

Υπερανάλυση Εικόνας με τη Μέθοδο Deep Unfolding και Χρήση Εικόνων Πολλαπλών Τύπων

Αλέξανδρος Καραμπάσης

Μεταπτυχιακή Εργασία Εξειδίκευσης



Ιωάννινα, Απρίλιος 2022



ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ Η/Υ & ΠΛΗΡΟΦΟΡΙΚΗΣ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF IOANNINA

Υπερανάλυση Εικόνας με τη Μέθοδο Deep Unfolding και Χρήση Εικόνων Πολλαπλών Τύπων

Η Μεταπτυχιακή Διπλωματική Εργασία

υποβάλλεται στην ορισθείσα

από τη Συνέλευση

του Τμήματος Μηχανικών Η/Υ και Πληροφορικής

Εξεταστική Επιτροπή

από τον

Αλέξανδρο Καραμπάση

ως μέρος των υποχρεώσεων για την απόκτηση του

ΔΙΠΛΩΜΑΤΟΣ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

ΣΤΗ ΜΗΧΑΝΙΚΗ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΥΠΟΛΟΓΙΣΤΙΚΩΝ

ΣΥΣΤΗΜΑΤΩΝ

ΜΕ ΕΙΔΙΚΕΥΣΗ

ΣΤΗΝ ΕΠΙΣΤΗΜΗ ΚΑΙ ΜΗΧΑΝΙΚΗ ΔΕΔΟΜΕΝΩΝ

Πανεπιστήμιο Ιωαννίνων

Πολυτεχνική Σχολή

Ιωάννινα 2022

Εξεταστική επιτροπή:

- **Κόντης Λυσίμαχος-Παύλος**, Καθηγητής Τμήμα Μηχανικών Η/Υ και Πληροφορικής, Πανεπιστήμιο Ιωαννίνων (Επιβλέπων)
- **Νίκου Χριστόφορος**, Καθηγητής Τμήμα Μηχανικών Η/Υ και Πληροφορικής, Πανεπιστήμιο Ιωαννίνων
- **Παρσόπουλος Κωνσταντίνος**, Καθηγητής Τμήμα Μηχανικών Η/Υ και Πληροφορικής, Πανεπιστήμιο Ιωαννίνων

ΑΦΙΕΡΩΣΗ

Αφιερώνω την μεταπτυχιακή μου εργασία στην οικογένεια μου και σε όσους με στήριξαν.

ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να ευχαριστήσω την Ε. Τσιλιγιάννη για την πολύτιμη βοήθεια και καθοδήγηση της κατά την διάρκεια της μεταπτυχιακής εργασίας. Επίσης, θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή κ. Α. Κόντη για την συμβολή του στην εργασία αλλά και την εμπιστοσύνη που μου έδειξε αναθέτοντας μου αυτό το θέμα της μεταπτυχιακής εργασίας.

ΠΕΡΙΕΧΟΜΕΝΑ

Κατάλογος Σχημάτων	iv
Κατάλογος Πινάκων	vi
Περίληψη	vii
Extended Abstract	viii
ΚΕΦΑΛΑΙΟ 1 Εισαγωγή	1
1.1 Αντικείμενο της Διπλωματικής Εργασίας.....	1
1.2 Οργάνωση του Τόμου.....	4
ΚΕΦΑΛΑΙΟ 2 Χρήση Βαθιών Νευρωνικών Δικτύων για Αντίστροφα	
Προβλήματα στην Απεικόνιση	5
2.1 Εισαγωγή.....	6
2.2 Αντίστροφο Πρόβλημα	6
2.3 Η Χρήση των Νευρωνικών Δικτύων για Αντίστροφα Προβλήματα στην Επεξεργασία Εικόνας.....	9
2.4 Συνελικτικά Νευρωνικά Δίκτυα (CNN)	10
2.5 Αντιστοίχιση από Άκρη σε Άκρη με την Βοήθεια των Συνελικτικών Νευρωνικών Δικτύων.....	10
2.6 Αύξηση του βάθους των συνελικτικών νευρωνικών δικτύων	12
2.7 Χρήση των Συνελικτικών Νευρωνικών Δικτύων Κωδικοποίησης και Αποκωδικοποίησης.....	14
2.8 Διαδικασία Εκπαίδευσης των Βαθιών Νευρωνικών Δικτύων για Αντίστροφα Προβλήματα.....	15
2.9 Μάθηση Αναπαραστάσεων Υψηλότερου Επιπέδου με την Χρήση Δικτύων CNN Κωδικοποίησης-Αποκωδικοποίησης.....	17
2.10 Το Μοντέλο SRCNN.....	18

ΚΕΦΑΛΑΙΟ 3	Νευρωνικά Δίκτυα και Αναλυτικές Μεθοδοι	21
3.1	Αραιή Αναπαράσταση.....	22
3.1.1	Ο αλγόριθμος ISTA	23
3.1.2	Το μοντέλο LISTA	24
3.2	Εφαρμογή του αλγορίθμου LISTA στην υπερανάλυση εικόνας	25
3.3	Συνελικτική Προσεγγιστική Αραιή Αναπαράσταση.....	27
3.3.1	Συνελικτική Αραιή Κωδικοποίηση	27
3.3.2	Το μοντέλο ACSC.....	28
ΚΕΦΑΛΑΙΟ 4	Μοντελα Πολλαπλών Τύπων (Multimodal Model)	30
4.1	Εισαγωγή	30
4.2	Βαθιά Μάθηση για Αντίστροφη Προβλήματα με Χρήση Πλευρικής Πληροφορίας	31
4.2.1	Αραιή προσέγγιση με πλευρική πληροφορία με τη μέθοδο deep unfolding.....	32
4.2.2	LeSITA για ανακατασκευή σήματος με πλευρική πληροφορία	34
4.3	Συνελικτική Αραιή Κωδικοποίηση για Υπερανάλυση Εικόνας με Χρήση Εικόνων Πολλαπλών Τύπων.....	34
4.4	Υπερανάλυση Εικόνας με το Δίκτυο LMCSC	36
ΚΕΦΑΛΑΙΟ 5	Πειραματικά αποτελεσματα	38
5.1	Λεπτομέρειες Υλοποίησης	38
5.1.1	Κατασκευή συνόλων δεδομένων εκπαίδευσης και ελέγχου	38
5.1.2	Ορισμός και αρχικοποίηση των μεταβλητών του νευρωνικού δικτύου	39
5.1.3	Εκπαίδευση δικτύου.....	39
5.2	Διαδικασία Ελέγχου	40
5.2.1	Πειραματικά δεδομένα	40
5.2.2	Χρήση της RGB εικόνας ως εικόνα καθοδήγησης.....	40
5.2.3	Χρήση του καναλιού Y ως εικόνα καθοδήγησης	41

5.2.4	Οπτικά αποτελέσματα	43
5.2.5	Οπτικά αποτελέσματα για συντελεστή υπερανάλυσης $\times 2$	44
5.2.6	Οπτικά αποτελέσματα για συντελεστή υπερανάλυσης $\times 3$	48
5.2.7	Οπτικά αποτελέσματα για συντελεστή υπερανάλυσης $\times 4$	52
5.2.8	Οπτικά αποτελέσματα για συντελεστή υπερανάλυσης $\times 6$	56
ΚΕΦΑΛΑΙΟ 6	Τελικά Συμπεράσματα και Πιθανοί Τρόποι Βελτίωσης	60
	Βιβλιογραφία	62

ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ

Σχήμα 2.1 Προς τα Εμπρός Πρόβλημα (αριστερά) και το Αντίστροφο Πρόβλημα (δεξιά).....	7
Σχήμα 2.2 Πλήρως Συνδεδεμένο Δίκτυο.....	9
Σχήμα 2.3 Συνελικτικό Νευρωνικό Δίκτυο Τριών Επιπέδων.....	12
Σχήμα 2.4 Συνελικτικό Νευρωνικό Δίκτυο με Χρήση Υπολειπόμενων Μπλοκ	13
Σχήμα 2.5 Κωδικοποιητής- Αποκωδικοποιητής με Χρήση Συνελικτικών Νευρωνικών Δικτύων.....	15
Σχήμα 2.6 Χρήση Δύο Συναρτήσεων Κόστους για Καλύτερο Αποτέλεσμα	17
Σχήμα 2.7 Το Μοντέλο των C.Dong και C.C.Loy[Don2015].....	20
Σχήμα 3.1 Δίκτυο LISTA	25
Σχήμα 3.2 Υπερανάλυση Εικόνας με το Δίκτυο LISTA.....	27
Σχήμα 3.3 Το Μοντέλο ACSC	29
Σχήμα 4.1 Το Δίκτυο LeSITA	33
Σχήμα 4.2 Το Δίκτυο LMCSC.....	36
Σχήμα 4.3 Το Δίκτυο LMCSC για Υπερανάλυση Εικόνας.....	36
Σχήμα 5.1 Αποτελέσματα του Μοντέλου για Συντελεστή Υπερανάλυσης x2.....	44
Σχήμα 5.2 Αποτελέσματα του Μοντέλου για Συντελεστή Υπερανάλυσης x2.	45
Σχήμα 5.3 Αποτελέσματα του Μοντέλου για Συντελεστή Υπερανάλυσης x2.....	46
Σχήμα 5.4 Αποτελέσματα του Μοντέλου για Συντελεστή Υπερανάλυσης x2.....	47
Σχήμα 5.5 Αποτελέσματα του Μοντέλου για Συντελεστή Υπερανάλυσης x3.	48
Σχήμα 5.6 Αποτελέσματα του Μοντέλου για Συντελεστή Υπερανάλυσης x3.	49

Σχήμα 5.7 Αποτελέσματα του Μοντέλου για Συντελεστή Υπερανάλυσης x_3	50
Σχήμα 5.8 Αποτελέσματα του Μοντέλου για Συντελεστή Υπερανάλυσης x_3	51
Σχήμα 5.9 Αποτελέσματα του Μοντέλου για Συντελεστή Υπερανάλυσης x_4 . Σ.....	52
Σχήμα 5.10 Αποτελέσματα του Μοντέλου για Συντελεστή Υπερανάλυσης x_4	53
Σχήμα 5.11 Αποτελέσματα του Μοντέλου για Συντελεστή Υπερανάλυσης x_4	54
Σχήμα 5.12 Αποτελέσματα του Μοντέλου για Συντελεστή Υπερανάλυσης x_4	55
Σχήμα 5.13 Αποτελέσματα του Μοντέλου για Συντελεστή Υπερανάλυσης x_6	56
Σχήμα 5.14 Αποτελέσματα του Μοντέλου για Συντελεστή Υπερανάλυσης x_6	57
Σχήμα 5.15 Αποτελέσματα του Μοντέλου για Συντελεστή Υπερανάλυσης x_6	58
Σχήμα 5.16 Αποτελέσματα του Μοντέλου για Συντελεστή Υπερανάλυσης x_6	59

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

Πίνακας 5.1 Συγκεντρωτικά Αποτελέσματα PSNR για Εικόνες Καθοδήγησης RGB	41
---	----

Πίνακας 5.2 Συγκεντρωτικά Αποτελέσματα PSNR για Εικόνες Καθοδήγησης Y	42
---	----

Πίνακας 5.3 Σύγκριση των Δύο Δικτύων ανάλογα με τις Εικόνες Καθοδήγησης	43
---	----

ΠΕΡΙΛΗΨΗ

Αλέξανδρος Καραμπάσης, Δ.Μ.Σ. στη Μηχανική Δεδομένων και Υπολογιστικών Συστημάτων, Τμήμα Μηχανικών Η/Υ και Πληροφορικής, Πολυτεχνική Σχολή, Πανεπιστήμιο Ιωαννίνων, Απρίλιος 2022

Υπερανάλυση εικόνας με τη μέθοδο deep unfolding και χρήση εικόνων πολλαπλών τύπων.

Επιβλέπων: Κόντης Λυσίμαχος-Παύλος, Καθηγητής

Η παρούσα διπλωματική εργασία πραγματεύεται την υπερανάλυση εικόνας με την βοήθεια των μοντέλων πολλαπλών τύπων. Η υπερανάλυση εικόνας είναι ένα γνωστό αντίστροφο πρόβλημα της μηχανικής όρασης για το οποίο υπάρχουν διάφορες προσεγγίσεις, μεταξύ των οποίων στατιστικές μέθοδοι, αλγόριθμοι που στηρίζονται σε αραιές αναπαραστάσεις και μέθοδοι που χρησιμοποιούν βαθιά μάθηση. Μια πρόσφατη κατηγορία έρευνας στη βαθιά μάθηση για αντίστροφα προβλήματα αποτελεί το deep unfolding δηλαδή η ανάπτυξη ενός επαναληπτικού αλγορίθμου με τη μορφή βαθιών νευρωνικών δικτύων. Πρακτικές εφαρμογές, όπως η ιατρική απεικόνιση, χρειάζονται εικόνες πολλαπλών τύπων που αποτυπώνουν την ίδια σκηνή, επομένως, μια άλλη προσέγγιση είναι η κοινή χρήση πολλαπλών τύπων εικόνας. Το πρόβλημα της υπερανάλυσης εικόνας με την βοήθεια μοντέλων πολλαπλών τύπων αναφέρεται στην ανακατασκευή μιας εικόνας υψηλής ανάλυσης από μια εικόνα χαμηλής ανάλυσης χρησιμοποιώντας μια εικόνα καθοδήγησης άλλου τύπου, που ονομάζεται πλευρική πληροφορία. Στόχος της παρούσας διπλωματικής εργασίας είναι η ανάπτυξη ενός μοντέλου πολλαπλών τύπων που θα ενσωματώνει στοιχεία από την θεωρία αραιών αναπαραστάσεων και θα επιτρέπει την αποτελεσματική ενσωμάτωση του τύπου καθοδήγησης για την επίτευξη μιας αποδοτικής υπερανάλυσης εικόνας. Σε αντίθεση με τα περισσότερα μοντέλα νευρωνικών δικτύων πολλαπλών τύπων η αρχιτεκτονική του μοντέλου που αναπτύχθηκε στην παρούσα εργασία καθιστά το μοντέλο ερμηνεύσιμο, με την έννοια ότι το μοντέλο εκτελεί βήματα παρόμοια με έναν επαναληπτικό αλγόριθμο, ενώ παρέχει τη δυνατότητα θεωρητικής μελέτης του μοντέλου με χρήση της θεωρίας αραιών αναπαραστάσεων.

EXTENDED ABSTRACT

Alexandros Karampasis, M.Sc. in Data and Computer Systems Engineering,
Department of Computer Science and Engineering, School of Engineering,
University of Ioannina, Greece, April 2022

Image super-resolution with the deep unfolding method and the use of
multimodal images.

Advisor: Kondi Lisimachos P., Professor

This thesis deals with image super-resolution using multimodal models. Image super-resolution is a common inverse problem of computer vision for which a plethora of different approaches have been demonstrated over the past years, including statistical methods, algorithms based on sparse representation and methods utilizing machine learning. A recent research in deep learning for inverse problems considers deep unfolding i.e., the unfolding of an iterative algorithm into the form of a DNN. Practical applications, such as medical imaging, involve different image modalities capturing the same scene, therefore, another approach in imaging is the joint use of multiple image modalities. The problem of multimodal image super-resolution refers to the reconstruction of a high resolution image from an lower resolution image using a guidance image from another modality, also referred to as side information. In this present work, we present an multimodal model which incorporates sparse priors and enables efficient integration of the guidance modality into the solution of super-resolution image. In contrast to most other neural network models, the architecture of the model make the model interpretable, in the sense that the model performs steps similar to an iterative algorithm, while also enabling its theoretical study using sparse representation theory

ΚΕΦΑΛΑΙΟ 1

ΕΙΣΑΓΩΓΗ

1.1 Αντικείμενο της διπλωματικής εργασίας

1.2 Οργάνωση του τόμου

1.1 Αντικείμενο της Διπλωματικής Εργασίας

Κάθε χρόνο η τεχνολογία ολοένα και περισσότερο εισβάλλει στην ζωή μας. Οι τεχνολογικοί κολοσσοί είναι σε ένα συνεχή ανταγωνισμό μεταξύ τους ώστε να προλάβει κάποιος να παράγει και να προωθήσει στην αγορά ένα νέο προϊόν. Στόχος τους είναι αυτό το προϊόν να είναι όσο το δυνατόν καλύτερο από την προηγούμενη έκδοσή του, να είναι δηλαδή όσο το δυνατόν “έξυπνότερο”.

Η δυνατότητα για τη δημιουργία έξυπνων προϊόντων είναι συνυφασμένη με την ανάπτυξη του επιστημονικού κλάδου της τεχνητής νοημοσύνης (AI - artificial intelligence). Με τον όρο τεχνητή νοημοσύνη αναφερόμαστε στο σύνολο των εργαλείων που χρησιμοποιούνται για να εκπαιδεύσουν ένα υπολογιστικό σύστημα να συμπεριφέρεται με “ευφυία” με αυτόματο τρόπο [Ben2009]. Η εκπαίδευση αυτή πραγματοποιείται μέσω αλγορίθμων μάθησης οι οποίοι εξάγουν πρότυπα από διαθέσιμα δεδομένα και τα εφαρμόζουν σε νέα δεδομένα, παρέχοντας έτσι τη δυνατότητα στο υπολογιστικό σύστημα να πραγματοποιήσει προβλέψεις. Η τεχνική αυτή ονομάζεται μηχανική μάθηση (machine learning). Τα τεχνητά νευρωνικά δίκτυα (Artificial neural networks) αποτελούν έναν αναδυόμενο τομέα

της μηχανικής μάθησης και της τεχνητής νοημοσύνης, τα οποία μέρα με την μέρα εξελίσσονται, και παράγουν εντυπωσιακά αποτελέσματα στην επιστήμη και στην βιομηχανία. Εφαρμογές στις οποίες συναντάμε τη χρήση τεχνητών νευρωνικών δικτύων περιλαμβάνουν οικιακές συσκευές, έξυπνα κινητά τηλέφωνα, συστήματα αναγνώριση φωνής, ακόμα και αυτο-οδηγούμενα οχήματα [Ben2009].

Ένα επιστημονικό πεδίο στο οποίο χρησιμοποιείται ευρέως η τεχνητή νοημοσύνη είναι και η μηχανική όραση (Computer Vision). Με τον όρο μηχανική όραση [Luc2018] ονομάζουμε την τεχνολογία που δίνει την δυνατότητα σε ένα υπολογιστικό σύστημα να αναγνωρίζει τι υπάρχει σε μια εικόνα ή ένα βίντεο ώστε να μπορεί να την αναλύσει και να την επεξεργαστεί. Για να γίνει η αλγοριθμική επεξεργασία της εικόνας και η εξαγωγή πληροφορίας από αυτήν η τεχνολογία της μηχανικής όρασης μετατρέπει την εικόνα σε δυαδικά, ψηφιακά δεδομένα. Μια από τις σύγχρονες μεθόδους που χρησιμοποιούνται σε προβλήματα μηχανικής όρασης είναι η σε βάθος μάθηση (deep learning).

Με τον ορό σε βάθος μάθηση [Ben2009], εννοούμε την χρήση πολύ-επίπεδων νευρωνικών δικτύων (βαθιά νευρωνικά δίκτυα) που αξιοποιούν τις τελευταίες εξελίξεις στην υπολογιστική ισχύ και τις βελτιωμένες τεχνικές εκπαίδευσης για να μάθουν από διαθέσιμα δεδομένα μεγάλου όγκου. Η σε βάθος μάθηση είναι μια μέθοδος που εφαρμόζεται σε πολλούς τομείς ψηφιακής επεξεργασίας σήματος όπως για παράδειγμα η επεξεργασία φωνής και εικόνας.

Η παρούσα μεταπτυχιακή εργασία εστιάζει στην υπερανάλυση εικόνας, ένα ενεργό θέμα στον τομέα της επεξεργασίας εικόνας. Με τον όρο ανάλυση εικόνας εννοούμε το πλήθος των εικονοστοιχείων (pixels) σε κάποια συγκεκριμένη περιοχή. Μια εικόνα με υψηλή ανάλυση (HR- High Resolution) έχει υψηλή πυκνότητα εικονοστοιχείων η οποία συμβάλει σε μια πιο λεπτομερή απεικόνιση. Σκοπός της υπερανάλυσης εικόνας είναι η ανακατασκευή μιας εικόνας υψηλής ανάλυσης από μια εικόνα χαμηλής ανάλυσης. Η υπερανάλυση εικόνας βοηθάει στο να ξεπεράσουμε τους εγγενείς περιορισμούς ανάλυσης που προκύπτουν από τους αισθητήρες απεικόνισης χαμηλού κόστους. Επίσης προσφέρει μια λύση για την βελτίωση της ποιότητας ήδη υπάρχων εικόνων οι οποίες έχουν παραχθεί από εξοπλισμό απεικόνισης παλαιού τύπου. Οι εικόνες με υψηλή ανάλυση είναι απαραίτητες για την βελτίωση εφαρμογών που απαιτούν την ανθρώπινη ερμηνεία και ανάλυση, καθώς επίσης και σε εφαρμογές που χρειάζεται μηχανική αντίληψη

για αναπαράσταση μοντέλων. Για παράδειγμα η χρήση μιας εικόνας υψηλής ανάλυσης μπορεί να βοηθήσει τους γιατρούς να κάνουν μια σωστή διάγνωση, ή την αστυνομία να εντοπίσει μια πινακίδα ενός αυτοκινήτου ή ένα πρόσωπο πιο εύκολα και γρήγορα. Συμβάλει επίσης στη βελτίωση της αναγνώρισης προτύπων στον τομέα της μηχανικής όρασης.

Όμως η χρήση των εικόνων με υψηλή ανάλυση δεν είναι εύκολη διαδικασία καθώς τα HR συστήματα έχουν υψηλό κόστος. Για την μείωση του κόστους μια εικόνα μπορεί να συμπιεστεί, αλλά αυτό έχει σαν αποτέλεσμα η λεπτομέρεια και η ποιότητά της να μειωθεί. Η λύση στο πρόβλημα αυτό δίνεται από τις τεχνικές επεξεργασίας σημάτων οι οποίες αξιοποιούν τις εικόνες με χαμηλότερη ανάλυση για την δημιουργία εικόνων υψηλής ανάλυσης. Σκοπός της παρούσας εργασίας είναι η δημιουργία μιας εικόνας υψηλής ανάλυσης από μία εικόνα χαμηλής ανάλυσης (υπερανάλυση εικόνας) μέσω μιας σχέσης μεταξύ της εικόνας χαμηλής ανάλυσης και της εικόνας υψηλής και με την βοήθεια μιας πλευρικής πληροφορίας χρησιμοποιώντας ένα βαθύ εκπαιδεύσιμο νευρωνικό δίκτυο πολλαπλών εισόδων.

Η υπερανάλυση εικόνας(image super resolution) μπορεί να εφαρμοστεί σε αρκετά παραδείγματα της καθημερινής μας ζωής. Για παράδειγμα μπορεί να χρησιμοποιηθεί στα κινητά τηλέφωνα ώστε να βελτιωθεί η ποιότητας της φωτογραφίας ή του βίντεο που παράγει η κάμερα του μειώνοντας έτσι το κόστος παραγωγής καθώς και με έναν φακό χαμηλής ποιότητας μπορούν να επιτευχθούν εικόνες ικανοποιητικής ανάλυσης. Ένα άλλο παράδειγμα εντοπίζεται στην ιατρική επιστήμη. Εικόνες από ιατρικά δεδομένα, όπως απο ακτινογραφίες, υπερηχογραφήματα ή μικροσκόπια, μπορούν να έχουν υψηλή ποιότητα ώστε να εξαχθεί με μεγαλύτερη ακρίβεια πληροφορία από αυτά τα δεδομένα με σκοπό να υποβοηθηθεί η διάγνωση και η θεραπεία ασθενειών. Επιπλέον μελλοντικά η υπερανάλυση εικόνας μπορεί να εφαρμοστεί και δυναμικά σε εικόνες, όπως για παράδειγμα σε ένα ηλεκτρονικό παιχνίδι όπου με την κατάλληλη επεξεργασία θα μπορεί να δημιουργείται εικόνα με καλύτερη ανάλυση και ο χρήστης να έχει μια καλύτερη εμπειρία με χαμηλότερο κόστος.

