



ΠΑΝΕΠΙΣΤΗΜΙΟ
ΙΩΑΝΝΙΝΩΝ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ



ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΜΗΧΑΝΙΚΩΝ Η/Υ
& ΔΙΚΤΥΩΝ

ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

“ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΕΙΚΟΝΩΝ ΧΑΜΗΛΟΥ ΦΩΤΙΣΜΟΥ ΜΕ
ΧΡΗΣΗ ΜΕΘΟΔΩΝ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ”

Ελευθέριος Κ. Μαστραπάς

Επιβλέπων: Πέτρος Καρβέλης
Επίκουρος Καθηγητής

Αρτα, ~~Ιούλιος~~ ~~Φεβρουάριος~~, 2022

Ορισμός στυλ: ΠΠ 1: Στηλοθέτες: 15,98 εκ., Δεξιά, Οδηγός:
...

Low Light Image Enhancement
Using Machine Learning Methods

Εγκρίθηκε από τριμελή εξεταστική επιτροπή

Αρτα, 07/2022 Τόπος, Ημερομηνία

ΕΠΙΤΡΟΠΗ ΑΞΙΟΛΟΓΗΣΗΣ

1. Επιβλέπων καθηγητής

Πέτρος Καρβέλης Όνομα-Επίθετο,

Επικουρος Καθηγητής τίτλος, βαθμίδα

2. Μέλος επιτροπής

Χρυσόστομος Στύλιος Όνομα-Επίθετο,

Καθηγητής τίτλος, βαθμίδα

3. Μέλος επιτροπής

Ιωάννης Τσούλος Όνομα-Επίθετο,

Αναπληρωτής Καθηγητής τίτλος, βαθμίδα

© Μαστραπάς, Ελευθέριος, 2022~~±~~.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Δήλωση μη λογοκλοπής

Δηλώνω υπεύθυνα και γνωρίζοντας τις κυρώσεις του Ν. 2121/1993 περί Πνευματικής Ιδιοκτησίας, ότι η παρούσα μεταπτυχιακή εργασία είναι εκ 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.

Μαστραπός, Ελευθέριος

Υπογραφή

ΕΥΧΑΡΙΣΤΙΕΣ

Η εκπόνηση της διπλωματικής εργασίας πραγματοποιήθηκε κατά το ακαδημαϊκό έτος 2021~~θ-~~ 2022~~+~~ με την παρουσία, υποστήριξη και συμπαράσταση ενός συνόλου ανθρώπων τους οποίους θα ήθελα να ευχαριστήσω. Τις θερμές μου ευχαριστίες στον επιβλέποντα Δρ. Καρβέλη Πέτρο, για την ανάθεση του θέματος, την καθοδήγηση και την υποστήριξη που μου παρείχε καθ' όλη την διάρκεια διεκπεραίωσης της παρούσας μεταπτυχιακής διπλωματικής εργασίας. Θα ήθελα επίσης να ευχαριστήσω την οικογένειά μου, για την ανιδιοτελή και ανυπολόγιστη υποστήριξη που μου παρείχε όλα αυτά τα χρόνια..

ΠΕΡΙΛΗΨΗ

Η μηχανική όραση αποτελεί ένα ραγδαίως αναπτυσσόμενο πεδίο, λόγω της αλματώδους εξέλιξης της τεχνητής νοημοσύνης και της βαθιάς μάθησης. Τα τεχνητά νευρωνικά δίκτυα και συγκεκριμένα τα δίκτυα βαθιάς μάθησης βοήθησαν στην ανάπτυξη εφαρμογών μηχανικής όρασης, όπως είναι η αναγνώριση αντικειμένων. Η αναγνώριση αντικειμένων σε εικόνες χαμηλού φωτισμού αποτελεί μία από τις μεγαλύτερες προκλήσεις της μηχανικής όρασης. Η περίληψη (στην ελληνική γλώσσα) αποτελεί μια συνοπτική παρουσίαση των κύριων στοιχείων και συμπερασμάτων της εργασίας, μαζί με μια σύντομη αναφορά στους στόχους και τις μεθοδολογίες που ακολουθήθηκαν. Δεν θα πρέπει να υπερβαίνει τις 300 λέξεις και έχει συγκεντρώσει έντονο ερευνητικό ενδιαφέρον. Η χρήση ενός ανιχνευτή αντικειμένων, σε εικόνες χαμηλού φωτισμού, μπορεί να οδηγήσει σε δυσάρεστα αποτελέσματα. Στην παρούσα εργασία περιγράφουμε τον τρόπο με τον οποίο τα συνελκτικά νευρωνικά δίκτυα πετυχαίνουν την ενίσχυση εικόνων χαμηλού φωτισμού που περιέχονται στο ExDark σετ δεδομένων και λύνουν το πρόβλημα της αναγνώρισης των αντικειμένων που υπάρχουν σε αυτές. Επίσης, έχει πραγματοποιηθεί εκτεταμένη βιβλιογραφική ανασκόπηση του τρόπου λειτουργίας της μεθόδου Kind και του αλγορίθμου YOLO, που χρησιμοποιούνται για ενίσχυση εικόνων και αναγνώριση αντικειμένων αντίστοιχα. Τέλος, παρουσιάζουμε την αποτελεσματικότητα του αλγορίθμου YOLO σε εικόνες χαμηλού φωτισμού, οι οποίες έχουν ενισχυθεί με χρήση της μεθόδου KinD.

Λέξεις-κλειδιά: Καταχωρίστε τρεις έως πέντε λέξεις ή φράσεις κλειδιά στα ελληνικά που περιγράφουν το θέμα της εργασίας: Μηχανική Όραση, Αναγνώριση Αντικειμένων, Μηχανική Μάθηση, Συνελκτικά Νευρωνικά Δίκτυα.

ABSTRACT

Μορφοποιήθηκε: Αγγλικά (Ηνωμένων Πολιτειών)

Computer vision is a rapidly growing field due to the leaps and bounds of artificial intelligence and deep learning. Artificial neural networks and specifically deep learning networks have helped develop computer vision applications such as object detection. Object detection in low-light images is one of the biggest challenges in computer vision and has gathered intense research interest. Using an object detector, in low-light images, can lead to unpleasant results. In this work, we describe how convolutional neural networks achieve the enhancement of low-light images contained in the ExDark dataset and solve the problem of recognizing the objects present in them. Also, an extensive literature review has been performed on the workings of the Kind method and the YOLO algorithm, which are used for image enhancement and object detection respectively. Finally, we demonstrate the effectiveness of the YOLO algorithm on low-light images, which have been enhanced using the KinD method. Η περίληψη στην αγγλική γλώσσα. Δεν θα πρέπει να υπερβαίνει τις 300 λέξεις.

Keywords: -Computer Vision, Object Detection, Machine Learning, Convolutional Neural Networks Κατηγορίστε τις λέξεις κλειδιά σύμφωνα με το πλαίσιο της εργασίας.

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΕΥΧΑΡΙΣΤΙΕΣ.....	iv*
ΠΕΡΙΛΗΨΗ.....	v*
ABSTRACT.....	vi*
ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ.....	vii*
ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ.....	xviii
ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ.....	xix*
ΠΙΝΑΚΑΣ ΣΥΝΤΟΜΟΓΡΑΦΙΩΝ.....	xviii
ΚΕΦΑΛΑΙΟ 1 – ΕΙΣΑΓΩΓΗ.....	1945
1.1 Τεχνητή Νοημοσύνη.....	1945
1.2 Κατηγορίες Τεχνητής Νοημοσύνης.....	16
1.3 Μηχανική Μάθηση.....	16
1.4 Κατηγορίες Μάθησης.....	18
1.5 Βαθιά Μάθηση.....	22
1.6 Τεχνητά Νευρωνικά Δίκτυα.....	22
1.7 Εκπαίδευση Δικτύου Βαθιάς Μάθησης.....	27
1.7.1 Βήματα Εκμάθησης και Αλγόριθμοι Βελτιστοποίησης ενός ΤΝΔ.....	27
ΚΕΦΑΛΑΙΟ 2 - ΜΗΧΑΝΙΚΗ ΟΡΑΣΗ.....	30
2.1 Εισαγωγή.....	30
2.2 Εφαρμογές της ΜΟ.....	30
2.3 Εικόνα - Αναπαράσταση Εικόνων.....	33

Μορφοποιήθηκε: Αγγλικά (Ηνωμένων Πολιτειών)

Μορφοποιήθηκε: Αγγλικά (Ηνωμένων Πολιτειών)

Μορφοποιήθηκε: Αριστερά

Μορφοποιήθηκε: Αριστερά, Διάστημα Μετά: 12 στ.

Μορφοποιήθηκε: Ελληνικά

Μορφοποιήθηκε: Αριστερά, Δεξιά: 0,5 εκ.

Μορφοποιήθηκε: Γραμματοσειρά: 11 στ., Όχι Έντονα

2.4 Προεπεξεργασία Εικόνας.....	35
2.5 Εξαγωγή Χαρακτηριστικών.....	36
2.6 Συνελκτικά Νευρωνικά Δίκτυα.....	38
2.7 Συνέλιξη σε Έγχρωμες Εικόνες.....	42
ΚΕΦΑΛΑΙΟ 3 - Επεξεργασία Εικόνων Χαμηλού Φωτισμού.....	44
3.1 Εισαγωγή.....	44
3.2 Εικόνες Χαμηλού Φωτισμού.....	44
3.3 Μέθοδοι Ενίσχυσης Εικόνων Χαμηλού Φωτισμού.....	45
3.4 Η Μέθοδος Ενίσχυσης KinD.....	49
ΚΕΦΑΛΑΙΟ 4 - Αναγνώριση Αντικειμένων.....	62
4.1 Εισαγωγή.....	62
4.1.1 Ιστορική Αναδρομή.....	63
4.2 Αλγόριθμοι Αναγνώρισης Αντικειμένων.....	66
ΚΕΦΑΛΑΙΟ 5 - Υλοποίηση και Αποτελέσματα.....	78
5.1 Το Σετ Δεδομένων.....	78
5.2 Αποτελέσματα της Μεθόδου KinD.....	82
5.3 Αποτελέσματα του Μοντέλου YOLO.....	87
5.4 Ανάλυση Αποτελεσμάτων.....	91
ΚΕΦΑΛΑΙΟ 6 - Συμπεράσματα.....	93
ΒΙΒΛΙΟΓΡΑΦΙΑ.....	95
ΕΥΧΑΡΙΣΤΙΕΣ — iv	

Μορφοποιήθηκε: Ελληνικά

Μορφοποιήθηκε: Γραμματοσειρά: 11 στ., Όχι Έντονα

Μορφοποιήθηκε: Αριστερά, Διάστημα Μετά: 12 στ.

Μορφοποιήθηκε: Ελληνικά

Μορφοποιήθηκε: Ελληνικά

Μορφοποιήθηκε: Βασικό, Πλήρης

Μορφοποιήθηκε: Προεπιλεγμένη γραμματοσειρά, Γραμματοσειρά: (Προεπιλεγμένη) + Σώμα (Calibri), Ορθογραφικός και γραμματικός έλεγχος

Μορφοποιήθηκε: Γραμματοσειρά: (Προεπιλεγμένη) Times New Roman, Ορθογραφικός και γραμματικός έλεγχος

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

<u>Πίνακας 3.1. Ανακεφαλαίωση τεχνικών και μεθόδων βαθιάς μάθησης για την ενίσχυση εικόνων χαμηλού φωτισμού</u>	47
<u>Πίνακας 3.2. Αρχιτεκτονική Στρώματος Αποσύνθεσης</u>	51
<u>Πίνακας 3.3. Αρχιτεκτονική ΣΝΔ Επαναφοράς</u>	53
<u>Πίνακας 3.4. Αρχιτεκτονική ΣΝΔ Ρύθμισης Φωτισμού</u>	54
<u>Πίνακας 3.5. Ποσοτική Σύγκριση Μεθόδων Ενίσχυσης</u>	56
<u>Πίνακας 5.1. Εικόνες Χαμηλού Φωτισμού από Δημοσιευμένα Σετ Δεδομένων που περιέχονται στο ExDark</u>	78
<u>Πίνακας 5.2. Αριθμός Εικόνων ανά κλάση Αντικειμένου</u>	79
<u>Πίνακας 5.3. Αποτελέσματα του μοντέλου Yolo</u>	86
<u>Πίνακας 5.4. Αποτελέσματα του μοντέλου Yolo</u>	86
<u>Πίνακας 5.5. Μετρήσεις των αρχικών εικόνων</u>	91
<u>Πίνακας 5.6. Μετρήσεις μετά την εφαρμογή της μεθόδου KinD</u>	92
<u>Πίνακας 2. [Τίτλος]</u>	αρ. σελίδας

ΚΑΤΑΛΟΓΟΣ ΔΙΑΓΡΑΜΜΑΤΩΝ/ΕΙΚΟΝΩΝ

Άρρητο Εικόνα 1.1 Αναπαράσταση αλπίων που υπολογίζονται από τον αλγόριθμο Κεφάλαιο 1	17
Εικόνα 1.2 Κατηγορίες Μηχανικής Μάθησης.....	18
Εικόνα 1.3 Μοντέλο πρόβλεψης τιμής πώλησης αυτοκινήτου.....	19
Εικόνα 1.4 Ομαδοποίηση δεδομένων μετά την εφαρμογή του αλγορίθμου <i>k-means</i>	20
Εικόνα 1.5 Οπτική αναπαράσταση της ενισχυτικής μάθησης.....	21
Εικόνα 1.6 Αναπαράσταση βιολογικού και τεχνητού νευρώνα.....	22
Εικόνα 1.7 Μαθηματική οπτικοποίηση ενός <i>TNA</i>	23
Εικόνα 1.8 Η συνάρτηση <i>Heaviside</i>	24
Εικόνα 1.9 Νευρωνικό Δίκτυο Βαθιάς Μάθησης.....	24
Εικόνα 1.10 <i>TNA</i> με στρώμα εισόδου τριών κόμβων.....	25
Άρρητο Εικόνα 2.1 Χάρτης Κεφάλαιο 2	29
Εικόνα 2.2 Εύρεση και αναγνώριση αντικειμένων σε πραγματικό χρόνο.....	31
Εικόνα 2.3 Αναπαράσταση ασπρόμαυρης εικόνας.....	33
Εικόνα 2.4 Αναπαράσταση έγχρωμης εικόνας.....	34
Εικόνα 2.5 Εξαγωγή χαρακτηριστικών και ταξινόμηση αντικειμένου με χρήση <i>TNA</i>	36
Εικόνα 2.6 Αρχιτεκτονική Συνελκτικού Νευρωνικού Δικτύου.....	38
Εικόνα 2.7 Συνέλιξη με φίλτρο αναγνώρισης κορυφών σε ασπρόμαυρη εικόνα.....	40
Εικόνα 2.8 Μετατροπή ενός $4*4$ χάρτη χαρακτηριστικών σε $1*1$	40
Εικόνα 2.9 Μετατροπή ενός $3D$ πίνακα σε διάνυσμα με χρήση συνάρτησης συγκέντρωσης.....	41
Εικόνα 2.10 Συνέλιξη σε χρωματισμένες εικόνες.....	42

<u>Εικόνα 3.1 Παραδείγματα Εικόνων Χαμηλού Φωτισμού.....</u>	<u>44</u>
<u>Εικόνα 3.2 Μία συνοπτική αναφορά των σύγχρονων μεθόδων ενίσχυσης εικόνων χαμηλού φωτισμού με χρήση τεχνικών βαθιάς μάθησης.....</u>	<u>45</u>
<u>Εικόνα 3.3 Στατιστική ανάλυση μεθόδων βαθιάς μάθησης.....</u>	<u>45</u>
<u>Εικόνα 3.4 Οπτικά αποτελέσματα της Kind.....</u>	<u>48</u>
<u>Εικόνα 3.5 Αρχιτεκτονική της μεθόδου Kind.....</u>	<u>49</u>
<u>Εικόνα 3.6 Οι χάρτες ανακλαστικότητας και φωτισμού b), c) όπως προκύπτουν από την εικόνα a).....</u>	<u>50</u>
<u>Εικόνα 3.7 Αποσύνθεση Εικόνων Χαμηλού και Υψηλού Φωτισμού.....</u>	<u>52</u>
<u>Εικόνα 3.8 Επαναφορά Ανακλαστικότητας.....</u>	<u>54</u>
<u>Εικόνα 3.9 Παραδείγματα Εικόνων LOL.....</u>	<u>55</u>
<u>Εικόνα 3.10 Οπτική Σύγκριση Μεθόδων Ενίσχυσης.....</u>	<u>56</u>
<u>Εικόνα 3.11 Οπτική Σύγκριση Μεθόδων Ενίσχυσης.....</u>	<u>57</u>
<u>Εικόνα 3.12 Οπτική Σύγκριση Μεθόδων Ενίσχυσης.....</u>	<u>58</u>
<u>Εικόνα 3.13 Οπτική Σύγκριση Μεθόδων Ενίσχυσης.....</u>	<u>59</u>
<u>Εικόνα 3.14 Οπτική Σύγκριση Μεθόδων Ενίσχυσης.....</u>	<u>60</u>
<u>Εικόνα 4.1 Διαφορές AA – Αναγνώρισης Εικόνας.....</u>	<u>61</u>
<u>Εικόνα 4.2 Σύγκριση ακρίβειας Αλγορίθμων AA.....</u>	<u>64</u>
<u>Εικόνα 4.3 Σύγκριση Ταχύτητας Αλγορίθμων AA.....</u>	<u>65</u>
<u>Εικόνα 4.4 Αρχιτεκτονική R-CNN αλγόριθμου.....</u>	<u>66</u>
<u>Εικόνα 4.5 Αρχιτεκτονική Fast R-CNN αλγόριθμου.....</u>	<u>67</u>
<u>Εικόνα 4.6 Αρχιτεκτονική SSD αλγόριθμου.....</u>	<u>68</u>

<u>Εικόνα 4.7 Αρχιτεκτονική VGG.....</u>	<u>68</u>
<u>Εικόνα 4.8 Υπολογισμός δείκτη IoU.....</u>	<u>70</u>
<u>Εικόνα 4.9 Παραδείγματα IoU.....</u>	<u>70</u>
<u>Εικόνα 4.10 Το μοντέλο YOLO.....</u>	<u>71</u>
<u>Εικόνα 4.11 Βήματα εξαγωγής πρόβλεψης του YOLO.....</u>	<u>72</u>
<u>Εικόνα 4.12 Αρχιτεκτονική TNA του αλγόριθμου YOLO.....</u>	<u>73</u>
<u>Εικόνα 5.1 Παραδείγματα Εικόνων ExDark με Ετικέτες Αντικειμένων.....</u>	<u>77</u>
<u>Εικόνα 5.2 Στατιστική Ανάλυση του ExDark.....</u>	<u>80</u>
<u>Εικόνα 5.3 Παραδείγματα Εικόνων Χαμηλού Φωτισμού Διαφόρων Τύπων στο ExDark....</u>	<u>81</u>

[Διαγράψτε αυτή τη σελίδα αν δεν τη χρειάζεστε. Αν έχει εφαρμογή, η αρίθμηση των Διαγραμμάτων/Εικόνων γίνεται με βάση τον αριθμό του κεφαλαίου που ανήκουν π.χ. Διάγραμμα 1.2, Διάγραμμα 1.2, Διάγραμμα 3.1]

ΠΙΝΑΚΑΣ ΣΥΝΤΟΜΟΓΡΑΦΙΩΝ

Μορφοποιήθηκε: Αγγλικά (Ηνωμένων Πολιτειών)

Μορφοποιήθηκε: Αγγλικά (Ηνωμένων Πολιτειών)

<u>TN</u>	<u>Τεχνητή Νοημοσύνη</u>
<u>AI</u>	<u>Artificial Intelligence</u>
<u>ANI</u>	<u>Artificial Narrow Intelligence</u>
<u>AGI</u>	<u>Artificial General Intelligence</u>
<u>ASI</u>	<u>Artificial Super Intelligence</u>
<u>MM</u>	<u>Μηχανική Μάθηση</u>
<u>ML</u>	<u>Machine Learning</u>
<u>DL</u>	<u>Deep Learning</u>
<u>TNΔ</u>	<u>Τεχνητό Νευρωνικό Δίκτυο</u>
<u>DNN</u>	<u>Deep Neural Network</u>
<u>MSE</u>	<u>Mean Squared Error</u>
<u>MO</u>	<u>Μηχανική Όραση</u>
<u>CV</u>	<u>Computer Vision</u>
<u>NLP</u>	<u>Natural Language Processing</u>
<u>ΣΝΔ</u>	<u>Συνελκτικό Νευρωνικό Δίκτυο</u>
<u>CNN</u>	<u>Convolutional Neural Network</u>
<u>YOLO</u>	<u>You Only Look Once</u>
<u>SSD</u>	<u>Single Shot Detector</u>

GAN.....Generative Adversarial Network

ΑΠ.....Αναγνώριση Προσώπου

RGB.....Red Blue Green

SNR.....Signal to Noise Ratio

SL.....Supervised Learning

UL.....Unsupervised Learning

RL.....Reinforcement Learning

SSL.....Semi Supervised Learning

ZSL.....Zero Shot Learning

ReLU.....Rectified Linear Unit

ΑΑ.....Αναγνώριση Αντικειμένων

OD.....Object Detection

MAP.....Mean Average Precision

R-CNN.....Region – Based Convolutional Neural Networks

IoU.....Intersection over Union

ExDark.....Exclusively Dark Dataset

AI.....Artificial Intelligence

ΚΕΦΑΛΑΙΟ 1 - ΕΙΣΑΓΩΓΗ

1.1 Τεχνητή Νοημοσύνη

Ο όρος **τεχνητή νοημοσύνη, TN (Artificial Intelligence, AI)**, αναφέρεται στον κλάδο της πληροφορικής, ο οποίος ασχολείται με τη σχεδίαση και την υλοποίηση αλγορίθμων που μιμούνται στοιχεία της ανθρώπινης συμπεριφοράς, τα οποία υπονοούν έστω και στοιχειώδη ευφυΐα: μάθηση, προσαρμοστικότητα, εξαγωγή συμπερασμάτων, κατανόηση από συμφραζόμενα, επίλυση προβλημάτων κλπ. Ο Τζον Μακάρθι όρισε τον τομέα αυτόν ως *«επιστήμη και μεθοδολογία της δημιουργίας νοημόνων μηχανών»*.

Η τεχνητή νοημοσύνη αποτελεί σημείο τομής μεταξύ πολλαπλών επιστημών όπως της πληροφορικής, της ψυχολογίας, της φιλοσοφίας, της νευρολογίας, της γλωσσολογίας και της επιστήμης μηχανικών, με στόχο τη σύνθεση ευφυούς συμπεριφοράς, με στοιχεία συλλογιστικής, μάθησης και προσαρμογής στο περιβάλλον, ενώ συνήθως εφαρμόζεται σε μηχανές ή υπολογιστές ειδικής κατασκευής. Διαιρείται στη συμβολική τεχνητή νοημοσύνη, η οποία επιχειρεί να εξομοιώσει την ανθρώπινη νοημοσύνη αλγοριθμικά, χρησιμοποιώντας σύμβολα και λογικούς κανόνες υψηλού επιπέδου, και στην υποσυμβολική τεχνητή νοημοσύνη, η οποία προσπαθεί να αναπαράγει την ανθρώπινη ευφυΐα χρησιμοποιώντας στοιχειώδη αριθμητικά μοντέλα, που συνθέτουν επαγωγικά νοήμονες συμπεριφορές με τη διαδοχική αυτοοργάνωση απλούστερων δομικών συστατικών («συμπεριφορική τεχνητή νοημοσύνη»), προσομοιώνουν πραγματικές βιολογικές διαδικασίες όπως η εξέλιξη των ειδών και η λειτουργία του εγκεφάλου («υπολογιστική νοημοσύνη»), ή αποτελούν εφαρμογή στατιστικών μεθοδολογιών σε προβλήματα TN. [1]

1.2 Κατηγορίες Τεχνητής Νοημοσύνης

Ο διαχωρισμός μεταξύ των διαφόρων κατηγοριών της τεχνητής νοημοσύνης προκύπτει από το βαθμό πολυπλοκότητας του προβλήματος.

- **Artificial Narrow Intelligence (ANI – Στενή Τεχνητή Νοημοσύνη)** : Στην κατηγορία αυτή εντάσσονται πρακτικές ΤΝ, οι οποίες λύνουν το πρόβλημα με πολύ περιορισμένες ικανότητες. Κλασσικό παράδειγμα της κατηγορίας αυτής αποτελεί το λογισμικό Siri της Apple, που χρησιμοποιείται στα κινητά τηλέφωνα – smartphones.
- **Artificial General Intelligence (AGI – Γενική Τεχνητή Νοημοσύνη)** : Αναφέρεται σε ένα υπολογιστικό σύστημα – μηχανή, το οποίο έχει παρόμοιες ικανότητες με τον άνθρωπο. Κλασσικό παράδειγμα αποτελεί το Pillo Robot, το οποίο μπορεί να διαγνώσει κάποια αρρώστια και να προτείνει φαρμακευτική αγωγή.
- **Artificial Super Intelligence (ASI – Σουπερ Τεχνητή Νοημοσύνη)** : Αναφέρεται σε μηχανές που υλοποιούν εφαρμογές, τις οποίες ο άνθρωπος αδυνατεί να υλοποιήσει. Η πρώτη προσπάθεια υλοποίησης εφαρμογής AI αυτής της κατηγορίας ήταν το Alpha 2 ρομπότ με το οποίο έγινε προσπάθεια διαχείρισης ενός έξυπνου σπιτιού (smart home).

Οι περισσότερες εφαρμογές AI που υπάρχουν σήμερα εντάσσονται στην κατηγορία της ANI. AGI και ASI βρίσκονται ακόμη σε στάδιο ανάπτυξης. [5]

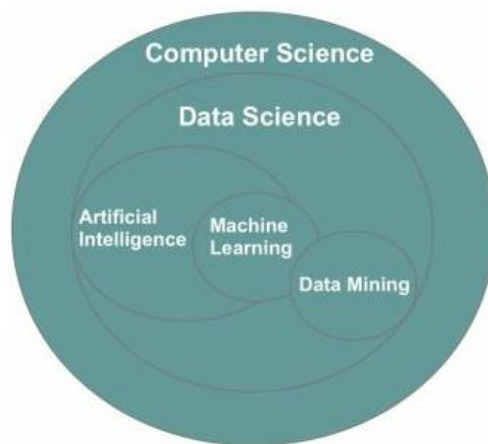
1.3 Μηχανική Μάθηση

Η **μηχανική μάθηση (Machine Learning, ML)**, είναι υποπεδίο της επιστήμης των υπολογιστών που αναπτύχθηκε από τη μελέτη της αναγνώρισης προτύπων και της υπολογιστικής θεωρίας μάθησης στην ΤΝ. Το 1959, ο Άρθουρ Σαμουελ ορίζει τη μηχανική μάθηση ως "Πεδίο μελέτης που δίνει στους υπολογιστές την ικανότητα να μαθαίνουν, χωρίς να έχουν ρητά προγραμματιστεί". Η μηχανική μάθηση διερευνά τη μελέτη και την κατασκευή αλγορίθμων που μπορούν να μαθαίνουν από τα δεδομένα και να κάνουν

προβλέψεις σχετικά με αυτά. Τέτοιοι αλγόριθμοι λειτουργούν κατασκευάζοντας μοντέλα από πειραματικά δεδομένα, προκειμένου να κάνουν προβλέψεις βασισμένες στα δεδομένα ή να εξάγουν αποφάσεις που εκφράζονται ως το αποτέλεσμα.

