ΔΗΜΟΚΡΙΤΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΡΑΚΗΣ

ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ

Τμήμα Ηλεκτρολόγων Μηχανικών & Μηχανικών Ηλεκτρονικών Υπολογιστών

Τομέας Ηλεκτρονικής Και Τεχνολογίας Συστημάτων Πληροφορικής

Εργαστήριο Συστημάτων Αυτομάτου Ελέγχου Και Ρομποτικής

Διπλωματική Εργασία

Ανάπτυξη Ταξινομητών με χρήση Συνελικτικών Νευρωνικών Δικτύων για Κατηγοριοποίηση με Ελάχιστα Παραδείγματα

Γεώργιος Λεπίδας

Επιβλέπων Καθηγητής: Ιωάννης Μπούταλης

Ξάνθη, Οκτώβριος 2022

Ευχαριστίες

Πρώτα απ’ όλα θα ήθελα να ευχαριστήσω τους δύο υπέροχους γονείς μου, Δημήτρη και Ντίνα για την αγάπη και την στήριξη που μου παρέχουν σε όλα τα επίπεδα, από τότε που ήρθα σε αυτήν την ζωή, την αδερφή μου και πρότυπό μου για το θάρρος της, Χρύσα, την γιαγιά μου Τζένη, η οποία με διδάσκει καθημερινά πως να ζω, και όλη την οικογένεια και φίλους μου οι οποίοι με βοηθούν να γίνομαι καλύτερος άνθρωπος. Επίσης, θα ήθελα να ευχαριστήσω τους ανθρώπους εκείνους οι οποίοι βρεθήκαν και υπήρξαν σημαντικό κομμάτι της ζωής μου, για όλες τις στιγμές και την διαμόρφωση μου, αλλά και κάποιους «αγνώστους» οι οποίοι θα βρίσκονται για πάντα στην καρδιά μου.

Στα πλαίσια αυτής της διπλωματικής, θα ήθελα να εκφράσω την θερμή ευγνωμοσύνη μου στον κύριο Ιωάννη Μπούταλη, επιβλέποντα της εργασίας και καθηγητή μου, για την πολύτιμη βοήθεια και την υποστηρικτική στάση του, την εμπιστοσύνη του προς το πρόσωπο μου καθώς και την ευκαιρία που μου προσέφερε να γνωρίσω και να εφαρμόσω ανεκτίμητες γνωστικές έννοιες. Ακόμη, θα ήθελα να ευχαριστήσω θερμά τον υποψήφιο διδάκτορα αλλά και αδελφικό μου φίλο, Σωκράτη Γκέλιο, ο οποίος ήταν δίπλα μου σε όλα τα στάδια υλοποίησης αυτής της εργασίας, καθώς και το Δημοκρίτειο Πανεπιστήμιο Θράκης και το τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών, στο οποίο είχα την τύχη να φοιτήσω.

Περίληψη

Πυρήνα της παρούσας διπλωματικής αποτελεί πρωτίστως η μελέτη της έννοιας της Μάθησης Λίγων Λήψεων (Few-Shot Learning) αλλά και ενός από τα βασικότερα είδη τεχνητών νευρωνικών δικτύων, των συνελικτικών (Convolutional Neural Networks-CNNs). Οι παραπάνω έννοιες εντάσσονται στον επιστημονικό τομέα της Μηχανικής Μάθησης (Machine Learning) η οποία αποτελεί αναπόσπαστο κομμάτι της Τεχνητής Νοημοσύνης (Artificial Intelligence).

Το Few-Shot Learning υπάγεται στην ευρύτερη περιοχή της Μηχανικής Μάθησης, συνιστά έναν από τους πιο προκλητικούς τομείς αυτής και ανερχόμενο πεδίο έρευνας και εξέλιξης. Ο λόγος για τα παραπάνω έγκειται στο γεγονός της διαφοροποίησης από τους «παραδοσιακούς» τρόπους μηχανικής μάθησης, όπου για την επίλυση της εκάστοτε διεργασίας αξιοποιούνται τεράστιες βάσεις δεδομένων. Σε αντίθεση με αυτές τις περιπτώσεις, στο Few-Shot Learning, το εκπαιδευόμενο μοντέλο καλείται να μάθει και να αναγνωρίζει έχοντας στην διάθεση του ελάχιστα έως καθόλου δεδομένα, προσομοιώνοντας ακόμη περισσότερο τον ανθρώπινο τρόπο μάθησης.

Τα Συνελικτικά Νευρωνικά Δίκτυα είναι ένα από τα πιο χρησιμοποιούμενα είδη τεχνητών νευρωνικών δικτύων, βρίσκουν ευρεία εφαρμογή σε ποικίλες διεργασίες (αναγνώριση προσώπου, εύρεση και αναγνώριση αντικειμένου, επεξεργασία ομιλίας και άλλες) και είναι το επικρατέστερο στο επιστημονικό πεδίο της Μηχανικής Όρασης. Θεμέλιο της λειτουργίας τους αποτελεί η μαθηματική πράξη της συνέλιξης και είναι εμπνευσμένα από τα κύτταρα του οπτικού φλοιού των έμβιων όντων.

Κεντρικό πεδίο διερεύνησης της παρούσας διπλωματικής, είναι η καταγραφή και η σύγκριση των επιδόσεων διάφορων συνελικτικών νευρωνικών δικτύων αλλά και η ανάπτυξη νέων «ταξινομητών» (classifiers), για το πρόβλημα της κατηγοριοποίησης εικόνων μέσα από ελάχιστα παραδείγματα (Few-shot Image Classification).

Πιο συγκεκριμένα, η διερεύνηση βασίζεται στον αλγόριθμο Generation-0 της εργασίας “Self-supervised Knowledge Distillation for Few-shot Learning” των Jathushan Rajasegaran et al. [1], της οποίας ο κώδικας (https://github.com/brjathu/SKD) αποτελεί και την βάση των πειραμάτων. Στον αλγόριθμο Generation-0 της [1], συνδυάζεται ο «παραδοσιακός» τρόπος μάθησης με την επίλυση μίας επιπλέον βοηθητικής διεργασίας κατά την εκπαίδευση του μοντέλου, της οποίας τα δεδομένα δημιουργούνται με την χρήση αύτο-εποπτευόμενης μάθησης (Self-supervised Learning). Μέσω αυτής της προσέγγισης, οι Jathushan Rajasegaran et al. [1], κατάφεραν να δημιουργήσουν ισχυρότερα μοντέλα για την διεργασία του Few-Shot Image Classification.

Κατά την εκπόνηση αυτής διπλωματικής διενεργήθηκαν τρεις βασικές περιπτώσεις συγκρίσεων όπου μελετώνται:

- ποια είναι η επίδραση του βάθους των συνελικτικών δικτύων Resnets [31] στις επιδόσεις του αλγορίθμου Generation-0 της [1] στο Few-Shot Image Classification Task,
- ποια είναι η αντίστοιχη επίδραση του πλάτους των δικτύων, και η εξέταση της υπόθεσης για το αν θα επέλθει βελτίωση με την προσθήκη ενός SE block [32] στο δίκτυο, καθώς και
- η υπόθεση για το αν μία διαφορετική βοηθητική διεργασία και προσέγγιση αύτο-εποπτευόμενης μάθησης, και ειδικότερα με την χρησιμοποίηση ενός Στοχαστικού Αυτοκωδικοποιητή (Variational Autoencoder) [63], θα επιφέρει βελτιώσεις στο Few-Shot Image Classification.

Abstract

The main purpose of this diploma thesis is to study the Few-Shot Learning concept and one of the most basic Artificial Neural Networks type, Convolutional Neural Networks (CNNs). The above terms come under the purview of Machine Learning, which is an integral part of the Artificial Intelligence field.

Few-Shot Learning constitutes one of the biggest challenges and one of the most upcoming and evolutional research fields in this domain. This is due to the structural differences between the traditional machine learning methods and the few-shot learning methods. In traditional learning methods, a huge amount of data is used for the model's training, while in few-shot learning methods the model can learn with little or no data at all, and in that way simulate human learning.

Convolutional Neural Networks is one of the most used types of artificial neural networks, applied to a variety of tasks (such as face recognition, object detection and recognition, natural language processing, and more), with predominant use in Computer Vision applications. Their functionality is based on the mathematical operation of convolution, and they are inspired by the visual brain cells of living creatures.

The principal objective of this diploma thesis is the performance monitoring and comparison of diverse convolutional classifiers, as is the creation and development of new models for the Few-Shot Image Classification task.

More specifically, the study is based on the Generation-0 algorithm, introduced by Jathushan Rajasegaran et al in the "Self-supervised Knowledge Distillation for Few-shot Learning" [1] research paper. The official code implementation of their work (https://github.com/brjathu/SKD) [1] constitutes the experimental basis for this diploma thesis. Generation-0 algorithm is a method where the traditional learning is combined with the addition and solution by the model during its training of an auxiliary task where the algorithm uses Self-Supervised Learning to create the auxiliary task data. With their approach, Jathushan Rajasegaran et al. [1] achieved the creation of more powerful models for the Few-shot Image Classification task.

During this thesis implementation, a total of three comparative cases are herein made focusing on:

- the effect of the convolutional networks Resnets [31] depth to the Few-Shot Image Classification performance with the utilization of Generation-0 algorithm [1],
- the effect of the convolutional networks Resnets [31] width to the same task, as well as the exploration of the assumption that the addition of a SE block [32] to the network would boost the models' performance, and lastly,
- the assumption that a different auxiliary Self-supervised task, specifically with the usage of a Variational Autoencoder [63], would lead to improvements to the Few-Shot Image Classification task.

Περιεχόμενα

Εισαγωγή

0.1 Few-Shot Learning

0.1.1 Το Έναυσμα Δημιουργίας

Διανύουμε μία εποχή, όπου η Τεχνητή Νοημοσύνη αναπτύσσεται ραγδαία, με την ύπαρξη ήδη κάποιων εντυπωσιακών αποτελεσμάτων και περιπτώσεων όπου οι Μηχανές επικρατούν των ανθρώπων σε διάφορα πεδία. Τέτοια παραδείγματα αποτελούν, το χαρακτηριστικό πρόγραμμα AlphaGo [33], το οποίο κατάφερε να νικήσει επαγγελματίες πρωταθλητές στο αρχαίο παιχνίδι Go, αλλά και το συνελικτικό δίκτυο ResNet [31], το οποίο επέδειξε καλύτερες επιδόσεις από τους ανθρώπους στο πρόβλημα κατηγοριοποίησης εικόνων της συλλογής δεδομένων ImageNet. Επιπροσθέτως, πλέον εντάσσονται στην ζωή μας όλο και περισσότερα «έξυπνα εργαλεία» όπως οι ισχυρές μηχανές αναζήτησης, η καθοδήγηση μέσω φωνής, τα αυτοκίνητα αυτόνομης οδήγησης και τα βιομηχανικά Robots με σκοπό την διευκόλυνση μας στις διάφορες πτυχές της καθημερινότητας.

Τα παραπάνω οφέλη της Τεχνητής Νοημοσύνης οφείλονται σε πολύ μεγάλο βαθμό στην εξέλιξη του τεχνολογικού εξοπλισμού (όπως η δημιουργία ισχυρών καρτών γραφικών GPUs) αλλά και στις τεράστιες συλλογές δεδομένων που έχουμε πλέον στην διάθεση μας. Αναφορικά, το πρόγραμμα AlphaGo [33] εκπαιδεύτηκε αξιοποιώντας μία βάση δεδομένων, η οποία περιείχε περίπου 30 εκατομμύρια καταγεγραμμένες κινήσεις ειδικών επαγγελματιών, σύμφωνα με αντίστοιχα δεδομένα - συνθήκες του παιχνιδιού. Έτσι, μπορούμε να δούμε πως η Τεχνητή Νοημοσύνη και οι Αλγόριθμοι Μηχανικής Μάθησης αποκτούν την δυνατότητα επίδειξης εξαιρετικών επιδόσεων, όταν τους τροφοδοτήσουμε με ένα τεράστιο πλήθος δεδομένων.

Αντίθετα, σε περιπτώσεις όπου τα δεδομένα είναι ελάχιστα, παρουσιάζουν τεράστιες δυσκολίες γενίκευσης. Από την άλλη πλευρά, οι άνθρωποι έχουμε μία έφεση στο να μαθαίνουμε γρήγορα κι εύκολα, έχοντας στην διάθεση μας πολύ λίγα δεδομένα. Για παράδειγμα, για ένα παιδάκι, το οποίο γνωρίζει την πράξη της πρόσθεσης, θα του ήταν εύκολο να αξιοποιήσει αυτήν την γνώση και να αφομοιώσει και την πράξη του πολλαπλασιασμού μέσα από μόλις λίγα παραδείγματα (όπως τα 2 × 3 = 2 + 2 + 2 και 1 × 3 = 1 + 1 + 1), ή θα αρκούσε να του δείξουμε μία φωτογραφία ελέφαντα, ώστε στο μέλλον να είναι σε θέση να τους διακρίνει ανάμεσα από πολλά άλλα είδη ζώων. Εμπνευσμένη από την ευκολία που παρουσιάζουμε εμείς οι άνθρωποι στην ανάπτυξη δεξιοτήτων κατέχοντας ελάχιστη πληροφορία, αλλά και την αδυναμία των μηχανών σε αντίστοιχες περιπτώσεις, η Τεχνητή Νοημοσύνη στράφηκε προς την ανάπτυξη μοντέλων με σκοπό την βελτίωση τους και σε τέτοιες συνθήκες. Το Few-Shot Learning συνιστά το αποτέλεσμα αυτής της νέας κατεύθυνσης.

0.1.2 Η Χρησιμότητα και ο Ρόλος του στην Τεχνητή Νοημοσύνη

Παρά το γεγονός της εγγενούς δυσκολίας ανάπτυξης της, η Μάθηση Λίγων Λήψεων κατέχει σημαντικό ρόλο στον τομέα της Μηχανικής Μάθησης και της Τεχνητής Νοημοσύνης, καθώς παρέχει καινοτόμες οπτικές και επεκτείνει τις δυνατότητες επίλυσης για νέα πιο σύνθετα προβλήματα.

Συγκεκριμένα μέσω της Μάθησης Λίγων Λήψεων:

- προστίθεται ένα ακόμη κριτήριο όσον αφορά τον βαθμό στον οποίο οι μηχανές προσεγγίζουν τον ανθρώπινο τρόπο σκέψης,

- παρέχεται η δυνατότητα μάθησης σε περιπτώσεις όπου τα δεδομένα είναι δυσεύρετα,
- ελαχιστοποιείται η καταβολή της ανθρώπινης προσπάθειας όσον αφορά την απόκτηση των δεδομένων και μειώνεται το υπολογιστικό κόστος.

0.1.2.1 Το Few-Shot Learning ως Τρόπος Αξιολόγησης των Μηχανών (Testbed)

Η γεφύρωση του κενού μεταξύ Τεχνητής Νοημοσύνης και ανθρώπου, αποτελεί σημαντικό εγχείρημα και η Μάθηση Λίγων Λήψεων κατέχει τον ρόλο ενός εκ των βασικότερων κριτηρίων του βαθμού επίτευξης αυτού. Η λογική είναι ότι η επαρκής ανταπόκριση μίας μηχανής σε ένα πρόβλημα Μάθησης Λίγων Λήψεων, συνεπάγεται και αποτελεσματικότερη προσέγγιση του ανθρώπινου τρόπου λειτουργίας.

Για την επεξήγηση του παραπάνω συλλογισμού, μπορούμε να λάβουμε υπόψη το πρόβλημα παραγωγής χαρακτήρων (Character Generation Task) [34], όπου η μηχανή καλείται να παράγει νέους χειρόγραφους χαρακτήρες, έχοντας στην διάθεση της ελάχιστα παραδείγματα. Για την αντιμετώπιση αυτού το προβλήματος, στην [34], περιλαμβάνεται μία προσέγγιση όπου το μοντέλο επιδιώκει την απόκτηση περεταίρω γνώσης, αποσυνθέτοντας σε μικρότερα κομμάτια τους ελάχιστους χειρόγραφους χαρακτήρες που έχει στην διάθεση του, προκειμένου να τα χρησιμοποιήσει για την σύνθεση ενός νέου χαρακτήρα.

Ο παραπάνω τρόπος μάθησης είναι πολύ συγγενικός με την συλλογιστική διαδικασία που ακολουθείται και από τους ανθρώπους. Παρόμοιες εφαρμογές έχουν υλοποιηθεί και σε άλλους τομείς όπως η Μηχανική Όραση (Computer Vision) [1, 25, 27, 28, 29], η Ρομποτική (Robotics) [68, 69, 70], η Επεξεργασία Φυσικής Γλώσσας (Natural Language Processing) [71] και η Επεξεργασία Ήχου (Acoustic Signal Processing) [72, 73, 74, 75].

0.1.2.2 Η Δυνατότητα Μάθησης για Σπάνιες Περιπτώσεις

Περιπτώσεις όπου η Μάθηση Λίγων Λήψεων αποτελεί επιτακτική ανάγκη, είναι όταν η πρόσβαση στην εποπτευόμενη πληροφορία (δηλαδή στις ετικέτες των δεδομένων) είναι δυσχερής εξαιτίας θεμάτων ιδιωτικότητας, ασφάλειας ή ηθικής.

Παράδειγμα τέτοιας περίπτωσης αποτελεί η ανακάλυψη φαρμάκων [35], όπου τα πειράματα τα οποία αποδίδουν τα διαθέσιμα δεδομένα, ελλοχεύουν κινδύνους καθώς ενδέχεται να επιφέρουν τοξικές επιπτώσεις και επιδιώκεται η ελαχιστοποίηση της υλοποίησης αυτών.

0.1.2.3 Ελαχιστοποίηση Καταβολής της Ανθρώπινης Προσπάθειας για την Απόκτηση των Δεδομένων και Μείωση του Υπολογιστικού Κόστους

Τα περισσότερα δεδομένα που υπάρχουν στον κόσμο, είναι μη κατηγοριοποιημένα. Η απόδοση της ετικέτας σε αυτά, συνιστά μία επίπονη διαδικασία και απαιτεί σημαντική καταβολή ανθρώπινης προσπάθειας.

Για παράδειγμα, η αντιστοίχιση των ετικετών για μία τεράστια συλλογή δεδομένων όπως είναι η ImageNet[1] αποτελεί εξαντλητική, χρονοβόρα και κοστοβόρα διαδικασία. Η ανάπτυξη της Μάθησης Λίγων Λήψεων, μπορεί να συμβάλει και να οδηγήσει στην απαλλαγή της παραπάνω διαδικασίας.

Επίσης, η μείωση και η αποφυγή της χειροκίνητης κατηγοριοποίησης δεδομένων, αλλά και οι πολύ μικρότερες σε όγκο συλλογές δεδομένων που αξιοποιούνται στα προβλήματα Few-Shot Learning, συνεισφέρουν στην μείωση του απαιτούμενου υπολογιστικού κόστους.

[1] Η συλλογή δεδομένων ImageNet αποτελεί μία από τις σημαντικότερες και μεγαλύτερες συλλογές δεδομένων της Μηχανικής Όρασης, η οποία συγκροτεί και βάση αναφοράς για διάφορες διεργασίες με κυριότερη την Κατηγοριοποίηση Εικόνων (Image Classification).

0.2 Συνελικτικά Νευρωνικά Δίκτυα

Τα Συνελικτικά Νευρωνικά Δίκτυα αποτελούν συγχρόνως το πιο χρησιμοποιούμενο είδος Νευρωνικών Δικτύων, την πιο προκλητική περιοχή σε επίπεδο έρευνας και εφαρμογών αλλά και το βασικό εργαλείο του επιστημονικού πεδίου της Μηχανικής Όρασης. Η λειτουργία τους είναι εμπνευσμένη από τον οπτικό φλοιό του ανθρώπινου εγκεφάλου και μονοπωλούν το ενδιαφέρον σε εφαρμογές που αφορούν την επεξεργασία βίντεο και εικόνας.

Υποδέχονται κατά κύριο λόγο εισόδους σε μορφή εικόνας, αλλά μέσω διάφορων μετασχηματισμών, μπορούν να χρησιμοποιηθούν με ικανοποιητική απόδοση και σε προβλήματα των οποίων οι είσοδοι αφορούν σειριακά δεδομένα. Έτσι, βρίσκουν εφαρμογή σε οποιαδήποτε διεργασία της οποίας τα δεδομένα εισόδου μπορούν να μετατραπούν σε δομή εικόνας προκειμένου να εξαχθεί ένα χρήσιμο συμπέρασμα. Παραδείγματος χάριν, η οπτικοποίηση της κυματομορφής ενός ηχητικού δεδομένου μπορεί να συμβάλλει σημαντικά στην εξαγωγή ωφέλιμων συμπερασμάτων για τις ιδιότητες του συγκεκριμένου ήχου.

0.2.1 Ιστορία και Αρχιτεκτονικές Συνελικτικών Νευρωνικών Δικτύων

Οι πρώτες επιτυχημένες εφαρμογές των Συνελικτικών Δικτύων εμφανιστήκαν το 1998 με την χρήση της αρχιτεκτονικής LeNet (Yann LeCun, Leon Bottou, Yoshua Bengio και Patrick Haffner) [36]. Έπειτα, η μελέτη τους επανήλθε δυναμικά στο προσκήνιο, το 2012 με την χρήση της αρχιτεκτονικής AlexNet (Alex Krizhevsky, Ilya Sutskever και Geoff Hinton) [37] με ευρείες εφαρμογές πάνω στην Μηχανική Όραση.

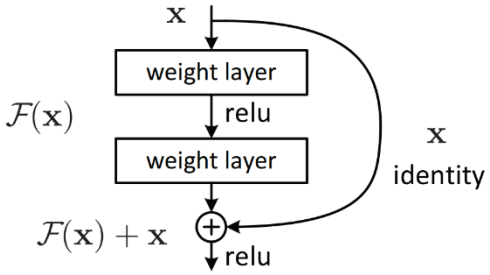
Έκτοτε, χάρη στις υψηλές επιδόσεις που επιδείχθηκαν, το πεδίο των Συνελικτικών Νευρωνικών Δικτύων ακμάζει ραγδαία, και έχουν δημιουργηθεί ποικίλες αρχιτεκτονικές οι οποίες πειραματίζονται με τα συνελικτικά επίπεδα, συνδυάζουν τις διάφορες λειτουργίες και προσθέτουν νέες, με σκοπό την υψηλότερη απόδοση στις ήδη υπάρχουσες διεργασίες αλλά και την αντιμετώπιση νέων προβλημάτων.

Παρακάτω, ακολουθούν μερικές αξιοσημείωτες αρχιτεκτονικές Συνελικτικών Νευρωνικών Δικτύων.

- LeNet-5 [36]: Πρόκειται για μία από τις πρώτες επιτυχημένες εφαρμογές Συνελικτικών Νευρωνικών Δικτύων (1998 Yann LeCun). Χρησιμοποιήθηκε για να διαβάζει κυρίως ταχυδρομικούς κώδικες, ψηφία και γενικότερα παρόμοιου είδους απλά χαρακτηριστικά.
- AlexNet [37]: Παρόμοια αρχιτεκτονική με το LeNet [36], αλλά βαθύτερο, μεγαλύτερο και με καινοτομία την προσθήκη συνεχόμενων συνελικτικών επιπέδων, χωρίς ενδιάμεσα να παρεμβάλλεται κάποιο επίπεδο Pooling όπως επικρατούσε μέχρι τότε.
- ZF Net [38]: Πρόκειται για μία αναβάθμιση της αρχιτεκτονικής AlexNet [37], η οποία προτάθηκε από τους Matthew Zeiler και Rob Fergus, τροποποιώντας τις υπερπαραμέτρους της αρχιτεκτονικής του δικτύου. Τέτοιες τροποποιήσεις συνιστούν η αύξηση του μεγέθους των μεσαίων συνελικτικών στρωμάτων, και η μείωση του βήματος (stride) και του μεγέθους των φίλτρων του πρώτου συνελικτικού στρώματος.
- VGGNets [39]: Αφορά μία σειρά προτάσεων Συνελικτικών Νευρωνικών Δικτύων από την Ομάδα Οπτικής Γεωμετρίας (Visual Geometry Group) του πανεπιστήμιου της Οξφόρδης, όπως τα VGG-11, VGG-11-LRN, VGG 13, VGG-16, και VGG-19. Βασική συνεισφορά αυτών, είναι το γεγονός ότι αποδείχθηκε πειραματικά η σημαντικότητα του βάθους των συνελικτικών δικτύων για τις υψηλές επιδόσεις.
- GoogLeNets [40, 48,49, 50]: Πρόκειται για αρχιτεκτονικές (Συνελικτικά Δίκτυα Inceptions) που εισήχθησαν κατά κύριο λόγο από μέλη της Google. Το βασικό τους χαρακτηριστικό είναι ότι συμπεριλαμβάνουν «Μονάδες Σημείων Εκκίνησης» (Inception Modules). Κύρια συμβολή αυτών αποτέλεσε η δραστική μείωση του αριθμού παραμέτρων του δικτύων. Στα Συνελικτικά Δίκτυα

Inceptions συμπεριλαμβάνονται τέσσερις βασικές εκδοχές, αυτές των Inception v1 [40], Inception v2 [48, 49], Inception v3 [49] και Inception v4 [50].

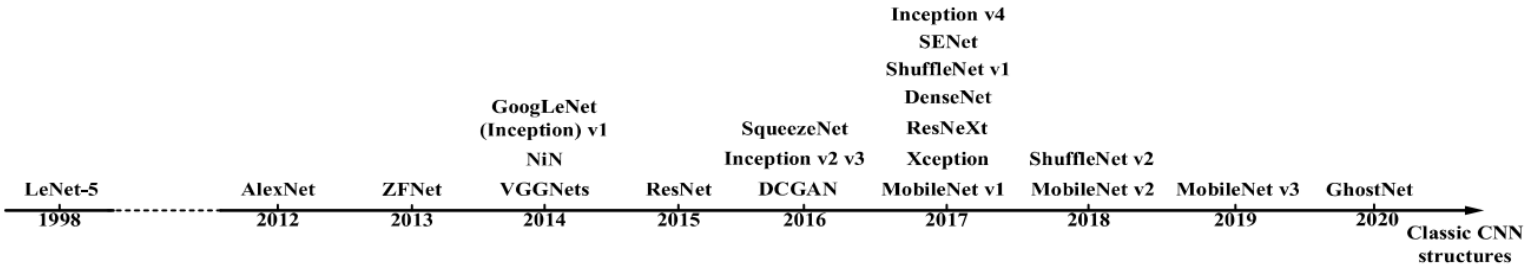
- ResNets [31]: Αρχιτεκτονικές οι οποίες αναπτύχθηκαν από τους Kaiming He, Xiangyu Zhang, Shaoqing Ren, και Jian Sun, ερευνητές της Microsoft. Η σημαντική τους ιδιαιτερότητα είναι η προσθήκη των ειδικών εμφωλευμένων συνδέσεων (residual connections), η εντονότερη χρήση Κανονικοποίησης Παρτιδών (Batch Normalization) και η ελάττωση των Πλήρως Συνδεδεμένων Επίπεδων (Fully Connected Layer) στο τέλος των δικτύων. Χάρη στην ιδιόμορφη αρχιτεκτονική τους μειώθηκε το φαινόμενο των Εξαφανιζόμενων και Εκτινασσόμενων Παραγώγων (Vanishing and Exploding Gradients Problem), και επιδείχθηκαν πολύ καλά αποτελέσματα στην ταχύτητα εκπαίδευσης αλλά και τις επιδόσεις στα διάφορα Datasets αναφοράς.



Σχήμα 0.2.1 - 1: Απεικόνιση των ειδικών συνδέσεων (Residual Connections) που λαμβάνουν χώρα στα Συνελικτικά Νευρωνικά Δίκτυα Resnets.

Πηγή: Deep Residual Learning for Image Recognition - Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun - Microsoft Research [31]

- DCGAN [41]: Πρόκειται για Συνελικτικό Παραγωγικό Αντιπαραθετικό Δίκτυο το οποίο προτάθηκε από τους A. Radford, L. Metz, και S. Chintala για την εφαρμογή στην συλλογή δεδομένων Κατανόησης Σκηνικών Μεγάλης Κλίμακας (Large-scale Scene Understanding – LSUN dataset).
- MobileNets [42, 43, 44]: Αφορά μία σειρά από ελαφριά μοντέλα, τα οποία προτάθηκαν από την Google, για χρήση και ενσωμάτωση σε μικρότερες συσκευές όπως τα κινητά τηλέφωνα. Χρησιμοποιούν «ογκικές» συνελίξεις (depthwise-κατά βάθος) και υψηλού επιπέδου τεχνικές για την δημιουργία «λεπτών» βαθιών Νευρωνικών Δικτύων. Υπάρχουν τρεις βασικές εκδόσεις αυτού του είδους, η MobileNet v1 [42], η MobileNet v2 [43] και η MobileNet v3 [44].
- ShuffleNets [45, 46]: Είναι μία σειρά από Συνελικτικά Νευρωνικά Δίκτυα η οποία προτάθηκε από την εταιρεία MEGVII για την επίλυση της μη επαρκούς υπολογιστικής ισχύος των κινητών τηλεφώνων. Αυτά τα μοντέλα συνδυάζουν ομαδικές συνελίξεις ενός σημείου (pointwise group convolution), το ανακάτεμά - αναδιάταξη των καναλιών και άλλες τεχνικές οι οποίες επιτυγχάνουν την δραστική μείωση του υπολογιστικού κόστους με πολύ μικρές κυρώσεις στις επιδόσεις. Οι βασικές εκδοχές αυτού του είδους δικτύων είναι η ShuffleNet v1 [45] και η ShuffleNet v2 [46].
- GhostNet [47]: Πρόκειται για άλλη μία αποτελεσματική προσπάθεια μείωσης του υπολογιστικού Κόστους των Συνελικτικών Νευρωνικών Δικτύων που προήλθε από τους Kai Han et al. Η υλοποίηση περιλαμβάνει πέρα από την κλασική λειτουργία της συνέλιξης, και λειτουργίες γραμμικών μετασχηματισμών με σκοπό την περικοπή συνελικτικών αποτελεσμάτων τα οποία αποδεικνύονται περιττά και ονομάζονται «φαντάσματα» (“ghosts”).



Σχήμα 0.2.1 - 2: Απεικόνιση διαφόρων αρχιτεκτονικών Συνελικτικών Νευρωνικών Δικτύων στο χρονολογικό φάσμα, σύμφωνα με την ημερομηνία δημοσίευσης τους.

Πηγή: A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects - Zewen Li, Wenjie Yang, Shouheng Peng, Fan Liu, Member, IEEE [51]

0.3 Εισαγωγή στο Πειραματικό Κομμάτι και στον Αλγόριθμο Generation-0 της [1]

Η εργασία των Jathushan Rajasegaran et al. [1], της οποίας η υλοποίηση αποτελεί την βάση αυτής της διπλωματικής εργασίας χωρίζεται σε δύο «γενιές» εκπαίδευσης, τις Generation-0 και Generation-1. Στην συγκεκριμένη πειραματική ανάλυση επιλέχθηκε το πρώτο σκέλος μάθησης (Generation-0).

0.3.1 Περιγραφή του Αλγορίθμου Generation-0

Πρόκειται για προσέγγιση η οποία αφορά το πρόβλημα της κατηγοριοποίησης εικόνας μέσα από ελάχιστα παραδείγματα με την επιστράτευση της Αύτο-εποπτευόμενης Μάθησης (Self-Supervised Learning) και μίας βοηθητικής διεργασίας (Auxiliary Task).

Στο κομμάτι αυτό της [1], οι ερευνητές επικεντρώθηκαν στην βελτίωση των δυνατοτήτων του μοντέλου για μία καλή «αναπαράσταση χαρακτηριστικών» (feature representation), συμπεριλαμβάνοντας στην εκπαίδευση αυτού μία επιπλέον βοηθητική διεργασία. Δηλαδή, η εκπαίδευση αποτελείται από δύο διαφορετικές διεργασίες,

- την βασική η οποία είναι η συνήθης, δηλαδή αυτή της ορθής κατηγοριοποίησης της εκάστοτε εικόνας στην αντίστοιχη ετικέτα της,
- και την βοηθητική, η οποία είναι μία διεργασία αύτο-εποπτευόμενης μάθησης.

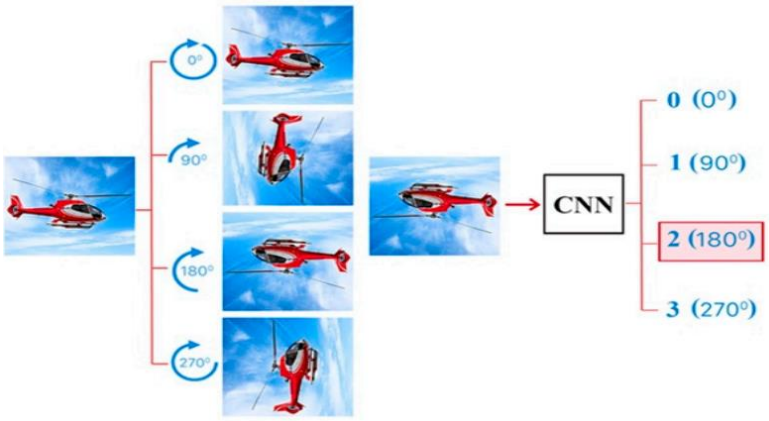
Ειδικότερα, για την βοηθητική διεργασία, ο αλγόριθμος δημιουργεί επιπρόσθετα δεδομένα για κάθε παρτίδα δεδομένων μέσα από την ίδια παρτίδα. Τα επιπρόσθετα δεδομένα προκύπτουν από περιστροφή όλων των δεδομένων της εκάστοτε παρτίδας κατά 90°, κατά 180° και κατά 270°. Στην συνέχεια τα παραχθέντα δεδομένα προστίθενται σε μία συνεπτυγμένη λίστα δεδομένων με τα ήδη υπάρχοντα (με περιστροφή 0°), συνοδευόμενα από τις αντίστοιχες ετικέτες κατηγοριοποίησης της κλάσης τους αλλά και από τις ετικέτες κατηγοριοποίησης του βαθμού περιστροφής τους (0°,90°,180°,270°), τα οποία συνιστούν τα τελικά επεξεργασμένα δεδομένα. Το μοντέλο τροφοδοτείται με τα τελικά δεδομένα προκειμένου να εκπαιδευτεί και στις δύο διεργασίες: στην κατηγοριοποίηση κλάσης των τελικών επεξεργασμένων δεδομένων καθώς και στην πρόβλεψη του βαθμού περιστροφής αυτών.

Τα αλγοριθμικά βήματα της ακριβούς διαδικασίας απεικονίζονται στο Σχήμα 0.3.1 - 1, ενώ το Σχήμα 0.3.1 - 2 αναπαριστά μία αποτύπωση της επεξεργασίας που λαμβάνει χώρα για την εν λόγω βοηθητική διεργασία.

Require: f_Φ, f_Θ, f_Ψ, $\mathcal{D}$
  for e iterations **do** ▷ *Generation Zero training*
    while $\mathcal{B} \sim \mathcal{D}$ **do**
      $\mathbf{x}^{90}, \mathbf{x}^{180}, \mathbf{x}^{270} \leftarrow \text{rotate}(\mathbf{x})$
      $\hat{\mathbf{x}} \leftarrow \{\mathbf{x}, \mathbf{x}^{90}, \mathbf{x}^{180}, \mathbf{x}^{270}\}$, and $\hat{\mathbf{y}} \leftarrow \{\mathbf{y}, \mathbf{y}, \mathbf{y}, \mathbf{y}\}$
      $\hat{\mathbf{r}} \leftarrow \{\mathbf{0}, \mathbf{1}, \mathbf{2}, \mathbf{3}\}$ ▷ where $\mathbf{0}$ is an all zero vector with length m
      $\hat{\mathbf{v}} \leftarrow f_\Phi(\hat{\mathbf{x}})$, $\hat{\mathbf{p}} \leftarrow f_\Theta(\hat{\mathbf{v}})$, $\hat{\mathbf{q}} \leftarrow f_\Psi(\hat{\mathbf{p}})$
      $\mathcal{L}_0 \leftarrow \mathcal{L}_{ce}(\hat{\mathbf{p}}, \hat{\mathbf{y}}) + \alpha \cdot \mathcal{L}_{ss}(\hat{\mathbf{q}}, \hat{\mathbf{r}})$
      $\{\Phi, \Theta, \Psi\} \leftarrow \{\Phi, \Theta, \Psi\} - \nabla_{\{\Phi, \Theta, \Psi\}} \mathcal{L}_0$

Σχήμα 0.3.1 - 1: Απεικόνιση των αλγοριθμικών βημάτων της διαδικασίας εκπαίδευσης για την Generation-0.

Πηγή: "Self-supervised Knowledge Distillation for Few-shot Learning" Jathushan Rajasegaran, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Mubarak Shah [1]



Σχήμα 0.3.1 - 2: Απεικόνιση της βοηθητικής διεργασίας αύτο-εποπτευόμενης μάθησης (self-supervised auxiliary task).

Πηγή: Survey on Self-Supervised Learning: Auxiliary Pretext Tasks and Contrastive Learning Methods in Imaging - Saleh Albelwi [52]

Αξίζει να σημειωθεί ότι η εκπαίδευση δεν διαφοροποιείται για διαφορετικές τιμές των n και k για οποιοδήποτε n-ways – k-shot πρόβλημα (όπως στις [17, 18]), καθώς πρόκειται για προσέγγιση που επιδιώκει την αποτύπωση μιας καλής αναπαράστασης [του τελικού χαρακτηριστικού (feature) του μοντέλου - Feature Learning].

Μέσα από πειράματα που συμπεριλαμβάνουν αυτό το είδος προσέγγισης, οι ερευνητές της [1], επιβεβαίωσαν ότι αυτός ο τρόπος εκπαίδευσης δίνει ώθηση στην απόδοση των μοντέλων για το πρόβλημα κατηγοριοποίησης εικόνων με ελάχιστες λήψεις.

0.3.2 Σχετική Έρευνα – Παρόμοιες Εργασίες στη Διεθνή Βιβλιογραφία

Στην συγκεκριμένη παράγραφο θα παρουσιαστούν σχετικές δημοσιευμένες εργασίες στην διεθνή βιβλιογραφία παρόμοιες με το κομμάτι (Generation-0) της εργασίας [1], το οποίο αποτελεί την βάση του πειραματικού μέρους αυτής της διπλωματικής εργασίας.

0.3.2.1 Αύτο-εποπτευόμενη Μάθηση και Βοηθητικές Διεργασίες

Η επίλυση των βοηθητικών διεργασιών οι οποίες δημιουργούνται με την Αύτο-εποπτευόμενη Μάθηση απαιτούν την κατανόηση υψηλού επιπέδου από το μοντέλο, τοποθετώντας το σε μία θέση που είναι ικανό να εξάγει ισχυρές αναπαραστάσεις με σημαντικές πληροφορίες για το εκάστοτε αντιμετωπιζόμενο πρόβλημα. Οι βασικές διαφορές στις υπάρχουσες τεχνικές αυτού του είδους έγκεινται στον τρόπο με τον οποίο ο αλγόριθμος δημιουργεί τις ετικέτες για την βοηθητική διεργασία μέσα από τα δεδομένα.

Στην εργασία των S. Gidaris et al. με τίτλο “Unsupervised representation learning by predicting image rotations” [4], μαθαίνονται χρήσιμες αναπαραστάσεις μέσα από την πρόβλεψη του βαθμού περιστροφής των εισερχόμενων εικόνων. Οι Doersch et al. στην έρευνα τους “Unsupervised visual representation learning by context prediction” [5] εκπαιδεύουν ένα συνελικτικό νευρωνικό δίκτυο στην πρόβλεψη της σχετικής θέσης μεταξύ δύο κομματιών της εικόνας, με την λογική ότι για να επιτευχθεί κάτι τέτοιο, προϋποτίθεται το αντικείμενο της εικόνας να έχει αναγνωριστεί. Η βάση της ίδιας ιδέας επεκτείνεται ακόμη περισσότερο από τους M. Noroozi και P. Favaro στην εργασία με τίτλο “Unsupervised learning of visual representations by solving jigsaw puzzles” [6] όπου το μοντέλο υποβάλλεται στην πρόβλεψη πολλαπλών τέτοιων σχετικών θέσεων και την επίλυση του παζλ ενός μέρους της εικόνας. Επιπροσθέτως, ο χρωματισμός της εικόνας και η καταμέτρηση αντικειμένων επιστρατεύονται ως βοηθητικές διεργασίες στις εργασίες των R. Zhang et al “Colorful image colorization” [7] και M. Noroozi et al. “Representation learning by learning to count” [8] αντίστοιχα. Οι X. Zhai et al. στην εργασία “Self-supervised semi-supervised learning” [9] προτείνουν μία προσέγγιση αύτο-εποπτευόμενης μάθησης εμπνευσμένη από την [4], ως αντιμετώπιση σε ένα πρόβλημα ημί-εποπτευόμενης φύσης το οποίο εμπεριέχει και κατηγοριοποιημένα και μη-κατηγοριοποιημένα δεδομένα.

Πολύ συγγενικές με τον τρόπο προσέγγισης της Generation-0 της [1] είναι οι εργασίες “Discriminative unsupervised feature learning with convolutional neural networks” [10] των A. Dosovitskiy et al, “Deep clustering for unsupervised learning of visual features” [11] των M. Caron et al. και “A simple framework for contrastive learning of visual representations” [12] των T. Chen et al., όπου επιδιώκεται η εκμάθηση του διαχωρισμού αναπαραστάσεων μεταξύ διάφορων επεξεργασιών και μετασχηματισμών των δεδομένων (augmented data) από τα ίδια τα δεδομένα. Η διαφορά έγκειται στο γεγονός ότι στην [1], οι ερευνητές χρησιμοποιούν τα augmented δεδομένα με σκοπό όχι τον διαχωρισμό των αναπαραστάσεων, αλλά την επέκταση αυτών μέσα από τα προσαυξημένα δεδομένα έτσι ώστε ο χώρος μίας κλάσης να δύναται να μαθευτεί μέσα από λίγα δεδομένα.

0.3.2.2 Few Shot Learning

Όσον αφορά την σχετική δουλειά στο κομμάτι του Few-Shot Learning, επειδή αποτελεί τον πυρήνα της παρούσας διπλωματικής εργασίας, στο θεωρητικό σκέλος εμπεριέχεται ένα κεφάλαιο αφιερωμένο σε αυτό και τις συνιστώσες του, όπου συμπεριλαμβάνεται και η παράθεση των κυριότερων έρευνών που έχουν λάβει χώρα. Έτσι, στην παράγραφο αυτήν θα αρκεστούμε σε μία πολύ σύντομη και στοχευμένη αναφορά στις διάφορες προσεγγίσεις που είναι συγγενικές.

Αυτές εκτείνονται από το metric Learning και την εκμάθηση μίας μετρικής [13, 14, 15, 16], μέχρι το Meta Learning [17, 18, 19, 20, 21] και την εκμάθηση μίας δυνατής κωδικοποίησης (embedding learning) [22, 23, 24]. Αξιοσημείωτο είναι ότι στην εργασία [24] γίνεται ο ισχυρισμός ότι η επιτυχία των Meta-Learning μεθόδων έγκειται στην ικανότητα εξαγωγής μίας δυνατής αναπαράστασης των features που τις διακατέχει και όχι στο ίδιο το Meta-Learning.

0.3.2.3 Συνδυασμός Self-supervision και Βοηθητικών Διεργασιών για το Few-Shot Learning

Η πρόσφατη έρευνα των Da Chen et al. με τίτλο “Self-Supervised Learning For Few-Shot Image Classification” [25], συνδυάζει την χρησιμότητα μίας βοηθητικής διεργασίας (η οποία είναι η μεγιστοποίηση κοινών πληροφοριών στα features από augmented εικόνες) και την αύτο-εποπτευόμενη μάθηση με το meta-Learning για το πρόβλημα του Few-Shot Image Classification. Οι Nassim Ait Ali Braham et al. στην εργασία τους “Self Supervised Learning For Few Shot Hyperspectral Image Classification” [27] χρησιμοποιούν τον αύτο-εποπτευόμενο αλγόριθμο Barlow-Twins [26] για Few-Shot κατηγοριοποίηση υπερσπεκτρικών εικόνων[2].

Στην εργασία “Conditional Self-Supervised Learning for Few-Shot Classification” [28], των Yuexuan An et al., προτείνεται μία μέθοδος αύτο-εποπτευόμενης μάθησης (Conditional Self-Supervised Learning - CSS) η οποία καθοδηγείται από τις ετικέτες ενός προ-εκπαιδευμένου δικτύου (pre-trained network) με σκοπό την απόκτηση ενός πολύ καλού feature για το πρόβλημα της κατηγοριοποίησης λίγων λήψεων.

Τέλος, σχεδόν ίδια προσέγγιση με αυτήν της Generation-0 της [1], αποτελεί η εργασία των S. Gidaris et al. “Boosting few-shot visual learning with self-supervision” [29], η όποια διαφέρει στον τρόπο χρησιμοποίησης της βοηθητικής διεργασίας και της αύτο-εποπτευόμενης μάθησης, αλλά και στην αρχιτεκτονική όπου επιλέγεται ένας παράλληλος σχεδιασμός, σε αντίθεση με την [1] στην οποία προτιμάται μία σειριακή υλοποίηση.

[2] Εικόνες οι οποίες περιέχουν πληροφορία στο ηλεκτρομαγνητικό φάσμα που χρησιμοποιούνται για την εξαγωγή χρήσιμων πληροφοριών.

0.4 Διάρθρωση της Εργασίας

Πριν από την πειραματική ανάλυση, παρατίθεται ένα θεωρητικό σκέλος κατά το οποίο έγινε η προσπάθεια να συμπεριληφθούν οι χρήσιμες θεωρητικές έννοιες. Πιο συγκεκριμένα, στο Κεφάλαιο 1, γίνεται μία ανάλυση πάνω στα Συνελικτικά Νευρωνικά Δίκτυα, και στα θεωρητικά χαρακτηριστικά που εφαρμόστηκαν σε αυτήν την εργασία, ενώ στο Κεφάλαιο 2 επιδιώχθηκε να γίνει μία περιήγηση σε όλο το φάσμα του Few-Shot Learning, καθώς είναι το κεντρικό θέμα της παρούσας εργασίας. Στην συνέχεια, ακολουθεί το 3^{ο} Κεφάλαιο και πειραματικό σκέλος, όπου λαμβάνει χώρα η παρουσίαση των πειραμάτων, ο τρόπος διεξαγωγής αυτών, τα αποτελέσματα τους και τα συμπεράσματα με βάση αυτά.

I

Θεωρητικό Μέρος

Κεφάλαιο 1ο

Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks)

1.1 Είδη Μάθησης

Υπάρχουν διάφορα είδη μάθησης τα οποία χρησιμοποιούνται στην Μηχανική Μάθηση κι αυτά εξαρτώνται από τα δεδομένα και την φύση του προβλήματος. Για να εκτιμήσουμε τα πλεονεκτήματα και τα μειονεκτήματα των διάφορων ειδών μάθησης θα πρέπει να αντιληφθούμε από τι τύπο, μορφή και τι δομή χαρακτηρίζονται τα δεδομένα που έχουμε στο εκάστοτε πρόβλημα. Υπάρχουν δύο τύποι δεδομένων, τα δεδομένα με ετικέτα (κατηγοριοποιημένα - labeled data) και τα δεδομένα χωρίς ετικέτα (μη-κατηγοριοποιημένα - unlabeled data).

Τα κατηγοριοποιημένα δεδομένα περιέχουν τόσο την είσοδο όσο και την επιθυμητή έξοδο για κάθε είσοδο σε πλήρως αναγνώσιμη μορφή για την μηχανή, αλλά η απόκτηση τους απαιτεί μεγάλη καταβολή ανθρώπινης προσπάθειας. Από την άλλη, τα μη κατηγοριοποιημένα δεδομένα δεν συνοδεύονται από κάποια πληροφορία για την ζητούμενη έξοδο. Πρόκειται για ακατέργαστα δεδομένα τα οποία παρόλο που μας απαλλάσσουν από την ανάγκη ανθρώπινης προσπάθειας για την απόκτηση τους, απαιτούν πολύ πιο σύνθετες λύσεις στην αντιμετώπιση των διάφορων προβλημάτων. Παρότι υπάρχουν προβλήματα τα οποία μπορεί να είναι δύσκολα στην κατηγοριοποίηση τους, έχοντας στην διάθεση μας δεδομένα και των δύο ειδών, οι πιο συνηθισμένοι χρησιμοποιούμενοι τρόποι μάθησης είναι τρεις:

1. Η Εποπτευόμενη Μάθηση (Supervised Learning),
2. Η Μη Εποπτευόμενη Μάθηση (Unsupervised Learning),
3. Η Ενισχυτική Μάθηση (Reinforcement Learning).

Υπάρχουν και οι τρόποι Υβριδικής Μάθησης, οι οποίοι αφορούν προβλήματα των οποίων οι συλλογές αποτελούνται συγχρόνως και από κατηγοριοποιημένα και από μη-κατηγοριοποιημένα δεδομένα. Δύο γνωστοί τύποι Υβριδικής Μάθησης είναι η ημί-εποπτευόμενη μάθηση (Semi-Supervised Learning), και η αύτο-εποπτευόμενη μάθηση (Self-Supervised Learning), κατά τους οποίους ακολουθούνται ιδιαίτερες διαδικασίες προσέγγισης.

Στην ενότητα αυτήν αναφέρονται και επεξηγούνται τα τρία παραπάνω κύρια είδη μάθησης, αλλά και η Αύτο-εποπτευόμενη Μάθηση, η οποία αξιοποιείται και επιστρατεύεται στον βασικό αλγόριθμο για την παρούσα διπλωματική εργασία, (αλγόριθμος Generation-0 της [1]).

1.1.1 Εποπτευόμενη Μάθηση (Supervised Learning)

Η εποπτευόμενη μάθηση είναι ένα από τα πιο γνωστά είδη κλασσικής μάθησης του Machine Learning. Σε αυτό το είδος μάθησης, μία απαίτηση που τίθεται είναι τα δεδομένα να είναι κατηγοριοποιημένα και φυσικά οι ετικέτες των δεδομένων αυτές να είναι οι ορθές (η επιθυμητή έξοδος-στόχος να είναι σωστή για

κάθε δεδομένο εισόδου). Υπό αυτήν την προϋπόθεση, η εποπτευόμενη μάθηση συγκροτεί ένα εξαιρετικά ισχυρό εργαλείο.

Στην εποπτευόμενη μάθηση, δίνεται στο μοντέλο μία συλλογή δεδομένων εκπαίδευσης (training dataset) για να εργαστεί με αυτήν. Η συγκεκριμένη συλλογή συνοδεύεται από τις αντίστοιχες ετικέτες-στόχους για κάθε δεδομένο και συνήθως αποτελεί υποσύνολο της ευρύτερης συλλογής δεδομένων (dataset) της εκάστοτε διεργασίας στην οποία καλείται να ανταποκριθεί το μοντέλο.

Έτσι, το μοντέλο έχοντας στην διάθεση του τα δεδομένα και τις αντίστοιχες ετικέτες, εγκαθιδρύει μια σχέση αιτίου - αποτελέσματος μεταξύ αυτών. Στο τέλος της εκπαίδευσης, το μοντέλο έχει αποκτήσει καλή γνώση για τον τρόπο με τον οποίο δομούνται τα δεδομένα και την ζητούμενη σχέση μεταξύ εισόδων (δεδομένων) και εξόδων (ετικετών).

Με το που ολοκληρωθεί η εκπαίδευση, θα έχουμε στην διάθεση μας ένα μοντέλο του οποίου έχουν τροποποιηθεί κατάλληλα οι παράμετροι ώστε να ανταποκρίνεται επαρκώς στην παραπάνω σχέση. Το μοντέλο αυτό είναι έτοιμο για χρήση και μπορεί να δοκιμαστεί για την πρόβλεψη τιμών σε νέα άγνωστα παρόμοια δεδομένα (test set) με αυτά της συλλογής δεδομένων στην οποία εκπαιδεύτηκε.

1.1.2 Μη Εποπτευόμενη Μάθηση (Unsupervised Learning)

Η μη εποπτευόμενη μηχανική μάθηση αποτελεί επίσης μέρος της κλασσικής μάθησης κι έχει ένα μεγάλο πλεονέκτημα, αυτό της ικανότητας να δουλεύει με μη-κατηγοριοποιημένα δεδομένα. Πρακτικά, το γεγονός πως τα δεδομένα είναι μη κατηγοριοποιημένα έχει δύο πολύ σημαντικά προνόμια. Πρώτον, αποφεύγεται η επίπονη ανθρώπινη προσπάθεια για να προσδοθεί στα δεδομένα μία ετικέτα. Δεύτερον, σπάνια έχουμε έλλειψη δεδομένων καθώς υπάρχει μεγάλη πληθώρα, αφού τα περισσότερα δεδομένα στον πραγματικό κόσμο είναι μη-κατηγοριοποιημένα. Συνεπώς, οι συλλογές δεδομένων είναι πολύ μεγαλύτερες.

