ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ - ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ

ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ

ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΠΟΛΟΓΙΣΤΩΝ



ΤΟΜΕΑΣ: ΗΛΕΚΤΡΟΝΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

ΕΡΓΑΣΤΗΡΙΟ: ΕΡΓΑΣΤΗΡΙΟ ΕΠΕΞΕΡΓΑΣΙΑΣ ΣΗΜΑΤΩΝ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του φοιτητή του Τμήματος Ηλεκτρολόγων Μηχανικών και Τεχνολογίας Υπολογιστών της Πολυτεχνικής Σχολής του Πανεπιστημίου Πατρών

ΒΑΣΙΛΕΙΟΥ ΤΣΙΓΚΑ ΒΕΛΕΝΤΖΑ ΤΟΥ ΑΝΔΡΕΑ

ΑΡΙΘΜΟΣ ΜΗΤΡΩΟΥ: 227946

Θέμα

Βελτίωση φασματικής ανάλυσης RGB εικόνων

Επιβλέπων

Καθηγητής Κωνσταντίνος Μπερμπερίδης

Πάτρα, Μάιος 2024

Πανεπιστήμιο Πατρών, Τμήμα Ηλεκτρολόγων Μηχανικών και Τεχνολογίας Υπολογιστών
Τσίγκας-Βελέντζας Βασίλειος
© 2024 – Με την επιφύλαξη παντός δικαιώματος Το σύνολο της εργασίας αποτελεί πρωτότυπο έργο, παραχθέν από τον Τσίγκα-Βελέντζα Βασίλειο και δεν παραβιάζει δικαιώματα τρίτων καθ' οποιονδήποτε τρόπο. Αν η εργασία περιέχει υλικό, το οποίο δεν έχει παραχθεί από τον ίδιο, αυτό είναι ευδιάκριτο και αναφέρεται ρητώς εντός του κειμένου της εργασίας ως προϊόν εργασίας τρίτου, σημειώνοντας με παρομοίως σαφή τρόπο τα στοιχεία ταυτοποίησής του, ενώ παράλληλα βεβαιώνει πως στην περίπτωση χρήσης αυτούσιων γραφικών αναπαραστάσεων, εικόνων, γραφημάτων κλπ., έχει λάβει τη χωρίς περιορισμούς άδεια του κατόχου των πνευματικών δικαιωμάτων για την συμπερίληψη και επακόλουθη δημοσίευση του υλικού αυτού.

ΠΙΣΤΟΠΟΙΗΣΗ

Πιστοποιείται ότι η διπλωματική εργασία με θέμα: «Βελτίωση φασματικής ανάλυσης RGB εικόνων»

Του φοιτητή του τμήματος Ηλεκτρολόγων Μηχανικών και Τεχνολογίας Υπολογιστών, της Πολυτεχνικής Σχολής του Πανεπιστημίου Πατρών: Τσίγκα - Βελέντζα Βασίλη με Αριθμό Μητρώου 227946 παρουσιάστηκε δημοσίως στο Τμήμα Ηλεκτρολόγων Μηχανικών και Τεχνολογίας Υπολογιστών του Πανεπιστημίου Πατρών στις 3 Ιουλίου 2024 και εξετάστηκε και εγκρίθηκε από την ακόλουθη Εξεταστική Επιτροπή:

	ΥΠΟΓΡΑΦΕΣ
Κωνσταντίνος Μπερμπερίδης, καθηγητής Πανεπιστημίου Πατρών, Επιβλέπων	
Γεώργιος Μουστακίδης, ομότιμος καθηγητής Πανεπιστημίου Πατρών, Συνεξεταστής	

Πάτρα, Ιούλιος 2024

Περίληψη

Στην παρούσα διπλωματική εργασία προτείνεται μια νέα μέθοδος ανακατασκευής πολυφασματικών εικόνων. Σκοπός της είναι να επιτύχει ικανοποιητικά αποτελέσματα εκτίμησης, έχοντας παράλληλα τις μικρότερες δυνατές χωρικές και χρονικές απαιτήσεις από τους πόρους του συστήματος.

Βασικό χαρακτηριστικό των πολυφασματικών σημάτων που αξιοποιήσαμε για να οδηγηθούμε σε απλά και μικρά μοντέλα, είναι η αραιότητα τους. Με χρήση της PCA απομονώσαμε τις σημαντικότερες λεπτομέρειες των υψηλής διάστασης σημάτων, δημιουργώντας έτσι μια ενδιάμεση συμπιεσμένη περιγραφή και τεμαχίζοντας το πρόβλημα σε δύο αρκετά απλούστερες απεικονίσεις. Για την πρώτη απεικόνιση, από τον RGB χώρο σε αυτόν της PCA, σχεδιάστηκαν δύο αρχιτεκτονικές νευρωνικών δικτύων. Χρησιμοποιήσαμε τις αρχιτεκτονικές των Multilayer Perceptron και Radial Basis Function δικτύων που αποτελούν τις απλούστερες μεθόδους μη γραμμικής απεικόνισης. Για την δεύτερη απεικόνιση, από τον χώρο της PCA στον πολυφασματικό, χρησιμοποιήσαμε τον αντίστροφο μετασχηματισμό της PCA. Η ευκολία αντιστροφής είναι και ο βασικότερος λόγος επιλογής της συγκεκριμένης μεθόδου. Εξετάσαμε επίσης την συμπεριφορά του συστήματος στην περίπτωση που δεν γνωρίζουμε τον αντίστροφο μετασχηματισμό αλλά τον προσεγγίζουμε με ένα ζεύγος συζευγμένων λεξικών, συχνή περίπτωση στις μη γραμμικές μεθόδους μείωσης διάστασης.

Τελικά η προσέγγιση μας ξεπερνά τις παραδοσιακές μεθόδους ανακατασκευής πολυφασματικών εικόνων σε αποτέλεσμα έχοντας παράλληλα πολλαπλάσιες φορές μικρότερο και γρηγορότερο σύστημα.

Λέξεις-κλειδιά: Φασματικές υπογραφές, Νευρωνικά Δίκτυα, Συζευγμένα λεξικά, Μείωση διάστασης.

Abstract

Multispectral analysis of natural scenes can provide much more detailed information than an ordinary RGB camera. The richer information provided has been beneficial to numerous applications, such as smart agriculture, medicine, meteorology etc. In this thesis we present an efficient lightweight method for accurately reconstructing the multispectral information from a single RGB image. Our method is based on the fundamental sparsity of natural signals in order to split the mapping into two simpler sub-problems. Using PCA we can find the most important components of the high dimensional spectral signatures and thus we create a reduced representation. That leads us to a two-step reconstruction approach. At first, we map the RGB signals into the respected reduced representations using a small neural network. We then reconstruct the final multispectral signal using the inverse PCA transform. The simplification provided by this approach leads us to much better reconstruction results than most lightweight methods having also multiple times smaller and faster system.

Keywords: Spectral signatures, Neural networks, Dictionary coupling, Dimensionality reduction.

Ευχαριστίες

Έχοντας πλέον ολοκληρώσει τη διπλωματική μου, θα ήθελα να ευχαριστήσω όλα τα μέλη του εργαστηρίου Ψηφιακής Επεξεργασίας Σημάτων και Τηλεπικοινωνιών του τμήματος Μηχανικών Η/Υ και Πληροφορικής για το εξαιρετικό κλίμα συνεργασίας κατά την εκπόνηση της διπλωματικής μου εργασίας.

Ιδιαιτέρως θα ήθελα να ευχαριστήσω τον καθηγητή μου κ. Κωνσταντίνο Μπερμπερίδη, ο οποίος με καθοδήγησε σε κάθε βήμα της εργασίας, δίνοντάς μου κίνητρα να εργάζομαι πάντα για το καλύτερο δυνατό αποτέλεσμα, για όλες τις ευκαιρίες που μου έδωσε να μάθω νέα πράγματα και να εξελιχθώ, και για την εμπιστοσύνη που μου έδειξε καθ' όλη την πορεία της διπλωματικής παρά την ασυνέχεια που εμφάνισε η ολοκλήρωση της.

Ευχαριστώ στη συνέχεια τον Αλέξανδρο Γκίλλα για την προθυμία του να με βοηθήσει σε κάθε πρόβλημα που συναντούσα, μικρό ή μεγάλο, καθώς για την καθοδήγησή του.

Ένα τεράστιο ευχαριστώ στην μητέρα μου, που με στήριξε στις επιλογές μου, σε όλη την πορεία μου ως φοιτητής και σε κάθε εύκολη ή δύσκολη στιγμή ανεξαρτήτως κόστους. Θα αρκεστώ σε μία πρόταση γιατί αλλιώς θα έπρεπε να γράψω ολόκληρο βιβλίο.

Τέλος, θέλω να ευχαριστήσω όλους τους κοντινούς ανθρώπους που μου στάθηκαν, ιδιαίτερα τον Μηνά, τον Παναγιώτη, τον Γιάννη και τον Γιώργο για όλες τις συζητήσεις, τις συμβουλές και την επιμονή τους όταν οι συνθήκες δεν το επέτρεπαν να σταθώ στα πόδια μου.

Περιεχόμενα

1 Εισαγωγή.....1
2 Θεωρία-Μέθοδος.....4
2.1 Φάσμα ηλεκτρομαγνητικής ακτινοβολίας.....5
2.1.1 Ορατό φως.....6
2.1.2 Πολυφασματική ανάλυση.....7
2.1.3 Ανακατασκευή πολυφασματικών εικόνων από RGB αισθητήρες.....9
2.2 Θεωρία αραιών αναπαραστάσεων.....10
2.2.1 Αραιότητα φυσικών σημάτων.....10
2.2.2 Ασθενώς ορισμένα γραμμικά συστήματα.....10
2.2.3 Κριτήρια κανονικοποίησης.....11
2.2.4 Η ℓ_0 ψευδονόρμα.....11
2.2.5 Pursuit αλγόριθμοι.....12
2.2.6 Ο αλγόριθμος Orthogonal Matching Pursuit.....13
2.2.7 Εκμάθηση λεξικών.....15
2.2.8 Ο αλγόριθμος K-SVD.....15
2.3 Τεχνικές μείωσης διάστασης.....19
2.3.1 Principal Component Analysis.....19
2.4 Μηχανική μάθηση.....22
2.4.1 Μη επιβλεπόμενη μάθηση.....23
2.4.2 Ενισχυτική μάθηση.....23
2.4.3 Επιβλεπόμενη μάθηση.....24
2.4.4 Νευρωνικά δίκτυα.....25
2.4.5 Δομή βιολογικού νευρώνα.....25
2.4.6 Μαθηματική δομή τεχνητού νευρώνα.....26
2.4.7 Αρχιτεκτονικές νευρωνικών δικτύων.....28
2.4.8 Multilayer Perceptron.....28
2.4.9 Εκπαίδευση δικτύων.....29
2.4.10 Ο αλγόριθμος Gradient Descent.....29
2.4.11 Ο αλγόριθμος Backpropagation of Error.....30
2.4.12 Ο αλγόριθμος Adam.....32

2.4.13 Radial Basis Function networks……33
2.4.14 Ο αλγόριθμος Levenberg-Marquardt……34
2.4.15 Συναρτήσεις απώλειας……36
3 Υλοποίηση……37
3.1 Αραιότητα πολυφασματικών σημάτων……38
3.2 Ορισμός προβλήματος……40
3.3 Εξαγωγή δεδομένων……42
3.3.1 Πολυφασματικά δεδομένα……42
3.3.2 Σύνθεση RGB δεδομένων……43
3.3.3 Μείωση διάστασης πολυφασματικών δεδομένων……44
3.4 Απεικόνιση από τον RGB χώρο στον χώρο της PCA……46
3.4.1 Radial Basis Function neural network……46
3.4.2 Multilayer Perceptron……48
3.5 Απεικόνιση από τον χώρο της PCA στον πολυφασματικό χώρο……50
3.5.1 Αντίστροφος μετασχηματισμός της PCA……50
3.5.2 Πρώτη προτεινόμενη μέθοδος επίλυσης……52
3.5.3 Συζευγμένα λεξικά (Dictionary Coupling)……53
3.5.4 Δεύτερη προτεινόμενη μέθοδος επίλυσης……54
3.6 Ανακατασκευή βασισμένη σε patches……56
3.6.1 Αραιότητα πολυφασματικών παραθύρων……56
3.6.2 Απεικονίσεις μεταξύ των χώρων……57
3.6.3 Τρίτη προτεινόμενη μέθοδος επίλυσης……59
4 Αποτελέσματα……60
4.1 Αποτελέσματα εκπαίδευσης……61
4.1.1 Εκπαίδευση Radial Basis Function neural network……61
4.1.2 Εκπαίδευση Multilayer Perceptron……62
4.1.3 Εκπαίδευση συζευγμένων λεξικών……64
4.1.4 Χρόνοι σύγκλισης διαφορετικών αρχιτεκτονικών……65
4.2 Αποτελέσματα ανακατασκευής……67
4.2.1 Μελέτη της επίδρασης του αριθμού των εκτιμώμενων συνιστωσών στο σφάλμα εκτίμησης……67
4.2.2 Μελέτη της επίδρασης της κάμερας στο σφάλμα εκτίμησης……69
4.2.3 Σύγκριση με άλλες σύγχρονες μεθόδους……71
4.2.4 Χρονική και χωρική σύγκριση μεθόδων……72
4.2.5 Οπτικά αποτελέσματα ανακατασκευής ολόκληρων εικόνων……73
5 Συμπεράσματα……82
Κατάλογος σχημάτων……84
Κατάλογος πινάκων……89
Βιβλιογραφία……90

ΚΕΦΑΛΑΙΟ 1

Εισαγωγή

Η αξιοποίηση της ανακλώμενης πληροφορίας από ένα αντικείμενο αποτελεί την βασικότερη λειτουργία της ανθρώπινης όρασης. Παρά το περιορισμένο εύρος του φάσματος που γίνεται αντιληπτό, το ανθρώπινο μάτι κατορθώνει να διαχωρίσει αντικείμενα με ακρίβεια, αξιοποιώντας την φασματική πληροφορία ενός μικρού συνόλου συντελεστών. Η τεχνολογική πρόοδος οδήγησε με τα χρόνια στην ανάπτυξη πολυφασματικών αισθητήρων οι οποίοι ξεπερνούν τον περιορισμό της ανθρώπινης όρασης, αποτυπώνοντας την φασματική πληροφορία που αντανακλάται σε ένα αρκετά μεγαλύτερο εύρος συχνοτήτων. Αυτό το νέο πρίσμα που μας παρέχει η πολυφασματική ανάλυση μας δίνει την δυνατότητα να αποτυπώσουμε με μοναδικό τρόπο τις διάφορες χημικές, γεωμετρικές κλπ. ιδιότητες ενός υλικού δημιουργώντας έτσι μια φασματική υπογραφή(spectral signature) [1]. Οι υπογραφές αυτές προσφέρουν έναν αξιόπιστο τρόπο μελέτης, ανίχνευσης και κατηγοριοποίησης των διαφόρων υλικών και βρίσκουν έτσι εφαρμογή σε πολλούς τομείς όπως στην έξυπνη γεωργία [2], αρχαιολογία [3], ιατρική διάγνωση [4], ανάλυση κειμένου [5], έλεγχο τροφίμων [6], απεικόνιση μεταλλευμάτων στην επιφάνεια της γης [7] κ.α. Το μεγάλο κόστος και μέγεθος όμως μιας πολυφασματικής κάμερας καθιστά απαγορευτική την χρήση της στο μεγαλύτερο κομμάτι των εφαρμογών. Αυτό το πρόβλημα οδήγησε τους ερευνητές στο να εξετάσουν κατά πόσο μπορούμε να εκτιμήσουμε με ακρίβεια την πολυφασματική πληροφορία μιας φυσικής σκηνής χρησιμοποιώντας χαμηλότερης ανάλυσης RGB αισθητήρες. Στο βαθμό που αυτό μπορεί να γίνει ικανοποιητικά, μας δίνεται η δυνατότητα να επεκτείνουμε δραστικά και με πολύ φθηνό τρόπο το πεδίο εφαρμογών της πολυφασματικής ανάλυσης.

Η αύξηση της υπολογιστικής ισχύος σε συνδυασμό με τον μεγάλο όγκο δεδομένων που προκύπτουν οδήγησε τα τελευταία χρόνια στην εμφάνιση πολλών μεθόδων εκπαίδευσης για την επίλυση τέτοιου είδους αντίστροφων προβλημάτων.

Μοντέλα βασισμένα στην μηχανική μάθηση(Machine Learning) καθώς και μέθοδοι βασισμένες στα συζευγμένα λεξικά(Dictionary Coupling) έχουν δώσει αξιόλογα αποτελέσματα στο πρόβλημα της ανακατασκευής πολυφασματικών εικόνων από τις αντίστοιχες RGB. Μια σύνοψη των καλύτερων μεθόδων βρίσκεται στην [8]. Χοντρικά μπορούμε να ταξινομήσουμε τις μεθόδους σε δύο κατηγορίες, στις prior based spectral reconstruction methods που βασίζονται σε συζευγμένα λεξικά και μικρού μεγέθους feedforward νευρωνικά δίκτυα [9, 10, 11]. Το μέγεθος αυτών των μοντέλων είναι αισθητά πιο μικρό γεγονός που τα καθιστά ικανοποιητικά για εφαρμογές με έντονο περιορισμό σε μνήμη και κατανάλωση ενέργειας. Και τις data-driven deep learning methods που βασίζονται σε γνωστές αρχιτεκτονικές μεγάλων νευρωνικών δικτύων(U-Nets, Gan models, Dense networks, Attention networks) όπου το πλήθος των συντελεστών ξεπερνά το εκατομμύριο [12, 13, 14]. Η βασική προσέγγιση όλων αυτών των μεθόδων είναι πως επιχειρούν να ανακατασκευάσουν τα πολυφασματικά pixel άμεσα από τα αντίστοιχα RGB επιλύοντας ένα έντονα ασθενώς ορισμένο πρόβλημα γεγονός που περιορίζει σε έναν βαθμό την απόδοση τους.

Σκοπός της παρούσας εργασίας είναι να εξετάσουμε αν με την αξιοποίηση μιας ενδιάμεσης συμπιεσμένης αναπαράστασης και την τμηματική ανακατασκευή των φασματικών υπογραφών μπορούμε να βελτιώσουμε την τελική συμπεριφορά του συστήματος μας. Πιο συγκεκριμένα με χρήση της PCA συμπιέζουμε την πολυφασματική πληροφορία δημιουργώντας περιγραφές αρκετά μικρότερης διάστασης από την αρχική που όμως διατηρούν το μέγιστο δυνατό ποσοστό πληροφορίας. Αυτό έχει ως αποτέλεσμα η εκτίμηση μας να γίνεται σε δύο βήματα. Πρώτα απεικονίζουμε το RGB περιεχόμενο στις αντίστοιχες PCA συνιστώσες. Το πρόβλημα αυτό έχει αρκετά μικρότερη διαφορά διάστασης από την αρχική απεικόνιση RGB σε Hyperspectral. Και σε δεύτερο χρόνο ανακατασκευάζουμε το τελικό πολυφασματικό pixel από τους αντίστοιχους συντελεστές της PCA, οι οποίοι όμως έχουν σχεδιαστεί ώστε να αποτυπώνουν το αναγκαίο ποσοστό πληροφορίας. Ο τεμαχισμός του αρχικού προβλήματος μας δίνει δύο αρκετά απλούστερα υποπροβλήματα που μπορούμε να επιλύσουμε συνδυάζοντας βασικές αρχιτεκτονικές που έχουν λειτουργήσει ικανοποιητικά και να συνδυάσουμε τα αποτελέσματα τους.

Για την απεικόνιση από τον RGB χώρο στον χώρο της PCA αναπτύχθηκαν δύο αρχιτεκτονικές βασισμένες σε feedforward νευρωνικά δίκτυα. Η μία βασίζεται στην αρχιτεκτονική του Multilayer Perceptron, και χρησιμοποιεί τον αλγόριθμο εκπαίδευσης SGD(Stochastic Gradient Descent) ενώ η δεύτερη βασίζεται στην αρχιτεκτονική των Radial Basis Function neural networks και χρησιμοποιεί τον αλγόριθμο εκπαίδευσης Levenberg-Marquardt. Τα δίκτυα προσαρμόστηκαν να ελαχιστοποιούν το μέσο τετραγωνικό σφάλμα (Mean Square Error) της εκτίμησης και το σύνολο των δεδομένων που χρησιμοποιήθηκαν έγινε σε αντιστοιχία με άλλες μεθόδους ώστε να είναι εφικτή η σύγκριση. Για το δεύτερο σκέλος της εκτίμησης, δηλαδή για την απεικόνιση από τον χώρο της PCA στον χώρο των φασματικών υπογραφών, αναπτύχθηκε αρχικά μια μέθοδο βασισμένη στον αντίστροφο μετασχηματισμό της PCA. Η επιλογή ενός γραμμικού ορθογώνιου μετασχηματισμού μας δίνει την δυνατότητα της απευθείας αντιστροφής και αποτέλεσε τον βασικό λόγο επιλογής της συγκεκριμένης μεθόδου. Στην συνέχεια εξετάσαμε την συμπεριφορά του συστήματος στην περίπτωση που δεν είναι γνωστός ο αντίστροφος μετασχηματισμός αλλά προσεγγίζεται μέσα από ένα ζεύγος συζευγμένων λεξικών. Μέσω ενός λεξικού χαμηλής ανάλυσης εντοπίζουμε την αραιή περιγραφή του διανύσματος της PCA και με χρήση των αντίστοιχων στηλών του λεξικού υψηλής

ανάλυσης συνθέτουμε το πολυφασματικό pixel. Τέλος επεκτάθηκε η λειτουργία των μοντέλων μας από τα pixel στα patches της εικόνας με σκοπό να εξετάσουμε σε ποιο βαθμό μπορούμε με τις εν λόγω αρχιτεκτονικές να αξιοποιήσουμε τις χωρικές συσχετίσεις μιας εικόνας. Το σύνολο των μοντέλων αξιολογήθηκε με βάση τις μετρικές της ρίζας του μέσου τετραγωνικού σφάλματος (Root Mean Square Error), της ρίζας του μέσου τετραγωνικού απόλυτου σφάλματος (Root Mean Square Relative Error) και το μέσο απόλυτο σχετικό σφάλμα (Mean Absolute Relative Error).

Το μεγαλύτερο μέρος της επεξεργασίας των δεδομένων πραγματοποιήθηκε στο περιβάλλον της MATLAB. Για την μείωση της διάστασης των δεδομένων με χρήση της PCA χρησιμοποιήθηκε το πακέτο drtoolbox [15]. Το δίκτυο Radial Basis Function neural network αναπτύχθηκε εξ ολοκλήρου με custom συναρτήσεις ενώ η εκπαίδευση του πραγματοποιήθηκε στην cloud έκδοση της MATLAB ώστε να ξεπεραστούν οι περιορισμοί (επεξεργαστική ισχύ,RAM) ενός συμβατικού PC. Η εκπαίδευση των λεξικών πραγματοποιήθηκε με χρήση των πακέτων ompbox10 και ksvdbox13 [16] και πάλι στο περιβάλλον της MATLAB. Οι μέθοδοι που βασίστηκαν στην αρχιτεκτονική του Multilayer Perceptron σχεδιάστηκαν στο περιβάλλον google colaboratory με χρήση της Python και αξιοποιώντας την βιβλιοθήκη Pytorch [17].

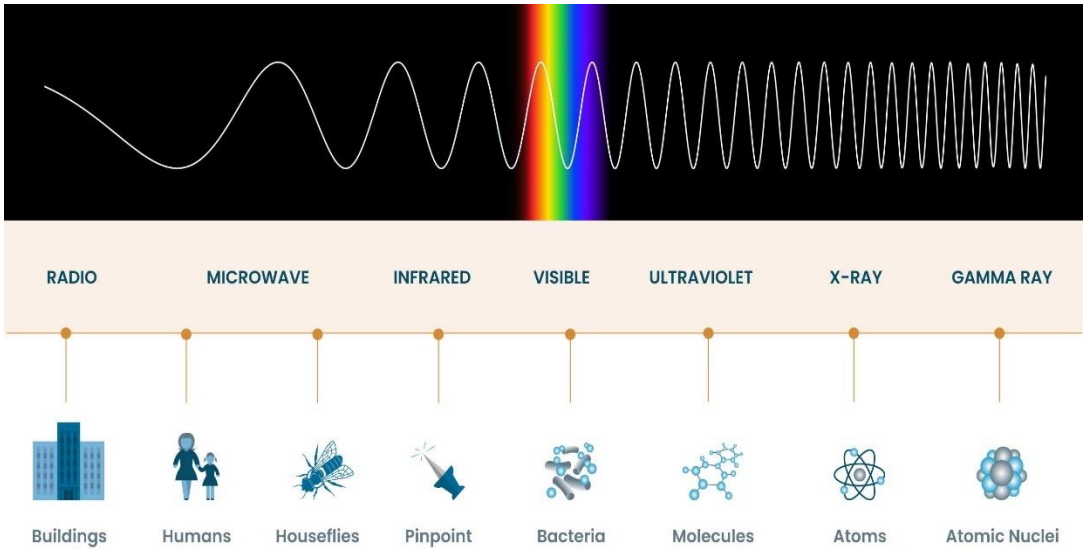
ΚΕΦΑΛΑΙΟ 2

Θεωρία-Μέθοδος

Στο κεφάλαιο αυτό περιγράφονται αναλυτικά όλες οι θεωρητικές έννοιες πάνω στις οποίες στηρίζεται η συγκεκριμένη εργασία. Αρχικά αναλύονται τα χαρακτηριστικά της ηλεκτρομαγνητικής ακτινοβολίας, η προσαρμογή της ανθρώπινης όρασης σε ένα μέρος του φάσματος της καθώς και η λειτουργία της πολυφασματικής ανάλυσης. Στη συνέχεια παρουσιάζεται η θεωρία των αραιών αναπαραστάσεων, ο αλγόριθμος OMP(Orthogonal Matching Pursuit) για την εξαγωγή αραιών περιγραφών και ο αλγόριθμος K-SVD για την εκπαίδευση λεξικών. Ακολουθεί μια εισαγωγή στις τεχνικές μείωσης διάστασης και η αναλυτική παρουσίαση της PCA(Principal Component Analysis). Τέλος έχουμε μια εκτενή παρουσίαση της θεωρίας της μηχανικής μάθησης, των χαρακτηριστικών των αρχιτεκτονικών του Multilayer Perceptron και του Radial Basis Function neural networks, των αλγορίθμων εκπαίδευσης Stochastic Gradient Descent και Levenberg-Marquardt καθώς και των μετρικών αξιολόγησης RMSE, MARE και rRMSE που χρησιμοποιήθηκαν.

2.1 Φάσμα ηλεκτρομαγνητικής ακτινοβολίας

Στον σύγχρονο τεχνολογικό κόσμο που ζούμε, οι εφαρμογές που βασίζονται στην ηλεκτρομαγνητική ακτινοβολία κυριαρχούν όσο ποτέ άλλοτε. Από τα ραδιοκύματα που μεταφέρουν πληροφορία σε τεράστιες αποστάσεις μέχρι τις ακτίνες-Χ που απορροφώνται τοπικά και απεικονίζουν την δομή των οστών μας, πληθώρα εφαρμογών έχουν αλλάξει δραστικά την καθημερινότητα μας. Βασικό χαρακτηριστικό που καθορίζει την συμπεριφορά των ηλεκτρομαγνητικών κυμάτων και την αλληλεπίδραση τους με την ύλη είναι η συχνότητα και το μήκος κύματος τους. Ως ηλεκτρομαγνητικό φάσμα καλούμε λοιπόν το σύνολο των συχνοτήτων, τα αντίστοιχα μήκη κύματος και τις ενέργειες των φωτονίων για τις διάφορες συχνότητες της ηλεκτρομαγνητικής ακτινοβολίας. Το εύρος των συχνοτήτων ξεκινά από το 1 Hertz και μπορεί να ξεπεράσει τα 10^25 Hertz με τα αντίστοιχα μήκη κύματος να κυμαίνονται από κάποιες χιλιάδες χιλιόμετρα μέχρι ένα κλάσμα του μεγέθους του πυρήνα του ατόμου. Το τεράστιο αυτό φάσμα τεμαχίζεται σε περιοχές(μπάντες) συχνοτήτων όπως φαίνεται και στο σχήμα 2.1 που καλούνται ραδιοκύματα, μικροκύματα, υπέρυθρη ακτινοβολία, ορατό φως, υπεριώδης ακτινοβολία. ακτίνες Χ και ακτίνες Γάμα.

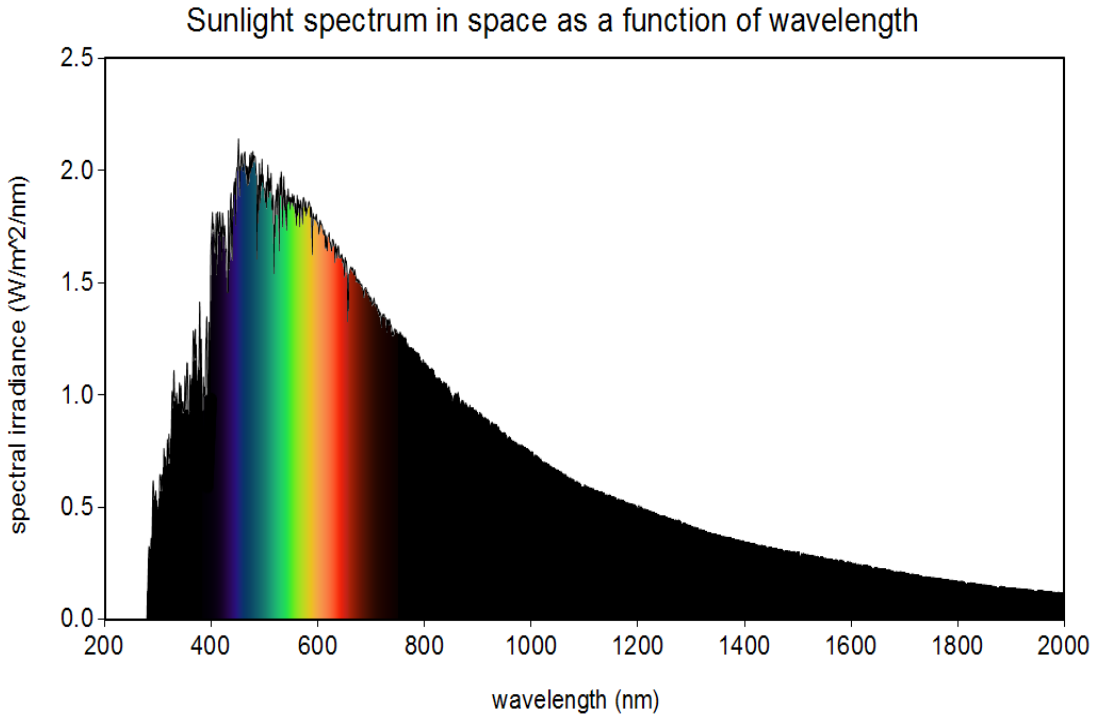


Σχήμα 2.1: Φάσμα ηλεκτρομαγνητικής ακτινοβολίας.

Τα ηλεκτρομαγνητικά κύματα που ταξινομούνται σε κάθε μπάντα συχνοτήτων εμφανίζουν διαφορές ως προς το πως παράγονται, πως αλληλοεπιδρούν με την ύλη και κατ' επέκταση πως χρησιμοποιούνται στην εφαρμογή.

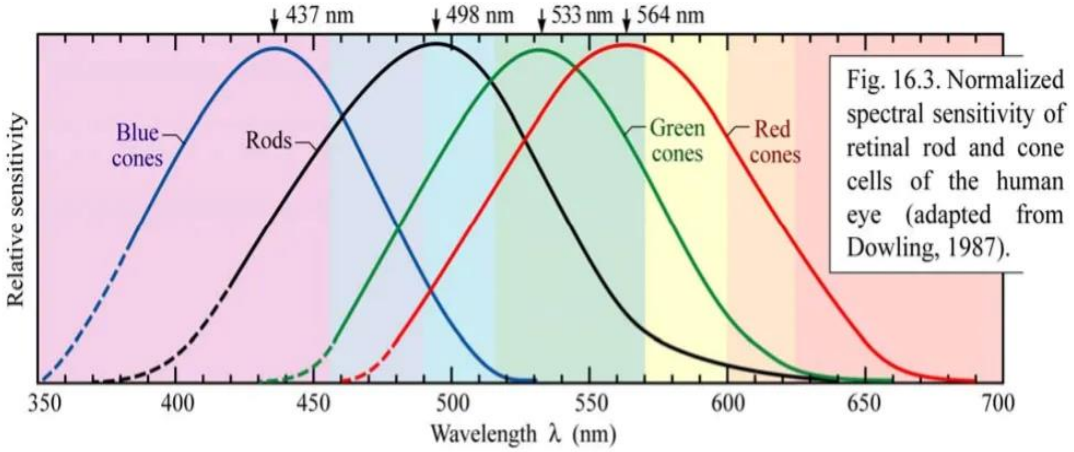
2.1.1 Ορατό φως

Ως ορατό φως καλούμε την μπάντα συχνοτήτων που γίνεται αντιληπτή από την ανθρώπινη όραση. Αποτελείται από μήκη κύματος στην περιοχή 380-750 nm που αντιστοιχεί στις συχνότητες 400-790 THertz. Εάν εξετάσουμε το διάγραμμα πυκνότητας εκπεμπόμενης ενέργειας του ηλίου που φαίνεται στο σχήμα 2.2 καταλαβαίνουμε ότι η όραση μας προσαρμόστηκε εξελικτικά στην περιοχή που υπάρχει η μέγιστη δυνατή πληροφορία προς αξιοποίηση.



Σχήμα 2.2: Πυκνότητα εκπεμπόμενης ισχύος του ηλίου ως προς το μήκος κύματος.

Εάν τώρα κάνουμε ένα ακόμη βήμα και εξετάσουμε της καμπύλες ευαισθησίας του ανθρώπινου ματιού που φαίνονται στο σχήμα 2.3 παρατηρούμε ότι το μάτι μας όχι απλά προσαρμόστηκε στην μικρή μπάντα συχνοτήτων του ορατού φωτός, αλλά δίνει κυρίως έμφαση σε τρία μόλις μήκη κύματος που αντιστοιχούν στα χρώματα κόκκινο, πράσινο και μπλε.

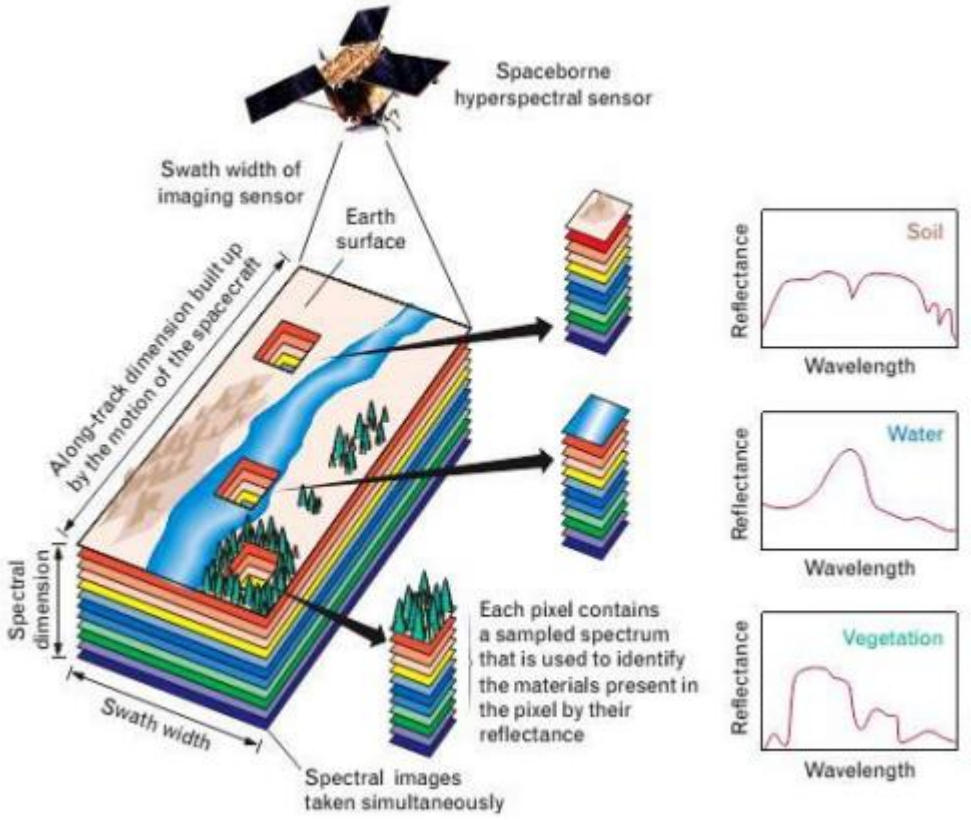


Σχήμα 2.3: Σχετική ευαισθησία ματιού στο φάσμα του ορατού φωτός.

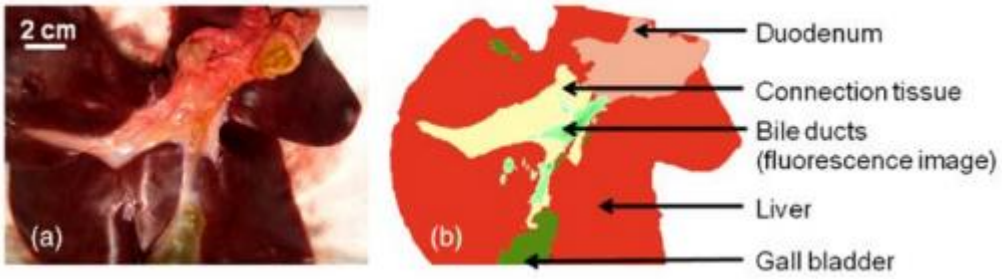
Η ανθρώπινη όραση λοιπόν περιορίζεται σε ένα πολύ μικρό τμήμα του ηλεκτρομαγνητικού φάσματος και διαχωρίζει αντικείμενα με βάση τη φασματική πληροφορία που βρίσκεται σε αυτό [18]. Τα συστήματα καταγραφής εικόνων με την σειρά τους, σχεδιάστηκαν να αποτυπώνουν την αναγκαία για τον άνθρωπο πληροφορία γεγονός που τα περιόρισε στις τρείς RGB συνιστώσες του ορατού φωτός.

2.1.2 Πολυφασματική ανάλυση

Το σύνολο της ανακλώμενης πληροφορίας από ένα υλικό δεν περιορίζεται όμως ούτε στο φάσμα του ορατού φωτός, ούτε στις τρείς RGB συνιστώσες που αναφέραμε. Εάν εξετάσουμε την φασματική του απόκριση(spectral radiance) σε ένα μεγαλύτερο εύρος συχνοτήτων παρατηρούμε ότι τα ιδιαίτερα χαρακτηριστικά του, όπως η χημική του σύσταση, η γεωμετρία των μορίων του, αποτυπώνονται με μοναδικό τρόπο δημιουργώντας έτσι μια φασματική υπογραφή(spectral signature). Αυτές οι υπογραφές μπορούν να χρησιμοποιηθούν για την ταξινόμηση, αξιολόγηση και γενικότερη μελέτη της συμπεριφοράς πολλών υλικών βρίσκοντας έτσι εφαρμογή σε πολλούς τομείς όπως αναφέρθηκε και νωρίτερα. Συνεπώς προκύπτει η ανάγκη δημιουργίας πολυφασματικών καμερών με σκοπό την αξιοποίηση της ανακλώμενης πληροφορίας σε ένα τεράστιο εύρος εφαρμογών. Στα σχήματα που ακολουθούν παρουσιάζονται εφαρμογές της πολυφασματικής ανάλυσης στην γεωργία και την ιατρική.



Σχήμα 2.4: Υπερφασματική απεικόνιση και ταξινόμηση περιοχών βλάστησης, νερού και εδάφους.



Σχήμα 2.5: (a) Δομή χοληφόρου ιστού.(b) Ταξινόμηση διαφορετικών περιοχών με βάση την φασματική τους υπογραφή.

2.1.3 Ανακατασκευή πολυφασματικών εικόνων από RGB αισθητήρες

Με την μέχρι τώρα ανάλυση μας γίνεται ξεκάθαρη η ανάγκη απόκτησης και αξιοποίησης της πολυφασματικής πληροφορίας που ανακλάται από τα διάφορα αντικείμενα. Οι πολυφασματικές κάμερες ωστόσο εμφανίζουν πολύ υψηλή περιπλοκότητα γεγονός που εκτοξεύει το κόστος και το μέγεθος τους και τις καθιστά απαγορευτικές για μεγάλο πλήθος εφαρμογών. Συνεπώς προκύπτει το ερώτημα του κατά πόσο μπορούμε να εκτιμήσουμε με ακρίβεια το πολυφασματικό περιεχόμενο μιας φυσικής σκηνής αξιοποιώντας χαμηλού κόστους RGB αισθητήρες. Πιο αυστηρά αν θεωρήσουμε $Hs: R^p, p > 3$ την φασματική αντανάκλαση ενός pixel $Camera: R^{3xp}$ την απόκριση της κάμερας και $Rgb: R^3$ τις τιμές των RGB καναλιών το πρόβλημα παίρνει την μορφή:

$$Rgb = Camera * Hs \quad (1)$$

Καλούμαστε λοιπόν να αντιστρέψουμε ένα έντονα ασθενώς ορισμένο πρόβλημα όπου το πλήθος των αγνώστων είναι αρκετά μεγαλύτερο από το πλήθος των εξισώσεων. Εάν η εκτίμηση γίνει σε ικανοποιητικά πλαίσια μπορούμε να επεκτείνουμε δραστικά το πεδίο εφαρμογών της πολυφασματικής ανάλυσης με πολύ μικρό κόστος.

2.2 Θεωρία αραιών αναπαραστάσεων

Στον τομέα της επεξεργασίας σημάτων βασικός στόχος είναι η αποδοτική καταγραφή, επεξεργασία, αποθήκευση, μετάδοση και εξαγωγή πληροφορίας από διάφορα είδη δεδομένων. Οι ραγδαίες τεχνολογικές εξελίξεις, χρόνο με το χρόνο, αυξάνουν κατά πολύ τον όγκο και την περιπλοκότητα των δεδομένων που προκύπτουν και γίνονται εφικτό να επεξεργαστούν. Παράλληλα αυξάνεται και η υπολογιστική ισχύ των συσκευών με αποτέλεσμα όλο και περιπλοκότερες μέθοδοι να γίνονται υλοποιήσιμες τόσο σε offline επίπεδο όσο και σε επίπεδο πραγματικού χρόνου. Σκοπός λοιπόν είναι να βρούμε τρόπους να βλέπουμε μέσα από ένα πιο απλό πρίσμα τα πολυδιάστατα πια δεδομένα με στόχο τον πιο εύκολο χειρισμό τους. Σε αυτή την κατεύθυνση αναπτύχθηκαν οι μέθοδοι βασισμένες στις αραιές αναπαραστάσεις.