Η μηχανική μάθηση είναι στενά συνδεδεμένη και συχνά συγχέεται με την υπολογιστική στατιστική δεδομένων – επιστήμη δεδομένων (data science), ένας κλάδος, που επίσης επικεντρώνεται στην πρόβλεψη μέσω της χρήσης των υπολογιστών. Έχει ισχυρούς δεσμούς με την μαθηματική βελτιστοποίηση, η οποία παρέχει μεθόδους, τη θεωρία και τομείς εφαρμογής. Η Μηχανική μάθηση εφαρμόζεται σε μια σειρά από υπολογιστικές εργασίες, όπου τόσο ο σχεδιασμός όσο και ο ρητός προγραμματισμός των αλγορίθμων είναι ανέφικτος. Παραδείγματα εφαρμογών αποτελούν τα φίλτρα spam (spam filtering), η οπτική αναγνώριση χαρακτήρων (OCR), οι μηχανές αναζήτησης και η υπολογιστική όραση. Η Μηχανική μάθηση μερικές φορές συγχέεται με την εξόρυξη δεδομένων (data mining), όπου η τελευταία επικεντρώνεται περισσότερο στην εξερευνητική ανάλυση των δεδομένων, γνωστή και ως μη επιτηρούμενη μάθηση.

Στο πεδίο της ανάλυσης δεδομένων, η μηχανική μάθηση είναι μια μέθοδος που χρησιμοποιείται για την επινόηση πολύπλοκων μοντέλων και αλγορίθμων που οδηγούν στην πρόβλεψη. Τα αναλυτικά μοντέλα επιτρέπουν στους ερευνητές, τους επιστήμονες δεδομένων, τους μηχανικούς και τους αναλυτές να παράγουν αξιόπιστες αποφάσεις και αποτελέσματα και να αναδείξουν αλληλοσυσχετίσεις μέσω της μάθησης από ιστορικές σχέσεις και τάσεις στα δεδομένα. [1][3]

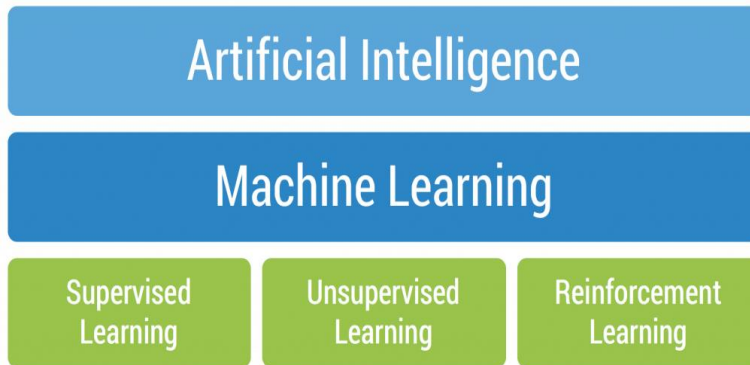


Εικόνα 1.1. Αναπαράσταση των σχέσεων μεταξύ πεδίων της επιστήμης των υπολογιστών.[4]

1.4 Κατηγορίες Μάθησης

Στη μηχανική μάθηση ενσωματώνονται αρκετοί στατιστικοί αλγόριθμοι και η επιλογή του σωστού αλγορίθμου, είναι μία συνεχής πρόκληση για όποιον εργάζεται στο συγκεκριμένο πεδίο. Οι εργασίες μηχανικής μάθησης συνήθως ταξινομούνται σε τρεις μεγάλες κατηγορίες, ανάλογα με τη φύση του εκπαιδευτικού «σήματος» ή την «ανατροφοδότηση» που είναι διαθέσιμα σε ένα σύστημα εκμάθησης. Αυτές είναι:

- **Επιτηρούμενη μάθηση (supervised learning):** Στην επιτηρούμενη μάθηση, βασικός σκοπός είναι η κατασκευή ενός μοντέλου από γνωστά δεδομένα με ετικέτα, το οποίο θα μας βοηθήσει να κάνουμε προβλέψεις από μελλοντικά καινούρια δεδομένα. Ο όρος επιτηρούμενη αναφέρεται στο σετ δεδομένων όπου η επιθυμητή έξοδος είναι ήδη γνωστή. Η κατηγορία αυτή της μηχανικής μάθησης επικεντρώνεται στη σχέση μεταξύ μεταβλητών και γνωστών δεδομένων με ετικέτα.
- **Μη επιτηρούμενη μάθηση (unsupervised learning):** Χωρίς να παρέχεται κάποια εμπειρία στον αλγόριθμο μάθησης, πρέπει να βρεί την δομή των δεδομένων εισόδου και να εξάγει χρήσιμες πληροφορίες. Στην περίπτωση της μη επιτηρούμενης μάθησης, δεν είναι όλες οι μεταβλητές και τα μοτίβα δεδομένων κατηγοριοποιημένα και το υπολογιστικό σύστημα πρέπει να ανακαλύψει κρυφά μοτίβα και να δημιουργήσει δεδομένα με ετικέτα (labels).
- **Ενισχυτική μάθηση (reinforcement learning):** Ένα πρόγραμμα υπολογιστή αλληλεπιδρά με ένα δυναμικό περιβάλλον στο οποίο πρέπει να επιτευχθεί ένας συγκεκριμένος στόχος (όπως η οδήγηση ενός οχήματος), χωρίς κάποιος δάσκαλος να του λέει ρητά αν έχει φτάσει κοντά στο στόχο του. Ένα άλλο παράδειγμα είναι να μάθει να παίζει ένα παιχνίδι εναντίον κάποιου αντιπάλου. [3][4]



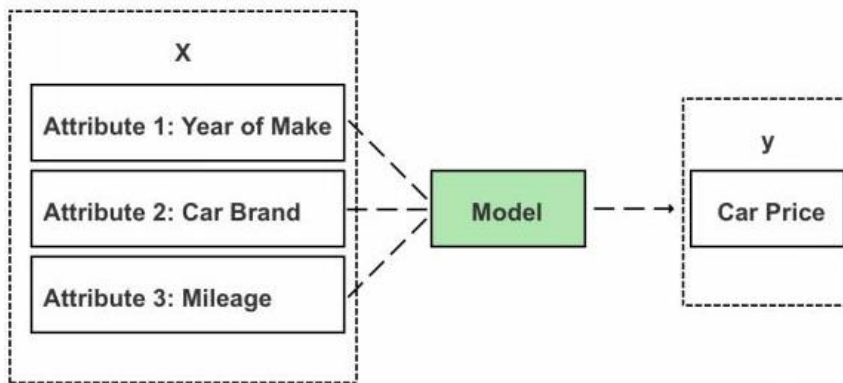
Εικόνα 1.2. Κατηγορίες Μηχανικής Μάθησης. [google]

1.4.1 Επιτηρούμενη Μάθηση – Supervised Learning

Στην επιτηρούμενη μάθηση, το πλήθος δεδομένων N , είναι μία συλλογή από χαρακτηρισμένα παραδείγματα (παραδείγματα με ετικέτα – labeled examples) $\{(x_i, y_i)\}_{i=1}^N$. Κάθε στοιχείο x_i του πλήθους N ονομάζεται διάνυσμα με χαρακτηριστικά (feature vector), για το οποίο διάνυσμα έχουμε μία επιπλέον τιμή που περιγράφει την κατηγορία/ετικέτα του δεδομένου και συμβολίζεται με y_i . Ένα διάνυσμα με χαρακτηριστικά είναι ένα διάνυσμα διάστασης $j = 1, \dots, D$, στο οποίο κάθε διάσταση περιέχει μία τιμή, η οποία με τη σειρά της περιγράφει το παράδειγμα με ετικέτα (labeled example). Η τιμή αυτή ονομάζεται χαρακτηριστικό (feature) και συμβολίζεται με $x^{(j)}$. Για παράδειγμα, αν η μεταβλητή x αναπαριστά ένα μεταχειρισμένο αυτοκίνητο, τότε το πρώτο χαρακτηριστικό $x^{(1)}$, περιέχει το έτος κατασκευής, το δεύτερο χαρακτηριστικό $x^{(2)}$, περιέχει τη μάρκα του, το τρίτο χαρακτηριστικό $x^{(3)}$, περιέχει τον αριθμό χιλιομέτρων κ.ο.κ. Για όλα τα παραδείγματα στο πλήθος δεδομένων N , οι τιμές $x^{(j)}$ περιέχουν πάντα κάποιο είδος πληροφορίας. Η επιγραφή (label) y_i είναι ένα στοιχείο, το οποίο περιέχεται σε ένα πεπερασμένο αριθμό κλάσεων $\{1, 2, \dots, C\}$, ή μπορεί να είναι ένας πραγματικός αριθμός ή ένα διάνυσμα, πίνακας, γράφημα κ.ο.κ.

Ο αλγόριθμος της επιτηρούμενης μάθησης υλοποιείται εισάγοντας στο υπολογιστικό σύστημα ένα δείγμα δεδομένων (sample data) με διάφορα χαρακτηριστικά ($X^{(j)}$) και τη σωστή τιμή για το κάθε χαρακτηριστικό ($Y^{(j)}$). Σκοπός του αλγόριθμου είναι να

χρησιμοποιήσει αυτό το δείγμα δεδομένων και να παράξει ένα μοντέλο – συνάρτηση, το οποίο θα δέχεται ως είσοδο διανύσματα με χαρακτηριστικά (feature vectors) και θα παράγει στην έξοδο πληροφορίες που προσδίδουν μία ετικέτα στα συγκεκριμένα διανύσματα. Το γεγονός ότι η έξοδος και οι τιμές των χαρακτηριστικών είναι ήδη γνωστές, χαρακτηρίζει το σετ δεδομένων ως «με ετικέτα» (labeled). Ο αλγόριθμος ανακαλύπτει κανόνες και μοτίβα τα οποία μπορεί να υπάρχουν στα δεδομένα και δημιουργεί ένα μοντέλο. Ως μοντέλο ορίζουμε εκείνη τη μαθηματική συνάρτηση, η οποία παράγει ένα αποτέλεσμα το οποίο επαληθεύει τα δεδομένα με ετικέτα. Μόλις το μοντέλο υλοποιηθεί, τότε μπορεί να εφαρμοσθεί σε νέα δεδομένα και θα εξεταστεί η ακρίβειά του. Αν περάσει επιτυχώς την εξέταση ακρίβειας, τότε είναι έτοιμο να εφαρμοσθεί σε δεδομένα του πραγματικού κόσμου.

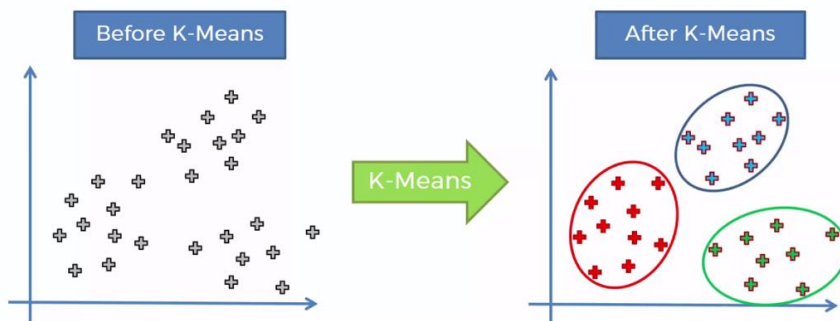


Εικόνα 1.3. Μοντέλο πρόβλεψης τιμής πώλησης αυτοκινήτου.[4]

1.4.2 Μη Επιτηρούμενη Μάθηση – Unsupervised Learning

Στην μη επιτηρούμενη μάθηση, το σετ δεδομένων αποτελεί μία συλλογή από παραδείγματα χωρίς ετικέτα $\{x_i\}_{i=1}^N$, όπου x ένα διάνυσμα με χαρακτηριστικά (feature vector). Σκοπός της μη επιτηρούμενης μάθησης είναι η δημιουργία ενός μοντέλου που δέχεται ως είσοδο ένα διάνυσμα x και είτε το μετατρέπει σε ένα άλλο διάνυσμα, είτε σε μία τιμή για την επίλυση ενός προβλήματος. Στην επιτηρούμενη μάθηση δεν έχουν ετικέτα όλα τα δεδομένα και οι μεταβλητές, οπότε οι αλγόριθμοι μη επιτηρούμενης μάθησης προσπαθούν να εντοπίσουν κρυμμένα μοτίβα και δομές στα δεδομένα, έτσι ώστε να τα ταξινομήσουν και να δημιουργήσουν ετικέτες.

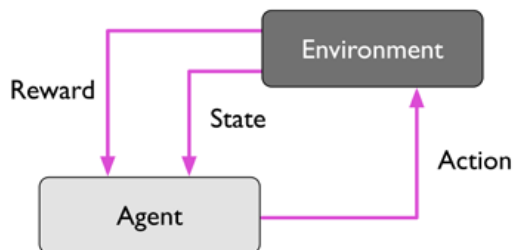
Ο αλγόριθμος *k-means*[14] είναι ένα κλασικό παράδειγμα ομαδοποίησης (clustering), ο οποίος χρησιμοποιείται στη μη επιτηρούμενη μάθηση. Ο συγκεκριμένος αλγόριθμος ομαδοποιεί δεδομένα που έχουν παρόμοια χαρακτηριστικά και ανακαλύπτει μοτίβα, τα οποία αρχικά ήταν άγνωστα.



Εικόνα 1.4. Ομαδοποίηση δεδομένων μετά την εφαρμογή του αλγορίθμου *k-means*. [google]

1.4.3 Ενισχυτική Μάθηση – Reinforcement Learning

Στην ενισχυτική μάθηση, οι αλγόριθμοι «ζούνε» μέσα σε ένα περιβάλλον, στο οποίο μπορούν να εκτελούν ενέργειες, ανάλογα με την κατάσταση του περιβάλλοντος. Διαφορετικές ενέργειες δίνουν διαφορετικά έπαθλα, τα οποία οδηγούν σε διαφορετικές καταστάσεις. Σκοπός της ενισχυτικής μάθησης είναι οι αλγόριθμοι να καταφέρουν να αντιληφθούν και να καταλάβουν μία πολιτική (policy). Η πολιτική είναι μία συνάρτηση, όμοια με το μοντέλο στην επιτηρούμενη μάθηση, η οποία δέχεται ως είσοδο ένα διάνυσμα με χαρακτηριστικά και εξάγει μία βέλτιστη ενέργεια, η οποία εκτελείται σε αυτή την κατάσταση του περιβάλλοντος. Η πολιτική αυτή είναι βέλτιστη, όταν δέχεται το επιθυμητό μέσο έπαθλο. Σε αντίθεση με την επιτηρούμενη και την μη επιτηρούμενη μάθηση, η ενισχυτική μάθηση διαρκώς εξελίσσει και βελτιώνει το μοντέλο – πολιτική της.[3]



Εικόνα 1.5. Οπτική αναπαράσταση της ενισχυτικής μάθησης. [google]

1.5 Βαθιά Μάθηση

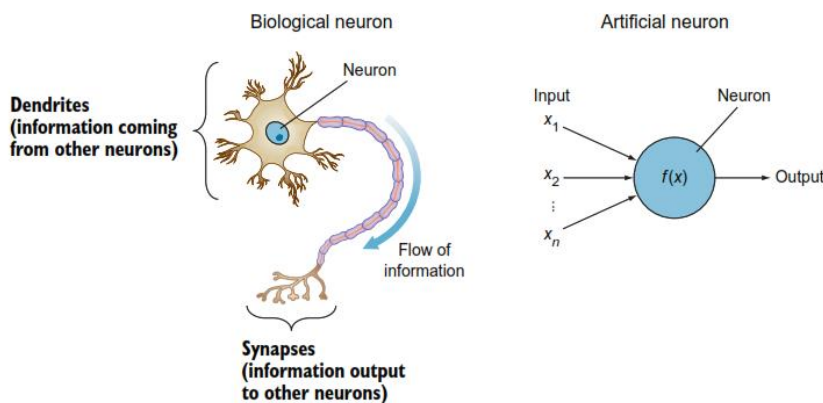
Η **βαθιά μάθηση (Deep Learning, DL)**, είναι το υποπεδίο της μηχανικής μάθησης που εστιάζει στη γνώση σημαντικών χαρακτηριστικών από δεδομένα και ειδικότερα σε πολύπλοκα δεδομένα ή δομές δεδομένων όπως βίντεο και εικόνες. Παρότι η βαθιά μάθηση έγινε διάσημη τα τελευταία χρόνια, η θεωρία της εξελίσσεται από τη δεκαετία του 1940[15]. Μάλιστα η ονομασία της ποικίλλει στο πέρασμα των ετών, καθώς στη δεκαετία του 1970 ονομαζόταν κυβερνητική (cybernetics), στα μέσα του 1980 ονομαζόταν συνδετικότητα (connectionism) και σήμερα έχει καταλήξει σε βαθιά μάθηση ή νευρωνικά δίκτυα.

Η βαθιά μάθηση μιμείται τη λειτουργία του ανθρώπινου εγκεφάλου. Επεξεργάζεται δεδομένα με κάποιες υπολογιστικές μονάδες που λέγονται νευρώνες, τα οποία διατάσσονται σε άλλους τομείς, που λέγονται συνδέσεις (layers). Ο κύριος αλγόριθμος της βαθιάς μάθησης είναι τα νευρωνικά δίκτυα (neural networks).

1.6 Τεχνητά Νευρωνικά Δίκτυα

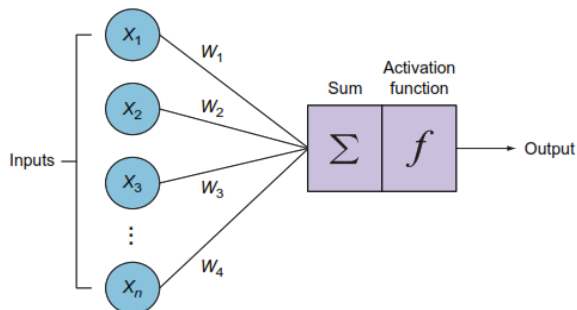
Ένα **τεχνητό νευρωνικό δίκτυο, ΤΝΔ**, είναι ένας αλγόριθμος τεχνητής νοημοσύνης που εμπνέεται από τον τρόπο λειτουργίας του ανθρώπινου εγκεφάλου. Ουσιαστικά ένα ΤΝΔ μιμείται τον τρόπο σύνδεσης και λειτουργίας των νευρώνων του ανθρώπινου εγκεφάλου. Ο ανθρώπινος εγκέφαλος περιέχει διασυνδεδεμένους νευρώνες με δενδρίτες, οι οποίοι δέχονται ερεθίσματα (inputs) και τα μετατρέπουν σε ηλεκτρικά σήματα (outputs), τα οποία διαδίδονται μέσω του νευρικού συστήματος σε άλλους νευρώνες και στον εγκέφαλο. Ένα

ΤΝΔ αναπαριστά μία μαθηματική συνάρτηση, η οποία δέχεται ως εισόδους (ερεθίσματα – inputs) μία συλλογή από τανυστές και προβλέπει έναν ή περισσότερους τανυστές ως έξοδο (σήμα – outputs). Οι μαθηματικές εξισώσεις με τις οποίες συνδέονται οι εισοδοί με τις εξόδους είναι αποτέλεσμα της αρχιτεκτονικής του ΤΝΔ, η οποία είναι προσαρμόσιμη στο πρόβλημα που χρειάζεται επίλυση ή στο είδος δεδομένων που είναι προς επεξεργασία. Άλλη αρχιτεκτονική χρησιμοποιείται σε δομημένα δεδομένα και άλλη σε μη δομημένα όπως είναι τα δεδομένα εικόνας και ήχου.



Εικόνα 1.6. Αναπαράσταση βιολογικού και τεχνητού νευρώνα.[6]

Ένα ΤΝΔ αποτελείται από διασυνδεδεμένους νευρώνες που ονομάζονται κόμβοι (nodes), οι οποίοι αλληλεπιδρούν μεταξύ τους μέσω αξόνων που ονομάζονται κορυφές (edges). Οι κόμβοι σε ένα ΤΝΔ διατάσσονται σε στρώματα (layers). Το πρώτο στρώμα (input layer) περιέχει ακατέργαστα δεδομένα (raw data) όπως αριθμητικές τιμές ή κείμενο ή εικόνες, τα οποία διαχωρίζονται στους κόμβους. Κάθε κόμβος στέλνει την πληροφορία στον κόμβο εξόδου (output layer) μέσω των κορυφών. Κάθε κορυφή έχει ένα αριθμητικό βάρος, το οποίο διαφέρει από κορυφή σε κορυφή και μπορεί να αλλάζει και να προσαρμόζεται με βάση την εμπειρία. Το βάρος αυτό υπάρχει για να εξασφαλίζει τη σωστή λειτουργία της συνάρτησης ενεργοποίησης έτσι ώστε να επιτυγχάνεται το κατώφλι επιτυχίας (threshold).



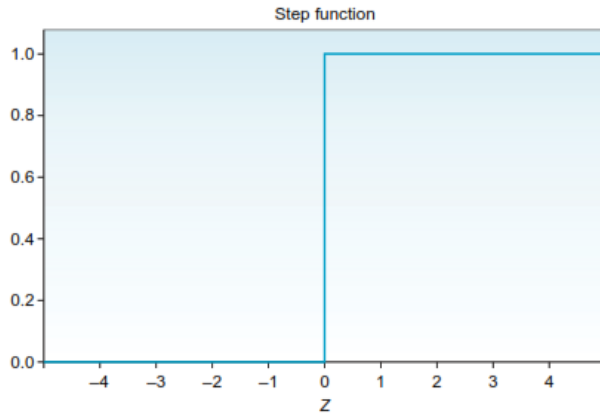
Εικόνα 1.7. Μαθηματική οπτικοποίηση ενός ΤΝΔ.[6]

Τα X_1, X_2, \dots, X_n συμβολίζουν τις μεταβλητές εισόδου και τα W_1, W_2, \dots, W_n συμβολίζουν τα βάρη που αντιστοιχούν σε κάθε μεταβλητή. Η τιμή εξόδου (output) υπολογίζεται ως εξής :

$$z = \sum X_i \cdot W_i + b \text{ (bias)}$$

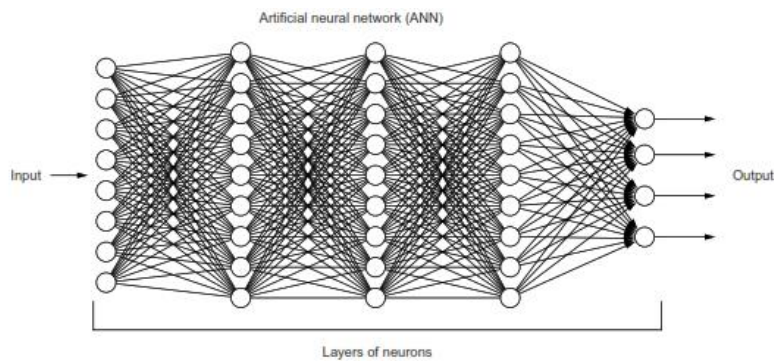
$$Z = x_1 \cdot w_1 + x_2 \cdot w_2 + x_3 \cdot w_3 + \dots + x_n \cdot w_n + b$$

Σε ένα ΤΝΔ η τιμή εξόδου υπολογίζεται από το άθροισμα του γινομένου των εισόδων με τα βάρη τους συν έναν όρο b (bias - παράμετρος), που είναι όμοιος με τις παραμέτρους που χρησιμοποιούμε στην γραμμική παλινδρόμηση (linear regression). Στη συνέχεια, η συνάρτηση ενεργοποίησης λαμβάνει αυτό το άθροισμα και το συγκρίνει με το επιθυμητό κατώφλι τιμών και ενεργοποιεί ή απενεργοποιεί τον νευρώνα. Η απλούστερη συνάρτηση ενεργοποίησης που χρησιμοποιείται στα ΤΝΔ είναι η βηματική συνάρτηση Heaviside $H(x)$ (step function) που παράγει δυαδικές τιμές (0 ή 1). Η συνάρτηση Heaviside συγκρίνει το άθροισμα γινομένου και αν αυτό είναι ≥ 0 , τότε ενεργοποιεί τον νευρώνα (έξοδος = 1). Αν το άθροισμα είναι < 0 , τότε απενεργοποιεί τον νευρώνα (έξοδος = 0).



Εικόνα 1.8. $H(x) = \begin{cases} 0 & \text{για } x < 0 \\ 1 & \text{για } x \geq 0 \end{cases}$ [6]

Στην περίπτωση των ΤΝΔ βαθιάς μάθησης (**Deep Neural Networks - DNNs**), ανάμεσα από τον κόμβο εισόδου και τον κόμβο εξόδου, μεσολαβούν κρυφά ενδιάμεσα στρώματα (hidden layers), τα οποία συνδέουν την είσοδο με την έξοδο ενώ ταυτόχρονα υλοποιούν πράξεις και μετατροπές στα δεδομένα εισόδου, πριν αυτά φτάσουν στην έξοδο. Η λειτουργικότητα και η επιτυχία ενός DNN καθορίζεται από την καταλληλότερη συνάρτηση ενεργοποίησης, η οποία τροποποιεί τα δεδομένα και τις τιμές τους σε κάθε κόμβο.[7]

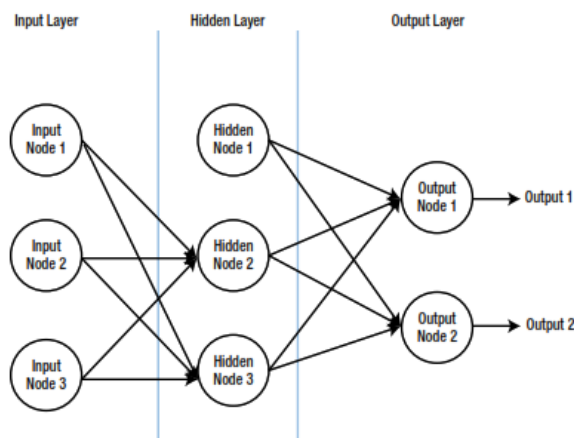


Εικόνα 1.97. Νευρωνικό Δίκτυο Βαθιάς Μάθησης. [6]

1.6.1 Αρχιτεκτονική ενός ΤΝΔ Βαθιάς Μάθησης

Τα απλά ΤΝΔ πετυχαίνουν ακριβείς προβλέψεις σε γραμμικά σετ δεδομένων, αλλά αν αυξηθεί η πολυπλοκότητα των δεδομένων, τότε αυξάνεται η πολυπλοκότητα των ΤΝΔ και χρησιμοποιούμε αρχιτεκτονική βαθιάς μάθησης. Τα βασικά συστατικά μιας αρχιτεκτονικής ΤΝΔ βαθιάς μάθησης είναι τα εξής :

- *Στρώμα εισόδου (input layer)* : Το στρώμα αυτό δέχεται τιμές μεταβλητών – δεδομένα – διανύσματα χαρακτηριστικών ως είσοδο.
- *Κρυφά στρώματα (hidden layers)* : Στο στρώμα αυτό, στοιβάζονται οι νευρώνες και ενώνουν την είσοδο του ΤΝΔ με την έξοδο, ενώ ταυτόχρονα πραγματοποιούν μετασχηματισμούς των δεδομένων εισόδου.
- *Κορυφές – συνδέσεις βαρών (edges)* : Από τη θεωρία γραφημάτων είναι γνωστές ως κορυφές που ενώνουν τους κόμβους. Ο αριθμός των κόμβων και των κορυφών ενός ΤΝΔ καθορίζεται από την εργασία που είναι προς εκτέλεση. Σε κάθε κορυφή ανατίθεται μία τιμή βάρους , η οποία καθορίζει την επίδραση που θα έχει το βάρος στην έξοδο.
- *Στρώμα εξόδου (output layer)* : Περιέχει τις τιμές των μεταβλητών εισόδου που προέβλεψε το ΤΝΔ.[5][6][7]



Εικόνα 1.10-8. ΤΝΔ με στρώμα εισόδου τριών κόμβων.[5]

1.7 Εκπαίδευση Δικτύου Βαθιάς Μάθησης

Τα απλά ΤΝΔ είναι ιδανικά για υλοποιήσεις γραμμικών μοντέλων και χρησιμοποιούν την ανθρώπινη λογική για να εκπαιδευτούν, δηλαδή τη λογική της δοκιμής και του λάθους και την απόκτηση εμπειρίας.

