ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ Μ/Υ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
ΣΧΟΛΗ ΝΑΥΤΙΛΙΑΣ ΚΑΙ ΒΙΟΜΗΧΑΝΙΑΣ
ΤΜΗΜΑΤΟΣ ΒΙΟΜΗΧΑΝΙΚΗΣ ΔΙΟΙΚΗΣΗΣ & ΤΕΧΝΟΛΟΓΙΑΣ
ΔΙΑΠΑΝΕΠΙΣΤΗΜΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ
ΣΠΟΥΔΩΝ«ΤΕΧΝΟ-ΟΙΚΟΝΟΜΙΚΑ ΣΥΣΤΗΜΑΤΑ»

ΔΙΕΠΙΣΤΗΜΟΝΙΚΟ – ΔΙΑΠΑΝΕΠΙΣΤΗΜΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«ΤΕΧΝΟ-ΟΙΚΟΝΟΜΙΚΑ ΣΥΣΤΗΜΑΤΑ»

Αξιολόγηση Τεχνικών Ανίχνευσης και Αναγνώρισης Προσώπων με Τεχνητή Νοημοσύνη για Συστήματα

ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Ιωάννης Δ. Κατσιγιάννης

Επιβλέπων : Ιωάννης Αναγνωστόπουλος
Καθηγητής, Πανεπιστήμιο Θεσσαλίας

Αθήνα, Οκτώβριος 2025

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ Μ/Υ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
ΣΧΟΛΗ ΝΑΥΤΙΛΙΑΣ ΚΑΙ ΒΙΟΜΗΧΑΝΙΑΣ
ΤΜΗΜΑΤΟΣ ΒΙΟΜΗΧΑΝΙΚΗΣ ΔΙΟΙΚΗΣΗΣ & ΤΕΧΝΟΛΟΓΙΑΣ
ΔΙΑΠΑΝΕΠΙΣΤΗΜΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ
ΣΠΟΥΔΩΝ «ΤΕΧΝΟ-ΟΙΚΟΝΟΜΙΚΑ ΣΥΣΤΗΜΑΤΑ»

Αξιολόγηση Τεχνικών Ανίχνευσης και Αναγνώρισης Προσώπων με Τεχνητή Νοημοσύνη για Συστήματα Ασφάλειας

ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Ιωάννης Δ. Κατσιγιάννης

Επιβλέπων : Ιωάννης Αναγνωστόπουλος
Καθηγητής, Πανεπιστήμιο Θεσσαλίας

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 21η Οκτωβρίου 2025.

.....	.....	.....
Ιωάννης Αναγνωστόπουλος	Μάριος Κόνιαρης	Γεράσιμος Ραζής
Καθηγητής, Παν. Θεσσαλιας	Μέλος ΕΔΙΠ, ΕΜΠ	Ακαδημαϊκός Υπότροφος , Παν. Θεσσαλίας

Αθήνα, Οκτώβριος 2025

.....
Ιωάννης Δ. Κατσιγιάννης
Κάτοχος Μεταπτυχιακού Προγράμματος «Τεχνο-οικονομικά Συστήματα» της Σχολής Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών Ε.Μ.Π

Περίληψη

Η παρούσα διπλωματική εργασία εστιάζει στη μελέτη και σύγκριση ευρέως χρησιμοποιούμενων τεχνολογιών υπολογιστικής όρασης, όπως το YOLO (You Only Look Once), καθώς και σε τεχνικές αναγνώρισης προσώπου. Στόχος είναι να διερευνηθεί η ακρίβεια, η ταχύτητα και η αξιοπιστία αυτών των τεχνολογιών σε περιβάλλοντα επιτήρησης. Για τις ανάγκες της έρευνας, πραγματοποιήθηκαν πειραματικές αξιολογήσεις με τη χρήση διαφορετικών συνόλων δεδομένων, τα οποία προέρχονται από ποικίλες πηγές: ένα κλειστό κύκλωμα παρακολούθησης, αποθηκευμένο βίντεο σε σκληρό δίσκο υπολογιστή, καθώς και από ταυτόχρονη χρήση καμερών τόσο από υπολογιστή όσο και από κινητό τηλέφωνο. Στο πλαίσιο αυτό, εξετάζεται η καταλληλόλητα κάθε τεχνικής σε διαφορετικά σενάρια, όπως η παρακολούθηση πλήθους και η ταυτοποίηση ατόμων. Η μελέτη αυτή συνεισφέρει ουσιαστικά στην κατανόηση των πλεονεκτημάτων και των περιορισμών κάθε μεθόδου, παρέχοντας τεκμηριωμένη καθοδήγηση για την επιλογή του κατάλληλου συστήματος σε εφαρμογές που σχετίζονται με την ασφάλεια.

Λέξεις-Κλειδιά: AI, ML, Deep Learning, CNN, RNN, Face_recognition (dlib),YOLO, Keras, ArcFace ONNX, CelebA, WIDER_FACE, MTCNN, OpenCV

Abstract

This thesis explores and compares widely adopted computer vision technologies such as YOLO (You Only Look Once)—alongside various face recognition techniques. Its primary goal is to evaluate their accuracy, speed, and reliability within surveillance contexts. To achieve this, experimental assessments are conducted using datasets from multiple sources, including footage from a closed surveillance system, a video stored locally on a computer's hard drive, and real-time video streams captured simultaneously from a computer and a mobile phone camera. The analysis focuses on how well each technique performs across different scenarios, such as crowd monitoring and individual identification. By highlighting the strengths and limitations of each method, this research offers well-founded insights and guidance for selecting the most suitable system for security-related applications.

Keywords: AI, ML, Deep Learning, CNN, RNN, Face_recognition (dlib), YOLO, Keras, ArcFace ONNX, CelebA, WIDER_FACE, MTCNN, OpenCV

Ευχαριστίες

Θα ήθελα να εκφράσω τις πιο θερμές μου ευχαριστίες στον επιβλέποντα καθηγητή μου, κ. Γεράσιμο Ραζή, ο οποίος μου έδωσε τη δυνατότητα να εξερευνήσω ένα διαφορετικό επιστημονικό πεδίο, αυτό της επιστημονικής όρασης, καθώς και για την πολύτιμη καθοδήγηση, την αμέριστη υποστήριξη και τις εύστοχες παρατηρήσεις του καθ' όλη τη διάρκεια της εκπόνησης της διπλωματικής μου εργασίας. Η συμβολή του υπήρξε καθοριστική τόσο σε επιστημονικό όσο και σε προσωπικό επίπεδο.

Ευχαριστώ θερμά τον Αξιωματικό Επιχειρήσεων, κ. Στέλιο Μαργέτη, για τη διευκόλυνση και την ουσιαστική βοήθεια στην επίλυση προβλήματος που ανέκυψε κατά την εκπόνηση της διπλωματικής μου εργασίας, με τη χαρακτηριστική παρότρυνση: «Κάθε φορά θα αναρωτιέσαι τι θέλεις να ψάξει να βρει το script σου».

Ιδιαίτερη ευγνωμοσύνη οφείλω στη σύζυγό μου, Στεύη Μαφούνη, για την αγάπη, την υπομονή και την αδιάκοπη στήριξή της σε κάθε βήμα αυτής της πορείας. Χωρίς τη δική της ενθάρρυνση και πίστη στις δυνατότητές μου, η ολοκλήρωση αυτής της προσπάθειας θα ήταν πολύ πιο δύσκολη.

Θερμά ευχαριστώ την οικογένειά μου για τη συνεχή συμπαράσταση, την εμπιστοσύνη και το ηθικό στήριγμα που μου προσφέρουν όλα αυτά τα χρόνια. Η παρουσία τους αποτελεί για μένα σταθερό σημείο αναφοράς και πηγή δύναμης. Τέλος, αφιερώνω τη διπλωματική μου εργασία και συνολικά την ακαδημαϊκή μου πορεία μέχρι σήμερα στη γιαγιά μου. **Η αγάπη, οι αξίες και η έμπνευση που μου προσέφερε αποτελούν θεμέλιο για κάθε μου βήμα.**

Περιεχόμενα

1. Εισαγωγή **13**

Κεφάλαιο 2: Θεωρητικό Υπόβαθρο **14**

2.1Εισαγωγή στην Υπολογιστική Όραση 14

2.2 Τεχνικές Επεξεργασίας Εικόνας 15

2.2.1 Εικόνα διαβαθμίσεων του γκρι (Grayscale Image): 16

2.2.2 Έγχρωμη Εικόνα (Colored Image) 17

2.2.3 Σύγκριση μεταξύ εικόνων διαβαθμίσεων του γκρι και έγχρωμων εικόνων 19

2.3 Φιλτράρισμα Εικόνων (Image Filtering) 20

2.3.1 Χωρικό Φιλτράρισμα (Spatial Filtering) 21

2.3.2 Προ επεξεργασία με Φίλτρα Εικόνας 22

2.4 Παραγωγή Θορύβου με Κατανομή Gauss (Gaussian Distribution) 23

2.4.1 Εφαρμογή Θορύβου στην Εικόνα 24

2.5 Περιστροφή Εικόνας (Image Rotation) 25

2.6 Μέθοδοι Προ επεξεργασίας 26

2.7 Μηχανική Μάθηση 27

2.8 Αρχιτεκτονική και Αρχές Νευρωνικών Δικτύων 27

2.8.1 Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks - CNNs) 29

2.8.2 Ανίχνευση Αντικειμένων με χρήση CNNs 30

2.8.3 Κατηγορίες Μεθόδων Ανίχνευσης Αντικειμένων με CNNs 31

2.9 Βιβλιοθήκες Υπολογιστικής Όρασης 38

2.9.1 Face_recognition (dlib) ως Βιβλιοθήκη Υπολογιστικής Όρασης 39

2.9.2 YOLOv8 – Μια σύγχρονη προσέγγιση στην ανίχνευση Προσώπων 40

2.10 Βιβλιοθήκες Αναγνώρισης προσώπων 42

2.10.1 Βιβλιοθήκη face_recognition (dlib) 43

2.10.2 ArcFace (Onnxruntime) 43

Κεφάλαιο 3: Σχετικά Υφιστάμενα Εργαλεία και Παρόμοιες Εργασίες **44**

3.1 Παρόμοιες Επιστημονικές Εργασίες 44

3.2 Συγκριτική Ανάλυση Υλοποιήσεων 47

3.3 Σημεία Υπεροχής της Παρούσας Εργασίας 48

3.4 Συμπερασματικά 49

Κεφάλαιο 4:Υλοποίηση Εφαρμογής (Τεχνικά + Τεκμηρίωση Κώδικα) **50**

4.1 Ανάλυση εφαρμογών 50

Σύστημα Ανίχνευσης & Αναγνώρισης Προσώπων με Βιβλιοθήκη face_recognition 50

Σύστημα Ανίχνευσης με YOLO & Αναγνώρισης Προσώπων με ArcFace ONNX 57

4.2 Εγκατάσταση εφαρμογών 64

1. face_detection9C.py 64
2. face detection15D0.py 65

Κεφάλαιο 5: Αξιολόγηση μεθόδων **66**

5.1 Αξιολόγηση του Συστήματος Ανίχνευσης και Αναγνώρισης Προσώπων με CelebA Dataset 67

5.1.1 Αξιολόγηση του Συστήματος Ανίχνευσης Προσώπων με CelebA Dataset 67

5.1.2 Αξιολόγηση Συστήματος Αναγνώρισης Προσώπων με CelebA 69

5.2Συνδυαστική Αξιολόγηση των δύο καλύτερων Συστημάτων με CelebA 70

Κεφάλαιο 6: Συμπεράσματα και Μελλοντικές Επεκτάσεις **72**

6.1 Συγκριτική Ανάλυση: face_detection9C.py vs face_detection15D0.py 72

6.2 Συμπεράσματα 73
6.3 Μελλοντικές Επεκτάσεις 74
Βιβλιογραφία **76**

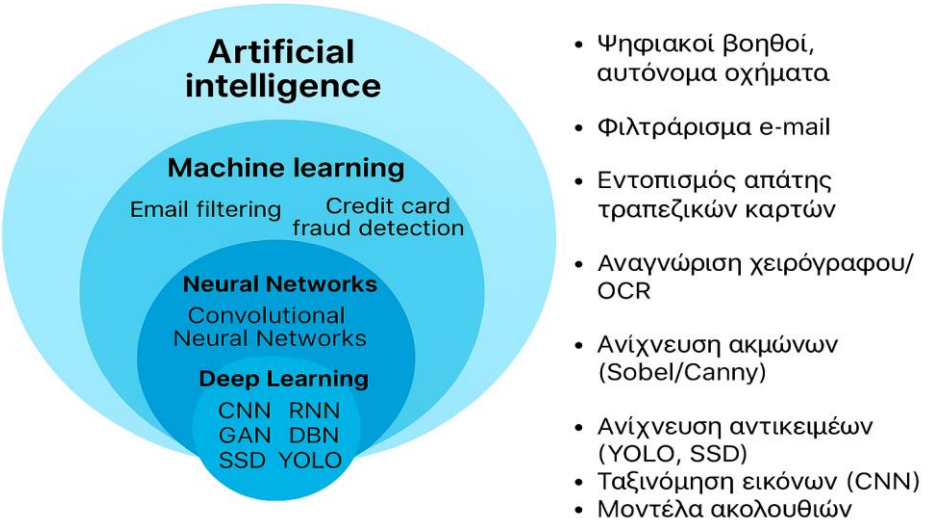
1. Εισαγωγή

Η παρούσα εργασία ασχολείται με τον σχεδιασμό και την ανάπτυξη ενός συστήματος αναγνώρισης προσώπων, το οποίο αξιοποιεί σύγχρονες τεχνικές μηχανικής μάθησης και τεχνητής νοημοσύνης. Στόχος είναι η δημιουργία μιας ολοκληρωμένης λύσης η οποία θα μπορεί να λειτουργεί σε πραγματικό χρόνο, με υψηλή ακρίβεια και αξιοπιστία, ώστε να συμβάλλει στην ενίσχυση της ασφάλειας σε ιδιωτικούς και δημόσιους χώρους. Η χρήση τέτοιων συστημάτων είναι κρίσιμη, καθώς σχετίζεται με την αποτροπή μη εξουσιοδοτημένων προσβάσεων και την έγκαιρη αντιμετώπιση καταστάσεων που μπορεί να είναι επικίνδυνες. Για τον λόγο αυτό, δίνεται ιδιαίτερη έμφαση τόσο στη σωστή τεχνική υλοποίηση όσο και στη φιλικότητα ως προς τον χρήστη. Το σύστημα έχει σχεδιαστεί έτσι ώστε να είναι εύχρηστο και κατανοητό ακόμη και από άτομα χωρίς εξειδικευμένες γνώσεις, προσφέροντας παράλληλα δυνατότητες για συνεχή ενημέρωση και επέκταση της βάσης δεδομένων με νέα πρόσωπα. Η ανθρωποκεντρική του σχεδίαση και η ευελιξία προσαρμογής σε διαφορετικά περιβάλλοντα εφαρμογής το καθιστούν ένα εργαλείο που μπορεί να αξιοποιηθεί σε βάθος χρόνου για σκοπούς προστασίας, ελέγχου πρόσβασης και διαχείρισης απειλών. Σε πρακτικό επίπεδο, η εργασία αυτή συμβάλλει με την ανάπτυξη ενός λειτουργικού συστήματος που συνδυάζει τεχνικές όπως τα συνελικτικά νευρωνικά δίκτυα (CNN), το YOLOv8 και τη βιβλιοθήκη Face Detection(dlib), ενώ η αποτελεσματικότητά του ελέγχεται με το σύνολο δεδομένων (Celeba,WIDER_FACE). Παράλληλα, περιλαμβάνει τεχνική τεκμηρίωση, συγκριτική ανάλυση μεθόδων και προτάσεις για μελλοντική αξιοποίηση και βελτίωση. Η εργασία είναι οργανωμένη σε έξι ενότητες. Στη δεύτερη ενότητα παρουσιάζεται το θεωρητικό υπόβαθρο, με αναφορά στις βασικές έννοιες της υπολογιστικής όρασης, τις τεχνικές επεξεργασίας εικόνας και τις αρχές της μηχανικής μάθησης. Στην τρίτη ενότητα εξετάζονται παρόμοιες επιστημονικές εργασίες και σχετικά εργαλεία, προκειμένου να αναδειχθούν τα σημεία διαφοροποίησης και υπεροχής της παρούσας μελέτης. Στην τέταρτη ενότητα περιγράφεται αναλυτικά η υλοποίηση του συστήματος και παρατίθεται η απαραίτητη τεχνική τεκμηρίωση. Ακολουθεί η πέμπτη ενότητα, όπου παρουσιάζεται η αξιολόγηση του συστήματος με το σύνολο δεδομένων (Celeba, WIDER_FACE) και τα αποτελέσματα. Τέλος, η έκτη ενότητα περιλαμβάνει τα συμπεράσματα της μελέτης, συνοψίζει τις συνεισφορές της εργασίας και προτείνει κατευθύνσεις για μελλοντικές επεκτάσεις.

Κεφάλαιο 2: Θεωρητικό Υπόβαθρο

2.1Εισαγωγή στην Υπολογιστική Όραση

Η Υπολογιστική Όραση (Computer Vision) αποτελεί έναν ραγδαία αναπτυσσόμενο κλάδο της Πληροφορικής, ο οποίος εστιάζει στο πώς οι υπολογιστές μπορούν να «βλέπουν» και να κατανοούν πληροφορίες από ψηφιακές εικόνες ή βίντεο. Ουσιαστικά, επιχειρεί να προσομοιώσει τη λειτουργία της ανθρώπινης όρασης, επιτρέποντας στα υπολογιστικά συστήματα να αναγνωρίζουν και να ερμηνεύουν οπτικά δεδομένα με σκοπό την εξαγωγή νοήματος. Με την υποστήριξη της Τεχνητής Νοημοσύνης (Artificial Intelligence – AI), η Υπολογιστική Όραση έχει βρει εφαρμογή σε πλήθος τομέων, όπως η αναγνώριση αντικειμένων, η ανάλυση σκηνών (π.χ. σε αθλητικούς αγώνες) και η αυτόνομη πλοήγηση οχημάτων. Ο τεράστιος όγκος οπτικών δεδομένων που παράγεται καθημερινά έχει συμβάλει σημαντικά στην ταχεία πρόοδο του πεδίου αυτού. Οι σύγχρονες εφαρμογές της Υπολογιστικής Όρασης στηρίζονται κυρίως σε τεχνικές Μηχανικής Μάθησης (Machine Learning), με κυρίαρχο παράδειγμα τα Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks – CNN).



Τα νευρωνικά δίκτυα είναι αλγόριθμοι εμπνευσμένοι από τον τρόπο λειτουργίας του ανθρώπινου εγκεφάλου, καθώς επεξεργάζονται πληροφορίες μέσω διαδοχικών επιπέδων (layers) τεχνητών «νευρώνων». Κύριος στόχος της Υπολογιστικής Όρασης είναι η ερμηνεία των εικόνων με τη χρήση υπολογιστικών αλγορίθμων, με σκοπό την αναγνώριση και κατηγοριοποίηση αντικειμένων, τη μέτρηση γεωμετρικών χαρακτηριστικών και τη λήψη αποφάσεων βασισμένων στο περιεχόμενο της εικόνας. Τα νευρωνικά δίκτυα διαδραματίζουν καθοριστικό ρόλο στη διαδικασία αυτή, καθώς επιτρέπουν την αυτόματη λήψη αποφάσεων από το σύστημα, με ελάχιστη ανθρώπινη παρέμβαση. Η ικανότητα των μηχανών να αναγνωρίζουν αντικείμενα, να ανιχνεύουν πρότυπα και να κατανοούν το περιεχόμενο μιας εικόνας με αυτοματοποιημένο τρόπο βρίσκεται στον πυρήνα της Υπολογιστικής Όρασης.

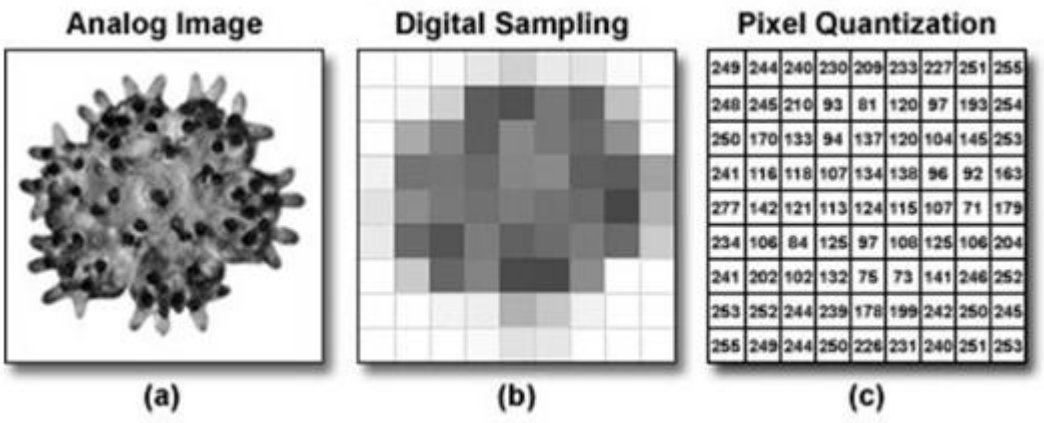
Στο πλαίσιο αυτό, η εικόνα θεωρείται μία από τις βασικές μαθηματικές αναπαραστάσεις των οπτικών δεδομένων. Μια ψηφιακή εικόνα περιγράφεται ως ένας διδιάστατος πίνακας, όπου κάθε στοιχείο αντιστοιχεί σε μία τιμή εικονοστοιχείου (pixel value). Μαθηματικά, μπορεί να παρασταθεί ως συνάρτηση της μορφής:

$$g:R^2 \rightarrow R$$

$$z = g(x,y), \qquad x, y, z \in R$$

Όπου τα x, y είναι οι χωρικές συντεταγμένες κάθε εικονοστοιχείου και η τιμή της συνάρτησης $z = g(x,y)$ αντιπροσωπεύει την ένταση του εικονοστοιχείου. Αξίζει να σημειωθεί ότι οι μηχανές μπορούν να κατανοήσουν μόνο ψηφιακές εικόνες, οι οποίες αποτελούν διακριτές αναπαραστάσεις αναλογικών εικόνων.

Η παρακάτω εικόνα αναπαριστά αυτό που καταλαβαίνει ο υπολογιστής.



Επεξήγηση της Εικόνας:

(a) Αναλογική Εικόνα (Analog Image): Η αρχική εικόνα με συνεχείς τιμές έντασης.

(b) Ψηφιακή Δειγματοληψία (Digital Sampling): Η εικόνα διασπάται σε ένα δισδιάστατο πλέγμα από εικονοστοιχεία (pixels), όπου μετράτε η ένταση σε κάθε κελί.

(c) Κβαντοποίηση Εικονοστοιχείων (Pixel Quantization): Οι τιμές έντασης κάθε pixel στρογγυλοποιούνται και αποθηκεύονται ως ακέραιοι αριθμοί, σχηματίζοντας έναν πίνακα αριθμών.

2.2 Τεχνικές Επεξεργασίας Εικόνας

Σε πολλές εφαρμογές της μηχανικής, όπως η ανάλυση και η αναγνώριση αντικειμένων σε μηχανολογικά σχέδια, η χρήση εικόνων διαβαθμίσεων του γκρι μπορεί να ενισχύσει σημαντικά την αποτελεσματικότητα των αλγορίθμων ανίχνευσης αντικειμένων (object detection). Αυτό οφείλεται στο γεγονός ότι οι εικόνες σε αποχρώσεις του γκρι επιτρέπουν την πιο ακριβή ανάδειξη των βασικών χαρακτηριστικών της εικόνας, όπως τα περιγράμματα και οι γραμμές – στοιχεία ιδιαίτερα κρίσιμα στην ανάλυση τεχνικών σχεδίων. Για τον λόγο αυτό,

είναι συχνά απαραίτητο να προηγείται η εφαρμογή ενός φίλτρου μετατροπής από έγχρωμη σε ασπρόμαυρη εικόνα (color-to-grayscale image filter), πριν ξεκινήσει η ανάλυση. Με αυτόν τον τρόπο, ο αλγόριθμος ανίχνευσης μπορεί να εστιάσει στις ουσιώδεις δομές και γεωμετρικά σχήματα του σχεδίου, χωρίς να επηρεάζεται από το χρώμα, το οποίο συνήθως δεν παίζει σημαντικό ρόλο σε τεχνικά ή μηχανολογικά περιβάλλοντα.

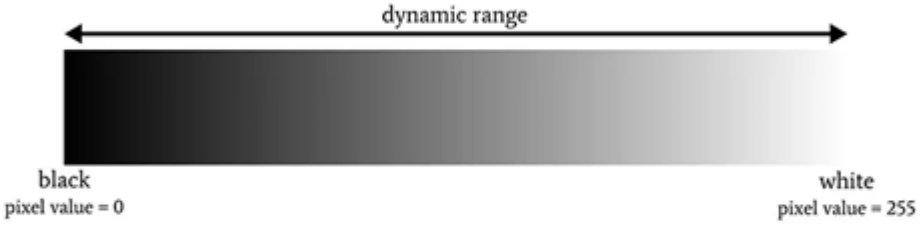
2.2.1 Εικόνα διαβαθμίσεων του γκρι (Grayscale Image):

Μια εικόνα διαβαθμίσεων του γκρι (ή αλλιώς grayscale) είναι ένας τύπος ψηφιακής εικόνας που περιλαμβάνει μόνο αποχρώσεις του γκρι και όχι χρώμα, δηλαδή είναι μονοχρωματική. Σε αυτό το είδος εικόνας, κάθε εικονοστοιχείο (pixel) εκφράζεται με μία μόνο τιμή έντασης φωτεινότητας (intensity value), η οποία δείχνει πόσο φωτεινό ή σκοτεινό είναι το συγκεκριμένο σημείο.

Η τιμή έντασης για κάθε pixel κυμαίνεται συνήθως από 0 έως 255:

- Τιμή **0** σημαίνει απόλυτο μαύρο.
- Τιμή **255** σημαίνει καθαρό λευκό.
- Ενδιάμεσες τιμές αντιστοιχούν σε διαφορετικές αποχρώσεις του γκρι, με τις μικρότερες τιμές να είναι πιο σκούρες και τις μεγαλύτερες πιο ανοιχτές.

Για παράδειγμα, σε μια φωτογραφία διαβαθμίσεων του γκρι από έναν τοίχο με σκιά, τα pixels που βρίσκονται στη σκιά θα έχουν χαμηλές τιμές (πιο κοντά στο μαύρο), ενώ τα pixels που φωτίζονται περισσότερο θα έχουν μεγαλύτερες τιμές (πιο κοντά στο λευκό). Έτσι, δημιουργείται η αίσθηση του βάθους και της φωτεινότητας μόνο με αποχρώσεις του γκρι.



Το παραπάνω διάγραμμα δείχνει το δυναμικό εύρος (Dynamic range) μιας εικόνας διαβαθμίσεων του γκρι:

Στα αριστερά (pixel value = 0) έχουμε το απόλυτο μαύρο.
Στα δεξιά (pixel value = 255) έχουμε το απόλυτο λευκό.
Οι ενδιάμεσες τιμές αντιστοιχούν σε αποχρώσεις του γκρι.

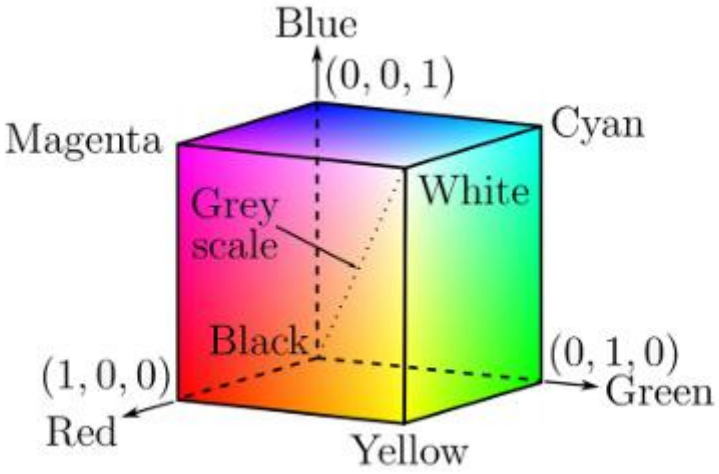
2.2.2 Έγχρωμη Εικόνα (Colored Image)

Σε αντίθεση με την εικόνα διαβαθμίσεων του γκρι, η έγχρωμη εικόνα αποτελείται από πολλά χρωματικά κανάλια, τα οποία συνήθως είναι το Κόκκινο (Red), το Πράσινο (Green) και το Μπλε (Blue), γνωστά ως μοντέλο RGB. Σε μια τέτοια εικόνα, κάθε εικονοστοιχείο (pixel) περιγράφεται από τρεις ξεχωριστές τιμές, μία για κάθε χρωματικό κανάλι (R, G, B).

Η κάθε μία από αυτές τις τιμές εκφράζει την ένταση του αντίστοιχου χρώματος στο συγκεκριμένο pixel, με τιμές που συνήθως κυμαίνονται από 0 έως 255:

- Τιμή **0** σημαίνει καθόλου συμμετοχή του συγκεκριμένου χρώματος.
- Τιμή **255** σημαίνει μέγιστη ένταση για το συγκεκριμένο χρώμα.

Ο συνδυασμός των τριών αυτών τιμών (προσθετική ανάμιξη) δημιουργεί το τελικό χρώμα του pixel. Έτσι, το κάθε pixel μιας έγχρωμης εικόνας μπορεί να πάρει μια μεγάλη ποικιλία αποχρώσεων. Για παράδειγμα, ένα pixel με τιμές R=255, G=0, B=0 εμφανίζεται ως καθαρό κόκκινο, ενώ ένα pixel με τιμές R=0, G=255, B=0 είναι καθαρό πράσινο. Αν όλα τα κανάλια έχουν την ίδια τιμή, όπως R=128, G=128, B=128, τότε το pixel είναι μια απόχρωση του γκρι. Αυτό δείχνει πώς με τη χρήση τριών καναλιών μπορούμε να δημιουργήσουμε οποιοδήποτε χρώμα στην ψηφιακή εικόνα. Γεωμετρικά, το κάθε pixel μιας έγχρωμης εικόνας μπορεί να αναπαρασταθεί ως σημείο σε έναν τρισδιάστατο χώρο, όπου κάθε άξονας αντιστοιχεί σε ένα χρωματικό κανάλι (R, G, B). Ολόκληρος ο χώρος αυτός σχηματίζει έναν κύβο χρωμάτων (color cube), μέσα στον οποίο βρίσκονται όλα τα δυνατά χρώματα που μπορεί να πάρει ένα pixel.



Επεξήγηση σχήματος

Κόκκινο (Red): (1, 0, 0)

Πράσινο (Green): (0, 1, 0)

Μπλε (Blue): (0, 0, 1)

Κυανό (Cyan), Ματζέντα (Magenta), Κίτρινο (Yellow), Άσπρο (White), Μαύρο (Black): Προκύπτουν από διάφορους συνδυασμούς τιμών στα κανάλια RGB.

Αποχρώσεις του γκρι (Grey scale): Βρίσκονται στη διαγώνιο του κύβου (όπου R = G = B), δηλαδή όταν όλα

τα χρώματα έχουν ίδια τιμή, προκύπτει κάποια απόχρωση του γκρι.

Με αυτόν τον τρόπο, κάθε pixel μιας έγχρωμης εικόνας μπορεί να θεωρηθεί ως διάνυσμα τριών διαστάσεων, όπου η θέση του μέσα στον κύβο καθορίζει το τελικό του χρώμα.

Grayscale Operation

Η διαδικασία μετατροπής μιας έγχρωμης εικόνας σε εικόνα διαβαθμίσεων του γκρι αποτελεί μια συνηθισμένη λειτουργία στην επεξεργασία εικόνας, η οποία υλοποιείται ως γραμμικό φίλτρο. Η διαδικασία αυτή στοχεύει στη μείωση της πληροφορίας της εικόνας από τα τρία χρωματικά κανάλια RGB (Κόκκινο, Πράσινο, Μπλε) σε ένα μόνο κανάλι έντασης ή φωτεινότητας (grayscale). Αυτό επιτυγχάνεται μέσω του υπολογισμού ενός σταθμισμένου μέσου όρου των τιμών των καναλιών R, G, και B για κάθε pixel. Ο μαθηματικός τύπος που χρησιμοποιείται είναι ο εξής:

$$f(i, j) = w_1 \cdot R(i,j) + w_2 \cdot G(i,j) + w_3 \cdot B(i,j)$$

όπου:

- f(i,j)f(i, j)f(i,j) είναι η τελική τιμή έντασης του pixel στη θέση (i,j)(i, j)(i,j) στην εικόνα διαβαθμίσεων του γκρι,
- R(i,j),G(i,j),B(i,j)R(i, j), G(i, j), B(i, j)R(i,j),G(i,j),B(i,j) είναι οι τιμές των καναλιών κόκκινου, πράσινου και μπλε, αντίστοιχα,
- w1,w2,w3w_1, w_2, w_3w1,w2,w3 είναι τα βάρη που χρησιμοποιούνται για κάθε κανάλι.

Η πιο διαδεδομένη εκδοχή βασίζεται στη φωτεινότητα (luminance-based weighting) και χρησιμοποιεί τα εξής βάρη:

- w1=0.2989 για το κόκκινο (Red)
- w2=0.5870 για το πράσινο (Green)
- w3=0.1140 για το μπλε (Blue)

Με τον τρόπο αυτό, δίνεται μεγαλύτερη βαρύτητα στο πράσινο κανάλι, καθώς ο ανθρώπινος οπτικός μηχανισμός είναι πιο ευαίσθητος σε αυτό το χρώμα. Αφού υπολογιστεί ο σταθμισμένος μέσος όρος για κάθε pixel, η τιμή που προκύπτει αποτελεί τη φωτεινότητα του pixel στην τελική εικόνα διαβαθμίσεων του γκρι. Η τελική grayscale εικόνα αναπαρίσταται ως ένας διδιάστατος πίνακας τιμών έντασης.

Σύγκριση Εικόνων

Έγχρωμη εικόνα (Colored Image): Αριστερά, εμφανίζεται η αρχική εικόνα με τα τρία κανάλια RGB.

Εικόνα διαβαθμίσεων του γκρι (Grayscale Image): Δεξιά, παρουσιάζεται η ίδια εικόνα μετά τη μετατροπή της, όπου κάθε pixel έχει μόνο μία τιμή έντασης που εκφράζει τη φωτεινότητά του.

2.2.3 Σύγκριση μεταξύ εικόνων διαβαθμίσεων του γκρι και έγχρωμων εικόνων

Οι έγχρωμες εικόνες προσφέρουν μια πιο ρεαλιστική και φυσική απεικόνιση του περιβάλλοντος, αφού το χρώμα είναι σημαντικό στοιχείο της ανθρώπινης αντίληψης. Ωστόσο, στις εφαρμογές της υπολογιστικής όρασης, συχνά προτιμώνται οι εικόνες διαβαθμίσεων του γκρι (grayscale), καθώς παρουσιάζουν αρκετά πλεονεκτήματα έναντι των έγχρωμων εικόνων.

Βασικά Πλεονεκτήματα Εικόνων Διαβαθμίσεων του Γκρι

1. **Απλότητα Επεξεργασίας**
   Οι εικόνες διαβαθμίσεων του γκρι περιέχουν μόνο ένα κανάλι πληροφορίας, δηλαδή μία τιμή έντασης για κάθε pixel. Αυτό τις καθιστά πιο εύκολες στην ανάλυση και στην ανίχνευση χαρακτηριστικών, όπως ακμές, σχήματα και υφές, σε σύγκριση με τις έγχρωμες που έχουν τρία κανάλια (R, G, B).

