



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ**  
**UNIVERSITY OF PIRAEUS**

---

**ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ**

**Π.Μ.Σ “Μεγάλα Δεδομένα & Αναλυτική” του Τμήματος Ψηφιακών Συστημάτων**

**ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ**

**«Αναγνώριση Παικτών και Ομάδων σε Βίντεο  
Ποδοσφαιρικών Αγώνων με Αλγορίθμους Βαθιάς Μάθησης»  
(Player Object Detection in Football Video Games using Deep  
Learning)**

**ΘΕΟΛΟΓΟΣ Π. ΚΩΝΣΤΑΝΤΙΝΟΣ**

**Επιβλέπων: Ηλίας Μαγκλογιάννης**

**ΠΕΙΡΑΙΑΣ ΣΕΠΤΕΜΒΡΙΟΣ 2021**

# «Αναγνώριση Παικτών και Ομάδων σε Βίντεο Ποδοσφαιρικών Αγώνων με Αλγορίθμους Βαθιάς Μάθησης» (Player Object Detection in Football Video Games using Deep Learning)

ΘΕΟΛΟΓΟΣ ΚΩΝΣΤΑΝΤΙΝΟΣ

Αριθμός Μητρώου :ME1809

Copyright © Θεολόγος Π. Κωνσταντίνος, 2021

Με επιφύλαξη παντός δικαιώματος.

All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα.

-

Οι απόψεις και θέσεις που περιέχονται σε αυτήν την εργασία εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Πανεπιστημίου Πειραιώς.

## ΠΡΟΛΟΓΟΣ

Η εκπόνηση της παρούσας διπλωματικής εργασίας πραγματοποιήθηκε στο πλαίσιο του Προγράμματος Μεταπτυχιακών Σπουδών στα Μεγάλα Δεδομένα & Αναλυτική του Τμήματος Ψηφιακών Συστημάτων του Πανεπιστημίου Πειραιά. Επιθυμώ να εκφράσω τις ευχαριστίες μου σε όλους εκείνους που συνέβαλαν άμεσα ή έμμεσα στην ολοκλήρωση της διπλωματικής μου εργασίας και κατά συνέπεια των μεταπτυχιακών σπουδών μου, σε μια δύσκολη περίοδο για εμένα και όλο τον κόσμο. Αρχικά θα ήθελα να ευχαριστήσω τον επιβλέποντα μου, Καθηγητή Ηλία Μαγκλογιάννη. Η υποστήριξη και διαθεσιμότητά του καθ' όλη τη διάρκεια της εκπόνησης της εργασίας αποτέλεσε σπουδαία βοήθεια, συμβάλλοντας ποικιλοτρόπως στην ολοκλήρωσή της, παρέχοντας μεταξύ άλλων πολύτιμες συμβουλές και καθοδήγηση όπου κρίθηκε αναγκαίο. Τέλος, νιώθω την ανάγκη να ευχαριστήσω την οικογένειά μου για την αγάπη κατανόηση και υποστήριξή τους, χωρίς τις οποίες δεν θα ήταν δυνατή η ολοκλήρωση της παρούσας εργασίας.

## Πίνακας Περιεχομένων

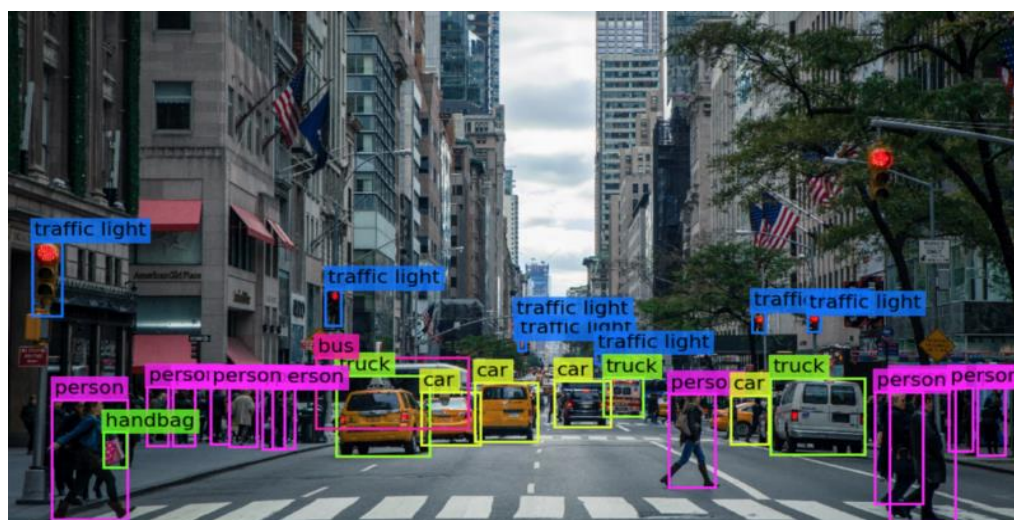
ΠΡΟΛΟΓΟΣ.....	3
ΚΕΦΑΛΑΙΟ 1 – ΕΙΣΑΓΩΓΗ.....	5
I. Ορισμός του προβλήματος.....	5
II. Θεματική περιοχή.....	7
III. Εφαρμογές Υπολογιστικής Όρασης στον Αθλητισμό.....	9
ΚΕΦΑΛΑΙΟ 2 – ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ – MACHINE LEARNING.....	11
I. Υπολογιστικές μηχανές & Νοημοσύνη.....	12
II. Τύποι προβλημάτων & εργασιών.....	12
III. Προσεγγίσεις Μηχανικής Μάθησης.....	14
IV. Βαθιά Μάθηση – Deep Learning.....	15
V. Ανίχνευση Αντικειμένων- Ανακεφαλαίωση.....	17
VI. Yolo Αλγόριθμος.....	18
VII. Το διάνυσμα προβλέψεων – The predictions vector.....	20
VIII. Η λειτουργία απώλειας – The loss function.....	22
ΚΕΦΑΛΑΙΟ 3 – ΜΗΧΑΝΙΚΗ ΟΡΑΣΗ - COMPUTER VISION.....	23
I. Πως λειτουργεί η Τεχνητή Όραση.....	25
II. Η εξέλιξη της Τεχνητής Όρασης.....	28
III. Τα πλεονεκτήματα & τα όρια της Τεχνητής Όρασης.....	30
ΚΕΦΑΛΑΙΟ 4 – ΣΥΜΠΕΡΑΣΜΑΤΑ.....	32
I. Αξιολόγηση εφαρμογής.....	33
ΚΕΦΑΛΑΙΟ 5 - Μελλοντικές βελτιώσεις.....	42
ΒΙΒΛΙΟΓΡΑΦΙΑ.....	43

# 1. ΕΙΣΑΓΩΓΗ

## I. Ορισμός του Προβλήματος

Από όποια οπτική γωνία και αν το παρακολουθείς, το ποδόσφαιρο, δεν μπορείς παρά να συμφωνήσεις ότι περικλείει άπλετο πάθος και συναίσθημα. Έχει λεχθεί ότι οι σύγχρονοι ποδοσφαιρικοί αγώνες είναι σαν σύγχρονες μονομαχίες σε αρένα. Εύκολα μπορούμε να αντιληφθούμε την αίσθηση που αφήνει ο σύγχρονος αθλητισμός, σαν αρχαίος πόλεμος σε ομιχλώδη χωράφια που αντηχεί στις βεράντες και στα γεμάτα στάδια όλου του κόσμου.

Όντας λάτρης του ποδοσφαίρου και των αθλημάτων από μικρή ηλικία δεν θα μπορούσα να παραλείψω, την συστηματική ενασχόληση με το παιχνίδι και ότι αυτό περικλείει στον ελεύθερο μου χρόνο. Λόγω της αυξανόμενης επαγγελματικοποίησης των ποδοσφαιρικών κλαμπ, ένα τεράστιο οικοσύστημα έχει σχηματιστεί γύρω από το άθλημα. Καθώς το ποδόσφαιρο είναι μια μεγάλη επιχείρηση και τα λάθη μπορούν να κοστίσουν εκατομμύρια στους συλλόγους, οι ομάδες αξιοποιούν στο έπακρο τους πόρους τους και προσδιορίζουν τους τομείς που επιδέχονται βελτίωση, όλοι θέλουν μερίδιο από το δέντρο με τα 'χρυσά μήλα'- κλαμπ, μέσα μαζικής ενημέρωσης, μάνατζερ και άλλες εταιρείες.



Εικόνα 1: Yolo result

Ο σύγχρονος πρωταθλητισμός αναγκάζει όποιον ασχολείται με αυτόν να βρίσκει νέες εναλλακτικές που θα μπορέσουν να του προσφέρουν πλεονέκτημα έναντι του εκάστοτε αντιπάλου του. Επομένως δεν θα μπορούσε να μην συμπεριλαμβάνεται σε αυτά το ποδόσφαιρο, εκεί που καθημερινά διακινούνται εκατομμύρια χρήματα σε όλο τον κόσμο. Ένα άθλημα που ο αδύναμος μπορεί να κερδίσει τον δυνατό ακόμα κι αν δεν έχει την ίδια οικονομική ευρωστία. Εκεί εμφανίζονται άνθρωποι – επιστήμονες – λάτρεις του αθλήματος που βρίσκουν ποικίλους τρόπους και εργαλεία για να μικρύνουν αυτή την ψαλίδα.

Ανέκαθεν σε όλους τους τομείς της ζωής μας η 'τακτική' διαδραματίζει σημαίνοντα ρόλο. Στο σύγχρονο επαγγελματικό ποδόσφαιρο η εξειδίκευση σε αυτό τον τομέα είναι αδιαμφησβήτητη. Χρησιμοποιώντας τη δύναμη των υπολογιστών θα προσπαθήσουμε να παρουσιάσουμε ένα τρόπο ανάλυσης που θα χρησιμοποιεί τεχνικές ανίχνευσης της κίνησης

καθώς και εικονικές αναπαραστάσεις δίνοντας τη δυνατότητα καλύτερης αντίληψης του αθλήματος σε απλούς φαν του αθλήματος.

Η ανάλυση ενός ποδοσφαιρικού αγώνα και η μελέτη της απόδοσης των ομάδων έχει απασχολήσει εδώ και αρκετά χρόνια τους ανθρώπους που ασχολούνται με την προπονητική διαδικασία. Πλέον δεν αρκεί η απλή παρακολούθηση ενός αγώνα ή οι σημειώσεις που κρατούν οι προπονητές κατά τη διάρκεια του, ώστε να δοθούν σωστές και έγκυρες απαντήσεις σε σημαντικά ερωτήματα σχετικά με το παιχνίδι και τους αντιπάλους.

Η φράση «μια εικόνα ίσον με 1000 λέξεις» εκφράζει ακριβώς το θέμα αυτό.

## ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ

Η δυνατότητα της τεχνολογίας των υπολογιστών έδωσε τη ευκαιρία για λεπτομερέστερη αξιολόγηση των αθλητικών συμπεριφορών και έδωσε το έναυσμα για τη δημιουργία πολλών προγραμμάτων βίντεο-ανάλυσης. Τα προγράμματα αυτά βασίζονται στη χρήση των υπολογιστών και σχεδιάστηκαν για να καταγράφουν τα τεχνικό-τακτικά στοιχεία των ποδοσφαιριστών με σκοπό να εξαχθούν συμπεράσματα για την απόδοση της ομάδας συνολικά, αλλά και των παικτών ατομικά. Το μυστικό για τη βελτίωση της απόδοσης στο γήπεδο ξεκινά με την κατανόηση του χάσματος μεταξύ του παρόντος επιπέδου αποτελεσματικότητας μιας ομάδας και του επιπέδου που θέλουν να φτάσει. Η βίντεο-ανάλυση βοηθάει τόσο τους ποδοσφαιριστές στο έργο τους όσο και τον προπονητή

Πιο συγκεκριμένα με την βίντεο-ανάλυση μπορεί:

- να **αυξηθεί η πληροφόρηση** των ποδοσφαιριστών μέσα από οπτικά ερεθίσματα, όπως σωστές εικόνες και κινήσεις άλλων παικτών σε κάποιον αγώνα
- να **αυξηθεί η παρακίνηση** τους καθώς δίνει την δυνατότητα τόσο στον ποδοσφαιριστή, όσο και στον προπονητή να παρατηρήσουν τον εαυτό τους, έχοντας μια πραγματική εικόνα για το πώς συμπεριφέρονται αγωνιστικά και όχι απλά να υποθέτουν
- να **διορθώνουν τα τεχνικά τους λάθη** διότι, ότι κι αν γίνεται μέσα σε έναν αγώνα, δεν μπορεί να επαναληφθεί εάν δεν έχει καταγραφεί ο συγκεκριμένος αγώνας και η συγκεκριμένη φάση
- να **παρατηρήσουμε τον αγώνα** και τις διάφορες αγωνιστικές εκτελέσεις των ποδοσφαιριστών, τόσες φορές, όσες είναι απαραίτητες για να βγάλουμε τα σωστά συμπεράσματα
- να **βελτιωθεί η επικοινωνία** ποδοσφαιριστή-προπονητή καθώς αφήνει τους παίκτες βλέποντας τα λάθη τους να γίνουν οι ίδιοι κριτές των εαυτών τους και μετά μέσα από συζήτηση να καταλήξουν στα σωστά συμπεράσματα.

Από τη πλευρά του προπονητή, η χρήση της βίντεο-ανάλυσης μπορεί να αυξήσει την αποδοτικότητα του καθώς μπορεί να διακρίνει αδυναμίες της ίδιας της ομάδας τόσο σε ατομικό, όσο και σε ομαδικό επίπεδο με σκοπό να γίνει επιπλέον καθορισμός των προπονητικών και αγωνιστικών στόχων της ομάδας. Εν συνεχεία να εντοπίσει ανάγκες για μεταγραφές ποδοσφαιριστών που έχει ανάγκη η ομάδα.

Από την άλλη ένα από τα σημαντικότερα οφέλη της βίντεο-ανάλυσης είναι η παρατήρηση της αντίπαλης ομάδας. Ο προπονητής οφείλει να έχει προσωπική γνώμη για τους παίκτες της αντιπάλου ομάδας να γνωρίζει τα ιδιαίτερα χαρακτηριστικά τους και να μπορεί να αναλύει τη φιλοσοφία ανάπτυξης του επιθετικού και αμυντικού παιχνιδιού των αντιπάλων. Η προβολή βίντεο αντιπροσωπευτικού αγώνα με κινήσεις του αντιπάλου από διάφορα παιχνίδια δίνουν τις πληροφορίες που χρειάζονται οι παίκτες, καθώς μπορούν να αντιληφθούν καλύτερα τον τρόπο της αντίδρασης των αντιπάλων. Οι εντολές που οφείλει να δώσει ο προπονητής στους παίκτες του πρέπει να είναι σαφείς και πλήρως κατατοπιστικές χωρίς να μεταφέρει τους δικούς του προβληματισμούς και ανησυχίες επιβαρύνοντας τους με επιπλέον άγχος. Η τεχνολογία με τη μορφή του βίντεο έκανε την

παρακολούθηση της αντίπαλης ομάδας - «κατασκοπεία» φθηνότερη και ευκολότερη αλλά υπάρχει και ο κίνδυνος, πολλοί προπονητές να ξοδεύουν απεριόριστο χρόνο στη παρακολούθηση των αντιπάλων και των αδυναμιών τους παρά να διδάσκουν τους παίκτες τους. Ποτέ δεν ήταν πιο εύκολο να δημιουργήσουμε και να αποκτήσουμε πρόσβαση σε ποδοσφαιρικά δεδομένα. Αλλά το κύριο ζήτημα είναι πως χρησιμοποιούμε τα δεδομένα που διαθέτουμε. Έως ένα βαθμό, μια λεπτομερής ανάλυση της προηγούμενης απόδοσης του παίκτη μπορεί να δώσει τις προβλέψεις για το μέλλον - ποιοι παίκτες θα αγοραστούν, σχηματισμούς που λειτουργούν καλύτερα εναντίον ενός συγκεκριμένου αντιπάλου, καθώς και η πιθανότητα ενός σουτ να μετατραπεί σε γκολ από οποιοδήποτε σημείο του γηπέδου.

### «Πρέπει πάντα να κάνεις τις σωστές ερωτήσεις.»

Φυσικά, με τόσες πολλές μεταβλητές σε κάθε αγώνα ποδοσφαίρου, τα analytics δεν έχουν τη μαγική σφαίρα. Οι στατιστικές προβλέψεις μπορούν να αλλοιωθούν από ένα απλό ανθρώπινο λάθος - ένα χαμένο σουτ, ένα κακό πέρασμα ή μια αμφισβητούμενη απόφαση διαιτητή. Ή μπορούν εξίσου εύκολα να καθοριστούν από μια στιγμή 'μαγείας' από έναν παίκτη παγκόσμιας κλάσης. Οι ομάδες παρακολουθούν τους δικούς τους παίκτες μέσω αισθητήρων που φορούν οι αθλητές και αναλύουν τους αντιπάλους τους χρησιμοποιώντας στατιστικά δεδομένα από μεγάλες εταιρίες στον κλάδο όπως είναι οι: Opta, Statsbomb, Wyscout, SportRadar, Prozone, Instat.