Στην εποπτευόμενη μάθηση, οι ετικέτες των δεδομένων είναι αυτές που παρέχουν την δυνατότητα της εξεύρεσης της ακριβούς φυσικής σχέσης μεταξύ της εισόδου και της επιθυμητής εξόδου. Στην μη-εποπτευόμενη μάθηση όμως οι ετικέτες δεν είναι διαθέσιμες, καθιστώντας απαραίτητη την δημιουργία εσωτερικών τρόπων και δομών για την ανακάλυψη των μοτίβων και των συσχετίσεων μεταξύ των δεδομένων.

Αυτή η δημιουργία των εσωτερικών δομών καθιστά τους αλγορίθμους μη εποπτευόμενης μάθησης πολυποίκιλους. Αντί για ένα προκαθορισμένο πρόβλημα, οι αλγόριθμοι μη εποπτευόμενης μάθησης μπορούν να τροποποιούν δυναμικά τις εσωτερικά δημιουργημένες δομές προσαρμόζοντας τες στα δεδομένα ώστε να επιλύουν προβλήματα διαφορετικού είδους από αυτά που επιλύονται στην εποπτευόμενη μάθηση. Για παράδειγμα, μία κοινή διεργασία μη-εποπτευόμενης μάθησης είναι η κατηγοριοποίηση παρόμοιων δεδομένων σε groups[3] που ονομάζεται «Συσταδοποίηση» (Clustering).

1.1.3 Ενισχυτική Μάθηση (Reinforcement Learning)

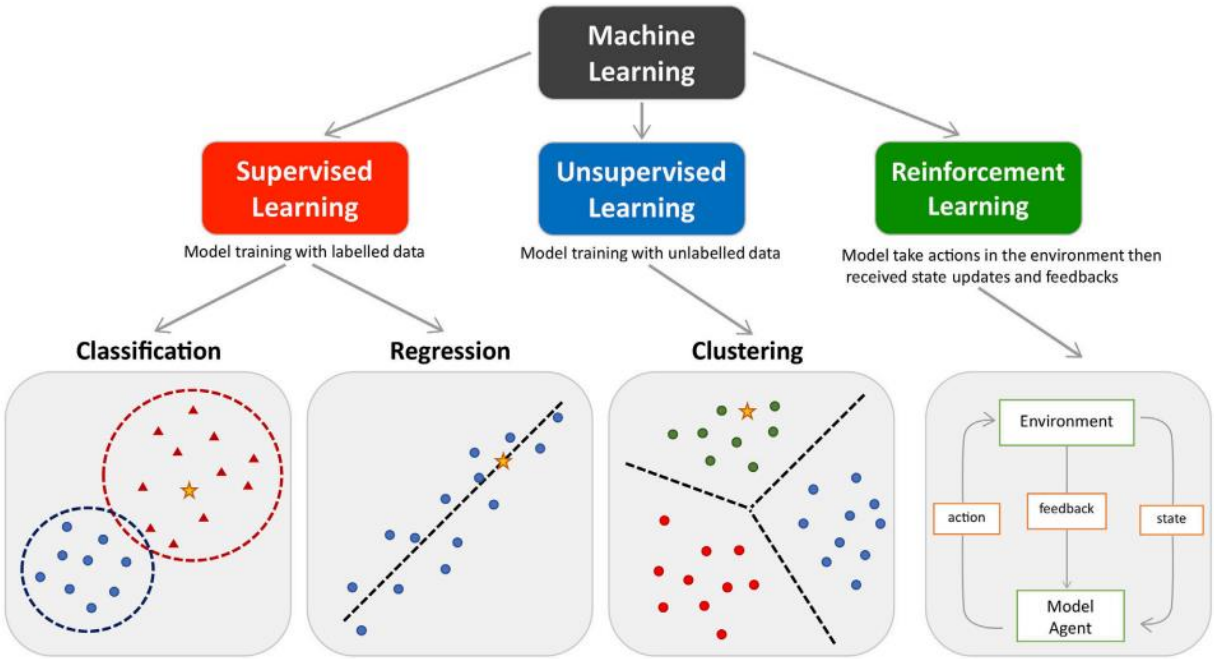
Η ενισχυτική μηχανική μάθηση διαφοροποιείται από τα παραπάνω είδη κλασσικής μάθησης, καθώς είναι ευθέως εμπνευσμένη από τον τρόπο με τον οποίο μαθαίνουμε εμείς οι άνθρωποι μέσα από τα δεδομένα. Σε αυτό το είδος μάθησης, τα μοντέλα βελτιώνονται μέσα από νέες καταστάσεις χρησιμοποιώντας την μέθοδο δοκιμής-σφάλματος. Οι επιθυμητές έξοδοι ενθαρρύνονται ή «ενισχύονται», ενώ οι μη επιθυμητές έξοδοι αποθαρρύνονται ή «τιμωρούνται».

Βασιζόμενη στην γενική ιδέα της «ψυχολογικής κατάστασης», η ενισχυτική μάθηση τοποθετεί τα μοντέλα μέσα σε ένα εργασιακό περιβάλλον το οποίο περιέχει έναν διερμηνέα και ένα σύστημα ανταμοιβής. Σε

[3] Για τις ομάδες αυτές των κατηγοριοποιημένων δεδομένων, στην διεθνή βιβλιογραφία χρησιμοποιείται ο όρος Clusters.

κάθε επανάληψη του αλγορίθμου μάθησης, η έξοδος δίνεται στον διερμηνέα ο οποίος αναλαμβάνει να αξιολογήσει το πόσο επιθυμητή είναι η συγκεκριμένη έξοδος.

Σε περίπτωση που το μοντέλο αποδίδει την σωστή έξοδο ή πλησιάζει προς μία ικανοποιητική προσέγγιση αυτής, ο διερμηνέας ενισχύει προς αυτήν την κατεύθυνση παρέχοντας στο μοντέλο ένα είδος επιβράβευσης από το σύστημα ανταμοιβής. Από την άλλη, όταν εξάγεται από το μοντέλο μία τιμή η οποία είναι μακριά από την επιθυμητή, τότε ο αλγόριθμος αναγκάζεται να επαναλάβει την διαδικασία μέχρις ότου να βρεθεί μία καλύτερη έξοδος. Η ανταμοιβή που παρέχεται στο μοντέλο έπειτα από κάθε επανάληψη του αλγορίθμου μάθησης είναι άμεσα συνδεδεμένη με την αποτελεσματικότητα της εξόδου. Ένα παράδειγμα ενισχυτικής μηχανικής μάθησης είναι η μεγιστοποίηση των κερδισμένων πόντων σε ένα παιχνίδι στο οποίο υπάρχει η δυνατότητα πολλών και διάφορων κινήσεων (όπου οι πόντοι επιτελούν τον ρόλο του συστήματος ανταμοιβής).



Σχήμα 1.1 - 1: Απεικόνιση των τριών βασικότερων ειδών μάθησης και παραδειγμάτων διεργασιών στις οποίες αυτά λαμβάνουν χώρα.

Πηγή: ResearchGate

1.1.4 Αύτο-Εποπτευόμενη Μάθηση (Selfsupervised Learning)

Η αύτο-εποπτευόμενη μάθηση (Self-supervised Learning) αποτελεί έναν συνδυασμό της εποπτευόμενης και της μη-εποπτευόμενης μάθησης. Αφορά έναν αποτελεσματικό τρόπο προσέγγισης, σε περιπτώσεις όπου τα κατηγοριοποιημένα δεδομένα είναι δυσεύρετα. Το μεγάλο προτέρημα της λειτουργίας της είναι η ανεξάρτητη δημιουργία ετικετών για αυτά τα δεδομένα, μέσα από τα ίδια τα δεδομένα (σύμφωνα με την δομή και τα χαρακτηριστικά τους), τις οποίες εν συνεχεία μπορεί να αξιοποιήσει με τον συνήθη εποπτευόμενο τρόπο μάθησης, για την αντιμετώπιση του εκάστοτε προβλήματος. Αξιοσημείωτο χαρακτηριστικό, αποτελεί το γεγονός ότι συνοδεύεται συνήθως από μία επιπρόσθετη βοηθητική διεργασία της οποίας η επίλυση δίνει ώθηση στις επιδόσεις των μοντέλων για την βασική διεργασία.

Έτσι, η Αύτο-Εποπτευόμενη Μάθηση αποτελεί ένα από τα πιο ανερχόμενα πεδία μελέτης της Μηχανικής Μάθησης καθώς:

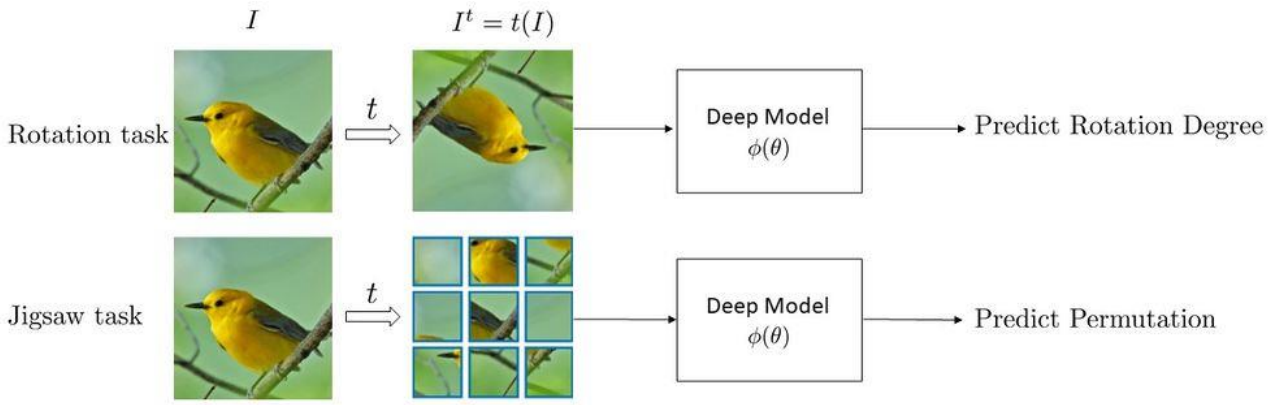
- Συμβάλλει στην μείωση της καταβολής της ανθρώπινης προσπάθειας και του απαιτούμενου κόστους για την απόκτηση κατηγοριοποιημένων δεδομένων.

- Ελαχιστοποιεί την επίπονη προετοιμασία των δεδομένων πριν από την διαδικασία εκπαίδευσης των μοντέλων (διαχωρισμός, φιλτράρισμα, αξιολόγηση και επεξεργασία προκειμένου να έρθουν στην επιθυμητή μορφή).
- Το ευρύτερο πλαίσιο λειτουργίας της τείνει να μεταφέρει τις ανθρώπινες γνωστικές ικανότητες και στις μηχανές, συνεισφέροντας στην κατεύθυνση της προσέγγισης του ανθρώπινου τρόπου λειτουργίας.

Πρόκειται για μέθοδο που βρίσκει εφαρμογή σε προβλήματα Επεξεργασίας Φυσικής Γλώσσας (Natural Language Processing), Μηχανικής Όρασης (Computer Vision) αλλά και σε βιομηχανικές περιοχές όπως ο εντοπισμός ρητορικής μίσους (Hate Speech Detection) από την Facebook [87] και η ανάλυση ιατρικών εικόνων (Medical Imaging Analysis) από την Google [88].

Παρά τα υποσχετικά προνόμια που παρέχει η Αύτο-Εποπτευόμενη μάθηση, συναντώνται και κάποιες προκλήσεις-δυσκολίες οι οποίες απαιτούν ιδιαίτερη προσοχή, όπως:

- Η ορθή δημιουργία ετικετών για τα δεδομένα, καθώς η συμπερίληψη άστοχων ετικετών στην βοηθητική διεργασία, μπορεί να επιφέρει αντίθετα αποτελέσματα στην επίδοση του εκπαιδευόμενου μοντέλου.
- Το απαιτούμενο υπολογιστικό κόστος αλλά και η διάρκεια της εκπαίδευσης αυξάνονται, καθώς πλέον κατά την διαδικασία της εκπαίδευσης συμπεριλαμβάνονται οι επιπλέον διεργασίες της δημιουργίας ετικετών από τον αλγόριθμο για τα δεδομένα αλλά και της εκπαίδευσης του μοντέλου σε αυτά.
- Η κατάλληλη επιλογή της βοηθητικής διεργασίας είναι κομβική. Για παράδειγμα, η συμπερίληψη ενός Selfsupervised Αυτοκωδικοποιητή στο πρόβλημα αναδημιουργίας εικόνων υψηλής ευκρίνειας ενδέχεται να βλάψει τις επιδόσεις του μοντέλου, καθώς είναι πιθανό ο Αυτοκωδικοποιητής να αναπαράγει και τον ανεπιθύμητο θόρυβο.



Σχήμα 1.1.4 - 1: Απεικόνιση δύο Selfsupervised Βοηθητικών διεργασιών. Στην πρώτη περίπτωση δημιουργείται μία περιστροφή του δεδομένου και το μοντέλο καλείται να προβλέψει σωστά τον βαθμό περιστροφής, ενώ στην δεύτερη περίπτωση εναλλάσσονται οι θέσεις διάφορων περιοχών της εικόνας και το μοντέλο εκπαιδεύεται στον ακριβή προσδιορισμό αυτών των εναλλαγών.

Πηγή: Math Wiki Server

1.2 Δομή και Λειτουργία Συνελικτικών Δικτύων

1.2.1 Δομή και Επίπεδα Συνελικτικών Νευρωνικών Δικτύων

Η συνήθης δομή ενός χαρακτηριστικού Συνελικτικού Νευρωνικού Δικτύου, απαρτίζεται από ποικίλα Επίπεδα ή "Blocks" (Convolutional Layers - Convolutional Blocks) τα οποία επιτελούν διάφορες λειτουργίες, με την πιο αξιοσημείωτη από αυτές, την λειτουργία της συνέλιξης, από την οποία πήρε και το όνομα του το συγκεκριμένο είδος Τεχνητών Νευρωνικών Δικτύων.

Τα επίπεδα αυτά ονομάζονται Συνελικτικά Επίπεδα[4] [5] και είναι τα εξής:

- Επίπεδο Εισόδου (Input Layer)
- Επίπεδο Συνέλιξης (Convolutional Layer)
- Επίπεδο Κανονικοποίησης Παρτίδων (Batch Normalization Layer)
- Επίπεδο Συνάρτησης Ενεργοποίησης (Activation Function Layer)
- Επίπεδο Pooling (Pooling Layer)
- Πλήρως Συνδεδεμένο Επίπεδο (Fully Connected Layer)
- Επίπεδο Dropout
- Επίπεδο Εξόδου ή Επίπεδο Συνάρτησης Κόστους (Output Layer – Cost Function Layer)

Το Επίπεδο Εισόδου αφορά την είσοδο και το γεγονός ότι αυτή εισέρχεται στο δίκτυο με τις φυσικές της διαστάσεις, γεγονός πολύ σημαντικό για τις επιδόσεις των μοντέλων σε προβλήματα εικόνων. Η λειτουργία του επιπέδου Συνάρτησης Ενεργοποίησης είναι αυτή που προσδίδει στην λειτουργία του δικτύου την έννοια της συνέχειας και της μη-γραμμικότητας καθιστώντας το πολύ ισχυρότερο, ενώ το επίπεδο Εξόδου ή Συνάρτησης Κόστους αφορά το τελικό παραχθέν αποτέλεσμα το οποίο είτε χρησιμοποιείται για την αξιολόγηση του μοντέλου είτε δίνεται στην εκάστοτε Συνάρτηση Κόστους για την εκπαίδευση του δικτύου.

Τα πλήρως Συνδεδεμένα επίπεδα βρίσκονται στο τέλος των Συνελικτικών Νευρωνικών Δικτύων, και πρόκειται για το σημείο του δικτύου εκείνο, στο οποίο συμβαίνει η μετατροπή των χωρικών δεδομένων επεξεργασίας σε δεδομένα μίας διάστασης (ευθυγράμμιση), και ουσιαστικά τα επεξεργασμένα δεδομένα εισέρχονται σε ένα κλασικό Πρόσω-Τροφοδοτούμενο Νευρωνικό Δίκτυο (Feedforward Neural Network-FNN προσκολλημένο στο τέλος του Συνελικτικού Δικτύου) προκειμένου να γίνει η αντίστοιχη πρόβλεψη.

Τα επίπεδα Συνέλιξης και Pooling, καθώς και οι λειτουργίες που λαμβάνουν χώρα στα επίπεδα Κανονικοποίησης Παρτιδών (Batch Normalization Layer) και Dropout επεξηγούνται αναλυτικά στις επόμενες παραγράφους.

Μερικά σημαντικά σχόλια για τα Συνελικτικά Επίπεδα είναι:

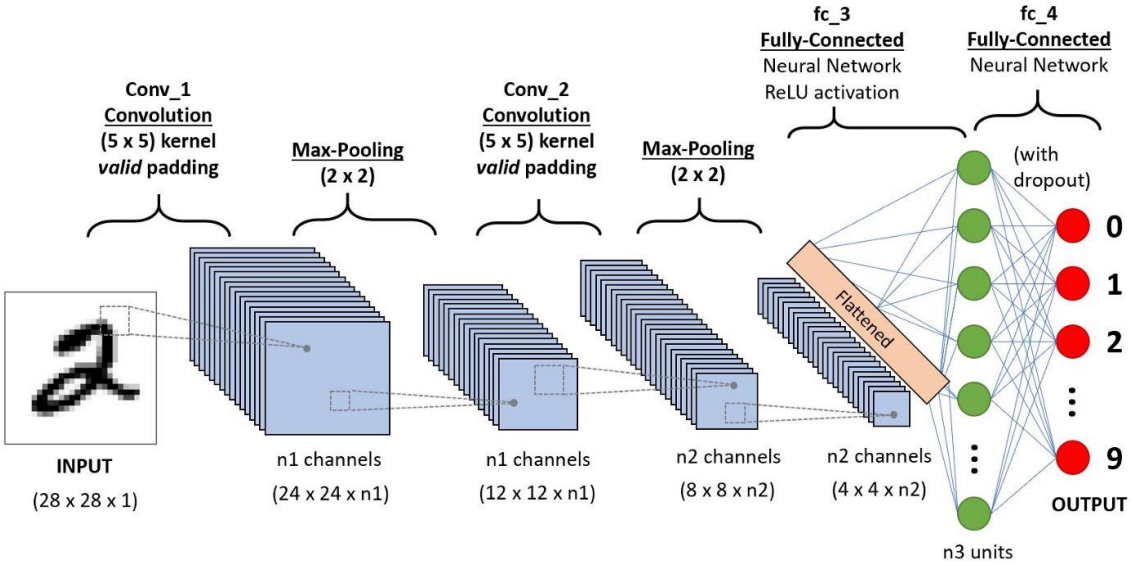
- Η είσοδος συνιστά χωρικό δεδομένο (μήκος-πλάτος εικόνας & βάθος ή αριθμός καναλιών) κι εισέρχεται με τις φυσικές της διαστάσεις στο δίκτυο. Οι λειτουργίες που λαμβάνουν χώρα στα

[4] Η λειτουργία της συνέλιξης συμπεριλαμβάνεται μόνο στο Επίπεδο Συνέλιξης. Η προέλευση του ονόματος των Συνελικτικών Επιπέδων δεν προέρχεται από την πράξη της συνέλιξης αλλά από το είδος δικτύων.

[5] Ο όρος των Συνελικτικών Επιπέδων – Blocks έχει δύο έννοιες στην βιβλιογραφία. Μπορεί να συμπεριλαμβάνει είτε ένα επίπεδο λειτουργίας (πχ της συνέλιξης), είτε περισσότερα επίπεδα λειτουργίας (πχ της συνέλιξης, της Κανονικοποίησης Παρτίδων, της συνάρτησης ενεργοποίησης και του επιπέδου Pooling). Στο θεωρητικό αυτό σκέλος θα εννοείται ότι κάθε συνελικτικό επίπεδο αντιστοιχίζεται σε μία μόνο λειτουργία.

διάφορα συνελικτικά επίπεδα, ως επί το πλείστον αποδίδουν επίσης χωρικά αποτελέσματα (με εξαίρεση τα τελευταία επίπεδα, όπου γίνεται η ευθυγράμμιση).

- Κατά την επεξεργασία των δεδομένων εισόδου και όσο προχωράμε βαθύτερα προς τα ενδιάμεσα συνελικτικά επίπεδα, τα δεδομένα επεξεργασίας αποκτούν μικρότερες διαστάσεις μήκους-πλάτους και μεγαλύτερη διάσταση βάθους (περισσότερα κανάλια επεξεργασίας).
- Μερικά Συνελικτικά Επίπεδα επιτελούν απλά μία λειτουργία, χωρίς να περιέχουν καθόλου παραμέτρους προς εκμάθηση (όπως τα επίπεδα Pool και των Συναρτήσεων Ενεργοποίησης, σε αντίθεση με τα Επίπεδα Συνέλιξης και τα Πλήρως Συνδεδεμένα τα οποία περιλαμβάνουν τις προς τροποποίηση παραμέτρους του δικτύου).
- Κάποια επίπεδα μπορούν να εμπεριέχουν επιπλέον υπερπαραμέτρους (όπως τα επίπεδα Συνέλιξης, τα επίπεδα Pool και τα Πλήρως Συνδεδεμένα επίπεδα, εν αντιθέσει με το επίπεδο της Συνάρτησης Ενεργοποίησης).



Σχήμα 1.2.1 - 1: Απεικόνιση ενός Συνελικτικού Νευρωνικού Δικτύου με σαφή διάκριση των επιπέδων του και συμπερίληψη των διαστάσεων των δεδομένων πριν και μετά από κάθε επίπεδο για ένα πρόβλημα κατηγοριοποίησης χειρόγραφων αριθμητικών ψηφίων (MNIST Handwritten Digits Classification Problem).

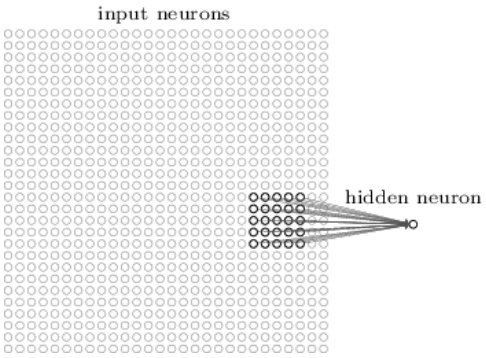
Πηγή: Towardsdatascience

1.2.2 Το Επίπεδο και η Λειτουργία της Συνέλιξης

Στα Συνελικτικά Νευρωνικά Δίκτυα η εικόνα εισέρχεται στο δίκτυο με τις φυσικές της διαστάσεις χωρίς να απαιτείται η ευθυγράμμιση των pixels της, προκειμένου να εφαρμόζει στο στρώμα εισόδου όπως συμβαίνει στα κλασικά Τεχνητά Νευρωνικά Δίκτυα. Έτσι, συνήθως έχουμε μία είσοδο είτε δύο διαστάσεων (μήκος και πλάτος εικόνας) είτε περισσότερων (βάθος - αναλόγως αν η εικόνα χαρακτηρίζεται από περισσότερα του ενός κανάλια, η οποία είναι και η συχνότερη περίπτωση). Η ορθή επεξεργασία των δεδομένων σε αυτές τις διαστάσεις οφείλεται στο Συνελικτικό Επίπεδο και αποτελεί βασικό αίτιο της βελτίωσης των επιδόσεων στις εφαρμογές εικόνων.

Στα συνελικτικά επίπεδα, δεν υπάρχει σύνδεση μεταξύ όλων των νευρώνων του εκάστοτε στρώματος (είτε εισόδου, είτε κρυμμένου) με το επόμενο όπως συμβαίνει στα Πρόσω - Τροφοδοτούμενα Νευρωνικά Δίκτυα. Αντί αυτού για κάθε έναν νευρώνα του επόμενου στρώματος αντιστοιχίζεται και μία συγκεκριμένη περιοχή νευρώνων του προηγούμενου στρώματος. Οι περιοχές αυτές ονομάζονται Πεδία Υποδοχής

(Receptive Fields) και σε αυτές λαμβάνει χώρα η λειτουργία της Συνέλιξης, προκειμένου να καταχωρηθεί στον κάθε νευρώνα του επόμενου στρώματος, ένα αποτέλεσμα.

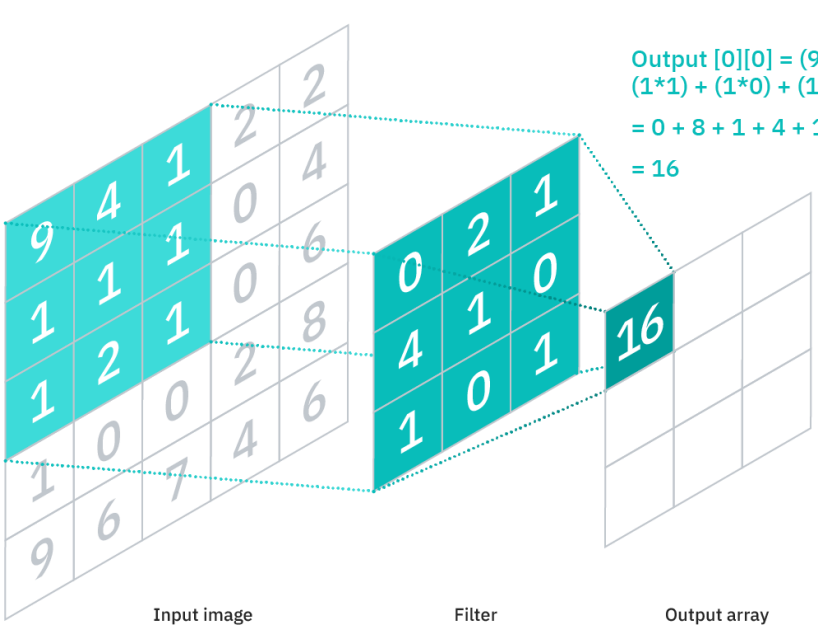


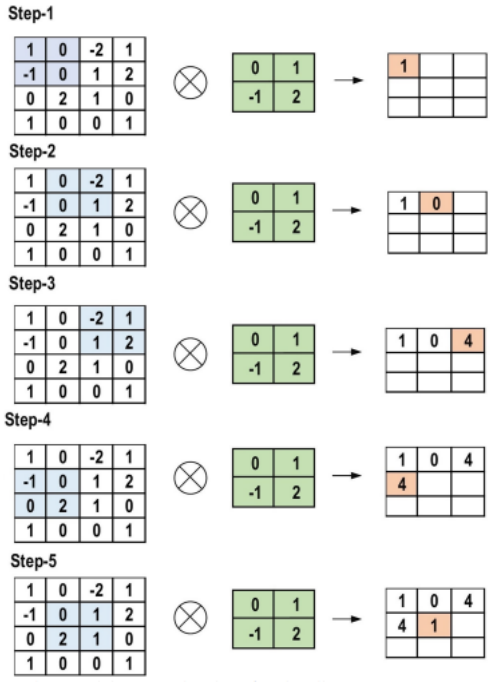
Σχήμα 1.2.2 - 1: Απεικόνιση σύνδεσης ενός συγκεκριμένου Πεδίου Υποδοχής της εισόδου με έναν νευρώνα του επόμενου στρώματος. Η είσοδος χωρίζεται σε συγκεκριμένο αριθμό Πεδίων Υποδοχής έτσι ώστε να συμπεριλαμβάνονται όλα τα στοιχεία - νευρώνες της. Για κάθε ένα από αυτά τα Πεδία Υποδοχής αντιστοιχίζεται ένας νευρώνας στο επόμενο στρώμα.[6]

Πηγή: Neural Networks and Deep Learning - Michael Nielsen (Online & Interactive Book)

Η πράξη της συνέλιξης[7], μέσω της οποίας πραγματοποιείται αυτή η αντιστοιχία, αφορά τον πολλαπλασιασμό στοιχείου-στοιχείου μεταξύ δύο πινάκων και την συμπερίληψη όλων αυτών των γινόμενών σε ένα ενιαίο στοιχείο μέσω της πρόσθεση τους. Οι δύο πίνακες μεταξύ των οποίων λαμβάνει χώρα αυτή η πράξη, προκειμένου να αντιστοιχιστεί στον νευρώνα του επόμενου στρώματος το ζητούμενο αποτέλεσμα, είναι ο πίνακας του εκάστοτε πεδίου υποδοχής και ένας πίνακας-φίλτρο. Τα φίλτρα [filters - ή αλλιώς πυρήνες (kernels)] περιλαμβάνουν τις παραμέτρους του δικτύου (έναν αριθμό βαρών και μία πόλωση το καθένα), οι οποίες μαθαίνονται κατά την διάρκεια της εκπαίδευσης, και το μέγεθος τους ισούται με αυτό των πεδίων υποδοχής.

Η ακριβής λειτουργία της συνέλιξης απεικονίζεται στο Σχήμα 1.2.2 - 2, ενώ στο Σχήμα 1.2.2 - 3 αναδεικνύεται κι ένας ενδεικτικός τρόπος «σάρωσης» της εισόδου μέσω της πράξης της συνέλιξης.





[6] Από το Σχήμα 1.2.2 - 1 μπορεί να γίνει ευδιάκριτος κι ο λόγος για τον οποίο συμβαίνει η μείωση των διαστάσεων μήκους-πλάτους της εικόνας όσο προχωράμε βαθύτερα στα επίπεδα ενός Συνελικτικού Νευρωνικού Δικτύου.

[7] Για λόγους απλοποίησης, σε πρώτο στάδιο, γίνεται η επεξήγηση της λειτουργίας της συνέλιξης για ένα δεδομένο εισόδου δύο διαστάσεων. Στην συνέχεια, στην παράγραφο **Ογκική - Σε Βάθος Συνελικτική Επεξεργασία** εξηγούνται και οι περιπτώσεις για δεδομένα εισόδου αποτελούμενα από τρεις διαστάσεις.

Σχήμα 1.2.2 - 2 & Σχήμα 1.2.2 - 3 :

Σχήμα 1.2.2 - 2 (Αριστερά): Απεικόνιση της πράξης της συνέλιξης μεταξύ ενός πεδίου υποδοχής της εισόδου (σκιαγραφημένη μπλε περιοχή στον αριστερό πίνακα) και ενός φίλτρου (ο μεσαίος πίνακας). Οι τιμές των στοιχείων του αριστερού πίνακα είναι η πληροφορία της εισόδου, ενώ οι τιμές των στοιχείων του φίλτρου αντιστοιχίζονται στα βάρη (παράμετροι του Συνελικτικού Νευρωνικού Δικτύου) τα οποία μεταβάλλονται κατά την εκπαίδευση.

Τα υπόλοιπα στοιχεία της εξόδου προκύπτουν από επαναλαμβανόμενες συνελίξεις μέσω των υπόλοιπων περιοχών υποδοχής και του ίδιου φίλτρου, οι οποίες λαμβάνουν χώρα μέχρις ότου «σαρωθεί» ολόκληρη η είσοδος.

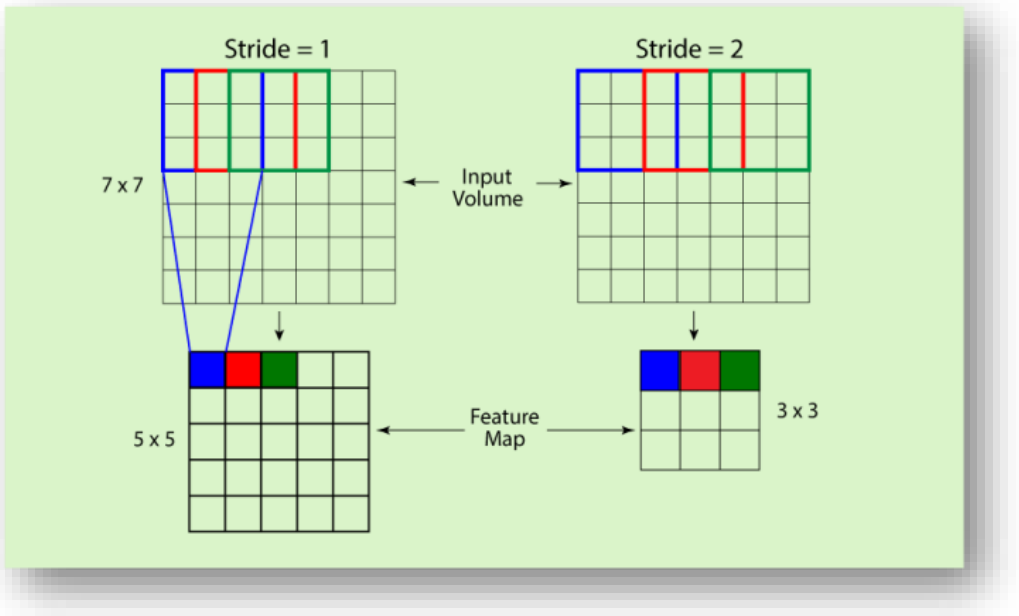
Πηγή: ibm.com

Σχήμα 1.2.2 - 3 (Δεξιά): Μία ενδεικτική απεικόνιση του τρόπου με τον οποίο «σαρώνεται» η εικόνα, μέσω της πράξης της συνέλιξης μεταξύ φίλτρου και εκάστοτε πεδίου υποδοχής, και η αντιστοιχία του κάθε συνελικτικού αποτελέσματος σε συγκεκριμένο στοιχείο εξόδου (Feature Map).

Πηγή: Review of deep learning: concepts, CNN architectures, challenges, applications, future directions - Laith Alzubaidi, Jinglan Zhang, Amjad J. Humaidi, Ayad Al-Dujaili, Ye Duan, Omran Al-Shamma, J. Santamaría, Mohammed A. Fadhel, Muthana Al-Amidie and Laith Farhan [53]

Όσον αφορά την «σάρωση» της εικόνας, τα φίλτρα «κυλούν» κατά μήκος και πλάτος της εισόδου και επιτελούν την πράξη της συνέλιξης μέχρις ότου να ολοκληρωθούν όλα τα στοιχεία της εισόδου. Η «σάρωση» αυτή συμπεριλαμβάνει δύο βασικά χαρακτηριστικά:

- Stride (άλμα): Αφορά το βήμα σάρωσης της εικόνας – άλμα που συμβαίνει προκειμένου να εκτελεστεί η συνέλιξη του επόμενου πεδίου υποδοχής με το συνελικτικό φίλτρο. Για παράδειγμα για stride = 1 το πεδίο υποδοχής κυλιέται κατά μία σειρά στοιχείων (είτε δεξιά - είτε κάτω), για stride = 2 το πεδίο υποδοχής μεταφέρεται κατά δύο σειρές στοιχείων (είτε δεξιά - είτε κάτω) κι ούτω καθεξής.
- Padding (γέμισμα): Τα στοιχεία της εισόδου που βρίσκονται στις άκρες συμμετέχουν σε λιγότερες συνελίξεις συγκριτικά με τα ενδιάμεσα. Εξαιτίας αυτού και προκειμένου να μην χαθεί η απεικονιστική πληροφορία των στοιχείων αυτών, χρησιμοποιείται το τέχνασμα του padding, το οποίο επεκτείνει το μέγεθος του δεδομένου που τίθεται υπό επεξεργασία, με επιπρόσθετες τιμές περιμετρικά αυτού (οι τιμές αυτές είναι συνήθως 0, το οποίο αποτελεί το λεγόμενο “Zero Padding”). Η λειτουργία του Padding παρέχει επίσης δυνατότητα ελέγχου του μεγέθους της εξόδου, έπειτα από την ολοκλήρωση της σάρωσης.



Σχήμα 1.2.2 - 4: Απεικόνιση του τρόπου με τον οποίο «σαρώνεται» η είσοδος, για τιμές του stride ίσες με 1 και 2. Ο όρος Feature Map (χαρτογράφηση χαρακτηριστικών) αντιστοιχίζεται στην έξοδο.

Πηγή: Oreilly.com

1.2.3 Ογκική - Σε Βάθος Συνελικτική Επεξεργασία

Παραπάνω, επεξηγήθηκε η διαδικασία της Συνέλιξης για περιπτώσεις όπου η είσοδος αποτελείται από δύο διαστάσεις (μήκος & πλάτος). Σε περιπτώσεις όπου η είσοδος είναι τρισδιάστατη (η πιο συνήθης περίπτωση), τα φίλτρα έχουν το ίδιο βάθος με την είσοδο, και η διαδικασία των συνεχών συνελίξεων για την σάρωση της εισόδου, όπως αυτή περιγράφηκε, λαμβάνει χώρα ανά κανάλι.

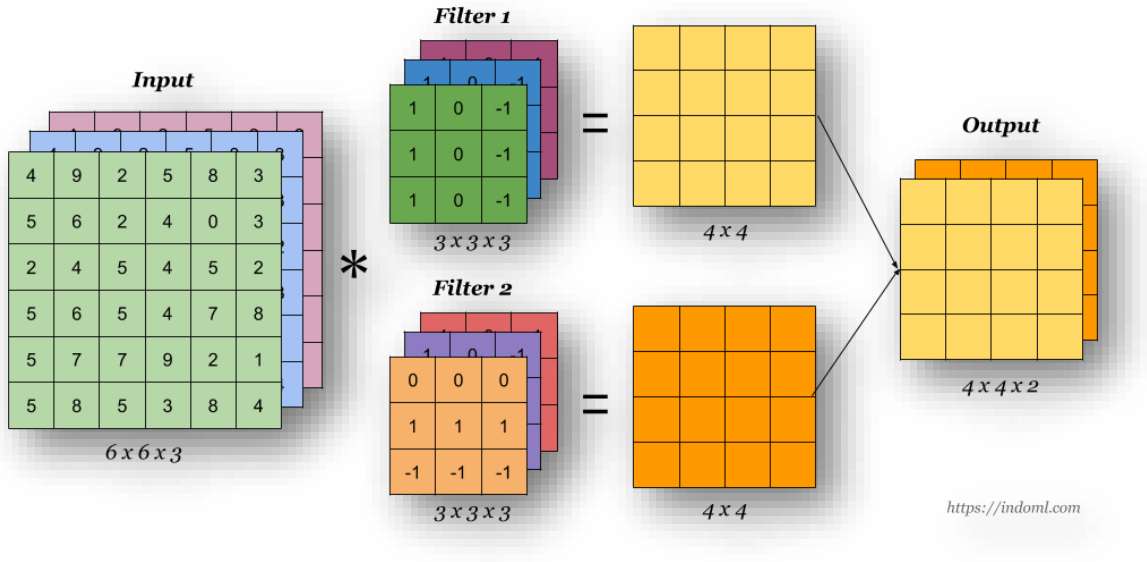
Παραδείγματος χάριν, για μία είσοδο-εικόνα με διαστάσεις 256 x 256 x 3, τα φίλτρα θα αποτελούνται κι αυτά από 3 κανάλια, και η διαδικασία εφαρμόζεται για το πρώτο κανάλι του φίλτρου με το πρώτο κανάλι της εισόδου, για το δεύτερο κανάλι του φίλτρου με το δεύτερο κανάλι της εισόδου και αντίστοιχα και για το τρίτο. Έτσι, έχουν προκύψει τρεις πίνακες - αποτελέσματα από την διαδικασία σάρωσης συνελίξεων, ένας για κάθε κανάλι. Στην συνέχεια, οι τρεις αυτοί πίνακες προστίθενται ανά στοιχείο δίνοντας τα παραγόμενα στοιχεία τα οποία συνιστούν το αποτέλεσμα της λειτουργίας της συνέλιξης μεταξύ της εισόδου και του εκάστοτε τρισδιάστατου φίλτρου.

Σε ένα συνελικτικό επίπεδο, μπορούν να χρησιμοποιούνται πολλά τέτοια φίλτρα. Ο αριθμός των φίλτρων που χρησιμοποιούνται σε κάθε επίπεδο, αντιστοιχίζεται στο βάθος του ογκικού αποτελέσματος[8] [9] που θα προκύψει έπειτα από την συνελικτική επεξεργασία της εισόδου του. Τα φίλτρα, είναι για τα δίκτυα ένας

[8] Οι άλλες δύο διαστάσεις του αποτελέσματος που θα προκύψει μετά την επεξεργασία, εξαρτώνται από την δομή της λειτουργίας που υλοποιείται (συνέλιξη ή pooling) και τα χαρακτηριστικά αυτών (το μέγεθος της εισόδου, μέγεθος των φίλτρων-παραθύρων που χρησιμοποιούνται, την τιμή του stride, την ύπαρξη padding κτλ)

[9] Μία συνήθης λειτουργία είναι η χρήση ενός συνελικτικού φίλτρου του οποίου οι διαστάσεις είναι 1x1xN, όπου N ο αριθμός των καναλιών του εκάστοτε δεδομένου που πρόκειται να επεξεργαστεί. Ουσιαστικά η λειτουργία αυτή, επιτελεί συνέλιξη σε όλα τα στοιχεία του δεδομένου και κατά βάθος (στα κανάλια), και η χρησιμότητα αυτής είναι η δυνατότητα ρύθμισης του βάθους της εξόδου που προκύπτει από το εκάστοτε συνελικτικό επίπεδο (όσα φίλτρα 1x1 χρησιμοποιηθούν, τόσα θα είναι και τα κανάλια της αντίστοιχη εξόδου). Χρησιμοποιείται στην Βαθιά Μάθηση για την ελάττωση των καναλιών και συνεπώς την μείωση των παραμέτρων του δικτύου.

τρόπος ανίχνευσης χρήσιμης πληροφορίας ή χαρακτηριστικών της εκάστοτε εισόδου, με σκοπό την ορθή τελική πρόβλεψη - κατηγοριοποίηση της εικόνας.



Σχήμα 1.2.3 - 1: Απεικόνιση του παραχθέντος αποτελέσματος, από την πράξη της συνέλιξης μεταξύ ενός δεδομένου τριών καναλιών και δύο συνελικτικών τρισδιάστατων φίλτρων - πυρήνων (ο αριθμός των καναλιών των φίλτρων είναι ο ίδιος με τον αριθμό των καναλιών της εισόδου). Το αποτέλεσμα της εξόδου περιλαμβάνει δύο κανάλια (βάθος = 2), το καθένα από τα οποία αποτελεί το αποτέλεσμα της συνέλιξης μεταξύ της εισόδου και του εκάστοτε φίλτρου όπως αυτό περιγράφηκε. Δηλαδή, το βάθος της εξόδου του παραχθέντος αποτελέσματος θα είναι ίσο με τον αριθμό των φίλτρων, τα οποία συμπεριλαμβάνονται στο εκάστοτε συνελικτικό επίπεδο.

Πηγή: indoml.com

1.2.4 Λειτουργία Επιπέδου Pooling

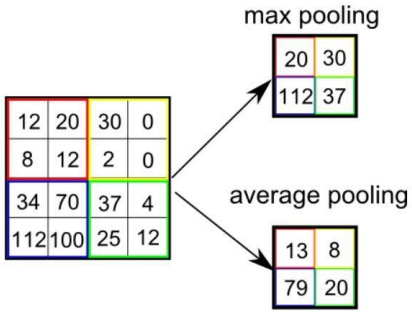
Η διαδικασία του Pooling, συνήθως λαμβάνει χώρα μετά τα επίπεδα συνέλιξης. Πρόκειται για μηχανισμό, ο οποίος πραγματοποιείται με πολύ παρόμοιο τρόπο με αυτόν των επίπεδων συνέλιξης, αλλά η λειτουργία διαφέρει. Δουλειά των επιπέδων Pooling είναι να απλοποιούν την πληροφορία των αποτελεσμάτων της συνέλιξης (Feature Maps), επιφέροντας μία δραστική περικοπή-μείωση στις διαστάσεις του εκάστοτε αποτελέσματος που έχει προκύψει από την συνελικτική λειτουργία, συντηρώντας όμως την χρήσιμη πληροφορία (Downsampling).

Έτσι, το υπολογιστικό κόστος επεξεργασίας των δεδομένων γίνεται μικρότερο, μειώνονται οι πιθανότητες του φαινομένου Overfitting και επέρχονται αποτελέσματα καλύτερης γενίκευσης από τα συνελικτικά μοντέλα. Η λογική βασίζεται στην ιδιότητα των συνελικτικών φίλτρων τα οποία εντοπίζουν χαρακτηριστικά της εικόνας. Εφόσον ένα χαρακτηριστικό έχει βρεθεί, έχει μεγαλύτερη σημασία η αλληλουχία και η συνύπαρξη με τα υπόλοιπα χαρακτηριστικά που εξετάζονται από τα φίλτρα, παρά η ακριβής τοποθεσία αυτού.

Η διαδικασία μείωσης συμβαίνει μέσω της σάρωσης της εξόδου, η οποία προήλθε από το συνελικτικό επίπεδο, από ένα φίλτρο-παράθυρο του οποίου το μέγεθος ποικίλει όπως ακριβώς συμβαίνει και στα επίπεδα συνέλιξης. Βασική διαφορά είναι ότι το παράθυρο δεν εμπεριέχει παραμέτρους προς εκμάθηση, αλλά υλοποιεί μία συνάρτηση βασισμένη στα στοιχεία του αντίστοιχου σημείου υποδοχής (δηλαδή, της

εκάστοτε εξεταζόμενης περιοχής). Οι πιο συνήθεις συναρτήσεις - μηχανισμοί μείωσης των διαστάσεων είναι δύο:

- Max Pooling: Αντιστοιχίζεται το στοιχείο με την μέγιστη τιμή της εξεταζόμενης περιοχής.
- Average Pooling: Αντιστοιχίζεται ο μέσος όρος όλων των τιμών που εμπεριέχονται στο εκάστοτε πεδίο.



Σχήμα 1.2.4 - 1: Απεικόνιση των δύο λειτουργιών Max Pooling και Average Pooling, σε ένα δεδομένο επεξεργασίας 4x4 για ένα Pooling παράθυρο 2x2.

Πηγή: Towardsdatascience

Για τα τρισδιάστατα δεδομένα, η διαδικασία λαμβάνει χώρα ανά κανάλι όπως ακριβώς και στην περίπτωση των συνελικτικών επιπέδων. Κατά κύριο λόγο έχει εδραιωθεί ο μηχανισμός Max Pooling, καθώς έχει επιδείξει καλύτερα πειραματικά αποτελέσματα.

1.2.5 Λειτουργία Batch Normalization

Προκειμένου να αναλυθεί με σαφήνεια η λειτουργία Batch Normalization, κρίθηκε σκόπιμη και η επεξήγηση της Κανονικοποίησης Εισόδων (Data Normalization), από την οποία αυτή είναι εμπνευσμένη αλλά και της οποίας αποτελεί γενίκευση για όλο το βάθος των δικτύων.

1.2.5.1 Κανονικοποίηση Εισόδων-Δεδομένων (Data Normalization)

Αφορά τεχνική η οποία μετατρέπει την κατανομή των δεδομένων στην Κανονική - Γκαουσιανή Κατανομή, δηλαδή σε Κατανομή με μέσο όρο 0 και τυπική απόκλιση 1. Αυτό επιτυγχάνεται αφαιρώντας από κάθε δεδομένο εισόδου τον μέσο όρο των δεδομένων και διαιρώντας με την τυπική απόκλιση αυτών, δηλαδή:

$$x_i' = \frac{x_i - m}{\sigma}$$

Όπου:

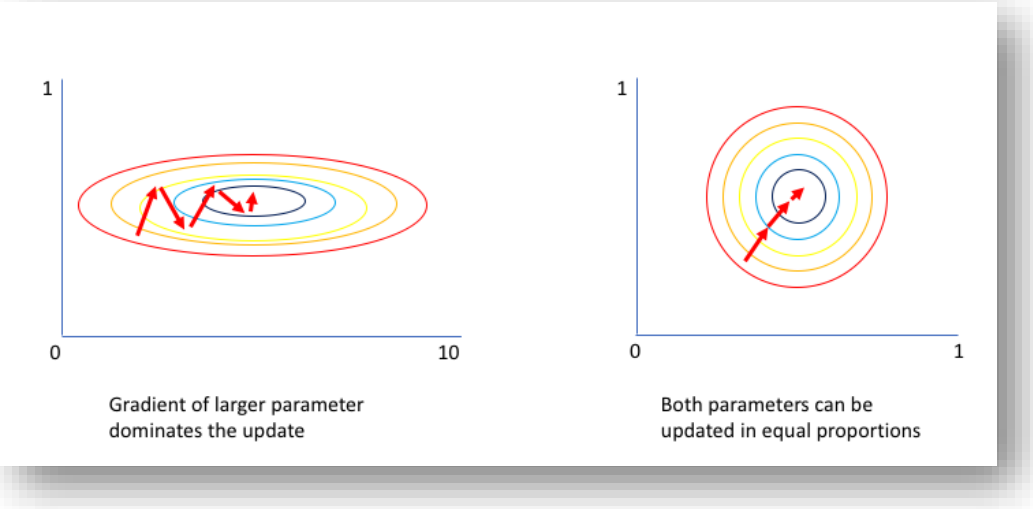
x_i' : το νέο τροποποιημένο δεδομένο,

x_i: το αρχικό δεδομένο εισόδου,

m: ο μέσος όρος της κατανομής όλων των δεδομένων εισόδου,

σ: η τυπική απόκλιση της κατανομής των δεδομένων εισόδου.

Ο λόγος που καθιστά αυτήν την τεχνική πάρα πολύ χρήσιμη, είναι ότι στις περιπτώσεις όπου η κατανομή είναι ανισόρροπη, το δίκτυο καλείται να δημιουργήσει επίσης ανισόρροπες παραμέτρους προκειμένου να αντισταθμίσει την ανισορροπία της κατανομής. Αυτό, έχει σαν αποτέλεσμα ο Αλγόριθμος Βελτιστοποίησης να επηρεάζεται περισσότερο από κάποιες παραμέτρους, καθυστερώντας την σύγκλιση προς το ζητούμενο ελάχιστο σημείο της Συνάρτησης Κόστους. Η Κανονικοποίηση των τιμών, εξομαλύνει σε μεγάλο βαθμό αυτό το πρόβλημα.



Σχήμα 1.2.5.1 - 1: Απεικόνιση του αποπροσανατολισμού του Αλγορίθμου Βελτιστοποίησης εξαιτίας της έλλειψης ισορροπίας. Έπειτα από την κανονικοποίηση η σύγκλιση είναι ταχύτερη.

Πηγή: jeremyjordan.me

1.2.5.2 Κανονικοποίηση Παρτίδων (Batch Normalization)

Η Κανονικοποίηση Δεδομένων-Εισόδων συντελεί στην ισορροπία των βαρών των νευρώνων του πρώτου στρώματος του δικτύου. Στην Κανονικοποίηση Παρτίδων[10] ακολουθείται ακριβώς η ίδια λογική με την Κανονικοποίηση των Δεδομένων, μόνο που σε αυτήν την περίπτωση προβλέπεται η ισορροπία όλων των βαρών και για τα υπόλοιπα στρώματα. Αυτό συμβαίνει, καθώς κανονικοποιούνται οι έξοδοι των νευρώνων, οι οποίες πρόκειται να χρησιμοποιηθούν ως είσοδοι στο επόμενο στρώμα, διαδικασία η οποία τελείται καθ' όλο το βάθος του δικτύου.

Η Κανονικοποίηση Παρτίδων δεν μετατρέπει τις εξόδους των νευρώνων πάντα σε μία αμιγώς Γκαουσιανή Κατανομή, αλλά σε μία υβριδική μορφή αυτής. Είναι δομημένη με τέτοιον τρόπο, ώστε η υβριδική κατανομή στην οποία μετατρέπει τις εξόδους ενός στρώματος νευρώνων, να είναι αυτήν για την οποία βελτιστοποιείται η ταχύτητα σύγκλισης και συνεπώς ελαχιστοποιείται ο χρόνος εκπαίδευσης.