2.2.1 Αραιότητα φυσικών σημάτων

Βασική θεώρηση για την λειτουργία αυτών των μεθόδων είναι ότι τα σήματα που εμφανίζονται στην πράξη έχουν μια εγγενή αραιότητα. Αν φανταστούμε μια εικόνα, το σύνολο όλων των διαφορετικών συνδυασμών στις τιμές των pixel της απαρτίζουν τον χώρο όλων των διαφορετικών εικόνων. Από αυτές η συντριπτική πλειοψηφία θα είναι εικόνες θορύβου, με ασυσχέτιστα διάσπαρτα pixel παντού. Μόνο ένα μικρό υποσύνολο αυτών θα εμφανίζουν δομή και θα περιγράφουν φυσικά αντικείμενα. Μπορούμε λοιπόν να εντοπίσουμε αυτές τις βασικές δομές και να περιγράψουμε το σύνολο των εικόνων ως αραιούς συνδυασμούς αυτών των βασικών στοιχείων.

Μέθοδοι βασισμένες στις αραιές αναπαραστάσεις έχουν βρει εφαρμογή σε πολλούς τομείς της υπολογιστικής όρασης, της επεξεργασίας σημάτων και της μηχανικής μάθησης όπως image denoising, super-resolution, inverse problems, motion and data segmentation, inpainting, background modeling, image classification κ.α. [19] .

2.2.2 Ασθενώς ορισμένα γραμμικά συστήματα

Το πρόβλημα μας εντάσσεται στην γενικότερη κατηγορία των γραμμικών συστημάτων. Πιο συγκεκριμένα αποτελεί ένα ασθενώς ορισμένο πρόβλημα, όπου το πλήθος των αγνώστων είναι αρκετά μεγαλύτερο από το σύνολο των εξισώσεων. Δηλαδή

$$Ax = b \; με \; A \in R^{nxm} \; και \; n < m \tag{2}$$

Συστήματα τέτοιας μορφής είτε δεν έχουν καμία λύση, στην περίπτωση όπου το διάνυσμα b δεν βρίσκεται στον χώρο των στηλών του πίνακα A, είτε έχουν άπειρες λύσεις. Κάνοντας την παραδοχή ότι ο A είναι ενας full rank πίνακας, δηλαδή υπάρχει ένα υποσύνολο n γραμμικώς ανεξάρτητων στηλών του που περιγράφουν όλο τον χώρο R^n, εξασφαλίζουμε την ύπαρξη λύσης.

Το πρόβλημα λοιπόν καταλήγει να έχει λύση αλλά μέσα από ένα άπειρο σύνολο λύσεων. Σκοπός λοιπόν είναι να επιστρατεύσουμε ένα σύνολο κριτηρίων κανονικοποίησης για να αξιολογούμε τις διαφορετικές επιλογές. Αν θεωρήσουμε μια συνάρτηση $J(x)$ η οποία αξιολογεί το πόσο επιθυμητή είναι μια λύση, δίνοντας μικρότερες τιμές στις καλύτερες λύσεις, τότε ορίζουμε το πρόβλημα βελτιστοποίησης (P_j).

$$(P_j): \quad \min_x J(x) \quad sublect\ to \quad b = Ax \tag{3}$$

Με δεδομένο ότι το σύνολο των διανυσμάτων x που ικανοποιούν το γραμμικό μας σύστημα είναι άπειρο, το κριτήριο $J(x)$ και η ελαχιστοποίηση του καθορίζουν πως θα κινηθεί ο αλγόριθμος για να επιλέξει το καλύτερο δυνατό διάνυσμα και τι μορφή τελικά αυτό θα έχει. Η επιλογή λοιπόν διαφορετικών μετρικών $J(x)$ μας δίνει την δυνατότητα να οδηγήσουμε το σύστημα μας σε λύσεις με εντελώς διαφορετικά χαρακτηριστικά, εκμεταλλευόμενοι τις ιδιότητες των εκάστοτε συναρτήσεων.

2.2.3 Κριτήρια κανονικοποίησης

Το πιο διαδεδομένο κριτήριο κανονικοποίησης είναι η ελαχιστοποίηση της ℓ_2 νόρμας. Η ευκλείδεια νόρμα εκφράζει την ενέργεια μιας λύσης και συνεπώς τον εν λόγω κριτήριο βασίζεται στην παραδοχή ότι τα συστήματα και οι διαδικασίες του φυσικού κόσμου τείνουν συχνά προς καταστάσεις που ελαχιστοποιούν την ενέργεια τους. Το πρόβλημα έτσι ανάγεται σε ένα κοίλο πρόβλημα ελαχίστων τετραγώνων με μοναδική λύση που υπολογίζεται αναλυτικά μέσα από τον ψευδοαντίστροφο του A. Το πρόβλημα ορίζεται ως

$$(P_2): \quad \min_x \|x\|_2^2 \quad sublect\ to \quad b = Ax \tag{4}$$

Και η λύση του

$$x_s = A^+ b = A^T (AA^T)^{-1} b \tag{5}$$

Η απλότητα της ελαχιστοποίησης της ℓ_2 νόρμας οδήγησε αφενός στην διάδοση της μεθόδου, δεν αποτελεί όμως την καθολικά καλύτερη λύση για όλα τα προβλήματα. Η λύση με την ελάχιστη ενέργεια τείνει να είναι πυκνή κάτι που έρχεται σε αντίθεση με τα μοντέλα αραιών αναπαραστάσεων που θέλουμε να εκμεταλλευτούμε. Συνεπώς καλούμαστε να εφαρμόσουμε διαφορετικά μέτρα ελαχιστοποίησης που θα προάγουν τις αραιές λύσεις έναντι των υπολοίπων και να εξετάσουμε σε πιο βαθμό αυτές είναι μοναδικές.

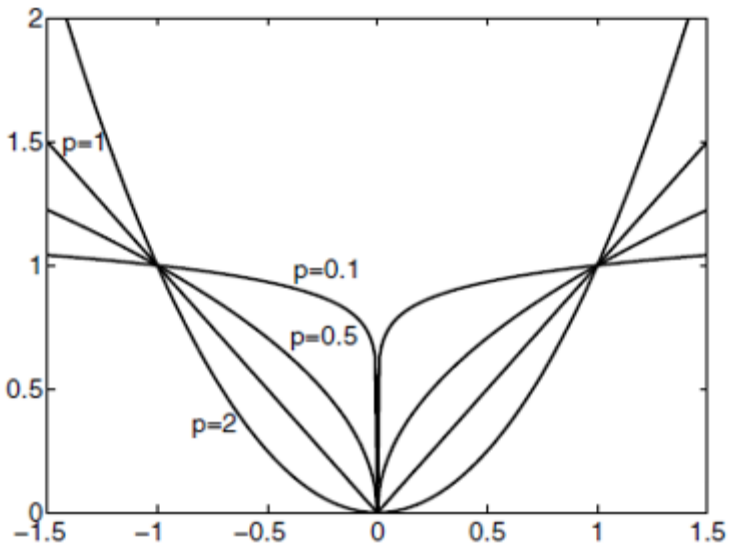
2.2.4 Η ℓ_0 ψευδονόρμα

Ένας απλός και διαισθητικός τρόπος να μετρήσουμε την αραιότητα ενός διανύσματος είναι να μετρήσουμε το σύνολο των μη μηδενικών στοιχείων του. Ένα διάνυσμα είναι αραιό όταν ένας μικρός αριθμός στοιχείων του είναι μη μηδενικά στοιχεία [20]. Ορίζουμε λοιπόν την ℓ_0 ψευδονόρμα:

$$\|x\|_0 = \#\{i: x_i \neq 0\} \tag{6}$$

Όπου αν $\|x\|_0 = k \ll m$ το x είναι αραιό.

Στο σχήμα 2.6 φαίνεται η συμπεριφορά της συνάρτησης $\|x\|^p$ για διαφορετικές τιμές του p. Καθώς $p \to 0$ η $\|x\|^p$ παίρνει της μορφή μιας συνάρτησης δείκτη η οποία για $x = 0$ δίνει 0 και 1 αλλού. Αν εφαρμοστεί στα στοιχεία ενός διανύσματος τότε πρακτικά απαριθμεί τα μη μηδενικά στοιχεία του.



Σχήμα 2.6: Στο σχήμα φαίνεται η συμπεριφορά της συνάρτησης $\|x\|^p$ για διαφορετικές τιμές του p. Καθώς $p \to 0$ η $\|x\|^p$ παίρνει της μορφή μιας συνάρτησης δείκτη η οποία για $x = 0$ δίνει 0 και 1 αλλού.

Ορίζουμε λοιπόν το πρόβλημα P_0 που προκύπτει από το γενικό πρόβλημα βελτιστοποίησης P_j αν επιλέξουμε σαν κριτήριο κανονικοποίησης $J(x) = J_0(x) = \|x\|_0$.

$$(P_0): \quad \min_x \|x\|_0^0 \quad sublect\ to \quad b = Ax \tag{7}$$

Το πρόβλημα P_0 σε αντίθεση με το P_2 δεν είναι κοίλο συνεπώς δεν υπάρχει καμμία εγγύηση για την μοναδικότητα της υποψήφιας λύσης του. Ο δε χώρος των k αραιών λύσεων του, με δεδομένο πως τα διαφορετικά k αραιά διανύσματα μπορεί να συνθέτουν τη λύση επιλέγοντας διαφορετικές στήλες του A, είναι διακριτός και ασυνεχής. Συνεπώς δεν υπάρχει αποτελεσματικός τρόπος για να βρούμε την επιθυμητή λύση. Το πρόβλημα P_0 λοιπόν είναι ένα κλασσικό πρόβλημα συνδυαστικής αναζήτησης όπου για να βρούμε την λύση οφείλουμε να εξετάσουμε εξαντλητικά όλα τα k αραιά υποσύνολα x_s και τα αντίστοιχα υποσυστήματα A_s ώστε $b = A_s x_s$. Το σύνολο όλων των πιθανών k αραιών λύσεων είναι $\binom{n}{k}$. Η πολυπλοκότητα τέτοιων προβλημάτων είναι εκθετική ως προς το k και έχει αποδειχτεί ότι τα P_0 προβλήματα είναι γενικά NP hard [21].

2.2.5 Pursuit αλγόριθμοι

Σκοπός μας λοιπόν είναι να αναπτύξουμε μια αποδοτική και αξιόπιστη μέθοδο επίλυσης για το πρόβλημα P_0 παρά τη δυσκολία που εμφανίζουν αυτά τα προβλήματα. Μια βασική παρατήρηση που μπορούμε να κάνουμε για τη λύση x του προβλήματος P_0 είναι πως αποτελείται από δύο κομμάτια. Το support του διανύσματος, δηλαδή τις στήλες του πίνακα A που θα επιλέξουμε για να συνθέσουμε την λύση και τις μη μηδενικές τιμές του support, δηλαδή την ενέργεια κάθε επιλεχθείσας στήλης. Μπορούμε λοιπόν να εστιάσουμε στο support κατανοώντας πως μόλις βρεθεί οι τιμές των μη μηδενικών συντελεστών υπολογίζονται απλά με μια προσέγγιση ελαχίστων τετραγώνων. Καθώς το support είναι διακριτό οι αλγόριθμοι αναζήτησης του είναι και αυτοί διακριτοί. Όλη αυτή η συλλογιστική μας οδηγεί στην οικογένεια των άπληστων αλγόριθμων(greedy algorithms) που θα

περιγράψουμε στην συνέχεια.

Μια εναλλακτική προσέγγιση για την επίλυση του προβλήματος P_0, η οποία δεν αναλύεται στην παρούσα εργασία αλλά αξίζει να αναφερθεί, είναι να θεωρήσουμε το διάνυσμα $x \in R^m$ συνεχές. Η εξομάλυνση(smoothing) της συνάρτησης σφάλματος $\|x\|_0$, μας περνά πια στο συνεχές πεδίο βελτιστοποίησης. Οι μέθοδοι αυτοί ονομάζονται μέθοδοι χαλάρωσης(relaxation methods) και καταφέρνουν να διατηρήσουν από τη μία την απαραίτητη αραιότητα της λύσης μας και να αντλήσουν από την άλλη τα κυρτά χαρακτηριστικά που εμφανίζουν οι νόρμες μεγαλύτερης τάξης [22].

2.2.6 Ο αλγόριθμος Orthogonal Matching Pursuit

Η βασική ιδέα όπως αναφέραμε βασίζεται στην εύρεση του support της λύσης x που είναι διακριτό. Αν θεωρήσουμε την πιο απλή περίπτωση όπου μόλις μια στήλη του $A \in R^{nxm}$ συνθέτει τη λύση x αρκεί να πραγματοποιήσουμε ένα σύνολο από m τεστ για την εύρεση της, ένα για κάθε στήλη. Το j-οστό τεστ πραγματοποιείται ελαχιστοποιώντας την $\epsilon(j) = \|a_j z_j - b\|_2^2$, που μας οδηγεί στο $z_j^* = a_j^T b / \|a_j\|_2^2$ όπου a_j είναι μια από τις m στήλες του A και z_j η αντίστοιχη ενέργεια της στήλης. Αντικαθιστώντας στην αρχική εξίσωση σφάλματος καταλήγουμε:

$$\begin{aligned} \epsilon(j) = \min_{z_j} \|a_j z_j - b\|_2^2 &= \left\| \frac{a_j^T b}{\|a_j\|_2^2} a_j - b \right\|_2^2 \\ &= \|b\|_2^2 - \frac{2\left(a_j^T b\right)^2}{\|a_j\|_2^2} + \frac{\left(a_j^T b\right)^2}{\|a_j\|_2^2} \\ &= \|b\|_2^2 - \frac{\left(a_j^T b\right)^2}{\|a_j\|_2^2} \end{aligned} \quad (8)$$

Αν το σφάλμα είναι μηδέν έχουμε βρει την κατάλληλη λύση. Η σχέση $\left(a_j^T b\right)^2$ υπολογίζει το εσωτερικό γινόμενο της κάθε στήλης του A με το διάνυσμα b και τελικά το σφάλμα μηδενίζεται για την στήλη που είναι παράλληλη στο b. Η διαδικασία αυτή απαιτεί $O(mn)$ flops που μπορεί να θεωρηθεί ικανοποιητικό για ρεαλιστικές εφαρμογές. Αν τώρα θεωρήσουμε πως το διάνυσμα b είναι γραμμικός συνδυασμός k_o στηλών του A και γενικεύοντας την προηγούμενη προσέγγιση καταλήγουμε ότι χρειαζόμαστε να εξετάσουμε όλα τα $\binom{m}{k_0}$ υποσύνολα των k_o στηλών του A και να αξιολογήσουμε το κάθε ένα. Η διαδικασία αυτή χρειάζεται $O(m^{k_o} n {k_o}^2)$ flops που είναι απαγορευτικά αργή.

Μια άπληστη στρατηγική εγκαταλείπει την λογική της εξαντλητικής αναζήτησης όλων των πιθανών υποσυνόλων και βασίζεται σε μια σειρά τοπικά βέλτιστων βημάτων. Ξεκινώντας από το $x^0 = 0$ κατασκευάζει επαναληπτικά μια προσέγγιση κ-όρων x^k διατηρώντας ένα σετ ενεργών στηλών, το οποίο αρχικά είναι κενό, και σε κάθε βήμα επεκτείνεται κατά μία στήλη. Η στήλη που επιλέγεται είναι αυτή που κάνει τη μέγιστη μείωση του υπολοίπου του σφάλματος, το οποίο υπολογίζεται από την προσέγγιση ελαχίστων τετραγώνων του διανύσματος b από τις μέχρι τώρα ενεργές στήλες. Σε κάθε επανάληψη υπολογίζεται το υπόλοιπο του σφάλματος και ο αλγόριθμος τερματίζει μόλις ξεπεραστεί ένα προκαθορισμένο κατώφλι ή φτάνοντας στο επιθυμητό βαθμό αραιότητας . Ο αλγόριθμος Orthogonal Matching

Pursuit βασίζεται ακριβώς σε αυτή την λογική, όπως και οι περισσότεροι άπληστοι αλγόριθμοι, με κάποιες μικρές αλλαγές.

Task: Approximate the solution of (P_0): $\min_{\mathbf{x}} \|\mathbf{x}\|_0$ subject to $\mathbf{Ax} = \mathbf{b}$.
Parameters: We are given the matrix $\mathbf{A}$, the vector $\mathbf{b}$, and the error threshold ϵ_0.
Initialization: Initialize $k = 0$, and set

- The initial solution $\mathbf{x}^0 = \mathbf{0}$.
- The initial residual $\mathbf{r}^0 = \mathbf{b} - \mathbf{Ax}^0 = \mathbf{b}$.
- The initial solution support $\mathcal{S}^0 = Support\{\mathbf{x}^0\} = \emptyset$.

Main Iteration: Increment k by 1 and perform the following steps:

- **Sweep:** Compute the errors $\epsilon(j) = \min_{z_j} \|\mathbf{a}_j z_j - \mathbf{r}^{k-1}\|_2^2$ for all j using the optimal choice $z_j^* = \mathbf{a}_j^T \mathbf{r}^{k-1} / \|\mathbf{a}_j\|_2^2$.
- **Update Support:** Find a minimizer, j_0 of $\epsilon(j)$: $\forall\, j \notin \mathcal{S}^{k-1}$, $\epsilon(j_0) \leq \epsilon(j)$, and update $\mathcal{S}^k = \mathcal{S}^{k-1} \cup \{j_0\}$.
- **Update Provisional Solution:** Compute $\mathbf{x}^k$, the minimizer of $\|\mathbf{Ax} - \mathbf{b}\|_2^2$ subject to $Support\{\mathbf{x}\} = \mathcal{S}^k$.
- **Update Residual:** Compute $\mathbf{r}^k = \mathbf{b} - \mathbf{Ax}^k$.
- **Stopping Rule:** If $\|\mathbf{r}^k\|_2 < \epsilon_0$, stop. Otherwise, apply another iteration.

Output: The proposed solution is $\mathbf{x}^k$ obtained after k iterations.

Σχήμα 2.7: Orthogonal Matching Pursuit - ένας άπληστος αλγόριθμος για την προσέγγιση της λύσης του προβλήματος P_0.

Το σχήμα παρουσιάζει αναλυτικά τον αλγόριθμο OMP [23]. Η ονομασία προέρχεται από την βιβλιογραφία του τομέα της επεξεργασίας σημάτων. Η διαδικασία όμως χρησιμοποιούταν νωρίτερα σε αρκετούς τομείς με άλλη ονομασία.

Ο αλγόριθμος ξεκινά με την αρχικοποίηση της λύσης $x^0 = 0$, του υπόλοιπου $r^0 = b$ και του support της λύσης $S^0 = Support\{x^0\} = \emptyset$. Η συνάρτηση σφάλματος έχει την ακόλουθη μορφή που μοιάζει αρκετά με την αρχική εξίσωση (8) :

$$\begin{aligned}\epsilon(j) = \min_{z_j} \left\|a_j z_j - r^{k-1}\right\|_2^2 &= \left\| \frac{a_j^T r^{k-1}}{\|a_j\|_2^2} a_j - r^{k-1} \right\|_2^2 \\ &= \|r^{k-1}\|_2^2 - \frac{2\left(a_j^T r^{k-1}\right)^2}{\|a_j\|_2^2} + \frac{\left(a_j^T r^{k-1}\right)^2}{\|a_j\|_2^2} \\ &= \|r^{k-1}\|_2^2 - \frac{\left(a_j^T r^{k-1}\right)^2}{\|a_j\|_2^2} \end{aligned} \quad (9)$$

Στη συνέχεια υπολογίζει το εσωτερικό γινόμενο μεταξύ της κάθε στήλης a_j του πίνακα A και του υπόλοιπου r^{k-1} και επιλέγει αυτή με το μεγαλύτερο. Αναζητά ουσιαστικά την στήλη του πίνακα A που μπορεί να περιγράψει το μεγαλύτερο ποσοστό πληροφορίας από το υπόλοιπο και δεν ανήκει στα στοιχεία που έχουμε επιλέξει μέχρι τώρα. Ανανεώνει το S^κ με τον δείκτη j της επιλεχθείσας στήλης. Επαναυπολογίζει όλους του συντελεστές της λύσης x^κ μέσω της ελαχιστοποίησης της σχέσης $\|Ax - b\|_2^2$ ως προς το x,έτσι ώστε το support της να είναι το σύνολο των στηλών S^κ. Αν θεωρήσουμε το πίνακα A_{S^k} ως ένα πίνακα με διαστάσεις $n \times |S^k|$ που περιέχει τις στήλες του A που ανήκουν στο support του x, το πρόβλημα που αρχικά θέλαμε να λύσουμε ανάγεται στην ελαχιστοποίηση της σχέσης $\left\|A_{S^k}x_{S^k} - b\right\|_2^2$, όπου x_{S^k} το είναι τα μη μηδενικά στοιχεία του διανύσματος x. Η λύση του προκύπτει μηδενίζοντας την παράγωγο αυτής της τετραγωνικής σχέσης.

$$A_{S^k}^T\left(A_{S^k}x_{S^k} - b\right) = -A_{S^k}^T r^k = 0 \tag{10}$$

Χρησιμοποιήσαμε την σχέση για το υπόλοιπο στην κ-οστή επανάληψη, $r^k = b - Ax^k = b - A_{S^k}x_{S^k}$. Η παραπάνω σχέση επιβάλει οι στήλες του A που ανήκουν το support S^κ να είναι ορθογώνιες ως προς το σφάλμα r^k. Αυτό με τη σειρά του οδηγεί τον αλγόριθμο στον να μην επιλέξει καμία από αυτές τις στήλες στα επόμενα βήματα. Από αυτή ακριβώς την απαίτηση, δηλαδή το υπόλοιπο να είναι ορθογώνιο ως προς τις επιλεχθέντες στήλες, προκύπτει και το όνομα της μεθόδου.

Έχουμε αναπτύξει λοιπόν μια διαδικασία η οποία μέσω μιας σειράς τοπικά βέλτιστων βημάτων εντοπίζει μια αραιή περιγραφή μέσα στο ζητούμενο όριο σφάλματος. Οι τοπικά βέλτιστες επιλογές δεν είναι και οι καθολικά καλύτερες συνεπώς δεν υπάρχει εγγύηση ότι θα βρούμε το ολικό ελάχιστο της σχέσης . Συμβιβαζόμαστε όμως σε μια ικανοποιητική προσέγγιση η οποία αν είναι αρκετά αραιή αποδεικνύεται ότι βρίσκεται αρκετά κοντά στην βέλτιστη που δεν μπορούμε να βρούμε [24].

2.2.7 Εκμάθηση λεξικών

Μέχρι τώρα περιγράψαμε πως μπορούμε να εξάγουμε μια αραιή αναπαράσταση. Για να εκμεταλλευτούμε την τάση των φυσικών σημάτων να περιγράφονται αραιά πρέπει να σχεδιάσουμε ένα υπερορισμένο λεξικό(overcomplete dictionary). Το λεξικό περιέχει έναν μεγάλο αριθμό πρωτότυπων σημάτων-ατόμων, και μπορεί να περιγράψει μια τεράστια γκάμα σημάτων ως αραιούς συνδυασμούς αυτών των ατόμων [25]. Πολύ συχνά βλέπουμε την εφαρμογή τέτοιων μεθόδων σε προβλήματα συμπίεσης, εξαγωγής χαρακτηριστικών και κανονικοποίησης αντίστροφων προβλημάτων μεταξύ άλλων. Η σχεδίαση ενός λεξικού που ταιριάζει σχετικά καλά στα δεδομένα μπορεί να γίνει εύκολα με τη χρήση του προκαθορισμένου πίνακα ενός γνωστού μετασχηματισμού όπως ο Wavelet, Fourier, Dct, Curvelet κ.α. Αυτοί οι μετασχηματισμοί είναι γνωστό ότι εντοπίζουν την αραιότητα στα φυσικά σήματα και αντιστρέφονται εύκολα. Με την εκπαίδευση όμως ενός λεξικού σε ένα σύνολο παραδειγμάτων καταφέρνουμε να πετύχουμε καλύτερα αποτελέσματα συμπίεσης και ανακατασκευής αφού γίνεται δυνατή η πολύ καλύτερη προσαρμογή στις διάφορες λεπτομέρειες των δεδομένων. Για τον σκοπό αυτό χρησιμοποιείται ο αλγόριθμος K-SVD.

2.2.8 Ο αλγόριθμος K-SVD

Με τη χρήση ενός υπερορισμένου λεξικού $D \in R^{n\times K}$ που περιέχει K πρωτότυπα άτομα σε κάθε στήλη $\{d_j\}_{j=1}^{K}$, μπορούμε να αποσυνθέσουμε ένα σήμα $y \in R^n$ σε ένα γραμμικό συνδυασμό αυτών των ατόμων. Η αναπαράσταση του y μπορέι να έιναι είτε ακριβής $y = Dx$ είτε προσέγγιση $y \approx Dx$ που ικανοποιεί ένα όριο σφάλματος $\|y - Dx\|_2 \leq \epsilon$. Το διάνυσμα $x \in R^K$ που περιέχει την αραιή αναπαράσταση του σήματος y προκύπτει μέσα από την επίλυση του προβλήματος P_0:

$$(P_0): \quad \min_x \|x\|_0 \quad subject\ to \quad y = Dx$$

Ή

$$(P_{0,\epsilon}): \quad \min_x \|x\|_0 \quad subject\ to \quad \|y - Dx\|_2 \leq \epsilon \tag{11}$$

Σκοπός λοιπόν του K-SVD είναι να δημιουργήσει ένα λεξικό D που να περιγράφει το σύνολο των δεδομένων y_i μέσα από αραιούς συνδυασμούς των στηλών του [26]. Αποτελεί ουσιαστικά μια γενίκευση του γνωστού αλγορίθμου k-means. Στην περίπτωση του clustering, ο αλγόριθμος k-means μαθαίνει ένα σύνολο διανυσμάτων $\{d_k\}_{k=1}^{K}$ με σκοπό κάθε δείγμα να περιγράφεται από ένα μόνο από αυτά τα διανύσματα(αυτό με την μικρότερη ευκλείδεια απόσταση). Μπορούμε να σκεφτούμε αυτή τη λογική ως μια ακραία περίπτωση αραιής αναπαράστασης όπου μόλις ένα άτομο επιτρέπεται για την αναπαράσταση των σημάτων και επιπλέον η ενέργεια του κάθε ατόμου είναι ένα. Στην περίπτωση των αραιών αναπαραστάσεων κάθε παράδειγμα περιγράφεται ως ένας γραμμικός συνδυασμός αρκετών ατόμων που αποτελεί μια γενίκευση της προσέγγισης του k-means.

Πιο αυστηρά ο αλγόριθμος προσπαθεί να ελαχιστοποιήσει την σχέση:

$$\min_{D,X}\{\|Y - DX\|_F^2\}\ subject\ to \quad \forall i, \|x_i\|_0 \leq T_0 \tag{12}$$

Δεν μπορούμε να λύσουμε ταυτόχρονα και ως προς τους δύο αγνώστους, γι' αυτό χρησιμοποιούμαι μια στρατηγική δύο βημάτων όπου στην μια επανάληψη ανανεώνονται οι αραιές αναπαραστάσεις και στην επόμενη τα άτομα του λεξικού.

Για τον υπολογισμό των αραιών αναπαραστάσεων X θεωρούμε το λεξικό D σταθερό. Το πρόβλημα ανάγεται έτσι στην εύρεση της αραιής αναπαράστασης του κάθε διανύσματος y_i. Η σχέση (12) μπορεί να τροποποιηθεί στην:

$$\|Y - DX\|_F^2 = \sum_{i=1}^{N} \|y_i - Dx_i\|_2^2 \tag{13}$$

Με αυτό τον τρόπο αποσυνθέτουμε την εξίσωση σε N διακριτά προβλήματα της μορφής

$$\min_{x_i}\{\|y_i - Dx_i\|_F^2\}\ subject\ to \quad \|x_i\|_0 \leq T_0 \tag{14}$$
$$for\ i = 1,2,\dots,N.$$

Οι αραιές αναπαραστάσεις x_i μπορούν εξαχθούν με χρήση κάποιου pursuit αλγορίθμου όπως ο OMP που περιγράψαμε νωρίτερα. Αν το T_0 είναι αρκετά μικρό η λύση είναι μια καλή προσέγγιση της ιδανικής λύσης που δεν μπορούμε να υπολογίσουμε.

Task: Find the best dictionary to represent the data samples $\{\mathbf{y}_i\}_{i=1}^N$ as sparse compositions, by solving

$$\min_{\mathbf{D},\mathbf{X}} \left\{\|\mathbf{Y}-\mathbf{D}\mathbf{X}\|_F^2\right\} \quad \text{subject to} \quad \forall i,\ \|\mathbf{x}_i\|_0 \le T_0.$$

Initialization : Set the dictionary matrix $\mathbf{D}^{(0)} \in \mathbb{R}^{n\times K}$ with ℓ^2 normalized columns. Set $J = 1$.
Repeat until convergence (stopping rule):

- *Sparse Coding Stage*: Use any pursuit algorithm to compute the representation vectors $\mathbf{x}_i$ for each example $\mathbf{y}_i$, by approximating the solution of

$$i = 1, 2, \ldots,\ N, \quad \min_{\mathbf{x}_i} \left\{\|\mathbf{y}_i - \mathbf{D}\mathbf{x}_i\|_2^2\right\} \quad \text{subject to} \quad \|\mathbf{x}_i\|_0 \le T_0.$$

- *Codebook Update Stage*: For each column $k = 1, 2, \ldots, K$ in $\mathbf{D}^{(J-1)}$, update it by
  - Define the group of examples that use this atom, $\omega_k = \{i |\ 1 \le i \le N,\ \mathbf{x}_T^k(i) \ne 0\}$.
  - Compute the overall representation error matrix, $\mathbf{E}_k$, by

$$\mathbf{E}_k = \mathbf{Y} - \sum_{j\ne k} \mathbf{d}_j \mathbf{x}_T^j.$$

  - Restrict $\mathbf{E}_k$ by choosing only the columns corresponding to ω_k, and obtain $\mathbf{E}_k^R$.
  - Apply SVD decomposition $\mathbf{E}_k^R = \mathbf{U}\boldsymbol{\Delta}\mathbf{V}^T$. Choose the updated dictionary column $\tilde{\mathbf{d}}_k$ to be the first column of $\mathbf{U}$. Update the coefficient vector $\mathbf{x}_R^k$ to be the first column of $\mathbf{V}$ multiplied by $\boldsymbol{\Delta}(1,1)$.
- Set $J = J + 1$.

Σχήμα 2.8: Ο αλγόριθμος K-SVD.

Στο πρώτο στάδιο περιγράψαμε την ανανέωση των αναπαραστάσεων X με το D σταθερό. Η διαδικασία της ανανέωσης του λεξικού D με σταθερές τις αναπαραστάσεις X είναι αρκετά πιο περίπλοκη. Αν υποθέσουμε ότι και το λεξικό D και οι αναπαραστάσεις X είναι σταθερά και πως θέλουμε να εξετάσουμε μόνο μια στήλη του d_k και τους αντίστοιχους συντελεστές της κ-οστης σειράς του X, x_T^k. Η αντικειμενική συνάρτηση παίρνει την μορφή:

$$\|Y - DX\|_F^2 \quad = \left\| Y - \sum_{j=1}^{K} d_j x_T^j \right\|_F^2$$

$$= \left\| \left(Y - \sum_{j \neq k}^{K} d_j x_T^j \right) - d_k x_T^k \right\|_F^2$$
$$= \left\| E_k - d_k x_T^k \right\|_F^2 \quad (15)$$

Αποσυνθέτουμε με αυτό τον τρόπο τον πολλαπλασιασμό DX σε άθροισμα K $rank$ 1 πινάκων. Μεταξύ αυτών, οι $K-1$ όροι παραμένουν σταθεροί και εξετάζουμε τον k-οστο. Ο πίνακας E_k περιέχει το σφάλμα όλων των N παραδειγμάτων όταν το k-οστο άτομο d_k αφαιρείται από την εκτίμηση τους.

Ορίζουμε ω_k ως το σύνολο των δεικτών i των παραδειγμάτων y_i που χρησιμοποιούν το άτομο d_k στην ανακατασκευή τους, δηλαδή τα μη μηδενικά στοιχεία του x_T^k.

$$\omega_k = \left\{ i \; 1 \leq i \leq K, x_T^k(i) \neq 0 \right\} \quad (16)$$

Ορίζουμε τον πίνακα Ω_k μεγέθους $N \times |\omega_k|$,με τιμές 1 στις θέσεις $(\omega_k(i), i)$ και 0 οπουδήποτε αλλού. Όταν πολλαπλασιάσουμε $x_R^k = x_T^k \Omega_k$ συρρικνώνουμε το διάνυσμα γραμμή x_T^k πετώντας τα μη μηδενικά στοιχεία του και δημιουργούμε το διάνυσμα x_R^k μήκους $|\omega_k|$. Με τον ίδιο τρόπο, ο πολλαπλασιασμός $Y_k^R = Y\Omega_k$ δημιουργεί ένα πίνακα μεγέθους $n \times |\omega_k|$ που περιέχει το υποσύνολο των παραδειγμάτων που χρησιμοποιούν το άτομο d_k στην ανακατασκευή τους. Και στην περίπτωση του $E_k^R = E_k \Omega_k$ απομονώνουμε μόνο το σφάλμα των παραδειγμάτων που εμπεριέχουν και πάλι το εν λόγω άτομο. Ουσιαστικά είναι το σύνολο της πληροφορίας που θέλουμε να περιγράψουμε με το k-οστο άτομο. Μπορούμε να τροποποιήσουμε τώρα την σχέση (15) και να ελαχιστοποιήσουμε τελικά την:

$$\left\| E_k \Omega_k - d_k x_T^k \Omega_k \right\|_F^2 = \left\| E_k^R - d_k x_R^k \right\|_F^2 \quad (17)$$

Έχουμε απομονώσει έτσι τα στοιχεία στα οποία πραγματικά συμμετέχει το άτομο για να τα παράξει και σκοπός είναι να το μεταβάλουμε ώστε να περιγράφει το μεγαλύτερο ποσοστό πληροφορίας που μπορεί. Αυτό το πετυχαίνει με την αποσύνθεση του πίνακα $E_k^R = U\Delta V^T$ με χρήση του SVD . Θέτουμε λοιπόν τη λύση $\widehat{d_k}$ ως την πρώτη στήλη του πίνακα U και το διάνυσμα συντελεστών x_R^k ως την πρώτη στήλη του πίνακα V πολλαπλασιασμένη με την τιμή $\Delta(1,1)$. Επιλέγουμε το άτομο και την ενέργεια με βάση την καλύτερη $rank$ 1 προσέγγιση που δίνει ο SVD. Η διαδικασία επαναλαμβάνεται για κάθε άτομο.

Ο αλγόριθμός τελικά σε πρώτο χρόνο εξάγει τις αραιές αναπαραστάσεις θεωρώντας σταθερό το λεξικό και στη συνέχεια βελτιστοποιεί το λεξικό με βάση την παραπάνω διαδικασία. Εφαρμόζοντας επαναληπτικά τα δύο βήματα θα συγκλίνουμε τελικά σε ένα μοντέλο που θα περιγράφει ικανοποιητικά και με ακρίβεια τα δεδομένα μας.

2.3 Τεχνικές μείωσης διάστασης

Όπως αναφέραμε και νωρίτερα, τα πραγματικά δεδομένα, όπως τα σήματα ομιλίας, οι ψηφιακές φωτογραφίες και τα fmri scans έχουν πολύ μεγάλη διάσταση. Για να μπορέσουμε να χειριστούμε τέτοιου είδους δεδομένα αποδοτικά η διάσταση τους πρέπει να μειωθεί. Οι μέθοδοι που θα περιγράψουμε μετασχηματίζουν τα πολυδιάστατα δεδομένα σε περιγραφές χαμηλότερης διάστασης όπου όμως διατηρούν το μεγαλύτερο κομμάτι της πληροφορίας. Ιδανικά η μειωμένη αναπαράσταση πρέπει να έχει ίδιο μέγεθος με την εσωτερική διάσταση των δεδομένων, η οποία είναι ο μικρότερος αριθμός παραμέτρων που χρειάζονται για να περιγράψουμε τις παρατηρούμενες ιδιότητες τους [27]. Οι τεχνικές μείωσης διάστασης είναι χρήσιμες σε πολλούς τομείς, καθώς περιορίζουν τις αρνητικές επιπτώσεις των πολυδιάστατων υποχωρών όπως η κατάρα της διαστασιμότητα [28]. Ως αποτέλεσμα, εφαρμόζονται μεταξύ άλλων σε προβλήματα κατηγοριοποίησης, οπτικοποίησης και συμπίεσης πολυδιάστατων δεδομένων. Οι πιο διαδεδομένες μέθοδοι είναι η Pca, factor analysis, classical scaling.

Η μέθοδος που θα περιγράψουμε ανήκει στις κυρτές τεχνικές, που βασίζονται στην βελτιστοποίηση μιας κυρτής συνάρτησης [29]. Αρκετά συχνά η αντικειμενική συνάρτηση έχει την μορφή ενός Rayleigh quotient, δηλαδή $\Phi(Y) = \frac{Y^T AY}{Y^T BY}$. Προβλήματα αυτής της μορφής είναι γνωστό πως βελτιστοποιούνται λύνοντας ένα γενικευμένο πρόβλημα ιδιοτιμών. Πιο συγκεκριμένα πραγματοποιούν ένα eigendecomposition του πίνακα που περιέχει τη συνδιακύμανση των διαστάσεων. Αυτές οι τεχνικές ονομάζονται επίσης full spectral.

2.3.1 Principal Component Analysis

Η PCA ορίζεται ως ένας ορθογώνιος γραμμικός μετασχηματισμός που μεταφέρει τα δεδομένα μας σε ένα νέο σύστημα συντεταγμένων έτσι ώστε το μεγαλύτερο ποσοστό της διασποράς να αποτυπώνεται με τη σειρά στην πρώτη συντεταγμένη(το πρώτο principal component), το αμέσως επόμενο στην δεύτερη κ.ο.κ [30].

Θεωρούμε έναν πίνακα δεδομένων X με διάσταση $n \times p$ όπου ανά στήλη εκφράζεται ένα χαρακτηριστικό του πειράματος(η μέση τιμή είναι μηδέν ανά στήλη) και σε κάθε γραμμή μια επανάληψη του.

Μαθηματικά, ο μετασχηματισμός ορίζεται από ένα σύνολο l p-διαστατων διανυσμάτων που απεικονίζουν την κάθε γραμμή του πίνακα X σε ένα νέο διάνυσμα που περιέχει τις τιμές των principal components $t_{(i)} = (t_1, \dots, t_l)$ και υπολογίζεται

$$t_{k(i)} = x_{(i)} * w_{(k)} \quad for \ i = 1, \dots, n \quad k = 1, \dots, l \tag{18}$$

Ο υπολογισμός γίνεται με τέτοιο τρόπο ώστε οι μεταβλητές $t_1, \dots, t_l$ να αποτυπώνουν το μεγαλύτερο ποσοστό της διασποράς των δεδομένων που βρίσκονται στις γραμμές του X και κάθε διάνυσμα w να είναι μοναδιαίο. Αν επιλέξουμε η διάσταση l των μετασχηματισμένων δεδομένων να είναι μικρότερη της αρχικής $l < p$, έχουμε μειώση την διάσταση της αναπαράστασης.

Για να πετύχουμε μεγιστοποίηση της διασποράς μετά την εφαρμογή του μετασχηματισμού, το πρώτο διάνυσμα $w_{(1)}$ πρέπει να ικανοποιεί την:

$$w_{(1)} = arg\max_{\|w\|=1}\left\{\sum_i (t_1)^2_{(i)}\right\} = arg\max_{\|w\|=1}\left\{\sum_i (x_{(i)} * w)^2\right\}$$

Και σε μορφή πινάκων γράφεται ισοδύναμα

$$w_{(1)} = arg\max_{\|w\|=1}\{\|Xw\|^2\} = arg\max_{\|w\|=1}\{w^T X^T Xw\}$$

Επειδή το $w_{(1)}$ είναι μοναδιαίο διάνυσμα ισχύει επίσης

$$w_{(1)} = arg\max_{\|w\|=1}\left\{\frac{w^T X^T Xw}{w^T w}\right\} \quad (19)$$

Η σχέση (19) που θέλουμε να μεγιστοποιήσουμε παίρνει την μορφή ενός Rayleigh quotient. Έχει αποδειχτεί ότι για έναν θετικά ημιορισμένο πίνακα όπως ο $X^T X$ το μέγιστο είναι η μεγαλύτερη ιδιοτιμή του πίνακα, και συμβαίνει όταν το διάνυσμα $w_{(1)}$ είναι το αντίστοιχο ιδιοδιάνυσμα. Βρίσκουμε έτσι το πρώτο διάνυσμα του μετασχηματισμού που περιγράφει την μεγαλύτερη δυνατή διασπορά.