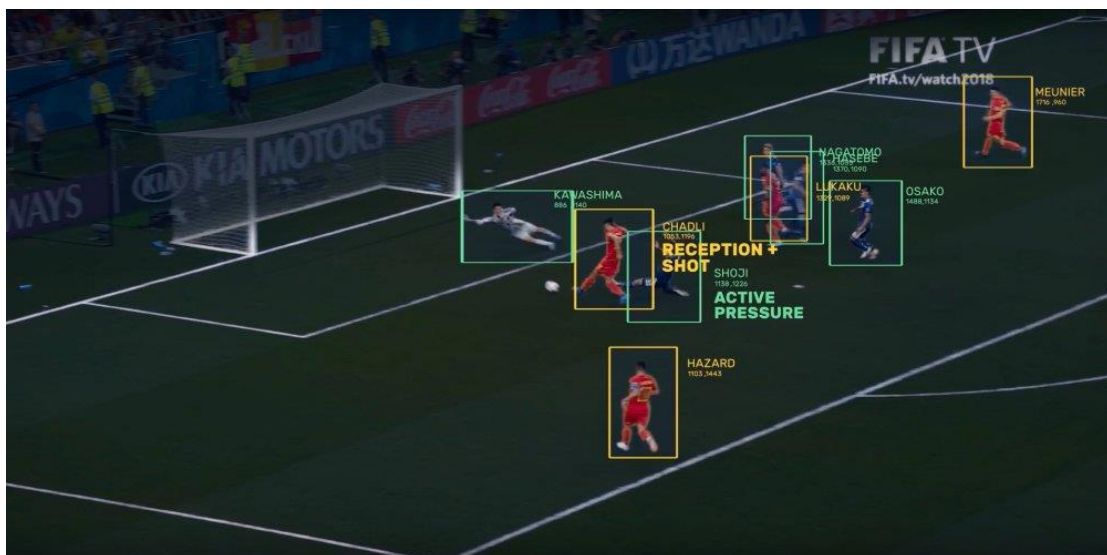


Εικόνα 2: Γνωστοί πάροχοι αθλητικών δεδομένων



### III. Οι εφαρμογές της Υπολογιστικής όρασης στον αθλητισμό

Στον αθλητισμό, η τεχνητή νοημοσύνη ήταν σχεδόν άγνωστη πριν από μερικά χρόνια, αλλά σήμερα με μεθόδους όπως το deep learning και το computer vision δείχνουν το δρόμο σε μια σειρά εφαρμογών στην αθλητική βιομηχανία. Σε προπονήσεις και σε ορισμένους αγώνες, που δεν έχουν τηλεοπτική κάλυψη, ο αναλυτής απόδοσης μπορεί να αποκτήσει μόνο περιορισμένο αριθμό οπτικών γωνιών του βίντεο. Αυτό το βίντεο περιορίζεται στην παροχή οπτικοποιημένης, της κίνησης του παίκτη, παρά λεπτομερούς ανάλυσής του. Τα δεδομένα και οι πληροφορίες που συλλέγονται από το βίντεο απαιτούν από τον αναλυτή να αφιερώσει πολλές ώρες χειροκίνητα, σημειώνοντας και συλλέγοντας γεγονότα καθώς επαναλαμβάνει το βίντεο. Σενάρια όπως αυτό είναι που βρίσκει εφαρμογή η τεχνική του computer vision και μπορεί να γεφυρώσει το χάσμα μεταξύ του αθλητικού γεγονότος και των αναλυτικών πληροφοριών, προσφέροντας νέους τρόπους για τη συλλογή δεδομένων και την απόκτηση πολύτιμης ανάλυσης μέσω αυτοματοποιημένων συστημάτων που εντοπίζουν και τμηματοποιούν κάθε παίκτη που μας ενδιαφέρει να παρακολουθήσουμε κατά την διάρκεια του βίντεο.



**Εικόνα 3:Στιγμιότυπο από το κανάλι της FIFA**

Στο πλαίσιο του αθλητισμού, το υλικό λαμβάνεται συνήθως μέσω μιας ή περισσότερων καμερών που είναι εγκατεστημένες σε κοντινή απόσταση από το σημείο όπου πραγματοποιείται η δράση (δηλ. τα περιθώρια ενός γηπέδου προπόνησης ή σε ένα γήπεδο κατά τη διάρκεια ενός αγώνα). Αυτό μπορεί να αποτελέσει πρόκληση για ορισμένες εφαρμογές της όρασης του υπολογιστή για να ανιχνεύσουν με ακρίβεια, την ακριβή θέση των αντικειμένων ή την κατεύθυνση κίνησης τους, καθώς ενδέχεται να μην καταλάβουν τις διάφορες διαμορφώσεις που χρησιμοποιούνται για τη λήψη των διαφορετικών βίντεο που τους παρουσιάζονται. Δηλαδή για την εκπαίδευση των μοντέλων ή ταξινόμηση νέων, αόρατων εικόνων.

Συνήθως, η ρύθμιση της κάμερας στα συστήματα παρακολούθησης πολλαπλών καμερών για κίνηση-παρακολούθηση μπάλας ή παικτών είναι αρκετά δαπανηρή. Για κάμερες σταθερής γωνίας, αυτό θα μπορούσε να γίνει μέσω βαθμονόμησης της σκηνής. Ωστόσο, οι

κάμερες μετάδοσης των αθλητικών γεγονότων, παρουσιάζουν κάποιες πρόσθετες προκλήσεις σχετικά με το ότι αλλάζουν συχνά τη μετατόπιση, την κλίση και το ζουμ. Αυτός ο δυναμισμός έπρεπε να ληφθεί υπόψη με τη χρήση αισθητήρων στη βάση της κάμερας και του φακού για τη μέτρηση των ρυθμίσεων ζουμ και εστίασης και να είναι σε θέση να συσχετίσουν τις πρώτες τιμές από τους κωδικοποιητές φακών στο εστιακό μήκος.

Η απόκτηση πρόσβασης σε αυτόν τον προηγμένο εξοπλισμό μαγνητοσκόπησης δεν αποτελεί συχνά επιλογή για τα περισσότερα τμήματα, καθώς όπως προαναφέραμε είναι αισθητά ακριβός από άλλες λύσεις ανάλυσης απόδοσης. Περιορίζοντας την ικανότητά, μη εύρωστων αθλητικών συλλόγων να εφαρμόσουν προηγμένη παρακολούθηση παικτών.

Η υπολογιστική όραση έχει λύσει εν μέρει αυτούς τους περιορισμούς. Με την εφαρμογή της επεξεργασίας εικόνας, τα συστήματα υπολογιστικών οράσεων μπορούν πλέον να διακρίνουν μεταξύ του εδάφους, των παικτών και άλλων αντικειμένων στο προσκήνιο. Μέθοδοι όπως η εξαίλιψη του εδάφους με βάση το χρώμα σε γήπεδα με ομοιόμορφες επιφάνειες επιτρέπουν στα μοντέλα όρασης να ανιχνεύουν τις ζώνες ενός γηπέδου, να παρακολουθούν κινούμενους παίκτες και να αναγνωρίζουν την μπάλα.

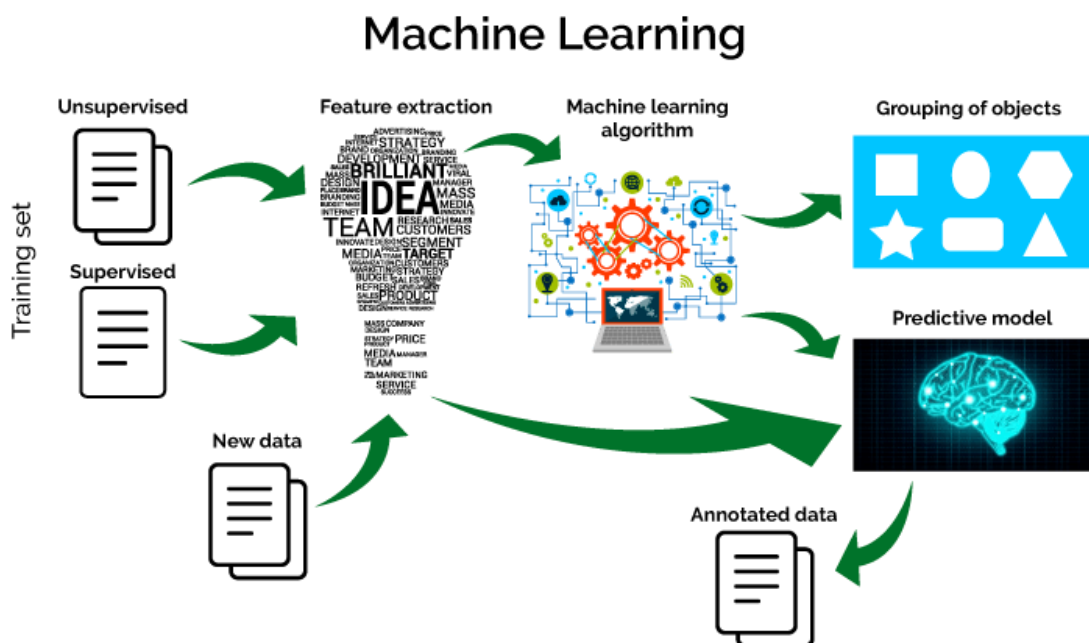
Για παράδειγμα, οι αλγόριθμοι τμηματοποίησης (clustering) με βάση το χρώμα που χρησιμοποιείται για την ανίχνευση του γρασιδιού από το πράσινο χρώμα του και τη μεταχείριση ως φόντο του πλαισίου εικόνας ή βίντεο, όπου οι παίκτες και τα αντικείμενα κινούνται. Επιπλέον, στατικά πλάνα έχουν χρησιμοποιηθεί επίσης ως μέθοδοι διαφοροποίησης εικόνας και αφαίρεσης φόντου για την ανίχνευση της κίνησης των τμηματικών παικτών προσκηνίου στο φόντο της εικόνας.

## 2. Μηχανική μάθηση - Machine Learning

Machine Learning / Μηχανική μάθηση είναι το υποπεδίο της επιστήμης των υπολογιστών που αναπτύχθηκε από τη μελέτη της αναγνώρισης προτύπων και της υπολογιστικής θεωρίας μάθησης στην τεχνητή νοημοσύνη. Η μηχανική μάθηση διερευνά τη μελέτη και την κατασκευή αλγορίθμων που μπορούν να μαθαίνουν από τα δεδομένα και να κάνουν προβλέψεις σχετικά με αυτά. Τέτοιοι αλγόριθμοι λειτουργούν κατασκευάζοντας μοντέλα από πειραματικά δεδομένα, προκειμένου να κάνουν προβλέψεις βασιζόμενες στα δεδομένα ή να εξάγουν αποφάσεις που εκφράζονται ως το αποτέλεσμα.

Η μηχανική μάθηση είναι στενά συνδεδεμένη και συχνά συγχέεται με την υπολογιστική στατιστική, ένας κλάδος, που επίσης επικεντρώνεται στην πρόβλεψη μέσω της χρήσης των υπολογιστών. Έχει ισχυρούς δεσμούς με την μαθηματική βελτιστοποίηση, η οποία της παρέχει μεθόδους, την θεωρία και τομείς εφαρμογής.

Η Μηχανική μάθηση εφαρμόζεται σε μια σειρά από υπολογιστικές εργασίες, όπου τόσο ο σχεδιασμός όσο και ο ρητός προγραμματισμός των αλγορίθμων είναι ανέφικτος. Παραδείγματα εφαρμογών αποτελούν τα φίλτρα spam (spam filtering), η οπτική αναγνώριση χαρακτήρων (OCR), οι μηχανές αναζήτησης και η υπολογιστική όραση. Η Μηχανική μάθηση μερικές φορές συγχέεται με την εξόρυξη δεδομένων, όπου η τελευταία επικεντρώνεται περισσότερο στην εξερευνητική ανάλυση των δεδομένων, γνωστή και ως μη επιτηρούμενη μάθηση.



Εικόνα 4: Διαδικασία Μηχανικής Μάθησης

### I. Υπολογιστικές μηχανές και Νοημοσύνη

Ο Tom M. Mitchell πρότεινε έναν πιο επίσημο ορισμό που χρησιμοποιείται ευρέως: «Ένα πρόγραμμα υπολογιστή λέγεται ότι μαθαίνει από εμπειρία  $E$  ως προς μια κλάση εργασιών  $T$  και ένα μέτρο επίδοσης  $P$ , αν η επίδοσή του σε εργασίες της κλάσης  $T$ , όπως αποτιμάται από το μέτρο  $P$ , βελτιώνεται με την εμπειρία  $E$ ».

Αυτός ο ορισμός είναι σημαντικός για τον καθορισμό της μηχανικής μάθησης σε βασικό λειτουργικό πλαίσιο παρά με γνωστικούς όρους, ακολουθώντας έτσι την πρόταση του Alan Turing στην εργασία του «Υπολογιστικές μηχανές και Νοημοσύνη», ότι το ερώτημα αν μπορούν οι μηχανές να σκεφτούν, μπορεί να αντικατασταθεί με το ερώτημα αν μπορούν οι μηχανές να κάνουν αυτό που εμείς (ως σκεπτόμενες οντότητες) μπορούμε να κάνουμε.

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

## II. Τύποι προβλημάτων και εργασιών

Οι διεργασίες μηχανικής μάθησης συνήθως ταξινομούνται σε τρεις μεγάλες κατηγορίες, ανάλογα με τη φύση του εκπαιδευτικού «σήματος» ή την «ανατροφοδότηση» που είναι διαθέσιμα σε ένα σύστημα εκμάθησης. Αυτές είναι:

**Επιτηρούμενη μάθηση** (αλλιώς επιβλεπόμενη μάθηση ή μάθηση με επίβλεψη) (supervised learning): Το υπολογιστικό πρόγραμμα δέχεται τις παραδειγματικές εισόδους καθώς και τα επιθυμητά αποτελέσματα από έναν «δάσκαλο», και ο στόχος είναι να μάθει έναν γενικό κανόνα προκειμένου να αντιστοιχίσει τις εισόδους με τα αποτελέσματα.

**Μη επιτηρούμενη μάθηση** (αλλιώς μάθηση χωρίς επίβλεψη (unsupervised learning): Χωρίς να παρέχεται κάποια εμπειρία στον αλγόριθμο μάθησης, πρέπει να βρει την δομή των δεδομένων εισόδου. Η Μη Επιτηρούμενη μάθηση μπορεί να είναι αυτοσκοπός (ανακαλύπτοντας κρυμμένα μοτίβα σε δεδομένα) ή μέσο για ένα τέλος (χαρακτηριστικό της μάθησης).

**Ενισχυτική μάθηση:** Ένα πρόγραμμα υπολογιστή αλληλοεπιδρά με ένα δυναμικό περιβάλλον στο οποίο πρέπει να επιτευχθεί ένας συγκεκριμένος στόχος (όπως η οδήγηση ενός οχήματος), χωρίς κάποιος δάσκαλος να του λέει ρητά αν έχει φτάσει κοντά στο στόχο του. Ένα άλλο παράδειγμα είναι να μάθει να παίζει ένα παιχνίδι εναντίον κάποιου αντιπάλου.

Μεταξύ της επιτηρούμενης και της μη επιτηρούμενης μάθησης είναι ημι-επιτηρούμενη μάθηση, όπου ο δάσκαλος δίνει ένα ελλιπές εκπαιδευτικό σήμα: ένα σύνολο εκπαίδευσης με κάποια (συχνά πολλά) από τα αποτελέσματα στόχους να λείπουν. Η Μεταγωγή είναι μια ειδική περίπτωση της αρχής αυτής, όπου το σύνολο των καταστάσεων του προβλήματος είναι γνωστό κατά το χρόνο εκμάθησης, όμως ένα μέρος των στόχων λείπουν.

Μεταξύ άλλων κατηγοριών μηχανικής μάθησης, υπάρχει ακόμα η διαδικασία εκμάθησης (meta learning) που μαθαίνει στην μηχανή (να αναπτύσσει) τις δικές της επαγωγικές μεθόδους, βασιζόμενο στην προηγούμενη εμπειρία.

Η Αναπτυξιακή μάθηση (Developmental robotics), η οποία έχει αναπτυχθεί για την εκμάθηση από ρομπότ, δημιουργεί τη δική της ακολουθία μαθησιακών καταστάσεων, ώστε το ρομπότ συσσωρευτικά αποκτά ποικιλία δεξιοτήτων μέσω της αυτόνομης αυτοεξερεύνησης και της κοινωνικής αλληλεπίδρασης με ανθρώπους εκπαιδευτές και χρησιμοποιώντας μηχανισμούς καθοδήγησης, όπως η ενεργητική μάθηση, η ωρίμανση και η μίμηση.

Μια άλλη κατηγοριοποίηση των προβλημάτων μηχανικής μάθησης προκύπτει όταν κάποιος θεωρήσει το επιθυμητό αποτέλεσμα του συστήματος μηχανικής μάθησης:

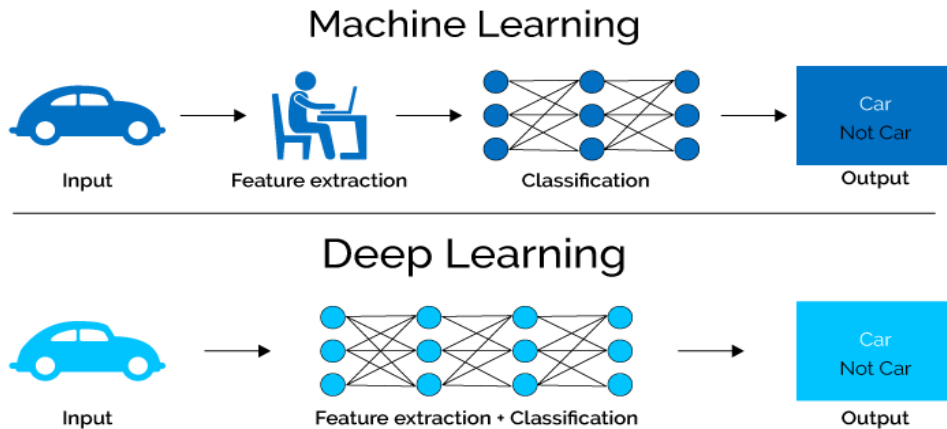
Στην **ταξινόμηση**, τα δεδομένα εισόδου χωρίζονται σε δύο ή περισσότερες κλάσεις, και η μηχανή πρέπει να κατασκευάσει ένα μοντέλο, το οποίο θα αντιστοιχίζει τα δεδομένα σε μία ή περισσότερες (multi-label ταξινόμηση) κλάσεις. Αυτό συνήθως εμπίπτει στην επιτηρούμενη μάθηση. Τα φίλτρα Spam είναι ένα παράδειγμα ταξινόμησης, όπου οι εισοδοί είναι τα emails ή άλλα μηνύματα και οι κλάσεις είναι “spam” και “όχι spam”.

Στην **παλινδρόμηση**, επίσης πρόβλημα επιτηρούμενης μάθησης, τα αποτελέσματα είναι συνεχή και όχι διακριτά.

Στην **συσταδοποίηση**, ένα σύνολο εισόδων πρόκειται να χωριστεί σε ομάδες. Σε αντίθεση με την ταξινόμηση, οι ομάδες δεν είναι γνωστές εκ των προτέρων, καθιστώντας αυτόν τον διαχωρισμό τυπική εργασία μη επιτηρούμενης μάθησης.

Στην εκτίμηση πυκνότητας βρίσκει την κατανομή των δεδομένων εισόδου σε κάποιο χώρο.

