



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

Βελτίωση μεθόδων μεταφοράς στυλ με επαύξηση δεδομένων για γενίκευση πεδίου

Μελέτη και υλοποίηση

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΤΟΥ

ΝΙΚΟΛΑΟΥ Α. ΣΠΑΝΟΥ

Επιβλέπων: Στέφανος Κόλλιας
Καθηγητής Ε.Μ.Π.

Αθήνα, Οκτώβριος 2023



ΕΘΝΙΚΟ ΜΕΤΕΩΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

Βελτίωση μεθόδων μεταφοράς στυλ με επαύξηση δεδομένων για γενίκευση πεδίου

Μελέτη και υλοποίηση

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΤΟΥ

ΝΙΚΟΛΑΟΥ Α. ΣΠΑΝΟΥ

Επιβλέπων: Στέφανος Κόλλιας
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 26ης Οκτωβρίου 2023.

(Υπογραφή)

(Υπογραφή)

(Υπογραφή)

.....
Στέφανος Κόλλιας
Καθηγητής Ε.Μ.Π.

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

.....
Αθανάσιος Βουλόδημος
Επικουρος Καθηγητής Ε.Μ.Π.

Αθήνα, Οκτώβριος 2023

Copyright © – All rights reserved. Με την επιφύλαξη παντός δικαιώματος.
Νικόλαος Σπανός, 2023.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα.

Το περιεχόμενο αυτής της εργασίας δεν απηχεί απαραίτητα τις απόψεις του Τμήματος, του Επιβλέποντα, ή της επιτροπής που την ενέκρινε.

ΔΗΛΩΣΗ ΜΗ ΛΟΓΟΚΛΟΠΗΣ ΚΑΙ ΑΝΑΛΗΨΗΣ ΠΡΟΣΩΠΙΚΗΣ ΕΥΘΥΝΗΣ

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

(Υπογραφή)

.....
Νικόλαος Σπανός

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

26ης Οκτωβρίου 2023

Περίληψη

Τα συστήματα βαθιάς μάθησης έχουν οδηγήσει σε τεράστιες βελτιώσεις στον τομέα της υπολογιστικής όρασης. Όμως, σε τομείς όπως ο ιατρικός, η έλλειψη οργανωμένων συνόλων δεδομένων έχει οδηγήσει σε μεγάλα προβλήματα απόδοσης. Σε συνδυασμό με αυτήν την έλλειψη, έχει παρατηρηθεί ότι η απόδοση συστημάτων βαθιάς μάθησης σε πραγματικές εφαρμογές είναι μειωμένη, λόγω του γεγονότος ότι δεν ανταποδίδουν επαρκώς σε δεδομένα που προέρχονται από άγνωστα πεδία. Αυτό συμβαίνει διότι κατά την εκπαίδευση γίνεται παραδοχή ότι τα δεδομένα εκπαίδευσης και ελέγχου βρίσκονται στο ίδιο πεδίο, ενώ σε πραγματικές εφαρμογές, όπως σε κατάτμηση ιατρικών εικόνων, η παρουσία κρυφών μεταβλητών μεταβάλλει τις εικόνες σε μεγάλο βαθμό. Ακόμα και εάν παρουσιάζουν το ίδιο σημασιολογικό περιεχόμενο, διαφορές στον τρόπο ή στο όργανο λήψης της εικόνας μπορούν να οδηγήσουν σε λανθασμένα αποτελέσματα. Μεγάλο μέρος της έρευνας αιχμής τα τελευταία χρόνια στον τομέα της βαθιάς μάθησης εστιάζει στην δημιουργία μεθόδων γενίκευσης πεδίου, με τις οποίες από λιγοστά δεδομένα μπορούν τα δίκτυα βαθιάς μάθησης να γενικεύσουν σε άγνωστα πεδία. Ο απώτερος σκοπός τέτοιων μεθόδων είναι η επίλυση του προβλήματος γενίκευσης και η δημιουργία πρακτικών και εύχρηστων εργαλείων, που θα βρουν εφαρμογή σε πολλούς τομείς της καθημερινότητας. Στόχος της διπλωματικής εργασίας είναι η διερεύνηση σύγχρονων μεθόδων γενίκευσης πεδίου και η ανάπτυξη ενός βελτιωμένου συστήματος, το οποίο εκμεταλλεύεται περίπλοκες μεθόδους επαύξησης δεδομένων σε συνδυασμό με τεχνικές μεταφοράς στυλ και ανταγωνιστικής μάθησης και η εφαρμογή του σε ιατρικές εικόνες.

Λέξεις Κλειδιά

Βαθιά μάθηση, γενίκευση πεδίου, μεταφορά στυλ, κατάτμηση εικόνας, ανταγωνιστική επαύξηση δεδομένων, επαύξηση δεδομένων στο επίπεδο εισόδου, επαύξηση δεδομένων στο επίπεδο χαρακτηριστικών

Abstract

Deep learning systems have led to huge improvements in computer vision. However, in areas such as the medical area, the lack of organised datasets has led to major performance issues for these systems. Coupled with this lack of data, it has been observed that deep learning systems do not respond adequately to data from unknown domains. This is because during training it is assumed that the training and control data are in the same domain, whereas in realistic applications, such as in medical image segmentation, the presence of latent variables alters the images in a large degree. Even if they present the same semantic content, differences in the way or instrument of image acquisition can lead to erroneous results. Much of the cutting edge research in recent years in the field of deep learning focuses on the creation of domain generalization methods, whereby from sparse data deep learning networks can generalize to unknown domains. The ultimate goal of such methods is to solve the generalization problem and create practical and easy-to-use tools, that can find practical uses in every-day life. The aim of this thesis is to investigate state-of-the-art methods for domain generalization and develop an improved system that exploits sophisticated data augmentation methods combined with style transfer and competitive learning techniques and apply it to medical images.

Keywords

Deep learning, domain generalisation, style transfer, image segmentation, adversarial data augmentation, input level augmentation, feature-space augmentation

στην οικογένεια μου

Ευχαριστίες

Αρχικά, θέλω να ευχαριστήσω με όλη μου την καρδιά την οικογένεια μου που τόσα χρόνια στέκεται δίπλα μου σε όλες τις δυσκολίες που έχω συναντήσει. Θέλω να ευχαριστήσω, επίσης, όλους τους φίλους μου, χωρίς την βοήθεια των οποίων δεν θα βρισκόμουν εδώ που είμαι και δεν θα είχα καταφέρει να ξεπεράσω τις αντιξοότητες του ταξιδιού. Τέλος, θέλω να ευχαριστήσω τους καθηγητές μου, κ. Γιώργο Στάμου, κ. Αθανάσιο Βουλόδημο και κ. Στέφανο Κόλλια και τους υπεύθυνους της διπλωματικής μου εργασίας, την κ. Παρασκευή Τζούβελη, τον Αναστάση Αρσένο και την Παρασκευή Θεοφίλου για την απίστευτη βοήθεια που μου προσέφεραν καθ'όλη την διάρκεια της συνεργασίας μας. Σας ευχαριστώ όλους.

Αθήνα, Οκτώβριος 2023

Νικόλαος Σπανός

Περιεχόμενα

Περίληψη	1
Abstract	3
Ευχαριστίες	7
1 Εισαγωγή	15
1.1 Αντικείμενο της διπλωματικής	16
1.2 Οργάνωση του τόμου	16
I Θεωρητικό Μέρος	17
2 Θεωρητικό υπόβαθρο	19
2.1 Κατάτμηση Εικόνων	19
2.1.1 Είδη Κατάτμησης Εικόνας	20
2.1.2 Το OOD πρόβλημα	21
2.2 Γενίκευση Πεδίου	22
2.3 Μεταφορά Στυλ	23
2.4 Μοντέλα βαθιάς μάθησης - Νευρωνικά Δίκτυα	24
2.4.1 CNN-Συνελκτικά Νευρωνικά Δίκτυα	24
2.4.2 FCN-Πλήρως Συνελκτικά Δίκτυα	28
2.4.3 Μοντέλα Encoder-Decoder	29
2.4.4 Μοντέλα FTN-STN	30
3 Σχετική βιβλιογραφία μεθόδων γενίκευσης πεδίου σε εικόνες	33
3.1 Σχετικές εργασίες	33
3.1.1 Προσαρμογή Πεδίου	33
3.1.2 Ανταγωνιστική Εκπαίδευση	34
3.1.3 Μεταφορά Στυλ	35
3.1.4 Επαύξηση στο επίπεδο των χαρακτηριστικών	35
3.1.5 Εφαρμογές Βαθιάς Μάθησης	36
II Πρακτικό Μέρος	37
4 Σύνολα Δεδομένων	39
4.1 Δεδομένα Προστάτη	39

4.2	Δεδομένα Καρδιάς	40
4.3	Προεπεξεργασία Δεδομένων	42
5	Υλοποίηση	47
5.1	Επαύξηση Δεδομένων	47
5.1.1	Μέθοδοι επαύξησης σε επίπεδο εισόδου	47
5.1.2	Μέθοδοι επαύξησης σε επίπεδο χαρακτηριστικών	49
6	Παρουσίαση και Σχολιασμός Πειραματικών Αποτελεσμάτων	55
6.1	Μετρικές Αξιολόγησης	55
6.2	Μεθοδολογία Ελέγχου	55
6.3	Αποτελέσματα στα δεδομένα προσάτη	56
6.4	Αποτελέσματα στα δεδομένα καρδιάς	56
III	Επίλογος	59
7	Επίλογος	61
7.1	Συμπεράσματα	61
7.2	Μελλοντικές Επεκτάσεις	61
	Παραρτήματα	63
A'	Διαγράμματα Μοντέλων	65
A'.1	Βασική Αρχιτεκτονική	65
A'.2	MixStyle	65
A'.3	DSU	66
A'.4	MaxStyle	66
A'.5	MaxStyle /w Pixmix	66
	Βιβλιογραφία	78
	Συνομογραφίες - Αρκτικόλεξα - Ακρωνύμια	79
	Απόδοση ξενόγλωσσων όρων	81

Κατάλογος Σχημάτων

2.1	Κατάτμηση εικόνας σε ένα σύνθετο αστικό σενάριο για την αυτόνομη οδήγηση [1].	19
2.2	Παράδειγμα Κατάτμησης Περιπτώσεων [2]	20
2.3	Παράδειγμα Σηματολογικής Κατάτμησης [2]	21
2.4	Δύο παραδείγματα που δείχνουν ότι εκπαιδευμένα Unet μοντέλα είναι ευάλωτα σε προσεκτικά υπολογισμένες διαταραχές που προστίθεται στην αρχική εικόνα [3].	22
2.5	Διαδικασία μεταφοράς στυλ από πίνακα ζωγραφικής σε εικόνες του πραγματικού κόσμου [4]	24
2.6	Παράδειγμα της διαδικασίας ενός δικτύου CNN [5]	25
2.7	Παράδειγμα του αποτελέσματος συνέλιξης ενός φίλτρου [5]	26
2.8	Παράδειγμα της διαδικασίας pooling [5]	27
2.9	Αρχιτεκτονικές VGG, ResNet [6]	27
2.10	Παράδειγμα διαδικασίας ενός δικτύου FCN. Διαθέτουν την δυνατότητα να μάθουν αποτελεσματικά να κάνουν πυκνές προβλέψεις για εργασίες ανά εικονοστοιχείο [7]	28
2.11	Αρχιτεκτονική U-Net (παράδειγμα για 32x32 εικονοστοιχεία στη χαμηλότερη ανάλυση). Κάθε μπλε πλαίσιο αντιστοιχεί σε χάρτη χαρακτηριστικών πολλαπλών καναλιών. Ο αριθμός των καναλιών συμβολίζεται στην κορυφή του πλαισίου. Το μέγεθος x-y παρέχεται στο κάτω αριστερό άκρο του πλαισίου. Λευκό κουτιά αντιπροσωπεύουν αντιγραμμένους χάρτες χαρακτηριστικών. Τα βέλη υποδηλώνουν τις διάφορες λειτουργίες [8]	29
2.12	Μια απεικόνιση της αρχιτεκτονικής FCN με βάση τον κωδικοποιητή-αποκωδικοποιητή για σημασιολογική κατάτμηση [9]	30
2.13	Το δίκτυο FTN-STN [10]	31
3.1	Παραδείγματα εικόνων από τρία σημεία αναφοράς γενίκευσης πεδίου που εκδηλώνουν διαφορετικούς τύπους μετατόπισης πεδίου. Στο (α), η μετατόπιση τομέα αντιστοιχεί κυρίως σε αλλαγές στο στυλ της γραμματοσειράς, το χρώμα και το φόντο. Στο (β), είναι σαφείς οι μεροληψίες που αφορούν συγκεκριμένα σύνολα δεδομένων, οι οποίες οφείλονται σε αλλαγές σε περιβάλλον/σκηνικό και την οπτική γωνία. Στο (γ), οι αλλαγές στο στυλ της εικόνας είναι ο κύριος λόγος για τη μετατόπιση τομέα [11].	33
3.2	Παρουσίαση του αποτελέσματος πρόσθεσης θορύβου στην πρόβλεψη ενός δικτύου ταξινόμησης εικόνας [12]	34

3.3	Παράδειγμα δημιουργίας δειγμάτων για προσαρμογή πεδίου με ρυθμιζόμενες υπερπαραμέτρους [13]	35
3.4	Παράδειγμα εφαρμογής μεθόδου Mixup [14]	36
4.1	Σύνολο Εικόνων των συνόλων δεδομένων	40
4.2	Εικόνα από το σύνολο δεδομένων	41
4.3	Κατάτμηση Εικόνας	41
4.4	Θορυβοποιήμενες εκδοχές εικόνας	42
4.5	Απεικόνιση του μη-ομογενούς πεδίου στις εικόνες μαγνητικής τομογραφίας. Η μορφή του είναι παρόμοια με ενός ανομοιόμορφου επιπέδου πάνω στην εικόνα [15]	44
4.6	Η πρώτη στήλη δίνει διαστάσεις σαγματικές όψεις των μεταθανάτιων ιπποκάμπων σε τρία άτομα ("r" για τον δεξιό ιππόκαμπο, "l" για τον αριστερό ιππόκαμπο). Η δεύτερη και η τρίτη στήλη δίνουν τη διορθωμένη εικόνα με τη χρήση του N3MNI και το αντίστοιχο πεδίο μεροληψίας, ενώ τα αποτελέσματα με τη χρήση του N4ITK παρέχονται στις στήλες 4 και 5 [15].	45
5.1	Παραδείγματα τυπικών μεθόδων επαύξησης δεδομένων στο επίπεδο της εισόδου [16], [17], [18], [19]	48
5.2	Η διαδικασία της επαύξησης δεδομένων PixMix. [20]	48
5.3	Η μορφή της ανάμειξης εικόνων μέσα σε μία παρτίδα. Ακόμα και εικόνες της ίδιας παρτίδας μπορούν να οδηγήσουν σε καινούργια στυλ με αρκετές επαναλήψεις [21]	50
5.4	Η διαδικασία επαύξησης με χρήση μοντελοποίησης των στατιστικών ως πολυδιάστατες Γκαουσιανές κατανομές [22]	51
5.5	Αριστερά: Η αρχιτεκτονική MaxStyle. Ο υποβοηθητικός αποκωδικοποιητής αλλάζει τις παραμέτρους MixStyle στο εσωτερικό της κύριας αρχιτεκτονικής, οδηγώντας σε καινούργια στυλ. Δεξιά: Εικόνα Εισόδου και Επαυξημένη Εικόνα. Από την συσχέτιση των παραγώγων παρατηρούμε την διατήρηση του σημασιολογικού περιεχομένου [23].	53
A.1	Βασική αρχιτεκτονική κωδικοποιητή-αποκωδικοποιητή	65
A.2	Αρχιτεκτονική με εισαγωγή MixStyle blocks στα αρχικά επίπεδα για επαύξηση των χαρτών χαρακτηριστικών	65
A.3	Αρχιτεκτονική με εισαγωγή DSU blocks στα αρχικά επίπεδα για επαύξηση των χαρτών χαρακτηριστικών	66
A.4	Αρχιτεκτονική με εισαγωγή MaxStyle blocks στα αρχικά επίπεδα για επαύξηση των χαρτών χαρακτηριστικών. Ο βοηθητικός αποκωδικοποιητής και τα σφάλματα φαίνονται από κάτω.	66
A.5	Αρχιτεκτονική με εισαγωγή MaxStyle blocks στα αρχικά επίπεδα για επαύξηση των χαρτών χαρακτηριστικών και επαύξηση δεδομένων στο επίπεδο εισόδου PixMix	67

Κατάλογος Πινάκων

6.1	Πίνακας σύγκρισης μεθόδων για adversarial network στα δεδομένα προσάτη σε αρχιτεκτονική FCN-16	56
6.2	Πίνακας σύγκρισης μεθόδων για adversarial network στα δεδομένα προσάτη σε αρχιτεκτονική FCN-64	56
6.3	Πίνακας σύγκρισης μεθόδων για adversarial network στα δεδομένα καρδιάς σε αρχιτεκτονική FCN-16 για την αριστερή κοιλία	57
6.4	Πίνακας σύγκρισης μεθόδων για adversarial network στα δεδομένα καρδιάς σε αρχιτεκτονική FCN-16 για το μυοκάρδιο	57
6.5	Πίνακας σύγκρισης μεθόδων για adversarial network στα δεδομένα καρδιάς σε αρχιτεκτονική FCN-16 για την δεξιά κοιλία	57

Κεφάλαιο **1**

Εισαγωγή

Η τεχνητή νοημοσύνη έχει γνωρίσει μια αξιοσημείωτη άνοδο τα τελευταία χρόνια, φέρνοντας επανάσταση σε πολλές πτυχές της επιστήμης. Αυτή η άνοδος μπορεί να αποδοθεί στις εξελίξεις στη μηχανική μάθηση, στην αυξημένη υπολογιστική ισχύ και στην αφθονία των δεδομένων που είναι διαθέσιμα για την εκπαίδευση των συστημάτων τεχνητής νοημοσύνης. Χάρη σε αυτές τις προόδους, καινούργια συστήματα, επανομαζόμενα συστήματα βαθιάς μάθησης, έχουν λάβει κύρια θέση στις εφαρμογές τεχνητής νοημοσύνης.

Η βαθιά μάθηση, δηλαδή μεγάλα μοντέλα τεχνητής νοημοσύνης που αναλύουν δεδομένα με μεγάλη λεπτομέρεια, έχει αλλάξει ριζικά τον τομέα της υπολογιστικής όρασης τα τελευταία χρόνια, έχοντας εκτελέσει τεράστια πρόοδο στην επεξεργασία εικόνων. Όμως, τα σύγχρονα μοντέλα που παρουσιάζουν συνήθως βέλτιστη απόδοση είτε κάνουν ορισμένες παραδοχές στην εκπαίδευση που δεν συμβαδίζουν με ρεαλιστικές συνθήκες εφαρμογής τέτοιων συστημάτων είτε έχουν εκπαιδευτεί σε μεγάλο όγκο δεδομένων, γεγονός που δεν είναι πάντα εφικτό.

Ένα από τα πιο βασικά θέματα στην εκπαίδευση εργαλείων ικανά για εφαρμογή σε διάφορους τομείς είναι η γενίκευση των μοντέλων. Τα συστήματα αυτά πρέπει να είναι ικανά να αντιστέκονται σε τυχόν αλλοιώσεις των δεδομένων, τις οποίες πιθανόν να μην έχουν δει κατά τη διάρκεια της εκπαίδευσης. Τέτοιες αλλοιώσεις μπορεί να προέρχονται από παράγοντες οι οποίοι δεν είναι ελέγξιμοι, όπως για παράδειγμα η αλλαγή καιρού σε εφαρμογές αυτόνομης οδήγησης. Εάν τέτοιες ακραίες περιπτώσεις δεν έχουν ληφθεί υπόψη κατά την διάρκεια της εκπαίδευσης, τότε μπορούν να οδηγήσουν σε σοβαρή πτώση της απόδοσης κατά την εφαρμογή.

Η έλλειψη δεδομένων είναι επίσης ένα σοβαρό πρόβλημα για την εφαρμογή μοντέλων μηχανικής μάθησης. Το παραπάνω πρόβλημα που αναφέρθηκε συνήθως επιλύεται με την χρήση μεγάλου όγκου ποικίλων δεδομένων εκπαίδευσης. Για πολλούς τομείς, όπως ο ιατρικός, όμως, η συλλογή δεδομένων από πολλαπλές πηγές μπορεί να είναι κοστοβόρα ή και αδύνατη, καθώς δεν υπάρχει πληθώρα ελεύθερων προς χρήση δεδομένων.

Η ανάγκη για πιο ανεπτυγμένες μεθόδους γενίκευσης μοντέλων είναι μεγάλη και η σύγχρονη επιστήμη έχει βρει τρόπους με τους οποίους είναι δυνατόν συστήματα τεχνητής νοημοσύνης να γενικεύουν σε άγνωστα δεδομένα, χωρίς να τα έχουν δει κατά την εκπαίδευση. Η μεταφορά στυλ και η ανταγωνιστική μάθηση είναι δύο από αυτές τις μεθόδους και έχουν οδηγήσει σε μεγάλη βελτίωση. Όμως, υπάρχει ανάγκη για περαιτέρω βελτίωση, καθώς η απόδοση είναι ακόμα χαμηλή και το κόστος και ο χρόνος εκπαίδευσης αρκετά μεγάλος.

Σε τέτοιες περιπτώσεις γίνεται χρήση μεθόδων επαύξησης δεδομένων, μέθοδοι, δηλαδή, που μεταβάλλουν το περιεχόμενο της εικόνας με σκοπό την αύξηση στιβαρότητας του μοντέλου.

Χρησιμοποιώντας, λοιπόν, μεθόδους επαύξησης δεδομένων μπορεί κάποιος να βελτιώσει την απόδοση, χωρίς κόστος στον χρόνο και στους πόρους εκπαίδευσης. Περίπλοκες επαυξήσεις, όμως, αποφεύγονται να χρησιμοποιηθούν στον τομέα της υγείας, λόγω της επίδρασης που έχουν στο σημασιολογικό περιεχόμενο των εικόνων. Σε πολλές έρευνες, όμως, έχει αποδειχθεί ότι η εισαγωγή μεγάλης πολυπλοκότητας κατά την εκπαίδευση μπορεί να μας οδηγήσει σε σημαντικά καλύτερα αποτελέσματα γενίκευσης. Οπότε, είναι ενδιαφέρον να διερευνήσουμε το πως τέτοιες περίπλοκες επαυξήσεις, σε συνδυασμό, με σύγχρονες μεθόδους γενίκευσης πεδίου μπορούν να μας οδηγήσουν σε βελτιωμένη απόδοση, χωρίς αύξηση των υπολογιστικών πόρων.

1.1 Αντικείμενο της διπλωματικής

Αντικείμενο της διπλωματικής είναι η διερεύνηση χρήσης περίπλοκων επαυξήσεων δεδομένων σε ιατρικές εικόνες, σε συνδυασμό με σύγχρονες μεθόδους μεταφοράς στυλ και ανταγωνιστικής μάθησης, με σκοπό την δημιουργία ενός πιο στιβαρού συστήματος σε άγνωστα δεδομένα και σε αλλοιώσεις των εικόνων. Απώτερος σκοπός είναι η δημιουργία μιας συνολικής, καινούργιας επαύξησης, η οποία μπορεί να συμβάλλει με υπολογιστικά φθινό τρόπο στην εκπαίδευση χρηστών εργαλείων για τον ιατρικό τομέα. Τέτοια συστήματα θα μπορούσαν να αντικαταστήσουν αναλογικές μεθόδους στην τμηματοποίηση ιατρικών εικόνων και να λάβουν συμβουλευτικό ρόλο για τους ειδικούς, ώστε να μειωθούν τα σφάλματα διάγνωσης και να εδραιωθεί μια υψηλής ποιότητας περίθαλψη.

1.2 Οργάνωση του τόμου

Η εργασία αυτή είναι οργανωμένη σε επτά κεφάλαια. Στο Κεφάλαιο 1 δίνει η εισαγωγή της εργασίας. Στο Κεφάλαιο 2 δίνεται το θεωρητικό υπόβαθρο των βασικών τεχνολογιών που σχετίζονται με τη διπλωματική αυτή. Αρχικά περιγράφεται ο τομέας της κατάτμησης εικόνας, στην συνέχεια το πρόβλημα το οποίο αντιμετωπίζουν τα δίκτυα βαθιάς μάθησης, όσον αφορά τη γενίκευση σε άγνωστα πεδία, στη συνέχεια τα είδη μεθόδων επίλυσης και τέλος οι βασικές αρχιτεκτονικές που χρησιμοποιούνται για την επεξεργασία εικόνας. Στο Κεφάλαιο 3 αρχικά περιγράφονται οι σχετικές με το θέμα εργασίες και στη συνέχεια δίνεται ο στόχος της συγκεκριμένης εργασίας. Στο Κεφάλαιο 4 παρουσιάζεται τα συνόλα δεδομένων που χρησιμοποιήθηκαν και τα ειδικά χαρακτηριστικά τους. Στην συνέχεια, περιγράφεται η προεπεξεργασία των δεδομένων και τα προβλήματα που αντιμετωπίζονται με αυτήν στις ιατρικές εικόνες. Το Κεφάλαιο 5 παρουσιάζει με μεγαλύτερη λεπτομέρεια τα εργαλεία που χρησιμοποιήθηκαν πάνω στην έρευνα της εργασίας ενώ στο Κεφάλαιο 6 παρουσιάζονται τα πειραματικά αποτελέσματα. Τέλος στο Κεφάλαιο 7 δίνεται η συνεισφορά αυτής της διπλωματικής εργασίας, καθώς και μελλοντικές επεκτάσεις.