Στην παρούσα διπλωματική εργασία αναλύεται ένα deep unfolding μοντέλο πολλαπλών τύπων για υπερανάλυση εικόνας το οποίο χρησιμοποιεί την πλευρική πληροφορία από μια εικόνα καθοδήγησης για την επίτευξη του στόχου. Η

πλευρική πληροφορία αποτελεί την αραιή αναπαράσταση της εικόνας καθοδήγησης. Η αραιή αναπαράσταση ενός σήματος παρέχει τη δυνατότητα μιας συμπαγούς κωδικοποίησης της πληροφορίας που περιέχεται στο σήμα. Ο σχεδιασμός ενός αλγορίθμου προσδιορισμού της αραιής αναπαράστασης ενός σήματος μέσω νευρωνικών δικτύων μειώνει σημαντικά το κόστος των απαιτούμενων υπολογισμών. Συνεπώς, το λαμβανόμενο μοντέλο συνδυάζει τα πλεονεκτήματα των αναλυτικών μοντέλων με τις δυνατότητες γρήγορης επεξεργασίας που προσφέρουν τα νευρωνικά δίκτυα.

1.2 Οργάνωση του Τόμου

Η παρούσα διπλωματική εργασία χωρίζεται στα εξής επιμέρους κεφάλαια.

Στο δεύτερο κεφάλαιο ορίζονται τα αντίστροφα προβλήματα και περιγράφεται η χρήση διαφορετικών βαθιών νευρωνικών δικτύων για την επίλυση αντίστροφων προβλημάτων στην απεικόνιση.

Στο τρίτο κεφάλαιο περιγράφονται οι αναλυτικές μέθοδοι που χρησιμοποιούνται για την επίλυση αντίστροφων προβλημάτων στην απεικόνιση αλλά και το πώς μπορούμε να τις εκφράσουμε με την μορφή βαθιών νευρωνικών δικτύων.

Στο τέταρτο κεφάλαιο γίνεται αναφορά στα μοντέλα πολλαπλών τύπων. Δίνεται ο ορισμός και η ανάλυση τους και παρουσιάζεται η χρήση τους στα αντίστροφα προβλήματα απεικόνισης. Επίσης παρουσιάζεται και το μοντέλο της παρούσας εργασίας που αποτελεί μοντέλο πολλαπλών τύπων.

Στο πέμπτο κεφάλαιο παρουσιάζονται τα αποτελέσματα του μοντέλου μας τόσο σε επίπεδο μαθηματικών πινάκων με την χρήση της μετρικής PSNR όσο και σε οπτικό επίπεδο καθώς έτσι γίνεται πραγματικά αντιληπτό το αποτέλεσμα του δικτύου αυτής της μελέτης.

Στο έκτο κεφάλαιο αναφέρονται τα τελικά συμπεράσματα που προέκυψαν από αυτήν την μελέτη και προτείνονται πιθανοί τρόποι βελτίωσης του μοντέλου υπερανάλυσης εικόνας.

Στο έβδομο και τελευταίο κεφάλαιο παρατίθεται η βιβλιογραφία που χρησιμοποιήθηκε στην μεταπτυχιακή εργασία.

ΚΕΦΑΛΑΙΟ 2

ΧΡΗΣΗ ΒΑΘΙΩΝ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ ΓΙΑ ΑΝΤΙΣΤΡΟΦΑ ΠΡΟΒΛΗΜΑΤΑ ΣΤΗΝ ΑΠΕΙΚΟΝΙΣΗ

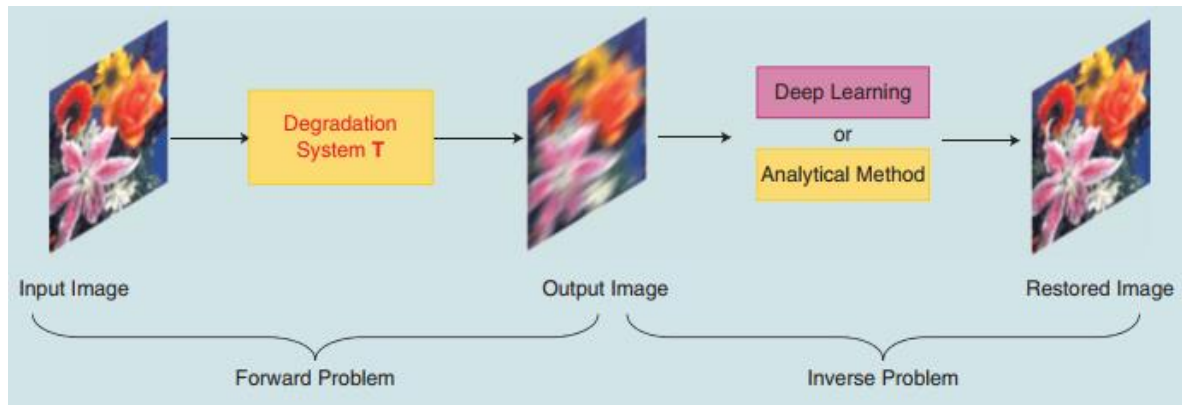
-
- 2.1 Εισαγωγή
 - 2.2 Αντίστροφο πρόβλημα
 - 2.3 Η χρήση των νευρωνικών δικτύων για αντίστροφα προβλήματα στην επεξεργασία εικόνας
 - 2.4 Συνελικτικά νευρωνικά δίκτυα (CNN)
 - 2.5 Αντιστοίχιση από άκρη σε άκρη με την βοήθεια των συνελικτικών νευρωνικών δικτύων
 - 2.6 Αύξηση του βάθους των συνελικτικών νευρωνικών δικτύων
 - 2.7 Χρήση των συνελικτικών νευρωνικών δικτύων κωδικοποίησης και αποκωδικοποίησης
 - 2.8 Διαδικασία εκπαίδευσης των βαθιών νευρωνικών δικτύων για αντίστροφα προβλήματα
 - 2.9 Μάθηση αναπαραστάσεων υψηλότερου επιπέδου με την χρήση δικτύων CNN κωδικοποίησης-αποκωδικοποίησης
 - 2.10 Το μοντέλο SRCNN
-

2.1 Εισαγωγή

Παραδοσιακά για την επίλυση προβλημάτων απεικόνισης, όπως είναι για παράδειγμα η υπερανάλυση εικόνας και η αποκατάσταση εικόνας έχουν χρησιμοποιηθεί αναλυτικές μέθοδοι. Με την πάροδο του χρόνου και με την βοήθεια της τεχνολογικής εξέλιξης ο τομέας της μηχανικής μάθησης και η βαθιά μάθηση αξιοποιούνται όλο και περισσότερο για την επίλυση τέτοιων προβλημάτων απεικόνισης καταφέροντας μάλιστα πολλές φορές να έχουν και καλύτερο αποτέλεσμα από αυτό των αναλυτικών μεθόδων. Σε αντίθεση με τις αναλυτικές μεθόδους που το πρόβλημα είναι ορισμένο και χρησιμοποιείται καθορισμένη λύση, στα βαθιά νευρωνικά δίκτυα η λύση για το αντίστροφο πρόβλημα μαθαίνεται μέσα από ένα μεγάλο σύνολο δεδομένων.

2.2 Αντίστροφο Πρόβλημα

Στις περισσότερες εφαρμογές ένα παρατηρούμενο σήμα y μπορεί να μοντελοποιηθεί ως έξοδος ενός συστήματος T , του οποίου η είσοδος δηλώνεται από το x [Luc2018]. Τόσο η είσοδος όσο και η έξοδος μπορούν να είναι πολυδιάστατα σήματα, όπως είναι για παράδειγμα οι εικόνες που είναι δυοδιάστατα σήματα. Ένα σύστημα θεωρείται προς τα εμπρός μοντέλο (forward model) όταν για ένα δοθέν σήμα εισόδου x και υποθέτοντας ότι έχουμε γνώση για το σύστημα αποφασίζουμε την έξοδο y . Αντίθετα όταν ψάχνουμε την είσοδο x για μια δοθείσα έξοδο y και έχουμε γνώση του συστήματος τότε αυτό ονομάζεται αντίστροφο πρόβλημα (inverse problem). Στην Σχήμα 2.1 παρουσιάζονται γραφικά τα δύο προβλήματα.



Σχήμα 2.1 Προς τα Εμπρός Πρόβλημα (αριστερά) και το Αντίστροφο Πρόβλημα (δεξιά)

Τα αντίστροφα προβλήματα χαρακτηρίζονται από την δυσκολία της επίλυσης τους καθώς εξαρτώνται από τις ιδιότητες της αντιστοίχισης T , καθιστώντας το σύστημα να είναι κακώς ορισμένο. Παραδείγματα αντίστροφων προβλημάτων, ή όπως αλλιώς λέγονται προβλημάτων αποκατάστασης, είναι η ανάκτηση, η συμπίεστική δειγματοληψία (compressive sensing – CS), η εκτίμηση κίνησης και βάθους και η υπερανάλυση εικόνας, πρόβλημα με το οποίο ασχολείται και η παρούσα διπλωματική. Η επίλυση των αντίστροφων προβλημάτων είναι ένα θέμα στο οποίο έχουν γίνει και γίνονται ακόμα έρευνες.

Ένας πιθανός τρόπος επίλυσης τους είναι οι αναλυτικές τεχνικές. Με τις αναλυτικές τεχνικές ορίζεται το προς τα εμπρός μοντέλο, επιλέγονται τα σωστά κριτήρια για την επίτευξη της λύσης και επιλέγεται μια προσεγγιστική λύση. Οι αναλυτικές τεχνικές χωρίζονται σε ντετερμινιστικές και στοχαστικές. Στις ντετερμινιστικές επιλέγεται ένα κριτήριο βελτιστοποίησης όπως είναι για παράδειγμα το σφάλμα νόρμας-2 $\|y - T(x)\|^2$. Έπειτα μια εκ των προτέρων γνώση ενσωματώνεται στην λύση και ένας επιπλέον όρος συμπεριλαμβάνεται στην συνάρτηση βελτιστοποίησης όπως είναι η ομαλότητα της λύσης. Στις στοχαστικές τεχνικές όλες οι άγνωστες ποσότητες αντιμετωπίζονται ως στοχαστικές και χρησιμοποιείται μια μέγιστη πιθανοφάνεια ή μια εκ των υστέρων ή μια πλήρως (ιεραρχική) προσέγγιση bayesian. Στην τελευταία περίπτωση η εκ των προτέρων γνώση που αποκτούμε εισάγεται στα ειδικά πιθανοτικά μοντέλα.

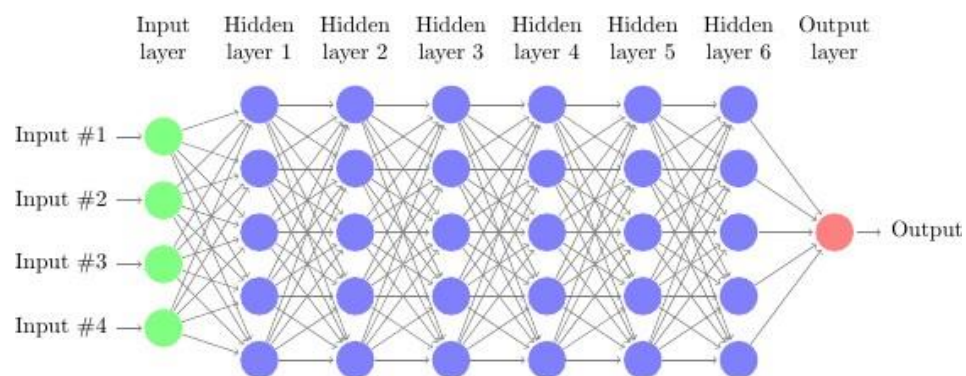
Μια άλλη πιθανή προσέγγιση λύσης των αντιστρόφων προβλημάτων είναι η χρήση βαθιών νευρωνικών δικτύων. Με τα βαθιά νευρωνικά δίκτυα προσπαθούμε να ελαχιστοποιήσουμε την σχέση $\|\mathbf{x} - g_\phi(\mathbf{y})\|^2$ για μία συνάρτηση $g_\phi(\cdot)$ που παίζει τον ρόλο του T^{-1} . Αυτή η συνάρτηση χρησιμοποιείται για να μάθει το δίκτυο τις παραμέτρους του ϕ . Αναλυτικότερα για την χρήση των βαθιών νευρωνικών δικτύων θα αναφερθούμε σε επόμενο κεφάλαιο.

Έτσι τίθεται το ερώτημα σχετικά με τα πλεονεκτήματα της κάθε τεχνικής, δηλαδή ποια από τις δύο προσεγγίσεις πρέπει να χρησιμοποιηθεί. Στην προσέγγιση των βαθιών νευρωνικών δικτύων το υπολογιστικό κόστος μετατίθεται στην διαδικασία της μάθησης ενώ η διαδικασία του ελέγχου υλοποιείται από ένα προς τα εμπρός μοντέλο καθιστώντας την έτσι υπολογιστικά αποδοτική. Αντίθετα, η προσέγγιση των αναλυτικών μεθόδων βασίζεται σε προσεγγίσεις βελτιστοποίησης, οι οποίες είναι υπολογιστικά πιο πολύπλοκες. Γενικά, όσο πιο περίπλοκη είναι η μοντελοποίηση του αντίστροφου προβλήματος, τόσο πιο απαιτητική είναι η διαδικασία βελτιστοποίησης. Όμως, σε περίπτωση που υπάρχει εκ των προτέρων πληροφορία τότε η προσέγγιση των αναλυτικών μεθόδων έχουν σαφή προβάδισμα μιας και αυτή η γνώση αποτελεί σημαντικό κομμάτι αυτής της προσέγγισης. Αντίθετα, στα νευρωνικά δίκτυα η ενσωμάτωση αυτής της πληροφορίας θεωρείται δύσκολη.

Ένα ακόμα ερώτημα σχετικά με τις δύο τεχνικές είναι κάτω από ποιες συνθήκες η λύση των νευρωνικών δικτύων είναι πιο ακριβής από αυτή των αναλυτικών μεθόδων. Δυστυχώς δεν υπάρχει ακριβής απάντηση σε αυτό το ερώτημα. Στην περίπτωση που ο τελεστής είναι γνωστός και αντιστρέψιμος και υπάρχει λίγο ή καθόλου θόρυβος στα δεδομένα τότε δεν κερδίζουμε και πολλά περισσότερα στην λύση με νευρωνικά δίκτυα. Από την άλλη πλευρά όταν ο τελεστής δεν είναι γνωστός ή δεν μπορεί να μοντελοποιηθεί με μαθηματική ακρίβεια ή αποτελείται από την συνένωση πολλών τελεστών (για παράδειγμα το σήμα είναι και συμπιεσμένο και θολό) τότε η λύση των νευρωνικών δικτύων υπερέχει αυτής των αναλυτικών μεθόδων.

2.3 Η Χρήση των Νευρωνικών Δικτύων για Αντίστροφα Προβλήματα στην Επεξεργασία Εικόνας

Ένα βαθύ νευρωνικό δίκτυο μπορεί να περιγραφεί ως μια πολυεπίπεδη στοίβα απλών επιπέδων, κάθε ένα από τα οποία μετατρέπει την είσοδο του σε μια νέα αναπαράσταση, η οποία είναι και η έξοδος του επιπέδου, και στην συνέχεια η έξοδος του του επιπέδου χρησιμοποιείται ως είσοδος στο επόμενο επίπεδο. Ένα παράδειγμα ενός πλήρους συνδεδεμένου νευρωνικού δικτύου, που ονομάζεται αλλιώς και πολυεπίπεδο perceptron (MPL) ,παρουσιάζεται στο Σχήμα 2.2.



Σχήμα 2.2 Πλήρως Συνδεδεμένο Δίκτυο

Εξαιτίας της ικανότητας των βαθιών νευρωνικών δικτύων να εκτελούν συμπεράσμα γρήγορης προώθησης, προτιμάται η προσέγγισή της εκπαίδευση των βαθιών νευρωνικών δικτύων για μάθηση μιας αντιστοίχισης μεταξύ μιας παρατήρησης y και της αντίστοιχης ανακατασκευής της $\hat{x}$. Η πιο διάσημη αρχιτεκτονική για τον σκοπό της αντιστοίχισης είναι αυτή των πλήρως συνδεδεμένων νευρωνικών δικτύων. Επιπλέον τα πλήρως συνδεδεμένα νευρωνικά δίκτυα μπορούν να χρησιμοποιηθούν και για άλλους σκοπούς. Για παράδειγμα στην έρευνα των Zhang και Salari [ZhS2005] χρησιμοποιούνται για αφαίρεση θορύβου στις εικόνες στον τομέα των κυματιδίων. Επίσης, οι Burger και άλλοι [BSH2012] χρησιμοποίησαν τα πλήρως συνδεδεμένα νευρωνικά δίκτυα για να κάνουν απευθείας αντιστοίχιση των εικόνων με θόρυβο στα αντίστοιχα κομμάτια με μη θόρυβο, ενώ μπορούν να χρησιμοποιηθούν για την αφαίρεση των αστοχιών

που παράγονται κατά την αντίστροφη συνέλιξη όπως έγινε στην έρευνα των Schuler και άλλων [Sch2013].

2.4 Συνελικτικά Νευρωνικά Δίκτυα (CNN)

Τα συνελικτικά νευρωνικά δίκτυα (ΣΝΔ) ή convolutional neural network (CNN) στα αγγλικά, είναι νευρωνικά δίκτυα τα οποία πραγματοποιούν συνέλιξη της εισόδου με προκαθορισμένα φίλτρα. Εκτός από συνέλιξη εκτελούν διαδικασίες pooling, κανονικοποίηση και εισαγωγή μη γραμμικότητας. Οι πληροφορίες των φίλτρων όπως μέγεθος, διαστάσεις και βάθος, δηλαδή αριθμών καναλιών, αρχικοποιούνται εξ αρχής και διαφέρουν ανάλογα με το πρόβλημα που έχουμε να λύσουμε. Οι τιμές εισόδου και εξόδου είναι διακριτές [Παπ2013].

Για την εκπαίδευση των συνελικτικών νευρωνικών δικτύων μπορούμε να χρησιμοποιήσουμε είτε πλήρως συνδεδεμένα δίκτυα είτε τοπικά συνδεδεμένα δίκτυα. Στην αρχιτεκτονική της πλήρους σύνδεσης (fully connected), κάθε νευρώνας στα κρυφά επίπεδα συνδέεται με την έξοδο όλων των νευρώνων του προηγούμενου επιπέδου και παράγει τα χαρακτηριστικά ολόκληρης της εικόνας. Αυτό όμως είναι εφικτό να συμβεί σε εικόνες σχετικά μικρές, δηλαδή μέχρι 28 X 28, αφού σε μεγαλύτερες το υπολογιστικό κόστος αυξάνεται κάνοντας τον αλγόριθμο αισθητά πιο αργό. Σε αυτήν την περίπτωση χρησιμοποιείται η αρχιτεκτονική των τοπικών συνδεδεμένων δικτύων. Σε αυτήν την αρχιτεκτονική ένας νευρώνας στο κρυφό επίπεδο μπορεί να συνδεθεί με ένα μικρό πλήθος νευρώνων εισόδου, περιορίζοντας έτσι τις συνδέσεις μεταξύ του επιπέδου εισόδου και του κρυφού επιπέδου.

2.5 Αντιστοίχιση από Άκρη σε Άκρη με την Βοήθεια των Συνελικτικών Νευρωνικών Δικτύων

Ένα πλήρως συνδεδεμένο δίκτυο το οποίο έχει μεγάλο πλήθος νευρώνων στα κρυμμένα του επίπεδα έχει την ικανότητα να αναπαριστά οποιαδήποτε λειτουργία θέλουμε να μάθει. Παρόλα αυτά, όταν τα δεδομένα μας είναι υψηλά δομημένα όπως είναι για παράδειγμα οι εικόνες ή τα βίντεο τότε το πιο συχνό

μοντέλο που χρησιμοποιούμε είναι αυτό του συνελικτικού νευρωνικού δικτύου. Ο λόγος που επιλέγονται τα συνελικτικά νευρωνικά δίκτυα για την επεξεργασία εικόνων είναι ότι μπορούν εύκολα να εξάγουν στατιστικά από την είσοδο που δέχονται και να τα χρησιμοποιήσουν για την επίλυση του αντίστροφου προβλήματος.

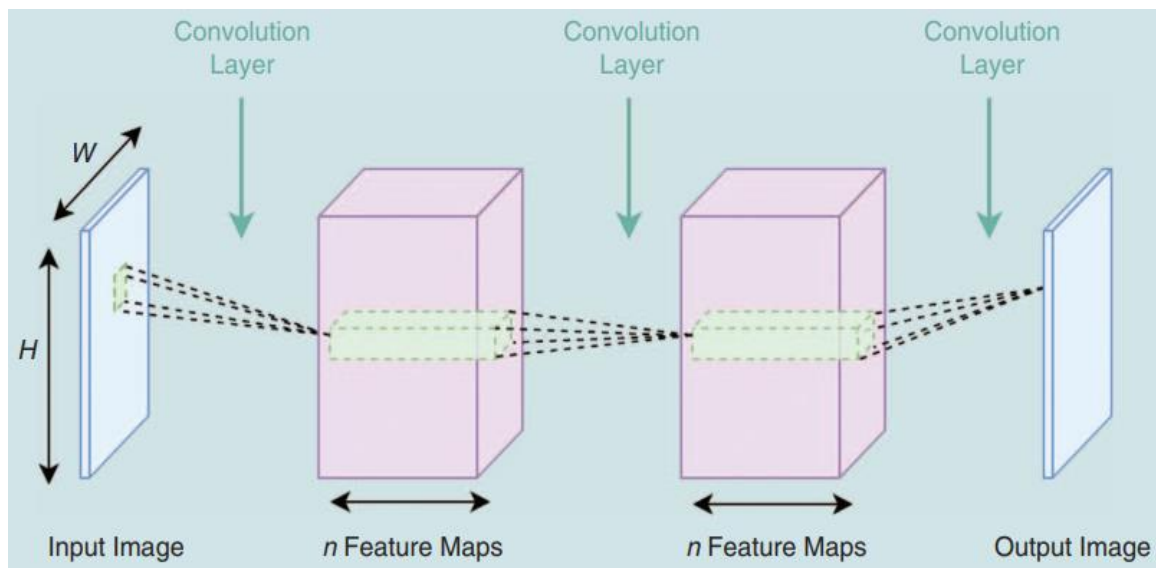
Η διαφορά μεταξύ των πλήρως συνδεδεμένων νευρωνικών δικτύων και των συνελικτικών είναι ότι τα συνελικτικά εκτελούν συνελίξεις στα προηγούμενα επίπεδα. Εκτελώντας μια συνέλιξη με έναν πυρήνα παράγουμε χαρακτηριστικά αντιστοίχισης ενώ μπορούμε να εκτελέσουμε συνέλιξη με πολλούς πυρήνες ώστε να παράγουμε πολλαπλά χαρακτηριστικά αντιστοίχισης τα οποία θα τα χρησιμοποιήσουμε μαζί για να παράγουμε μια νέα αναπαράσταση για την είσοδο.

Τα οφέλη των συνελικτικών νευρωνικών δικτύων στην επίλυση των αντίστροφων προβλημάτων είναι πολλαπλά. Αρχικά επειδή τα βάρη των πυρήνων είναι καθορισμένα, το δίκτυο έχει να μάθει λιγότερες παραμέτρους. Η μείωση του αριθμού των παραμέτρων απλοποιεί το πρόβλημα της βελτιστοποίησης. Επιπλέον, η λειτουργία της συνελίξης προσφέρει συμφέρουσες ιδιότητες όταν χρησιμοποιούνται εικόνες, όπως είναι για παράδειγμα η μεταβλητότητα μετάφρασης και η τοποθεσία. Επιπρόσθετα, έχει αποδειχθεί ότι τα συνελικτικά νευρωνικά δίκτυα είναι ιδιαίτερα κατάλληλα για την εκμάθηση ενδιαφερουσών αναπαραστάσεων από εικόνες. Ακόμα εξάγουν σημαντική πληροφορία από μια παρατηρούμενη εικόνα y η οποία μετά χρησιμοποιείται για να παραχθεί η ανακατασκευή της $\hat{x}$. Τέλος έχει αποδειχθεί ότι οι αρχιτεκτονικές βασισμένες στην συνέλιξη, αν σχεδιαστούν προσεκτικά τότε μπορούν να έχουν ομοιότητες με τις αναλυτικές μεθόδους, καθιστώντας τις αρχιτεκτονικές αυτές ισχυρά εργαλεία επίλυσης αντίστροφων προβλημάτων στην απεικόνιση.