Σε προβλήματα μείωσης διαστασιμότητας (dimensionality reduction), τα δεδομένα απλοποιούνται και αντιστοιχίζονται σε ένα χώρο λιγότερων διαστάσεων. Το στατιστικό μοντέλο θεμάτων (Topic modeling) είναι ένα σχετικό πρόβλημα, όπου η μηχανή καλείται να βρει έγγραφα που καλύπτουν παρόμοια θέματα από ένα σύνολο εγγράφων γραμμένων σε φυσική γλώσσα.



**Εικόνα 5: Σύγκριση των 2 τεχνικών**

### III. Προσεγγίσεις μηχανικής μάθησης

#### Εκμάθηση με δέντρο απόφασης

Η εκμάθηση με δέντρο απόφασης χρησιμοποιεί ένα δέντρο απόφασης ως προγνωστικό μοντέλο, το οποίο αντιστοιχίζει παρατηρήσεις σχετικά με ένα στοιχείο σε συμπεράσματα σχετικά με την τιμή στόχο του αντικειμένου.

#### Εκμάθηση με Κανόνες συσχέτισης

Η εκμάθηση με κανόνες συσχέτισης είναι μια μέθοδος ανακάλυψης ενδιαφερουσών σχέσεων μεταξύ των μεταβλητών σε μεγάλες βάσεις δεδομένων.

#### Τεχνητά νευρωνικά δίκτυα

Ένας αλγόριθμος εκμάθησης Τεχνητού νευρωνικού δικτύου, που συνήθως ονομάζεται “νευρωνικό δίκτυο” (NN), είναι ένας αλγόριθμος μάθησης, που εμπνέεται από τη δομή και τις λειτουργικές πτυχές των βιολογικών νευρωνικών δικτύων.

Η δομή των υπολογισμών βασίζεται σε μια ομάδα εσωτερικά διασυνδεδεμένων τεχνητών νευρώνων, οι οποίοι επεξεργάζονται την πληροφορία και εκτελούν υπολογισμούς επικοινωνώντας μεταξύ τους. Τα σύγχρονα νευρωνικά δίκτυα είναι εργαλεία μη γραμμικής στατιστικής μοντελοποίησης δεδομένων.

Συνήθως χρησιμοποιούνται για τη μοντελοποίηση σύνθετων σχέσεων μεταξύ δεδομένων εισόδου και εξόδου, για την ανακάλυψη προτύπων στα δεδομένα, ή για τον εντοπισμό στατιστικής δομής σε μία άγνωστη κοινή κατανομή πιθανότητας μεταξύ των παρατηρούμενων μεταβλητών.

## IV. Βαθιά Μάθηση - Deep Learning

Η πτώση των τιμών του υλικού των τελευταίων ετών καθώς και η ανάπτυξη των GPU για προσωπική χρήση, οδήγησε στην ανάπτυξη της ιδέας της Βαθιάς Μάθησης. Αυτή η προσέγγιση προσπαθεί να μοντελο-ποιήσει τον τρόπο που ο ανθρώπινος εγκέφαλος επεξεργάζεται το φως και τον ήχο και τα μετατρέπει σε όραση και ακοή. Ορισμένες επιτυχείς εφαρμογές της Βαθιάς μάθησης είναι η μηχανική όραση και η αναγνώριση ομιλίας.

### Επαγωγικός λογικός προγραμματισμός

Ο Επαγωγικός λογικός προγραμματισμός (ILP) είναι μια προσέγγιση που διέπει την μάθηση και χρησιμοποιεί λογικό προγραμματισμό ως τρόπο παρουσίασης των παραδειγμάτων εισόδου, του γνωστικού υποβάθρου και των υποθέσεων.

Δεδομένης μιας κωδικοποίησης του γνωστικού υποβάθρου και ενός συνόλου παραδειγμάτων που παρουσιάζονται σαν λογική βάση γεγονότων, το σύστημα ΕΛΠ παράγει το υποτιθέμενο λογικό πρόγραμμα που περιέχει όλα τα θετικά και κανένα αρνητικό παράδειγμα.

Ο επαγωγικός προγραμματισμός είναι ένας σχετικός τομέας που λαμβάνει υπόψιν κάθε είδος προγραμματιστικής γλώσσας για την αναπαράσταση υποθέσεων (και όχι μόνο λογικό προγραμματισμό), όπως τα συναρτησιακά προγράμματα.

### Μηχανές διανυσμάτων υποστήριξης

Οι μηχανές διανυσμάτων υποστήριξης είναι ένα σύνολο μεθόδων επιτηρούμενης μάθησης που χρησιμοποιούνται για την ταξινόμηση και την παλινδρόμηση. Σ' αυτήν την περίπτωση δίνεται ένα σύνολο παραδειγμάτων εκπαίδευσης και κάθε φορά δηλώνεται σε ποια από τις δύο κατηγορίες ανήκει το παράδειγμα.

Μία μηχανή διανυσμάτων υποστήριξης κατασκευάζει ένα μοντέλο που προβλέπει αν το νέο παράδειγμα εμπίπτει στην μία κατηγορία ή την άλλη.

### Ομαδοποίηση

Η ομαδοποίηση είναι η διαδικασία κατά την οποία ένα σύνολο παρατηρήσεων χωρίζεται σε υποσύνολα έτσι ώστε οι παρατηρήσεις που ανήκουν στην ίδια ομάδα (cluster) είναι όμοιες, σύμφωνα με κάποιο ή κάποια προκαθορισμένα κριτήρια, ενώ οι παρατηρήσεις που προέρχονται από διαφορετικά υποσύνολα είναι ανόμοιες.

Διαφορετικές τεχνικές κατηγοριοποίησης οδηγούν σε διαφορετικές υποθέσεις σχετικά με τη δομή των δεδομένων, οι οποίες συχνά καθορίζονται από κάποιο μέτρο ομοιότητας και

αξιολογούνται για παράδειγμα ως προς την εσωτερική συνοχή (ομοιότητα μεταξύ των μελών του ίδιου cluster) και το διαχωρισμό ανάμεσα σε διαφορετικές ομάδες.

Άλλες μέθοδοι βασίζονται στην εκτιμώμενη πυκνότητα και την συνεκτικότητα των γραφημάτων. Η ομαδοποίηση είναι μία μέθοδος μη επιτηρούμενης μάθησης και μία τεχνική η οποία χρησιμοποιείται επίσης στην στατιστική ανάλυση δεδομένων.

## **Δίκτυα Bayes**

Ένα δίκτυο Bayes, ένα δίκτυο εμπιστοσύνης ή ένα άκυκλο γραφικό μοντέλο είναι ένα πιθανοθεωρητικό γραφικό μοντέλο που απεικονίζει ένα σύνολο τυχαίων μεταβλητών και την μεταξύ τους υποθετική ανεξαρτησία διαμέσου ενός κατευθυνόμενου άκυκλου γράφου.

Για παράδειγμα, ένα δίκτυο Bayes μπορεί να αναπαραστήσει την πιθανοθεωρητική σχέση μεταξύ ασθενειών και συμπτωμάτων. Δεδομένων των συμπτωμάτων, το δίκτυο μπορεί να χρησιμοποιηθεί για να υπολογίσει τις πιθανότητες παρουσίας διαφόρων ασθενειών.

## **Ενισχυτική μάθηση**

Η Ενισχυτική μάθηση ασχολείται με το πώς ένα υποκείμενο (πράκτορας) θα πρέπει να δράσει σε ένα περιβάλλον, έτσι ώστε να μεγιστοποιηθεί κάποια έννοια μακροπρόθεσμης ανταμοιβής.

Οι αλγόριθμοι ενισχυτικής μάθησης προσπαθούν να βρουν μια πολιτική που αντιστοιχίζει τις καταστάσεις του περιβάλλοντος με τις ενέργειες που ο πράκτορας θα πρέπει να επιτελέσει σε αυτές τις καταστάσεις.

Η ενισχυτική μάθηση διαφέρει από τα προβλήματα επιτηρούμενης μάθησης αφού τα σωστά ζεύγη δεδομένων εισόδου/εξόδου ζεύγη δεν παρουσιάστηκαν ποτέ, ούτε οι βέλτιστες δυνατές ενέργειες έχουν ρητά διορθωθεί.

## **Εκμάθηση με μέτρο ομοιότητας**

Σε αυτή την κατηγορία προβλημάτων δίνονται στην μηχανή μάθησης ζεύγη παραδειγμάτων που θεωρούνται όμοια και ζεύγη που θεωρούνται ανόμοια. Τότε η μηχανή μάθησης πρέπει να μάθει μια συνάρτηση ομοιότητας (ή μια συνάρτηση μετρικής απόστασης), που μπορεί να προβλέψει αν δύο καινούρια αντικείμενα είναι όμοια. Πρόκειται για μια τεχνική που χρησιμοποιείται σε συστήματα σύστασης.



## V. Ανίχνευση Αντικειμένου - Ανακεφαλαίωση

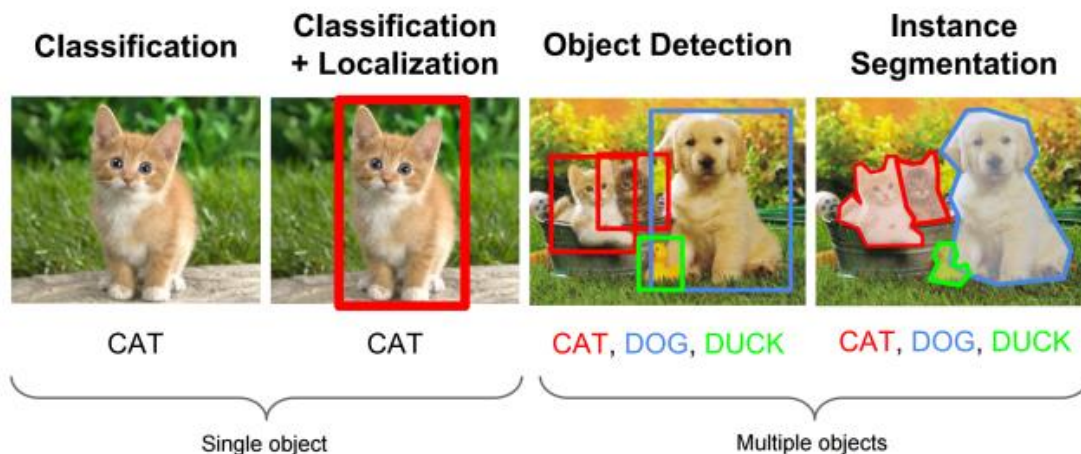
Για να κατανοήσουμε τη σημασία της έννοιας της ανίχνευσης αντικειμένων είναι προαπαιτούμενο να ξεκινήσουμε με την ταξινόμηση(classification) της εικόνας. Περνάει από κάποια επίπεδα σταδιακής πολυπλοκότητας.

Η **ταξινόμηση εικόνας(1)** στοχεύει στην κατηγοριοποίηση μιας εικόνας σε μια από τις διάφορες κατηγορίες (π.χ. αυτοκίνητο, σκύλος, γάτα, άνθρωπος κ.λπ.), απαντώντας ουσιαστικά στην ερώτηση «τι είναι αυτή η εικόνα». Κάθε εικόνα έχει μόνο μια κατηγορία στην οποία ορίζεται.

Ο **εντοπισμός αντικειμένων (2)** μας επιτρέπει να εντοπίσουμε το αντικείμενό μας στην εκάστοτε εικόνα οπότε η ερώτησή μας αλλάζει σε «τι είναι και που είναι;»

Σε ένα πραγματικό σενάριο πρέπει να κοιτάξουμε πέρα από την απλή ανίχνευση ενός αντικειμένου και να εξετάσουμε την πολλαπλή ανίχνευση σε μια εικόνα. Για παράδειγμα, ένα αυτοκινούμενο αυτοκίνητο πρέπει να βρει τη θέση άλλων αυτοκινήτων, φανάρια, πινακίδες, ανθρώπους και να λάβει τα κατάλληλα μέτρα βάσει αυτών των πληροφοριών.

Η **ανίχνευση αντικειμένων (3)** παρέχει τα εργαλεία για να το κάνει ακριβώς αυτό - εύρεση όλων των αντικειμένων σε μια εικόνα και σχεδίαση των λεγόμενων πλαισίων οριοθέτησης γύρω από αυτά. Υπάρχουν επίσης ορισμένες καταστάσεις όπου θέλουμε να βρούμε τα ακριβή όρια των αντικειμένων μας στη διαδικασία που ονομάζεται τμηματοποίηση στιγμιοτύπων.



Εικόνα 6:Επιλογές κατηγοριοποίησης

Υπάρχουν αρκετοί διαφορετικοί αλγόριθμοι που βοηθούν στην ανίχνευση των αντικειμένων και χωρίζονται σε δύο ομάδες:

1. **Αλγόριθμοι με βάση την ταξινόμηση.** Εφαρμόζονται σε δύο στάδια. Πρώτον, επιλέγουν περιοχές ενδιαφέροντος για μια εικόνα και δεύτερον ταξινομούν αυτές τις περιοχές χρησιμοποιώντας συνελκτικά νευρωνικά δίκτυα. Αυτή η λύση μπορεί να

είναι αργή επειδή πρέπει να εκτελέσουμε προβλέψεις για κάθε επιλεγμένη περιοχή. Ένα ευρέως γνωστό παράδειγμα αυτού του τύπου αλγορίθμων είναι το συνελεκτικό νευρωνικό δίκτυο (RCNN) και τα “ξαδέλφια” του Fast-RCNN, Faster-RCNN και η τελευταία προσθήκη στην οικογένεια: Mask-RCNN. Άλλο ένα επίσης γνωστό παράδειγμα είναι το RetinaNet.

2. **Αλγόριθμοι με βάση την παλινδρόμηση** - αντί να επιλέγουν τα ενδιαφέροντα μέρη μιας εικόνας, προβλέπουν τάξεις και πλαίσια οριοθέτησης (bounding boxes) για ολόκληρη την εικόνα σε μία προσπέλαση του αλγορίθμου. Τα δύο πιο γνωστά παραδείγματα αυτής της ομάδας είναι ο YOLO (You Only Look Once) και SSD (Single Shot Multibox Detector). Χρησιμοποιούνται συνήθως για την ανίχνευση αντικειμένων σε πραγματικό χρόνο καθώς, σε γενικές γραμμές, ανταλλάσσουν λίγη ακρίβεια για μεγάλες βελτιώσεις στην ταχύτητα της ανίχνευσης.

## VI. Yolo αλγόριθμος

Ο αλγόριθμος που χρησιμοποιήσαμε στην συγκεκριμένη εργασία είναι ο λεγόμενος και ως YOLO (You Only Look Once). Ανιχνεύει αντικείμενα σε πραγματικό χρόνο και είναι ένας από τους πιο αποτελεσματικούς στην ανίχνευση αυτών. Η ανίχνευση αντικειμένων είναι μια κρίσιμη ικανότητα που μας δίνει τη δυνατότητα της τεχνολογίας των αυτόνομων οχημάτων. Ο YOLO πρωτοεμφανίστηκε στον τομέα της υπολογιστικής όρασης στα μέσα του 2015 από τον ερευνητή Joseph Redmon και συγκεκριμένα στη δημοσίευσή του “You Only Look Once: Unified, Real-Time Object Detection,”. Άμεσα υπήρξε ενδιαφέρον στην επιστημονική κοινότητα για το νέο αλγόριθμο.

Η αναγνώριση αντικειμένων υπήρξε ανέκαθεν ένα από τα κλασσικά προβλήματα στον τομέα της υπολογιστικής όρασης. Συγκεκριμένα ποια αντικείμενα βρίσκονται μέσα σε μια εικόνα και σε ποιο σημείο αυτής. Το πρόβλημα της ανίχνευσης των αντικειμένων είναι πιο περίπλοκο από την ταξινόμηση (classification) η οποία μπορεί επίσης να αναγνωρίσει αντικείμενα, αλλά δεν υποδεικνύει που βρίσκονται μέσα στην εκάστοτε εικόνα. Επιπλέον, η ταξινόμηση δεν λειτουργεί σε εικόνες που περιέχουν περισσότερα από ένα αντικείμενα.

Επομένως εκτός από την απλή ταξινόμηση εικόνας, υπάρχουν πολλά άλλα συναρπαστικά προβλήματα στην υπολογιστική όραση, με την ανίχνευση αντικειμένων να είναι ένα από τα πιο ενδιαφέροντα. Συνήθως συσχετίζεται με αυτοκινούμενα αυτοκίνητα όπου τα συστήματα συνδυάζουν την όραση του υπολογιστή, το LIDAR (Light Detection And Ranging) και άλλες τεχνολογίες για τη δημιουργία πολυδιάστατης αναπαράστασης του δρόμου με όλους τους συμμετέχοντες σε αυτόν (πεζοί, σήματα, άλλα οχήματα).

Η ανίχνευση αντικειμένων χρησιμοποιείται συνήθως στην παρακολούθηση βίντεο, ειδικά στην παρακολούθηση πλήθους για την αποτροπή τρομοκρατικών επιθέσεων, την καταμέτρηση ατόμων για γενικές στατιστικές ή την ανάλυση της εμπειρίας των πελατών μέσα στα εμπορικά κέντρα.

## Πως λειτουργεί ο Yolo;

Ο YOLO είναι δημοφιλής επειδή επιτυγχάνει υψηλή ακρίβεια ενώ ταυτόχρονα μπορεί να τρέχει σε πραγματικό χρόνο. Ο αλγόριθμος «κοιτάζει μόνο μία φορά» στην εικόνα με την έννοια ότι απαιτεί μόνο ένα “πέρασμα” μέσω του νευρικού δικτύου για να κάνει προβλέψεις. Μετά από την ανάλυση η οποία διασφαλίζει ότι ο αλγόριθμος ανίχνευσης αντικειμένων ανιχνεύει μόνο μία φορά το κάθε αντικείμενο, τότε εξάγει αναγνωρισμένα αντικείμενα μαζί με τα πλαίσια οριοθέτησής τους τα λεγόμενα και ως bounding boxes.