Συγκεκριμένα, για όλες τις εξόδους των νευρώνων των κρυφών στρωμάτων ενός δικτύου ακολουθείται η εξής διαδικασία,

$$h_{ij}^{norm} = \frac{h_{ij} - m_j}{\sigma_j}$$

Όπου,

h_{ij}^{norm} : η κανονικοποιημένη έξοδος του i-στού νευρώνα του j-στού στρώματος,

h_{ij} : η μη-κανονικοποιημένη έξοδος του i-στού νευρώνα του j-στού στρώματος,

m_j : ο μέσος όρος της κατανομής των εξόδων όλων των νευρώνων του j-στού στρώματος,

σ_j : η τυπική απόκλιση της κατανομής των εξόδων όλων των νευρώνων του j-στού στρώματος,

[10] Το όνομα Batch, προκύπτει από το γεγονός ότι ο μέσος όρος m_j και η τυπική απόκλιση σ_j που χρησιμοποιούνται για την κανονικοποίηση, δεν προκύπτουν από τις εξόδους των στρωμάτων όλων των δεδομένων, αλλά από τις εξόδους των στρωμάτων της εκάστοτε παρτίδας δεδομένων.

και στην συνέχεια,

$$h_{ij}^{final} = \gamma * h_{ij}^{norm} + \beta$$

Όπου,

h_{ij}^{final} : η τελική-υβριδική κανονικοποιημένη έξοδος του i-στού νευρώνα του j-στού στρώματος

h_{ij}^{norm} : η κανονικοποιημένη έξοδος του i-στού νευρώνα του j-στού στρώματος του παραπάνω σταδίου,

γ, β : προστιθέμενες παράμετροι της Τεχνικής Κανονικοποίησης Παρτιδών οι οποίες μαθαίνονται από το δίκτυο[11].

Πέρα από την ταχύτητα σύγκλισης, η οποία βελτιώνεται εμφατικά, ειδικότερα για περιπτώσεις σύνθετων προβλημάτων όπου απαιτείται ένας μεγάλος αριθμός εποχών εκπαίδευσης αλλά και πάρα πολλά δεδομένα, μερικά ακόμη βασικά πλεονεκτήματα της Κανονικοποίησης είναι τα εξής:

- Αποτρέπεται το πρόβλημα των «εξομαλυμένων παραγώγων» (Vanishing Gradient Problem).
- Ελέγχεται αποτελεσματικά η μη-ικανοποιητική αρχικοποίηση βαρών.
- Μειώνονται οι πιθανότητες για Overfitting, καθώς η κανονικοποίηση των τιμών των εισόδων σε κάθε επίπεδο, επιτελεί και ρόλο ομαλοποίησης (Regularization Effect).
- Είναι εύκολο να ενσωματωθεί αλλά και να αφαιρεθεί προγραμματιστικά από τα δίκτυα, καθώς χρησιμοποιείται ως συνιστώσα αυτών κατά την δημιουργία τους.

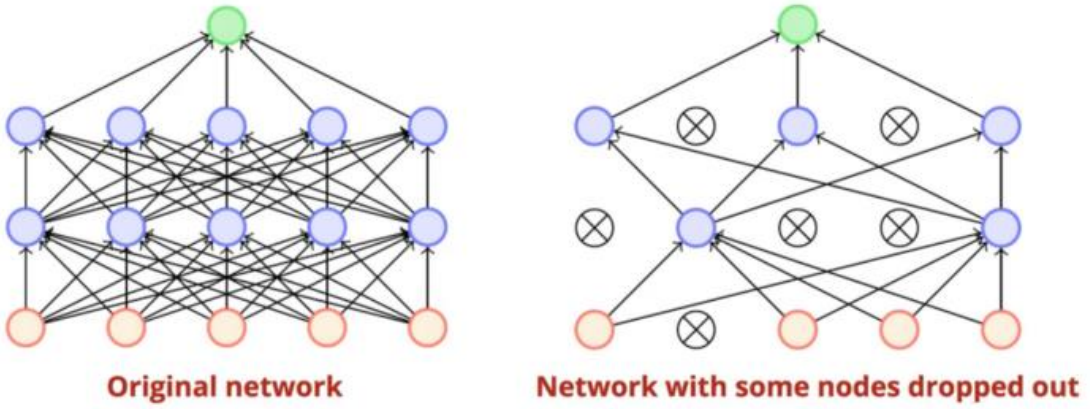
1.2.6 Λειτουργία Επιπέδου Dropout

Κατά την διάρκεια μία εποχής της εκπαίδευσης «πετιούνται» από το δίκτυο κάποιοι νευρώνες, δηλαδή είναι σαν να απενεργοποιούνται αυτοί και όλες οι συνδέσεις τους. Το δίκτυο αναγκάζεται να προσαρμόζει τις παραμέτρους του παρά την απώλεια κάποιων δομικών συνιστωσών του, προσδίδοντας έναν βαθμό ανεξαρτησίας μεταξύ των νευρώνων.

Σε κάθε εποχή, απενεργοποιούνται και διαφορετικοί νευρώνες[12] και έτσι είναι σαν κάθε φορά να εκπαιδεύονται διαφορετικά δίκτυα. Κατά την διαδικασία της αξιολόγησης λαμβάνεται υπόψη ο μέσος όρος αυτών των δικτύων καθώς χρησιμοποιούνται όλοι οι νευρώνες. Πρόκειται για μία πολύ χρησιμοποιούμενη τακτική, η οποία επιφέρει σημαντικές βελτιώσεις και συμβάλλει αισθητά στην αποφυγή φαινομένων «υπερμοντελοποίησης» (Overfitting).

[11] Σε αντίθεση με τις τιμές του μέσου όρου m_j και της τυπικής απόκλισης σ_j οι οποίες υπολογίζονται ανά παρτίδα, οι προστιθέμενες παράμετροι γ,β είναι γενικές και μαθαίνονται από όλες τις παρτίδες δεδομένων.

[12] Το αν θα παραμείνει ενεργοποιημένος ένας νευρώνας την εκάστοτε εποχή, εξαρτάται από μία πιθανότητα *p*, η οποία ορίζεται από τον προγραμματιστή. Συνεπώς, η πιθανότητα απενεργοποίησης νευρώνα ισούται με *1-p*.



Σχήμα 1.2.6 - 1: Απεικόνιση της απενεργοποίησης κάποιων νευρώνων του δικτύου για μία εποχή, κατά την διάρκεια της εκπαίδευσης σύμφωνα με την τεχνική Dropout.

Πηγή: Towardsdatascience

1.2.7 Χρήσιμα Συμπεράσματα: Ιδιότητες και Πλεονεκτήματα

Παρακάτω παρατίθενται χρήσιμα συμπεράσματα για τα Συνελικτικά Νευρωνικά Δίκτυα, μερικά βασικά πλεονεκτήματα αλλά και η πρόκληση, με βάση την λειτουργία τους και τον τρόπο που είναι δομημένα:

(1) Μπορούμε να πούμε πως κατ' αντιστοιχία με τον νευρώνα ενός κλασικού Νευρωνικού Δικτύου, η πράξη της συνέλιξης αντικαθιστά τον όρο wx του γραμμικού μετασχηματισμού wx+b και η είσοδος x_i σε έναν νευρώνα είναι για τα συνελικτικά δίκτυα το εκάστοτε πεδίο υποδοχής που συνιστά μία ολόκληρη περιοχή της εικόνας.

(2) Το γεγονός αυτό επιφέρει μείωση στον αριθμό των παραμέτρων του δικτύου και δίνει την δυνατότητα στα Συνελικτικά Δίκτυα να εξετάζουν τις εικόνες ανά περιοχές και σύμφωνα με την δομή τους, τα στοιχεία που υπάρχουν μέσα αυτές αλλά και την αλληλουχία αυτών των στοιχείων, μπορεί να εξαχθεί ένα γενικότερο χαρακτηριστικό-συμπέρασμα για την συνολική εικόνα, όπως ακριβώς συμβαίνει και με την ανθρώπινη όραση. Το δίκτυο συλλέγει διάφορα τέτοια χαρακτηριστικά (και έπειτα από αυτά εξάγει κι άλλα ακόμη πιο ειδικά, από τα οποία εξάγει κι άλλα κι ούτω καθεξής – σύμφωνα με το βάθος του δικτύου) με σκοπό να τα αξιοποιήσει για την ακριβή κατηγοριοποίηση του περιεχομένου της εικόνας.

Ο τρόπος λειτουργίας αυτός καθιστά τα Συνελικτικά Νευρωνικά Δίκτυα ικανά να αναγνωρίζουν πιο σύνθετες απεικονίσεις, εκμεταλλευόμενα τις συσχετίσεις - εξαρτήσεις που υπάρχουν μεταξύ των γειτονικών στοιχείων της εκάστοτε εικόνας και ανιχνεύοντας με αυτόν τον τρόπο την πιθανή υπάρχουσα τοπική πληροφορία.

(3) Υπάρχει διαμοιρασμός παραμέτρων (Paremeter Sharing), καθώς επαναχρησιμοποιούνται τα ίδια φίλτρα για την ανάγνωση κατά μήκος και κατά πλάτος της εικόνας. Η κοινή χρήση των φίλτρων οδηγεί σε περαιτέρω μείωση του αριθμού των παραμέτρων. Η μείωση του αριθμού των παραμέτρων απαιτεί λιγότερη μνήμη και επιταχύνει την σύγκλιση προς το επιθυμητό σημείο κατά την διαδικασία της εκπαίδευσης.

(4) Πρόκληση για τα Συνελικτικά Νευρωνικά Δίκτυα αποτελεί το γεγονός ότι έρχονται αντιμέτωπα με το πρόβλημα των «Εξαφανιζόμενων και Εκτινασσόμενων Παραγώγων» (Vanishing and Exploding Gradients).

1.3 Αυτοκωδικοποιητές (Autoencoders)

Οι Autoencoders (Αυτοκωδικοποιητές), είναι ένα είδος δικτύου το οποίο είναι σχεδιασμένο με τέτοιον τρόπο ώστε να αναπαράγει στην έξοδο του την είσοδο που δέχεται. Βασικό σκοπό τους, αποτελεί η εκμάθηση και η ποιοτική δημιουργία μίας «συμπιεσμένης» αναπαράστασης (latent representation) των δεδομένων.

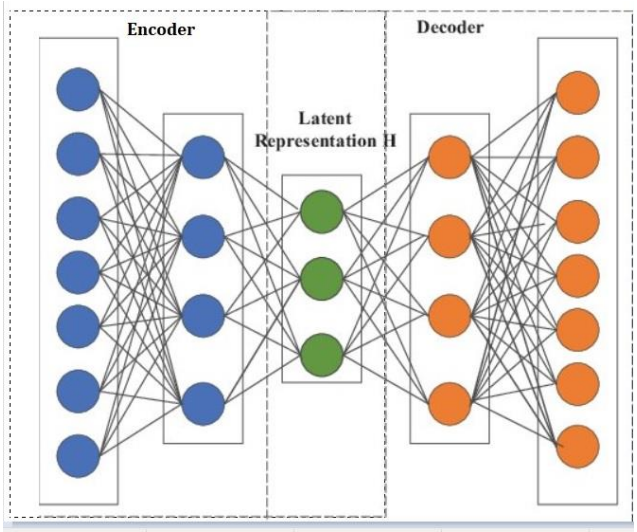
Περιπτώσεις όπου βρίσκουν εφαρμογή οι Αυτοκωδικοποιητές είναι:

- Ως παραγωγικά μοντέλα (Generative Models),
- Σε διεργασίες κατηγοριοποίησης (Classification Problems),
- Σε προβλήματα Clustering («Συσταδοποίηση»),
- Σε περιπτώσεις εντοπισμού ανωμαλιών των δεδομένων (Anomaly Detection),
- Στα συστήματα προτάσεων (Recommendation Systems),
- Σε περιπτώσεις όπου είναι απαραίτητη ή μπορεί να αξιοποιηθεί η μείωση των διαστάσεων (Dimensionality Reduction).

1.3.1 Δομή και Λειτουργία

Ένας Αυτοκωδικοποιητής απαρτίζεται από τα εξής δομικά στοιχεία:

- Το δίκτυο του Κωδικοποιητή (Encoder), το οποίο κωδικοποιεί την είσοδο, σε μία συμπιεσμένη και ουσιώδη αναπαράσταση (latent representation),
- Το ενδιάμεσο στρώμα Bottleneck (ή latent layer) το οποίο περιλαμβάνει την κωδικοποιημένη αναπαράσταση και εμπεριέχεται συγχρόνως και στο δίκτυο Κωδικοποιητή και στο δίκτυο Αποκωδικοποιητή, και
- Το δίκτυο του Αποκωδικοποιητή (Decoder), το οποίο αναλαμβάνει την ανακατασκευή της εισόδου μέσα από την κωδικοποιημένη αναπαράσταση που αποδίδει το ενδιάμεσο στρώμα αποσκοπώντας στην όσο το δυνατόν ποιοτικότερη ανάκτηση.[13]



Σχήμα 1.3.1 - 1: Γενική απεικόνιση της δομής ενός Αυτοκωδικοποιητή.

Πηγή: Researchgate – uploaded by Hieu Mac

[13] Τα δίκτυα του Κωδικοποιητή και του Αποκωδικοποιητή αποτελούνται κατά κύριο λόγο από Τεχνητά Νευρωνικά Δίκτυα, όπου στο δίκτυο του Αποκωδικοποιητή λαμβάνουν χώρα οι αντίστροφες λειτουργίες από αυτές του δικτύου του Κωδικοποιητή. Σε περιπτώσεις προβλημάτων Μηχανικής Όρασης τα δίκτυα αυτά μπορούν να είναι Συνελικτικά.

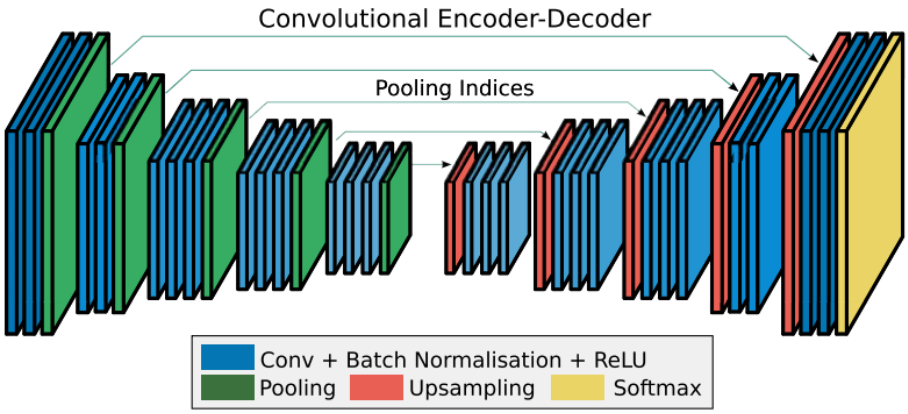
Ένας επίσημος ορισμός για την λειτουργία των Αυτοκωδικοποιητών, δόθηκε από τον Baldi το 2012, στην εργασία "Autoencoders, unsupervised learning, and deep architectures" [85], όπου έθεσε το πρόβλημα ως την εκμάθηση δύο συναρτήσεων $A: R^n \rightarrow R^p$ (κωδικοποιητής) και $B: R^p \rightarrow R^n$ (αποκωδικοποιητής), έτσι ώστε να ελαχιστοποιείται ο παράγοντας

$$E[\, \Delta(x, B \circ A(x)\,)\,]$$

Όπου,

- x: οι είσοδοι,
- E: η προσδοκία όσον αφορά την κατανομή των εισόδων x,
- $B \circ A(x)$: το αποτέλεσμα του Αυτοκωδικοποιητή έπειτα από την κωδικοποίηση και την αποκωδικοποίηση των εισόδων μέσα από τις συναρτήσεις A, B,
- $\Delta(x, B \circ A(x))$: η Απώλεια Ανακατασκευής (Reconstruction Loss).[14]

Η πρόκληση των εν λόγω μοντέλων είναι να διακρίνονται από επαρκή χωρητικότητα, η οποία θα τους προσδίδει την ικανότητα ποιοτικής αναδημιουργίας της εισόδου, και συγχρόνως το μέγεθος του ενδιάμεσου στρώματος να είναι όσο το δυνατόν μικρότερο χωρίς να έχει απώλειες ουσιαστικής πληροφορίας.



Σχήμα 1.3.1 - 2: Απεικόνιση της δομής ενός Συνελικτικού Αυτοκωδικοποιητή.

Πηγή: SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation - Vijay Badrinarayanan, Alex Kendall, Roberto Cipolla [92]

1.3.2 Είδη Αυτοκωδικοποιητών

Σύμφωνα με την εργασία "Autoencoders" των Dor Bank, Noam Koenigstein και Raja Giryes [86], οι Αυτοκωδικοποιητές μπορούν να διακριθούν στους:

1. Ομαλοποιημένους Αυτοκωδικοποιητές (Regularized Autoencoders), και στους
2. Στοχαστικούς Αυτοκωδικοποιητές (Variational Autoencoders-VAE) [63].

1.3.2.1 Ομαλοποιημένοι Αυτοκωδικοποιητές (Regularized Autoencoders)

[14] Συνήθως για την απώλεια Ανακατασκευής χρησιμοποιείται η μετρική L2-Norm.

Δημιουργήθηκαν για τις περιπτώσεις όπου το πλήθος των νευρώνων του ενδιάμεσου στρώματος ήταν μεγαλύτερο από το συνολικό μέγεθος των εικόνων εισόδου. Σε αυτές τις περιπτώσεις ελλοχεύει ο κίνδυνος της εμφάνισης του φαινομένου Overfitting, και της αποτύπωσης της «ταυτοτικής» συνάρτησης (identity function) από τον Encoder. Για την αποφυγή αυτού, χωρίς να είναι απαραίτητη η καταφυγή σε ενδιάμεσα στρώματα μικρότερων διαστάσεων και η δημιουργία ενός Bottleneck επιπέδου, χρησιμοποιήθηκαν τακτικές ομαλοποίησης (regularization methods) οι οποίες αναγκάζουν τον Αυτοκωδικοποιητή να μαθαίνει διαφορετικές αναπαραστάσεις από τις εισόδους. Οι Ομαλοποιημένοι Αυτοκωδικοποιητές διακρίνονται στις παρακάτω περιπτώσεις.

1. Αραιοί Αυτοκωδικοποιητές (Sparse Autoencoders)

Στο συγκεκριμένο είδος χρησιμοποιούνται δύο τρόποι ομαλοποίησης.

- Ο 1ος είναι η προσθήκη L1-Ομαλοποίησης (L1-Regularization) για τις ενεργοποιήσεις των νευρώνων του ενδιάμεσου στρώματος, δηλαδή ο ζητούμενος όρος ελαχιστοποίησης διαμορφώνεται σε

$$E[\Delta(x, B \circ A(x)] + \lambda \sum_i a_i$$

Όπου,

  - λ: ο παράγοντας της L1 ομαλοποίησης,
  - a_i: οι ενεργοποιήσεις των νευρώνων του ενδιάμεσου στρώματος (latent layer).

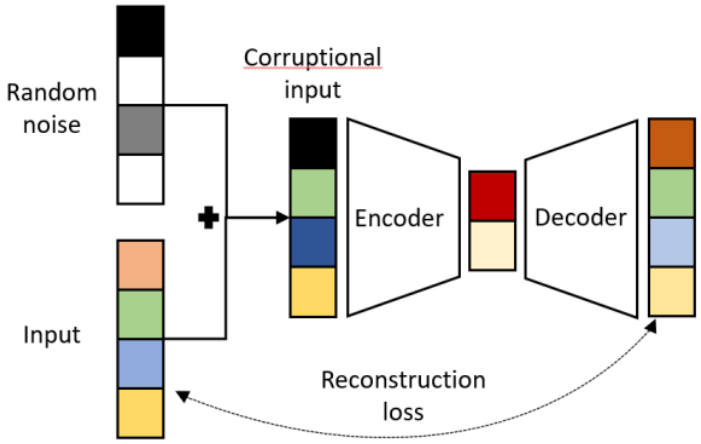
- Ο 2ος είναι η προσθήκη ομαλοποίησης μέσω της απόστασης KL-divergence, η οποία υποκαθιστά την L1 της προηγούμενης περίπτωσης. Εδώ, θεωρείται ότι οι έξοδοι των νευρώνων του ενδιάμεσου στρώματος ανήκουν σε μία κατανομή Bernouli p, σε κάθε Batch μετριέται η εμπειρική κατανομή των εξόδων ως $\hat{p}_j = \frac{1}{m}\sum_i a_i(x)$ (όπου m: το μέγεθος του Batch) , και ο αντίστοιχος παράγοντας ομαλοποίησης προκύπτει μέσω της απόστασης KL-divergence αυτών. Δηλαδή, ο ανάλογος όρος γίνεται:

$$E[\Delta(x, B \circ A(x)] + \sum_j KL(p||\hat{p}_j)$$

Όπου, το άθροισμα $\sum_j KL(p||\hat{p}_j)$ αφορά την παραπάνω απόσταση για όλους τους νευρώνες του ενδιάμεσου στρώματος latent.

2. Αυτοκωδικοποιητές Εξάλειψης Θορύβου (Denoising Autoencoders)

Πρόκειται για εύρωστους Αυτοκωδικοποιητές ομαλοποίησης, οι οποίοι χρησιμοποιούνται για την απομάκρυνση του θορύβου. Στην περίπτωση αυτήν, προστίθεται θόρυβος στις εισόδους που δίνονται στον Αυτοκωδικοποιητή, ο οποίος εκπαιδεύεται στην ανάκτηση των πρωτότυπων εισόδων έχοντας στην διάθεση του μόνο τις αλλοιωμένες είσοδούς.



Σχήμα 1.3.2.1 - 1: Γενική απεικόνιση της χρήσης ενός Denoising Autoencoder.

Πηγή: Autoencoders - Dor Bank, Noam Koenigstein, Raja Giryes [86]

3. «Αντιθετικοί» Αυτοκωδικοποιητές (Contractive Autoencoders)

Σε αντίθεση με τους Denoising Autoencoders, όπου η έμφαση δίνεται στο να παραβλέπονται οι τροποποιήσεις της εισόδου (θόρυβος), στους «Αντιθετικούς» Αυτοκωδικοποιητές επιδιώκεται να υπάρχει ευαισθησία όσον αφορά τις τροποποιήσεις οι οποίες ενδέχεται να ενέχουν σημαντικό ρόλο στην ανακατασκευή. Για τον σκοπό αυτόν, προστίθεται στο συνολικό κόστος η μετρική L2-norm του Ιακωβιανού πίνακα που αφορά το ενδιάμεσο στρώμα του Αυτοκωδικοποιητή. Ο Ιακωβιανός πίνακας περιλαμβάνει την παράγωγο κάθε νευρώνα h_j για κάθε είσοδο x_i της x, δηλαδή, $J_{ji} = \nabla_{x_i} h_j(x_i)$. Έτσι, ο όρος ελαχιστοποίησης διαμορφώνεται σε:

$$E[\Delta(x, B \circ A(x)] + \lambda \, \|J_A(x)\|_2^2$$

Όπου,

- λ: ο παράγοντας της L2 ομαλοποίησης,
- $\|J_A(x)\|_2^2$: η L2-norm του Ιακωβιανού πίνακα του ενδιάμεσου στρώματος του Αυτοκωδικοποιητή.

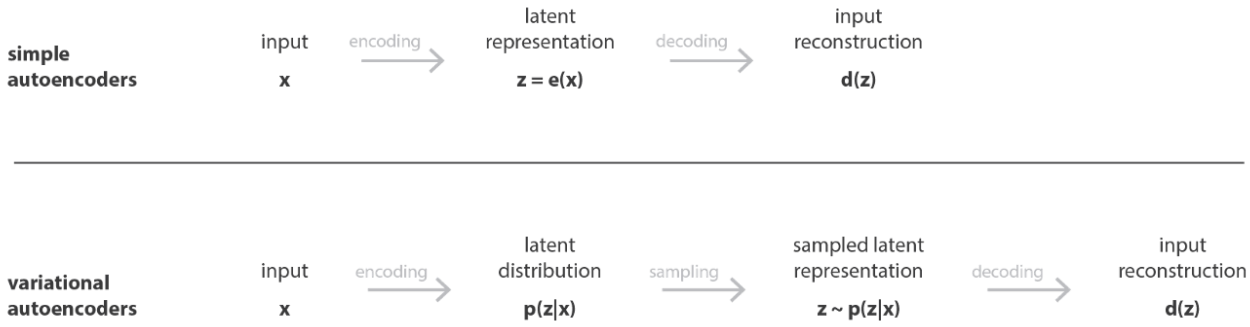
Η ελαχιστοποίηση του όρου της L2-norm του Ιακωβιανού πίνακα, αναγκάζει την τροποποίηση των παραμέτρων προς την αντίθετη κατεύθυνση από αυτήν της ελαχιστοποίησης του όρου ανακατασκευής, καθώς η ελαχιστοποίηση αυτού οδηγεί στις όσο το δυνατόν κοντινότερες σε ομοιότητα αναπαραστάσεις των εισόδων στο ενδιάμεσο στρώμα, γεγονός που δυσκολεύει την ανάκτηση από τον Decoder. Η ιδέα αυτής της υλοποίησης είναι, ότι οι περιττές ομοιότητες των εισόδων οι οποίες δεν είναι σημαντικές για την ανακατασκευή τους εξαφανίζονται από την επίδραση του όρου της L2-Norm του Ιακωβιανού πίνακα, ενώ οι απαραίτητες θα παραμείνουν χάρη στην επιρροή που έχουν στον όρο της ανακατασκευής.

1.3.2.2 Στοχαστικοί Αυτοκωδικοποιητές (Variational Autoencoders-VAEs)

Μία σημαντικότατη βελτίωση όσον αφορά τις δυνατότητες αναπαράστασης των Αυτοκωδικοποιητών επήλθε με τα μοντέλα Στοχαστικών Αυτοκωδικοποιητών [63]. Κατά την εκπαίδευση αυτών των μοντέλων, χρησιμοποιείται ισχυρή ομαλοποίηση ώστε να εξασφαλιστεί μία επαρκής κωδικοποίηση για την παραγωγή ακόμη και νέων δεδομένων (σε απλές περιπτώσεις Αυτοκωδικοποιητών, χρησιμοποιείται μόνο η απαραίτητη ομαλοποίηση προκειμένου αυτοί να ανταποκρίνονται στην ανακατασκευή των συγκεκριμένων δεδομένων). Ο όρος “Variational” προέρχεται από την συσχέτιση μεταξύ αυτής της ισχυρής ομαλοποίησης (Regularization) και του τομέα της «Στοχαστικής Συμπερασματολογίας» (Variational Inference).

1.3.2.2.1 Κωδικοποίηση

Ο τρόπος με τον οποίο επιτυγχάνεται η παραπάνω ισχυρή ομαλοποίηση επικεντρώνεται στην μέθοδο κωδικοποίησης που χρησιμοποιείται σε αυτό το είδος Αυτοκωδικοποιητών. Σε αυτήν την περίπτωση το ενδιάμεσο στρώμα δεν περιλαμβάνει μία κωδικοποίηση με πληροφορίες για τις εισόδους x, αλλά για την κατανομή αυτών p(x). Έτσι, η πληροφορία του latent επιπέδου περιλαμβάνει μία ενδιάμεση κατανομή p(z|x), στην οποία εμπεριέχονται οι αντίστοιχες κατανομές πιθανοτήτων για το κάθε χαρακτηριστικό (νευρώνα). Κατά την αποκωδικοποίηση, η ανακατασκευή γίνεται σύμφωνα με τις πιθανότητες αυτών των χαρακτηριστικών. Με τον τρόπο αυτόν δίνεται και στον αποκωδικοποιητή μία πλήρως ομαλοποιημένη είσοδος προς αποκωδικοποίηση. Έτσι, αναμένουμε οι είσοδοι που βρίσκονται κοντά σύμφωνα με την κατανομή τους, να έχουν και κοντινές αποκωδικοποιήσεις ανακατασκευής.



Σχήμα 1.3.2.2.1 - 1: Διαφορά του τρόπου λειτουργίας «απλών» και στοχαστικών Αυτοκωδικοποιητών.

Πηγή: Towardsdatascience

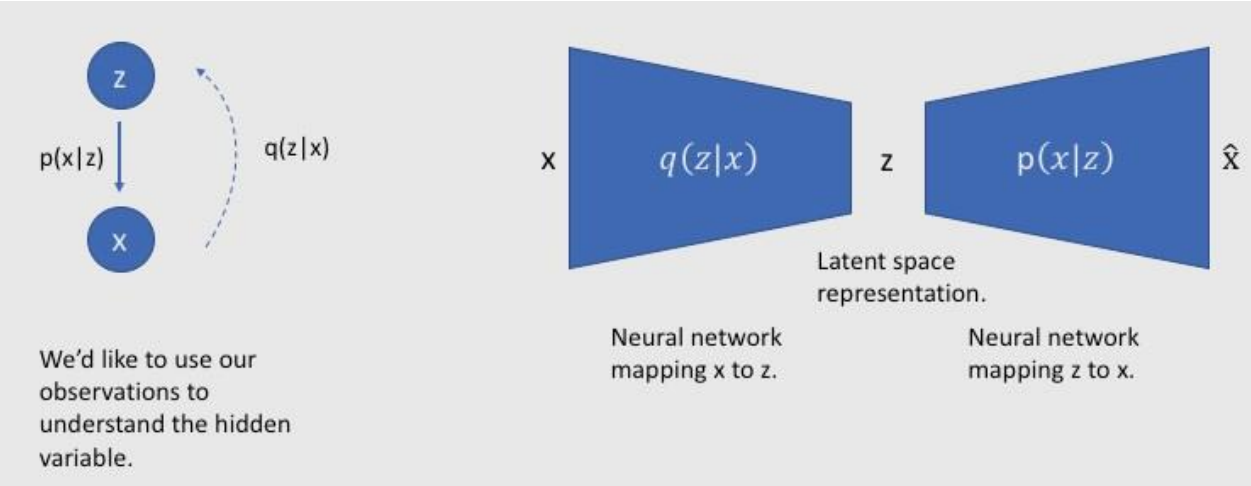
Για την κατανομή p(z|x) χρησιμοποιείται μία προσέγγιση q(z|x), καθώς σύμφωνα με την θεωρία του Bayes, προκύπτει ότι αφορά μη διαχειρίσιμη κατανομή. Για την μεγιστοποίηση της ομοιότητας των δύο αυτών κατανομών, αρκεί να ελαχιστοποιηθεί η μεταξύ τους απόσταση KL-divergence. Προκειμένου να ελαχιστοποιηθεί αυτή η απόκλιση, η απόσταση KL-divergence των δύο κατανομών συμπεριλαμβάνεται στο συνολικό κόστος.

Έτσι το συνολικό κόστος για τους VAE, διαμορφώνεται ως:

$$L(x,\hat{x}) + \sum_j KL(q_j(z|x)||p(z))$$

Όπου,

- $L(x,\hat{x})$: το κόστος Ανακατασκευής,
- $\sum_j KL(q_j(z|x)||p(z))$: η συνολική διαφορά μεταξύ της τεχνητής προσέγγισης q(z|x) και της πραγματικής p(z)[15] για κάθε χαρακτηριστικό-νευρώνα j του ενδιάμεσου στρώματος.



Σχήμα 1.3.2.2.1 - 2: Απεικόνιση της λογικής της λειτουργίας των VAE.

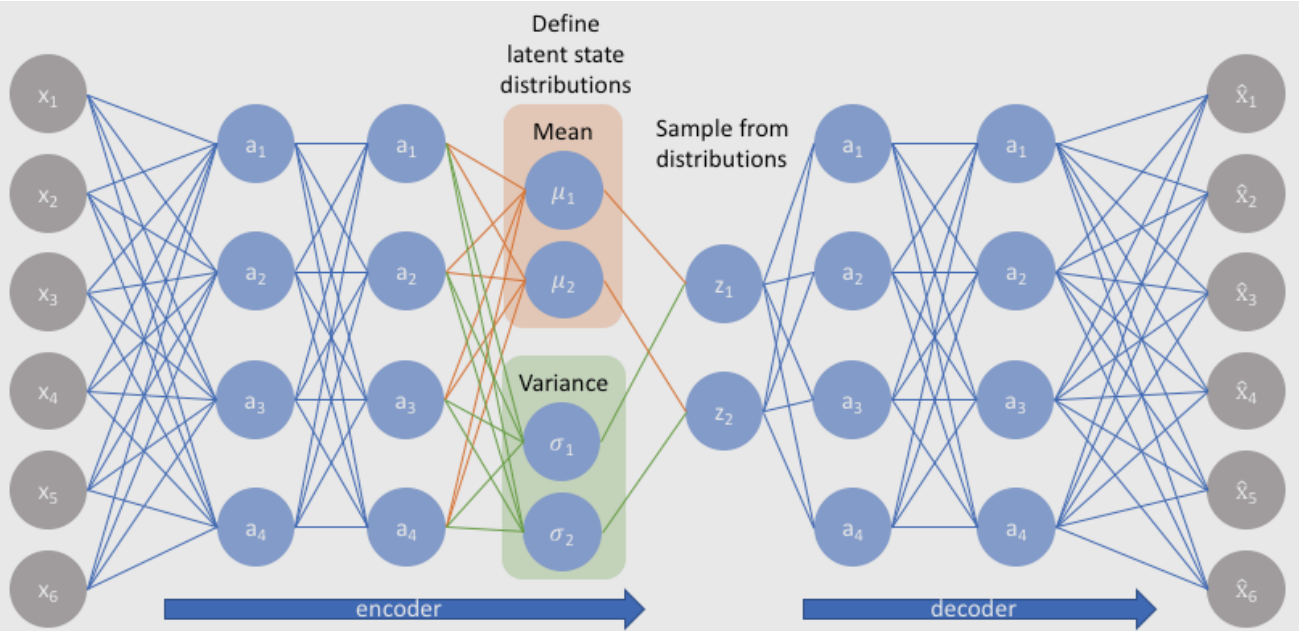
Πηγή: Jeremyjordan.me

[15] Συνήθως υποτίθεται ότι η πραγματική κατανομή p(z) ακολουθεί την Κανονική-Γκαουσιανή μορφή κατανομής.

1.3.2.2.2 Υλοποίηση

Προκειμένου το ενδιάμεσο στρώμα να εμπεριέχει μία κωδικοποίηση που αφορά πλέον κατανομή χαρακτηριστικών και όχι απλά χαρακτηριστικά δεδομένων, απαιτούνται οι αντίστοιχες ποσότητες περιγραφής αυτών. Καθώς υποτίθεται πως η κατανομή p(z) που επιδιώκεται να προσεγγιστεί ακολουθεί την Γκαουσιανή μορφή, οι απαραίτητες ποσότητες για την προσέγγιση q(z) αυτής είναι ο μέσος όρος και η τυπική απόκλιση για τις κατανομές του κάθε χαρακτηριστικού.

Για την αποκωδικοποίηση, λαμβάνεται από τον Decoder ένα δείγμα σύμφωνα με τις κατανομές των κωδικοποιημένων χαρακτηριστικών, ο οποίος πραγματοποιεί την τελική ανακατασκευή που δίνεται ως είσοδος στο Κόστος Ανακατασκευής προκειμένου να χρησιμοποιηθεί από τον Αλγόριθμο «Οπισθοδιάδοσης» (Backpropagation Algorithm) για την κατάλληλη τροποποίηση των παραμέτρων του μοντέλου.



Σχήμα 1.3.2.2.2 - 1: Απεικόνιση του τρόπου κωδικοποίησης μίας Γκαουσιανής-Κανονικής κατανομής.

Πηγή: Jeremyjordan.me

Τέλος, ο τρόπος δημιουργίας του δείγματος μέσα από τον αποκωδικοποιητή γίνεται μέσω ενός τεχνάσματος που ονομάζεται “reparameterization trick”, όπου επιλέγεται μέσα από μία Γκαουσιανή μονάδα μία ποσότητα ε, η οποία στην συνέχεια μετατοπίζεται σύμφωνα με τον μέσο όρο των κατανομών των χαρακτηριστικών του επιπέδου κωδικοποίησης και μετασχηματίζεται στην ζητούμενη κλίμακα σύμφωνα με την αντίστοιχη τυπική απόκλιση. Πιο συγκεκριμένα, το δείγμα z(x) που λαμβάνεται προς αποκωδικοποίηση προκύπτει ως

$$z(x) = m(x) + \sigma(x)\varepsilon \quad \mu\varepsilon \quad \varepsilon \sim N(0, I).$$

Όπου,

- m(x): ο μέσος όρος για κάθε χαρακτηριστικό του latent layer,
- σ(x): η τυπική απόκλιση για κάθε χαρακτηριστικό του latent layer, και
- ε: η επιλεγόμενη ποσότητα μέσα από μία Γκαουσιανή Μονάδα.

Κεφάλαιο 2ο

Μάθηση Λίγων Λήψεων (Few Shot Learning)

2.1 Η Έννοια του Few-Shot Learning

2.1.1 Τεχνικός Ορισμός & Γενική Αντιμετώπιση της Μάθησης Λίγων Λήψεων

2.1.1.1 Επιστημονικό Πεδίο Ένταξης και Τεχνικός Ορισμός

Το Few Shot Learning αφορά έναν συγκεκριμένο τομέα της ευρύτερης περιοχής της Μηχανικής Μάθησης. Στις εφαρμογές Μηχανικής Μάθησης, απαιτείται ένας μεγάλος αριθμός δεδομένων με εποπτευόμενη πληροφορία - κατηγοριοποιημένων δεδομένων προκειμένου να επιτευχθούν ικανοποιητικές επιδόσεις. Η απόκτηση αυτών των δεδομένων όμως, πολλές φορές καθίσταται δύσκολη ή ακόμη κι αδύνατη. Η μάθηση Λίγων λήψεων είναι μία ειδική περίπτωση Μηχανικής Μάθησης, η οποία αποσκοπεί στις υψηλές επιδόσεις για διεργασίες όπου η συλλογή δεδομένων εκπαίδευσης είναι πάρα πολύ μικρή.

Το 1997, ο Tom M. Mitchell εισήγαγε έναν τεχνικό ορισμό για την Μηχανική Μάθηση [54], ο οποίος είναι ο εξής:

> Μία μηχανή θεωρείται πως μαθαίνει από εμπειρία E που αφορά ένα σύνολο διεργασιών T και μέτρηση απόδοσης P, όταν η απόδοση στις διεργασίες T, όπως αυτή μετριέται από την απόδοση P βελτιώνεται με την εμπειρία E.

Κατ' αντιστοιχία με τον ορισμό του Tom M. Mitchell, ένας αντίστοιχος ορισμός δόθηκε και για το Few Shot Learning, από τους Yaqing Wang et al. στην εργασία "Generalizing from a Few Examples: A Survey on Few-Shot Learning" [55], ο οποίος είναι ο παρακάτω:

> Η Μάθηση Λίγων Λήψεων είναι ένας τύπος συγκεκριμένων προβλημάτων Μηχανικής Μάθησης (ορισμένα από μία εμπειρία E, για μία διεργασία T, με αντίστοιχη απόδοση P), όταν στην αντίστοιχη εμπειρία E, εμπεριέχεται ένας περιορισμένος αριθμός παραδειγμάτων με εποπτευόμενη πληροφορία (κατηγοριοποιημένων) , για την διεργασία T.

2.1.1.2 Η Χρήση Προγενέστερης Γνώσης Ως Αντιμετώπιση

Ο πιο συνήθης τρόπος, προκειμένου να αντιμετωπιστεί η δυσκολία των αλγορίθμων Μηχανικής Μάθησης, σε διεργασίες όπου η εμπειρία E περιέχει λίγα κατηγοριοποιημένα δεδομένα, είναι η επιδίωξη του

συνδυασμού των περιορισμένων δεδομένων της Ε, με μία «προγενέστερη γνώση» (Prior Knowledge) η οποία είναι χρήσιμη για την διεργασία Τ.

Με τον όρο προγενέστερη γνώση, αναφερόμαστε σε οποιαδήποτε χρήσιμη πληροφορία μπορεί να κατέχει ή να αποκτήσει το μοντέλο μάθησης, σχετικά με την λειτουργία Τ που θέλει να υλοποιήσει, προτού το τροφοδοτήσουμε με τα ελάχιστα διαθέσιμα δείγματα.[16]

2.1.2 Είδη Διεργασιών

Στην παράγραφο αυτήν, αναφέρονται τα πιο συνήθη και βασικά είδη διεργασιών που περιλαμβάνονται στην Μάθηση Λίγων Λήψεων.

Οι περισσότερες ερευνητικές εργασίες και εφαρμογές Few-Shot Learning αφορούν κυρίως Προβλήματα Εποπτευόμενης Μάθησης (Few-Shot Supervised Learning), και κατά κύριο λόγο διεργασίες Κατηγοριοποίησης (Classification) και Παλινδρόμησης (Regression).

2.1.2.1 Κατηγοριοποίηση με Ελάχιστα Παραδείγματα (Few-Shot Classification)

Πρόκειται για προβλήματα κατηγοριοποίησης, όπου στα δεδομένα μας έχουμε ελάχιστες ετικέτες για την κάθε κλάση. Εφαρμογές κατηγοριοποίησης με ελάχιστα παραδείγματα, λαμβάνουν χώρα κυρίως σε προβλήματα κατηγοριοποίησης εικόνας (Image Classification), χαρακτηρισμού περιεχομένου από μία σύντομη περιγραφή κείμενου (Sentiment Classification) και αναγνώρισης αντικειμένου (Object Recognition).

Μία καθιερωμένη γενική μορφή των προβλημάτων Κατηγοριοποίησης Ελάχιστων Παραδειγμάτων, είναι αυτή της n-επιλογών (κλάσεων) και k-λήψεων (δειγμάτων) (γνωστή ως n-way – k-shot), με την συλλογή δεδομένων εκπαίδευσης να περιλαμβάνει σύνολο $I = k \times n$ κατηγοριοποιημένα δεδομένα.

Όπου,

- n: ο αριθμός των διαθέσιμων κλάσεων για το συγκεκριμένο πρόβλημα κατηγοριοποίησης,
- k: ο αριθμός των παραδειγμάτων που έχουμε διαθέσιμα, για κάθε μία από τις n κλάσεις.
- I: Το συνολικό πλήθος των δεδομένων εκπαίδευσης με ετικέτα, του συγκεκριμένου προβλήματος κατηγοριοποίησης.

Για τα μοντέλα Few-Shot της βιβλιογραφίας, συνηθίζεται να ελέγχονται οι επιδόσεις για k≤5. Όταν k=5, λέγεται ότι το πρόβλημα είναι 5-shot, όταν k=4 λέγεται ότι το πρόβλημα είναι 4-shot κι ούτω καθεξής. Για k=1, έχουμε ένα βασικό πρόβλημα (και μία από τις επικρατέστερες περιπτώσεις ελέγχου), το οποίο ονομάζεται 1-shot, και σε αυτήν την περίπτωση έχουμε μόνο ένα διαθέσιμο δεδομένο για κάθε κλάση.

Υπάρχει και η περίπτωση για k=0 (0-shot), όπου η πληροφορία παρέχεται από άλλες πηγές (όπως παραδείγματος χάριν σε μορφή κειμένου) προκειμένου να είναι δυνατή η προσέγγιση του προβλήματος.[17]

[16] Ένα χαρακτηριστικό παράδειγμα, όπου γίνεται χρήση προγενέστερης γνώσης είναι η Μάθηση Bayesian (Bayesian Learning). Η Μάθηση Bayesian συνδυάζει την παρεχόμενη πληροφορία που βρίσκεται στην συλλογή δεδομένων D_{train}, με μία προγενέστερη πιθανοτική κατανομή, η οποία έχει αποκτηθεί σε προηγούμενο στάδιο.

[17] Συνήθως, στις περιπτώσεις 0-shot Learning χρησιμοποιούνται διαφορετικής μορφής-πηγής δεδομένα για την απόκτηση πληροφορίας και την αντιμετώπιση των εκάστοτε διεργασιών. Παραδείγματος χάριν, η πληροφορία για ένα πρόβλημα κατηγοριοποίησης εικόνων μπορεί να προέρχεται είτε από ηχητικά δεδομένα, είτε από δεδομένα κειμένου είτε από κάποια άλλη μορφή δεδομένου (modality). Ο τρόπος αντιμετώπισης αυτός ονομάζεται Multi-modal Learning και μπορεί να αξιοποιηθεί και να επεκταθεί και σε άλλων ειδών προβλήματα.

Σχήμα 2.1.2.1 - 1: Παράδειγμα απεικόνισης μίας περίπτωσης κατηγοριοποίησης εικόνων με ελάχιστα παραδείγματα, σύμφωνα με το πρότυπο προβλημάτων n-ways k-shot, για 3 κλάσεις (τρία είδη: γάτες -πρόβατα - γουρούνια) με δύο διαθέσιμα παραδείγματα για την κάθε μία (δηλαδή n=3 και k=2, 3-ways – 2-shot).

Αριστερά η συλλογή δεδομένων εκπαίδευσης με τα ελάχιστα διαθέσιμα παραδείγματα και δεξιά οι ζητούμενες εικόνες προς κατηγοριοποίηση.[18]

Πηγή: velog.io

2.1.2.2 Παλινδρόμηση με ελάχιστα παραδείγματα (Few-shot regression)

Είναι τα προβλήματα που αποσκοπούν στην εύρεση μίας ζητούμενης συνάρτησης *f*, η οποία να ικανοποιεί την σχέση *y=f(x)* έχοντας μόνο περιορισμένα ζεύγη τιμών *x*-*y*.

2.1.2.3 Reinforcement και Cross-Domain Few Shot Learning

Επιπλέον, στην βιβλιογραφία συναντώνται περιπτώσεις ενισχυτικής Μάθησης Λίγων Λήψεων (Few-Shot Reinforcement Learning) [70], στις οποίες ζητούμενο είναι η εύρεση μίας καλής τακτικής κινήσεων, μέσα από περιορισμένη πληροφορία όσον αφορά τις κινήσεις-δράσεις που θα ακολουθήσει το μοντέλο και τις ανταμοιβές-αξιολογήσεις που θα επέλθουν σύμφωνα με αυτές.

Μία προσφάτως πολύ αναπτυσσόμενη περιοχή είναι τα προβλήματα Cross-Domain FSL (Προβλήματα Μάθησης Λίγων Λήψεων Διασταυρωμένων Περιοχών). Πρόκειται για θεματική, η οποία προέρχεται από την «Μάθηση Μέσω Μεταφοράς» (Transfer Learning – έννοια που επεξηγείται στην επόμενη ενότητα). Τα προβλήματα Cross-Domain FSL συνδυάζουν τα χαρακτηριστικά του Transfer Learning και του Few-Shot Learning.

2.1.3 Σχετικές Μορφές Μάθησης

Σε αυτήν την παράγραφο, αναφέρονται και επεξηγούνται μερικές συγγενικές περιοχές Μάθησης με την Μάθηση Λίγων Λήψεων. Πρόκειται για είδη μάθησης τα οποία είναι σημαντικά για την κατανόηση του Few-Shot Learning, καθώς χρησιμοποιούνται παράλληλα και επικουρικά στην επίλυση των προβλημάτων του. Ο τρόπος αυτός χρησιμοποίησης τους, πολλές φορές συγχέει τα διαχωριστικά όρια μεταξύ αυτών των περιοχών και της Μάθησης Λίγων Λήψεων, καθιστώντας τον διαχωρισμό τους δυσδιάκριτο.

[18] Εναλλακτικοί όροι οι οποίοι συναντώνται συχνά στην βιβλιογραφία, για την συλλογή δεδομένων εκπαίδευσης και την συλλογή δεδομένων αξιολόγησης, σε προβλήματα λίγων λήψεων, είναι το support set (συλλογή υποστήριξης) και τα queries (ερωτήματα), αντίστοιχα.

Έτσι, προκειμένου να γίνει πιο σαφής αυτός ο διαχωρισμός, πέρα από την επεξήγηση των συγκεκριμένων ειδών μάθησης, όπου κρίνεται απαραίτητο, επισημαίνονται και τα βασικά στοιχεία διαφοροποίησης τους από το Few-Shot Learning.

2.1.3.1 Μάθηση Αδύναμης Επίβλεψης (Weakly Supervised Learning)

Πρόκειται για διεργασίες Τ, όπου η εμπειρία των μοντέλων Ε χαρακτηρίζεται από αδύναμη επίβλεψη. Περιπτώσεις αδύναμης επίβλεψης έχουμε σε προβλήματα όπου οι ετικέτες των δεδομένων είναι ελλιπείς, ανακριβείς, λανθασμένες ή έχουν τροποποιηθεί εξαιτίας της ύπαρξης θορύβου. Οι πιο κοντινές διεργασίες αυτού του είδους μάθησης με την Μάθηση Λίγων Λήψεων είναι αυτές όπου μόνο μία μικρή ποσότητα δεδομένων εμπεριέχει πληροφορία εποπτείας. Σύμφωνα με τον βαθμό αναγκαιότητας της ανθρώπινης παρέμβασης, η Μάθηση Αδύναμης Επίβλεψης διακρίνεται σε:

- Ημί-Εποπτευόμενη Μάθηση (Semi-Supervised Learning)

  Στο συγκεκριμένο είδος μάθησης το μοντέλο εκπαιδεύεται σε έναν μικρό αριθμό κατηγοριοποιημένων δεδομένων και συνήθως έναν μεγάλο αριθμό μη-κατηγοριοποιημένων δεδομένων. Εφαρμογές αυτού του τρόπου μάθησης λαμβάνουν χώρα στην κατηγοριοποίηση κειμένου και ιστοσελίδων (text & webpage classification).

  Ειδική κατηγορία της Ημί-Εποπτευόμενης Μάθησης συνιστά η μάθηση Θετικών και μη κατηγοριοποιημένων δεδομένων (Positive-Unlabeled Learning) κατά την οποία στην συλλογή δεδομένων περιλαμβάνονται είτε θετικά κατηγοριοποιημένα δεδομένα είτε μη-κατηγοριοποιημένα δεδομένα. Ο συγκεκριμένος τρόπος μάθησης βρίσκει εφαρμογή στα μέσα κοινωνικής δικτύωσης, όσον αφορά τις προτάσεις φίλων. Στην περίπτωση αυτή η μόνη γνωστή πληροφορία είναι η λίστα φίλων του εκάστοτε χρήστη (θετική κατηγοριοποίηση), ενώ η σχέση με τους υπόλοιπους χρήστες είναι άγνωστη.

- Ενεργός Μάθηση (Active Learning)

  Αφορά ένα διαδραστικό είδος μάθησης, στο οποίο μπορεί να ζητηθεί άμεσα από την πηγή προέλευσης των δεδομένων η κατηγοριοποίηση ενός συγκεκριμένου μη-κατηγοριοποιημένου δεδομένου. Χρησιμοποιείται σε εφαρμογές όπως ο εντοπισμός πεζού (pedestrian detection) όπου η γνωστοποίηση των ετικετών των δεδομένων είναι αρκετά δαπανηρή.

Εκ φύσεως, η Μάθηση Αδύναμης Επίβλεψης αναφέρεται σε διεργασίες κατηγοριοποίησης και παλινδρόμησης με αδύναμη εποπτεία, ενώ η Μάθηση Λίγων Λήψεων αφορά ένα ευρύτερο πλαίσιο προβλημάτων που συμπεριλαμβάνει και διεργασίες ενισχυτικής Μάθησης. Επίσης, στην Μάθηση Αδύναμης Επίβλεψης με ημιτελή εποπτεία χρησιμοποιούνται κατά κύριο λόγο μη κατηγοριοποιημένα δεδομένα ως επιπλέον πληροφορία για την εμπειρία Ε, ενώ στην Μάθηση Λίγων Λήψεων η προγενέστερη γνώση μπορεί να προέρχεται από διάφορες πηγές, και δεν περιορίζεται μόνο στην χρήση και την αξιοποίηση των μη κατηγοριοποιημένων δεδομένων. Συνεπώς, μπορούμε να πούμε ότι η Μάθηση Αδύναμη Επίβλεψης αφορά ένα πολύ συγκεκριμένο σύνολο προβλημάτων Few-Shot Learning, όπου

(1) η συλλογή δεδομένων είναι μικρή,
(2) η προγενέστερη γνώση που επιδιώκει το μοντέλο να αποκτήσει και να αξιοποιήσει μετέπειτα, προέρχεται αυστηρά από τα μη κατηγοριοποιημένα δεδομένα και
(3) το πρόβλημα που καλείται να αντιμετωπίσει το μοντέλο αφορά διεργασία κατηγοριοποίησης ή παλινδρόμησης.

