Εθνικό Μετσόβιο Πολυτεχνείο

Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Τομέας Τεχνολογίας Πληροφορικής και Υπολογιστών

Ανακατασκεύη εικόνας με βαθιά μάθηση

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΓΕΩΡΓΙΟΣ ΧΑΒΑΛΕΣ

Επιβλέπων: Ανδρέας-Γεώργιος Σταφυλοπάτης

Καθηγητής Ε.Μ.Π.

Συνεπιβλέπων: Γεώργιος Σιόλας

Εργαστηριακό και Διδακτικό Προσωπικό Ε.Μ.Π.

Αθήνα, Σεπτέμβριος 2022

Εθνικό Μετσόβιο Πολυτεχνείο

Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Τομέας Τεχνολογίας Πληροφορικής και Υπολογιστών

Ανακατασκεύη εικόνας με βαθιά μάθηση

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΓΕΩΡΓΙΟΣ ΧΑΒΑΛΕΣ

Επιβλέπων: Ανδρέας-Γεώργιος Σταφυλοπάτης

Καθηγητής Ε.Μ.Π.

Συνεπιβλέπων: Γεώργιος Σιόλας

Εργαστηριακό και Διδακτικό Προσωπικό Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 22η Σεπτεμβρίου 2022.

(Υπογραφή)	*(Υπογραφή)*	*(Υπογραφή)*
.....	.....	.....
Ανδρέας-Γεώργιος Σταφυλοπάτης	*Γεώργιος Στάμου*	*Στέφανος Κόλιας*
Καθηγήτης Ε.Μ.Π.	*Καθηγητής Ε.Μ.Π*	*Καθηγητής Ε.Μ.Π*

Αθήνα, Σεπτέμβριος 2022

Εθνικό Μετσόβιο Πολυτεχνείο

Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Τομέας Τεχνολογίας Πληροφορικής και Υπολογιστών

(Υπογραφή)

.....

Γεώργιος Χαβαλές

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

Περίληψη

Η ανακατασκευή εικονών με περιοχές που λείπουν ανήκει στην κατηγορία προβλημάτων της όρασης υπολογιστών. Η λύση του προβλήματος πρέπει να γεμίζει τις περιοχές της εικόνας που λείπουν με πίξελ,έχοντας συνοχή με τις ήδη υπάρχοντες περιοχές και χωρίς να εισάγεται θόρυβος.

Χάρη στην ραγδαία εξέλιξη της μηχανικής μάθησης και την συνεχή εφεύρεση νέων μοντέλων και συστημάτων, έχει σημειωθεί μεγάλη πρόοδος στην αντιμετώπιση του προβλήματος. Σύγχρονες μέθοδοι συνδυάζουν την χρήση συνελικτικών νευρωνικών δικτύων (CNN) με γεννητικά ανταγωνιστικά δίκτυα (GAN). Παρ'ολα αυτά, λόγω του συνυπολογισμού των μη έγκυρων περιοχών κατά την πράξη της συνέλιξης, συχνά παρατηρούνται ασυνέπειες στα τελικά αποτελέσματα, όπως χρωματική διαφορά,θολότητα και θόρυβος.

Στην παρούσα διπλωματική εργασία προτείνουμε την χρήση μιας αρχιτεκτονικής,η οποία χρησιμοποιεί μερικές συνελίξεις. Αυτές κατά την πράξη της συνέλιξης λαμβάνουν υπ'όψην μόνο τα έγκυρα πίξελ της εικόνας και όχι τις περιοχές που λείπουν. Επιπλέον, όσο η είσοδος προχωράει βαθύτερα στο δίκτυο,ανανεώνεται αυτόματα και η μάσκα της εισόδου (οι κρυμμένες περιοχές). Το μοντέλο μας ονομάζεται PConv, εκπαιδεύεται στο σύνολο δεδομένων Places2 και όλες οι εικόνες έχουν διαστάσεις 512x512, δηλαδή είναι υψηλών διαστάσεων για την δυσκολία του προβλήματος. Στο τέλος,παραθέτουμε τα αποτελέσματα του δικτύου μας, τα οποία είναι εξαιρετικά τόσο ποιοτικά όσο και ποσοτικά.

Λέξεις κλειδιά

Μηχανική μάθηση, όραση υπολογιστών, συνελικτικό νευρωνικό δίκτυο, ανακατασκευή εικόνας με περιοχές που λείπουν, μερικώς συνελικτικά νευρωνικά επίπεδα.

Abstract

The problem of image inpainting, or else reconstruction of images with holes, belongs in the field of computer vision. The solution of the problem must fill the missing regions of the image with pixels, having consistency with the surrounding regions and without introducing noise.

Thanks to the rapid evolution of machine learning and the continuous development of new models and systems, there has been big progress in solving the image inpainting problem. Recent methods combine the use of convolutional neural networks (CNN) with generative adversial networks (GAN). However, applying convolutions that include invalid pixels from the holes of the image leads to artifacts and inconsistencies in the results, such as color discrepancy, blurriness and noise.

In the current diploma work, we propose the use of an architecture that uses partial convolutions, where the convolution operation uses only the valid pixels of the image. Furthermore, as the input advances deeper in the network, the mask gets updated and is erased by the decoder part. Our model's name is PConv. We train it in the Places2 dataset and all the images and masks have 512x512 size, which is thought to be high resolution considering the difficulty of the problem. In the end, we cite the results of our model, which are remarkable both in quality and quantitatively.

Key words

machine learning, computer vision, convolutional neural network, image inpainting, partial convolutional neural layers.

Ευχαριστίες

Με αυτήν την διπλωματική εργασία ολοκληρώνονται οι προπτυχιακές σπουδές μου στην σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Εθνικού Μετσοβίου Πολυτεχνείου. Επομένως, θα ήθελα να ευχαριστήσω όλους όσους με υποστήριξαν τόσο στην εκπόνηση της διπλωματικής μου εργασίας όσο και και στο σύνολο των σπουδών μου.

Αρχικά, θα ήθελα να ευχαριστήσω τον κύριο Ανδρέα Γεώργιο Σταφυλοπάτη, που είναι ο επιβλέποντας καθηγητής της διπλωματικής εργασίας, για την ευκαιρία και την εμπιστοσύνη που μου έδωσε να εκπονήσω αυτήν την εργασία. Επιπροσθέτως,θα ήθελα να ευχαριστήσω τον συνεπιβλέποντα κύριο Γεώργιο Σιόλα, για την καθοδήγηση και την βοήθεια που μου προσέφερε,καθώς και για τις ενδιαφέρουσες συζητήσεις που είχαμε τόσο πάνω στο θέμα όσο και στον ευρύτερο επιστημονικό χώρο. Τέλος, θα ήθελα να ευχαριστήσω τους κύριους Στέφανο Κόλλια και Γεώργιο Στάμου, για την τιμή που μου έκαναν με την παρουσία τους στην τριμελή επιτροπή εξέτασης.

Επιπλέον, θα ήθελα να ευχαριστήσω την οικογένεια μου και τους φίλους μου,που με υποστήριξαν όλα τα χρόνια της σχολής και με βοήθησαν να υπερκεράσω κάθε εμπόδιο που παρουσιάστηκε στην πορεία.

Γεώργιος Χαβαλές

Αθήνα, 22 Σεπτεμβρίου 2022

Περιεχόμενα

Κεφάλαιο 1

Εισαγωγή

1.1 Αντικείμενο της διπλωματικής εργασίας

Από το 1959 που ο Άρθουρ Σάμουελ εισήγαγε τον όρο «Μηχανική Μάθηση», αυτή και η τεχνητή νοημοσύνη έχουν εξελιχθεί ραγδαία. Στην σημερινή εποχή τα μοντέλα μηχανικής μάθησης έχουν προοδεύσει σε μεγάλο βαθμό και συχνά υπερτερούν των ανθρώπων στην επίλυση των προβλημάτων. Πλέον, τα προβλήματα της μηχανικής μάθησης χωρίζονται σε πολλές κατηγορίες, όπως είναι η όραση υπολογιστών, η επεξεργασία γλώσσας,η ταξινόμηση ιατρικών εικονών, αναγνώριση ομιλίας, αυτοματοποίηση και άλλα.

Πιο συγκεκριμένα, η όραση υπολογιστών είναι ένα επιστημονικό πεδίο της τεχνητής νοημοσύνης ,το οποίο ασχολείται με την δυνατότητα των υπολογιστών να κατανοήσουν σε υψηλό επίπεδο εικόνες και βίντεο. Η όραση υπολογιστών χωρίζεται σε πολλές υποκατηγορίες.Μερικές από τις πιο γνωστές κατηγορίες είναι η ταξινόμηση εικονών,η αναγνώριση αντικειμένων,η αναβάθμιση εικονών σε ψηλότερη ανάλυση,η αυτόματη οδήγηση οχημάτων και η ανακατασκευή εικονών. Στην τελευταία κατηγορία εντάσσεται και το πρόβλημα του image inpainting.

Το image inpainting είναι το πρόβλημα της ανακατασκευής εικονών από τις οποίες λείπουν περιοχές. Έχει πολλές εφαρμογές στην επεξεργασία εικόνας, όπως την αφαίρεση αντικειμένων ή την επεξεργασία προσώπου. Ο κύριος στόχος αυτού του προβλήματος είναι η σύνθεση οπτικά ρεαλιστικών πίξελ για τις περιοχές που λείπουν,τα οποία ανταποκρίνονται στο νόημα της εικόνας και έχουν συνοχή με τα ήδη υπάρχοντα πίξελ.

Το πρώτο έργο που επιχείρησε να επιλύσει το πρόβλημα αυτό ήταν το [1] και παρουσιάστηκε το 1999. Από τότε η εξέλιξη τόσο των υπολογιστικών συστημάτων όσο και των δικτύων μηχανικής μάθησης έχουν συμβάλει στην παραγωγή μοντέλων που δίνουν εξαιρετικά αποτελέσματα, τα οποία πολλές φορές είναι πανομοιότυπα με την γνήσια εικόνα. Ο τρόπος λέιτουργίας αυτών των δικτύων είναι ο εξής: Εφαρμόζουμε μία μάσκα στην αρχική γνήσια εικόνα, κρύβοντας έτσι περιοχές της.Αυτή είναι η είσοδος του δικτύου. Η έξοδος του θα είναι η πρόβλεψη του, έχοντας γεμίσει τις περιοχές που λείπουν. Ένα τέτοιο παράδειγμα φαίνεται στην εικόνα 1. Η πρώτη εικόνα είναι η γνήσια εικόνα, η δεύτερη εικόνα είναι η μάσκα που εφαρμόζουμε στην γνήσια εικόνα, η τρίτη εικόνα είναι η

είσοδος του δικτύου που καλείται να πραγματοποιήσει το image inpainitng και η τέταρτη εικόνα είναι η πρόβλεψη-έξοδος του δικτύου.

Εικόνα 1 Παράδειγμα image inpainting.

Στην παρούσα διπλωματική εργασία επιχειρούμε την κατασκευή ενός μοντέλου που μπορεί να δόσει ικανοποιητικά αποτελέσματα στο πρόβλημα της ανακατασκευής εικόνων με περιοχές που λείπουν. Αυτό το καταφέρνουμε με την αναπαραγωγή του μοντέλου που προτείνεται στο έργο [2]. Το εκπαιδεύουμε πάνω στο σύνολο δεδομένων Places2, το οποίο περιέχει τεράστιο πλήθος και ποικιλία εικόνων. Την αρχιτεκτονική και τον τρόπο λειτουργίας του μοντέλου μας,καθώς και τα αποτελέσματα του τα αναλύουμε σε επόμενα κεφάλαια.

1.2 Δομή της διπλωματικής εργασίας

Η διπλωματική μας εργασία χωρίζεται σε 6 κεφάλαια. Στο κεφάλαιο 1 έγινε μια εισαγωγή στο πρόβλημα που καλούμαστε να επιλύσουμε. Στο κεφάλαιο 2 καλύπτεται το απαραίτητο θεωρητικό υπόβαθρο δικτύων μηχανικής μάθησης, προκειμένου να γίνουν κατανοητές οι αρχιτεκτονικές που μελετούνται στην συνέχεια. Στο κεφάλαιο 3 εισάγονται και αναλύονται περιληπτικά διάφορες αρχιτεκτονικές από προηγούμενα έργα πάνω στο πρόβλημα του image inpainting. Στο κεφάλαιο 4 παρουσιάζεται η αρχιτεκτονική του μοντέλου μας και ο τρόπος λειτουργίας του. Το κεφάλαιο 5 πραγματεύεται τα σύνολα δεδομένων και μασκών που χρησιμοποιούμε, τον τρόπο εκπαίδευσης και αξιολόγησης του μοντέλου, καθώς και τα πειραματικά αποτελέσματα. Τέλος, στο κεφάλαιο 6 περιέχονται τα συμπεράσματα της διπλωματικής εργασίας καθώς και μελλοντικές προσεγγίσεις, τόσο για το πρόβλημα του image inpainting όσο και για το μοντέλο μας.

Κεφάλαιο 2

Θεωρητικό Υπόβαθρο

2.1 Μηχανική Μάθηση

Η μηχανική μάθηση αποτελεί υποεπίπεδο της επιστήμης των υπολογιστών, που αναπτύχθηκε από την μελέτη της αναγνώρισης προτύπων και της υπολογιστικής θεωρίας μάθησης στην τεχνητή νοημοσύνη. [3]Η μηχανική μάθηση διερευνά τη μελέτη και την κατασκευή αλγορίθμων που μπορούν να μαθαίνουν από τα δεδομένα και να κάνουν προβλέψεις σχετικά με αυτά. Τέτοιοι αλγόριθμοι λειτουργούν κατασκευάζοντας μοντέλα από πειραματικά δεδομένα, προκειμένου να κάνουν προβλέψεις βασιζόμενες στα δεδομένα ή να εξάγουν αποφάσεις που εκφράζονται ως το αποτέλεσμα. Το 1959, ο Άρθουρ Σάμουελ ορίζει πρώτος τη μηχανική μάθηση ως "Πεδίο μελέτης που δίνει στους υπολογιστές την ικανότητα να μαθαίνουν, χωρίς να έχουν ρητά προγραμματιστεί" [4]. Το 1997 ο Tom M. Mitchell έδωσε έναν πιο επίσημο ορισμό που χρησιμοποιείται ευρέως: «Ένα πρόγραμμα υπολογιστή λέγεται ότι μαθαίνει από εμπειρία E ως προς μια κλάση εργασιών T και ένα μέτρο επίδοσης P, αν η επίδοσή του σε εργασίες της κλάσης T, όπως αποτιμάται από το μέτρο P, βελτιώνεται με την εμπειρία E». [5] Αυτός ο ορισμός είναι σημαντικός για τον καθορισμό της μηχανικής μάθησης σε βασικό λειτουργικό πλαίσιο παρά με γνωστικούς όρους, ακολουθώντας έτσι την πρόταση του Alan Turing στην εργασία του «Υπολογιστικές μηχανές και Νοημοσύνη», ότι το ερώτημα αν μπορούν οι μηχανές να σκεφτούν, μπορεί να αντικατασταθεί με το ερώτημα αν μπορούν οι μηχανές να κάνουν αυτό που εμείς (ως σκεπτόμενες οντότητες) μπορούμε να κάνουμε. [6]

Η μηχανική μάθηση συνδέεται και συγχέεται συχνά με την υπολογιστική στατιστική,η οποία αποτελεί κλάδο που επίσης επικεντρώνεται στην πρόβλεψη με την χρήση υπολογιστών. Η μελέτη της μαθηματικής βελτιστοποίησης παρέχει μεθόδους,θεωρία και τομείς εφαρμογής στον τομέα της μηχανικής μάθησης. Επίσης, η μηχανική μάθηση συνδέεται με την εξόρυξη δεδομένων, η οποία αποτελεί παραπλήσιο τομέα επιστήμης και επικεντρώνεται στην εξερευνητική ανάλυση των δεδομένων,γνωστή και ώς μη επιβλεπόμενη μάθηση.

Στο πεδίο της ανάλυσης δεδομένων, η μηχανική μάθηση είναι μια μέθοδος που χρησιμοποιείται για την επινόηση πολύπλοκων μοντέλων και αλγορίθμων που οδηγούν στην πρόβλεψη. Τα αναλυτικά μοντέλα επιτρέπουν στους ερευνητές, τους επιστήμονες δεδομένων, τους μηχανικούς και τους αναλυτές να παράγουν αξιόπιστες αποφάσεις και αποτελέσματα και να αναδείξουν αλληλοσυσχετίσεις μέσω της μάθησης από ιστορικές σχέσεις και τάσεις στα δεδομένα.

