



ΠΑΝΕΠΙΣΤΗΜΙΟ
ΙΩΑΝΝΙΝΩΝ

ΣΧΟΛΗ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ ΤΟΥ
ΠΑΝΕΠΙΣΤΗΜΙΟΥ ΙΩΑΝΝΙΝΩΝ
ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

ΑΝΑΒΑΘΜΙΣΗ ΕΙΚΟΝΑΣ ΜΕ ΧΡΗΣΗ ΜΕΘΟΔΩΝ
ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ

ΔΙΟΝΥΣΗΣ ΛΑΜΠΡΟΣ ΚΟΥΚΟΒΙΝΗΣ
ΘΕΟΔΟΣΗΣ ΒΕΝΑΡΔΟΣ

Επιβλέπων: Νικόλαος Γιαννακέας

Αναπληρωτής Καθηγητής

Άρτα, Μάιος 2025

IMAGE UPSCALING USING ARTIFICIAL INTELLIGENCE

Εγκρίθηκε από τριμελή εξεταστική επιτροπή

Άρτα, 23/05/2025

ΕΠΙΤΡΟΠΗ ΑΞΙΟΛΟΓΗΣΗΣ

1. Επιβλέπων καθηγητής

Νικόλαος Γιαννακέας,

Αναπληρωτής Καθηγητής

2. Μέλος επιτροπής

Αλέξανδρος Τζάλλας,

Αναπληρωτής Καθηγητής

3. Μέλος επιτροπής

Δημήτριος Δημόπουλος,

Εντεταλμένος Καθηγητής

© Κουκοβίνης, Διονύσης Λάμπρος και Βενάρδος Θεοδόσης, 2025.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

ΔΗΛΩΣΗ ΜΗ ΛΟΓΟΚΛΟΠΗΣ

Δηλώνω υπεύθυνα και γνωρίζοντας τις κυρώσεις του Ν. 2121/1993 περί Πνευματικής Ιδιοκτησίας, ότι η παρούσα πτυχιακή εργασία είναι εξ ολοκλήρου αποτέλεσμα δικής μου ερευνητικής εργασίας, δεν αποτελεί προϊόν αντιγραφής ούτε προέρχεται από ανάθεση σε τρίτους. Όλες οι πηγές που χρησιμοποιήθηκαν (κάθε είδους, μορφής και προέλευσης) για τη συγγραφή της περιλαμβάνονται στη βιβλιογραφία.

Διονύσης Λάμπρος Κουκοβίνης

Θεοδόσης Βενάρδος

Υπογραφή:

Υπογραφή:

ΕΥΧΑΡΙΣΤΙΕΣ

Η παρούσα πτυχιακή εργασία με θέμα «ΑΝΑΒΑΘΜΙΣΗ ΕΙΚΟΝΑΣ ΜΕ ΧΡΗΣΗ ΜΕΘΟΔΩΝ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ» πραγματοποιήθηκε στο πλαίσιο της πτυχιακής εργασίας του τμήματος Μηχανικών Πληροφορικής του Πανεπιστημίου Ιωαννίνων. Θα θέλαμε να εκφράσουμε τις θερμές ευχαριστίες μας στον επιβλέποντα καθηγητή κύριο Νικόλαο Γιαννακέα, ο οποίος μας εμπιστεύτηκε την παρούσα πτυχιακή εργασία. Τέλος, θα θέλαμε να ευχαριστήσουμε τους φίλους μας και τους συμφοιτητές μας για όλα αυτά τα χρόνια που περάσαμε μαζί αναφέροντας ονομαστικά τον καθένα από αυτούς.

Είναι οι Άγγελος, Αλέξανδρος, Αντώνης, Βασίλης, Βασίλης, Δανάη, Δήμητρα, Διονύσης, Κωσταντίνος, Παντελής, Σπύρος.

Θεοδόσιος Βενάρδος 2025

Διονύσιος – Λάμπρος Κουκοβίνης 2025

ΠΕΡΙΛΗΨΗ

Στην εργασία μας ασχοληθήκαμε με την εξέλιξη των φωτογραφιών και των εικόνων και με τη μεγάλη πρόοδο που έχει σημειωθεί στην επεξεργασία τους, εστιάζοντας κυρίως στην αναλυτική κατανόηση του τρόπου με τον οποίο η τεχνητή νοημοσύνη (AI) έχει επηρεάσει αυτή τη διαδικασία. Στη διάρκεια της εργασίας μας, αναφερόμαστε στην ιστορική εξέλιξη της φωτογραφίας, τη μετάβαση από τις παραδοσιακές μεθόδους επεξεργασίας, και τις σημαντικές εξελίξεις που έχουν λάβει χώρα με τη χρήση υπολογιστών και τεχνολογιών AI όπως τα Convolutional Neural Networks (CNN) και τα Generative Adversarial Networks (GAN). Η φωτογραφία, ως μέσο καταγραφής και αναπαράστασης της πραγματικότητας, έχει περάσει από σημαντικά επίπεδα ανάπτυξης, που ξεκινούν από τις πρώτες χημικές διαδικασίες του 19ου αιώνα, όπως το daguerreotype, και φτάνουν στις ψηφιακές φωτογραφίες που βλέπουμε σήμερα.

Η εξέλιξη της φωτογραφίας αποτέλεσε σταδιακή μετάβαση από την αναλογική στην ψηφιακή τεχνολογία, και αυτό είχε αντίκτυπο στην επεξεργασία των εικόνων. Οι αρχικές τεχνικές επεξεργασίας, όπως το φιλμ, αντικαταστάθηκαν από ψηφιακά εργαλεία που επιτρέπουν στους χρήστες να τροποποιούν την εικόνα με ακριβή και άμεσο τρόπο. Με τη χρήση υπολογιστών και ψηφιακών τεχνικών επεξεργασίας, η δυνατότητα τροποποίησης εικόνας έγινε πιο προσβάσιμη και ευέλικτη, δίνοντας τη δυνατότητα στους φωτογράφους, αλλά και σε άλλους χρήστες, να δημιουργούν νέες μορφές και εφέ. Στην εργασία αυτή, αναλύουμε και τη συμβολή της τεχνητής νοημοσύνης στην επεξεργασία εικόνας, κάτι που έχει επαναστατήσει την προσέγγιση της επεξεργασίας και της αναπαραγωγής φωτογραφιών. Ειδικότερα, η χρήση νευρωνικών δικτύων έχει βελτιώσει δραματικά τις δυνατότητες επεξεργασίας, όπως η βελτίωση της ποιότητας εικόνας, η δημιουργία νέων εφέ και η ακριβής αναγνώριση και αναπαραγωγή αντικειμένων.

Τα CNN χρησιμοποιούνται εκτενώς στην αναγνώριση εικόνας και στην ανάλυση τους. Η βασική λειτουργία τους είναι να αναγνωρίζουν μοτίβα και χαρακτηριστικά σε εικόνες με τέτοιο τρόπο ώστε να μπορούν να αναγνωρίσουν αντικείμενα, πρόσωπα και άλλες σημαντικές λεπτομέρειες στην εικόνα. Αναφερόμαστε και στα GAN τα οποία είναι μια καινοτόμος τεχνολογία που επιτρέπει τη δημιουργία νέων εικόνων από υπάρχουσες εικόνες μέσω μιας διαδικασίας generative learning. Με τη χρήση αυτής της τεχνολογίας, είναι δυνατόν να δημιουργηθούν εξαιρετικά ρεαλιστικές εικόνες, που μπορούν να φαίνονται εντελώς αυθεντικές. Στην παρούσα εργασία, ασχολούμαστε επίσης με ένα πείραμα που πραγματοποιήθηκε για να εξετάσουμε την απόδοση της τεχνολογίας ESRGAN (Enhanced

Super-Resolution Generative Adversarial Network), μια μέθοδος που χρησιμοποιείται για τη βελτίωση της ανάλυσης μιας εικόνας. Το πείραμα συγκρίνει δύο εικόνες προτού και μετά την επεξεργασία με τη συγκεκριμένη τεχνολογία, προκειμένου να αναλυθεί η ποιότητα και η ακριβής αναπαραγωγή των λεπτομερειών.

Τα αποτελέσματα του πειράματος δείχνουν ότι η τεχνολογία ESRGAN μπορεί να βελτιώσει σημαντικά την ανάλυση των εικόνων, αποκαθιστώντας μικρές λεπτομέρειες και προσφέροντας καθαρότερη εικόνα με μεγαλύτερη πιστότητα στα αρχικά χαρακτηριστικά της. Συνολικά, η εργασία αυτή ασχολείται με τις σημαντικές τεχνολογικές εξελίξεις που αφορούν τις εικόνες και τη φωτογραφία, αναλύοντας την ιστορική εξέλιξη, την επανάσταση της ψηφιακής τεχνολογίας και τη συμβολή της τεχνητής νοημοσύνης στην επεξεργασία εικόνων. Η εργασία εξετάζει επίσης τη χρήση προηγμένων τεχνολογιών όπως τα CNN και τα GAN στην επεξεργασία εικόνων, δίνοντας ιδιαίτερη έμφαση στη χρήση της ESRGAN για την αποκατάσταση της ανάλυσης των εικόνων.

Λέξεις Κλειδιά: Εικόνα, Απεικόνιση, Επεξεργασία φωτογραφιών, Μέθοδοι επεξεργασίας εικόνων, Τεχνητή Νοημοσύνη, Συνελκτικά Νευρωνικά Δίκτυα, Γενετικά Ανταγωνιστικά Δίκτυα, ESRGAN Μελέτη

ABSTRACT

In our project, we focused on the evolution of photos and images and the significant progress made in their processing, primarily concentrating on the detailed understanding of how artificial intelligence (AI) has influenced this process. Throughout our work, we discuss the historical evolution of photography, the transition from traditional processing methods, and the significant advancements that have occurred through the use of computers and AI technologies such as Convolutional Neural Networks (CNN) and Generative Adversarial Networks (GAN). Photography, as a medium of recording and representing reality, has gone through significant stages of development, starting with the early chemical processes of the 19th century, such as the daguerreotype, and progressing to the digital photographs we see today. The evolution of photography marked a gradual transition from analog to digital technology, which had a direct impact on image processing.

Early processing techniques, such as film, were replaced by digital tools that allow users to modify images in precise and immediate ways. With the use of computers and digital processing techniques, the ability to modify images became more accessible and flexible, enabling photographers and other users to create new forms and effects. In this work, we also analyze the contribution of artificial intelligence to image processing, a field that has revolutionized the approach to editing and reproducing photographs. Specifically, the use of neural networks has dramatically improved processing capabilities, such as enhancing image quality, creating new effects, and accurately recognizing and reproducing objects.

CNNs are widely used in image recognition and analysis. Their primary function is to recognize patterns and features in images in such a way that they can identify objects, faces, and other important details within the image. We also refer to GANs, which are innovative technology that allows for the creation of new images from existing ones through a process known as generative learning. Using this technology, it is possible to generate highly realistic images that can appear completely authentic. In this work, we also conducted an experiment to examine the performance of the ESRGAN (Enhanced Super-Resolution Generative Adversarial Network) technology, a method used to enhance the resolution of an image. The experiment compares two images before and after processing with this technology to analyze the quality and accuracy of the detailed reproduction. The results of the experiment show that ESRGAN technology can significantly improve the resolution of images, restoring fine details and providing a clearer image with greater fidelity to the original characteristics.

Overall, this work addresses the significant technological developments related to images and photography, analyzing the historical evolution, the revolution of digital technology, and the contribution of artificial intelligence to image processing. The work also examines the use of

advanced technologies like CNNs and GANs in image processing, with a particular focus on the application of ESRGAN for enhancing image resolution.

Keywords: Image, Depiction, Picture editing, Image processing methods, Artificial Intelligence, Convolutional Neural Networks, Generative Adversarial Networks, ESRGAN Research

ΠΕΡΙΕΧΟΜΕΝΑ

ΔΗΛΩΣΗ ΜΗ ΛΟΓΟΚΛΟΠΗΣ	5
ΕΥΧΑΡΙΣΤΙΕΣ	6
ΠΕΡΙΛΗΨΗ	7
ABSTRACT	9
ΠΕΡΙΕΧΟΜΕΝΑ.....	11
ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ	14
1. ΕΙΣΑΓΩΓΗ	16
1.1 Εικόνα	16
1.2 Ανακάλυψη της τεχνητής νοημοσύνης	17
1.3 Μηχανική μάθηση	18
2. ΙΣΤΟΡΙΚΗ ΑΝΑΔΡΟΜΗ	20
2.1 Η πρώτη εικόνα	20
2.2 Απεικόνιση	21
2.3 Η ανακάλυψη της Φωτογραφίας	23
2.4 Η εφεύρεση του Φιλμ	26
3. ΨΗΦΙΑΚΗ ΕΠΟΧΗ	30
3.1 Ψηφιοποίηση	30
3.1.1 Ψηφιοποίηση εικόνας	32
3.1.2 Εξέλιξη και εκβιομηχάνιση	33
3.1.3 Τα βασικά βήματα ψηφιοποίησης εικόνας	34
3.2 Ψηφιακή εικόνα	35
3.2.1 Η Δομή μίας ψηφιακής εικόνας	36
3.2.2 Μορφές ψηφιακών εικόνων	37
3.3 Ψηφιακές κάμερες	38
3.3.1 Οι πρώτες εμπορικές κυκλοφορίες και η παροδική εξέλιξή τους	40
3.3.2 Ραγδαία εμπορική ανάπτυξη	41

3.3.3 Κινητές συσκευές με φωτογραφικές δυνατότητες	42
3.3.4 Σύγχρονη τεχνολογία και οι τελευταίες εξελίξεις	42
3.4 Επεξεργασία εικόνας.....	43
3.4.1 Εφαρμογές της ψηφιακής επεξεργασίας εικόνας	44
3.4.2 Τεχνικές ψηφιακής επεξεργασίας εικόνας	45
3.4.3 Ανακατασκευή και αποκατάσταση εικόνας.....	47
4. ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ	50
4.1 Η αρχή της τεχνητής νοημοσύνης.....	50
4.2 Επεξεργασία εικόνας με την βοήθεια της τεχνητής νοημοσύνης	53
4.3 Κλιμάκωση εικόνας (Image Upscaling)	56
4.4 Μέθοδοι κλιμάκωσης εικόνας (Image Upscaling)	57
4.5 Ανάλυση βασικών μεθόδων Image Upscaling	59
4.5.1 Nearest Neighbor Interpolation	59
4.5.2 Bilinear Interpolation.....	61
4.5.3 Bicubic Interpolation	62
4.5.4 Lanczos Interpolation	63
5. Εισαγωγή στα Convolutional Neural Networks.....	66
5.1 CNN'S.....	66
5.2 Αριθμητική ανάλυση CNN.....	69
5.3 Ανάλυση μεθόδων των Νευρωνικών Δικτύων (Neural Networks)	74
5.3.1 SRCNN (Super-Resolution Convolutional Neural Network)	74
5.3.2 GAN (Generative Adversarial Networks).....	76
5.3.3 SRGAN (Super-Resolution Generative Adversarial Network).....	78
5.3.4 ESRGAN (Enhanced SRGAN)	81
5.3.5 Αναφορά επιπλέον μεθόδων Image Upscaling.....	84
6. Image Upscaling με την χρήση του ESRGAN	87
6.1 Ανάλυση και επίλυση προκλήσεων του ESRGAN.....	87
6.2 Βήματα εγκατάστασης και ανάλυση κώδικα ESRGAN σε Python	89

6.3 Αποτελέσματα Upscaling με ESRGAN	92
6.4 SNR (Signal to noise ratio)	95
6.5 Αποτελέσματα πειράματος.....	96
BIBΛΙΟΓΡΑΦΙΑ	98

ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

Εικόνα 1 Artificial Intelligence	18
Εικόνα 2 Machine Learning	19
Εικόνα 3 Τοιχογραφία	20
Εικόνα 4 Camera Obscura	21
Εικόνα 5 Leonardo Davinci Obscura	22
Εικόνα 6 Η πρώτη φωτογραφία	23
Εικόνα 7 Photograph Developing	25
Εικόνα 8 Evolution of photography	29
Εικόνα 9 Digitalization	31
Εικόνα 10 MRI	34
Εικόνα 11 Grayscale to RGB	35
Εικόνα 12 Image Formats	38
Εικόνα 13 Steven Sasson	39
Εικόνα 14 Fujix DS-1P	40
Εικόνα 15 Early digital compact cameras	41
Εικόνα 16 J-Phone	42
Εικόνα 17 Nikon Z6 II	43
Εικόνα 18 Adobe Products	45
Εικόνα 19 Fourier Filter	47
Εικόνα 20 Alan Turing and the Turing machine	50
Εικόνα 21 Kasparov vs Deep Blue	52
Εικόνα 22 Deep Learning	54
Εικόνα 23 CNN Layers	66
Εικόνα 24 CNN RELU	68
Εικόνα 25 Object Identification	73
Εικόνα 26 RRDB ESRGAN test.py	89
Εικόνα 27 Define paths	90
Εικόνα 28 Import Model	90
Εικόνα 29 Image processing	91
Εικόνα 30 Image pre processing	91
Εικόνα 31 Model Application and Output export	92
Εικόνα 32 LR image 1 62x90	93
Εικόνα 33 HR image 248x368	93

Εικόνα 34 LR image 125x120.....	94
Εικόνα 35 HR image 500x480.....	94
Εικόνα 36 SNR code in MATLAB.....	95
Εικόνα 37 SNR dB results.....	96
Εικόνα 38 SNR Visualization	96
Εικόνα 39 dB Stats For SNR	97

ΚΕΦΑΛΑΙΟ 1

1. ΕΙΣΑΓΩΓΗ

1.1 Εικόνα

Η εικόνα και η φωτογραφία έχουν αναδειχθεί ως θεμελιώδη εργαλεία για την ανθρωπότητα, επηρεάζοντας καθοριστικά την κατανόηση και την αλληλεπίδρασή μας με τον κόσμο γύρω μας. Έως και την ψηφιακή εποχή, η ικανότητα της εικόνας να αποτυπώνει τον εξωτερικό κόσμο, έχει συμβάλλει πραγματικά στην εξέλιξη της ανθρωπότητας. Η ανθρώπινη ανάγκη για καταγραφή και επικοινωνία παρέμεινε αμετάβλητη, αν και οι τεχνικές και τα εργαλεία που χρησιμοποιούνται έχουν εξελιχθεί δραματικά. Η ανακάλυψη της φωτογραφίας τον 19ο αιώνα αποτέλεσε επαναστατική πρόοδο στην καταγραφή και διάδοση της εικόνας. Με την εφεύρεση της φωτογραφικής μηχανής, οι άνθρωποι απέκτησαν τη δυνατότητα να αποτυπώνουν στιγμές με ακρίβεια και λεπτομέρεια που προηγουμένως ήταν αδύνατη. Η φωτογραφία ανέδειξε νέες δυνατότητες για την καλλιτεχνική έκφραση, την τεκμηρίωση ιστορικών γεγονότων και την ανάπτυξη επιστημονικών μεθόδων.

Σήμερα, η ψηφιακή φωτογραφία και η επεξεργασία εικόνας έχουν εισέλθει στην καθημερινότητά μας, προσφέροντας αμέτρητες ευκαιρίες και δυνατότητες. Η ψηφιακή τεχνολογία έχει καταστήσει την εικόνα πιο προσιτή και εύχρηστη. Η ικανότητα να αποθηκεύουμε, να επεξεργαζόμαστε και να διανέμουμε εικόνες έχει ανοίξει νέες δυνατότητες σε τομείς όπως η ιατρική απεικόνιση, η επιστημονική έρευνα, η εκπαίδευση, και η ψυχαγωγία. Στην ιατρική, για παράδειγμα, οι ψηφιακές εικόνες επιτρέπουν την ακριβή διάγνωση και παρακολούθηση της υγείας των ασθενών, προσφέροντας λεπτομερή απεικόνιση εσωτερικών οργάνων και ιστών που βοηθά στη λήψη κρίσιμων αποφάσεων. Στην εκπαίδευση, οι εικόνες και οι φωτογραφίες γίνονται αναπόσπαστο μέρος της διδασκαλίας, διευκολύνοντας την κατανόηση σύνθετων εννοιών και ενισχύοντας τη μαθησιακή εμπειρία. Στη διαφήμιση και τα μέσα ενημέρωσης, η ικανότητα δημιουργίας εντυπωσιακών και ελκυστικών εικόνων είναι καθοριστική για την προσέγγιση και την αλληλεπίδραση με το κοινό.

Η ψηφιακή επεξεργασία έχει επαναστατήσει την ικανότητα των φωτογράφων και των δημιουργών περιεχομένου να σχεδιάζουν εικόνες που τραβούν την προσοχή και προκαλούν συναισθηματικό αντίκτυπο. Συνολικά, η εικόνα και η φωτογραφία έχουν εξελιχθεί σε αναπόσπαστα εργαλεία της σύγχρονης κοινωνίας, επιτρέποντας την ακριβή καταγραφή, ανάλυση και επικοινωνία πληροφοριών. Καθώς η τεχνολογία συνεχίζει να προχωρά, οι δυνατότητες που προσφέρουν οι εικόνες αναμένεται να γίνουν ακόμη πιο ισχυρές και καθοριστικές για την κατανόηση και την ανάπτυξη της ανθρωπότητας. [\[1\]](#) [\[2\]](#) [\[3\]](#) [\[4\]](#)

1.2 Ανακάλυψη της τεχνητής νοημοσύνης

Η τεχνητή νοημοσύνη (AI) έχει επιφέρει επαναστατικές αλλαγές σε πολλούς τομείς της τεχνολογίας, με την επεξεργασία εικόνας να αποτελεί έναν από τους πιο δυναμικούς τομείς εφαρμογής της. Καθώς οι υπολογιστές γίνονται ολοένα και πιο ικανοί στην ανάλυση και κατανόηση οπτικών δεδομένων, η τεχνητή νοημοσύνη προσφέρει νέες δυνατότητες που ενισχύουν την ποιότητα, την ακρίβεια και την αποδοτικότητα της επεξεργασίας εικόνας. Η επεξεργασία εικόνας με τη βοήθεια της τεχνητής νοημοσύνης αξιοποιεί αλγορίθμους μηχανικής μάθησης, νευρωνικά δίκτυα και άλλες προηγμένες τεχνολογίες για να βελτιώσει και να αναλύσει εικόνες με τρόπο που προηγουμένως ήταν αδύνατος.

Οι βασικές μέθοδοι περιλαμβάνουν την αναγνώριση αντικειμένων, την ανάλυση εικόνας, τη βελτίωση ποιότητας εικόνας, και την αυτόματη κατηγοριοποίηση εικόνων. Η αναγνώριση αντικειμένων, για παράδειγμα, επιτρέπει στα συστήματα AI να εντοπίζουν και να κατηγοριοποιούν αντικείμενα σε εικόνες με εκπληκτική ακρίβεια. Αυτό έχει σημαντικές εφαρμογές σε τομείς όπως η αυτόνομη οδήγηση, όπου η ακριβής αναγνώριση οδόσημων, πεζών και άλλων οχημάτων είναι κρίσιμη για την ασφάλεια των συστημάτων. Η βελτίωση της ποιότητας εικόνας είναι ένα άλλο σημαντικό πεδίο εφαρμογής της AI. Μέθοδοι όπως το image upscaling (κλιμάκωση εικόνας) επιτρέπουν την αύξηση της ανάλυσης εικόνας χωρίς να χάνεται η ποιότητα, χάρη σε τεχνικές όπως τα νευρωνικά δίκτυα και οι αλγόριθμοι deep learning. Αυτές οι τεχνικές χρησιμοποιούνται ευρέως στην ιατρική απεικόνιση για την ανάλυση ιατρικών εικόνων, όπως ακτινογραφίες και MRI, παρέχοντας πιο λεπτομερείς και ακριβείς πληροφορίες για τη διάγνωση και τη θεραπεία ασθενειών. Η τεχνητή νοημοσύνη επιτρέπει επίσης την αυτόματη κατηγοριοποίηση εικόνων, βελτιώνοντας τη διαχείριση και αναζήτηση δεδομένων σε μεγάλες βάσεις δεδομένων. Αυτή η ικανότητα έχει σημαντικές εφαρμογές στη διαχείριση ψηφιακών αρχείων, στη βελτίωση των εργαλείων αναζήτησης εικόνας και στην οργάνωση φωτογραφικών συλλογών.

Η συνδυασμένη χρήση των τεχνολογιών AI με τις παραδοσιακές τεχνικές επεξεργασίας εικόνας προσφέρει νέες δυνατότητες για τη δημιουργία, την ανάλυση και την εφαρμογή εικόνας σε πολλές περιοχές. Από την ενίσχυση της καλλιτεχνικής δημιουργίας έως την ενίσχυση της επιστημονικής έρευνας και της επιχειρηματικής ανάλυσης, η επεξεργασία εικόνας με την τεχνητή νοημοσύνη επιταχύνει την καινοτομία και επαναστατεί την κατανόηση του οπτικού κόσμου. [\[9\]\[10\]](#)



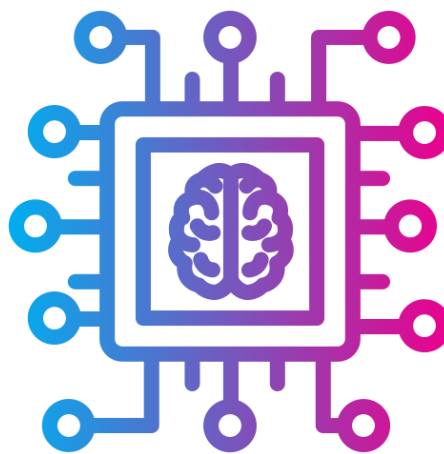
Εικόνα 1 Artificial Intelligence

1.3 Μηχανική μάθηση

Το deep learning, ως υποκατηγορία της μηχανικής μάθησης, αξιοποιεί σύνθετα νευρωνικά δίκτυα για να αναγνωρίσει και να αναλύσει πρότυπα σε δεδομένα με μοναδική ακρίβεια και αποδοτικότητα. Η εφαρμογή αυτής της τεχνολογίας στην επεξεργασία εικόνας έχει δημιουργήσει νέες προοπτικές, επηρεάζοντας πολλούς τομείς όπως η ιατρική, η ασφάλεια, η βιομηχανία και η ψυχαγωγία. Στον πυρήνα του deep learning βρίσκονται τα νευρωνικά δίκτυα, και συγκεκριμένα τα βαθιά νευρωνικά δίκτυα (deep neural networks), τα οποία περιλαμβάνουν πολλές στρώσεις (Layers) από νευρώνες που επεξεργάζονται τα δεδομένα σε διάφορα επίπεδα, ανιχνεύοντας πολύπλοκα μοτίβα και σχέσεις που δεν είναι άμεσα προφανείς. Για παράδειγμα, σε μια εικόνα, τα πρώτα επίπεδα του δικτύου μπορεί να ανιχνεύσουν βασικά στοιχεία όπως άκρα και υφές, ενώ τα βαθύτερα επίπεδα μπορούν να αναγνωρίσουν πιο σύνθετα αντικείμενα ή σκηνές. Ενδεικτικά, η αναγνώριση αντικειμένων χρησιμοποιεί νευρωνικά δίκτυα για να εντοπίσει και να κατηγοριοποιήσει αντικείμενα σε εικόνες με μεγάλη ακρίβεια. Αυτή η τεχνολογία είναι κρίσιμη για πληθώρα εφαρμογών. Στην

ιατρική απεικόνιση, το deep learning ενισχύει τη διάγνωση ασθενειών μέσω της ανάλυσης ιατρικών εικόνων, όπως ακτινογραφίες, MRI και CT scans.

Τα νευρωνικά δίκτυα μπορούν να αναγνωρίσουν ανωμαλίες και πρότυπα που είναι δύσκολο να ανιχνευθούν από το ανθρώπινο μάτι, βελτιώνοντας την ακρίβεια των διαγνώσεων και επιτρέποντας πιο πρώιμη παρέμβαση στις ασθένειες. Στον τομέα της ψυχαγωγίας, το deep learning έχει επαναστατήσει τη δημιουργία και την επεξεργασία περιεχομένου. Από την αναγνώριση προσώπων σε φωτογραφίες και βίντεο μέχρι τη δημιουργία ρεαλιστικών εφέ και την επεξεργασία εικόνας σε πραγματικό χρόνο, οι δυνατότητες που προσφέρει είναι εντυπωσιακές. Συνολικά, το deep learning αναμορφώνει την επεξεργασία εικόνας, καθιστώντας την πιο ακριβή, αποδοτική και προσαρμόσιμη. Οι δυνατότητες που προσφέρει η τεχνολογία αυτή ανοίγουν νέους δρόμους σε διάφορους τομείς, προσφέροντας προοδευτικές λύσεις για την ανάλυση και την κατανόηση των οπτικών δεδομένων. Καθώς η τεχνολογία συνεχίζει να εξελίσσεται, αναμένονται ακόμη πιο επαναστατικές εξελίξεις στην επεξεργασία εικόνας, που θα φέρουν νέες ευκαιρίες και προκλήσεις στην επιστήμη και την τεχνολογία. [\[23\]](#)[\[24\]](#)[\[25\]](#)



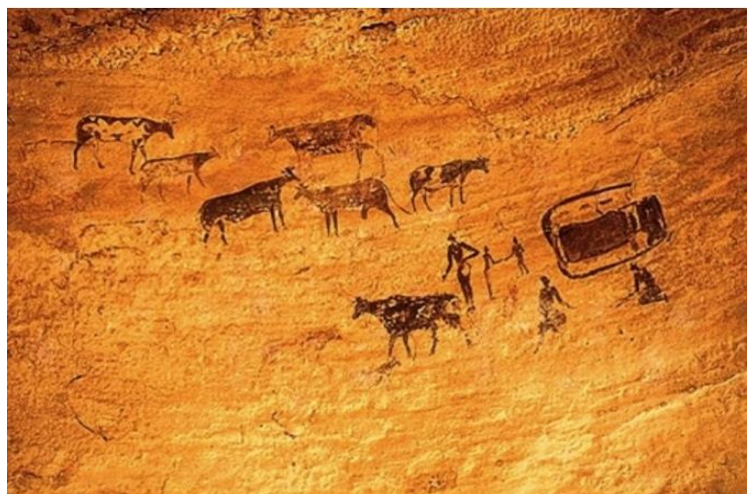
Εικόνα 2 Machine Learning

ΚΕΦΑΛΑΙΟ 2

2. ΙΣΤΟΡΙΚΗ ΑΝΑΔΡΟΜΗ

2.1 Η πρώτη εικόνα

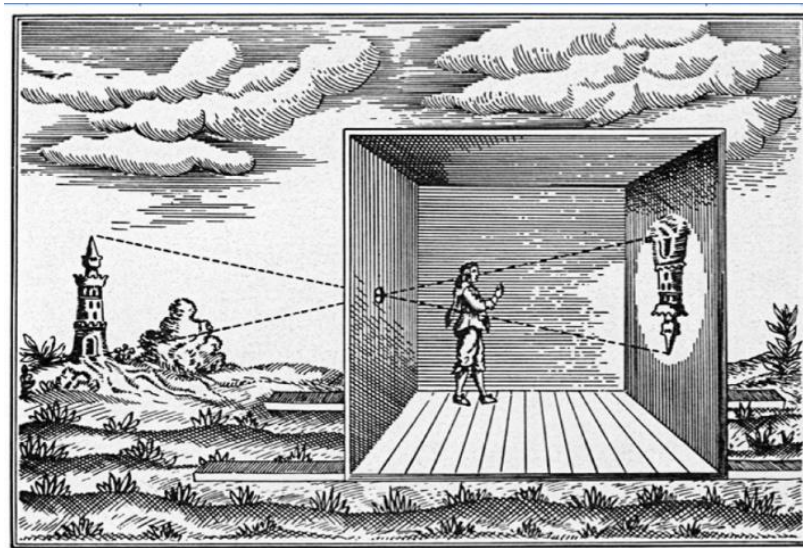
Η εικόνα, από την αρχαιότητα, μέχρι και σήμερα αποτελεί κεντρικό στοιχείο της ανθρώπινης έκφρασης, επικοινωνίας και κατανόησης. Η ιστορία της εικόνας αντικατοπτρίζει την εξέλιξη της κοινωνίας, της τέχνης, του ανθρώπινου πνεύματος και κυρίως της τεχνολογίας. Από την αρχαιότητα ο άνθρωπος χρησιμοποιούσε τις εικόνες για να καταγράψει τον κόσμο και τις εμπειρίες της καθημερινότητάς του. Αρχαίες προσπάθειες της ζωγραφικής όπως οι τοιχογραφίες των σπηλαίων και η γλυπτική αποτελούν σημαντικά παραδείγματα για την ανάγκη της αναζήτησης της εικονικής απεικόνισης. Κατά τον Μεσαίωνα, ανακαλύπτουμε παραδείγματα της εξέλιξης της εικόνας, μέσα από δημιουργήματα καλλιτεχνών, οι οποίοι ξεκίνησαν να αναπτύσσουν τεχνικές για να αποδώσουν το φως, τη σκιά και την έκφραση σε μία διάσταση. Δίνοντας την δυνατότητα αντίληψης της απεικόνισης της πραγματικότητας. Μέσα από βιβλία και τοιχογραφίες εκκλησιών, μπορούσε ακόμα και ο μέσος άνθρωπος να διακρίνει πως το έργο των καλλιτεχνών μπορεί να απεικονίσει ένα υπάρχον αντικείμενο ή μία πραγματική ύπαρξη. [\[1\]\[2\]\[3\]\[4\]](#)



Εικόνα 3 Τοιχογραφία

2.2 Απεικόνιση

Το πρώτο εργαλείο καταγραφής εικόνας ονομάζεται στα λατινικά *camera obscura*, το οποίο και μεταφράζεται ως ‘σκοτεινός θάλαμος’. Ο «σκοτεινός θάλαμος» είναι γνωστός ως έννοια από την εποχή των αρχαίων φιλόσοφων Mozi και του Αριστοτέλη και σε μετέπειτα χρόνια του Ιρακινού επιστήμονα Alhazen. Η πρώτη αναφορά στην αρχή λειτουργίας του ‘σκοτεινού θαλάμου’ έγινε από τον Κινέζο φιλόσοφο Mozi. Ο Mozi αναφέρεται σε αυτή τη διάταξη ως ‘συλλεκτική πλάκα’. Ο Έλληνας φιλόσοφος Αριστοτέλης (384-322 π.Χ.) κατανοούσε την οπτική αρχή του ‘σκοτεινού δωματίου’.

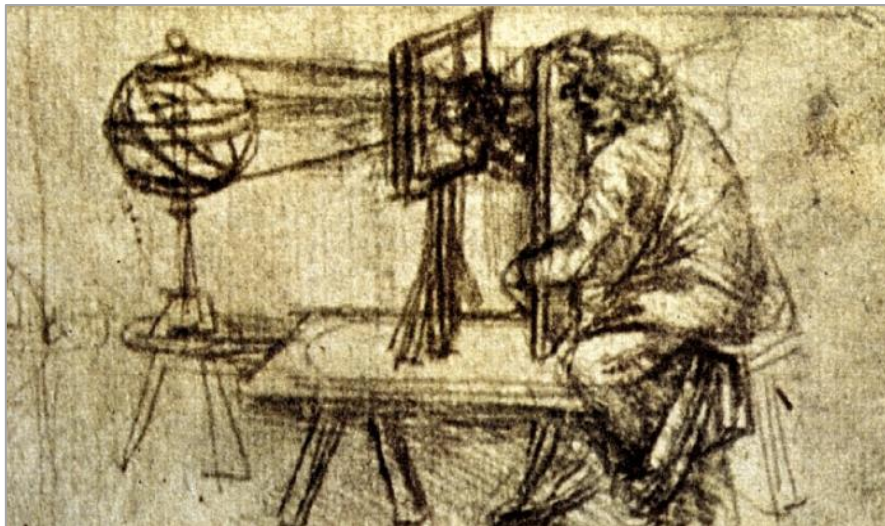


Εικόνα 4 Camera Obscura

Κατά τη διάρκεια μιας ηλιακής έκλειψης και καθώς ο Ήλιος ήταν μερικώς σκοτεινός, παρατήρησε το σχήμα ημισελήνου του Ήλιου να προβάλλεται στο έδαφος μέσα από τις τρύπες σε ένα κόσκινο και μέσα από τα κενά μεταξύ των φύλλων ενός δέντρου. Σύμφωνα με τον Ευκλείδη (περίπου 300 π.Χ.) ο ‘σκοτεινός θάλαμος’ παρουσιάζεται ως απόδειξη ότι το φως ταξιδεύει σε ευθεία γραμμή. Ο επιστήμονας Alhazen έγραψε πολλαπλές θεωρίες βασιζόμενος στον σκοτεινό θάλαμο. Ασχολήθηκε κυρίως με την οπτική αντίληψη των ανθρώπων. Επίσης πρόσφερε πολλές συμβολές στην εξέλιξη των κατόπτρων και στις διόπτρες βασιζόμενος στην μελέτη της αντανάκλασης και του αντικατοπτρισμού της εικόνας που διαμορφώνεται από τις ακτίνες του φωτός.