Όταν τα συνελικτικά νευρωνικά δίκτυα χρησιμοποιούνται για διακριτές εργασίες, όπως είναι η ταξινόμηση, κάθε συνέλιξη που πραγματοποιείται σε κάθε βήμα οδηγεί στην μείωση της χωρικής έκτασης των χαρακτηριστικών αντιστοίχισης έτσι ώστε στο τελευταίο επίπεδο να αναπαριστά τις χαμηλής διάστασης κατηγορίες των αντίστοιχων εικόνων εισόδου. Όμως στα αντίστροφα προβλήματα εικόνων η έξοδος του δικτύου θα πρέπει να είναι μια εικόνα υψηλής διάστασης, συνήθως ίδιας διάστασης με την εικόνα εισόδου. Για αυτό τον λόγο όταν χρησιμοποιούμε τα συνελικτικά νευρωνικά δίκτυα για την επίλυση ενός

αντίστροφου προβλήματος θα πρέπει να διατηρούμε τις διαστάσεις των χαρακτηριστικών αντιστοίχισης ίσες με τις διαστάσεις της εισόδου του συνελικτικού επιπέδου το οποίο είναι εφικτό με τη χρήση κατάλληλης συμπλήρωσης (padding) με μηδενικά. Ένα τέτοιο δίκτυο παρουσιάζεται στο Σχήμα 2.3.

Όπως έχει αναφερθεί το πρόβλημα της υπερανάλυσης εικόνας ανήκει στα αντίστροφα προβλήματα. Ο Dong και άλλοι [Don2015] χρησιμοποίησαν ένα δίκτυο συνέλιξης τριών επιπέδων το οποίο δεχόταν σαν είσοδο τμήμα (patch) μιας εικόνα παρεμβολής χαμηλής ανάλυσης (Low resolution) και παράγαγε το αντίστοιχο κομμάτι υψηλής ανάλυσης (High resolution) με ικανοποιητικά αποτελέσματα.



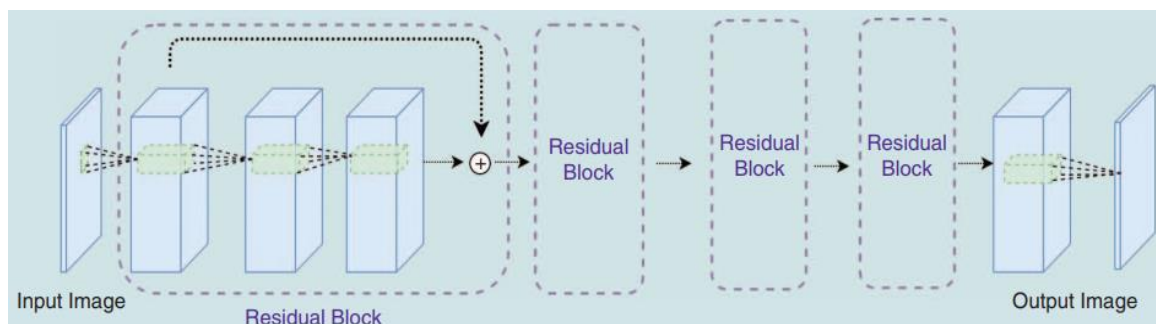
Σχήμα 2.3 Συνελικτικό Νευρωνικό Δίκτυο Τριών Επιπέδων

2.6 Αύξηση του βάθους των συνελικτικών νευρωνικών δικτύων

Έχει αποδειχθεί από έρευνες, όπως αυτή των Hradis και άλλων [Hra2015], ότι η εκπαίδευση βαθύτερων νευρωνικών δικτύων παράγει αποτελέσματα πολύ καλύτερης ποιότητας συγκριτικά με τα αποτελέσματα ενός πιο ρηχού δικτύου. Μπορεί να γίνει η υπόθεση ότι η κατασκευή νευρωνικών δικτύων με μεγαλύτερο βάθος παρέχει στα δίκτυα υψηλότερη ισχύ στην αναπαράσταση αλλά και ότι αυξάνει το συνολικό δεκτικό πεδίο των μοντέλων με αποτέλεσμα να βελτιώνει

την απόδοση στον τομέα της ανάκτησης εικόνας. Βέβαια τα παλιότερα χρόνια ήταν δύσκολο να πραγματοποιηθεί εκπαίδευση σε ένα νευρωνικό δίκτυο με πάρα πολλά επίπεδα καθώς και περιορισμένη υπολογιστική δύναμη υπήρχε αλλά και οι βάσεις δεδομένων των εικόνων δεν ήταν πολύ μεγάλες. Όμως, αυτό έχει αλλάξει τα τελευταία χρόνια. Ο λόγος είναι ότι πλέον υπάρχει πρόσβαση σε μεγάλες βάσεις δεδομένων, υπάρχουν συστήματα με ισχυρή υπολογιστική δύναμη τα οποία σε συνδυασμό με τις αποτελεσματικές συναρτήσεις ενεργοποίησης βοηθούν στην εκπαίδευση βαθιών νευρωνικών δικτύων.

Η χρήση των υπολειπόμενων μπλοκ (residual block) [He2016] έχει παίξει σημαντικό ρόλο στην εκπαίδευση πολύ βαθιών νευρωνικών δικτύων. Αντί να γίνει μάθηση για μια νέα συνάρτηση αντιστοίχισης από το ένα επίπεδο στο άλλο, τα υπολειπόμενα μπλοκ μαθαίνουν ένα υπόλοιπο μεταξύ δύο ή περισσότερων επιπέδων προσθέτοντας μια σύνδεση παράλειψης μεταξύ της εισόδου του υπολειπόμενου μπλοκ και στην έξοδο του. Επειδή η εκπαίδευση των υπόλοιπων είναι ευκολότερη εργασία συγκριτικά με την εκπαίδευση αντιστοίχισης μεταξύ δύο επιπέδων τα βαθιά νευρωνικά δίκτυα με την βοήθεια υπολειπόμενων μπλοκ θεωρείται μια σταθερή και ταυτόχρονα ικανοποιητική εναλλακτική για την εκπαίδευση βαθιών νευρωνικών δικτύων. Στο Σχήμα 2.4 παρουσιάζεται γραφικά ένα δίκτυο συνελίξης που χρησιμοποιεί υπολειπόμενα μπλοκ.



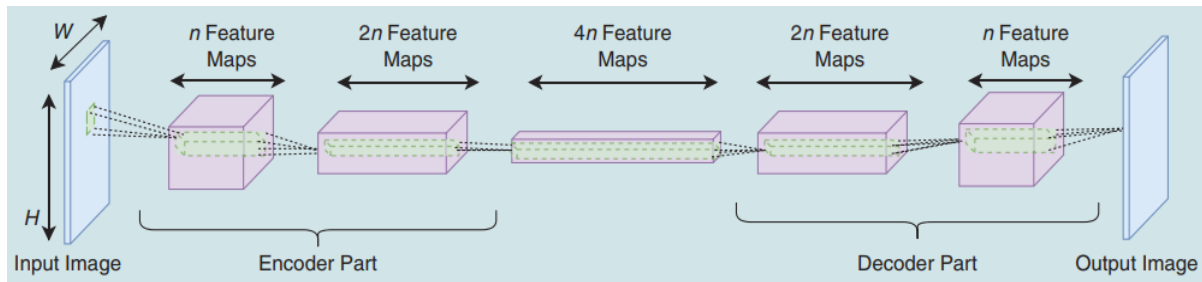
Σχήμα 2.4 Συνελικτικό Νευρωνικό Δίκτυο με Χρήση Υπολειπόμενων Μπλοκ

Επιπλέον υπάρχει η εναλλακτική οι συνδέσεις παράλειψης, αντί να χρησιμοποιηθούν μεταξύ της εισόδου και της εξόδου των υπολειπόμενων μπλοκ, να εισαχθούν μεταξύ του επιπέδου εισόδου και του επιπέδου εξόδου του νευρωνικού δικτύου. Η εναλλακτική αυτή εφαρμόζεται κυρίως σε προβλήματα αποκατάστασης εικόνας όπου η εικόνα εισόδου και εξόδου έχουν παρόμοια

πληροφορία. Για παράδειγμα, στο πρόβλημα της υπερανάλυσης εικόνας η χρησιμοποίηση συνδέσεων παράλειψης μεταξύ του επιπέδου εισόδου και του επιπέδου εξόδου έχει σαν αποτέλεσμα να οδηγεί το δίκτυο να μαθαίνει τις λεπτομέρειες της εικόνας. Έτσι σε ένα δίκτυο συνέλιξης που χρησιμοποιείται για υπερανάλυση εικόνας αν προστεθεί μια σύνδεση παράλειψης μεταξύ του κομματιού εισόδου χαμηλής ανάλυσης και του κομματιού εξόδου υψηλής ανάλυσης, το δίκτυο μαθαίνει να προβλέπει τα χαμένα στοιχεία υψηλής συχνότητας από το χαμηλής ποιότητας κομματιού αντί να μάθει μια νέα αντιστοίχιση μεταξύ του κομματιού χαμηλής και υψηλής ποιότητας.

2.7 Χρήση των Συνελικτικών Νευρωνικών Δικτύων Κωδικοποίησης και Αποκωδικοποίησης

Όπως αναφέρθηκε και παραπάνω όταν χρησιμοποιούμε συνελικτικά νευρωνικά δίκτυα διατηρούμε τις διαστάσεις των χαρακτηριστικών αντιστοίχισης ίσες με τις διαστάσεις της εισόδου του συνελικτικού επιπέδου. Όμως υπάρχει η επιλογή κάποιος να επιλέξει να πραγματοποιήσει υποδειγματοληψία στα χαρακτηριστικά αντιστοίχισης σε κάθε βήμα συνέλιξης μέχρι ένα επίπεδο συμφόρησης και, στη συνέχεια, να πραγματοποιήσει υπερ-δειγματοληψία έτσι ώστε το μέγεθος τους να είναι ίσο με το μέγεθος της εξόδου. Η διαδικασία αυτή παρουσιάζεται γραφικά στο Σχήμα 2.5. Αυτή η ιδέα εφαρμόζεται συνήθως σε εργασίες τμηματοποίησης και πρόβλεψης βάθους. Το πρώτο μέρος του δικτύου ονομάζεται κωδικοποιητής (encoder), και μαθαίνει μια αφηρημένη αναπαράσταση της εικόνας εισόδου. Έπειτα αυτή η αναπαράσταση χρησιμοποιείται στο δεύτερο τμήμα του δικτύου για την παραγωγή μιας εικόνας εξόδου, το οποίο ονομάζεται αποκωδικοποιητής (decoder).



Σχήμα 2.5 Κωδικοποιητής- Αποκωδικοποιητής με Χρήση Συνελικτικών Νευρωνικών Δικτύων

Όμως η χρήση κωδικοποιητή οδηγεί σε σημαντική απώλεια πληροφορίας λεπτομέρειας στην εικόνα εξόδου, καθώς σε κάθε βήμα συνέλιξης ο κωδικοποιητής συμπιέζει χωρική πληροφορία από τα χαρακτηριστικά αντιστοίχισης. Μία λύση σε αυτό το πρόβλημα είναι η εισαγωγή συνδέσεων παράλειψης μεταξύ του χαμηλότερου επιπέδου συνέλιξης που χρησιμοποιείται για την μείωση της διάστασης και του αντίστοιχου ψηλότερου επιπέδου συνέλιξης για την αύξηση διάστασης. Αυτή η αρχιτεκτονική ονομάζεται U-net architecture [RFB2012] και έχει χρησιμοποιηθεί σε αντίστροφα προβλήματα όπως είναι η αφαίρεση θορύβου από την εικόνα και η οπτική ροή.

2.8 Διαδικασία Εκπαίδευσης των Βαθιών Νευρωνικών Δικτύων για Αντίστροφα Προβλήματα

Προηγουμένως αναφέρθηκαν διάφορες αρχιτεκτονικές που μπορούν να χρησιμοποιηθούν για την επίλυση των αντίστροφων προβλημάτων. Ωστόσο, η απλή επιλογή μιας αρχιτεκτονικής νευρωνικού δικτύου, είτε είναι για παράδειγμα πολυεπίπεδο perceptron είτε είναι συνελικτικό νευρωνικό δίκτυο, με τυχαίες παραμέτρους φ θα αποτύχει στην επίλυση του αντίστροφου προβλήματος. Η επιλογή της αρχιτεκτονικής ορίζει ένα σύνολο λειτουργικών σχέσεων που πρέπει να μάθει το δίκτυο. Η διαδικασία εκπαίδευσης έχει αυτόν τον ρόλο, δηλαδή τον προσδιορισμό της βέλτιστης σχέσης για μια συγκεκριμένη εργασία.

Η εκπαίδευση των βαθιών νευρωνικών δικτύου με επίβλεψη, απαιτεί την ύπαρξη ενός συνόλου δεδομένων μεγάλου μεγέθους που θα περιέχει ζεύγη (y, x) . Συγκεκριμένα στο πλαίσιο των αντίστροφων προβλημάτων, αυτά τα ζεύγη

μπορούν να δημιουργηθούν συνθετικά διαφθείροντας την αρχική εικόνα x με μια λειτουργία μετασχηματισμού T . Επιπλέον εκτός από το σύνολο δεδομένων η διαδικασία της εκπαίδευσης απαιτεί και μια συνάρτηση απώλειας/κόστους. Όταν αναφερόμαστε στα αντίστροφα προβλήματα αυτή η συνάρτηση είναι συνήθως η συνάρτηση ελάχιστων τετραγώνων (MSE – Mean Square Error), μιας και χειριζόμαστε τα αντίστροφα προβλήματα σαν προβλήματα παλινδρόμησης. Ο τύπος της συνάρτησης ελάχιστων τετραγώνων δίνεται από την σχέση (2.1):

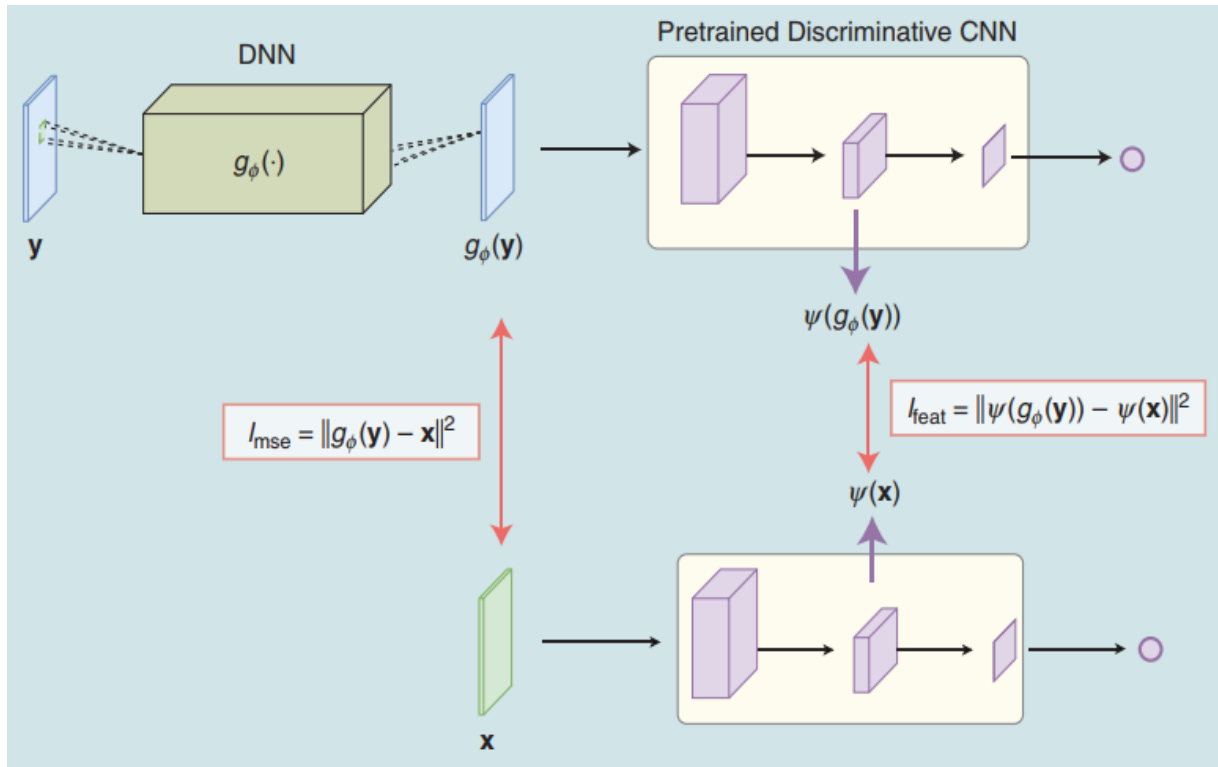
$$l_{mse} = \|g_{\phi}(y) - x\|^2 \quad (2.1)$$

όπου $g_{\phi}(y)$ είναι το αποτέλεσμα του δικτύου και x η αρχική εικόνα.

Σε κάθε βήμα της διαδικασίας εκπαίδευσης, οι παράμετροι z του μοντέλου ενημερώνονται από έναν αλγόριθμο βελτιστοποίησης, όπως είναι ο αλγόριθμος κατάβασης δυναμικού (Gradient Descent). Στους αλγορίθμους βελτιστοποίησης χρειάζεται να οριστούν κάποιοι υπερ παράμετροι όπως είναι ο ρυθμός εκπαίδευσης και το βάρος της κανονικοποίησης (regularization strength). Μόλις το μοντέλο εκπαιδευτεί, έχει μάθει τις βέλτιστες παραμέτρους και την βέλτιστη σχέση συσχέτισης μεταξύ της εισόδου και της εξόδου. Έτσι μπορούμε να εκτιμήσουμε την εικόνα εξόδου $\hat{x}$ από μια παρατηρούμενη εικόνα εισόδου υπολογίζοντας $\hat{x} = g_{\phi}(y)$. Αυτό το απλό και αποτελεσματικό βήμα συμπεράσματος είναι ένα από τα σημαντικότερα πλεονεκτήματα των νευρωνικών δικτύων έναντι των παραδοσιακών αναλυτικών μεθόδων που μπορεί να απαιτούν πολύπλοκες διαδικασίες συμπερασμάτων.

Μια άλλη προσέγγιση σχετικά με τις συναρτήσεις κόστους είναι ότι μπορούμε να έχουμε παραπάνω από μια. Ο λόγος είναι ότι το γεγονός πως η συνάρτηση ελάχιστων τετραγώνων είναι υπολογιστικά ελκυστική, δεν αντικατοπτρίζει πλήρως τη διαφορά στην οπτική ποιότητα μεταξύ δύο εικόνων. Η χρήση της συνάρτησης ελάχιστων τετραγώνων ως της μοναδικής συνάρτησης κόστους για την επίλυση αντίστροφων προβλημάτων δεν μπορεί να διακρίνει επαρκώς μεταξύ των ενδεχομένως πολλαπλών παρόμοιων λύσεων που το μοντέλο θα μπορούσε να παράγει. Έτσι μπορούμε να χρησιμοποιήσουμε την ευκλείδεια απόσταση $\psi(\cdot)$ στον χώρο των πίξελ ώστε να κανονικοποιήσει την λειότητα (oversmoothing) της συνάρτησης ελάχιστων τετραγώνων. Αυτός ο τρόπος παρουσιάζεται γραφικά από το Σχήμα 2.6 και περιγράφεται από την σχέση (2.2) :

$$l_{\text{feat}} = \|\psi(g_\phi(\mathbf{y})) - \psi(\mathbf{x})\|^2 \quad (2.2)$$



Σχήμα 2.6 Χρήση Δύο Συναρτήσεων Κόστους για Καλύτερο Αποτέλεσμα

2.9 Μάθηση Αναπαραστάσεων Υψηλότερου Επιπέδου με την Χρήση Δικτύων CNN Κωδικοποίησης-Αποκωδικοποίησης

Όπως είδαμε και προηγουμένως υπάρχουν δύο είδη συνελικτικών νευρωνικών δικτύων που μπορούμε να χρησιμοποιήσουμε για την επίλυση αντίστροφων προβλημάτων. Το πρώτο είδος αποτελείται από πολλαπλά επίπεδα συνέλιξης που παράγουν χαρακτηριστικά αντιστοίχισης με σταθερό μέγεθος. Ο άλλος τύπος συνελικτικού νευρωνικού δικτύου που έχει αναφερθεί, αποτελείται από μια αρχιτεκτονική κωδικοποίησης - αποκωδικοποίησης, στην οποία το χωρικό μέγεθος των χαρακτηριστικών αντιστοίχισης πρώτα μειώνεται και στη συνέχεια αυξάνεται ώστε να ταιριάζει με το μέγεθος της εικόνας εξόδου. Μια αρχική υπόθεση που μπορεί να γίνει είναι ότι στα αντίστροφα προβλήματα, όπου θέλουμε η εικόνα εισόδου και εξόδου να έχουν ίδιο μέγεθος, είναι μη αναγκαίο να μειώσουμε την διάσταση των χαρακτηριστικών αντιστοίχισης και έπειτα να την ξανά αυξήσουμε.

Όμως οι Johnson και άλλοι [JAF2012] απέδειξαν ότι η αρχιτεκτονική με τον κωδικοποιητή-αποκωδικοποιητή έχει αρκετά πλεονεκτήματα στα αντίστροφα προβλήματα. Αρχικά μειώνοντας τα χαρακτηριστικά αντιστοίχισης σε κάθε βήμα έχουμε λιγότερες αλγεβρικές πράξεις με αποτέλεσμα να βελτιώνεται αποτελεσματικότητα του συμπεράσματος. Δεύτερον, λόγω της μείωσης της διάστασης, το αποτελεσματικό δεκτικό πεδίο του δικτύου αυξάνεται σημαντικά. Με το δεκτικό πεδίο μιας μονάδας σε ένα νευρωνικό δίκτυο αναφερόμαστε στο μέγεθος του πεδίου προβολής που έχει η μονάδα πάνω από το επίπεδο εισόδου της. Στα συνελικτικά νευρωνικά δίκτυα, το δεκτικό πεδίο μιας μονάδας καθορίζεται από το πλάτος k του πυρήνα της συνέλιξης. Η χρήση διαδοχικών συνελίξεων σε ένα συνελικτικό νευρωνικό δίκτυο με κωδικοποιητή - αποκωδικοποιητή αυξάνει το συνολικό δεκτικό πεδίο του μοντέλου. Η έννοια του δεκτικού πεδίου στα συνελικτικά νευρωνικά δίκτυα είναι σημαντικής σημασίας για την επίλυση των αντίστροφων προβλημάτων καθώς ένα μεγάλο οπτικό πεδίο πάνω από την εικόνα εισόδου μπορεί να βελτιώσει σημαντικά την πρόβλεψη σε κάθε πιξελ της εικόνας εξόδου. Έτσι, μπορούμε να χρησιμοποιήσουμε τα συνελικτικά νευρωνικά δίκτυα με κωδικοποιητή-αποκωδικοποιητή όταν πρέπει να επιλύσουμε ένα αντίστροφο πρόβλημα που απαιτεί ένα μεγάλο οπτικό πεδίο. Έτσι μπορούμε να χρησιμοποιήσουμε ένα συνελικτικό δίκτυο με κωδικοποιητή - αποκωδικοποιητή ώστε να αποκτήσουμε μια αντιστοιχία μεταξύ της εικόνας εισόδου και μιας χρήσιμης αναπαράστασης και έπειτα να χρησιμοποιήσουμε αυτήν την αναπαράσταση και μέσω ενός αποκωδικοποιητή να ανακατασκευάσουμε την τελική εικόνα εξόδου.