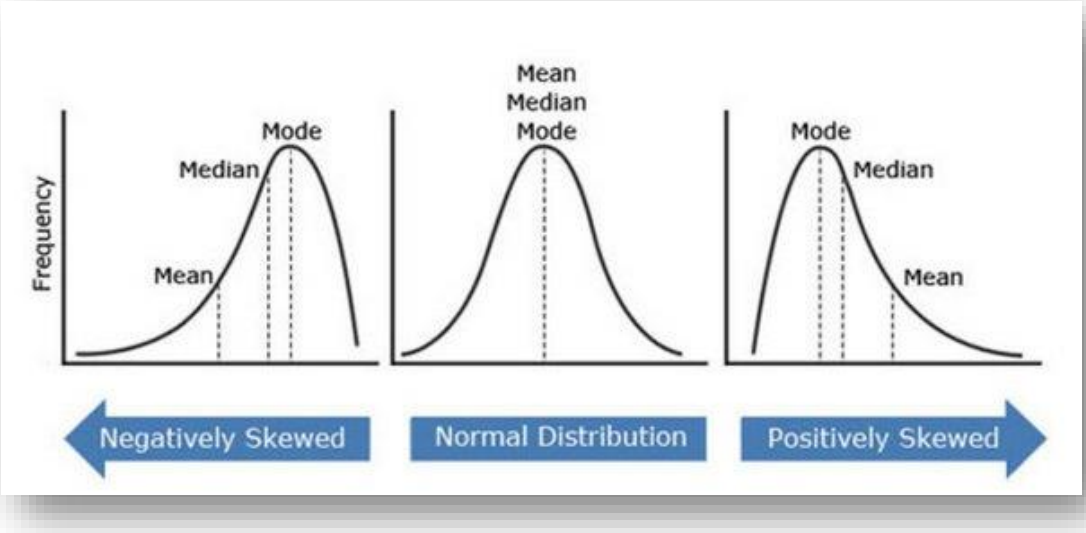
2.1.3.2 Μη-Ισορροπημένη Μάθηση (Imbalanced Learning)

Σε προβλήματα Μη-Ισορροπημένης Μάθησης, στην εμπειρία Ε του μοντέλου μάθησης συμπεριλαμβάνονται συλλογές δεδομένων όπου οι Κατανομές Πιθανοτήτων (μεταξύ των δεδομένων

εισόδου x και των ετικετών y) *p(x,y)* είναι ασύμμετρες. Αυτό συμβαίνει σε εφαρμογές όπου οι ετικέτες y των δεδομένων λαμβάνουν σπανίως μία θετική ή αρνητική τιμή, όπως αυτές του εντοπισμού απάτης (fraud detection) και της προσμονής καταστροφών (catastrophe anticipation).

Τα μοντέλα Μη-Ισορροπημένης Μάθησης εκπαιδεύονται και ελέγχονται σε ολόκληρες τις συλλογές δεδομένων οι οποίες χαρακτηρίζονται από ανισορροπία. Η ειδοποιός διαφορά με τα προβλήματα Μάθησης Λίγων Λήψεων είναι ότι ο έλεγχος γίνεται στα ελάχιστα διαθέσιμα κατηγοριοποιημένα παραδείγματα.

Η Μη-Ισορροπημένη Μάθηση μπορεί να φανεί χρήσιμη στα προβλήματα Μάθησης Λίγων Λήψεων, όταν αυτή χρησιμοποιείται ως πηγή προγενέστερης γνώσης.



Σχήμα 2.1.3.2 - 1: Απεικόνιση γραφικών διαφορών μεταξύ Ασύμμετρων Κατανομών Πιθανοτήτων και Κανονικής Γκαουσιανής Κατανομής. Αριστερά φαίνεται μία Ασύμμετρη Κατανομή Πιθανοτήτων με αρνητική ασυμμετρία, στην μέση αναπαρίσταται η Κανονική Γκαουσιανή Κατανομή και δεξιά απεικονίζεται μία Ασύμμετρη Κατανομή Πιθανοτήτων με θετική ασυμμετρία.

Επίσης, αναπαρίστανται και τα μεγέθη μέσης τιμής (μέσος όρος όλων των τιμών της κατανομής), ενδιάμεσης τιμής (μέσος ή μεσαία τιμή των μεταβλητών της κατανομής) και επικρατούσας τιμής (η τιμή που εμφανίζεται τις περισσότερες φορές στην κατανομή) για κάθε μία από τις παραπάνω κατανομές.

Πηγή: Quora

2.1.3.3 Μάθηση Μέσω Μεταφοράς (Transfer Learning)

Πρόκειται για την μεταφορά κεκτημένης γνώσης από ένα μοντέλο (πηγή γνώσης), το οποίο έχει εκπαιδευτεί σε μία διεργασία - περιοχή όπου υπάρχει αφθονία δεδομένων, σε ένα άλλο μοντέλο (αποδέκτης γνώσης), όπου τα δεδομένα είναι περιορισμένα.

Η Μάθηση Μεταφοράς έχει συνεισφέρει τα μέγιστα σε διάφορες εφαρμογές και διεργασίες όπως:

- προτάσεις εμπνευσμένες από παρόμοια πεδία ενδιαφέροντος (cross-domain recommendation),
- η μετάδοση γνώσης για τον εντοπισμό περιοχής WiFi σε διάφορες χρονικές περιόδους (WiFi localization across time periods),
- η μεταφορά δυνατοτήτων σε κινητές και διαστημικές συσκευές.

Ο πιο καθιερωμένος τρόπος χρησιμοποίησης του Transfer Learning λέγεται «προσαρμογή στην περιοχή» (Domain Adaptation) και αφορά την μεταφορά γνώσης για περιπτώσεις όπου η διεργασία του μοντέλου από το οποίο προέρχεται η γνώση και η διεργασία που καλείται να επιλύσει το μοντέλο που αποδέχεται την γνώση είναι η ίδια, αλλά διαφοροποιείται το πεδίο εφαρμογής. Σε αυτήν την περίπτωση, το μοντέλο – αποδέκτης, πρέπει να τροποποιήσει αντίστοιχα τις παράμετρούς του, προκειμένου να προσαρμόσει την γνώση που έλαβε στο ζητούμενο πεδίο εφαρμογής που καλείται να αντιμετωπίσει.

Παράδειγμα «προσαρμογής σε περιοχή», θα μπορούσε να αποτελέσει μία διεργασία χαρακτηρισμού περιεχομένου μέσα από μία σύντομη περιγραφή κείμενου (Sentiment Classification), όπου το μοντέλο της πηγής της γνώσης έχει εκπαιδευτεί στην κατηγοριοποίηση περιεχομένου για κείμενα που σχετίζονται με σχόλια θεατών πάνω σε ταινίες, ενώ η ζητούμενη περιοχή εφαρμογής του μοντέλου - αποδέκτη αφορά σχόλια χρηστών για καθημερινά αγαθά.

Η Μάθηση Μέσω Μεταφοράς συνιστά ένα εξαιρετικό εργαλείο απόκτησης προγενέστερης γνώσης και χρησιμοποιείται κατά κόρον στα προβλήματα Few-Shot Learning.

2.1.3.4 Μέτα-Μάθηση (Meta-Learning)

Στα προβλήματα Μέτα-Μάθησης, η απόδοση P της ζητούμενης διεργασίας T που καλείται ο αλγόριθμος να αντιμετωπίσει, βελτιώνεται μέσα από τον συνδυασμό της συλλογής δεδομένων που παρέχεται και της Μέτα-Γνώσης η οποία εξάγεται από την διαδικασία της Μέτα-Μάθησης.

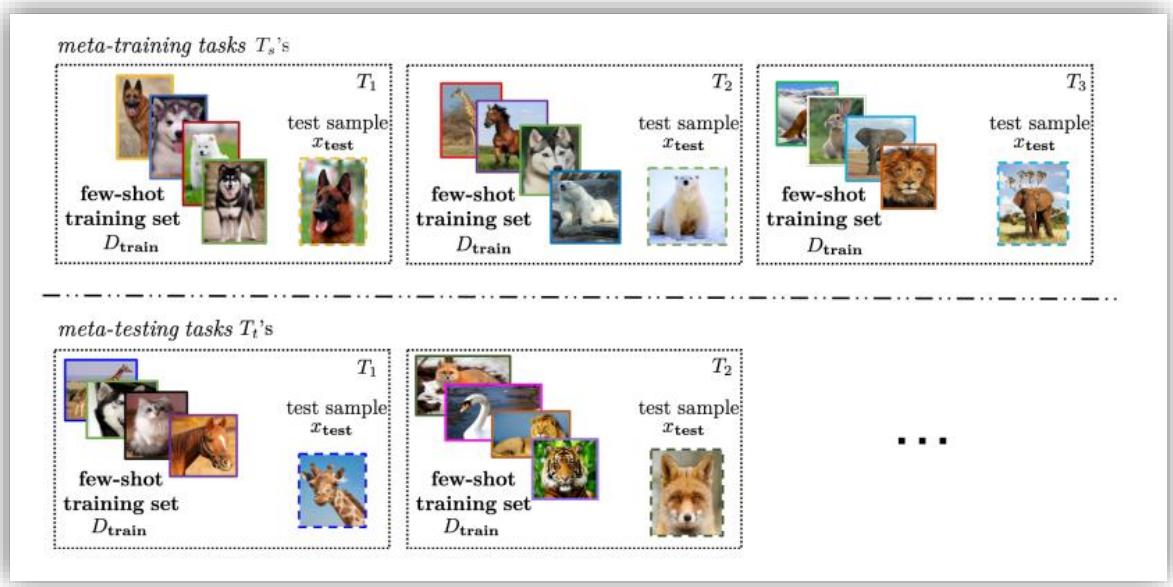
Πιο συγκεκριμένα, κατά την διαδικασία της Μέτα-Μάθησης, το μοντέλο μαθαίνει γενικές πληροφορίες (Μέτα-Γνώση) μέσα από ένα σύνολο σχετικών διεργασιών, τις οποίες στην συνέχεια καλείται να προσαρμόσει στην διεργασία T με την οποία έρχεται αντιμέτωπο.

Θα μπορούσαμε να πούμε, πως το μοντέλο «μαθαίνει να μαθαίνει» μέσω της τριβής του με τις διάφορες διεργασίες, και στην συνέχεια υποβάλλεται στην τελική διεργασία ενδιαφέροντος T.

Η Μέτα-Μάθηση έχει εφαρμοστεί επιτυχώς σε προβλήματα όπως:

- η εκμάθηση των αλγορίθμων βελτιστοποίησης (learning optimizers),
- η αντιμετώπιση της δυσκολίας προτάσεων για νέα δεδομένα στα συστήματα συστάσεων (cold-start problem in collaborative filtering), και
- η καθοδήγηση αυτόνομων προγραμμάτων μέσω φυσικής γλώσσας (guiding policies by natural language).

Οι μέθοδοι της Μέτα-Μάθησης εφαρμόζονται κατά κόρον στην αντιμετώπιση διεργασιών Μάθησης Λίγων Λήψεων, αξιοποιώντας την Μέτα-Γνώση που παράγεται από τις διάφορες διεργασίες ως προγενέστερη γνώση.



Σχήμα 2.1.3.4 - 1: Η χρήση της Μέτα-Μάθησης στην αντιμετώπιση του προβλήματος Μάθησης Λίγων Λήψεων.

Στην φάση της Μέτα-Εκπαίδευσης (meta-training) το μοντέλο εκπαιδεύεται σε ένα σύνολο διεργασιών T_1, T_2, T_3 και προσαρμόζει τις παραμέτρους του ϑ, προκειμένου να ελαχιστοποιήσει το συνολικό κόστος το οποίο αντιπροσωπεύει κάθε μία από αυτές. Έτσι εξάγεται η Μέτα-Γνώση, η οποία χρησιμοποιείται ως πηγή προγενέστερης γνώσης κατά την φάση του Μέτα-ελέγχου.

Στην φάση του Μέτα-ελέγχου (στο σημείο αυτό ξεκινάει το πρόβλημα Μάθησης Ελάχιστων Λήψεων), το μοντέλο συντηρεί την κεκτημένη Μέτα-Γνώση και επιπλέον εκπαιδεύεται στα ελάχιστα κατηγοριοποιημένα δεδομένα. Το σφάλμα που θα προκύψει από αυτά, ονομάζεται σφάλμα ελέγχου Μέτα-Μάθησης (meta-learning testing error).

Πηγή: Generalizing from a Few Examples: A Survey on Few-Shot Learning Yaqing Wang, Quanming Yao, James T. Kwok, and Lionel M. Ni [55]

2.1.4 Επικρατέστερα Datasets

Στα πρώιμα στάδια της ανάπτυξης του Few Shot Learning, οι συλλογές δεδομένων αναφοράς κατασκευάζονταν χειροκίνητα, σύμφωνα με την δομή του προβλήματος και την ζητούμενη εξεταζόμενη περιοχή ελέγχου των επιδόσεων. Έτσι, για παράδειγμα σε περιπτώσεις που ήθελαν να εξετάσουν ένα πρόβλημα n-ways – k-shot επιλέγονταν k παραδείγματα για κάθε μία από τις n κλάσεις. Ο τρόπος αυτός ελέγχου, δεν ανταποκρίνεται πλήρως όμως στα προβλήματα πραγματικών δεδομένων και πλέον μετά από κάποια χρόνια ανάπτυξης της περιοχής της Μάθησης Λίγων Λήψεων έχουν δημιουργηθεί γενικά και αξιόπιστα datasets που χρησιμοποιούνται ως πρότυπα για τον ορθότερο έλεγχο στα προβλήματα few-shot.

Στην παράγραφο αυτήν παρουσιάζονται τα πιο χρησιμοποιούμενα Datasets σύμφωνα με την εργασία [56] τα οποία αποτελούν μέτρο των επιδόσεων και τρόπο αξιολόγησης και σύγκρισης μεταξύ των διαφόρων προσεγγίσεων στα προβλήματα λίγων λήψεων.

Οι 8 πιο καθιερωμένες συλλογές δεδομένων, με φθίνουσα σειρά χρησιμοποίησης σύμφωνα με τις πιο πρόσφατες δημοσιευμένες έρευνες στο Few Shot Learning (2017-2021), είναι:

- CUB-200-2011: Πρόκειται για επέκταση της συλλογής CUB-200. Περιέχει ένα σύνολο από 11.788 εικόνες, στις οποίες αναπαρίστανται 200 διαφορετικά είδη πτηνών, συνοδευόμενες από ποικίλες πληροφορίες για αυτά, επιστημονικές λεκτικές περιγραφές, σημειώσεις και επισημάνσεις για τα διάφορα χαρακτηριστικά τους.
- MiniImageNet: Η δημιουργία του MiniImageNet, προήλθε από την επιλογή 100 τυχαίων κλάσεων μέσα από την τεράστια συλλογή ImageNet. Περιλαμβάνει σύνολο 60.000 εικόνες, διαφόρων αντικειμένων, 100 κλάσεις με 600 δείγματα για την κάθε μία.

- Omniglot: Η συλλογή Omniglot περιέχει σύνολο 1623 χειρόγραφους χαρακτήρες, προερχόμενους από 20 διαφορετικές πηγές - ανθρώπους, μέσα από 50 διαφορετικά αλφάβητα.
- Meta-DataSet: Αφορά μία συλλογή δεδομένων, η οποία αποτελείται από πολλαπλές συλλογές δεδομένων διαφόρων κατανομών-πηγών. Δεν περιορίζεται στο πρότυπο n-ways – k-shot, προσομοιώνοντας με αυτόν τον τρόπο προβλήματα ρεαλιστικών σεναρίων. Αντλεί πληροφορίες από 10 διαφορετικές συλλογές δεδομένων οι οποίες είναι οι: ILSVRC-2012, Omniglot, Aircraft, CUB-200-2011, Describable Textures, Quick Draw, Fungi, VGG Flower, Traffic Signs και MSCOCO.
- PASCAL-5i: Πρόκειται για dataset που δημιουργήθηκε για χρήση στο πρόβλημα «κατηγοριοποίησης - ομαδοποίησης στοιχείων εικόνας» (Image Segmentation), μέσα από την συλλογή PASCALVOC 2012 και εποπτευόμενη πληροφορία αντλημένη μέσα από την έρευνα [91] των Bharath Hariharan et al. Περιέχει 4 ομαδοποιήσεις, οι οποίες περιλαμβάνουν 5 κλάσεις και τις αντίστοιχες ετικέτες για αυτές.
- Paris-Lille-3D: Πρόκειται για dataset που χρησιμοποιείται σε διεργασίες κατηγοριοποίησης αταξινόμητων σημείων 3-διαστάσεων (Point Cloud Classification & Segmentation), του οποίου τα δεδομένα αποκτήθηκαν μέσα από κινητούς «χαρτογραφητές» (Mobile Laser Scanning). Τα τρισδιάστατα σημεία (Point Clouds) αφορούν τις χαρτογραφήσεις 2 χιλιομέτρων για τις πόλεις Παρίσι και Λιλ. Περιέχει 50 κλάσεις, στις οποίες οι ετικέτες αποδόθηκαν μία-μία με το χέρι.
- N-Digit MNIST: Αφορά παραλλαγή του γνωστού dataset χειρόγραφων ψηφίων MNIST. Για n=2 τα ψηφία εκτείνονται από 0 μέχρι 99, για n=3 από 0 μέχρι 999 κι ούτω καθεξής, επεκτείνοντας με αυτόν τον τρόπο εύκολα των αριθμό των κλάσεων και συντηρώντας παράλληλα την απλότητα των χαρακτηριστικών.
- SUN397: Συλλογή δεδομένων η οποία αφορά την αναγνώριση σκηνικού (Scene UNderastanding - SUN). Περιλαμβάνει 130.519 εικόνες στις οποίες υπάρχουν 899 κλάσεις.

Με τις τρεις πρώτες να υπερισχύουν κατά το μεγαλύτερο ποσοστό ερευνών και τις τρεις τελευταίες να εμφανίζονται σε πολύ περιορισμένο αριθμό εργασιών.

Σχήμα 2.1.4 - 1: Απεικόνιση ενός μικρού δείγματος για τις τρεις επικρατέστερες συλλογές δεδομένων στο Few-Shot Learning.

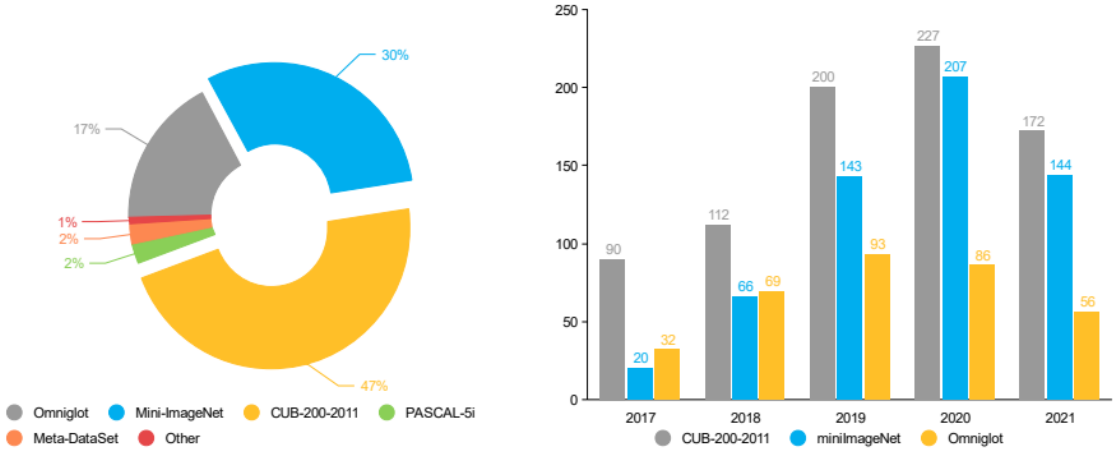
Πηγή: A survey of few-shot learning in smart agriculture: developments, applications, and challenges - Jiachen Yang, Xiaolan Guo, Yang Li, Francesco Marinello, Sezai Ercisli, Zhuo Zhang [57]

Επιπλέον, για την βελτιστοποίηση της αξιολόγησης των προβλημάτων Cross-Domain FSL, το 2020 παρατέθηκε μία εργασία από τους Yunhui Guo et al. [58], η οποία προτείνει την συλλογή αναφοράς BSCD-FSL, που περιλαμβάνει δυσεύρετα αλλά πραγματικά δεδομένα προβλημάτων, τα οποία μέχρι στιγμής ήταν σε έλλειψη στην διεθνή βιβλιογραφία. Η συλλογή αυτή αποτελείται από 4 διαφορετικές συλλογές,

- Την συλλογή CropDiseases, μία συλλογή δεδομένων για ασθένειες φυτών,
- Την συλλογή EuroSAT, η οποία αφορά δορυφορικές φωτογραφίες,

- Την ISIC, που περιέχει φωτογραφίες που απεικονίζουν δερματικές παθήσεις, και
- Την ChestX, η οποία αποτελείται από ακτινογραφίες στήθους.

Η δυσκολία για τις παραπάνω συλλογές, είναι αυξανόμενη σύμφωνα με την σειρά που παρατέθηκαν.



Σχήμα 2.1.4 - 2: Αριστερά, απεικονίζεται το ποσοστό χρησιμοποίησης της κάθε μίας από τις παραπάνω συλλογές δεδομένων στο Few Shot Learning για τις δημοσιευμένες εργασίες του διαστήματος 2017-2021. Δεξιά, απεικονίζεται ένα ιστόγραμμα που περιλαμβάνει τον αριθμό των εργασιών για τις τρεις επικρατέστερες συλλογές δεδομένων για το ίδιο χρονικό διάστημα.

Πηγή: A Comprehensive Survey of Few-shot Learning: Evolution, Applications, Challenges, and Opportunities - Yisheng Song, Ting Wang, Subrota K Mondal, Jyoti Prakash Sahoo [56]

2.1.5 Μαθηματική Μοντελοποίηση της Κεντρικής Πρόκλησης του Few-Shot Learning

Όσο αποδοτικό και αν είναι ένα μοντέλο Μηχανικής Μάθησης, πάντα θα υπάρχει ένα σφάλμα μεταξύ προβλέψεων και πραγματικών τιμών. Σε αυτήν την ενότητα, θα αναλύσουμε την επιπλέον δυσκολία που εισάγεται στα προβλήματα Μάθησης Λίγων Λήψεων. Η εν λόγω ανάλυση, ανταποκρίνεται πλήρως στα προβλήματα κατηγοριοποίησης και παλινδρόμησης, αλλά προσφέρει και μία βασική οπτική για την κατανόηση της ενισχυτικής Μάθησης Λίγων Λήψεων (Few-Shot Reinforcement Learning).

2.1.5.1 Ελαχιστοποίηση Εμπειρικού Ρίσκου

Δοθείσης μίας υπόθεσης h, επιζητούμε την ελαχιστοποίηση του Εκτιμώμενου Ρίσκου R, το οποίο είναι το συνολικό Κόστος-Απώλεια συναρτήσει της *p(x,y)*, και εκφράζεται ως:

$$R(h) = \int l(h(x), y)\, dp(x, y) = E[l(h(x), y)]$$

Όπου,

- h : η υπόθεση που αποδίδει ο αλγόριθμος Μηχανικής Μάθησης,
- x : τα δεδομένα εισόδου του μοντέλου Μηχανικής Μάθησης,
- y : οι ετικέτες που αντιστοιχούν στα δεδομένα εισόδου του μοντέλου Μηχανικής Μάθησης,
- $\mathrm{p}(\mathrm{x}, \mathrm{y})$: η πραγματική Κατανομή Πιθανοτήτων μεταξύ των δεδομένων εισόδου x και ετικετών y,
- $\mathrm{l}(\mathrm{h}(\mathrm{x}), \mathrm{y})$: Η χρησιμοποιούμενη συνάρτηση Κόστους-Απώλειας που μετρά το σφάλμα μεταξύ της τιμής της υπόθεσης που αποδίδει ο αλγόριθμος Μηχανικής Μάθησης και των πραγματικών ετικετών των δεδομένων, y.

Με βάση το γεγονός ότι η πραγματική Κατανομή Πιθανοτήτων μεταξύ των δεδομένων εισόδου x και ετικετών y, p(x,y) είναι άγνωστη, ως προσέγγιση του Εκτιμώμενου Ρίσκου R χρησιμοποιείται το Εμπειρικό Ρίσκο R_I. Το Εμπειρικό Ρίσκο αποτελεί τον μέσο όρο των απωλειών για τα I δείγματα που υπάρχουν στην συλλογή δεδομένων εκπαίδευσης D_{train}.

Πιο συγκεκριμένα, το εμπειρικό ρίσκο εκφράζεται ως:

$$R_I(h) = \frac{1}{I}\sum_{i=1}^{I} l(h(x_i, y_i))$$

Όπου, I: ο αριθμός δειγμάτων που έχουμε στο συγκεκριμένο πρόβλημα (στις περιπτώσεις Μάθησης Λίγων Λήψεων, το I είναι πολύ μικρό)

Για λόγους απεικόνισης και διευκόλυνσης της επεξήγησης, παρατίθενται οι παρακάτω έννοιες με τους αντίστοιχους συμβολισμούς τους:

- $\mathcal{H}$: ο χώρος υποθέσεων που συμπεριλαμβάνει τις διαθέσιμες συναρτήσεις, δηλαδή αυτές που δύνανται να αποτυπωθούν από το μοντέλο Μηχανικής Μάθησης.
- $\hat{h}$: η ιδανική υπόθεση που επιδιώκει να αποδώσει το μοντέλο Μηχανικής Μάθησης, δηλαδή η συνάρτηση που ελαχιστοποιεί το εκτιμώμενο Ρίσκο *R(h).*
- h^*: η κοντινότερη δυνατή προσέγγιση της υπόθεσης $\hat{h}$ μέσα από τον διαθέσιμο χώρο υποθέσεων.
- h_I : η υπόθεση μέσα από τον χώρο υποθέσεων για την οποία το εμπειρικό ρίσκο $R_I(h)$ λαμβάνει την μικρότερη τιμή.
- h : η υπόθεση η οποία αποδίδεται από το μοντέλο Μηχανικής Μάθησης

Η βέλτιστη υπόθεση $\hat{h}$ που επιδιώκει να αποδώσει το μοντέλο μηχανικής Μάθησης είναι πάντοτε άγνωστη. Συνεπώς, η ρεαλιστική προσέγγιση του προβλήματος είναι η απόδοση μίας υπόθεσης *h* η οποία να ανήκει στον χώρο υποθέσεων και να είναι όσο το δυνατόν πιο κοντά σε αυτήν. Η υπόθεση αυτή είναι η h^*, και είναι αυτήν που ελαχιστοποιεί το εκτιμώμενο Ρίσκο *R(h).* Όσον αφορά, την h_I αποτελεί την καλύτερη δυνατή υπόθεση που μπορούμε να έχουμε με τα I δεδομένα, και είναι αυτή η οποία ελαχιστοποιεί το Εμπειρικό Ρίσκο $R_I(h)$. Για λόγους απλοποίησης θα υποθέσουμε ότι οι $\hat{h}$, h^* και h_I είναι μοναδικές.

Έτσι, το συνολικό σφάλμα μπορεί να αναλυθεί ως εξής:

$$E\left[R(h_I) - R\left(\hat{h}\right)\right] = \underbrace{E\left[R(h^*) - R\left(\hat{h}\right)\right]}_{\boldsymbol{\varepsilon}_{app}(\mathcal{H})} + \underbrace{E\left[R(h_I) - R(h^*)\right]}_{\boldsymbol{\varepsilon}_{est}(\mathcal{H},\mathrm{I})}$$

Η αναμενόμενη τιμή του παραπάνω γενικού σφάλματος ισχύει για οποιαδήποτε μέγεθος συλλογής δεδομένων εκπαίδευσης D_{train}. Όπως μπορούμε να δούμε, το συνολικό σφάλμα διακρίνεται σε δύο όρους σφαλμάτων:

a. το σφάλμα προσέγγισης $\boldsymbol{\varepsilon}_{\mathbf{app}}(\mathcal{H})$, το οποίο αποτελεί ένα μέτρο του βαθμού στον οποίο μπορεί να πλησιάσει ο χώρος υποθέσεων του μοντέλου την ιδανική υπόθεση $\hat{h}$[19] και
b. το σφάλμα εκτίμησης $\boldsymbol{\varepsilon}_{\mathbf{est}}(\mathcal{H}, I)$ το οποίο εισάγεται εξαιτίας της προσπάθειας της ελαχιστοποίησης του εμπειρικού ρίσκου αντί του εκτιμώμενου.

[19] Όσο πιο κοντά ο χώρος υποθέσεων στην επιζητούμενη ιδανική συνάρτηση $\hat{h}$, τόσο πιο μικρό το σφάλμα προσέγγισης. Το συγκεκριμένο σφάλμα όμως μπορεί μόνο να ελαχιστοποιηθεί και όχι να μηδενιστεί.

Συνεπώς, το συνολικό σφάλμα καθορίζεται από τον διαθέσιμο χώρο υποθέσεων $\mathcal{H}$ που χαρακτηρίζει το μοντέλο μάθησης και από τον αριθμό I των κατηγοριοποιημένων δεδομένων που είναι διαθέσιμα για την εκπαίδευση του.

Όπως θα δούμε αναλυτικά στην επόμενη ενότητα, η ελαχιστοποίηση του παραπάνω σφάλματος μπορεί να μελετηθεί υπό τις τρεις παρακάτω οπτικές:

i. υπό την οπτική των δεδομένων, τα οποία συνιστούν την διαθέσιμη συλλογή D_{train},
ii. υπό την οπτική του μοντέλου, καθώς οι δυνατότητες και η χωρητικότητα αυτού είναι που καθορίζει τον χώρο υποθέσεων $\mathcal{H}$ και
iii. υπό την οπτική του αλγορίθμου χρησιμοποίησης, καθώς αυτός είναι υπεύθυνος για την εύρεση της καλύτερης δυνατής συνάρτησης $h_I \in \mathcal{H}$ σύμφωνα με τα διαθέσιμα δεδομένα της συλλογής εκπαίδευσης D_{train}.

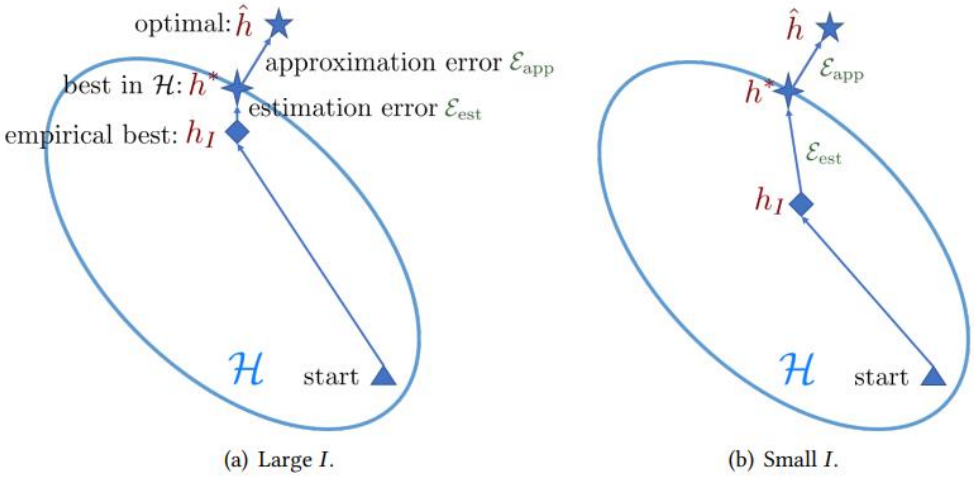
2.1.5.2 Η Δυσκολία Μείωσης του Εμπειρικού Ρίσκου στα Προβλήματα Μάθησης Λίγων Λήψεων

Γενικά, η επίτευξη της μείωσης του σφάλματος εκτίμησης $\boldsymbol{\varepsilon_{est}}(\mathcal{H}, I)$ μπορεί να επέλθει εάν αυξήσουμε το πλήθος των παραδειγμάτων που έχουμε διαθέσιμα στην συλλογή δεδομένων. Όταν υπάρχει ένας επαρκής αριθμός κατηγοριοποιημένων δεδομένων, είναι ασφαλές να ισχυριστούμε πως το Εμπειρικό Ρίσκο αποτελεί μία αρκετά καλή προσέγγιση του Εκτιμώμενου, και συνεπώς η συνάρτηση h_I που αποδίδεται από το μοντέλο είναι αρκετά καλή προσέγγιση της βέλτιστης συνάρτησης h^* του χώρου υποθέσεων.

Όπως αναφέραμε όμως, βασικό χαρακτηριστικό των προβλημάτων Μηχανικής Μάθησης Λίγων Λήψεων είναι ότι η συλλογή δεδομένων που έχουμε στην διάθεση μας είναι πολύ περιορισμένη. Συνεπώς, σίγουρα τα διαθέσιμα δεδομένα δεν συνιστούν ένα επαρκές σύνολο ώστε να φτάσουμε με ασφάλεια σε ένα συμπέρασμα όπως το παραπάνω. Αντίθετα, αυτή η έλλειψη δεδομένων οδηγεί σε μεγάλη πιθανότητα αναξιοπιστίας μεταξύ του Εκτιμώμενου Ρίσκου και της προσέγγισης του (Εμπειρικό Ρίσκο).

Ο λόγος που συμβαίνει αυτό, είναι ότι το Εμπειρικό Ρίσκο εξαρτάται άμεσα από το διαθέσιμο δείγμα δεδομένων I, το οποίο στις περιπτώσεις του Few-Shot Learning είναι πολύ μικρό για να καταστήσει το Εμπειρικό Ρίσκο μία αξιόπιστη προσέγγιση. Οπότε, αν υποθέσουμε ότι το μοντέλο εκπαιδεύτηκε με τον παραδοσιακό τρόπο εποπτευόμενης μάθησης, είναι πολύ εύκολο η εξαγόμενη συνάρτηση του, h_I , να χαρακτηρίζεται από το φαινόμενο της «υπερμοντελοποίησης» (Overfitting) πάνω στα ελάχιστα δεδομένα που υπάρχουν διαθέσιμα. Έτσι, το πιθανότερο είναι αυτή η συνάρτηση (h_I) να απέχει μεγάλη απόσταση από την ζητούμενη συνάρτηση h^* του χώρου υποθέσεων.

Ουσιαστικά, πέρα από την εγγενή δυσκολία που υπάρχει σε όλα τα προβλήματα Μηχανικής Μάθησης (αυτή της βέλτιστης προσέγγισης της συνάρτησης h^* μέσα από τον χώρο υποθέσεων), τα προβλήματα Ελάχιστης Μάθησης προσθέτουν ένα ακόμη εμπόδιο (αυτό της βέλτιστης απόδοσης της h_I). Το γεγονός αυτό είναι που καθιστά τα προβλήματα Μάθησης Ελάχιστων Δεδομένων πρόκληση στο πεδίο της Μηχανικής Μάθησης και ακόμη δυσκολότερα από τα κλασικά προβλήματα.



(a) Large I. (b) Small I.

Σχήμα 2.1.5.2 - 1: Σύγκριση της προσέγγισης της βέλτιστης συνάρτησης $\hat{h}$ για τις περιπτώσεις όπου η συλλογή δεδομένων είναι (a) μεγάλη και (b) μικρή. Πιο συγκεκριμένα, απεικονίζονται μία ενδεικτική αναπαράσταση της θέσης των υποθέσεων $\hat{h}$, h^* και h_I και ο χώρος υποθέσεων $\mathcal{H}$ ο οποίος διαγράφεται από την περιοχή που περικλείει η ελλειπτική καμπύλη.

Στην αριστερή διαγραμματική αναπαράσταση, όπου η συλλογή δεδομένων αποτελείται από έναν μεγάλο αριθμό δειγμάτων, παρατηρούμε ότι η υπόθεση h_I που αποδίδεται μέσα από την ελαχιστοποίηση του εμπειρικού ρίσκου είναι αρκετά κοντά στην υπόθεση h^*, κι έτσι υπάρχει μεγάλη δυνατότητα περιορισμού του εκτιμώμενου σφάλματος $\boldsymbol{\varepsilon_{est}}$.

Αντίθετα στην δεξιά διαγραμματική αναπαράσταση, όπου αντιμετωπίζεται ένα πρόβλημα Μάθησης Λίγων Λήψεων, η καλύτερη δυνατή υπόθεση h_I με βάση την διαθέσιμη περιορισμένη συλλογή δεδομένων απέχει μεγάλη απόσταση από την ιδανική προσέγγιση h^* του χώρου υποθέσεων και η ελαχιστοποίηση του εκτιμώμενου σφάλματος $\boldsymbol{\varepsilon_{est}}$ αποτελεί πολύ πιο δύσκολο εγχείρημα.

Πηγή: Generalizing from a Few Examples: A Survey on Few-Shot Learning - Yaqing Wang, Quanming Yao, James T. Kwok, and Lionel M. Ni [55]

2.2 Κατηγοριοποίηση Προβλημάτων Μάθησης Λίγων Λήψεων

Προκειμένου να αντιμετωπίσουμε την παραπάνω πρόκληση, και να δημιουργήσουμε μοντέλα Μηχανικής Μάθησης τα οποία ανταποκρίνονται με αξιόλογες επιδόσεις στα προβλήματα Μάθησης Λίγων Λήψεων, είναι επιτακτική η ανάγκη χρησιμοποίησης κάποιου είδους προγενέστερης γνώσης.

Στην ενότητα αυτήν, παρατίθεται μία κατηγοριοποίηση των προβλημάτων του Few-Shot Learning, σύμφωνα με την αναλυτική εργασία “Generalizing from a Few Examples: A Survey on Few-Shot Learning” των Yaqing Wang et al [55]. Στην συνέχεια ακολουθούν ξεχωριστές επεξηγηματικές παράγραφοι για κάθε μία από αυτές τις κατηγορίες, όπου παρουσιάζονται οι επί μέρους υποκατηγορίες, η λειτουργία τους και παραδείγματα εφαρμογών των εκάστοτε τεχνικών που συμπεριλαμβάνονται σε αυτές.

Έτσι, σύμφωνα με την παραπάνω εργασία, ανάλογα με την προέλευση αλλά και το πεδίο εστίασης της προγενέστερης γνώσης που πρόκειται να χρησιμοποιηθεί, οι εργασίες Μάθησης Λίγων Λήψεων μπορούν να κατηγοριοποιηθούν σε τρεις περιοχές εξέτασης.

Οι περιοχές αυτές είναι:

1) τα δεδομένα, όπου πεδίο ενδιαφέροντος είναι η επέκταση της συλλογής δεδομένων,

2) το μοντέλο, όπου η προσοχή επικεντρώνεται στον διαθέσιμο χώρο υποθέσεων,
3) ο αλγόριθμος, όπου στόχος είναι η βελτίωση των μέθοδών αναζήτησης της βέλτιστης συνάρτησης h^* μέσα στον εκάστοτε χώρο υποθέσεων.

Δεδομένα

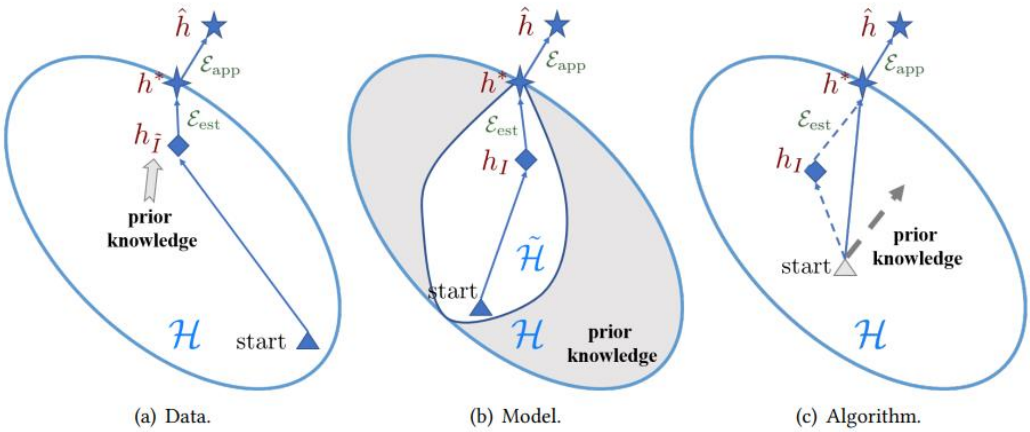
Η περιοχή όπου η προγενέστερη γνώση χρησιμοποιείται με σκοπό την αύξηση του μεγέθους της συλλογής δεδομένων από I σε $\tilde{I}$ με $\tilde{I} \gg I$. Έπειτα, μπορούν να χρησιμοποιηθούν με τους συνήθεις τρόπους κλασικοί αλγόριθμοι Μηχανικής Μάθησης στην νέα αποκτημένη και μεγαλύτερη συλλογή δεδομένων, προκειμένου να εξαχθεί μία αποδοτικότερη υπόθεση $h_{\tilde{I}}$ (η οποία θα βρίσκεται πιο κοντά στην h^*).

Μοντέλο

Χρησιμοποιούνται μέθοδοι οι οποίες αξιοποιούν προγενέστερη γνώση με σκοπό τον περιορισμό του χώρου υποθέσεων. Απώτερος σκοπός αυτού του περιορισμού είναι η επιλογή μίας καταλληλότερης υπόθεσης h_I. Ο νέος περιορισμένος χώρος υποθέσεων $\tilde{\mathcal{H}}$ οδηγεί σε μικρότερη χωρητικότητα-πολυπλοκότητα άρα και στην απαίτηση για εκμάθηση λιγότερων παραμέτρων. Έτσι, το εγχείρημα είναι να καταφέρουμε να περιορίσουμε κατάλληλα τον χώρο υποθέσεων των μοντέλων, σε τέτοιον βαθμό όπου τα λίγα διαθέσιμα δεδομένα θα αποτελούν ένα επαρκές σύνολο για την ρύθμιση των παραμέτρων των μοντέλων και την επίτευξη μίας υπόθεσης h_I η οποία θα διαφέρει ελάχιστα από την h^*.

Αλγόριθμος

Οι εργασίες που επικεντρώνονται στην βελτίωση του αλγορίθμου μάθησης, χρησιμοποιούν προγενέστερη γνώση στοχεύοντας στην ποιοτικότερη αναζήτηση των παραμέτρων ϑ του μοντέλου, οι οποίες θα οδηγήσουν σε μία αποδοτικότερη υπόθεση h_I (η οποία θα βρίσκεται όσο το δυνατόν κοντύτερα στην βέλτιστη συνάρτηση h^* του χώρου υποθέσεων). Η βελτίωση αυτής της στρατηγικής αναζήτησης, επιτυγχάνεται είτε μέσω καλύτερης αρχικοποίησης των παραμέτρων ϑ (έτσι ώστε η h που αποδίδεται από το μοντέλο προτού αυτό εκπαιδευτεί στα ελάχιστα δεδομένα, να βρίσκεται ήδη κοντύτερα στην ζητούμενη h^*), είτε δίνοντας μία καταλληλότερη κατεύθυνση στην τροποποίηση των παραμέτρων του μοντέλου, προκειμένου αυτό να πλησιάσει ορθότερα προς την επιθυμητή h^* του χώρου υποθέσεων. Στην περίπτωση αυτήν, τα βήματα αναζήτησης που θα ακολουθηθούν εξαρτώνται από την προγενέστερη γνώση που χρησιμοποιείται και τον τρόπο προσέγγισης της ελαχιστοποίησης του εμπειρικού ρίσκου.



Σχήμα 2.2 - 1: Απεικόνιση μορφής προγενέστερης γνώσης για τις παραπάνω διαχωρισμένες περιοχές μελέτης. (a)περιοχή εξέτασης των δεδομένων (b)περιοχή εξέτασης του μοντέλου (c)περιοχή εξέτασης του Αλγορίθμου.

Στην περίπτωση της αριστερής διαγραμματικής αναπαράστασης, η οποία αφορά την επέκταση της συλλογής δεδομένων, η προγενέστερη γνώση οπτικοποιείται μέσω της h_I η οποία μας οδηγεί σε μία καλύτερη προσέγγιση της συνάρτησης h^* (παρόμοια με το Σχήμα 2.1.5.2 - 1 για την περίπτωση όπου η συλλογή δεδομένων είναι μεγάλη).

Στην περίπτωση της μεσαίας διαγραμματικής αναπαράστασης, η οποία σχετίζεται με την περιοχή του Μοντέλου, η προγενέστερη γνώση αναπαρίσταται από την γκρίζα περιοχή η οποία αφαιρείται από τον αρχικό χώρο υποθέσεων, δημιουργώντας τον νέο περιορισμένο μικρότερο χώρο $\widetilde{\mathcal{H}}$ (μικρότερη ελλειπτική καμπύλη). Έτσι, οδηγούμαστε σε μικρότερη πολυπλοκότητα του μοντέλου, άρα λιγότερες παραμέτρους εκμάθησης και συνεπώς περισσότερες πιθανότητες μίας πιο εύστοχης υπόθεσης h_I.

Στην δεξιά διαγραμματική αναπαράσταση, όπου έχουμε την περίπτωση βελτίωσης του αλγόριθμου, η προγενέστερη γνώση συμβολίζεται: (i) με γκρι τρίγωνο (το οποίο ξεκινά πιο κοντά στην h^* από τις υπόλοιπες περιπτώσεις), για την περίπτωση βελτίωσης μέσω της καλύτερης αρχικοποίησης των παραμέτρων ϑ και (ii) με το γκρι διακεκομμένο βέλος, για την περίπτωση βελτίωσης μέσω της καλύτερης καθοδήγησης των βημάτων αναζήτησης του αλγορίθμου βελτιστοποίησης.

Πηγή: Generalizing from a Few Examples: A Survey on Few-Shot Learning Yaqing Wang, Quanming Yao, James T. Kwok, and Lionel M. Ni [55]

2.2.1 Εστίαση στα Δεδομένα

Στον τομέα των δεδομένων, οι διεργασίες Μάθησης Λίγων Λήψεων αποσκοπούν στην απόκτηση προγενέστερης γνώσης και την απόδοση μίας ικανοποιητικής συνάρτησης h_I μέσω της επέκτασης της Συλλογής Δεδομένων με διάφορους τρόπους.

Μία συνήθης μεθοδολογία για τον εμπλουτισμό της συλλογής δεδομένων, σε προβλήματα Μάθησης Λίγων Λήψεων, είναι μέσω χειροκίνητων τρόπων επεξεργασίας των υπαρχόντων περιορισμένων δεδομένων. Παραδείγματος χάριν, για ένα πρόβλημα όπου τα δεδομένα αφορούν εικόνες, τέτοιους τρόπους επεξεργασίας αποτελούν:

- η «μετάθεση» εικόνας (translation), η οποία αφορά την μετακίνηση του περιεχομένου της εικόνας, κατά έναν παράγοντα, είτε προς τον άξονα x, είτε προς τον άξονα y, είτε προς έναν συνδυασμό αυτών,
- το αναποδογύρισμα της εικόνας (flipping),
- η «κόμμωση» εικόνας (shearing), κατά την οποία η εικόνα υπόκειται σε παραμόρφωση του περιεχομένου της είτε προς τον άξονα x, είτε προς τον άξονα y,
- η κλιμάκωση της εικόνας (scaling), δηλαδή η μεγέθυνση – επέκταση της εικόνας ή η σμίκρυνση – συρρίκνωση της,
- η αντανάκλαση της εικόνας (reflection), όπου το περιεχόμενο της εικόνας αντικατοπτρίζεται ως προς κάποιον άξονα (συνήθως τον κατακόρυφο κεντρικό),
- η επιλογή τομέα της εικόνας (cropping), όπου επιλέγεται μία συγκεκριμένη περιοχή της, στην συνέχεια επαναπροσαρμόζεται στο μέγεθος της αρχικής εικόνας, και το αποτέλεσμα που αποδίδεται από αυτήν την διαδικασία συνιστά το νέο παραχθέν δείγμα, και
- η περιστροφή εικόνας (rotation), όπου το περιεχόμενο περιστρέφεται κατά μία γωνία φ.

Με τους παραπάνω τρόπους επεξεργασίας παράγονται νέες εικόνες, οι οποίες αποτελούν παραλλαγές της αρχικής και μπορούν να χρησιμοποιηθούν ως επιπρόσθετα δείγματα στην περιορισμένη συλλογή δεδομένων. Παρά την αποτελεσματικότητα των χειροκίνητων τρόπων επέκτασης της συλλογής δεδομένων, κατά την εφαρμογή αυτών εμφανίζονται και κάποια εμπόδια, όπως:

- το γεγονός ότι ο αναλυτικός σχεδιασμός τους απαιτεί την αφιέρωση πολλής ανθρώπινης ενέργειας αλλά και εξειδικευμένη γνώση πάνω στα συγκεκριμένα πεδία εφαρμογής τους,

- η κατασκευή αυτών υλοποιείται με σκοπό την χρήση σε συγκεκριμένη συλλογή δεδομένων και η επαναχρησιμοποίηση των ίδιων κανόνων σε διαφορετικές συλλογές δεδομένων είναι συνήθως αναποτελεσματική,
- επιδέχονται βελτιώσεις καθώς είναι πολύ δύσκολο για έναν άνθρωπο να καταφέρει να καταμετρήσει όλες τις πιθανές παραλλαγές που μπορούν να παραχθούν.

Έτσι, είναι αδύνατο να ισχυριστούμε πως οι χειροκίνητοι τρόποι επέκτασης της συλλογής δεδομένων επιλύουν ολοσχερώς το πρόβλημα Μάθησης Λίγων Λήψεων.

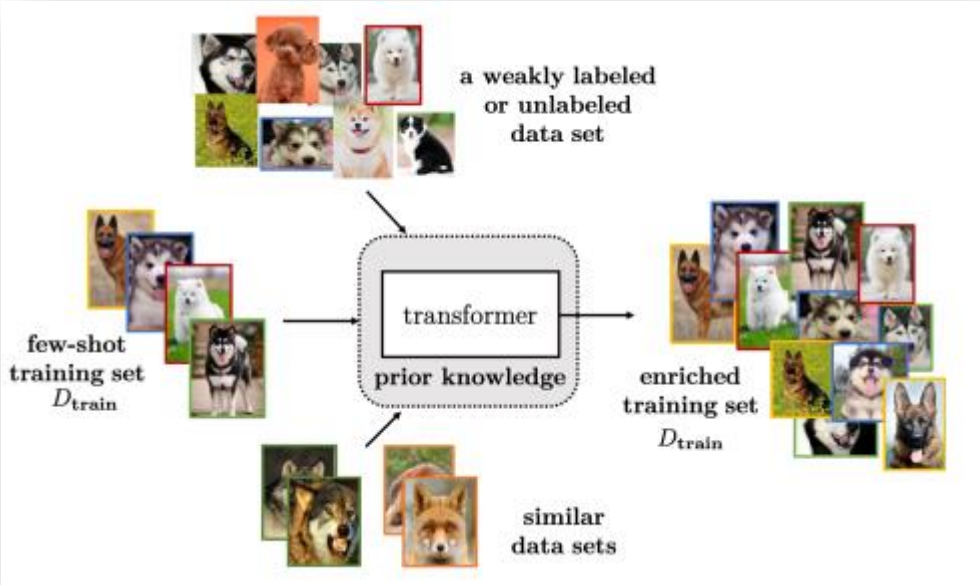
Παρακάτω παρουσιάζεται μία κατηγοριοποίηση πιο προηγμένων τρόπων εμπλουτισμού των συλλογών, με βάση την πηγή προέλευσης των δεδομένων που πρόκειται να προστεθούν στην διαθέσιμη D_{train}. Τα νέα δείγματα που πρόκειται να προστεθούν στην διαθέσιμη συλλογή, αποτελούν προϊόντα μετασχηματισμού των δεδομένων της εκάστοτε πηγής προέλευσης. Έτσι, οι κατηγορίες είναι τρεις και έχουν ως εξής:

Η επέκταση της συλλογής δεδομένων επιτυγχάνεται μέσω

1. μετασχηματισμού δεδομένων που βρίσκονται στην διαθέσιμη συλλογή εκπαίδευσης D_{train}.

2. μετασχηματισμού δεδομένων από μία αδυνάμως κατηγοριοποιημένη ή μη-κατηγοριοποιημένη συλλογή δεδομένων (weakly labeled or unlabeled dataset).

3. μετασχηματισμού δεδομένων από παραπλήσιες συλλογές δεδομένων.



Σχήμα 2.2.1 - 1: Απεικόνιση των τριών κατηγοριών εμπλουτισμού της συλλογής δεδομένων στα προβλήματα Μάθησης Λίγων Λήψεων με διάκριση την πηγή προέλευσης των προστιθέμενων δεδομένων. Ο μετασχηματιστής (transformer) $t(\cdot)$ δέχεται ως είσοδο τα δεδομένα με τις ετικέτες τους (x,y) από την εκάστοτε πηγή προέλευσης και παράγει από αυτές τα νέα δείγματα (x',y') τα οποία επεκτείνουν την συλλογή δεδομένων D_{train}.

Πηγή: Generalizing from a Few Examples: A Survey on Few-Shot Learning Yaqing Wang, Quanming Yao, James T. Kwok, and Lionel M. Ni [55]

2.2.1.1 Δημιουργία Δειγμάτων μέσα από την Συλλογή Δεδομένων D_{train}

Η συγκεκριμένη στρατηγική, μετασχηματίζει τα δεδομένα που βρίσκονται στην συλλογή δεδομένων D_{train} δημιουργώντας διάφορες παραλλαγές οι οποίες αντιστοιχούν στην ίδια ετικέτα.