2. **Μειωμένη Υπολογιστική Πολυπλοκότητα**
   Επειδή οι grayscale εικόνες έχουν λιγότερα δεδομένα ανά pixel, απαιτούν λιγότερη μνήμη και επεξεργαστική ισχύ. Έτσι, η ανάλυση και επεξεργασία τους γίνεται ταχύτερα και με μικρότερο κόστος για το σύστημα.

3. **Λιγότερος Θόρυβος**
   Στις έγχρωμες εικόνες, κάθε κανάλι μπορεί να εισάγει επιπλέον θόρυβο και διακυμάνσεις στην ένταση, κάτι που μπορεί να επηρεάσει αρνητικά τον εντοπισμό αντικειμένων. Οι εικόνες διαβαθμίσεων του γκρι είναι συνήθως πιο «καθαρές» και ομοιογενείς.

4. **Μεγαλύτερη Ανθεκτικότητα σε Φωτισμό**

   Οι grayscale εικόνες είναι συχνά πιο ανθεκτικές σε αλλαγές φωτισμού ή έκθεσης. Σε αντίθεση, στις έγχρωμες εικόνες, το χρώμα μπορεί να αλλοιωθεί σημαντικά λόγω αλλαγών στο φως.

5. **Εστίαση σε Υφή και Σχήμα**

   Σε πολλές εφαρμογές, όπως η αναγνώριση ακμών, μορφών ή μοτίβων, το χρώμα δεν είναι τόσο σημαντικό. Οι εικόνες διαβαθμίσεων του γκρι δίνουν πιο καθαρή πληροφορία για υφή και σχήμα, γι' αυτό και είναι ιδανικές για αυτού του τύπου αναλύσεις.

Για παράδειγμα, σε ένα σύστημα αυτόματης ανίχνευσης ακμών για αναγνώριση πινακίδων κυκλοφορίας, η ανάλυση γίνεται συνήθως πάνω σε εικόνες διαβαθμίσεων του γκρι, διότι ο αλγόριθμος χρειάζεται να εντοπίσει τα περιγράμματα των αντικειμένων, ανεξάρτητα από το χρώμα τους.

2.3 Φιλτράρισμα Εικόνων (Image Filtering)

Το φιλτράρισμα εικόνων είναι μια βασική διαδικασία προ-επεξεργασίας στην υπολογιστική όραση. Στόχος του είναι να μετατρέψει μια ψηφιακή εικόνα σε μια πιο απλή ή βελτιωμένη μορφή, ώστε να διευκολυνθεί η ανάλυση από αλγορίθμους.

Το φιλτράρισμα εικόνων εφαρμόζεται για διάφορους σκοπούς, όπως:

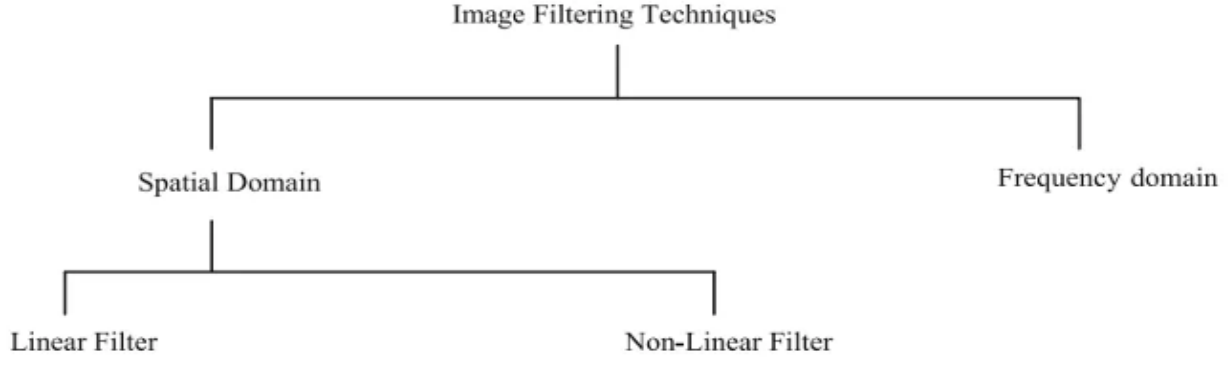
- **Αφαίρεση θορύβου**

  Ελαχιστοποιεί ανεπιθύμητες διακυμάνσεις ή παραμορφώσεις που υπάρχουν στην εικόνα, κάνοντας το τελικό αποτέλεσμα πιο «καθαρό».
- **Αφαίρεση ανεπιθύμητων χαρακτηριστικών**

  Απομακρύνει στοιχεία της εικόνας που δεν ενδιαφέρουν την ανάλυση, όπως μικρά σημεία ή σκιές.
- **Ενίσχυση συγκεκριμένων χαρακτηριστικών**

  Βελτιώνει σημαντικά στοιχεία της εικόνας, όπως τις ακμές ή τις υφές, ώστε να εντοπίζονται ευκολότερα από αλγορίθμους ανίχνευσης.

Με αυτό τον τρόπο, το φιλτράρισμα προετοιμάζει την εικόνα, καθιστώντας την καταλληλότερη για επόμενα στάδια επεξεργασίας και ανάλυσης.



2.3.1 Χωρικό Φιλτράρισμα (Spatial Filtering)

Το χωρικό φιλτράρισμα αποτελεί μια θεμελιώδη τεχνική στην επεξεργασία ψηφιακών εικόνων, η οποία βασίζεται στην εφαρμογή φίλτρων απευθείας στις τιμές των εικονοστοιχείων (pixels), λαμβάνοντας υπόψη τη χωρική τους διάταξη μέσα στην εικόνα. Η διαδικασία αυτή πραγματοποιείται συνήθως μέσω του μαθηματικού τελεστή της συνέλιξης (convolution), όπου ένας πυρήνας (kernel) «σαρώνει» το πλέγμα των pixels και υπολογίζει τη νέα τιμή κάθε pixel ως σταθμισμένο άθροισμα των τιμών των γειτονικών του εικονοστοιχείων. Με αυτόν τον τρόπο, το φιλτραρισμένο αποτέλεσμα ενσωματώνει την τοπική πληροφορία της εικόνας, όπως αυτή καθορίζεται από τα χαρακτηριστικά του πυρήνα. Ανάλογα με τη φύση και τη μορφή του πυρήνα, το χωρικό φιλτράρισμα μπορεί να εξομαλύνει θόρυβο, να εντοπίσει ακμές ή να αναδείξει συγκεκριμένα μοτίβα και δομές στην εικόνα.

Γραμμικά Φίλτρα (Linear Filters)

Τα γραμμικά φίλτρα εκτελούν μετασχηματισμούς πάνω στις τιμές έντασης των pixels, χρησιμοποιώντας ένα σύνολο βαρών (weights) που ορίζονται από τον πυρήνα (kernel) του φίλτρου. Η τελική τιμή κάθε pixel στη νέα εικόνα προκύπτει από τον σταθμισμένο συνδυασμό των τιμών των γειτονικών pixels. Ο υπολογισμός αυτός γίνεται με τη διαδικασία της συνέλιξης. Έστω ότι $f(i, j) \in Z^2$ είναι μια εικόνα και το φίλτρο ή πυρήνας (kernel) είναι το h, το οποίο περιλαμβάνει τα βάρη $h(k,l)$. Τότε, η έξοδος της συνέλιξης (convolutional output image) δίνεται ως εξής:

$$g = f \times h$$

$$g(i, j) = \sum f(i-k, j-l) \cdot h(k,l) \; N = \sum f(k,l) \cdot h(i-k, j-l)$$

Ανάλογα με τα βάρη (συντελεστές φίλτρου) του πυρήνα $\boldsymbol{h}$, μπορεί να έχουμε διαφορετική συνεισφορά κάθε εισερχόμενου pixel στην εικόνα $\boldsymbol{f}$, δημιουργώντας μια ποικιλία γραμμικών φίλτρων που έχουν σχεδιαστεί για να επιτελούν διάφορες διεργασίες επεξεργασίας εικόνας, όπως η εξομάλυνση, η ενίσχυση και η ανίχνευση ακμών.

Τέτοια παραδείγματα γραμμικών φίλτρων περιλαμβάνουν:

1) Gaussian filter

Το Gaussian φίλτρο εφαρμόζει μία εξομάλυνση στην εικόνα, δίνοντας μεγαλύτερο βάρος στα κοντινά pixels. Έτσι, μειώνει τον θόρυβο και θολώνει ήπια την εικόνα. Συχνά χρησιμοποιείται πριν από αλγορίθμους ανίχνευσης αντικειμένων για να αφαιρέσει μικρές ατέλειες. Για παράδειγμα μια θολή φωτογραφία προσώπου, όπου έχουν εξαλειφθεί τα μικρά στίγματα.

2) Sliding Average ή Box filter

Το φίλτρο αυτό αντικαθιστά την τιμή κάθε pixel με τον μέσο όρο των τιμών των γειτονικών του pixels. Αυτό οδηγεί επίσης σε εξομάλυνση (smoothing) και μείωση του θορύβου όπως η αφαίρεση του «κόκκου» (noise) σε

μια παλιά φωτογραφία.

3) Sobel filter

Το φίλτρο Sobel ενισχύει τις ακμές στην εικόνα, δηλαδή τα σημεία όπου αλλάζει απότομα η ένταση των pixels. Είναι πολύ χρήσιμο για τον εντοπισμό περιγραμμάτων και ορίων αντικειμένων όπως η ανίχνευση των περιγραμμάτων ενός αντικειμένου πάνω σε ένα τραπέζι, ώστε να ξεχωρίζει από το φόντο.

Μη-γραμμικά Φίλτρα (Non-Linear Filters)

Τα μη-γραμμικά φίλτρα συνιστούν μια κατηγορία τεχνικών επεξεργασίας εικόνων, στις οποίες η τιμή κάθε εικονοστοιχείου υπολογίζεται μέσω μη-γραμμικών συναρτήσεων των τιμών έντασης των γειτονικών pixels. Σε αντίθεση με τα γραμμικά φίλτρα, τα μη-γραμμικά φίλτρα αντιμετωπίζουν τον τοπικό πυρήνα ως έναν στατιστικό εκτιμητή, με στόχο τη βελτιστοποιημένη εκτίμηση της τιμής έντασης ενός pixel, ιδίως σε συνθήκες παρουσίας θορύβου. Οι πλέον διαδεδομένοι τύποι μη-γραμμικών φίλτρων περιλαμβάνουν:

1) Median filter

Το φίλτρο διάμεσου (median filter) αντικαθιστά την τιμή κάθε pixel με τη διάμεσο των τιμών έντασης ενός προκαθορισμένου συνόλου γειτονικών pixels. Η διαδικασία αυτή είναι ιδιαίτερα αποτελεσματική στην απομάκρυνση θορύβου τύπου “salt-and-pepper”, καθώς διατηρεί ανέπαφες τις ακμές και τις λεπτομέρειες της εικόνας, σε αντίθεση με τα γραμμικά φίλτρα που συχνά προκαλούν θόλωση (blurring).

2) Bilateral filter

Το διμερές φίλτρο (bilateral filter) υπολογίζει τον νέο τιμή έντασης κάθε pixel λαμβάνοντας υπόψη τόσο τη χωρική εγγύτητα όσο και τη διαφορά στην τιμή έντασης με τα γειτονικά pixels, εφαρμόζοντας έναν σταθμισμένο μέσο όρο. Με τον τρόπο αυτό, επιτυγχάνει τη μείωση του θορύβου, ενώ ταυτόχρονα διατηρεί τις ακμές της εικόνας, χαρακτηριστικό που το καθιστά κατάλληλο για εφαρμογές όπου απαιτείται υψηλή ποιότητα απεικόνισης.

Τα μη-γραμμικά φίλτρα είναι ιδανικά για τη διατήρηση λεπτομερειών στην εικόνα, αλλά έχουν αυξημένο υπολογιστικό κόστος λόγω της πολυπλοκότητάς τους. Γενικός κανόνας είναι ότι τα γρήγορα γραμμικά φίλτρα είναι αποτελεσματικά για απλές εργασίες επεξεργασίας εικόνας (όπως εξομάλυνση και ενίσχυση), όμως όταν η διατήρηση των λεπτομερειών της εικόνας είναι κρίσιμη, η χρήση μη-γραμμικών φίλτρων είναι πολύ πιο κατάλληλη.

2.3.2 Προ επεξεργασία με Φίλτρα Εικόνας

Η προ επεξεργασία των δεδομένων αποτελεί κρίσιμο στάδιο πριν από την εκπαίδευση οποιουδήποτε μοντέλου υπολογιστικής όρασης. Ο βασικός στόχος της προ επεξεργασίας είναι η βελτίωση της ποιότητας των εισόδων, ώστε να αυξηθεί η ακρίβεια και η αποδοτικότητα του μοντέλου στην ανίχνευση και αναγνώριση αντικειμένων. Ιδιαίτερα για σαρωμένες εικόνες (scanned images), συνιστάται η δημιουργία ενός συνόλου δεδομένων (dataset) που να αντανακλά ρεαλιστικές συνθήκες, περιλαμβάνοντας:

- **Εικόνες με προσθήκη θορύβου**

  Έτσι, το μοντέλο μαθαίνει να διαχειρίζεται μικρές ατέλειες ή παραμορφώσεις που συχνά υπάρχουν στις πραγματικές σαρώσεις.
- **Εικόνες με μικρές περιστροφές (±1°)**

  Με αυτό τον τρόπο, το σύστημα γίνεται πιο ανθεκτικό σε μικρές αλλαγές στην ευθυγράμμιση των εικόνων, οι οποίες είναι συχνές κατά τη διαδικασία της σάρωσης.

Για παράδειγμα, εάν θέλουμε να εκπαιδεύσουμε ένα μοντέλο για την αναγνώριση σχεδίων ή εγγράφων από σαρωμένες εικόνες, προσθέτουμε τεχνητό θόρυβο (noise) και εφαρμόζουμε μικρές περιστροφές (π.χ. +1° ή -1°) στις εικόνες εκπαίδευσης. Αυτό βοηθά το μοντέλο να μάθει να εντοπίζει τα αντικείμενα σωστά ακόμη και αν οι εικόνες είναι ελαφρώς παραμορφωμένες ή έχουν ατέλειες, οδηγώντας σε πιο αξιόπιστα αποτελέσματα σε πραγματικά δεδομένα.

2.4 Παραγωγή Θορύβου με Κατανομή Gauss (Gaussian Distribution)

Η τεχνητή παραγωγή θορύβου μέσω της Κανονικής Κατανομής (Gaussian Distribution) αποτελεί βασική τεχνική προσομοίωσης ρεαλιστικών συνθηκών κατά τη συλλογή και επεξεργασία εικόνων. Σε πραγματικές συνθήκες, οι εικόνες συχνά περιέχουν θόρυβο λόγω χαμηλού φωτισμού ή ατελειών του αισθητήρα καταγραφής. Σε μια εφαρμογή εκπαίδευσης μοντέλου υπολογιστικής όρασης, προστίθεται Gaussian θόρυβος σε εικόνες με μέση τιμή μ=0 και συγκεκριμένη τυπική απόκλιση σ, ώστε το μοντέλο να μάθει να αναγνωρίζει αντικείμενα ακόμα και όταν τα δεδομένα περιέχουν ρεαλιστικές ατέλειες και παραμορφώσεις.

Η κατανομή Gauss ορίζεται ως εξής:

$$G(z) = (1 / \sigma\sqrt{2\pi}) \cdot e^{(-[(z-\mu)^2 / 2\sigma^2])}$$

Για την ειδική περίπτωση όπου ο μέσος όρος είναι μηδέν ($\mu = 0$), γράφεται ως:

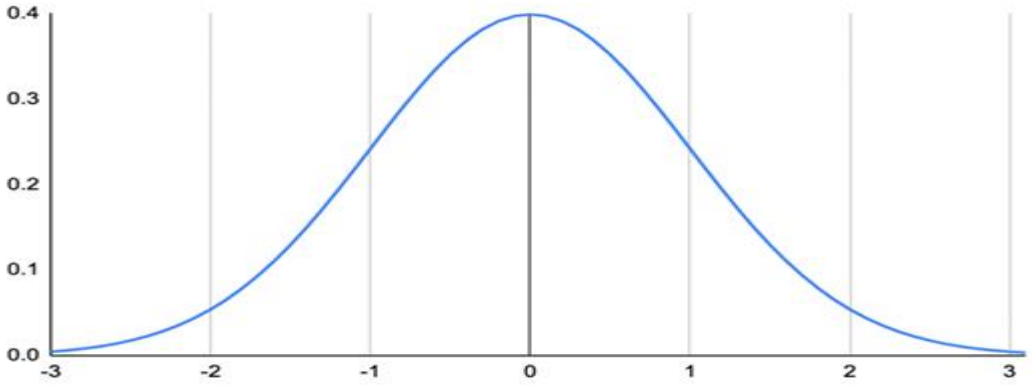
$$Nz(0,\sigma^2) = (1 / \sigma\sqrt{2\pi}) \cdot e^{(-[z^2 / 2\sigma^2])}$$

Όπου:

μ: ο μέσος όρος (mean)

σ: η τυπική απόκλιση (Standard deviation)

Όταν $\mu = 0$, διασφαλίζεται ότι ο προστιθέμενος θόρυβος δεν θα έχει καθαρή επίδραση στη συνολική φωτεινότητα της εικόνας.



2.4.1 Εφαρμογή Θορύβου στην Εικόνα

Χρησιμοποιώντας την κατανομή Gauss, μπορούμε να δημιουργήσουμε τυχαίες τιμές (θόρυβο) με τυχαίες πιθανότητες, τις οποίες προσθέτουμε στις τιμές έντασης των pixels, με μια γραμμική πράξη:

$$I\ nois\ y = I + N(0, \sigma 2)$$

Όπου:

I nois y: η τελική «θορυβώδης» εικόνα

I: η αρχική εικόνα

$N(0, \sigma^2)$: δείγμα θορύβου από την κανονική κατανομή με μέσο 0 και διασπορά σ^2

Όσο μεγαλύτερη είναι η τυπική απόκλιση ($\boldsymbol{\sigma}$) τόσο πιο έντονος θα είναι ο θόρυβος στην εικόνα, και το αντίστροφο. Για τιμή $\boldsymbol{\sigma = 0.5}$ το αποτέλεσμα φαίνεται στην παρακάτω εικόνα:

Επεξήγηση εικόνας:

Χωρίς Θόρυβο (Without Noise): Αριστερά, η αρχική καθαρή εικόνα.

Με Θόρυβο (With Noise): Δεξιά, η ίδια εικόνα με προσθήκη θορύβου (σ = 0.5), ο παραγόμενος θόρυβος είναι αρκετά αισθητός και μπορεί να παρατηρηθεί ως «κόκκος» ή παραμόρφωση στην εικόνα. Αυτή η τεχνική

χρησιμοποιείται ευρέως στη δημιουργία συνόλων δεδομένων για την εκπαίδευση μοντέλων υπολογιστικής όρασης, βοηθώντας τα να γίνουν πιο ανθεκτικά σε θορυβώδεις ή ατελείς πραγματικές εικόνες.

2.5 Περιστροφή Εικόνας (Image Rotation)

Η περιστροφή (Rotation) μιας εικόνας κατά μια συγκεκριμένη γωνία γύρω από το κέντρο της, αν και δεν αποτελεί φίλτρο, είναι μια ιδιαίτερα χρήσιμη μετασχηματιστική ενέργεια (augmentation technique) που χρησιμοποιείται συχνά στην προ επεξεργασία δεδομένων στην υπολογιστική όραση. Ο σκοπός της είναι να αυξήσει την ποικιλία των δεδομένων εκπαίδευσης, καθιστώντας το μοντέλο πιο ανθεκτικό και ακριβές σε διαφορετικούς προσανατολισμούς των αντικειμένων. Η διαδικασία της περιστροφής επιτυγχάνεται μέσω του μετασχηματισμού των διδιάστατων συντεταγμένων κάθε pixel, χρησιμοποιώντας έναν πίνακα περιστροφής (rotation matrix). Για μια εικόνα με διαστάσεις ύψος h και πλάτος w, η περιστροφή πραγματοποιείται γύρω από το κέντρο της εικόνας, δηλαδή το σημείο (w/2,h/2)(w/2, h/2)(w/2,h/2) για οποιαδήποτε γωνία θ δίνεται ως εξής:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} cos\theta & -sin\theta & \frac{w}{2}(1-cos\theta)+\frac{h}{2}sin\theta \\ sin\theta & cos\theta & -\frac{w}{2}sin\theta+\frac{h}{2}(1-cos\theta) \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

Όπου:

(x,y)(x, y)(x,y): Αρχικές συντεταγμένες του pixel
(x′,y′)(x', y')(x′,y′): Νέες συντεταγμένες μετά την περιστροφή
w: Πλάτος εικόνας
h: Ύψος εικόνας
θ: Γωνία περιστροφής (σε ακτίνια)

Οι δύο πρώτες γραμμές του πίνακα εκτελούν ταυτόχρονα την περιστροφή και τη μετάθεση (translation), ώστε η εικόνα να παραμείνει στην ίδια αρχική θέση, ενώ η τελευταία γραμμή αποτελεί ομογενή μετασχηματισμό (homogeneous transformation) για να διατηρηθεί αμετάβλητη η τρίτη διάσταση. Αν έχουμε μια εικόνα μεγέθους 200x200 pixels και επιθυμούμε να την περιστρέψουμε κατά 5° γύρω από το κέντρο της, εφαρμόζουμε τον παραπάνω μετασχηματισμό σε κάθε pixel. Αυτό δημιουργεί μια νέα εικόνα, όπου όλα τα σημεία έχουν περιστραφεί συμμετρικά γύρω από το κέντρο. Η τεχνική αυτή χρησιμοποιείται για να βοηθήσει το μοντέλο να μάθει να αναγνωρίζει αντικείμενα ανεξάρτητα από τη γωνία που εμφανίζονται.

2.6 Μέθοδοι Προ επεξεργασίας

Η εφαρμογή αλγορίθμων Οπτικής Αναγνώρισης Χαρακτήρων (Optical Character Recognition – OCR) για την αυτόματη ανάγνωση αριθμητικών διαστάσεων από μηχανολογικά σχέδια, όπως το σχέδιο τοποθέτησης άξονα (Shaft Arrangement Plan), αποτελεί πλέον αναπόσπαστο στοιχείο της αυτοματοποίησης στη βιομηχανική πληροφορική και στη μηχανολογική τεκμηρίωση. Η αποτελεσματικότητα των συστημάτων OCR εξαρτάται άμεσα από την ποιότητα της εισερχόμενης εικόνας. Για τον λόγο αυτό, προηγούνται συνήθως στάδια προ επεξεργασίας της εικόνας, τα οποία περιλαμβάνουν την εφαρμογή ειδικών φίλτρων επεξεργασίας εικόνας με σκοπό τη βελτίωση της ευκρίνειας και της διακριτότητας των προς αναγνώριση χαρακτήρων. Ειδικότερα, χρησιμοποιούνται:

Φίλτρα ανίχνευσης ακμών (όπως τα Sobel ή Canny), τα οποία ενισχύουν τις περιοχές της εικόνας όπου εντοπίζονται απότομες αλλαγές στην ένταση, καθιστώντας τα όρια των χαρακτήρων περισσότερο ευδιάκριτα για το OCR.
Φίλτρα μείωσης θορύβου (όπως τα Median ή Gaussian), τα οποία απομακρύνουν τις ανεπιθύμητες διακυμάνσεις της φωτεινότητας και εξαλείφουν παραμορφώσεις που δυσχεραίνουν τη διαδικασία αναγνώρισης.

Φίλτρο Ανίχνευσης Ακμών Sobel

Το φίλτρο Sobel αποτελεί μία από τις πιο διαδεδομένες μεθόδους ανίχνευσης ακμών. Υλοποιείται μέσω συνέλιξης της εικόνας με δύο πυρήνες (kernels) διαστάσεων 3x3: έναν για την ανίχνευση οριζόντιων ακμών και έναν για την ανίχνευση κάθετων ακμών. Οι πυρήνες Sobel, όπως ορίζονται από τους Feldman και συν., υπολογίζουν το τοπικό βαθμωτό (gradient) της τιμής έντασης των εικονοστοιχείων.

Παράδειγμα Εφαρμογής Φίλτρου Sobel

- **Αριστερά:** Εικόνα χωρίς την εφαρμογή φίλτρου (No filter), όπου τα όρια μεταξύ αντικειμένων δεν είναι ευκρινώς διακριτά.
- **Δεξιά:** Εικόνα μετά την εφαρμογή του φίλτρου Sobel-Feldman (Sobel-Feldman Filter), στην οποία οι ακμές, όπως τα περιγράμματα του πλοίου και της θάλασσας, καθίστανται σαφώς ορατές, ενώ οι περιοχές με ήπιες μεταβολές στην ένταση καταστέλλονται.

Επιπρόσθετα, η OpenCV περιλαμβάνει πρόσβαση σε προ εκπαιδευμένα μοντέλα βαθιάς μάθησης (deep Learning), τα οποία μπορούν να χρησιμοποιηθούν για προηγμένες εφαρμογές ανίχνευσης και αναγνώρισης αντικειμένων, με τη δυνατότητα περαιτέρω προσαρμογής (fine-tuning) σε εξειδικευμένα δεδομένα και σενάρια. Στη παρόν εργασία, η OpenCV αξιοποιείται κυρίως για: διαχείριση πηγών βίντεο (κάμερες, αρχεία, RTSP), προ επεξεργασία (resize, Gaussian, μετατροπές χρωματικού χώρου), ενίσχυση τοπικής αντίθεσης (CLAHE).

2.7 Μηχανική Μάθηση

Η μηχανική μάθηση (Machine Learning), όπως ορίστηκε αρχικά από τον Arthur Samuel, αφορά την ικανότητα των υπολογιστικών συστημάτων να «μαθαίνουν» και να βελτιώνουν την απόδοσή τους σε συγκεκριμένες εργασίες χωρίς να έχουν προγραμματιστεί ρητά για κάθε δυνατό ενδεχόμενο. Ένας πιο σύγχρονος και ευρέως αποδεκτός ορισμός προτείνεται από τον Tom Mitchell (1997), ο οποίος διατυπώνει (Mitchell, T. M. (1997). *Machine Learning*. New York: McGraw-Hill):

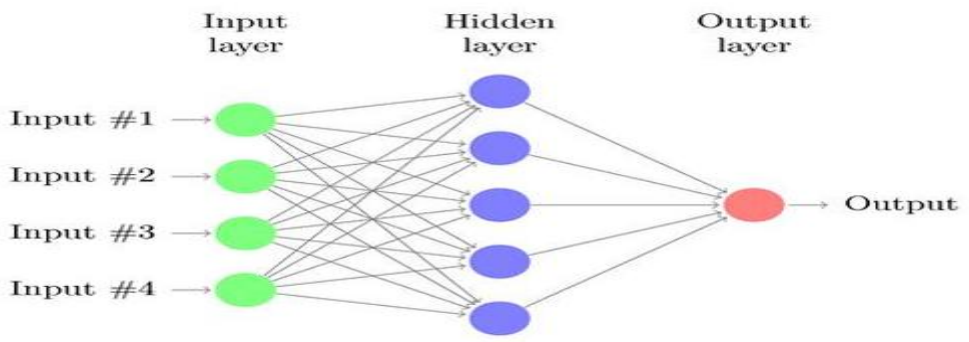
"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E."

Δηλαδή

"Ένα πρόγραμμα υπολογιστή μαθαίνει από εμπειρία E ως προς μια κατηγορία εργασιών T και μέτρο απόδοσης P, εάν η απόδοσή του στις εργασίες αυτές, όπως μετριέται από το P, βελτιώνεται με την εμπειρία E."

2.8 Αρχιτεκτονική και Αρχές Νευρωνικών Δικτύων

Τη δεκαετία του 1980–1990 η εξέλιξη των μονάδων γραφικών επεξεργασίας (GPUs) επέτρεψε την επιτάχυνση της εκπαίδευσης και την ανάπτυξη πολύπλοκων και ολοκληρωμένων δικτύων. Τα τεχνητά νευρωνικά δίκτυα δομούνται από πολλαπλούς νευρώνες τύπου περσέπτρον (Perceptrons), οι οποίοι οργανώνονται σε επίπεδα (layers). Το διάγραμμα που ακολουθεί απεικονίζει τη βασική αρχιτεκτονική ενός νευρωτικού δικτύου:



Περιγραφή του Διαγράμματος

Input layer (Στρώμα εισόδου): Περιλαμβάνει τα δεδομένα εισόδου (π.χ. Input #1, #2, #3, #4).
Hidden layer (Κρυφό στρώμα): Ενδιάμεσα επίπεδα όπου πραγματοποιείται η επεξεργασία των δεδομένων μέσω ενεργοποιήσεων νευρώνων. Κάθε νευρώνας σε αυτό το επίπεδο λαμβάνει σήματα από όλους τους νευρώνες του επιπέδου εισόδου.
Output layer (Στρώμα εξόδου): Το τελικό επίπεδο που παρέχει το αποτέλεσμα του δικτύου.
Κάθε είσοδος συνδέεται με όλους τους νευρώνες του κρυφού στρώματος, και κάθε νευρώνας του κρυφού στρώματος συνδέεται με το στρώμα εξόδου. Αυτή η διασύνδεση ονομάζεται πλήρως συνδεδεμένο δίκτυο (fully connected network).
Ένα τεχνητό νευρωνικό δίκτυο (Artificial Neural Network) αποτελείται από τρία βασικά είδη στρωμάτων (layers):

Στρώμα Εισόδου (Input Layer)

Το στρώμα εισόδου αποτελείται από νευρώνες που λαμβάνουν τα αρχικά δεδομένα του προβλήματος και τα προωθούν προς τα επόμενα επίπεδα του νευρωνικού δικτύου. Ο αριθμός των νευρώνων στο στρώμα αυτό εξαρτάται από τη διαστατικότητα των δεδομένων εισόδου. Για παράδειγμα, σε εφαρμογές επεξεργασίας εικόνας, το στρώμα εισόδου περιλαμβάνει τόσους νευρώνες όσοι είναι και τα εικονοστοιχεία (pixels) της εικόνας.

Κρυφά Στρώματα (Hidden Layers)

Τα κρυφά στρώματα βρίσκονται ανάμεσα στο στρώμα εισόδου και το στρώμα εξόδου ενός νευρωνικού δικτύου. Ονομάζονται «κρυφά» επειδή οι ενεργοποιήσεις (activations) που παράγουν δεν είναι άμεσα ορατές ούτε στον τελικό χρήστη ούτε στο εξωτερικό περιβάλλον. Αυτά τα στρώματα παίζουν κρίσιμο ρόλο, καθώς επιτρέπουν στο δίκτυο να μαθαίνει σύνθετα, αφηρημένα και υψηλού επιπέδου χαρακτηριστικά από τα δεδομένα, γεγονός που το καθιστά ικανό να επιλύει περίπλοκα προβλήματα. Όταν το δίκτυο περιλαμβάνει πολλά κρυφά στρώματα, τότε χαρακτηρίζεται ως βαθύ νευρωνικό δίκτυο (Deep Neural Network – DNN). Τέτοιου τύπου δίκτυα διαθέτουν ενισχυμένη ικανότητα αναπαράστασης και μάθησης, επιτρέποντας την κατανόηση περίπλοκων δομών και συσχετίσεων μέσα στα δεδομένα.

Στρώμα Εξόδου (Output Layer)

Το στρώμα εξόδου αποτελεί το τελευταίο επίπεδο ενός νευρωνικού δικτύου και είναι υπεύθυνο για την παραγωγή των τελικών αποτελεσμάτων, δηλαδή των προβλέψεων του μοντέλου. Ο αριθμός των νευρώνων που περιλαμβάνει καθορίζεται από τη φύση του προβλήματος και τη διαστατικότητα της επιθυμητής εξόδου. Για παράδειγμα, σε ένα πρόβλημα ταξινόμησης αντικειμένων, το στρώμα εξόδου περιλαμβάνει τόσους νευρώνες όσες και οι κατηγορίες στις οποίες μπορεί να ανήκει ένα αντικείμενο. Κάθε νευρώνας αντιστοιχεί σε μία κατηγορία και η τιμή του εκφράζει την πιθανότητα ή τον βαθμό εμπιστοσύνης (confidence score) της πρόβλεψης του μοντέλου για αυτήν την κατηγορία.

Η αύξηση του αριθμού των κρυφών στρωμάτων επιτρέπει στα νευρωνικά δίκτυα να εντοπίζουν όλο και πιο σύνθετα μοτίβα στα δεδομένα, βελτιώνοντας την ικανότητα γενίκευσης και την ακρίβεια των προβλέψεων του μοντέλου.

2.8.1 Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks - CNNs)

Τα Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks – CNNs) παρουσιάστηκαν για πρώτη φορά από τον Yann LeCun και τους συνεργάτες του τη δεκαετία του 1980, με χαρακτηριστικό παράδειγμα τη δημοσίευση “Backpropagation Applied to Handwritten Zip Code Recognition” (LeCun et al., 1989). Ο LeCun διαπίστωσε ότι τα πλήρως συνδεδεμένα νευρωνικά δίκτυα (Fully Connected Neural Networks), όπως τα βαθιά νευρωνικά δίκτυα (Deep Neural Network – DNNs) και τα απλά εμπρόσθια δίκτυα (Feed Forward ANNs), δεν είναι κατάλληλα για την επεξεργασία δεδομένων σε μορφή πλέγματος, όπως οι δισδιάστατες εικόνες. Ο κύριος λόγος είναι ότι ο τεράστιος αριθμός παραμέτρων (weights) που απαιτείται, λόγω της σύνδεσης κάθε εισόδου με κάθε νευρώνα στα κρυφά στρώματα, αυξάνει σημαντικά την υπολογιστική πολυπλοκότητα και καθιστά την εκπαίδευση αναποτελεσματική. Για να αντιμετωπίσει αυτό το ζήτημα, ο LeCun εμπνεύστηκε από τον βιολογικό οπτικό φλοιό (visual cortex) του ανθρώπινου εγκεφάλου. Ο οπτικός φλοιός είναι οργανωμένος ιεραρχικά σε πολλαπλά στρώματα, τα οποία επεξεργάζονται σταδιακά διαφορετικές πτυχές του οπτικού ερεθίσματος. Τα πρώτα στρώματα ανιχνεύουν βασικά χαρακτηριστικά, όπως ακμές και γωνίες, ενώ τα βαθύτερα στρώματα συνδυάζουν αυτές τις πληροφορίες για να αναγνωρίσουν πιο σύνθετα σχήματα, αντικείμενα και σκηνές. Η βασική ιδέα που μεταφέρθηκε στα CNNs είναι η ύπαρξη πολλών επιπέδων αφαίρεσης (levels of abstraction), όπου κάθε επίπεδο εξάγει χαρακτηριστικά ολοένα και μεγαλύτερης πολυπλοκότητας από τα δεδομένα εισόδου. Παράλληλα, εισήχθη η τεχνική της «μοιρασιάς βαρών» (weight sharing), σύμφωνα με την οποία οι ίδιοι πυρήνες (kernels) χρησιμοποιούνται για την εξαγωγή χαρακτηριστικών από διαφορετικές περιοχές της εικόνας. Αυτή η στρατηγική μειώνει δραστικά τον συνολικό αριθμό παραμέτρων που πρέπει να εκπαιδευτούν, καθώς ο αριθμός των βαρών δεν εξαρτάται πλέον γραμμικά από το μέγεθος της εισόδου. Αυτή η καινοτόμος προσέγγιση έκανε τα CNNs εξαιρετικά αποδοτικά στην ανάλυση και επεξεργασία οπτικών δεδομένων, δημιουργώντας τη βάση για τη ραγδαία εξέλιξη των τεχνολογιών υπολογιστικής όρασης.