Έπειτα από πολλά χρόνια, το εργαλείο αυτό δημιουργήθηκε και η καταγραφή της εικόνας έγινε εφικτή. Πρόκειται για την πρώτη οπτική συσκευή που προβάλλει το είδωλο του περιβάλλοντος χώρου σε μία οθόνη. Το οπτικό αυτό όργανο αποτελείται από ένα κουτί το οποίο είναι φτιαγμένο έτσι ώστε αποτρέπει την είσοδο από το φως, με μία μόνο χαρακτηριστική οπή στην μία πλευρά του κουτιού. Από το εξωτερικό περιβάλλον, το φως των αντικειμένων διαπερνά

μέσα στο φωτοστεγές κουτί μέσω της οπής, αντανakλάει σε μία εσωτερική επιφάνεια και αντικατοπτρίζει την εξωτερική εικόνα. Τα αντικείμενα βέβαια φαινόταναν αντεστραμμένα αλλά το σημαντικό ήταν πως η συγκεκριμένη αναπαραγωγή της εικόνας είχε διατηρημένα τα χρώματα του εξωτερικού περιβάλλοντος. Έπειτα την εικόνα αυτή μπορούσαν να την προβάλουν σε χαρτί ζωγραφίζοντάς την, με σκοπό να δημιουργήσουν μία εξαιρετικά ακριβής αναπαράσταση. Σε συνδυασμό με πρόσθετα κάτοπτρα υπό γωνία και χαρτί αντιγραφής στο πίσω μέρος του κουτιού φτιαγμένο σε γυάλινη επιφάνεια, κατάφεραν να αντιστρέψουν την εικόνα σε ορθή μορφή, κάνοντας την εξωτερική εικόνα του περιβάλλοντος να προβάλλεται κατευθείαν χωρίς να χρειαστεί να την αντιστρέψουν. Για παράδειγμα το 1490 ο παγκοσμίως γνωστός εφευρέτης και καλλιτέχνης Leonardo Da Vinci κατάφερε να καταγράψει με την χρήση του σκοτεινού θαλάμου ζωγραφισμένες ανθρώπινες μορφές.



Εικόνα 5 Leonardo Davinci Obscura

Το 1700 οι camera obscura έχουν γίνει πλέον συνηθισμένα αντικείμενα που χρησιμοποιούσαν για σχεδιαγράμματα γνωστοί καλλιτέχνες όπως ο Rembrandt, ο Vermeer και ο Caravaggio. Μία πολύ σημαντική παρατήρηση που δημιουργήθηκε με την διαδικασία καταγραφής εικόνας μέσω του σκοτεινού θαλάμου είναι πως, όσο πιο μικρή είναι η οπή τόσο αυξάνεται και η ευκρίνεια του αντικειμένου που προσπαθούσαν να απεικονίσουν, αλλά παράλληλα μετατρέπονταν και πιο σκοτεινή. Η επιδείνωση της ευκρίνειας γίνεται λόγω της περίθλασης. Η περίθλαση είναι φαινόμενο της διάδοσης των κυμάτων.

Το φαινόμενο της περίθλασης παρατηρήθηκε για πρώτη φορά από τον Ιταλό φυσικό Francesco Grimaldi το 1665, ο οποίος παρατήρησε ότι οι ακτίνες φωτός που διαπερνούσαν μία λεπτή σχισμή δημιουργούσαν μία λεπτή φωτεινή ζώνη λίγο πλατύτερη από την σχισμή. Το φαινόμενο αυτό το ερμήνευσε ως κάμψη του φωτός λόγω των άκρων των σχισμών και το ονόμασε ως το γνωστό φαινόμενο της περίθλασης. Παράλληλα μία πολύ σημαντική ανακάλυψη για την

εξέλιξη της φωτογραφίας έγινε από τον Johann Heinrich Schulze, το 1724, οποίος παρατήρησε πως αντικείμενα που έχουν αναμιχθεί με Νιτρικό Άργυρο (Silver Nitrate), αντιδράνε με την άμεση έκθεση στο φως. Δυστυχώς όμως οι εικόνες ξεθώριαζαν σχεδόν αμέσως. Έπειτα ο James Gregory παρατήρησε εικόνα περίθλασης που προερχόταν από φως που έπεφτε σε φτερό πτηνού. Το φτερό ήταν η πρώτη διάταξη παρατήρησης εικόνας περίθλασης που χρησιμοποιήθηκε. Το 1801 ο Thomas Young πραγματοποίησε το πείραμα με τις δυο σχισμές που επιβεβαίωσε την κυματική φύση του φωτός. [\[1\]\[2\]\[3\]\[4\]](#)

2.3 Η ανακάλυψη της Φωτογραφίας

Οι συνεχείς προσπάθειες για την καταγραφή εικόνας, μας φέρνουν στον 19^ο αιώνα. Πιο συγκεκριμένα, το 1826 ο Joseph Nicéphore Niepce κατάφερε να αποθανατίσει τις πρώτες οριακά εμφανείς στο μάτι φωτογραφίες. Βέβαια όλες είχαν πολύ θολά αποτελέσματα και η διαδικασία της καταγραφής απαιτούσε μεγάλο χρονικό διάστημα. Για να καταγραφεί μία φωτογραφία θα έπρεπε η έκθεση της εικόνας στην κάμερα να διαρκεί πάνω από 8 ώρες και σε πολλές περιπτώσεις θα έπρεπε να περιμένουν και μέρες ολόκληρες. Την τεχνική αυτή την ονόμασε Heliography (Ηλιογραφία). Η πρώτη φωτογραφία που καταγράφηκε από τον Niepce, είναι ένα τοπίο έξω από το σπίτι του σε ασπρόμαυρη μορφή και καταφέρνει να αναδείξει τις σκιές που σχηματίζονται στους τοίχους των κτηρίων από τις ακτίνες του ήλιου. Η έκθεση της φωτογραφίας στην κάμερα, για να εμφανιστεί διήρκεσε μερικές ημέρες.



Εικόνα 6 Η πρώτη φωτογραφία

Μετά τον θάνατό του, συνέχισε το έργο του ο συνétaίρός του, ο Γάλλος Louis Daguerre. Ο Daguerre κατάφερε να δημιουργήσει το πρώτο πρακτικό διάγραμμα με την διαδικασία καταγραφής εικόνας. Δημιούργησε την δική του χημική κάμερα, το 1833, την οποία και

ονόμασε Daguerreotype. Μετά από μερικά χρόνια, το 1839, η γαλλική κυβέρνηση αγόρασε την Daguerreotype κάμερα και την έκανε προσβάσιμη στον υπόλοιπο κόσμο, κάνοντάς την, την πρώτη κάμερα στον κόσμο που πωλήθηκε στο κοινό. Το όνομα του Daguerre υπάρχει ακόμα και σήμερα γραμμένο στον πύργο του Eiffel, ως ιστορικό πρόσωπο. Σε αυτήν την χρονική περίοδο καταγράφηκε και η πρώτη φωτογραφία με άνθρωπο. Χρειάστηκε έκθεση στον ήλιο περίπου 7 λεπτά και η φιγούρα του ανθρώπου στην φωτογραφία φαίνεται κανονικά, διότι εκείνη την στιγμή ήταν ακίνητος καθώς γυάλιζε τα παπούτσια του σε έναν πλανόδιο τσαγκάρη.

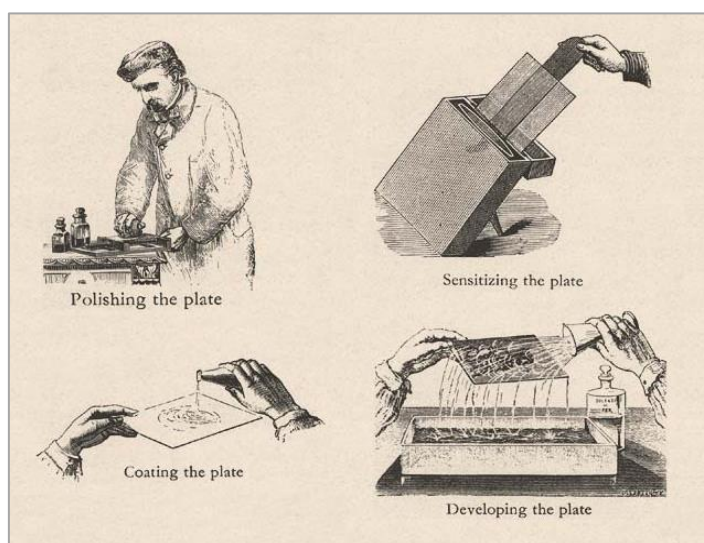
Η πρώτη αυτοπρόσωπη φωτογραφία καταγράφηκε από τον Robert Cornelius και τότε ξεκίνησε η εποχή που δημιουργήθηκε το επάγγελμα του φωτογράφου. Τότε ανοίχτηκαν και τα πρώτα photography studios, για όλον το κόσμο που είχε την οικονομική δυνατότητα να πάει και να φωτογραφηθεί. Το 1840 ο Άγγλος Henry Fox Talbot δημιούργησε την δική του κάμερα, την Calotype κάμερα. Η Calotype camera χρησιμοποιούσε την αρνητική απόχρωση του φωτός (negative).

Ο Talbot εμπνευσμένος από τον Johann Heinrich Schulze, πειραματίστηκε με την χημική αντίδραση που αποκτάει το νιτρικό άργυρο στο φως. Ξεκίνησε χρησιμοποιώντας φύλλα φυτών. Τα τοποθετούσε σε χαρτί που ήταν επικαλυμμένο σε διάλυμα νιτρικού άργυρου και αλατιού και το άφηνε στον ήλιο. Ένα αντικείμενο που είναι επικαλυμμένο με το διάλυμα αυτό, με την έκθεση στον ήλιο, σκουραίνει. Τοποθετώντας λοιπόν το φύλλο πάνω στο χαρτί, το φως δεν κατάφερνε να περάσει μέσα από το φύλλο, έχοντας ως αποτέλεσμα, να αποτυπώνεται στο χαρτί το σχήμα του φύλλου, κάνοντας τις καταγραφές αυτές ως τις πρώτες αρνητικές φωτογραφίες. Αυτό το φαινόμενο έδωσε την έμπνευση στον Talbot, να χρησιμοποιήσει την αντίστροφη διαδικασία. Δηλαδή, χρησιμοποίησε ως μοντέλο καταγραφής την αρνητική φωτογραφία σε μία αυτοσχέδια camera obscura, τοποθετώντας μέσα στο κουτί ένα ακόμα κομμάτι χαρτιού επικαλυμμένο στο διάλυμα του αργύρου. Έτσι λοιπόν το φως μετέτρεπε τις σκοτεινές περιοχές των φωτογραφιών σε φωτεινές και αντιστρόφως τις φωτεινές σε σκοτεινές. Αυτή είναι η ιδιαιτερότητα που είχε η Calotype camera, δίνοντας στην ουσία την δυνατότητα της αντιγραφής της φωτογραφίας σε όσα αντίτυπα επιθυμούσε. Δυστυχώς όμως ο Talbot θέλησε να δώσει στον κόσμο την κάμερα του (Calotype) επί πληρωμή, σε αντίθεση με τον Daguerre που επέτρεψε στην γαλλική κυβέρνηση να δώσει την κάμερά του δωρεάν, πράγμα το οποίο δεν άρεσε στους ανθρώπους που δεν είχαν την οικονομική αυτή δυνατότητα. Η διαδικασία καταγραφής όμως αυτή, που μετατρέπει τις αρνητικές απεικονίσεις σε θετικές, συνεχίστηκε να χρησιμοποιείται από όλο τον κόσμο μέχρι την ανακάλυψη της ψηφιακής φωτογραφίας.

Η λέξη 'φωτογραφία' χρησιμοποιήθηκε και έμεινε στην ιστορία, από τον Sir John Herschel, ο οποίος δημοσίευσε πολλά επιστημονικά άρθρα πάνω στην εξέλιξη της φωτογραφίας, εμπνευσμένη από τις ελληνικές λέξεις «Φως» και «Ζωγραφική». Το 1850, η φορητικότητα

των καμερών έγινε εφικτή καθώς το μέγεθος των συσκευών είχε μειωθεί σημαντικά. Αυτό έδωσε την δυνατότητα στους φωτογράφους να δουλεύουν εν κίνηση. Επίσης με την πάροδο του χρόνου, η απαιτούμενη διάρκεια έκθεσης του αντικειμένου ή του προσώπου στην κάμερα, είχε μειωθεί στα περίπου 3 λεπτά. Ακόμα και σε αυτά τα λίγα λεπτά όμως, ο άνθρωπος που φωτογραφιζόταν θα έπρεπε να ήταν απολύτως ακίνητος, για αυτό και όλες οι φωτογραφίες της εποχής απεικονίζουν ανθρώπους που δεν χαμογελούσαν. Σε πολλές περιπτώσεις μάλιστα για να είναι απολύτως ακίνητοι οι άνθρωποι κατά την διάρκεια της φωτογράφισης, τους δίνανε ηρεμιστικά και στηρίγματα για την μέση και τον σβέρκο.

Το 1871 ο φυσικός και φωτογράφος Richard Leach Maddox, εφηύρε τις στεγνές πλάκες ζελατίνης (Gelatin Dry Plate), οι οποίες ήταν η αρχή του γνωστού φιλμ. Η διαδικασία που εφηύρε ο Maddox, είναι βασισμένη στην διαδικασία του Talbot με τις αρνητικές φωτογραφίες, με την διαφορά όμως ότι για να πραγματοποιηθεί μία καταγραφή, δεν χρειάζεται η άμεση έκθεση στο φως του ήλιου. Οι πλάκες αυτές είναι εμποτισμένες σε διάλυμα από νιτρικό άργυρο, βρώμιο και ζελατίνη, οι οποίες αποθηκευόντουσαν μέσα στην κάμερα. Κάθε μία πλάκα περιείχε χαρτί με την αρνητική εκτύπωση. Κάθε φορά που βγάζανε φωτογραφία, η εικόνα αποθηκευόταν στην πλάκα χωρίς τον κίνδυνο αλλοίωσης λόγω της χημικής αντίδρασης του διαλύματος. Αυτό σημαίνει πως μπορούσαν να βγάλουν πολλαπλές αρνητικές φωτογραφίες και να τις επεξεργάζονται αργότερα. Μετά την φωτογράφιση βγάζανε τις αρνητικές φωτογραφίες από τις πλάκες, με προσοχή σε δωμάτιο χωρίς φωτισμό και τις τοποθετούσαν σε μία μικρή δεξαμενή, όπου στην συνέχεια τις ξεπλένανε με άλλες χημικές ουσίες όπως νιτρικό άργυρο. Έπειτα τις αφήνανε στην έκθεση τεχνητού φωτός να στεγνώσουν. Έτσι λοιπόν ανακαλύφθηκε η διαδικασία ανάπτυξης των φωτογραφιών (Developing). Η κάμερα πλέον ήταν σημαντικά μικρότερη σε μέγεθος και δεν χρειαζόταν η χρήση τρίποδου. [\[1\]\[2\]\[3\]\[4\]](#)



Εικόνα 7 Photograph Developing

2.4 Η εφεύρεση του Φιλμ

Η Kodak, μια από τις πιο σημαντικές εταιρείες στην ιστορία της φωτογραφίας, ιδρύθηκε από τον George Eastman και τον Henry A. Strong το 1888 στο Ρότσεστερ της Νέας Υόρκης. Η ιστορία της Kodak είναι αναμφισβήτητα συνδεδεμένη με την ιστορία της φωτογραφίας, καθώς η εταιρεία έπαιξε καθοριστικό ρόλο στην ανάπτυξη της φωτογραφικής τεχνολογίας και στην καθιέρωση της φωτογραφίας ως ένα μαζικά προσιτό και δημοφιλές μέσο. Όπως προαναφέραμε, Την εποχή εκείνη, η φωτογραφία ήταν μια περίπλοκη και δαπανηρή διαδικασία που απαιτούσε βαριά και δυσκίνητα εξοπλισμό και χημικά. Το 1879, ο Eastman εφεύρε μια μηχανή για την εφαρμογή ενός λεπτού στρώματος ζελατίνης στο φωτογραφικό χαρτί, διευκολύνοντας έτσι την παραγωγή φωτογραφικών πλακών. Αυτή η εφεύρεση αποτέλεσε το πρώτο βήμα προς την εμπορική του επιτυχία.

Η μεγαλύτερη συμβολή της Kodak στη φωτογραφία ήταν η εφεύρεση και η εμπορική διάθεση του ρολού φιλμ. Το 1889, ο Eastman ανέπτυξε το πρώτο ευέλικτο ρολό φιλμ που μπορούσε να χρησιμοποιηθεί με τις φωτογραφικές μηχανές της Kodak. Το ρολό φιλμ έκανε τη φωτογραφία πιο πρακτική και προσιτή, επιτρέποντας στους φωτογράφους να τραβούν πολλές φωτογραφίες χωρίς να χρειάζεται να αλλάζουν πλάκες κάθε φορά. Η κάμερα αυτή ήταν μικρή, εύκολη στη χρήση και περιείχε ένα ρολό φιλμ που μπορούσε να τραβήξει 100 φωτογραφίες. Το σλόγκαν της εταιρείας, "You press the button, we do the rest" (Εσείς πατάτε το κουμπί, εμείς κάνουμε τα υπόλοιπα), αναδείκνυε τη φιλοσοφία της Kodak να κάνει τη φωτογραφία προσιτή σε όλους. Στις αρχές του 20ού αιώνα, η Kodak είχε ήδη εδραιωθεί ως ηγετική εταιρεία στη βιομηχανία της φωτογραφίας. Το 1900, η εταιρεία παρουσίασε την κάμερα Brownie, μια απλή και φθηνή φωτογραφική μηχανή που κόστιζε μόλις ένα δολάριο.

Η Brownie έγινε εξαιρετικά δημοφιλής και εισήγαγε εκατομμύρια ανθρώπους στον κόσμο της φωτογραφίας. Καθ' όλη τη διάρκεια του 20ού αιώνα, η Kodak συνέχισε να καινοτομεί και να επεκτείνεται. Η εταιρεία εισήγαγε νέες τεχνολογίες όπως τα έγχρωμα φιλμ Kodachrome το 1935 και το Ektachrome το 1946. Η Kodak διαδραμάτισε επίσης σημαντικό ρόλο στη βιομηχανία κινηματογράφου, προμηθεύοντας φιλμ σε στούντιο παραγωγής ταινιών. Παρά την επιτυχία της, η Kodak αντιμετώπισε σημαντικό ανταγωνισμό, ειδικά από εταιρείες όπως η Fuji.

Η έλευση της ψηφιακής φωτογραφίας στα τέλη του 20ού αιώνα αποτέλεσε μια μεγάλη πρόκληση για την Kodak, καθώς η εταιρεία ήταν αργή να προσαρμοστεί στις νέες τεχνολογίες. Παρόλο που η Kodak ήταν από τις πρώτες εταιρείες που ανέπτυξαν ψηφιακές φωτογραφικές μηχανές, η μετάβαση από το αναλογικό στο ψηφιακό περιεχόμενο ήταν αργή

και δύσκολη. Η εταιρεία δεν κατάφερε να προσαρμοστεί γρήγορα στις νέες αγορές και έχασε σημαντικό μερίδιο από τον ανταγωνισμό. Το 2012, η Kodak κατέθεσε αίτηση για πτώχευση, αδυνατώντας να αντέξει τον ανταγωνισμό και την τεχνολογική μετάβαση. Ωστόσο, μετά από αναδιάρθρωση, η εταιρεία επέστρεψε στην αγορά, εστιάζοντας πλέον σε επαγγελματικές λύσεις εκτύπωσης, λογισμικό και ανανεώσιμες πηγές τεχνολογίας.

Η κληρονομιά της Kodak είναι αναμφισβήτητη. Η εταιρεία έπαιξε καθοριστικό ρόλο στην εξέλιξη της φωτογραφίας και έκανε τη φωτογραφία προσιτή σε εκατομμύρια ανθρώπους παγκοσμίως. Οι καινοτομίες της, από το ρολό φιλμ μέχρι τις φωτογραφικές μηχανές Brownie, καθόρισαν την πορεία της φωτογραφίας για πάνω από έναν αιώνα. Παράλληλα το 1914, Γερμανός μηχανικός και εφευρέτης Όσκαρ Μπαρνάκ (Oskar Barnack) ανέπτυξε την πρώτη λειτουργική πρωτότυπη φωτογραφική μηχανή 35mm, γνωστή ως Ur-Leica (από το "Ur", που σημαίνει "πρωτότυπος" στα γερμανικά, και "Leica", συντομογραφία του "Leitz Camera"). Η Ur-Leica ήταν μια επαναστατική καινοτομία για την εποχή της. Ήταν μικρή, ελαφριά και φορητή, επιτρέποντας στους φωτογράφους να τραβούν φωτογραφίες με ευκολία και να καταγράφουν στιγμές σε υψηλή ποιότητα. Η Ur-Leica χρησιμοποίησε φιλμ 35mm, το οποίο ήταν το ίδιο φιλμ που χρησιμοποιούσαν οι κινηματογραφικές κάμερες της εποχής. Αυτό επέτρεψε στους φωτογράφους να τραβούν πολλές φωτογραφίες σε μια σειρά χωρίς να χρειάζεται να αλλάζουν φιλμ. Η μηχανή ήταν εξοπλισμένη με έναν φακό υψηλής ποιότητας και ένα σύστημα κλείστρου που επέτρεπε τη ρύθμιση της ταχύτητας και της έκθεσης.

Το 1925 παρουσίασε την πρώτη εμπορική φωτογραφική μηχανή Leica I. Η Leica I έγινε αμέσως δημοφιλής, καθιερώνοντας το φιλμ 35mm ως το πρότυπο για τη φωτογραφία. Η Ur-Leica και οι μετέπειτα εκδόσεις της συνέβαλαν στην επανάσταση της φωτογραφίας, καθιστώντας τη φορητή φωτογραφία προσιτή σε επαγγελματίες και ερασιτέχνες φωτογράφους. Ο σχεδιασμός και οι καινοτομίες της Ur-Leica επηρέασαν βαθιά την εξέλιξη των φωτογραφικών μηχανών και έθεσαν τις βάσεις για τη σύγχρονη φωτογραφία. Η Ur-Leica βρίσκεται σήμερα σε μουσεία και συλλογές, αναγνωρισμένη ως ένα από τα σημαντικότερα ορόσημα στην ιστορία της φωτογραφίας.

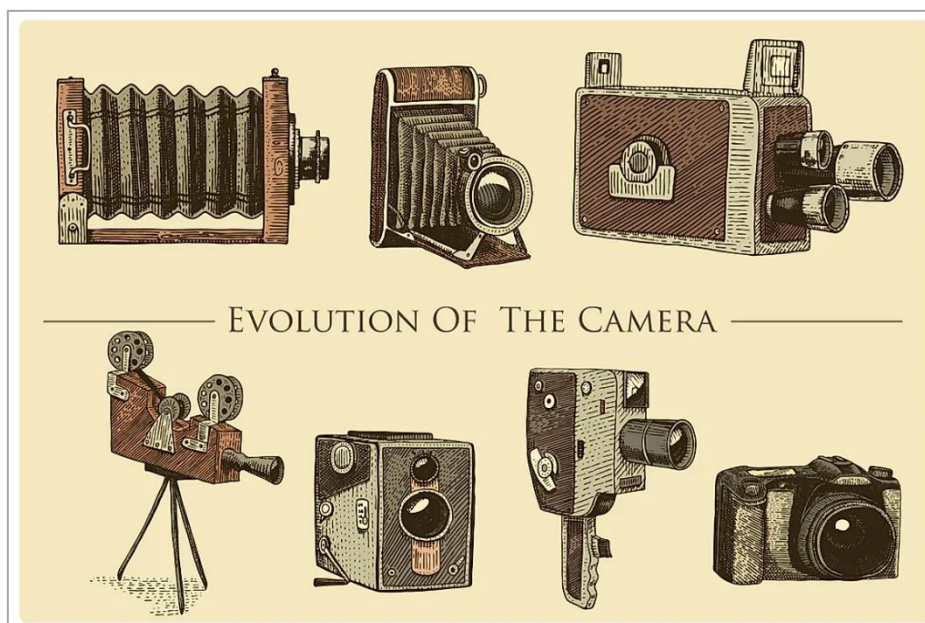
Έπειτα από μία δεκαετία, το 1936 εφευρέθηκε η πρώτη Single Lens Reflex (SLR) κάμερα ήταν η Kine Exakta, που παρουσιάστηκε το 1936 από την Ihagee, μια εταιρεία με έδρα τη Δρέσδη της Γερμανίας. Η Kine Exakta ήταν η πρώτη κάμερα που χρησιμοποίησε φιλμ 35mm και έθεσε τις βάσεις για την ανάπτυξη των μετέπειτα SLR καμερών. Η Kine Exakta είχε έναν καθρέπτη που επέτρεπε στον φωτογράφο να βλέπει τη σκηνή απευθείας μέσα από τον φακό. Χρησιμοποιούσε, όπως οι προγενέστερες κάμερες, φιλμ 35mm, το οποίο ήταν το ίδιο που

χρησιμοποιούσαν οι κινηματογραφικές κάμερες, καθιστώντας την ελαφριά και φορητή και διέθετε μηχανικό κλείστρο που επέτρεπε ακριβή έλεγχο της ταχύτητας του κλείστρου. Η τεχνολογία και ο σχεδιασμός της επηρέασαν σημαντικά τις μεταγενέστερες φωτογραφικές μηχανές. Μετά τον πόλεμο, αρκετές εταιρείες, όπως η Nikon και η Canon, βελτίωσαν τον σχεδιασμό των SLR καμερών, οδηγώντας στην καθιέρωση των SLR καμερών ως το κυρίαρχο είδος φωτογραφικών μηχανών για επαγγελματική και ερασιτεχνική χρήση. Η Nikon ιδρύθηκε το 1917 στην Ιαπωνία. Η εταιρεία προήλθε από τη συγχώνευση τριών μικρότερων εταιρειών οπτικής και από την αρχή, επικεντρώθηκε στην κατασκευή φακών και άλλων οπτικών συστημάτων για στρατιωτικές και βιομηχανικές χρήσεις.

Το 1957, παρουσίασε την πρώτη της Single Lens Reflex (SLR) κάμερα, την Nikon F, η οποία καθόρισε τα πρότυπα για επαγγελματικές φωτογραφικές μηχανές. Η Nikon F ήταν ανθεκτική, αξιόπιστη και προσέφερε ένα ευρύ φάσμα εναλλακτικών φακών και αξεσουάρ, καθιστώντας την αγαπημένη επιλογή για επαγγελματίες φωτογράφους. Σημαντική επιπρόσθετη εφεύρεση της Nikon είναι το σύστημα F-Mount, το οποίο παραμένει σε χρήση μέχρι σήμερα, επιτρέποντας τη συμβατότητα με μια τεράστια γκάμα φακών. Η επιτυχία της Nikon F οδήγησε την εταιρεία να συνεχίσει να καινοτομεί και να επεκτείνει τη σειρά των προϊόντων της. Συνέχισε να εξελίσσει την αρχική της σειρά SLR με τις Nikon F2 (1971) και Nikon F3 (1980), οι οποίες ενσωμάτωναν βελτιώσεις στον σχεδιασμό και τις λειτουργίες. Εκτός από τις επαγγελματικές SLR, η Nikon παρουσίασε και πιο προσιτές σειρές για ερασιτέχνες φωτογράφους, όπως η σειρά Nikkormat.

Ο κύριος παράγοντας που τεκμηριώθηκε από πολλούς ως η αρχή της Nikon, ήταν η Canon. Η Canon ιδρύθηκε το 1933 ως Precision Optical Instruments Laboratory από τους Takeshi Mitarai, Goro Yoshida, Saburo Uchida και Takeo Maeda. Το αρχικό τους όραμα ήταν να δημιουργήσουν μια φωτογραφική μηχανή υψηλής ποιότητας που να μπορεί να ανταγωνιστεί τα γερμανικά μοντέλα της εποχής, όπως η Leica και η Contax. Το 1934, παρουσίασαν την πρώτη τους φωτογραφική μηχανή, την Kwanon η οποία αν και δεν μπήκε ποτέ σε μαζική παραγωγή, έθεσε τα θεμέλια για την ανάπτυξη της πρώτης εμπορικής φωτογραφικής μηχανής της εταιρείας. Η πρώτη εμπορικά διαθέσιμη φωτογραφική μηχανή της εταιρείας ήταν η Hansa Canon, η οποία κυκλοφόρησε το 1936. Αυτή η φωτογραφική μηχανή χρησιμοποιούσε φακούς της Nippon Kogaku K.K. (που αργότερα έγινε η Nikon), δημιουργώντας έτσι μία από τις πρώτες συνεργασίες στην ιστορία της φωτογραφίας. Μετά τον Β' Παγκόσμιο Πόλεμο, η Canon ξεκίνησε να επεκτείνεται διεθνώς και να διαφοροποιεί τα προϊόντα της. Στη δεκαετία του 1950, η Canon εισήγαγε την Canon IVSb, μια τηλεμετρική κάμερα που ήταν δημοφιλής στους επαγγελματίες φωτογράφους. Η σειρά Canon 7, που

κυκλοφόρησε το 1961, ήταν επίσης σημαντική, καθώς ήταν από τις πρώτες κάμερες με ενσωματωμένο φωτόμετρο. Επίσης, άρχισε να επεκτείνει τη γκάμα των προϊόντων της, συμπεριλαμβάνοντας εκτυπωτές και άλλες συσκευές γραφείου. Η δεκαετία του 1970 και του 1980 ήταν μια περίοδος σημαντικής ανάπτυξης και καινοτομίας για την Canon στον τομέα των SLR φωτογραφικών μηχανών. Η πρώτη επαγγελματική SLR κάμερα της Canon, η F-1, κυκλοφόρησε το 1971 και έγινε γρήγορα δημοφιλής στους επαγγελματίες φωτογράφους. Μόλις πέντε χρόνια μετά, το 1976, εφευρέθηκε η Canon AE-1, η οποία ήταν η πρώτη κάμερα που ενσωμάτωσε μικροεπεξεργαστή, φέρνοντας την αυτόματη έκθεση σε μαζική αγορά. Αυτή η κάμερα έγινε μία από τις πιο δημοφιλείς SLR όλων των εποχών. [\[1\]\[2\]\[3\]\[4\]](#)



Εικόνα 8 Evolution of photography

ΚΕΦΑΛΑΙΟ 3

3. ΨΗΦΙΑΚΗ ΕΠΟΧΗ

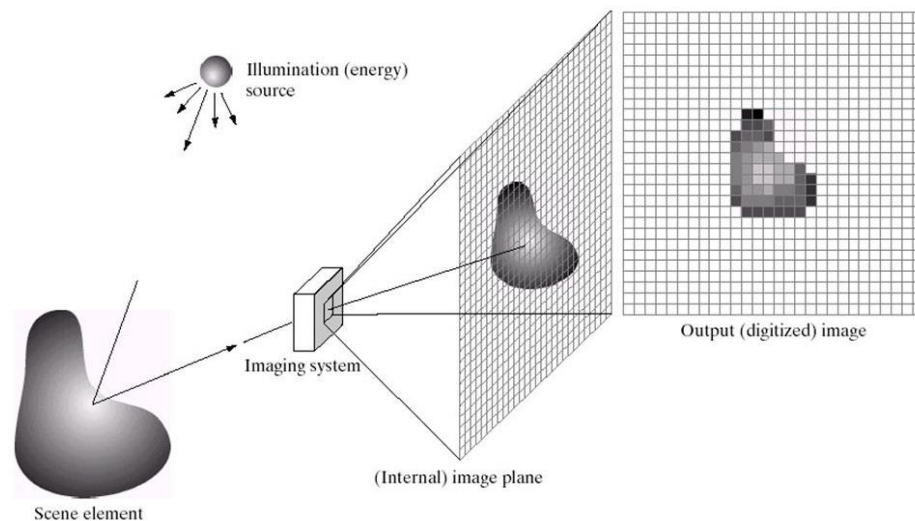
3.1 Ψηφιοποίηση

Η Ψηφιοποίηση ή αλλιώς Digitization, είναι η προσέγγιση της μετατροπής αναλογικών πληροφοριών ή δεδομένων σε ψηφιακή μορφή, με τη βοήθεια ηλεκτρονικών μέσων. Τα δεδομένα αυτά μπορεί να είναι έγγραφα, εικόνες, βίντεο, ήχοι ακόμα και άλλα ακόμα. Στις μέρες μας, η ψηφιοποίηση φαίνεται να παίζει σημαντικό ρόλο σε πολλούς τομείς της καθημερινότητάς μας, όπως στον τομέα της εκπαίδευσης, της υγείας, των επιχειρήσεων και πολλών άλλων ακόμα. Ωστόσο, παρόλο της συχνής εμφάνισής της, η ψηφιοποίηση φέρνει και μερικά από τα οφέλη της όπως για παράδειγμα, η εύκολη πρόσβαση, στην οποία η προσβασιμότητα στα δεδομένα είναι αρκετά πιο εύκολη, αρκεί να υπάρχει πρόσβαση σε κάποιο δίκτυο. Επίσης, η αποθήκευση και η συντήρηση των αρχείων σε ψηφιακή μορφή, όπου τα έγγραφα και αρχεία δεν φθείρονται όπως στη φυσική μορφή τους, με την πάροδο του χρόνου και έτσι μπορούν να είναι διαθέσιμα για αρκετά χρόνια έως και για πάντα, όπως επίσης και η αναζήτηση και ανάκτηση των δεδομένων που οι πληροφορίες είναι εύκολο να αναζητηθούν μέσω αρκετών μηχανών αναζήτησης ή ακόμα και βάσεων δεδομένων.

Παρόλα αυτά τα οφέλη της ψηφιοποίησης τα βρίσκουμε και στο κομμάτι των επιχειρήσεων που η συνεργασία και διαμερισμός των αρχείων κάνει την ροή της εργασίας ακόμα πιο εύκολη. Πέρα από τα οφέλη και τον σημαντικό ρόλο στη καθημερινότητα των ανθρώπων, η ψηφιοποίηση φαίνεται να εφαρμόζεται και σε έγγραφα, όπου αρκετές εταιρείες όσο και οργανισμοί μετατρέπουν τα ψηφιακά τους έγγραφα σε ψηφιοποιημένα, με τη διαχείριση των εγγράφων να είναι ακόμα πιο εύκολη όπως και η ανάκτηση των δεδομένων να είναι ακόμα ταχύτερη. Επιπλέον, φαίνεται να κάνει την εφαρμογή της και σε αρχαιολογικά ευρήματα, όπου οι πολιτιστικοί οργανισμοί έχουν την δυνατότητα να ψηφιοποιήσουν μουσειακά αντικείμενα με έργα τέχνης, με αποτέλεσμα να μπορεί το κοινό να τα προβάλει μέσω διαδικτυακών εκθέσεων. Παρόλα αυτά στον τομέα της υγείας, τα ιατρικά αρχεία έχουν αρχίσει να ψηφιοποιούνται με αποτέλεσμα να μπορούν τόσο οι γιατροί όσο και οι ασθενείς, έχουν ταχύτερη και ευκολότερη πρόσβαση σε ιατρικά αρχεία και έγγραφα, ενώ με αυτό τον τρόπο διευκολύνεται η ανάλυση των δεδομένων για ερευνητικούς σκοπούς.

Όπως στους παραπάνω τομείς, έτσι και στον τομέα της εκπαίδευσης η ψηφιοποίηση κάνει την εφαρμογή της σε βιβλιοθήκες και εκπαιδευτικά ιδρύματα ψηφιοποιώντας βιβλία, άρθρα και εκπαιδευτικό υλικό, με αποτέλεσμα οι εκπαιδευτικοί και οι μαθητές να έχουν μία εύκολη και ταχύτερη πρόσβαση σε αυτά. Παρόλο που η ψηφιοποίηση έχει μπει για τα καλά στη καθημερινότητα των ανθρώπων με τις θετικές εξελίξεις της, είναι αρκετές οι φορές που υπάρχουν κάποιοι περιορισμοί όσο και κίνδυνοι. Ανάμεσα τους και η ασφάλεια των ψηφιακών δεδομένων, τα οποία είναι ευάλωτα στο διαδίκτυο και απαιτούν υψηλά μέτρα προστασίας. Επίσης, με τη πάροδο των χρόνων και με την τεχνολογία να αναπτύσσεται ραγδαία όλο και περισσότερο, υπάρχει κίνδυνος οι ψηφιοποιημένες μορφές αρχείων και εγγράφων, να μην υποστηρίζονται με αποτέλεσμα να χαθεί η προσβασιμότητα τους. Τέλος εκτός των παραπάνω, είναι σημαντικό να αναφέρουμε και το κόστος, πόσο μάλλον και στις μεγάλες συλλογές αρχείων και δεδομένων.