2.2.1.2 Δημιουργία Δειγμάτων μέσα από μία Αδυνάμως Κατηγοριοποιημένη ή Μη-Κατηγοριοποιημένη Συλλογή Δεδομένων

Πρόκειται για πολιτική η οποία επεκτείνει την D_{train} επιλέγοντας δεδομένα ίδιων ετικετών με αυτών του εκάστοτε προβλήματος Μάθησης Λίγων Λήψεων, από τεράστιες συλλογές οι οποίες είναι αδυνάμως κατηγοριοποιημένες ή μη-κατηγοριοποιημένες. Τέτοιες συλλογές δεδομένων μπορούν να αποτελέσουν οι φωτογραφίες που λαμβάνονται από μία απλή κάμερα επιτήρησης ή η βιντεοσκόπηση μίας ομιλίας. Και στις δύο περιπτώσεις υπάρχει πολλή και διάχυτη μη κατηγοριοποιημένη πληροφορία όπως δρόμοι, αυτοκίνητα, πεζοί στο παράδειγμα των φωτογραφιών και μία σειρά χειρονομιών του ομιλητή στην περίπτωση της καταγραφής.

Η απόκτηση τέτοιων συλλογών συνιστά εύκολη διαδικασία καθώς δεν απαιτείται κάποια ανθρώπινη προσπάθεια για την απόδοση ετικέτας. Επιπλέον, λόγω της ευκολίας στην πρόσβαση τους και χάρη στο μέγεθος τους, περιλαμβάνουν πληθώρα παραλλαγών των οποίων η προσθήκη στην D_{train} παρέχει την δυνατότητα διαμόρφωσης μίας πολύ καλύτερης εικόνας για την κατανομή πιθανοτήτων $p(x,y)$. Το κεντρικό ζήτημα είναι η εύρεση του τρόπου με τον οποίο θα επιλεχθούν τα δείγματα από αυτές τις συλλογές, για την ορθή επέκταση των δεδομένων του προβλήματος Μάθησης Λίγων Λήψεων (δηλαδή ποια δεδομένα θα μετασχηματιστούν και με τι κριτήριο θα επιλέγονται).

2.2.1.3 Δημιουργία Δειγμάτων μέσα από παραπλήσιες Συλλογές Δεδομένων

Η στρατηγική αυτή αυξάνει το μέγεθος της D_{train} συλλέγοντας ζευγάρια δεδομένων εισόδου – ετικετών από εφάμιλλες συλλογές δεδομένων μεγαλύτερης διάστασης. Ο βαθμός εμπλουτισμού της διαθέσιμης D_{train} από την παραπλήσια βοηθητική συλλογή βασίζεται σε ένα συντελεστή ομοιότητας μεταξύ των δειγμάτων που περιέχονται στις διαφορετικές συλλογές. Λαμβάνοντας υπόψη ότι τα νέα δείγματα προέρχονται από μία διαφορετική συλλογή δεδομένων, η άμεση τοποθέτηση τους στην D_{train} μπορεί να οδηγήσει σε αναξιοπιστίες.

Ο όρος του μετασχηματισμού-μετασχηματιστή ενέχει διαφορετική σημασία σύμφωνα με την κατηγορία που εμπίπτει η εκάστοτε εφαρμογή της επέκτασης των δεδομένων. Στην περίπτωση όπου πηγή προέλευσης της επέκτασης είναι τα δεδομένα της D_{train}, ο μετασχηματιστής επιτελεί τον ρόλο μίας συνάρτησης η οποία εφαρμόζεται σε αυτά. Στην περίπτωση όπου ο εμπλουτισμός προέρχεται από δεδομένα μίας αδυνάμως κατηγοριοποιημένης ή μη-κατηγοριοποιημένης συλλογή, αναλαμβάνει χρέη «ταξινομητή» για τα μη κατηγοριοποιημένα δεδομένα. Ενώ στην περίπτωση όπου τα επιπρόσθετα δείγματα απορρέουν από μία παραπλήσια μεγαλύτερη συλλογή, αποτελεί έναν μηχανισμό πρόσμιξης των διαφορετικών συλλογών για την παραγωγή της τελικής συλλογής δεδομένων.

Η επιλογή της στρατηγικής του εμπλουτισμού εξαρτάται από τα δεδομένα που έχουμε στην διάθεση μας και από τον τύπο της εφαρμογής.

Γενικά, η αύξηση των δεδομένων είναι ένας ευθύς και αποτελεσματικός τρόπος αντιμετώπισης του προβλήματος Μάθησης Λίγων Λήψεων. Φυσικό περιορισμό αποτελεί το γεγονός ότι οι παραπάνω στρατηγικές απαιτούν την προσαρμογή σε κάθε συλλογή δεδομένων ξεχωριστά, χωρίς να παρέχεται η ευελιξία επαναχρησιμοποίησης τους για διαφορετικές συλλογές. Επίσης, οι εφαρμογές που επικεντρώνονται στην επέκταση των δεδομένων, μέχρι στιγμής αφορούν κυρίως διεργασίες οι οποίες λαμβάνουν ως δεδομένα εισόδου εικόνες. Οι εφαρμογές εμπλουτισμού των συλλογών σε προβλήματα τα οποία δέχονται εισόδους δεδομένα κειμένου ή ήχου είναι ελάχιστες, καθώς πρόκειται για δομές δεδομένων που είναι πολύ πιο δύσκολο να παραχθούν και να αξιολογηθούν.

2.2.2 Εστίαση στο Μοντέλο

Το ζήτημα της Μάθησης Λίγων Λήψεων είναι ότι η συνάρτηση h^* που αποδίδεται από τα μοντέλα μάθησης αποδεικνύεται εξαιρετικά απλοϊκή για τις απαιτήσεις των ρεαλιστικών προβλημάτων. Έτσι, προτιμάται η αξιοποίηση ενός μεγαλύτερου χώρου υποθέσεων για την μείωση του σφάλματος προσέγγισης $\boldsymbol{\varepsilon_{app}}(\mathcal{H})$ (δηλαδή, την μείωση της απόστασης μεταξύ των h^* και $\hat{h}$). Η επιδίωξη όμως για απόδοση μίας καταλληλότερης υπόθεσης h^* επιφέρει την αύξηση του σφάλματος εκτίμησης $\boldsymbol{\varepsilon_{est}}(\mathcal{H}, I)$. Η δημιουργία ενός νέου μικρότερου χώρου υποθέσεων $\widetilde{\mathcal{H}}$ μπορεί να αντισταθμίσει αυτήν της αύξηση καθώς ο αλγόριθμος θα έχει καλύτερες πιθανότητες απόδοσης μίας υπόθεσης h_I , η οποία θα βρίσκεται κοντύτερα στην βέλτιστη δυνατή h^*.

Κατά την αντιμετώπιση του προβλήματος Μάθησης Λίγων Λήψεων από την σκοπιά του μοντέλου, ακολουθείται ακριβώς αυτή η προσέγγιση. Έτσι, στην παράγραφο αυτήν, εξετάζονται οι στρατηγικές με τις οποίες μπορεί να επιτευχθεί η απόδοση μίας καταλληλότερης h_I μέσω της δημιουργίας ενός νέου περιορισμένου χώρου $\widetilde{\mathcal{H}}$. Με διάκριση την μορφή προγενέστερης γνώσης η οποία χρησιμοποιείται για αυτήν την επιδίωξη, οι μέθοδοι διαχωρίζονται στις εξής κατηγορίες:

- Μάθηση Πολλαπλών Διεργασιών (Multitask Learning),
- Μάθηση «Ενσωμάτωσης» (Embedding Learning),
- Μάθηση με Εξωτερική Μνήμη (Learning with External Memory),
- Παραγωγική Μοντελοποίηση (Generative Modeling).

2.2.2.1 Μάθηση Πολλαπλών Διεργασιών (Multitask Learning)

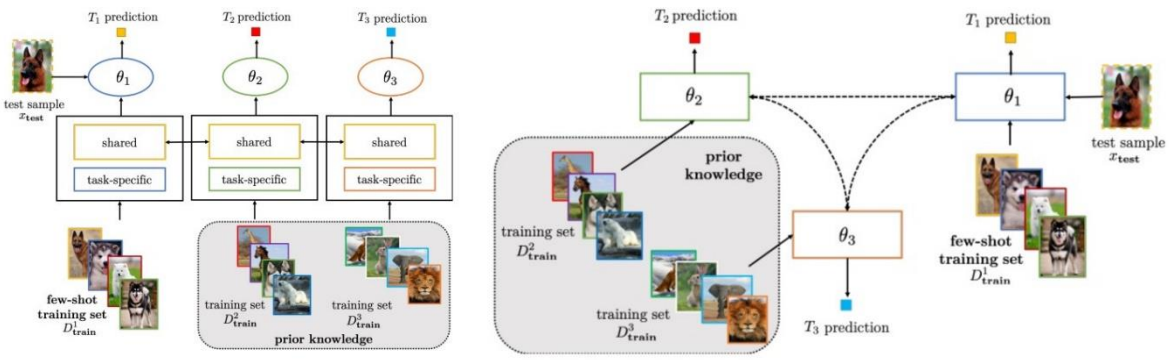
Στην κατηγορία αυτήν, προκειμένου να επιτευχθεί η μείωση του χώρου υποθέσεων H, χρησιμοποιείται ως προγενέστερη γνώση ένα σύνολο από διάφορες, αλλά σχετικές διεργασίες με αυτήν του προβλήματος Μάθησης Λίγων Λήψεων που καλούμαστε να αντιμετωπίσουμε. Η εκμάθηση αυτών των διεργασιών οδηγεί συγχρόνως στην εξαγωγή γενικών αλλά και ειδικότερων πληροφοριών που αφορούν την εκάστοτε διεργασία.

Έστω ότι έχουμε τις C σχετικές διεργασίες T_1, T_2, … , T_C απ' τις οποίες κάποιες περιέχουν λίγα δείγματα και άλλες περισσότερα. Σε κάθε διεργασία T_c, συμπεριλαμβάνεται μία συλλογή δεδομένων $D_c = \{D_{Train}^{\ c}, D_{Test}^{\ c}\}$, όπου $D_{Train}^{\ c}$ η συλλογή δεδομένων εκπαίδευσης, και , $D_{Test}^{\ c}$ η συλλογή δεδομένων αξιολόγησης. Οι διεργασίες που περιέχουν τα ελάχιστα δείγματα χαρακτηρίζονται ως «διεργασίες στόχου» ενώ αυτές με τα επαρκή χαρακτηρίζονται ως «διεργασίες πηγής».

Ζητούμενο της Μάθησης Πολλαπλών Διεργασιών είναι η διαμόρφωση των κατάλληλων παραμέτρων ϑ_c για κάθε διεργασία T_c. Κατά την εκπαίδευση του μοντέλου, οι παράμετροι ϑ_c της αντίστοιχης υπόθεσης h_c, περιορίζονται από όλες τις υπόλοιπες διεργασίες. Με τον τρόπο αυτόν επιτυγχάνεται και η δημιουργία του νέου περιορισμένου χώρου $\widetilde{\mathcal{H}}$.

Σύμφωνα με τον τρόπο περιορισμού των παραμέτρων ϑ_c διαχωρίζουμε τις συγκεκριμένες μεθόδους στις παρακάτω δύο κατηγορίες:

i. Διαμοιρασμός Παραμέτρων (Parameter Sharing): Στην μέθοδο αυτή γίνεται χρήση των ίδιων παραμέτρων ϑ_c , ή ενός συνόλου αυτών των παραμέτρων, για δύο ή περισσότερες διεργασίες.
ii. «Δέσιμο» Παραμέτρων (Parameter Tying): Αφορά μία στρατηγική η οποία εξαναγκάζει την δημιουργία παρόμοιων παραμέτρων ϑ_c , για διαφορετικές διεργασίες.



Σχήμα 2.2.2.1 - 1: Απεικόνιση του περιορισμού των παραμέτρων θ_c για το πρόβλημα του Few-Shot Learning.

Στο αριστερό σκέλος του σχήματος (αριστερή εικόνα), ο περιορισμός επιτυγχάνεται μέσω διαμοιρασμού (περίπτωση Parameter Sharing), ενώ στο δεξί σκέλος (δεξιά εικόνα) ο περιορισμός επιτυγχάνεται μέσω της «δέσμευσης – δεσίματος» (περίπτωση Parameter Tying).

Πηγή: Generalizing from a Few Examples: A Survey on Few-Shot Learning Yaqing Wang, Quanming Yao, James T. Kwok, and Lionel M. Ni [55]

2.2.2.2 Μάθηση «Ενσωμάτωσης» (Embedding Learning)

Κατά την μέθοδο της Μάθησης της «Ενσωμάτωσης», κάθε δείγμα $x_i \in X \subseteq R^d$ κωδικοποιείται σε μίας μικρότερων διαστάσεων (ενσωματωμένη) μορφή $z_i \in Z \subseteq R^m$, όπου $m < d$. Σκοπός αυτής της στρατηγικής είναι ο ευδιάκριτος διαχωρισμός των παρόμοιων και των διαφορετικών δειγμάτων μέσα από την αποτύπωση των κωδικοποιήσεων τους στο νέο χώρο R^m.

Σε αυτήν την περίπτωση, ως προγενέστερη γνώση χρησιμοποιείται η εκάστοτε συνάρτηση κωδικοποίησης, και η επίτευξη του περιορισμού του χώρου υποθέσεων H προκύπτει από την αναγωγή του προβλήματος από τον χώρο X στον μικρότερων διαστάσεων χώρο Z. Επίσης, για την δημιουργία της συνάρτησης κωδικοποίησης μπορεί να αντληθεί και πιο συγκεκριμένη πληροφορία, για το εκάστοτε πρόβλημα αντιμετώπισης λίγων λήψεων, μέσα από την διαθέσιμη συλλογή D_{train}.

Η Μάθηση Ενσωμάτωσης αποτελείται από τις παρακάτω συνιστώσες:

- Μία συνάρτηση *f*, η οποία αποτελεί τον μηχανισμό κωδικοποίησης των δειγμάτων αξιολόγησης $x_{test} \in D_{test}$ στον χώρο Z.
- Μία συνάρτηση *g*, η οποία αποτελεί τον μηχανισμό κωδικοποίησης των δειγμάτων εκπαίδευσης $x_i \in D_{train}$ στον χώρο Z.
- Μία συνάρτηση ομοιότητας *s,* η οποία μετρά τον βαθμό συγγένειας μεταξύ των $f(x_{test})$ και $g(x_i)$ στον χώρο Z.

Η κλάση η οποία πρόκειται να ανατεθεί στο κάθε δείγμα x_{test}, είναι αυτή του δεδομένου x_i, για το οποίο η κωδικοποίηση $g(x_i)$ συσχετίζεται περισσότερο με την κωδικοποίηση $f(x_{test})$ σύμφωνα με την συνάρτηση ομοιότητας *s.*[20]

Ανάλογα με τον τρόπο με τον οποίο λειτουργούν οι συναρτήσεις κωδικοποίησης για τις διάφορες διεργασίες, και τον τρόπο που ποικίλουν οι παράμετροι αυτών, οι μέθοδοι «Ενσωματωμένης» Μάθησης διακρίνονται σε περιπτώσεις όπου χρησιμοποιούνται:

Μοντέλα άμεσα Σχετιζόμενης με την Διεργασία Κωδικοποίησης (Task-Specific Embedding Models)

[20] Παρότι είναι δυνατή η χρησιμοποίηση μίας συνάρτησης κωδικοποίησης για τα δεδομένα εκπαίδευσης και για τα δεδομένα ελέγχου, σύμφωνα με τα ερευνητικά αποτελέσματα ενδείκνυται περισσότερο η χρήση δύο διαφορετικών συναρτήσεων.

Είναι τα μοντέλα, των οποίων οι συναρτήσεις κωδικοποίησης προέρχονται αποκλειστικά μέσα από την συλλογή δεδομένων του προβλήματος που καλούνται να αντιμετωπίσουν.

Μοντέλα Ανεξάρτητης από την Διεργασία Κωδικοποίησης (Task-Invariant Embedding Models)

Στις μεθόδους ανεξάρτητης από την διεργασία κωδικοποίησης εκπαιδεύεται ένα μοντέλο γενικής κωδικοποίησης μέσα από μία μεγάλη συλλογή δεδομένων, το οποίο δεν αξιοποιεί καθόλου τα δεδομένα της συλλογής D_{train} του προβλήματος Λίγων Λήψεων, παρά μόνο για να τα κωδικοποιήσει. Αυτό σημαίνει πως το μοντέλο κατά κανόνα επιλύει γενικές διεργασίες και δεν εκπαιδεύεται καθόλου στα δεδομένα του εκάστοτε προβλήματος αντιμετώπισης. Το πρώτο few-shot embedding μοντέλο, κωδικοποιούσε τα δείγματα με την χρήση ενός «πυρήνα» [59], ενώ αργότερα δημιουργήθηκαν και πιο σύνθετες κωδικοποιήσεις όπως η προσέγγιση των Σιαμαίων Συνελικτικών Δικτύων [13].

Παρά το γεγονός ότι οι παράμετροι του χρησιμοποιούμενου μοντέλου κωδικοποίησης δεν επηρεάζονται από την συλλογή D_{train} της διεργασίας Λίγων Λήψεων, έχουν γίνει πολλές απόπειρες προσομοίωσης του προβλήματος ελαχίστων λήψεων, διασπώντας τις πολλών κλάσεων συλλογές δεδομένων σε περισσότερες συλλογές δεδομένων με λιγότερες κλάσεις. Το μοντέλο εκπαιδεύεται σε αυτές τις περισσότερες συλλογές δεδομένων με τις λιγότερες πλέον κλάσεις με σκοπό να αποκτήσει μία καλή δυνατότητα γενίκευσης για προβλήματα μάθησης λίγων λήψεων.

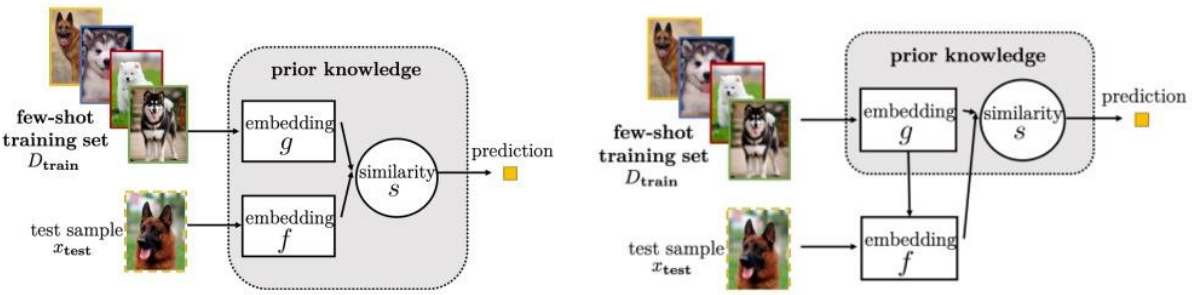
Σε μία από τις πρώτες απόπειρες, μαθαίνονταν μία γραμμική συνάρτηση κωδικοποίησης [60]. Προσφάτως, έχουν δημιουργηθεί πιο σύνθετες μέθοδοι ανεξάρτητης από την διεργασία κωδικοποίησης μέσω της διαδικασίας της μέτα-μάθησης, όπως:

- Τα Δίκτυα Αντιστοίχισης [15] (Matching Networks) και παραλλαγές αυτών,
- Τα Δίκτυα Πρωτοτύπων [16] (Prototypical Networks) και οι διάφορες παραλλαγές αυτών,
- Τα Δίκτυα Συσχέτισης [14] (Relation Networks),
- Οι επαναλαμβανόμενης προσοχής «Συγκριτές» [61] (Attentive Recurrent Comparators).

Μοντέλα Υβριδικής Κωδικοποίησης (Hybrid Embedding Models)

Αφορά έναν συνδυασμό των δύο παραπάνω τρόπων κωδικοποίησης. Στις μεθόδους Υβριδικής κωδικοποίησης το εκπαιδευμένο μοντέλο ανεξάρτητης από την διεργασία κωδικοποίησης προσαρμόζει τις παραμέτρους του σύμφωνα με την συλλογή δεδομένων D_{train} του συγκεκριμένου προβλήματος μάθησης λίγων λήψεων (μέσω της συνάρτησης κωδικοποίησης f των δειγμάτων αξιολόγησης).

Είναι μία στρατηγική η οποία επέφερε αξιοσημείωτες βελτιώσεις στα προβλήματα Few-Shot Learning καθώς οι ιδιαιτερότητες και οι ιδιομορφίες που υπάρχουν στα ελάχιστα διαθέσιμα δεδομένα είναι κομβικές για την επίλυση της εκάστοτε διεργασίας (όπως για παράδειγμα στην μάθηση που αφορά δεδομένα σπανίων περιπτώσεων).



Σχήμα 2.2.2.2 - 1: Απεικόνιση δύο μοντέλων μάθησης ενσωμάτωσης για την λύση του προβλήματος Few-Shot Learning, όπου η αριστερή περίπτωση εντάσσεται στα Μοντέλα Ανεξάρτητης από την Διεργασία Κωδικοποίησης, ενώ η δεξιά περίπτωση στα Μοντέλα Υβριδικής Κωδικοποίησης.

Όπως απεικονίζεται, στο μοντέλο Υβριδικής Κωδικοποίησης, η συνάρτηση κωδικοποίησης f των δεδομένων ελέγχου προέρχεται μέσα από τα δεδομένα του προβλήματος Μάθησης Λίγων Λήψεων, σε αντίθεση με το μοντέλο Ανεξάρτητης από την Διεργασία Κωδικοποίησης, όπου η f εμπεριέχεται στην χρησιμοποιούμενη προγενέστερη γνώση.

Πηγή: Generalizing from a Few Examples: A Survey on Few-Shot Learning Yaqing Wang, Quanming Yao, James T. Kwok, and Lionel M. Ni [55]

2.2.2.3 Μάθηση με Εξωτερική Μνήμη (Learning With External Memory)

Πρόκειται για πολύ όμοια στρατηγική με αυτήν της Μάθησης «Ενσωμάτωσης», όσον αφορά τον τρόπο λειτουργίας. Η διαφορά έγκειται στο γεγονός της χρήσης μίας βοηθητικής εξωτερικής μνήμης.

Οι μνήμες που χρησιμοποιούνται σε αυτές τις περιπτώσεις είναι τύπου κλειδιού - τιμής (key – value). Δηλαδή, κάθε θέση των μνημών αυτών περιέχει δύο τιμές. Η πρώτη (κλειδί – key) περιέχει την κωδικοποίηση ενός δεδομένου εκπαίδευσης $f(x_i)$, ενώ η δεύτερη (τιμή – value) περιλαμβάνει μία τιμή αντιστοίχισης σε αυτήν (η οποία περιέχει πληροφορία για την ετικέτα y_i).

Τα δεδομένα x_{test} κωδικοποιούνται μέσω της συνάρτησης f, αποδίδοντας την κωδικοποίηση $f(x_{test})$, η οποία συγκρίνεται με τις διάφορες κωδικοποιήσεις $f(x_i)$ που εμπεριέχονται στην μνήμη. Στην συνέχεια με βάση τις συγκρίσεις και σύμφωνα με μία συνάρτηση ομοιότητας s, εξάγεται ένας συντελεστής συνεισφοράς για κάθε μία από τις θέσεις μνήμης.

Έπειτα, σχηματίζεται ένα υβριδικό δεδομένο που αντιστοιχίζεται στο δείγμα ελέγχου x_{test}, σύμφωνα με τον συντελεστή συνεισφοράς που εξάχθηκε, το οποίο εισέρχεται σε έναν «ταξινομητή» (όπως είναι η συνάρτηση ενεργοποίησης Softmax) προκειμένου να γίνει η πρόβλεψη - κατηγοριοποίηση του.[21]

Ως προγενέστερη γνώση χρησιμοποιείται η συνάρτηση κωδικοποίησης f και ο τρόπος αλληλεπίδρασης με την εξωτερική μνήμη. Η επίτευξη του περιορισμού του χώρου υποθέσεων Η επιτυγχάνεται χάριν στην κατ' αποκλειστικότητα αναπαράσταση των δεδομένων αξιολόγησης x_{test}, μέσα από τα περιεχόμενα της μνήμης.

Αξίζει να σημειωθεί, πως η χρήση μνήμων και η αλληλεπίδραση με αυτές είναι αρκετά δαπανηρή, συνεπώς το μέγεθος τους δεν πρέπει να είναι πολύ μεγάλο. Με βάση αυτόν τον περιορισμό, αλλά και το γεγονός ότι τα στοιχεία της μνήμης είναι αποκλειστικά αυτά που διαμορφώνουν την αναπαράσταση των δεδομένων x_{test}, η επιλογή των στοιχείων που θα συμπεριληφθούν στην μνήμη είναι κομβικής σημασίας.

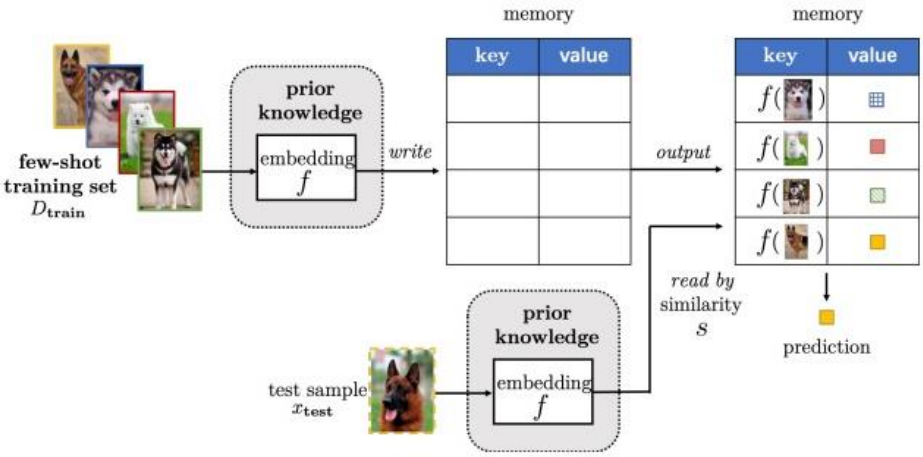
Σύμφωνα με την λειτουργικότητα της μνήμης, τα προβλήματα Μάθησης Ελάχιστων Λήψεων αυτής της κατηγορίας διαχωρίζονται σε περιπτώσεις όπου η μνήμη συμπεριλαμβάνει:

- «Επαναβελτιωμένες» Αναπαραστάσεις (Refining Representations)

[21] Δηλαδή, το αποτέλεσμα της συνάρτησης κωδικοποίησης $f(x_{test})$, δεν χρησιμοποιείται ως άμεσο ερώτημα κατηγοριοποίησης σύμφωνα με μία συνάρτηση ομοιότητας s με την αντίστοιχη $g(x_{test})$, όπως γινόταν στην περίπτωση της Μάθησης «Ενσωμάτωσης», αλλά ως ερώτημα στα περιεχόμενα της μνήμης προκειμένου να προκύψει το δεδομένο που θα χρησιμοποιηθεί ως το βασικό ερώτημα κατηγοριοποίησης.

Η μνήμη συμπεριλαμβάνει στοιχεία που οδηγούν στην αποτελεσματικότερη αναπαράσταση των ελάχιστων δεδομένων.

- «Επαναβελτιωμένες» Παραμέτρους (Refining Parameters): Η μνήμη περιέχει στοιχεία που θα βοηθήσουν άμεσα στην αποτελεσματικότερη τροποποίηση των παραμέτρων ϑ.



Σχήμα 2.2.2.3 - 1: Απεικόνιση αντιμετώπισης του προβλήματος Μάθησης Λίγων Λήψεων μέσω Μάθησης «Ενσωμάτωσης» με χρήση Εξωτερικής Μνήμης, όπου για κάθε δεδομένο x_i της περιορισμένης συλλογής D_{train}, η μνήμη καταχωρεί τις κωδικοποιήσεις των δεδομένων $f(x_i)$ στην θέση του κλειδιού και τις ετικέτες y_i στην θέση της τιμής, προκειμένου να γίνει η πρόβλεψη.

Πηγή: Generalizing from a Few Examples: A Survey on Few-Shot Learning Yaqing Wang, Quanming Yao, James T. Kwok, and Lionel M. Ni [55]

2.2.2.4 Παραγωγική Μοντελοποίηση (Generative Modeling)

Οι μέθοδοι παραγωγικής μοντελοποίησης προσεγγίζουν την πιθανοτική κατανομή $p(x)$[22] των εισερχόμενων δεδομένων x_i, με την βοήθεια προγενέστερης γνώσης. Η επίτευξη της προσέγγισης αυτής, συνήθως συνεπάγεται και την γνώση-προσέγγιση των κατανομών $p(x|y)$ και $p(y)$.

Οι μέθοδοι αυτής της κατηγορίας μπορούν να αντιμετωπίσουν διάφορες διεργασίες όπως:

- η παραγωγή-δημιουργία νέων δεδομένων (Data Generation),
- η αναγνώριση δεδομένων (Data Recognition),
- η ανακατασκευή δεδομένων (Data Reconstruction), και
- το αναποδογύρισμα εικόνας (Image Flipping).

Κατά την στρατηγική αυτήν, λαμβάνεται ως δεδομένο πως οι παρατηρήσεις x_i προέρχονται από μία πιθανοτική κατανομή $p(x; \theta)$, όπου θ οι παράμετροι που καθορίζουν την κατανομή $p(x)$. Συνήθως στις μεθόδους της παραγωγικής μοντελοποίησης, εντάσσεται μία βοηθητική ενδιάμεση μεταβλητή z, η οποία αντιστοιχίζεται στην κατανομή $p(z; \gamma)$, όπου γ οι παράμετροι που καθορίζουν την κατανομή $p(z)$. Η επιλογή της μεταβλητής z γίνεται με τέτοιον τρόπο, ώστε για τις εισόδους x_i να ικανοποιείται η σχέση:

$$p(x) = \int p(x|z;\ \theta)\, p(z;\gamma)\, dz$$

[22] Ο όρος πιθανοτική κατανομή μιας τυχαίας μεταβλητής x, συμπεριλαμβάνει τις τιμές που μπορεί να λάβει η μεταβλητή x, και την συχνότητα εμφάνισης αυτών.

Δηλαδή, η ζητούμενη κατανομή p(x) να προσεγγίζεται ικανοποιητικά με την βοήθεια (της κατανομής) της ενδιάμεσης μεταβλητής z.

Ως προγενέστερη γνώση στις μεθόδους παραγωγικής μοντελοποίησης χρησιμοποιείται η κατανομή p(z; γ) και οι παράμετροι γ, των οποίων η εκμάθηση επιτυγχάνεται μέσα από άλλες συλλογές δεδομένων. Ο περιορισμός του χώρου υποθέσεων H, επέρχεται μέσω του συνδυασμού της προγενέστερης αυτής γνώσης και της αξιοποίησης της διαθέσιμης συλλογής δεδομένων του προβλήματος ελάχιστων λήψεων, D_{train}.

Οι μέθοδοι παραγωγικής μοντελοποίησης ομαδοποιούνται σε τρία είδη, σύμφωνα με το αν η βοηθητική μεταβλητή z αντιπροσωπεύει:

1.Αποσυνθεμένες Συνιστώσες (Decomposable Components)

Στην κατηγορία αυτήν, η βοηθητική μεταβλητή z επιτελεί τον ρόλο αποσυνθεμένων συνιστωσών. Με τον όρο αποσυνθεμένες συνιστώσες εννοούμε τον διαχωρισμό των δομικών στοιχείων που απαρτίζουν τα δεδομένα. Παρά το γεγονός πως στα προβλήματα λίγων λήψεων τα δείγματα είναι ελάχιστα, μπορούν να μοιράζονται κοινά χαρακτηριστικά με δείγματα από άλλες συλλογές.

Για παράδειγμα, σε ένα πρόβλημα αναγνώρισης προσώπου με ελάχιστες διαθέσιμες φωτογραφίες, η εύρεση παρόμοιων προσώπων δεν συνιστά και πολύ εύκολη διαδικασία. Από την άλλη, η εύρεση φωτογραφιών στις οποίες αναπαρίστανται παρόμοια χαρακτηριστικά προσώπου (μάτια, μύτη, στόμα κι ούτω καθεξής) αποτελεί έναν πολύ προσιτό στόχο. Στην προκειμένη περίπτωση, τα προσωπικά χαρακτηριστικά είναι αυτά που αποτελούν τις αποσυνθεμένες συνιστώσες.

Με την ύπαρξη ενός επαρκούς αριθμού δειγμάτων, μπορούν εύκολα να εκπαιδευτούν μοντέλα για την εκμάθηση των αποσυνθεμένων χαρακτηριστικών. Έπειτα, εκκρεμεί ο κατάλληλος συνδυασμός αυτών, για την παραγωγή μίας παρόμοιας εικόνας με αυτές της συλλογής D_{train}, και η ανάθεση της κλάσης στην οποία αυτή ανήκει. Πρόκειται για μία στρατηγική ερμηνεύσιμη και αναγνώσιμη από τον άνθρωπο, καθώς η επιλογή των αποσυνθεμένων συνιστωσών για το εκάστοτε πρόβλημα, υλοποιείται από τον ίδιο.

2.Βοηθητικές Πιθανοτικές Κατανομές από παρόμοιες Διεργασίες (Groupwise Shared Prior)

Πολύ συχνά, οι παρόμοιες διεργασίες χαρακτηρίζονται και από παρόμοιες πιθανοτικές κατανομές των δεδομένων τους, γεγονός που μπορεί να αξιοποιηθεί στα προβλήματα Λίγων Λήψεων. Παραδείγματος χάριν, έστω τρία προβλήματα με συλλογές δεδομένων που περιέχουν «πορτοκαλί γάτες», «λεοπαρδάλεις» και «τίγρεις της Βεγγάλης» αντίστοιχα. Θα υποθέσουμε ότι η διεργασία με την συλλογή που περιλαμβάνει τις τίγρεις της Βεγγάλης συγκροτεί ένα πρόβλημα λίγων λήψεων, καθώς τα δεδομένα της αφορούν ένα είδος ζώου που βρίσκεται υπό εξαφάνιση.

Για την προσέγγιση της πιθανοτικής κατανομής p(x) των δεδομένων που αποτυπώνουν τις τίγρεις της Βεγγάλης, θα μπορούσε να αξιοποιηθεί ως προγενέστερη γνώση (ενδιάμεση μεταβλητή z) η εκμάθηση των πιθανοτικών κατανομών των συλλογών με τις πορτοκαλί γάτες και τις λεοπαρδάλεις καθώς πρόκειται για τρεις συγγενικές κατηγορίες.

3.Παραμέτρους Δικτύων Συμπερασμάτων (Parameters of Inference Networks)

Σε αυτήν την κατηγορία λαμβάνεται ως βάση, ότι η εύρεση των βέλτιστων τιμών των θ (παράμετροι οι οποίες καθορίζουν την κατανομή p(x)) ισοδυναμεί με την μεγιστοποίηση της πιθανοτικής κατανομής p(z|x; θ, γ), για την οποία ισχύει:

$$p(z|x;\ \theta,\gamma\ ) = \frac{p(x,z;\ \theta,\gamma\ )}{p(x;\gamma\ )} = \frac{p(x\ |z;\ \theta)\ p(z;\gamma\ )}{\int p(x\ |z;\ \theta)\ p(z;\gamma\ )\ dz}$$

Λόγω της ύπαρξης του ολοκληρώματος στον παρονομαστή της παραπάνω ισότητας, ο υπολογισμός της $p(z|x;\ \theta,\gamma)$ καθίσταται δύσκολο εγχείρημα. Την δυσκολία αυτήν έρχεται να κατευνάσει μία νέα εναλλακτική κατανομή q(z; δ), η οποία μαθαίνεται μέσα από τα «δίκτυα συμπερασμάτων» [62] (Inference Networks) και καλείται να προσεγγίσει την $p(z|x;\ \theta,\gamma)$.

Στην προκειμένη περίπτωση η μεταβλητή z, δεν αντιστοιχίζεται σε κάποιο νοηματικό περιεχόμενο όπως στις προηγούμενες κατηγορίες, αλλά στις παραμέτρους του εκάστοτε «δικτύου συμπερασμάτων».

Παραδείγματα δικτύων συμπερασμάτων, τα οποία υϊοθετήθηκαν για το πρόβλημα λίγων λήψεων μέσω της παραγοντικής μοντελοποίησης, συγκροτούν:

- οι Στοχαστικοί Αυτοκωδικοποιητές [63] (Variational Autoencoder),
- τα «Αύτο-παλινδρομικά» Μοντέλα [64] (Autoregressive Models), και
- τα Παραγωγικά Αντιπαραθετικά Δίκτυα [65] (Generative Adversarial Networks - GANs).

2.2.3 Εστίαση στον Αλγόριθμο

Οι στρατηγικές που εμπεριέχονται στην αντιμετώπιση των Προβλημάτων Ελάχιστων Λήψεων υπό το πρίσμα του Αλγορίθμου, αποσκοπούν στην καλύτερη αναζήτηση των παραμέτρων θ, για την όσο το δυνατόν καλύτερη προσέγγιση της υπόθεσης h^*. Κατά τον Αλγόριθμο «Βαθμιαίας Καθόδου» (Gradient Decent Algorithm) οι παράμετροι θ ενός δικτύου, κατά την επανάληψη t, τροποποιούνται ως εξής:

$$\theta_t = \theta_{t-1} - a_t \nabla_{\theta_{t-1}} l(h(x_t; \theta_{t-1}), y_t)$$

Όπου

θ_t: η νέα τιμή της εκάστοτε τροποποιούμενης παραμέτρου θ,

θ_{t-1}: η προηγούμενη τιμή της εκάστοτε τροποποιούμενης παραμέτρου θ,

α_t: το χρησιμοποιούμενο βήμα - ρυθμός μάθησης (Learning Rate-Step Size) του αλγορίθμου βελτιστοποίησης (Optimization Algorithm),

$\nabla_{\theta_{t-1}} l(h(x_t; \theta_{t-1}), y_t)$: η βαθμίδα της αντίστοιχης Συνάρτησης Κόστους ως προς την συγκεκριμένη παράμετρο θ.

Αρχικοποιώντας την παράμετρο θ στην τιμή θ_0, η εξίσωση τροποποίησης του αλγορίθμου γίνεται:

$$\theta_t = \theta_0 + \sum_{i=1}^{t} \Delta\theta_{i-1}$$

Στα προβλήματα Ελαχίστων Λήψεων, τα δεδομένα δεν είναι επαρκή ώστε να τροποποιηθούν αξιόπιστα οι παράμετροι θ. Οι μέθοδοι αυτής της στρατηγικής, χρησιμοποιούν προγενέστερη γνώση για την αποτελεσματικότερη τροποποίηση των παραμέτρων των μοντέλων.

Σύμφωνα με τον τρόπο με τον οποίο επηρεάζεται η τεχνική αναζήτησης των βέλτιστων παραμέτρων θ, διακρίνουμε της μεθόδους αυτής της τεχνικής σε τρεις κατηγορίες:

1. Ρύθμιση Υπαρχόντων Παραμέτρων (Refining Existing Parameters): Μαθαίνονται οι αρχικές παράμετροι θ_0 από άλλες διεργασίες, και στην συνέχεια ρυθμίζονται με βάση την περιορισμένη συλλογή δεδομένων D_{train}.
2. Ρύθμιση Μέτα-Μαθημένων Παραμέτρων (Refining Meta-Learned Parameter): Οι αρχικές παράμετροι θ_0 μέτα-μαθαίνονται από ένα πλήθος διεργασιών οι οποίες είναι από την ίδια

κατανομή διεργασιών με το εκάστοτε αντιμετωπιζόμενο πρόβλημα, και έπειτα γίνεται ρύθμιση αυτών με βάση τα διαθέσιμα δεδομένα της D_{train}.

3. Μάθηση του Αλγορίθμου Βελτιστοποίησης (Learning the Optimizer): Πρόκειται για στρατηγική όπου μέτα-μαθαίνονται τα χαρακτηριστικά του Αλγορίθμου βελτιστοποίησης, όπως ο ρυθμός μάθησης που θα χρησιμοποιηθεί και οι κατευθύνσεις τροποποίησης οι οποίες πρόκειται να ακολουθηθούν.

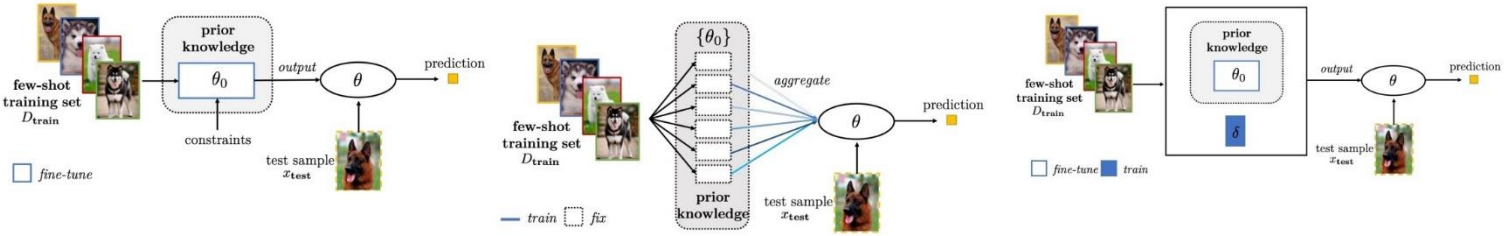
2.2.3.1 Ρύθμιση Υπαρχόντων Παραμέτρων (Refining Existing Parameters)

Στην στρατηγική αυτή λαμβάνονται ως αρχικές παράμετροι θ_0 , οι παράμετροι ενός προ-εκπαιδευμένου μοντέλου και έπειτα γίνεται ρύθμιση αυτών στις αντίστοιχες συλλογές δεδομένων των προβλημάτων αντιμετώπισης. Η λογική είναι ότι οι εν λόγω αρχικές παράμετροι θ_0 ανταποκρίνονται σε μεγάλης κλίμακας - γενικά χαρακτηριστικά και είναι εύκολο με μόλις λίγες επαναλήψεις εκπαίδευσης να προσαρμοστούν στα ειδικά δεδομένα της εκάστοτε διεργασίας.

Στην περιοχή αυτήν συναντώνται τρεις προσεγγίσεις:

1. Ρύθμιση των Παραμέτρων μέσω Τεχνικών Ομαλοποίησης (Fine-Tuning Existing Parameter by Regularization)
   Μετά την απόκτηση και κατά την ρύθμιση - προσαρμογή των παραμέτρων θ_0 στα ελάχιστα διαθέσιμα δεδομένα, ελλοχεύει μεγάλη πιθανότητα εμφάνισης του φαινομένου Overfitting. Προκειμένου να αποφευχθεί αυτό, η ρύθμιση γίνεται μέσω τεχνικών Ομαλοποίησης (Regularization Methods) όπως
   - η Έγκαιρη Διακοπή (Early Stopping),
   - η επιλεκτική τροποποίηση των παραμέτρων θ_0 ,
   - η ομαδοποίηση των σχετιζόμενων παραμέτρων και η σύγχρονη ανανέωση αυτών, ή
   - η χρήση ενός Δικτύου Παλινδρόμησης (Model Regression Network).
2. Πρόσμιξη μέσα από ένα Σύνολο Παραμέτρων (Aggregating a Set of Parameters)
   Η προσέγγιση αυτή συναντάται όταν δεν υπάρχουν κατάλληλες θ_0 για αρχικοποίηση, αλλά υπάρχουν εκπαιδευμένα μοντέλα με αντίστοιχες παραμέτρους σε παρόμοιες διεργασίες. Έτσι, αντί για συγκεκριμένες παραμέτρους θ_0, παρέχεται ένα σύνολο αρχικών παραμέτρων, και η συλλογή δεδομένων D_{train} χρησιμοποιείται για την εύρεση του κατάλληλου συνδυασμού αυτών.
3. Ρύθμιση των υπαρχόντων Παραμέτρων με την προσθήκη Νέων Παραμέτρων (Fine-Tuning Existing Parameter with New Parameters)
   Εισάγονται νέες βοηθητικές παράμετροι, προκειμένου να μαθευτούν και να προσαρμοστούν κατάλληλα οι αρχικές θ_0 στο πρόβλημα Μάθησης Λίγων Λήψεων.

Κατά την χρησιμοποίηση των έτοιμων παραμέτρων θ_0, μειώνεται σε μεγάλο βαθμό η δυσκολία της αναζήτησης της κατάλληλης υπόθεσης h^* μέσα από τον χώρο υποθέσεων $\mathcal{H}$, αλλά και το υπολογιστικό κόστος που απαιτείται. Παρ' όλ' αυτά, καθώς αυτές οι παράμετροι προέρχονται από ένα διαφορετικό πρόβλημα, υπάρχει ο κίνδυνος να θυσιάζεται η αποτελεσματικότητα εις βάρος της ταχύτητας.



Σχήμα 2.2.3.1 - 1: Απεικόνιση των παραπάνω τριών προσεγγίσεων ρύθμισης υπαρχόντων παραμέτρων στην επίλυση του προβλήματος μάθησης λίγων λήψεων. Το κυκλικό κομμάτι που εμπεριέχει τις τελικές παραμέτρους θ είναι κοινό και για τις τρεις προσεγγίσεις και αναπαριστά το σκέλος του ελέγχου-αξιολόγησης του προβλήματος.

Αριστερά, απεικονίζεται η μέθοδος ρύθμισης των παραμέτρων θ_0 μέσω τεχνικών ομαλοποίησης.

Στην μέση, αναπαρίσταται η μέθοδος πρόσμιξης μέσα από ένα σύνολο παραμέτρων. Ως προγενέστερη γνώση παρέχεται ένα σύνολο αρχικοποιημένων παραμέτρων θ_0 και το μοντέλο καλείται να βρει τον καταλληλότερο συνδυασμό - πρόσμιξη αυτών.

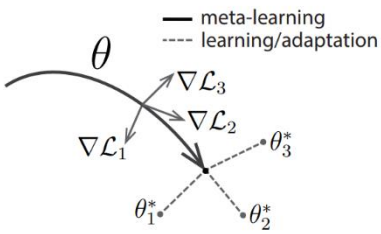
Δεξιά, φαίνεται η τεχνική ρύθμισης των υπαρχόντων παραμέτρων με την προσθήκη νέων. Οι νέες βοηθητικές προστιθέμενες παράμετροι συμβολίζονται ως δ και συνεισφέρουν στην εκμάθηση-οδήγηση των βέλτιστων τελικών παραμέτρων θ.

Πηγή: Generalizing from a Few Examples: A Survey on Few-Shot Learning Yaqing Wang, Quanming Yao, James T. Kwok, and Lionel M. Ni [55]

2.2.3.2 Ρύθμιση Μέτα-Μαθημένων Παραμέτρων (Refining Meta-Learned Parameter)

Οι μέθοδοι αυτής της προσέγγισης χρησιμοποιούν την Μέτα-Μάθηση για την απόκτηση των αρχικών παραμέτρων θ_0. Η λογική είναι, να αποκτηθούν μέσα από ένα σύνολο πολλών διεργασιών τέτοιες παράμετροι θ_0, ώστε να είναι εύκολο με μόλις λίγες επαναλήψεις του αλγόριθμου βελτιστοποίησης να πλησιάσουν τις επιθυμητές παραμέτρους θ^* και να προσαρμοστούν ικανοποιητικά στο πρόβλημα μάθησης λίγων λήψεων.

Χαρακτηριστική συμβολή, είναι αυτή των Chelsea Finn, Pieter Abbeel και Sergey Levine με την μέθοδο MAML (Model-Agnostic Meta Learning) [17].



Σχήμα 2.2.3.2 - 1: Απεικόνιση της ευελιξίας προσαρμογής των μετα-μαθημένων παραμέτρων θ_0 σε τρεις διαφορετικές διεργασίες T_1, T_2 ,T_3. Η προσαρμογή λαμβάνει χώρα σε μόλις λίγες επαναλήψεις βελτιστοποίησης, σύμφωνα με την εκάστοτε κλίση της συνάρτησης κόστους για κάθε μία από αυτές τις διεργασίες. Η κουκίδα στην μέση συμβολίζει τις μέτα-μαθημένες αρχικοποιημένες παραμέτρους θ_0, ενώ οι θ_1^*, θ_2^* , θ_3^* αναπαριστούν τις βέλτιστες παραμέτρους για κάθε διεργασία αντίστοιχα.

Πηγή: Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks - Chelsea Finn, Pieter Abbeel, Sergey Levine [17]

Μερικές βελτιστοποιήσεις οι οποίες επήλθαν χάριν στην μέθοδο MAML είναι:

- Η άντληση συγκεκριμένων πληροφοριών σύμφωνα με την διεργασία αντιμετώπισης για την παροχή καλύτερων μετα-μαθημένων αρχικών παραμέτρων θ_0,
- Η διευθέτηση της αβεβαιότητας που επέρχεται εξαιτίας της χρήσης μέτα-μαθημένων παραμέτρων θ_0,
- Η βελτίωση της διαδικασίας προσαρμογής στο πρόβλημα.

2.2.3.3 Μάθηση του Αλγορίθμου Βελτιστοποίησης (Learning the Optimizer)

Οι μέθοδοι της προηγούμενης παραγράφου αποκτούσαν μέσω της μέτα-μάθησης τις αρχικές παραμέτρους θ_0, και γινόταν προσαρμογή στο πρόβλημα αντιμετώπισης, μέσω των κλασσικών αλγορίθμων βελτιστοποίησης σύμφωνα με την διαθέσιμη συλλογή D_{train}.

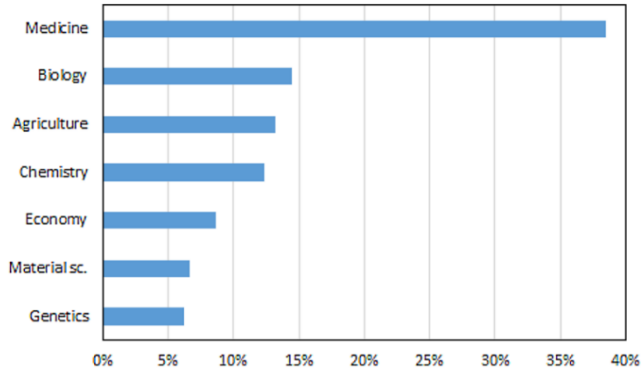
Αντίθετα, στις μεθόδους αυτής της στρατηγικής [66, 67], αντί για την χρησιμοποίηση των συνηθισμένων αλγορίθμων βελτιστοποίησης, μαθαίνεται η ακριβής μεταβολή των παραμέτρων θ. Έτσι, απαλλασσόμαστε από την επιλογή του αλγορίθμου βελτιστοποίησης αλλά και την ανάγκη της κατάλληλης επιλογής του βήματος-ρυθμού μάθησης του (Learning Rate).

Ουσιαστικά, χρησιμοποιείται μέτα-μάθηση για την απόκτηση του Αλγορίθμου Βελτιστοποίησης, και αντί για τις παραμέτρους θ_0, εδώ δίνεται ως προγενέστερη γνώση ένας μέτα-μαθημένος αλγόριθμος βελτιστοποίησης ο οποίος έχει εξαχθεί από εφάμιλλες διεργασίες με την αντιμετωπιζόμενη. Στην συνέχεια ακολουθεί παρόμοια προσαρμογή της προγενέστερης γνώσης πάνω στα διαθέσιμα δεδομένα του εκάστοτε αντιμετωπιζόμενου προβλήματος.

2.3 Εφαρμογές

Όπως είδαμε, η Μάθηση Λίγων Λήψεων, ενέχει κομβικό ρόλο σε προβλήματα που αποτελούνται από σπάνια - δυσεύρετα δεδομένα, ή όπου κρίνεται απαραίτητη η ελαχιστοποίηση της καταβολής ανθρώπινης προσπάθειας για την απόκτηση δεδομένων και της μείωσης του υπολογιστικού κόστους.

Έτσι, χάρη στην δομή της, τα χαρακτηριστικά, αλλά και την σημαντικότητα της στις ανάγκες της εποχής, η Μάθηση Λίγων Λήψεων εκτείνεται σε ένα ευρύ φάσμα πρακτικών εφαρμογών και επιστημονικών περιοχών, όπως είναι η Φαρμακευτική, η Βιολογία, η Γεωργία, η Χημεία, η Οικονομία, η Επιστήμη των Υλικών και η Γενετική.