1.7.1 Βήματα Εκμάθησης και Αλγόριθμοι Βελτιστοποίησης ενός ΤΝΔ

Η λογική εκμάθησης ενός απλού ΤΝΔ είναι η εξής :

- I. Διαδικασία προώθησης (**feedforward process**) : Ο νευρώνας υπολογίζει το άθροισμα συμπεριλαμβανομένων των βαρών και εφαρμόζει τη συνάρτηση ενεργοποίησης κάνοντας μία πρόβλεψη \hat{y} .

$$\hat{y} = \text{activation}(\sum x_i \cdot w_i + b)$$

Υπάρχουν κι άλλες συναρτήσεις ενεργοποίησης που χρησιμοποιούνται συχνά :

$$\text{Sigmoid activation}(x) = \frac{1}{1 + e^{-x}}$$

$$\text{ReLU activation}(x) = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}$$

$$\text{Tanh activation}(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

- II. Υπολογισμός σφάλματος μέσω της σύγκρισης της τιμής της πρόβλεψης \hat{y} και της σωστής ετικέτας y :

$$\text{error} = y - \hat{y}$$

Η παραπάνω σχέση αποτελεί μία γενική αποτύπωση του όρου του σφάλματος. Στην πραγματικότητα χρησιμοποιούνται συναρτήσεις σφάλματος (**error functions**) για να ποσοτικοποιήσουν τη λανθασμένη πρόβλεψη του ΤΝΔ, σε σχέση με την αναμενόμενη τιμή εξόδου. Οι συναρτήσεις σφάλματος διαδραματίζουν καθοριστικό ρόλο στην επίτευξη της μέγιστης βελτιστοποίησης. Η επιλογή της κατάλληλης

συνάρτησης σφάλματος γίνεται ανάλογα με το είδος της προβλεπόμενης τιμής εξόδου.

- Αν η προβλεπόμενη τιμή λαμβάνει συνεχείς τιμές, τότε χρησιμοποιείται το μέσο τετραγωνικό σφάλμα (**mean squared error – MSE**) με συνάρτηση :

$$J_{\theta} = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2$$

$$\hat{y}_i = \eta_{\theta}(x_i)$$

όπου y_i είναι η πραγματική τιμή, \hat{y}_i η πρόβλεψη, η_{θ} η συνάρτηση ενεργοποίησης, x_i η τιμή εισόδου και m ο αριθμός των σειρών στο σετ δεδομένων.

- Αν η προβλεπόμενη τιμή λαμβάνει διακριτές τιμές, τότε χρησιμοποιείται η κατηγορηματική διασταυρούμενη εντροπία (**categorical cross – entropy**) με συνάρτηση :

$$-\frac{1}{m} \sum_{j=1}^C \sum_{i=1}^m y_i \log(p_i)$$

- Αν η προβλεπόμενη τιμή λαμβάνει δυαδικές διακριτές τιμές, τότε χρησιμοποιείται η δυαδική διασταυρούμενη εντροπία (**binary cross – entropy**) με συνάρτηση :

$$-\frac{1}{m} \sum_{i=1}^m (y_i \log(p_i) + (1 - y_i) \log(1 - p_i))$$

όπου y είναι η πραγματική τιμή, p η πρόβλεψη, m ο συνολικός αριθμός των σημείων – δεδομένων και C ο συνολικός αριθμός των κλάσεων.

- III. Αλλαγή των τιμών των αρχικών βαρών, έτσι ώστε να μειωθεί η τιμή του σφάλματος. Η επιδιόρθωση επιτυγχάνεται με υλοποίηση του αλγορίθμου της οπισθοδιάδοσης (**backpropagation algorithm**[16]), καθώς οι τιμές των βαρών αναβαθμίζονται προς τα πίσω, με βάση το σφάλμα που έχει υπολογισθεί στο στρώμα εξόδου. Μόλις υπολογισθεί το σφάλμα εφαρμόζεται η μέθοδος της βαθμιαίας κατάβασης για επανυπολογισμό των βαρών. Η μέθοδος της βαθμιαίας κατάβασης (**gradient**

descent) είναι μία πολύ γνωστή τεχνική βελτιστοποίησης που χρησιμοποιείται σε πολλά μοντέλα μηχανικής μάθησης και διενεργείται μέχρι να επιτευχθεί η μέγιστη βελτιστοποίηση, δηλαδή το ελάχιστο κόστος.

- IV. Επανάληψη. Ο αριθμός των επαναλήψεων καθορίζεται από την ελαχιστοποίηση του σφάλματος. Σκοπός είναι το σφάλμα να τείνει στο 0, κάτι που σημαίνει πως το TNΔ έχει επιτύχει πρόβλεψη πολύ κοντά στη σωστή τιμή. Στο σημείο αυτό γίνεται καταγραφή των τιμών των βαρών για μελλοντική χρήση.

ΚΕΦΑΛΑΙΟ 2 – ΜΗΧΑΝΙΚΗ ΟΡΑΣΗ

2.1 Μηχανική Όραση

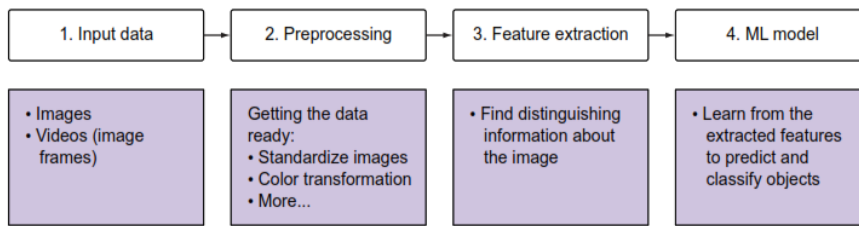
Η **Μηχανική Όραση, MO (Computer Vision - CV)**, είναι το πεδίο της επιστήμης των υπολογιστών, το οποίο ασχολείται με τη μελέτη της πολυπλοκότητας του συστήματος της ανθρώπινης όρασης και βοηθά τους υπολογιστές να αναγνωρίσουν και να επεξεργαστούν δεδομένα σε εικόνες και βίντεο, όπως ακριβώς ο άνθρωπος. Χάρη στην αλματώδη πρόοδο στο πεδίο της βαθιάς μάθησης και των ΤΝΔ, η ΜΟ έχει προχωρήσει αρκετά βήματα και έχει καταφέρει να ξεπεράσει τον άνθρωπο σε εφαρμογές που σχετίζονται με την αναγνώριση αντικειμένων. Καταλυτικό ρόλο σε αυτή την πρόοδο διαδραματίζει και ο τεράστιος όγκος δεδομένων που παράγουμε σήμερα (έχει υπολογιστεί πως περίπου 3 δισεκατομμύρια εικόνες κοινοποιούνται καθημερινά), τα οποία στη συνέχεια χρησιμοποιούνται για να εκπαιδεύσουμε τα ΤΝΔ και να αναπτύξουμε περαιτέρω την ΜΟ. Ταυτόχρονα, η υπολογιστική δύναμη που απαιτείται για αυτού του είδους τις εφαρμογές είναι πια προσιτή.

Τα ΤΝΔ βαθιάς μάθησης ανέπτυξαν περαιτέρω το πεδίο της ΜΟ και σε συνεργασία με το πεδίο της επεξεργασίας φυσικής γλώσσας (**Natural Language Processing - NLP**) προέκυψαν τεχνικές ανάλυσης κειμένου μέσω των εικόνων και των βίντεο. Επίσης τα ΤΝΔ βαθιάς μάθησης τελειοποίησαν τη λήψη φωτογραφιών, καθώς με τη χρήση τους μπορούμε πια να εξαφανίσουμε τυχόν θορύβους που εμφανίζονται στο υπόβαθρο, μπορούμε επίσης να τονίσουμε κάποια σημεία ή αντικείμενα σε εικόνες και τέλος μπορούμε να σταθεροποιήσουμε τη λήψη «κουνημένων» φωτογραφιών.

2.2 Εφαρμογές της Μηχανικής Υπολογιστικής Όρασης

Οι εφαρμογές της ΜΟ που χρησιμοποιούνται στον πραγματικό κόσμο και συγκεκριμένα στις επιχειρήσεις, στην ψυχαγωγία, στις μετακινήσεις και στην υγεία, καθορίζουν το πόσο σημαντική είναι η ΜΟ. Καθοριστικός παράγοντας στην ανάπτυξη αυτών των εφαρμογών αποτελεί η πληροφορία που εξάγεται από τα έξυπνα τηλέφωνα (smartphones), τις κάμερες

ασφαλείας, τις κάμερες οδικής κυκλοφορίας ή οποιαδήποτε άλλη συσκευή εικόνας. Η πληροφορία αυτή εισάγεται σε ένα υπολογιστικό σύστημα όρασης, το οποίο με τη σειρά του αναλύει και επεξεργάζεται αυτή την πληροφορία – εικόνα, υλοποιώντας συγκεκριμένη σειρά βημάτων που ονομάζεται αλληλουχία MO (**Computer Vision Pipeline**).



Εικόνα 2.1. Αλληλουχία βημάτων της MO.[6]

2.2.1 Ταξινόμηση Εικόνων

Στην ταξινόμηση εικόνων (image classification), το υπολογιστικό σύστημα αναθέτει μία ετικέτα σε μία εικόνα, η οποία ετικέτα προέρχεται από ήδη χαρακτηρισμένο σετ εικόνων, δηλαδή το υπολογιστικό σύστημα ταξινομεί μία άγνωστη σε αυτό εικόνα, σύμφωνα με την εμπειρία που έχει αποκτήσει ύστερα από εκπαίδευση σε εικόνες με ετικέτα (labeled images). Η ταξινόμηση εικόνων χρησιμοποιεί ΤΝΔ και συγκεκριμένα **Συνελκτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks – CNNs [17])** και χρησιμοποιείται σε εφαρμογές όπως :

- Διάγνωση Καρκίνου του Πνεύμονα
- Αναγνώριση Σημάτων Οδικής Κυκλοφορίας

2.2.2 Εντοπισμός και Αναγνώριση Αντικειμένων

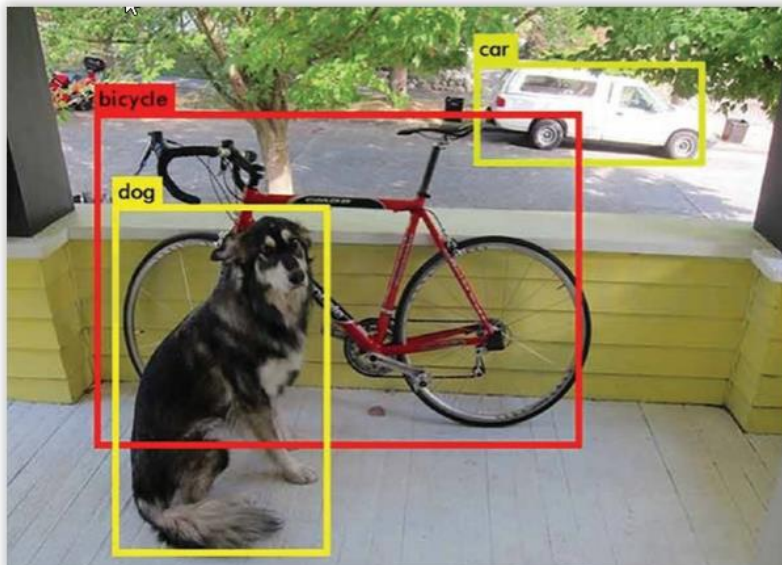
Η ταξινόμηση εικόνων είναι πολύ χρήσιμη στην αναγνώριση εικόνων, στις οποίες υπάρχει μόνο ένα αντικείμενο, αλλά αν στοχεύουμε να φτάσουμε το επίπεδο της ανθρώπινης όρασης, θα πρέπει να αναπτύξουμε πιο πολύπλοκα ΤΝΔ και υπολογιστικά συστήματα, έτσι

ώστε να μπορούν να αναγνωρίζουν περισσότερα αντικείμενα σε μία εικόνα. Τέτοια συστήματα αναγνώρισης αντικειμένων είναι τα :

- **YOLO** (You Only Look Once) [\[34\]](#)
- **SSD** (Single Shot Detector) [\[21\]](#)

Η αναγνώριση αντικειμένων είναι κομβικής σημασίας τεχνολογία για την εξέλιξη της βιομηχανίας των αυτοκινούμενων οχημάτων και ειδικότερα στην ανάπτυξη της ασφάλειάς τους και της λειτουργικότητάς τους. Τα αυτοκινούμενα οχήματα είναι εξοπλισμένα με αισθητήρες (sensors) και κάμερες για να συλλέγουν μεγάλο όγκο δεδομένων εικόνας από το περιβάλλον γύρω τους, τα οποία καθορίζουν τη συμπεριφορά του αυτοκινήτου στο δρόμο.

Επιπροσθέτως, η αναγνώριση αντικειμένων έχει εξαιρετικές εφαρμογές στον τομέα της υγείας και της βιοϊατρικής, καθώς οι φωτογραφίες των ανθρώπινων οργάνων επεξεργάζονται με αλγορίθμους MO για να είναι φωτεινότερες και ευκρινέστερες, έτσι ώστε να υπάρχει καλύτερη πρόληψη σε περιπτώσεις καρκινικών μεταλλάξεων.



Εικόνα 2.2. Εύρεση και αναγνώριση αντικειμένων σε πραγματικό χρόνο.[5]

2.2.3 Δημιουργία Εικόνων

Το 2014 αποτέλεσε χρονιά σταθμός για την εξέλιξη των TND και των CNNs, καθώς ο Ian Goodfellow εφηύρε τα **GANs (Generative Adversarial Networks [18])**, τα οποία αποτελούν εξέλιξη των CNNs και μπορούν από ένα γνωστό σετ εικόνων να δημιουργήσουν νέες εικόνες. Η χρησιμότητα των GANs είναι πάρα πολύ σημαντική, καθώς θα βοηθήσει στον ανασχηματισμό παλιών έργων τέχνης, τα οποία έχουν αλλοιωθεί λόγω παλαιότητας.

2.2.4 Αναγνώριση Προσώπων

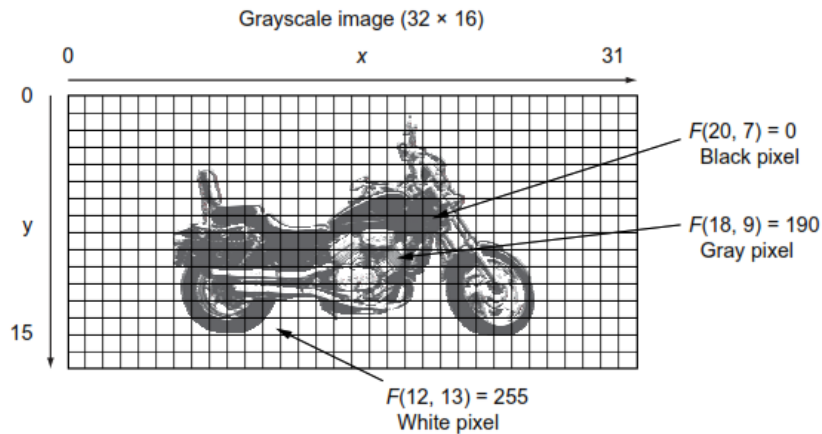
Η **αναγνώριση προσώπων**, **ΑΠ** μας επιτρέπει την αναγνώριση ενός ατόμου σε μία φωτογραφία ή σε ένα βίντεο και χωρίζεται σε δύο στάδια :

- Ταυτοποίηση Προσώπου
- Επαλήθευση Προσώπου

Η τεχνολογία αυτή είναι χρήσιμη σε συστήματα συναγερμού, όπου η αναγνώριση προσώπου είναι πολύ σημαντική για το ξεκλείδωμά τους, όπως πχ. στα έξυπνα τηλέφωνα.

2.3 Εικόνα – Αναπαράσταση Εικόνων

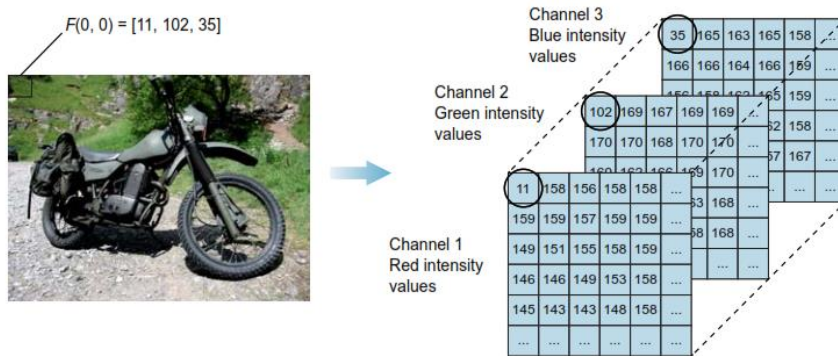
Στον πυρήνα της ΜΟ συναντάμε την εικόνα. Μία ψηφιακή εικόνα, με επέκταση “JPEG” ή “jpg” ή “PNG”, είναι ένα σύνολο από εικονοστοιχεία, τα οποία αποτελούν θεμελιώδη στοιχεία μιας εικόνας. Ουσιαστικά μία εικόνα είναι ένα πλέγμα από εικονοστοιχεία που το καθένα παίρνει τιμές ανάλογα με τη φωτεινότητά του. Στην κλίμακα του γκρι, σε κάθε εικονοστοιχείο αναλογεί μία τιμή μεταξύ του 0 και του 255, όπου στην τιμή 0 αντιστοιχεί το μαύρο, στη τιμή 255 το λευκό και σε οποιαδήποτε ενδιάμεση τιμή το γκρι. Όσο μικρότερη είναι η τιμή του εικονοστοιχείου, τόσο πιο σκοτεινό είναι και όσο μεγαλύτερη η τιμή του, τόσο πιο φωτεινό.



Μία εικόνα έχει ύψος(y) * πλάτος(x) * βάθος(c), όπου y ο αριθμός των σειρών (rows), όπου x ο αριθμός των στηλών (columns) και c ο αριθμός των καναλιών. Στις ασπρόμαυρες εικόνες το $c = 1$, ενώ στις πολύχρωμες το $c = 3$, ένα για το κόκκινο, ένα για το μπλέ και ένα για το πράσινο (αναπαράσταση RGB). Συνεπώς κάθε εικόνα μπορεί να αναπαρασταθεί και ως μία συνάρτηση δύο μεταβλητών $F(x,y)$ ή ως ένα διάνυσμα δύο διαστάσεων ή ως ένας πίνακας με πεπερασμένο αριθμό από γραμμές και στήλες.

Στην εικόνα 2.3 απεικονίζεται μία ασπρόμαυρη εικόνα, η οποία έχει διαστάσεις 32*16, που σημαίνει πως η εικόνα έχει πλάτος 32 pixels και ύψος 16 pixels και συνολικά 512 pixels. Επίσης παρατηρούμε και την αναπαράσταση τριών διαφορετικών pixel, καθώς και την τιμή τους. Για παράδειγμα το pixel που έχει $x = 20$ και $y = 7$ αναπαριστάται ως $F(20,7)$ και παίρνει την τιμή 0, αφού είναι μαύρο.

Στις έγχρωμες εικόνες η αναπαράσταση των pixel δεν καθορίζεται μόνο από μία τιμή, αλλά από τρεις, μία για το κόκκινο (Red), μία για το πράσινο (Green) και μία για το μπλέ (Blue), το οποίο είναι γνωστό ως σύστημα RGB.



Εικόνα 2.4. Αναπαράσταση έγχρωμης εικόνας [6].

Στην εικόνα 2.4 απεικονίζεται μία πολύχρωμη εικόνα και η αναπαράστασή της σε πίνακα 3 διαστάσεων (3D matrix), μία διάσταση για κάθε κανάλι. Στον πρώτο πίνακα παρατηρούμε την αναπαράσταση του κόκκινου καναλιού με τις τιμές του κάθε pixel (0 έως 255), στον δεύτερο το πράσινο και στον τρίτο το μπλέ κανάλι. Το pixel στην κορυφή αριστερά αντιστοιχεί σε πράσινο γρασίδι, γι αυτό και η τιμή του pixel στο πράσινο είναι 102 και στα άλλα χρώματα 11 και 35 αντίστοιχα.

2.4 Προεπεξεργασία Εικόνας

Στην επεξεργασία εικόνας, αρκετά συχνά απαιτείται προεπεξεργασία – καθαρισμός των δεδομένων πριν αυτά εισαχθούν στο μοντέλο εκμάθησης, διότι τα δεδομένα αρχικά είναι ακατάστατα και προέρχονται από διαφορετικές πηγές. Με αυτό τον τρόπο επιτυγχάνεται ευκολότερη και ακριβέστερη ανάλυση των δεδομένων από το μοντέλο μηχανικής μάθησης. Το ίδιο συμβαίνει και με τις εικόνες, οι οποίες προεπεξεργάζονται ανάλογα με το πρόβλημα που πρόκειται να αντιμετωπίσουμε. Η προεπεξεργασία εικόνων έχει στόχο να μειώσει την πολυπλοκότητα και να αυξήσει την ακρίβεια του εφαρμοζόμενου αλγόριθμου.

2.4.1 Τεχνικές Προεπεξεργασίας Εικόνας

Τεχνικές προεπεξεργασίας εικόνας αποτελούν η αλλαγή μεγέθους των εικόνων (image resizing), η γεωμετρική και χρωματική μεταμόρφωση, η μετατροπή μιας χρωματισμένης εικόνας σε ασπρόμαυρη κ.α.

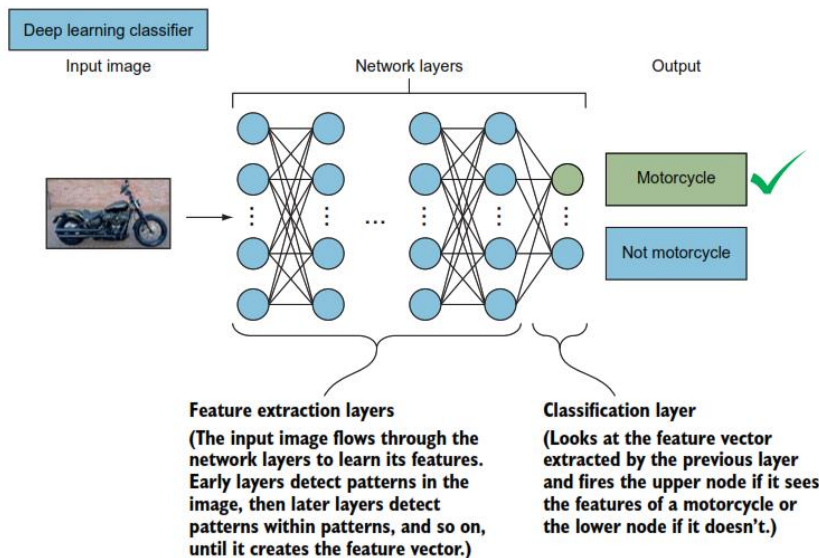
- **Μετατροπή έγχρωμης εικόνας σε ασπρόμαυρη:** Η τεχνική αυτή χρησιμοποιείται σε προβλήματα αναγνώρισης συγκεκριμένων αντικειμένων σε φωτογραφίες. Οι χρωματισμένες φωτογραφίες περιέχουν περισσότερες πληροφορίες από τις ασπρόμαυρες, με αποτέλεσμα να αυξάνεται η πολυπλοκότητα και να χρειάζεται περισσότερη μνήμη. Μετατρέποντας τη χρωματισμένη εικόνα σε ασπρόμαυρη, αυτομάτως μειώνεται ο αριθμός των καναλιών, άρα και των εικονοκυτάρρων.
- **Κανονικοποίηση εικόνων:** Η τεχνική αυτή χρησιμοποιείται όταν είναι απαραίτητη η μετατροπή – κλιμάκωση μεγέθους των εικόνων, για να έχουν ιδανικά ύψη και πλάτη πριν εισαχθούν στον αλγόριθμο εκμάθησης.
- **Αύξηση δεδομένων εικόνας:** Η τεχνική αυτή περιέχει εμπλουτισμό του σετ δεδομένων με τροποποιημένες εκδόσεις των εικόνων που ήδη περιέχονται στο σετ. Με αυτό τον τρόπο, ο αλγόριθμος εκμάθησης εκπαιδεύεται με παραλλαγές διάφορων εικόνων και έχει ως αποτέλεσμα την αναγνώριση αντικειμένων με μεγάλη ακρίβεια.
- **Άλλες τεχνικές:** Σε κάποιες εργασίες μπορεί να χρειαστεί η εφαρμογή και άλλων τεχνικών προεπεξεργασίας εικόνας, όπως είναι η απομάκρυνση χρώματος από το υπόβαθρο, για να μειωθεί ο θόρυβος. Άλλες εργασίες μπορεί να χρειαστούν φωτισμό εικόνας, ειδικά σε εικόνες που λαμβάνονται σε σκοτεινά περιβάλλοντα.

2.5 Εξαγωγή Χαρακτηριστικών

Η εξαγωγή χαρακτηριστικών είναι βασικό συστατικό της αλληλουχίας βημάτων της ΜΟ και της υλοποίησης ενός αλγόριθμου βαθιάς μάθησης. Ένα μοντέλο ΤΝΔ προσπαθεί να αναγνωρίσει αντικείμενα, ανακαλύπτοντας συγκεκριμένα χαρακτηριστικά σε μία εικόνα. Ως χαρακτηριστικό ορίζεται μία μετρήσιμη ιδιότητα στα δεδομένα μιας εικόνας, η οποία αφορά μόνο τη συγκεκριμένη εικόνα και την καθιστά διαφορετική από τις υπόλοιπες εικόνες ενός σετ δεδομένων. Αυτή η ιδιότητα – χαρακτηριστικό μπορεί να είναι κάποιο

συγκεκριμένο χρώμα σε κάποιο εικονοστοιχείο ή κάποια γραμμή ή κάποια άκρη ή γενικά κάποιο τμήμα σε μία εικόνα.

Η ακρίβεια ενός μοντέλου βαθιάς μάθησης εξαρτάται σε μεγάλο βαθμό από τα χαρακτηριστικά τα οποία του παρέχουμε για εκπαίδευση. Σε ένα ΤΝΔ βαθιάς μάθησης, η εξαγωγή χαρακτηριστικών γίνεται αυτόματα από το νευρωνικό δίκτυο. Η εικόνα εισάγεται στο ΤΝΔ και στη συνέχεια, καθώς αυτή διέρχεται μέσα από τα στρώματα (layers), εκπαιδεύει το δίκτυο στο να αναγνωρίζει συγκεκριμένα μοτίβα και στη συνέχεια να εξάγει χαρακτηριστικά για την κάθε εικόνα. Ένα δίκτυο βαθιάς μάθησης εκπαιδεύεται σε αναγνώριση χαρακτηριστικών μέσω της αναπροσαρμογής των βαρών. Όταν η εικόνα εισάγεται αρχικά στο ΤΝΔ, τότε αυτό αρχικοποιεί τα βάρη σε τυχαίες τιμές και καθώς η εκπαίδευση προχωρά, το δίκτυο αναπροσαρμόζει τις τιμές των βαρών ανάλογα με τα μοτίβα που παρατηρεί στις εικόνες. Τα συχνότερα μοτίβα λαμβάνουν μεγαλύτερες τιμές βαρών, απ' ότι τα χαρακτηριστικά με μικρή επιρροή στο αποτέλεσμα.