2.2 Κατηγορίες Μηχανικής Μάθησης

Οι εργασίες μηχανικής μάθησης χωρίζονται κυρίως σε τρεις μεγάλες κατηγορίες, ανάλογα με την φύση του εκπαιδευτικού «σήματος» ή την «ανατροφοδότηση» που είναι διαθέσιμα σε ένα σύστημα εκμάθησης. Αυτές είναι η επιβλεπόμενη μάθηση (supervised learning), η μη επιβλεπόμενη μάθηση (unsupervised learning) και η ενισχυτική μάθηση (reinforcement learning). Οι τρείς κατηγορίες αναλύονται περεταίρω παρακάτω. [7]

2.2.1 Επιβλεπόμενη μάθηση

Η επιβλεπόμενη μάθηση, αλλίως γνωστή και ως επιτηρούμενη μάθηση και supervised learning, είναι μια κατηγορία της μηχανικής μάθησης,η οποία στοχεύει στον χαρακτηρισμό δεδομένων με βάση κάποια δεδομένα εκπαίδευσης. [8] Τα δεδομένα εκπαίδευσης αποτελούνται από ένα σύνολο παραδειγμάτων τα οποία χρησιμοποιούνται για εκπαίδευση μοντέλων.
Στην επιβλεπόμενη μάθηση, κάθε παράδειγμα αποτελείται από ένα σύνολο εισόδου (συνήθως ένα διάνυσμα από χαρακτηριστικά) και μια επιθυμητή τιμή εξόδου. Οι αλγόριθμοι επιβλεπόμενης μάθησης αναλύουν τα δεδομένα εκπαίδευσης και παράγουν ένα μοντέλο το οποίο μπορεί να χρησιμοποιηθεί για να χαρακτηρίσει νέα παραδείγματα. Το βέλτιστο σενάριο επιτρέπει στον αλγόριθμο να καθορίσει σωστά την ετικέτα της κατηγορίας για άγνωστα μέχρι τώρα παραδείγματα.

Η επιβλεπόμενη μάθηση χωρίζεται κυρίως σε δύο υπο-κατηγορίες, την ταξινόμηση (classification) και την παλινδρόμηση (regression).

- *Ταξινόμηση(Classification):* Η ταξινόμηση είναι το πρόβλημα στο οποίο υπάρχουν μια ή περισσότερες κατηγορίες,με βάση ένα σετ εκπαίδευσης δεδομένων που ανήκουν στις κατηγορίες, και ο αλγόριθμος μας προσπαθεί να προσδιορίσει σε ποιά από τις κατηγορίες αντιστοιχεί μια νέα παρατήρηση.
- *Παλινδρόμηση(Regression):* Η παλινδρόμηση είναι μια ευρέως χρησιμοποιούμενη στατιστική τεχνική μοντελοποίησης για την έρευνα της συσχέτισης μεταξύ μίας εξαρτώμενης μεταβλητής και μιας ή περισσότερων ανεξάρτητων μεταβλητών . Χρησιμοποιείται με σκοπό την εκχώρηση δεδομένων σε μία πραγματική μεταβλητή πρόβλεψης, όπως ισχύει και στην περίπτωση της κατηγοριοποίησης όταν είναι διακριτή, αλλιώς καλείται παλινδρόμηση αν η μεταβλητή είναι συνεχής. Ένα παράδειγμα εφαρμογής της παλινδρόμησης είναι η πρόβλεψη ζήτησης ενός νέου προϊόντος συναρτήσει των δαπανών διαφήμισης του προϊόντος.

2.2.2 Μη Επιβλεπόμενη Μάθηση

Η μη επιβλεπόμενη μάθηση, αλλίως γνωστή και ως μη επιτηρούμενη μάθηση και unsupervised learning, αποτελεί κατηγορία της μηχανικής μάθησης, στόχος της οποίας είναι η ανακάλυψη πιθανής δομής του κόσμου των δεδομένων, μέσω της εκπαίδευσης του αλγορίθμου με μη χαρακτηρισμένα δεδομένα.

Η πιο γνωστή υπο-κατηγορία μη επιβλεπόμενης μάθησης είναι η συσταδοποίηση (clustering).

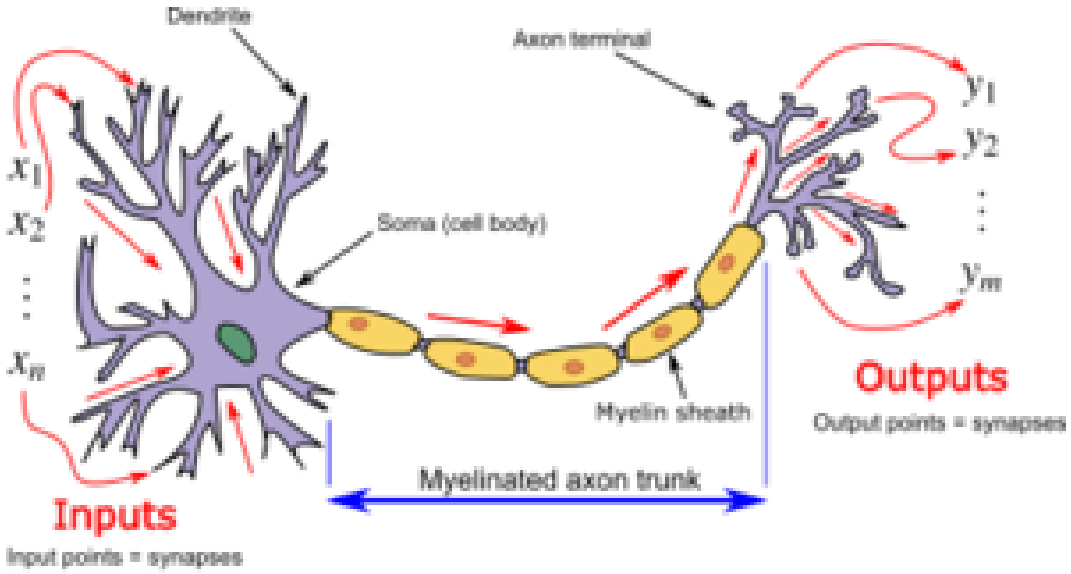
- *Συσταδοποίηση(Clustering):* Η συσταδοποίηση είναι η διαδικασία κατά την οποία ένα σύνολο δεδομένων χωρίζεται σε ομάδες,με βάση την ομοιότητα των χαρακτηριστικών τους.

2.2.3 Ενισχυτική Μάθηση

Η ενισχυτική μάθηση, αλλίως γνωστή και ως reinforcement learning, είναι κατηγορία της μηχανικής μάθησης, στην οποία ο αλγόριθμος μάθησης βασίζεται στην επιβράβευση των επιθυμητών ενεργειών και στην τιμωρία των ανεπιθύμητων. Σκοπός του είναι η μεγιστοποίηση της επιβράβευσης σε κάθε βήμα του αλγορίθμου. Παραδείγματα προβλημάτων ενισχυτικής μάθησης είναι ο έλεγχος κίνησης ρομπότ, η αυτόματη οδήγηση ή παρκάρισμα οχήματος και η εκμάθηση παιχνιδιών. [9]

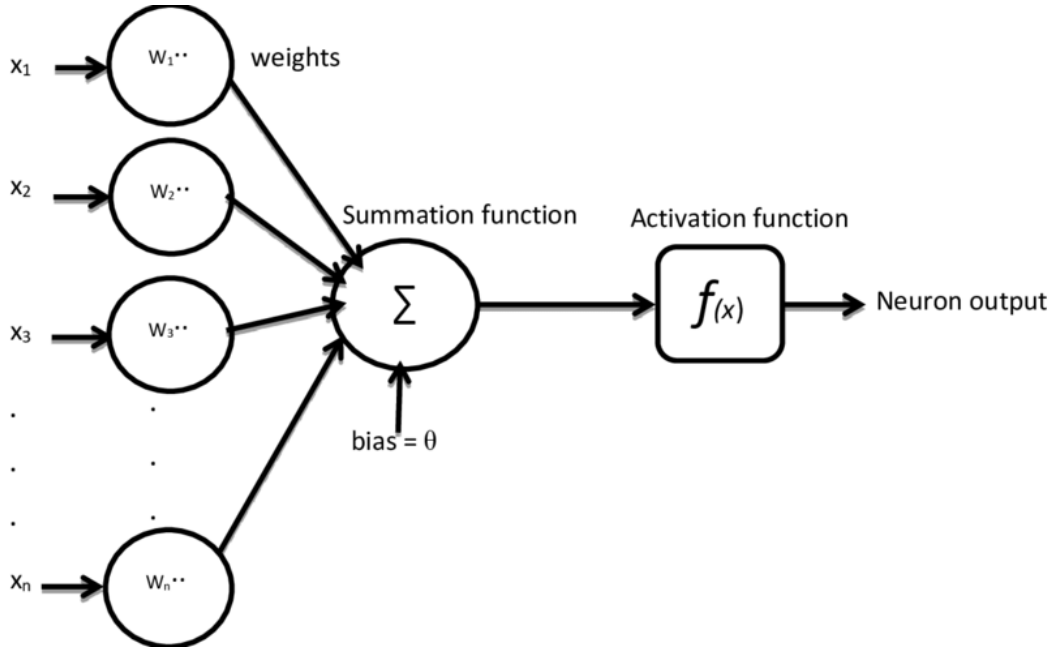
2.3 Νευρωνικά Δίκτυα

Τα νευρωνικά δίκτυα (neural networks), ή αλλιώς τεχνητά νευρωνικά δίκτυα (artificial neural networks), [10] είναι υπολογιστικά συστήματα εμπνευσμένα από τα βιολογικά νευρικά δίκτυα που υπάρχουν στους ζωντανούς οργανισμούς(2). [11]



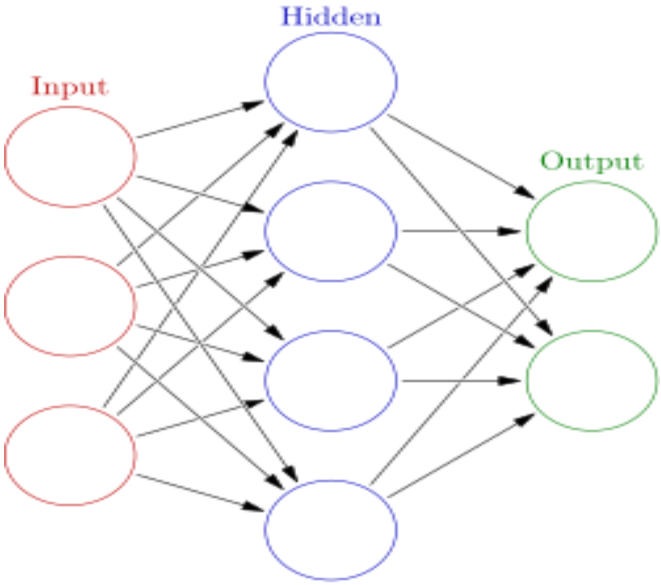
Εικόνα 2 Βιολογικός νευρώνας. Πηγή: https://upload.wikimedia.org/wikipedia/commons/4/44/Neuron3.png

Ένα νευρωνικό δίκτυο απαρτίζεται από συνδεδεμένους κόμβους,οι οποίοι ονομάζονται τεχνητοί νευρώνες και μοντελοποιούν τους νευρώνες ενός βιολογικού εγκεφάλου. Κάθε σύνδεση, σαν τις συνάψεις του βιολογικού εγκεφάλου, έχει την δυνατότητα να εκπέμψει ένα σήμα στους άλλους νευρώνες. Ένας τεχνητός νευρώνας λαμβάνει σήματα,τα επεξεργάζεται και στέλνει σήμα στους συνδεδεμένους σε αυτόν νευρώνες. Οι συνδέσεις ονομάζονται ακμές.Κάθε ακμή έχει ένα βάρος, η τιμή του οποίου προσαρμόζεται κατά την εκπαίδευση.Το βάρος αυξάνει ή μειώνει την ισχύ του σήματος σε μια σύνδεση. Το σήμα σε μία σύνδεση είναι ενας πραγματικός αριθμός και η έξοδος ενός νευρώνα υπολογίζεται από μια μη γραμμική συνάρτηση του αθροίσματος της πόλωσης (bias) και των γινομένων της κάθε εισόδου του με το αντίστοιχο βάρος της σύνδεσης από την οποία προήρθε. Η έξοδος του νευρώνα συγκρίνεται με ένα κατώφλι μέσω της συνάρτησης ενεργοποίησης. Εαν η τιμή της εξόδου είναι μεγαλύτερη από το κατώφλι, ο νευρώνας ενεργοποιείται. Η παραπάνω δομή που περιγράφει έναν νευρώνα φαίνεται και στην εικόνα 3.



Εικόνα 3 Τεχνητός νευρώνας. Πηγή:https://www.researchgate.net/profile/Douw-Boshoff/publication/328733599/figure/fig2/AS:734165188218886@1552050029499/The-structure-of-the-artificial-neuron.png

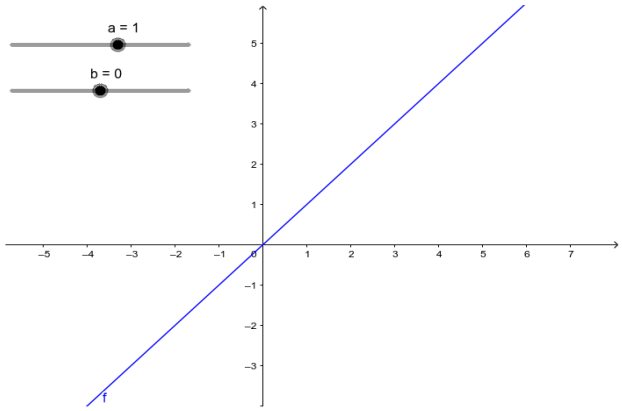
Τυπικά, οι νευρώνες είναι χωρισμένοι σε επίπεδα,όπως βλέπουμε στην εικόνα 4.Τα σήματα εισάγονται στο πρώτο επίπεδο και «ταξιδεύουν» μέχρι και το τελευταίο,το οποίο δίνει την τελική έξοδο.



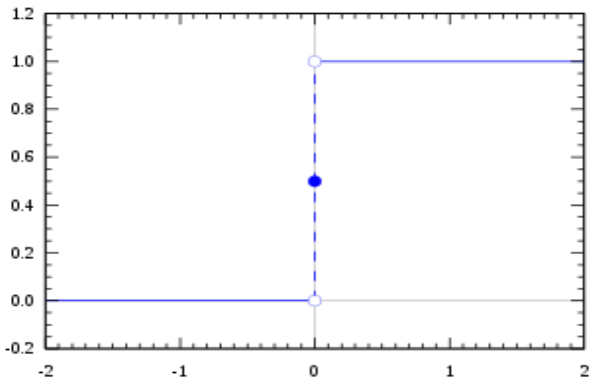
Εικόνα 4Τεχνητό Νευρωνικό Δίκτυο

Παρακάτω παραθέτονται οι πιο γνωστές συναρτήσεις ενεργοποίησης.

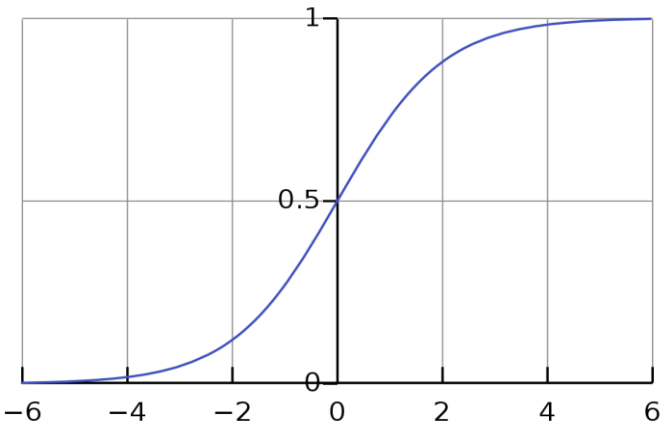
- Γραμμική συνάρτηση: Η γραμμική συνάρτηση ορίζεται από τον τύπο $\varphi(x) = x$.



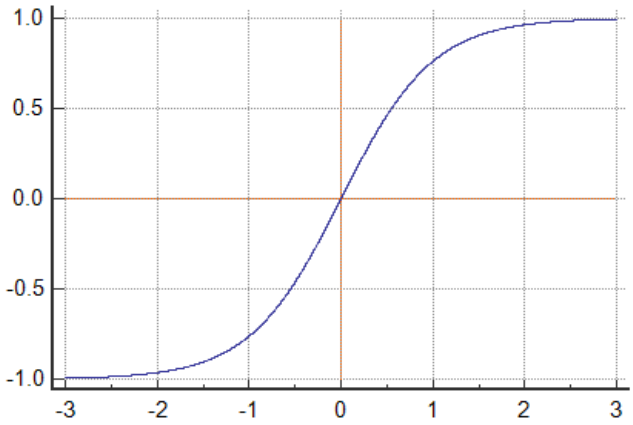
- Βηματική συνάρτηση: Η βηματική συνάρτηση ορίζεται από τον τύπο $\varphi(x) = \begin{cases} x, x \geq 0 \\ 0, x < 0 \end{cases}$.



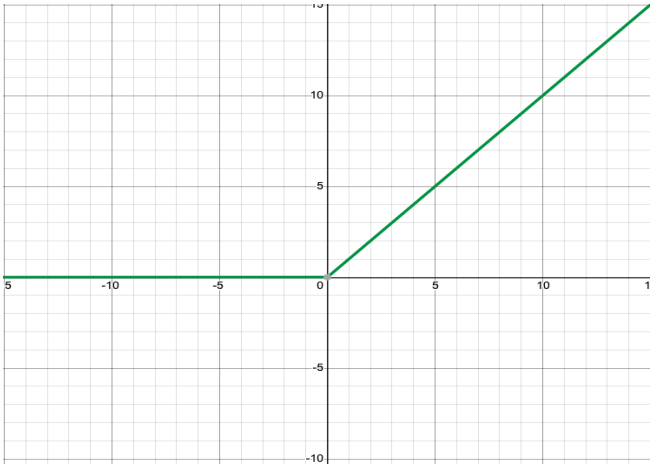
- Σιγμοειδής συνάρτηση: Η σιγμοειδής συνάρτηση μοιάζει με τον χαρακτήρα S και ορίζεται από τον τύπο $\varphi(x) = \frac{1}{1+e^{-x}}$.