Αν συνεχίσουμε την διαδικασία το k-οστο στοιχείο βρίσκεται αν αρχικά αφαιρέσουμε τα $k-1$ principal components από το X :

$$\hat{X}_k = X - \sum_{s=1}^{k-1} Xw_{(s)}w_{(s)}^T \quad (20)$$

Και στη συνέχεια βρούμε το διάνυσμα $w_{(k)}$ που εξάγει τη μέγιστη διασπορά από τον νέο πίνακα δεδομένων $\hat{X}_k$:

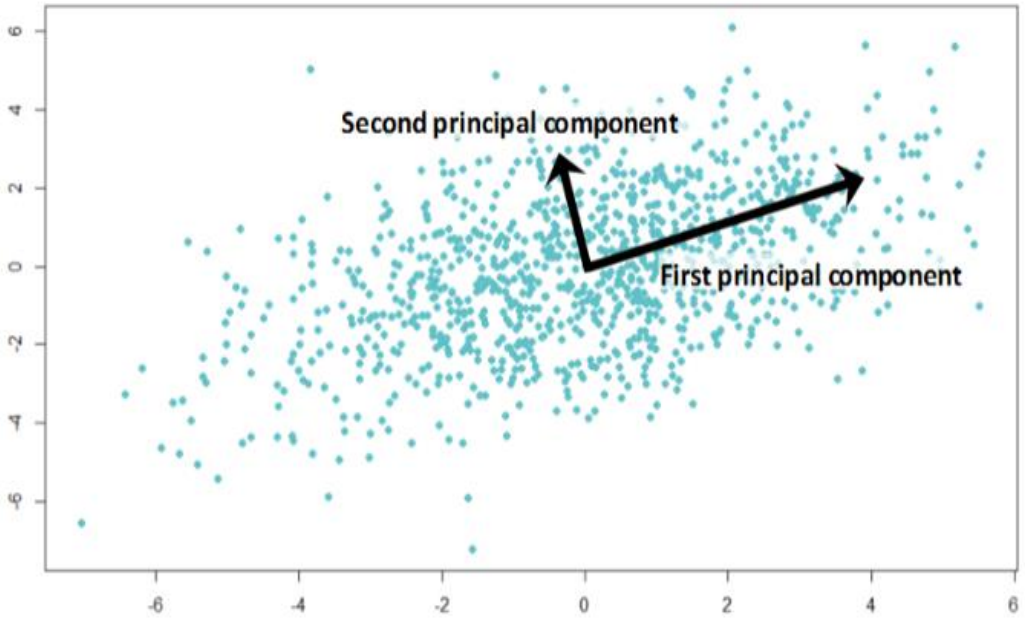
$$w_{(k)} = arg\max_{\|w\|=1}\left\{\left\|\hat{X}_k w\right\|^2\right\} = arg\max_{\|w\|=1}\left\{\frac{w^T \hat{X}_k{}^T \hat{X}_k w}{w^T w}\right\} \quad (21)$$

Αποδεικνύεται ότι αυτή η διαδικασία μας δίνει και τα υπόλοιπα ιδιοδιανύσματα του πίνακα $X^T X$.

Τελικά ο μετασχηματισμός των δεδομένων παίρνει την μορφή

$$T = XW \quad (22)$$

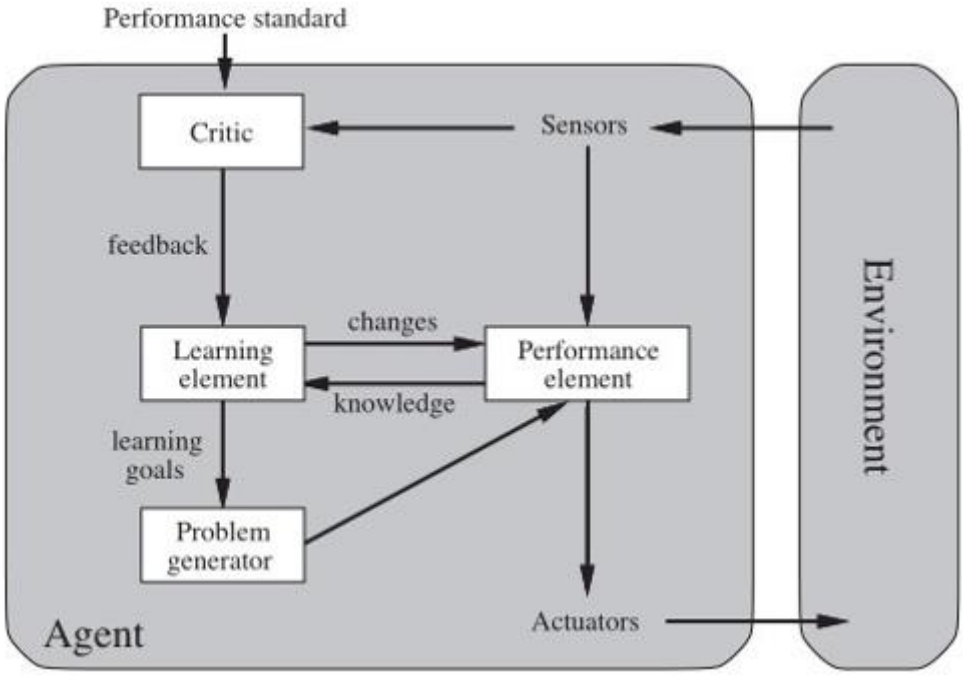
Ο W είναι ένας πίνακας pxl όπου η κάθε στήλη του περιέχει ένα ιδιοδιάνυσμα του $X^T X$. Ο πίνακας $X^T X$ αποτελεί μια εκτίμηση του πίνακα συνδιακύμανσης των δεδομένων, ο οποίος περιγράφει τις αποστάσεις και τις συσχετίσεις μεταξυ των σημείων, και μέσω της αποσύνθεσης του στα l πρώτα ιδιοδιανύσματα του αποτυπώνουμε το μεγαλύτερο ποσοστό της διασποράς των δεδομένων χρησιμοποιώντας μικρότερη αναπαράσταση.



Σχήμα 2.9: Σχηματική αναπαράσταση των δύο κυρίαρχων συνιστωσών(Principal components) σε ένα σύνολο δισδιάστατων δεδομένων.

2.4 Μηχανική μάθηση

Στην προσπάθεια μας να περιγράψουμε την διαδικασία της μάθησης, από το εκάστοτε σύστημα που σχεδιάζουμε, οφείλουμε να δώσουμε κάποιους ορισμούς. Τα συστήματα στα οποία αναφερόμαστε, που έχουν και την δυνατότητα να μάθουν ονομάζονται πράκτορες. Πράκτορας είναι οτιδήποτε μπορεί να θεωρηθεί ότι αντιλαμβάνεται το περιβάλλον του(environment) μέσω αισθητήρων(sensors) και επενεργεί σε αυτό το περιβάλλον μέσω μηχανισμών δράσης(actuators). Η ιδέα λοιπόν πίσω από τη μάθηση είναι ότι οι αισθήσεις δεν πρέπει να χρησιμοποιούνται απλώς για άμεση δράση, αλλά και για την βελτίωση της ικανότητας του πράκτορα να ενεργεί στο μέλλον. Η μάθηση συμβαίνει καθώς ο πράκτορας παρατηρεί τις αλληλεπιδράσεις του με τον κόσμο και τις διαδικασίες λήψεως αποφάσεων που ακολουθεί. Η μάθηση μπορεί να κυμαίνεται από την απλή απομνημόνευση των εμπειριών, μέχρι την δημιουργία ολόκληρων επιστημονικών θεωριών [31] .



Σχήμα 2.10: Αλληλεπίδραση πράκτορα με το περιβάλλον του.

Ένας πράκτορας μάθησης μπορεί να θεωρηθεί ότι περιλαμβάνει ένα στοιχείο εκτέλεσης(performance element), το οποίο αποφασίζει τι ενέργειες να κάνει, και ένα στοιχείο μάθησης(learning element) που τροποποιεί το στοιχείο εκτέλεσης έτσι ώστε να λαμβάνει καλύτερες αποφάσεις. Οι ερευνητές της μηχανικής μάθησης έχουν εφεύρει μια μεγάλη ποικιλία στοιχείων μάθησης. Για να τα κατανοήσουμε, είναι χρήσιμο να δούμε με ποιον τρόπο η σχεδίαση τους επηρεάζεται από το περιβάλλον στο οποίο θα εφαρμοστούν. Η σχεδίαση ενός στοιχείου μάθησης επηρεάζεται από τρία κύρια ζητήματα:

- Ποιες συνιστώσες του στοιχείου εκτέλεσης πρέπει να μαθευτούν.

- Τι ανάδραση διατίθεται για την μάθηση αυτών των συνιστωσών.
- Ποια αναπαράσταση χρησιμοποιείται για τις συνιστώσες.

Ο τύπος της ανάδρασης που διατίθεται για την μάθηση είναι συνήθως ο πιο σημαντικός παράγοντας για τον προσδιορισμό της φύσης του μαθησιακού προβλήματος που αντιμετωπίζει ο πράκτορας. Ο τομέας της μηχανικής μάθησης διακρίνει συνήθως τρείς περιπτώσεις: επιβλεπόμενη μάθηση(supervised learning), μη επιβλεπόμενη μάθηση(unsupervised learning), και ενισχυτική μάθηση(reinforcement learning).

2.4.1. Μη επιβλεπόμενη μάθηση

Η μη επιβλεπόμενη μάθηση, χρησιμοποιεί αλγορίθμους μηχανικής μάθησης για να αναλύσει και να κατηγοριοποιήσει unlabeled δεδομένα. Αυτοί οι αλγόριθμοι ανακαλύπτουν κρυφά πρότυπα(hidden patterns) ή συγκεντρώσεις στα δεδομένα(clusters) χωρίς την ανάγκη για ανθρώπινη παρέμβαση [32]. Η δυνατότητα αυτών των αλγορίθμων να εντοπίζουν ομοιότητες και διαφορές στα δεδομένα τους καθιστούν ιδανικούς για την ανάλυση δεδομένων. Βασικές κατηγορίες τέτοιων αλγορίθμων είναι οι αλγόριθμοι ανάλυσης συστάδων(cluster analysis), και μείωσης διάστασης(PCA,SVD). Αυτή η κατηγορία μάθησης βρίσκει ευρεία εφαρμογή σε προβλήματα εκτίμησης της συνάρτησης πυκνότητας πιθανότητας(pdf) σε προβλήματα στατιστικής.

2.4.2 Ενισχυτική μάθηση

Η ενισχυτική μάθηση είναι ένας τομέας της μηχανικής μάθησης που ασχολείται με το πως ένας ευφυής πράκτορας παίρνει αποφάσεις και τις υλοποιεί σε ένα περιβάλλον με στόχο να μεγιστοποιήσει μια ανταμοιβή. Πιο αυστηρά η ενισχυτική μάθηση είναι το να μαθαίνεις τι να κάνεις, πως να απεικονίσεις μια κατάσταση σε μια ενέργεια, έτσι ώστε να μεγιστοποιήσεις ένα σήμα ανταμοιβής. Ο μαθητευόμενος δεν γνωρίζει ποιες ενέργειες πρέπει να επιλέξει, αλλά οφείλει να ανακαλύψει ποιες από αυτές δίνουν τη μεγαλύτερη ανταμοιβή μέσα από μια σειρά δοκιμών [33]. Όπως και η μη επιβλεπόμενη μάθηση δεν χρειάζεται συγκεκριμένες τιμές εξόδου. Υλοποιείται τυπικά με την μορφή αλυσίδων αποφάσεων Markov.



Σχήμα 2.11: Σχηματική αναπαράσταση της αλληλεπίδρασης της ενισχυτικής μάθησης με το περιβάλλον.

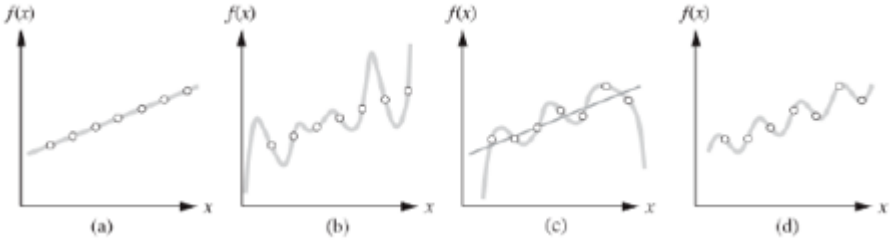
2.4.3 Επιβλεπόμενη μάθηση

Η επιβλεπόμενη μάθηση εφαρμόζεται σε προβλήματα όπου τα δεδομένα μας αποτελούνται από labeled παραδείγματα, δηλαδή κάθε παράδειγμα εκπαίδευσης περιέχει τις αντίστοιχες τιμές εισόδου(features) και εξόδου(label) [34]. Ο σκοπός της επιβλεπόμενης μάθησης είναι να μάθει μια συνάρτηση η οποία απεικονίζει τα δεδομένα εισόδου(inputs) στο αντίστοιχο label εξόδου(output), βασισμένο στα ζευγάρια εισόδου-εξόδου που χρησιμοποιούμαι κατά την εκπαίδευση. Τέτοιοι αλγόριθμοι αναλύουν τα δεδομένα εισόδου και παράγουν μια συνάρτηση απεικόνισης η οποία στη συνέχεια μπορεί να χρησιμοποιηθεί για την απεικόνιση νέων άγνωστων παραδειγμάτων. Στην ιδανική περίπτωση, ο αλγόριθμος μπορεί μετά το πέρας της εκπαίδευσης να αντιστοιχίσει με ακρίβεια τα άγνωστα δεδομένα εισόδου στις αντίστοιχες κατηγορίες(classification) ή να εκτιμήσει τις αντίστοιχες πραγματικές τιμές εξόδου(regression).

Πιο αυστηρά ο στόχος της επιβλεπόμενης μάθησης είναι: με δεδομένο ένα σύνολο N παραδειγμάτων εισόδου εξόδου της μορφής

$$(x_1\, y_1), (x_2\, y_2), .., (x_N\, y_N)$$

όπου κάθε ένα δημιουργείται από μια άγνωστη συνάρτηση της μορφής $y = f(x)$ να προσεγγίσουμε την συνάρτηση f μέσω της συνάρτησης h , μέσα από χώρο πιθανών συναρτήσεων H, με όσο το δυνατόν μεγαλύτερη ακρίβεια γίνεται.



Σχήμα 2.12: Παραδείγματα ζευγών $(x, f(x))$ και υποθέσεις διαφορετικών βαθμών πολυωνύμου (a) 1ου βαθμού, (b) 7ου βαθμού, (c) ένα διαφορετικό dataset και μια γραμμική εκτίμηση, (d) ημιτονοειδής εκτίμηση.

Στο σχήμα 2.12 βλέπουμε ένα παράδειγμα όπου ζεύγη σημείων της μορφής $(x, f(x))$, με την συνάρτηση $f(x)$ να είναι άγνωστη, προσεγγίζονται από μια συνάρτηση h. Η συνάρτηση h επιλέγεται από τον χώρο H που αποτελείται από πολυώνυμα διαφορετικής τάξης. Στην περίπτωση όπου διαφορετικές συναρτήσεις h ταιριάζουν στα δεδομένα μας με την ίδια ακρίβεια προκύπτει το πρόβλημα της επιλογής μεταξύ των διαφορετικών υποθέσεων. Μια βασική αρχή είναι να επιλέγουμε την υπόθεση με την μικρότερη πολυπλοκότητα(Okam's razor).

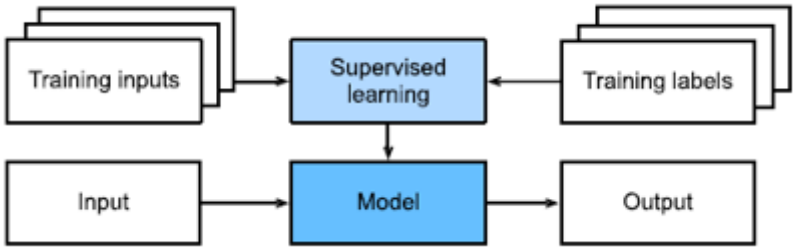
Στο συγκεκριμένο παράδειγμα η επιβλεπόμενη μάθηση πραγματοποιείται με την επιλογή της h^* που είναι πιο πιθανή με βάση τα δεδομένα μας:

$$h^* = \underset{h \in H}{arg\max} P(h/data) \tag{23}$$

Η εξίσωση (23) μέσω του κανόνα του Bayes γίνεται

$$h^* = \underset{h \in H}{arg\max} P(h|data)\, P(h) \quad (24)$$

Η a priori πιθανότητα $P(h)$ είναι μεγάλη για μικρότερου βαθμού πολυώνυμα και μικραίνει όσο ο βαθμός τους μεγαλώνει. Με αυτό τον τρόπο δίνεται προτεραιότητα σε απλούστερες λύσεις και μόνο σε περίπτωση που το απαιτούν τα δεδομένα θα οδηγηθούμε σε περιπλοκότερα συστήματα. Γενικότερα υπάρχει ένα tradeoff μεταξύ της εκφραστικότητας ενός χώρου υποθέσεων, και της πολυπλοκότητας της εύρεσης μια καλής υπόθεσης μέσα στον χώρο αυτό.



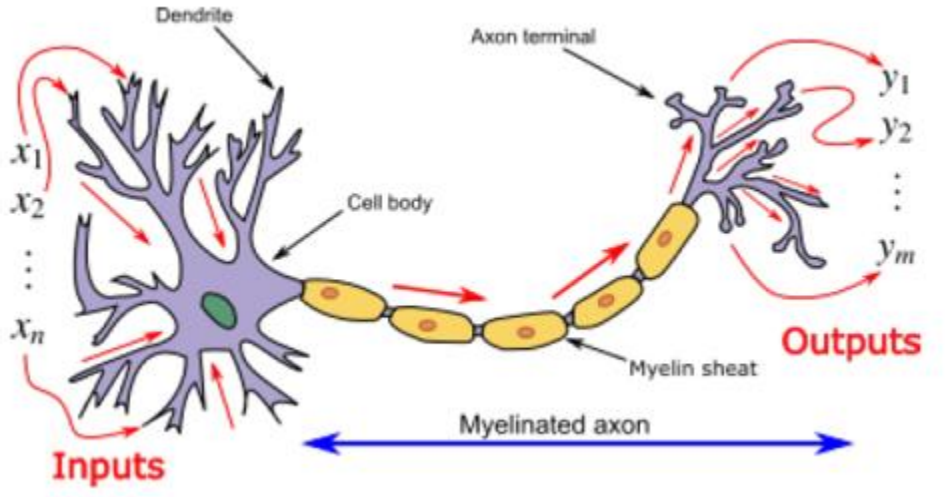
Σχήμα 2.13: Σχηματική αναπαράσταση της επιβλεπόμενης μάθησης.

2.4.4 Νευρωνικά δίκτυα

Τα τεχνητά νευρωνικά δίκτυα(Artificial Neural Networks), η πιο απλά τα νευρωνικά δίκτυα είναι υπολογιστικές μηχανές εμπνευσμένες από τα βιολογικά δίκτυα νευρώνων που βρίσκονται στους εγκεφάλους των ζώων [35]. Τα συστήματα αυτά εμφανίζουν την ικανότητα να μαθαίνουν από παραδείγματα που τους παρέχουμε και καταλήγουν έτσι να επιλύουν προβλήματα που δεν ήταν σαφώς προγραμματισμένα να το κάνουν εξαρχής [36]. Η τεράστια προσαρμοστικότητα τους σε διαφορετικά προβλήματα οδηγεί όλο και περισσότερους τους ερευνητές στην χρήση τους. Τα τελευταία χρόνια εμφανίζονται όλο και περισσότερες εφαρμογές σε τομείς από την ιατρική και τα μαθηματικά μέχρι την γλωσσολογία και τις γνωσιακές επιστήμες [37] .

2.4.5 Δομή βιολογικού νευρώνα

Ένα νευρωνικό δίκτυο βασίζεται σε μια συλλογή διασυνδεδεμένων νευρώνων ή κόμβων που αποκαλούνται τεχνητοί νευρώνες, και οι οποίοι μοντελοποιούν σε έναν βαθμό την λειτουργία των πραγματικών νευρώνων σε έναν εγκέφαλο. Κάθε σύνδεση , όπως και οι συνάψεις σε έναν εγκέφαλο, μεταδίδουν σήματα σε άλλους νευρώνες. Η σύνδεση μεταξύ τεχνητών νευρώνων περιγράφει την βασική βιολογική σύνδεση μεταξύ άξονα-σύναψης-δενδρίτη όπου σήματα προηγούμενων νευρώνων προωθούνται στους επόμενους μέσω των εν λόγω συνδέσεων. Ένας βιολογικός νευρώνας λαμβάνει το ηλεκτρικό σήμα, το επεξεργάζεται και στην συνέχεια το οδηγεί στους επόμενους νευρώνες με τους οποίους συνδέεται μέσω κατάλληλων χημικών διεργασιών. Τα δύο βασικά βήματα στην λειτουργία ενός νευρώνα λοιπόν είναι η επεξεργασία των σημάτων εισόδου και η δημιουργία ενός σήματος εξόδου που δίνεται ως είσοδο στα επόμενα επίπεδα. Τόσο η συσχέτιση των εισόδων ενός νευρώνα όσο και η τελική ενεργοποίηση της εξόδου του μπορούν να αποτυπωθούν μαθηματικά με χρήση κατάλληλων συναρτήσεων και να περάσουμε έτσι από το πεδίο της βιολογίας στο πεδίο της θεωρίας υπολογισμού.



Σχήμα 2.14: Νευρώνας και μυελιωμένος άξονας, με τη ροή του σήματος από τις εισόδους στους δενδρίτες στις εξόδους στον άξονα.

2.4.6 Μαθηματική δομή τεχνητού νευρώνα

Οι βασικές λειτουργίες ενός τεχνητού νευρώνα είναι σε πρώτο βαθμό η επεξεργασία του σήματος εισόδου και σε δεύτερο η τελική ενεργοποίηση του. Η αρχική επεξεργασία πραγματοποιείται με τη χρήση ενός βεβαρημένου αθροίσματος, όπου κάθε σήμα εισόδου πολλαπλασιάζεται με ένα αντίστοιχο βάρος για να καθοριστεί το πόση σημασία έχει στον καθορισμό της εξόδου. Σε δεύτερο χρόνο υπολογίζεται μια μη γραμμική συνάρτηση ενεργοποίησης η οποία δίνει τελικά τη δυνατότητα στο σύστημα να προσεγγίσει μη γραμμικές συναρτήσεις. Αν θεωρήσουμε ένα σήμα εισόδου N σημείων x_n τότε η όλη διαδικασία περιγράφεται μαθηματικά ως:

$$y = f\left(\sum (x_1 \cdot w_1 + x_2 \cdot w_2 + \cdots + x_n \cdot w_n)\right) \tag{25}$$

Όπου

x_n = οι είσοδοι του νευρώνα
w_n = τα βάρη του νευρώνα
$f(x)$ = η συνάρτηση ενεργοποίησης
y = η έξοδος του νευρώνα

Από την εξίσωση (25) παρατηρούμε ότι οι παράμετροι που καθορίζουν την συμπεριφορά ενός νευρώνα είναι, πρώτον τα βάρη του w_n που προσδιορίζουν το σε ποιες εισόδους θα δώσει σημασία ο νευρώνας αλλά και πόση. Οι συντελεστές αυτοί καθορίζονται κατά την εκπαίδευση του συστήματος. Δεύτερον η συνάρτηση ενεργοποίησης $f(x)$ η οποία καθορίζει την μορφή της εξόδου του συστήματος και έτσι η επιλογή της μπορεί να αλλάξει ανάλογα με το πρόβλημα που αντιμετωπίζουμε.

Συναρτήσεις ενεργοποίησης νευρώνων:

- Βηματικές συναρτήσεις: Η έξοδος αυτής της συνάρτησης είναι δυαδική και εξαρτάται από το αν το βεβαρημένο άθροισμα στο σήμα εισόδου ξεπερνά ή όχι ένα προκαθορισμένο κατώφλι θ. Πρακτικά εκτελεί μια διαίρεση του χώρου των εισόδων του νευρώνα με τη χρήση ενός υπερεπιπέδου.

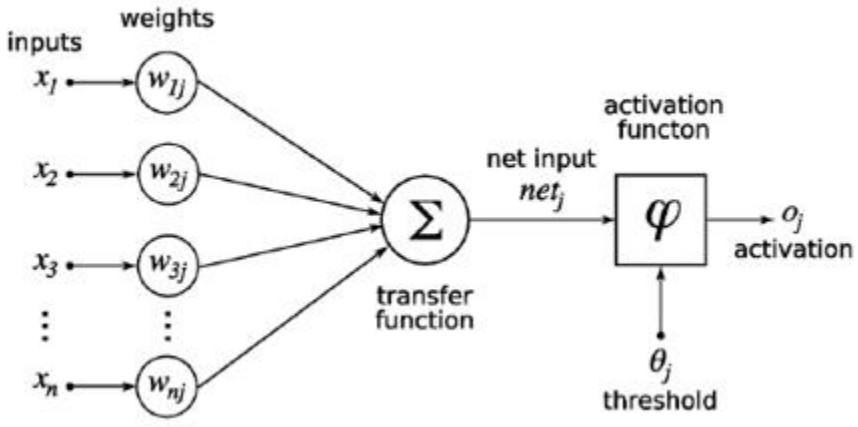
$$y = \begin{Bmatrix} 0\ if\ \sum(x_n * y_n) < \theta \\ 1\ if\ \sum(x_n * y_n) > \theta \end{Bmatrix} \tag{26}$$

- Σιγμοειδείς συναρτήσεις: πρόκειται για φραγμένες, διαφορίσιμες, μη γραμμικές συναρτήσεις που χρησιμοποιούνται κυρίως για την υπολογιστική τους απλότητα. Ένα παράδειγμα είναι η logistic function που έχει την ακόλουθη μορφή:

$$y = \frac{1}{1 + e^{-x}} = \frac{e^x}{e^x + 1} \tag{27}$$

- Συναρτήσεις γραμμικής ανόρθωσης: αποτελούν την πλέον διαδεδομένη κατηγορία συναρτήσεων ενεργοποίησης για πολλούς τύπους νευρωνικών δικτύων, έχουν υπολογιστική απλότητα και αντιμετωπίζουν προβλήματα διαφόρισης που εμφανίζουν οι σιγμοειδείς συναρτήσεις(vanishing gradient κ.α.) Ο μαθηματικός ορισμός έχει ως εξής:

$$y = \max(0, x) \tag{28}$$



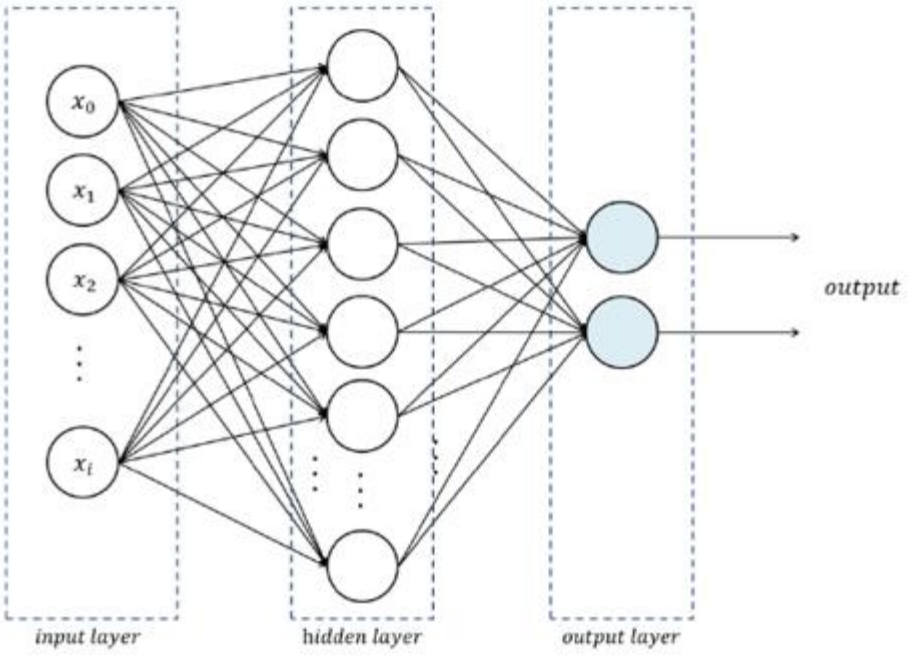
Σχήμα 2.15: Μαθηματική δομή τεχνητού νευρώνα.

2.4.7 Αρχιτεκτονικές νευρωνικών δικτύων

Έχοντας ορίσει λοιπόν την στοιχειώδη συμπεριφορά ενός νευρώνα περνάμε στην ανάλυση της οργάνωσης ενός νευρωνικού δικτύου. Βασικό χαρακτηριστικό , όπως και στην περίπτωση των νευρικών κυττάρων του εγκεφάλου, είναι η οργάνωση τους σε δίκτυα. Πιο συγκεκριμένα ένα ANN αποτελείται από επίπεδα(layers) νευρώνων όπου οι έξοδοι ενός επιπέδου αποτελούν τις εισόδους του επόμενου. Ο ακριβής τρόπος σύνδεσης τους καθορίζεται από την εκάστοτε αρχιτεκτονική. Υπάρχουν περιπτώσεις όπου όλες οι έξοδοι γίνονται είσοδοι στα επόμενα επίπεδα, σε αυτή την περίπτωση μιλάμε για πλήρως διασυνδεδεμένες αρχιτεκτονικές. Υπάρχουν περιπτώσεις όπου προηγούμενες έξοδοι επιστρέφουν ως είσοδοι στο σύστημα μέσω σημάτων ανατροφοδότησης,(recursive neural networks) που σκοπός είναι να εντοπίζουν πιθανές συσχετίσεις σε ακολουθίες δεδομένων όπως τα ηχητικά σήματα. Ακόμη υπάρχουν αρχιτεκτονικές όπου τα βάρη εισόδου δεν έχουν μονοδιάστατη μορφή αλλά μεγαλύτερη διάσταση με σκοπό την εφαρμογή τους σε σήματα μεγαλύτερης πολυπλοκότητας(εικόνα,βίντεο) και καλούνται συνελικτικά νευρωνικά δίκτυα(CNN). Καταλαβαίνουμε λοιπόν ότι υπάρχει πληθώρα τρόπων διασύνδεσης, ενεργοποίησης και εκπαίδευσης των δικτύων αυτών για να επιλέξουμε ανάλογα με τις ανάγκες του κάθε προβλήματος. Η βασική όμως αρχή είναι πως επιλέγοντας ένα αρκετά μεγάλο σύνολο νευρώνων, ανεξάρτητα από το πόσα επίπεδα υπάρχουν και πως συνδέονται, μπορούμε να περιγράψουμε οποιαδήποτε μη γραμμική σχέση μεταξύ των δεδομένων εισόδου και εξόδου οσοδήποτε κοντά επιθυμούμε [38]. Το θεώρημα αυτό ονομάζεται universal approximation theorem και εξηγεί ακριβώς για ποιο λόγο οι εφαρμογές νευρωνικών δικτύων αποτελούν τις καλύτερες λύσεις μέχρι τώρα σε μια τεράστια γκάμα προβλημάτων που εντάσσονται γενικότερα στα curve fitting προβλήματα.

2.4.8 Multilayer Perceptron

Η απλούστερη δομή νευρωνικών δικτύων είναι ο Multilayer Perceptron [39]. Αποτελεί ένα πλήρως διασυνδεδεμένο εμπρόσθιας διάδοσης σύστημα όπου η πληροφορία ρέει από το επίπεδο εισόδου προς την έξοδο. Αποτελείται από τουλάχιστον τρία επίπεδα, το επίπεδο εισόδου , το κρυφό επίπεδο και το επίπεδο εξόδου όπως φαίνεται και στο σχήμα 2.16. Το επίπεδο εισόδου δέχεται το σήμα που θέλουμε να χειριστούμε. Το επίπεδο εξόδου είναι υπεύθυνο για λειτουργίες όπως η κατηγοριοποίηση(classification) και η πρόβλεψη(prediction). Τα ενδιάμεσα κρυφά επίπεδα συνδέουν την είσοδο με την έξοδο μέσα από ένα σύνολο γραμμικών μετασχηματισμών και μη γραμμικών ενεργοποιήσεων. Στην περίπτωση που ο αριθμός των κρυφών επιπέδων είναι μεγάλος μιλάμε για βαθιά νευρωνικά δίκτυα. Ο MLP μπορεί να χειριστεί μη γραμμικά διαχωρίσιμα δεδομένα και να προσεγγίσει ικανοποιητικά πραγματικές μη γραμμικές συναρτήσεις.



Σχήμα 2.16: Σχηματική δομή ενός Multilayer Perceptron.

2.4.9 Εκπαίδευση δικτύων

Η εκπαίδευση ενός νευρωνικού δικτύου πραγματοποιείται μέσω της αλλαγής των βαρών του δικτύου χρησιμοποιώντας την πληροφορία του σφάλματος που έχουμε στην εκτίμηση των δεδομένων εκπαίδευσης. Για κάθε ένα σύνολο από τα παραδείγματα εκπαίδευσης μπορούμε μέσω μια συνάρτησης σφάλματος να αξιολογήσουμε την απόσταση της εξόδου του δικτύου από την επιθυμητή τιμή. Καταλήγουμε έτσι σε ένα πρόβλημα βελτιστοποίησης όπου οι συντελεστές του δικτύου πρέπει να μεταβληθούν για να ελαχιστοποιήσουμε το μέσο αναμενόμενο σφάλμα.

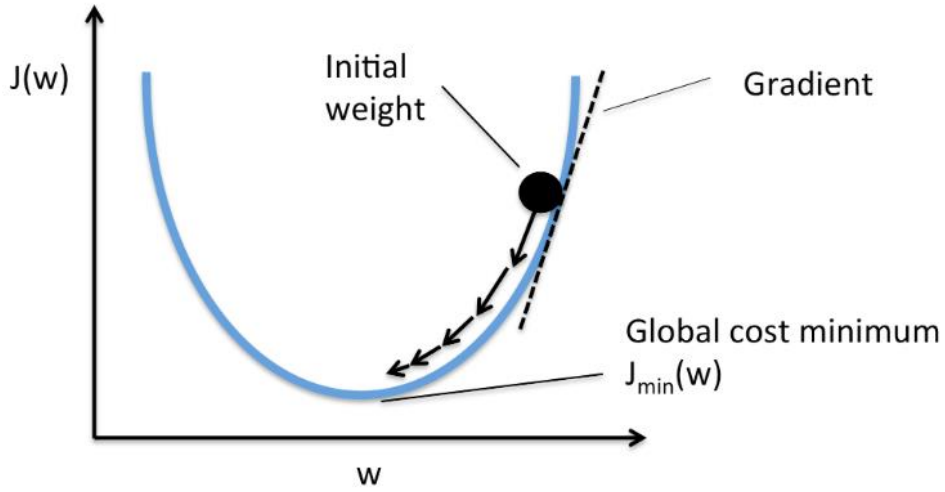
2.4.10 Ο αλγόριθμος Gradient Descent

Ο αλγόριθμος Gradient Descent αποτελεί μία πρώτης τάξης επαναληπτική μέθοδο που χρησιμοποιείται κατά κόρον σε προβλήματα βελτιστοποίησης και κατ' επέκταση στην εκπαίδευση νευρωνικών δικτύων. Η μέθοδος βασίζεται στην παρατήρηση πως αν έχουμε μια πολυδιάστατη συνάρτηση $F(x)$ που ορίζεται και είναι διαφορίσιμη στην γειτονιά ενός διανύσματος w , τότε η $F(x)$ μειώνεται γρηγορότερα αν από το αρχικό σημείο w κινηθούμε κατά την αρνητική τιμή της παραγώγου της $F(x)$ στο σημείο αυτό,$-\nabla F(w)$.

$$w_{n+1} = w_n - \gamma \nabla F(w_n) \tag{29}$$

Συνεπώς καταφέρνουμε με τη χρήση μιας συνάρτησης σφάλματος να ορίσουμε πόσο λάθος κάνει ένα δίκτυο και με την αξιολόγηση της παραγώγου να βρούμε την κατεύθυνση ως προς

την οποία αν το αλλάξουμε θα αρχίσει επαναληπτικά να μειώνει το σφάλμα.



Σχήμα 2.17: Γραφική αναπαράσταση του αλγορίθμου Gradient Descent.

2.4.11 Ο αλγόριθμος Backpropagation of error

Όπως αναφέραμε η εκπαίδευση ενός νευρωνικού δικτύου ανάγεται σε ένα πρόβλημα βελτιστοποίησης. Ο υπολογισμός της κλίσης της συνάρτησης σφάλματος ως προς τον εκάστοτε συντελεστή του δικτύου καθορίζει την κατεύθυνση ελαχιστοποίησης της. Για να υπολογίσουμε όμως την κλίση ως προς τα βάρη όλων των επιπέδων ενός δικτύου από την είσοδο μέχρι την έξοδο του επιστρατεύουμε τον αλγόριθμο της οπισθοδρομικής διάδοσης του σφάλματος(backpropagation of error) [40]. Ο αλγόριθμος αυτός με χρήση του κανόνα της αλυσίδας του Leibnitz υπολογίζει επαναληπτικά την κλίση της συνάρτησης σφάλματος ως προς τους συντελεστές κάθε επιπέδου(layer) του δικτύου, ξεκινώντας από το επίπεδο εξόδου και πηγαίνοντας προς τα πίσω. Πιο αναλυτικά αν ορίσουμε:

$$E = L(t, y) \tag{30}$$

Όπου

L είναι το σφάλμα για την έξοδο y και την επιθυμητή τιμή t
t είναι η επιθυμητή τιμή για το αντίστοιχο δείγμα εκπαίδευσης
y είναι η έξοδος του δικτύου για το αντίστοιχο δείγμα εισόδου.

Για κάθε νευρώνα j , η έξοδος του o_j ορίζεται ώς

$$o_j = \varphi(net_j) = \varphi\left(\sum_{k=1}^{n} w_{kj}\, x_k\right) \tag{31}$$

Όπου η συνάρτηση φ είναι μη γραμμική και διαφορίσιμη στην περιοχή ενεργοποίησης της. Η είσοδος net_j σε έναν νευρώνα είναι το βεβαρημένο άθροισμα των εξόδων o_k των προηγούμενων νευρώνων. Αν ο νευρώνας βρίσκεται στο πρώτο επίπεδο τότε οι είσοδοι o_k

είναι απλά οι είσοδοι x_k του δικτύου. Η μεταβλητή w_{kj} εκφράζει το βάρος που συνδέει τον νευρώνα k του προηγούμενου επιπέδου με τον νευρώνα j του τωρινού επιπέδου.
Ο υπολογισμός της μερικής παραγώγου του σφάλματος ως προς τους συντελεστές πραγματοποιείται με χρήση του κανόνα της αλυσίδας δύο φορές:

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial o_j} \cdot \frac{\partial o_j}{\partial w_{ij}} = \frac{\partial E}{\partial o_j} \cdot \frac{\partial o_j}{\partial net_j} \cdot \frac{\partial net_j}{\partial w_{ij}} \tag{32}$$

Στον τελευταίο παράγοντα της σχέσης μόνο ένας όρος, αυτός της εσωτερικής κατάστασης net_j, εξαρτάται από το βάρος w_{ij} έτσι έχουμε:

$$\frac{\partial net_j}{\partial w_{ij}} = \frac{\partial\left(\sum_{k=1}^{n} w_{kj}\, o_k\right)}{\partial w_{ij}} = \frac{\partial w_{ij} o_i}{\partial w_{ij}} = o_i \tag{33}$$

Αν ο νευρώνας βρίσκεται στο πρώτο επίπεδο τότε το o_i είναι απλά η είσοδος x_i.
Η παράγωγος του νευρώνα j ως προς την είσοδο του είναι απλά η παράγωγος της συνάρτησης ενεργοποίησης. Αυτός είναι και ο λόγος χρήσης διαφορίσιμων συναρτήσεων.

$$\frac{\partial o_j}{\partial net_j} = \frac{\partial \varphi\left(net_j\right)}{\partial net_j} \tag{34}$$

Για τον πρώτο παράγοντα της σχέσης ο υπολογισμός είναι απλός αν ο νευρώνας βρίσκεται στο επίπεδο εξόδου γιατί τότε ισχύει $o_j = y$ και

$$\frac{\partial E}{\partial o_j} = \frac{\partial E(y)}{\partial y} \tag{35}$$

Αν όμως ο νευρώνας j βρίσκεται σε κάποιο εσωτερικό επίπεδο του δικτύου, ο υπολογισμός της παραγώγου του E ως προς το o_j δεν είναι τόσο προφανής. Αν θεωρήσουμε την E ως μια συνάρτηση με εισόδους όλους τους νευρώνες $L = \{u, v, \dots, w\}$ που δέχονται ως είσοδο την έξοδο του νευρώνα j τότε

$$\frac{\partial E\left(o_j\right)}{\partial o_j} = \frac{\partial E(net_u, net_v, \dots, net_w)}{\partial o_j} \tag{36}$$

Και υπολογίζοντας την ολική παράγωγο ως προς το o_j, προκύπτει μια αναδρομική σχέση για την παράγωγο:

$$\frac{\partial E}{\partial o_j} = \sum_{\ell \in L}\left(\frac{\partial E}{\partial net_\ell} \cdot \frac{\partial net_\ell}{\partial o_j}\right) = \sum_{\ell \in L}\left(\frac{\partial E}{\partial o_\ell} \cdot \frac{\partial o_\ell}{\partial net_\ell} \cdot \frac{\partial net_\ell}{\partial o_j}\right) = \sum_{\ell \in L}\left(\frac{\partial E}{\partial o_\ell} \cdot \frac{\partial o_\ell}{\partial net_\ell} \cdot w_{j\ell}\right) \tag{37}$$

Καταλήγουμε έτσι να μπορούμε να υπολογίσουμε την παράγωγο ως προς το o_j αν όλες οι παράγωγοι ως προς τις εξόδους o_ℓ του επόμενου επιπέδου, που βρίσκονται πιο κοντά στο επίπεδο εξόδου, είναι γνωστές.

Αντικαθιστώντας τις σχέσεις στην έχουμε:

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial o_j} \cdot \frac{\partial o_j}{\partial net_j} \cdot \frac{\partial net_j}{\partial w_{ij}} = \frac{\partial E}{\partial o_j} \cdot \frac{\partial o_j}{\partial net_j} \cdot o_i$$

$$\frac{\partial E}{\partial w_{ij}} = o_i \cdot \delta_j$$

Με

$$\delta_j = \frac{\partial E}{\partial o_j} * \frac{\partial o_j}{\partial net_j} = \begin{cases} \frac{\partial l(o_j, t)}{\partial o_j} * \frac{d\varphi(net_j)}{dnet_j} \; if\ j\ is\ an\ output\ neuron, \\ \left(\sum_{\ell \in \mathcal{L}} w_{j\ell}\, \delta_\ell\right) * \frac{d\varphi(net_j)}{dnet_j} \; if\ j\ is\ an\ inner\ neuron. \end{cases} \quad (38)$$

Έχοντας υπολογίσει την κλίση της συνάρτησης σφάλματος ως προς το κάθε ξεχωριστό βάρος w_{ij} μπορούμε τελικά, με χρήση του gradient descent και με την επιλογή ενός κατάλληλου learning rate $\eta > 0$, να βρούμε τις διαφορές Δw_{ij} που αν προστεθούν στο κάθε βάρος θα μειώσουν το συνολικό σφάλμα του συστήματος.

$$\Delta w_{ij} = -\frac{\eta \partial E}{\partial w_{ij}} = -\eta o_i \delta_j \quad (39)$$

2.4.12 Ο αλγόριθμος Adam

Έχουμε καθορίσει λοιπόν τόσο την γενική λογική πίσω από την βελτιστοποίηση ενός νευρωνικού δικτύου μέσω του αλγορίθμου Gradient Descent όσο και τον τρόπο υπολογισμού της απαιτούμενης κλίσης για δίκτυα με βάθος μέσω του αλγορίθμου Backpropagation of error. Το βασικότερο πρόβλημα που εμφανίζει ο Gradient Descent είναι πως το learning rate η παραμένει σταθερό κατά την εκπαίδευση. Σε περιοχές όπου η συνάρτηση σφάλματος έχει έντονη κλίση χρειάζεται να επιλέξουμε μικρό learning rate ώστε να αποφύγουμε πιθανά άλματα που θα μας οδηγήσουν ίσως μακριά από την περιοχή ελαχίστου που βρισκόμαστε ή πιθανότατα σε μια ταλάντωση γύρω από αυτό. Αυτό έχει ως συνέπεια κοντά στο ελάχιστο, που η παράγωγος τείνει στο μηδέν, να κινούμαστε απειροελάχιστα και να αυξάνεται έτσι κατά πολύ ο χρόνος σύγκλισης του αλγορίθμου(vanishing gradient). Στο πρόβλημα αυτό στην περίπτωση του MLP απαντάμε με τον αλγόριθμο Adam(Adaptive Moment Estimation) [41] .