Σχήμα 2.3 - 1: Ιστόγραμμα απεικόνισης του ποσοστού των δημοσιευμένων εργασιών Few-Shot Learning για κάθε μία από τις παραπάνω επιστημονικές περιοχές.

Πηγή: A survey of few-shot learning in smart agriculture: developments, applications, and challenges - Jiachen Yang, Xiaolan Guo, Yang Li, Francesco Marinello, Sezai Ercisli, Zhuo Zhang [57]

Στην ενότητα αυτήν παρουσιάζονται εφαρμογές του Few Shot Learning, στα διάφορα πεδία ανάπτυξης και εφαρμογών της Μηχανικής Μάθησης. Πιο συγκεκριμένα, αναφέρονται παραδείγματα αξιοποίησης του, στις περιοχές:

- Της Μηχανικής Όρασης (Computer Vision),
- Της Ρομποτικής (Robotics),
- Της Επεξεργασίας Φυσικής Γλώσσας (Natural Language Processing),
- Της Επεξεργασίας Ήχου (Acoustic Signal Processing), και

- Άλλων (Επιπρόσθετων και Μη-Κατηγοριοποιημένων).

Επίσης, όπου κρίνεται σκόπιμο, παρατίθενται και οι συγκεκριμένες εργασίες οι οποίες έχουν εισαγάγει τις εκάστοτε εφαρμογές.

Από τις παραπάνω περιοχές, η Μηχανική Όραση είναι αυτή η οποία πρωταγωνιστεί στις πρακτικές εφαρμογές Few-Shot Learning με εμφατική διαφορά πλήθους δημοσιευμένων εργασιών από τις υπόλοιπες. Αυτό φυσικά δεν αναιρεί σε καμία περίπτωση την σημαντικότητα και την συμβολή της Μάθησης Λίγων Λήψεων στην ανάπτυξη και των άλλων περιοχών.

2.3.1 Μηχανική Όραση (Computer Vision)

Πρόκειται για την περιοχή εφαρμογών, στην οποία συναντάται το μεγαλύτερο πλήθος εργασιών Μάθησης Λίγων Λήψεων[23]. Οι δύο περιοχές διεργασιών Μηχανικής Όρασης, στις οποίες βρίσκει μεγαλύτερη εφαρμογή η Μάθηση Λίγων Λήψεων είναι:

1. η Οπτική Αναγνώριση Χαρακτήρων (Optical Character Recognition), και
2. η Κατηγοριοποίηση Εικόνων (Image Classification).

Οι επιδόσεις που έχουν επιτευχθεί σε συλλογές δεδομένων αναφοράς (όπως η συλλογή Omniglot – για την Αναγνώριση Χαρακτήρων και η συλλογή MiniImageNet – για την Κατηγοριοποίηση Εικόνων), είναι ήδη πολύ αξιόλογες, καθιστώντας τις περεταίρω βελτιώσεις πραγματική πρόκληση.

Μερικές ακόμη εφαρμογές Μάθησης Λίγων Λήψεων, από την πληθώρα διεργασιών της Μηχανικής Όρασης, με μικρότερο αριθμό δημοσιευμένων ερευνών συγκριτικά με τις διεργασίες της Οπτικής Αναγνώρισης Χαρακτήρων και της Κατηγοριοποίησης Εικόνων, αλλά εξίσου αξιοσημείωτες είναι:

- η Αναγνώριση Αντικειμένου (Object Recognition),
- η Διάδοση Τύπου Γραμματοσειράς (Font Style Transfer),
- η «Ιχνηλάτηση» Αντικειμένου (Object Tracking),
- η Εξαγωγή Περιγραφής (Phrase Grounding),
- η Ανάκτηση Παρόμοιων Εικόνων (Image Retrieval),
- η Μέτρηση Εμφανίσεων Αντικειμένου σε Εικόνα (Specific Object Counting in Images),
- η Αναγνώριση Τοποθεσίας Σκηνικού (Scene Location Recognition),
- η Αναγνώριση Χειρονομιών (Gesture Recognition),
- η Ονομασία Συγκεκριμένων Σημείων (Part Labeling),
- η Παραγωγή νέας Εικόνας (Image Generation),
- η Αντιστοίχιση Εικόνας από μία Περιοχή σε μία άλλη Περιοχή (Image Translation Across Domains),
- η Αποτύπωση σχήματος για Τρισδιάστατα Αντικείμενα (Shape View Reconstruction For 3D Objects),
- η Απόδοση Λεζάντας σε Εικόνες και οι Απαντήσεις σε Οπτικές Ερωτήσεις (Image Captioning and Visual Question Answering).

Τέλος, η Μάθηση Λίγων Λήψεων βρίσκει επιτυχή εφαρμογή και σε διεργασίες Μηχανικής Όρασης, των οποίων τα δεδομένα αφορούν βίντεο. Ενδεικτικά παραδείγματα αποτελούν:

- η Πρόβλεψη Κίνησης (Motion Prediction),
- η Κατηγοριοποίηση Video (Video Classification),
- η Αντιστοίχιση Ενέργειας σε Τόπο (Action Localization),

[23] Η Μηχανική Όραση αποτελεί γενικά ένα πολυδιάστατο πεδίο ενδιαφέροντος και εφαρμογών κι έναν από τους πιο αναπτυσσόμενους τομείς της Βαθιάς και Μηχανικής Μάθησης.

- η «Επαναταυτοποίηση» Ατόμου (Person Re-identification),
- ο Εντοπισμός Γεγονότος (Event Detection), και
- η «Κατάτμηση» Αντικειμένου (Object Segmentation).

2.3.2 Ρομποτική (Robotics)

Προκειμένου τα Robots να προσεγγίσουν αποτελεσματικότερα την ανθρώπινη συμπεριφορά, καλούνται να είναι σε θέση να γενικεύουν αποδοτικά μέσα από λίγες αναπαραστάσεις - επιδείξεις. Συνεπώς, η Μάθηση Λίγων Λήψεων έχει παίξει κομβικό ρόλο και στην εξέλιξη της Ρομποτικής.

Πρώιμα παραδείγματα εφαρμογών Few Shot Learning πάνω στον τομέα της ρομποτικής αποτελούν:

- η εκμάθηση της κίνησης ρομποτικού βραχίονα χρησιμοποιώντας Μάθηση Μίμησης (Imitating Learning) μέσα από μία μόνο επίδειξη (One-shot Demonstration) [68],
- η εκμάθηση ενεργειών χειρισμού μέσα από λίγες επιδείξεις, με την βοήθεια ενός δασκάλου - διερμηνέα, ο οποίος διορθώνει τις λανθασμένες ενέργειες [69].

Πέρα από την μίμηση των χρηστών όμως, τα Robots μπορούν να βελτιώσουν την συμπεριφορά τους και μέσω της αλληλεπίδρασης με αυτούς. Το 2016, παρατέθηκαν βοηθητικές στρατηγικές εκμάθησης, μέσα από ελάχιστες αλληλεπιδράσεις με τον χρήστη, με την χρήση Ενισχυτικής Μάθησης Λίγων Λήψεων (Few Shot Reinforcement Learning) [70].

Άλλα παραδείγματα εφαρμογών Μάθησης Λίγων Λήψεων στην Ρομποτική συμπεριλαμβάνουν:

- τα Robots Πολλαπλών Βραχιόνων (Multi-armed Bandits),
- την Οπτική Πλοήγηση (Visual Navigation) και
- τον Συνεχή Έλεγχο (Continuous Control).

Εφαρμογές, οι οποίες λαμβάνουν χώρα και επεκτείνονται όλο και περισσότερο σε πιο δυναμικά περιβάλλοντα (αλληλεπίδρασης με τον χρήστη).

2.3.3 Επεξεργασία Φυσικής Γλώσσας (Natural Language Processing)

Η Μάθηση Λίγων Λήψεων πλέον έχει επεκταθεί και στον τομέα της Επεξεργασίας της Φυσικής Γλώσσας. Παραδείγματα εφαρμογής της στην Φυσική Γλώσσα συγκροτούν:

- η Συντακτική Ανάλυση (Parsing),
- η Μετάφραση (Translation),
- η Συμπλήρωση Προτάσεων (Sentence Completion), όπου προστίθεται περιεχόμενο στα κενά προτάσεων χρησιμοποιώντας μία λέξη μέσα από την διαθέσιμη Συλλογή Δεδομένων,
- η Κατηγοριοποίηση της Ποιότητας σύντομων Κριτικών (Sentiment Classification from short reviews), όπου μία κριτική κατηγοριοποιείται ανάλογα με το περιεχόμενο της ως θετική, ουδέτερη, ή αρνητική,
- η Κατηγοριοποίηση Πρόθεσης Χρήστη μέσα από τις απαντήσεις του σε διάλογο (User Intent Classification for dialog systems),
- η Πρόβλεψη Ποινικών Ετυμηγοριών (Criminal Charge Prediction),
- οι Διεργασίες Ομοιότητας Λέξεων, όπως η δημιουργία υποκατάστατων λέξεων (Word Similarity Tasks, such as nonce definition), και
- η Κατηγοριοποίηση Κειμένου Πολλαπλών Κλάσεων (An Multi-Label Text Classification).

Το 2018 μάλιστα, δημιουργήθηκε μία συλλογή δεδομένων, η οποία ονομάζεται FewRel, ερχόμενη να αντιμετωπίσει την έλλειψη των benchmarks για διεργασίες Μάθησης Λίγων Λήψεων στον τομέα της Επεξεργασίας Φυσικής Γλώσσας [71].

2.3.4 Επεξεργασία Ήχου (Acoustic Signal Processing)

Μία πρώτη προσπάθεια εφαρμογής Μάθησης Λίγων Λήψεων στον τομέα Επεξεργασίας Ήχου αφορούσε την αναγνώριση ειπωμένων λέξεων μέσα από μόνο ένα παράδειγμα [72].

Μεταγενέστερα, η προσοχή στράφηκε επιπλέον στην σύνθεση φωνής κι ακουστικής ομιλίας (Voice Synthesis). Μία πρόσφατη εφαρμογή [73], είναι η κλωνοποίηση-μίμηση της φωνής του χρήστη (ίδια τονικότητα, χροιά και άλλα χαρακτηριστικά φωνής) μέσα από ελάχιστα ηχητικά δεδομένα. Η εφαρμογή αυτή μπορεί να φανεί χρήσιμη:

- στην παραγωγή προσωπικής φωνής, η οποία μπορεί να χρησιμοποιηθεί στην παροχή συμβουλών πλοήγησης, σε εφαρμογές χαρτών, κατά την διαδικασία της οδήγησης,
- στην εξιστόρηση παραμυθιών σε μικρά παιδιά, με την φωνή γονιών, η οποία θα παράγεται από μία έξυπνη οικιακή συσκευή.

Προσφάτως, κατέστη δυνατή η αποτύπωση μίας ομιλίας ενός χρήστη, με την φωνή ενός άλλου χρήστη χρησιμοποιώντας μόνο μία λήψη ηχητικού ή γραπτού δείγματος [74] (A. Tjandra, S. Sakti, και S. Nakamura, 2018 – “Machine speech chain with one-shot speaker adaptation”), αλλά και η μετατροπή ομιλίας από μία γλώσσα σε μία άλλη [75].

2.3.5 Άλλες Εφαρμογές

Μία πρόσφατη προσπάθεια στο πλαίσιο ιατρικοφαρμακευτικών εφαρμογών με λίγες λήψεις, είναι αυτή των H. Altae-Tran, B. Ramsundar, A. S. Pappu, και V. Pande με τίτλο “Low data drug discovery with one-shot learning” [35], στην οποία παρουσιάζεται πως μέσω της μάθησης μίας λήψης (1-Shot Learning) μπορεί να μειωθεί δραστικά η ποσότητα των δεδομένων που απαιτούνται για την ανακάλυψη φαρμάκων.

Αξιοσημείωτες εργασίες Μάθησης Λίγων Λήψεων συνιστούν και οι [76, 77, 78, 79], οι οποίες αφορούν διεργασίες «παλινδρόμησης και προσέγγισης – προσαρμογής πάνω σε συγκεκριμένη καμπύλη» (Regression-Curve Fitting), αλλά και η εργασία των T. Ramalho και M. Garnelo (2019) με τίτλο “Adaptive posterior learning: Few-shot learning with a surprise-based memory module” [80], η οποία αποσκοπεί στην κατανόηση «αριθμητικών αναλογιών» (Number Analogy) με χρήση λογικών συλλογισμών για την εκτέλεση υπολογισμών.

Επιπροσθέτως όσον αφορά την αναζήτηση βέλτιστης αρχιτεκτονικής μοντέλου (Model Architecture Search), στις εργασίες [81, 82, 83] προτείνεται και μελετάται η Αναζήτηση Αρχιτεκτονικής Μίας Λήψης (One-shot Architecture Search – OAS) για το πρόβλημα εύρεσης του βέλτιστου Βαθέως Νευρωνικού Δικτύου.

Σε αντίθεση με την τυχαία Αναζήτηση (Random Search) και την Αναζήτηση Πλέγματος (Grid Search) οι οποίες απαιτούν πολλαπλές επαναλήψεις έως ότου να βρεθεί η καλύτερη αρχιτεκτονική, η Αναζήτηση Αρχιτεκτονικής Μίας Λήψης μπορεί να καταλήξει σε μία αποδοτική αντιστοίχιση, εκπαιδεύοντας το «υπερδίκτυο» (Supernet) μόλις μία φορά.

Τις παθογένειες των μεθόδων Αναζήτησης Αρχιτεκτονικής Μίας Λήψης, έρχεται να βελτιώσει η Αναζήτηση Αρχιτεκτονικής Ελάχιστων Λήψεων (Few-shot Architecture Search) που παρατίθεται στην εργασία των Yiyang Zhao, Linnan Wang, Yuandong Tian, Rodrigo Fonseca, και Tian Guo, το 2021, με τίτλο “Few-shot Neural Architecture Search” [84].

II

Πειραματικό Μέρος

Κεφάλαιο 3ο

Παρουσίαση Πειραμάτων

3.1 Δόμηση και Προετοιμασία

Τα πειράματα της παρούσας διπλωματικής βασίζονται στην εργασία “Self-supervised Knowledge Distillation for Few-shot Learning” των Jathushan Rajasegaran et al. [1]. Το προγραμματιστικό περιβάλλον που χρησιμοποιήθηκε για την υλοποίηση αυτών είναι η ψηφιακή online πλατφόρμα του Kaggle (https://www.kaggle.com/), ενώ η εξαγωγή των γραφημάτων των συγκρίσεων έγινε με την βοήθεια του αναλυτικού εργαλείου καταγραφής πειραμάτων WandB (https://wandb.ai/site). Η υλοποίηση του κώδικα της [1] είναι βασισμένη στην βιβλιοθήκη Pytorch και υπάρχει διαθέσιμη στο https://github.com/brjathu/SKD.

Στην ενότητα αυτήν, παρατίθενται ο ακριβής τρόπος που δομήθηκαν τα πειράματα, τα χαρακτηριστικά και οι υπερπαράμετροι αυτών. Τα παρακάτω είναι κοινά για όλες τις συγκρίσεις που έλαβαν χώρα, ενώ οποιαδήποτε διαφοροποίηση ή προσθήκη επισημαίνεται στην αντίστοιχη επεξηγηματική παράγραφο.

3.1.1 Συγκρίσεις

Στην παρούσα διπλωματική, διενεργήθηκαν τρεις βασικές περιπτώσεις συγκρίσεων στις οποίες εξετάστηκαν:

1. Ο τρόπος με τον οποίο επηρεάζει το βάθος των συνελικτικών δικτύων Resnets [31] στις επιδόσεις των μοντέλων στο πρόβλημα του Few-Shot Image Classification, για τον αλγόριθμο Generation-0 της [1].
2. Ο τρόπος με τον οποίο επιδρά το πλάτος των συνελικτικών δικτύων Resnets [31] και η υπόθεση για τον αν επέρχεται βελτίωση στις επιδόσεις με την προσθήκη και την χρήση ενός SEblock [32] στην αρχιτεκτονική των δικτύων για την ίδια διεργασία.
3. Η υπόθεση για το αν αυξάνεται η επίδοση των μοντέλων στο πρόβλημα κατηγοριοποίησης εικόνας μέσα από ελάχιστα παραδείγματα, με την βοήθεια μίας διαφορετικής βοηθητικής διεργασίας αύτο-εποπτευόμενης μάθησης και συγκεκριμένα με την χρησιμοποίηση ενός Στοχαστικού Αυτοκωδικοποιητή (Variational Autoencoder) [63].

3.1.2 Κατηγορίες Few-Shot

Κάθε σύγκριση εξετάζεται στα δύο πιο δημοφιλή προβλήματα Few Shot Learning, σύμφωνα με την διεθνή βιβλιογραφία, δηλαδή το 5-ways – 1-shot και το 5-ways – 5-shot.

3.1.3 Datasets

Η εργασία [1] περιλαμβάνει σύνολο 4 datasets, δύο προέρχονται από την συλλογή ILSVRC2012 (MiniImageNet & TieredImageNet) και δύο από την συλλογή CIFAR-100 (CIFAR-FS & FC100). Τα TieredImageNet και FC100 έχουν την ιδιαιτερότητα ότι δομούν τις κλάσεις σε «σούπερ-κλάσεις»

προκειμένου να αποφευχθούν οι συγκαλύψεις που μπορεί να προκύψουν στις κλάσεις μεταξύ των Validation & Test sets με το Train set, γεγονός που τα καθιστά πιο απαιτητικά.

Οι συλλογές δεδομένων αναφοράς (datasets-benchmarks), που επιλέχθηκαν για τις συγκρίσεις είναι δύο, η CIFAR-FS και η FC100.

Οι λόγοι που επιλέχθηκαν τα συγκεκριμένα datasets είναι:

- ότι αποτελούν μία πραγματική πρόκληση για το πρόβλημα του Few-Shot Image Classification, χάρη στην μικρή ανάλυση των εικόνων (32x32),
- το γεγονός ότι οι 8 συνολικά εποχές εκπαίδευσης που υλοποιούνται στην [1] για όλες τις συλλογές πλην της CIFAR-FS (την οποία εξετάζουν για 65 εποχές) είναι σπάνιο (βασικός λόγος για τον οποίο συμπεριλήφθηκε η CIFAR-FS στα πειράματα της παρούσας διπλωματικής),
- την επιθυμία να συμπεριληφθεί και μία συλλογή αναφοράς η οποία χρησιμοποιεί τον διαχωρισμό των κλάσεων σε «σούπερ-κλάσεις» (βασικός λόγος για τον οποίο επιλέχθηκε η FC100),
- χάρη στο βολικό μέγεθος και τις «ελαφρύτερες» απαιτήσεις τους σε υπολογιστική ισχύ συγκριτικά με τις υπόλοιπες.

Και τα δύο datasets είναι βασισμένα στην συλλογή CIFAR-100, περιλαμβάνουν 60.000 εικόνες ανάλυσης 32x32 (100 κλάσεις των 600 εικόνων) και ο διαχωρισμός των Train Set, Validation Set και Test Set είναι κατ' αντιστοιχία 64, 16, 20 για την CIFAR-FS και 60, 20, 20 για την FC-100.

CIFAR-FS Dataset (CIFAR100 few-shots): Πρόκειται για dataset το οποίο εισάχθηκε στην έρευνα "Meta-learning with differentiable closed-form solvers" των Bertinetto et al. [2] λόγω της ανάγκης της ύπαρξης μίας πιο απαιτητικής-σύνθετης συλλογής αναφοράς από την Omniglot, για το πρόβλημα του Few Shot Learning, και συγχρόνως μίας πιο «ελαφριάς» από την MiniImageNet (σε επίπεδο του απαιτούμενου χρόνου σύγκλισης). Έτσι οι ερευνητές της [2], προτείναν το CIFAR-FS dataset, το οποίο είναι αποτέλεσμα τυχαίας δειγματοληψίας μέσα από την συλλογή CIFAR-100, με τα ίδια ακριβώς κριτήρια με τα οποία δημιουργήθηκε η συλλογή MiniImageNet μέσα από την ILSVRC2012. Η ομοιότητα μεταξύ των κλάσεων της συλλογής αλλά και το μέγεθος των εικόνων που αυτή περιέχει την καθιστούν μία απαιτητική πρόκληση για το πρόβλημα του Few-Shot Learning παρέχοντας συγχρόνως την δυνατότητα γρήγορης προσπέλασης.

FC-100 Dataset (Fewshot-CIFAR100): Αφορά συλλογή η οποία εισάχθηκε στην εργασία "TADAM: Task dependent adaptive metric for improved few-shot learning" των Oreshkin et al. [3]. Περιλαμβάνει τα ίδια πλεονεκτήματα με το CIFAR-FS (γρήγορη προσπέλαση, απαιτητικό benchmark), με την εξής επιπρόσθετη ιδιαιτερότητα: Προκειμένου να αποφευχθεί η πιθανότητα επικάλυψης της ίδιας κλάσης μεταξύ των Train Set και των Validation Set και Test Set (καθώς η αντιστοίχιση τους βασίζεται σε δειγματοληψία), οι 100 κλάσεις ομαδοποιούνται σε 20 «σούπερ-κλάσεις». Από αυτές οι 12 αναλογούν στο Train Set, 4 αντιστοιχούν στο Validation Set και 4 για το Test Set. Κάθε «σούπερ-κλάση» περιλαμβάνει 5 κλάσεις. Με τον τρόπο αυτόν διασφαλίζεται ότι ο έλεγχος στο πρόβλημα Few-Shot Learning γίνεται αυστηρά σε κλάσεις στις οποίες το μοντέλο βλέπει για πρώτη φορά, καθιστώντας την εν λόγω συλλογή αναφοράς ακόμη πιο απαιτητική.

3.1.4 Εκπαίδευση – Στοιχεία & Υπερπαράμετροι των πειραμάτων

Σε αυτήν την παράγραφο παρουσιάζονται τα στοιχεία εκπαίδευσης που χρησιμοποιήθηκαν στα διεξαγόμενα πειράματα, τα οποία είναι σε πλήρη ευθυγράμμιση με αυτά της υλοποίησης της [1] και όπου υπάρχουν διαφοροποιήσεις επισημαίνονται.

Ο αλγόριθμος βελτιστοποίησης που χρησιμοποιείται είναι ο Αλγόριθμος «Στοχαστικής Βαθμιαίας Καθόδου» (Stochastic Gradient Decent - SGD) με χαρακτηριστικά: Learning Rate = 0.05 , momentum = 0.9, weight decay = 0000.5.

Για το CIFAR-FS dataset, τα μοντέλα εκπαιδεύονται για 65 εποχές σύνολο, από τις οποίες οι 5 τελευταίες χαρακτηρίζονται από προσαρμογή - μείωση του Learning Rate υπό έναν παράγοντα 0.1. Στο dataset FC100, οι ερευνητές της [1] επιλέγουν εκπαίδευση για μόλις 8 εποχές, ενώ για τα πειράματα αυτής της εργασίας ο αριθμός των εποχών που επιλέχθηκαν είναι 20, από τις οποίες οι 8 πρώτες έχουν τα ίδια χαρακτηριστικά με αυτά της [1] (προκειμένου να είναι δυνατή η ευθυγράμμιση των συμπερασμάτων), και οι υπόλοιπες 12 χαρακτηρίζονται από μείωση του Learning Rate υπό έναν παράγοντα 0.1 και είναι καθαρά για επιπλέον εποπτική πληροφορία.

Το training batch size είναι 64, το συνολικό Loss είναι $L = Lce + aLss$, όπου για το Lce της βασικής διεργασίας χρησιμοποιείται το Κόστος Διασταυρωμένης Εντροπίας (nn.CrossEntropyLoss), ενώ για την απώλεια βοηθητικής διεργασίας Lss χρησιμοποιείται το Δυαδικό Κόστος Διασταυρωμένης Εντροπίας με logits (nn.functional.binary_cross_entropy_with_logits).

Για την υπερπαράμετρο – συντελεστή συνεισφοράς a, η οποία καθορίζει τον βαθμό συμμετοχής του βοηθητικού κόστους, χρησιμοποιήθηκε η βέλτιστη τιμή a = 2, η οποία προέκυψε έπειτα από μελέτη και tuning των ερευνητών της [1].

Όσον αφορά το Data Augmentation, σε ευθυγράμμιση με τις προηγούμενες εργασίες [17, 23, 20], η επεξεργασία δεδομένων που λαμβάνει χώρα κατά την εκπαίδευση είναι η εξής:

- τυχαία περικοπή (Random Crop),
- μορφοποίηση χρωμάτων (Color Jittering),
- το τυχαίως οριζόντιο αναποδογύρισμα (Random Horizontal Flip) και
- κανονικοποίηση με βάση τα κανάλια.

Αξίζει να σημειωθεί, ότι αναμένουμε τα δικά μας πειραματικά αποτελέσματα των επιδόσεων, να είναι χαμηλότερα από τα αναγραφόμενα της [1], καθώς στα πειράματα που υλοποιήθηκαν, το Validation Set αξιοποιείται καθαρά για εποπτική πληροφορία και ως συγκριτική ευθυγράμμιση με το Test Set, ενώ στα δημοσιοποιημένα αποτελέσματα της [1] έχει αξιοποιηθεί για περαιτέρω εκπαίδευση.

3.1.5 Αξιολόγηση στο Few-Shot Task – Στοιχεία & Υπερπαράμετροι Ελέγχου

Στην παράγραφο αυτήν επεξηγείται ο τρόπος με τον οποίο λαμβάνει χώρα η αξιολόγηση στο Few-Shot Task, τα στοιχεία και οι υπερπαράμετροι που τον χαρακτηρίζουν.

Προκειμένου να γίνει ο έλεγχος, επιλέγονται μέσα από το Test Set, τα Support Set και Query Set {D_{supp}, D_{query}}. Το D_{supp} αποτελείται από ζευγάρια εικόνων-ετικετών {x_{supp}, y_{supp}}, ενώ το D_{query} μόνο από τις εικόνες x_{query}. Τα x_{supp} και x_{query} δίνονται ως είσοδοι στο εκπαιδευμένο μοντέλο προκειμένου να παραχθούν τα αντίστοιχα feature embeddings v_{supp} και v_{query}. Στην συνέχεια, έπειτα από κανονικοποίηση των feature embeddings σε μία μοναδιαία σφαίρα [23] χρησιμοποιείται ένας απλός «ταξινομητής λογιστικής παλινδρόμησης» (logistic regression classifier) [2, 23] προκειμένου να γίνει η κατηγοριοποίηση των εικόνων x_{query} σύμφωνα με τα παραχθέντα από το μοντέλο feature embeddings v_{supp} και v_{query}.

Ο έλεγχος λαμβάνει χώρα για 600 επεισόδια-εκτελέσεις αξιολόγησης και η αντιπροσωπευτική τιμή ελέγχου είναι ο μέσος όρος αυτών με διάστημα εμπιστοσύνης 95%. Ο αριθμός των εικόνων queries της x_{query} είναι 15 εικόνες ανά επεισόδιο. Επιπλέον, για κάθε εικόνα περιλαμβάνονται 5 διαφορετικές μετασχηματισμένες εκδοχές αυτής (n_aug_support_samples=5) και για το D_{query} και για το D_{supp}. Οι μετασχηματισμοί που λαμβάνουν χώρα προκειμένου να σχηματιστούν αυτές οι διαφορετικές εκδοχές είναι οι ίδιοι ακριβώς με αυτούς που γίνονται στο στάδιο εκπαίδευσης για το D_{supp}, ενώ για το D_{query} δεν χρησιμοποιείται μόνο η μορφοποίηση χρωμάτων.

3.2 Συγκρίσεις

Στην ενότητα αυτήν, παρουσιάζονται και αναλύονται τα αποτελέσματα των τριών βασικών περιπτώσεων συγκρίσεων, υπό τις παραπάνω συνθήκες.

3.2.1 Σύγκριση 1η

Στην παρούσα σύγκριση, εξετάζεται ποια είναι η επίδραση του βάθους των συνελικτικών δικτύων Resnets [31] στο πρόβλημα του Few-Shot Image Classification του αλγορίθμου Generation-0 της [1]. Βασικοί λόγοι που συντέλεσαν στην επιλογή των Resnets είναι:

- η συχνότητα εμφάνισης τους στις διάφορες εργασίες,
- οι αξιοσημείωτες επιδόσεις τους,
- καθώς και το γεγονός ότι ένα από αυτά (και συγκεκριμένα το Resnet-12), αποτελεί μοντέλο "backbone" για την καταγραφή των επιδόσεων πρωτίστως στην εργασία [1], αλλά και σε άλλες, όπως τις [3, 21, 23, 30].

3.2.1.1 Περιγραφή Μοντέλων Σύγκρισης

Γενικά

Τα μοντέλα Resnets [31] είναι δίκτυα τα οποία περιλαμβάνουν την σειριακή τοποθέτηση πολλαπλών επιπέδων λειτουργίας συνοδευόμενη από την σύνδεση αυτών, με τις λεγόμενες «συνδέσεις παράλειψης» (shortcut connections). Πρόκειται για προσέγγιση η οποία εισάχθηκε από τους Kaiming He, Xiangyu Zhang, Shaoqing Ren και Jian Sun και αποσκοπεί στην επίλυση διάφορων ζητημάτων. Οι He et al. στην μελέτη με τίτλο "Deep Residual Learning for Image Recognition" [31] παραθέτουν ισχυρά πειραματικά αποτελέσματα τα οποία δείχνουν πως συγκριτικά με προηγούμενες προσεγγίσεις, η δόμηση της αρχιτεκτονικής των δικτύων με αυτόν τον τρόπο:

- Διευκολύνει την διαδικασία της εκπαίδευσης των βαθιών νευρωνικών δικτύων, προσδίδοντας ευκολότερη βελτιστοποίηση και σύγκλιση,
- Παρέχει πολύ μικρότερη υπολογιστική πολυπλοκότητα συγκριτικά ακόμη και με πιο ρηχά "plain" δίκτυα (χωρίς shortcut connections),
- Επιτυγχάνει βαθιές αρχιτεκτονικές και επιφέρει σημαντική αύξηση στις επιδόσεις των μοντέλων με την αύξηση του βάθους, σε αντίθεση με τα "plain" δίκτυα και το πρόβλημα του κορεσμού που εμφανιζόταν σε προηγούμενες μελέτες.

Στην έρευνα τους [31], οι He et al., παραθέτουν μελέτη από 100 μέχρι 1000 στρώματα βάθους για αυτά τα δίκτυα και επιτυγχάνουν σημαντικές βελτιώσεις και διακρίσεις στους διαγωνισμούς ILSVRC και COCO 2015, όπου καταφέρνουν να κατακτήσουν την 1η θέση στις διεργασίες εντοπισμού και απόδοση ακριβούς θέσης στην συλλογή αναφοράς ImageNet (detection & localization), και στις διεργασίες εντοπισμού και διαχωρισμού στην συλλογή αναφοράς COCO (detection & segmentation).

Ειδικά

Πιο συγκεκριμένα γίνεται σύγκριση των επιδόσεων μεταξύ των αρχιτεκτονικών Resnet-12, μίας «ρηχής» έκδοχής των Resnets, του Resnet-50, μίας πιο «βαθιάς» αρχιτεκτονικής, και του Resnet-101, μίας ακόμη πιο «βαθιάς» εκδοχής αυτών των δικτύων.

Η υλοποίηση των δικτύων Resnets σύμφωνα με τα μοντέλα που περιλαμβάνονται στον κώδικα της [1] (https://github.com/brjathu/SKD), περιέχει τα εξής δομικά στοιχεία σε σειρά:

- 4 βασικές ομάδες Blocks επιπέδων, δηλαδή,
  - την ομάδα των Block-Α,
  - την ομάδα των Block-Β,
  - την ομάδα των Block-Γ,
  - και την ομάδα των Block-Δ
- 1 επίπεδο Average pooling (AdaptiveAvgPool2d)
- και το τελικό επίπεδο ευθυγράμμισης και αντιστοιχίας στις κλάσεις κατηγοριοποίησης, που είναι ένας γραμμικός μετασχηματιστής με 640 κόμβους που εξέρχονται από το τελευταίο Block-Δ (nn.Linear(640, self.num_classes)).
- για τον αλγόριθμο Generation-0, συμπεριλαμβάνεται σε σειρά ένας ακόμη γραμμικός μετασχηματιστής (nn.Linear(self.num_classes,4)) ο οποίος χρησιμοποιείται στην βοηθητική διεργασία του rotation.

Όσον αφορά την επεξήγηση των Blocks, πρόκειται για το ίδιο Block, το οποίο καλείται επαναλαμβανόμενα με 4 διαφορετικούς τρόπους. Παρακάτω αναφέρονται τα επίπεδα που εμπεριέχονται σε ένα Block-Α και στην συνέχεια αναφέρονται οι διαφοροποιήσεις των υπολοίπων τύπων Block από αυτό.

Ένα Block-Α περιλαμβάνει:

- 1ο Στρώμα επιπέδων το οποίο περιέχει:
  - 1 συνελικτικό επίπεδο 64 φίλτρων με πυρήνα διαστάσεων 3x3, με stride=1, padding=1 και χωρίς bias.
  - 1 επίπεδο BatchNormalization.
  - 1 συνάρτηση ενεργοποίησης LeakyRelu με negative_slope=0.1.
- 2ο Στρώμα επιπέδων το οποίο περιέχει:
  - 1 συνελικτικό επίπεδο 64 φίλτρων με πυρήνα διαστάσεων 3x3, με stride=1, padding=1 και χωρίς bias.
  - 1 επίπεδο BatchNormalization.
  - 1 συνάρτηση ενεργοποίησης LeakyRelu με negative_slope=0.1.
- 3ο Στρώμα επιπέδων το οποίο περιέχει:
  - 1 συνελικτικό επίπεδο 64 φίλτρων με πυρήνα διαστάσεων 3x3, με stride=1, padding=1 και χωρίς bias.
  - 1 επίπεδο BatchNormalization.
  - Την residual σύνδεση η οποία είναι το άθροισμα της εξόδου είτε με την αρχική είσοδο του Block-Α αυτούσια, είτε με την Downsampled επεξεργασία αυτής. Η Downsample επεξεργασία, λαμβάνει χώρα όποτε αλλάζει ο τύπος της ομάδας των Blocks, προκειμένου να υπάρχει ομοιογένεια στις διαστάσεις των προστιθέμενων tensors (και συγκεκριμένα στον αριθμό των φίλτρων).
    Το Downsample στρώμα απαρτίζεται από:
    - 1 συνελικτικό επίπεδο 64 φίλτρων με πυρήνα διαστάσεων 1x1, stride=1 και χωρίς bias.
    - 1 επίπεδο BatchNormalization.
  - 1 συνάρτηση ενεργοποίησης LeakyRelu με negative_slope=0.1.
  - 1 επίπεδο Maxpool με μέγεθος πυρήνα 2x2 και stride=2.
  - 1 επίπεδο Dropout ή Dropblock τα οποία ενεργοποιούνται υπό προϋποθέσεις με κοινό Droprate=0.1. (Το Dropblock ενεργοποιείται μόνο για τις ομάδες των Block-Γ και Block-Δ κι

εφόσον πρόκειται για το τελευταίο Block της ομάδας, σε κάθε άλλη περίπτωση ενεργοποιείται το Dropout).

Τα Block-Β, Block-Γ και Block-Δ περιλαμβάνουν ακριβώς τα ίδια στρώματα και επίπεδα, διαφοροποιούνται όμως στον αριθμό των χρησιμοποιούμενων φίλτρων στα συνελικτικά επίπεδα, τα οποία είναι 160, 320 και 640 αντίστοιχα. Επιπλέον, υπάρχουν κάποιες διαφοροποιήσεις στον τρόπο με τον οποίο γίνεται η εφαρμογή του Dropblock (τα Γ-Δ ενεργοποιούν το Dropblock επίπεδο, ενώ τα Α-Β ενεργοποιούν το Dropout).

Συνεπώς, για τα μοντέλα που χρησιμοποιήθηκαν σε αυτήν την σύγκριση:

Μοντέλο Resnet12

Τα δομικά επίπεδα του Resnet12 είναι αυτά που αναφέρθηκαν παραπάνω και η κάθε ομάδα έχει πλήθος 1 Block.

Μοντέλο Resnet50

Για τα δομικά επίπεδα του Resnet50 ισχύει ακριβώς το ίδιο, με την διαφορά ότι οι ομάδες των Α, Β, Γ, Δ απαρτίζονται από 3, 4, 6 και 3 Blocks αντίστοιχα.

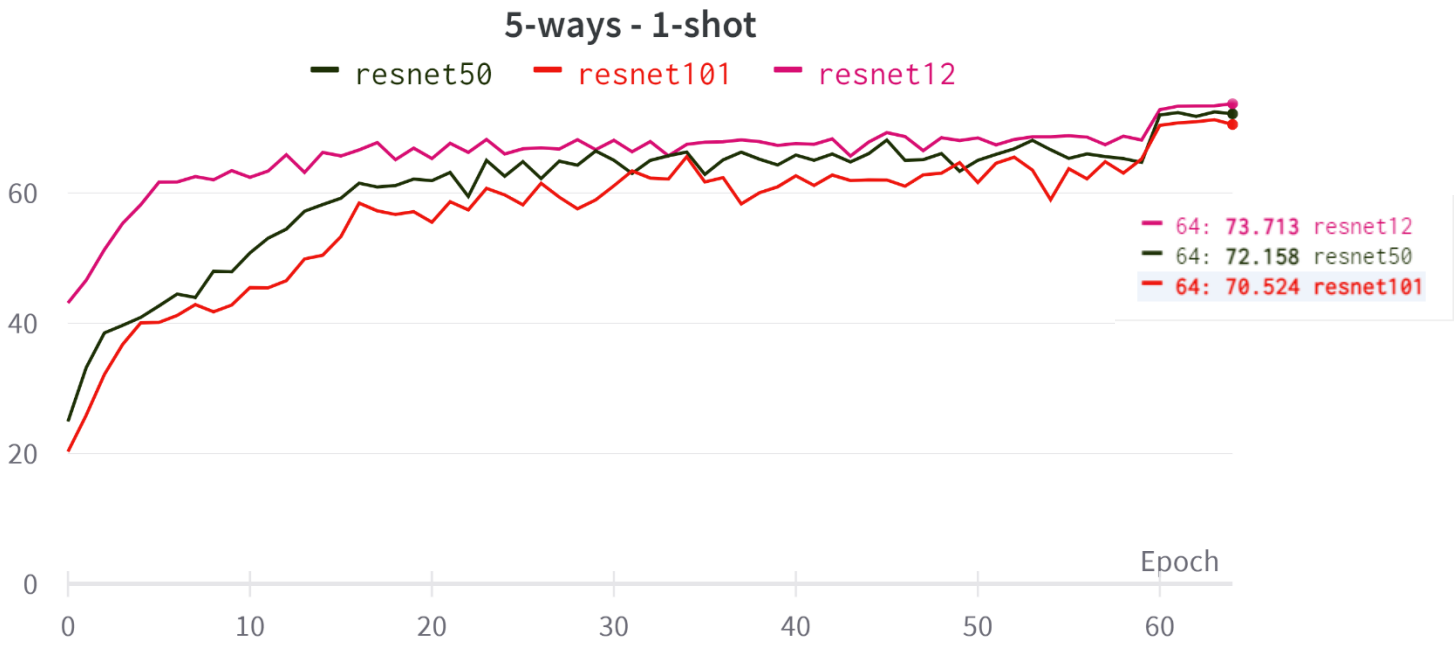
Μοντέλο Resnet101

Σε σύμπλευση με τα παραπάνω η διαφοροποίηση του μοντέλου Resnet101 είναι ότι αποτελείται από 3, 4, 23 και 3 Blocks αντίστοιχα για την κάθε ομάδα.

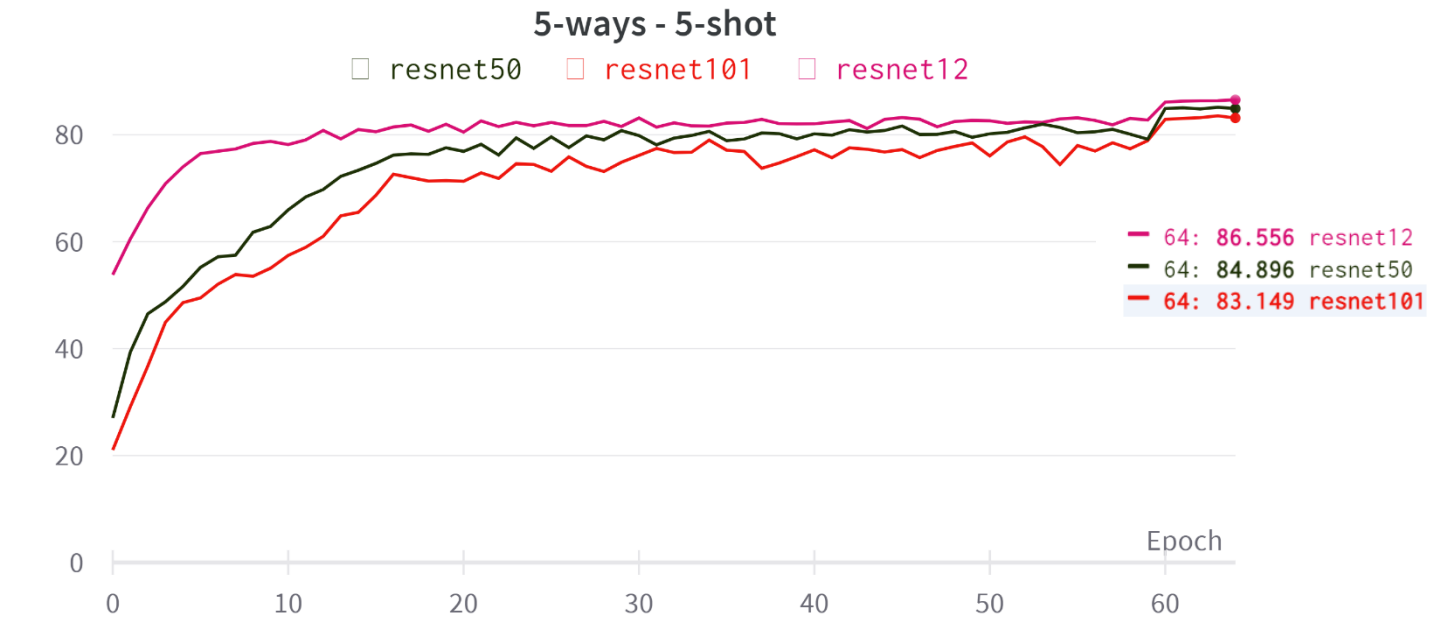
Τέλος, αξίζει να σημειωθεί ότι για κάθε μοντέλο, έλαβε χώρα αρχικοποίηση των παραμέτρων του δικτύου και συγκεκριμένα, για τα βάρη των συνελικτικών επιπέδων χρησιμοποιήθηκε η αρχικοποίηση Kaiming Normal, ενώ για τις παραμέτρους εκμάθησης γ, β των Batchnormalization επιπέδων, οι τιμές 0 και 1 αντίστοιχα.

3.2.1.2 Διαγράμματα Αποτελεσμάτων

CIFAR-FS dataset

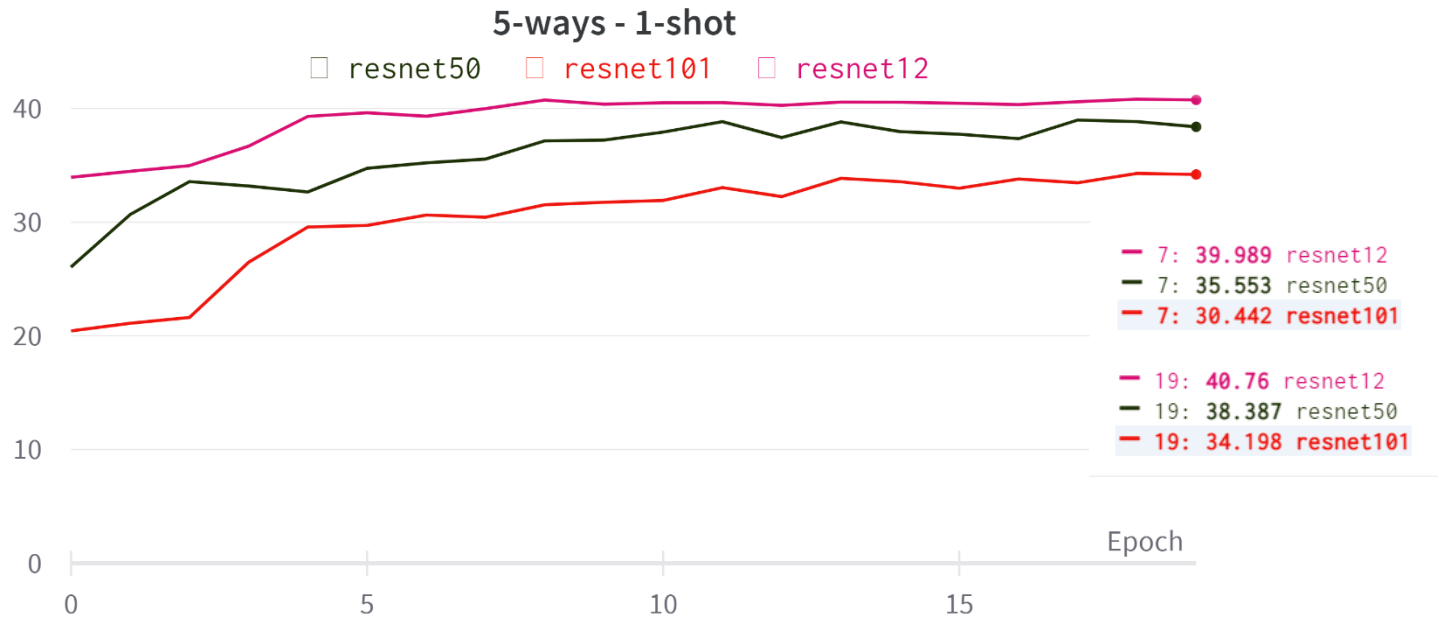


Σχήμα 3.2.1.2 - 1: Διάγραμμα σύγκρισης επίδρασης βάθους δικτύων 5-ways – 1-shot για το CIFAR-FS dataset.

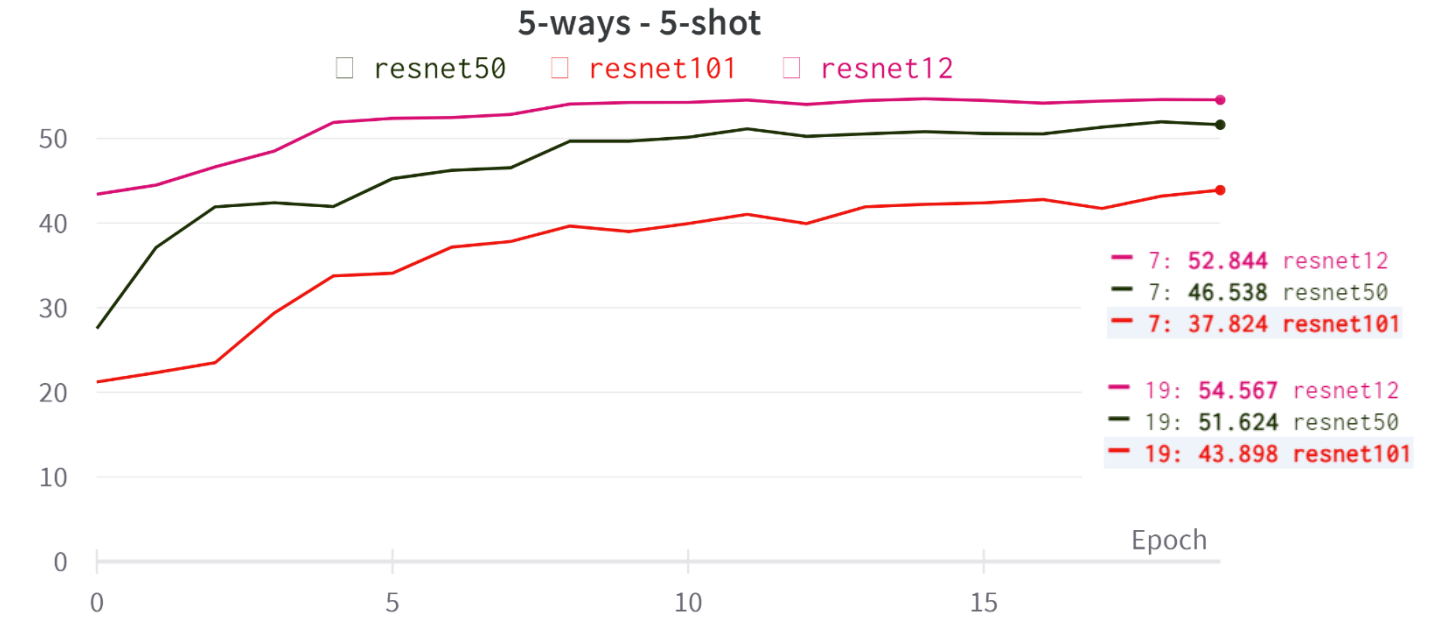


Σχήμα 3.2.1.2 - 2: Διάγραμμα σύγκρισης επίδρασης βάθους δικτύων 5-ways – 5-shot για το CIFAR-FS dataset.

FC100 dataset



Σχήμα 3.2.1.2 - 3: Διάγραμμα σύγκρισης επίδρασης βάθους δικτύων 5-ways – 1-shot για το FC100 dataset.



Σχήμα 3.2.1.2 - 4: Διάγραμμα σύγκρισης επίδρασης βάθους δικτύων 5-ways – 5-shot για το FC100 dataset.

3.2.1.3 Παρατηρήσεις-Συμπεράσματα

Στα διαγράμματα των Σχημάτων 3.2.1.2 - 1, 3.2.1.2 - 2, 3.2.1.2 - 3 και 3.2.1.2 - 4 απεικονίζεται αναλυτικά ο έλεγχος στο Few-Shot Image Classification Task. Στον οριζόντιο άξονα περιλαμβάνεται ο αριθμός της εκάστοτε εποχής εκπαίδευσης (με την αρίθμηση να ξεκινάει από το 0), ενώ στον κάθετο άξονα περιλαμβάνεται το ποσοστό απόδοσης που αντιστοιχίζεται στην κάθε εποχή. Η ακριβής τιμή του ποσοστού απόδοσης υπολογίζεται από την κανονικοποιημένη τιμή στην ποσοστιαία κλίμακα του 100 (%), του μέσου όρου των τιμών που προκύπτουν από τα 600 επεισόδια αξιολόγησης τα οποία λαμβάνουν χώρα έπειτα από την κάθε εποχή εκπαίδευσης. Δεξιά από τα διαγράμματα αναγράφονται οι ετικέτες με τις τιμές επίδοσης του κάθε μοντέλου (με σειρά καλύτερης επίδοσης από πάνω προς τα κάτω). Στο CIFAR-FS dataset περιλαμβάνονται μόνο οι ετικέτες με τις τελικές επιδόσεις των μοντέλων (65^{η} εποχή) ενώ στο FC100 περιλαμβάνονται και οι τελικές αλλά και της 8^{ης} εποχής, όπως αναφέρθηκε, σε συμφωνία με την εργασία [1].

Όσον αφορά το “backbone” μοντέλο της [1], το Resnet12, τα πειραματικά αποτελέσματα για το CIFAR-FS dataset είναι πολύ κοντά με τα αναγραφόμενα της [1] και ελαφρώς πιο χαμηλά όπως ήταν αναμενόμενο (συγκεκριμένα, για το 1-shot: 73.713$\pm$0.94 έναντι 74.5$\pm$0.9 [1] και για το 5-shot: 86.556$\pm$0.61 έναντι 88.0$\pm$0.6 [1]). Από την άλλη, δεν φαίνεται να συμβαίνει το ίδιο στο FC100 dataset όπου τα αποτελέσματα είναι αρκετά χαμηλότερα από τα αναγραφόμενα της [1] (συγκεκριμένα, για το 1-shot: 39.989$\pm$0.71 έναντι 45.3$\pm$0.8 [1] και για το 5-shot: 52.844$\pm$0.69 έναντι 62.2$\pm$0.7 [1]).

