Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Αγρονόμων Τοπογράφων Μηχανικών και Μηχανικών Γεωπληροφορικής

Κατάτμηση και παρακολούθηση αντικειμένων από δεδομένα εικόνων από κινητή πλατφόρμα σε περιβάλλον εμπορικού λιμένα

Διπλωματική Εργασία

Πρίφτης Ιωάννης

Αθήνα, Ιούνιος 2023

Περίληψη

Τα τελευταία χρόνια με την ανάπτυξη μεθόδων στο τομέα της τεχνητής νοημοσύνης η ανίχνευση αντικειμένων απασχολεί ολοένα και περισσότερο την επιστημονική κοινότητα. Η γνώση της θέσης ενός αντικειμένου είναι απαραίτητη σε διάφορες διαδικασίες αυτοματισμού.
Η συγκεκριμένη εργασία επικεντρώνεται στην αξιοποίηση αλγορίθμων της Όρασης Υπολογιστών και της Βαθιάς Μάθησης για την κατάτμηση και τη παρακολούθηση του αντικειμένου της αρπάγης. Η αρπάγη είναι ένα εξάρτημα που προσαρμόζεται σε γερανούς και επιτρέπει την μεταφορά κοντέινερ. Το συγκεκριμένο αντικείμενο συναντάται σε σημεία του εμπορικού δικτύου που γίνεται αλλαγή μέσου μεταφοράς των κοντέινερ όπως είναι λιμάνια και σταθμοί τρένων. Η ανίχνευση και η παρακολούθηση της αρπάγης είναι σημαντική σε εφαρμογές ασφαλείας σε βιομηχανικά περιβάλλοντα όπου γίνεται φορτοεκφόρτωση κοντέινερ συμβάλλοντας στην αποφυγή ατυχημάτων.
Αρχικά γίνεται αναφορά στο απαραίτητο θεωρητικό υπόβαθρο για την ανάπτυξη αλγορίθμων παρακολούθησης της αρπάγης. Συγκεκριμένα περιγράφονται οι βασικές αρχές λειτουργίας των τεχνητών νευρωνικών δικτύων που χρησιμοποιούνται από τεχνικές Βαθιάς Μάθησης. Έπειτα γίνεται ανάλυση της λειτουργίας βασικών μοντέλων νευρωνικών δικτύων που χρησιμοποιούνται για την ανίχνευση και κατάτμηση αντικειμένων. Επιπλέον παρατίθενται μερικές συμπληρωματικές έννοιες που βοηθούν τόσο στην κατανόηση ορισμένων σταδίων της διαδικασίας ανίχνευσης με τεχνητά νευρωνικά δίκτυα όσο και με χρήση τεχνικών Όρασης Υπολογιστών.
Σε επόμενο στάδιο της εργασίας περιγράφεται η μεθοδολογία που χρησιμοποιήθηκε για την ανάπτυξη και εφαρμογή αλγορίθμων. Συγκεκριμένα έγινε ανάπτυξη ενός αλγορίθμου που κάνει χρήση τεχνικών όρασης υπολογιστών και εφαρμογή δύο αλγορίθμων τεχνητών νευρωνικών δικτύων. Σε αυτό το κομμάτι παρατίθενται λεπτομέρειες για την μεθοδολογία που ακολουθήθηκε για τον σχεδιασμό του αλγόριθμου με βάση κλασικές μεθόδους όρασης υπολογιστών, της εκπαίδευσης και ρύθμισης βασικών παραμέτρων των νευρωνικών δικτύων Mask Rcnn και Yolact.
Στο επόμενο τμήμα της εργασίας γίνεται η αξιολόγηση των αλγορίθμων. Οι τρεις αλγόριθμοι αξιολογούνται με βάση της προδιαγραφές που έχουν οριστεί σε εισαγωγικό στάδιο της εργασίας. Η διαδικασία της αξιολόγησης αξιοποιεί δεδομένα από πραγματικό περιβάλλον εφαρμογής.
Τέλος γίνεται μία σύνοψη της όλης διαδικασίας από την εφαρμογή των αλγορίθμων έως και την αξιολόγηση των αποτελεσμάτων και την εξαγωγή του πορίσματος. Επιπλέον αναφέρονται ορισμένες πτυχές που έχουν περιθώριο βελτίωσης καθώς και ιδέες για μελλοντική εργασία.

NATIONAL TECHNICAL UNIVERSITY OF ATHENS

School of Rural, Surveying and Geoinformatics Engineering

Segmentation and tracking of objects based on image data from a mobile platform in commercial port environment

Diploma Thesis

Priftis Ioannis

Athens, June 2023

Abstract

In recent years, with the development of methods in the field of artificial intelligence, the detection of objects has increasingly attracted the interest of the scientific community. The information of the location of an object is essential in various processes that are automated.

This particular thesis focuses on the utilization of Computer Vision and Deep Learning algorithms for the 2D segmentation and tracking of the spreader. The spreader is an object that is attached to cranes and allows to transport containers. This particular object is found in trading network points where the mean of transportation of the containers changes such as in ports and rail stations. The detection and tracking of the spreader are crucial in safety applications that support accident prevention in industrial areas where loading and unloading of containers takes place.

First the necessary theoretical background for the development of the spreader tracking algorithms is provided. Specifically, there are references to the basic operating principles of Artificial Neural Networks that are being used in Deep Learning methods. Next the operation of basic Neural Network models that are being used for object detection and segmentation is analyzed. In addition, some additional concepts are listed that help to understand some stages of the object detection process using neural networks and using Computer Vision methods.

In the next section of the thesis, the methodology is described. Specifically, there was development of an algorithm that uses Computer Vision techniques and of two algorithms that use neural network techniques. This section lists the details about the methodology used for the design of the Computer Vision algorithm, as well as the training and setting of the basic parameters of the neural networks Mask Rcnn and Yolact.

In the next part of the thesis the algorithms are being evaluated. The three algorithms are being evaluated based on the requirements set on the introduction of the thesis. The evaluation process utilizes real world data to assess the outcomes of the detection and tracking algorithms.

Finally, a summary of the whole process is reported, from the use of the algorithms to the evaluation of the results and the extraction of the conclusions. In addition, some aspects that have room for improvement are mentioned as well as ideas for future work.

Περιεχόμενα

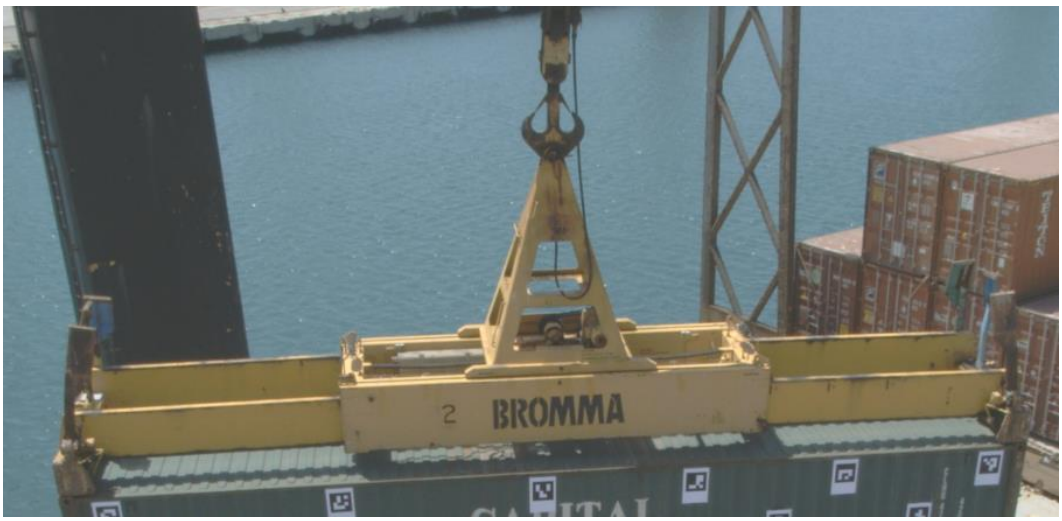
Πίνακας Σχημάτων

Πίνακας Πινάκων

Κεφάλαιο 1: Εισαγωγή

1.1 Κίνητρο

Στο σύγχρονο εμπόριο η πρακτική της μεταφοράς πρώτων υλών και προϊόντων με την χρήση κοντέινερ συνέβαλε στην δημιουργία ενός διεθνώς συστήματος μεταφοράς. Το σύστημα αυτό αξιοποιεί μεταξύ άλλων πλωτά, εναέρια και μέσα σταθερής τροχιάς για την μεταφορά των κοντέινερ. Κατά την διαδικασία της μεταφοράς των προϊόντων συχνά παρεμβάλλονται σημεία αλλαγής μέσου μεταφοράς όπως είναι τα λιμάνια όπου το κοντέινερ αλλάζουν μέσω από το πλωτό μέσο σε κάποιο άλλο μέσο. Στα σημεία αυτά αξιοποιούνται γερανοί εξοπλισμένοι με μία αρπάγη για την διαδικασία της φόρτωσης και εκφόρτωσης των κοντέινερ από το ένα μέσο στο άλλο. Η αρπάγη είναι μια μηχανική κατασκευή που εγκαθίσταται στους γερανούς και εφαρμόζει στις πλευρές του κοντέινερ δίνοντας την δυνατότητα να στον γερανό να σηκώσει το σύστημα αρπάγη-κοντέινερ.



Σχήμα 1: Αρπάγη

Η αυξανόμενη ανάγκη για μεταφορά αγαθών απαιτεί ταχείες τακτικές φόρτωσης- εκφόρτωσης που όμως συσχετίζονται με την ασφάλεια των εργατών. Αυτό οφείλεται στο ότι αυξάνουν την κούραση των χειριστών των γερανών και μπορούν να υπονομεύσουν την ασφάλεια στο εργασιακό περιβάλλον που γίνεται αυτή η διαδικασία. Σύμφωνα με έρευνα (Budiyanto & H.Fernanda, 2020) τα περισσότερα ατυχήματα σε περιβάλλον φόρτωσης – εκφόρτωσης κοντέινερ οφείλονται στον ανθρώπινο παράγοντα.

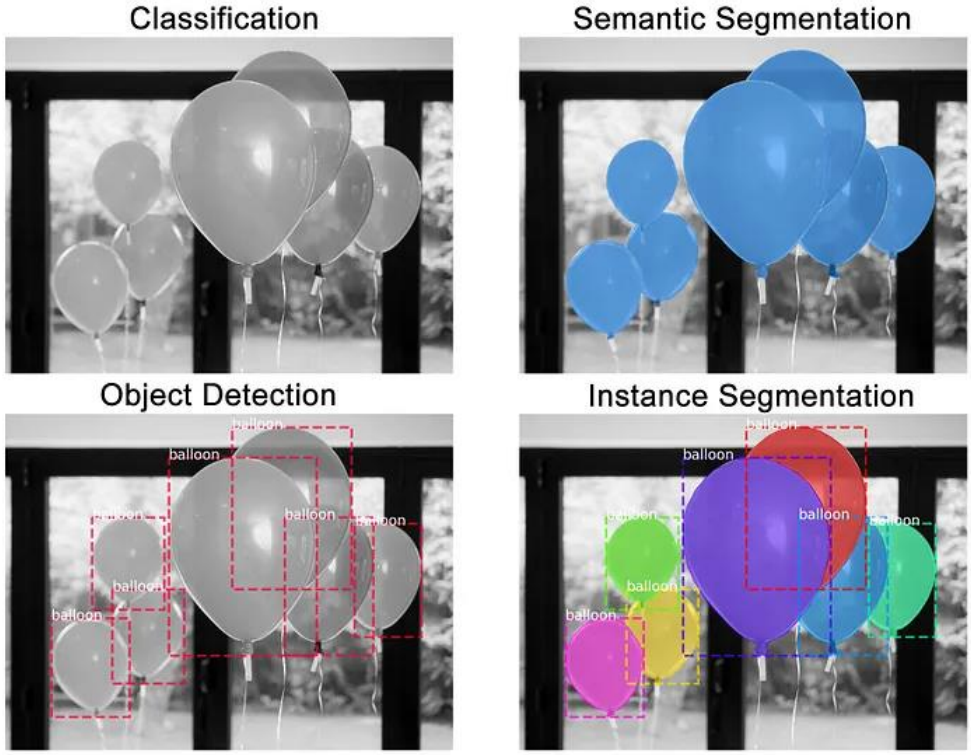
Η παρούσα εργασία εστιάζει στην αξιοποίηση εικόνων από κάμερα που είναι τοποθετημένη στο γερανό ώστε να παρακολουθείται η αρπάγη με τεχνικές όρασης υπολογιστών (Computer Vision) και με τεχνικές τεχνητών νευρωνικών δικτύων (Artificial Neural Networks). Η όραση υπολογιστών επιτρέπει στον υπολογιστή να εξάγει πληροφορία από δεδομένα όπως είναι εικόνες, σειρές βίντεο και άλλα οπτικά μέσα. Τα τεχνητά νευρωνικά δίκτυα επιτρέπουν στον υπολογιστή να μαθαίνει να εκπονεί διαδικασίες αφού πρώτα εκπαιδευτεί δίνοντας του δεδομένα και τα επιθυμητά αποτελέσματα των διαδικασιών αυτών. Κύριο κίνητρο για την εκπόνηση της εργασίας είναι η μείωση των ατυχημάτων σε περιβάλλον εργασίας όπου γίνεται η χρήση συστήματος γερανού-αρπάγης. Γνωρίζοντας την θέση της αρπάγης στο χώρο ανά πάσα στιγμή μπορεί να αποφευχθεί κάποιο ατύχημα που οφείλεται στον τραυματισμό εργατών από το κοντέινερ

ή από την ίδια την αρπάγη. Για την αποφυγή του ατυχήματος είναι δυνατό να αξιοποιηθούν τεχνολογίες Internet of Things (IoT), οι οποίες εφαρμόζονται όλο και περισσότερο σε σύγχρονα βιομηχανικά περιβάλλοντα όπου αυτοματοποιούνται διάφορες διαδικασίες. Η τεχνολογία αυτή αφορά διάφορες υπολογιστικές, ψηφιακές και μηχανικές συσκευές (αισθητήρες, υπολογιστές, κινητά τηλέφωνα, κ.α.) που είναι συνδεδεμένες και αλληλοεπιδρούν μεταξύ τους με την χρήση ενός δικτύου. Σε περίπτωση που η σχετική θέση ενός εργάτη και της αρπάγης είναι μικρή τότε μπορεί ένα αυτοματοποιημένο σύστημα να σημάνει το κίνδυνο και έτσι να απομακρυνθεί ο εργάτης ή και να σταματήσει ο χειρισμός του γερανού. Ένα τέτοιο σύστημα θα μπορούσε να αποτελέσει δικλείδα ασφαλείας ώστε να αποφευχθούν ατυχήματα που οφείλονται στον ανθρώπινο παράγοντα.

1.2 Σκοπός

Σκοπός της παρούσας εργασίας είναι η ανάπτυξη αλγορίθμου που με χρήση κάμερας θα πραγματοποιεί παρακολούθηση της αρπάγης. Υπάρχουν διαφορετικά είδη αλγορίθμων ανίχνευσης αντικειμένων τα οποία εξάγουν διαφορετικά παράγωγα. Συγκεκριμένα υπάρχουν:

- Αλγόριθμοι ταξινόμησης (Classification) που εξάγουν το πόρισμα για το αν στην εικόνα υπάρχει ή όχι αντικείμενο που ανήκει στις προς ανίχνευση κατηγορίες.
- Αλγόριθμοι σημασιολογικής κατάτμησης (Semantic Segmentation) που εξάγουν πολύγωνα που περιέχουν όλα τα αντικείμενα που ανήκουν σε μια προς ανίχνευση κλάση.
- Αλγόριθμοι ανίχνευσης αντικειμένων (Object Detection) που εξάγουν ένα πλαίσιο που περικλείει το καθένα αντικείμενο που ανιχνεύθηκε και την κλάση στην οποία ανήκει.
- Αλγόριθμοι εννοιολογικής κατάτμησης (Instance Segmentation) που εξάγουν μια μάσκα, δηλαδή ένα πολύγωνο που περιέχει το κάθε αντικείμενο που ανιχνεύθηκε και την κλάση στην οποία ανήκει



Σχήμα 2: Διαφορετικά παράγωγα αλγορίθμων ανίχνευσης (Πηγή : (engineering.matterport, 2018)

Ο αλγόριθμος που θα χρησιμοποιηθεί θα πρέπει να κάνει εννοιολογική κατάτμηση της αρπάγης καθώς είναι θεμιτό να απομονωθεί η αρπάγη από το περιβάλλον. Η απλή ανίχνευση αντικειμένου με πλαίσιο δεν απομονώνει το αντικείμενο ανίχνευσης από το περιβάλλον καθώς εντός του πλαισίου ενδέχεται να υπάρχει και τμήμα άλλου αντικειμένου.

Ο αλγόριθμος αυτός θα συμβάλει στην πρόληψη ατυχημάτων που σχετίζονται με τα κοντέινερ. Επιλέγεται να γίνει ανίχνευση και κατάτμηση της αρπάγης και όχι των κοντέινερ διότι τα κοντέινερ παρουσιάζουν πολλές διαφοροποιήσεις τόσο στο μέγεθος όσο και στο χρώμα. Συνεπώς η ανίχνευση των κοντέινερ είναι πιο περίπλοκη διαδικασία. Επιπλέον η αρπάγη μπορεί να μην μεταφέρει κάποιο κοντέινερ κατά τη διάρκεια της κίνησής της, παρόλα αυτά είναι απαραίτητη η παρακολούθησή της για την αποφυγή ατυχημάτων λόγω σύγκρουσης.

1.3 Απαιτήσεις

Ο παραπάνω αλγόριθμος θα πρέπει να έχει ορισμένα χαρακτηριστικά ώστε να είναι αξιοποιήσιμος για την εφαρμογή του σε ένα βιομηχανικό περιβάλλον όπως είναι αυτό του λιμένα. Αρχικά θα πρέπει να μπορεί να λειτουργεί σε πραγματικό χρόνο ώστε η ανίχνευση του αντικειμένου να είναι συνεχής. Έπειτα θα πρέπει να είναι λειτουργικός σε ένα δυναμικό περιβάλλον. Αυτό σημαίνει να μπορεί να ανιχνεύει την αρπάγη παρόλο που αλλάζουν οι συνθήκες φωτισμού. Η αλλαγή των συνθηκών φωτισμού μπορεί να οφείλεται στην αλλαγή των καιρικών συνθηκών, την ώρα της ημέρας και την εποχή του χρόνου.

Κεφάλαιο 2: Ανάλυση τρέχουσας κατάστασης και σχετικό υπόβαθρο

2.1 Στοιχεία στατιστικής

- Πείραμα τύχης: Το πείραμα, δηλαδή μία διαδικασία της οποία το αποτέλεσμα δεν μπορεί να προβλεφθεί. Σε ένα πείραμα τύχης το σύνολο όλων των δυνατών αποτελεσμάτων καλείτε δειγματικός χώρος (Ω).

- Πιθανότητα: Έστω ένας δειγματικός χώρος Ω που περιέχει k στοιχεία. Επίσης το γεγονός Α περιέχει ένα υποσύνολο λ του δειγματικού χώρου ($\lambda \leq k$). Η πιθανότητα Ρ του γεγονότος Α είναι το κλάσμα που έχει αριθμητή το σύνολο των στοιχείων του γεγονότος Α και παρονομαστή το σύνολο των στοιχείων του δειγματικού χώρου που ανήκει το Α.

$$P(A) = \frac{\lambda}{k}$$

- Τυχαία Μεταβλητή: Η τυχαία μεταβλητή είναι μία συνάρτηση η οποία παίρνει τυχαίες τιμές. Χρησιμοποιείται συχνά για να περιγράψει τα αποτελέσματα ενός πειράματος. Για παράδειγμα κατά το πείραμα ρίψης ενός ζαριού η πιθανότητα να έρθει ο αριθμός 6 είναι $P(X = 6) = {}^{1}/_{6}$.

  Μια τυχαία μεταβλητή μπορεί να είναι διακριτή όταν το παίρνει τιμές σε ένα αριθμήσιμο σύνολο και τότε ισχύει ότι: $\sum_{k=1}^{\infty} P(X = x_k) = 1$
  Μια τυχαία μεταβλητή είναι συνεχής όταν παίρνει τιμές από μη αριθμήσιμο σύνολο και τότε ισχύει: $\int_{-\infty}^{\infty} f(x)dx = 1$

- Συνάρτηση μάζας πιθανότητας: Συνάρτηση που περιγράφει την κατανομή μιας διακριτής τυχαίας μεταβλητής.

$$f(x_i) = P(X = x_i)$$

- Συνάρτηση πυκνότητας πιθανότητας: Πρόκειται για συνάρτηση που δίνει την πιθανότητα μία συνεχής τυχαία μεταβλητή να βρίσκεται μεταξύ σε ένα κλειστό διάστημα [a,b]

$$P(a < X \leq b) = \int_a^b f(x)dx$$

- Συνάρτηση κατανομής πιθανότητας: Δίνει την πιθανότητα μία συνεχής τυχαία μεταβλητή να παίρνει τιμές μικρότερες ή ίσες από μία τιμή

$$P(X \leq x) = \int_{-\infty}^{x} f(w)dw$$

- Μέση τιμή (μ ή ΕΧ): Η μέση τιμή αποτελεί στοιχειώδες στατιστικό μέτρο για μία τυχαία μεταβλητή. Πρόκειται για μία γενίκευση του μέσου όρου και δείχνει το κέντρο βάρος γύρω από το οποίο παίρνει τιμές η τυχαία μεταβλητή.
  Για μία διακριτή τυχαία μεταβλητή ισχύει $\mu = \sum_i x_i \cdot P(X = x_i)$.

Για μία συνεχή τυχαία μεταβλητή ισχύει $\mu = \int_{-\infty}^{\infty} x \cdot f(x)dx$

- Διασπορά ή Διακύμανση (VarX ή σ^2): Στατιστικό μέτρο που εκφράζει το πόσο απέχουν οι τιμές της τυχαίας μεταβλητής από τη μέση τιμή. Η διασπορά ορίζεται ως $\sigma^2 = VarX = E(X-\mu)^2$. Η τιμή σ, όπου $\sigma = \sqrt{VarX}$ ονομάζεται τυπική απόκλιση
  Για μία διακριτή τυχαία μεταβλητή ισχύει $\sigma^2 = \sum_i (x_i - \mu)^2 \cdot P(X = x_i)$.
  Για μία συνεχή τυχαία μεταβλητή ισχύει $\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 \cdot f(x)dx$

- Συνδιακύμανση (Covariecne): Η συνδιακύμανση σαν στατιστικό μέτρο δείχνει το πως σχετίζονται μεταξύ τους δύο τυχαίες μεταβλητές X και Y. Όταν οι συνδιακύμανση δύο τυχαίων μεταβλητών είναι ίση με το μηδέν τότε αυτές θεωρούνται ασυσχέτιστες.
  $$Cov(X,Y) = E(X-\mu_X)(Y-\mu_Y)$$
  Για X=Y προκύπτει ο τύπος της διασποράς.

- Πολυδιάστατη τυχαία μεταβλητή: Αντί για μία τυχαία μεταβλητή είναι δυνατό να υπάρχει διάνυσμα που οι συντελεστές του να είναι τυχαίες μεταβλητές
  $$X = (X_1, X_2, \cdots, X_n)^T$$
  Αντίστοιχα ορίζονται τα διανύσματα πιθανότητας , μέσης τιμής και διακύμανσης
  $$P(X) = (P(X) = x_1, P(X) = x_2, \dots, P(X) = x_n)^T$$
  $$EX = (EX_1, EX_2, \cdots, EX_n)^T$$
  $$VarX = (EX_1 - \mu_1)^2, (EX_2 - \mu_2)^2, \cdots, (EX_n - \mu_n)^2)^T$$

- Πίνακας συνδιακύμανσης: Ορίζεται ως ένας πίνακας που δίνει την διακύμανση ανάμεσα σε τυχαίες μεταβλητές. Έστω δύο πολυδιάστατες τυχαίες μεταβλητές $X = (X_1, X_2, \cdots, X_n)^T$ και $Y = (Y_1, Y_2, \cdots, Y_n)^T$
  $$\Sigma = Cov(X,Y) = \begin{bmatrix} E(X_1-\mu_1)(Y_1-\mu_1) & \cdots & E(X_1-\mu 1)(Y_n-\mu_n) \\ \vdots & \ddots & \vdots \\ E(X_n-\mu_n)(Y_1-\mu_1) & \cdots & E(X_n-\mu_n)(Y_n-\mu_n) \end{bmatrix}$$

2.2 Φίλτρο Kalman

Το φίλτρο Kalman είναι ένας αλγόριθμος που χρησιμοποιείται για να εκτιμήσει την κατάσταση ενός συστήματος (System State) με βάση παρατηρήσεις πάνω στο σύστημα. Με τη χρήση του φίλτρου Kalman μπορεί να εκτιμηθεί η μελλοντική κατάσταση του συστήματος με βάση της προηγούμενες καταστάσεις.

