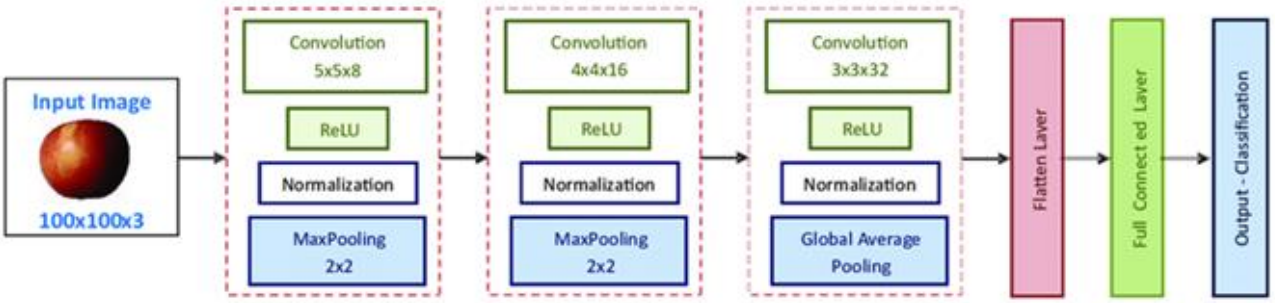
ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ
ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΤΟΠΟΓΡΑΦΙΑΣ & ΓΕΩΠΛΗΡΟΦΟΡΙΚΗΣ
ΠΜΣ: ΓΕΩΧΩΡΙΚΕΣ ΤΕΧΝΟΛΟΓΙΕΣ

«Σύγκριση Μοντέλων Βαθιάς Εκμάθησης Μηχανής για την Ανίχνευση και την Αναγνώριση Φρούτων από Εικόνες»



Μεταπτυχιακή Διπλωματική Εργασία

Θεοδοσία Βαρδουλάκη

Α.Μ. 1709

Τριμελής Επιτροπή

Γραμματικόπουλος Λ.
Πέτσα Ε.
Καλησπεράκης Η.

University of West Attica

School of Engineering

Department of Surveying & Geoinformatics Engineering

MSc: Geospatial Technologies

Fruit Image Classification and Detection Methods

Master Thesis

Vardoulaki Theodosia

November 2021

ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ
ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΤΟΠΟΓΡΑΦΙΑΣ & ΓΕΩΠΛΗΡΟΦΟΡΙΚΗΣ
ΠΜΣ: ΓΕΩΧΩΡΙΚΕΣ ΤΕΧΝΟΛΟΓΙΕΣ

«Σύγκριση Μοντέλων Βαθιάς Εκμάθησης Μηχανής για την Ανίχνευση και την Αναγνώριση Φρούτων από Εικόνες»

Η μεταπτυχιακή διπλωματική εργασία εξετάστηκε επιτυχώς από την κάτωθι Εξεταστική Επιτροπή:

α/α	ΟΝΟΜΑ ΕΠΩΝΥΜΟ	ΒΑΘΜΙΔΑ/ΙΔΙΟΤΗΤΑ	ΨΗΦΙΑΚΗ ΥΠΟΓΡΑΦΗ
1.	Λάζαρος Γραμματικόπουλος	Επίκουρος Καθηγητής	Lazaros Grammatikopoulos Digitally signed by Lazaros Grammatikopoulos Date: 2022.03.29 23:12:19 +03'00'
2.	Ελένη Πέτσα	Καθηγήτρια	Eleni Petsa Digitally signed by Eleni Petsa Date: 2022.03.30 00:45:51 +03'00'
3.	Ηλίας Καλησπεράκης	Αγρονόμος Τοπογράφος Μηχανικός ΕΜΠ Διδάκτωρ Μηχανικός ΕΜΠ	

ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΜΕΤΑΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ

Η κάτωθι υπογεγραμμένη Βαρδουλάκη Θεοδοσία του Εμμανουήλ, με αριθμό μητρώου GST1709 φοιτήτρια του Προγράμματος Μεταπτυχιακών Σπουδών «Γεωχωρικές Τεχνολογίες» του Τμήματος Μηχανικών Τοπογραφίας και Γεωπληροφορικής της Σχολής Μηχανικών του Πανεπιστημίου Δυτικής Αττικής, δηλώνω ότι:

«Είμαι συγγραφέας αυτής της μεταπτυχιακής εργασίας και ότι κάθε βοήθεια την οποία είχαγια την προετοιμασία της, είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, οι όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε ακριβώς είτε παραφρασμένες, αναφέρονται στο σύνολό τους, με πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο. Επίσης, βεβαιώνω ότι αυτή η εργασία έχει συγγραφεί από μένα αποκλειστικά και αποτελεί προϊόν πνευματικής ιδιοκτησίας τόσο δικής μου, όσο και του Ιδρύματος.

Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του πτυχίου μου».

Η Δηλούσα

Βαρδουλάκη Θεοδοσία

Λάζαρος Γραμματικόπουλος
Επίκουρος Καθηγητής

Ελένη Πέτσα
Καθηγήτρια

Lazaros Grammatikopoulos
Digitally signed by Lazaros Grammatikopoulos
Date: 2022.03.29 23:12:56 +03'00'

Πρόλογος

Η παρούσα μεταπτυχιακή διπλωματική εργασία εκπονήθηκε στο πλαίσιο της ολοκλήρωσης των μεταπτυχιακών μου σπουδών στο μεταπτυχιακό πρόγραμμα σπουδών Γεωχωρικές Τεχνολογίες του Τμήματος Μηχανικών Τοπογραφίας και Γεωπληροφορικής στο Πανεπιστήμιο Δυτικής Αττικής.

Σε αυτό το σημείο θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου, κύριο Γραμματικόπουλο Λάζαρο, για τη βοήθεια, την καθοδήγηση και το ενδιαφέρον που έδειξε σε όλη τη διάρκεια της παρούσας εργασίας. Στην συνέχεια ευχαριστώ τον αδερφό μου για την πολύτιμη βοήθεια και στήριξη του σε όλη τη διάρκεια της διπλωματικής. Τέλος, ευχαριστώ την οικογένεια μου, τον Γιώργο και τους φίλους μου που πάντα είναι δίπλα μου και με στηρίζουν.

