες και τα τυχαία βάρη, με fine-tuning σε όλα τα επίπεδα του μοντέλου

Στο CIFAR-10, παρατηρούμε αύξηση σε σχέση με την αρχική απόδοση των τυχαίων βαρών:

- 0.67% για το rotation (από 80.42% σε 80.96%),
- 2.52% για το colorization (από 80.42% σε 82.45%),
- 3.38% για το inpainting (από 80.42% σε 83.14%).

Στο CIFAR-100, παρατηρούμε αύξηση:

- 20.69% για το rotation (από 42.19% σε 50.92%),
- 32.52% για το colorization (από 42.19% σε 55.91%),
- 32.80% για το inpainting (από 42.19% σε 56.03%).

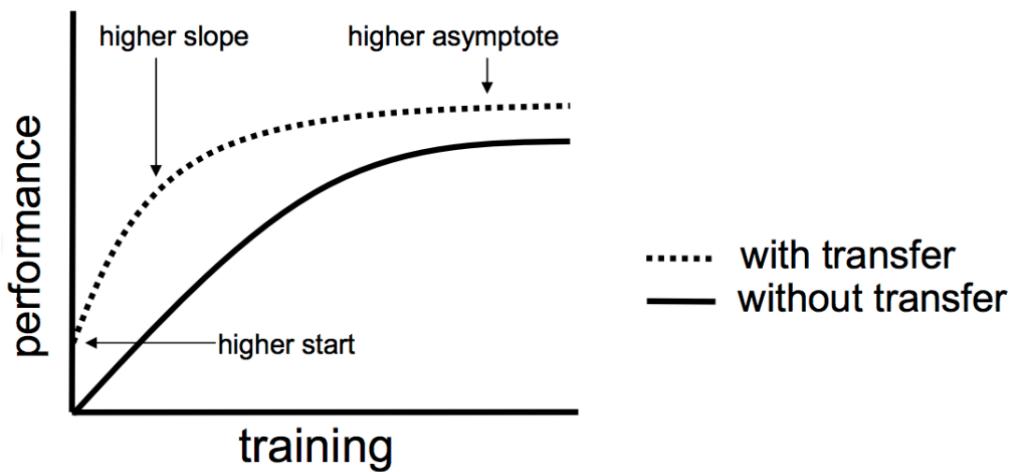
Σύμφωνα με τα αποτελέσματα, βλέπουμε ότι με το fine-tuning σε όλα τα επίπεδα του μοντέλου, οι τελικές επιδόσεις των proxy tasks είναι ελάχιστα καλύτερες από αυτές των random weights. Αυτό συμβαίνει διότι τα βάρη, είτε είναι τυχαία είτε προεκπαιδευμένα, χρησιμοποιούνται μόνο για αρχικοποίηση και στη συνέχεια αλλάζουν κατά τη διάρκεια της εκπαίδευσης. Παρ' όλα αυτά, τα pre-trained weights βοηθούν το μοντέλο να συγκλίνει ελαφρώς καλύτερα.

Ταχύτητα επίτευξης ικανοποιητικής επίδοσης με fine-tuning σε όλα τα επίπεδα του μοντέλου

Σύμφωνα με τα αποτελέσματα, είναι εμφανές στα διαγράμματα που απεικονίζουν την ακρίβεια ανά εποχή ότι η ταχύτητα με την οποία τα proxy tasks επιτρέπουν στο μοντέλο να φτάσει σε υψηλά επίπεδα απόδοσης είναι πολύ μεγαλύτερη σε σχέση με τα τυχαία βάρη. Συγκεκριμένα, στο CIFAR-10, με τα τρία proxy tasks (rotation prediction, colorization, inpainting) έχουμε επιτύχει accuracy 75% στις πρώτες 5 εποχές (epochs) και με τα τρία pretrained weights, ενώ φτάνουμε στο 80% στις πρώτες 10 εποχές. Μετά από εκεί, το καθένα συγκλίνει αντίστοιχα στις τελικές τιμές που αναφέραμε παραπάνω. Αντίθετα, με τα random weights, στην 5η εποχή το accuracy είναι περίπου 40-45% και στις 10 εποχές φτάνει περίπου στο 61%.