Έστω ότι για ένα σύστημα υπάρχει μία εξίσωση διαφορών που περιγράφει την κατάσταση του συστήματος. Στην εξίσωση αυτή γίνεται η υπόθεση ότι η κατάσταση του συστήματος προκύπτει από την προηγούμενη κατάσταση του συστήματος και έχει την μορφή:

$$x_k = A \cdot x_{k-1} + B \cdot u_{k-1} + w_{k-1}$$

Όπου x_k : διάνυσμα κατάστασης
A: πίνακας μετάβασης από την χρονική στιγμή k-1στην χρονική στιγμή k
B: πίνακας δεδομένων ελέγχου $u_κ$
w_k: διάνυσμα θορύβου

Για το συγκεκριμένο σύστημα υπάρχει διάνυσμα μετρήσεων το οποίο είναι γραμμικός συνδυασμός της γνωστής κατάστασης του συστήματος και περιγράφεται από μία εξίσωση της μορφής:

$$z_k = H \cdot x_k + v_k$$

Όπου z_k : το διάνυσμα μετρήσεων
H: πίνακας μετρήσεων
v_k: διάνυσμα θορύβου

Δύο σημαντικές έννοιες που χρησιμοποιούνται στις συναρτήσεις του αλγορίθμου για να αναφερθούν στις μεταβλητές του διανύσματος κατάστασης x_k και του πίνακα συνδιακύμανσης σφάλματος P_k είναι οι εκ των προτέρων εκτιμήσεις (a priori) $x_{\bar{k}}, P_{\bar{k}}$ και οι βελτιωμένες εκτιμήσεις (a posteriori) $\hat{x}_k, \hat{P}_k$.

Ο αλγόριθμος του φίλτρου Kalman αποτελεί μια επαναληπτική διαδικασία δύο σταδίων. Στο πρώτο στάδιο γίνονται οι a priori εκτιμήσεις της κατάστασης του συστήματος και του πίνακα συνδιακύμανσης σφάλματος με την χρήση των εξισώσεων:

$$\bar{x}_k = A \cdot \hat{x}_{k-1} + B \cdot u_{k-1}$$

$$\bar{P}_k = A \cdot \hat{P}_{k-1} \cdot A^T + Q$$

Όπου $\hat{x}_{k-1}$: η a posteriori προηγούμενη κατάσταση του συστήματος
$\hat{P}_{k-1}$: ο a posteriori πίνακας συνδιακύμανσης σφάλματος
Q: πίνακας συνδιακύμανσης του θορύβου του διανύσματος w_k

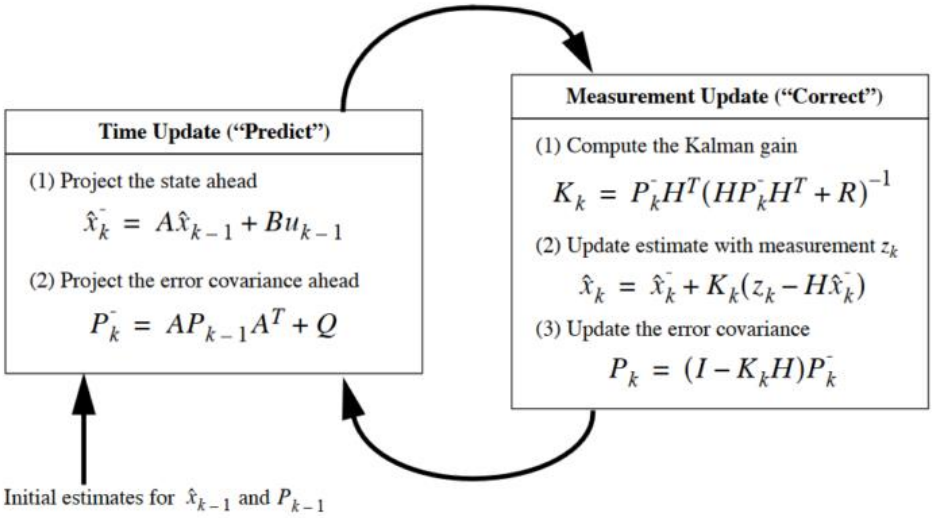
Στο δεύτερο στάδιο αρχικά υπολογίζεται ο πίνακας κέρδους Kalman K_k (Kalman Gain). Έπειτα ο πίνακας K_k αξιοποιείται για τον υπολογισμό των a posteriori τιμών της κατάστασης του συστήματος και του πίνακα συνδιακύμανσης σφάλματος με την χρήση των εξισώσεων:

$$K_k = \bar{P}_k \cdot H^T \cdot (H \cdot \bar{P}_k \cdot H^T + R)^{-1}$$

$$\hat{x}_k = \bar{x}_k + K_k(z_k - H \cdot \bar{x}_k)$$

$$\hat{P}_k. = (I - K_k \cdot H) \cdot \bar{P}_k$$

Όπου R: πίνακας συνδιακύμανσης του θορύβου του διανύσματος v_k



Σχήμα 3: Φίλτρο Kalman (Πηγή: (Welch & Bishop, 2006))

Μία από τις πολλές εφαρμογές του φίλτρου Kalman είναι η παρακολούθηση αντικειμένου σε εικόνες από σειρές βίντεο. Τέτοιος αλγόριθμος χρησιμοποιείτε στη συγκεκριμένη εργασία (Βλέπε 3.1.2). Σε αυτή την περίπτωση χρησιμοποιείται ένα φίλτρο Kalman 2 διαστάσεων για την πρόβλεψη των συντεταγμένων (x,y) του υπό παρακολούθηση αντικειμένου. Από την φυσική είναι γνωστές οι εξισώσεις της κίνησης:

$$x_k = x_{k-1} + v_{k-1} \cdot \Delta t + \frac{1}{2} \cdot a_{k-1} \cdot \Delta t^2$$

$$v_k = v_{k-1} + a_{k-1} \cdot \Delta t$$

Όπου x : η θέση
v: η ταχύτητα
a: η επιτάχυνση
Δt: το χρονικό διάστημα μεταξύ της στιγμής k-1 και k

Χρησιμοποιώντας τις εξισώσεις της κίνησης για να περιγράψουμε το σύστημα που έχει διάνυσμα κατάστασης $\mathbf{x}_k$ προκύπτει:

$$\boldsymbol{x}_k = \begin{bmatrix} x_\kappa \\ y_\kappa \\ vx_\kappa \\ vy_\kappa \end{bmatrix} = \begin{bmatrix} x_{k-1} + vx_{k-1} \cdot \Delta t + \frac{1}{2} \cdot ax_{k-1} \cdot \Delta t^2 \\ y_{k-1} + vy_{k-1} \cdot \Delta t + \frac{1}{2} \cdot ay_{k-1} \cdot \Delta t^2 \\ vx_k = vx_{k-1} + ax_{k-1} \cdot \Delta t \\ vy_k = vy_{k-1} + ay_{k-1} \cdot \Delta t \end{bmatrix}$$

άρα

$$\boldsymbol{x}_k = \begin{bmatrix} x_\kappa \\ y_\kappa \\ vx_\kappa \\ vy_\kappa \end{bmatrix} = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_{\kappa-1} \\ y_{\kappa-1} \\ vx_{\kappa-1} \\ vy_{\kappa-1} \end{bmatrix} + \begin{bmatrix} \frac{1}{2} \cdot \Delta t^2 & 0 \\ 0 & \frac{1}{2} \cdot \Delta t^2 \\ \Delta t & 0 \\ 0 & \Delta t \end{bmatrix} \cdot \begin{bmatrix} ax_{k-1} \\ ay_{k-1} \end{bmatrix}$$

άρα

$$\boldsymbol{x}_k = \begin{bmatrix} x_\kappa \\ y_\kappa \\ vx_\kappa \\ vy_\kappa \end{bmatrix} = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \boldsymbol{x}_{k-1} + \begin{bmatrix} \frac{1}{2} \cdot \Delta t^2 & 0 \\ 0 & \frac{1}{2} \cdot \Delta t^2 \\ \Delta t & 0 \\ 0 & \Delta t \end{bmatrix} \cdot a_{k-1}$$

Άρα για το συγκεκριμένο σύστημα ο πίνακας μετάβασης A και ο πίνακας δεδομένων ελέγχου B είναι:

$$A = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \qquad B = \begin{bmatrix} \frac{1}{2} \cdot \Delta t^2 & 0 \\ 0 & \frac{1}{2} \cdot \Delta t^2 \\ \Delta t & 0 \\ 0 & \Delta t \end{bmatrix}$$

Όσο αναφορά το μοντέλο μετρήσεων ισχύει από τις εξισώσεις του φίλτρου Kalman ότι:

$$z_k = H \cdot \boldsymbol{x}_k + v_k$$

άρα

$$z_k = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} x_\kappa \\ y_\kappa \\ vx_\kappa \\ vy_\kappa \end{bmatrix} + v_k$$

Οι τιμές x_k, y_k αντιστοιχούν στις εικονοσυντεταγμένες και είναι οι μόνες μετρήσεις που γίνονται πάνω στην εικόνα. Ο πίνακας μετρήσεων παίρνει συντελεστές διάφορους του μηδενός μόνο για τιμές που αποτελούν μετρήσεις και για αυτό διαμορφώνεται ως:

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

Τέλος μένει να υπολογιστούν ο πίνακας οι πίνακες συμμεταβλητότητας Q και R :

$$Q = \begin{bmatrix} \sigma_x^2 & 0 & \sigma_x \cdot \sigma_{vx} & 0 \\ 0 & \sigma_y^2 & 0 & \sigma_y \cdot \sigma_{vy} \\ \sigma_{vx} \cdot \sigma_x & 0 & \sigma_{vx}^2 & 0 \\ 0 & \sigma_{vy} \cdot \sigma_y & 0 & \sigma_{vy}^2 \end{bmatrix}$$

$$R = \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{bmatrix}$$

Έχοντας υπολογίσει τους παραπάνω πίνακες είναι δυνατό να υπολογιστεί με την χρήση του αλγόριθμου Kalman μία πρόβλεψη για τις εικονοσυντεταγμένες x_k, y_k που αναφέρονται στο καρέ κ από τις εικονοσυντεταγμένες x_{k-1}, y_{k-1} που αναφέρονται στο προηγούμενο καρέ κ -1.

2.3 Νευρωνικά δίκτυα και βαθιά μάθηση

2.3.1 Μηχανική μάθηση

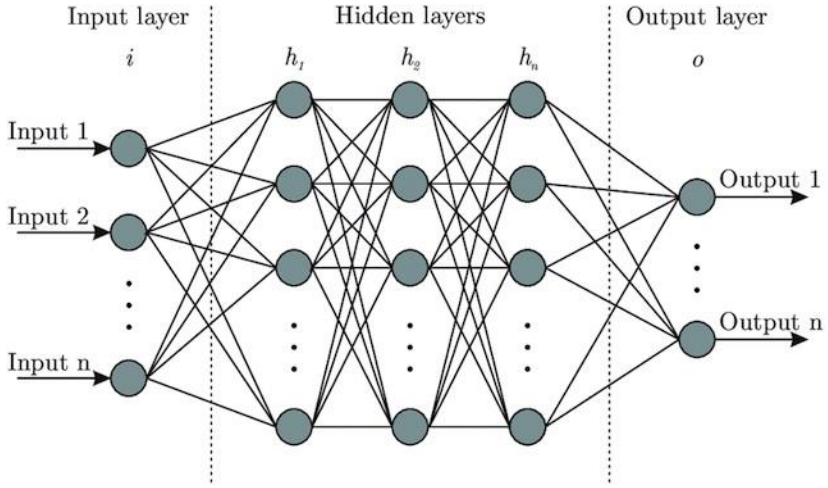
Η μηχανική μάθηση (Machine Learning) αποτελεί τμήμα της ευρύτερης προγραμματιστικής ενότητας της τεχνητής νοημοσύνης (A.I) που προσομοιάζει τον τρόπο με τον οποίο οι άνθρωποι μαθαίνουν.

Η βαθιά μάθηση (Deep Learning) αποτελεί υποκατηγορία της μηχανικής μάθησης. Τα μοντέλα βαθιάς μάθησης αξιοποιούν μεταξύ άλλων αλγορίθμων τα τεχνητά νευρωνικά δίκτυα για την επίλυση ορισμένων προβλημάτων όπως είναι η ανίχνευση αντικειμένων σε εικόνες που δεν είναι εύκολο να αντιμετωπιστούν με κλασσικές προγραμματιστικές μεθόδους.

2.3.2 Νευρωνικά δίκτυα

Τα τεχνητά νευρωνικά δίκτυα είναι αλγόριθμοι εμπνευσμένοι από την λειτουργία των βιολογικών νευρωνικών δικτύων οι οποίοι δίνουν την δυνατότητα στον υπολογιστή να μάθει από παραδείγματα. Αποτελούνται από ένα δίκτυο διασυνδεδεμένων κόμβων που ονομάζονται τεχνητοί νευρώνες.

Νευρωνικό δίκτυο είναι ένα δίκτυο διασυνδεδεμένων τεχνητών νευρώνων οι οποίοι είναι οργανωμένοι σε επίπεδα. Το πρώτο επίπεδο νευρώνων ονομάζεται επίπεδο εισαγωγής (Input layer) και οι νευρώνες του δέχονται σαν είσοδο τα δεδομένα που θα επεξεργαστεί το δίκτυο. Το τελευταίο επίπεδο νευρώνων ονομάζεται επίπεδο εξαγωγής (Output layer) και οι έξοδοι των νευρώνων αυτού του επιπέδου είναι το αποτέλεσμα της επεξεργασίας του δικτύου. Τα ενδιάμεσα επίπεδα ονομάζονται κρυφά επίπεδα (Hidden layers).



Σχήμα 4: Τεχνητό νευρωνικό δίκτυο (Πηγή: (dev.to, 2023))

Για να γίνει αντιληπτή η λειτουργία των τεχνητών νευρωνικών δικτύων κανείς πρέπει να κατανοήσει τον τρόπο λειτουργίας του βασικού τους δομικού συστατικού, του τεχνητού νευρώνα. Η λειτουργία ενός τεχνητού νευρώνα είναι να επεξεργάζεται τις τιμές εισόδου και να υπολογίζει μία τιμή εξόδου. Ο υπολογισμός της τιμής εξόδου γίνεται με την χρήση της συνάρτησης ενεργοποίησης. Υπάρχουν διάφορα είδη τεχνητών νευρώνων που χρησιμοποιούν διαφορετικές συναρτήσεις ενεργοποίησης για τον υπολογισμό της τιμής εξόδου. Η συναρτησιακή μορφή του τεχνητού νευρώνα είναι:

$$y = f\left(b + \sum_{j}^{n} x_j \cdot w_j\right)$$

όπου : y : η τιμή εξόδου
w_j : το βάρος της τιμής εισόδου j
x_j : η τιμής εισόδου j
b : η τιμή κατωφλιού

2.3.2.1 Συναρτήσεις ενεργοποίησης

Όπως προαναφέρθηκε η συνάρτηση ενεργοποίησης χρησιμοποιείται για τον υπολογισμό της τιμής εξόδου ενός τεχνητού νευρώνα. Μερικές από τις συναρτήσεις ενεργοποίησης που συναντώνται πιο συχνά σε τεχνητά νευρωνικά δίκτυα είναι :

- Η σιγμοειδής συνάρτησης (Sigmoid Function):

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

- Η συνάρτηση ανορθωμένης γραμμικής μονάδας - ReLU (Rectified Linear Unit):

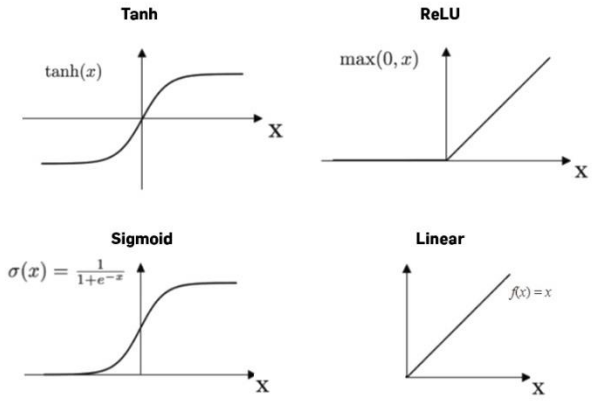
$$ReLU(x) = \begin{cases} x, x > 0 \\ 0, x \leqslant 0 \end{cases}$$

- Η συνάρτηση υπερβολικής εφαπτομένης (Hyperbolic tangent):

$$tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$$

- Η συνάρτηση γραμμικής ενεργοποίησης (Linear activation):

$$LA(x) = x$$



Σχήμα 5: Συναρτήσεις ενεργοποίησης (Πηγή: (machine-learning.paperspace, 2023))

2.3.2.2 Διαδικασία εκπαίδευσης

Η εκπαίδευση ενός τεχνικού νευρωνικού δικτύου είναι μια διαδικασία κατά την οποία οι μεταβλητές του μοντέλου (βάρη και τιμές κατωφλίου) μεταβάλλονται με απώτερο σκοπό τα δεδομένα εξόδου του μοντέλου

να προσεγγίζουν τις επιθυμητές τους τιμές.

Στην πρώτη αλληλεπίδραση του μοντέλου με τα δεδομένα εκπαίδευσης οι μεταβλητές του μοντέλου είναι τυχαίες (ή έχουν οριστεί με κάποια άλλη διαδικασία). Στην συνέχεια υπολογίζονται στο στάδιο της εμπρόσθιας διάδοσης (Forward propagation) οι έξοδοι όλων των νευρώνων του δικτύου από το πρώτο κρυφό επίπεδο μέχρι και το επίπεδο εξόδου. Έπειτα με την χρήση της συνάρτησης κόστους (Loss function) γίνεται σύγκριση των αληθών δεδομένων (Ground truth) των δεδομένων εκπαίδευσης με την πρόβλεψη του μοντέλου και υπολογίζεται το συνολικό κόστος.

Σκοπός είναι η μείωση του κόστους με την πάροδο των επαναλήψεων του μοντέλου, δηλαδή οι προβλέψεις του μοντέλου να προσεγγίζουν τις επιθυμητές. Για να μειωθεί το συνολικό κόστος χρειάζεται να ανανεωθούν οι παράμετροι του μοντέλου. Για την μείωση του κόστους αξιοποιείται μία συνάρτηση βελτιστοποίησης που υπολογίζει με βάση τις παραγώγους της συνάρτησης κόστους ως προς τις παραμέτρους του δικτύου πόσο πρέπει να μεταβληθούν. Κατά την διαδικασία οπίσθιας διάδοσης (Backward propagation) γίνεται ενημέρωση των παραμέτρων του δικτύου με βάση τις τιμές που υπολογίστηκαν με την συνάρτηση βελτιστοποίησης.

Τα στάδια της εμπρόσθιας διάδοσης για τον υπολογισμό της εξόδου του μοντέλου και του συνολικού κόστους και της οπίσθιας διάδοσης για την ανανέωση των παραμέτρων του μοντέλου επαναλαμβάνονται κατά την διαδικασία της εκπαίδευσης. Το συνολικό κόστος θεωρητικά μειώνεται με την πάροδο των επαναλήψεων εφόσον έχουν επιλεχθεί οι κατάλληλες παράμετροι για το μοντέλο.

2.3.2.3 Συναρτήσεις κόστους

Συναρτήσεις κόστους υπάρχουν πολλές και η επιλογή της κατάλληλης γίνεται βάση της φύσης του προβλήματος που εκπαιδεύεται να λύνει το τεχνητό νευρωνικό δίκτυο. Μερικές από αυτές είναι :

- Mean Squared Error (MSE):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2$$

- Mean Absolute Error (MAE):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|^2$$

- Binary Cross-Entropy (CE)

$$CE = -\frac{1}{n}\sum_{i=1}^{n}\left(y_i \cdot log(p_i) + (1 - y_i) \cdot log(1 - p_i)\right)$$

- Categorical Cross-Entropy (CE)

$$CE = -\frac{1}{n}\sum_{j=1}^{C}\sum_{i=1}^{m} y_{ij} \cdot log\left(p_{ij}\right) \quad , \quad C \geq 2$$

- Κανονικοποιημένη εκθετική συνάρτηση (Softmax): Η συνάρτηση ενεργοποίησης softmax χρησιμοποιείται σε προβλήματα ταξινόμησης. Η συνάρτηση softmax εκφράζει την έξοδο του νευρώνα ως ένα διάνυσμα διαστάσεων όσες και οι K κλάσεις της προς ταξινόμηση εξόδου. Κάθε στοιχείο i του διανύσματος είναι η πιθανότητα να ανήκει στην το στοιχείο i στην αντίστοιχη κλάση.

$$Softmax(x)_i = \frac{e_i^x}{\sum_{j=i}^{K} e_i^x}$$

2.3.2.4 Αλγόριθμοι βελτιστοποίησης

Στόχος του αλγόριθμου βελτιστοποίησης είναι ελαχιστοποίηση του συνολικού κόστους του μοντέλου κατά την διάρκεια της εκπαίδευσης. Η συνάρτηση κόστους ορίζει μία ν-διάστατη επιφάνεια στον χώρο. Ο αλγόριθμος βελτιστοποίησης αναζητά το ελάχιστο σημείο της επιφάνειας αυτής στο οποίο οι παράμετροι του δικτύου θα είναι οι κατάλληλες ώστε οι προβλέψεις του δικτύου να έχουν το μικρότερο δυνατό κόστος.

Οι αλγόριθμοι βελτιστοποίησης που χρησιμοποιούνται πιο συχνά βασίζονται στον αλγόριθμο Κλίσης Καθόδου (Gradient Decent). Αυτοί οι αλγόριθμοι υπολογίζουν την κλίση για ένα δοσμένο σημείο της ν-διάστατης επιφάνειας που ορίζει η συνάρτηση κόστους και μεταβάλλουν της παραμέτρους του δικτύου προς την αντίθετη κατεύθυνση. Έτσι το νέο σημείο στην επόμενη επανάληψη βρίσκεται πιο κοντά στο θεωρητικά ελάχιστο της επιφάνειας, στο οποίο θα ελαχιστοποιείται και το κόστος.

2.3.2.5 Υπερπαράμετροι

Οι υπερπαράμετροι ενός νευρωνικού δικτύου είναι μεταβλητές που ορίζουν τη δομή του και επηρεάζουν την διαδικασία της εκπαίδευσης. Μερικές από τις πιο σημαντικές είναι:

- Ρυθμός εκμάθησης (Learning rate): Ο ρυθμός εκμάθησης ορίζει το πόσο μεγάλη θα είναι η αλλαγή στις παραμέτρους του μοντέλου. Η συνάρτηση βελτιστοποίησης υπολογίζει μία τιμή κατά την οποία πρέπει να μεταβληθεί η κάθε παράμετρος του δικτύου. Η μεταβολή αυτή πολλαπλασιάζεται με τον ρυθμό εκμάθησης συνεπώς ο ρυθμός εκμάθησης επηρεάζει το πόσο σημαντική θα είναι αυτή η μεταβολή. Αν ο ρυθμός εκμάθησης είναι πολύ μεγάλος τότε οι μεταβολές των παραμέτρων είναι μεγάλες και μπορεί η συνάρτηση βελτιστοποίησης να μην καταφέρνει να ελαχιστοποιήσει το κόστος. Αν ρυθμός εκμάθησης είναι πολύ μικρός τότε οι μεταβολές των παραμέτρων είναι μικρές και μπορεί η σύγκλιση του μοντέλου να καθυστερήσει πολύ ή και το μοντέλο να συγκλίνει σε κάποιο τοπικό ελάχιστο της επιφανείας της συνάρτησης κόστους και όχι στο ολικό ελάχιστο. Συνεπώς η επιλογή σωστού ρυθμού εκμάθησης είναι πολύ σημαντική για την διαδικασία της εκπαίδευσης. Συνήθως η τιμή αυτή μειώνεται με κάποια πολιτική κατά την διάρκεια της εκπαίδευσης.

- Ορμή (Momentum): Η ορμή χρησιμοποιεί την τελευταία ανανέωση των παραμέτρων κατά τον υπολογισμό της νέας ανανέωσης των παραμέτρων. Έτσι :

$$dw_n + m \cdot dw_{n-1} \ , \ 0 \leqslant m < 1$$

όπου : dw_n : η νέα ανανέωση της παραμέτρου w
dw_{n-1} : η προηγούμενη ανανέωση της παραμέτρου w
m: το μέτρο της ορμής

- Φθορά βαρών (Weight Decay): Συντελεστής που αξιοποιείται για την αποφυγή της υπερπροσαρμογής του δικτύου στα δεδομένα εκπαίδευσης (βλ. 2.3.2.6). Η τιμή αυτή του συντελεστή συμβάλει στην αποφυγή μεγάλων τιμών στα βάρη του μοντέλου που είναι θεμιτό για να αποφευχθεί η υπερπροσαρμογή του δικτύου στα δεδομένα εκπαίδευσης.