Η τάση προς την πλήρη ψηφιοποίηση αυξάνεται σημαντικά μέρα με τη μέρα και με τη βελτίωση των τεχνολογιών όπως το cloud computing, η τεχνητή νοημοσύνη και το Internet Of Things, τα έξυπνα συστήματα θα επιτρέψουν ακόμη και πιο προηγμένη ανάλυση και διαχείριση των ψηφιακών δεδομένων, ανοίγοντας ορίζοντες για νέους τρόπους συνεργασίας, τόσο της καινοτομίας όσο και της ανάπτυξης.



Εικόνα 9 Digitalization

Η ψηφιοποίηση δεν είναι απλώς μια τεχνολογική διαδικασία, αλλά μια μετασχηματιστική δύναμη που επηρεάζει όλους τους παράγοντες της ανθρώπινης ζωής ξεκινώντας από την εικόνα. [\[3\]](#)[\[4\]](#)[\[5\]](#)

3.1.1 Ψηφιοποίηση εικόνας

Η ψηφιοποίηση της εικόνας ξεκίνησε με αποτέλεσμα την αποθήκευση και επεξεργασία τόσο οπτικών όσο και οπτικοακουστικών δεδομένων από υπολογιστές. Η μετάβαση από την αναλογική στην ψηφιακή μορφή εικόνας προέκυψε κυρίως από τεχνολογικές εξελίξεις στα μαθηματικά, την επιστήμη των υπολογιστών, την ηλεκτρονική και τη μηχανική εικόνας, στις οποίες είχε βαθιές επιρροές σε διάφορους τομείς όπως η τηλεπικοινωνία, η τέχνη και η επιστήμη. Τα πρώτα βήματα της ψηφιοποίησης της εικόνας φαίνεται να έγιναν κατά τις δεκαετίες του 1950 και 1960, όταν οι ερευνητές ξεκίνησαν να πειραματίζονται με τον τρόπο τον οποίο οι εικόνες θα μπορούσαν να αναπαραχθούν και να αναλυθούν από τους ηλεκτρονικούς υπολογιστές.

Το πρώτο ψηφιακό σύστημα εικόνας σχεδιάστηκε για να χρησιμοποιηθεί κυρίως για στρατιωτικούς αλλά και επιστημονικούς σκοπούς, όπως για την ανάλυση δορυφορικών εικόνων και για την βελτίωση των ακτινογραφιών για τις ιατρικές εφαρμογές. Μία από τις αρχικές εφαρμογές της ψηφιοποίησης της εικόνας έγινε για την ανάπτυξη των συστημάτων σάρωσης, καθώς στις αρχές της δεκαετίας του 1950, οι επιστήμονες και ερευνητές χρησιμοποίησαν σαρωτές εικόνας για τη μετατροπή των αναλογικών εικόνων σε ψηφιακά δεδομένα, όπου θα μπορούσαν να αναλυθούν στη συνέχεια από τους ηλεκτρονικούς υπολογιστές. Με τη πάροδο των χρόνων, καθώς η τεχνολογία άρχισε να αναπτύσσεται σημαντικά, η δυνατότητα ψηφιακής αναπαράστασης άρχισε να αποτυπώνεται με μεγαλύτερη ακρίβεια.

Από τις αρχές της δεκαετίας του 1960, όπως αρκετοί οργανισμοί έτσι και η NASA εφάρμοσαν τη ψηφιακή επεξεργασία εικόνας, με κύριο στόχο τους να βελτιώσουν την ανάλυση των εικόνων από το διάστημα. Οι δορυφόροι της NASA άρχισαν να αναπαράγουν εικόνες χαμηλής ανάλυσης και με βάση μαθηματικών μεθόδων οι επιστήμονες και ερευνητές προσπάθησαν να βελτιώσουν τις εικόνες αυτές. Έτσι και δημιουργήθηκαν οι πρώτες βάσεις για την επεξεργασία εικόνας, στις οποίες σαν κύρια τεχνική τους ήταν τον φιλτράρισμα του θορύβου και η αύξηση της αντίθεσης. Καθώς και μία από της σημαντικότερες ανακαλύψεις για την εποχή εκείνη, ήταν η εισαγωγή του αλγορίθμου ταχείας μετασχηματισμού Fourier ή αλλιώς Fast Fourier Transforming (FFT), από τους James Cooley και John Turkey το 1965. Με βάση του αλγορίθμου αυτού ο υπολογισμός των μετασχηματισμών Fourier ήταν ταχύτερος, με τη χρήση τους για την βελτίωση και ανάλυση των εικόνων, καθώς και την αναπαράσταση εικόνων σε διαφορετικές συχνότητες, διευκολύνοντας έτσι την απομάκρυνση του θορύβου.

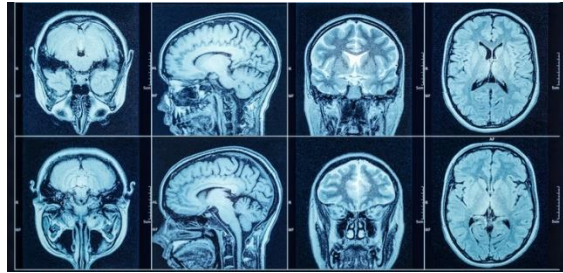
Κατά τη δεκαετία του 1970 η ψηφιοποίηση της εικόνας εξελίχθηκε ακόμα περισσότερο με την ανάπτυξη των ψηφιακών καμερών. Πιο συγκεκριμένα, η πρώτη ψηφιακή φωτογραφική μηχανή έκανε την εμφάνισή της το 1975 από τον μηχανικό της εταιρείας Kodak, Ste. Στη ψηφιακή αυτή κάμερα χρησιμοποίησαν έναν αισθητήρα Charge Coupled Device (CCD) για τη λήψη και αποθήκευση της εικόνας σε ψηφιακή μορφή. Αν και η ανάλυση της εικόνας αυτής ήταν αρκετά χαμηλή, μόλις 0.01 megapixel, αποτέλεσε το πρώτο βήμα αλλά και βάση για τη σημαντική αυτή πρόοδο στη ψηφιοποίηση εικόνων. [\[3\]\[4\]\[5\]](#)

3.1.2 Εξέλιξη και εκβιομηχάνιση

Κατά τη δεκαετία του 1980, καθώς τα υπολογιστικά συστήματα και η μείωση του κόστους αποθήκευσης δεδομένων ξεκίνησαν να αναπτύσσονται, μαζί τους εφαρμόστηκε και η μαζική χρήση της ψηφιοποίησης της εικόνας. Πέρα από τους αρχικούς τομείς όπου και αναπτύχθηκε η ψηφιακή επεξεργασία εικόνας, άρχισε να εφαρμόζεται και σε άλλους τομείς της βιομηχανίας όπως στην ιατρική, στις επιστημονικές έρευνες, στη ψηφιακή και πολιτιστική κληρονομιά ακόμα και στον τομέα της φωτογραφίας και ψυχαγωγίας των ανθρώπων. Ειδικότερα, στον τομέα της ιατρικής η ψηφιοποίηση της εικόνας επέτρεψε να αναπτυχθεί και να διαδοθεί η μαγνητική τομογραφία (Magnetic Resonance Imaging - MRI) καθώς και η αξονική τομογραφία (Computed Tomography - CT), οι οποίες είναι βασισμένες σε ψηφιακά δεδομένα για την ανάλυση και καταγραφή εικόνων του ανθρώπινου σώματος. Επίσης στον τομέα των επιστημονικών ερευνών, χρησιμοποιείται για την ανάλυση δεδομένων σε μελέτες δορυφορικών ή μικροσκοπικών εικόνων, ώστε να πετύχουν την παρακολούθηση του περιβάλλοντος ή για τη μελέτη βιολογικών και χημικών φαινομένων αντίστοιχα.

Σχετικά με τη ψηφιακή τέχνη η ψηφιοποίηση των εικόνων έκανε την εμφάνισή της σε μουσεία και βιβλιοθήκες ώστε να προστατευτούν αλλά και να διαδοθούν έργα τέχνης αλλά και σπουδαία αρχεία. Αυτό εφαρμόστηκε για τον λόγο του ότι οι αναλογικές εικόνες μπορούν να υποστούν φθορές με την πάροδο των χρόνων, ενώ αντίθετα οι ψηφιακές αυτές αναπαραγωγές μπορούν να διατηρηθούν για πάρα πολλά χρόνια και να είναι προσβάσιμες από όλο τον κόσμο. Ένας ακόμα βασικός τομέας της ψηφιοποίησης είναι στη βιομηχανία του κινηματογράφου, της φωτογραφίας, ακόμα και των βιντεοπαιχνιδιών, καθώς εικόνες υψηλής ανάλυσης και ευκρίνειας αλλά και προηγμένες τεχνικές επεξεργασίας εφαρμόζονται για να την ανάπτυξη οπτικών αλλά και ακουστικών εφέ για την υψηλή ποιότητα γραφικών.

[\[3\]\[4\]\[5\]](#)



Εικόνα 10 MRI

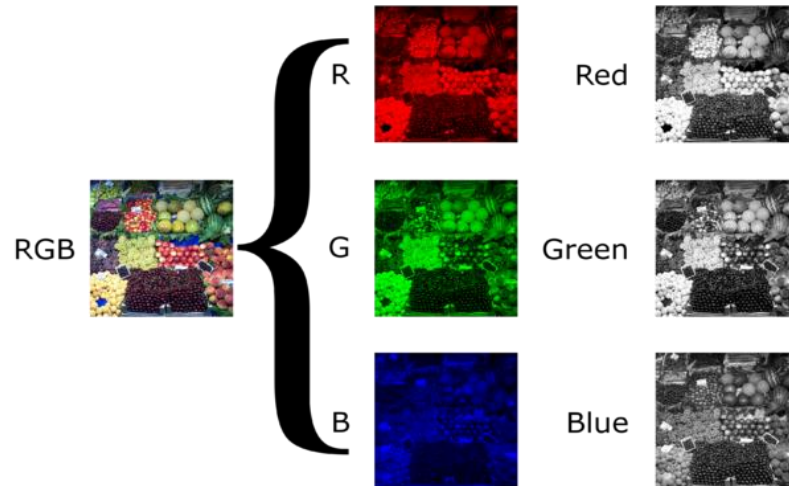
3.1.3 Τα βασικά βήματα ψηφιοποίησης εικόνας

Με τη πάροδο των χρόνων η ψηφιοποίηση της εικόνας ανέπτυξε σημαντικά τέσσερα βασικά βήματα, τη δειγματοληψία, η οποία αναφέρεται στη διαδικασία που η εικόνα χωρίζεται σε ένα σύνολο από σημεία pixels), τα οποία βρίσκονται σε ένα πλέγμα, όπου όσο υψηλότερη είναι η ανάλυση μίας εικόνας, τόσο μεγαλύτερος είναι ο αριθμός των pixels, το οποίο συνεπάγεται και με καλύτερη λεπτομέρεια. Η ανάλυση μετράτε σε κουκίδες ανά ίντσα (DPI –Dots Per Inch), επομένως μία εικόνα με υψηλό DPI θα περιέχει περισσότερες πληροφορίες, άρα θα είναι καθαρότερη και με καλύτερη ευκρίνεια. Ένα ακόμα βήμα της, είναι η κβαντοποίηση, στην οποία κάθε pixel αντιστοιχεί σε ένα χρώμα ή σε ένα βαθμό φωτεινότητας.

Ουσιαστικά η κβαντοποίηση είναι η διαδικασία ανάθεσης μίας διακριτής τιμής σε κάθε pixel, τα οποία είτε θα είναι με βάση του χρωματικού μοντέλου RGB (Red - Green - Blue), είτε με βάση της γκριζας κλίμακας στη περίπτωση ασπρόμαυρων εικόνων. Στη περίπτωση του μοντέλου RGB οι τιμές που λαμβάνονται είναι από 0 έως 255, με αποτέλεσμα να δημιουργούνται περίπου 16,7 εκατομμύρια συνδυασμοί χρωμάτων. Επιπλέον, μία εικόνα όταν ψηφιοποιηθεί μπορεί να αποθηκευτεί σε αρκετές μορφές αρχείου ανάλογα με τη χρήση της και τις απαιτήσεις για αποθηκευτικό χώρο, οι πιο σύνηθες μορφές είναι η JPEG, η PNG και η TIFF κ.α.. Φυσικά, κάθε μία από τις μορφές αυτές έχει και τα πλεονεκτήματα αλλά και τα μειονεκτήματά της. Για παράδειγμα, η μορφή αρχείου JPEG χρησιμοποιεί συμπίεση με απώλειες για τη μείωση του μεγέθους του αρχείου, ενώ η μορφή PNG χρησιμοποιεί συμπίεση χωρίς απώλειες, διατηρώντας έτσι την αρχική ποιότητα της εικόνας.

Τέλος, ένα ακόμα βήμα είναι η επεξεργασία της ψηφιοποιημένης εικόνας, καθώς μόλις ολοκληρωθεί η ψηφιοποίηση της μπορεί με την βοήθεια κάποιου λογισμικού επεξεργασίας εικόνας να υποβληθεί σε διάφορες μορφές. Η επεξεργασία αυτή μπορεί να είναι ως προς τη μείωση θορύβου, είτε ως προς τη διόρθωση των χρωμάτων και της φωτεινότητας, είτε για τη βελτίωση της ευκρίνειας όπως και άλλες ακόμα τεχνικές επεξεργασίας ώστε να αποδοθεί

καλύτερη η παρουσίαση ή ανάλυση της εικόνας. Επίσης, εκτός από τις τεχνικές επεξεργασίας, με τη ψηφιοποίηση είναι δυνατόν να γίνει και ανακατασκευή εικόνων, κάνοντας νέες παραλλαγές από αυτές των πρωτότυπων εικόνων. [6]



Εικόνα 11 Grayscale to RGB

3.2 Ψηφιακή εικόνα

Η ψηφιακή εικόνα είναι ένα σύνολο από δεδομένα, τα οποία αναπαριστούν μία οπτική πληροφορία, όπως για παράδειγμα μία γραφική παράσταση ή μία φωτογραφία, όπου έχει τη μορφή που δίνει τη δυνατότητα να επεξεργαστεί από μία ψηφιακή συσκευή είτε από έναν υπολογιστή. Η βασική μονάδα μέτρησης μίας ψηφιακής εικόνας είναι τα pixels ή αλλιώς εικονοστοιχεία, τα οποία είναι μικροσκοπικά τετράγωνα χρωματισμών, όπου όταν ενωθούν χιλιάδες ή εκατομμύρια pixels μεταξύ τους μας δίνουν μία πλήρη ψηφιακή εικόνα. Η ανάλυση μίας ψηφιακής εικόνας βασίζεται στον αριθμό των εικονοστοιχείων (pixels) που περιέχει η εικόνα ανά ίντσα.

Η υψηλής ανάλυσης εικόνα αντιπροσωπεύει περισσότερα pixels και επομένως μεγαλύτερη λεπτομέρεια, ενώ αντίθετα εικόνες με χαμηλή ανάλυση μπορεί να εμφανίζονται με θόλωση (pixelation) ή τραχιές όταν προβάλλονται σε μεγάλες οθόνες ή ακόμα κατά την εκτύπωσή τους σε μεγάλες διαστάσεις. Επίσης, η ποιότητα της εικόνας εξαρτάται και στο βάθος χρώματος (Color Depth), δηλαδή από το πόσα bits χρησιμοποιούνται για την αναπαράσταση του χρώματος σε κάθε ένα pixel. Για παράδειγμα, μία εικόνα με βάθος χρώματος 24-bit (True Color) μπορεί να αναπαραστήσει πάνω από δεκαέξι (16) εκατομμύρια χρώματα, επιτρέποντας πολύ ρεαλιστικές αλλά και ζωντανές εικόνες.

Πλέον, οι ψηφιακές εικόνες έχουν γίνει απαραίτητες σε πολλούς τομείς των σύγχρονων αναγκών, από την ψυχαγωγία μέσω της τέχνης της φωτογραφίας, του κινηματογράφου και της γραφιστικής μέχρι των επιστημονικών ερευνών και της ιατρικής. Έτσι, η αποθήκευση και η μετάδοση των οπτικών πληροφοριών γίνεται με μεγαλύτερη ακρίβεια και ευελιξία, ενώ η δυνατότητα της επεξεργασίας και της διανομής τους με βάση μία ποικιλία μορφών τις καθιστούν κατάλληλες για τη χρήση εφαρμογών, από απλά γραφικά έως σύνθετα επιστημονικά δεδομένα, όπου πλέον αυτά υλοποιούνται με μεγαλύτερη ευκολία [3][4][5]

3.2.1 Η Δομή μίας ψηφιακής εικόνας

Μία ψηφιακή εικόνα είναι ένας πίνακας από χιλιάδες ή εκατομμύρια pixels, όπου καθένα από αυτά έχουν μία πληροφορία με βάση το χρωματισμό αλλά και τη φωτεινότητά τους, στο σημείο που βρίσκονται. Οι μορφές που μπορούν να εκφράσουν τα χρώματα των pixels είναι κυρίως είναι του RGB (Red-Green-Blue), όπου η αναλογία τους καθορίζει την απόχρωση του κάθε pixel. Για παράδειγμα, σε περίπτωση που τα κανάλια RGB έχουν τιμή 255 ή 0, το pixel αυτό θα έχει χρώμα είτε λευκό, είτε μαύρο αντίστοιχα. Παρόλα αυτά, οι εικόνες χωρίζονται σε δύο κύριες κατηγορίες, στις Ραστερικές και στις Διανυσματικές.

Όπως τα αρχεία PNG, JPG και BMP, έτσι και οι raster εικόνες (raster images) είναι ψηφιακές εικόνες που αποτελούνται από ένα πλέγμα εικονοστοιχείων (pixels), οι οποίες βασίζονται στην ανάλυση (resolution), δηλαδή τον αριθμό των εικονοστοιχείων που περιέχουν ανά ίντσα (dots per inch – dpi). Κάθε pixel περιέχει μία πληροφορία σχετικά με το χρώμα και την φωτεινότητά του, δημιουργώντας μία συνολική εικόνα με βάση το συνδυασμό όλων των pixels. Έτσι, όσο περισσότερα εικονοστοιχεία (pixels) έχει μία εικόνα, τόσο μεγαλύτερη και πιο λεπτομερής θα φαίνεται η ανάλυσή της, αλλά όσο καλύτερη ανάλυση έχει η εικόνα, τόσο μεγαλύτερο θα είναι το μέγεθος του αρχείου. Οι raster images χρησιμοποιούνται κυρίως για φωτογραφίες και λεπτομερείς εικόνες, λόγω της δυνατότητάς τους στο να αποτυπώσουν πιο σύνθετα χρώματα αλλά και με περισσότερη λεπτομέρεια. Ωστόσο, αν μία raster εικόνα μεγεθυνθεί σε υπερβολικό βαθμό, τείνει να χάσει την ποιότητά της αλλά και να εμφανίσει θόλωση (pixelation).

Σε αντίθεση με τις raster εικόνες, οι οποίες βασίζονται σε pixels, οι διανυσματικές εικόνες (vector images), βασίζονται σε μαθηματικές εξισώσεις και γεωμετρικά σχήματα όπως γραμμές, καμπύλες αλλά και πολύγωνα, αντί για pixels. Έτσι οι διανυσματικές εικόνες, όπως και τα αρχεία εικόνας SVG (Scalable Vector Graphics), μπορούν να σμικρύνονται ή να μεγεθύνονται αντίστοιχα σε πολύ μεγάλο βαθμό χωρίς να μεταβάλλεται η ποιότητα της

εικόνας τους, καθώς τα σχήματα ανασχεδιάζονται κάθε φορά με βάση μαθηματικών υπολογισμών. Για τον λόγο αυτό, οι διανυσματικές εικόνες μπορούν να χρησιμοποιηθούν ιδανικά σε λογότυπα, γραφήματα αλλά και σε γραφικά σχέδια που απαιτούν καθαρές γραμμές και ευκρινείς άκρες σε οποιοδήποτε μέγεθος, ανεξάρτητα της ανάλυσης της οθόνης ή της κλίμακας. Κοινές μορφές διανυσματικών εικόνων είναι τα SVG (Scalable Vector Graphics), AI (Artificial intelligence) και EPS (Encapsulated PostScript). Ωστόσο, το μέγεθος των διανυσματικών εικόνων είναι αρκετά μικρότερο από αυτό των raster εικόνων, ειδικότερα όταν περιέχουν απλά σχήματα, για τον λόγο ότι δεν είναι ιδανικές για τη δημιουργία σύνθετων εικόνων με πολλές λεπτομέρειες, όπως φωτογραφίες. [\[3\]\[4\]\[5\]](#)

3.2.2 Μορφές ψηφιακών εικόνων

Υπάρχουν πολλές μορφές αρχείων στις οποίες μπορούν να αποθηκευτούν οι ψηφιακές εικόνες ανάλογα με τον σκοπό και τις ανάγκες της χρησιμότητάς τους. Οι πιο διαδεδομένες μορφές αποθήκευσης είναι οι JPEG, οι PNG, οι GIF, οι SVG, οι BMP, οι TIFF, οι WEBP και οι HEIC. Οι JPEG (Joint Photographic Experts Group) είναι μία από τις πιο κοινές μορφές εικόνων, καθώς χρησιμοποιείται κυρίως για φωτογραφίες αλλά και για εικόνες με πολλά χρώματα. Αυτή η μορφή χρησιμοποιεί συμπίεση με απώλειες, όπου αυτό σημαίνει πως κάποιες εικόνες χάνουν τη μεγάλη λεπτομέρεια τους για να μπορέσει η εικόνα να έχει μικρότερο μέγεθος. Αντίθετα, στη PNG (Portable Network Graphics) μορφή, τα αρχεία έχουν μεγαλύτερο μέγεθος σε σχέση με τα JPEG καθώς η μορφή αυτή υποστηρίζει διαφάνεια και η ποιότητα της εικόνας είναι καλύτερη και χωρίς απώλειες στην ανάλυση τους. Χρησιμοποιείται κυρίως τόσο για γραφικά, όσο και στη τέχνη της γραφιστικής.

Η GIF (Graphics Interchange Format) μορφή χρησιμοποιείται κυρίως για απλές εικόνες και κινούμενα γραφικά (Animation), καθώς έχει περιορισμένη παλέτα χρωμάτων και υποστηρίζει μόνο 256 χρώματα, αλλά μπορεί να υποστηρίξει κινούμενα σχέδια. Αντιθέτως, δεν είναι τόσο κατάλληλη μορφή για φωτογραφίες. Μία ακόμα μορφή που χρησιμοποιείται για γραφικά και εικονογράμματα και πιο συγκεκριμένα για διανυσματικά γραφικά, είναι η SVG (Scalable Vector Graphics) μορφή, η οποία είναι βασισμένη σε HTML και CSS κώδικα. Είναι κατάλληλη για εικόνες που απαιτούν διατήρηση ποιότητας ανεξαρτήτως μεγέθους, καθώς το μέγεθος τους μπορεί να κλιμακωθεί χωρίς να επηρεαστεί η ποιότητα και η ανάλυση της εικόνας. Επίσης, έχουμε τη BMP (Bitmap) μορφή, στην οποία τα αρχεία δεν συμπιέζονται και το μέγεθος του κάθε αρχείου είναι μεγαλύτερο και χωρίς απώλειες ποιότητα.

Η χρήση της μορφής αυτής είναι αρκετά σπάνια συγκριτικά με άλλες μορφές, λόγω του υψηλού μεγέθους των αρχείων. Ακόμα μία μορφή, η οποία χρησιμοποιείται κυρίως για επαγγελματικές εκτυπώσεις και για αρχεία υψηλής ποιότητας και ανάλυσης είναι η TIFF (Tagged Image File Format), όπου τα αρχεία που αποθηκεύει μπορεί να είναι χωρίς συμπίεση ή με συμπίεση αλλά χωρίς απώλειες στη ποιότητά τους. Ωστόσο, η Google ανέπτυξε μία μορφή την WEBP, για τη βελτιστοποίηση των εικόνων στο διαδίκτυο. Η μορφή αυτή συνδυάζει τα πλεονεκτήματα των μορφών JPEG και PNG αντίστοιχα, προσφέροντας μικρότερο μέγεθος αρχείων χωρίς μεγάλη απώλεια ποιότητας και ανάλυσης. Τέλος, έχουμε και τη μορφή HEIC (High Efficiency Image Coding η οποία χρησιμοποιείται κυρίως στις συσκευές της Apple, προσφέροντας υψηλή ποιότητα και ανάλυση των εικόνων, αλλά με μικρότερο μέγεθος αρχείου. Φυσικά, αυτές είναι μερικές από τις πιο κοινές και πιο διαδεδομένες μορφές εικόνων, όπου κάθε μία από αυτές χρησιμοποιείται ανάλογα με τις ανάγκες της χρήσης τους. [\[3\]](#)[\[4\]](#)[\[5\]](#)

	BMP	JPG	PNG	GIF
Compressed	FALSE	TRUE	TRUE	TRUE
Lossless	TRUE	FALSE	TRUE	TRUE
Transparency	FALSE	FALSE	TRUE	TRUE
Translucency	FALSE	FALSE	TRUE	FALSE
Recommended for photographs	FALSE	TRUE	FALSE	FALSE
Recommended for static graphics/icons	FALSE	FALSE	TRUE	FALSE
Recommended for animated graphics/icons	FALSE	FALSE	FALSE	TRUE

Εικόνα 12 Image Formats

3.3 Ψηφιακές κάμερες

Με βάση τις καινοτομίες που αναπτύχθηκαν στον τομέα της φωτογραφίας αλλά και τις ηλεκτρονικές τεχνολογίες που έκαναν την εμφάνισή τους κατά τις δεκαετίες του 1960 και 1970, έτσι και οι ψηφιακές φωτογραφικές μηχανές και κάμερες άρχισαν να αναπτύσσονται και να εξελίσσονται το ίδιο σημαντικά. Κατά τη δεκαετία του 1960 και με το πέρασμα των επόμενων δεκαετιών, η ανάπτυξη των ψηφιακών φωτογραφικών μηχανών αλλά και καμερών αναπτύχθηκε με γρήγορους ρυθμούς, καθώς έγινε μεγάλη πρόοδος στον τομέα της

τεχνολογίας και οι ανάγκες για νέες μεθόδους αποτύπωσης των εικόνων άρχισαν να αυξάνονται ραγδαία.

Μία από τις πιο σημαντικές εφευρέσεις την εποχή εκείνη, όπου έκανε την εμφάνισή της στη ψηφιακή φωτογραφία ήταν ο αισθητήρας CCD (Charge–Coupled Device), όπου και αναπτύχθηκε το 1969 από τους Willard Boyle και George E. Smith στα εργαστήρια της Bell Labs. Οι δύο αυτοί επιστήμονες ανακάλυψαν μία νέα μέθοδο για την εποχή εκείνη, ώστε να γίνει δυνατή η μετατροπή των φωτεινών σημάτων σε ψηφιακή μορφή μέσω του αισθητήρα CCD, κάτι που έδινε τη δυνατότητα αποτύπωσης των εικόνων σε ψηφιακή μορφή αντί για τα γνωστά Film. Έτσι, η πρώτη ψηφιακή φωτογραφική μηχανή κατασκευάστηκε το 1975 από τον Στίβεν Σάσον (Steve Sasson), ο οποίος ήταν μηχανικός στην εταιρεία Kodak στην εποχή εκείνη όπου η ψηφιακή τεχνολογία ήταν ακόμα αν όχι στα αρχικά, σίγουρα στα πρώτα βήματά της. Ο Σάσον πλέον έδωσε τις βάσεις για να δημιουργηθεί μία νέα βιομηχανία, όπου και αναπτύχθηκε με γρήγορους αλλά και σημαντικούς ρυθμούς μέσα στις επόμενες δεκαετίες, αλλάζοντας έτσι τελείως τις τεχνικές αλλά και γενικότερα τη τέχνη της φωτογραφίας. Η ψηφιακή αυτή φωτογραφική μηχανή, ζύγιζε περίπου 3,6 κιλά και είχε το μέγεθος μίας μικρής συσκευής κουζίνας, όπου μπορούσε να αποτυπώσει μόνο ασπρόμαυρες εικόνες με ανάλυση 0,01 megapixel (100 x 100 pixels) και στη συνέχεια να τις αποθηκεύσει σε κασέτα. Για τη καταγραφή και τη μεταφορά της εικόνας σε υπολογιστή, η συσκευή αυτή χρειαζόταν περίπου 23 δευτερόλεπτα. Παρόλο που η συσκευή αυτή του Σάσον δεν ήταν κατάλληλη για εμπορική χρήση, μπορεί να θεωρηθεί πως ήταν η βάση για την αρχή της ψηφιακής εικόνας, προετοιμάζοντας μας για τις δυνατότητες της τεχνολογίας με τη πάροδο των χρόνων.

[\[3\]](#)[\[4\]](#)[\[5\]](#)

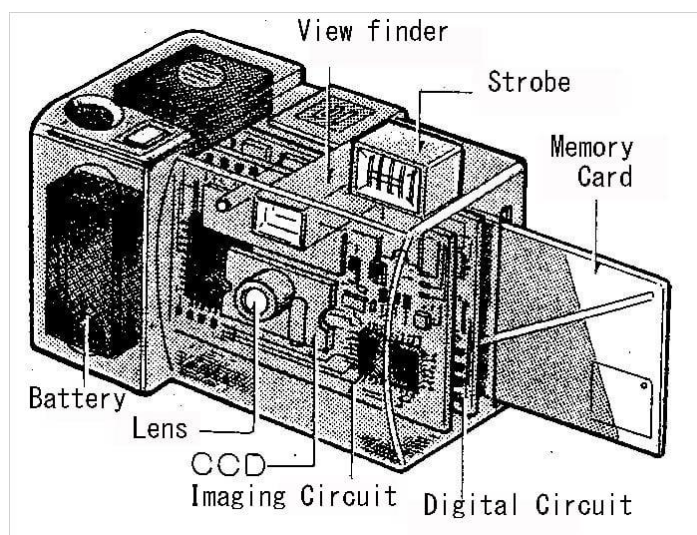


Εικόνα 13 Steven Sasson

3.3.1 Οι πρώτες εμπορικές κυκλοφορίες και η παροδική εξέλιξή τους

Παρόλο που ο Στίβεν Σάσον είχε κάνει μία εντυπωσιακή για την εποχή εφεύρεση και ενώ οι αναλογικές φωτογραφικές μηχανές συνέχιζαν να είναι αρκετά διαδεδομένες εκείνη την εποχή, η εταιρεία Kodak δεν αξιοποίησε άμεσα την τεχνολογία αυτή του Σάσον. Κατά τη δεκαετία του 1980, αρκετές από τις εταιρείες ξεκίνησαν να πειραματίζονται με βάση τη ψηφιακή τεχνολογία. Ωστόσο, το 1981 πιο συγκεκριμένα, η εταιρεία Sony παρουσίασε τη Manica, μία ηλεκτρονική φωτογραφική μηχανή η οποία αποθήκευε τις φωτογραφίες της σε μαγνητικό δίσκο. Βέβαια, η Manica δεν ήταν πλήρως ψηφιακή κάμερα καθώς οι φωτογραφίες της δεν αποθηκεύονταν ακόμη σε ψηφιακή, αλλά σε αναλογική μορφή, παρόλα αυτά έθεσε τα θεμέλια για την αρχή της ηλεκτρονικής αποθήκευση στην καταγραφή φωτογραφιών και γενικότερα εικόνων.

Μερικά χρόνια μετά και πιο συγκεκριμένα το 1988, η εταιρεία Fujifilm παρουσίασε τη πρώτη όπως θεωρείται πλήρως ψηφιακή φωτογραφική μηχανή, την Fujix DS-1P, η οποία αποθήκευε τις φωτογραφίες σε flash μνήμη. Η συγκεκριμένη κάμερα ήταν ένα βήμα πιο μπροστά για την εποχή εκείνη, καθώς οι φωτογραφίες μπορούσαν να αποθηκεύονται αλλά και να μεταφέρονται χωρίς αναλογικά μέσα. Παρόλο που η καινοτομία της κάμερας αυτής ήταν αρκετά μπροστά για την εποχή της, η Fujix DS-1P δεν κυκλοφόρησε αρκετά, λόγω του υψηλού της κόστους, αλλά αποτέλεσε μία σημαντική εξέλιξη για την ανάπτυξη ψηφιακών καμερών στο μέλλον. [\[3\]](#)[\[4\]](#)[\[5\]](#)



Εικόνα 14 Fujix DS-1P

3.3.2 Ραγδαία εμπορική ανάπτυξη

Κατά τα τέλη της δεκαετίας του 1990 και ως προς τις αρχές της δεκαετίας του 2000, η φωτογραφική βιομηχανία έκανε ραγδαία ανάπτυξη επάνω στον κλάδο των ψηφιακών φωτογραφικών μηχανών. Καθώς η τεχνολογία άρχισε να μπαίνει για τα καλά στη ζωή των ανθρώπων έτσι και οι τιμές στις μηχανές αυτές άρχισαν να μειώνονται και να είναι περισσότερο προσιτές οικονομικά. Με τη σειρά τους μερικές από τις πιο γνωστές εταιρείες του χώρου, όπως η Nikon, η Sony, η Canon και η Olympus ξεκίνησαν να εισάγουν στην αγορά νέες σειρές και μοντέλα ψηφιακών φωτογραφικών μηχανών με ακόμα πιο υψηλή ανάλυση, μεγαλύτερη χωρητικότητα αποθήκευσης και πιο εύκολη χρήση ως προς τον χρήστη.

Η αυξανόμενη δημοτικότητα των φωτογραφικών αυτών μηχανών συνοδεύτηκε από σημαντικές βελτιώσεις στην ποιότητα των εικόνων και στην ευκολία χρήσης των συσκευών. Έτσι, καθώς οι ψηφιακές φωτογραφικές μηχανές έγιναν πιο δημοφιλείς για την εποχή εκείνη, η τεχνολογία τους εξελίχθηκε προσφέροντας καλύτερη ποιότητα εικόνας και πιο άμεση και απλή χρήση. Η ανάπτυξη του διαδικτύου και η ευρεία χρήση των ηλεκτρονικών υπολογιστών έπαιξαν σημαντικό ρόλο για την εμπορική επιτυχία της ψηφιακής φωτογραφίας. Αυτά τα τεχνολογικά επιτεύγματα έπαιξαν καθοριστικό ρόλο στην ώθηση της ψηφιακής φωτογραφίας στην αγορά. Στη συνέχεια και με βάση της χρήσης του διαδικτύου, οι χρήστες είχαν τη δυνατότητα να αποθηκεύσουν, να επεξεργαστούν αλλά και να μοιραστούν τις ψηφιακές τους φωτογραφίες διαδικτυακά και αυτό εμφάνισε έναν νέο τρόπο όπου οι χρήστες χρησιμοποιούσαν τις φωτογραφίες τους. Όπως επίσης, πολλές από τις πιο γνωστές έως και σήμερα εφαρμογές, όπως το Photoshop και Lightroom της εταιρείας Adobe αλλά και πλατφόρμες όπως το Flickr και στη συνέχεια το Facebook και το Instagram, δημιούργησαν νέους δρόμους στη χρήση των ψηφιακών φωτογραφιών και εικόνων. [\[3\]](#)[\[4\]](#)[\[5\]](#)



Εικόνα 15 Early digital compact cameras

3.3.3 Κινητές συσκευές με φωτογραφικές δυνατότητες

Καθώς η τεχνολογία άρχισε να ακμάζει κάθε χρόνο όλο και περισσότερο, παρουσιάστηκε η εισαγωγή των καμερών στα κινητά τηλέφωνα κατά τις αρχές προς τα μέσα της δεκαετίας του 2000. Μάλιστα, το πρώτο κινητό τηλέφωνο με ενσωματωμένη κάμερα ήταν το J-Phone της εταιρείας Sharp, όπου κυκλοφόρησε στην Ιαπωνία το 2000 και είχε τη δυνατότητα να βγάζει φωτογραφίες με ανάλυση 0,11 megapixel. Λίγα χρόνια μετά και πιο συγκεκριμένα το 2007 η εταιρεία Apple παρουσίασε το πρώτο της smartphone, το iPhone το οποίο με τη σειρά του έφερε ραγδαία αλλαγή στη χρήση της φωτογραφίας. Παρόλα αυτά, η ποιότητα των καμερών στα κινητά τηλέφωνα άρχισε να αναπτύσσεται σημαντικά με τη πάροδο των χρόνων με αποτέλεσμα να αντικαταστήσουν τις ψηφιακές φωτογραφικές μηχανές, ειδικότερα για τις καθημερινές φωτογραφίες. Οι χρήστες πλέον με τον τρόπο και τις συσκευές αυτές έχουν τη δυνατότητα να βγάζουν και να μοιράζονται τις φωτογραφίες μέσα σε μόλις λίγα λεπτά, δημιουργώντας έτσι ένα νέο είδος φωτογραφίας, τη κινητή φωτογραφία. Με την εξέλιξη αυτή η βιομηχανία των ψηφιακών φωτογραφικών μηχανών επηρεάστηκε αρνητικά σε μεγάλο βαθμό, οδηγώντας έτσι σε πτώση των πωλήσεων των παραδοσιακών ψηφιακών φωτογραφικών μηχανών, καθώς πλέον ένα αρκετά μεγάλο ποσοστό των χρηστών, χρησιμοποιούν τη κινητή τους συσκευή για τις φωτογραφίες τους. [\[3\]\[4\]\[5\]](#)



Εικόνα 16 J-Phone

3.3.4 Σύγχρονη τεχνολογία και οι τελευταίες εξελίξεις

Σήμερα, οι ψηφιακές φωτογραφικές μηχανές συνεχίζουν να εξελίσσονται σημαντικά όπως επίσης να παρουσιάζουν αρκετά καλά χαρακτηριστικά, όπως η υψηλή ανάλυσης, η προηγμένη ανάλυση χρωμάτων και το δυναμικό εύρος, όπως και έχουν τη δυνατότητα για λήψη HDR (High Dynamic Range) η οποία είναι μία τεχνολογία που χρησιμοποιείται για την

αναπαραγωγή εικόνων αλλά και βίντεο με μεγαλύτερη αντίθεση, πιο λεπτομερείς στις σκοτεινές περιοχές αλλά και πιο ζωντανά χρώματα. Επίσης, έχουμε και τη τεχνητή νοημοσύνη (Artificial Intelligence - AI) η οποία είναι ένας τομέας της επιστήμης των ηλεκτρονικών υπολογιστών, όπου εστιάζει στη δημιουργία συστημάτων και φωτογραφικών μηχανών ικανών να εκτελούν εργασίες. Έτσι, οι επαγγελματικές αυτές μηχανές, όπως για παράδειγμα οι DSLR αλλά και οι mirrorless, παρέχουν υψηλές δυνατότητες που ικανοποιούν μέχρι και τις ανάγκες φωτογράφων αλλά και κινηματογραφιστών.