Αντίστοιχα, στο CIFAR-100, παρατηρούμε ότι με το rotation πετυχαίνουμε 41% στα πρώτα 10 εποχές, ενώ με το colorization και το inpainting η μέγιστη απόδοση φτάνει στο 55% ήδη από τις πρώτες 10 εποχές. Τα random weights, ωστόσο, παραμένουν στο 11% κατά τις πρώτες 10 εποχές.

Οι καμπύλες που βλέπουμε στα διαγράμματα ταιριάζουν απόλυτα με την παρακάτω εικόνα, η οποία παρουσιάζεται στο paper "An analysis of transfer learning for domain mismatched text-independent speaker verification" από τους Chunlei Zhang, Shivesh Ranjan, John H.L. Hansen [54]. Η εικόνα δείχνει ότι, γενικότερα στο transfer learning, στην αρχή της εκπαίδευσης έχουμε μια υψηλότερη αρχική τιμή (higher start), στις πρώτες εποχές παρατηρείται μια μεγαλύτερη κλίση (higher slope), και προς τα τέλη της εκπαίδευσης παρατηρείται μια υψηλότερη ασύμπτωτη τιμή (higher asymptote). Αυτή η ανάλυση επιβεβαιώνει τη λογική των μετρήσεών μας, η οποία αποτυπώνεται ξεκάθαρα στα διαγράμματα, δείχνοντας ότι τα pretrained weights προσφέρουν ταχύτερη και υψηλότερη επίδοση σε σχέση με τα τυχαία βάρη.



Εικόνα 3.6 - 1: Αναπαράσταση της διαφοράς μεταξύ transfer learning και εκπαίδευσης από τυχαία αρχικοποίηση.

(Πηγή: Zhang, C., Ranjan, S., & Hansen, J. H. (2018, June). An Analysis of Transfer Learning for Domain Mismatched Text-independent Speaker Verification. In *Odyssey* (pp. 181-186).)

Αυτή η διαφορά γίνεται ακόμη πιο εμφανής στα διαγράμματα που δείχνουν τη διαφορά στην ακρίβεια ανάμεσα σε κάθε proxy task και τα random weights, αναδεικνύοντας την ταχύτερη επίτευξη υψηλών επιδόσεων που επιτυγχάνεται με τη χρήση των pretrained weights.

Έμμεση διεργασία με τις καλύτερες επιδόσεις

Όλα τα pretrained weights από τα αντίστοιχα proxy tasks παρουσίασαν καλύτερες επιδόσεις σε σχέση με τα random weights. Η σειρά από το καλύτερο προς το χειρότερο ήταν: 1) Image Inpainting, 2) Image Colorization, 3) Image Rotation Prediction.

Αυτό μας δείχνει ότι το Image Inpainting ήταν το πιο αποδοτικό task, πιθανώς επειδή είναι μια σύνθετη διεργασία που απαιτεί από το μοντέλο να μάθει πώς να συμπληρώνει κενά στις εικόνες, διατηρώντας τη συνοχή τους. Αυτή η διαδικασία αναγκάζει το δίκτυο να κατανοήσει τόσο το περιεχόμενο της εικόνας όσο και το ευρύτερο πλαίσιο (context) της, ενισχύοντας έτσι την ικανότητά του να αναπαριστά πολύπλοκες σχέσεις και χαρακτηριστικά.