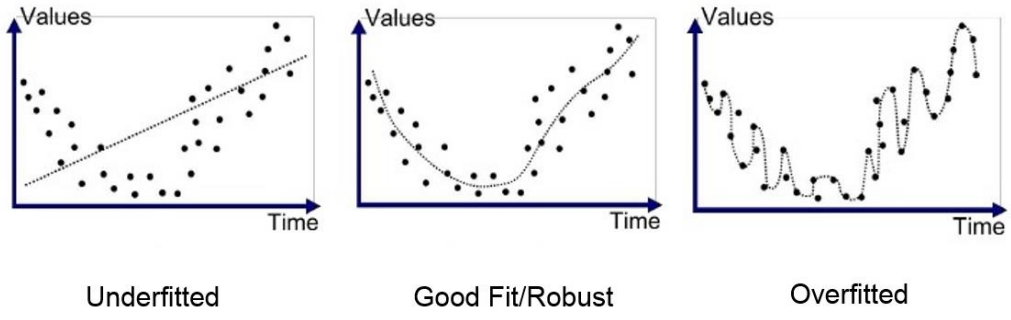
- Εποχές (Epoch): Κατά την διάρκεια μίας εποχής εκπαίδευσης το νευρωνικό δίκτυο έχει δει όλα τα δεδομένα εκπαίδευσης μία φορά.

- Μέγεθος παρτίδας (Batch size): Το μέγεθος παρτίδας είναι το υποσύνολο των δεδομένων εκπαίδευσης που θα χρησιμοποιηθεί από το δίκτυο για τον υπολογισμό του συνολικού κόστους. Όσο πιο μεγάλο είναι το μέγεθος παρτίδας τόσο πιο κανονικοποιημένη είναι η τιμή του κόστους που υπολογίζεται. Ωστόσο η χρήση μεγαλύτερου μεγέθους παρτίδας απαιτεί μεγάλο μέγεθος μνήμης.

2.3.2.6 Υπερπροσαρμογή

Στόχος της εκπαίδευσης ενός νευρωνικού δικτύου είναι το τελικό μοντέλο να μπορεί να κάνει αξιόπιστες προβλέψεις για δεδομένα εισόδου που διαφέρουν σε ένα βαθμό από τα δεδομένα με τα οποία εκπαιδεύτηκε το μοντέλο. Η ικανότητα αυτή του μοντέλου ονομάζεται γενίκευση.

Ένα συχνό πρόβλημα στην διαδικασία εκπαίδευσης μοντέλων είναι η υπερπροσαρμογή στα δεδομένα εκπαίδευσης (Overfiting). Η υπερπροσαρμογή είναι μια κατάσταση όπου τα χαρακτηριστικά που έχει μάθει να αναγνωρίζει το δίκτυο αφορούν σε τόσο μεγάλο βαθμό τα δεδομένα εκπαίδευσης που αν τα δεδομένα εισόδου διαφοροποιηθούν λίγο (διαφορετική γωνία θέασης αντικειμένου, θόρυβος) η έξοδος του μοντέλου αλλάζει και αποκλίνει από την αναμενόμενη. Έτσι ένα μοντέλο που έχει υπερπροσαρμοστεί δεν μπορεί να γενικεύσει. Αντίστοιχα υπάρχει και η έννοια της υποπροσαρμογής (Underfiting) όταν το μοντέλο δεν αποδίδει ικανοποιητικά στα δεδομένα εκπαίδευσης και κατ' επέκταση ούτε σε δεδομένα που διαφέρουν σε ένα βαθμό από αυτά. Το underfiting είναι σημάδι ότι το μοντέλο δεν προσαρμόζεται στο είδος των δεδομένων που καλείτε να επεξεργαστεί.



Σχήμα 6: Τρία μοντέλα όπου το αριστερά παρουσιάζει undefiting, το δεξιά overfiting και το κεντρικό είναι το θεμιτό που παρουσιάζει γενίκευση (Πηγή: (medium, 2023))

Για την αποφυγή της υπερπροσαρμογης ενός μοντέλου υπάρχουν διάφορες τεχνικές που ονομάζονται τεχνικές ομαλοποίησης. Οι πιο συνηθισμένες είναι οι L1 και L2. Έχει παρατηρηθεί ότι όταν το μοντέλο έχει υπερπροσαρμοστεί οι τιμές των βαρών είναι μεγάλες. Μεγάλες τιμές στα βάρη σημαίνει ότι μικρές διαφορές στα δεδομένα εισόδου μπορούν να οδηγήσουν σε μεγάλες μεταβολές στα δεδομένα εξόδου. Οι τεχνικές ομαλοποίησης L1 και L2 εισάγουν έναν όρο στην συνάρτηση κόστους με σκοπό την μείωση των τελικών βαρών του μοντέλου και συνεπώς την αποφυγή της υπερπροσαρμογής:

- L1 (κανονικοποίηση L1): Η τεχνική ομαλοποίησης L1 προσθέτει στην συνάρτηση κόστους το άθροισμα απόλυτων τιμών των βαρών του δικτύου πολλαπλασιασμένα με ένα συντελεστή λ.

$$L1 = Loss + \lambda \cdot \sum_{i=1}^{n} |wi|$$

- L2 (κανονικοποίηση L2): Η τεχνική ομαλοποίησης L2 προσθέτει στην συνάρτηση κόστους το άθροισμα των τετραγώνων των βαρών του δικτύου πολλαπλασιασμένα με ένα συντελεστή λ. Ο συντελεστής λ αποτελεί την υπερπαράμετρο της φθοράς βαρών.

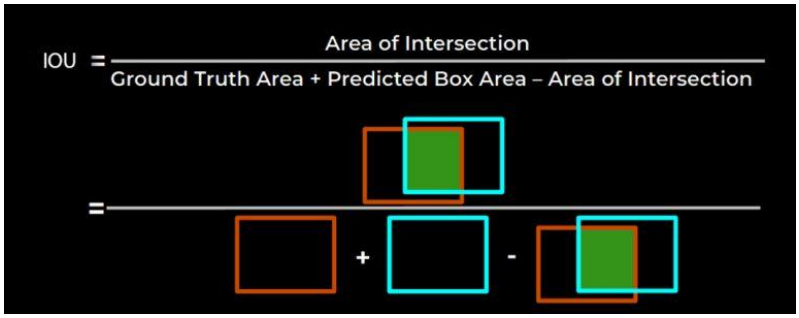
$$L2 = Loss + \lambda \cdot \sum_{i=1}^{n} wi^2$$

Στις παραπάνω τεχνικές όσο μεγαλύτερη είναι η τιμή της παραμέτρου λ τόσο μεγαλύτερη είναι η αύξηση του κόστους. Έτσι το δίκτυο για να πετύχει την σύγκλιση κατά την διαδικασία της εκπαίδευσης πρέπει να διατηρεί τις τιμές των βαρών χαμηλές.

2.3.2.7 Μέτρα απόδοσης

Τα μέτρα απόδοσης χρησιμοποιούνται για την αξιολόγηση νευρωνικών δικτύων που πραγματοποιούν ανίχνευση αντικειμένων (Object detection) ή και εννοιολογική κατάτμηση (Instance segmentation). Τα πιο διαδεδομένα μέτρα απόδοσης είναι η μέση ακρίβεια (mAP) και η μέση επανάκληση (mAR) που αξιοποιούν τις τιμές του μέτρου Intersection Over Union (IOU).

- Intersection Over Union (δείκτης Jaccard):
  Πρόκειται για δείκτη που χρησιμοποιείται για την αξιολόγηση μοντέλων ανίχνευσης αντικειμένων. Ο δείκτης υπολογίζεται ως το κλάσμα που έχει αριθμητή την τομή και παρονομαστή την ένωση των πλαισίων της πρόβλεψης του μοντέλου με τα αληθή πλαίσια (Ground truth). Όσο μεγαλύτερη είναι η τιμή του δείκτη τόσο πιο κοντά στα αληθή δεδομένα είναι τα πλαίσια της πρόβλεψης του μοντέλου.



Σχήμα 7: Δείκτης IoU για πλαίσια (Πηγή: (learnopencv, 2023)

Στην περίπτωση της εννοιολογικής κατάτμησης τα μοντέλα εξάγουν μάσκες, δηλαδή πολύγωνα εντός των οποίων προβλέπεται ότι υπάρχει ένα αντικείμενο που ανήκει σε μία κλάση. Ο δείκτης Jaccard για την υπό εκτίμηση μάσκα υπολογίζεται αντίστοιχα ως η τομή προς την ένωση τους της υπό εκτίμηση μάσκας με την αληθή μάσκα (Ground truth).

Σχήμα 8: Σχηματική αναπαράσταση των συντελεστών του μέτρου IoU για εννοιολογική κατάτμηση (Πηγή: (learnopencv, 2023))

Οι τιμές του μέτρου IoU είναι στο εύρος [0,1]. Όσο πιο μεγάλη είναι η τιμή τόσο περισσότερο ταυτίζεται η υπό εκτίμηση μάσκα με την αληθή μάσκα.
Ορίζοντας ένα κατώφλι για μέτρο IOU μπορούμε να κατατάξουμε την υπό εκτίμηση μάσκα που εξάγει ένα μοντέλο εννοιολογικής κατάτμησης σε:.

- Αληθώς θετική (True Positive): Όταν η τιμή του μέτρου IoU είναι μεγαλύτερη ή ίση με το κατώφλι και το αντικείμενο ανήκει στη κλάση που προβλέπει το μοντέλο
- Ψευδώς θετική (False Positive): Όταν η τιμή του μέτρου IoU είναι μεγαλύτερη ή ίση με το κατώφλι και το αντικείμενο ανήκει στη κλάση που προβλέπει το μοντέλο .
- Ψευδώς αρνητική (False Negative): Όταν το μοντέλο προβλέπει ότι το αντικείμενο δεν ανήκει σε μία κλάση και το αντικείμενο όντως ανήκει σε αυτή την κλάση
- Αληθώς αρνητική (True Negative): Όταν το μοντέλο προβλέπει ότι το αντικείμενο δεν ανήκει σε μία κλάση και το αντικείμενο όντως δεν ανήκει σε αυτή την κλάση

- Μέση Ακρίβεια mAP: Η Ακρίβεια (Precision) είναι το μέτρο απόδοσης που ορίζεται ως το σύνολο των ορθά ανιχνευμένων μασκών προς το σύνολο των ανιχνευμένων μασκών. Με την χρήση μέτρου IoU υπολογίζεται αν η μάσκα είναι αληθώς θετική, ψευδώς θετική, ψευδώς αρνητική ή αληθώς αρνητική.

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives}$$

Στην συνέχεια για το σύνολο των κλάσεων K υπολογίζεται το μέτρο της μέσης Ακρίβειας mAP ως:

$$mAP = \frac{1}{K} \cdot \sum_{i=1}^{K} AP_i$$

- Μέση Επανάκληση mAR: Η Επανάκληση (Recall) είναι το μέτρο απόδοσης που ορίζεται ως το σύνολο των ορθά ανιχνευμένων μασκών προς το σύνολο των αληθών μασκών. Με την χρήση μέτρου IoU υπολογίζεται αν η μάσκα είναι αληθώς θετική, ψευδώς θετική, ψευδώς αρνητική ή αληθώς αρνητική.

$$Recall = \frac{TruePositives}{TruePositives + FalseNegatives}$$

Στην συνέχεια για το σύνολο των κλάσεων K υπολογίζεται το μέτρο της μέσης επανάκλησης mAR ως:

$$mAR = \frac{1}{K} \cdot \sum_{i=1}^{K} AR_i$$

2.4 Χρήσιμες έννοιες

- Support Vector machine (SVM): Πρόκειται για τύπο τεχνητού νευρωνικού δικτύου που χρησιμοποιείται για την ταξινόμηση δεδομένων. Τα SVM προσπαθούν να βρουν ένα υπερεπίπεδο που διαχωρίζει τα δεδομένα.

- Augmentation: Το augmentation είναι τεχνική παραγωγής συνθετικών δεδομένων με σκοπό να αποφευχθεί η υπερπροσαρμογή και να επιτευχθεί η γενίκευση. Τα νέα δεδομένα αξιοποιούνται για την επέκταση του συνόλου των δεδομένων εκπαίδευσης με δεδομένα που προσομοιάζουν συνθήκες που πιθανώς δεν υπάρχουν στα αρχικά δεδομένα εκπαίδευσης. Τα συνθετικά δεδομένα παράγονται από τα δεδομένα εκπαίδευσης αφού πρώτα υποστούν μετασχηματισμούς όπως:
  - Αφινικούς μετασχηματισμούς που στρέφουν, μεταθέτουν και αλλάζουν κλίμακα
  - Φωτομετρικούς μετασχηματισμούς που αλλάζουν τη φωτεινότητα, την αντίθεση και τον κορεσμό στα δεδομένα
  - Περικοπή τμήματος δεδομένων

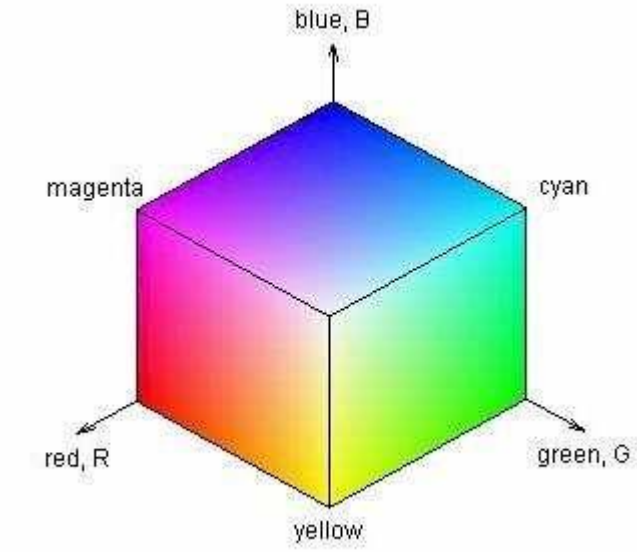
- Non Max Supression (NMS): Η τεχνική μη μέγιστης καταστολής (NMS) είναι ένα σύνολο από αλγορίθμους που χρησιμοποιείται στην προγραμματιστική ενότητα της όρασης υπολογιστών. Χρησιμοποιείται για την επιλογή μίας οντότητας (πλαίσιο ή μάσκα) από πολλές οντότητας που επικαλύπτονται.

Σχήμα 9: Αποτέλεσμα αλγορίθμου NMS όπου προκύπτει ένα τελικό πλαίσιο από πολλά επικαλυπτόμενα πλαίσια (Πηγή: (learnopencv, 2023))

- Χρωματικός χώρος RGB: Δεδομένου ότι το ανθρώπινο μάτι έχει κύτταρα ευαίσθητα στο μπλε, το πράσινο το κόκκινο τμήμα της ακτινοβολίας είναι θεωρητικά πιθανό να περιγραφή κάθε ορατό χρώμα ως συνδυασμός αυτών των τριών χρωμάτων. Ο χρωματικός χώρος RGB (Red, Green, Blue) αναλύει τα χρώματα ως ένα διάνυσμα που η πρώτη συνιστώσα είναι η ένταση του κόκκινου, η

δεύτερη του πράσινου και η τρίτη του μπλε.

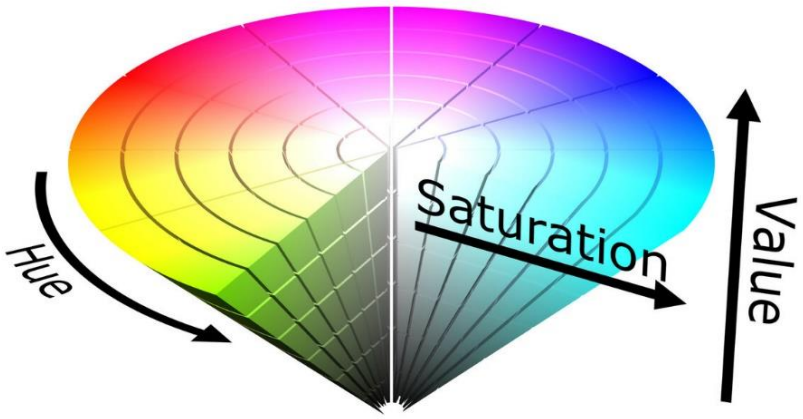
Πρόκειται για τον πιο διαδεδομένο χρωματικό χώρο και χρησιμοποιείται από στις περισσότερες οθόνες συσκευών που συναντώνται στην καθημερινότητα.



Σχήμα 10: Αναπαράσταση του χρωματικού χώρου RGB ως σύστημα τριών διαστάσεων όπου κάθε άξονας αντιστοιχεί σε μία συνιστώσα red, green και blue (Πηγή: (researchgate, 2023))

- Χρωματικός χώρος HSV: Ο χρωματικός χώρος HSV χρησιμοποιείται για να περιγράψει χρώματα χρησιμοποιώντας τις τρεις συνιστώσες :
    - της απόχρωσης (Hue)
    - του κορεσμού (Saturation).
    - της τιμής φωτεινότητας (Value)

Η απόχρωση αναφέρεται στο χρώμα και μπορεί να αναπαρασταθεί ως ένα κύκλος όπου οι τιμές των μοιρών αναφέρονται σε κάποια συγκεκριμένη απόχρωση. Ο κορεσμός αναφέρεται στο ποσοστό του γκρι χρώματος σε ένα χρώμα. Όσο πιο μικρή είναι η τιμή του κορεσμού τόσο πιο κοντά στο γκρι είναι η απόχρωση. Η τιμή της φωτεινότητας αναφέρεται περιγράφει το πόσο έντονο ή φωτεινό είναι ένα χρώμα. Όσο μειώνεται η τιμή της φωτεινότητας τόσο τείνει προς το μαύρο το χρώμα.



Σχήμα 11: Ο χρωματικός χώρος HSV (Πηγή: (wikimedia, 2023))

Ο χρωματικός χώρος HSV χρησιμοποιείται σε εφαρμογές ανίχνευσης αντικειμένων με βάση το χρώμα καθώς περιγράφει τα χρώματα με τρόπο παρόμοιο με τον τρόπο που τα αντιλαμβάνεται ο άνθρωπος. Επιπλέον ιδιότητα του συγκεκριμένου τρόπου περιγραφής των χρωμάτων είναι ότι απομονώνει την πληροφορία του χρώματος από την πληροφορία της φωτεινότητας. Στον χρωματικό χώρο HSV αν δύο αποχρώσεις διαφέρουν μόνο κατά την πληροφορία της φωτεινότητας τότε έχουν ίδια τιμή Hue και διαφορετική τιμή Value ενώ στον χρωματικό χώρο RGB οι δύο αυτές τιμές θα έχουν διαφορετικές και τις τρεις συνιστώσες Red, Green και Blue.

- Selective Search: Ο αλγόριθμος Selective Search χρησιμοποιείται για την εξαγωγή προτεινόμενων περιοχών από μία εικόνα. Οι προτάσεις περιοχών χρησιμοποιούνται στην ανίχνευση αντικειμένων για να ψάξει ο εκάστοτε αλγόριθμος εντός αυτών για το υπό αναζήτηση αντικείμενο. Ο αλγόριθμος Selective Search αρχικά χωρίζει την εικόνα σε πάρα πολλές υποπεριοχές με βάση την πυκνότητα των εικονοστοιχείων. Στην συνέχεια ενώνει τις όμοιες γειτονικές περιοχές δημιουργώντας έτσι μεγαλύτερες περιοχές. Η διαδικασία αυτή επαναλαμβάνεται έως ότου προκύψουν οι τελικές περιοχές από την διαρκή συνένωση όμοιων γειτονικών περιοχών. Για να θεωρηθούν δύο γειτονικές περιοχές όμοιες ο αλγόριθμος υπολογίζει δείκτες ομοιότητας με βάση:
    - το χρώμα
    - την υφή
    - το μέγεθος
    - το σχήμα

  Στην συνέχεια υπολογίζεται η τελική ομοιότητα των δύο περιοχών ως γραμμικός συνδυασμός των παραπάνω ομοιοτήτων και αν η τιμή της είναι μεγαλύτερη από ένα κατώφλι τότε οι περιοχές θεωρούνται όμοιες.

2.5 CNN

Τα συνελικτικά νευρωνικά δίκτυα (CNN) αποτελούν ειδική κατηγορία δικτύων μηχανικής μάθησης που εξειδικεύονται στην επίλυση θεμάτων ταξινόμησης που αφορούν εικόνες και άλλα δεδομένα δομής κανάβου. Βασικό πλεονέκτημα των CNN σε σχέση με ένα πλήρως συνδεδεμένο νευρωνικό δίκτυο είναι η ανεκτικότητα σε γεωμετρικούς μετασχηματισμούς των δεδομένων εισόδου και η μειωμένη έκταση του δικτύου συγκριτικά με ένα αντίστοιχο πλήρως συνδεδεμένο νευρωνικό δίκτυο που θα χρειαστεί περισσότερους κόμβους για να επεξεργαστεί τα ίδια δεδομένα εισόδου.

2.5.1 Λειτουργία και Αρχιτεκτονική

Τα συνελικτικά νευρωνικά δίκτυα αποτελούνται από δύο κυρίως τμήματα . Στο πρώτο τμήμα εξάγονται χαρακτηριστικά από τα δεδομένα εισόδου τα οποία αξιοποιούνται στο δεύτερο τμήμα για την εξαγωγή του τελικού συμπεράσματος.

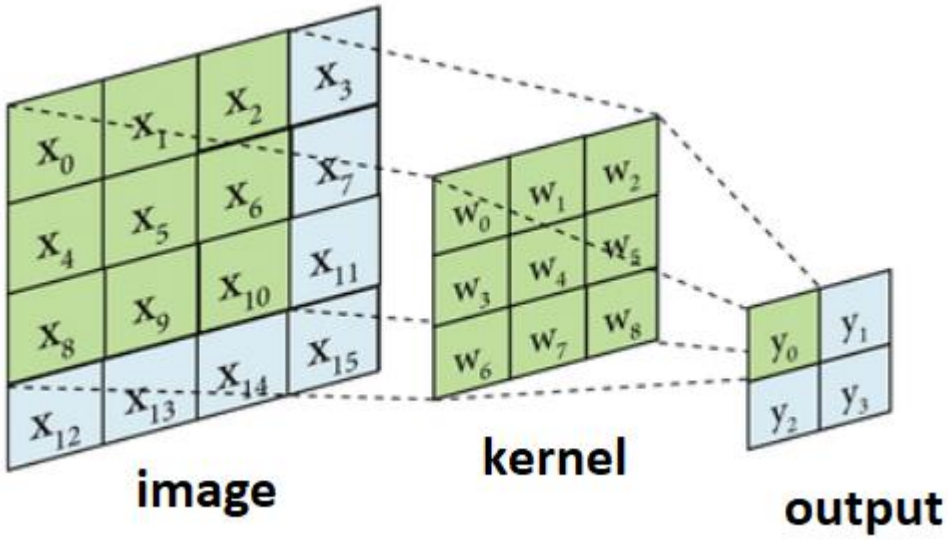
Η αρχιτεκτονική του εκάστοτε συνελικτικού νευρωνικού δικτύου διαφέρει ανάλογα τη φύση του εκάστοτε δικτύου και της ιδιομορφίας των δεδομένων εισόδου. Παρόλα αυτά η αρχή λειτουργίας τους παραμένει ίδια. Βασικοί τελεστές των συνελικτικών νευρωνικών δικτύων είναι τα:

- συνελικτικά επίπεδα (Convolutional Layers)
- τα επίπεδα μη-γραμμικοποίησης (Relu Layers)
- επίπεδα συνάθροισις / χωρικής υποδειγματολειψίας (Pooling Layers)

Η πράξη της συνέλιξης εκφράζει έναν πολλαπλασιασμό πινάκων. Ένας πίνακας (kernel) που έχει μέγεθος μικρότερο από το μέγεθος της εικόνας πολλαπλασιάζεται κάθε φορά με ένα υποσύνολο της εικόνας ίδιου μεγέθους με τον kernel. Η διαδικασία ξεκινάει από το πάνω αριστερά μέρος της εικόνας και ο kernel διατρέχει την εικόνα μετακινούμενος κάθε φορά κατά μία ποσότητα που ονομάζεται stride.

$$y(m,n) = x(m,n) * h(m,n) = \sum_{j=-\infty}^{\infty} \sum_{i=-\infty}^{\infty} x(i,j) \cdot h(m-i, n-j)$$

Όπου x(m,n): η υποπεριοχή της εικόνας και το μέγεθος του kernel
h(m,n):ο πίνακας της συνέλιξης ή φίλτρο



Σχήμα 12: Η πράξη της συνέλιξης (Πηγή: (medium, 2021))

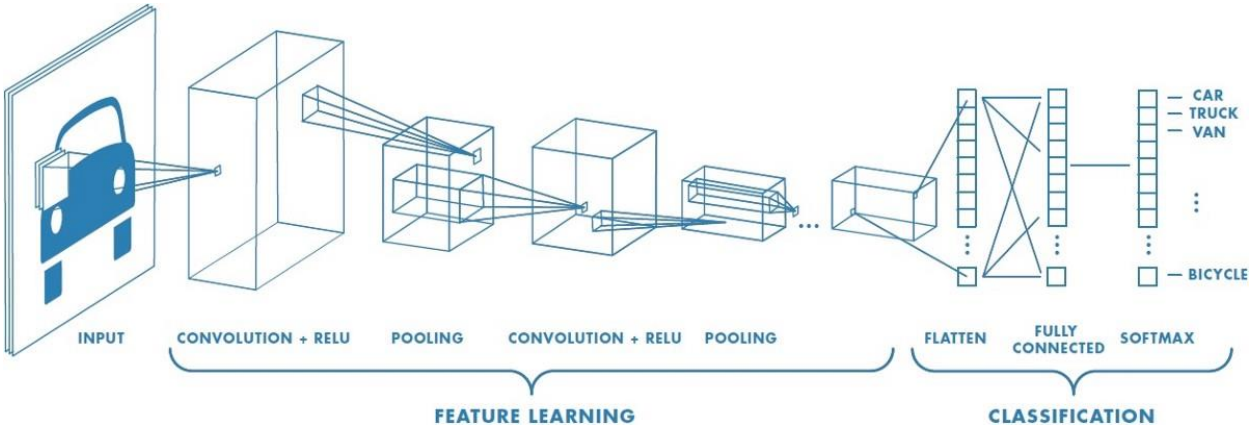
Ένα συνελικτικό επίπεδο που αποτελείται από ένα σύνολο νευρώνων οι οποίοι έχουν τα ίδια σύνολα βαρών και άρα εξάγουν τα ίδια χαρακτηριστικά από όλες τις υποπεριοχές των δεδομένων εισόδου . Έτσι η έξοδος του κάθε νευρώνα για την υποπεριοχή που εφαρμόζεται είναι αντίστοιχη με το αποτέλεσμα της συνέλιξης με ένα πυρήνα, τα στοιχεία του οποίου θα είναι τα σύνολα βαρών του νευρώνα, που εφαρμόζεται στην ίδια υποπεριοχή της εικόνας. Η πράξη της συνέλιξης δίνει στο δίκτυο την ανεκτικότητα σε μετασχηματισμούς στροφής των δεδομένων εισόδου. Το σύνολο των εξόδων των νευρώνων ενός συνελικτικού επιπέδου ονομάζεται χάρτης χαρακτηριστικών και περιέχει χαρακτηριστικά τα οποία εξήχθησαν από τα δεδομένα εισόδου με την συνέλιξη του εκάστοτε πυρήνα.