Μέρος I

Θεωρητικό Μέρος

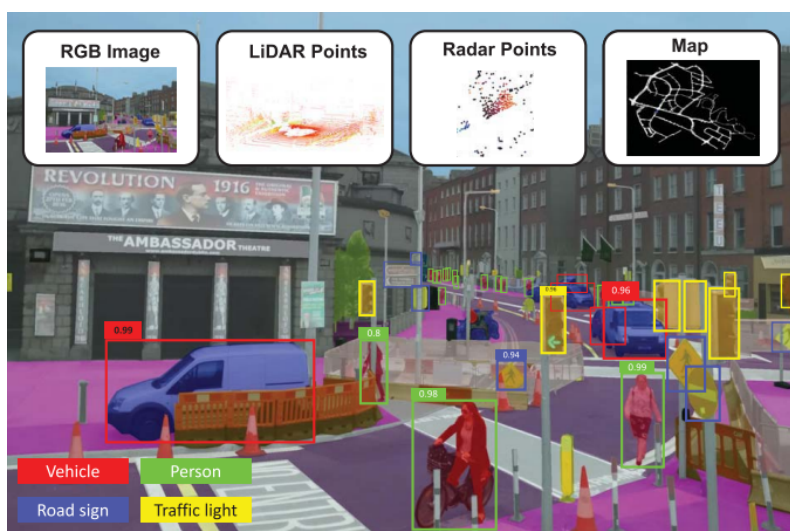
Κεφάλαιο 2

Θεωρητικό υπόβαθρο

Στο κεφάλαιο αυτό παρουσιάζεται το θεωρητικό υπόβαθρο που είναι απαραίτητο για να γίνουν κατανοητά το περιεχόμενο της διπλωματικής εργασίας και οι μέθοδοι που χρησιμοποιούνται στο πειραματικό μέρος.

2.1 Κατάτμηση Εικόνων

Η κατάτμηση εικόνων αποτελεί μία από τις κυριότερες εφαρμογές των συστημάτων βαθιάς μάθησης στην υπολογιστική όραση, κατά την οποία μια εικόνα χωρίζεται ανά εικονοστοιχείο σε διαφορετικές περιοχές ενδιαφέροντος [24]. Η κατάτμηση διαδραματίζει κεντρικό ρόλο σε ένα ευρύ φάσμα εφαρμογών, όπως για παράδειγμα στην ανάλυση ιατρικών εικόνων από ακτινογραφίες ή και άλλες εικόνες [25], [26], σε αυτόνομα οχήματα [27] (π.χ., πλοήγηση σε επιφάνειες και ανίχνευση πεζών) και εφαρμογές ασφαλείας και ελέγχου από δορυφορικές εικόνες [28]. Η αρχική εικόνα μετατρέπεται σε ψηφιακή και με την χρήση ενός εκπαιδευμένου νευρωνικού δικτύου λαμβάνονται τα χαρακτηριστικά της, από τα οποία πηγάζει το τελικό αποτέλεσμα.



Σχήμα 2.1: Κατάτμηση εικόνας σε ένα σύνθετο αστικό σενάριο για την αυτόνομη οδήγηση [1].

2.1.1 Είδη Κατάτμησης Εικόνας

Μπορούμε να χωρίσουμε την κατάτμηση εικόνας σε δύο βασικές κατηγορίες:

- Κατάτμηση Περιπτώσεων (Instance Segmentation)
- Σημασιολογική Κατάτμηση (Semantic Segmentation)

Κατάτμηση Περιπτώσεων

Η κατάτμηση περιπτώσεων αφορά την αναγνώριση και τμηματοποίηση διαφορετικών αντικειμένων μέσα σε μία εικόνα με βάση την εμφάνισή τους ή το περιεχόμενό τους [29] (Σχήμα 2.2). Οι τεχνικές για την εκτέλεση κατάτμησης περιπτώσεων μπορούν να χωριστούν σε δύο κατηγορίες:

- Μέθοδος "από κάτω προς τα πάνω": Ξεκινά με την ανίχνευση μεμονωμένων εικονοστοιχείων σε μια εικόνα και στη συνέχεια τα ομαδοποιεί για να σχηματίσει ένα αντικείμενο.
- Μέθοδος από πάνω προς τα κάτω: Ανιχνεύει ολόκληρη τη σκηνή στην εικόνα, εντοπίζει μεμονωμένα αντικείμενα και τα τμηματοποιεί



Σχήμα 2.2: Παράδειγμα Κατάτμησης Περιπτώσεων [2]

Σημασιολογική Κατάτμηση

Η σημασιολογική κατάτμηση είναι η διαδικασία ανάθεσης μιας ετικέτας σε κάθε εικονοστοιχείο μιας εικόνας εισόδου [29] (Σχήμα 2.3).



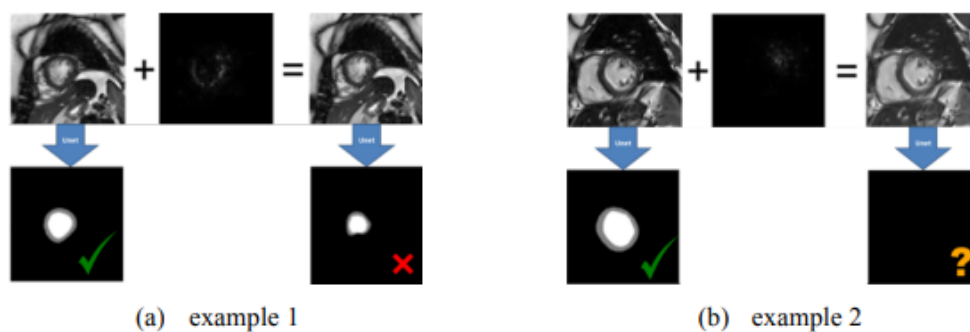
Σχήμα 2.3: Παράδειγμα Σημασιολογικής Κατάτμησης [2]

2.1.2 Το OOD πρόβλημα

Τα μοντέλα βαθιάς μάθησης συνήθως κάνουν την παραδοχή ότι τα δεδομένα εκπαίδευσης και τα δεδομένα στα οποία θα εφαρμοστούν βρίσκονται στην ίδια κατανομή, δηλαδή η μέση τιμή και η τυπική απόκλιση των χαρακτηριστικών των εικόνων αυτών είναι ίδιες [30], [31], [32], [33], [34]. Όμως, στην πραγματικότητα, πολλοί εξωγενείς παράγοντες, από τον τρόπο που λήφθηκε η εικόνα μέχρι και το περιβάλλον, προκαλούν μεταβολές στο πεδίο της εικόνας, τις λεγόμενες μεταβάσεις πεδίου (domain shifts) [35].

Έστω ότι οι εισοδοί και οι έξοδοι ενός μοντέλου συνδέονται μέσω της από κοινού συνάρτησης κατανομής πιθανότητας $P(X, Y)$, όπου X οι εισοδοί και Y οι έξοδοι. Συνήθως οι εισοδοί της εκπαίδευσης και της επικύρωσης προέρχονται από την ίδια κατανομή, οπότε δεν υπάρχουν μεταβολές στην από κοινού συνάρτηση κατά την εκπαίδευση. Όμως, όταν γίνει εφαρμογή του μοντέλου σε κάποιο ρεαλιστικό επίπεδο, τα δεδομένα που χρησιμοποιούνται για την εκπαίδευση, την επικύρωση και τη δοκιμή προέρχονται από την υπό συνθήκη κατανομή $P(X, Y|Z \in U)$ όπου Z είναι κάποια τυχαία μεταβλητή που μπορεί να μην παρατηρούμε, η οποία δεν είναι ανεξάρτητη από Y και X , και U είναι ένα κατάλληλο υποσύνολο της Z [36].

Παραδείγματος χάριν, στον ιατρικό τομέα, το πρόβλημα έγγυται στο γεγονός ότι σε επίπεδο εικόνας δύο MRI ακτινογραφίες οι οποίες έχουν ληφθεί από δύο διαφορετικά μηχανήματα μπορεί να παρουσιάζουν μεγάλες αποκλίσεις, ακόμα και αν αναφέρονται στο ίδιο περιεχόμενο. Αυτό οφείλεται στην φύση των μηχανημάτων λήψης ιατρικών εικόνων να προσθέτουν θόρυβο και τεχνουργήματα στις εικόνες και λόγω των διαφορών στο μαγνητικό πεδίο, που, ακόμα και με την κατάλληλη προεπεξεργασία, μπορεί να οδηγήσουν σε σημαντικές διαφορές στο επίπεδο της εκπαίδευσης ενός μοντέλου τεχνητής νοημοσύνης [3],[37].



Σχήμα 2.4: Δύο παραδείγματα που δείχνουν ότι εκπαιδευμένα U-Net μοντέλα είναι ευάλωτα σε προσεκτικά υπολογισμένες διαταραχές που προστίθεται στην αρχική εικόνα [3].

Ως αποτέλεσμα αυτού του προβλήματος, η απόδοση των μοντέλων βαθιάς μάθησης πέφτει δραματικά σε περιπτώσεις όπου υπάρχουν μεγάλες διαφορές πεδίων [31], [38], [39], γεγονός που αποτρέπει την πρακτική και διαδεδομένη εφαρμογή τους σε μεγάλη κλίμακα. Μια προφανής λύση για το πρόβλημα αυτό είναι η χρήση εικόνων από ποικιλία πεδίων, ώστε το μοντέλο να γίνει στιβαρό σε τέτοιες αποκλίσεις. Όμως, η λύση αυτή πολλές φορές είναι ανέφικτη, καθώς σε δεν υπάρχει όγκος και ποικιλία δεδομένων ελεύθερα για χρήση [37]. Για αυτόν τον λόγο, στο τομέα της βαθιάς μάθησης έχουν υιοθετηθεί διαφορετικές μέθοδοι προσέγγισης επίλυσης του προβλήματος χωρίς την ανάγκη εκπαίδευσης πάνω σε υπέρτομο σύνολο δεδομένων, οι λεγόμενες μέθοδοι γενίκευσης πεδίου (domain generalization).

2.2 Γενίκευση Πεδίου

Για να ξεπεραστεί, λοιπόν, το πρόβλημα της μετάβασης πεδίου, καθώς και η απουσία δεδομένων, γίνεται κυρίως χρήση μεθόδων γενίκευσης πεδίου πάνω σε μοντέλα βαθιάς μάθησης. Η γενίκευση πεδίου είναι ένα πρόβλημα μηχανικής μάθησης που αποσκοπεί στην εκμάθηση ενός μοντέλου που μπορεί να γενικεύσει σε άγνωστους τομείς, λαμβάνοντας υπόψη μόνο επισημασμένα δεδομένα από ένα σύνολο αρχικών πεδίων κατά τη διάρκεια της εκπαίδευσης [11]. Ο στόχος της γενίκευσης πεδίου είναι η εκμάθηση μιας αναπαράστασης αναλλοίωτης ως προς το πεδίο, η οποία αποτυπώνει την κοινή υποκείμενη δομή σε διαφορετικούς πεδία, ενώ παράλληλα είναι ανθεκτική σε παραλλαγές που αφορούν συγκεκριμένα πεδία. Η γενίκευση πεδίου είναι ιδιαίτερα χρήσιμη σε σενάρια όπου το πεδίο-στόχος είναι άγνωστος ή μη προσβάσιμος κατά τη διάρκεια της εκπαίδευσης, όπως στην ιατρική διάγνωση ή την αυτόνομη οδήγηση.

Η ιστορία των μεθόδων γενίκευσης πεδίου χρονολογείται από τις αρχές της δεκαετίας του 2000, όταν το πρόβλημα εισήχθη για πρώτη φορά ως πρόβλημα μηχανικής μάθησης από τους Blanchard et al [40]. Σε αντίθεση με άλλα συναφή προβλήματα μάθησης, όπως η προσαρμογή πεδίου ή η μάθηση μεταφοράς, η γενίκευση πεδίου εξετάζει τα σενάρια όπου τα δεδομένα-στόχος είναι απρόσιτα κατά τη διάρκεια της μάθησης του μοντέλου. Το κίνητρο πίσω από την γενίκευση πεδίου προέρχεται από μια ιατρική εφαρμογή που ονομάζεται αυτόματη πύλη δεδομένων ροής κυτταρομετρίας [40]. Στόχος ήταν ο σχεδιασμός αλγορίθμων για την αυτοματοποίηση της διαδικασίας ταξινόμησης των κυττάρων σε δείγματα αίματος

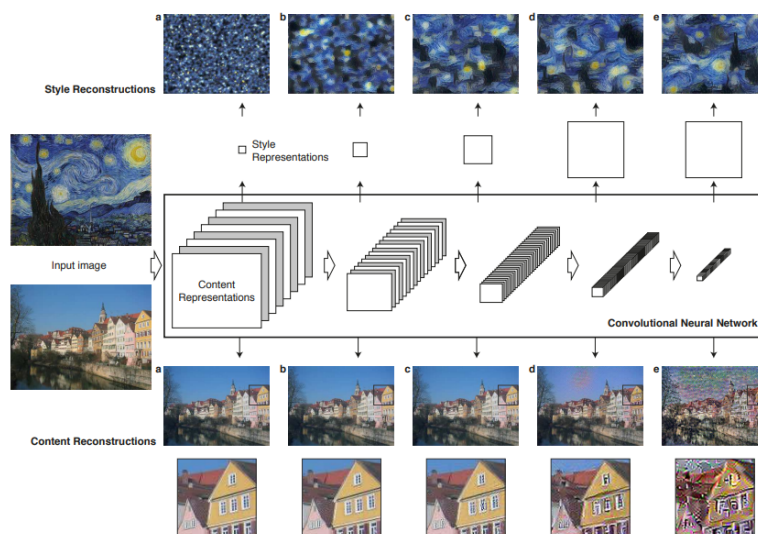
ασθενών με βάση διαφορετικές ιδιότητες, π.χ. για τη διάκριση μεταξύ λεμφοκυττάρων και μη λεμφοκυττάρων [40]. Μια τέτοια τεχνολογία θα μπορούσε να είναι ζωτικής σημασίας για τη διευκόλυνση της διάγνωσης της υγείας των ασθενών, δεδομένου ότι η χειροκίνητη ταξινόμηση είναι εξαιρετικά χρονοβόρα και απαιτεί ειδική εμπειρογνωμοσύνη σε συγκεκριμένο τομέα.

Η γενίκευση πεδίου έχει μελετηθεί εκτενώς στη βιβλιογραφία, με πολυάριθμες μεθόδους που έχουν αναπτυχθεί σε διάφορους τομείς εφαρμογών. Αυτές οι μέθοδοι μπορούν να κατηγοριοποιηθούν σε γενικές γραμμές σε τέσσερις ομάδες: μέθοδοι βασισμένες σε χαρακτηριστικά, μέθοδοι βασισμένες σε μετρικές, μέθοδοι βασισμένες σε μοντέλα και μέθοδοι βασισμένες σε μετα-εκπαίδευση, κάποιες από τις οποίες θα αναλυθούν περαιτέρω στο κομμάτι της σχετικής εργασίας. Οι μέθοδοι βασισμένες σε χαρακτηριστικά στοχεύουν στην εκμάθηση μιας αναπαράστασης χαρακτηριστικών, αμετάβλητων ως προς το πεδίο, προσθέτοντας όρους κανονικοποίησης στη συνάρτηση απώλειας ή χρησιμοποιώντας δίκτυα διαχωρισμού πεδίων. Οι μέθοδοι που βασίζονται στη μετρική στοχεύουν στην εκμάθηση ενός μετρικού χώρου που είναι αναλλοίωτος σε μεταβολές που αφορούν το πεδίο, ενώ οι μέθοδοι που βασίζονται στο μοντέλο στοχεύουν στην εκμάθηση ενός μοντέλου που είναι ανθεκτικό στη μετατόπιση του πεδίου. Οι μέθοδοι που βασίζονται στη μετα-εκπαίδευση στοχεύουν στην εκμάθηση ενός μετα-μαθητή που μπορεί να προσαρμοστεί γρήγορα σε νέα πεδία με λίγα δείγματα με ετικέτες.

2.3 Μεταφορά Στυλ

Η μεταφορά στυλ είναι μια καινούργια τεχνική στην όραση υπολογιστών και την επεξεργασία εικόνας που επιτρέπει τη μετατροπή του καλλιτεχνικού στυλ μιας εικόνας, διατηρώντας παράλληλα το περιεχόμενό της [41]. Η έννοια αυτή απέκτησε ευρεία προσοχή με την εισαγωγή των αλγορίθμων νευρωνικής μεταφοράς στυλ, οι οποίοι αξιοποιούν τη δύναμη των βαθιών νευρωνικών δικτύων για τη δημιουργία οπτικά εντυπωσιακών και καλλιτεχνικών εικόνων. Στον πυρήνα της μεταφοράς στυλ βρίσκεται η ιδέα του διαχωρισμού και του επανασυνδυασμού του περιεχομένου και του στυλ δύο διαφορετικών εικόνων για την παραγωγή μιας νέας και μοναδικής σύνθεσης (Σχήμα 2.4).

Η διαδικασία μεταφοράς στυλ περιλαμβάνει συνήθως δύο κύρια συστατικά: την εικόνα περιεχομένου και την εικόνα στυλ. Η εικόνα περιεχομένου περιέχει τα αντικείμενα και τα στοιχεία που θα διατηρηθούν στο τελικό αποτέλεσμα, ενώ η εικόνα στυλ ενσωματώνει τα στυλιστικά χαρακτηριστικά που θέλουμε να εφαρμόσουμε στο περιεχόμενο [42]. Τα νευρωνικά δίκτυα διαδραματίζουν καθοριστικό ρόλο σε αυτή τη διαδικασία, αναλύοντας την εικόνα περιεχομένου για να εξάγουν τα χαρακτηριστικά και την εικόνα στυλ για να αποτυπώσουν τα στυλιστικά χαρακτηριστικά της. Αυτά τα χαρακτηριστικά χρησιμοποιούνται στη συνέχεια για τη δημιουργία μιας νέας εικόνας που συνδυάζει το περιεχόμενο από την εικόνα περιεχομένου με τις υφολογικές ιδιότητες της εικόνας στυλ [42].



Σχήμα 2.5: Διαδικασία μεταφοράς στυλ από πίνακα ζωγραφικής σε εικόνες του πραγματικού κόσμου [4]

Χρησιμοποιώντας εικόνες που παρουσιάζουν διαφορές στα πεδία τους, μπορούμε να χρησιμοποιήσουμε την μεταφορά στυλ για να δημιουργήσουμε εικόνες διαφορετικών πεδίων, εμπλουτίζοντας το σύνολο δεδομένων και βελτιώνοντας την γενίκευση του μοντέλου από ένα και μόνο σύνολο δεδομένων [23]. Οι ποσότητες που χρησιμοποιούνται κυρίως είναι η μέση τιμή και η διακύμανση των χαρτών χαρακτηριστικών, καθώς είναι τα κύρια μεγέθη στα οποία αποθηκεύεται στα επίπεδα ενός νευρωνικού δικτύου η πληροφορία για το στυλ μιας εικόνας [21].

2.4 Μοντέλα βαθιάς μάθησης - Νευρωνικά Δίκτυα

Θεωρείται φρόνημο σε αυτό το σημείο να δοθούν βασικές πληροφορίες για τα κύρια μοντέλα βαθιάς μάθησης που χρησιμοποιούνται στον τομέα επεξεργασίας εικόνας.

2.4.1 CNN-Συνελικτικά Νευρωνικά Δίκτυα

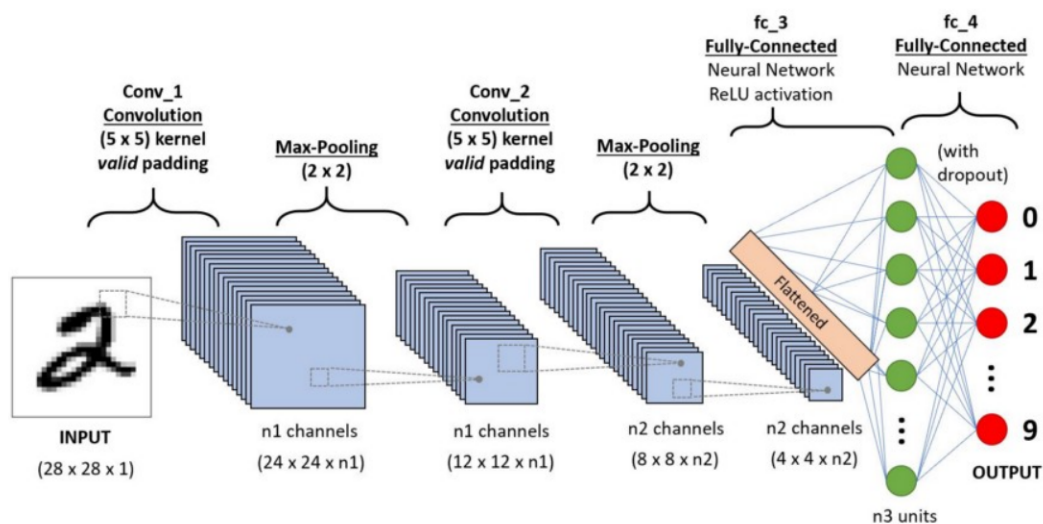
Η βασική αρχιτεκτονική για την επεξεργασία εικόνων είναι τα συνελικτικά νευρωνικά δίκτυα (CNN). Τα CNNs δημιουργήθηκαν για να αντιμετωπίσουν την αυξανόμενη πολυπλοκότητα και το αυξανόμενο μέγεθος των συνολών δεδομένων και να αντικαταστήσουν τα δίκτυα πρόσθειας μετάβασης, τα οποία δημιουργούσαν πολλές συνδέσεις μεταξύ των επιπέδων και κατέληξαν δύσχρηστα [43]. Τα CNNs έχουν σχεδιαστεί για να μαθαίνουν αυτόματα και προσαρμοστικά μοτίβα και χαρακτηριστικά από εικόνες και άλλες δομές δεδομένων που μοιάζουν με πλέγματα. Αυτό το επιτυγχάνουν μέσω στρωμάτων συνελίξεων και πράξεων συγκέντρωσης, τα οποία τους επιτρέπουν να ανιχνεύουν ιεραρχίες χαρακτηριστικών, από απλές ακμές έως σύνθετα μέρη αντικειμένων. Αυτή η ιεραρχική εκμάθηση χαρακτηριστικών καθιστά CNNs ιδιαίτερα κατάλληλα για εργασίες όπως η ταξινόμηση εικόνων, η ανίχνευση αντικειμένων, η αναγνώριση προσώπου και άλλα [44]. Εκτός από την αποτελεσματικότητά τους στην όραση υπολογιστών, τα CNNs έχουν εφαρμοστεί σε διάφορους τομείς, συμπερι-

λαμβανομένης της επεξεργασίας φυσικής γλώσσας, της ανάλυσης ιατρικών εικόνων, ακόμη και σε παιχνίδια.

Με περισσότερες λεπτομέρειες, ο αλγόριθμος CNN βασίζεται σε διάφορες ενότητες που είναι δομημένες σε μια συγκεκριμένη ροή εργασίας [5], οι οποίες παρατίθενται ως εξής:

- Εικόνα εισόδου
- Στρώμα συνέλιξης (πυρήνας)
- Στρώμα συγκέντρωσης
- Ταξινόμηση - Στρώμα πλήρους σύνδεσης
- Αρχιτεκτονικές

Ένα δίκτυο CNN δέχεται μια εικόνα ως είσοδο, διακρίνει τα αντικείμενά της με βάση τρία χρωματικά επίπεδα και προσδιορίζει διάφορους χρωματικούς χώρους και μετρά, επίσης, τις διαστάσεις της εικόνας [5].

