



Εθνικό Μετσόβιο Πολυτεχνείο

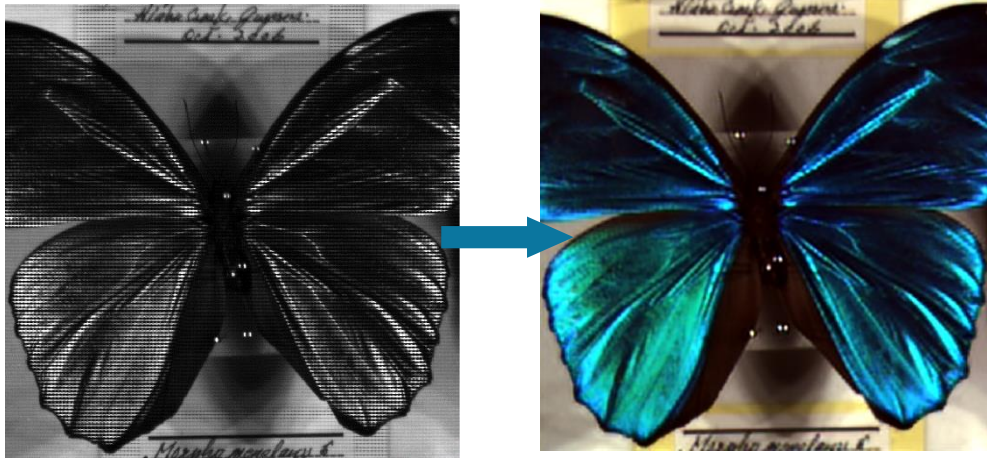
Σχολή Αγρονόμων και Τοπογράφων Μηχανικών -
Μηχανικών Γεωπληροφορικής

Τομέας Τοπογραφίας

Εργαστήριο Τηλεπισκόπησης

Διπλωματική Εργασία:

«Απομοσαϊκοποίηση πολυφασματικών δεδομένων με χρήση
Συνελκτικών Νευρωνικών Δικτύων»



Συντάξασα: Ασπρολούπου Μυρσίνη Παναγιώτα (Α.Μ. rs18066)

Επιβλέπων: Καράντζαλος Κωνσταντίνος, Καθηγητής

Αθήνα, Ιούλιος 2024

Τριμελής Εξεταστική Επιτροπή

Καράντζαλος Κωνσταντίνος

Παπαδοπούλου Μαρία

Τσιχριντζής Βασίλειος

Καθηγητής

Καθηγήτρια

Καθηγητής



National Technical University of Athens

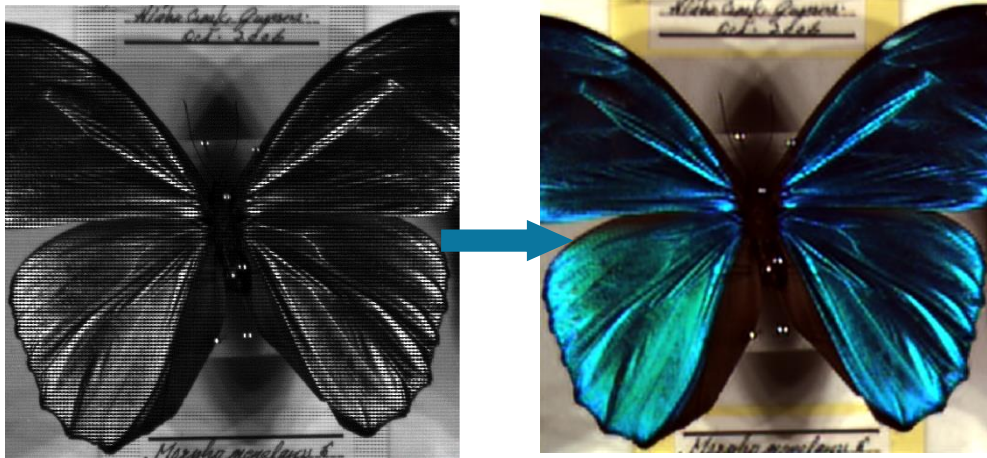
School of Rural, Surveying and Geoinformatics
Engineering

Department of Topography

Remote Sensing Laboratory

Diploma Thesis:

**«Demosaiicing of multispectral data using Convolutional Neural
Networks»**



Author: Asprolopou Myrsini Panagiota (A.M. rs18066)

Supervisor: Karantzalos Konstantinos, Professor

Athens, July 2024

Three-member Committee

Karantzalos Konstantinos

Papadopoulou Maria

Tsixritzis Vasileios

Professor

Professor

Professor

Ευχαριστίες

Αρχικά, επιθυμώ να ευχαριστήσω θερμά τον υπεύθυνο καθηγητή της Διπλωματικής μου κ. Κωνσταντίνο Καραντζαλο που στάθηκε δίπλα μου εισακούγοντας τις κλίσεις μου και μεριμνώντας για την επιλογή ενός θέματος με προοπτικές. Έπειτα, θα ήθελα να εκφράσω την ειλικρινή μου ευγνωμοσύνη στον συνεπιβλέποντα της διπλωματικής μου Δρ. Βαλσάμη Ντούσκο με την συνεισφορά του να είναι καίρια όντας παρών τόσο στην εκτενέστερη κατανόηση του θεωρητικού υποβάθρου όσο και στο κομμάτι των ιδεών και της επίλυσης προβλημάτων. Ένα ιδιαίτερο ευχαριστώ αξίζει στον φίλο μου και συνάδελφο μου Σωτήριο Σπανό, όπου η βοήθεια του ήταν καταλυτική στο τμήμα της υλοποίησης καθώς η τεχνολογική του κατάρτιση και το επιστημονικό του υπόβαθρο ήταν επιδραστικά για την ολοκλήρωση της παρούσας ΔΕ. Δεν μπορεί να λησμονηθεί, ο οικογενειακός μου κύκλος που με στηρίζει σε κάθε βήμα της ζωής μου, αποτελώντας πάντα ένα σταθερό θεμέλιο. Τέλος, ευχαριστώ θερμά το αγόρι μου και τους στενούς μου φίλους που με υποστηρίζουν και εμπυχώνουν καθ' όλη τη διάρκεια της ακαδημαϊκής μου πορείας και ιδίως κατά την εκπόνηση της διπλωματικής μου εργασίας, όπου αποτέλεσαν αναντικατάστατο καταφύγιο.

Περίληψη

Η χρήση πολυφασματικών δεδομένων αποτελεί ένα πολύτιμο εργαλείο που αποδεικνύεται χρήσιμο σε πολλά επιστημονικά πεδία, για αυτό και η μελέτη τους αποτελεί αντικείμενο πολλών πρωτότυπων ερευνητικών εργασιών. Η διαχείριση τέτοιου είδους δεδομένων ποικίλει ανάλογα τη μορφολογία των αισθητήρων που χρησιμοποιούνται και τη σκοπιμότητα κάθε εφαρμογής. Στην προκειμένη περίπτωση, το πρόβλημα που καλούμαστε να αντιμετωπίσουμε είναι αυτό του MSFA demosaicing, δηλαδή της ανακατασκευής πολυφασματικών εικόνων σε πλήρη χωρική ανάλυση για δεδομένα που έχουν ληφθεί από πολυφασματικούς αισθητήρες επίπεδης διάταξης.

Ένα τέτοιο πρόβλημα διαχειρίζεται σαν μία super-resolution inverse εργασία. Η προσέγγιση που αναπτύσσεται στην παρούσα διπλωματική συνιστά μία δισκελή μεθοδολογία, με τελικό στόχο την αναδόμηση πολυφασματικών εικόνων που θα προσομοιάζουν σε πολύ καλό βαθμό τα πραγματικά δεδομένα. Αναλυτικά, στο πρώτο στάδιο προσέγγισης αυτής της λύσης χρησιμοποιείται το training free μοντέλο που στηρίζεται στη λογική του Deep Image Prior, που αποτελεί μία εργασία βελτιστοποίησης, δηλαδή γίνεται επίλυση ανά εικόνα, και εξάγει αποτελέσματα εξαιρετικής ανάλυσης. Σε επόμενο στάδιο τα δεδομένα αυτά χρησιμοποιούνται σαν δεδομένα εκπαίδευσης για το επόμενο μοντέλο που είναι δομής τύπου U-Net που αξιοποιεί χαρακτηριστικά attention και recurrent blocks. Το δεύτερο μοντέλο στοχεύει στη μίμηση της διαδικασίας απομοσαϊκοποίησης (demosaicing) των δεδομένων όπως επιτελεί το MD²IP προσεγγίζοντας το δυνατόν περισσότερο την απόδοση του αλλά με ταχύτερη απόκριση. Η παραπάνω μεθοδολογία αξιολογείται με ποσοτικά και ποιοτικά κριτήρια και αποδεικνύει την μεγάλη επιτυχία της συγκεκριμένης πρότασης. Επομένως, με ευκολία μπορεί να ειπωθεί ότι αποτελεί ένα ολοκληρωμένο framework που μπορεί να αξιοποιηθεί μελλοντικά από χρήστες για την επεξεργασία αντίστοιχων πολυφασματικών δεδομένων που ακολουθούν εφάμιλλο τρόπο καταγραφής της πληροφορίας.

Λέξεις Κλειδιά: Τεχνητή Νοημοσύνη, Βαθιά Μάθηση, Όραση υπολογιστών, Πολυφασματικές εικόνες, Απομοσαϊκοποίηση MSFA, Αποκατάσταση εικόνας, Προγενέστερη γνώση εικόνας

Abstract

The use of multispectral data is a valuable tool that proves useful in many scientific fields, which is why their study is the subject of many original research papers. The management of such data varies depending on the morphology of the sensors used and the purpose of each application. In this case, the problem we are called to address is that of MSFA demosaicing, that is the reconstruction of multispectral images at full spatial resolution for data acquired from multispectral flat array sensors.

Such a problem is handled as a super-resolution inverse task. The approach developed in this thesis constitutes a two-part methodology, with the ultimate goal of reconstructing multispectral images that closely resemble real data. Specifically, in the first stage of this solution, a training-free model based on the logic of Deep Image Prior is used, which is an optimization task, meaning it is solved on a per-image basis, and yields results of excellent resolution. In the next stage, these data are used as training data for the next model, which has a U-Net type structure that leverages attention features and recurrent blocks. The second model aims to mimic the demosaicing process of the data as performed by the MD²IP, approximating its performance as closely as possible but with faster response. The aforementioned methodology is evaluated using quantitative and qualitative criteria and demonstrates the great success of this approach. Therefore, it can easily be said that it constitutes a comprehensive framework that can be utilized in the future by users for processing similar multispectral data that follow a comparable method of information recording.

Key words: Artificial Intelligence, Deep Learning, Computer vision, Multispectral images, MSFA Demosaicing, Image Restoration, Image Prior

Περιεχόμενα

Ευχαριστίες	9
Περίληψη	10
Abstract	11
Ευρετήριο Σχημάτων	14
Ευρετήριο Πινάκων	17
Συνοτομογραφίες.....	18
ΚΕΦΑΛΑΙΟ 1 ^ο	19
1.1. Εισαγωγή στο αντικείμενο μελέτης.....	19
1.2. Κύριες προκλήσεις.....	20
1.3. Συνεισφορά.....	22
1.4. Διάρθρωση Διπλωματικής Εργασίας.....	23
ΚΕΦΑΛΑΙΟ 2 ^ο	24
2.1. Επιστημονικό / Θεωρητικό Υπόβαθρο	24
2.1.1. Εισαγωγή στη Μηχανική Μάθηση	24
2.1.2. Εισαγωγή στη Βαθιά Μάθηση	26
2.1.3. Συνελκτικά Νευρωνικά Δίκτυα (CNNs)	28
2.1.4. Εκπαίδευση & Βελτιστοποίηση (Training & Optimization).....	29
2.1.5. Μετρικές Αξιολόγησης.....	32
2.1.6. MSFA	33
2.1.7. Demosaicing.....	36
2.1.8. U-Net.....	38
2.1.9. Super Resolution	41
2.2. Ανάλυση Ερευνητικών Εργασιών	44
2.2.1. A Deep Joint Network for Multispectral Demosaicking Based on Pseudo-Panchromatic Images.....	44
2.2.2. Deep Image Prior	48

2.2.3. MD ² IP: TRAINING-FREE MULTISPECTRAL DEMOSAICING WITH DEEP IMAGE PRIORS.....	51
ΚΕΦΑΛΑΙΟ 3 ^ο – 1 ^η μεθοδολογία (MD ² IP)	53
3.1. Σύνολο Δεδομένων	53
3.2. Προεπεξεργασία δεδομένων	58
3.3. Multispectral Demosaicing Deep Image Prior (MD ² IP).....	62
3.3.1. Ανάλυση μεθοδολογίας MD ² IP.....	62
3.3.2. Αποτελέσματα και αξιολόγηση του μοντέλου MD ² IP	66
ΚΕΦΑΛΑΙΟ 4 ^ο – 2 ^η μεθοδολογία (U-Net)	81
4.1. Προεπεξεργασία δεδομένων	81
4.2. Υλοποίηση Μοντέλου τύπου U-Net.....	87
4.3. Πειράματα	91
4.3.1. Δομή Πειραμάτων.....	91
4.3.2. Αντιπροσωπευτικότητα συνόλου Αξιολόγησης (Validation set)	92
4.3.3. Υπερπαραμέτροι Μοντέλων	94
4.3.4. Υλοποιημένα Μοντέλα – Διάταξη Πειραμάτων	97
4.4. Αξιολόγηση.....	100
4.4.1. Ποσοτική Αξιολόγηση	106
4.4.2. Ποιοτική Αξιολόγηση	111
ΚΕΦΑΛΑΙΟ 5 ^ο - Συμπεράσματα - Προτάσεις	119
Βιβλιογραφικές αναφορές.....	122

Ευρετήριο Σχημάτων

Σχήμα 2.1: Αναπαράσταση ενός Τεχνητού Νευρωνικού Δικτύου (ANN). (Πηγή: <https://medium.com/mit-6-s089-intro-to-quantum-computing/quantum-neural-networks-7b5bc469d984>).....σελ.19

Σχήμα 2.2: Τυπική δομή ενός Συνελικτικού Νευρωνικού Δικτύου. (Πηγή: <https://nafizshahriar.medium.com/what-is-convolutional-neural-network-cnn-deep-learning-b3921bdd82d5>).σελ.21

Σχήμα 2.3: ΧΙΜΕΑ's xiQ USB3.0 camera στηρία ενσωματώθηκαν οι αισθητήρες υπερφασματικής απεικόνισης (HSI) της IMEC. (Πηγή: <https://www.ximea.com/en/usb3-vision-camera/xiq>).σελ.27

Σχήμα 2.4: Υπερφασματική απεικόνιση χρησιμοποιώντας την ΧΙΜΕΑ xiSpec κάμερα. (Πηγή: <https://www.ximea.com/support/attachments/5981/SpectroNet-2016-03-Ximea-V02.pdf>).σελ.27

Σχήμα 2.5: Δημιουργία εικόνας MSFA της μορφής NxN, επιλέγοντας από τα κανάλια της πολυφασματικής εικόνας πλήρους ανάλυσης τα pixel εκείνα που ταιριάζουν στη θέση των αντίστοιχων καναλιών στον MSFA. Πηγή: [3].σελ.28

Σχήμα 2.6: Η τυπική διάταξη συστοιχίας φίλτρων Bayer και ο πυρήνας συνέλιξης για τη σχεδίαση συστοιχίας φίλτρων 2x2 τύπου Bayer. Πηγή [5].σελ.29

Σχήμα 2.7: Αρχιτεκτονική UNet μοντέλου. (Πηγή: <https://towardsdatascience.com/unet-line-by-line-explanation-9b191c76baf5>).σελ.33

Σχήμα 2.8: IMSFA, $\hat{I}M$, Iks , $\hat{I}kWB$ and $\hat{I}k$ συμβολίζονται ως η ακατέργαστη ψηφιδωτή MSFA εικόνα, εκτιμώμενη PPI, «αραιή μωσαϊκή» εικόνα, προσωρινή demosaicked εικόνα and τελική demosaicked εικόνα για κάθε κανάλι k , αντίστοιχα. Η ροή εργασιών του DPD-Net αποτελείται από δύο δίκτυα που ονομάζονται DPG-Net και DDM-Net, αντίστοιχα. Πηγή: [1].σελ.37

Σχήμα 2.9: Multispectral Filter Array (MSFA) 4x4, όπως προκύπτει από την IMEC κάμερα. Πηγή: [1].σελ.38

Σχήμα 2.10: Super-resolution χρησιμοποιώντας deep image prior και άλλες μεθόδους που απαιτούν (SRResNet,) ή όχι εκπαίδευση (Bicubic). Πηγή: [2].σελ.41

Σχήμα 2.11: Μετάβαση από την μη αρχικοποιημένη έξοδο του δικτύου στην τελική εικόνα εξόδου \hat{I} μετά το πέρας των επαναλήψεων, για την εικόνα Butterfly (εμφάνιση σε RGB). Πηγή [3].σελ.52

Σχήμα 3.1: Πολυφασματικές εικόνες που απαρτίζουν το σύνολο δεδομένων CAVE. (Πηγή: <https://www.cs.columbia.edu/CAVE/databases/multispectral/>).....σελ.45

Σχήμα 3.2: Ενδεικτικά 15 σκηνές από το σύνολο δεδομένων TokyoTech-31band. (Πηγή: <http://www.ok.sc.e.titech.ac.jp/res/MSI/MSIdata31.html>).σελ.47

Σχήμα 3.3: Ενδεικτικές σκηνές από το σύνολο δεδομένων TokyoTech-59band. (Πηγή: <http://www.ok.sc.e.titech.ac.jp/res/MSI/MSIdata59.html>).σελ.48

Σχήμα 3.4: Ενδεικτικές σκηνές από το σύνολο δεδομένων Manchester 2015. (Πηγή: https://personalpages.manchester.ac.uk/staff/d.h.foster/Local_Illumination_HSIs/Local_Illumination_HSIs_2015.html).σελ.49

Σχήμα 3.5: Η πραγματική εικόνα MSF \hat{A} που κατασκευάστηκε βάση της πολυφασματικής λήψης «balloons», του dataset CAVE.σελ.52

Σχήμα 3.6: Η πραγματική εικόνα MSF \hat{A} που κατασκευάστηκε βάση της πολυφασματικής λήψης «Butterfly», του dataset TokyoTech-31band.σελ.52

Σχήμα 3.7: Η πραγματική εικόνα MSF \hat{A} που κατασκευάστηκε βάση της πολυφασματικής λήψης «cloth3», του dataset TokyoTech-59band.σελ.53

Σχήμα 3.8: Η πραγματική εικόνα MSF \hat{A} που κατασκευάστηκε βάση της πολυφασματικής λήψης «Gualtar Columns», του dataset Manchester 2015.σελ.53

Σχήμα 3.9: Αποτελέσματα μετρικών PSNR και SSIM για το σετ δεδομένων Cave για δύο ενδεικτικές εικόνες.σελ.60

Σχήμα 3.10: Αποτελέσματα μετρικών PSNR και SSIM για το σετ δεδομένων Manchester 2015 για δύο ενδεικτικές εικόνες.σελ.61

Σχήμα 3.11: Αποτελέσματα μετρικών PSNR και SSIM για το σετ δεδομένων TokyoTech-31band για δύο ενδεικτικές εικόνες.σελ.63

Σχήμα 3.12: Αποτελέσματα μετρικών PSNR και SSIM για το σετ δεδομένων TokyoTech-59band για δύο ενδεικτικές εικόνες.σελ.64

Σχήμα 3.13: Απεικόνιση των μέσων τιμών των μεγίστων PSNR merged των εικόνων για καθένα από τα 4 σετ δεδομένων.σελ.65

Σχήμα 3.14: Απεικόνιση των μέσων τιμών των SSIM merged των εικόνων για καθένα από τα 4 σετ δεδομένων.σελ.66

Σχήμα 3.15: Διαφοροποιήσεις στη γραμματοσειρά (λέπτυνση γραμμών), εμφάνιση φαινομένου τύπου «hallo». Έγχρωμη εικόνα BMP (όπως έχει αποδοθεί από την MS εικόνα), MS πραγματική εικόνα όπως προέκυψε με το έγχρωμο σύνθετο που εστιάζει σε περιοχή ενδιαφέροντος, MS συνθετική εικόνα που εστιάζει σε περιοχή ενδιαφέροντος.σελ.69

Σχήμα 3.16: Διαφοροποιήσεις σε ακμές. Έγχρωμη εικόνα BMP (όπως έχει αποδοθεί από την MS εικόνα), MS πραγματική εικόνα όπως προέκυψε με το έγχρωμο σύνθετο που εστιάζει σε περιοχή ενδιαφέροντος, MS συνθετική εικόνα που εστιάζει σε περιοχή ενδιαφέροντος.σελ.70

Σχήμα 3.17: Δυσκολία αναπαράστασης λεπτών γραμμών, ευρύτερη επικράτηση θολότητας-εξομάλυνσης. Φυσικό έγχρωμο σύνθετο RGB, MS πραγματική εικόνα όπως προέκυψε με το έγχρωμο σύνθετο που εστιάζει σε περιοχή ενδιαφέροντος, MS συνθετική εικόνα που εστιάζει σε περιοχή ενδιαφέροντος.σελ.70

Σχήμα 3.18 Φυσικό έγχρωμο σύνθετο RGB, MS πραγματική εικόνα όπως προέκυψε με το έγχρωμο σύνθετο που εστιάζει σε περιοχή ενδιαφέροντος, MS συνθετική εικόνα που εστιάζει σε περιοχή ενδιαφέροντος.σελ.71

Σχήμα 4.1: Πολυφασματική απεικόνιση της πραγματικής (GT) εικόνας “Gualtar Villa” από το σετ δεδομένων «M15» (πάνω) και 4 τμήματα αυτής διάστασης 500x500 εικονοστοιχεία (κάτω).σελ.74

Σχήμα 4.2: Διάγραμμα που αναπαριστά την ομοιότητα μεταξύ 1) Bil. MD²IP -GT, 2) Bil. GT-GT & 3) MD²IP -GT μέσω της μετρικής PSNR.σελ.77

Σχήμα 4.3: Ζευγάρι παρεμβαλλόμενων πολυφασματικών εικόνων (GT- MD²IP) όπως προέκυψαν από τη διαδικασία της διγραμμικής παρεμβολής στην εικόνα «Fan» από το σύνολο δεδομένων «TT31».σελ.78

Σχήμα 4.4: Ζευγάρι παρεμβαλλόμενων πολυφασματικών εικόνων (GT- MD²IP) όπως προέκυψαν από τη διαδικασία της διγραμμικής παρεμβολής στην εικόνα «Chart» από το σύνολο δεδομένων «TT59».σελ.79

Σχήμα 4.5: Δομή AttR2U-Net δικτύου που χρησιμοποιήθηκε για την υλοποίηση του μοντέλου μας. Πηγή: [17].σελ.81

Σχήμα 4.6: Σχηματική αναπαράσταση της προτεινόμενης attention gate (AG) μεθοδολογίας που φιλτράρει τα χαρακτηριστικά που διαδίδονται μέσω των skip connections και σύνδεση της με το διάγραμμα των blocks του προτεινόμενου μοντέλου κατάτμησης Attention U-Net, κατά την ερευνητική μελέτη [16].σελ.82

Σχήμα 4.7: Αντιπαραβολή απόδοσης μεταξύ του συνόλου δεδομένων και του τυχαίου συνόλου αξιολόγησης, βάση της μετρικής PSNR σε τρεις κατηγορίες 1) Bil. MD²IP -GT, 2) Bil. GT-GT & 3) MD²IP -GT.σελ.86

Σχήμα 4.8: Διαγράμματα 1) Training & Validation Loss, 2) μετρικής PSNR για το 1ο μοντέλο.σελ.94

Σχήμα 4.9: Διαγράμματα 1) Training & Validation Loss, 2) μετρικής PSNR για το 2ο μοντέλο.σελ.96

Σχήμα 4.10: Διαγράμματα 1) Training & Validation Loss, 2) μετρικής PSNR για το 3ο μοντέλο.σελ.97

Σχήμα 4.11: Απόδοση δικτύου τύπου U-Net σε σχέση με τα δεδομένα εισόδου και το μοντέλο που εκπαιδεύτηκε.σελ.99

Σχήμα 4.12: Απόδοση Μοντέλων (M1, M2, M3) και MD²IP επί του συνόλου αξιολόγησης, ως προς τις μετρικές PSNR&SSIM.σελ.100

Σχήμα 4.13: Αντιπαραβολή πολυφασματικής εικόνας «Bom Jesus Bush _patch_1» σε RGB: (15, 7, 1) στα διαφορετικά στάδια επεξεργασίας του προτεινόμενου framework.....σελ.105

Σχήμα 4.14: Αντιπαραβολή πολυφασματικής εικόνας «Butterfly2_patch_0» σε RGB: (15, 7, 1) στα διαφορετικά στάδια επεξεργασίας του προτεινόμενου framework.σελ.106

Σχήμα 4.15: Αντιπαραβολή πολυφασματικής εικόνας «Chart _patch_0» σε RGB: (15, 7, 1) στα διαφορετικά στάδια επεξεργασίας του προτεινόμενου framework.σελ.107

Σχήμα 4.16: Αντιπαραβολή πολυφασματικής εικόνας «Egyptian statue _patch_0» σε RGB: (15, 7, 1) στα διαφορετικά στάδια επεξεργασίας του προτεινόμενου framework.σελ.108

Σχήμα 4.17: Αντιπαραβολή πολυφασματικής εικόνας «Bom Jesus Marigolds _patch_3» σε RGB: (15, 7, 1) στα διαφορετικά στάδια επεξεργασίας του προτεινόμενου framework.σελ.109

Ευρετήριο Πινάκων

Πίνακας 3.1: Πληροφορίες λήψης εικόνων για το σύνολο δεδομένων CAVE. (Πηγή: <https://www.cs.columbia.edu/CAVE/databases/multispectral/>).....σελ.46

Πίνακας 3.2: Ο πίνακας πολυφσματικού φίλτρου διάταξης 4x4 (MSFA) της IMEC.σελ.51

Πίνακας 4.1.: Τα δεδομένα εκπαίδευσης που χρησιμοποιούνται για καθένα από τα 3 μοντέλα που θα υλοποιηθούν.σελ.84

Πίνακας 4.2: Υπερπαράμετροι που χρησιμοποιήθηκαν κατά την εφαρμογή του 1ου πειράματος για το 1ο μοντέλο.σελ.90

Πίνακας 4.3: Υπερπαράμετροι που χρησιμοποιήθηκαν κατά την εφαρμογή του 2ου πειράματος για το 1ο μοντέλο.σελ.90

Πίνακας 4.4: Υπερπαράμετροι που χρησιμοποιήθηκαν κατά την εφαρμογή του 2ου πειράματος για το 1ο μοντέλο.σελ.90

Πίνακας 4.5: Υπερπαράμετροι που χρησιμοποιήθηκαν κατά την εφαρμογή του 1ου πειράματος για το 2ο μοντέλο.σελ.91

Πίνακας 4.6: Υπερπαράμετροι που χρησιμοποιήθηκαν κατά την εφαρμογή του 1ου πειράματος για το 3ο μοντέλο.σελ.92

Πίνακας 4.7: Σύγκριση σε διαφορετικά στάδια επεξεργασίας των δεδομένων (validation set), σύμφωνα με την μετρική PSNR.σελ.102

Πίνακας 4.8: Σύγκριση σε διαφορετικά στάδια επεξεργασίας των δεδομένων (validation set), σύμφωνα με την μετρική SSIM.σελ.102

Συντομογραφίες

ΔΕ	Διπλωματική Εργασία
ΝΔ	Νευρωνικό Δίκτυο
FCN	Fully Connected Network
FCN	Fully Convolutional Network
MS	Multispectral
MSFA	Multispectral Filter Array
PPI	Pseudo-Panchromatic Image
CNN	Convolutional Neural Network
PSNR	Peak Signal-to-Noise Ratio
SSIM	Structural Similarity Index Measure
DIP	Deep Image prior
MD ² IP	Multispectral Demosaicing with Deep Image Priors
BFA	Bayer filter array
CFA	Color Filter Array
SR	Super-Resolution

ΚΕΦΑΛΑΙΟ 1^ο

1.1. Εισαγωγή στο αντικείμενο μελέτης

Η ανακατασκευή εικόνας (Image Restoration) είναι μία από τις βασικότερες προκλήσεις που αντιμετωπίζονται στην επεξεργασία εικόνων. Αποσκοπεί στην ανάκτηση ή βελτίωση της ποιότητας μίας εικόνας που έχει υποβαθμιστεί λόγω πολλών παραγόντων όπως ο θόρυβος, η παραμόρφωση ή το θόλωμα. Στόχος της συγκεκριμένης εργασίας είναι η κατά τον δυνατόν καλύτερη ανακατασκευή της αρχικής-καθαρής εικόνας. Αυτό μπορεί να επιτευχθεί με διάφορες τεχνικές αποκατάστασης όπως το φιλτράρισμα, η κανονικοποίηση, η αποσυνέλιξη (deconvolution), και φυσικά με σύγχρονες μεθόδους μηχανικής και βαθιάς μάθησης. Ανάλογα την εφαρμογή επιλέγεται και η κατάλληλη τεχνική με στόχο τη μείωση θορύβου, τη διόρθωση χρωμάτων, τη συμπλήρωση και επισκευή κατεστραμμένων τμημάτων, την ενίσχυση λεπτομερειών κ.α.. Τέλος, όπως γίνεται σαφές η διαδικασία αυτή βρίσκει εφαρμογή σε πολυάριθμους τομείς από τη φωτογραφία, την ιατρική και δορυφορική απεικόνιση μέχρι την εγκληματολογία.

Στην παρούσα διπλωματική εργασία θα αξιοποιηθεί το πεδίο έρευνας της όρασης υπολογιστών από τη σκοπιά της απομοσαϊκοποίησης (demosaiicing) εικόνων από αισθητήρες επίπεδης διάταξης φίλτρων (MSFA) με τη χρήση Συνελκτικών Νευρωνικών Δικτύων (Convolutional Neural Networks ή CNNs). Τα CNNs επιλέχθηκαν έναντι των παραδοσιακών μεθόδων παρεμβολής καθώς είναι πολύ πιο αποτελεσματικά (λιγότερα σφάλματα (artifacts)) και μαθαίνουν καλύτερα πολύπλοκα μοτίβα και σχέσεις σε δεδομένα εικόνας. Επίσης, εκπαιδεύονται για να προβλέψουν τα πλήρη φασματικά δεδομένα από υποδειγματοληπτικές μετρήσεις που έχει καταγράψει ο αισθητήρας. Μία τέτοια προσέγγιση αναμένεται να έχει ιδιαίτερο ενδιαφέρον μιας και συνδυάζει τις τεχνολογίες απεικόνισης με προηγμένες μεθόδους μηχανικής μάθησης. Ακόμα πρόκειται για έναν επιστημονικό κλάδο που έχει απασχολήσει αρκετά και κατά το παρελθόν όπως αποδεικνύουν και οι επιστημονικές μελέτες [1], [3] που μελετήθηκαν βιβλιογραφικά. Επιπρόσθετα, το γεγονός ότι το αντικείμενο μελέτης έχει λειτουργικό χαρακτήρα σε τόσο ευρύ εφαρμογές· από καθημερινές όπως η ψυχαγωγία έως και πιο εντατικές σε επίπεδο βιομηχανικής κλίμακας, το καθιστά επιπλέον ελκυστικό κεντρίζοντας το προσωπικό μου ενδιαφέρον.

Αφορμή ενασχόλησης με το συγκεκριμένο αντικείμενο της διπλωματικής μου εργασίας αποτέλεσε η ερευνητική μελέτη [3] που ήταν υπό εξέλιξη από την ομάδα του εργαστηρίου τηλεπισκόπησης της ΣΑΤΜ – ΜΓ. Η ομάδα αυτή ασχολήθηκε με το demosaicing των MSFA εικόνων σε ένα σύνολο δεδομένων το «TokyoTech-31band», που η ανάλυση του συμπεριλαμβάνεται στην παρούσα ΔΕ. Χάρη σε αυτή την ερευνητική εργασία συνειδητοποιήσα την προσφορά μίας εργασίας demosaicing καθώς πρόκειται για ένα πολύ καλό εργαλείο προεπεξεργασίας εικόνων χρήσιμο σε μελλοντικούς ερευνητές πολλών επιστημονικών πεδίων.

Στόχος της παρούσας ΔΕ είναι όπως προαναφέρθηκε η απομοσαϊκοποίηση MSFA εικόνων με σκοπό την ανάκτηση πολυφασματικών εικόνων πλήρους ανάλυσης. Αυτό επιτυγχάνεται με δύο μεθόδους. Αρχικά, εφαρμόστηκε η μεθοδολογία demosaicing της ερευνητικής δημοσίευσης [3] -αναλύεται εκτενώς στην υποενότητα 2.2.3.- που στηρίζεται σε ένα μη επιβλεπόμενο μοντέλο και βασίζεται στη λογική των Deep Image Priors (DIP). Το μοντέλο αυτό βασίζεται στην εγγενή ιδιότητα των συνελκτικών δικτύων να λειτουργούν σαν «image priors» και να παράγουν, χωρίς εκπαίδευση σε άλλα δεδομένα, μια αξιόπιστη ανακατασκευή ή αποκατάσταση της εικόνας. Παράγωγα αυτής της διαδικασίας, στην περίπτωση της απομοσαϊκοποίησης είναι οι full resolution demosaiced εικόνες, με μοναδικό μειονέκτημα τον εκτεταμένο χρόνο βελτιστοποίησης που απαιτείται για την ανακατασκευή της εκάστοτε εικόνας. Ακολούθως, οι ανακατασκευασμένες εικόνες (όπως προέκυψαν από το MD²IP) θα αποτελούν τα «label» για το νέο δίκτυο τύπου U-Net που χρησιμοποιήσαμε και σαν δεδομένα εισόδου θα εκχωρούνται οι παρεμβαλλόμενες εικόνες που θα στηρίζονται στις πρώτες. Με το δίκτυο αυτό επιδιώκεται η εκπαίδευση και επιτάχυνση της διαδικασίας βελτιστοποίησης που υλοποιεί το MD²IP. Συνεπώς, παράγωγα του δικτύου μας είναι και πάλι οι πολυφασματικές εικόνες πλήρους ανάλυσης που θα μιμούνται τον τρόπο ανακατασκευής του MD²IP μοντέλου.

1.2. Κύριες προκλήσεις

Μία από τις πρώτες προκλήσεις που κλήθηκα να αντιμετωπίσω είναι η εύρεση κατάλληλων σετ δεδομένων που να απαρτίζονται από πολυφασματικές εικόνες που καταγράφουν στο ηλεκτρομαγνητικό φάσμα 470-630nm. Το εύρος αυτό μας ενδιαφέρει καθώς είναι αυτό που καλύπτει η XIMEA xiSpec “snapshot mosaic” υπερφασματική κάμερα με αισθητήρα απεικόνισης της IMEC, η οποία θα αποτελέσει το πρότυπο

απεικόνισης πολυφασματικών εικόνων της μορφής 4x4 MSFA που θα προσομοιάσουν τα δεδομένα μας.

Ένα άλλο ζήτημα που απασχόλησε αφορούσε την επιλογή της διαμόρφωσης των δεδομένων ώστε να είναι αξιοποιήσιμα από το νευρωνικό δίκτυο το οποίο θα επιλεγόταν. Πιο συγκεκριμένα, με απασχόλησαν τα κατάλληλα στάδια προεπεξεργασίας στο σύνολο των δεδομένων που απαιτούνταν για να διατρεχθούν και τα δύο μοντέλα που εφαρμόστηκαν. Ειδικότερα, δοκίμασα διαφορετικά format των εικόνων και ακόμα πειραματίστηκα με διαφορετικές εκδοχές των αρχικών και παράγωγων δεδομένων (βλ. υποκεφάλαια §3.2 και §4.1) προκειμένου τα μοντέλα μου να αποδώσουν κατά το βέλτιστο δυνατό. Εν προκειμένω, είναι σημαντικό να σημειωθεί ότι λόγω του μικρού όγκου δεδομένων που αντλήθηκαν και της δυσκολίας δημιουργίας νέων δεδομένων εκπαίδευσης από τα διαθέσιμα δεδομένα (data augmentation) -με στόχο τη μέγιστη δυνατή γενίκευση σε δεδομένα ελέγχου και νέα δεδομένα- οι εικόνες μας διακρίθηκαν μόνο σε σετ εκπαίδευσης και αξιολόγησης και όχι ελέγχου. Τα δεδομένα αξιολόγησης αυτά αξιοποιήθηκαν στην αμερόληπτη αξιολόγηση της προσαρμογής του μοντέλου αλλά και στην εκτίμηση των πραγματικών επιδόσεων πρόβλεψης του μοντέλου, κάτι τέτοιο βέβαια σε μία εργασία βελτιστοποίησης (που στοχεύει στην ελαχιστοποίηση του σφάλματος στο σετ εκπαίδευσης) είναι σύνηθες.

Ακόμη σε δεύτερο επίπεδο αναζητήθηκε η κατάλληλη αρχιτεκτονική που να επιταχύνει τη διαδικασία βελτιστοποίησης που υλοποιεί το μοντέλο που στηρίζεται στο MD²IP. Σε αυτό το σημείο είναι σημαντικό να υπογραμμιστεί ότι δεν υπάρχει κάποιο προεκπαιδευμένο δίκτυο με έτοιμα - αρχικοποιημένα βάρη, που θα μπορούσε να χρησιμοποιηθεί σαν “backbone” στο ίδιο πρόβλημα στην παρούσα φάση. Για αυτό το λόγο αρχικά χρησιμοποιήθηκε ένα απλό μοντέλο αρχιτεκτονικής τύπου U-net όπου δοκιμάστηκαν πληθώρα υπερπαραμέτρων και έγιναν κάποιες στοχευμένες αλλαγές ως προς τη δομή του που θεωρήθηκαν ότι θα βελτιώσουν την απόδοση του. Παρά τις εντατικές αυτές προσπάθειες όμως το μοντέλο φάνηκε να μην εκπαιδεύεται σε ικανοποιητικό βαθμό επιτυγχάνοντας ενδεικτικά μετρικές κατά το στάδιο του validation -όπου αξιολογείται το μοντέλο και βελτιστοποιούνται οι υπερπαραμέτροι του- που δεν ξεπερνούσαν την PSNR≈25, που όπως θα αναφερθεί και στην πορεία δεν συνιστά ικανοποιητική τιμή. Συνεπώς, η παραπάνω δυσκολία απόδοσης του μοντέλου με ώθησε σε εκτενέστερη αναζήτηση μίας αρχιτεκτονικής που θα υπόσχεται καλύτερη απόκριση στα δεδομένα αξιολόγησης. Έτσι, λοιπόν, κατέληξα στη χρήση ενός πιο προηγμένου μοντέλου τύπου U-net που χρησιμοποιεί attention και recurrent blocks. Κλείνοντας, σε αυτή την περίπτωση όμοια με τις προηγούμενες φορές ακολουθήσε μία σειρά από

πειράματα όπου έγιναν ποικίλες αλλαγές τόσο ως προς τα δεδομένα εισόδου του δικτύου αλλά και ως προς τις υπερπαραμέτρους του.

1.3. Συνεισφορά

Η φύση του αντικειμένου ενασχόλησης της απομοσαϊκοποίησης δηλαδή της διαδικασίας ανακατασκευής μίας πλήρως έγχρωμης εικόνας από ελλιπή έγχρωμα δείγματα που εξάγονται από ένα αισθητήρα εικόνας μπορεί να προεκταθεί ως ένα πρόβλημα απόκτησης εικόνων υψηλής ανάλυσης από αρχικά δεδομένα χαμηλότερης (super resolution). Αυτό ερμηνεύεται από τη σκοπιά ότι τα εικονοστοιχεία στην πρώτη περίπτωση του demosaicing να μην έχουν χωρική συσχέτιση μεταξύ τους (εν αντιθέσει με το super resolution) αλλά για την ανακατασκευή της εικόνας σε πλήρη χωρική ανάλυση απαιτείται η συμπλήρωση της πληροφορίας στα ενδιάμεσα pixels που είναι άγνωστη. Μία τέτοια λοιπόν επίλυση είναι εμφανές ότι έχει άμεση συνεισφορά σε πληθώρα επιστημονικών πεδίων που διαχειρίζονται δεδομένα εικόνας και βίντεο. Η σημαντικότερη ίσως προσφορά της παρούσας εργασίας έγκειται στη διαχείριση πολυφασματικών δεδομένων που συνεπάγονται πολύ μεγαλύτερη πολυπλοκότητα από ότι οι συνήθεις προσεγγίσεις που αφορούν την ανακατασκευή εικόνων από αισθητήρα που βασίζεται στο Bayer Filter Array (BFA) σε τρία μόλις κανάλια. Μάλιστα, η προτεινόμενη μεθοδολογία πλεονεκτεί καθώς δεν εξαρτάται από ένα συγκεκριμένο MSFA ground truth data pattern και επιπλέον είναι ασυσχέτιστη με τα μήκη κύματος που καταγράφουν τα εκάστοτε κανάλια. Συνεπώς, με υπόδειγμα τη δική μας επεξεργασία και ανάλυση των δεδομένων που ακολουθούν τη δομή ενός 4x4 MSFA μπορούν μελλοντικοί ερευνητές να γενικεύσουν ακόμα και σε υπερφασματικά δεδομένα.

Παράλληλα, αντλήθηκαν και επεξεργάστηκαν με ενιαίο τρόπο 4 διαφορετικά μεταξύ τους σύνολα δεδομένων τα οποία προσομοίωσαν τη δομή 4x4 MSFA εικόνων όπως θα προέκυπταν από τη XIMEA xiSpec “snapshot mosaic”, δηλαδή μίας υπερφασματικής κάμερας με ευρύ εφαρμογή στον κλάδο της επεξεργασίας εικόνων. Επομένως, υλοποιήθηκε ένα σετ δεδομένων όμοιο με αυτά που θα παρήγαγαν πολυφασματικοί αισθητήρες επίπεδης διάταξης φίλτρων που μπορεί να αξιοποιηθεί σαν στάδιο προεπεξεργασίας για μελλοντική εκπαίδευση μοντέλων διαφορετικών αρχιτεκτονικών για να βελτιώσουν την ανάλυση των εικόνων. Κλείνοντας, η παρούσα εργασία συνεισφέρει επιστημονικά παρέχοντας ένα συνολικό μεθοδολογικό πλαίσιο (framework) που στοχεύει στον απομοσαϊσμό ή όπως θα αναφέρεται στο εξής στο

demosaiicing και παρουσιάζει πολύ υποσχόμενα αποτελέσματα βάση της ποσοτικής και ποιοτικής του αξιολόγησης.

1.4. Διάρθρωση Διπλωματικής Εργασίας

Τα κύρια μέρη της διπλωματικής εργασίας αναπτύσσονται στα Κεφάλαια που ακολουθούν και πιο συγκεκριμένα:

- Στο 2^ο Κεφάλαιο παρουσιάζεται το επιστημονικό υπόβαθρο που είναι χρήσιμο για την εισαγωγή σε βασικές έννοιες επί του αντικείμενου ενασχόλησης και ορισμένες από τις ερευνητικές εργασίες που μελετήθηκαν και ήταν καταλυτικές για την εξέλιξη της παρούσας εργασίας.
- Στο 3^ο Κεφάλαιο αναφέρονται τα χαρακτηριστικά από τα σύνολα δεδομένων που έχουν ανακτηθεί καθώς και τα στάδια προεπεξεργασίας τους ώστε να αποτελέσουν τα δεδομένα αληθείας για το 1^ο μοντέλο που διατρέχεται, που στηρίζεται στο Deep Image Prior (DIP). Ακολουθεί ανάλυση της μεθόδου και αξιολόγηση αυτής με ποσοτικό και ποιοτικό τρόπο.
- Το 4^ο Κεφάλαιο ασχολείται με την μετέπειτα επεξεργασία των παράγωγων δεδομένων όπως προέκυψαν από την 1^η μεθοδολογία (MD²IP). Έπειτα, αναπτύσσεται η υλοποίηση του μοντέλου μας τύπου U-Net -που συνιστά τη 2^η μεθοδολογία- και συνοδεύεται με την απαραίτητη ποιοτική και ποσοτική αξιολόγηση του. Παράλληλα συγκρίνεται η επίδοση του με αυτή που θα είχε το 1^ο μοντέλο σε σχέση με τα πραγματικά δεδομένα και γίνεται μία γενική αποτίμηση, όπου εκεί έγκειται η βασικότερη προσφορά της παρούσας εργασίας.
- Το 5^ο και τελευταίο κεφάλαιο απαρτίζεται από τα συμπεράσματα που εξάγονται, περιπτώσεις αστοχίας, προοπτικές βελτίωσης και μελλοντικούς στόχους.

ΚΕΦΑΛΑΙΟ 2^ο

2.1. Επιστημονικό / Θεωρητικό Υπόβαθρο

2.1.1. Εισαγωγή στη Μηχανική Μάθηση

Ως τεχνητή νοημοσύνη (AI) μπορεί να οριστεί το πεδίο στο οποίο αναπτύσσονται συστήματα και μηχανές που υλοποιούν εργασίες που υπό κανονικές συνθήκες προϋποθέτουν ανθρώπινη νοημοσύνη ή αλλιώς κάθε τεχνική που επιτρέπει ένα πρόγραμμα Η/Υ να μιμείται την ανθρώπινη συμπεριφορά. Οι κύριοι τομείς στους οποίους απευθύνεται το AI αφορούν την επίλυση προβλημάτων όπως η ταξινόμηση, η παλινδρόμηση, η ομαδοποίηση, η πρόβλεψη (forecasting), η λήψη αποφάσεων, η ανίχνευση αντικειμένων κ.α. . Αποτελείται από υποπεδία όπως η μηχανική και βαθιά μάθηση, η όραση υπολογιστών, η ρομποτική κ.α. όμως στη προκειμένη διπλωματική εργασία θα ασχοληθούμε μόνο με τις δύο πρώτες υποκατηγορίες.

Η Μηχανική Μάθηση (ML) όπως αποδόθηκε από τον Arthur Samuel (1959) “συνιστά πεδίο μελέτης που δίνει στους υπολογιστές την ικανότητα να μαθαίνουν χωρίς να έχουν ρητά προγραμματιστεί”. Αξιοποιεί μοτίβα και χαρακτηριστικά που υπολογίζονται από τα δεδομένα εισόδου (input) τα οποία μεταφράζονται μέσω κατάλληλων αλγορίθμων για την υλοποίηση των εκάστοτε εργασιών και παράλληλα έχει τη δυνατότητα να βελτιώνει την απόδοση των αποτελεσμάτων με την πάροδο του χρόνου.

Η μηχανική μάθηση μπορεί να κατηγοριοποιηθεί στις εξής κατηγορίες ανάλογα με τον τρόπο λειτουργίας των αλγορίθμων:

- *Επιβλεπόμενη Μάθηση:* Η μέθοδος εκπαίδευσης στηρίζεται σε δεδομένα εισόδου όπου κάθε παρατήρηση έχει αντιστοιχιστεί σε μία κατηγορία (κλάση). Αυτά ονομάζονται και δεδομένα αληθείας (groundtruth data). Με αυτόν τον τρόπο αναπτύσσεται το μοντέλο κατά το οποίο ο αλγόριθμος μαθαίνει να αντιστοιχίζει δεδομένα εισόδου σε προκαθορισμένα δεδομένα εξόδου (labels-κλάση) και στη συνέχεια να κάνει προβλέψεις σε νέες παρατηρήσεις. Το σύνολο των δεδομένων αληθείας διακρίνεται σε δύο κατηγορίες τα δεδομένα εκπαίδευσης που συνιστούν ~70% και τα δεδομένα ελέγχου ~30%. Δύο ευρείς εφαρμογές

της επιβλεπόμενης μάθησης αποτελούν η ταξινόμηση (classification) - όπου ο αλγόριθμος προβλέπει μία ετικέτα για κάθε στοιχείο εισόδου- και η παλινδρόμηση (regression) - όπου ο αλγόριθμος προβλέπει συνεχείς αριθμητικές τιμές-.

- Μη Επιβλεπόμενη Μάθηση: Σε αυτήν την περίπτωση το μοντέλο δεν δέχεται προσημασμένο -σε κατηγορίες- dataset. Αντί αυτού ο αλγόριθμος εκπαιδεύεται προσπαθώντας χωρίς ρητή καθοδήγηση από τον χρήστη να αντιληφθεί τη “βασική” δομή και τα μοτίβα των δεδομένων εισόδου. Αυτού του είδους η μάθηση χρησιμοποιείται σε διαδικασίες ομαδοποίησης (clustering) -όπου ο αλγόριθμος ομαδοποιεί παρόμοιες παρατηρήσεις- και μείωσης διάστασης (dimensionality reduction) -όπου ο αλγόριθμος μειώνει τον αριθμό των χαρακτηριστικών ενώ παράλληλα διατηρεί τα απαραίτητα χαρακτηριστικά τους-. Τέλος, η μη επιβλεπόμενη ταξινόμηση μπορεί να αποτελέσει χρήσιμο εργαλείο καθώς μπορεί να υποδείξει τον βέλτιστο αριθμό κλάσεων πριν χρησιμοποιηθεί μία επιβλεπόμενη ταξινόμηση
- Ημί-Επιβλεπόμενη Μάθηση: Η συγκεκριμένη μέθοδος εκπαίδευσης όπως υποδηλώνει και το όνομα της χρησιμοποιεί ως δεδομένα εισόδου τόσο με ετικέτα (label) όσο και χωρίς, δηλαδή αποτελεί κάτι σαν υβριδική τεχνική μεταξύ των δύο παραπάνω μεθόδων μηχανικής μάθησης. Αναλυτικότερα, ο αλγόριθμος εκπαιδεύεται με μικρό πλήθος δεδομένων που έχουν αντιστοιχιστεί σε κλάσεις και με αρκετό μεγαλύτερο όγκο δεδομένων που δεν έχουν κατηγοριοποιηθεί. Τα μεν πρώτα καθοδηγούν την εκπαιδευτική διαδικασία του μοντέλου ενώ τα δεύτερα του παρέχουν την “ελευθερία” να αναγνωρίσει στοιχεία δομής και μοτίβα που εντοπίζονται στα δεδομένα. Τέλος, η ημι-επιβλεπόμενη μάθηση φαίνεται ιδιαίτερα χρήσιμη σε περιπτώσεις όπου η απόκτηση επαρκούς ποσότητας προσημασμένων δεδομένων είναι απαγορευτική οικονομικά αλλά και για λόγους δυσκολίας ενώ η απόκτηση δεδομένων χωρίς label είναι αρκετά πιο εφικτή. &
- Ενισχυτική Μάθηση: Αποτελεί έναν τύπο μηχανικής μάθησης που βασίζεται σε συνεχείς αποφάσεις και αλληλεπιδράσεις με το περιβάλλον. Προσομοιάζει κατά μία έννοια τα στάδια ανάπτυξης του

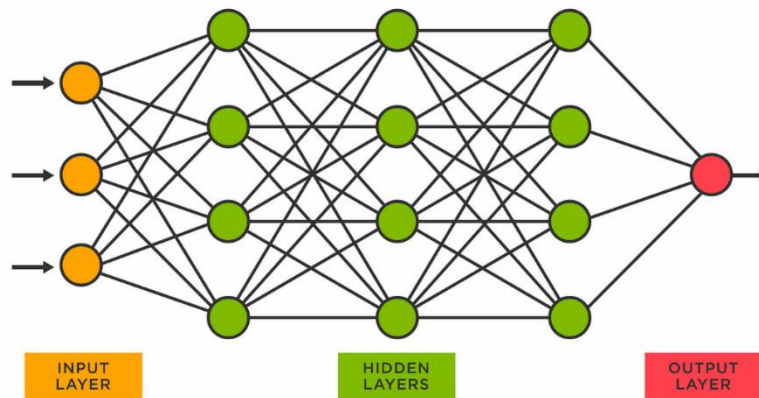
ανθρώπου καθώς μιμείται τη διαδικασία εκμάθησης trial and error. Πρόκειται, δηλαδή, για μία “αυτό-εκπαιδευόμενη” μέθοδο που στοχεύει μέσω παρατηρήσεων στις καλύτερες δυνατές αποφάσεις προκειμένου να λάβει τις περισσότερες “επιβραβεύσεις” και τις λιγότερες “τιμωρίες” αντίστοιχα. Η συγκεκριμένη μέθοδος εκπαίδευσης έχει εφαρμοστεί ευρέως σε κλάδους ρομποτικής, ηλεκτρονικών παιχνιδιών και αυτόματης οδήγησης που χαρακτηρίζονται από ένα περιβάλλον που μεταβάλλεται συνεχώς.