Ο λόγος των χαμηλότερων προσδοκιών σε επιδόσεις, είναι, όπως αναφέρθηκε, η μη-αξιοποίηση των δεδομένων του Validation Set κατά την διαδικασία της εκπαίδευσης. Με αφορμή την μεγάλη απόκλιση των αποτελεσμάτων από αυτά της [1] για το FC100 dataset, διενεργήθηκε το ίδιο ακριβώς πείραμα για το μοντέλο Resnet12 στο FC100 dataset, με αξιοποίηση του Validation Set στην εκπαιδευτική διαδικασία, προκειμένου να εξακριβωθεί εάν αυτό είναι το αίτιο που δημιουργεί την τόσο μεγάλη απόκλιση. Τα αποτελέσματα επιβεβαίωσαν αυτόν τον ισχυρισμό, καθώς οι τελικές επιδόσεις ήταν: στο 1-shot: 46.956$\pm$0.84 έναντι 45.3$\pm$0.8 [1] και στο 5-shot: 60.7$\pm$0.73 έναντι 62.2$\pm$0.7 ([1]).[24]

Συνεπώς, το FC100 dataset φαίνεται να είναι αρκετά ευαίσθητο στις μεταβολές (άρα και στον θόρυβο), γεγονός που δεν θα το καθιστούσε καλή επιλογή για την εξαγωγή συμπερασμάτων εφόσον κάποιος βασιστεί μόνο σε αυτό. Επομένως, ο ρόλος του για αυτήν αλλά και τις επόμενες συγκρίσεις είναι κυρίως επικουρικός.

Παρά την παραπάνω απόκλιση στα αποτελέσματα του FC100 dataset, φαίνεται πως η αύξηση του βάθους του δικτύου Resnet για τον αλγόριθμο Generation-0 της [1], έχει αρνητική επίδραση στις επιδόσεις του Few-Shot Image Classification. Σύμφωνα με τα διαγράμματα των Σχημάτων 3.2.1.2 - 1, 3.2.1.2 - 2, 3.2.1.2 - 3 και 3.2.1.2 - 4 φαίνεται πως το Resnet12 υπερτερεί σε επίδοση των Resnet50 και Resnet101, και το Resnet50 υπερτερεί του Resnet101. Μάλιστα, αυτό συμβαίνει και για το 5-ways – 1-shot task αλλά και για το 5-ways – 5-shot task και επαληθεύεται και στα δύο χρησιμοποιούμενα datasets. Επίσης, μπορούμε να παρατηρήσουμε πως ειδικά στο FC100 dataset, η διαφορά επίδοσης του Resnet12 από το Resnet101 είναι μεγάλη, συγκεκριμένα, για το 5-ways – 1-shot: 39.989$\pm$0.71 (Resnet12) έναντι 30.442$\pm$0.74 (Resnet101) ($\sim$10% διαφορά, συγκριτικά με το αντίστοιχο $\sim$3.5% που συναντάται στο CIFAR-FS dataset), και για το 5-ways – 5-shot: 52.844$\pm$0.69 (Resnet12) έναντι 37.824$\pm$0.7 (Resnet101) ($\sim$15% διαφορά, συγκριτικά και πάλι με το αντίστοιχο $\sim$3.5% του CIFAR-FS) γεγονός που επιβεβαιώνει και πάλι την ευαισθησία του FC100 dataset.

[24] Το ίδιο πείραμα τελέστηκε και για το CIFAR-FS dataset, και τα αποτελέσματα ήταν ακόμη πιο κοντά με τα αναγραφόμενα της [1] (συγκεκριμένα: 1-shot: 75.144$\pm$0.91 έναντι 74.5$\pm$0.9 [1] και 5-shot: 87.062$\pm$0.62 έναντι 88.0$\pm$0.6 [1]).

Ο Πίνακας 3.2.1.2 - 1 συνοψίζει τα τελικά αποτελέσματα αυτής της σύγκρισης και για τις δύο χρησιμοποιούμενες συλλογές δεδομένων αναφοράς.

	CIFAR-FS		FC-100 (8η Εποχή)	
Μοντέλο	5-ways - 1-shot	5-ways - 5-shot	5-ways - 1-shot	5-ways - 5-shot
Resnet12	**73.713±0.94**	**86.556±0.61**	**39.989±0.71**	**52.844±0.69**
Resnet50	72.158±0.96	84.896±0.63	35.553±0.73	46.538±0.73
Resnet101	70.524±0.97	83.149±0.67	30.442±0.74	37.824±0.7

Πίνακας 3.2.1.2 - 1:

Συγκεντρωτικός πίνακας των τελικών αποτελεσμάτων των μοντέλων, όσον αφορά την διερεύνηση της επίδρασης του βάθους των δικτύων.

3.2.2 Σύγκριση 2η

Η εν λόγω σύγκριση αποσκοπεί στην διερεύνηση της επίδρασης του πλάτους (αριθμός των συνελικτικών καναλιών) του δικτύου Resnet12 στο Few-Shot Image Classification Task για τον αλγόριθμο Generation-0 της [1]. Επιπλέον, εξετάζεται και η υπόθεση για το αν θα επέλθει βελτίωση στον ίδιο αλγόριθμο, εξαιτίας της συμπερίληψης ενός SE block [32] στα συνελικτικά Blocks που περιλαμβάνει το δίκτυο.

Τα μοντέλα που χρησιμοποιήθηκαν σε αυτήν την σύγκριση είναι:

- Τα Resnet12, Resnet12-HalfFeatures και Resnet12-QuarterFeatures όσον αφορά την εξαγωγή συμπερασμάτων με βάση το πλάτος.
- Το Seresnet12 όσον αφορά την διερεύνηση της υπόθεσης της συμπερίληψης του SE Block στο δίκτυο Resnet12.

Οι λόγοι που επιλέχθηκε το Resnet12 ως αφετηρία για τις παραπάνω αξιολογήσεις είναι:

- η συχνότητα εμφάνισης του στις έρευνες,
- το γεγονός ότι αποτελεί μοντέλο “backbone” της [1],
- το γεγονός ότι κατέγραψε τις καλύτερες επιδόσεις για τον συγκεκριμένο αλγόριθμο σύμφωνα με την 1η σύγκριση,
- το βολικό μέγεθος κι οι χαμηλότερες απαιτήσεις του δικτύου σε υπολογιστική ισχύ.

3.2.2.1 Περιγραφή Μοντέλων Σύγκρισης

Γενικά

Καθώς τα μοντέλα Resnets [31] έχουν ήδη αναφερθεί, στην παράγραφο αυτή ακολουθεί μόνο μία σύντομη αναφορά στα μοντέλα SENets [32].

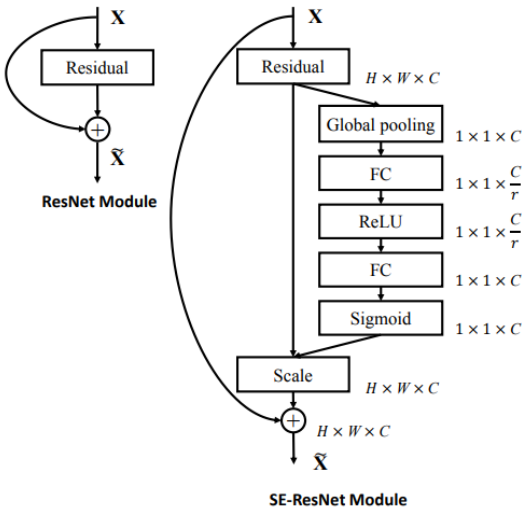
Μοντέλα SENets

Πρόκειται για μία κατηγορία που εμπεριέχει την προσθήκη μίας επιπλέον μονάδας που ονομάζεται “Squeeze-and-Excitation Block” (SE block). Τα SENets και τα SE blocks προτάθηκαν στην εργασία “Squeeze-and-Excitation Networks” των Jie Hu et al. [32]. Η κεντρική ιδέα είναι η προσθήκη ενός υπολογιστικού μηχανισμού προσοχής, ο οποίος επιδιώκει την βαθμονόμηση της σημαντικότητας των Features που εξάγονται στα ενδιάμεσα στάδια ενός συνελικτικού νευρωνικού δικτύου. Παρά την απλότητα και την ευκολία της προσθήκης ενός τέτοιου Block, φάνηκε ότι ρηχά μοντέλα μπορούν να πετύχουν παρόμοιες επιδόσεις με αυτές που επιτυγχάνουν πολύ πιο βαθιά δίκτυα, εξοικονομώντας υπολογιστική ισχύ και απαλείφοντας από πολλά επίπεδα νευρώνων, προσθέτοντας μόλις λίγες παραμέτρους.

Ένα SE block συνήθως περιλαμβάνει:

- 1 Global Pooling επίπεδο,
- 1 Fully Connected επίπεδο (όπου λαμβάνει χώρα και μία σύμπτυξη του αριθμού των καναλιών),
- 1 συνάρτηση ενεργοποίησης ReLu για μη-γραμμικότητα,
- 1 Fully Connected επίπεδο (όπου επαναφέρεται ο αριθμός των καναλιών),
- 1 Sigmoid συνάρτηση ενεργοποίησης.

Το αποτέλεσμα του SE Block περιλαμβάνει τιμές από το 0 μέχρι το 1 για κάθε κανάλι (οι οποίες επιτελούν τον ρόλο συντελεστή βαρύτητας), και στην συνέχεια πολλαπλασιάζεται με το αποτέλεσμα του κάθε συνελικτικού επιπέδου προσδίδοντας έτσι την τελική αποτύπωση η οποία συνυπολογίζει και την σημαντικότητα συμβολής του κάθε καναλιού-feature.



Σχήμα 3.2.2.1 - 1: Απεικόνιση της μετατροπής ενός απλού residual block σε SE-residual block.

Πηγή: Squeeze-and-Excitation Networks - Jie Hu, Li Shen, Samuel Albanie, Gang Sun, Enhua Wu. [32]

Ειδικά

Τα μοντέλα που χρησιμοποιήθηκαν σε αυτήν την σύγκριση είναι τα Resnet12, Resnet12-HalfFeatures, Resnet12-QuarterFeatures και Seresnet12. Στην συνέχεια αναφέρονται τα ακριβή χαρακτηριστικά τους.

Μοντέλο Resnet12

Πρόκειται για το ίδιο μοντέλο που χρησιμοποιήθηκε και περιγράφηκε στην 1η Σύγκριση. Εδώ, καθώς πρόκειται να εξεταστεί η επίδραση του πλάτους αξίζει να σημειωθεί, πως ο αριθμός των φίλτρων που χρησιμοποιούν οι ερευνητές της [1] στα residual blocks, δεν είναι ακριβώς ο ίδιος με τις συνήθεις υλοποιήσεις. Πιο συγκεκριμένα για το Resnet12 όπως αναφέρθηκε στην 1η σύγκριση ο αριθμός αυτός για τα 4 χρησιμοποιούμενα Residual Blocks είναι 64, 160, 320, 640 αντίστοιχα, ενώ η επικρατούσα υλοποίηση αποτελείται από 64, 128, 256, 512 κατ' αντιστοιχία.[25]

Μοντέλο Resnet12-HalfFeatures

Πρόκειται για το Resnet12 που χρησιμοποιήθηκε στην 1η σύγκριση. Συμπεριλαμβάνει ακριβώς τα ίδια χαρακτηριστικά τα οποία αναλύθηκαν στην προηγούμενη ενότητα, με την διαφοροποίηση ότι ο αριθμός

[25] Τελέστηκε και ένα επιπλέον πείραμα στο CIFAR-FS dataset, με τον συνήθη αριθμό features του Resnet12 (64, 128, 256, 512), προκειμένου να φανούν οι διαφορές από την υλοποίηση των ερευνητών της [1]. Τα αποτελέσματα έδειξαν να μην διαφέρουν και πολύ για τις δύο υλοποιήσεις, με μοναδική αξιοσημείωτη παρατήρηση ότι η υλοποίηση της [1] επιδεικνύει ελαφρώς καλύτερα αποτελέσματα στο 1-shot task (διαφορά της τάξης ~0.6%). Συγκεκριμένα είχαμε, για το 1-shot: 73.093±0.93 έναντι 73.713±0.93 (υλοποίηση Resnet 12 της [1]) και για το 5-shot: 86.358±0.61 έναντι 86.556±0.61 (υλοποίηση Resnet12 της [1]).

των συνελικτικών φίλτρων που χρησιμοποιήθηκε στα Residual Blocks υποδιπλασιάστηκε (συγκεκριμένα από 64, 160, 320, 640 για τα Block τύπου Α, Β, Γ, Δ αντίστοιχα, έγινε 32, 80, 160, 320).

Μοντέλο Resnet12-QuarterFeatures

Πρόκειται πάλι για το Resnet12, με τα χαρακτηριστικά που αναλύθηκαν παραπάνω. Η διαφοροποίηση εδώ, είναι ότι ο αριθμός των συνελικτικών φίλτρων που χρησιμοποιήθηκε στα Residual Blocks υποτετραπλασιάστηκε (συγκεκριμένα από 64, 160, 320, 640 για τα Block τύπου Α, Β, Γ, Δ αντίστοιχα, έγινε 16, 40, 80, 160).

Μοντέλο Seresnet12

Η δομή του μοντέλου Seresnet12 είναι ακριβώς η ίδια με αυτή του Resnet12, με την επιπλέον προσθήκη ενός SE Block.

Το ακριβές SE Block που προστίθεται αποτελείται από τα εξής επίπεδα σε σειρά:

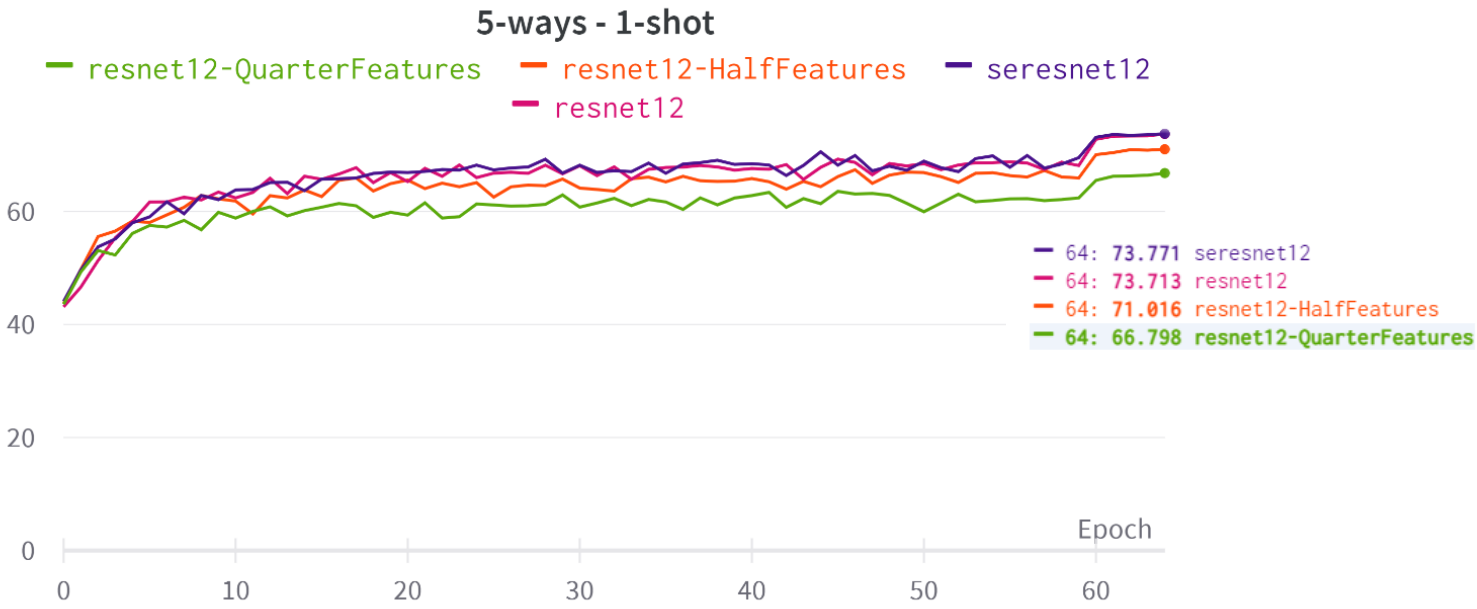
- 1 επίπεδο Average Pooling (nn.AdaptiveAvgPool2d),
- 1 επίπεδο Fully Connected το οποίο συμπεριλαμβάνει την σύμπτυξη των καναλιών (nn.Linear(channel, channel // reduction)),
- 1 συνάρτηση ενεργοποίησης ReLu με αντικατάσταση των τιμών (Inplace=True),
- 1 επίπεδο Fully Connected το οποίο ανακτά το μέγεθος των καναλιών (nn.Linear(channel // reduction, channel)),
- 1 συνάρτηση ενεργοποίησης Sigmoid.

Το προϊόν που παράγεται από το SE Block πολλαπλασιάζεται με την είσοδο του, και το τελικό αποτέλεσμα εμπεριέχει την τελική αποτύπωση του κάθε Feature-καναλιού σύμφωνα με τον συντελεστή σημαντικότητας της συμβολής αυτού που παράχθηκε από το SE block. Ο βαθμός σύμπτυξης (reduction) που χρησιμοποιήθηκε είναι ίσος με 16.

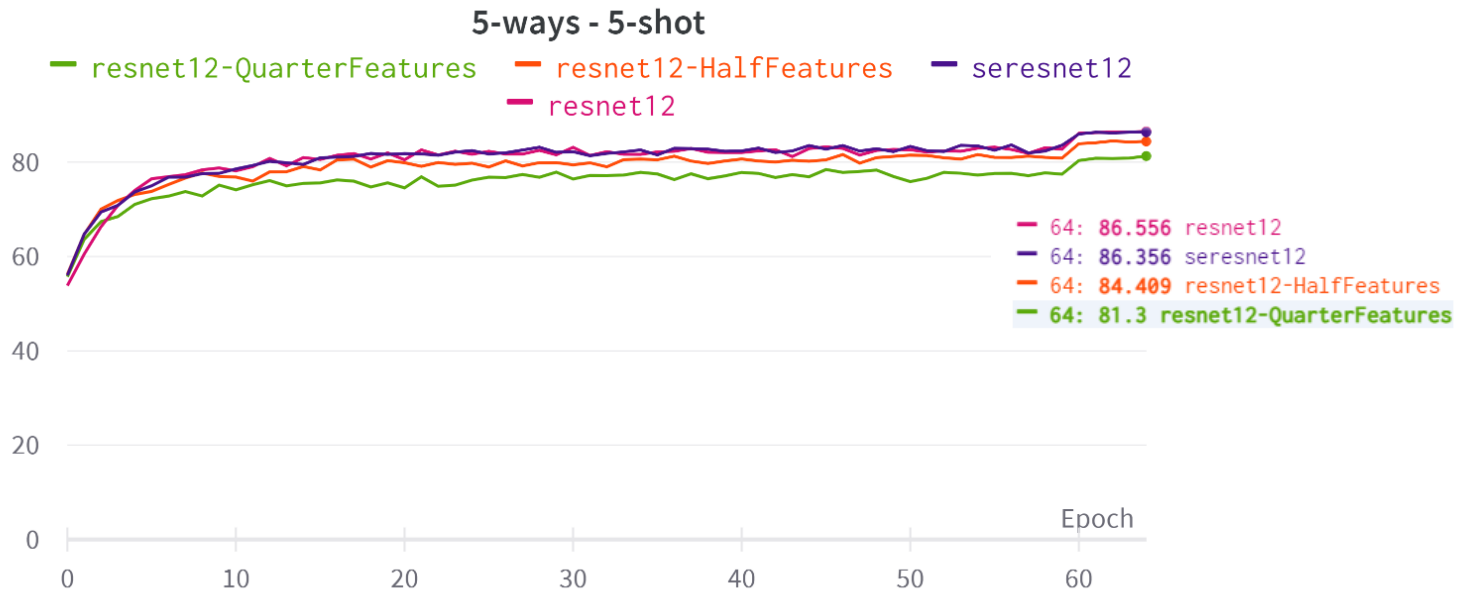
Το SE Block εφαρμόζεται σε κάθε 3^{ο} στρώμα επιπέδων από κάθε Block (όλων των τύπων ομάδων) που περιλαμβάνει το δίκτυο αμέσως μετά το επίπεδο του BatchNormalization και ακριβώς πριν το επίπεδο της residual σύνδεσης.

3.2.2.2 Διαγράμματα Αποτελεσμάτων

CIFAR-FS dataset

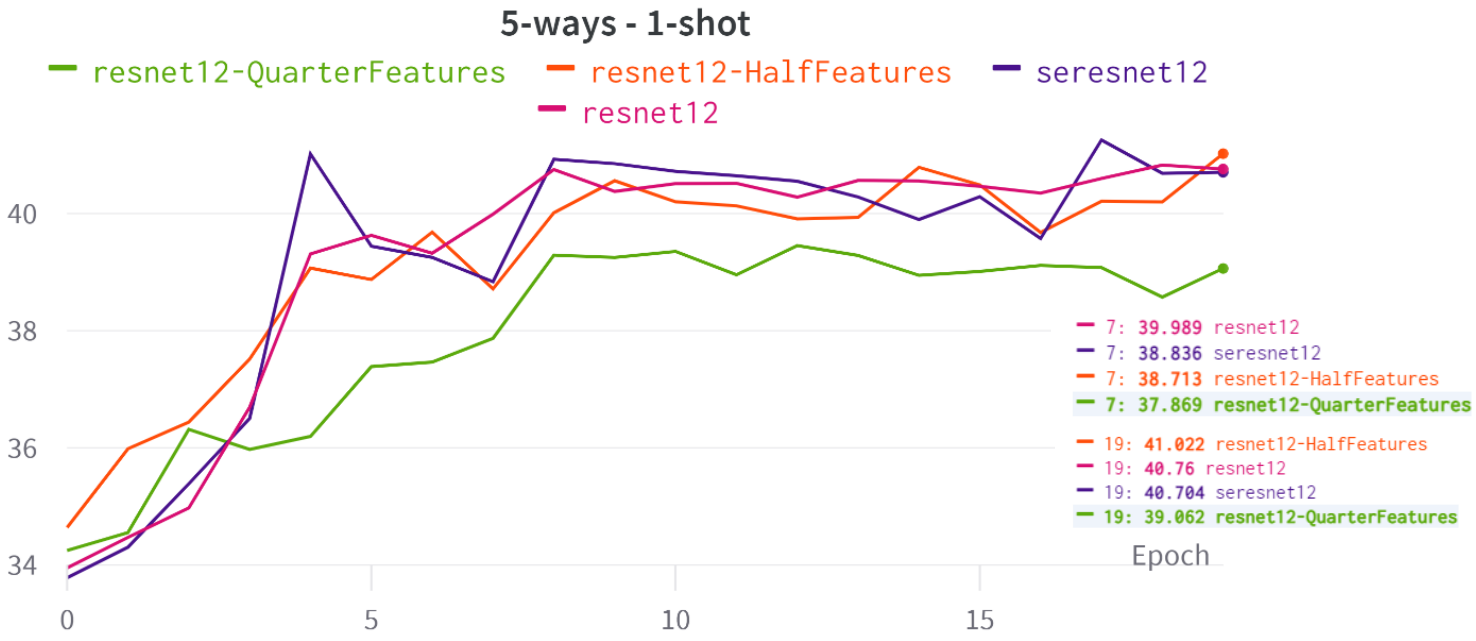


Σχήμα 3.2.2.2 - 1: Διάγραμμα σύγκρισης επίδρασης πλάτους των δικτύων και της συμπερίληψης ενός SE-block 5-ways – 1-shot για το CIFAR-FS dataset.

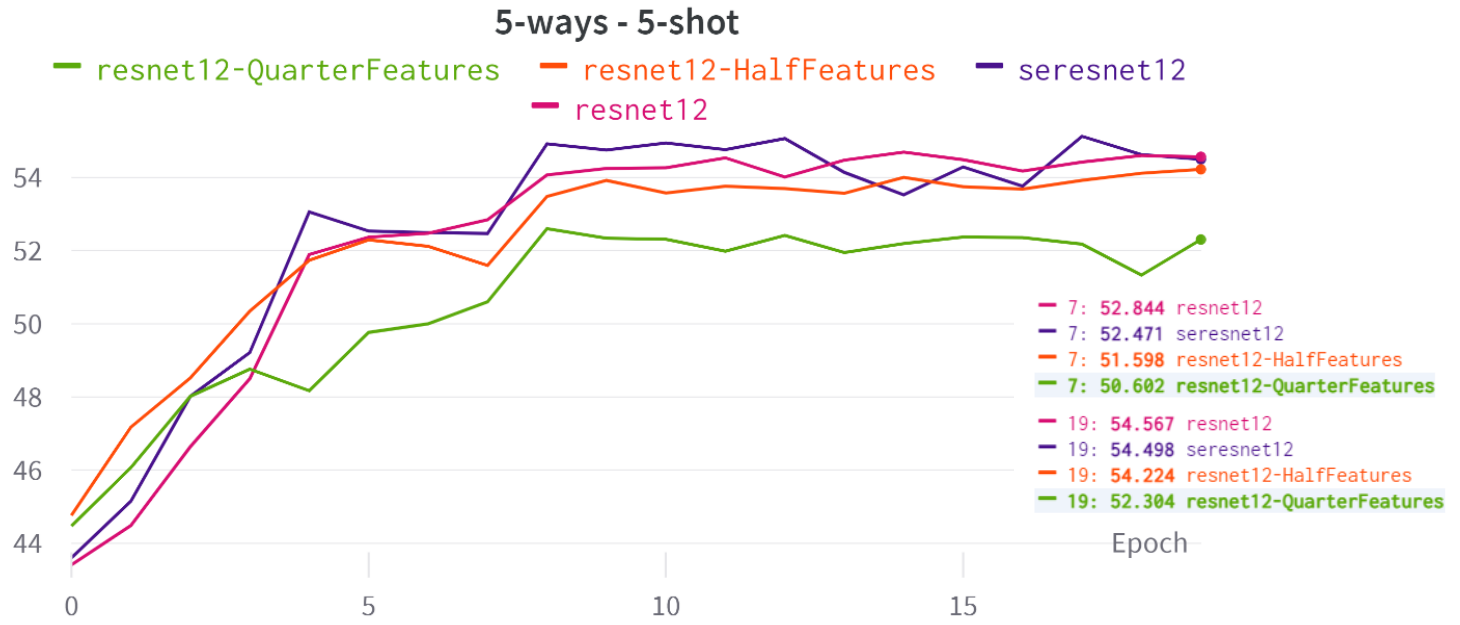


Σχήμα 3.2.2.2 - 2: Διάγραμμα σύγκρισης επίδρασης πλάτους των δικτύων και της συμπερίληψης ενός SE-block 5-ways – 5-shot για το CIFAR-FS dataset.

FC100 dataset



Σχήμα 3.2.2.2 - 3: Διάγραμμα σύγκρισης επίδρασης πλάτους των δικτύων και της συμπερίληψης ενός SE-block 5-ways – 1-shot για το FC100 dataset.



Σχήμα 3.2.2.2 - 4: Διάγραμμα σύγκρισης επίδρασης πλάτους των δικτύων και της συμπερίληψης ενός SE-block 5-ways – 5-shot για το FC100 dataset.

3.2.2.3 Παρατηρήσεις-Συμπεράσματα:

Στα διαγράμματα των Σχημάτων 3.2.2.2 - 1, 3.2.2.2 - 2, 3.2.2.2 - 3, και 3.2.2.2 - 4 αναπαρίσταται ο έλεγχος στο Few-Shot Image Classification Task σύμφωνα με τον αλγόριθμο Generation-0 της [1] για τα δίκτυα που προαναφέραμε. Όσον αφορά τα στοιχεία αναπαράστασης των αξόνων, καθώς και τις αναγραφόμενες ετικέτες επιδόσεων των μοντέλων για κάθε dataset, ακολουθείται πάλι, ακριβώς η ίδια γραμμή που τέθηκε κατά την 1η σύγκριση.

Σχετικά με την μελέτη της επίδρασης του μεγέθους του πλάτους των δικτύων, φαίνεται ότι η μείωση του αριθμού των συνελικτικών φίλτρων είναι αντιστρόφως ανάλογη με τις επιδόσεις των μοντέλων στο Few-Shot Task του αλγορίθμου Generation-0 της [1]. Και στα δύο Datasets, επιβεβαιώνεται σαφής διάκριση του μοντέλου που έχει τα περισσότερα features στα Residual Blocks του.

Αναλυτικότερα, για το CIFAR-FS dataset, το οποίο όπως αναφέρθηκε είναι το πλέον αξιόπιστο για την εξαγωγή συμπερασμάτων, έχουμε:

- για το 1-shot: **Resnet12** > Resnet12-HalfFeatures > Resnet12-QuarterFeatures με αντίστοιχες τιμές επιδόσεων 73.713±0.94, 71.016±0.9, και 66.798±0.94, και
- για το 5-shot: **Resnet12** > Resnet12-HalfFeatures > Resnet12-QuarterFeatures με τις τιμές των επιδόσεων των μοντέλων να διαμορφώνονται σε 86.556±0.61, 84.409±0.66, και 81.3±0.67 αντίστοιχα.

Η κατάταξη **Resnet12** > Resnet12-HalfFeatures > Resnet12-QuarterFeatures φαίνεται πως επαληθεύεται και στην περίπτωση του FC100 dataset, όπου:

- στο 1-shot επιδείχθηκαν τιμές επιδόσεων 39.989±0.71, 38.713±0.74 και 37.869±0.73 για τα Resnet12, Resnet12-HalfFeatures και Resnet12-QuarterFeatures αντίστοιχα, ενώ
- στο 5-shot, οι τιμές διαμορφώθηκαν ανάλογα σε 52.844±0.69, 51.598±0.74 και 50.602±0.7.

Αξίζει να αναφερθεί, πως η διαφορά του Resnet12 από το τελευταίο σε επίδοση Resnet12-QuarterFeatures για το CIFAR-FS dataset είναι ~6.9%, καθόλου αμελητέα αν λάβουμε υπόψη την μη-ευαισθησία που διακρίνει την συγκεκριμένη συλλογή. Συνεπώς, μπορούμε να συμπεράνουμε, πως η αύξηση του μεγέθους του πλάτους των δικτύων (τουλάχιστον μέχρι το σημείο που έχει εφαρμοστεί στην [1]) επιφέρει βελτίωση στην επίδοση του μοντέλου Resnet12 στο Few-Shot Image Classification task του Generation-0 της [1]. Τα συνολικά ευρήματα συνοψίζονται στον Πίνακα 3.2.2.3 - 1.

	CIFAR-FS		FC-100 (8η Εποχή)	
Μοντέλο	5-ways - 1-shot	5-ways - 5-shot	5-ways - 1-shot	5-ways - 5-shot
Resnet12	**73.713±0.94**	**86.556±0.61**	**39.989±0.71**	**52.844±0.69**
Resnet12-HalfFeatures	71.016±0.9	84.409±0.66	38.713±0.74	51.598±0.74
Resnet12-QuarterFeatures	66.798±0.94	81.3±0.67	37.869±0.73	50.602±0.7

Πίνακας 3.2.2.3 - 1: Συγκεντρωτικός πίνακας των τελικών αποτελεσμάτων των μοντέλων, όσον αφορά την διερεύνηση της επίδρασης του πλάτους των δικτύων.

Όσον αφορά την προσθήκη του SE block στο Resnet12, τα αποτελέσματα έδειξαν πως δεν φαίνεται να έχει κάποια σημαντική συμβολή στην βελτίωση της επίδοσης για το Few-Shot Image Classification του Generation-0 της [1]. Οι επιδόσεις των Resnet12 και Seresnet12 είναι πάρα πολύ κοντά μεταξύ τους, και με βάση το CIFAR-FS dataset, τα πειράματα έδειξαν πως στο 1-shot task υπερτερεί το Seresnet12, ενώ στο 5-shot task υπερτερεί το Resnet12. Παρ' όλ' αυτά οι διαφορές στις επιδόσεις μεταξύ των δύο είναι

αμελητέες, γεγονός που δεν επιτρέπει την εξαγωγή κάποιου ασφαλούς συμπεράσματος.[26] Συγκεκριμένα, για το CIFAR-FS dataset έχουμε, για το 1-shot: 73.713±0.94 (Resnet12) έναντι 73.771±0.91 (Seresnet12), ενώ για το 5-shot: 86.556±0.61 (Resnet12) έναντι 86.356±0.62 (Seresnet12).

Τέλος, αναφορικά και μόνο, για το FC-100 dataset[27], οι επιδόσεις που καταγράφηκαν έχουν ως εξής, για το 1-shot: 39.989±0.71 (Resnet12) έναντι 38.836±0.75 (Seresnet12), ενώ για το 5-shot: 52.844±0.69 (Resnet12) έναντι 52.471±0.73 (Seresnet12).

Ο πίνακας 3.2.2.3 - 2 συνοψίζει τα όλα τα αποτελέσματα όσον αφορά την μελέτη της συνεισφοράς του προστιθέμενου SE Block.

	CIFAR-FS		FC-100 ($8^{η}$ Εποχή)	
Μοντέλο	5-ways - 1-shot	5-ways - 5-shot	5-ways - 1-shot	5-ways - 5-shot
Resnet12	73.713±0.94	**86.556±0.61**	**39.989±0.71**	**52.844±0.69**
Seresnet12	**73.771±0.91**	86.356±0.62	38.836±0.75	52.471±0.73

Πίνακας 3.2.2.3 - 2: Συγκεντρωτικός πίνακας των τελικών αποτελεσμάτων των μοντέλων, όσον αφορά την συνεισφορά του προστιθέμενου SE Block.

[26] Εάν έπρεπε αναγκαστικά να εξαχθεί κάποιο συμπέρασμα μεταξύ των δύο επιδόσεων για τα δίκτυα Resnet12 και Seresnet 12, αυτό θα ήταν ότι το Resnet12 επιδεικνύει ελαφρώς καλύτερα αποτελέσματα.

[27] Η ευαισθησία του FC-100 dataset επιβεβαιώθηκε και με την πραγματοποίηση ενός ακόμη πειράματος το οποίο εκτελέστηκε δύο φορές με ακριβώς τις ίδιες συνθήκες και υπερπαραμέτρους. Πιο συγκεκριμένα, για το μοντέλο Resnet18 τα αποτελέσματα στο 1-shot task διαφέραν έως ~1.8% (37.533±0.74 έναντι 35.716±0.7). Αξίζει να σημειωθεί ότι η διαφορά της τάξης ~1.8% είναι ικανή να στερήσει και να «κρύψει» την πληροφορία αξιοσημείωτων επιδράσεων όπως αυτή του πλάτους των δικτύων. Για παράδειγμα θα ήταν πολύ δύσκολο να ανιχνευτεί η διαφορά του μοντέλου Resnet12-HalfFeatures από το Resnet12-QuarterFeatures η οποία είναι μόλις ~0.9% για το 1-shot task, χωρίς την αξιολόγηση στο CIFAR-FS Dataset όπου σημειώθηκε διαφορά της τάξης ~4.2%.

3.2.3 Σύγκριση 3η

Στην παρούσα σύγκριση εξετάζεται η υπόθεση για το αν αυξάνονται οι επιδόσεις των μοντέλων στο πρόβλημα του Few-Shot Image Classification, με την χρήση μίας άλλης βοηθητικής self-supervised διεργασίας στην οποία χρησιμοποιείται συγκεκριμένα ένας VAE (Variational Autoencoder - Στοχαστικός Αυτοκωδικοποιητής) [63].

3.2.3.1 Επεξήγηση της διαδικασίας

Η βάση και η λογική του αλγορίθμου που χρησιμοποιήθηκε είναι παρόμοια με αυτήν της Generation-0 της [1], δηλαδή κατά την εκπαίδευση του μοντέλου συμπεριλαμβάνεται μία διαφορετική αυτήν την φορά, βοηθητική διεργασία αύτο-εποπτευόμενης μάθησης. Συγκεκριμένα, η ιδέα είναι, πλην της βασικής διεργασίας κατηγοριοποίησης των εικόνων που λαμβάνει το μοντέλο, να το υποβάλλουμε συγχρόνως και στο να τις ανακατασκευάζει, και να εξετάσουμε το αν ο τρόπος αυτός εκπαίδευσης θα επιφέρει βελτιώσεις στις επιδόσεις για το Few-Shot Image Classification.

Η υλοποίηση της ιδέας έχει ως εξής, ο αλγόριθμος χρησιμοποιεί το δίκτυο του Κωδικοποιητή (Encoder) του VAE για κάθε παρτίδα δεδομένων, από το οποίο εξάγεται το feature κατηγοριοποίησης, και το δίκτυο του Αποκωδικοποιητή (Decoder) του VAE για την εξαγωγή ενός νέου δείγματος (ανακατασκευασμένων) εικόνων. Το feature του κωδικοποιητή χρησιμοποιείται στην βασική συνάρτηση κόστους (διεργασία του Classification), ενώ το νέο δείγμα εικόνων χρησιμοποιείται στην βοηθητική συνάρτηση (διεργασία Reconstruction). Έτσι το μοντέλο εκπαιδεύεται συγχρόνως στην ορθή κατηγοριοποίηση των εικόνων και στην ορθή ανακατασκευή αυτών. Και σε αυτήν την περίπτωση, η συνάρτηση Κόστους αποτελείται από τις δύο αντίστοιχες συνιστώσες, κι ορίζεται ως εξής:

$$L = a\, L_{Classification} + \beta\, L_{Reconstruction}$$

Όπου,

- ο όρος $\alpha\, L_{Classification}$ αντιστοιχίζεται στην βασική διεργασία, δηλαδή αυτή της ορθής αντιστοίχισης της εκάστοτε εικόνας στην ανάλογη ετικέτα, με α τον αντίστοιχο συντελεστή συνεισφοράς, και
- $\beta\, L_{Reconstruction}$ ο όρος της προστιθέμενης βοηθητικής διεργασίας, που αφορά την ανακατασκευή της εικόνας από τον VAE, με β τον αντίστοιχο συντελεστή συνεισφοράς.

3.2.3.2 Υλοποίηση του Αλγορίθμου και Μοντέλα Σύγκρισης

Για την υλοποίηση του παραπάνω αλγορίθμου χρησιμοποιήθηκε ως μοντέλο βάσης η αρχιτεκτονική του Resnet18 και ο αντίστοιχος VAE, που συμπεριλαμβάνεται στο Repository του Julian Stastny (https://github.com/julianstastny/VAE-ResNet18-PyTorch) [90], όπου μπορεί να ανατρέξει κανείς για να δει και την ακριβή αρχιτεκτονική. Τα χαρακτηριστικά και οι υπερπαράμετροι που χρησιμοποιήθηκαν για τα πειράματα αυτής της σύγκρισης είναι ακριβώς τα ίδια με τις προηγούμενες συγκρίσεις, τα οποία έχουν επεξηγηθεί παραπάνω και είναι σε πλήρη σύμπλευση με την υλοποίηση του αλγορίθμου Generation-0 της [1].

Πέρα από την βασική υπόθεση, που είναι το αν η επίδοση των μοντέλων αυξάνεται με την χρησιμοποίηση του παραπάνω αλγορίθμου, στην εν λόγω σύγκριση συμπεριλήφθηκε και το μοντέλο της selfsupervised διαδικασίας της [1] (rotation) προκειμένου να αποκτήσουμε και μία εικόνα σύγκρισης μεταξύ των δύο selfsupervised προσεγγίσεων.

Έτσι, τα μοντέλα που συμπεριλαμβάνονται σε αυτήν την σύγκριση είναι:

- ένα μοντέλο χωρίς καμία selfsupervised διαδικασία, χτισμένο ακριβώς πάνω στον κώδικα της [1], όπου η συνάρτηση κόστους αποτελείται μόνο από τον όρο της διεργασίας του Classification (resnet18-softmax),
- ένα μοντέλο στο οποίο εμπεριέχεται και η εν λόγω selfsupervised βοηθητική διεργασία με την χρησιμοποίηση του VAE της υλοποίησης της [90] (resnet18-VAE), και
- ένα μοντέλο που εκπαιδεύτηκε με την selfsupervised διεργασία του rotation, ακριβώς υπό τις συνθήκες που έχουν ορίσει οι Jathushan Rajasegaran et al. στον αλγόριθμο Generation-0 της [1] (resnet18-rotation).

Όσον αφορά τα χαρακτηριστικά των παραπάνω μοντέλων,

- μοντέλο resnet18-softmax: η μόνη διαφοροποιήση από τον κώδικα της [1] είναι η προσθήκη του μοντέλου resnet18 της [90] και η αφαίρεση της βοηθητικής διεργασίας του rotation.
- μοντέλο resnet18-rotation: εφαρμόστηκε στον κώδικα της [1], η αρχιτεκτονική resnet18 (του encoder του VAE) της υλοποίησης του Stastny [90], αυτή την φορά με την συμπερίληψη της selfsupervised βοηθητικής διεργασίας του rotation.
- μοντέλο resnet18-VAE: πέρα από την τοποθέτηση της αρχιτεκτονικής του Julian Stastny [90], το σύστημα διαφοροποιείται,
    1) στο είδος της χρησιμοποιούμενης βοηθητικής συνάρτησης κόστους όπου επιλέχθηκε η MAE (Mean Absolute Error – L1 Loss) έναντι της BCE (Binary Cross Entropy with logits) που χρησιμοποιείται στο μοντέλο του rotation,
    2) στον συνδυασμό των συντελεστών συνεισφοράς των χρησιμοποιούμενων συναρτήσεων κόστους, όπου επιλέχθηκε ο συνδυασμός ($\alpha = 0.25, \beta = 0.75$) για την βασική και βοηθητική διεργασία αντίστοιχα, έναντι των ($\alpha = 1, \beta = 0$) του μοντέλου resnet18-softmax και των ($\alpha = 1, \beta = 2$) του μοντέλου resnet18-rotation.
    3) στον τρόπο από τον οποίο προκύπτει το χρησιμοποιούμενο feature, το οποίο αξιοποιείται στην βασική διεργασία της κατηγοριοποίησης και στην τελική αξιολόγηση στο Few-Shot Image Classification Task. Στην συγκεκριμένη περίπτωση, για την εξαγωγή του feature χρησιμοποιήθηκε ένα παράλληλο προστιθέμενο γραμμικό στρώμα (nn.Linear) στο σημείο ακριβώς πριν από το ενδιάμεσο στρώμα του VAE (latent layer), σε αντίθεση με την υλοποίηση της [1] η οποία είναι σειριακή και το γραμμικό στρώμα από το οποίο προκύπτει το τελικό feature του μοντέλου είναι το ίδιο που δίνει και τις εισόδους στο αντίστοιχο γραμμικό layer του rotation που χρησιμοποιείται για την πρόβλεψη του βαθμού περιστροφής.

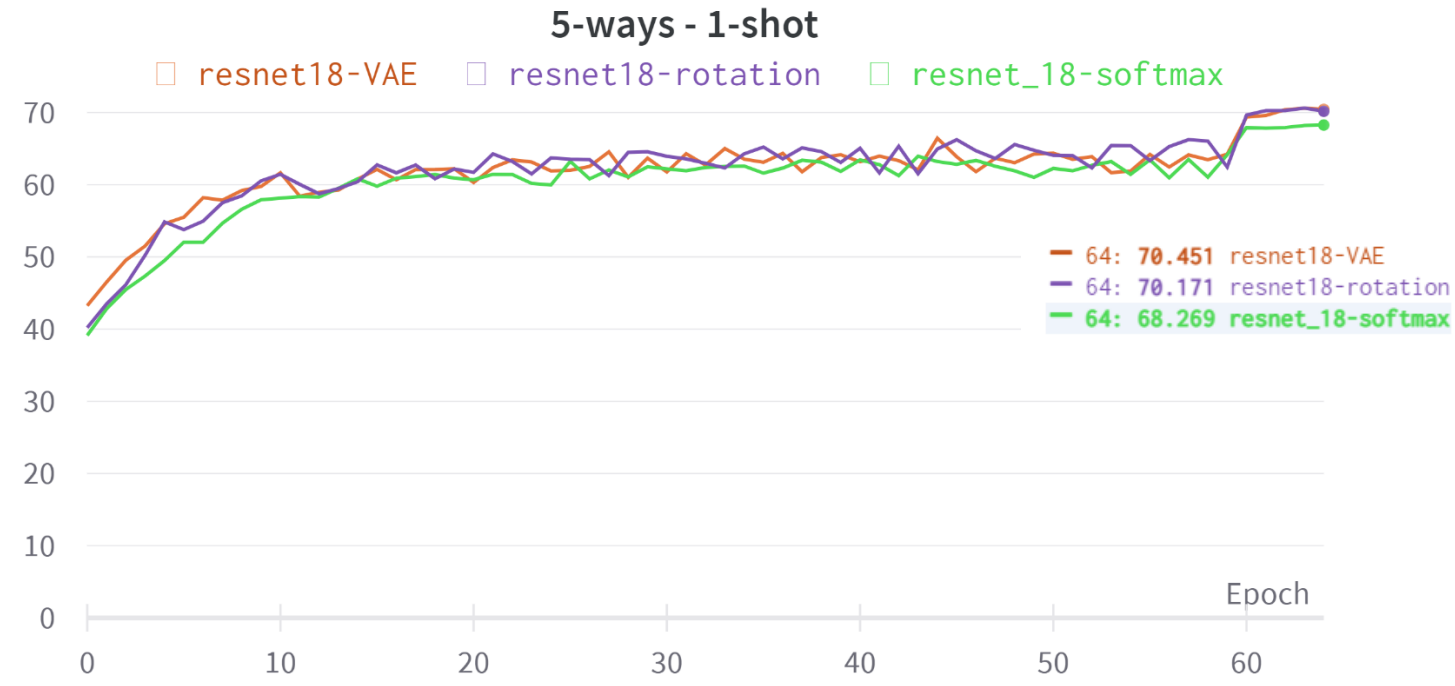
    Επίσης αξίζει να σημειωθεί, ότι ο αριθμός των νευρώνων του ενδιάμεσου στρώματος αναπαράστασης (latent layer) του VAE, το οποίο κωδικοποιεί την πληροφορία των εισόδων, τέθηκε ίσος με 64.

Η επιλογή των χαρακτηριστικών και των υπερπαραμέτρων της υλοποίησης του μοντέλου Resnet18-VAE είναι βασισμένη σε μελέτη η οποία αναλύεται σε επόμενη παράγραφο.

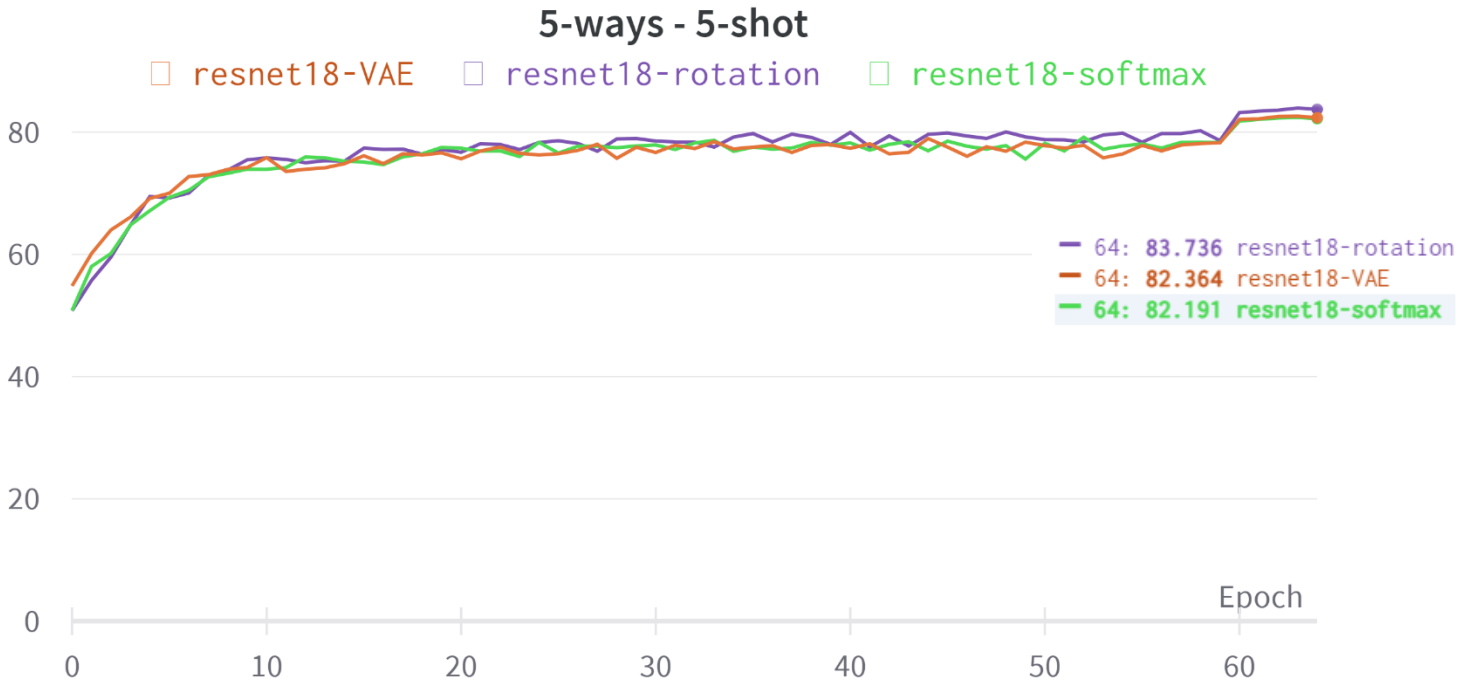
Κατά την σύγκριση αυτήν, η μελέτη εστιάζεται στο CIFAR-FS dataset καθώς σύμφωνα με τα πειράματα είναι το πλέον αξιόπιστο. Επίσης συμπληρωματικά και μόνο, η ίδια σύγκριση λαμβάνει χώρα και στο FC100 dataset, παρά την ευαισθησία που το διακατέχει, προκειμένου να εξεταστεί αν τα συμπεράσματα που προέκυψαν κατά την διερεύνηση στο CIFAR-FS dataset μεταφέρονται και σε αυτήν την συλλογή αναφοράς.

3.2.3.3 Διαγράμματα Αποτελεσμάτων και Συμπεράσματα

CIFAR-FS dataset



Σχήμα 3.2.3.3 - 1: Διάγραμμα σύγκρισης επίδρασης της συμπερίληψης Selfsupervision και του είδους αυτής 5-ways – 1-shot για το CIFAR-FS dataset.



Σχήμα 3.2.3.3 - 2: Διάγραμμα σύγκρισης επίδρασης της συμπερίληψης Selfsupervision και του είδους αυτής 5-ways – 5-shot για το CIFAR-FS dataset.

Σύμφωνα με τα διαγράμματα των Σχημάτων 3.2.3.3 - 1 και 3.2.3.3 - 2: [28]

- στο 1-shot Task η κατάταξη των επιδόσεων είναι **resnet18-VAE** (70.451%) > resnet18-rotation (70.171%) > resnet18-softmax (68.269%), δηλαδή το μοντέλο της VAE selfsupervised διεργασίας, κατάφερε να υπερισχύσει και του μοντέλου με την selfsupervised διεργασία του rotation και του μοντέλου χωρίς καθόλου selfsupervision. Μάλιστα, μπορούμε να διακρίνουμε ότι τα δύο μοντέλα με τις selfsupervised διεργασίες έχουν ένα σαφές προβάδισμα έναντι του μοντέλου που δεν συμπεριλαμβάνει κάποια βοηθητική διεργασία.
- Στο 5-shot Task, η κατάταξη των επιδόσεων αναδιαμορφώθηκε ως **resnet18-rotation** (83.736%) > resnet18-VAE (82.364%) > resnet18-softmax (82.191%), με το μοντέλο του rotation να έχει μία σαφή υπερίσχυση έναντι των άλλων δύο, και το μοντέλο χωρίς καθόλου selfsupervision να καταλαμβάνει και πάλι την 3η θέση.

Προκειμένου να εξαχθεί ένα πιο βέβαιο συμπέρασμα είναι θεμιτό να συνυπολογίσουμε τις επιδόσεις των παραπάνω μοντέλων και στο Validation Set. Για τον σκοπό αυτόν, δημιουργήθηκε ο Πίνακας 3.2.3.3 - 1, ο οποίος συνοψίζει όλα τα απαραίτητα αποτελέσματα.

	Test Set		Validation Set	
	5-ways - 1-shot	5-ways - 5-shot	5-ways - 1-shot	5-ways - 5-shot
Μοντέλο χωρίς selfsupervision (resnet18-softmax)	68.269 ± 0.95	82.191 ± 0.63	60.709 ± 0.97	75.129 ± 0.76
Μοντέλο με VAE selfsupervision (resnet18-VAE)	**70.451 ± 0.91**	82.364 ± 0.65	62.142 ± 1.04	75.851 ± 0.74
Μοντέλο με Rotation selfsupervision (resnet18-rotation)	70.171 ± 0.92	**83.736 ± 0.63**	**63.798 ± 1.05**	**77.682 ± 0.72**

Πίνακας 3.2.3.3 - 1: Συγκεντρωτικός πίνακας των επιδόσεων των παραπάνω μοντέλων για τις διεργασίες του 1-shot, 5-shot και για το Test Set και για το Validation Set.