2.10 Το Μοντέλο SRCNN

Μια σχετικά νέα έρευνα [Don2015] το 2015 έδωσε μία νέα οπτική πάνω στο θέμα. Για την ακρίβεια, οι C.Dong, C.C.Loy, K.He και X.Tang παρουσίασαν ένα μοντέλο το οποίο αποτελείται αποκλειστικά από συνελίξεις καθιστώντας τα συνελικτικά νευρωνικά δίκτυα (CNN) ως την πρώτη επιλογή για κατασκευή τεχνητού νευρωνικού δικτύου που έχει ως στόχο την υπερανάλυση εικόνας.

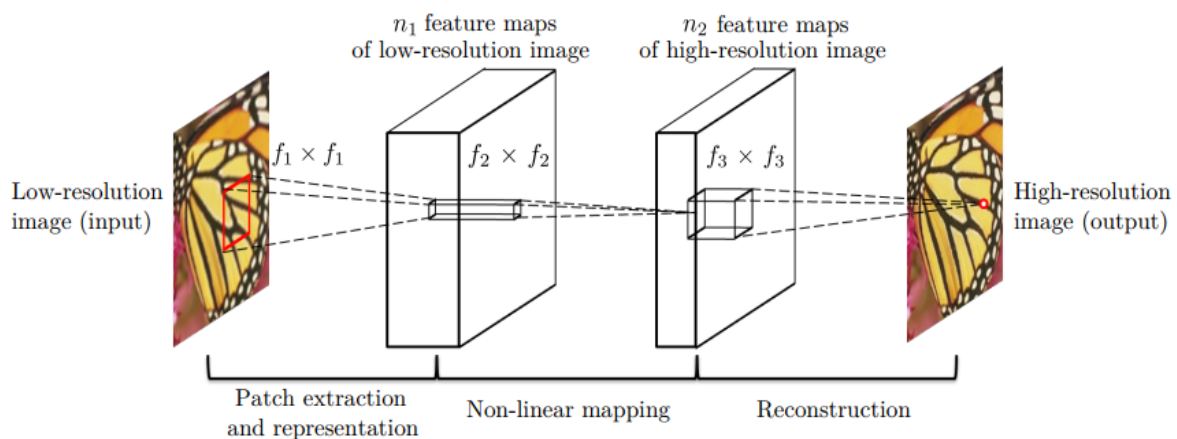
Συγκεκριμένα, αναπτύχθηκε ένα μοντέλο το οποίο είχε μόνο συνελίξεις και πολύ λίγη προ-επεξεργασία (pre processing) και επεξεργασία μετά την υπόθεση του μοντέλου(post processing). Το δίκτυο χωρίζεται σε τρία μέρη. Στο πρώτο μέρος γίνεται η εξαγωγή χαρακτηριστικών από την εικόνα, το δεύτερο μέρος πραγματοποιεί την μη γραμμική απεικόνιση (non- linear mapping) και στο τρίτο μέρος υλοποιείται η ανακατασκευή της εικόνας. Παρακάτω παρουσιάζονται αναλυτικά τα τρία αυτά μέρη και οι διεργασίες που έγιναν:

ΕΞΑΓΩΓΗ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ : Σε αυτό το μέρος εκτελείται η διαδικασία κατά την οποία εξάγονται τα patches από την εικόνα με την χαμηλή ανάλυση. Για να γίνει αυτό εκτελείται μια συνέλιξη μεταξύ της εικόνας εισόδου (χαμηλή ανάλυση) και ενός φίλτρου $W1$. Έπειτα προστίθεται σε αυτό μια πόλωση $B1$ (bias) και τέλος εφαρμόζεται σε αυτό το αποτέλεσμα η συνάρτηση ενεργοποίησης της Ανορθωμένης Γραμμικής Συνάρτησης Ράμπας (ReLU) που έχει αναφερθεί σε προηγούμενο κεφάλαιο. Το φίλτρο έχει μέγεθος πυρήνα $c \times f1 \times f1$ όπου c είναι τα κανάλια της εικόνας εισόδου (1 για της ασπρόμαυρες εικόνες, 3 για την έγχρωμες RGB εικόνες), $f1$ είναι το χωρικό μέγεθος του φίλτρου, ενώ εφαρμόζονται $n1$ τέτοια φίλτρα. Το $B1$ είναι ένας πίνακας διάστασης $n1$. Το αποτέλεσμα αυτού του μέρους είναι $n1$ χαρακτηριστικά απεικόνισης της εικόνας εισόδου. Συγκεντρωτικά το επίπεδο συνοψίζεται σαν τύπος : $F1(Y) = \max(0; W1 * Y + B1)$, με Y την εικόνα εισόδου.

NON – LINEAR MAPPING: Στο δεύτερο μέρος το αποτέλεσμα της προηγούμενης εξαγωγής γίνεται εκ νέου συνέλιξη με ένα φίλτρο $W2$. Έπειτα ακολουθεί η ίδια διαδικασία, δηλαδή προστίθεται σε αυτό μια πόλωση $B2$ και στο τελικό αποτέλεσμα εφαρμόζεται η συνάρτηση ενεργοποίησης ReLU. Εδώ το φίλτρο $W2$ έχει μέγεθος πυρήνα $n1 \times f2 \times f2$, όπου $n1$ είναι το μέγεθος από το προηγούμενο επίπεδο ενώ το $f2$ ισοδυναμεί με 1. Εφαρμόζονται $n2$ τέτοια φίλτρα και αντίστοιχα το $B2$ είναι πίνακας διάστασης $n2$. Μπορούν να εφαρμοστούν πολλά τέτοια επίπεδα με σκοπό να αυξηθεί η μη γραμμικότητα του δικτύου, με κόστος προφανώς ότι χρειάζεται περισσότερος χρόνος για εκπαίδευση. Το $f2$ μπορεί να πάρει άλλη τιμή αλλά πλέον δεν θα μιλάμε για patch της εικόνας αλλά για ένα patch των χαρακτηριστικών της εικόνας. Συγκεντρωτικά το επίπεδο συνοψίζεται σαν τύπος:

$F2(Y) = \max(0; W2 * F1(Y) + B2)$, με Y το αποτέλεσμα του προηγούμενου επιπέδου.

ΑΝΑΚΑΤΑΣΚΕΥΗ ΕΙΚΟΝΑΣ (RECONSTRUCTION): Στο τελευταίο μέρος πραγματοποιείται η ανακατασκευή της εικόνας με καλύτερη πλέον ανάλυση. Όπως και με τα παραπάνω επίπεδα, εκτελείται συνέλιξη με ένα φίλτρο $W3$ με πυρήνα $n2 \times f3 \times f3$, όπου $n2$ το μέγεθος από το προηγούμενο επίπεδο ενώ $f3$ το χωρικό μέγεθος του φίλτρου. Έπειτα στο αποτέλεσμα της συνέλιξης προστίθεται μια πόλωση $B3$ η οποία στην ουσία είναι ένας πίνακας διάστασης c , δηλαδή όσο είναι το νούμερο των καναλιών της εικόνας (1 ή 3). Η διαφορά με τα προηγούμενα επίπεδα είναι ότι εδώ δεν εφαρμόζεται καμία συνάρτηση ενεργοποίησης αλλά επιστρέφεται το αποτέλεσμα της πρόσθεσης όπως φαίνεται και από τον τύπο $F(Y) = W3 * F2(Y) + B3$, με Y το αποτέλεσμα του δεύτερου μέρους. Στο Σχήμα 2.7 εμφανίζεται συνοπτικά η αρχιτεκτονική του δικτύου που χρησιμοποιείται.



Σχήμα 2.7 Το Μοντέλο των C.Dong και C.C.Loy[Don2015]

Με την παραπάνω έρευνα έγινε αντιληπτό ότι τα συνελικτικά νευρωνικά δίκτυα είναι πολύ χρήσιμα στην επεξεργασία εικόνας και στην παραγωγή καλύτερης και πιο ποιοτικής εικόνας. Έπειτα από αυτήν την έρευνα έχουν προταθεί διάφορες παραλλαγές, όμως υπάρχει ένα μεγάλο κοινό σε όλες τις προσπάθειες. Όλες έχουν σαν βάση τα συνελικτικά νευρωνικά δίκτυα (CNN).

ΚΕΦΑΛΑΙΟ 3

ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ ΚΑΙ ΑΝΑΛΥΤΙΚΕΣ ΜΕΘΟΔΟΙ

-
- 3.1 Αραιή αναπαράσταση
 - 3.2 Ο αλγόριθμος ISTA
 - 3.3 Εφαρμογή του αλγορίθμου LISTA στην υπερανάλυση εικόνας
 - 3.4 Συνελικτική προσεγγιστική αραιή αναπαράσταση
-

Όπως αναφέρθηκε παραπάνω η επιλογή της αρχιτεκτονικής που θα χρησιμοποιήσουμε παίζει σημαντικό ρόλο στις λειτουργίες που θα μάθει το μοντέλο. Ένα ισχυρό πλεονέκτημα των αναλυτικών μεθόδων αντί των νευρωνικών δικτύων είναι ότι μπορούμε να ενσωματώσουμε τις γνώσεις μας σχετικά με τις άγνωστες ποσότητες στη διαδικασία λύσης κατά την ανάπτυξη αναλυτικών μεθόδων, ενώ στα νευρωνικά δίκτυα έχουμε πολύ μικρό έλεγχο. Για αυτόν τον λόγο ερευνώνται επιλογές ώστε να ενσωματώσουμε στα νευρωνικά δίκτυα λειτουργίες μάθησης που υλοποιούνται στις αναλυτικές μεθόδους. Ένα τέτοιο παράδειγμα είναι ο αλγόριθμος LISTA [GrL2010], ο οποίος είναι [HRW2014] μια απλή περίπτωση ενός εξελισσόμενου αλγορίθμου που έχει σκοπό τον συνδυασμό των πλεονεκτημάτων τόσο των αναλυτικών μεθόδων όσο και των νευρωνικών δικτύων. Η βασική ιδέα είναι ότι ξεκινάει με μια αναλυτική προσέγγιση και έναν σχετικό αλγόριθμο συμπερασμάτων και έπειτα ξεδιπλώνει τις επαναλήψεις του

αλγόριθμου ως επίπεδα σε ένα βαθύ δίκτυο. Αφού καθοριστεί το μέγεθος του δικτύου, το δίκτυο εκπαιδεύεται για να εκτελεί ακριβή συμπεράσματα. Πιο συγκεκριμένα, ο αλγόριθμος LISTA αποτελεί μια προσεγγιστική μέθοδο για τον υπολογισμό αραιών αναπαραστάσεων μέσω ενός πολυ-επίπεδου νευρωνικού δικτύου. Ο Wang και άλλοι [Wan2015], χρησιμοποίησαν αυτήν την τεχνική ώστε να συνδυάσουν συνελικτικά νευρωνικά δίκτυα και αραιή αναπαράσταση για πρόβλημα της υπερανάλυσης εικόνας και κατάφεραν να έχουν ικανοποιητικά αποτελέσματα. Για τον αλγόριθμο LISTA και τις αραιές αναπαραστάσεις θα αναφερθούμε αναλυτικότερα αργότερα.

3.1 Αραιή Αναπαράσταση

Η αραιή αναπαράσταση (Sparse Representation) ή η αραιή κωδικοποίηση (Sparse Coding) [Tso2017] [Ela2010], αποτελεί μία διαδικασία μη εποπτευόμενης μάθησης και χρησιμοποιείται ευρέως στη μηχανική μάθηση, τη νευροεπιστήμη, την επεξεργασία σημάτων και την στατιστική, καθώς αποτελεί ένα ερευνητικό πεδίο με αξιοσημείωτα αποτελέσματα τόσο σε απλά προβλήματα όπως απομάκρυνσης θορύβου, σύνθεση υφής, επεξεργασίας ήχου κ.α., όσο και σε πιο σύνθετα όπως προβλήματα ταξινόμησης και ομαδοποίησης. Σκοπός της είναι η εύρεση ‘αραιών’ λύσεων σε γραμμικά συστήματα.

Στην αριθμητική ανάλυση ένας πίνακας ονομάζεται αραιός όταν τα περισσότερα στοιχεία του είναι μηδέν, δηλαδή όταν το αποτέλεσμα της διαίρεσης μεταξύ των μηδενικών στοιχείων του πίνακα και των συνολικών στοιχείων του πίνακα είναι μεγαλύτερο του 0.5 [BuR2014]. Επομένως με τον όρο αραιότητα αναφερόμαστε στην αναπαράσταση μιας πληροφορίας με όσο το δυνατόν λιγότερους μη μηδενικούς συντελεστές.

Μία μορφή του προβλήματος της αραιής κωδικοποίησης αφορά την εύρεση της καλύτερης αναπαράστασης –με βάση το σφάλμα ανακατασκευής– ενός δεδομένου σήματος εισόδου ως γραμμικού συνδυασμού ενός περιορισμένου αριθμού σημάτων ενός λεξικού (dictionary), που ονομάζονται άτομα (atoms).

Στις αραιές αναπαραστάσεις χρησιμοποιούνται γραμμικά συστήματα στα οποία οι άγνωστοι συντελεστές είναι περισσότεροι από ότι οι εξισώσεις. Πιο

συγκεκριμένα [Ela2010][SMF2015] [RIG2014], για ένα σήμα y μήκους m , το οποίο παρουσιάζεται σε μορφή πίνακα $y \in R^m$, ορίζεται η αναπαράσταση $a \in R^N$ του y ως προς ένα υπερπλήρες λεξικό $D \in R^{m \times N}$, με $m < N$, σύμφωνα με την σχέση (3.1).

$$y = Da \quad (3.1)$$

Το γραμμικό σύστημα που περιγράφεται από τη σχέση (3.1) μπορεί να έχει καμία ή άπειρες λύσεις. Θεωρούμε ότι $\text{rank}(D) = m$ ώστε να αποκλείσουμε την περίπτωση της μη ύπαρξης λύσης. Ωστόσο, για να οδηγηθούμε σε μοναδική λύση θα πρέπει να επιβάλλουμε επιπλέον περιορισμούς. Ένας τέτοιος περιορισμός είναι η αραιότητα της λύσης. Δηλαδή αναζητούμε μια λύση με λίγους μη μηδενικούς συντελεστές. Μαθηματικά ο περιορισμός αυτός εκφράζεται μέσω της ψευδο-νόρμας ℓ_0 η οποία ορίζεται ως το πλήθος των μη μηδενικών στοιχείων ενός διανύσματος. Συνεπώς, οδηγούμαστε στο πρόβλημα ελαχιστοποίησης:

$$\min_a \frac{1}{2} \|Da - y\|_2^2 + \lambda \|a\|_0 \quad (3.2)$$

με λ κατάλληλη σταθερά. Το πρόβλημα αυτό είναι NP-hard. Ένα ισοδύναμο πρόβλημα μπορεί να προκύψει αντικαθιστώντας την νόρμα ℓ_0 με την νόρμα ℓ_1 [CaP2009]. Έτσι οδηγούμαστε στο πρόβλημα ελαχιστοποίησης :

$$\min_a \frac{1}{2} \|Da - y\|_2^2 + \lambda \|a\|_1 \quad (3.3)$$

όπου $\|a\|_1 = \sum_{i=1}^N |a_i|$ είναι η ℓ_1 -νόρμα που επιβάλλει αραιότητα στην λύση.

Για την σχέση (3.3) έχουν αναπτυχθεί πολλοί αριθμητικοί αλγόριθμοι επίλυσης [TrW2010]. Το βασικό τους μειονέκτημα ωστόσο είναι το υψηλό υπολογιστικό κόστος.

3.1.1 Ο αλγόριθμος ISTA

Η σχέση (3.3) για τον προσδιορισμό μιας αραιής αναπαράστασης μπορεί να γραφεί και με τον γενικό τύπο βελτιστοποίησης προβλημάτων:

$$\min_a f(a) + \lambda g(a) \quad (3.4)$$

όπου $f(a) = \frac{1}{2} \|Da - y\|_2^2$ και $g(a) = \lambda \|a\|_1$. Μια κατηγορία αλγορίθμων για την επίλυση της σχέσης (3.4) είναι οι αλγόριθμοι proximal. Οι αλγόριθμοι αυτοί βασίζονται στον τελεστή proximal που ορίζεται από τη σχέση (3.5) :

$$\text{Prox}_{\mu g}(a) = \arg \min_v \left\{ \frac{1}{2} \|v - a\|^2 + \theta \|v\|_1 \right\} \quad (3.5)$$

με $\theta = \frac{\lambda}{L}$, όπου L κατάλληλη σταθερά [PNS2014].

Ο τελεστής proximal για την σχέση (3.3) δίνεται από τη σχέση (3.6):

$$\text{Prox}_{\mu g}(a_i) = [h_\theta(a)]_i = \text{sign}(a_i) \cdot \max(|a_i| - \theta, 0) \quad (3.6)$$

Σε κάθε επανάληψη οι proximal μέθοδοι πραγματοποιούν τον εξής υπολογισμό [Bac2011]

$$a^{t+1} = \text{Prox}_{\frac{\lambda}{L}g} \left(a^t - \frac{1}{L} \nabla f(a^t) \right) \quad (3.7)$$

Από τη σχέση (3.2) έχουμε $\nabla f(a) = D^T(Da - y)$, άρα η σχέση (3.7) μπορεί να μετατραπεί στην σχέση (3.8), δηλαδή στον γνωστό αλγόριθμο ISTA (Iterative Soft Thresholding Algorithm) όπως αυτός έχει περιγραφεί στο [CoW2005] .

$$a^{t+1} = h_\theta \left(a^t - \frac{1}{L} D^T (Da^t - y) \right) \quad (3.8)$$

3.1.2 Το μοντέλο LISTA

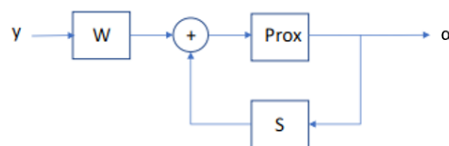
Μελέτες πάνω στον αλγόριθμο ISTA και την αραιή αναπαράσταση οδήγησαν την είσοδο του αλγορίθμου ISTA στα νευρωνικά δίκτυα. Πιο συγκεκριμένα, η σχέση (3.8) μπορεί να πάρει την μορφή της σχέσης (3.9):

$$\begin{aligned} a^{t+1} &= h_\theta \left(a^t - \frac{1}{L} D^T (Da^t - y) \right) \\ &= h_\theta \left(a^t - \frac{1}{L} D^T Da^t + \frac{1}{L} D^T y \right) \\ &= h_\theta \left(I - \frac{1}{L} D^T D a^t + \frac{1}{L} D^T y \right) \end{aligned} \quad (3.9)$$

Αν θέσουμε $S = I - \frac{1}{L} D^T D$, $W = \frac{1}{L} D^T$ τότε καταλήγουμε στην σχέση (3.10):

$$a^{t+1} = h_\theta (Sa^t + Wy) \quad (3.10)$$

Η σχέση (3.10) υλοποιήθηκε από τους K. Gregor και Y. Le Cun [GrL2010] ως ένα feed forward νευρωνικό δίκτυο και ως ένα αναδρομικό (recurrent) νευρωνικό δίκτυο. Μια σχηματική αναπαράσταση αυτών των νευρωνικών δικτύων παρουσιάζεται στο Σχήμα 3.1. Εκπαιδεύοντας το νευρωνικό δίκτυο μπορεί να επιτευχθεί η μάθηση ενός τελεστή κωδικοποίησης ο οποίος μπορεί να αντιστοιχεί ένα διάνυσμα εισόδου y στην αντίστοιχη αραιή αναπαράστασή του α . Ο αλγόριθμος που λαμβάνεται με αυτόν τον τρόπο αναφέρεται ως LISTA (Learned ISTA). Όπως θα αναφερθεί στο επόμενο υποκεφάλαιο η εκπαίδευση ενός δικτύου που χρησιμοποιεί τον αλγόριθμο LISTA μπορεί να χρησιμοποιηθεί για υπερανάλυση εικόνας.



Σχήμα 3.1 Δίκτυο LISTA

3.2 Εφαρμογή του αλγορίθμου LISTA στην υπερανάλυση εικόνας

Λαμβάνοντας υπόψιν ότι η αραιή αναπαράσταση μπορεί να υλοποιηθεί αποτελεσματικά με την βοήθεια νευρωνικών δικτύων μέσω του αλγορίθμου LISTA, έγιναν έρευνες για την δημιουργία ενός πολυ-επίπεδου νευρωνικού δικτύου το οποίο θα χρησιμοποιεί την αραιή αναπαράσταση για υπερανάλυση εικόνας [Yan2010]. Μια τέτοια έρευνα πραγματοποιήθηκε από τους Z. Wang, D. Liu, J. Yang, W. Han, T. Huang [Wan2015], οι οποίοι συνδύασαν τα συνελκτικά νευρωνικά δίκτυα και την αραιή αναπαράσταση για υπερανάλυση εικόνας. Η μέθοδος αυτή αποτελεί την βάση την παρούσας εργασίας.

Όπως οι περισσότεροι μέθοδοι για υπερανάλυση εικόνας έτσι και το συγκεκριμένο μοντέλο δέχεται σαν είσοδο μια εικόνα χαμηλής ανάλυσης I_y και σαν έξοδο του δικτύου παράγει την ίδια εικόνα I_x με καλύτερη ανάλυση. Ειδικότερα, η εικόνα εισόδου I_y λαμβάνεται μετά από εφαρμογή κυβικής παρεμβολής σε δύο διαστάσεις στην δοθείσα εικόνα χαμηλής ανάλυσης. Το

νευρωνικό δίκτυο για την υπερανάλυση εικόνας χωρίζεται στα τρία παρακάτω μέρη:

ΕΞΑΓΩΓΗ ΤΩΝ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΤΗΣ ΕΙΚΟΝΑΣ ΜΕ ΤΗΝ ΧΑΜΗΛΗ ΑΝΑΛΥΣΗ: Για την εξαγωγή των χαρακτηριστικών χρησιμοποιείται ένα επίπεδο συνέλιξης (convolution layer) Η πιο συγκεκριμένα πραγματοποιείται συνέλιξη της εικόνας I_y (LR) με m_y φίλτρα χωρικού μεγέθους $S_y \times S_y$. Επιπλέον σαν φίλτρο εισόδου χρησιμοποιείται ο αριθμός των καναλιών της εικόνας. Με αυτόν τον τρόπο δημιουργούνται patches εισόδου με χωρικό μεγέθους $S_y \times S_y$ των οποίων τα χαρακτηριστικά αναπαράστασης y θα έχουν μέγεθος m_y .

ΔΙΚΤΥΟ LISTA: Τα χαρακτηριστικά αναπαράστασης y που δημιουργήθηκαν από το προηγούμενο επίπεδο χρησιμοποιούνται σαν είσοδος σε αυτό το επίπεδο, δηλαδή στο επίπεδο όπου υλοποιείται το δίκτυο LISTA. Ο αλγόριθμος LISTA που αναφέρθηκε στην ενότητα 3.2.2 εκτελείται επαναληπτικά για t επίπεδα ώστε να υπολογίσει την αραιή αναπαράσταση $a \in R^n$ για κάθε χαρακτηριστικό. Ο αριθμός t καθορίζεται πριν την εκπαίδευση του δικτύου και δεν μεταβάλλεται κατά την διάρκεια της. Πιο συγκεκριμένα τα χαρακτηριστικά αναπαράστασης πολλαπλασιάζονται με ένα βάρος $W \in R^{n \times m_y}$, το οποίο εκπαιδεύεται ταυτόχρονα με το δίκτυο, και έπειτα το αποτέλεσμα δίνεται ως όρισμα σε μια συνάρτηση ενεργοποίησης h_θ :

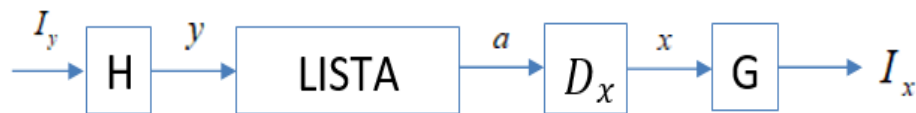
$$[h_\theta(\mathbf{a})]_i = \text{sign}(a_i) \cdot \max(|a_i| - \theta, 0) \quad (3.11)$$

όπου $\theta \in R^n$ το κατώφλι ενεργοποίησης, το οποίο εκπαιδεύεται και αυτό ταυτόχρονα με το δίκτυο, και a_i είναι το i -οστό στοιχείο του διανύσματος a . Το αποτέλεσμα της συνάρτησης ενεργοποίησης αποτελεί την αραιή αναπαράσταση των χαρακτηριστικών. Αυτό συμβαίνει όταν $t=1$. Στην περίπτωση που $t>1$ επαναλαμβάνεται η διαδικασία του πολλαπλασιασμού και της συνάρτησης ενεργοποίησης με την διαφορά ότι μετά την πρώτη ολοκλήρωση του αλγορίθμου στην θέση των χαρακτηριστικών χρησιμοποιείται η $t-1$ αραιή αναπαράσταση που έχει υπολογιστεί. Τέλος, αφού έχουν ολοκληρωθεί οι t επαναλήψεις του αλγορίθμου, η αραιή αναπαράσταση a που έχει υπολογιστεί πολλαπλασιάζεται

με έναν πίνακα που αποτελεί το λεξικό $D_x \in R^{m_x \times n}$, το οποίο θα χρησιμοποιηθεί για την αντιστοίχιση της αραιής αναπαράστασης με την εικόνα υψηλότερης ανάλυσης. Ο πολλαπλασιασμός δημιουργεί τα patches x μεγέθους $S_x \times S_x = m_x$ για την ανακατασκευή της εικόνας με υψηλή ανάλυση.