- Υπερβολική εφαπτομένη: Η υπερβολική εφαπτομένη ορίζεται από τον τύπο $\varphi(x) = \tanh x$.

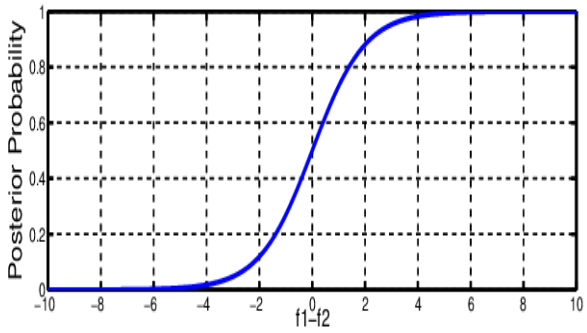


- ReLU(Rectified Linear Unit): Η ReLU χρησιμοποιείται κυρίως στα κρυφά επίπεδα ενός νευρωνικού δικτύου και ορίζεται από τον τύπο $\varphi(x) = \max(0, x)$.



- Softmax: Η softmax χρησιμοποιείται συνήθως στο τελευταίο επίπεδο ενός νευρωνικού δικτύου και ορίζεται από τον τύπο

$$(x_i) = \frac{e^{x_i}}{\sum_{n=1}^{\infty} e^{x_n}} .$$



2.4 Συνελικτικά Νευρωνικά Δίκτυα

Η βαθειά μάθηση αποτελεί μία κλάση αλγορίθμων μηχανικής μάθησης, οι οποίοι χρησιμοποιούν πολλά επίπεδα προκειμένου να εξάγουν σταδιακά υψηλού επιπέδου χαρακτηριστικά από την είσοδο. Μια από τις πιο μοντέρνες και διαδεδομένες κατηγορίες μοντέλων βαθειάς μάθησης είναι τα συνελικτικά νευρωνικά δίκτυα. Τα συνελικτικά νευρωνικά δίκτυα, αλλίως γνωστά ως Convolutional neural networks(CNN), είναι μία κλάση νευρωνικών δικτύων που χρησιμοποιείται κυρίως σε προβλήματα όρασης υπολογιστών. [12] Τα συνελικτικά νευρωνικά δίκτυα εφαρμόζονται σε

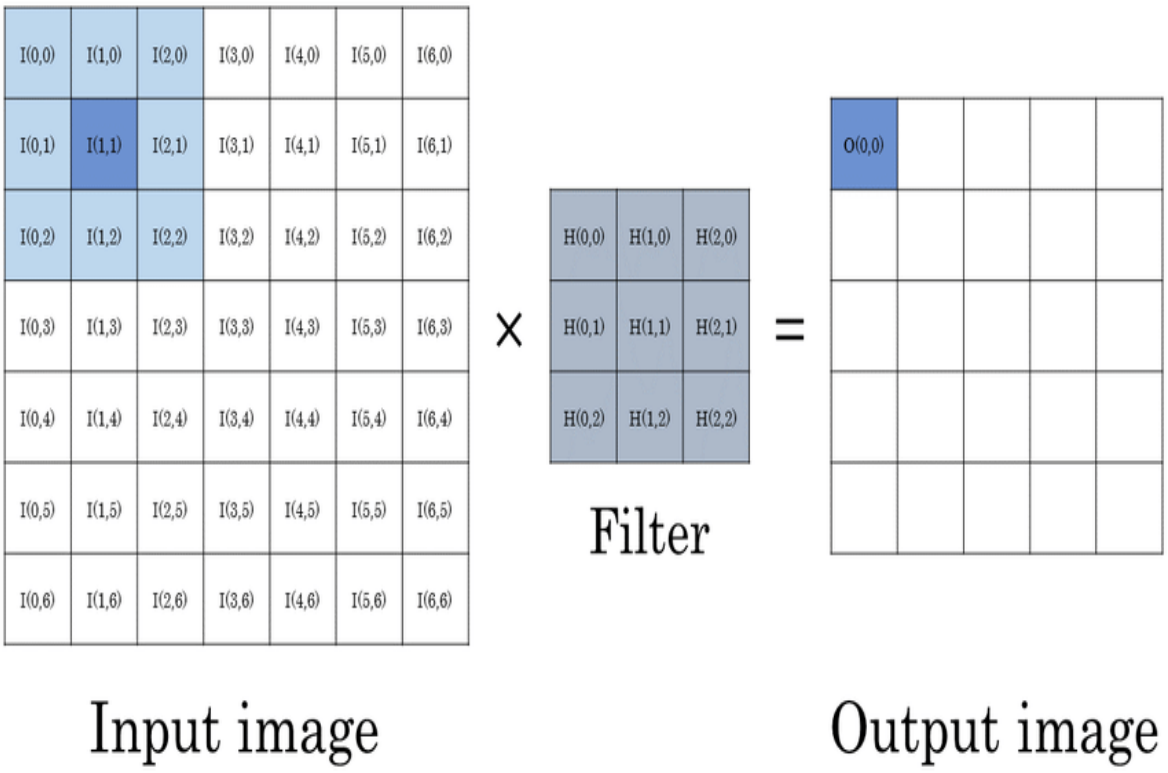
προβλήματα αναγνώρισης εικόνας και βίντεο,ταξινόμησης εικόνας, τιτλοποίησης εικόνας, επεξεργασίας φυσικής γλώσσας κ.α. . Ένα συνελικτικό νευρωνικό δίκτυο εξάγει τα χαρακτηριστικά και τις εξαρτήσεις της εισόδου εφαρμόζοντας κατάλληλα φίλτρα, τα οποία βρίσκονται στο κρυφό επίπεδο και προσαρμόζονται κατά την διαδικασία της εκπαίδευσης.

Τα συνελικτικά δίκτυα έχουν πολλά επίπεδα και μοιάζουν αρκετά με τα κανονικά νευρωνικά δίκτυα.Η διαφορά τους είναι ότι ενώ τα κανονικά νευρωνικά δίκτυα αποτελούνται από πλήρως συνδεδεμένα επίπεδα(fully connected), τα συνελικτικά νευρωνικά δίκτυα αποτελούνται από το επίπεδο της εισόδου,το επίπεδο της εξόδου και το κρυφό επίπεδο,που περιέχει πολλά συνελικτικά επίπεδα, επίπεδα υποδειγματοληψίας, επίπεδα κανονικοποίησης και πλήρως συνδεδεμένα επίπεδα. [13]

2.4.1 Συνελικτικό Επίπεδο

Σε συνελικτικά νευρωνικά δίκτυα που έχουν ως είσοδο εικόνα, η είσοδος έχει διαστάσεις (μήκος εικόνας) x (πλάτος εικόνας) x (πλήθος καναλιών εικόνας) . Αφού περάσει από ένα συνελικτικό επίπεδο, η εικόνα μετατρέπεται σε έναν χάρτη χαρακτηριστικών, ή αλλιώς χάρτη ενεργοποίησης, με διαστάσεις (μήκος χάρτη χαρακτηριστικών) x (πλάτος χάρτη χαρακτηριστικών) x (πλήθος καναλιών χάρτη χαρακτηριστικών). Η μετατροπή από εικόνα σε χάρτη χαρακτηριστικών γίνεται με την βοήθεια ενός φίλτρου. Το φίλτρο, ή αλλιώς πυρήνας, έχει διαστάσεις (μήκος φίλτρου) x (πλάτος φίλτρου) x (πλήθος καναλιών εικόνας). Όπως φαίνεται και στην εικόνα 5, το φίλτρο πολλαπλασιάζεται με ένα κομμάτι της εικόνας ίδιων διαστάσεων και τα αποτελέσματα τοποθετούνται στον χάρτη ενεργοποίησης. Το φίλτρο αρχίζει από την πάνω αριστερή θέση της εικόνας και υπολογίζει την συνέλιξη τους. Στην συνέχεια, το φίλτρο μετακινείται προς τα δεξιά με βήμα ίσο ή μεγαλύτερο του 1.Σε κάθε βήμα υπολογίζεται η συνέλιξη του φίλτρου και του αντίστοιχου κομματιού της εικόνας και τοποθετείται στον χάρτη χαρακτηριστικών. Όταν το φίλτρο φτάσει στο δεξιότερο μέρος της εικόνας, επιστρέφει στο αριστερότερο και κάνει ένα βήμα προς τα κάτω, ίδιας τιμής με το βήμα με το οποίο κινείται δεξιά. Αυτή η διαδικασία συνεχίζεται έως ότου το φίλτρο φτάσει στο κάτω δεξιότερο σημείο της εικόνας. Ένα συνελικτικό επίπεδο ακολουθείται από ένα επίπεδο ενεργοποίησης,το οποίο συνήθως δεν αναφέρεται.Η πιο

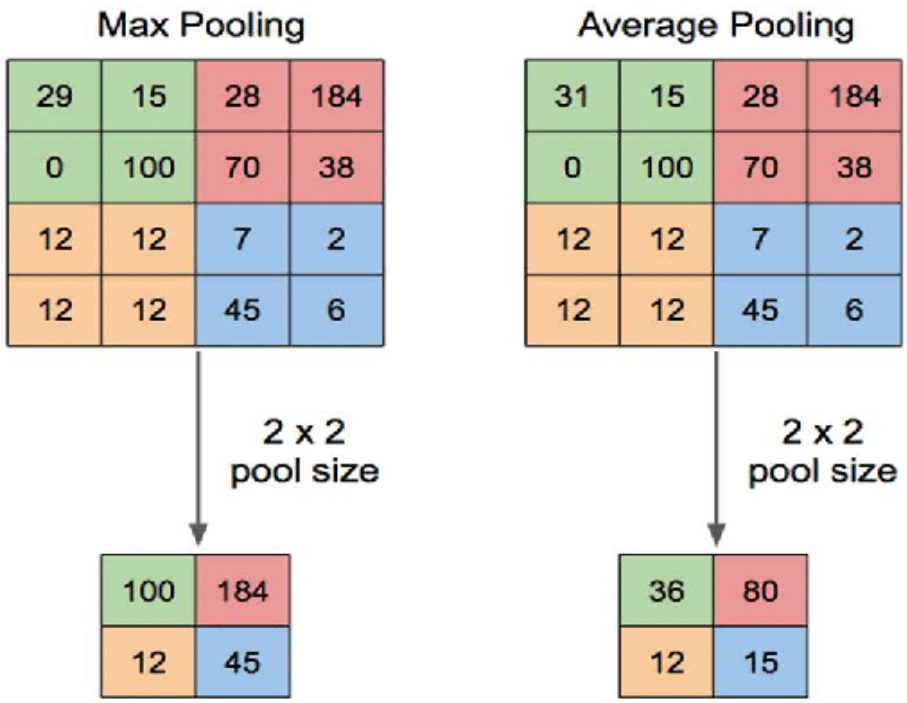
διαδεδομένη συνάρτηση ενεργοποίησης είναι η ReLU.Τα επίπεδα ενεργοποίησης εισάγουν μη-γραμμικότητα στο δίκτυο. ([14], [15])



Εικόνα 5 Παράδειγμα συνέλιξης εισόδου με φίλτρο. Πηγή:https://www.researchgate.net/figure/Image-convolution-with-an-input-image-of-size-7-7-and-a-filter-kernel-of-size-3-3_fig1_318849314

2.4.2 Επίπεδο Υποδειγματοληψίας

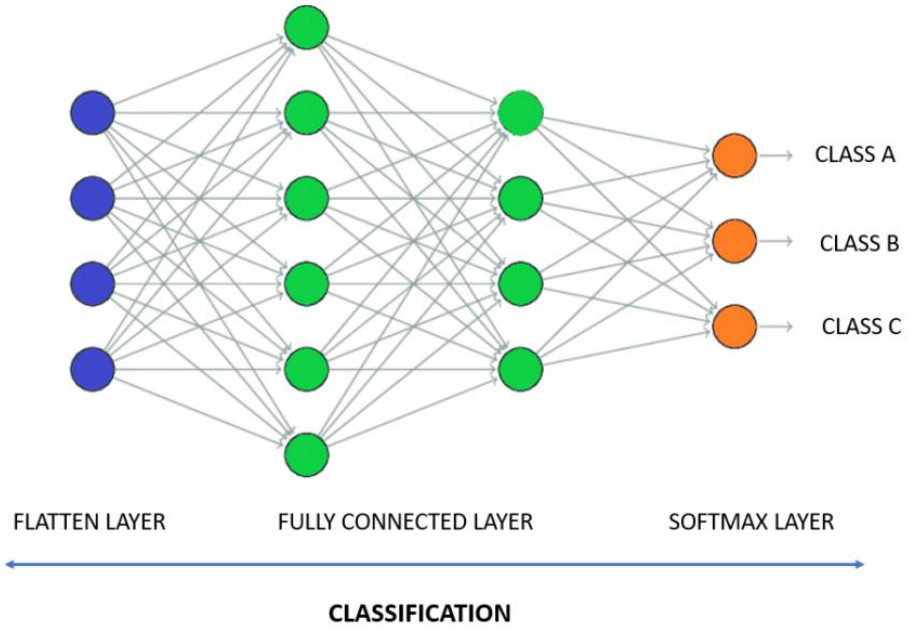
Το επίπεδο υποδειγματοληψίας, ή αλλιώς pooling layer, είναι υπεύθυνα για την μείωση της διάστασης των δεδομένων. Τα επίπεδα υποδειγματοληψίας τοποθετούνται συνήθως ανάμεσα σε συνελικτικά επίπεδα. Όπως φαίνεται και στην εικόνα 6, η είσοδος του επιπέδου υποδειγματοληψίας χωρίζεται σε ορθογώνια, ανάλογα με τις διαστάσεις και το βήμα του φίλτρο του επιπέδου δειγματοληψίας.Για παράδειγμα, στην εικόνα 6 το φίλτρο έχει διαστάσεις 2x2 και βήμα 2.Τα κύρια φίλτρα που χρησιμοποιούνται είναι το φίλτρο μέγιστου όρου(max pooling) και μέσου όρου(average pooling).([14], [16], [17])



Εικόνα 6 Παράδειγμα υποδειγματοληψίας μεγίστου και μέσου όρου. Πηγή:https://www.researchgate.net/figure/Illustration-of-Max-Pooling-and-Average-Pooling-Figure-2-above-shows-an-example-of-max_fig2_333593451

2.4.3 Πλήρως Συνδεδεμένο Επίπεδο

Το πλήρως συνδεδεμένο επίπεδο, ή αλλιώς fully connected layer, είναι συνήθως το τελεύταιο επίπεδο σε ένα συνελικτικό νευρωνικό δίκτυο. Οι νευρώνες ενός πλήρως συνδεδεμένου δικτύου συνδέονται με όλες τις ενεργοποιήσεις του προηγούμενου επιπέδου. Η είσοδος του πλήρως συνδεδεμένου επιπέδου πρέπει πρώτα να έχουν μετατραπεί σε έναν μονοδιάστατο πίνακα.Αυτήν την δουλειά την κάνει το flatten layer που τοποθετείται πριν από το πλήρως συνδεδεμένο επίπεδο.Τα πλήρως συνδεδεμένα επίπεδα ακολουθούνται από ένα επίπεδο ενεργοποίησης,το οποίο δεν αναφέρεται συνήθως. Η πιο συνηθισμένη συνάρτηση ενεργοποίησης είναι η softmax. Στην εικόνα 7 φαίνεται ένα παράδειγμα πλήρως συνδεδεμένου επιπέδου που κατηγοριοποιεί την είσοδο σε 3 κλάσεις. [14]



Εικόνα 7 Παράδειγμα πλήρως συνδεδεμένου επιπέδου. Πηγή: https://indiantechwarrior.com/fully-connected-layers-in-convolutional-neural-networks/