Από την άλλη, το Image Rotation Prediction αποδείχθηκε πιο "αδύναμο", καθώς είναι μια πιο απλή διεργασία που εστιάζει μόνο στον προσανατολισμό των αντικειμένων. Αν και παρέχει κάποια χρήσιμη πληροφορία στο μοντέλο, δεν είναι τόσο πλούσιο σε χαρακτηριστικά όσο τα άλλα δύο tasks, με αποτέλεσμα η πληροφορία που μεταφέρει να είναι λιγότερο χρήσιμη στο downstream task, όπως φαίνεται από τις ελαφρώς αυξημένες, αλλά περιορισμένες, επιδόσεις του.

3.7 Επίλογος και Μελλοντικές Επεκτάσεις

Συνοψίζοντας τα αποτελέσματα αυτής της διπλωματικής εργασίας, καταλήγουμε στο συμπέρασμα ότι η Αυτο-Εποπτευόμενη Μάθηση μέσω της χρήσης έμμεσων διεργασιών προσφέρει σημαντικά πλεονεκτήματα στην εκπαίδευση νευρωνικών δικτύων για ταξινόμηση εικόνων. Συγκεκριμένα, τα προεκπαιδευμένα βάρη από τα proxy tasks αποδείχθηκαν πιο αποδοτικά σε σχέση με τα τυχαία βάρη, τόσο όσον αφορά τις τελικές επιδόσεις όσο και την ταχύτητα εκπαίδευσης. Η διεργασία Image Inpainting αναδείχθηκε ως η πιο αποτελεσματική, ακολουθούμενη από το Image Colorization, ενώ το Image Rotation Prediction παρουσίασε τα χαμηλότερα αποτελέσματα, αν και παραμένει καλύτερο από την εκπαίδευση με τυχαία βάρη.

Η προσέγγιση αυτή προσφέρει σημαντικά πλεονεκτήματα, καθώς επιτρέπει την εκμετάλλευση μη επισημασμένων δεδομένων, μειώνοντας τις ανάγκες για εξειδικευμένο labeling και επιταχύνοντας τη διαδικασία εκπαίδευσης. Το γεγονός ότι τα προεκπαιδευμένα βάρη επέτρεψαν στο μοντέλο να συγκλίνει ταχύτερα και να πετύχει υψηλότερες επιδόσεις με λιγότερους πόρους υπογραμμίζει την αξία της μεταφοράς μάθησης σε προβλήματα που δεν έχουν πολλά επισημασμένα δεδομένα.

Όσον αφορά τις μελλοντικές επεκτάσεις της παρούσας εργασίας, θα μπορούσε να εξεταστεί η εφαρμογή πιο σύνθετων αρχιτεκτονικών νευρωνικών δικτύων ή και η ενσωμάτωση attention mechanisms, όπως τα Transformer-based μοντέλα, για τη βελτίωση της απόδοσης σε πιο απαιτητικά tasks. Παράλληλα, η εφαρμογή των έμμεσων διεργασιών σε μεγαλύτερα και πιο ετερογενή σύνολα δεδομένων θα μπορούσε να δώσει σημαντικά αποτελέσματα σχετικά με την γενικευσιμότητα των χαρακτηριστικών που μαθαίνουν τα μοντέλα.

Μια ακόμη σημαντική κατεύθυνση για περαιτέρω μελέτη θα ήταν η εξέταση της προσαρμογής των έμμεσων διεργασιών σε δεδομένα διαφορετικής φύσης, όπως δεδομένα βίντεο ή ακόμα και δεδομένα αισθητήρων, για τη βελτίωση της ανάλυσης πολυδιάστατων και διαδοχικών πληροφοριών. Τέλος, η χρήση πιο προηγμένων τεχνικών fine-tuning, όπως το progressive fine-tuning, θα μπορούσε να οδηγήσει σε μεγαλύτερη αποδοτικότητα και οικονομία πόρων, ειδικά σε περιπτώσεις όπου τα δεδομένα του downstream task είναι περιορισμένα.

Η έρευνα σε αυτό τον τομέα παρουσιάζει σημαντικές προοπτικές, και η παρούσα εργασία προσέφερε μία λεπτομερή αξιολόγηση των δυνατοτήτων της Αυτο-Εποπτευόμενης Μάθησης με τη χρήση έμμεσων διεργασιών στη βελτίωση της απόδοσης νευρωνικών δικτύων για την ταξινόμηση εικόνων.