Ο Adam προσαρμόζει διαφορετικά learning rates για κάθε παράμετρο χρησιμοποιώντας τις εκτιμήσεις των ροπών πρώτης και δεύτερης τάξης του gradient. Από τη μία, με χρήση της μεθόδου momentum, ο αλγόριθμος θυμάται τις προηγούμενες αλλαγές και υπολογίζει ένα εκθετικά βεβαρημένο μέσο όρο ώστε όταν οι αλλαγές ευθυγραμμίζονται να επιταχύνεται η σύγκλιση προς αυτή την κατεύθυνση. Από την άλλη υπολογίζοντας έναν βεβαρημένο μέσο όρο των τετραγώνων των προηγούμενων gradients, αξιολογούμε το πόσο απότομη είναι η συνάρτηση σφάλματος στην περιοχή που βρισκόμαστε, και προσαρμόζουμε το βήμα ανάλογα. Αν η τιμή είναι μεγάλη μικραίνουμε το βήμα ώστε να αποφύγουμε πιθανή υπέρβαση του ελαχίστου που αναζητούμε. Ακολουθούν οι εξισώσεις για τον υπολογισμό των επιθυμητών αλλαγών.

$$m_w^{(t+1)} = \beta_1 m_w^{(t)} + (1-\beta_1)\nabla_w L^{(t)} \quad (40)$$

$$v_w^{(t+1)} = \beta_2 v_w^{(t)} + (1-\beta_2)\left(\nabla_w L^{(t)}\right)^2 \quad (41)$$

$$\hat{m}_w = \frac{m_w^{(t+1)}}{(1-\beta_1^t)} \quad (42)$$

$$\hat{v}_w = \frac{v_w^{(t+1)}}{(1-\beta_2^t)} \quad (43)$$

$$w^{(t+1)} = w^t - \eta \frac{\hat{m}_w}{\sqrt{\hat{v}_w} + \epsilon} \quad (44)$$

Όπου το ϵ είναι μια μικρή τιμή, συνήθως $\epsilon = 10^{-8}$, που χρησιμοποιείται ώστε να αποφύγουμε την διαίρεση με το 0. Οι παράγοντες β_1, β_2 καλούνται συντελεστές λήθης και καθορίζουν σε πόσες παρελθοντικές τιμές θα δώσει σημασία αλλά και πόση τελικά θα είναι αυτή. Συνήθεις τιμές είναι $\beta_1 = 0.9$, $\beta_2 = 0.999$.

2.4.13 Radial Basis Function networks

Τα Radial Basis Function networks είναι ένα είδος νευρωνικών δικτύων που χρησιμοποιούν ακτινικές συναρτήσεις βάσης ως συνάρτηση ενεργοποίησης [42] . Η έξοδος τους είναι ένας γραμμικός συνδυασμός από τις εξόδους των ακτινικών συναρτήσεων βάσης που εφαρμόζονται στα δεδομένα εισόδου. Εμφανίζουν και αυτά πληθώρα εφαρμογών όπως προσέγγιση συναρτήσεων, πρόβλεψη χρονοσειρών, ταξινόμηση και έλεγχο συστημάτων μεταξύ άλλων.

Τα RBF δίκτυα αποτελούνται από 3 επίπεδα: το επίπεδο εισόδου, ένα κρυφό επίπεδο με μη γραμμικές ακτινικές συναρτήσεις βάσης και ένα γραμμικό επίπεδο εξόδου. Η είσοδος περιγράφεται ως ένα διάνυσμα εισόδου πραγματικών αριθμών διάστασης n ,$x \in \mathbb{R}^n$. Η έξοδος του δικτύου είναι τις πιο πολλές φορές μια βαθμωτή συνάρτηση του διανύσματος εισόδου της μορφής $\varphi: \mathbb{R}^n \to \mathbb{R}$ και υπολογίζεται:

$$\varphi(x) = \sum_{i=1}^{N} a_i \, \rho(\|x - c_i\|) \quad (45)$$

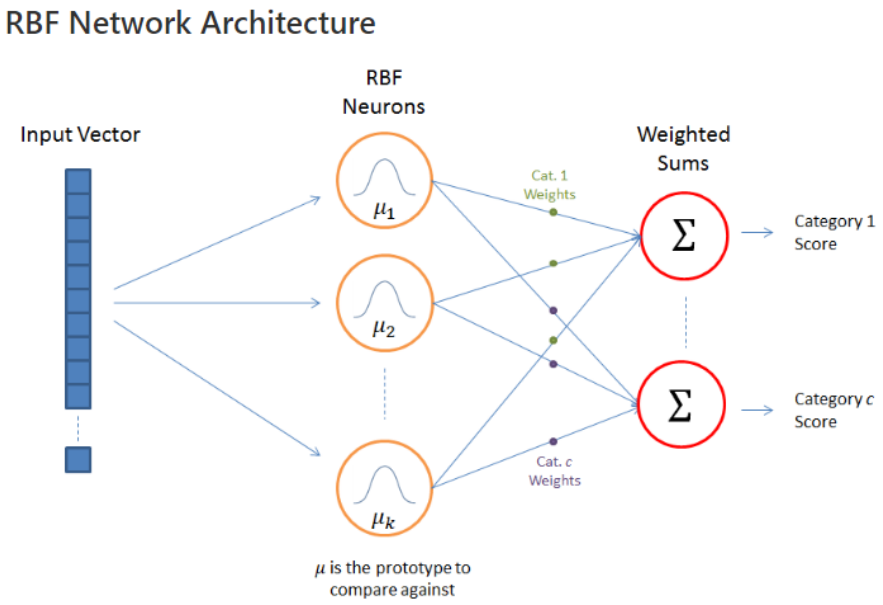
Όπου N είναι ο αριθμός των νευρώνων του κρυφού επιπέδου, c_i είναι το διάνυσμα-κέντρο του νευρώνα i , και a_i είναι το αντίστοιχο βάρος του νευρώνα στο γραμμικό επίπεδο εξόδου. Η μορφή του μπορεί να τροποποιηθεί και στην έξοδο να υπολογίζει πολλαπλούς γραμμικούς συνδυασμούς των προηγούμενων εξόδων δίνοντας μια διανυσματική συνάρτηση ως έξοδο της μορφής $\varphi: \mathbb{R}^n \to \mathbb{R}^k$. Οι συναρτήσεις εξαρτώνται μόνο από την απόσταση από το διάνυσμα-κέντρο και είναι ακτινικά συμμετρικές ως προς αυτό. Στην βασική του μορφή, όλες οι είσοδοι συνδέονται με κάθε νευρώνα του κρυφού επιπέδου. Η νόρμα επιλέγεται συνήθως να είναι η ευκλείδια απόσταση(αν και η απόσταση mahalanobis εμφανίζει καλύτερα αποτελέσματα σε προβλήματα αναγνώρισης προτύπων) και οι ακτινικές συναρτήσεις βάσεις είναι γκαουσιανές:

$$\rho(\|x - c_i\|) = \exp[-\beta_i \|x - c_i\|^2] \quad (46)$$

Οι γκαουσιανές συναρτήσεις βάσεις είναι τοπικές υπό την έννοια ότι:

$$\lim_{\|x\| \to \infty} \rho(\|x - c_i\|) = 0 \quad (47)$$

Δηλαδή η ενεργοποίηση ενός νευρώνα επηρεάζεται ελάχιστα από τα διανύσματα εισόδου που βρίσκονται μακριά από το κέντρο του εν λόγω νευρώνα. Οι συντελεστές α_i, β_i και c_i καθορίζουν την τελική προσέγγιση του συστήματος και διαμορφώνονται κατά την εκπαίδευση.



Σχήμα 2.18: Σχηματική δομή ενός Radial basis neural network.

2.4.14 Ο αλγόριθμος Levenberg-Marquardt

Για την εκπαίδευση τέτοιων δικτύων μπορούμε να χρησιμοποιήσουμε την βασική λογική του gradient descent αλγορίθμου. Όμως λόγω του μικρού μεγέθους τους μπορούν να χρησιμοποιηθούν υπολογιστικά πολυπλοκότερες μέθοδοι που ξεπερνούν κάποια από τα βασικά προβλήματα που εμφανίζονται. Για τον σκοπό αυτό έχει ο αναπτυχθεί ο αλγόριθμος Levenberg-Marquardt ο οποίος σε προβλήματα μεσαίου μεγέθους ξεπερνά σε απόδοση τον βασικο gradient descent αλγόριθμο [43] .

Όπως αναφέραμε το βασικό πρόβλημα του gradient descent πηγάζει από το σταθερό step size. Θα θέλαμε ιδανικά σε περιοχές με έντονη κλίση, μακριά από το ελάχιστο που αναζητούμε, να παίρνει μικρότερες τιμές και να εξομαλύνουμε έτσι τα μεγάλα άλματα, και σε περιοχές κοντά στη λύση να μεγαλώνει για να αντισταθμίσει την μικρή της παραγώγου

και να φτάσει έτσι γρηγορότερα στο ελάχιστο. Λύση σε αυτό το πρόβλημα δίνει ο αλγόριθμος Gauss-Newton ο οποίος εκτιμά την τιμή της δεύτερης παραγώγου, δηλαδή την κυρτότητα στην περιοχή που βρισκόμαστε, και μέσω μιας προσέγγισης δεύτερης τάξης, προσαρμόζει αντίστοιχα τα βήματα. Πιο αναλυτικά:

Έστω μια συνάρτηση της μορφής $f(x,\beta)$ όπου χ τα δεδομένα εισόδου και β οι συντελεστές που θέλουμε να προσδιορίσουμε. Σκοπός μας είναι να υπολογίσουμε μια μεταβολή δ ώστε οι νέοι συντελεστές $\beta+\delta$ να βελτιώνουν την εκτίμηση του συστήματος μας. Παίρνοντας την γραμμική προσέγγιση της $f(x,\beta+\delta)$ στην περιοχή του β έχουμε:

$$f(x_i,\beta+\delta) \approx f(x_i,\beta) + J_i\delta \tag{48}$$

Όπου

$$J_i = \frac{\partial f(x_i,\beta)}{\partial \beta} \tag{49}$$

Είναι η κλίση της f ως προς το β.
Το άθροισμα των τετραγωνικών διαφορών $S(\beta)$ παίρνει την ελάχιστη τιμή του όταν η παράγωγος του μηδενίζεται. Η παραπάνω προσέγγιση πρώτης τάξης μας δίνει:

$$\begin{aligned} S(\beta+\delta) &\approx \|y - f(\beta) - J\delta\|^2 \\ &= |y - f(\beta) - J\delta|^{\mathrm{T}}|y - f(\beta) - J\delta| \\ = |y - f(\beta)|^{\mathrm{T}}|y - f(\beta)| - |y - f(\beta)|^{\mathrm{T}}J\cdot\delta - (J\cdot\delta)^{\mathrm{T}}|y - f(\beta)| + \delta^{\mathrm{T}}J^{\mathrm{T}}J\delta \\ &= |y - f(\beta)|^{\mathrm{T}}|y - f(\beta)| - 2|y - f(\beta)|^{\mathrm{T}}J\cdot\delta + \delta^{\mathrm{T}}J^{\mathrm{T}}J\delta \end{aligned} \tag{50}$$

Παίρνοντας την παράγωγο του $S(\beta+\delta)$ ως προς το δ και θέτοντας το αποτέλεσμα ίσο με το μηδέν καταλήγουμε στη σχέση

$$(J^TJ)\delta = J^T[y - f(\beta)] \tag{51}$$

Όπου J είναι η ιακωβιανή μήτρα. Ο πίνακας J δεν είναι συνήθως τετραγωνικός αλλά της μορφής mxn όπου n είναι ο αριθμός των συντελεστών(η διάσταση του β). Η προσέγγιση του Hessian πίνακα (J^TJ) δημιουργεί έναν τετραγωνικό πίνακα nxn καταλήγοντας έτσι σε ένα γραμμικό σύστημα n αγνώστων και n εξισώσεων. Η προσφορά των Levenberg-Marquardt είναι η αντικατάσταση της εξίσωσης (45) με την:

$$\left(J^{\mathrm{T}}J + \lambda I\right)\delta = J^T[y - f(\beta)] \tag{52}$$

Η ως προς το δ που είναι και το ζητούμενο

$$\delta = \left(J^{\mathrm{T}}J + \lambda I\right)^{-1}J^T[y - f(\beta)] \tag{53}$$

Η προσθήκη του παράγοντα λI φροντίζει ώστε ο πίνακας πάντα να αντιστρέφεται. Σε περιοχές μακριά από τη λύση που η κυρτότητα δεν είναι έντονη ή μεταβάλλεται ο πίνακας J^TJ παίρνει μικρές τιμές έτσι μεγαλώνοντας το λ κάνουμε βήμα με βάση τον gradient descent. Πλησιάζοντας το ελάχιστο όμως χρησιμοποιούμε την πληροφορία της κυρτότητας αυτής της περιοχής, μικραίνουμε το λ, και μέσω μιας δευτεροβάθμιας προσέγγισης του σφάλματος στην εν λόγω περιοχή κάνουμε μεγαλύτερη "βουτιά" για να βρούμε το ελάχιστο.

2.4.15 Συναρτήσεις Απώλειας

Οι αλγόριθμοι που περιγράψαμε βασίζονται στην επιλογή μιας συνάρτησης απώλειας, η οποία αξιολογεί την απόσταση της εξόδου του συστήματος μας από την επιθυμητή, και στην ελαχιστοποίηση της ώστε να πετύχουμε την καλύτερη δυνατή προσαρμογή. Η χρήση διαφορετικών μετρικών μας δίνει την δυνατότητα να αποτυπώσουμε με διαφορετικούς τρόπους την επιθυμητή ομοιότητα των δεδομένων μας, και να αξιολογήσουμε έτσι πιο σφαιρικά την γενική συμπεριφορά των συστημάτων μας.

Μέσο απόλυτο σχετικό σφάλμα

Το μέσο απόλυτο σχετικό σφάλμα(MARE) υπολογίζει έναν μέσο όρο των απόλυτων τιμών των σχετικών διαφορών μεταξύ των δεδομένων. Υπολογίζει ουσιαστικά ένα ποσοστό λάθους ανά δείγμα γεγονός που το καθιστά ανεξάρτητο από την κλίμακα των δεδομένων. Το διάστημα τιμών είναι $D = [0\,,1]$ και υπολογίζεται από την σχέση:

$$MARE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - x_i}{x_i}\right| \tag{54}$$

Ρίζα μέσου τετραγωνικού σχετικού σφάλματος

Η ρίζα του μέσου τετραγωνικού σχετικού σφάλματος (relative Root Mean Square Error) είναι μια μετρική αξιολόγησης της απόστασης μεταξύ ενός συνόλου μεταβλητών στόχου X και των αντίστοιχων εκτιμήσεων Y. Για ένα σύνολο από n παρατηρήσεις υπολογίζεται από την εξίσωση (55).. Το πεδίο τιμών της είναι και αυτό $D = [0\,,1]$.

$$rRMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{(y_i - x_i)}{x_i}\right)^2} \tag{55}$$

Ρίζα μέσου τετραγωνικού σφάλματος

Το RMSE είναι η ρίζα του μέσου όρου των τετραγώνων των διαφορών μεταξύ των μεταβλητών μας. Έχει πεδίο ορισμού στην ίδια κλίμακα με τα δεδομένα μας, δηλαδή $D = [0\,,+\infty)$. Υπολογίζεται από την σχέση:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2} \tag{56}$$

ΚΕΦΑΛΑΙΟ 3

Υλοποίηση

Στο κεφάλαιο 2 αναλύθηκε με λεπτομέρεια το θεωρητικό υπόβαθρο στο οποίο βασίστηκε η εργασία. Εδώ περιγράφεται η προσέγγιση που επιλέχθηκε για την επίλυση του προβλήματος της ανακατασκευής πολυφασματικών εικόνων από τις αντίστοιχες RGB. Αρχικά αναλύεται η αραιότητα των πολυφασματικών σημάτων που αποτελεί και το βασικότερο στοιχείο της μεθόδου μας. Στην συνέχεια ορίζονται οι χώροι των διαφορετικών αναπαραστάσεων και οι απεικονίσεις μεταξύ αυτών. Ακολουθεί η διαδικασία εξαγωγής των φασματικών υπογραφών, η σύνθεση των αντίστοιχων RGB και PCA δεδομένων καθώς και η κανονικοποίηση τους. Συνεχίζουμε με τη περιγραφή των αρχιτεκτονικών Multilayer Perceptron και Radial Basis Function neural networks που αναπτύχθηκαν για την απεικόνιση από τον RGB χώρο στον χώρο της PCA. Έπειτα παρουσιάζονται οι τελικές μέθοδοι για την ανακατασκευή των πολυφασματικών pixel από τα εκτιμώμενα διανύσματα της PCA. Μία μέθοδος βασισμένη σε συζευγμένα λεξικά(Dictionary Coupling) και μια στον αντίστροφο μετασχηματισμό της PCA. Τέλος επεκτείνεται η λειτουργία των μοντέλων μας από τα pixel στα patches της εικόνας με χρήση μη επικαλυπτόμενων παραθύρων και εξετάζεται η συμπεριφορά τους για διαφορετικές διαστάσεις 3x3,5x5.

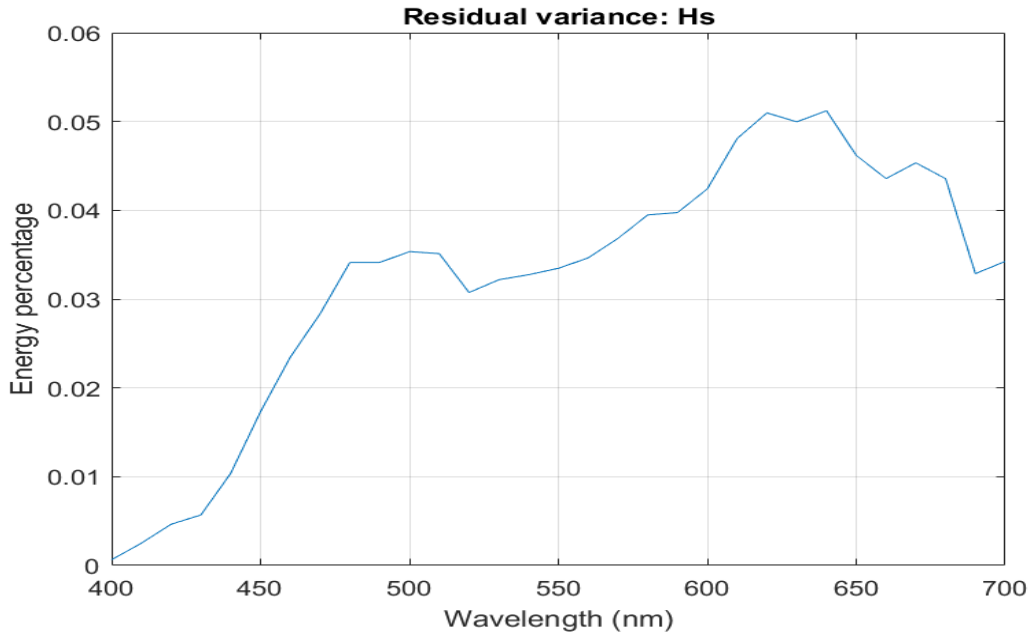
3.1 Αραιότητα των πολυφασματικών σημάτων

Το βασικότερο χαρακτηριστικό των πολυφασματικών σημάτων που επιλέξαμε να αξιοποιήσουμε είναι η αραιότητα τους. Δηλαδή, υπάρχει μια εσωτερική γεωμετρία των δεδομένων όπου αν εντοπιστεί μας δίνεται η δυνατότητα να περιγράψουμε τις πολυφασματικές υπογραφές με αρκετά μικρότερη διάσταση διατηρώντας το μέγιστο δυνατό ποσοστό πληροφορίας. Η υπόθεση αυτή επιβεβαιώνεται από την [44]. Αφού λοιπόν υπάρχει μια μικρότερη αναπαράσταση ικανή να συμπιέσει τα δεδομένα, η εύρεση της και η ευκολία αντιστροφής της καθορίζουν την τελική συμπεριφορά του συστήματος μας.

Για να μελετήσουμε την επίδραση της PCA στη μείωση της διάστασης υπολογίζουμε την εναπομένουσα διασπορά (residual variance) ανά κανάλι. Με δεδομένο πως η διασπορά εκφράζει την ενέργεια του κάθε καναλιού, ένας απλός τρόπος να εκφράσουμε τη σημασία του είναι να δούμε το ποσοστό ενέργειας που περιέχει ως προς το σύνολο της ενέργειας του σήματος και στις 31 διαστάσεις. Πιο αναλυτικά ο τύπος παίρνει την μορφή:

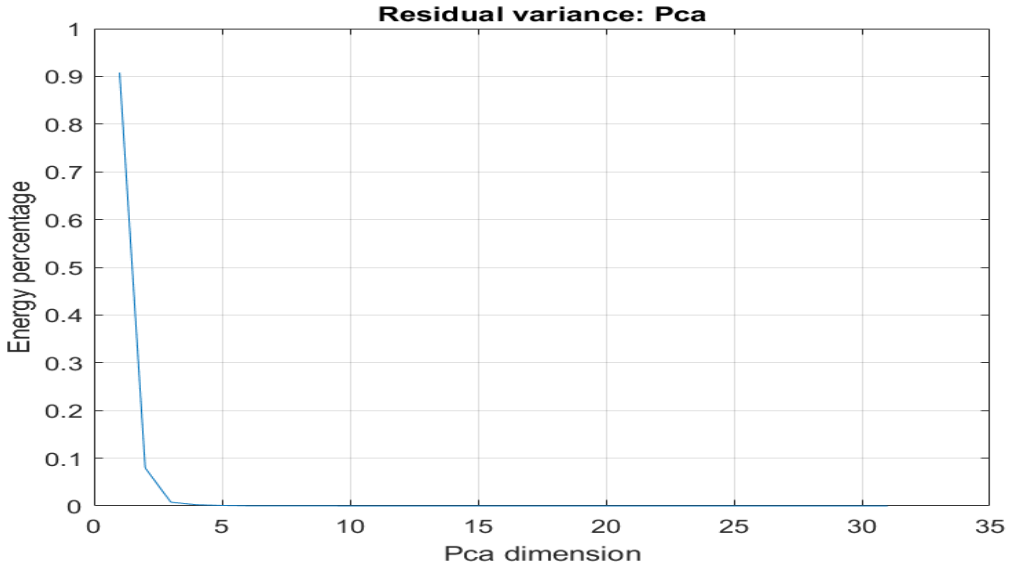
$$residual\ variance(i) = \frac{Std_i}{\sum_{j=1}^{31} Std_j}\ for\ i = 1,2,\dots,31 \tag{57}$$

Αν υπολογίσουμε την σχέση στα αρχικά διανύσματα παίρνουμε το διάγραμμα .

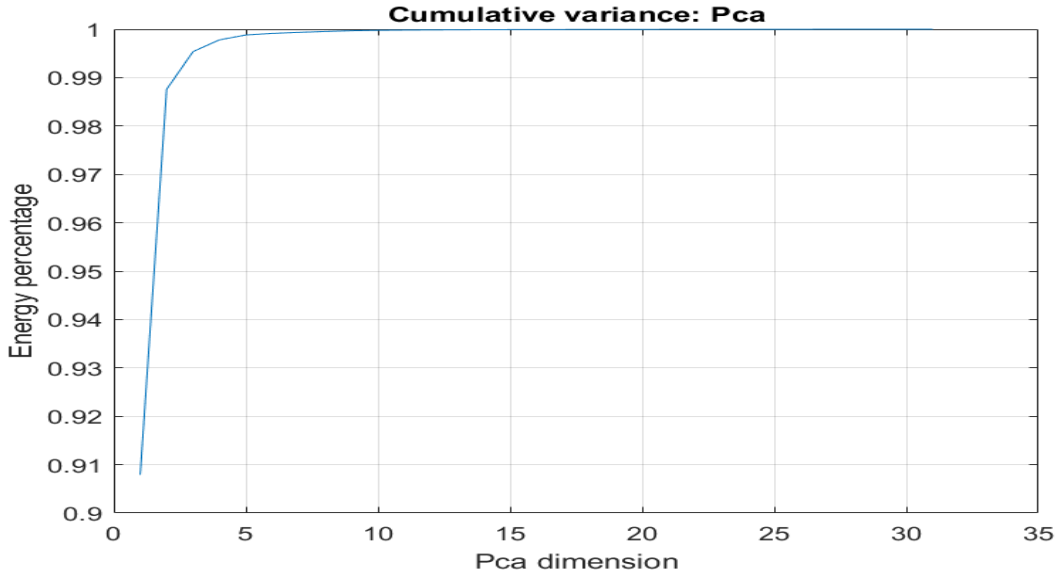


Σχήμα 3.1: Διάγραμμα εναπομένουσας διασποράς (residual variance) των αρχικών πολυφασματικών δεδομένων.

Στα αρχικά ασυμπίεστα δεδομένα παρατηρούμε πως η ενέργεια απλώνεται σε ένα μικρό εύρος τιμών, από 2% μέχρι 5%. Έχουμε δηλαδή μια σχετικά ομοιόμορφη κατανομή ενέργειας στα 31 κανάλια των φασματικών υπογραφών. Αν εφαρμόσουμε τώρα την PCA και υπολογίσουμε τόσο το αντίστοιχο διάγραμμα όσο και το αθροιστικό έχουμε.



Σχήμα 3.2: Διάγραμμα εναπομένουσας διασποράς (residual variance) μετά την εφαρμογή της PCA. Η ενέργεια συγκεντρώνεται στους πρώτους σημαντικότερους συντελεστές.



Σχήμα 3.3: Αθροιστικό διάγραμμα κατανομής ενέργειας.

Με την εφαρμογή της PCA παρατηρούμε ότι πράγματι το δεδομένα εμφανίζουν υψηλό βαθμό συμπίεσης. Με μόλις ένα στοιχείο κατορθώνουμε να περιγράψουμε το 90.7% της πληροφορίας και με τα οχτώ πρώτα το 99.95%. Εντοπίζοντας λοιπόν τις βασικότερες συνιστώσες των δεδομένων μπορούμε να απλοποιήσουμε αισθητά την περιγραφή τους, και να το εκμεταλλευτούμε αυτό ώστε να απλοποιήσουμε τις απεικονίσεις στο πρόβλημα μας.

3.2 Ορισμός προβλήματος

Όπως αναφέραμε και στο κεφάλαιο 2 σκοπός της παρούσας εργασίας είναι η ανακατασκευή πολυφασματικών εικόνων από τις αντίστοιχες RGB. Πιο συγκεκριμένα μια RGB εικόνα προκύπτει από την πολυφασματική αν εφαρμόσουμε σε αυτή την συνάρτηση ευαισθησίας (sensitivity function) μιας γνωστής κάμερας. Έστω $Hs \in R^{31}$ τα διάνυσμα των φασματικών υπογραφών, $Rgb \in R^3$ τα αντίστοιχα RGB pixel και $Camera \in R^{3x31}$ η φασματική απόκρίση της κάμερας ο μετασχηματισμός ορίζεται ως:

$$Rgb = Camera * Hs \tag{58}$$

Οι RGB εικόνες προκύπτουν λοιπόν από την προβολή των φασματικών υπογραφών στον RGB χώρο μέσω της απόκρισης της κάμερας. Σκοπός λοιπόν είναι να αντιστρέψουμε την επίδραση της κάμερας επιλύοντας ένα έντονα ασθενώς ορισμένο πρόβλημα $3 \rightarrow 31$. Πληθώρα μεθόδων έχουν αναπτυχθεί προς αυτήν την κατεύθυνση, με βασικό τους χαρακτηριστικό ότι προσπαθούν να ανακατασκευάσουν τα πολυφασματικά pixel απευθείας από τα αντίστοιχα RGB. Αν ορίσουμε ως $F(rgb): R^3 \rightarrow R^{31}$ μια συνάρτηση απεικόνισης από τον RGB χώρο στον πολυφασματικό τότε οι προσεγγίσεις αυτές παίρνουν την μορφή:

$$F(Rgb) = Hs \tag{59}$$

Τα αποτελέσματα τέτοιων μεθόδων είναι ικανοποιητικά, περιορίζονται όμως από τη μεγάλη διαφορά διάστασης μεταξύ των δύο χώρων. Η δυσκολία της απευθείας απεικόνισης οδηγεί πολλές από τις προτεινόμενες μεθόδους σε τεράστια μοντέλα τα οποία έχουν μεγάλες υπολογιστικές απαιτήσεις τόσο κατά την λειτουργία τους όσο και κατά την εκπαίδευση τους.

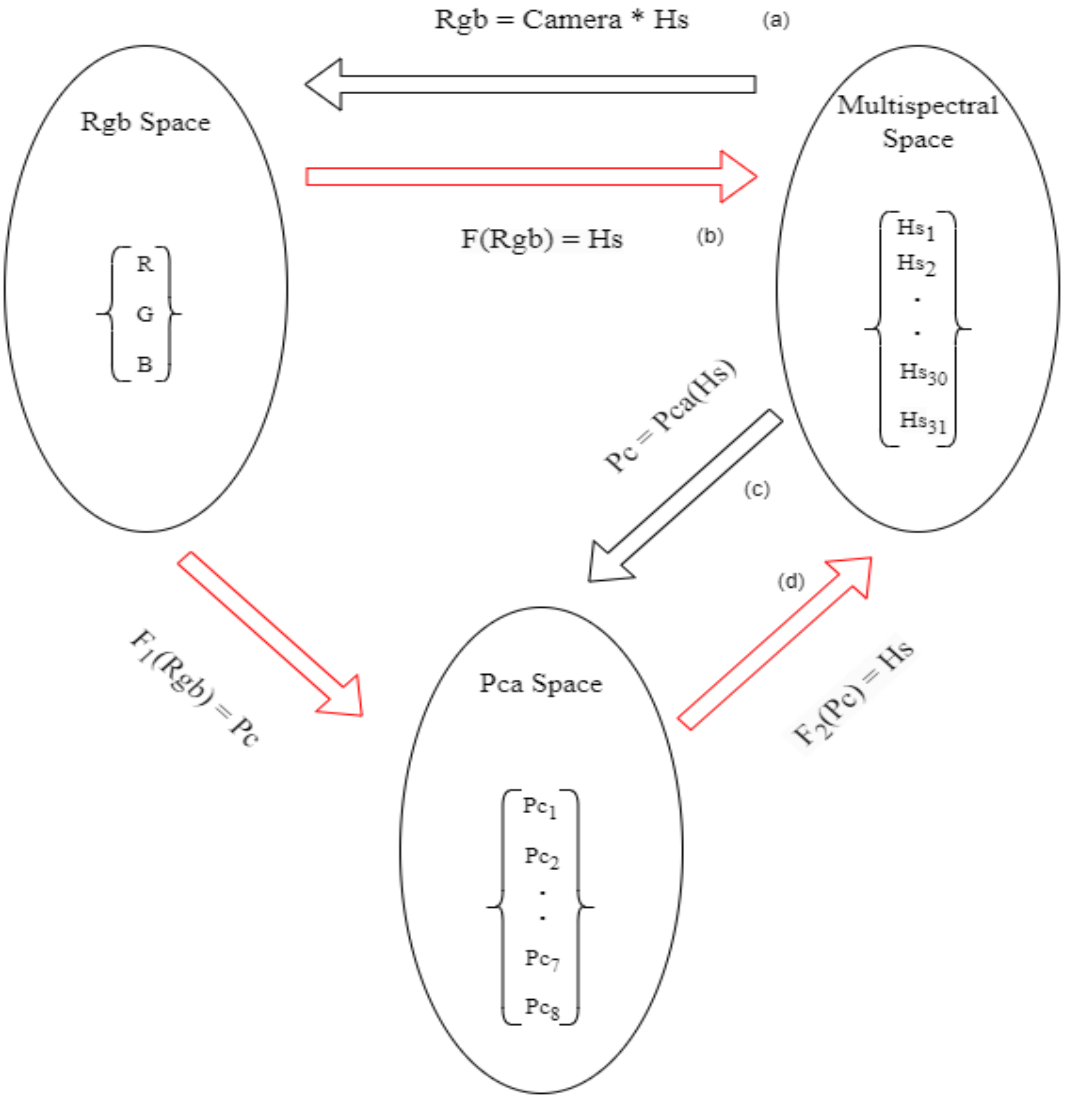
Για την δική μας προσέγγιση, αξιοποιήσαμε όπως αναφέραμε και προηγουμένως την αραιότητα των πολυφασματικών σημάτων. Με χρήση της PCA δημιουργούμε μια ενδιάμεση αναπαράσταση, που απομονώνει τις σημαντικότερες συνιστώσες των δεδομένων και μας δίνει την δυνατότητα να τεμαχίσουμε την τελική ανακατασκευή. Αν θεωρήσουμε $Pc \in R^8$ τις ενδιάμεσες αναπαραστάσεις που προκύπτουν από την μείωση διάστασης τότε το μοντέλο παίρνει την μορφή:

$$F_1(Rgb) = Pc \quad where\ F_1: R^3 \rightarrow R^8 \tag{60}$$

Και

$$F_2(Pc) = Hs \quad where\ F_2: R^8 \rightarrow R^{31} \tag{61}$$

Με αυτόν τον τρόπο δημιουργούμε δύο απλούστερα υποπροβλήματα. Αντί της απευθείας απεικόνισης $3 \rightarrow 31$ επιλύουμε αρχικά ένα πρόβλημα $3 \rightarrow 8$, απεικονίζοντας τα RGB δεδομένα στα αντίστοιχα embeddings της PCA. Και στην συνέχεια ανακατασκευάζουμε τα τελικά πολυφασματικά pixel με βάση τις εκτιμήσεις του πρώτου βήματος, χρησιμοποιώντας όμως ένα σύνολο συντελεστών που περιέχει την απαιτούμενη πληροφορία για την ικανοποιητική εκτίμηση τους. Για την προσέγγιση των συναρτήσεων F_1 και F_2 χρησιμοποιήθηκαν βασικές αρχιτεκτονικές feedforward νευρωνικών δικτύων και μεθόδων βελτιστοποίησης (dictionary coupling) που έχουν χρησιμοποιηθεί κατά κόρον στην βιβλιογραφία. Στο σχήμα 3.1 παρουσιάζονται οι σχέσεις μεταξύ των διάφορων χώρων και οι απεικονίσεις που προσεγγίζονται από τις συναρτήσεις F, F_1, F_2.

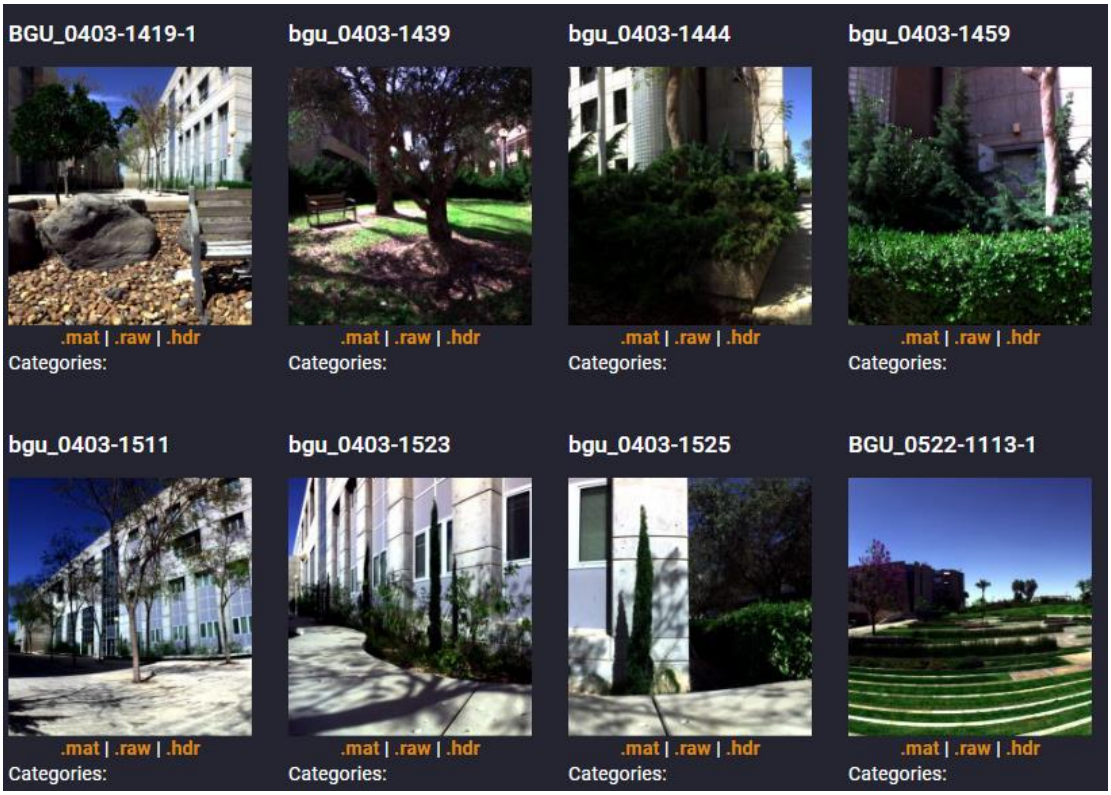


Σχήμα 3.4: Σχηματική αναπαράσταση των απεικονίσεων μεταξύ των χώρων. (a) Ο RGB χώρος προκύπτει από την προβολή των πολυφασματικών υπογραφών μέσω της απόκρισης της κάμερας. (b) Απευθείας ανακατασκευή των πολυφασματικών υπογραφών μέσω μια συνάρτησης απεικόνισης F(Rgb) = Hs. (c) Εύρεση ενδιάμεσης αναπαράστασης με χρήση της PCA. (d) Τμηματική ανακατασκευή μέσω των συναρτήσεων απεικόνισης $F_1(Rgb) = Pc, F_2(Pc) = Hs$.

3.3 Εξαγωγή δεδομένων

3.3.1 Πολυφασματικά δεδομένα

Για την εκπαίδευση και την αξιολόγηση των μοντέλων χρησιμοποιήθηκε η βάση πολυφασματικών εικόνων ICVL dataset (Interdisciplinary Computational Vision Laboratory) [9]. Αποτελείται από 200 πολυφασματικές εικόνες που τραβήχτηκαν με χρήση της υπερφασματικής κάμερας Specim PS Kappa DX4 hyperspectral camera και ενός περιστρεφόμενου τραπεζιού(rotary stage) για χωρική σάρωση. Οι εικόνες αποτελούνται από 1392x1300 pixel και 519 φασματικές μπάντες. Πραγματοποιήθηκε δηλαδή μια δειγματοληψία στο διάστημα 400-1000 nm σε βήματα των 1.25 nm. Για τον ευκολότερο χειρισμό και τη σύγκριση με άλλες μεθόδους παρέχεται και μια downsampled εκδοχή με 31 μπάντες συχνοτήτων όπου το διάστημα 400-700 nm(ορατό φως) έχει μετρηθεί σε διάστημα 10 nm. Η διάσταση λοιπόν των εικόνων είναι 1392x1300x31. Η αναπαράσταση της κάμερας είναι στα 12 bit έτσι το διάστημα τιμών του κάθε πολυφασματικού pixel είναι το [0 4095]. Τα δεδομένα διαιρούνται με την τιμή 4095 ώστε να κανονικοποιηθούν στο διάστημα [0 1].



Σχήμα 3.5: Χαρακτηριστικά παραδείγματα από το dataset πολυφασματικών εικόνων ICVL(Interdisciplinary Computer Vision Laboratory).

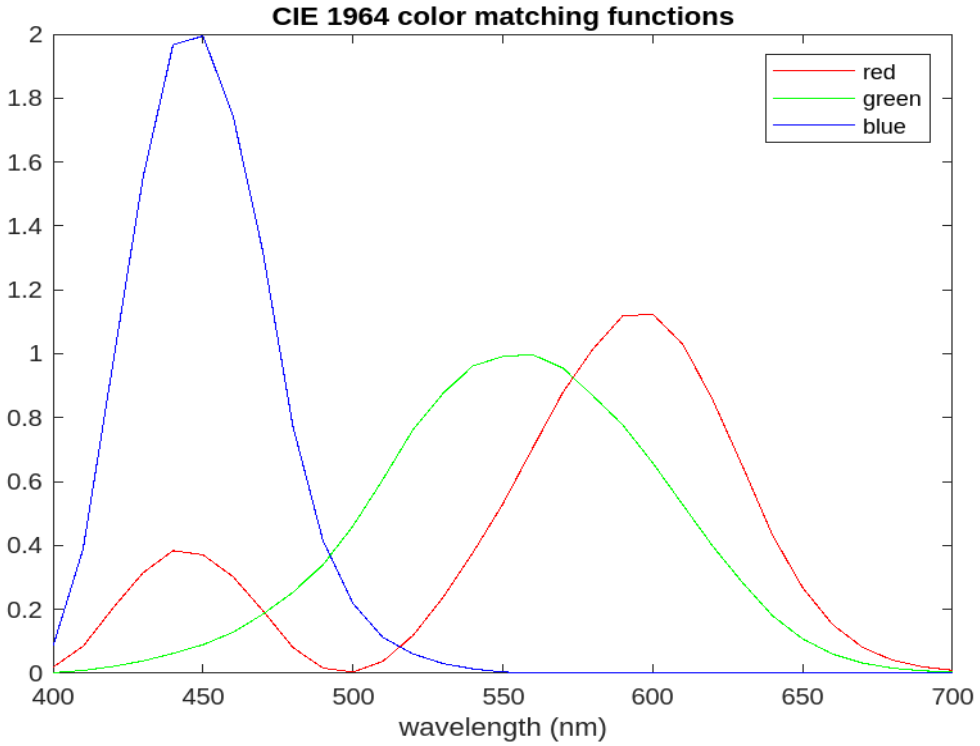
Για την δημιουργία των training και test sets χωρίζουμε τυχαία τις εικόνες σε δύο σύνολα 100 εικόνων. Από τις 100 εικόνες εκπαίδευσης αντλούμε 2000 τυχαία pixel από την

κάθε μία δημιουργώντας έτσι ένα σύνολο 200000 δεδομένων εκπαίδευσης. Σκοπός είναι να δημιουργήσουμε τις ίδιες πειραματικές συνθήκες ώστε να συγκρίνουμε με άλλες σύγχρονες μεθόδους.