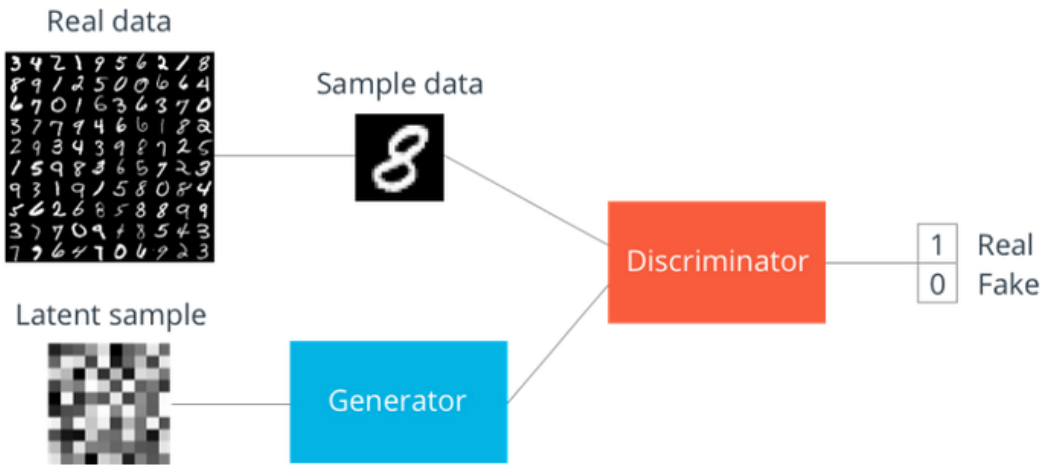
2.4.4 Επίπεδο κανονικοποίησης δέσμης

Τα επίπεδα κανονικοποίησης δέσμης, ή αλλιώς batch normalization layers, είναι επίπεδα που χρησιμοποιούνται για να επιταχύνουν την εκπαίδευση νευρωνικών δικτύων και να την κάνουν πιο «σταθερή», κανονικοποιώντας την είσοδο των επιπέδων. Το επίπεδο κανονικοποίησης δέσμης βοηθάει να μετριαστεί το πρόβλημα «*internal covariate shift*», στο οποίο η αρχικοποίηση των παραμέτρων και η αλλαγές στην κατανομή των εισόδων κάθε επιπέδου επηρρεάζουν τον βαθμό μάθησης του δικτύου. Τα επίπεδα κανονικοποίησης δέσμης τοποθετούνται συνήθως μετά από συνελικτικά επίπεδα και πριν την συνάρτηση ενεργοποίησης. [18]

2.5 Παραγωγικά Ανταγωνιστικά Δίκτυα

Τα Παραγωγικά Ανταγωνιστικά Δίκτυα, ή αλλιώς παραγωγικά αντιπαλικά δίκτυα ή generative adversarial networks (GAN), είναι μία κατηγορία συστημάτων μηχανικής μάθησης που εφευρέθηκε από τον Ian Goodfellow και τους συναδέρφους του το 2014. Ουσιαστικά, δύο νευρωνικά δίκτυα

ανταγωνίζονται σε ένα παιχνίδι. Το δίκτυο μαθαίνει να παράγει νέα δεδομένα με ίδια στατιστικά στοιχεία με το σύνολο εκπαίδευσης. Το παραγωγικό δίκτυο (γεννήτορας - generator) δημιουργεί υποψήφιους ενώ το διευκρινιστικό δίκτυο (διευκρινιστής - discriminator) τους αξιολογεί. Ο γεννήτορας αρχικά παράγει τυχαίο θόρυβο και με βάση την απόκριση που λαμβάνει από τον διευκρινιστή,τελικά μαθαίνει την επιθυμητή κατανομή των δεδομένων. Ο διευκρινιστής συγκρίνει την παραγώμενη κατανομή με την πραγματική κατανομή. Και τα δύο δίκτυα αποτελούνται από πολυεπίπεδα perceptrons. Ο διευκρινιστής εκπαιδεύεται με σκοπό να μεγιστοποιήσει την πιθανότητα ανάθεσης σωστής ετικέτας (αληθινής/ψευδής εικόνας) στην είσοδο του. Ο γεννήτορας εκπαιδεύεται με σκοπό να ελαχιστοποιήσει την πιθανότητα ο διευκρινιστής να αναγνωρίσει την έξοδο του γεννήτορα ως ψεύτικη, δηλαδή για να ξεγελάσει τον διευκρινιστή. Τα δύο δίκτυα εκπαιδεύονται στοχεύοντας να βρεθεί το σύστημα σε ισορροπία, δηλαδή ο γεννήτορας να παράγει δείγματα που να μην διακρίνονται από τα αληθινά και ο διευκρινιστής να μην μπορεί να κρίνει εάν είναι αληθινά ή ψεύτικα. Ένα παράδειγμα παραγωγικού ανταγωνιστικού δικτύου φαίνεται στην εικόνα 8.



Εικόνα 8 Παράδειγμα ενός Παραγωγικού Ανταγωνιστικού Δικτύου. Πηγή: https://towardsai.net/p/l/understand-generative-adversarial-network-gan-in-deep-learning

Κεφάλαιο 3

Σχετικό έργο

Τα υπάρχοντα έργα για το προβλημα του image inpainting μπορούν να χωριστούν κυρίως σε 2 κατηγορίες. Η πρώτη κατηγορία αντιπροσωπεύει παραδοσιακές μεθόδους που χρησιμοποιούν χαρακτηριστικά χαμηλού επιπέδου. Η δεύτερη κατηγορία επιχειρεί να λύσει το πρόβλημα του inpainting με προσεγγίσεις μάθησης, όπως με την εκπαίδευση βαθειών συνελικτικών νευρωνικών δικτύων με σκοπό την πρόβλεψη των πίξελ των περιοχών της εικόνας που λείπουν.

3.1 Παραδοσιακές μέθοδοι

Οι παραδοσιακές μέθοδοι με την σειρά τους χωρίζονται σε 2 κύριες κατηγορίες,τις μεθόδους που βασίζονται σε «διάχυση» (diffusion) και τις μεθόδους που βασίζονται σε «μπάλλωμα» (patch).

Οι μέθοδοι που βασίζονται στην διάχυση μεταδίδουν γειτονικές πληροφορίες στις περιοχές της εικόνας που λείπουν([19], [20], [21]). Στο έργο [20] προτείνεται η χρήση ισόφωτων και λαπλασιανής διαδικασίας διάχυσης για να μεταδωθούν τα πίξελ αυτόματα προς όλες τις κατευθύνσεις στις περιοχές που λείπουν.Για να επιτευχθεί αυτό εισάγειται ένας εκτιμητής ομαλότητας κατά τον υπολογισμό και η μεταδιδόμενη πληροφορία βρίσκεται στην κατεύθυνση των ισοφώτων. Αυτός ο αλγόριθμος είναι αποτελεσματικός σε εικόνες με μικρές ρωγμές λόγω της ανισοτροπικής διάχυσης, αλλά οδηγεί σε ένα εφέ θολώματος όταν υπάρχουν μεγαλύτερες μάσκες. Στο [22] εισάγεται μια γρήγορη προσέγγιση που χρησιμοποιεί έναν γκαουσσιανό πυρήνα διάχυσης με εκτιμήσεις που βασίζονται στην ανοχή θολών περιοχών από την ανθρώπινη όραση σε περιοχές με ακμές υψηλής αντίθεσης. Ο αλγόριθμος περιέχει επαναλαμβανόμενες συνελίξεις της εικόνας με τον γκαουσσιανό πυρήνα. Σε αυτήν την διαδικασία υπολογίζεται ο μέσος όρος των γειτονικών πίξελ ,που είναι αντίστοιχο της ισοτροπικής διάχυσης, χρησιμοποιώντας μία γραμμική ισότητα θερμότητας (ισοτροπική διάχυση) σαν φράγματα διάχυσης για τον χειρισμό της επανασύνδεσης των ακριανών περιοχών. Παρ'όλα αυτά, τα αποτελέσματα είναι λίγο θολά

χωρίς να έχει καθορίσει νωρίτερα ο χρήστης τα φράγματα διάχυσης. Γενικά,οι μέθοδοι διάχυσης εισάγουν ορθά πίξελ από τις οριακές περιοχές μέσα στις περιοχές που λείπουν, γεμίζοντας τες με όμοιο ή ίδιο χρώμα. Είναι ικανές να αντιμετωπίσουν το πρόβλημα του inpainting σε εικόνες με μικρές «γρατζουνιές» και ευθείες γραμμές. Παρ'όλα αυτά, όταν οι μάσκες είναι μεγαλύτερες, οι περιοχές που γεμίζουν είναι θολές. Επομένως, δεν είναι αποτελεσματικές στις περισσότερες περιπτώσεις.

Οι μέθοδοι που βασίζονται σε «μπάλλωμα» (patch) γεμίζουν τις περιοχές που λείπουν (στόχος) αντιγράφοντας πληροφορίες από παρόμοιες περιοχές (πηγή) της ίδιας εικόνας ή μίας συλλογής εικόνων. Στο έργο [1] οι Efros και Leung πρότειναν μια πρωτοποριακή μέθοδο, η οποία έθεσε τα θεμέλια για τις μεθόδους που βασίζονται σε «μπάλλωμα». Αυτή η μέθοδος χρησιμοποιεί μοντελοποίηση MRF για να εντοπίσει την κατανομή των πίξελ και ένα νέο κομμάτι πληροφορίας σχηματίζεται στις περιοχές που λείπουν ψάχνοντας υπάρχοντα κομμάτια πληροφορίας για να βρεθούν τέτοια με παρόμοια πίξελ. Αυτή η διαδικασία συλλαμβάνει όλα τα γειτονικά πίξελ για να δημιουργήσει ένα νέο κομμάτι από πίξελ συνθέτοντας τα ένα πίξελ την φορά.Η διαδικασία είναι επαναληπτική και χρησιμοποιεί ένα μπάλλωμα από γνωστές τιμές πίξελ από ένα ήδη γνωστό μπάλλωμα του προηγούμενου βήματος. Οι περιορισμοί είναι η ασυνέχειες, που προκαλούν το νέο κομμάτι να μην έχει ομοιομορφία. Στο έργο [23] χρησιμοποιούνται δείγματα από υποψήφια κομμάτια για να σχηματίσουν ένα παρόμοιο μπάλλωμα με διαφορετική διάσταση μέσω της συρραφής. Αυτή η τεχνική δειγματοληπτεί τετράγωνα κομμάτια από το αρχικό δείγμα, για να σχηματίσει μια γραμμή από πίξελ που αποτελούν την τελική εικόνα. Για να επιτευχθεί αυτό, το επόμενο κομμάτι που θα συρραφεί στην εικόνα προέρχεται από ένα σετ από υποψήφια κομμάτια. Τα υποψήφια κομμάτια κρατούνται σε SSD, όπου και υπολογίζονται τα σκόρ τους και επιλέγεται αυτό με το καλύτερο. Γενικά, οι μέθοδοι που βασίστηκαν στο [1] είναι υπολογιστικά ακριβές και αργές, αφού πρέπει να υπολογιστούν σκορ ομοιότητας για κάθε ζευγάρι πηγής-στόχου. Το προηγούμενο πρόβλημα αντιμετωπίστηκε στο έργο [24] ,στο οποίο χρησιμοποιείται ένας γρήγορος αλγόριθμος πεδίου πλησιέστερου γείτονα, γνωστός και ώς PatchMatch. Γενικά οι μέθοδοι που βασίζονται σε «μπάλλωμα» υποθέτουν ότι η περιοχή που λείπει μπορεί να βρεθεί κάπου αλλού στην εικόνα.Παρ'ολα αυτά, αυτή η υπόθεση δεν ισχύει πάντα και οι μέθοδοι δυσκολεύονται να ανακατασκευάσουν περιοχές που είναι τοπικά

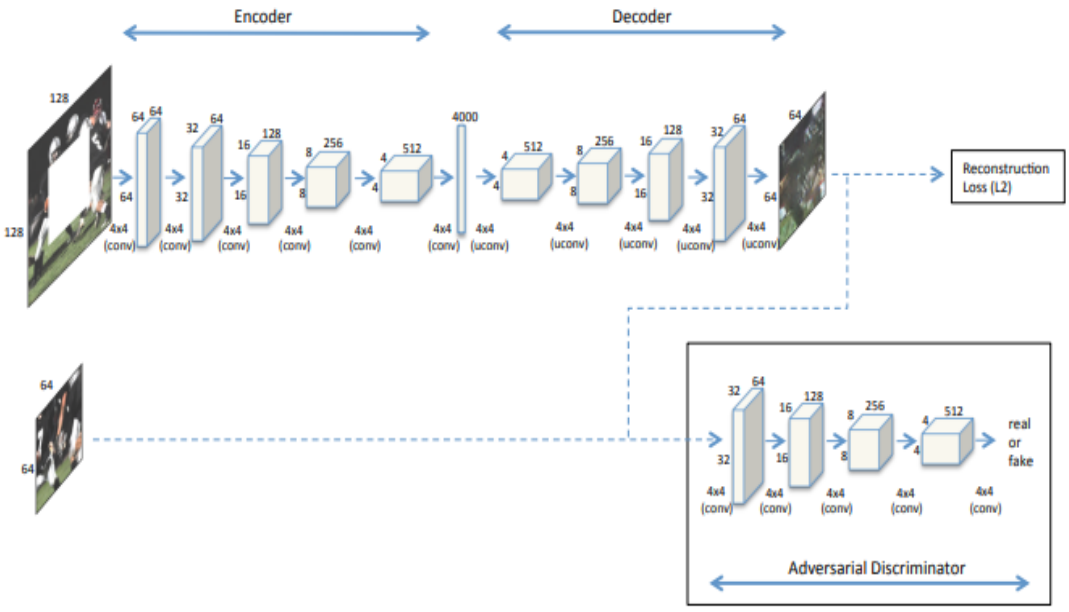
μοναδικές. Επομένως, σε συνδυασμό ότι είναι αργές και υπολογιστικά ακριβές, αυτές οι μέθοδοι δεν είναι ιδανικές.

3.2 Σύγχρονες Μέθοδοι

Σε πιο πρόσφατες έρευνες, η χρήση συνελικτικών νευρωνικών δικτύων (CNN) και παραγωγικών ανταγωνιστικών δικτύων (GAN) έχει γίνει η κύρια μέθοδος για να αντιμετωπιστεί το πρόβλημα του image inpainting. Η χρήση συνελικτικών νευρωνικών δικτύων σε συνδυασμό με ανταγωνιστική εκπαίδευση έχουν παράξει εξαιρετικά αποτελέσματα, με μεγάλη ομοιότητα εισόδου και εξόδου.

Το έργο [25] εισήγαγε την χρήση των συνελικτικών νευρωνικών δικτύων στο πρόβλημα του image inpainting, πλαισιώνοντας το υπολογιστικό κομμάτι μέσα σε ένα στατιστικό πλαίσιο παλινδρόμησης αντί για εκτίμηση πυκνότητας. Ουσιαστικά, αποτελεί ένα συνελικτικό νευρωνικό δίκτυο που μαθαίνει να αφαιρεί τον θόρυβο από την εικόνα με καθαρές εικόνες στις οποίες έχει εισαχθεί θόρυβος κατά την εκπαίδευση. ‘Ομως, αυτή η μέθοδος περιορίζεται μόνο σε εικόνες ενός καναλιού χρώματος και στην αφαίρεση μόνο θορύβου του είδους «salt and pepper», ενώ ταυτόχρονα απαιτεί σημαντικό κόστος υπολογισμού. Η προηγούμενη μέθοδος βελτιώθηκε στο [26] ,το οποίο προτείνει τον συνδυασμό αραιής κωδικοποίησης και βαθειών νευρωνικών δικτύων ώς έναν αυτο-κωδικοποιητή αφαίρεσης θορύβου (denoising auto-encoder), για τον χειρισμό ασυνεπών κατεστραμένων πίξελ στις περιοχές της αρχικής εικόνας που λείπουν. Ο συνδυασμός αυτός ξεπέρασε το εμπόδιο του υπολογιστικού κόστους και εμπόδισε πίξελ με θόρυβο να τροφοδοτηθούν στον αλγόριθμο, αλλά βασίζεται στην επιβλεπώμενη μάθηση και μπορεί να χειριστεί μόνο εικόνες που απαιτούν μικρό βαθμό αφαίρεσης θορύβου.

Οι προηγόυμενες μέθοδοι βασίστηκαν αποκλειστικά στην χρήση συνελικτικών νευρωνικών δικτύων. Οι μέθοδοι που ακολουθούν συνδυάζουν συνελικτικά νευρωνικά δίκτυα με παραγωγικά ανταγωνιστικά δίκτυα. Το συνελικτικό νευρωνικό δίκτυο λειτουργεί συνήθως ως κωδικοποιητής-αποκωδικοποιητής και αποτελεί τον γεννήτορα, ενώ ο διευκρινιστής είναι υπεύθυνος για την βελτίωση των αποτελεσμάτων και των παραμέτρων του γεννήτορα. Στην εικόνα 9 φαίνεται ένα παράδειγμα αρχιτεκτονικής κωδικοποιητή-αποκωδικοποιητή σε συνδυασμό με έναν διευκρινιστή.Η εικονιζόμενη αρχιτεκτονική ανήκει στο έργο [27].