Μετά τα συνελικτικά επίπεδα ακολουθούν τα επίπεδα μη γραμμικοποίησης (Relu). Σε αυτά τα επίπεδα εφαρμόζεται η συνάρτηση ενεργοποίησης ReLU η οποία μηδενίζει την τιμή του χάρτη χαρακτηριστικών αν αυτή είναι αρνητική. Τα επίπεδα αυτά προσδίδουν μία μη γραμμικότητα στο νευρωνικό δίκτυο που είναι επιθυμητό για να παραχθούν μη γραμμικά όρια μεταξύ των χαρακτηριστικών.

Τα επίπεδα συνάθροισης / χωρικής υποδειγματοληψίας (Pooling) δέχονται σαν δεδομένο εισόδου ένα χάρτη χαρακτηριστικών και εφαρμόζοντας ένα πλέγμα σταθερών διαστάσεων στην εικόνα εξάγουν την μέγιστη τιμή των εικονοστοιχείων εντός του πλέγματος. Σκοπός αυτών των επιπέδων είναι να μειώσουν σταδιακά το μέγεθος των χαρτών χαρακτηριστικών .

Στο πρώτο τμήμα ενός συνελικτικού δικτύου τα δεδομένα εισόδου διέρχονται από μια αλληλουχία επιπέδων συνέλιξης, ReLU και χωρικής υποδειγματοληψίας μέχρι οι παράγωγοι χάρτες χαρακτηριστικών να έχουν κατάλληλο μέγεθος, αρκετά μικρότερο από τα δεδομένα εισόδου. Στην συνέχεια οι τελικοί χάρτες χαρακτηριστικών μετατρέπονται σε διάνυσμα μίας διάστασης.

Στο δεύτερο τμήμα τα στοιχεία του διανύσματος μίας διάστασης εισέρχονται σε ένα πλήρως συνδεδεμένο νευρωνικό δίκτυο το οποίο ταξινομεί τελικά την εικόνα.



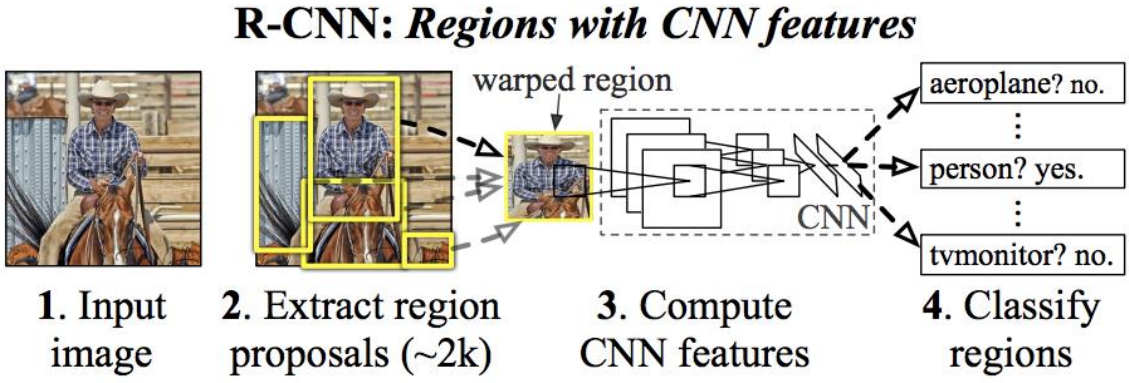
Σχήμα 13: Αρχιτεκτονική ενός CNN. Στο πρώτο μέρος του δικτύου γίνεται η εξαγωγή των χαρακτηριστικών (Feature Learning) και στο δεύτερο μέρος η εξαγωγή της κλάσης που ανήκει το αντικείμενο (Classification) (Πηγή: (medium, 2023))

2.6 RCNN, Fast-RCNN, Faster-RCNN, Mask RCNN

Τα CNN δίκτυα αναγνωρίζουν ένα κυρίαρχο αντικείμενο στην εικόνα. Για την επίλυση του προβλήματος της ανίχνευσης πολλαπλών αντικειμένων που ενδεχομένως να ανήκουν σε διαφορετικές κλάσεις δημιουργήθηκαν μεταγενέστερα δίκτυα που βασίζονται στην λειτουργία των CNN.

2.6.1 RCNN

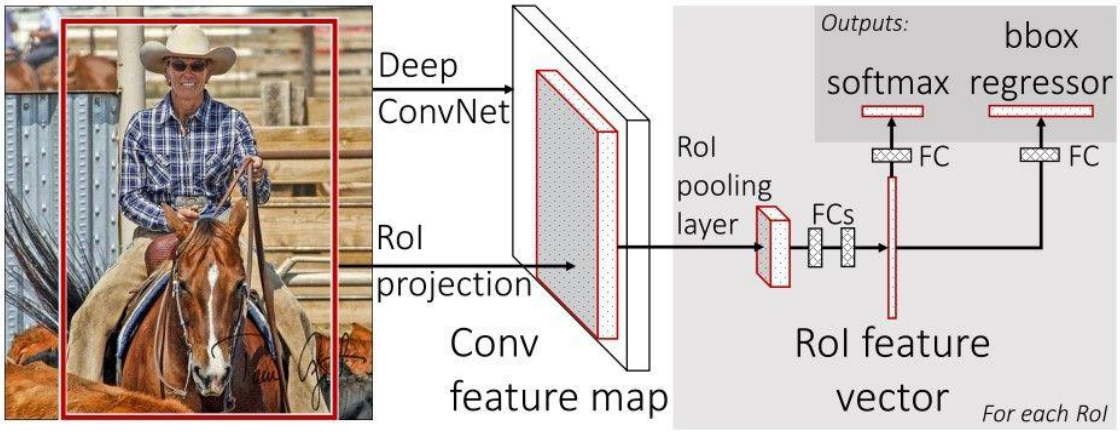
Μία πρώτη προσέγγιση είναι τα δίκτυα βαθιάς μηχανικής μάθησης για ανίχνευση αντικειμένων Region-CNN (R-CNN) (Girshick, et al., 2014). Αρχικά από την εικόνα εισόδου εξάγονται 2000 προτεινόμενες περιοχές με την χρήση του αλγορίθμου Selective Search (βλ. 2.4) που αξιοποιεί ιδιότητες όπως είναι το χρώμα, η υφή και το μέγεθος για να εξάγει τις προτεινόμενες περιοχές. Κάθε μία από τις προτεινόμενες περιοχές που μπορεί να περιέχει αντικείμενο μετατρέπεται σε μία εικόνα συγκεκριμένων διαστάσεων και επεξεργάζεται από ένα συνελικτικό νευρωνικό δίκτυο για την παραγωγή χάρτη χαρακτηριστικών. Τέλος οι χάρτες χαρακτηριστικών εισέρχονται σαν δεδομένα εισόδου σε ένα SVM για την ταξινόμηση του αντικειμένου και σε ένα δίκτυο παλινδρόμησης για την εξαγωγή του πλαισίου οριοθέτησης. Το R-CNN μοντέλο ανιχνεύει πολλαπλά αντικείμενα σε μία εικόνα σε σχέση με το απλό CNN. Παρόλα αυτά το συγκεκριμένο μοντέλο είναι αρκετά αργό και απαιτητικό σε υπολογιστικούς πόρους λόγο. Αυτό οφείλεται στο γεγονός ότι καθεμία από της 2000 προτεινόμενες περιοχές πρέπει να επεξεργαστεί από το CNN για την ανίχνευση σε μία εικόνα μόνο.



Σχήμα 14: Το μοντέλο RCNN δέχεται σαν δεδομένο εισόδου μία εικόνα από την οποία εξάγονται προτεινόμενες περιοχές οι οποίες μορφοποιούνται σε συγκεκριμένο μέγεθος και εισέρχονται σαν δεδομένο εισόδου στο CNN για την ταξινόμηση τους. (Πηγή: (Girshick, et al., 2014))

2.6.2 Fast- RCNN

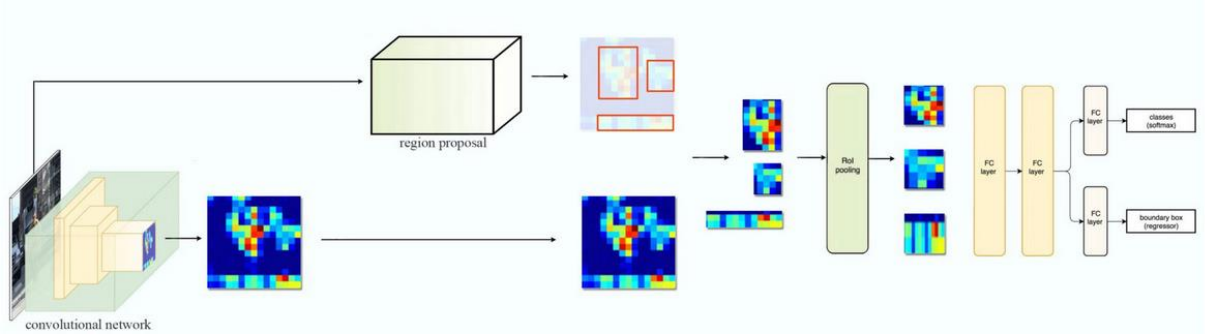
Το μοντέλο Fast-RCNN (Girshick, 2015) χρησιμοποιείται επίσης για την ανίχνευση αντικειμένων βελτιώνοντας μερικά από τα μειονεκτήματα του μοντέλου R-CNN. Το μοντέλο Fast-RCNN δέχεται σαν δεδομένα εισόδου την εικόνα και κάποιες προτεινόμενες περιοχές που μπορεί να περιέχουν αντικείμενο. Οι προτεινόμενες περιοχές προέρχονται από τον αλγόριθμο Selective Search (βλ. 2.4). Αρχικά τροφοδοτεί την εικόνα σε ένα CNN για την παραγωγή ενός ενιαίου χάρτη χαρακτηριστικών αντί να τροφοδοτεί 2000 προτάσεις περιοχών σε ένα CNN. Στην συνέχεια μέσω ενός επιπέδου χωρικής υποδειγματοληψίας (ROI pooling) από κάθε προτεινόμενη περιοχή εξάγεται ένα διάνυσμα χαρακτηριστικών συγκεκριμένου μεγέθους από τον ενιαίο χάρτη χαρακτηριστικών. Έπειτα το διάνυσμα αυτό εισέρχεται σαν δεδομένο εισόδου σε ένα πλήρως συνδεδεμένο νευρωνικό δίκτυο με δύο διακλαδώσεις. Μια διακλάδωση softmax που υπολογίζει την πιθανότητα του αντικειμένου να ανήκει σε κάποια προκαθορισμένη κλάση και μία διακλάδωση που μέσω παλινδρόμησης υπολογίζει τις συντεταγμένες του πλαισίου που εμπεριέχει το αντικείμενο. Βασικό πλεονέκτημα του μοντέλου Fast RCNN είναι πιο γρήγορο και λιγότερο απαιτητικό σε υπολογιστικούς πόρους από το RCNN καθώς η εικόνα εισέρχεται μία φορά από το CNN αντί για 2000 ξεχωριστά τμήματα της εικόνας να διέρχονται από το CNN. Η εξαγωγή προτεινόμενων περιοχών είναι ένα τμήμα της διαδικασίας που χρειάζεται βελτίωση καθώς ο αλγόριθμος Selective Search αποτελεί μια χρονοβόρα διαδικασία.



Σχήμα 15: Το μοντέλο Fast-RCNN (Πηγή: (Girshick, 2015))

2.6.3 Faster- RCNN

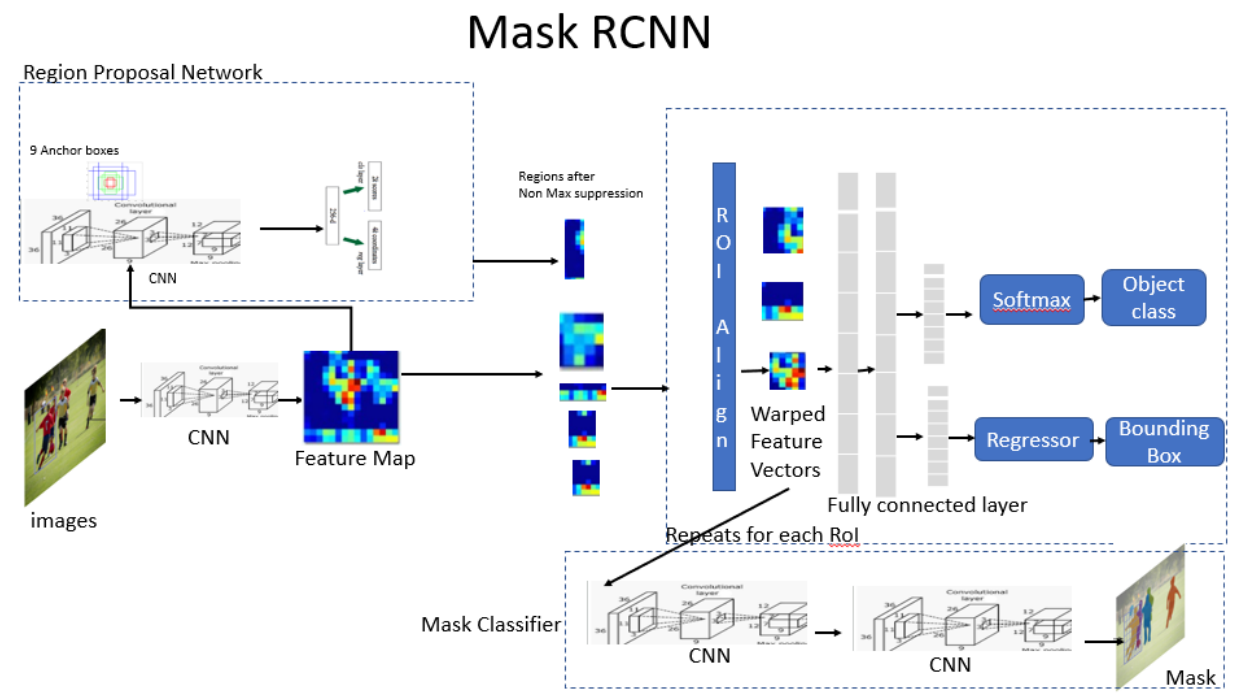
Το μοντέλο Faster-RCNN (Ren, et al., 2017) αποτελεί μια βελτιωμένη έκδοση του μοντέλου Fast-RCNN που επιλύει το θέμα της αργής εξαγωγής προτεινόμενων περιοχών. Αρχικά η εικόνα διέρχεται από ένα συνελικτικό δίκτυο και παράγεται ένας χάρτης χαρακτηριστικών. Το συγκεκριμένο CNN ονομάζεται δίκτυο κορμού (backbone) και είναι ο κύριος τρόπος εξαγωγής χαρακτηριστικών από την εικόνα εισόδου. Επιπλέον η εικόνα διέρχεται από ένα συνελικτικό δίκτυο παραγωγής προτεινόμενων περιοχών (RPN) που χρησιμοποιεί τα ίδια συνελικτικά επίπεδα με το δίκτυο κορμού ώστε η διαδικασία αυτή να γίνεται παράλληλα με την διαδικασία παραγωγής του χάρτη χαρακτηριστικών. Στο RPN ένας πυρήνας σταθερών διαστάσεων διατρέχει την εικόνα κατά την διαδικασία της συνέλιξης και εφαρμόζονται στο κέντρο του πυρήνα διαφορετικά πλαίσια. Τα πλαίσια αυτά ποικίλουν σε κλίμακες και λόγους διαστάσεων ώστε να μπορούν να ανιχνευθούν αντικείμενα διαφορετικών μεγεθών και σχημάτων. Τα πλαίσια αυτά διέρχονται από ένα πλήρως συνδεδεμένο δίκτυο με δύο κλάδους. Ο ένας κλάδος υπολογίζει την πιθανότητα να περιέχει αντικείμενο το πλαίσιο και ο άλλος κλάδος υπολογίζει τις συντεταγμένες του πλαισίου που εμπεριέχει το αντικείμενο. Οι περιοχές με την μεγαλύτερη πιθανότητα να περιέχουν αντικείμενο είναι οι προτεινόμενες περιοχές του δικτύου RPN. Στην συνέχεια ακολουθείτε η ίδια διαδικασία με το μοντέλο Fast-RCNN με τη μόνη διαφορά ότι αξιοποιούνται οι προτεινόμενες περιοχές του δικτύου RPN αντί για να παραχθούν με τον αλγόριθμο Selective Search. Η αλλαγή αυτή στον τρόπο παραγωγής προτεινόμενων περιοχών κάνει το μοντέλο Faster-RCNN σημαντικά πιο γρήγορο από το μοντέλο Fast-RCNN.



Σχήμα 16: Το μοντέλο Faster-RCNN (Πηγή: (Ren, et al., 2017))

2.6.4 Mask-RCNN

Το μοντέλο Mask-RCNN (HE., et al., 2017) αποτελεί ενισχυμένη έκδοση του μοντέλου Faster RCNN καθώς πέρα από την ανίχνευση αντικειμένων πραγματοποιεί και εννοιολογική κατάτμηση (Instance segmentation). Πρόκειται για αλγόριθμο ανίχνευσης και εννοιολογικής κατάτμησης δύο σταδίων. Αρχικά στο πρώτο στάδιο το μοντέλο όπως και το προγενέστερο Faster-RCNN παράγει ένα χάρτη χαρακτηριστικών από την εικόνα εισόδου μέσο ενός δικτύου κορμού και παράλληλα το συνελικτικό δίκτυο RPN εξάγει προτεινόμενες περιοχές. Έπειτα στο δεύτερο στάδιο χρησιμοποιείται ένα δίκτυο ευθυγραμμισμένης χωρικής υποδειγματοληψίας (ROI align) για την εξαγωγή χαρακτηριστικών που βρίσκονται εντός των προτεινόμενων περιοχών από τον ενιαίο χάρτη χαρακτηριστικών. Η χρήση του δικτύου ευθυγραμμισμένης χωρικής υποδειγματοληψίας (ROI align) διατηρεί περισσότερη πληροφορία από το τμήμα του ενιαίου χάρτη χαρακτηριστικών που βρίσκεται εντός της κάθε προτεινόμενης περιοχής σε σχέση με το αντίστοιχο δίκτυο χωρικής υποδειγματοληψίας (ROI pool) που χρησιμοποιούν τα μοντέλα Fast-RCNN και Faster-RCNN. Στην συνέχεια τα χαρακτηριστικά αυτά εισέρχονται σαν δεδομένα εισόδου σε ένα πλήρως συνδεδεμένο δίκτυο δύο κλάδων για την ταξινόμηση και την εύρεση του πλαισίου που εμπεριέχει το αντικείμενο όπως και στο μοντέλο Fast-RCNN. Παράλληλα τα χαρακτηριστικά αυτά εισέρχονται σε ένα συνελικτικό δίκτυο που εξάγει την μάσκα του αντικειμένου. Έτσι το τελικό αποτέλεσμα του μοντέλου για κάθε αντικείμενο που ανιχνεύεται είναι το πλαίσιο που το εμπεριέχει, η μάσκα του και η κλάση στην οποία ανήκει.



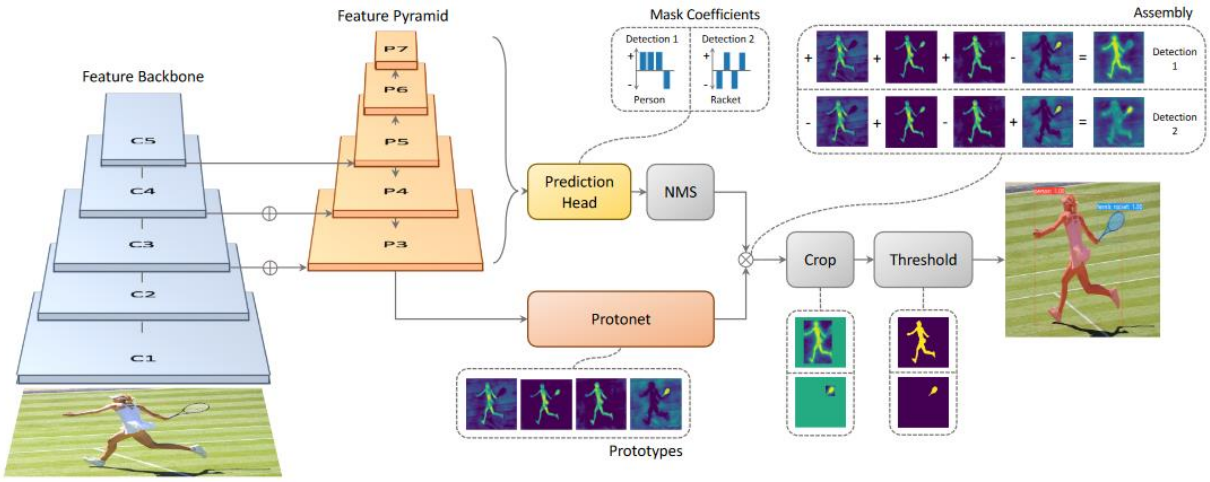
Σχήμα 17: Το μοντέλο Mask-RCNN (Πηγή: (HE., et al., 2017))

2.7 Yolact

Το μοντέλο Yolact (You Only Look at CoefficienTs) (Bolya, et al., 2019) προέρχεται από τα μοντέλα ανίχνευσης αντικειμένων της οικογένειας YOLO και πραγματοποιεί εννοιολογική κατάτμηση πέρα από την ανίχνευση αντικειμένων. Πρόκειται για μοντέλο ανίχνευσης και εννοιολογικής κατάτμησης ενός σταδίου. Αρχικά η εικόνα εισέρχεται σε ένα συνελικτικό δίκτυο κορμού (backbone) για την παραγωγή χαρτών χαρακτηριστικών οι οποίοι ταυτόχρονα επεξεργάζονται από ένα δίκτυο FPN για εξαγωγή χαρακτηριστικών σε διάφορες χωρικές κλίμακες για να είναι εφικτή η καλύτερη ανίχνευση αντικειμένων διαφόρων μεγεθών.