Οι mirrorless φωτογραφικές μηχανές ή αλλιώς Mirrorless Interchangeable Lens Cameras (MILC), είναι μία νέα κατηγορία ψηφιακών φωτογραφικών μηχανών που δεν χρησιμοποιούν το παραδοσιακό σύστημα καθρέπτη όπως στις DSLR μηχανές και έχουν γίνει ιδιαίτερα δημοφιλείς, για το λόγο ότι προσφέρουν την ποιότητα αλλά και την ευελιξία των DSLR φωτογραφικών μηχανών σε μικρότερο μέγεθος. Επιπλέον, η τεχνολογία LiDAR (Light Detection And Ranging και οι πολλαπλοί φακοί στις κάμερες των κινητών τηλεφώνων, βελτιώνουν τη ποιότητα αλλά και την ανάλυση σε βάθος της κινητής φωτογραφίας ώστε οι φωτογραφίες αυτές να είναι πιο καθαρές. Καθώς η τεχνολογία συνεχίζει να εξελίσσεται με γρήγορους ρυθμούς, οι ψηφιακές φωτογραφικές μηχανές και κάμερες βρίσκονται στην αιχμή της τεχνολογικής προόδου, διαδραματίζοντας έτσι σημαντικό αλλά και βασικό ρόλο τόσο στη τέχνη όσο και στη καθημερινότητα των ανθρώπων. [\[3\]](#)[\[4\]](#)[\[5\]](#)



Εικόνα 17 Nikon Z6 II

3.4 Επεξεργασία εικόνας

Η ψηφιακή επεξεργασία εικόνας είναι η διαδικασία με την οποία μπορεί να βελτιωθεί ή να τροποποιηθεί γενικότερα μία ψηφιακή εικόνα, χρησιμοποιώντας υπολογιστικά εργαλεία και λογισμικά. Είναι μία ισχυρή τεχνολογία όπου έχει αλλάξει για τα καλά τον τρόπο με τον οποίο μπορούν να αναλυθούν οι εικόνες και φωτογραφίες σε διάφορους τομείς της

καθημερινότητας του ανθρώπου. Η τεχνολογία αυτή κάνει την εμφάνισή της σε ένα μεγάλο πλήθος από εφαρμογές οι οποίες μπορούν να χρησιμοποιηθούν είτε για ψυχαγωγία, είτε στον τομέα της φωτογραφίας και της ιατρική για την παρουσίαση διαγνώσεων μέχρι και στον τομέα της ασφάλειας. Οι βασικές αρχές της ψηφιακής επεξεργασίας εικόνας βασίζονται κυρίως με τη χρήση αλγορίθμων. Όπως γνωρίζουμε οι εικόνες αποτελούνται από πάρα πολύ μικρά στοιχεία, τα pixels, στα οποία κάθε pixel περιέχει μία πληροφορία με βάση το χρώμα αλλά και την ένταση του φωτός. Αυτό που υλοποιεί λοιπόν η ψηφιακή επεξεργασία εικόνας είναι με τη βοήθεια εφαρμογών, να τροποποιήσει τις τιμές αυτές των pixels για να αναπαράγει την επιθυμητή αλλαγή, είτε ως προς την ένταση του φωτός, είτε ως προς τα χρώματα, είτε ακόμα και ως προς των συνδυασμό και των δύο. [\[5\]\[6\]\[9\]](#)

3.4.1 Εφαρμογές της ψηφιακής επεξεργασίας εικόνας

Όπως είναι εμφανές και με τη σταδιακή ανάπτυξη της τεχνολογίας και των εφαρμογών μέρα με τη μέρα, έχουν δημιουργηθεί αρκετές εφαρμογές για την επεξεργασία ψηφιακών εικόνων σε διάφορους τομείς της καθημερινότητας του ανθρώπου όπως στον τομέα της φωτογραφίας και των γραφιστικών, όπου η ψηφιακή επεξεργασία χρησιμοποιείται για τη βελτίωση της ποιότητας μίας εικόνας ή και της διόρθωσης των χρωμάτων της, ακόμα και για τη δημιουργία εφέ ή και για την αφαίρεση ανεπιθύμητων στοιχείων ή μερών μιας εικόνας. Οι πιο δημοφιλείς εφαρμογές για την επεξεργασία εικόνων είναι το Photoshop και Lightroom από την εταιρεία Adobe και το GIMP από την εταιρεία Softonic, οι οποίες επιτρέπουν στους χρήστες να έχουν τη δυνατότητα να επεξεργαστούν σε πολύ μεγάλο βαθμό μία εικόνα, είτε να εφαρμόσουν φίλτρα, είτε να κόψουν και να αναπροσαρμόσουν μία εικόνα. Επίσης, έχουν δημιουργηθεί και εφαρμογές στις οποίες γίνεται χρήση της επεξεργασίας εικόνων με βάση τον τομέα της κινηματογραφίας και της ψυχαγωγίας, όπου στη παραγωγή ταινιών, χρησιμοποιούνται τεχνικές για τη προσθήκη ειδικών εφέ ή και για την επεξεργασία σκηνών και σκηνικών, ακόμα και για 3D χαρακτήρων. Τα παραπάνω υλοποιούνται με τα CGI (Computer – Generated Imagery), όπου είναι ένα δημοφιλή παράδειγμα ψηφιακής επεξεργασίας και που έχει βοηθήσει σημαντικά σε όλα τα μονοπάτια της βιομηχανίας του κινηματογράφου. Εκτός απ' τη φωτογραφία και τη ψυχαγωγία, τη ψηφιακή επεξεργασία εικόνων τη συναντάμε και κατά την ιατρική απεικόνιση, στην οποία χρησιμοποιείται επεξεργασία για τη βελτίωση των ακτινογραφιών, των υπερηχογραφήμάτων και των μαγνητικών τομογράφων (MRI), δίνοντας τη δυνατότητα στους γιατρούς να έχουν πιο καθαρές και ακριβείς εικόνες των διαγνώσεων. Εκτός όλων αυτών, η ψηφιακή επεξεργασία

κάνει αισθητή τη παρουσία της τόσο στην ασφάλεια όσο και στην επιτήρηση, όπου βοηθάει στην βελτίωση της ποιότητας της εικόνας αλλά και των βίντεο για την ταυτοποίηση ατόμων σε διάφορα εγκλήματα, αναγνωρίζοντας πρόσωπα μέσα από κάμερες ασφαλείας. Τέλος, για την ανάλυση αλλά και τη παρακολούθηση περιβαλλοντικών φαινομένων έχουμε τις δορυφορικές εικόνες, οι οποίες χρησιμοποιούνται για τη μελέτη της γης αλλά και του περιβάλλοντος, ώστε να μπορούν να αποκαλύψουν αλλαγές όπως την αποψίλωση των δασών και την εξάπλωση των πόλεων. [5][6][9]



Εικόνα 18 Adobe Products

3.4.2 Τεχνικές ψηφιακής επεξεργασίας εικόνας

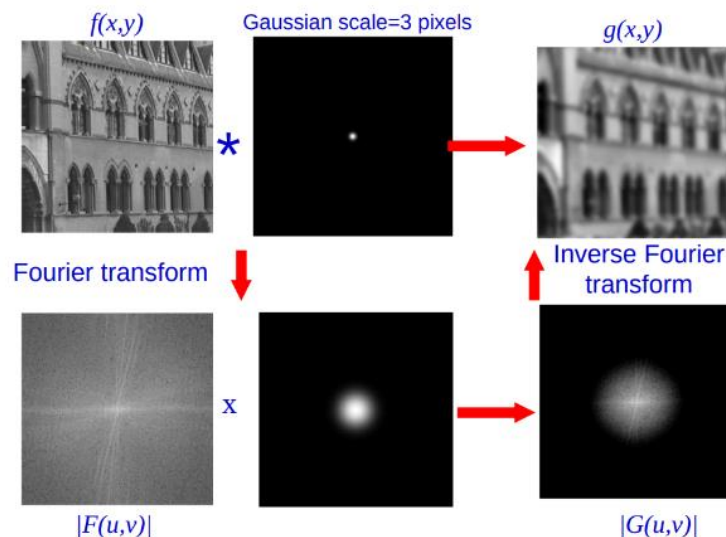
Η ψηφιακή επεξεργασία εικόνας έχει αρχίσει με τη πάροδο του χρόνου να αλλάζει τον τρόπο με τον οποίο βλέπουμε αλλά και αναλύουμε τις εικόνες καθώς η τεχνολογία ακμάζει και γίνεται όλο και ισχυρότερη και αναμένεται στο μέλλον να επεκταθεί ακόμα περισσότερο.

Για να εφαρμοστεί επεξεργασία σε μία εικόνα, θα χρειαστεί να συμπεριληφθούν κάποιες τεχνικές όπως το φιλτράρισμα, το οποίο χρησιμοποιείται για να βελτιώσει ή και να τροποποιήσει μία εικόνα, είτε ως προς τη μείωση του θορύβου, είτε ως προς την ενίσχυση λεπτομερειών και ανίχνευση ακμών όπως το Computer Vision, είτε ως προς τη συμπίεση εικόνων όπως οι JPEG χρησιμοποιούν φίλτρα DCT, είτε ακόμα και για ιατρικές απεικονίσεις όπως είναι το MRI. Η τεχνική τους φιλτραρίσματος γίνεται πράξη μέσα από μία συνέλιξη με έναν πυρήνα (kernel), όπου στη πραγματικότητα αυτό που συμβαίνει είναι να εφαρμόζεται ένας πίνακας πάνω σε μία εικόνα και μέσω μαθηματικών πράξεων να δημιουργείται το φίλτρο.

Φυσικά, υπάρχουν αρκετοί τύποι φιλτραρίσματος όπως το Χωρικό Φιλτράρισμα (Spatial Filtering), το οποίο διαχωρίζεται στα Χαμηλοπερατά Φίλτρα (Low-pass Filters), όπου βελτιώνουν την εικόνα ως προς τη μείωση του θορύβου και τις υψηλές συχνότητες και στα Υψιπερατά Φίλτρα (High-pass Filters), όπου βελτιώνουν την εικόνα ενισχύοντας τις λεπτομέρειες και τις ακμές της. Ένας ακόμα τύπος φιλτραρίσματος είναι το Φιλτράρισμα στο πεδίο συχνοτήτων (Frequency Filtering), στο οποίο με βάση το Μετασχηματισμό Fourier (Fourier Transform), η εικόνα παίρνει τη μορφή συχνότητας και στη συνέχεια εφαρμόζονται αντίστοιχα τα Χαμηλοπερατά, τα Υψιπερατά και τα Ζωνοπερατά Φίλτρα (Band-pass Filters). Ωστόσο, μία ακόμα τεχνική επεξεργασίας, στην οποία εφαρμόζεται φιλτράρισμα μόνο σε συγκεκριμένες περιοχές μιας εικόνας, χωρίς να επηρεαστεί ολόκληρη την εικόνα, γίνεται με το Επιλεκτικό Φιλτράρισμα (Selective Filtering).

Υλοποιώντας τη διαδικασία αυτή, μπορεί να επιτευχθεί βελτίωση αλλά και επεξεργασία συγκεκριμένων χαρακτηριστικών και στοιχείων, όπως οι ακμές, οι λεπτομέρειες αλλά και οι περιοχές με υψηλό θόρυβο. Επίσης, ένας ακόμα τύπος φιλτραρίσματος, ο οποίος εφαρμόζεται κυρίως στις δυαδικές εικόνες ώστε να αναλυθούν τα περιγράμματα και τα σχήματα μίας εικόνας γίνεται με βάση τα Μορφολογικά Φίλτρα (Morphological Filtering) και διαχωρίζονται σε τρεις λειτουργίες. Οι λειτουργίες αυτές είναι η Διαστολή (Dilation), στην οποία επεκτείνονται τα λευκά σημεία μίας εικόνας, η Συστολή (Erosion) όπου γίνεται αντίστοιχα το ανάστροφο της διαστολής, με αποτέλεσμα να μειώνονται τα λευκά σημεία και τέλος το Άνοιγμα (Opening) αλλά και Κλείσιμο (Closing), τα οποία χρησιμοποιούνται για την μείωση αλλά και απομάκρυνση θορύβου. Τέλος, μία ακόμα σημαντική και γνωστή κατηγορία φίλτρων και φιλτραρίσματος είναι τα Προσαρμοστικά Φίλτρα (Adaptive Filtering), όπου ανάλογα τα χαρακτηριστικά της περιοχής της εικόνας προσαρμόζουν κατάλληλα τη συμπεριφορά τους. Φυσικά τα φίλτρα αυτά διαχωρίζονται σε διάφορους τύπους, πιο συγκεκριμένα στο Προσαρμοστικό Φίλτρο Μέσης Τιμής (Adaptive Mean Filter) στο οποίο προσαρμόζεται το μέγεθος του παραθύρου (Kernel) με βάση το επίπεδο θορύβου που βρίσκεται σε μία περιοχή της εικόνας και χρησιμοποιείται κυρίως για την μείωση του Γκαουσιανού Θορύβου (Gaussian Noise), όπου ο θόρυβος προστίθεται τυχαία στα pixel της εικόνας, είτε συγκεντρώνεται κανονικά γύρω από τη μέση τιμή και όσο απομακρύνουμε την εικόνα τόσο μειώνεται και αυτός, είτε προστίθεται επάνω στο αρχικό pixel με αποτέλεσμα να δημιουργείται ένα νέο pixel ή ακόμα, κάνει μία ομοιόμορφη επίδραση στην οποία επηρεάζεται ολόκληρη η εικόνα και όχι μέρη της. Αυτό οφείλεται συνήθως σε εικόνες με χαμηλό φωτισμό και παρουσιάζεται αρκετές φορές σε αισθητήρες κάμερας λόγω διάφορων θερμικών διακυμάνσεων ή σε μετάδοση σήματος ή σε επεξεργασία εικόνας μορφής JPEG.

Όπως και για την αποθορυβοποίηση και βελτίωση θολών εικόνων από θόρυβο, χρησιμοποιείται το Προσαρμοστικό Φίλτρο Wiener (Wiener Filter), το οποίο υπολογίζει την καλύτερη εκτίμηση σε ένα pixel με τη διακύμανση της εικόνας στο σημείο που βρίσκεται το pixel αυτό. Αντίθετα, για την ανίχνευση ακμών χωρίς ενίσχυση θορύβου αλλά και για την αύξηση της ευκρίνειας (Sharpness), λιγότερο σε ομαλές περιοχές και περισσότερο σε περιοχές με έντονες αλλαγές χρησιμοποιούνται τα Προσαρμοστικά Υψηλερατά Φίλτρα (Adaptive High pass Filters). Τέλος, για την αφαίρεση ανεπιθύμητου θορύβου όπου εμφανίζεται ως επαναλαμβανόμενο μοτίβο σε μία εικόνα και μπορεί να φιλτράρει μόνο τις συχνότητες που σχετίζονται με θόρυβο χρησιμοποιείται το Προσαρμοστικό Φιλτράρισμα Συχνότητας (Adaptive Frequency Filtering). [5][6][7][9]



Εικόνα 19 Fourier Filter

3.4.3 Ανακατασκευή και αποκατάσταση εικόνας

Πέρα απ' τη διαδικασία φιλτραρίσματος για τη βελτίωση και αποθορυβοποίηση της εικόνας, υπάρχουν και οι διαδικασίες της ανακατασκευής αλλά και αποκατάστασης μίας εικόνας, οι οποίες κάνουν χρήση τόσο για τη βελτίωση όσο και για την αναδημιουργία εικόνων που έχουν χαθεί είτε καταστραφεί είτε και υποβαθμιστεί.

Για την αποκατάσταση αλλά και βελτίωση της εικόνας από θόρυβο, παραμορφώσεις, ακόμα και για την επαναφορά λεπτομερειών μπορούν να χρησιμοποιηθούν κάποιες τεχνικές διαδικασίες όπως στο Φιλτράρισμα Θορύβου (Noise Reduction), στο οποίο μπορεί να χρησιμοποιηθεί η τεχνική του Μέσου Φίλτρου (Median Filter) για την αφαίρεση του θορύβου χωρίς να μεταβληθούν οι ακμές ή το Bilateral Filter για να μειωθεί ο θόρυβος. Ακόμα μία

γνωστή διαδικασία είναι η Αποθόλωση (De-Blurring), καθώς με τη χρήση του Φίλτρου Wiener (Wiener Filter) μπορεί να μειώσει έως και να αφαιρέσει το θάμπωμα από μία εικόνα καθώς βασίζεται στη συχνότητα μιας εικόνας.

Επίσης, με τη χρήση του αλγορίθμου Richardson Lucy Algorithm, υλοποιούνται επαναληπτικοί υπολογισμοί ώστε να βελτιωθεί όσο το δυνατό περισσότερο η ευκρίνεια της εικόνας. Παρόλα αυτά, είναι εξίσου σημαντική για την αποκατάσταση μιας εικόνας η Ανάκτηση Χρωμάτων (Color Restoration) ώστε να βελτιθούν ξεθωριασμένες ή ασπρόμαυρες εικόνες. Αυτό μπορεί να επιτευχθεί με τη τεχνική του Deep Learning Colorization, που κάνει την εμφάνισή της με τη χρήση της τεχνητής νοημοσύνης, προσθέτοντας χρώματα και αποχρώσεις σε παλιές αλλά και ασπρόμαυρες εικόνες, όπως επίσης και με τη τεχνική του Histogram Equalization, που διορθώνει τη φωτεινότητα σε μέρη της εικόνας. Φυσικά, ένα ακόμα πολύ σημαντικό ρόλο για τη διόρθωση αλλά και βελτίωση εικόνων από χαμηλή ποιότητα λόγω συμπίεσης γίνεται με την Απομάκρυνση Συμπίεσης (Compression Artifacts Removal), στην οποία με τη χρήση του Deep CNN (Convolutional Neural Networks) η εικόνα αποκαθίσταται με πολύπλοκα μοτίβα και να κατασκευαστούν εκ νέου χαμένες αλλά και αλλοιωμένες λεπτομέρειες, όπως επίσης και η τεχνική του Μετασχηματισμού Κυμάτων (Wavelet Transform) που έχει τη δυνατότητα να επαναφέρει λεπτομέρειες που έχουν χαθεί λόγω της συμπίεσης. Τέλος, με τη διαδικασία της Βελτίωσης της Ανάλυσης (Super - Resolution), υπάρχει η δυνατότητα μια εικόνα να αυξήσει την ανάλυσή της και να βελτιώσει τις λεπτομέρειές της με βάση τη παραδοσιακή μέθοδο της αύξησης της ανάλυσης Bicubic Interpolation, όπως επίσης και με τη τεχνική του SRGAN (Super- Resolution GANs) όπου τα νευρωνικά δίκτυα δημιουργούν εικόνες από χαμηλής ποιότητας δεδομένα με υψηλές αναλύσεις.

Αντίστοιχα η ανακατασκευή αφορά κυρίως εικόνες που έχουν χαθεί, είτε καταστραφεί, είτε έχει υποβαθμιστεί ένα μέρος τους. Η διαδικασία αυτή μπορεί να κάνει χρήση τόσο σε παλιές φωτογραφίες και δορυφορικές εικόνες όσο και σε ιατρικές απεικονίσεις, με αποτέλεσμα να βελτιωθεί όσο το δυνατό περισσότερο η χαμηλή ποιότητα της ανάλυσής τους. Όπως στη διαδικασία της αποκατάστασης, έτσι και στην ανακατασκευή έχουμε μερικές διαδικασίες για να επιτευχθεί η διαδικασία αυτή, όπως είναι η Παρεμβολή (Interpolation Methods), όπου χρησιμοποιείται για την αύξηση της ανάλυσης εικόνας αλλά και για την ανακατασκευή ελλειπόντων pixels, με τη χρήση τεχνικών όπως είναι το Nearest Neighbor Interpolation όπου δημιουργεί ένα μπλοκ από pixels ανάλογα με το πλαίσιο της εικόνας. Παρόλα αυτά σαν μία ακόμα τεχνική είναι και το Bicubic Interpolation, ή οποία μπορεί να προκαλέσει σε υπερβολικό βαθμό επεξεργασία σε μία εικόνα αλλά έχει καλύτερη ομαλότητα απ' ό τι έχει το

Bilinear Interpolation, όπου περιορίζει σε έναν βαθμό τις μεταβολές χρωμάτων της εικόνας αλλά υπάρχει πιθανότητα να προκαλέσει θόλωση.

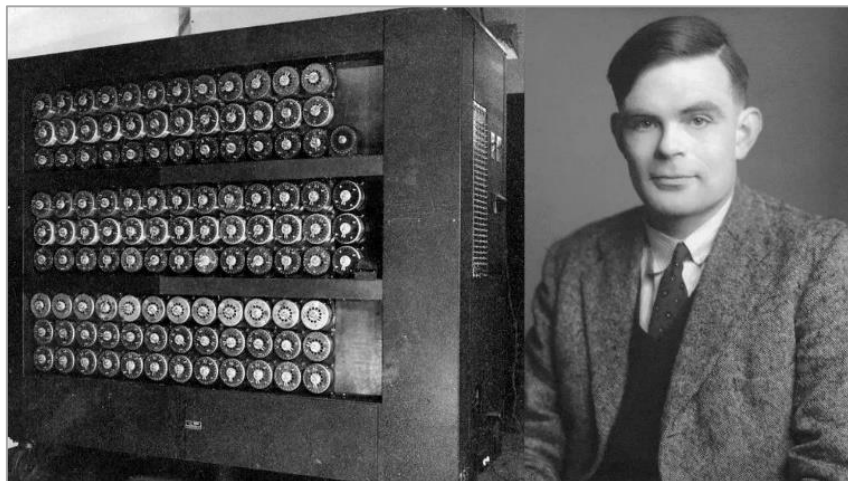
Μία ακόμα σημαντική διαδικασία της ανακατασκευής είναι η Συμπλήρωση Κενών Περιοχών (In-painting), όπου χρησιμεύει στην ανακατασκευή περιοχών των εικόνων που έχουν φθαρεί, είτε έχουν αφαιρεθεί αντικείμενα, είτε είναι φωτογραφίες που έχουν σκιστεί. Βέβαια η συμπλήρωση και ανακατασκευή αυτή γίνεται με τη βοήθεια έξυπνων αλγορίθμων και τεχνικών όπως το Exemplar Based In-painting, το οποίο αναγνωρίζει και αντιγράφει παρόμοιες περιοχές από pixels της εικόνας και έτσι συμπληρώνει τις περιοχές. Ακόμα, με τη χρήση του αλγορίθμου Patch Match Algorithm ελέγχονται και αντικαθίστανται τα χαμένα pixels της εικόνας με τις καλύτερες αντιστοιχίες από τα σημεία της, όπως και το GAN-based In-painting (Generative Adversarial Networks) χρησιμοποιείται για την ρεαλιστική ανακατασκευή περιοχών της εικόνας. Τέλος, μία ακόμα σημαντική διαδικασία, η οποία βελτιώνει τις λεπτομέρειες μιας εικόνας με χαμηλή ανάλυσης και παρασκευάζει μια εκδοχή της με υψηλότερη και βελτιωμένη ανάλυσή της με τη χρήση του AI, η διαδικασία αυτή είναι η Αύξηση Ανάλυσης (Super-Resolution). [\[6\]\[7\]\[8\]\[9\]](#)

ΚΕΦΑΛΑΙΟ 4

4. ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ

4.1 Η αρχή της τεχνητής νοημοσύνης

Η έννοια της τεχνητής νοημοσύνης (AI) ξεκινά από την ανθρώπινη επιθυμία να κατασκευαστούν μηχανές που μπορούν να σκέφτονται και να δρουν όπως ο άνθρωπος. Από την αρχαιότητα βλέπουμε σημαντικές ενδείξεις πως οι άνθρωποι εξέφραζαν την επιθυμία της επικοινωνίας με τις μηχανές. Όμως, η πραγματική έρευνα για την τεχνητή νοημοσύνη ξεκίνησε μόλις τον 20ό αιώνα, όταν οι επιστήμονες είχαν πλέον την απαραίτητη θεωρητική βάση και τεχνολογικές δυνατότητες. Τα πρώτα βήματα ξεκίνησαν με τον Άλαν Τούρινγκ (Alan Turing), έναν από τους θεμελιωτές της υπολογιστικής επιστήμης. Το 1950, στο διάσημο άρθρο του "Computing Machinery and Intelligence", ο Τούρινγκ πρότεινε το ερώτημα "Μπορούν οι μηχανές να σκέφτονται;", και στη συνέχεια περιέγραψε το Turing Test, έναν τρόπο να αξιολογηθεί αν μια μηχανή μπορεί να εμφανίσει νοημοσύνη παρόμοια με εκείνη των ανθρώπων. Το τεστ παραμένει ένα από τα βασικά σημεία αναφοράς στη μελέτη της AI.



Εικόνα 20 Alan Turing and the Turing machine

Το 1956, το συνέδριο του Dartmouth στις ΗΠΑ θεωρείται η επίσημη αρχή της έρευνας για την τεχνητή νοημοσύνη ως επιστημονικού πεδίου. Κατά τη διάρκεια αυτού του συνεδρίου, ο Τζον Μακάρθι (John McCarthy) εισήγαγε τον όρο "τεχνητή νοημοσύνη". Οι επιστήμονες που

συμμετείχαν πίστευαν ότι σε λίγες δεκαετίες θα μπορούσαν να κατασκευάσουν μηχανές που θα μιμούνταν τον ανθρώπινο εγκέφαλο και τη νοημοσύνη. Κατά τις δεκαετίες του 1960 και 1970, η έρευνα επικεντρώθηκε στην ανάπτυξη λογισμικών που θα μπορούσαν να επιλύουν προβλήματα και να παίζουν παιχνίδια στρατηγικής, όπως το σκάκι. Τα πρώτα προγράμματα AI, όπως το Logic Theorist (1955) και το General Problem Solver (1957), είχαν τη δυνατότητα να επιλύουν απλά μαθηματικά και λογικά προβλήματα. Ωστόσο, αυτές οι πρώτες απόπειρες σύντομα έδειξαν ότι η ανάπτυξη μιας πραγματικά "έξυπνης" μηχανής ήταν πολύ πιο δύσκολη από ό,τι αρχικά φανταζόταν.

Τα προγράμματα εκείνης της εποχής δεν μπορούσαν να αντιμετωπίσουν σύνθετες καταστάσεις και απαιτούσαν τεράστια υπολογιστική ισχύ για την εποχή, πράγμα το οποίο προφανώς και δεν ήταν διαθέσιμο. Τις δεκαετίες του 1970 και 1980, η έρευνα της τεχνητής νοημοσύνης AI πέρασε μια περίοδο που είναι γνωστή ως "χειμώνας της τεχνητής νοημοσύνης". Παρά τις αρχικές αισιόδοξες προβλέψεις, οι έρευνες δεν απέδωσαν τα αναμενόμενα αποτελέσματα. Τα υπάρχοντα υπολογιστικά συστήματα ήταν ανεπαρκή για την υποστήριξη των πολύπλοκων αλγορίθμων που απαιτούνταν, ενώ και οι ίδιες οι θεωρίες της νοημοσύνης και της μάθησης αποδείχτηκαν πιο περίπλοκες από ό,τι είχε προβλεφθεί. Οι χρηματοδοτήσεις μειώθηκαν και οι επιστήμονες στράφηκαν σε άλλους τομείς. Παρόλα αυτά, η έρευνα δεν σταμάτησε εντελώς. Το 1980, η ανάπτυξη των συστημάτων εμπειρογνομόνων έδωσε νέα πνοή στην AI. Τα συστήματα αυτά ήταν ειδικευμένα λογισμικά που μπορούσαν να λαμβάνουν αποφάσεις σε συγκεκριμένα πεδία, όπως η ιατρική διάγνωση ή η χρηματοοικονομική ανάλυση, βασιζόμενα σε μια βάση γνώσεων και σε προκαθορισμένους κανόνες.

Η πραγματική εξέλιξη της τεχνητής νοημοσύνης, ήρθε τη δεκαετία του 1990 με την άνοδο της μηχανικής μάθησης (machine learning). Αντί να προγραμματίζουν λεπτομερώς τους κανόνες για κάθε πιθανή κατάσταση, οι επιστήμονες ανακάλυψαν ότι θα μπορούσαν να αναπτύξουν συστήματα που μαθαίνουν από δεδομένα. Η ιδέα ήταν ότι μια μηχανή θα μπορούσε να βελτιώνεται αυτόματα, αναγνωρίζοντας μοτίβα σε μεγάλα σύνολα δεδομένων, από ειδικές βάσεις δεδομένων. Αρκετά αξιοσημείωτο, είναι και η νίκη του υπολογιστή της IBM Deep Blue απέναντι στον παγκόσμιο πρωταθλητή σκακιού Garry Kasparov, το 1997. Η νίκη αυτή έδειξε ότι οι μηχανές μπορούσαν πλέον να ανταγωνιστούν τον ανθρώπινο νου σε σύνθετα παιχνίδια στρατηγικής, μια ένδειξη των δυνατοτήτων της AI.

*Εικόνα 21 Kasparov vs Deep Blue*

Με την άνοδο του διαδικτύου και τη ραγδαία αύξηση της επεξεργαστικής ισχύος των υπολογιστών, η ΑΙ μπήκε σε μια νέα εποχή. Το deep learning, μια υποκατηγορία της μηχανικής μάθησης, βασίζεται σε τεχνητά νευρωνικά δίκτυα, εμπνευσμένα από τη λειτουργία του ανθρώπινου εγκεφάλου. Η βαθιά μάθηση επέτρεψε στην ΑΙ να διαχειριστεί και να αναλύσει τεράστια ποσά δεδομένων, κάτι που ήταν απαραίτητο για την ανάπτυξη συστημάτων όπως η αυτόματη αναγνώριση εικόνας, η φωνητική αναγνώριση, και η επεξεργασία φυσικής γλώσσας. Το 2012, το νευρωνικό δίκτυο AlexNet κέρδισε έναν διαγωνισμό αναγνώρισης εικόνων, αποδεικνύοντας την υπεροχή της βαθιάς μάθησης σε συγκεκριμένα προβλήματα. Από τότε, τα τεχνητά νευρωνικά δίκτυα έγιναν η βάση για πολλές εφαρμογές της ΑΙ, όπως οι αυτόνομες οδήγησεις, η αναγνώριση ομιλίας και τα chatbots. Τα τελευταία χρόνια, η ΑΙ έχει εισέλθει σε κάθε πτυχή της καθημερινής μας ζωής.

Οι ψηφιακοί βοηθοί όπως η Siri και η Alexa, τα αυτόνομα οχήματα, και οι προσωπικές συστάσεις που παρέχονται από πλατφόρμες όπως το Netflix ή το Amazon, βασίζονται σε αλγορίθμους τεχνητής νοημοσύνης, όπου δίνουν στον χρήστη την δυνατότητα να χρησιμοποιήσει προϊόντα που το ίδιο το ΑΙ θεωρεί «προτεινόμενα», βασιζόμενο στις ήδη υπάρχον προβολές του χρήστη. Η ΑΙ χρησιμοποιείται επίσης εκτενώς στον τομέα της υγείας, της χρηματοδότησης, της εκπαίδευσης και της βιομηχανίας. Ένας σημαντικός τομέας έρευνας σήμερα είναι η ηθική της ΑΙ. Καθώς οι μηχανές γίνονται ολοένα και πιο έξυπνες, προκύπτουν ζητήματα σχετικά με τη χρήση τους, την προστασία της ιδιωτικότητας, και την ευθύνη για τις αποφάσεις που λαμβάνουν. Η τεχνητή νοημοσύνη έχει διανύσει έναν μακρύ δρόμο από τις πρώτες θεωρητικές προτάσεις του Άλαν Τούρινγκ. Από την αναπαραγωγή ανθρώπινων ικανοτήτων σε παιχνίδια μέχρι τις σύγχρονες εφαρμογές βαθιάς μάθησης, η ΑΙ έχει εξελιχθεί σε έναν από τους πιο σημαντικούς τεχνολογικούς τομείς της εποχής μας. Οι

προκλήσεις και οι δυνατότητες παραμένουν τεράστιες, με τη συνεχή πρόοδο να υπόσχεται ακόμα πιο καινοτόμες και πρωτοποριακές εφαρμογές στο μέλλον. [\[10\]\[11\]](#)

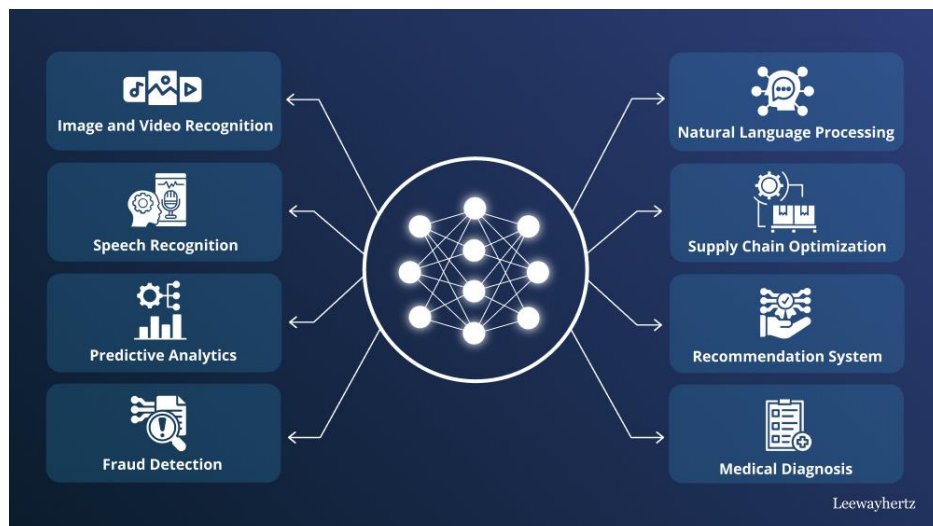
4.2 Επεξεργασία εικόνας με την βοήθεια της τεχνητής νοημοσύνης

Στην συγκεκριμένη ενότητα θα ασχοληθούμε κυρίως με την έρευνα της επεξεργασίας εικόνων με την βοήθεια της τεχνητής νοημοσύνης. Η επεξεργασία εικόνας με τη χρήση τεχνητής νοημοσύνης (AI) εξέλιξε τον τομέα της ψηφιακής επεξεργασίας και ανάλυσης εικόνων, παρέχοντας εργαλεία που υπερβαίνουν κατά πολύ τις δυνατότητες των παραδοσιακών τεχνικών. Από τη βελτίωση της ποιότητας εικόνας μέχρι την αυτοματοποιημένη ανάλυση δεδομένων, οι αλγόριθμοι μηχανικής μάθησης και βαθιάς μάθησης (deep learning) έχουν καταστήσει εφικτή την επεξεργασία εικόνων με ακρίβεια, ταχύτητα και ευελιξία που δεν ήταν δυνατές στο παρελθόν.