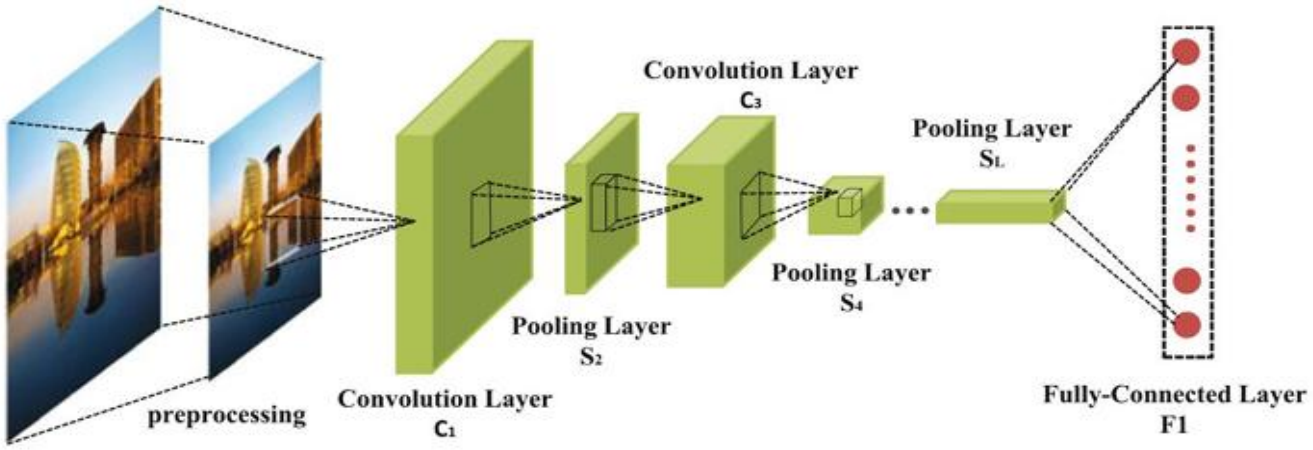
Δομή του CNN

Η αρχιτεκτονική ενός Συνελικτικού Νευρωνικού Δικτύου (Convolutional Neural Network – CNN) συνίσταται από πολλαπλά διαδοχικά στρώματα, το καθένα από τα οποία διαδραματίζει διακριτό ρόλο στη διαδικασία της εξαγωγής και της ερμηνείας χαρακτηριστικών από την εικόνα εισόδου. Τα βασικά δομικά στοιχεία ενός CNN περιλαμβάνουν:

- **Συνελικτικά στρώματα (Convolutional layers):** Υπεύθυνα για την εξαγωγή χαρακτηριστικών (feature extraction) μέσω εφαρμογής φίλτρων (kernels) στην εικόνα.
- **Στρώματα υπό-δειγματοληψίας ή συμπύκνωσης (Pooling layers):** Χρησιμοποιούνται για τη μείωση

της διαστατικότητας των δεδομένων, συγκρατώντας τις πιο σημαντικές πληροφορίες.

- **Πλήρως συνδεδεμένα στρώματα (Fully Connected layers):** Αναλαμβάνουν την τελική ταξινόμηση ή πρόβλεψη βασιζόμενα στα εξαγόμενα χαρακτηριστικά.



Συνελικτικά Στρώματα (Convolutional Layers)

Τα συνελικτικά στρώματα αποτελούν τον ακρογωνιαίο λίθο της λειτουργίας ενός CNN. Κάθε συνελικτικό στρώμα εφαρμόζει ένα σύνολο φίλτρων (kernels), τα οποία εκπαιδεύονται κατά τη διαδικασία μάθησης, με στόχο την ανίχνευση συγκεκριμένων μοτίβων (όπως ακμές, υφές ή σύνθετα σχήματα) στα δεδομένα εισόδου. Ως είσοδο δέχονται έναν πίνακα τιμών έντασης εικονοστοιχείων (pixels) της εικόνας, συνήθως σε δισδιάστατη μορφή.

Η βασική λειτουργία του Συνελικτικού στρώματος περιγράφεται ως εξής:

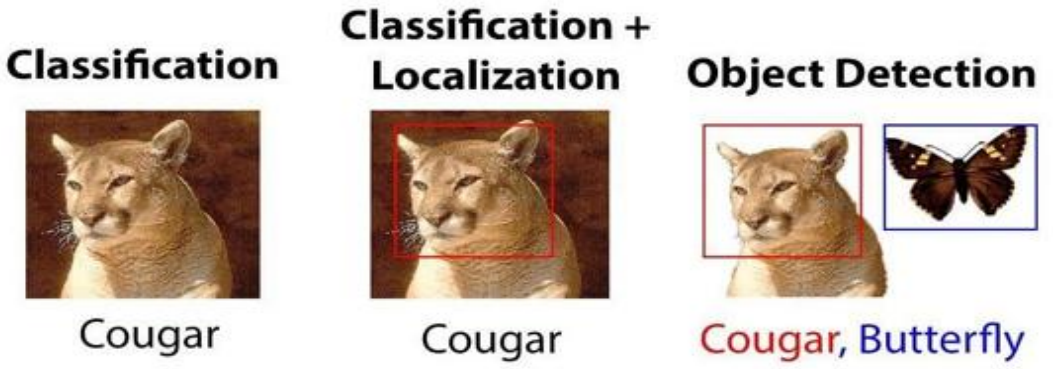
- Ένας μικρός πυρήνας (kernels ή φίλτρο) με συγκεκριμένες διαστάσεις (π.χ. 3x3 ή 5x5) «γλιστρά» (slides) κατά μήκος της εικόνας εισόδου.
- Σε κάθε θέση, οι τιμές του πυρήνα πολλαπλασιάζονται στοιχειομετρικά με τις αντίστοιχες τιμές της περιοχής της εικόνας που καλύπτει το φίλτρο.
- Το άθροισμα των πολλαπλασιασμών παράγει μια νέα τιμή, η οποία αποθηκεύεται στον αντίστοιχο χάρτη χαρακτηριστικών (feature map).
- Η διαδικασία επαναλαμβάνεται, καλύπτοντας ολόκληρη την είσοδο.

2.8.2 Ανίχνευση Αντικειμένων με χρήση CNNs

Η ανίχνευση αντικειμένων (object detection) αποτελεί ένα από τα θεμελιώδη και πιο απαιτητικά προβλήματα της υπολογιστικής όρασης. Ο πυρήνας του προβλήματος περιλαμβάνει δύο βασικές διεργασίες:

- Εντοπισμό (Identification): Αναγνώριση της ύπαρξης και του τύπου (κατηγορίας) των αντικειμένων εντός μιας εικόνας ή ακολουθίας εικόνων (βίντεο).
- Καθορισμό της θέσης (Localization): Ακριβής προσδιορισμός της θέσης και των ορίων (boundary extent) κάθε αντικειμένου μέσα στην εικόνα.

Ο στόχος των συστημάτων ανίχνευσης αντικειμένων είναι διττός. Πρώτον να ταξινομούν (classify) σωστά τα αντικείμενα που υπάρχουν στην εικόνα σε προκαθορισμένες κατηγορίες και δεύτερον να εντοπίζουν με ακρίβεια τη θέση κάθε αντικειμένου, η οποία αποδίδεται συνήθως με ένα ορθογώνιο πλαίσιο (bounding box) που οριοθετεί το αντικείμενο. Το αποτέλεσμα της ανίχνευσης αντικειμένων είναι ένα σύνολο προβλέψεων, όπου για κάθε αντικείμενο αναφέρονται η κατηγορία και η θέση του μέσα στην εικόνα. Η ανίχνευση αντικειμένων παραμένει μια ιδιαίτερα δύσκολη πρόκληση στην υπολογιστική όραση, καθώς απαιτεί τον ακριβή εντοπισμό και ταξινόμηση αντικειμένων που μπορεί να διαφέρουν σε μέγεθος, προσανατολισμό, σχήμα, να παρουσιάζουν μερική απόκρυψη (occlusion), να βρίσκονται σε πολύπλοκα ή θορυβώδη φόντα και να απεικονίζονται υπό διαφορετικές συνθήκες φωτισμού. Τα τελευταία χρόνια, η εμφάνιση και η ραγδαία πρόοδος των συνελικτικών νευρωνικών δικτύων (Convolutional Neural Networks – CNNs) έχει συμβάλει καταλυτικά στην πρόοδο της ανίχνευσης αντικειμένων. Τα CNNs έχουν τη δυνατότητα να μαθαίνουν αυτόματα τα κατάλληλα χαρακτηριστικά απευθείας από τα ακατέργαστα δεδομένα εισόδου (raw pixels) και να συνδυάζουν διαδικασίες εξαγωγής χαρακτηριστικών και ταξινόμησης σε ένα ενιαίο μοντέλο.



Επεξήγηση Εικόνας

- Classification (Ταξινόμηση): Το μοντέλο αναγνωρίζει το είδος του αντικειμένου (π.χ. Cougar).
- Classification + Localization (Ταξινόμηση και Εντοπισμός): Το μοντέλο εντοπίζει και την περιοχή που βρίσκεται το αντικείμενο (με bounding box).
- Object Detection (Ανίχνευση Αντικειμένων): Το μοντέλο αναγνωρίζει και εντοπίζει ταυτόχρονα πολλαπλά αντικείμενα στην εικόνα (π.χ. Cougar και Butterfly), το καθένα με το δικό του bounding box και ετικέτα.

2.8.3 Κατηγορίες Μεθόδων Ανίχνευσης Αντικειμένων με CNNs

Οι σύγχρονες μέθοδοι ανίχνευσης αντικειμένων που βασίζονται σε συνελικτικά νευρωνικά δίκτυα (CNNs) διακρίνονται σε δύο βασικές κατηγορίες:

- Δι-σταδιακές μέθοδοι (2-stage methods):

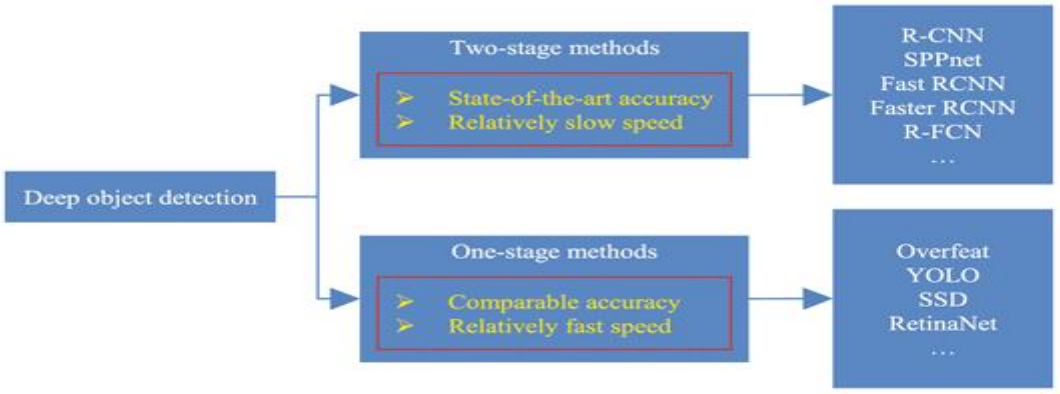
  Στις μεθόδους αυτές, το σύστημα αρχικά εντοπίζει περιοχές ενδιαφέροντος (object proposals), δηλαδή

περιοχές της εικόνας που είναι πιθανό να περιέχουν κάποιο αντικείμενο. Στη συνέχεια, κάθε υποψήφια περιοχή ταξινομείται ως προς την κατηγορία του αντικειμένου. Τυπικό παράδειγμα δι-σταδιακής μεθόδου είναι το Faster R-CNN.

- Μονοσταδιακές μέθοδοι (one-stage methods):

  Σε αυτές τις μεθόδους, η εξαγωγή των προτάσεων αντικειμένων και η ταξινόμησή τους πραγματοποιούνται ταυτόχρονα, σε ένα ενιαίο στάδιο. Δημοφιλή παραδείγματα αποτελούν τα YOLO (You Only Look Once) και SSD (Single Shot Detector).

Η επιλογή της κατάλληλης μεθόδου εξαρτάται από τις απαιτήσεις της εκάστοτε εφαρμογής. Σε εφαρμογές όπου προτεραιότητα αποτελεί η ακρίβεια (π.χ. ιατρική διάγνωση), προτιμώνται οι δι-σταδιακές μέθοδοι. Σε εφαρμογές όπου απαιτείται πραγματικός χρόνος ή επεξεργασία μεγάλου όγκου δεδομένων (π.χ. αυτόνομα οχήματα, βιομηχανία), οι Μονοσταδιακές μέθοδοι είναι συχνά καταλληλότερες.



.

Δι-σταδιακές Μέθοδοι Ανίχνευσης Αντικειμένων (Two-Stage Methods)

Οι δι-σταδιακές μέθοδοι αποτελούν μία από τις βασικές προσεγγίσεις στην ανίχνευση αντικειμένων με χρήση συνελικτικών νευρωνικών δικτύων (CNNs).

Η διαδικασία συνίσταται σε δύο κύρια στάδια:

1. Στάδιο πρότασης περιοχών (Region Proposal stage): Δημιουργείται ένα σύνολο υποψήφιων περιοχών εντός της εικόνας, οι οποίες είναι πιθανό να περιέχουν αντικείμενα.
2. Στάδιο ανίχνευσης αντικειμένων (object detection stage): Οι προτεινόμενες περιοχές αξιολογούνται περαιτέρω, ταξινομούνται σε προκαθορισμένες κατηγορίες και γίνεται ακριβής προσδιορισμός των πλαισίων τους.

Η βασική αρχή των δι-σταδιακών μεθόδων είναι ο διαχωρισμός της πρότασης πιθανών περιοχών από την τελική διαδικασία ταξινόμησης και εντοπισμού, μειώνοντας σημαντικά τον συνολικό αριθμό περιοχών που πρέπει να εξετάσει ο ταξινομητής. Αυτό οδηγεί σε μεγαλύτερη ανθεκτικότητα στο φόντο και υψηλότερη ακρίβεια, καθώς η προσοχή του μοντέλου επικεντρώνεται σε περιοχές με πραγματικό ενδιαφέρον.

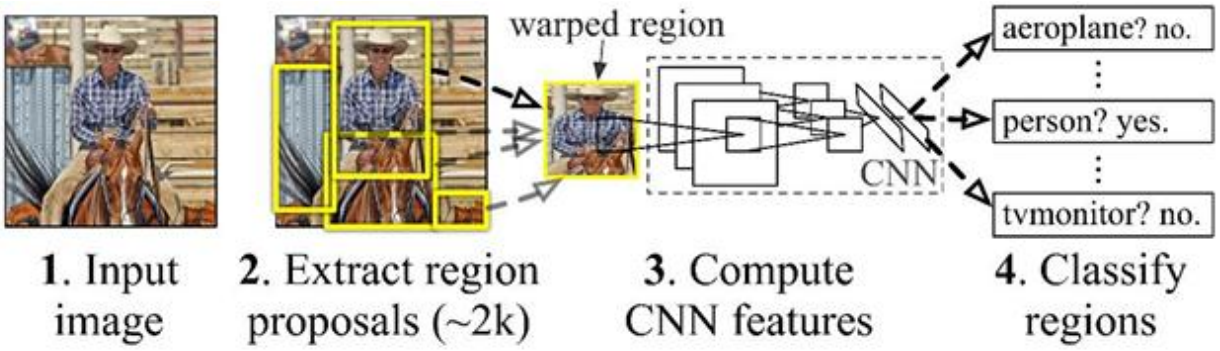
A) R-CNN (Regional Convolutional Neural Network)

Το R-CNN (Girshick et al., 2014) υπήρξε η πρώτη επιτυχημένη δι-σταδιακή προσέγγιση για την ανίχνευση

αντικειμένων με CNNs η οποία λειτουργεί ως εξής:

1. Πρόταση περιοχών: Εφαρμόζεται ο αλγόριθμος Selective Search για τμηματοποίηση της εικόνας και ομαδοποίηση παρόμοιων περιοχών, παράγοντας πολλαπλά bounding boxes ως υποψήφιες περιοχές.
2. Εξαγωγή χαρακτηριστικών: Κάθε υποψήφια περιοχή περνά από ένα προ εκπαιδευμένο συνελικτικό δίκτυο (π.χ. AlexNet), από όπου εξάγονται χαρακτηριστικά (feature vectors).
3. Ταξινόμηση: Τα εξαγόμενα χαρακτηριστικά ταξινομούνται σε κατηγορίες μέσω ανεξάρτητων γραμμικών SVMs, ένα για κάθε κατηγορία αντικειμένου.
4. Βελτίωση πλαισίων: Εφαρμόζεται bounding box regression ώστε να βελτιωθούν οι συντεταγμένες των πλαισίων και να προσεγγίζουν καλύτερα τα πραγματικά όρια των αντικειμένων.

Παρότι το R-CNN σημείωσε σημαντική επιτυχία (π.χ. mAP ≈ 21% στο PASCAL VOC2010), εμφανίζει υπολογιστικές αδυναμίες λόγω του υψηλού κόστους εξαγωγής χαρακτηριστικών για κάθε πρόταση περιοχής και της έλλειψης ενιαίας βελτιστοποίησης των σταδίων.

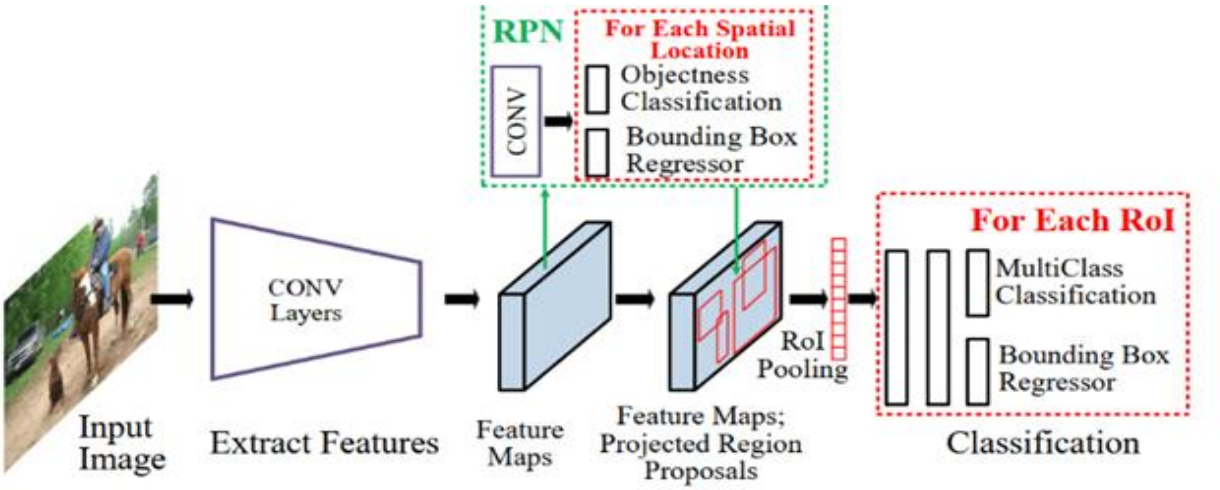


Επεξήγηση Διαγράμματος

- **Input image:** Είσοδος της αρχικής εικόνας.
- **Extract region proposals (~2k):** Εξαγωγή περίπου 2000 υποψήφιων περιοχών με bounding boxes (π.χ. μέσω Selective Search).
- **Compute CNN features:** Κάθε υποψήφια περιοχή μετασχηματίζεται και περνάει από ένα CNN για εξαγωγή χαρακτηριστικών.
- **Classify Region:** Κάθε περιοχή ταξινομείται ως προς το αν περιέχει αντικείμενο συγκεκριμένης κατηγορίας (π.χ. “person”, “aeroplane”, “tvmonitor”).

B) Faster R-CNN

Το **Faster R-CNN** (Ren et al., 2015) αποτέλεσε ουσιαστική βελτίωση του R-CNN, εισάγοντας το Region Proposal Network (RPN), ένα επιπλέον συνελικτικό δίκτυο που δημιουργεί προτάσεις περιοχών απευθείας από τα εξαγόμενα χαρακτηριστικά του CNN backbone, καταργώντας έτσι την ανάγκη για εξωτερικούς αλγορίθμους πρότασης (π.χ. Selective Search). Με τη χρήση της τεχνικής RoI (Region of Interest) Pooling, τα χαρακτηριστικά που εξάγονται από το CNN backbone μοιράζονται τόσο με το RPN όσο και με το δίκτυο ανίχνευσης αντικειμένων, εξασφαλίζοντας από κοινού βελτιστοποίηση και σημαντική μείωση του χρόνου επεξεργασίας. Το αποτέλεσμα είναι η επίτευξη υψηλότερης ακρίβειας με πολύ μεγαλύτερη ταχύτητα, γεγονός που οδήγησε στην καθιέρωση του Faster R-CNN ως πρότυπο για τις δι-σταδιακές μεθόδους ανίχνευσης αντικειμένων.



Επεξήγηση Διαγράμματος

- **Input Image:** Είσοδος της αρχικής εικόνας.
- **Extract Features:** Εξαγωγή χαρακτηριστικών μέσω των συνελικτικών στρωμάτων (CONV Layers).
- **Feature Maps:** Οι χάρτες χαρακτηριστικών (feature maps) περνούν στο Region Proposal Network (RPN), όπου για κάθε χωρική θέση υπολογίζεται:
    1. Αν υπάρχει αντικείμενο (objectness classification)
    2. Η θέση του αντικειμένου (bounding box regressor)
- **RoI Pooling:** Οι προτάσεις περιοχών (Region proposals) προβάλλονται στους χάρτες χαρακτηριστικών και μετατρέπονται μέσω της διαδικασίας RoI Pooling.
- **Classification:** Για κάθε RoI, γίνεται:
    1. **Πολύ κατηγορική ταξινόμηση (Multiclass Classification)**
    2. **Βελτίωση των συντεταγμένων των bounding boxes (Bounding Box Regressor)**

Μονοσταδιακές Μέθοδοι για Ανίχνευση Αντικειμένων (One-Stage Methods)

Σε αντίθεση με τις δι-σταδιακές μεθόδους, οι Μονοσταδιακές μέθοδοι (one-stage methods) έχουν ως στόχο την ταυτόχρονη πρόβλεψη τόσο της θέσης όσο και της κατηγορίας των αντικειμένων απευθείας από την εικόνα εισόδου. Αυτό οδηγεί σε σημαντική αύξηση της ταχύτητας ανίχνευσης, διατηρώντας παράλληλα σε μεγάλο βαθμό υψηλή ακρίβεια. Οι δύο κυριότεροι αλγόριθμοι της κατηγορίας αυτής είναι το SSD (Single Shot Detector) και το YOLO (You Only Look Once).

A) One-Stage Methods - SSD (Single Shot Detector)

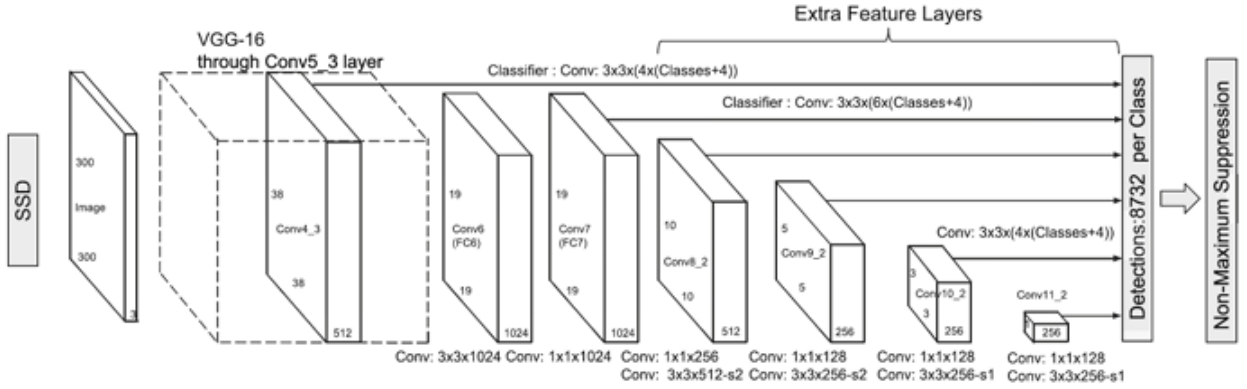
Το SSD αποτελεί μια από τις πιο αξιόπιστες και ταχείες προσεγγίσεις για την ανίχνευση αντικειμένων. Η αρχιτεκτονική του περιλαμβάνει, ένα βασικό συνελικτικό δίκτυο (base CNN), συνήθως ένα προ εκπαιδευμένο δίκτυο όπως το VGG-16 και μια σειρά από επιπλέον συνελικτικά στρώματα (extra feature layers), τα οποία

εξάγουν χαρακτηριστικά σε διαφορετικές κλίμακες και μεγέθη.

Η βασική ιδέα του SSD βασίζεται στη χρήση προκαθορισμένων anchor boxes (ή default boxes) με διάφορες κλίμακες και αναλογίες διαστάσεων, τα οποία εφαρμόζονται σε κάθε επίπεδο πρόβλεψης του δικτύου. Για κάθε anchor box:

- Το SSD προβλέπει πιθανότητες κατηγορίας (Class probabilities) μέσω της συνάρτησης SoftMax στα αποτελέσματα των συνελικτικών φίλτρων.
- Υπολογίζει επίσης μετατοπίσεις (offsets) των bounding boxes σε σχέση με το ground truth, ώστε να προσαρμόζει δυναμικά τη θέση κάθε anchor box στο πραγματικό αντικείμενο.

Μετά την πρόβλεψη, εφαρμόζεται η διαδικασία Non-Maximum Suppression (NMS) για την απομάκρυνση επικαλυπτόμενων ανιχνεύσεων και τη διατήρηση μόνο των bounding boxes με το υψηλότερο confidence score για κάθε αντικείμενο.



Επεξήγηση Διαγράμματος

- **VGG-16 through Conv5_3 layer:** Το βασικό CNN που εξάγει τα αρχικά χαρακτηριστικά από την εικόνα.
- **Extra Feature Layers:** Επιπλέον συνελικτικά στρώματα διαφορετικής ανάλυσης που ανιχνεύουν αντικείμενα διαφορετικών μεγεθών.
- **Detections per Class & Non-Maximum Suppression:** Τα τελικά αποτελέσματα ανίχνευσης μετά το φιλτράρισμα NMS.

B) One-Stage Methods - YOLO (You Only Look Once)

Το YOLO (You Only Look Once) παρουσιάστηκε για πρώτη φορά από τον Joseph Redmon και τους συνεργάτες του το 2016, στο άρθρο ***"You Only Look Once: Unified, Real-Time Object Detection".*** Η αρχιτεκτονική του αποτελείται από 24 συνελικτικά στρώματα (convolutional layers) και 2 πλήρως συνδεδεμένα στρώματα (fully connected layers), με τα πρώτα 20 στρώματα να είναι προ εκπαιδευμένα στο ImageNet, ώστε να επιτυγχάνεται αποδοτική εξαγωγή χαρακτηριστικών.

Η καινοτομία του YOLO έγκειται στον διαμερισμό της εικόνας εισόδου σε ένα πλέγμα διαστάσεων KxK. Κάθε

κελί του πλέγματος είναι υπεύθυνο για την ανίχνευση αντικειμένων των οποίων το κέντρο βρίσκεται μέσα σε αυτό. Για κάθε κελί:

- Προβλέπονται **BBB** bounding boxes με αντίστοιχα Confidence scores, που εκφράζουν τόσο την πιθανότητα ύπαρξης αντικειμένου όσο και την ακρίβεια εντοπισμού της θέσης (μέσω του IoU σε σχέση με το ground truth).
- Κατά την εκπαίδευση, το YOLO αναθέτει σε κάθε predictor την πρόβλεψη συγκεκριμένων αντικειμένων, επιλέγοντας εκείνον με το υψηλότερο IoU σε σχέση με το πραγματικό πλαίσιο. Αυτό οδηγεί τους predictors να εξειδικεύονται σε συγκεκριμένες κατηγορίες, μεγέθη ή αναλογίες αντικειμένων, βελτιώνοντας σημαντικά την ανάκληση (recall) του συστήματος.

Για παράδειγμα:

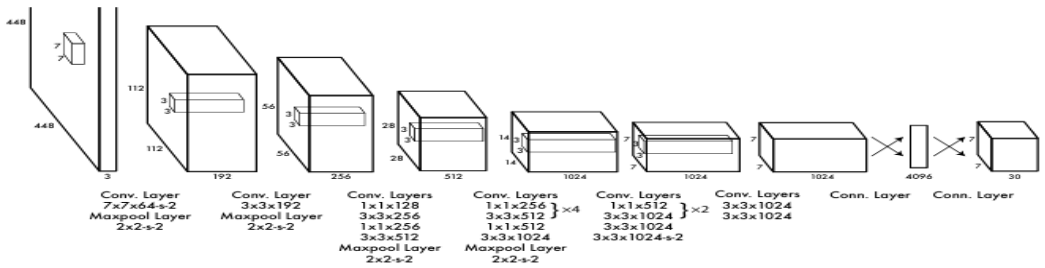
Ας υποθέσουμε ότι έχουμε μια εικόνα από έναν δρόμο, όπου φαίνονται ένα αυτοκίνητο, ένας πεζός και ένα ποδήλατο. Το YOLO χωρίζει την εικόνα σε ένα πλέγμα μεγέθους **7×7** (δηλαδή 49 κελιά). Κάθε κελί είναι υπεύθυνο να εντοπίζει αντικείμενα των οποίων το κέντρο βρίσκεται μέσα σε αυτό το κελί.

- Στο κελί που περιέχει το κέντρο του αυτοκινήτου, το δίκτυο προβλέπει **π.χ. 2 ή 3 bounding boxes (BBB)**. Για κάθε bounding box δίνει:
    1. **Συντεταγμένες** (θέση και μέγεθος του πλαισίου).
    2. **Confidence score**, που δείχνει πόσο σίγουρο είναι ότι υπάρχει αντικείμενο.
    3. **IoU (Intersection over Union)**, που δείχνει πόσο καλά ταιριάζει το προβλεπόμενο πλαίσιο με το πραγματικό.

- Το ίδιο συμβαίνει για τον πεζό και το ποδήλατο, με τα αντίστοιχα κελιά να είναι υπεύθυνα για τις προβλέψεις τους.

Κατά την εκπαίδευση, για κάθε πραγματικό αντικείμενο (ground truth) ανατίθεται σε ένα συγκεκριμένο predictor το καθήκον της πρόβλεψης, με βάση ποιο bounding box έχει το μεγαλύτερο IoU. Έτσι, με τον καιρό, κάποιοι predictors «μαθαίνουν» να εξειδικεύονται σε μεγάλα αντικείμενα (π.χ. αυτοκίνητα), άλλοι σε μικρότερα (π.χ. πεζοί ή ποδήλατα). Αυτό αυξάνει την ακρίβεια και την ικανότητα ανάκλησης του συστήματος.

Όπως και το SSD, το YOLO χρησιμοποιεί τη μέθοδο Non-Maximum Suppression (NMS) για να απομακρύνει τα επικαλυπτόμενα ή λανθασμένα bounding boxes, διατηρώντας μόνο τα πιο αξιόπιστα πλαίσια για κάθε αντικείμενο. Σήμερα, το YOLO –και ιδιαίτερα οι πιο πρόσφατες εκδόσεις του– θεωρείται μία από τις πιο σύγχρονες και αποδοτικές προσεγγίσεις για την ανίχνευση αντικειμένων, με ευρεία χρήση τόσο στην έρευνα όσο και σε βιομηχανικές εφαρμογές.



Επεξήγηση εικόνας:

Η επεξεργασία ξεκινά με εικόνα εισόδου διαστάσεων 448×448×3 (RGB). Η επιλογή της συγκεκριμένης ανάλυσης δεν είναι τυχαία το μεγαλύτερο μέγεθος, σε σχέση με συνήθεις εισόδους ταξινόμησης όπως τα 224×224, επιτρέπει την καλύτερη διατήρηση χωρικής πληροφορίας και, κατά συνέπεια, ακριβέστερο εντοπισμό μικρών αντικειμένων. Ακολουθεί η συνελικτική ενότητα όπου κατά το πρώτο συνελικτικό στρώμα χρησιμοποιεί φίλτρα 7×7 με 64 κανάλια και βήμα (stride) 2 (Conv Layer: 7×7 kernel, 64 Filters, stride 2), ενώ αμέσως μετά εφαρμόζεται μέγιστη υπό δειγματοληψία (MaxPooling) 2×2 με stride 2. Ο συνδυασμός αυτός μειώνει τις διαστάσεις από 448×448 σε 112×112 και αρχίζει να ανιχνεύει πρωτογενή μοτίβα όπως ακμές και απλά σχήματα. Ακολουθεί δεύτερο συνελικτικό στρώμα με φίλτρα 3×3 και 192 κανάλια, καθώς και νέο MaxPooling 2×2 (stride 2), με αποτέλεσμα η αναπαράσταση να γίνεται 56×56×192. Τα επόμενα συνελικτικά μπλοκ (3–5) οργανώνονται ως εναλλαγές συνελίξεων 1×1 και 3×3, με στόχο την εξισορρόπηση υπολογιστικού κόστους και χωρικής κατανόησης. Οι συνελίξεις 1×1 λειτουργούν ως συμπιεστές του χώρου χαρακτηριστικών, επιτρέποντας τη βαθύτερη επεξεργασία χωρίς υπερβολική αύξηση παραμέτρων, ενώ οι 3×3 εμπλουτίζουν τη χωρική πληροφορία. Σε κάθε μπλοκ αυξάνεται προοδευτικά ο αριθμός των καναλιών (128 → 256 → 512 → 1024) και παρεμβάλλεται MaxPooling, με αποτέλεσμα οι χωρικές διαστάσεις να συρρικνώνονται σταδιακά σε 56×56 → 28×28 → 14×14 → 7×7. Το καταληκτικό συνελικτικό μπλοκ αποδίδει έναν χάρτη χαρακτηριστικών 7×7×1024, δηλαδή μια συμπυκνωμένη αλλά πλούσια αναπαράσταση της εικόνας, πάνω στην οποία στηρίζονται οι μεταγενέστερες προβλέψεις εντοπισμού. Μετά την ολοκλήρωση της συνελικτικής ενότητας, η έξοδος 7×7×1024 = 50.176 τιμών επιπεδοποιείται (flatten) και τροφοδοτείται σε ένα πλήρως συνδεδεμένο στρώμα με 4096 νευρώνες. Το στρώμα αυτό λειτουργεί ως συνδετικός κρίκος ανάμεσα στα εξαγόμενα χωρικά χαρακτηριστικά και την τελική πρόβλεψη, συνδυάζοντας πληροφορίες από όλο το πεδίο της εικόνας. Στη συνέχεια, η αναπαράσταση περνά σε ένα δεύτερο πλήρως συνδεδεμένο στρώμα εξόδου, το οποίο έχει μέγεθος S×S×(B×5 + C), (S = 7 είναι το μέγεθος του πλέγματος της εικόνας (7×7), B = 2 είναι ο αριθμός των bounding boxes που προβλέπονται ανά κελί, C είναι ο αριθμός των κατηγοριών (π.χ. 20 για το dataset PASCAL VOC)). Συνεπώς, ο αριθμός των εξόδων είναι 7X7X(2X5+20)=7X7X30=1470 το μέγεθος φαίνεται και στο τελευταίο τμήμα του σχήματος (Conv → FC → 30), όπου τα «30» αντιστοιχούν στα (B×5 + C) ανά κελί. Για παράδειγμα εικόνα εισόδου διαστάσεων 448x448 στην οποία απεικονίζονται τρία αντικείμενα-στόχοι, ένα αυτοκίνητο, ένας πεζός και ένα ποδήλατο. Το μοντέλο διαμερίζει την είσοδο σε πλέγμα S x S με S=7 και εκχωρεί σε κάθε κελί την ευθύνη ανίχνευσης αντικειμένων των οποίων το κέντρο εμπίπτει εντός του κελιού. Για κάθε κελί, το YOLO παράγει B=2 προβλέψεις πλαισίων (bounding boxes) μαζί με βαθμολογίες βεβαιότητας (confidence scores) και πιθανότητες εκτιμήσεις της κάθε κατηγορίας. Στο κελί που περιέχει το κέντρο του αυτοκινήτου, οι δύο

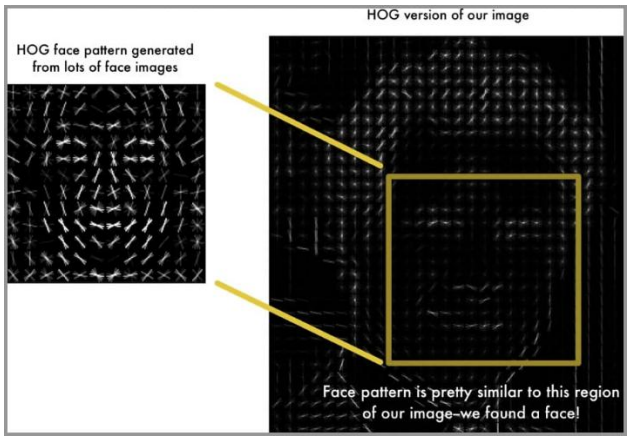
προβλέψεις αξιολογούνται ως προς το Intersection over Union (IoU) με το αντίστοιχο ground truth πλαίσιο όπου επιλέγεται εκείνη με τη μέγιστη τιμή, π.χ. IoU=0.8. Η επιλεγμένη πρόβλεψη παρέχει τις παραμετροποιημένες συντεταγμένες (x,y,w,h)—με x,y σχετικές ως προς το κελί και w,h κανονικοποιημένες ως προς το μέγεθος της εικόνας καθώς και υψηλό confidence (ενδεικτικά 0,9). Η πιθανότητα κατηγορίας για το αυτοκίνητο μεγιστοποιείται, υποδεικνύοντας ορθή ταξινόμηση. Ανάλογη διαδικασία λαμβάνει χώρα για τον πεζό και το ποδήλατο στα κελιά όπου βρίσκονται τα κέντρα τους. Συχνά ένας από τους δύο predictors «ειδικεύεται» εμπειρικά σε μικρότερα αντικείμενα, επιτυγχάνοντας υψηλότερο IoU. Μετά τη συγκέντρωση όλων των υποψηφίων πλαισίων από τα 49 κελιά, εφαρμόζεται Non-Maximum Suppression (NMS) ανά κατηγορία, με κατώφλια σε confidence/IoU, ώστε να εξαλειφθούν πλεονάζουσες επικαλύψεις και να διατηρηθούν οι πλέον αξιόπιστες ανιχνεύσεις. Το τελικό αποτέλεσμα συνίσταται σε τρεις καθαρές προβλέψεις μία για κάθε αντικείμενο (αυτοκίνητο, πεζός, ποδήλατο)οι οποίες αποδίδουν ταυτόχρονα χωρική εντόπιση και ετικέτα κατηγοριοποίησης.