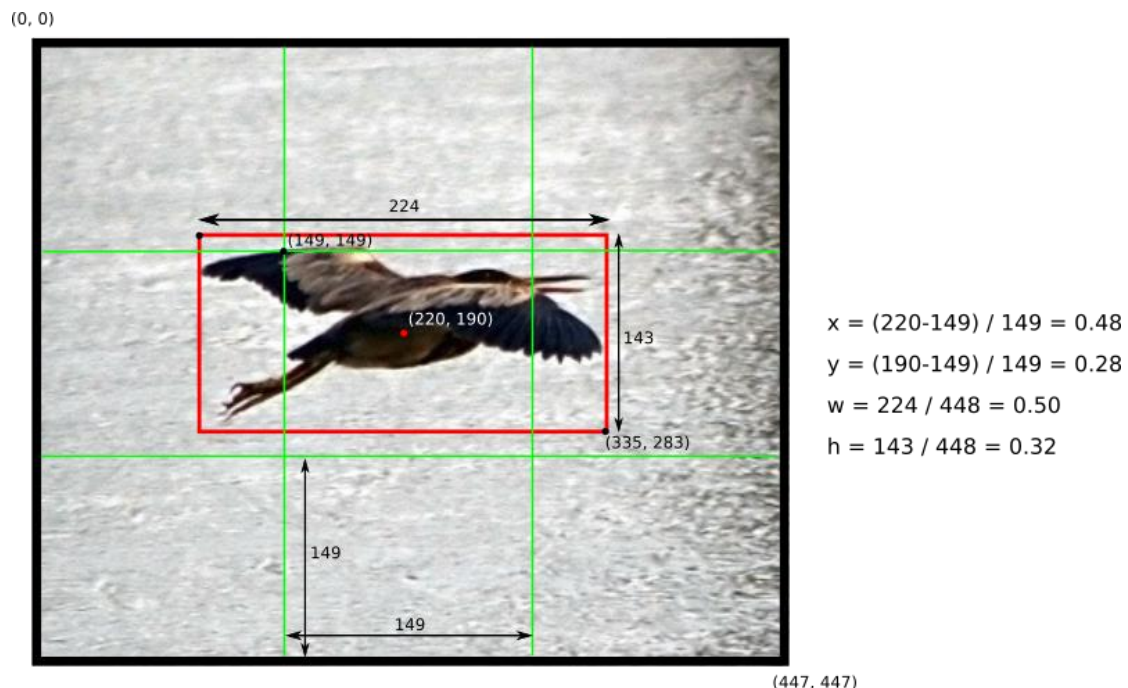
Με το YOLO, ένα CNN(convolutional neural network) προβλέπει ταυτόχρονα πολλαπλά πλαίσια οριοθέτησης και πιθανότητες κλάσης για το καθένα από αυτά. Εκπαιδεύεται σε πλήρεις εικόνες και βελτιστοποιεί άμεσα την απόδοση ανίχνευσης του. Αυτό το μοντέλο έχει πολλά οφέλη σε σχέση με άλλες μεθόδους ανίχνευσης αντικειμένων. Κάποια από αυτά είναι:

- Πραγματικά πολύ γρήγορος
- Διαβάζει ολόκληρη την εικόνα κατά τη διάρκεια του train & test του μοντέλου με συνέπεια να κωδικοποιεί πληροφορίες για τις classes με βάση την εμφάνισή τους.
- Μαθαίνει τις γενικευμένες αναπαραστάσεις των αντικειμένων έτσι ώστε όταν εκπαιδεύεται σε φυσικές εικόνες και δοκιμάζεται, ο αλγόριθμος ξεπερνά σε πολλές περιπτώσεις άλλες κορυφαίες μεθόδους ανίχνευσης.
- Είναι ανοικτού κώδικα και δίνει την ευχέρεια σε τροποποιήσεις

## VII. Το διάνυσμα προβλέψεων (The Predictions Vector)

Το πρώτο βήμα για την κατανόηση του YOLO είναι πώς κωδικοποιεί το αποτέλεσμά του. Η εικόνα εισαγωγής χωρίζεται σε πλέγμα κελιών  $S \times S$ . Για κάθε αντικείμενο που υπάρχει στην εικόνα, ένα κελί πλέγματος λέγεται είναι «υπεύθυνο» για την πρόβλεψή του. Αυτό είναι το κελί στο οποίο πέφτει το κέντρο του αντικειμένου. Κάθε κελί πλέγματος προβλέπει πλαίσια οριοθέτησης (B) καθώς και πιθανότητες κατηγορίας (Γ). Η πρόβλεψη πλαισίου οριοθέτησης έχει 5 στοιχεία:  $(x, y, w, h, \text{confidence})$ . Οι συντεταγμένες  $(x, y)$  αντιπροσωπεύουν το κέντρο του κουτιού, σε σχέση με τη θέση του κελιού πλέγματος (ας έχουμε υπόψιν μας, πως εάν το κέντρο του κουτιού δεν πέσει μέσα στο κελί πλέγματος, από αυτό το κελί δεν είναι υπεύθυνο για την αστοχία αυτή).

Αυτές οι συντεταγμένες κανονικοποιούνται για να πέσουν μεταξύ 0 και 1. Οι διαστάσεις του κουτιού  $(w, h)$  εμφανίζουν επίσης τιμές σε  $[0, 1]$ , σε σχέση με το μέγεθος της εικόνας.



**Εικόνα 7: Παράδειγμα για τον υπολογισμό των συντεταγμένων κουτιού σε μια εικόνα 448x448 με  $S = 3$ . Σημειώστε πώς υπολογίζονται οι συντεταγμένες  $(x, y)$  σε σχέση με το κεντρικό κελί πλέγματος**

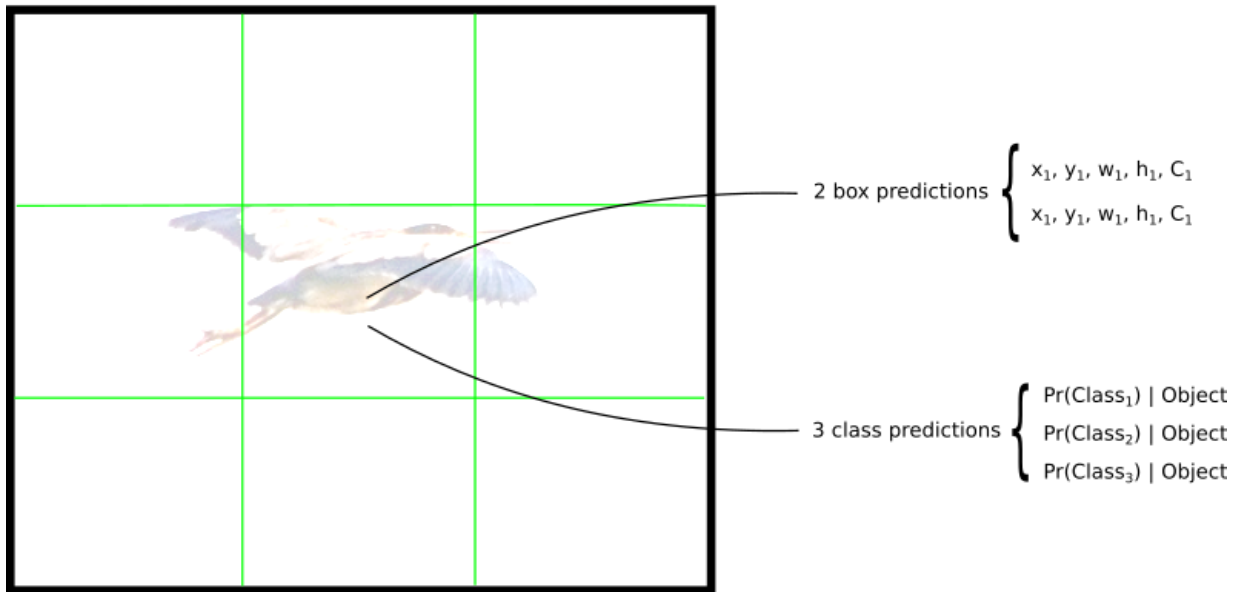
Υπάρχει ακόμη ένα στοιχείο για την πρόβλεψη των ορίων (bounding boxes), που είναι η βαθμολογία εμπιστοσύνης ή αλλιώς «confidence score».

Επίσης ορίζουμε την εμπιστοσύνη (confidence) ως  $\text{Pr}(\text{Object}) * \text{IOU}(\text{pred}, \text{true})$ . Εάν δεν υπάρχει αντικείμενο σε αυτό το κελί, η βαθμολογία εμπιστοσύνης πρέπει να είναι μηδέν. Διαφορετικά, θέλουμε το σκορ εμπιστοσύνης να ισούται με τη διασταύρωση έναντι ένωσης (IOU – Intersection over Union) μεταξύ του προβλεπόμενου κουτιού και της αλήθειας.

Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi (2015). 'You Only Look Once: Unified, Real-Time Object Detection'.

Ανακτήθηκε από το Cornell University library Archives 9 May 2016

Είναι επίσης απαραίτητο να προβλεφθούν οι πιθανότητες της κλάσης,  $\Pr(\text{Class}(i) | \text{Object})$ . Αυτή η πιθανότητα εξαρτάται από το κελί πλέγματος που περιέχει ένα αντικείμενο. Στην πράξη, αυτό σημαίνει ότι εάν δεν υπάρχει αντικείμενο στο κελί πλέγματος, η συνάρτηση απώλειας δεν θα εμφανίσει για λάθος την πρόβλεψη. Το δίκτυο προβλέπει μόνο ένα σύνολο πιθανοτήτων κλάσης ανά κελί, ανεξάρτητα από τον αριθμό των κουτιών B. Αυτό καθιστά συνολικά τις πιθανότητες κλάσης  $S \times S \times C$ .



Εικόνα 8: Κάθε κελί πλέγματος κάνει προβλέψεις πλαισίου οριοθέτησης B και προβλέψεις κατηγορίας Γ ( $S = 3$ ,  $B = 2$  και  $C = 3$ )

## VIII. Η Λειτουργία απώλειας – The Loss Function

Αυτή η εξίσωση υπολογίζει την απώλεια που σχετίζεται με την θέση της προβλεπόμενης οριοθέτησης (x, y). Η συνάρτηση υπολογίζει ένα άθροισμα για κάθε προβλέψιμο κουτί οριοθέτησης (j = 0 .. B) κάθε κελιού πλέγματος (i = 0 .. S ^ 2). Το  $\mathbb{1}_{obj}$  ορίζεται ως εξής:

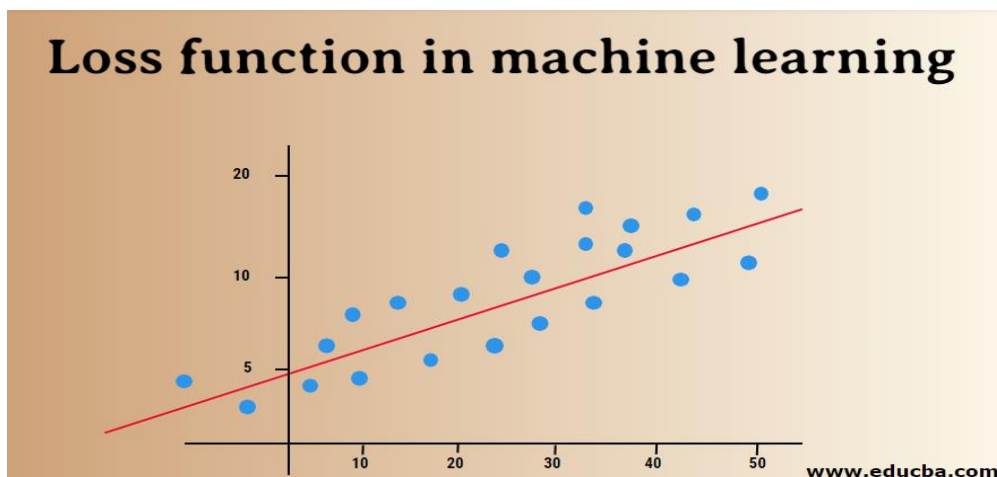
- 1, Εάν ένα αντικείμενο υπάρχει στο κελί πλέγματος i και ο προγνωστικός παράγοντας κουτιού οριοθέτησης jth είναι «υπεύθυνος» για αυτήν την πρόβλεψη
- 0, διαφορετικά

$$\lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{obj} (x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2$$

### Εικόνα 9:YOLO Loss Function

Ο ορισμός που δίνεται για την πρόβλεψη ενός αντικειμένου και πως γνωρίζουμε ποια τιμή είναι σημαντική, παραθέτετε ως εξής:

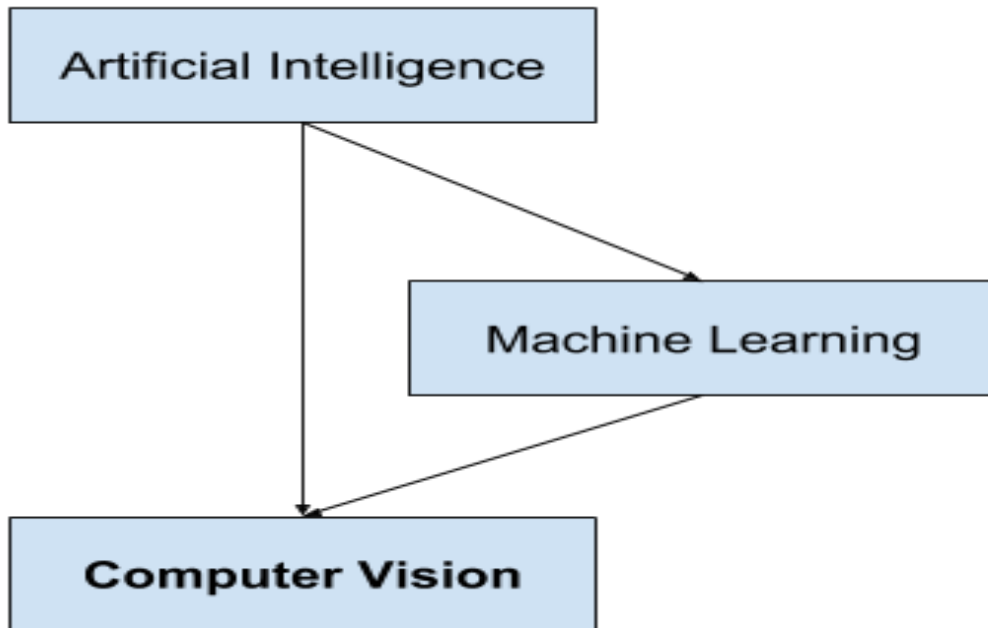
Ο YOLO προβλέπει πολλαπλά πλαίσια οριοθέτησης ανά κελί πλέγματος. Κατά την εκπαίδευσή του, θέλουμε μόνο ένας προγνωστικός παράγοντας οριοθέτησης να είναι υπεύθυνος για κάθε αντικείμενο. Αναθέτουμε έτσι έναν «υπεύθυνο» για την πρόβλεψη ενός αντικειμένου βάσει του οποίου η πρόβλεψη έχει το υψηλότερο τρέχον IOU ακρίβειας στην εκπαίδευση της ταξινόμησης.



Εικόνα 10: Γραφική αναπαράσταση απώλειας λειτουργίας

### 3. Μηχανική Όραση - Computer Vision

Ένας από τους πιο ισχυρούς και συναρπαστικούς τύπους τεχνητής νοημοσύνης είναι η μηχανική όραση που σχεδόν σίγουρα έχει κάποιος βιώσει με πολλούς τρόπους χωρίς καν να το γνωρίζει. Η όραση υπολογιστών είναι το πεδίο της επιστήμης των υπολογιστών που επικεντρώνεται στην αναπαραγωγή τμημάτων της πολυπλοκότητας του συστήματος ανθρώπινης όρασης και στη δυνατότητα των υπολογιστών να αναγνωρίζουν και να επεξεργάζονται αντικείμενα σε εικόνες και βίντεο με τον ίδιο τρόπο που το πραγματοποιούν οι άνθρωποι.



Εικόνα 11: Αλληλεπίδραση των τομέων

Τεχνητή όραση είναι η μετατροπή των δεδομένων από μια φωτογραφική μηχανή ή μια κάμερα, σε ένα συμπέρασμα ή μια νέα αναπαράσταση. Οι μετατροπές αυτές γίνονται για την επίτευξη ορισμένων στόχων. Μια νέα αναπαράσταση μπορεί να είναι η μετατροπή μιας έγχρωμης εικόνας σε ασπρόμαυρη ή η αφαίρεση της κίνησης της κάμερας από μια ακολουθία εικόνων. Επειδή είμαστε οπτικά πλάσματα, είναι εύκολο να μας δημιουργηθεί η λανθασμένη εντύπωση ότι η τεχνητή όραση είναι κάτι απλό. Πόσο δύσκολο είναι να βρεθεί, για παράδειγμα, ένα αυτοκίνητο σε μια εικόνα;

Η αρχική σας διαίσθηση ίσως είναι αρκετά παραπλανητική. Ο ανθρώπινος εγκέφαλος διαχωρίζει το σήμα της όρασης σε πολλά κανάλια, τα οποία δίνουν διαφορετικά είδη πληροφοριών. Έχει ένα σύστημα αναγνώρισης το οποίο αναδεικνύει ποια είναι τα σημαντικά τμήματα της εικόνας για εξέταση και μειώνει την σημασία των άλλων τμημάτων. Αυτό που είναι δεδομένο είναι ότι η διαδικασία της όρασης αποτελείται από πολύ σύνθετες παραμέτρους οι οποίες, μέχρι στιγμής, είναι ελάχιστα κατανοητές. Υπάρχουν διάσπαρτες συνειρμικές εισαγωγές από τους αισθητήρες ελέγχου των μυών των ματιών και όλες τις άλλες ανθρώπινες αισθήσεις που επιτρέπουν στον εγκέφαλο να λειτουργεί βάση συσχετίσεων, οι οποίες δημιουργήθηκαν έπειτα από χιλιάδες χρόνια εξέλιξης του ανθρώπινου είδους. Τα συστήματα ανατροφοδότησης πληροφοριών στον εγκέφαλο

επενεργούν σε όλα τα στάδια της επεξεργασίας συμπεριλαμβανομένου του hardware των αισθητήρων, δηλαδή τα μάτια, τα οποία ελέγχουν μηχανικά το φωτισμό μέσω της ίριδας και την λήψη της εικόνας στην επιφάνεια του αμφιβληστροειδή.

Σε ένα σύστημα τεχνητής όρασης όμως, ο ηλεκτρονικός υπολογιστής λαμβάνει ένα σύνολο από αριθμούς από μια κάμερα και αυτό είναι όλο. Ως επί το πλείστον, δεν υπάρχει ενσωματωμένη αναγνώριση προτύπων, δεν υπάρχει αυτόματος έλεγχος εστίασης και διαφράγματος, και δεν υπάρχουν συσχετίσεις από αιώνες εξέλιξης. Τα συστήματα όρασης βρίσκονται ακόμη σε νηπιακό στάδιο.