ΑΝΑΚΑΤΑΣΚΕΥΗ ΕΙΚΟΝΑΣ ΜΕ ΥΨΗΛΗ ΑΝΑΛΥΣΗ (RECONSTRUCTION):

Η ανακατασκευή της εικόνας με υψηλότερη ανάλυση υλοποιείται μέσω ενός επιπέδου συνέλιξης G . Ειδικότερα, τα ανακτημένα patches x του προηγούμενου επιπέδου εκτελούν συνέλιξη με φίλτρο που έχει m_x κανάλια και χωρικό μέγεθος $S_g \times S_g$. Η συνέλιξη αυτή θα τοποθετήσει τα ανακτημένα patches όλα μαζί στις κατάλληλες θέσεις ώστε να παραχθεί η εικόνα I_x με την υψηλή ανάλυση (HR). Η εικόνα αυτή αποτελεί και την τελική έξοδο του δικτύου. Η συνοπτική περιγραφή της μεθόδου παρουσιάζεται στο Σχήμα 3.2.



Σχήμα 3.2 Υπερανάλυση Εικόνας με το Δίκτυο LISTA

3.3 Συνελικτική Προσεγγιστική Αραιή Αναπαράσταση

3.3.1 Συνελικτική Αραιή Κωδικοποίηση

Λόγω του μεγάλου μεγέθους των εικόνων, οι αραιές τεχνικές μοντελοποίησης εφαρμόζονται συνήθως σε κομμάτια από τις εικόνες και όχι σε ολόκληρες τις εικόνες. Όμως καθώς οι υπολογισμοί γίνονται σε ανεξάρτητα κομμάτια αυτό έχει σαν αποτέλεσμα να μην λαμβάνεται υπόψιν η συνοχή των εικονοστοιχείων στα επικαλυπτόμενα κομμάτια [Gu2015]. Η συνελικτική αραιή κωδικοποίηση - Approximate Convolutional Sparse Coding (ACSC) [Zei2010] είναι μια εναλλακτική προσέγγιση, η οποία μπορεί να εφαρμοστεί άμεσα σε ολόκληρη την εικόνα. Έστω

$\mathbf{Y} \in \mathbb{R}^{n_1 \times n_2}$ η εικόνα εισόδου μας, τότε η συνελικτική αραιή κωδικοποίηση μπορεί να υπολογιστεί λύνοντας το πρόβλημα (3.12) :

$$\min_{\mathbf{U}_i} \frac{1}{2} \left\| \mathbf{Y} - \sum_{i=1}^k \mathbf{D}_i * \mathbf{U}_i \right\|_F^2 + \lambda \sum_{i=1}^k \|\mathbf{U}_i\|_1 \quad (3.12)$$

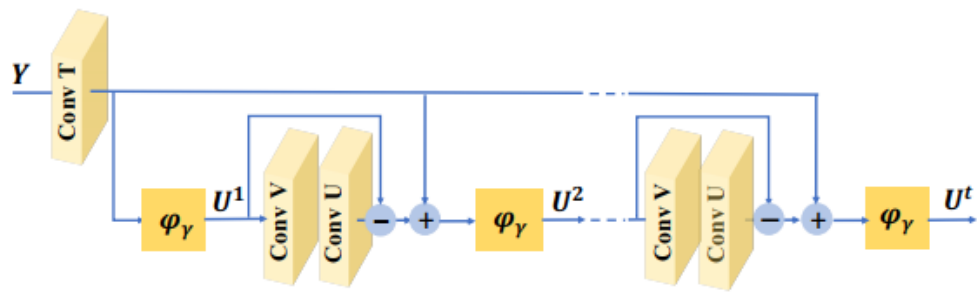
όπου $\|\mathbf{A}\|_F = \sqrt{\sum_i \sum_j |a_{ij}|^2}$ είναι η νόρμα Forbenius, $\mathbf{D}_i \in \mathbb{R}^{p_1 \times p_2}$, $i=1, \dots, k$ είναι τα άτομα ενός συνελικτικού λεξικού $\mathbf{D} \in \mathbb{R}^{p_1 \times p_2 \times k}$ και $\mathbf{U}_i \in \mathbb{R}^{n_1 \times n_2}$, $i=1, \dots, k$ είναι τα αραιά χαρακτηριστικά αντιστοίχισης συναφή με το $\mathbf{D}$. Η νόρμα ℓ_1 υπολογίζει το άθροισμα των απόλυτων τιμών των στοιχείων του $\mathbf{U}_i$. Αποδοτικές λύσεις για το πρόβλημα της σχέσης (3.12) δίνονται στις έρευνες [Woh2014] και [HHW2015].

3.3.2 Το μοντέλο ACSC

Οι αναλυτικές προσεγγίσεις για αραιή προσέγγιση είναι συνήθως εξοπλισμένες με θεωρητικές εγγυήσεις με αποτέλεσμα να έχουμε υψηλό υπολογιστικό κόστος. Όπως αναφέρθηκε και στο κεφάλαιο 3.2.1, για την αντιμετώπιση της μεγάλης πολυπλοκότητας της αριθμητικής βελτιστοποίησης έχει προταθεί ο αλγόριθμος ISTA δηλαδή η σχέση (3.8). Όπως έχει αναλυθεί ο αλγόριθμος ISTA μπορεί να γραφεί σε μορφή νευρωνικού δικτύου, οδηγώντας μας στον αλγόριθμο LISTA δηλαδή στην σχέση (3.10). Σε επόμενο επίπεδο, ένα συνελικτικό LISTA δίκτυο σχεδιασμένο για τον υπολογισμό της συνελικτικής αραιής κωδικοποίησης προτάθηκε στην έρευνα [SrG2018]. Ένα τέτοιο δίκτυο (ACSC) παρουσιάζεται γραφικά από το Σχήμα 3.3 και περιγράφεται από την σχέση (3.13) :

$$\mathbf{U}^t = \phi_\gamma \left(\mathbf{U}^{t-1} - \mathcal{T} * \mathcal{V} * \mathbf{U}^{t-1} + \mathcal{G} * \mathbf{Y} \right) \quad (3.13)$$

όπου ϕ_γ ένας proximal τελεστής που δίνεται από την σχέση (3.6) και $\mathcal{T}$, $\mathcal{V}$, $\mathcal{G}$ κατάλληλες παράμετροι που ορίζονται παρόμοια με το LISTA.



Σχήμα 3.3 Το Μοντέλο ACSC

ΚΕΦΑΛΑΙΟ 4

ΜΟΝΤΕΛΑ ΠΟΛΛΑΠΛΩΝ ΤΥΠΩΝ (MULTIMODAL MODEL)

-
- 4.1 Εισαγωγή
 - 4.2 Βαθιά μάθηση για αντίστροφη προβλήματα με χρήση πλευρικής πληροφορίας
 - 4.3 Συνελικτική αραιή κωδικοποίηση για υπερανάλυση εικόνας με χρήση εικόνων πολλαπλών τύπων
 - 4.4 Υπερανάλυση εικόνας με το δίκτυο LMCSC
-

4.1 Εισαγωγή

Στα γραμμικά αντίστροφα προβλήματα στόχος μας είναι η ανάκτηση ενός σήματος εξόδου από ένα σήμα εισόδου με θόρυβο. Συνήθως, η ανάκτηση βασίζεται σε πολύπλοκες μεθόδους αριθμητικής βελτιστοποίησης. Η πρόσφατες προσεγγίσεις εκτελούν την ανάπτυξη ενός αριθμητικού αλγορίθμου σε μια μορφή νευρωνικού δικτύου, με αποτέλεσμα τη σημαντική μείωση της υπολογιστικής πολυπλοκότητας [TsD2019], όπως έχουμε δει και προηγουμένως. Μια νέα ιδέα σχετικά με την ανάκτηση του σήματος είναι ότι μπορούμε να χρησιμοποιήσουμε μια πληροφορία σήματος που σχετίζεται με το σήμα εισόδου, την λεγόμενη πλευρική πληροφορία (Side information –SI) και να αναπτύξουμε ένα βαθύ

μοντέλο που ενσωματώνει το SI. Ένα τέτοιο μοντέλο, το οποίο θα αναλυθεί παρακάτω στο κεφάλαιο, προτάθηκε από τους E.Tsiligianni, I.Marinavi, B. Cornelis και N.Deligianis [Mar2019] το οποίο δέχεται σαν είσοδο σχεδόν υπέρυθρες (near infrared – NIR) εικόνες και σαν βοηθητικό σήμα χρησιμοποιούνται πολύχρωμες (RGB) εικόνες.

4.2 Βαθιά Μάθηση για Αντίστροφη Προβλήματα με Χρήση Πλευρικής Πληροφορίας

Σε αυτήν την υπο ενότητα θα αναλύσουμε το μοντέλο που προτάθηκε από τους E.Tsiligianni, I.Marinavi, B. Cornelis και N.Deligianis [Mar2019], καθώς αποτελεί την βάση για το δίκτυο που υλοποιήθηκε σε αυτήν την εργασία.

Αρχικά, εκτός από τις παρατηρήσεις του σήματος-στόχου y , έχουμε επίσης πρόσβαση σε πλευρική πληροφορία SI, δηλαδή ένα σήμα z που σχετίζεται με το άγνωστο x . Υποθέτουμε ότι τα $x \in \mathbb{R}^n$ και $z \in \mathbb{R}^d$ έχουν παρόμοιες αραιές αναπαραστάσεις $\alpha \in \mathbb{R}^k$, $w \in \mathbb{R}^k$ ως προς τα λεξικά $D_x \in \mathbb{R}^{n \times k}$, $D_z \in \mathbb{R}^{d \times k}$ με $n \leq k$, $d \leq k$ αντίστοιχα. Ειδικότερα, υποθέτουμε ότι α και w είναι όμοια υπό τον κανόνα της ℓ_1 -νόρμας έτσι ώστε το $\|\alpha - w\|_1$ να είναι μικρό. Η υπόθεση αυτή ισχύει για συζευγμένες αραιές αναπαραστάσεις, δηλαδή για τις παραστάσεις με μερικώς κοινό φορέα (support) και ορισμένους παρόμοιους μη μηδενικούς συντελεστές. Έπειτα το α μπορεί να υπολογιστεί από το πρόβλημα ελαχιστοποίησης $\ell_1 - \ell_1$:

$$\min_{\alpha} \frac{1}{2} \|\Phi D_x \alpha - y\|_2^2 + \lambda (\|\alpha\|_1 + \|\alpha - w\|_1) \quad (4.1)$$

Στις εφαρμογές πολλαπλών τύπων, η εύρεση των συζευγμένων αραιών αναπαραστάσεων περιλαμβάνει τη μάθηση λεξικών και πολύπλοκες μεθόδους βελτιστοποίησης. Το μοντέλο αυτού του κεφαλαίου αποτελεί μια αποτελεσματική προσέγγιση βασισμένη σε ένα νέο μοντέλο πολλαπλών τύπων βαθιάς εξέλιξης για την εκμάθηση συζευγμένων αναπαραστάσεων του σήματος-στόχου και του SI, καθώς και για την ανακατασκευή με SI. Η προσέγγισή αυτή βασίστηκε σε έναν προσεγγιστικό αλγόριθμο για τη λύση της σχέσης (4.1).

4.2.1 Αραιή προσέγγιση με πλευρική πληροφορία με τη μέθοδο deep unfolding

Η σχέση (4.1) είναι της μορφής $f(\alpha) = \frac{1}{2} \|F\alpha - y\|_2^2$, $F := \Phi D_x, F \in \mathbb{R}^{m \times k}$, και $g(\alpha) = \|\alpha\|_1 + \|\alpha - w\|_1$. Ο τελεστής proximal για την g μπορεί να οριστεί από την σχέση (4.2).

$$\xi_\mu(u) = \arg \min_v \left\{ \frac{1}{2} \|v - u\|_2^2 + \mu (\|v\|_1 + \|v - w\|_1) \right\} \quad (4.2)$$

όπου $\mu = \frac{\lambda}{L}$ και $L > 0$ είναι ένα ανώτερο όριο στη σταθερά Lipschitz της ∇f .

Στην σχέση (4.2) όλοι οι όροι είναι διαχωρίσιμοι επομένως όπως έχει αποδειχθεί [TsD2019]:

Για $w_i \geq 0, i = 1, \dots, k$:

$$\xi_\mu(u_i) = \begin{cases} u_i + 2\mu, & u_i < -2\mu \\ 0, & -2\mu \leq u_i \leq 0 \\ u_i, & 0 < u_i < w_i \\ w_i, & w_i \leq u_i \leq w_i + 2\mu \\ u_i - 2\mu, & u_i > w_i + 2\mu \end{cases} \quad (4.3)$$

Για $w_i < 0, i = 1, \dots, k$:

$$\xi_\mu(u_i) = \begin{cases} u_i + 2\mu, & u_i < w_i - 2\mu \\ w_i, & w_i - 2\mu \leq u_i \leq w_i \\ u_i, & w_i < u_i < 0 \\ 0, & 0 \leq u_i \leq 2\mu \\ u_i - 2\mu, & u_i > 2\mu \end{cases} \quad (4.4)$$

Θέτοντας $\nabla f(\alpha) = F^\top (F\alpha - y)$ η σχέση (4.2) μπορεί να πάρει την μορφή της σχέσης (4.5):

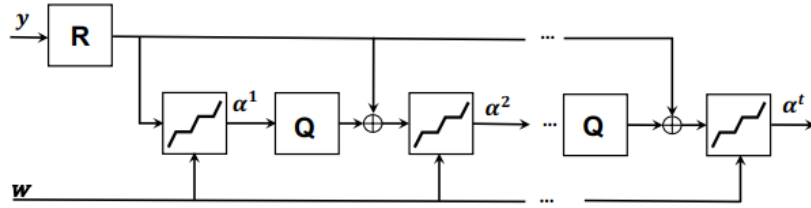
$$\alpha^t = \xi_\mu \left(\alpha^{t-1} - \frac{1}{L} F^\top (F\alpha^{t-1} - y) \right), \quad \alpha^0 = 0 \quad (4.5)$$

η οποία ονομάζεται Side-Information-driven iterative soft Thresholding Algorithm (SITA).

Η σχέση (4.5) μπορεί να μετατραπεί σε μορφή νευρωνικού δικτύου αν αντικαταστήσουμε $Q = I - \frac{1}{L} F^\top F$, $R = \frac{1}{L} F^\top$, οδηγώντας μας έτσι στην σχέση (4.6):

$$\alpha^t = \xi_\mu(Q\alpha^{t-1} + Ry) \quad (4.6)$$

Η σχέση (4.6) είναι παρόμοια με αυτήν του αλγορίθμου LISTA (σχέση (3.10)) με την διαφορά ότι έχουν διαφορετικούς proximal τελεστές. Μπορεί να υλοποιηθεί από ένα recurrent ή feed forward νευρωνικό δίκτυο, με τους proximal συντελεστές των σχέσεων (4.3) και (4.4) να χρησιμοποιούνται ως μη γραμμικές συναρτήσεις ενεργοποίησης που ενσωματώνουν την πλευρική πληροφορία SI. Ένα τέτοιο δίκτυο ονομάζεται LeSITA και παρουσιάζεται γραφικά στην Σχήμα 4.1.



Σχήμα 4.1 Το Δίκτυο LeSITA

Το δίκτυο μπορεί να εκπαιδευτεί χρησιμοποιώντας J ζεύγη αραιών κωδικών $\{\alpha_{(j)}, w_{(j)}\}_{j=1}^J$ που αντιστοιχούν σε J αντίστοιχα αραιά σήματα $\{x_{(j)}, z_{(j)}\}_{j=1}^J$ και μια συνάρτηση κόστους της μορφής :

$$\mathcal{L} = \left\| \sum_{j=1}^J \alpha_{(j)} - \hat{\alpha}_{(j)} \right\|_2^2 \quad (4.7)$$

όπου $\hat{\alpha}$ είναι η εκτίμηση εξόδου. Η μάθηση έχει ως αποτέλεσμα έναν γρήγορο αραιό τελεστή προσέγγισης που αντιστοιχίζει απευθείας το διάνυσμα παρατήρησης εισόδου y σε έναν αραιό κώδικα α με τη βοήθεια της πλευρικής πληροφορίας w .

4.2.2 LeSITA για ανακατασκευή σήματος με πλευρική πληροφορία

Μπορεί να δημιουργηθεί ένα δίκτυο LeSITA που χρησιμοποιεί πλευρική πληροφορία για την ανακατασκευή ενός σήματος όπως φαίνεται γραφικά και στην Εικόνα 11. Στον κύριο κλάδο, ένας κωδικοποιητής LeSITA υπολογίζει μια κρυμμένη αναπαράσταση α του διανύσματος παρατήρησης y που υπολογίζεται από την σχέση (4.8):

$$y = \Phi x + e \quad (4.8)$$

σύμφωνα με την σχέση (4.6). Για την ανακατασκευή του άγνωστου σήματος χρησιμοποιείται ένας γραμμικός αποκωδικοποιητής, δηλαδή, $\hat{x} = Da$ με $D \in \mathbb{R}^{n \times k}$ ένα εκπαιδευσιμο λεξικό. Ο ρόλος του κλάδου που περιέχει το δίκτυο με την πλευρική πληροφορία είναι η βελτίωση της διαδικασίας κωδικοποίησης παρέχοντας μια εκ των προτέρων γνώση. Το μοντέλο εκπαιδεύεται με ℓ_2 συνάρτηση κόστους $\mathcal{L} = \left\| \sum_{j=1}^J x_{(j)} - \hat{x}_{(j)} \right\|_2^2$ με $x_{(j)}$ να είναι το j -ιστό δείγμα σήματος εισόδου και $\hat{x}_{(j)}$ η αντίστοιχη εκτίμηση εξόδου.

4.3 Συνελικτική Αραιή Κωδικοποίηση για Υπερανάλυση Εικόνας με Χρήση Εικόνων Πολλαπλών Τύπων

Το πρόβλημα της εύρεσης συνελικτικών αραιών κωδικών μιας εικόνας προορισμού $Y \in \mathbb{R}^{n_1 \times n_2}$ με τη βοήθεια μιας εικόνας άλλου τύπου $\Omega \in \mathbb{R}^{n_1 \times n_2}$ μπορεί να γραφεί με την μορφή της σχέσης (4.9) :

$$\min_{\alpha} \frac{1}{2} \left\| Y - \sum_{i=1}^k d_i * \alpha_i \right\|_2^2 + \lambda \left(\sum_{i=1}^k \|\alpha_i\|_1 + \sum_{i=1}^k \|\alpha_i - z_i\|_1 \right) \quad (4.9)$$

όπου $\alpha_i \in \mathbb{R}^{n_1 \times n_2}, i=1, \dots, k$ είναι τα αραιά χαρακτηριστικά αντιστοίχισης της εικόνας Y σε συσχέτιση με ένα λεξικό $D \in \mathbb{R}^{p_1 \times p_2 \times k}$ με άτομα $d_i \in \mathbb{R}^{p_1 \times p_2}$, και $z_i \in \mathbb{R}^{n_1 \times n_2}, i=1, \dots, k$ είναι τα αραιά χαρακτηριστικά αντιστοίχισης της πλευρικής πληροφορίας Ω σε συσχέτιση με ένα λεξικό με άτομα b_i τέτοια ώστε $\Omega = \sum_{i=1}^k b_i * z_i$.

Η σχέση (4.9) μπορεί να γραφτεί στην μορφή της σχέσης (3.12) χάρη στην

γραμμικές ιδιότητες της συνέλιξης. Έτσι έχουμε ένα αραιό λεξικό $A \in \mathbb{R}^{(n_1-p_1+1)(n_2-p_2+1) \times kn_1n_2}$ ως συνένωση των πινάκων Toeplitz που ξετυλίγει τα d_i , $x \in \mathbb{R}^{kn_1n_2}$ και $s \in \mathbb{R}^{kn_1n_2}$ είναι οι διανυσματικές μορφές των αραιών χαρακτηριστικών αντιστοίχισης των εικόνων στόχου και της πλευρικής πληροφορίας, αντίστοιχα. Έτσι μια καλή προσέγγιση λύσης της σχέσης (4.9) είναι η λύση του proximal αλγορίθμου που δίνεται από την σχέση (4.10) :

$$x^{t+1} = \xi_\mu \left(x^t - \frac{1}{L} A^T A x^t + \frac{1}{L} A^T y; s \right) \quad (4.10)$$

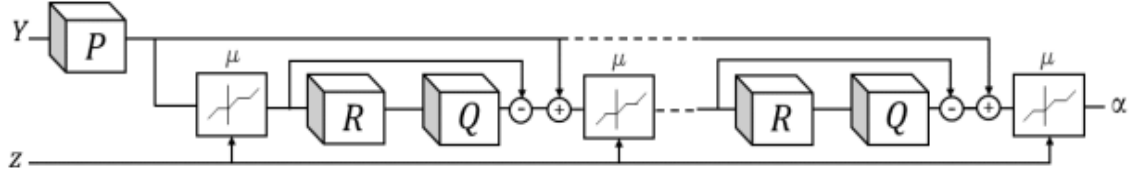
Όμως αυτός ο τρόπος λύσης δεν είναι υπολογιστικά αποδοτικός, καθώς η αραιή συνελικτική κωδικοποίηση χειρίζεται ολόκληρη την εικόνα με αποτέλεσμα να αυξηθεί η διαστατικότητα της σχέσης (4.2). Έτσι μπορούμε να γράψουμε την σχέση (4.10) σε συνελικτική μορφή με τον εξής τρόπο: Λαμβάνοντας υπόψη την αντιστοιχία μεταξύ της ειδικής μορφής του A και των συνελικτικών λεξικών, ο πολλαπλασιασμός του A με τους αραιούς κωδικούς x στην σχέση (4.10) ισοδυναμεί με τον υπολογισμό του αθροίσματος των συνελικτικών όρων της σχέσης (4.9). Καθώς ο ανάστροφος του πίνακα Toeplitz είναι πάλι Toeplitz, οι όροι A^T της σχέσης (4.10) μπορούν να θεωρηθούν και συνελικτικοί όροι. Έτσι αν αντικατασταθούν στην σχέση (4.10) οι πολλαπλασιασμοί πινάκων με συνελίξεις τότε οδηγούμαστε στην σχέση (4.11) :

$$\alpha^{t+1} = \xi_\mu \left(\alpha^t - \tilde{D} * D * \alpha^t + \tilde{D} * Y; z \right) \quad (4.11)$$

όπου $\tilde{D}$ ένα συνελικτικό λεξικό αντίστοιχο με το A^T . Ο επαναληπτικός αλγόριθμος που δίνεται από την σχέση (3.9) μπορεί να γραφτεί στην μορφή ενός βαθιού περίπλοκου νευρικού δικτύου, όπου κάθε στάδιο του δικτύου υπολογίζει τους αραιές αντιστοιχίες σύμφωνα με την σχέση (4.12)

$$\alpha^{t+1} = \xi_\mu \left(\alpha^t - Q * R * \alpha^t + P * Y; z \right) \quad (4.12)$$

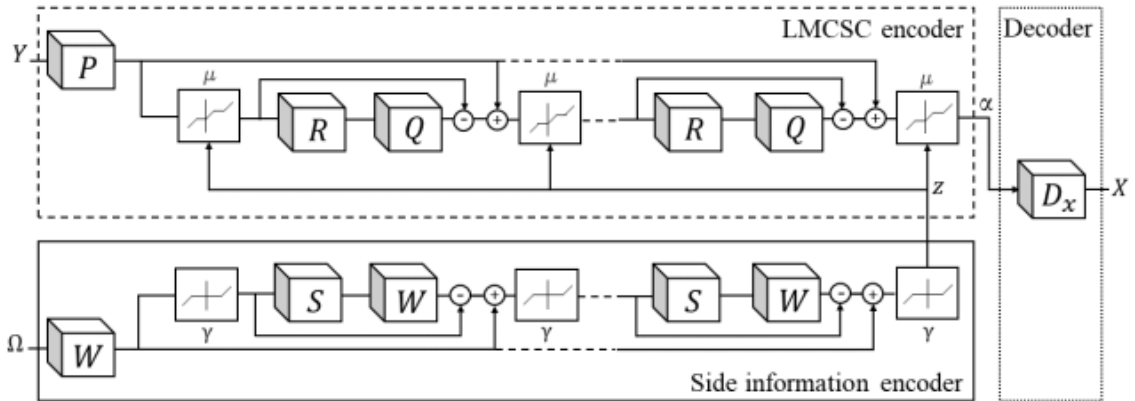
όπου Q, R, P είναι εκπαιδύσιμα συνελικτικά επίπεδα και μ μια εκπαιδύσιμη παράμετρος. Ένα τέτοιο δίκτυο ονομάζεται Learned Multimodal Convolutional Sparse Coding (LMCSC) και παρουσιάζεται γραφικά στο Σχήμα 4.2.