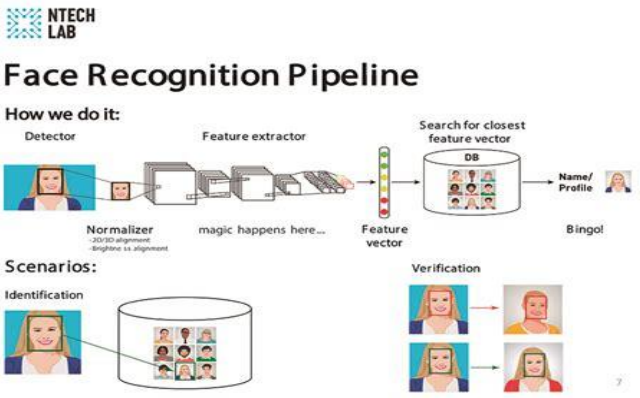
2.9 Βιβλιοθήκες Υπολογιστικής Όρασης

Στο παρόν κεφάλαιο εξετάζονται διεξοδικά οι δύο θεμελιώδεις βιβλιοθήκες στις οποίες ερείδεται η παρούσα διπλωματική εργασία, η face_recognition (dlib) και το YOLOv8. Η επιλογή τους τεκμηριώνεται από τον συμπληρωματικό χαρακτήρα των λειτουργιών τους αφ’ ενός η face_recognition (dlib) παρέχει ώριμο και ευρέως αποδεκτό σύνολο μεθόδων για την επεξεργασία μιας εικόνας, αφ’ ετέρου το YOLOv8 ενσωματώνει σύγχρονες αρχές βαθιάς μάθησης για αποδοτική και αξιόπιστη ανίχνευση αντικειμένων σε πραγματικό χρόνο. Ειδικότερα, η face_recognition (dlib) αξιοποιήθηκε στα κρίσιμα στάδια προ επεξεργασίας και μετά επεξεργασίας, εφαρμόζοντας σύμφωνα με το κεφάλαιο 2.2 της παρούσας εργασίας (α) χωρικό γραμμικό φιλτράρισμα Gaussian για αποθορυβοποίηση-εξομάλυνση επί του μειωμένης κλίμακας καρέ, (β) στοχευμένες μετατροπές χρωματικού χώρου για δια λειτουργικότητα με το σύστημα αναγνώρισης προσώπων και για τοπική επεξεργασία περιοχών ενδιαφέροντος, (γ) ενίσχυση τοπικής αντίθεσης μέσω CLAHE για ανάδειξη λεπτομερειών, και (δ) οπτικοποίηση αποτελεσμάτων με σχεδίαση ορθογώνιων πλαισίων και εμφάνισης μηνυμάτων. Αντίστοιχα, το YOLOv8 χρησιμοποιήθηκε ως κύριος μηχανισμός ανίχνευσης, υλοποιώντας μονοσταδιακή (one-stage) ροή πρόβλεψης (βλ. Κεφ. 2.4). Στην παρούσα υλοποίηση αξιοποιείται προ εκπαιδευμένο μοντέλο YOLOv8n-face (Ultralytics), με κατώφλι εμπιστοσύνης κατά την εκτέλεση (conf ≥ 0.5). Οι τελικές προβλέψεις επιστρέφουν ορθογώνια πλαίσια (bounding boxes) και βαθμολογίες εμπιστοσύνης, ενώ σε συνεργασία με το ArcFace (για σκοπούς ταυτοποίησης) πραγματοποιείται εξαγωγή χαρακτηριστικών (embeddings) και ταυτοποίηση προσώπων.

2.9.1 Face_recognition (dlib) ως Βιβλιοθήκη Υπολογιστικής Όρασης

Ο λειτουργικός ρόλος της βιβλιοθήκης face_recognition (dlib) είναι σαφώς ορισμένος μετά την προ επεξεργασία των καρέ, αναλαμβάνοντας (α) τον εντοπισμό περιοχών προσώπου και (β) την κωδικοποίηση-ταυτοποίηση των αντίστοιχων απεικονίσεων. Ο εντοπισμός υλοποιείται με τη συνάρτηση face_locations. Δεδομένου ότι στον παρόντα κώδικα η συνάρτηση καλείται χωρίς ρητό ορισμό μοντέλου, χρησιμοποιείται η προεπιλεγμένη μέθοδος HOG (Histogram of Oriented Gradients). Η επιλογή αυτή είναι σκόπιμη, καθώς ο «εντοπιστής» HOG απαιτεί ελάχιστη ισχύς και επαρκή ευστάθεια της CPU (Central Processing Unit), επιτρέποντας επεξεργασία σε πραγματικό χρόνο σε τυπικές υπολογιστικές διαμορφώσεις (π.χ. φορητός υπολογιστής), χωρίς εμφανή υστέρηση στην απόδοση των πλαισίων εντοπισμού. Για κάθε ανιχνευμένη περιοχή προσώπου, η face encodings παράγει ένα διάνυσμα χαρακτηριστικών 128 διαστάσεων (embedding), το οποίο λειτουργεί ως συμπαγής αναπαράσταση ταυτότητας. Τα παραγόμενα embeddings συγκρίνονται με τα προϋπάρχοντα embeddings αναφοράς (τα οποία έχουν εξαχθεί εκ των προτέρων από δομημένο αποθετήριο ανά άτομο). Η σύγκριση βασίζεται στην ευκλείδεια απόσταση δηλαδή οι μικρότερες τιμές υποδηλώνουν μεγαλύτερη ομοιότητα. Η τελική απόφαση λαμβάνεται με κανόνα πλησιέστερου γείτονα υπό σταθερό κατώφλι ανοχής 0,60. Ενδεικτικά, τιμές περί το 0,38 αντιστοιχούν συνήθως σε ισχυρή ταύτιση, τιμές 0,57 χαρακτηρίζονται οριακές, ενώ τιμές 0,60 οδηγούν σε απόδοση της ετικέτας Unknown. Με τον τρόπο αυτό επιτυγχάνεται ελεγχόμενη ισορροπία μεταξύ ψευδώς θετικών και ψευδώς αρνητικών αναγνωρίσεων, χωρίς την ανάγκη εισαγωγής πρόσθετου, πιο σύνθετου ταξινομητή. Οι πρακτικές παράμετροι βελτιστοποίησης εντάσσονται στο ίδιο πλαίσιο. Όταν τα πρόσωπα εμφανίζονται σε μικρή κλίμακα (π.χ. μεγάλες αποστάσεις από την κάμερα), η ικανότητα εντοπισμού βελτιώνεται με αύξηση του Number of times to upsample, με αποτέλεσμα μια μικρή επιβάρυνση στον χρόνο επεξεργασίας ανά καρέ. Σε δύσκολες συνθήκες φωτισμού (υπό φωτισμός, έντονες σκιές), η προ επεξεργασία με ήπια ομαλοποίηση και ορθή μετατροπή σε RGB συντελεί στη σταθερότητα των embeddings. Τέλος, η ποιότητα και η ποικιλία του συνόλου στο δομημένο αποθετήριο ανά άτομο είναι κρίσιμες. Από τρείς έως δέκα καθαρές, μετωπικές απεικονίσεις ανά άτομο, σε διαφορετικές συνθήκες φωτισμού και έκφρασης, βελτιώνουν σημαντικά τη γενίκευση και τη συνεκτικότητα των αποφάσεων.
Συνολικά, η βιβλιοθήκη face_recognition παρέχει έναν αξιόπιστο και αποδοτικό μηχανισμό εντοπισμού (μέσω HOG) και ταυτοποίησης (μέσω 128-D embeddings και κατωφλίου απόστασης), ο οποίος είναι διαφανής ως προς τις υποκείμενες επιλογές παραμέτρων και επαρκώς αποδοτικός για εφαρμογές πραγματικού χρόνου σε συμβατικά υπολογιστικά περιβάλλοντα.





2.9.2 YOLOv8 – Μια σύγχρονη προσέγγιση στην ανίχνευση Προσώπων

Το YOLOv8 είναι η νεότερη και πιο εξελιγμένη έκδοση της οικογένειας YOLO. κυκλοφόρησε τον Ιανουάριο του 2023 από την Ultralytics LLC και από τότε εξελίσσεται διαρκώς, με συχνές ανανεώσεις και μια ιδιαίτερα δραστήρια κοινότητα που δοκιμάζει, συγκρίνει και προτείνει βελτιώσεις. Μέχρι στιγμής, η βασική τεκμηρίωση προέρχεται από το επίσημο αποθετήριο της Ultralytics στο GitHub (https://github.com/ultralytics/ultralytics), καθώς δεν έχει δημοσιευθεί ακόμη σχετικό άρθρο σε επιστημονικό περιοδικό με κριτές. Ως προς την αρχιτεκτονική, το YOLOv8 οργανώνεται σε δύο κύρια μέρη. Πρώτον, το backbone όπου αποτελείται από μια ακολουθία συνελικτικών στρωμάτων που εξάγουν χαρακτηριστικά της εικόνας σε πολλαπλές κλίμακες και συμπυκνώνουν την πληροφορία σε διαφορετικά επίπεδα ανάλυσης. Αυτό επιτρέπει στο σύστημα να εντοπίζει αντικείμενα πολύ μικρά αλλά και πολύ μεγάλα, χωρίς να χάνει κρίσιμες λεπτομέρειες. Δεύτερον, το head του YOLOv8 όπου το ανασχεδιασμένο Head συνδυάζει τα χαρακτηριστικά από το backbone και παράγει τις τελικές προβλέψεις. Σε αυτό το στάδιο εφαρμόζονται βελτιωμένες συναρτήσεις απώλειας για τα bounding boxes, τις κατηγορίες και το objectness (δηλαδή την πιθανότητα ότι ένα πλαίσιο όντως περιέχει αντικείμενο), με στόχο μεγαλύτερη ακρίβεια και σταθερότερη εκπαίδευση.

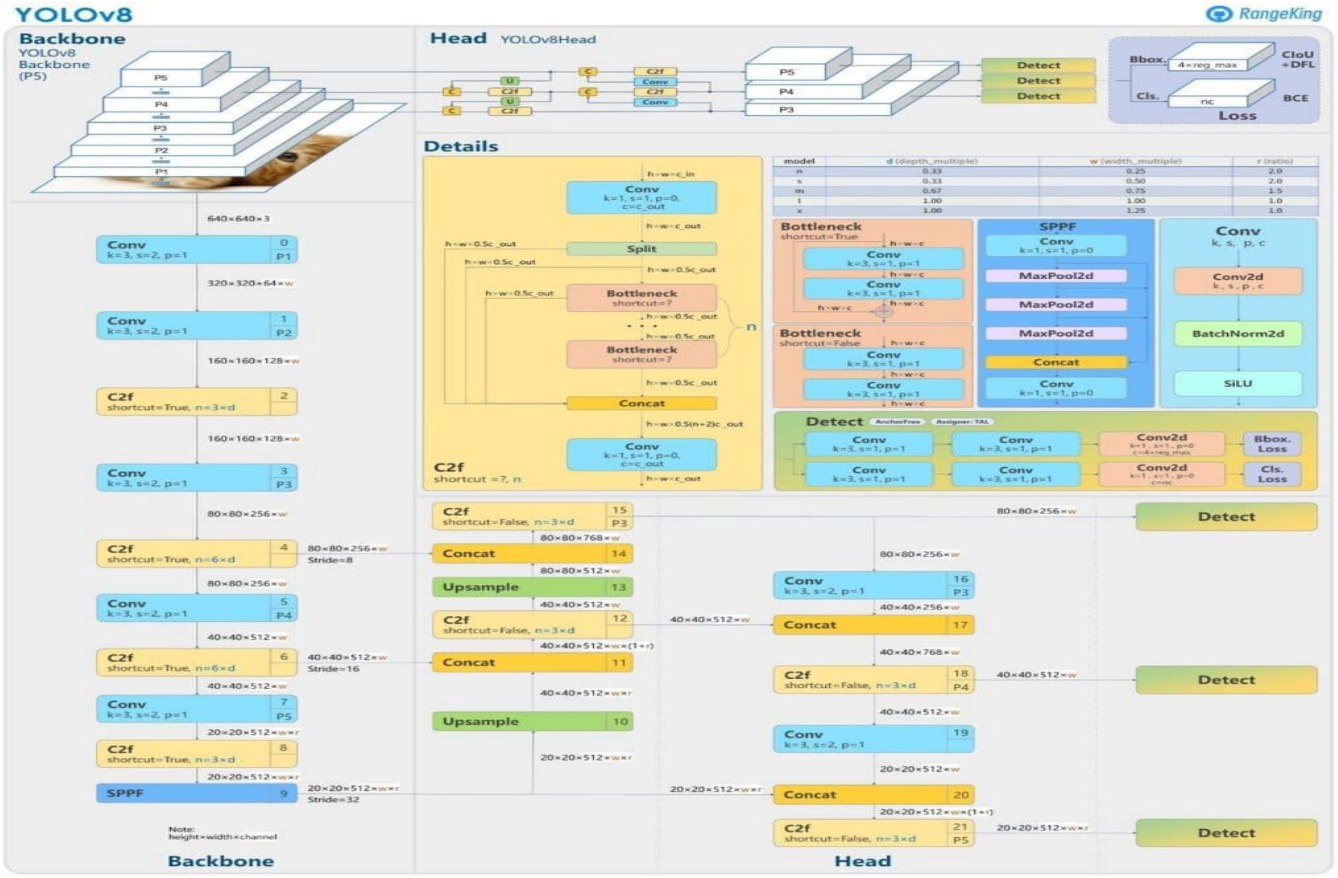
Βασικές καινοτομίες και βελτιώσεις του YOLOv8

Κεντρική επιλογή του YOLOv8 είναι η μετάβαση σε anchor-free head. Σε αντίθεση με παλαιότερες εκδόσεις που στηρίζονταν σε anchor boxes, εδώ οι προβλέψεις δεν εξαρτώνται από προκαθορισμένα άγκιστρα (anchor boxes*)*. Με τον όρο προκαθορισμένα άγκιστρα (anchor boxes) αναφερόμαστε σε ορθογώνια πλαίσια-πρότυπα με συγκεκριμένα μεγέθη, τα οποία τοποθετούνται προληπτικά σε κάθε κελί του πλέγματος της εικόνας και λειτουργούν ως αφετηρίες για την πρόβλεψη των τελικών bounding boxes. Τα anchor-based μοντέλα μαθαίνουν διορθώσεις πάνω σε αυτά τα πρότυπα, γεγονός που τα καθιστά ευαίσθητα στην επιλογή των αρχικών μεγεθών-αναλογιών. Στο YOLOv8, η μετάβαση σε anchor-free head αίρει αυτή την εξάρτηση με αποτέλεσμα το μοντέλο να προβλέπει άμεσα τη θέση και το μέγεθος του πλαισίου, βελτιώνοντας την προσαρμοστικότητα σε αντικείμενα με ασυνήθιστες αναλογίες (πολύ ψηλές και στενές κολώνες φωτισμού) και ενισχύοντας την ανίχνευση πολύ μικρών στόχων. Επιπλέον, η απουσία anchors απλοποιεί και επιταχύνει το στάδιο Non-

Maximum Suppression, με θετικό αντίκτυπο στην ταχύτητα και στη συνολική ακρίβεια. Ενδεικτικά, σε σκηνές με μεγάλη ποικιλία μεγεθών και μορφών (π.χ. λήψεις από Drone με ανθρώπους, οχήματα και κτίρια στο ίδιο καρέ), η anchor-free προσέγγιση προσφέρει πιο σταθερή απόδοση σε σχέση με λύσεις που εξαρτώνται από λίστες προεπιλεγμένων κουτιών-προτύπων. Αυτό το καθιστά πιο ευέλικτο όταν τα αντικείμενα έχουν ασυνήθιστες αναλογίες (π.χ. πολύ μεγάλο ή πολύ μικρό ύψος σε σχέση με το πλάτος) και, πρακτικά, απλοποιεί και επιταχύνει το στάδιο Non-Maximum Suppression (NMS), με θετική επίδραση τόσο στην ταχύτητα όσο και στην ακρίβεια. Παράλληλα, η εσωτερική αρχιτεκτονική έχει ανανεωθεί με πιο σύγχρονα συνελικτικά δομικά στοιχεία. Η αλλαγή αυτή δεν αποσκοπεί σε αύξηση της πολυπλοκότητας, αλλά σε ουσιαστικότερη εξαγωγή χαρακτηριστικών με μετρημένο υπολογιστικό κόστος—δηλαδή καλύτερη πληροφορία προς το head, χωρίς την αύξηση των απαιτήσεων. Κατά την εκπαίδευση, το YOLOv8 εφαρμόζει τη στρατηγική mosaic augmentation, όπου συνδυάζονται τέσσερις διαφορετικές εικόνες σε μία. Αυτό αναγκάζει το μοντέλο να μάθει να ανιχνεύει αντικείμενα σε νέα περιβάλλοντα, υπό μερική απόκρυψη (occlusion) ή σε συνθήκες πολύπλοκου φόντου, αυξάνοντας τη γενίκευση και την ανθεκτικότητα του δικτύου.

Απόδοση και αξιολόγηση

Σε καθιερωμένα σύνολα δεδομένων καταγράφονται ανταγωνιστικές επιδόσεις. Σύμφωνα με τον Jacob Solawetz και τα επίσημα benchmarks της Ultralytics, στο COCO αναφέρονται τιμές mAP γύρω στο 53,9%, υψηλότερες από προηγούμενες εκδόσεις της οικογένειας YOLO, ενώ στο Roboflow-100 (100 υποσύνολα, 224.714 εικόνες) το YOLOv8 εμφανίζει ιδιαίτερα καλή συμπεριφορά σε κατηγορίες όπως τα “documents” και σε ποικίλες εφαρμογές ανίχνευσης-ταξινόμησης. Συνολικά, ο συνδυασμός anchor-free σχεδιασμού, ανανεωμένων συνελικτικών δομών και σύγχρονων στρατηγικών εκπαίδευσης (π.χ. mosaic augmentation) εξηγεί την ισχυρή του παρουσία σε πληθώρα datasets και πραγματικών εφαρμογών. Η ταχύτητα σε πραγματικό χρόνο, η προσαρμοστικότητα σε πολύπλοκα δεδομένα και η ενεργή του υποστήριξη καθιστούν το YOLOv8 τεκμηριωμένα ασφαλή επιλογή για σύγχρονες υλοποιήσεις υπολογιστικής όρασης.



Επεξήγηση Διαγράμματος

- **Backbone (YOLOv8 Backbone):** Περιλαμβάνει διαδοχικά conv και C2f layers που επεξεργάζονται την εικόνα σε διαδοχικά επίπεδα ανάλυσης, εξάγοντας χαρακτηριστικά από διαφορετικές κλίμακες.
- **SPPF και άλλες μονάδες (Details):** Περιλαμβάνουν δομές όπως Bottleneck, Concat, SPPF, που αυξάνουν τη δυνατότητα του μοντέλου να μαθαίνει σύνθετα χαρακτηριστικά.
- **YOLOv8 Head:** Παίρνει ως είσοδο τα χαρακτηριστικά από τα διάφορα επίπεδα του backbone (P3, P4, P5) και πραγματοποιεί την τελική ανίχνευση (Detect layers).
- **Loss Functions:** Το loss περιλαμβάνει BCE (Binary Cross Entropy) για κατηγορία (CLS), objectness (OBJ) και bounding box loss (BBox).

2.10 Βιβλιοθήκες Αναγνώρισης προσώπων

Την αναγνώριση προσώπου μπορούμε να τη σκεφτούμε σαν μια απλή διαδικασία ταύτισης, πρώτα εντοπίζουμε πού υπάρχει πρόσωπο στο καρέ και στη συνέχεια προσπαθούμε να απαντήσουμε στην ερώτηση «ποιος είναι;». Η παρούσα εργασία αξιοποιεί ένα σύνολο βιβλιοθηκών ανοιχτού κώδικα για την υλοποίηση του συστήματος αναγνώρισης προσώπων. Υιοθετούνται δύο «αρχιτεκτονικές» προσεγγίσεις, οι οποίες μοιράζονται την ίδια υποδομή όσων αφορά την είσοδο των δεδομένων με OpenCV αλλά και κατά την εμφάνιση των αποτελεσμάτων με NumPy, διαφέρουν όμως ως προς τον τρόπο με τον οποίο αντλούνται τα χαρακτηριστικά του προσώπου (embeddings) και λαμβάνεται η τελική απόφαση. Η πρώτη προσέγγιση υλοποιείται με έναν ενιαίο τρόπο

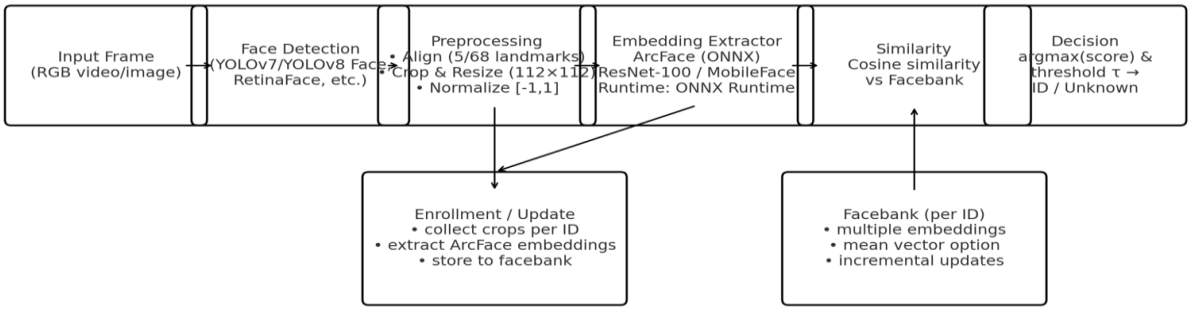
αξιοποιώντας τη βιβλιοθήκη face_recognition στην ουσία ο ίδιος πυρήνας (dlib) παρέχει ανίχνευση προσώπου, εξαγωγή τα embeddings 128 διαστάσεων και στο τέλος συσχετίζει τα πρόσωπα στα οποία έγινε η ανίχνευση με τα γνωστά πρόσωπα τα οποία υπάρχουν στο αρχείο του χρήστη. Από την άλλη μεριά η δεύτερη προσέγγιση είναι μια πιο σύνθετη διαδικασία κατά την οποία αξιοποιούνται δύο βιβλιοθήκες η YOLOv8-face για ανίχνευση και η ArcFace για την εξαγωγή embeddings μέσω ONNX Runtime. Η απόφαση βασίζεται σε ομοιότητα συνημίτονού (cosine similarity) έναντι μιας «τράπεζας προσώπων» (facebank).

2.10.1 Βιβλιοθήκη face_recognition (dlib)

Στη πρώτη προσέγγιση, η face_recognition λειτουργεί ως μία ενιαία λύση για ανίχνευση και αναγνώριση προσώπων. Η «εντολή» face_locations εντοπίζει τα πρόσωπα στο καρέ, ενώ η face encodings εξάγει για κάθε εντοπισμό ένα αποτύπωμα (embedding) 128 διαστάσεων. Η σύγκριση με το σύνολο αναφοράς, το οποίο είναι προ-οργανωμένο σε φακέλους ανά ταυτότητα, πραγματοποιείται με τις συναρτήσεις face_distances και compare_faces. Η τελική απόφαση αντιστοιχεί ουσιαστικά σε ταξινόμηση όπου επιλέγεται το ελάχιστο της ευκλείδειας απόστασης και εφαρμόζεται κατώφλι ανοχής (ενδεικτικά, 0,60). Αποστάσεις κάτω από το κατώφλι παράγουν θετική ταυτοποίηση ενώ αποστάσεις άνω του κατωφλίου ταξινομούνται ως Unknown. Η συγκεκριμένη βιβλιοθήκη επιτρέπει λειτουργική υλοποίηση ακόμη και σε CPU, με ελάχιστη παραμετροποίηση και ταχεία ενσωμάτωση στον κώδικα.

2.10.2 ArcFace (Onnxruntime)

Στη δεύτερη προσέγγιση, η εξαγωγή των embeddings πραγματοποιείται μέσω ArcFace σε μορφή ONNX, με εκτέλεση από το Onnxruntime. Πριν την προώθηση στο μοντέλο, η αποκοπή προσώπου (crop) δειγματοληπτείται στην απαιτούμενη ανάλυση και κανονικοποιείται. Το μοντέλο επιστρέφει ένα κανονικοποιημένο διάνυσμα χαρακτηριστικών, επί του οποίου υπολογίζεται το συνημίτονο ομοιότητας (cosine similarity) έναντι των αναφορών που διατηρούνται στο facebank (τράπεζα προσώπων). Για κάθε ταυτότητα μπορούν να αποθηκεύονται πολλαπλά embeddings ή, εναλλακτικά, ο μέσος όρος τους. Η απόφαση λαμβάνεται με βάση το μέγιστο σκορ ομοιότητας και την υπέρβαση ενός κατωφλίου. Τιμές κάτω από το όριο του κατωφλίου χαρακτηρίζονται ως “Unknown”. Η παραπάνω διάταξη αποδίδει ιδιαίτερα όταν το σύνολο ταυτοτήτων διευρύνεται ή όταν οι συνθήκες λήψης είναι απαιτητικές.



Ανακεφαλαίωση

Η OpenCV αποτελεί βιβλιοθήκη υπολογιστικής όρασης γενικού σκοπού, κατάλληλη για επεξεργασία και ανάλυση εικόνας/βίντεο. Στο πλαίσιο της παρούσας πτυχιακής εργασίας, η OpenCV χρησιμοποιείται για τη διαχείριση των πηγών (είσοδο/έξοδο από κάμερες και αρχεία βίντεο), προ επεξεργασία των καρέ, ενίσχυση τοπικής αντίθεσης και απεικόνιση των τελικών αποτελεσμάτων. Η ανίχνευση θέσης και η αναγνώριση ταυτότητας προσώπων υλοποιούνται σε συνεργασία με τη βιβλιοθήκη Face recognition, η οποία παράγει διανύσματα χαρακτηριστικών (encodings) και υποστηρίζει τη σύγκρισή τους με σύνολο «γνωστών» προσώπων. Από την άλλη μεριά το YOLO (You Only Look Once) αποτελεί μονοσταδιακό ανιχνευτή αντικειμένων (one-stage Detector) και παρέχει, σε πραγματικό χρόνο, περιγράμματα εντοπισμού (bounding boxes), και βαθμό βεβαιότητας (confidence) για πολλαπλά αντικείμενα σε μία εικόνα ή ακολουθία καρέ. Για τη ταυτοποίηση προσώπου, το YOLO συνδυάζεται με μοντέλα embeddings (π.χ. ArcFace). Στην παρούσα υλοποίηση αξιοποιείται το YOLOv8-face για ανίχνευση και το ArcFace (ONNX) για εξαγωγή embeddings και τελική ταυτοποίηση, με ρυθμιζόμενο κατώφλι ομοιότητας (threshold) και facebank για αποθήκευση αναφορών. Μία από τις κυριότερες διαφορές μεταξύ της Βιβλιοθήκης Face recognition και του YOLO είναι η αποδοτικότητα διότι το πρώτο κάνει χρήση CPU ενώ το δεύτερο GPU.

Κεφάλαιο 3: Σχετικά Υφιστάμενα Εργαλεία και Παρόμοιες Εργασίες

3.1 Παρόμοιες Επιστημονικές Εργασίες

Η πτυχιακή εργασία του Σ. Κανελλόπουλου, με τίτλο «Αναγνώριση Προσώπου με χρήση Συνελικτικών Νευρωνικών Δικτύων» (Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών, 2019), αποτελεί χαρακτηριστικό παράδειγμα εφαρμογής προ εκπαιδευμένων συνελικτικών νευρωνικών δικτύων (CNNs) στο πρόβλημα της αναγνώρισης προσώπων. Στην εργασία αξιοποιούνται αρχιτεκτονικές όπως InceptionV3, VGG16, AlexNet και ResNet50, ενώ για την υλοποίηση χρησιμοποιείται η βιβλιοθήκη Keras σε συνδυασμό με προ εκπαιδευμένα βάρη από το ImageNet. Αυτό επιτρέπει την ταχύτερη ανάπτυξη του μοντέλου μέσω τεχνικών Transfer Learning και Fine Tuning. Ως σύνολο δεδομένων επιλέχθηκε το Labeled Faces in the Wild (LFW), που περιλαμβάνει πάνω από 13.000 εικόνες. Το προ επεξεργαστικό στάδιο περιλάμβανε ευθυγράμμιση των προσώπων με βάση τα μάτια και το στόμα (align dlib), μετασχηματισμό των εικόνων, ενίσχυση δεδομένων (data augmentation) και επιλογή των 10 ατόμων με τις περισσότερες διαθέσιμες εικόνες. Το dataset χωρίστηκε σε 70% για εκπαίδευση, 15% για επικύρωση και 15% για δοκιμή, ενώ χρησιμοποιήθηκε cross-validation και παρακολούθηση της ακρίβειας σε κάθε epoch. Το πρόβλημα της αναγνώρισης αντιμετωπίστηκε ως εποπτευόμενη ταξινόμηση πολλών κλάσεων, με το τελικό στρώμα (SoftMax layer) να προβλέπει την ταυτότητα του προσώπου. Στην εργασία περιγράφονται επίσης loss functions όπως το ArcFace και το Triplet Loss, αν και δεν εφαρμόστηκαν στην πράξη. Τα αποτελέσματα κατέδειξαν πολύ υψηλή ακρίβεια, ξεπερνώντας το 97% στο validation set,

προσφέροντας ένα ιδιαίτερα χρήσιμο παράδειγμα εφαρμογής τεχνικών deep Learning σε στατικό σύνολο δεδομένων εικόνων.

Ένα άλλο πρόσφατο παράδειγμα αποτελεί η εργασία των Anjeana & Anusudha (2024) με τίτλο *"*Real time face recognition System based on YOLO and InsightFace*"* που δημοσιεύθηκε στο περιοδικό Multimedia Tools and Applications, 19 September 2023 . Στην εργασία αυτή προτείνεται ένα σύστημα αναγνώρισης προσώπων σε πραγματικό χρόνο, το οποίο συνδυάζει το YOLOv7 για την ανίχνευση προσώπων και το InsightFace για την εξαγωγή χαρακτηριστικών και την αναγνώριση. Το μοντέλο παρουσιάζει ικανοποιητικές επιδόσεις σε περιβάλλοντα με καλές συνθήκες φωτισμού και ορατότητας, αναδεικνύοντας τη δυνατότητα αξιοποίησης των Deep Learning τεχνικών σε πραγματικές εφαρμογές. Οι περιορισμοί που αναφέρονται από τους συγγραφείς, παρότι το σύστημα λειτουργεί σε πραγματικό χρόνο και δείχνει υψηλή ακρίβεια, οι ερευνητές υπογραμμίζουν ότι η αναγνώριση προσώπων σε περιπτώσεις μερικής απόκρυψης (π.χ. μάσκες, γυαλιά) παραμένει ιδιαίτερα δύσκολη και δεν έχει αντιμετωπιστεί αποτελεσματικά. Επιπλέον, δεν γίνεται αναφορά σε σενάρια ανοιχτού συνόλου ταυτοτήτων, όπου το σύστημα πρέπει να αναγνωρίζει άγνωστα πρόσωπα, ούτε εξετάζεται η δυνατότητα δυναμικής ενημέρωσης της βάσης δεδομένων προσώπων (facebank). Τέλος, η έρευνα επικεντρώνεται σε μία πηγή εισόδου χωρίς να αξιολογεί την επίδοση σε περιβάλλοντα με πολλαπλές κάμερες ή ροές βίντεο.