Εικόνα 2.5. Εξαγωγή χαρακτηριστικών και ταξινόμηση αντικειμένου με χρήση ΤΝΔ.[6]

Αν για παράδειγμα έχουμε ένα σετ δεδομένων που περιέχει 15.000 εικόνες από αυτοκίνητα, που η καθεμία έχει πλάτος 1000 και ύψος 1000. Κάποιες από αυτές μπορεί να μην έχουν καθόλου φόντο, κάποιες μπορεί να έχουν ασπρόμαυρο φόντο, κάποιες πολύχρωμο, κάποιες

άλλες μπορεί να έχουν βαρύ φόντο με πολλά αχρειαστα δεδομένα. Όταν αυτό το σετ δεδομένων δοθεί ως είσοδος στους αλγόριθμους εξαγωγής χαρακτηριστικών, τότε θα απαλλαγούμε από όλα αυτά τα αχρειαστα δεδομένα και θα φιλτραριστούν μόνο τα χρήσιμα χαρακτηριστικά, οπότε η αναγνώριση εικόνων από το ΤΝΔ θα είναι ακριβέστερη και επιτυχημένη.[6]

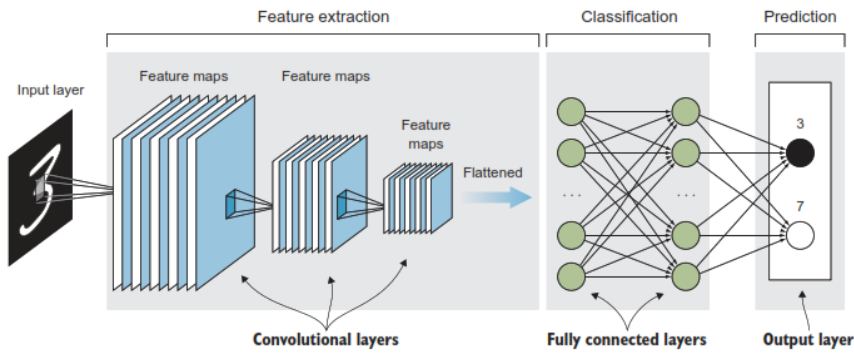
2.6 Συνελκτικά Νευρωνικά Δίκτυα

Τα Συνελκτικά Νευρωνικά Δίκτυα, ΣΝΔ (Convolutional Neural Networks – CNNs) αποτελούν υποκατηγορία των ΤΝΔ βαθιάς μάθησης και προσπαθούν να επιλύσουν εφαρμογές της ΜΟ, που δεν κατάφεραν να επιλύσουν τα παραδοσιακά ΤΝΔ. Αποτελούν την προτιμότερη αρχιτεκτονική ΤΝΔ για υλοποίηση εφαρμογών ΜΟ, όπως η ταξινόμηση εικόνων, η αναγνώριση κορυφών και αντικειμένων σε εικόνα και βίντεο, αλλά και σε εφαρμογές ενίσχυσης σκοτεινών εικόνων, όπως το αντικείμενο που μελετά η παρούσα διπλωματική εργασία.

2.6.1 Αρχιτεκτονική Συνελκτικών Νευρωνικών Δικτύων

Η αρχιτεκτονική των ΣΝΔ βασίζεται στην αρχιτεκτονική των ΤΝΔ βαθιάς μάθησης, με τη διαφορά ότι στα ΣΝΔ δε χρησιμοποιούμε διασυνδεδεμένα κρυφά στρώματα, αλλά συνελκτικά στρώματα, τα οποία χρησιμεύουν στην εξαγωγή χαρακτηριστικών. Η αρχιτεκτονική των ΣΝΔ αποτελείται από τα εξής επίπεδα :

- *Στρώμα Εισόδου (input layer)*
- *Συνελκτικά Στρώματα (convolutional layers)*
- *Διασυνδεδεμένα Στρώματα (connected layers)*
- *Στρώμα Εξόδου (output layer)*



Εικόνα 2.6. Αρχιτεκτονική Συνελκτικού Νευρωνικού Δικτύου.[6]

Η εικόνα εισέρχεται στο στρώμα εισόδου και στη συνέχεια περνά στα συνελκτικά στρώματα για να υποστεί μείωση στις διαστάσεις της. Ακολούθως εισέρχεται στα διασυνδεδεμένα στρώματα για να ταξινομηθούν τα χαρακτηριστικά της εικόνας και τέλος, το στρώμα εξόδου του ΣΝΔ δίνει την πρόβλεψη.

2.6.2 Συνέλιξη

Στα μαθηματικά η συνέλιξη (**convolution**) είναι ένας πολλαπλασιασμός μεταξύ δύο πινάκων ή δύο συναρτήσεων, ο οποίος παράγει ένα τρίτο πίνακα ή μία τρίτη συνάρτηση. Στα ΣΝΔ, ο πρώτος πίνακας είναι η εικόνα εισόδου (**input image**) και ο δεύτερος πίνακας είναι ένα φίλτρο συνέλιξης, ο οποίος ονομάζεται φίλτρο ή πυρήνας (**kernel**).

Έστω ο πίνακας A διαστάσεων $P \times Q$, ο οποίος αναπαριστά μία εικόνα, με στοιχεία διάφορες τιμές της φωτεινότητας των εικονοκυττάρων (**pixels**):

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

Και ο πίνακας B :

1	2
3	4

Ουσιαστικά ο πίνακας B (φίλτρο) είναι ένας πίνακας διαστάσεων $M \times N$, με στοιχεία τις τιμές των βαρών, τα οποία αρχικοποιούνται τυχαία. Με χρήση μεθόδων βελτιστοποίησης, το ΣΝΔ εκπαιδεύεται και αν χρειαστεί τις διορθώνει μέχρι να επιτευχθεί η μέγιστη βελτιστοποίηση, άρα και οι βέλτιστες τιμές των βαρών. Το φίλτρο ή πίνακας B, με την πράξη της συνέλιξης «σαρώνει» όλη την εικόνα ή πίνακα A. Η πράξη εκφράζεται μαθηματικά ως :

$$\Gamma_{ij} = \sum_{m=1}^M \sum_{n=1}^N A_{i-m,j-n} B_{mn}$$

Η πράξη αυτή είναι γνωστή ως πράξη της συνέλιξης και συμβολίζεται συνήθως ως :

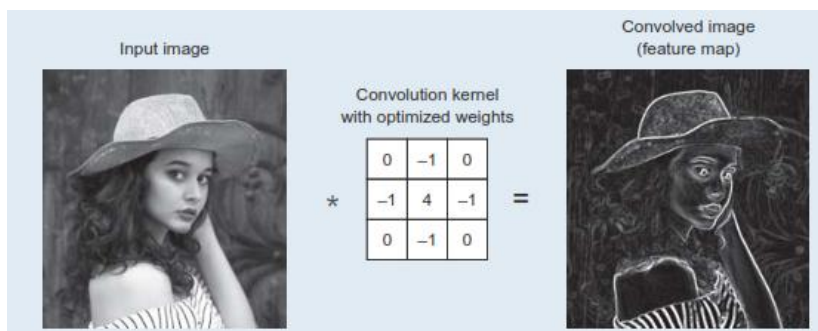
$$\Gamma_{ij} = (\mathbf{B} * \mathbf{A})(i, j)$$

Εκτελώντας την πράξη της συνέλιξης μεταξύ των πινάκων A, B, προκύπτει ο πίνακας Γ :

44	54	64
84	94	104
124	134	144

Ο πίνακας Γ είναι το αποτέλεσμα της συνέλιξης του A με τον B πίνακα και ονομάζεται *χάρτης χαρακτηριστικών ή χάρτης ενεργοποίησης*.

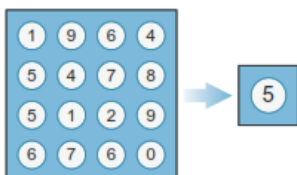
Η λογική των φίλτρων και της συνέλιξης γενικότερα, είναι η εκπαίδευση του ΤΝΔ και η ακριβέστερη εξαγωγή χαρακτηριστικών από τις εικόνες εισόδου. Με τη χρήση των φίλτρων εκπαιδεύουμε το ΤΝΔ στο να αφαιρεί τα αχρείαστα δεδομένα των εικόνων και να εστιάζει στα χρήσιμα δεδομένα – χαρακτηριστικά. Το μέγεθος των φίλτρων κυμαίνεται μεταξύ των διαστάσεων 2x2 και 5x5. Όσο μικρότερα τα φίλτρα, τόσο μεγαλύτερη η λεπτομέρεια. Στην παρακάτω εικόνα εφαρμόζεται ένα φίλτρο 3x3 που ανιχνεύει κορυφές, όμως υπάρχουν φίλτρα που ανιχνεύουν οριζόντιες κορυφές, άλλα που ανιχνεύουν κατακόρυφες κορυφές, άλλα που ανιχνεύουν πολύπλοκα σχήματα κ.ο.κ.[6], [7].

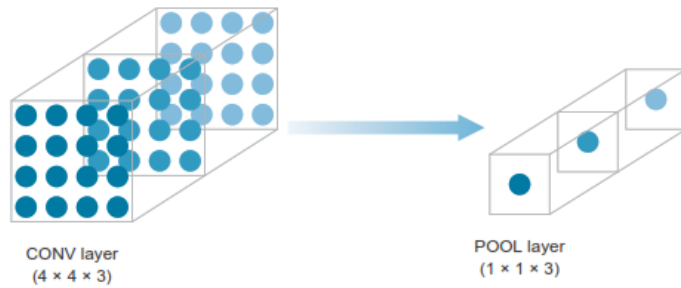


Εικόνα 2.7. Συνέλιξη με φίλτρο αναγνώρισης κορυφών σε ασπρόμαυρη εικόνα.[6]

2.6.3 Συγκέντρωση (Pooling)

Τα στρώματα συγκέντρωσης (**pooling layers**) προστίθενται στα συνελκτικά στρώματα ενός ΣΝΔ για να μειώσουν τις διαστάσεις τους. Η χρήση των στρωμάτων συγκέντρωσης είναι απαραίτητη σε πολύπλοκα σετ δεδομένων, στα οποία έχουν εφαρμοσθεί πάρα πολλά συνελκτικά φίλτρα, με αποτέλεσμα τα ενδιάμεσα συνελκτικά στρώματα να είναι πολύ μεγάλα. Συνεπώς η πρόσθεση των στρωμάτων συγκέντρωσης ενισχύει την εξαγωγή χαρακτηριστικών από το ΣΝΔ, αφού μειώνει δραστικά τις διαστάσεις των εικόνων. Η πράξη της συγκέντρωσης εφαρμόζει μία συνάρτηση μεγιστοποίησης ή μέσης τιμής στα συνελκτικά στρώματα, έτσι ώστε να μειώσει το σύνολο των παραμέτρων που έχουν προκύψει, μειώνοντας ταυτόχρονα την πολυπλοκότητα των υπολογισμών. Η διαφορά των στρωμάτων συγκέντρωσης με τα συνελκτικά φίλτρα είναι ότι τα πρώτα δεν περιέχουν τιμές βαρών που εφαρμόζονται στην αρχική εικόνα.

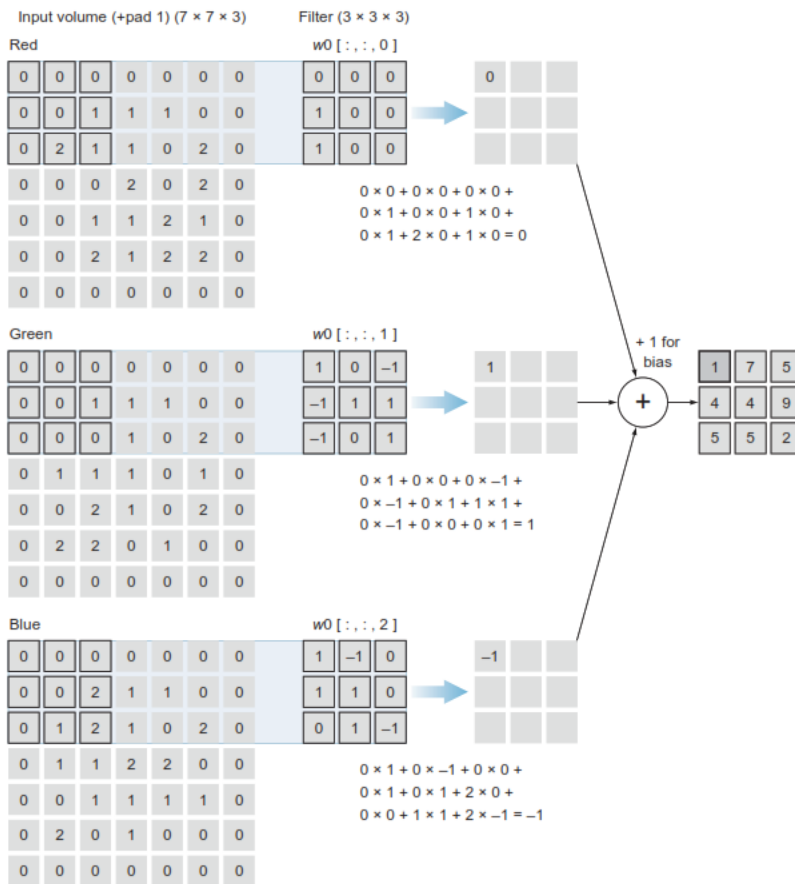




Εικόνα 2.9. Μετατροπή ενός 3D πίνακα σε διάνυσμα με χρήση συνάρτησης συγκέντρωσης.[6]

2.7 Συνέλιξη σε Έγχρωμες Εικόνες

Οι χρωματισμένες εικόνες αναγνωρίζονται από τα υπολογιστικά συστήματα ως πίνακες τριών διαστάσεων με ύψος, πλάτος και βάθος. Στην περίπτωση των RGB (Red, Green, Blue) εικόνων το βάθος είναι 3, δηλαδή ένα κανάλι για κάθε χρώμα, οπότε συνολικά για κάθε εικόνα έχουμε μία συσσώρευση από τρεις πίνακες των δύο διαστάσεων. Η συνέλιξη σε αυτές τις εικόνες είναι παρόμοια με τη συνέλιξη των εικόνων κλίμακας του γκρι, δηλαδή των εικόνων που έχουν βάθος ίσο με 1. Για να επιτύχουμε συνέλιξη ακολουθούμε την ίδια διαδικασία με πριν, με τη διαφορά ότι το τελικό άθροισμά μας είναι υπολογισμένο 3 φορές, μία για κάθε κανάλι χρώματος.



Εικόνα 2.10. Συνέλιξη σε χρωματισμένες εικόνες.[6]

- Σε κάθε κανάλι εφαρμόζεται συγκεκριμένο φίλτρο.
- Κάθε φίλτρο εφαρμόζεται πάνω στην εικόνα, πολλαπλασιάζει την τιμή του εικονοκυττάρου και στο τέλος προστίθενται όλα μαζί για να προκύψει η τιμή του εικονοκυττάρου κάθε καναλιού μετά τη συνέλιξη.
- Ακολούθως, προστίθενται και οι τρεις τιμές των εικονοκυττάρων και προκύπτει η τιμή του εικονοκυττάρου στον πίνακα χαρακτηριστικών (feature map).
- Τέλος, εφαρμόζεται το φίλτρο σε κάθε εικονοκύτταρο κάθε καναλιού της εικόνας, μέχρι να υπολογισθεί ο συνολικός πίνακας χαρακτηριστικών με όλες τις τιμές των εικονοκυττάρων μετά τη συνέλιξη.

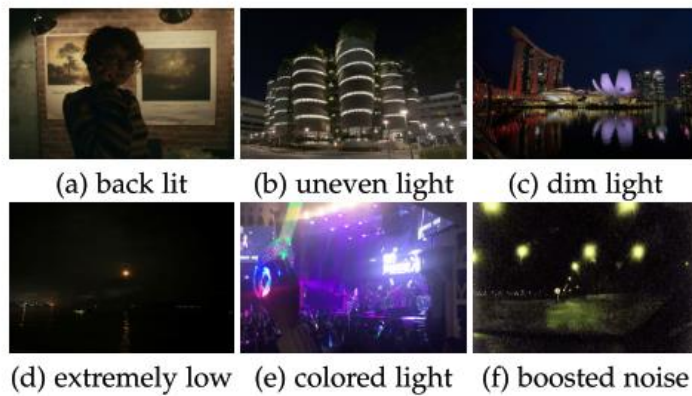
ΚΕΦΑΛΑΙΟ 3 – Επεξεργασία Εικόνων Χαμηλού Φωτισμού

3.1 Εισαγωγή

Οι εικόνες και τα βίντεο αναπαριστούν τα πραγματικά σκηνικά με αναλυτική και υψηλής ποιότητας πληροφορία. Από τη σύλληψη και την επεξεργασία των δεδομένων εικόνας και βίντεο, τα υπολογιστικά συστήματα μπορούν να εκτελέσουν ποικίλες εργασίες, όπως αναγνώριση αντικειμένων, ταξινόμηση, τρισδιάστατο ανασχηματισμό, οι οποίες χρησιμοποιούνται εκτενώς σε καθημερινές εφαρμογές, όπως είναι η αυτοματοποιημένη οδήγηση και η παρακολούθηση χώρων και εκτάσεων. Όμως, η αποτελεσματικότητα αυτών των εφαρμογών, βασίζεται κυρίως στην ποιότητα των εικόνων και βίντεο, δηλαδή των δεδομένων εισόδου. Όταν τα δεδομένα εισόδου είναι υψηλής ποιότητας, τόσο μεγαλύτερη θα είναι και η επιτυχής εκτέλεση της υπηρεσίας – εφαρμογής, ειδάλως θα υπάρξει κακή εκτέλεση. Τυπικό παράδειγμα δεδομένων χαμηλής ποιότητας, αποτελούν οι φωτογραφίες και τα βίντεο που λαμβάνονται σε περιβάλλοντα χαμηλού ή κακού φωτισμού, καθώς υπάρχει έντονη απώλεια πληροφορίας και απροσδόκητος θόρυβος (noise). Η χρήση τέτοιων δεδομένων οδηγεί σε μείωση της αποτελεσματικότητας των υπάρχοντων αλγορίθμων υπολογιστικής όρασης και γι' αυτό το λόγο αναπτύσσονται τα τελευταία χρόνια αλγόριθμοι ενίσχυσης εικόνων χαμηλού φωτισμού.

3.2 Εικόνες Χαμηλού Φωτισμού

Η έλλειψη φωτός κατά τη διάρκεια της λήψης μιας φωτογραφίας, οδηγεί σε απώλεια σημαντικής πληροφορίας και απροσδόκητου θορύβου, άρα και μειωμένου SNR (signal to noise ratio). Οι εικόνες αυτές, υποφέρουν από οπίσθιο, άνισο και ανεπαρκή φωτισμό (backlit – uneven light) καθώς και από έντονη παρουσία θορύβου. Η χρήση τέτοιων εικόνων χαμηλού φωτισμού οδηγεί σε μειωμένη απόδοση των περισσότερων αλγορίθμων υπολογιστικής όρασης.



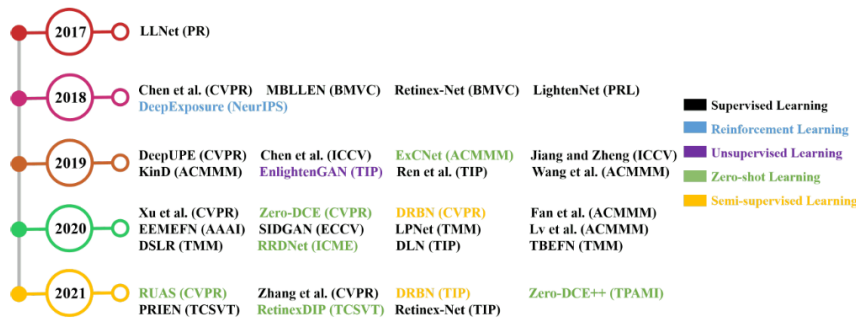
Εικόνα 3.1. Παραδείγματα Εικόνων Χαμηλού Φωτισμού.[10]

Τα τελευταία χρόνια έχουν αναπτυχθεί αρκετές τεχνικές βελτίωσης, οι οποίες φαίνεται να αντιμετωπίζουν πιο αποτελεσματικά το πρόβλημα έναντι των φυσικών τρόπων αύξησης του SNR όπως είναι το άνοιγμα της οπής της φωτογραφικής μηχανής, αύξηση του χρόνου έκθεσης και χρήση του φλας. Όλοι οι φυσικοί τρόποι αντιμετώπισης έχουν σημαντικά πλεονεκτήματα, αλλά και χαρακτηριστικά μειονεκτήματα όπως είναι η αύξηση της θολούρας. Η ραγδαίως αυξανόμενη χρήση των έξυπνων τηλεφώνων (smartphones) για φωτογράφιση σε συνθήκες χαμηλού φωτισμού, έχει στρέψει το ερευνητικό ενδιαφέρον στην ανάπτυξη μεθόδων και αλγορίθμων ενίσχυσης των εικόνων χαμηλού φωτισμού.

3.3 Μέθοδοι Ενίσχυσης Εικόνων Χαμηλού Φωτισμού

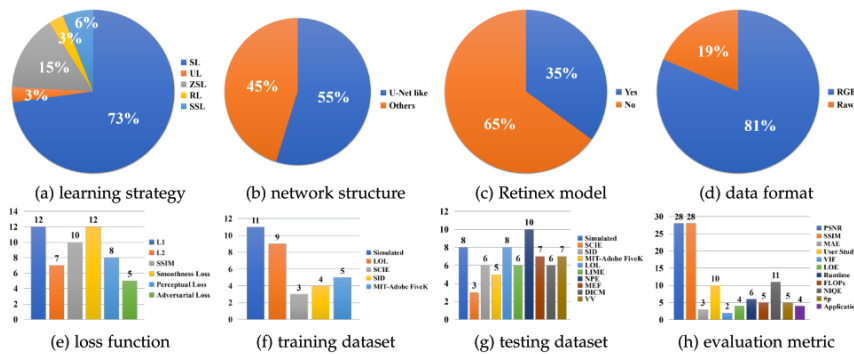
Η ενίσχυση εικόνων χαμηλού φωτισμού στοχεύει στην βελτίωση εικόνων, οι οποίες λαμβάνονται σε περιβάλλοντα χαμηλού φωτισμού. Ιστορικά έχουν αναπτυχθεί χρήσιμες και παραδοσιακές μέθοδοι ενίσχυσης, οι οποίες περιλαμβάνουν μεθόδους που βασίζονται στην εξίσωση ιστογραμμάτων (histograms equalization)[12] και στη θεωρία Retinex[11]. Τα τελευταία 3 χρόνια, η αλματώδης εξέλιξη του πεδίου της βαθιάς μάθησης, έχει οδηγήσει στην ανάπτυξη μεθόδων που βασίζονται στην τεχνική των νευρωνικών δικτύων, καθώς οι τεχνικές αυτές έχουν μεγαλύτερη ακρίβεια, ταχύτητα και στιβαρότητα.

Με σχόλια [PK1]: δεν έχει αναφέρει τι γινόνταν πιο παλιά που πηγαίναμε με Histogram equalization και άλλες τεχνικές βάλε εικόνες το μάστερ αφορά εικόνα.



Εικόνα 3.2. Μία συνοπτική αναφορά των σύγχρονων μεθόδων ενίσχυσης εικόνων χαμηλού φωτισμού με χρήση τεχνικών βαθιάς μάθησης [10].

Από το 2017 και ύστερα οι μέθοδοι ενίσχυσης σκοτεινών εικόνων, οι οποίες βασίζονται σε τεχνικές βαθιάς μάθησης, έχουν επικρατήσει. Οι στρατηγικές μάθησης που χρησιμοποιούνται περιλαμβάνουν τεχνικές επιτηρούμενης μάθησης (supervised learning – SL), μη επιτηρούμενης μάθησης (unsupervised learning – UL), ενισχυτικής μάθησης (reinforcement learning – RL), ημιεπιτηρούμενης μάθησης (semi supervised learning – SSL) και μηδενικής μάθησης (zero shot learning – ZSL).



Εικόνα 3.3. Στατιστική ανάλυση μεθόδων βαθιάς μάθησης. [10]

Παρατηρώντας την εικόνα 3.3, διαπιστώνουμε ότι το 73% των μεθόδων, βασίζεται στην επιτηρούμενη μάθηση (SL) και αυτό εξηγείται από το γεγονός ότι η επιτηρούμενη μάθηση είναι σχετικά εύκολη και εφαρμόσιμη σε σετ δεδομένων όπως το LOL[13] και το SID[8]

που είναι και σετ δεδομένων – ζεύγος (paired datasets). Ωστόσο οι SL μέθοδοι εμφανίζουν κάποιες προκλήσεις :

- Δυσκολία συλλογής μεγάλων σετ δεδομένων – ζεύγος πραγματικών εικόνων χαμηλού φωτισμού
- Η εκπαίδευση ενός δικτύου βαθιάς μάθησης σε δεδομένα – ζεύγη (paired data) οδηγεί σε περιορισμένη γενίκευση των πραγματικών εικόνων που έχουν διαφορετική φωτεινότητα.

Προκειμένου να επιλυθούν αυτές οι προκλήσεις που δημιουργούνται μέσω της SL, παρατηρούμε ότι γίνεται χρήση και των υπολοίπων μεθόδων μάθησης, όπως η UL, RL, SSL και ZSL, οι οποίες με τη σειρά τους πάσχουν από κάποιους περιορισμούς, όπως :

- Μη υλοποίηση σταθερής εκπαίδευσης, καθώς αποφεύγεται η αναγνώριση της διαφοράς μεταξύ χρωμάτων (SSL και UL)
- Σχεδίαση και υλοποίηση ενός αποτελεσματικού μηχανισμού βράβευσης (RL)

Στον παρακάτω πίνακα 1 παρατίθενται όλες οι μέθοδοι βαθιάς μάθησης που έχουν αναπτυχθεί μέχρι σήμερα, καθώς και οι τεχνικές μάθησης τους, οι συναρτήσεις σφάλματος, τα δεδομένα εκπαίδευσης και τα δεδομένα δοκιμής, όπως επίσης οι μέθοδοι αξιολόγησης και η πλατφόρμα υλοποίησης της κάθε τεχνικής [10].