Έπειτα η ροή των χαρτών χαρακτηριστικών στο δίκτυο σπάει σε δύο κλάδους που λειτουργούν παράλληλα. Ο πρώτος κλάδος αποτελείται από ένα πλήρως συνελικτικό δίκτυο που δέχεται σαν δεδομένα εισόδου τους χάρτες χαρακτηριστικών και εξάγει ένα αριθμό K πρότυπες μάσκες. Ο δεύτερος κλάδος διακλαδίζεται σε τρεις επιμέρους κλάδους όπου υπολογίζεται για κάθε χάρτη χαρακτηριστικών το πλαίσιο που περιέχει το αντικείμενο, η κλάση στην οποία ανήκει το αντικείμενο και K συντελεστές συμμεταβλητότητας όσες και οι αντίστοιχες K πρότυπες μάσκες. Τα πλαίσια που στην συνέχεια διέρχονται από έναν αλγόριθμο μη μέγιστης καταστολής (NMS) και προκύπτουν οι τελικές προβλέψεις του δεύτερου κλάδου.

Στην συνέχεια για την παραγωγή των μασκών για γίνεται περικοπή με βάση το πλαίσιο και γραμμικός συνδυασμός των πρότυπων μασκών του πρώτου κλάδου με συντελεστές συμμεταβλητότητας αυτούς που υπολογίστηκαν στον δεύτερο κλάδο. Έτσι οι πρότυπες μάσκες που περιέχουν το αντικείμενο θα έχουν μεγαλύτερη συμμετοχή στην παραγωγή της τελικής μάσκας.



Σχήμα 18: Το μοντέλο Yolact (Πηγή: (Bolya, et al., 2019))

Κεφάλαιο 3: Σχεδίαση, μεθοδολογία και υλοποίηση

Για την επίλυση του προβλήματος της ανίχνευσης και παρακολούθησης της αρπάγης σε πραγματικό χρόνο αξιοποιήθηκαν τρεις αλγόριθμοι:

- Ο πρώτος αλγόριθμος κάνει χρήση της ιδιότητας του χρώματος της αρπάγης για την ανίχνευση της
- Ο δεύτερος αλγόριθμος κάνει χρήση του νευρωνικού δικτύου δύο σταδίων Mask Rcnn
- Ο τρίτος αλγόριθμος κάνει χρήση του νευρωνικού δικτύου ενός σταδίου Yolact

Χρονολογικά ο πρώτος αλγόριθμος αποτελεί την πρώτη προσέγγιση για την επίλυση του προβλήματος της ανίχνευσης. Ωστόσο η αξιολόγηση (βλ. 4.1) έδειξε ότι ο αλγόριθμος δεν καλύπτει τις προδιαγραφές που ορίστηκαν. Έτσι αποφασίστηκε στις επόμενες προσεγγίσεις να δοκιμαστούν δίκτυα ανίχνευσης αντικειμένων και εννοιολογικής κατάτμησης ενός και δύο σταδίων.

3.1 Αλγόριθμος Color Track

Μια πρώτη προσέγγιση στην επίλυση του προβλήματος του εντοπισμού και παρακολούθησης της αρπάγης ήταν η ανάπτυξη του αλγορίθμου Color Track. Ο παραπάνω αλγόριθμος κάνει χρήση βασικών τεχνικών όρασης υπολογιστών που αξιοποιούν την ιδιότητα του χρώματος για την εξαγωγή μίας μάσκας και ενός πλαισίου που περιέχουν την αρπάγη. Η ανάπτυξη του συγκεκριμένου αλγορίθμου έγινε διότι η ανίχνευση με τις τεχνικές αυτές έχει χαμηλές απαιτήσεις σε υπολογιστικούς πόρους και συνεπώς ο αλγόριθμος θα μπορεί να τρέχει σε πραγματικό χρόνο χωρίς ιδιαίτερες απαιτήσεις σε hardware. Επιπλέον η υλοποίηση ενός τέτοιου αλγορίθμου είναι αρκετά πιο απλή από την εκπαίδευση και προσαρμογή ενός νευρωνικού δικτύου βαθιάς μάθησης.

3.1.1 Εύρεση τιμών HSV

Για την ανίχνευση της αρπάγης με βάση το χρώμα χρειάστηκε αρχικά να βρεθούν αρχικές τιμές στον χρωματικό χώρο HSV της αρπάγης. Για να επιτευχθεί αυτό δημιουργήθηκε αλγόριθμος που δέχεται σαν δεδομένο εισόδου μία σειρά βίντεο και εκτελώντας μία επαναληπτική διαδικασία κάνει τις εξής διαδικασίες για κάθε καρέ:

- Εμφανίζει το αρχικό καρέ στον χρωματικό χώρο RGB
- Μετατρέπει το αρχικό καρέ στον χρωματικό χώρο HSV
- Εμφανίζει 6 συρόμενες μπάρες όπου ο χρήστης μπορεί να ρυθμίσει χειροκίνητα τις τιμές για το κάτω και άνω όριο των τιμών Hue, Saturation και Value
- Δημιουργείτε μάσκα, δηλαδή δυαδική εικόνα όπου την τιμή 1 παίρνουν όλα τα εικονοστοιχεία του αρχικού καρέ που βρίσκονται εντός των δύο ορίων HSV που ορίστηκαν από τις συρόμενες μπάρες
- Εμφανίζεται η μάσκα

Ο αλγόριθμος αυτός είναι δυναμικός καθώς με τις μεταβολές των τιμών HSV με τις συρόμενες μπάρες αλλάζει και η μάσκα που παράγεται. Αφού έγιναν πειραματισμοί με διαρκείς μεταβολές των μπαρών και συνεπώς τροποποίησης του εύρους HSV βρέθηκαν οι τιμές των μπαρών για τις οποίες εντός της παραγόμενης μάσκας εμφανίζεται το αντικείμενο της αρπάγης όσο το δυνατόν πιο απομονωμένο από το περιβάλλον για τα περισσότερα καρέ.

Πίνακας 1:Όρια HSV

	Hue	Saturation	Value
Κάτω όριο	14	50	75
Πάνω όριο	20	95	255

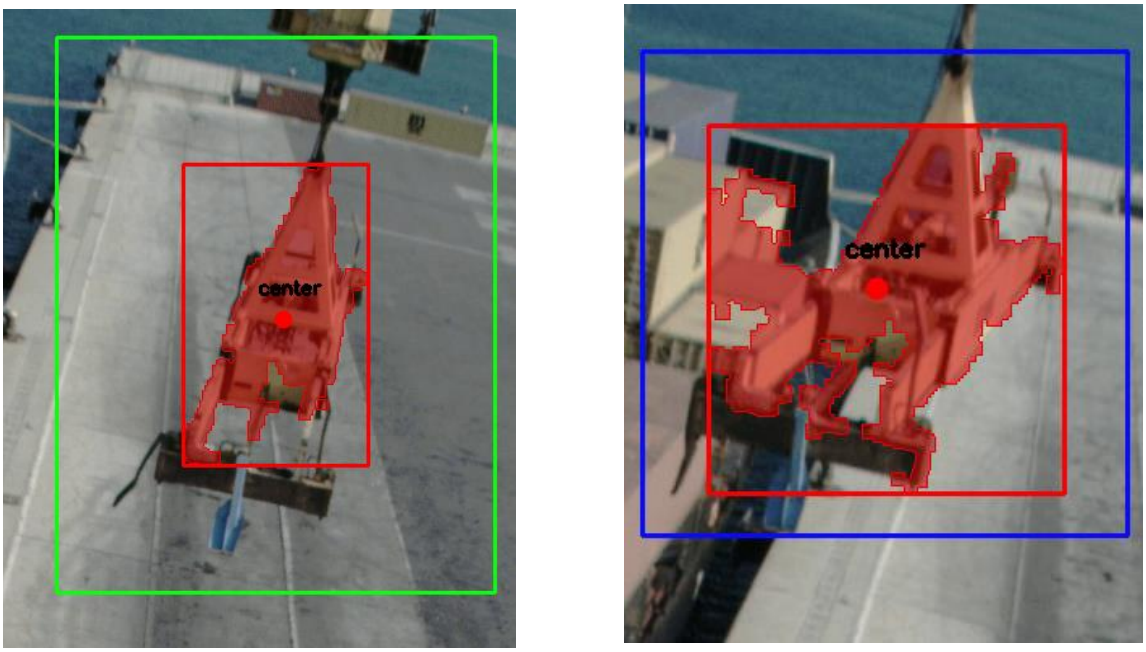
Από τις τρείς ιδιότητες του χρωματικού χώρου HSV οι ιδιότητες του Hue και του Saturation είναι πιο καθοριστικές στην απομόνωση της αρπάγης . Η ιδιότητα του Value είναι λιγότερο καθοριστική και για αυτό έχει και μεγαλύτερο εύρος από τις άλλες. Συγκεκριμένα το εύρος της τιμής Value έχει περιοριστεί στις πολύ χαμηλές τιμές διότι όσο μειώνεται η τιμή τόσο προσεγγίζουν το μαύρο οι αποχρώσεις και αρχίζουν να εισέρχονται σκουρόχρωμα αντικείμενα εντός της μάσκας όπως είναι αντικείμενα υπό σκιά.

3.1.2 Λειτουργία Color Track

Η αρχικοποίηση του αλγορίθμου γίνεται χειροκίνητα. Στο πρώτο καρέ ο χρήστης δίνει ένα πλαίσιο που περιέχει την αρπάγη. Στην συνέχεια ο αλγόριθμος μέσα σε αυτό το πλαίσιο, που αποτελεί υποπεριοχή της εικόνας, εντοπίζει την μάσκα και το πλαίσιο που περιβάλει την αρπάγη ακολουθώντας την διαδικασία εντοπισμού που περιγράφεται παρακάτω.

Έπειτα για τα επόμενα 20 καρέ ο αλγόριθμος διευρύνει το πλαίσιο που βρέθηκε η αρπάγη στο προηγούμενο καρέ κατά 100 εικονοστοιχεία . Το διευρυμένο αυτό πλαίσιο χρησιμοποιείται ως υποπεριοχή της εικόνας (Σχήμα 19a) για να ψάξει ξανά την αρπάγη στο νέο καρέ. Εδώ γίνεται η υπόθεση ότι η αρπάγη δεν θα μετακινηθεί σε μεγάλη απόσταση από με το πέρασμα ενός καρέ. Ταυτόχρονα υπολογίζεται η τιμή ενός δισδιάστατου φίλτρου Kalman για τις τιμές κέντρου της μάσκας που περιέχει την αρπάγη. Στον συγκεκριμένο αλγόριθμο χρησιμοποιήθηκε υλοποίηση δισδιάστατου φίλτρου Kalman (RahmadSadli) που λειτουργεί όπως αυτό που περιγράφεται στην θεωρία (βλ. 2.2).

Για τα υπόλοιπα καρέ το πλαίσιο που βρέθηκε στο προηγούμενο καρέ διευρύνεται κατά 100 εικονοστοιχεία και μετατοπίζεται κατά απόσταση d. Το διευρυμένο και μετατοπισμένο αυτό πλαίσιο αποτελεί την υποπεριοχή που θα ψάξει ο αλγόριθμος την αρπάγη στο επόμενο καρέ (Σχήμα 19b). Η απόσταση d είναι η απόσταση μεταξύ του κέντρου της μάσκας του προηγούμενου καρέ και της εκτίμησης του κέντρου της μάσκας από το φίλτρο Kalman. Η πρόβλεψη του φίλτρου Kalman δεν χρησιμοποιείται για την ανίχνευση στις πρώτες 20 επαναλήψεις διότι οι πρώτες προβλέψεις του φίλτρου συνήθως δεν είναι οι αναμενόμενες . Για αυτό τον λόγο ο αλγόριθμος αναζητά την αρπάγη σε μία κοντινή περιοχή από αυτή που την είχε εντοπίσει στο προηγούμενο καρέ για τις πρώτες 20 επαναλήψεις μέχρι να βελτιωθούν οι προβλέψεις του φίλτρου Kalman.



(a) *(b)*

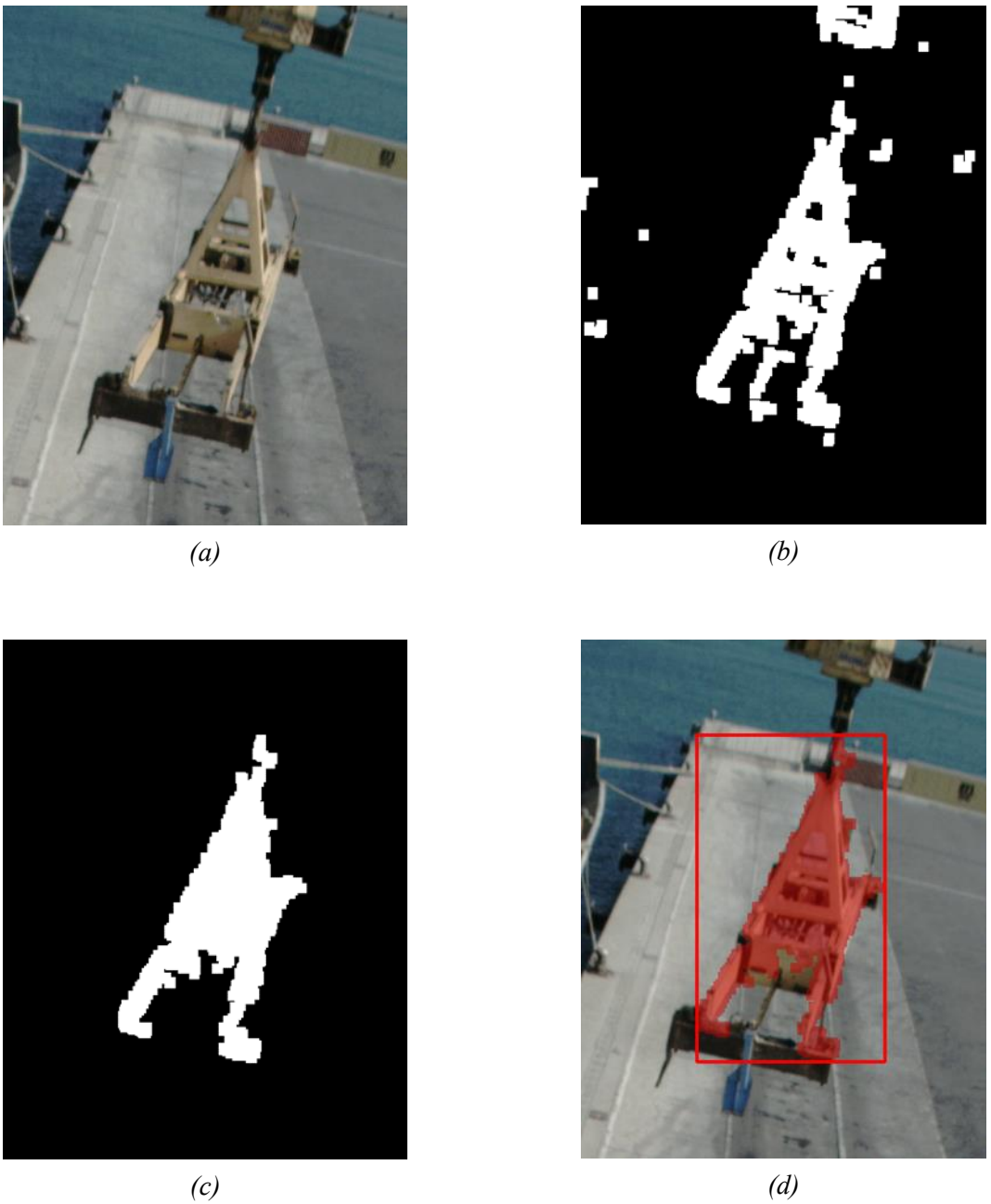
Σχήμα 19: Περιοχή αναζήτησης για ένα από τα πρώτα 20 καρέ με πράσινο χρώμα (a) και περιοχή αναζήτησης για ένα από τα επόμενα καρέ με μπλε χρώμα (b)

Η διαδικασία του εντοπισμού αντικειμένων γίνεται με τα ακόλουθα βήματα:

1) Αρχικά η περιοχή ανίχνευσης (τμήμα εικόνας) μετατρέπεται από τον χρωματικό χώρο RGB στον χρωματικό χώρο HSV
2) Στην συνέχεια δημιουργείτε μία μάσκα, δηλαδή μία δυαδική εικόνα όπου οι τιμές των εικονοστοιχείων που είναι εντός του εύρους ανίχνευσης στον χρωματικό χώρο HSV παίρνουν την τιμή 1 και οι υπόλοιπες τιμές την τιμή 0 (Σχήμα 20b).
3) Έπειτα γίνεται εξαγωγή των περιγραμμάτων (contours) των περιοχών εντός της μάσκας που έχουν τιμή 1.
4) Από τα περιγράμματα αυτά επιλέγεται το περίγραμμα που έχει το μεγαλύτερο εμβαδό. Εδώ γίνεται η παραδοχή ότι εντός της περιοχής ανίχνευσης το μεγαλύτερο αντικείμενο που θα είναι εντός του εύρους HSV θα είναι η αρπάγη
5) Στην συνέχεια δημιουργείται η νέα μάσκα όπου τα εικονοστοιχεία που βρίσκονται εντός του μεγαλύτερου περιγράμματος παίρνουν τιμή 1 και τα υπόλοιπα τιμή 0 (Σχήμα 20c).
6) Από την νέα μάσκα εξάγεται και το πλαίσιο που περιέχει το πολύγωνο της μάσκα αλλά και το κέντρο του πολυγώνου της μάσκας. Οι συντεταγμένες αυτού του κέντρου θα χρησιμοποιηθούν για την πρόβλεψη από το φίλτρο Kalman της μελλοντικής θέσης του κέντρου.
7) Τέλος το πλαίσιο, η μάσκα και το κέντρο προβάλλονται στο αρχικό καρέ (Σχήμα 20d)

Τα βήματα 3-5 γίνονται για αφαιρεθούν τα μικροαντικείμενα από την τελική μάσκα. Ουσιαστικά από την μάσκα που δημιουργήθηκε στο βήμα 2 παραμένει εντός μάσκας μόνο το μεγαλύτερο ενιαίο αντικείμενο. Η διαδικασία ανίχνευσης της αρπάγης σε υποπεριοχή της εικόνας γίνεται για να μειωθεί το υπολογιστικό κόστος. Επιπλέον με την υπόθεση ότι η αρπάγη σε κάθε νέο καρέ θα βρίσκεται κοντά στην προηγούμενη

της θέση αποφεύγεται η ανίχνευση σε ολόκληρο το καρέ. Η ανίχνευση σε ολόκληρο το καρέ εκτός από την αύξηση του υπολογιστικού κόστους θα αύξανε και την λανθασμένη ανίχνευση αντικειμένων καθώς σε μεγαλύτερη περιοχή ανίχνευσης υπάρχει μεγαλύτερη πιθανότητα να υπάρχουν και άλλα αντικείμενα εντός του εύρους HSV όπως είναι τα κοντέινερ. Επιπλέον ανάλογα με την θέση της αρπάγης προς το φακό της κάμερας θα μπορούσε κάποιο κοντέινερ εντός του εύρους HSV να εμφανίζεται μεγαλύτερο από την αρπάγη και επομένως το φιλτράρισμα των ανιχνευμένων περιοχών με βάση το μέγεθος, όπως γίνεται στα βήματα 3-5 δεν θα ήταν εφικτό.



(a) *(b)*

(c) *(d)*

Σχήμα 20: Διαδικασία ανίχνευσης όπου η αρχική περιοχή (a), η μάσκα των αντικειμένων που βρίσκονται εντός εύρους HSV (b), η τελική μάσκα με το μεγαλύτερο αντικείμενο (c), το τελικό αποτέλεσμα (d)

3.2 Software και Hardware

Για την ανάπτυξη και χρήση των αλγορίθμων χρησιμοποιήθηκε υπολογιστής με κεντρική μονάδα επεξεργασίας (CPU) Intel Core i9-11900K και κάρτα γραφικών (GPU) Nvidia RTX 3070 με 8GB VRAM και 5888 CUDA CORES. Για την ανίχνευση της αρπάγης χρησιμοποιήθηκε το μοντέλο Mask Rcnn που πραγματοποιεί εννοιολογική κατάτμηση της εικόνας. Αρχικά έγιναν δοκιμές με υλοποίηση του Mask Rcnn (Abdulla) που κάνει χρήση της βιβλιοθήκης Tensorflow. Τα αποτελέσματα ήταν ικανοποιητικά με περιθώρια βελτίωσης. Επιπλέον έγινε δοκιμή της υλοποίησης του Mask Rcnn που εμπεριέχεται στο API Detectron2 (Wu, et al.). Η συγκεκριμένη υλοποίηση κάνει χρήση της βιβλιοθήκης PyTorch που μειώνει αισθητά τους χρόνους εκπαίδευσης και ανίχνευσης σε σχέση με την προηγούμενη υλοποίηση. Έτσι αποφασίστηκε να χρησιμοποιηθεί η υλοποίηση του API Detectron2 για την δημιουργία της εφαρμογής. Η υλοποίηση του αλγορίθμου Yolact (Bolya, et al.) κάνει επίσης χρήση της βιβλιοθήκης PyTorch.

3.3 Δεδομένα

Τα αρχικά δεδομένα για την ανάπτυξη αλγορίθμων ήταν σειρές βίντεο από λιμένα στις οποίες φαίνεται η αρπάγη σε λειτουργία. Από κάθε σειρά βίντεο έγινε εξαγωγή των καρέ (frames) και αλλαγή διαστάσεων από 1928x1448 σε 1024x768 με χρήση αλγορίθμου υποδειγματοληψίας (downsample) με μέθοδο απόδοσης χρωμάτων τη δικυβική παρεμβολή (bicubic interpolation). Η αλλαγή των διαστάσεων έγινε διότι από τις πρώτες δοκιμαστικές εκπαιδεύσεις των μοντέλων παρατηρήθηκε ότι η μνήμη της κάρτας γραφικών δεν επαρκεί για να εκπαιδευτεί το δίκτυο με την πλήρη διάσταση των εικόνων. Αυτό οφείλεται στο γεγονός ότι όσο μεγαλώνει η διάσταση της εικόνας τόσο μεγαλώνουν και οι διαστάσεις των χαρτών χαρακτηριστικών που δημιουργούνται από τα διάφορα επίπεδα συνέλιξης του δικτύου κορμού (Backbone).

Στην συνέχεια δημιουργήθηκαν σετ δεδομένων τα οποία περιείχαν διάφορα καρέ από τα παραπάνω βίντεο ώστε να υπάρχουν καρέ που απεικονίζεται η αρπάγη σε ποικίλες συνθήκες φωτισμού και γωνίες στροφής ως προς την κάμερα. Έπειτα στα καρέ του κάθε σετ δεδομένων έγινε χειροκίνητη ψηφιοποίηση της αρπάγης με μορφή πολυγώνου. Τα δεδομένα της ψηφιοποίησης αποθηκεύονται σε αρχείο ascii τύπου json όπου καταγράφονται πληροφορίες όπως είναι οι συντεταγμένες του πολυγώνου της αρπάγης, το όνομα του καρέ στο οποίο αναφέρονται και οι διαστάσεις του καρέ. Το σετ δεδομένων που δημιουργήθηκε για την εκπαίδευση των μοντέλων της συγκεκριμένης εφαρμογής περιέχει καρέ που προέρχονται από 5 διαφορετικές σειρές βίντεο.

Κάθε σετ δεδομένων περιέχει υποφακέλους με το σετ εκπαίδευσης (training set) και το σετ ελέγχου (validation set). Το training set περιέχει εικόνες (188) που προορίζονται για την εκπαίδευση του μοντέλου και το αρχείο asci τύπου json με τις πληροφορίες των ψηφιοποιήμενων πολυγώνων. Το validation set περιέχει επίσης αρχείο asci τύπου json με τα δεδομένα της ψηφιοποίησης και εικόνες (36) που προορίζονται για τον έλεγχο του μοντέλου σε εικόνες που δεν έχει επεξεργαστεί κατά την διάρκεια της εκπαίδευσης και την παραγωγή μετρητικών στοιχείων για την αξιολόγηση του μοντέλου. Επιπλέον υπάρχει ένα σετ αξιολόσησης (test set) το οποίο περιέχει 3238 εικόνες. Το test set αποτελείτε από εικόνες που προέρχονται από διάφορες σειρές βίντεο, μερικές από τις οποίες δεν συμμετείχαν στην δημιουργία των train set και validation set. Σκοπός των δεδομένων test είναι να αξιολογηθεί η απόδοση του μοντέλου σε πραγματικές συνθήκες.