2.1.2. Εισαγωγή στη Βαθιά Μάθηση

Η βαθιά μάθηση αποτελεί με τη σειρά της υποπεδίο της μηχανικής μάθησης και στηρίζεται στην εκμάθηση αναπαραστάσεων υψηλού επιπέδου. Το βασικό «δομικό» της στοιχείο είναι ο νευρώνας ή perceptron που πρακτικά υλοποιεί μία δυαδική ταξινόμηση και αποτελεί το απλούστερο νευρωνικό δίκτυο (ΝΔ). Οι νευρώνες αυτοί είναι οργανωμένοι σε στρώσεις που διακρίνονται σε νευρώνες εισόδου (input layer), ένα ή περισσότερα κρυφά επίπεδα (hidden layer) και νευρώνες εξόδου (output layer), υλοποιώντας ένα τεχνητό νευρωνικό δίκτυο. Το πλήθος των νευρώνων ανά στρώση μπορεί να είναι διαφορετικό. Επίσης, οι κρυφές στρώσεις είναι τα ενδιάμεσα στάδια μεταξύ input και output που γίνονται οι υπολογισμοί και η εξαγωγή χαρακτηριστικών από τα δεδομένα και παράλληλα αυξάνουν την τάξη μη γραμμικότητας του δικτύου.

Κύριο χαρακτηριστικό της βαθιάς μάθησης είναι η χρήση τεχνητών νευρωνικών δικτύων (ANNs) που συνιστούν υπολογιστικά μοντέλα που έχουν εμπνευστεί από βιολογικά νευρωνικά δίκτυα. Με άλλα λόγια αποτελείται από πολλαπλά επίπεδα διασυνδεδεμένων νευρώνων, όπου σε κάθε επίπεδο εξάγεται πληροφορία και μεταβιβάζεται στο επόμενο επιτρέποντας με αυτό τον τρόπο στο μοντέλο να εκπαιδευτεί στην εξαγωγή χαρακτηριστικών. Έτσι λοιπόν, γίνεται λόγος για μία πλήρως συνδεδεμένη στρώση (layer) ή όπως ορίζεται από τη βιβλιογραφία Fully Connected ή dense. Κάποια από τα μεγάλα πλεονεκτήματα των συγκεκριμένων δικτύων αφορούν την διαλειτουργικότητα τους μέσα από την μεταφορά μάθησης εντός του ίδιου χώρου προτύπων ή και τον επανασυντονισμό (fine tuning) ώστε να προσαρμοστούν τα προεκπαιδευμένα δίκτυα σε διαφορετικούς αλλά εφάμιλλους χώρους προτύπων σε σχέση με το αρχικό δίκτυο. Τέλος, τα ΝΔ χρησιμοποιούνται ευρέως σε προβλήματα ταξινόμησης και ομαδοποίησης δεδομένων, σε περιπτώσεις παλινδρόμησης κ.α..

Ένα απλοϊκό ΝΔ αναπαρίσταται και στο Σχήμα 2.1 που αποτελείται από κόμβους όπου καθένας αντιστοιχεί σε ένα νευρώνα. Συγκεκριμένα, οι πορτοκαλί είναι νευρώνες εισόδου, οι πράσινοι κρυφά επίπεδα και ο ροζ νευρώνας εξόδου ενώ παρατηρείται ότι τα επίπεδα είναι πλήρως συνδεδεμένα. Εδώ έχει ενδιαφέρον να αναφερθεί ότι κάθε βέλος μεταξύ δύο επιπέδων αναπαριστά ένα βάρος και ότι όλα τα βέλη που παρουσιάζονται θα μπορούσαν να γραφούν σε ένα κοινό πίνακα βαρών καθώς αποτελούν ένα γραμμικό μετασχηματισμό μεταξύ δύο διανυσμάτων με τις αντίστοιχες διαστάσεις. Το μη γραμμικό κομμάτι του συστήματος είναι στην έξοδο των κόμβων.



Σχήμα 2.1: Αναπαράσταση ενός Τεχνητού Νευρωνικού Δικτύου (ANN).
(Πηγή: <https://medium.com/mit-6-s089-intro-to-quantum-computing/quantum-neural-networks-7b5bc469d984>).

Κάποιες γνωστές αρχιτεκτονικές βαθιάς μάθησης είναι τα Convolutional Neural Networks (CNNs) για εργασίες που αφορούν την επεξεργασία εικόνων, τα Recurrent Neural Networks (RNNs) για διαδοχική επεξεργασία δεδομένων και τα Generative Adversarial Networks (GANs) για τη δημιουργία συνθετικών δεδομένων, τα οποία τείνουν να αντικατασταθούν από τα Diffusion Models. Τέλος, η βαθιά μάθηση πλεονεκτεί μιας και μέσω αυτής μπορούν να χρησιμοποιηθούν μεγαλύτερα σετ δεδομένων αλλά και δίκτυα (περισσότερες παράμετροι/βάρη). Συνεπώς, μπορούν να αναπαρασταθούν πιο αντιπροσωπευτικά σύνθετα ζητήματα της πραγματικότητας - εισάγοντας την έννοια της μη γραμμικότητας- επιτυγχάνοντας μάλιστα πολύ υψηλά επίπεδα ακρίβειας, αν και σε σχέση με μικρότερα δίκτυα η εκπαίδευση τείνει να είναι πιο χρονοβόρα.

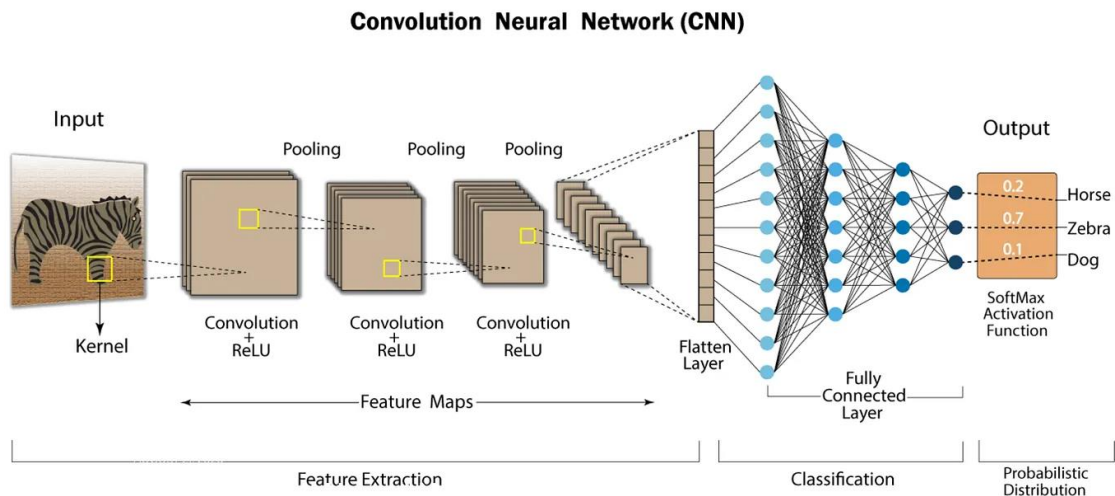
Στην παρούσα εργασία όπου τα προς μελέτη δεδομένα αφορούν εικόνες θα δοθεί έμφαση στο πεδίο της Υπολογιστικής Όρασης (Computer Vision, CV) που προσπαθεί να κατανοήσει και να ερμηνεύσει τέτοιου είδους πληροφορία. Μάλιστα, η όραση υπολογιστών στοχεύει στη δημιουργία «τεχνητής αντίληψης» (machine perception) μέσω της αξιοποίησης ψηφιακών δεδομένων εικόνων και βίντεο. Η βαθιά μάθηση υλοποιεί τέτοιου είδους εργασίες με τη συνδρομή των Συνελικτικών Νευρωνικών Δικτύων (CNNs), όπως προαναφέρθηκαν.

2.1.3. Συνελικτικά Νευρωνικά Δίκτυα (CNNs)

Τα Συνελικτικά Νευρωνικά Δίκτυα αποτελούν μία κλάση των δικτύων βαθιάς μάθησης και χρησιμοποιούνται σε εφαρμογές της όρασης υπολογιστών που αφορούν τον εντοπισμό αντικειμένων, την ταξινόμηση και την αναγνώριση-τμηματοποίηση εικόνων (image recognition-segmentation) και γενικότερα σε περιπτώσεις όπου τα δεδομένα εισόδου αποτελούν εικόνες. Περιέχουν πολλές στρώσεις συμπεριλαμβανομένου συνελικτικές (convolutional layers), συγκέντρωσης (pooling layers), και πλήρως συνδεδεμένες (fully connected layers). Αναλυτικότερα, οι συνελικτικές στρώσεις των CNNs βρίσκονται ανάμεσα στα hidden layers και δέχονται σαν είσοδο τανυστές 2ης ή 3ης τάξης, αν πρόκειται για greyscale και RGB εικόνα αντίστοιχα, και εξάγουν πληροφορία υψηλού επιπέδου μέσω πινάκων χαρακτηριστικών. Στο τέλος κάθε συνελικτικής στρώσης προστίθεται ο δείκτης μεροληψίας (bias) και εφαρμόζεται μία μη γραμμική συνάρτηση ενεργοποίησης με πιο συνηθισ τις ReLU, SoftMax, Sigmoid κ.α.. Οι τελευταίες προσθέτουν μη γραμμικότητα στο μοντέλο και το εκπαιδεύουν σε σύνθετα μοτίβα δεδομένων.

Σε ένα CNN έπειτα από μία συνελικτική στρώση ακολουθεί ένα επίπεδο συγκέντρωσης (pooling layer) που μειώνει τις χωρικές διαστάσεις κάθε «συνελιγμένης εικόνας» (feature map) μειώνοντας το μήκος και το πλάτος της διατηρώντας όμως ίδιο τον αριθμό των καναλιών. Με αυτό τον τρόπο επιδιώκεται μείωση του overfitting και του υπολογιστικού κόστους συντηρώντας την πλούσια σημασιολογική πληροφορία. Το τμήμα του δικτύου στο οποίο διατρέχονται τα επίπεδα συνέλιξης και pooling συνιστούν το τμήμα εξαγωγής χαρακτηριστικών. Αυτό αφού «ισοπεδωθεί» (δηλαδή μετατραπεί από τανυστή 3^{ης} σε 1^η τάξη) διαδέχεται μία ή περισσότερες Πλήρως Συνδεδεμένες (Fully Connected) στρώσεις, που αποτελούν το τμήμα ταξινόμησης. Στόχος των συγκεκριμένων στρώσεων είναι να εκπαιδεύσουν περαιτέρω το μοντέλο σε μη γραμμικούς συνδυασμούς των δεδομένων και να τα ταξινομήσουν τελικά σε

διαφορετικές κατηγορίες. Ο αριθμός των κλάσεων ταυτίζεται αριθμητικά με τον αριθμό των κόμβων στη στρώση εξόδου. Οι κόμβοι αυτοί τοποθετούνται προς κατάταξη στο τελευταίο τμήμα πιθανοτικής κατανομής (probabilistic distribution) εφαρμόζοντας τη συνάρτηση ενεργοποίησης Softmax, όπως αναπαρίσταται στο παράδειγμα του Σχήματος 2.2.



Σχήμα 2.2: Τυπική δομή ενός Συνελικτικού Νευρωνικού Δικτύου.

(Πηγή: <https://nafizshahriar.medium.com/what-is-convolutional-neural-network-cnn-deep-learning-b3921bdd82d5>).

2.1.4. Εκπαίδευση & Βελτιστοποίηση (Training & Optimization)

Η εκπαίδευση σε ένα μοντέλο μηχανικής μάθησης ερμηνεύεται ως η διαδικασία κατά την οποία ένα δίκτυο μαθαίνει να εντοπίζει και να αναγνωρίζει μοτίβα και να κάνει προβλέψεις βασισμένα στα δεδομένα εισόδου. Σαν είσοδο δέχεται συνήθως ένα σύνολο δεδομένων που αποτελείται από «ζευγάρια» στοιχεία εισόδου-εξόδου, από τα οποία μέσω επαναληπτικών διαδικασιών προσπαθεί να αντιληφθεί τη συσχέτιση τους μέσω μοτίβων-χαρακτηριστικών. Στόχος είναι μέσω μιας επαναληπτικής διαδικασίας εκμάθησης forward-backward (back propagation) να εκτιμηθούν εκείνες οι παράμετροι (βάρη (weights) και biases) του νευρωνικού δικτύου που θα επιτυγχάνουν το μικρότερο κόστος ώστε να είναι σε θέση να κάνει εύστοχες προβλέψεις σε νέα δεδομένα που δεν έχει ξαναδεί. Η διαδικασία αυτή συνεχούς εκπαίδευσης του μοντέλου ώστε να επιτυγχάνει όλο και καλύτερες προβλέψεις γίνεται με τη χρήση αλγορίθμων βελτιστοποίησης. Συνεπώς, γίνεται σαφές ότι η βελτιστοποίηση αποτελεί κρίσιμο

στοιχείο κατά την εκπαιδευτική διαδικασία καθώς καθορίζει πόσο αποτελεσματικά θα μαθαίνει το μοντέλο από τα δεδομένα.

Η βελτιστοποίηση συνδέεται άρρηκτα με τη διαχείριση περίπλοκων και μη γραμμικών χώρων καθώς είναι υπεύθυνη για τη προσαρμογή των παραμέτρων ενός νευρωνικού δικτύου -με τη χρήση μη γραμμικών συναρτήσεων ενεργοποίησης (activation functions)- ώστε να μπορούν να κατανοήσουν και να αναπαραστήσουν σύνθετα μοτίβα δεδομένων. Ένα από τα πρώτα στάδια στη βελτιστοποίηση είναι ο προσδιορισμός μίας συνάρτησης κόστους που θα καταγράφει πόσο καλά αποδίδει το μοντέλο. Επομένως, στόχος είναι η ελαχιστοποίηση του σφάλματος ή αλλιώς η εύρεση του ολικού ελαχίστου ή κάποιου τοπικού ελαχίστου της συνάρτησης κόστους. Όταν αναφερόμαστε, σε σφάλμα εννοείται η απόκλιση των προβλέψεων του μοντέλου σε σχέση με τα δεδομένα αληθείας (ground truth). Λόγω του συνδυασμού μη γραμμικών συναρτήσεων ενεργοποίησης που χρησιμοποιούνται, η συνάρτηση κόστους τείνει να μην είναι κυρτή (non-convex) στην γενική περίπτωση, δηλαδή να παρουσιάζει πολλά τοπικά ελάχιστα και σημεία «σέλλας», όπου η παράγωγος δηλαδή είναι ίση με μηδέν.

Προκειμένου οι αλγόριθμοι βελτιστοποίησης να μπορούν να διαχειριστούν τέτοιου είδους χώρους χρησιμοποιούν τεχνικές όπως η κανονικοποίηση, προσαρμοστικοί ρυθμοί εκμάθησης, momentum κ.α.. Ο πιο κοινός αλγόριθμος είναι ο Gradient Descent που προσαρμόζει τις παραμέτρους του μοντέλου με βάση την κατεύθυνση της παραγώγου της συνάρτησης κόστους (εύρεση διεύθυνσης της μεγαλύτερης κλίσης) μέχρις ότου αυτή να μηδενιστεί, να επέλθει σύγκλιση του δικτύου και οι παράμετροι να αποκτήσουν την βέλτιστη τιμή τους. Άλλοι ευρέως γνωστοί αλγόριθμοι βελτιστοποίησης είναι οι (Stochastic) Gradient Descent (SGD) και οι παραλλαγές του Adam, RMSprop, AdaGrad που προσαρμόζουν τον βαθμό εκμάθησης κάθε παραμέτρου σύμφωνα με προγενέστερες κλίσεις, γεγονός που μπορεί να βελτιώσει τον ρυθμό σύγκλισης και την ευρωστία του μοντέλου.

Ακόμα, στα πλαίσια της βελτιστοποίησης εφαρμόζονται τεχνικές κανονικοποίησης (πχ. Early Stopping, L1, L2 regularization, Dropout, Data Augmentation κ.α.) που χρησιμοποιούνται για να αποφευχθούν «αχρείαστα» σύνθετα μοντέλα, δηλαδή για την αποφυγή της υπερπροσαρμογής (overfitting) του μοντέλου στα δεδομένα -κατά την οποία το μοντέλο προσαρμόζεται όχι μόνο στα δεδομένα εκπαίδευσης αλλά και στο θόρυβο που αυτά συνεπάγονται-. Αυτό επιτυγχάνεται μέσω επιβολής «ποιμών» στη συνάρτηση κόστους. Με αυτόν τον τρόπο το μοντέλο μαθαίνει πιο απλοϊκά μοτίβα και μειώνει την πιθανότητα προσαρμογής του μοντέλου στον θόρυβο που περιέχεται στα δεδομένα εκπαίδευσης. Έτσι επιτρέπεται στο μοντέλο η

μέγιστη δυνατή γενίκευση στα δεδομένα ελέγχου και σε νέα δεδομένα. Σχετικά με το Early Stopping (πρόωρη διακοπή) πρόκειται για μία άλλη τεχνική που χρησιμοποιείται για να αποφευχθεί το overfitting και αποσκοπεί στην ελαχιστοποίηση του σφάλματος γενίκευσης. Συγκεκριμένα, επεμβαίνει στο στάδιο εκπαίδευσης και σταματάει τη λειτουργία του μοντέλου όταν η απόδοση του validation set ξεκινά να μειώνεται, γεγονός που υποδεικνύει ότι το μοντέλο κάνει overfit στα δεδομένα εκπαίδευσης. Πρόκειται για μία υπερπαράμετρο του δικτύου που ορίζει το πλήθος των εποχών μετά από τις οποίες θα σταματήσει η εκπαίδευση του, μία διαδικασία που γίνεται χειροκίνητα ή αυτόματα.

Μία ακόμα υπερπαράμετρος του δικτύου αποτελεί ο βαθμός εκμάθησης ή αλλιώς learning rate που ορίζει το βήμα με το οποίο αλλάζουν οι παράμετροι του δικτύου σε κάθε επανάληψη στη μέθοδο καταβιβασμού κλίσης (Gradient Descent). Πρακτικά, υποδεικνύει το ρυθμό εκμάθησης του μοντέλου. Η επιλογή της κατάλληλης τιμής δεν είναι προκαθορισμένη -αν και υπάρχει μια τυπική τιμή 10^{-3} - 10^{-4} - και είναι κρίσιμη για την επιτυχία της διαδικασίας βελτιστοποίησης. Αν ο βαθμός εκμάθησης είναι μεγάλος τότε η βελτιστοποίηση μπορεί να αποκλίνει σημαντικά από το ολικό ελάχιστο και να υπάρχει ένα φαινόμενο «ταλάντωσης» ενώ αν είναι πολύ μικρή η σύγκλιση θα είναι αργή κάτι που συνεπάγεται μεγάλο υπολογιστικό κόστος.

Συνοψίζοντας, το στάδιο εκπαίδευσης περιλαμβάνει τη συνολική διαδικασία εκμάθησης ενός μοντέλου να κάνει προβλέψεις και η βελτιστοποίηση αποτελεί αναπόσπαστο κομμάτι αυτής προσπαθώντας να εντοπίσει τον καλύτερο συνδυασμό παραμέτρων που θα ελαχιστοποιήσουν τη συνάρτηση κόστους. Αυτό επιτυγχάνεται προσαρμόζοντας σε κάθε επανάληψη τις παραμέτρους με τέτοιο τρόπο ώστε η απόδοση του μοντέλου διαρκώς να βελτιώνεται. Συνεπώς, γίμεται σαφές ότι η κατάλληλη επιλογή αλγορίθμου βελτιστοποίησης και υπερπαραμέτρων θα παίξουν καθοριστικό ρόλο στην ταχύτητα σύγκλισης του δικτύου και την ανθεκτικότητα του μοντέλου. Τέλος να σημειωθεί ότι αυτές οι δύο έννοιες της βελτιστοποίησης και της μηχανικής μάθησης έχουν διαφορετικό στόχο καθώς η πρώτη στοχεύει στην ελαχιστοποίηση του σφάλματος στο σετ εκπαίδευσης (συνήθως δεν ορίζεται σετ αξιολόγησης/ελέγχου) ενώ η δεύτερη αποσκοπεί επιπλέον και στην ταυτόχρονη μείωση της διαφοράς στο σφάλμα μεταξύ του σετ εκπαίδευσης και του σετ ελέγχου.

2.1.5. Μετρικές Αξιολόγησης

Στην βαθιά μάθηση υπάρχει πληθώρα διαφορετικών τρόπων αξιολόγησης των μοντέλων ανάλογα την εφαρμογή. Ενδεικτικά παραδείγματα αποτελούν οι μετρικές της ακρίβειας (accuracy, precision), το recall, average precision (AP) που χρησιμοποιούνται για προβλήματα ταξινόμησης, παλινδρόμησης, ανίχνευσης αντικειμένων κ.α.. Αυτές και πολλές άλλες μετρικές αξιολόγησης μπορούν να εφαρμοστούν για την εκτίμηση του μοντέλου αλλά για την παρούσα εργασία δεν αποτελούν τον βέλτιστο τρόπο αξιολόγησης. Για αυτό το λόγο αναζητήθηκαν δείκτες που ενδείκνυνται για την επεξεργασία εικόνων και συγκρίνουν ποιοτικά και δομικά χαρακτηριστικά μεταξύ μιας εικόνας αναφοράς και της αντίστοιχης επεξεργασμένης-ανακατασκευασμένης εικόνας, όπως είναι η μετρική PSNR και SSIM.

➤ PSNR metric

Στο πεδίο της επεξεργασίας εικόνων (κατ' επέκταση και βίντεο) και των αλγορίθμων συμπίεσης συχνά χρησιμοποιείται η μετρική Peak Signal-to-Noise Ratio (PSNR) που μετρά την αναλογία μεταξύ της μέγιστης δυνατής ισχύος ενός σήματος και της ισχύος του αλλοιωτικού θορύβου, που επηρεάζει την ποιότητα αναπαράστασης του. Εν προκειμένω, η μετρική PSNR υπολογίζει τη διαφορά στις ψηφιακές τιμές των pixel μεταξύ της εικόνας αναφοράς (ground truth image) και της υπολογισμένης εικόνας (output) όπως αυτή έχει προκύψει από την εκάστοτε διαδικασία επεξεργασίας. Οι τυπικές τιμές του PSNR κυμαίνονται από 30 σε 50 db, όπου όσο υψηλότερη η τιμή τόσο καλύτερη η ποιότητα της ανακατασκευασμένης εικόνας. Ενδεικτικά αναφέρεται ότι τιμές <20 db θεωρούνται μη αποδεκτές ενώ >40 db είναι εξαιρετικές.

Ο τύπος σύμφωνα με τον οποίο υπολογίζεται η μετρική PSNR είναι ο ακόλουθος:

$$PSNR = 10 \log_{10} \frac{HWK}{\|x - \tilde{x}\|^2} ,$$

όπου x η εικόνα αναφοράς και \tilde{x} η υπολογισμένη εικόνα, H , W και K είναι ο αριθμός των γραμμών, στηλών και των καναλιών του x αντίστοιχα.

➤ SSIM metric

Σε διεργασίες που αφορούν την επεξεργασία των εικόνων όπου η ποιότητα των παραγόμενων δεδομένων είναι πρωταρχικής σημασίας όπως σε εφαρμογές συμπίεσης, αποθορυβοποίησης και ανακατασκευής εικόνας η μετρική Structural Similarity Index Measure (SSIM) είναι ευρέως διαδεδομένη. Πρακτικά πρόκειται για μία ακόμη μετρική που χρησιμοποιείται για να αξιολογήσει την ομοιότητα ως προς τα δομικά στοιχεία μεταξύ δύο σημάτων εικόνων. Η ποσοτική αξιολόγηση που γίνεται στην SSIM βασίζεται σε τρεις παραμέτρους την 1) φωτεινότητα, 2) την αντίθεση και 3) τη δομική πληροφορία των προς σύγκριση εικόνων. Θεωρητικά η εικόνα αναφοράς θα είναι αυτή που θα έχει τη βέλτιστη ποιότητα και η εικόνα που αντιπαρατίθεται θα είναι αυτή που θα έχει προκύψει από μία από τις παραπάνω εφαρμογές. Συνεπώς, η SSIM θα υποδείξει την ποιότητα του παράγωγου σήματος εικόνας. Κυμαίνεται σε τιμές μεταξύ -1 και 1, όπου 1 υποδεικνύει την τέλεια ομοιότητα. Ακόμα να σημειωθεί ότι η SSIM λόγω της δημοφιλίας της συχνά συγκρίνεται με άλλες μετρικές όπως οι MSE και PSNR. Αξίζει να υπογραμμιστεί ότι εν αντιθέσει με τη PSNR που εστιάζει σε διαφορές ανά pixel η SSIM λαμβάνει υπόψη συνολικά τη δομική πληροφορία των εικόνων καθώς η ιδέα πίσω από αυτό είναι ότι τα εικονοστοιχεία έχουν υψηλές αλληλεξαρτήσεις ειδικά όταν είναι κοντά χωρικά.

Ο τύπος σύμφωνα με τον οποίο υπολογίζεται η μετρική SSIM είναι ο ακόλουθος:

$$SSIM = \frac{(2\mu_{\bar{x}}\mu_x + k_1)(2\sigma_{\bar{x}x} + k_2)}{(\mu_{\bar{x}}^2 + \mu_x^2 + k_1)(\sigma_{\bar{x}}^2 + \sigma_x^2 + k_2)}$$

όπου $\mu_{\bar{x}}$, μ_x , $\sigma_{\bar{x}}^2$ και σ_x^2 είναι η μέση τιμή και η διασπορά της εικόνας αναφοράς και της υπολογισμένης εικόνας αντίστοιχα ενώ η $\sigma_{\bar{x}x}$ είναι η συνδιασπορά. Εφόσον οι εικόνες που επεξεργαζόμαστε είναι πολυφασματικές υπολογίζεται μία τιμή SSIM για κάθε κανάλι ξεχωριστά και τελικά παρουσιάζεται για κάθε εικόνα μία τιμή μετρικής που έχει προκύψει από τη μέση τιμή των επιμέρους.

2.1.6. MSFA

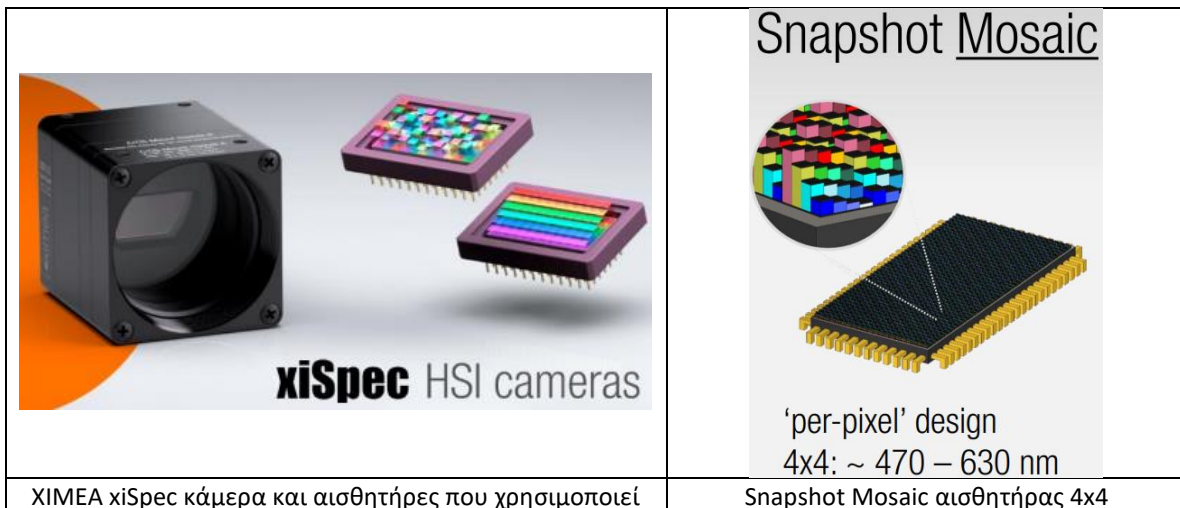
Παλαιότερα η λήψη πολυφασματικών δεδομένων γινόταν με αυτόματο ή χειροκίνητο τρόπο αντικαθιστώντας τα οπτικά φίλτρα στους αισθητήρες των καμερών, που αποτελούσε μία χρονοβόρα και περιοριστική διαδικασία μιας και η απόκτηση

δυναμικών σκηνών ήταν πολύ δύσκολη. Αυτό το πρόβλημα παρακάμφθηκε μέσω των single shot cameras που διέθεταν ένα σύστημα πολυφασματικών αισθητήρων επίπεδης διάταξης φίλτρων ή αλλιώς Multispectral Filter Array (MSFA) χάρη στο οποίο ήταν εφικτή η λήψη πολυφασματικής πληροφορίας ταυτόχρονα -εν αντιθέσει με το σύνηθες Bayer filter array (BFA) ικανό να ανιχνεύει μόνο τιμές ορατού φάσματος (RGB) σε κάθε pixel-. Αναλυτικότερα, ένα MSFA είναι συνήθως ένα τετραγωνικό πλέγμα $N \times N$ που προκύπτει ως ένα άθροισμα φασματικών φίλτρων που είναι κατανομημένα σύμφωνα με ένα επαναλαμβανόμενο χωρικά πρότυπο και βρίσκονται πάνω από ένα μονοχρωματικό αισθητήρα, όπως είναι ο CMOS. Τέτοιου είδους κάμερες, λοιπόν, παράγουν raw MSFA εικόνες, όπου σε κάθε εικονοστοιχείο έχει καταγραφεί η φασματική πληροφορία από ένα μόνο κανάλι και έχουν αγνοηθεί όλες οι άλλες.

Ένα τέτοιο χαρακτηριστικό παράδειγμα μοντέλου που παράγει MSFA εικόνες αποτελεί η XIMEA xiSpec “snapshot mosaic” υπερφασματική κάμερα (που απεικονίζεται στο Σχήμα 2.3) και χρησιμοποιεί τον αισθητήρα υπερφασματικής απεικόνισης της IMEC. Διαθέτει 4 αισθητήρες: Linescan και Snapshot Mosaic που στηρίζονται στους standard CMOS αισθητήρες, με τα υπερφασματικά φίλτρα να προστίθενται σε επίπεδο πλακέτας. Εξ αυτών η παρούσα εργασία εστιάζει στον αισθητήρα που καταγράφει με μοτίβο μωσαϊκού 4×4 που επαναλαμβάνεται διαρκώς στην επιφάνεια του αισθητήρα, όπως αναπαρίσταται στο Σχήμα 2.4. Ο πίνακας 4×4 διαθέτει 16 παρεμβαλόμενα φίλτρα, δηλαδή καταγράφει 16 φασματικά κανάλια στο εύρος 470-630 nm. Γενικά, εφαρμόζοντας φασματικά φίλτρα στενής ζώνης (narrow-band) σε επίπεδο εικονοστοιχείου η τεχνολογία της IMEC επιτρέπει λύσεις υπερφασματικής απεικόνισης με εξαιρετικά συμπαγείς διαστάσεις, χαμηλό βάρος, υψηλή αξιοπιστία και αποδεδειγμένη ικανότητα μαζικής παραγωγής με συγκριτικά χαμηλό κόστος και όγκο, κάτι που την κατατάσσει παγκοσμίως ως τη μικρότερη υπερφασματική κάμερα, ιδανική για ενσωματωμένα συστήματα όρασης [7].



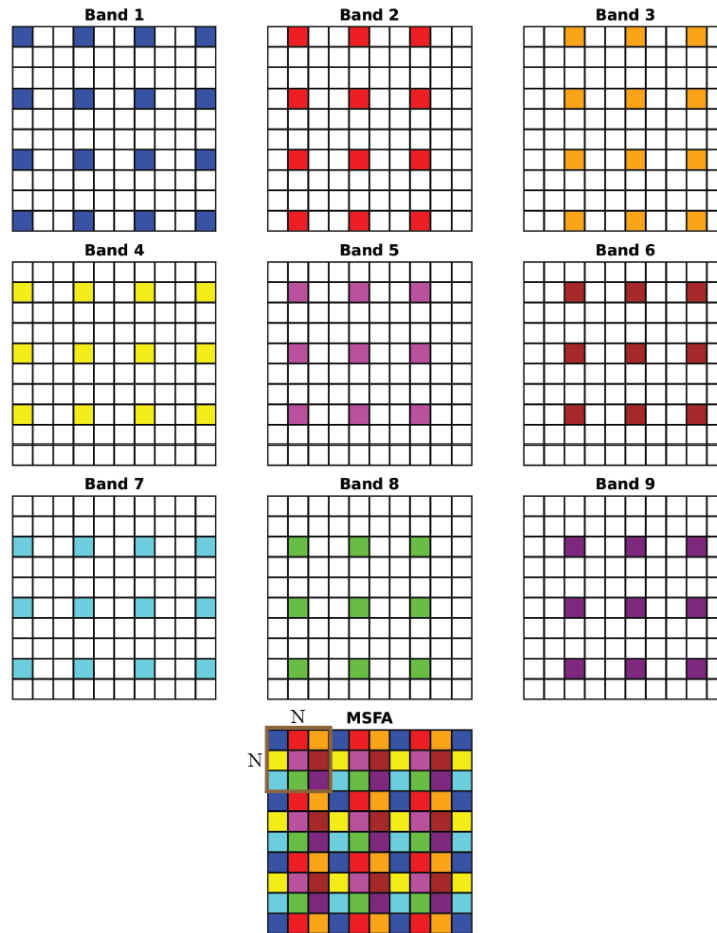
Σχήμα 2.3: XIMEA's xiQ USB3.0 camera στηρία ενσωματώθηκαν οι αισθητήρες υπερφασματικής απεικόνισης (HSI) της IMEC. (Πηγή: <https://www.ximea.com/en/usb3-vision-camera/xiq>).



Σχήμα 2.4: Υπερφασματική απεικόνιση χρησιμοποιώντας την XIMEA xiSpec κάμερα. (Πηγή: <https://www.ximea.com/support/attachments/5981/SpectroNet-2016-03-Ximea-V02.pdf>).

Στο Σχήμα 2.5 που ακολουθεί οπτικοποιείται η δημιουργία μίας εικόνας MSFA της μορφής $N \times N$, επιλέγοντας από τα κανάλια της πολυφασματικής εικόνας πλήρους ανάλυσης τα ri pixel εκείνα που ταιριάζουν στη θέση των αντίστοιχων καναλιών στον MSFA. Με άλλα λόγια η πληροφορία που συγκεντρώνεται για το εκάστοτε κανάλι από κάθε φίλτρο καταχωρείται από μία καθορισμένη κατανομή εικονοστοιχείων που διατηρούν μία τακτική απόσταση, αυτή δηλαδή που προβλέπεται από το μοτίβο MSFA που συνήθως πρόκειται για ένα τετραγωνικό πλέγμα $N \times N$. Συνεπώς κάθε εικόνα που αντιστοιχίζεται με ένα κανάλι έχει χωρικές διαστάσεις που είναι N φορές μικρότερες σε σχέση με την αρχική (raw) εικόνα. Σε επόμενο στάδιο, με μια απλή διαδικασία μπορεί

να απομονωθεί η φασματική πληροφορία ανά κανάλι με επαναληψιμότητα τέτοια που θα ορίζει το πρότυπο του MSFA (πχ. αν η κάμερα οπτικοποιεί με μοτίβο 4x4 τότε η πληροφορία που θα εξαχθεί για κάθε κανάλι θα επαναλαμβάνεται ανά 4). Με άλλα λόγια, όπως είναι προφανές, για να ανακατασκευαστεί η raw MSFA σε μία πολυφασματική πλήρης ανάλυσης εικόνα είναι απαραίτητη η διαδικασία του demosaicing.

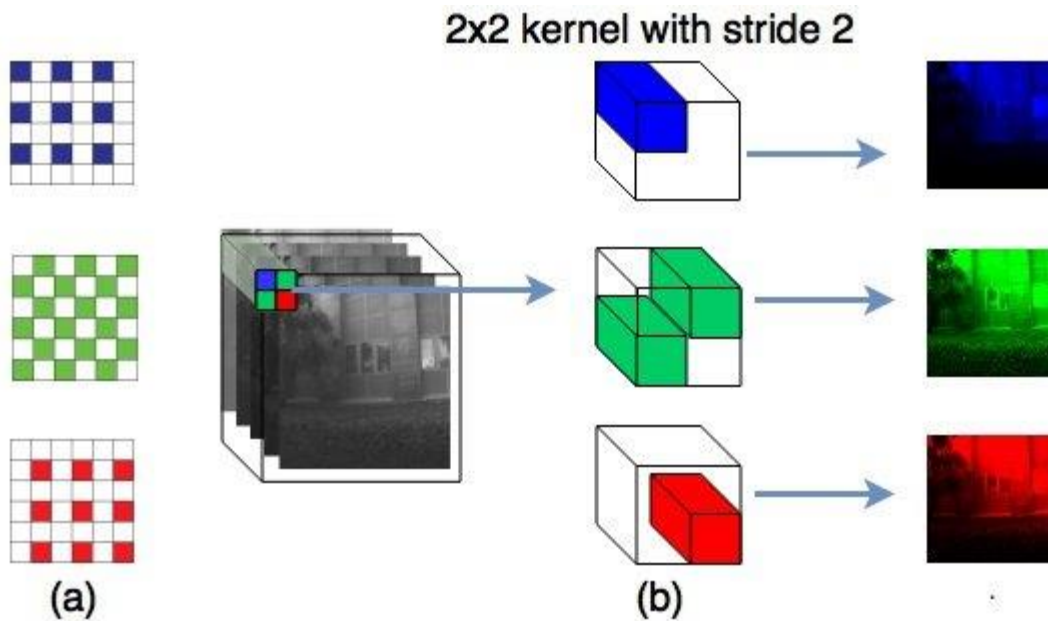


Σχήμα 2.5: Δημιουργία εικόνας MSFA της μορφής NxN, επιλέγοντας από τα κανάλια της πολυφασματικής εικόνας πλήρους ανάλυσης τα pixel εκείνα που ταιριάζουν στη θέση των αντίστοιχων καναλιών στον MSFA. Πηγή: [3].

2.1.7. Demosaicing

Με τον όρο demosaicing ορίζεται η διαδικασία ανακατασκευής μίας πλήρως έγχρωμης εικόνας από ελλιπή έγχρωμα δείγματα που εξάγονται από ένα αισθητήρα

εικόνας που επικαλύπτεται από μία συστοιχία έγχρωμων φίλτρων (color filter array ή CFA). Αυτή η διαδικασία είναι απαραίτητη στις ψηφιακές φωτογραφικές μηχανές και γενικά στις συσκευές απεικόνισης που χρησιμοποιούν έναν μόνο αισθητήρα εικόνας. Υπάρχει πληθώρα αλγορίθμων από απλούς μέχρι πιο σύνθετους άρα πιο υπολογιστικά απαιτητικούς. Η κατάλληλη επιλογή επηρεάζει την ποιότητα της εικόνας όσο και τις υπολογιστικές απαιτήσεις. Ο πιο χαρακτηριστικός και συνήθης τύπος CFA είναι το φίλτρο Bayer filter array (BFA), που αναπαρίσταται και στο Σχήμα 2.6, που διατάσσει τα χρώματα σύμφωνα με ένα συγκεκριμένο μοτίβο κόκκινου(25%), πράσινου (50%) και μπλε (25%). Εδώ να σημειωθεί ότι η υπερίσχυση των πράσινων εικονοστοιχείων αποδίδεται στην ευαισθησία του ανθρώπινου ματιού να αντιλαμβάνεται το πράσινο χρώμα, έτσι επιλέγεται να δοθεί υψηλότερη ανάλυση για το φωτεινό τμήμα της εικόνας. Οι εφαρμογές της συγκεκριμένης μεθόδου είναι ευρείς από τις ψηφιακές κάμερες που έχουν υιοθετήσει αισθητήρες απεικόνισης που χρησιμοποιούν Bayer filter, τα σύγχρονα «έξυπνα» κινητά τηλέφωνα που πέρα από demosaicing εφαρμόζουν επιπλέον βήματα μετά-επεξεργασίας για να βελτιώσουν της ποιότητα της εικόνας καθώς και στην επιστημονική απεικόνιση (σε τηλεσκόπια και μικροσκόπια) όπου η υψηλή ακρίβεια είναι ζητούμενο. Κλείνοντας, γίνεται σαφές ότι το demosaicing είναι μία ουσιαστική διαδικασία στην ψηφιακή απεικόνιση που μετατρέπει τα ακατέργαστα (raw) δεδομένα του αισθητήρα σε μία πλήρη και ακριβή έγχρωμη εικόνα.



Σχήμα 2.6: Η τυπική διάταξη συστοιχίας φίλτρων Bayer και ο πυρήνας συνέλιξης για τη σχεδίαση συστοιχίας φίλτρων 2x2 τύπου Bayer. Πηγή [5].

2.1.8. U-Net

Το UNet είναι ένα συνελκτικό νευρωνικό δίκτυο (CNN) που χρησιμοποιείται για εργασίες σημασιολογικής κατάτμησης στην επεξεργασία εικόνων και γενικότερα στο ευρύ πεδίο της όρασης υπολογιστών. Αναπτύχθηκε από τους Olaf Ronneberger, Philipp Fischer, and Thomas Brox το 2015 στο paper εν ονόματι "U-Net: Convolutional Networks for Biomedical Image Segmentation" για να χρησιμοποιηθεί για σημασιολογική κατάτμηση σε εικόνες βιοϊατρικής. Εδώ είναι σκόπιμο να σημειωθεί ότι στην περίπτωση της βιοϊατρικής δεν αρκεί μόνο η ανίχνευση μίας ασθένειας αλλά απαιτείται και ο εντοπισμός της περιοχής που παρουσιάζει την ανωμαλία. Αυτό επιτυγχάνεται μέσω του UNet καθώς υλοποιεί ταξινόμηση ανά εικονοστοιχείο της εικόνας. Επομένως, τα δεδομένα εισόδου και εξόδου έχουν το ίδιο μέγεθος.

Στο σημείο αυτό κρίνεται σκόπιμο να σημειωθεί ότι η Σημασιολογική Κατάτμηση (Semantic Segmentation) αποτελεί μία τεχνική του πεδίου της Όρασης Υπολογιστών και αφορά την κατάταξη ανά εικονοστοιχείο μίας εικόνας σε προκαθορισμένες κατηγορίες. Τέτοιου είδους κατηγορίες μπορεί να είναι άνθρωποι, αυτοκίνητα, κτίρια κ.α.. Επομένως γίνεται σαφής η σπουδαιότητα της σε τομείς όπως η αυτόνομη οδήγηση, ανάλυση δορυφορικών ακόμα και εικόνων ιατρικού περιεχομένου. Στόχος, λοιπόν, είναι η δημιουργία ενός πυκνού χάρτη τμηματοποίησης ανά pixel μίας εικόνας (μεγέθους συνήθως όσο η εικόνα εισόδου), με κάθε εικονοστοιχείο κατάλληλα ταξινομημένο σε μία συγκεκριμένη κλάση. Για αυτό το λόγο αποκαλείται συχνά και ως πυκνή πρόβλεψη «dense prediction». Συχνά μάλιστα τείνει να συγχέεται με την παραδοσιακή διαδικασία ταξινόμησης -που αφορά την απόδοση μίας κατηγορίας για κάθε εικόνα- ενώ στην πραγματικότητα αυτό που υλοποιεί μέσω της «κατάτμησης» σε πιο ουσιαστικά χαρακτηριστικά είναι να παρέχει μία πιο λεπτομερή ανάλυση και κατανόηση επί των δεδομένων της εικόνας. Μία τέτοια αρχιτεκτονική χρησιμοποιείται στον κλάδο της βαθιάς μάθησης και ιδίως με τη χρήση συνελκτικών δικτύων με μεγάλη επιτυχία με πιο αξιοσημείωτες εφαρμογές μεταξύ άλλων τα Πλήρως Συνδεδεμένα Δίκτυα ή Fully Convolutional Networks (FCN) και το U-Net, όπως θα αναλυθεί παρακάτω.

Λαμβάνοντας υπόψη όσα προαναφέρθηκαν γίνεται σαφές ότι το UNet ως CNN στοχεύει στη ταξινόμηση εικόνας -δηλαδή δέχεται σαν «είσοδο» εικόνα και εξάγει μία ετικέτα- και πιο συγκεκριμένα «μάσκες κατάτμησης» (segmentation masks/maps) όπου κάθε εικονοστοιχείο ταξινομείται σε μία από τις προκαθορισμένες κατηγορίες του δικτύου. Πιο συγκεκριμένα, σαν δεδομένα εισόδου λαμβάνει συνήθως εικόνες διαφόρων format όπως grayscale, RGB, πολυφασματικές, προεπεξεργασμένες (δηλαδή

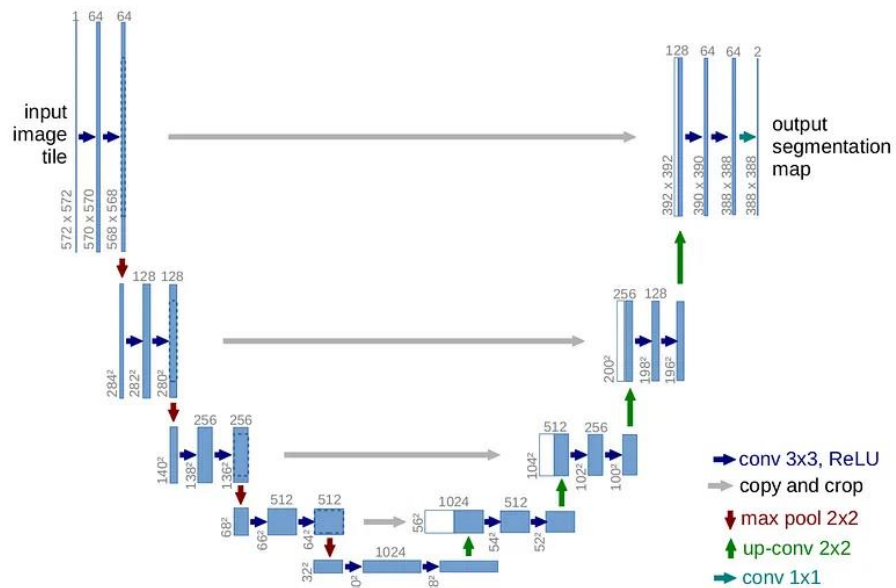
εικόνες που μπορεί να έχουν υποστεί κάποια επεξεργασία όπως κανονικοποίηση, αλλαγή διαστάσεων, augmentation) κ.α. ανάλογα την εκάστοτε περίπτωση.

Η δομή του έχει ένα χαρακτηριστικό σχήμα σαν “U” εξού και το όνομα του και πρόκειται για μία αρχιτεκτονική της μορφής encoder-decoder. Όσον αφορά το κομμάτι της αρχιτεκτονικής του αποτελείται από δύο «μονοπάτια» αυτά της συμπίεσης (contraction path) και της επέκτασης (expanding path) που καθένα από αυτά απαρτίζεται από τέσσερα (4) blocks που ενώνονται επιμέρους με skip connections. Αναλυτικότερα, πρόκειται για ένα Πλήρως Συνελικτικό Δίκτυο ή Fully Convolutional Network (FCN) που αποτελείται αποκλειστικά από συνελικτικά επίπεδα και δεν περιέχει καθόλου πυκνά (dense) στρώματα. Αυτό αποτελεί σκόπιμη επιλογή σχεδιασμού για να επιτρέπει στο δίκτυο να δέχεται εικόνες αυθαίρετων μεγεθών μιας και σε αντίθετη περίπτωση ένα πλήρως συνδεδεμένο δίκτυο ή Fully Connected Network (FCN) απαιτεί ένα προκαθορισμένο αριθμό νευρώνων και συνδέσεων μεταξύ των δεδομένων εισόδου και των νευρώνων, κάτι που προκαθορίζεται από το μέγεθος των δεδομένων εισόδου.

- Το πρώτο τμήμα συντελείται από τον **encoder** ή αλλιώς μονοπάτι συμπίεσης, κατά το οποίο γίνεται η εξαγωγή της πληροφορίας μέσω πινάκων χαρακτηριστικών (feature maps). Αποτελείται πρακτικά από διαδοχικά επίπεδα συνέλιξης (convolutional layers) και συγκέντρωσης (pooling layers) που σταδιακά αυξάνουν το βάθος της εικόνας (bands) και μικραίνουν τη χωρική ανάλυση της, αντίστοιχα. Γενικά και οι δύο διαδικασίες και κυρίως τα επίπεδα συγκέντρωσης έχουν ως αποτέλεσμα να μειώσουν το μέγεθος της εικόνας. Κοινώς κάνουν αυτό που ονομάζεται down sampling που πρακτικά σημαίνει μετάπτωση μίας εικόνας υψηλής σε χαμηλή ανάλυση. Με αυτόν τον τρόπο όσο «προχωράει» το δίκτυο πιο βαθιά παρότι το μέγεθος της εικόνας μικραίνει το receptive field (η περιοχή της εικόνας εισόδου που καλύπτει το φίλτρο (πχ. συνέλιξης) κάθε φορά που τη διατρέχει) αυξάνεται. Έτσι, γίνεται η εξαγωγή χαρακτηριστικών από διαφορετικά επίπεδα (levels of abstraction). Αντίστοιχα, όσο ο αριθμός των καναλιών ή αλλιώς το βάθος της εικόνας αυξάνει αυτό επιτρέπει την εξαγωγή πιο σύνθετων χαρακτηριστικών. Τέλος, μπορεί να προκύψει το συμπέρασμα ότι μέσω πράξεων υποδειγματολειψίας της εικόνας το μοντέλο μπορεί να αντιληφθεί τι υπάρχει στην εικόνα αλλά χάνει την χωρική διάσταση της πληροφορίας.
- Το δεύτερο τμήμα επιτελεί το ρόλο του **decoder** στο δίκτυο ή αλλιώς το συμμετρικό μονοπάτι επέκτασης που χρησιμοποιείται για τον ακριβή εντοπισμό

των προς μελέτη χαρακτηριστικών μέσω transposed convolutions σε συνδυασμό με κλασικές συνελίξεις. Με άλλα λόγια κάνει κατά έναν τρόπο την αντίστροφη διαδικασία από τον encoder καθώς εκτελεί αναδειγματοληψία (upscaling) στον χάρτη χαρακτηριστικών διατηρώντας τη χωρική πληροφορία με απώτερο σκοπό την ανακατασκευή της μάσκας κατάτμησης. Εδώ, λοιπόν, το βάθος της εικόνας μειώνεται και αυξάνονται οι διαστάσεις της. Αυτό το στάδιο είναι απαραίτητο καθώς όπως προαναφέρθηκε στην σημασιολογική κατάτμηση πέρα από την ανίχνευση χαρακτηριστικών μας ενδιαφέρει και η χωρική πληροφορία. Σύμφωνα με βιβλιογραφία η χωρική διάσταση μπορεί να ανακτηθεί με διάφορες τεχνικές up-sampling με την πιο σύγχρονη και έμπιστη να είναι τα επίπεδα αντίστροφης συνέλιξης (transposed convolutional layers). Αυτά λειτουργούν ακριβώς αντίθετα από τα κλασικά συνελικτικά και δέχονται σαν είσοδο εικόνα χαμηλής ανάλυσης και τη μετατρέπουν σε υψηλής.