Σχήμα 4.2 Το Δίκτυο LMCSC

4.4 Υπερανάλυση Εικόνας με το Δίκτυο LMCSC

Το μοντέλο LMCSC μπορεί να χρησιμοποιηθεί σε βαθύ δίκτυο μάθησης για υπερανάλυση εικόνας. Ένα τέτοιο δίκτυο ερευνήθηκε [Mar2019] και αποτελούταν από 3 υπο δίκτυα. Το πρώτο υπο δίκτυο είναι ένας κωδικοποιητής LMCSC που παράγει περίπλοκες κρυμμένες αναπαραστάσεις της εικόνας-στόχου με τη βοήθεια πλευρικών πληροφοριών. Το δεύτερο είναι ένας συνελικτικός κωδικοποιητής LISTA που παράγει αραιές αναπαραστάσεις της εικόνας καθοδήγησης, δηλαδή την πλευρική πληροφορία. Τέλος το τρίτο υπο δίκτυο είναι ένας αποκωδικοποιητής συνέλιξης που παράγει την υπεραναλυμένη εικόνα. Ένα τέτοιο δίκτυο παρουσιάζεται γραφικά στο Σχήμα 4.3.



Σχήμα 4.3 Το Δίκτυο LMCSC για Υπερανάλυση Εικόνας

Σκοπός του κωδικοποιητή LMCSC είναι η μάθηση συνελικτικών αραιών χαρακτηριστικών αντιστοίχισης $\alpha \in \mathbb{R}^{n_1 \times n_2 \times k}$ για την εικόνα χαμηλής ανάλυσης χρησιμοποιώντας τα συνελικτικά αραιά χαρακτηριστικά αντιστοίχισης $z \in \mathbb{R}^{n_1 \times n_2 \times k}$ της πλευρικής πληροφορίας. Η επιθυμητή έξοδος του προτεινόμενου μοντέλου υπερανάλυσης εικόνας είναι η εικόνα υψηλής ανάλυσης του ίδιου

τύπου με την εικόνα εισόδου χαμηλής ανάλυσης. Ακόμα το δίκτυο εκπαιδεύεται από άκρη σε άκρη ελαχιστοποιώντας το σφάλμα ανακατασκευής μεταξύ της ανακατασκευασμένης εικόνας και την εικόνας υψηλής ανάλυσης που είναι στο ίδιο τύπο. Τέλος αξίζει να σημειωθεί ότι αυτό το δίκτυο δημιουργείται με την υπόθεση ότι η εικόνα χαμηλής ανάλυσης και η εικόνα υψηλής ανάλυσης μοιράζονται τα ίδια συνελικτικά χαρακτηριστικά αντιστοίχισης χρησιμοποιώντας διαφορετικά συνελικτικά λεξικά.

ΚΕΦΑΛΑΙΟ 5

ΠΕΙΡΑΜΑΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ

5.1 Λεπτομέρειες υλοποίησης

5.2 Διαδικασία ελέγχου

5.1 Λεπτομέρειες Υλοποίησης

5.1.1 Κατασκευή συνόλων δεδομένων εκπαίδευσης και ελέγχου

Για την εκπαίδευση και την αξιολόγηση της απόδοσης του δικτύου χρησιμοποιήθηκε η βάση δεδομένων¹ EPFL RGB-NIR, η οποία αποτελείται από 477 πολύχρωμες (RGB) και Near-infrared (NIR) εικόνες. Από αυτές, οι 8 χρησιμοποιούνται για τον έλεγχο του δικτύου και οι υπόλοιπες αποτελούν την βάση δεδομένων για εκπαίδευση. Οι εικόνες που χρησιμοποιούνται για πλευρική βοήθεια είναι οι RGB, ενώ οι NIR εικόνες αποτελούν τις εικόνες εισόδου και στόχου του δικτύου. Οι εικόνες χαμηλής ανάλυσης κατασκευάζονται με υπο-δειγματοληψία των εικόνων υψηλής ανάλυσης και έπειτα με υπερ-δειγματοληψία

¹ https://ivrl.epfl.ch/supplementary_material/cvpr11/

αυτών χρησιμοποιώντας δικυβική παρεμβολή. Επιπλέον παίρνουμε κομμάτια 56×56 από αυτές τις εικόνες για να αυξηθεί το μέγεθος της βάσης δεδομένων για εκπαίδευση. Όσο αφορά της εικόνες καθοδήγησης υλοποιήθηκαν δύο πειράματα. Στο ένα πείραμα χρησιμοποιήθηκαν οι RGB εικόνες, ενώ στο δεύτερο πείραμα πρώτα μετατράπηκαν στον χρωματικό χώρο YCbCr και έπειτα επιλέχθηκε μόνο το κανάλι χρώματος Y, δηλαδή της φωτεινότητας (luminance).

5.1.2 Ορισμός και αρχικοποίηση των μεταβλητών του νευρωνικού δικτύου

Για το δίκτυο της εργασίας χρησιμοποιήθηκαν 3 στάδια unfolding τόσο στον κλάδο του κωδικοποιητή LMCSC όσο και στον κλάδο της πλευρικής πληροφορίας. Για τα πειράματα που χρησιμοποιούνται οι εικόνες RGB σαν εικόνες καθοδήγησης, σαν μέγεθος των P, Q και W επιλέχθηκε το $7 \times 7 \times 3 \times 40$ και για μέγεθος των R, S και D_x επιλέχθηκε το $7 \times 7 \times 40 \times 3$. Για τα πειράματα που χρησιμοποιούνται οι εικόνες του χρωματικού καναλιού Y σαν εικόνες καθοδήγησης, σαν μέγεθος των P, Q και W επιλέχθηκε το $7 \times 7 \times 1 \times 40$ και για μέγεθος των R, S και D_x επιλέχθηκε το $7 \times 7 \times 40 \times 1$. Τα βάρη όλων των επιπέδων αρχικοποιούνται τυχαία χρησιμοποιώντας την κατανομή gaussian με τυπική απόκλιση 0,01, ενώ η εκπαιδευσιμη παράμετρος μ αρχικοποιείται στην τιμή 0,1.

5.1.3 Εκπαίδευση δικτύου

Για την εκπαίδευση του δικτύου χρησιμοποιούνται εικόνες μεγέθους 56×56 οι οποίες εισέρχονται σε δεσμίδες (batches) μεγέθους 128. Σαν συνάρτηση κόστους χρησιμοποιείται η συνάρτηση ελάχιστων τετραγώνων (MSE) μεταξύ της εξόδου του δικτύου και της εικόνας στόχου NIR. Για την βελτιστοποίηση του κόστους χρησιμοποιείται ο αλγόριθμος βελτιστοποίησης Adam, ο οποίος είναι μια παραλλαγή του αλγορίθμου κατάβασης δυναμικού (Gradient Descent). Το δίκτυο εκπαιδεύτηκε για 2000 επαναλήψεις (epochs). Για την δημιουργία του μοντέλου χρησιμοποιήθηκε η βιβλιοθήκη νευρωνικών δικτύων Tensorflow της Google, ενώ για την επιτάχυνση την διαδικασίας εκπαίδευσης χρησιμοποιήθηκε η πλατφόρμα παράλληλου προγραμματισμού CUDA που αναπτύχθηκε από την Nvidia. Η διαδικασία της εκπαίδευσης υλοποιήθηκε από έναν υπολογιστή που είχε τα εξής

χαρακτηριστικά: επεξεργαστής AMD Ryzen™ 7 1700X, κάρτα γραφικών NVIDIA GEFORCE GTX 1070 Ti για τους υπολογισμούς. Ο χρόνος της εκπαίδευσης είναι περίπου 12 ώρες.

5.2 Διαδικασία Ελέγχου

5.2.1 Πειραματικά δεδομένα

Για τον έλεγχο της αποτελεσματικότητας του δικτύου πραγματοποιήσαμε μια σειρά πειραμάτων. Πιο συγκεκριμένα τα πειράματα που πραγματοποιήσαμε χωρίζονται σε δύο κατηγορίες, σε πειράματα με μεταβλητό συντελεστή υπερανάληψης και σε πειράματα με διαφορετικό τύπο εικόνων καθοδήγησης. Και στις δυο περιπτώσεις που χρησιμοποιούνται οι διαφορετικοί τύποι εικόνων καθοδήγησης χρησιμοποιούνται οι ίδιοι συντελεστές υπερανάληψης. Παρακάτω παρουσιάζονται αναλυτικά να πειράματα ανά τύπο.

5.2.2 Χρήση της RGB εικόνας ως εικόνα καθοδήγησης

Για αυτά τα πειράματα χρησιμοποιήθηκαν αυτούσιες οι εικόνες RGB της βάσης δεδομένων ως πλευρική πληροφορία. Έπειτα πραγματοποιήθηκαν τέσσερα διαφορετικά είδη του μοντέλου ανάλογα με τον συντελεστή υπερανάληψης που εφαρμόστηκε στις εικόνες εκπαίδευσης. Το πρώτο μοντέλο εκπαιδεύεται με $x2$ συντελεστή υπερανάληψης, το δεύτερο με $x3$ συντελεστή υπερανάληψης και το τρίτο με $x4$ συντελεστή υπερανάληψης και το τέταρτο με $x6$ συντελεστή υπερανάληψης. Τα αποτελέσματα που παρουσιάζουμε είναι ο μέσος όρος PSNR (Peak Signal-to-Noise Ratio) των αποτελεσμάτων ελέγχου. Στους παρακάτω πίνακες παρουσιάζονται συγκεντρωτικά τα αποτελέσματα υπερανάληψης για όλους τους συντελεστές υπερανάληψης (κλίμακες). Επίσης αναφέρονται και τα αποτελέσματα προηγούμενων ερευνών που χρησιμοποίησαν ίδια σύνολα δεδομένων, για μια άμεση σύγκριση. Πιο συγκεκριμένα συγκρίθηκαν με τα μοντέλα μονού τύπου: i) CSCN [Liu2016] ii) ACSC [SrG2018] και τα μοντέλα πολλαπλού iii) DJF [Li2016], που επεκτείνει ένα μοντέλο vanilla CNN σε ένα μοντέλο πολλαπλών τύπων, iv) το μοντέλο CDLSR [Son2019] που είναι ένα μοντέλο πολλαπλών τύπων αραιής κωδικοποίησης και το μοντέλο LMCSC [Mar2019] που αποτέλεσε την βάση για το μοντέλο μας. Για όλα τα μοντέλα δεν

υπάρχουν αποτελέσματα για συντελεστή υπερανάλυσης x3, ενώ για το μοντέλο CDLSR δεν υπάρχουν αποτελέσματα και συντελεστή υπερανάλυσης x4.

Πίνακας 5.1 Συγκεντρωτικά Αποτελέσματα PSNR για Εικόνες Καθοδήγησης RGB

Upsampling	X2	X3	X4	X6
CSCN [Liu2016]	34.76	-	29.14	26.86
ACSC [SrG2018]	34.93	-	29.03	26.85
DJF [Li2016]	36.03	-	31.83	30.75
CDLSR [Son2019]	-	-	33.54	31.32
LMCSC [Mar2019]	38.88	-	34.32	32.46
Το δικό μας LMCSC με εικόνες καθοδήγησης RGB	37.25	34.92	34.06	33.12

Όπως φαίνεται στον Πίνακα 5.1 το μοντέλο μας που χρησιμοποιεί εικόνες καθοδήγησης RGB παρουσιάζει βελτίωση στα αποτελέσματα PSNR σε σχέση με τις προηγούμενες έρευνες, εκτός του μοντέλου LMCSC [Mar2019], σε όλους τους συντελεστές υπερανάλυσης. Το μοντέλο LMCSC [Mar2019] που αποτελεί και την βάση του μοντέλου μας συνεχίζει να έχει καλύτερα αποτελέσματα στους συντελεστές υπερανάλυσης x2 και x4. Αξιοσημείωτη είναι ωστόσο η βελτίωση του μοντέλου μας και στον μεγαλύτερο συντελεστή υπερανάλυσης καθώς στον συντελεστή υπερανάλυσης x6 επιτυγχάνει καλύτερο αποτέλεσμα PSNR και από το μοντέλο LMCSC [Mar2019] αφού η αύξηση του PSNR είναι 0.66 dB.

5.2.3 Χρήση του καναλιού Y ως εικόνα καθοδήγησης

Για αυτά τα πειράματα χρησιμοποιήθηκαν ως πλευρική πληροφορία οι εικόνες της βάσης δεδομένων που αποτελούνται μόνο από το Y κανάλι των εικόνων που είναι στον χρωματικό χώρο YCbCr της βάσης δεδομένων ως πλευρική πληροφορία. Η σύγκριση των αποτελεσμάτων PSNR γίνεται πάλι για τους ίδιους

συντελεστές υπερανάλυσης και με τις ίδιες έρευνες που χρησιμοποιήθηκαν στο υποκεφάλαιο 5.2.2. Επίσης γίνεται και η σύγκριση του μοντέλου όταν αυτό χρησιμοποιεί για εικόνες καθοδήγησης εικόνες RGB και όταν χρησιμοποιεί εικόνες του καναλιού Y.

Πίνακας 5.2 Συγκεντρωτικά Αποτελέσματα PSNR για Εικόνες Καθοδήγησης Y

Upsampling	X2	X3	X4	X6
CSCN [Liu2016]	34.76	-	29.14	26.86
ACSC [SrG2018]	34.93	-	29.03	26.85
DJF [Li2016]	36.03	-	31.83	30.75
CDLSR [Son2019]	-	-	33.54	31.32
LMCSC [Mar2019]	38.88	-	34.32	32.46
Το δικό μας LMCSC με εικόνες καθοδήγησης Y	37.82	35.15	34.50	33.39

Στον Πίνακα 5.2 παρουσιάζονται τα αποτελέσματα από το μοντέλο μας όταν αυτό χρησιμοποιεί για εικόνες καθοδήγησης μόνο το κανάλι Y. Όπως παρατηρείται, και σε αυτήν την περίπτωση το μοντέλο μας παρουσιάζει καλύτερα αποτελέσματα PSNR σε σχέση με τις προηγούμενες έρευνες σε όλους τους συντελεστές υπερανάλυσης. Η μόνη εξαίρεση είναι ότι στον συντελεστή υπερανάλυσης x2 το μοντέλο LMCSC [Mar2019] συνεχίζει να έχει το καλύτερο αποτέλεσμα από όλες τις έρευνες. Πιο συγκεκριμένα, στον συντελεστή x4 η αύξηση του PSNR είναι 0.18 dB ενώ στον συντελεστή x6 η αύξηση του PSNR είναι 0.93 dB.

Παρατηρώντας τους δύο πίνακες μπορούμε εύκολα να παρατηρήσουμε πως το μοντέλο της παρούσας εργασίας παρουσιάζει μια βελτιωμένη πρόταση του

μοντέλου LMCSC [Mar2019] πολλαπλών τύπων για υπερανάλυση εικόνων ειδικότερα σε μεγαλύτερους συντελεστές υπερανάλυσης.

Στον Πίνακα 5.3 παρουσιάζεται η σύγκριση του μοντέλου όταν αυτό χρησιμοποιεί για εικόνες καθοδήγησης εικόνες RGB και όταν χρησιμοποιεί εικόνες του καναλιού Y. Είναι αντιληπτό πως όταν χρησιμοποιείται το κανάλι Y απόδοση του μοντέλου είναι καλύτερη σε όλους τους συντελεστές υπερανάλυσης.

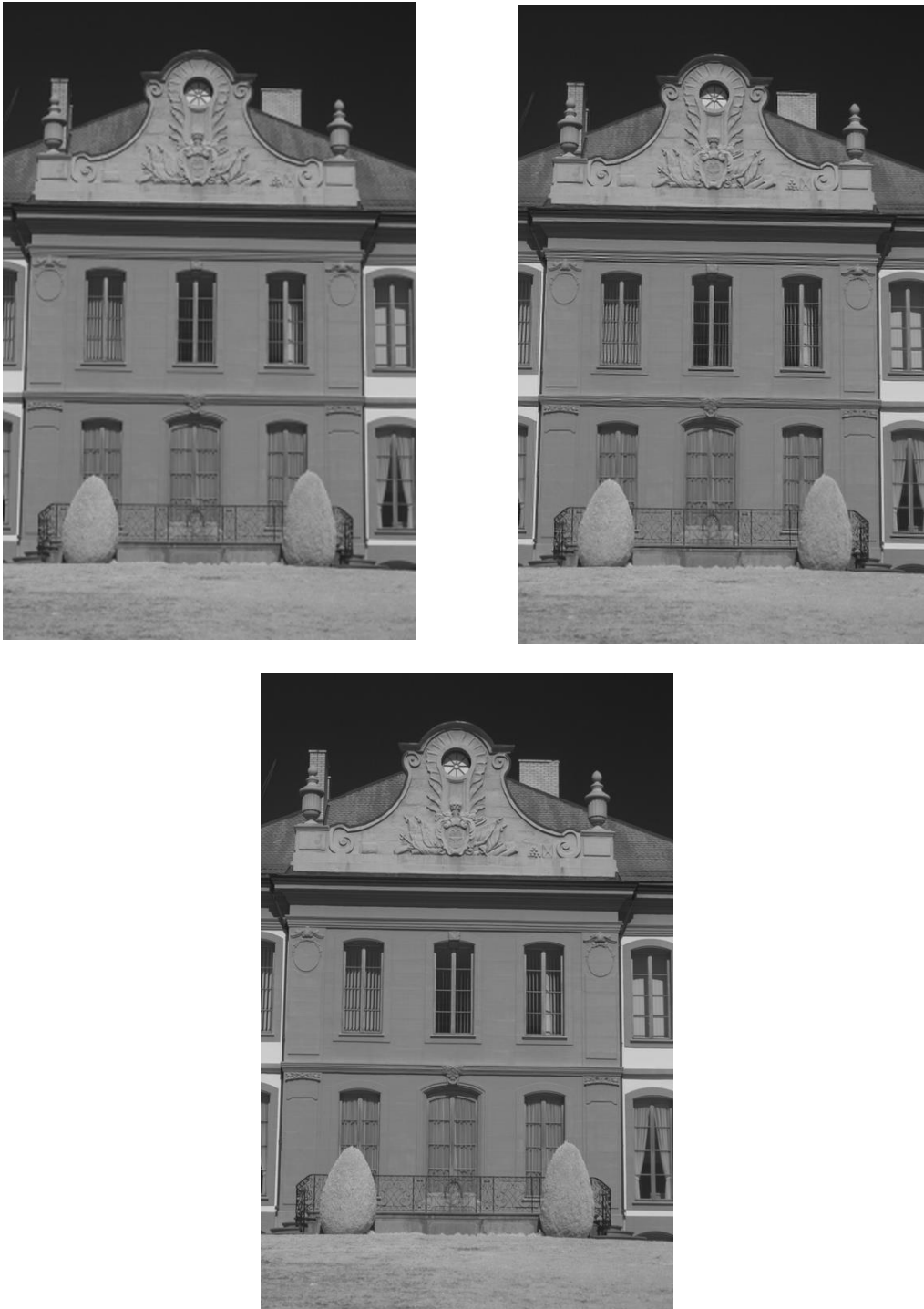
Πίνακας 5.3 Σύγκριση των Δύο Δικτύων ανάλογα με τις Εικόνες Καθοδήγησης

Upsampling	LMCSC με εικόνες καθοδήγησης RGB	LMCSC με εικόνες καθοδήγησης Y
X2	37.25	37.82
X3	34.92	35.15
X4	34.06	34.50
X6	33.12	33.39

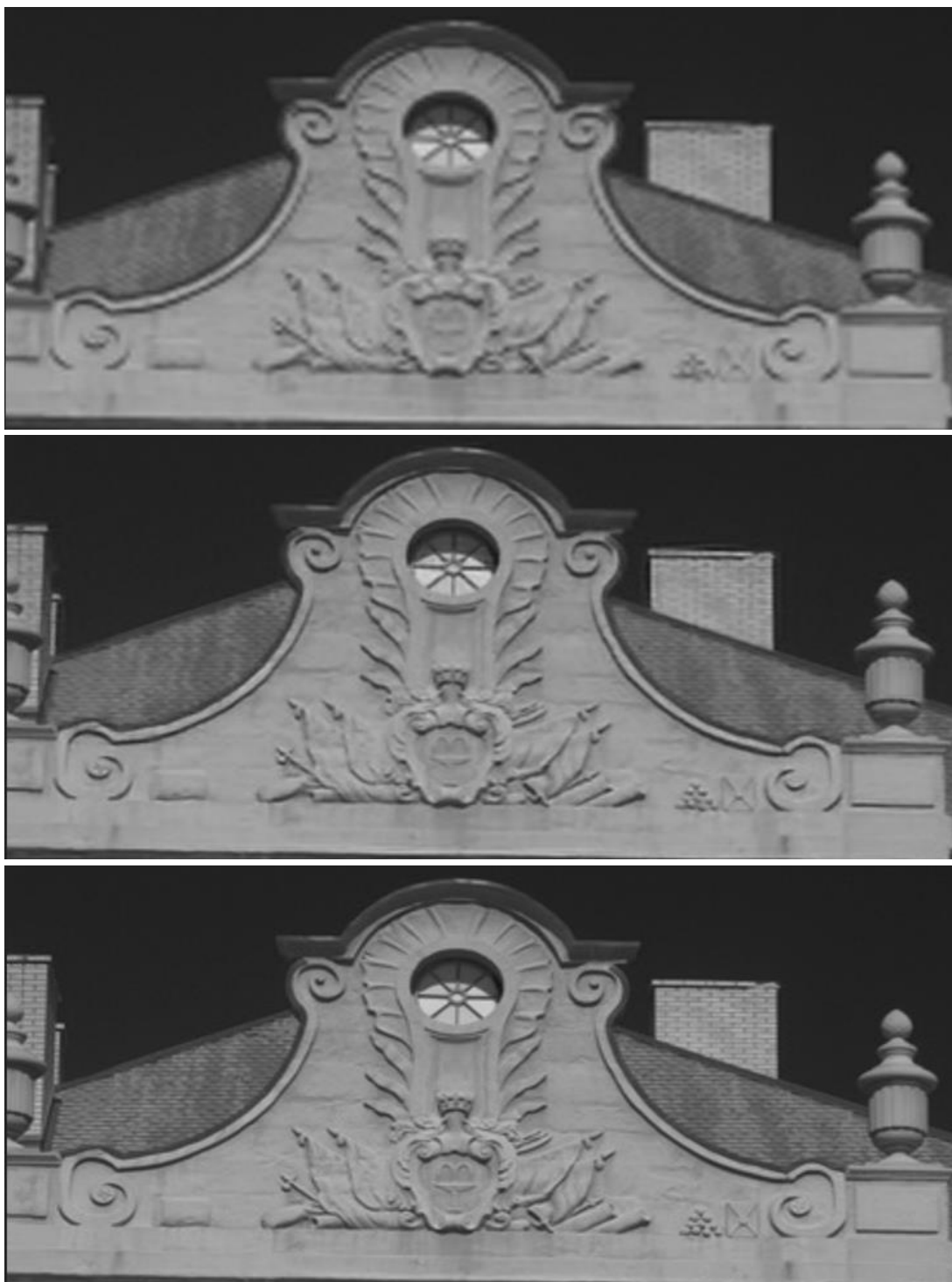
5.2.4 Οπτικά αποτελέσματα

Εκτός από τον αριθμητικό έλεγχο του μοντέλου, θεωρείται αναγκαίος και ο οπτικός έλεγχος, αφού έτσι γίνεται πραγματικά αντιληπτό το αποτέλεσμα του δικτύου. Για τον οπτικό έλεγχο χρησιμοποιήθηκε το μοντέλο που χρησιμοποιεί για εικόνες καθοδήγησης εικόνες του καναλιού Y, καθώς αυτό εμφάνισε την καλύτερη απόδοση. Παρακάτω παρουσιάζονται τα αποτελέσματα του μοντέλου για τους συντελεστές υπερανάλυσης x2, x3, x4, x6. Οι εικόνες είναι από την βάση δεδομένων EPFL RGB-NIR και συγκεκριμένα είναι η εικόνα “oldbuilding-0030” και η εικόνα “urban-0017”.