Πίνακας 2:Σετ δεδομένων

Τύπος δεδομένων	Εικόνες
Training	188
Validation	36
Test	3238

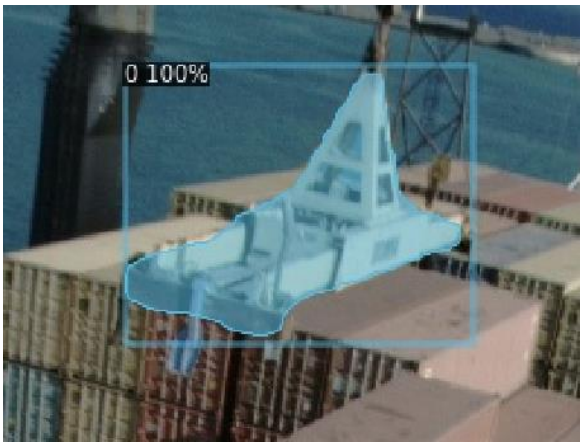
3.4 Εκπαίδευση Mask RCNN

Για την διαδικασία της εκπαίδευσης του μοντέλου MaskRCNN χρησιμοποιήθηκε το παραπάνω σετ δεδομένων με 188 εικόνες στο σετ εκπαίδευσης και 36 εικόνες στο σετ ελέγχου. Χρησιμοποιήθηκε σαν δίκτυο κορμού (backbone) το δίκτυο resnet50 καθώς είχε τις μικρότερες τιμές χρόνου πρόβλεψης σύμφωνα με τον πίνακα σύγκρισης μοντέλων στην επίσημη σελίδα του Detectron2. Σαν αρχικές τιμές για τα βάρη χρησιμοποιήθηκαν τα βάρη από μοντέλο εκπαιδευμένο στο σετ δεδομένων COCO με την χρήση του παραπάνω δικτύου κορμού.

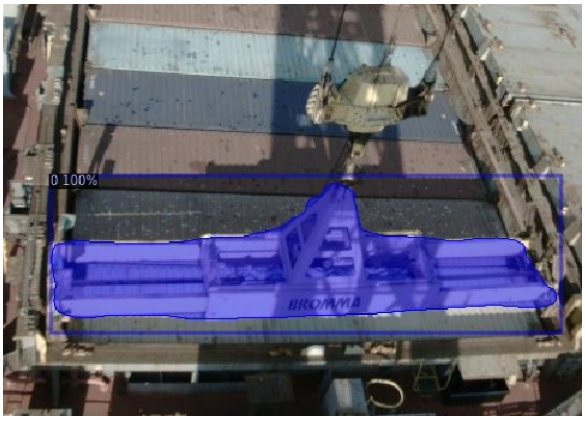
Η εκπαίδευση του μοντέλου έγινε για 100 εποχές και οι υπερπαράμετροι του μοντέλου τέθηκαν 0.9 για την ορμή (momentum), 10^{-4} για την φθορά των βαρών (weight decay) και μέγεθος παρτίδας 2 εικόνες. Η υπερπαράμετρος του ρυθμού εκμάθησης (learning rate) ρυθμίστηκε έτσι ώστε να είναι 10^{-4} για τις πρώτες 30 εποχές, 10^{-5} για τις επόμενες 30 εποχές και 10^{-6} για τις τελευταίες 40 εποχές. Επιπλέον ο αλγόριθμος ρυθμίστηκε έτσι ώστε να κάνει επικαιροποίηση (validation) στο τέλος της κάθε εποχής και να αποθηκεύει τα βάρη κάθε 5 εποχές. Η ρύθμιση αυτή έγινε έτσι ώστε αν παρατηρηθεί με την πάροδο των εποχών μείωση της απόδοσης του μοντέλου να θεωρηθεί ως τελικό μοντέλο αυτό που έχει την μικρότερη τιμή ολικού κόστους (total loss) πριν την πτωτική πορεία της απόδοσης. Με αυτό τον τρόπο αν παρατηρηθεί υπερπροσαρμογή του μοντέλου ως τελικό μοντέλο θα θεωρηθεί το τελευταίο αποθηκευμένο μοντέλο πριν την υπερπροσαρμογή. Επιπλέον για την αποφυγή της υπερπροσαρμογής του μοντέλου έγινε η χρήση παράγωγων δεδομένων (augmentation). Στα δεδομένα έγιναν φωτομετρικοί μετασχηματισμοί αυξομείωσης της φωτεινότητας, του κορεσμού και της αντίθεσης. Επίσης έγιναν μετασχηματισμοί γύρω από τον άξονα y.

Μετά την διαδικασία της εκπαίδευσης παρατηρήθηκε ότι οι προβλέψεις του μοντέλου στα δεδομένα validation είναι οι αναμενόμενες και ανιχνεύεται η αρπάγη σε όλα τα καρέ. Βέβαια το μοντέλο αξιολογείτε καλύτερα με το σετ δεδομένων test που περιέχει περισσότερα δείγματα (βλ. 4.2). Ακόμα παρατηρήθηκε ότι οι συναρτήσεις της ολικού κόστους (Total loss) και της κόστους μάσκας (Mask loss) ακολουθούν πτωτική πορεία. Η συνάρτηση μέσης ακρίβειας ακολουθεί ανοδική πορεία για τις πρώτες εποχές και μετά σταθεροποιείται χωρίς να μειώνεται με την πάροδο των εποχών. Η εικόνα αυτή των συναρτήσεων είναι θεμιτή και το παραπάνω μοντέλο χρησιμοποιήθηκε για την εφαρμογή με βάση το Mask RCNN.

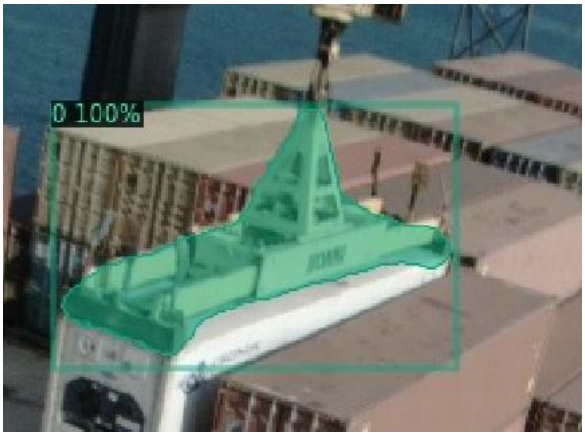
(a)



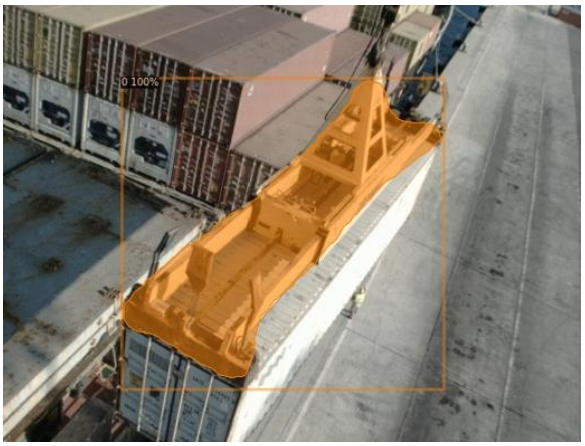
(b)



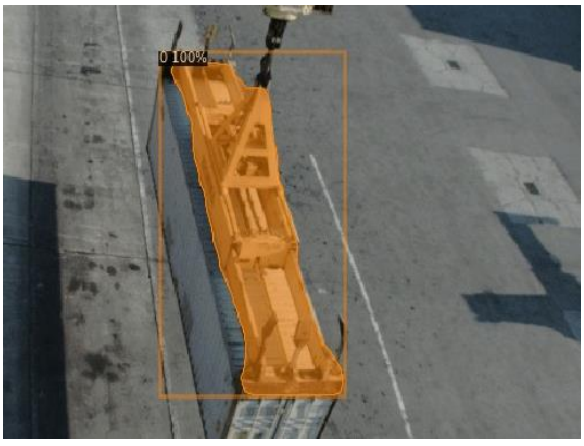
(c)



(d)

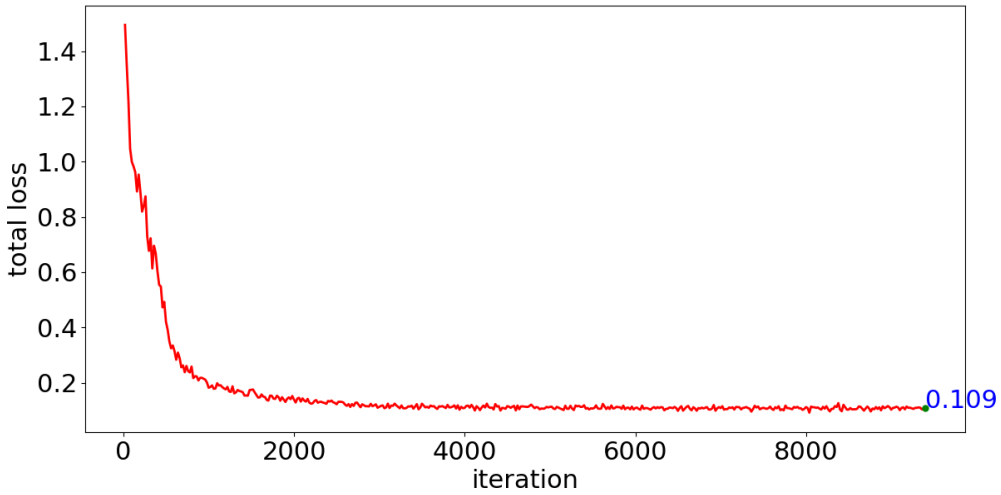


(e)

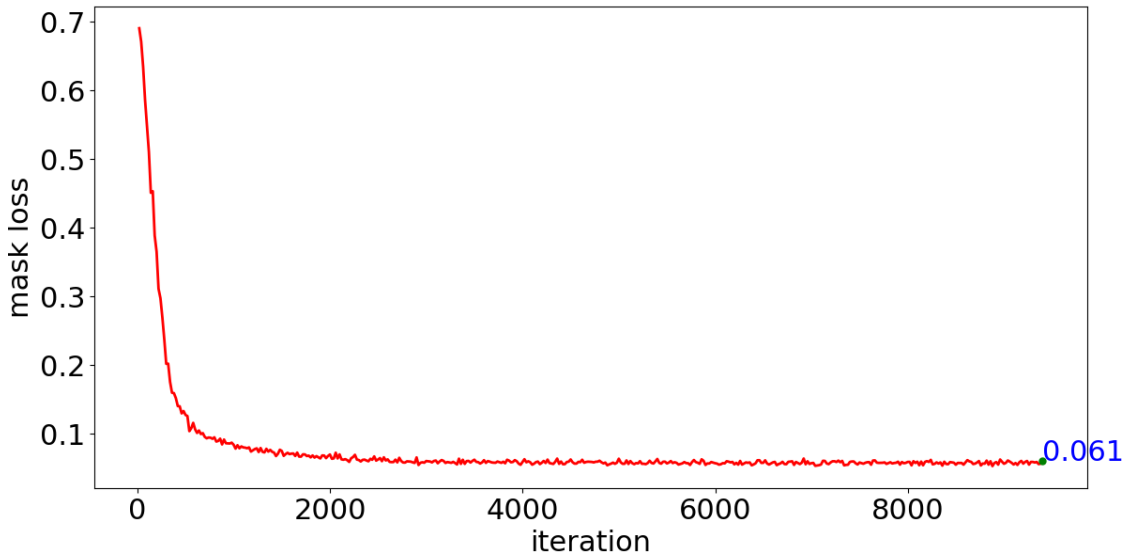


(f)

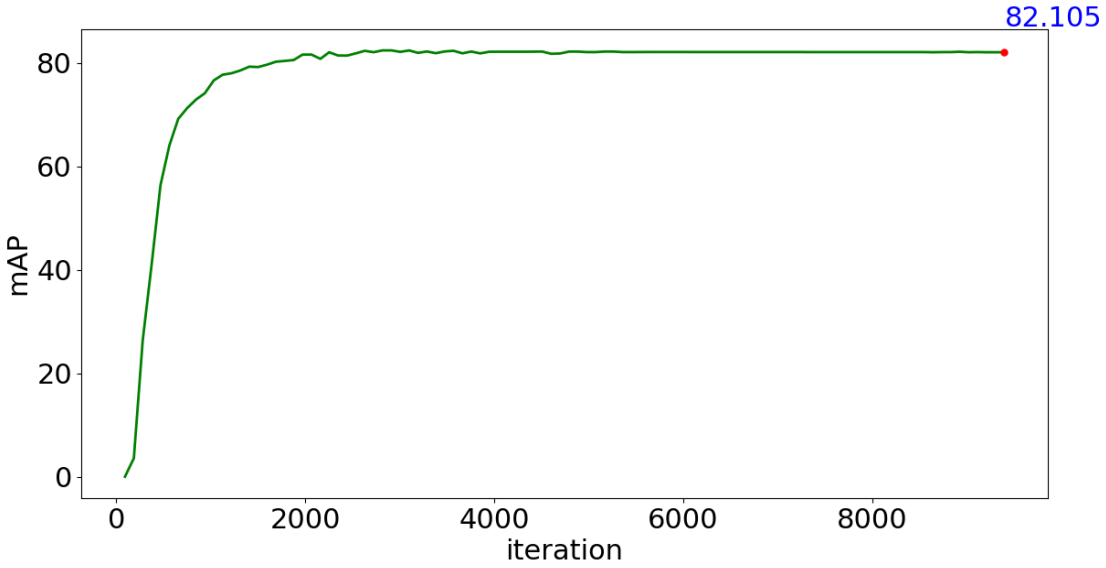
Σχήμα 21: Προβλέψεις μοντέλου Mask Rcnn σε δεδομένα Validation



Σχήμα 22: Συνάρτηση συνολικού κόστους Mask Rcnn



Σχήμα 23: Συνάρτηση κόστους μάσκας Mask Rcnn



Σχήμα 24: Μέτρο μέσης ακρίβειας (mAP) Mask Rcnn

3.5 Εκπαίδευση Yolact

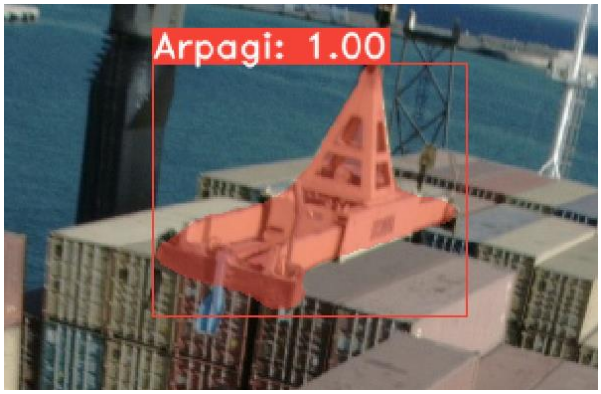
Για την διαδικασία της εκπαίδευσης του μοντέλου Yolact χρησιμοποιήθηκε το ίδιο σετ δεδομένων με 188 εικόνες στο σετ εκπαίδευσης και 36 εικόνες στο σετ ελέγχου. Σαν δίκτυο κορμού (backbone) χρησιμοποιήθηκε και σε αυτή τη περίπτωση το δίκτυο resnet 50 καθώς είχε τη μικρότερη τιμή χρόνου πρόβλεψης το πίνακα σύγκρισης μοντέλων της επίσημης σελίδας του Yolact. Σαν αρχικές τιμές για τα βάρη χρησιμοποιήθηκαν τα βάρη από μοντέλο εκπαιδευμένο στο σετ δεδομένων COCO με την χρήση του παραπάνω δικτύου κορμού.

Η εκπαίδευση του μοντέλου έγινε για 100 εποχές και οι υπερπαράμετροι του μοντέλου τέθηκαν 0.9 για την ορμή (momentum) , 10^{-4} για την φθορά των βαρών (weight decay) και μέγεθος παρτίδας 1 εικόνα. Η υπερπαράμετρος του ρυθμού εκμάθησης (learning rate) ρυθμίστηκε έτσι ώστε να είναι 10^{-4} για τις πρώτες 30 εποχές, 10^{-5} για τις επόμενες 30 επτοχές και 10^{-6} για τις τελευταίες 40 εποχές. Επιπλέον ο αλγόριθμος ρυθμίστηκε έτσι ώστε να κάνει επικαιροποίηση (validation) στο τέλος της κάθε εποχής και να αποθηκεύει τα βάρη κάθε 5 εποχές. Η ρύθμιση αυτή έγινε έτσι ώστε αν παρατηρηθεί με την πάροδο των εποχών μείωση της απόδοσης του μοντέλου να θεωρηθεί ως τελικό μοντέλο αυτό που έχει την μικρότερη τιμή συνολικού κόστους (total loss) πριν την πτωτική πορεία της απόδοσης. Η ρύθμιση αυτή όπως προαναφέρθηκε γίνεται για να αποφευχθεί η υπερπροσαρμογή του μοντέλου. Αν παρατηρηθεί από τις συναρτήσεις κόστους (loss) και από το μέτρο απόδοσης (mAP) υπερπροσαρμογή του μοντέλου με την πάροδο των εποχών τότε ως τελικό μοντέλο θεωρείτε το τελευταίο αποθηκευμένο μοντέλο πριν αρχίσει η υπερπροσαρμογή. Επιπλέον για την αποφυγή της υπερπροσαρμογής του μοντέλου έγινε η χρήση παράγωγων δεδομένων (augmentation). Στα δεδομένα έγιναν φωτομετρικοί μετασχηματισμοί αυξομείωσης της φωτεινότητας, του κορεσμού και της αντίθεσης. Επίσης έγιναν μετασχηματισμοί γύρω από τον άξονα y και τυχαίας περικοπής τμήματος αντικειμένου.

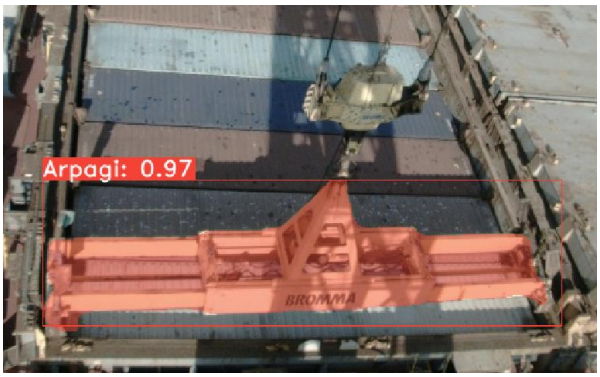
Παρατηρώντας τις συναρτήσεις για το ολικό κόστος (Total loss) και κόστος μάσκας (Mask loss) φαίνεται

ότι το μοντέλο μετά τις 10000 περίπου επαναλήψεις δεν μειώνει περαιτέρω το κόστος. Η συνάρτηση του μέτρου απόδοσης mAP φαίνεται να αυξάνεται με αργό ρυθμό μετά τις 10000 επαναλήψεις. Επιπλέον όπως και με το μοντέλο Mask Rcnn έγιναν προβλέψεις για το validation set για μία πρώτη εκτίμηση της ποιότητας πρόβλεψης και τα αποτελέσματα ήταν ικανοποιητικά. Για αυτό αποφασίστηκε ως τελικό μοντέλο να θεωρηθεί το μοντέλο της τελευταίας επανάληψης.

(a)

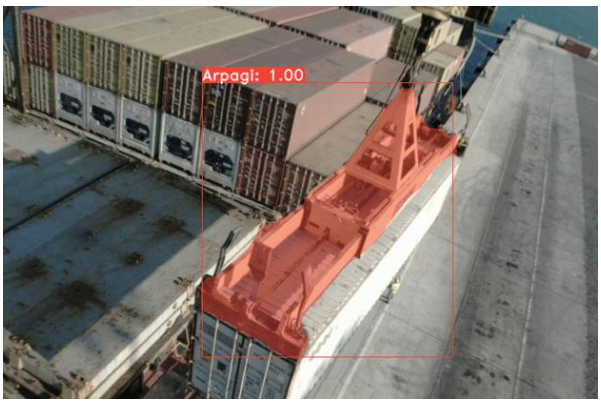


(b)

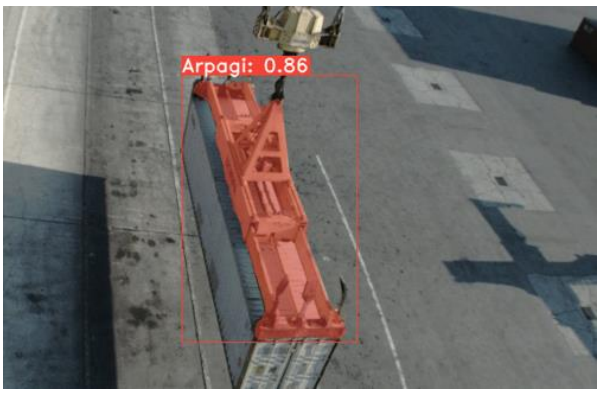


(c)

(d)

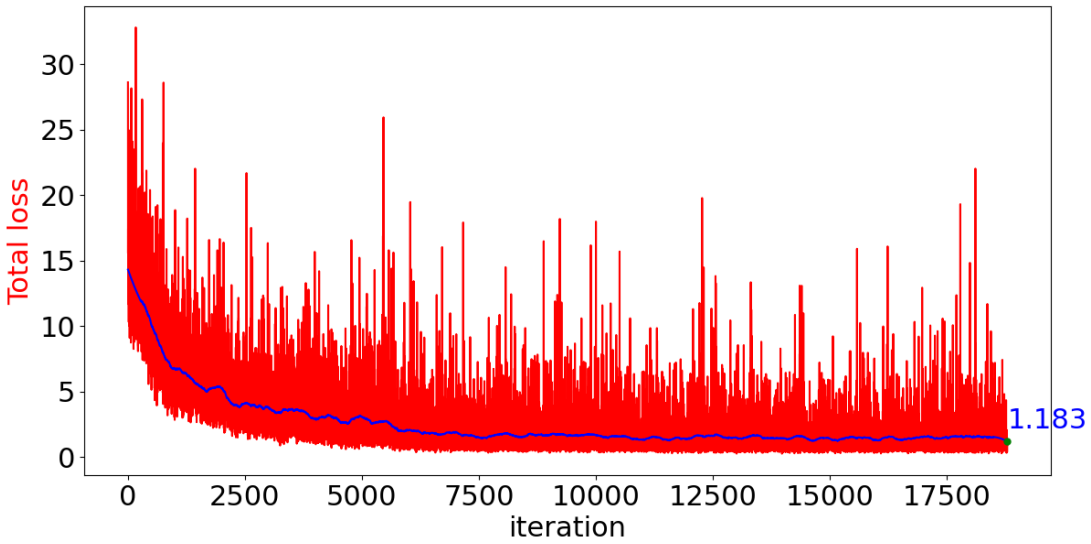


(e)

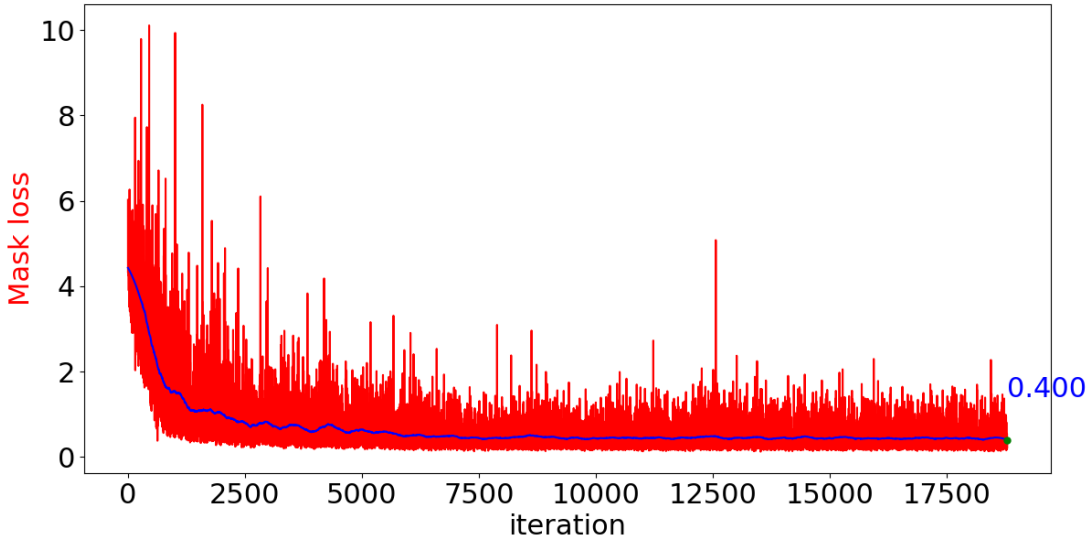


(f)