- Εδώ να σημειωθεί ότι σε κάθε στάδιο του decoder χρησιμοποιούνται **skip connections**, σύμφωνα με τα οποία γίνεται «επίθεση» του αποτελέσματος που προκύπτει από τα επίπεδα αντίστροφης συνέλιξης με τους πίνακες χαρακτηριστικών όπως προκύπτουν από το αντίστοιχο στάδιο του encoder. Πιο αναλυτικά, ένα skip connection όπως το όνομα προδίδει παραλείπει ορισμένα από τα επίπεδα του νεωρωνικού δικτύου και τροφοδοτεί το output ενός επιπέδου του encoder στο input του αντίστοιχου επιπέδου του decoder, δηλαδή γίνεται μεταφορά αυτούσιων χαρακτηριστικών ανά κλίμακα. Παρατηρώντας το Σχήμα 2.7 που ακολουθεί τα skip connections αναπαρίστανται με βέλη. Συνεπώς, μέσω αυτών επιτυγχάνεται η άμεση μεταφορά χαρακτηριστικών από το ένα «μονοπάτι» στο άλλο που έχει σαν αποτέλεσμα την παραγωγή καλύτερων μασκών κατάτμησης (segmentation masks).



Σχήμα 2.7: Αρχιτεκτονική UNet μοντέλου.

(Πηγή: <https://towardsdatascience.com/unet-line-by-line-explanation-9b191c76baf5>).

2.1.9. Super Resolution

Το Super Resolution (SR) είναι η διαδικασία απόκτησης μίας ή περισσότερων εικόνων υψηλής ανάλυσης από μία ή περισσότερες παρατηρήσεις χαμηλής ανάλυσης [8]. Πρόκειται δηλαδή για μία τεχνική που χρησιμοποιείται στην επεξεργασία εικόνων και στοχεύει στην ενίσχυση της ανάλυσης μίας εικόνας, αυξάνοντας τόσο το μέγεθος (χωρική ανάλυση) όσο και τις λεπτομέρειες που απεικονίζει. Οι πρακτικές εφαρμογές που έχει βρει σε πραγματικά και σύγχρονα προβλήματα είναι πολυάριθμες, από δορυφορικές και εναέριες απεικονίσεις -για ποικίλες εφαρμογές όπως χαρτογράφηση, περιβαλλοντική παρακολούθηση-, στην επεξεργασία ιατρικών εικόνων (πχ. MRI, CT scans) για καλύτερες διαγνώσεις και αναλύσεις, σε βελτίωση βίντεο για ψυχαγωγικούς ρόλους ή για λόγους επιτήρησης από κάμερες ασφαλείας, σε ανάλυση και ενίσχυση εικόνας προσώπου, δακτυλικών αποτυπωμάτων, κειμένου, σε αντικείμενα συγκοινωνιακά και ασφαλείας όπως οι πινακίδες κυκλοφορίας για λόγους επιτήρησης κ.α.. Το SR δεν πρέπει να συγχέεται με παρόμοιες τεχνικές όπως η 1) παρεμβολή (interpolation) καθώς σε αυτή την περίπτωση οι λεπτομέρειες υψηλής συχνότητας δεν αποκαθίστανται, 2) η αποκατάσταση (restoration), όπου η ποιότητα της εξαγόμενης

πληροφορίας είναι βελτιωμένη αλλά το μέγεθος των input output εικόνων είναι το ίδιο και 3) η απόδοση εικόνας (image rendering).

Το SR μπορεί να διακριθεί σε δύο διαφορετικές τύπους τεχνικών: τη μονοεικονική ή Single-Image Super Resolution (SISR) και την πολυεικονική ή Multi-Image Super Resolution (MISR). Στην πρώτη περίπτωση γίνεται λόγος για ενίσχυση της ανάλυσης μίας εικόνας που μπορεί να επιτευχθεί με διάφορες μεθόδους από παραδοσιακούς αλγορίθμους μέχρι και προσεγγίσεις machine learning. Χαρακτηριστικά παραδείγματα των πρώτων είναι οι μέθοδοι παρεμβολής που είναι απλές τεχνικές όπως ο εγγύτερος γείτονας, η διγραμμική και δικυβική παρεμβολή που αυξάνουν το μέγεθος της εικόνας υπολογίζοντας τις τιμές των εικονοστοιχείων βάση των γειτονικών. Οι παραπάνω προσεγγίσεις είναι απλές αλλά συχνά οι παράγωγες σκηνές είναι θολές. Άλλο ένα παράδειγμα παραδοσιακών αλγορίθμων είναι οι προσεγγίσεις με βάση το μοντέλο όπως edge-preserving smoothing and total variation μέθοδοι, που χρησιμοποιούν μαθηματικά μοντέλα για την ανακατασκευή εικόνων υψηλής ανάλυσης αλλά είναι υπολογιστικά απαιτητικές και χωρίς απαραίτητα καλύτερα αποτελέσματα.

Όσον αφορά, τις προσεγγίσεις του machine learning -για την επίλυση προβλημάτων super resolution- διακρίνονται σε τρεις κατηγορίες:

- Convolutional Neural Networks (CNNs): Μοντέλα Deep Learning και πιο συγκεκριμένα CNNs όπως τα SRCNN (Super-Resolution Convolutional Neural Network), FSRCNN (Fast Super-Resolution Convolutional Neural Network) και VDSR (Very Deep Super Resolution) έχουν επιδείξει αξιόλογες επιδόσεις. Αυτά εκπαιδεύονται σε μεγάλα σετ δεδομένων από ζευγάρια χαμηλής και υψηλής ανάλυσης εικόνων για να προβλέπουν και να παράγουν σε επόμενο στάδιο ένα high resolution output από ένα low resolution input.
- Generative Adversarial Networks (GANs): Τα GANs όπως το ESRGAN (Enhanced Super-Resolution Generative Adversarial Network) χρησιμοποιούν ένα generator network για τη δημιουργία εικόνων υψηλής ανάλυσης και ένα discriminator network που διαχωρίζει τις πραγματικές εικόνες υψηλής ανάλυσης με αυτές που δημιουργήθηκαν από το δίκτυο. Αυτού του είδους η «ανταγωνιστική» εκπαίδευση συμβάλει στη δημιουργία πιο ρεαλιστικών και λεπτομερών εικόνων.
- Recurrent Neural Networks (RNNs): Τα RNNs χρησιμοποιούνται για να μοντελοποιήσουν τις εξαρτήσεις μεταξύ διαφορετικών τμημάτων της εικόνας, γεγονός που βοηθάει στη παραγωγή super-resolved εικόνων υψηλότερης ανάλυσης.

Τέλος, σχετικά με την τεχνική Multi-Image Super Resolution (MISR) καταφέρνει να συνδυάζει την πληροφορία από πολλές εικόνες της ίδιας σκηνής που λαμβάνονται από διαφορετική γωνία λήψης ή χρονικές στιγμές. Για την εν λόγω τεχνική επικαλούνται μεθοδολογίες όπως Image Registration που στοχεύει στην ευθυγράμμιση των εικόνων πριν από τον συνδυασμό τους και Fusion Methods όπως το φιλτράρισμα με μέσο όρο, διάμεση τιμή ή πιο σύνθετους αλγορίθμους σύντηξης/συγχώνευσης (fusion) για τη συγχώνευση της πληροφορίας.

Παρά το συνεχώς εξελισσόμενο πεδίο του super resolution με σημαντικές προόδους χάρη στη μηχανική μάθηση και τεχνικών βαθιάς μάθησης συναντώνται ορισμένες προκλήσεις που κάθε νέα μέθοδος καλείται να αντιμετωπίσει. Μία από αυτές είναι τα τεχνουργήματα ή «artifacts» που παρατηρούνται σε εικόνες υψηλής ανάλυσης. Ακόμα, έχει ενδιαφέρον ο βαθμός γενίκευσης του μοντέλου ώστε παρότι εκπαιδευμένο σε συγκεκριμένο dataset να μπορεί να ανταποκρίνεται καλά σε διάφορα είδη εικόνων. Η υπολογιστική απόδοση, δηλαδή η διαρκής βελτίωση της ταχύτητας και της αποτελεσματικότητας των αλγορίθμων SR, ώστε να μπορούν να εφαρμοστούν σε δεδομένα σε πραγματικό χρόνο, καθώς και η ενσωμάτωση και διαλειτουργικότητα του SR με άλλες AI τεχνολογίες όπως η παράλληλη ανίχνευση αντικειμένων αποτελούν μελλοντικούς στόχους και κατεύθυνση για το συγκεκριμένο κλάδο, που έχει έρθει για να δώσει λύσεις σε πολλούς τομείς βελτιώνοντας την ανάλυση της εικόνας.

2.2. Ανάλυση Ερευνητικών Εργασιών

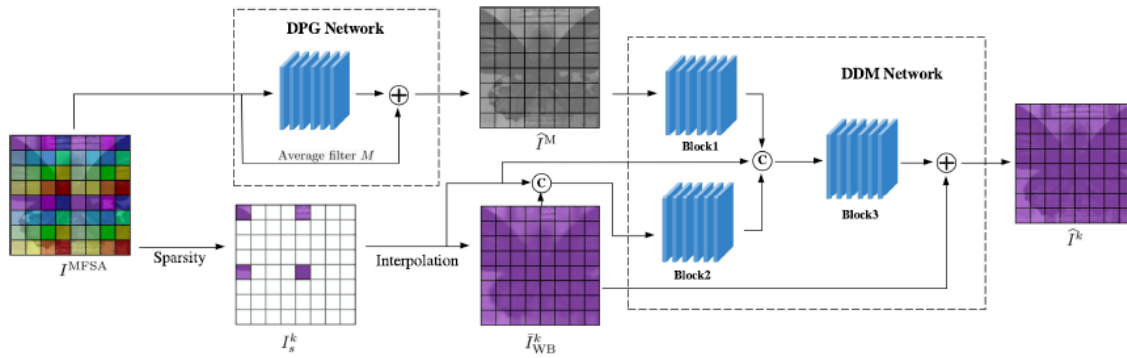
2.2.1. A Deep Joint Network for Multispectral Demosaicking Based on Pseudo-Panchromatic Images

Η παρούσα εργασία αναπτύσσει μία νέα μεθοδολογία απομοσαϊκοποίησης (demosaicing) πολυφασματικών εικόνων όπως έχουν προκύψει από μία single sensor camera. Όπως προαναφέρθηκε στην παράγραφο του θεωρητικού υποβάθρου οι εν λόγω κάμερες εξάγουν τη χωρική και ραδιομετρική πληροφορία με τη μορφή MSFA. Για το σκοπό αυτό λοιπόν αναπτύχθηκε μία αρχιτεκτονική CNN βαθιάς μάθησης που στηρίχτηκε στη μέθοδο των PPIs εν ονόματι Deep PPI Demosaicking Network (DPD-Net) που συνδυάζει τα οφέλη παραδοσιακών (πχ. PPI) και data-driven μεθοδολογιών (πχ. CNN) και υπερτερεί σε επίπεδο απόδοσης ως προς αυτές, σύμφωνα με ποιοτική και ποσοτική αξιολόγηση. Ακόμα, στο πλαίσιο της παρούσας εργασίας σχεδιάστηκε μία συνάρτηση κόστους, μία διαδικασία εκπαίδευσης και επεξεργασίας των δεδομένων - που πηγάζουν από τέσσερα dataset- προκειμένου να επιτευχθεί η διαδικασία του demosaicing.

Όσον αφορά την αρχιτεκτονική DPD-Net όπως αναφέρθηκε νωρίτερα συνδυάζει τα πλεονεκτήματα των PPIs και των μεθόδων μηχανικής μάθησης. Αναλυτικότερα, τα Pseudo-panchromatic images PPIs φαίνεται από βιβλιογραφική μελέτη πολύ χρήσιμα στη διαδικασία απομοσαϊκοποίησης καθώς ενσωματώνουν την πληροφορία υψηλών συχνοτήτων από όλα τα κανάλια και βοηθούν στην παρεμβολή καθενός subsampled band. Αντίστοιχα τα Συνελικτικά Νευρωνικά Δίκτυα (CNNs) συμβάλλουν στη βελτιστοποίηση των υπό μελέτη εικόνων λαμβάνοντας υπόψη τη μη γραμμικότητα. Σχετικά με τη δομή του δικτύου το DPD-Net αποτελείται από δύο υπό-δίκτυα το Deep PPI Generation Network (DPG-Net) και το Deep Demosaic Network (DDM-Net). Το πρώτο υπολογίζει τα PPIs \hat{I}^M ενώ το δεύτερο παράγει τις τελικές demosaicked εικόνες \hat{I}^k όπως έχουν προκύψει από τα εκτιμώμενα PPIs \hat{I}^M και τις αρχικές demosaicked εικόνες I_{WB}^k (που αποτελούν παράγωγο διαδικασίας παρεμβολής επί των αραιών μωσαϊκών εικόνων Iks).

Έπειτα κρίνεται σημαντικό να σχολιαστεί η αρχιτεκτονική που αναπτύχθηκε. Αναλυτικότερα, όπως προαναφέρθηκε το DPD-Net συντίθεται από δύο υπό-δίκτυα τα DPG-Net και DDM-Net. Το DPG-Net δέχεται σαν input την εκάστοτε MSFA εικόνα και σαν output βγάζει την ψευδο-παγχρωματική εικόνα PPI. Αντίστοιχα το DDM-Net δέχεται τρία inputs: (1) την εκτιμώμενη PPI (παράγωγο του DPG-Net), (2) την «αραιή

μωσαϊκή» εικόνα για το εκάστοτε κανάλι και (3) την προσωρινή (ενδιάμεση) demosaicked εικόνα (όπως αυτή έχει προκύψει από διγραμμική παρεμβολή από τη (2)). Το παράγωγο του DDM-Net είναι η τελική demosaicked εικόνα για κάθε κανάλι. Η παραπάνω αρχιτεκτονική του DPD-Net αναπαρίσταται στο Σχήμα 2.8.



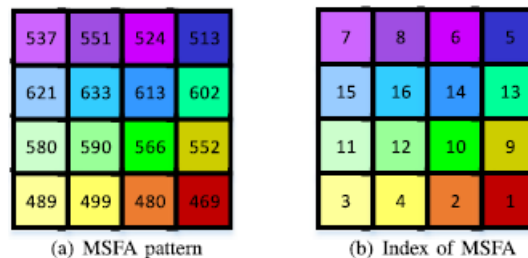
Σχήμα 2.8: I^{MSFA} , \hat{I}^M , I_s^k , \hat{I}_{WB}^k and \hat{I}^k συμβολίζονται ως η ακατέργαστη ψηφιδωτή MSFA εικόνα, εκτιμώμενη PPI, «αραιή μωσαϊκή» εικόνα, προσωρινή demosaicked εικόνα and τελική demosaicked εικόνα για κάθε κανάλι k , αντίστοιχα. Η ροή εργασιών του DPD-Net αποτελείται από δύο δίκτυα που ονομάζονται DPG-Net και DDM-Net, αντίστοιχα. Πηγή: [1].

Σε ένα από τα πρώτα στάδια γίνεται η παραδοχή ότι οι πληροφορίες υψηλής συχνότητας των PPIs έχουν υψηλή θετική συσχέτιση με τις demosaicked εικόνες, κάτι που επαληθεύεται για τα περισσότερα pixels. Μάλιστα, παρατηρείται ενδεικτικά μεταξύ του PPI και του 10^{00} καναλιού συντελεστής συσχέτισης ίσος με 0,944 και αποδεικνύεται περαιτέρω ότι τα εικονοστοιχεία με ισχυρή συσχέτιση είναι επιπλέον και χωρικά σχετιζόμενα. Καταληκτικά, οι PPI φαίνονται χρήσιμες για την ανακατασκευή των demosaicked εικόνων διερευνώντας τους συσχετισμούς από τα γειτονικά pixels.

Γενικά, η ψευδο-παγχρωματική εικόνα Pseudo-Panchromatic Image (PPI) προκύπτει ανά εικονοστοιχείο ως η μέση τιμή όλων των φασματικών καναλιών μίας πολυφασματικής εικόνας. Στην προκειμένη περίπτωση στο υπό-δίκτυο του DPG-Net υπολογίζεται αρχικά μία ενδιάμεση PPI εφαρμόζοντας ένα σταθμισμένο μέσο φίλτρο (M) συνελίσσοντας την raw mosaic image I^{MSFA} , η οποία ως είναι επόμενο έχει χάσει την πληροφορία ψηλής συχνότητας και αποτελεί μία πιο θολή αναπαράσταση της πραγματικής I^M . Για το σκοπό αυτό αναπτύχθηκε το DPG-Net, εμπνευσμένο από το Very Deep Super-Resolution (VDSR) Network, που χρησιμοποιήθηκε για να εξαχθούν τα χαρακτηριστικά της εικόνας επαναφέροντας την πληροφορία υψηλών συχνοτήτων. Αυτό επιτεύχθηκε χάρη σε ένα block convolutional layer. Τέλος, τα δύο παράγωγα συνδυάζονται προκειμένου να παραχθεί η τελική PPI εικόνα.

Παράλληλα, από την MSFA εικόνα μπορεί εύκολα να εξαχθεί για κάθε κανάλι η λεγόμενη *sparse mosaic image* που συντίθεται από τα εικονοστοιχεία για τα οποία έχει καταγραφεί τιμή για το εκάστοτε κανάλι ενώ για όλα τα άλλα δεν υπάρχει πληροφορία. Έπειτα, σε αυτή εφαρμόζεται διγραμμική παρεμβολή και παράγεται μία προσωρινή *demosaicked* εικόνα \bar{I}^k_{WB} , που παρουσιάζει *coherent artifacts*. Με απώτερο σκοπό να οξυνθούν οι ακμές της \bar{I}^k_{WB} και να μεταφερθεί η πληροφορία υψηλής συχνότητας των PPIs στην τελική *demosaicked* εικόνα κατασκευάστηκε το υπό—δίκτυο DDM-Net. Αυτό με τη σειρά του απαρτίζεται από 3 blocks -που συντίθενται από συνελκτικές στρώσεις- με τα δύο πρώτα να εξυπηρετούν στην εξαγωγή χαρακτηριστικών και χρωμάτων από τις \hat{I}^M και \bar{I}^k_{WB} . Τα χαρακτηριστικά αυτά συνενώνονται με την I^k_s και αποτελούν είσοδο για το τρίτο block, που υπολογίζει τη διαφορά μεταξύ της πραγματικής πολυφασματικής εικόνας \hat{I}^k και της εξομαλυμένης \bar{I}^k_{WB} . Η διαφορά αυτή προστίθεται στην \bar{I}^k_{WB} και έτσι προκύπτει η τελική εικόνα \hat{I}^k . Είναι σημαντικό να σημειωθεί εδώ ότι αυτή η ροή εργασιών αποτελεί μία μονοκάναλη διαδικασία που επαναλαμβάνεται k φορές, δηλαδή όσες τα κανάλια που συνθέτουν την υπό μελέτη πολυφασματική εικόνα.

Στην εργασία αυτή χρησιμοποιήθηκαν 4 σετ πολυφασματικών δεδομένων και αυτά είναι τα CAVE, TokyoTeck-31band (TT31), TokyoTeck-59band (TT59) και το University of Manchester 2015 (M15) εκ των οποίων το TT31 αξιοποιήθηκε ως δεδομένα ελέγχου (35 εικόνες) και τα υπόλοιπα σαν δεδομένα εκπαίδευσης (69 εικόνες). Τα δεδομένα αυτά προσομοιάστηκαν με εικόνας 4x4 MSFA που είναι η δομή των πολυφασματικών εικόνων όπως λαμβάνονται από τις XIMEA's xiSpec and IMEC's "snapshot mosaic" κάμερες (βλ. Σχήμα 2.9).



Σχήμα 2.9: Multispectral Filter Array (MSFA) 4x4, όπως προκύπτει από την IMEC κάμερα. Πηγή: [1].

Σχετικά με την συνάρτηση κόστους που κατασκευάστηκε είναι χρήσιμο να αναφερθούν τα εξής. Με απώτερο σκοπό την ελαχιστοποίηση των σφαλμάτων κατά την ανακατασκευή του σήματος στα δίκτυα DPG-Net και DDM-Net χρησιμοποιείται μία

συνάρτηση κόστους που συνδυάζει το μέσο τετραγωνικό σφάλμα (MSE) και το wavelet loss. Επίσης, προκειμένου να διατηρηθεί μία συνέπεια μεταξύ της «αραιής μωσαϊκής» εικόνας I_s^k και της πραγματικής πολυφασματικής \hat{I}^k προστίθεται επιπλέον ο όρος του mosaic loss.

Για την αξιολόγηση της προτεινόμενης ροής εργασιών χρησιμοποιήθηκαν ως μετρικές ποσοτικές αξιολόγησης οι PSNR, SSIM & SAM. Το προτεινόμενο δίκτυο είναι αυτό που κατέγραψε τις καλύτερες μετρικές συνολικά συγκρινόμενο τόσο με παραδοσιακές μεθόδους (πχ. Bilinear, IID, PPID) όσο και μεθόδους βασισμένες στη βαθιά μάθηση (πχ. ResNet-3D, DGRI & MCAN) και παρουσίασε τα λιγότερα τεχνουργήματα/στρεβλώσεις. Ακόμα, η υπεροχή του DPD-Net αποδίδεται αφενός στο γεγονός ότι συνολικά το δίκτυο και οι συναρτήσεις κόστους κατασκευάστηκαν λαμβάνοντας υπόψη τη δομή των MSFA και της διαδικασίας demosaicking που θα ακολουθήσει και αφετέρου στις υψηλής ποιότητας PPIs που μπορούν να υπολογιστούν πολύ καλά από το DPG-Net. Επιπρόσθετα, η τεχνική «channel adaptivity» που επιτρέπει ένα μόλις δίκτυο να μπορεί να εκπαιδευτεί και να εφαρμοστεί για όλα τα κανάλια καθώς και η συνάρτηση mosaic loss βελτιώνουν σημαντικά την ποιότητα ανακατασκευής της εικόνας (demosaicking quality). Κλείνοντας, η προτεινόμενη αρχιτεκτονική DPD-Net υπερτερεί έναντι υπαρχόντων μεθόδων multispectral demosaicking σύμφωνα με ποσοτικές μετρικές αλλά και με σύγκριση των εικόνων ενώ επιπλέον αποδεικνύεται συμβατή τόσο σε simulated MSFA εικόνες όσο και σε πραγματικές εικόνες που λήφθηκαν από MSFA κάμερα.

2.2.2. Deep Image Prior

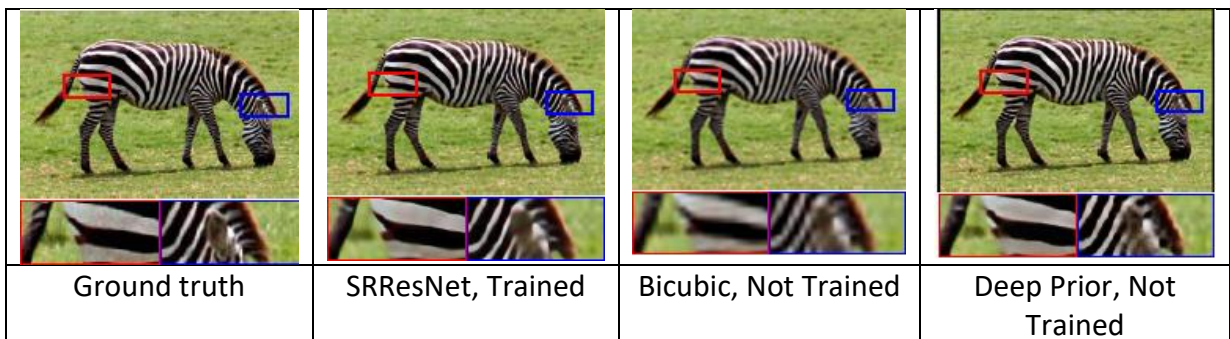
Σε αυτή την ερευνητική εργασία του Dmitry Ulyanov αναπτύσσεται ένα μη επιβλεπόμενο μοντέλο που χρησιμοποιεί αρχιτεκτονική generator δικτύου, «εκπαιδεύεται» ανά εικόνα και παράγει το λεγόμενο «image prior». Εδώ κρίνεται σκόπιμο να αναλυθεί τι είναι το prior -μιας και πρόκειται για κύρια έννοια πάνω στην οποία αναπτύσσεται η παρούσα εργασία-. Πρόκειται, λοιπόν, για τις εγγενείς-προγενέστερες γνώσεις ή υποθέσεις σχετικά με τη δομή και τις ιδιότητες των φυσικών εικόνων. Εδώ τίθεται η ιδέα ότι οι εικόνες του πραγματικού κόσμου μοιράζονται ορισμένα χαρακτηριστικά όπως είναι η ομαλότητα, η συνέχεια και η αραιότητα. Τα priors χρησιμοποιούνται σε διαδικασίες ανακατασκευής, ενίσχυσης εικόνων και λειτουργούν σαν περιορισμοί στους αλγόριθμους επεξεργασίας εικόνων για να παράξουν πιο ρεαλιστικά και οπτικά ελκυστικά αποτελέσματα, ιδίως σε περιπτώσεις αποθορυβοποίησης (denoising), inpainting κ.α..

Γενικά, τα βαθιά συνελκτικά δίκτυα χρησιμοποιούνται ευρέως για εργασίες παραγωγής και ανακατασκευής εικόνων και η μεγάλη τους επιτυχία αποδίδεται στην ικανότητα τους να μαθαίνουν ρεαλιστικά image priors από ένα μεγάλο δείγμα εικόνων. Σε αυτή την μελέτη προβάλλεται ότι η δομή ενός generator network είναι επαρκής για να ανιχνεύσει σημαντικό μέρος της πληροφορίας και των χαρακτηριστικών μίας εικόνας χωρίς να έχει προηγηθεί διαδικασία εκμάθησης. Συγκεκριμένα, αναπτύσσεται ένα νευρωνικό generator δίκτυο, στο οποίο εισάγεται κάθε φορά μία υποβαθμισμένη εικόνα, με τυχαία αρχικοποιημένα βάρη που λειτουργούν σαν παράμετροι της ανακατασκευασμένης εικόνας και βελτιστοποιούνται προκειμένου να προσεγγίσουν τη βέλτιστη restored image. Γίνεται επομένως σαφές, ότι πρόκειται για μία μονοεικονική διαδικασία ανακατασκευής όπου η μόνη πληροφορία που χρειάζεται για να επιλυθεί το εκάστοτε πρόβλημα περιέχεται στην μοναδική υποβαθμισμένη εικόνα και τη δομή του δικτύου που ορίζεται από τον χρήστη. Εδώ δίνεται έμφαση ότι καμία πτυχή του δικτύου δεν εκπαιδεύεται από τα δεδομένα αλλά η μόνη προγενέστερη γνώση πηγάζει από τη δομή του δικτύου -που επιβάλλει ένα ισχυρό prior- που είναι εξίσου σημαντική με τα δεδομένα εκπαίδευσης. Η καινοτομία της παρούσας εργασίας έγκειται ακριβώς σε αυτόν τον λόγο, ότι δηλαδή ερευνά το prior που προέρχεται από ένα deep convolutional generative network ανεξάρτητο από την εκπαίδευση των παραμέτρων του δικτύου από τις εικόνες.

Με άλλα λόγια το δίκτυο αυτό μπορεί να χρησιμοποιηθεί ως handcrafted prior με εξαιρετικά αποτελέσματα σε τυπικά inverse problems, όπως denoising, inpainting, super-resolution, συγκρινόμενα με αποτελέσματα επιβλεπόμενων εναλλακτικών μεθόδων, που κατασκευάζουν καλά image priors. Χαρακτηριστικό παράδειγμα αποτελεί αυτό του Σχήματος 2.10 που ακολουθεί όπου αντιμετωπίζεται το πρόβλημα του super resolution. Στη συγκεκριμένη περίπτωση χρησιμοποιείται ένα τυχαία αρχικοποιημένο ConvNet για να αυξηθούν οι χωρικές διαστάσεις της εικόνας με τυχαιοδειγματοληπτικούς τρόπους (upsample) χρησιμοποιώντας τη δομή του σαν

image prior. Αυτό λειτουργεί παρόμοια με το bicubic upsample -μιας και οι δύο αρχιτεκτονικές δεν προαπαιτούν εκπαίδευση- αλλά το Deep Image prior (DIP) παράγει πολύ πιο καθαρά και με «κοφτερές» ακμές αποτελέσματα. Μάλιστα, παρατηρείται να συγκρίνεται με σύγχρονες μεθόδους super resolution που χρησιμοποιούν ConvNets που εκπαιδεύονται σε μεγάλο όγκο δεδομένων έχοντας ανταγωνιστικά αποτελέσματα.

Ακόμα, μπορεί να χρησιμοποιηθεί για την αντιστροφή αναπαραστάσεων βαθιών νευρωνικών (invert deep neural representations) καθώς και για την αποκατάσταση ζευγαριών εικόνων τύπου flash-no flash. Πέρα των παραπάνω εφαρμογών, η παρούσα εργασία εστιάζει στο inductive bias που παρουσιάζεται στις τυπικές αρχιτεκτονικές ενός generator δικτύου. Με άλλα λόγια, αναφέρεται στη προγενέστερη γνώση και τις υποθέσεις που κάνει ο αλγόριθμος όταν πρόκειται να επιλέξει και να κάνει προβλέψεις ανάμεσα σε δύο υποθέσεις-μοντέλα, όταν δηλαδή υπάρχει ασάφεια στα δεδομένα κάτι που είναι εξαιρετικά χρήσιμο σε περιπτώσεις όπου τα δεδομένα περιέχουν θόρυβο ή είναι ατελή. Επιπλέον, μία αξιοσημείωτη προσφορά της συγκεκριμένης εργασίας είναι ότι γεφυρώνει το κενό μεταξύ δύο μεθόδων ανακατασκευής εικόνων: των εκπαιδευμένων μεθόδων που χρησιμοποιούν βαθιά συνελκτικά δίκτυα και αυτών που δεν έχουν εκπαιδευτεί και βασίζονται σε handcrafted image priors, όπως η ομοιότητα με τον εαυτό (self-similarity).



Σχήμα 2.10: Super-resolution χρησιμοποιώντας deep image prior και άλλες μεθόδους που απαιτούν (SRResNet,) ή όχι εκπαίδευση (Bicubic). Πηγή: [2].

Όσον αφορά τη μεθοδολογία που αναπτύχθηκε το deep generator network f μπορεί να θεωρηθεί ως μία παραμετρική συνάρτηση $\mathbf{x} = \mathbf{f}_\theta(\mathbf{z})$ που δέχεται σαν δεδομένο εισόδου ένα κρυμμένο διάνυσμα \mathbf{z} (ίδιων διαστάσεων με την corrupted εικόνα x_0) και το αντιστοιχίζει σε μία ανακατασκευασμένη εικόνα \mathbf{x} . Η αντιστοίχιση αυτή αποτελείται από μετασχηματισμούς που προκύπτουν από την αρχιτεκτονική του δικτύου και τα διαρκώς εκπαιδευόμενα βάρη, επιτρέποντας με αυτόν τον τρόπο τη μοντελοποίηση σύνθετων κατανομών εικόνων. Εδώ είναι σημαντικό να σημειωθεί ότι η βελτιστοποίηση του δικτύου αντί να γίνει στον χώρο της εικόνας γίνεται στο χώρο της παραμέτρου θ , που περιλαμβάνει τα βάρη και το bias των φίλτρων, μέσω της mapping συνάρτησης g . Το δίκτυο έχει μία συγκεκριμένη δομή αλλά τα φίλτρα (linear convolution, upsampling and non-linear activation functions) εναλλάσσονται ανάλογα το reconstruction task. Ακόμα, το δίκτυο γνωρίζει για τη low-level δομή των φυσικών

εικόνων, γεγονός που το καθιστά ικανό να αποκαταστήσει μία κατεστραμμένη εικόνα, συνεπώς μπορεί να αξιοποιηθεί στην επίλυση inverse problems.

Παράλληλα, το δίκτυο αντί να διαχειρίζεται κατανομές αναπτύσσει εφαρμογές ελαχιστοποίησης ενέργειας που δίνονται από τον ακόλουθο τύπο:

$$\mathbf{x}^* = \min_{\mathbf{x}} E(\mathbf{x}; \mathbf{x}_0) + R(\mathbf{x}),$$

- όπου $E(\mathbf{x}; \mathbf{x}_0)$ είναι όρος που εξαρτάται από την εφαρμογή και συχνά υπαγορεύεται άμεσα. Αναπαριστά την πιστότητα των δεδομένων, δηλαδή εκφράζει πόσο καλά η παράγωγη λύση (ανακατασκευασμένη εικόνα \mathbf{x}) προσεγγίζει τα παρατηρούμενα δεδομένα (κατεστραμμένη εικόνα \mathbf{x}_0). Για παράδειγμα στη περίπτωση μίας εικόνας με θόρυβο (\mathbf{x}_0) ο όρος $E(\mathbf{x}; \mathbf{x}_0)$ μετράει τη διαφορά με την αποθορυβοποιημένη εικόνα (\mathbf{x}). και
- $R(\mathbf{x})$ που είναι ο κανονικοποιητής (regularizer) που εισάγει τη γνώση του prior και τυχόν περιορισμούς ως προς τη λύση προκειμένου να ακολουθεί ορισμένες ιδιότητες των φυσικών εικόνων, όπως η ομαλότητα, η αραιότητα και άλλα δομικά χαρακτηριστικά. Ο regularizer είναι απαραίτητος προκειμένου να αποφευχθεί το overfitting και η λύση να είναι ουσιαστική. Εν αντιθέσει με τον όρο $E(\mathbf{x}; \mathbf{x}_0)$, ο $R(\mathbf{x})$ δεν συνδέεται με κάποια συγκεκριμένη εφαρμογή λόγω της κανονικότητας των εικόνων που συνεπάγεται.

Μέσω της ελαχιστοποίησης επιτυγχάνεται η εύρεση της βέλτιστης λύσης που θα ισορροπεί την προσαρμογή των δεδομένων και την τήρηση των προγενέστερων γνώσεων και περιορισμών. Τέλος, επειδή είναι πολύ δύσκολη η βελτιστοποίηση και των δύο όρων ο $R(\mathbf{x})$ δεν λαμβάνεται υπόψη και χρησιμοποιείται μόνο το prior που βρίσκεται στη παραμετροποίηση του δικτύου. Λαμβάνοντας υπόψη τα παραπάνω, γίνεται σαφές ότι το δίκτυο που συνιστάται στην παρούσα εργασία εν ονόματι «Deep Image Prior» αντιστέκεται να εξάγει «άσχημα» αποτελέσματα και μάλιστα τείνει σε εικόνες με πιο φυσικά χαρακτηριστικά.

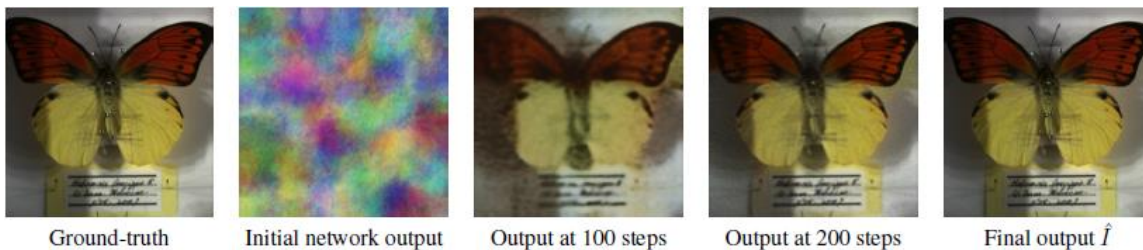
Επιπρόσθετα, βάση ποικίλων πειραμάτων που υλοποιήθηκαν αποδείχθηκε ότι η επιλογή της κατάλληλης αρχιτεκτονικής δικτύου επηρεάζει την διαδικασία βελτιστοποίησης σε εργασίες ανακατασκευής εικόνων. Μάλιστα τα εν λόγω πειράματα έτρεξαν χρησιμοποιώντας encoder-decoder / U-Net-like «hourglass» αρχιτεκτονικές με skip connections, όπου τα z και x έχουν τις ίδιες χωρικές διαστάσεις -με το z να είναι ένας σταθερός τυχαία αρχικοποιημένος 3D τανυστής.- και το δίκτυο διαθέτει αρκετές εκατομμύρια παραμέτρους.

2.2.3. MD²IP: TRAINING-FREE MULTISPECTRAL DEMOSAICING WITH DEEP

IMAGE PRIORS

Η συγκεκριμένη ερευνητική δημοσίευση επικαλείται τη μέθοδο του Deep Image Prior (που αναλύθηκε στην ενότητα 2.2.2. της παρούσας διπλωματικής) για να προτείνει μία μέθοδο demosaicing -που δεν εκπαιδεύεται- και μπορεί να αξιοποιηθεί στη περίπτωση των MSFA εικόνων, που αποτελούν παράγωγα των snapshot multispectral (MS) καμερών. Στην παρούσα δημοσίευση η διαδικασία του MSFA demosaicing αποσκοπεί στην αποκατάσταση πλήρους ανάλυσης (Full resolution) πολυφασματικών εικόνων που ακολουθούν τη δομή MSFA. Όσον αφορά το Deep Image Prior προέρχεται από τα ευρέως γνωστά δίκτυα παραγωγής εικόνων (image generation networks) και χρησιμοποιείται για την ανακατασκευή της MSFA εικόνας ενώ εστιάζει στην κατάλληλη κατασκευή της δομής ενός δικτύου για να λειτουργήσει σαν image prior.

Συγκεκριμένα, προτείνεται μία training free μεθοδολογία που εκμεταλλεύεται την ικανότητα ενός τυχαία αρχικοποιημένου βαθύ νευρωνικού δικτύου να λειτουργεί σαν κατάλληλο image prior. Εδώ να σημειωθεί ότι πρόκειται για μία μεθοδολογία που βασίζεται σε ένα μόνο δείγμα (instance-based). Το πλεονέκτημα της παρούσας μεθόδου είναι ότι δεν προαπαιτεί την δύσκολη απόκτηση δεδομένων υψηλής ποιότητας, όπου τα κανάλια θα πρέπει να αντιστοιχίζονται με αυτά του εκάστοτε filter array που θα επιλέγεται αλλά αντ' αυτού μπορεί να χρησιμοποιηθεί σε αυθαίρετες διαμορφώσεις πολυφασματικών filter arrays. Αναλυτικότερα, αναπτύσσεται μία αρχιτεκτονική από Fully Convolutional Network που ενσωματώνει προγενέστερη γνώση σχετικά με την εικόνα και έχει σαν στόχο να βελτιστοποιήσει τις παραμέτρους του δικτύου -χρησιμοποιώντας σαν είσοδο έναν ομοιόμορφα κατανεμημένο τανυστή z (που λειτουργεί σαν αρχική είσοδο του δικτύου)- μέχρι να προσεγγιστεί ικανοποιητικά το image prior. Η βελτιστοποίηση και συνεπώς η ανανέωση των παραμέτρων γίνεται για προκαθορισμένο αριθμό επαναλήψεων και έως ότου να ελαχιστοποιηθεί το mean squared error loss. Το βελτιστοποιημένο παράγωγο του δικτύου είναι η demosaicked εικόνα. Πιο λεπτομερώς χρησιμοποιήθηκε ένα δίκτυο της μορφής encoder-decoder (Enc-Dec) με skip connections. Μία ενδεικτική υλοποίηση της παραπάνω μεθοδολογίας που ακολουθήθηκε παρατηρείται στο Σχήμα 2.11.



Σχήμα 2.11: Μετάβαση από την μη αρχικοποιημένη έξοδο του δικτύου στην τελική εικόνα εξόδου \hat{I} μετά το πέρας των επαναλήψεων, για την εικόνα Butterfly (εμφάνιση σε RGB). Πηγή [3].

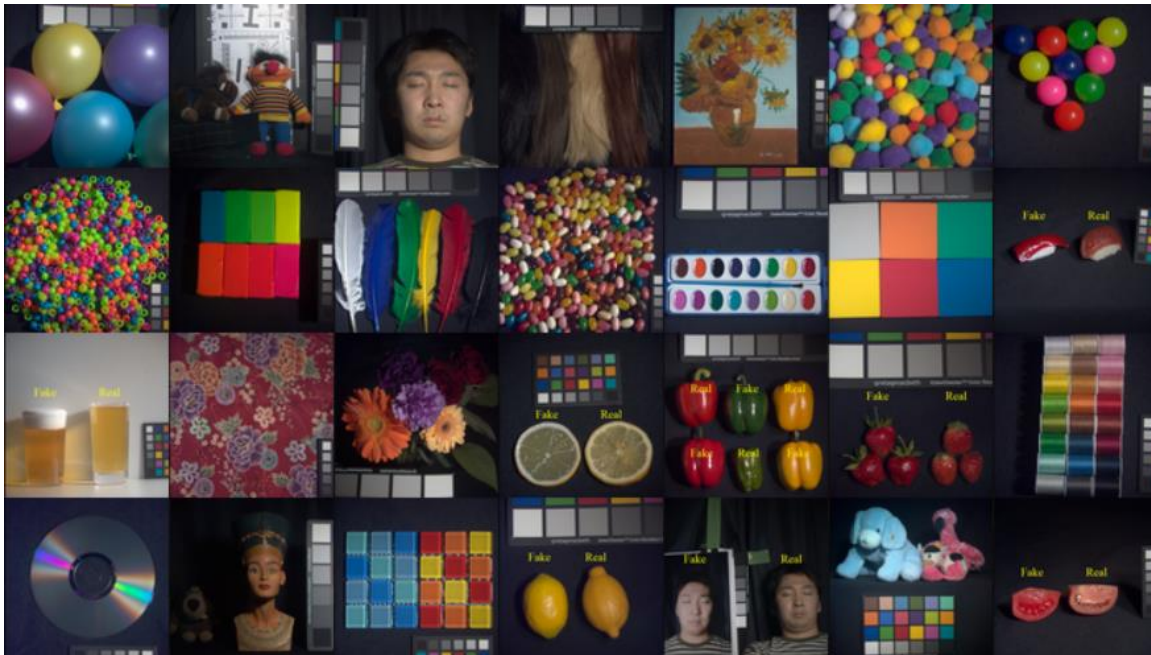
Παράλληλα, πολυάριθμες μεθοδολογίες έχουν αναπτυχθεί για την επίλυση του MSFA demosaicing από απλές μεθόδους παρεμβολής (πχ. Bilinear) μέχρι μεθόδους βαθιάς μάθησης βασισμένες στο super resolution. Αυτές που αντιπαραβάλλονται στην προκειμένη περίπτωση με την προτεινόμενη μεθοδολογία MD²IP είναι μέθοδοι MSFA demosaicing που δεν εκπαιδεύονται: bilinear interpolation, Iterative Intensity Difference (IID) και Pseudo-Panchromatic Image Difference (PPID) και αυτές που εκπαιδεύονται σε demosaicing μοντέλα για συγκεκριμένες διαμορφώσεις MSFA arrays όπως οι DGRI, MCAN και DPD. Πραγματοποιήθηκε, λοιπόν, ποιοτική και ποσοτική αξιολόγηση με τις παραπάνω μεθόδους στο σύνολο δεδομένων «TokyoTech31» που αποτελείται από υπερφασματικές εικόνες που καταγράφουν φασματική πληροφορία σε 31 κανάλια. Μάλιστα το συγκεκριμένο dataset είναι ένα από αυτά που χρησιμοποιήθηκαν και στην ερευνητική εργασία που αναλύθηκε στο υπεδάφιο 2.2.1. Σύμφωνα με τις αξιολογήσεις φαίνεται η MD²IP να παρουσιάζει καλύτερες επιδόσεις σε σχέση ακόμα και με σύγχρονες μεθόδους που έχουν ακολουθήσει εκπαίδευση με την επιπλέον ευελιξία να μην εξαρτάται από ένα συγκεκριμένο MSFA ground truth data pattern και μοναδικό μειονέκτημα τον εκτεταμένο χρόνο ανακατασκευής της demosaicked εικόνας, γεγονός σύνηθες για τις μεθόδους που στηρίζονται στο deep image prior.

ΚΕΦΑΛΑΙΟ 3^ο – 1^η μεθοδολογία (MD²IP)

3.1. Σύνολο Δεδομένων

Στην παρούσα διπλωματική εργασία σαν πειραματικά δεδομένα αντλήθηκαν όμοια με αυτά που επικαλούνται στην ερευνητική εργασία που αναλύθηκε στο υπεδάφιο 2.2.1. του 2^{ου} Κεφαλαίου με όνομα «A Deep Joint Network for Multispectral Demosaicking Based on Pseudo-Panchromatic Images», δηλαδή επιδιώχθηκαν σύνολα δεδομένων που θα καταγράφουν πληροφορία στο φάσμα 460-640nm, που είναι αυτό που ακολουθεί η ΧΙΜΕΑ's xiSpec κάμερα. Πιο αναλυτικά χρησιμοποιήθηκαν 4 σετ πολυφασματικών δεδομένων που συγκεντρώθηκαν από προγενέστερες ερευνητικές δημοσιεύσεις [5], [6], [7], [8] και τα χαρακτηριστικά των οποίων θα αναλυθούν παρακάτω. Τα δεδομένα αυτά θα αποτελούν -έπειτα από κατάλληλη προεπεξεργασία που θα αναλυθεί σε επόμενη ενότητα- το σύνολο στο οποίο θα στηριχθεί η παρούσα ΔΕ για το στάδιο εκπαίδευσης και ελέγχου. Αναλυτικότερα, αυτά είναι τα εξής:

➤ CAVE



Σχήμα 3.1: Πολυφασματικές εικόνες που απαρτίζουν το σύνολο δεδομένων CAVE.
(Πηγή: <https://www.cs.columbia.edu/CAVE/databases/multispectral/>).

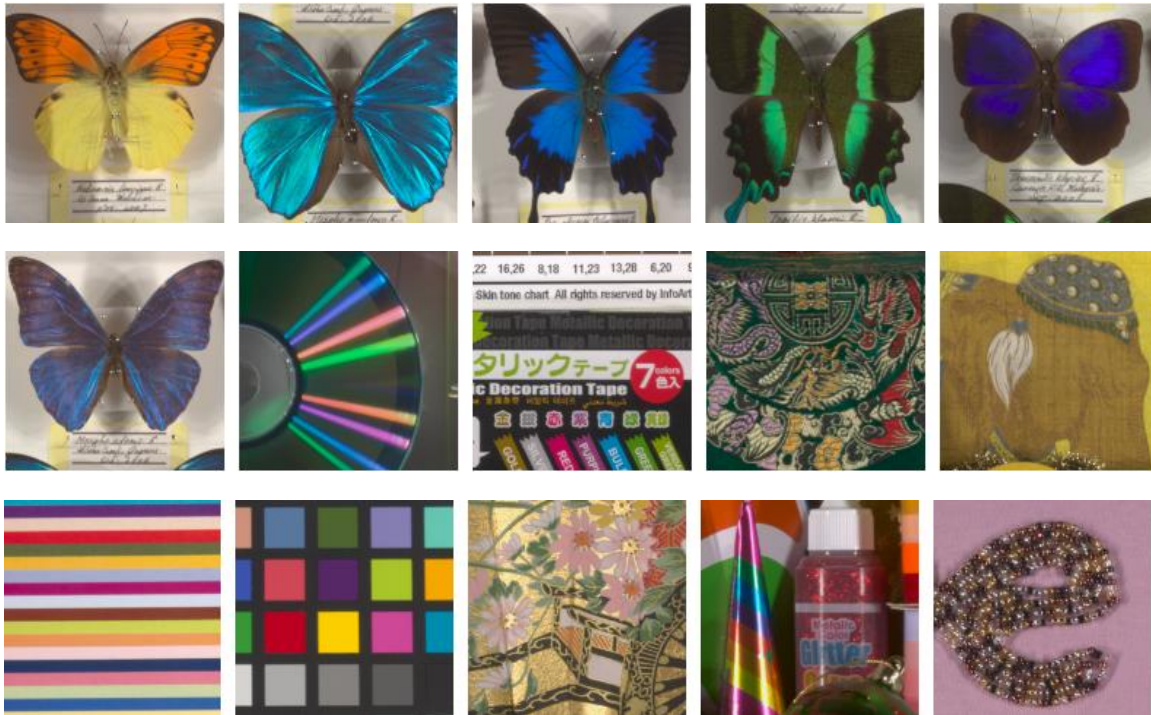
Το CAVE dataset αποτελείται από 32 πολυφασματικές εικόνες διάστασης 512x512 pixels, με 31 φασματικά κανάλια που καλύπτουν μήκη κύματος στο εύρος των 400-700nm, με βήμα 10nm. Οι εικόνες αυτές αναπαριστούν τις τιμές ανακλαστικότητας (reflectance values) των αντικειμένων που βρίσκονται σε κάθε σκηνή. Με άλλα λόγια η τιμή του εικονοστοιχείου κάθε εικόνας προκύπτει από «υπολογισμούς επί της ληφθείσας πολυφασματικής εικόνας με βαθμονομημένες εκτιμήσεις της ηλιακής ακτινοβολίας και της φασματικής απόκρισης της κάμερας». Συνεπώς, κάθε πολυφασματική εικόνα του συνόλου δεδομένων πρόκειται για μία κοντινή και όχι ακριβή εκτίμηση της πραγματικής ανακλαστικότητας της σκηνής.

Ακόμα, οι 32 πολυφασματικές εικόνες διακρίνονται σε 5 κατηγορίες (αντικείμενα, δέρμα και μαλλιά, ζωγραφιές, φαγητά και ποτά, πραγματικά και ψεύτικα αντικείμενα). Κάθε εικόνα αποτελείται από ένα συμπιεσμένο (zip) αρχείο με καθένα από αυτά να απαρτίζεται από 31 (.png) αρχεία όσα και τα κανάλια που καταγράφει που ακολουθείται με τον αντίστοιχο δείκτη στην ονοματοδοσία του αρχείου (πχ. object_ms_01.png, με το «01» να υποδεικνύει ότι πρόκειται για την πρώτη εικόνα που καταγράφηκε σε μήκος κύματος 400nm κ.ο.κ.). Τέλος το zip αρχείο περιλαμβάνει ένα έγχρωμο σύνθετο κατά το πρότυπο sRGB που έχει ληφθεί υπό συνθήκες που προσομοιάζουν φυσικό φωτισμό (D65). Τέλος, παρατίθεται ο Πίνακας 3.1 που περικλείει όλες τις πληροφορίες λήψης των εικόνων:

Camera	Cooled CCD camera (Apogee Alta U260)
Resolution	512 x 512 pixel
Filter	VariSpec liquid crystal tunable filter
Illuminant	CIE Standard Illuminant D65
Range of wevelength	400nm - 700nm
Steps	10nm
Number of bands	31 band
Focal length	f/1.4
Focus	Fixed (focused using 550nm image)
Image format	PNG (16bit)

Πίνακας 3.1: Πληροφορίες λήψης εικόνων για το σύνολο δεδομένων CAVE.
(Πηγή: <https://www.cs.columbia.edu/CAVE/databases/multispectral/>).