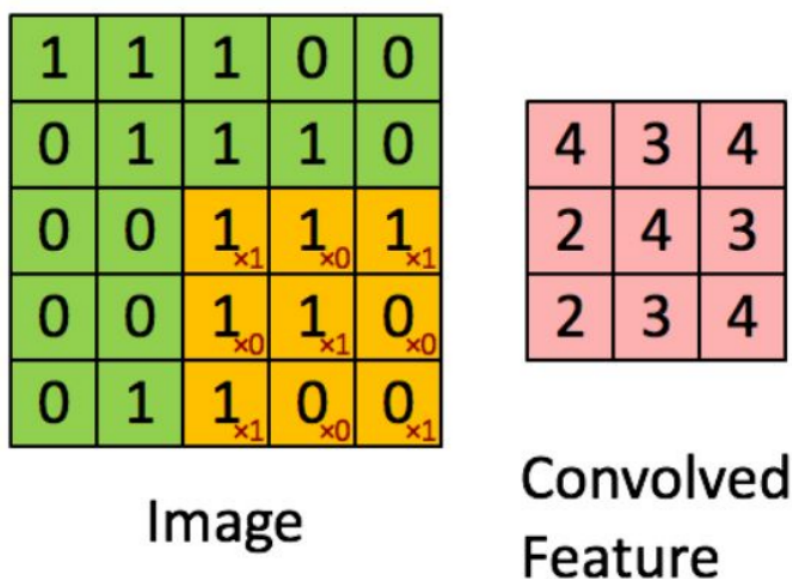


Σχήμα 2.6: Παράδειγμα της διαδικασίας ενός δικτύου CNN [5]

Επίπεδο Συνέλιξης

Ο πυρήνας του CNN λειτουργεί με βάση τον ακόλουθο τύπο: Διαστάσεις εικόνας = $n1 \times n2 \times 1$ όπου $n1$ = ύψος, $n2$ = πλάτος και 1 = αριθμός καναλιών, όπως RGB. Στην παρακάτω εικόνα, το πράσινο τμήμα δείχνει τον τύπο $5 \times 5 \times 1$ [5]. Το κίτρινο πλαίσιο εξελίσσεται από το πρώτο πλαίσιο μέχρι το τελευταίο, εκτελώντας τη συνελκτική πράξη σε κάθε πίνακα 3×3 .

Ο στόχος αυτής της λειτουργίας συνέλιξης είναι να ληφθούν όλα τα υψηλού επιπέδου χαρακτηριστικά της εικόνας. Τα χαρακτηριστικά υψηλού επιπέδου μπορούν να περιλαμβάνουν και τις ακμές της εικόνας [5]. Αυτό το επίπεδο δεν περιορίζεται μόνο στα χαρακτηριστικά υψηλού επιπέδου- εκτελεί επίσης μια λειτουργία σε χαρακτηριστικά χαμηλού επιπέδου, όπως το χρώμα και ο προσανατολισμός της κλίσης [5].



Σχήμα 2.7: Παράδειγμα του αποτελέσματος συνέλιξης ενός φίλτρου [5]

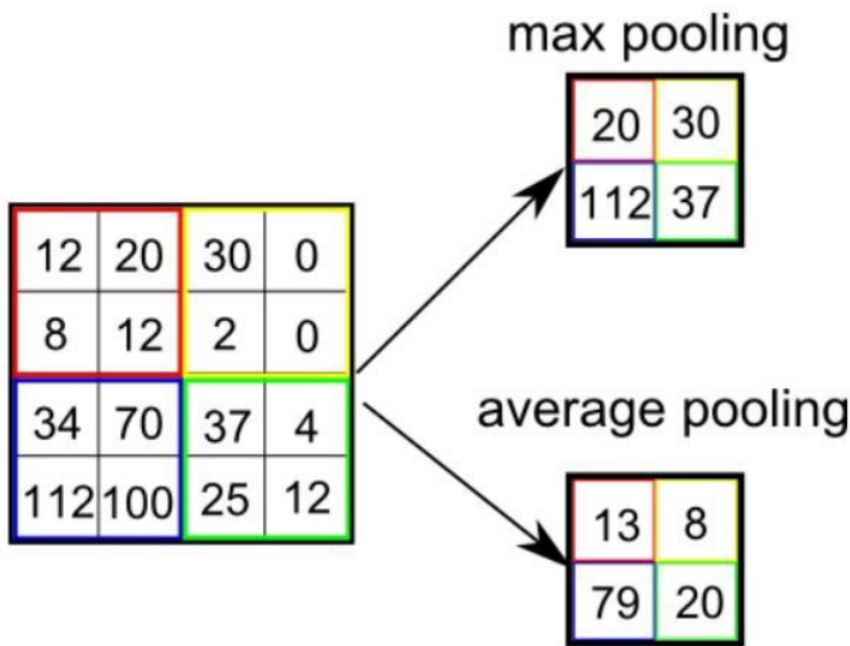
Ο στόχος των στρωμάτων γεμίματος είναι να μειώσουν τη διαστατικότητα της εικόνας που βρίσκεται στην αρχική εικόνα εισόδου και να αυξήσουν τη διαστατικότητα ή, σε ορισμένες περιπτώσεις, να την αφήσουν αμετάβλητη, ανάλογα με την απαιτούμενη έξοδο.

Επίπεδο Pooling

Ο πρωταρχικός στόχος του στρώματος Pooling είναι να μειώσει το χωρικό μέγεθος του των χαρακτηριστικών του συνελκτικού επιπέδου. Λειτουργεί για τη μείωση της απαιτούμενης υπολογιστικής ισχύος για την επεξεργασία των δεδομένων με τη μέθοδο της μείωσης της διάστασης [5]. Επιπλέον, είναι επίσης ευεργετικό για την εξαγωγή των κυρίαρχων χαρακτηριστικών, τα οποία είναι κατά βάση περιστροφικά καθώς και αναλλοίωτα ως προς τη θέση, οπότε απαιτείται η διατήρηση της διαδικασίας αποτελεσματικά. Υπάρχουν κυρίως δύο διαφορετικοί τύποι Pooling, οι οποίοι είναι οι εξής:

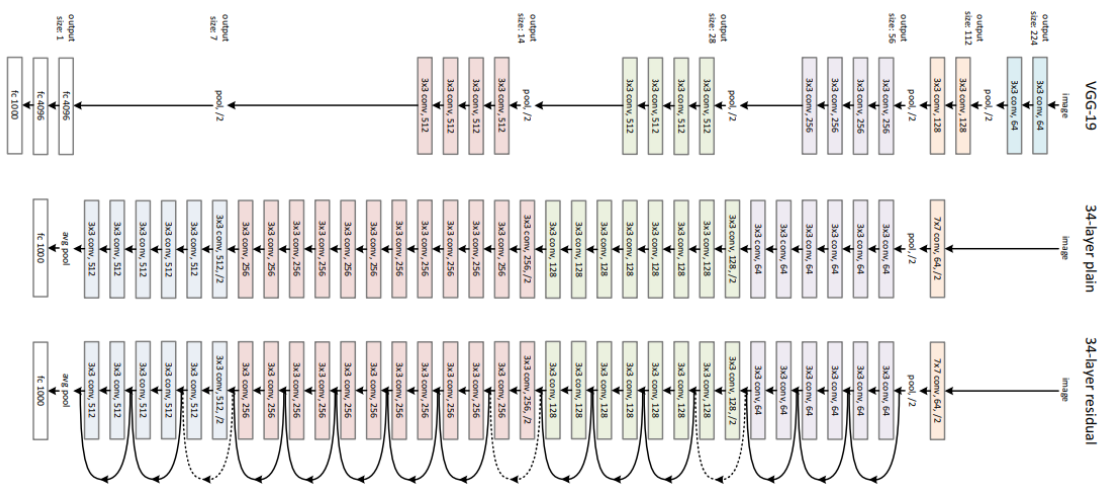
- Μέγιστη Συγκέντρωση (Max Pooling): Το Max Pooling παρέχει τη μέγιστη τιμή εντός της καλυμμένης εικόνας από τον πυρήνα.
- Μέση Συγκέντρωση (Average Pooling): Η Μέση Συγκέντρωση παρέχει και επιστρέφει τη μέση τιμή εντός της καλυμμένης εικόνας από τον πυρήνα.

Με την πάροδο των ετών, τα CNN έχουν σημειώσει αξιοσημείωτες βελτιώσεις τόσο στην αρχιτεκτονική όσο και στις επιδόσεις. Η πρόοδος μπορεί να αποδοθεί σε διάφορους βασικούς παράγοντες, όπως μεγαλύτερα και ποικίλα σύνολα δεδομένων, βελτιωμένες τεχνικές βελτιστοποίησης (π.χ. διακοπή, ομαλοποίηση παρτίδων, βελτιωμένες μεθόδους επαύξησης δεδομένων) και βαθύτερες αρχιτεκτονικές δικτύων. Η ανάπτυξη προ-εκπαιδευμένων μοντέλων, όπως το VGG [44], το ResNet [6] και το Inception [45], επέτρεψε στους ερευνητές και



Σχήμα 2.8: Παράδειγμα της διαδικασίας pooling [5]

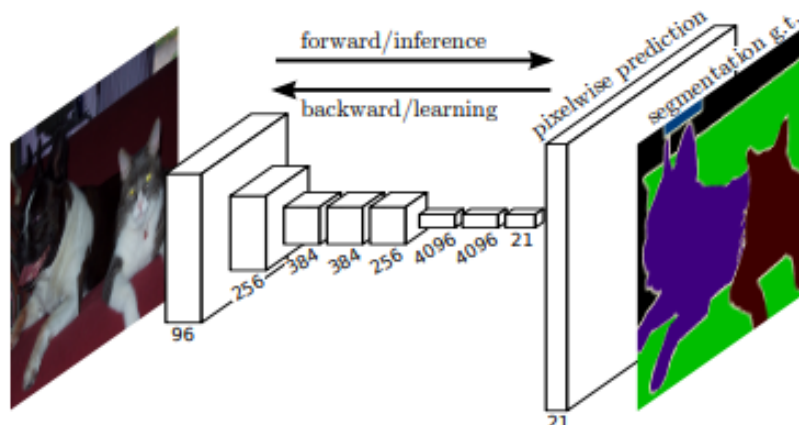
τους επαγγελματίες να αξιοποιήσουν τη εκμάθηση μεταφοράς, όπου τα μοντέλα που εκπαιδεύτηκαν σε μια εργασία μπορούν να ρυθμιστούν λεπτομερώς για μια άλλη, με αποτέλεσμα να έχουμε σημαντική αύξηση της αποδοτικότητας και της ακρίβειας [45]. Επιπλέον, οι επιταχυντές υλικού, όπως οι κάρτες γραφικών (GPUs) και οι κάρτες επεξεργασίας τένσων (TPUs), έχουν διαδραματίσει καθοριστικό ρόλο στην επιτάχυνση της εκπαίδευσης των CNNs, καθιστώντας δυνατή την αντιμετώπιση πιο σύνθετων εργασιών και μεγαλύτερων συνόλων δεδομένων [43].



Σχήμα 2.9: Αρχιτεκτονικές VGG, ResNet [6]

2.4.2 FCN-Πλήρως Συνελικτικά Δίκτυα

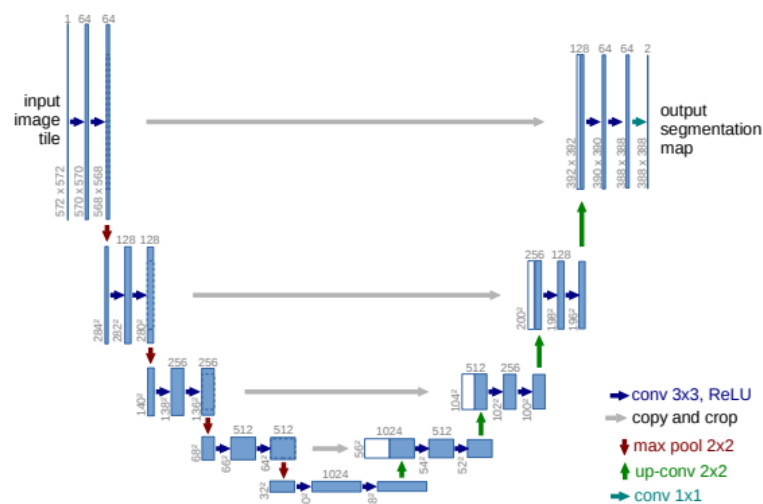
Τα CNNs χρησιμοποιήθηκαν στο τομέα της κατάτμησης εικόνας υπό αλλαγμένη μορφή, τα πλήρως συνελικτικά δίκτυα (Fully Convolutional Network - FCN). Σε αντίθεση με τα παραδοσιακά CNNs, τα FCNs είναι ειδικά προσαρμοσμένα για εργασίες σημασιολογικής κατάτμησης σε επίπεδο εικονοστοιχείου, όπου ο στόχος είναι η ταξινόμηση κάθε εικονοστοιχείου μιας εικόνας σε προκαθορισμένες κλάσεις [7]. Τα FCNs διαφέρουν από τα CNNs αντικαθιστώντας τα πλήρως συνδεδεμένα στρώματα με στρώματα συνέλιξης, γεγονός που συμβάλλει στη διατήρηση της χωρικής πληροφορίας σε όλο το δίκτυο. Παράλληλα, μέσω της χρήσης συνδέσεων παράλειψης στις οποίες οι χάρτες χαρακτηριστικών από τα τελευταία στρώματα του μοντέλου αναβαθμίζονται και συγχωνεύονται με χάρτες χαρακτηριστικών των προηγούμενων στρωμάτων, το μοντέλο συνδυάζει σημασιολογικές πληροφορίες (από βαθιά, χονδροειδή στρώματα) και πληροφορίες εμφάνισης (από ρηχά, λεπτά στρώματα) σε προκειμένου να παράγει ακριβείς και λεπτομερείς τμηματοποιήσεις.



Σχήμα 2.10: Παράδειγμα διαδικασίας ενός δικτύου FCN. Διαθέτουν την δυνατότητα να μάθουν αποτελεσματικά να κάνουν πυκνές προβλέψεις για εργασίες ανά εικονοστοιχείο [7]

Μια αξιοσημείωτη αρχιτεκτονική εμπνευσμένη από τα FCN είναι το U-Net, γνωστό για την αρχιτεκτονική του σε σχήμα U που αποτελείται τόσο από συρρικνούμενα όσο και από επεκτεινόμενα μονοπάτια, καθιστώντας το ιδιαίτερα αποτελεσματικό για εργασίες όπως η κατάτμηση εικόνων [8] (Σχήμα 2.7). Το δίκτυο προτάθηκε αρχικά για την τμηματοποίηση βιολογικών εικόνων μικροσκοπίας. Το δίκτυο και η εκπαιδευτική στρατηγική βασίστηκαν στη χρήση επαύξησης δεδομένων για την εκμάθηση από τις πολύ λίγες σχολιασμένες εικόνες αποτελεσματικά. Η U-Net αρχιτεκτονική αποτελείται από δύο μέρη, ένα συμβαλλόμενο μονοπάτι για τη σύλληψη του πλαισίου, και ένα συμμετρικό επεκτατικό μονοπάτι που επιτρέπει τον ακριβή εντοπισμό [8]. Το τμήμα της δειγματοληψίας προς τα κάτω ή της συρρίκνωσης έχει μια αρχιτεκτονική που μοιάζει με FCN και εξάγει χαρακτηριστικά με συνέλιξεις 3×3 . Το τμήμα δειγματοληψίας προς τα πάνω ή επέκτασης μέρος χρησιμοποιεί ανοδική συνέλιξη (ή αποσυμπίεση), μειώνοντας το αριθμό των χαρτών χαρακτηριστικών, αυξάνοντας παράλληλα τις διαστάσεις τους. Χάρτες χαρακτηριστικών από το τμήμα του δικτύου με τη δειγματοληψία προς τα κάτω αντιγράφονται στο τμήμα ανοδικής δειγματοληψίας για να αποφευχθεί η απώλεια πληροφορίας των προτύπων [8]. Τέλος, μια συνέλιξη 1×1 επεξεργάζεται τους χάρτες

χαρακτηριστικών για τη δημιουργία ενός χάρτη τμηματοποίησης που κατηγοριοποιεί κάθε εικονοστοιχείο της εικόνας εισόδου.



Σχήμα 2.11: Αρχιτεκτονική U-Net (παράδειγμα για 32x32 εικονοστοιχεία στη χαμηλότερη ανάλυση). Κάθε μπλε πλαίσιο αντιστοιχεί σε χάρτη χαρακτηριστικών πολλαπλασίων καναλιών. Ο αριθμός των καναλιών συμβολίζεται στην κορυφή του πλαισίου. Το μέγεθος x-y παρέχεται στο κάτω αριστερό άκρο του πλαισίου. Λευκό κουτιά αντιπροσωπεύουν αντιγραμμένους χάρτες χαρακτηριστικών. Τα βέλη υποδηλώνουν τις διάφορες λειτουργίες [8]

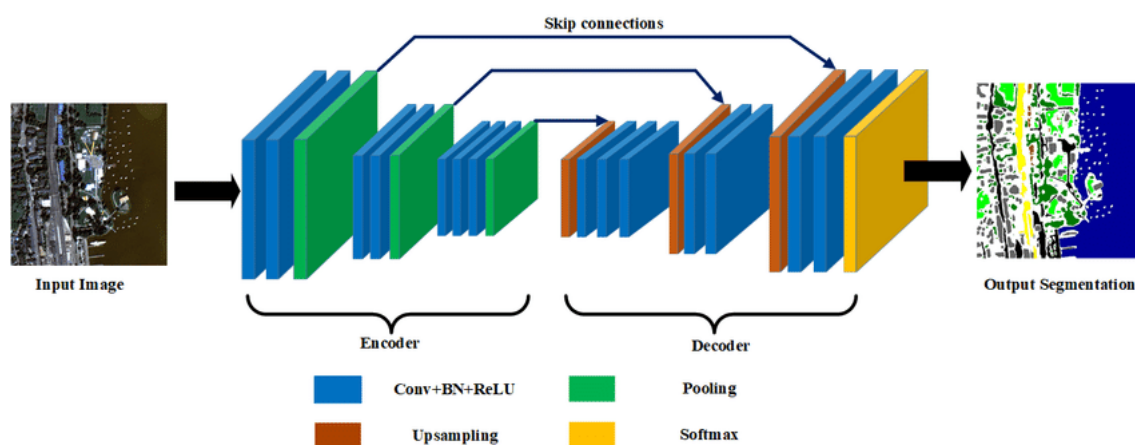
Στον ιατρικό τομέα, τα FCN και παραλλαγές όπως το U-Net έχουν βρει εκτεταμένη εφαρμογή, όπως σε εφαρμογές κατάτμησης όγκων εγκεφάλου [46], τμηματοποίηση με επίγνωση της περίπτωσης [47], κατάτμηση καρκίνων του δέρματος [48] και κατάτμηση της ίριδας [49]. Έφεραν επανάσταση στην ανάλυση ιατρικών εικόνων, επιτρέποντας την ακριβή και αυτοματοποιημένη κατάτμηση οργάνων και όγκων σε εικόνες από διάφορες μεθόδους [50], όπως μαγνητική τομογραφία, αξονικές τομογραφίες και ακτίνες X [8]. Τα FCN επιτρέπουν στους επαγγελματίες της ιατρικής να εξοικονομούν χρόνο και να κάνουν ακριβέστερες διαγνώσεις παρέχοντας λεπτομερείς χάρτες ανατομικών δομών ή παθολογικών περιοχών.

Ωστόσο, παρά τη δημοτικότητα και την αποτελεσματικότητά του, το συμβατικό μοντέλο FCN έχει ορισμένους περιορισμούς - δεν είναι αρκετά γρήγορο για εξαγωγή συμπερασμάτων σε πραγματικό χρόνο, δεν λαμβάνει υπόψη του τις πληροφορίες του παγκόσμιου πλαισίου σε ένα αποδοτικό και δεν είναι εύκολα μεταβιβάσιμο σε τρισδιάστατες εικόνες. Μεταγενέστερες αρχιτεκτονικές αποσκοπούσαν στην διόρθωση αυτών των προβλημάτων.

2.4.3 Μοντέλα Encoder-Decoder

Οι αρχιτεκτονικές κωδικοποιητή-αποκωδικοποιητή είναι μια ισχυρή κατηγορία μοντέλων νευρωνικών δικτύων που χρησιμοποιούνται ευρέως σε διάφορους τομείς, όπως η όραση υπολογιστών, η επεξεργασία φυσικής γλώσσας και άλλα. Αυτές οι αρχιτεκτονικές είναι ιδιαίτερα αξιολογικές για την ικανότητά τους να μειώνουν την πληροφορία σε ένα χώρο χαμηλότερης διάστασης χρησιμοποιώντας έναν κωδικοποιητή και στη συνέχεια να ανακατασκευάζουν ή να χρησιμοποιούν αυτή τη μειωμένη πληροφορία χρησιμοποιώντας έναν αποκωδικοποιητή [51],[52].

Στο πλαίσιο της επεξεργασίας εικόνας, ο συστατικός κωδικοποιητής της αρχιτεκτονικής λαμβάνει μια είσοδο υψηλής διάστασης, όπως μια εικόνα, και την ανάγει προοδευτικά σε μια αναπαράσταση χαμηλότερης διάστασης που συχνά αναφέρεται ως λανθάνων χώρος ή στενωπός (latent space) [51],[52]. Αυτή η διαδικασία περιλαμβάνει μια σειρά από επίπεδα συνελικτικής και ομαδοποίησης που συλλαμβάνουν όλο και πιο αφηρημένα και συμπαγή χαρακτηριστικά από την είσοδο. Αυτές οι αναπαραστάσεις χαμηλότερης διάστασης διατηρούν βασικές πληροφορίες σχετικά με την είσοδο, ενώ απορρίπτουν περιττές λεπτομέρειες, καθιστώντας το έναν αποτελεσματικό τρόπο εξαγωγής ουσιαστικών χαρακτηριστικών από πολύπλοκα δεδομένα.



Σχήμα 2.12: Μια απεικόνιση της αρχιτεκτονικής FCN με βάση τον κωδικοποιητή-αποκωδικοποιητή για σημασιολογική κατάτμηση [9]

Το στοιχείο αποκωδικοποιητή, από την άλλη πλευρά, λαμβάνει τις μειωμένες πληροφορίες από τον κωδικοποιητή και τις χρησιμοποιεί για να ανακατασκευάσει την αρχική είσοδο ή να παράγει μια έξοδο με τα επιθυμητά χαρακτηριστικά. Για εργασίες που σχετίζονται με εικόνες, ο αποκωδικοποιητής αποτελείται συνήθως από συνελικτικά στρώματα που αναβαθμίζουν τη λανθάνουσα αναπαράσταση, επεκτείνοντάς την σταδιακά στις αρχικές διαστάσεις εισόδου [51],[52]. Αυτή η ανακατασκευασμένη έξοδος μπορεί να εξυπηρετήσει διάφορους σκοπούς, από την αποθορυβοποίηση εικόνας και την αφαίρεση χρώματος έως τη μεταφορά στυλ και την υπερανάλυση. Σε άλλες εφαρμογές, όπως η μηχανική μετάφραση στην επεξεργασία φυσικής γλώσσας, ο αποκωδικοποιητής χρησιμοποιεί τις πληροφορίες χαμηλότερης διάστασης για να παράγει ακολουθίες κειμένου ή άλλα δεδομένα. Συνολικά, οι αρχιτεκτονικές κωδικοποιητή-αποκωδικοποιητή παρέχουν ένα ευέλικτο πλαίσιο για τη συμπίεση και τη χρήση πληροφοριών, επιτρέποντας ένα ευρύ φάσμα εργασιών όπου η αποτελεσματική εξαγωγή και ανακατασκευή χαρακτηριστικών είναι απαραίτητη [51],[52].

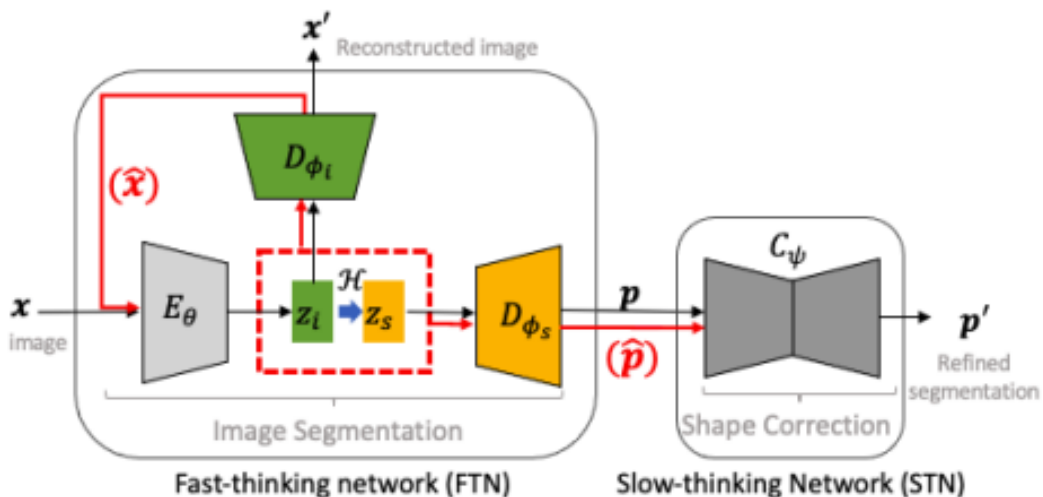
2.4.4 Μοντέλα FTN-STN

Πρόσφατες έρευνες έχουν βρει μεθόδους στις οποίες κατάλληλες αρχιτεκτονικές μπορούν να βοηθήσουν στην αύξηση σιθαρότητας προς τις αλλαγές πεδίων των εικόνων. Ένα τέτοιο πλαίσιο είναι τα συνεργατικά δίκτυα FTN-STN (Fast Thinking Network-Slow Thinking Network), τα οποία βασίζονται στην λειτουργία του εγκεφάλου με γρήγορο τρόπο να

παίρνει απλές αποφάσεις αυτόματα και είναι πιο επιρρεπές σε σφάλματα και με αργό τρόπο να επεξεργάζεται περίπλοκες αποφάσεις [10], [53].