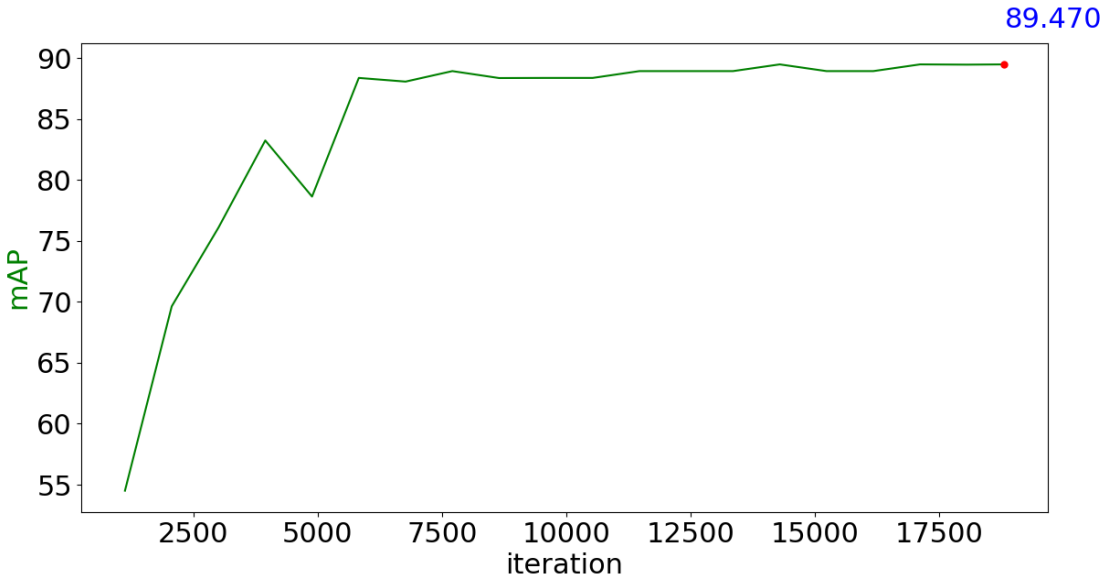
Σχήμα 25: Προβλέψεις μοντέλου Yolact σε δεδομένα Validation



Σχήμα 26: Συνάρτηση συνολικού κόστους Yolact



Σχήμα 27: Συνάρτηση κόστους μάσκας Yolact

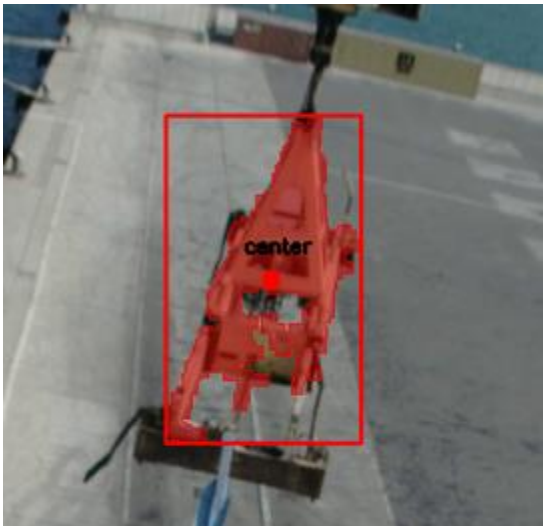


Σχήμα 28: Μέτρο μέσης ακρίβειας (mAP) Yolact

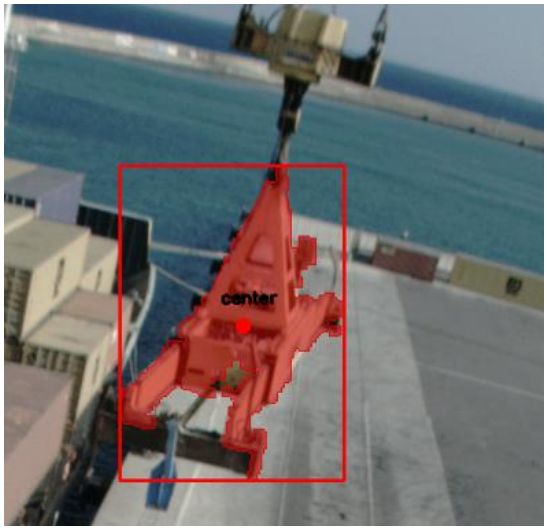
Κεφάλαιο 4: Αποτελέσματα και Αξιολόγηση συστήματος/εφαρμογή

4.1 Αξιολόγηση αλγορίθμου Color Track

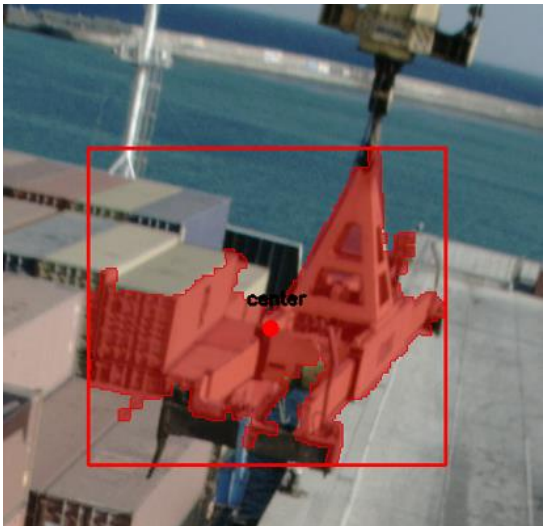
Ο αλγόριθμος color track ήταν η πρώτη προσέγγιση στην επίλυση του προβλήματος της ανίχνευσης της αρπάγης σε πραγματικό χρόνο. Για την αξιολόγηση του χρησιμοποιήθηκε σειρά βίντεο αντί για το σετ δεδομένων test. Αυτό γίνεται διότι οι εικόνες στο σετ δεδομένων test προέρχονται από διάφορες σειρές βίντεο και δεν έχουν μια συνέχεια στην θέση που απεικονίζεται η αρπάγη. Αυτό δημιουργεί πρόβλημα στην χρήση του φίλτρου Kalman που για να παράγει ορθές εκτιμήσεις για την επόμενη θέση της αρπάγης πρέπει οι προηγούμενες θέσεις τις να έχουν μία συνέχεια όπως και στις πραγματικές συνθήκες που θα κληθεί ο αλγόριθμος να παρακολουθεί την αρπάγη.



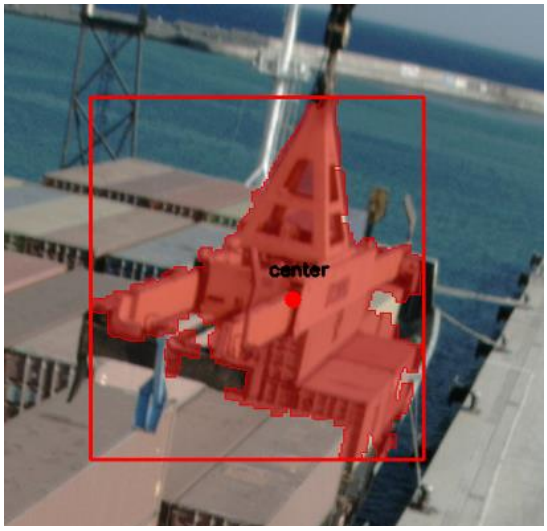
(a)



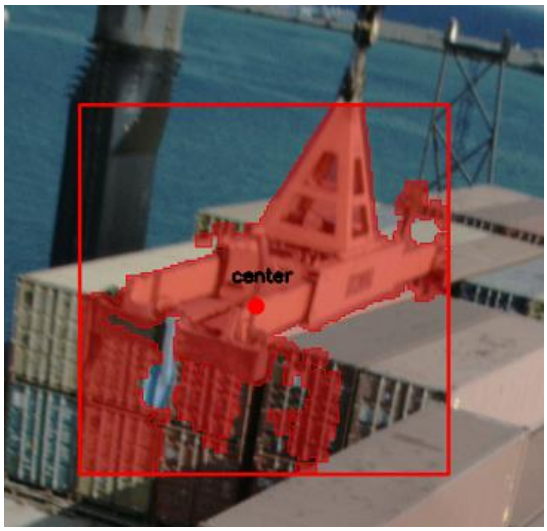
(b)



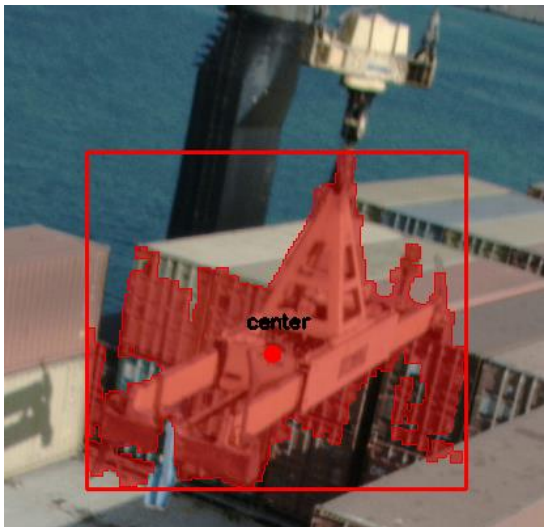
(c)



(d)

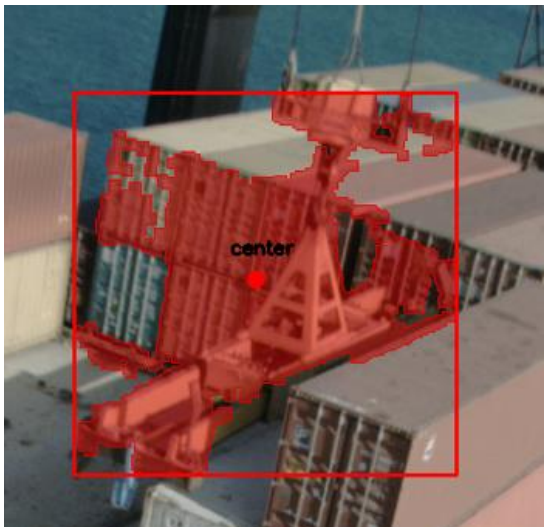


(e)



(f)

(g)



(h)

Σχήμα 29: Αποτελέσματα ανίχνευσης αλγορίθμου Color Track

Παρατηρώντας τα αποτελέσματα του αλγορίθμου φαίνεται ότι η ποιότητα τις ανίχνευσης δεν είναι καλή διότι η μάσκα που παράγει το μοντέλο δεν έχει ομαλά όρια. Αυτό οφείλεται στον τρόπο που παράγεται η μάσκα και πιθανός με τη εφαρμογή κατάλληλων μορφολογικών φίλτρων να επιλυθεί.

Το κυριότερο πρόβλημα είναι η λανθασμένη ανίχνευση αντικειμένων. Σε περιπτώσεις όπου το πλαίσιο στο οποίο αναζητά ο αλγόριθμος την αρπάγη δεν περιέχει αντικείμενα με παρόμοιο χρώμα στον χρωματικό χώρο HSV η ανίχνευση είναι σωστή (Σχήμα 29b). Ωστόσο σε περιπτώσεις που εντός αυτής της περιοχής βρίσκονται αντικείμενα όπως είναι τα κοντέινερ που έχουν παρόμοιο χρώμα με την αρπάγη τότε η τελική μάσκα που ανιχνεύει ο αλγόριθμος είναι λανθασμένη. Αυτό το αποτέλεσμα εκτός από το γεγονός ότι δεν

είναι επιθυμητό επηρεάζει και τις προβλέψεις του φίλτρου Kalman καθώς μετατοπίζεται το κέντρο της μάσκας που περιέχει την αρπάγη , τις συντεταγμένες του οποίου καλείται να εκτιμήσει το φίλτρο.

4.2 Αξιολόγηση μοντέλων Mask Rcnn και Yolact

Η ανίχνευση με την χρήση βασικών τεχνικών όρασης υπολογιστών φαίνεται πως δεν επαρκεί για την εφαρμογή. Έτσι αποφασίστηκε να χρησιμοποιηθούν τα νευρωνικά δίκτυα Mask Rcnn και Yolact για την ανίχνευση της αρπάγης. Η αξιολόγηση των μοντέλων γίνεται λαμβάνοντάς υπόψη τις κύριες προδιαγραφές τις εφαρμογής, δηλαδή:

- ανίχνευση σε δυναμικό περιβάλλον
- ανίχνευση σε πραγματικό χρόνο

4.2.1 Ανίχνευση σε δυναμικό περιβάλλον

Για να εξεταστεί η ποιότητα της ανίχνευσης έγιναν προβλέψεις των δύο μοντέλων για το σετ δεδομένων Test που όπως προαναφέρθηκε δημιουργήθηκε από σειρές βίντεο που δεν αξιοποιήθηκαν όλες για την δημιουργία των σετ δεδομένων train και validation. Το κάθε μοντέλο πέρα από τη μάσκα και το πλαίσιο που περιβάλλει το αντικείμενο εξάγει και την τιμή της πιθανότητας το αντικείμενο να είναι αρπάγη. Το confidence score είναι μια τιμή κατωφλίου πάνω από την οποία θα εμφανίζονται οι προβλέψεις του μοντέλου. Για τα δύο μοντέλα χρησιμοποιήθηκε χαμηλό confidence score κατά την αξιολόγηση με σκοπό να εμφανίζεται η αρπάγη ακόμα και όταν δεν φαίνεται ολόκληρη αλλά τμήμα της.

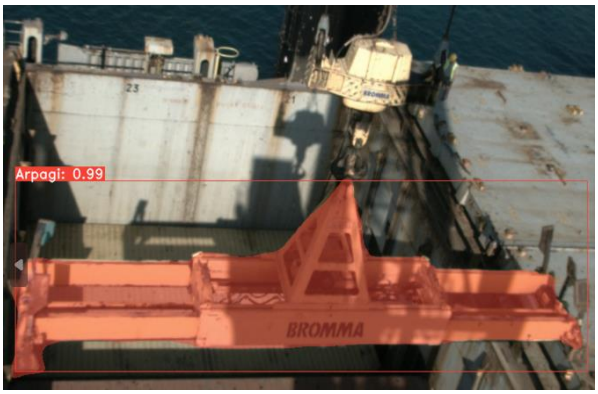
Παρατηρώντας τις προβλέψεις προκύπτει ότι οι μάσκες που παράγει το μοντέλο Yolact είναι πιο λεπτομερείς από τις αντίστοιχες του μοντέλου Mask Rcnn. Οι μάσκες του μοντέλου Yolact καλύπτουν πιο σωστά τα όρια της αρπάγης ενώ του Mask Rcnn έχουν μια κυματοειδή μορφή και σε αρκετές περιπτώσεις αφήνουν μικρά τμήματα της αρπάγης εκτός μάσκας ή βάζουν τμήματα του φόντου εντός μάσκας. Αυτό είναι αναμενόμενο καθώς όπως προαναφέρθηκε το Yolact χρησιμοποιεί χαμηλότερα επίπεδα του συνελικτικού δικτύου εξαγωγής χαρακτηριστικών (FPN) για την παραγωγή των μασκών και συνεπώς οι πρότυπες μάσκες έχουν μεγαλύτερη ανάλυση.

Yolact

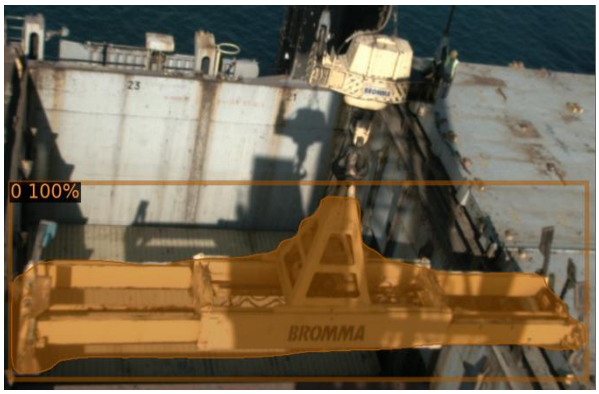
(a)

Mask Rcnn

(b)


Arpagi: 0.99
BROMMA

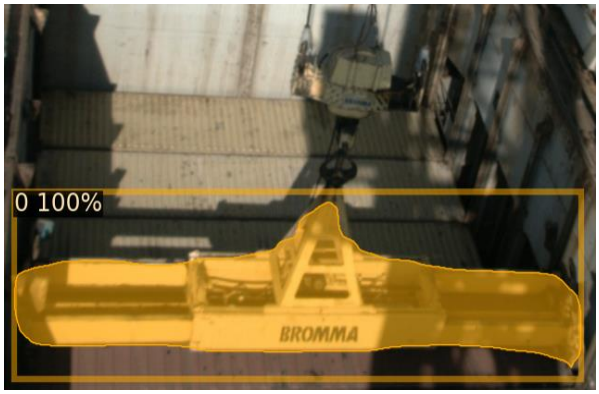
(c)


0 100%
BROMMA

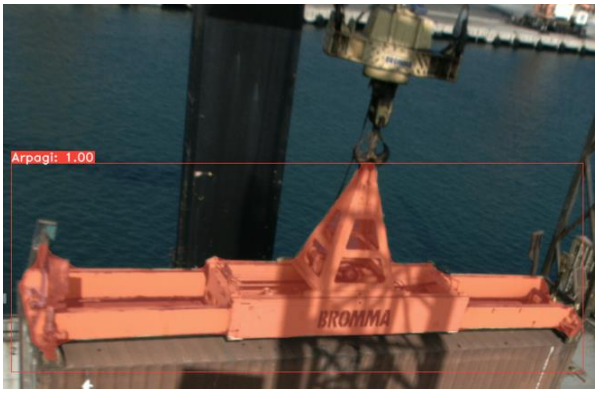
(d)


Arpagi: 0.87
BROMMA

(e)


0 100%
BROMMA

(f)


Arpagi: 1.00
BROMMA

(g)


0 100%
BROMMA

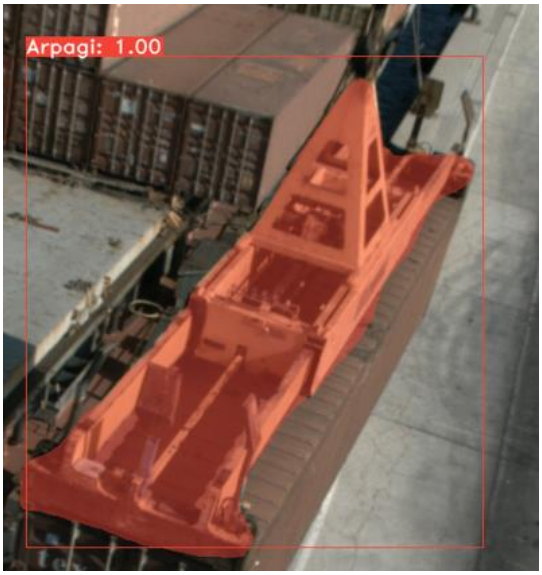
(h)


Arpagi: 1.00
BROMMA

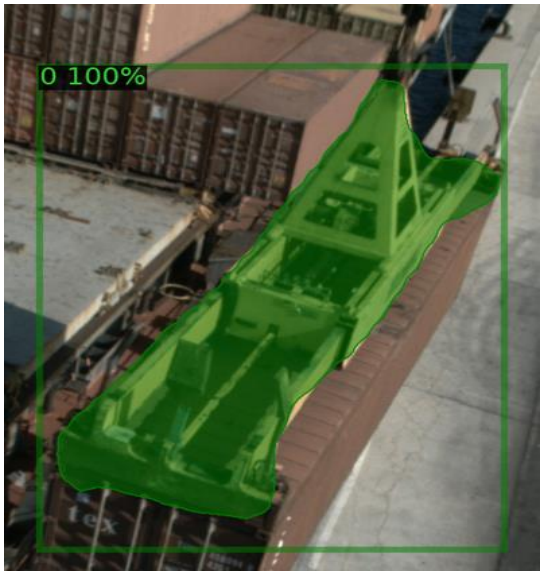
(i)


0 100%
BROMMA

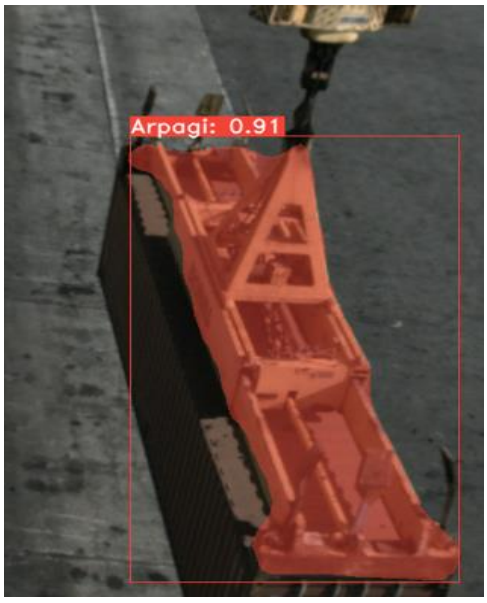
(j)


Arpagi: 1.00

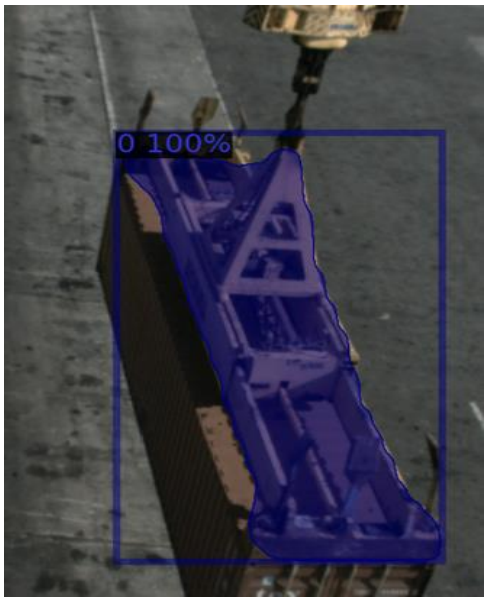
(k)


0 100%

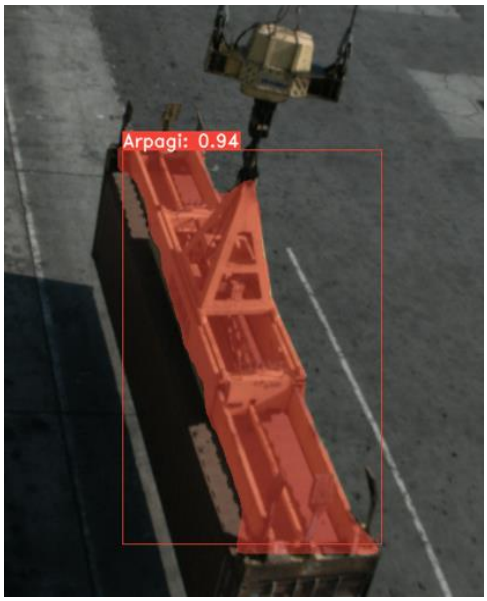
(l)


Arpagi: 0.91

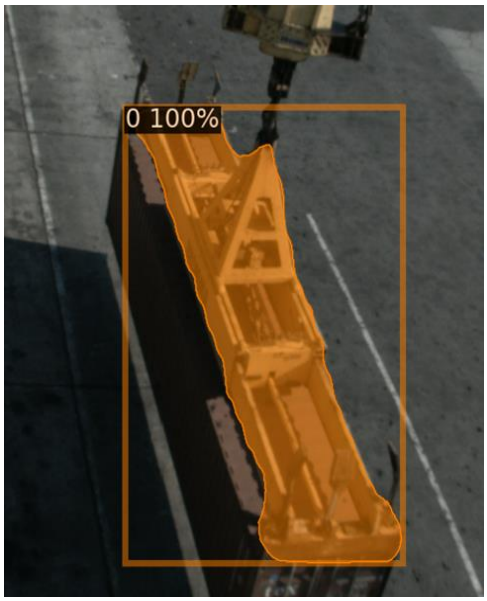
(m)


0 100%

(n)



(o)



(p)

Σχήμα 30: Σύγκριση ποιότητας μάσκας των μοντέλων Yolact (αριστερά) και Mask Rcnn (δεξιά)

Από τις προβλέψεις των δύο μοντέλων πέρα από την ποιότητα των μασκών παρατηρούνται και ορισμένες αδυναμίες των μοντέλων όσο αναφορά το δυναμικό περιβάλλον όπως είναι:

- Κακή ποιότητα ανίχνευσης σε μεγάλες τιμές φωτεινότητας :
  Σε μία περίπτωση όπου η φωτεινότητα της εικόνας είναι υψηλή η ποιότητα των προβλέψεων είναι κακή. Συγκεκριμένα το μοντέλο Mask Rcnn κάνει καλύτερη πρόβλεψη από το μοντέλο Yolact (Σχήμα 31). Τα δύο μοντέλα έχουν ρυθμιστεί να κάνουν την διαδικασία augmentation για την φωτεινότητα φαίνεται αλλά η διαδικασία αυτή αυξάνει την φωτεινότητα με μία πιθανότητα 0.5 και με ένα συγκεκριμένο εύρος τιμών. Άρα υπάρχει πιθανότητα οι τιμές που επιλέχθηκαν τυχαία από τον αλγόριθμο του Yolact να μην είναι αρκετά υψηλές ώστε να προσομοιάζει η τελική εικόνα μεγάλα μεγέθη φωτεινότητας ενώ αντίστοιχα οι τιμές που επιλέχθηκαν τυχαία από τον αλγόριθμο του Mask Rcnn να ήταν. Αυτό θα μπορούσε να λυθεί με τροποποίηση του αλγορίθμου του Yolact ή με την χειροκίνητη διεύρυνση του σετ δεδομένων εκπαίδευσης ώστε να περιλαμβάνει περισσότερες εικόνες με διάφορα εύρη φωτεινότητας.