Ένα ακόμη πρόσφατο παράδειγμα αποτελεί η εργασία των Imran, Ahmed, Hasan, Ahmed & Alyami (2024) με τίτλο Face Engine: A Tracking-Based Framework for Real-Time Face Recognition in Video Surveillance System (SN Computer Science, 2024). Στη μελέτη αυτή προτείνεται ένα πλαίσιο πραγματικού χρόνου αναγνώρισης προσώπων, ειδικά σχεδιασμένο για εφαρμογές επιτήρησης, το οποίο ενσωματώνει τεχνικές παρακολούθησης (tracking) και χρονικής ψηφοφορίας (temporal voting). Ο συνδυασμός αυτός επιτρέπει την αξιόπιστη ταυτοποίηση προσώπων σε συνθήκες κίνησης και μεταβαλλόμενου φωτισμού. Κεντρικό στοιχείο της μελέτης αποτελεί η διαδικασία βελτιστοποίησης εισόδου (Input optimization), η οποία περιλαμβάνει κανονικοποίηση και ευθυγράμμιση των προσώπων, φιλτράρισμα εικόνων χαμηλής ποιότητας ή με πολλαπλά άτομα, καθώς και επιλογή αντιπροσωπευτικών δειγμάτων με διαφορετικές εκφράσεις, γωνίες και συνθήκες φωτισμού. Μέσω αυτής της προσέγγισης μειώνεται η ανάγκη για μεγάλο αριθμό εικόνων ανά άτομο, αφού το σύστημα μπορεί να επιτύχει υψηλή απόδοση ακόμη και με 5–10 εικόνες ανά άτομο-ταυτότητα, περιορίζοντας τον κίνδυνο υπερπροσαρμογής και ενισχύοντας τη γενίκευση σε νέες συνθήκες. Παράλληλα, εισάγεται η έννοια των προσαρμοστικών κατωφλίων (Adaptive threshold), τα οποία ρυθμίζονται δυναμικά ανάλογα με την ποιότητα των εισερχόμενων εικόνων, τη σταθερότητα των αποτελεσμάτων σε διαδοχικά καρέ και την πολυπλοκότητα της βάσης δεδομένων. Έτσι, το σύστημα μπορεί να είναι πιο «αυστηρό» σε περιβάλλοντα με υψηλή ανάλυση και καλή ευκρίνεια, αλλά και πιο «ευέλικτο» σε περιπτώσεις χαμηλής ποιότητας ή ταχύτατης κίνησης, περιορίζοντας τις λανθασμένες θετικές αναγνωρίσεις και αυξάνοντας την αξιοπιστία. Τα αποτελέσματα των πειραμάτων καταδεικνύουν ότι η χρήση των τεχνικών αυτών οδήγησε σε αύξηση της ακρίβειας από 86,36% σε 90,91%, ενώ σε συγκριτική αξιολόγηση το FaceEngine υπερείχε σαφώς έναντι εναλλακτικών μεθόδων όπως το OpenCV+Fisher LDA (82%) και το YOLOv3+Dlib (86,3%), επιτυγχάνοντας βελτίωση κατά 10,87% και 5,34% αντίστοιχα. Παράλληλα, οι δοκιμές με «θολές» εικόνες ή με «παρείσακτα»

δεδομένα ανέδειξαν τη σημασία της βελτιστοποίησης και της δυναμικής ρύθμισης των thresholds, καθώς χωρίς αυτές η ακρίβεια υποχωρούσε έως και στο 72,73%.

Επίσης ακόμη ένα πρόσφατο παράδειγμα αποτελεί η δημοσίευση των Vasavi, Bora, Murali, Jacob & Raj (2025) με τίτλο Robust Face Recognition System for Extensive Distances in Uncontrolled Environments on Edge Device (PowerTech Journal, Vol. 49, No. 2). Η μελέτη αυτή εστιάζει στην πρόκληση της αναγνώρισης προσώπων σε μεσαίες και μεγάλες αποστάσεις (αναγνώριση από Drones), όπου παραδοσιακές μέθοδοι εμφανίζουν σημαντικούς περιορισμούς. Οι συγγραφείς προτείνουν μια ενισχυμένη αρχιτεκτονική βασισμένη στον αλγόριθμο MTCNN για την ανίχνευση προσώπων και στο ArcFace με backbone ResNet-50 για την εξαγωγή και αναγνώριση χαρακτηριστικών, αξιοποιώντας margin-based loss functions για μεγαλύτερη διαχωριστικότητα μεταξύ κατηγοριών. Η εκπαίδευση πραγματοποιήθηκε με χειροκίνητα συλλεγμένες εικόνες σε ποικίλες συνθήκες φωτισμού, καθώς και με το ευρύ σύνολο δεδομένων Celeba (>200.000 εικόνες), ώστε να ενισχυθεί η γενικευσιμότητα του μοντέλου. Η αξιολόγηση πραγματοποιήθηκε μέσω μιας συστηματικής ανάλυσης τύπου πειραματικής διαδικασία αξιολόγησης (ablation study), όπου οι ερευνητές αφαιρούν ή τροποποιούν επιμέρους τμήματα ενός μοντέλου (π.χ. layers, attention blocks, loss functions) για να δουν πόσο συμβάλλει κάθε τμήμα ξεχωριστά στην τελική απόδοση, με στόχο τη διερεύνηση της συνεισφοράς κάθε επιμέρους συνιστώσας στην τελική απόδοση του συστήματος. Τα πειραματικά αποτελέσματα καταδεικνύουν ότι η αφαίρεση των dense layers συνεπάγεται μείωση της ακρίβειας κατά 3,5%, γεγονός που αναδεικνύει τον ρόλο τους στη διατήρηση λεπτομερών χωρικών χαρακτηριστικών σε σενάρια χαμηλής ανάλυσης. Η απουσία attention modules (CBAM/SE blocks) οδήγησε σε περαιτέρω πτώση της απόδοσης κατά 4,3%, επιβεβαιώνοντας τη σημασία των μηχανισμών προσοχής στην ενίσχυση των κρίσιμων περιοχών του προσώπου και στον περιορισμό του θορύβου υποβάθρου. Αντίστοιχα, η μη εφαρμογή τεχνικών upsampling (το upsampling είναι ένα βήμα προ επεξεργασίας που «βελτιώνει» την ποιότητα των μικρών ή μακρινών προσώπων ώστε το μοντέλο να τα αναγνωρίζει πιο αξιόπιστα) σε εικόνες χαμηλής ανάλυσης επέφερε μείωση της ακρίβειας κατά 7%, υπογραμμίζοντας την ανάγκη αποκατάστασης λεπτομερειών για αξιόπιστη αναγνώριση σε μεσαίες και μεγάλες αποστάσεις. Τέλος, η αντικατάσταση της ArcFace loss με έναν απλό SoftMax classifier (SoftMax είναι ο κλασικός τελικός ταξινομητής στα περισσότερα νευρωνικά δίκτυα) οδήγησε σε απώλεια απόδοσης της τάξης του 5,8%, επιβεβαιώνοντας τον καθοριστικό ρόλο των margin-based loss functions στην ενίσχυση της διαχωριστικότητας μεταξύ διαφορετικών ταυτοτήτων. Συνολικά, η ανάλυση τεκμηριώνει ότι κάθε επιμέρους τεχνική συμβάλλει ουσιαστικά στη βελτίωση της συνολικής ακρίβειας του μοντέλου, αναδεικνύοντας τη σημασία ενός ολιστικού σχεδιασμού. Τα πειραματικά αποτελέσματα τεκμηριώνουν την υπεροχή του προτεινόμενου πλαισίου (αυτό που σχεδίασαν οι Vasavi, Bora, Murali, Jacob & Raj (2025) στη δημοσίευση) σε σχέση με συμβατικές αρχιτεκτονικές (ένα απλό MTCNN + ResNet101 χωρίς τις πρόσθετες βελτιώσεις). Συγκεκριμένα, το πλήρες μοντέλο πέτυχε συνολική ακρίβεια 91,6% με F1-score 0,915, υπερβαίνοντας κατά σχεδόν 9 ποσοστιαίες μονάδες τη βασική υλοποίηση MTCNN+ResNet101 (82,4%). Σε ελεγχόμενα σενάρια αξιολόγησης, όπου ελαχιστοποιούνται οι παράγοντες θορύβου, η επίδοση ανήλθε στο 97%, καταδεικνύοντας το θεωρητικό ανώτατο όριο της προτεινόμενης αρχιτεκτονικής. Η βελτίωση αυτή δεν είναι τυχαία αλλά απορρέει από τη ενσωμάτωση επιμέρους τεχνικών όπως dense Connection (διατηρούν κρίσιμες χωρικές λεπτομέρειες), τα attention mechanisms (εστιάζουν στις διακριτικές περιοχές του προσώπου) και οι τεχνικές upsampling

(αποκαθιστούν πληροφορία σε περιπτώσεις χαμηλής ανάλυσης), ενώ η ArcFace loss διασφαλίζει σαφή διαχωριστικότητα μεταξύ διαφορετικών ταυτοτήτων μέσω margin-based εκπαίδευσης. Η συμβολή της μελέτης έγκειται στην ανάδειξη της αξίας του ολιστικού συνδυασμού αυτών των στοιχείων, καθώς κάθε συνιστώσα αποδείχθηκε καθοριστική για την τελική απόδοση.

Τέλος ένα ακόμη παράδειγμα παρόμοιο με τη παρούσα εργασία, το οποίο εστιάζει στην πρακτική αξιοποίηση της αναγνώρισης προσώπων σε εκπαιδευτικά περιβάλλοντα αποτελεί η εργασία των Faruque, Siddiqui & Noor (2023) με τίτλο Face Recognition-Based Mass Attendance Using YOLOv5 and ArcFace (Springer, LNICST, Vol. 490). Οι συγγραφείς επιχειρούν να αυτοματοποιήσουν τη διαδικασία καταγραφής παρουσιών σε αίθουσες διδασκαλίας, διερευνώντας την επίδοση διαφορετικών αλγορίθμων ανίχνευσης και αναγνώρισης προσώπων σε πραγματικές συνθήκες. Συγκεκριμένα, συγκρίνονται μέθοδοι ανίχνευσης όπως οι HaarCascade, SSD, MTCNN και YOLOv5, καθώς και μέθοδοι αναγνώρισης όπως οι LBPH, FaceNet, Deep Face και ArcFace, με δεδομένα που συλλέχθηκαν από 120 φοιτητές σε πραγματικό περιβάλλον τάξης. Η διαδικασία προ επεξεργασίας περιλάμβανε ευθυγράμμιση των προσώπων, αλλαγή μεγέθους και δημιουργία συνόλου δεδομένων με διαφορετικές συνθήκες φωτισμού, γωνίες λήψης και διαφοροποιήσεις στην εμφάνιση (π.χ. χρήση χιτζάμπ απο φοιτήτριες). Για την εκπαίδευση αξιοποιήθηκαν προ εκπαιδευμένα μοντέλα ώστε να μειωθεί ο κίνδυνος υπερπροσαρμογής δηλαδή ότι οι ερευνητές δεν εκπαίδευσαν το μοντέλο τους εξ’ αρχής αποκλειστικά στο δικό τους dataset (που ήταν σχετικά μικρό, ~120 φοιτητές με 50 εικόνες ο καθένας), αν το έκαναν το νευρωνικό δίκτυο θα μπορούσε να μάθει «πολύ καλά» τις συγκεκριμένες εικόνες των φοιτητών, αλλά να μην μπορεί να γενικεύσει σε νέα δεδομένα (δηλαδή να αποτυγχάνει όταν βλέπει νέα πρόσωπα ή νέες συνθήκες φωτισμού-γωνίας). Γι’ αυτό χρησιμοποίησαν προ εκπαιδευμένα μοντέλα (pretrained models), όπως το ArcFace που είχε ήδη εκπαιδευτεί σε μεγάλα σύνολα δεδομένων προσώπων. Έτσι, το δικό τους μοντέλο ξεκινά με «γενικές» γνώσεις για την αναγνώριση προσώπων και χρειάζεται μόνο fine-tuning στο ειδικότερο dataset της τάξης. Η πειραματική αξιολόγηση κατέδειξε ότι ο συνδυασμός YOLOv5 για την ανίχνευση και ArcFace για την αναγνώριση υπερτερούσε σημαντικά έναντι των άλλων μοντέλων, προσφέροντας ταυτόχρονα υψηλή ακρίβεια και ταχύτητα σε πραγματικές συνθήκες μαζικής παρακολούθησης. Όπως προκύπτει από τα ευρήματα, οι κλασικές μέθοδοι (HaarCascade–LBPH) υστερούν σε επίδοση, ενώ τα σύγχρονα βαθιά νευρωνικά δίκτυα (MTCNN–Deep Face, YOLOv5–ArcFace) εμφανίζουν σαφώς βελτιωμένα ποσοστά. Ειδικότερα, ο συνδυασμός YOLOv5 και ArcFace αποτελεί την πιο αποδοτική λύση, με ακρίβεια ανίχνευσης και αναγνώρισης που αγγίζει το 95% και το 94,74% αντίστοιχα, γεγονός που καθιστά το προτεινόμενο σύστημα κατάλληλο για εφαρμογές μεγάλης κλίμακας, όπως η αυτόματη παρακολούθηση παρουσιών σε τάξεις ή σεμινάρια.

3.2 Συγκριτική Ανάλυση Υλοποιήσεων

Η μελέτη του Κανελλόπουλου (2019) επικεντρώνεται στην αξιοποίηση προ εκπαιδευμένων συνελικτικών δικτύων (InceptionV3, VGG16, AlexNet, ResNet50) και τεχνικών μεταφοράς μάθησης (Transfer Learning, Fine Tuning) σε στατικό σύνολο δεδομένων (LFW). Η προσέγγιση αυτή αποδεικνύει τη δυνατότητα των CNNs να επιτυγχάνουν υψηλά ποσοστά ακρίβειας (>97%), ωστόσο περιορίζεται σε σενάρια στατικών εικόνων, χωρίς να λαμβάνει υπόψη ζητήματα πραγματικού χρόνου ή προσαρμογής σε διαφορετικά περιβάλλοντα. Οι Anjeana &

Anusudha (2024) προτείνουν ένα σύστημα πραγματικού χρόνου που συνδυάζει YOLOv7 για ανίχνευση και InsightFace για αναγνώριση. Το μοντέλο παρουσιάζει ικανοποιητική απόδοση υπό καλές συνθήκες φωτισμού, αλλά οι συγγραφείς αναγνωρίζουν περιορισμούς σε περιπτώσεις μερικής απόκρυψης και απουσίας πολλαπλών εισόδων, ενώ δεν εξετάζονται σενάρια ανοιχτών ταυτοτήτων ή δυναμικής ενημέρωσης βάσης δεδομένων. Η εργασία των Imran, Ahmed, Hasan, Ahmed & Alyami (2024) με τίτλο Face Engine εισάγει ένα πλαίσιο βασισμένο σε tracking και temporal voting, ενσωματώνοντας παράλληλα τεχνικές βελτιστοποίησης εισόδου και προσαρμοστικών κατωφλίων. Τα αποτελέσματα δείχνουν σημαντική βελτίωση στην ακρίβεια (από 86,36% σε 90,91%) και μειωμένες απαιτήσεις ως προς τον αριθμό εικόνων ανά ταυτότητα. Ωστόσο, η εφαρμογή παραμένει προσανατολισμένη σε συστήματα επιτήρησης και δεν παρέχει λύσεις για δυναμική διαχείριση βάσης προσώπων ή προσαρμογή σε περιβάλλοντα πολλαπλών καμερών. Οι Vasavi, Bora, Murali, Jacob & Raj (2025) με τίτλο Robust Face Recognition System for Extensive Distances in Uncontrolled Environments on Edge Device εστιάζουν στην αναγνώριση προσώπων σε μεγάλες αποστάσεις και σε μη ελεγχόμενα περιβάλλοντα, συνδυάζοντας MTCNN και ArcFace/ResNet-50. Μέσω εκτενούς ablation study, αναδεικνύεται η συμβολή dense connections, attention modules, upsampling και margin-based loss στην επίτευξη συνολικής ακρίβειας 91,6% (έναντι 82,4% του MTCNN+ResNet101). Η εργασία παρέχει σημαντική τεκμηρίωση για συνθήκες αποστάσεων και edge deployment, αλλά δεν εξετάζει ζητήματα πολλαπλών ροών εισόδου ή δυναμικής ενημέρωσης της βάσης δεδομένων. Η μελέτη των Faruque, Siddiqui & Noor (2023) αξιοποιεί το πλαίσιο YOLOv5–ArcFace σε εφαρμογή μαζικής καταγραφής παρουσιών σε εκπαιδευτικά περιβάλλοντα. Η πειραματική σύγκριση με άλλους αλγορίθμους (HaarCascade, SSD, MTCNN, LBPH, FaceNet, Deep Face) ανέδειξε την υπεροχή του συνδυασμού (ανίχνευση 95%, αναγνώριση 94,74%). Παρά την υψηλή ακρίβεια, το σύστημα παραμένει προσανατολισμένο σε περιβάλλοντα με μια πηγή εισόδου και δεν αντιμετωπίζει ζητήματα ανοιχτού συνόλου ταυτοτήτων ή δυναμικής παραμετροποίησης αυτών. Σε αντιπαραβολή, η παρούσα εργασία προτείνει ένα ολοκληρωμένο σύστημα πραγματικού χρόνου με πολλαπλές πηγές εισόδου (USB, Wi-Fi, RTSP, αρχεία βίντεο), το οποίο ενσωματώνει YOLOv8n-face για ανίχνευση και ArcFace (ONNX) για αναγνώριση, με δυναμική διαχείριση facebank και παραμετροποίηση κατωφλίων ανά πηγή εισόδου. Η προσέγγιση αυτή επιχειρεί να καλύψει τα κενά των προηγούμενων εργασιών, συνδυάζοντας την ταχύτητα, την ακρίβεια και την επεκτασιμότητα σε ένα λειτουργικό πλαίσιο που ανταποκρίνεται στις απαιτήσεις πραγματικών εφαρμογών μεγάλης κλίμακας.

3.3 Σημεία Υπεροχής της Παρούσας Εργασίας

Η καινοτομία της παρούσας εργασίας δεν εδράζεται αποκλειστικά στην αξιοποίηση σύγχρονων αλγορίθμων, όπως το YOLOv8n-face για την ανίχνευση και το ArcFace ONNX για την αναγνώριση προσώπων, αλλά στον τρόπο με τον οποίο τα επιμέρους υποσυστήματα ενσωματώνονται σε ένα συνολικό, λειτουργικό και επεκτάσιμο πλαίσιο. Ο σχεδιασμός του συστήματος στοχεύει στην αξιοπιστία, την αποτελεσματικότητα και την πρακτική αξιοποίηση σε εφαρμογές πραγματικού χρόνου, υπερβαίνοντας τον ερευνητικό χαρακτήρα και προσεγγίζοντας τις απαιτήσεις ενός ολοκληρωμένου προϊόντος.

Ιδιαίτερη υπεροχή αποτελεί η δυνατότητα της δυναμικής παραμετροποίησης του κατωφλίου αναγνώρισης (threshold Tuning), η οποία επιτρέπει την προσαρμογή του συστήματος σε διαφορετικά περιβάλλοντα και ανάγκες εφαρμογής. Παράλληλα, προβλέπεται ενεργός συμμετοχή του χρήστη στη διαδικασία ενημέρωσης της βάσης δεδομένων, μέσω απλών ενεργειών, όπως η επιλογή καταχώρισης ενός νέου προσώπου. Η λειτουργικότητα αυτή ενισχύει τον αυτοπροσαρμοστικό χαρακτήρα του συστήματος και συμβάλλει στη συνεχή βελτίωση της απόδοσής του. Επιπλέον, η βάση γνωστών προσώπων (facebank) ανανεώνεται αυτόματα στο τέλος κάθε εκτέλεσης, γεγονός που εξασφαλίζει την διαρκή επέκταση και εμπλουτισμό της, προσδίδοντας στο σύστημα χαρακτηριστικά «μνήμης» και αυξανόμενης αποδοτικότητας. Έτσι, η υλοποίηση υπερβαίνει τον περιορισμό των στατικών συνόλων δεδομένων, καθιστώντας το σύστημα περισσότερο ευέλικτο και προσαρμοστικό. Τέλος, στη παρούσα μελέτη η ArcFace loss (score = 1 - cosine(emb, ref_emb)) εφαρμόζεται πρακτικά δηλαδή χρησιμοποιούμε ArcFace-εκπαιδευμένο encoder που παράγει κανονικοποιημένα embeddings. Η ταύτιση γίνεται με ομοιότητα συνημίτονου μεταξύ του embedding του εισόδου και κάθε embedding στο facebank. Για παράδειγμα, θεωρούμε ένα facebank με δύο ταυτότητες, Ελένη και Νίκος , για τις οποίες διαθέτουμε L2-κανονικοποιημένα embeddings. Για ένα νέο πρόσωπο, ο ArcFace encoder παράγει επίσης ένα κανονικοποιημένο embedding. Στη συνέχεια υπολογίζουμε τη συνημιτονική ομοιότητα (cosine similarity) του embedding αυτού με κάθε διάνυσμα στο facebank. Η ταυτότητα που αντιστοιχεί στη μέγιστη ομοιότητα προτείνεται ως αντίστοιχη της πραγματικής ταυτότητας και γίνεται αποδεκτή μόνο εφόσον η τιμή υπερβαίνει ένα προκαθορισμένο κατώφλι (π.χ. 0,50). Με όρους εφαρμογής, αν προκύψουν τιμές ομοιότητας SΕλένη=0,997 και SΝίκος=0,092, τότε επιλέγεται η Ελένη και, δεδομένου ότι 0,997>0,5 το πρόσωπο χαρακτηρίζεται ως γνωστό, διαφορετικά αναγνωρίζεται ως Unknown. Με αυτόν τον τρόπο, η διαδικασία ταύτισης βασίζεται σε κανονικοποιημένα embeddings και με σαφή κανόνα απόφασης μέσω κατωφλίου. Η επιλογή αυτή ενισχύει την επιστημονική εγκυρότητα και τον λειτουργικό χαρακτήρα της εργασίας, προσφέροντας μια ολοκληρωμένη προσέγγιση στο πεδίο της αναγνώρισης προσώπων.

3.4 Συμπερασματικά

Οι παραπάνω επιστημονικές εργασίες δείχνουν τη μετάβαση από στατικές εφαρμογές μεταφοράς μάθησης (Κανελλόπουλος, 2019) σε λύσεις πραγματικού χρόνου (Anjeana & Anusudha, 2024), με βελτιώσεις μέσω tracking/temporal voting (Imran κ.ά., 2024) και ειδικές ρυθμίσεις για δύσκολα περιβάλλοντα και αποστάσεις (Vasavi κ.ά., 2025). Παρά τα θετικά αποτελέσματα, παρατηρούμε ότι υπάρχουν κενά στη μερική απόκρυψη του προσώπου, στη διαχείριση των αγνώστων προσώπων, στη διαχείριση μοναδικών ροών εισόδου και στην έλλειψη δυναμικής ενημέρωσης βάσης προσώπων. Μελέτες εφαρμογών (Faruque κ.ά., 2023) επιβεβαιώνουν την πρακτική αξία συνδυασμών τύπου YOLO–ArcFace, αλλά με τους ίδιους περιορισμούς. Η παρούσα εργασία προτείνει ένα πρακτικό και επεκτάσιμο σύστημα που υποστηρίζει πολλαπλές εισόδους (USB, Wi-Fi, RTSP, αρχεία VIDEO), συνδυάζει YOLOv8n-face για γρήγορη ανίχνευση με ArcFace (ONNX) για αναγνώριση, διαθέτει δυναμική facebank με απλή προσθήκη νέων προσώπων και αυτόματη ενημέρωση, εφαρμόζει ρύθμιση κατωφλίου ανά πηγή εισόδου και τέλος λαμβάνει αποφάσεις με cosine similarity πάνω σε ArcFace embeddings

(όχι με εκ νέου «ArcFace loss» στην αναγνώριση). Συνολικά, συνδυάζει ταχύτητα, ακρίβεια και ευελιξία, υποστηρίζει Open-set λογική και συνεχή βελτίωση της βάσης δεδομένων των προσώπων.

Κεφάλαιο 4:Υλοποίηση Εφαρμογής (Τεχνικά + Τεκμηρίωση Κώδικα)

Η παρούσα εργασία επικεντρώνεται στην ανάπτυξη και αξιολόγηση μιας διαδικασίας αναγνώρισης και παρακολούθησης ανθρώπων σε πραγματικό χρόνο, αξιοποιώντας νευρωνικά δίκτυα και τεχνολογίες υπολογιστικής όρασης. Στόχος είναι η διερεύνηση, ανάλυση και σύγκριση ευρέως διαδεδομένων μεθόδων, όπως το YOLO (You Only Look Once) και διάφορες τεχνικές αναγνώρισης προσώπου, με έμφαση στην ακρίβεια, την ταχύτητα επεξεργασίας και την αξιοπιστία τους σε πραγματικές συνθήκες επιτήρησης. Για την επίτευξη των στόχων, πραγματοποιείται πειραματική αξιολόγηση των επιλεγμένων τεχνολογιών σε ποικίλα σύνολα δεδομένων: βίντεο από κλειστά κυκλώματα παρακολούθησης (CCTV), αρχεία βίντεο αποθηκευμένα τοπικά σε υπολογιστή, καθώς και δεδομένα που συλλέγονται ταυτόχρονα από πολλαπλές κάμερες (σταθερού υπολογιστή, κινητού τηλεφώνου κτλ.). Μέσα από συστηματική ανάλυση των αποτελεσμάτων, εξετάζεται η καταλληλόλητα κάθε μεθόδου σε διαφορετικά σενάρια, όπως η παρακολούθηση πλήθους και η ταυτοποίηση συγκεκριμένων ατόμων. Η εργασία συμβάλλει στην εμβάθυνση της κατανόησης των πλεονεκτημάτων και των περιορισμών κάθε τεχνολογίας, προσφέροντας τεκμηριωμένη καθοδήγηση για την επιλογή της κατάλληλης λύσης σε εφαρμογές ασφάλειας και επιτήρησης. Τέλος, η ανάπτυξη και η υλοποίηση της εφαρμογής ακολούθησαν μια σειρά σαφώς ορισμένων σταδίων, τα οποία παρουσιάζονται αναλυτικά στα επόμενα κεφάλαια.

4.1 Ανάλυση εφαρμογών

Σύστημα Ανίχνευσης & Αναγνώρισης Προσώπων με Βιβλιοθήκη face_recognition

Τεχνική Περιγραφή & Υλοποίηση

Ανάγνωση και διαχείριση γνωστών προσώπων

Το face_detection9c.py συνδυάζει τη βιβλιοθήκη OpenCV για χειρισμό βίντεο/εικόνας και την ισχυρή βιβλιοθήκη face_recognition για αναγνώριση προσώπων μέσω deep learning. Η εφαρμογή αναζητά όλους τους διαθέσιμους φακέλους «Known_faces», όπου κάθε φάκελος αντιστοιχεί σε ένα αναγνωρισμένο άτομο όπου περιέχει φωτογραφίες του. Για κάθε εικόνα, παράγεται ένα μοναδικό face encoding το οποίο αποθηκεύεται με το όνομα του ατόμου. Η διαδικασία διασφαλίζει ότι το σύστημα μπορεί να διαχειριστεί πολλαπλές φωτογραφίες

ανά άτομο, βελτιώνοντας την ακρίβεια της αναγνώρισης, αφού λαμβάνει υπόψη ποικιλίες γωνιών, φωτισμού και εκφράσεων του προσώπου.

Εισαγωγή πηγών βίντεο

Ο χρήστης επιλέγει πηγή για ανίχνευση προσώπων, η οποία μπορεί να είναι όπως φαίνεται και στη παρακάτω εικόνα:

1. Ενσωματωμένη ή εξωτερική κάμερα υπολογιστή/κινητού (π.χ. μέσω USB, Wi-Fi, ή εφαρμογής όπως DroidCam)
2. Αποθηκευμένο βίντεο (π.χ. .mp4, .avi) που βρίσκεται στον φάκελο video του συστήματος
3. Κάμερα συστήματος παρακολούθησης μέσω **RTSP** (Real Time Streaming Protocol)

```
Επιλέξτε πηγή βίντεο για την ανίχνευση:
1: Χρήση κάμερας υπολογιστή και εξωτερικής κάμερας (π.χ., κινητό μέσω RTSP)
2: Χρήση αποθηκευμένου βίντεο
3: Χρήση κλειστού συστήματος παρακολούθησης (RTSP)
Δώσε επιλογή (1, 2 ή 3): 1

Επιλέξτε κάμερα για χρήση:
1: Κάμερα υπολογιστή
2: Εξωτερική κάμερα (π.χ., μέσω κινητό)
3: Και οι δύο κάμερες
Δώσε επιλογή (1, 2 ή 3): _
```

Διαχείριση και επεξεργασία καρέ σε πραγματικό χρόνο

Κάθε καρέ από το βίντεο/κάμερα περνάει από τα εξής στάδια επεξεργασίας:

1. **Προ επεξεργασία**

   Αλλαγή μεγέθους (για αύξηση της ταχύτητας), μείωση θορύβου με κατάλληλα φίλτρα και μετατροπή σε χρωματικό χώρο RGB.
2. **Ανίχνευση προσώπων**

   Με τους αλγορίθμους της βιβλιοθήκης face_recognition εντοπίζονται οι περιοχές του καρέ όπου υπάρχουν πρόσωπα.
3. **Εξαγωγή face encodings**

   Για κάθε ανιχνευμένο πρόσωπο υπολογίζεται το αντίστοιχο αριθμητικό encoding, που αποτυπώνει τα μοναδικά χαρακτηριστικά του μέσω deep learning.
4. **Σύγκριση με γνωστά πρόσωπα**

   Τα νέα encodings συγκρίνονται με τα αποθηκευμένα encodings των γνωστών προσώπων, ώστε να γίνει ταυτοποίηση.

5. **Ανάθεση ονόματος ή χαρακτηρισμός ως "Unknown"**

   Αν η απόσταση μεταξύ encodings βρίσκεται κάτω από το καθορισμένο όριο (συνήθως 0.6), το πρόσωπο αναγνωρίζεται και εμφανίζεται το όνομά του· διαφορετικά χαρακτηρίζεται ως "Unknown".

Οπτική ανατροφοδότηση & ενίσχυση εικόνας

Για κάθε ανιχνευόμενο πρόσωπο:

1. **Οπτική σήμανση**

   Σχεδιάζεται πράσινο πλαίσιο γύρω από το πρόσωπο και προβάλλεται το όνομα ή ο χαρακτηρισμός «Unknown».
2. **Τοπική μεγέθυνση (zoom-in)**

   Εφαρμόζεται μεγέθυνση στην περιοχή του προσώπου για καλύτερη ανάλυση και πιο καθαρή οπτική παρουσίαση.
3. **Εμφάνιση μετρικών απόστασης**

   Προβάλλονται οι τιμές «απόστασης» (distance) μεταξύ των encodings, ώστε να διευκολύνονται οι διαγνωστικοί έλεγχοι και η ρύθμιση των παραμέτρων.
4. **Βελτίωση αντίθεσης με CLAHE**

   Χρησιμοποιείται η τεχνική CLAHE (Contrast Limited Adaptive Histogram Equalization) για βελτίωση της εικόνας σε δύσκολες συνθήκες φωτισμού. Η CLAHE ενισχύει την τοπική αντίθεση μιας εικόνας, ιδιαίτερα όταν είναι σκοτεινή ή έχει χαμηλή συνολική αντίθεση.

Η CLAHE είναι μια τεχνική επεξεργασίας εικόνας που βελτιώνει την τοπική αντίθεση μιας εικόνας, ειδικά όταν αυτή είναι σκοτεινή ή έχει χαμηλή αντίθεση. Χρησιμοποιείται για να βελτιώνει λεπτομέρειες σε σκοτεινά ή θαμπά σημεία, για να μειώνει την εμφάνιση έντονων τεχνητών περιοχών (artefacts) που συχνά προκαλεί η απλή εξισορρόπηση ιστογράμματος και τέλος είναι ιδανική για ιατρικές εικόνες, νυχτερινές λήψεις ή εικόνες με μεγάλη διακύμανση φωτεινότητας.

CLAHE λειτουργία βήμα προς βήμα

1. **Τμηματοποίηση**

   Η εικόνα χωρίζεται σε μικρές περιοχές (tiles/blocks).
2. **Εξισορρόπηση ιστογράμματος ανά περιοχή**

   Σε κάθε tile εφαρμόζεται εξισορρόπηση ιστογράμματος, ενισχύοντας την τοπική αντίθεση.
3. **Περιορισμός αντίθεσης (Contrast Limiting)**

   Αν ορισμένες τιμές φωτεινότητας εμφανίζονται υπερβολικά συχνά, η CLAHE τις «κόβει» (clipping) για να αποφεύγονται υπερβολές και να μη διογκώνεται ο θόρυβος.

4. **Ομαλή συγχώνευση περιοχών**

   Οι γειτονικές περιοχές συγχωνεύονται με παρεμβολή, ώστε να μην προκύπτουν απότομα όρια μεταξύ των tiles.

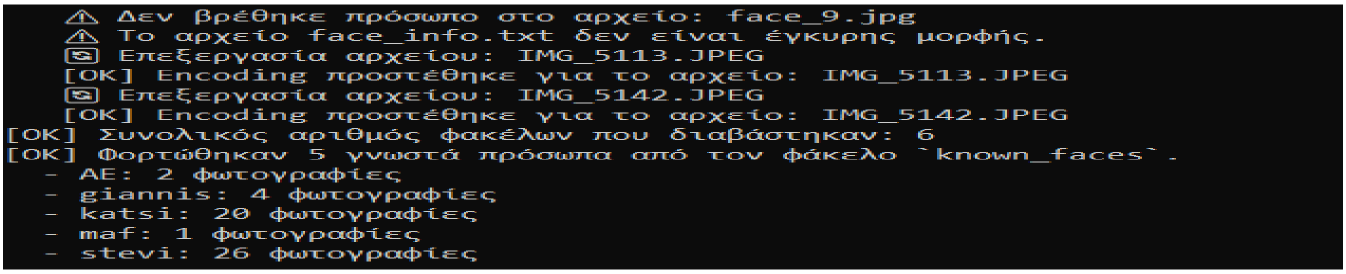
Υποστήριξη πολλαπλών πηγών βίντεο ταυτόχρονα

Το σύστημα υποστηρίζει:

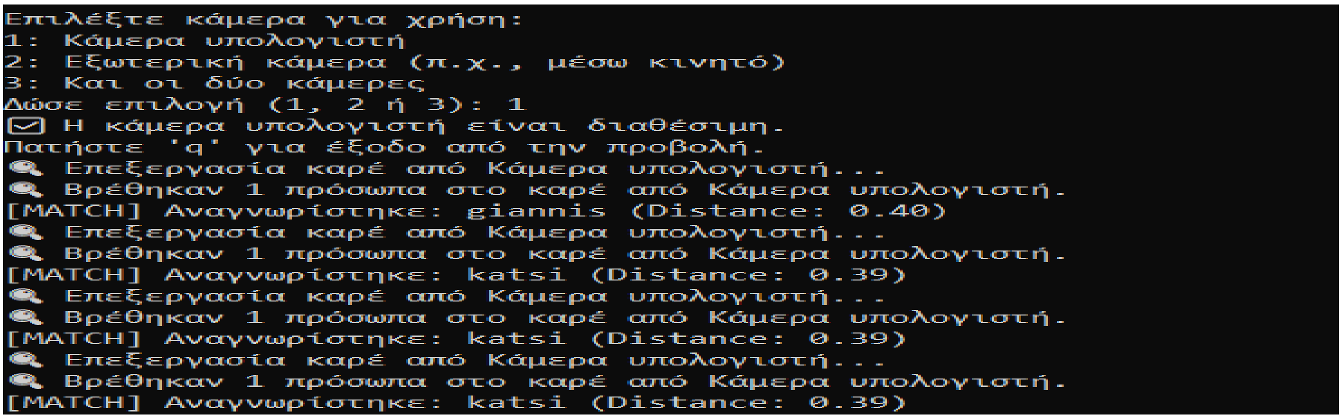
1. Επεξεργασία και αναγνώριση από πολλαπλές κάμερες ή streams ταυτόχρονα, με ανεξάρτητα παράθυρα παρουσίασης (π.χ. κάμερα laptop + κάμερα κινητού + RTSP).
2. Ειδικό χειρισμό για εικονικές κάμερες (όπως DroidCam), επιτρέποντας ακόμα και χρήση κινητού τηλεφώνου ως κάμερα αναγνώρισης.