Περίληψη

Η γεωργία ως συνιστώσα του πρωτογενή τομέα είναι απαραίτητη για τον άνθρωπο, διότι ο ίδιος εξαρτάται άμεσα από αυτήν για την παραγωγή τροφίμων. Επομένως, απαιτείται συνεχής προσφορά και παραγωγή γεωργικών αγαθών για να ικανοποιηθεί η ζήτηση του αυξανόμενου παγκόσμιου πληθυσμού. Για το λόγο αυτό, ολόκληρη η αλυσίδα του αγροδιατροφικού τομέα αντιμετωπίζει αυξανόμενες προκλήσεις, οι οποίες απαιτούν την εφαρμογή νέων καινοτόμων τεχνολογιών για τη βελτίωση της παραγωγικότητάς της με την παράλληλη διατήρηση της ποιότητας των παραγόμενων αγαθών. Τα φρούτα ως αντικείμενο μελέτης της παρούσας διπλωματικής είναι επίσης ένα βασικό γεωργικό αγαθό για την ανθρώπινη ζωή. Κατά συνέπεια, η έρευνα για την ανάπτυξη εφαρμογών για την ταξινόμηση, ανίχνευση και τον ποιοτικό έλεγχο των φρούτων είναι σημαντική για αρκετούς οικονομικούς τομείς, τόσο για τις αγορές χονδρικής και λιανικής, όσο και για τις βιομηχανίες μεταποίησης. Η όραση υπολογιστών και η εκμάθηση μηχανής είναι από τα πιο χρησιμοποιούμενα τεχνολογικά εργαλεία στον αγροβιομηχανικό τομέα, τόσο στην αυτόματη συγκομιδή φρούτων, στις μηχανές διαλογής φρούτων και στη σάρωση φρούτων στο εμπόριο. Η βαθιά εκμάθηση μηχανής (Deep Learning) ως τομέας της εκμάθησης μηχανής αποτελεί μια πρόσφατη, σύγχρονη τεχνική επεξεργασίας εικόνας και ανάλυσης δεδομένων, με πολλά υποσχόμενα αποτελέσματα και μεγάλες δυνατότητες. Καθώς η βαθιά εκμάθηση μηχανής έχει εφαρμοστεί ήδη με επιτυχία σε διάφορους τομείς, πρόσφατα έχει αρχίσει να χρησιμοποιείται επίσης στον τομέα της γεωργίας. Η ανίχνευση φρούτων είναι ένα κρίσιμο συστατικό για τον αυτοματισμό της γεωργίας σε συνθήκες χωραφιού. Με την ακριβή γνώση των μεμονωμένων θέσεων των φρούτων στο χωράφι, είναι δυνατή η εκτίμηση της σοδειάς και η χαρτογράφηση του, οι οποίες είναι σημαντικές διαδικασίες για τους καλλιεργητές καθώς διευκολύνουν την αποτελεσματική χρήση των πόρων (λίπανση, πότισμα)και με αυτό το τρόπο μπορεί να γίνει ζωνοποίηση της καλλιέργειας και να γίνονται στοχευμένες επεμβάσεις σε αυτή βελτιώνοντας την παραγωγικότητα ανά περιοχή και εξοικονομώντας πόρους. Επίσης ο ακριβής εντοπισμός των φρούτων είναι απαραίτητο συστατικό για την ανάπτυξη ενός αυτοματοποιημένου ρομποτικού συστήματος συγκομιδής, το οποίο μπορεί να βοηθήσει στην άμβλυνση των κοπιαστικών και εντατικών εργασιών σε έναν οπωρώνα.

Στο παραπάνω πλαίσιο εντάσσεται και το αντικείμενο της παρούσας διπλωματικής, όπου μελετάται το πρόβλημα της ταξινόμησης και ανίχνευσης τριών κατηγοριών φρούτων σε εικόνες RGB που περιέχουν διαφορετικά είδη φρούτων χρησιμοποιώντας μοντέλα βαθιάς εκμάθησης μηχανής (deep learning). Στόχος είναι να αποκτηθεί καλύτερη κατανόηση των δυνατοτήτων και των περιορισμών των τεχνικών βαθιάς εκμάθησης μηχανής καθώς και να διερευνήσουν το πώς οι αρχιτεκτονικές βαθιάς εκμάθησης μηχανής μπορούν να αποτελέσουν τμήμα μιας εφαρμογής αξιοποιήσιμης στην πραγματική ζωή. Η μεθοδολογία της παρούσας εργασίας μπορεί να χωριστεί σε δυο διακριτά μέρη της ταξινόμησης και της ανίχνευσης φρούτων με μοντέλα deep learning. Σε καθένα από αυτά τα μέρη επιλέχθηκαν αρχιτεκτονικές και μοντέλα deep learning από τη μελέτη της βιβλιογραφίας, όπου χρησιμοποιούνται για την επίλυση αντίστοιχων προβλημάτων και εφαρμογών. Για το τμήμα της ταξινόμησης επιλέχθηκαν να χρησιμοποιηθούν 3 μοντέλα MobileNet, Resnet50 και VGG16. Ενώ για το τμήμα της ανίχνευσης φρούτων επιλέχθηκαν οι αρχιτεκτονικές SSD, Faster R-CNN σε συνδυασμό με το μοντέλο ResNet50 και YOLO v3. Τα μοντέλα εκπαιδεύτηκαν και αξιολογήθηκαν σε ένα ελεύθερα διαθέσιμο σετ δεδομένων από τον ιστότοπο Kaggle. Τα πειράματα έγιναν με τη χρήση του TensorFlow, όπου είναι μια βιβλιοθήκη ανοιχτού κώδικα η οποία προσφέρει τη δυνατότητα τόσο της χρήση και της εκπαίδευσης έτοιμων μοντέλων εκμάθησης μηχανής όσο και τη δημιουργία νέων μοντέλων από τον χρήστη. Στη διαδικασία των πειραμάτων έγινε εμβάθυνση στις τεχνικές transfer learning και fine tuning, όπως και σε τεχνικές προεπεξεργασίας όπως η επαύξηση δεδομένων (data augmentation), όπου βοηθούν τα μοντέλα να έχουν καλύτερες αποδόσεις ειδικά αν τα δεδομένα εκπαίδευσης είναι περιορισμένα καθώς εξοικονομούν υπολογιστικούς πόρους. Τέλος συγκρίθηκαν οι αποδόσεις των μοντέλων τόσο στην ταξινόμηση όσο και στην ανίχνευση των εξεταζόμενων κατηγοριών φρούτων.