3.3.2 Σύνθεση RGB δεδομένων

Για την σύνθεση των RGB δεδομένων χρησιμοποιήθηκαν οι CIE 1964 color matching functions [45] που αποτυπώνουν την χρωματική αίσθηση που αντιλαμβάνεται η όραση ενός μέσου ανθρώπου. Αποτελούν μια χρωματική αναπαράσταση ανεξάρτητη της εκάστοτε συσκευής και έτσι χρησιμοποιούνται ως σημείο αναφοράς για να ορίσουν μια πληθώρα χρωματικών χώρων. Αν λοιπόν θεωρήσουμε το σύνολο των πολυφασματικών δεδομένων μας ως ένα πίνακα όπου σε κάθε στήλη έχουμε και μια πολυφασματική υπογραφή $Hs: R^{31x200000}$ και την αντίστοιχη απόκριση ευαισθησίας ως $CIE: R^{3x31}$ τότε τα RGB $Rgb: R^{3x200000}$ δεδομένα μας προκύπτουν από τον πολλαπλασιασμό:

$$Rgb = CIE * Hs \tag{62}$$



Σχήμα 3.6: Φασματική απόκριση του πρότυπου CIE 1964 color matching functions. Κάθε συνιστώσα απεικονίζεται με το αντίστοιχο χρώμα Red,Green,Blue.

Στην συνέχεια κανονικοποιούμε τα RGB διανύσματα ώστε να έχουν μηδενική μέση τιμή και διασπορά ένα. Για τον σκοπό αυτό υπολογίζουμε την μέση τιμή και την διασπορά για κάθε μια από τις Red, Green και Blue συνιστώσες των δεδομένων εκπαίδευσης. Οι τιμές αυτές εφαρμόζονται τόσο στα δεδομένα εκπαίδευσης όσο και στα τεστ δεδομένα. Αν θεωρήσουμε rgb_i το κάθε ένα από τα 3 κανάλια $mrgb_i$ την μέση τιμή του καναλιού και $stdrgb_i$ την διασπορά του τότε η κανονικοποίηση παίρνει την μορφή

$$\frac{rgb_i - mrgb_i}{stdrgb_i} \quad where\ i = 1,2,3 \tag{63}$$

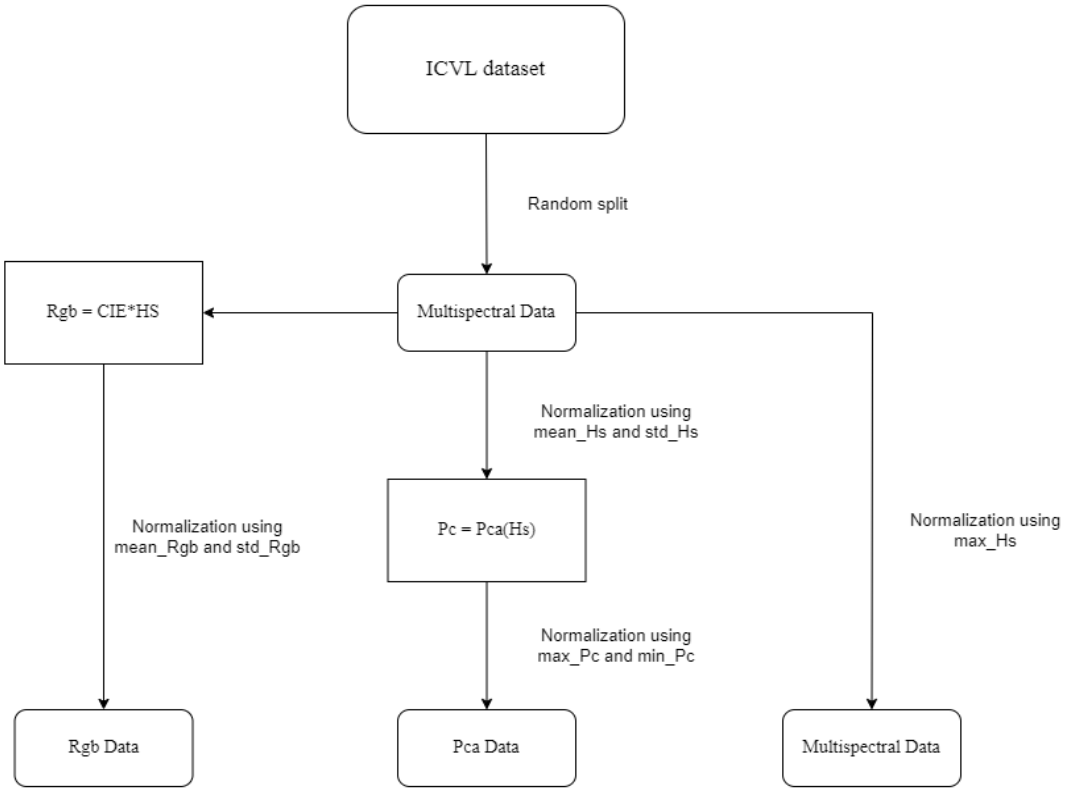
3.3.3 Μείωση διάστασης πολυφασματικών δεδομένων

Για τον υπολογισμό της PCA αρχικά κανονικοποιούμε τα δεδομένα ώστε κάθε κανάλι να έχει μηδενική μέση τιμή και διασπορά ένα. Αν θεωρήσουμε hs_i το κάθε ένα από τα 31 κανάλια, mhs_i την μέση τιμή του καναλιού και $stdhs_i$ την διασπορά του τότε η κανονικοποίηση παίρνει την μορφή

$$\frac{hs_i - mhs_i}{stdhs_i} \quad where\ i = 1,2,\dots,31 \tag{64}$$

Οι τιμές mhs_i και $stdhs_i$ υπολογίζονται από τα 200000 δεδομένα εκπαίδευσης και εφαρμόζονται τόσο σε αυτά όσο και στα τεστ δεδομένα. Αντίστοιχα ο μετασχηματισμός των τεστ δεδομένων γίνεται με τον πίνακα W που υπολογίστηκε από την εφαρμογή της PCA στα δεδομένα εκπαίδευσης. Τέλος υπολογίζουμε την μέγιστη και ελάχιστη τιμή των συμπιεσμένων διανυσμάτων εκπαίδευσης και μετασχηματίζουμε το σύνολο των embeddings στο διάστημα [0 1] μέσω της σχέσης

$$\frac{Pc - minPc}{maxPc - minPc} \tag{65}$$



Σχήμα 3.7: Μπλόκ διάγραμμα εξαγωγής δεδομένων. Απεικονίζονται αναλυτικά οι μετασχηματισμοί επί των πολυφασματικών δεδομένων καθώς και οι κανονικοποιήσεις πριν από κάθε βήμα. Καταλήγουμε με ένα σύνολο δεδομένων για κάθε χώρο αναπαράστασης Rgb,Pca,Multispectral.

3.4 Απεικόνιση από τον RGB χώρο στον χώρο της PCA

Για το πρώτο βήμα της εκτίμησης, δηλαδή για την συνάρτηση απεικόνισης $F_1: R^3 \rightarrow R^8$ των RGB δεδομένων στα αντίστοιχα embeddings της PCA χρησιμοποιήθηκαν βασικές αρχιτεκτονικές feedforward νευρωνικών δικτύων και συγκρίθηκαν τα αποτελέσματα τους. Η βασική θεωρία τόσο για τον Multilayer Perceptron όσο και για τα Radial Basis Function neural networks αναπτύχθηκε στο κεφάλαιο 2. Βασικό λόγος που επιλέξαμε τις συγκεκριμένες αρχιτεκτονικές είναι η απλότητα στην προσέγγιση τους που οδηγεί σε μικρά μοντέλα, ιδανικά για εφαρμογές με περιορισμό των υπολογιστικών πόρων. Και τα δύο μοντέλα μας προσπαθούν να ελαχιστοποιήσουν το μέσο τετραγωνικό σφάλμα (Mean Square Error) της εκτίμησης. Για την αξιολόγηση τους χρησιμοποιήθηκαν οι μετρικές της μέσης απόλυτης τιμής της σχετικής διαφοράς (Mean Absolute Relative Error), η ρίζα του μέσου τετραγωνικού σφάλματος (Root Mean Square Error) και η ρίζα του μέσου τετραγωνικού σχετικού σφάλματος (relative Root Mean Square Error). Στην συνέχεια περιγράφονται η αρχιτεκτονική, οι υπερπαράμετροι και οι αλγόριθμοι εκπαίδευσης που χρησιμοποιήθηκαν για τις καλύτερες υλοποιήσεις της κάθε κατηγορίας.

3.4.1 Radial Basis Function neural network

Τα Radial Basis Function neural networks είναι μια απλή feedforward αρχιτεκτονική που αποτελείται από 3 επίπεδα. Το επίπεδο εισόδου περιέχει τα διανύσματα που δίνονται ως είσοδο στο δίκτυο μας, στην περίπτωση μας έχουμε τα 3 κανάλια Red Green και Blue του κάθε pixel. Για το διάστημα τιμών των εν λόγω δειγμάτων δοκιμάστηκαν διαφορετικά σχήματα κανονικοποίησης. Εδώ περιγράφεται το επικρατέστερο όπου στο κάθε κανάλι αφαιρούμε την μέση τιμή και διαιρούμε με την διασπορά. Ο χώρος εισόδου λοιπόν είναι ο 3D χώρος με τα διανύσματα να είναι κεντραρισμένα γύρω από την τιμή 0.

Στην συνέχεια έχουμε το κρυφό επίπεδο που πραγματοποιεί την μη γραμμική ενεργοποίηση του δικτύου μας. Αποτελείται από γκαουσιανούς πυρήνες όπου ο κάθε ένας υπολογίζει την απόσταση του διανύσματος εισόδου από ένα διάνυσμα κέντρο c_i. Ουσιαστικά έχουμε έναν τεμαχισμό του χώρου εισόδου όπου διαφορετικά σημεία ενεργοποιούν διαφορετικά κέντρα με την διασπορά s_i αντίστοιχα να καθορίζει την ευαισθησία του πυρήνα, δηλαδή πόσο μακριά από το κέντρο θέλουμε να αντιδρά. Με δεδομένο πως δουλεύουμε στον 3D χώρο μπορούμε να φανταστούμε του πυρήνες ως σφαίρες όπου το κέντρο και η ακτίνα τους είναι οι παράγοντες προς εκπαίδευση. Με αυτόν τον τρόπο γεωμετρικά κοντινές συγκεντρώσεις δεδομένων θα επηρεάσουν τους ίδιους πυρήνες ενώ διαφορετικές συγκεντρώσεις δεδομένων θα αποτυπωθούν σε εντελώς διαφορετικά διανύσματα ενεργοποίησης. Το σύνολο των κρυφών νευρώνων είναι μια παράμετρος που προσδιορίζεται με μια trial and error λογική. Εδώ θα αναφέρουμε την καλύτερη υλοποίηση που βασίστηκε σε μια αρχιτεκτονική 20 νευρώνων.

Αφού υπολογιστούν οι μη γραμμικές ενεργοποιήσεις καταλήγουμε σε ένα διάνυσμα 20 τιμών. Το τελικό επίπεδο εξόδου πραγματοποιεί έναν γραμμικό μετασχηματισμό του διανύσματος για να παραχθεί το τελικό αποτέλεσμα. Η διάσταση τόσο του γραμμικού πίνακα όσο και του bias καθορίζονται από την διάσταση εξόδου του συστήματος. Τα αποτελέσματα για διαφορετικές διαστάσεις εξόδου δίνονται στο επόμενο κεφάλαιο, εδώ θα αναφέρουμε την υλοποίηση με το καλύτερο αποτέλεσμα που είναι διάστασης 8. Τόσο τα 20 κέντρα c_i με τις αντίστοιχες διασπορές s_i όσο και οι πίνακες μετασχηματισμού W, b υπολογίζονται κατά την εκπαίδευση. Οι σχέσεις του δικτύου παίρνουν την μορφή:

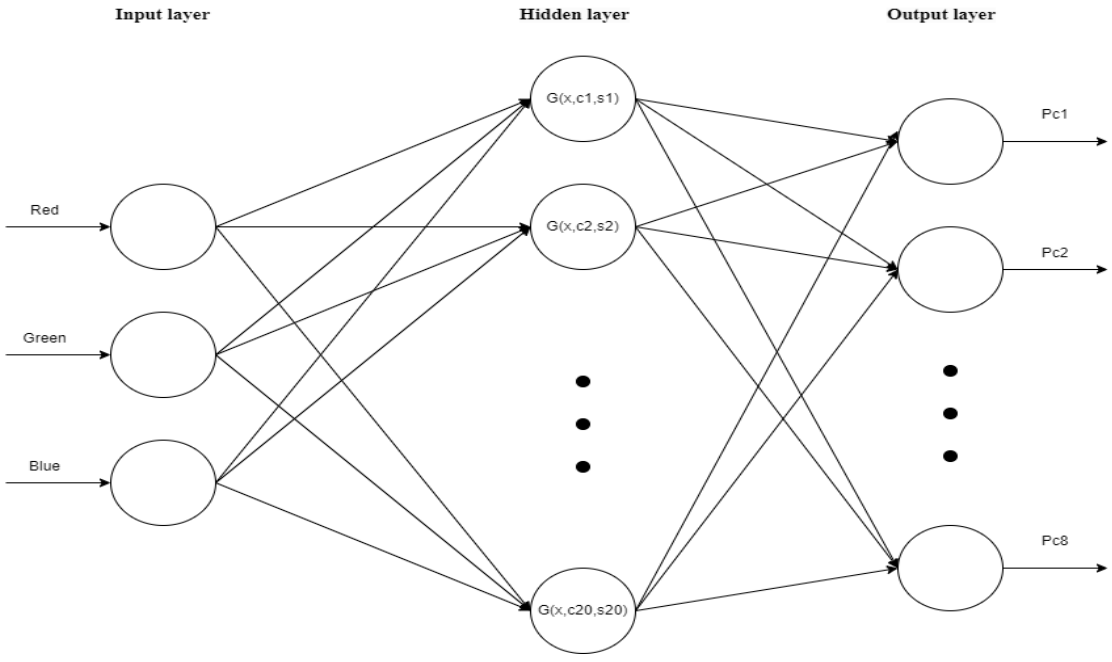
$$G = e^{\frac{\|Rgb - c_i\|^2}{2s_i^2}} \quad i = 1,..,20 \tag{66}$$

Αρχικά υπολογίζουμε την έξοδο των 20 γκαουσιανών πυρήνων.

$$Pc = WG + b \tag{67}$$

Και στην συνέχεια προβάλουμε το διάνυσμα της μη γραμμικής ενεργοποίησης G στην έξοδο.

Για την εκπαίδευση του δικτύου χρησιμοποιήθηκε ο αλγόριθμος Levemberg-Marquardt που περιγράφεται στο κεφάλαιο 2. Επειδή ο εν λόγω αλγόριθμος δεν αποτελεί την πιο συχνή επιλογή για την εκπαίδευση δικτύων δεν ήταν δυνατό να χρησιμοποιηθεί κάποιο από τα γνωστά πακέτα των βιβλιοθηκών της Pytorch και Tensorflow έτσι ο αλγόριθμος αναπτύχθηκε στην Matlab εξ ολοκλήρου από την αρχή. Λόγω της απλής αρχιτεκτονικής του δικτύου δεν υπάρχει ανάγκη για διάδοση του σφάλματος σε προηγούμενα επίπεδα γεγονός που απλοποιεί σημαντικά τους υπολογισμούς. Η κλίση της συνάρτησης σφάλματος ως προς τους συντελεστές c_i, s_i, W, b υπολογίζεται απευθείας από την στιγμιαία εκτίμηση του σφάλματος. Η ακριβής μορφή και θέση των παραγώγων στο μητρώο της ιακωβιανής αφήνονται ως σχόλια στον κώδικα. Ως προς τις υπερπαραμέτρους του δικτύου, ο συντελεστής εκπαίδευσης αρχικοποιήθηκε στο 0.0001. Ο αλγόριθμος όμως τροποποιεί τον συντελεστή συνεπώς δεν επηρεάζεται η σύγκλιση. Τέλος τα διάφορα δίκτυα εκπαιδεύθηκαν για 100 εποχές όπου σε κάθε μία χρησιμοποιούνται όλα τα δεδομένα για τον υπολογισμό των αλλαγών του δικτύου, με αποτέλεσμα να μην υπάρχει ανάγκη για χρήση batch.



Σχήμα 3.8: Σχηματική δομή του Radial Basis Function neural network. Στην είσοδο δίνεται ένα RGB pixel και εκτιμάται το περιεχόμενο των 8 κυρίαρχων συνιστωσών.

3.4.2 Multilayer Perceptron

Ο Multilayer Perceptron είναι η πιο διαδεδομένη feedforward αρχιτεκτονική νευρωνικών δικτύων. Αποτελείται από ένα σύνολο κρυφών επιπέδων όπου το κάθε ένα πραγματοποιεί έναν γραμμικό μετασχηματισμό και μία μη γραμμική ενεργοποίηση στην έξοδο του προηγούμενου επιπέδου, με σκοπό να προσεγγίσει μια μη γραμμική σχέση μεταξύ των διανυσμάτων εισόδου και εξόδου. Κατά τον σχεδιασμό του μοντέλου δοκιμάστηκαν πολλαπλές αρχιτεκτονικές και διαφορετικές επιλογές υπερπαραμέτρων. Εδώ θα αναφερθούμε στην καλύτερη η οποία προέκυψε μέσω μιας trial and error λογικής.

Αρχικά έχουμε το επίπεδο εισόδου στο οποίο δίνονται τα κανονικοποιημένα RGB διανύματα. Η κανονικοποίηση που επιλέχθηκε ήταν και πάλι η αφαίρεση της μέσης τιμής και της διασποράς του κάθε καναλιού. Στην συνέχεια έχουμε μια σειρά από 3 κρυφά επίπεδα με διαστάσεις 64, 32, 16. Κάθε ένα από τα κρυφά επίπεδα εμπεριέχει μια μη γραμμική συνάρτηση ενεργοποίησης. Δοκιμάστηκαν κατά σειρά οι συναρτήσεις $ReLU, tanh$ και η $sigmoid$ με τα καλύτερα αποτελέσματα να προκύπτουν από την $ReLU$. Το επίπεδο εξόδου αποτελείται από έναν γραμμικό μετασχηματισμό χωρίς κάποια συνάρτηση ενεργοποίησης με σκοπό να προσεγγίσει τις πραγματικές τιμές του διανύσματος στόχου και η διάσταση του είναι 8. Το δίκτυο μας φαίνεται στο σχήμα 3.9. Για την εκπαίδευση χρησιμοποιήθηκαν οι αλγόριθμοι Stochastic Gradient Descent with momentum και Adam optimizer. Ο συντελεστής εκπαίδευσης αρχικοποιήθηκε στην τιμή μ = 0.0001 και εξετάστηκε η επίδραση ενός learning rate scheduler που τροποποιούσε την τιμή του συντελεστή κατά την διάρκεια της. Τα καλύτερα αποτελέσματα προέκυψαν από τον Adam optimizer χωρίς κάποια εμφανή διαφορά από την χρήση του scheduler. Το batch size εξετάστηκε για διαφορετικές τιμές batch_size=10,25,50,100. Τα καλύτερα αποτελέσματα προέκυψαν για την τιμή 25. Η εκπαίδευση πραγματοποιήθηκε για 100 εποχές. Για τους συντελεστές β_1, β_2 εξετάστηκαν διαφορετικές τιμές. Η τιμή του $\beta_1 = 0.8$ οδήγησε σε ομαλότερη σύγκλιση ενώ η τιμή του β_2 παρέμεινε στην αρχική τιμή 0.999 καθώς προκαλούσε έντονη αστάθεια κατά την εκπαίδευση.

Αν θεωρήσουμε W_i, b_i τους πίνακες των μετασχηματισμών κάθε επιπέδου τότε η σχέση που περιγράφει το δίκτυο είναι:

$$z_1 = ReLU(W_1 * Rgb + b_1) \tag{68}$$

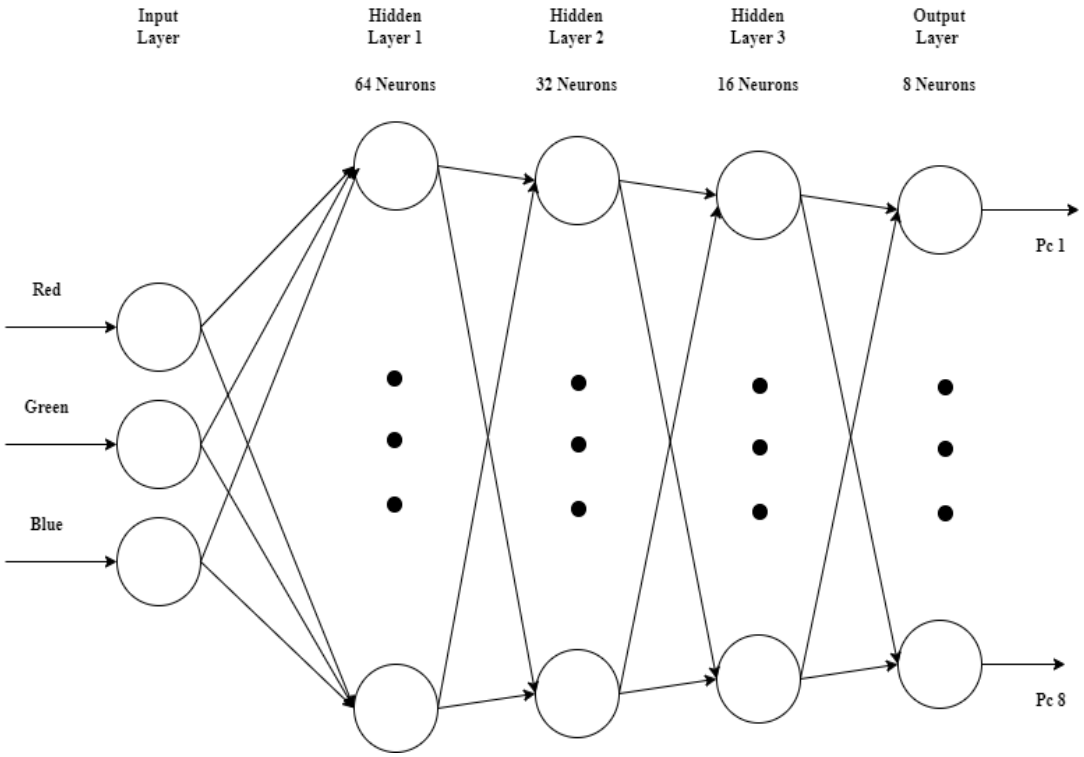
Στο πρώτο επίπεδο εφαρμόζουμε τον μετασχηματισμό στα RGB δεδομένα και χρησιμοποιούμε την $ReLU$.

$$z_2 = ReLU(W_2 * z_1 + b_2) \tag{69}$$
$$z_3 = ReLU(W_3 * z_2 + b_3) \tag{70}$$

Στα δύο επόμενα κρυφά επίπεδα εφαρμόζουμε τους αντίστοιχους μετασχηματισμούς και χρησιμοποιούμε την $ReLU$ στις εξόδους του προηγούμενου επιπέδου.

$$Pc = W_4 * z_3 + b_4 \tag{71}$$

Στο επίπεδο εξόδου εφαρμόζουμε τον τελευταίο μετασχηματισμό χωρίς μη γραμμική ενεργοποίηση για να εκτιμήσουμε το περιεχόμενο των κυρίαρχων συνιστωσών.



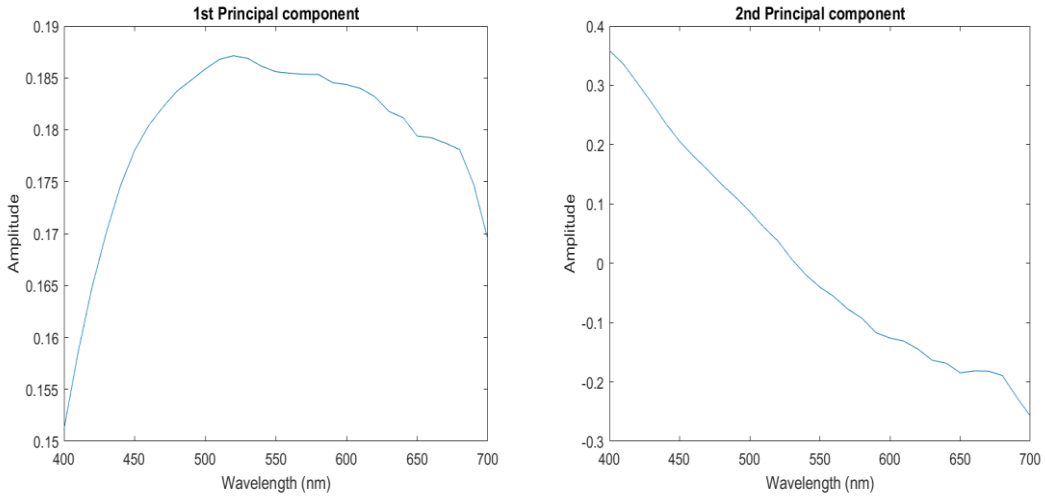
Σχήμα 3.9: Σχηματική δομή του Multilayer Perceptron. Στην είσοδο δίνεται ένα Rgb pixel και εκτιμάται το περιεχόμενο των 8 κυρίαρχων συνιστωσών.

3.5 Απεικόνιση από τον χώρο της PCA στον πολυφασματικό χώρο

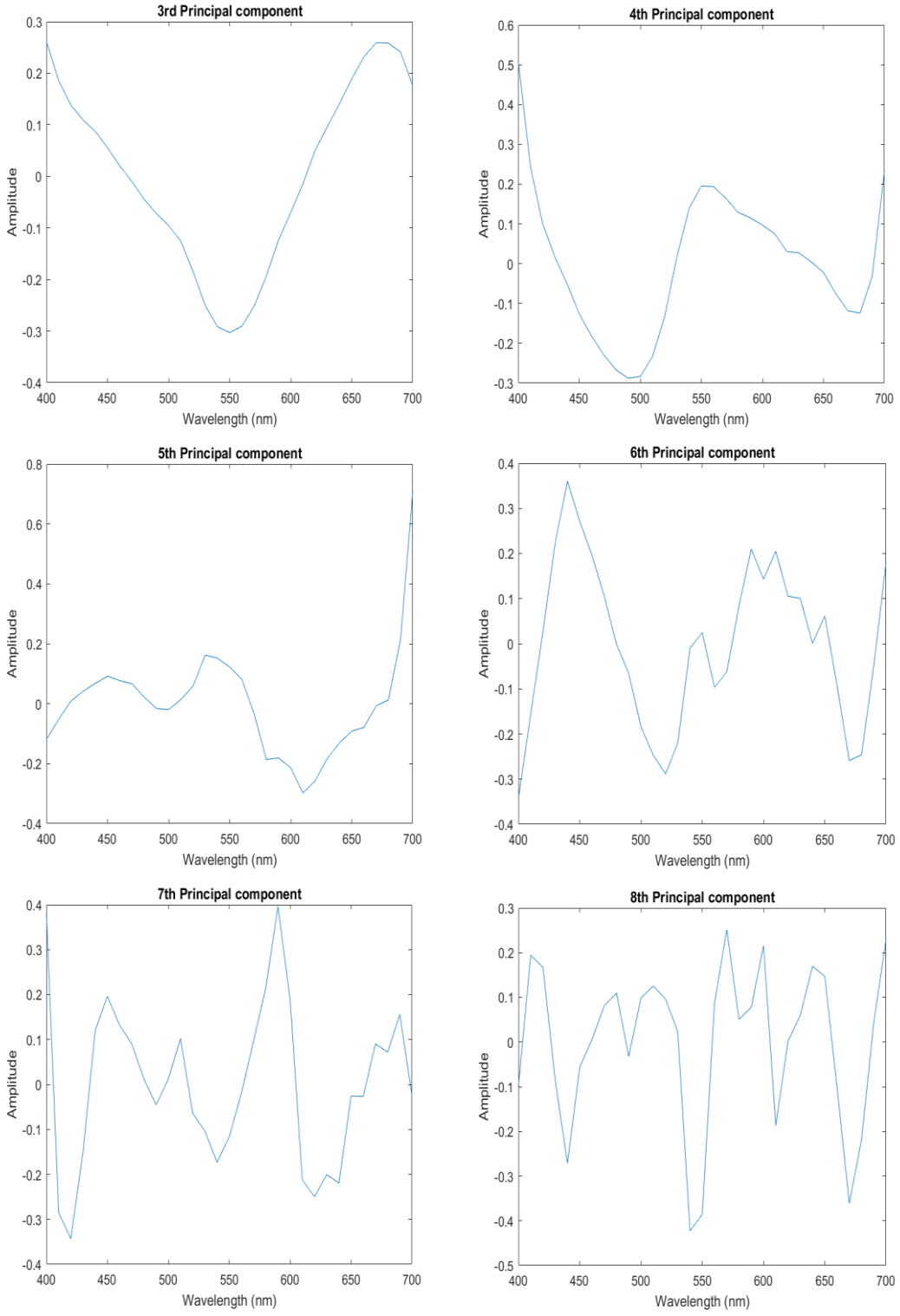
3.5.1 Αντίστροφος μετασχηματισμός της PCA

Για την συνάρτηση απεικόνισης $F_2(Pc) = Hs, F_2(Pc): R^8 \rightarrow R^{31}$ από τον χώρο της PCA στον χώρο των πολυφασματικών δεδομένων αναπτύχθηκαν δύο μέθοδοι που θα περιγράψουμε στην συνέχεια. Ένα από τα βασικά πλεονεκτήματα της PCA, σε σχέση ιδιαίτερα με μη γραμμικές μεθόδους μείωσης διάστασης, είναι πως μπορεί να αντιστραφεί σχετικά εύκολα. Η PCA ουσιαστικά αποσυνθέτει κάθε φασματική υπογραφή στις 8 κυρίαρχες συνιστώσες της που μάλιστα είναι και κάθετες μεταξύ τους. Συνεπώς μπορεί να αντιστραφεί άμεσα με χρήση του ανάστροφου μετασχηματισμού W^T. Τα 8-διάστατα διανύσματα που προκύπτουν, περιγράφουν την ενέργεια κάθε κυρίαρχης συνιστώσας στην σύνθεση του αντίστοιχου πολυφασματικού pixel. Εφόσον έχουμε πετύχει ικανοποιητική ακρίβεια στην εκτίμηση των embeddings από τα αντίστοιχα RGB pixel, με την χρήση των νευρωνικών δικτύων που αναπτύχθηκαν, μπορούμε να αντιστρέψουμε απευθείας υπολογίζοντας τον γραμμικό συνδυασμό των κυρίαρχων συνιστωσών με βάση τις εκτιμήσεις μας.

Από το σχήμα 3.10 παρατηρούμε ότι οι πρώτες και βασικότερες συνιστώσες είναι χαμηλόσυχνες. Λίγα και απλά στοιχεία αποτυπώνουν το μεγαλύτερο κομμάτι της πληροφορίας. Καθώς όμως προχωράμε προκύπτουν περιπλοκότερες συνιστώσες που ναι μεν έχουν μικρή συμμετοχή αλλά καθορίζουν τις λεπτομέρειες του τελικού σχήματος. Το φαινόμενο αυτό παρατηρείται σε πολλούς μετασχηματισμούς που αποσυνθέτουν φυσικά σήματα σε ένα σύνολο βασικών συνιστωσών, είτε είναι μετασχηματισμοί σε περιοδικές κυρίαρχες ή κυρίαρχες συνιστώσες(Fourier,Pca), και προκύπτει από την αραιότητα τους.



Σχήμα 3.10(α): Οι 8 κυρίαρχες συνιστώσες των δεδομένων.



.

Σχήμα 3.10(β): Οι 8 κυρίαρχες συνιστώσες των δεδομένων.

3.5.2 Πρώτη προτεινόμενη μέθοδος επίλυσης

Καταλήγουμε έτσι στην πρώτη προτεινόμενη μέθοδο επίλυσης. Για κάθε RGB pixel χρησιμοποιούμε ένα από τα νευρωνικά δίκτυα για να εκτιμήσουμε το περιεχόμενο των 8 κυρίαρχων συνιστωσών. Στην συνέχεια αντιστρέφουμε την κανονικοποίηση χρησιμοποιώντας τις min,max τιμές των emdeddings της PCA. Με χρήση του πίνακα W^T υπολογίζουμε τον γραμμικό συνδυασμό των 8 κυρίαρχων συνιστωσών. Και τέλος αντιστρέφουμε την τελική κανονικοποίηση χρησιμοποιώντας τις τιμές της μέσης τιμής και της διασποράς των 31 πολυφασματικών καναλιών.

Αρχικά εκτιμάμε το περιεχόμενο της PCA με τις σχέσεις:

$$F_{net}(Rgb) = Pc' \tag{72}$$

και

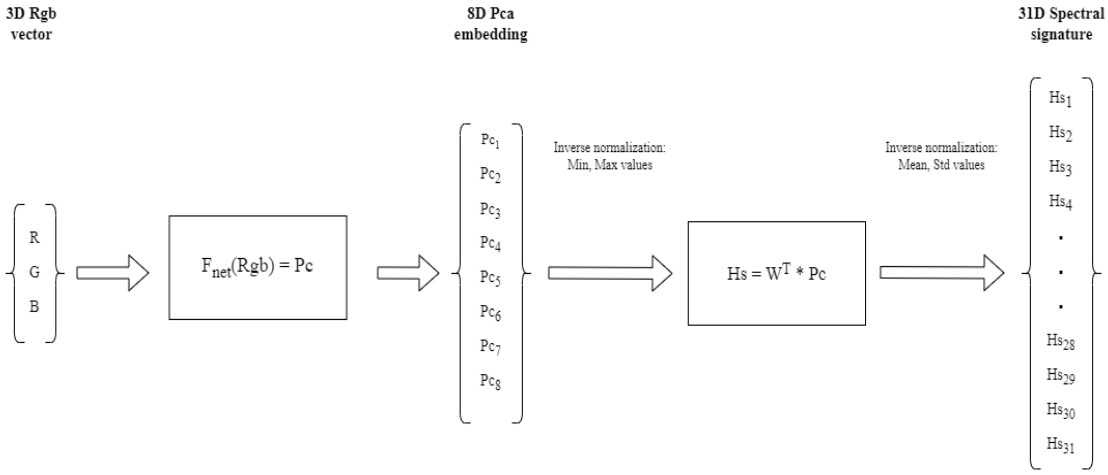
$$Pc = \ Pc'(maxPc - minPc) + \min Pc \tag{73}$$

Και τελικά ανακατασκευάζουμε το πολυφασματικό pixel με τις σχέσεις:

$$hs' = W^T * Pc \tag{74}$$

Και στην συνέχεια

$$Hs_i = hs_i' * Std_i + m_i \ \ where\ i = 1, \dots, 31 \tag{75}$$



Σχήμα 3.11: Μπλόκ διάγραμμα πρώτης προτεινόμενης μεθόδου επίλυσης.

3.5.3 Συζευγμένα λεξικά (Dictionary Coupling)

Στη συνέχεια εξετάσαμε την συμπεριφορά της μεθόδου στην περίπτωση που ο αντίστροφος μετασχηματισμός δεν είναι γνωστός αλλά προσεγγίζεται μέσα από ένα ζεύγος συζευγμένων λεξικών (dictionary coupling). Η μέθοδος αυτή περιγράφει τις βασικές συσχετίσεις δύο διαφορετικών χώρων μέσα από το πρίσμα δύο προσαρμοσμένων λεξικών. Μας δίνεται έτσι η δυνατότητα να εντοπίσουμε και να ταιριάξουμε τα βασικότερα στοιχεία για την αναπαράσταση του κάθε χώρου και να μεταβούμε έτσι από το έναν στον άλλο με ικανοποιητική ακρίβεια. Σκοπός μας είναι να εξετάσουμε αν μπορούμε να αντιστρέψουμε την διαδικασία της μείωσης διάστασης ικανοποιητικά, ώστε να επεκτείνουμε τα μοντέλα μας σε μη γραμμικές μεθόδους όπου δεν υπάρχει συνήθως αντίστροφος μετασχηματισμός.

Πιο αναλυτικά, αν θεωρήσουμε $Pc \in R^{8x200000}$ το σύνολο των embeddings της PCA για την εκπαίδευση, $Hs \in R^{31x200000}$ το σύνολο των αντίστοιχων πολυφασματικών υπογραφών, $D_h \in R^{31xdictDim}$, $D_l \in R^{8xdictDim}$ τα λεξικά υψηλής και χαμηλής ανάλυσης και $G \in R^{dictDimx200000}$ τις αραιές περιγραφές των διανυσμάτων με βάση το αντίστοιχο λεξικό τότε το πρόβλημα ανάγεται στην ελαχιστοποίηση της σχέσης:

$$\min_{D_h, D_l, G} \{\|Pc - D_l G\|_F^2 + \|Hs - D_h G\|_F^2\} \; subject\ to \;\; \forall i, \|g_i\|_0 \leq T_0 \qquad (76)$$

Κατορθώνουμε έτσι να δημιουργήσουμε δύο λεξικά τα οποία περιγράφουν το σύνολο των δεδομένων εκπαίδευσης με όσο το δυνατόν μικρότερο μέσο τετραγωνικό σφάλμα και χρησιμοποιώντας κοινές αραιές περιγραφές G. Η απαίτηση για κοινές αραιές περιγραφές οδηγεί τα λεξικά στο να ταιριάξουν τα αντίστοιχα άτομα τους, να εντοπίσουν και να συσχετίσουν έτσι τις σημαντικότερες λεπτομέρειες των δύο χώρων, δίνοντας μας τη δυνατότητα να μεταφερθούμε από τον ένα στον άλλο.

Η ελαχιστοποίηση της σχέσης (76) πραγματοποιείται ως προς τους συντελεστές D_h, D_l, G γεγονός που καθιστά αδύνατη την απευθείας επίλυση του προβλήματος. Αρχικά επικεντρωνόμαστε στο ένα σκέλος του προβλήματος δηλαδή στη σχέση $\|Hs - D_h G\|_F^2$. Η προσαρμογή του λεξικού D_h και ο υπολογισμός των αραιών περιγραφών G γίνεται με τη χρήση του αλγορίθμου K-SVD που περιγράψαμε αναλυτικά στο κεφάλαιο 2. Μόλις ολοκληρωθεί η εκπαίδευση καταλήγουμε σε ένα λεξικό υψηλής ανάλυσης που περιγράφει ικανοποιητικά τα δεδομένα Hs μέσα από τις T_0-αραιές περιγραφές που υπολογίζονται από τον αλγόριθμο OMP. Η διάσταση του λεξικού(dictDim) και ο παράγοντας αραιότητας T_0 αποτελούν τις παραμέτρους του μοντέλου που καθορίζονται κατά την εκπαίδευση μέσω μιας trial and error λογικής.

Αφού λοιπόν έχουμε υπολογίσει τα D_h και G μεταφερόμαστε στην σχέση $\|Pc - D_l G\|_F^2$ όπου τόσο το G όσο και τα δεδομένα μας Pc είναι γνωστά. Η ελαχιστοποίηση της σχέσης προκύπτει άμεσα με τον υπολογισμό του ψευδοαντίστροφου δηλαδή:

$$D_l = (G^T G)^{-1} G^T * Pc \qquad (77)$$

Με αυτόν τον τρόπο έχουμε εκπαιδεύσει ένα λεξικό υψηλής ανάλυσης D_h να περιγράφει τα δεδομένα μέσα από τις αραιές αναπαραστάσεις G και στην συνέχεια επιλύσαμε ως προς το λεξικό χαμηλής ανάλυσης χρησιμοποιώντας αυτές τις περιγραφές.

Επειδή ο αλγόριθμος OMP που θα εφαρμοστεί στο λεξικό D_l απαιτεί οι στήλες του να είναι μοναδιαία διανύσματα διαιρούμε το κάθε άτομο του λεξικού με την νόρμα του. Αυτό έχεις ως αποτέλεσμα οι αραιές περιγραφές να αλλάξουν. Διαιρούμε λοιπόν και τις αντίστοιχες στήλες του D_h για να διατηρήσουμε την αναλογία των συντελεστών G. Αν θεωρήσουμε D_{l_i} το κάθε άτομο του λεξικού χαμηλής ανάλυσης και D_{h_i}τα αντίστοιχα του λεξικού υψηλής ανάλυσης η κανονικοποίηση παίρνει την μορφή:

$$\frac{D_{l_i}}{norm(D_{l_i})}, \frac{D_{h_i}}{norm(D_{l_i})} \quad for\ i = 1, \dots, dictSize \tag{78}$$

Για τον παράγοντα αραιότητας και το μέγεθος του λεξικού δοκιμάστηκαν διαφορετικοί συνδυασμοί τιμών που παρατίθενται σε διάγραμμα του επόμενου κεφαλαίου. Η καλύτερη λύση δόθηκε για μέγεθος λεξικού $dictSize = 2048$ και $T_0 = 12$.