Το προτεινόμενο πλαίσιο αποτελείται από ένα δίκτυο ταχείας σκέψης, το οποίο δεδομένης μιας εικόνας x , εξάγει χαρακτηριστικά σχήματος z_s ειδικά για την εργασία τμηματοποίησης και τα χαρακτηριστικά πλαισίου z_l της εικόνας για να εκτελέσει την εργασία ανακατασκευής της εικόνας [10]. Το δίκτυο αποτελείται από έναν κωδικοποιητή E_θ , ένα αποσυζεύκτη χαρακτηριστικών H και δύο αποκωδικοποιητές D_{ϕ_s} και D_{ϕ_l} για τις εργασίες τμηματοποίησης και ανακατασκευής εικόνας. Ο αποσυζεύκτης κώδικα H εφαρμόζεται στα z_l χαρακτηριστικά ώστε οι πληροφορίες που δεν σχετίζονται με την εργασία να απενεργοποιούνται στο z_s . Ο αποσυζεύκτης H χρησιμοποιεί μια στοίβα από δύο επίπεδα συνελκτικής ανάλυσης ακολουθούμενα από μια συνάρτηση ενεργοποίησης ReLU [10].

Το δίκτυο βραχείας σκέψης είναι ένα δίκτυο αυτόματου κωδικοποιητή αποθρομβοποίησης C_ψ , το οποίο διορθώνει την κατάτμηση που προβλέπεται από το δίκτυο ταχείας σκέψης χρησιμοποιώντας ένα μαθημένο προηγούμενο σχήμα που κωδικοποιείται στο δίκτυο [10].



Σχήμα 2.13: Το δίκτυο FTN-STN [10]

Το δίκτυο εκπαιδεύεται σε τρία σφάλματα, το σφάλμα ανακατασκευής, το σφάλμα κατάτμησης και το σφάλμα σχήματος στην περίπτωση που χρησιμοποιείται το δίκτυο βραχείας σκέψης [10]. Επίσης, χρησιμοποιείται τυχαία κάποια μορφή επαύξησης δεδομένων στο λανθάνων επίπεδο για να αποφευχθούν περιπτώσεις υπερπροσαρμογής. Η λανθάνουσα αύξηση δεδομένων χώρου εκτελεί συγκάλυψη σε επίπεδο καναλιού και χώρου για τον λανθάνοντα κώδικα που μαθαίνεται από το FTN με τυχαίο και στοχευμένο τρόπο. Η ανακατασκευή εικόνων με μασκοποιημένους λανθάνουσες κώδικες παράγει ένα ποικίλο σύνολο απαιτητικών εικόνων και αλλοιωμένους χάρτες κατάτμησης για την ενίσχυση της εκπαίδευσης και των δύο δικτύων [10].

Κεφάλαιο 3

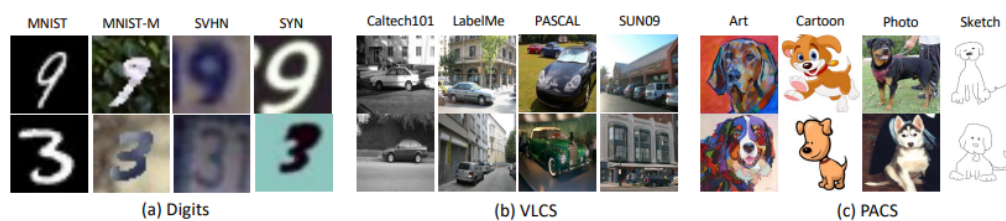
Σχετική βιβλιογραφία μεθόδων γενίκευσης πεδίου σε εικόνες

Στο κεφάλαιο αυτό αρχικά γίνεται μια περιγραφή των διαφόρων μεθόδων που εφαρμόζονται για την γενίκευση πεδίου, που είχαν επίδραση και στην έρευνα της διπλωματικής εργασίας.

3.1 Σχετικές εργασίες

3.1.1 Προσαρμογή Πεδίου

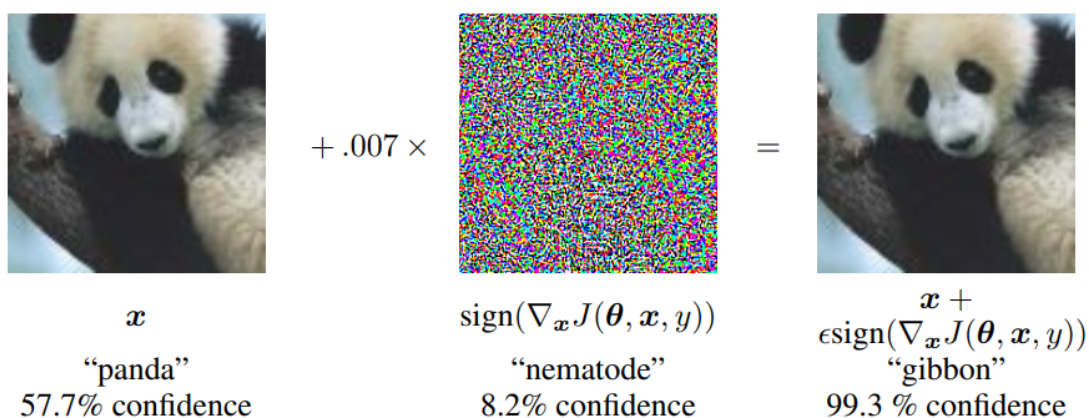
Μία από τις πιο συνηθισμένες μεθόδους γενίκευσης πεδίου είναι η προσαρμογή πεδίου (domain adaptation), η οποία έχει διερευνηθεί σε εκτενή βαθμό [54], [55], [56], [57], [58]. Η προσαρμογή πεδίου αφορά την προσέγγιση της άγνωστης κατανομής, όμως κάνει την παραδοχή ότι υπάρχουν αραιώς επισημασμένα ή ακόμα και μη επισημασμένα δεδομένα για την εκπαίδευση [59], [60], [61], [62]. Το πρόβλημα που παρουσιάζεται εδώ είναι ότι δεν γίνεται να γίνει εφαρμογή σε περιπτώσεις που δεν έχουμε δεδομένα από διαφορετικά πεδία. Υπάρχουν ορισμένες έρευνες που έχουν ασχοληθεί με μη-επιβλεπόμενη κατάτμηση εικόνας [63],[64], όμως χρειάζεται αρκετά μεγάλος όγκος δεδομένων για να παρουσιάσει καλή απόδοση, γεγονός που δεν ισχύει για πολλές εφαρμογές.



Σχήμα 3.1: Παραδείγματα εικόνων από τρία σημεία αναφοράς γενίκευσης πεδίου που εκδηλώνουν διαφορετικούς τύπους μετατόπισης πεδίου. Στο (α), η μετατόπιση τομέα αντιστοιχεί κυρίως σε αλλαγές στο στυλ της γραμματοσειράς, το χρώμα και το φόντο. Στο (β), είναι σαφείς οι μεροληψίες που αφορούν συγκεκριμένα σύνολα δεδομένων, οι οποίες οφείλονται σε αλλαγές σε περιβάλλον/σκηνικό και την οπτική γωνία. Στο (γ), οι αλλαγές στο στυλ της εικόνας είναι ο κύριος λόγος για τη μετατόπιση τομέα [11].

3.1.2 Ανταγωνιστική Εκπαίδευση

Η ανταγωνιστική εκπαίδευση είναι μια τεχνική μηχανικής μάθησης που έχει σχεδιαστεί για την ενίσχυση της στιβαρότητας των μοντέλων βαθιάς μάθησης. Αντλεί έμπνευση από αντίπαλα παραδείγματα, τα οποία είναι εισοδοί σκόπιμα σχεδιασμένες για να παραπλανήσουν ένα νευρωνικό δίκτυο. Στην ανταγωνιστική εκπαίδευση, το μοντέλο εκτίθεται σε αυτά τα αντίξοα παραδείγματα κατά τη διάρκεια της διαδικασίας εκπαίδευσης, αναγκάζοντάς το να μάθει να είναι πιο ανθεκτικό σε τέτοιες διαταραχές. Η βασική ιδέα προήλθε από την ανακάλυψη ότι η πρόσθεση ενός ανεπαίσθητου θορύβου σε εικόνες μπορεί να οδηγήσει ένα μοντέλο βαθιάς μάθησης στο να βγάλει λάθος αποτελέσματα, παρά το γεγονός ότι δεν φαίνεται κάποια αλλαγή στο ανθρώπινο μάτι [12].



Σχήμα 3.2: Παρουσίαση του αποτελέσματος πρόσθεσης θορύβου στην πρόβλεψη ενός δικτύου ταξινόμησης εικόνας [12]

Ο πρωταρχικός ρόλος της ανταγωνιστικής εκπαίδευσης στην αύξηση της ανθεκτικότητας έγκειται στην ικανότητά της να βελτιώνει την απόδοση ενός μοντέλου όταν αντιμετωπίζει αντίξοες εισόδους ή άγνωστες κατανομές δεδομένων. Με την ενσωμάτωση αντίπαλων παραδειγμάτων στο σύνολο δεδομένων εκπαίδευσης, το μοντέλο μαθαίνει να αναγνωρίζει και να προσαρμόζεται σε λεπτές αλλαγές και αποκλίσεις στα δεδομένα εισόδου, καθιστώντας το λιγότερο ευάλωτο σε χειραγώγηση και εκμετάλλευση. Αυτό όχι μόνο ενισχύει την ικανότητα του μοντέλου να υπερασπίζεται έναντι εχθρικών επιθέσεων, αλλά και βελτιώνει τη συνολική απόδοση της γενίκευσής του. Η εκπαίδευση με ανταγωνιστικό τρόπο έχει βρει εφαρμογές σε διάφορους τομείς, όπως η όραση υπολογιστών, η επεξεργασία φυσικής γλώσσας και η ενισχυτική μάθηση.

Η ανταγωνιστική εκπαίδευση έχει χρησιμοποιηθεί κυρίως στα μοντέλα παραγωγικής ανταγωνιστικής εκπαίδευσης (GANs) στην οποία λύνεται το πρόβλημα ελαχιστοποίησης της κατανομής μέσω ενός παιγνίου δύο παικτών *minimax* [65]. Αυτό επιτυγχάνεται με την εκμάθηση ενός διαχωριστή για τη διάκριση μεταξύ πραγματικών και παραγόμενων ψεύτικων εικόνων, ενθαρρύνοντας παράλληλα τη γεννήτρια να ξεγελάσει τον διαχωριστή [65]. Παράλληλα, ως αποτέλεσμα αυτής της χρήσης για την αύξηση στιβαρότητας των μοντέλων, χρησιμοποιείται ευρέως ως μέθοδος γενίκευσης πεδίου [66], [67], [68], [69]. Σε σύγχρονες μεθόδους, η ανταγωνιστική μάθηση μπορεί να γίνεται και χωρίς την χρήση κάποιου διαχωριστή, αλλά με τον “ανταγωνισμό” μεταξύ δύο διαφορετικών σφαλμάτων κατά την εκπαίδευση

[23].

3.1.3 Μεταφορά Στυλ

Η μεταφορά στυλ είναι μια τεχνική στην όραση υπολογιστών που περιλαμβάνει τη δημιουργία μιας νέας εικόνας συνδυάζοντας το περιεχόμενο μιας εικόνας με το στυλ μιας άλλης εικόνας. Στόχος της μεταφοράς στυλ είναι η δημιουργία μιας εικόνας που διατηρεί το περιεχόμενο της αρχικής εικόνας, ενώ εφαρμόζει το οπτικό στυλ μίας άλλης [70]. Χάρη στις πρόσφατες εξελίξεις στη μεταφορά στυλ [71], [72], έχει εισαχθεί ευρέως ως μέθοδος γενίκευσης πεδίου. Αρκετές μέθοδοι γενίκευσης [73], [74], [75] χρησιμοποιούν έτοιμα μοντέλα μεταφοράς στυλ, όπως το AdaIN [71] ή δημιουργούν δίκτυα τα οποία μαθαίνουν από δείγματα να κάνουν επαύξηση δεδομένων με συγκεκριμένα στυλ [23],[72]. Παράλληλα, αξιοποιούνται εξωτερικά στυλ, για να αυξηθεί η ποικιλία των δεδομένων εκπαίδευσης [76]. Είναι μία αρκετά χρήσιμη τεχνική, καθώς μπορούν να δημιουργηθούν εικόνες που αναπαριστούν διαφορετικά πεδία και να οδηγηθούμε σε καλύτερη γενίκευση των μοντέλων, ακόμα και με λίγα δεδομένα.

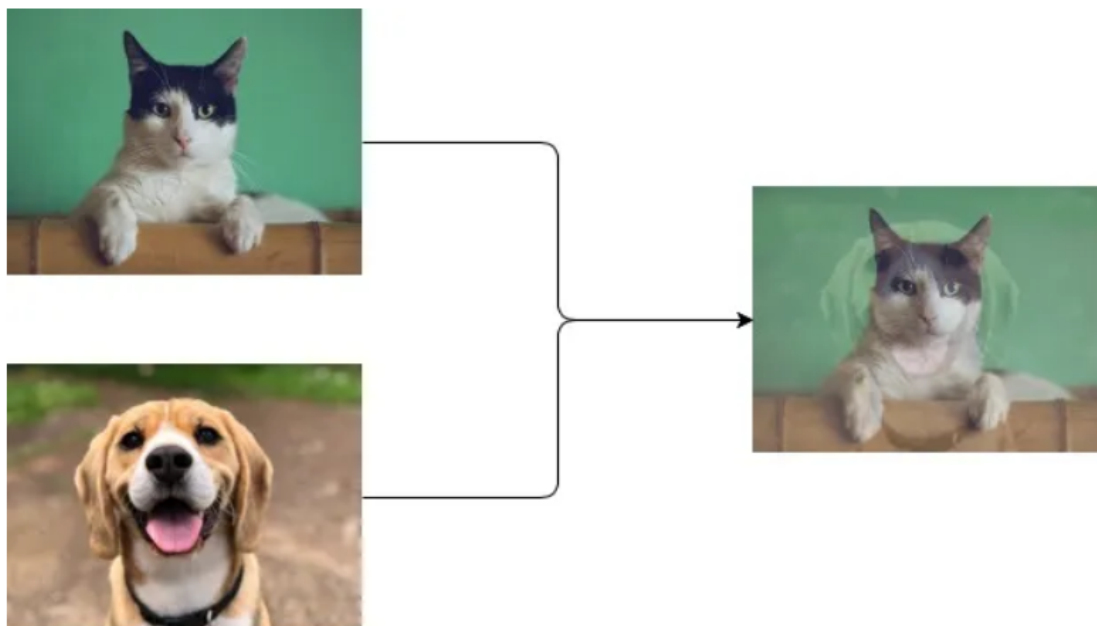


Σχήμα 3.3: Παράδειγμα δημιουργίας δειγμάτων για προσαρμογή πεδίου με ρυθμιζόμενες υπερπαραμέτρους [13]

3.1.4 Επαύξηση στο επίπεδο των χαρακτηριστικών

Η επαύξηση στο επίπεδο των χαρακτηριστικών έχει αναδειχθεί ως μια πρωτοποριακή τεχνική ειδικά προσαρμοσμένη για την αντιμετώπιση των προκλήσεων της γενίκευσης πεδίου [21], [16]. Στο πλαίσιο της γενίκευσης πεδίου, η επαύξηση με βάση τα χαρακτηριστικά αξιοποιεί την παραδοχή ότι τα στατιστικά στοιχεία των χαρακτηριστικών του CNN περιλαμβάνουν πληροφορίες που σχετίζονται με τον τομέα [21], [16]. Τα MixStyle και Mixup είναι δύο χαρακτηριστικά παραδείγματα αυτής της προσέγγισης. Το MixStyle εισάγει την επαύξηση στυλ με την ανάμειξη στατιστικών χαρακτηριστικών CNN που προέρχονται από περιπτώσεις που ανήκουν σε διαφορετικά πεδία. Με αυτόν τον τρόπο, τα μοντέλα προσαρμόζονται αποτελεσματικότερα σε πεδία που δεν είχαν προηγουμένως δει, προσδίδοντας διαφορετικά οπτικά στυλ και χαρακτηριστικά [21]. Το Mixup, από την άλλη πλευρά εκτελεί επαύξηση σε επίπεδο εικονοστοιχείου, επεκτείνοντας την επιρροή του στο χώρο των χαρακτηριστικών, επιτρέποντας έτσι την αρμονική συγχώνευση περιπτώσεων που προέρχονται από διαφορετικούς τομείς σε επίπεδο χαρακτηριστικών [16]. Η ανάμειξη επεκτείνει την κατανομή εκπαίδευσης ενσωματώνοντας την προηγούμενη γνώση ότι η γραμμική παρεμβολή των διανυσμάτων χαρακτηριστικών πρέπει να οδηγούν σε γραμμικές παρεμβολές των σχετικών στόχων. Με την ενσωμάτωση στρατηγικών επαύξησης βασισμένων σε χαρακτηριστικά όπως το MixStyle και το Mixup, οι ερευνητές έχουν προωθήσει τον τομέα της γενίκευσης πεδίου, εξοπλίζοντας τα

μοντέλα με την ικανότητα να αποδίδουν αξιόπιστα σε διάφορους τομείς χωρίς την ανάγκη εκτεταμένης εκπαίδευσης σε συγκεκριμένα πεδία.



Σχήμα 3.4: Παράδειγμα εφαρμογής μεθόδου Mixup [14]

3.1.5 Εφαρμογές Βαθιάς Μάθησης

Αρχιτεκτονικές βαθιών νευρωνικών δικτύων έχουν υλοποιηθεί και χρησιμοποιηθεί σε διάφορες εφαρμογές από μέλη του Εργαστηρίου Συστημάτων Τεχνητής Νοημοσύνης και Μάθησης του ΕΜΠ. Ειδικότερα επιβλεπόμενες τεχνικές CNN και CNN-RNN έχουν εφαρμοστεί για κατηγοριοποίηση αντικειμένων, στην ιατρική διάγνωση νευροεκφυλιστικών ασθενειών, όπως η νόσος του Πάρκινσον [77, 78, 79, 80, 81] ή η Covid-19 [82, 83, 84, 85, 86], περιλαμβάνοντας κατάτμηση 2-Δ ή 3-Δ εικόνων [87]. Έμφαση έχει δοθεί στην διαφάνεια και στην προσαρμογή των μοντέλων [88, 89] αλλά και στην ανάπτυξη πλέον σύνθετων αρχιτεκτονικών, μπαΐσειανών, με κάψουλες και αβεβαιότητα [90, 91, 92]. Βαθιές ημι- και αυτο-επιβλεπόμενες 3-Δ νευρωνικές αρχιτεκτονικές, αλλά και αρχιτεκτονικές κωδικοποιητή-αποκωδικοποιητή έχουν εφαρμοστεί στην ανίχνευση βλαβών σε πυρηνικούς αντιδραστήρες [93, 83, 94], στην πρόβλεψη της παραγωγής στον αγροτικό τομέα [95, 96], και στην αναγνώριση και σύνθεση συναισθήματος [97, 98, 89, 99]. Άλλες εφαρμόζονται σε προβλήματα αλληλεπίδρασης ανθρώπου-υπολογιστή [100, 101, 102], και ανάλυσης εικόνων [103, 104, 105, 106, 107, 108, 109].

Μέρος 

Πρακτικό Μέρος

Κεφάλαιο 4

Σύνολα Δεδομένων

Στο κεφάλαιο αυτό θα παρουσιαστούν τα σύνολα δεδομένων που χρησιμοποιήθηκαν στα πλαίσια της διπλωματικής εργασίας και οι διαδικασίες που ακολουθήθηκαν για την προεπεξεργασία των δεδομένων εκπαίδευσης.

4.1 Δεδομένα Προστάτη

Η πρώτη από τις δύο εφαρμογές που εξετάστηκε το δίκτυο ήταν στην κατάτμηση προστάτη από εικόνες μαγνητικής τομογραφίας. Τα δεδομένα προέρχονται από πολλαπλά σύνολα δεδομένων, και αποτελούνται συνολικά από 141 πολυπαραμετρικές έρευνες, δηλαδή βελτιωμένες εκδοχές MRI που σκοπεύουν στην καλύτερη οπτικοποίηση του αντικειμένου, οι περισσότερες από τις οποίες αποτελούν έρευνες T2-βαρύτητας για προστάτη.

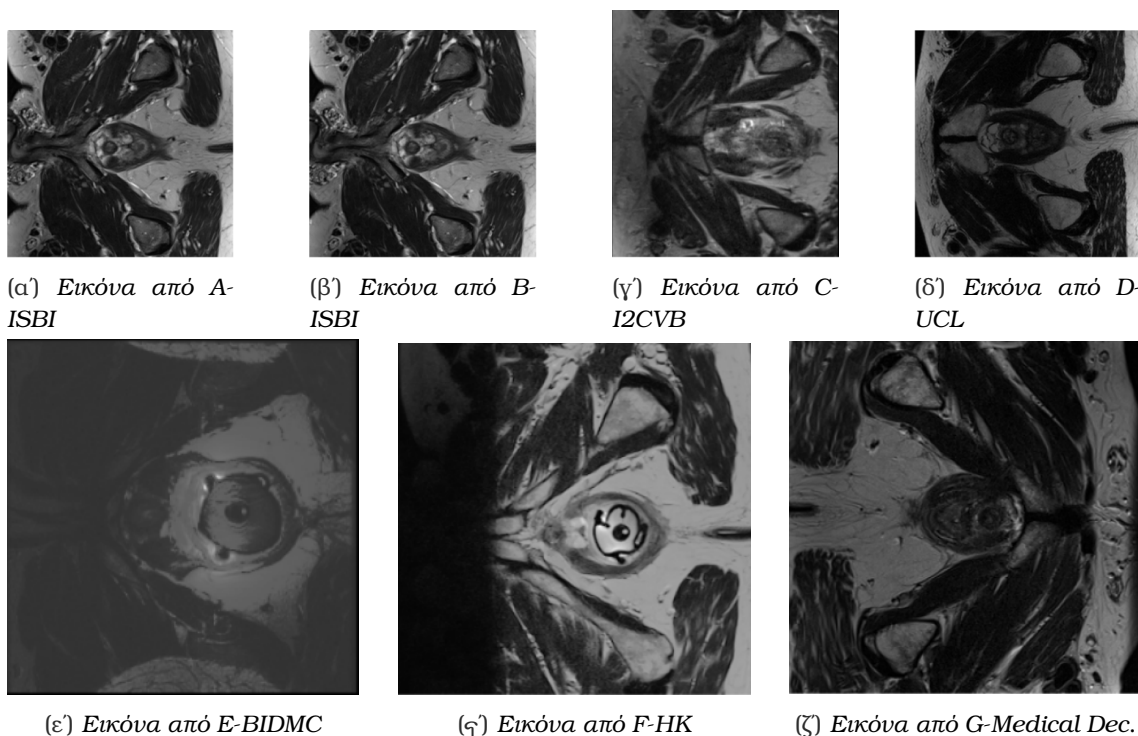
Η τεχνική T2 αναφέρεται στην ρύθμιση του μαγνητικού τομογράφου για να δώσει έμφαση στους χρόνους χαλάρωσης των ιστών. Ως "T2" ορίζεται ο χρόνος που χρειάζονται τα ενεργοποιημένα πρωτόνια των ιστών για να επαναεθυγραμμιστούν, αφού ενεργοποιηθούν από ραδιοκύματα. Ιστοί με υψηλούς T2 χρόνους εμφανίζονται με σκούρο χρώμα, ενώ οι ιστοί με χαμηλούς T2 χρόνους εμφανίζονται πιο φωτεινοί. Οι εικόνες αυτές είναι εξαιρετικά χρήσιμες στο να αναγνωρίζουν απαλούς ιστούς και όργανα που διατηρούν υγρές ουσίες και χρησιμοποιούνται συχνά στην ανίχνευση ανωμαλιών.