Οι ερευνητές της μηχανικής όρασης έχουν αναπτύξει μαθηματικές τεχνικές για την ανάκτηση του τρισδιάστατου σχήματος και την εμφάνιση των αντικειμένων σε εικόνες. Πλέον έχουμε αξιόπιστες τεχνικές για τον ακριβή υπολογισμό τρισδιάστατων μοντέλων ενός περιβάλλοντος με χιλιάδες επικαλυπτόμενες φωτογραφίες. Λαμβάνοντας υπόψη ένα αρκετά μεγάλο σύνολο απόψεων για ένα συγκεκριμένο αντικείμενο μπορούμε να δημιουργήσουμε ένα τρισδιάστατο μοντέλο επιφανειών χρησιμοποιώντας τη στεροσκοπική αντιστοίχιση. Μπορούμε να εντοπίσουμε έναν άνθρωπο που κινείται σε ένα σύνθετο φόντο. Μπορούμε ακόμη, με μέτρια επιτυχία, να προσπαθούμε να βρούμε και να ονομάσουμε όλους τους ανθρώπους σε μια φωτογραφία με το συνδυασμό προσώπου, ντυσίματος και μαλλιών.

Ωστόσο, παρ' όλες αυτές τις εξελίξεις, το όνειρο της κατοχής ενός υπολογιστή που ερμηνεύει μια εικόνα στο ίδιο επίπεδο με ένα δίχρονο παιδί (για παράδειγμα, να μετράει όλα τα ζώα σε μια εικόνα) παραμένει άπιαστο. Γιατί η όραση είναι τόσο δύσκολη; Εν μέρει, αυτό συμβαίνει γιατί η όραση είναι ένα αντίστροφο πρόβλημα, στο οποίο προσπαθούμε να ανακτήσουμε κάποια άγνωστα πράγματα λόγω των μη επαρκών πληροφοριών για τον πλήρη προσδιορισμό της λύσης. Πρέπει λοιπόν να καταφύγουμε σε μοντέλα βασισμένα στη φυσική και στις πιθανότητες για να αποσαφηνιστούν ενδιάμεσες λύσεις.

Τα μοντέλα που χρησιμοποιούνται στη τεχνητή όραση συνήθως αναπτύσσονται στη φυσική (ραδιομετρία, οπτικά και σχεδιασμός αισθητήρα) και στα γραφικά υπολογιστών. Και τα δύο αυτά πεδία μοντελοποιούν το πώς κινούνται και απεικονίζονται τα αντικείμενα, το πώς το φως αντανακλάται στις επιφάνειες τους, σκορπίζεται από την ατμόσφαιρα, διαθλάται μέσα από τους φακούς της κάμερας και τέλος προβάλλεται πάνω σε ένα επίπεδο (ή κοίλο) επίπεδο εικόνας. Στη τεχνητή όραση, προσπαθούν να κάνουν το αντίστροφο, δηλαδή, να περιγραφεί ο κόσμος που βλέπουμε σε μία ή περισσότερες εικόνες και να ανακατασκευάσουν τις ιδιότητες του, όπως το σχήμα, το φωτισμό και τη διανομή χρώματος. Είναι εκπληκτικό το ότι οι άνθρωποι και τα ζώα το κάνουν αυτό αβίαστα, ενώ οι αλγόριθμοι τεχνητής όρασης έχουν πιθανότητα σφάλματος. Τα καλά νέα είναι ότι η τεχνητή όραση χρησιμοποιείται σήμερα σε μια ευρεία ποικιλία εφαρμογών οι οποίες περιλαμβάνουν:

- **Τρισδιάστατο μοντέλο κτιρίων** (φωτομετρία): πλήρως αυτοματοποιημένη κατασκευή τρισδιάστατων μοντέλων από αεροφωτογραφίες.
- **Ιατρική απεικόνιση**: ταξινόμηση κυττάρων, ανίχνευση όγκων, αναπαράσταση ανθρώπινων οργάνων, εγχείρηση με ρομπότ.
- **Εφαρμογές αυτοκινήτου**: αυτονομία οχημάτων, παρακολούθηση επαγρύπνησης του οδηγού, αποφυγή εμποδίων.
- **Αθλητισμός**: Τακτική ανάλυση των αθλητών



- **Βιομηχανικοί αυτοματισμοί:** βιομηχανική επιθεώρηση (ανίχνευση βλαβών και επιθεώρηση), διάβασμα barcode και ετικετών, ταξινόμηση εμπορευμάτων.
- **Ρομποτική:** αυτόματος καθορισμός θέσης, αποφυγή εμποδίων, έξυπνα ρομπότ αλληλοεπιδρούν με τον άνθρωπο και τον εξυπηρετούν.
- **Ασφάλεια:** (ίριδα, δαχτυλικά αποτυπώματα, αναγνώριση προσώπου), επιτήρηση ύποπτων δραστηριοτήτων ή συμπεριφορών.
- Βάσεις δεδομένων εικόνας και βίντεο

## I. Πως λειτουργεί η Τεχνητή Όραση;

Ένα από τα σημαντικότερα ανοιχτά ερωτήματα τόσο στην Νευροεπιστήμη όσο και στη Μηχανική Μάθηση είναι: το πως λειτουργεί ο ανθρώπινος εγκέφαλος και πως μπορούμε να τον προσεγγίσουμε όσο το δυνατόν περισσότερο με τους δικούς μας αλγόριθμους. Η πραγματικότητα έγκειται στο ότι υπάρχουν πολύ λίγες λειτουργικές και περιεκτικές θεωρίες του πως υπολογίζει ο ανθρώπινος εγκέφαλος. Παρά το γεγονός ότι τα Νευρωνικά Δίκτυα είναι αυτά που προσπαθούν να «μιμηθούν» τον τρόπο λειτουργίας του εγκεφάλου κανείς δεν είναι αρκετά βέβαιος αν αυτό είναι αληθές.

Το ίδιο παράδοξο ισχύει και για την Τεχνητή Όραση, δεδομένου ότι δεν είμαστε πολύ ακριβής για το πώς ο εγκέφαλος και τα μάτια αλληλοεπιδρούν με σκοπό την επεξεργασία των εικόνων. Είναι δύσκολο να αναφέρουμε, πόσο σωστά οι αλγόριθμοι χρησιμοποιούν τις δικές μας εσωτερικές διανοητικές διαδικασίες.

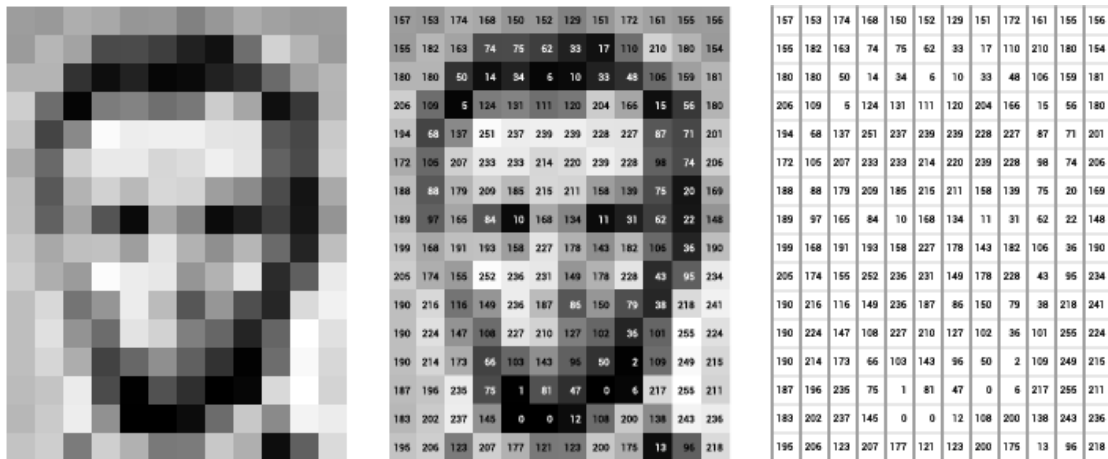
Στο σημείο αυτό να αναφέρουμε ότι η υπολογιστική όραση αφορά κυρίως στην αναγνώριση προτύπων. Επομένως ένας τρόπος για να εκπαιδεύσουμε ένα υπολογιστή για να κατανοήσει τα οπτικά δεδομένα είναι να τον τροφοδοτήσουμε με αρκετές χιλιάδες εικόνες, εκατομμύρια ει δυνατόν που να έχουν ετικέτες με την ονομασία η διάφορα χαρακτηριστικά στοιχεία. Στη συνέχεια θα υποβάλλονται σε διάφορες τεχνικές λογισμικού ή σε συγκεκριμένους αλγόριθμους που θα επιτρέπουν στον υπολογιστή να κινηθεί σε μοτίβα που θα του ορίζουν οι σχετιζόμενες ετικέτες.

Επί παραδείγματι εάν τροφοδοτήσουμε τον υπολογιστή με ένα εκατομμύριο εικόνες ενός αθλητή, θα τις υποβάλει σε επεξεργασία από αλγόριθμους που θα επιτρέπουν την ανάλυση στα χρώματα της εκάστοτε εικόνας-φωτογραφίας, το σχήμα, την απόσταση μεταξύ των σχημάτων εκεί που τα ίδια τα αντικείμενα οριοθετούνται μεταξύ τους.

Έτσι προσδιορίζεται ένα προφίλ που κατηγοριοποιεί του τι «σημαίνει» αθλητής. Όταν ολοκληρωθεί η διαδικασία ο υπολογιστής (θεωρητικά) θα μπορεί να χρησιμοποιήσει την

εμπειρία του σε περίπτωση που τροφοδοτηθεί εκ νέου με εικόνες και του ζητηθεί να εντοπίσει αν εμφανίζεται στην εικόνα κάποιος αθλητής.

Παρακάτω θα παρουσιάσουμε μια απλή απεικόνιση της κλίμακας του γκρι και πως αυτή αποθηκεύεται. Η φωτεινότητα κάθε εικονοστοιχείου αντιπροσωπεύεται από ένα και μόνο αριθμό 8-bit, του οποίου το εύρος ξεκινάει από 0(μαύρο) και καταλήγει έως το 255(λευκό):



Εικόνα 12:Αρίθμηση εικόνας

Στην πραγματικότητα, οι τιμές των pixel αποθηκεύονται σχεδόν καθολικά, σε έναν μονοδιάστατο πίνακα. Αυτός ο τρόπος αποθήκευσης δεδομένων εικόνας ενδέχεται να έρχεται σε αντίθεση με ότι μπορεί να έχουμε στο μυαλό μας για την λειτουργία του, δεδομένου ότι τα δεδομένα σίγουρα φαίνεται να είναι δισδιάστατα όταν εμφανίζονται. Ωστόσο, αυτό συμβαίνει, διότι η μνήμη του υπολογιστή αποτελείται πλήρως από μια συνεχώς αυξανόμενη γραμμική λίστα χωρικών διευθύνσεων.

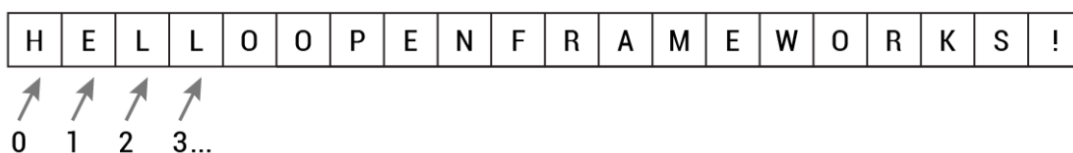
How the pixels look:

H	E	L	L	O
O	P	E	N	F
R	A	M	E	W
O	R	K	S	!

How the pixels are numbered:

0	1	2	3	4
5	6	7	8	9
10	11	12	13	14
15	16	17	18	19

How the pixels are stored in computer memory:



Εικόνα 13: Pixel και διαμόρφωσή τους






Είναι σωστό να αναφερθεί ότι ο υπολογιστής διαβάζει τα χρώματα σαν μια σειρά από 3 μεταβλητές - κόκκινο, πράσινο και μπλε (RGB) στην ίδια κλίμακα από 0-255. Επομένως, κάθε pixel έχει στην πραγματικότητα 3 τιμές για να αποθηκεύσει ο υπολογιστής εκτός από τη θέση της εικόνας.



## How to create colors with RGB?

Combine parts of the three primary colors **red**, **green** and **blue**.

Each of the primary colors can have a value in the range from 0 to 255.

					
R:	255	0	0	0	255
G:	0	255	0	0	255
B:	0	0	255	0	255

© Presentable

Εικόνα 14: Δημιουργία χρωμάτων με RGB

Εύκολα γίνεται αντιληπτό ότι απαιτείται μεγάλο μέρος της μνήμης για μία εικόνα και πολλά εικονοστοιχεία μέσα σε αυτή, για να μπορέσει να τα επεξεργαστεί ένας αλγόριθμος. Αλλά για να εκπαιδεύσουμε ένα μοντέλο με πραγματική ακρίβεια ειδικά όταν μιλάτε για “deep learning”, συνήθως χρειάζεται δεκάδες χιλιάδες εικόνες και όσο περισσότερες τόσο καλύτερο δυνατό και το αποτέλεσμα.

## II. Η εξέλιξη της Τεχνητής όρασης

Πριν από την έλευση του “deep learning”, οι εργασίες που μπορούσε να εκτελέσει η όραση του υπολογιστή ήταν πολύ περιορισμένες και απαιτούσαν πολύ χειροκίνητη κωδικοποίηση και προσπάθεια από προγραμματιστές και άλλους ανθρώπινους χειρισμούς. Για παράδειγμα, εάν κάποιος ήθελε να πραγματοποιήσει μια αναγνώριση προσώπου, θα πρέπει να εκτελέσει τα ακόλουθα βήματα:

1. **Δημιουργία βάσης δεδομένων:** Απαιτούνταν να έχουν τραβηχτεί μεμονωμένες εικόνες όλων των θεμάτων που θέλαμε να παρακολουθήσουμε σε μια συγκεκριμένη μορφή.
2. **Σχολιασμός εικόνων:** Στη συνέχεια, για κάθε μεμονωμένη εικόνα, θα πρέπει να εισαγάγουμε πολλά βασικά σημεία δεδομένων, όπως η απόσταση μεταξύ των ματιών, πλάτος της γέφυρας της μύτης, απόσταση μεταξύ άνω χείλους και μύτης και δεκάδες άλλες μετρήσεις που καθορίζουν τα μοναδικά χαρακτηριστικά κάθε ατόμου.
3. **Λήψη νέων εικόνων:** Μετέπειτα, θα πρέπει να τραβήξουμε νέες εικόνες, είτε από φωτογραφίες είτε από περιεχόμενο βίντεο. Και τότε έπρεπε να γίνει ξανά από η διαδικασία μέτρησης, σημειώνοντας τα βασικά σημεία της εικόνας. Έπρεπε επίσης να λάβουμε υπόψη τη γωνία που τραβήχτηκε η εικόνα.

Μετά από όλη τη χειροκίνητη εργασία, η εφαρμογή θα μπορούσε τελικά να συγκρίνει τις μετρήσεις στη νέα εικόνα με αυτές που είναι αποθηκευμένες στη βάση δεδομένων και να πει εάν αντιστοιχούσε σε οποιοδήποτε από τα προφίλ που παρακολουθούσε. Στην πραγματικότητα, υπήρχε λιγότερος αυτοματισμός και το μεγαλύτερο μέρος της εργασίας έγινε χειροκίνητα. Το περιθώριο σφάλματος ήταν ακόμα αυξημένο.

Η μηχανική μάθηση παρέχει μια διαφορετική προσέγγιση για την επίλυση προβλημάτων τεχνητής όρασης στον υπολογιστή. Με τη μηχανική εκμάθηση, οι προγραμματιστές δεν χρειάζονται πλέον να κωδικοποιούν με μη αυτόματο τρόπο κάθε κανόνα στις εφαρμογές της. Αντί αυτού, προγραμμάτισαν τα «χαρακτηριστικά» τους, σε μικρότερες εφαρμογές που θα μπορούσαν να ανιχνεύσουν συγκεκριμένα μοτίβα μέσα στις εικόνες. Μετέπειτα χρησιμοποίησαν έναν αλγόριθμο στατιστικής μάθησης όπως η γραμμική παλινδρόμηση, η λογιστική παλινδρόμηση, τα δέντρα αποφάσεων ή η μηχανές διανυσμάτων υποστήριξης (SVM) για να ανιχνεύσουν μοτίβα και να ταξινομήσουν εικόνες καθώς επίσης να εντοπίσουν αντικείμενα σε αυτές.

Η μηχανική μάθηση βοήθησε στην επίλυση πολλών προβλημάτων που ήταν δύσκολο να επιλυθούν με τα προ υπάρχοντα εργαλεία και τις προσεγγίσεις της τότε κλασικής ανάπτυξης λογισμικού. Για παράδειγμα, πριν από χρόνια, οι μηχανικοί λογισμικού κατάφεραν να δημιουργήσουν ένα πρόγραμμα που θα μπορούσε να προβλέψει καλύτερα τα ποσοστά επιβίωσης του καρκίνου του μαστού από τους ειδικούς ιατρούς. Ωστόσο, η κατασκευή των «χαρακτηριστικών» του λογισμικού απαιτούσε τις προσπάθειες δεκάδων μηχανικών και εμπειρογνομώνων για τον καρκίνο του μαστού και χρειάστηκε πολύς χρόνος για να αναπτυχθεί.

Το deep learning παρείχε μια θεμελιωδώς διαφορετική προσέγγιση για τη μηχανική μάθηση. Η εκ βάθρων μάθηση βασίζεται σε νευρωνικά δίκτυα, μια λειτουργία γενικού σκοπού που μπορεί να λύσει οποιοδήποτε πρόβλημα αντιπροσωπεύεται μέσω παραδειγμάτων. Όταν δημιουργείται ένα νευρωνικό δίκτυο με πολλαπλά παραδείγματα ενός συγκεκριμένου είδους δεδομένων, είναι σε θέση να εξαγάγει πολλά κοινά μοτίβα μεταξύ αυτών των παραδειγμάτων και να τα μετατρέψει σε μαθηματική εξίσωση, που θα βοηθήσει στην ταξινόμηση μελλοντικών πληροφοριών σε κομμάτια.