	Method	Learning	Network Structure	Loss Function	Training Data	Testing Data	Evaluation Metric	Format	Platform	Retinex
2017	LLNet [1]	SL	SSDA	SRR loss	simulated by Gamma Correction & Gaussian Noise	simulated self-selected	PSNR SSIM	RGB	Theano	
	2018	LightenNet [5]	SL	four layers	L_2 loss	simulated by random illumination values	simulated self-selected	PSNR MAE SSIM User Study	RGB	Caffe MATLAB
Retinex-Net [4]		SL	multi-scale network	L_1 loss smoothness loss invariable reflectance loss	simulated by adjusting histogram	self-selected	-	RGB	TensorFlow	✓
	MBLEN [3]	SL	multi-branch fusion	SSIM loss region loss perceptual loss	simulated by Gamma Correction & Poisson Noise	simulated self-selected	PSNR SSIM AB VIF LOE TOMI	RGB	TensorFlow	
	SCIE [6]	SL	frequency decomposition	L_2 loss L_1 loss SSIM loss	SCIE	SCIE	PSNR FSIM Runtime FLOPs	RGB	Caffe MATLAB	
	Chen et al. [2]	SL	U-Net	L_1 loss	SID	SID	PSNR SSIM	raw	TensorFlow	
	Deepexposure [25]	RL	policy network GAN	deterministic policy gradient adversarial loss	MIT-Adobe FiveK	MIT-Adobe FiveK	PSNR SSIM	raw	TensorFlow	
2019	Chen et al. [8]	SL	siamese network	L_1 loss self-consistency loss	DRV	DRV	PSNR SSIM MAE	raw	TensorFlow	
	Jiang and Zheng [9]	SL	3D U-Net	L_1 loss	SMOID	SMOID	PSNR SSIM MSE	raw	TensorFlow	
	DeepUPE [60]	SL	illumination map	L_1 loss color loss smoothness loss reflectance similarity loss illumination smoothness loss mutual consistency loss L_1 loss L_2 loss SSIM loss texture similarity loss illumination adjustment loss	retouched image pairs	MIT-Adobe FiveK	PSNR SSIM User Study	RGB	TensorFlow	✓
	KinD [11]	SL	three subnetworks U-Net	illumination smoothness loss mutual consistency loss L_1 loss L_2 loss SSIM loss texture similarity loss illumination adjustment loss	LOL	LOL LIME NPE MEF	PSNR SSIM LOE NIQE	RGB	TensorFlow	✓
	Wang et al. [10]	SL	two subnetworks pointwise Conv	L_1 loss	simulated by camera imaging model MIT-Adobe FiveK with Gamma correction & Gaussian noise	IP10 FNFS MFL LOL NPE	PSNR SSIM NIQE	RGB	Caffe	✓
	Ren et al. [12]	SL	U-Net like network RNN dilated Conv	L_2 loss perceptual loss adversarial loss	simulated self-selected	DFPD	PSNR SSIM Runtime	RGB	Caffe	
	EnlightenGAN [26]	UL	U-Net like network	adversarial loss self feature preserving loss	unpaired real images	NPE LIME MEF DICM VV BBD-100K ExDARK	User Study NIQE Classification	RGB	PyTorch	
	ExCNet [27]	ZSL	fully connected layers	energy minimization loss	real images	$I_{Exp,D}$	User Study CDRQA LOD	RGB	PyTorch	
2020	Zero-DCE [28]	ZSL	U-Net like network	spatial consistency loss exposure control loss color consistency loss illumination smoothness loss SSIM loss perceptual loss adversarial loss	SICE	SICE NPE LIME MEF DICM VV DARK FACE	User Study PI PSNR SSIM MAE Runtime Face detection	RGB	PyTorch	
	DRBN [33]	SSL	recursive network	SSIM loss perceptual loss adversarial loss	LOL images selected by MOS	LOL	PSNR SSIM SSM-GC	RGB	PyTorch	
	Lv et al. [15]	SL	U-Net like network	Huber loss SSIM loss perceptual loss illumination smoothness loss mutual smoothness loss reconstruction loss	simulated by a retouching module	LOL SICE DeepUPE	User Study PSNR SSIM VIF LOE NIQE #F Runtime Face detection	RGB	TensorFlow	✓
	Fan et al. [14]	SL	four subnetworks U-Net like network feature modulation	illumination smoothness loss cross entropy loss consistency loss SSIM loss gradient loss ratio learning loss	simulated by illumination adjustment, slight color distortion, and noise simulation	simulated self-selected	PSNR SSIM NIQE	RGB	-	✓
	Xu et al. [57]	SL	frequency decomposition U-Net like network	L_2 loss perceptual loss	SID in RGB	SID in RGB self-selected	PSNR SSIM	RGB	PyTorch	
	EEMFN [16]	SL	U-Net like network edge detection network	L_1 loss weighted cross-entropy loss	SID	SID	PSNR SSIM	raw	TensorFlow PaddlePaddle	
	DLN [19]	SL	residual learning interactive factor back projection network	SSIM loss total variation loss	simulated by illumination adjustment, slight color distortion, and noise simulation	simulated LOL	User Study PSNR SSIM NIQE	RGB	PyTorch	
	LPNet [18]	SL	pyramid network	L_1 loss perceptual loss luminance loss	LOL SID in RGB MIT-Adobe FiveK	LOL SID in RGB MIT-Adobe FiveK MEF NPE DICM VV	PSNR SSIM NIQE #F FLOPs Runtime PSNR SSIM	RGB	PyTorch	
	SIDGAN [17]	SL	U-Net	CycleGAN loss	SIDGAN	SIDGAN	PSNR SSIM TPSNR TSSIM ATWE	raw	TensorFlow	
	RRDNet [29]	ZSL	three subnetworks	retinex reconstruction loss texture enhancement loss noise estimation loss	-	NPE LIME MEF DICM	NIQE CFCQI	RGB	PyTorch	✓
	TBEFN [20]	SL	three stages U-Net like network	SSIM loss perceptual loss smoothness loss	SCIE LOL	SCIE LOL DICM MEF NPE VV	PSNR SSIM NIQE Runtime #F FLOPs	RGB	TensorFlow	✓
	DSLK [21]	SL	Laplacian pyramid U-Net like network	L_2 loss Laplacian loss color loss	MIT-Adobe FiveK	MIT-Adobe FiveK self-selected	PSNR SSIM NIQMC NIQE BTMCI CAHDC	RGB	PyTorch	
2021	RUAS [32]	ZSL	neural architecture search	cooperative loss similar loss total variation loss	LOL MIT-Adobe FiveK	LOL MIT-Adobe FiveK	PSNR SSIM Runtime #F FLOPs	RGB	PyTorch	✓
	Zhang et al. [22]	SL	U-Net	L_1 loss consistency loss	simulated by illumination adjustment and noise simulation	simulated self-selected	User Study PSNR SSIM AB MAAD WE User Study PI	RGB	PyTorch	
	Zero-DCE++ [30]	ZSL	U-Net like network	spatial consistency loss exposure control loss color consistency loss illumination smoothness loss perceptual loss detail loss quality loss L_1 loss L_2 loss SSIM loss	SICE	SICE NPE LIME MEF DICM VV DARK FACE	User Study PI PSNR SSIM #F MAE Runtime Face detection FLOPs	RGB	PyTorch	
	DRBN [34]	SSL	recursive network	perceptual loss	LOL	LOL	PSNR SSIM SSM-GC	RGB	PyTorch	
	Retinex-Net [24]	SL	three subnetworks	L_1 loss SSIM loss total variation loss reconstruction loss	simulated by adjusting histogram	LOL simulated NPE DICM VV	PSNR SSIM UQI OSS User Study	RGB	PyTorch	✓
	RetinexDIP [31]	ZSL	encoder-decoder networks	illumination-consistency loss reflectance loss illumination smoothness loss	-	DICM ExDark Fusion LIME NASA NPE VV	NIQE NIQMC CFCQI	RGB	PyTorch	✓
	FRIEN [23]	SL	recursive network	SSIM loss	MEF LOL simulated by adjusting histogram	LOL LIME NPE MEF VV	PSNR SSIM LOE TMQI	RGB	PyTorch	

Πίνακας 1. Ανακεφαλαίωση τεχνικών και μεθόδων βαθιάς μάθησης για την ενίσχυση εικόνων χαμηλού φωτισμού.[10]

3.4 Η Μέθοδος Ενίσχυσης Kind – Kindling the Darkness

3.4.1 Εισαγωγή

Στο γενικό πεδίο της ΜΟ και συγκεκριμένα στα πεδία της επεξεργασίας και της αποθρομβοποίησης εικόνων υπάρχουν αρκετές προτεινόμενες μέθοδοι τα τελευταία χρόνια και ειδικότερα αυτές που χρησιμοποιούν τεχνολογία ΤΝΔ. Μία από αυτές τις μεθόδους προτάθηκε από τους Zhang Y., Zhang J., Guo X. [9] και ουσιαστικά ενισχύει τη φωτεινότητα των σκοτεινών περιοχών των φωτογραφιών. Για την υλοποίησή της χρησιμοποιείται ένα ΣΝΔ, το οποίο είναι εμπνευσμένο από τη θεωρία Retinex και διαχωρίζει την εικόνα σε δύο άλλα στοιχεία. Το ένα στοιχείο χρησιμοποιείται για τη ρύθμιση της φωτεινότητας, ενώ το άλλο για την αφαίρεση των σκοτεινών συστατικών. Στην αριστερή στήλη της εικόνας 3.4 παρατίθενται τρεις εικόνες που ελήφθησαν σε διαφορετικές συνθήκες φωτισμού, ενώ στη δεξιά στήλη παρατίθενται τα αποτελέσματα ύστερα από εφαρμογή της μεθόδου ενίσχυσης Kind.

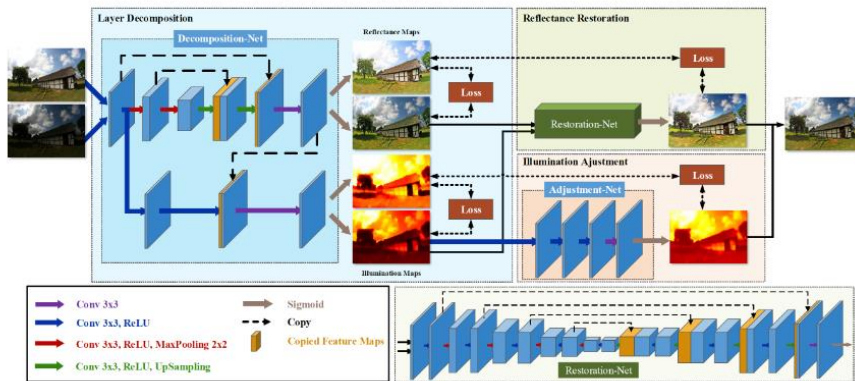


Εικόνα 3.4. Οπτικά αποτελέσματα της Kind.[9]

3.4.2 Μεθοδολογία

Τα αποτελέσματα μιας μεθόδου ενίσχυσης καθορίζονται αρχικά από την ικανότητά της να αφαιρεί τις υποβαθμίσεις που δημιουργεί ο χαμηλός φωτισμός σε μία εικόνα και δευτερευόντως, από την ικανότητά της να ρυθμίζει τις συνθήκες φωτισμού. Για το σκοπό αυτό, το ΣΝΔ που χρησιμοποιείται στη μέθοδο Kind αποτελείται από τρία τμήματα:

- Τμήμα – Στρώμα Αποσύνθεσης (*Decomposition Layer*)
- Επαναφορά Ανακλαστικότητας (*Reflectance Restoration*)
- Ρύθμιση Φωτισμού (*Illumination Adjustment*)



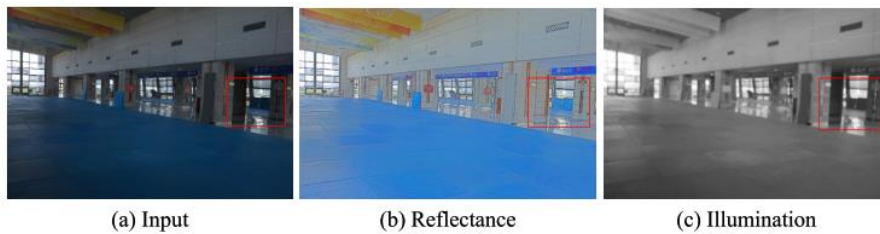
Εικόνα 3.5. Αρχιτεκτονική της μεθόδου Kind.[9]

3.4.2.1 Στρώμα Αποσύνθεσης

Η κλασική θεωρία Retinex μοντελοποιεί την ανθρώπινη αντίληψη των χρωμάτων. Η βασική υπόθεση της συγκεκριμένης θεωρίας είναι ότι κάθε εικόνα I προκύπτει από τη σύνθεση δύο στοιχείων, της ανακλαστικότητας R και του φωτισμού L , τα οποία συνδέονται μεταξύ τους με τη σχέση :

$$I = R \circ L,$$

όπου \circ η πράξη του πολλαπλασιασμού μεταξύ των πινάκων \mathbf{R} , \mathbf{L} ίδιων διαστάσεων. Η ανακλαστικότητα \mathbf{R} περιγράφει τις φυσικές ιδιότητες των ληφθέντων εικόνων – αντικειμένων, οι οποίες δεν εξαρτώνται από τις συνθήκες φωτισμού και ο φωτισμός \mathbf{L} αναπαριστά τη φωτεινότητα των αντικειμένων.



Εικόνα 3.6. Οι χάρτες ανακλαστικότητας και φωτισμού $b)$, $c)$ όπως προκύπτουν από την εικόνα $a)$. [13]

Στο στρώμα αποσύνθεσης της μεθόδου kind χρησιμοποιείται ένα ΣΝΔ, το οποίο χωρίζει την εικόνα σε \mathbf{R} και \mathbf{L} . Στο στάδιο της εκπαίδευσης, το ΣΝΔ παίρνει ζεύγη εικόνων χαμηλού και κανονικού φωτισμού $[\mathbf{I}_l, \mathbf{I}_h]$ και εκπαιδεύεται στην εξαγωγή της ανακλαστικότητας \mathbf{R} από εικόνες διαφορετικής φωτεινότητας $[\mathbf{R}_l, \mathbf{R}_h]$. Η εκπαίδευση πραγματοποιείται υπό τον περιορισμό ότι τόσο οι εικόνες χαμηλού φωτισμού, όσο και οι εικόνες κανονικού φωτισμού, ιδανικά (χωρίς υποβαθμίσεις) παρουσιάζουν την ίδια ανακλαστικότητα \mathbf{R} , δηλαδή τα \mathbf{R}_l και \mathbf{R}_h τείνουν να είναι ίσα. Επιπροσθέτως, το ζεύγος του φωτισμού $[\mathbf{L}_l, \mathbf{L}_h]$ θα βρίσκεται σε αμοιβαία συμφωνία.

Από την εικόνα 3.5 παρατηρούμε ότι αρχικά χρησιμοποιείται ένα συνελκτικό στρώμα 3×3 για την εξαγωγή χαρακτηριστικών από τις εικόνες που δόθηκαν ως είσοδος. Στη συνέχεια χρησιμοποιούνται αρκετά συνελκτικά στρώματα 3×3 με συνάρτηση ενεργοποίησης ReLU [18], όπως και φίλτρα 2×2 , για αναπαράσταση της RGB εικόνας σε ανακλαστικότητα (\mathbf{R}) και φωτισμό (\mathbf{L}). Στο επόμενο βήμα χρησιμοποιείται η σιγμοειδής συνάρτηση [18] για να περιορίσει τα \mathbf{R} , \mathbf{L} στο $[0, 1]$.

Inputs	Operator	Kernel	Output Channels	Stride	Output Name
RGB	Conv&ReLU	3 × 3	32	1	Decom_conv1
Decom_conv1	Max Pooling	2 × 2	32	2	Decom_pool1
Decom_pool1	Conv&ReLU	3 × 3	64	1	Decom_conv2
Decom_conv2	Max Pooling	2 × 2	64	2	Decom_pool2
Decom_pool2	Conv&ReLU	3 × 3	128	1	Decom_conv3
Decom_conv3	Deconv	2 × 2	64	2	Decom_up1
Decom_up1, Decom_conv2	Concat	-	128	-	Decom_concat1
Decom_concat1	Conv&ReLU	3 × 3	64	1	Decom_conv4
Decom_conv4	Deconv	2 × 2	32	2	Decom_up2
Decom_up2, Decom_conv1	Concat	-	64	-	Decom_concat2
Decom_concat2	Conv&ReLU	3 × 3	32	1	Decom_conv5
Decom_conv5	Conv	3 × 3	3	1	Decom_conv6
Decom_conv6	Sigmoid	-	3	-	Decom_Reflectance
Decom_conv1	Conv&ReLU	3 × 3	32	1	Decom_i_conv1
Decom_i_conv1, Decom_conv5	Concat	-	64	-	Decom_i_conv2
Decom_i_conv2	Conv	3 × 3	1	1	Decom_i_conv3
Decom_i_conv3	Sigmoid	-	1	-	Decom_Illumination

Πίνακας 2. Αρχιτεκτονική Στρώματος Αποσύνθεσης.[9]

Στο τελικό βήμα του στρώματος αποσύνθεσης, εφαρμόζεται η συνάρτηση κόστους (\mathcal{L}^{LD}) στις εικόνες εξόδου του στρώματος της ανακλαστικότητας και του στρώματος φωτισμού, η οποία περιορίζει τα σφάλματα:

$$\mathcal{L}^{LD} := \mathcal{L}_{rec}^{LD} + 0.01\mathcal{L}_{rs}^{LD} + 0.08\mathcal{L}_{is}^{LD} + 0.1\mathcal{L}_{mc}^{LD},$$

- $\mathcal{L}_{rec}^{LD} = \|I_l - R_l \circ L_l\|_1 + \|I_h - R_h \circ L_h\|_1$
- $\mathcal{L}_{rs}^{LD} := \|R_l - R_h\|_2^2$
- $\mathcal{L}_{is}^{LD} = \left\| \frac{\nabla L_l}{\max(|\nabla I_l|, \epsilon)} \right\|_1 + \left\| \frac{\nabla L_h}{\max(|\nabla I_h|, \epsilon)} \right\|_1, \epsilon = 0.01$
- $\mathcal{L}_{mc}^{LD} = \|M \circ \exp(-c \cdot M)\|_1, M = |\nabla L_l| + |\nabla L_h|$



a) Εικόνα Χαμηλού Φωτισμού

b) Εικόνα Κανονικού Φωτισμού



c) Χάρτης Φωτισμού της a)



d) Χάρτης Φωτισμού της b)



e) Χάρτης Ανακλαστικότητας της a)



f) Χάρτης Ανακλαστικότητας της b)

Εικόνα 3.7. Αποσύνθεση Εικόνων Χαμηλού και Υψηλού Φωτισμού [9].

Στην αριστερή στήλη της εικόνας 3.6 παρατηρούμε την αποσύνθεση μιας εικόνας χαμηλού φωτισμού (1^η σειρά) σε χάρτη φωτισμού (2^η σειρά) και χάρτη ανακλαστικότητας (3^η σειρά). Στη δεξιά στήλη παρατηρούμε την αποσύνθεση της ίδιας εικόνας με κανονικό φωτισμό.

3.4.2.2 Επαναφορά Ανακλαστικότητας

Οι χάρτες ανακλαστικότητας των εικόνων χαμηλού φωτισμού πάσχουν από υψηλές υποβαθμίσεις. Σκοπός της μεθόδου kind είναι η εξάλειψη αυτών των υποβαθμίσεων και το πετυχαίνει αυτό χρησιμοποιώντας ένα ΣΝΔ επαναφοράς (πίνακας 3), το οποίο χρησιμοποιεί περισσότερα ενδιάμεσα στρώματα απ' ό,τι το αρχικό ΣΝΔ αποσύνθεσης.

Inputs	Operator	Kernel	Output Channels	Stride	Output Name
Decom_i_conv3, Decom_conv5	Concat	-	33	-	RE_concat1
RE_concat1	Conv&ReLU	3 × 3	32	1	RE_conv1_1
RE_conv1_1	Conv&ReLU	3 × 3	32	1	RE_conv1_2
RE_conv1_2	Max Pooling	2 × 2	32	2	RE_pool1
RE_pool1	Conv&ReLU	3 × 3	64	1	RE_conv2_1
RE_conv2_1	Conv&ReLU	3 × 3	64	1	RE_conv2_2
RE_conv2_2	Max Pooling	2 × 2	64	2	RE_pool2
RE_pool2	Conv&ReLU	3 × 3	128	1	RE_conv3_1
RE_conv3_1	Conv&ReLU	3 × 3	128	1	RE_conv3_2
RE_conv3_2	Max Pooling	2 × 2	128	2	RE_pool3
RE_pool3	Conv&ReLU	3 × 3	256	1	RE_conv4_1
RE_conv4_1	Conv&ReLU	3 × 3	256	1	RE_conv4_2
RE_conv4_2	Max Pooling	2 × 2	256	2	RE_pool4
RE_pool4	Conv&ReLU	3 × 3	512	1	RE_conv5_1
RE_conv5_1	Conv&ReLU	3 × 3	512	1	RE_conv5_2
RE_conv5_2	Deconv	2 × 2	256	2	RE_up1
RE_up1, RE_conv4_2	Concat	-	512	-	RE_concat2
RE_concat2	Conv&ReLU	3 × 3	256	1	RE_conv6_1
RE_conv6_1	Conv&ReLU	3 × 3	256	1	RE_conv6_2
RE_conv6_2	Deconv	2 × 2	128	2	RE_up2
RE_up2, RE_conv3_2	Concat	-	256	-	RE_concat3
RE_concat3	Conv&ReLU	3 × 3	128	1	RE_conv7_1
RE_conv7_1	Conv&ReLU	3 × 3	128	1	RE_conv7_2
RE_conv7_2	Deconv	2 × 2	64	2	RE_up3
RE_up3, RE_conv2_2	Concat	-	128	-	RE_concat4
RE_concat4	Conv&ReLU	3 × 3	64	1	RE_conv8_1
RE_conv8_1	Conv&ReLU	3 × 3	64	1	RE_conv8_2
RE_conv8_2	Deconv	2 × 2	32	2	RE_up4
RE_up4, RE_conv1_2	Concat	-	64	-	RE_concat5
RE_concat5	Conv&ReLU	3 × 3	32	1	RE_conv9_1
RE_conv9_1	Conv&ReLU	3 × 3	256	1	RE_conv9_2
RE_conv9_2	Conv	3 × 3	3	1	RE_conv10
RE_conv10	Sigmoid	-	3	-	RE_reflectance

Πίνακας 3. Αρχιτεκτονική ΣΝΔ Επαναφοράς.

Το ΣΝΔ επαναφοράς δέχεται ως εισόδους τους χάρτες ανακλαστικότητας και φωτισμού και στην έξοδό του εφαρμόζει τη συνάρτηση κόστους:

$$\mathcal{L}^{RR} := \|\hat{R}_x - R_b\|_2^2 - \text{SSIM}(\hat{R}_x, R_b) + \|\nabla \hat{R} - \nabla R_b\|_2^2$$



- Μορφοποιήθηκε:** Αγγλικά (Ηνωμένων Πολιτειών)
- Μορφοποιήθηκε:** Γραμματοσειρά: 14 στ., Αγγλικά (Ηνωμένων Πολιτειών)
- Μορφοποιήθηκε:** Γραμματοσειρά: 14 στ.
- Μορφοποιήθηκε:** Γραμματοσειρά: 14 στ.
- Μορφοποιήθηκε:** Γραμματοσειρά: (Προεπιλεγμένη) Cambria Math, 14 στ.
- Μορφοποιήθηκε:** Γραμματοσειρά: (Προεπιλεγμένη) Cambria Math, 14 στ.
- Μορφοποιήθηκε:** Γραμματοσειρά: (Προεπιλεγμένη) Cambria Math, 14 στ.
- Μορφοποιήθηκε:** Γραμματοσειρά: 14 στ.
- Μορφοποιήθηκε:** Γραμματοσειρά: 14 στ., Αγγλικά (Ηνωμένων Πολιτειών)
- Μορφοποιήθηκε:** Γραμματοσειρά: 14 στ.
- Μορφοποιήθηκε:** Αγγλικά (Ηνωμένων Πολιτειών)
- Μορφοποιήθηκε:** Αγγλικά (Ηνωμένων Πολιτειών)



Εικόνα 3.8. Επαναφορά Ανακλαστικότητα. [9]

Στην εικόνα 3.8 παρατηρούνται τα αποτελέσματα του στρώματος επαναφοράς ανακλαστικότητας (2^η σειρά) σε εικόνες χαμηλού φωτισμού (1^η σειρά) διαφορετικών υποβαθμίσεων.

3.4.2.3 Ρύθμιση Φωτισμού

Στο τελευταίο βήμα της μεθόδου *kind* απαιτείται ένας μηχανισμός, ο οποίος θα μετατρέπει μία εικόνα χαμηλού φωτισμού L_s σε εικόνα επιθυμητού φωτισμού L_t . Η αρχιτεκτονική του ΣΝΔ ρύθμισης βασίζεται στο λόγο:

$$\alpha = \frac{L_t}{L_s}$$

ο οποίος χρησιμοποιείται στο στάδιο της εκπαίδευσης ως εξής:

- Αν $\alpha > 1$, τότε πρέπει να αυξηθεί η φωτεινότητα
- Αν $\alpha \leq 1$, τότε δεν απαιτείται ρύθμιση

Η τιμή του λόγου α δίνεται ως μέρος της εισόδου στο ΣΝΔ, το οποίο αποτελείται από 3 συνελκτικά στρώματα και 1 σιγμοειδές όπως φαίνεται στον πίνακα 4:

Inputs	Operator	Kernel	Output Channels	Stride	Output Name
Decom_illumination, Ratio	Concat	-	2	-	Adjust_concat1
Adjust_concat1	Conv&ReLU	3 × 3	32	1	Adjust_conv1
Adjust_conv1	Conv&ReLU	3 × 3	32	1	Adjust_conv2
Adjust_conv2	Conv&ReLU	3 × 3	32	1	Adjust_conv3
Adjust_conv3	Conv	3 × 3	1	1	Adjust_conv4
Adjust_conv4	Sigmoid	-	1	-	Adjust_illumination

Πίνακας 4. Αρχιτεκτονική ΣΝΔ Ρύθμισης Φωτισμού.

Στην έξοδο του ΣΝΔ εφαρμόζεται η συνάρτηση κόστους:

$$\mathcal{L}^{IA} := \|\hat{\mathbf{L}} - \mathbf{L}_t\|_2^2 + \|\nabla \hat{\mathbf{L}}\| - \|\nabla \mathbf{L}_t\|_2^2,$$

όπου \mathbf{L}_t είναι το \mathbf{L}_h ή το \mathbf{L}_l και $\hat{\mathbf{L}}$ είναι ο χάρτης ρυθμισμένης φωτεινότητας της εισόδου \mathbf{L}_t .

3.4.3 Συγκριτικά Αποτελέσματα της Μεθόδου Kind

Στα παρακάτω οπτικά αποτελέσματα της kind έχει χρησιμοποιηθεί για εκπαίδευση του ΤΝΔ το σετ δεδομένων LOL [13], το οποίο περιέχει 500 ζευγάρια εικόνων χαμηλού και κανονικού φωτισμού. Η εκπαίδευση του ΤΝΔ έχει πραγματοποιηθεί σε σύστημα το οποίο αποτελείται από μία κάρτα γραφικών 2080Ti GPU και επεξεργαστή Intel Core i7-8700 3.20 GHz CPU.