5.2.5 Οπτικά αποτελέσματα για συντελεστή υπερανάλυσης $\times 2$



Σχήμα 5.1 Αποτελέσματα του Μοντέλου για Συντελεστή Υπερανάλυσης $\times 2$. Πάνω αριστερά η Αρχική Εικόνα (LR), πάνω δεξιά η Εικόνα που Παράγει το Μοντέλο (estimated) και κάτω η Εικόνα Υψηλής Ανάλυσης (HR).



Σχήμα 5.2 Αποτελέσματα του Μοντέλου για Συντελεστή Υπερανάλυσης $\times 2$. Στην Πρώτη Γραμμή Παρουσιάζονται η Εικόνα Εισόδου (LR), στην Δεύτερη Γραμμή η Εικόνα που Παράγει το Μοντέλο (estimated) και στην Τρίτη Γραμμή η Εικόνα Υψηλής Ανάλυσης (HR).



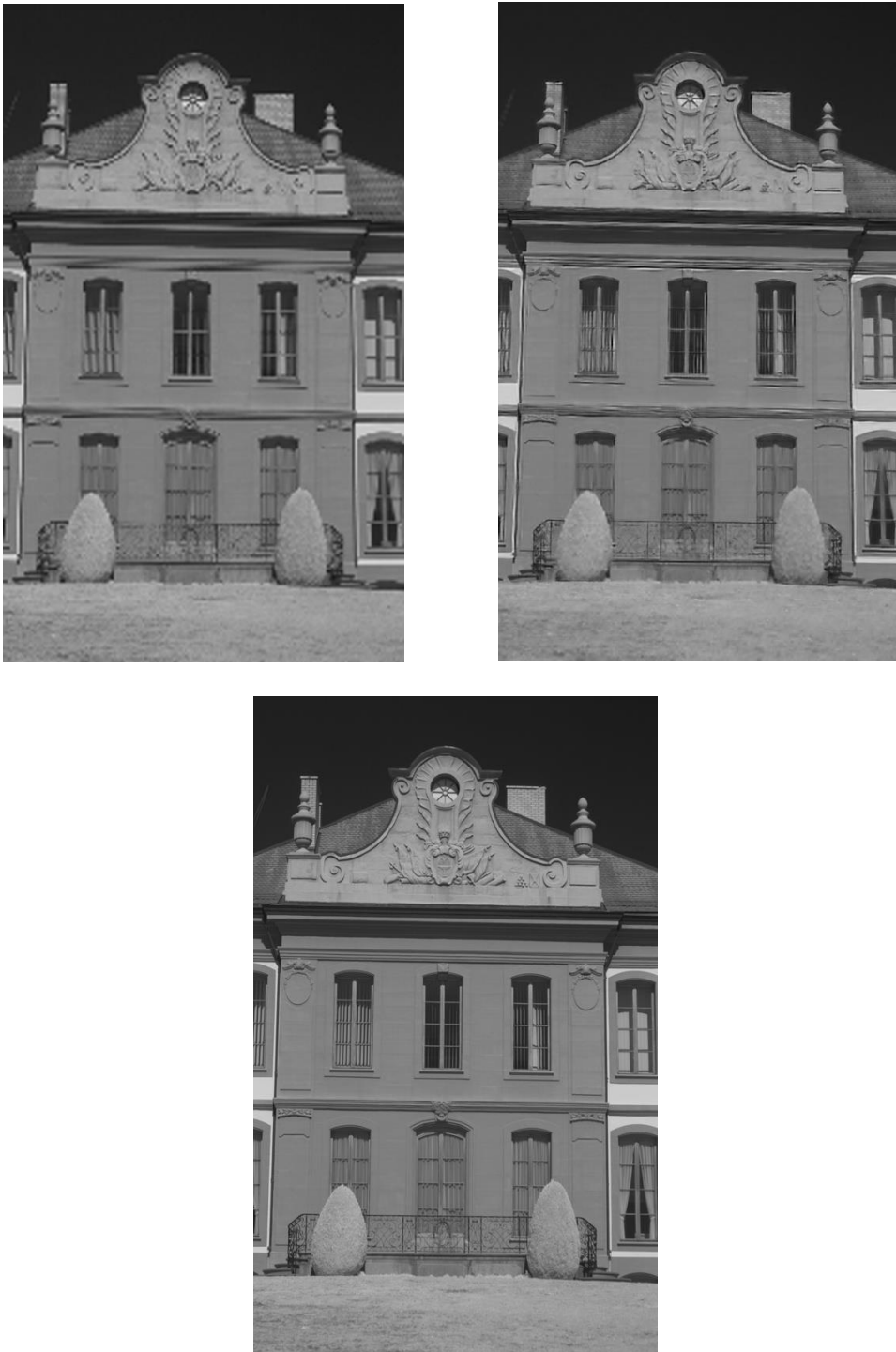
Σχήμα 5.3 Αποτελέσματα του Μοντέλου για Συντελεστή Υπερανάλυσης x2. Στην Πρώτη Γραμμή Παρουσιάζονται η Εικόνα Εισόδου (LR), στην Δεύτερη Γραμμή η Εικόνα που Παράγει το Μοντέλο (estimated) και στην Τρίτη Γραμμή η Εικόνα Υψηλής Ανάλυσης (HR).



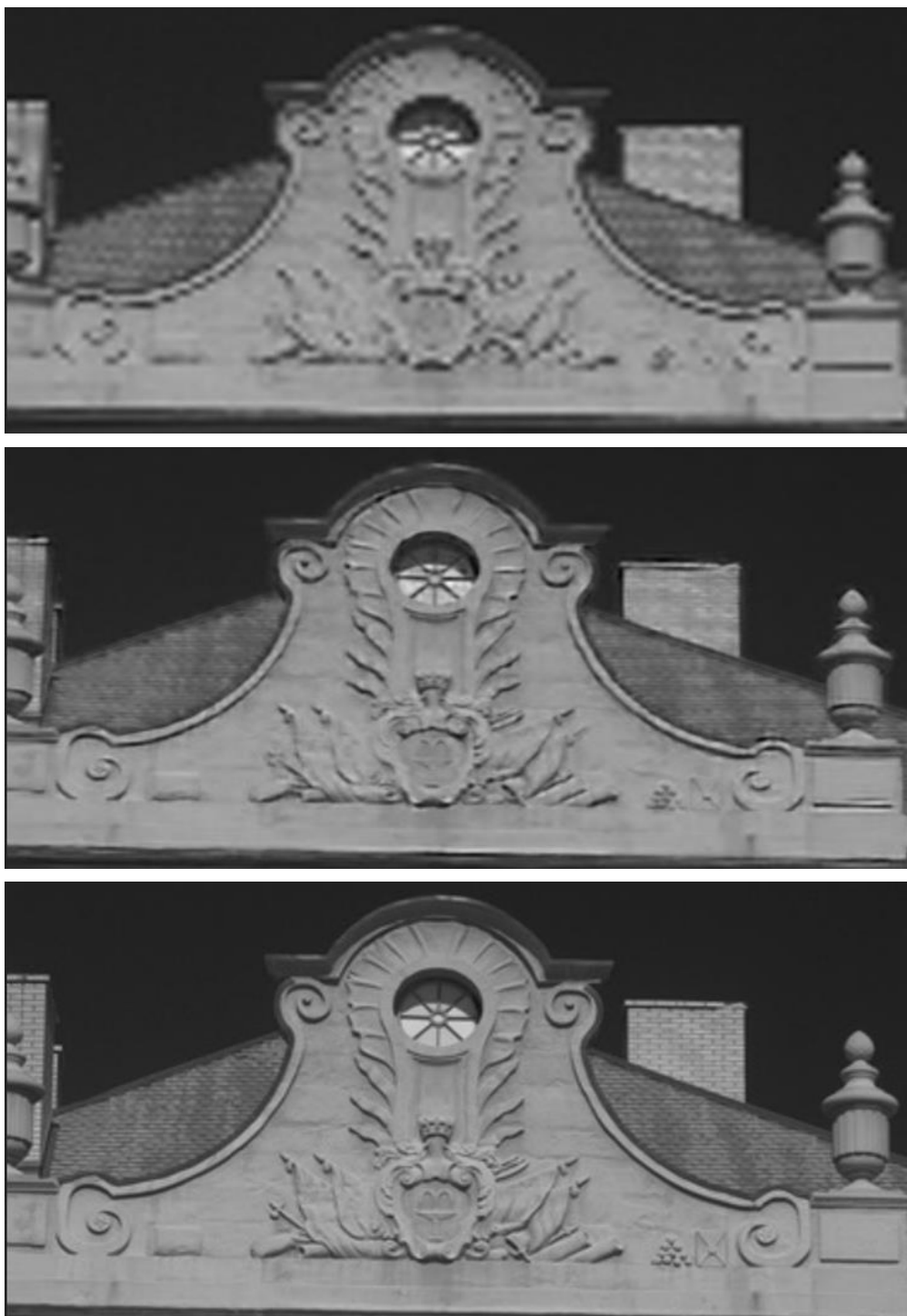
Σχήμα 5.4 Αποτελέσματα του Μοντέλου για Συντελεστή Υπερανάλυσης $\times 2$. Πάνω αριστερά η Αρχική Εικόνα (LR), πάνω δεξιά η Εικόνα που Παράγει το Μοντέλο (estimated) και κάτω η Εικόνα Υψηλής Ανάλυσης (HR).

Παρατηρώντας τις παραπάνω εικόνες μπορούμε να συμπεράνουμε ότι στον συντελεστή υπερανάλυσης $\times 2$, το μοντέλο μας ναι μεν βελτιώνει την ποιότητα της εικόνας αλλά αυτό δεν είναι ορατό όταν έχουμε ολόκληρες τις εικόνες (Σχήμα 5.1 και Σχήμα 5.3). Όμως στην περίπτωση που κάνουμε μεγέθυνση στις εικόνες (Εικόνα Σχήμα 5.2 και Σχήμα 5.4) βλέπουμε ότι η ποιότητα της εκτιμώμενης εικόνας είναι αρκετά σημαντική.

5.2.6 Οπτικά αποτελέσματα για συντελεστή υπερανάλυσης $\times 3$



Σχήμα 5.5 Αποτελέσματα του Μοντέλου για Συντελεστή Υπερανάλυσης $\times 3$. Πάνω αριστερά η Αρχική Εικόνα (LR), πάνω δεξιά η Εικόνα που Παράγει το Μοντέλο (estimated) και κάτω η Εικόνα Υψηλής Ανάλυσης (HR).



Σχήμα 5.6 Αποτελέσματα του Μοντέλου για Συντελεστή Υπερανάλυσης $\times 3$. Στην Πρώτη Γραμμή Παρουσιάζεται η Εικόνα Εισόδου (LR), στην Δεύτερη Γραμμή η Εικόνα που Παράγει το Μοντέλο (estimated) και στην Τρίτη Γραμμή η Εικόνα Υψηλής Ανάλυσης (HR).



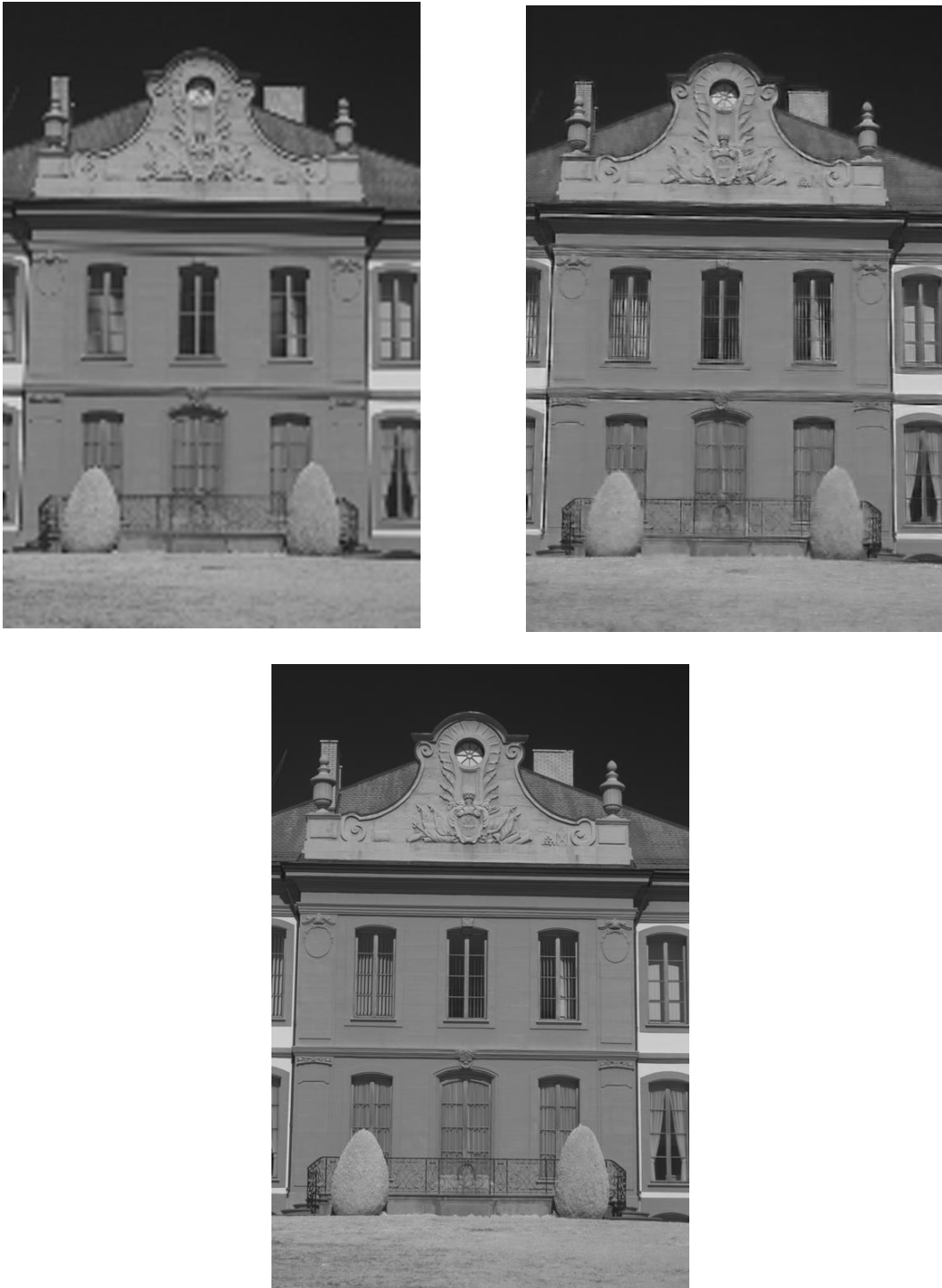
Σχήμα 5.7 Αποτελέσματα του Μοντέλου για Συντελεστή Υπερανάλυσης $\times 3$. Στην Πρώτη Γραμμή Παρουσιάζεται η Εικόνα Εισόδου (LR), στην Δεύτερη Γραμμή η Εικόνα που Παράγει το Μοντέλο (estimated) και στην Τρίτη Γραμμή η Εικόνα Υψηλής Ανάλυσης (HR).



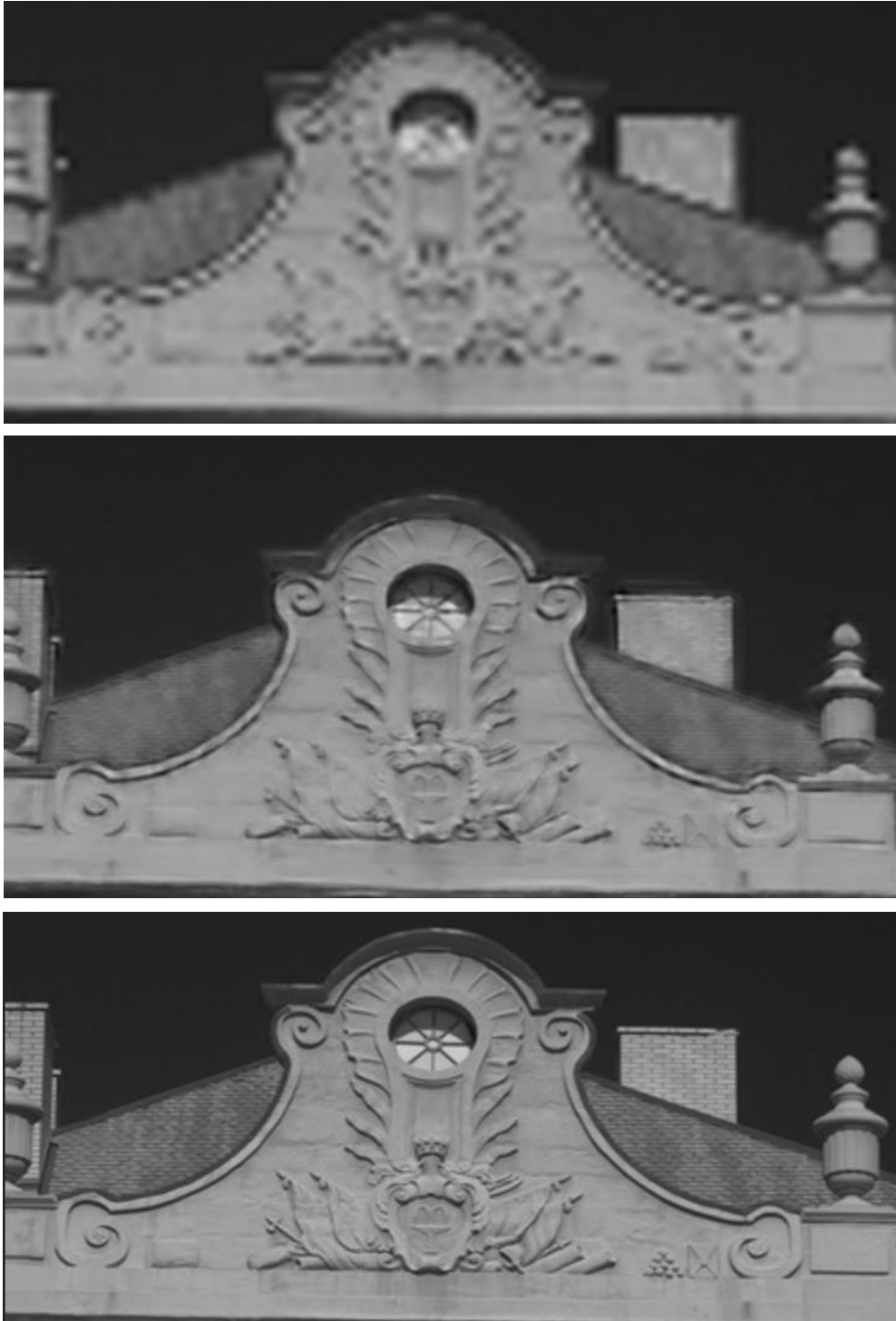
Σχήμα 5.8 Αποτελέσματα του Μοντέλου για Συντελεστή Υπερανάλυσης $\times 3$. Πάνω αριστερά η Αρχική Εικόνα (LR), πάνω δεξιά η Εικόνα που Παράγει το Μοντέλο (estimated) και κάτω η Εικόνα Υψηλής Ανάλυσης (HR).

Σχετικά με τον συντελεστή υπερανάλυσης $\times 3$ ισχύουν παρόμοια συμπεράσματα με αυτά του συντελεστή υπερανάλυσης $\times 2$. Στις ολόκληρες εικόνες (Σχήμα 5.5 και Σχήμα 5.7) παρατηρούμε μια αύξηση στην ποιότητα της εκτιμώμενης εικόνας συγκριτικά με την εικόνας εισόδου, ενώ στις εικόνες που προκύπτουν μετά από μεγέθυνση (Σχήμα 5.6 και Σχήμα 5.8) η διαφορά της ποιότητας μεταξύ της εκτιμώμενης και την εικόνας εισόδου είναι πολύ εντυπωσιακή.

5.2.7 Οπτικά αποτελέσματα για συντελεστή υπερανάλυσης $\times 4$



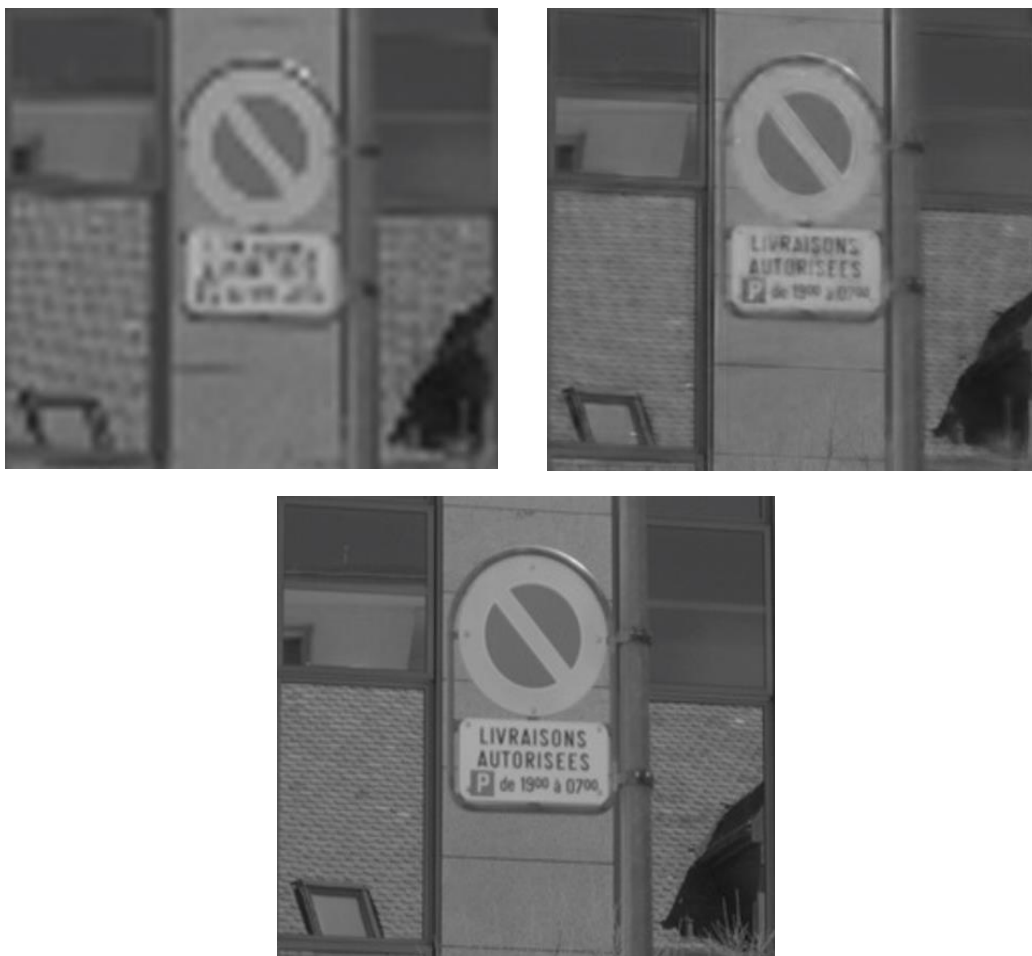
Σχήμα 5.9 Αποτελέσματα του Μοντέλου για Συντελεστή Υπερανάλυσης $\times 4$. Στην Πρώτη Γραμμή Παρουσιάζεται η Εικόνα Εισόδου (LR), στην Δεύτερη Γραμμή η Εικόνα που Παράγει το Μοντέλο (estimated) και στην Τρίτη Γραμμή η Εικόνα Υψηλής Ανάλυσης (HR).