Διαγνωστικά μηνύματα & στατιστικά

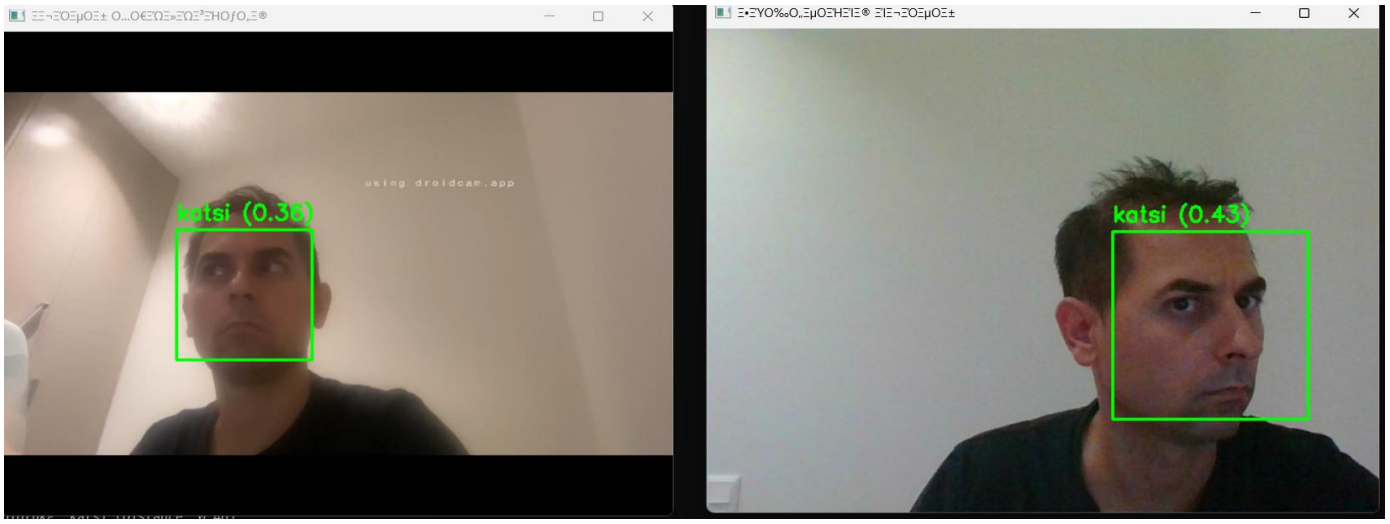
Καθ’ όλη τη διάρκεια λειτουργίας εμφανίζονται διαγνωστικά μηνύματα σχετικά με τη φόρτωση γνωστών προσώπων, την κατάσταση των καμερών/βίντεο, την αναγνώριση ή μη κάθε προσώπου. Υπολογίζεται και προβάλλεται ο συνολικός αριθμός γνωστών και αγνώστων προσώπων, καθώς και στατιστικά για τις αποστάσεις των μη αναγνωρισμένων προσώπων.



```
   ⚠ Δεν βρέθηκε πρόσωπο στο αρχείο: face_9.jpg
   ⚠ Το αρχείο face_info.txt δεν είναι έγκυρης μορφής.
   🔄 Επεξεργασία αρχείου: IMG_5113.JPEG
   [OK] Encoding προστέθηκε για το αρχείο: IMG_5113.JPEG
   🔄 Επεξεργασία αρχείου: IMG_5142.JPEG
   [OK] Encoding προστέθηκε για το αρχείο: IMG_5142.JPEG
[OK] Συνολικός αριθμός φακέλων που διαβάστηκαν: 6
[OK] Φορτώθηκαν 5 γνωστά πρόσωπα από τον φάκελο `known_faces`.
  - AE: 2 φωτογραφίες
  - giannis: 4 φωτογραφίες
  - katsi: 20 φωτογραφίες
  - maf: 1 φωτογραφίες
  - stevi: 26 φωτογραφίες
```



```
Επιλέξτε κάμερα για χρήση:
1: Κάμερα υπολογιστη
2: Εξωτερική κάμερα (π.χ., μέσω κινητό)
3: Και οι δύο κάμερες
Δώσε επιλογή (1, 2 ή 3): 1
✅ Η κάμερα υπολογιστή είναι διαθέσιμη.
Πατήστε 'q' για έξοδο από την προβολή.
🔍 Επεξεργασία καρέ από Κάμερα υπολογιστή...
🔍 Βρέθηκαν 1 πρόσωπα στο καρέ από Κάμερα υπολογιστή.
[MATCH] Αναγνωρίστηκε: giannis (Distance: 0.40)
🔍 Επεξεργασία καρέ από Κάμερα υπολογιστή...
🔍 Βρέθηκαν 1 πρόσωπα στο καρέ από Κάμερα υπολογιστή.
[MATCH] Αναγνωρίστηκε: katsi (Distance: 0.39)
🔍 Επεξεργασία καρέ από Κάμερα υπολογιστή...
🔍 Βρέθηκαν 1 πρόσωπα στο καρέ από Κάμερα υπολογιστή.
[MATCH] Αναγνωρίστηκε: katsi (Distance: 0.39)
🔍 Επεξεργασία καρέ από Κάμερα υπολογιστή...
🔍 Βρέθηκαν 1 πρόσωπα στο καρέ από Κάμερα υπολογιστή.
[MATCH] Αναγνωρίστηκε: katsi (Distance: 0.39)
```



Τεχνολογίες και Εξαρτήσεις

Το σύστημα αναπτύχθηκε σε γλώσσα προγραμματισμού Python, η οποία λειτουργεί ως κύρια γλώσσα προγραμματισμού χάρη στο πλούσιο οικοσύστημα βιβλιοθηκών για υπολογιστική όραση και μηχανική μάθηση. Ο πυρήνας της επεξεργασίας εικόνας/βίντεο υλοποιείται με τη βιβλιοθήκη OpenCV, η οποία υποστηρίζει την πρόσβαση στην κάμερα, τον χειρισμό ροών βίντεο και βασικές απεικονιστικές διεργασίες (π.χ. μετατροπές χρωματικού χώρου, φιλτραρίσματα, σχολιασμούς καρέ). Για την ανίχνευση, εξαγωγή χαρακτηριστικών (encodings) και ταυτοποίηση προσώπων αξιοποιείται η βιβλιοθήκη face_recognition, η οποία ενσωματώνει σύγχρονες τεχνικές βαθιάς μάθησης για υπολογισμό αποστάσεων χαρακτηριστικών και σύγκριση ταυτοτήτων. Η βιβλιοθήκη NumPy υποστηρίζει τις αριθμητικές πράξεις και τη διαχείριση πινάκων/διανυσμάτων υψηλής απόδοσης, ενώ η argparse επιτρέπει την παραμετροποίηση της εφαρμογής μέσω γραμμής εντολών. Με απλά λόγια μπορούμε να πούμε η NumPy εκτελεί το «βαρύ» αριθμητικό κομμάτι γρήγορα και με ακρίβεια ενώ η argparse ορίζει «τι» θα εκτελέσουμε και «με ποιες ρυθμίσεις». Για την οργανωμένη αποθήκευση των κωδικοποιήσεων του προσώπου ανά άτομο χρησιμοποιείται η δομή defaultdict από το πακέτο collections, διευκολύνοντας την ευέλικτη χαρτογράφηση μεταξύ ταυτότητας και πολλαπλών encodings.

Εγχειρίδιο Χρήσης της εφαρμογής

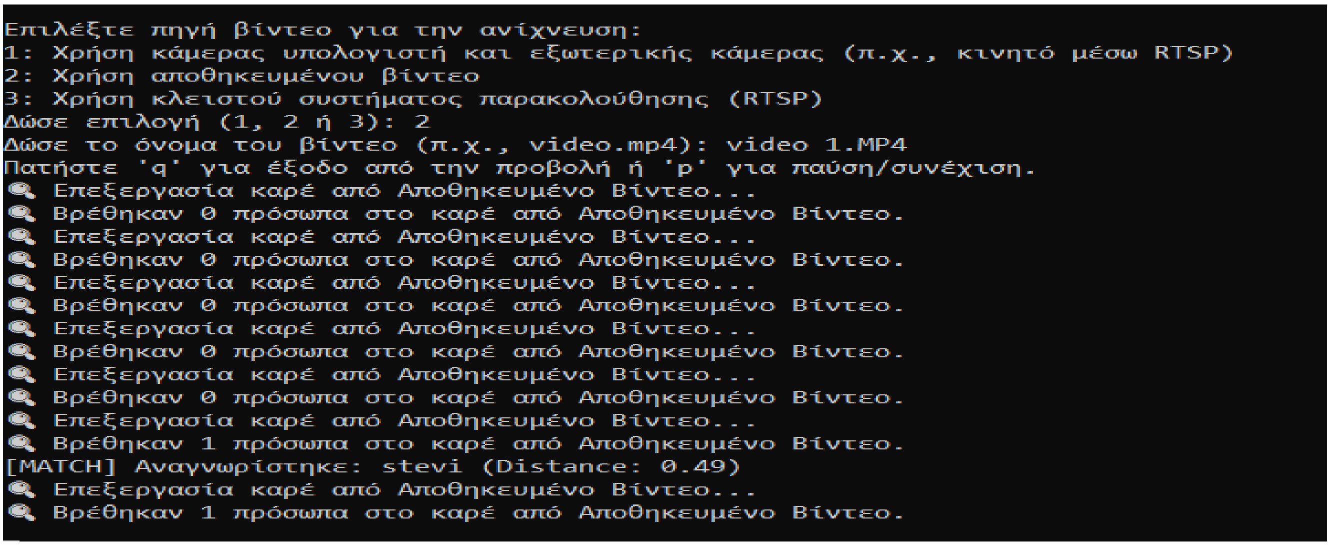
Η εφαρμογή έχει σχεδιαστεί για την ανίχνευση και αναγνώριση προσώπων σε πραγματικό χρόνο, αξιοποιώντας τη βιβλιοθήκη OpenCV για τη λήψη και προβολή βίντεο και τη βιβλιοθήκη face_recognition για την εξαγωγή και σύγκριση χαρακτηριστικών. Πριν την εκτέλεση, ο χρήστης πρέπει να έχει οργανώσει τον φάκελο Known_faces όπως φαίνεται στο κεφάλαιο (4.2 Εγκατάσταση εφαρμογών – Υπο-παράγραφος 1.5 Δομή φακέλου). Μέσα σε αυτόν το φάκελο έχουν δημιουργηθεί ξεχωριστεί υπό φάκελοι για κάθε άτομο που θέλει να αναγνωρίζεται από το σύστημα για παράδειγμα, ένας φάκελος με όνομα «Μαρία» και ένας με όνομα «Γιάννης». Σε κάθε φάκελο τοποθετούνται μερικές καθαρές φωτογραφίες του αντίστοιχου προσώπου, κατά προτίμηση από διαφορετικές γωνίες και με καλό φωτισμό. Οι φωτογραφίες γίνονται δεκτές σε μορφή .jpg, .jpeg ή .png. Παράλληλα, υπάρχει ο φάκελος video , στον οποίο ο χρήστης μπορεί να αποθηκεύσει αρχεία βίντεο για δοκιμές. Η εκκίνηση γίνεται μέσα από τη γραμμή εντολών με την εντολή python face_detectio9c.py, στη συνέχεια φορτώνονται και επεξεργάζονται τα γνωστά πρόσωπα από τον παραπάνω φάκελο όπου παράγεται ένα μοναδικό face encoding για κάθε φωτογραφία του ατόμου.

```
C:\Users\dioni\Desktop\YOLO PROJECT>python face_detection9C.py
[OK] Φορτώνονται γνωστά πρόσωπα από τον φάκελο: C:\Users\dioni\Desktop\YOLO PROJECT\Known_faces
[SCAN] Επεξεργασία φακέλου για το άτομο: AE
   ⚠ Το αρχείο embedding_1.npy δεν είναι έγκυρης μορφής.
   ⚠ Το αρχείο embedding_3.npy δεν είναι έγκυρης μορφής.
   🔄 Επεξεργασία αρχείου: face_1.jpg
   [OK] Encoding προστέθηκε για το αρχείο: face_1.jpg
   🔄 Επεξεργασία αρχείου: face_3.jpg
   [OK] Encoding προστέθηκε για το αρχείο: face_3.jpg
[SCAN] Επεξεργασία φακέλου για το άτομο: giannis
   🔄 Επεξεργασία αρχείου: 71870398641__B08AC05B-FBE5-4AB2-9464-F36D6952266F.fullsizerender.JPEG
   [OK] Encoding προστέθηκε για το αρχείο: 71870398641__B08AC05B-FBE5-4AB2-9464-F36D6952266F.fullsizerender.JPEG
```

Αμέσως μετά, εμφανίζεται ένα μενού που ζητά από τον χρήστη να επιλέξει την πηγή εισόδου για την ανίχνευση. Υπάρχουν τρεις εναλλακτικές πηγές η χρήση της κάμερας του υπολογιστή ή κάποιας εξωτερικής κάμερας (με σύνδεση μέσω Wi-Fi ή USB) ή η επιλογή να τρέξουν και οι δύο κάμερες ταυτόχρονα σε διαφορετικά παράθυρα, η προβολή αποθηκευμένου βίντεο από τον φάκελο video και τέλος η σύνδεση με κλειστό κύκλωμα παρακολούθησης.

```
Επιλέξτε πηγή βίντεο για την ανίχνευση:
1: Χρήση κάμερας υπολογιστή και εξωτερικής κάμερας (π.χ., κινητό μέσω RTSP)
2: Χρήση αποθηκευμένου βίντεο
3: Χρήση κλειστού συστήματος παρακολούθησης (RTSP)
Δώσε επιλογή (1, 2 ή 3): 1

Επιλέξτε κάμερα για χρήση:
1: Κάμερα υπολογιστή
2: Εξωτερική κάμερα (π.χ., μέσω κινητό)
3: Και οι δύο κάμερες
Δώσε επιλογή (1, 2 ή 3):
```



```
Επιλέξτε πηγή βίντεο για την ανίχνευση:
1: Χρήση κάμερας υπολογιστή και εξωτερικής κάμερας (π.χ., κινητό μέσω RTSP)
2: Χρήση αποθηκευμένου βίντεο
3: Χρήση κλειστού συστήματος παρακολούθησης (RTSP)
Δώσε επιλογή (1, 2 ή 3): 2
Δώσε το όνομα του βίντεο (π.χ., video.mp4): video 1.MP4
Πατήστε 'q' για έξοδο από την προβολή ή 'p' για παύση/συνέχιση.
🔍 Επεξεργασία καρέ από Αποθηκευμένο Βίντεο...
🔍 Βρέθηκαν 0 πρόσωπα στο καρέ από Αποθηκευμένο Βίντεο.
🔍 Επεξεργασία καρέ από Αποθηκευμένο Βίντεο...
🔍 Βρέθηκαν 0 πρόσωπα στο καρέ από Αποθηκευμένο Βίντεο.
🔍 Επεξεργασία καρέ από Αποθηκευμένο Βίντεο...
🔍 Βρέθηκαν 0 πρόσωπα στο καρέ από Αποθηκευμένο Βίντεο.
🔍 Επεξεργασία καρέ από Αποθηκευμένο Βίντεο...
🔍 Βρέθηκαν 0 πρόσωπα στο καρέ από Αποθηκευμένο Βίντεο.
🔍 Επεξεργασία καρέ από Αποθηκευμένο Βίντεο...
🔍 Βρέθηκαν 0 πρόσωπα στο καρέ από Αποθηκευμένο Βίντεο.
🔍 Επεξεργασία καρέ από Αποθηκευμένο Βίντεο...
🔍 Βρέθηκαν 1 πρόσωπα στο καρέ από Αποθηκευμένο Βίντεο.
[MATCH] Αναγνωρίστηκε: stevi (Distance: 0.49)
🔍 Επεξεργασία καρέ από Αποθηκευμένο Βίντεο...
🔍 Βρέθηκαν 1 πρόσωπα στο καρέ από Αποθηκευμένο Βίντεο.
```

Κατά την εκτέλεση, κάθε εισερχόμενο καρέ υφίσταται μια σειρά από επεξεργασίες. Πρώτα μειώνεται το μέγεθός του στο μισό για ταχύτερη επεξεργασία και εφαρμόζεται φίλτρο Gaussian για μείωση του θορύβου. Στη συνέχεια η εικόνα μετατρέπεται σε RGB ώστε να είναι συμβατή με το face_recognition το οποίο εντοπίζει τα πρόσωπα και εξάγει τα αντίστοιχα διανύσματα χαρακτηριστικών (encodings). Τα νέα encodings συγκρίνονται με τα ήδη αποθηκευμένα από τον φάκελο γνωστών προσώπων. Αν η τιμή απόστασης είναι μικρότερη από το όριο 0.60, το πρόσωπο θεωρείται αναγνωρισμένο και εμφανίζεται το όνομά του στην οθόνη μαζί με την τιμή της απόστασης. Αντίθετα, αν το αποτέλεσμα βρίσκεται πάνω από το κατώφλι, το πρόσωπο χαρακτηρίζεται ως «Unknown». Παράλληλα, το πρόγραμμα κάνει zoom και ενίσχυση αντίθεσης (CLAHE) στην περιοχή του προσώπου ώστε να βελτιώνεται η ορατότητα. Κατά τη διάρκεια της χρήσης του script, ο χρήστης έχει στη διάθεσή του μερικούς απλούς χειρισμούς από το πληκτρολόγιο, με το πλήκτρο q τερματίζει την εκτέλεση και κλείνει τα παράθυρα, ενώ στην περίπτωση που παίζει ένα αποθηκευμένο βίντεο, με το πλήκτρο p μπορεί να το παγώσει και να το συνεχίσει. Στο τερματικό εμφανίζονται διαγνωστικά μηνύματα, όπως πόσα πρόσωπα εντοπίστηκαν σε κάθε καρέ ή αν υπήρξε επιτυχής ταύτιση με κάποιο γνωστό πρόσωπο. Στο τέλος, το πρόγραμμα εκτυπώνει τον αριθμό των αγνώστων προσώπων και τις αποστάσεις ομοιότητας που καταγράφηκαν. Η σωστή λειτουργία του συστήματος προϋποθέτει επαρκή αριθμό και ποιότητα φωτογραφιών για κάθε γνωστό πρόσωπο. Αν η αναγνώριση δεν είναι επιτυχής, συνίσταται να προστεθούν περισσότερες φωτογραφίες ή να βελτιωθεί ο φωτισμός στον χώρο λήψης.

Services in Action – Περίπτωση Χρήσης Εφαρμογής

Σενάριο: Έλεγχος πρόσβασης στην κεντρική είσοδο μιας μεγάλης εταιρίας

Η εφαρμογή μπορεί να αξιοποιηθεί αποτελεσματικά σε πραγματικές συνθήκες ελέγχου πρόσβασης, όπως για παράδειγμα στην κεντρική είσοδο μιας μεγάλης εταιρίας. Στην περίπτωση αυτή, η διοίκηση φροντίζει να δημιουργήσει εκ των προτέρων τον φάκελο Known_faces, στον οποίο καταχωρούνται οι φωτογραφίες όλων των εργαζομένων που έχουν δικαίωμα εισόδου στο κτίριο. Κάθε εργαζόμενος διαθέτει τον δικό του υπό φάκελο

με αρκετές φωτογραφίες από διαφορετικές γωνίες και με ποικιλία εκφράσεων, ώστε το σύστημα να μπορεί να αναγνωρίζει με μεγαλύτερη ακρίβεια τα πρόσωπα ακόμη και σε συνθήκες διαφορετικού φωτισμού. Συνήθως στην είσοδο υπάρχει μία σταθερή κάμερα παρακολούθησης, αλλά μπορεί να χρησιμοποιηθεί και δεύτερη κάμερα (π.χ. εξωτερική ή συνδεδεμένη μέσω δικτύου με RTSP). Μετά την επιλογή, ανοίγει αυτόματα το παράθυρο προβολής με τη ζωντανή εικόνα από τον χώρο. Καθώς οι εργαζόμενοι εισέρχονται στο κτίριο, η κάμερα καταγράφει το πρόσωπό τους. Το πρόγραμμα μειώνει το μέγεθος του καρέ, το επεξεργάζεται για να μειώσει τον θόρυβο και στη συνέχεια αναγνωρίζει τα χαρακτηριστικά του προσώπου με τη βιβλιοθήκη face_recognition. Αν το πρόσωπο που ανιχνεύεται ταιριάζει με κάποιο από τα ήδη αποθηκευμένα στον φάκελο Known faces τότε στην οθόνη εμφανίζεται το όνομα του εργαζομένου μαζί με μια ένδειξη βαθμού ομοιότητας (π.χ. «Μαρία (0.42)»). Η παρουσία αυτή επιβεβαιώνει την ταυτότητά του και επιτρέπει την απρόσκοπτη είσοδό του. Αντίθετα, στην περίπτωση που εμφανιστεί άτομο που δεν έχει δηλωθεί στο σύστημα, θα το χαρακτηρίσει ως Unknown. Επίσης παρέχονται και πρακτικοί χειρισμοί μέσω του πληκτρολογίου στον χρήστη. Με το πάτημα του πλήκτρου q γίνεται άμεσος τερματισμός της εφαρμογής, ενώ στην περίπτωση που χρησιμοποιείται προ-ηχογραφημένο βίντεο υπάρχει δυνατότητα παύσης και συνέχισης μέσω του πλήκτρου p. Κατά τη διάρκεια της εκτέλεσης, στο τερματικό εμφανίζονται μηνύματα που ενημερώνουν για την πρόοδο της αναγνώρισης, τον αριθμό των προσώπων που ανιχνεύθηκαν σε κάθε καρέ, αλλά και τις τιμές ομοιότητας. Με αυτόν τον τρόπο, η κεντρική είσοδος της εταιρίας αποκτά ένα έξυπνο και αυτοματοποιημένο σύστημα ελέγχου ταυτότητας. Η πρόσβαση των εργαζομένων γίνεται ταχύτερη και ασφαλέστερη, μειώνοντας με αυτό το τρόπο την υποκειμενικότητα κατά τη διάρκεια των ελέγχων, ενώ παράλληλα καταγράφονται συστηματικά οι απόπειρες εισόδου από άγνωστα πρόσωπα.

Σύστημα Ανίχνευσης με YOLO & Αναγνώρισης Προσώπων με ArcFace ONNX

Τεχνική Περιγραφή & Υλοποίηση

Ανάγνωση και διαχείριση γνωστών προσώπων

Το script βασίζεται στη φιλοσοφία του facebank. Κατά την αρχικοποίηση, διαβάζονται από τον φάκελο Known_faces όλα τα διαθέσιμα δείγματα προσώπων ανά ταυτότητα. Για κάθε δείγμα, υπολογίζεται embedding μέσω ArcFace ONNX και οργανώνεται σε dictionary, όπου κάθε ταυτότητα συνδέεται με λίστα embeddings. Επιπρόσθετα, τα embeddings αποθηκεύονται (ως .npy αρχεία και .npz αρχεία συμπίεσης) με σκοπό τη γρήγορη φόρτωση και την αποτροπή επανυπολογισμού σε επόμενες εκτελέσεις. Ιδιαίτερα σημαντική είναι η δυνατότητα δυναμικής επέκτασης με νέα πρόσωπα τα οποία μπορούν να καταχωρούνται άμεσα στο facebank κατά την ανίχνευση, μαζί με το αντίστοιχο cropped snapshot, εξασφαλίζοντας αυτό-εκπαίδευση και εμπλουτισμό της βάσης δεδομένων.

Εισαγωγή πηγών βίντεο

Η αρχιτεκτονική καλύπτει το πλήρες φάσμα εισερχόμενων πηγών:

1. Ενσωματωμένες και εξωτερικές κάμερες
2. Βίντεο αρχείων (με περιήγηση σε τοπικό φάκελο)
3. RTSP streams από συστήματα παρακολούθησης με δυνατότητα διαχείρισης πολλών streams ταυτόχρονα

Υποστηρίζει ταυτόχρονη επεξεργασία πολλών πηγών βίντεο. Κάθε πηγή διαχειρίζεται ως ανεξάρτητο VideoCapture instance, με ξεχωριστό παράθυρο προβολής, διατηρώντας πλήρη λειτουργικότητα για όλες τις διαθέσιμες επιλογές (κάμερες, RTSP, αρχεία). Η λογική αυτή το καθιστά κατάλληλο για σενάρια πραγματικής επιτήρησης με πολλαπλά streams και αυξημένες απαιτήσεις. Κάθε πηγή ανοίγει μέσω κατάλληλης βοηθητικής συνάρτησης, η οποία παρέχει διαγνωστική ανατροφοδότηση διαθεσιμότητας, και ο χειριστής καθοδηγείται με αλληλεπιδραστικό μενού επιλογών. Επιπρόσθετα εμφανίζονται ενημερωτικά μηνύματα της πηγής και της αριθμητικής τιμής του threshold.

```
Επιλέξτε πηγή βίντεο για την ανίχνευση:
1: Κάμερα υπολογιστή / Εξωτερική / Και οι δύο
2: Αποθηκευμένο βίντεο
3: Κλειστό σύστημα παρακολούθησης (RTSP)
Δώσε επιλογή (1, 2 ή 3): 1
[DEBUG] Αναζήτηση threshold στο: C:\Users\dioni\Desktop\YOLO PROJECT\threshold_1.txt
[DEBUG] Περιεχόμενο threshold αρχείου: '0.3'
[CONFIG] Χρήση threshold από αρχείο C:\Users\dioni\Desktop\YOLO PROJECT\threshold_1.txt: 0.3
```

Διαχείριση και επεξεργασία καρέ σε πραγματικό χρόνο

Για κάθε καρέ που λαμβάνεται από τις πηγές, εντοπίζονται πρόσωπα με YOLOv8 και κάθε bounding box περνάει ελέγχους ποιότητας και στατιστικών τιμών pixel (mean, std) για αποφυγή false positives από σκοτεινά/κενά crops. Για κάθε valid crop, εξάγεται embedding με ArcFace, κανονικοποιείται και συγκρίνεται με κάθε embedding του facebank (cosine similarity). Η διαδικασία αναγνώρισης βασίζεται στο αν το μέγιστο score υπερβαίνει το ορισμένο threshold, το οποίο ρυθμίζεται δυναμικά (με βάση το context της πηγής ή με argument/αρχείο). Αν ανιχνευθεί "Unknown", το snapshot αποθηκεύεται, εμφανίζεται στον χρήστη και μπορεί να καταχωρηθεί ως νέο πρόσωπο, αυτομάτως ενημερώνοντας και retraining το facebank χωρίς τερματισμό της εφαρμογής.

Οπτική ανατροφοδότηση & ενίσχυση εικόνας

Η υλοποίηση παρέχει πολύ επίπεδη οπτική ανατροφοδότηση. Τα αναγνωρισμένα πρόσωπα σημειώνονται με πλαίσια, labels και scores, συμπεριλαμβανομένων πληροφοριών threshold και πηγής. Για κάθε "Unknown" πρόσωπο εμφανίζεται αυτόματα popup με snapshot, το οποίο συνοδεύεται από προτροπή προς τον χρήστη για άμεση ονομασία ή παράλειψη. Ειδικοί οπτικοί δείκτες (π.χ., πράσινα πλαίσια και ετικέτες) σηματοδοτούν καρέ ή crops χαμηλής ποιότητας ή άδειες περιοχές, ενισχύοντας τη διαφάνεια του pipeline και επιτρέποντας γρήγορη διάγνωση σφαλμάτων.

Διαγνωστικά μηνύματα & στατιστικά

Η υλοποίηση παρέχει εκτενείς διαγνωστικές αναφορές και στατιστικά σε όλα τα στάδια της λειτουργίας. Από την αρχικοποίηση των παραμέτρων εισόδου με διαγνωστικά μηνύματα ανοίγματος πηγών και επιτυχούς/αποτυχημένης σύνδεσης, με συνεχείς ενημερώσεις κατά τη δημιουργία ή φόρτωση του facebank, retrain, και αποθήκευσης embeddings, με στατιστικές πληροφορίες και scores αναγνώρισης, εμφανιζόμενες σε πραγματικό χρόνο και τέλος με Warnings και debug prints για σκοτεινά, άδεια ή εσφαλμένα crops.

Τεχνολογίες και Εξαρτήσεις

Η υλοποίηση αξιοποιεί μια πληθώρα σύγχρονων βιβλιοθηκών και τεχνολογιών. Το OpenCV αναλαμβάνει τη διαχείριση και προ επεξεργασία εικόνας/βίντεο. Η ανίχνευση προσώπων πραγματοποιείται με Ultralytics YOLOv8, ενώ για την αναγνώριση αυτών αξιοποιείται το ArcFace σε μορφή ONNX, εκτελούμενο μέσω onnxruntime με υποστήριξη CUDA ή CPU. Τα δεδομένα και τα χαρακτηριστικά διαχειρίζονται αποδοτικά με NumPy, ενώ η αποθήκευση και η ταξινόμηση γίνονται σε JSON. Πριν από τη σύγκριση, τα διανύσματα κανονικοποιούνται με scikit-learn, και η ομοιότητα υπολογίζεται μέσω Scipy, με έμφαση στη συσχέτιση συνημίτονου (cosine similarity). Τέλος, το PyTorch (torch) χρησιμοποιείται για τον έλεγχο της διαθεσιμότητας GPU.

Εκκίνηση και βασικοί χειρισμοί

Η εφαρμογή εκκινείτε μέσω γραμμής εντολών με την εντολή «python face_detection15D0.py».

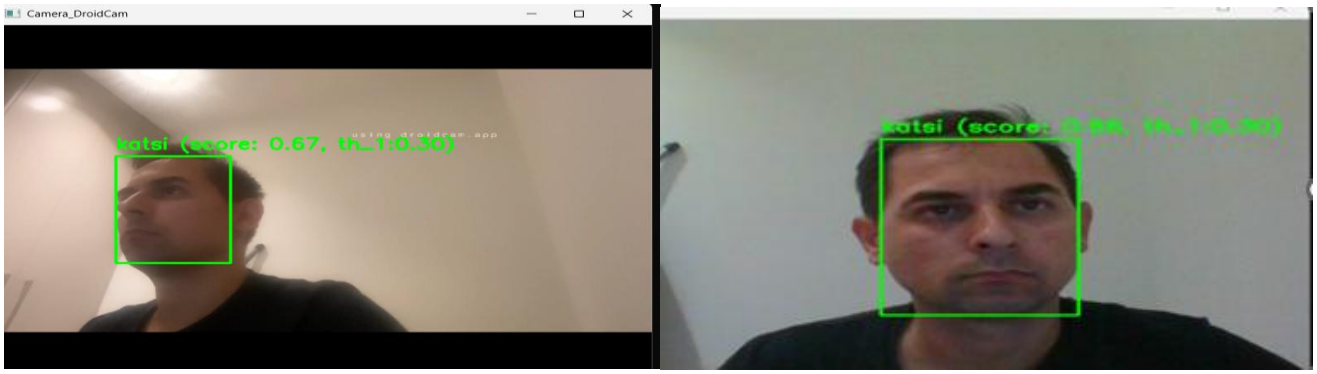
```
C:\Users\dioni\Desktop\YOLO PROJECT>python face_detection15D0.py
☑ Το facebank φορτώθηκε.
🔄 Υπολογισμός embeddings για όλα τα γνωστά πρόσωπα από τον φάκελο Known_faces...
[INFO] Διαβάζονται και υπολογίζονται embeddings για όλα τα γνωστά πρόσωπα από: C:\Users\dioni\Desktop\YOLO PROJECT\Known_faces
[INFO] Ολοκληρώθηκε η δημιουργία embeddings για όλα τα γνωστά πρόσωπα. Βρέθηκαν: ['AE', 'giannis', 'jamal', 'katsi', 'maf', 'stevi']
☑ Τα embeddings των γνωστών προσώπων ενημερώθηκαν από τον φάκελο Known_faces.
[INFO] Το module δυναμικού threshold δεν βρέθηκε - χρήση σταθερών τιμών
```

Μετά την εκκίνηση, ο χρήστης καλείται να επιλέξει την πηγή εισόδου εικόνας ή βίντεο.

```
Επιλέξτε πηγή βίντεο για την ανίχνευση:
1: Κάμερα υπολογιστή / Εξωτερική / Και οι δύο
2: Αποθηκευμένο βίντεο
3: Κλειστό σύστημα παρακολούθησης (RTSP)
Δώσε επιλογή (1, 2 ή 3): 1
[DEBUG] Αναζήτηση threshold στο: C:\Users\dioni\Desktop\YOLO PROJECT\threshold_1.txt
[DEBUG] Περιεχόμενο threshold αρχείου: '0.3'
[CONFIG] Χρήση threshold από αρχείο C:\Users\dioni\Desktop\YOLO PROJECT\threshold_1.txt: 0.3
```

Υπάρχει η δυνατότητα αξιοποίησης της κάμερας του υπολογιστή ή μιας άλλης εξωτερικής κάμερας ή ακόμη και την επιλογή να τρέχουν παράλληλα σε διαφορετικά παράθυρα ταυτόχρονα και οι δύο κάμερες καθώς επίσης και την επιλογή να τρέξει ένα αρχείο βίντεο από τον φάκελο του υπολογιστή που έχει επιλέξει ο χρήστης ή

σύνδεσης με ροές RTSP που προέρχονται από κλειστά κυκλώματα παρακολούθησης. Η λειτουργία του συστήματος βασίζεται σε μία ακολουθία βημάτων που διασφαλίζουν την ανίχνευση, την αναγνώριση και την ενσωμάτωση νέων ταυτοτήτων. Αρχικά, κάθε εισερχόμενο βίντεο υποβάλλεται σε επεξεργασία με το μοντέλο YOLOv8-face, το οποίο εντοπίζει πρόσωπα στα οποία εμφανίζονται με ορθογώνια πλαίσια και συνοδεύονται από ετικέτες ταυτοποίησης και εξάγει τις αντίστοιχες περιοχές ενδιαφέροντος.