➤ TokyoTech-31band (TT31)



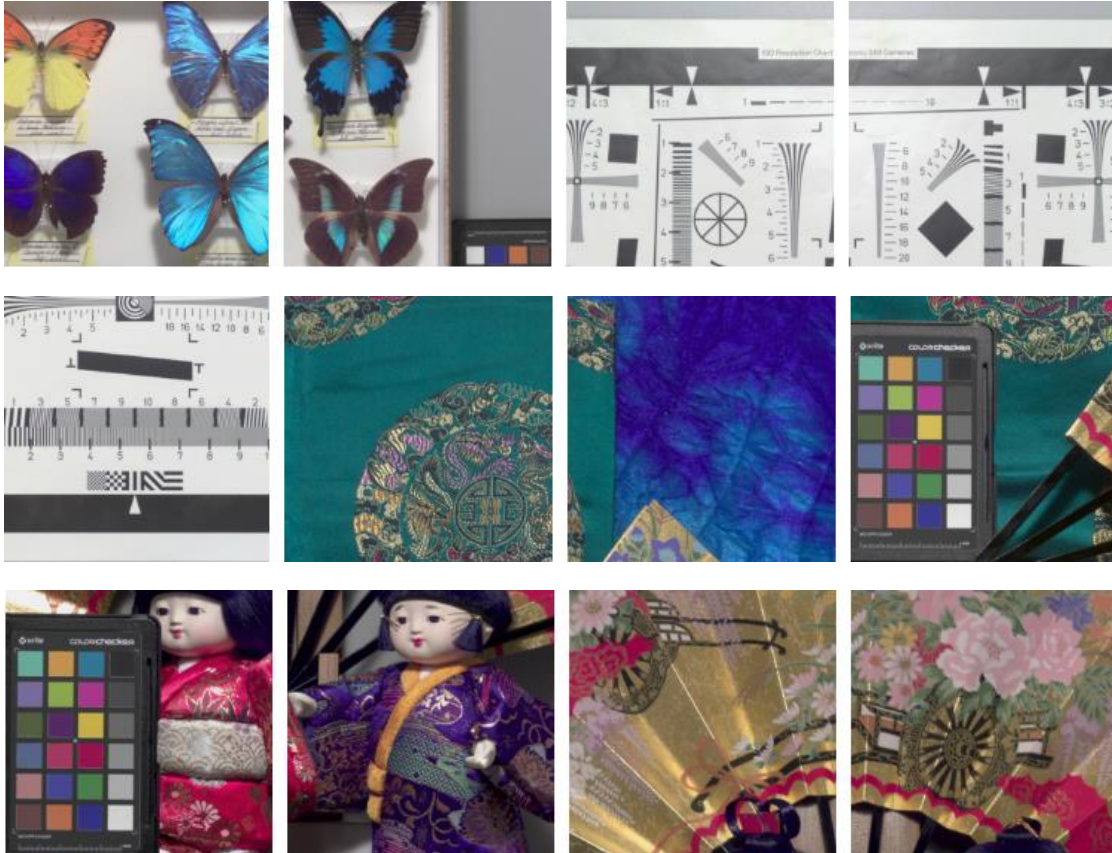
Σχήμα 3.2: Ενδεικτικά 15 σκηνές από το σύνολο δεδομένων TokyoTech-31band.
(Πηγή: <http://www.ok.sc.e.titech.ac.jp/res/MSI/MSIdata31.html>).

Το παρόν σύνολο δεδομένων συνιστάται από 30 υπερφασματικές εικόνες με εύρος κάλυψης στο ηλεκτρομαγνητικό φάσμα μεταξύ 420-720nm με διαστήματα των 10nm. Δηλαδή κάθε εικόνα αποτελείται από 31 φασματικά κανάλια, όπως προδίδει και το όνομα του, TokyoTech-31band και έχει διάσταση 500x500 pixels. Οι υπερφασματικές εικόνες λήφθηκαν χρησιμοποιώντας μία μονοχρωματική (monochrome) κάμερα με VariSpec liquid crystal tunable φίλτρο, που πρακτικά μεταφέρει και εμποδίζει ορισμένα μήκη κύματος. Πρόκειται για ένα σύνολο δεδομένων που έχει βαθμονομηθεί χρησιμοποιώντας colorchart και οι τιμές που έχουν αποθηκευτεί ανά εικονοστοιχείο αντιστοιχούν σε μορφή φασματικής ανακλαστικότητας.

Οι εικόνες αναπαριστούν πολύχρωμα αντικείμενα πλούσια σε υφή και εν προκειμένω χρήσιμα σε διαδικασίες αξιολόγησης υπερφασματικών εικόνων και γενικότερα σε εργασίες αποκατάστασης (restoration tasks). Εδώ να σημειωθεί ότι στη σελίδα όπου διατίθεται το TokyoTech-31band παρέχονται επιπλέον 5 εικόνες με διάφορα charts, που αντλήθηκαν και αξιοποιήθηκαν ομοίως με τα υπόλοιπα. Οι συγκεκριμένες εικόνες διατηρούν ίδιο αριθμό καναλιών ίσο με 31 αλλά οι διαστάσεις διαφοροποιούνται ακόμα και μεταξύ τους. Τέλος, για μελλοντικούς χρήστες που θέλουν

να προσομοιάσουν την διαδικασία λήψης υπερφασματικών εικόνων με αυθαίρετη ευαισθησία κάμερας και συνθήκες φωτισμού τους παρέχεται ένα υπόδειγμα κώδικα για να το κάνουν, κάτι που προσωπικά θεωρείται σημαντική προσφορά άξια αναφοράς.

➤ **TokyoTech-59band (TT59)**



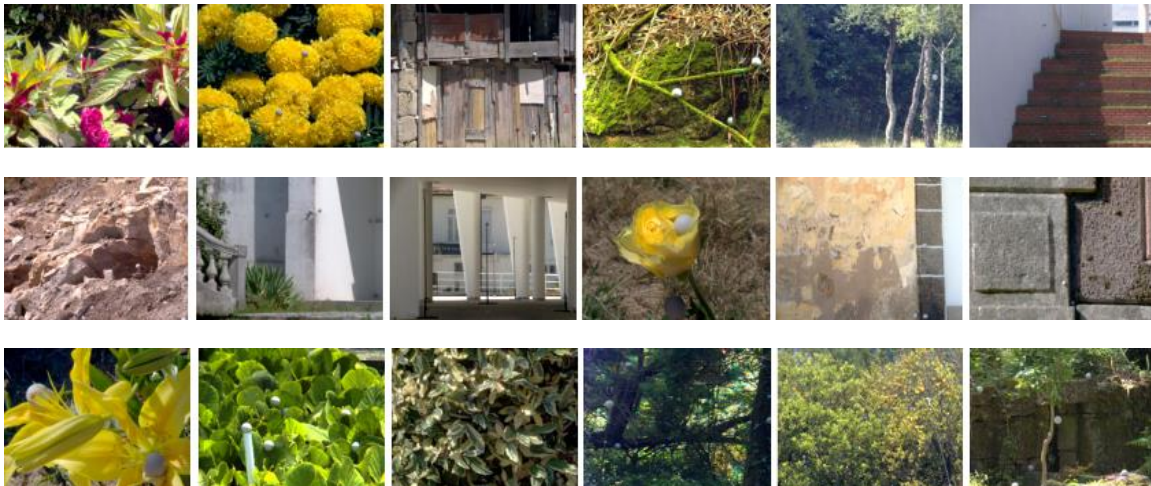
Σχήμα 3.3: Ενδεικτικές σκηνές από το σύνολο δεδομένων TokyoTech-59band.
(Πηγή: <http://www.ok.sc.e.titech.ac.jp/res/MSI/MSIdata59.html>).

Το TokyoTech-59band (TT59) αποτελείται από 40 υπερφασματικές εικόνες εκ των οποίων μόνο οι 16 είναι διαθέσιμες για ερευνητικούς σκοπούς λόγω περιορισμών αδειας. Καλύπτουν το ηλεκτρομαγνητικό φάσμα μεταξύ 420-1000nm με διαστήματα των 10nm και αποτελούνται από 59 κανάλια και έχουν διάσταση 512x512 εικονοστοιχεία. Οι εικόνες λήφθηκαν χρησιμοποιώντας τη μονοχρωματική κάμερα και δύο VariSpec tunable φίλτρα και πιο συγκεκριμένα το VIS για 420-650nm και το SNIR για 650-1000nm. Ομοίως με το TokyoTech-31band οι υπερφασματικές εικόνες

καταγράφουν τιμές ανακλαστικότητας χάρη στη διαδικασία βαθμονόμησης που στηρίχτηκε σε ένα colorchart όπου οι τιμές ανακλαστικότητας είναι γνωστές.

Οι υπερφασματικές εικόνες καλύπτουν το ορατό και το εγγύς υπέρυθρο εύρος του φάσματος και απεικονίζουν ποικίλα πολύχρωμα αντικείμενα με έντονη υφή και επομένως μπορούν να φανούν χρήσιμες για την αξιολόγηση RGB/NIR ή και υπερφασματικών απεικονίσεων και ευρύτερα σε διαδικασίες αποκατάστασης. Τέλος μιας και οι εικόνες όπως προαναφέρθηκε αντιστοιχούν σε τιμές ανακλαστικότητας σε περίπτωση που κάποιος μελλοντικός χρήστης θελήσει να προσομοιάσει την διαδικασία λήψης RGB/NIR/υπερφασματικών εικόνων με αυθαίρετη ευαισθησία κάμερας και συνθήκες φωτισμού του παρέχεται ένα υπόδειγμα κώδικα.

➤ Manchester 2015 (M15)



Σχήμα 3.4: Ενδεικτικές σκηνές από το σύνολο δεδομένων Manchester 2015.
(Πηγή:https://personalpages.manchester.ac.uk/staff/d.h.foster/Local_Illumination_HSIs/Local_Illumination_HSIs_2015.html).

Το τέταρτο και τελευταίο σετ δεδομένων είναι το Manchester 2015. Αυτό αποτελούνταν από 30 εικόνες υπερφασματικής ακτινοβολίας που καλύπτουν το φάσμα 400-720nm με διαστήματα των 10nm, με καθεμιά από αυτές να συνιστάται από 33 κανάλια διαστάσεων (περίπου) 1024x1344 εικονοστοιχεία. Είναι σημαντικό να υπογραμμιστεί ότι εδώ οι τιμές ανά pixel αναπαριστούσαν φασματική ακτινοβολία και όχι ανακλαστικότητα όπως όλα τα προηγούμενα, δηλαδή οι εικόνες που δίνονται πρόκειται για τις αρχικές λήψεις (raw grayscale images) και άνευ προεπεξεργασίας. Τα

δεδομένα παρέχονται σε μορφή (.mat) αρχείων και επιπλέον κάθε εικόνα συνοδεύεται από ένα έγχρωμο (RGB) BMP αρχείο εικόνας και ένα αρχείο (.pdf) που αναφέρει εκτενώς πληροφορίες που αφορούν κάθε λήψη (όνομα αρχείου και μορφή αναπαράστασης σκηνής, γεωγραφική αναφορά, τοποθεσία, ώρα & γεωμετρία λήψης, μεταεπεξεργασία κ.α.).

Στις εικόνες αυτές μελετώνται φυσικές σκηνές και η λήψη τους πραγματοποιήθηκε στη περιοχή Minho της Πορτογαλίας κατά την ανοιξιάτικη και καλοκαιρινή περίοδο στο διάστημα 2002-2003. Ενδιαφέρον παρουσιάζει το χρονικό διάστημα καθώς και η λεπτομέρεια καταγραφής των σκηνών καθώς φρόντισαν να γίνουν υπό συνθήκες καλού φωτισμού (περίοδο που υπάρχει καθαρός ουρανός, ιδανικά χωρίς νεφώσεις), με εξαίρεση πέντε (5) εξ αυτών, μιας και η εργασία για την οποία θα αξιοποιούνταν αρχικά εστίαζε στην καταγραφή τοπικών διαφοροποιήσεων του φωτισμού στο φάσμα. Γενικά για όλες τις σκηνές παρατηρήθηκε χαμηλή αναλογία σήματος-θορύβου (signal-to-noise ratio) στα άκρα του φάσματος 400-720nm όπου μάλιστα η ανθρώπινη όραση είναι λιγότερο ευαίσθητη σε σχέση με τα ενδιάμεσα μήκη κύματος. Συνεπώς αυτά τα μήκη κύματος έχουν μικρότερη συνεισφορά ως προς το χρώμα και την τελική εικόνα που γίνεται αντιληπτή από τον άνθρωπο σε φυσικές σκηνές. Αυτός ο δείκτης συγκρίνει πόση χρήσιμη πληροφορία -κοινώς σήμα- καταγράφεται σε σχέση με ανεπιθύμητη που είναι ο θόρυβος, παρόλα αυτά δεν απορρίφθηκαν δεδομένα μήκους κύματος.

3.2. Προεπεξεργασία δεδομένων

Αφού έγινε η λήψη των τεσσάρων σετ δεδομένων που αναπτύχθηκαν παραπάνω στην Ενότητα 3.1. κρίθηκε σκόπιμη η ενιαία διαχείριση και μετατροπή τους σε έναν κοινό τύπο αρχείου, το (.tif). Συνεπώς οι σκηνές του CAVE dataset από συμπιεσμένη μορφή (.png) μετατράπηκαν σε (.tif) και με παρόμοιο τρόπο τα (.mat) αρχεία των υπολοίπων συνόλων δεδομένων. Επίσης, για λόγους ομοιομορφίας επικράτησε για κάθε εικόνα η ονοματοδοσία "band_1, band_2" κ.ο.κ..

Οι snapshot multispectral (MS) κάμερες έχουν προσελκύσει το ενδιαφέρον λόγω της ευελιξίας και της ευκολίας στη χρήση τους σε διάφορες εφαρμογές σχετιζόμενες με πολυφασματικές απεικονίσεις. Αυτές όπως έχει προαναφερθεί προκύπτουν εφαρμόζοντας πολυφασματικά φίλτρα σε αισθητήρες απεικόνισης, με παρόμοιο τρόπο όπως γίνεται με τα Bayes color filter arrays σε RGB κάμερες. Στην προκειμένη

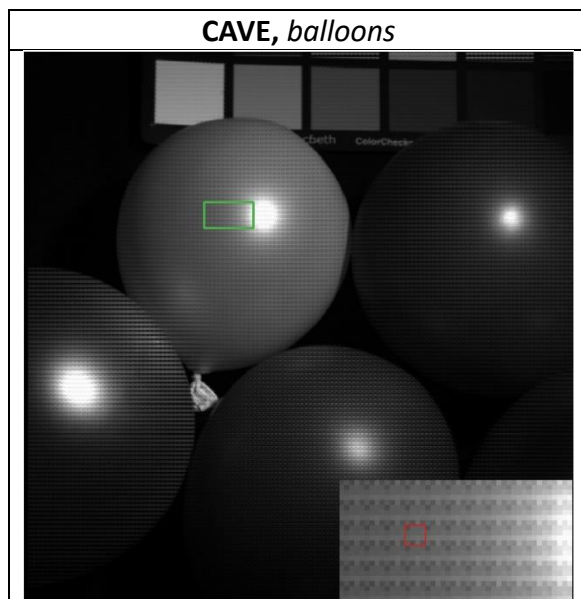
περίπτωση τα δεδομένα μας θεωρείται ότι προσομοιάζουν πολυφασματικές εικόνες όπως προκύπτουν από πολυφασματικούς αισθητήρες επίπεδης διάταξης φίλτρων της μορφής 4x4 MSFA, δηλαδή ακολουθούν το μοτίβο εκείνο που στηρίζεται η XIMEA xiSpec “snapshot mosaic” υπερφασματική κάμερα με αισθητήρα απεικόνισης της IMEC.

Πιο αναλυτικά, τα φασματικά κανάλια της IMEC πάνω στα οποία αναπτύχθηκε η MSFA κεντράρονται στα μήκη κύματος που προβάλλονται στον Πίνακα 3.2. Τα σύνολο δεδομένων που έχουν συλλεχθεί όμως καταγράφουν φασματική πληροφορία σε μήκη κύματος πολλαπλάσια του x10 στο εύρος 400-1000nm, όπως έχει προαναφερθεί. Επομένως, για να γίνει μία τέτοια προσέγγιση και να αναχθούν τα πολυφασματικά δεδομένα στο μοτίβο MSFA της IMEC κάμερας χρειάζεται να γίνει αναδειγματοληψία (resampling) σε 16 κανάλια. Για να γίνει αυτό -για κάθε dataset- αρχικά απομονώθηκαν 19 κανάλια στο ηλεκτρομαγνητικό φάσμα 460, 470, 480...640nm που ήταν απαραίτητα για να υπολογιστεί η γραμμική παρεμβολή και συνεπώς να εκτιμηθούν τα τελικά 16 κανάλια, που αποτελούν τα δεδομένα αληθείας μας (groundtruth data), που θα συμμετέχουν στην κατασκευή του MSFA.

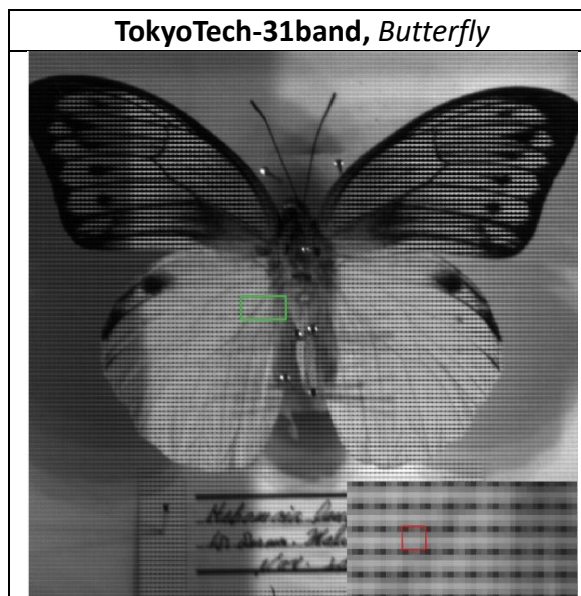
λ7=537	λ8=551	λ6=524	λ5=513
λ15=621	λ16=633	λ14=613	λ13=602
λ11=580	λ12=590	λ10=566	λ9=552
λ3=489	λ4=499	λ2=480	λ1=469

Πίνακας 3.2: Ο πίνακας πολυφασματικού φίλτρου διάταξης 4x4 (MSFA) της IMEC.

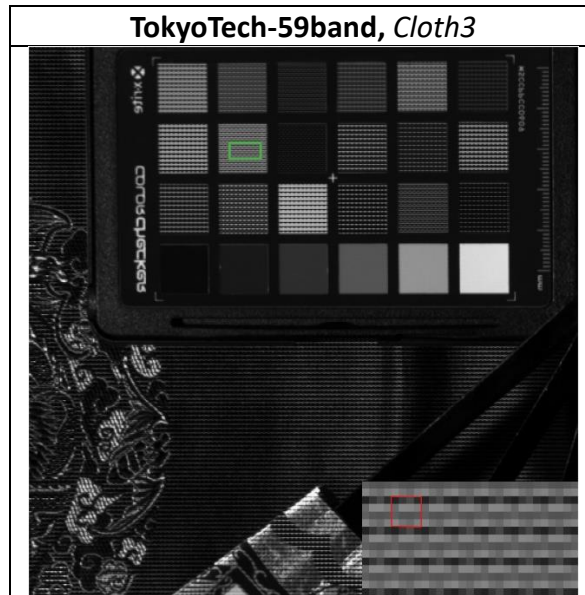
Αφού υπολογίστηκαν οι τιμές ανακλαστικότητας για καθένα από τα 16 πολυφασματικά κανάλια ακολούθησε η κατασκευή της MSFA. Για τη διαδικασία αυτή υλοποιήθηκε πρακτικά υποδειγματοληψία των πολυφασματικών καναλιών πλήρους ανάλυσης καθώς σε μία εικόνα τύπου MSFA κάθε εικονοστοιχείο καταγράφει φασματική πληροφορία από ένα κανάλι. Αυτό το στάδιο οπτικοποιείται στα Σχήματα 3.5, 3.6, 3.7 και 3.8 που ακολουθούν. Συγκεκριμένα, σε κάθε Σχήμα αναπαρίσταται μία εικόνα ανά σετ δεδομένων που θα αντιπροσωπεύει στο εξής την πραγματική εικόνα MSFĀ, σύμφωνα με την οποία θα γίνει η βελτιστοποίηση του δικτύου και θα υπολογίζεται το loss function. Σε κάθε Σχήμα επιλέγεται μία σχετικά ομοιογενής περιοχή που επισημαίνεται με πράσινο πλαίσιο στην οποία γίνεται εκτενέστερη εστίαση για να παρατηρηθεί το 4x4 πρότυπο που ακολουθεί η MSFĀ, που περιγράφεται από το κόκκινο πλαίσιο.



Σχήμα 3.5: Η πραγματική εικόνα MSF \hat{A} που κατασκευάστηκε βάση της πολυφασματικής λήψης «balloons», του dataset CAVE.



Σχήμα 3.6: Η πραγματική εικόνα MSF \hat{A} που κατασκευάστηκε βάση της πολυφασματικής λήψης «Butterfly», του dataset TokyoTech-31band.



Σχήμα 3.7: Η πραγματική εικόνα MSFĀ που κατασκευάστηκε βάση της πολυφασματικής λήψης «cloth3», του dataset TokyoTech-59band.



Σχήμα 3.8: Η πραγματική εικόνα MSFĀ που κατασκευάστηκε βάση της πολυφασματικής λήψης «Gualtar Columns», του dataset Manchester 2015.

Ακολουθως, πριν διατρεχθούν τα δεδομένα από το MD²IP δίκτυο κρίθηκε σκόπιμο να γίνει πρώτα κανονικοποίηση των τιμών στο διάστημα 0-1 προκειμένου να είναι στο ίδιο εύρος μιας και οι αρχικές τιμές για κάθε σετ δεδομένων είναι διαφορετικές. Για την ακρίβεια, αυτό επιτεύχθηκε διαιρώντας σύμφωνα με τη ραδιομετρική ανάλυση (bit) του εκάστοτε σετ δεδομένων πχ. στο «Cave» dataset που καταγράφει την πληροφορία σε 16bit οι τιμές διαιρέθηκαν με το 2¹⁶ κ.ο.κ.. Αυτή η προεργασία θα φανεί χρήσιμη σε μετέπειτα διαδικασία όταν θα αποτελέσουν δεδομένα εισόδου στο δίκτυο τύπου U-Net που θα εκπαιδευτεί στη συνέχεια.

3.3. Multispectral Demosaicing Deep Image Prior (MD²IP)

3.3.1. Ανάλυση μεθοδολογίας MD²IP

Στην παρούσα ενότητα θα αναλυθεί η μεθοδολογία που ακολουθήθηκε για την ανακατασκευή του μωσαϊκού των πολυφασματικών εικόνων από τις MSFA εικόνες, που λαμβάνονται σαν παράγωγα από μία snapshot camera. Συγκεκριμένα, εφαρμόζουμε μία training-free μεθοδολογία που αξιοποιεί την ικανότητα ενός τυχαία αρχικοποιημένου βαθύ νευρωνικού δικτύου να λειτουργεί σαν image prior, όπως αποδείχθηκε στην ερευνητική δημοσίευση [2]. Με άλλα λόγια διαχειριζόμαστε το πρόβλημα του MSFA demosaicing σαν ένα πρόβλημα αντιστροφής και το Deep Image Prior (DIP) -που προέρχεται από την ευρύ οικογένεια των image generator networks- αποτελεί το μέσο για τον υπολογισμό των ανακατασκευασμένων MSFA εικόνων [3]. Εδώ είναι σημαντικό να αναφερθεί, ότι δεδομένης της βιβλιογραφικής ανάλυσης που προηγήθηκε σε προηγούμενο κεφάλαιο των ερευνητικών εργασιών (βλ §2.2.2. & §2.2.3) καθώς και του θεωρητικού υποβάθρου (§2.1) θα παραλειφθεί η εκτενής ανάλυση εννοιών που έχουν αναπτυχθεί ήδη.

Αυτό που εκμεταλλευόμαστε στην προκειμένη περίπτωση είναι η ιδιότητα των Συνελικτικών Νευρωνικών Δικτύων να μοντελοποιούν καλύτερα πραγματικές – φυσικές εικόνες σε σχέση με το θόρυβο, που θεωρείται γραμμική παραμόρφωση της εικόνας, δηλαδή λειτουργούν σαν priors για το χώρο των εικόνων. Αυτό σημαίνει ότι το ίδιο το δίκτυο λόγω της δομής του και των συνελίξεων έχει μία τάση να έχει μικρότερη αντίσταση στο να αναπαράγει πραγματικές εικόνες από το να αναπαράγει θόρυβο, δηλαδή δυσκολεύεται να κάνει overfit στο θόρυβο. Έτσι, εκμεταλλευόμαστε το γεγονός αυτό βελτιστοποιώντας τις παραμέτρους του δικτύου με στόχο τη γνωστή πληροφορία (δομή) της εικόνας και την αφαίρεση του θορύβου (ενώ μπορεί να χρησιμοποιηθεί σε άλλες εφαρμογές, όπως inpainting). Συνεπώς, το δίκτυο επιτελεί μία διαδικασία βελτιστοποίησης ανά εικόνα αλλά επειδή αργεί να ανακατασκευάσει τον (υψηλής συχνότητας) θόρυβο σε σχέση με τη δομή της εικόνας λειτουργεί σαν prior. Τέλος, αυτό που αξίζει να υπογραμμιστεί είναι ότι το prior είναι το ίδιο το δίκτυο μαζί με τις παραμέτρους του. Μάλιστα, η λύση δεν αναζητείται στον χώρο της εικόνας x αλλά σε έναν άλλο χώρο θ , που είναι ο χώρος των παραμέτρων του δικτύου.

Όσον αφορά την εργασία που εκτελούμε το demosaicing γίνεται άμεσα αντιληπτό δεδομένου ότι και οι RGB cameras βασίζονται σε αυτή την αρχή για να παράξουν τις εικόνες από τον αισθητήρα σε τρία κανάλια, αφού στηρίζονται στη δομή

του Bayer matrix. Σε κάτι αντίστοιχο αλλά υπολογιστικά πιο απαιτητικό εργαζόμαστε και εμείς καθώς από μία MSFA με κάρναβο 4x4 παράγουμε 16 κανάλια. Η εργασία αυτή θεωρείται πιο απαιτητική καθώς η υποδειγματοληψία (subsampling) είναι πιο αραιή, δηλαδή όσο πιο μεγάλο είναι το $N \times N$ παράθυρο τόσο πιο δύσκολη είναι η ανακατασκευή του γιατί η πληροφορία που έχει καταγραφεί από τον αισθητήρα είναι σημαντικά πιο αραιή. Κλείνοντας, για να γίνει σαφές αυτό που επιτελούμε είναι demosaicing που μπορεί να ερμηνευτεί σαν super-resolution σε όλα τα κανάλια (αν και η συγκεκριμένη εργασία δεν είναι το τυπικό παράδειγμα SR) καθώς δεδομένου ότι δεν υπάρχει επαρκής πληροφορία για όλα τα εικονοστοιχεία για να καλυφθεί η ίδια χωρική διάσταση της αρχικής εικόνας ανά κανάλι χρειάζεται να γίνει super-resolution.

Ανακεφαλαιώνοντας η μεθοδολογία που αναπτύσσεται στην παρούσα διπλωματική εργασία αφορά έναν αλγόριθμο demosaicing που είναι **Instance based**, δηλαδή χειρίζεται κάθε φορά από μία εικόνα και **Training free** με την έννοια ότι δεν υπάρχει προγενέστερο στάδιο εκπαίδευσης από άλλα δεδομένα με το ίδιο distribution. Συνεπώς, χρησιμοποιώντας το prior και την MSFA εικόνα που είναι subsample των 16 (στην προκειμένη περίπτωση) πολυφασματικών καναλιών μπορεί να παραχθεί η αντίστοιχη πολυφασματική εικόνα πλήρης ανάλυσης. Εδώ να σημειωθεί ότι όταν γίνεται λόγος για prior (δηλαδή συμπλήρωση πληροφορίας που είναι άγνωστη) αναφερόμαστε σε inverse problems. Γίνεται σαφές, λοιπόν, ότι σαν δεδομένα εισόδου του αλγορίθμου ορίζονται το δίκτυο (σαν δομή πρακτικά) και η MSFA εικόνα.

Όσον αφορά το κομμάτι της υλοποίησης είναι σκόπιμο να αναφερθεί ότι η παραπάνω ιδιότητα των CNN -να λειτουργούν σαν priors- αξιοποιείται στο πρόβλημα των πολυφασματικών αισθητήρων επίπεδης διάταξης φίλτρων ή αλλιώς MSFA. Αναλυτικότερα, το MD²IP χρησιμοποιείται για λόγους MSFA demosaicing με σκοπό την ανακατασκευή όλων των φασματικών καναλιών στην πλήρη χωρική ανάλυση. Αυτό γίνεται ως εξής:

- Αρχικά ορίζουμε τη δομή ενός πλήρως συνελκτικού δικτύου τύπου U-net, που αποτελεί το prior μας. Η διαφοροποίηση με ένα τυπικό δίκτυο U-net έγκειται στο γεγονός ότι επεμβαίνουμε στον αριθμό των συνελκτικών στρωμάτων ανά επίπεδο (εν αντιθέσει με ένα κλασσικό U-net που έχει συγκεκριμένο αριθμό) και επιπλέον στα skip connections που έχουν convolutions αντί να είναι direct connections. Εδώ έχει ενδιαφέρον να αναφερθεί ότι βιβλιογραφικά έχουν χρησιμοποιηθεί και άλλα δίκτυα όπως το προεκπαιδευμένο ResNet αλλά δεν απέδωσαν τόσο καλά αποτελέσματα. Μάλιστα, στη [\[2\]](#) αναφέρεται ρητά ότι η

κατάλληλη επιλογή αρχιτεκτονικής επηρεάζει σημαντικά τη διαδικασία βελτιστοποίησης σε εργασίες ανακατασκευής εικόνων.

- Έπειτα δειγματοληπούμε τυχαίο θόρυβο (υπό την μορφή τανυστή Z) σαν input του δικτύου και του ζητάμε μέσω διαδοχικών επαναλήψεων με backpropagation να ελαχιστοποιήσουμε το loss μεταξύ της συνθετικής ($MSF\bar{A}$) και της πραγματικής εικόνας ($MSF\hat{A}$). Μέσω αυτής της διαδικασίας που διατρέχει το MD²IP μπορούν να κατασκευαστούν όλα τα κανάλια χωρίς να έχει προηγηθεί training. Πρόκειται δηλαδή για μία εργασία βελτιστοποίησης ανά εικόνα όπου το δίκτυο και οι παράμετροι του λειτουργούν σαν prior και ανακατασκευάζουν πολύ καλά τα κανάλια, όπως αποδεικνύεται και από τα αποτελέσματα στην ενότητα 3.3.2. που ακολουθεί μέσω ποσοτικής και ποιοτικής αξιολόγησης.
- Σε επόμενο στάδιο ορίζουμε **output του δικτύου** να είναι τα 16 κανάλια συγκεκριμένης διάστασης. Το δίκτυο ανατρέχει στην εκάστοτε αρχική $MSF\hat{A}$ όπου γίνεται η βελτιστοποίηση (optimization) μέσω της διαδικασίας του backpropagation και ρυθμίζει τις παραμέτρους του δικτύου ώστε να δημιουργήσουν την πολυφασματική εικόνα πλήρης ανάλυσης με 16 κανάλια. Στη διαδικασία αυτή χρησιμοποιείται σαν συνάρτηση κόστους (loss function) το μέσο τετραγωνικό σφάλμα (Mean Squared Error). Στο loss του δικτύου συμμετέχουν οι $MSF\hat{A}$, και οι αρχικές πολυφασματικές εικόνες πλήρης ανάλυσης με τα 16 κανάλια (Multispectral Groundtruth), με τις δεύτερες να συμβάλλουν στο στάδιο ποιοτικής αξιολόγησης της μεθόδου σαν validation loss. Η σύγκριση του loss function γίνεται μεταξύ της συνθετικής $MSF\bar{A}$ και της πραγματικής $MSF\hat{A}$ εικόνας. Τέλος, σε όλα τα κανάλια που συμμετέχουν στην πολυφασματική απεικόνιση συμπληρώνεται πληροφορία για κάθε εικονοστοιχείο ώστε να ανταποκρίνονται στη χωρική ανάλυση της αρχικής $MSF\hat{A}$ εικόνας, που έχει προκύψει σαν «υποδειγματοληψία» των καναλιών αυτών.

Υπενθυμίζεται ότι οι Multispectral Groundtruth είναι γνωστές από τα σετ δεδομένων που λήφθηκαν και προσομοιάστηκαν με το μοτίβο MSFA 4x4 στο στάδιο του preprocessing που αναλύθηκε παραπάνω. Κρίσιμο σημείο της διαδικασίας που αναλύουμε είναι ότι οι εν λόγω εικόνες (MS Groundtruth), δεν θα τις είχαμε υπό «κανονικές» συνθήκες όπου τα δεδομένα μας θα ήταν από snapshot πολυφασματική κάμερα όπως η IMEC, παρά μόνο την πραγματική $MSF\hat{A}$.

- Ακολουθεί ο μετασχηματισμός από 16 κανάλια σε MSFA grid, που συνιστά την τελική συνθετική εικόνα (MSFĀ). Εδώ υπεισέρχονται και οι μετρικές αξιολόγησης (PSNR, SSIM) που υπολογίζονται για κάθε επανάληψη. Για καθεμία από αυτές καταγράφονται δύο τιμές μία που αφορά την απόκλιση μεταξύ των (i) MSFA και των (ii) 16 φασματικών καναλιών. Συνεπώς, σημειώνονται οι μετρικές PSNR MSFA, PSNR merged, SSIM MSFA και SSIM merged. Εδώ επισημαίνεται πάλι ότι η μέθοδος είναι training free, δηλαδή το αποτέλεσμα της PSNR merged λειτουργεί σαν validation loss, δηλαδή δεν επεμβαίνει στη βελτιστοποίηση του αλγορίθμου, μιας και είναι κάτι που δεν θα παρεχόταν σαν δεδομένο στην πραγματικότητα. Αντιθέτως, η μετρική PSNR MSFA θα είναι πάντα υπολογίσιμη, αφού προκύπτει από τη διαφορά της συνθετικής εικόνας που κατασκευάζεται από το δίκτυο με την πραγματική (αυτή που δίνει μία snapshot κάμερα). Ιδανικά αν εξασφαλιστεί μία καλή επίδοση στις μετρικές που αφορούν την MSFA αυτό συνεπάγεται βελτίωση και στις μετρικές merged, που μπορεί να είναι γνωστές μόνο σαν validation set καθώς θα προϋπέθετε να διατίθεται η 16κάναλη πολυφασματική εικόνα, που στην πραγματικότητα δεν την έχουμε σε προβλήματα που αντιμετωπίζουμε με κάμερες που υιοθετούν τη δομή των MSFAs.
- Ένα ακόμα κρίσιμο κομμάτι του αλγορίθμου είναι ο αριθμός των επαναλήψεων (Iterations) που θα υλοποιηθούν μιας και δεν υπάρχει κάποια στρατηγική πρόταση. Μία τέτοια επιλογή αποτελεί υπερπαραμέτρο του δικτύου που έπειτα από ορισμένους πειραματισμούς καταλήξαμε στις 5000 επαναλήψεις, καθώς σε αυτό το εύρος μπορούσαμε να έχουμε πλήρη εποπτεία της πορείας σύγκλισης του αλγορίθμου μέσω των μετρικών αξιολόγησης. Τέλος, η πολυφασματική εικόνα που επιλέγεται να είναι το τελικό παράγωγο του δικτύου είναι αυτή που σημειώνει τη βέλτιστη τιμή μετρικής PSNR merged.

Καταληκτικά, η συγκεκριμένη μεθοδολογία μπορεί να εφαρμοστεί σε μια οποιαδήποτε snapshot κάμερα του εμπορίου (silios, Ximea, Imec) όπου υπό διαφορετικές συνθήκες θα έπρεπε να είναι γνωστά τα φίλτρα σε κάθε κάμερα και θα χρειαζόταν να κατασκευαστεί εξειδικευμένο σετ δεδομένων για καθεμία από αυτές. Αντιθέτως, τώρα το MD²IP λειτουργεί για κάθε περίπτωση αρκεί να γνωρίζεις το μοτίβο επαναληψιμότητας του αισθητήρα, όπου μάλιστα δεν χρειάζεται να είναι γνωστά ποια είναι ακριβώς τα κανάλια αυτά αρκεί να είναι συνεπής η θέση που καταγράφεται πληροφορία για το καθένα.

3.3.2. Αποτελέσματα και αξιολόγηση του μοντέλου MD²IP

Στα πλαίσια της παρούσας διπλωματικής διατριβής επιλέγεται ο αλγόριθμος να αποθηκεύει την εικόνα εκείνη που στην πάροδο των 5000 επαναλήψεων (συνθήκη τερματισμού) σημειώνει την καλύτερη επίδοση στην μετρική PSNR merged. Υπενθυμίζεται, ότι το δίκτυο που εφαρμόζεται στηρίζεται στο Deep Image Prior που αποσκοπεί στη βελτιστοποίηση, που με τη σειρά της έχει στόχο το overfit. Επομένως, είναι λογικό να επιδιώξουμε το βέλτιστο αποτέλεσμα για κάθε πολυφασματική εικόνα βάση της κατάλληλης μετρικής. Με άλλα λόγια, η επιλογή αυτή είναι στοχευμένη καθώς έτσι θα εξαχθούν τα καλύτερα δυνατά αποτελέσματα δεδομένης της δομής του δικτύου. Συνεπώς θα εξαχθεί η πολυφασματική εικόνα που προσεγγίζει και προσομοιάζει στον μέγιστο βαθμό την πραγματική πολυφασματική εικόνα 16 καναλιών και κατ' επέκταση την πραγματική MSFĀ. Ευρύτερα, στόχος του δικτύου είναι μέσω της συνάρτησης κόστους να ελαχιστοποιηθεί η διαφορά μεταξύ συνθετικής και πραγματικής MSFA εικόνας (minimize (MSFĀ - MSFĀ)) και κατ' επέκταση αυτό να εκφραστεί μέσω της μεγιστοποίησης της αντίστοιχης μετρικής (maximize (PSNR MSFA)). Επίσης, η επιλογή αυτή θα εξυπηρετήσει σημαντικά στην μεθοδολογία που ακολουθεί στο 4^ο Κεφάλαιο καθώς τα δεδομένα εξόδου του MD²IP θα αποτελέσουν τα δεδομένα εκπαίδευσης για την εκπαίδευση ενός άλλου δικτύου. Οπότε, είναι σκόπιμο το νέο δίκτυο που θα εκπαιδευτεί να έχει τα βέλτιστα δεδομένα εισόδου για να κριθεί με τη σειρά του η επίδοση του. Τέλος, στη συγκεκριμένη ενότητα θα γίνει αξιολόγηση του μοντέλου που διακρίνεται σε δύο μέλη: αυτό της ποσοτικής και ποιοτικής αξιολόγησης μέσω γραφημάτων και εικόνων προς απόδειξη και τεκμηρίωση της επιτυχίας των αποτελεσμάτων που εξήχθησαν.

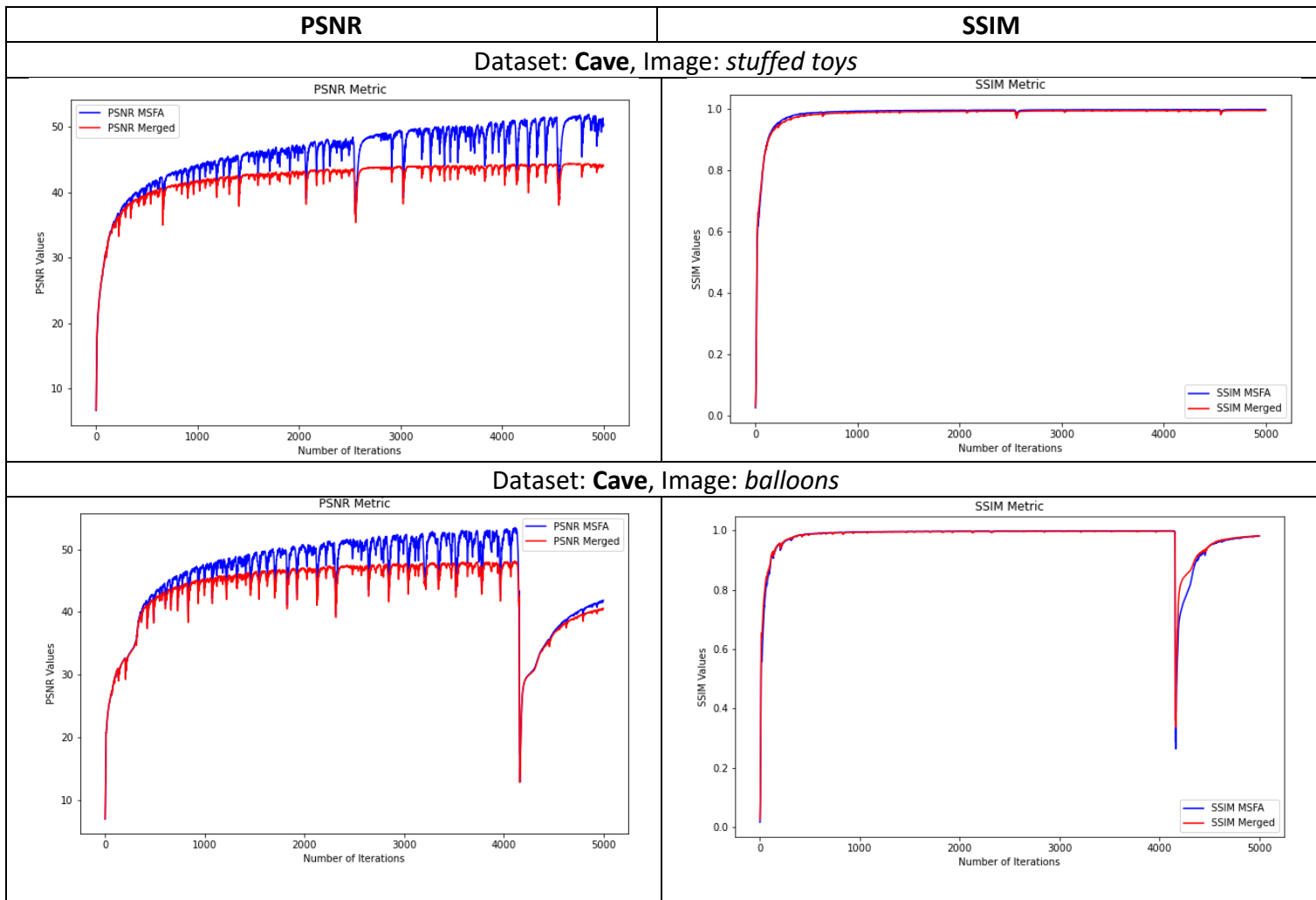
Ποσοτική Αξιολόγηση

Για το στάδιο της ποσοτικής αξιολόγησης κατασκευάστηκαν διαγράμματα που οπτικοποιούν την εξέλιξη των μετρικών PSNR και SSIM. Συγκεκριμένα, στα Σχήματα 3.9, 3.10, 3.11 και 3.12 που ακολουθούν παρατηρούνται οι αποδόσεις των μετρικών PSNR και SSIM από δύο αντιπροσωπευτικές εικόνες για κάθε dataset. Συνολικά, γίνεται άμεσα αντιληπτό ότι σε κάθε μετρική η τάση καθώς και ο αριθμός των επαναλήψεων στις οποίες αρχίζει να συγκλίνει ο αλγόριθμος είναι κοινές.

Αναλυτικότερα, στο Σχήμα 3.9 παρουσιάζονται τα διαγράμματα του σετ δεδομένων «Cave» για τις εικόνες «*stuffed toys*» και «*balloons*». Αρχικά, πέρα από την κοινή τάση που είναι εμφανής και για τις δύο μετρικές φαίνεται ότι μεταξύ PSNR MSFA

και PSNR Merged διατηρείται μετά τις 2000 επαναλήψεις μία σταθερή απόκλιση τιμών της τάξης 5-8 μονάδες ενώ οι τιμές για την PSNR Merged είναι πολύ υψηλές στο εύρος των 40+ μονάδων. Ταυτόχρονα υπάρχει μία τάση «μμητισμού» της μίας ως προς την άλλη, κάτι που αναφέρθηκε και νωρίτερα. Δηλαδή, είναι επόμενο όταν κατά την βελτιστοποίηση και την αλλαγή των παραμέτρων του δικτύου η συνθετική MSF \hat{A} συγκλίνει προς την αρχική MSF \bar{A} τότε και τα αντίστοιχα φασματικά κανάλια της MS να προσεγγίζονται μεταξύ τους. Επιπλέον, είναι εντυπωσιακή η ταχύτητα σύγκλισης της SSIM, η οποία καταγραφεί μία απότομη άνοδο στις πρώτες 300 μόλις επαναλήψεις όπου προσεγγίζει τη μονάδα (1) και έπειτα συγκλίνει. Ακόμα παρατηρείται ακόμα μεγαλύτερη ταύτιση τιμών μεταξύ SSIM MSFA και SSIM Merged σε σχέση με τις μετρικές που αφορούν την PSNR.

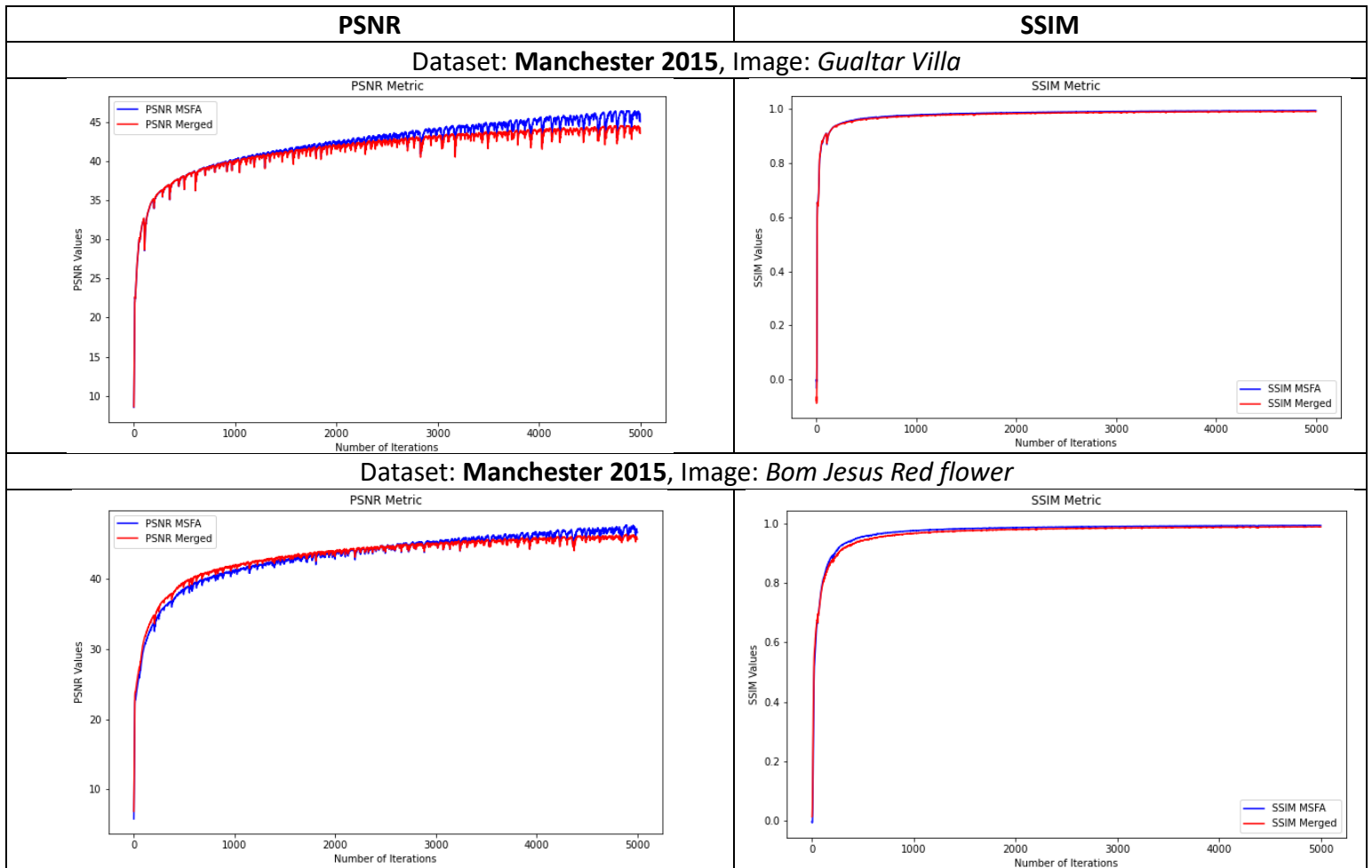
Σχετικά με την δεύτερη εικόνα που παρατίθεται για το σετ δεδομένων «Cave» και πιο συγκεκριμένα για την εικόνα «balloons» έχει ενδιαφέρον να σχολιαστεί η σημαντική πτώση τιμών που καταγράφεται και αφορά αμφότερες τις μετρικές. Για να γίνει σαφές σημειώθηκε μεγάλη πτώση από την επανάληψη 4150 και έπειτα μέχρι το σημείο καμπής που ήταν στην 4168 επανάληψη από όπου ξεκίνησε εκ νέου να εκπαιδεύεται ο αλγόριθμος με ταχύτετους ρυθμούς. Η διαφορά τιμών που καταγράφει η PSNR Merged είναι 31,1 μονάδες ενώ για την SSIM merged 0,63. Το συγκεκριμένο συμβάν μπορεί να ερμηνευτεί από τη δομή του αλγορίθμου σύμφωνα με την οποία όταν οι τιμές των μετρικών συγκλίνουν για πεπερασμένο αριθμό επαναλήψεων τότε οι παράμετροι του δικτύου αρχικοποιούνται εκ νέου με τυχαίο τρόπο ώστε να αποφευχθεί ο «εγκλωβισμός» του μοντέλου σε κάποιο τοπικό ελάχιστο ή σημείο «σέλλας», όπου η παράγωγος της συνάρτησης κόστους δε μεταβάλλεται. Με αυτόν τον τρόπο το δίκτυο -προσθέτοντας εκ νέου το στοιχείο της τυχειότητας (μέσω θορύβου) και στηριζόμενοι στο γεγονός ότι το MD²IP κάθε φορά που θα διατρέχεται θα έχει παρόμοια αποτελέσματα με μικρές αποκλίσεις μεταξύ τους- ωθεί το μοντέλο να επανυπολογίσει τις παραμέτρους μέσω της μεθόδου της βελτιστοποίησης και εν τέλει ιδανικά να συγκλίνει στο ολικό ελάχιστο της συνάρτησης κόστους, κοινώς να προσομοιαστεί στο βέλτιστο βαθμό η πολυφασματική εικόνα με τα 16 κανάλια που εξάγει με αυτά που έχουμε ως δεδομένα αληθείας. Τέλος, έχει ενδιαφέρον η παρατήρηση ότι τα σημεία «ορόσημα» ως προς τον αριθμό των επαναλήψεων, δηλαδή αυτό της απότομης πτώσης και της επανέναρξης της εκπαίδευσης του δικτύου συμφωνούν απόλυτα μεταξύ των μετρικών PSNR και SSIM, κάτι που βέβαια αναμενόταν. Το συγκεκριμένο φαινόμενο όπου το μοντέλο επανεκπαιδεύεται εσωτερικά παρατηρήθηκε σε αρκετές εικόνες κυρίως στο Cave dataset για αυτό κρίθηκε απαραίτητο να σημειωθεί.



Σχήμα 3.9: Αποτελέσματα μετρικών PSNR και SSIM για το σετ δεδομένων Cave για δύο ενδεικτικές εικόνες.