Βιβλιογραφία

Βιβλια:

- Μπούταλης, Ι. & Συρακούλης, Γ. (2010). Υπολογιστική Νοημοσύνη και εφαρμογές. Κρίκος – Αφοί Παπαμάρκου Ο.Ε.
- Goodfellow, I. (2016). Deep learning. MIT Press Ltd
- Hecht-Nielsen, R. (1989). *Neurocomputing*. Addison-Wesley Longman Publishing Co., Inc..

Αναφορές:

- [1] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278-2324.
- [2] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.
- [3] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- [4] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).
- [5] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1-9).
- [6] Tan, M. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. *arXiv preprint arXiv:1905.11946*.
- [7] He, K., Fan, H., Wu, Y., Xie, S., & Girshick, R. (2020). Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 9729-9738).
- [8] Doersch, C., Gupta, A., & Efros, A. A. (2015). Unsupervised visual representation learning by context prediction. In *Proceedings of the IEEE international conference on computer vision* (pp. 1422-1430).
- [9] Noroozi, M., & Favaro, P. (2016, September). Unsupervised learning of visual representations by solving jigsaw puzzles. In *European conference on computer vision* (pp. 69-84). Cham: Springer International Publishing.
- [10] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248-255). ieee.
- [11] Zhang, R., Isola, P., & Efros, A. A. (2016). Colorful image colorization. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part III*

14 (pp. 649-666). Springer International Publishing.

- [12]Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., & Efros, A. A. (2016). Context encoders: Feature learning by inpainting. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2536-2544).
- [13]Gidaris, S., Singh, P., & Komodakis, N. (2018). Unsupervised representation learning by predicting image rotations. *arXiv preprint arXiv:1803.07728*.
- [14]Krizhevsky, A., & Hinton, G. (2009). Learning multiple layers of features from tiny images.
- [15]Pan, S. J., & Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10), 1345-1359.
- [16]Long, M., Cao, Y., Wang, J., & Jordan, M. (2015, June). Learning transferable features with deep adaptation networks. In *International conference on machine learning* (pp. 97-105). PMLR.
- [17]Vaswani, A. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*.
- [18]McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5, 115-133.
- [19]Kingma, D. P. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- [20]Ioffe, S. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*.
- [21]Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1), 1929-1958.
- [22]Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6), 386.
- [23]Lowe, D., & Broomhead, D. (1988). Multivariable functional interpolation and adaptive networks. *Complex systems*, 2(3), 321-355.
- [24]Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *nature*, 323(6088), 533-536.
- [25]Watkins, C. J., & Dayan, P. (1992). Q-learning. *Machine learning*, 8, 279-292.
- [26]Rummery, G. A., & Niranjan, M. (1994). *On-line Q-learning using connectionist systems* (Vol. 37, p. 14). Cambridge, UK: University of Cambridge, Department of Engineering.
- [27]Mnih, K. (2015). Mnih V., Kavukcuoglu K., Silver D., Rusu Aa, Veness J., Bellemare MG, et al. *Human-level control through deep reinforcement learning*, *Nature*, 518(7540), 529-533.
- [28]Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1), 267-288.
- [29]Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55-67.
- [30]Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *science*, 313(5786), 504-507.