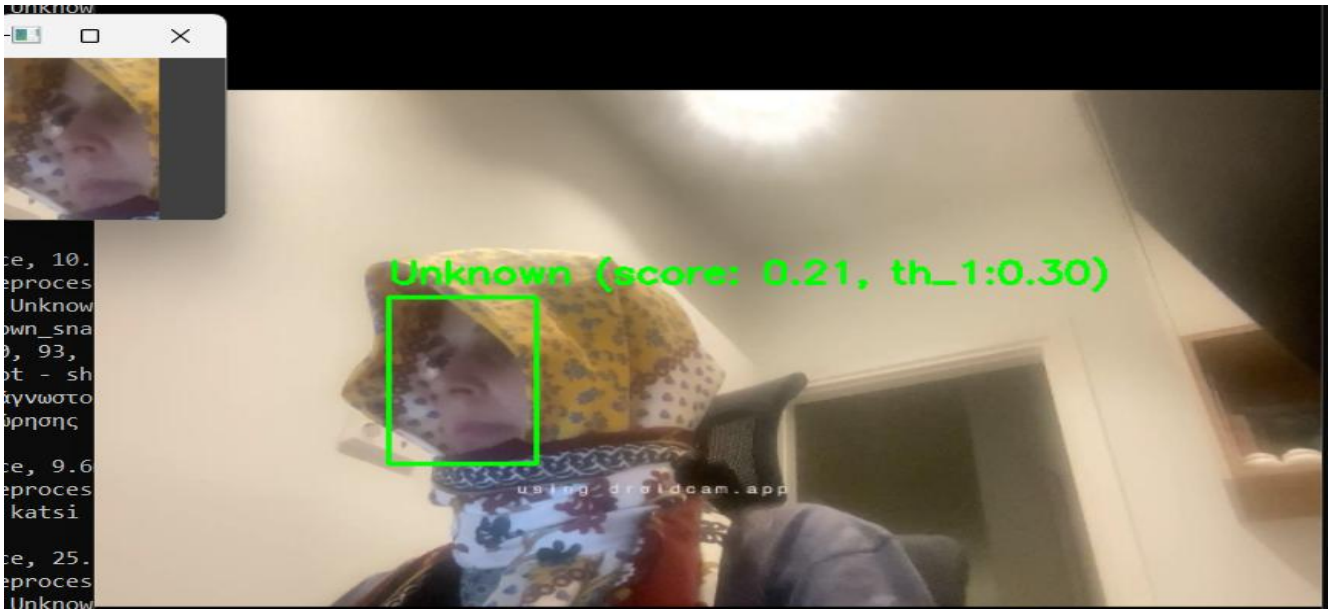


Στις ετικέτες αυτές εμφανίζεται είτε το όνομα του προσώπου είτε η ένδειξη Unknown, μαζί με την τιμή του δείκτη ομοιότητας και το αντίστοιχο κατώφλι αναγνώρισης.

```
0: 480x640 1 face, 20.9ms
Speed: 2.4ms preprocess, 20.9ms inference, 4.5ms postprocess per image at shape (1, 3, 480, 640)
[INFO] Πρόσωπο: katsi | Score: 0.66 | Threshold: 0.30
```

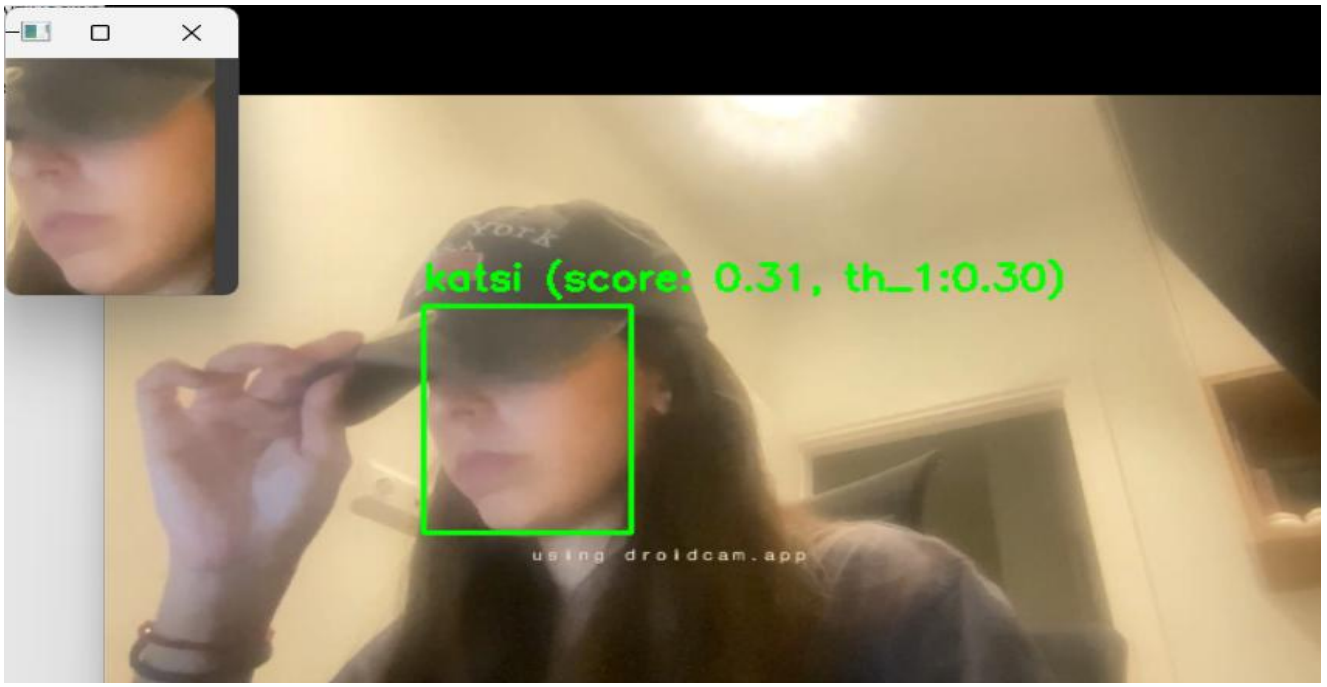
Επομένως για κάθε αποκομμένο τμήμα εικόνας προσώπου, εφαρμόζεται το μοντέλο ArcFace σε περιβάλλον ONNX, ώστε να παραχθεί ένα αποτύπωμα (embedding). Το αποτύπωμα αυτό συγκρίνεται με όλα τα διαθέσιμα embeddings που είναι αποθηκευμένα στο facebank μέσω του δείκτη συνημίτονου γωνίας (cosine similarity). Ως δείκτης ταυτοποίησης λαμβάνεται το μέγιστο σκορ ομοιότητας. Αν το σκορ αυτό είναι μεγαλύτερο ή ίσο από το προκαθορισμένο κατώφλι “threshold”, το πρόσωπο αναγνωρίζεται και του αποδίδεται η αντίστοιχη ετικέτα. Αντιθέτως, όταν το σκορ είναι χαμηλότερο του κατωφλίου, το πρόσωπο χαρακτηρίζεται ως Unknown.

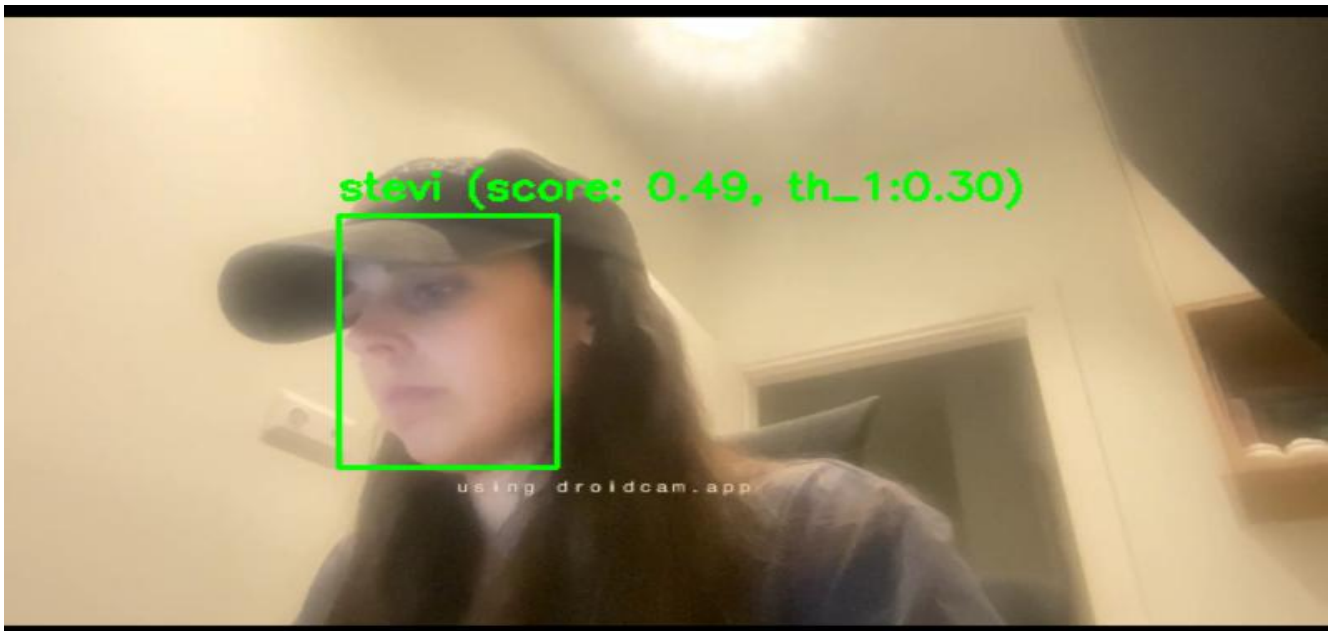
```
0: 480x640 1 face, 9.0ms
Speed: 1.5ms preprocess, 9.0ms inference, 1.8ms postprocess per image at shape (1, 3, 480, 640)
[INFO] Πρόσωπο: Unknown | Score: 0.25 | Threshold: 0.30
[SNAPSHOT] Unknown_snapshots\Camera_DroidCam_unknown_0.25_20250924_225452_191234.jpg
crop shape: (113, 85, 3) min/max: 28 230
Εμφάνιση snapshot - shape: (113, 85, 3) min: 28 max: 230
➡ Εντοπίστηκε άγνωστο πρόσωπο. Πάτησε 'n' για να το καταχωρήσεις, 'q' για έξοδο ή οποιοδήποτε άλλο πλήκτρο για παράλειψη.
Όνομα: stevi
[INFO] Facebank τώρα περιέχει 101 embeddings για 'stevi'.
[SYNC] Το embedding αποθηκεύτηκε και στον φάκελο Known_faces: C:\Users\dioni\Desktop\YOLO PROJECT\Known_faces\stevi\embedding_101.npy
[SYNC] Το crop αποθηκεύτηκε και στον φάκελο Known_faces: C:\Users\dioni\Desktop\YOLO PROJECT\Known_faces\stevi\face_101.jpg
[INFO] Διαβάζονται και υπολογίζονται embeddings για όλα τα γνωστά πρόσωπα από: C:\Users\dioni\Desktop\YOLO PROJECT\Known_faces
[INFO] Ολοκληρώθηκε η δημιουργία embeddings για όλα τα γνωστά πρόσωπα. Βρέθηκαν: ['AE', 'jamal', 'katsi', 'maf', 'stevi']
🔄 Έγινε retrain όλων των γνωστών προσώπων στο facebank.
✅ Το πρόσωπο 'stevi' προστέθηκε στο facebank.
```



Σε αυτή την περίπτωση μπορεί να αποθηκευτεί στιγμιότυπο και embedding, ενώ δίνεται η δυνατότητα στον χρήστη να αποδώσει όνομα και να το προσθέσει άμεσα στο facebank.

Αυτό μπορεί να πραγματοποιηθεί μέσω βασικών εντολών από το πληκτρολόγιο με το πλήκτρο n όπου παρέχεται η δυνατότητα προσθήκης ταυτότητας του αγνώστου προσώπου και άμεσης προσθήκης του στο facebank δίνοντας τη δυνατότητα στο χρήστη να εκπαιδεύσει το μοντέλο, με το πλήκτρο r εκτελείται εκ νέου υπολογισμός (retrain) του facebank βάσει των δεδομένων που βρίσκονται στον φάκελο Known_faces στη περίπτωση που αντιληφθεί ο χρήστης κατά τη διάρκεια όπου τρέχει το script ότι δεν έγινε σωστός υπολογισμός των embeddings και τέλος με το πλήκτρο q ολοκληρώνεται η συνεδρία και τερματίζεται το πρόγραμμα.





Κατά τον τερματισμό της εφαρμογής πραγματοποιείται retrain στο facebank, ο οποίος αξιοποιεί τόσο τα δεδομένα του φακέλου Known_faces όσο και τα στιγμιότυπα και τα ενσωματωμένα διανύσματα που έχουν αποθηκευτεί στον φάκελο Unknown snapshots. Στην τελευταία περίπτωση, κάθε ανώνυμο στιγμιότυπο λαμβάνει μια μοναδική ετικέτα, ώστε να διασφαλίζεται η συνέπεια και η αξιοπιστία του συστήματος. Τέλος είναι σημαντικό να αναφερθεί ότι τη διάρκεια που τρέχει το script εμφανίζονται τα ανάλογα διαγνωστικά μηνύματα ώστε ο χρήστης να έχει πλήρη έλεγχο και ενημέρωση.

```
[EXIT] Έγινε έξοδος από όλες τις κάμερες.
[INFO] Ξεκινά retrain για όλα τα γνωστά και άγνωστα πρόσωπα...
🔄 Έγινε retrain όλων των προσώπων και unknown στο facebank.
[INFO] Ολοκληρώθηκε retrain για όλα τα γνωστά και άγνωστα πρόσωπα.
☑ Το πρόγραμμα τερματίστηκε επιτυχώς.
```

Services in Action - Περίπτωση Χρήσης Εφαρμογής

Σενάριο: Έλεγχος πρόσβασης στην κεντρική είσοδο μιας μεγάλης εταιρίας

Η προτεινόμενη εφαρμογή μπορεί να αξιοποιηθεί αποτελεσματικά σε ένα περιβάλλον μεγάλης επιχείρησης, όπου απαιτείται συστηματικός και ασφαλής έλεγχος ταυτότητας κατά την είσοδο των εργαζομένων. Σκοπός της συγκεκριμένης περίπτωσης χρήσης είναι η αυτοματοποίηση της διαδικασίας αναγνώρισης προσώπων στην κεντρική είσοδο, η ταχεία και αδιάβλητη επιβεβαίωση της ταυτότητας των εργαζομένων και η δυνατότητα επανεκπαίδευσης του μοντέλου από τον κεντρικό διαχειριστή του συστήματος. Κατά την έναρξη της λειτουργίας, εκκινείτε το πρόγραμμα και ξεκινά την απεικόνιση σε πραγματικό χρόνο των ζωντανών λήψεων. Καθώς οι εργαζόμενοι πλησιάζουν την είσοδο, εντοπίζει τα πρόσωπά τους και δημιουργεί τις αντίστοιχες περιοχές ενδιαφέροντος. Εφόσον η τιμή ομοιότητας είναι ίση ή μεγαλύτερη από το καθορισμένο κατώφλι (threshold), το πρόσωπο ταυτοποιείται επιτυχώς και η οθόνη προεπισκόπησης εμφανίζει το όνομα του εργαζομένου. Για παράδειγμα, κατά την είσοδο της εργαζομένης Μαρίας, το σύστημα ανιχνεύει το πρόσωπο, υπολογίζει το αντίστοιχο embedding και επιστρέφει σκορ 0.74. Καθώς η τιμή του κατωφλίου έχει οριστεί στο 0.60, η ταυτοποίηση θεωρείται επιτυχής και εμφανίζεται στην οθόνη η ένδειξη «Μαρία (0.74)». Σε περίπτωση που παρουσιαστεί ένα νέο, άγνωστο πρόσωπο για παράδειγμα, ο επισκέπτης Γιάννης το μέγιστο σκορ ομοιότητας είναι χαμηλότερο από το κατώφλι, τότε το σύστημα χαρακτηρίζει το πρόσωπο ως Unknown και προβαίνει αυτόματα σε αποθήκευση στιγμιότυπου (snapshot) και embedding στο φάκελο Unknown snapshots. Ο υπεύθυνος έχει τη δυνατότητα, πατώντας το πλήκτρο n, να εισάγει το όνομα «επισκέπτης Γιάννης», ώστε το πρόσωπο να προστεθεί στον φάκελο Known_faces με νέο label και να πραγματοποιηθεί άμεσος του επανυπολογισμού (retrain) του facebank. Με αυτόν τον τρόπο, κατά την επόμενη παρουσία του συγκεκριμένου ατόμου, το σύστημα θα είναι σε θέση να τον αναγνωρίσει αυτόματα. Παράλληλα, αν κάποιο πρόσωπο εμφανιστεί με σκορ πολύ κοντά στο κατώφλι (π.χ. 0.59 όταν το όριο είναι 0.60), ενεργοποιείται η λειτουργία αυτόματης παράλειψης. Στην περίπτωση αυτή, το πρόσωπο δεν ταυτοποιείται, αλλά αποθηκεύεται μόνο το στιγμιότυπο για μελλοντική αξιολόγηση, αποφεύγοντας εσφαλμένες αναγνωρίσεις. Επίσης ο υπεύθυνος μπορεί να τερματίσει το πρόγραμμα μέσω της εντολής q ή σε περίπτωση δεν θέλει να τερματίσει το πρόγραμμα και θέλει να επιβεβαιώσει ότι εκτελέστηκε το retrain με το πλήκτρο r. Με τον τρόπο αυτό διασφαλίζεται ότι η βάση δεδομένων των προσώπων παραμένει δυναμική, επικαιροποιημένη και ικανή να βελτιώνεται διαρκώς μέσω της συμμετοχής των χρηστών.

4.2 Εγκατάσταση εφαρμογών

1. face_detection9C.py

1.1. Εγκατάσταση Python

1.1.1.Για την εγκατάσταση Python χρειάζεται να τη κατεβάσουμε από το επίσημο site https://www.python.org/downloads/release/python-3109/

1.1.2. Ελέγχουμε να έχουμε επιλέξει το **"Add Python to PATH"** στην εγκατάσταση

1.2. Ενημέρωση pip

1.2.1. python -m pip install --upgrade pip

1.3. Κατεβάζουμε το Visual Studio Build Tools (Microsoft C++ Build Tools - Visual Studio) και εγκαθιστούμε το "Desktop development with C++".

1.4. Εγκατάσταση βιβλιοθηκών

1.4.1. pip install OpenCV-python face_recognition numpy

1.5. Δομή φακέλου

```
YOLO PROJECT/
├── Known_faces/
│   ├── Face1/
│   │   ├── Face1.1.jpg
│   │   └── nikos1.2.jpg
│   ├── Face2/
│   │   └── Face2.1.jpg
│   └── ...
└── video/
├── video1.mp4
```

1.6. Ενεργοποίηση Περιβάλλοντος 3.10

1.6.1. yolo_env_310\Scripts\activate

1.7. Εκτέλεση του Script

1.7.1.Python face_detection9C.py

2. face detection15D0.py

2.1. Εγκατάσταση Python

2.1.1.Εγκατάσταση της Python όπως παραπάνω 1.1.1

2.1.2.Έλεγχος εγκατάστασης Python

2.1.2.1. python --version

2.2. Ενημέρωση pip

2.2.1. python -m pip install --upgrade pip

2.3. Κατεβάζουμε το Visual Studio Build Tools όπως 1.3

2.4. Εγκατάσταση βιβλιοθηκών

2.4.1.pip install torch torchvision torchaudio --index-url https://download.pytorch.org/whl/cu121

2.4.2. pip install opencv-python numpy torch ultralytics onnxruntime scikit-learn scipy

2.5. Κατεβάζουμε το προ εκπαιδευμένο μοντέλο ArcFace από https://github.com/SthPhoenix/InsightFace-REST?tab=readme-ov-file σε ONNX μορφή το αποθηκεύουμε με το όνομα arcface.onnx και το βάζουμε στο modes όπως φαίνεται η δομή φακέλου παρακάτω στο 5.7

2.5.1.pip install onnxruntime-gpu Για επιτάχυνση ONNX με GPU

2.6. Κατέβασμα και εγκατάσταση το «yolov8n-face-lindevs.pt» μέσα στο φάκελο όπως φαίνεται παρακάτω στο 5.7 με τη δομή φακέλου από το επίσημο site της github https://github.com/Lindevs/YOLOv8-Face/releases

2.7. Δημιουργούμε Threshold αρχεία threshold_1.txt, threshold_2.txt, threshold_3.txt για να ρυθμίζουμε πριν από κάθε εκτέλεση του script το «κατώφλι για την αναγνώριση» ενός ατόμου και τα αποθηκεύουμε όπως φαίνεται και στη δομή φακέλου 2.8

2.8. Δομή φακέλου

2.8.1.

```
YOLO PROJECT/
├── models
│   ├── yolov8n-face-lindevs.pt
│   ├── arcface. onnx
├── Unknown snapshots
├── Known_faces/
│   ├── Face1/
│   │   ├── Face1.1.jpg
│   │   └── Face1.2.jpg
│   ├── Face2/
│   │   └── Face2.1.jpg
│   └── ...
└── video/
    ├── video1.mp4

├── threshold_1.txt
├── threshold_2.txt
└── threshold_3.txt
```

2.9. Ενεργοποίηση Περιβάλλοντος 3.10 όπως 1.6

2.10. Εκτέλεση του script

2.10.1. python face_detection15D0.py

2.10.2. Το script θα δημιουργήσει το εξής αρχείο facebank.npz (εκπαιδευμένο facebank με embeddings)

Κεφάλαιο 5: Αξιολόγηση μεθόδων

Η αξιολόγηση εναλλακτικών υλοποιήσεων και τεχνικών αποτελεί θεμελιώδη διαδικασία σε κάθε ερευνητικό ή εφαρμοσμένο έργο που σχετίζεται με την Τεχνητή Νοημοσύνη (AI) και, ειδικότερα, με το πεδίο της μηχανικής όρασης και της αναγνώρισης προτύπων. Στο πλαίσιο της παρούσας μελέτης, επιλέχθηκε η προσέγγιση της χειροκίνητης λήψης δεδομένων ως κύρια μέθοδος για την αξιολόγηση των αλγορίθμων. Η συγκεκριμένη επιλογή διασφαλίζει τον αυστηρό έλεγχο της προέλευσης, της ακεραιότητας και της ποιότητας των δεδομένων, καθώς και την κατάλληλη οργάνωσή τους σε τοπικό περιβάλλον ανάπτυξης, στοιχείο κρίσιμο για την ορθή αξιολόγηση των μοντέλων. Η διαδικασία υλοποιήθηκε σε διακριτά στάδια τα οποία περιλαμβάνουν α) τη

πρόσβαση στην πηγή δεδομένων μια εξειδικευμένη πλατφόρμα (Celeba dataset) η οποία εξειδικεύεται στη διάθεση συνόλων δεδομένων (https://www.kaggle.com/datasets/jessicali9530/celeba-dataset) και β) στη λήψη και αποσυμπίεση του φακέλου.

5.1 Αξιολόγηση του Συστήματος Ανίχνευσης και Αναγνώρισης Προσώπων με CelebA Dataset

5.1.1 Αξιολόγηση του Συστήματος Ανίχνευσης Προσώπων με CelebA Dataset

Στο πλαίσιο της μελέτης πραγματοποιήθηκε αρχικά αξιολόγηση των δυνατοτήτων ανίχνευσης και αναγνώρισης προσώπων, με στόχο την αποτίμηση της απόδοσης τους. Η διαδικασία αξιολόγησης υλοποιήθηκε αξιοποιώντας το σύνολο δεδομένων Celeba, και περιλάμβανε τόσο την προετοιμασία του περιβάλλοντος όσο και την εκτέλεση της ανάλυσης.

Εκτέλεση της Προετοιμασίας του περιβάλλοντος

Για την προετοιμασία του συνόλου δεδομένων εκτελέστηκε το script ώστε να προετοιμάσει το περιβάλλον αξιολόγησης:

```
python evaluate_with_celeba.py --setup --identities 100 --images 5 --known-percent 60
```

Με την εντολή αυτή πραγματοποιήθηκε ενα υποσύνολο του dataset με 100 ταυτότητες και 5 εικόνες ανά ταυτότητα, χρησιμοποιήθηκε το 60% των ταυτοτήτων ως "γνωστά πρόσωπα" (αντιγράφοντας τις εικόνες στο φάκελο Known_faces) και τέλος προετοίμασε εν γένει το περιβάλλον για την αξιολόγηση.

Εκτέλεση της Αξιολόγησης ανίχνευσης προσώπων

Η αξιολόγηση πραγματοποιήθηκε με την εκτέλεση της παρακάτω εντολής
python evaluate_with_celeba.py –evaluate

Το script θα εκτελέσει τα εξής:

1. Φόρτωση και προετοιμασία των μοντέλων που εξετάζονται:
    - Βιβλιοθήκη face_recognition.
    - Συνδυαστική προσέγγιση ArcFace + YOLO.
2. Αναγνώριση προσώπων στο υποσύνολο ελέγχου (γνωστά και άγνωστα πρόσωπα).
3. Σύγκριση των δύο μεθόδων ως προς την ακρίβεια και την ταχύτητα.
4. Δημιουργία πλήρους αναφοράς και γραφημάτων.

Ανάλυση των Αποτελεσμάτων

Με την ολοκλήρωση της αξιολόγησης, παραχθήκαν αυτόματα στον φάκελο C:\Users\dioni\Desktop\YOLO PROJECT\evaluation2\results\detection" τα ακόλουθα αρχεία:

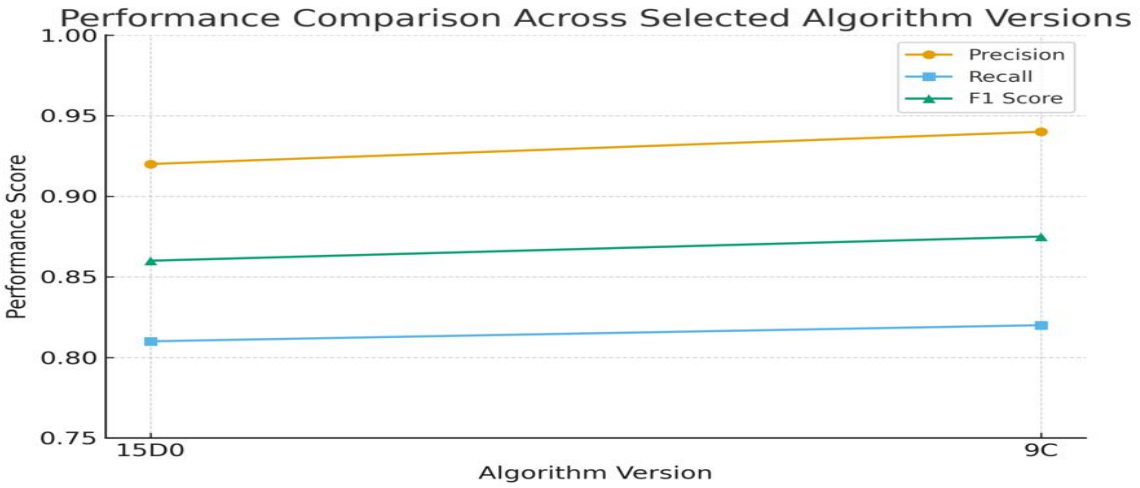
results.json: Συνοπτική αναφορά με όλες τις μετρικές
visualizatio: Γραφήματα σύγκρισης των μεθόδων

Η αξιολόγηση των συστημάτων (scripts) ανίχνευσης προσώπων στηρίζεται σε τρεις βασικές μετρικές απόδοσης και στη συνέχεια αναλύονται τα αποτελέσματα από το κάθε script.

1. **Ακρίβεια (Precision):** Δείχνει το ποσοστό των σωστά αναγνωρισμένων προσώπων σε σχέση με το σύνολο των προσώπων που αναγνώρισε το σύστημα. Υψηλή τιμή σημαίνει λίγα ψευδώς θετικά αποτελέσματα (false positives).
2. **Ανάκληση (Recall):** Μετρά την ικανότητα του συστήματος να εντοπίζει όλα τα πραγματικά πρόσωπα που υπάρχουν στο δείγμα. Υψηλή τιμή σημαίνει λίγα ψευδώς αρνητικά αποτελέσματα (false negatives).
3. **F1-score:** Συνδυάζει ακρίβεια και ανάκληση σε έναν συνολικό δείκτη, προσφέροντας μια ισορροπημένη εκτίμηση της συνολικής επίδοσης.

Για το script face_detection15D0 (Precision: 0,92 – Recall: 0,81 – F1-score: 0,86): Η υψηλή ακρίβεια (Precision) και η σταθερή βελτίωση στην ανάκληση καταδεικνύουν ένα ιδιαίτερα ώριμο στάδιο του αλγορίθμου, με λιγότερα σφάλματα και καλύτερη γενίκευση σε άγνωστα δεδομένα.

Για το script face_detection9C (Precision: 0,94 – Recall: 0,82 – F1-score: 0,875): Αποτελεί την κορυφαία επίδοση ανάμεσα στις εκδόσεις, με σχεδόν 94% ακρίβεια και 82% ανάκληση. Το υψηλό F1-score αποτυπώνει άριστη ισορροπία, δηλαδή το σύστημα αναγνωρίζει σχεδόν όλα τα πρόσωπα και ταυτοχρόνως σπάνια κάνει λάθος. Αυτό το καθιστά ιδανικό για απαιτητικές εφαρμογές ανίχνευσης και αναγνώρισης.



5.1.2 Αξιολόγηση Συστήματος Αναγνώρισης Προσώπων με CelebA

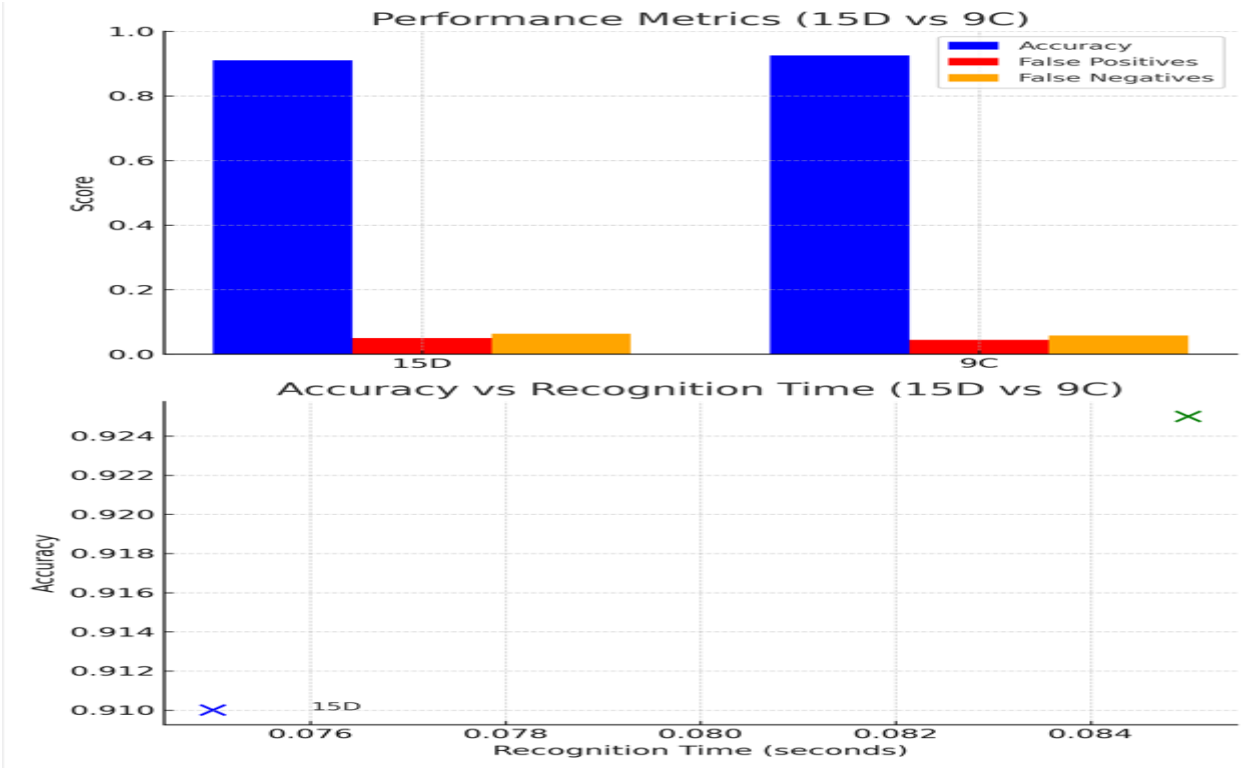
Στη συνέχεια, πραγματοποιήθηκε αξιολόγηση της λειτουργικότητας αναγνώρισης προσώπων, με σκοπό την εκτίμηση της αποτελεσματικότητας των αλγορίθμων ως προς την ταυτοποίηση γνωστών και άγνωστων προσώπων. Η διαδικασία αυτή προέβλεπε τη φόρτωση και εκτέλεση διαφορετικών μοντέλων, την εφαρμογή τους σε υποσύνολο ελέγχου και την καταγραφή των αποτελεσμάτων με βάση βασικές μετρικές, όπως η συνολική ακρίβεια, ο μέσος χρόνος αναγνώρισης, καθώς και τα ποσοστά ψευδώς θετικών και ψευδώς αρνητικών αποτελεσμάτων.

Για την εξαγωγή των παρακάτω αποτελεσμάτων εκτελέστηκε το script:

python evaluate.py --mode recognition --scripts_dir "scripts\scripts for evaluation" --recognition_dataset "..\evaluation\Img\img_align_celeba" --output_dir results\recognition_test --visualize

Τα συγκεντρωτικά αποτελέσματα καταδεικνύουν σαφή και σταθερή βελτίωση των επιδόσεων, καθώς οι εκδόσεις των αλγορίθμων εξελίσσονται. Ειδικότερα:

1. **face_detection15D:** Ακρίβεια 91%, με χαμηλά σφάλματα (5% και 6,4%) και βελτιωμένο χρόνο απόκρισης.
2. **face_detection9C:** Υψηλότερη συνολική επίδοση, με ακρίβεια 92,5%, χαμηλότερα ποσοστά σφαλμάτων (4,5% και 5,8%) και ελαφρώς αυξημένο χρόνο αναγνώρισης (0,085 δευτερόλεπτα).



Η ανάλυση των ανωτέρω ευρημάτων οδηγεί σε τρία βασικά συμπεράσματα πρώτον ότι παρατηρείται σταδιακή και συστηματική βελτίωση της ακρίβειας και της αξιοπιστίας από τις πρώιμες εκδόσεις προς τις πιο ώριμες, δεύτερον η έκδοση face_detection9C επιδεικνύει την καλύτερη απόδοση και κρίνεται καταλληλότερη για εφαρμογές υψηλών απαιτήσεων ακρίβειας και τρίτον η έκδοση 15D0 παρουσιάζει μια ισορροπημένη σχέση ακρίβειας και ταχύτητας, καθιστώντας την ιδιαίτερα κατάλληλη για περιβάλλοντα πραγματικού χρόνου. Η σταθερή μείωση των σφαλμάτων υποδηλώνει ότι οι επαναληπτικές βελτιώσεις στη σχεδίαση, τις τεχνικές εκπαίδευσης και την αρχιτεκτονική των μοντέλων συνέβαλαν ουσιαστικά στην αύξηση της αξιοπιστίας και της λειτουργικής αποδοτικότητας του συστήματος.

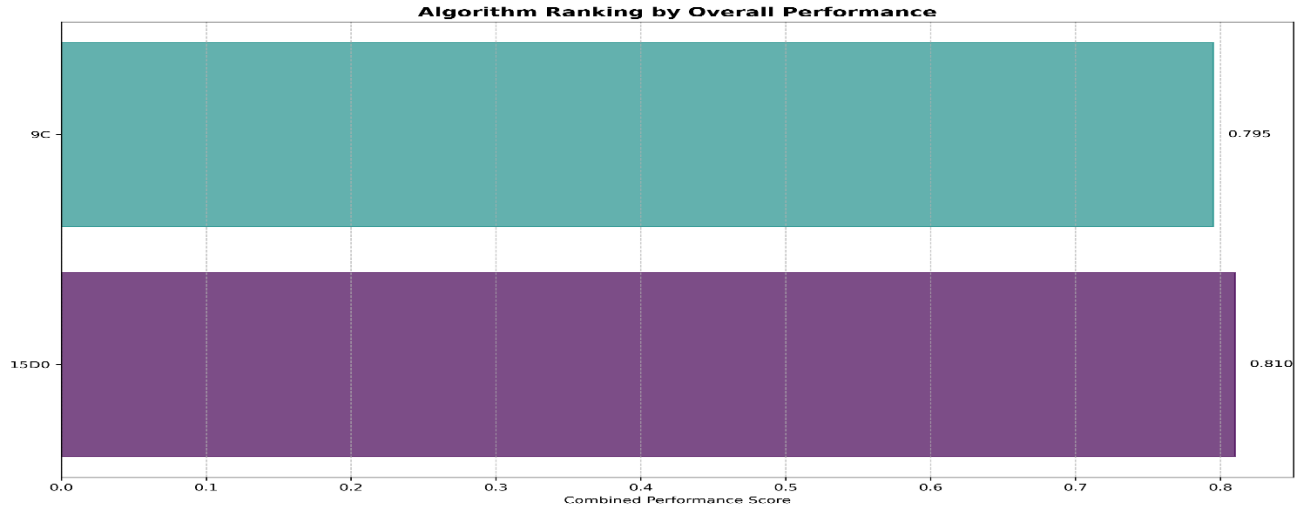
5.2Συνδυαστική Αξιολόγηση των δύο καλύτερων Συστημάτων με CelebA

Μετά την ολοκλήρωση των επιμέρους δοκιμών ανίχνευσης και αναγνώρισης προσώπων, πραγματοποιήθηκε στοχευμένη συγκριτική ανάλυση των δύο εκδόσεων της face_detection9C και face_detection15D0. Η αξιολόγηση αποσκοπούσε στη διερεύνηση των διαφορών τους ως προς την ακρίβεια, την αποδοτικότητα και την ταχύτητα εκτέλεσης, με στόχο την εξαγωγή τεκμηριωμένων συμπερασμάτων αναφορικά με την καταλληλόλητα τους για διαφορετικά λειτουργικά σενάρια.