Ακολουθεί το Σχήμα 3.10 όπου παρουσιάζονται τα διαγράμματα του σετ δεδομένων «Manchester 2015» για τις εικόνες «*Gualtar Villa*» και «*Bom Jesus Red flower*». Ομοίως με το Cave οι τιμές που σημειώνουν οι μετρικές είναι πολύ υψηλές αφού η PSNR κυμαίνεται στις 40-45 μονάδες και η SSIM τείνει στο 1, δηλαδή γίνεται λόγος για πλήρη ομοιότητα ως προς τα δομικά στοιχεία μεταξύ δύο σημάτων εικόνων. Στις εν λόγω εικόνες μάλιστα η μετρική PSNR MSFA και PSNR merged φαίνεται να συγκλίνουν μετά τις 3000 επαναλήψεις και σε σχέση με το προηγούμενο σετ δεδομένων παρατηρείται ότι -παρότι καθυστερούν λίγο περισσότερο να συγκλίνουν οι εικόνες του M15- οι γραφικές αναπαραστάσεις τους οριακά ταυτίζονται ιδίως μέχρι το σημείο σύγκλισης. Από τις 3000 επαναλήψεις και έπειτα φαίνεται ότι το μοντέλο

καταφέρνει να προσεγγίσει καλύτερη επίδοση για την MSFA και οριακά μικρότερη για τις merged (κανάλια πολυφασματικής εικόνας). Τέλος, έχει ενδιαφέρον το γεγονός ότι στην εικόνα «*Bom Jesus Red flower*» η μετρική PSNR merged ξεπερνάει παροδικά μέχρι τις 2800 επαναλήψεις την PSNR MSFA, δηλαδή σχεδόν μέχρι τη στιγμή όπου το μοντέλο αρχίζει να σταθεροποιείται όπου επανέρχεται η εκτιμώμενη ισορροπία. Σε αυτό το σημείο κρίνεται σημαντικό να θυχθεί ότι οι τόσο υψηλές αποδόσεις του αλγορίθμου σε αυτά τα δύο σετ δεδομένων αποδίδεται αρχικά στην αρχιτεκτονική του δικτύου και ως έναν βαθμό στην σχετικά χαμηλότερη ποιότητα των εικόνων και την παρουσία θορύβου σε αυτές σε σχέση με αυτή των πολυφασματικών εικόνων που θα ακολουθήσουν από τα dataset TT31 και TT59. Κάτι τέτοιο θα καταστήσει μελλοντικά πιο δύσκολη την ομοιότητα μεταξύ συνθετικών και πραγματικών δεδομένων και θα δοκιμάσει ακόμη περισσότερο την απόδοση του MD²IP.

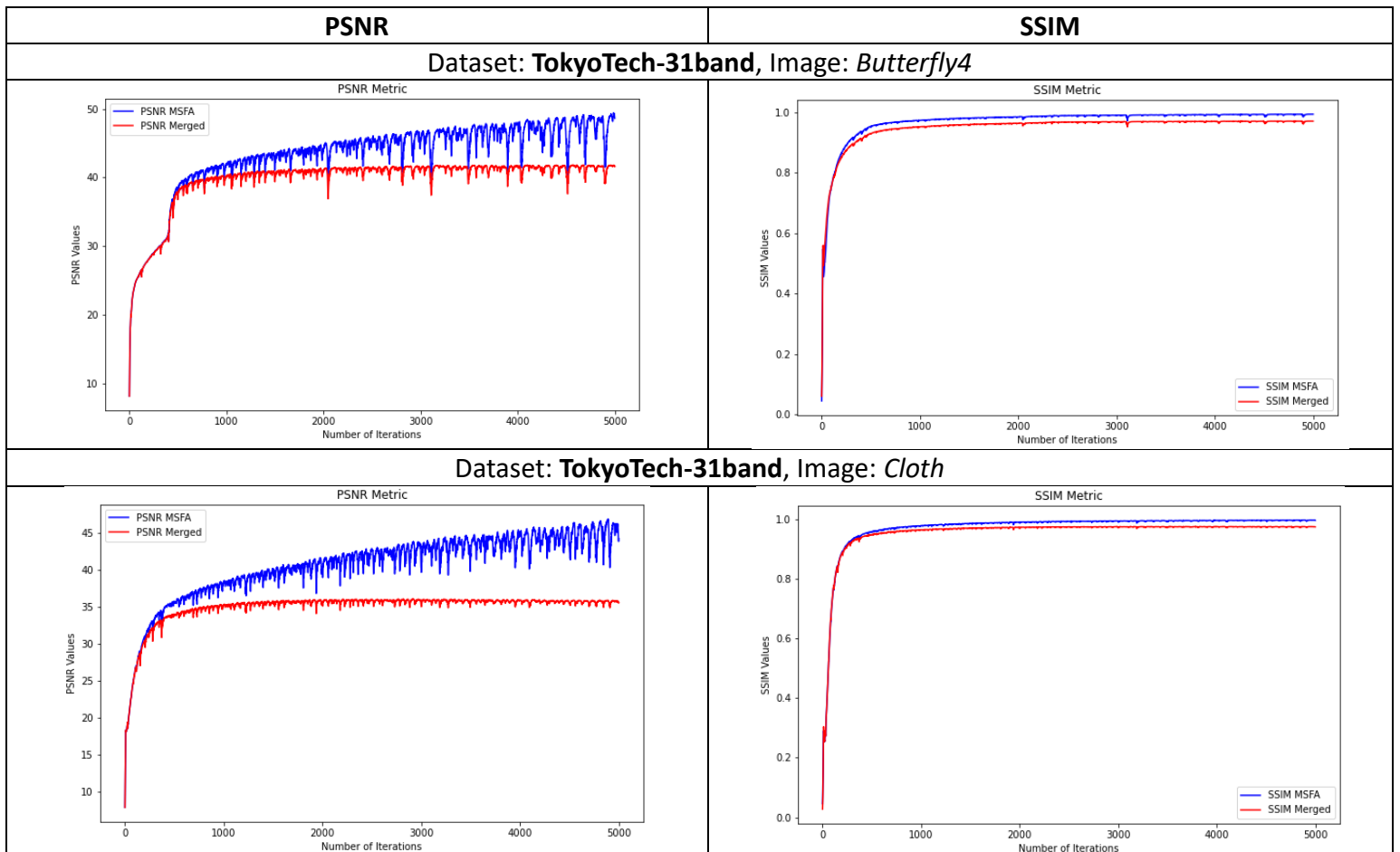


Σχήμα 3.10: Αποτελέσματα μετρικών PSNR και SSIM για το σετ δεδομένων Manchester 2015 για δύο ενδεικτικές εικόνες.

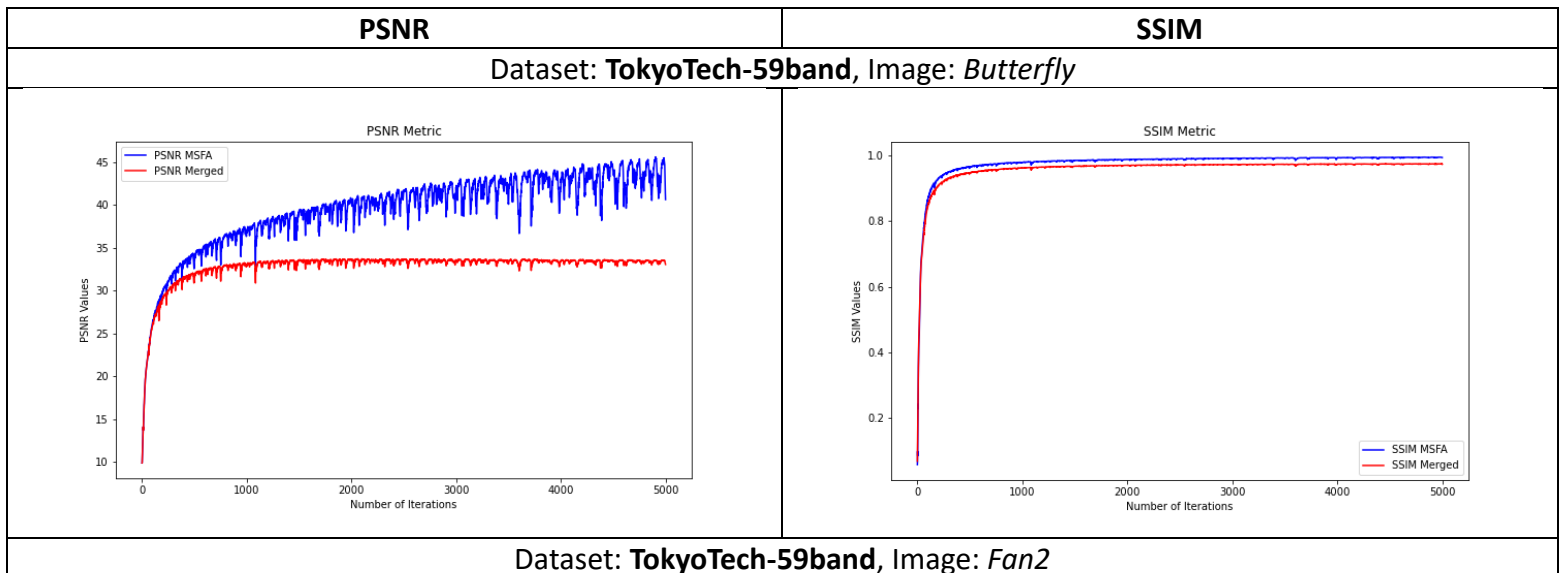
Στο Σχήμα 3.11 που παρατίθεται ακολούθως έχουν επιλεγεί για το σετ δεδομένων «TokyoTech-31band» τα διαγράμματα που αφορούν τις εικόνες «*Butterfly4*» και «*Cloth*». Αντίστοιχα στο Σχήμα 3.12 παρουσιάζονται τα διαγράμματα του σετ δεδομένων «TokyoTech-59band» για τις εικόνες «*Butterfly*» και «*Fan2*». Σε αυτά τα δύο dataset οι συμπεριφορές των μετρικών παρουσιάζουν πολλά κοινά, γεγονός που μπορεί να αποδοθεί στο ότι οι εικόνες που έχουν εξαχθεί πηγάζουν αμφότερες από μονοχρωματική κάμερα με VariSpec liquid crystal tunable φίλτρα. Όσον αφορά την SSIM υπάρχει σύγκλιση των τιμών νωρίτερα από τις 1000 επαναλήψεις και προσεγγίζονται τιμές κοντά στη μονάδα, κοινώς το μοντέλο φαίνεται να έχει αποδώσει τα βέλτιστα.

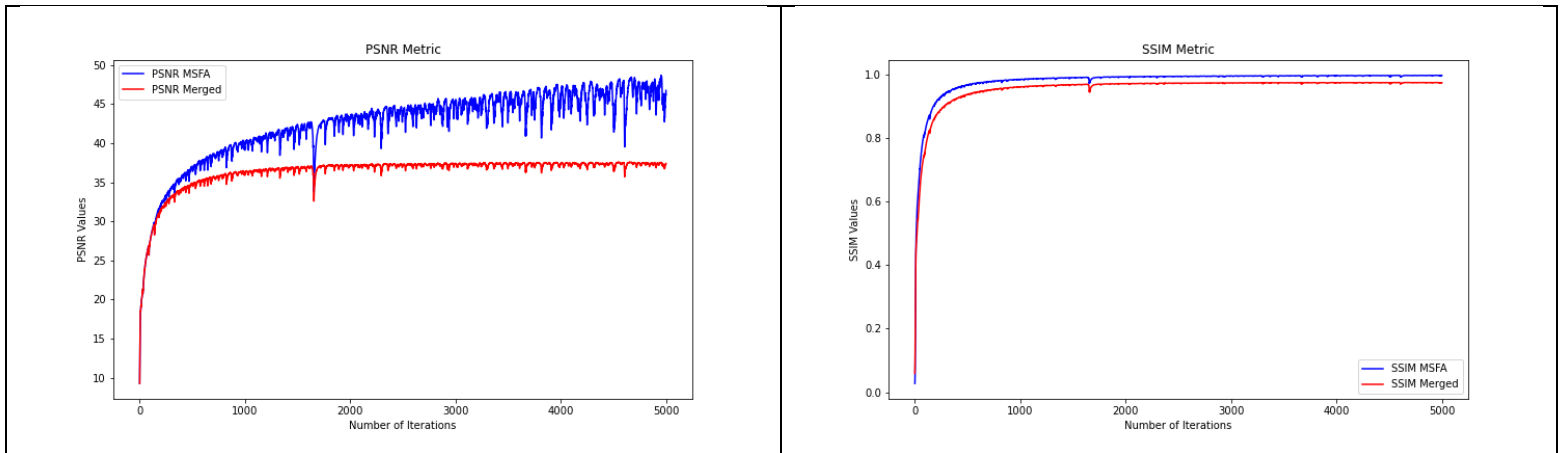
Στα TT31 και TT59 παρατηρούνται σχετικά μικρότερες τιμές ως προς την PSNR merged της τάξης 30-35, με εξαίρεση την εικόνα «*Butterfly4*» που σημειώνει μέχρι και 40. Κάτι τέτοιο καταλογίζεται στην καλύτερη ευκρίνεια των εικόνων γεγονός που συνεπάγεται μεγαλύτερες απαιτήσεις για το δίκτυο ώστε να ανταποκριθεί και να προσεγγίσει κατά τον βέλτιστο βαθμό την πραγματική εικόνα. Παράλληλα, είναι αντιληπτή η ταχύτερη σύγκλιση των μετρικών PSNR merged σε σχέση με όλα τα υπόλοιπα dataset, στις μόλις 1000 επαναλήψεις. Με βραδύτερο ρυθμό συγκλίνει και η μετρική PSNR MSFA που ακολουθεί μέχρι τέλους ανοδική πορεία και προσεγγίζει πολύ υψηλές τιμές στο εύρος 45-50. Επομένως, το μοντέλο επιτελεί το σκοπό του αποδίδοντας πολύ ικανοποιητικά ακόμα και σε πιο απαιτητικές εικόνες υψηλής ανάλυσης.

Ακόμα, είναι εμφανής η μεγαλύτερη απόκλιση όσον αφορά τις επιμέρους μετρικές της PSNR, δηλαδή μεταξύ των PSNR MSFA και PSNR merged. Αυτό αποδίδεται και πάλι ως ένα βαθμό στην υψηλότερη ανάλυση των εικόνων που υπάρχουν στα TT31, TT59. Επομένως αναμένεται το μοντέλο να «φιτάρει» καλύτερα στην PSNR MSFA από ότι στην PSNR Merged αφού αφενός έχει εκπαιδευτεί βάση της συνάρτησης κόστους που στοχεύει στην ελαχιστοποίηση της διαφοράς μεταξύ της πραγματικής εικόνας $MSF\hat{A}$ με την συνθετική $MSF\bar{A}$ και αφετέρου τα κανάλια που απαρτίζουν την MS έχουν προκύψει από το $1/16$ της πληροφορίας που είναι καταγεγραμμένη στην MSFA μιας και η τελευταία αποτελεί υποδειγματοληψία της MS.



Σχήμα 3.11: Αποτελέσματα μετρικών PSNR και SSIM για το σετ δεδομένων TokyoTech-31band για δύο ενδεικτικές εικόνες.



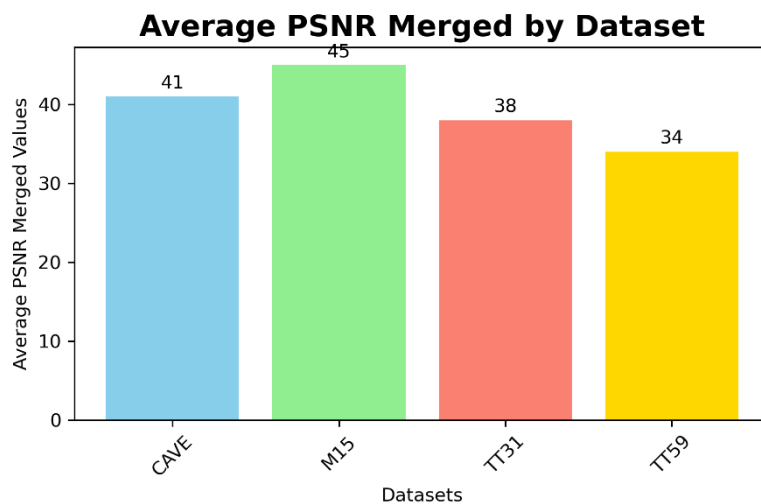


Σχήμα 3.12: Αποτελέσματα μετρικών PSNR και SSIM για το σετ δεδομένων TokyoTech-59band για δύο ενδεικτικές εικόνες.

Ευρύτερα από όλα τα Σχήματα που παρατέθηκαν (3.9, 3.10, 3.11 και 3.12), έχουν ενδιαφέρον οι υψηλές τιμές που κατέγραψαν οι μετρικές από την περίοδο σύγκλισης και έπειτα που κυμαίνονται μεταξύ 40-50 για την PSNR και έως πολύ κοντά στη μονάδα για την SSIM, κάτι που δείχνει το βαθμό προσέγγισης και ομοιότητας ακόμα και ως προς τα δομικά χαρακτηριστικά των εικόνων. Όσον αφορά τη μετρική SSIM παρατηρείται σε όλα τα Σχήματα συνολικά ακόμα μεγαλύτερη ταύτιση των γραφικών παραστάσεων και σύγκλιση ήδη από τις 1000 επαναλήψεις.

Συνολικά ο αλγόριθμος αποδεικνύει ότι επιτελεί τον στόχο της βελτιστοποίησης ανά εικόνα καθώς όλες οι μετρικές σημειώνουν πολύ υψηλές τιμές και η ποιότητα της ανακατασκευασμένης εικόνας είναι πολύ ικανοποιητική. Μάλιστα παρατηρείται ότι σε στιγμιότυπα όπου η PSNR MSFA παρουσιάζει τοπικά ελάχιστα ή μέγιστα στον ίδιο αριθμό επαναλήψεων παρουσιάζει ίδια συμπεριφορά και η PSNR Merged. Γεγονός που μας βοηθάει να γενικεύσουμε σε μελλοντικές εργασίες -που θα διαθέτουμε αποκλειστικά σαν δεδομένο την πραγματική εικόνα MSF \hat{A} , όπως θα έχει προκύψει από πολυφασματικούς αισθητήρες επίπεδης διάταξης φίλτρων- άρα θα είναι υπολογίσιμη μόνο η PSNR MSFA και αντίστοιχα η SSIM MSFA. Σε αυτές τις περιπτώσεις μπορεί να ισχυριστεί, βάση των παραπάνω γραφημάτων, ότι όταν επιλέγει μία τιμή PSNR MSFA που αντιστοιχεί σε τοπικό μέγιστο (T.M.) αυτό είναι πολύ πιθανό να είναι T.M. και για τη PSNR Merged.

Σε επόμενο στάδιο θεωρήθηκε ενδιαφέρον να υπολογιστεί μία μέση τιμή από τις μέγιστες PSNR merged που σημειώθηκαν ανά σετ δεδομένων. Με άλλα λόγια, επιλέχθηκαν οι $\max(\text{PSNR merged})$ ανά εικόνα -κοινώς η καλύτερη μετρική που σημείωσε η εκάστοτε εικόνα και αποτέλεσε κριτήριο για την αποθήκευση της τελικά σαν παράγωγο του MD²IP -. Στη συνέχεια επί αυτών των τιμών υπολογίστηκε μία μέση τιμή για κάθε dataset. Αναλυτικότερα, το Cave σημείωσε μέση τιμή ή αλλιώς $\text{Avg}(\text{Max}(\text{PSNR Merged}))$: 41, το Manchester 2015: 45(44,75 για ακρίβεια), το TokyoTech-31band: 38 και το TokyoTech-59band: 34. Οι παραπάνω υπολογισμοί οπτικοποιούνται στο Σχήμα 3.13 που ακολουθεί.

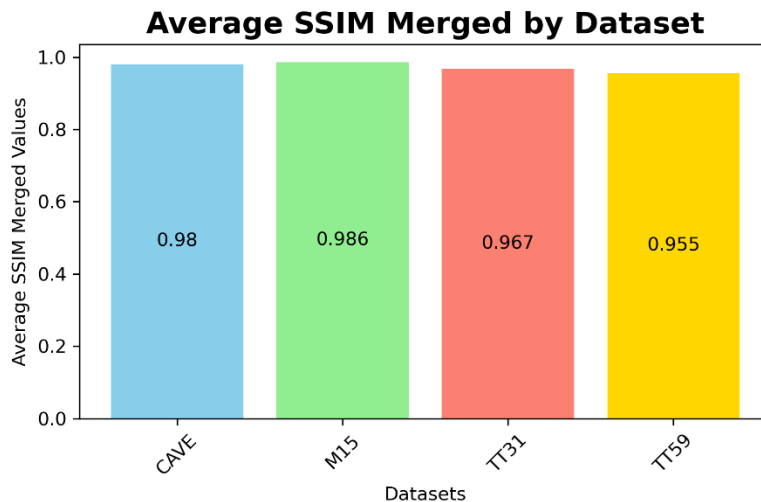


Σχήμα 3.13: Απεικόνιση των μέσων τιμών των μεγίστων PSNR merged των εικόνων για καθένα από τα 4 σετ δεδομένων.

Γενικά, οι τιμές που σημειώνονται ενδεικτικά στην κεφαλίδα της μπάρας κάθε dataset είναι πολύ καλές όπως αναμενόταν και από την ανάλυση που προηγήθηκε με ορισμένα αντιπροσωπευτικά διαγράμματα και φαίνεται να συμφωνούν. Επίσης, στο Σχήμα αυτό παρατηρείται ότι τα πρώτα δύο σετ δεδομένων κατέγραψαν συνολικά αισθητά μεγαλύτερες τιμές PSNR merged και αυτό αποδίδεται στη χαμηλότερη ποιότητα των εικόνων που περιέχονται στα συγκεκριμένα dataset και φυσικά στην παρουσία θορύβου που συνεπάγεται πιο θολές εικόνες με λιγότερη πληροφορία και συνολικά την αίσθηση μίας πιο ομαλής εικόνας με εξασθενημένες ακμές. Αυτό ως είναι επόμενο καθιστά την αναπαράσταση της πολυφασματικής εικόνας αρκετά πιο «εύκολη» αφού το δίκτυο δεν μπορεί να εντρυφήσει σε εκτενέστερες λεπτομέρειες άρα η πληροφορία που προσπαθεί να συνθέσει είναι αρκετά πιο απλοϊκή. Συνεπώς, για το σύνολο των εικόνων στα Cave και Manchester 2015 επιτυγχάνονται στην πάροδο των

5000 επαναλήψεων πολύ υψηλές τιμές PSNR merged που αντικατοπτρίζονται από το παραπάνω διάγραμμα.

Εξ' όλων των σετ δεδομένων αυτό που καταγράφει τη μέγιστη τιμή είναι το M15, κάτι που δεν είναι καθόλου τυχαίο διότι είναι αυτό που έχει πολύ υποδεέστερη ποιότητα εικόνων λόγω παρουσίας θορύβου και επιπλέον η κατανομή των κανονικοποιημένων τιμών ανακλαστικότητας στο εύρος 0-1 (αφού διαιρέθηκαν σύμφωνα με τη ραδιομετρική ανάλυση του σετ δεδομένων) είναι πολύ κακή. Κάτι τέτοιο πιθανό να βελτιωνόταν με κάποιο scaling των τιμών κάτι που δεν υλοποιήθηκε καθώς αποφασίστηκε τα δεδομένα από όλα τα σετ να διαχειριστούν με τον ίδιο τρόπο. Για όλους τους λόγους που προαναφέρθηκαν είναι λογικό να εμφανίζει την υψηλότερη μέση τιμή για τη μετρική PSNR merged. Αντίστοιχα για τα TT31 και TT59 αν και η μέση τιμή είναι μικρότερη (38 και 34 αντίστοιχα) δεν παύει να αποτελεί υψηλή τιμή με πολύ καλά αποτελέσματα. Προφανώς, λόγω της καλύτερης ευκρίνειας των εικόνων που περιλαμβάνουν αποτελούν μία περαιτέρω πρόκληση για το μοντέλο του MD²IP, που παρά τις επιπλέον υπολογιστικές απαιτήσεις που συνεπάγονται οι εικόνες αυτές καταφέρνει να συγκλίνει πολύ σύντομα και να σημειώσει πολύ αξιόλογα ποσοτικά αποτελέσματα.



Σχήμα 3.14: Απεικόνιση των μέσων τιμών των SSIM merged των εικόνων για καθένα από τα 4 σετ δεδομένων.

Ακολουθώς, επαναλήφθηκε η ίδια διαδικασία για τη μετρική SSIM με τη διαφορά ότι αυτή τη φορά υπολογίστηκε η μέση τιμή των SSIM merged που αντιστοιχούν στη μέγιστη τιμή της εκάστοτε PSNR merged (δηλαδή βρίσκονται στην ίδια γραμμή εγγραφής, κοινώς στην ίδια επανάληψη). Ο λόγος που επιλέχθηκαν οι SSIM

merged που είναι στην ίδια επανάληψη με την μέγιστη PSNR merged και όχι οι μέγιστες τιμές της πρώτης είναι για να οπτικοποιηθεί πόσο καλά αποδίδει η αποθηκευμένη εικόνα που έχει εξαχθεί από το MD²IP και ως προς τη μετρική SSIM και επιπλέον είναι σημαντικό να αναφερθεί ότι η μέγιστη SSIM δεν είναι απαραίτητο ότι θα αντιστοιχεί στη μέγιστη PSNR. Όσον αφορά τις μέσες τιμές που υπολογίστηκαν το Cave σημείωσε μέση τιμή ή αλλιώς Avg (SSIM Merged): 0,98, το Manchester 2015: 0,986, το TokyoTech-31band: 0,967 και το TokyoTech-59band: 0,955. Οι υπολογισμοί αυτοί παρουσιάζονται παραστατικά μέσω του διαγράμματος του Σχήματος 3.14.

Αυτό που παρατηρείται και στην περίπτωση της μετρικής SSIM merged είναι οι πάρα πολύ υψηλές αποδόσεις του μοντέλου στο σύνολο των δεδομένων, γεγονός που επαληθεύει και τα διαγράμματα που προηγήθηκαν στα Σχήματα 3.9, 3.10, 3.11 και 3.12 που όπως σχολιάστηκαν τείνουν στη μονάδα, κάτι που μεταφράζεται σε άριστη ομοιότητα -ακόμα και ως προς τα δομικά στοιχεία των εικόνων- μεταξύ συνθετικών και πραγματικών καναλιών που συνθέτουν μία πολυφασματική εικόνα. Όπως και στο Σχήμα 3.13 τη μεγαλύτερη τιμή σημειώνει το M15, έπεται το Cave και ακολουθούν τα TT31 και TT59. Οι λόγοι που παρατηρείται αυτή η τάση είναι κοινοί με αυτούς που προαναφέρθηκαν.

Ποιοτική Αξιολόγηση

Στην παράγραφο αυτή θα αξιολογηθεί το μοντέλο με ποιοτικό τρόπο. Πιο συγκεκριμένα δημιουργούνται έγχρωμα σύνθετα που προσπαθούν να προσεγγίσουν το φυσικό RGB -δεδομένου των 16 καναλιών που διατίθενται- για να είναι οπτικά άμεσα κατανοητά από τον αναγνώστη και να συγκλίνουν στην πραγματικότητα όπως την αντιλαμβάνεται το ανθρώπινο μάτι. Τα σύνθετα αυτά κατασκευάζονται τόσο για τις πραγματικές πολυφασματικές εικόνες που αποτελούν και δεδομένα αληθείας για το δίκτυο μας (όσον αφορά τους υπολογισμούς της ποσοτικής αξιολόγησης και όχι κατά την εκπαίδευση) όσο και για τις αντίστοιχες συνθετικές εικόνες που αποτελούν παράγωγα του MD²IP. Παρακάτω, θα παρατεθούν ορισμένα ενδεικτικά παραδείγματα για κάθε σετ δεδομένων προκειμένου να αξιολογηθεί η απόδοση του μοντέλου.

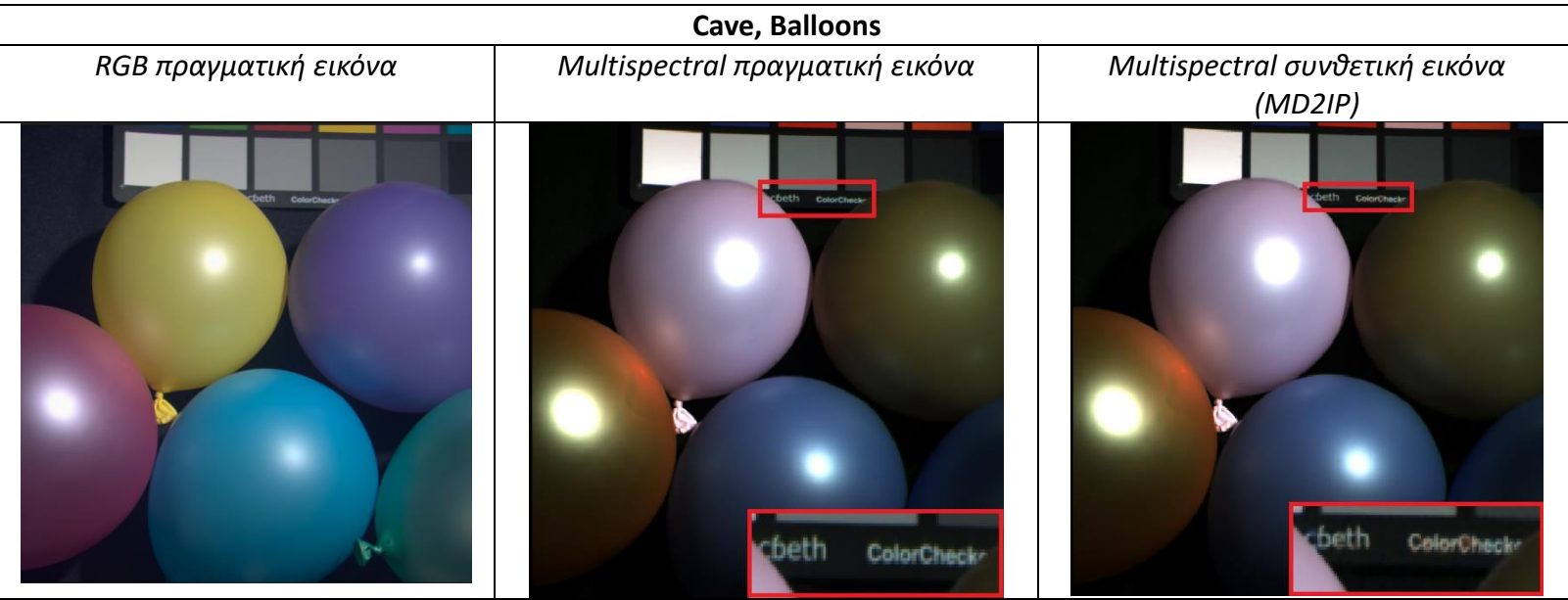
Πριν παρατεθούν προς σύγκριση τα ζευγάρια πραγματικών και συνθετικών πολυφασματικών εικόνων είναι σημαντικό να αναφερθεί το πώς κατασκευάστηκαν. Αναλυτικότερα, αξιοποιώντας τα φασματικά κανάλια της IMEC (βλ. Σχήμα 2.9) κατασκευάζεται ένα έγχρωμο σύνθετο αξιοποιώντας τα κανάλια που αντιστοιχούν καλύτερα στα μήκη κύματος των Red (620-750nm), Green (495-570nm) & Blue(450-

495nm). Επειδή δεν υπάρχει ακριβής αντιστοιχία για τα δεδομένα μήκη κύματος έγιναν διάφορες απόπειρες μέχρις ότου να προσεγγιστεί η αρχική RGB εικόνα του dataset. Σε αυτή τη διαδικασία αντιμετωπίστηκαν ορισμένες δυσκολίες δεδομένου ότι ορισμένα μήκη κύματος όπως αυτό στο πράσινο «εύρος» επικαλύπτεται με το κίτρινο ή βρίσκεται κοντά στα άκρα άλλων και ακόμα από τα 16 κανάλια δεν υπήρχε κάποιο που να καταγράφει πληροφορία στο εύρος του Red, πέραν του 15^{ου} (621nm) και 16^{ου} καναλιού (633nm) και αυτά στις αρχές του φάσματος. Αυτό σήμαινε ότι η παρουσία του κόκκινου θα ήταν ελλιπής και επομένως θα επηρεαζόταν συνολικά το οπτικό αποτέλεσμα της εικόνας. Όλα τα παραπάνω συνεπάγονταν την προσεκτική επιλογή των καναλιών για την εξισορρόπηση της αναπαράστασης. Τελικώς, χρησιμοποιήθηκαν τα μήκη κύματος – κανάλια εκείνα που ήταν αντιπροσωπευτικά για το RGB υπό την έννοια ότι βρίσκονταν στο κεντρικό μήκος κύματος από κάθε εύρος για την αναπαράσταση κάθε χρώματος και αυτά ήταν τα RGB: 15(621nm), 7(537nm), 1(469nm).

Ακολούθως, θα παρουσιαστούν προς σύγκριση οι MS εικόνες (πραγματικές με συνθετικές) προκειμένου να αξιολογηθεί η απόδοση του μοντέλου. Εδώ είναι σημαντικό να αναφερθεί ότι δεν ήταν εύκολο να διακριθούν αποκλίσεις μεταξύ των πολυφασματικών εικόνων γεγονός που επαληθεύει και την ενότητα της ποσοτικής αξιολόγησης. Παρόλα αυτά επιλέχθηκαν να εξεταστούν αντιπροσωπευτικά παραδείγματα για κάθε σετ δεδομένων που εμφάνισαν ενδιαφέρον και παρατίθενται στα Σχήματα 3.15, 3.16, 3.17 και 3.18. Συγκεκριμένα, για κάθε σετ δεδομένων παρατηρείται μία τριάδα εικόνων που αντιστοιχεί αριστερά στην έγχρωμη εικόνα BMP, όπως έχει αποδοθεί από την MS εικόνα, στη μέση είναι η πολυφασματική πραγματική εικόνα και δεξιά η αντίστοιχη MS συνθετική εικόνα (παράγωγο του MD²IP δικτύου) όπως προέκυψαν με το έγχρωμο σύνθετο που προαναφέρθηκε (RGB:15, 7, 1). Οι δύο τελευταίες εστιάζουν μέσω του κόκκινου πλαισίου σε μία περιοχή ενδιαφέροντος, που μεγεθύνεται επιπλέον στο κάτω άκρο της εικόνας, όπου διακρίνονται κάποιες διαφοροποιήσεις άξιες σχολιασμού.

Στο Σχήμα 3.15 παρατίθεται η εικόνα «balloons» που ανήκει στο σετ δεδομένων «Cave» η οποία εκ πρώτης όψεως δεν έχει καταφέρει να προσεγγίσει με ιδιαίτερη επιτυχία την έγχρωμη BMP εικόνας πέραν των μπλε αποχρώσεων παρά τις πολυάριθμες απόπειρες. Ωστόσο εκεί που θα εστιαστεί το ενδιαφέρον είναι σε τυχόν αποκλίσεις που έχουν η πραγματική με τη συνθετική MS εικόνα. Αυτή που γίνεται σαφής είναι στο παράθυρο που εστιάζεται στο κάτω δεξί τμήμα των εικόνων στην αναγραφή του «ColorChecker», δηλαδή του πίνακα που χρησιμοποιείται σε φωτογραφικές λήψεις και βίντεο για τη βαθμονόμηση και τον έλεγχο της ακρίβειας των χρωμάτων. Συγκεκριμένα, στα αρχικά δεδομένα η γραμματοσειρά είναι πιο διαυγής και

τα γράμματα πιο έντονα και ξεκάθαρα σε σχέση με την παραγόμενη εικόνα που τα γράμματα είναι πιο λεπτά και παρουσιάζουν περιμετρικά ένα πορτοκαλί περίγραμμα, τύπου «hallo». Παρόλα αυτά το κείμενο που αναγράφεται πάνω στον πίνακα βαθμονόμησης εξακολουθεί να είναι εμφανές και κατανοητό. Κατά τα άλλα δεν παρατηρήθηκε κάποια άλλη διαφοροποίηση.



Σχήμα 3.15: Διαφοροποιήσεις στη γραμματοσειρά (λέπτυνση γραμμάτων), εμφάνιση φαινομένου τύπου «hallo». Έγχρωμη εικόνα BMP (όπως έχει αποδοθεί από την MS εικόνα), MS πραγματική εικόνα όπως προέκυψε με το έγχρωμο σύνθετο που εστιάζει σε περιοχή ενδιαφέροντος, MS συνθετική εικόνα που εστιάζει σε περιοχή ενδιαφέροντος.

Έπεται το Σχήμα 3.16 που εξετάζει την εικόνα «Lilly Closeup» που ανήκει στο σετ δεδομένων «Manchester 2015». Αυτή τη φορά αν και το αντικείμενο αναφοράς (λουλούδι) έχει προσεγγιστεί ως προς τη χρωματική παλέτα, το υπόβαθρο δεν είχε αντίστοιχη επιτυχία στο έγχρωμο σύνθετο μιας και υπερίσχυσε το μπλε χρώμα αντί του καφέ (συνδυασμός των RGB) κάτι που αποδίδεται σε ένα βαθμό στην απουσία του κόκκινου χρώματος. Όσον αφορά τυχόν αποκλίσεις αυτές που έγιναν αντιληπτές ήταν στο κοτσάνι της Lilly και αφορούσαν ακμές που δεν ήταν τόσο «αιχμηρές» και συνολικά το στιγμιότυπο αυτό που εστιάστηκε ήταν σχετικά θολό με διακριτές δε εναλλαγές στο κέντρο του άνθους.

Manchester 2015 (M15), Lilly Closeup

RGB πραγματική εικόνα

Multispectral πραγματική εικόνα

Multispectral συνθετική εικόνα (MD2IP)



Σχήμα 3.16: Διαφοροποιήσεις σε ακμές.

Έγχρωμη εικόνα BMP (όπως έχει αποδοθεί από την MS εικόνα), MS πραγματική εικόνα όπως προέκυψε με το έγχρωμο σύνθετο που εστιάζει σε περιοχή ενδιαφέροντος, MS συνθετική εικόνα που εστιάζει σε περιοχή ενδιαφέροντος.

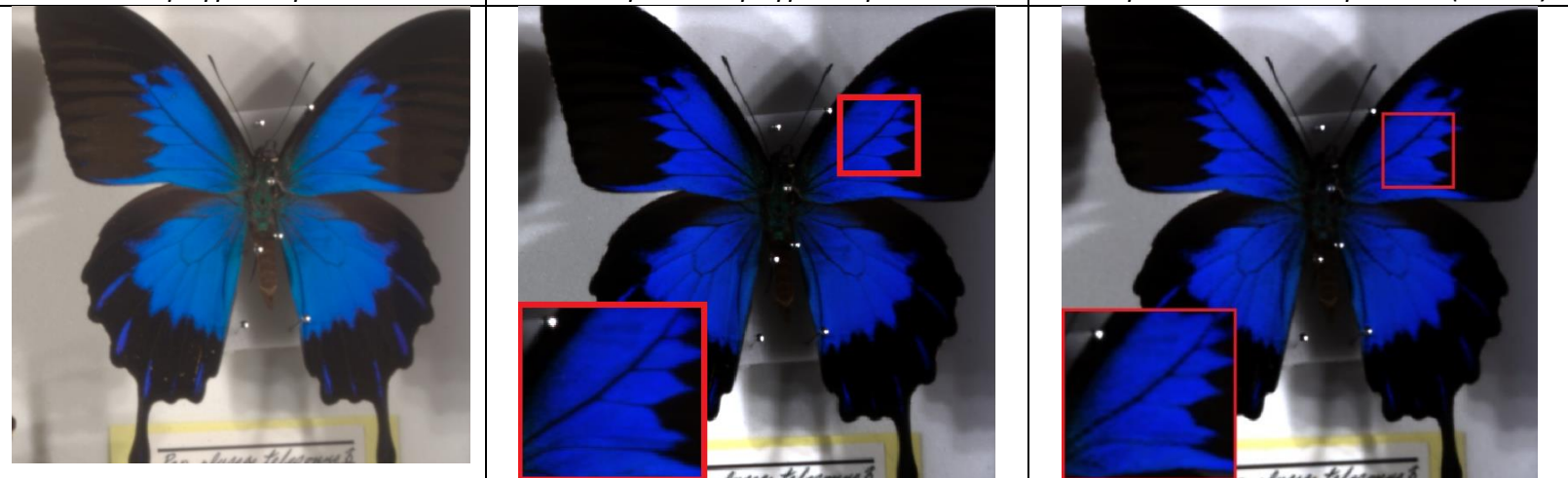
Στη συνέχεια, στο Σχήμα 3.17 παρουσιάζεται η εικόνα «*Butterfly3*» που ανήκει στο σετ δεδομένων «TokyoTech-31band». Εδώ είναι εμφανής η πλήρης συμφωνία κατά τη χρωματική απόδοση μεταξύ και των τριών εικόνων, κάτι που οφείλεται στην παρουσία βασικών χρωμάτων όπως το μπλε, μαύρο και κίτρινο (στο κάτω άκρο της εικόνας) που μπορούν να αποδοθούν ικανοποιητικά από το έγχρωμο σύνθετο. Στο συγκεκριμένο παράδειγμα οι αποκλίσεις που παρατηρούνται αφορούν και πάλι τις ακμές και πιο συγκεκριμένα τα γραμμικά μοτίβα που παρατηρούνται στην πραγματική εικόνα είναι λιγότερο διακριτά στη σύνθετη όπου έχει επικρατήσει μία πιο εξομαλυμένη απεικόνιση.

TokyoTech-31band (TT31), Butterfly3

RGB πραγματική εικόνα

Multispectral πραγματική εικόνα

Multispectral συνθετική εικόνα (MD²IP)



Σχήμα 3.17: Δυσκολία αναπαράστασης λεπτών γραμμών, ευρύτερη επικράτηση θολότητας-εξομάλυνσης. Φυσικό έγχρωμο σύνθετο RGB, MS πραγματική εικόνα όπως προέκυψε με το έγχρωμο σύνθετο που εστιάζει σε περιοχή ενδιαφέροντος, MS συνθετική εικόνα που εστιάζει σε περιοχή ενδιαφέροντος.

Ακολουθεί το Σχήμα 3.18 όπου οπτικοποιείται η εικόνα «Spray» που ανήκει στο σετ δεδομένων «TokyoTech-59band». Σε αυτή την περίπτωση οι αποδόσεις των έγχρωμων σύνθετων είναι υπέρ το δέον πιστές στην αρχική έγχρωμη RGB εικόνα. Οι αποκλίσεις που διαφοροποιούνται έρχονται να επαληθεύσουν τις παρατηρήσεις που σημειώθηκαν στα προηγούμενα παραδείγματα. Αναλυτικότερα, στην ετικέτα που εντοπίζεται στο μπουκάλι Spray υπάρχει πληθώρα πληροφορίας που αποτελείται από πυκνή γραμματοσειρά με αλφαριθμητικά στοιχεία που εξαρχής από την πραγματική εικόνα ήδη δεν είναι πλήρως κατανοητά. Ως επόμενο, η παραγόμενη πολυφασματική εικόνα δυσκολεύεται περισσότερο στην απόδοση αυτής της πληροφορίας λόγω του πυκνού νοήματος και των λεπτών γραμμών που το χαρακτηρίζουν. Σαν επακόλουθο, γίνεται λέπτυνση επιπλέον των γραμμών και επικρατεί ένα πορτοκαλί περίγραμμα γύρω από αυτές.



Σχήμα 3.18 Διαφοροποιήσεις στη γραμματοσειρά (λέπτυνση αλφαριθμητικού κειμένου), εμφάνιση φαινομένου τύπου «hallo». Φυσικό έγχρωμο σύνθετο RGB, MS πραγματική εικόνα όπως προέκυψε με το έγχρωμο σύνθετο που εστιάζει σε περιοχή ενδιαφέροντος, MS συνθετική εικόνα που εστιάζει σε περιοχή ενδιαφέροντος.

Συνοψίζοντας, λαμβάνοντας υπόψη όσα θίχθηκαν στην παρούσα ενότητα της ποιοτικής αξιολόγησης διαπιστώνεται ότι το training free μοντέλο του MD²IP που υλοποιήθηκε αποδίδει σε εξαιρετικό βαθμό στην εργασία της βελτιστοποίησης.

Συγκεκριμένα, έρχεται να επιβεβαιώσει την ενότητα της ποσοτικής αξιολόγησης όπου οι μετρικές σημείωσαν άριστες επιδόσεις με την SSIM να προσεγγίζει οριακά μονάδα και την PSNR μέχρι και 50. Επιπρόσθετα, τα αποτελέσματα των ανακατασκευασμένων πολυφασματικών εικόνων είναι άριστης ποιότητας με τα τεχνουργήματα (artifacts) που παρουσιάζονται να αποτελούν οριακά πταίσματα καθώς δεν επηρεάζουν αισθητά την ανάλυση της εικόνας ενώ πολλές λεπτομέρειες εξακολουθούν να είναι διακριτές. Κλείνοντας, το γεγονός αυτό επιδέχεται μεγαλύτερης αναγνώρισης ιδίως αν αναλογιστούμε ότι η συνθετική εικόνα έχει στηριχθεί αποκλειστικά στην MSFĀ -που αποτελεί υποδειγματοληψία της πλήρους MS εικόνας, έχοντας το 1/16 της συνολικής πληροφορίας για κάθε κανάλι- συνεπώς τα αποτελέσματα της παρούσας μεθόδου είναι πάρα πολύ καλά.

ΚΕΦΑΛΑΙΟ 4^ο – 2^η μεθοδολογία (U-Net)

4.1. Προεπεξεργασία δεδομένων

Για την ενότητα αυτή θα χρειαστεί να γίνουν ορισμένα ακόμα βήματα επεξεργασίας των εικόνων που παρέχονται από τα 4 σετ δεδομένων που αντλήσαμε. Στόχος είναι να κατασκευάσουμε για καθεμία εικόνα τμήματα αυτών (patches) με χωρική διάσταση 500x500 εικονοστοιχεία. Αυτή η διαδικασία θα υλοποιηθεί για τις πολυφασματικές εικόνες πλήρους ανάλυσης (κοινώς για τα 16 κανάλια που συνθέτουν την καθεμιά) και επιπλέον για τις MSFA εικόνες (αφορά το δίκτυο τύπου U-Net που απέδωσε κακές μετρικές και θα σχολιαστεί παρακάτω). Πιο συγκεκριμένα, η διαδικασία αυτή της κατάτμησης των εικόνων θα υλοποιηθεί τόσο για τις πραγματικές πολυφασματικές εικόνες όσο και για τις συνθετικές όπως έχουν προκύψει από το μοντέλο του MD²IP (που διατρήχθηκε και αναλύθηκε εκτενώς στο 3^ο Κεφάλαιο) και ομοίως για τις πραγματικές MSFĀ και συνθετικές MSFĀ εικόνες. Όλα τα παραπάνω υπό τμήματα των εικόνων θα αποτελέσουν δεδομένα εισόδου για το δίκτυο που θα εφαρμοστεί σε αυτό το Κεφάλαιο. Εδώ είναι σημαντικό να αναφερθεί ότι η κατάτμηση αυτή έγινε αφού είχε προηγηθεί κανονικοποίηση των τιμών ανακλαστικότητας στο διάστημα 0-1, διαιρώντας σε κάθε περίπτωση σετ δεδομένων ανάλογα τη ραδιομετρική του ανάλυση.

Η παραπάνω κατάτμηση των εικόνων είναι απαραίτητη διότι για να εκπαιδεύσεις ένα δίκτυο τα δεδομένα εισόδου πρέπει να είναι ίδιων διαστάσεων. Αυτό γίνεται είτε εσωτερικά στην κλάση όπου καλεί το dataset σε μορφή batches είτε σαν προεπεξεργασία, που είναι και αυτό που επιλέχθηκε. Όσον αφορά το μέγεθος των patches η επιλογή 500x500 προκύπτει αφενός από την ιδέα ότι δεν μπορεί να εισαχθούν εικόνες πολύ μεγάλες γιατί το δίκτυο θα γίνει «βαρύ» και αφετέρου ορισμένες εικόνες από τα σετ δεδομένων (TokyoTech-31band) είχαν ήδη αυτό το μέγεθος σαν αρχική διάσταση. Επιπλέον, παρότι το σύνηθες είναι δυνάμεις του 2 (πχ. 512x512) θα έπρεπε στις εικόνες διάστασης 500x500 να γίνει Padding, κάτι που θα δυσκόλευε τη διαδικασία επεξεργασίας.

Πιο αναλυτικά αυτό που υλοποιήθηκε είναι ότι αρχικές εικόνες που ήταν αρκετά μεγαλύτερες και πιο συγκεκριμένα γίνεται λόγος για τις εικόνες του dataset “Manchester 2015” κατατμήστηκαν σε περισσότερες με χαρακτηριστικό παράδειγμα αυτό του Σχήματος 4.1. Στην περίπτωση μάλιστα του «M15» κάθε εικόνα διάστασης

1024×1344 αντιστοιχίζεται σε τέσσερις 500×500 που αποτελούν τμήματα της πρώτης και η υπόλοιπη πληροφορία απορρίπτεται καθώς διαφορετικά θα έπρεπε να γίνει *upsampling* στο τμήμα της εικόνας 24×344 που θα είχε πολύ αραιή πληροφορία και θα περιέπλεκε την εκπαίδευση του μοντέλου μας σε επόμενο στάδιο. Αντίστοιχα στα *datasets* που η αρχική διάσταση 512×512 ήταν οριακά μεγαλύτερη όπως στα *Cave* και *TokyoTech-59band* περικόπηκαν οι πλεονάζουσες (12) γραμμές και στήλες ώστε να αποθηκευτούν τμήματα εικόνων που συμφωνούν με τη διάσταση των 500×500 εικονοστοιχείων. Συνεπώς, προκύπτει ένα νέο σύνολο δεδομένων για τις πραγματικές και συνθετικές πολυφασματικές εικόνες (*MS_GT* και *MS_MD²IP* αντίστοιχα) και τις αντίστοιχες *MSFA* εικόνες, όπου έκαστο αποτελείται από 214 *patches*.



Σχήμα 4.1: Πολυφασματική απεικόνιση της πραγματικής (GT) εικόνας “Gualtar Villa” από το σετ δεδομένων «M15» (πάνω) και 4 τμήματα αυτής διάστασης 500×500 εικονοστοιχεία (κάτω).

Αξίζει να επισημανθεί στο σημείο αυτό ότι τα δεδομένα που κατασκευάστηκαν σε τμήματα παραπάνω θα αξιοποιηθούν ανά ζεύγη για την εκπαίδευση δύο διαφορετικών μοντέλων. Με άλλα λόγια οι πραγματικές MSFA εικόνες θα αποτελούν δεδομένα εκπαίδευσης για το 1^ο μοντέλο και οι αντίστοιχες πραγματικές πολυφασματικές (MS_GT) θα είναι τα «label» του. Το μοντέλο αυτό μέσω μετρικών (training loss function αλλά και validation PSNR) θα μας υποδείξει πόσο καλά εκπαιδεύεται στο να εκτελεί την εργασία του demosaicing, έχοντας δεδομένα εκπαίδευσης (labels) υψηλής ποιότητας. Με παρόμοιο τρόπο το 2^ο μοντέλο θα έχει δεδομένα εκπαίδευσης τις συνθετικές εικόνες MSFA (όπως είχαν προκύψει από τα παράγωγα του MD²IP) και σαν «label» τις συνθετικές πολυφασματικές εικόνες πλήρους ανάλυσης (MS_MD²IP). Η απόδοση του 2^{ου} μοντέλου θα δείξει αν το μοντέλο μαθαίνει να λύνει το πρόβλημα του demosaicing εξίσου καλά με το 1^ο μοντέλο, που είναι και η επιδίωξη του παρόντος μεθοδολογικού πλαισίου.