Εικόνα 9 Παράδειγμα αρχικετκονικής κωδικοποιητή-αποκωδικοποιητή σε συνδυασμό με GAN. Πηγή: [27]

Το έργο [27] καινοτόμησε ως προς την ανταγωνιστική εκπαίδευση, ένα απο-άκρη-σε-άκρη δίκτυο βασισμένο στα συνελικτικά νευρωνικά δίκτυα, για την πρόβλεψη του περιεχομένου που λείπει μιας αυθαίρετης περιοχής εικόνας βάσει τον περιβάλλοντα χώρο με ρεαλιστική έξοδο. Η μέθοδος αυτή, γνωστή ώς κωδικοποιητής γενικού πλαισίου (context encoder), χρησιμοποιεί μια αρχιτεκτονική κωδικοποιητή-αποκωδικοποιητή. Ο κωδικοποιητής μετατρέπει εικόνες με διαστάσεις 128x128 με μια κεντρική τρύπα με διαστάσεις 64x64 σε χαρακτηριστικά χαμηλού επιπέδου και ο αποκωδικοποιητής μετατρέπει τα χαρακτηριστικά σε μια εικόνα με διαστάσεις 64x64. Το δίκτυο χρησιμοποιεί L_2 απώλεια ανακατασκευής των πίξελ και παραγωγική ανταγωνιστική απώλεια. Ωστόσο, αυτή η μέθοδος περιορίζεται σε εικόνες χαμηλής ανάλυσης και συχνά οι παραγώμενες εικόνες είναι θολές. Επίσης, ο διευκρινιστής επικεντρώνεται κυρίως στις περιοχές της εικόνας που λείπουν και δεν λαμβάνει υπόψην το συνολικό πλαίσιο της εικόνας. Αργότερα, το [28] βελτιώνει την προηγούμενη μέθοδο εισάγοντας παγκόσμιους και τοπικούς διευκρινιστές ως ανταγωνιστική απώλεια. Επιπρόσθετα, μείωσε τον αριθμό των επιπέδων υποδειγματοληψείας και αντικατέστησε το πλήρως συνδεδεμένο επίπεδο με διευρυμένες συνελίξεις (dilated convolutions), η χρήση των οποίων αυξάνει τα πεδία λήψης των νευρώνων στην έξοδο. Η απόδοση του δικτύου βελτιώθηκε με τις παραπάνω αλλαγές, αλλά ο χρόνος εκπαίδευσης αυξήθηκε σημαντικά (2 μήνες εκπαίδευση με χρήση τεσσάρων GPUs)

εξαιτίας των πολύ αραιών φίλτρων που δημιουργήθηκαν από μεγάλους παράγοντες διεύρυνσης.

Το έργο [29] προτείνει μια προσέγγιση σύνθεσης πολλαπλής κλίμακας νευρωνικού μπαλλώματος (multiscale neural patch synthesis approach). Χρησιμοποιεί ένα ήδη εκπαιδευμένο VGG δίκτυο [30] ώς διευκρινιστή για να βελτιώσει την έξοδο του context encoder [27], ελαχιστοποιώντας την διαφορά χαρακτηριστικών του φόντου της εικόνας, αυξάνοντας όμως την πολυπλοκότητα και το υπολογιστικό κόστος του δικτύου. Το έργο [31] εκπαιδεύει ένα παραγωγικό ανταγωνιστικό δίκτυο που ψάχνει για κωδικοποιήσεις της αλλοιωμένης εικόνας στον χώρο χαρακτηριστικών για να ανακτήσει την περιοχή της εικόνας που λείπει βάσει τα περιβάλλοντα χαρακτηριστικά της εικόνας σαν αναφορά. Έτσι, με αυτήν την κωδικοποίηση ο αλγόριθμος ανακατασκευάζει την αρχική εικόνα. Το έργο [32] εισάγει για πρώτη φορά ένα δίκτυο που αποτελείται από δύο μέρη (coarse-to-fine) και χρησιμοποίησε τις διευρυμένες συνελίξεις του [28].Το πρώτο μέρος είναι ένας κωδικοποιητής-αποκωδικοποιητής που παράγει μια χονδρικά ανακτημένη εικόνα, γεμίζοντας τις περιοχές που λείπουν με προβλέψεις, και το δεύτερο μέρος καθαρίζει το αποτέλεσμα με την βοήθεια των διευρυμένων συνελίξεων. Το δίκτυο χρησιμοποιεί απώλεια ανακατασκευής και δύο απώλειες Wasserstein GAN.

Κεφάλαιο 4

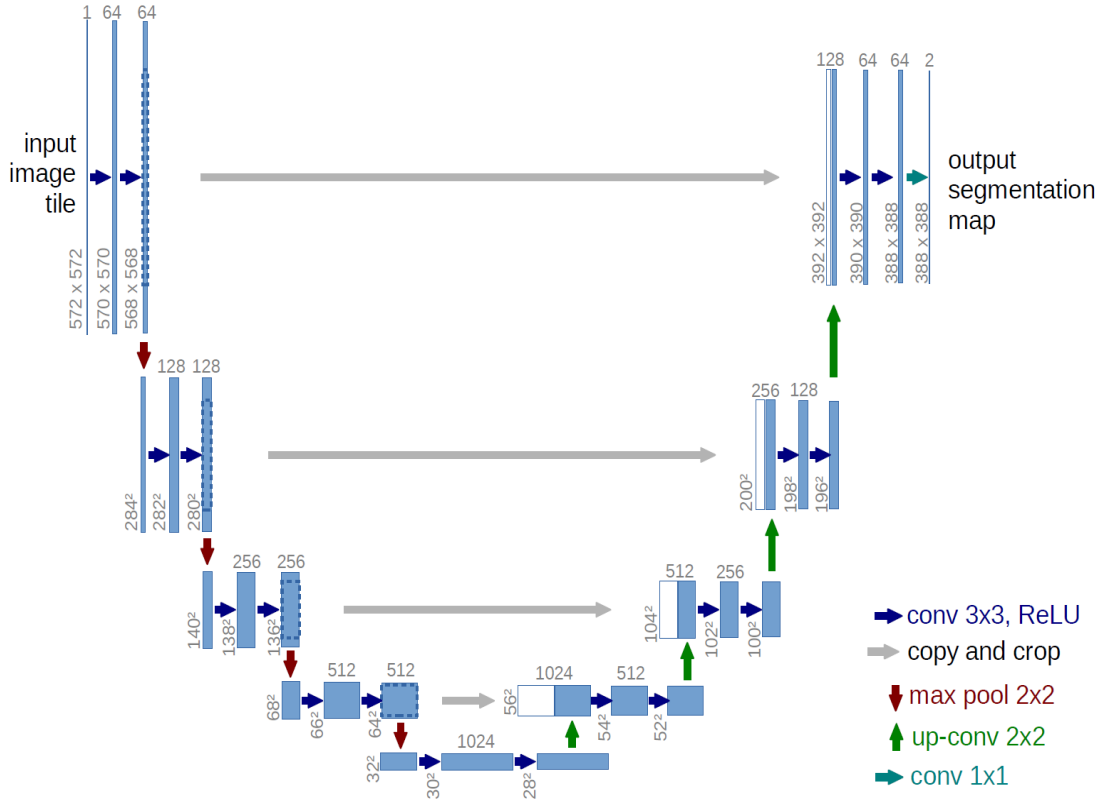
Προσέγγιση

Στην παρούσα διπλωματική εργασία επιχειρούμε να αναπαράξουμε το έργο των Guilin Liu ,Fitsum A. Reda, Kevin J. Shih, Ting-Chun Wang ,Andrew Tao και Bryan Catanzaro, στο οποίο προτείνεται ένα νευρωνικό δίκτυο βασισμένο στο μοντέλο U-net, το οποίο ,αντί για τα συνηθισμένα συνελικτικά νεωρωνικά επίπεδα, χρησιμοποιεί μερικές συνελίξεις (partial convolutions). [2]

4.1 U-Net

Το μοντέλο U-Net είναι ένα συνελικτικό νευρωνικό δίκτυο, το οποίο αναπτύχθηκε για τμηματοποίηση εικόνων βιοϊατρικής. Το δίκτυο είναι βασισμένο στο πλήρες συνελικτικό δίκτυο [33] και τροποποιήθηκε ώστε να προσφέρει πιο ακριβείς τμηματοποιήσεις με λιγότερες εικόνες εκπαίδευσης. Το U-Net αντικατέστησε τα επίπεδα υποδειγματοληψίας (pooling layers) με επίπεδα υπερδειγματοληψίας (upsampling layers), ώστε η έξοδος του δικτύου να είναι πιο ακριβής και να έχει μεγαλύτερη ανάλυση. Επιπλέον, στο κομμάτι της υπερδειγματοληψίας υπάρχει μεγάλος αριθμός από κανάλια χαρακτηριστικών, τα οποία επιτρέπουν στο δίκτυο να διαδόσει πληροφορίες περιβάλλοντος στα επίπεδα υψηλότερης ανάλυσης. Το κομμάτι διαστολής είναι συμμετρικό προς το κομμάτι συστολής του δικτύου και έχει το σχήμα του γράμματος U, από όπου πήρε και το όνομα του.Το δίκτυο δεν περιέχει πλήρως συνδεδεμένα επίπεδα και χρησιμοποιεί μόνο τα έγκυρα κομμάτια κάθε συνέλιξης. Για παράδειγμα, ο χάρτης τμηματοποίησης περιέχει μόνο τα πίξελ για τα οποία το πλήρες πλαίσιο είναι διαθέσιμο κατά την είσοδο της εικόνας στο δίκτυο. Το κομμάτι συστολής ακολουθεί την αρχιτεκτονική ενός τυπικού συνελικτικού νευρωνικού δικτύου. Αποτελείται από επαναλαμβανόμενες εφαρμογές από δύο 3x3 συνελίξεων , κάθε μια από τις οποίες ακολουθείται από μια συνάρτηση ενεργοποίησης ReLU και ένα επίπεδο υποδειγματοληψίας 2x2 μεγίστου όρου (max pooling) με βήμα 2. Σε κάθε βήμα υποδειγματοληψίας, ο αριθμός των καναλιών χαρακτηριστικών διπλασιάζεται. Στο κομμάτι διαστολής, κάθε βήμα αποτελείται από μια υπερδειγματοληψία (upsampling) του χάρτη χαρακτηριστικών, μία 2x2 συνέλιξη, που υποδιπλασιάζει των αριθμό τον καναλιών χαρακτηριστικών, μία ένωση με τον αντίστοιχο κομμένο χάρτη

χαρακτηριστικών από το κομμάτι συστολής και δύο 3x3 συνελίξεις, κάθε μία ακολουθούμενη από μία συνάρτηση ενεργοποίησης ReLU. Το κόψιμο του χάρτη χαρακτηριστικών από το κομμάτι συστολής είναι απαραίτητο λόγω της απώλειας των οριακών πίξελ της εικόνας μετά από κάθε συνέλιξη. Στο τέλος του δικτύου υπάρχει ένα επίπεδο 1x1 συνέλιξης το οποίο αντιστοιχεί το διάνυσμα χαρακτηριστικών στον επιθυμητό αριθμό κλάσεων. Στο σύνολο, το δίκτυο αποτελείται από 23 συνελικτικά νευρωνικά επίπεδα. Η αρχιτεκτονική του δικτύου φαίνεται στην εικόνα 10. Το αριστερό μέρος είναι το κομμάτι συστολής και το δεξί μέρος είναι το κομμάτι διαστολής. [34]



Εικόνα 10 Αρχιτεκτονική του δικτύου U-Net. Πηγή: [34]

4.2 Μερικώς Συνελικτικά Νευρωνικά Επίπεδα

Στο μοντέλο μας χρησιμοποιούμε στακαρισμένες μερικές συνελίξεις και βήματα ανανέωσης μάσκας για την αντιμετώπιση του προβλήματος του image inpainting. Οι δύο παραπάνω λειτουργείες αποτελούν απο κοινού το μερικώς συνελικτικό επίπεδο (partial convolution layer). Έστω ότι **W**

είναι τα βάρη του συνελικτικού φίλτρου και **b** η αντίστοιχη πόλωση. **X** είναι οι τιμές των χαρακτηριστικών/πίξελ του παρόντος παραθύρου συνέλιξης που καλύπτει το φίλτρο και **M** είναι η αντίστοιχη δυαδική μάσκα. Η μερική συνέλιξη σε κάθε παραθύρου δίνεται από την ακόλουθη σχέση:

$$x' = \begin{cases} W^T\ (X\ \odot M)\dfrac{sum(1)}{sum(M)} + b\ , if\ sum(M) > 0 \\ 0, \qquad otherwise \end{cases}$$

όπου το σύμβολο ⊙ δηλώνει τον πολλαπλασιασμό στοιχείων και το 1 έχει ίδιες διαστάσεις με το M αλλά όλα τα στοιχεία είναι 1. Είναι προφανές ότι οι τιμές τις εξόδου βασίζονται μόνο στις περιοχές της εισόδου που δεν καλύπτονται από την μάσκα.Για αυτόν τον λόγο, το επίπεδο ονομάστηκε μερικώς συνελικτικό (partial convolution), αφού χρησιμοποιεί μέρος της εισόδου αντί για ολόκληρη την είσοδο. Ο συντελεστής κλιμάκωσης sum(1)/sum(M) εφαρμόζει κατάλληλη κλιμάκωση για την προσαρμογή των διάφορων πληθών έγκυρων περιοχών.

Μετά από κάθε εφαρμογή μερικής συνέλιξης, ενημερώνουμε την μάσκα ώς εξής: εάν $x' \neq 0$, δηλαδή εάν η συνέλιξη μπόρεσε να θέσει την έξοδο της σε τουλάχιστον μία έγκυρη τιμή εισόδου, τότε θέτουμε αυτήν την περιοχή ως έγκυρη. Αυτό φαίνεται από την ακόλουθη σχέση:

$$m' = \begin{cases} 1, if\ sum(M) > 0 \\ 0, \qquad otherwise \end{cases}$$

Η παραπάνω σχέση μπορεί να εφαρμοστεί σε οποιαδήποτε δομή βαθειάς μάθησης ώς μέρος του εμπρόσθειου περάσματος (forward pass). Μετά από επαρκείς αλλεπάλληλες εφαρμογές μερικής συνέλιξης, οποιαδήποτε μάσκα θα αποτελείται τελικά μόνο απο τιμές 1, εάν η είσοδος είχε έγκυρα πίξελ. Η ενημέρωση της μάσκας πραγματοποιείται χρησιμοποιώντας ένα απλό συνελικτικό δίκτυο που δεν εκπαιδεύεται,με ίσους πυρήνες με το μερικώς συνελικτικό επίπεδο του αντίστοιχου επιπέδου, με όλα τα βάρη του να ισούνται με 1 και χωρίς πόλωση.

4.3 Αρχιτεκτονική Μοντέλου

Το μοντέλο μας υιοθετεί την αρχιτεκτονική του U-Net, αντικαθιστώντας όλα τα συνελικτικά επίπεδα με μερικώς συνελικτικά επίπεδα.Επίσης, στο στάδιο αποκωδικοποιητή/διαστολής χρησιμοποιούμε υπερδειγματοληψία πλησιέστερου γείτονα (nearest neighbor upsampling). Οι σύνδεσμοι μεταξύ κωδικοποιητή και αποκωδικοποιητή ενώνουν δύο χάρτες χαρακτηριστικών και δύο μάσκες αντίστοιχα, που χρησιμοποιούνται στην συνέχεια ως είσοδοι του επόμενου μερικώς συνελικτικού επιπέδου. Η είσοδος του τελευταίου μερικώς συνελικτικού επιπέδου θα περιέχει την ένωση της αρχικής εικόνας εισόδου με μάσκα και την αρχική μάσκα με την έξοδο του προηγούμενου επιπέδου,κάνοντας εφικτή την αντιγραφή έγκυρων πίξελ. Περισσότερες λεπτομέρειες για το μοντέλο μας φαίνονται στον πίνακα της εικόνας 11.

Module Name	Filter Size	# Filters/Channels	Stride/Up Factor	BatchNorm	Nonlinearity
PConv1	7×7	64	2	-	ReLU
PConv2	5×5	128	2	Y	ReLU
PConv3	5×5	256	2	Y	ReLU
PConv4	3×3	512	2	Y	ReLU
PConv5	3×3	512	2	Y	ReLU
PConv6	3×3	512	2	Y	ReLU
PConv7	3×3	512	2	Y	ReLU
PConv8	3×3	512	2	Y	ReLU
NearestUpSample1		512	2	-	-
Concat1(w/ PConv7)		512+512		-	-
PConv9	3×3	512	1	Y	LeakyReLU(0.2)
NearestUpSample2		512	2	-	-
Concat2(w/ PConv6)		512+512		-	-
PConv10	3×3	512	1	Y	LeakyReLU(0.2)
NearestUpSample3		512	2	-	-
Concat3(w/ PConv5)		512+512		-	-
PConv11	3×3	512	1	Y	LeakyReLU(0.2)
NearestUpSample4		512	2	-	-
Concat4(w/ PConv4)		512+512		-	-
PConv12	3×3	512	1	Y	LeakyReLU(0.2)
NearestUpSample5		512	2	-	-
Concat5(w/ PConv3)		512+256		-	-
PConv13	3×3	256	1	Y	LeakyReLU(0.2)
NearestUpSample6		256	2	-	-
Concat6(w/ PConv2)		256+128		-	-
PConv14	3×3	128	1	Y	LeakyReLU(0.2)
NearestUpSample7		128	2	-	-
Concat7(w/ PConv1)		128+64		-	-
PConv15	3×3	64	1	Y	LeakyReLU(0.2)
NearestUpSample8		64	2	-	-
Concat8(w/ Input)		64+3		-	-
PConv16	3×3	3	1	-	-

Εικόνα 11 Αρχιτεκτονική του μοντέλου μας Πηγή: [2]

Ως PConv ορίζεται το μερικώς συνελικτικό επίπεδο με το αντίστοιχο μέγεθος φίλτρου,αριθμό φιλτρών και βήμα του φίλτρου. Τα επίπεδα PConv1-8 αποτελούν τον κωδικοποιητή, ενώ τα επίπεδα PConv9-16 τον αποκωδικοποιητή.Η στήλη με τίτλο BatchNorm δείχνει εάν το αντίστοιχο PConv ακολουθείται από ένα επίπεδο κανονικοποίησης δέσμης (Batch Normalization layer). Η στήλη Nonlinearity υποδεικνύει έαν το επίπεδο κανονικοποίησης δέσμης (εφόσων αυτό χρησιμοποιείται) ακολουθείται από κάποιο επίπεδο μη-γραμμικότητας, καθώς και το είδος του επιπέδου αυτού.Παρατηρούμε ότι στον κωδικοποιητή χρησιμοποιείται η συνάρτηση ReLU, ενώ στον αποκωδικοποιητή η συνάρτης LeakyReLU. Οι σύνδεσμοι μεταξύ κωδικοποιητή και αποκωδικοποιητή φαίνονται από τα Concat1-8, τα οποία ενώνουν την έξοδο του προηγουμένου επιπέδου υπερδειγματοληψίας πλησιέστερου γείτονα με τα αποτελέσματα του αντίστοιχου PConv# που αναφέρεται από τον κωδικοποιητή.