- [31]Bengio, Y., Lamblin, P., Popovici, D., & Larochelle, H. (2006). Greedy layer-wise training of deep networks. *Advances in neural information processing systems*, 19.
- [32]Masci, J., Meier, U., Cireşan, D., & Schmidhuber, J. (2011). Stacked convolutional auto-encoders for hierarchical feature extraction. In *Artificial Neural Networks and Machine Learning–ICANN 2011: 21st International Conference on Artificial Neural Networks, Espoo, Finland, June 14–17, 2011, Proceedings, Part I* 21 (pp. 52-59). Springer Berlin Heidelberg.
- [33]Ng, A., & Autoencoder, S. (2011). CS294A Lecture notes. *Dosegljivo: https://web. stanford. edu/class/cs294a/sparseAutoencoder_2011new. pdf*. [Dostopano 20. 7. 2016].
- [34]Vincent, P., Larochelle, H., Bengio, Y., & Manzagol, P. A. (2008, July). Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th international conference on Machine learning* (pp. 1096-1103).
- [35]Kingma, D. P. (2013). Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.
- [36]Sakurada, M., & Yairi, T. (2014, December). Anomaly detection using autoencoders with nonlinear dimensionality reduction. In *Proceedings of the MLSDA 2014 2nd workshop on machine learning for sensory data analysis* (pp. 4-11).
- [37]Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.
- [38]Redmon, J. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*.
- [39]Ren, S. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *arXiv preprint arXiv:1506.01497*.
- [40]Kalchbrenner, N., Grefenstette, E., & Blunsom, P. (2014). A convolutional neural network for modelling sentences. *arXiv preprint arXiv:1404.2188*.
- [41]Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical image analysis*, 42, 60-88.
- [42]Bojarski, M., Del Testa, D., Dworakowski, D., Firner, B., Flepp, B., Goyal, P., ... & Zieba, K. (2016). End to end learning for self-driving cars. *arXiv preprint arXiv:1604.07316*.
- [43]Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems*, 27.
- [44]Simonyan, K., & Zisserman, A. (2014). Two-stream convolutional networks for action recognition in videos. *Advances in neural information processing systems*, 27.
- [45]Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... & Chintala, S. (2019). Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32.
- [46]Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Zheng, X. (2016). {TensorFlow}: a system for {Large-Scale} machine learning. In *12th USENIX symposium on operating systems design and implementation (OSDI 16)* (pp. 265-283).
- [47]Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12, 2825-2830.

- [48]Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., ... & Rush, A. M. (2020, October). Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations* (pp. 38-45).
- [49]Howard, J., & Gugger, S. (2020). Fastai: a layered API for deep learning. *Information*, 11(2), 108.
- [50]Devlin, J. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- [51]Yenduri, G., Ramalingam, M., Selvi, G. C., Supriya, Y., Srivastava, G., Maddikunta, P. K. R., ... & Gadekallu, T. R. (2024). Gpt (generative pre-trained transformer)—a comprehensive review on enabling technologies, potential applications, emerging challenges, and future directions. *IEEE Access*.
- [52]Griffin, G., Holub, A., & Perona, P. (2007). *Caltech-256 object category dataset* (Vol. 10). Pasadena: Technical Report 7694, California Institute of Technology.
- [53]Zhang, C., Ranjan, S., & Hansen, J. H. (2018, June). An Analysis of Transfer Learning for Domain Mismatched Text-independent Speaker Verification. In *Odyssey* (pp. 181-186).
- [54]Patel, C., Shah, D., & Patel, A. (2013). Automatic number plate recognition system (anpr): A survey. *International Journal of Computer Applications*, 69(9).
- [55]Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., Manzagol, P. A., & Bottou, L. (2010). Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of machine learning research*, 11(12).
- [56]Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020, November). A simple framework for contrastive learning of visual representations. In *International conference on machine learning* (pp. 1597-1607). PMLR.
- [57]Wu, Z., Xiong, Y., Yu, S. X., & Lin, D. (2018). Unsupervised feature learning via non-parametric instance discrimination. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3733-3742).
- [58]Blog, A. K. (2015). The unreasonable effectiveness of recurrent neural networks. URL: <http://karpathy.github.io/2015/05/21/rnn-effectiveness/> dated May, 21, 31.
- [59]Yue-Hei Ng, J., Hausknecht, M., Vijayanarasimhan, S., Vinyals, O., Monga, R., & Toderici, G. (2015). Beyond short snippets: Deep networks for video classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4694-4702).