3.5.4 Δεύτερη προτεινόμενη μέθοδος επίλυσης

Καταλήγουμε έτσι στην δεύτερη προτεινόμενη μέθοδο επίλυσης. Στο πρώτο βήμα και πάλι εκτιμάμε το περιεχόμενο των 8 κυρίαρχων συνιστωσών με χρήση ενός από τα νευρωνικά δίκτυα που σχεδιάσαμε. Στην συνέχεια όμως αντί να αντιστρέψουμε απευθείας χρησιμοποιούμε το ζεύγος των συζευγμένων λεξικών. Με χρήση του λεξικού χαμηλής ανάλυσης D_l και του αλγορίθμου OMP εντοπίζουμε την αραιή περιγραφή G, δηλαδή τις στήλες του D_l που περιγράφουν το διάνυσμα της PCA, και στην συνέχεια συνθέτουμε το υψηλής ανάλυσης σήμα χρησιμοποιώντας τις αντίστοιχες στήλες του λεξικού D_h :

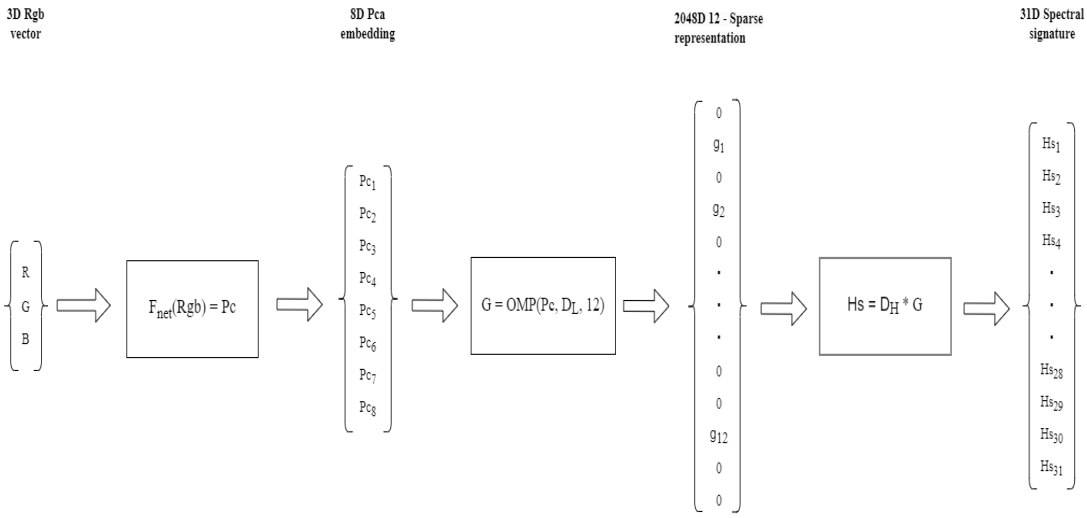
Άρα στο πρώτο βήμα και πάλι για κάθε RGB διάνυσμα εκτιμάμε το περιεχόμενο της PCA με χρήση ενός νευρωνικού δικτύου: $F_{net}(Rgb) = Pc$.

Στη συνέχεια εντοπίζουμε την αραιή περιγραφή G με χρήση του αλγορίθμου OMP και του λεξικού D_l.

$$G = omp(Pc, D_l, 12), \qquad extract\ 12 - sparse\ solution \tag{79}$$

Και συνθέτουμε το πολυφασματικό pixel χρησιμοποιώντας τις αντίστοιχες στήλες του λεξικού D_h .

$$Hs = D_h * G, \ \ synthesize\ Hs\ using\ G \tag{80}$$



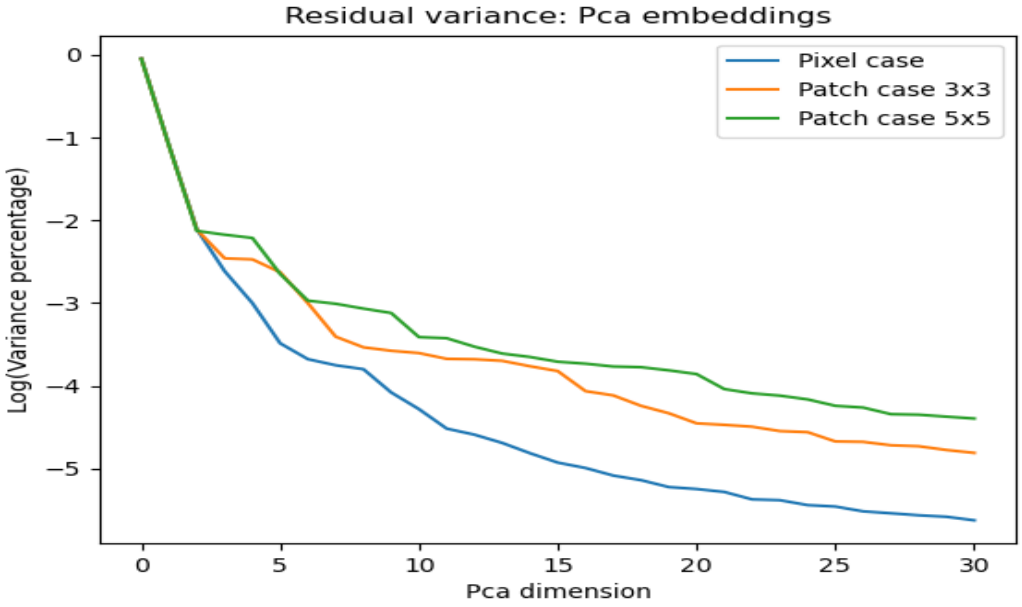
Σχήμα 3.12: Μπλοκ διάγραμμα δεύτερης προτεινόμενης μεθόδου επίλυσης.

3.6 Ανακατασκευή βασισμένη σε patches

Ένα βασικό χαρακτηριστικό της υλοποίησης μας είναι πως εφαρμόζεται σε pixel της εικόνας. Το σύνολο της πληροφορίας δηλαδή που αξιοποιεί εξαντλείται στις συσχετίσεις των τριών τιμών των RGB καναλιών. Οι εικόνες όμως αποτελούν χωρικά σήματα όπου το κάθε pixel εξαρτάται έντονα από την γειτονιά του. Την χωρική αυτή συσχέτιση δεν την αποτυπώνουμε στην μέχρι τώρα υλοποίηση, συνεπώς μια πιθανή βελτίωση που δοκιμάσαμε ήταν να σχεδιάσουμε ένα σύστημα βασισμένο σε patches δηλαδή μικρά τμήματα της εικόνας.

3.6.1 Αραιότητα πολυφασματικών παραθύρων

Για την δημιουργία των σετ εκπαίδευσης και αξιολόγησης ακολουθούμε την ίδια διαδικασία όπου οι εικόνες χωρίζονται σε δύο σύνολα των 100 εικόνων τυχαία. Από τις 100 εικόνες εκπαίδευσης αντλούμε 2000 τυχαία patches δημιουργώντας έτσι ένα σύνολο 200.000 δεδομένων εκπαίδευσης. Για την λήψη του dataset δημιουργήθηκε custom συνάρτηση η οποία αρχικά περιορίζει τη χωρική ανάλυση της εικόνας, ώστε να χωράει ένας ακέραιος αριθμός από patches σε κάθε της διάσταση και στη συνέχεια επιλέγει τυχαία μεταξύ τους. Επιλέξαμε μη επικαλυπτόμενα παράθυρα διάστασης 3x3 και 5x5 δηλαδή γειτονιές των 9 pixel(για την περίπτωση 3x3), των 25 pixel (για την περίπτωση 5x5) και τοποθετούμε τα κανάλια σε σειρά. Τα διανύσματα εισόδου εξόδου αποκτούν έτσι διαστάσεις 27(9*3) και 279(9*31) για την 3x3 περίπτωση και 75(25*3), 775(25*31) για την 5x5 περίπτωση αντίστοιχα. Οι κανονικοποιήσεις παραμένουν ως έχουν με μόνη διαφορά ότι η μέση τιμή και η διασπορά τόσο για το Rgb patch όσο και το πολυφασματικό patch υπολογίζονται για 27, 279 και 75, 775 κανάλια αντίστοιχα. Στην συνέχεια εφαρμόζουμε την PCA και τυπώνουμε τα διαγράμματα της εναπομένουσας διασποράς για να εξετάσουμε το βαθμό συμπίεσης και το σύνολο των συντελεστών που θα χρησιμοποιηθούν για την ανακατασκευή.



Σχήμα 3.13: Διάγραμμα εναπομένουσας διασποράς(residual variance) μετά την εφαρμογή της PCA. Η ενέργεια συγκεντρώνεται και πάλι στους πρώτους σημαντικότερους συντελεστές.

Από το διάγραμμα παρατηρούμε ότι ο βαθμός συμπίεσης των patches είναι αρκετά υψηλός. Με μόλις ένα στοιχείο περιγράφουμε το 90.02% της πληροφορίας στην 3x3 περίπτωση και το 89.1% στην 5x5. Ενώ μιλάμε για 9 και 25 φορές περισσότερα pixel με μόλις ένα στοιχείο αποτυπώνουμε το μεγαλύτερο ποσοστό της ενέργειας του σήματος, με παρόμοιο βαθμό συμπίεσης όπως και στην pixel περίπτωση (90.7% το πρώτο στοιχείο). Στη συνέχεια τα τρία προφίλ συμπίεσης εμφανίζουν διαφορές μόνο με τη χρήση της λογαριθμικής απεικόνισης. Οι περιπτώσεις των παραθύρων όπως είναι λογικό έχουν μεγαλύτερη κατανομή ενέργειας στους υπόλοιπους συντελεστές λόγω της αυξημένης πολυπλοκότητας τους, με την 5x5 περίπτωση να αποτελεί την δυσκολότερη περίπτωση μείωσης της διάστασης. Για την μελέτη του βαθμού συμπίεσης υπολογίσαμε πόσα στοιχεία χρειάζεται η μειωμένη αναπαράσταση ώστε να περιγράψει το 99.95% της πληροφορίας του σήματος. Για την pixel περίπτωση χρειάζονται 8 στοιχεία δηλαδή σε σχέση με το αρχικό μέγεθος του διανύσματος (31) έχουμε το 25.8% του μεγέθους. Για την 3x3 περίπτωση χρειαζόμαστε 20 στοιχεία δηλαδή το 7.1% του αρχικού μεγέθους. Και για την περίπτωση 5x5 χρειαζόμαστε 38 στοιχεία δηλαδή το 4.9% του μεγέθους. Η βελτίωση στην συμπίεση προκύπτει από το γεγονός ότι τα φυσικά αντικείμενα που αποτυπώνουμε σε εικόνες έχουν περιεχόμενο χαμηλής συχνότητας. Η φωτεινότητα σε μια μικρή περιοχή αλλάζει πολύ σιγά έτσι τα pixel μοιάζουν πολύ, γεγονός που οδηγεί στην ευκολότερη ομαδοποίηση τους. Μάλιστα αυξάνοντας το παράθυρο των δεδομένων έχουμε ακόμη περισσότερα όμοια pixel και επιτυγχάνουμε ακόμη μεγαλύτερους βαθμούς συμπίεσης.

3.6.2 Απεικονίσεις μεταξύ των χώρων

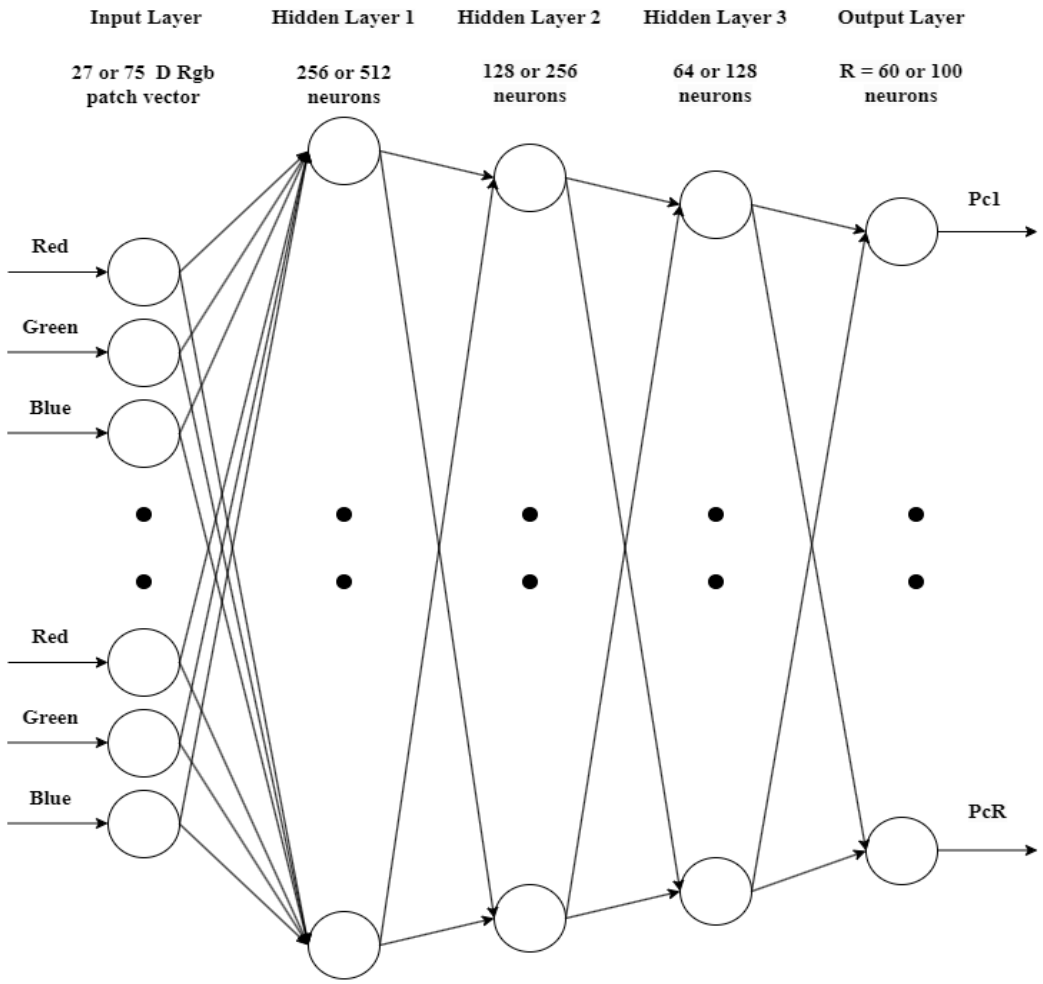
Για την απεικόνιση από τον RGB χώρο εισόδου στον χώρο της PCA αναπτύχθηκε πάλι μια υλοποίηση βασισμένη στον MLP. Η εκπαίδευση ενός RBF δικτύου ήταν αδύνατη καθώς η ιακωβιανή ξεπερνούσε κατά πολύ το μέγιστο μέγεθος πίνακα που μπορούσε να χειριστεί η Matlab. Εδώ παρατηρούμε τον περιορισμό που έχει ο αλγόριθμος Levenberg-Marquardt και η πλήρης εκτίμηση του Hessian πίνακα, αφού καθίσταται αδύνατη η υλοποίηση του για ένα σχετικά μικρό δίκτυο το οποίο οι βασικοί αλγόριθμοι εκπαίδευσης για έναν MLP χειρίζονται με ευκολία.

Η κανονικοποίηση των δεδομένων εισόδου γίνεται αφαιρώντας τη μέση τιμή και διαιρώντας με την διασπορά για κάθε ένα από τα 27 ή 75 κανάλια εισόδου. Τα διανύσματα της PCA κανονικοποιούνται και πάλι στο [0 1] με χρήση των min max τιμών. Τα καλύτερα αποτελέσματα και για τις δύο περιπτώσεις τα πήραμε με χρήση του Adam optimizer, με batch size 50, $\beta_1 = 0.75$, $\beta_2 = 0.999$, $m = 0.0001$ ενώ η εκπαίδευση κράτησε για 120 εποχές.

Η καλύτερη αρχιτεκτονική για την 3x3 περίπτωση αποτελείται από 3 κρυφά επίπεδα με διαστάσεις 256, 128, 64 και συνάρτηση ενεργοποίησης την $ReLU$ ενώ το επίπεδο εξόδου πραγματοποιεί και πάλι έναν γραμμικό μετασχηματισμό χωρίς μη γραμμική ενεργοποίηση και η διάσταση του είναι 60 με σκοπό να εκτιμήσει τις 60 κυρίαρχες συνιστώσες .

Η καλύτερη αρχιτεκτονική για την 5x5 περίπτωση αποτελείται από 3 κρυφά επίπεδα με διαστάσεις 512, 256, 128 και συνάρτηση ενεργοποίησης την $ReLU$ ενώ το επίπεδο εξόδου πραγματοποιεί και πάλι έναν γραμμικό μετασχηματισμό χωρίς μη γραμμική ενεργοποίηση και η διάσταση του είναι 100 με σκοπό να εκτιμήσει τις 100 κυρίαρχες συνιστώσες. Ακολουθούν το σχήμα των αρχιτεκτονικών.

Για την απεικόνιση από τον χώρο της PCA στον πολυφασματικό, αναπτύχθηκε μια μέθοδος βασισμένη στον αντίστροφο μετασχηματισμό της PCA. Με βάση την εκτίμηση των 60 ή 100 κυρίαρχων συνιστωσών συνθέτουμε το πολυφασματικο patch απευθείας. Η μέθοδος των συζευγμένων λεξικών είχε πολύ υψηλούς χρόνους εκπαίδευσης και πολύ μεγάλα μεγέθη λεξικών χωρίς να επιτυγχάνει αξιόλογα αποτελέσματα, συνεπώς δεν επεκτάθηκε και στην περίπτωση των patches.



Σχήμα 3.14: Σχηματική δομή του Multilayer Perceptron για την patch περίπτωση. Στην είσοδο δίνονται οι RGB τιμές ενός patch διάστασης 27 ή 75 και εκτιμάται το περιεχόμενο των 60 ή 100 κυρίαρχων συνιστωσών.

3.6.3 Τρίτη προτεινόμενη μέθοδος επίλυσης

Καταλήγουμε έτσι στην τρίτη προτεινόμενη μέθοδο επίλυσης. Για κάθε RGB patch διάστασης είτε 27 είτε 75 χρησιμοποιούμε ένα από τα νευρωνικά δίκτυα για να εκτιμήσουμε το περιεχόμενο των 60 και 100 κυρίαρχων συνιστωσών αντίστοιχα. Στην συνέχεια αντιστρέφουμε την κανονικοποίηση χρησιμοποιώντας τις min max τιμές των emdeddings της PCA. Με χρήση του πίνακα W^T υπολογίζουμε τον γραμμικό συνδυασμό των 60 ή 100 κυρίαρχων συνιστωσών. Και τέλος αντιστρέφουμε την τελική κανονικοποίηση χρησιμοποιώντας τις τιμές τις μέσης τιμής και της διασποράς των 279 ή 775 πολυφασματικών καναλιών αντίστοιχα. Εάν θεωρήσουμε Rgb τα RGB patches, Pc_R τις συμπιεσμένες περιγραφές και Hs_C τα πολυφασματικά patches.

Αρχικά εκτιμάμε το περιεχόμενο της PCA με τις σχέσεις:

$$F_{net}(Rgb) = Pc'_R \tag{81}$$

και

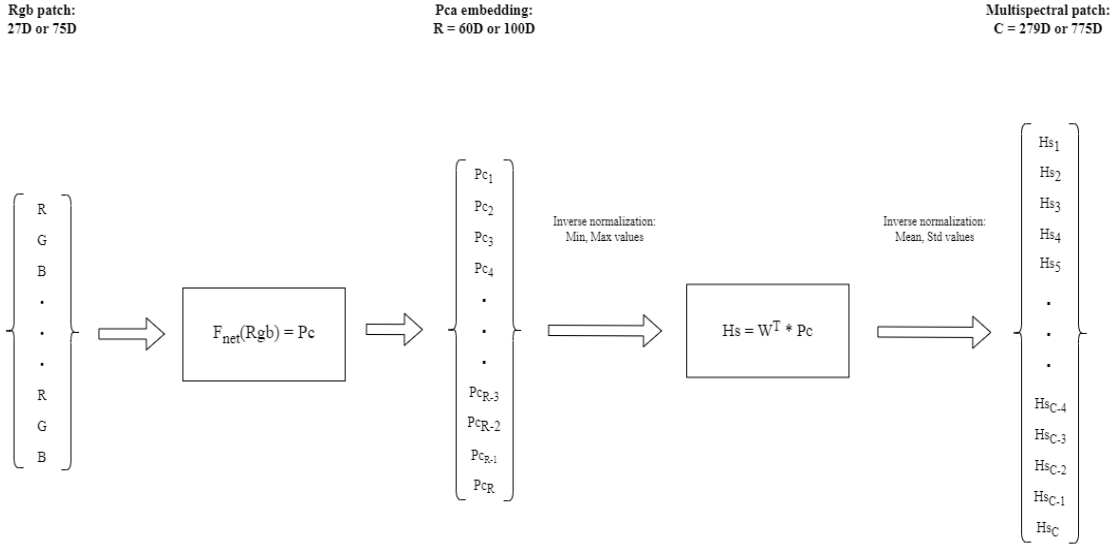
$$Pc_R = Pc'_R \ (maxPc_R - minPc_R) + \min Pc_R \tag{82}$$

Και τελικά ανακατασκευάζουμε το πολυφασματικό patch με τις σχέσεις:

$$hs' = W^T * Pc_R \tag{83}$$

Και στην συνέχεια

$$Hs_c = hs'_i * Std_i + m_i \ \ where\ i = 1, \dots, 279\ or\ 1, \dots, 775 \tag{84}$$



Σχήμα 3.15: Μπλόκ διάγραμμα τρίτης προτεινόμενης μεθόδου επίλυσης.

ΚΕΦΑΛΑΙΟ 4

Αποτελέσματα

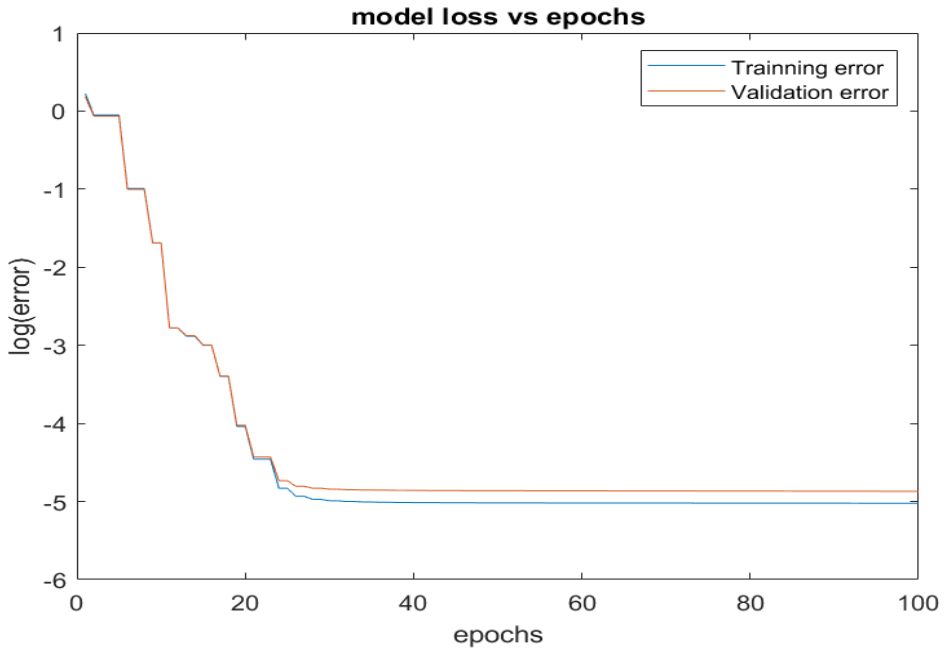
Στα κεφάλαια 2 και 3, αναλύθηκε το θεωρητικό υπόβαθρο της παρούσας εργασίας καθώς και το σύνολο των υλοποιήσεων που αναπτύχθηκαν για την εκτίμηση της πολυφασματικής πληροφορίας μιας φυσικής σκηνής από εικόνες χαμηλότερης φασματικής ανάλυσης(RGB). Σε αυτό το κεφάλαιο, παρουσιάζεται η πορεία της εκπαίδευσης των εκάστοτε αρχιτεκτονικών καθώς και τα αποτελέσματα εκτίμησης για διάφορα πειράματα που πραγματοποιήθηκαν. Αρχικά παρουσιάζονται τα αποτελέσματα εκπαίδευσης των νευρωνικών δικτύων τόσο για την περίπτωση εφαρμογής σε pixel όσο και σε patches αλλά και η συμπεριφορά των συζευγμένων λεξικών για διαφορετικά μεγέθη λεξικού και διαφορετικούς βαθμούς αραιότητας. Στην συνέχεια συγκρίνονται οι χρόνοι εκπαίδευσης της εκάστοτε αρχιτεκτονικής. Ακολουθεί η μελέτη της επίδρασης του αριθμού των εκτιμώμενων συνιστωσών και της φασματικής απόκρισης της κάμερας στο τελικό σφάλμα του συστήματος. Παρουσιάζονται τα αποτελέσματα του τεστ σε ολόκληρες εικόνες, με σκοπό την σύγκριση με κάποιες από τις καλύτερες σύγχρονες μεθόδους που εμφανίζονται στην βιβλιογραφία και πραγματοποιείται επίσης η χρονική και χωρική σύγκριση των αρχιτεκτονικών. Τέλος παρουσιάζονται τα οπτικά αποτελέσματα ανακατασκευής μιας ολόκληρης εικόνας και αναλύεται η λειτουργία των μεθόδων.

4.1 Αποτελέσματα εκπαίδευσης

Σε αυτό το σημείο, αρχικά παρουσιάζονται τα διαγράμματα που δείχνουν την συμπεριφορά των μοντέλων μας κατά την διάρκεια της εκπαίδευσης, για τις διάφορες αρχιτεκτονικές νευρωνικών δικτύων που αναπτύχθηκαν. Στη συνέχεια, εξετάζεται η επίδραση που έχει το μέγεθος του λεξικού και ο βαθμός αραιότητας στην τελική εκτίμηση. Και τέλος αναλύονται οι διαφορετικοί αλγόριθμοι μάθησης που υλοποιήθηκαν και συγκρίνονται οι τελικοί χρόνοι εκπαίδευσης για το κάθε μοντέλο.

4.1.1 Εκπαίδευση Radial Basis Function neural network

Για την εκπαίδευση των RBF δικτύων σχεδιάστηκε, όπως αναφέραμε και νωρίτερα, custom συνάρτηση που υλοποιεί τον αλγόριθμο Levenberg-Marquardt. Οι κώδικες αναπτύχθηκαν σε MATLAB ενώ η εκπαίδευση πραγματοποιήθηκε στην cloud έκδοση της. Η περιπλοκότητα του αλγορίθμου οδήγησε σε μεγάλους χρόνους εκπαίδευσης και σε απαγορευτικά μεγάλες απαιτήσεις σε μνήμη ώστε να επεκταθεί η εφαρμογή των δικτύων σε patches.

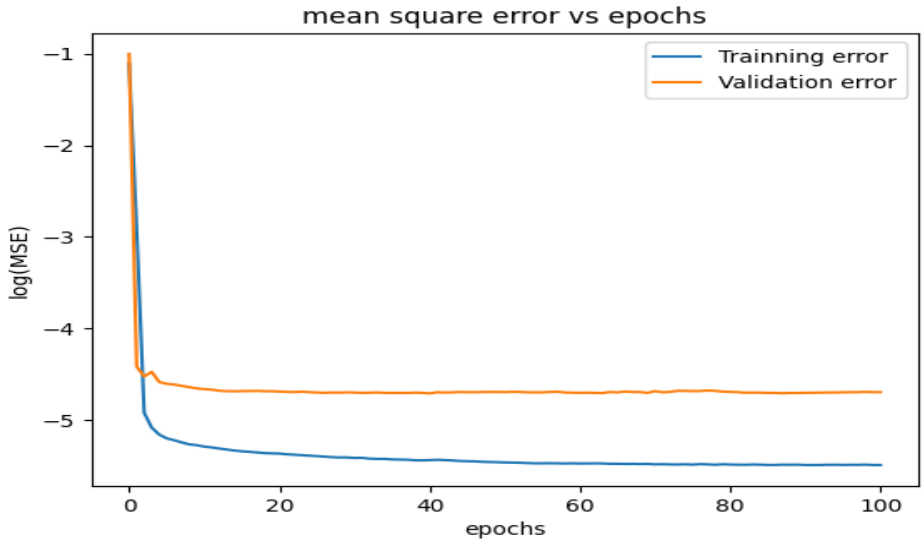


Σχήμα 4.1: Διάγραμμα σύγκλισης αλγορίθμου Levenberg-Marquardt για την εκπαίδευση του Radial Basis neural network. Στο διάγραμμα περιλαμβάνεται ο λογάριθμος του σφάλματος ως προς τις εποχές εκπαίδευσης τόσο για τα παραδείγματα εκπαίδευσης όσο και για τα δεδομένα του validation set.

Στο σχήμα 4.1 παρουσιάζεται το διάγραμμα σύγκλισης του αλγορίθμου Levenberg-Marquardt για την εκπαίδευση του Radial Basis Function δικτύου. Ο αλγόριθμος ελαχιστοποιεί το μέσο τετραγωνικό σφάλμα της εκτίμησης. Στο διάγραμμα παρουσιάζεται ο λογάριθμος του σφάλματος ως προς τις εποχές εκπαίδευσης, με σκοπό να γίνουν εμφανής οι διαφορές. Με μπλε χρώμα απεικονίζεται το σφάλμα εκτίμησης των δεδομένων εκπαίδευσης ενώ με κόκκινο το σφάλμα σε ένα validation set άγνωστων παραδειγμάτων. Η χρήση του validation set γίνεται για να εξετάσουμε αν ο αλγόριθμος γενικεύει την γνώση του σε άγνωστα παραδείγματα και για αυτό το λόγο, το σφάλμα είναι λίγο μεγαλύτερο από αυτό των παραδειγμάτων που χρησιμοποιεί για να μάθει. Βασικό χαρακτηριστικό της καμπύλης είναι η τμηματική συμπεριφορά του αλγορίθμου. Ο LM μόλις εντοπίσει κυρτότητα στην περιοχή της συνάρτησης σφάλματος αξιοποιεί την πληροφορία και με μια δευτεροβάθμια εκτίμηση της συνάρτησης κάνει βουτιά για να βρει το ελάχιστο. Μόλις η εκτίμηση του Hessian πίνακα μέσω της σχέσης $J^T J$ δώσει μικρές τιμές, αυξάνει σταδιακά το συντελεστή εκπαίδευσης m, ώστε να πραγματοποιήσει ένα βήμα με όλο και μεγαλύτερη βαρύτητα στον gradient descent και να ξεφύγει από την συγκεκριμένη περιοχή. Αυτό παρατηρείται στις περιοχές που η καμπύλη σφάλματος είναι παράλληλη στον άξονα των εποχών. Μόλις η κυρτότητα αλλάξει και ο πίνακας $J^T J$ αρχίσει και πάλι να δίνει μεγαλύτερες τιμές, μειώνεται ο όρος m και κυριαρχεί και πάλι το βήμα με βάση την δευτεροβάθμια προσέγγιση που ορμά στο ελάχιστο.

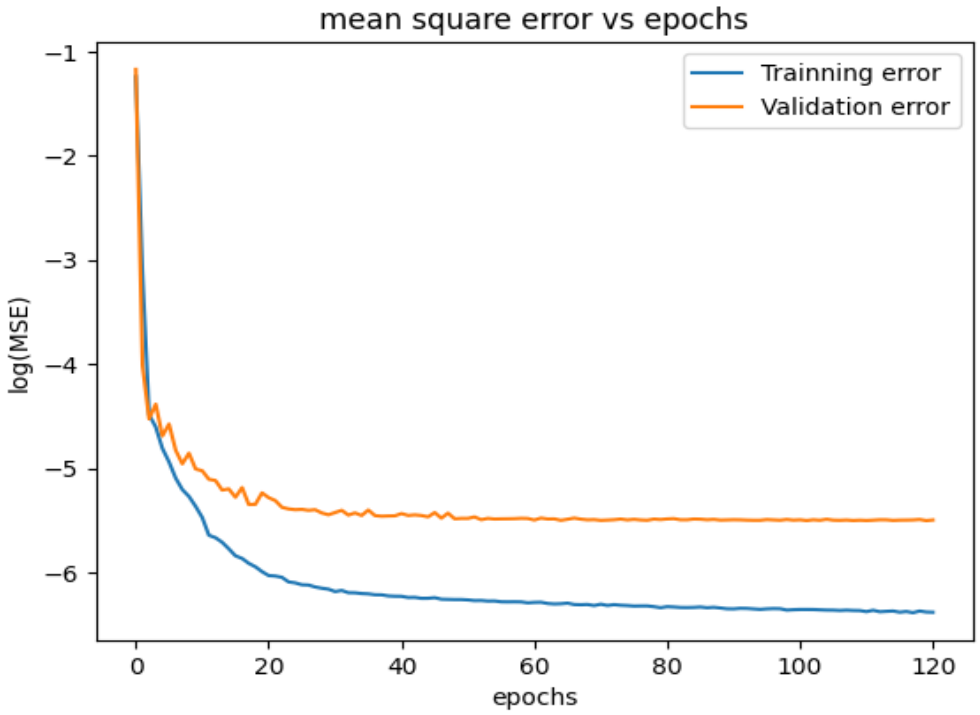
4.1.2 Εκπαίδευση Multilayer Perceptron

Για την εκπαίδευση των Multilayer Perceptron δικτύων η διαδικασία ήταν αρκετά πιο απλή καθώς το πακέτο της Pytorch παρέχει τρομερή ευκολία στον σχεδιασμό και στον χειρισμό των δικτύων. Επίσης η παραλληλοποίηση των διαδικασιών και η χρήση GPU επιταχύνουν πολύ την εκπαίδευση γεγονός που αποτυπώνεται και στους χρόνους που παρουσιάζονται στη συνέχεια. Τα δίκτυα εκπαιδεύθηκαν να μειώνουν το μέσο τετραγωνικό σφάλμα της εκτίμησης και με χρήση ενός validation set εξετάζουμε και πάλι το κατά πόσο γενικεύει ο αλγόριθμος τη γνώση του σε άγνωστα παραδείγματα. Σε αντίθεση με τον Levenberg-Marquardt, που σε κάθε εποχή πραγματοποιεί μία αλλαγή των συντελεστών του δικτύου χρησιμοποιώντας όλη την πληροφορία των παραδειγμάτων εκπαίδευσης, ο αλγόριθμος Adam τεμαχίζει το dataset σε batches και πραγματοποιεί αλλαγές εκτιμώντας το gradient από ένα μικρότερο σύνολο διανυσμάτων. Αυτό του δίνει τη δυνατότητα να τρέξει αποδοτικά σε τεράστια δίκτυα και σε ικανοποιητικούς χρόνους κάτι που δεν ήταν εφικτό με τον αλγόριθμο LM.



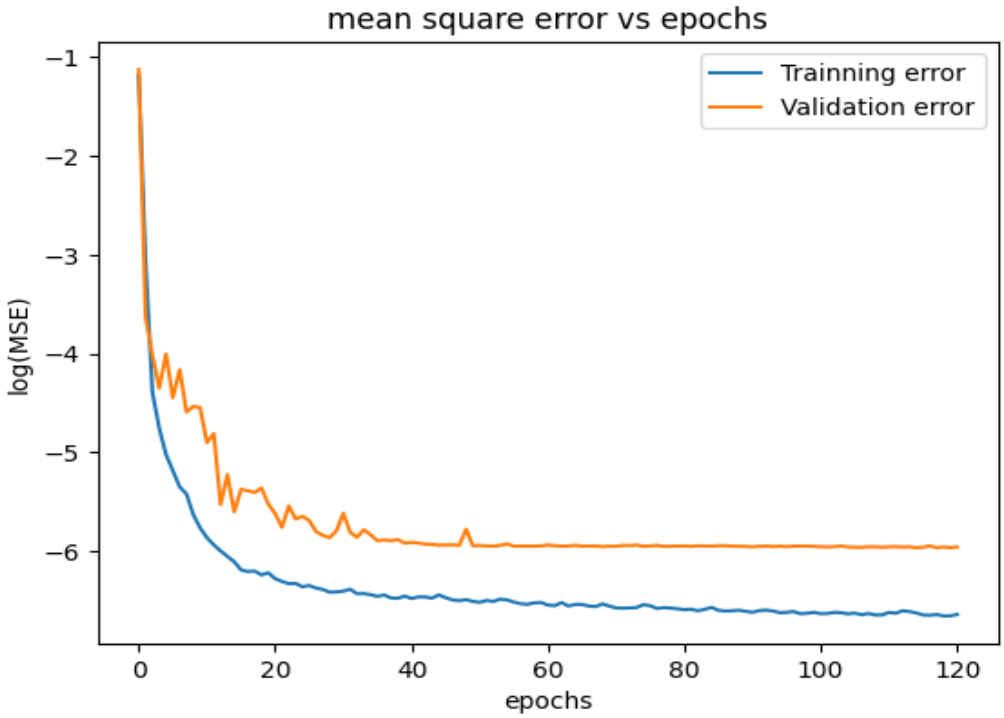
Σχήμα 4.2: Διάγραμμα σύγκλισης του αλγορίθμου Adam για την εκπαίδευση του Multilayer Perceptron δικτύου για την περίπτωση που εφαρμόζεται σε pixel. Στο διάγραμμα περιλαμβάνεται ο λογάριθμος του σφάλματος ως προς τις εποχές εκπαίδευσης τόσο για τα παραδείγματα εκπαίδευσης όσο και για τα δεδομένα του validation set.

Από το σχήμα 4.2 παρατηρούμε ότι η σύγκλιση είναι αρκετά ομαλή και δεν εμφανίζεται η τμηματική συμπεριφορά του προηγούμενου διαγράμματος. Χρησιμοποιήθηκε και πάλι η λογαριθμική απεικόνιση για να φανούν οι διαφορές, με το σφάλμα στα άγνωστα παραδείγματα του validation set να είναι πιο υψηλό όπως προβλέπεται. Χαρακτηριστικό της καμπύλης είναι πως ενώ το σφάλμα στα δεδομένα εκπαίδευσης συνεχίζει και μειώνεται το σύστημα δεν μαθαίνει κάτι καινούργιο για τα άγνωστα.



Σχήμα 4.3: Διάγραμμα σύγκλισης του αλγορίθμου Adam για την εκπαίδευση του Multilayer Perceptron δικτύου για την περίπτωση που εφαρμόζεται σε patches διάστασης 3x3. Στο διάγραμμα περιλαμβάνεται ο λογάριθμος του σφάλματος ως προς τις εποχές εκπαίδευσης τόσο για τα παραδείγματα εκπαίδευσης όσο και για τα δεδομένα του validation set.

Αντίστοιχη συμπεριφορά εμφανίζεται και στην εκπαίδευση του MLP για την περίπτωση των 3x3 patches όπως φαίνεται στο σχήμα 4.3. Η μόνη εμφανής διαφορά είναι πως τόσο το σφάλμα, τόσο των άγνωστων δεδομένων του validation set όσο και των δεδομένων εκπαίδευσης, εμφανίζουν διακυμάνσεις που είναι απόρροια της δυσκολότερης απεικόνισης που επιδιώκει το δίκτυο να προσεγγίσει. Η μείωση του συντελεστή β_1 στην τιμή 0.75 οδήγησε σε ομαλότερα αποτελέσματα ενώ η οποιαδήποτε αλλαγή στον συντελεστή β_2 έκανε πιο έντονες διακυμάνσεις και στις δύο περιπτώσεις(3x3,5x5).

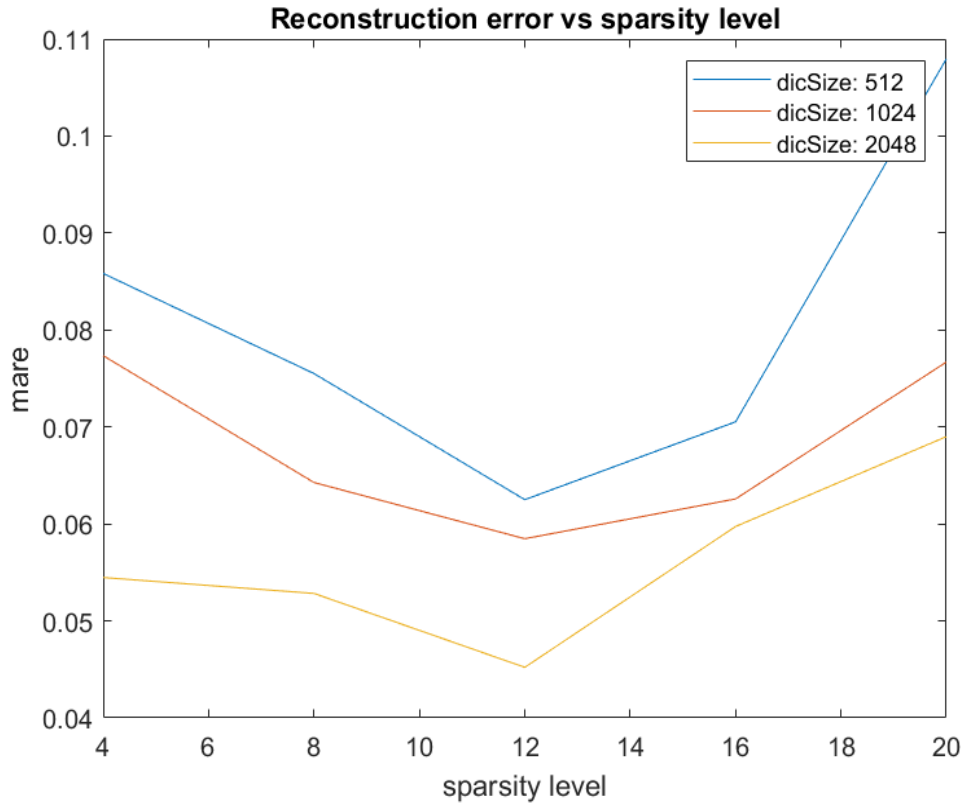


Σχήμα 4.4: Διάγραμμα σύγκλισης του αλγορίθμου Adam για την εκπαίδευση του Multilayer Perceptron δικτύου για την περίπτωση που εφαρμόζεται σε patches διάστασης 5x5. Στο διάγραμμα περιλαμβάνεται ο λογάριθμος του σφάλματος ως προς τις εποχές εκπαίδευσης τόσο για τα παραδείγματα εκπαίδευσης όσο και για τα δεδομένα του validation set.

Η ίδια συμπεριφορά εμφανίζεται και στην περίπτωση των patches 5χ5 με τις διακυμάνσεις να είναι ακόμη πιο έντονες όπως βλέπουμε στο σχήμα 4.4.