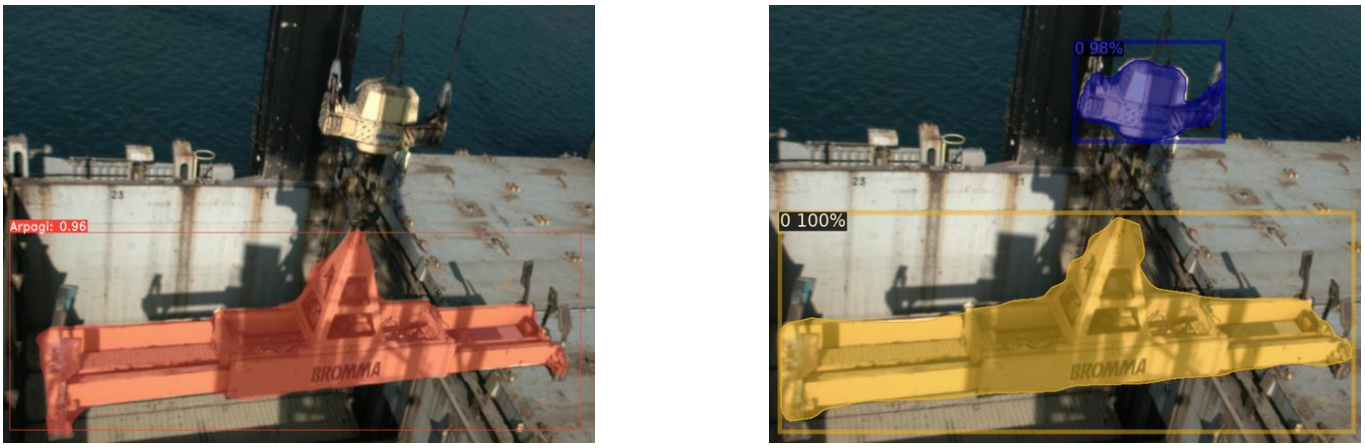
Yolact **Mask Rcnn**

(a) (b)

Σχήμα 31: Σύγκριση ποιότητας αποτελεσμάτων σε συνθήκες υψηλού φωτισμού μοντέλων Yolact (αριστερά) και Mask Rcnn (δεξιά)

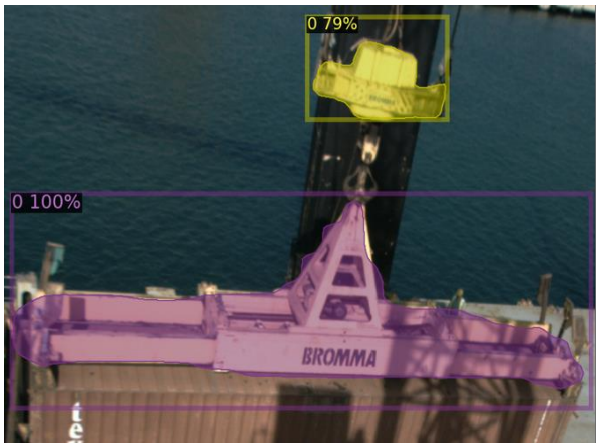
- Λανθασμένη ανίχνευση αντικειμένου:
  Σε αρκετές περιπτώσεις το μοντέλο Mask Rcnn αναγνωρίζει ως αρπάγη αντικείμενο που δεν είναι η αρπάγη, πράγμα δεν συμβαίνει με την περίπτωση του μοντέλου Yolact (Σχήμα 32). Αυτό θα μπορούσε να βελτιωθεί με την αύξηση του confidence score πάρα πολύ, η τιμή να τείνει στο 100%. Η συγκεκριμένη πρακτική όμως θα δημιουργούσε πρόβλημα σε περιπτώσεις που φαίνεται τμήμα της αρπάγης και το μοντέλο δίνει πιθανότητα το αντικείμενο που ανίχνευσε να είναι αρπάγη μικρότερη του 100%.

Yolact **Mask Rcnn**



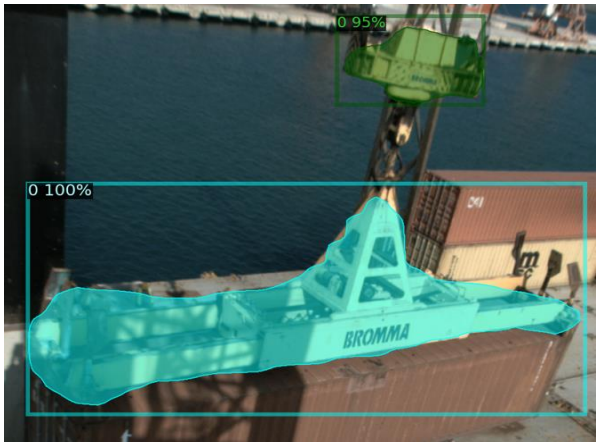
(a) (b)

(c)



(d)

(e)

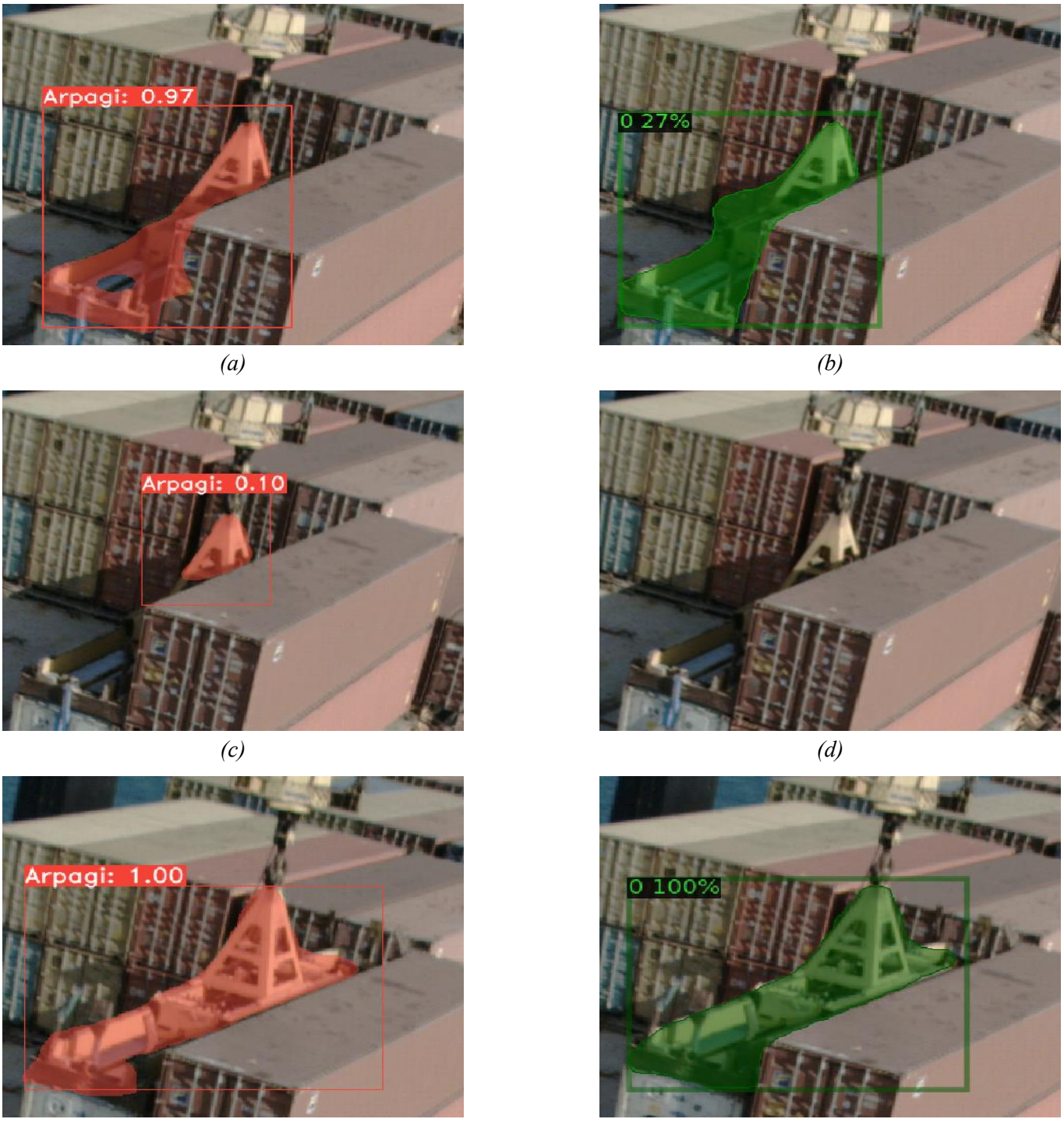


(f)

Σχήμα 32: Σύγκριση λανθασμένης ανίχνευσης των αποτελεσμάτων των μοντέλων Yolact (αριστερά) και Mask Rcnn (δεξιά)

Στο σετ δεδομένων Test υπάρχει μία περίπτωση όπου η θέα της αρπάγης μπλοκάρετε από ένα κοντέινερ και παρατηρείτε ότι και τα δύο μοντέλα ‘χάνουν’ την αρπάγη. Το μοντέλο Yolact όμως την εντοπίζει σε μερικά καρέ πριν χαθεί τελείως ενώ το μοντέλο Mask Rcnn όχι. Αυτό οφείλεται στο γεγονός ότι το μοντέλο Yolcat έχει εκπαιδευτεί με augmentation περικοπής τμήματος αντικειμένου και έτσι ο αλγόριθμος έχει μάθει να αναγνωρίζει το συγκεκριμένο τμήμα της αρπάγης. Στις ανιχνεύσεις του τμήματος της αρπάγης είναι εμφανές γιατί επιλέχθηκαν χαμηλά confidence score για τα δύο μοντέλα. Στην πρώτη περίπτωση το μοντέλο Yolact (Σχήμα 33a) εντοπίζει την αρπάγη με πιθανότητα 97% ενώ το μοντέλο Mask Rcnn (Σχήμα 33b) μόλις με 27%. Σε επόμενο καρέ το μοντέλο Yolact (Σχήμα 33c) εντοπίζει την αρπάγη με πιθανότητα 10% ενώ το μοντέλο Mask Rcnn (Σχήμα 33d) δεν καταφέρνει να εντοπίσει την αρπάγη.

Yolact **Mask Rcnn**



(a) *(b)*

(c) *(d)*

(e) *(f)*

Σχήμα 33: Σύγκριση σε συνθήκες όπου φαίνεται τμήμα της αρπάγης των αποτελεσμάτων των μοντέλων Yolact (αριστερά) και Mask Rcnn (δεξιά)

4.2.2 Ανίχνευση σε πραγματικό χρόνο

Οι σειρές βίντεο που δόθηκαν σαν δεδομένα έχουν μέγεθος καρέ 1928x1448 και ταχύτατα μετάδοσης 15 καρέ/δευτερόλεπτο (FPS). Για να εξετασθεί αν ένα μοντέλο μπορεί να πραγματοποιήσει ανίχνευση σε πραγματικό χρόνο πρέπει ο χρόνος πρόβλεψης να μετατραπεί σε FPS. Στον συνολικό χρόνο εξαγωγής προβλέψεων υπολογίζεται και η διαδικασία της οπτικοποίησης των αποτελεσμάτων. Η μετατροπή αυτή γίνεται εύκολα με την εξίσωση:

$$FPS = \frac{1}{inference\ time\ (s)}$$

όπου inference time: ο συνολικός χρόνος πρόβλεψης σε s

Ο αλγόριθμος Color Track πετυχαίνει περίπου 15 fps λειτουργώντας με την χρήση της CPU (βλ. 3.2). Αυτό σημαίνει ότι μπορεί να τρέξει σε πραγματικό χρόνο με χαμηλές υπολογιστικές απαιτήσεις, δηλαδή δεν χρειάζεται ξεχωριστή κάρτα γραφικών. Όσον αφορά το μοντέλο Yolact πετυχαίνει σημαντικά πιο μικρούς χρόνους πρόβλεψης σε σχέση με το μοντέλο Mask Rcnn για το συγκεκριμένο μέγεθος καρέ. Η τιμή του χρόνου πρόβλεψης για το μοντέλο Yolact αντιστοιχεί σε περίπου 20 FPS ενώ το Mask Rcnn πετυχαίνει περίπου 7 FPS. Συγκρίνοντας τις τιμές αυτές με την τιμή 15FPS που έχει η σειρά βίντεο από την δεδομένη κάμερα προκύπτει ότι μόνο το μοντέλο Yolact μπορεί να λειτουργήσει σε πραγματικό χρόνο.

Πίνακας 3:Χρόνοι ανίχνευσης

Model	Inference time (ms)	FPS
Yolact	48.309	20.7
Color Track	67.568	14.8
Mask Rcnn	144.928	6.9

Κεφάλαιο 5: Συμπεράσματα και μελλοντική εργασία

Συνοψίζοντας, στόχος της συγκεκριμένης εργασίας ήταν να δημιουργηθεί εφαρμογή που θα μπορεί να ανιχνεύει την αρπάγη σε πραγματικό χρόνο και σε ένα δυναμικό περιβάλλον. Το πρόβλημα προσεγγίστηκε με τρεις τρόπους:

1. Αλγόριθμος Color Track
2. Mask Rcnn
3. Yolact

Ο αλγόριθμος Color Track αξιοποιεί βασικές τεχνικές όρασης υπολογιστών για την ανίχνευση της αρπάγης. Συγκεκριμένα αξιοποιείται η ιδιότητα του χρώματος για την ανίχνευση της αρπάγης και χρησιμοποιείται ο χρωματικός χώρος HSV για την καλύτερη απομόνωση του χρώματος της αρπάγης. Επίσης ο αλγόριθμος χρησιμοποιεί ένα φίλτρο Kalman δύο διαστάσεων ώστε να προβλέπει την επόμενη θέση της αρπάγης και να ψάχνει στην περιοχή που ορίζει η πρόβλεψη να ανίχνευση μελλοντικά την αρπάγη. Η αξιολόγηση έδειξε ότι η ποιότητα των προβλέψεων δεν είναι επιθυμητή καθώς σε πολλές περιπτώσεις κοντέινερ και άλλα αντικείμενα έχουν χρώμα που βρίσκεται εντός του εύρους χρωμάτων που ανιχνεύει ο αλγόριθμος. Παρόλα αυτά ο αλγόριθμος τρέχει σε πραγματικό χρόνο και με την χρήση μόνο CPU.

Στην δεύτερη προσέγγιση αξιοποιήθηκε το τεχνητό νευρωνικό δίκτυο Mask Rcnn εκπαιδεύτηκε για την ανίχνευση της αρπάγης. Το Mask Rcnn είναι δίκτυο βαθιάς μάθησης δύο σταδίων. Στο πρώτο στάδιο εξάγει προτάσεις περιοχών και στο δεύτερο ανιχνεύει τα αντικείμενα που ενδεχομένως βρίσκονται σε αυτές τις περιοχές. Η αξιολόγηση του έδειξε ότι έχει καλύτερη ποιότητα προβλέψεων από αυτή του αλγορίθμου Color Track. Σε ορισμένες περιπτώσεις όμως στην ανίχνευση υπάρχουν θέματα όπως είναι η λανθασμένη ανίχνευση αντικειμένου και η ποιότητα των γραμμών των ορίων που παρουσιάζει μία κυματοειδή μορφή. Το μεγαλύτερο μειονέκτημα όμως που παρουσιάζει το Mask Rcnn για τα δεδομένα της συγκεκριμένης εφαρμογής είναι ότι δεν πετυχαίνει ανίχνευση σε πραγματικό χρόνο για εικόνα μεγέθους όσο και το καρέ της κάμερας που χρησιμοποιήθηκε για την ανίχνευση.

Η τελευταία προσέγγιση έγινε με την χρήση του τεχνητού νευρωνικού δικτύου Yolact. Το συγκεκριμένο δίκτυο πρόκειται για δίκτυο βαθιάς μάθησης ενός σταδίου όπου χρησιμοποιεί όλη την εικόνα για την ανίχνευση του αντικειμένου, γλιτώνοντας έτσι το στάδιο των προτάσεων περιοχών. Το δίκτυο Yolact εκπαιδεύτηκε στο ίδιο σετ δεδομένων με το δίκτυο Mask Rcnn. Η αξιολόγηση δείχνει ότι η ποιότητα των προβλέψεων του είναι η επιθυμητή για την εφαρμογή. Τα όρια της ανίχνευσης είναι πιο κοντά στα πραγματικά και δεν παρουσιάζει πρόβλημα λανθασμένης ανίχνευσης για το ίδιο σετ δεδομένων που δοκιμάστηκε και το μοντέλο Mask Rcnn. Μονάχα σε μία περίπτωση που οι συνθήκες φωτεινότητας ήταν εξαιρετικά υψηλές η ποιότητα της ανίχνευσης το μοντέλου Yolact υστερεί σε σχέση με το μοντέλο Mask Rcnn. Επιπλέον το μοντέλο Yolact πετυχαίνει ανίχνευση σε πραγματικό χρόνο. Συνεπώς από την διαδικασία της αξιολόγησης προκύπτει ότι το μοντέλο Yolact είναι το κατάλληλο και πληροί τις προδιαγραφές για την εφαρμογή.

Στα πλαίσια μιας διπλωματικής εργασίας παράγοντες όπως είναι ο χρόνος διεκπεραίωσης επηρεάζουν την διεξοδικότητα με την οποία έχει εκτελεστεί το εκάστοτε ζητούμενο. Έτσι και στη συγκεκριμένη εργασία υπάρχουν περιθώρια βελτίωσης σε μερικούς τομείς, που θα μπορούσαν να αποτελέσουν και αντικείμενο για μελλοντική εργασία.

Στην διαδικασία της εκπαίδευσης των νευρωνικών δικτύων μπορούν να γίνουν ορισμένες βελτιώσεις. Αρχικά μπορεί να διευρυνθεί το σετ δεδομένων εκπαίδευσης ώστε να περιλαμβάνει ακόμα περισσότερη διαφοροποίηση στις στροφές και τις γωνίες θέασης της αρπάγης άλλα και στις δυναμικές συνθήκες που οφείλονται στις εναλλαγές των καιρικών συνθηκών και τις φωτεινότητας. Έτσι το μοντέλο θα γενικεύει καλύτερα και θα βελτιωθεί η ποιότητα της ανίχνευσης. Επιπλέον πειραματισμός με το είδος και το πλήθος των μετατροπών των δεδομένων εκπαίδευσης κατά την διαδικασία του augmentation θα συμβάλουν επίσης στην καλύτερη γενίκευση του μοντέλου. Ακόμα μπορούν να εφαρμοστούν τεχνικές για την μείωση του χρόνου πρόβλεψης των μοντέλων ώστε να επιτυγχάνουν ανίχνευση σε πραγματικό χρόνο και για σειρές βίντεο με περισσότερα FPS.

Όσον αφορά τον αλγόριθμο Color Track θα μπορούσε να αξιοποιηθεί κάποιος αλγόριθμος παρακολούθησης (tracker algorithm) για την πρόβλεψη της θέσης της αρπάγης και μετά να γίνεται η εξαγωγή της μάσκας με την χρήση της πληροφορίας του χρώματος. Επιπλέον αυτός ο αλγόριθμος θα μπορούσε να χρησιμοποιηθεί υβριδικά με το μοντέλο Mask Rcnn. Θα μπορούσε να χρησιμοποιηθεί η πρόβλεψη του μοντέλου για την αρχικοποίηση της θέσης της αρπάγης στον αλγόριθμο Color Track. Έπειτα θα μπορούσε να προγραμματιστεί στον αλγόριθμο η δυναμική ανανέωση του εύρους στο οποίο ο αλγόριθμος αναζητά την αρπάγη με βάση το χρώμα.

Κεφάλαιο 6: Αναφορές

Βιβλιογραφία

1. Abdulla, W., 2017. *Github.* [Ηλεκτρονικό] Available at: https://github.com/matterport/Mask_RCNN [Πρόσβαση 31 5 2023].
2. augmentedstartups, 2022. [Ηλεκτρονικό] Available at: https://www.augmentedstartups.com/blog/Object-detection-vs-Object-segmentation [Πρόσβαση 8 6 2023].
3. Bolya, D., Zhou, C., Xiao, F. & Lee, Y. J., 2019. *github.* [Ηλεκτρονικό] Available at: https://github.com/dbolya/yolact [Πρόσβαση 31 5 2023].
4. Bolya, D., Zhou, C., Xiao, F. & Lee, Y. J., 2019. Yolact: Real-time instance. *IEEE International Conference on Computer (ICCV).*
5. Budiyanto, M. A. & H.Fernanda, 2020. Risk assessment of work. *J. of Mar. Sci. and Eng.*.
6. dev.to, 2023. [Ηλεκτρονικό] Available at: https://dev.to/dattran1999/how-neural-networks-work-dma [Πρόσβαση 31 05 2023].
7. engineering.matterport, 2018. [Ηλεκτρονικό] Available at: https://engineering.matterport.com/splash-of-color-instance-segmentation-with-mask-r-cnn-and-tensorflow-7c761e238b46 [Πρόσβαση 18 06 2023].
8. Girshick, R., 2015. Fast R-CNN. *2015 IEEE International Conference on Computer Vision (ICCV),* pp. 1440-1448.
9. Girshick, R., Donahue, J., Darrell, T. & Malik, J., 2014. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. *IEEE Conference on Computer Vision and Pattern Recognition,* pp. 580-587.
10. HE., K., Gkioxari., G. & P.Dollar, 2017. Mask R-CNN. *IEEE International Conference on Computer Vision (ICCV),* pp. 2980-2988.
11. learnopencv, 2023. [Ηλεκτρονικό] Available at: https://learnopencv.com/intersection-over-union-iou-in-object-detection-and-segmentation/ [Πρόσβαση 31 5 2023].
12. learnopencv, 2023. [Ηλεκτρονικό] Available at: https://learnopencv.com/intersection-over-union-iou-in-object-detection-and-segmentation/ [Πρόσβαση 31 5 2023].
13. learnopencv, 2023. [Ηλεκτρονικό] Available at: https://learnopencv.com/non-maximum-suppression-theory-and-implementation-in-pytorch/

[Πρόσβαση 31 5 2023].

14. machine-learning.paperspace, 2023. [Ηλεκτρονικό] Available at: https://machine-learning.paperspace.com/wiki/activation-function [Πρόσβαση 31 5 2023].
15. medium, 2021. [Ηλεκτρονικό] Available at: https://medium.com/@thepyprogrammer/2d-image-convolution-with-numpy-with-a-handmade-sliding-window-view-946c4acb98b4 [Πρόσβαση 7 6 2023].
16. medium, 2023. [Ηλεκτρονικό] Available at: https://medium.com/@waiyan.nn18/what-is-over-fitting-how-to-avoid-57cd72fa7e8 [Πρόσβαση 27 5 2023].
17. medium, 2023. [Ηλεκτρονικό] Available at: https://medium.com/analytics-vidhya/building-blocks-of-convolutional-neural-network-e641b6772008 [Πρόσβαση 31 5 2023].
18. RahmadSadli, χ.χ. *github.* [Ηλεκτρονικό] Available at: https://github.com/RahmadSadli/2-D-Kalman-Filter#readme [Πρόσβαση 6 6 2023].
19. Ren, S., He, K., Girshick, R. & Sun, J., 2017. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* pp. 1137-1149.
20. researchgate, 2023. [Ηλεκτρονικό] Available at: https://www.researchgate.net/figure/The-RGB-color-space-visualized-as-a-cube_fig3_228719004 [Πρόσβαση 8 6 2023].
21. Welch, G. & Bishop, G., 2006. [Ηλεκτρονικό] [Πρόσβαση 6 3 2023].
22. wikimedia, 2023. [Ηλεκτρονικό] Available at: https://commons.wikimedia.org/wiki/File:HSV_color_solid_cone.png [Πρόσβαση 31 5 2023].
23. Wu, Y. και συν., 2019. *github.* [Ηλεκτρονικό] Available at: https://github.com/facebookresearch/detectron2 [Πρόσβαση 31 5 2023].