Η διαδικασία αυτή είναι μια πολύ αποτελεσματική μέθοδος για την τεχνητή όραση. Στις περισσότερες περιπτώσεις, η δημιουργία ενός καλού αλγορίθμου εκ βάθρων μάθησης καταλήγει στη συλλογή ενός μεγάλου όγκου εκπαιδευμένων-επισημασμένων(labeled) δεδομένων καθώς και τη ρύθμιση των παραμέτρων όπως ο τύπος και ο αριθμός των επιπέδων των νευρωνικών δικτύων και των κύκλων εκπαίδευσης. Σε σύγκριση με προηγούμενους τύπους μηχανικής μάθησης, η εκ βάθρων είναι ευκολότερη και ταχύτερη για ανάπτυξη και δημιουργία.

Οι περισσότερες από τις τρέχουσες εφαρμογές τεχνητής όρασης, όπως η ανίχνευση καρκίνου, τα αυτοκινούμενα αυτοκίνητα και η αναγνώριση προσώπου κάνουν χρήση της έννοιας του deep learning. Επομένως μαζί με τα νευρωνικά δίκτυα έχουν προσπεράσει το εννοιολογικό φάσμα σε πρακτικές εφαρμογές χάρη στη διαθεσιμότητα και την πρόοδο σε πόρους υλικού και υπολογιστικού νέφους(cloud).



Εικόνα 14: Γνωμικό

### III. Τα πλεονεκτήματα και τα όρια της Τεχνητής Όρασης

Στο σημείο αυτό θα χρειαστεί να αναφέρουμε κάποια από τα πολλά πλεονεκτήματα που αποκτάμε με την χρήση της Τ.Ο. καθώς επίσης και τα όριά της με τα τωρινά δεδομένα. Ευκολά όλα αυτά μπορούν να αλλάξουν καθώς η πρόοδος που επιτυγχάνεται στον κλάδο είναι τεράστια και συνεχής.

**Ταχύτερη και απλούστερη διαδικασία** - Τα συστήματα τεχνητής όρασης των υπολογιστών, μπορούν να εκτελούν μονότονες, επαναλαμβανόμενες εργασίες με ταχύτερο ρυθμό, καθιστώντας την όλη διαδικασία απλούστερη.

**Ακριβές αποτέλεσμα** - Δεν είναι μυστικό ότι τα μηχανήματα δεν κάνουν ποτέ λάθος. Ομοίως, τα συστήματα τεχνητής όρασης με δυνατότητες επεξεργασίας εικόνας θα διαπράξουν μηδενικά λάθη, σε αντίθεση με τους ανθρώπους. Τελικώς, τα προϊόντα ή οι υπηρεσίες που παρέχονται δεν θα είναι μόνο γρήγορα αλλά και υψηλής ποιότητας.

**Μείωση κόστους** - Με τους υπολογιστές να αναλαμβάνουν ευθύνες για την εκτέλεση δυσκίνητων εργασιών, τα λάθη θα ελαχιστοποιούνται, μη δίνοντας χώρο σε ελαττωματικά προϊόντα ή υπηρεσίες. Ως αποτέλεσμα, οι εταιρείες μπορούν να εξοικονομήσουν πολλά χρήματα που διαφορετικά θα δαπανούσαν για τον καθορισμό ελαττωματικών διαδικασιών και προϊόντων.

Ενώ υπάρχουν μεγάλες φιλοδοξίες και προσμονή για τις δυνατότητες που μας παρέχει η Τεχνητή όραση των υπολογιστών, καθώς καμία νέα τεχνολογία-καινοτομία δεν είναι απαλλαγμένη από ελαττώματα το ίδιο ισχύει και για τα συστήματα Τεχνητής όρασης των υπολογιστών. Τα μειονεκτήματα της όρασης του υπολογιστή αφορούν ένα μεγάλο ζήτημα στη σύγχρονη εποχή: την ιδιωτικότητα. Η κινητήρια δύναμη που καθιστά την τεχνολογία αυτή τόσο αποτελεσματική είναι η ίδια, που γεννά ζητήματα στους καταναλωτές και δημιουργεί αμφιβολίες για το αν πρέπει να συνεχίζουμε να τη χρησιμοποιούμε. Συγκεντρώνοντας και μαθαίνοντας από χιλιάδες και χιλιάδες φωτογραφίες, βίντεο και άλλα κομμάτια πληροφοριών, όλα όσα κάνουμε αποθηκεύονται κάπου στο διαδίκτυο, που ανήκουν σε εταιρείες ή είναι ορατά σε όλους.

Με την ικανότητα που έχει να αναγνωρίζει τα πρόσωπα των ανθρώπων, καθώς και να εντοπίζει τον τόπο και τις συνήθειες τους, η Τ.Ο. του υπολογιστή έχει αλλάξει το μέλλον της ιδιωτικής ζωής. Καθώς η συγκεκριμένη τεχνολογία τεχνητής νοημοσύνης γίνεται όλο και πιο ευρέως διαδεδομένη, οι χρήστες θα πρέπει να συνειδητοποιήσουν καλύτερα το είδος των δεδομένων που βάζουν προς χρήση.

Παρακάτω παραθέτονται μερικοί περιορισμοί που διαθέτει η τεχνολογία:

- **Έλλειψη ειδικών** - Η τεχνολογία τεχνητής όρασης περιλαμβάνει τη χρήση AI(Artificial Intelligence) και ML(Machine Learning). Για να εκπαιδύσουμε ένα σύστημα όρασης υπολογιστή που υποστηρίζεται από AI και ML, οι εταιρείες πρέπει να έχουν μια ομάδα

επαγγελματιών με τεχνική εμπειρία. Χωρίς αυτούς, δεν είναι δυνατή η δημιουργία ενός συστήματος που να μπορεί να αναλύει και να επεξεργάζεται τις πιθανές περιβάλλουσες λεπτομέρειες.

- **Χρειάζεται τακτική παρακολούθηση** - Τι γίνεται αν ένα σύστημα όρασης υπολογιστή καταρρεύσει ή έχει τεχνικό πρόβλημα; Για να διασφαλιστεί ότι αυτό δεν θα συμβεί, οι εταιρείες πρέπει να αποκτήσουν μια ειδική ομάδα για τακτική παρακολούθηση και αξιολόγηση. Παρά τους ισχύοντες περιορισμούς τους, τα συστήματα τεχνητής όρασης των υπολογιστών μπορούν να φέρουν στις εταιρείες τεράστιες ευκαιρίες για να αυξήσουν τις ροές εσόδων τους, να επιτύχουν τους στόχους παραγωγικότητας και να απλοποιήσουν τις διαδικασίες εργασίας. Ωστόσο ακόμα παρακολουθούμε την επιφάνεια των δυνατοτήτων όρασης του υπολογιστή. Το μέλλον μας επιφυλάσσει αρκετά.

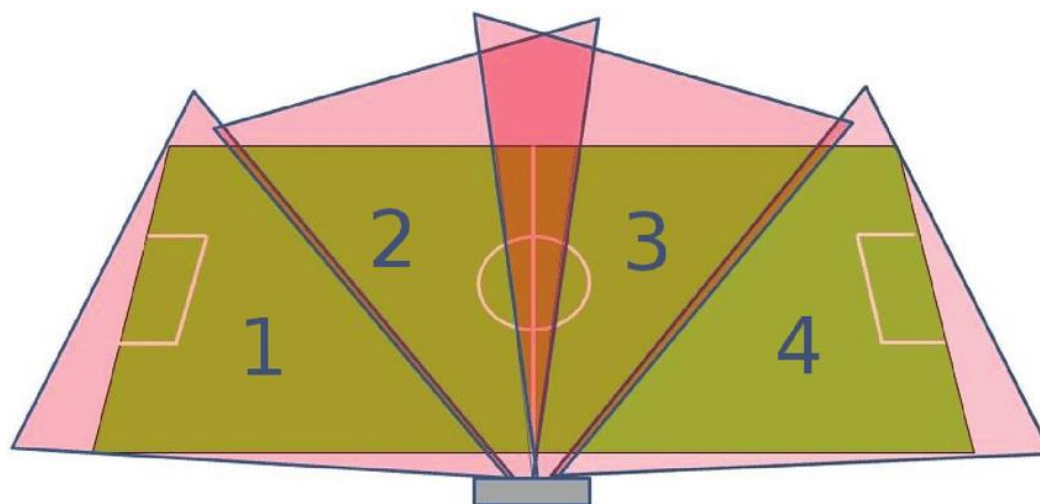
Αντί επιλόγου πρέπει να αναφέρουμε ότι η Τεχνητή Όραση(Computer Vision) του υπολογιστή έχει τη δυνατότητα να αλλάξει το τοπίο κάθε πεδίου που μπορούμε να φανταστούμε. Η ζωή του μέσου πολίτη μπορεί να διευκολυνθεί με ευκολότερες αγορές, επαυξημένη πραγματικότητα για καλύτερα ενημερωμένες αποφάσεις και μεγαλύτερη συνδεσιμότητα με τον κόσμο συνολικά, τόσο φυσική όσο και ψηφιακή.

Η ουσία είναι ότι υπάρχουν πλεονεκτήματα και μειονεκτήματα. Αν και είναι μια απίστευτη τεχνολογία, αξίζει να είμαστε επιφυλακτικοί και να δρούμε με σκεπτικισμό. Οι νέες πρόοδοι δεν είναι ποτέ σωστές από την πρώτη στιγμή της χρήσης τους.

## 4. Συμπεράσματα

Στην παρούσα εργασία είχαμε ως στόχο την καλύτερη δυνατή χρήση των αλγορίθμων πάνω σε βίντεο αγώνων ποδοσφαίρου με σκοπό την ανίχνευση κίνησης. Επιδίωξη ήταν να δοθεί όσο το δυνατόν πιο απλή και κατανοητή η όλη μετακίνηση στον χώρο, που με γυμνά μάτια στην ροή του αγώνα είναι δύσκολο να το αντιληφθεί ο τηλεθεατής. Ένας αθλητικός αγώνας μπορεί να αναπαρασταθεί από μια συλλογή διαδοχικών καταστάσεων και τις αλληλεπιδράσεις τους. Η κατάσταση του παιχνιδιού μπορεί να περιγραφεί χρησιμοποιώντας ένα σύνολο χαρακτηριστικών που περικλείουν τις θέσεις παικτών, τη φυσική τους απεικόνιση ή διάφορα μοντέλα κίνησης. Τότε ο στόχος της παρακολούθησης πολλών παικτών είναι η εκτίμηση της κατάστασης του παιχνιδιού  $x$ , τη στιγμή  $t$ , δεδομένου ενός συνόλου παρατηρήσεων  $z$  μέχρι την παρούσα στιγμή. Η περίπτωση που εξετάσαμε ήταν ο διεθνής αγώνας ανάμεσα στον Ολυμπιακό και την Άλκμααρ.

Μια από τις πιο σημαντικές αποφάσεις που πρέπει να λάβουμε υπόψιν είναι η δυσκολία που αντιμετωπίζουμε στη διαμόρφωση της κάμερας. Στις ζωντανές μεταδώσεις τα πλάνα που λαμβάνουμε έχουν ληφθεί μέσω κάμερας που έχει την pan-tilt-zoom λειτουργία. Επομένως επιλέξαμε η μετάδοση να είναι μονοκάμερη προκειμένου να μην απαιτείται από τον αλγόριθμο να υπολογίζει συνεχώς την νέα θέση του παίκτη.



Εικόνα 15: Στήσιμο καμερών σε γήπεδο

Αυτό έχει ως αποτέλεσμα την ταχύτερη ανίχνευση. Ο ακριβής και αποτελεσματικός εντοπισμός παικτών και μπάλας είναι το σημείο κλειδί οποιασδήποτε λύσης που αποσκοπεί στην αυτοματοποίηση της ανάλυσης των βίντεο όπως στην περίπτωση μας.

Το ποδόσφαιρο είναι ένα από τα πιο δημοφιλή αθλήματα παγκοσμίως και μπορεί να τραβήξει την προσοχή εκατομμυρίων τηλεθεατών σε ένα και μόνο παιχνίδι στα υψηλότερα πρωταθλήματα. Εκατομμύρια βλέμματα καρφωμένα στις ίδιες εικόνες που δείχνουν 22 παίκτες να παλεύουν για την κατοχή μιας μπάλας, το δύσκολο είναι να παρουσιάσουμε τις λεπτομέρειες, αντλώντας όσο το δυνατόν περισσότερες γνώσεις από τη ροή του βίντεο ενός ποδοσφαιρικού αγώνα που έχει καταγραφεί από μία μόνο κάμερα που μοιάζει με μετάδοση.



Είναι μακριά από την πραγματικότητα να πιστεύουμε ότι με την απλή προσέγγιση της αφαίρεσης του φόντου θα μπορέσουμε να κρατήσουμε μόνο τα pixels που ανήκουν στους παίκτες. Αυτό δεν μπορεί να λειτουργήσει διότι στην εξίσωση μπαίνουν και άλλες μεταβλητές όπως η σκιά από τον ήλιο και ο ανεπαρκής φωτισμός σε σημεία του γηπέδου. Η μέθοδος που χρησιμοποιήσαμε επιτρέπει την αποτελεσματικότερη επιλογή για την ανίχνευση παικτών-μπάλας σε μακρινά σουτ από υψηλής ευκρίνειας βίντεο. Επομένως καταλήξαμε στην χρήση του αλγορίθμου YOLO που αποφέρει καλύτερα αποτελέσματα.

Η τελική επιλογή του αλγορίθμου έγινε με γνώμονα μελλοντικό project λόγω του ότι με τις απαραίτητες αναβαθμίσεις σε κάρτα γραφικών υπάρχει η δυνατότητα και για live ανάλυση.

## I. ΑΞΙΟΛΟΓΗΣΗ ΕΦΑΡΜΟΓΗΣ

Το πρώτο πράγμα που μπορεί να παρατηρήσουμε όταν προσεγγίζουμε ένα πρόβλημα όπως αυτό από τη σκοπιά του ML είναι ότι είναι πολύ δύσκολο να βρεθούν διαθέσιμα δεδομένα με αξιοπρεπή ποιότητα. Ως εκ τούτου, η χρήση του πιο γνωστού ανιχνευτή από την μεριά μας, του YoloV4 ήταν μονόδρομος. Αρχικά χρησιμοποιήσαμε τον αλγόριθμο μαζί με τα 80 είδη προ-εκπαιδευμένων κλάσεων πάνω στο διαθέσιμο βίντεο του αγώνα, αναμένοντας την αρχική αναγνώριση των παικτών.

```
# Convert darknet weights to tensorflow model
python save_model.py --model yolov4

# Run yolov4 deep sort object tracker on video
python object_tracker.py --video ./data/video/test.mp4 --output ./outputs/demo.avi --model yolov4

# Run yolov4 deep sort object tracker on webcam (set video flag to 0)
python object_tracker.py --video 0 --output ./outputs/webcam.avi --model yolov4
```

**Εικόνα 16: Κώδικας**

Δεδομένου ότι δώσαμε μεγάλη προτεραιότητα στην ακρίβεια έναντι της ταχύτητας, τροφοδοτήσαμε το YOLO με την αρχική εικόνα ανάλυσης χρησιμοποιώντας ένα συρόμενο παράθυρο για να κάνουμε το δίκτυο να λειτουργεί σε ολόκληρο το πλαίσιο, κομμάτι - κομμάτι. Τα αποτελέσματα που αποκτήσαμε με αυτόν τον τρόπο μας επιτρέπουν να εντοπίζουμε παίκτες/δισαιτητή και μπάλα με συνέπεια.

Από την δημιουργία του ο κώδικας έχει προεπιλεγεί για την παρακολούθηση και των 80 περίπου τάξεων από το σύνολο δεδομένων coco, σε αυτό εκπαιδεύεται το προ-εκπαιδευμένο μοντέλο YOLOv4. Ωστόσο, μπορούμε εύκολα να προσαρμόσουμε μερικές γραμμές κώδικα για να παρακολουθήσουμε οποιαδήποτε 1 ή συνδυασμό των 80 κλάσεων. Είναι εξαιρετικά απλό να φιλτράρουμε μόνο την κατηγορία ατόμων ή μόνο την κατηγορία αυτοκινήτου που είναι πιο συνηθισμένα.

Για να φιλτράρουμε μια συγκεκριμένη επιλογή τάξεων το μόνο που χρειάζεται να κάνουμε είναι να σχολιάσουμε τη γραμμή 159 και να διαγράψουμε τη γραμμή 162 από το αρχείο `object_tracker.py`

```
155 # by default allow all classes in .names file
156 #allowed_classes = list(class_names.values())
157
158 # custom allowed classes (uncomment line below to customize tracker for only people)
159 allowed_classes = ['person', 'car']
```

**Εικόνα 17: Κώδικας**

Αφού φιλτράραμε κάποιες κατηγορίες, τον τροποποιήσαμε να αναζητάει μόνο την κατηγορία 'person' προκειμένου να μην εμφανιστεί περαιτέρω θόρυβος στο αποτέλεσμα μας. Στόχος μας ήταν η όσο πιο ακριβής αναγνώριση των παικτών καθ' όλη την διάρκεια του βίντεο.



**Εικόνα 18: Παρακολούθηση παίκτη χρησιμοποιώντας τον Yolov3 και OpenCV**

Εδώ θα χρειαστεί να αναφέρουμε ότι ο αλγόριθμος παρουσίαζε πρόβλημα όταν κάποιος παίκτης εξαφανιζόταν από την πλήρη εικόνα και αφαιρούσε την αρχική του αρίθμηση στον tracker. Επομένως επιλέξαμε για την ανάλυση παίκτες που εμφανίζονται καθ' όλη την διάρκεια του βίντεο. Περικόψαμε το αρχικό αρχείο στα 50'' προκειμένου να ελαχιστοποιήσουμε τον χρόνο που χρειάζεται ο αλγόριθμος για να σκανάρει όλα τα διαθέσιμα ρικελ. Μετέπειτα σχεδιάσαμε τα κουτιά για τους εντοπισμένους παίκτες και εμφανίσαμε τις "ουρές" τους για τα προηγούμενα δέκα καρέ. Στη συνέχεια, θελήσαμε να παρακολουθήσουμε μέσω της συσκευή αναπαραγωγής και να τις εκχωρήσουμε μοναδικές ταυτότητες σε κάθε παίκτη.