Η επεξεργασία εικόνας με AI περιλαμβάνει τη χρήση αλγορίθμων και μοντέλων βαθιάς μάθησης για την ανάλυση, επεξεργασία και βελτίωση ψηφιακών εικόνων. Αυτοί οι αλγόριθμοι είναι εκπαιδευμένοι σε τεράστιες ποσότητες δεδομένων, ώστε να μπορούν να κατανοούν και να αναλύουν μοτίβα στις εικόνες, όπως η αναγνώριση αντικειμένων, η ανάλυση χρωμάτων, και η πρόβλεψη κίνησης. Χρησιμοποιούνται επίσης για την αποκατάσταση εικόνων, την αναβάθμιση της ανάλυσης και την εφαρμογή δημιουργικών φίλτρων που βελτιώνουν την οπτική ποιότητα. Μία από τις πιο σημαντικές εφαρμογές της AI στην επεξεργασία εικόνας είναι η αναγνώριση αντικειμένων. Οι αλγόριθμοι βαθιάς μάθησης, όπως τα συγκλίνοντα νευρωνικά δίκτυα (CNNs), χρησιμοποιούνται για την αναγνώριση αντικειμένων σε εικόνες, με εκπληκτική ακρίβεια. Οι εφαρμογές αυτές έχουν επεκταθεί σε τομείς όπως η αυτόνομη οδήγηση, η ανάλυση ιατρικών εικόνων, και η παρακολούθηση σε πραγματικό χρόνο σε βιομηχανικά περιβάλλοντα. Οι αλγόριθμοι AI χρησιμοποιούνται επίσης για την αποκατάσταση φθαρμένων ή χαμηλής ανάλυσης εικόνων, όπως ιστορικές φωτογραφίες ή εικόνες που έχουν υποστεί θόρυβο. Η τεχνική αυτή επιτρέπει την αποκατάσταση των λεπτομερειών σε εικόνες χαμηλής ανάλυσης ή φθοράς, βελτιώνοντας την ποιότητα και την ευκρίνεια τους.

Ένα από τα εργαλεία που χρησιμοποιούνται για τέτοιες εφαρμογές είναι το GAN (Generative Adversarial Networks), το οποίο είναι ικανό να αναπαράγει την αρχική ποιότητα της εικόνας χρησιμοποιώντας μηχανική μάθηση. Η τεχνική super resolution επιτρέπει την αναβάθμιση της ανάλυσης των εικόνων χωρίς την απώλεια ποιότητας. Χρησιμοποιώντας αλγόριθμους

βαθιάς μάθησης, όπως το ESRGAN (Enhanced Super-Resolution Generative Adversarial Network), οι εικόνες μπορούν να γίνουν πιο καθαρές και πιο λεπτομερείς, ακόμα και αν προέρχονται από χαμηλής ανάλυσης δεδομένα. Αυτή η τεχνική είναι εξαιρετικά σημαντική για εφαρμογές όπως η παρακολούθηση από δορυφόρους και η ανάλυση ιατρικών δεδομένων, όπου η λεπτομέρεια είναι κρίσιμη. Οι εικόνες που περιέχουν θόρυβο, όπως αυτές που έχουν ληφθεί σε χαμηλό φωτισμό, μπορούν να καθαριστούν και να βελτιωθούν μέσω αλγορίθμων βαθιάς μάθησης. Η αφαίρεση θορύβου επιτυγχάνεται με τη χρήση νευρωνικών δικτύων που έχουν εκπαιδευτεί να αναγνωρίζουν τον θόρυβο και να τον αφαιρούν χωρίς να αλλοιώνουν τις λεπτομέρειες της εικόνας. Τα εργαλεία αυτά χρησιμοποιούνται ευρέως σε εφαρμογές όπως η ψηφιακή φωτογραφία και η επεξεργασία βίντεο. Η τεχνητή νοημοσύνη χρησιμοποιείται επίσης για τη δημιουργία και επεξεργασία εικόνων με καλλιτεχνικό και δημιουργικό τρόπο.



Εικόνα 22 Deep Learning

Τα μοντέλα όπως το DeepDream της Google και το Style GAN έχουν τη δυνατότητα να δημιουργούν εικόνες από το μηδέν ή να μετατρέπουν υπάρχουσες εικόνες σε εντελώς νέες, εφαρμόζοντας διαφορετικά στυλ και τεχνικές. Αυτές οι τεχνικές χρησιμοποιούνται ευρέως σε καλλιτεχνικές εφαρμογές και γραφικό σχεδιασμό. Ένας από τους τομείς όπου η επεξεργασία εικόνων με AI έχει κάνει τη μεγαλύτερη διαφορά είναι η ιατρική. Η AI χρησιμοποιείται για την ανάλυση ιατρικών εικόνων όπως ακτινογραφίες, αξονικές τομογραφίες, και μαγνητικές τομογραφίες, βοηθώντας τους γιατρούς να εντοπίσουν παθολογίες με μεγαλύτερη ακρίβεια και ταχύτητα. Επιπλέον, χρησιμοποιείται για τη διάγνωση σοβαρών ασθενειών, όπως ο καρκίνος, μέσω της ανάλυσης κυττάρων και ιστών.

Η επεξεργασία εικόνας με ΑΙ μπορεί να ανιχνεύσει προβλήματα που ενδεχομένως να μην είναι ορατά με το ανθρώπινο μάτι, βελτιώνοντας έτσι την πρώιμη διάγνωση και θεραπεία. Στη βιομηχανία, η ΑΙ βοηθά στην αυτοματοποίηση διαδικασιών όπως η παρακολούθηση ποιότητας προϊόντων και η ανάλυση δεδομένων από κάμερες παρακολούθησης. Οι αλγόριθμοι επεξεργασίας εικόνας μπορούν να αναγνωρίσουν ελαττώματα σε προϊόντα, να εντοπίσουν αντικείμενα σε εργοστασιακές γραμμές παραγωγής, και να βελτιστοποιήσουν την αποτελεσματικότητα των διαδικασιών. Οι τεχνικές αυτές μειώνουν το ανθρώπινο λάθος και βελτιώνουν την αποδοτικότητα. Η επεξεργασία εικόνας με τη χρήση ΑΙ έχει επίσης σημαντικό αντίκτυπο στην ψυχαγωγία και τα μέσα ενημέρωσης. Η δημιουργία οπτικών εφέ (VFX) για ταινίες και τηλεοπτικές σειρές έχει γίνει πιο γρήγορη και πιο αποδοτική με τη βοήθεια ΑΙ. Επιπλέον, τα εργαλεία επεξεργασίας βίντεο και εικόνας χρησιμοποιούνται για την αναβάθμιση παλαιών ταινιών και φωτογραφιών σε υψηλότερες αναλύσεις, καθώς και για τη δημιουργία εικονικών κόσμων σε βιντεοπαιχνίδια. Στον τομέα της ασφάλειας, η επεξεργασία εικόνας με ΑΙ χρησιμοποιείται ευρέως για την αναγνώριση προσώπων και την παρακολούθηση μέσω καμερών. Τα συστήματα αυτά έχουν τη δυνατότητα να ανιχνεύουν ύποπτες δραστηριότητες και να ειδοποιούν τις αρχές σε πραγματικό χρόνο.

Η αναγνώριση προσώπου είναι επίσης χρήσιμη σε εφαρμογές όπως τα συστήματα ασφαλείας για κινητά τηλέφωνα και άλλες ψηφιακές συσκευές. Παρά τα εντυπωσιακά αποτελέσματα, η επεξεργασία εικόνας με τη χρήση ΑΙ δεν είναι απαλλαγμένη από προκλήσεις. Ένα από τα βασικά ζητήματα είναι η ανάγκη για μεγάλα σύνολα δεδομένων προκειμένου να εκπαιδευτούν οι αλγόριθμοι, κάτι που μπορεί να είναι δύσκολο σε περιπτώσεις όπου τα δεδομένα είναι περιορισμένα ή ευαίσθητα, όπως στην ιατρική. Επιπλέον, προκύπτουν ζητήματα που αφορούν την ηθική και την προστασία της ιδιωτικότητας, ειδικά όταν η ΑΙ χρησιμοποιείται για την παρακολούθηση και την αναγνώριση προσώπων. Το μέλλον της επεξεργασίας εικόνας με ΑΙ είναι εξαιρετικά υποσχόμενο. Οι συνεχείς εξελίξεις στους αλγόριθμους βαθιάς μάθησης και η αυξανόμενη υπολογιστική ισχύς θα οδηγήσουν σε ακόμα πιο εξελιγμένα συστήματα επεξεργασίας εικόνας. Από την ενίσχυση της αυτοματοποίησης έως τη δημιουργία πλήρως ρεαλιστικών εικονικών κόσμων, οι εφαρμογές της ΑΙ στην επεξεργασία εικόνας θα συνεχίσουν να εξελίσσονται και να επεκτείνονται σε νέους τομείς. [\[11\]](#)[\[12\]](#)[\[13\]](#)

4.3 Κλιμάκωση εικόνας (Image Upscaling)

Αρχικά, οι παραδοσιακές μέθοδοι upscaling βασίζονταν σε απλές τεχνικές όπως η bicubic interpolation, η οποία υπολογίζει τα νέα pixels βάσει των γειτονικών pixels. Ωστόσο, αυτές οι μέθοδοι συχνά οδηγούσαν σε θολές και μη ρεαλιστικές εικόνες. Με την ανάπτυξη της τεχνητής νοημοσύνης και των νευρωνικών δικτύων, έχουν εμφανιστεί πιο προηγμένες μέθοδοι upscaling, όπως τα CNNs και τα GANs, που έχουν βελτιώσει σημαντικά την ποιότητα των αναβαθμισμένων εικόνων. Το image upscaling, στην ουσία, είναι η διαδικασία αύξησης της ανάλυσης μιας εικόνας, προκειμένου να δημιουργηθεί μια μεγαλύτερη εκδοχή της χωρίς να χάνεται η ποιότητα της εικόνας. Η κύρια πρόκληση της διαδικασίας αυτής είναι να προσδιοριστούν και να προστεθούν νέα εικονοστοιχεία (pixels) στην εικόνα με τέτοιο τρόπο ώστε να διατηρηθεί η λεπτομέρεια και η ευκρίνεια. Το upscaling χρησιμοποιείται ευρέως σε διάφορους τομείς, όπως η φωτογραφία, η κινηματογραφία, τα βιντεοπαιχνίδια και η ιατρική απεικόνιση.

Η διαδικασία image upscaling περιλαμβάνει διάφορα στάδια. Αρχικά, η εικόνα αναλύεται για να εντοπιστούν οι λεπτομέρειες και τα χαρακτηριστικά της. Οι αλγόριθμοι χρησιμοποιούν προηγμένα μαθηματικά μοντέλα για να υπολογίσουν ποια pixels πρέπει να προστεθούν, ώστε να διατηρηθεί η ποιότητα. Στο τελευταίο στάδιο, η νέα εικόνα κατασκευάζεται με τα υπολογισμένα pixels, προσφέροντας μια μεγαλύτερη και πιο καθαρή εκδοχή της αρχικής εικόνας. Η επεξεργασία εικόνας καλύπτει ένα ευρύτερο φάσμα τεχνικών και σκοπών σε σύγκριση με το upscaling. Ακολουθούν οι βασικές διαφορές. Ο κύριος σκοπός της κλιμάκωσης της εικόνας (image upscaling) είναι η αύξηση της ανάλυσης και της ποιότητας της εικόνας. Επικεντρώνεται κυρίως στην προσθήκη νέων λεπτομερειών και στην αποφυγή της θολούρας. Ο σκοπός της επεξεργασίας εικόνας με τις απλές παραδοσιακές μεθόδους είναι η διόρθωση χρώματος, η αφαίρεση θορύβου, η προσθήκη φίλτρων, ή τη κοπή και αναδιάταξη στοιχείων. Το image upscaling χρησιμοποιεί συγκεκριμένες τεχνικές, όπως η bicubic interpolation ή αλγόριθμους τεχνητής νοημοσύνης (π.χ. SRCNN, ESRGAN), που εστιάζουν στην αύξηση της ανάλυσης, σε αντίθεση με την απλή επεξεργασία εικόνας η οποία περιλαμβάνει ποικιλία τεχνικών όπως η ρύθμιση φωτεινότητας, η αντίθεση, η εφαρμογή φίλτρων και η διόρθωση ελαττωμάτων.

Οι μέθοδοι αυτές μπορεί να είναι χειροκίνητες ή αυτοματοποιημένες. Φυσικά με την χρήση του image upscaling το αποτέλεσμα είναι μια εικόνα με μεγαλύτερη ανάλυση και λεπτομέρειες. Ωστόσο, αν δεν γίνει σωστά, μπορεί να οδηγήσει σε θολές ή ανεπαρκείς εικόνες. Χρησιμοποιείται σε τομείς όπως η ιατρική απεικόνιση, η φωτογραφία, οι ταινίες, και

τα βιντεοπαιχνίδια, όπου είναι κρίσιμη η ποιότητα των εικόνων. Και στις δύο περιπτώσεις κύριο συστατικό για το καλύτερο αποτέλεσμα, έχει να κάνει με την ποιότητα της αρχικής εικόνας. Όσο πιο καθαρό είναι το Original Sample, τόσο καλύτερα αποτελέσματα θα έχουμε.

Η τελευταία δεκαετία έχει φέρει επανάσταση στη διαδικασία του image upscaling, με την εμφάνιση αλγορίθμων τεχνητής νοημοσύνης. Τα Συγκροτήματα Νευρωνικών Δικτύων (Convolutional Neural Networks) έχουν αποδειχθεί εξαιρετικά αποτελεσματικά για την αναβάθμιση εικόνας, επιτρέποντας στα μοντέλα να μάθουν λεπτομέρειες από μεγάλες βάσεις δεδομένων. Τα Γεννητικά Αντίστροφα Δίκτυα (Generative Adversarial Networks) προσφέρουν τη δυνατότητα παραγωγής ρεαλιστικών εικόνων μέσω της συνεργασίας δύο δικτύων: ενός γεννητικού και ενός διακριτικού. Οι προ-εκπαιδευμένοι αλγόριθμοι (Pre-trained models) μπορούν να χρησιμοποιηθούν για γρήγορη και αποδοτική εκπαίδευση, επιτρέποντας σε ειδικούς και μη ειδικούς να επιτύχουν υψηλής ποιότητας αναβάθμιση εικόνας χωρίς την ανάγκη για εκτενή εκπαίδευση. Θα τα αναλύσουμε περισσότερο σε επόμενο κεφάλαιο. Γενικότερα, το image upscaling και η επεξεργασία εικόνας είναι δύο θεμελιώδεις διαδικασίες που εξυπηρετούν διαφορετικούς σκοπούς στον κόσμο της ψηφιακής εικόνας. Ενώ το upscaling επικεντρώνεται στην αύξηση της ανάλυσης και της ποιότητας των εικόνων, η επεξεργασία εικόνας καλύπτει ένα ευρύτερο φάσμα αλλαγών και τροποποιήσεων. Η κατανόηση αυτών των διαφορών είναι κρίσιμη για την επιλογή των κατάλληλων μεθόδων και τεχνικών, ανάλογα με τις ανάγκες του εκάστοτε έργου. Η εξέλιξη της τεχνολογίας και οι νέες μέθοδοι που προκύπτουν καθιστούν και τις δύο διαδικασίες ολοένα και πιο σημαντικές σε πολλούς τομείς, από την τέχνη μέχρι την επιστήμη. [\[12\]\[13\]\[14\]](#)

4.4 Μέθοδοι κλιμάκωσης εικόνας (Image Upscaling)

Οι αλγόριθμοι τεχνητής νοημοσύνης (AI) για image upscaling έχουν φέρει επανάσταση στη διαδικασία αναβάθμισης της ανάλυσης εικόνας, βελτιώνοντας την ποιότητα και την ευκρίνεια των ψηφιακών εικόνων. Η σημαντικότερη μέθοδος που θα ασχοληθούμε είναι με βάση την τεχνολογία των Convolutional Neural Networks (CNNs). Τα CNNs είναι από τους πιο διαδεδομένους αλγόριθμους για την επεξεργασία εικόνας. Χρησιμοποιούνται σε μοντέλα που εκπαιδεύονται για να αναγνωρίζουν και να αναπαράγουν λεπτομέρειες εικόνας κατά τη διάρκεια της διαδικασίας upscaling. Δουλεύουν με τη δημιουργία χαρακτηριστικών (features) σε διάφορα επίπεδα, επιτρέποντας τη διαχείριση σύνθετων προτύπων και λεπτομερειών. Αρχικά το SRCNN είναι ένα από τα πρώτα CNNs που σχεδιάστηκε ειδικά για super resolution. Ο αλγόριθμος αυτός χρησιμοποιεί ένα τριών επιπέδων δίκτυο για να

αναλύσει και να ενισχύσει τις εικόνες, εστιάζοντας στην αποκατάσταση λεπτομερειών που χάνονται σε χαμηλές αναλύσεις. Το ESRGAN (Enhanced Super-Resolution Generative Adversarial Networks) είναι μια προχωρημένη έκδοση του GAN που χρησιμοποιείται για super-resolution. Συνδυάζει δύο δίκτυα—το γεννητικό και το διακριτικό—για να παράγει εικόνες υψηλής ανάλυσης με ρεαλιστικές λεπτομέρειες. Χρησιμοποιεί επίσης τεχνικές όπως perceptual loss για να βελτιώσει την ποιότητα των παραγόμενων εικόνων. Σε συνέχεια, Το FSRCNN (fast Super Resolution Convolutional Neural Networks) είναι μια βελτιωμένη εκδοχή του SRCNN που έχει σχεδιαστεί για ταχύτερη επεξεργασία. Χρησιμοποιεί μια διαφορετική αρχιτεκτονική που επιτρέπει την ταχύτερη εκπαίδευση και λειτουργία, χωρίς να θυσιάζει την ποιότητα.

Όπως αναφέραμε και προηγουμένως, έχουμε και τα Generative adversarial Networks Τα GANs είναι ευρέως γνωστά για τη δυνατότητά τους να δημιουργούν νέες εικόνες από το μηδέν. Στην επεξεργασία εικόνας, χρησιμοποιούνται για να προσφέρουν ρεαλιστικές λεπτομέρειες κατά το upscaling. Τα GANs λειτουργούν με τη δημιουργία ενός μοντέλου που μπορεί να διακρίνει μεταξύ "πραγματικών" και "ψεύτικων" εικόνων, βελτιώνοντας συνεχώς την παραγωγή τους.

Η τεχνική Deep Image Prior βασίζεται στην ιδέα ότι ένα δίκτυο μπορεί να εκπαιδευτεί μόνο με τη χρήση της εικόνας που θέλουμε να αναβαθμίσουμε. Ο Deep Image Prior επιτρέπει την επεξεργασία εικόνας χωρίς να χρειάζεται εκτενή εκπαίδευση σε μεγάλες βάσεις δεδομένων, χρησιμοποιώντας απλώς την εικόνα ως εισαγωγή. Φυσικά ένα από τα σημαντικότερα εργαλεία της τεχνητής νοημοσύνης είναι και οι Auto Encoders. Οι auto encoders είναι αλγόριθμοι που μαθαίνουν να αναγνωρίζουν και να αναπαράγουν πληροφορίες εικόνας. Χρησιμοποιούνται για την αναβάθμιση της ανάλυσης εικόνας, καθώς μπορούν να συλλάβουν τις βασικές δομές και λεπτομέρειες κατά τη διαδικασία εκπαίδευσης.

Έπειτα υπάρχουν και τα RNNs και συγκεκριμένα η μέθοδος SRRNN (Super-Resolution Using Recurrent Neural Networks). Αυτή η τεχνική εκμεταλλεύεται την ικανότητα των RNNs να διαχειρίζονται σειρές δεδομένων. Χρησιμοποιείται κυρίως σε περιπτώσεις όπου η χρονική διάσταση των εικόνων είναι σημαντική, όπως σε βίντεο.

Ο αλγόριθμος PRSR (Pixell Recursive Super Resolution) βασίζεται στη διαδικασία δειγματοληψίας εικονοστοιχείων (pixels) και χρησιμοποιεί τη θεωρία των πιθανοτήτων για να δημιουργήσει νέα εικονοστοιχεία που είναι συνεπή με την ήδη υπάρχουσα εικόνα. Αυτό επιτρέπει τη δημιουργία εικόνων υψηλής ανάλυσης με πιο ρεαλιστικά αποτελέσματα. Και η μέθοδος Self-Supervised Learning, επιτρέπει στα μοντέλα να μάθουν από τις ίδιες τις εικόνες

χωρίς την ανάγκη για επεξεργασία ετικετών (labels). Χρησιμοποιεί τεχνικές όπως η δημιουργία παραλλαγών μιας εικόνας για να εκπαιδεύσει το μοντέλο να αναγνωρίζει και να αποκαθιστά λεπτομέρειες.

Εν κατακλείδι, η επεξεργασία εικόνας με τη βοήθεια τεχνητής νοημοσύνης έχει ανοίξει νέες προοπτικές για το image upscaling, με τις τεχνικές αυτές να βελτιώνουν σημαντικά την ποιότητα των ψηφιακών εικόνων. Καθώς οι αλγόριθμοι συνεχίζουν να εξελίσσονται, αναμένονται ακόμη πιο εντυπωσιακά αποτελέσματα στον τομέα αυτό, με εφαρμογές που εκτείνονται από τη φωτογραφία και το design μέχρι τη ιατρική και τη βιομηχανία. [\[13\]\[14\]](#)

4.5 Ανάλυση βασικών μεθόδων Image Upscaling

Όπως αναφέραμε και στα προηγούμενα κεφάλαια, η γενική ιδέα για το πως λειτουργεί η διαδικασία του image upscaling είναι κατά βάση η μεγέθυνση των γειτονικών εικονοστοιχείων. Κατά το upscaling μιας εικόνας, πρέπει να προστεθούν νέα pixels για να καλυφθεί ο μεγαλύτερος χώρος της νέας εικόνας. Η διαδικασία αυτή ονομάζεται με τους μαθηματικούς ορισμούς ως παρεμβολή (interpolation). Το image upscale interpolation είναι μια τεχνική που χρησιμοποιείται για την αύξηση της ανάλυσης μιας εικόνας, δηλαδή τη μεγέθυνσή της, προσθέτοντας νέα εικονοστοιχεία (pixels) μεταξύ των ήδη υπαρχόντων. Το interpolation προσπαθεί να εκτιμήσει την τιμή των νέων pixels με βάση τις τιμές των γειτονικών pixels. Υπάρχουν διάφορες μέθοδοι για την εκτίμηση αυτή, τις οποίες και θα αναλύσουμε με αριθμητικά παραδείγματα. [\[15\]\[16\]\[17\]](#)

4.5.1 Nearest Neighbor Interpolation

Η μέθοδος nearest neighbor (πλησιέστερου γείτονα) για το image upscaling είναι μια απλή και γρήγορη τεχνική για την αύξηση της ανάλυσης μιας εικόνας. Στην ουσία, η μέθοδος αυτή επιλέγει την τιμή του πλησιέστερου pixel για να γεμίσει νέα κενά pixel που δημιουργούνται κατά τη μεγέθυνση της εικόνας. Δεν γίνεται καμία εκτίμηση ή εξομάλυνση των τιμών μεταξύ των pixels, κάτι που την καθιστά πολύ γρήγορη αλλά μπορεί να δημιουργήσει μια “pixelated” όψη.

Για την καλύτερη κατανόηση της μεθόδου, χρησιμοποιούμε ένα απλό αριθμητικό παράδειγμα για μία εικόνα 2 X 2.

Pixel position	Value
(1,1)	100
(1,2)	150
(2,1)	200
(2,2)	250

Και θέλουμε να μεγαλώσουμε την συγκεκριμένη φωτογραφία στο διπλάσιο μέγεθος, δηλαδή 4 x 4, με την συγκεκριμένη μέθοδο Nearest Neighbor. Στην ουσία, δημιουργούμε έναν πίνακα μεγέθους 4x4 και για κάθε νέο εικονοστοιχείο pixel αντιγράφουμε το αντίστοιχο γειτονικό pixel από τον αρχικό πίνακα.

New Position	Copied Pixels	Original pixel position
(1,1)	100	(1,1)
(1,2)	100	(1,1)
(1,3)	150	(1,2)
(1,4)	150	(1,2)
(2,1)	100	(1,1)
(2,2)	100	(1,1)
(2,3)	150	(1,2)
(2,4)	150	(1,2)
(3,1)	200	(2,1)
(3,2)	200	(2,1)
(3,3)	250	(2,2)
(3,4)	250	(2,2)
(4,1)	200	(2,1)
(4,2)	200	(2,1)
(4,3)	250	(2,2)
(4,4)	250	(2,2)

Με την εφαρμογή λοιπόν της image upscaling, nearest neighbor μέθοδο, η νέα μας εικόνα θα έχει της εξής μορφή.

	1	2	3	4
1	100	100	150	150
2	100	100	150	150
3	200	200	250	250
4	200	200	250	250

Είναι μία απλή και γρήγορη διαδικασία και δεν χρειάζεται μεγάλη υπολογιστική ισχύ αλλά η ποιότητα της εικόνας μπορεί να αλλοιωθεί σε μεγάλο βαθμό, λόγω της απλής αντιγραφής των pixel. [\[18\]](#)[\[19\]](#)

4.5.2 Bilinear Interpolation

Η bilinear interpolation είναι μια μέθοδος upscaling εικόνας που χρησιμοποιεί γραμμική παρεμβολή μεταξύ των γειτονικών pixels για να υπολογίσει τις τιμές των νέων pixels σε μια μεγαλύτερη εικόνα. Η μέθοδος αυτή λαμβάνει υπόψη τα τέσσερα γειτονικά pixels και χρησιμοποιεί τη μέση τιμή τους για να δημιουργήσει πιο ομαλές μεταβάσεις, σε αντίθεση με τη μέθοδο nearest neighbor, που απλώς αντιγράφει την τιμή του πλησιέστερου pixel.

Χρησιμοποιώντας έναν πίνακα 2x2 διαστάσεων μπορούμε να αναλύσουμε την μέθοδο αυτή, με την χρήση απλών αριθμητικών τιμών.

Pixel position	Value
(1,1)	100
(1,2)	150
(2,1)	200
(2,2)	250

Για κάθε νέο pixel, υπολογίζουμε τη μέση τιμή χρησιμοποιώντας τα τέσσερα κοντινότερα pixels από την αρχική εικόνα. Η bilinear interpolation κάνει σταδιακή μετάβαση μεταξύ των τιμών αυτών, δημιουργώντας πιο ομαλές μεταβάσεις στην τελική εικόνα.

Για τα γωνιακά pixel, παίρνουμε τις τιμές των γωνιακών από τον αρχικό πίνακα 2x2, δηλαδή **(1, 1) -> 100, (1, 4) -> 150, (4, 1) -> 200, (4, 4) -> 250.**

Για τα ενδιάμεσα pixel, υπολογίζουμε τη μέση τιμή με βάση τα τέσσερα πλησιέστερα pixels της αρχικής εικόνας. Για παράδειγμα για το pixel, (1,2) στην 4x4 εικόνα, θα χρησιμοποιήσουμε τα αρχικά pixels (1, 1) και (1, 2), οπότε

- **(1,2) = (100 + 150) / 2 = 125**

Και για το pixel (2,1),

- **(2,1) = (100 + 200) / 2 = 150**

Για το pixel (2, 2), που βρίσκεται στο κέντρο, θα λάβουμε υπόψη και τα τέσσερα αρχικά pixels (1, 1), (1, 2), (2, 1), (2, 2), δηλαδή,

- **(2,2) = (100 + 150 + 200 + 250) / 2 = 175**

Με την εφαρμογή της μεθόδου bilinear interpolation, η νέα εικόνα 4x4 θα έχει τις εξής τιμές:

	1	2	3	4
1	100	125	150	150
2	150	175	200	200
3	200	225	250	250

4	200	225	250	250
---	-----	-----	-----	-----

Η μέθοδος bilinear interpolation δημιουργεί πιο ομαλές εικόνες σε σύγκριση με τη μέθοδο nearest neighbor και με καλύτερη ποιότητα όταν γίνεται upscaling, καθώς δημιουργεί πιο φυσικές μεταβάσεις μεταξύ των pixels, αλλά παρότι είναι πιο ομαλή, η bilinear interpolation δεν διατηρεί όλες τις λεπτομέρειες της εικόνας και μπορεί να οδηγήσει σε θόλωση σε σχέση με πιο προηγμένες μεθόδους [20]

4.5.3 Bicubic Interpolation

Η μέθοδος bicubic interpolation είναι μια πιο εξελιγμένη μέθοδος upscaling εικόνας σε σύγκριση με τις nearest neighbor και bilinear interpolation. Χρησιμοποιεί μια μεγαλύτερη περιοχή γειτονικών pixels (16 συνολικά σε ένα πλέγμα 4x4 γύρω από το pixel που θέλουμε να υπολογίσουμε) και λαμβάνει υπόψη όχι μόνο τις τιμές των pixels αλλά και τις παραγώγους τους. Αυτό οδηγεί σε πιο ομαλή εικόνα με καλύτερη λεπτομέρεια και λιγότερη θόλωση.

Για την καλύτερη κατανόηση θα χρησιμοποιήσουμε ένα απλό αριθμητικό παράδειγμα με έναν πίνακα 2x2 και θα εφαρμόσουμε την μέθοδο Bicubic Interpolation για την μεγέθυνση της εικόνας σε 4x4.

Pixel position	Value
(1,1)	100
(1,2)	150
(2,1)	200
(2,2)	250

Η bicubic interpolation υπολογίζει τις τιμές των νέων pixels χρησιμοποιώντας ένα μαθηματικό μοντέλο που λαμβάνει υπόψη τις τιμές και τις παραγώγους (προσεγγίζοντας τις κλίσεις) των γειτονικών pixels. Αυτό γίνεται για να εξασφαλιστεί μια ομαλότερη και πιο φυσική μετάβαση μεταξύ των pixels. Δηλαδή για κάθε νέο pixel, χρησιμοποιούμε τα 16 γειτονικά pixels (4x4 πλέγμα) γύρω από αυτό, και εφαρμόζουμε μια κυβική συνάρτηση (cubically-weighted average) για να υπολογίσουμε τη νέα τιμή. Χρησιμοποιώντας λοιπόν τον παραπάνω αρχικό πίνακα, για αρχή όπως και με την μέθοδο Bilinear Interpolation, πρώτα θα εξετάσουμε τα γωνιακά pixel Τα νέα pixel για τον πίνακα θα έχουν τις τιμές, (1,1) -> 100, (1,4) -> 150, (4,1) -> 200, (4,4) -> 250.

Έπειτα προχωράμε στα ενδιάμεσα pixel. Για να υπολογίσουμε τα ενδιάμεσα pixels, χρησιμοποιούμε μια κυβική συνάρτηση με βάση τις τιμές των γειτονικών pixels και τις κλίσεις τους. Ένα απλό παράδειγμα για το pixel. Οπότε το αποτέλεσμα θα είναι:

- Για το **(2,2)** = **(100 + 150 + 200 + 250) / 4 = 175**
- Για το **(1,2)** = **((100 x w1) + (150 x w2) / (w1 + w2) = 125**, όπου **w1** και **w2** είναι οι κυβικές συντελεστές που εξαρτώνται από την απόσταση του pixel από τα γειτονικά του.

Με την εφαρμογή της μεθόδου bicubic interpolation, η νέα εικόνα 4x4 μπορεί να έχει τις εξής τιμές

	1	2	3	4
1	100	125	150	150
2	150	175	200	200
3	200	225	250	250
4	200	225	250	250

Η μέθοδος bicubic interpolation δημιουργεί πιο ομαλές και λεπτομερείς εικόνες σε σχέση με τις προηγούμενες μεθόδους (nearest neighbor, bilinear και παρέχουν καλύτερη ανάλυση στην μεγεθυμένη εικόνα με μειωμένα επίπεδα θόλωσης και τεχνητών γραμμών. Το αρνητικό είναι ότι η μέθοδος απαιτεί περισσότερη υπολογιστική ισχύ, λόγω της πολυπλοκότητας των υπολογισμών. Η bicubic interpolation προσφέρει την καλύτερη ισορροπία μεταξύ ποιότητας και υπολογιστικής ισχύος, γι' αυτό και χρησιμοποιείται συχνά σε προγράμματα επεξεργασίας εικόνας [\[21\]](#)[\[22\]](#)

4.5.4 Lanczos Interpolation

Η μέθοδος Lanczos είναι μια τεχνική interpolation που χρησιμοποιείται για το image upscaling και προσφέρει ακόμα καλύτερα αποτελέσματα σε σχέση με άλλες μεθόδους όπως η nearest neighbor, bilinear και bicubic. Χρησιμοποιεί μια windowed sinc function που λαμβάνει υπόψη μεγαλύτερο αριθμό γειτονικών pixels, με σκοπό να διατηρήσει περισσότερες λεπτομέρειες και να μειώσει τις ατέλειες. Η Lanczos interpolation βασίζεται στη χρήση της συνάρτησης sinc, η οποία είναι ιδανική για την ανακατασκευή σήματος. Στην πράξη, η sinc χρησιμοποιείται με ένα παράθυρο, δηλαδή περιορίζεται η συνάρτηση σε έναν συγκεκριμένο αριθμό γειτονικών pixels για να μειωθούν οι υπολογιστικές απαιτήσεις και να περιοριστούν τα φαινόμενα κυματισμών.

Για μία ακόμα φορά, για την αριθμητική κατανόηση της μεθόδου θα χρησιμοποιήσουμε τον ίδιο πίνακα 2x2, με τις παρακάτω αριθμητικές τιμές.

Pixel position	Value
(1,1)	100
(1,2)	150
(2,1)	200
(2,2)	250

Και θέλουμε να κάνουμε upscaling αυτής της εικόνας σε μια 4x4 εικόνα χρησιμοποιώντας τη μέθοδο Lanczos με έναν παράγοντα "a" = 2, δηλαδή χρησιμοποιώντας ένα παράθυρο που υπολογίζει τις τιμές με βάση τα 4 πλησιέστερα pixels.

Για κάθε νέο pixel που προκύπτει από το upscaling, η Lanczos συνάρτηση χρησιμοποιεί τα γειτονικά pixels και εφαρμόζει την εξής φόρμουλα:

- $L(x) = \text{sinc}(x) \times \text{sinc}(a / x)$, όπου η ημίτονη συνάρτηση $\sin(x)$:
- $\text{Sin}(x) = \sin(\pi x) / \pi x$

Αυτή η συνάρτηση εφαρμόζεται σε έναν ορισμένο αριθμό γειτονικών pixels (π.χ., 4x4 ή 3x3), και το αποτέλεσμα είναι η τιμή του νέου pixel. Η Lanczos μέθοδος προσφέρει καλύτερη διατήρηση λεπτομερειών επειδή εκμεταλλεύεται περισσότερα pixels από τη γειτονιά του σημείου που θέλουμε να υπολογίσουμε, σε σύγκριση με τις πιο απλές μεθόδους.

Σύμφωνα λοιπόν με τον παραπάνω πίνακα, ας υποθέσουμε ότι θέλουμε να υπολογίσουμε το νέο pixel στη θέση (2, 2) της εικόνας. Για τον υπολογισμό αυτόν, θα χρησιμοποιήσουμε τα 4 κοντινότερα pixels και θα εφαρμόσουμε τη Lanczos συνάρτηση. Για παράδειγμα, για το pixel (2, 2), αν εφαρμόσουμε την παραπάνω συνάρτηση με τα γύρω pixels, μπορούμε να υπολογίσουμε την τιμή του ως εξής:

- $\text{Pixel}(2,2) = 100 \times L(0.5) + 150 \times L(1.5) + 200 \times L(2.5) + 250 \times L(3.5)$

Αυτό το αποτέλεσμα θα δώσει μια βελτιωμένη τιμή για το pixel (2, 2), λαμβάνοντας υπόψη τις συνεισφορές των γύρω pixels με βάση την sinc συνάρτηση. Αν υπολογίσουμε απλοποιημένα με συντελεστές που αντιπροσωπεύουν τις κοντινές θέσεις, μπορούμε να υποθέσουμε για τη θέση (2, 2) ότι το νέο pixel θα είναι ένας συνδυασμός των αρχικών τιμών με βάρη που προκύπτουν από τη συνάρτηση Lanczos.