4.4 Συναρτήσεις απώλειας

Οι συναρτήσεις απώλειας (loss functions) του μοντέλου μας στοχεύουν τόσο την ακρίβεια ανακατασκευής των πίξελ όσο και την συνολική σύνθεση, όπως για παράδειγμα πόσο ομαλά η περιοχή της εικόνας που προβλέπει το μοντέλο ταιριάζει με το περιβαλλόν νόημα.

Δοθέντος ως είσοδο μία εικόνα με τρύπα $\mathbf{I_{in}}$, αρχική δυαδική μάσκα **M**, όπου οι περιοχές της εικόνας που λείπουν έχουν τιμή 0, την πρόβλεψη/έξοδο του δικτύου $\mathbf{I_{out}}$ και την ολόκληρη αρχική εικόνα $\mathbf{I_{gt}}$ (ground truth), ορίζουμε τις ακόλουθες συναρτήσεις απώλειας που στοχεύουν σε κάθε πίξελ:

$$L_{hole} = \frac{1}{N_{I_{gt}}} ||(1 - M) \odot (I_{out} - I_{gt})||_1,$$

$$L_{valid} = \frac{1}{N_{I_{gt}}} ||M \odot (I_{out} - I_{gt})||_1.$$

Το $N_{I_{gt}}$ δηλώνει τον αριθμό των στοιχείων στην I_{gt}, δηλαδή $N_{I_{gt}}$= CxHxW, όπου C είναι το μέγεθος των καναλιών της $I_{gt,}$ H είναι το ύψος της I_{gt} και W το πλάτος της I_{gt}. Αυτές είναι οι συναρτήσεις απώλειας στην έξοδο του δικτύου για τα πίξελ των περιοχών που λείπουν και των υπόλοιπων περιοχών αντίστοιχα.

Η επόμενη συνάρτηση απώλειας που ορίζουμε είναι η απώλεια αντίληψης (perceptual loss), η οποία συστάθηκε πρώτα στο έργο [35]:

$$L_{perceptual} = \sum_{p=0}^{P-1} \frac{||\Psi_p^{I_{out}} - \Psi_p^{I_{gt}}||_1}{N_{\Psi_p^{I_{gt}}}} + \sum_{p=0}^{P-1} \frac{||\Psi_p^{I_{comp}} - \Psi_p^{I_{gt}}||_1}{N_{\Psi_p^{I_{gt}}}}.$$

Το I_{comp} είναι ουσιαστικά η έξοδος I_{out}, αλλά θέτοντας τις περιοχές που δεν καλύπτονταν αρχικά από την μάσκα στις αντίστοιχες περιοχές της I_{gt}. Το $N_{\Psi_p^{I_{gt}}}$ είναι ο αριθμός των στοιχείων στην $\Psi_p^{I_{gt}}$. Η απώλεια αντίληψης υπολογίζει τις L^1 αποστάσεις των I_{out} και I_{comp} από την I_{gt} , έχοντας πρώτα προβάλει αυτές τις εικόνες σε χώρους χαρακτηριστικών υψηλότερου επιπέδου χρησιμοποιώντας έναν προεκπαιδευμένο στο ImageNet VGG-16 [36]. Το $\Psi_p^{I_*}$ είναι ο χάρτης ενεργοποίησης του p-οστού επιλεγμένου επιπέδου της εικόνας I_*. Χρησιμοποιούμε μόνο τα επίπεδα pool1, pool2 και pool3 για την απώλεια μας.

Στην συνολική απώλεια του δικτύου περιέχεται και η απώλεια στυλ, η οποία είναι παρόμοια με την απώλεια αντίληψης, έχοντας πρώτα εκτελέσει μια αυτοσυσχέτιση κάθε χάρτη χαρακτηριστικών (gram matrix) πριν εφαρμόσουμε τις $L^{1.}$ Οι δύο τύποι είναι οι ακόλουθοι:

$$L_{style_{out}} = \sum_{p=0}^{P-1} \frac{1}{C_p C_p} ||K_p \left(\left(\Psi_p^{I_{out}}\right)^T (\Psi_p^{I_{out}}) - \left(\Psi_p^{I_{gt}}\right)^T (\Psi_p^{I_{gt}}) \right)||_1$$

$$L_{style_{comp}} = \sum_{p=0}^{P-1} \frac{1}{C_p C_p} ||K_p \left(\left(\Psi_p^{I_{comp}}\right)^T (\Psi_p^{I_{comp}}) - \left(\Psi_p^{I_{gt}}\right)^T (\Psi_p^{I_{gt}}) \right)||_1$$

Τα χαρακτηριστικά υψηλού επιπέδου $\Psi_p^{I_*}$ έχουν διάσταση $(H_pW_p)xC_p$, με αποτέλεσμα έναν πίνακα αυτοσυσχέτισης C_pxC_p (gram matrix). Το K_p είναι ο παράγοντας κανονικοποίησης $1/H_pW_pC_p$ του p-οστού επιλεγμένου επιπέδου.Παρατηρούμε ότι, όπως και στην απώλεια αντίληψης, περιλαμβάνουμε όρους απώλειας και για την I_{out} και για την I_{comp} στην απώλεια στυλ.

Η τελευταία συνάρτηση απώλειας του μοντέλου μας είναι η απώλεια συνολικής παρέκλισης (total variation). Αυτή είναι υπεύθυνη για την ποινή

εξομάλυνσης στο R,όπου το R είναι η περιοχή διαστολής ενός πίξελ στην περιοχή που λείπει από την εικόνα. Ο τύπος της είναι ο ακόλουθος:

$$L_{tv} = \sum_{(i,j)\in R,(i,j+1)\in R} \frac{||I_{comp}^{i,j+1} - I_{comp}^{i,j}||_1}{N_{I_{comp}}} + \sum_{(i,j)\in R,(i+1,j)\in R} \frac{||I_{comp}^{i+1,j} - I_{comp}^{i,j}||_1}{N_{I_{comp}}}$$

Το $N_{I_{comp}}$ δηλώνει τον αριθμό των στοιχείων στην $I_{comp.}$

Η συνολική συνάρτηση απώλειας είναι ο συνδυασμός των παραπάνω απωλειών και δίνεται από τον ακόλουθο τύπο:

$$L_{total} = L_{valid} + 6L_{hole} + 0.05L_{perceptual} + 120\left(L_{style_{out}} + L_{style_{comp}}\right) + 0.1L_{tv} \ .$$

Τα βάρη των όρων απώλειας αποφασίστηκαν από τους συγγραφείς κάνοντας μία αναζήτηση υπερπαραμέτρων σε 100 εικόνες.

Κεφάλαιο 5 Πειραματική διαδικασία και αποτελέσματα

5.1 Σύνολα Δεδομένων

Οι συγγραφείς για την εκπαίδευση και την αξιολόγηση του μοντέλου χρησιμοποιούν 3 σύνολα δεδομένων, το ImageNet [37], το Places2 [38] και το CelebA-HQ [39]. Εμείς στα πλαίσια αυτής της διπλωματικής εργασίας χρησιμοποιούμε μόνο το σύνολο δεδομένων Places2, λόγω περιορισμένων υπολογιστικών πόρων.

Το Places2 σύνολο δεδομένων σχηματίστηκε ακολουθώντας τις αρχές της ανθρώπινης οπτικής γνώσης. Στόχος του είναι το χτίσιμο ενός πυρήνα οπτικής γνώσης η οποία μπορεί να χρησιμοποιηθεί για την εκπαίδευση συστημάτων τεχνητής νοημοσύνης για υψηλού επιπέδου προβλήματα όρασης, όπως είναι η αναγνώριση αντικειμένων, η ταξινόμηση, η κατανόηση του θέματος της σκηνής και άλλα. Συνολικά αποτελείται από 10 εκατομμύρια εικόνες,οι οποίες είναι ταξινομημένες σε πάνω απο 400 κατηγορίες, όπως δωμάτια, δρόμους και γήπεδα. Κάθε κατηγορία αποτελείται από 5000 έως 30000 εικόνες εκπαίδευσης. Λόγω της μεγάλης ποικιλίας κατηγοριών και εικονών, τα συνελικτικά νευρωνικά δίκτυα που εκπαιδεύονται μαθαίνουν βαθειά χαρακτηριστικά σκηνών για δίαφορα προβλήματα αναγνώρισης σκηνών.

Εμείς,συγκεκριμένα χρησιμοποιούμε το Places365-Standard σύνολο δεδομένων, το οποίο είναι ήδη χωρισμένο σε σύνολα εκπαίδευσης και αξιολόγησης. Το σύνολο εκπαίδευσης περιέχει 1.8 εκατομμύριο εικόνες από 365 κατηγορίες σκηνών,ενώ το σύνολο αξιολόγησης περιέχει 328 χιλιάδες εικόνες. [40]

Εκτός από τα σύνολα εκπαίδευσης και αξιολόγησης του Places2, χρησιμοποιούμε το σύνολο δεδομένων ακανόνιστων μασκών, το οποίο παρέχεται από τους συγγραφείς στην ιστοσελίδα της Nvidia. Παλαιότερα έργα δημιουργούσαν τρύπες στις εικόνες εισόδου αφαιρώντας τυχαία ορθογώνιες περιοχές της εικόνας. Όμως,αυτός ο τρόπος δημιουργίας μασκών είναι ελλιπής, καθώς δεν υπάρχει ποικιλία στα σχήματα και στα μεγέθη των μασκών. Έτσι, οι συγγραφείς επιχείρησαν να φτιάξουν μάσκες συλλέγοντας τυχαίες τρύπες από αυθαίρετα σχήματα. Βρήκαν τέτοια σχήματα παίρνοντας από δύο συνεχόμενα πλαίσια από βίντεο περιοχές που κρύφτηκαν ή περιοχές που ήταν κρυμμένες και εμφανίστηκαν. Συνολικά παρήγαγαν 55,116 μάσκες για την εκπαίδευση και 24,866 μάσκες για την αξιολόγηση. Κατά την εκπαίδευση, πριν την είσοδο της εικόνας στο

μοντέλο, επιλέγεται μία μάσκα τυχαία από τις 55,116 και εφαρμόζονται πάνω της τυχαία διαστολή,περιστροφή και κόψιμο της εικόνας. Για το σύνολο μασκών αξιολόγησης, οι συγγραφείς πήραν τις 24,866 μάσκες και τους εφάρμοσαν τυχαία διαστολή, περιστροφή και κόψιμο. Τελικά, χώρισαν τις μάσκες σε 6 κατηγορίες ανάλογα με την αναλογία μάσκας και εικόνας, τις (0.01, 0.1], (0.1, 0.2], (0.2, 0.3], (0.3, 0.4], (0.4, 0.5] και (0.5, 0.6]. Επιπρόσθετα, κάθε κατηγορία, περιέχει 1000 μάσκες με τρύπες κοντά στα όρια της εικόνας και 1000 μάσκες με τρύπες τουλάχιστον 50 πίξελ μακριά από τα όρια της εικόνας. Συνολικά, επομένως το σύνολο αξιολόγησης περιέχει 12,000 μάσκες. Στα πλαίσια αυτής της διπλωματικής εργασίας, χρησιμοποιούμε μόνο το σύνολο μασκών αξιολόγησης και για την εκπαίδευση και για την αξιολόγηση, το οποίο είναι διαθέσιμο από τους συγγραφείς. [41] Όλες οι εικόνες και οι μάσκες εκπαίδευσης και αξιολόγησης έχουν μέγεθος 512x512.

5.2 Εκπαίδευση

Η εκπαίδευση του δικτύου μας πραγματοποιήθηκε στο ARIS. ARIS είναι το όνομα του ελληνικού υπερυπολογιστή, ο οποίος παρέχεται και λειτουργείται από το GRNET στην Αθήνα. Ο ARIS αποτελείται από 532 υπολογιστκούς κόμβους. Εμείς εκπαιδεύσαμε το δίκτυο μας σε μία GPU NVIDIA Tesla k40m . Το μοντέλο μας χρησιμοποιεί τον Adam για την βελτιστοποίηση του δικτύου. Ως είσοδο στο μοντέλο μας χρησιμοποιούμε δέσμες εικονών (batch) των 6.

Όπως έχουμε προαναφέρει, στο δίκτυο μας χρησιμοποιούμε επίπεδα κανονικοποίησης δέσμης (batch normalization). Παρ'όλα αυτα, οι τρύπες των μασκών παρουσιάζουν πρόβλημα για την κανονικοποίηση δέσμης, επειδή η μέση τιμή και η διασπορά που θα υπολογιστούν θα περιλαμβάνουν και πίξελ από τις τρύπες της μάσκας, οπότε θα είχε νόημα να μην τα λαμβάνουμε υπ'όψην. Ωστόσο, οι μάσκες ανανεώνονται και γεμίζουν όσο πιο βαθειά πάμε στο μοντέλο μας και συνήθως έχουν εξαφανιστεί εντελώς όταν φτάνουν στο στάδιο του αποκωδικοποιητή.

Έτσι, χρησιμοποιούμε κανονικοποίηση δέσμης με την παρουσία των τρυπών στην αρχική εκπαίδευση έχοντας βαθμό μάθησης 0.0002 (learning rate). Στην συνέχεια, τελειοποιούμε (finetune) το μοντέλο μας εκπαιδεύοντας το με βαθμό μάθησης 0.00005 και παγώνοντας τις παραμέτρους κανονικοποίησης δέσμης στο κομμάτι του κωδικοποιητή του

δικτύου μας. Τα επίπεδα κανονικοποίησης δέσμης παραμένουν ενεργά στο κομμάτι του αποκωδικοποιητή του δικτύου. Με αυτόν τον τρόπο, αποφεύγουμε το πρόβλημα με τους λανθασμένους υπολογισμούς μέσης τιμής και διασποράς, ενώ ταυτόχρονα το μοντέλο μας επιτυγχάνει ταχύτερη σύγκλιση. Η αρχική εκπαίδευση του δικτύου μας με βαθμό μάθησης 0.0002 διήρκησε 24 μέρες,ενώ η τελειοποίηση του δικτύου με βαθμό μάθησης 0.0005 διήρκησε 5 μέρες.

5.3 Μετρικές Αξιολόγησης

Είναι γνωστό ότι τα νευρωνικά δίκτυα ή τα γεννητικά ανταγωνιστικά δίκτυα τα οποία ανακατασκευάζουν εικόνες είναι δύσκολο να αξιολογηθούν στο αποτέλεσμα του χωρίς την συμμετοχή ανθρώπινου παράγοντα,καθώς επίσης και λόγω της ύπαρξης πολλών πιθανών λύσεων. Παρ'όλα αυτά, τα περισσότερα έργα πάνω στο image inpainting χρησιμοποιούν τις μετρικές L1 error, PSNR και SSIM για την αξιολόγηση των μοντέλων τους. Για αυτόν τον λόγο, και εμείς χρησιμοποιούμε αυτές.

Η μετρική L1 error, γνωστή και ως L1 απώλεια ή απώλεια απόλυτου σφάλματος, είναι η απόλυτη διαφορά μεταξύ της πρόβλεψης ενός μοντέλου και της πραγματικής τιμής. Υπολογίζεται η διαφορά μεταξύ πραγματικής εικόνας και πρόβλεψης και στο τέλος βρίσκεται η μέση τιμή όλων των διαφορών. Στόχος μας είναι η ελαχιστοποίηση της μετρικής L1. Ο τύπος είναι ο ακόλουθος:

$$L1 = \frac{\sum_{i=1}^{n} |y_{gt} - y_{predicted}|}{n}$$

,όπου y_{gt} είνα η πραγματική εικόνα, $y_{predicted}$ είναι η πρόβλεψη του μοντέλου μας και n είναι το σύνολο των εικονών του συνόλου αξιολόγησης.