Τα παραπάνω δεδομένα, δηλαδή τα τμήματα των πολυφασματικών εικόνων πλήρους ανάλυσης και οι MSFA εικόνες, είναι επαρκή δεδομένα εισόδου για την αρχιτεκτονική τύπου U-net που επιλέχθηκε με σκοπό να μιμηθεί τη διαδικασία demosaicing που επιτελεί το MD²IP, που δεν είναι άλλη από τη βελτιστοποίηση. Αυτό το μοντέλο όμως δεν φαίνεται να εκπαιδεύεται ικανοποιητικά -παρά τα εκτεταμένα πειράματα που έγιναν- σύμφωνα με την ποσοτική αξιολόγηση του μοντέλου, η οποία κατέστησε σαφές ότι το μοντέλο δεν εκπαιδεύεται επαρκώς. Το γεγονός αυτό οδήγησε στην αναζήτηση μίας καλύτερης αρχιτεκτονικής με πιο σύγχρονες προδιαγραφές που στηρίζεται σε attention και recurrent blocks. Σαν δεδομένα εκπαίδευσης για το συγκεκριμένο μοντέλο έγιναν διάφοροι πειραματισμοί με τα labels να παραμένουν προφανώς σταθερά. Αρχικά, κατασκευάστηκαν sparse πολυφασματικές εικόνες που προέκυψαν από τις MSFA, δηλαδή εικόνες με αραιή πληροφορία ανάλογη του μοτίβου της MSFA (4x4 στην προκειμένη περίπτωση, δηλαδή καταγράφεται πληροφορία ανά 4 εικονοστοιχεία) για καθένα από τα 16 κανάλια. Η διαδικασία αυτή διατρήθηκε για τα δεδομένα εκπαίδευσης που αφορούν και τα δύο μοντέλα που προαναφέρθηκαν, δηλαδή για τις πραγματικές και συνθετικές MSFA εικόνες.

Σε επόμενο στάδιο επιδιώκοντας ακόμα καλύτερα αποτελέσματα επί τις sparse εικόνες πραγματοποιήθηκε διγραμμική παρεμβολή (bilinear interpolation). Έτσι προέκυψαν πολυφασματικές παρεμβαλλόμενες εικόνες για τις οποίες θα υπάρχει 1 προς 1 αντιστοιχία με τις εικόνες που θα αποτελούν τα label, κάτι που θεωρήθηκε ότι θα συνδράμει στην καλύτερη εκπαίδευση του μοντέλου. Πράγματι, αυτό επαληθεύτηκε αλλά κάτι τέτοιο θα αναλυθεί εκτενέστερα στα επόμενα υποκεφάλαια. Σχετικά με τη διγραμμική παρεμβολή (bilinear interpolation) πρόκειται για μία μαθηματική μέθοδο

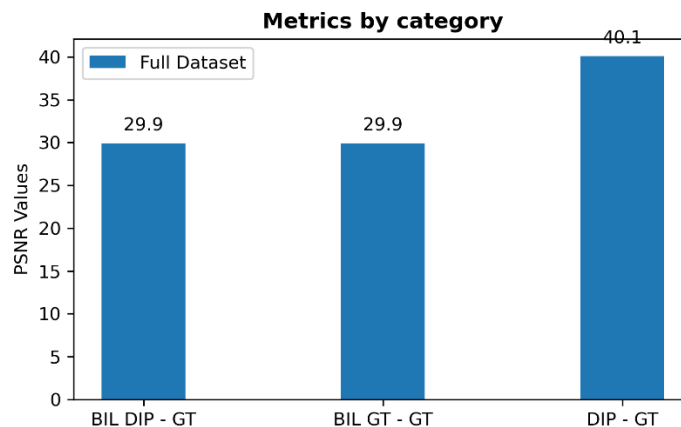
που χρησιμοποιείται για την εκτίμηση τιμών σε ένα διδιάστατο «πλέγμα». Στηρίζεται στη γραμμική παρεμβολή με τη διαφορά ότι εφαρμόζεται σε δύο διαστάσεις έναντι μίας, δηλαδή τόσο κατά τον άξονα x όσο και κατά τον y . Πρόκειται για μία τεχνική που εξισορροπεί την υπολογιστική απόδοση με την ακρίβεια. Για αυτό το λόγο έχει ευρύ εφαρμογή σε τομείς όπως η επεξεργασία εικόνων, τα Γεωγραφικά Συστήματα Πληροφοριών (ΓΣΠ) ή GIS για αναπροσαρμογή της κλίμακας των εικόνων, η εξομάλυνση δεδομένων, σε περιπτώσεις με ελλιπή σημεία δεδομένων για τη συμπλήρωση τους κ.α..

Συγκεκριμένα, στην επεξεργασία εικόνων η διγραμμική παρεμβολή πλεονεκτεί σε σχέση με άλλες απλούστερες μεθόδους όπως του εγγύτερου γείτονα ενώ λειτουργεί ως ένα είδος χαμηλοπερατού φίλτρου (Low-pass filter). Ορισμένα από τα οφέλη είναι η ομαλότητα καθώς έχει την ιδιότητα να αναπαράγει με μεγάλη ακρίβεια τις χαμηλές συχνότητες και να εξομαλύνει τις υψηλές συχνότητες, δηλαδή αναπαρίσταται η γενική δομή της εικόνας και τυχόν ακμές και λεπτομέρειες υφής μειώνονται. Ως απόρροια αυτού είναι τυχόν διαφορές στην ένταση να είναι πιο ήπιες και σταδιακές άρα προκύπτουν ομαλές χρωματικά περιοχές χωρίς λεπτομέρειες. Ακόμα, μειώνει την εμφάνιση τεχνουργημάτων και στρεβλώσεων, αποδίδει σε περιπτώσεις κλιμάκωσης και αποτελεί μία γρήγορη και εύκολη επιλογή σε σχέση με πιο σύνθετες μεθόδους όπως η τριγραμμική ή η κυματοειδής παρεμβολή που είναι πολύτιμο χαρακτηριστικό σε περιπτώσεις που απαιτείται μία άμεση και αποτελεσματική λύση. Αυτό την καθιστά χρήσιμη σε εφαρμογές πραγματικού χρόνου όπου η απόδοση είναι κρίσιμος παράγοντας. Συνοψίζοντας, η διγραμμική παρεμβολή συνιστά μία μέθοδο που παράγει αξιολογα αποτελέσματα με βελτιωμένη ποιότητα (καθώς έχει τη δυνατότητα να μειώσει την εμφάνιση θορύβου) και ομαλότητα.

Στην προκειμένη περίπτωση τα παράγωγα της θα αξιοποιηθούν σαν δεδομένα εκπαίδευσης για τα δύο μοντέλα που θα διατρεχθούν, στηριζόμενοι στην ιδέα ότι το δίκτυο θα μπορέσει να κατανοήσει και να αντιστοιχίσει καλύτερα τα δεδομένα εκπαίδευσης με τα labels και με τη σειρά του να εκπαιδευτεί πιο αποδοτικά εξάγοντας πολυφασματικές εικόνες που θα προσεγγίζουν τα δεδομένα αληθείας. Ιδιαίτερα σημαντικό είναι να ειπωθεί ότι η παρεμβολή αναπαράγει πολύ καλά τις χαμηλές συχνότητες, όπως προαναφέρθηκε, επομένως τα παράγωγα που αφορούν τόσο τις πραγματικές παρεμβαλλόμενες πολυφασματικές εικόνες (BIL_GT) όσο και τις αντίστοιχες συνθετικές (BIL_MD²IP) αναμένεται να μην έχουν ουσιαστικές διαφορές μεταξύ τους. Αυτό οφείλεται επιπλέον στο γεγονός ότι το training free μοντέλο που στηρίζεται στο MD²IP που εφαρμόστηκε στο 3^ο Κεφάλαιο χαρακτηρίζεται από την καλή του επίδοση στην αναπαράσταση υψηλών συχνοτήτων και την εξαιρετική στις χαμηλές συχνότητες. Συνεπώς, και στις δύο περιπτώσεις από τη διαδικασία της παρεμβολής θα

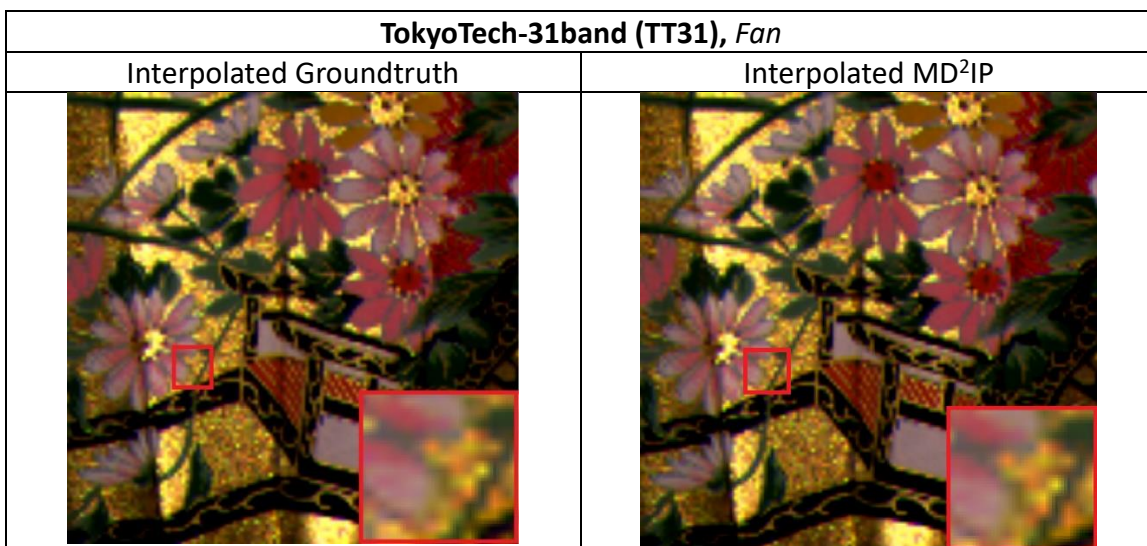
συντηρηθούν οι χαμηλές συχνότητες ενώ οι υψηλές θα μειωθούν σημαντικά άρα προσδοκάται ότι οι «Interpolated» δεν θα παρουσιάζουν οπτικές διαφορές μεταξύ τους παρά μόνο σε επίπεδο εικονοστοιχείων που θα είναι ελάχιστες.

Η παραπάνω σκέψη επαληθεύεται από το διάγραμμα του Σχήματος 4.2 που ακολουθεί, το οποίο απαρτίζεται από τρεις στήλες οι οποίες καταγράφουν τη μετρική PSNR που αφορά το σύνολο δεδομένων (απαρτίζεται και από τα 4 dataset, δηλαδή αντιστοιχεί στα 214 patches) σε τρία διαφορετικά στάδια επεξεργασίας. Ειδικότερα, η πρώτη μπάρα αφορά τη μετρική PSNR που σημείωσαν οι συνθετικές παρεμβαλλόμενες εικόνες – παράγωγα του MD²IP- (BIL MD²IP) σε σχέση με τις πραγματικές πολυφασματικές (GT), η δεύτερη μπάρα αναπαριστά τον βαθμό ομοιότητας μεταξύ των πραγματικών παρεμβαλλόμενων εικόνων (BIL GT) με τις πραγματικές πολυφασματικές (GT) και η τρίτη την ομοιότητα μεταξύ των παράγωγων συνθετικών πολυφασματικών εικόνων (MD²IP) και των πραγματικών (GT). Πράγματι, λοιπόν, βάση του παρακάτω διαγράμματος γίνεται εμφανές βάση της μετρικής PSNR, που αποτελεί δείκτη ομοιότητας μεταξύ δύο εικόνων, ότι οι παρεμβαλλόμενες εικόνες είτε έχουν προκύψει από το μοντέλο του MD²IP είτε από τις πραγματικές και έχοντας κατά νου ότι η διγραμμική παρεμβολή αποτελεί ένα χαμηλοπερατό φίλτρο, δεν παρουσιάζουν καμία διαφορά από τις πραγματικές αφού ισχύει PSNR=29,9 και για τις δύο περιπτώσεις. Τέλος, η τρίτη στήλη σημειώνει πολύ μεγαλύτερη τιμή PSNR=40,4 από τις προηγούμενες κάτι που αποδίδεται στην ιδιότητα των παράγωγων συνθετικών εικόνων (MS_MD²IP) να διατηρούν μέρος των υψηλών συχνοτήτων (εν αντιθέσει με τη παρεμβολή που αποκαθιστά μόνο τις χαμηλές) και συνεπώς να προσεγγίζει περισσότερο τις πραγματικές.

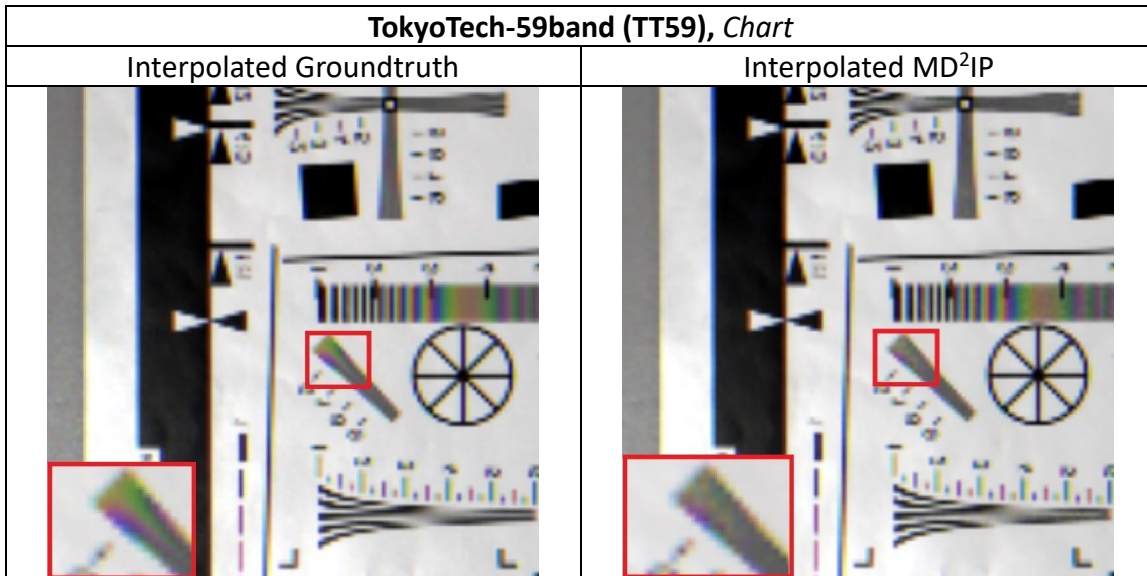


Σχήμα 4.2: Διάγραμμα που αναπαριστά την ομοιότητα μεταξύ 1) Bil. MD²IP-GT, 2) Bil. GT-GT & 3) MD²IP-GT μέσω της μετρικής PSNR.

Για να γίνει επιπλέον σαφές όσα αναφέρθηκαν επιλέγεται στα Σχήματα 4.3 και 4.4 που ακολουθούν να παρουσιαστούν ενδεικτικά κάποια ζευγάρια εικόνων που προέκυψαν από διγραμμική παρεμβολή. Σε αυτά αριστερά παρατίθεται η πραγματική πολυφασματική παρεμβαλλόμενη εικόνα που έχει προκύψει από τα groundtruth και δεξιά η αντίστοιχη παράγωγη του MD²IP. Γενικά, όπως παρατηρείται τα ζεύγη εικόνων δεν παρουσιάζουν διαφορές συνολικά με γυμνό μάτι παρά μόνο σε επίπεδο εικονοστοιχείου και πιο συγκεκριμένα όταν γίνει μεγάλη εστίαση σε κάποιο παράθυρο που επισημαίνεται με κόκκινο πλαίσιο, όπου υπάρχουν μικρές αποκλίσεις στις ψηφιακές τιμές. Μάλιστα, αυτό που εντοπίστηκε και έχει ουσία να επισημανθεί είναι οι εξομαλυμένες ακμές και μικρές διαφοροποιήσεις στην υφή που δίνουν την αίσθηση μίας πιο ομαλής εικόνας αλλά μόνο εφόσον εστιάσει ο παρατηρητής αρκετά. Κάτι τέτοιο βέβαια προσδοκόταν σύμφωνα με τη δομή του δικτύου MD²IP, που αναπαράγει σε πολύ καλό βαθμό τις χαμηλές συχνότητες και εξομαλύνει τις αιχμηρές ακμές. Στο Σχήμα 4.3 παρά την εστίαση και την εκτεταμένη προσπάθεια εύρεσης κάποιας διαφοροποίησης αυτό κατέστη αδύνατο, κάτι που ίσχυε για την πλειονότητα των συνόλων δεδομένων. Ωστόσο, το Σχήμα 4.4 μπορεί να θεωρηθεί «outlier» που αναζητήθηκε επαρκώς, μιας και βρέθηκαν ορισμένα σημεία με εμφανείς διαφοροποιήσεις με πιο χαρακτηριστική αυτή που επισημαίνεται στο περιγραφόμενο πλαίσιο. Στο παράθυρο θέασης που εστιάζει το Σχήμα επισημαίνεται κυρίως χρωματική απόκλιση μεταξύ των παραγωγών παρεμβαλλόμενων εικόνων που πιθανώς να οφείλεται στην πυκνή γραμμική πληροφορία που είχε εξασθενήσει ήδη από τη μεθοδολογία του MD²IP.



Σχήμα 4.3: Ζευγάρι παρεμβαλλόμενων πολυφασματικών εικόνων (GT-MD²IP) όπως προέκυψαν από τη διαδικασία της διγραμμικής παρεμβολής στην εικόνα «Fan» από το σύνολο δεδομένων «TT31».



Σχήμα 4.4: Ζευγάρι παρεμβαλλόμενων πολυφασματικών εικόνων (GT-MD²IP) όπως προέκυψαν από τη διαδικασία της διγραμμικής παρεμβολής στην εικόνα «Chart» από το σύνολο δεδομένων «TT59».

4.2. Υλοποίηση Μοντέλου τύπου U-Net

Για την διαδικασία της εκπαίδευσης επιλέχθηκε να υλοποιηθεί ένα συνελκτικό δίκτυο τύπου U-Net για να εκμεταλλευτούμε την εξαγωγή καλών χαρακτηριστικών από τον encoder αλλά και τις δυνατότητες ανακατασκευής της εικόνας του decoder. Εκτός αυτού η δυνατότητα «μεταφοράς» χαρακτηριστικών διαφορετικών κλιμάκων μέσα από τα skip connections (διασυνδέσεις τύπου concatenate ανάμεσα στον encoder και τον decoder) κρίθηκε ότι θα βοηθήσει το μοντέλο στην κατανόηση του MSFA pattern. Για την υλοποίηση του συγκεκριμένου δικτύου έπρεπε να γίνουν κάποιες βασικές σχεδιαστικές επιλογές. Δηλαδή, το πόσα upsample & downsample θα υλοποιηθούν αλλά και πόσες διασυνδέσεις θα έχουν. Εκτός αυτού κάθε επίπεδο που καταλήγει σε μέθοδο pooling operation (downsample) ή transposed convolution layer (upsample) περιλαμβάνει και απλά συνελκτικά επίπεδα σε συνδυασμό με batch normalization και activation function που έπεται στη συνέχεια.

Σε αυτό το σημείο κρίνεται σκόπιμο να θιχθούν κάποια ενδιαφέροντα μεθοδολογικά στοιχεία. Αρχικά σχετικά με τις δυνατότητες που υπάρχουν για αναδειγματοληψία ή κοινώς upsampling προσφέρονται μέθοδοι που διακρίνονται σε δύο κατηγορίες τις παραμετρικές και τις μη παραμετρικές. Στις μη παραμετρικές ενδεικτικά υπάρχουν οι υπολογιστικές όπως είναι η διγραμμική και η τριγραμμική ενώ στις μη υπολογιστικές συγκαταλέγεται η μέθοδος του εγγύτερου γείτονα. Στις

παραμετρικές συνήθως χρησιμοποιείται η *transposed convolution*, αλλά υπάρχουν και άλλες πιο σύνθετες επιλογές όπως η *Content-Aware ReAssembly of FEatures (CARAFE)*.

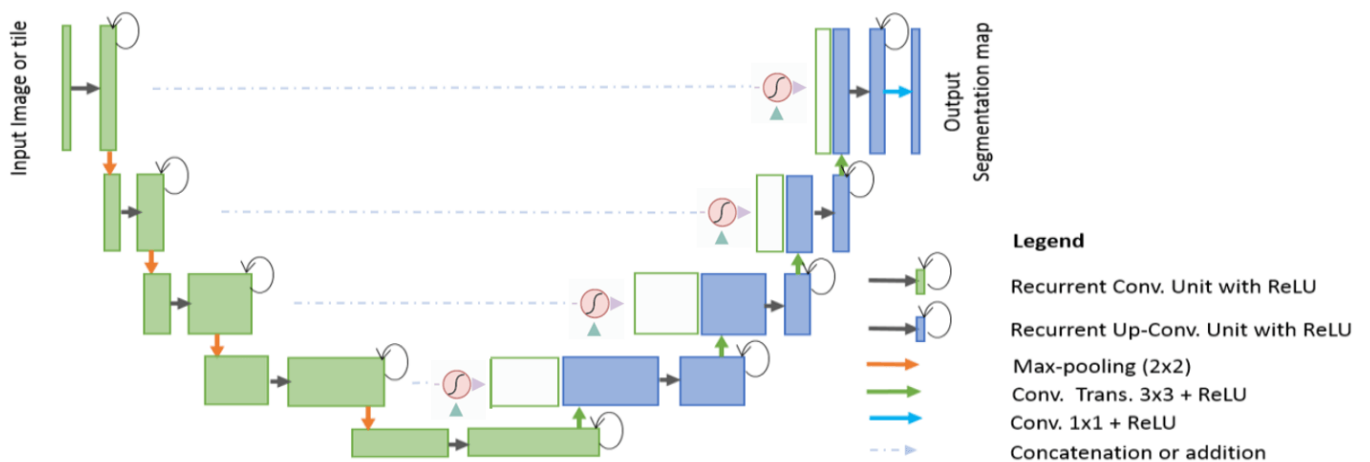
Το πλεονέκτημα τέτοιων μεθόδων είναι ότι η διαδικασία του *upsampling* γίνεται πιο αποτελεσματική μιας και το μοντέλο «μαθαίνει» τις βέλτιστες τιμές για τα *kernel* της συνέλιξης ώστε να επιτύχει την βέλτιστη αναδειγματοληψία με στόχο την αύξηση των χωρικών διαστάσεων της εικόνας. Αντίστοιχα, το μεγαλύτερο μειονέκτημα τέτοιων μεθόδων είναι ότι αυξάνεται ο αριθμός των παραμέτρων του μοντέλου. Αυτό έχει ως αποτέλεσμα η εκπαίδευση του να καθίσταται πιο δύσκολη αλλά και σε ορισμένες περιπτώσεις να δυσχεραίνεται η διαδικασία του *back propagation* μιας και σε πολύ βαθιά δίκτυα οι μερικές παράγωγοι -όταν υπάρχουν πολλά συνεχόμενα *operations*- είναι πιθανό να μηδενιστούν. Παρόλα αυτά στην δική μας περίπτωση όπου το τελικό μοντέλο θα είναι αρκετά μικρό τέτοια ζητήματα δεν μας προβληματίζουν. Κατά συνέπεια μπορούμε να χρησιμοποιήσουμε μια τέτοια μέθοδο και πιο συγκεκριμένα επιλέγονται τα *transposed convolution* μιας και αποτελούν τη συνήθη ιδανική επιλογή στα περισσότερα μοντέλα.

Όσον αφορά το *downsampling* υπάρχουν αντίστοιχα επιλογές με τις πιο συνήθεις να είναι το *operation pooling* (μέσο, μέγιστο και καθολικό) αλλά και συνελκτικές προσεγγίσεις. Στην προκειμένη περίπτωση επιλέχθηκε να μην χρησιμοποιηθεί το *average pooling* καθώς έχει την τάση να «θολώνει» τα παράγωγα, γεγονός που θα οδηγούσε στην απόκτηση μερικών σχετικά καλών αποτελεσμάτων με αρκετό θόρυβο στον αντίποδα ενώ το μοντέλο θα δυσκολευόταν στην ανάκτηση των υψηλών συχνοτήτων. Αντίστοιχα η μέθοδος του *max pooling* απορρίφθηκε διότι «διατηρώντας» αποκλειστικά την μέγιστη τιμή από κάθε *kernel* χάνεται σημαντική πληροφορία που ειδικά σε εφαρμογές που αφορούν *demosaiicing* είναι σημαντικό να διατηρηθεί. Έτσι, η καλύτερη επιλογή φαίνεται να είναι μια συνέλιξη που θα μεταβάλλει τις διαστάσεις του *activation map* χρησιμοποιώντας την παράμετρο *stride*, η οποία καθορίζει το «βήμα» κάθε συνελκτικού φίλτρου.

Όσον αφορά τα επίπεδα (*layers*) που περιλαμβάνουν τις «πράξεις» που γίνονται πριν από κάθε δειγματοληψία (*downsampling, upsampling*) η βέλτιστη επιλογή ήταν μάλλον και η πιο συνήθης με μια πράξη συνέλιξης (3×3) ακολουθούμενη από το *batch normalization* και μια *activation function*. Αυτή η τριάδα επαναλήφθηκε δύο φορές εντός κάθε επιπέδου. Ακόμα, σαν συνάρτηση ενεργοποίησης χρησιμοποιήθηκε η *ReLU*-ενώ δοκιμάστηκαν επιπλέον η *leaky ReLU* αλλά και η *SELU*- καθότι κατά την διαδικασία των δοκιμών του μοντέλου απέδιδε καλύτερα από τις υπόλοιπες εναλλακτικές. Συνοψίζοντας, τα μοντέλα με τα χαρακτηριστικά που παρουσιάστηκαν παραπάνω δοκιμάστηκαν για την επίλυση του προβλήματος *demosaiicing* χωρίς ιδιαίτερη επιτυχία,

δηλαδή χωρίς να παρουσιάζουν καλύτερες μετρικές από αυτές που αναμένονταν από το sanity check. Ως sanity check λογίζεται η επίτευξη μετρικών PSNR και SSIM καλύτερες από αυτές που επιτυγχάνονται με απλές μεθόδους interpolation (bilinear & trilinear) για αυτό προέκυψε η ανάγκη ενός πιο «σύνθετου» δικτύου.

Σε αυτή την κατεύθυνση, κρίθηκε μιας και γενικότερα τα δίκτυα τύπου U-Net παρουσιάζουν πολύ καλά αποτελέσματα σε «image to image tasks» να αναζητηθεί ένα δίκτυο που θα διατηρεί την πλειονότητα των χαρακτηριστικών (από αρχιτεκτονικής άποψης) που αναφέρθηκαν προσθέτοντας επιπλέον κάποιους μηχανισμούς «ενίσχυσης». Αναλυτικότερα, οι μηχανισμοί που επιλέχθηκαν ήταν τα: Attention αλλά και recurrent layers με residual connections. Η γενική μορφή του δικτύου φαίνεται στο Σχήμα 4.5 που ακολουθεί. Αυτές οι πρακτικές παρουσιάστηκαν στις δημοσιεύσεις [14], [15], [16] και η βασική υλοποίηση τους υπάρχει στο αποθετήριο [17]. Επειδή οι συγκεκριμένοι μηχανισμοί αποτελούν προσθήκες που δεν παρουσιάζονται στα ίδια τμήματα του δικτύου, δηλαδή ο μηχανισμός attention εφαρμόζεται στα skip connections ενώ τα recurrent layers εντοπίζονται κατά την “κατάβαση” και “ανάβαση” του δικτύου, θα παρουσιαστούν ξεχωριστά.

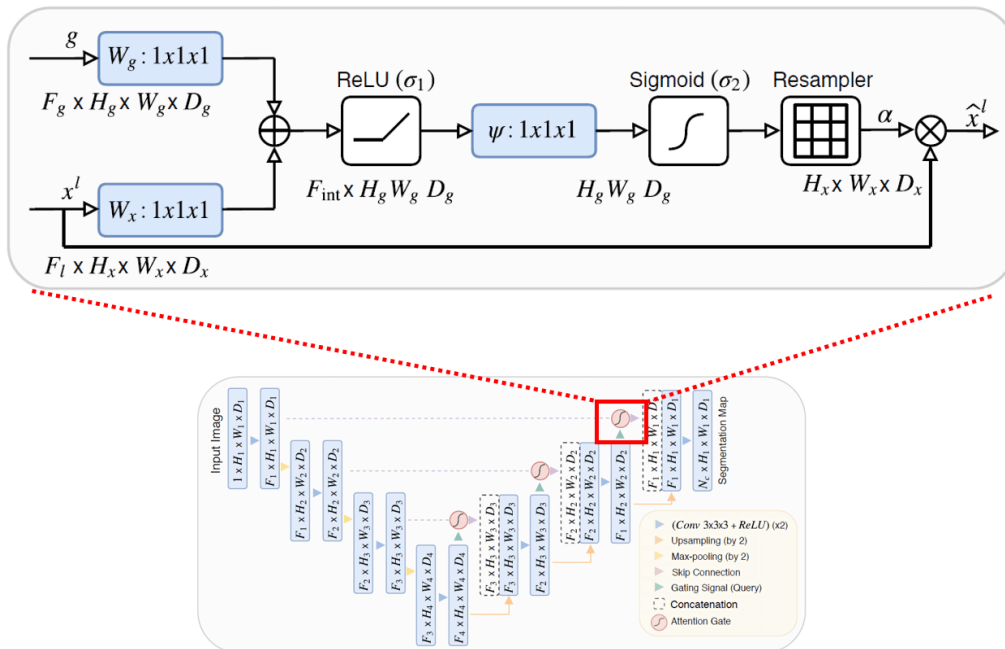


Σχήμα 4.5: Δομή Attr2U-Net δικτύου που χρησιμοποιήθηκε για την υλοποίηση του μοντέλου μας. Πηγή: [17].

Πριν γίνει ανάλυση των μηχανισμών που εφαρμόστηκαν στο δίκτυο μας κρίνεται απαραίτητο να τονιστεί ότι εφόσον το αρχικό δίκτυο τύπου U-Net ήταν μικρότερο και πιο γρήγορο υπήρχε περιθώριο να επιλεγθούν τα convolution layers σαν μέθοδος downsampling αφού επέφεραν καλύτερη επίλυση. Ωστόσο στη νέα αρχιτεκτονική που χρησιμοποιήθηκε με τα recurrent layers -που είναι πιο αργή- ως μέθοδος υποδειγματοληψίας (downsampling) προτιμήθηκε το max pooling. Η επιλογή αυτή ήταν

σκόπιμη διότι αποτελούσε μία σαφέστερα πιο γρήγορη διαδικασία και εν τέλει δεν παρατηρήθηκαν μεγάλες διαφορές. Επίσης, επιδιώχθηκε ο αριθμός των παραμέτρων του μοντέλου να διατηρηθεί το δυνατόν μικρότερος, κάτι που συνεπάγεται μικρότερη υπολογιστική ισχύ, γεγονός που δεν ήταν εφικτό με τη χρήση συνελιξων καθώς εισάγουν πολλές νέες παραμέτρους εν αντιθέσει με τα pooling operations. Όσον αφορά τα υπόλοιπα χαρακτηριστικά που αφορούν τη δομή της αρχιτεκτονικής του δικτύου διατηρήθηκαν ως είχαν.

Σχετικά με τη χρήση μηχανισμών Attention (attention gate) στα δίκτυα τύπου U-Net αποτελεί μια νέα τάση που επαληθεύεται από πληθώρα ερευνητικών εργασιών [16], [18], [19], [20]. Η βασική ιδέα της συγκεκριμένης μεθόδου φαίνεται στο Σχήμα 4.6, όπου χρησιμοποιούνται συντελεστές βάρους οι οποίοι βασίζονται τόσο στις τιμές του layer που “έρχεται” από το τμήμα του encoder (skip connection) όσο και σε αυτές που “έρχονται” από τον decoder (upsample). Οι συντελεστές αυτοί (κυμαίνονται στο εύρος 0 έως 1) πολλαπλασιάζουν τα activation maps που προκύπτουν από τα skip connection και εξαρτώνται από βάρη τα οποία μαθαίνονται (προσεγγίζονται) κατά την διάρκεια της εκπαίδευσης. Συνεπώς γίνεται σαφές ότι η συγκεκριμένη μεθοδολογία δεν ακολουθεί την τυπική τακτική αθροίσματος (concatenate) των activation maps μεταξύ encoder και decoder layer.



Σχήμα 4.6: Σχηματική αναπαράσταση της προτεινόμενης attention gate (AG) μεθοδολογίας που φιλτράρει τα χαρακτηριστικά που διαδίδονται μέσω των skip connections και σύνδεση της με το διάγραμμα των blocks του προτεινόμενου μοντέλου κατάτμησης Attention U-Net, κατά την ερευνητική μελέτη [16].

Ακολουθως, σχετικά με τα recurrent επίπεδα ακολουθούν μια πολύ απλή λογική όπου αντί να υπάρχουν στην σειρά πράξεις συνέλιξης με διαφορετικά βάρη οι πράξεις αυτές υλοποιούνται με τα ίδια βάρη για έναν προκαθορισμένο αριθμό επαναλήψεων (κάτι που συνιστά υπερπαραμέτρο του δικτύου). Με την συγκεκριμένη μεθοδολογία ειδικά σε συνδυασμό με τον μηχανισμό Attention ο στόχος των ερευνητών που τη κατασκεύασαν αλλά και ο προσωπικός είναι η αύξηση του receptive field του μοντέλου. Αυτό πρακτικά σημαίνει ότι το μοντέλο θα μπορεί να «δει» μεγαλύτερο τμήμα της εικόνας. Αυτή η μεθοδολογία αποτελεί μία αρκετά έξυπνη επίλυση που ειδικά για εφαρμογές που παρουσιάζουν κάποιο επαναλαμβανόμενο μοτίβο, όπως στη περίπτωση μας στην εργασία του demosaicing, είναι εύστοχη. Κλείνοντας, το μεγάλο πλεονέκτημα αυτής της μεθόδου είναι ότι δεν αυξάνει τις παραμέτρους του δικτύου ενώ ταυτόχρονα μεγαλώνει το receptive field, κάτι το οποίο επιτυγχάνεται τυπικά με περισσότερα φίλτρα συνέλιξης.

Κλείνοντας, αναφορικά με τα residual χρησιμοποιούν την κλασική πλέον τακτική που εφαρμόστηκε αρχικά στις αρχιτεκτονικές τύπου ResNet [21] ώστε να δώσουν μια νέα δίοδο στις παραγώγους. Σε αυτήν τη μεθοδολογία οι συνέλιξεις εφαρμόζονται όχι στον βασικό άξονα του δικτύου αλλά παράλληλα στα υπολειπόμενα τμήματα (residuals). Η τεχνική αυτή βασίζεται στον προσδιορισμό των μεταβολών που πρέπει να εφαρμοστούν στο σήμα σε κάθε Layer προσθετικά χωρίς όμως να γίνεται πλήρης μετασχηματισμός του σήματος σε κάθε Layer.

4.3. Πειράματα

4.3.1. Δομή Πειραμάτων

Στην ενότητα αυτή αναλύονται τα πειράματα που υλοποιήθηκαν χρησιμοποιώντας το μοντέλο τύπου U-Net που αναλύθηκε εκτενώς στην ενότητα 4.2.. Τα πειράματα αυτά διεξάγονται για τρία μοντέλα με την ειδοποιός διαφορά να έγκειται στα δεδομένα εκπαίδευσης -έναντι των δύο που είχαν εφαρμοστεί αρχικά στην περίπτωση του πιο απλού μοντέλου τύπου U-Net που σημείωσε μη ικανοποιητικά αποτελέσματα-. Ειδικότερα, στο 1^ο μοντέλο ως δεδομένα εκπαίδευσης εκχωρούνται οι παρεμβαλλόμενες συνθετικές πολυφασματικές εικόνες (Interpolated MD²IP) με δεδομένα αληθείας (Labels), δηλαδή τιμές που πρέπει να προβλέψει το μοντέλο και αναπαριστούν την αναμενόμενη έξοδο για δεδομένη είσοδο, τις αντίστοιχες συνθετικές (MS MD²IP). Οι ίδιες ετικέτες (Labels) χρησιμοποιούνται και από το 3^ο μοντέλο με

δεδομένα εισόδου τις παρεμβαλλόμενες πραγματικές (Interpolated GT). Τα μοντέλα αυτά διατρέχονται για να παρατηρηθεί πόσο ευαίσθητο είναι το μοντέλο τύπου U-Net στην αλλαγή δεδομένων εισόδου και επιπλέον για να εξεταστεί η επίδοσή του σε εικόνες που θεωρήθηκε αρχικά ότι θα συγκρατούν πληροφορία καλύτερης ανάλυσης (Interpolated GT) έχοντας διατηρήσει μέρος των υψηλών συχνοτήτων, κάτι που βέβαια όπως αποδείχθηκε στην ενότητα της προεπεξεργασίας δεν συνέβη. Τα πειράματα αυτά πραγματοποιήθηκαν πριν υπολογιστούν οι μετρικές οπότε εξετάστηκαν επιπλέον διότι στην προκειμένη περίπτωση τα παρεμβαλλόμενα πραγματικά δεδομένα σαν δεδομένα εισόδου αναμενόταν να αποδώσουν μερικώς καλύτερα αποτελέσματα, κάτι που θα φανεί στην παράγραφο της Αξιολόγησης. Το αποτέλεσμα που θα εξαχθεί από τις παρούσες δοκιμές δεν αποτελεί πανάκια για κάθε εφαρμογή και θα έχει ενδιαφέρον να μελετηθεί αφού λογικά τα δεδομένα που έχουν προκύψει από πραγματικά εικονοστοιχεία (Groundtruth) προσδοκάται να έχουν καλύτερη επίδοση από ότι τα συνθετικά (παράγωγα του MD²IP). Τέλος, το 2^ο και τελευταίο μοντέλο λαμβάνει δεδομένα εισόδου τις παρεμβαλλόμενες πραγματικές και Labels τις πραγματικές (MS GT). Το συγκεκριμένο μοντέλο υλοποιείται για να παρατηρηθεί η απόδοση του δικτύου σε πραγματικά δεδομένα και να συγκριθεί στη συνέχεια με αυτή των προηγούμενων μοντέλων. Πιο συγκεκριμένα, διερευνάται κατά πόσο τα παράγωγα του MD²IP μπορούν να λειτουργήσουν εξίσου καλά ως δεδομένα αληθείας σε σχέση με τα πραγματικά δεδομένα. Για να γίνει περισσότερο κατανοητή η παραπάνω κατανομή των δεδομένων εκπαίδευσης ανά τα μοντέλα κατασκευάστηκε ο Πίνακας 4.1.

	Δεδομένα Εκπαίδευσης Μοντέλων	
	<u>Δεδομένα Εισόδου</u>	<u>Δεδομένα αληθείας</u>
1^ο Μοντέλο	MSF \bar{A} (MD ² IP)	MS MD ² IP
2^ο Μοντέλο	MSF \hat{A} (GT)	MS GT
3^ο Μοντέλο	MSF \hat{A} (GT)	MS MD ² IP

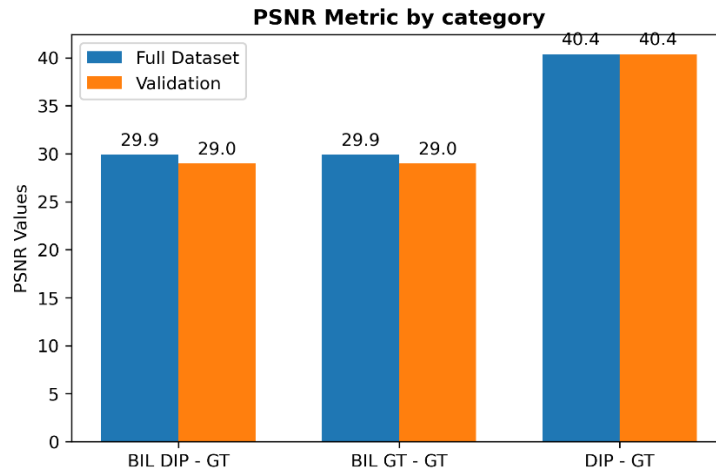
Πίνακας 4.1.: Τα δεδομένα εκπαίδευσης που χρησιμοποιούνται για καθένα από τα 3 μοντέλα που θα υλοποιηθούν.

4.3.2. Αντιπροσωπευτικότητα συνόλου Αξιολόγησης (Validation set)

Αρχικά, είναι σημαντικό να αναφερθεί ότι τα δεδομένα που συλλέχθηκαν και προεπεξεργάστηκαν διακρίθηκαν σε σετ εκπαίδευσης και σετ αξιολόγησης ενώ σετ ελέγχου δεν δημιουργήθηκε λόγω του συνόλου δεδομένων που ήταν περιορισμένο. Λόγω του μικρού όγκου δεδομένων αποφασίστηκε το σύνολο αξιολόγησης να συνιστά

περίπου το 10% του συνόλου και πιο συγκεκριμένα αποτελείται από 20 εικόνες τυχαία επιλεγμένες από το δίκτυο. Εδώ, κρίνεται απαραίτητο να γίνει μία ανάλυση για το κατά πόσο αντιπροσωπευτικό ήταν το δείγμα των δεδομένων που συνέστησαν το validation set, γεγονός που θα υποδείξει και το κατά πόσο αντικειμενικά είναι τα αποτελέσματα του μοντέλου μας. Για το σκοπό αυτό κατασκευάστηκε το γράφημα του Σχήματος 4.7. όπου αντιπαραβάλλονται οι μετρικές που σημείωσαν το σύνολο δεδομένων (απαρτίζεται από 214 εικόνες) και το υποσύνολο του σετ αξιολόγησης (20 εικόνες) σε διαφορετικές περιπτώσεις προεπεξεργασίας. Το διάγραμμα αυτό συμπληρώνει το Σχήμα 4.2 που αναλύθηκε στο προηγούμενο υποκεφάλαιο, οπότε θα παρακαμφθεί η εκτενής εξήγηση του.

Έτσι, λοιπόν, από το γράφημα που ακολουθεί γίνεται σαφές ότι το σύνολο αξιολόγησης όχι μόνο είναι αντιπροσωπευτικό του συνόλου αλλά και πιο προκλητικό-δύσκολο για την απόδοση του μοντέλου αφού φαίνεται η διαφορά ψηφιακών τιμών μεταξύ πραγματικών πολυφασματικών με τις παρεμβαλλόμενες που καταγράφει η μετρική PSNR να είναι μερικώς μικρότερη. Πιο λεπτομερώς για την κατηγορία όπου η PSNR υπολογίζει τη διαφορά μεταξύ παρεμβαλλόμενων συνθετικών εικόνων με τις πραγματικές (BIL MD²IP – GT) το σύνολο δεδομένων σημειώνει τιμή ίση με PSNR (Full Dataset) =29,9 ενώ το σύνολο δεδομένων αξιολόγησης μόλις PSNR (Validation set) =29,0 και αντίστοιχα στη κατηγορία σύγκρισης των παρεμβαλλόμενων πραγματικών με τις πραγματικές (BIL GT – GT) τα σύνολα δεδομένων σημειώνουν ίδιες τιμές με την προηγούμενη κατηγορία. Αυτό σημαίνει ότι για το σύνολο αξιολόγησης έχουν επιλεχθεί εικόνες που αποκλίνουν περισσότερο από τις πραγματικές άρα θα είναι πιο δύσκολο να μιμηθούν από το δίκτυο τύπου U-Net που θα χρησιμοποιηθεί. Τέλος, όσον αφορά την σύγκριση των συνθετικών πολυφασματικών εικόνων με τις πραγματικές παρατηρείται σύμπνοια στη μετρική, αφού PSNR=40,4 και στις δύο περιπτώσεις, κάτι που βέβαια επηρεάζει τον βαθμό εκμάθησης άρα και την απόδοση του μοντέλου κατά το στάδιο εκπαίδευσης αφού οι συνθετικές εικόνες συνιστούν τα δεδομένα αληθείας. Παράλληλα η validation PSNR αποτελεί δείκτη του κατά πόσο το μοντέλο μπορεί να προσεγγίσει την απόδοση των συνθετικών εικόνων (παραγώγων της μεθοδολογίας MD²IP) που είναι και ο στόχος της παρούσας μεθοδολογίας.



Σχήμα 4.7: Αντιπαραβολή απόδοσης μεταξύ του συνόλου δεδομένων και του τυχαίου συνόλου αξιολόγησης, βάση της μετρικής PSNR σε τρεις κατηγορίες 1) Bil. MD²IP-GT, 2) Bil. GT-GT & 3) MD²IP-GT.

4.3.3. Υπερπαράμετροι Μοντέλων

Στα πλαίσια των πειραμάτων πραγματοποιήθηκαν πολλές δοκιμές προς αναζήτηση των βέλτιστων υπερπαραμέτρων που θα αποδώσουν τα καλύτερα δυνατά αποτελέσματα. Στην ενότητα αυτή θα παρουσιαστεί μέρος των συνολικών πειραμάτων που υλοποιήθηκαν καθώς πολλά εξ αυτών εξετάστηκαν για να παρατηρηθεί η απόκριση του δικτύου στις εκάστοτε αλλαγές. Συνήθως, σε κάθε νέο πείραμα εφαρμόζεται μία αλλαγή που αφορά μία υπερπαραμέτρο για να είναι εμφανής η επίδραση της στην εκπαίδευση του μοντέλου. Κατά τη διάρκεια των πειραμάτων δοκιμάστηκαν διάφοροι συνδυασμοί υπερπαραμέτρων που θεωρήθηκαν κατάλληλοι για τη φύση του προβλήματος από συναρτήσεις ενεργοποίησης, αρχικοποίηση βαρών, ρυθμό εκμάθησης κ.α. Επίσης, σε πολλά από αυτά επενέβη χειροκίνητα με αποτέλεσμα τον πρόωρο τερματισμό τους ως μία λύση αποφυγής του overfitting όταν η συνάρτηση κόστους άρχιζε να μειώνεται ή αλλιώς όταν η PSNR στο στάδιο του validation σταμάτησε να βελτιώνεται.

Ακόμα, πριν προβούμε στην αναλυτική παρουσίαση των υπερπαραμέτρων κάθε πειράματος είναι κρίσιμο να σημειωθεί ότι το 1^ο μοντέλο είναι υπεύθυνο για την βελτιστοποίηση των υπερπαραμέτρων του δικτύου επομένως είναι λογικό να εξετάσουμε πολλά μοντέλα-πειράματα σε σχέση με τα άλλα δύο. Επομένως, στο παρών κεφάλαιο θα αναπτυχθούν τα χαρακτηριστικά από πέντε μόλις πειράματα που θα συνοδεύονται με την απαραίτητη αξιολόγηση. Για την ακρίβεια θα παρατεθούν, τα τρία πρώτα πειράματα που αντιστοιχίζονται στο 1^ο μοντέλο, το οποίο απαρτίζεται από

δεδομένα εκπαίδευσης βασισμένα σε συνθετικά δεδομένα, ένα για το 2^ο μοντέλο που απαρτίζεται από δεδομένα εκπαίδευσης βασισμένα σε πραγματικά δεδομένα και ένα για το 3^ο μοντέλο που έχει δεδομένα εκπαίδευσης πραγματικά και δεδομένα αληθείας συνθετικά δεδομένα.

Καταρχάς, πριν ξεκινήσει η ανάλυση των υπερπαραμέτρων για κάθε πείραμα έχει σημασία να θυχθεί ότι οι υπερπαραμέτροι του δικτύου που διαπιστώθηκε ότι επηρεάζουν περισσότερο την απόδοση του είναι ο τρόπος αρχικοποίησης των βαρών και ο βελτιστοποιητής, κοινώς optimizer. Πράγματι, η αρχικοποίηση βαρών είναι καθοριστικής σημασίας για ένα μοντέλο καθώς αποτελεί δείκτη της ταχύτητας σύγκλισης και της απόδοσης του μοντέλου. Πρόκειται για ένα στάδιο όπου ορίζονται οι αρχικές τιμές για τα βάρη του δικτύου (παραμέτροι) και υλοποιείται πριν την έναρξη της εκπαίδευσης. Για αυτήν την υπερπαραμέτρο έγιναν δοκιμές με βασικές μεθόδους όπως η κανονική κατανομή, η Xavier και η He (Kaiming) μεταξύ αυτών. Λίγα λόγια σχετικά με την καθεμία:

- Στην κανονική κατανομή τα βάρη αρχικοποιούνται με τυχαίο τρόπο από μία κανονική κατανομή που έχει μέση τιμή το μηδέν (0) και μικρή διασπορά.
- Στην αρχικοποίηση Xavier ή αλλιώς (Glorot initialization) τα βάρη ρυθμίζονται σύμφωνα με τον αριθμό των νευρώνων εισόδου (n_{in}) και εξόδου (n_{out}). Εδώ τα βάρη αντλούνται από μία κατανομή με μέση τιμή 0 και διακύμανση $\sigma^2=2/(n_{in} + n_{out})$. Ενδείκνυται σε περιπτώσεις όπου χρησιμοποιούνται συναρτήσεις ενεργοποίησης όπως η tanh ή η σιγμοειδής (sigmoid). και η
- He (Kaiming) αρχικοποίηση που μοιάζει με αυτή του Xavier αλλά είναι κατασκευασμένη κατάλληλα για ενεργοποιήσεις ReLU και παραλλαγές της. Ομοίως, προτείνει την άντληση τυχαίων βαρών που ακολουθούν κανονική κατανομή με μέση τιμή 0 και διασπορά $2/n_{in}$, με n_{in} ο αριθμός των εισόδων στον κάθε νευρώνα.

Επιπρόσθετα, μετά την αρχικοποίηση βαρών ακολουθεί η εύρεση κατάλληλου βελτιστοποιητή (optimizer) που στοχεύει στην ελαχιστοποίηση της συνάρτησης κόστους μέσω της κατάλληλης προσαρμογής των παραμέτρων του μοντέλου. Για αυτό το σκοπό επιλέχθηκαν οι βελτιστοποιητές Adam και ο RMSprop.

- Ο Adam optimizer (Adaptive Moment Estimation) αποτελεί έναν από τους πιο ευρύ σε χρήση βελτιστοποιητές που χρησιμοποιείται κατά την εκπαίδευση νευρωνικών δικτύων λόγω της αποδοτικότητας και της ευκολίας χρήσης του ενώ είναι αποτελεσματικός ακόμα και σε μεγάλα σύνολα δεδομένων. Συνδυάζει τα

πλεονεκτήματα δύο άλλων αλγορίθμων των Adagrad και RMSprop και χαρακτηρίζεται από δύο βασικές αρχικές αυτές του προσαρμοστικού ρυθμού εκμάθησης και της ορμής. Με άλλα λόγια δίνει τη δυνατότητα επιλογής διαφορετικού ρυθμού εκμάθησης για κάθε παράμετρο επιτρέποντας έτσι τη ταχύτερη σύγκλιση και την αποφυγή τοπικών ελαχίστων. Αυτό επιτυγχάνεται και μέσω της ορμής που αξιοποιεί την εκθετική εξομάλυνση των προηγούμενων παραγώγων (gradients).

- Ο RMSprop με τη σειρά του αποτελεί μία βελτίωση του Adagrad αλγορίθμου και αναπτύχθηκε για την αντιμετώπιση προβλημάτων που ανακύπτουν λόγω σταθερού ρυθμού εκμάθησης και ιδίως όταν οι παράγωγοι (gradients) παρουσιάζουν μεγάλες διακυμάνσεις. Τέλος, στον RMSprop χρησιμοποιείται ένας εκθετικός κινητός μέρος όρος των τετραγωνικών παραγώγων. Ο εν λόγω βελτιστοποιητής επιλέχθηκε ακόμα διότι βιβλιογραφικά έχει φανεί ότι σε δίκτυα που χρησιμοποιούν attention ο optimizer αυτός αποδίδει καλύτερα.