Υποθέτοντας λοιπόν ότι το βάρος για τα πλησιέστερα pixels L, είναι 0.5, τότε στην συγκεκριμένη περίπτωση, η νέα τιμή θα είναι

- $\text{Pixel}(2,2) = 100 \times 0.25 + 150 \times 0.5 + 200 \times 0.75 + 250 \times 1 = 50 + 75 + 150 + 250 = 525$

Αυτό, προφανώς, είναι ένα απλοποιημένο παράδειγμα της βασικής ιδέας, αν και η Lanczos μέθοδος είναι πολύ πιο ακριβής από αυτή την απλοποιημένη εκδοχή. Η Lanczos interpolation προσφέρει πολύ υψηλή ποιότητα εικόνας για το upscaling, με διατήρηση των λεπτομερειών και λιγότερο "θόρυβο" και μειώνει τα φαινόμενα θολώματος και κυματισμών που εμφανίζονται με άλλες μεθόδους.

Τα αρνητικά βεβαίως είναι ότι απαιτεί περισσότερη υπολογιστική ισχύ από άλλες απλούστερες μεθόδους, λόγω των πιο περίπλοκων υπολογισμών και μπορεί να παράγει κάποιους τεχνητούς κυματισμούς στις περιοχές με μεγάλες αντιθέσεις, αν και είναι πιο ελεγχόμενοι από άλλες μεθόδους.

Συγκριτικά λοιπόν με βάση και των προηγούμενων μεθόδων συμπεραίνουμε πως:

Interpolation Methods	Quality	PC Power Requirements	Use
Neighbor	Low	Very Low	Simple upscaling, no details
Bilinear	Medium	Low	Simple quality, some details
Bicubic	High	Medium	Great quality, less blur issues
Lanczos	Very High	High	Great detail retainment

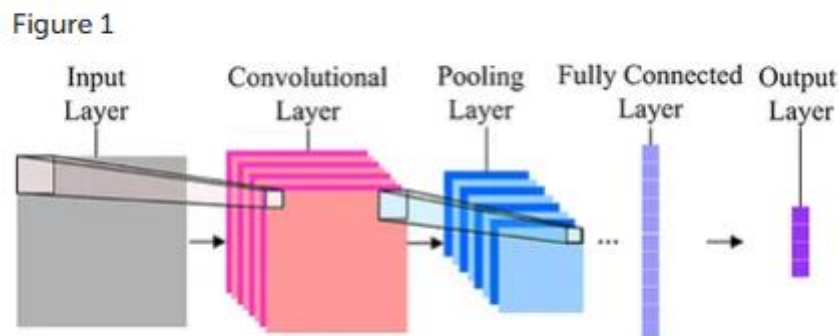
Η Lanczos interpolation είναι μία από τις καλύτερες μεθόδους για upscaling, ειδικά όταν θέλουμε να διατηρήσουμε τις λεπτομέρειες σε εικόνες υψηλής ανάλυσης. [\[23\]](#)[\[24\]](#)[\[25\]](#)

ΚΕΦΑΛΑΙΟ 5

5. Εισαγωγή στα Convolutional Neural Networks

5.1 CNN'S

Τα Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks - CNNs) είναι ένας τύπος νευρωνικών δικτύων που έχουν σχεδιαστεί με μορφή πλέγματος (Grid), όπως είναι οι εικόνες. Τα CNNs είναι εξαιρετικά αποτελεσματικά για εργασίες αναγνώρισης προτύπων, όπως η αναγνώριση αντικειμένων, η ταξινόμηση εικόνων, και η ανάλυση βίντεο. Χρησιμοποιούν στρώματα συνελίξεων (Convolutional Layers) για να εντοπίσουν αυτόματα χαρακτηριστικά από τα δεδομένα εισόδου, χωρίς την ανάγκη χειροκίνητης εξαγωγής χαρακτηριστικών. Τα συνελικτικά στρώματα είναι το κεντρικό χαρακτηριστικό των CNN. Σκοπός τους είναι να εξάγουν, αυτόματα, χαρακτηριστικά από τα δεδομένα εισόδου (Input), όπως γωνίες (edges), υφές (textures) και περίπλοκα μοτίβα (complex patterns).



Εικόνα 23 CNN Layers

Το πρώτο εργαλείο είναι τα φίλτρα (Filters ή Kernels). Τα φίλτρα είναι μικρά πλέγματα δεδομένων, π.χ. διαστάσεων 3x3 ή 5x5, που σαρώνουν την είσοδο και ανιχνεύουν συγκεκριμένα χαρακτηριστικά. Κάθε φίλτρο είναι υπεύθυνο για την ανίχνευση ενός διαφορετικού τύπου χαρακτηριστικού. Κατά τη σάρωση του φίλτρου πάνω από την εικόνα, πραγματοποιείται στοιχείο-προς-στοιχείο πολλαπλασιασμός των τιμών του φίλτρου με τις αντίστοιχες τιμές της εικόνας. Τα αποτελέσματα συνοψίζονται για να δημιουργήσουν ένα νέο

χαρακτηριστικό χάρτη (map feature), που περιγράφει την παρουσία αυτού του χαρακτηριστικού σε διάφορα σημεία της εικόνας. Τα φίλτρα κινούνται στην εικόνα εισόδου (Input) όσο πραγματοποιούν την συνελκτική σάρωση. Ο ρυθμός με τον οποίο κινείται ένα φίλτρο ονομάζεται διαβάθμιση (Stride). Η συμπλήρωση (padding) προστίθεται στις άκρες της εικόνας για να διατηρηθεί η αρχική της διάσταση μετά από τη συνέλιξη. Μετά τη συνέλιξη, εφαρμόζεται μια μη γραμμική συνάρτηση ενεργοποίησης σε κάθε τιμή του χάρτη χαρακτηριστικών.

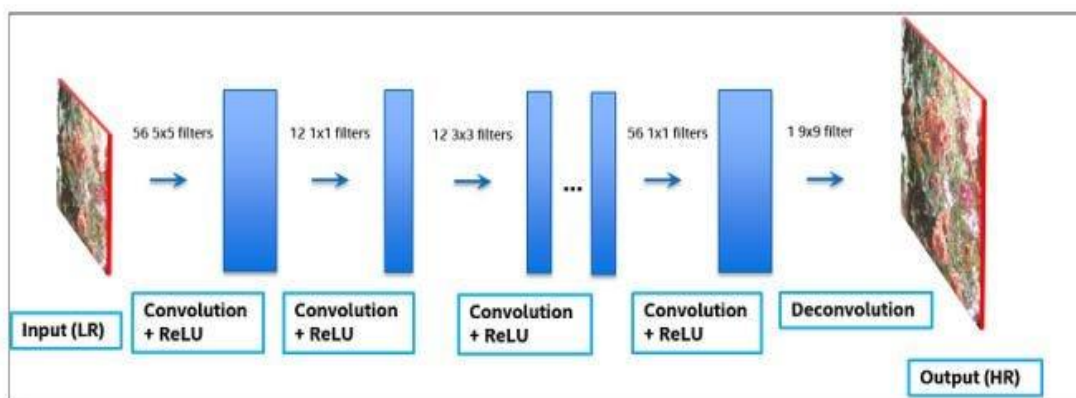
Η πιο κοινή είναι η ReLU (Rectified Linear Unit), η οποία αντικαθιστά τις αρνητικές τιμές με το μηδέν. Η προσθήκη μη γραμμικότητας επιτρέπει στο CNN να μάθει πιο περίπλοκα χαρακτηριστικά. Η συγκεκριμένη συνάρτηση ονομάζεται, συνάρτηση ενεργοποίησης (Activation Function). Βασικό στοιχείο των CNN είναι η χρήση πολύπλοκων στρώματων (Layers), ξεκινώντας με τα Pooling Layers (Στρώματα υποδειγματοληψίας).

Τα Pooling Layers μειώνουν τις διαστάσεις των map features, μειώνοντας έτσι την υπολογιστική πολυπλοκότητα του μοντέλου και καθιστώντας το πιο ανθεκτικό σε μικρές μετατοπίσεις στην είσοδο, όπου, το ονομαζόμενο Max Pooling, παίρνει τη μέγιστη τιμή σε ένα μικρό παράθυρο (π.χ. 2x2) από κάθε map feature και τη χρησιμοποιεί ως εκπρόσωπο σε αντίθεση με το Average Pooling, το οποίο, αντί της μέγιστης τιμής, χρησιμοποιεί τον μέσο όρο των τιμών σε κάθε παράθυρο. Τα στρώματα υποδειγματοληψίας (Pooling Layers) βοηθούν στη μείωση του μεγέθους των δεδομένων, αλλά διατηρούν τα σημαντικά χαρακτηριστικά, καθιστώντας το μοντέλο πιο αποδοτικό. Μετά τα Convolutional και Pooling Layers, τα map features απλώνονται σε έναν μονοδιάστατο πίνακα και συνδέονται με πλήρως συνδεδεμένα στρώματα (Fully Connected Layers). Σε αυτά τα στρώματα, κάθε νευρώνας συνδέεται με κάθε νευρώνα του προηγούμενου στρώματος. Αυτά τα στρώματα συνήθως χρησιμοποιούνται προς το τέλος του δικτύου και είναι υπεύθυνα για τη λήψη τελικών αποφάσεων, όπως η κατηγοριοποίηση σε διάφορες κλάσεις. Τέλος το τελικό Layer είναι το στρώμα εξόδου (Output Layer), το οποίο σε προβλήματα ταξινόμησης εικόνων συνήθως χρησιμοποιεί τη συνάρτηση softmax για να παράγει μια πιθανότητα για κάθε κατηγορία. Το αποτέλεσμα είναι η κατηγορία με τη μεγαλύτερη πιθανότητα.

Η χρήση φίλτρων και η συνελκτική λειτουργία μειώνει σταδιακά τις διαστάσεις των δεδομένων, επιτρέποντας στο δίκτυο να εκπαιδευτεί γρηγορότερα και πιο αποτελεσματικά. Αυτή η διαδικασία γίνεται με την εκμετάλλευση των χωρικών σχέσεων των δεδομένων εισόδου. Τα ίδια φίλτρα εφαρμόζονται σε ολόκληρη την εικόνα, κάτι που μειώνει σημαντικά τον αριθμό των παραμέτρων που πρέπει να εκπαιδευτούν, σε σύγκριση με τα πλήρως

συνδεδεμένα στρώματα. Η διεργασία αυτή ονομάζεται παραμετρική μείωση (Parameter Sharing).

Υπάρχουν πολλές διαφορετικές αρχιτεκτονικές CNNs που έχουν αναπτυχθεί με την πάροδο του χρόνου, μία από τις πιο διάσημες είναι η LeNet. Ένα από τα πρώτα CNNs, σχεδιασμένο από τον Yann LeCun για την αναγνώριση χειρόγραφων ψηφίων. Αποτελείται από λίγα συνελκτικά και πλήρως συνδεδεμένα στρώματα, και ήταν πολύ επιτυχημένο στην αναγνώριση χαρακτήρων. Το 2012 έγινε εξίσου διάσημη, η αρχιτεκτονική AlexNet, κερδίζοντας τον διαγωνισμό ImageNet, επιτυγχάνοντας σημαντική μείωση του σφάλματος στην ταξινόμηση εικόνων. Εισήγαγε τη χρήση πολλών συνελκτικών στρωμάτων και ReLU ενεργοποιήσεων. Το VGGNet χαρακτηρίζεται από τη χρήση πολλών συνελκτικών στρωμάτων με μικρά φίλτρα (3x3). Ενώ είναι αρκετά πιο εξειδικευμένο από το AlexNet, παραμένει σχετικά απλό και έχει χρησιμοποιηθεί ευρέως στην έρευνα και την πρακτική εφαρμογή. Αξιοσημείωτη αρχιτεκτονική είναι η ResNet. Η ResNet εισήγαγε τις υπολειμματικές συνδέσεις (residual connections), οι οποίες επιτρέπουν την εκπαίδευση πολύ βαθύτερων δικτύων χωρίς να αντιμετωπίζουν προβλήματα υποβάθμισης της ακρίβειας. Η πιο πρόσφατη και μία από τις πιο διαδεδομένες αρχιτεκτονικές είναι η GoogleNet. Η GoogleNet παρουσίασε το Inception module, το οποίο εφαρμόζει πολλαπλές συνελκτικές πράξεις με διαφορετικά μεγέθη φίλτρων στο ίδιο επίπεδο και συνδυάζει τα αποτελέσματα. Αυτό μειώνει το υπολογιστικό κόστος χωρίς να θυσιάζεται η ακρίβεια.



Εικόνα 24 CNN RELU

Τα Convolutional Neural Networks (CNNs), είναι ένα από τα πιο κύρια και βασικά εργαλεία στον τομέα της μηχανικής μάθησης (Machine Learning), επιτρέποντας την αυτόματη ανάλυση και κατανόηση οπτικών δεδομένων. Τα CNNs χρησιμοποιούνται ευρέως στην ταξινόμηση εικόνων, όπως για την αναγνώριση προσώπων, αντικειμένων ή σκηνών. Με την ανίχνευση αντικειμένων, τα CNNs μπορούν όχι μόνο να αναγνωρίσουν ποια αντικείμενα

υπάρχουν σε μια εικόνα, αλλά και να εντοπίσουν την ακριβή τους θέση. Χρησιμοποιούνται στην ανάλυση ιατρικών εικόνων, όπως η ανίχνευση όγκων σε MRI ή η αναγνώριση παθήσεων σε ακτινογραφίες. Βέβαια, εκτός από εικόνες, τα CNNs μπορούν να επεκταθούν και στην ανάλυση βίντεο, αναγνωρίζοντας αντικείμενα ή σκηνές σε κάθε καρέ και παρακολουθώντας την κίνηση.

Αν και αρχικά σχεδιάστηκαν για εικόνες, τα CNNs έχουν εφαρμοστεί και σε κείμενα για καθήκοντα όπως η ανάλυση συναισθήματος και η ταξινόμηση κειμένων, με την επανομαζόμενη διαδικασία Natural Language Process (Επεξεργασία Φυσικής Γλώσσας). Παρόλο που τα CNNs είναι πολύ αποτελεσματικά, υπάρχουν κάποιες προκλήσεις που πρέπει να αντιμετωπιστούν. Τα CNNs απαιτούν μεγάλες ποσότητες δεδομένων για να εκπαιδευτούν αποτελεσματικά, ιδιαίτερα για περίπλοκα καθήκοντα. Οι πιο εξειδικευμένες αρχιτεκτονικές, όπως η VGG και η ResNet, μπορούν να είναι εξαιρετικά υπολογιστικά ακριβές, απαιτώντας σημαντική επεξεργαστική ισχύ και χρόνο εκπαίδευσης. Για να ξεπεραστούν αυτές οι προκλήσεις, νέες τεχνικές και αρχιτεκτονικές συνεχίζουν να αναπτύσσονται. Ορισμένες βελτιώσεις περιλαμβάνουν τη χρήση προ-εκπαιδευμένων μοντέλων, τη μεταφορά μάθησης (transfer learning) και την εφαρμογή μοντέλων συμπίεσης (model compression) για να μειωθεί η πολυπλοκότητα. Τα Συνελκτικά Νευρωνικά Δίκτυα αποτελούν μια από τις πιο σημαντικές τεχνολογίες στη σύγχρονη μηχανική μάθηση και έχουν συμβάλει δραστικά στην εξέλιξη των εφαρμογών επεξεργασίας εικόνας και βίντεο. Με την αυτόματη εξαγωγή χαρακτηριστικών και τη βαθιά ικανότητα μάθησης, τα CNNs έχουν καταφέρει να ξεπεράσουν τις παραδοσιακές μεθόδους στην ανάλυση οπτικών δεδομένων και να ανοίξουν το δρόμο για νέες, επαναστατικές εφαρμογές στην τεχνητή νοημοσύνη. [\[26\]\[27\]\[28\]](#)

5.2 Αριθμητική ανάλυση CNN

Για την πιο εύκολη κατανόηση, θα αναλύσουμε την image upscaling διαδικασία που χρησιμοποιούν τα Νευρωνικά Δίκτυα, με την χρήση όσο πιο απλουστευμένων τιμών γίνεται. Ας υποθέσουμε αρχικά ότι έχουμε μία grayscale εικόνα 5x5 pixel, όπου κάθε αριθμός αντιστοιχεί με την ένταση/βάθος (intensity) ενός pixel.

- Εισαγωγή εικόνας **Input** 5x5

1	2	0	1	3
0	1	1	2	2

2	2	0	0	1
1	0	1	2	3
3	1	2	1	0

Αυτή είναι η είσοδος του δικτύου. Κάθε αριθμός αντιπροσωπεύει την ένταση φωτός του αντίστοιχου pixel. Τώρα, θα χρησιμοποιήσουμε ένα φίλτρο (kernel) μεγέθους 3x3 για να εντοπίσουμε χαρακτηριστικά στην εικόνα. Το φίλτρο είναι το εξής:

- Φίλτρο **Kernel** 3x3

1	0	-1
1	0	-1
1	0	-1

Αυτό το φίλτρο έχει σχεδιαστεί για να ανιχνεύει κάθετες άκρες στην εικόνα. Θα εφαρμόσουμε τη convolution process μετακινώντας το φίλτρο πάνω στην εικόνα. Για κάθε θέση του φίλτρου, θα υπολογίσουμε το γινόμενο κάθε στοιχείου του φίλτρου με το αντίστοιχο στοιχείο της εικόνας και θα προσθέσουμε τα γινόμενα. Ας δούμε πώς θα εφαρμόσουμε τη συνελίξη σε μια συγκεκριμένη θέση, αρχίζοντας από το πάνω αριστερό μέρος της εικόνας (καλύπτουμε τα πρώτα 3x3 pixels).

- **Convolution Process Application**

Πολλαπλασιάζουμε τα πρώτα 3x3 με το Kernel

1	2	0
0	1	1
2	2	0

1*1	+	0*2	+	(-1)*0	=	1
1*0	+	0*1	+	(-1)*1	=	-1
1*2	+	0*2	+	(-1)*0	=	2

Το άθροισμα των παραπάνω τιμών (**Output**) είναι: $1 + (-1) + 2 = 2$, το οποίο σημαίνει ότι το πρώτο στοιχείο του νέου map feature θα είναι το ψηφίο 2. Επαναλαμβάνουμε τη διαδικασία για κάθε θέση του φίλτρου στην εικόνα, πηγαίνοντας κάθε φορά μία θέση δεξιά.

Οπότε από τον αρχικό πίνακα 5X5, επιλέγουμε, ξεκινώντας από την δεύτερη στήλη, τα επόμενα 9 pixel και εφαρμόζουμε πάλι το φίλτρο kernel.

2	0	1
1	1	2
2	0	0

2*1	+	0*0	+	1*(-1)	=	1
1*1	+	1*0	+	2*(-1)	=	-1
2*1	+	0*0	+	0*(-1)	=	2

Εφαρμόζοντας το φίλτρο 3X3, από την δεύτερη θέση του αρχικού πίνακα, παίρνουμε πάλι αποτέλεσμα **Output = 2**. Προχωράμε λοιπόν στο επόμενο φίλτρο 3X3, ξεκινώντας από την 3^η στήλη και εφαρμόζουμε την ίδια διαδικασία.

0	1	3
1	2	2
0	0	1

0*1	+	1*0	+	3*(-1)	=	-3
1*1	+	2*0	+	2*(-1)	=	-1
0*1	+	0*0	+	1*(-1)	=	-1

Η έξοδος στην συγκεκριμένη περίπτωση είναι **Output = -5**. Η συγκεκριμένη διαδικασία με την εφαρμογή του φίλτρου Kernel επαναλαμβάνεται σε κάθε pixel της εικόνας, που μπορεί φυσικά να εφαρμοστεί, δηλαδή για συνολικά 9 θέσεις (3X3 εξόδους). Συνεχίζοντας λοιπόν για κάθε πιθανό 3x3 φίλτρο, με τους ίδιους υπολογισμούς η τελική μας εικόνα είναι η εξής:

2	2	-5
4	2	-3
2	3	0

- **ReLU Function**

Μετά την όλη διαδικασία συνέλιξης (Convolution), θα εφαρμόσουμε την συνάρτηση **ReLU(x) = max(0,x)**. Η ReLU (Rectified Linear Unit) αντικαθιστά τις αρνητικές τιμές με το 0, ώστε να εισάγουμε μη γραμμικότητα στο μοντέλο. Εφαρμόζοντας τη ReLU στο τελευταίο map feature:

2	2	0
4	2	0
2	3	0

- **Max Pooling**

Για να απλοποιήσουμε περαιτέρω τα δεδομένα και να μειώσουμε τις διαστάσεις, εφαρμόζουμε Max Pooling με παράθυρο 2x2 και χρησιμοποιούμε το Kernel που είδαμε και σε προηγούμενο βήμα. Αυτό σημαίνει ότι θα πάρουμε τη μέγιστη τιμή από κάθε grid 2x2 του map feature.

First 2x2 area:

2	2
4	2

Second 2x2 area:

4	2
2	3

Τελική εικόνα εφαρμόζοντας την διαδικασία Max Pooling όπου επιλέγουμε την μέγιστη τιμή pixel σε κάθε σημείο του τρέχων πίνακα:

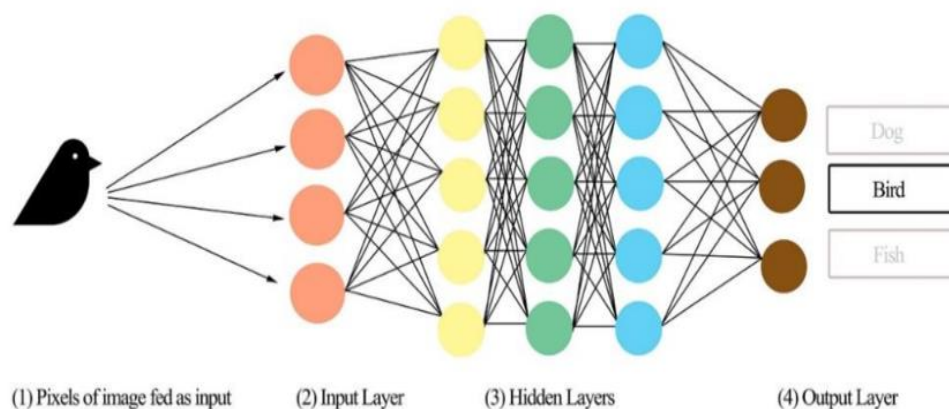
4	2
4	3

Τα αποτελέσματα του Max Pooling είναι όπως είδαμε σε μορφή πίνακα 2x2 διαστάσεων. Πριν σταλθεί το αποτέλεσμα στα Fully Connected Layers, θα πρέπει να μετατραπεί σε μονοδιάστατο πίνακα δεδομένων. Η διαδικασία αυτή ονομάζεται Flattening. Επομένως, σύμφωνα με το παραπάνω παράδειγμα, τα Fully Connected Layers θα λάβουν τον εξής μονοδιάστατο πίνακα:

4	2	4	3
---	---	---	---

Αυτός ο πίνακας εισέρχεται στα Fully Connected Layers, όπου χρησιμοποιείται για την τελική πρόβλεψη. Τα Fully Connected Layers χρησιμοποιούν την τιμή του κάθε pixel του πίνακα και παράγουν την τελική πρόβλεψη ως αποτέλεσμα.

Σε αυτό το υπό-κεφάλαιο, παρουσιάσαμε με απλά αριθμητικά παραδείγματα, πως λειτουργεί ένα Convolutional Neural Network, χρησιμοποιώντας μία απλή Grayscale 5X5 εικόνα. Εφαρμόσαμε την διαδικασία Convolution με φίλτρο 3X3(Kernel), έπειτα την συνάρτηση ReLU όπου οι αρνητικές τιμές μετατρέπονται σε μηδενικές, στην συνέχεια την διαδικασία υποδειγματοληψίας Max Pooling για την μείωση των διαστάσεων του πίνακα σε μορφή 2X2 και τέλος την μετατροπή του αποτελέσματος σε μονοδιάστατο πίνακα (Flattening). Αυτά είναι τα βασικά βήματα που χρησιμοποιούν τα CNN για να παράγουν την τελική πρόβλεψη. [\[29\]](#)[\[30\]](#)[\[31\]](#)[\[32\]](#)



Εικόνα 25 Object Identification

5.3 Ανάλυση μεθόδων των Νευρωνικών Δικτύων (Neural Networks)

5.3.1 SRCNN (Super-Resolution Convolutional Neural Network)

Ένα από τα πιο δημοφιλή νευρωνικά δίκτυα, που εφαρμόζεται για το κομμάτι της υπέρ-ανάλυσης εικόνας (super resolution), είναι το Super Resolution Convolutional Neural Network (SRCNN). Super Resolution είναι η διαδικασία κατά την οποία αυξάνεται η ανάλυση μιας εικόνας χαμηλής ανάλυσης για να δημιουργηθεί μια εικόνα υψηλότερης ανάλυσης, διατηρώντας τη λεπτομέρεια και την ποιότητα της αρχικής εικόνας. Το SRCNN αποτελείται από διαδοχικά συνελκτικά στρώματα που μαθαίνουν πώς να βελτιώνουν την ποιότητα της εικόνας και να ανακατασκευάζουν τις λεπτομέρειες που χάνονται όταν η εικόνα γίνεται downscale (συμπίεση). Το μοντέλο εκπαιδεύεται να μάθει μια αναπαράσταση της αρχικής εικόνας υψηλής ανάλυσης και να την αναπαράγει από την εικόνα χαμηλής ανάλυσης.

Η αρχιτεκτονική του SRCNN ξεκινάει με το Pre-Upscaling. Του δίνουμε μία εικόνα LR (Low resolution), την οποία επεξεργάζεται και μεγεθύνει με την bilinear interpolation μέθοδο, που αναλύσαμε και σε προηγούμενο κεφάλαιο. Έπειτα εφαρμόζεται ένα convolutional filter, για να εξάγει νέα χαρακτηριστικά από την LR εικόνα στην οποία έχει ήδη εφαρμοστεί το Pre-Upscaling. Έτσι το μοντέλο μας εκπαιδεύεται ή αλλιώς μαθαίνει, την πιο αντιπροσωπευτική μορφή της high resolution εικόνας HR. Η διαδικασία αυτή ονομάζεται Feature Extraction. Οι εξαγόμενες πληροφορίες περνούν από ένα Layer και συγκεκριμένα ένα Non-Linear Mapping Layer, το οποίο είναι υπεύθυνο για την συσχέτιση της Low Resolution εικόνας με την High Resolution εικόνα και στο τέλος γίνεται ανακατασκευή της τελικής HR εικόνας (Reconstruction), με βάση αυτή την συσχέτιση.

Για την περεταίρω κατανόηση και ανάλυση του SRCNN, θα δοκιμάσουμε ένα απλό αριθμητικό παράδειγμα. Θα χρησιμοποιήσουμε μια μικρή εικόνα και θα προσπαθήσουμε να την μετατρέψουμε από χαμηλή ανάλυση (LR) σε υψηλή ανάλυση (HR). Έστω ότι παίρνουμε μία LR εικόνα, όπου έχει διαστάσεις 3X3.

10	20	30
40	50	60
70	80	90

Στην συνέχεια πραγματοποιείται η διαδικασία Pre-scaling. Ουσιαστικά η αρχική μας εικόνα μεγεθύνεται, προσθέτοντας την ενδιάμεση τιμή των pixel. Η τεχνική αυτή γίνεται με βάση

την bilinear interpolation για να βρεθεί η ενδιάμεση τιμή. Το αποτέλεσμα στον νέο πίνακα διαστάσεων 5X5 είναι:

10	15	20	25	30
25	32	40	47	55
40	47	50	52	60
55	62	70	77	85
70	75	80	85	90

Όπως και με την βασική διαδικασία των CNN, για κάθε 3X3 κομμάτι της εικόνας θα εφαρμοστεί ένα kernel φίλτρο για να υπολογιστεί το γινόμενο του κάθε γειτονικού pixel και στην συνέχεια προσθέτουμε τα αποτελέσματα. Στην περίπτωση που το φίλτρο μας έχει την μορφή

0.2	0.5	0.2
0.5	1	0.5
0.2	0.5	0.2

Το αποτέλεσμα για την πρώτη δειγματοληψία της εικόνας που βρίσκεται μέσα στις πρώτες 3X3 διαστάσεις του πίνακα είναι το εξής:

$$\text{Output} = (0.2 \times 10) + (0.5 \times 15) + (0.2 \times 20) + (0.5 \times 25) + (1.0 \times 32) + (0.5 \times 40) + (0.2 \times 40) + (0.5 \times 47) + (0.2 \times 50) = 120.4$$

10	15	20
25	32	40
40	47	50

Η τιμή της νέας εικόνας στο συγκεκριμένο position είναι 120.4. Έπειτα συνεχίζεται η ίδια διαδικασία, με το ίδιο φίλτρο σε όλες τις συντεταγμένες του πίνακα. Αφού εφαρμοστεί το φίλτρο για όλα τα pixel της εικόνας, τότε εφαρμόζεται η συνάρτηση ReLU, όπως είδαμε και σε προηγούμενα κεφάλαια, η οποία αντικαθιστά ότι πιθανόν αρνητική τιμή υπάρχει με την τιμή μηδέν. Στο τέλος το τελικό layer της μεθόδου SRCNN παίρνει τιμές που προέκυψαν και ανακατασκευάζει για να παραχθεί η τελική High resolution εικόνα. Το αποτέλεσμα είναι μια εικόνα υψηλότερης ανάλυσης από την αρχική, με περισσότερες λεπτομέρειες και ομαλότερη

μετάβαση ανάμεσα στις τιμές των pixels. Το SRCNN χρησιμοποιεί τα Convolutional Filters και Non Linear Layers για να ενισχύσει την ποιότητα της εικόνας χαμηλής ανάλυσης και να παράγει μια πιο λεπτομερή έκδοση υψηλής ανάλυσης. [\[31\]\[32\]](#)

5.3.2 GAN (Generative Adversarial Networks)

Τα Generative Adversarial Networks (GANs) και τα Convolutional Neural Networks (CNNs) είναι και τα δύο ισχυρά εργαλεία στον τομέα του Machine Learning, και μάλιστα χρησιμοποιούνται συχνά μαζί, ειδικά σε εφαρμογές που αφορούν εικόνες. Η σχέση τους προκύπτει από την αρχιτεκτονική τους και τους τύπους προβλημάτων που λύνουν. Τα Generative Adversarial Networks εισαχθήκαν από τον ερευνητή Ian Goodfellow και από τότε έχουν διαμορφώσει σημαντικά τον τομέα της τεχνητής νοημοσύνης και του Deep Learning. Τα GANs έχουν εφαρμογές σε πολλούς τομείς, όπως η επεξεργασία και δημιουργία εικόνων, ο συνθετικός ήχος, καθώς και τα βίντεο. Η καινοτομία τους έγκειται στον τρόπο εκπαίδευσης και λειτουργίας τους, που βασίζεται στην «κρίση» μεταξύ δύο νευρωνικών δικτύων που ονομάζονται Generator και Discriminator.

Τα GANs αποτελούνται από δύο διακριτά νευρωνικά δίκτυα που αλληλοεπιδρούν δυναμικά. Η λειτουργία τους είναι βασισμένη σε έναν μηχανισμό όπου τα δύο δίκτυα μαθαίνουν να βελτιώνονται μέσω της αντιπαράθεσης και ανατροφοδότησης. Ο Generator είναι ένα νευρωνικό δίκτυο που έχει ως στόχο να δημιουργήσει δεδομένα, όπως εικόνες ή ήχους, που μοιάζουν με τα πραγματικά δεδομένα που υπάρχουν στο σύνολο εκπαίδευσης. Αρχικά, παίρνει τυχαίο θόρυβο ως είσοδο και τον «μετατρέπει» σε δείγματα δεδομένων. Στην αρχή της εκπαίδευσης, τα δείγματα αυτά είναι πολύ αδρά και σχεδόν καθόλου ρεαλιστικά. Όσο όμως το training συνεχίζεται, ο Generator προσαρμόζει τα δεδομένα του ώστε να γίνονται όλο και πιο πειστικά, προσπαθώντας να «ξεγελάσει» τον Discriminator.

Ο Discriminator, είναι το δίκτυο που μαθαίνει να διακρίνει αν τα δεδομένα που λαμβάνει είναι «πραγματικά», από την διαδικασία του training ή «ψεύτικα» από τον Generator. Στόχος του είναι να εντοπίζει με ακρίβεια ποιες εικόνες είναι πλαστές και ποιες όχι, βελτιώνοντας διαρκώς τις δυνατότητές του να εντοπίζει τον Generator. Κατά την διαδικασία του training του GAN, ο Discriminator ενημερώνεται συνεχώς ώστε να διακρίνει καλύτερα τα πλαστά δεδομένα, ενώ ο Generator προσπαθεί να παράγει δεδομένα που δεν θα μπορεί να διακρίνει ο Discriminator. Με την πάροδο του χρόνου, τα δύο δίκτυα συμβάλλουν, στο να εξελίσσονται συνεχώς, δημιουργώντας έναν ατελείωτο ανταγωνισμό, έως ότου ο Generator φτάσει σε σημείο να δημιουργεί δεδομένα που είναι δύσκολο να διακριθούν από τα πραγματικά.

Το training πραγματοποιείται μέσω αλγορίθμων βελτιστοποίησης, όπως το stochastic gradient descent (SGD). Το Stochastic Gradient Descent (SGD), ή Στοχαστική Βαθμιδωτή Κατάβαση, είναι ένας αλγόριθμος βελτιστοποίησης που χρησιμοποιείται ευρέως στον τομέα του machine learning και στα neural networks. Είναι μια μέθοδος για τη ρύθμιση των παραμέτρων ενός μοντέλου, έτσι ώστε να ελαχιστοποιηθεί η πιθανότητα σφάλματος πρόβλεψης, η οποία μετράει το πόσο μακριά βρίσκεται η πρόβλεψη του μοντέλου από τις πραγματικές τιμές. Λόγω της επεξεργασίας μικρών υποσυνόλων, το SGD είναι πολύ πιο γρήγορο από την παραδοσιακή βαθμιδωτή κατάβαση, ιδιαίτερα σε μεγάλα σύνολα δεδομένων. Είναι ένας σχετικά απλός αλγόριθμος που χρησιμοποιείται εύκολα σε διάφορα μοντέλα και αρχιτεκτονικές νευρωνικών δικτύων.

Τα GANs χρησιμοποιούνται για τη δημιουργία συνθετικών εικόνων προσώπων, αντικειμένων ή σκηνών που μοιάζουν απόλυτα ρεαλιστικές. Χρησιμοποιούνται για την αφαίρεση θορύβου, την αποκατάσταση παλιών ή κατεστραμμένων εικόνων, καθώς και για την ανακατασκευή σκηνών με βάση ελλιπή δεδομένα. Επίσης εξαιρετικά ενδιαφέρον είναι ότι χρησιμοποιούνται και σε συστήματα που απαιτούν ικανότητα κατανόησης και δημιουργίας, όπως η αναγνώριση προσώπου και η σύνθεση ανθρώπινων χαρακτηριστικών σε φωνητικά και γραπτά δεδομένα. Βέβαια, παρά την αποτελεσματικότητά τους, τα GANs έρχονται με ορισμένες προκλήσεις και περιορισμούς. Ένας από τους σημαντικότερους περιορισμούς είναι η ασταθής εκπαίδευση, καθώς είναι δύσκολο να διατηρηθεί μια ισορροπία μεταξύ του Generator και του Discriminator. Εάν η ισορροπία διαταραχθεί, μπορεί να οδηγήσει σε μη ρεαλιστικά δεδομένα ή σε αναποτελεσματική εκπαίδευση. Επίσης, το training των GANs είναι αρκετά απαιτητικό σε πόρους, απαιτώντας μεγάλο αριθμό δεδομένων και υπολογιστική ισχύ για να επιτευχθούν τα επιθυμητά αποτελέσματα. Τα GANs αποτελούν μια από τις σημαντικότερες εξελίξεις στην τεχνητή νοημοσύνη, ανοίγοντας νέους ορίζοντες στη δημιουργία και βελτίωση δεδομένων, εικόνων και ήχου. Ο μηχανισμός ανταγωνισμού μεταξύ του Generator και του Discriminator τα καθιστά εξαιρετικά αποτελεσματικά στη δημιουργία δεδομένων υψηλής ποιότητας. Τα Gans χρησιμοποιούν την τεχνική Binary Cross-Entropy Loss, για την εκπαίδευση (training) των Generator και Discriminator. Η Binary Cross-Entropy Loss (BCE Loss) μετρά τη διαφορά μεταξύ των πραγματικών τιμών (0 ή 1) και των προβλέψεων (πιθανότητες μεταξύ 0 και 1). Η μαθηματική της έκφραση είναι:

$$L(y, f(x)) = -[y * \log(f(x)) + (1 - y) * \log(1 - f(x))]$$

Όπου y είναι η πραγματική δυαδική ετικέτα (Label) 0 ή 1 και $f(x)$ είναι η προβλεπόμενη πιθανότητα της κατηγορίας μεταξύ 0 ή 1.