Συνολικά για την έρευνα χρησιμοποιήθηκαν 7 διαφορετικά σύνολα δεδομένων:

- NCI-ISBI-2013: Δύο διαφορετικά σύνολα δεδομένων από τον διαγωνισμό του 2013, που έγινε από το πρόγραμμα απεικόνισης καρκίνου του Εθνικού Ινστιτούτου Καρκίνου (NCI) σε συνεργασία με την Διεθνή Εταιρεία Βιοϊατρικής Απεικόνισης (ISBI). Οι μισές εικόνες ελήφθησαν σε 1.5T (Philips Achieva) με ενδο-ορθικό πηνίο δέκτη (από το Ιατρικό κέντρο της Βοστώνης) και οι άλλες μισές σε 3T (Siemens TIM) με επιφανειακό πηνίο (από το Ιατρικό Κέντρο του Πανεπιστημίου Radboud Nijmegen της Ολλανδίας) [110], [111], [112]. Οι εικόνες, λοιπόν, χωρίζονται ανάλογα την πρόελευση τους για τις δοκιμές σε OOD εικόνες.
- I2CVB: Ένα σύνολο δεδομένων από την Πρωτοβουλία για τη Συνεργατική Συγκριτική Αξιολόγηση της Όρασης Υπολογιστών (Initiative for Collaborative Computer Vision Benchmarking, I2CVB) [110], [111], [113]. Οι εικόνες ελήφθησαν από έναν 3T ανιχνευτή Siemens και με τις ακόλουθες μεθόδους για καλύτερη αντιπροσώπευση:

- Μαγνητική τομογραφία με στάθμιση T2 (T2-W).
 - Μαγνητική τομογραφία ενισχυμένης δυναμικής αντίθεσης (DCE).
 - Μαγνητική τομογραφία σταθμισμένης διάχυσης (DWI).
 - Φασματοσκοπική απεικόνιση μαγνητικού συντονισμού (MRSI).
- PROMISE12: Τρία σύνολα δεδομένων από διαφορετικά ιατρικά κέντρα και διαφορετικές μεθόδους λήψης από τον διαγωνισμό PROMISE12 για την τμηματοποίηση ιατρικών εικόνων προστάτη [110], [111], [114].
 - Medical Decathlon Dataset: Ένα σύνολο δεδομένων με πολυπαραμετρικές έρευνες από τον διαγωνισμό Medical Decathlon Dataset, ο οποίος περιέχει συνολικά 10 διαφορετικά σύνολα δεδομένων για διαφορετικά όργανα του ανθρώπινου σώματος. Το συγκεκριμένο σύνολο δεδομένων χρησιμοποιήθηκε για την εκπαίδευση και τον έλεγχο σε εφαρμογές εντός κατανομής [115].

Παρατίθενται εικόνες από κάθε dataset για την δίκαιη σύγκριση μεταξύ των εικόνων:



Σχήμα 4.1: Σύνολο Εικόνων των συνόλων δεδομένων

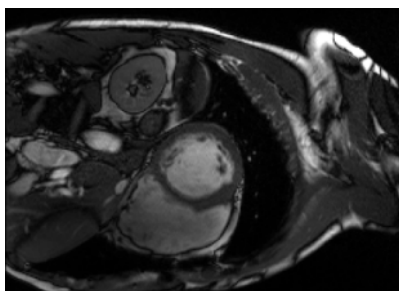
4.2 Δεδομένα Καρδιάς

Η δεύτερη εφαρμογή πάνω στην οποία εξετάστηκε το δίκτυο ήταν στην αναγνώριση καρδιάς από εικόνες μαγνητικής τομογραφίας. Τα δεδομένα προέρχονται από το MM (Multi-Centre, Multi-Vendor and Multi-Disease) cardiac segmentation challenge dataset [116], και αποτελούνται συνολικά από 100 MRI έρευνες για καρδιά. Το σύνολο δεδομένων διαθέτει ποικιλία εικόνων από διαφορετικούς φορείς (νοσοκομεία της Ισπανίας, της Γερμανίας και του

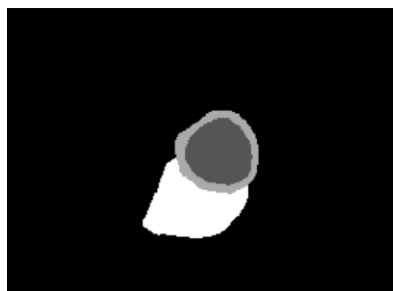
Καναδά) με τέσσερα διαφορετικά μηχανήματα και διαθέτει, επίσης, εικόνες από άτομα που πάσχουν από υπερτροφικές και διασταλτικές καρδιομυοπάθειες. Παρέχονται, επίσης, πέρα από τις τετραδιάστατες (4D) εικόνες οι τρισδιάστατες εικόνες που αντιστοιχούν στις περιόδους τελοδιαστολής και τελοσυστολής, τις χρονικές στιγμές, δηλαδή, που η καρδιά έχει τον μέγιστο όγκο στην διαστολή και τον ελάχιστο όγκο στην συστολή. Η εκπαίδευση του μοντέλου έγινε πάνω σε αυτές τις δύο χρονικές στιγμές, καθώς διαθέτουν χρήσιμη πληροφορία για τις αντλητικές ικανότητες της καρδιάς και για ποικιλία της μορφολογίας του οργάνου κατά την εκπαίδευση. Η αξιολόγηση της απόδοσης του μοντέλου έγινε πάνω σε παραποιημένες μορφές των δεδομένων και, συγκεκριμένα, πάνω σε:

- **Motion:** Τυχαία "κίνηση", που συνήθως εμφανίζεται φυσιολογικά σε MRI εικόνες λόγω της κίνησης των οργάνων,
- **Spike:** Τυχαία spike artifacts, γνωστά και ως Herringbone artifacts, δημιουργούν ραβδώσεις σε διαφορετικές κατευθύνσεις του χώρου της εικόνας, λόγω των spikes που δημιουργούνται στο ηλεκτρομαγνητικό πεδίο του μηχανήματος,
- **Ghosting:** Τυχαία μετατόπιση ενός ειδώλου της εικόνας, που συνήθως οφείλεται στην καρδιακή κίνηση, στην κίνηση του ασθενή κατά την διάρκεια της εξέτασης ή στην ροή του αίματος,
- **Bias Field:** Τυχαίες διακυμάνσεις της έντασης της εικόνας που συνήθως οφείλονται στην ανομοιογένεια του πεδίου που παράγει το μηχάνημα MRI

Παρατίθενται παράδειγμα μιας εικόνας από το σύνολο δεδομένων, καθώς και την μορφή της τμηματοποίησης για τις 3 διαφορετικές περιοχές (δεξιά κοιλία, μυοκάρδιο, αριστερή κοιλία):

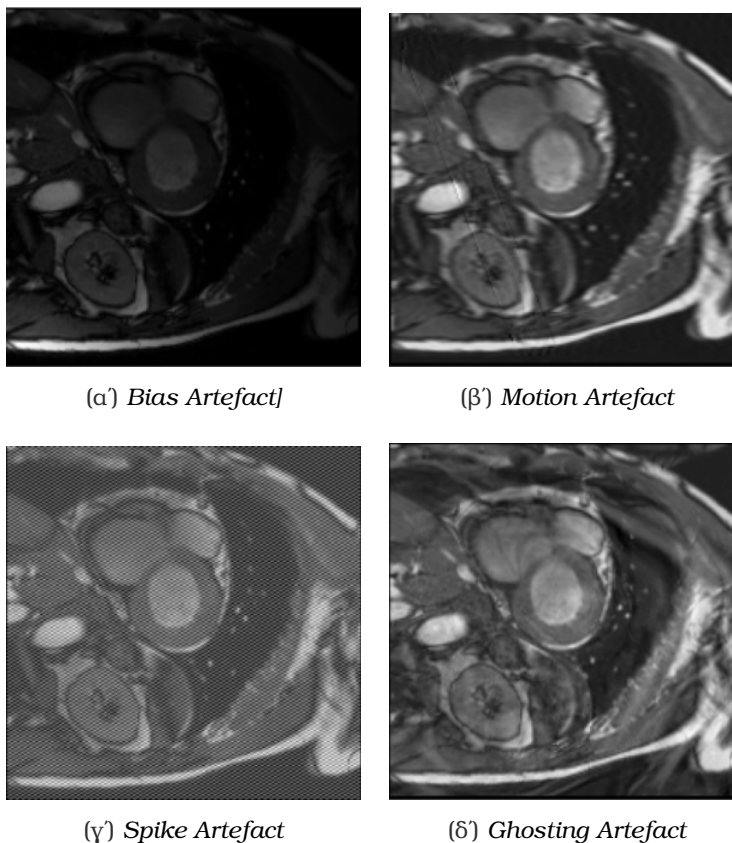


Σχήμα 4.2: Εικόνα από το σύνολο δεδομένων



Σχήμα 4.3: Κατάτμηση Εικόνας

Παρατίθενται παραδείγματα των θορυβοποιημένων μορφών των εικόνων:



Σχήμα 4.4: Θορυβοποιημένες εκδοχές εικόνας

4.3 Προεπεξεργασία Δεδομένων

Οι ιατρικές εικόνες, και ειδικότερα οι MRI εικόνες απαιτούν υψηλή προεπεξεργασία, καθώς διαθέτουν υψηλά επίπεδα θορύβου και η αρχική μορφή τους είναι γενικά αρκετά προβληματική για την επεξεργασία από τα υπολογιστικά συστήματα. Συνήθως ερευνούνται από επαγγελματίες με γνώση του πώς να επεξεργαστούν και να αναγνωρίσουν τα περιεχόμενα, χωρίς να χρειάζεται να χρησιμοποιήσουν κάποιο λογισμικό. Πέρα από τα καθιερωμένα βήματα που γίνονται στην προεπεξεργασία εικόνων, τα οποία θα αναλυθούν στην συνέχεια, γίνεται χρήση ορισμένων πολύ σημαντικών αλγορίθμων για την επεξεργασία ιατρικών εικόνων, των οποίων η σημασία δικαιολογεί να έχουν ξεχωριστό τμήμα σε αυτήν την εργασία. Οπότε, αρχικά, θα μιλήσουμε για τον αλγόριθμο N3 και μετά για τον αλγόριθμο N4, που αποτελεί την άμεση βελτίωση του προηγούμενου και τον αλγόριθμο που χρησιμοποιήθηκε σε αυτήν την εργασία.

N3 Αλγόριθμος

Λόγω της φύσης των MRI εικόνων, δηλαδή ότι τις λαμβάνουμε χρησιμοποιώντας έντονο μαγνητικό πεδίο (με τιμές της τάξης των 1.5 και 3T), παρουσιάζεται κατά την λήψη τους μη-ομοιογενής θόρυβος χαμηλών συχνοτήτων, ο οποίος είναι γνωστός ως πεδίο μεροληψίας

(bias field) [117]. Για αυτόν τον λόγο κατά την διαδικασία της προεπεξεργασίας χρειαζόμαστε έναν αλγόριθμο διόρθωσης αυτού του πεδίου θορύβου. Ο N3 αποτελεί έναν μη-παραμετρικό μη-ομοιόμορφο αλγόριθμο κανονικοποίησης εικόνων και, μέχρι πρόσφατα, αποτελούσε τον εδρεύοντα αλγόριθμο για την συγκεκριμένη εργασία. Ο αλγόριθμος λειτουργεί υποθέτοντας ότι οι εικόνες εισόδου είναι της μορφής

$$u(x) = v(x)f(x) + n(x)$$

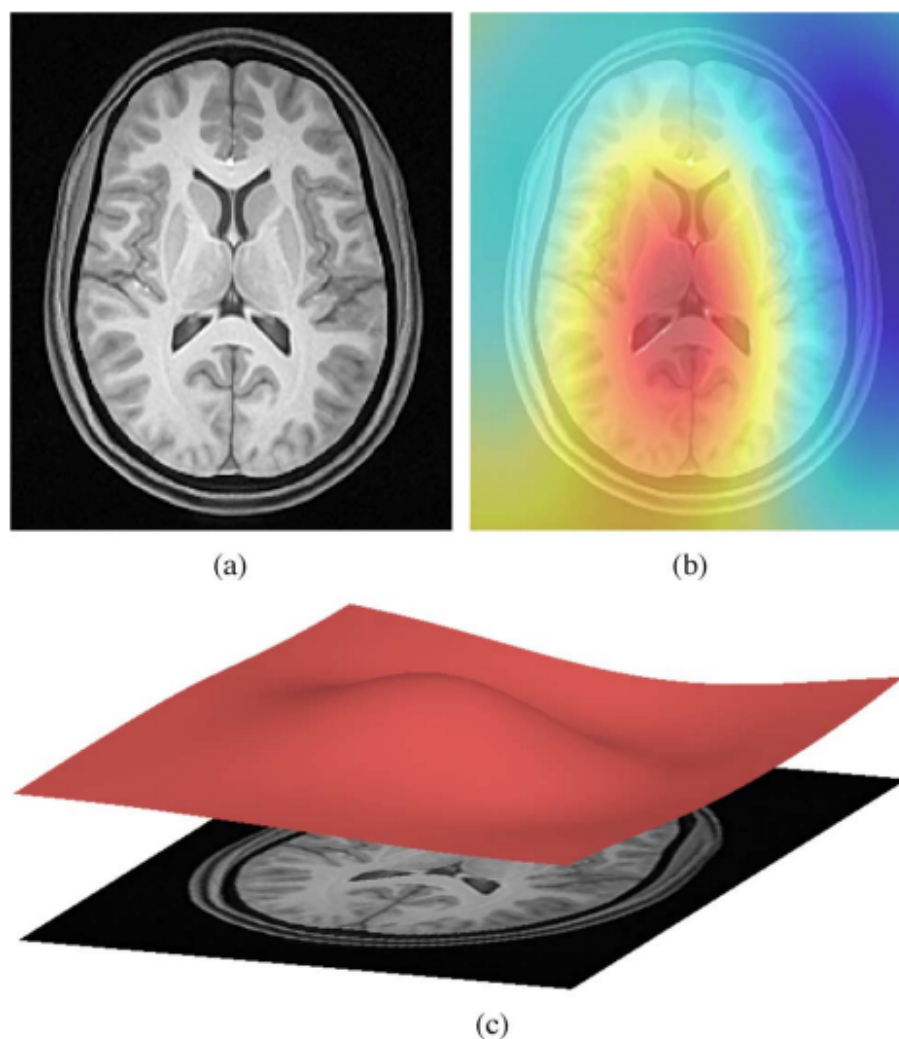
όπου u η αλλοιωμένη εικόνα, v η καθαρή εικόνα, f το πεδίο θορύβου μεροληψίας και n ο ανεξάρτητος Γκαουσιανός θόρυβος [117]. Υποθέτοντας ότι έχει γίνει χρήση ενός φίλτρου για να απορρίψουμε τον λευκό θόρυβο από την εικόνα, θεωρούμε ότι

$$\hat{u}(x) = \hat{v}(x) + \hat{f}(x)$$

όπου $\hat{u} = \log(u)$. Αναδρομικά ο αλγόριθμος λαμβάνει προσεγγίσεις για το πεδίο μεροληψίας χρησιμοποιώντας ένα δράστη λείανσης, πιο συγκεκριμένα ένα προσεγγιστή B-spline [117]. Ως "spline" ορίζεται μια συνάρτηση που εξηγείται ανά μέρη με πολυωνυμικές συναρτήσεις. Χρησιμοποιείται στην λείανση των εικόνων, στην παρεμβολή δεδομένων και στη προσέγγιση πολύπλοκων σχημάτων και ταιριάσματος καμπυλών. Τα συνδεόμενα σημεία των γραμμικών συναρτήσεων ονομάζονται κόμποι (knots). Τα B-Splines αποτελούν την βάση των splines και κάθε spline μπορεί να περιγραφθεί ως γραμμικός συνδυασμός πολλών B-Splines. Πρακτικά, τα splines αποτελούν έναν συνδυασμό εύκαμπτων μπάντων που ελέγχονται από τα σημεία σύνδεσης. Χρησιμοποιώντας, λοιπόν, αυτές τις συναρτήσεις για να προσεγγίσουμε το πεδίο μεροληψίας, η εικόνα διορθώνεται με τον παρακάτω αναδρομικό αλγόριθμο:

$$\hat{v}^n = \hat{u} - \hat{f}_e^n = \hat{u} - S\{\hat{u} - E[\hat{u}|\hat{u}^{n-1}]\}$$

Ανά αναδρομή γίνεται μία εκτίμηση του πεδίου μεροληψίας χρησιμοποιώντας την προσέγγιση του B-Spline της αναμενόμενης τιμής της καθαρής εικόνας, δεδομένου της εκτίμησης της προηγούμενης αναδρομής. Η εκτίμηση αυτή αφαιρείται από την αλλοιωμένη εικόνα για να λάβουμε, εν τέλει, την διορθωμένη [117]. Παρατηρούμε, επίσης, ότι ο αλγόριθμος δουλεύει μόνο με την εικόνα εισόδου και την εκτίμηση που κάνει αναδρομικά, δηλαδή δεν χρειάζεται κάποια πρόωρη γνώση για το πεδίο ούτε για την ίδια την εικόνα που λαμβάνει.



Σχήμα 4.5: Απεικόνιση του μη-ομογενούς πεδίου στις εικόνες μαγνητικής τομογραφίας. Η μορφή του είναι παρόμοια με ενός ανομοιόμορφου επιπέδου πάνω στην εικόνα [15]

N4 Αλγόριθμος

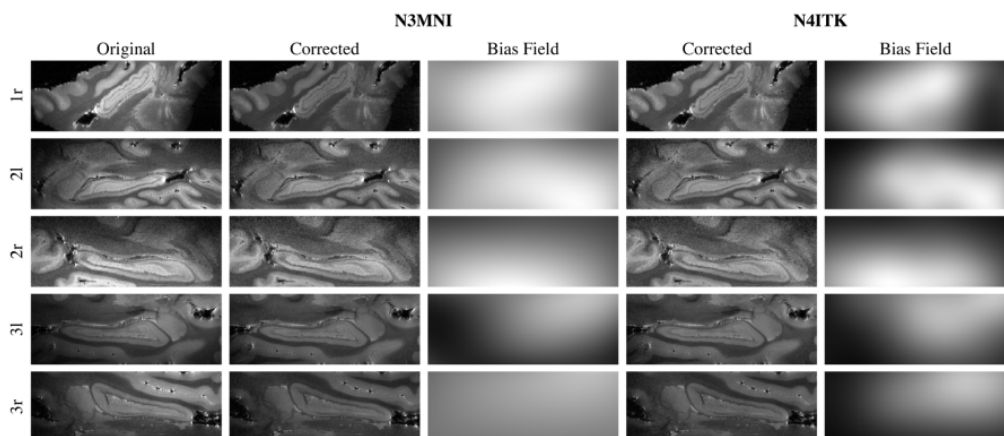
Ο N3, λόγω της ευκολίας χρήσης του, έγινε ο πιο διαδεδομένος αλγόριθμος για την διόρθωση του πεδίου μεροληψίας και αποδεικνύεται επανειλημμένως ότι ξεπερνάει τους υπόλοιπους αλγορίθμους [15]. Όμως, αυτή η ευρεία χρήση του αλγορίθμου κατέληξε στο να δημιουργήσει μία στασιμότητα στον ερευνητικό τομέα για την εύρεση βελτιωμένων αλγορίθμων. Με αφορμή αυτήν την στασιμότητα, έγιναν ορισμένες διορθώσεις και βελτιώσεις στα αναδρομικά τμήματα του αλγορίθμου και δημιουργήθηκε ο N4 αλγόριθμος. Οι συγγραφείς εκτέλεσαν ορισμένες πολύ σημαντικές αλλαγές στον αναδρομικό αλγόριθμο, ο οποίος, πλέον, είναι ο εξής:

$$\hat{u}^n = \hat{u}^{n-1} - \hat{f}_r^n = \hat{u}^{n-1} - S^* \{ \hat{u}^{n-1} - E[\hat{u} | \hat{u}^{n-1}] \}$$

Η πρώτη σημαντική αλλαγή είναι η προσέγγιση του απομείναντος πεδίο θορύβου, σε αντίθεση με την προσέγγιση του συνολικού από τον N3 αλγόριθμο. Η δεύτερη αλλαγή είναι η αντικατάσταση της αλλοιωμένης εικόνας με την προσεγγιστικά διορθωμένη σε όλα τα βήματα του αλγορίθμου, ο οποίος είναι και ο λόγος που υπολογίζεται το απομείναντα πεδίο, καθώς η

εικόνα αναδρομικά διορθώνεται όλο και περισσότερο. Η τρίτη και τελευταία αλλαγή είναι η χρήση ενός ανανεωμένου B-Spline προσεγγιστή, ο οποίος επιτρέπει για μικρότερη απόσταση στα σημεία ελέγχου για να φιλοξενήσει μεγαλύτερες εντάσεις πεδίου χωρίς την πιθανότητα αλγοριθμικής αποτυχίας, αποφεύγει την ανάγκη για μια τεχνητή παράμετρο κανονικοποίησης και επιτρέπει τον προσδιορισμό μιας σταθμισμένης περιφερειακής μάσκας για πιθανή χρήση στο πλαίσιο μιας επαναληπτικής πλαίσιο τμηματοποίησης. Τα πρόσθετα πλεονεκτήματα περιλαμβάνουν ταχύτερους χρόνους εκτέλεσης λόγω της παραλληλοποίησης του αλγορίθμου προσέγγισης B-spline και μιας στρατηγικής προσέγγισης πολλαπλών αναλύσεων για ιεραρχική προσαρμογή διαδοχικά υψηλότερων επιπέδων συχνότητας διαμόρφωσης του πεδίου πόλωσης.

Από τις πειραματικές διαδικασίες, παρατηρήθηκε ότι η απόδοση του N4 αλγορίθμου βελτιώνονταν όσο αυξανόταν η ένταση του πεδίου και του Γκαουσιανού θορύβου, καθώς και με την μείωση της απόστασης μεταξύ των συναρτήσεων splines. Οπότε, σε γενικές γραμμές, ο N4 αποτέλεσε μια σημαντική βελτίωση του N3, καθώς η υποστήριξη προσέγγισης σε πολλαπλές αναλύσεις του πεδίου εικόνας και η βελτιωμένη αναδρομική διαδικασία υπολογίζει πιο αποτελεσματικά το επίπεδο θορύβου και διορθώνει την εξάρτηση του N3 αλγορίθμου από την τυπική απόκλιση, λόγω της χασοτικής συμπεριφοράς του αναδρομικού αλγορίθμου (Σχήμα 4.2).



Σχήμα 4.6: Η πρώτη στήλη δίνει διαδιάστατες σαγματικές όψεις των μεταθανάτιων ιπποκάμπων σε τρία άτομα ("r" για τον δεξιό ιππόκαμπο, "l" για τον αριστερό ιππόκαμπο). Η δεύτερη και η τρίτη στήλη δίνουν τη διορθωμένη εικόνα με τη χρήση του N3MNI και το αντίστοιχο πεδίο μεροληψίας, ενώ τα αποτελέσματα με τη χρήση του N4ITK παρέχονται στις στήλες 4 και 5 [15].

Rescaling

Η ένταση των εικόνων έγινε επανακλιμάκωση όπως αναφέρεται παρακάτω:

$$(x - x_2)/(x_{98} - x)$$

όπου x_2 και x_{98} είναι το 2ο και 98ο εκατοστημόριο έντασης της κάθε εικόνας και x η εικόνα. Η διαδικασία της επανακλιμάκωσης μας βοηθάει στο να διατηρούμε πληροφορία και

να έχουμε παρόμοια επίπεδα έντασης στις εικόνες, ειδικά στις MRI εικόνες, όπου η ένταση διαθέτει μεγάλο κομμάτι της πληροφορίας.

Resizing

Ένα σύνηθες βήμα στην προεπεξεργασία των εικόνων είναι η αλλαγή μεγέθους. Η αλλαγή μεγέθους συνηθίζεται στις εφαρμογές βαθιάς μάθησης που αφορούν εικόνες, καθώς μας βοηθάει στο να φέρουμε όλες τις εικόνες σε ένα σταθερό, κοινό μέγεθος και στο να μειώσουμε το κόστος της εκπαίδευσης. Παράλληλα, αποδεικνύεται πως έχουμε καλύτερη γενίκευση σε μοντέλα εικόνων, όταν τα δεδομένα έχουν μικρή ανάλυση. Στην συγκεκριμένη εφαρμογή η αλλαγή μεγέθους έγινε σε τρισδιάστατο επίπεδο, μεταβάλλοντας τις αποστάσεις των ογκοστοιχείων και ύστερα περικοπή στις δισδιάστατες εικόνες για να έχουμε σταθερό μέγεθος. Για τις εικόνες προσάτη, οι εικόνες απέκτησαν μέγεθος $0.627 \times 0.627 \times 3.6 \text{mm}^3$ στον τρισδιάστατο χώρο και ύστερα έγινε περικοπή σε μέγεθος 224×224 . Για τις καρδιακές εικόνες, στον τρισδιάστατο χώρο έχουμε μέγεθος $1.367 \times 1.367 \text{mm}^2$ και ύστερα περικοπή σε μέγεθος 192×192 στο κέντρο της εικόνας, καθώς η καρδιά συνηθίζει να βρίσκεται στο κέντρο μιας MRI εικόνας.