Όσον αφορά την σύγκριση μεταξύ των δύο Selfsupervised διεργασιών, τα συγκεντρωτικά αποτελέσματα του Πίνακα 3.2.3.3 - 1 δείχνουν ότι η διεργασία του rotation υπερισχύει σε όλες τις περιπτώσεις πλην της 1-shot του Test Set, στην οποία όμως υπολείπεται για μόλις ~0.3%, διαφορά πολύ μικρότερη από αυτές στις οποίες επικρατεί στις υπόλοιπες περιπτώσεις. Συγκεκριμένα στο 5-shot σημειώνει ~1.4% και ~1.9% καλύτερη επίδοση, για το Test Set και το Validation Set αντίστοιχα, ενώ για το 1-shot του Validation Set επικρατεί με ~1.6%. Συνεπώς, η χρησιμοποίηση του Selfsupervision με την διεργασία του rotation επιδεικνύει καλύτερα αποτελέσματα στο Few-shot Image Classification Task, συγκριτικά με αυτά της βοηθητικής διεργασίας του VAE.

Τέλος, αναφορικά με την αρχική υπόθεση της παρούσας σύγκρισης, η συνολική πληροφορία που αποδίδεται από τον Πίνακα 3.2.3.3 - 1 δείχνει πως το μοντέλο resnet18-VAE επικρατεί του μοντέλου resnet18-softmax σε όλες τις περιπτώσεις. Συγκεκριμένα, στο 1-shot task σημειώνει βελτιώσεις βαθμού ~2.2% και ~1.4% για το Test και το Validation Set αντίστοιχα. Στο 5-shot task δεν φαίνεται να εμφανίζεται ο ίδιος βαθμός προόδου, παρ' όλ' αυτά και πάλι επιδεικνύονται καλύτερα αποτελέσματα, συγκεκριμένα κατά ~0.2% για το Test Set και κατά ~0.7% για το Validation Set.

[28] Εδώ αξίζει να σημειωθεί πως για τα στοιχεία αναπαράστασης των διαγραμμάτων των Σχημάτων 3.2.3.3 - 1 & 3.2.3.3 - 2 ακολουθείται η ίδια γραμμή με τις δύο πρώτες περιπτώσεις συγκρίσεων, δηλαδή στον οριζόντιο άξονα περιλαμβάνεται ο αριθμός της εκάστοτε εποχής εκπαίδευσης (με την αρίθμηση να ξεκινάει από το 0), ενώ στον κάθετο άξονα περιλαμβάνεται το ποσοστό απόδοσης που αντιστοιχίζεται στην κάθε εποχή. Υπενθυμίζουμε ότι η ακριβής τιμή του ποσοστού απόδοσης υπολογίζεται από την κανονικοποιημένη τιμή στην ποσοστιαία κλίμακα του 100 (%), του μέσου όρου των τιμών που προκύπτουν από τα 600 επεισόδια αξιολόγησης τα οποία λαμβάνουν χώρα έπειτα από την κάθε εποχή εκπαίδευσης.

Επομένως, συνυπολογίζοντας τα συνολικά ευρήματα, μπορούμε να συμπεράνουμε πως η προσθήκη της βοηθητικής selfsupervised διεργασίας του VAE επέφερε βελτιώσεις στο πρόβλημα κατηγοριοποίησης εικόνων μέσα από ελάχιστα παραδείγματα. Εν κατακλείδι, σύμφωνα με όλα τα παραπάνω, η τελική κατάταξη των μοντέλων για το πρόβλημα του Few-Shot είναι resnet18-rotation > renet18-VAE > resnet18-softmax.[29]

FC100 dataset

Λαμβάνοντας υπόψη την ευαισθησία που χαρακτηρίζει την εν λόγω συλλογή αναφοράς, κρίθηκε πως δεν χρειάζεται να γίνει ανάλυση στις ακριβείς τιμές των επιδόσεων των μοντέλων. Αντί αυτού, θεωρήθηκε προτιμότερο το ζήτημα να εξεταστεί μακροσκοπικά και η εξέταση των παραπάνω συμπερασμάτων να προκύψει από τις μορφές καμπυλών των επιδόσεων των μοντέλων στις διεργασίες few-shot. Για τον σκοπό αυτό, παρατίθεται το Σχήμα 3.2.3.3 - 1, το οποίο περιέχει ένα συγκεντρωτικό σύνολο διαγραμμάτων, όπου συμπεριλαμβάνονται όλες οι περιπτώσεις που μπορούν να δώσουν αυτήν την πληροφορία.

Σχήμα 3.2.3.3 - 1: Διαγράμματα σύγκρισης επίδρασης της συμπερίληψης Selfsupervision και του είδους αυτής, για όλες τις πιθανές περιπτώσεις για το FC100 dataset.

Σύμφωνα με τα γραφήματα του Σχήματος 3.2.3.3 - 1, μπορούμε να διακρίνουμε ότι η πράσινη καμπύλη, η οποία αντιστοιχίζεται στο μοντέλο που δεν εμπεριέχει καμία βοηθητική διεργασία αύτο-εποπτευόμενης μάθησης (resnet18-softmax), είναι χαμηλότερα από αυτές των άλλων δύο μοντέλων για το Test Set. Το ίδιο φαίνεται να συμβαίνει και για το Validation Set, παρότι βρίσκεται αρκετά κοντά με το μοντέλο που εμπεριέχει την VAE selfsupervised διεργασία (πορτοκαλί καμπύλη). Επιπλέον, σε όλες τις περιπτώσεις, υπάρχει σαφής υπερίσχυση του μοντέλου της rotation selfsupervised διεργασίας (resnet18-rotation - μωβ καμπύλη), εκτός από την περίπτωση 1-shot του Test Set όπου το προβάδισμα του έναντι του μοντέλου με την VAE selfsupervised διεργασία, δεν είναι και τόσο ευδιάκριτο.

Συνεπώς, για το συνολικό πρόβλημα του few-shot classification, σύμφωνα με τα παραπάνω ευρήματα, φαίνεται πως η κατάταξη (resnet18-rotation > resnet18-VAE > resnet18-softmax) που εξάχθηκε από την μελέτη του CIFAR-FS dataset επαληθεύεται και στην περίπτωση του FC100-dataset.

[29] Άξιο επισήμανσης είναι ότι η σχέση της απόδοσης των μοντέλων στο Few-Shot Task και της γενικής ευστοχίας αυτών δεν είναι ευθέως ανάλογη, καθώς για την παραπάνω περίπτωση, στο Validation Accuracy η κατάταξη των μοντέλων διαμορφώθηκε ως resnet18-softmax > resnet18-rotation > resnet18-VAE.

3.2.3.4 Μελέτη για την ρύθμιση και την επιλογή των Υπερπαραμέτρων του μοντέλου Resnet18-VAE

Κατά την υλοποίηση της ιδέας της βοηθητικής διεργασίας με την χρήση του VAE, δημιουργήθηκαν τρία ερωτήματα όσον αφορά την ρύθμιση του συστήματος, τα οποία είναι:

1. Ποια είναι η συνάρτηση κόστους που θα χρησιμοποιηθεί για την βοηθητική διεργασία,
2. Τι τιμές θα δοθούν στους συντελεστές συνεισφοράς της βασικής και την βοηθητικής συνάρτησης κόστους (δηλαδή, η ρύθμιση της υπερπαραμέτρων α, β που εμπεριέχονται στο συνολικό κόστος),
3. Με ποιον τρόπο θα προκύψει το τελικό feature του μοντέλου και τι αριθμό νευρώνων θα ορίσουμε για το ενδιάμεσο στρώμα αναπαράστασης (latent dimension) του Αυτοκωδικοποιητή.

Παρακάτω παρουσιάζεται η μελέτη και ο τρόπος με τον οποίο προσεγγίστηκαν τα παραπάνω ερωτήματα. Συγκεκριμένα, πραγματοποιήθηκαν πειράματα στην βάση του κώδικα της [1], με τα ίδια χαρακτηριστικά και υπερπαραμέτρους συστήματος του αλγόριθμου Generation-0 που έχουν επεξηγηθεί στην αρχή αυτού του κεφαλαίου. Τα πειράματα έλαβαν χώρα στο CIFAR-FS dataset, το οποίο είναι καταλληλότερο για την εξαγωγή συμπερασμάτων, και η επιλογή και η ρύθμιση των υπερπαραμέτρων έγινε με βασικό κριτήριο τις επιδόσεις των μοντέλων στα 1-shot και 5-shot tasks του Test Set, ενώ επικουρικά χρησιμοποιήθηκαν και οι αντίστοιχες επιδόσεις στο Validation Set. Ως σημείο εκκίνησης, επιλέχθηκε η διερεύνηση των ερωτημάτων να πραγματοποιηθεί με την σειρά που αυτά παρουσιάστηκαν παραπάνω.

3.2.3.4.1 Επιλογή Συνάρτησης Κόστους Βοηθητικής Διεργασίας

Στην διερεύνηση για την επιλογή της συνάρτησης κόστους της βοηθητικής διεργασίας οι πιθανές συναρτήσεις προς επιλογή ήταν τρεις:

1. η MAE (Mean Absolute Error – L1-Loss),
2. η MSE (Mean Squared Error – L2-loss) η οποία σύμφωνα με την μελέτη [86] πάνω στους Αυτοκωδικοποιητές, είναι η πιο συνήθης χρησιμοποιούμενη για αυτές τις περιπτώσεις, και
3. η BCE (Binary Cross Entropy), η οποία είναι και η προτεινόμενη από τα documentations του Pytorch όσον αφορά την ανακατασκευή των Autoencoders.

Ως συνθήκες αφετηρίας χρησιμοποιήθηκαν:

- η συνεισφορά της βασικής και βοηθητικής συνάρτησης κόστους να είναι η ίδια (δηλαδή a = 1 και β = 1),
- για το σημείο από το οποίο εξάγεται το τελικό Feature του μοντέλου, προστέθηκε ένα παράλληλο γραμμικό στρώμα (nn.Linear) ακριβώς πριν από το latent layer του VAE,
- ο αριθμός των νευρώνων του ενδιάμεσου στρώματος είναι ίσος με 20 (δηλαδή, latent dimension=20), που αποτελεί και την default τιμή στην υλοποίηση του Julian Stastny [90].

Αρχικά διενεργήθηκαν πειράματα για την εξέταση των επιδόσεων των συναρτήσεων MAE και MSE. Η BCE, προκειμένου να επιτελέσει την λειτουργία για την οποία προορίζεται, προϋποθέτει οι τιμές της να βρίσκονται στο διάστημα από 0 μέχρι 1 (δηλαδή στο εύρος [0,1]). Αυτό συμβαίνει για την έξοδο που δίνει ο VAE της [90], αλλά δεν συμβαίνει για τις εικόνες που δέχεται το δίκτυο. Αιτία αυτού αποτελεί ότι κατά την υλοποίηση του αλγορίθμου της [1], στο Data Augmentation που γίνεται, λαμβάνει χώρα και η κανονικοποίηση των εικόνων ανά κανάλι (Channel Normalization) γεγονός που δημιουργεί εισόδους εικόνων με τιμές και εκτός του εύρους [0,1]. Έτσι, επιστρατεύτηκαν δύο τρόποι αντιμετώπισης αυτού του ζητήματος και την ορθή χρησιμοποίηση της BCE.

1. Το πέρασμα από μία σιγμοειδή συνάρτηση (nn.Sigmoid) των εικόνων ακριβώς πριν δοθούν στην BCE (παρ' όλ' αυτά οι είσοδοι που δίνονται στον Αυτοκωδικοποιητή δεν είναι στο εύρος [0,1]).

2. Η αφαίρεση της κανονικοποίησης ανά κανάλι από το Data Augmentation που λαμβάνει χώρα, προκειμένου και οι είσοδοι που δίνονται στον Αυτοκωδικοποιητή να είναι στο εύρος [0,1].

Έτσι, πραγματοποιήθηκαν δύο πειράματα για την συνάρτηση κόστους BCE, ένα με την χρησιμοποίηση του σιγμοειδούς στρώματος όπως αναφέραμε και ένα απλά χωρίς Channel Normalization. Μετέπειτα διενεργήθηκαν κι άλλα δύο πειράματα για τις συναρτήσεις κόστους MAE και MSE προκειμένου να εξεταστούν και οι επιδόσεις αυτών στις ίδιες συνθήκες του 2ου πειράματος της BCE (δηλαδή, χωρίς Channel Normalization, δίνοντας όμως εισόδους στον Αυτοκωδικοποιητή που κυμαίνονται στο εύρος [0,1]).

Τα αποτελέσματα συνοψίζονται στον Πίνακα 3.2.3.4.1 - 1, όπου ακολουθεί το τελικό συμπέρασμα και ο ακριβής τρόπος επιλογής της συνάρτησης κόστους για την βοηθητική διεργασία.

	Test Set		Validation Set	
Συνάρτηση Κόστους Βοηθητικής Διεργασίας	5-ways - 1-shot	5-ways - 5-shot	5-ways - 1-shot	5-ways - 5-shot
BCE (με Sigmoid)	68.956 ± 0.89	82.104 ± 0.63	60.829 ± 1.02	75.26 ± 0.76
MAE	69.242 ± 0.9 (2η θέση)	**82.353 ± 0.65** (1η θέση)	61.227 ± 1.01 (2η θέση)	75.584 ± 0.76 (3η θέση)
MSE	68.511 ± 0.92	81.767 ± 0.65	60.829 ± 1.03	75.067 ± 0.78
BCE (χωρίς Channel Normalization)	68.651 ± 0.9	81.993 ± 0.63	61.136 ± 1.03	75.36 ± 0.78
MAE (χωρίς Channel Normalization)	68.887 ± 0.91 (4η θέση)	81.849 ± 0.67 (5η θέση)	**61.344 ± 1.02** (1η θέση)	**76.211 ± 0.75** (1η θέση)
MSE (χωρίς Channel Normalization)	**69.249 ± 0.89** (1η θέση)	82.249 ± 0.65 (2η θέση)	60.913 ± 1.03 (3η θέση)	75.587 ± 0.76 (2η θέση)

Πίνακας 3.2.3.4.1 - 1: Συγκεντρωτικός πίνακας των επιδόσεων των βοηθητικών συναρτήσεων Κόστους υπό τις συνθήκες που εξηγήθηκαν παραπάνω.

Τις δύο καλύτερες επιδόσεις στα 5-shot και 1-shot tasks στο Test Set σημείωσαν οι συναρτήσεις MAE και MSE-χωρίς Channel Normalization. Επιλέχθηκε η MAE καθώς επέδειξε την καλύτερη επίδοση στο 5-shot task του Test Set, ενώ και στο 1-shot task καταλαμβάνει την 2η θέση με διαφορά δεκαδικών ψηφίων. Σε αυτήν την απόφαση συντέλεσε και το Validation Set, όπου ξεχώρισε η MAE-χωρίς Channel Normalization, η οποία όμως είναι αρκετά χαμηλά στην κατάταξη όσον αφορά το Test Set. Και εδώ, οι MAE και MSE-χωρίς Channel Normalization καταλαμβάνουν τις επόμενες δύο θέσεις. Η MAE καταλαμβάνει την 2η θέση στο 1-shot task, ενώ στο 5-shot χάνει την 2η θέση από την MSE-χωρίς Channel Normalization και πάλι με διαφορά δεκαδικών ψηφίων. Συνολικά, κρίθηκε ότι για το συγκεκριμένο Few-Shot task τις καλύτερες επιδόσεις επέδειξε η MAE.

3.2.3.4.2 Επιλογή των τιμών των συντελεστών συνεισφοράς

Στην παράγραφο αυτή παρατίθεται η μελέτη για την βέλτιστη επιλογή των τιμών των συντελεστών συνεισφοράς α και β, για την βασική (Classification Task) και την βοηθητική διεργασία (Reconstruction Task) αντίστοιχα.

Για τα πειράματα αυτής της διερεύνησης, επιλέχθηκαν:

- ως βοηθητική συνάρτηση ανακατασκευής η συνάρτηση MAE, η οποία επέδειξε τις καλύτερες επιδόσεις σύμφωνα με την μελέτη της προηγούμενης παραγράφου,
- το σημείο από το οποίο προκύπτει το τελικό Feature του μοντέλου προέρχεται και πάλι από ένα προστιθέμενο παράλληλο γραμμικό layer (nn.Linear) πριν από το latent layer του VAE, και

- ο αριθμός των νευρώνων του latent είναι ίσος με 20 (latent dimension=20) όπως στην υλοποίηση [90].

Οι συνδυασμοί των α, β που διερευνήθηκαν για τις επιδόσεις στο Few-Shot Task για τον εν λόγω αλγόριθμο του Αυτοκωδικοποιητή είναι οι εξής:

1. α = 1 και β = 2, ως έμπνευση από την [1], καθώς πρόκειται για τους συνδυασμούς που χρησιμοποιούνται στην selfsupervised διεργασία (rotation) του αλγόριθμου Genaration-0.
2. α = 1 και β = 1, για να ελεγχθεί η επίδοση της ίδιας συνεισφοράς των δύο διεργασιών,

Και επιπλέον, προκειμένου να εξαχθεί ένα συμπέρασμα για το ποια από τις δύο συναρτήσεις επηρεάζει περισσότερο τις τελικές επιδόσεις τέθηκαν και οι συνδυασμοί,

3. α = 0.75 και β = 0.25,
4. α = 0.25 και β = 0.75, και
5. α = 0.5 και β = 0.5.

Ο λόγος που πραγματοποιήθηκε ένα επιπλέον πείραμα με την ίδια συνεισφορά των δύο διεργασιών (α = 0.5, β = 0.5) είναι για να υπάρχει πιο σαφής σύγκριση μεταξύ των τριών τελευταίων, καθώς στην περίπτωση όπου οι συντελεστές είναι μικρότεροι της μονάδας, η πραγματική τιμή του συνολικού κόστους συρρικνώνεται, γεγονός που επηρεάζει σε έναν βαθμό την όλη διαδικασία.

Στην συνέχεια παρατίθεται ο Πίνακας 3.2.3.4.2 - 1 ο οποίος συνοψίζει τα αποτελέσματα των παραπάνω περαμάτων και σχολιάζεται ο τρόπος με τον οποίο επιλέχθηκε ο βέλτιστος συνδυασμός των συντελεστών συνεισφοράς.

	Test Set		Validation Set	
(α, β)	5-ways - 1-shot	5-ways - 5-shot	5-ways - 1-shot	5-ways - 5-shot
(1, 2)	69.098 ± 0.89	81.842 ± 0.65	61.069 ± 1.02	75.698 ± 0.75
(1, 1)	69.242 ± 0.9 (4η θέση)	**82.353 ± 0.65** (1η θέση)	61.227 ± 1.01 (3η θέση)	75.584 ± 0.76 (3η θέση)
(0.75, 0.25)	69.242 ± 0.9	82.109 ± 0.66	60.824 ± 1.06	75.136 ± 0.79
(0.25, 0.75)	**70.024 ± 0.91** (1η θέση)	82.3 ± 0.65 (2η θέση)	**62.033 ± 1.05** (1η θέση)	**75.824 ± 0.77** (1η θέση)
(0.5, 0.5)	69.318 ± 0.93	82.104 ± 0.66	61.582 ± 1.03	75.516 ± 0.75

Πίνακας 3.2.3.4.2 - 1: Συγκεντρωτικός πίνακας των επιδόσεων, των διάφορων συνδυασμών για τους συντελεστές συνεισφοράς βασικής και βοηθητικής συνάρτησης κόστους υπό τις παραπάνω συνθήκες.

Σχετικά με την σύγκριση των πειραμάτων στα οποία θέσαμε συντελεστές μικρότερης της μονάδας ((0.75, 0.25) , (0.25, 0.75), (0.5, 0.5)), φαίνεται πως η βοηθητική Loss επηρεάζει περισσότερο την απόδοση στο συγκεκριμένο Few-Shot Classification Task απ' ότι η βασική. Όπως μπορούμε να διακρίνουμε από τον Πίνακα 3.2.3.4.2 - 1, η σειρά κατάταξης (για τις τρεις αυτές περιπτώσεις) είναι **(0.25 , 0.75)** > (0.5 , 0.5) > (0.75 , 0.25) σε όλες τις συνθήκες, εκτός από την περίπτωση του 5-shot για το Test set όπου εναλλάσσεται η 2η με την 3η θέση (με διαφορά λίγων μόλις δεκαδικών ψηφίων), δηλαδή, **(0.25 , 0.75)** > (0.75 , 0.25) > (0.5 , 0.5).

Τα κριτήρια για την τελική επιλογή του συνδυασμού των συντελεστών α, β είναι όπως και προηγουμένως κατά κύριο λόγο οι επιδόσεις στα 1-shot και 5-shot Tasks του Test Set, και λαμβάνεται υπόψη και πάλι βοηθητικά το Validation Set. Από τα παραπάνω πειράματα, συμπεραίνουμε πως την καλύτερη συνολική επίδοση για το συγκεκριμένο πρόβλημα Few-Shot, επιδεικνύει ο συνδυασμός συντελεστών (α = 0.25, β =

0.75). Ο συγκεκριμένος συνδυασμός κατέγραψε την 1η καλύτερη επίδοση σε όλες τις περιπτώσεις πλην του 5-shot Task του Test Set όπου κατατάχθηκε στην 2η θέση, με διαφορά λίγων δεκαδικών ψηφίων από τον 1ο συνδυασμό (α = 1, β = 1), ο οποίος στις υπόλοιπες περιπτώσεις κατέλαβε χαμηλότερες θέσεις.

Τέλος, αξίζει να σημειωθεί, ότι σύμφωνα με την αντίστοιχη μελέτη που έγινε στην εργασία [1] των Jathushan Rajasegaran et al. και τα αποτελέσματα αυτής (πίνακας-3 [1]), φαίνεται πως η κατάλληλη ρύθμιση της συνεισφοράς της βοηθητικής συνάρτησης κόστους, μπορεί να δώσει ώθηση έως και ~1.5% στις επιδόσεις στο 1-shot task, για το CIFAR-FS dataset, ένα αξιοσημείωτο ποσοστό με βάση τα αποτελέσματα των πειραμάτων που έχουμε πραγματοποιήσει.

3.2.3.4.3 Επιλογή του τρόπου εξαγωγής του τελικού Feature και της ρύθμισης της υπερπαραμέτρου latent dimension

Στην παράγραφο αυτήν επιδιώκεται η εύρεση

- του καλύτερου τρόπου εξαγωγής του τελικού feature, που χρησιμοποιείται στην βασική διεργασία του Classification, και
- της βέλτιστης τιμής του αριθμού των νευρώνων του ενδιάμεσου στρώματος του VAE.

Τα χαρακτηριστικά που χρησιμοποιήθηκαν για αυτά τα πειράματα είναι αυτά που προέκυψαν από τις δύο προηγούμενες αναζητήσεις, δηλαδή

- η συνάρτηση MAE ως βοηθητική συνάρτηση κόστους, και
- ο συνδυασμός των συντελεστών συνεισφοράς α = 0.25 και β=0.75, για την βασική και την βοηθητική συνάρτηση αντίστοιχα.

3.2.3.4.3.1 Επιλογή τρόπου εξαγωγής Feature

Αρχικά, έλαβε χώρα μία σύγκριση μεταξύ ενός μοντέλου παράλληλης υλοποίησης (ακριβώς όπως αυτή που επιλέχθηκε στο σημείο εκκίνησης για την ευρύτερη μελέτη των υπερπαραμέτρων) και ενός μοντέλου σειριακής υλοποίησης, σε αντιστοιχία με την αρχιτεκτονική της [1].

Στην περίπτωση της σειριακής υλοποίησης, το feature που χρησιμοποιείται στην βασική διεργασία προκύπτει από το τελευταίο και γραμμικό στρώμα του κωδικοποιητή του VAE (δηλαδή το latent layer). Η υπερπαράμετρος latent dimension τέθηκε ίση με 64, καθώς ο αριθμός των νευρώνων από το οποίο προκύπτει το feature κατηγοριοποίησης πρέπει να αντιστοιχίζεται στον αριθμό των κλάσεων του Training Set, (ο οποίος για το CIFAR-FS dataset είναι 64).

Αξίζει να σημειωθεί πως το μοντέλο της παράλληλης υλοποίησης, όπου το Feature εξάγεται από ένα προστιθέμενο παράλληλο γραμμικό στρώμα, είναι αυτό που χρησιμοποιήθηκε και προηγουμένως, με το latent dimension του VAE σε αυτήν την περίπτωση να είναι ίσο με 20. Τα αποτελέσματα αυτής της σύγκρισης συνοψίζονται στον Πίνακα 3.2.3.4.3.1 - 1.

	Test Set		Validation Set	
Υλοποίηση	5-ways - 1-shot	5-ways - 5-shot	5-ways - 1-shot	5-ways - 5-shot
Παράλληλη	**70.024 ± 0.91**	**82.3 ± 0.65**	**62.033 ± 1.05**	75.824 ± 0.77
Σειριακή (όπως στην [1])	68.916 ± 0.92	81.884 ± 0.66	60.616 ± 1.04	**75.896 ± 0.72**

Πίνακας 3.2.3.4.3.1 - 1: Συγκεντρωτικός πίνακας αποτελεσμάτων για σύγκριση μεταξύ παράλληλης και σειριακής υλοποίησης.

Επιλέχθηκε η παράλληλη υλοποίηση, καθώς όπως μπορούμε να δούμε από τον Πίνακα 3.2.3.4.3.1 - 1, υπερισχύει και στις δύο περιπτώσεις του Test Set και ειδικότερα στο 1-shot Task έχει μία αξιοσημείωτη διαφορά. Η μόνη περίπτωση όπου επικράτησε η σειριακή υλοποίηση είναι αυτή του 5-ways – 5-shot του Validation Set, όπου η διαφορά της από την παράλληλη υλοποίηση είναι ελάχιστη.

3.2.3.4.3.2 Ρύθμιση της τιμής της υπερπαραμέτρου latent dimension

Στην συγκεκριμένη παράγραφο παρατίθεται ο τρόπος με τον οποίο επιλέχθηκε ο αριθμός των νευρώνων του ενδιάμεσου στρώματος του VAE που χρησιμοποιήθηκε για την υλοποίηση. Ως συνέχεια της αρχικής συλλογιστικής πορείας, τα χαρακτηριστικά των πειραμάτων αυτής της διερεύνησης, είναι αυτά που μέχρι στιγμής έχουν παρουσιάσει τις καλύτερες επιδόσεις, δηλαδή

- η συνάρτηση MAE ως βοηθητική συνάρτηση κόστους,
- ο συνδυασμός των συντελεστών συνεισφοράς α = 0.25 και β=0.75, για την βασική και την βοηθητική συνάρτηση αντίστοιχα, και
- το feature της κατηγοριοποίησης προκύπτει από έναν παράλληλο γραμμικό κλάδο (nn.Linear) τοποθετημένο ακριβώς πριν από το latent layer του VAE (παράλληλη υλοποίηση).

Για την μελέτη της τιμής της υπερπαραμέτρου latent dimension, τέθηκαν οι εξής περιπτώσεις πειραμάτων:

- ένα πείραμα με latent dimension=256, καθώς η χαμηλή default τιμή 20 νευρώνων για την κωδικοποίηση όλης της πληροφορίας εικόνων μεγέθους 32x32, έθεσε αμφιβολίες για την ποιότητα της ανακατασκευής των εικόνων από τον VAE,
- ένα πείραμα με latent dimension=128 και
- ένα πείραμα με latent dimension=64, προκειμένου να εξεταστεί η επίδοση και σε ενδιάμεσες τιμές αριθμού νευρώνων (μεταξύ του 20 και του 256),
- ένα πείραμα με latent dimension=20, η οποία είναι η default τιμή στην υλοποίηση του Julian Stastny [90], και
- ένα πείραμα με latent dimension=10, καθώς η υλοποίηση του VAE του Stastny [90] έχει πραγματοποιηθεί για ανακατασκευή εικόνων μεγέθους 64x64, ενώ στην δική μας περίπτωση οι εικόνες εισόδου είναι 32x32, αλλά και για να έχουμε και μία εικόνα της επίδρασης αυτού όπου ο αριθμός latent των νευρώνων είναι μικρότερος του 20.

Τα αποτελέσματα για τα παραπάνω πειράματα συνοψίζονται στον Πίνακα 3.2.3.4.3.2 - 1.

	Test Set		Validation Set	
Αριθμός Νευρώνων Latent	5-ways - 1-shot	5-ways - 5-shot	5-ways - 1-shot	5-ways - 5-shot
256	69.862 ± 0.92	82.256 ± 0.65	62.089 ± 1.07 ($3^{η}$ θέση)	75.962 ± 0.77 ($2^{η}$ θέση)
128	70.273 ± 0.92 ($2^{η}$ θέση)	82.318 ± 0.66 ($2^{η}$ θέση)	**62.18 ± 1.02** ($1^{η}$ θέση)	**75.976 ± 0.75** ($1^{η}$ θέση)
64	**70.451 ± 0.91** ($1^{η}$ θέση)	**82.364 ± 0.65** ($1^{η}$ θέση)	62.142 ± 1.04 (2η θέση)	75.581 ± 0.74 ($3^{η}$ θέση)
20	70.024 ± 0.91 ($3^{η}$ θέση)	82.3 ± 0.65 ($3^{η}$ θέση)	62.033 ± 1.05	75.824 ± 0.77
10	69.611 ± 0.92	81.649 ± 0.65	61.233 ± 1.04	75.169 ± 0.76

Πίνακας 3.2.3.4.3.2 - 1: Συγκεντρωτικός πίνακας αποτελεσμάτων για σύγκριση μεταξύ των μοντέλων που χαρακτηρίζονται από διάφορους αριθμούς νευρώνων για το ενδιάμεσο στρώμα του VAE.

Όπως μπορούμε να διακρίνουμε από τον Πίνακα 3.2.3.4.3.2 - 1 τις δύο καλύτερες επιδόσεις επέδειξαν τα μοντέλα τα οποία χαρακτηρίζονται από ενδιάμεσες τιμές 64 και 128 νευρώνων του latent layer. Επιλέχθηκε το μοντέλο με latent dimension ίσο με 64, το οποίο υπερτερεί και στις δύο διεργασίες του Test Set που είναι και το βασικό κριτήριο. Αξίζει να σημειωθεί πως τα μοντέλα με τις ακραίες τιμές νευρώνων 256, και 10 κατέγραψαν τις λιγότερο καλές επιδόσεις, ενώ οι επιδόσεις του μοντέλου με την τιμή νευρώνων που επιλέχθηκε στο σημείο αφετηρίας αυτής της διερεύνησης (latent dimension = 20), κυμάνθηκε στις ενδιάμεσες θέσεις.

3.3 Επίλογος και Μελλοντική Εργασία

3.3.1 Απολογισμός

Στην παραπάνω πειραματική ανάλυση, χρησιμοποιήθηκε ως βάση ο selfsupervised αλγόριθμος Generation-0 της εργασίας [1] των Jathushan Rajasegaran et al. και διενεργήθηκαν τρεις βασικές συγκρίσεις προκειμένου να εξεταστούν οι επιδόσεις συνελικτικών μοντέλων στο πρόβλημα του Few-Shot Image Classification.

Στην 1η σύγκριση εξετάστηκε ποια είναι η επίδραση του βάθους των δικτύων για τον εν λόγω αλγόριθμο και βρέθηκε πως όσο μικρότερο το βάθος των δικτύων τόσο καλύτερα είναι τα πειραματικά αποτελέσματα που αποδίδονται.

Στην 2η σύγκριση, εξετάστηκε ποια είναι η επίδραση του πλάτους των δικτύων (αριθμός των χρησιμοποιούμενων συνελικτικών φίλτρων) για τον εν λόγω αλγόριθμο, και τα πειραματικά αποτελέσματα οδηγήσαν στο συμπέρασμα πως όσο πιο πλατύ είναι το δίκτυο τόσο καλύτερες είναι οι επιδόσεις που αυτό επιδεικνύει. Επιπλέον, ελέγχθηκε και ο τρόπος με τον οποίο επηρεάζει η βαθμονόμηση της συμβολής των καναλιών μέσω της προσθήκης ενός SE block [32], όπου δεν παρατηρήθηκε κάποια ιδιαίτερη επίδραση στις επιδόσεις.

Στην 3η σύγκριση, διερευνήθηκε η υπόθεση για το αν θα επέλθει βελτίωση στο Few-Shot Image Classification Task με την συμπερίληψη μίας διαφορετικής Selfsupervised βοηθητικής διεργασίας και συγκεκριμένα με την χρησιμοποίηση ενός VAE (Variational Autoencoder) [63]. Τα πειραματικά ευρήματα υπέδειξαν ότι αυτός ο τρόπος εκπαίδευσης έχει θετική επίδραση στις επιδόσεις των μοντέλων χωρίς όμως να συνεισφέρει στον ίδιο βαθμό που συνεισφέρει η Selfsupervised διαδικασία του Rotation του αλγορίθμου Generation-0 της [1].

3.3.2 Μελλοντική Ενασχόληση

Ως μελλοντικές κατευθύνσεις της παρούσας πειραματικής ανάλυσης αυτής της διπλωματικής εργασίας μπορούν να ακολουθηθούν διάφορες οδοί.

Στο μοτίβο των δύο πρώτων συγκρίσεων θα μπορούσε η μελέτη των συνελικτικών μοντέλων να εμπλουτιστεί περεταίρω με την τοποθέτηση και την διερεύνηση δικτύων όπως τα VGGNets [39], τα GoogLeNets [40, 48,49, 50], τα ShuffleNets [45, 46], τα MobileNets [42, 43, 44] κι ούτω καθεξής, εξετάζοντας την επίδραση, των διάφορων στοιχείων που διακατέχουν αυτά τα μοντέλα, στις επιδόσεις. Επιπλέον, η μελέτη θα μπορούσε να επεκταθεί και στην διερεύνηση διαφορετικών νευρωνικών δικτύων πέραν των συνελικτικών, όπως είναι οι Μετασχηματιστές (Transformers). Οι Μετασχηματιστές αποτελούν δίκτυα, τα οποία παρότι ξεκίνησαν να εφαρμόζονται σε διεργασίες Επεξεργασίας Φυσικής Γλώσσας (Natural Language

Processing), συνιστούν ένα ταχέως ανερχόμενο είδος δικτύων και σε εφαρμογές Μηχανικής Όρασης (Computer Vision) επιδεικνύοντας ανταγωνιστικότατες επιδόσεις, και συνεπώς θα είχε ενδιαφέρον μία αντίστοιχη διερεύνηση και για αυτό το είδος μοντέλων στο Few-Shot Image Classification Task.

Στο πλαίσιο της 3ης σύγκρισης, συνέχιση αυτής της εργασίας θα μπορούσε να αποτελέσει οποιαδήποτε προσθήκη διαφορετικής βοηθητικής Selfsupervised διεργασίας, ή ακόμη και η υλοποίηση της ίδιας ιδέας με διαφορετικά χαρακτηριστικά και υπερπαραμέτρους όσον αφορά την τοποθέτηση του VAE. Δηλαδή, για την επιλογή και την ρύθμιση των υπερπαραμέτρων της υλοποίησης θα μπορούσε είτε να επιλεχθεί ένα διαφορετικό σημείο αφετηρίας της μελέτης είτε να ακολουθηθεί μία διαφορετική ροή διερεύνησης για τις αντίστοιχες επιλογές, η οποία είναι πιθανό να οδηγήσει σε μία αποδοτικότερη υλοποίηση. Μία πρόταση θα ήταν η πραγματοποίηση της ίδιας διερεύνησης μόνο που ως συνάρτηση βοηθητικού κόστους θα χρησιμοποιούνταν η συνάρτηση MSE (χωρίς Channel Normalization), καθώς η συγκεκριμένη επιλογή αποτέλεσε μία πραγματικά δύσκολη απόφαση για την παρούσα υλοποίηση.

Τέλος, ο κώδικας της [1] παρέχει την ευχέρεια συνδυασμού του εν λόγω αλγορίθμου με μία ή περισσότερες ερευνητικές ιδέες και μπορεί να αποτελέσει την προγραμματιστική βάση πάνω στην οποία μπορεί να χτίσει και να δημιουργήσει κανείς, καθώς πρόκειται για μία όμορφη και προσιτή υλοποίηση. Έτσι, μία άλλη εναλλακτική μελλοντικής ενασχόλησης θα μπορούσε να είναι ο συνδυασμός της εργασίας της [1] με οποιαδήποτε άλλη ερευνητική προσέγγιση - εργασία, με σκοπό την δημιουργία ενός πολύ αποδοτικού μοντέλου για το Few-Shot Image Classification Task. Πρόταση ενός τέτοιου συνδυασμού μπορεί να αποτελέσει η εργασία των Daniel Shalam και Simon Korman με τίτλο “The Self-Optimal-Transport Feature Transform” [89], εφόσον φυσικά η δομή του κώδικα της το επιτρέπει, η οποία αποτελεί μία πολύ πρόσφατη και διακεκριμένη τεχνική σε διάφορες διεργασίες μία εκ των οποίων είναι και το Few-Shot Image Classification.

Βιβλιογραφία

Βιβλία

- Συρακούλης, Γ., & Μπούταλης, Ι. (2010). Υπολογιστική Νοημοσύνη & Εφαρμογές.
- Nielsen, M. A. (2015). Neural networks and deep learning (Vol. 25). San Francisco, CA, USA: Determination press. http://neuralnetworksanddeeplearning.com/
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press.

Αναφορές

[1] Rajasegaran, J., Khan, S., Hayat, M., Khan, F. S., & Shah, M. (2020). Self-supervised knowledge distillation for few-shot learning. arXiv preprint arXiv:2006.09785.

[2] Bertinetto, L., Henriques, J. F., Torr, P. H., & Vedaldi, A. (2018). Meta-learning with differentiable closed-form solvers. arXiv preprint arXiv:1805.08136.

[3] Oreshkin, B., Rodríguez López, P., & Lacoste, A. (2018). Tadam: Task dependent adaptive metric for improved few-shot learning. Advances in neural information processing systems, 31.

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

[5] 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).

[6] Noroozi, M., & Favaro, P. (2016, October). Unsupervised learning of visual representations by solving jigsaw puzzles. In European conference on computer vision (pp. 69-84). Springer, Cham.

[7] Zhang, R., Isola, P., & Efros, A. A. (2016). Colorful image colorization. CoRR abs/1603.08511 (2016). arXiv preprint arXiv:1603.08511.

[8] Noroozi, M., Pirsiavash, H., & Favaro, P. (2017). Representation learning by learning to count. In Proceedings of the IEEE international conference on computer vision (pp. 5898-5906).

[9] Zhai, X., Oliver, A., Kolesnikov, A., & Beyer, L. (2019). S4l: Self-supervised semi-supervised learning. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 1476-1485).

[10] Dosovitskiy, A., Springenberg, J. T., Riedmiller, M., & Brox, T. (2014). Discriminative unsupervised feature learning with convolutional neural networks. Advances in neural information processing systems, 27.

[11] Caron, M., Bojanowski, P., Joulin, A., & Douze, M. (2018). Deep clustering for unsupervised learning of visual features. In Proceedings of the European conference on computer vision (ECCV) (pp. 132-149).

[12] 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.

[13] Koch, G., Zemel, R., & Salakhutdinov, R. (2015, July). Siamese neural networks for one-shot image recognition. In ICML deep learning workshop (Vol. 2, p. 0).

[14] Sung, F., Yang, Y., Zhang, L., Xiang, T., Torr, P. H., & Hospedales, T. M. (2018). Learning to compare: Relation network for few-shot learning. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1199-1208).

[15] Vinyals, O., Blundell, C., Lillicrap, T., & Wierstra, D. (2016). Matching networks for one shot learning. Advances in neural information processing systems, 29.

[16] Snell, J., Swersky, K., & Zemel, R. (2017). Prototypical networks for few-shot learning. Advances in neural information processing systems, 30.

[17] Finn, C., Abbeel, P., & Levine, S. (2017, July). Model-agnostic meta-learning for fast adaptation of deep networks. In International conference on machine learning (pp. 1126-1135). PMLR.

[18] Li, Z., Zhou, F., Chen, F., & Li, H. (2017). Meta-sgd: Learning to learn quickly for few-shot learning. arXiv preprint arXiv:1707.09835."

[19] Flennerhag, S., Rusu, A. A., Pascanu, R., Visin, F., Yin, H., & Hadsell, R. (2019). Meta-learning with warped gradient descent. arXiv preprint arXiv:1909.00025.

[20] Rusu, A. A., Rao, D., Sygnowski, J., Vinyals, O., Pascanu, R., Osindero, S., & Hadsell, R. (2018). Meta-learning with latent embedding optimization. arXiv preprint arXiv:1807.05960.

[21] Lee, K., Maji, S., Ravichandran, A., & Soatto, S. (2019). Meta-learning with differentiable convex optimization. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 10657-10665).

[22] Dhillon, G. S., Chaudhari, P., Ravichandran, A., & Soatto, S. (2019). A baseline for few-shot image classification. arXiv preprint arXiv:1909.02729.

[23] Tian, Y., Wang, Y., Krishnan, D., Tenenbaum, J. B., & Isola, P. (2020, August). Rethinking few-shot image classification: a good embedding is all you need?. In European Conference on Computer Vision (pp. 266-282). Springer, Cham.

[24] Raghu, A., Raghu, M., Bengio, S., & Vinyals, O. (2019). Rapid learning or feature reuse? towards understanding the effectiveness of maml. arXiv preprint arXiv:1909.09157.

[25] Chen, D., Chen, Y., Li, Y., Mao, F., He, Y., & Xue, H. (2021, June). Self-supervised learning for few-shot image classification. In ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 1745-1749). IEEE.

[26] Zbontar, J., Jing, L., Misra, I., LeCun, Y., & Deny, S. (2021, July). Barlow twins: Self-supervised learning via redundancy reduction. In International Conference on Machine Learning (pp. 12310-12320). PMLR.

[27] Braham, N. A. A., Mou, L., Chanussot, J., Mairal, J., & Zhu, X. X. (2022). Self Supervised Learning for Few Shot Hyperspectral Image Classification. arXiv preprint arXiv:2206.12117.

[28] An, Y., Xue, H., Zhao, X., & Zhang, L. (2021). Conditional Self-Supervised Learning for Few-Shot Classification. In IJCAI (pp. 2140-2146).

[29] Gidaris, S., Bursuc, A., Komodakis, N., Pérez, P., & Cord, M. (2019). Boosting few-shot visual learning with self-supervision. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 8059-8068).

[30] Mishra, N., Rohaninejad, M., Chen, X., & Abbeel, P. (2017). A simple neural attentive meta-learner. arXiv preprint arXiv:1707.03141.

[31] 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).

[32] Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7132-7141).

[33] Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., ... & Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. nature, 529(7587), 484-489.

[34] Lake, B. M., Salakhutdinov, R., & Tenenbaum, J. B. (2015). Human-level concept learning through probabilistic program induction. Science, 350(6266), 1332-1338.

[35] Altae-Tran, H., Ramsundar, B., Pappu, A. S., & Pande, V. (2017). Low data drug discovery with one-shot learning. ACS central science, 3(4), 283-293.

[36] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.

[37] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. Communications of the ACM, 60(6), 84-90.

[38] Zeiler, M. D., & Fergus, R. (2014, September). Visualizing and understanding convolutional networks. In European conference on computer vision (pp. 818-833). Springer, Cham.

[39] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[40] 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).

[41] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2020). Generative adversarial networks. Communications of the ACM, 63(11), 139-144.

[42] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.

[43] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4510-4520).

[44] Howard, A., Sandler, M., Chu, G., Chen, L. C., Chen, B., Tan, M., ... & Adam, H. (2019). Searching for mobilenetv3. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 1314-1324).

[45] Zhang, X., Zhou, X., Lin, M., & Sun, J. (2018). Shufflenet: An extremely efficient convolutional neural network for mobile devices. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 6848-6856).

[46] Ma, N., Zhang, X., Zheng, H. T., & Sun, J. (2018). Shufflenet v2: Practical guidelines for efficient cnn architecture design. In Proceedings of the European conference on computer vision (ECCV) (pp. 116-131).

[47] Han, K., Wang, Y., Tian, Q., Guo, J., Xu, C., & Xu, C. (2020). Ghostnet: More features from cheap operations. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 1580-1589).

[48] Ioffe, S., & Szegedy, C. (2015, June). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In International conference on machine learning (pp. 448-456). PMLR.

[49] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2818-2826).

[50] Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017, February). Inception-v4, inception-resnet and the impact of residual connections on learning. In Thirty-first AAAI conference on artificial intelligence.

[51] Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2021). A survey of convolutional neural networks: analysis, applications, and prospects. IEEE transactions on neural networks and learning systems.

[52] Albelwi, S. (2022). Survey on Self-Supervised Learning: Auxiliary Pretext Tasks and Contrastive Learning Methods in Imaging. Entropy, 24(4), 551.

[53] Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., ... & Farhan, L. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. Journal of big Data, 8(1), 1-74.

[54] Mitchell, T. M., & Mitchell, T. M. (1997). Machine learning (Vol. 1, No. 9). New York: McGraw-hill.

[55] Wang, Y., Yao, Q., Kwok, J. T., & Ni, L. M. (2020). Generalizing from a few examples: A survey on few-shot learning. ACM computing surveys (csur), 53(3), 1-34.

[56] Song, Y., Wang, T., Mondal, S. K., & Sahoo, J. P. (2022). A Comprehensive Survey of Few-shot Learning: Evolution, Applications, Challenges, and Opportunities. arXiv preprint arXiv:2205.06743.

[57] Yang, J., Guo, X., Li, Y., Marinello, F., Ercisli, S., & Zhang, Z. (2022). A survey of few-shot learning in smart agriculture: developments, applications, and challenges. Plant Methods, 18(1), 1-12.

[58] Guo, Y., Codella, N. C., Karlinsky, L., Codella, J. V., Smith, J. R., Saenko, K., ... & Feris, R. (2020, August). A broader study of cross-domain few-shot learning. In European conference on computer vision (pp. 124-141). Springer, Cham.

[59] Fink, M. (2004). Object classification from a single example utilizing class relevance metrics. Advances in neural information processing systems, 17.

[60] Tang, K. D., Tappen, M. F., Sukthankar, R., & Lampert, C. H. (2010, June). Optimizing one-shot recognition with micro-set learning. In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (pp. 3027-3034). IEEE.

[61] Shyam, P., Gupta, S., & Dukkipati, A. (2017, July). Attentive recurrent comparators. In International conference on machine learning (pp. 3173-3181). PMLR.

[62] Zhang, C., Bütepage, J., Kjellström, H., & Mandt, S. (2018). Advances in variational inference. IEEE transactions on pattern analysis and machine intelligence, 41(8), 2008-2026.

[63] Kingma, D. P., & Welling, M. (2014, April). Stochastic gradient VB and the variational auto-encoder. In Second International Conference on Learning Representations, ICLR (Vol. 19, p. 121).

[64] Van den Oord, A., Kalchbrenner, N., Espeholt, L., Vinyals, O., & Graves, A. (2016). Conditional image generation with pixelcnn decoders. Advances in neural information processing systems, 29.

[65]] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2020). Generative adversarial networks. Communications of the ACM, 63(11), 139-144.

[66] Ravi, S., & Larochelle, H. (2016). Optimization as a model for few-shot learning.

[67] Andrychowicz, M., Denil, M., Gomez, S., Hoffman, M. W., Pfau, D., Schaul, T., ... & De Freitas, N. (2016). Learning to learn by gradient descent by gradient descent. Advances in neural information processing systems, 29.

[68] Wu, Y., & Demiris, Y. (2010, May). Towards one shot learning by imitation for humanoid robots. In 2010 IEEE International Conference on Robotics and Automation (pp. 2889-2894). IEEE.

[69] Abdo, N., Kretzschmar, H., Spinello, L., & Stachniss, C. (2013, May). Learning manipulation actions from a few demonstrations. In 2013 IEEE International Conference on Robotics and Automation (pp. 1268-1275). IEEE.

[70] Hamaya, M., Matsubara, T., Noda, T., Teramae, T., & Morimoto, J. (2016, May). Learning assistive strategies from a few user-robot interactions: Model-based reinforcement learning approach. In 2016 IEEE International Conference on Robotics and Automation (ICRA) (pp. 3346-3351). IEEE.

[71] Han, X., Zhu, H., Yu, P., Wang, Z., Yao, Y., Liu, Z., & Sun, M. (2018). Fewrel: A large-scale supervised few-shot relation classification dataset with state-of-the-art evaluation. arXiv preprint arXiv:1810.10147.

[72] Lake, B., Lee, C. Y., Glass, J., & Tenenbaum, J. (2014). One-shot learning of generative speech concepts. In Proceedings of the Annual Meeting of the Cognitive Science Society (Vol. 36, No. 36).

[73] Arik, S., Chen, J., Peng, K., Ping, W., & Zhou, Y. (2018). Neural voice cloning with a few samples. Advances in neural information processing systems, 31.

[74] Tjandra, A., Sakti, S., & Nakamura, S. (2018). Machine speech chain with one-shot speaker adaptation. arXiv preprint arXiv:1803.10525.

[75] Mohammadi, S. H., & Kim, T. (2018). Investigation of using disentangled and interpretable representations for one-shot cross-lingual voice conversion. arXiv preprint arXiv:1808.05294.

[76] Santoro, A., Bartunov, S., Botvinick, M., Wierstra, D., & Lillicrap, T. (2016, June). Meta-learning with memory-augmented neural networks. In International conference on machine learning (pp. 1842-1850). PMLR.

[77] Finn, C., Xu, K., & Levine, S. (2018). Probabilistic model-agnostic meta-learning. Advances in neural information processing systems, 31.

[78] Grant, E., Finn, C., Levine, S., Darrell, T., & Griffiths, T. (2018). Recasting gradient-based meta-learning as hierarchical bayes. arXiv preprint arXiv:1801.08930.

[79] Yoon, J., Kim, T., Dia, O., Kim, S., Bengio, Y., & Ahn, S. (2018). Bayesian model-agnostic meta-learning. Advances in neural information processing systems, 31.

[80] Ramalho, T., & Garnelo, M. (2019). Adaptive posterior learning: few-shot learning with a surprise-based memory module. arXiv preprint arXiv:1902.02527.

[81] Brock, A., Lim, T., Ritchie, J. M., & Weston, N. (2017). Smash: one-shot model architecture search through hypernetworks. arXiv preprint arXiv:1708.05344.

[82] Liu, H., Simonyan, K., & Yang, Y. (2018). Darts: Differentiable architecture search. arXiv preprint arXiv:1806.09055.

[83] Yao, Q., Xu, J., Tu, W. W., & Zhu, Z. (2020, April). Efficient neural architecture search via proximal iterations. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34, No. 04, pp. 6664-6671).

[84] Zhao, Y., Wang, L., Tian, Y., Fonseca, R., & Guo, T. (2021, July). Few-shot neural architecture search. In International Conference on Machine Learning (pp. 12707-12718). PMLR.

[85] Baldi, P. (2012, June). Autoencoders, unsupervised learning, and deep architectures. In Proceedings of ICML workshop on unsupervised and transfer learning (pp. 37-49). JMLR Workshop and Conference Proceedings.

[86] Bank, D., Koenigstein, N., & Giryes, R. (2020). Autoencoders. arXiv preprint arXiv:2003.05991.

[87] Lample, G., & Conneau, A. (2019). Cross-lingual language model pretraining. *arXiv preprint arXiv:1901.07291*.

[88] Azizi, S., Mustafa, B., Ryan, F., Beaver, Z., Freyberg, J., Deaton, J., ... & Norouzi, M. (2021). Big self-supervised models advance medical image classification. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 3478-3488).

[89] Shalam, D., & Korman, S. (2022). The Self-Optimal-Transport Feature Transform. arXiv preprint arXiv:2204.03065.

[90] Stastny (2019) VAE-ResNet18-PyTorch [Source Code]. https://github.com/julianstastny/VAE-ResNet18-PyTorch.

[91] Hariharan, B., Arbeláez, P., Girshick, R., & Malik, J. (2014, September). Simultaneous detection and segmentation. In European conference on computer vision (pp. 297-312). Springer, Cham.

[92] Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence, 39(12), 2481-2495.

Links

https://www.potentiaco.com/what-is-machine-learning-definition-types-applications-and-examples/

https://neptune.ai/blog/self-supervised-learning

https://cs231n.github.io/

https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

https://www.ibm.com/cloud/learn/convolutional-neural-networks

https://www.jeremyjordan.me/batch-normalization/

https://towardsdatascience.com/batch-normalization-and-dropout-in-neural-networks-explained-with-pytorch-47d7a8459bcd

https://www.jeremyjordan.me/variational-autoencoders/

https://towardsdatascience.com/understanding-variational-autoencoders-vaes-f70510919f73

https://www.borealisai.com/research-blogs/tutorial-2-few-shot-learning-and-meta-learning-i/