Εικόνα 3.9. Παραδείγματα Εικόνων LOL. [13]

Για την αξιολόγηση της μεθόδου έχουν χρησιμοποιηθεί τα σετ δεδομένων LOL [13], LIME [20], NPE [21] και MEF [22]. Για την ποσοτική σύγκριση χρησιμοποιήθηκαν 4 μετρήσεις, οι οποίες είναι οι PSNR, SSIM, LOE [21] και NIQE [23]. Οι υψηλές τιμές στα PSNR και SSIM δηλώνουν καλύτερη ποιότητα, ενώ στα LOE και NIQE οι επιθυμητές τιμές είναι οι χαμηλότερες. Οι μέθοδοι – ανταγωνιστές που έχουν χρησιμοποιηθεί για τη σύγκριση είναι οι BIMEF [24], SRIE [25], CRM [26], Dong [27], LIME [20], MF [28], RRM [29], Retinex – Net [13], GLAD [30], MSR [31] και NPE [21]. Τα αποτελέσματα παρατίθενται στον πίνακα 5:

Metrics	BIMEF [24]	CRM [26]	Dong [27]	LIME [20]	MF [28]	RRM[29]
PSNR	13.8753	17.2033	16.7165	16.7586	18.7916	13.8765
SSIM	0.5771	0.6442	0.5824	0.5644	0.6422	0.6577
LOE	1456.1	1757.7	1283.2	1909.5	2051.7	2025.5
LOE _{ref}	985.9	926.1	1391.5	1342.4	1042.1	958.7
NIQE	7.5150	7.6865	8.3157	8.3777	8.8770	5.8101
Metrics	SRIE [25]	Retinex-Net [13]	MSR [31]	NPE [21]	GLAD [36]	KinD
PSNR	11.8552	16.7740	13.1728	16.9697	19.7182	20.8665
SSIM	0.4979	0.5594	0.4787	0.5894	0.7035	0.8022
LOE	1745.4	2449.3	2589.4	2076.3	1795.5	2012.2
LOE _{ref}	1199.8	2201.7	2084.8	1643.1	1017.1	977.3
NIQE	7.2869	8.8785	8.1136	8.4390	6.4755	5.1461

Πίνακας 5. Ποσοτική Σύγκριση Μεθόδων Ενίσχυσης. [9]



a) Είσοδος

b) BIMEF [24]

c) SRIE [25]

d) GLAD [36]



e) CRM [26] f) LIME [20] g) MF [28]

h) Retinex – Net [13]



i) MSR [31]

j) NPE [21]

k) DONG [27]

l) Kind

Εικόνα 3.10. Οπτική Σύγκριση Μεθόδων Ενίσχυσης



a) Είσοδος b) BIMEF [24] c) SRIE [25] d) GLAD [36]



e) CRM [26] f) LIME [20] g) MF [28] h) Retinex - Net [13]



i) MSR [31] j) NPE [21] k) DONG [27] l) Kind

Εικόνα 3.11. Οπτική Σύγκριση Μεθόδων Ενίσχυσης



a) Είσοδος b) BIMEF [24] c) SRIE [25] d) GLAD [36]



e) CRM [26] f) LIME [20] g) MF [28] h) Retinex - Net [13]



i) MSR [31]

j) NPE [21]

k) DONG [27]

l) Kind

Εικόνα 3.12. Οπτική Σύγκριση Μεθόδων Ενίσχυσης



a) Είσοδος

b) BIMEF [24]

c) SRIE [25]

d) GLAD [36]



e) CRM [26]

f) LIME [20]

g) MF [28]

h) Retinex – Net [13]



i) MSR [31]

j) NPE [21]

k) DONG [27]

l) Kind

Εικόνα 3.13. Οπτική Σύγκριση Μεθόδων Ενίσχυσης



a) Είσοδος

b) BIMEF [24]

c) SRIE [25]

d) GLAD [36]



e) CRM [26]

f) LIME [20]

g) MF [28]

h) Retinex – Net [13]



i) MSR [31]

j) NPE [21]

k) DONG [27]

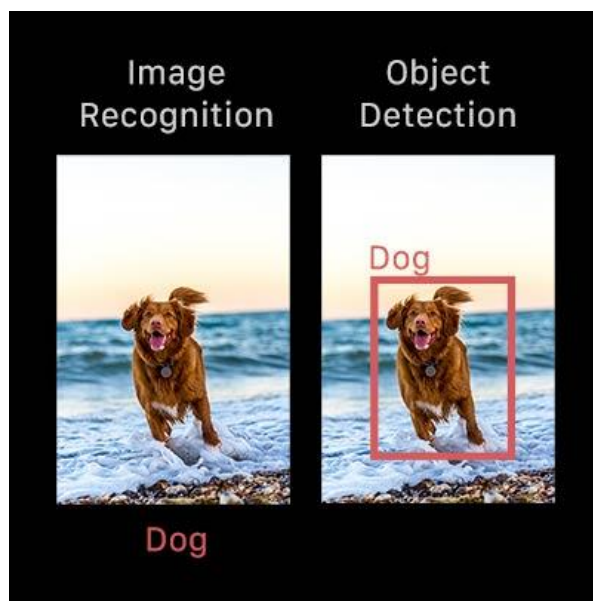
l) Kind

Εικόνα 3.14. Οπτική Σύγκριση Μεθόδων Ενίσχυσης

ΚΕΦΑΛΑΙΟ 4 – ΑΝΑΓΝΩΡΙΣΗ ΑΝΤΙΚΕΙΜΕΝΩΝ

4.1 Εισαγωγή

Η Αναγνώριση Αντικειμένων, ΑΑ (Object Detection – OD), είναι μία πολύ σημαντική και προχωρημένη εφαρμογή της ΜΟ, η οποία χρησιμοποιείται για να αναγνωρίζει στιγμιότυπα οπτικών αντικειμένων. Ο σκοπός της ΑΑ είναι να αναπτύσσει υπολογιστικά μοντέλα με χρήση ΤΝΔ, τα οποία υποδεικνύουν διάφορα οπτικά αντικείμενα σε ψηφιακές φωτογραφίες ή βίντεο. Συγκεκριμένα, η ΑΑ χαράσσει – ζωγραφίζει οριοθετημένα κουτιά γύρω από ανιχνεύσιμα αντικείμενα, τα οποία υποδεικνύουν το που βρίσκονται αυτά τα αντικείμενα μέσα σε μία φωτογραφία ή κινούμενη εικόνα. Συχνά η ΑΑ συγγέται με την αναγνώριση εικόνας (image recognition), αλλά είναι δύο διαφορετικές μέθοδοι. Η ΑΑ χρησιμοποιείται για να αναγνωρίσει το αντικείμενο σε μία εικόνα, ενώ η αναγνώριση εικόνας χρησιμοποιείται για να προσδώσει ετικέτα σε μία εικόνα.



Εικόνα 4.1. Διαφορές ΑΑ – Αναγνώρισης Εικόνας. [google]

Η ΑΑ είναι ένα από τα θεμελιώδη προβλήματα της ΜΟ, καθώς αποτελεί τη βάση άλλων εφαρμογών της ΜΟ, όπως η τοποθέτηση λεζάντας σε φωτογραφία (image captioning), η παρακολούθηση αντικειμένου (object tracking), η αναγνώριση πεζών, η αναγνώριση προσώπου (face detection), η αναγνώριση κειμένου (text detection) και η αναγνώριση πινακίδων κυκλοφορίας.

Τα τελευταία χρόνια, η ραγδαία ανάπτυξη των τεχνικών βαθιάς μάθησης, σε συνδυασμό με την υπολογιστική ισχύ των καρτών γραφικών, έχει συμπαράσχει την ανάπτυξη και απόδοση της ΑΑ. Ένα ευρύ φάσμα εφαρμογών της ΜΟ έχει υλοποιηθεί με χρήση μεθόδων ΑΑ, όπως η παρακολούθηση ιατροφαρμακευτικής περίθαλψης, η αυτόνομη οδήγηση, η παρακολούθηση βίντεο και η ρομποτική όραση. Η ΑΑ μπορεί να υλοποιηθεί με παραδοσιακές τεχνικές επεξεργασίας εικόνας, είτε με χρήση ΤΝΔ βαθιάς μάθησης.

1. Οι Τεχνικές Επεξεργασίας Εικόνας δεν απαιτούν την ύπαρξη δεδομένων εκπαίδευσης και είναι από τη φύση τους ανεπιτήρητες (unsupervised), αλλά η απόδοσή τους είναι περιορισμένη σε πολύπλοκα σενάρια με εικόνες που περιέχουν μερικώς κρυφά αντικείμενα, εικόνες με μονοχρωματικό φόντο, εικόνες με έντονο φωτισμό ή πολλές σκιές κ.α.
2. Οι Τεχνικές Βαθιάς Μάθησης εξαρτώνται από την εκπαίδευση με επίβλεψη (supervised) και υπερτερούν σε πολύπλοκα σενάρια, όμως η απόδοσή τους εξαρτάται από τον όγκο των δεδομένων εκπαίδευσης. Για παράδειγμα, για να εκπαιδευτεί ένας αλγόριθμος ΑΑ με χρήση ΤΝΔ, απαιτείται η χρήση τουλάχιστον 500.000 χαρακτηρισμένων εικόνων (labeled images).

Σήμερα, η χρήση τεχνικών βαθιάς μάθησης στην ΑΑ είναι ευρέως αποδεκτή από τους ερευνητές και χρησιμοποιείται από εταιρίες υψηλής τεχνολογίας για την ανάπτυξη σχετικών προϊόντων.

4.1.1 Ιστορική Αναδρομή

Το πεδίο της ΑΑ αναπτύσσεται και εξελίσσεται τα τελευταία 20 χρόνια και μπορεί να διαχωριστεί σε δύο διαφορετικές ιστορικές περιόδους:

- Πριν το 2014 – Περίοδος της παραδοσιακής ΑΑ**

1. Αλγόριθμος Viola – Jones (2001)
2. Ανιχνευτής HOG (2006)

3. Deformable Part Model Detector, DPM (2008)

• Μετά το 2014 – Περίοδος της Βαθιάς Μάθησης

➤ Σημαντικότεροι αλγόριθμοι δύο επιπέδων:

1. RCNN και SPPNet (2014)
2. Fast RCNN και Faster RCNN (2015)
3. Mask R-CNN (2017)
4. Pyramid Networks/FPN (2017)
5. G-RCNN (2021)

➤ Σημαντικότεροι αλγόριθμοι ενός επιπέδου:

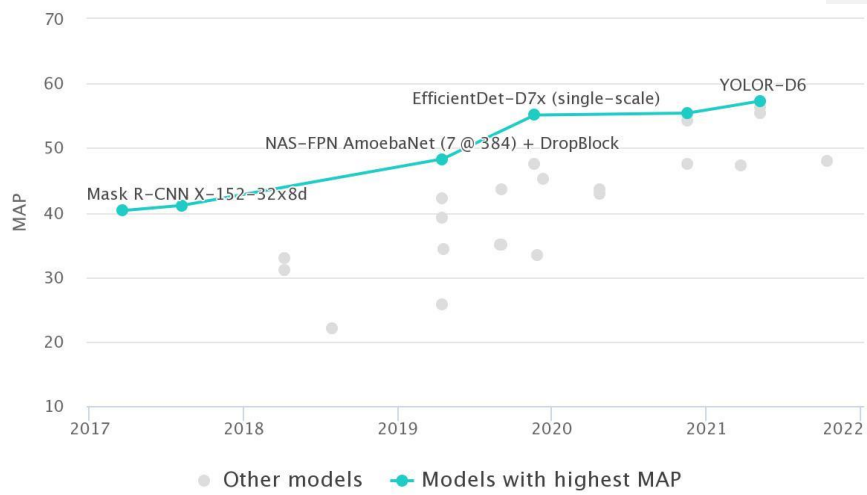
1. YOLO (2015)
2. SSD (2016)
3. RetinaNet (2017)
4. YOLOv3 (2018)
5. YOLOv4 (2020)
6. YOLOR (2021)

Οι ανιχνευτές δύο επιπέδων:

- Προτείνουν την περιοχή που βρίσκεται το αντικείμενο με χρήση συμβατικών μεθόδων MO ή TNA, την οποία ακολουθεί μία ταξινόμηση των αντικειμένων με βάση τα χαρακτηριστικά της προτεινόμενης περιοχής.
- Οι ανιχνευτές δύο επιπέδων πετυχαίνουν μεγαλύτερη ακρίβεια, αλλά απαιτείται μεγαλύτερο χρονικό διάστημα για την υλοποίησή τους, λόγω των πολλών επιπέδων.

Οι ανιχνευτές ενός επιπέδου:

- Δίνουν προτεραιότητα στην ταχύτερη εξαγωγή συμπεράσματος, αλλά δεν είναι αποδοτικοί στην αναγνώριση αντικειμένων με περίεργο σχήμα ή στην αναγνώριση ενός συνόλου από μικρά αντικείμενα.
- Οι δημοφιλέστεροι ανιχνευτές ενός επιπέδου είναι οι YOLO, SSD και RetinaNet.
- Το κύριο πλεονέκτημα των ανιχνευτών ενός επιπέδου είναι η ταχύτητά τους και η απλούστερη δομή τους.



Εικόνα 4.2. Σύγκριση ακρίβειας Αλγορίθμων ΑΑ. [source]

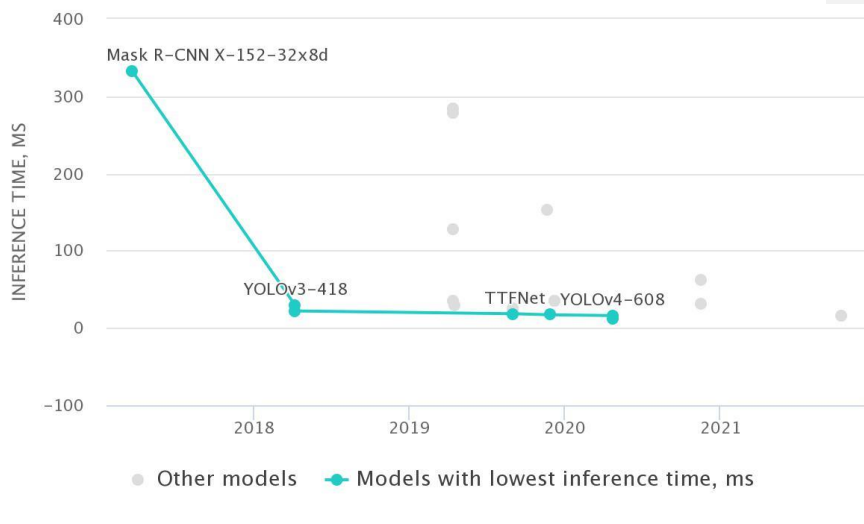
Στην εικόνα 4.2 παρουσιάζεται η ακρίβεια των καλύτερων αλγορίθμων ΑΑ για το 2021 στο σετ δεδομένων Microsoft COCO [MS COCO dataset] με χρήση του δείκτη MAP (Mean Average Precision). Ο καλύτερος αλγόριθμος είναι ο YOLOR και ακολουθείται από τον YOLOv4.

Μορφοποιήθηκε: Ελληνικά

Μορφοποιήθηκε: Ελληνικά

Μορφοποιήθηκε: Ελληνικά

Μορφοποιήθηκε: Ελληνικά



Εικόνα 4.3. Σύγκριση Ταχύτητας Αλγορίθμων ΑΑ. [source]

Μορφοποιήθηκε: Γραμματοσειρά: Πλάγια, Ελληνικά

Στην εικόνα 4.3 παρουσιάζεται η διαφορά ταχύτητας των διάφορων αλγορίθμων ΑΑ, με τον αλγόριθμο YOLOv4 να είναι ο ταχύτερος (16 ms) και να ακολουθούν οι TTFNet (18 ms) και YOLOv3 (29 ms).

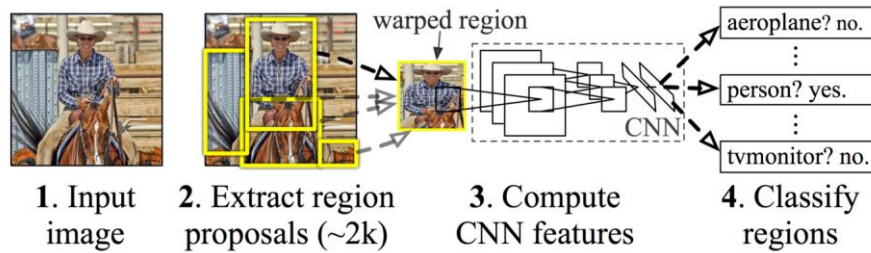
4.2 Αλγόριθμοι Αναγνώρισης Αντικειμένων

Οι διασημότεροι αλγόριθμοι ΑΑ είναι οι R-CNN (Region – Based Convolutional Neural Networks), Fast R-CNN και YOLO. Οι R-CNNs και Fast R-CNN ανήκουν στην οικογένεια των R-CNN αλγορίθμων, ενώ ο YOLO στην οικογένεια των SSD (Single – Shot Detector) αλγορίθμων.

4.2.1 Αλγόριθμος R-CNN

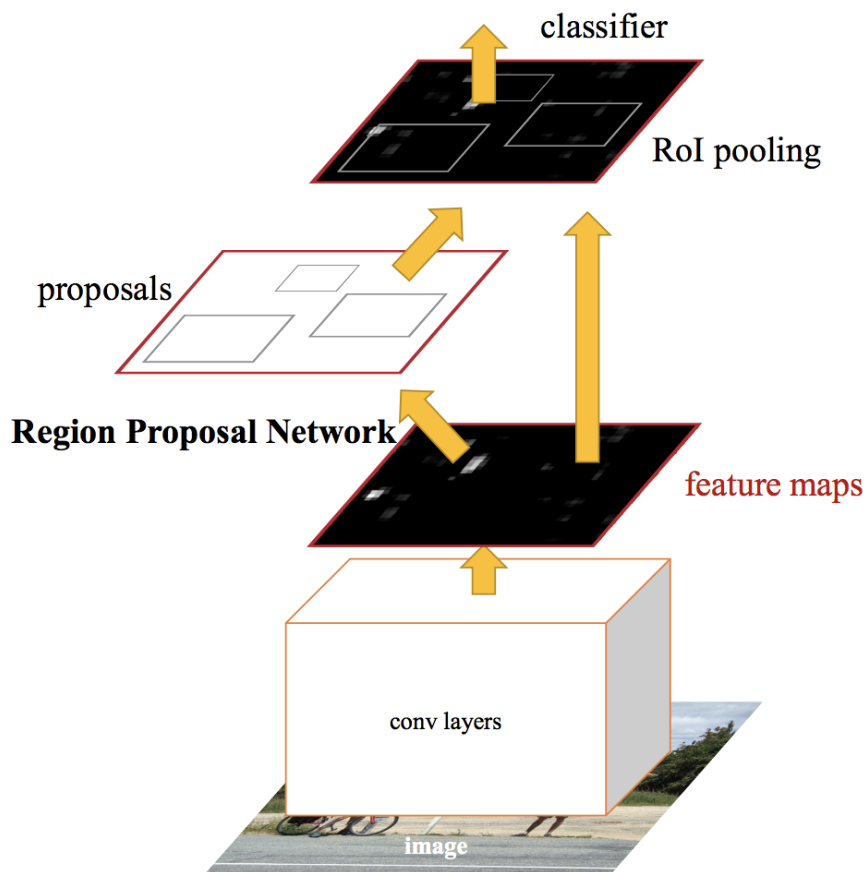
Τα ΣΝΔ (CNNs) είναι ΤΝΔ, τα οποία είναι ικανά να εξάγουν γάρτες χαρακτηριστικών από εικόνες, με σκοπό να επιτύχουν την ΑΑ. Έστω ότι σε μία εικόνα έχουμε 2000 αντικείμενα. Το βασικό πρόβλημα ενός μοντέλου βαθιάς μάθησης, θα είναι η αρχικοποίηση των οριοθετημένων κουτιών γύρω από αυτά τα αντικείμενα. Για το σκοπό αυτό, οι R-CNN [32] αλγόριθμοι χρησιμοποιούν μία μέθοδο επιλεκτικής αναζήτησης, έτσι ώστε να εξάγουν 2000

περιοχές (regions), τις οποίες δέχεται ως είσοδο το ΣΝΔ και στη συνέχεια τις ταξινομεί με χρήση κάποιου άλλου αλγόριθμου ταξινόμησης.



Εικόνα 4.4. Αρχιτεκτονική R-CNN αλγόριθμου. [32]

Το βασικό πρόβλημα του R-CNN είναι η ταχύτητά του, καθώς απαιτεί 47 δευτερόλεπτα για κάθε εικόνα, έτσι ώστε να εξάγει τις προτεινόμενες περιοχές. Για το λόγο αυτό αναπτύχθηκε ο Fast R-CNN αλγόριθμος, ο οποίος προσεγγίζει αρκετά τον R-CNN, μόνο που η εικόνα δίνεται ως είσοδος αρχικά στο ΣΝΔ για την εξαγωγή των χαρτών χαρακτηριστικών και των προτεινόμενων περιοχών. Οπότε η μέθοδος εφαρμόζεται σε εικόνες που έχουν ήδη οριοθετηθεί ως υποψήφιες με αντικείμενα και η επιλεκτική αναζήτηση δεν γίνεται στην αρχική γενική εικόνα.

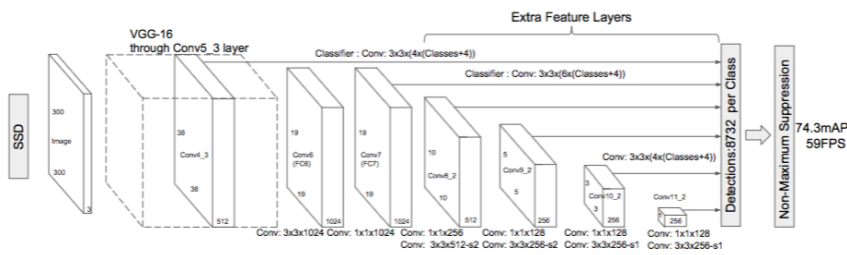


Εικόνα 4.5. Αρχιτεκτονική Fast R-CNN αλγόριθμου. [google]

4.2.2 Αλγόριθμος SSD - Single Shot Detector

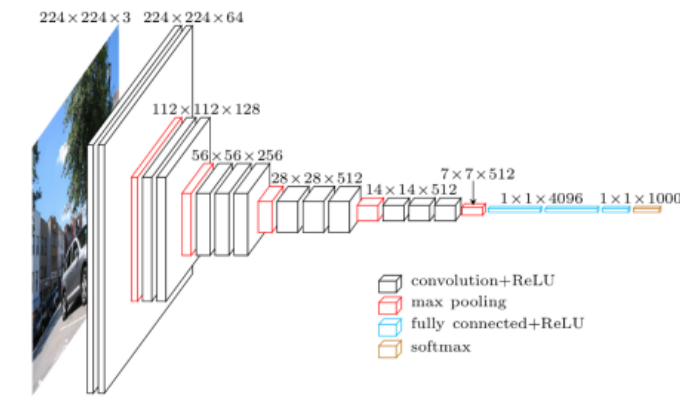
Ο αλγόριθμος SSD βασίζεται στη χρήση ΣΝΔ, τα οποία παράγουν ένα μεγάλο αριθμό οριοθετημένων κουτιών διαφορετικών διαστάσεων πάνω στην εικόνα. Οφείλει το όνομά του στο γεγονός ότι η ταξινόμηση και η ΑΑ επιτυγχάνεται ύστερα από μονό πέρασμα (single forward pass) μέσα από το ΤΝΔ. Το πρώτο βήμα του αλγόριθμου SSD, όπως και του YOLO είναι ο διαχωρισμός της εικόνας σε ένα πλέγμα κελιών, στο οποίο το κάθε κελί είναι υπεύθυνο για την πρόβλεψη ύπαρξης αντικείμενος εντός των ορίων του. Σε κάθε κελί

αντιστοιχεί ένας αριθμός από κουτιά, τα οποία διαφέρουν σε διαστάσεις. Ο σκοπός του κάθε κουτιού είναι να εντοπίσει ακόμα και τις πιο μικρές λεπτομέρειες εντός ενός κελιού της εικόνας. Κάθε κουτί λαμβάνει μία τιμή, ανάλογα με το πόσο εύστοχα έχει οριοθετηθεί εντός του το αντικείμενο.



Εικόνα 4.6. Αρχιτεκτονική SSD αλγόριθμου. [google]

Η αρχιτεκτονική SSD στηρίζεται στην αρχιτεκτονική VGG – 16 [33], η οποία είναι αρκετά αποδοτική σε ταξινόμηση εικόνων πολύ υψηλής ποιότητας.



Εικόνα 4.7. Αρχιτεκτονική VGG. [google]

4.2.3 Αλγόριθμος YOLO

Ο όρος YOLO [34] είναι μία συντομογραφία του όρου “You Only Look Once”. Ο YOLO είναι ένας αλγόριθμος που, σε πραγματικό χρόνο, ανιχνεύει και αναγνωρίζει αντικείμενα σε

μία εικόνα. Ο συγκεκριμένος αλγόριθμος χρησιμοποιεί ΣΝΔ για την ανίχνευση των αντικειμένων και απαιτεί μόνο μία διάδοση μέσω του ΣΝΔ (single forward propagation). Αυτό σημαίνει ότι η πρόβλεψη που επιτυγχάνει το ΣΝΔ, πραγματοποιείται με μία υλοποίηση του αλγόριθμου, σε αντίθεση με τους προγενέστερους αλγόριθμους, που επαναχρησιμοποιούσαν το ΣΝΔ. Η διαφορετική αυτή προσέγγιση επέτυχε εκπληκτικά αποτελέσματα στο πεδίο της ΑΑ και επέδειξε τη σημαντικότητα του συγκεκριμένου αλγόριθμου για τους εξής λόγους:

- Ο αλγόριθμος βελτίωσε την **ταχύτητα** ανίχνευσης, λόγω της υλοποίησής του σε πραγματικό χρόνο.
- **Υψηλή ακρίβεια** με σχετικά μικρά σφάλματα.
- Ο YOLO έχει εξαιρετικές **δυνατότητες εκμάθησης** που του δίνουν τη δυνατότητα να «μάθει» τις αναπαραστάσεις αντικειμένων και να τις εφαρμόσει στην ΑΑ.

4.2.3.1 Μεθοδολογία

Κάθε εικόνα διαχωρίζεται σε N όμοια πλέγματα - κελιά, που το καθένα έχει διαστάσεις $S_x S_y$. Το κάθε πλέγμα - κελί είναι υπεύθυνο για την ανίχνευση του αντικειμένου που μπορεί να περιέχει και προβλέπει ένα B οριοθετημένο κουτί μαζί με το σκορ αυτοπεποίθησης (confidence score) του B . Τα σκορ αυτοπεποίθησης αποτυπώνουν το βαθμό αυτοπεποίθησης του μοντέλου σχετικά με την ύπαρξη αντικειμένου εντός του ορίου τους, καθώς και την ακρίβεια πρόβλεψης. Η αυτοπεποίθηση ορίζεται ως :

$$\Pr(\text{Object}) \cdot IOU_{Pred}^{truth}$$

- $\Pr(\text{Object})$: η πιθανότητα ύπαρξης αντικειμένου στο κουτί.
- IOU_{Pred}^{truth} : ο δείκτης που αξιολογεί τα προβλεπόμενα κουτιά (predicted boxes) σε σχέση με τα πραγματικά (ground truth), μετρώντας την έκταση της επικάλυψης του ενός κουτιού πάνω στο άλλο.


Μορφοποιήθηκε: Γραμματοσειρά: 14 στ.

Μορφοποιήθηκε: Γραμματοσειρά: 14 στ.

Μορφοποιήθηκε: Γραμματοσειρά: 14 στ., Όχι Πλάγια

Μορφοποιήθηκε: Γραμματοσειρά: 14 στ.

Μορφοποιήθηκε: Γραμματοσειρά: 14 στ.

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$


Εικόνα 4.8. Υπολογισμός δείκτη IoU. [google]

Αν δεν υπάρχει κάποιο αντικείμενο στο πλέγμα – κελί, τότε το σκορ αυτοπεποίθησης ισούται με μηδέν, αλλιώς η αυτοπεποίθηση επιθυμούμε να είναι ίση με τον IoU, δηλαδή ίση με 1. Το ολικό όμως ταίριασμα μεταξύ των δύο κουτιών είναι μη εφικτό, οπότε αν $\text{IoU} > 0.5$, τότε έχει επιτευχθεί μία καλή πρόβλεψη.

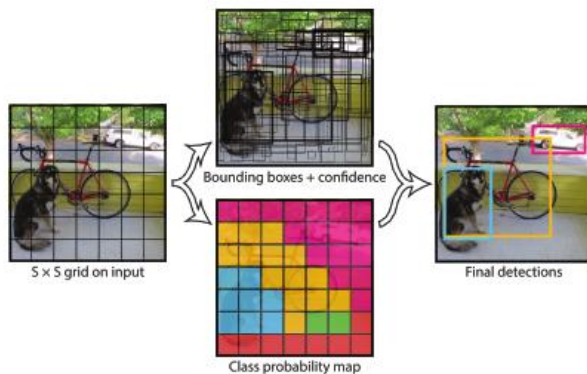


Εικόνα 4.9. Παραδείγματα IoU. [google]

Κάθε πλέγμα – κελί,

- προβλέπει B οριοθετημένα κουτιά που το καθένα έχει ένα σκορ αυτοπεποίθησης,
- ανιχνεύει μόνο ένα αντικείμενο, ανεξάρτητα του αριθμού των κουτιών B,
- προβλέπει C υποθετικές κλάσεις πιθανοτήτων.