Γεωμετρικές και Φωτομετρικές Μετατροπές

Τέλος, στις εικόνες εφαρμόστηκαν φωτομετρικά φίλτρα για τον περιορισμό ορισμένων συχνοτήτων και την διατήρηση των σημαντικών συχνοτήτων για τις μετρήσεις κατά την εκπαίδευση του δικτύου, καθώς και γεωμετρικά φίλτρα για την αλλαγή του σχήματος και της κατεύθυνσης των εικόνων, ώστε να βρίσκονται στην απαραίτητη μορφή για χρήση.

Κεφάλαιο 5

Υλοποίηση

Στο κεφάλαιο αυτό περιγράφονται οι κύριες μέθοδοι που χρησιμοποιήθηκαν στα πλαίσια της διπλωματικής εργασίας.

5.1 Επαύξηση Δεδομένων

Η επαύξηση δεδομένων αποτελεί την δημιουργία νέων δειγμάτων, αλλοιώνοντας τα δεδομένα με τρόπο που είναι διαφορετικά από τα αρχικά, αλλά διατηρούν την πληροφορία που περιέχουν [118]. Η μέθοδος είναι αρκετά συνηθισμένη, ειδικά στον τομέα της επεξεργασίας εικόνας, και βελτιώνει κατά μεγάλο το βαθμό την απόδοση των μοντέλων, εφόσον το αρχικό σύνολο δεδομένων διαθέτει αντιπροσωπευτικά δείγματα για την επιθυμητή εφαρμογή και αρκετή πληροφορία για να ολοκληρωθεί η διαδικασία της εκμάθησης. Στα πλαίσια των πειραμάτων έγινε χρήση μεθόδων επαύξησης των δειγμάτων στην είσοδο και μέθοδοι επαύξησης των χαρακτηριστικών των επιπέδων εντός της αρχιτεκτονικής του μοντέλου.

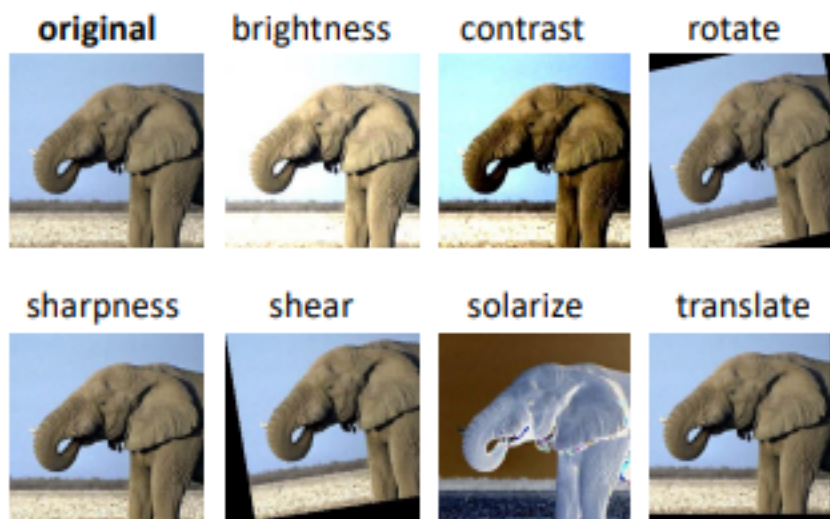
5.1.1 Μέθοδοι επαύξησης σε επίπεδο εισόδου

Οι πιο συνηθισμένες μέθοδοι στην επαύξηση δεδομένων αποτελούν απλές αλλαγές στην εικόνα και την ετικέτα πριν την είσοδο της στο μοντέλο. Ο κύριος σκοπός είναι ο εμπλουτισμός του συνόλου δεδομένων από παραλλαγές των ήδη υπάρχοντων εικόνων, έτσι ώστε το μοντέλο να εξαγάγει την σωστή πληροφορία για την λήψη αποφάσεων και να γίνεται πιο στιβαρό [16]. Οι μέθοδοι αυτές έχουν μεγάλη ποικιλία και το εύρος τους διαθέτει από αρκετά απλές έως αρκετά περίπλοκες επαυξήσεις.

Αυτές οι μέθοδοι είναι εύκολες στην χρήση και υλοποίηση, όμως δεν συνηθίζεται να οδηγούν σε μεγάλες βελτιώσεις της γενίκευσης ενός μοντέλο για περίπλοκες αρχιτεκτονικές και εφαρμογές και για αυτό επιλέγονται συνήθως ή πιο αναβαθμισμένες μέθοδοι επαύξησης ή μέθοδοι επαύξησης στο επίπεδο των χαρακτηριστικών.

PixMix

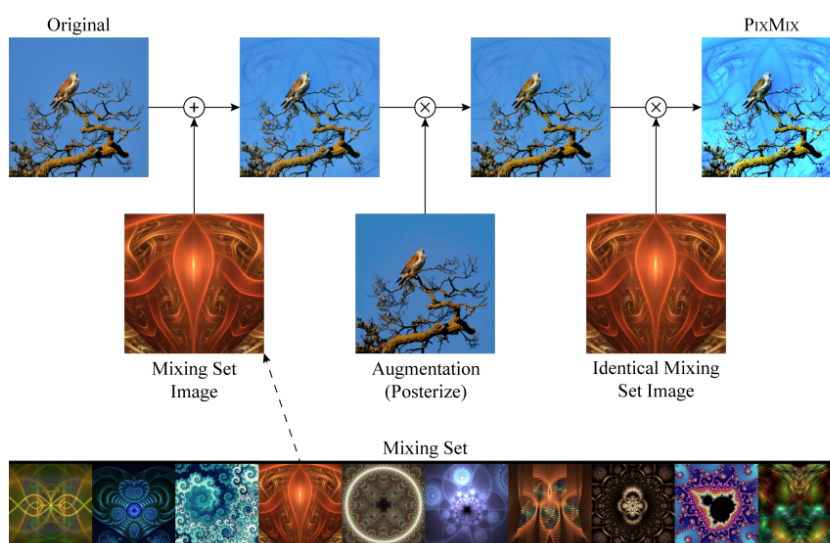
Στα πλαίσια της διπλωματικής εργασίας διερευνήθηκε η επίδραση μιας πιο περίπλοκης επαύξησης δεδομένων σε εφαρμογές κατάτμησης εικόνων. Η μεθοδολογία PixMix λειτουργεί αναμειγνύοντας τις εικόνες εισόδου με εικόνες φρακταλ, με σκοπό την εισαγωγή περίπλοκων εικόνων [20].



Σχήμα 5.1: Παραδείγματα τυπικών μεθόδων επαύξησης δεδομένων στο επίπεδο της εισόδου [16], [17], [18], [19]

Οι συγγραφείς υποστηρίζουν ότι οι απλές μέθοδοι επαύξησης δεδομένων αποφεύγουν να δημιουργήσουν δομικά περίπλοκες εικόνες. Παράλληλα, παρατηρείται συνήθως ότι η βελτιστοποίηση μιας εφαρμογής με μια συγκεκριμένη μέθοδο για κάποια μετρικής ασφαλείας συνήθως έρχεται εις βάρος άλλων μετρικών [20]. Για αυτόν τον λόγο, πειραματίστηκαν με την εισαγωγή δομικής πολυπλοκότητας μέσω ανάμειξης με εικόνες φρακταλ, ώστε να αποφανθούν εάν είναι εφικτό να υπάρξει μέθοδος επαύξησης δεδομένων που βελτιώνει πολλαπλές μετρικές.

Η ανάμειξη έγινε εκτελώντας αριθμητικές πράξεις μεταξύ των εικόνων εισόδου και ανάμειξης, αναφορικά προσθετικές και πολλαπλασιαστικές με τον γεωμετρικό μέσο, πολλαπλές φορές και σε συγκεκριμένη ένταση που καθορίζεται από εκθετικά βάρη και μια βήτα υπερπαράμετρο.



Σχήμα 5.2: Η διαδικασία της επαύξησης δεδομένων PixMix. [20]

Τα πειράματα ήταν θετικά και αποδείχτηκε ότι η επαύξηση δεδομένων έχει καλά αποτελέσματα σε πολλαπλές μετρικές ασφαλείας, καθώς και στην OOD απόδοση των μοντέλων [20]. Η εισαγωγή, λοιπόν, εσωτερικής πολυπλοκότητας βελτιώνει την απόδοση των μοντέλων και η διαφοροποίηση που παρέχουν οι εικόνες φράκταλ βελτιώνει την απόδοση των μετρικών, χωρίς να είναι εις βάρος άλλων μετρικών.

5.1.2 Μέθοδοι επαύξησης σε επίπεδο χαρακτηριστικών

Πρόσφατες έρευνες έχουν δείξει ότι οι μέθοδοι επαύξησης δεδομένων στο επίπεδο της εισόδου εν τέλει δεν οδηγούν τα μοντέλα βαθιάς μηχανικής μάθησης σε καλή γενίκευση [23], [119]. Οι σύγχρονες μέθοδοι για να καταπολεμήσουν αυτό το πρόβλημα υλοποιούν επαύξηση δεδομένων στα χαρακτηριστικά που παράγονται από τα ενδιάμεσα επίπεδα ενός μοντέλου, με σκοπό να βελτιωθεί η στιβαρότητα και η ικανότητα γενίκευσης. Στα πρόβληματα κατάτμησης εικόνας, όμως, δεν γινόταν χρήση τέτοιων μεθόδων επαύξησης, καθώς υπάρχει ο κίνδυνος να αλλοιωθεί το σημασιολογικό περιεχόμενο της εικόνας και να μην εκπαιδευτεί σωστά το μοντέλο [119]. Τα τελευταία χρόνια έχουν υλοποιηθεί μέθοδοι οι οποίες επιτυχώς αλλάζουν τα χαρακτηριστικά με τρόπο τέτοιο ώστε να διατηρείται το σημασιολογικό περιεχόμενο.

MixStyle

Η μέθοδος MixStyle είναι η πρώτη από αυτές τις μεθόδους που διερευνήθηκε στα πλαίσια της διπλωματικής εργασίας. Η μέθοδος εκτελεί επαύξηση δεδομένων διατηρώντας το σημασιολογικό περιεχόμενο των εικόνων, αναμειγνύοντας τα διαφορετικά οπτικά στυλ στο επίπεδο των χαρακτηριστικών [21]. Πρόσφατες έρευνες πάνω στην μεταφορά στυλ έχουν δείξει πως πληροφορίες για την μορφή μιας εικόνας διατηρούνται στα αρχικά επίπεδα ενός συνελκτικού νευρωνικού δικτύου στην μορφή των στατιστικών των χαρακτηριστικών [21]. Οι συγγραφείς υποστηρίζουν η ανάμειξη των στατιστικών διαφορετικών εικόνων θα οδηγήσει στην δημιουργία νέων στυλ, αυξάνοντας, έτσι, την ποικιλομορφία των δεδομένων κατά την εκπαίδευση, χωρίς την δημιουργία νέων εικόνων, αλλά με την αλλαγή των χαρακτηριστικών [21].

Η μέθοδος MixStyle εφαρμόζεται με λειτουργία plug-and-play και τοποθετείται ανάμεσα στα επίπεδα ενός δικτύου, αξιοποιώντας συνήθως τεχνικές mini-batching. Οι παρτίδες εισόδου δημιουργούνται είτε αναμειγνύοντας δείγματα από διαφορετικά πεδία όταν γνωρίζουμε το που κατατάσσονται είτε κάνοντας τυχαία δειγματοληψία από το σύνολο εκπαίδευσης [21]. Οι εικόνες δεν χρειάζεται απαραίτητα να είναι από διαφορετικές κατανομές, καθώς ακόμα και μέσα σε ένα πεδίο υπάρχουν υποπεδία οπότε και με δείγματα από το ίδιο πεδίο μπορούν να δημιουργηθούν καινούργια. Για κάθε παρτίδα x δημιουργείται μια παρτίδα αναφοράς \hat{x} , όπου η παρτίδα αναφοράς λαμβάνεται κάνοντας ανακάτεμα στην διάσταση της παρτίδας x τυχαία είτε ανάλογα με το πεδίο. Μετά το ανακάτεμα υπολογίζονται τα καινούργια στατιστικά των χαρακτηριστικών ως εξής:

$$\gamma_{mix} = \beta\sigma(x) + (1 - \beta)\sigma(\hat{x}),$$

$$\beta_{mix} = \hat{\beta}\mu(x) + (1 - \hat{\beta})\mu(\hat{x})$$

όπου $\sigma(x)$, $\mu(x)$ η τυπική απόκλιση και η μέση τιμή της παρτίδας αντίστοιχα, $\hat{\beta} \in \mathbb{R}$ είναι βάρη για κάθε ξεχωριστή περίπτωση που λαμβάνονται από μία κατανομή βήτα, δηλαδή $\hat{\beta} \sim \text{Beta}(a, a)$, όπου $a \in (0, \infty)$ και a υπερπαραμέτρος [21]. Αφού υπολογιστούν τα νέα στατιστικά, δημιουργείται καινούργιο στυλ βάσει της παρακάτω εξίσωσης:

$$\text{MixStyle}(x) = \gamma_{mix} \frac{x - \mu(x)}{\sigma(x)} + \beta_{mix}$$

$$\begin{aligned} \mathbf{x} &= [\textcircled{x_1} \textcircled{x_2} \textcircled{x_3} \textcircled{x_4} \textcircled{x_5} \textcircled{x_6}] \\ \tilde{\mathbf{x}} &= [\textcircled{x_6} \textcircled{x_1} \textcircled{x_5} \textcircled{x_3} \textcircled{x_2} \textcircled{x_4}] \end{aligned}$$

Σχήμα 5.3: Η μορφή της ανάμειξης εικόνων μέσα σε μία παρτίδα. Ακόμα και εικόνες της ίδιας παρτίδας μπορούν να οδηγήσουν σε καινούργια στυλ με αρκετές επαναλήψεις [21]

DSU

Η μέθοδος DSU (Domain Shifts with Uncertainty) αποτελεί μία πρόσφατη μέθοδο επαύξησης δεδομένων, η οποία ασχολείται κυρίως με την αντιμετώπιση της αβεβαιότητας που υπάρχει στα στατιστικά των χαρακτηριστικών [22]. Οι συγγραφείς υποστηρίζουν ότι οι περισσότερες σύγχρονες μέθοδοι δεν λαμβάνουν υπόψιν τις τυχαίες αποκλίσεις που υπάρχουν στα πεδία των χαρακτηριστικών, με αποτέλεσμα να υπάρχει υπερπροσαρμογή στο πεδίο της εκπαίδευσης [22]. Η μοντελοποίηση αυτής της αβεβαιότητας θα οδηγούσε στην μεγαλύτερη ποικιλομορφία των πεδίων εκπαίδευσης και την αύξηση σταθερότητας των μοντέλων βαθιάς μάθησης.

Η μέθοδος DSU, λοιπόν, μοντελοποιεί τα στατιστικά των χαρακτηριστικών ως πολυδιάστατες Γκαουσιανές κατανομές, δηλαδή με την υπόθεση ότι η μέση τιμή και η τυπική απόκλιση ακολουθούν κατανομές του τύπου $N(\mu, \Sigma_\mu^2)$ και $N(\sigma, \Sigma_\sigma^2)$. Αποτελεί plug-and-play μέθοδο, όπως και το MixStyle, και τοποθετείται ενδιάμεσου των επιπέδων ενός δικτύου [22]. Το κέντρο της κατανομής τοποθετείται στην αρχική τιμή και με απόκλιση δημιουργούνται οι στατιστικές αποκλίσεις. Η τυπική απόκλιση των κατανομών και, κατέπекταση, η αβεβαιότητα των στατιστικών υπολογίζεται ως εξής:

$$\Sigma_\mu^2 = \frac{1}{B} \sum_{b=1}^B (\mu(x) - \mathbb{E}[\mu(x)])^2,$$

$$\Sigma_\sigma^2 = \frac{1}{B} \sum_{b=1}^B (\sigma(x) - \mathbb{E}[\sigma(x)])^2$$

Ο υπολογισμός βασίζεται σε έρευνες που υποστηρίζουν ότι οι αποκλίσεις μεταξύ των χαρακτηριστικών περιέχουν έμμεση σημασιολογική σημασία και ότι οι κατευθύνσεις με μεγαλύτερες

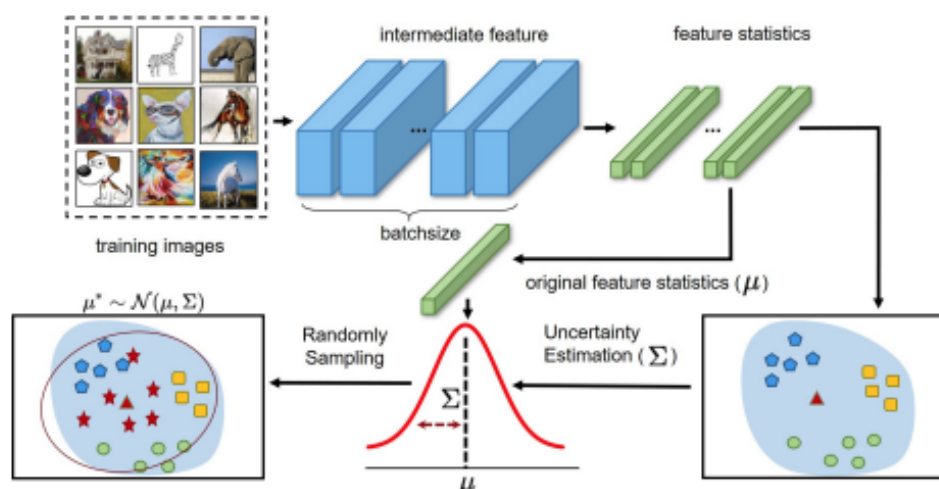
αποκλίσεις μπορούν να υποδηλώνουν δυνατότητες πιο πολύτιμων σημασιολογικών αλλαγών [22]. Αφού υπολογιστούν οι κατανομές, υπολογίζονται η καινούργια μέση τιμή και τυπική απόκλιση, ως εξής:

$$\gamma_{mix} = \sigma(x) + \epsilon_{\sigma} \Sigma_{\sigma}, \epsilon_{\sigma} \sim N(0, 1),$$

$$\beta_{mix} = \mu(x) + \epsilon_{\mu} \Sigma_{\mu}, \epsilon_{\mu} \sim N(0, 1).$$

Όπως και στην μέθοδο MixStyle, η τελική επαύξηση υπολογίζεται με την μετατροπή των χαρακτηριστικών:

$$DSU(x) = \gamma_{mix} \frac{x - \mu(x)}{\sigma(x)} + \beta_{mix}$$



Σχήμα 5.4: Η διαδικασία επαύξησης με χρήση μοντελοποίησης των στατιστικών ως πολλαδιάστατες Γκαουσιανές κατανομές [22]

MaxStyle

Η μέθοδος MaxStyle αποτελεί την πιο σύγχρονη μέθοδο που δοκιμάστηκε στα πλαίσια της διπλωματικής εργασίας, η οποία χρησιμοποιεί μια βοηθητική αρχιτεκτονική αποκωδικοποιητή για να παράγει ανταγωνιστικά καινούργια στυλ εικόνων [23]. Αποτελεί βελτίωση του MixStyle και σκοπεύει στο να επιλύσει ορισμένα προβλήματα που παρουσιάζει, τα οποία σύμφωνα με τους συγγραφείς είναι τα εξής:

- Δυσκολίες στην οπτικοποίηση και ερμηνεία των επαυξημένων χαρακτηριστικών σε χώρους υψηλών διαστάσεων
- Περιορισμένη ποικιλία των οπτικών στυλ λόγω χρήσης γραμμικής παρεμβολής
- Υποβέλτιστη επίδραση στην στιβαρότητα του μοντέλου καθώς δεν λαμβάνεται υπόψη η ευαισθησία των μοντέλων σε προβλήματα υπερπροσαρμογής

Η κύρια ιδέα της μεθόδου είναι η χρήση ενός υποβοηθητικού αποκωδικοποιητή σε συνδυασμό με την αρχιτεκτονική κωδικοποιητή-αποκωδικοποιητή, πάνω στον οποίο εφαρμόζεται η

επαύξηση δεδομένων MixStyle και ο σκοπός του είναι να παράξει τις επαυξημένες εικόνες και να αναπαράγει την αρχική εικόνα [23]. Το σφάλμα αναπαραγωγής τροφοδοτείται στον αρχικό κωδικοποιητή και οδηγεί σε βελτιωμένη σιθαρότητα για OOD δεδομένα. Σε συνδυασμό με αυτό τον αποκωδικοποιητή, η μέθοδος MaxStyle προσθέτει θόρυβο για να μεγαλώσει τον χώρο των οπτικών στυλ που δημιουργούνται και αξιοποιεί ανταγωνιστική εκπαίδευση για την εύρεση βέλτιστων γραμμικών συντελεστών για την επαύξηση δεδομένων και βέλτιστων συντελεστών για την πρόσθεση του θορύβου [23]. Δεδομένου ότι f είναι τα χαρακτηριστικά σε ένα επίπεδο του ενός συνελκτικού νευρωνικού δικτύου στον αποκωδικοποιητή D_ϕ , με την εικόνα x ως είσοδο, η μέθοδος MaxStyle εκτελεί επαύξηση δεδομένων ως εξής:

$$\text{MaxStyle}(f) = (\gamma_{mix} + \Sigma_\gamma \epsilon_\gamma) \odot f_{normalized} + (\beta_{mix} + \Sigma_\beta \epsilon_\beta),$$

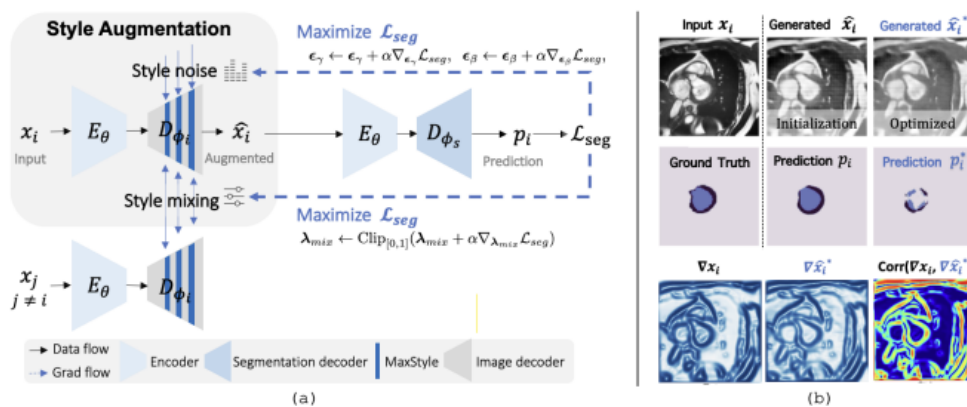
όπου $\gamma_{mix}, \beta_{mix}$ τα στυλ ανάμειξης και $\Sigma_\gamma \epsilon_\gamma, \Sigma_\beta \epsilon_\beta$ ο θόρυβος που προστίθεται για προσομοίωση των μεταβάσεων πεδίων. Για να βρίσκεται ο θόρυβος σε λογικά πλαίσια, δειγματοληπτείται από μία Γκαουσιανή κατανομή με παράγοντες Σ_β και Σ_γ , δηλαδή διακυμάνσεις που λαμβάνονται από τα στατιστικά μίας παρτίδας i . Οι διακυμάνσεις λαμβάνονται ως εξής:

$$\Sigma_\gamma = \sigma^2(\{\sigma(f_i)\}), \Sigma_\beta = \sigma^2(\{\mu(f_i)\}), \epsilon_\gamma, \epsilon_\beta \sim \mathcal{N}(0, 1)$$

Το ανταγωνιστικό κομμάτι της εκπαίδευσης γίνεται μεταβάλλοντας τους παράγοντες $\epsilon_\gamma, \epsilon_\beta, \beta_{mix}$, μεγιστοποιώντας το σφάλμα κατάτμησης με κλίση ανόδου, έτσι ώστε η συνθετική εικόνα να παράξει λανθασμένη πρόβλεψη. Η κλίση ανόδου γίνεται ως εξής:

$$\epsilon_\gamma \leftarrow \epsilon_\gamma + a \nabla_{\epsilon_\gamma} L_{seg}(\hat{p}, y), \epsilon_\beta \leftarrow \epsilon_\beta + a \nabla_{\epsilon_\beta} L_{seg}(\hat{p}, y), \text{Clip}_{[0,1]}(\beta_{mix} \leftarrow \beta_{mix} + a \nabla_{\beta_{mix}} L_{seg}(\hat{p}, y)),$$

ανανεώνοντας τις τιμές για τον θόρυβο και τον παράγοντα ανάμειξης που χρησιμοποιείται από την μέθοδο MixStyle. Για την εκπαίδευση χρησιμοποιείται και η αρχική εικόνα και η επαυξημένη, με σκοπό να ελαχιστοποιηθούν τα σφάλματα κατάτμησης και ανακατασκευής εικόνας [23]. Έτσι, οδηγούμαστε σε μία πιο σιθαρή εκπαίδευση και σε καλύτερο αποτέλεσμα απόδοσης σε άγνωστα πεδία δεδομένων.