Σχήμα 5.10 Αποτελέσματα του Μοντέλου για Συντελεστή Υπερανάλυσης $\times 4$. Πάνω αριστερά η Αρχική Εικόνα (LR), πάνω δεξιά η Εικόνα που Παράγει το Μοντέλο (estimated) και κάτω η Εικόνα Υψηλής Ανάλυσης (HR).



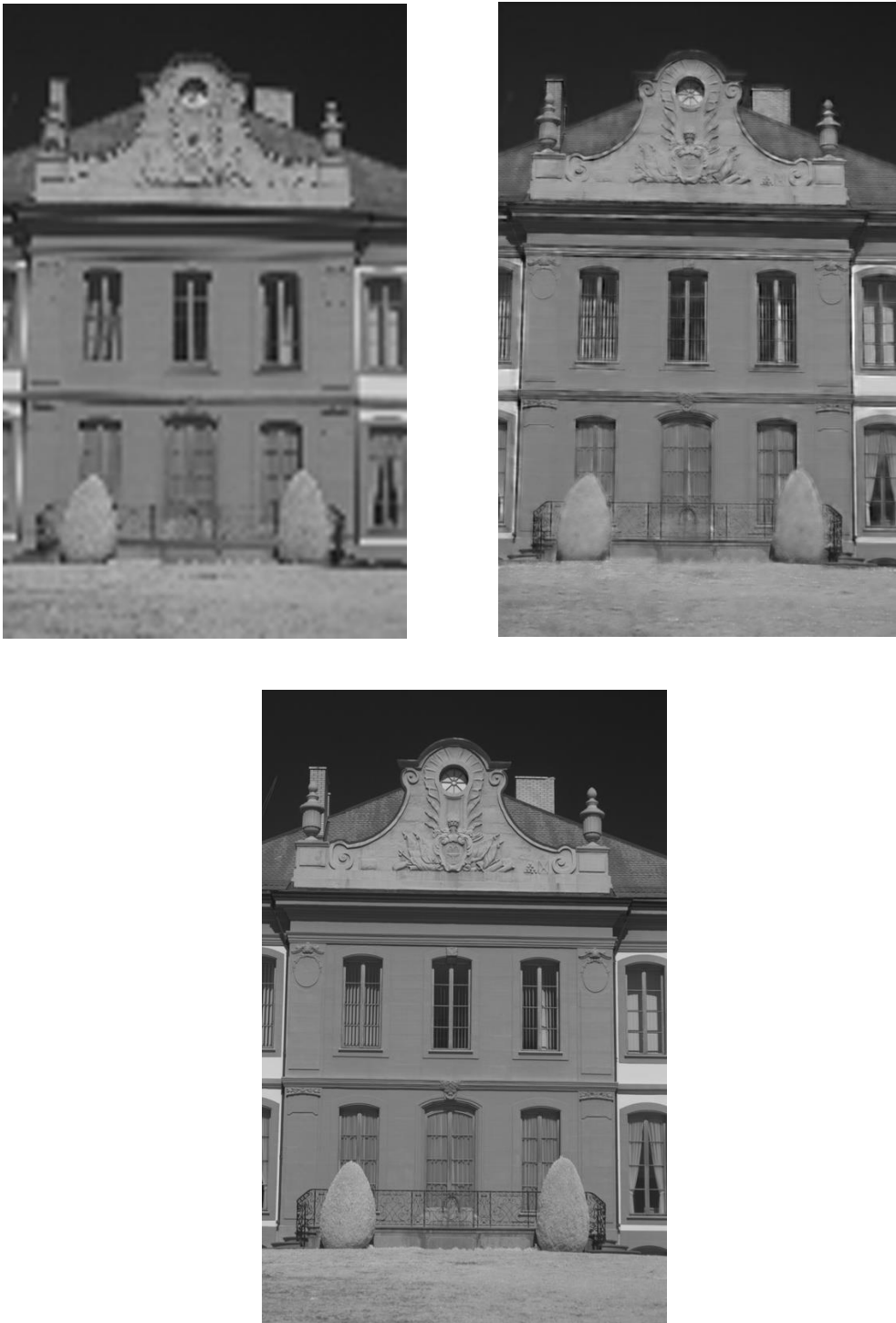
Σχήμα 5.11 Αποτελέσματα του Μοντέλου για Συντελεστή Υπερανάλυσης $\times 4$. Στην Πρώτη Γραμμή Παρουσιάζεται η Εικόνα Εισόδου (LR), στην Δεύτερη Γραμμή η Εικόνα που Παράγει το Μοντέλο (estimated) και στην Τρίτη Γραμμή η Εικόνα Υψηλής Ανάλυσης (HR).



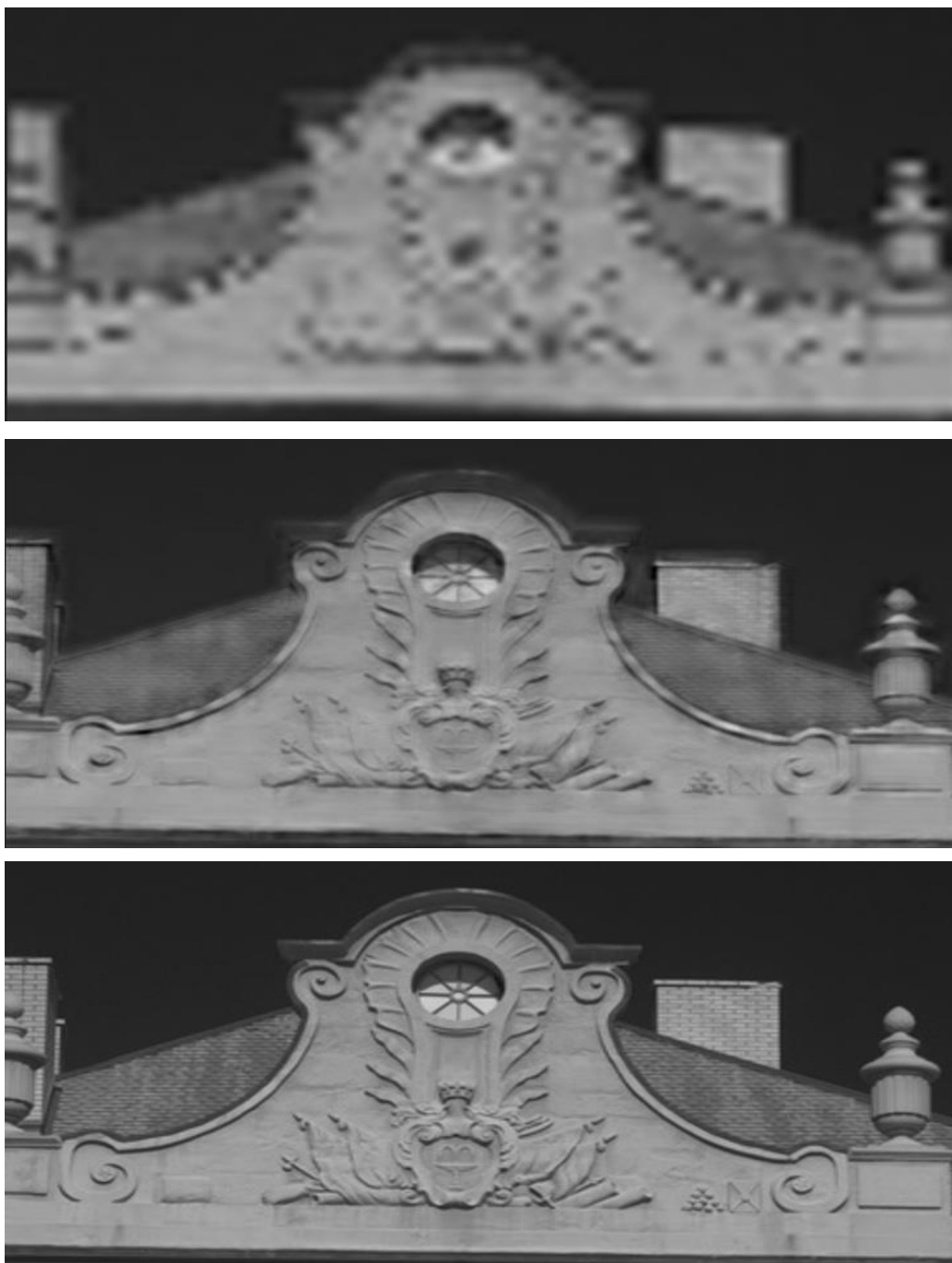
Σχήμα 5.12 Αποτελέσματα του Μοντέλου για Συντελεστή Υπερανάλυσης $\times 4$. Πάνω αριστερά η Αρχική Εικόνα (LR), πάνω δεξιά η Εικόνα που Παράγει το Μοντέλο (estimated) και κάτω η Εικόνα Υψηλής Ανάλυσης (HR).

Όσο αφορά τον συντελεστή υπερανάλυσης $\times 4$ μπορούμε πλέον να παρατηρήσουμε και την μεγάλη διαφορά στην ποιότητα μεταξύ της εικόνας εισόδου και την εκτιμώμενης και στις εικόνες κανονικού μεγέθους (Σχήμα 5.9 και Σχήμα 5.11). Αναφορικά με τις εικόνες μεγέθυνσης (Σχήμα 5.10 και Σχήμα 5.12) η ποιότητα της παραγόμενης εικόνας είναι θαυμαστική αν υπολογιστεί το ποια εικόνα χρησιμοποιείται σαν είσοδος στο μοντέλο μας.

5.2.8 Οπτικά αποτελέσματα για συντελεστή υπερανάλυσης $\times 6$



Σχήμα 5.13 Αποτελέσματα του Μοντέλου για Συντελεστή Υπερανάλυσης $\times 6$. Στην Πρώτη Γραμμή Παρουσιάζεται η Εικόνα Εισόδου (LR), στην Δεύτερη Γραμμή η Εικόνα που Παράγει το Μοντέλο (estimated) και στην Τρίτη Γραμμή η Εικόνα Υψηλής Ανάλυσης (HR).



Σχήμα 5.14 Αποτελέσματα του Μοντέλου για Συντελεστή Υπερανάλυσης $\times 6$. Πάνω αριστερά η Αρχική Εικόνα (LR), πάνω δεξιά η Εικόνα που Παράγει το Μοντέλο (estimated) και κάτω η Εικόνα Υψηλής Ανάλυσης (HR).



Σχήμα 5.15 Αποτελέσματα του Μοντέλου για Συντελεστή Υπερανάλυσης $\times 6$. Στην Πρώτη Γραμμή Παρουσιάζεται η Εικόνα Εισόδου (LR), στην Δεύτερη Γραμμή η Εικόνα που Παράγει το Μοντέλο (estimated) και στην Τρίτη Γραμμή η Εικόνα Υψηλής Ανάλυσης (HR).



Σχήμα 5.16 Αποτελέσματα του Μοντέλου για Συντελεστή Υπερανάλυσης $\times 6$. Πάνω αριστερά η Αρχική Εικόνα (LR), πάνω δεξιά η Εικόνα που Παράγει το Μοντέλο (estimated) και κάτω η Εικόνα Υψηλής Ανάλυσης (HR).

Στον συντελεστή υπερανάλυσης $\times 6$ τα αποτελέσματα είναι θεαματικά τόσο στις ολόκληρες εικόνες (Σχήμα 5.13 και Σχήμα 5.15) όσο και στις εικόνες μεγέθυνσης (Σχήμα 5.14 και Σχήμα 5.16). Οι εικόνες εισόδου του μοντέλου είναι πολύ κακής ποιότητας ενώ οι εκτιμώμενες εικόνες είναι αρκετά καθαρές και έχουν καλή ανάλυση. Στις εικόνες εισόδου δεν μπορούμε να διακρίνουμε γράμματα ή μοτίβα ενώ στις παραγόμενες εικόνες διακρίνονται πολύ έντονα φτάνοντας να είναι συγκρίσιμα με αυτά των αρχικών εικόνων υψηλής ανάλυσης.

ΚΕΦΑΛΑΙΟ 6

ΤΕΛΙΚΑ ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΠΙΘΑΝΟΙ ΤΡΟΠΟΙ ΒΕΛΤΙΩΣΗΣ

Η υπερανάλυση εικόνας αποτελεί ένα ενεργό θέμα στον τομέα της επεξεργασίας εικόνας. Ο λόγος είναι ότι μπορεί να εφαρμοστεί σε αρκετά παραδείγματα τόσο της καθημερινής μας ζωής όσο και σε πιο εξειδικευμένες καταστάσεις, όπως είναι η ιατρική. Η χρήση των εικόνων με υψηλή ανάλυση δεν είναι εύκολη διαδικασία καθώς τα HR συστήματα έχουν υψηλό κόστος και έτσι πρέπει να χρησιμοποιούνται συμπιεσμένες εικόνες που έχουν μειωμένη λεπτομέρεια και ποιότητα. Επομένως η υπερανάλυση εικόνας είναι ιδιαίτερα σημαντική καθώς μπορούμε να δημιουργήσουμε μια εικόνα υψηλής ανάλυσης από μία εικόνα χαμηλής ανάλυσης.

Ωστόσο, το πρόβλημα της υπερανάλυσης εικόνας δεν έχει μοναδική λύση και επομένως η μελέτη και κατανόηση των προηγούμενων ερευνών είναι απαραίτητη για την βελτίωση των προτεινόμενων λύσεων, αφού ο συνδυασμός κάποιων μεθόδων μπορεί να οδηγήσει σε καλύτερα αποτελέσματα. Στην συγκεκριμένη εργασία προτείνεται ένα deep unfolding μοντέλο πολλαπλών τύπων που χρησιμοποιεί πλευρική πληροφορία για την επίτευξη υπερανάλυσης εικόνων τύπου NIR και τα αποτελέσματα που παράγονται είναι αρκετά ικανοποιητικά. Σε αντίθεση με τα περισσότερα μοντέλα νευρωνικών δικτύων τα οποία δύσκολα μπορούν να ερμηνευθούν, η αρχιτεκτονική του μοντέλου που αναπτύχθηκε στην παρούσα εργασία καθιστά την ερμηνεία του μοντέλου εύκολη, ενώ παρέχει τη δυνατότητα θεωρητικής μελέτης του μοντέλου με χρήση της θεωρίας αραιών αναπαραστάσεων. Η αραιή αναπαραστάση ενός σήματος παρέχει τη δυνατότητα μιας συμπαγούς κωδικοποίησης της πληροφορίας που περιέχεται στο σήμα. Ο

σχεδιασμός ενός αλγορίθμου προσδιορισμού της αραιής αναπαράστασης ενός σήματος μέσω νευρωνικών δικτύων μειώνει σημαντικά το κόστος των απαιτούμενων υπολογισμών. Συνεπώς, το λαμβανόμενο μοντέλο συνδυάζει τα πλεονεκτήματα των αναλυτικών μοντέλων με τις δυνατότητες γρήγορης επεξεργασίας που προσφέρουν τα νευρωνικά δίκτυα.

Πραγματοποιήθηκαν πειράματα στα οποία ο συντελεστής υπερανάλυσης ήταν διαφορετικός. Χρησιμοποιήθηκαν οι συντελεστές υπερανάλυσης x_2 , x_3 , x_4 και x_6 . Όσο μεγαλύτερος είναι ο συντελεστής υπερανάλυσης τόσο χαμηλότερη ανάλυση έχει η εικόνα εισόδου του δικτύου καθώς η εικόνα εισόδου παράγεται από μια υπο-δειγματοληψία της αρχικής εικόνας υψηλής ανάλυσης βαθμού x_2 , x_3 , x_4 και x_6 αντίστοιχα. Μέσα από τα πειράματα παρατηρήθηκε ότι όσο μεγαλύτερος είναι ο βαθμός υπερανάλυσης τόσο πιο εντυπωσιακά αποτελέσματα πετυχαίνει το δίκτυο της εργασίας μας. Με άλλα λόγια, στις εικόνες εισόδου που έχουν υποστεί χαμηλού βαθμού υπο-δειγματοληψία η εκτιμώμενη εικόνα του δικτύου να μην έχει καλύτερη ποιότητα από ότι η εικόνα εισόδου αλλά δεν είναι τόσο ορατή. Αντίθετα στις εικόνες που έχουν υποστεί μεγάλου βαθμού υπο-δειγματοληψία, η εικόνα που παράγει το δίκτυο έχει πολύ μεγάλη διαφορά στην ποιότητα σε σχέση με την εικόνα εισόδου.

Μελλοντικά το μοντέλο μας μπορεί να εξελιχθεί ακόμα περισσότερο και στον τομέα της ποιότητας αλλά και ως προς τον χρόνο εκπαίδευσης, με χρήση νέων μεθόδων, όπως η υιοθέτηση διαφορετικής μεθόδου εξαγωγής των χαρακτηριστικών των εικόνων όπως οι ροπές, είτε διαφορετικός τρόπος παραγωγής της αραιής κωδικοποίησης είτε με περαιτέρω μελέτη πάνω στον τομέα της αραιής αναπαράστασης. Επιπλέον μπορεί να μελετηθεί η επιρροή των εικόνων καθοδήγησης χρησιμοποιώντας διαφορετικούς χρωματικούς χώρους.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- [Θρα2015] Θράκα, Σοφία. Ανακατασκευή υπερφασματικών εικόνων υψηλής ανάλυσης με τεχνικές compressed sensing. Diss. 2015
- [Παπ2013] Παπαδόπουλος, Αθανάσιος. Συνελικτικά νευρωνικά δίκτυα στην υπολογιστική όραση. Diss. 2016.
- [Τσο2017] Τσουρούνης, Δημήτριος. Βαθιά αραιή κωδικοποίηση. Diss. 2017.
- [Bac2011] Bach, Francis, et al. "Optimization with sparsity-inducing penalties." arXiv preprint arXiv:1108.0775 (2011).
- [Ben2009] Bengio, Yoshua. Learning deep architectures for AI. Now Publishers Inc, 2009.
- [BSH2012] Burger, Harold C., Christian J. Schuler, and Stefan Harmeling. "Image denoising: Can plain neural networks compete with BM3D?." 2012 IEEE conference on computer vision and pattern recognition. IEEE, 2012.
- [BuR2014] Bunch, James R., and Donald J. Rose, eds. Sparse matrix computations. Academic Press, 2014.
- [CaP2009] Candès, Emmanuel J., and Yaniv Plan. "Near-ideal model selection by ℓ_1 minimization." The Annals of Statistics 37.5A (2009): 2145-2177.
- [CoW2005] Combettes, Patrick L., and Valérie R. Wajs. "Signal recovery by proximal forward-backward splitting." Multiscale modeling & simulation 4.4 (2005): 1168-1200.
- [Don2015] Dong, Chao, et al. "Image super-resolution using deep convolutional networks." IEEE transactions on pattern analysis and machine intelligence 38.2 (2015): 295-307.
- [Ela2010] Elad, Michael. Sparse and redundant representations: from theory to applications in signal and image processing. Vol. 2. No. 1. New York: Springer, 2010.
- [GrL2010] Gregor, Karol, and Yann LeCun. "Learning fast approximations of sparse coding." Proceedings of the 27th international conference on international conference on machine learning. 2010.

[Gu2015] Gu, Shuhang, et al. "Convolutional sparse coding for image super-resolution." Proceedings of the IEEE International Conference on Computer Vision. 2015.

[He2016] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[HHW2015] Heide, Felix, Wolfgang Heidrich, and Gordon Wetzstein. "Fast and flexible convolutional sparse coding." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015.

[HRW2014] Hershey, John R., Jonathan Le Roux, and Felix Weninger. "Deep unfolding: Model-based inspiration of novel deep architectures." arXiv preprint arXiv:1409.2574 (2014).

[Hra2015] Hradiš, Michal, et al. "Convolutional neural networks for direct text deblurring." Proceedings of BMVC. Vol. 10. No. 2. 2015.

[JAF2012] Johnson, Justin, Alexandre Alahi, and Li Fei-Fei. "Perceptual losses for real-time style transfer and super-resolution." European conference on computer vision. Springer, Cham, 2016.

[Liu2016] Liu, Ding, et al. "Robust single image super-resolution via deep networks with sparse prior." IEEE Transactions on Image Processing 25.7 (2016): 3194-3207.

[Li2016] Li, Yijun, et al. "Deep joint image filtering." European Conference on Computer Vision. Springer, Cham, 2016.

[Luc2018] Lucas, Alice, et al. "Using deep neural networks for inverse problems in imaging: beyond analytical methods." IEEE Signal Processing Magazine 35.1 (2018): 20-36.

[Mar2019] Marivani, Iman, et al. "Learned multimodal convolutional sparse coding for guided image super-resolution." 2019 IEEE International Conference on Image Processing (ICIP). IEEE, 2019.

[PNS2014] Parikh, Neal, and Stephen Boyd. "Proximal algorithms." Foundations and Trends in optimization 1.3 (2014): 127-239.

[RIG2014] Rish, Irina, and Genady Grabarnik. Sparse modeling: theory, algorithms, and applications. CRC press, 2014.

[RFB2012] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015

[TrW2010] Tropp, Joel A., and Stephen J. Wright. "Computational methods for sparse solution of linear inverse problems." Proceedings of the IEEE 98.6 (2010): 948-958.

[TsD2019] Tsiliogianni Evaggelia, and Nikos Deligiannis. "Deep coupled-representation learning for sparse linear inverse problems with side information." IEEE Signal Processing Letters 26.12 (2019): 1768-1772.

[Sch2013] Schuler, Christian J., et al. "A machine learning approach for non-blind image deconvolution." Proceedings of the IEEE conference on computer vision and pattern recognition. 2013.

[SrG2018] Sreter, Hillel, and Raja Giryes. "Learned convolutional sparse coding." 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2018.

[SMF2015] Starck, Jean-Luc, Fionn Murtagh, and Jalal Fadili. Sparse image and signal processing: Wavelets and related geometric multiscale analysis. Cambridge university press, 2015.

[Son2019] Song, Pingfan, et al. "Multimodal image super-resolution via joint sparse representations induced by coupled dictionaries." IEEE Transactions on Computational Imaging 6 (2019): 57-72.

[Woh2014] Wohlberg, Brendt. "Efficient convolutional sparse coding." 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2014.

[Wan2015] Wang, Zhaowen, et al. "Deep networks for image super-resolution with sparse prior." Proceedings of the IEEE international conference on computer vision. 2015.

[Yan2010] Yang, Jianchao, et al. "Image super-resolution via sparse representation." IEEE transactions on image processing 19.11 (2010): 2861-2873.

[Zei2010] Zeiler, Matthew D., et al. "Deconvolutional networks." 2010 IEEE Computer Society Conference on computer vision and pattern recognition. IEEE, 2010.

[ZhS2005] Zhang, Shuangteng, and Ezzatollah Salari. "Image denoising using a neural network based non-linear filter in wavelet domain." Proceedings.(ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005. Vol. 2. IEEE, 2005.