4.1.3 Εκπαίδευση συζευγμένων λεξικών

Για την περίπτωση των συζευγμένων λεξικών εξετάστηκαν διαφορετικά μεγέθη λεξικού και διαφορετικοί βαθμοί αραιότητας. Η εκπαίδευση πραγματοποιήθηκε για 40 επαναλήψεις. Σε πρώτο χρόνο υπολογίζονται οι αραιές αναπαραστάσεις με χρήση του OMP και το λεξικό σταθερό και στη συνέχεια διορθώνονται τα άτομα του λεξικού με εφαρμογή του K-SVD και τις αναπαραστάσεις σταθερές. Οι 40 επαναλήψεις επιλέχθηκαν σε αντιστοιχία με άλλες μεθόδους, πέραν των 10 όμως η διαφορά στο τελικό σφάλμα δεν ήταν αισθητή.



Σχήμα 4.5: Διάγραμμα σφάλματος ανακατασκευής για διαφορετικά μεγέθη λεξικού και διαφορετικούς βαθμούς αραιότητας.

Από το σχήμα 4.5 παρατηρούμε ότι η αύξηση του μεγέθους του λεξικού μειώνει το σφάλμα για όλα τα επίπεδα αραιότητας. Η χρήση περισσότερων ατόμων δίνει τη δυνατότητα στο λεξικό να απλώσει και να διαχωρίσει ακόμη καλύτερα τις λεπτομέρειες των δεδομένων. Η αύξηση του βαθμού αραιότητας και αυτή με τη σειρά της βελτιώνει την τελική ανακατασκευή με το ελάχιστο να εμφανίζεται στην τιμή $T_0 = 12$. Περαιτέρω αύξηση των ατόμων που χρησιμοποιούνται όμως αυξάνει το σφάλμα, καθώς και τους χρόνους ανακατασκευής και εκπαίδευσης. Αυτό αποτελεί μια μορφή overfitting και μας δείχνει ότι η αύξηση της περιπλοκότητας ενός συστήματος δεν σημαίνει υποχρεωτικά βελτίωση της τελικής συμπεριφοράς του.

4.1.4 Χρόνοι σύγκλισης διαφορετικών αρχιτεκτονικών

Στον πίνακα 4.1 παρουσιάζονται οι χρόνοι σύγκλισης για την εκπαίδευση της κάθε διαφορετικής αρχιτεκτονικής. Η χρήση διαφορετικών υπολογιστικών συστημάτων και βιβλιοθηκών δεν καθιστά τόσο αντικειμενική αυτή τη μέτρηση αλλά δίνει μια γενική ιδέα για την ευκολία εκπαίδευσης του κάθε μοντέλου. Στην περίπτωση του αλγορίθμου Levenberg-Marquardt ο κώδικας αναπτύχθηκε από το μηδέν γεγονός που τον καθιστά μη βέλτιστο σε σχέση με την βιβλιοθήκη Pytorch που χρησιμοποιήθηκε για τις αρχιτεκτονικές των MLP

δικτύων. Η μεγάλη υπολογιστική πολυπλοκότητα του αλγορίθμου όμως, θα δημιουργούσε όπως και να έχει μεγαλύτερους χρόνους εκπαίδευσης. Βασική παρατήρηση πάντως είναι πως η χρήση GPU και η παραλληλοποίηση των διαδικασιών στην περίπτωση της Pytorch μειώνει πολύ τους χρόνους εκπαίδευσης τόσο σε σύγκριση με τις CPU based εκπαιδεύσεις των λεξικών και του Radial Basis Function δικτύου όσο και μεταξύ των διαφορετικών αρχιτεκτονικών των MLP δικτύων. Χαρακτηριστικό παράδειγμα αποτελούν τα αρκετά μεγαλύτερα δίκτυα για τις patch based περιπτώσεις όπου παρόλο που το σύνολο των συντελεστών των δικτύων, το μέγεθος των διανυσμάτων αλλά και των εποχών είναι αρκετά μεγαλύτερα, η χρήση μεγαλύτερου batch δεδομένων οδηγεί σε μικρότερους χρόνους εκπαίδευσης από την απλή pixel περίπτωση.

Model architecture	Epochs/iterations	Time (sec)
Radial Basis neural network	100	5868
Multilayer Perceptron (pixel case)	100	1418
Multilayer Perceptron (patch case 3x3)	120	1130
Multilayer Perceptron (patch case 5x5)	120	1189
Dictionaries	40	2337

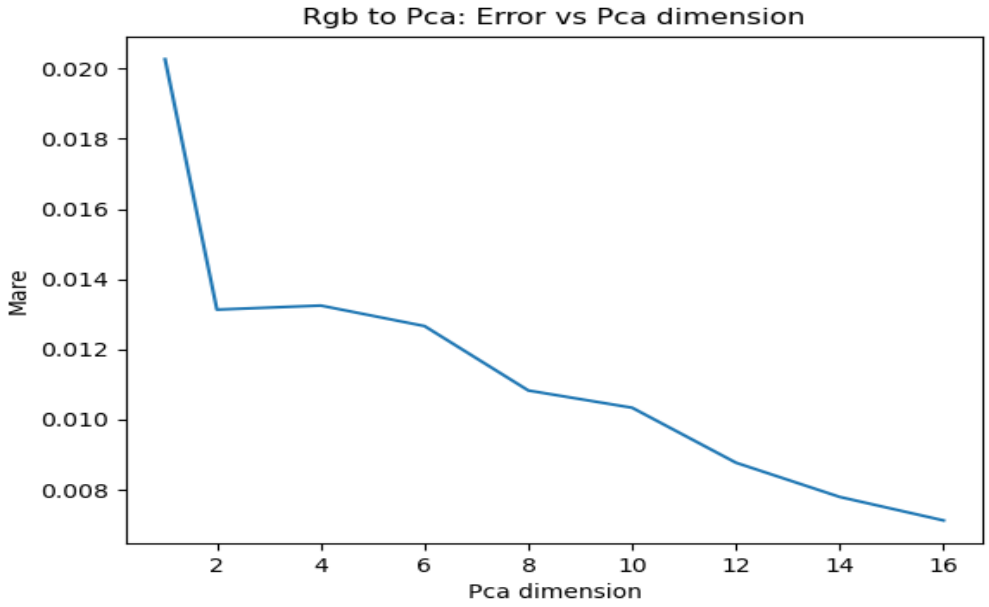
Πίνακας 4.1: Σύγκριση εποχών και χρόνων εκπαίδευσης της εκάστοτε αρχιτεκτονικής.

4.2 Αποτελέσματα ανακατασκευής

Σε αυτό το σημείο παραθέτουμε τα αποτελέσματα ανακατασκευής όλων των μεθόδων, και μελετάμε την συμπεριφορά των συστημάτων μας μέσα από μια σειρά πειραμάτων. Αρχικά εξετάζουμε την επίδραση του αριθμού των εκτιμώμενων συνιστωσών αλλά και της της απόκρισης ευαισθησίας της κάμερας στο τελικό σφάλμα. Στην συνέχεια πραγματοποιούμε ένα τεστ σε ολόκληρες εικόνες με σκοπό τη σύγκριση με κάποιες από τις καλύτερες σύγχρονες μεθόδους που αναφέρονται στην [8]. Ακολουθεί η χρονική και χωρική σύγκριση των μεθόδων. Τέλος εξετάζουμε τα οπτικά αποτελέσματα ανακατασκευής για όλες τις μεθόδους ανακατασκευάζοντας μια τυχαία εικόνα από το dataset.

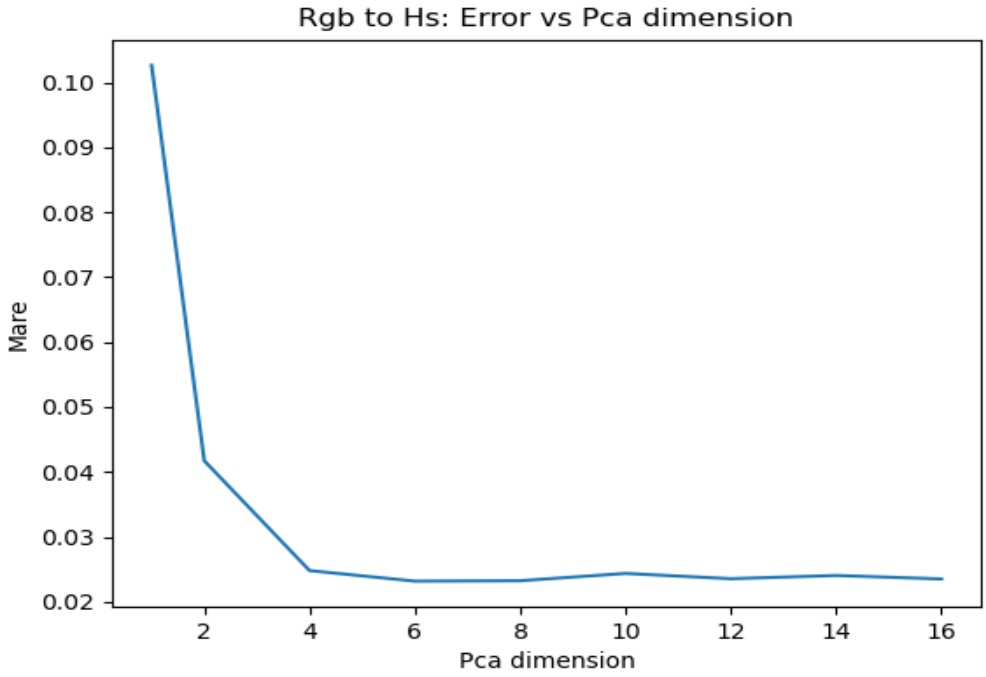
4.2.1 Μελέτη της επίδρασης του αριθμού των εκτιμώμενων συνιστωσών στο σφάλμα εκτίμησης

Για την μελέτη της επίδρασης του αριθμού των εκτιμώμενων συνιστωσών στο σφάλμα εκτίμησης, σχεδιάσαμε διαφορετικά δίκτυα που ανακατασκευάζουν τα πολυφασματικά σήματα εκτιμώντας από 1 μέχρι 16 κυρίαρχες συνιστώσες. Υπολογίσαμε το σφάλμα τόσο για το πρώτο βήμα της εκτίμησης όσο και για τη συνολική απεικόνιση. Τα σχήματα παρουσιάζονται στη συνέχεια.



Σχήμα 4.6: Διάγραμμα σφάλματος για την απεικόνιση από τον RGB χώρο στον χώρο της PCA ως συνάρτηση του αριθμού των εκτιμώμενων κυρίαρχων συνιστωσών (Principal Components).

Από το σχήμα 4.6 παρατηρούμε ότι το σφάλμα εκτίμησης του πρώτου βήματος μειώνεται συνεχώς όσο αυξάνουμε την διάσταση των διανυσμάτων της PCA. Κάθε επόμενο στοιχείο έχει όλο και μικρότερη ενέργεια, συνεπώς το σύστημα αρκεί να εντοπίσει μόνο την μέση τιμή της συγκεκριμένης συνιστώσας. Επιτυγχάνει έτσι όλο και μικρότερο μέσο σφάλμα χωρίς όμως να εντοπίζει τις διαφοροποιήσεις της, οι οποίες αποτυπώνουν τη ζητούμενη πληροφορία. Η αύξηση της διάστασης των εκτιμώμενων κυρίαρχων συνιστωσών λοιπόν, ρίχνει πολύ το σφάλμα εκτίμησης του πρώτου βήματος, ειδικά στα δίκτυα που εφαρμόζονται σε patches και εκτιμούν 60 και 100 συνιστώσες αντίστοιχα, αλλά το τελικό σφάλμα ανακατασκευής παραμένει κοντά για όλες τις υλοποιήσεις.

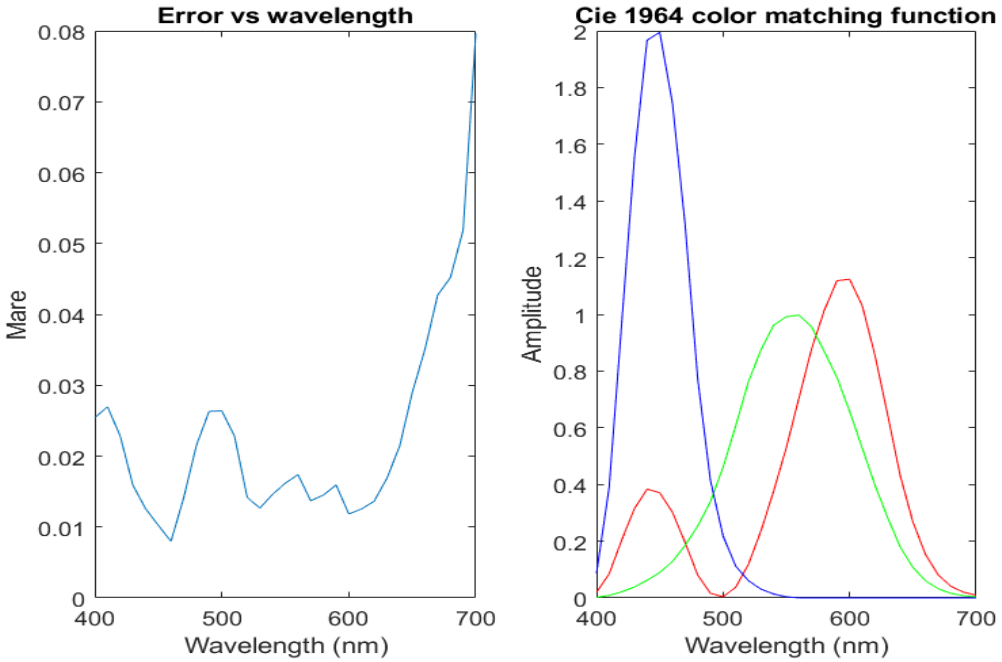


Σχήμα 4.7: Διάγραμμα σφάλματος για την απεικόνιση από τον RGB χώρο στον πολυφασματικό ως συνάρτηση του αριθμού των εκτιμώμενων κυρίαρχων συνιστωσών (Principal Components).

Στο σχήμα 4.7 φαίνεται το συνολικό σφάλμα ανακατασκευής του συστήματος για διαφορετικό αριθμό κυρίαρχων συνιστωσών. Με μόλις μια συνιστώσα εκτιμάται ικανοποιητικά το 90% της πολυφασματικής πληροφορίας, και καθώς αυξάνουμε τον αριθμό επιτυγχάνουμε όλο και καλύτερα αποτελέσματα. Η καλύτερη εκτίμηση πραγματοποιείται για διάσταση 6-8 και καθώς προσθέτουμε στοιχεία το σφάλμα εμφανίζει ένα είδος κορεσμού. Αυτό οφείλεται στην όλο και μικρότερη πληροφορία που περιέχει το κάθε στοιχείο, που από την μία η συνεισφορά του φθίνει εκθετικά και δεν περιέχει αρκετή πληροφορία για το σήμα μας, από την άλλη δεν εκτιμάται ικανοποιητικά στο πρώτο βήμα και δεν αποτυπώνονται ορθά οι ζητούμενες διαφοροποιήσεις του αλλά μόνο η μέση τιμή του όπως αναφέραμε.

4.2.2 Μελέτη της επίδρασης της κάμερας στο σφάλμα εκτίμησης

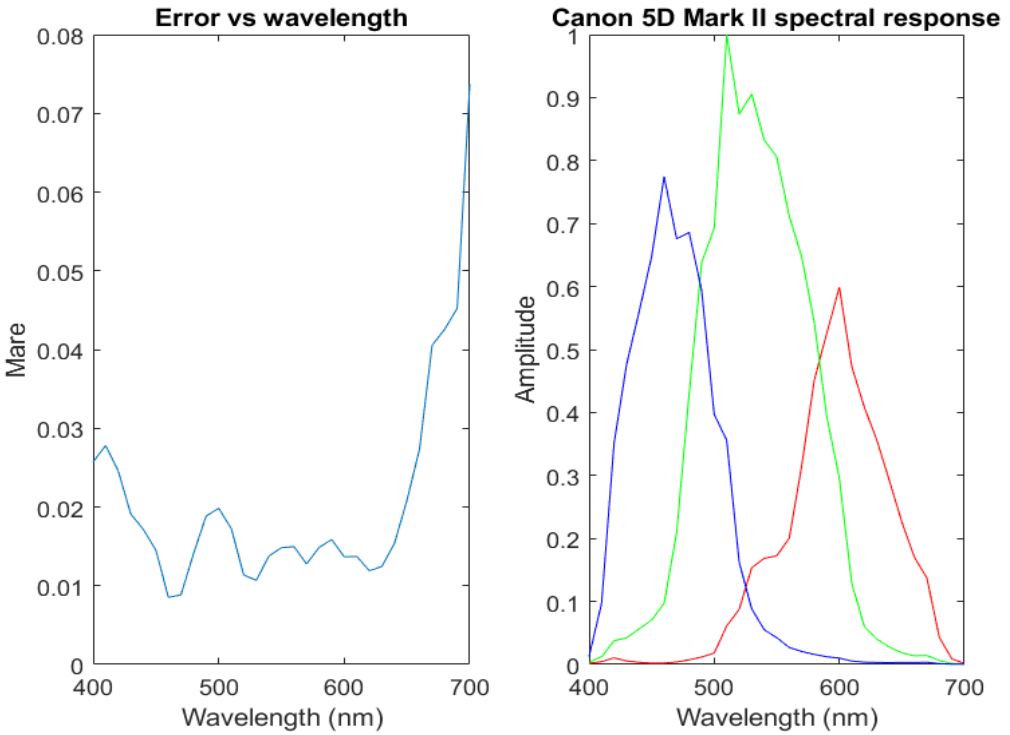
Στην συνέχεια εξετάσαμε της επίδραση της απόκρισης ευαισθησίας της κάμερας στο τελικό σφάλμα ανακατασκευής. Υπολογίσαμε το μέσο απόλυτο σχετικό σφάλμα (MARE) ανά μήκος κύματος και τυπώνουμε τα αποτελέσματα σε αντιπαραβολή με την συνάρτηση ευαισθησίας.



Σχήμα 4.8: Διάγραμμα μέσου απόλυτου σχετικού σφάλματος ανά μήκος κύματος σε αντιπαραβολή με τις CIE 1964 color matching functions.

Στο σχήμα 4.8 παρατηρούμε το μέσο απόλυτο σχετικό σφάλμα (MARE) ανά μήκος κύματος σε σύγκριση με τις Color Matching functions που χρησιμοποιήσαμε για να συνθέσουμε τα δεδομένα. Αμέσως γίνεται εμφανές ότι η επιλογή της απόκρισης ευαισθησίας επηρεάζει πολύ έντονα το τελικό σφάλμα. Το μεγαλύτερο σφάλμα εμφανίζεται για μήκη κύματος στα 650-700 nm, όπου και τα τρία RGB κανάλια σβήνουν. Η πληροφορία δεν αποτυπώνεται στις αντίστοιχες RGB τιμές και δεν γίνεται εφικτό το σύστημα να εκτιμήσει ικανοποιητικά τις αντίστοιχες συνιστώσες. Αντίστοιχα υψηλό σφάλμα εμφανίζεται στην περιοχή των 400 nm, όπου και πάλι τα τρία κανάλια εμφανίζουν μικρές τιμές αλλά και στην περιοχή γύρω από τα 500 nm όπου το Red κανάλι έχει σβήσει ενώ και τα Blue και Green έχουν σχετικά μικρές τιμές. Αντίστοιχα οι περιοχές με ελάχιστο σφάλμα είναι οι περιοχές όπου η απόκριση των RGB καναλιών παίρνει μέγιστες τιμές και υπάρχει ταυτόχρονη συνεισφορά από τουλάχιστον άλλο ένα κανάλι. Είναι εμφανές λοιπόν πως οι γνωστές αποκρίσεις διαφόρων καμερών, που προσεγγίζουν τη γενικότερη συμπεριφορά της ανθρώπινης όρασης, αδιαφορούν για συγκεκριμένες περιοχές του ορατού φάσματος γεγονός που δυσκολεύει το σύστημα στην τελική ανακατασκευή τους. Η πληροφορία αυτών των

περιοχών δεν αποτυπώνεται ικανοποιητικά στις RGB τιμές καθιστώντας έτσι την ανακατασκευή τους δυσκολότερη. Στη συνέχεια εξετάσαμε την συμπεριφορά της μεθόδου μας για την απόκριση ευαισθησίας της κάμερας Canon 5D Mark II.



Σχήμα 4.9: Διάγραμμα μέσου απόλυτου σχετικού σφάλματος ανά μήκος κύματος σε αντιπαραβολή με φασματική απόκριση της κάμερας Canon 5D Mark II.

Από το σχήμα 4.9 παρατηρούμε την ίδια συμπεριφορά στις περιοχές κοντά στα 400 και 700 nm όπου οι τιμές των καναλιών είναι μικρές. Χαρακτηριστική διαφορά εμφανίζεται όμως στην περιοχή γύρω από τα 500 nm όπου η εξασθένιση του Red καναλιού οδηγεί και πάλι σε αύξηση του σφάλματος, λόγω της έντονης παρουσίας όμως του Green καναλιού το μέσο σφάλμα είναι μικρότερο.

4.2.3 Σύγκριση με άλλες σύγχρονες μεθόδους

Για την σύγκριση με άλλες μεθόδους σύγχρονες μεθόδους αρχικά εντοπίσαμε τις καλύτερες εργασίες του τομέα με βάση την [8], η οποία αποτελεί την συστηματικότερη μελέτη των καλύτερων μεθόδων ανακατασκευής πολυφασματικών εικόνων. Χοντρικά ταξινομεί τις μεθόδους σε δύο κατηγορίες. Τις prior based spectral reconstruction μεθόδους που βασίζονται σε αρχιτεκτονικές συζευγμένων λεξικών (Dictionary Coupling), Local Linear Embeddings, Gaussian Processes, Feedforward networks κλπ. Αυτές οι μέθοδοι είναι αρκετά μικρότερες σε μέγεθος και χρόνους ανακατασκευής και αποτελούν ιδανικές επιλογές για εφαρμογές που έχουν περιορισμούς στους υπολογιστικούς πόρους. Και τις data driven deep learning μεθόδους που βασίζονται σε γνωστές αρχιτεκτονικές μεγάλων νευρωνικών δικτύων όπως GANs, Dense networks, Attention networks κλπ. Το πλήθος των συντελεστών ξεπερνά το εκατομμύριο γεγονός που τις καθιστά απαγορευτικές για μεγάλο πλήθος εφαρμογών. Για την σύγκριση, με βάση και την δική μας στόχευση σε μικρού μεγέθους μοντέλα, επιλέξαμε από την κατηγορία των prior based spectral reconstruction μεθόδων το βασικότερο benchmark, την μέθοδο του Arad [9], αλλά την καλύτερη εργασία, την μέθοδο A+ [10]. Ανακατασκευάσαμε λοιπόν και τις 100 άγνωστες εικόνες του test set και υπολογίσαμε τις μετρικές σφάλματος σε αντιστοιχία με το πείραμα της εργασίας [10]. Επίσης, επειδή η μέθοδος του Arad βασίζεται σε ένα ζεύγος λεξικών απευθείας από τον RGB χώρο στον πολυφασματικό, υλοποιήσαμε και εμείς την μέθοδο για να υπολογίσουμε το σφάλμα στο δικό μας split δεδομένων.

Model architecture	MARE	rRMSE	RMSE(8bit)
Arad [9]	0.0507	0.0873	3.24
A+ [10]	0.0382	0.0599	2.00
A+ 3x3	0.0344	0.0584	1.96
Arad (Ours)	0.0851	0.1511	4.23
1st proposed method - RBF	0.0229	0.0391	**1.66**
1st proposed method – MLP	**0.0223**	**0.0373**	1.74
2nd proposed method	0.0452	0.0681	2.37
3rd proposed method – 3x3	0.0234	0.0396	1.84
3rd proposed method – 5x5	0.0239	0.0404	1.87

Πίνακας 4.2: Αποτελέσματα ανακατασκευής για τις διάφορες αρχιτεκτονικές και σύγκριση με κάποιες από τις καλύτερες μεθόδους για το σύνολο των 100 εικόνων αξιολόγησης.

Από τα αποτελέσματα του πίνακα 4.2 αρχικά παρατηρούμε πως η δική μας υλοποίηση της μεθόδου του Arad εμφανίζει χειρότερο αποτέλεσμα από αυτό που αναφέρεται στην [10]. Αυτό οφείλεται στο διαφορετικό split δεδομένων που χρησιμοποιήθηκε. Η δειγματοληψία στην περίπτωση μας πραγματοποιείται με τυχαίο τρόπο ως προς το σύνολο των 100 εικόνων. Αντίθετα η δειγματοληψία στην [10] πραγματοποιήθηκε τυχαία ως προς τα διαφορετικά

υποσύνολα του dataset. Με αυτόν τον τρόπο εξασφαλίζουν μια πιο δίκαιη κατανομή της πληροφορίας μεταξύ training και testing ειδικά για ορισμένα μικρά υποσύνολα εικόνων. Στην δική μας περίπτωση μπορεί ορισμένα υποσύνολα να μην περιγράφονται ικανοποιητικά κατά την εκπαίδευση και να δυσκολεύουν έτσι τις υλοποιήσεις στην τελική ανακατασκευή.

Για τις μεθόδους μας παρατηρούμε ότι η προσέγγιση μας, με τον τεμαχισμό του προβλήματος, λειτουργεί καλύτερα από τον Arad για όλες τις περιπτώσεις. Χαρακτηριστικά, ακόμη και η δεύτερη προτεινόμενη μέθοδος που δεν θεωρεί γνωστό τον μετασχηματισμό έχει σχεδόν το μισό σφάλμα σε σχέση με την δική μας υλοποίηση της μεθόδου . Η εκτίμηση λοιπόν των διανυσμάτων της PCA και η αντιστροφή με χρήση λεξικών δίνει καλύτερο αποτέλεσμα από ένα coupling λεξικών απευθείας από τον RGB χώρο στον πολυφασματικό. Για τις περιπτώσεις που χρησιμοποιούμε τον αντίστροφο μετασχηματισμό, δηλαδή για την πρώτη και την τρίτη προτεινόμενη μέθοδο, παρατηρούμε ότι έχουμε ακόμη καλύτερο αποτέλεσμα και ξεπερνάμε σε απόδοση και την μέθοδο A+. Όλες οι υλοποιήσεις μας έχουν κοντινά αποτελέσματα γεγονός που επιβεβαιώνει την γενικότερη λειτουργία της προσέγγισης. Η χρήση γραμμικού ορθογώνιου μετασχηματισμού που οδηγεί σε ευκολότερη αντιστροφή και η τμηματική ανακατασκευή των φασματικών υπογραφών ξεπερνά τις σημαντικότερες σύγχρονες μεθόδους της κατηγορίας. Το σχετικό σφάλμα μάλιστα σε σχέση με την δικιά μας υλοποίηση της μεθόδου του Arad, που είναι το βασικό benchmark και αποτελεί μέρος των πειραμάτων όλων των εργασιών, είναι 4 φορές μικρότερο. Τέλος παρατηρούμε ότι τα δίκτυα που βασίστηκαν σε patches δεν αντλούν κάτι από τις χωρικές συσχετίσεις και καταλήγουν να έχουν παρόμοιο ή και ελάχιστα χειρότερο αποτέλεσμα την απλή pixel περίπτωση παρά την αύξηση της περιπλοκότητας τους.

4.2.4 Χρονική και χωρική σύγκριση μεθόδων

Στην συνέχεια για τη χρονική και χωρική σύγκριση των μεθόδων υπολογίσαμε το χρόνο ανακατασκευής μια εικόνας αλλά και το πλήθος των παραμέτρων για όλες τις μεθόδους.

Model architecture	Image reconstruction time (sec)	Number of parameters
A+ [10]	4.3 times faster than Arad	98304
Arad [9]	18.720	17408
1st proposed method - RBF	4.847	496
1st proposed method – MLP	**2.529**	3248
2nd proposed method	202.896	80080
3rd proposed method – 3x3	11.830	68960
3rd proposed method – 5x5	5.354	293536

Πίνακας 4.3: Χρόνοι ανακατασκευής μιας εικόνας και πλήθος παραμέτρων για την κάθε αρχιτεκτονική.

Από τον πίνακα 4.3 για την χρονική σύγκριση των μεθόδων υπολογίσαμε τον χρόνο ανακατασκευής μια εικόνας τόσο για τις προτεινόμενες μεθόδους μας όσο και για την μέθοδο του Arad. Στην [10] αναφέρετε η σχετική επιτάχυνση της μεθόδου A+ σε σύγκριση και πάλι με την μέθοδο του Arad. Χρησιμοποιήσαμε αυτά τα αποτελέσματα ως οδηγό για να

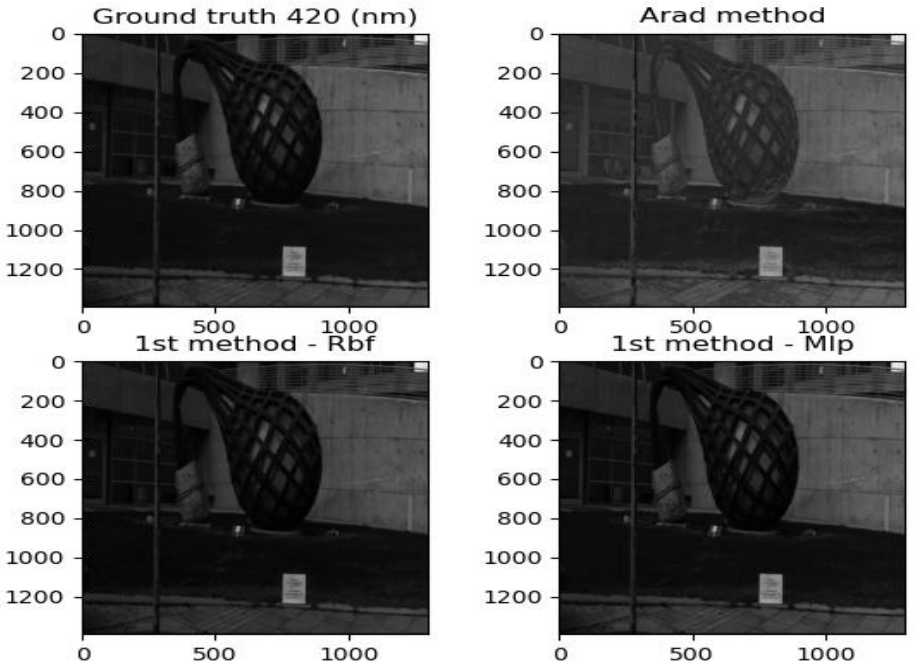
συγκρίνουμε όλες τις μεθόδους μεταξύ τους. Παρατηρούμε λοιπόν ότι η δεύτερη προτεινόμενη μέθοδος εμφανίζει απαγορευτικά μεγάλους χρόνους ανακατασκευής με το ζεύγος των λεξικών να αποτελεί το bottleneck του συστήματος. Η επαναληπτική προβολή των διανυσμάτων της PCA εξαντλητικά σε κάθε ένα από τα 2048 άτομα του λεξικού καθυστερεί πολύ το σύστημα ειδικά σε σχέση με την περίπτωση χρήσης του αντίστροφου μετασχηματισμού. Οι μέθοδοι λοιπόν που βασίζονται σε συζευγμένα λεξικά θα εμφανίζουν το ίδιο πρόβλημα και η αύξηση είτε του βαθμού αραιότητας είτε του μεγέθους των λεξικών, που αποτελεί απαραίτητο βήμα για την καλύτερη ανακατασκευή, εκτοξεύει του χρόνους ανακατασκευής σε σύγκριση με άλλες με μεθόδους. Για την μεθόδους που βασίζονται στον αντίστροφο μετασχηματισμό βλέπουμε πολύ μεγάλη επιτάχυνση όπως περιμέναμε. Χαρακτηριστικό μάλιστα είναι πως στην patch περίπτωση η αύξηση του μεγέθους των παραθύρων μειώνει έντονα τους χρόνους. Θα μπορούσαμε λοιπόν να έχουμε παρόμοιους ή και καλύτερους χρόνους από την pixel περίπτωση εάν τα αποτελέσματα των εν λόγω δικτύων είχαν ικανοποιητικό αποτέλεσμα ώστε να εξετάσουμε περιπτώσεις διάστασης 7x7,9x9 κλπ. Τελικά το καλύτερο αποτέλεσμα το έχει η πρώτη προτεινόμενη μέθοδος που φτάνει να είναι 7.48 γρηγορότερη του Arad και 1.73 φορές γρηγορότερη του A+. συνεπώς η μέθοδος μας ξεπερνά και σε χρόνο τις βασικότερες εργασίες.

Για την χωρική σύγκριση των μεθόδων υπολογίσαμε το πλήθος των συντελεστών της κάθε μεθόδου. Από την πίνακα βλέπουμε ότι όλες οι υλοποιήσεις μας έχουν μικρότερο μέγεθος από τον A+ εκτός από την τελευταία. Εφόσον η πρώτη προτεινόμενη μέθοδος έχει τα καλύτερα αποτελέσματα τόσο σε σφάλμα όσο και σε χρόνο ανακατασκευής θα αναλύσουμε αυτήν την περίπτωση. Το RBF δίκτυο είχε με διαφορά το μικρότερο μέγεθος καταλήγοντας σε 198 φορές μικρότερο μοντέλο. Ενώ το MLP δίκτυο έχει 32 φορές μικρότερο μέγεθος.

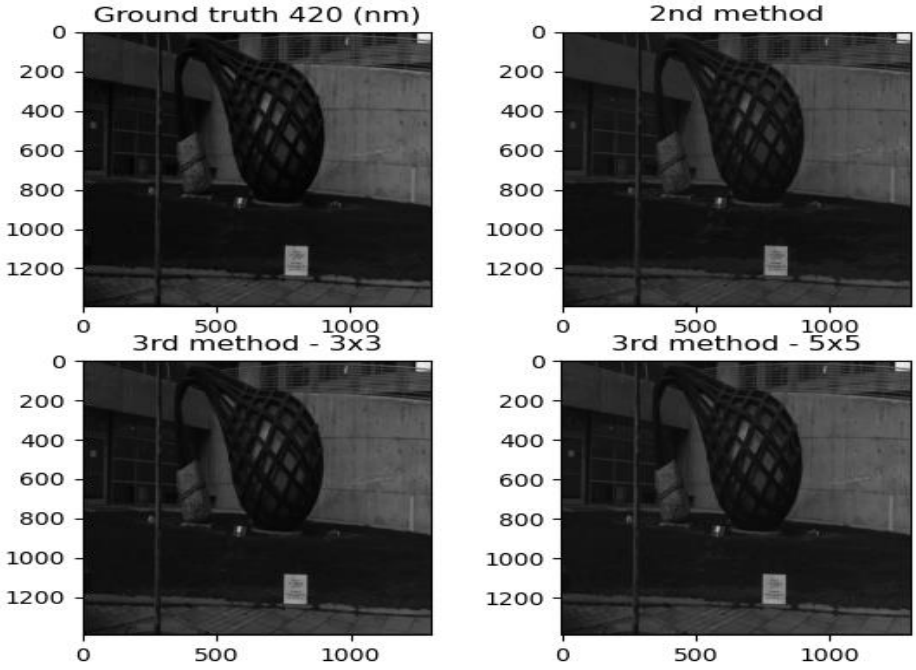
4.2.5 Οπτικά αποτελέσματα ανακατασκευής ολόκληρων εικόνων

Τέλος εξετάσαμε την λειτουργία των μεθόδων στην ανακατασκευή μιας ολόκληρης εικόνας. Για το σκοπό αυτό ανακατασκευάσαμε μια τυχαία εικόνα από το dataset και απεικονίσαμε τα αποτελέσματα αλλά και το αντίστοιχο σφάλμα για τρία διαφορετικά μήκη κύματος. Η εικόνα που επιλέξαμε είναι η bulb_0822-0909 και τα αντίστοιχα μήκη κύματος ήταν στα 420 nm, 550 nm και 650 nm.

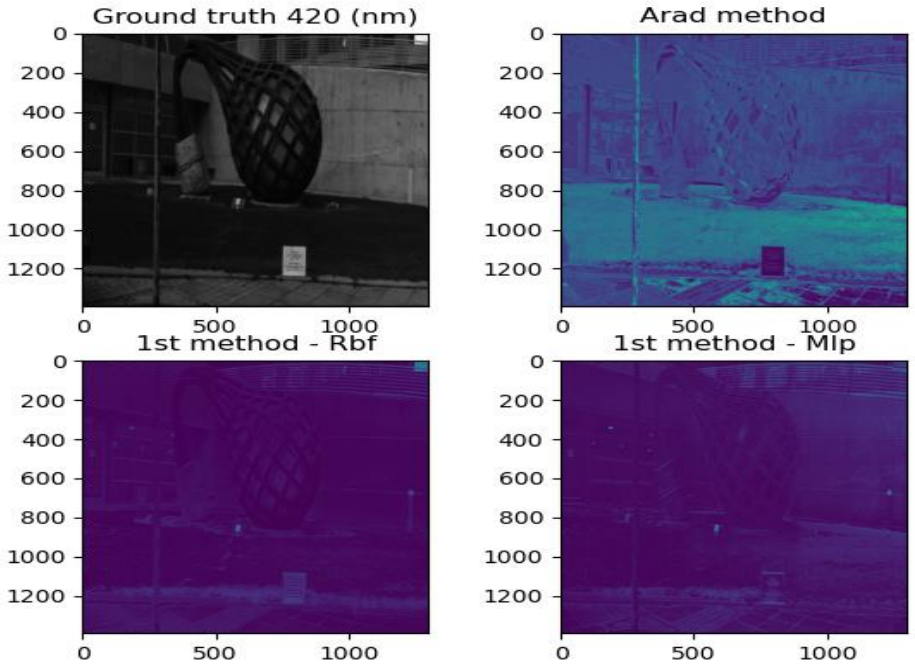
Για το κανάλι των 420 nm έχουμε:



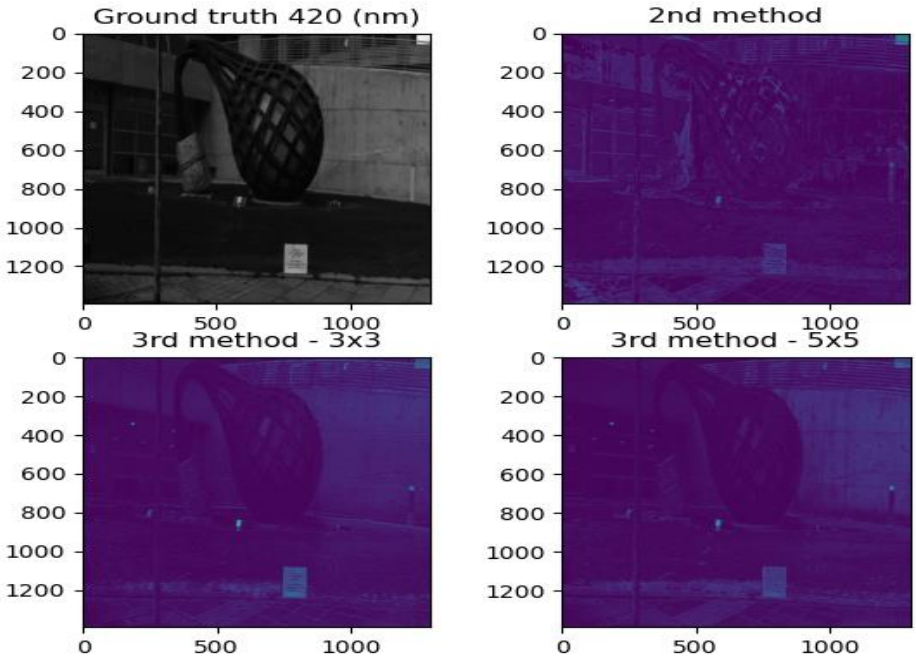
Σχήμα 4.10: Αποτελέσματα εκτίμησης για το κανάλι των 420 nm για την μέθοδο του Arad και για την πρώτη προτεινόμενη μέθοδο.



Σχήμα 4.11: Αποτελέσματα εκτίμησης για το κανάλι των 420 nm για την δεύτερη και τρίτη προτεινόμενη μέθοδο.



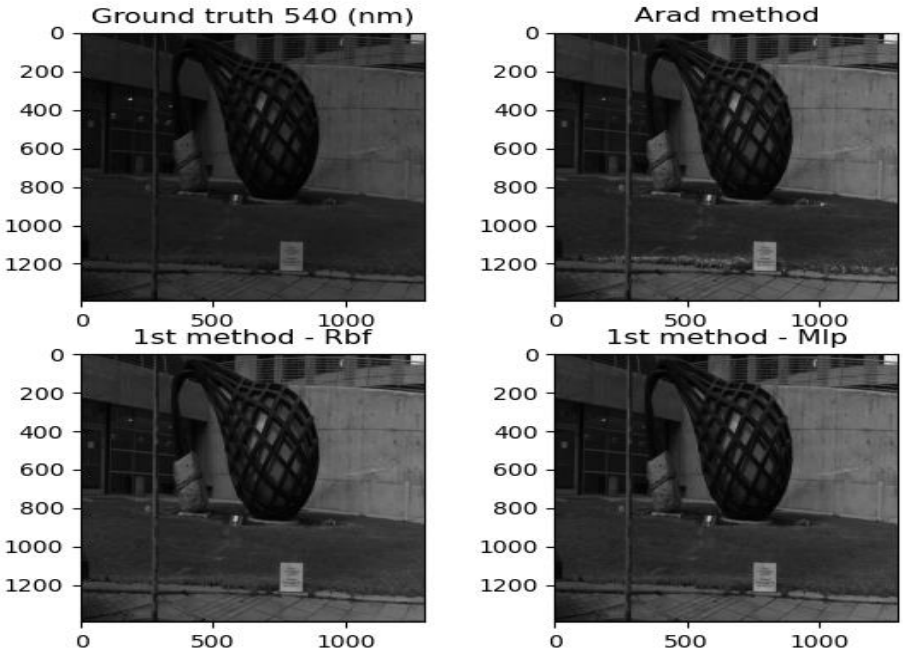
Σχήμα 4.12: Σφάλμα ανακατασκευής για το κανάλι των 420 nm για την μέθοδο του Arad και για την πρώτη προτεινόμενη μέθοδο.