Σχήμα 5.5: Αριστερά: Η αρχιτεκτονική MaxStyle. Ο υποβοηθητικός αποκωδικοποιητής αλληλάζει τις παραμέτρους MixStyle στο εσωτερικό της κύριας αρχιτεκτονικής, οδηγώντας σε καινούργια στυλ. Δεξιά: Εικόνα Εισόδου και Επαυξημένη Εικόνα. Από την συσχέτιση των παραγώγων παρατηρούμε την διατήρηση του σημασιολογικού περιεχομένου [23].

Κεφάλαιο 6

Παρουσίαση και Σχολιασμός Πειραματικών Αποτελεσμάτων

6.1 Μετρικές Αξιολόγησης

Για τον υπολογισμό των μετρικών έγινε χρήση του μέσου Sørensen–Dice coefficient, το οποίο μας δείχνει το πόσο δύο εικόνες μοιάζουν ανά εικονοστοχείο η μία με την άλλη, πάνω σε κάθε σύνολο δεδομένων [120]. Το dice score για δύο σύνολα X, Y υπολογίζεται ως εξής:

$$\text{Dice score} = \frac{2 \cdot |X \cap Y|}{|X| + |Y|} \quad (6.1)$$

που για εικόνες η τομή είναι τα κοινά εικονοστοιχεία.

6.2 Μεθοδολογία Ελέγχου

Για την χρήση του εργαλείου μπορούμε να ελέγξουμε τις εξής υπερπαραμέτρους:

- **Severity:** Η ένταση με την οποία θα εκτελεστούν οι αναμειξεις που έχουν επιλεγθεί,
- **Probability:** Η πιθανότητα με την οποία επιλέγεται να γίνει augment η εικόνα ή να γίνει μίξη με κάποια εικόνα από το σύνολο εικόνων ανάμειξης ή με την ίδια την επαυξημένη εικόνα,
- **Rounds:** Το πλήθος των αναμειξεων που θα εφαρμοστούν σε μία εικόνα.

Καθώς το PixMix χρησιμοποιείται κυρίως για classification tasks, οι προεπιλεγμένες υπερπαραμέτροι είναι αρκετά υπερσυντονισμένες για μια εργασία τμηματοποίησης, οπότε για την συγκεκριμένη εφαρμογή θέσαμε την πιθανότητα στο 0.5, την ένταση στο 0.5, και το πλήθος σε ένα με τρία, επιλεγμένο τυχαία. Παράλληλα, παρατηρήθηκε πως η υψηλού βαθμού εφαρμογή του συστήματος οδηγούσε σε πτώση της απόδοσης, λόγω του ότι ο βοηθητικός κωδικοποιητής δεν μπορούσε να αντιληφθεί την σωστή πληροφορία. Οπότε, για να αυξηθεί η σιβαρότητα του δικτύου, καθώς και η ποικιλία των δεδομένων κατά την εκπαίδευση, η συνολική διαδικασία PixMix εκτελούνταν για μία παρτίδα με πιθανότητα 0.3, ενώ με πιθανότητα 0.7 οι εικόνες εισάγονται στο δίκτυο με απλή διαταραχή θορύβου στυλ για αποφυγή της υπερπροσαρμογής.

6.3 Αποτελέσματα στα δεδομένα προστάτη

Η εκπαίδευση έγινε για 600 εποχές με μέγεθος παρτίδας ίσο με 20 για την αρχιτεκτονική FCN-16 και ίσο με 15 για την αρχιτεκτονική FCN-64. Στα παρακάτω πινακάκια παρουσιάζουμε τα αποτελέσματα των πειραμάτων για τις πολλαπλές μεθόδους επαύξησις δεδομένων στις δύο αρχιτεκτονικές. Με μαύρα γράμματα εμφανίζεται η μέθοδος με καλύτερη απόδοση σε κάθε σύνολο δεδομένων:

Datasets	OOD						IID
	A	B	C	D	E	F	G
Baseline	0.8795	0.7552	0.7766	0.7678	0.2952	0.242	0.838
MixStyle	0.875	0.6934	0.8002	0.7335	0.4313	0.4076	0.8449
DSU	0.8821	0.5438	0.8154	0.8405	0.5856	0.6403	0.8311
MaxStyle	0.8966	0.8186	0.8219	0.8337	0.528	0.7145	0.8549
MaxStyle w/PixMix	0.8958	0.8387	0.8239	0.8542	0.6461	0.6455	0.8636

Πίνακας 6.1: Πίνακας σύγκρισης μεθόδων για adversarial network στα δεδομένα προστάτη σε αρχιτεκτονική FCN-16

Datasets	OOD						IID
	A	B	C	D	E	F	G
Baseline	0.9035	0.6785	0.7929	0.7914	0.4574	0.6085	0.8363
MixStyle	0.9108	0.6889	0.8258	0.8243	0.438	0.5538	0.8725
MaxStyle	0.905	0.8039	0.8448	0.8603	0.4577	0.5599	0.8803
MaxStyle w/PixMix	0.9072	0.8619	0.8567	0.8847	0.7523	0.7723	0.8635

Πίνακας 6.2: Πίνακας σύγκρισης μεθόδων για adversarial network στα δεδομένα προστάτη σε αρχιτεκτονική FCN-64

Και για τις δύο αρχιτεκτονικές η πρόσθεση της μεθόδου PixMix είχε θετικά αποτελέσματα, ξεπερνώντας το προηγούμενο μοντέλο. Επίσης, είχε μεγάλη αύξηση της τάξης του 10% για την OOD απόδοση στην αρχιτεκτονική FCN-64. Από τις δύο αρχιτεκτονικές την καλύτερη απόδοση παρουσίασε η FCN-64, παρά το γεγονός ότι οι αρχικές μέθοδοι δεν βελτίωσαν πολύ το μέσο dice score.

6.4 Αποτελέσματα στα δεδομένα καρδιάς

Η εκπαίδευση έγινε για 600 εποχές με μέγεθος παρτίδας ίσο με 15 για την αρχιτεκτονική FCN-16. Στα παρακάτω πινακάκια παρουσιάζουμε τα αποτελέσματα των πειραμάτων για τις πολλαπλές μεθόδους επαύξησις δεδομένων. Με μαύρα γράμματα εμφανίζεται η μέθοδος με καλύτερη απόδοση σε κάθε σύνολο δεδομένων. Η απόδοση στην συγκεκριμένη εφαρμογή αφορά την στιβαρότητα του μοντέλου προς τα διάφορα τεχνουργήματα που αναφέραμε σε προηγούμενο κεφάλαιο.

Datasets	Corruption			
	RandomBias	RandomSpike	RandomMotion	RandomGhosting
Baseline	0.9199	0.01	0.9158	0.9111
MaxStyle	0.9151	0.5792	0.9183	0.91
MaxStyle w/PixMix	0.9189	0.6251	0.9161	0.9088

Πίνακας 6.3: Πίνακας σύγκρισης μεθόδων για adversarial network στα δεδομένα καρδιάς σε αρχιτεκτονική FCN-16 για την αριστερή κοιλία

Datasets	Corruption			
	RandomBias	RandomSpike	RandomMotion	RandomGhosting
Baseline	0.855	0.01	0.8359	0.8373
MaxStyle	0.8492	0.4835	0.8418	0.8364
MaxStyle w/PixMix	0.8482	0.495	0.8388	0.8336

Πίνακας 6.4: Πίνακας σύγκρισης μεθόδων για adversarial network στα δεδομένα καρδιάς σε αρχιτεκτονική FCN-16 για το μυοκάρδιο

Datasets	Corruption			
	RandomBias	RandomSpike	RandomMotion	RandomGhosting
Baseline	0.8317	0.01	0.8279	0.8162
MaxStyle	0.8319	0.4492	0.8387	0.8266
MaxStyle w/PixMix	0.8342	0.5098	0.8297	0.8027

Πίνακας 6.5: Πίνακας σύγκρισης μεθόδων για adversarial network στα δεδομένα καρδιάς σε αρχιτεκτονική FCN-16 για την δεξιά κοιλία

Παρατηρούμε ότι η κύρια βελτίωση είναι στο RandomSpike σύνολο δεδομένων, ενώ τα υπόλοιπα τεχνουργήματα παραμένουν σε απόδοση ίδιας τάξης.

Μέρος 

Επίλογος

Κεφάλαιο **7**

Επίλογος

7.1 Συμπεράσματα

Σε μεθόδους όπου τα δεδομένα είναι λιγοστά, η ανταγωνιστική μάθηση συμβάλλει σε μεγάλο βαθμό στην αποφυγή της υπερπροσαρμογής του μοντέλου και στην καλύτερη γενίκευση χρησιμοποιώντας ανταγωνιστικά παραδείγματα για να υπολειτουργήσει το μοντέλο σε εικόνες που παρουσιάζουν την ίδια σημασιολογική πληροφορία. Όμως, διαθέτει το πρόβλημα ότι είναι χρονοβόρα και κοστοβόρα, καθώς χρειάζεται εκπαίδευση πολλών εποχών για να συγκλίνει τα ανταγωνιστικά συλ των παραγόμενων εικόνων και επειδή το επιπρόσθετο σύστημα καταναλώνει αρκετή μνήμη.

Παράλληλα, η επίδραση των μεθόδων επαύξησης δεδομένων στο επίπεδο των χαρακτηριστικών οδηγούν σε μεγάλη βελτίωση, καθώς επιτρέπουν την άμεση επίδραση πάνω στο συλ μιας εικόνας από τα μεγέθη με τα οποία αυτό αντιλαμβάνεται την πληροφορία. Η πρόσθηκη περίπλοκων επαυξήσεων δεδομένων μπορεί να οδηγήσει σε καλύτερα αποτελέσματα, καθώς εκμεταλλεύεται ταυτόχρονα την εκμάθηση σε διαφορετικά πεδία, μέσω των ίδιων των μετασχηματισμών που εκτελούνται πάνω στις εικόνες, και την δημιουργία ανταγωνιστικών παραδειγμάτων από τις ίδιες τις επαυξήσεις μέσω του βοηθητικού αποκωδικοποιητή. Έτσι, η επαύξηση μπορεί ταυτόχρονα να επωφεληθεί σε μεγάλο βαθμό το δίκτυο που εκτελεί την κεντρική εφαρμογή και το σύστημα ανταγωνιστικής μάθησης στο να αλλάξει τις παραμέτρους για πιο ανταγωνιστικά συλ.

Όμως, η μεγάλη εφαρμογή της επαύξησης μπορεί να οδηγήσει σε προβλήματα εκπαίδευσης και στην παραγωγή παραδειγμάτων που δεν διατηρούν το σημασιολογικό περιεχόμενο, γεγονός που θέλουμε να αποφύγουμε. Για αυτό, εφαρμόζοντας την με έναν μετριασμένο τρόπο, ώστε να μην επηρεάσει σε μεγάλο βαθμό το ανταγωνιστικό κομμάτι της εκπαίδευσης, μπορεί ένα σύστημα να επωφεληθεί από την δημιουργία καλύτερων, πιο περίπλοκων επαυξημένων εικόνων, χωρίς κόστος στον χρόνο εκπαίδευσης και στους πόρους που χρειάζονται και να οδηγηθεί σε καλύτερη γενίκευση. Αυτός είναι ο απώτερος σκοπός για να μπορούν τέτοια εργαλεία να εφαρμοστούν σε πρακτικό επίπεδο.

7.2 Μελλοντικές Επεκτάσεις

Το σύστημα που αναπτύχθηκε στα πλαίσια αυτής της διπλωματικής εργασίας θα μπορούσε να βελτιωθεί και να επεκταθεί περαιτέρω, τουλάχιστον ως προς τρεις κατευθύνσεις.

Συγκεκριμένα, αναφέρονται τα ακόλουθα :

- Χρήση της επαύξησης σε πιο περίπλοκες αρχιτεκτονικές. Το συνολικό μοντέλο επαύξησης μπορεί να λειτουργήσει ως βοηθητικό σε κάποια αρχιτεκτονική που δεν ενσωματώνει με τέτοιο ευκολία τον βοηθητικό κωδικοποιητή. Για παράδειγμα, θα μπορούσε να λειτουργήσει ως εργαλείο επαύξησης ενός συνόλου δεδομένων με το οποίο θα εκπαιδευτεί ένας Vision Transformer. Στόχος εδώ είναι να εκπαιδευσουμε το σύστημα παραγωγής επαυξημένων εικόνων όσον το δυνατό καλύτερο σε ένα υποσύνολο του κεντρικού, και να το χρησιμοποιήσουμε για να επαυξήσουμε την διαδικασία προσαρμογής μιας αρχιτεκτονικής που διαθέτει μεγαλύτερες δυνατότητες επεξεργασίας εικόνων.
- Χρήση πολλαπλών μεθόδων επαύξησης δεδομένων ταυτόχρονα. Αξίζει να εξεταστεί το πως ο συνδυασμός πολλών διαφορετικών, σύγχρονων μεθόδων μπορεί να οδηγήσει στην προσέγγιση περισσότερων άγνωστων πεδίων, για να είναι όσον το δυνατόν καλύτερη η γενίκευση του μοντέλου.
- Αξιολόγηση του συστήματος σε διαφορετική εφαρμογή. Θα ήταν φρόνιμο να ελεγχθούν τα συμπεράσματα της εργασίας σε ένα μεγαλύτερης κλίμακας σύνολο δεδομένων, όπως σύνολα δεδομένων αυτόνομης οδήγησης ή σε γενικές εφαρμογές ταξινόμησης, για να διερευνηθούν καλύτερα οι δυνατότητες της αρχιτεκτονικής.
- Εφαρμογή του συστήματος σε ρεαλιστικές συνθήκες. Σημαντικό κομμάτι της διαδικασίας δημιουργίας ενός πρακτικού εργαλείου είναι η συνεργασία με τους ειδικούς. Ιδανικό θα ήταν να υπάρξει μία συνεργασία με γιατρούς και να διερευνηθεί το κατά πόσο το εργαλείο στην μορφή που βρίσκεται τώρα είναι χρήσιμο και τι παραπάνω θα χρειαζόταν για να είναι εφαρμόσιμο.

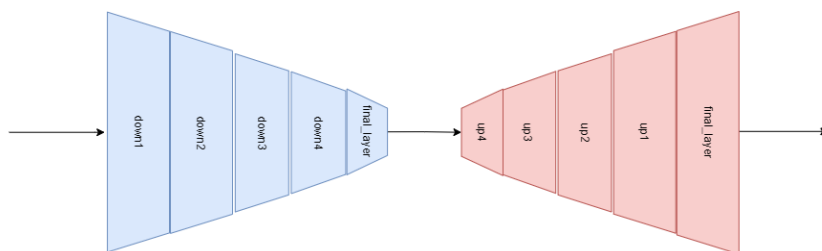
Παραρτήματα

Διαγράμματα Μοντέλων

Στο κεφάλαιο αυτό παρουσιάζονται με μεγαλύτερη λεπτομέρεια σε διαγραμματική μορφή οι αρχιτεκτονικές που χρησιμοποιήθηκαν στο πειραματικό μέρος.

A'.1 Βασική Αρχιτεκτονική

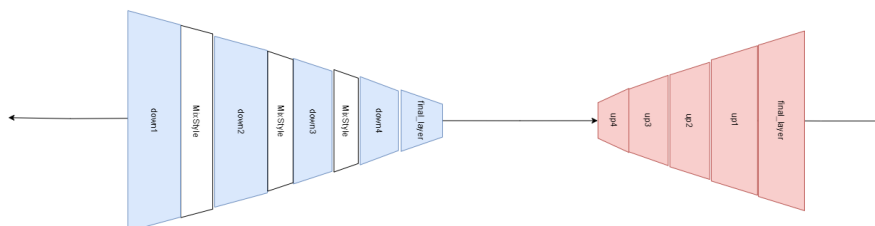
Η βασική αρχιτεκτονική κωδικοποιητή-αποκωδικοποιητή. Οι δύο αρχιτεκτονικές (FCN-16, FCN-64) διαφοροποιούνται στο πλήθος των χαρακτηριστικών. Η δεύτερη αρχιτεκτονική διαθέτει τέσσερις φορές παραπάνω παραμέτρους για την ανάλυση. Κάθε μέρος διαθέτει πέντε επίπεδα.



Σχήμα A'.1: Βασική αρχιτεκτονική κωδικοποιητή-αποκωδικοποιητή

A'.2 MixStyle

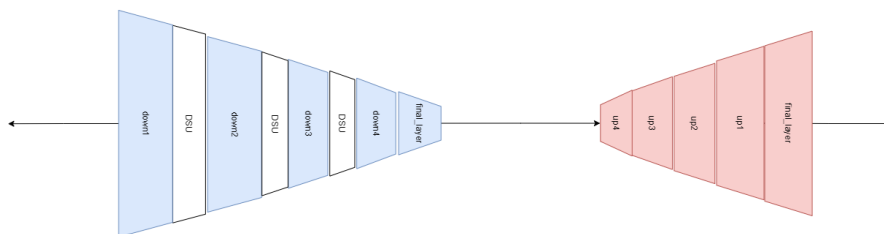
Η αρχιτεκτονική MixStyle. Τα blocks MixStyle εισάγονται ενδιάμεσα των τριών πρώτων επιπέδων, καθώς εκεί βρίσκεται η πληροφορία για το στυλ των εικόνων και για τα πεδία.



Σχήμα A'.2: Αρχιτεκτονική με εισαγωγή MixStyle blocks στα αρχικά επίπεδα για επαύξηση των χαρτών χαρακτηριστικών

Α.3 DSU

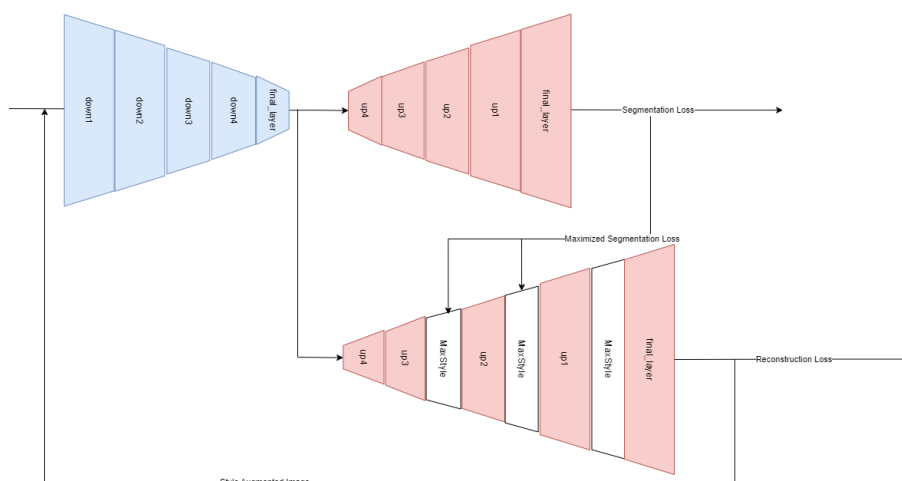
Η αρχιτεκτονική DSU. Τα blocks DSU λειτουργούν παρόμοια με την αρχιτεκτονική MixStyle, όμως μοντελοποιούν τα χαρακτηριστικά, δηλαδή την μέση τιμή και την διακύμανση, ως πολυδιάστατες γκαουσιανές κατανομές.



Σχήμα Α.3: Αρχιτεκτονική με εισαγωγή DSU blocks στα αρχικά επίπεδα για επαύξηση των χαρτών χαρακτηριστικών

Α.4 MaxStyle

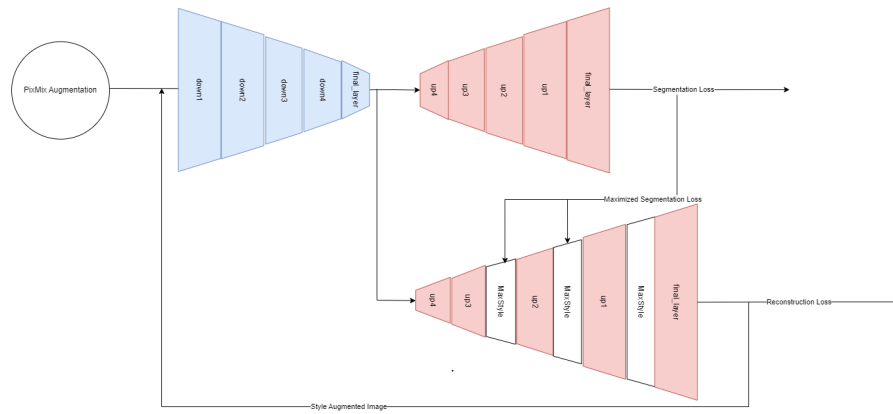
Η αρχιτεκτονική MaxStyle. Τα επίπεδα MaxStyle τοποθετούνται τα τελευταία επίπεδα του αποκωδικοποιητή για πιο ποιοτική επαύξηση των εικόνων της παρτίδας. Οι παράμετροι μίξης αναβαθμίζονται βάσει της μεγιστοποίησης του σφάλματος κατάτμησης και ο υποβοηθητικός αποκωδικοποιητής εκπαιδεύεται βάσει του σφάλματος ανακατασκευής.



Σχήμα Α.4: Αρχιτεκτονική με εισαγωγή MaxStyle blocks στα αρχικά επίπεδα για επαύξηση των χαρτών χαρακτηριστικών. Ο βοηθητικός αποκωδικοποιητής και τα σφάλματα φαίνονται από κάτω.

Α.5 MaxStyle /w Pixmix

Η συνολική αρχιτεκτονική. Προσθετικά πάνω στην αρχιτεκτονική τοποθετούμε την επαύξηση PixMix. Η επαύξηση εφαρμόζεται με πιθανότητα 30% σε ολόκληρη την παρτίδα για να μπορεί να επωφεληθεί ο βοηθητικός αποκωδικοποιητής από την μίξη μεταξύ των εικόνων.



Σχήμα Α.5: Αρχιτεκτονική με εισαγωγή MaxStyle blocks στα αρχικά επίπεδα για επαύξηση των χαρτών χαρακτηριστικών και επαύξηση δεδομένων στο επίπεδο εισόδου PixMix

Βιβλιογραφία

- [1] Di Feng, Christian Haase-Schütz, Lars Rosenbaum, Heinz Hertlein, Claudius Gläser, Fabian Timm, Werner Wiesbeck και Klaus Dietmayer. *Deep Multi-Modal Object Detection and Semantic Segmentation for Autonomous Driving: Datasets, Methods, and Challenges*. *IEEE Transactions on Intelligent Transportation Systems*, 22(3):1341–1360, 2021.
- [2] *What is Image Segmentation: The Basics and Key Techniques | Mindy Support Outsourcing – mindy-support.com*. <https://mindy-support.com/news-post/what-is-image-segmentation-the-basics-and-key-techniques/>. [Ημερομηνία Πρόσβασης 28-09-2023].
- [3] Wenjun Yan, Yuanyuan Wang, Shengjia Gu, Lu Huang, Fuhua Yan, Liming Xia και Qian Tao. *The Domain Shift Problem of Medical Image Segmentation and Vendor-Adaptation by Unet-GAN*, 2019.
- [4] Leon A. Gatys, Alexander S. Ecker και Matthias Bethge. *Image Style Transfer Using Convolutional Neural Networks*. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [5] *Introduction to Convolutional Neural Networks CNNs – aigents.co*. <https://aigents.co/data-science-blog/publication/introduction-to-convolutional-neural-networks-cnns>. [Ημερομηνία Πρόσβασης 06-10-2023].
- [6] Kaiming He, Xiangyu Zhang, Shaoqing Ren και Jian Sun. *Deep Residual Learning for Image Recognition*, 2015.
- [7] Jonathan Long, Evan Shelhamer και Trevor Darrell. *Fully Convolutional Networks for Semantic Segmentation*, 2015.
- [8] Olaf Ronneberger, Philipp Fischer και Thomas Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation*, 2015.
- [9] Daifeng Peng, Yiyi Zhang και Guan. *End-to-End Change Detection for High Resolution Satellite Images Using Improved UNet++*. *Remote Sensing*, 11:1382, 2019.
- [10] Chen Chen, Kerstin Hammernik, Cheng Ouyang, Chen Qin, Wenjia Bai και Daniel Rueckert. *Cooperative Training and Latent Space Data Augmentation for Robust Medical Image Segmentation*, 2021.