Παράλληλα, μία άλλη επιλογή που έχει ενδιαφέρον να σχολιαστεί αφορά τις συναρτήσεις ενεργοποίησης, που εισάγουν τη μη γραμμικότητα στο μοντέλο και επιτρέπουν την εκμάθηση και αναπαράσταση πιο περίπλοκων σχέσεων στα δεδομένα. Ειδικότερα, αυτές που δοκιμάστηκαν ήταν οι ReLU, Leaky ReLU και SeLU.

- Η ReLU αποτελεί μία καλή προεπιλογή σαν συνάρτηση ενεργοποίησης για τα περισσότερα προβλήματα, λόγω της απλότητας και της αποτελεσματικότητας της, κάτι που δικαιολογεί την ευρύ της χρήση. Στηρίζεται στη σχέση που δίνει ως έξοδο το μέγιστο μεταξύ του μηδενός και της τιμής εισόδου, δηλ. ισχύει $\text{ReLU}(x) = \max(0, x)$. Πλεονεκτεί ακόμα μειώνοντας το πρόβλημα της εξαφάνισης των «gradients» συμβάλλοντας στη ταχύτερη σύγκλιση ενώ μειονεκτεί λόγω του προβλήματος των «νεκρών νευρώνων» που ανακύπτει, όπου νευρώνες σταματούν να ενεργοποιούνται για όλες τις εισόδους.
- Η Leaky ReLU αποτελεί μία παραλλαγή της ReLU με επιπλέον πλεονέκτημα ότι επιλύει το πρόβλημα των νεκρών νευρώνων. Δίνεται από τη σχέση: $\text{Leaky ReLU}(x) = \max(\alpha x, x)$.
- Τέλος, η SeLU είναι μία εξελιγμένη παραλλαγή της ELU, που λειτουργεί καλύτερα σε βαθιά δίκτυα. Συμβάλλει όπως και οι προηγούμενοι αλγόριθμοι στην επίλυση του προβλήματος εξαφάνισης των «gradients» και ακόμα έχει την ιδιότητα της αυτοκανονικοποίησης. Αυτό σημαίνει ότι οι έξοδοι των νευρώνων τείνουν να παρουσιάζουν μέση τιμή κοντά στο 0 και διασπορά στη μονάδα. Το

χαρακτηριστικό αυτό απαλλάσσει από την ανάγκη εφαρμογών κανονικοποίησης όπως είναι το batch normalization. Στα μειονεκτήματα της κατατάσσεται το μεγάλο υπολογιστικό της κόστος λόγω της χρήσης εκθετικών υπολογισμών σε αρνητικές τιμές δεδομένων εισόδου.

4.3.4. Υλοποιημένα Μοντέλα – Διάταξη Πειραμάτων

Ας ξεκινήσουμε, με τα πρώτα τρία πειράματα που αποτελούν εφαρμογή του 1^{ου} μοντέλου, όπως προαναφέρθηκε. Σε αυτό το πείραμα αλλά και σε όλα όσα ακολουθήσουν τα δεδομένα εκπαίδευσης (inputs & labels) αλλά και τα δεδομένα εξόδου είναι 16, αφού γίνεται λόγος για πολυφασματικές εικόνες που καταγράφουν πληροφορία σε 16 φασματικά κανάλια. Ακόμα, ορίζεται συνθήκη τερματισμού ο αριθμός εποχών ίσος με 300 αν και σε περιπτώσεις όπου η συνάρτηση κόστους σταματάει να φθίνει επιλέγεται να τερματιστεί πρόωρα η διαδικασία εκπαίδευσης.

Μία άλλη υπερπαράμετρο που πειραματιστήκαμε ήταν ο ρυθμός εκμάθησης (Learning Rate - LR), όπου δοκιμάστηκαν μικρότερες και μεγαλύτερες τιμές και καταλήξαμε ότι η βέλτιστη απόδοση ίσχυε για αρχική τιμή $LR=0,01$. Ακόμα, στην διαδικασία αυτή της εκπαίδευσης του μοντέλου χρησιμοποιήθηκε ένας αλγόριθμος ο «LR scheduler» που πρακτικά προσαρμόζει το LR ρυθμίζοντας το βήμα όπου αυτός μεταβάλλεται με κατεύθυνση την εύρεση ελαχίστου στη συνάρτηση κόστους. Η σωστή επιλογή ενός scheduler επηρεάζει σημαντικά τον ρυθμό σύγκλισης και βελτιώνει την απόδοση του μοντέλου. Έτσι, γίνεται σαφές ότι και οι υπερπαράμετροι που ορίζονται σε έναν scheduler είναι ιδιόζουσας σημασίας. Αυτοί είναι το 1) patience: που αντιστοιχεί στο πλήθος των εποχών που κάνει «υπομονή» το μοντέλο όταν το training loss σταματάει να φθίνει και το 3) βήμα κατά το οποίο διαφοροποιείται - μειώνεται κάθε φορά ο ρυθμός εκμάθησης. Χάρη, λοιπόν στη χρήση ενός scheduler μπορούν να αποφευχθούν τοπικά ελάχιστα και να μειωθεί η υπερπροσαρμογή των δεδομένων (overfitting) παρέχοντας καλύτερη γενίκευση.

1. Στο πρώτο (1^ο) πείραμα οι υπερπαράμετροι που ορίστηκαν είναι αυτοί που συνοψίζονται στον Πίνακα 4.2. Όπως προαναφέρθηκε, εδώ ορίζονται ως αριθμός εποχών οι 300, η αρχική τιμή ρυθμού εκμάθησης είναι 0,01 που μεταβάλλεται χάρη στον scheduler κατά τη διάρκεια της εκπαίδευσης με patience 15. Ακόμα, η αρχικοποίηση των βαρών έγινε σύμφωνα με την κανονική

κατανομή, για βελτιστοποιητή και συνάρτηση ενεργοποίησης εφαρμόστηκαν οι «Adam» και «ReLU» αντίστοιχα.

1^ο Πείραμα – 1^ο ΜΟΝΤΕΛΟ	
Αριθμός Εποχών	300
Ρυθμός Εκμάθησης	0,01 (initial LR)
Patience	15
Optimizer	Adam
Activation Function	ReLU
Weight Initialization	Normal Distribution

Πίνακας 4.2: Υπερπαράμετροι που χρησιμοποιήθηκαν κατά την εφαρμογή του 1^{ου} πειράματος για το 1^ο μοντέλο.

2. Στο 2^ο πείραμα που ακολούθησε επί του 1^{ου} Μοντέλου κρίνεται σκόπιμο να αναφερθούν μόνο οι υπερπαράμετροι που διαφοροποιούνται από το 1^ο, που στην προκειμένη περίπτωση αφορά μόνο τη μέθοδο αρχικοποίησης βαρών. Συνεπώς, οι υπερπαράμετροι που συνθέτουν το 2^ο πείραμα είναι αυτοί του Πίνακα 4.3.

2^ο Πείραμα – 1^ο ΜΟΝΤΕΛΟ	
Αριθμός Εποχών	300
Ρυθμός Εκμάθησης	0,01 (initial LR)
Patience	15
Optimizer	Adam
Activation Function	ReLU
Weight Initialization	Kaiming

Πίνακας 4.3: Υπερπαράμετροι που χρησιμοποιήθηκαν κατά την εφαρμογή του 2^{ου} πειράματος για το 1^ο μοντέλο.

3. Στο 3^ο πείραμα και τελευταίο που εφαρμόστηκε για το 1^ο Μοντέλο και που θα παρουσιαστεί στα πλαίσια της παρούσας ΔΕ η διαφοροποίηση σε σχέση με τα προηγούμενα μοντέλα έγκειται και πάλι στην αλλαγή του τρόπου αρχικοποίησης των βαρών. Αυτή τη φορά επιλέχθηκε η αρχικοποίηση Xavier ή αλλιώς (Glorot initialization). Το σύνολο των υπερπαραμέτρων που συνθέτουν το πείραμα αυτό περιέχονται στον Πίνακα 4.4 που ακολουθεί.

3^ο Πείραμα – 1^ο ΜΟΝΤΕΛΟ	
Αριθμός Εποχών	300
Ρυθμός Εκμάθησης	0,01 (initial LR)
Patience	15
Optimizer	Adam
Activation Function	ReLU

Weight Initialization	Xavier
------------------------------	---------------

Πίνακας 4.4: Υπερπαράμετροι που χρησιμοποιήθηκαν κατά την εφαρμογή του 2^{ου} πειράματος για το 1^ο μοντέλο.

Τώρα, θα παρουσιαστεί ένα πείραμα που εκπροσωπεί το 2^ο Μοντέλο, που υπενθυμίζεται ότι αποτελείται από δεδομένα εκπαίδευσης βασισμένα σε πραγματικά δεδομένα και πιο συγκεκριμένα λαμβάνει σαν είσοδο τις παρεμβαλλόμενες πραγματικές εικόνες και σαν δεδομένα αληθείας τις πραγματικές πολυφασματικές. Για το μοντέλο αυτό έγιναν όμοια με όλα τα προηγούμενα πολλές δοκιμές στις υπερπαραμέτρους αλλά εδώ θα παρουσιαστεί ένα μόλις πείραμα με υπερπαραμέτρους ίδιες με αυτές του 2^{ου} Πειράματος που αναπτύχθηκε για το 1^ο Μοντέλο. Αναλυτικά οι υπερπαραμέτροι που επιλέχθηκαν αναρτώνται στον Πίνακα 4.5. Η επιλογή αυτή δεν είναι τυχαία καθώς παρατηρήθηκε ότι ο συγκεκριμένος συνδυασμός υπερπαραμέτρων αποδίδουν το βέλτιστο κατά την εκπαίδευση του δικτύου. Επίσης, είναι σημαντικό να αναφερθεί ότι το παρόν μοντέλο(2^ο) αναμένεται να σημειώσει καλύτερες μετρικές στο στάδιο της αξιολόγησης σε σχέση με το 1^ο Μοντέλο εξάγοντας εικόνες που προσεγγίζουν περισσότερο τις πραγματικές, δεδομένου ότι εκπαιδεύεται σε εικόνες καλύτερης ανάλυσης, αλλά αυτό θα αναπτυχθεί σε επόμενο υποκεφάλαιο.

1ο Πείραμα – 2ο ΜΟΝΤΕΛΟ	
Αριθμός Εποχών	300
Ρυθμός Εκμάθησης	0,01 (initial LR)
Patience	15
Optimizer	Adam
Activation Function	ReLU
Weight Initialization	Kaiming

Πίνακας 4.5: Υπερπαράμετροι που χρησιμοποιήθηκαν κατά την εφαρμογή του 1^{ου} πειράματος για το 2^ο μοντέλο.

Όμοια, για το 3^ο μοντέλο, που χρησιμοποιεί τα ίδια δεδομένα αληθείας με το πρώτο αλλά ως είσοδο δέχεται τις παρεμβαλλόμενες πραγματικές εικόνες που εισάγονται και στο 2^ο, θα παρουσιαστεί ένα πείραμα. Αποφασίστηκε να διατρεχθεί το μοντέλο αυτό σε αντιπαράθεση με το πρώτο προκειμένου να εξεταστεί η επίδοση του, που αναμένεται να είναι καλύτερη έχοντας δεδομένα εισόδου που έχουν διατηρήσει μέρος των υψηλότερων συχνοτήτων. Αυτό μένει να εξεταστεί στην ενότητα της Αξιολόγησης που θα ακολουθήσει. Οι υπερπαραμέτροι και σε αυτή την περίπτωση μοιάζουν με αυτές του 2^{ου} πειράματος που υλοποιήθηκε για το 1^ο μοντέλο με τη διαφορά ότι η υπερπαραμέτρος (patience) του ρυθμού εκμάθησης πήρε την τιμή 10.

Αναλυτικά, ο Πίνακας 4.6 εκθέτει το σύνολο των υπερπαραμέτρων που ορίστηκαν για το εν λόγω πείραμα.

1ο Πείραμα – 3ο ΜΟΝΤΕΛΟ	
Αριθμός Εποχών	300
Ρυθμός Εκμάθησης	0,01 (initial LR)
Patience	10
Optimizer	Adam
Activation Function	ReLU
Weight Initialization	Kaiming

Πίνακας 4.6: Υπερπαραμέτροι που χρησιμοποιήθηκαν κατά την εφαρμογή του 1^{ου} πειράματος για το 3^ο μοντέλο.

Από όλα τα παραπάνω πειράματα γίνεται σαφής μία επαναληψιμότητα ως προς την επιλογή των υπερπαραμέτρων και αυτό διότι επιλέχθηκαν να παρουσιαστούν τα πειράματα εκείνα με τις καλύτερες επιδόσεις. Με άλλα λόγια, δεν θα εξεταστεί το σύνολο των πειραμάτων καθώς πολλά από αυτά αποτελούν αρχικά τεστ που υλοποιήθηκαν για προσωπική παρατήρηση και προκειμένου να διακριθεί ποιοι υπερπαραμέτροι συμβάλλουν και «σαμποτάρουν» αντίστοιχα την εκπαίδευση του εκάστοτε μοντέλου. Συνεπώς, ο συνδυασμός των υπερπαραμέτρων που μας εξυπηρετούν είναι η ReLU σαν συνάρτηση ενεργοποίησης, ο Adam βελτιστοποιητής, ο αρχικός ρυθμός εκμάθησης με τιμή 0,01 που μεταβάλλεται με τη χρήση scheduler και σαν αρχικοποίηση βαρών ο Kaiming που φαίνεται να σημειώνει καλύτερες μετρικές τόσο στη συνάρτηση κόστους όσο και στην PSNR, για αυτό και επιλέχθηκε να εφαρμοστεί σε όλα τα μοντέλα.

4.4. Αξιολόγηση

Έχοντας ολοκληρώσει τα πειράματα που αναλύθηκαν για τα τρία μοντέλα έπεται μία σημαντική διαδικασία. Αυτή δεν είναι άλλη από μία εφαρμογή κανονικοποίησης, το early stopping, σύμφωνα με το οποίο επεμβαίνουμε χειροκίνητα με σκοπό τον τερματισμό της εκπαίδευσης προκειμένου να αποφευχθεί η υπερπροσαρμογή του μοντέλου μας στα δεδομένα εκπαίδευσης. Με άλλα λόγια κατά το overfitting, το μοντέλο μαθαίνει τόσο καλά τα δεδομένα εκπαίδευσης που σαν αποτέλεσμα θα εξάγει καλά αποτελέσματα μόνο για σύνολα δεδομένων που μοιάζουν σε αυτά. Η συνθήκη τερματισμού στηρίζεται σε δύο παράγοντες· αφενός το σημείο

όπου η συνάρτηση κόστους σταματάει να φθίνει και αφετέρου την εποχή εκείνη όπου παράλληλα και η μετρική PSNR, που καταγράφει τις διαφορές στις ψηφιακές τιμές μεταξύ δύο εικόνων, ξεκινάει να συγκλίνει σε κάποια τιμή.

Η διαδικασία αυτή θα πραγματοποιηθεί για καθένα από τα τρία μοντέλα για να εξεταστεί η απόδοσή τους. Στην περίπτωση του πρώτου μοντέλου, βάση του οποίου γίνεται βελτιστοποίηση των υπερπαραμέτρων, θα επιλεγεί ένα από τα τρία πειράματα, αυτό δηλαδή που θα σημειώσει καλύτερη απόκριση κατά την αξιολόγηση (validation). Αυτό σημαίνει ότι θα επιλεγεί το μοντέλο εκείνο που θα αποδώσει καλύτερα κατά το validation loss και ταυτόχρονα η μετρική PSNR στο σύνολο των δεδομένων αξιολόγησης θα καταγράψει τις μεγαλύτερες τιμές. Παρατηρώντας, λοιπόν, την απόδοση των πειραμάτων επιλέγεται το 2^ο Πείραμα σαν το καλύτερο που θα αντιπροσωπεύσει το 1^ο Μοντέλο.

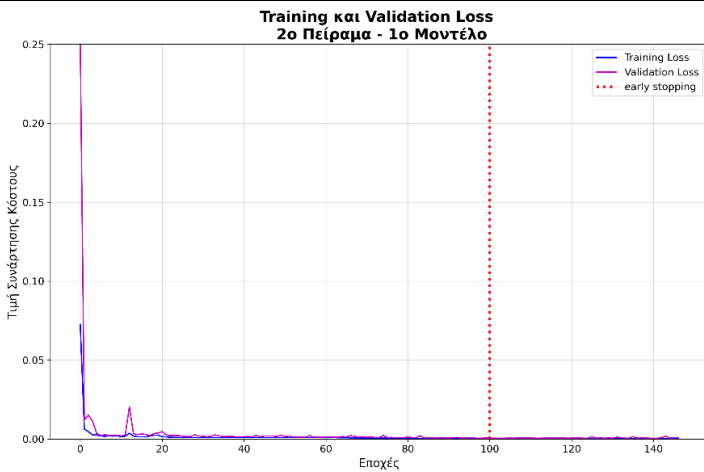
Εδώ πριν αξιολογηθούν τα αποτελέσματα των μοντέλων είναι απαραίτητο να αναλυθεί η λογική που εφαρμόστηκε για την πρόωρη διακοπή της εκπαίδευσης σε καθένα από αυτά. Καταρχάς, για κάθε μοντέλο κατασκευάστηκαν δύο διαγράμματα το ένα αφορούσε την εξέλιξη της συνάρτησης κόστους για το στάδιο της εκπαίδευσης και της αξιολόγησης αντίστοιχα (που οπτικοποιήθηκαν σε κοινό διάγραμμα) και το δεύτερο την πορεία της μετρικής PSNR στο στάδιο της αξιολόγησης. Οι training και validation loss είναι δύο μετρικές πολύ χρήσιμες κατά την εκπαίδευση ενός μοντέλου και αποτελούν κριτήρια αξιολόγησης του κατά τη διάρκεια και έπειτα από το στάδιο εκπαίδευσης ενώ μπορούν να υποδείξουν αν το μοντέλο υπερ- ή υπο-προσαρμόζεται. Ειδικά, η συνάρτηση κόστους -που υπολογίζεται κατά το στάδιο της εκπαίδευσης- (Training Loss) υποδεικνύει πόσο καλά μαθαίνει το μοντέλο από τα δεδομένα εκπαίδευσης. Από την άλλη, το Validation Loss σηματοδοτεί πόσο καλά μπορεί να γενικεύσει το μοντέλο σε νέα δεδομένα που δεν έχει ξαναδεί. Ο τερματισμός της εκπαίδευσης στηρίχθηκε και στα δύο διαγράμματα λαμβάνοντας υπόψη τη ζώνη εκείνη όπου τα Loss Function σταμάτησαν να φθίνουν και ταυτόχρονα η PSNR σταμάτησε να βελτιώνεται.

Ειδικότερα, για το πρώτο μοντέλο που εκπαιδεύτηκε με δεδομένα εισόδου τις παρεμβλλόμενες συνθετικές εικόνες κατασκευάστηκαν τα διαγράμματα που αναρτώνται στο Σχήμα 4.8. Στην πάνω σειρά φαίνονται οι συναρτήσεις κόστους και οι τιμές της μετρικής καθόλη τη διάρκεια των εποχών εκπαίδευσης ενώ στην από κάτω σειρά παρατίθενται σε εστίαση η περιοχή ενδιαφέροντος, που θα εξεταστεί για να οριστεί ο πρόωρος τερματισμός του μοντέλου, δηλαδή θα εκτελεστεί το early stopping. Το zoom αυτό κρίνεται απαραίτητο καθώς στην αρχική κλίμακα δεν είναι ευκρινής η

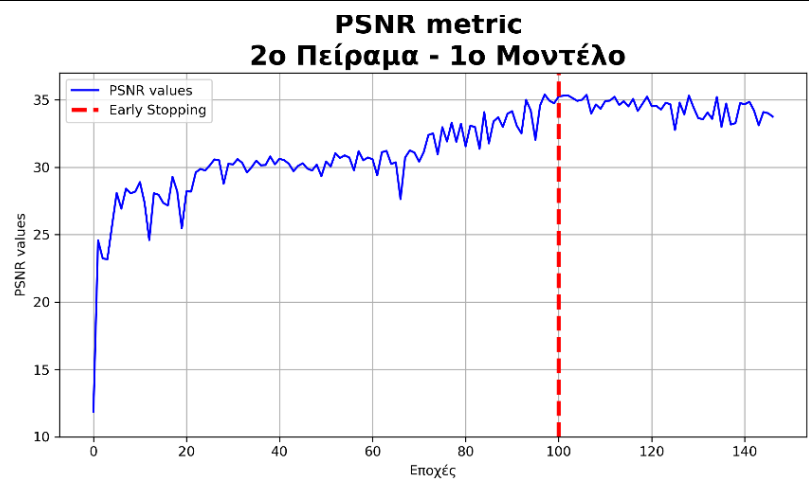
συμπεριφορά των γραφημάτων. Εδώ είναι καίριας σημασίας να σημειωθεί ότι το μοντέλο έχει εκπαιδευτεί ώστε να αποθηκεύει τα βάρη κάθε 5 εποχές (checkpoint).

Ελέγχοντας τα διαγράμματα φαίνεται ότι γύρω στις 100 εποχές οι συναρτήσεις κόστους (σε στάδιο εκπαίδευσης και αξιολόγησης) παύουν να μειώνονται αισθητά. Σταδιακά μάλιστα το μοντέλο μετά την 130^η εποχή αρχίζει να υπερπροσαρμόζεται στα δεδομένα. Αντίστοιχα, στο διάγραμμα της μετρικής PSNR στην 100 εποχή παρουσιάζεται η μέγιστη τιμή με PSNR=35,39 και από εκεί και έπειτα φαίνεται να σταματάει να βελτιώνεται. Ακόμα, δεδομένου ότι στην PSNR η 100^η εποχή βρίσκεται στην περιοχή σύγκλισης θα είναι και αυτή που θα επιλεγεί για να τερματιστεί η εκπαίδευση του μοντέλου μας και τα αποτελέσματα σε αυτό το σημείο θα αξιοποιηθούν για την αξιολόγηση του.

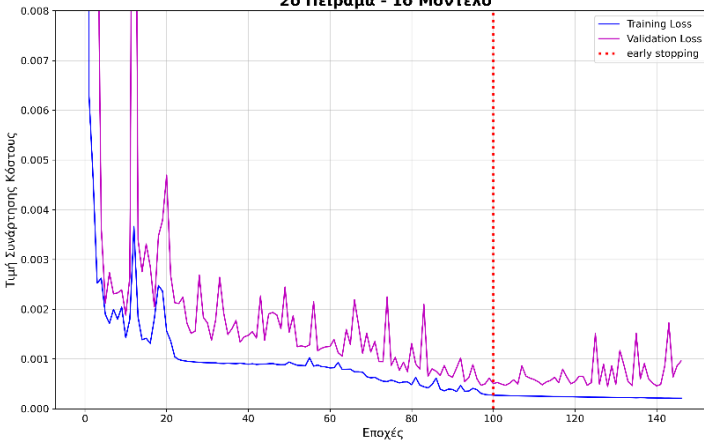
Training & Validation Loss



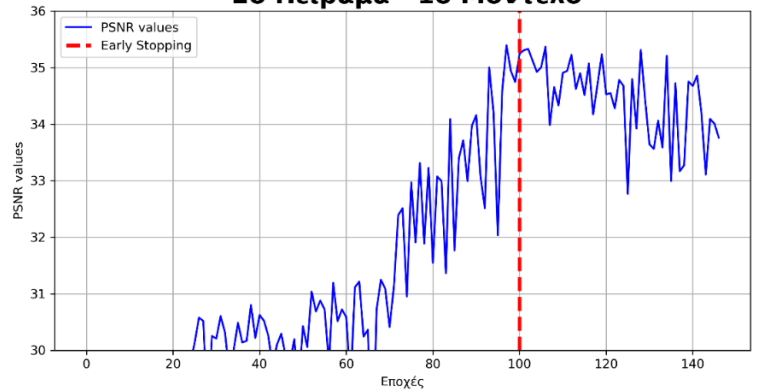
PSNR metric



**Training και Validation Loss
2ο Πείραμα - 1ο Μοντέλο**



**PSNR metric
2ο Πείραμα - 1ο Μοντέλο**



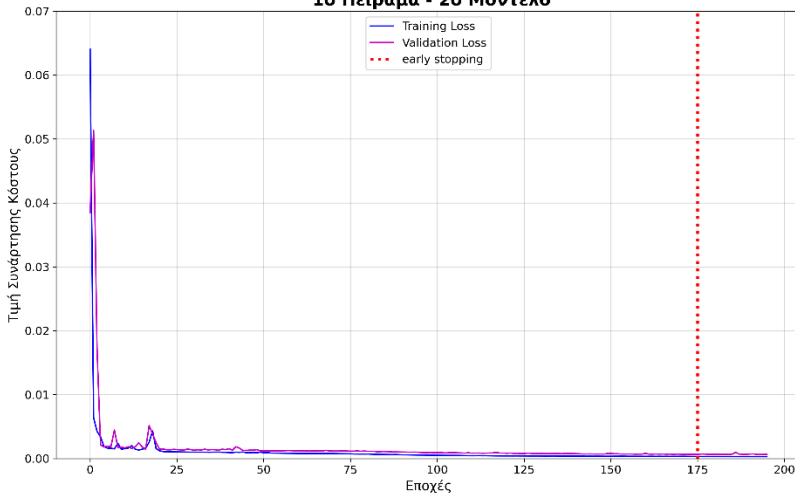
Σχήμα 4.8: Διαγράμματα 1) Training & Validation Loss, 2) μετρικής PSNR για το 1^ο μοντέλο.

Για το δεύτερο μοντέλο, αντίστοιχα, κατασκευάστηκαν τα διαγράμματα συνάρτησης κόστους και PSNR (βλ. Σχήμα 4.9). Όσον αφορά την εποχή που θα επιλεγεί για να γίνει το early stopping στην περίπτωση αυτή φαίνεται από την εποχή 175 και έπειτα ότι οι συναρτήσεις κόστους έχουν πολύ μικρή μεταβολή και οι τιμές μειώνονται αργά. Εκεί, η PSNR αποκτά τιμή $PSNR=33.41$ και έπεται ένα διάστημα με διαρκή σκαμπανεβάσματα στις τιμές, κάτι που σηματοδοτεί ότι η εκπαίδευση πρέπει να σταματήσει. Ακόμα, στην 175^η εποχή παρότι μπορεί να μη σημειώνεται η καλύτερη τιμή PSNR γίνεται εμφανές και από το διάγραμμα ότι περίξ της τιμής 33 σταματάει να βελτιώνεται η μετρική, οπότε αυτή η εποχή επιλέγεται για την εκτέλεση του early stopping.

Το μοντέλο αυτό ανακαλείται ότι δέχεται ως δεδομένα εισόδου τις παρεμβαλλόμενες πραγματικές εικόνες και τις πραγματικές πολυφασματικές (ως δεδ. Αληθείας). Λόγω αυτού, σημειώνεται ότι οι loss functions αργούν αισθητά να σταματήσουν τη φθίνουσα πορεία τους και η PSNR να συγκλίνει σε κάποια τιμή, σε σχέση με το προηγούμενο μοντέλο. Το συμβάν αυτό μπορεί να αποδοθεί στην επιπλέον πληροφορία υψηλής συχνότητας που φέρουν οι πραγματικές εικόνες. Βέβαια, παρά τη βραδύτητα σύγκλισης η μετρική PSNR καταγράφει υποδεέστερες τιμές, δηλαδή οι demosaiced εικόνες δεν προσεγγίζουν τόσο καλά τις πραγματικές πολυφασματικές. Αυτό, αρχικά δεν αναμενόταν αλλά από την άλλη μπορεί να αποδοθεί στη δυσκολία του δικτύου να αναπαράγει τις υψηλές συχνότητες και ευρύτερα εικόνες με τόση πληροφορία. Ακόμα, πρέπει να ληφθεί υπόψη ότι τα δεδομένα εισόδου του συγκεκριμένου δικτύου (BIL GT) όπως αποδείχθηκε και από το γράφημα του Σχήματος 4.7. εν τέλει δεν είχαν συγκρατήσει επιπλέον πληροφορία από αυτά του πρώτου μοντέλου (BIL MD²IP), οπότε η μόνη διαφοροποίηση τους έγκειται στα Labels και την επιπλέον πληροφορία υψηλών συχνοτήτων που συντηρούν. Τέλος, είναι πιθανό η κατανομή των δεδομένων του MD²IP να είναι πιο κατανοητή από αυτή των GT, για αυτό και αποδίδει καλύτερα το μοντέλο στην πρώτη περίπτωση.

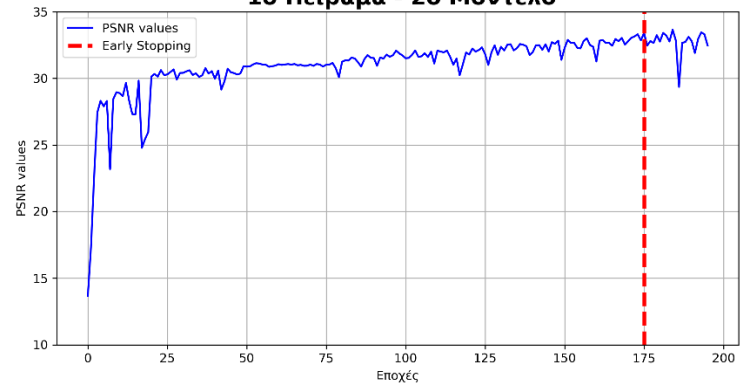
Training & Validation Loss

Training και Validation Loss
1ο Πείραμα - 2ο Μοντέλο

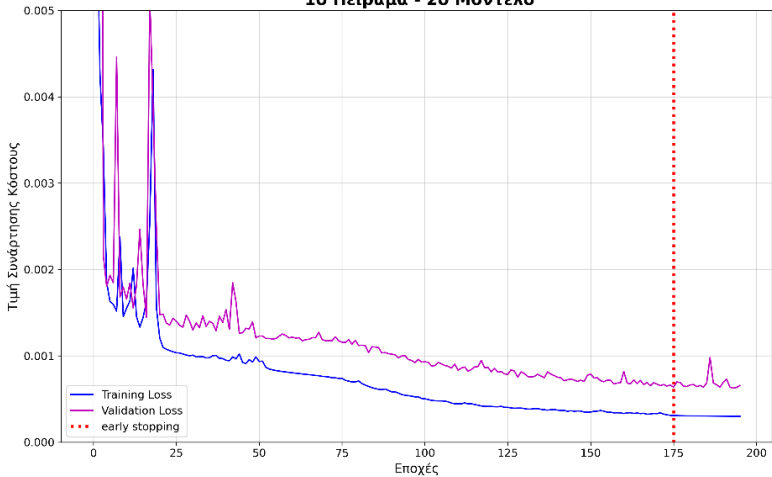


PSNR metric

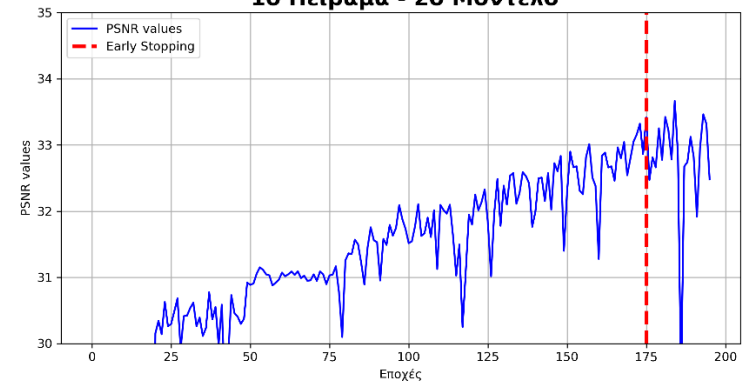
PSNR metric
1ο Πείραμα - 2ο Μοντέλο



Training και Validation Loss
1ο Πείραμα - 2ο Μοντέλο



PSNR metric
1ο Πείραμα - 2ο Μοντέλο



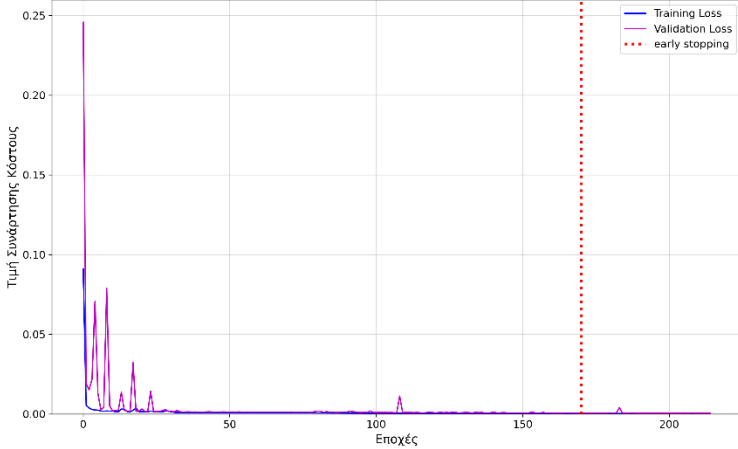
Σχήμα 4.9: Διαγράμματα 1) Training & Validation Loss, 2) μετρικής PSNR για το 2^ο μοντέλο.

Μία ευρύ παρατήρηση που επαληθεύει τη θεωρία είναι ότι το training & validation loss εμφανίζουν σταθερή συμπεριφορά από την άποψη ότι το validation σχεδόν πάντα σημειώνει μεγαλύτερες τιμές σε σχέση με αυτές του training για κάθε εποχή. Όμοια, για το τρίτο μοντέλο τα διαγράμματα που συντέθηκαν παρουσιάζονται στο Σχήμα 4.10. Να σημειωθεί ότι το μοντέλο αυτό έχει τα ίδια δεδομένα αληθείας με το πρώτο αλλά σαν δεδομένα εισόδου δέχεται τις παρεμβαλλόμενες πραγματικές εικόνες έναντι των συνθετικών. Κατά συνέπεια αυτού όπως και στο δεύτερο μοντέλο αναμένεται βραδύτερη σύγκλιση της PSNR και των συναρτήσεων κόστους αντίστοιχα. Πράγματι σε αυτό το μοντέλο οι τιμές στη συνάρτηση κόστους μειώνονται πολύ αργά από την 170^η εποχή και η PSNR συγκλίνει γύρω από τη τιμή $PSNR(170^{\text{η}}) = 34,10$. Έτσι η 170^η εποχή είναι αυτή που επιλέγεται να γίνει το early stopping. Τέλος, στο μοντέλο

αυτό παρατηρείται υποδεέστερη απόδοση στη μετρική PSNR από ότι στο πρώτο μοντέλο. Η διαφοροποίηση από το πρώτο έγκειται στα δεδομένα εισόδου και πιο συγκεκριμένα στην επιπλέον πολυπλοκότητα που φέρουν οι παρεμβαλλόμενες πραγματικές εικόνες, αφού έχουν προκύψει ως παράγωγα εικόνων με υψηλότερες συχνότητες και περισσότερη σύνθεση σε σχέση με τις παρεμβαλλόμενες συνθετικές. Ως επόμενο, αφενός το μοντέλο δυσκολεύεται να τις διαχειριστεί και αργεί να συγκλίνει και αφετέρου τα τελικά αποτελέσματα είναι υποδεέστερα (θα μπορούσε να έχει να κάνει με την γενική κατανομή των δεδομένων αλλά αυτό αποτελεί μια πιο σύνθετη εκτίμηση).

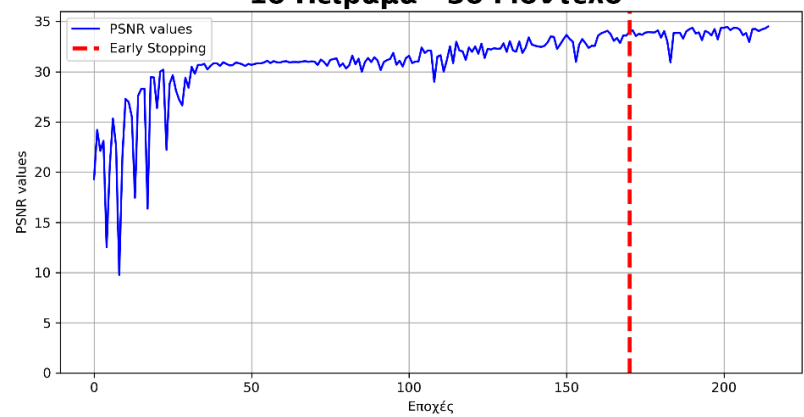
Training & Validation Loss

Training και Validation Loss
1ο Πείραμα - 3ο Μοντέλο

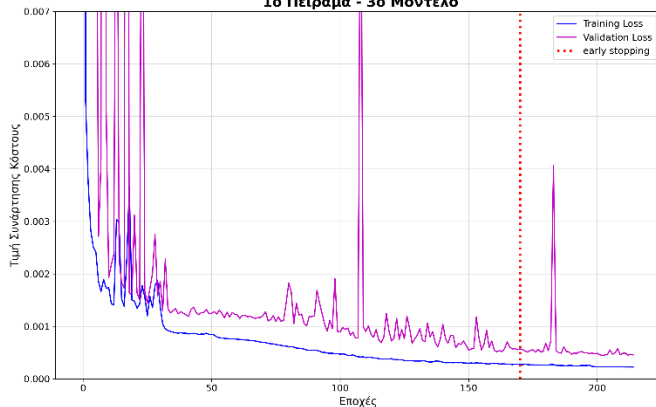


PSNR metric

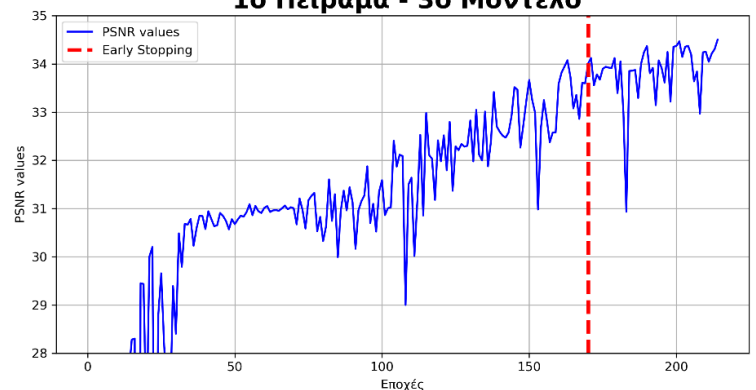
PSNR metric
1ο Πείραμα - 3ο Μοντέλο



Training και Validation Loss
1ο Πείραμα - 3ο Μοντέλο



PSNR metric
1ο Πείραμα - 3ο Μοντέλο



Σχήμα 4.10: Διαγράμματα 1) Training & Validation Loss, 2) μετρικής PSNR για το 3^ο μοντέλο.

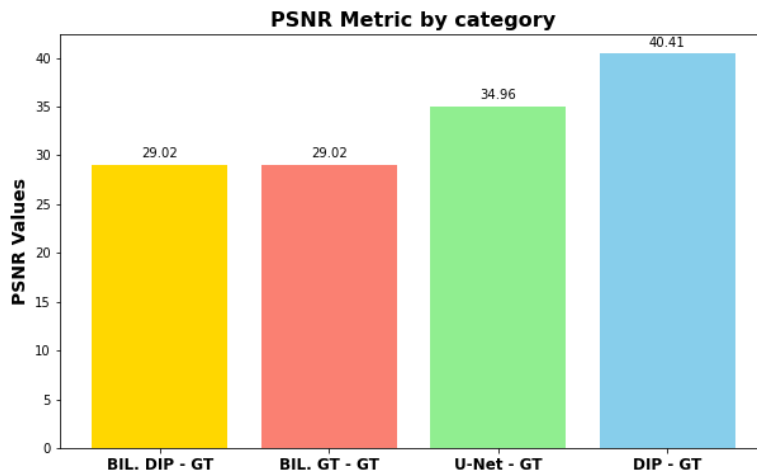
4.4.1. Ποσοτική Αξιολόγηση

Επιπρόσθετα, για το 1^ο μοντέλο πέρα της μετρικής PSNR που αφορούσε τη σύγκριση μεταξύ των αποτελεσμάτων του validation set με τα παράγωγα του MD²IP (που αποτέλεσαν δεδομένα εκπαίδευσης του μοντέλου) υπολογίστηκε επιπλέον και η μετρική που συνέκρινε τα αποτελέσματα του U-Net με τις πραγματικές εικόνες. Η PSNR αυτή είναι πολύ χρήσιμη για την αντικειμενική σύγκριση της με τα υπόλοιπα μοντέλα. Έτσι λοιπόν προκύπτει το επόμενο διάγραμμα του Σχήματος 4.11. Σε αυτό αναπαρίστανται οι μετρικές PSNR που αφορούν το σύνολο των εικόνων -που αξιοποιήθηκαν ως σύνολο αξιολόγησης (validation set)- και πιο συγκεκριμένα των παρεμβαλλόμενων συνθετικών και πραγματικών δεδομένων, των παραγώγων του μοντέλου μας (U-Net) και του μοντέλου (MD²IP) που εφαρμόστηκε στο 3^ο κεφάλαιο σε σχέση με τις πραγματικές εικόνες. Το διάγραμμα με μπάρες οπτικοποιεί την απόδοση του μοντέλου σε σχέση με τα δεδομένα εισόδου και το μοντέλο που εκπαιδεύτηκε.

Βάση αυτού, λοιπόν, διαπιστώνεται ότι η μετρική που σημειώνουν οι παρεμβαλλόμενες εικόνες (συνθετικές και πραγματικές) αποτελούν το «κατώφλι» ή αλλιώς το ελάχιστο όριο που θα φτάσει το μοντέλο καθώς για τιμή μικρότερη της PSNR=29,02 θα θεωρείται ότι το μοντέλο παρουσιάζει αδυναμίες. Αντίστοιχα, η τιμή της μετρικής που καταγράφει η 4^η στήλη αποτελεί το άνω όριο καθώς πρόκειται για την επίδοση του μοντέλου του MD²IP, βάση του οποίου εκπαιδεύεται το μοντέλο. Από το Σχήμα 4.11 φαίνεται ότι το μοντέλο μας σημείωσε PSNR=34,96, δηλαδή βελτιώθηκε η απόδοση του κατά 6 μονάδες σε σχέση με τη διγραμμική παρεμβολή, παράγωγα της οποίας χρησιμοποιήθηκαν ως δεδομένα εισόδου. Ως είναι επόμενο, αυτό σημαίνει ότι το μοντέλο δούλεψε και εκπαιδεύτηκε σε πολύ καλό βαθμό να αναπαράγει πολυφασματικές εικόνες πλήρους ανάλυσης. Εδώ, έχει ενδιαφέρον να σχολιαστεί ότι η απόδοση του M1 που καταγράφηκε από τις PSNR₁(U-Net - MD²IP)=35,39 και PSNR₁(U-Net - GT)=34,96 παρουσιάζουν πολύ μικρή απόκλιση ενώ αναμενόταν η πρώτη μετρική να σημείωνε σημαντικά μεγαλύτερη τιμή από τη δεύτερη.

Ταυτόχρονα, από το άνω όριο PSNR (MD²IP - GT) υπάρχει απόκλιση 5,4 μονάδων που συμβολίζει ότι το U-Net έχει πάει καλά και έχει περιθώριο βελτίωσης. Αυτό βέβαια δεν είναι τόσο εύκολο να επιτευχθεί καθώς αρχικά ένα μοντέλο δεν μπορεί να γίνει καλύτερο από τα δεδομένα αληθείας ή κοινώς από το μοντέλο που βασίστηκε για να εκπαιδευτεί παρά στην καλύτερη περίπτωση να προσεγγίσει αρκετά την απόδοση του και αυτό αν πρόκειται για ένα μεγάλο σύνολο δεδομένων. Δευτερευόντως είναι δύσκολο να αναπαραχθούν οι υψηλές συχνότητες, οπότε η

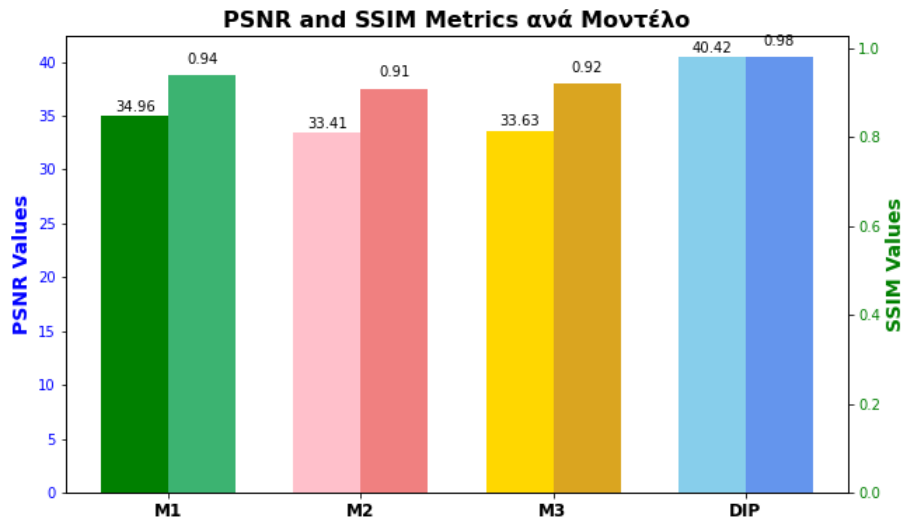
χαμηλότερη επίδοση του U-Net σε σχέση με το MD²IP είναι απολύτως δικαιολογημένη, αφού τα δεδομένα από το δεύτερο συντηρούν μέρος αυτών. Κλείνοντας, είναι κρίσιμο να σημειωθεί ότι το σύνολο των δεδομένων που έχει συλλεχθεί για την παρούσα ΔΕ θεωρείται μικρό άρα αντιμετωπίζουμε μία δύσκολη πρόκληση ενώ επιπλέον δεν υπάρχει κάποιο προεκπαιδευμένο δίκτυο με έτοιμα - αρχικοποιημένα βάρη, που θα μπορούσε να χρησιμοποιηθεί σαν “backbone” στο ίδιο πρόβλημα στην παρούσα φάση. Λαμβάνοντας υπόψη όλα όσα έχουν θιχθεί στην παρούσα παράγραφο κρίνεται ότι το δίκτυο με τη συγκεκριμένη αρχιτεκτονική υλοποιεί την εκπαιδευτική διαδικασία για την οποία χρησιμοποιήθηκε εξάγοντας πολύ καλά αποτελέσματα που μένει να σχολιαστούν και σύμφωνα με ποιοτικά κριτήρια αξιολόγησης.



Σχήμα 4.11: Απόδοση δικτύου τύπου U-Net σε σχέση με τα δεδομένα εισόδου και το μοντέλο που εκπαιδεύτηκε.

Συμπληρωματικά, για λόγους εκτενέστερης κατανόησης και οπτικοποίησης των αποτελεσμάτων κατασκευάστηκε το διάγραμμα του Σχήματος 4.12 στο οποίο συνοψίζεται η απόδοση των τριών μοντέλων που υλοποιήθηκαν καθώς και των αποτελεσμάτων του MD²IP μοντέλου που διατρήθηκε στο προηγούμενο κεφάλαιο. Σαφέστερα, υπολογίζονται οι μετρικές PSNR και SSIM που συγκρίνουν τα αποτελέσματα του δικτύου (U-Net και MD²IP αντίστοιχα) με τις πραγματικές πολυφασματικές εικόνες (GT), προκειμένου να γίνει σαφές κατά πόσο τα παράγωγα των μοντέλων προσεγγίζουν τα πραγματικά δεδομένα. Γίνεται, επομένως, εύκολο αντιληπτό ότι το Μοντέλο 1 θα είναι αυτό στο οποίο θα στηριχθεί η παρούσα ενότητα καθώς αποδίδει καλύτερα από το 3^ο Μοντέλο αφού σημειώνει $PSNR_1=34,96$ έναντι της $PSNR_3=33,63$ και $SSIM_1=0,94$ έναντι της $SSIM_3=0,92$.

Όσον αφορά το 2^ο Μοντέλο, όπως έχει προαναφερθεί κατασκευάστηκε προκειμένου να συγκριθεί με το 1^ο, δεδομένου ότι θα σημείωνε καλύτερη επίδοση. Εντούτοις κάτι τέτοιο δε συνέβη όπως έχει ήδη αναφερθεί, σημειώνοντας τιμές αμφότερα για τις μετρικές υποδεέστερες του 1^{ου} και 3^{ου} μοντέλου και πιο λεπτομερώς ισχύει ότι $PSNR_2=33,41$ και $SSIM_2=0,91$, πιθανώς διότι η κατανομή των δεδομένων εκπαίδευσης δεν είναι τόσο κατανοητή από το μοντέλο και δεν μπορεί να μιμηθεί τις υψηλές συχνότητες των GT (που έχουν περισσότερες λεπτομέρειες από αυτές του MD²IP), κάτι που λογικά θα επιλυόταν αν υπήρχε μεγαλύτερο σύνολο δεδομένων. Κλείνοντας, η απόδοση του MD²IP δικτύου όπως προβάλλεται και από το Σχήμα 4.11 σημειώνει τις μεγαλύτερες τιμές κατά την ποσοτική αξιολόγηση με $PSNR=40,42$ και $SSIM=0,98$, κάτι που προβλεπόταν εξ αρχής μιας και πρόκειται για την επίδοση του μοντέλου στο οποίο στηρίχθηκε το δίκτυο τύπου U-Net που χρησιμοποιήσαμε για την εκπαίδευση του.



Σχήμα 4.12: Απόδοση Μοντέλων (M1, M2, M3) και MD²IP επί του συνόλου αξιολόγησης, ως προς τις μετρικές PSNR&SSIM.

Παράλληλα, θεωρήθηκε ότι θα είχε ενδιαφέρον να εξεταστεί η επίδοση κάθε εικόνας ξεχωριστά, που συνθέτουν το σύνολο αξιολόγησης, σε σχέση με τα πραγματικά πολυφασματικά δεδομένα βάση των μετρικών PSNR και SSIM. Για αυτό το λόγο κατασκευάστηκαν οι Πίνακες 4.7 και 4.8 όπου αντιπαραβάλλονται οι τιμές των μετρικών για τρία διαφορετικά μοντέλα-στάδια επεξεργασίας από όπου έχουν διέλθει τα δεδομένα. Με άλλα λόγια, προβάλλεται για κάθε εικόνα η τιμή της μετρικής στο

στάδιο όπου έχει εκτελεστεί διγραμμική παρεμβολή και έχουν παραχθεί τα δεδομένα εισόδου των μοντέλων, στο στάδιο όπου έχουν εξαχθεί τα αποτελέσματα του μοντέλου μας (τύπου U-Net) και στα παράγωγα του MD²IP δικτύου. Από τους Πίνακες διαπιστώνεται αρχικά ότι το validation set συντίθεται από εικόνες όπου η πλειονότητα τους προέρχεται από το σύνολο δεδομένων M15 συνιστώντας 10/20 εικόνες. Κάτι τέτοιο δεν είναι καθόλου τυχαίο καθώς οι εικόνες αυτές ήταν εκείνες που κατακερματιστήκαν περισσότερο σε μικρότερα τμήματα (patches) και συνιστούσαν πάνω από το μισό των συνολικών δεδομένων, συνιστώντας τα 120/214 τμήματα εικόνων. Έπειτα, ακολουθούν 4/20 εικόνες που επιλέχθηκαν τυχαία από τα TT31 και CAVE και μόλις 2/20 από το TT59. Επομένως, πράγματι το σύνολο αξιολόγησης προέκυψε από τυχαία επιλογή από το σύνολο των δεδομένων.