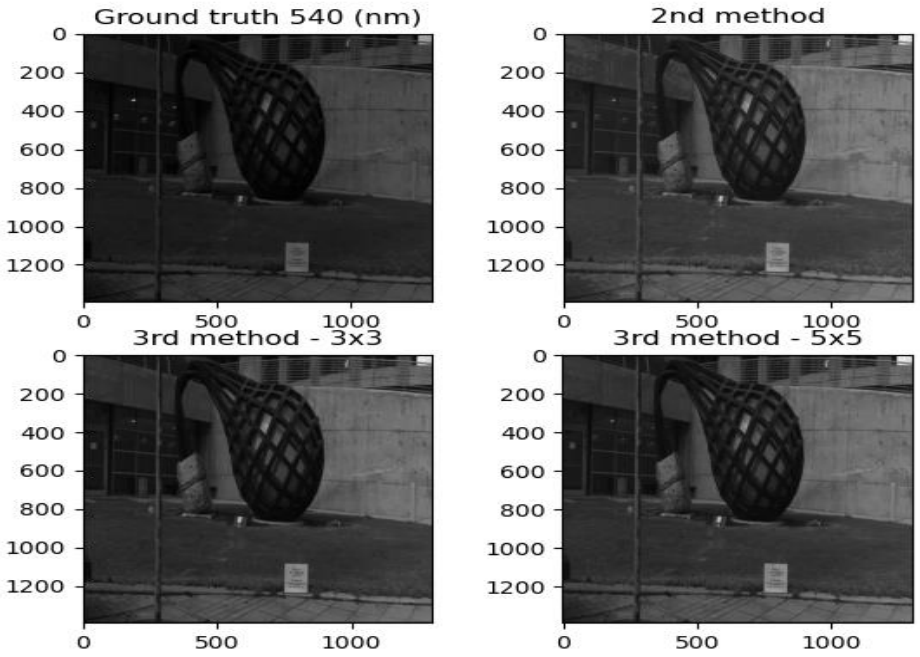
Σχήμα 4.13: Σφάλμα ανακατασκευής για το κανάλι των 420 nm για την δεύτερη και τρίτη προτεινόμενη μέθοδο.

Από τα σχήματα 4.10 και 4.11 για το κανάλι των 420 nm παρατηρούμε πως η μέθοδος του Arad εμφανίζει με διαφορά τα χειρότερα αποτελέσματα με τις διαφορές στη φωτεινότητα να γίνονται απευθείας ορατές. Οι μέθοδοι μας εμφανίζουν όλες καλύτερο αποτέλεσμα χωρίς εμφανής διαφορές στην εικόνα παρά την δυσκολία που εμφανίζει στην ανακατασκευή το καναλιού των 420 nm. Από τα σχήματα 4.12 και 4.13 που περιέχουν τα διαγράμματα σφάλματος παρατηρούμε αρχικά το έντονο σφάλμα σε όλη την εικόνα από την μέθοδο του Arad. Η πρώτη προτεινόμενη μέθοδος εμφανίζει το καλύτερο αποτέλεσμα. Στο σφάλμα εμφανίζεται η δομή της εικόνας συνεπώς υπάρχει ένα υπόλοιπο πληροφορίας που δεν κατόρθωσε να περιγράψει το σύστημα αλλά είναι το μικρότερο από όλες τις υλοποιήσεις ειδικά για την υλοποίηση που βασίστηκε στο RBF δίκτυο. Η δεύτερη προτεινόμενη μέθοδος, που δεν θεωρεί γνωστό τον μετασχηματισμό, έχει το περισσότερο σφάλμα από τις υλοποιήσεις μας αλλά σε καμία περίπτωση συγκρίσιμο με του Arad. Τέλος η τρίτη προτεινόμενη μέθοδος έχει ικανοποιητικό αποτέλεσμα αλλά δεν ξεπερνά τις υλοποιήσεις που βασίστηκαν σε pixel παρά την αύξηση της πολυπλοκότητας και την ελευθερία που δώσαμε στο σύστημα να αξιοποιήσει τις χωρικές συσχετίσεις της εικόνας.

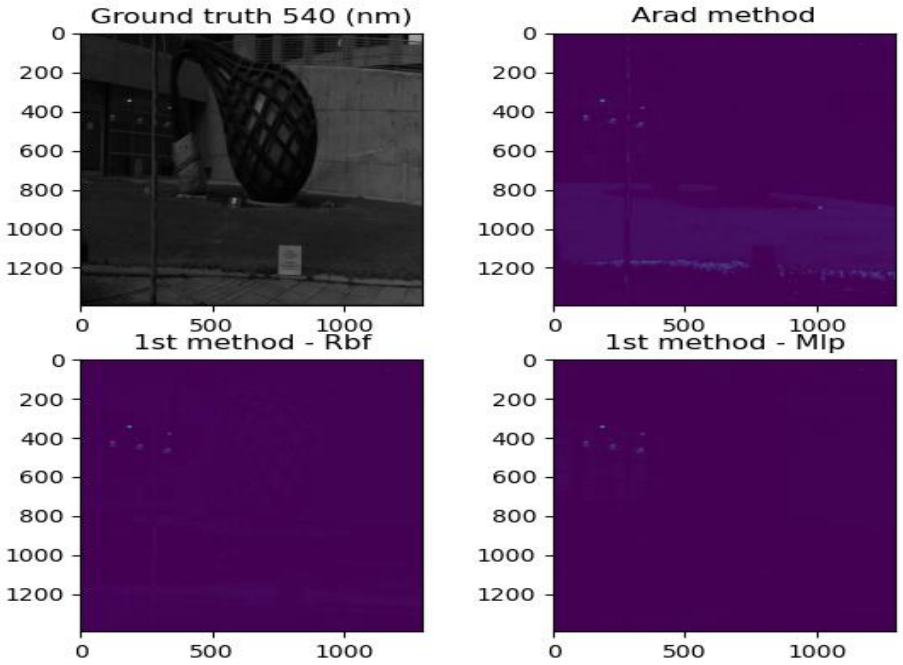
Για το κανάλι των 540 nm έχουμε:



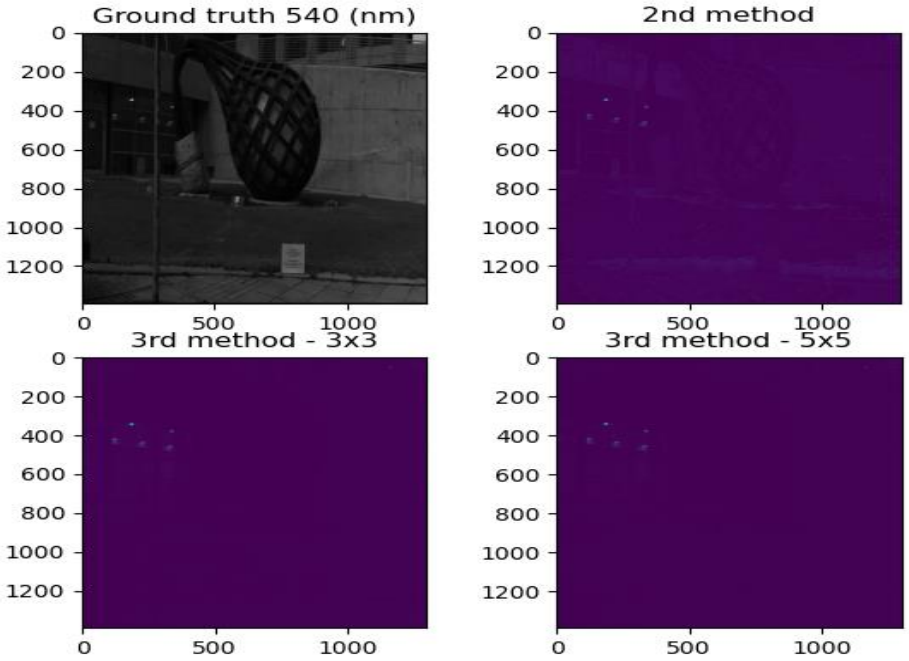
Σχήμα 4.14: Αποτελέσματα εκτίμησης για το κανάλι των 540 nm για την μέθοδο του Arad και για την πρώτη προτεινόμενη μέθοδο.



Σχήμα 4.15: Αποτελέσματα εκτίμησης για το κανάλι των 540 nm για την δεύτερη και τρίτη προτεινόμενη μέθοδο.



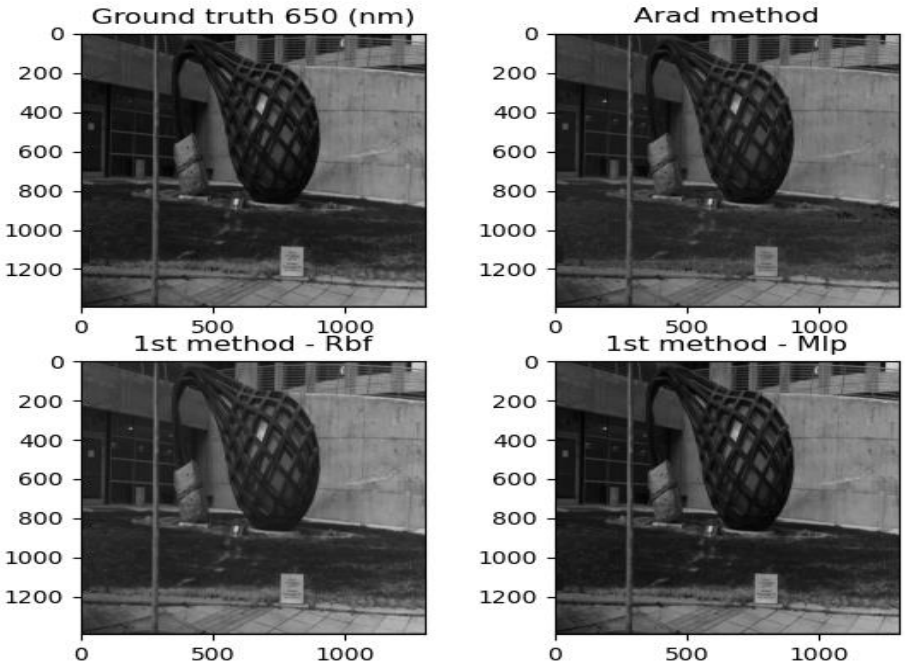
Σχήμα 4.16: Σφάλμα ανακατασκευής για το κανάλι των 540 nm για την μέθοδο του Arad και για την πρώτη προτεινόμενη μέθοδο.



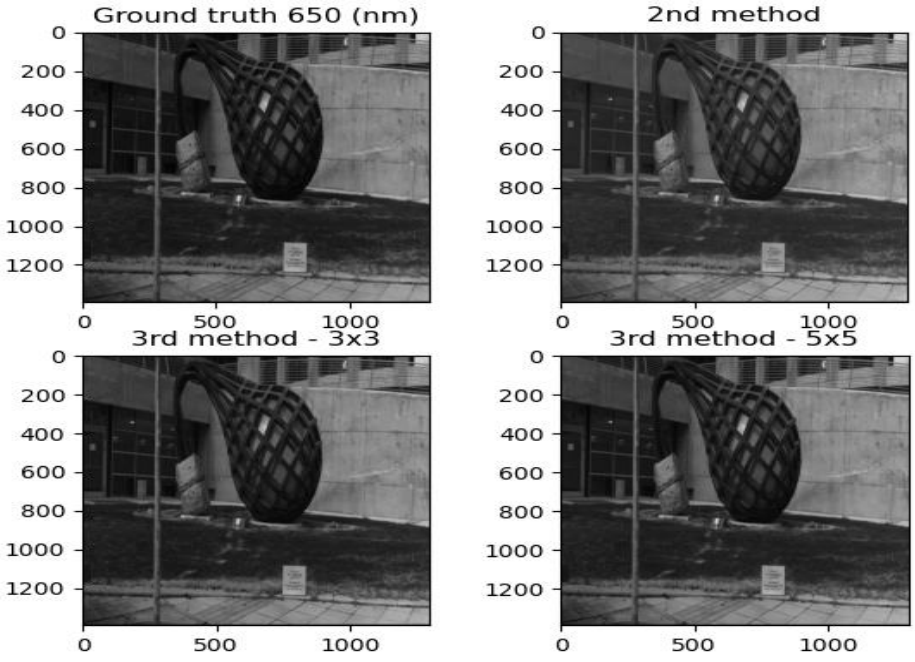
Σχήμα 4.17: Σφάλμα ανακατασκευής για το κανάλι των 540 nm για την δεύτερη και τρίτη προτεινόμενη μέθοδο.

Για την περίπτωση των 540 nm έχουμε ικανοποιητικά αποτελέσματα από όλες τις μεθόδους καθώς αποτελεί ένα από τα εύκολα κανάλια ανακατασκευής αφού η πληροφορία του αποτυπώνεται στις RGB τιμές από την απόκριση της κάμερας. Από τα σχήματα 4.14 και 4.15 δεν γίνονται εμφανής διαφορές. Από τα σχήματα 4.16 και 4.17 που περιέχουν τα διαγράμματα σφάλματος παρατηρούμε ότι η μέθοδος του Arad έχει το χειρότερο αποτέλεσμα αφού σχηματίζεται η δομή της εικόνας και ακολουθεί η δεύτερη προτεινόμενη μέθοδος. Ένα υπόλοιπο πληροφορίας συνεπώς δεν αποτυπώθηκε αλλά είναι αρκετά μικρό αφού δεν αναγνωρίζουμε πλήρως την εικόνα αλλά σχηματίζονται τα όρια των τμημάτων της όπως το γρασίδι και το έδαφος χωρίς ιδιαίτερες λεπτομέρειες. Για την πρώτη και τρίτη προτεινόμενη μέθοδο έχουμε το καλύτερο αποτέλεσμα χωρίς να μπορούμε από το υπόλοιπο του σφάλματος να εντοπίσουμε τη δομή της εικόνας.

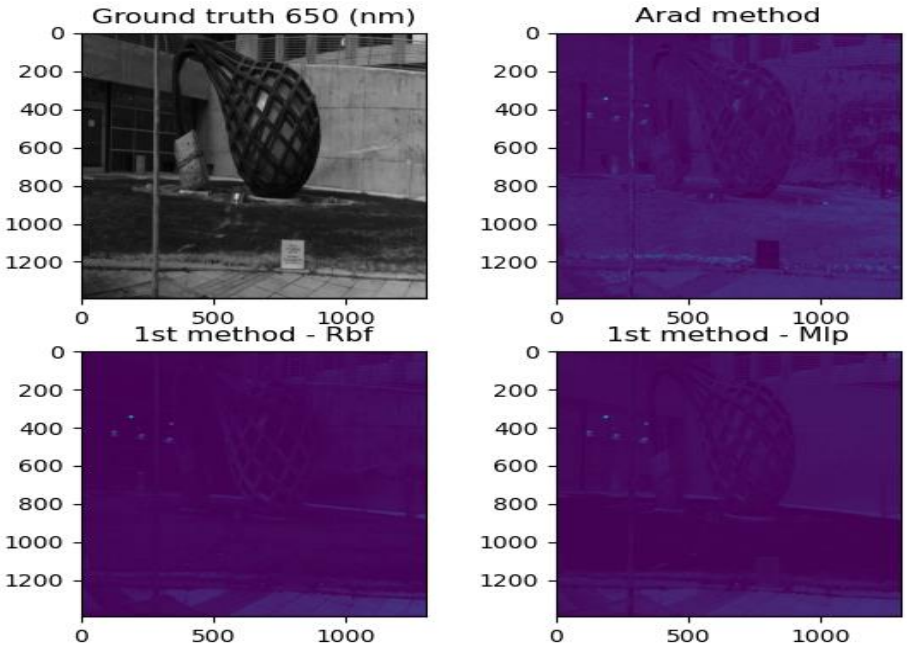
Για το κανάλι των 650 nm έχουμε:



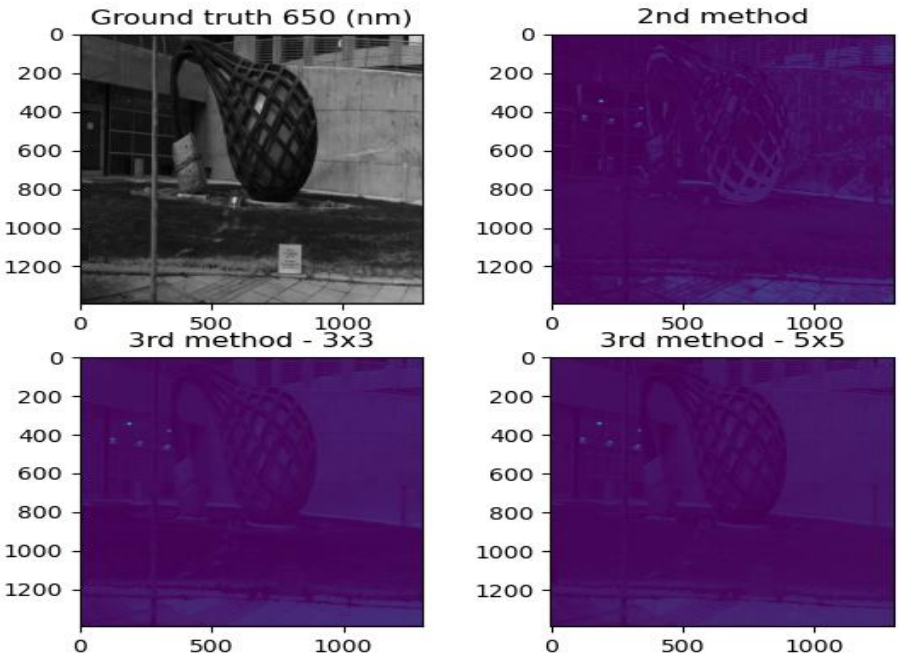
Σχήμα 4.18: Αποτελέσματα εκτίμησης για το κανάλι των 650 nm για την μέθοδο του Arad και για την πρώτη προτεινόμενη μέθοδο.



Σχήμα 4.19: Αποτελέσματα εκτίμησης για το κανάλι των 650 nm για την δεύτερη και τρίτη προτεινόμενη μέθοδο.



Σχήμα 4.20: Σφάλμα ανακατασκευής για το κανάλι των 650 nm για την μέθοδο του Arad και για την πρώτη προτεινόμενη μέθοδο.



Σχήμα 4.21: Σφάλμα ανακατασκευής για το κανάλι των 650 nm για την δεύτερη και τρίτη προτεινόμενη μέθοδο.

Τέλος για την περίπτωση τω 650 nm που αποτελεί και το δυσκολότερο κανάλι από τα σχήματα 4.18 και 4.19 δεν εντοπίζουμε εύκολα διαφορές. Από τα σχήματα 4.20 και 4.21 που περιέχουν τα διαγράμματα σφάλματος παρατηρούμε ότι και πάλι η μέθοδος του Arad έχει τα χειρότερα αποτελέσματα και ακολουθεί η δεύτερη προτεινόμενη μέθοδος. Η πρώτη προτεινόμενη μέθοδος έχει το καλύτερο αποτέλεσμα, ένα υπόλοιπο πληροφορίας εμφανίζεται στα σφάλματα αλλά ειδικά για την περίπτωση των RBF δικτύων οριακά εντοπίζουμε πλήρως την δομή της. Για την τρίτη προτεινόμενη μέθοδο παρατηρούμε σχετικά ικανοποιητικά αποτελέσματα αλλά χειρότερα και πάλι από την pixel περίπτωση με τη δομή της εικόνας να φαίνεται ξεκάθαρα στα διαγράμματα σφάλματος.

Τελικά η προσέγγιση μας με τον τεμαχισμό του προβλήματος δίνει ιδιαίτερα ικανοποιητικά αποτελέσματα και μικρά και γρήγορα συστήματα. Ακόμη και η $2^{η}$ προτεινόμενη μέθοδος δίνει αξιόλογο αποτέλεσμα καλύτερο σε όλες τις περιπτώσεις από το βασικότερο benchmark. Οι μέθοδοι που βασίζονται στην αντίστροφη PCA εμφανίζουν τα καλύτερα αποτελέσματα και οπτικά μικρές διαφορές από την επιθυμητή εικόνα ακόμη και για τα κανάλια που η πληροφορία δεν αποτυπώνεται στις αντίστοιχες RGB τιμές. Αλλά μεταξύ των υλοποιήσεων της πρώτης και τρίτης προτεινόμενης μεθόδου τα συστήματα που βασίστηκαν σε patches όχι απλά δεν ξεπερνούν αυτά της πρώτης αλλά εμφανίζουν και λίγο χειρότερα αποτελέσματα.

ΚΕΦΑΛΑΙΟ 5

Συμπεράσματα

Το πρόβλημα της ανακατασκευής πολυφασματικών (multispectral imaging) εικόνων αποτελεί έναν σχετικά καινούργιο τομέα στον οποίο στρέφονται όλο και περισσότεροι ερευνητές τα τελευταία χρόνια. Η επέκταση των πεδίων εφαρμογής της πολυφασματικής ανάλυσης αποτελεί αναγκαία οδό σε πολλούς τομείς τόσο της επιστήμης όσο και της τεχνολογίας με αποτέλεσμα τεράστιο πλήθος εφαρμογών και μεθόδων να αναδύονται τόσο σε θεωρητικό όσο και πρακτικό επίπεδο.

Οι βασικότερες μέθοδοι προσεγγίσουν το πρόβλημα χρησιμοποιώντας αλγορίθμους μάθησης, αξιοποιώντας τις στατιστικές ιδιότητες των όλο και μεγαλύτερων dataset που γίνονται προσβάσιμα και αποτυπώνοντας τες κυρίως σε αρχιτεκτονικές νευρωνικών δικτύων και συζευγμένων λεξικών.

Η προσέγγιση μας που βασίζεται στον τεμαχισμό του προβλήματος με χρήση μεθόδου μείωσης διάστασης κατόρθωσε να ξεπεράσει τα καλύτερα αποτελέσματα των prior based spectral reconstruction τεχνικών δημιουργώντας ταυτόχρονα ένα μικρό μοντέλο με πολύ μικρούς χρόνους ανακατασκευής. Αυτό το καθιστά ιδανικό για εφαρμογές με έντονο περιορισμό σε μνήμη, επεξεργαστική ισχύ και κατανάλωση ενέργειας. Μια πιθανή κατεύθυνση βελτίωσης είναι να εξετάσουμε άλλες μεθόδους

μείωσης διάστασης που όμως παρέχουν και μια μέθοδο ανακατασκευής των συμπιεσμένων περιγραφών. Η εφαρμογή της PCA είχε ακριβώς αυτό το πλεονέκτημα κάτι που δεν εμφανίζεται απαραίτητα στις σύνθετες μη γραμμικές μεθόδους μείωσης διάστασης.

Βασικό χαρακτηριστικό επίσης των prior based spectral reconstruction methods είναι πως δεν κατορθώνουν να εκμεταλλευτούν τις χωρικές συσχετίσεις των εικόνων και να αντλήσουν πληροφορία από την γειτονιά των pixel με ικανοποιητικό τρόπο. Τα δίκτυα MLP που αναπτύξαμε για την περίπτωση των μη επικαλυπτόμενων παραθύρων δεν κατορθώνουν να ξεπεράσουν την απλή pixel περίπτωση παρά την αύξηση της περιπλοκότητας τους. Μια πιθανή κατεύθυνση είναι να επεκτείνουμε το σύνολο των δεδομένων εκπαίδευσης, λόγω του μεγάλου αριθμού παραμέτρων που έχουν και του συνόλου των δεδομένων που απαιτούν, κάτι που δεν ήταν εύκολο να γίνει στο περιβάλλον google colab λόγω του περιορισμού στην μνήμη που μας παρέχεται. Παράλληλα μπορούμε εξετάσουμε μια προσέγγιση βασισμένη σε συνελικτικά δίκτυα (CNN) που θα εντοπίζουν τις χωρικές συσχετίσεις και θα εκτιμούν ίσως ακόμη καλύτερα τα embeddings της PCA.

Τέλος βασικός περιορισμός σε όλες τις μεθόδους ήταν η απόκριση της κάμερας, η οποία δεν αποτυπώνει συγκεκριμένες περιοχές του φάσματος με αποτέλεσμα να μην γίνεται ικανοποιητικά η ανακατασκευή τους. Μια ακόμη κατεύθυνση είναι λοιπόν να σχεδιάσουμε τις επιθυμητές αποκρίσεις, που διευκολύνουν το πρόβλημα της εκτίμηση της πολυφασματικής πληροφορίας από τις τρείς μόλις RGB τιμές, και στο βαθμό που μπορούν να παραχθούν μαζικά και με μικρό κόστος να βελτιώσουμε κατά πολύ τα τελικά αποτελέσματα όλων των μεθόδων.

Κατάλογος σχημάτων

2.1 Φάσμα ηλεκτρομαγνητικής ακτινοβολίας……..5

2.2 Πυκνότητα εκπεμπόμενης ισχύος του ηλίου ως προς το μήκος κύματος…..6

2.3 Σχετική ευαισθησία ματιού στο φάσμα του ορατού φωτός……..……...7

2.4 Υπερφασματική απεικόνιση και ταξινόμηση περιοχών βλάστησης, νερού και εδάφους……...8

2.5 (a) Δομή χοληφόρου ιστού.(b) Ταξινόμηση διαφορετικών περιοχών με βάση την φασματική τους υπογραφή……..…8

2.6 Στο σχήμα φαίνεται η συμπεριφορά της συνάρτησης $\|x\|^p$ για διαφορετικές τιμές του p. Καθώς $p \to 0$ η $\|x\|^p$ παίρνει της μορφή μιας συνάρτησης δείκτη η οποία για $x = 0$ δίνει 0 και 1 αλλού……...12

2.7 Orthogonal Matching Pursuit - ένας άπληστος αλγόριθμος για την προσέγγιση της λύσης του προβλήματος P_0 ...……14

2.8 Ο αλγόριθμος K-SVD……..17

2.9 Σχηματική αναπαράσταση των δύο κυρίαρχων συνιστωσών(Principal components) σε ένα σύνολο δυδιάστατων δεδομένων……21

2.10 Αλληλεπίδραση πράκτορα με το περιβάλλον του……...22

2.11 Σχηματική αναπαράσταση της αλληλεπίδρασης της ενισχυτικής μάθησης με το περιβάλλον……..23

2.12 Παραδείγματα ζευγών $(x, f(x))$ και υποθέσεις διαφορετικών βαθμών πολυωνύμου (a) 1ου βαθμού, (b) 7ου βαθμού, (c) ένα διαφορετικό dataset και μια γραμμική εκτίμηση, (d) ημιτονοειδής εκτίμηση……..…….24

2.13 Σχηματική αναπαράσταση της επιβλεπόμενης μάθησης…….25

2.14 Νευρώνας και μυελιωμένος άξονας, με τη ροή του σήματος από τις εισόδους στους δενδρίτες στις εξόδους στον άξονα…….26

2.15 Μαθηματική δομή τεχνητού νευρώνα…….27

2.16 Σχηματική δομή ενός Multilayer Perceptron……29

2.17 Γραφική αναπαράσταση του αλγορίθμου Gradient Descent……30

2.18 Σχηματική δομή ενός Radial basis neural network……...34

3.1 Διάγραμμα εναπομένουσας διασποράς (residual variance) των αρχικών πολυφασματικών δεδομένων……...38

3.2 Διάγραμμα εναπομένουσας διασποράς (residual variance) μετά την εφαρμογή της PCA. Η ενέργεια συγκεντρώνεται στους πρώτους σημαντικότερους συντελεστές……39

3.3 Αθροιστικό διάγραμμα κατανομής ενέργειας……...39

3.4 Σχηματική αναπαράσταση των απεικονίσεων μεταξύ των χώρων. (a) Ο RGB χώρος προκύπτει από την προβολή των πολυφασματικών υπογραφών μέσω της απόκρισης της κάμερας. (b) Απευθείας ανακατασκευή των πολυφασματικών υπογραφών μέσω μια συνάρτησης απεικόνισης F(Rgb) = Hs. (c) Εύρεση ενδιάμεσης αναπαράστασης με χρήση της PCA. (d) Τμηματική ανακατασκευή μέσω των συναρτήσεων απεικόνισης $F_1(Rgb) = Pc, F_2(Pc) = Hs$……...41

3.5 Χαρακτηριστικά παραδείγματα από το dataset πολυφασματικών εικόνων ICVL(Interdisciplinary Computer Vision Laboratory)……42

3.6 Φασματική απόκριση του πρότυπου CIE 1964 color matching functions.

Κάθε συνιστώσα απεικονίζεται με το αντίστοιχο χρώμα Red,Green,Blue.43

3.7 Μπλόκ διάγραμμα εξαγωγής δεδομένων. Απεικονίζονται αναλυτικά οι μετασχηματισμοί επί των πολυφασματικών δεδομένων καθώς και οι κανονικοποιήσεις πριν από κάθε βήμα. Καταλήγουμε με ένα σύνολο δεδομένων για κάθε χώρο αναπαράστασης Rgb,Pca,Hyperspectral……45

3.8 Σχηματική δομή του Radial Basis Function neural network. Στην είσοδο δίνεται ένα RGB pixel και εκτιμάται το περιεχόμενο των 8 κυρίαρχων συνιστωσών ……...47

3.9 Σχηματική δομή του Multilayer Perceptron. Στην είσοδο δίνεται ένα RGB pixel και εκτιμάται το περιεχόμενο των 8 κυρίαρχων συνιστωσών…….49

3.10 Οι 8 κυρίαρχες συνιστώσες των δεδομένων……51

3.11 Μπλόκ διάγραμμα πρώτης προτεινόμενης μεθόδου επίλυσης……52

3.12 Μπλόκ διάγραμμα δεύτερης προτεινόμενης μεθόδου επίλυση……55

3.13 Διάγραμμα εναπομένουσας διασποράς(residual variance) μετά την εφαρμογή της PCA. Η ενέργεια συγκεντρώνεται και πάλι στους πρώτους σημαντικότερους συντελεστές……56

3.14 Σχηματική δομή του Multilayer Perceptron για την patch περίπτωση. Στην είσοδο δίνονται οι RGB τιμές ενός patch διάστασης 27 ή 75 και εκτιμάται το περιεχόμενο των 60 ή 100 κυρίαρχων συνιστωσών……..…….58

3.15 Μπλόκ διάγραμμα τρίτης προτεινόμενης μεθόδου επίλυσης…….59

4.1 Διάγραμμα σύγκλισης αλγορίθμου Levenberg-Marquardt για την εκπαίδευση του Radial Basis neural network. Στο διάγραμμα περιλαμβάνεται ο λογάριθμος του σφάλματος ως προς τις εποχές εκπαίδευσης τόσο για τα παραδείγματα εκπαίδευσης όσο και για τα δεδομένα του validation set…….61

4.2 Διάγραμμα σύγκλισης του αλγορίθμου Adam για την εκπαίδευση του Multilayer Perceptron δικτύου για την περίπτωση που εφαρμόζεται σε pixel. Στο διάγραμμα περιλαμβάνεται ο λογάριθμος του σφάλματος ως προς τις εποχές εκπαίδευσης τόσο για τα παραδείγματα εκπαίδευσης όσο και για τα δεδομένα του validation set……....62

4.3 Διάγραμμα σύγκλισης του αλγορίθμου Adam για την εκπαίδευση του

Multilayer Perceptron δικτύου για την περίπτωση που εφαρμόζεται σε patches διάστασης 3x3. Στο διάγραμμα περιλαμβάνεται ο λογάριθμος του σφάλματος ως προς τις εποχές εκπαίδευσης τόσο για τα παραδείγματα εκπαίδευσης όσο και για τα δεδομένα του validation set……....63

4.4 Διάγραμμα σύγκλισης του αλγορίθμου Adam για την εκπαίδευση του Multilayer Perceptron δικτύου για την περίπτωση που εφαρμόζεται σε patches διάστασης 3x3. Στο διάγραμμα περιλαμβάνεται ο λογάριθμος του σφάλματος ως προς τις εποχές εκπαίδευσης τόσο για τα παραδείγματα εκπαίδευσης όσο και για τα δεδομένα του validation set……...64

4.5 Διάγραμμα σφάλματος ανακατασκευής για διαφορετικά μεγέθη λεξικού και διαφορετικούς βαθμούς αραιότητας……..……...65

4.6 Διάγραμμα σφάλματος για την απεικόνιση από τον RGB χώρο στον χώρο της PCA ως συνάρτηση του αριθμού των εκτιμώμενων κυρίαρχων συνιστωσών (Principal Components)……67

4.7 Διάγραμμα σφάλματος για την απεικόνιση από τον RGB χώρο στον πολυφασματικό ως συνάρτηση του αριθμού των εκτιμώμενων κυρίαρχων συνιστωσών (Principal Components)……...68

4.8 Διάγραμμα μέσου απόλυτου σχετικού σφάλματος ανά μήκος κύματος σε αντιπαραβολή με τις CIE 1964 color matching functions……69

4.9 Διάγραμμα μέσου απόλυτου σχετικού σφάλματος ανά μήκος κύματος σε αντιπαραβολή με φασματική απόκριση της κάμερας Canon 5D Mark II……..70

4.10 Αποτελέσματα εκτίμησης για το κανάλι των 420 nm για την μέθοδο του Arad και για την πρώτη προτεινόμενη μέθοδο……74

4.11 Αποτελέσματα εκτίμησης για το κανάλι των 420 nm για την δεύτερη και τρίτη προτεινόμενη μέθοδο……74

4.12 Σφάλμα ανακατασκευής για το κανάλι των 420 nm για την μέθοδο του Arad και για την πρώτη προτεινόμενη μέθοδο……75

4.13 Σφάλμα ανακατασκευής για το κανάλι των 420 nm για την δεύτερη και τρίτη προτεινόμενη μέθοδο……75

4.14 Αποτελέσματα εκτίμησης για το κανάλι των 540 nm για την μέθοδο του Arad και για την πρώτη προτεινόμενη μέθοδο……...76

4.15 Αποτελέσματα εκτίμησης για το κανάλι των 540 nm για την δεύτερη και τρίτη προτεινόμενη μέθοδο……77

4.16 Σφάλμα ανακατασκευής για το κανάλι των 540 nm για την μέθοδο του Arad και για την πρώτη προτεινόμενη μέθοδο……77

4.17 Σφάλμα ανακατασκευής για το κανάλι των 540 nm για την δεύτερη και τρίτη προτεινόμενη μέθοδο……..78

4.18 Αποτελέσματα εκτίμησης για το κανάλι των 650 nm για την μέθοδο του Arad και για την πρώτη προτεινόμενη μέθοδο……79
4.19 Αποτελέσματα εκτίμησης για το κανάλι των 650 nm για την δεύτερη και τρίτη προτεινόμενη μέθοδο……....79

4.20 Σφάλμα ανακατασκευής για το κανάλι των 650 nm για την μέθοδο του Arad και για την πρώτη προτεινόμενη μέθοδο……..……80

4.21 Σφάλμα ανακατασκευής για το κανάλι των 650 nm για την δεύτερη και τρίτη προτεινόμενη μέθοδο……..80

Κατάλογος πινάκων

4.1 Σύγκριση εποχών και χρόνων εκπαίδευσης της εκάστοτε αρχιτεκτονικής.66
4.2 Αποτελέσματα ανακατασκευής για τις διάφορες αρχιτεκτονικές και σύγκριση με κάποιες από τις καλύτερες μεθόδους για το σύνολο των 100 εικόνων αξιολόγησης……...71
4.3 Χρόνοι ανακατασκευής μιας εικόνας και πλήθος παραμέτρων για την κάθε αρχιτεκτονική……72

Βιβλιογραφία

[1] Khan, M. J., Khan, H. S., Yousaf, A., Khurshid, K., & Abbas, A. (2018). Modern trends in hyperspectral image analysis: A review. *Ieee Access*, *6*, 14118-14129.

[2] Govender, M., Chetty, K., & Bulcock, H. (2007). A review of hyperspectral remote sensing and its application in vegetation and water resource studies. *Water Sa*, *33*(2), 145-151.

[3] Liang, H. (2012). Advances in multispectral and hyperspectral imaging for archaeology and art conservation. *Applied Physics A*, *106*, 309-323.

[4] Afromowitz, M. A., Callis, J. B., Heimbach, D. M., DeSoto, L. A., & Norton, M. K. (1988). Multispectral imaging of burn wounds: a new clinical instrument for evaluating burn depth. *IEEE Transactions on Biomedical Engineering*, *35*(10), 842-850.

[5] Padoan, R., Steemers, T. A., Klein, M., Aalderink, B., & De Bruin, G. (2008). Quantitative hyperspectral imaging of historical documents: technique and applications. *Art Proceedings*, 25-30.

[6] Feng, Y. Z., & Sun, D. W. (2012). Application of hyperspectral imaging in food safety inspection and control: a review. *Critical reviews in food science and nutrition*, *52*(11), 1039-1058.

[7] Clark, R. N., & Swayze, G. A. (1995, January). Mapping minerals, amorphous materials, environmental materials, vegetation, water, ice and snow, and other materials: the USGS Tricorder algorithm. In *JPL, Summaries of the Fifth Annual JPL Airborne Earth Science*

Workshop. Volume 1: AVIRIS Workshop.

[8] Zhang, J., Su, R., Fu, Q., Ren, W., Heide, F., & Nie, Y. (2022). A survey on computational spectral reconstruction methods from RGB to hyperspectral imaging. *Scientific reports*, *12*(1), 11905.

[9] Arad, B., & Ben-Shahar, O. (2016). Sparse recovery of hyperspectral signal from natural RGB images. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part VII 14* (pp. 19-34). Springer International Publishing.

[10] Aeschbacher, J., Wu, J., & Timofte, R. (2017). In defense of shallow learned spectral reconstruction from RGB images. In *Proceedings of the IEEE International Conference on Computer Vision Workshops* (pp. 471-479).

[11] Jia, Y., Zheng, Y., Gu, L., Subpa-Asa, A., Lam, A., Sato, Y., & Sato, I. (2017). From RGB to spectrum for natural scenes via manifold-based mapping. In *Proceedings of the IEEE international conference on computer vision* (pp. 4705-4713).

[12] Li, J., Wu, C., Song, R., Li, Y., & Liu, F. (2020). Adaptive weighted attention network with camera spectral sensitivity prior for spectral reconstruction from RGB images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (pp. 462-463).

[13] Han, X. H., Shi, B., & Zheng, Y. (2018, August). Residual HSRCNN: Residual hyper-spectral reconstruction CNN from an RGB image. In *2018 24th International Conference on Pattern Recognition (ICPR)* (pp. 2664-2669). IEEE.

[14] Li, J., Wu, C., Song, R., Li, Y., & Liu, F. (2020). Adaptive weighted attention network with camera spectral sensitivity prior for spectral reconstruction from RGB images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (pp. 462-463).

[15] Van der Maaten, L., Postma, E. O., & van den Herik, H. J. (2007). Matlab toolbox for dimensionality reduction. *MICC, Maastricht University*.

[16] Rubinstein, R., Zibulevsky, M., & Elad, M. (2008). *Efficient implementation of the K-SVD algorithm using batch orthogonal matching pursuit* (No. CS Technion report CS-2008-08). Computer Science Department, Technion.

[17] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... & Chintala, S. (2019). Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems, 32*.

[18] Land, E. H., & McCann, J. J. (1971). Lightness and retinex theory. *Josa*, *61*(1), 1-11.

[19] Wright, J., Ma, Y., Mairal, J., Sapiro, G., Huang, T. S., & Yan, S. (2010). Sparse representation for computer vision and pattern recognition. *Proceedings of the IEEE*, *98*(6), 1031-1044.

[20] Rubinstein, R., Bruckstein, A. M., & Elad, M. (2010). Dictionaries for sparse representation modeling. *Proceedings of the IEEE*, *98*(6), 1045-1057.

[21] Davis, G., Mallat, S., & Avellaneda, M. (1997). Adaptive greedy approximations. *Constructive approximation*, *13*, 57-98.

[22] Donoho, D. L., & Elad, M. (2003). Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ1 minimization. *Proceedings of the National Academy of Sciences*, *100*(5), 2197-2202.

[23] Chen, S., Billings, S. A., & Luo, W. (1989). Orthogonal least squares methods and their application to non-linear system identification. *International Journal of control*, *50*(5), 1873-1896.

[24] Tropp, J. A. (2004). Greed is good: Algorithmic results for sparse approximation. *IEEE Transactions on Information theory*, *50*(10), 2231-2242.

[25] Mairal, J., Ponce, J., Sapiro, G., Zisserman, A., & Bach, F. (2008). Supervised dictionary learning. *Advances in neural information processing systems*, *21*.

[26] Aharon, M., Elad, M., & Bruckstein, A. (2006). K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on signal processing*, *54*(11), 4311-4322.

[27] Fukunaga, K. (2013). *Introduction to statistical pattern recognition*. Elsevier.

[28] Jimenez, L. O., & Landgrebe, D. A. (1998). Supervised classification in high-dimensional space: geometrical, statistical, and asymptotical properties of multivariate data. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *28*(1), 39-54.

[29] Van Der Maaten, L., Postma, E., & Van den Herik, J. (2009). Dimensionality reduction: a comparative. *J Mach Learn Res*, *10*(66-71).

[30] Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical transactions of the royal society A: Mathematical, Physical and Engineering Sciences*, *374*(2065), 20150202.

[31] Russell, P. N. (2010). Artificial intelligence: a modern approach by stuart. *Russell and Peter Norvig contributing writers, Ernest Davis...[et al.]*.

[32] Barlow, H. B. (1989). Unsupervised learning. *Neural computation*, *1*(3), 295-311.

[33] Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement learning: A survey. *Journal of artificial intelligence research*, *4*, 237-285.

[34] Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (Vol. 4, No. 4, p. 738). New York: springer.

[35] Hardesty, L. (2017). Explained: neural networks. *MIT News*, *14*.

[36] Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, *79*(8), 2554-2558.

[37] Angra, S., & Ahuja, S. (2017, March). Machine learning and its applications: A review. In *2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC)* (pp. 57-60). IEEE.

[38] Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, *2*(4), 303-314.

[39] Gardner, M. W., & Dorling, S. R. (1998). Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric environment*, *32*(14-15), 2627-2636.

[40] Amari, S. I. (1993). Backpropagation and stochastic gradient descent method. *Neurocomputing*, *5*(4-5), 185-196.

[41] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

[42] Ghosh, J., & Nag, A. (2001). An overview of radial basis function networks. *Radial basis*

function networks 2: new advances in design, 1-36.

[43] Yu, H., & Wilamowski, B. M. (2018). Levenberg–marquardt training. In *Intelligent systems* (pp. 12-1). CRC Press.

[44] Maloney, L. T. (1986). Evaluation of linear models of surface spectral reflectance with small numbers of parameters. *JOSA A*, *3*(10), 1673-1683.

[45] Jiang, J., Liu, D., Gu, J., & Süsstrunk, S. (2013, January). What is the space of spectral sensitivity functions for digital color cameras?. In *2013 IEEE Workshop on Applications of Computer Vision (WACV)* (pp. 168-179). IEEE.