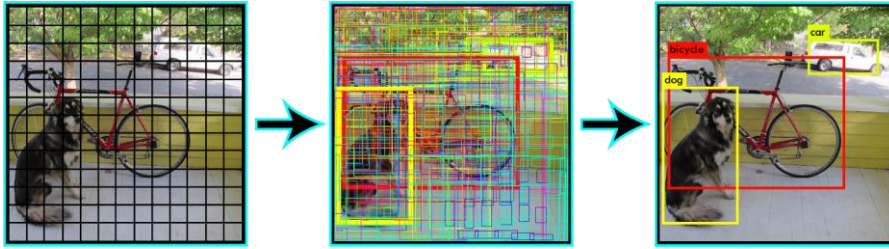
Για παράδειγμα, για την αξιολόγηση του PASCAL VOC [35], ο YOLO χρησιμοποιεί 7x7 πλέγματα (SxS), 2 οριοθετημένα κουτιά (B) και 20 κλάσεις πιθανοτήτων (C).



Εικόνα 4.10. Το μοντέλο YOLO. [34]

Κάθε οριοθετημένο κουτί αποτελείται από 5 προβλέψεις: x , y , w , h και το σκορ αποπεποίησης του κουτιού. Οι (x,y) αποτελούν τις συντεταγμένες του κέντρου του κουτιού σε σχέση με τα άκρα του κουτιού. Το πλάτος w και το ύψος h προβλέπονται σε σχέση με τις διαστάσεις της εικόνας. Τα x,y,w και h παίρνουν τιμές μεταξύ 0 και 1. Κάθε κελί έχει 20 κλάσεις πιθανοτήτων, οι οποίες αντιστοιχούν στην πιθανότητα το ανιχνεύσιμο αντικείμενο να ανήκει σε μία συγκεκριμένη κλάση αντικειμένων. Έτσι, η πρόβλεψη του YOLO είναι ένας τανυστής σχήματος $(S, S, B \times 5 + C) = (7, 7, 2 \times 5 + 20) = (7, 7, 30)$.

Η πρόβλεψη του τανυστή $(7, 7, 30)$ γίνεται με χρήση ενός ΣΝΔ, το οποίο μειώνει τις χωρικές διαστάσεις μιας εικόνας σε 7x7 με 1024 κανάλια εξόδου. Ακολούθως, ο YOLO εφαρμόζει γραμμική παλινδρόμηση (linear regression) με τη βοήθεια 2 συνδεδεμένων στρωμάτων, τα οποία εξάγουν τα 7x7x2 οριοθετημένα κουτιά με τις προβλέψεις, όπως φαίνεται στο μεσαίο κουτί της εικόνας 4.11. Στο δεξί κουτί της εικόνας 4.11 φαίνονται οι τελικές προβλέψεις (σκορ μεγαλύτερο του 0.25).



Εικόνα 4.11. Βήματα εξαγωγής πρόβλεψης του YOLO. [source]

Το σκορ αυτοπεποίθησης κάθε κλάσης για κάθε κουτί υπολογίζεται ως:

$$Pr(class_i) \cdot IoU = (Pr(Object) \cdot IoU) \cdot Pr(class_i/object)$$

Όπου

- $Pr(Object)$: η πιθανότητα ύπαρξης αντικειμένου στο κουτί.
- $Pr(class_i/object)$: η δοθείσα πιθανότητα το αντικείμενο να ανήκει σε υπάρχουσα κλάση αντικειμένων.
- $Pr(class_i)$: η πιθανότητα το αντικείμενο να ανήκει στην κλάση

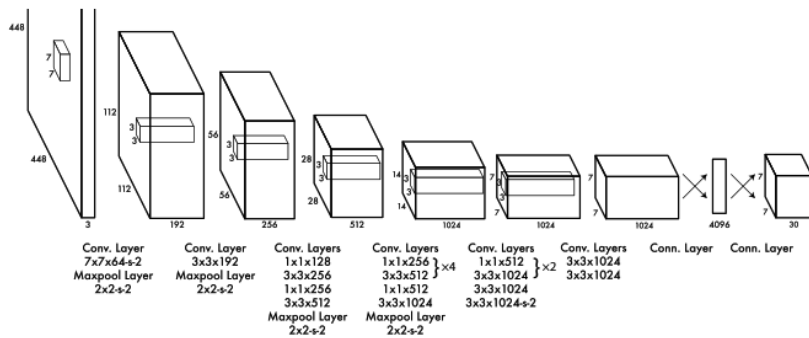
4.2.3.2 Αρχιτεκτονική ΤΝΔ του YOLO

Ο YOLO αποτελείται από 24 συνελκτικά στρώματα και 2 πλήρως συνδεδεμένα στρώματα, τα οποία έπονται των συνελκτικών. Ορισμένα συνελκτικά στρώματα χρησιμοποιούνε 1x1 στρώματα για να μειώσουν το βάθος των χαρτών χαρακτηριστικών. Το τελευταίο συνελκτικό στρώμα δίνει στην έξοδο τον ταυστή (7, 7, 1024), ο οποίος στη συνέχεια αλλάζει διαστάσεις. Τέλος, χρησιμοποιούνται τα 2 πλήρως συνδεδεμένα στρώματα για την εφαρμογή της γραμμικής παλινδρόμησης, η οποία δίνει στην έξοδό της 7x7x30 παραμέτρους, δηλαδή 2 προβλέψεις για κάθε τοποθεσία στην εικόνα.

Μορφοποιήθηκε: Ελληνικά

Αλλαγή κωδικού πεδίου

Μορφοποιήθηκε: Ελληνικά



Εικόνα 4.12. Αρχιτεκτονική ΤΝΔ του αλγόριθμου YOLO. [34]

Ο YOLO προβλέπει πολλά οριοθετημένα κουτιά για κάθε πλέγμα – κελί, αλλά το ένα από αυτά θα πρέπει να είναι υπεύθυνο για το αντικείμενο. Για το σκοπό αυτό επιλέγουμε το κουτί με το μεγαλύτερο IoU. Η τακτική αυτή οδηγεί σε εξειδίκευση των προβλέψεων για κάθε κουτί, με αποτέλεσμα οι προβλέψεις να γίνονται όλο και πιο ακριβέστερες. Η εξειδίκευση αυτή πραγματοποιείται με χρήση μιας συνάρτησης κόστους, η οποία αποτελείται από:

- Το κόστος ταξινόμησης (classification loss),
- Το κόστος τοπικοποίησης (localization loss),
- Το κόστος αυτοπεποίθησης (confidence loss).

Αν ένα αντικείμενο ανιχνευθεί, τότε το κόστος ταξινόμησης υπολογίζεται ως:

$$\sum_{i=0}^{s^2} 1_i^{obj} \cdot \sum (P_i(c) - \hat{P}_i(c))^2$$

όπου

$1_i^{obj} = 1$ αν το αντικείμενο εμφανίζεται στο κελί i , αλλιώς είναι 0.

$\hat{P}_i(c)$ δηλώνει την υποθετική κλάση πιθανότητας για την κλάση C σε ένα κελί i .

Το κόστος τοπικοποίησης μετράει το σφάλμα μεγέθους του κουτιού, καθώς και το σφάλμα της περιοχής που εμφανίστηκε:

- Μορφοποιήθηκε: Γραμματοσειρά: Πλάγια
- Μορφοποιήθηκε: Γραμματοσειρά: Πλάγια, Ελληνικά
- Μορφοποιήθηκε: Ελληνικά
- Μορφοποιήθηκε: Ελληνικά
- Μορφοποιήθηκε: Γραμματοσειρά: Πλάγια
- Μορφοποιήθηκε: Γραμματοσειρά: Πλάγια, Ελληνικά
- Μορφοποιήθηκε: Γραμματοσειρά: Πλάγια
- Μορφοποιήθηκε: Γραμματοσειρά: Πλάγια, Ελληνικά

$$\lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right] \\ + \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{obj}} \left[(\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2 \right]$$

όπου

$1_{ij}^{\text{obj}} = 1$, αν το j κουτί στο κελί i είναι υπεύθυνο για την ανίχνευση αντικειμένου, αλλιώς είναι 0.

λ_{coord} είναι ο συντελεστής που αυξάνει το βάρος για το κόστος των συντεταγμένων του κουτιού.

Αν το αντικείμενο ανιχνεύθηκε μέσα σε ένα κουτί, το κόστος αυτοπεποίθησης μετρά την αναγνωρισιμότητά του:

$$\sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{obj}} (C_i - \hat{C}_i)^2$$

όπου

C_i είναι το σκορ αυτοπεποίθησης του κουτιού j στο κελί i .

$1_{ij}^{\text{obj}} = 1$, αν το j κουτί στο κελί i είναι υπεύθυνο για να ανιχνεύσει το αντικείμενο, αλλιώς είναι ίσο με 0.

Αν το αντικείμενο δεν έχει ανιχνευθεί μέσα στο κουτί, τότε το κόστος αυτοπεποίθησης είναι:

$$\lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{noobj}} (C_i - \hat{C}_i)^2$$

Μορφοποιήθηκε: Ελληνικά

Μορφοποιήθηκε: Γραμματοσειρά: Πλάγια

Μορφοποιήθηκε: Ελληνικά

Μορφοποιήθηκε: Ελληνικά

Μορφοποιήθηκε: Ελληνικά

όπου

1_{ij}^{noobj} είναι το συμπλήρωμα του 1_{ij}^{obj}

Μορφοποιήθηκε: Ελληνικά

C_i είναι το σκορ αυτοπεποίθησης του κουτιού j στο κελί i .

λ_{noobj} είναι ο συντελεστής που μειώνει τα βάρη κόστους, αν ανιχνευθεί φόντο.

Ο λ_{noobj} είναι απαραίτητος, διότι τα περισσότερα κουτιά σε μία εικόνα δεν περιέχουν κάποιο αντικείμενο. Άρα υπάρχει κίνδυνος το μοντέλο να εκπαιδευτεί στο να ανιχνεύει φόντο χωρίς αντικείμενα. Με το συντελεστή λ_{noobj} αποφεύγουμε αυτόν τον κίνδυνο.

Η τελική **συνάρτηση κόστους** υπολογίζει συνολικά όλα τα κόστη:

$$\begin{aligned} & \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right] \\ & + \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{obj}} \left[(\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2 \right] \\ & + \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{obj}} (C_i - \hat{C}_i)^2 \\ & + \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{noobj}} (C_i - \hat{C}_i)^2 \\ & + \sum_{i=0}^{S^2} \mathbb{1}_i^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2 \end{aligned}$$

4.2.3.3 Πλεονεκτήματα και Περιορισμοί

Πλεονεκτήματα:

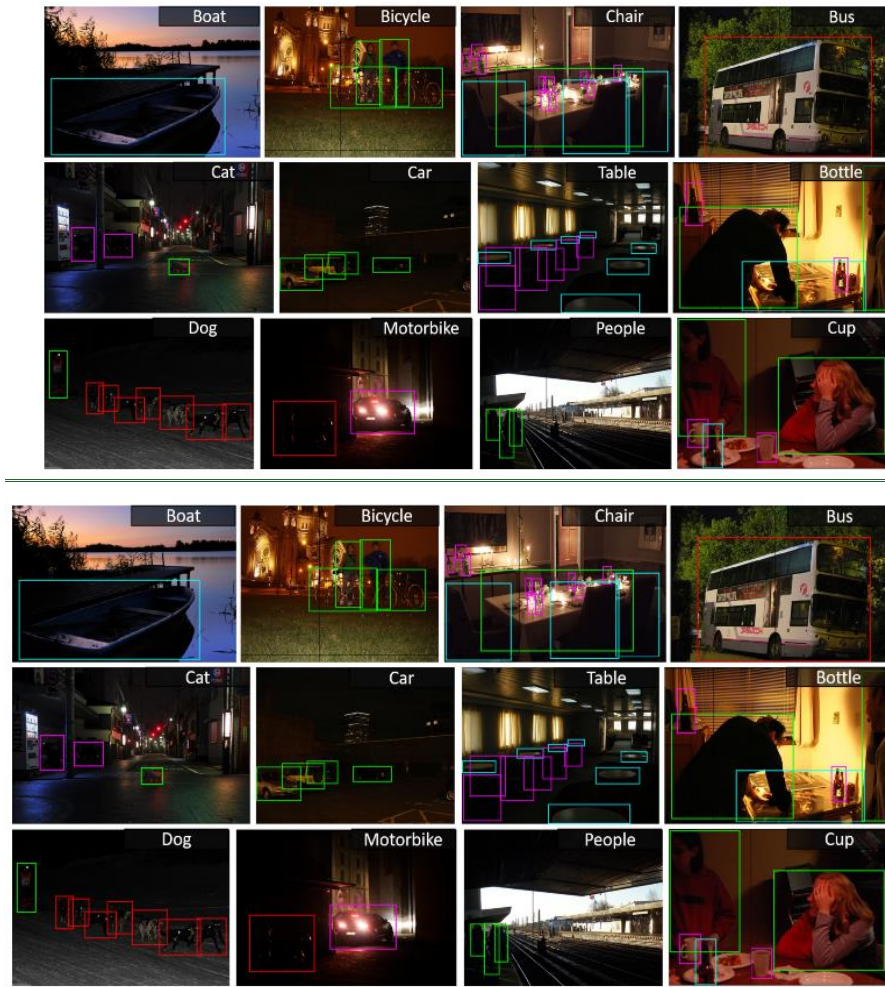
- Είναι γρήγορος και ενδείκνυται για επεξεργασία σε πραγματικό χρόνο.
- Οι προβλέψεις πραγματοποιούνται από ένα μόνο ΤΝΔ.
- Είναι πιο γενικευμένος και πιο αποδοτικός έναντι άλλων μεθόδων τόσο σε φυσικές εικόνες, είτε σε εικόνες άλλου τύπου, όπως τα έργα τέχνης.
- Ανιχνεύει αντικείμενα σε κάθε πλέγμα – κελί ξεχωριστά.

Περιορισμοί:

- Εφόσον το μοντέλο μαθαίνει να προβλέπει οριοθετημένα κουτιά από τα δεδομένα, δυσκολεύεται να επιτύχει προβλέψεις σε διαφοροποιημένες εικόνες – αντικείμενα.
- Στο στάδιο της εκπαίδευσης, η συνάρτηση κόστους αντιμετωπίζει τα σφάλματα με τον ίδιο τρόπο, είτε έχουμε μικρά κουτιά, είτε μεγάλα. Ένα μικρό σφάλμα σε ένα μεγάλο κουτί είναι γενικά αβλαβές, αλλά ένα μικρό σφάλμα σε ένα μικρό κουτί έχει μεγαλύτερη επίδραση στο *IoU*.
- Επιβάλλει ισχυρούς χωρικούς περιορισμούς στις προβλέψεις των οριοθετημένων κουτιών, καθώς κάθε πλέγμα – κελί προβλέπει 2 κουτιά και έχει μόνο μία κλάση. Αυτό έχει ως αποτέλεσμα, τον μικρό αριθμό αντικειμένων που μπορεί το μοντέλο να προβλέψει. Επίσης, με αυτή τη μέθοδο δυσκολεύεται να αναγνωρίσει μικρά αντικείμενα που εμφανίζονται σε ομάδες, όπως ένα σμήνος πουλιών.

ΚΕΦΑΛΑΙΟ 56 – ΥΛΟΠΟΙΗΣΗ ΚΑΙ ΑΠΟΤΕΛΕΣΜΑΤΑ

56.1 Το Σετ Δεδομένων



Εικόνα 5.1. Παραδείγματα Εικόνων ExDark με Ετικέτες Αντικειμένων. [35]

Στην παρούσα εργασία έχει χρησιμοποιηθεί το Exclusively Dark Dataset – ExDark [35], το οποίο περιέχει τη μεγαλύτερη συλλογή εικόνων χαμηλού φωτισμού και συγκεκριμένα 7.363 εικόνες που ελήφθησαν σε περιβάλλοντα χαμηλού φωτισμού έως το λυκόφως. Οι εικόνες

προέρχονται από το διαδίκτυο και από μηχανές αναζήτησης όπως flickr.com, Photo-bucket.com, Imgur.com, Devianart.com, Gettyimages.com και αναζήτηση μέσω google. Επίσης περιέχεται και δείγμα εικόνων από σετ δεδομένων όπως τα PASCAL VOC, ImageNet και Microsoft COCO (Πίνακας 1).

Dataset		Low-light image
Microsoft COCO	Training	149 (0.18%)
	Validation	163 (0.4%)
	Testing 2014	138 (0.34%)
	Testing 2015	115 (0.14%)
	Total	565 (0.23%)
ImageNet	Training 2012	255 (0.02%)
	Validation 2012	38 (0.08%)
	Testing 2012	51 (0.05%)
	Validation 2013	12 (0.26%)
	Testing 2013	22 (0.23%)
	Training 2014	72 (0.12%)
	Total	450 (0.03%)
PASCAL VOC	2007	123 (1.24%)
	2008	72 (1.66%)
	2009	43 (1.58%)
	2010	50 (1.43%)
	2011	48 (1.32%)
	2012	17 (0.79%)
	Total	353 (1.34%)
Exclusively Dark		7,363 (100%)

Πίνακας 1. Εικόνες Χαμηλού Φωτισμού από Δημοσιευμένα Σετ Δεδομένων που περιέχονται στο ExDark. [35]

Επίσης, περιέχονται 12 κλάσεις αντικειμένων, οι οποίες σημειώνονται πάνω στις εικόνες, αλλά και στα οριοθετημένα κουτιά των αντικειμένων που εντοπίζονται στις εικόνες αυτές. Οι κλάσεις αυτές περιέχουν *ποδήλατα, βάρκες, καρέκλες, λεωφορεία, γάτες, αυτοκίνητα, τραπέζια, μπουκάλια, σκύλους, μοτοσυκλέτες, ανθρώπους και φλυτζάνια*. Η ετικετοποίηση – σημείωση των εικόνων – δεδομένων πραγματοποιείται σε 2 στάδια. Στο πρώτο στάδιο οι εικόνες ταξινομούνται σε 12 κλάσεις, ανεξάρτητα με το αν το αντικείμενο βρίσκεται στο μεγαλύτερο μέρος της εικόνας ή όχι. Στο δεύτερο στάδιο, η ετικετοποίηση – σημείωση των αντικειμένων γίνεται με βάση τα οριοθετημένα κουτιά που περικλείουν στο εσωτερικό τους τα αντικείμενα. Κάθε αντικείμενο που βρίσκεται εντός του κουτιού κατηγοριοποιείται σε μία εκ των 12 κλάσεων με χρήση του *Piotr's Computer Vision Matlab toolbox (source)*.

Μορφοποιήθηκε: Ελληνικά

Μορφοποιήθηκε: Ελληνικά

Μορφοποιήθηκε: Ελληνικά

Μορφοποιήθηκε: Ελληνικά

Μορφοποιήθηκε: Ελληνικά

Μορφοποιήθηκε: Ελληνικά

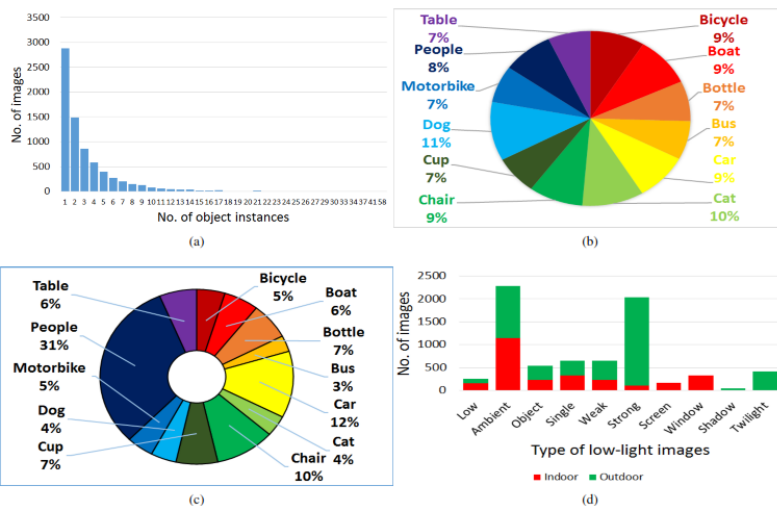
Μορφοποιήθηκε: Ελληνικά

Αλλαγή κωδικού πεδίου

Dataset	Exclusively Dark	Microsoft COCO
Class	Number of Image	Number of image
Bicycle	652	603
Boat	679	650
Bottle	547	650
Bus	527	564
Car	638	650
Cat	735	650
Chair	648	651
Cup	519	650
Dog	801	650
Motorbike	503	644
People	609	650
Table	505	650
Total	7,363	7,662

Πίνακας 2. Αριθμός Εικόνων ανά κλάση Αντικειμένου. [35]

Στην εικόνα 5.1 απεικονίζεται η στατιστική ανάλυση των εικόνων και στις περισσότερες από αυτές περιέγεται ένα ~~στιγμιότυπο περιστατικό~~ ενός αντικειμένου. Υπάρχει όμως και ένας αξιοσημείωτος αριθμός εικόνων που περιέχουν πολλά ~~στιγμιότυπα περιστατικά~~ αντικειμένων και συγκεκριμένα σε μία εικόνα βρέθηκαν 58 ετικέτες μέσα στα οριοθετημένα κουτιά (2a). Επίσης παρατηρούμε ότι οι περισσότερες σημειώσεις αντικειμένων αφορούν την κλάση των ανθρώπων (2c). Σε 23.710 σημειώσεις αντικειμένων, υπάρχουν 7.460 άνθρωποι, το οποίο ενισχύει την αναγνώριση πεζών.



Εικόνα 5.23. Στατιστική Ανάλυση του ExDark. (α) Αριθμός Αντικειμένων ανά εικόνα, (β) Ποσοστό Κλάσης ανά εικόνα, (γ) Ποσοστό Κλάσης στο ExDark, (δ) Τύποι Φωτεινότητας Εικόνων

Μορφοποιήθηκε: Ελληνικά

περιβάλλοντα εσωτερικού και εξωτερικού χώρου:

- **Low** : Εικόνες με πολύ χαμηλή φωτεινότητα.
- **Ambient** : Εικόνες με αδύναμη φωτεινότητα.
- **Object** : Εικόνες με αντικείμενα έντονα φωτισμένα και με το γύρω περιβάλλον σκοτεινό.
- **Single** : Εικόνες που η μοναδική πηγή φωτός είναι ορατή.
- **Weak** : Εικόνες με πολλές αλλά αδύναμες ορατές πηγές φωτός.
- **Strong** : Εικόνες με πολλές ορατές και σχετικά έντονα φωτεινές πηγές φωτός.
- **Screen** : Εικόνες εσωτερικού χώρου με ορατές φωτεινές οθόνες (οθόνες υπολογιστών, τηλεοράσεις).
- **Window** : Εικόνες εσωτερικού χώρου με φωτεινά παράθυρα ως πηγές φωτός.
- **Shadow** : Εικόνες εξωτερικού χώρου τραβηγμένες στο φως της ημέρας, αλλά τα αντικείμενα βρίσκονται σε σκιά.
- **Twilight** : Εικόνες εξωτερικού χώρου τραβηγμένες στο λυκόφως.

Μορφοποιήθηκε: Ελληνικά

Μορφοποιήθηκε: Γραμματοσειρά: Όχι Έντονα, Ελληνικά

Μορφοποιήθηκε: Ελληνικά

~~Εικόνα 5.1. Παραδείγματα Εικόνων ExDark με Ετικέτες Αντικειμένων. [35]~~

Μορφοποιήθηκε: Ελληνικά

Εικόνα 5.32. Παραδείγματα Εικόνων Χαμηλού Φωτισμού Διαφόρων Τύπων στο ExDark.
[35]

5.2 Αποτελέσματα της μεθόδου Kind

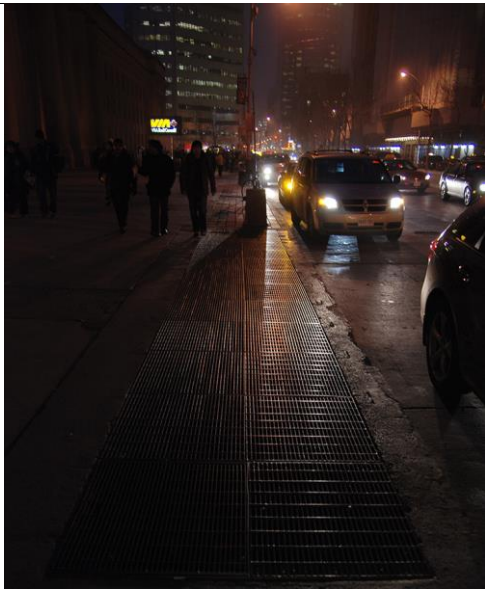
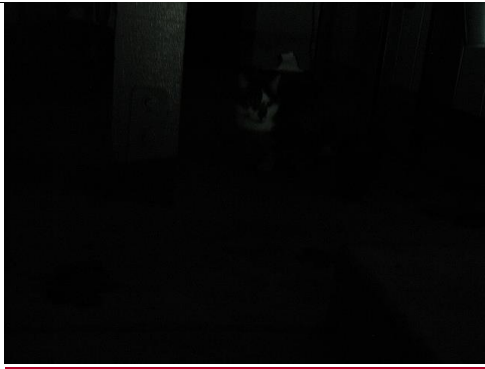
Η υλοποίηση της μεθόδου πραγματοποιήθηκε σε υπολογιστικό σύστημα που αποτελείται από:

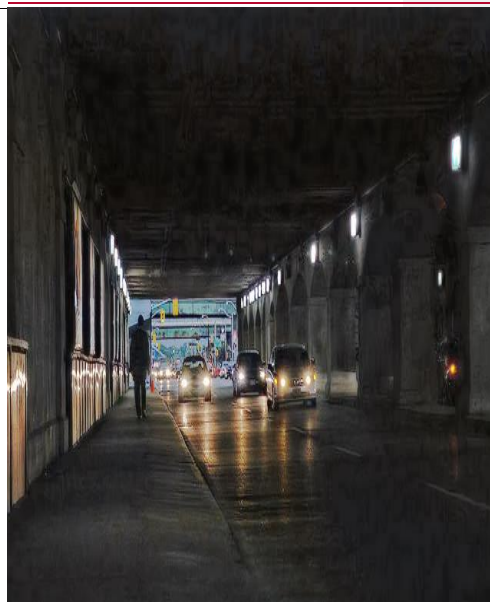
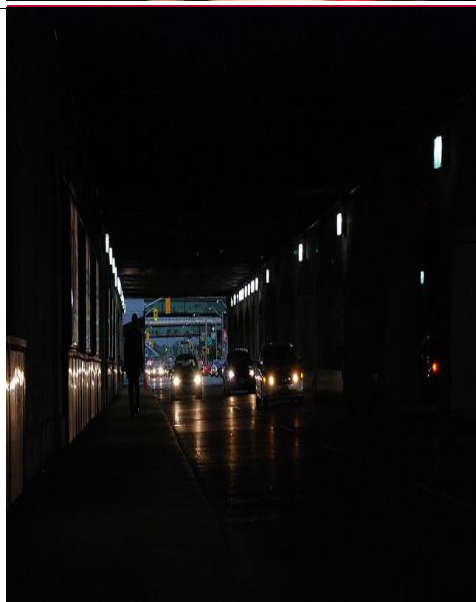
- CPU: Intel Core i9 – 7960X 2.80 GHz
- GPU: NVIDIA QUADRO P6000 24 GB GDDR5X
- RAM: 64 Gb (2x32)
- Operating System: UBUNTU 10

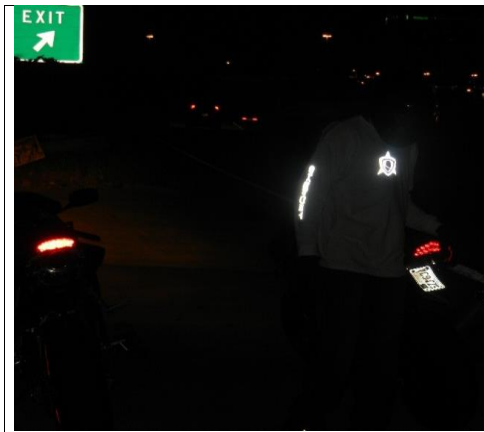
και είγαμε τα ενδεικτικά αποτελέσματα του παρακάτω πίνακα 3.