- [11] Kaiyang Zhou, Ziwei Liu, Yu Qiao, Tao Xiang και Chen Change Loy. *Domain Generalization: A Survey*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, σελίδες 1–20, 2022.
- [12] Ian J. Goodfellow, Jonathon Shlens και Christian Szegedy. *Explaining and Harnessing Adversarial Examples*, 2015.
- [13] Yabin Zhang, Minghan Li, Ruihuang Li, Kui Jia και Lei Zhang. *Exact Feature Distribution Matching for Arbitrary Style Transfer and Domain Generalization*. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, σελίδες 8035–8045, 2022.
- [14] Tim Cheng. *Enhancing Neural Networks with Mixup in PyTorch – towards-datascience.com*. <https://towardsdatascience.com/enhancing-neural-networks-with-mixup-in-pytorch-5129d261bc4a>. [Ημερομηνία Πρόσβασης 23-10-2023].
- [15] Nicholas J. Tustison, Brian B. Avants, Philip A. Cook, Yuanjie Zheng, Alexander Egan, Paul A. Yushkevich και James C. Gee. *N4ITK: Improved N3 Bias Correction*. *IEEE Trans. Med. Imaging*, 29(6):1310–1320, 2010.
- [16] Hongyi Zhang, Moustapha Cisse, Yann N. Dauphin και David Lopez-Paz. *mixup: Beyond Empirical Risk Minimization*, 2018.
- [17] Terrance DeVries και Graham W. Taylor. *Improved Regularization of Convolutional Neural Networks with Cutout*, 2017.
- [18] Ekin D. Cubuk, Barret Zoph, Dandelion Mane, Vijay Vasudevan και Quoc V. Le. *AutoAugment: Learning Augmentation Policies from Data*, 2019.
- [19] Daniel Ho, Eric Liang, Ion Stoica, Pieter Abbeel και Xi Chen. *Population Based Augmentation: Efficient Learning of Augmentation Policy Schedules*, 2019.
- [20] Dan Hendrycks, Andy Zou, Mantas Mazeika, Leonard Tang, Bo Li, Dawn Song και Jacob Steinhardt. *PixMix: Dreamlike Pictures Comprehensively Improve Safety Measures*, 2022.
- [21] Kaiyang Zhou, Yongxin Yang, Yu Qiao και Tao Xiang. *Domain Generalization with MixStyle*, 2021.
- [22] Xiaotong Li, Yongxing Dai, Yixiao Ge, Jun Liu, Ying Shan και Ling Yu Duan. *Uncertainty Modeling for Out-of-Distribution Generalization*, 2022.
- [23] Chen Chen, Zeju Li, Cheng Ouyang, Matt Sinclair, Wenjia Bai και Daniel Rueckert. *MaxStyle: Adversarial Style Composition for Robust Medical Image Segmentation*, 2022.
- [24] Shervin Minaee, Yuri Boykov, Fatih Porikli, Antonio Plaza, Nasser Kehtarnavaz και Demetri Terzopoulos. *Image Segmentation Using Deep Learning: A Survey*, 2020.

- [25] Risheng Wang, Tao Lei, Ruixia Cui, Bingtao Zhang, Hongying Meng και Asoke K. Nandi. *Medical image segmentation using deep learning: A survey*. *IET Image Processing*, 16(5):1243–1267, 2022.
- [26] Georgios Sapountzakis, Paraskevi Antonia Theofilou και Paraskevi Tzouveli. *Covid-19 Detection From X-Rays Images Using Deep Learning Methods*. *2023 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW)*, σελίδες 1–5, 2023.
- [27] Senay Cakir, Marcel Gauß, Kai Häppeler, Yassine Ounajjar, Fabian Heinle και Reiner Marchthaler. *Semantic Segmentation for Autonomous Driving: Model Evaluation, Dataset Generation, Perspective Comparison, and Real-Time Capability*, 2022.
- [28] Eleftherios Lymperopoulos, Paraskevi Tzouveli και Stefanos Kollias. *Satellite image super-resolution for forest localization*. *2023 International Conference on Machine Intelligence for GeoAnalytics and Remote Sensing (MIGARS)*, τόμος 1, σελίδες 1–4, 2023.
- [29] Sangsun Moon. *What is Image Segmentation? - Definition, Types, Use case, Deep Learning, Trends* – *thedatahunt.com*. <https://www.thedatahunt.com/en-insight/image-segmentation>. [Ημερομηνία Πρόσβασης 13-09-2023].
- [30] Jose G Moreno-Torres, Troy Raeder, Raúl Alaiz-Rodríguez, Nitesh V Chawla και Francisco Herrera. *A unifying view on dataset shift in classification*. *Pattern Recognition*, 2012.
- [31] Benjamin Recht, Rebecca Roelofs, Ludwig Schmidt και Vaishaal Shankar. *Do ImageNet classifiers generalize to ImageNet?* *International Conference on Machine Learning (ICML)*, 2019.
- [32] Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira και Jennifer Wortman Vaughan. *A theory of learning from different domains*. *Machine Learning*, 2010.
- [33] Rohan Taori, Abhishek Dave, Vaishaal Shankar, Nicholas Carlini, Benjamin Recht και Ludwig Schmidt. *Measuring robustness to natural distribution shifts in image classification*. *NeurIPS*, 2020.
- [34] Gilles Blanchard, Akshay Anand Deshmukh, Ugur Dogan, Gary Lee και Clayton Scott. *Domain generalization by marginal transfer learning*. *Journal of Machine Learning Research (JMLR)*, 2021.
- [35] Hidetoshi Shimodaira. *Improving predictive inference under covariate shift by weighting the log-likelihood function*. *Journal of Statistical Planning and Inference*, 90(2):227–244, 2000.
- [36] *Domain shift* – *statlect.com*. <https://www.statlect.com/machine-learning/domain-shift>. [Ημερομηνία Πρόσβασης 12-09-2023].

- [37] Hao Guan και Mingxia Liu. *Domain Adaptation for Medical Image Analysis: A Survey*. *IEEE Transactions on Biomedical Engineering*, 69(3):1173–1185, 2022.
- [38] Dan Hendrycks και Thomas Dietterich. *Benchmarking neural network robustness to common corruptions and perturbations*. *International Conference on Learning Representations (ICLR)*, 2019.
- [39] Jie Yang, Kun Zhou, Yufei Li και Zhiyuan Liu. *Generalized out-of-distribution detection: A survey*. *arXiv preprint arXiv:2110.11334*, 2021.
- [40] Gilles Blanchard, Gyemin Lee και Clayton Scott. *Generalizing from Several Related Classification Tasks to a New Unlabeled Sample*. *Advances in Neural Information Processing Systems* J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira και K.Q. Weinberger, επιμελητές, τόμος 24. Curran Associates, Inc., 2011.
- [41] Leon A. Gatys, Alexander S. Ecker και Matthias Bethge. *A Neural Algorithm of Artistic Style*, 2015.
- [42] *Neural Style Transfer: Using Deep Learning to Generate Art – v7labs.com*. <https://www.v7labs.com/blog/neural-style-transfer>. [Ημερομηνία Πρόσβασης 12-09-2023].
- [43] Alex Krizhevsky, Ilya Sutskever και Geoffrey E Hinton. *ImageNet Classification with Deep Convolutional Neural Networks*. *Advances in Neural Information Processing Systems* F. Pereira, C.J. Burges, L. Bottou και K.Q. Weinberger, επιμελητές, τόμος 25. Curran Associates, Inc., 2012.
- [44] Karen Simonyan και Andrew Zisserman. *Very Deep Convolutional Networks for Large-Scale Image Recognition*, 2015.
- [45] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke και Andrew Rabinovich. *Going Deeper with Convolutions*, 2014.
- [46] Guotai Wang, Wenqi Li, Sebastien Ourselin και Tom Vercauteren. *Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks*. *International MICCAI Brainlesion Workshop*, σελίδες 178–190, 2017.
- [47] Yang Li, Hao Qi, Jiyan Dai, Xiaodan Ji και Yichen Wei. *Fully convolutional instance-aware semantic segmentation*. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, σελίδες 2359–2367, 2017.
- [48] Yutong Yuan, Ma Chao και Yung Che Lo. *Automatic skin lesion segmentation using deep fully convolutional networks with Jaccard distance*. *IEEE Transactions on Medical Imaging*, 36(9):1876–1886, 2017.
- [49] Ning Liu, Hui Li, Meng Zhang, Jing Liu, Zhenan Sun και Tieniu Tan. *Accurate iris segmentation in non-cooperative environments using fully convolutional networks*. *2016 International Conference on Biometrics (ICB)*, σελίδες 1–8, 2016.

- [50] Natalia Salpea, Paraskevi Tzouveli και Dimitrios Kollias. *Medical Image Segmentation: A Review of Modern Architectures*. *Computer Vision - ECCV 2022 Workshops* Leonid Karlinsky, Tomer Michaeli και Ko Nishino, επιμελητές, σελίδες 691–708, Cham, 2023. Springer Nature Switzerland.
- [51] Diederik P Kingma και Max Welling. *Auto-Encoding Variational Bayes*, 2022.
- [52] Pascal Vincent, H. Larochelle, Isabelle Lajoie, Yoshua Bengio και Pierre Antoine Manzagol. *Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion*. *J. Mach. Learn. Res.*, 11:3371–3408, 2010.
- [53] Daniel Kahneman. *Thinking, fast and slow*. Farrar, Straus and Giroux, New York, 2011.
- [54] Zhihe Lu, Yongxin Yang, Xiatian Zhu, Cong Liu, Yi Zhe Song και Tao Xiang. *Stochastic Classifiers for Unsupervised Domain Adaptation*. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.
- [55] Kuniaki Saito, Kohei Watanabe, Yoshitaka Ushiku και Tatsuya Harada. *Maximum Classifier Discrepancy for Unsupervised Domain Adaptation*, 2018.
- [56] Yaroslav Ganin και Victor Lempitsky. *Unsupervised Domain Adaptation by Back-propagation*, 2015.
- [57] Mingsheng Long, Yue Cao, Jianmin Wang και Michael I. Jordan. *Learning Transferable Features with Deep Adaptation Networks*, 2015.
- [58] Ziwei Liu, Zhongqi Miao, Xingang Pan, Xiaohang Zhan, Dahua Lin, Stella X. Yu και Boqing Gong. *Open Compound Domain Adaptation*. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.
- [59] Boqing Gong, Yuan Shi, Fei Sha και Kristen Grauman. *Geodesic flow kernel for unsupervised domain adaptation*. *2012 IEEE Conference on Computer Vision and Pattern Recognition*, σελίδες 2066–2073, 2012.
- [60] Mingsheng Long, Han Zhu, Jianmin Wang και Michael I. Jordan. *Unsupervised Domain Adaptation with Residual Transfer Networks*, 2017.
- [61] Yogesh Balaji, Rama Chellappa και Soheil Feizi. *Normalized Wasserstein Distance for Mixture Distributions with Applications in Adversarial Learning and Domain Adaptation*, 2019.
- [62] Guoliang Kang, Lu Jiang, Yi Yang και Alexander G Hauptmann. *Contrastive Adaptation Network for Unsupervised Domain Adaptation*, 2019.
- [63] Xide Xia και Brian Kulis. *W-Net: A Deep Model for Fully Unsupervised Image Segmentation*, 2017.

- [64] Asako Kanezaki. *Unsupervised Image Segmentation by Backpropagation*. 2018 *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, σελίδες 1543–1547, 2018.
- [65] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville και Yoshua Bengio. *Generative Adversarial Nets*. *Advances in Neural Information Processing Systems*. Ghahramani, M. Welling, C. Cortes, N. Lawrence και K.Q. Weinberger, επιμελητές, τόμος 27. Curran Associates, Inc., 2014.
- [66] Mohammad Mahfujur Rahman, Clinton Fookes, Mahsa Baktashmotlagh και Sridha Sridharan. *Correlation-aware adversarial domain adaptation and generalization*. *Pattern Recognition*, 100:107124, 2020.
- [67] Isabela Albuquerque, João Monteiro, Mohammad Darvishi, Tiago H. Falk και Ioannis Mitliagkas. *Generalizing to unseen domains via distribution matching*, 2021.
- [68] Zhun Deng, Frances Ding, Cynthia Dwork, Rachel Hong, Giovanni Parmigiani, Prasad Patil και Pragya Sur. *Representation via Representations: Domain Generalization via Adversarially Learned Invariant Representations*, 2020.
- [69] Rui Shao, Xiangyuan Lan, Jiawei Li και Pong C. Yuen. *Multi-Adversarial Discriminative Deep Domain Generalization for Face Presentation Attack Detection*. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, σελίδες 10015–10023, 2019.
- [70] *Papers with Code - Style Transfer – paperswithcode.com*. <https://paperswithcode.com/task/style-transfer>. [Ημερομηνία Πρόσβασης 20-09-2023].
- [71] Xun Huang και Serge Belongie. *Arbitrary Style Transfer in Real-time with Adaptive Instance Normalization*, 2017.
- [72] Lei Li, Veronika A. Zimmer, Wangbin Ding, Fuping Wu, Liqin Huang, Julia A. Schnabel και Xiahai Zhuang. *Random Style Transfer based Domain Generalization Networks Integrating Shape and Spatial Information*, 2020.
- [73] Kaiyang Zhou, Chen Change Loy και Ziwei Liu. *Semi-Supervised Domain Generalization with Stochastic StyleMatch*, 2021.
- [74] Nathan Somavarapu, Chih Yao Ma και Zsolt Kira. *Frustratingly Simple Domain Generalization via Image Stylization*, 2020.
- [75] Francesco Cappio Borlino, Antonio D’Innocente και Tatiana Tommasi. *Rethinking Domain Generalization Baselines*, 2021.
- [76] Xiangyu Yue, Yang Zhang, Sicheng Zhao, Alberto Sangiovanni-Vincentelli, Kurt Keutzer και Boqing Gong. *Domain Randomization and Pyramid Consistency: Simulation-to-Real Generalization without Accessing Target Domain Data*, 2022.

- [77] Athanasios Tagaris, Dimitrios Kollias και Andreas Stafylopatis. *Assessment of Parkinson's disease based on deep neural networks*. *International Conference on Engineering Applications of Neural Networks*, σελίδες 391–403. Springer, 2017.
- [78] Dimitrios Kollias, Athanasios Tagaris, Andreas Stafylopatis, Stefanos Kollias και Georgios Tagaris. *Deep neural architectures for prediction in healthcare*. *Complex & Intelligent Systems*, 4(2):119–131, 2018.
- [79] Athanasios Tagaris, Dimitrios Kollias, Andreas Stafylopatis, Georgios Tagaris και Stefanos Kollias. *Machine learning for neurodegenerative disorder diagnosis—survey of practices and launch of benchmark dataset*. *International Journal on Artificial Intelligence Tools*, 27(03):1850011, 2018.
- [80] Ilianna Kollia, Andreas Georgios Stafylopatis και Stefanos Kollias. *Predicting Parkinson's disease using latent information extracted from deep neural networks*. *2019 International Joint Conference on Neural Networks (IJCNN)*, σελίδες 1–8. IEEE, 2019.
- [81] James Wingate, Ilianna Kollia, Luc Bidaut και Stefanos Kollias. *Unified deep learning approach for prediction of Parkinson's disease*. *IET Image Processing*, 14(10):1980–1989, 2020.
- [82] Dimitrios Kollias, Anastasios Arsenos, Levon Soukissian και Stefanos Kollias. *Miacov19d: Covid-19 detection through 3-d chest ct image analysis*. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, σελίδες 537–544, 2021.
- [83] Anastasios Arsenos, Dimitrios Kollias και Stefanos Kollias. *A Large Imaging Database and Novel Deep Neural Architecture for Covid-19 Diagnosis*. *2022 IEEE 14th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)*, σελίδες 1–5. IEEE, 2022.
- [84] Dimitrios Kollias, Anastasios Arsenos και Stefanos Kollias. *AI-MIA: Covid-19 detection and severity analysis through medical imaging*. *Computer Vision—ECCV 2022 Workshops: Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part VII*, σελίδες 677–690. Springer, 2023.
- [85] Anastasios Arsenos, Andjoli Davidhi, Dimitrios Kollias, Panos Prassopoulos και Stefanos Kollias. *Data-Driven Covid-19 Detection Through Medical Imaging*. *2023 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW)*, σελίδες 1–5, 2023.
- [86] Philippos Skovelef Orfanoudakis, Paraskevi Tzouveli και Stefanos Kollias. *Effects of image quality and quantity on building a competitive COVID-19 diagnosis model*. *2021 IEEE International Conference on Big Data (Big Data)*, σελίδες 4405–4410, 2021.
- [87] Natalia Salpea, Paraskevi Tzouveli και Dimitrios Kollias. *Medical image segmentation: A review of modern architectures*. *European Conference on Computer Vision*, σελίδες 691–708. Springer, 2022.

- [88] Dimitrios Kollias, Miao Yu, Athanasios Tagaris, Georgios Leontidis, Andreas Stafylopatis και Stefanos Kollias. *Adaptation and contextualization of deep neural network models. 2017 IEEE symposium series on computational intelligence (SSCI)*, σελίδες 1–8. IEEE.
- [89] Dimitris Kollias, Y Vlaxos, M Seferis, Ilianna Kollia, Levon Sukissian, James Wingate και S Kollias. *Transparent adaptation in deep medical image diagnosis. International Workshop on the Foundations of Trustworthy AI Integrating Learning, Optimization and Reasoning*, σελίδες 251–267. Springer, 2020.
- [90] Fabio De Sousa Ribeiro, Francesco Calivá, Mark Swainson, Kjartan Gudmundsson, Georgios Leontidis και Stefanos Kollias. *Deep bayesian self-training. Neural Computing and Applications*, 32(9):4275–4291, 2020.
- [91] Fabio De Sousa Ribeiro, Georgios Leontidis και Stefanos Kollias. *Capsule routing via variational bayes. Proceedings of the AAAI Conference on Artificial Intelligence*, τόμος 34, σελίδες 3749–3756, 2020.
- [92] Fabio De Sousa Ribeiro, Georgios Leontidis και Stefanos Kollias. *Introducing routing uncertainty in capsule networks. Advances in Neural Information Processing Systems*, 33:6490–6502, 2020.
- [93] Francesco Caliva, Fabio Sousa De Ribeiro, Antonios Mylonakis, Christophe Demazière, Paolo Vinai, Georgios Leontidis και Stefanos Kollias. *A deep learning approach to anomaly detection in nuclear reactors. 2018 International joint conference on neural networks (IJCNN)*, σελίδες 1–8. IEEE, 2018.
- [94] Stefanos Kollias, Miao Yu, James Wingate, Aiden Durrant, Georgios Leontidis, Georgios Alexandridis, Andreas Stafylopatis, Antonios Mylonakis, Paolo Vinai και Christophe Demaziere. *Machine learning for analysis of real nuclear plant data in the frequency domain. Annals of Nuclear Energy*, 177:109293, 2022.
- [95] Bashar Alhnaity, Stefanos Kollias, Georgios Leontidis, Shouyong Jiang, Bert Schamp και Simon Pearson. *An autoencoder wavelet based deep neural network with attention mechanism for multi-step prediction of plant growth. Information Sciences*, 560:35–50, 2021.
- [96] Bashar Alhnaity, Simon Pearson, Georgios Leontidis και Stefanos Kollias. *Using deep learning to predict plant growth and yield in greenhouse environments. International Symposium on Advanced Technologies and Management for Innovative Greenhouses: GreenSys2019 1296*, σελίδες 425–432, 2019.
- [97] Dimitrios Kollias, Panagiotis Tzirakis, Alice Baird, Alan Cowen και Stefanos Zafeiriou. *Abaw: Valence-arousal estimation, expression recognition, action unit detection & emotional reaction intensity estimation challenges. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, σελίδες 5888–5897, 2023.

- [98] Andreas Psaroudakis και Dimitrios Kollias. *MixAugment & Mixup: Augmentation Methods for Facial Expression Recognition. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, σελίδες 2367–2375, 2022.
- [99] *A snake model for object tracking in natural sequences. Signal processing: image communication*, 19(3):219–238, 2004.
- [100] *Interactive content-based retrieval in video databases using fuzzy classification and relevance feedback. Proceedings IEEE International Conference on Multimedia Computing and Systems*, τόμος 2, σελίδες 954–958. IEEE, 1999.
- [101] Manolis Wallace, Ilias Maglogiannis, Kostas Karpouzis, George Kormentzas και Stefanos Kollias. *Intelligent one-stop-shop travel recommendations using an adaptive neural network and clustering of history. Information Technology & Tourism*, 6(3):181–193, 2003.
- [102] Miao Yu, Liyun Gong και Stefanos Kollias. *Computer vision based fall detection by a convolutional neural network. Proceedings of the 19th ACM international conference on multimodal interaction*, σελίδες 416–420, 2017.
- [103] *Fire: A fuzzy reasoning engine for imprecise knowledge.*
- [104] G Caridakis, A Raouzaïou, K Karpouzis και S Kollias. *Synthesizing gesture expressivity based on real sequences.*
- [105] S Kollias. *Towards modeling embodied conversational agent character profiles using appraisal theory predictions in expression synthesis. Applied intelligence*, 30:58–64, 2009.
- [106] Sofia Gourgari, Georgios Goudelis, Konstantinos Karpouzis και Stefanos Kollias. *Thetis: Three dimensional tennis shots a human action dataset. Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, σελίδες 676–681, 2013.
- [107] Phivos Mylonas, Evaggelos Spyrou, Yannis Avrithis και Stefanos Kollias. *Using visual context and region semantics for high-level concept detection. IEEE Transactions on Multimedia*, 11(2):229–243, 2009.
- [108] *Face detection in color images and video sequences. 2000 10th Mediterranean Electrotechnical Conference. Information Technology and Electrotechnology for the Mediterranean Countries. Proceedings. MeleCon 2000 (Cat. No. 00CH37099)*, τόμος 2, σελίδες 498–502. IEEE, 2000.
- [109] *A unified neural network approach to digital image halftoning. IEEE Transactions on signal processing*, 39(4):980–984, 1991.
- [110] Q. Liu και others. *Ms-net: Multi-site network for improving prostate segmentation with heterogeneous MRI data. IEEE Transactions on Medical Imaging (TMI)*, 2020.

- [111] Q. Liu, Q. Dou και P. A. Heng. *Shape-aware Meta-learning for Generalizing Prostate MRI Segmentation to Unseen Domains*. *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2020.
- [112] *NCI-ISBI 2013 Challenge - Automated Segmentation of Prostate Structures - The Cancer Imaging Archive (TCIA) Public Access - Cancer Imaging Archive Wiki* – [wiki.cancerimagingarchive.net](https://wiki.cancerimagingarchive.net/display/public/nci-isbi+2013+challenge+-+automated+segmentation+of+prostate+structures). <https://wiki.cancerimagingarchive.net/display/public/nci-isbi+2013+challenge+-+automated+segmentation+of+prostate+structures>. [Ημερομηνία Πρόσβασης 15-09-2023].
- [113] Guillaume Lemaitre, Robert Marti, Jordi Freixenet, Joan C Vilanova, Paul M Walker και Fabrice Meriaudeau. *Computer-Aided Detection and Diagnosis for prostate cancer based on mono and multi-parametric MRI: A Review*. *Computers in Biology and Medicine*, 60:8–31, 2015.
- [114] *PROMISE12 - Grand Challenge* – promise12.grand-challenge.org. <https://promise12.grand-challenge.org>. [Ημερομηνία Πρόσβασης 15-09-2023].
- [115] M. Antonelli και others. *The Medical Segmentation Decathlon*. *arXiv:2106.05735*, 2021.
- [116] Victor Manuel Campello και others. *Multi-Centre, Multi-Vendor and Multi-Disease Cardiac Segmentation: The M&Ms Challenge*. *IEEE Transactions on Medical Imaging*, 2021.
- [117] Christian Thode Larsen, J. Eugenio Iglesias και Koen Van Leemput. *N3 Bias Field Correction Explained as a Bayesian Modeling Method*. *Bayesian and graphical Models for Biomedical Imaging*. Jorge Cardoso, Ivor Simpson, Tal Arbel, Doina Precup και Annemie Ribbens, επιμελητές, σελίδες 1–12, Cham, 2014. Springer International Publishing.
- [118] Suorong Yang, Weikang Xiao, Mengcheng Zhang, Suhan Guo, Jian Zhao και Furao Shen. *Image Data Augmentation for Deep Learning: A Survey*, 2022.
- [119] Adrian Sandru, Mariana Iuliana Georgescu και Radu Tudor Ionescu. *Feature-level augmentation to improve robustness of deep neural networks to affine transformations*, 2022.
- [120] *Dice score - OECD.AI* – [oecd.ai](https://oecd.ai/en/catalogue/metrics/dice-score). <https://oecd.ai/en/catalogue/metrics/dice-score>. [Ημερομηνία Πρόσβασης 13-09-2023].

Συντομογραφίες - Αρκτικόλεξα - Ακρωνύμια

βλπ	βλέπε
κ.λπ.	και λοιπά
κ.ο.κ	και ούτω καθεξής
OOD	Out-of-Domain
ID	Intra Domain
MRI	Magnetic Resonance Imaging
CNN	Convolutional Neural Network
FCN	Fully Convolutional Network
FTN	Fast Thinking Network
STN	Slow Thinking Network

Απόδοση ξενόγλωσσων όρων

Απόδοση

γενίκευση πεδίου
προσαρμογή πεδίου
μεταφορά στυλ
παρτίδα
επαύξηση δεδομένων
επαύξηση χαρακτηριστικών
ανταγωνιστική εκμάθηση
κατάτμηση εικόνων
κατάτμηση περιπτώσεων
σημασιολογική κατάτμηση
εκμάθηση μεταφοράς

Ξενόγλωσσος όρος

domain generalization
domain adaptation
style transfer
batch
data augmentation
feature augmentation
adversarial training
image segmentation
instance segmentation
semantic segmentation
transfer learning