Υπάρχουν διάφορες μέθοδοι για την παρακολούθηση αντικειμένων επιλέξαμε να χρησιμοποιήσουμε τον αλγόριθμο SORT του Alex Bewley (απλή παρακολούθηση σε απευθείας σύνδεση και σε πραγματικό χρόνο). Η SORT είναι βασίζεται στην εφαρμογή ενός οπτικού πλαισίου παρακολούθησης πολλαπλών αντικειμένων και βασίζεται σε στοιχειώδεις συσχετίσεις δεδομένων και τεχνικές εκτίμησης κατάστασης . Έχει σχεδιαστεί για διαδικτυακές εφαρμογές παρακολούθησης όπου είναι διαθέσιμα μόνο παλαιότερα ή τρέχοντα καρέ και έτσι η μέθοδος παράγει ταυτότητες αντικειμένων εν κινήσει. Θα χρειαστεί να βρούμε 4 σταθερά points ως σημείο αναφοράς και να εντοπίσουμε τους συντονισμούς από το βίντεο και την παγωμένη εικόνα.

Αρχικά εντοπίζω 4 σημεία αναφοράς από το βίντεο συνήθως λίγο πιο μέσα από το περιθώριο της εικόνας και λαμβάνω τους συντονισμούς μέσω των pixel.

```
np.array([
    [1, 47], # Upper left
    [878, 54], # Upper right
    [1019, 544], # Lower right
    [1, 546] # Lower left
])
```

**Εικόνα 19: Κώδικας**

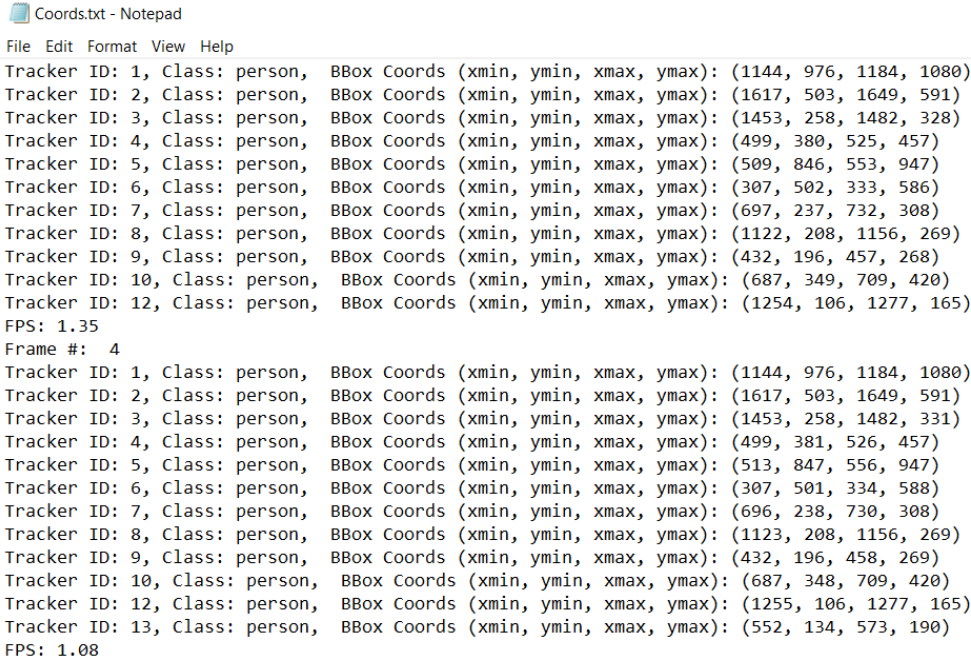
Δεν είδαμε πραγματικά πολύ σταθερά σημεία αναφοράς από το βίντεο, επομένως εντοπίσαμε περίπου 4 σημεία και σημειώσαμε αυτές τις θέσεις στην bird eye view οπτική και πήραμε τους αντίστοιχους συντονισμούς των pixel. Θα ήταν πιο ακριβές εάν τα σημεία αναφοράς είναι πιο ισχυρά-σταθερά.



```
np.array([
    [871, 37], # Upper left
    [1490, 39], # Upper right
    [1458, 959], # Lower right
    [1061, 955] # Lower left
])
```

**Εικόνα 20: Τα 4 σημεία αναφοράς που σημειώνονται με birds eye view(κόκκινη κηλίδα)**

Αποθηκεύσαμε σε μια λίστα κατά τη διάρκεια της διαδικασίας της ανίχνευσης και τη μετατρέψαμε σε πλαίσιο δεδομένων pandas για περαιτέρω ανάλυση. Όλα τα coordinates των παικτών αποθηκευόντουσαν απευθείας σε ένα txt αρχείο προκειμένου να χρησιμοποιηθούν για την μετέπειτα οπτικοποίηση των παικτών.

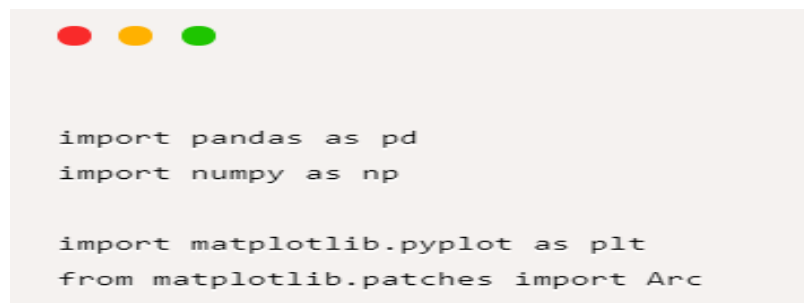


```
Coords.txt - Notepad
File Edit Format View Help
Tracker ID: 1, Class: person, BBox Coords (xmin, ymin, xmax, ymax): (1144, 976, 1184, 1080)
Tracker ID: 2, Class: person, BBox Coords (xmin, ymin, xmax, ymax): (1617, 503, 1649, 591)
Tracker ID: 3, Class: person, BBox Coords (xmin, ymin, xmax, ymax): (1453, 258, 1482, 328)
Tracker ID: 4, Class: person, BBox Coords (xmin, ymin, xmax, ymax): (499, 380, 525, 457)
Tracker ID: 5, Class: person, BBox Coords (xmin, ymin, xmax, ymax): (509, 846, 553, 947)
Tracker ID: 6, Class: person, BBox Coords (xmin, ymin, xmax, ymax): (307, 502, 333, 586)
Tracker ID: 7, Class: person, BBox Coords (xmin, ymin, xmax, ymax): (697, 237, 732, 308)
Tracker ID: 8, Class: person, BBox Coords (xmin, ymin, xmax, ymax): (1122, 208, 1156, 269)
Tracker ID: 9, Class: person, BBox Coords (xmin, ymin, xmax, ymax): (432, 196, 457, 268)
Tracker ID: 10, Class: person, BBox Coords (xmin, ymin, xmax, ymax): (687, 349, 709, 420)
Tracker ID: 12, Class: person, BBox Coords (xmin, ymin, xmax, ymax): (1254, 106, 1277, 165)
FPS: 1.35
Frame #: 4
Tracker ID: 1, Class: person, BBox Coords (xmin, ymin, xmax, ymax): (1144, 976, 1184, 1080)
Tracker ID: 2, Class: person, BBox Coords (xmin, ymin, xmax, ymax): (1617, 503, 1649, 591)
Tracker ID: 3, Class: person, BBox Coords (xmin, ymin, xmax, ymax): (1453, 258, 1482, 331)
Tracker ID: 4, Class: person, BBox Coords (xmin, ymin, xmax, ymax): (499, 381, 526, 457)
Tracker ID: 5, Class: person, BBox Coords (xmin, ymin, xmax, ymax): (513, 847, 556, 947)
Tracker ID: 6, Class: person, BBox Coords (xmin, ymin, xmax, ymax): (307, 501, 334, 588)
Tracker ID: 7, Class: person, BBox Coords (xmin, ymin, xmax, ymax): (696, 238, 730, 308)
Tracker ID: 8, Class: person, BBox Coords (xmin, ymin, xmax, ymax): (1123, 208, 1156, 269)
Tracker ID: 9, Class: person, BBox Coords (xmin, ymin, xmax, ymax): (432, 196, 458, 269)
Tracker ID: 10, Class: person, BBox Coords (xmin, ymin, xmax, ymax): (687, 348, 709, 420)
Tracker ID: 12, Class: person, BBox Coords (xmin, ymin, xmax, ymax): (1255, 106, 1277, 165)
Tracker ID: 13, Class: person, BBox Coords (xmin, ymin, xmax, ymax): (552, 134, 573, 190)
FPS: 1.08
```

**Εικόνα 21:Αποθηκευμένο αποτέλεσμα tracker**

Το επόμενο βήμα και αφού βρήκαμε τις σωστές διαστάσεις μέσω προσχεδιασμένου template της Orta δημιουργήσαμε ένα εικονικό γήπεδο προκειμένου να αναπαραστήσουμε τις κινήσεις των παικτών με την μορφή του heatmap. Διαλέξαμε να χρησιμοποιήσουμε τον “παίκτη 24” που είναι αριστερό μπασκετμπολίστα, προκειμένου να δώσουμε μια πιο ορθή ανάλυση του παίκτη μέσα από τις κινήσεις του στο γήπεδο.

Εφαρμόσαμε κάποια πακέτα δεδομένων όπως το pandas, numpy, seaborn, matplotlib που βοηθούν στη δημιουργία γηπέδων σε visualizations.



```
import pandas as pd
import numpy as np

import matplotlib.pyplot as plt
from matplotlib.patches import Arc
```

**Εικόνα 22: Κώδικας**

Δεν υπάρχει κάτι πραγματικά περίπλοκο σε αυτό μόλις χωρίσαμε τα συστατικά του μέρη. Το πρώτο βήμα είναι απλά μια συλλογή γραμμών και σχημάτων. Θα χρησιμοποιήσουμε τις

μετρήσεις συντεταγμένων και θα τις μετατρέψουμε στο επιθυμητό μέγεθος και σχέδιο ενός γηπέδου.

Πρώτα ορίσαμε ορισμένες μεταβλητές που θα χρησιμοποιήσουμε κατά τη διάρκεια της σχεδίασης. Και αρχίσαμε την οριοθέτηση του γηπέδου.

```
4 y_min = 0
5 y_max = 68
6
7 line_color = "grey"
8 line_thickness = 1.5
9 background = "w" # w = white
10 point_size = 20
11
12
13 pitch_x = [0,5.8,11.5,17,50,83,88.5,94.2,100] # σημεία x του γηπέδου
14
15 """
16 Εδώ εμφανίζονται διάφορες γραμμές στο μήκος του γηπέδου
17
18 [goal line, six yard box, penalty spot, edge of box,
19  halfway line, edge of box, penalty spot, six yard box, goal line]
20
21 """
22
23 pitch_y = [0, 21.1, 36.6, 50, 63.2, 78.9, 100]
24
25 """
26 Εδώ εμφανίζονται διάφορες γραμμές στο πλάτος του γηπέδου
27
28 [sideline, edge of box, six yard box, centre of pitch,
29  six yard box, edge of box, sideline ]
30
31 """
32
33 goal_y = [45.2, 54.8] # goal posts
34
```

**Εικόνα 23: Κώδικας**

Μετάπειτα πρέπει να τα μετατρέψουμε σε μέτρα, δεδομένου ότι αυτά είναι 100 x 100. Μπορούμε είτε να διαιρέσουμε τις συντεταγμένες με 100 και να τις πολλαπλασιάσουμε με το μέγιστο μέγεθος των x ή y μας, αλλά μπορούμε να τις απλοποιήσουμε πολλαπλασιάζοντας τις συντεταγμένες x και y με το λόγο της διαφοράς. δηλ.,  $105 / 100 = 1,05$ ,  $68 / 100 = 0,68$

Αντί να χρειαστεί να υπολογίσουμε αυτό σε κάθε τιμή, μπορούμε να χρησιμοποιήσουμε την έννοια της λίστας

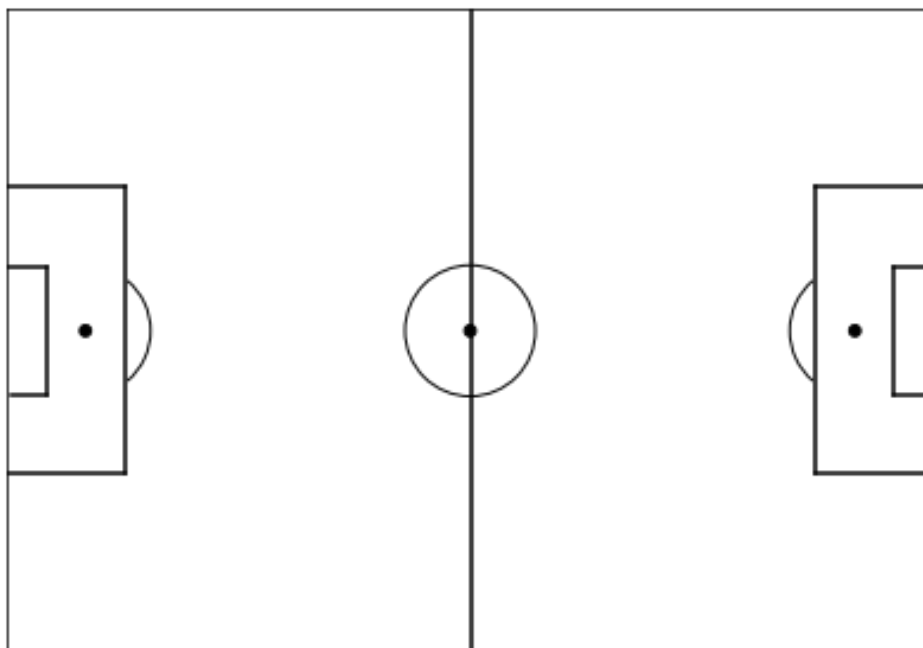
```
pitch_x = [item * 1.05 for item in pitch_x]

pitch_y = [item * 0.68 for item in pitch_y]

goal_y = [item * 0.68 for item in goal_y]
```

**Εικόνα 24: Κώδικας**

Στην τελικό αποτέλεσμα συνδυάσαμε και άλλες παραμέτρους όπως τα ημικύκλια στις περιοχές και στην σέντρα καθώς επίσης και τα σημεία του πέναλτι. Το γήπεδο ολοκληρωμένο έχει την παρακάτω εμφάνιση:



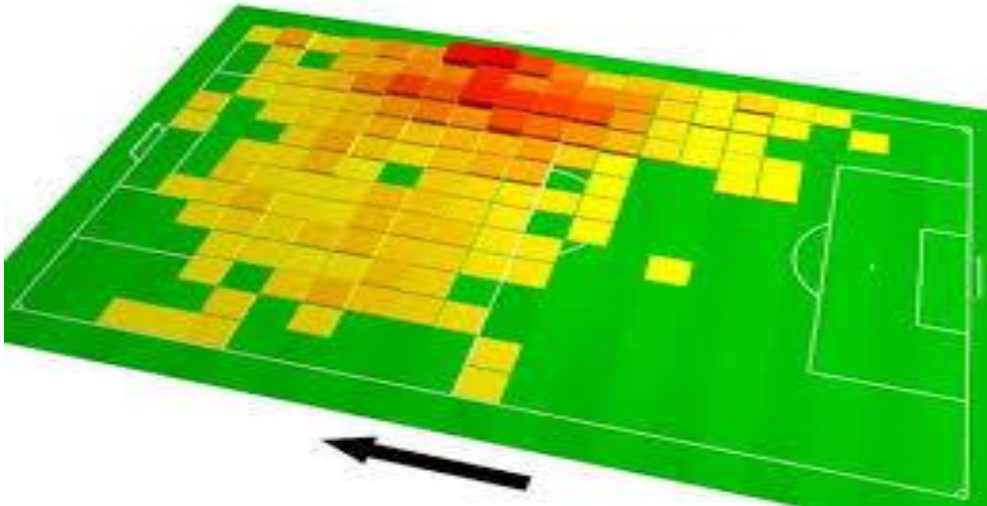
**Εικόνα 25: Τελικός σχεδιασμός γηπέδου**

Αφού σχεδιάσαμε το γήπεδο σειρά πήρε η εμφάνιση της κίνησης του παίκτη στο γήπεδο κατά την καταγραφή του βίντεο που αναλύσαμε και αυτό έγινε με τον παρακάτω τρόπο.

## HEATMAPS

Η χρήση των χαρτών heatmap είναι ευρέως γνωστή και ειδικότερα στον χώρο του αθλητισμού όπου χρησιμοποιούνται από αναλυτές και μέσα μαζικής ενημέρωσης για να απεικονίσουν την περιοχή μέσα στην οποία ήταν παρών ένας παίκτης. Δίνει την καλύτερη δυνατή οπτικοποίηση των κινήσεων καθώς παρουσιάζει τα σημεία στα οποία το αντικείμενο που παρακολουθούμε βρίσκεται πιο πολύ ώρα. Αυτό το επιτυγχάνουμε με την παλέτα των χρωμάτων, καθώς όσο πιο σκούρα η απόχρωση τόσο πιο πολύ ώρα πέρασε από αυτό το σημείο.

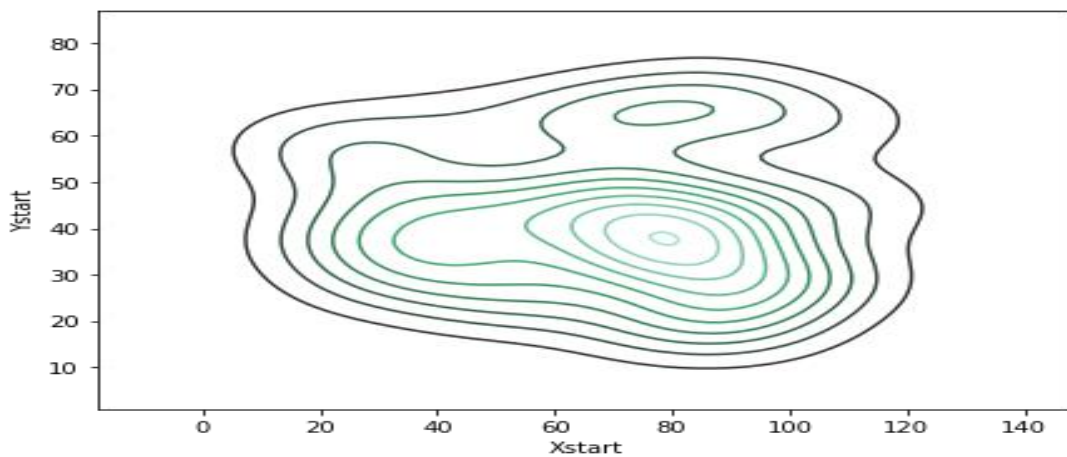
Ενώ μπορεί να υπάρχει κάποια συζήτηση για το πόσο είναι χρήσιμες (δεν μας παρουσιάζουν αν οι ενέργειες/οι κινήσεις είναι καλές ή κακές), Συχνά μπορεί να είναι πολύ αισθητικά ευχάριστες και ελκυστικές, εξ ου και η δημοτικότητά τους.