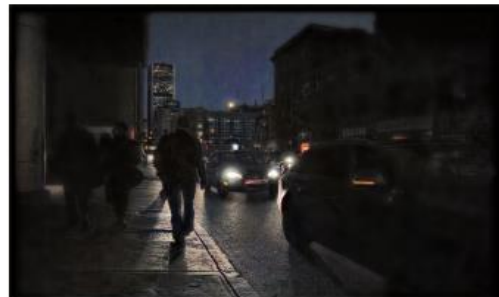
Αρχική Εικόνα

Ενισχυμένη Εικόνα



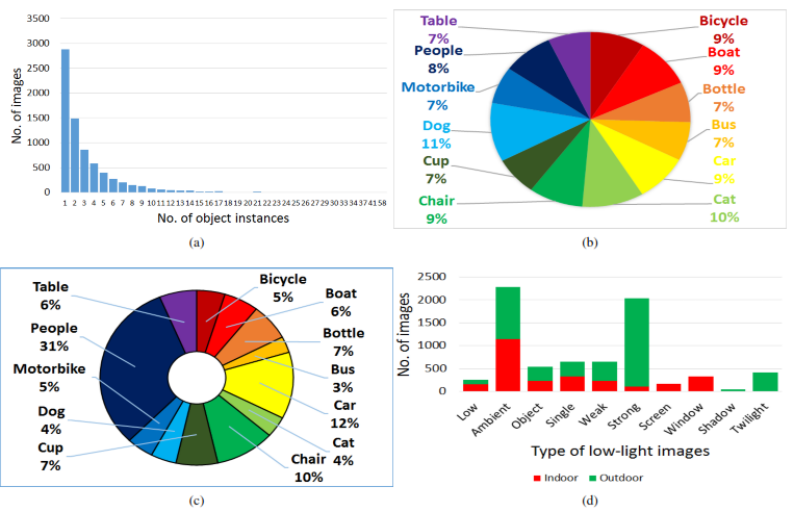








Πίνακας 3. Αποτελέσματα της μεθόδου Kind.



Μορφοποιήθηκε: Πλήρης

5.3 Αποτελέσματα του Μοντέλου YOLO *Εικόνα 5.3. Στατιστική Ανάλυση του ExDark. (α) Αριθμός Αντικειμένων ανά εικόνα, (β) Ποσοστό Κλάσης ανά εικόνα, (γ) Ποσοστό Κλάσης στο ExDark, (δ) Τύποι Φωτισμού Εικόνας*

Στην ενότητα αυτή περιλαμβάνονται τα οπτικά αποτελέσματα και οι συγκρίσεις της εφαρμογής του μοντέλου YOLO στις σκοτεινές και στις αντίστοιχες ενισχυμένες εικόνες. Σύμφωνα με τα αποτελέσματα του Πίνακα 4, οι ενισχυμένες εικόνες παράγουν καλύτερα

Μορφοποιήθηκε: Στοιχισμένο στο κέντρο

Μορφοποιήθηκε: Ελληνικά

Μορφοποιήθηκε: Ελληνικά

Μορφοποιήθηκε: Ελληνικά

Μορφοποιήθηκε: Ελληνικά

Μορφοποιήθηκε: Ελληνικά

Μορφοποιήθηκε: Ελληνικά

Μορφοποιήθηκε: Ελληνικά

Μορφοποιήθηκε: Ελληνικά

Μορφοποιήθηκε: Ελληνικά

Μορφοποιήθηκε: Ελληνικά

Μορφοποιήθηκε: Ελληνικά

Μορφοποιήθηκε: Ελληνικά

Μορφοποιήθηκε: Ελληνικά

Μορφοποιήθηκε: Ελληνικά

Μορφοποιήθηκε: Ελληνικά

Μορφοποιήθηκε: Ελληνικά

Μορφοποιήθηκε: Ελληνικά

Μορφοποιήθηκε: Ελληνικά

Μορφοποιήθηκε: Ελληνικά

Μορφοποιήθηκε: Ελληνικά

Μορφοποιήθηκε: Ελληνικά

Μορφοποιήθηκε: Ελληνικά

Μορφοποιήθηκε: Ελληνικά

Μορφοποιήθηκε: Ελληνικά

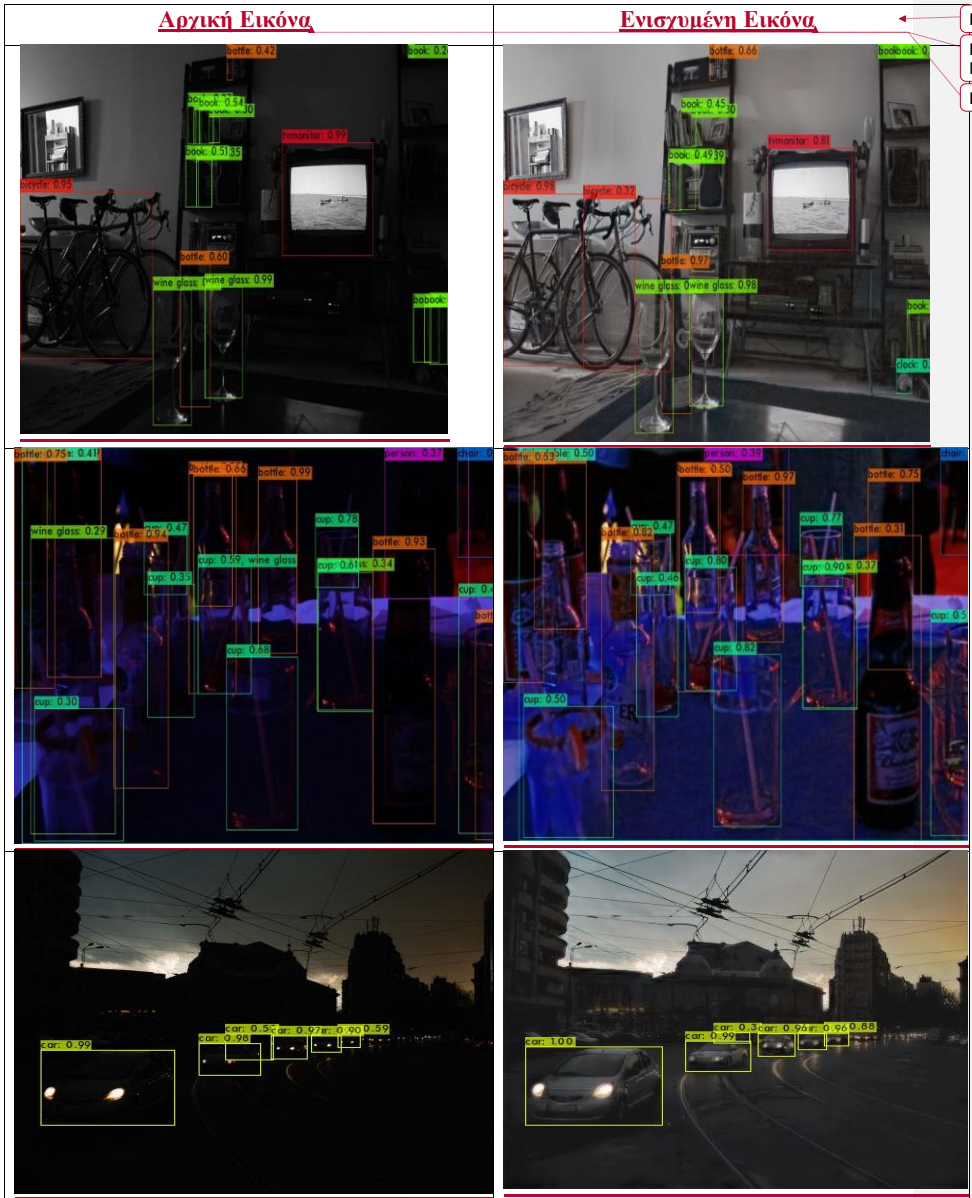
Μορφοποιήθηκε: Ελληνικά

Μορφοποιήθηκε: Ελληνικά

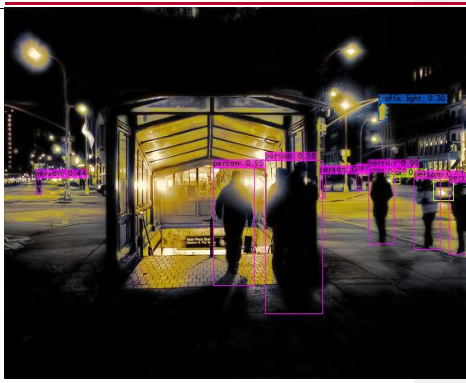
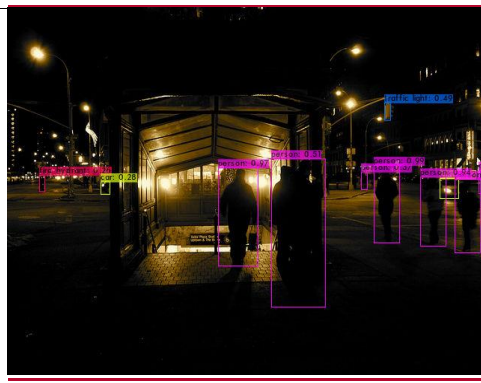
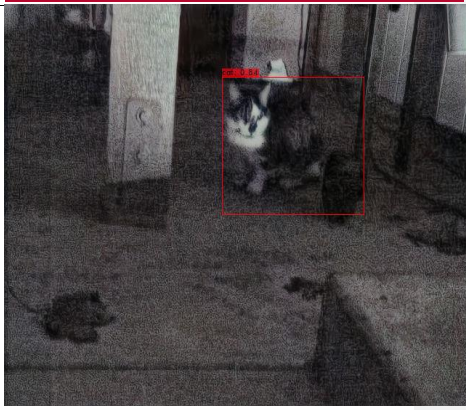
Μορφοποιήθηκε: Ελληνικά

Μορφοποιήθηκε: Ελληνικά

αποτελέσματα. Στις περισσότερες των περιπτώσεων, η μέθοδος KinD ενισχύει την αναγνωρισιμότητα των αντικειμένων με συνέπεια την ενισχυμένη απόδοση του YOLO.



Μορφοποιημένος πίνακας
 Μορφοποιήθηκε: Υπογράμμιση, Αγγλικά (Ηνωμένων Πολιτειών)
 Μορφοποιήθηκε: Υπογράμμιση





Μορφοποιήθηκε: Γραμματοσειρά: Έντονα, Ελληνικά

Μορφοποιήθηκε: Στοιχισμένο στο κέντρο

Μορφοποιήθηκε: Ελληνικά

Μορφοποιήθηκε: Γραμματοσειρά: Πλάγια

Πίνακας 4. Αποτελέσματα του μοντέλου YOLO.

5.4 Ανάλυση Αποτελεσμάτων

Στην ενότητα αυτή περιλαμβάνονται στατιστικές αναλύσεις που προέκυψαν από τα αποτελέσματα του YOLO. Τα αποτελέσματα που παρουσιάζονται στους παρακάτω πίνακες, υπολογίζονται με βάση την πληροφορία που έχει προκύψει από την αρχική εικόνα (groundtruth). Στην πρώτη στήλη αναγράφεται ο κωδικός της κάθε εικόνας από το ExDark.

- **Detection Rate:** Ο συνολικός αριθμός των ανιχνεύσεων του μοντέλου YOLO στην ενισχυμένη εικόνα, σε σύγκριση με τον αριθμό των ανιχνεύσεων του μοντέλου στην αρχική εικόνα.
- **Average Confidence:** Η μέση τιμή της αυτοπεποίθησης του YOLO σε κάθε εικόνα.
- **Missing Detections:** Ο αριθμός των ανιχνεύσεων, οι οποίες δεν καταγράφηκαν από τον YOLO, με βάση την αρχική πληροφορία της κάθε εικόνας.
- **Recall [source]:** Ο λόγος των επιτυχημένων προβλέψεων του μοντέλου προς το σύνολο των επιτυχημένων προβλέψεων και του συνόλου των αποτυχημένων ανιχνεύσεων. Οι τιμές του Recall έχουν εύρος από 0 έως 1, το οποίο μπορεί να μετατραπεί σε ποσοστό %.

$$\text{Recall} = \frac{\text{Correct Predictions}}{\text{Correct Predictions} + \text{Missing Detections}}$$

- **Precision [source]:** Ο λόγος των επιτυχημένων προβλέψεων του μοντέλου προς το σύνολο των επιτυχημένων προβλέψεων συν τις λανθασμένες προβλέψεις. Μία πρόβλεψη θεωρείται λανθασμένη, όταν αυτό που προβλέπει δεν ανήκει σε κάποια από τις γνωστές κλάσεις. Ο λόγος αυτός έχει επίσης εύρος από 0 έως 1, το οποίο μπορεί να μετατραπεί σε ποσοστό %.

$$\text{Precision} = \frac{\text{Correct Predictions}}{\text{Correct Predictions} + \text{Incorrect Predictions}}$$

Μορφοποιήθηκε: Ελληνικά
Μορφοποιήθηκε: Ελληνικά
Αλλαγή κωδικού πεδίου
Μορφοποιήθηκε: Ελληνικά
Μορφοποιήθηκε: Ελληνικά
Μορφοποιήθηκε: Γραμματοσειρά: 14 στ.
Μορφοποιήθηκε: Γραμματοσειρά: 14 στ.
Μορφοποιήθηκε: Γραμματοσειρά: 14 στ.
Μορφοποιήθηκε: Γραμματοσειρά: 14 στ., Αγγλικά (Ηνωμένων Πολιτειών)
Μορφοποιήθηκε: Γραμματοσειρά: 14 στ.
Μορφοποιήθηκε: Γραμματοσειρά: 14 στ., Αγγλικά (Ηνωμένων Πολιτειών)
Μορφοποιήθηκε: Γραμματοσειρά: 14 στ.
Μορφοποιήθηκε: Γραμματοσειρά: 14 στ., Αγγλικά (Ηνωμένων Πολιτειών)
Μορφοποιήθηκε: Αγγλικά (Ηνωμένων Πολιτειών)
Μορφοποιήθηκε: Γραμματοσειρά: Έντονα
Μορφοποιήθηκε: Ελληνικά
Αλλαγή κωδικού πεδίου
Μορφοποιήθηκε: Ελληνικά
Μορφοποιήθηκε: Ελληνικά
Μορφοποιήθηκε: Στοιχισμένο στο κέντρο
Μορφοποιήθηκε: Αγγλικά (Ηνωμένων Πολιτειών)
Μορφοποιήθηκε: Αγγλικά (Ηνωμένων Πολιτειών)
Μορφοποιήθηκε: Αγγλικά (Ηνωμένων Πολιτειών)
Μορφοποιήθηκε: Γραμματοσειρά: 14 στ., Έντονα, Πλάγια

Initial Images						
Image	N. Groundtruth Detections	Correct Predictions	Incorrect Predictions	Missing Detections	Recall	Precision
01367	10	10	0	0	100%	100%
01611	5	5	0	0	100%	100%
01764	15	13	2	0	100%	86.67%
02446	9	5	0	4	55.55%	100%
02594	8	8	0	0	100%	100%
02756	6	4	0	2	66.67%	100%
03052	3	1	1	1	50%	50%
03215	2	1	0	1	50%	100%
03278	1	1	0	0	100%	100%
06298	5	5	0	0	100%	100%
06365	9	9	0	0	100%	100%
06542	5	4	0	1	80%	100%
-	-	Total: 66	Total: 3	Total: 9	Avg: 83.52%	Avg: 94.72%

ΠΑΡΑΡΤΗΜΑ

Πίνακας 5. Μετρήσεις των αρχικών εικόνων.

KinD - Kindling the Darkness $\alpha = 5.0$						
Image	N. Groundtruth Detections	Correct Predictions	Incorrect Predictions	Missing Detections	Recall	Precision
01367	10	8	0	2	80%	100%
01611	5	5	0	0	100%	100%
01764	15	14	0	1	93.33%	100%
02446	9	5	0	4	55.55%	100%
02594	8	8	0	0	100%	100%
02756	6	4	0	2	66.7%	100%
03052	3	1	1	1	50%	50%
03215	2	1	0	1	50%	100%
03278	1	1	0	0	100%	100%
06298	5	5	0	0	100%	100%
06365	9	9	0	0	100%	100%
06542	5	5	0	0	100%	100%
-	-	Total: 66	Total: 1	Total Misses: 11	Avg: 82.96%	Avg: 95.83%

Πίνακας 6. Μετρήσεις μετά την εφαρμογή της μεθόδου KinD.

Τα αποτελέσματα των παραπάνω πινάκων καταδεικνύουν την σαφώς καλύτερη απόδοση του μοντέλου YOLO στις ενισχυμένες εικόνες έναντι των αρχικών σκοτεινών εικόνων. Αξιοσημείωτα είναι επίσης τα ποσοστά επιτυχίας του YOLO, ύστερα από την εφαρμογή της μεθόδου KinD για την ενίσχυση των σκοτεινών εικόνων.

ΚΕΦΑΛΑΙΟ 6 – ΣΥΜΠΕΡΑΣΜΑΤΑ

Η παρούσα ερευνητική εργασία επικεντρώθηκε στις τεχνικές της μηχανικής μάθησης που εφαρμόζονται στο πεδίο της ΜΟ και πιο συγκεκριμένα στο πεδίο της ΑΑ. Ιδιαίτερη έμφαση δόθηκε στη θεωρία των ΤΝΔ και συγκεκριμένα στα ΣΝΔ, που είναι δίκτυα βαθιάς μάθησης τα οποία χρησιμοποιούνται στη ΜΟ. Αναλύθηκε εκτενώς το ΣΝΔ της μεθόδου KinD, το οποίο είναι εμπνευσμένο από τη θεωρία Retinex και χρησιμοποιείται για την αποσύνθεση των εικόνων σε στρώματα ανακλαστικότητας και φωτισμού. Με την αποσύνθεση αυτή πετυχαίνουμε αρχικά την εκπαίδευση του δικτύου σε εικόνες χαμηλού φωτισμού και δευτερευόντως το ΣΝΔ, μέσω ενός δικτύου επαναφοράς, ρυθμίζει τη φωτεινότητα και εξάγει μία ενισχυμένη εικόνα. Διαπιστώσαμε ότι η μέθοδος KinD ενισχύει την εικόνα χαμηλού φωτισμού χωρίς το ρίσκο της παραμόρφωσης των χαρακτηριστικών της. Ωστόσο υπάρχει κίνδυνος υποβάθμισης της εικόνας, λόγω του αλγορίθμου αποθορυβοποίησης που περιλαμβάνει η μέθος KinD.

Στη συνέχεια αναλύσαμε την αρχιτεκτονική του μοντέλου YOLO, το οποίο εφαρμόζεται με ιδιαίτερη επιτυχία και ταχύτητα σε εφαρμογές αναγνώρισης αντικειμένων σε εικόνες και βίντεο. Αναλύσαμε την απόδοσή του σε εικόνες χαμηλού φωτισμού χωρίς κάποια μέθοδο ενίσχυσης και τη συγκρίναμε με την απόδοσή του σε εικόνες χαμηλού φωτισμού, οι οποίες ενισχύθηκαν αρχικά με τη μέθοδο KinD. Στα αποτελέσματα παρατηρήσαμε ότι οι αλγόριθμοι ΑΑ και συγκεκριμένα ο YOLO, αναγνώρισαν περισσότερα αντικείμενα στην ενισχυμένη εικόνα, σε σχέση με την αρχική σκοτεινή εικόνα. Επίσης παρατηρήθηκε και αύξηση της αυτοπεποίθησης του μοντέλου, σχετικά με την ταξινόμηση του αντικειμένου σε κάποια γνωστή κλάση.

Στην παρούσα εργασία πραγματοποιήθηκε σε πρώτο στάδιο η ενίσχυση των όλων των εικόνων χαμηλού φωτισμού μέσω της μεθόδου Kind και σε δεύτερο στάδιο η εφαρμογή του μοντέλου YOLO σε όλες τις ενισχυμένες εικόνες. Αυτή η μέθοδος υλοποίησης που χρησιμοποιήθηκε πραγματοποιείται σε δύο στάδια και απαιτεί αρκετό χρόνο και ενέργεια για την ολοκλήρωση της διαδικασίας. Προτείνεται σε επόμενες ερευνητικές εργασίες, η ανάπτυξη ενός ΤΝΔ, το οποίο θα πραγματοποιεί ενίσχυση και αναγνώριση ταυτόχρονα σε κάθε εικόνα. Δηλαδή ενίσχυση της κάθε εικόνας και ταυτόχρονη αναγνώριση και ταξινόμηση των αντικειμένων που βρίσκονται μέσα σε αυτή. Η αρχιτεκτονική αυτή θα επιταχύνει το χρόνο υλοποίησης και θα εξοικονομήσει πόρους και ενέργεια από το υπολογιστικό σύστημα.

[Υλικό που είναι ογκώδες, παραδείγματος χάριν αποτελέσματα ερωτηματολογίων, σχήματα, πίνακες κ.ά. που εμποδίζουν τον αναγνώστη να κατανοήσει τη συνέχεια του κειμένου, μπορεί να τοποθετηθεί σε παραρτήματα. Η αρίθμηση των παραρτημάτων γίνεται με κεφαλαία ελληνικά γράμματα Α, Β, Γ, ... ενώ σχήματα, σχέδια, πίνακες κ.λπ., που περιλαμβάνονται πρέπει να ονομάζονται Σχήμα Α1, Πίνακας Β2, κ.λπ.]

ΒΙΒΛΙΟΓΡΑΦΙΑ

- ✚ Σφάλμα! Η αναφορά της υπερ-σύνδεσης δεν είναι έγκυρη.
- ~~2-1.~~ https://el.wikipedia.org/wiki/Μηχανική_μάθηση
 - ~~3-2.~~ A. Burkov : The Hundred – Page Machine Learning Book
 - ~~4-3.~~ O. Theobald : Machine Learning For Absolute Beginners, 2nd Ed.
 - ~~5-4.~~ S. Gollapudi : Learn Computer Vision Using OpenCV
 - ~~6-5.~~ M. Elgendy : Deep Learning for Vision Systems
 - ~~7-6.~~ V K. Avvadevara, Y. Reddy : Modern Computer Vision with Pytorch
 - ~~8-7.~~ C. Chen, Q. Chen, J. Xu, V. Koltun : Learning to See in the Dark
 - ~~9-8.~~ Y. Zhang, J. Zhang, X. Guo : Kindling the Darkness : A Practical Low – Light Image Enhancer
 - ~~10-9.~~ C. Li, Chunle Guo, L. Han, J. Jiang, M. M. Cheng, J. Gu, C. Change Lov : Low – Light Image and Video Enhancement Using Deep Learning : A Survey
 - ~~11-10.~~ Edwin H. Land, John J. McCann : Lightness and Retinex Theory
 - ~~12-11.~~ M. Abdullah Al Wadud, Md. Hasanul Kabir, M. Ali Akber Dewan, O. Chae : A Dynamic Histogram Equalization for Image Contrast Enhancement
 - ~~13-12.~~ C. Wei, W. Wang, W. Yang, J. Liu : Deep Retinex Decomposition for Low – Light Enhancement
 - ~~14-13.~~ https://en.wikipedia.org/wiki/K-means_clustering
 - ~~15-14.~~ W. S. McCulloch, W. Pitts : A Logical Calculus of the Ideas Immanent in Nervous Activity
 - ~~16-15.~~ D. Rumelhart, G. Hinton, R. Williams : Learning Representations by Back – Propagating Errors
 - ~~17-16.~~ A. Krizhevsky, I. Sutskever, G. Hinton : ImageNet Classification with Deep Convolutional Neural Networks
 - ~~17.~~ I. Goodfellow, J. Abadie, M. Mirza, B. Xu, D. Warde – Farley, S. Ozair, A. Courville, Y. Benio : Generative Adversarial Nets
 - ~~18.~~ Z. Wang, A. C. Bovik, H.R. Sheikh, E. P. Simoncelli : Image Quality Assessment : From Error Visibility to Structural Similarity
 - ~~19.~~ X. Guo, Y. Li, H. Ling : Lime: Low-Light Image Enhancement via Illumination Map Estimation
 - ~~20.~~ W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.Y. Fu, A. Berg : SSD: Single Shot Detector
 - ~~20.~~ S. Wang, J. Zheng, H. Hu, B. Li : Naturalness Preserved Enhancement Algorithm for Non-Uniform Illumination Images

Μορφοποιήθηκε: Αγγλικά (Ηνωμένων Πολιτειών)

Μορφοποιήθηκε: Εσοχή: Αριστερά: 1,38 εκ., Χωρίς κουκκίδες ή αρίθμηση

Αλλαγή κωδικού πεδίου

22. [A. Mittal, R. Soundararajan, A. Bovik : Making a Completely Blind Image Quality Analyzer](#)
23. [Z. Ying, L. Ge, W. Gao : A Bio-Inspired Multi-Exposure Fusion Framework for Low-Light Image Enhancement](#)
24. [X. Fu, D. Zeng, Y. Huang, X. Zhang, X. Ding : A Weighted Variational Model for Simultaneous Reflectance and Illumination Estimation](#)
25. [Z. Ying, L. Ge, Y. Ren, R. Wang, W. Wang : A New Low-Light Image Enhancement Algorithm Using Camera Response Model](#)
26. [X. Dong, Y. Pang, J. Wen : Fast Efficient Algorithm for Enhancement of Low Lighting Video](#)
27. [X. Fu, D. Zeng, H. Yue, Y. Liao, X. Ding, J. Paislev : A Fusion-Based Enhancing Method for Weakly Illuminated Images](#)
28. [M. Li, J. Liu, W. Yang, X. Sun, Z. Guo : Structure-Revealing Low-Light Image Enhancement via Robust Retinex Model](#)
29. [W.Wang, W. Chen, W. Yang, J. Liu : Gladnet: Low-Light Enhancement Network with Global Awareness](#)
30. [D. J. Jobson, Z. Rahman, G. A. Woodell : A Multiscale Retinex for Bridging the Gap Between Color Images and the Human Observation of Scenes](#)
31. [R. Girshick, J. Donahue, T. Darrell, J. Malik : Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation Tech Report](#)
32. [K. Simonvan, A. Zisserman : Very Deep Convolutional Networks for Large – Scale Image Recognition](#)
33. [J. Redmon, S. Divvala, R. Girshick, A. Farhadi : You Only Look Once: Unified, Real – Time Object Detection](#)
34. [M. Everingham, S.M. Ali Eslami, L. Van Gool, C. Williams, J. Winn, A. Zisserman: The Pascal Visual Object Classes Challenge: A Retrospective](#)
35. [Y. P. Loh, C. S. Chan : Getting to Know Low – Light Images with The Exclusively Dark Dataset](#)
36. [https://el.wikipedia.org/wiki/%CE%9C%CE%B7%CF%87%CE%B1%CE%BD%CE%B9%CE%BA%CE%AE %CE%BC%CE%AC%CE%B8%CE%B7%CF%83%CE%B7](#)

30.–

[Οπισθόφυλλο. Κενή σελίδα]