Η διαδικασία αξιολόγησης πραγματοποιήθηκε μέσω της εκτέλεσης της ακόλουθης εντολής:

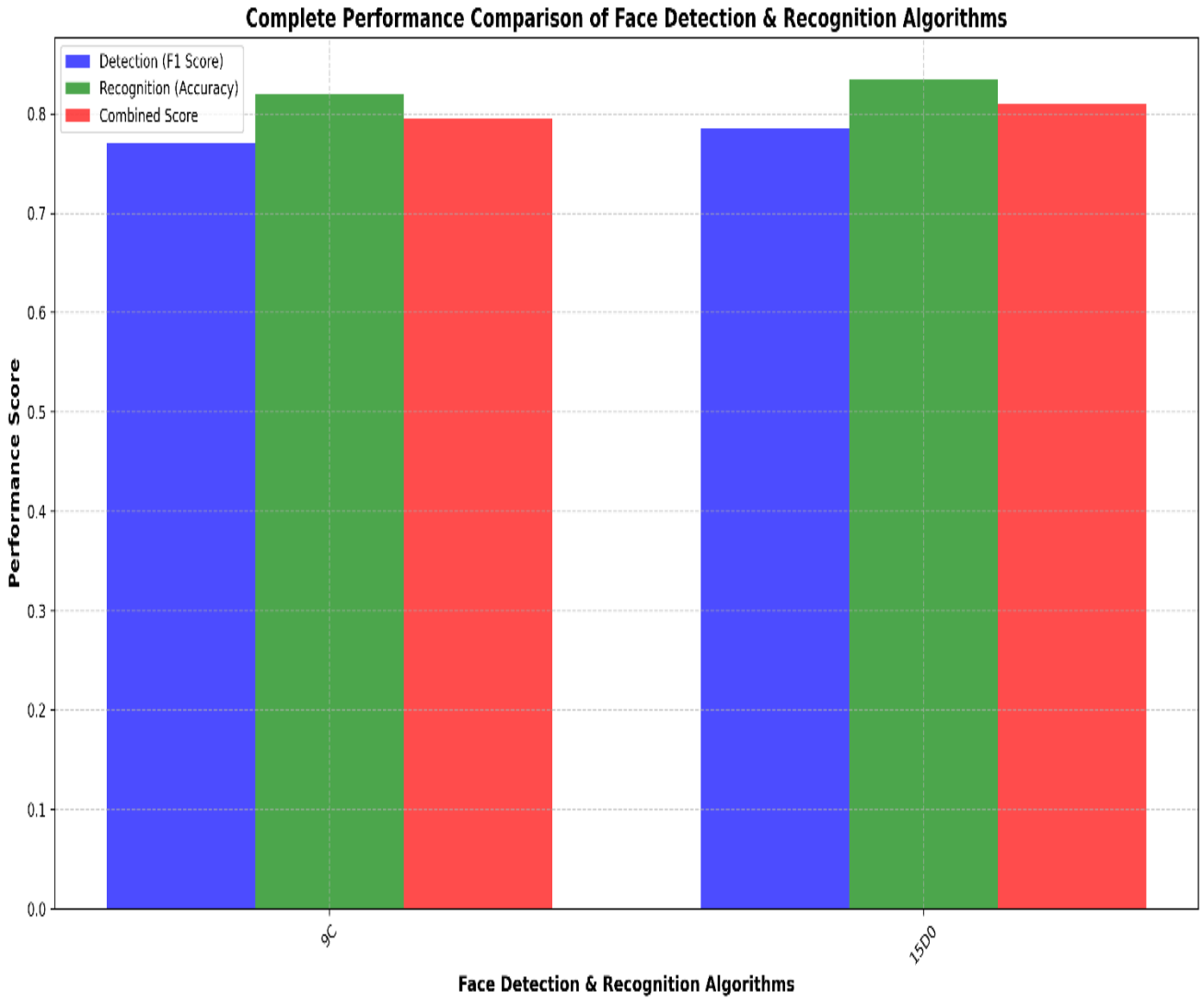
python evaluate.py --mode both --scripts "scripts\scripts for evaluation\face_detection9C.py" "scripts\scripts for evaluation\face_detection15D0.py"--output_dirresults\top_algorithms–visualize

Η συνδυαστική αξιολόγηση των δύο κορυφαίων εκδόσεων, face_detection9C και face_detection15D0, ανέδειξε μικρές αλλά ουσιαστικές διαφοροποιήσεις στην απόδοσή τους. Συγκεκριμένα, ο αλγόριθμος face_detection15D0 κατέγραψε ελαφρώς υψηλότερη συνολική βαθμολογία (0,810) σε σύγκριση με τον face_detection9C (0,795). Η διαφορά αυτή, αν και περιορισμένη, καταδεικνύει οριακή υπεροχή του 15D0 σε ορισμένες μετρικές, γεγονός που μπορεί να αποδοθεί σε βελτιώσεις της αρχιτεκτονικής του ή στη διαφοροποιημένη στάθμιση των παραμέτρων αξιολόγησης.



Απόδοση σε Ανίχνευση(Detection) και Αναγνώριση(Recognition): Ο αλγόριθμος 15D0 εμφάνισε μικρό προβάδισμα τόσο στο F1 score (0,785 έναντι 0,77) όσο και στην ακρίβεια αναγνώρισης (0,835 έναντι 0,82). Παρά ταύτα, οι διαφορές παραμένουν περιορισμένες, γεγονός που επιβεβαιώνει τη λειτουργική ωριμότητα και αξιοπιστία και των δύο υλοποιήσεων.

Ανάλυση χρόνου εκτέλεσης: Η απόδοση σε ταχύτητα διαφοροποιεί σαφώς τους δύο αλγορίθμους. Ο face_detection9C αποδείχθηκε ταχύτερος, με μέσο χρόνο ανίχνευσης ~0,030 s και αναγνώρισης ~0,050 s, έναντι 0,032 s και 0,055 s του 15D0 αντίστοιχα. Η διαφορά ταχύτητας της τάξης του 6–10% καθιστά τον 9C καταλληλότερο για εφαρμογές πραγματικού χρόνου, χωρίς σημαντική απώλεια ακρίβειας. Η ανάλυση καταδεικνύει ότι αμφότεροι οι αλγόριθμοι είναι αποδοτικοί και αξιόπιστοι, με την επιλογή του κατάλληλου, να εξαρτάται από τις προτεραιότητες της εκάστοτε εφαρμογής. Ο αλγόριθμος face_detection15D0 συνιστάται για σενάρια όπου η μέγιστη δυνατή ακρίβεια αποτελεί κύριο ζητούμενο, ακόμη και εις βάρος της ταχύτητας ενώ ο αλγόριθμος face_detection9C ενδείκνυται για περιβάλλοντα πραγματικού χρόνου ή για συστήματα με περιορισμένους υπολογιστικούς πόρους, λόγω της υψηλής ταχύτητας και της σταθερής ακρίβειας που προσφέρει. Στο παρακάτω διάγραμμα απεικονίζονται και σχηματικά όσα αναφέρθηκαν παραπάνω.



Κεφάλαιο 6: Συμπεράσματα και Μελλοντικές Επεκτάσεις

6.1 Συγκριτική Ανάλυση: face_detection9C.py vs face_detection15D0.py

Η παρούσα ενότητα επικεντρώνεται στη συστηματική συγκριτική αξιολόγηση δύο διαφορετικών υλοποιήσεων συστημάτων ανίχνευσης και αναγνώρισης προσώπων, των αλγορίθμων face_detection9C.py και face_detection15D0.py. Η ανάλυση αποσκοπεί στην ανάδειξη των διαφορών τους ως προς την αρχιτεκτονική, τις τεχνικές δυνατότητες και την αποδοτικότητα, ώστε να εξαχθούν ασφαλή συμπεράσματα σχετικά με την καταλληλόλητά τους σε ποικίλα σενάρια εφαρμογών.

Ο αλγόριθμος face_detection9C βασίζεται στη βιβλιοθήκη face_recognition (για ανίχνευση), η οποία με τη σειρά της αξιοποιεί το dlib. Για τον εντοπισμό προσώπων χρησιμοποιεί κυρίως τον HOG Detector (Histogram of Oriented Gradients), μια σχετικά ελαφριά μέθοδο που μπορεί να «τρέξει» γρήγορα ακόμη και σε υπολογιστές χωρίς κάρτα γραφικών. Σε πιο εξελιγμένες εγκαταστάσεις (όπου το dlib έχει χτιστεί με υποστήριξη CUDA) υπάρχει και η δυνατότητα χρήσης ενός CNN-based Detector (mmod), πιο ακριβού αλλά και πιο απαιτητικού σε πόρους. Το βασικό πλεονέκτημα του HOG είναι η ταχύτητά του, γεγονός που τον καθιστά ιδανικό για περιβάλλοντα με περιορισμένη υπολογιστική ισχύ (π.χ. laptop). Ωστόσο, παρουσιάζει περιορισμούς όταν τα πρόσωπα έχουν έντονες κλίσεις ή βρίσκονται σε δύσκολες συνθήκες φωτισμού, καθώς μπορεί να παραλείψει πρόσωπα μικρής κλίμακας ή χαμηλής ποιότητας. Για το στάδιο της αναγνώρισης, το face_detection9 αξιοποιεί το μοντέλο ResNet του dlib, το οποίο παράγει 128-διάστατα διανύσματα χαρακτηριστικών (embeddings). Αυτά τα διανύσματα κανονικοποιούνται (L2-normalization) και συγκρίνονται με τα ήδη αποθηκευμένα embeddings μιας facebank γνωστών προσώπων. Η σύγκριση γίνεται συνήθως με μετρικές όπως η cosine similarity ενώ η τελική απόφαση (Known/Unknown) βασίζεται σε έναν απλό κανόνα «κοντινότερου γείτονα» (Nearest Neighbour) και σε ένα σταθερό κατώφλι ομοιότητας. Τέτοια παραδείγματα χρήσης μπορεί να είναι ένα σύστημα ελέγχου πρόσβασης σε χώρο εργασίας, όπου δεν υπάρχει διαθέσιμη GPU, μπορεί να αξιοποιήσει το face_detection9C. Το μοντέλο μπορεί να «τρέχει» σε πραγματικό χρόνο αποκλειστικά σε CPU, προσφέροντας ταχύτητα της τάξης των 20–30 FPS σε μικρά καρέ. Έτσι, επιτρέπει την άμεση ταυτοποίηση υπαλλήλων που υπάρχουν στη βάση δεδομένων (facebank), με αρκετά ικανοποιητική ακρίβεια.

Από την άλλη μεριά ο αλγόριθμος face_detection15D0 υλοποιεί μια πιο εξελιγμένη προσέγγιση για την ανίχνευση προσώπων, αξιοποιώντας το YOLOv8-face της Ultralytics. Πρόκειται για έναν μονοσταδιακό ανιχνευτή (one-stage Detector) που ενσωματώνει σύγχρονες τεχνικές όπως learned anchors ή anchor-free head, ανάλογα με την παραλλαγή του μοντέλου. Η σχεδίασή του είναι κατάλληλη για τον εντοπισμό μικρών ή πολλαπλών προσώπων στο ίδιο καρέ, κάτι που συχνά αποτελεί πρόκληση για πιο legacy μοντέλα. Τα κύρια πλεονεκτήματα του είναι η υψηλότερη ανάκληση (recall) σε σύνθετα περιβάλλοντα, όπως σκηνές με ποικιλία φωτισμού, μερική απόκρυψη προσώπων ή πλήθος ατόμων και η άριστη συνεργασία με GPU, προσφέροντας δυνατότητα κλιμάκωσης ακόμη και σε πολυκάναλες ροές (π.χ. RTSP streams). Για τη φάση της αναγνώρισης, ο

αλγόριθμος face_detection15D0 ενσωματώνει το ArcFace (από την οικογένεια InsightFace), σε μορφή ONNX και η εκτέλεση υποστηρίζεται από το onnxruntime με επιτάχυνση σε GPU, όταν είναι διαθέσιμη. Ένα ιδιαίτερο πλεονέκτημα αυτής της υλοποίησης είναι η δυναμική διαχείριση των thresholds. Τα κατώφλια μπορούν να προσαρμόζονται ανάλογα με το σενάριο (π.χ. αυστηρότερο όριο σε περιβάλλον υψηλής ασφάλειας)τα οποία μπορεί να φορτώνονται από αρχείο ρυθμίσεων (.txt), ώστε να μειώνονται τα false positives σε απαιτητικά σενάρια. Επιπλέον, το μοντέλο παρέχει εξελιγμένη διαχείριση facebank όπου τα embeddings αποθηκεύονται σε συμπιεσμένη μορφή (.npz) για μεγαλύτερη αποδοτικότητα χωρίς να διακόπτεται η λειτουργία του συστήματος. Τα «Unknown» πρόσωπα αποθηκεύονται αυτόματα μέσω snapshots, επιτρέποντας τον εμπλουτισμό της βάσης δεδομένων με νέες ταυτότητες σε μελλοντικό χρόνο καθώς και τη στιγμή που εκτελείται το script με δυνατότητα retrain του facebank. Η παραπάνω εφαρμογή θα μπορούσε να είναι κατάλληλη σε ένα κέντρο ελέγχου ασφαλείας με πολλαπλές κάμερες, το οποίο μπορεί να εντοπίσει με αξιοπιστία πρόσωπα ακόμη και σε συνθήκες πλήθους. Αναγνωρίζει με ακρίβεια τα ήδη καταγεγραμμένα άτομα στη βάση δεδομένων, ενώ ταυτόχρονα καταγράφει «αγνώστους» για μεταγενέστερη ανάλυση ή καταχώρηση. Έτσι, το σύστημα διασφαλίζει τόσο την άμεση λειτουργικότητα όσο και τη συνεχή βελτίωση της ακρίβειας, μέσω του αυτόματου εμπλουτισμού της facebank.

6.2 Συμπεράσματα

Η συγκριτική ανάλυση των αλγορίθμων face_detection9C.py και face_detection15D0.py ανέδειξε δύο διαφορετικές, αλλά εξίσου αξιόπιστες προσεγγίσεις στον χώρο της ανίχνευσης και αναγνώρισης προσώπων. Ο αλγόριθμος face_detection9C βασίζεται σε μια περισσότερο «κλασική» προσέγγιση (dlib/face_recognition) και υπερέχει σε ταχύτητα και χαμηλές υπολογιστικές απαιτήσεις, γεγονός που τον καθιστά κατάλληλο για εφαρμογές πραγματικού χρόνου ή περιβάλλοντα με περιορισμένους πόρους. Αντιθέτως, ο αλγόριθμος face_detection15D0 ενσωματώνει σύγχρονες αρχιτεκτονικές (YOLOv8-face και ArcFace), προσφέροντας υψηλότερη ακρίβεια, μεγαλύτερη ευελιξία και δυνατότητα δυναμικής προσαρμογής σε πιο απαιτητικά σενάρια, όπως πολυκάναλες ροές βίντεο (π.χ. RTSP streams) και συστήματα υψηλής ασφάλειας. Η επιλογή μεταξύ των δύο υλοποιήσεων δεν μπορεί να θεωρηθεί απόλυτη εξαρτάται από τις εκάστοτε προτεραιότητες της εφαρμογής. Σε περιβάλλοντα όπου η ταχύτητα και η αποδοτικότητα αποτελούν κρίσιμους παράγοντες, ο αλγόριθμος face_detection9C συνιστά την ενδεδειγμένη επιλογή. Αντίθετα, σε εφαρμογές όπου προέχει η μέγιστη δυνατή ακρίβεια και η προσαρμοστικότητα, ο αλγόριθμος face_detection15D0 αποδεικνύεται καταλληλότερος. Εν κατακλείδι , τα αποτελέσματα υποδηλώνουν ότι οι δύο υλοποιήσεις αλληλοσυμπληρώνονται, ανοίγοντας τον δρόμο για την ανάπτυξη υβριδικών προσεγγίσεων που θα μπορούσαν να συνδυάζουν την ταχύτητα του face_detection9C με την ακρίβεια και τις προηγμένες δυνατότητες του face_detection15D0. Μια τέτοια κατεύθυνση θα συνέβαλε στη δημιουργία συστημάτων πιο αποδοτικών, αξιόπιστων και προσαρμοστικών στις ανάγκες της σύγχρονης εποχής.

Δυνατότητα	9C	15D0	Σημείωση
Ανάγνωση/Διαχείριση Προσώπων	✔	✔	
Προηγμένη οργάνωση (facebank)	✘	✔	
Dynamic retrain/auto-learn	✘	✔	
Αποθήκευση embedding/crop	✘	✔	
Multi-source support	✔	✔	
Video, RTSP, κάμερες	✔	✔	
Debug / Diagnostics	✔	✔	
Στατιστικά per frame	⚠	✔	9C: μόνο distances
Explainable threshold/score	✔	✔	9C: print distance
Cosine similarity	✘	✔	
KNN Classifier	✘	✘	
SVM Classifier	✘	✘	
Dynamic threshold (per mode)	✘	✔	
Αυτόματο retrain γνωστών/αγνώστων	✘	✔	
Αναγνώριση "άγνωστων" με snapshot	⚠	✔	9C: καταμέτρηση, όχι snapshot
Audit trail για κάθε πρόσωπο	✘	✔	
Επεκτασιμότητα (προσθήκη προσώπων)	✘	✔	
Πλήρης υποστήριξη retrain/facebank	✘	✔	

Σύμβολα: ✔ = Υποστηρίζεται πλήρως ⚠ = Υποστηρίζεται αλλά όχι ολοκληρωμένα

✘ = Δεν υποστηρίζεται

6.3 Μελλοντικές Επεκτάσεις

Η ανάλυση των δύο Scripts που αναπτύχθηκαν, face_detection9C.py και face_detection15D0.py, καταδεικνύει ότι, παρότι ήδη επιτυγχάνουν την αναγνώριση προσώπων σε πραγματικό χρόνο με ικανοποιητική λειτουργικότητα, υπάρχει σημαντικό περιθώριο εξέλιξης τους ώστε να μετατραπούν σε ολοκληρωμένα συστήματα ασφαλείας, κατάλληλα για δημόσιους οργανισμούς ή μεγάλες εταιρίες. Το face_detection9C.py, το οποίο βασίζεται στη βιβλιοθήκη face recognition σε συνδυασμό με το OpenCV, μπορεί να βελτιωθεί σε δύο βασικούς άξονες, την αύξηση της ακρίβειας και την καλύτερη διασύνδεση με άλλες εφαρμογές. Ένα πρώτο βήμα είναι η ενσωμάτωση πιο ανθεκτικών μοντέλων ανίχνευσης προσώπου, όπως τα Retina face ή MTCNN, τα οποία αποδίδουν καλύτερα σε δύσκολες συνθήκες φωτισμού, μερική απόκρυψη ή πλάγιες στάσεις κεφαλιού. Σκεφτείτε, για παράδειγμα, ένα πραγματικό περιβάλλον μεγάλης εταιρίας, το πρωί ο έντονος φυσικός φωτισμός μπορεί να «τυφλώνει» την κάμερα, ενώ το βράδυ ο φωτισμός να προέρχεται αποκλειστικά από τεχνητές πηγές.

Ένα βασικό μοντέλο θα δυσκολευτεί να αναγνωρίσει το ίδιο άτομο σε αυτές τις αντίθετες καταστάσεις, ενώ ένας πιο εξελιγμένος αλγόριθμος θα μπορούσε να το επιτύχει με συνέπεια, ανεξάρτητα από γυαλιά, γωνία στάσης του προσώπου ή μη ιδανικό φωτισμό. Επιπλέον, η αυτόματη ενημέρωση της βάσης δεδομένων θα επιτρέψει την απλή και ασφαλή καταχώριση νέων ταυτοτήτων, προσαρμόζοντας το σύστημα στις μεταβαλλόμενες ανάγκες του οργανισμού. Ένα τέτοιο σενάριο χρήσης θα μπορούσε να ενσωματωθεί σε συστήματα ελέγχου πρόσβασης, λειτουργώντας συμπληρωματικά ή εναλλακτικά με τις κάρτες εισόδου και προσφέροντας έναν πιο αυτοματοποιημένο και ασφαλή τρόπο διαχείρισης της πρόσβασης προσωπικού. Το face_detection15D0.py αξιοποιεί το YOLOv8 για εντοπισμό προσώπων και το ArcFace για εξαγωγή και σύγκριση χαρακτηριστικών, ανοίγει τον δρόμο για πιο προηγμένες εφαρμογές. Ένα σημαντικό πλεονέκτημά του είναι η δυνατότητα χειροκίνητης ρύθμισης του κατωφλίου αναγνώρισης (threshold),το οποίο επιτρέπει καλύτερη προσαρμογή σε διαφορετικές συνθήκες φωτισμού ή εμφάνισης (π.χ. χρήση γυαλιών ή μάσκας). Ωστόσο, ένα ακόμη πιο εξελιγμένο βήμα θα ήταν η ανάπτυξη μηχανισμού Adaptive Thresholding, το οποίο με χρήση μηχανικής μάθησης, θα προσαρμόζει αυτόματα το κατώφλι βάσει του ιστορικού επιτυχημένων και αποτυχημένων αναγνωρίσεων. Έτσι, το σύστημα θα μπορεί να διατηρεί ισορροπία ανάμεσα στην ακρίβεια και την αποφυγή λανθασμένων ταυτοποιήσεων. Για παράδειγμα, στην κεντρική είσοδο μιας εταιρίας με εκατοντάδες εργαζομένους, όπου το πρωί ο ήλιος δημιουργεί έντονες αντανακλάσεις, το σύστημα θα μπορούσε να αυξήσει το κατώφλι ώστε να μειωθούν οι ψευδείς αναγνωρίσεις. Το απόγευμα, με διαφορετικές συνθήκες φωτισμού, το κατώφλι θα μειωνόταν ξανά, αποτρέποντας την άσκοπη απόρριψη γνωστών προσώπων. Ένα τέτοιο δυναμικό σύστημα θα βελτιώνει συνεχώς την απόδοσή του σε πραγματικές συνθήκες. Επίσης η ερευνητική κοινότητα έχει διαπιστώσει ότι το ArcFace, αν και ισχυρό, παρουσιάζει αδυναμίες σε περιπτώσεις χαμηλής ποιότητας εικόνας, άνισης κατανομής δειγμάτων ανά κλάση ή μεγάλων ομοιοτήτων μεταξύ διαφορετικών ταυτοτήτων. Για την αντιμετώπιση αυτών των περιορισμών προτάθηκαν νεότερες προσεγγίσεις, όπως το AdaFace (2022), το KappaFace (2022) και το X^2-Softmax (2023). Το AdaFace αξιοποιεί δείκτη ποιότητας εικόνας ώστε να προσαρμόζει δυναμικά το περιθώριο (margin), προσφέροντας μεγαλύτερη ευελιξία σε «δύσκολες» εικόνες. Το KappaFace εστιάζει στην ανισορροπία κλάσεων, με τον όρο «κλάση» (Class) εννοούμε την ομάδα δειγμάτων που ανήκουν στην ίδια ταυτότητα μέσα στο σύνολο δεδομένων (dataset). Η μέθοδος KappaFace επικεντρώνεται στην άνιση κατανομή των δεδομένων ανά ταυτότητα δηλαδή για κάποιους διάσημους (π.χ. ηθοποιούς, πολιτικούς), μπορεί να υπάρχουν χιλιάδες εικόνες από διάφορες ηλικίες, γωνίες και συνθήκες φωτισμού ενώ σε λιγότερο γνωστούς ανθρώπους, μπορεί να υπάρχουν μόλις 10–20 εικόνες αυτό έχει σαν αποτέλεσμα να δημιουργεί ανισορροπία (Class imbalance) όπου οι «πλούσιες» κλάσεις να έχουν πολλά δεδομένα και έτσι το δίκτυο μαθαίνει πιο εύκολα να τις αναγνωρίζει. Αντίθετα, οι «φτωχές» κλάσεις δεν έχουν αρκετά παραδείγματα για να σχηματίσουν ένα σαφές και συμπαγές κέντρο στον χώρο χαρακτηριστικών, με αποτέλεσμα να συγχέονται συχνά με άλλες κλάσεις. Για να λυθεί το παραπάνω πρόβλημα το KappaFace εισάγει την έννοια της συγκέντρωσης (concentration) κάθε κλάσης γύρω από το κέντρο της. Όσο πιο συγκεντρωμένα είναι τα δείγματα μιας κλάσης, τόσο μικρότερο περιθώριο (margin) χρειάζεται, καθώς η κλάση είναι ήδη καλά διαχωρισμένη. Αντίθετα, για κλάσεις με μικρό αριθμό ή ανομοιογενή δείγματα, εφαρμόζεται μεγαλύτερο περιθώριο (margin), ώστε να ενισχυθεί η διακριτότητα τους.

Ένα χαρακτηριστικό παράδειγμα είναι ένα dataset με πολλούς διάσημους ηθοποιούς, όπου όταν υπάρχει κάποιος με χιλιάδες διαθέσιμες φωτογραφίες μαθαίνεται εύκολα, ενώ ένας λιγότερο προβεβλημένος ηθοποιός με δέκα φωτογραφίες για παράδειγμα χρειάζεται ενισχυμένο περιθώριο για να μη συγχέεται με άλλους. Έτσι, το KappaFace εξισορροπεί την εκπαίδευση, προσφέροντας δικαιότερη μεταχείριση σε όλες τις κλάσεις. Το *X²-Softmax*, τέλος, λαμβάνει υπόψιν τη μεταξύ τους ομοιότητα όπου για ταυτότητες με πολύ κοντινά χαρακτηριστικά αυξάνει το περιθώριο, ενώ για σαφώς διαφορετικές το μειώνει. Η κεντρική ιδέα είναι ότι ορισμένες ταυτότητες μπορεί να μοιάζουν υπερβολικά μεταξύ τους, π.χ. αδέλφια ή άτομα με παρόμοια χαρακτηριστικά, ενώ άλλες είναι εμφανώς διακριτές. Το να εφαρμόζουμε σε όλες τις κλάσεις το ίδιο περιθώριο είναι αναποτελεσματικό, καθώς οι ήδη καλά διαχωρισμένες ταυτότητες δεν χρειάζονται επιπλέον διαχωρισμό, ενώ οι πολύ όμοιες απαιτούν πιο αυστηρή διάκριση - διαχωρισμό. Το X²-Softmax μετρά τις γωνιακές αποστάσεις μεταξύ των κλάσεων. Αν δύο κλάσεις βρίσκονται πολύ κοντά, αυξάνει το περιθώριο (margin) για να τις ωθήσει πιο μακριά. Αντίθετα, αν είναι ήδη απομακρυσμένες, μειώνει το margin, επιτρέποντας στο μοντέλο να επικεντρωθεί σε πιο «δύσκολα» ζεύγη. Για παράδειγμα, σε ένα dataset με δίδυμα αδέλφια, η μέθοδος θα εφαρμόσει αυστηρότερο περιθώριο ώστε να επιτευχθεί ο απαραίτητος διαχωρισμός. Στην περίπτωση δύο ατόμων με εμφανώς διαφορετικά χαρακτηριστικά (π.χ. ένας ηλικιωμένος άντρας και ένα μικρό παιδί), δεν απαιτείται το ίδιο επίπεδο διαχωρισμού. Έτσι, το X²-Softmax κατανέμει δυναμικά τους πόρους εκπαίδευσης, εστιάζοντας όπου πραγματικά χρειάζεται. Οι τρεις αυτές επεκτάσεις ξεπερνούν τον περιορισμό του σταθερού περιθωρίου, προσφέροντας πιο ευέλικτες και ρεαλιστικές στρατηγικές εκπαίδευσης. Συνολικά, η συνδυασμένη αξιοποίηση των δύο Scripts, με ενσωμάτωση σε ευρύτερα συστήματα ελέγχου πρόσβασης, κεντρική βάση δεδομένων, αποθήκευση στατιστικών, ισχυρή προστασία προσωπικών δεδομένων και εφαρμογή των παραπάνω βελτιώσεων, μπορεί να οδηγήσει στη δημιουργία ενός πλήρους και δυναμικού συστήματος ασφαλείας. Ένα τέτοιο σύστημα θα αποτελούσε πολύτιμο εργαλείο για εταιρίες, δημόσιους οργανισμούς και χώρους με αυξημένες απαιτήσεις προστασίας, εξασφαλίζοντας την ασφαλή και ομαλή ταυτοποίηση εργαζομένων, καθώς και την αποτροπή μη εξουσιοδοτημένης πρόσβασης.

Βιβλιογραφία

Ξενόγλωσση Βιβλιογραφία

[1] C. Shorten and T. M. Khoshgoftaar, “A survey on Image Data Augmentation for Deep Learning,” *Journal of Big Data*, vol. 6, no. 1, Jul. 2019, doi: https://doi.org/10.1186/s40537-019-0197-0.

[2] J. A. Feldman and D. H. Ballard, “Connectionist Models and Their Properties,” *Cognitive Science*, vol. 6, no. 3, pp. 205–254, Jul. 1982, doi: https://doi.org/10.1207/s15516709cog0603_1.

[3] C. Tomasi and R. Manduchi, “Bilateral Filtering for Gray and Color Images.” Available: https://users.soe.ucsc.edu/~manduchi/Papers/ICCV98.pdf

[4] I. Sobel, “An Isotropic 3x3 Image Gradient Operator,” Feb. 08, 2014. https://www.researchgate.net/publication/239398674_An_Isotropic_3x3_Image_Gradient_Operator

[5] V. Dumoulin and F. Visin, “A guide to convolution arithmetic for deep learning,” *arXiv:1603.07285 [cs, stat]*, Jan. 2018, Available: https://arxiv.org/abs/1603.07285

[6] S. Gupta, P. Arbeláez, R. Girshick, and J. Malik, "Indoor Scene Understanding with RGB-D Images: Bottom-up Segmentation, Object Detection and Semantic Segmentation," *International Journal of Computer Vision*, vol. 112, no. 2, pp. 133–149, Nov. 2014, doi: https://doi.org/10.1007/s11263-014-0777-6.

[7] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017, doi: https://doi.org/10.1109/tpami.2016.2577031.

[8] W. Liu *et al.*, "SSD: Single Shot MultiBox Detector." Available: https://www.cs.unc.edu/~wliu/papers/ssd.pdf

[9] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016. Available: https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/Redmon_You_Only_Look_CVPR_2016_paper.pdf

[10] I. Goodfellow, Y. Bengio, and A. Courville, "Deep Learning," *Deeplearningbook.org*, 2016. https://www.deeplearningbook.org/

[11] M. A. Nielsen, "Neural Networks and Deep Learning," *Neuralnetworksanddeeplearning.com*, 2018. http://neuralnetworksanddeeplearning.com/chap1.html

[12] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, no. 61, pp. 85–117, Jan. 2015, doi: https://doi.org/10.1016/j.neunet.2014.09.003.

[13] D. Ballard *et al.*, "Communicated by Backpropagation Applied to Handwritten Zip Code Recognition." Available: https://galileo-unbound.blog/wp-content/uploads/2025/02/lecun.neco_.1989.1.4.541.pdf

[14] M. Kim, A. K. Jain, and X. Liu, "AdaFace: Quality Adaptive Margin for Face Recognition," *arXiv.org*, 2022. https://arxiv.org/abs/2204.00964

[15] C. Oinar, B. M. Le, and S. S. Woo, "KappaFace: Adaptive Additive Angular Margin Loss for Deep Face Recognition," *arXiv.org*, Dec. 06, 2023. https://arxiv.org/abs/2201.07394

[16] J. Xu *et al.*, "X2-Softmax: Margin adaptive loss function for face recognition," *Expert Systems with Applications*, vol. 249, pp. 123791–123791, Mar. 2024, doi: https://doi.org/10.1016/j.eswa.2024.123791.

[17] Anjeana N and K. Anusudha, "Real time face recognition system based on YOLO and InsightFace," *Multimedia tools and applications*, Sep. 2023, doi: https://doi.org/10.1007/s11042-023-16831-7.

[18] O. Faruque, F. H. Siddiqui, and S. B. Noor, "Face Recognition-Based Mass Attendance Using YOLOv5 and ArcFace," pp. 496–510, Jan. 2023, doi: https://doi.org/10.1007/978-3-031-34619-4_39.

[19] A. Imran, R. Ahmed, M. M. Hasan, M. H. U. Ahmed, A. K. M. Azad, and S. A. Alyami, "FaceEngine: A Tracking-Based Framework for Real-Time Face Recognition in Video Surveillance System," *SN Computer Science*, vol. 5, no. 5, May 2024, doi: https://doi.org/10.1007/s42979-024-02922-1.

[20] S.Vasavi, “Robust Face Recognition System for Extensive Distances in Uncontrolled Environments on Edge Device,” *Power System Technology*, vol. 49, no. 2, pp. 93–124, 2025, doi: https://powertechjournal.com/index.php/journal/article/view/1757.

[21] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” *arXiv.org*, Jun. 08, 2015. https://arxiv.org/abs/1506.02640

[22] J. KHALAF, “ARTIFICIAL INTELLIGENCE’S IMPACT ON ORGANIZATIONAL WORK AND MANAGEMENT,” *Unipd.it*, Jul. 2024, doi: https://hdl.handle.net/20.500.12608/68365.

Ελληνική Βιβλιογραφία

[1] Big, “Υπολογιστική Όραση (Computer Vision): Ορισμός και Εφαρμογές,” *Big Blue Data Academy*, Sep. 09, 2023. https://bigblue.academy/gr/upologistiki-orasi

[2] K. Βασίλη and K. Vasili, “Αναγνώριση αντικειμένων σε εναέρια υψηλής ανάλυσης δεδομένα βίντεο με συνελικτικά νευρωνικά δίκτυα,” *Ntua.gr*, 2019, doi: https://dspace.lib.ntua.gr/xmlui/handle/123456789/48535.

[3] Γ. Οικονόμου and G. Oikonomou, “Εντοπισμός και αναγνώριση τύπου αεροσκαφών σε δορυφορικές εικόνες: Δημιουργία συνόλου δεδομένων και εφαρμογή μοντέλων Deep Learning,” *Ntua.gr*, 2025, doi: https://dspace.lib.ntua.gr/xmlui/handle/123456789/60669.

[4] Σ. Κανελλόπουλος, “ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ Αναγνώριση Προσώπου με χρήση Συνελικτικών Νευρωνικών Δικτύων.” Available: https://pergamos.lib.uoa.gr/uoa/dl/object/2864679/file.pdf