Στα GANs, η BCE Loss χρησιμοποιείται διαφορετικά για του Discriminator και του Generator. Ο Discriminator προσπαθεί να διακρίνει τις πραγματικές εικόνες από τις συνθετικές που δημιουργεί ο Generator. Για πραγματικές εικόνες, η επιθυμητή ετικέτα είναι 1 και για συνθετικές εικόνες, η επιθυμητή ετικέτα είναι 0. Αυτό σημαίνει πως ο κύριος σκοπός του Generator είναι να ξεγελάσει τον Discriminator πως η πρόβλεψη της τιμής της ετικέτας της συνθετικής εικόνας είναι στην κατηγορία 0.

Το πλεονέκτημα της BCE Loss είναι πως είναι ιδανική για προβλήματα όπου οι προβλέψεις είναι πιθανότητες μεταξύ 0 και 1 και πως οι υπολογισμοί των εξισώσεων είναι εύκολη να υλοποιηθούν. Το βασικό πρόβλημα με την BCE είναι όπως βλέπουμε πάλι, η πιθανότητα που ο Discriminator είναι πολύ καλός και οι προβλέψεις του είναι σχεδόν 0 ή 1. Αυτό έχει ως αποτέλεσμα η BCE Loss να παράγει πολύ μικρές τιμές απώλειας, δυσκολεύοντας την εκπαίδευση του Generator και το αποτέλεσμα να μην είναι ακριβές. Η Binary Cross-Entropy Loss αποτελεί βασικό συστατικό για την εκπαίδευση των GAN, καθώς διαχειρίζεται τον ανταγωνισμό μεταξύ Generator και Discriminator. Ωστόσο, η χρήση της απαιτεί προσεκτική ρύθμιση των παραμέτρων του μοντέλου για να αποφευχθούν προβλήματα όπως η αστάθεια και η αργή εκπαίδευση. [\[34\]](#)[\[35\]](#)

5.3.3 SRGAN (Super-Resolution Generative Adversarial Network)

Τα Super-Resolution Generative Adversarial Networks (SRGANs) είναι μία από τις πιο σημαντικές και πρωτοποριακές εφαρμογές των Generative Adversarial Networks (GANs) στον τομέα βελτίωσης και επεξεργασίας εικόνας. Εισήχθησαν το 2017 και ο κύριος στόχος τους είναι η παραγωγή εικόνων υψηλής ανάλυσης (High-Resolution ή HR) από εικόνες χαμηλής ανάλυσης (Low-Resolution ή LR). Αυτό επιτυγχάνεται μέσω μιας μοναδικής αρχιτεκτονικής που συνδυάζει τον ανταγωνισμό μεταξύ δύο νευρωνικών δικτύων με την ικανότητα να συλλαμβάνει και να βελτιώνει λεπτομέρειες εικόνας. Τα SRGANs θεωρούνται ένα από τα πιο ισχυρά εργαλεία για αύξηση ανάλυσης εικόνων (image super-resolution), το οποίο βρίσκει εφαρμογές σε τομείς όπως η ιατρική απεικόνιση, η δορυφορική φωτογραφία, η επεξεργασία βίντεο και οι σύγχρονες οπτικές τεχνολογίες.

Ο σχεδιασμός των Super Resolution Generative Adversarial Networks, δημιουργήθηκε με σκοπό να καλυφθούν περιορισμοί, όπως οι κάμερες χαμηλής ανάλυσης όπου οι λεπτομέρειες δεν είναι αρκετά διακριτές και μπορεί να χαθούν, οι δορυφορικές εικόνες όπου η ανάλυση περιορίζεται από την απόσταση και την τεχνολογία του εξοπλισμού, αλλά ακόμα και για συστήματα παρακολούθησης (CCTV), όπου συχνά απαιτείται να βελτιωθεί η ανάλυση

προσώπων και αντικειμένων. Είναι προφανώς βασισμένα στην αρχιτεκτονική των απλών Generative Adversarial Networks, καθώς, κατά τον υπολογισμό, χρησιμοποιείται το ίδιο σύστημα αντιπαλότητας μεταξύ Generator και Discriminator. Μία από τις σημαντικότερες διαφορές μεταξύ των απλών GANs και των SRGANs βρίσκεται στις τεχνικές απώλειας δεδομένων (Loss Functions). Τα Gans χρησιμοποιούν την τεχνική Binary Cross-Entropy Loss, για την εκπαίδευση (training) των Generator και Discriminator, που αναφέραμε και στο προηγούμενο κεφάλαιο. Σε αντίθεση τα SRGAN χρησιμοποιούν διαφορετικές μεθόδους εκπαίδευσης των Generator και Discriminator.

Αρχικά τα SRGAN χρησιμοποιούν τα Residual Blocks. Η βασική ιδέα πίσω από τα Residual Blocks είναι η προστιθέμενη σύνδεση ή το residual connection. Αντί να μεταβιβάζεται η έξοδος ενός επιμέρους στρώματος απευθείας στο επόμενο στρώμα, η έξοδος αυτού του στρώματος προστίθεται στην αρχική είσοδο του μπλοκ. Αυτή η προσθήκη βοηθάει στο να παρακαμφθεί η ανάγκη για τα δίκτυα να μάθουν τη μετατόπιση από το μηδέν και δίνει τη δυνατότητα στο δίκτυο να μάθει πιο εύκολα τη "διαφορά" μεταξύ των χαρακτηριστικών, βελτιώνοντας την απόδοση στην εκπαίδευση. Στα SRGAN, τα Residual Blocks βοηθούν τον Generator να ανακτήσει τις σημαντικές λεπτομέρειες από την εικόνα υψηλής ανάλυσης (HR) που δεν είναι εύκολα ανακτήσιμες από τα παραδοσιακά δίκτυα, μειώνοντας τις απώλειες στην ποιότητα κατά τη διάρκεια της αναβάθμισης της ανάλυσης. Αυτή η προσθήκη της "αρχικής εισόδου" στον υπολογισμό της εξόδου αποτρέπει την απώλεια σημαντικών χαρακτηριστικών που μπορεί να προκύψει λόγω της βαθιάς δομής του δικτύου, κάνοντάς το πιο αποτελεσματικό στην εξαγωγή και ανασύνθεση λεπτομερειών εικόνας. Ας υποθέσουμε ότι έχουμε μια είσοδο x και θέλουμε να μάθουμε την έξοδο $F(x)$ ενός μπλοκ. Σε ένα **Residual Block**, η έξοδος υπολογίζεται ως **Output = $F(x) + x$** . Η πρόσθεση $F(x) + x$ βοηθάει στην εκμάθηση της υπολειπόμενης διαφοράς μεταξύ της αρχικής εικόνας και της επιθυμητής εξόδου, κάτι που κάνει τη διαδικασία εκπαίδευσης πιο αποδοτική.

Ένας βασικός τρόπος υπολογισμού της απώλειας στα Neural Networks είναι μέσω της Mean Squared Error (MSE), που μετρά τη διαφορά μεταξύ των pixel της παραγόμενης και της πραγματικής εικόνας. Ωστόσο, η MSE συνήθως δεν είναι ικανή να συλλάβει τις πιο αφηρημένες οπτικές και αντιληπτικές πληροφορίες, όπως υφή, σχήματα ή χαρακτηριστικά υψηλού επιπέδου που επηρεάζουν την ανθρώπινη αντίληψη της ποιότητας μιας εικόνας. Τα SRGAN Χρησιμοποιούν την Perceptual Loss. Η Perceptual Loss αποσκοπεί στο να μετρήσει την απώλεια μεταξύ των εικόνων με βάση τα χαρακτηριστικά που εξάγονται από τα υψηλότερα επίπεδα ενός pre-trained Convolutional Neural Network (CNN), όπως το VGG-19.

Το VGG-19 χρησιμοποιείται ευρέως ως pre-trained δίκτυο λόγω της ικανότητάς του να εξάγει ισχυρά χαρακτηριστικά εικόνας.

Το VGG-19 είναι pre-trained στο ImageNet, το οποίο είναι ένα μεγάλο σύνολο δεδομένων εικόνας που περιλαμβάνει 1000 κατηγορίες. Αυτό επιτρέπει στους χρήστες να εκμεταλλευτούν το pre-training (transfer learning) για διάφορες εφαρμογές. Δηλαδή, μπορεί να χρησιμεύσει ως feature extractor για άλλες εργασίες υπολογιστικής όρασης, όπως η αναγνώριση αντικειμένων, η ανίχνευση προσώπων, ή η ανάλυση σκηνών. Η απλή και αποτελεσματική αρχιτεκτονική του, με τη χρήση μικρών φίλτρων και πολλαπλών στρωμάτων, το καθιστά εύκολο στην κατανόηση και την εφαρμογή σε διάφορες εργασίες. Αν και έχει υψηλές απαιτήσεις σε υπολογιστική ισχύ, παραμένει ένα από τα πιο ισχυρά εργαλεία για την ανάλυση εικόνας.

Το VGG-19 χρησιμοποιεί 16 convolution layers για την εξαγωγή χαρακτηριστικών από τις εικόνες. Κάθε convolution χρησιμοποιεί φίλτρα 3x3, τα οποία βοηθούν στην ανίχνευση τοπικών χαρακτηριστικών όπως άκρα, γωνίες, και υφές. Όλα τα convolutional layers στο VGG-19 είναι ενεργοποιημένα με τη συνάρτηση ReLU, η οποία εισάγει μη γραμμικότητα και επιτρέπει στο δίκτυο να μάθει πιο σύνθετα χαρακτηριστικά. Τα Max-Pooling layers χρησιμοποιούνται μετά από κάθε δύο ή τρία convolutional layers για να μειώσουν τις διαστάσεις των χαρακτηριστικών (down-sampling), ενώ ταυτόχρονα διατηρούν τα σημαντικότερα χαρακτηριστικά της εικόνας. Αυτό γίνεται με τη χρήση ενός φίλτρου 2x2 που εφαρμόζεται σε κάθε επιμέρους χαρακτηριστικό, με βήμα 2 (Stride=2). Η διαδικασία αυτή μειώνει την ανάλυση των χαρακτηριστικών ενώ ενισχύει την υπολογιστική αποδοτικότητα. Στο τέλος το VGG-19 διαθέτει τρία πλήρως συνδεδεμένα στρώματα, τα οποία επιτρέπουν στο δίκτυο να κάνει την τελική πρόβλεψη ή ταξινόμηση. Τα δύο πρώτα στρώματα είναι 4096 νευρώνες το καθένα και το τελευταίο στρώμα έχει μέγεθος ανάλογο με τον αριθμό των κλάσεων που πρόκειται να ταξινομηθούν (για παράδειγμα, 1000 κλάσεις για το ImageNet). Κάθε ένα από τα πλήρως συνδεδεμένα στρώματα χρησιμοποιεί επίσης τη συνάρτηση ReLU και το τελευταίο στρώμα χρησιμοποιεί τη συνάρτηση Softmax για την παραγωγή πιθανοτήτων κατηγοριών.

Η ιδέα πίσω από την Perceptual Loss είναι ότι, αντί να συγκρίνουμε τις εικόνες pixel προς pixel, συγκρίνουμε τα χαρακτηριστικά τους που εξάγονται από ένα pre-trained δίκτυο. Για παράδειγμα, μπορούμε να συγκρίνουμε τα χαρακτηριστικά από τα πρώτα επίπεδα του VGG-19, τα οποία αντιστοιχούν σε απλά χαρακτηριστικά όπως άκρα, γραμμές και υφές, ή από τα ανώτερα επίπεδα του δικτύου που αναπαριστούν πιο αφηρημένα χαρακτηριστικά όπως

αντικείμενα ή μορφές. Η Perceptual Loss υπολογίζεται με βάση τη διαφορά των χαρακτηριστικών μεταξύ της παραγόμενης εικόνας και της πραγματικής εικόνας υψηλής ανάλυσης.

Τα GANs έχουν γενικό σκοπό τη δημιουργία συνθετικών δεδομένων που μοιάζουν με πραγματικά. Χρησιμοποιούνται για την παραγωγή εικόνων, κειμένων, βίντεο, ήχου και άλλων μορφών δεδομένων. Για παράδειγμα, μπορούν να δημιουργήσουν ρεαλιστικά πορτρέτα ανθρώπων που δεν υπάρχουν ή να συνθέσουν μουσική. Τα SRGANs εξειδικεύονται στη βελτίωση της ανάλυσης εικόνων. Ο κύριος στόχος τους είναι η παραγωγή εικόνων υψηλής ανάλυσης (High-Resolution ή HR) από αντίστοιχες εικόνες χαμηλής ανάλυσης (Low Resolution ή LR), εστιάζοντας στη διατήρηση της λεπτομέρειας και της ποιότητας. Οι εφαρμογές των SRGANs επικεντρώνονται αποκλειστικά στην ενίσχυση της ανάλυσης εικόνων και περιλαμβάνουν, ενίσχυση παλιών ή θολών φωτογραφιών, βελτίωση ανάλυσης εικόνων σε δορυφορική φωτογραφία, βελτίωση της ποιότητας βίντεο σε streaming πλατφόρμες, χρήση στην ιατρική απεικόνιση για καλύτερη ανάλυση σε MRI και CT scans και επαναφορά θολών ή κατεστραμμένων εικόνων. Στα SRGANs τα Residual Blocks και η Perceptual Loss συνεργάζονται για να παράγουν πιο ρεαλιστικές και φυσικές εικόνες υψηλής ανάλυσης. Τα Residual Blocks επιτρέπουν στον Generator να ανακτήσει καλύτερα τα χαρακτηριστικά της εικόνας χωρίς να διαταράξει τις υπάρχουσες πληροφορίες και η Perceptual Loss διασφαλίζει ότι η παραγόμενη εικόνα θα είναι αντιληπτικά όμοια με την πραγματική εικόνα υψηλής ανάλυσης, ακόμη και αν δεν είναι απολύτως ακριβής σε επίπεδο pixel. Αυτή η συνεργασία επιτρέπει στα SRGAN να βελτιώνουν την ποιότητα εικόνας με ρεαλιστικές λεπτομέρειες και υφή, διατηρώντας τις αισθητικές και αντιληπτικές πτυχές της εικόνας. [\[36\]](#)[\[37\]](#)[\[38\]](#)

5.3.4 ESRGAN (Enhanced SRGAN)

Τα ESRGAN (Enhanced Super-Resolution Generative Adversarial Networks) είναι μία εξελιγμένη μορφή των Super-Resolution Generative Adversarial Networks, που χρησιμοποιείται για την αναβάθμιση της ανάλυσης εικόνας. Η βασική διαφορά τους από τα παραδοσιακά SRGANs είναι η προσθήκη καινοτόμων βελτιώσεων που καθιστούν το ESRGAN πιο ικανό στην παραγωγή υψηλής ποιότητας, ρεαλιστικών εικόνων υψηλής ανάλυσης από εικόνες χαμηλής ανάλυσης. Το ESRGAN ενσωματώνει νέες τεχνικές και βελτιώσεις τόσο στη δομή του δικτύου όσο και στον τρόπο εκπαίδευσης, επιτυγχάνοντας σημαντική πρόοδο σε σύγκριση με τα παραδοσιακά SRGANs. Το ESRGAN είναι ένα τύπο Generative Adversarial Network (GAN) που χρησιμοποιεί την ιδέα των Residual Dense Blocks για να βελτιώσει την

ποιότητα της παραγόμενης εικόνας και την ταχύτητα της εκπαίδευσης. Όπως και τα παραδοσιακά SRGAN, το ESRGAN περιλαμβάνει δύο βασικά μέρη: τον Γεννήτορα Generator και τον Discriminator.

Στα ESRGAN, οι Residual Dense Blocks παίζουν καθοριστικό ρόλο στην παραγωγή εικόνας. Τα Dense Blocks είναι μια άλλη δημοφιλής τεχνική που χρησιμοποιείται σε σύγχρονα δίκτυα όπως το DenseNet (Densely Connected Convolutional Networks). Σε αντίθεση με τα παραδοσιακά convolutional δίκτυα, όπου κάθε επίπεδο δέχεται είσοδο μόνο από το προηγούμενο επίπεδο, σε ένα Dense Block, κάθε επίπεδο δέχεται είσοδο από όλα τα προηγούμενα επίπεδα του μπλοκ. Αυτή η τεχνική ενθαρρύνει την πλουσιότερη ροή πληροφορίας και επιτρέπει στο δίκτυο να μάθει πιο ισχυρά χαρακτηριστικά από τα δεδομένα. Ειδικότερα, το κάθε επίπεδο του Dense Block μπορεί να αξιοποιήσει τις πληροφορίες από τα προηγούμενα επίπεδα, επιτρέποντας μια πιο αποτελεσματική μάθηση και αυξάνοντας την ικανότητα του δικτύου να εξάγει και να κατανοήσει τα χαρακτηριστικά των δεδομένων.

Το Residual dense Blocks (RDB) συνδυάζει τα πλεονεκτήματα των Residual Blocks και των Dense Blocks για να δημιουργήσει μια πιο αποδοτική και ισχυρή μονάδα μάθησης. Ο συνδυασμός αυτών των δύο τεχνικών βοηθά το δίκτυο να μάθει καλύτερα τα χαρακτηριστικά των εικόνων με έναν πιο αποδοτικό και εύκολο τρόπο. Στο πρώτο Layer, το πρώτο Residual Block αναλύει την είσοδο, μαθαίνοντας τις βασικές πληροφορίες και τα χαρακτηριστικά της. Στη συνέχεια, η έξοδος του Residual Block εισέρχεται σε μια σειρά από Dense Blocks, τα οποία επιτρέπουν στο δίκτυο να αξιοποιήσει πληροφορίες από τα προηγούμενα επίπεδα του μπλοκ και να εξάγει πιο σύνθετα χαρακτηριστικά. Μετά την επεξεργασία στα Dense Blocks, το δίκτυο προσθέτει τις εξόδους αυτών των μπλοκ στην αρχική είσοδο μέσω μιας “βραχυκύκλωσης” (skip connection), ώστε το αποτέλεσμα να ενσωματώνει τόσο τις πρωτότυπες όσο και τις νέες πληροφορίες που έχουν εξαχθεί. Η διαδικασία αυτή ονομάζεται Residual Connection. Τα Residual Dense Blocks (RDBs) είναι μια άκρως σημαντική καινοτομία στην τεχνολογία των νευρωνικών δικτύων, ειδικά σε εφαρμογές όπως η υψηλής ανάλυσης αναβάθμιση εικόνας. Συνδυάζοντας τα πλεονεκτήματα των Residual Blocks και των Dense Blocks, τα RDBs βοηθούν στη δημιουργία πιο ισχυρών, πιο αποδοτικών και πιο ακριβών μοντέλων για την παραγωγή ρεαλιστικών εικόνων. Αυτή η αρχιτεκτονική έχει κάνει τεράστια πρόοδο στις επιδόσεις μοντέλων όπως το ESRGAN, κάνοντάς τα πιο αποτελεσματικά στην αναγνώριση, μάθηση και αναβάθμιση των εικόνων με μεγαλύτερη ακρίβεια.

Όπως και τα SRGANs, τα ESRGAN χρησιμοποιούν την Perceptual Loss για να υπολογίσουν την απώλεια μεταξύ της παραγόμενης και της πραγματικής εικόνας. Η βασική καινοτομία στο ESRGAN είναι ότι χρησιμοποιεί μια πιο εξελιγμένη μορφή αυτής της απώλειας, που επιτρέπει στο δίκτυο να διατηρήσει περισσότερες αντιληπτικές λεπτομέρειες και υφές, παρά την ελαχιστοποίηση της απώλειας στο επίπεδο pixel. Η Perceptual Loss υπολογίζεται χρησιμοποιώντας pre-trained δίκτυα CNN όπως το VGG, με σκοπό την ανίχνευση αφηρημένων χαρακτηριστικών που είναι σημαντικά για την ανθρώπινη αντίληψη. Ένα άλλο σημαντικό χαρακτηριστικό του ESRGAN είναι η βελτίωση στην εκπαίδευση του διακριτή. Αντί να χρησιμοποιεί τον παραδοσιακό τρόπο της binary cross-entropy loss για την εκπαίδευση του Discriminator, το ESRGAN εισάγει μια καινοτόμο τεχνική γνωστή ως Relativistic Average GAN. Η τεχνική αυτή αξιοποιεί τη σχετική πραγματικότητα μεταξύ των πραγματικών και των παραγόμενων εικόνων. Πιο συγκεκριμένα, ο Discriminator είναι αυτός ο οποίος χρησιμοποιείται η Relativistic Average GAN (RaGAN), η οποία υπολογίζει τη σχετική αλήθεια μεταξύ της πραγματικής και της παραγόμενης εικόνας, επιτρέποντας μια πιο ισχυρή εκπαίδευση και καλύτερη ποιότητα εικόνας. Στο RaGAN, αντί να χρησιμοποιείται η παραδοσιακή μέθοδος binary cross-entropy loss (όπου ο Discriminator προσπαθεί να κατηγοριοποιήσει τις εικόνες ως "πραγματικές" ή "ψεύτικες"), εισάγεται μια νέα μεθοδολογία, η οποία ενσωματώνει τις σχετικές (Relativistic) έννοιες στην απώλεια του διακριτή.

Τα αποτελέσματα του ESRGAN δείχνουν σημαντική βελτίωση σε σχέση με το SRGAN. Ο τρόπος που λειτουργεί το ESRGAN επιτρέπει την παραγωγή εικόνων που έχουν λιγότερους θορύβους (Artifacts) και πιο φυσικές λεπτομέρειες. Η χρήση των Residual Dense Blocks βοηθά στην καλύτερη ενσωμάτωση των χαρακτηριστικών της εικόνας, ενώ η χρήση του Relativistic Average κάνει τον διακριτή να γίνεται πιο ακριβής και ικανός να εντοπίζει μικρές διαφορές μεταξύ των πραγματικών και ψεύτικων εικόνων. Επίσης, η χρήση της Perceptual Loss βελτιώνει τη συνολική ποιότητα της εικόνας, αφού οι εικόνες που παράγονται μοιάζουν πιο κοντά στην αντίληψη του ανθρώπινου ματιού για τις υφές και τα σχήματα. Το ESRGAN έχει πολλές χρήσεις στον τομέα του super resolution, ειδικά στις περιπτώσεις όπου είναι απαραίτητη η παραγωγή εικόνας υψηλής ανάλυσης από εικόνες χαμηλής ανάλυσης. Ενδεικτικά, το ESRGAN χρησιμοποιείται για την αναβάθμιση εικόνας από ιατρικές σαρώσεις, όπως ακτινογραφίες ή MRI, για καλύτερη διάγνωση, για την αναβάθμιση δορυφορικών εικόνων σε υψηλότερη ανάλυση, ώστε να διακρίνονται καλύτερα οι λεπτομέρειες του εδάφους καθώς και για την αναβάθμιση της ποιότητας εικόνας σε βίντεο και κινούμενα σχέδια. Το ESRGAN είναι ένα ισχυρό εργαλείο για την παραγωγή εικόνας υψηλής ανάλυσης

από εικόνες χαμηλής ανάλυσης, με σημαντική βελτίωση σε σχέση με τα προηγούμενα μοντέλα όπως το SRGAN. Με τη χρήση τεχνικών όπως τα Residual Dense Blocks, η Perceptual Loss, και το Relativistic Average GAN, το ESRGAN παρέχει εικόνες με καλύτερη ποιότητα, λιγότερο θόρυβο και πιο φυσικές λεπτομέρειες. Οι βελτιώσεις αυτές το καθιστούν ιδανικό για εφαρμογές που απαιτούν super resolution εικόνας σε πραγματικές συνθήκες. Το συγκεκριμένο μοντέλο θα το μελετήσουμε και αναλύσουμε περισσότερο στο επόμενο κεφάλαιο με παράδειγμα σε μορφή πειράματος. [\[39\]\[40\]](#)

5.3.5 Αναφορά επιπλέον μεθόδων Image Upscaling

Φυσικά υπάρχουν πολλές CNN μέθοδοι και μοντέλα όσον αφορά τον τομέα του Image Upscaling, για τα οποία θα μιλήσουμε αναφορικά. Αρχικά, υπάρχει η μέθοδος LapSRN (Laplacian Pyramid Super-Resolution Network). Το LapSRN λειτουργεί προοδευτικά, ανακατασκευάζοντας την εικόνα υψηλής ανάλυσης μέσω ενός συστήματος πολλαπλών επιπέδων (pyramid). Αντί να αναβαθμίζει την εικόνα σε ένα μόνο βήμα, όπως γίνεται σε άλλες μεθόδους, το LapSRN βελτιώνει την εικόνα σε διάφορες κλίμακες. Η μέθοδος αναβαθμίζει την εικόνα σε μικρά βήματα. Σε κάθε επίπεδο της πυραμίδας, η εικόνα αναβαθμίζεται και προστίθενται λεπτομέρειες (residuals) για να προσεγγίσει καλύτερα την πραγματική εικόνα υψηλής ανάλυσης. Το LapSRN εφαρμόζει ξεχωριστή λειτουργία απώλειας (loss function) σε κάθε επίπεδο της πυραμίδας. Αυτό διασφαλίζει ότι το δίκτυο μαθαίνει να βελτιώνει τις λεπτομέρειες σε κάθε ανάλυση. Ουσιαστικά Αντί να προσπαθεί να ανακατασκευάσει ολόκληρη την εικόνα, το LapSRN μαθαίνει τις λεπτομέρειες που λείπουν (high frequency details) και τις προσθέτει στην αναβαθμισμένη εικόνα. Σε κάθε επίπεδο της πυραμίδας, η εικόνα αναβαθμίζεται χρησιμοποιώντας μεθόδους όπως η διαπλοκή (bicubic interpolation) ή οι αντιστροφες συνελίξεις (transposed convolution). Το μοντέλο προβλέπει τα υπολείμματα (residuals), δηλαδή τις λεπτομέρειες που πρέπει να προστεθούν στην αναβαθμισμένη εικόνα. Οι προβλέψεις υπολειμμάτων προστίθενται στην αναβαθμισμένη εικόνα, και τα στρώματα αυτά διασφαλίζουν ότι το τελικό αποτέλεσμα είναι υψηλής ποιότητας. Παρόλο που το LapSRN προσφέρει πολλά πλεονεκτήματα, έχει ορισμένους περιορισμούς. Μπορεί να δημιουργήσει τεχνητά αντικείμενα (artifacts) αν τα υπολείμματα δεν προβλέπονται σωστά. Δεν έχει την ίδια αντιληπτική ποιότητα (perceptual quality) με πιο σύγχρονες μεθόδους, όπως τα SRGAN ή ESRGAN.

Έπειτα ένα ακόμα γνωστό μοντέλο είναι το Deeply Recursive Convolution Network (DCRN). Το DCRN είναι ένα δίκτυο deep learning σχεδιασμένο για την ανάπτυξη και την επίλυση των

ζητημάτων όσον αφορά το super resolution των εικόνων. Βασίζεται σε μια καινοτόμα αρχιτεκτονική που χρησιμοποιεί την έννοια της αναδρομής (recursion), μειώνοντας την πολυπλοκότητα του μοντέλου και βελτιώνοντας την αποτελεσματικότητα της μάθησης. Το DCRN χρησιμοποιείται για τη μετατροπή εικόνων χαμηλής ανάλυσης σε εικόνες υψηλής ανάλυσης, ενισχύοντας τις λεπτομέρειες και τις υφές τους. Το πρώτο μέρος του μοντέλου ονομάζεται Feature Extraction, όπου λαμβάνει συγκεκριμένα χαρακτηριστικά από την εικόνα που παίρνει ως input. Το δίκτυο περιλαμβάνει ένα μικρό αριθμό convolution στρώματων, τα οποία εφαρμόζονται επαναληπτικά. Με αυτόν τον τρόπο, τα ίδια στρώματα εκπαιδεύονται να μαθαίνουν πιο σύνθετες αναπαραστάσεις της εικόνας σε κάθε βήμα. Είναι τα γνωστά Recursive Layers. Η αναδρομική χρήση μειώνει το συνολικό μέγεθος του μοντέλου χωρίς να μειώνει την απόδοση. Μετά την επεξεργασία των χαρακτηριστικών από τα recursive layers, το δίκτυο ανακατασκευάζει την εικόνα σε High Resolution μέσω ενός τελικού layer. Μπορεί να μην καταφέρει να αποδώσει τόσο καλά όσο πιο σύγχρονες, πιο σύνθετες μέθοδοι, όπως τα GANs ή τα ESRGANs, σε περιπτώσεις όπου απαιτείται εξαιρετική ποιότητα εικόνας, λόγω του ότι η loss function του είναι η Mean Squared Error που είδαμε και σε προηγούμενο κεφάλαιο στα GANs και προφανώς λόγω της αναδρομικής φύσης του.

Τέλος ιδιαίτερα αξιοσημείωτο είναι και το δίκτυο ScuNet (Spatially Conditioned U-Net). Το ScuNet είναι μια πρωτοποριακή αρχιτεκτονική deep learning που έχει σχεδιαστεί για την ανάκτηση και επεξεργασία εικόνας. Βασίζεται στο δημοφιλές μοντέλο U-Net, αλλά εισάγει χωρικά εξαρτημένες (spatially conditioned) πληροφορίες, καθιστώντας το ιδιαίτερα αποδοτικό σε προβλήματα όπου η χωρική μεταβλητότητα παίζει σημαντικό ρόλο. Το ScuNet έχει εφαρμοστεί σε διάφορα πεδία, όπως η αφαίρεση θορύβου (denoising), η από-θόλωση (deblurring), και η ανάλυση (super-resolution) σε εικόνες και βίντεο. Τα αρχικά στρώματα του δικτύου αναλύουν την είσοδο και εξάγουν χαρακτηριστικά χαμηλού επιπέδου από την εικόνα. Χρησιμοποιεί έναν Encoder, που μειώνει την ανάλυση της εικόνας ενώ διατηρεί τις σημαντικές πληροφορίες. Εδώ, χρησιμοποιούνται convolutional layers και functions ενεργοποίησης για τη συμπίεση των χαρακτηριστικών. Το επόμενο βήμα είναι μοναδικό για το συγκεκριμένο δίκτυο και ονομάζεται Spatial Conditioning. Εισάγεται ένας μηχανισμός που ενσωματώνει χωρικά εξαρτημένες πληροφορίες. Αυτό μπορεί να γίνει με τη χρήση επιπλέον καναλιών που περιγράφουν τη θέση κάθε pixel ή άλλες μεταβλητές που σχετίζονται με τη γεωμετρία της εικόνας. Για την ανακατασκευή της εικόνας χρησιμοποιείται ένας Decoder, που αυξάνει την ανάλυση της εικόνας, ανακατασκευάζοντας την αρχική δομή από τα συμπιεσμένα χαρακτηριστικά. Εδώ χρησιμοποιούνται λειτουργίες upsampling και συνδυάζονται τα χαρακτηριστικά χαμηλής και υψηλής ανάλυσης. Όπως και στο U-Net, έτσι

και το ScuNet περιλαμβάνει συνδέσεις παράκαμψης μεταξύ του κωδικοποιητή και του αποκωδικοποιητή (skip connections), που διατηρούν τις λεπτομέρειες της εικόνας κατά τη διάρκεια της διαδικασίας επεξεργασίας. Το τελευταίο στρώμα που παρέχει την τελική ανακατασκευασμένη εικόνα, είτε πρόκειται για καθαρή εικόνα (Denoised) είτε για εικόνα υψηλής ανάλυσης.

Όσον αφορά το τελικό κομμάτι της εργασίας, αποφασίσαμε να δώσουμε μεγαλύτερη βάση στο δίκτυο ESRGAN, αντί του SCUNET. Η επιλογή μεταξύ ScuNet και ESRGAN εξαρτάται από τις ανάγκες και τον στόχο του εκάστοτε έργου που αφορά την επεξεργασία εικόνας. Το SCUNET είναι ένα νεότερο δίκτυο που έχει σχεδιαστεί για γενικές εργασίες ανάκτησης εικόνας. Ενσωματώνει χωρικές πληροφορίες, προσαρμόζοντας την επεξεργασία στις ιδιαιτερότητες κάθε περιοχής της εικόνας. Είναι κατάλληλο κυρίως για εικόνες που έχουν θόρυβο ή θόλωση και δεν κατανέμονται ομοιόμορφα. Αντιθέτως το ESRGAN είναι ιδανικό για την βελτίωση ανάλυσης εικόνων χαμηλής ποιότητας, το ESRGAN παρέχει εξαιρετικά ρεαλιστικά αποτελέσματα. Ενσωματώνει τη Perceptual Loss και την Relativistic GAN για πιο φυσικές και λεπτομερείς ανακατασκευασμένες εικόνες. Και τα δύο μοντέλα είναι ισχυρά και πρωτοποριακά στον τομέα τους, αλλά η επιλογή εξαρτάται από τον συγκεκριμένο στόχο που ευελπιστούμε να πετύχουμε. [\[39\]\[40\]\[41\]](#)

ΚΕΦΑΛΑΙΟ 6

6. Image Upscaling με την χρήση του ESRGAN

6.1 Ανάλυση και επίλυση προκλήσεων του ESRGAN

Όπως αναφέραμε και σε προηγούμενα κεφάλαια, το ESRGAN, εντάσσεται στην κατηγορία των Generative Adversarial Networks (GANs), το οποίο είναι ένας τύπος μοντέλου machine learning που χρησιμοποιείται για τη δημιουργία δεδομένων, όπως εικόνες, ήχους, ή και κείμενα, τα οποία μοιάζουν με πραγματικά δεδομένα. Η κύρια λειτουργία του είναι η αντιπαλότητα των μηχανισμών Generator και Discriminator. Ο Generator, δημιουργεί δεδομένα που προσπαθούν να εμφανιστούν ως πραγματικά και ο Discriminator αξιολογεί εάν τα δεδομένα αυτά που έλαβε είναι πραγματικά ή παραγόμενα από τον Generator. Αυτή η διαδικασία συνεχίζεται μέχρι ο Generator να δημιουργεί δεδομένα τόσο ρεαλιστικά που ο Discriminator να μην μπορεί πλέον να τα διακρίνει από τα πραγματικά, δημιουργώντας μία συνεχή εξέλιξη των δυνατοτήτων τους.

Δυστυχώς υπάρχουν πολλές δυσκολίες και προκλήσεις για την πλήρη ανάλυση του ESRGAN. Το training ενός GAN είναι εξαιρετικά απαιτητική από άποψη υπολογιστικής ισχύος. Η συνδυασμένη εκπαίδευση του Generator και του Discriminator απαιτεί πολλές επαναλήψεις για να επιτευχθεί ισορροπία. Τα GANs, γενικά, είναι γνωστά για την αστάθεια κατά την εκπαίδευση, καθώς μπορούν να προκύψουν προβλήματα όπως το Mode Collapse, όπου ο Generator παράγει συνεχώς παρόμοια μεταξύ τους δεδομένα ή Oscillating Loss, όταν οι απώλειες του Generator και του Discriminator παύουν να κυμαίνονται χωρίς σύγκλιση. Επίσης η πολυπλοκότητα των Residual Dense Blocks καθιστούν την βελτιστοποίηση του μοντέλου δυσκολότερη και για να αποδώσει σωστά το ESRGAN, απαιτούνται μεγάλοι όγκοι High Resolution δεδομένων για το Training του. Πράγμα το οποίο είναι δύσκολο να αποκτηθεί χωρίς τον κατάλληλο εξοπλισμό με την ανάλογη υπολογιστή ισχύ. Η αναβάθμιση εικόνων υψηλής ανάλυσης σε πραγματικό χρόνο απαιτεί σύγχρονες GPU καθιστώντας το ESRGAN λιγότερο προσβάσιμο για χρήστες με περιορισμένο υλικό. Η δημιουργία ενός pre-trained (προ-εκπαιδευμένου) ESRGAN απαιτεί μεγάλους πόρους, τεχνική γνώση, και σταθερού training. Η ίδια η διαδικασία εκπαίδευσης των GAN, παραμένει μια από τις μεγαλύτερες προκλήσεις στον τομέα της τεχνητής νοημοσύνης.