Πιο συγκεκριμένα και από τους δύο Πίνακες γίνεται αντιληπτό, ότι το MD²IP (Deep Image Prior) συστηματικά υπερτερεί των άλλων δύο μεθόδων σε όλα τα τμήματα εικόνων, όπως και η προτεινόμενη μέθοδος που χρησιμοποιεί ένα μοντέλο Attention R2U-Net υπερτερεί της διγραμμικής παρεμβολής, κάτι που είχε παρατηρηθεί και επί του συνόλου δεδομένων αξιολόγησης στο Σχήμα 4.11. Επίσης, οι εικόνες που εμφανίζουν τις μεγαλύτερες τιμές και συνεπώς τις ισχυρότερες επιδόσεις είναι κατά βάση αυτές που προέρχονται από τα συνολικά δεδομένων CAVE και ιδίως από το M15, όπου η ποιότητα των εικόνων όπως είχε σχολιαστεί σε προηγούμενο υποκεφάλαιο δεν είναι τόσο καλή και υπάρχει θόρυβος, συνεπώς είναι πιο εύκολο να αναπαρασταθούν από το εκάστοτε δίκτυο. Χαρακτηριστικό τέτοιο παράδειγμα αποτελεί η εικόνα «Bom Jesus Bush _patch_1» από το M15, όπου η τιμή της PSNR βελτιώνεται από 36,48 (Διγραμμική Παρεμβολή) σε 44,96 (Προτεινόμενη Μέθοδος Attention R2U-Net) και 49,13 (MD²IP). Αντίστοιχα, για την μετρική SSIM είναι ορατή η πολύ καλή επίδοση αφού για την ίδια εικόνα ισχύει 0,96(Bilinear), 0,97(Attention R2U-Net), 0,99 (MD²IP). Ακόμα, παρατηρείται ότι εικόνες που επιλέχθηκαν από το TT59 παρουσιάζουν σημαντική βελτίωση και στις δύο μετρικές από τη διγραμμική παρεμβολή στη προτεινόμενη μέθοδο και ακόμη περισσότερο στο MD²IP. Λόγου χάριν, η εικόνα «Chart _patch_0» όπου στην PSNR έχει τιμή 19,64 (Bilinear), 26,42(Attention R2U-Net), 36,04 (MD²IP) και στην SSIM 0,77, 0,89 και 0,96 αντίστοιχα.

Σχετικά, με την SSIM είναι σκόπιμο να σημειωθεί ότι η μέση βελτίωση από τη διγραμμική παρεμβολή στην προτεινόμενη μέθοδο είναι 0,06 SSIM μονάδες και από το Attention R2U-Net στο MD²IP περίπου 0,04. Αυτό ερμηνεύεται με τον ίδιο τρόπο που εξηγήθηκε παραπάνω βάση της PSNR μετρικής, δηλαδή ότι το μοντέλο καταφέρνει να εκπαιδευτεί πολύ καλά βελτιώνοντας την επίδοση του αισθητά σε σχέση με τα δεδομένα εισόδου του και έχοντας περιθώριο βελτίωσης μέχρι να προσεγγίσει

περισσότερο την απόδοση του μοντέλου πάνω στο οποίο εκπαιδεύτηκε. Εν κατακλείδι, η ανάλυση των αποδόσεων των εικόνων αποτέλεσε μία ευκαιρία για εκτενέστερη εμβάθυνση και παρατήρηση της συμβολής των εικόνων τόσο μεμονωμένα όσο και σε επίπεδο dataset στις μεθόδους που εξετάστηκαν.

PSNR Metric				
IMAGE	DATASET	BILINEAR (MD ² IP)	OUR (Attention R2U-Net)	MD ² IP
Butterfly2_patch_0	TT59	26,40	33,27	40,35
Chart_patch_0	TT59	19,64	26,42	36,04
Bom Jesus Bush_patch_1	M15	36,48	44,46	49,13
Bom Jesus Marigolds_patch_0	M15	32,58	38,88	43,84
Bom Jesus Marigolds_patch_3	M15	28,10	34,40	40,29
Bom Jesus Red flower_patch_1	M15	31,62	37,15	42,44
Butterfly2_patch_0	TT31	23,24	28,17	32,71
Butterfly5_patch_0	TT31	28,95	35,88	44,94
Egyptian statue_patch_0	CAVE	33,37	38,77	44,21
face_patch_0	CAVE	31,91	37,92	43,61
jelly beans_patch_0	CAVE	23,84	30,36	37,08
Souto Wood Pile_patch_1	M15	26,37	35,30	40,16
sponges_patch_0	CAVE	27,71	33,27	43,61
Tape_patch_0	TT31	26,16	29,16	31,53
Tibaes Corridor_patch_3	M15	29,74	35,07	38,87
Tibaes Garden Entrance_patch_1	M15	32,97	37,71	39,73
Tibaes Garden_patch_0	M15	31,53	36,95	39,97
T-shirts_patch_0	TT31	20,53	26,40	31,77
Yellow Rose_patch_1	M15	36,22	41,09	45,41
Yellow Rose_patch_2	M15	32,96	38,48	42,62
AVERAGE		29.02	34.96	40.42

Πίνακας 4.7: Σύγκριση σε διαφορετικά στάδια επεξεργασίας των δεδομένων (validation set), σύμφωνα με την μετρική PSNR.

SSIM Metric				
IMAGE	DATASET	BILINEAR (MD ² IP)	OUR (Attention R2U-Net)	MD ² IP
Butterfly2_patch_0	TT59	0,89	0,95	0,98
Chart_patch_0	TT59	0,77	0,89	0,96
Bom Jesus Bush_patch_1	M15	0,96	0,97	0,99
Bom Jesus Marigolds_patch_0	M15	0,92	0,96	0,98
Bom Jesus Marigolds_patch_3	M15	0,83	0,95	0,98
Bom Jesus Red flower_patch_1	M15	0,93	0,96	0,99
Butterfly2_patch_0	TT31	0,80	0,87	0,94
Butterfly5_patch_0	TT31	0,92	0,96	0,99
Egyptian statue_patch_0	CAVE	0,95	0,96	0,99
face_patch_0	CAVE	0,95	0,97	0,99
jelly beans_patch_0	CAVE	0,81	0,92	0,98
Souto Wood Pile_patch_1	M15	0,83	0,95	0,98

sponges_patch_0	CAVE	0,94	0,97	0,99
Tape_patch_0	TT31	0,83	0,88	0,92
Tibaes Corridor_patch_3	M15	0,91	0,97	0,98
Tibaes Garden Entrance_patch_1	M15	0,95	0,98	0,98
Tibaes Garden_patch_0	M15	0,93	0,97	0,98
T-shirts_patch_0	TT31	0,60	0,82	0,93
Yellow Rose_patch_1	M15	0,95	0,98	0,99
Yellow Rose_patch_2	M15	0,96	0,99	0,99
AVERAGE		0.88	0.94	0.98

Πίνακας 4.8: Σύγκριση σε διαφορετικά στάδια επεξεργασίας των δεδομένων (validation set), σύμφωνα με την μετρική SSIM.

Συνοψίζοντας, οι μετρικές των μοντέλων που υλοποιήθηκαν κατέστησαν σαφές τα προτερήματα και τις στερήσεις της αρχιτεκτονικής του δικτύου που χρησιμοποιήσαμε. Σχετικά με το 1^ο Μοντέλο, που αποτελεί την προτεινόμενη επιλογή για την μεθοδολογία μας, αποδείχθηκε ότι σημείωσε εξαιρετική ακρίβεια στην εργασία της απομοσαϊκοποίησης (demosaiicing), επιτυγχάνοντας αντικειμενικά υψηλές τιμές ιδίως εφόσον συγκριθεί με την απόδοση των δεδομένων εισόδου και του μοντέλου MD²IP, που αποτέλεσε το μοντέλο στο οποίο στηρίχθηκε για την εκπαίδευση του. Μάλιστα είναι κρίσιμο να τονιστεί ότι παρά το μικρό όγκο δεδομένων που ανακτήθηκαν απέδωσε πολύ καλά βάση των μετρικών σύγκρισης ομοιότητας στο σύνολο των δεδομένων αξιολόγησης καταγράφοντας PSNR=34,96 και SSIM=0,94 σε σχέση με τα πραγματικά δεδομένα. Καταληκτικά, σύμφωνα με τα αποτελέσματα της ποσοτικής αξιολόγησης το βέλτιστο μοντέλο που υλοποιήθηκε κρίθηκε ως αρκετά επιτυχημένο με μελλοντικές βελτιώσεις που μπορούν να επιτευχθούν και θα αναλυθούν στο επόμενο κεφάλαιο (5^ο).

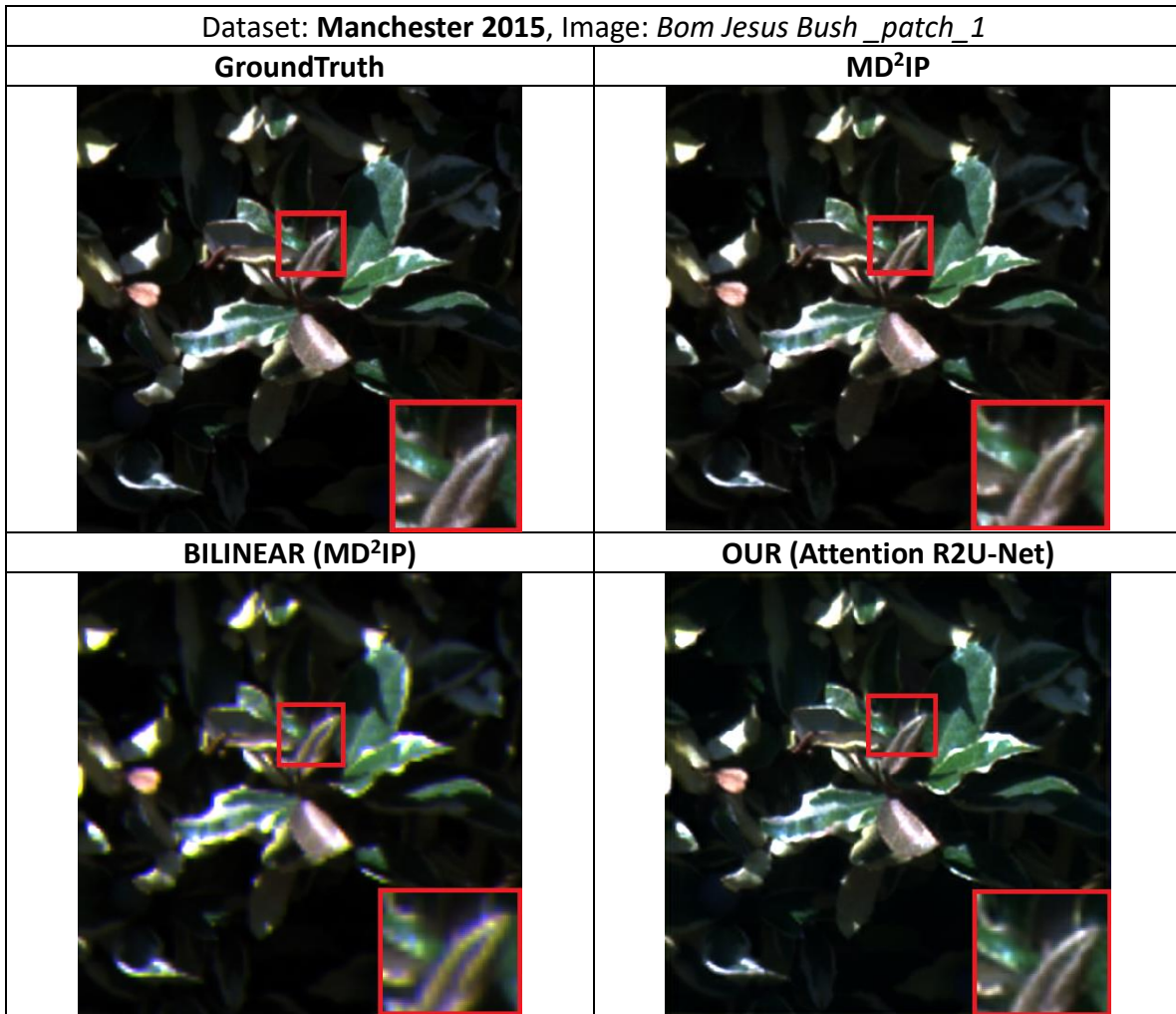
4.4.2. Ποιοτική Αξιολόγηση

Πέραν της ποσοτικής αξιολόγησης κρίνεται απαραίτητη και η αξιολόγηση της προτεινόμενης μεθοδολογίας και με ποιοτικά κριτήρια. Για αυτόν τον λόγο θα ακολουθήσουν μερικά παραδείγματα όπου στόχο έχουν να βοηθήσουν στην καλύτερη αντίληψη της απόδοσης του μοντέλου καθ' όλα τα στάδια επεξεργασίας στη ροή εργασιών που υλοποιήθηκε και προτείνεται. Αναπόσπαστο κομμάτι αυτής της διαδικασίας είναι φυσικά και ο εντοπισμός τυχόν σφαλμάτων-artifacts, η έκταση που καταλαμβάνουν καθώς και η σημαντικότητα αυτών. Σε αυτή την κατεύθυνση, θα δημιουργηθούν μερικές τετράδες εικόνων όπου θα γίνεται εστίαση σε κάποιο παράθυρο για εκτενέστερη ανάλυση διαφορών ή επισήμανσης κάποιων

χαρακτηριστικών που παρουσίασαν ενδιαφέρον. Στις τετράδες αυτές θα αναπαριστώνται (για μία εικόνα από το σύνολο αξιολόγησης) σε έγχρωμο σύνθετο RGB: (15, 7, 1) η πραγματική πολυφασματική εικόνα (**GroundTruth**), το output του MD²IP μοντέλου (**MD²IP**), όπως προέκυψε από το 3^ο Κεφάλαιο, το παράγωγο της διγραμμικής παρεμβολής επί των MD²IP δεδομένων (**BILINEAR (MD²IP)**) και η εικόνα που έχει προκύψει από την προτεινόμενη μεθοδολογία με τη χρήση ενός δικτύου τύπου U-Net (**Attention R2U-Net**).

Γενικά, λόγω της έλλειψης φασματικής πληροφορίας στο ορατό φάσμα είναι επόμενο τα έγχρωμα σύνθετα που θα κατασκευαστούν να μην μπορούν να αναπαραστήσουν πλήρως τη πραγματική πληροφορία από πλευράς φωτισμού και κορεσμού χρωμάτων με τρόπο που αντιλαμβάνεται το ανθρώπινο μάτι. Όσον αφορά τις αντιπαραβαλλόμενες εικόνες σε όλες είναι ευκρινές και κατανοητό το προς απεικόνιση αντικείμενο και είναι δεδομένο ότι η GroundTruth είναι η πιο ευκρινής εικόνα οπότε η σύγκριση έχει νόημα να γίνει σε σχέση με αυτήν. Πράγματι, όπως έγινε εμφανές και από τις μετρικές κατά την ποσοτική αξιολόγηση την ανάλυση των πραγματικών δεδομένων ακολουθεί η εικόνα - παράγωγο του MD²IP, έπεται η εικόνα από το Attention R2U-Net και τέλος αυτή της διγραμμικής παρεμβολής.

Πιο συγκεκριμένα, στο Σχήμα 4.13 που ακολουθεί οπτικοποιήθηκε η πολυφασματική εικόνα «Bom Jesus Bush_patch_1» (που αντιστοιχεί στο σύνολο δεδομένων M15) στα 4 στάδια που προαναφέρθηκαν. Στην προκειμένη περίπτωση παρότι τμήμα της εικόνας δεν είναι ευκρινές είναι εφικτό να ληφθούν κάποιες παρατηρήσεις που μπορούν να γενικευθούν. Εδώ θα εστιάσουμε στο βαθμό προσέγγισης και τις αποκλίσεις της προτεινόμενης μεθόδου με το MD²IP, αφού τα χαρακτηριστικά αυτού προσπαθεί το μοντέλο να μιμηθεί. Καταρχάς, υπάρχει χρωματική ομοιομορφία και ευρύτερα δεν παρατηρούνται ιδιαίτερες μεταβολές παρά μόνο μία ήπια εξομάλυνση στο σύνολο της εικόνας που τη καθιστά μερικώς πιο θολή. Ακόμα, το μόνο που μπόρεσε να εντοπιστεί είναι η δυσκολία του μοντέλου τύπου U-Net να διατηρήσει τις ακμές, κάτι που επισημαίνεται με το κόκκινο εστιασμένο παράθυρο σε όλες τις εικόνες του Σχήματος. Η διαφοροποίηση αυτή μπορεί να ερμηνευτεί από την ομαλοποίηση που επιτελείται. Οπότε, η παράγωγη εικόνα από την προτεινόμενη μεθοδολογία μπόρεσε να μιμηθεί εξαιρετικά τα χαρακτηριστικά των δεδομένων εκπαίδευσης που αποτελούν παράγωγα του MD²IP και των σχέσεων που τα συνδέουν, άρα επιτελούν αποτελεσματικά τη διαδικασία του demosaicing.

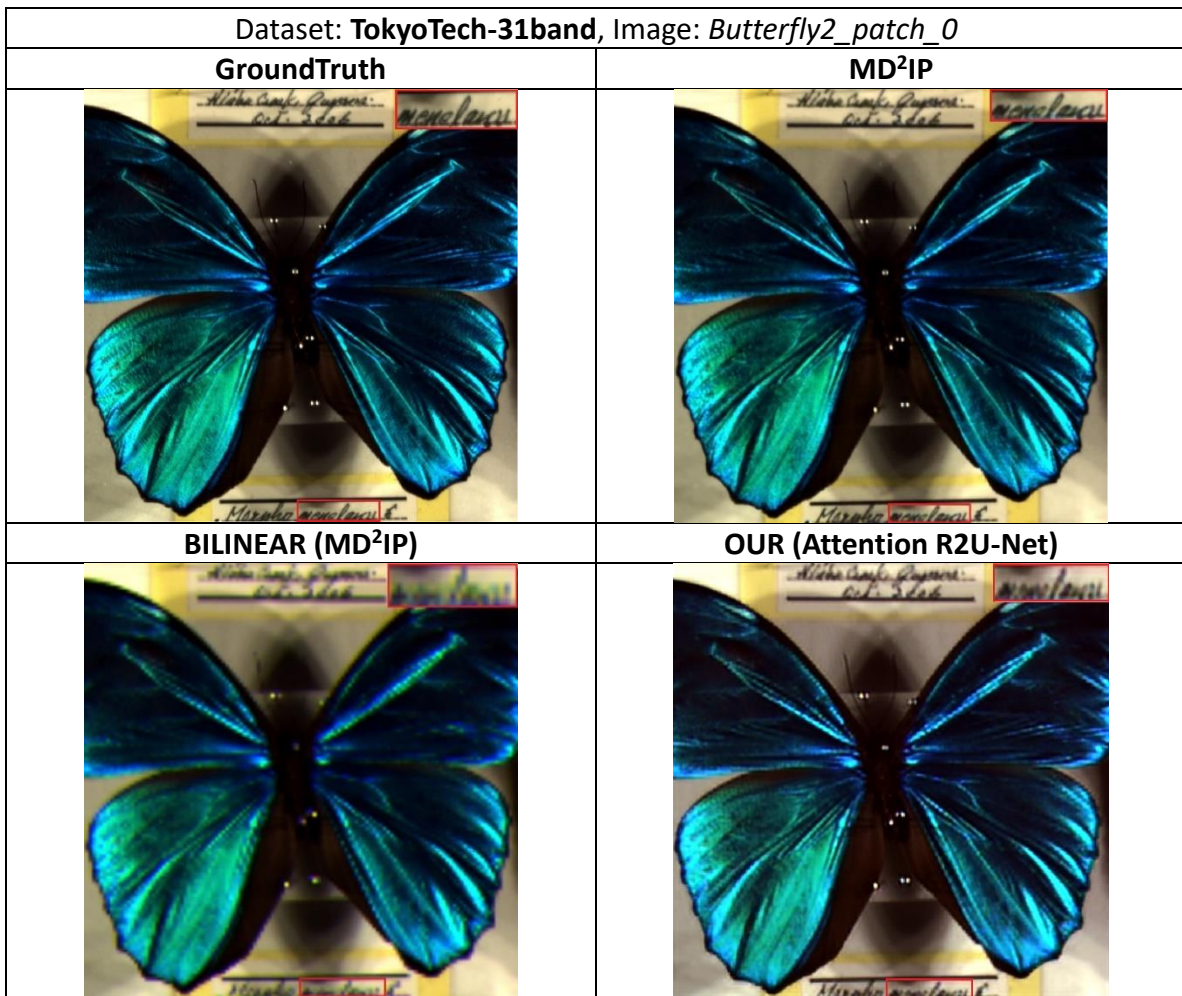


Σχήμα 4.13: Αντιπαραβολή πολυφασματικής εικόνας «Bom Jesus Bush _patch_1» σε RGB: (15, 7, 1) στα διαφορετικά στάδια επεξεργασίας του προτεινόμενου framework.

Αντίστοιχα, δημιουργήθηκε το Σχήμα 4.14 που προβάλλεται η εικόνα «Butterfly2_patch_0» από το σύνολο δεδομένων TT31. Η εικόνα αυτή μπορεί να χαρακτηριστεί πιο απαιτητική ερχόμενη από ένα dataset με μεγαλύτερη ευκρίνεια στις εικόνες. Εν τούτοις, σημειώνεται πολύ καλή αναπαράσταση της πολυφασματικής εικόνας από το Attention R2U-Net. Μάλιστα παρατηρείται ότι διατηρούνται σε πολύ καλό βαθμό τα γραμμικά στοιχεία της εικόνας -αν και λόγω ότι η προτεινόμενη μέθοδος δέχεται σαν δεδομένα εκπαίδευσης τις παρεμβαλλόμενες εικόνες, κοινώς εικόνες που έχουν συγκρατήσει μόνο τις χαμηλές συχνότητες- είναι επόμενο πολύ λεπτές γραμμές να ασθενούν και να εξομαλύνονται γίνοντας μερικές φορές μη διακριτές από το background. Μία άλλη ερμηνεία αυτού του τεχνουργήματος είναι ότι η συγκεκριμένη εικόνα περιέχει πιο σύνθετη πληροφορία που είναι δύσκολο να μιμηθεί. Το φαινόμενο αυτό είναι εμφανές πέρα από την πινακίδα στο κάτω μέρος της εικόνας με τη γραμματοσειρά που επισημαίνεται και σε κάποια μέρη των φτερών. Βέβαια, είναι

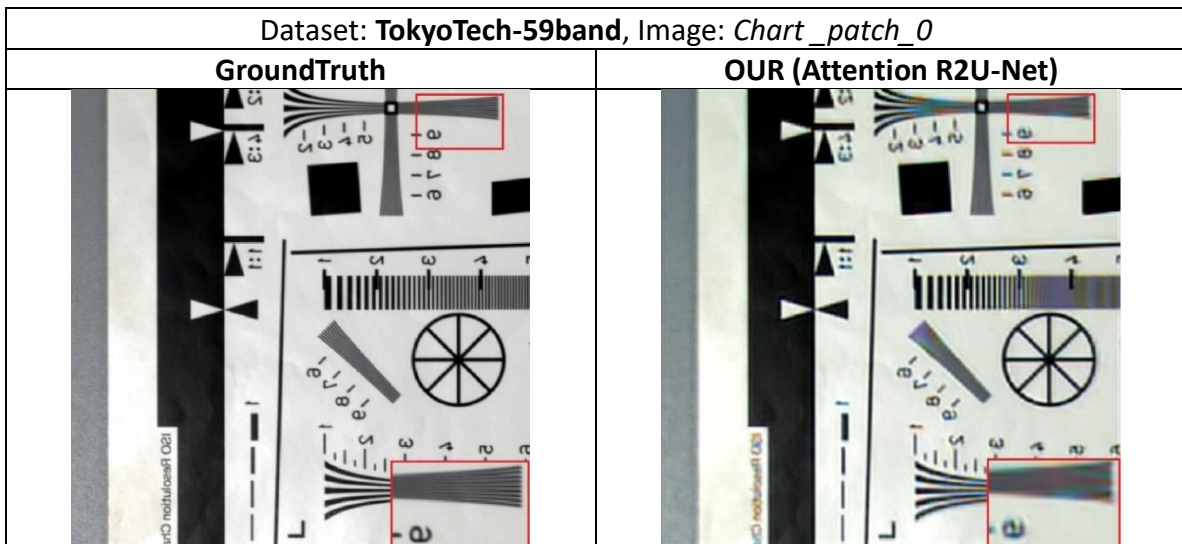
σκόπιμο να αναφερθεί ότι ήδη από την εικόνα του MD²IP παρατηρείται μία μικρή αλλοίωση των γραφικών χαρακτήρων και μία χρωματική διαφοροποίηση. Όντας τα παράγωγα του MD²IP· επιπλέον labels για το δίκτυο μας είναι λογικό το σφάλμα αυτό να μεταφέρεται στο U-Net ως ένα βαθμό και σε πολύ μεγαλύτερη κλίμακα στη παρεμβολή.

Τέλος, αυτό που μπορεί να διακριθεί είναι ότι τόσο στις εικόνες που εξήχθησαν από τη διγραμμική όσο και σε αυτές από το δίκτυο μας· παρότι συνολικά αναπαρέστησαν πολύ καλά τα χρώματα των GT & MD²IP αναγνωρίζονται κάποιες χρωματικές αποκλίσεις. Πιο συγκεκριμένα, φαίνεται να εξασθενεί η παρουσία του κίτρινου χρώματος (τόσο σε background όσο και στα φτερά που είναι λιγότερο πράσινα). Κάτι τέτοιο είναι πολύ λογικό αφού οι μεν (BIL.) αποτελούν δεδομένα εισόδου για τις δε (Attention R2U-Net), συνεπώς επηρεάζουν σημαντικά την εκπαιδευτική διαδικασία. Παρά τις μικρές αυτές διαφοροποιήσεις που εντοπίστηκαν κρίνεται ότι συνολικά η απόδοση του μοντέλου είναι πολύ καλή.



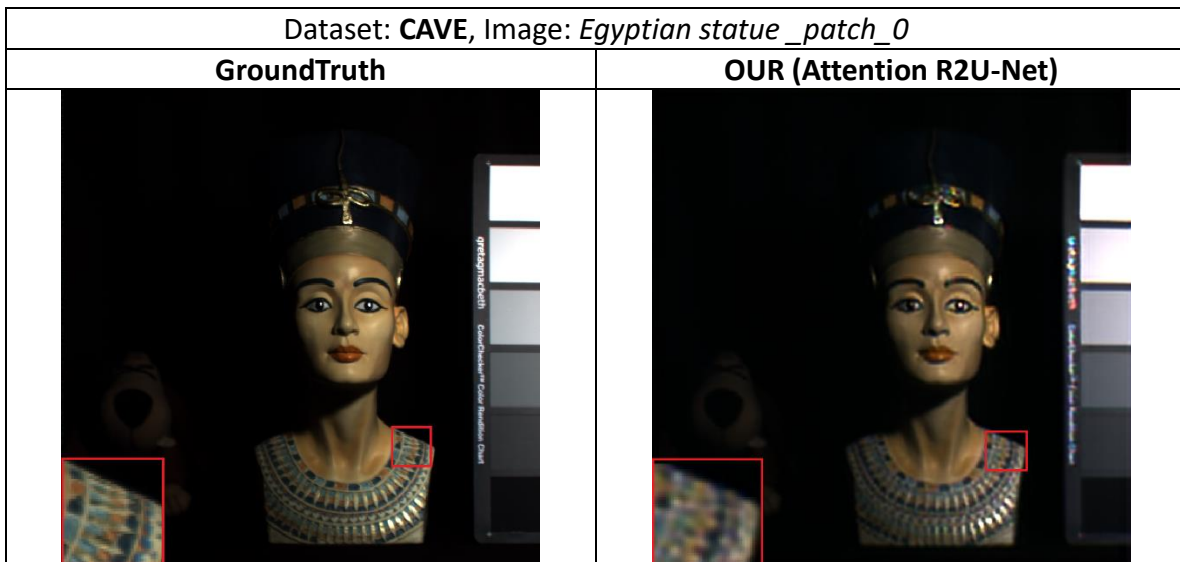
Σχήμα 4.14: Αντιπαραβολή πολυφασματικής εικόνας «Butterfly2_patch_0» σε RGB: (15, 7, 1) στα διαφορετικά στάδια επεξεργασίας του προτεινόμενου framework.

Παρακάτω, ακολουθούν μερικά ζευγάρια από εικόνες όπου αυτή τη φορά αντιπαραβάλλονται μόνο οι πραγματικές πολυφασματικές με αυτές που παράγει το μοντέλο μας που στηρίζεται σε Attention R2U-Net. Με αυτόν τον τρόπο μπορεί να γίνει άμεση σύγκριση ανάμεσα στα δεδομένα που ιδανικά θέλουμε να προσεγγίσουμε αλλά στην πραγματικότητα δεν είναι διαθέσιμα λόγω του τρόπου καταγραφής της πολυφασματικής πληροφορίας από τις single shot cameras, που διαθέτουν ένα σύστημα πολυφασματικών αισθητήρων επίπεδης διάταξης φίλτρων (MSFA). Το πρώτο ζευγάρι που παρουσιάζει ενδιαφέρον είναι αυτό του Σχήματος 4.15, όπου προβάλλεται η εικόνα «Chart_patch_0» από το σύνολο δεδομένων TT59. Παρότι, συνολικά παρατηρείται συμφωνία ως προς τα δομικά χαρακτηριστικά των εικόνων και το παράγωγο μας φαίνεται να έχει μιμηθεί πολύ καλά την πληροφορία της πραγματικής παράλληλα είναι ευκρινείς κάποιες αστοχίες του μοντέλου. Πιο λεπτομερώς, όπως επισημαίνεται και με το κόκκινο εστιασμένο παράθυρο γίνονται αντιληπτά κάποια συστηματικά τεχνουργήματα (artifacts) που εντοπίζονται σε λεπτά γραμμικά στοιχεία ιδίως όταν αυτά βρίσκονται σε κοντινή απόσταση. Αυτό όπως προαναφέρθηκε μπορεί να αποδοθεί στο ότι τα συγκεκριμένα δεδομένα είναι πιο απαιτητικά με περισσότερες λεπτομέρειες. Επιπλέον σε αυτού του είδους την «αλλοίωση» συμβάλει το γεγονός ότι γενικότερα το μοντέλο έχει την ιδιότητα να εξομαλύνει τα δεδομένα. Τέλος, είναι ορατή η μικρή χρωματική απόκλιση και πιο συγκεκριμένα η διαφορά στον τόνο του λευκού που υπάρχει ως υπόβαθρο.



Σχήμα 4.15: Αντιπαραβολή πολυφασματικής εικόνας «Chart_patch_0» σε RGB: (15, 7, 1) στα διαφορετικά στάδια επεξεργασίας του προτεινόμενου framework.

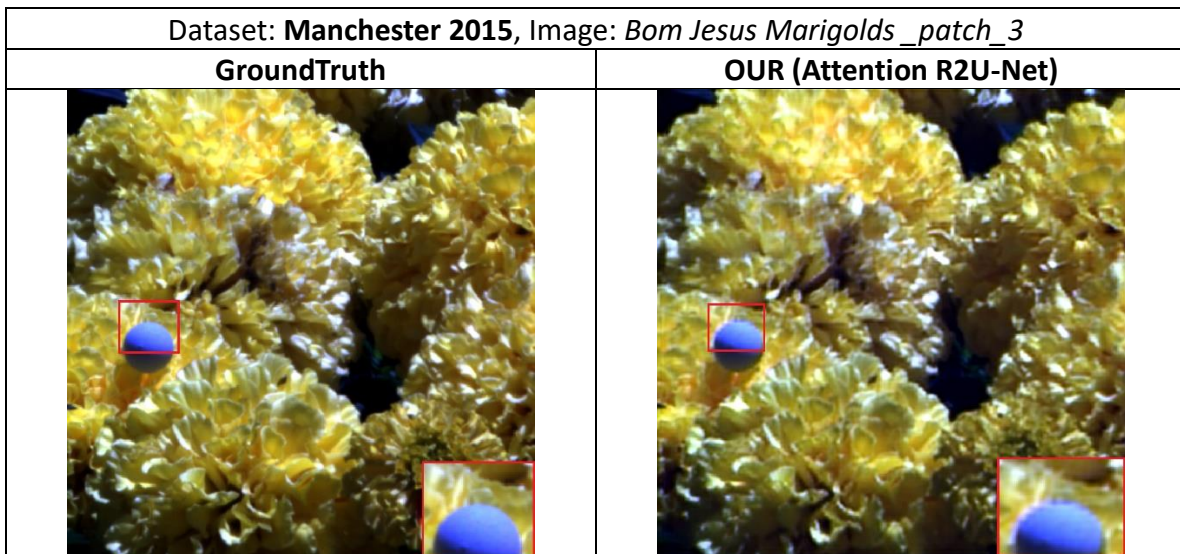
Ακολουθεί, το Σχήμα 4.16 με την εικόνα «Egyptian statue_patch_0» από το σύνολο δεδομένων CAVE. Αν και μέρος της πληροφορίας και εδώ δεν είναι ορατό μπορούν να γίνουν κάποιες παρατηρήσεις σχετικά με την απόδοση του μοντέλου. Αναλυτικότερα, φαίνεται και στις δύο εικόνες με ευκρίνεια το αιγυπτιακό άγαλμα και λεπτομέρειες αυτού καθώς και το colorchart, που υπάρχει στο πλάι. Μόνο εφόσον εστιάσουμε σημαντικά σε μέρη της εικόνας μπορούν να διακριθούν διαφορές και αυτές παρατηρούνται και πάλι σε μέρη όπου υπάρχει σύνθετη πληροφορία. Με άλλα λόγια στο κολάρο της αιγυπτιακής θεότητας όπου υπάρχουν έντονες μεταβολές σε υφή και χρώμα διαπιστώνεται μία δυσκολία διατήρησης του μοτίβου ενώ η ομαλοποίηση που συνεπάγεται το μοντέλο μας επιτείνει τις διαφορές στη συγκεκριμένη περίπτωση. Όπως και προηγουμένως με την εικόνα της πεταλούδας στο Σχήμα 4.14 η γραμματοσειρά που υπάρχει στο colorchart είναι παρομοίως απαιτητική να αναπαρασταθεί λόγω του βαθμού λεπτομέρειας και της πυκνής πληροφορίας που συνεπάγεται στη δεδομένη χωρική ανάλυση. Κατά τα άλλα είναι σημαντικό να επισημανθεί ότι η προτεινόμενη μέθοδος επιτυγχάνει να διατηρήσει τη δομή και σημαντικά χαρακτηριστικά μικρότερης κλίμακας.



Σχήμα 4.16: Αντιπαραβολή πολυφασματικής εικόνας «Egyptian statue_patch_0» σε RGB: (15, 7, 1) στα διαφορετικά στάδια επεξεργασίας του προτεινόμενου framework.

Μία τελευταία περίπτωση, που θα σχολιαστεί σε αυτήν την ενότητα της ποιοτικής αξιολόγησης είναι αυτή που βρίσκει αφορμή από την εικόνα «Bom Jesus Marigolds_patch_3» του συνόλου δεδομένων M15. Στο Σχήμα 4.17 που παρατίθεται οι δύο εικόνες εκ πρώτης όψεως δεν φαίνεται να έχουν διαφορές εν γένει. Αυτό είναι λογικό να αποδοθεί σε μεγάλο βαθμό στην ποικιλομορφία του άνθους που

αναπαριστά, εννοώντας τις πτυχώσεις που παρουσιάζει και φυσικά στις συνθήκες ηλιασμού που δημιουργούν σκιάσεις και περιπλέκουν επιπλέον την εικόνα. Ωστόσο, αν γίνει μία πιο λεπτομερής παρατήρηση υπό κλίμακα θα διακριθεί και πάλι μία αίσθηση θολότητας στα δεδομένα, κάτι που δεν επηρεάζει όπως έχει αποδειχθεί και σε όλα τα παραπάνω αποτελέσματα την συνολική αντίληψη μιας εικόνας. Επιπλέον, αυτό που αξίζει να σημειωθεί είναι το φαινόμενο τύπου «halo» που εμφανίζεται περιμετρικά των αντικειμένων και πιο συγκεκριμένα εντοπίστηκε γύρω από το μπλε μπαλάκι και σε ορισμένα πέταλα των λουλουδιών. Αυτό με άλλα λόγια μπορεί να γίνει κατανοητό σαν ένα λευκό περίγραμμα που περιβάλλει τα αντικείμενα. Πιθανώς αυτό να εμφανίζεται σε περιπτώσεις εικόνων όπου οι χρωματικές μεταβάσεις είναι απότομες και είναι απόρροια της περιπλοκότητας των εικόνων ή ακόμα και της ομαλοποίησης που επιφέρει η δομή του μοντέλου.



Σχήμα 4.17: Αντιπαραβολή πολυφασματικής εικόνας «Bom Jesus Marigolds_patch_3» σε RGB: (15, 7, 1) στα διαφορετικά στάδια επεξεργασίας του προτεινόμενου framework.

Ανακεφαλαιώνοντας, λαμβάνοντας υπόψη την αξιολόγηση με ποσοτικά και ποιοτικά κριτήρια αποδεικνύεται ότι τα πορίσματα από αυτές συνάδουν. Για να γίνει αυτό πιο σαφές, διαπιστώθηκε ότι το προτεινόμενο μοντέλο (Attention R2U-Net) που εκπαιδεύτηκε βάση του MD²IP μπόρεσε να διακρίνει τις σχέσεις που συνδέουν τα δεδομένα εισόδου και να εξαγάγει δεδομένα που αναπαριστούν εξαιρετικά τα παράγωγα του MD²IP και κατ' επέκταση μιμούνται τα πραγματικά πολυφασματικά δεδομένα πολύ καλά. Αυτό το συμπέρασμα επιβεβαιώνεται αφενός εκ πρώτης όψεως στις εικόνες αλλά και εφόσον γίνει εστίαση όπου οι αποκλίσεις δεν είναι ιδιαίτερα εμφανείς. Αυτό επιτείνεται σε εικόνες που σημείωσαν υψηλές τιμές στις μετρικές ήδη από το στάδιο

της διγραμμικής παρεμβολής και κατ' επέκταση ήταν πιο δύσκολο να εντοπιστούν μεταβολές, με χαρακτηριστικό παράδειγμα η εικόνα του Σχήματος 4.13. Μοναδικό ελάττωμα που παρουσιάστηκε και η επίδραση του έγινε ορατή σε αρκετές περιπτώσεις αφορούσε την εξομάλυνση που επιφέρει το μοντέλο. Η ομαλοποίηση αυτή εξηγείται από την ανάλυση των δεδομένων εισόδου, που έχουν συγκρατήσει κατά βάση τις χαμηλές συχνότητες και έχουν απορρίψει μέρος των υψηλών που ευθύνονται για τις λεπτομέρειες και τα λεπτά χαρακτηριστικά μίας εικόνας (πχ. αλλαγές στο χρώμα, τη φωτεινότητα, ακμές, μοτίβα και υφές). Τέλος, η στιβαρότητα του υλοποιημένου μοντέλου μας επιβεβαιώνεται από όλα τα παραδείγματα που παρατέθηκαν λεπτομερώς με εικόνες αλλά και μέσω πινάκων των μετρικών.

ΚΕΦΑΛΑΙΟ 5^ο - Συμπεράσματα - Προτάσεις

Η δομή της παρούσας διπλωματικής επιτάσσει τον συνολικό σχολιασμό του framework (μεθοδολογικού πλαισίου) που ακολουθήθηκε και πιο συγκεκριμένα των δύο μεθοδολογιών που εφαρμόστηκαν. Στο κεφάλαιο αυτό θα γίνει μία σύνοψη των συμπερασμάτων που θα αφορούν τα δύο υλοποιημένα μοντέλα και επιπλέον θα παρατεθούν προτάσεις για την εξέλιξη της παρούσας μεθοδολογίας. Με άλλα λόγια καλούμαστε να αντιμετωπίσουμε την εργασία του MSFA demosaicing που επιλύεται μέσω δύο μοντέλων που χρησιμοποιούνται. Το πρώτο στηρίζεται στο μοντέλο του Deep Image Prior (MD²IP) που αποτελεί μία εργασία βελτιστοποίησης ενώ το δεύτερο αφορά ένα μοντέλο τύπου U-Net που αξιοποιεί χαρακτηριστικά attention και recurrent blocks και προσπαθεί να μιμηθεί την εργασία του πρώτου.

Όσον αφορά το training free μοντέλο του MD²IP αναγνωρίζεται ότι πράγματι επιτελεί με πολύ μεγάλη επιτυχία το στόχο της βελτιστοποίησης ανά εικόνα. Η παραπάνω δήλωση επαληθεύεται από την ποσοτική και ποιοτική αξιολόγηση που πραγματοποιήθηκε. Στην περίπτωση της πρώτης οι μετρικές PSNR και SSIM που χρησιμοποιήθηκαν κατέγραψαν πολύ υψηλές τιμές. Παράλληλα, η ποιοτική αξιολόγηση ήρθε να συμπληρώσει την ποσοτική καθώς χάρη σε αυτήν έγινε σαφές ότι η ποιότητα των ανακατασκευασμένων εικόνων πλήρους ανάλυσης είναι αξιόπαινη με πολλές λεπτομέρειες να συντηρούνται ενώ ακόμα τυχόν τεχνουργήματα δεν επηρεάζουν αισθητά την ανάλυση της εικόνας. Σε αυτό το στάδιο της προτεινόμενης μεθοδολογίας θα μπορούσε να προταθεί σαν μελλοντική εργασία να αναζητηθεί μία μέθοδος τερματισμού των επαναλήψεων που υλοποιούνται από το δίκτυο. Αυτό θα μπορούσε να στηριχθεί στη συνάρτηση κόστους που υπολογίζεται μεταξύ της συνθετικής (MSF \bar{A}) και της πραγματικής (MSF \hat{A}) εικόνας. Έτσι, η βελτιστοποίηση μπορεί να σταματήσει όταν η loss function ελαχιστοποιηθεί ή σταματήσει να φθίνει. Με αυτόν τον τρόπο θα μπορούσε να μειωθεί αισθητά το υπολογιστικό και χρονικό κόστος που συνεπάγεται μια τέτοια εργασία.

Ακολούθως, το προτεινόμενο μοντέλο (Attention R2U-Net) σύμφωνα με αντίστοιχα στάδια αξιολόγησης είναι σαφές ότι εκπαιδεύτηκε πολύ καλά στην αναπαραγωγή πολυφασματικών εικόνων που έχουν προκύψει από μία διαδικασία απομοσαϊκοποίησης (που υλοποίησε πρώτα το MD²IP) σε λήψεις που προέρχονται από snapshot mosaic cameras και ακολουθούν τη δομή ενός 4x4 MSFA. Μάλιστα, το U-Net - με κατάλληλη ρύθμιση των υπερπαραμέτρων- πέτυχε την αναπαραστάση εικόνων

υψηλής ποιότητας σημειώνοντας μικρότερο υπολογιστικό κόστος συγκρινόμενο με το MD²IP.

Βέβαια, είναι απαραίτητο να επισημανθεί ότι παρότι η απόδοση του μοντέλου είναι αξιόλογη σε μελλοντικό επίπεδο μπορούν να γίνουν κάποιες επεμβάσεις που θα τη βελτιώσουν περαιτέρω. Συγκεκριμένα, το γεγονός ότι το σύνολο των δεδομένων ήταν μικρό κατέστησε την εκπαιδευτική διαδικασία περιοριστική καθώς τα δεδομένα εκπαίδευσης ήταν αρκετά περιορισμένα, γεγονός που δυσχέρανε τη γενίκευση του μοντέλου. Αυτό θα μπορούσε να επιλυθεί με εκτενέστερη αναζήτηση επιπλέον dataset που θα αποτελούνταν από πολυφασματικές εικόνες στο υπό μελέτη φάσμα 470-630nm και θα μπορούσαν να προστεθούν στα 4 ήδη ανακτημένα σύνολα δεδομένων. Δεδομένου όμως ότι η δημοσίευση τέτοιων δεδομένων είναι περιορισμένη θα μπορούσε να ληφθεί η πρωτοβουλία σε συνεργασία με την ομάδα του εργαστηρίου τηλεπισκόπησης να δημιουργηθεί ένα νέο dataset με τη χρήση της Silios camera, που υπάρχει σε διαθεσιμότητα. Στη περίπτωση αυτή θα λαμβάνονταν λήψεις σε δείγματα (αντικείμενα και σκηνές) υπό διαφορετικές συνθήκες. Εν τούτοις, μια τέτοια εργασία πέρα από την μετεπεξεργασία θα απαιτούσε επιπλέον φωτογραμμετρικούς ελέγχους, βαθμονόμηση και άλλες τεχνικές προκλήσεις. Όλες οι παραπάνω δράσεις θα συνέβαλαν στην αύξηση της στιβαρότητας του μοντέλου και ευρύτερα στις δυνατότητες γενίκευσης.

Επιπρόσθετα, η απόκλιση των παραγώγων του μοντέλου μας σε σχέση με αυτά του MD²IP σε μεγάλο βαθμό αποδίδεται στην ομαλοποίηση που επέφερε το δίκτυο τύπου U-Net. Αυτό όπως σχολιάστηκε και στο υπεδάφιο της αξιολόγησης θα μπορούσε να επιλυθεί παρέχοντας στο δίκτυο δεδομένα εισόδου υψηλότερης ανάλυσης έναντι αυτών που προέκυψαν από τη διγραμμική παρεμβολή, που είχαν απαλλαχτεί από μεγάλο μέρος των υψηλών συχνοτήτων, κοινώς των λεπτομερειών που εμφάνιζαν οι εικόνες. Μία λύση σε αυτή την κατεύθυνση θα ήταν να επιλεγεί μία πιο σύνθετη διαδικασία παρεμβολής (όπως η τριγραμμική ή η κυματοειδής παρεμβολή) που θα συνεπαγόταν μεγαλύτερο υπολογιστικό κόστος αλλά και καλύτερη ανάλυση δεδομένων.

Επίσης, είναι σκόπιμο να σημειωθεί ότι η θέση του μοντέλου που υλοποιήθηκε στο 4ο Κεφάλαιο (Attention R2U-Net) μπορεί να αντικατασταθεί με κάποιο άλλο δίκτυο, όπως το DPD-Net (που συντίθεται από δύο υπό-δίκτυα [1] και αναπτύσσεται στο υπεδάφιο 2.2.1.) όπου και πάλι θα αξιοποιούνταν τα παράγωγα του MD²IP ως δεδομένα εκπαίδευσης. Έχοντας πει αυτό, θα παρουσίαζε μεγάλο ενδιαφέρον ο πειραματισμός με νεότερες αρχιτεκτονικές που θα είναι πιο απλές και γρήγορες (όπως

το μοντέλο U-Net που δοκιμάστηκε αρχικά) και θα μπορούν να δώσουν λύση σε προβλήματα demosaicing. Χάρη σε αυτό, θα συγκριθούν τα αποτελέσματα από διαφορετικές αρχιτεκτονικές και έτσι θα διακριθεί το μοντέλο εκείνο που συνδυάζει τα βέλτιστα με τη μικρότερη δυνατή υπολογιστική ισχύ. Παρόλα αυτά, τόσο η ποιοτική όσο και η ποσοτική αξιολόγηση του προτεινόμενου μεθοδολογικού πλαισίου -με τις δύο υλοποιήσεις μοντέλων- αναδεικνύουν τη συνολική αποτελεσματικότητα της μεθοδολογίας που αναπτύχθηκε για την ανακατασκευή εικόνων που ακολουθούν τη δομή μίας MSFA. Εν κατακλείδι, παρά τις προτάσεις για περαιτέρω βελτιώσεις που μπορούν να επιτευχθούν θεωρείται ότι το προτεινόμενο framework προσφέρει μία καινοτόμα λύση με αξιόλογη επίδοση στον τομέα της επεξεργασίας εικόνων και προσφέρει ευρείς δυνατότητες εφαρμογής σε ποικίλα επιστημονικά πεδία.

Βιβλιογραφικές αναφορές

- [1] Liu, S., Zhang, Y., Chen, J., Lim, K. P., & Rahardja, S. (2022). A deep joint network for multispectral demosaicking based on pseudo-panchromatic images. *IEEE Journal of Selected Topics in Signal Processing*, 16(4), 622-635.
- [2] Ulyanov, D., Vedaldi, A., & Lempitsky, V. (2018). Deep image prior. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 9446-9454).
- [3] Spanos S, Ntouskos V, Karantzalos K (2024) Multispectral Demosaicing with Deep Image Priors. *International I Geoscience and Remote Sensing Symposium*.
- [4] Yasuma, F., Mitsunaga, T., Iso, D., & Nayar, S. K. (2010). Generalized assorted pixel camera: postcapture control of resolution, dynamic range, and spectrum. *IEEE transactions on image processing*, 19(9), 2241-2253.
- [5] Monno, Y., Kikuchi, S., Tanaka, M., & Okutomi, M. (2015). A practical one-shot multispectral imaging system using a single image sensor. *IEEE Transactions on Image Processing*, 24(10), 3048-3059.
- [6] Monno, Y., Teranaka, H., Yoshizaki, K., Tanaka, M., & Okutomi, M. (2018). Single-sensor RGB-NIR imaging: High-quality system design and prototype implementation. *IEEE Sensors Journal*, 19(2), 497-507.
- [7] Nascimento, S. M., Amano, K., & Foster, D. H. (2016). Spatial distributions of local illumination color in natural scenes. *Vision research*, 120, 39-44.
- [8] Nasrollahi, K., & Moeslund, T. B. (2014). Super-resolution: a comprehensive survey. *Machine vision and applications*, 25, 1423-1468. <https://doi.org/10.1007/s00138-014-0623-4> <https://rdcu.be/dJeSA>
- [9] Nie, S., Gu, L., Zheng, Y., Lam, A., Ono, N., & Sato, I. (2018). Deeply learned filter response functions for hyperspectral reconstruction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 4767-4776).
- [10] https://channelsystems.ca/sites/default/files/documents/xispec_hyperspectral_cameras_brochure_2.pdf (xiSpec Brochure)
- [11] <https://www.ximea.com/en/products/xilab-application-specific-oem-custom/hyperspectral-cameras-based-on-usb3-xispec>

- [12] <https://www.ximea.com/support/attachments/5981/SpectroNet-2016-03-Ximea-V02.pdf>
- [13] <https://www.ximea.com/files/xiSpec-whitepaper-hyperspectral-camera.pdf>
(Hyperspectral whitepaper)
- [14] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In Medical image computing and computer-assisted intervention—MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18 (pp. 234-241). Springer International Publishing.
- [15] Alom, M. Z., Hasan, M., Yakopcic, C., Taha, T. M., & Asari, V. K. (2018). Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. arXiv preprint arXiv:1802.06955.
- [16] Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., ... & Rueckert, D. (2018). Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999.
- [17] https://github.com/LeeJunHyun/Image_Segmentation?tab=readme-ov-file#readme
- [18] Trebing, K., Stańczyk, T., & Mehrkanoon, S. (2021). SmaAt-UNet: Precipitation nowcasting using a small attention-UNet architecture. Pattern Recognition Letters, 145, 178-186.
- [19] Islam, M., Vibashan, V. S., Jose, V. J. M., Wijethilake, N., Utkarsh, U., & Ren, H. (2020). Brain tumor segmentation and survival prediction using 3D attention UNet. In Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 5th International Workshop, BrainLes 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 17, 2019, Revised Selected Papers, Part I 5 (pp. 262-272). Springer International Publishing.
- [20] He, N., Fang, L., & Plaza, A. (2020). Hybrid first and second order attention Unet for building segmentation in remote sensing images. Science China Information Sciences, 63, 1-12.
- [21] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).