Η μετρική PSNR, ολογράφως Peak Signal-to-Noise Ratio, είναι μια έκφραση της αναλογίας μεταξύ της μέγιστης δυνατής ισχύος ενός σήματος και της ισχύος του θορύβου παραμόρφωσης που επηρεάζει την ποιότητα της αναπαράστασης της εικόνας. Η αξιολόγηση της αναπαράστασης μιας

εικόνας είναι υποκειμενική και από άτομο σε άτομο διαφέρει η επιλογή καλύτερης εικόνας. Επομένως, είναι αναγκαία η καθιέρωση ποσοτικών μετρικών για την αξιολόγηση των εξόδων μοντέλων κατασκευής και ενίσχυσης εικονών. Η PSNR χρησιμοποιείται κυρίως για την αξιολόγηση της ποιότητας ανακατασκευής εικονών μετά από συμπίεση. Σε αυτήν την περίπτωση, το σήμα είναι τα αρχικά δεδομένα και ο θόρυβος είναι η επιπλέον πρόβλεψη του μοντέλου. Όσο μεγαλύτερη τιμή PSNR έχουμε,τόσο καλύτερη είναι η πρόβλεψη του μοντέλου. Ο τύπος της μετρικής PSNR είναι ο ακόλουθος:

$$PSNR = 20\log_{10}(\frac{MAX_{y_{gt}}}{\sqrt{MSE}})$$

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}||y_{gt}(i,j) - y_{predicted}(i,j)||^2$$

όπου m και n είναι οι διαστάσεις της εικόνας, MSE είναι το μέσο τετραγωνικό σφάλμα μεταξύ πραγματικής εικόνας και πρόβλεψης και $MAX_{y_{gt}}$ είναι η μέγιστη τιμή σήματος που υπάρχει στην αρχική μας εικόνα. [42]

Η μετρική SSIM, ολογράφως structural similarity index measure, είναι μια αντιληπτική μετρική που υπολογίζει τον ποιοτικό υποβιβασμό μίας εικόνας κατά την επεξεργασία της, όπως συμπίεση ή στην περίπτωση μας πρόβλεψη περιοχών που λείπουν. Σε αντίθεση με τις μετρικές L1 και PSNR, η SSIM βασίζεται στις ορατές δομές μέσα στην εικόνα.Δεν υπολογίζει απόλυτα σφάλματα, αλλά αντιλαμβάνεται την πληροφορία των δομών από τις εξαρτήσεις των πίξελ με τα άλλα κοντινά τους πίξελ.Η SSIM υπολογίζεται σε διάφορα παράθυρα μιας εικόνας. Έστω ότι έχουμε το παράθυρο x της κανονικής εικόνας και το παράθυρο y της πρόβλεψης,τα οποία αντιστοιχούν στις ίδιες συντεταγμένες. Ο τύπος της SSIM ορίζεται από τον ακόλουθο τύπο:

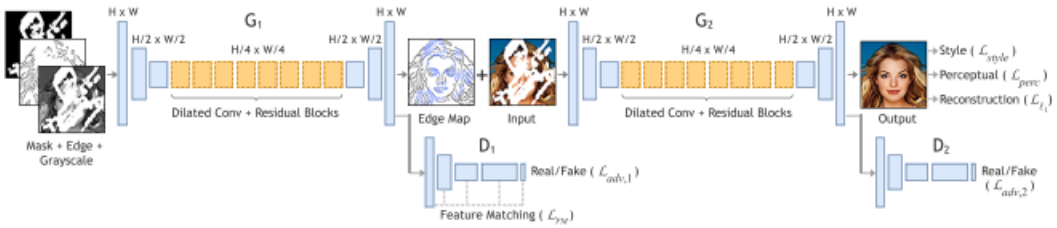
$$\mathrm{SSIM}(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

όπου μ_x,σ_x .μ_y,σ_y είναι οι μέσες τιμές και οι διασπορές των x και y αντίστοιχα, σ_{xy} η συνδιασπορά των x και y, $c_1 = (k_1L)^2$, $c_2 = (k_2L)^2$,με k_1=0.01, k_2 = 0.03 και L το δυναμικό εύρος των τιμών των πίξελ. Στόχος μας είναι η μεγιστοποίηση αυτής της μετρικής. [43]

Πιστεύεται από τον ακαδημαϊκό χώρο ότι η SSIM μετρική έχει καλύτερη απόδοση από τις υπόλοιπες μετρικές που υπολογίζουν απλώς το μέσο σφάλμα μεταξύ των πίξελ. [44] Παρ'όλα αυτά, στα προβλήματα ανακατασκευής εικόνας συνήθως χρησιμοποιούνται και οι τρείς μετρικές που αναφέραμε για την ποσοτική αξιολόγηση του μοντέλου και την σύγκριση με άλλα μοντέλα.

5.4 Ποσοτικές αξιολογήσεις

Σε αυτήν την ενότητα θα συγκρίνουμε την απόδοση του μοντέλου μας στις μετρικές της ενότητας 5.3 με άλλα state-of–the-art μοντέλα της εποχής του. Πιο συγκεκριμένα, ένα από τα μοντέλα είναι το δίκτυο CA του έργου [32], το οποίο αποτελεί ένα coarse-to-fine δίκτυο όπως είδαμε και στην ενότητα 3. Ένα άλλο μοντέλο είναι το δίκτυο EC (EdgeConnect) του έργου [45]. Αυτό το έργο χρησιμοποιεί δύο γεννήτορες. Ο πρώτος γεννήτορας προβλέπει τον χάρτη ακμών της εικόνας-εισόδου,ενώ ο δεύτερος παίρνει τον χάρτη ακμών,μαζί με την αρχική είσοδο και δίνει το τελικό αποτέλεσμα. Το μοντέλο EC φαίνεται αναλυτικότερα στην εικόνα 12. Τέλος, το τρίτο μοντέλο με το οποίο θα συγκρίνουμε το δίκτυο μας είναι το δίκτυο RN του έργου [46]. Το RN εισήγαγε πρώτο την ιδέα της κανονικοποίησης ανά περιοχή (region normalization), η οποία χρησιμοποιεί διαφορετική κανονικοποίηση για τις έγκυρες περιοχές της εικόνας και τις περιοχές της μάσκας, σε αντίθεση με την κανονικοποίηση δέσμης. Το RN χρησιμοποιεί μόνο τον γεννήτορα του EC, έχοντας αντικαταστήσει τα επίπεδα κανονικοποίησης δέσμης με επίπεδα κανονικοποίησης ανά περιοχή.



Εικόνα 12 Το μοντέλο EC. Πηγή: [45]

Τα αποτελέσματα αξιολόγησης του κάθε μοντέλου τα πήραμε από το αντίστοιχο έργο και είναι αυτά του συνόλου δεδομένων Places2. Τα αποτελέσματα φαίνονται στον πίνακα 1.

	CA	EC	RN	Το μοντέλο μας
L1 (%) ↓	8.6	3.86	2.7	**1.38**
SSIM ↑	0.818	0.823	0.823	**0.9414**
PSNR ↑	18.91	21.75	**25.10**	21.29

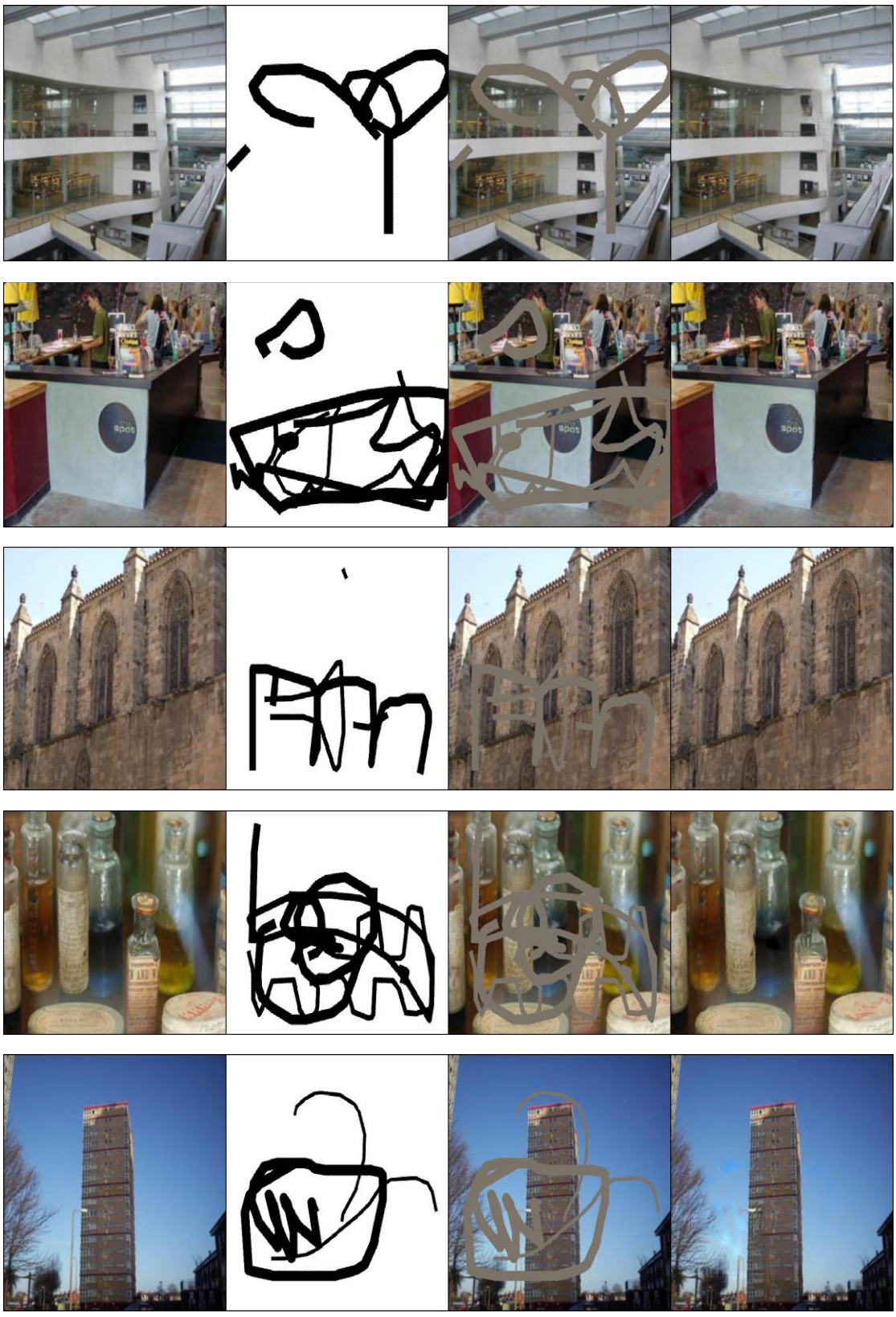
Πίνακας 1 Ποσοτικά αποτελέσματα για το σύνολο Places2 των μοντέλων Contextual Attention (CA), EdgeConnect (EC), Region Normalization (RN) και του δικού μας μοντέλου. Τα καλύτερα αποτελέσματα κάθε μετρικής έχουν επισημανθεί με σκούρο χρώμα. Το σύμβολο ↑ σημαίνει ότι η μεγαλύτερη τιμή νικάει, ενώ το ↓ ότι η μικρότερη τιμή νικάει.

Παρατηρούμε ότι το δικό μας μοντέλο πετυχαίνει την καλύτερη απόδοση ως προς τις μετρικές L1 error και SSIM, με αρκετή διαφορά μάλιστα από τα υπόλοιπα μοντέλα. Παρ'όλα αυτά, το μοντέλο RN υπερτερεί των υπολοίπων ως προς την μετρική PSNR. Αξίζει να σημειωθεί ,ωστόσο, ότι στα αντίστοιχα έργα αναφέρονται ψηλότερες τιμές PSNR για το μοντέλο μας (PConv), το οποίο επανεκπαίδευσαν οι συγγραφείς του κάθε έργου. Επομένως, η διαφορά στις τιμές της μετρικής PSNR του μοντέλου μας μπορεί να οφείλεται στον τρόπο υπολογισμού της μετρικής.

5.5 Ποιοτικά αποτελέσματα

Σε αυτήν την ενότητα παραθέτουμε διάφορες εξόδους του μοντέλου μας. Σε κάθε παράδειγμα, η πρώτη εικόνα είναι η πραγματική εικόνα, η δεύτερη εικόνα είναι η μάσκα, η τρίτη εικόνα είναι η πραγματική εικόνα με την

μάσκα εφαρμοσμένη πάνω της και αποτελεί την είσοδο του δικτύου και η τέταρτη εικόνα είναι η πρόβλεψη του δικτύου μας.



Παρατηρούμε στα παραπάνω αποτελέσματα ότι το δίκτυο μας δίνει εξαιρετκά αποτελέσματα, ακόμα και σε μεγάλες αναλογίες μάσκας-εικόνας. Βέβαια,σε κάποια σημεία μπορεί να υπάρχουν ασυνέπειες χρώματος όσο μεγαλώνει η μάσκα. Παρ'όλα αυτά, δεν έχουμε πάντα καλά αποτελέσματα. Όπως είναι λογικό, όσο μεγαλώνει η μάσκα, τόσο δυσκολότερο είναι το μοντέλο να κατανοήσει το νόημα της εικόνας. Ειδικά, όταν υπάρχουν μεγάλα κομμάτια μάσκας ενωμένα μεταξύ τους, όσο η είσοδος προχωράει πιο βαθειά στο δίκτυο, τα μερικώς συνελικτικά επίπεδα δεν θα αποκτούν καθόλου πληροφορία για τα κομμάτια μάσκας

αφού θα έχουν έξοδο 0 για αυτά τα σημεία. Έτσι, στην έξοδο του κωδικοποιητή, ο χάρτης χαρακτηριστικών υψηλού επιπέδου δεν θα έχει ουσιαστικές πληροφορίες για τις περιοχές που λείπουν από την αρχική είσοδο. Παρακάτω παραθέτουμε μερικά παραδείγματα με μεγάλα κομμάτια μάσκας. Φαίνεται ότι σε αυτά τα κομμάτια εισάγεται κυρίως θόρυβος από το δίκτυο παρά ουσιαστική πληροφορία.

Κεφάλαιο 6

Συμπεράσματα και Μελλοντικές Κατευθύνσεις

Στην παρούσα διπλωματική εργασία μελετήσαμε διάφορες αρχιτεκτονικές, οι οποίες αντιμετωπίζουν το πρόβλημα της ανακατασκευής εικόνων με περιοχές που λείπουν, ή αλλίως το πρόβλημα του image inpainting. Καταλήξαμε στην υιοθέτηση και αναπαραγωγή του μοντέλου PConv από το έργο [41].

Το μοντέλο μας ,όπως είδαμε και στην ενότητα 5.5, δίνει ικανοποιητικά αποτελέσματα, παρά την έλλειψη υπολογιστικών πόρων και την ελλιπή εκπαίδευση σε σχέση με τις αναφορές των συγγραφέων στο αντίστοιχο έργο. Σε συνδυασμό με το γεγονός ότι το σύνολο δεδομένων Places2 περιέχει σχεδόν δύο εκατομμύρια εικόνες με πληθώρα διαφορετικών σκηνών, κάτι που κάνει ιδιαίτερα δύσκολη την εκπαίδευση του μοντέλου μας, μπορούμε να συμπεράνουμε ότι αυτό είναι αποδοτικό με καλές επιδόσεις.

Όπως αναφέραμε και στην ενότητα 5.5, το μοντέλο μας δυσκολεύεται να γεμίσει μεγάλα κομμάτια μάσκας, λόγω του τρόπου λειτουργίας των μερικώς συνελικτικών επιπέδων και της ανανέωσης μάσκας μετά απο κάθε επίπεδο. Αξίζει ,όμως, να σημειωθεί ότι τα περισσότερα μοντέλα ,που αντιμετωπίζουν το πρόβλημα ανακατασκευής εικόνας με περιοχές που λείπουν, δεν δίνουν καλά αποτελέσματα όταν έρχονται αντιμέτωπα με μεγάλες μάσκες. Αυτό είναι λογικό, καθώς όταν η αναλογία μάσκας-εικόνας είναι μεγάλη κατά την είσοδο, τα μοντέλα αδυνατούν να εξάγουν αρκετά χαρακτηριστικά υψηλού επιπέδου από τις έγκυρες περιοχές της εικόνας στα βαθειά επίπεδα για ολική ανακατασκευή, με αποτέλεσμα οι περιοχές που γεμίζουν τις τρύπες να είναι θολές και γεμάτες θόρυβο. Ωστόσο, πραγματικές εφαρμογές που χρησιμοποιούν τέτοια μοντέλα, όπως είναι η εφαρμογές επεξεργασίας εικόνων (Photoshop), εφαρμόζουν μικρές μάσκες στις εικόνες στην πλειοψηφία των περιπτώσεων. Επομένως, οι κακές επιδόσες του μοντέλου μας σε μεγάλες μάσκες δεν επηρρεάζει την αποτελεσματικότητα του στα περισσότερα προβλήματα πραγματικού κόσμου.