Για το πειραματικό κομμάτι της εργασίας μας λοιπόν, θα χρησιμοποιήσουμε να από τα πιο διαδεδομένα pre trained μοντέλα για αναβάθμιση εικόνων, προσφέροντας εξαιρετική ποιότητα και ρεαλισμό, το [RRDB_ESRGAN_x4.pth](#). Το RRDB_ESRGAN_x4.pth είναι ένα προ-εκπαιδευμένο μοντέλο του ESRGAN (Enhanced Super Resolution Generative adversarial Network) που έχει εκπαιδευτεί για την αναβάθμιση εικόνων κατά 4 φορές της αρχικής τους ανάλυσης, π.χ. Μια εικόνα 128x128 μπορεί να αναβαθμιστεί σε 512x512. Το αρχείο αντιπροσωπεύει ένα ESRGAN με τη συγκεκριμένη αρχιτεκτονική RRDB (Residual-in-Residual Dense Block), που είναι η κύρια μονάδα του ESRGAN. Αυτή η αρχιτεκτονική χρησιμοποιείται για τη διατήρηση λεπτομερειών και τη βελτίωση της σταθερότητας κατά την διαδικασία του training. Η επέκταση .pth είναι η προεπιλεγμένη επέκταση για την αποθήκευση μοντέλων στο PyTorch, τη βιβλιοθήκη με την οποία έχει υλοποιηθεί το ESRGAN. Το αρχείο περιλαμβάνει παραμέτρους όπως οι διαστάσεις των φίλτρων, τα learning rates κ.λπ., που καθορίζουν τη συμπεριφορά του μοντέλου καθώς και πληροφορίες σχετικά με την αρχιτεκτονική του μοντέλου, όπως τα Residual Dense Blocks, τον αριθμό επιπέδων, και τον τρόπο που συνδέονται.

Το μοντέλο έχει εκπαιδευτεί σε μεγάλα σύνολα δεδομένων υψηλής ποιότητας, όπως το DIV2K dataset. Το DIV2K (Diverse 2K) είναι ένα σύνολο δεδομένων (Dataset) που χρησιμοποιείται ευρέως στην έρευνα της αναβάθμισης εικόνων (image super resolution) και άλλων εφαρμογών vision programming. Το σύνολο δεδομένων δημιουργήθηκε για να παρέχει εικόνες υψηλής ποιότητας και ποικιλίας, ώστε να βοηθήσει στην εκπαίδευση και την αξιολόγηση αλγορίθμων. Περιέχει 1000 εικόνες υψηλής ανάλυσης (2048 pixel), οι οποίες καλύπτουν μεγάλη ποικιλία θεμάτων (π.χ. τοπία, αντικείμενα κ.λπ.), ώστε να προάγεται η γενίκευση των αλγορίθμων. Οι εικόνες υψηλής ανάλυσης (HR) συνοδεύονται από υποδείγματα χαμηλής ανάλυσης (LR) που δημιουργούνται με διάφορες τεχνικές, όπως θόλωση ή μείωση διαστάσεων. Οι τεχνικές αυτές προσομοιώνουν τις προκλήσεις της αναβάθμισης εικόνων. Το DIV2K χρησιμοποιείται κυρίως σε εφαρμογές που σχετίζονται με την αναβάθμιση ανάλυσης εικόνων, όπως είναι το training στα GAN, SRGAN και ESRGAN. Το DIV2K μπορεί να ληφθεί δωρεάν από την ιστοσελίδα του οργανισμού NTIRE (New Trends in Image Restoration and Enhancement). Παρέχουν ρεαλιστικά δεδομένα που αντιπροσωπεύουν τις πραγματικές εφαρμογές, υποστηρίζουν τη σύγκριση αλγορίθμων μεταξύ διαφορετικών ερευνητικών ομάδων καθώς και βοηθάνε στην γενίκευση και στη βελτίωση της απόδοσης των μοντέλων.

6.2 Βήματα εγκατάστασης και ανάλυση κώδικα ESRGAN σε Python

Για την επίλυση λοιπόν της μεγάλης πολυπλοκότητας των GAN και λόγω της περιορισμένης υπολογιστικής ισχύος, θα χρησιμοποιήσουμε το pre trained model του ESRGAN, [RRDB_ESRGAN_x4.pth](#), που αναφέραμε και στο προηγούμενο κεφάλαιο. Μας δίνεται η δυνατότητα να το χρησιμοποιήσουμε για τους ερευνητικούς στόχους της εργασίας μας. Για υλοποιήσουμε όμως την έρευνά μας θα πρέπει να ακολουθήσουμε κάποια συγκεκριμένα βήματα. Θα πρέπει να εγκαταστήσουμε και να τροποποιήσουμε κάποια συγκεκριμένα εργαλεία.

```
1 import os.path as osp
2 import glob
3 import cv2
4 import numpy as np
5 import torch
6 import RRDBNet_arch as arch
7
8 model_path = 'models/RRDB_ESRGAN_x4.pth' # models/RRDB_ESRGAN_x4.pth OR models/RRDB_PSNR_x4.pth
9 device = torch.device('cpu') # if you want to run on CPU, change 'cuda' -> cpu
10 # device = torch.device('cpu')
11
12 test_img_folder = 'LR/*'
13
14 model = arch.RRDBNet(3, 3, 64, 23, gc=32)
15 model.load_state_dict(torch.load(model_path), strict=True)
16 model.eval()
17 model = model.to(device)
18
19 print('Model path {:s}. \nTesting...'.format(model_path))
20
21 idx = 0
22 for path in glob.glob(test_img_folder):
23     idx += 1
24     base = osp.splitext(osp.basename(path))[0]
25     print(idx, base)
26     # read images
27     img = cv2.imread(path, cv2.IMREAD_COLOR)
28     img = img * 1.0 / 255
29     img = torch.from_numpy(np.transpose(img[:, :, [2, 1, 0]], (2, 0, 1))).float()
30     img_LR = img.unsqueeze(0)
31     img_LR = img_LR.to(device)
32
33     with torch.no_grad():
34         output = model(img_LR).data.squeeze().float().cpu().clamp_(0, 1).numpy()
35         output = np.transpose(output[[2, 1, 0], :, :], (1, 2, 0))
36         output = (output * 255.0).round()
37         cv2.imwrite('results/{:s}_rlt.png'.format(base), output)
38
```

Εικόνα 26 RRDB ESRGAN test.py

Αρχικά εγκαταστήσαμε στον υπολογιστή μας Python, μέσω CMD. Έπειτα έπρεπε να εγκαταστήσουμε την βιβλιοθήκη PyTorch καθώς αυτή είναι η βιβλιοθήκη που έχει υλοποιηθεί το ESRGAN. Τέλος, κατεβάσαμε το pre trained μοντέλο, το οποίο υπάρχει διαθέσιμο στο Github, ως open source κώδικας, προσβάσιμος δωρεάν για οποιοδήποτε χρήστη επιθυμεί να τεστάρει και να αναλύσει το ESRGAN, για να κάνει upscale όσες εικόνες επιθυμεί. Μέσα στον φάκελο που γίνεται η εγκατάσταση μοντέλου, υπάρχει και ένα εκτελέσιμο πρόγραμμα με όνομα «test.py», το οποίο είναι και ο κώδικας που εκτελεί τις απαραίτητες εντολές για την διαδικασία Image Upscaling της low resolution εικόνας που εισάγει ο χρήστης. Ο κώδικας έχει ως εξής:

Αρχικά γίνεται εισαγωγή των απαραίτητων βιβλιοθηκών. Η `os.path`, είναι για την μεταφορά και διαχείριση των path των αρχείων και των φακέλων. Η `glob`, χρησιμοποιείται να βρει τα αρχεία με επεκτάσεις όπως jpeg κλπ.. Η `cv2`, σχετίζεται με την βιβλιοθήκη OpenCV, η οποία δίνει την δυνατότητα image reading and writing. Η `numpy`, είναι για αριθμητικούς υπολογισμούς και ειδικά για πίνακες. Η `torch`, που αναφέραμε και στο προηγούμενο κεφάλαιο, είναι για deep learning με PyTorch και τέλος η `RRDBNet_arch`, είναι το custom αρχείο που περιλαμβάνει το δίκτυο RRDBNet (Residual-in-Residual Dense Block Net).

```
model_path = 'models/RRDB_ESRGAN_x4.pth' # models/RRDB_ESRGAN_x4.pth OR models/RRDB_PSNR_x4.pth
device = torch.device('cpu') # if you want to run on CPU, change 'cuda' -> cpu
# device = torch.device('cuda')

test_img_folder = 'LR/*'
```

Εικόνα 27 Define paths

Η μεταβλητή `model_path`, ορίζει το σημείο που περιέχει τα pre trained μεγέθη (Weights) του μοντέλου. Η `device` ορίζει ποια συσκευή θα τρέξει τους υπολογισμούς. Στην δική μας περίπτωση θα εισάγουμε την CPU του υπολογιστή μας. Η `test_img_folder`, υποδεικνύει το path των εικόνων που θα υποβληθούν σε επεξεργασία.

```
model = arch.RRDBNet(3, 3, 64, 23, gc=32)
model.load_state_dict(torch.load(model_path), strict=True)
model.eval()
model = model.to(device)
```

Εικόνα 28 Import Model

Στο σημείο αυτό, δημιουργείται ένα αντικείμενο `arch.RRDBNet`, που σχετίζεται με το μοντέλο RRDBNet με ακριβείς παραμέτρους. Περιλαμβάνει 3 Channel inputs and outputs (RGB εικόνες), 64 features per layer, 23 Residual Dense Block levels και η παράμετρος `gc=32` είναι ο αριθμός των καναλιών στα ενδιάμεσα dense blocks. Η `model.load_state_dict`, φορτώνει τα μεγέθη του pre trained μοντέλου από το αρχείο. Η `model.eval()`, θέτει το μοντέλο σε evaluation mode. Η `model.to(device)` μεταφέρει το μοντέλο στην συσκευή (CPU or CUDA).

```
idx = 0
for path in glob.glob(test_img_folder):
    idx += 1
    base = osp.splitext(osp.basename(path))[0]
    print(idx, base)
```

Εικόνα 29 Image processing

Το script επεξεργάζεται όλες τις εικόνες από τον LR (Low Resolution) φάκελο χρησιμοποιώντας μία loop. Η `glob.glob(test_img_folder)`, βρίσκει όλα τα αρχεία που ταιριάζουν στο μοτίβο με όνομα LR. Η `osp.basename(path)` εξαγεί το όνομα του αρχείου από το path. Η `osp.splitext(...)[0]`, αφαιρεί την επέκταση των εικόνων και κρατάει μόνο το όνομά τους.

```
img = cv2.imread(path, cv2.IMREAD_COLOR)
img = img * 1.0 / 255
img = torch.from_numpy(np.transpose(img[:, :, [2, 1, 0]], (2, 0, 1))).float()
img_LR = img.unsqueeze(0)
img_LR = img_LR.to(device)
```

Εικόνα 30 Image pre processing

Η `cv2.imread(path, cv2.IMREAD_COLOUR)`, επεξεργάζεται την εικόνα ως έγχρωμη RGB εικόνα. Η `img * 1.0 / 255`, απλοποιεί τις τιμές των pixel σε [0, 1]. Η `np.transpose` αλλάζει τη διάταξη των διαστάσεων σε κανάλια, ύψος και πλάτος. Η `torch.from_numpy`, μετατρέπει την εικόνα σε PyTorch tensor. Η `img.unsqueeze(0)`, προσθέτει μία διάσταση (batch size =1). Η `img_LR.to(device)`, μεταφέρει την εικόνα στη συσκευή υπολογισμού.

```
with torch.no_grad():  
    output = model(img_LR).data.squeeze().float().cpu().clamp_(0, 1).numpy()  
    output = np.transpose(output[[2, 1, 0], :, :], (1, 2, 0))  
    output = (output * 255.0).round()  
    cv2.imwrite('results/{:s}_rlt.png'.format(base), output)
```

Εικόνα 31 Model Application and Output export

Η `torch.no_grad()`, απενεργοποιεί τον υπολογισμό gradients για εξοικονόμηση της μνήμης. Η `model(img_LR)`, εφαρμόζει το super resolution model στην LR εικόνα. Η `data.squeeze()` αφαιρεί την διάσταση batch που εισήγαγε για τους απαραίτητους υπολογισμούς και επιστρέφει τα δεδομένα της εικόνας. Η `clamp_(0,1)` εξασφαλίζει ότι οι τιμές του output παραμένουν στο [0,1]. Η `np.transpose` μετατρέπει τα κανάλια πίσω σε διάταξη OpenCV (Υψος, Πλάτος και Κανάλια). Η `output * 255.0`, μετατρέπει τις τιμές του output πίσω στο εύρος [0,255] και τέλος η `cv2.imwrite`, αποθηκεύει την επεξεργασμένη εικόνα στο φάκελο "results" με το ίδιο όνομα αλλά με την επέκταση `_rlt.png`.

Ουσιαστικά ο κώδικάς μας, εκτελεί Super Resolution image Upscaling σε χαμηλής ανάλυσης εικόνες που βρίσκονται στον φάκελο LR. Το μοντέλο που χρησιμοποιείται είναι το RRDBNet, μέρος του framework ESRGAN (Enhanced Super-Resolution Generative Adversarial Networks).

6.3 Αποτελέσματα Upscaling με ESRGAN

Για να εμφανίσουμε τα αποτελέσματα της έρευνά μας, τοποθετούμε Low Resolution εικόνες στον φάκελο LR, που δημιουργείτε κατά την εγκατάσταση το Repository. Έπειτα ο κώδικας κοιτάει και επιλέγει τις εικόνες από το φάκελο LR, και καλεί το pre trained μοντέλο RRDBNet, για να επεξεργαστεί τις εικόνες μας. Μετά την διαδικασία του Super Resolution Upscaling, ο κώδικας τοποθετεί την τελική εικόνα στον φάκελο results. Χρησιμοποιήσαμε δύο Low Resolution δείγματα, όπου η μία έχει διαστάσεις 125 X 120 και η άλλη έχει 62 X 90 και τα αποτελέσματα που λάβαμε είναι τα εξής:



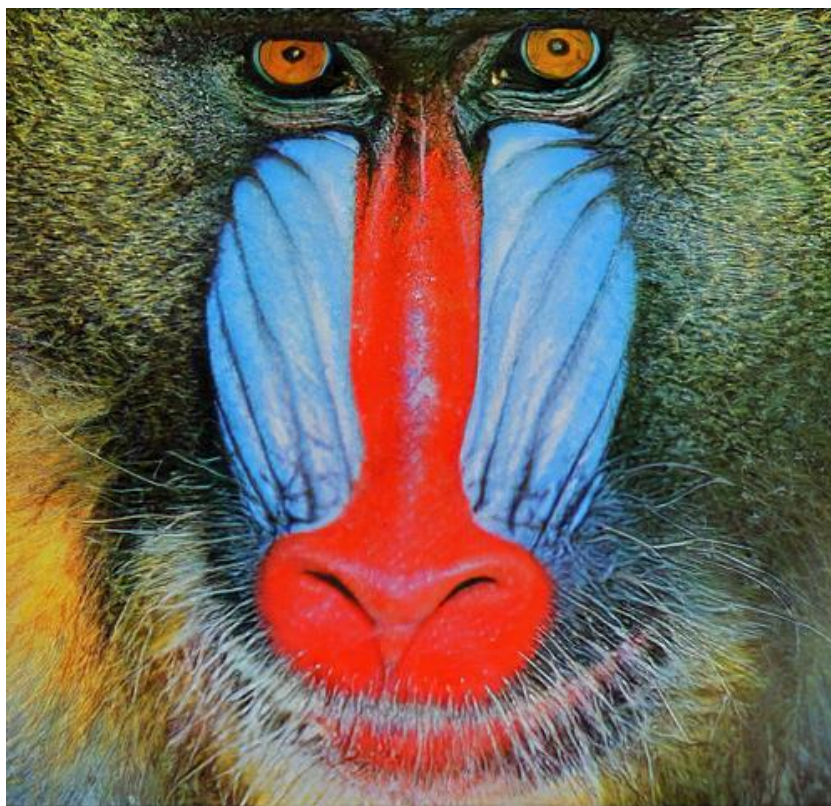
Εικόνα 32 LR image 1 62x90



Εικόνα 33 HR image 248x368



Εικόνα 34 LR image 125x120

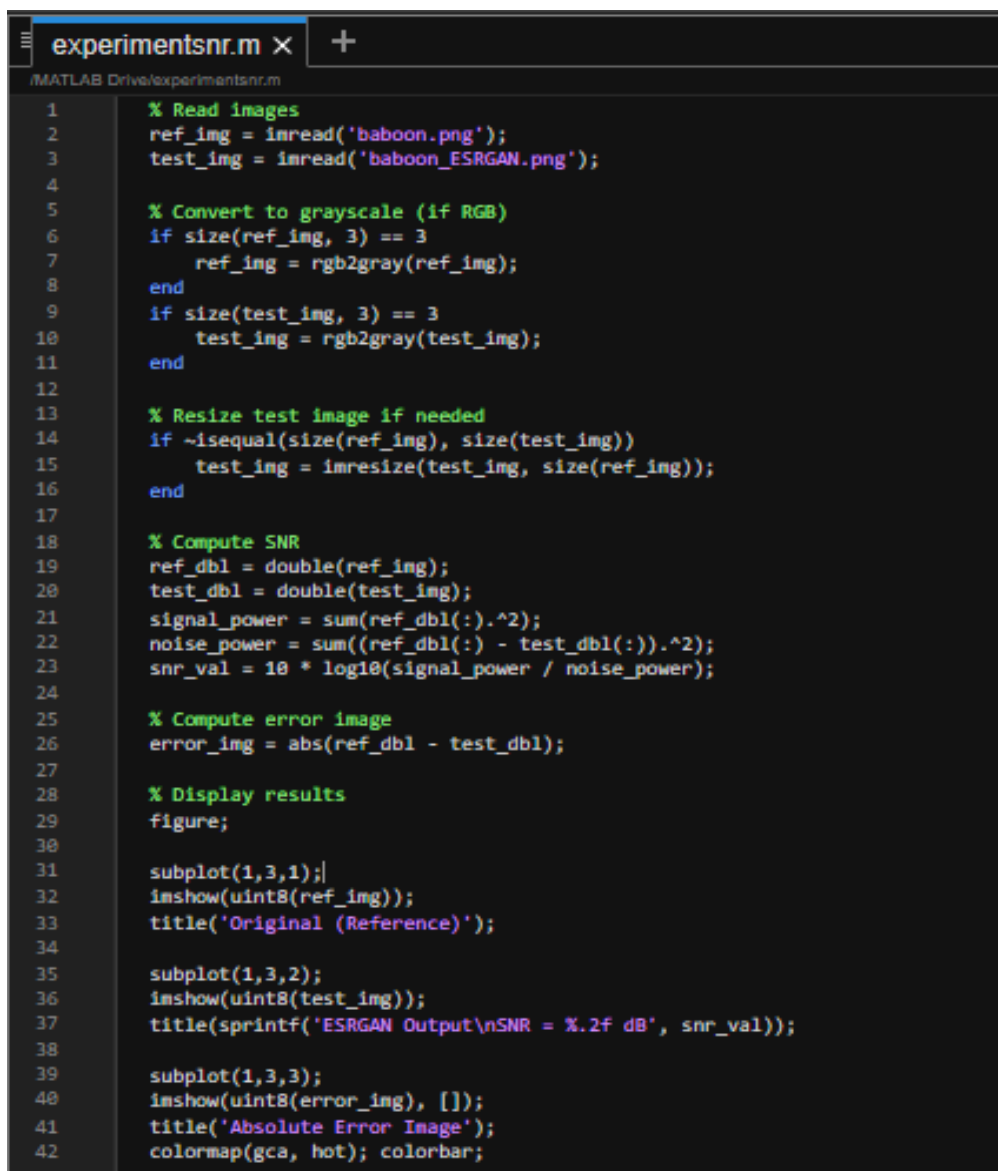


Εικόνα 35 HR image 500x480

Μπορούμε να διακρίνουμε πως ο κώδικας λειτούργησε με επιτυχία και οι εικόνες δέχτηκαν κανονικά Upscaling με το pre trained super resolution model. Αυξήθηκε το μέγεθός τους Χ4 και η ποιότητα της ανάλυσης στο μάτι φαίνεται εξαιρετική. Οι διαστάσεις της πρώτης εικόνας μετατράπηκαν από 62 X 90 σε 248 X 368 και η δεύτερη εικόνα μετατράπηκε κανονικά από 125 X 120 σε 500 X 480.

6.4 SNR (Signal to noise ratio)

Χρησιμοποιήσαμε λοιπόν το MATLAB για να τεστάρουμε την απόκλιση των εικόνων που επεξεργαστήκαμε με το ESRGAN image upscale πείραμα. Και ποιο συγκεκριμένα τον παρακάτω κώδικα που υπολογίζει το Signal to Noise ratio μεταξύ δύο εικόνων.

A screenshot of a MATLAB script editor window titled 'experimentsnr.m'. The code is written in a dark-themed editor and contains 42 lines of MATLAB code. The script reads two images ('baboon.png' and 'baboon_ESRGAN.png'), converts them to grayscale if they are RGB, resizes the test image to match the reference image, computes the SNR, and displays the results in a 3x1 subplot. The first subplot shows the original reference image, the second shows the ESRGAN output, and the third shows the absolute error image. The SNR value is printed in the title of the second subplot.

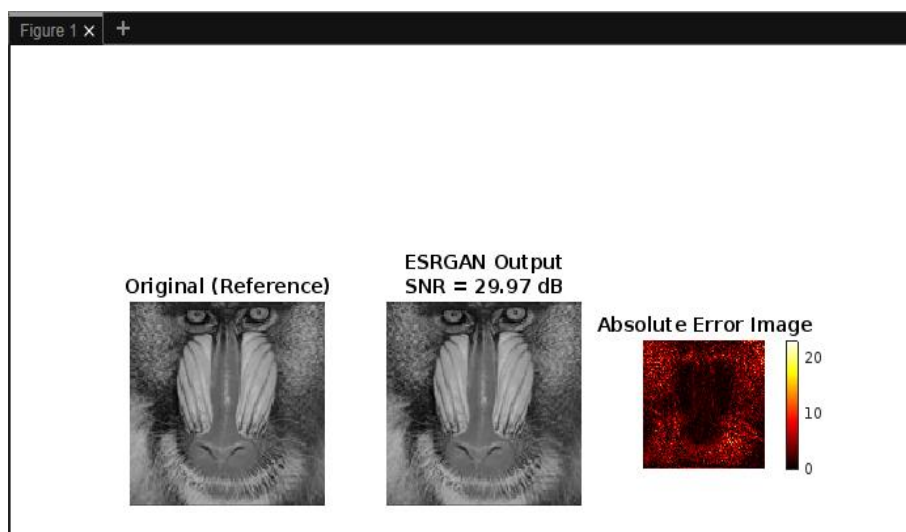
```
1 % Read images
2 ref_img = imread('baboon.png');
3 test_img = imread('baboon_ESRGAN.png');
4
5 % Convert to grayscale (if RGB)
6 if size(ref_img, 3) == 3
7     ref_img = rgb2gray(ref_img);
8 end
9 if size(test_img, 3) == 3
10     test_img = rgb2gray(test_img);
11 end
12
13 % Resize test image if needed
14 if ~isequal(size(ref_img), size(test_img))
15     test_img = imresize(test_img, size(ref_img));
16 end
17
18 % Compute SNR
19 ref_dbl = double(ref_img);
20 test_dbl = double(test_img);
21 signal_power = sum(ref_dbl(:).^2);
22 noise_power = sum((ref_dbl(:) - test_dbl(:)).^2);
23 snr_val = 10 * log10(signal_power / noise_power);
24
25 % Compute error image
26 error_img = abs(ref_dbl - test_dbl);
27
28 % Display results
29 figure;
30
31 subplot(1,3,1);
32 imshow(uint8(ref_img));
33 title('Original (Reference)');
34
35 subplot(1,3,2);
36 imshow(uint8(test_img));
37 title(sprintf('ESRGAN Output\nSNR = %.2f dB', snr_val));
38
39 subplot(1,3,3);
40 imshow(uint8(error_img), []);
41 title('Absolute Error Image');
42 colormap(gca, hot); colorbar;
```

Εικόνα 36 SNR code in MATLAB

Στην ουσία ο κώδικας αυτός εφαρμόζει τον SNR έλεγχο μεταξύ δύο εικόνων που θέλουμε να του δώσουμε. Στην συγκεκριμένη περίπτωση του δώσαμε το αρχικό png με όνομα “baboon.png” και την επεξεργασμένη εικόνα έπειτα από το πείραμα με το upscaling image ESRGAN, με όνομα “baboon_ESRGAN.png”. Αρχικά ο κώδικας αυτός μετατρέπει τις εικόνες σε (1 Channel) Grayscale εικόνες, καθώς το SNR μπορεί να εφαρμοστεί μόνο μέσω Grayscale εικόνων και στο τέλος μας δίνει το αποτέλεσμα σε dB και βάλαμε να μας δείχνει και την διαφορά του θορύβου μεταξύ των δύο εικόνων για μία καλύτερη αναπαράσταση.

```
Command Window
>> experimentsnr
SNR between baboon.png and baboon_ESRGAN.png: 29.97 dB
>> experimentsnr
>>
```

Εικόνα 37 SNR dB results



Εικόνα 38 SNR Visualization

6.5 Αποτελέσματα πειράματος

Στο πλαίσιο της παρούσας εργασίας μελετήθηκε η αποτελεσματικότητα της τεχνικής ESRGAN (Enhanced Super-Resolution Generative Adversarial Network) για image upscaling, δηλαδή τη βελτίωση της ανάλυσης εικόνων χαμηλής ανάλυσης. Για το πείραμα χρησιμοποιήθηκαν δύο εικόνες, εκ των οποίων η μία αποτέλεσε το αρχικό δείγμα χαμηλής ανάλυσης (input), ενώ η δεύτερη ήταν το αποτέλεσμα της επεξεργασίας με τον ανοιχτού κώδικα ESRGAN αλγόριθμο.

Στόχος του πειράματος ήταν η αξιολόγηση της ποιότητας της ανακατασκευασμένης εικόνας, μέσω της μέτρησης του SNR (Signal-to-Noise Ratio), το οποίο αποτελεί έναν ευρέως χρησιμοποιούμενο δείκτη για τον υπολογισμό της ποσότητας θορύβου που προστίθεται κατά την επεξεργασία εικόνας. Ο υπολογισμός του SNR έγινε μέσω κώδικα που υλοποιήθηκε στο περιβάλλον του MATLAB, συγκρίνοντας την αρχική εικόνα με την επεξεργασμένη. Οι εικόνες μετατράπηκαν σε Grayscale και έγιναν κατάλληλες μετατροπές τύπου και μεγέθους για την ακριβή σύγκριση των pixel. Τέλος, έγινε οπτική απεικόνιση των διαφορών μεταξύ των δύο εικόνων, συμπεριλαμβανομένης και της αποτύπωσης της εικόνας σφάλματος. Το αποτέλεσμα της ανάλυσης έδειξε ότι το SNR ανήλθε στα **29.97 dB**, τιμή που υποδηλώνει μέτρια προς υψηλή ποιότητα.

SNR (dB)	Quality
< 20 dB	Poor — very noisy
20–30 dB	Fair to good — some noise visible
> 30 dB	High quality — minimal noise

Εικόνα 39 dB Stats For SNR

Αν και η τιμή αυτή δεν υποδηλώνει τέλεια αποκατάσταση (ιδανικά SNR > 30 dB), εντούτοις δηλώνει ότι η εικόνα που προέκυψε μέσω ESRGAN είναι **αρκετά κοντά στην αρχική**, με **σχετικά μικρή προσθήκη θορύβου**. Είναι σημαντικό να σημειωθεί ότι οι GAN-based τεχνικές δεν στοχεύουν αποκλειστικά σε μέγιστο SNR, αλλά περισσότερο στη **πιο βέλτιστη οπτική ποιότητα (perceptual quality)**, δηλαδή στη δημιουργία εικόνας που αναπαρίσταται φυσική στον ανθρώπινο παρατηρητή, ακόμα κι αν ο θόρυβος είναι αριθμητικά υψηλότερος. Συνοψίζοντας, το ESRGAN πέτυχε σημαντική βελτίωση στην ανάλυση των εικόνων με ελέγχοντας τις μετατροπές στον θόρυβο. Η προσέγγιση αυτή μπορεί να χρησιμοποιηθεί αποτελεσματικά σε εφαρμογές όπως η αποκατάσταση παλιών φωτογραφιών.

ΒΙΒΛΙΟΓΡΑΦΙΑ

[1] Wikipedia, Photography

<https://el.wikipedia.org/wiki/%CE%A6%CF%89%CF%84%CE%BF%CE%B3%CF%81%CE%B1%CF%86%CE%AF%CE%B1>

[2] Culture Point | Αναρτήθηκε στον ιστότοπο στις 12/6/22

<https://culturepoint.gr/istoriki-ekseliksi-tis-fotografias/>

[3] Harmann Studios | Αναρτήθηκε στον ιστότοπο στις 8/1/25

<https://www.harmann.com/blog-post/the-evolution-of-photography-from-film-to-digital?utm>

[4] Britannica | Αναρτήθηκε στον ιστότοπο στις 25/4/25

<https://www.britannica.com/technology/photography>

[5] Wikipedia Information Age

https://en.wikipedia.org/wiki/Information_Age

[6] Raphael C. Gonzalez – Richard E. Woods, Ψηφιακή Επεξεργασία Εικόνας 3rd Edition

ΚΕΦΑΛΑΙΟ 6.2: ΧΡΩΜΑΤΙΚΑ ΜΟΝΤΕΛΑ

[7] Raphael C. Gonzalez – Richard E. Woods, Ψηφιακή Επεξεργασία Εικόνας 3rd Edition

ΚΕΦΑΛΑΙΟ 4.9: ΟΞΥΝΣΗ ΜΕ ΦΙΛΤΡΑ ΣΤΟ ΧΩΡΟ ΣΥΧΝΟΤΗΤΩΝ

[8] Raphael C. Gonzalez – Richard E. Woods, Ψηφιακή Επεξεργασία Εικόνας 3rd Edition

ΚΕΦΑΛΑΙΟ 5: ΑΠΟΚΑΤΑΣΤΑΣΗ ΚΑΙ ΑΝΑΚΑΤΑΣΚΕΥΗ ΕΙΚΟΝΑΣ

[9] Raphael C. Gonzalez – Richard E. Woods, Ψηφιακή Επεξεργασία Εικόνας 3rd Edition

ΚΕΦΑΛΑΙΟ 1.1: ΤΙ ΕΙΝΑΙ Η ΕΠΕΞΕΡΓΑΣΙΑ ΕΙΚΟΝΑΣ

[10] Geeks for Geeks, Digital Image Processing Basics | Αναρτήθηκε στον ιστότοπο στις 22/2/23

<https://www.geeksforgeeks.org/digital-image-processing-basics/>

[11] Geeks for Geeks, Difference between Low pass filter and High pass filter | Αναρτήθηκε στον ιστότοπο στις 15/3/23

<https://www.geeksforgeeks.org/difference-between-low-pass-filter-and-high-pass-filter/>

[12] IBM, The history of AI | Αναρτήθηκε στον ιστότοπο στις 21/2/24

<https://www.ibm.com/think/topics/history-of-artificial-intelligence>

[13] Wikipedia History of Artificial Intelligence

https://en.wikipedia.org/wiki/History_of_artificial_intelligence

[14] Nanonets, How AI Image Processing Transforms Business Operations | Αναρτήθηκε στον ιστότοπο στις 26/7/24

<https://nanonets.com/blog/ai-image-processing/>

[15] Geeks for geeks, AI in Image Processing | Αναρτήθηκε στον ιστότοπο στις 17/7/24

<https://www.geeksforgeeks.org/ai-in-image-processing/>

[16] Wikipedia, Image Scaling

https://en.wikipedia.org/wiki/Image_scaling

[17] Artstorefronts, The Basics of Image Upscaling: Understanding the Process and Pros and Cons | Αναρτήθηκε στον ιστότοπο στις 4/7/24

<https://support.artstorefronts.com/hc/en-us/articles/15091792050203-The-Basics-of-Image-Upscaling-Understanding-the-Process-and-Pros-and-Cons>

[18] Wikipedia, Comparison gallery of image scaling algorithms

https://en.wikipedia.org/wiki/Comparison_gallery_of_image_scaling_algorithms

[19] Medium, Near Neighbor Interpolation | Αναρτήθηκε στον ιστότοπο στις 27/3/24

<https://medium.com/%40akp83540/nearest-neighbor-interpolation-84c956ee56a3>

[20] X-Engineer Bilinear Interpolation

<https://x-engineer.org/bilinear-interpolation/>

[21] Cloudinary, Bicubic Interpolation | Αναρτήθηκε στον ιστότοπο στις 21/4/25

<https://cloudinary.com/glossary/bicubic-interpolation>

[22] Wikipedia, Bicubic Interpolation

https://en.wikipedia.org/wiki/Bicubic_interpolation

[23] PixInsight, Lanczos Interpolation

<https://pixinsight.com/doc/docs/InterpolationAlgorithms/InterpolationAlgorithms.html>

[24] Mazzo.li, Lanczos Interpolation explained | Αναρτήθηκε στον ιστότοπο στις 10/7/22

<https://mazzo.li/posts/lanczos.html>

[25] Wikipedia, Lanczos Interpolation

https://en.wikipedia.org/wiki/Lanczos_resampling

[26] V7labs, Image Processing: Techniques, Types & Application | Αναρτήθηκε στον ιστότοπο στις 3/8/22

<https://www.v7labs.com/blog/image-processing-guide>

[27] Upscale Media, A Step-by-Step Guide About the Image Upscaling Process

<https://www.upscale.media/blog/everything-you-want-to-know-about-the-image-upscaling-process>

[28] Intel, Convolutional Neural Networks

<https://www.intel.com/content/www/us/en/internet-of-things/computer-vision/convolutional-neural-networks.html>

[29] IBM, What are convolutional neural networks?

<https://www.ibm.com/think/topics/convolutional-neural-networks>

[30] Geeks for geeks, Math Behind Convolutional Neural Networks | Αναρτήθηκε στον ιστότοπο στις 30/8/24

<https://www.geeksforgeeks.org/math-behind-convolutional-neural-networks/>

[31] Medium, Breaking Down the Mathematics Behind CNN Models: A Comprehensive Guide | Αναρτήθηκε στον ιστότοπο στις 2/5/23

<https://medium.com/%40beingfarina/breaking-down-the-mathematics-behind-cnn-models-a-comprehensive-guide-1853aa6b011e>

[32] Medium, Gentle Dive into Math Behind Convolutional Neural Networks | Αναρτήθηκε στον ιστότοπο στις 13/4/19

<https://medium.com/data-science/gentle-dive-into-math-behind-convolutional-neural-networks-79a07dd44cf9>

[33] Medium, Superresolution explained | Αναρτήθηκε στον ιστότοπο στις 12/6/19

<https://medium.com/@manasbedmutha98/superresolution-explained-ep2-8c188de1edfc>

[34] Geeks for geeks, Generative Adversarial Network (GAN) | Αναρτήθηκε στον ιστότοπο στις 10/3/25

<https://www.geeksforgeeks.org/generative-adversarial-network-gan/>

[35] Overview of Gan Structure

https://developers.google.com/machine-learning/gan/gan_structure

- [36] Geeks for geeks, Super Resolution GAN (SRGAN) | Αναρτήθηκε στον ιστότοπο στις 4/7/22 <https://www.geeksforgeeks.org/super-resolution-gan-srgan/>
- [37] Analytics Vidhya, SRGANs: Bridging the Gap Between Low-res and High-res Images | Αναρτήθηκε στον ιστότοπο στις 12/7/23
<https://www.analyticsvidhya.com/blog/2023/06/srgans-bridging-the-gap-between-low-res-and-high-res-images/>
- [38] Digital Ocean, SRGAN: Super Resolution Generative Adversarial Networks | Αναρτήθηκε στον ιστότοπο στις 25/9/24
<https://www.digitalocean.com/community/tutorials/super-resolution-generative-adversarial-networks>
- [39] Medium, ESRGAN: Enhanced Super-Resolution Generative Adversarial Network | Αναρτήθηκε στον ιστότοπο στις 15/9/20
<https://medium.com/analytics-vidhya/esrgan-enhanced-super-resolution-generative-adversarial-network-using-keras-a34134b72b77>
- [40] Geeks for geeks, Image Super-Resolution with ESRGAN using PyTorch | | Αναρτήθηκε στον ιστότοπο στις 19/9/24
<https://www.geeksforgeeks.org/image-super-resolution-with-esrgan-using-pytorch/>
- [41] Wikipedia, Generative Adversarial Networks
https://en.wikipedia.org/wiki/Generative_adversarial_network
- [42] Github Repository, xinntao
<https://github.com/xinntao/Real-ESRGAN>
- [43] Matlab SNR
<https://www.mathworks.com/help/signal/ref/snr.html>