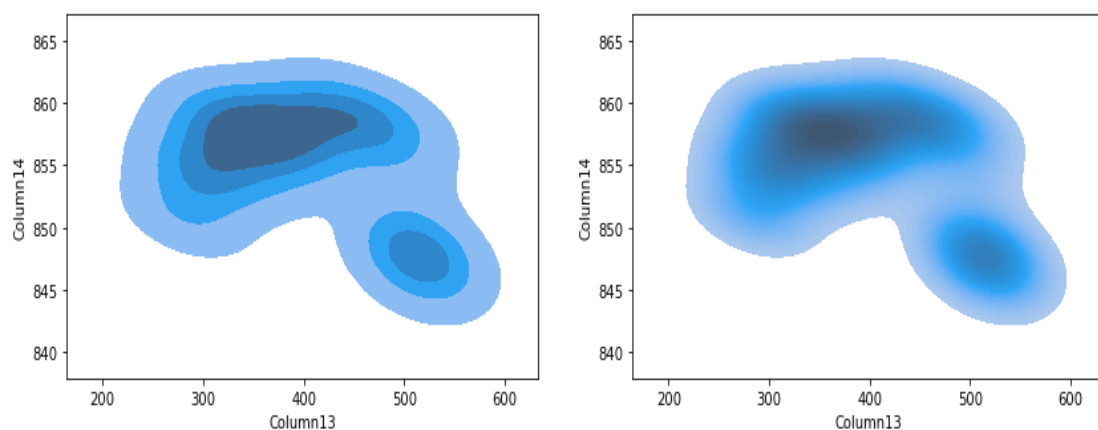
**Εικόνα 26: Πίνακας heatmap**

Η ενότητα Seaborn της Python καθιστά τη σχεδίαση ενός τακτοποιημένου συνόλου δεδομένων απίστευτα εύκολη διαδικασία με το «.kdeplot ()».

Η γραφική παράσταση με τις στήλες συντεταγμένων x και y ως ορίσματα μας εμφανίζει το παρακάτω αποτέλεσμα:



**Εικόνα 27: Οπτικοποίηση συντεταγμένων**

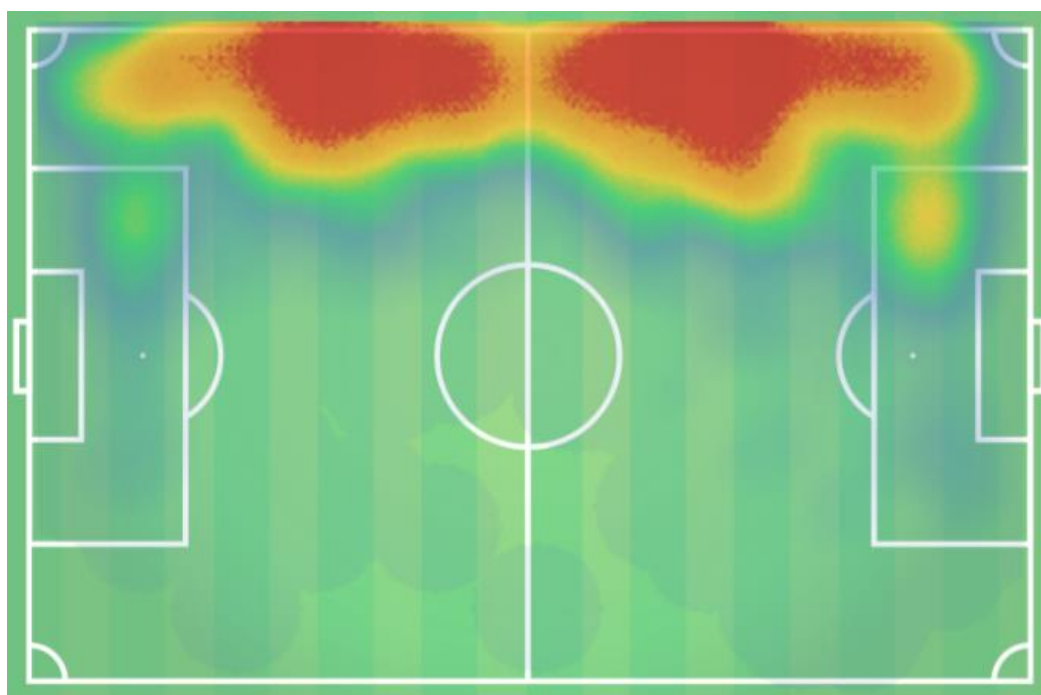


**Εικόνα 28: Οπτικοποίηση συντεταγμένων**

Τώρα που μπορούμε να προσαρμόσουμε το σχέδιο όπως μας φαίνεται κατάλληλο, απλώς θα χρειαστεί να προσθέσουμε τον χάρτη μας. Αλλάξαμε τις συντεταγμένες για να ταιριάζουν στα δεδομένα μας. Επίσης, λάβαμε υπόψιν τις γραμμές μας `xlim` και `ylim` - τις χρησιμοποιούμε για να ορίσουμε το μέγεθος του γηπέδου, έτσι ώστε ο θερμικός χάρτης να μην αναπτυχθεί και έξω από το σχεδιασμένο γήπεδο.

Επομένως έπειτα από την σύμπτυξη των παραπάνω γραφημάτων μπορούμε να παρουσιάσουμε τις κινήσεις του player 24 όπως τον όρισε ο αλγόριθμος και εμείς τον αναλύσαμε.

Το τελικό αποτέλεσμα εμφανίζεται στο παρακάτω διάγραμμα



**Εικόνα 29:Heatmap person 24**



Ακόμα και να μην γνωρίζαμε τη θέση του παίκτη γίνεται εύκολα αντιληπτό ότι ο περισσότερος χρόνος κυκλοφορίας του παίκτη είναι από την αριστερή πλευρά, και πιο συγκεκριμένα παράλληλα της πλάγιας γραμμής του γηπέδου. Ο player 24 εμφανίζει πολύπλευρη παρουσία στο γήπεδο καθώς το κόκκινο σημείο είναι μοιρασμένο και στις δύο πλευρές του γηπέδου που υποδηλώνει ένα επιθετικό ακραίο μπακ με δημιουργικές ικανότητες και αρκετή αθλητικότητα.

## 5. ΜΕΛΛΟΝΤΙΚΕΣ ΒΕΛΤΙΩΣΕΙΣ

Με τις πληροφορίες κίνησης των παικτών, είναι δυνατή η περαιτέρω ανάλυση, όπως η απόσταση και η ταχύτητα των παικτών. Η ταχύτητα εκτέλεσης της παρακολούθησης του παίκτη είναι περίπου 0,3 δευτερόλεπτα ανά καρέ με επεξεργαστή Intel i7 8<sup>th</sup> gen. Είναι δυνατόν να γίνει αυτό σε πραγματικό χρόνο χρησιμοποιώντας GPU για ορισμένες εφαρμογές, εάν είναι απαραίτητο. Επομένως μαζί με την live ανάλυση θα μπορούσε να δημιουργηθεί και ένα bird eye view template που θα εμφανίζει σε κανονική ροή αγώνα όλους τους παίκτες και τις κινήσεις τους μέσα στο γήπεδο. Το ποδόσφαιρο έχει να κάνει με τους χώρους, αλλά τα ποδοσφαιρικά στατιστικά βασίζονται κυρίως σε μεμονωμένες ενέργειες που δεν αφορούν το περιβάλλον. Ακόμα και σύνθετες μετρήσεις όπως τα expected goals συχνά εξετάζουν μόνο τις θέσεις της ίδιας της δράσης χωρίς να λαμβάνουν υπόψη τη γενική θέση της στιγμής που συνέβη. Αυτό ισχύει ιδιαίτερα όταν προσπαθούμε να αξιολογήσουμε τους αμυντικούς. Χρειαζόμαστε νέες μετρήσεις, πάνω από πράγματα όπως τα τάκλιν και τα κλεψίματα, επειδή η τοποθέτηση είναι η πιο σημαντική αμυντική ικανότητα για έναν αμυντικό - και δεν μετριέται καθόλου. Μια λύση σε αυτό το πρόβλημα είναι τα διαγράμματα Voronoi. Στα μαθηματικά, ένα διάγραμμα Voronoi είναι μια διαίρεση ενός επιπέδου σε περιοχές με βάση την απόσταση από σημεία σε ένα συγκεκριμένο υποσύνολο του επιπέδου. Αυτό το σύνολο σημείων (που ονομάζονται τοποθεσίες ή γεννήτριες) έχει καθοριστεί εκ των προτέρων και για κάθε γεννήτρια υπάρχει μια αντίστοιχη περιοχή που αποτελείται από όλα τα σημεία πιο κοντά στην γεννήτρια από ό, τι σε οποιοδήποτε άλλο. Πως είναι όμως κάτι τέτοιο χρήσιμο; Αν ο στόχος του αμυντικού είναι να φτάσει οπουδήποτε στο γήπεδο γρηγορότερα από τον αντίπαλο παίκτη, ο καλύτερος τρόπος για να το πετύχει αυτό είναι να είναι πιο κοντά από τον επιθετικό στις επικίνδυνες περιοχές. Αυτή δεν είναι μια νέα παρατήρηση καθώς ο Ντέιβιντ Σάμπτερ, συγγραφέας του Soccermetrics, μίλησε και έγραψε για αυτό, και ένας φοιτητής Φυσικής από την Κορέα έγραψε ένα paper χρησιμοποιώντας δεδομένα βιντεοπαιχνιδιών. Τελευταία, με όλο και περισσότερα δεδομένα θέσης να συλλέγονται από επαγγελματίες παρόχους δεδομένων, αυτή η ανάλυση έχει γίνει πιο συνηθισμένη.

## ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1]. Amir Hossein Alavi,(2017), “Investigation into Tracking Football Players from Video Streams Produced by Cameras Set Up for TV Broadcasting”, American Journal of Engineering Research (AJER), ISSN : 2278 - 3075 Website: [www.ijitee.org](http://www.ijitee.org) Volume-9 Issue-2S, DECEMBER 2019 Published by: Blue Eyes Intelligence Engineering and Sciences Publication.
- [2]. Takuma Narizuka & Yoshihiro Yamazaki,(2018), “Statistical properties for directional alignment and chasing of players in football games”, Waseda University, Shinjuku, Tokyo. EPL (Europhysics Letters), Volume 116, Number 6.
- [3]. Santiago, Catarina & Sousa, Armando & Reis, Luís & Estriga, Maria et al.,(2010), “Real Time Colour Based Player Tracking in Indoor Sports”, Computational Vision and Medical Image Processing, book: [Computational Vision and Medical Image Processing \(pp.17-35\)](#), ISBN: 978-94-007-0010-9
- [4]. Joseph Redmon & Ali Farhadi,(2018),“YOLOv3:An Incremental Improvement”, University of Washington. Available: <https://pjreddie.com/darknet/yolo/>
- [5]. Komorowski, J.; Kurzejamski, G. and Sarwas, G. (2020). FootAndBall: Integrated Player and Ball Detector. In Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 5: VISAPP, ISBN 978-989-758-402-2; ISSN 2184-4321, pages 47-56.
- [6]. Hawk-Eye,[Online].Available: [https://nihal111.github.io/hawk\\_eye/](https://nihal111.github.io/hawk_eye/)
- [7]. Zibatron, “Mapping Camera Coordinates to a 2D Floor Plan”,(2019), [Online]. Available: <https://zbigatron.com/mapping-camera-coordinates-to-a-2d-floor-plan/>
- [8]. Basile Roth, “A social distancing detector using a Tensorflow object detection model, Python and OpenCV”,(2017), [Online]. Available: <https://deepnote.com/@deepnote/A-social-distancing-detector-using-a-Tensorflow-object-detection-model-Python-and-OpenCV-KBcEvWejRjGyjy2YnxiP5Q> access at 2018
- [9]. CK, “How to track football players using Yolo, SORT and Opencv.”,(2020),[Online].Available: <https://towardsdatascience.com/how-to-track-football-players-using-yolo-sort-and-opencv-6c58f71120b8>
- [10]. Adrian Rosebrock, “Deep Learning with OpenCV”,(2017), [Online].Available: <https://www.pyimagesearch.com/2017/08/21/deep-learning-with-opencv/>
- [11]. Yolo – Darknet, “YOLO: Real-Time Object Detection”, [Online]. Available: <https://pjreddie.com/darknet/yolo/>

- [12]. Ricardo Tavares, "Using Voronoi Diagrams in Football",(2019), [Online]. Available: <https://medium.com/football-crunching/using-voronoi-diagrams-in-football-ca730ea81c05>
- [13]. Rohith Gandhi, "R-CNN, Fast R-CNN, Faster R-CNN, YOLO — Object Detection Algorithms,(2018), [Online]. Available: <https://towardsdatascience.com/r-cnn-fast-r-cnn-faster-r-cnn-yolo-object-detection-algorithms-36d53571365e>
- [14]. J.M.F. Rodrigues, P.J.S. Cardoso, T. Vilas, S. Bruno, P. Rodrigues, A.Belguinha and C. Gomes et al.,(2020), "A computer vision based web application for tracking soccer players", University of the Algarve, [Online]. Available: <https://core.ac.uk/download/pdf/61523875.pdf>
- [15]. Adrian Rosebrock, "Tracking multiple objects with OpenCV",(2018),[Online].Available: <https://www.pyimagesearch.com/2018/08/06/tracking-multiple-objects-with-opencv/>
- [16]. Surya Remanan, "Beginner's Guide to Object Detection Algorithms",(2019),[Online].Available: <https://medium.com/analytics-vidhya/beginners-guide-to-object-detection-algorithms-6620fb31c375>
- [17]. Kanan Vyas, "Player and football detection using Opencv-Python in FIFA match.",(2018),[Online].Available: <https://kananvyas.medium.com/player-and-football-detection-using-opencv-python-in-fifa-match-6fd2e4e373f0>
- [18]. Galang Imanta, "Deep Learning for Detecting Football Players Using Convolutional Neural Network, Tensorflow and OpenCV", (2019), [Online]. Available: <https://medium.com/galang-imanta/deep-learning-for-detecting-football-players-using-convolutional-neural-network-tensorflow-and-a0158251ed7b>
- [19]. Priya Dwivedi, "Highlight Action Area in Soccer using Tensorflow",(2019),[Online].Available: <https://towardsdatascience.com/highlight-action-area-in-soccer-using-tensorflow-1c59d644b404>
- [20]. Christopher Jenness, "NBA player tracking visualization and analysis",(2017),[Online].Available: <https://github.com/christopherjenness/NBA-player-movement>
- [21]. Park Chansung, "YOLOv2 object detection using Darkflow",(2018),[Online].Available: <https://towardsdatascience.com/yolov2-object-detection-using-darkflow-83db6aa5cf5f>
- [22]. Enrique a., "Detecting custom objects in images/video using YOLO with Darkflow",(2018),[Online].Available: <https://medium.com/coinmonks/detecting-custom-objects-in-images-video-using-yolo-with-darkflow-1ff119fa002f>

- [23]. Stephan Janssen, "The journey towards creating a Basketball mini-map", (2019), [Online]. Available: <https://www.linkedin.com/pulse/journey-towards-creating-basketball-mini-map-stephan-janssen/>
- [24]. Guillermo Martinez Arastey, "COMPUTER VISION IN SPORT", (2020), [Online]. Available: <https://www.sportperformanceanalysis.com/article/computer-vision-in-sport>
- [25]. John Burn-Murdoch, "How data analysis helps football clubs make better signings", Financial Times (2018), [Online]. Available: <https://www.ft.com/content/84aa8b5e-c1a9-11e8-84cd-9e601db069b8>
- [26]. Nicolo' Lucchesi, "Football Games Analysis from video stream with Machine Learning", (2021), [Online]. Available: <https://towardsdatascience.com/football-games-analysis-from-video-stream-with-machine-learning-745e62b36295>
- [27]. Praveen Pavithran, "Object Detection with Yolo Python and OpenCV- Yolo 2", (2020), [Online]. Available: <https://cloudxlab.com/blog/object-detection-yolo-and-python-pydarknet/>
- [28]. The AI Guy, "How to Build an Object Tracker Using YOLOv3, Deep SORT and TensorFlow", (2020), [Online]. Available: [https://www.youtube.com/watch?v=Cf1INvUsvkM&ab\\_channel=TheAIGuy](https://www.youtube.com/watch?v=Cf1INvUsvkM&ab_channel=TheAIGuy)
- [29]. Henrique Morimitsua, Isabelle Blochb, Roberto M. Cesar-Jr et al., "Exploring structure for long-term tracking of multiple objects in sports videos". Institute of Mathematics and Statistics, University of São Paulo, São Paulo, Brazil. Volume 159, June 2017, Pages 89-104.
- [30]. Michael Herrmann, Martin Hoerniga, Bernd Radiga et al., "Online Multi-player Tracking in Monocular Soccer Videos", Technische Universität München. AASRI Procedia 8 ( 2014 ) 30–37.
- [31]. Sebastian Gerke, Antje Linnemann, Karsten Müller et al., "Soccer player recognition using spatial constellation features and jersey number recognition, Elsevier Journal [Volume 159](#), June 2017, Pages 105-115.
- [32]. Varuna De Silva, Mike Caine 1, James Skinner, Safak Dogan, Ahmet Kondoz, Tilson Peter, Elliott Axtell, Matt Birnie and Ben Smith et al., "Player Tracking Data Analytics as a Tool for Physical Performance Management in Football: A Case Study from Chelsea Football Club Academy", Loughborough University. Vol. 10, No. 5, Article 59, Publication date: September 2019.
- [33]. Sermetcan Baysal, "MODEL FIELD PARTICLES WITH POSITIONAL APPEARANCE LEARNING FOR SPORTS PLAYER TRACKING", (2016), Bilkent University. [Online]. Available: <http://repository.bilkent.edu.tr/handle/11693/30128>

- [34]. M. Manafifarda, H. Ebadi b, H. Abrishami Moghaddam et al.,(2017), "A survey on player tracking in soccer videos", . Toosi University of Technology. DOI:10.1016/j.cviu.2017.02.002