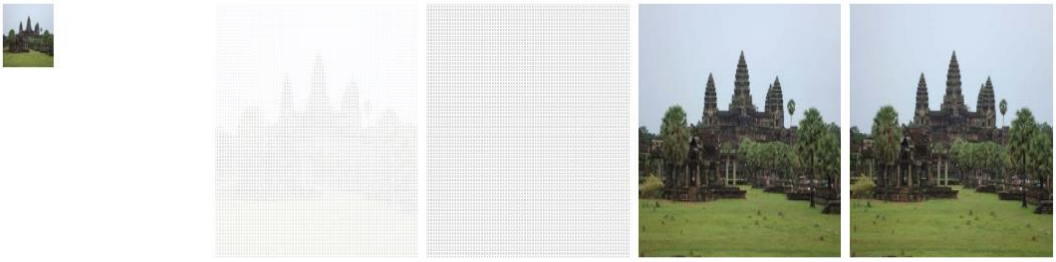
Μία επέκταση του μοντέλου μας θα ήταν η εκπαίδευση σε περισσότερα σύνολα δεδομένων. Πιο συγκεκριμένα, το μοντέλο μας μπορεί να εκπαιδευτεί σε κάποιο σύνολο δεδομένων που περιέχει εικόνες προσώπων, όπως είναι το CelebA-HQ [39]. Η εκπαίδευση με σύνολο δεδομένων

εικόνες προσώπων επιτρέπει στο μοντέλο να μάθει την κατανομή των προσώπων εύκολα. Κάτι τέτοιο θα έχει χρησιμότητα σε εφαρμογές επεξεργασίας εικονών (photoshop), οι οποίες γίνονται όλό και πιο δημοφιλείς, καθώς πολλοί πλέον επεξεργάζονται τις προσωπικές τους φωτογραφίες πριν τις ανεβάσουν στα μέσα κοινωνικής δικτύωσης.

Μία άλλη τροποποίηση του μοντέλου μας που μπορεί να βελτιώσει περεταίρω την απόδοση του είναι η αντικατάσταση των επιπέδων κανονικοποίησης δέσμης με επίπεδα κανονικοποίησης ανά περιοχή από το έργο [46]. Αυτά χωρίζονται σε δύο κατηγορίες, το βασικό και το εκπαιδεύσιμο. Το βασικό επίπεδο κανονικοποίησης ανά περιοχή (Basic RN) κανονικοποιεί τα πίξελ από τις έγκυρες περιοχές της εικόνας και από τις περιοχές της μάσκας ξεχωριστά, κάτι που αντιμετωπίζει το πρόβλημα αλλαγής της μέσης τιμής και της διαφοράς (mean and variance shift). Το εκπαιδεύσιμο επίπεδο κανονικοποίησης ανά περιοχή (Learnable RN) μαθαίνει να εντοπίζει αυτόματα πιθανές διεφθαρμένες και αδιαφθορες από μάσκα περιοχές για ξεχωριστή κανονικοποίηση και βελτιώνει την ένωση των δύο περιοχών εκτελώντας ολικό μετασχηματισμό συγγένειας (global affine transformation). Το βασικό επίπεδο κανονικοποίησης ανά περιοχή εφαρμόζεται στα πρώτα στάδια του δικτύου, δηλαδή στο στάδιο κωδικοποίησης, ενώ το εκπαιδεύσιμο επίπεδο κανονικοποίησης ανά περιοχή στα τελευταία στάδια του δικτύου, ή αλλιώς στο στάδιο αποκωδικοποίησης. Οι συγγραφείς του έργου υποστηρίζουν ότι τα επίπεδα κανονικοποίησης ανά περιοχή που κατασκεύασαν μπορούν να αντικαταστήσουν τα επίπεδα κανονικοποίησης δέσμης σε οποιοδήποτε δίκτυο πάνω στο image inpainting, δίνοντας καλύτερα αποτελέσματα,τόσο ποιοτικά όσο και ποσοτικά.

Επιπροσθέτως, το μοντέλο μας μπορεί να επεκταθεί από το πρόβλημα του image inpainting σε πρόβλημα αναβάθμισης εικόνων σε ψηλότερη ανάλυση.Αυτό μπορεί να επιτευχθεί μετακινώντας τα πίξελ της αρχικής εικόνας και προσθέτοντας τρύπες ενδιάμεσα. Έστω ότι έχουμε μια αρχική εικόνα I με ύψος H και πλάτος W και παράγοντα αναβάθμισης K, φτιάχνουμε την εικόνα I’ με διαστάσεις K*H και K*W, η οποία θα αποτελεί την είσοδο του δικτύου, ώς εξής: τοποθετούμε κάθε πίξελ της εικόνας I με συντεταγμένες (x,y) στην θέση (K*x + K/2, K*y + K/2) στην εικόνα I’ και θέτουμε 1 την τιμή αυτής της θέσης στην μάσκα. Έτσι, ουσιαστικά έχουμε μια εικόνα με διαστάσεις K*H και K*W, η οποία έχει πολλές μικρές τρύπες μεταξύ των έγκυρων πίξελ της αρχικής εικόνας. Όπως είναι γνωστό, το πρόβλημα μας είναι αποδοτικό στον χειρισμό μικρών τρυπών,οπότε μπορει να ανταπεξέλθει στο πρόβλημα

αναβάθμισης εικονών σε υψηλότερη ανάλυση. Στην εικόνα 13 βλέπουμε ένα παράδειγμα λειτουργίας του μοντέλου μας στο παραπάνω πρόβλημα. Η πρώτη εικόνα είναι η αρχική εικόνα χαμηλής ανάλυσης. Η δεύτερη και η τρίτη εικόνα είναι η είσοδος στο μοντέλο μας και η μάσκα αντίστοιχα, οι οποίες κατασκευάστηκαν με τον τρόπο που αναφέραμε. Η τέταρτη εικόνα είναι η έξοδος του δικτύου μας,ενώ η τελευταία είναι η γνήσια εικόνα



Εικόνα 13 Παράδειγμα λειτουργίας του μοντέλου στο πρόβλημα αναβάθμισης σε ψηλότερη ανάλυση. Πηγή: [41]

Βέβαια, πέρα από την βελτίωση του μοντέλου μας, υπάρχει μεγάλο πλήθος μοντέλων τα οποία ειδικεύονται στο image inpainting. Επιπλέον, η όραση υπολογιστών είναι ένας κλάδος που αναπτύσσεται με καταιγιστικούς ρυθμούς και συνεχώς δημοσιεύονται νέα έργα που καταρρίπτουν τα προηγούμενα τόσο ως προς τις ποσοτικές μετρικές όσο και στα ποιοτικά αποτελέσματα. Επομένως, είναι κατανοητό ότι οι επιλογές μοντέλων για το image inpainting είναι άφθονες και βελτιώνονται συνεχώς.

Βιβλιογραφία

References

[1] T. K. L. A. A. Efros, "Texture synthesis by nonparametric sampling," 1999.

[2] F. A. R. K. J. S. T.-C. W. ,. T. ,. B. C. Guilin Liu, "Image Inpainting for Irregular Holes Using," 2018.

[3] W. L. Hosch, "machine learning," [Online]. Available: http://www.britannica.com/EBchecked/topic/1116194/machine-learning.

[4] P. Simon, Too Big to Ignore: The Business Case for Big Data, 2013.

[5] T. Mitchell, Machine Learning, 1997.

[6] S. Harnad, "«The Annotation Game: On Turing (1950) on Computing, Machinery, and Intelligence»," 2008.

[7] S. J. Russell, " Artificial intelligence : a modern approach," 2002.

[8] A. R. A. T. Mehryar Mohri, " Foundations of Machine Learning," 2012.

[9] Ι. Βλαχάβας, Τεχνητή Νοημοσύνη, 2006.

[10] L. Hardesty, "Explained: Neural networks," *MIT News Office,* 2017.

[11] Z. Yang and Z. Yang, " Comprehensive Biomedical Physics," 2014.

[12] M. Valueva, N. Nagornov, P. Lyakhov, G. Valuev and N. Chervyakov, "Application of the residue number system to reduce hardware costs of the convolutional neural network implementation", 2020.

[13] "Convolutional neural network," [Online]. Available: https://en.wikipedia.org/wiki/Convolutional_neural_network.

[14] A. Rosebrock, "Convolutional Neural Networks (CNNs) and Layer Types," 14 May 2021. [Online]. Available: https://pyimagesearch.com/2021/05/14/convolutional-neural-networks-cnns-and-layer-types/.

[15] R. Venkatesan and B. Li, Convolutional Neural Networks in Visual Computing: A Concise Guide, 2017.

[16] D. Scherer, A. C. Müller and S. Behnke, ""Evaluation of Pooling Operations in Convolutional Architectures for Object Recognition"," 2010.

[17] D. Ciresan, U. Meier and J. Schmidhuber, "Multi-column deep neural networks for image classification.," 2012.

[18] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift", 2015.

[19] M. B. V. C. G. S. a. C. Ballester, "Filling-in by joint interpolation of vector fields," 2001.

[20] G. S. V. C. a. C. B. M. Bertalmio, "Image Inpainting," 2000.

[21] A. Telea, "An image inpainting technique based on the fast marching method," 2004.

[22] C. M.-S. Richard M.M.O.B.B., "Imaging and Image Processing," 2001.

[23] W. T. F. A. A. Efros, "Image quilting for texture synthesis and transfer," 2001.

[24] E. S. A. F. D. B. G. C. Barnes, "Patchmatch: A randomized correspondence algorithm," 2009.

[25] J. V. a. S. S., "Natural image denoising with convolutional networks," 2009.

[26] X. L. a. C. E. Xie J., "Image denoising and inpainting with deep neural networks," 2012.

[27] P. K. J. D. T. D. A. A. E. D. Pathak, "Context encoders: Feature learning by inpainting," 2016.

[28] E. S.-S. H. I. S. Iizuka, "Globally and locally consistent image completion," 2017.

[29] X. L. Z. L. E. S. O. W. H. L. C. Yang, "High-resolution image inpainting using multi-scale neural," 2016.

[30] K. S. a. A. Zisserman., "Very deep convolutional networks for large-scale image recognition," 2014.

[31] C. C. L. T.-Y. S. A. H.-J. M. ,. M. Yeh R.A., "Semantic image inpainting with deep generative models," 2017.

[32] J. L. Z. Y. J. S. X. L. X. H. T. Yu, " Generative image inpainting with contextual attention," 2018.

[33] L. J. D. T. Shelhamer E, "Fully Convolutional Networks for Semantic Segmentation," 2014.

[34] F. P. B. T. Ronneberger O, "U-Net: Convolutional Networks for Biomedical Image Segmentation," 2015.

[35] L. E. A. B. M. Gatys, "A neural algorithm of artistic style.," 2015.

[36] K. Z. A. Simonyan, "Very deep convolutional networks for large-scale image recognition," 2014.

[37] O. D. J. S. H. K. J. S. S. M. S. H. Z. A. K. A. B. M. B. A. F.-F. L. Russakovsky, " ImageNet Large Scale Visual Recognition Challenge.," 2015.

[38] B. L. A. K. A. O. A. T. A. Zhou, "Places: A 10 million image database for scene recognition.," *IEEE Transactions on Pattern Analysis and Machine Intelligence ,* 217.

[39] Z. L. P. L. X. W. X. Tang, "Large-scale CelebFaces Attributes (CelebA) Dataset," [Online]. Available: https://mmlab.ie.cuhk.edu.hk/projects/CelebA.html.

[40] "PLacs2 website," [Online]. Available: http://places2.csail.mit.edu/.

[41] F. A. R. K. J. S. T.-C. W. A. T. B. C. Guilin Liu, "Image Inpainting for Irregular Holes Using Partial Convolutions," Nvidia, 2018. [Online]. Available: https://nv-adlr.github.io/publication/partialconv-inpainting.

[42] "Peak Signal-to-Noise Ratio as an Image Quality Metric," 16 12 2020. [Online]. Available: https://www.ni.com/en-us/innovations/white-papers/11/peak-signal-to-noise-ratio-as-an-image-quality-metric.html.

[43] Z. Wang, A. Bovik, H. Sheikh and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing,* 2004.

[44] L. Zhang, L. Zhang, X. Mou and D. Zhang, "A comprehensive evaluation of full reference image quality assessment algorithms," *IEEE International Conference on Image Processing.,* 2012.

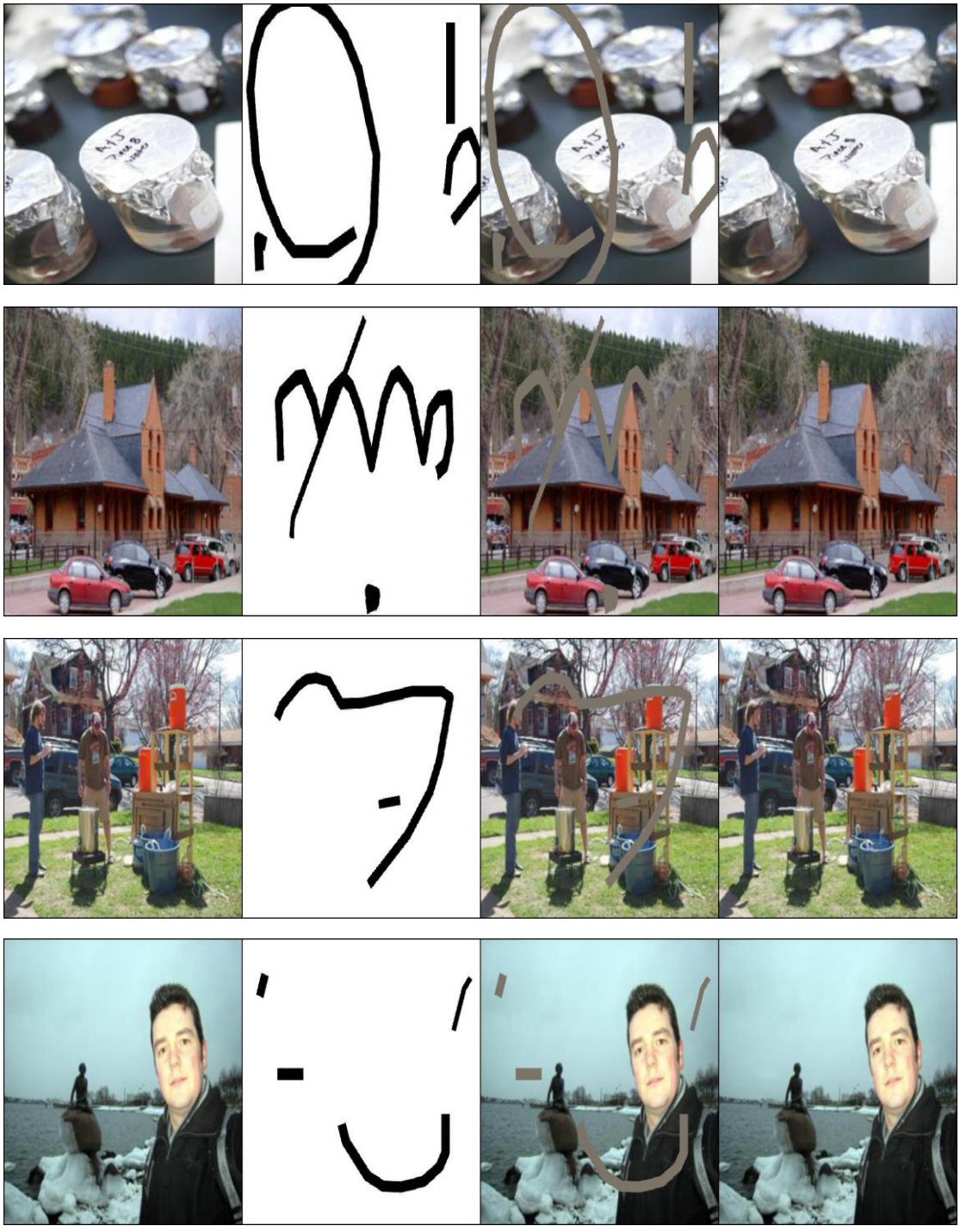
[45] E. N. T. J. F. Z. Q. M. E. Kamyar Nazeri, "EdgeConnect: Generative Image Inpainting with Adversarial Edge Learning," 2019.

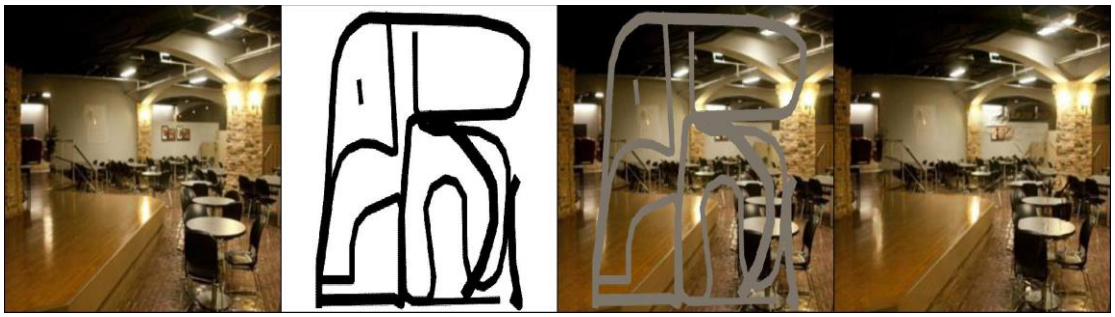
[46] Z. G. X. J. S. W. Z. C. W. L. Z. Z. S. L. Tao Yu, "Region Normalization for Image Inpainting".

[47] Z. L. J. Y. S. L. S. H. Jiahui Yu, "Generative Image Inpainting with Contextual Attention," 2018.

Παράρτημα

1.Περισσότερα καλά αποτελέσματα





2.Περισσότερα κακά αποτελέσματα