Abstract

Agriculture as a component of the primary sector is essential for man, because he is directly dependent on it for food production. Therefore, continuous supply and production of agricultural goods is required to meet the demand of the growing world population. For this reason, the entire agri-food sector is facing increasing challenges, which require the application of new innovative technologies to improve its productivity while maintaining the quality of the produced commodities. Fruit, as the subject of this dissertation, is also a basic agricultural product for human life. Consequently, research into application development for fruit sorting, detection and quality control is important for several economic sectors, concerning automatic fruit harvesting, fruit sorting machines and commercial fruit scanning. Computer vision and machine learning are among the most widely used technological tools in the agro-industrial sector, both in automatic fruit harvesting, machines of grading for fruits and commercial fruit scanning. Deep Learning, as a field of machine learning, is recent; it is a modern technique of image processing and data analysis, with many promising results and great potential. As deep learning has already been successfully applied in various fields, it has recently also started to be used in agriculture. Fruit tracking is a critical point for agricultural automation in field conditions. Knowing the accurate individual locations of the fruit in the field, it is possible to perform yield estimation and mapping. These procedures are significant for growers as they facilitate the efficient use of resources (fertilization, watering). In this way cultivators could zone their crops and do targeted interventions in it, improving productivity per area and saving resources. Additional, precise localization of the fruit is a necessary element of an automated robotic harvesting system, which can help mitigate one of the most labor tasks in an orchard.

The above context also includes the subject of this dissertation, which studies the problem of classification and detection of three categories of fruit in RGB images that contain different types of fruit using deep machine learning models. The aim is to gain a better understanding of the possibilities and limitations of deep learning techniques as well as to explore how deep machine learning architectures can be part of a real-life application. The methodology of the present work can be divided into two distinct parts of the classification and detection of fruits with deep learning models. In each of these parts, architectures and deep learning models were selected from the literature study, where they are used to face corresponding problems and applications. For the classification section, 3 models are used MobileNet, Resnet50 and VGG16 while for the object detection section SSD, Faster R-CNN and YOLO v3 are employed. The deep learning models were trained and evaluated in a freely available dataset of images from the Kaggle website. The dataset contains different type of fruits. The experiments were performed using TensorFlow, which is an open-source library that offers the possibility of both using and training ready-made machine learning models as well as creating new models by the user. During the training of the models, data augmentation, transfer learning and fine-tuning techniques were used, which are really help the models to have higher performances, especially if the training data is limited as they save computing resources. Finally, the accuracy of the classification and object detection models was evaluated using the testing part of the dataset.

Περιεχόμενα

1.Εισαγωγή

Στο κεφάλαιο αυτό γίνεται μια εισαγωγική αναφορά στις έννοιες που πραγματεύεται η παρούσα διπλωματική εργασία. Αναφέρονται στη συνέχεια το αντικείμενο της εργασίας, ο στόχος της και το κίνητρο για την ολοκλήρωση της. Τέλος παρουσιάζεται η δομή της εργασίας.

1.1 Γενικές Έννοιες

Η γεωργία ακριβείας μπορεί να οριστεί ως η εφαρμογή σύγχρονων τεχνολογιών για την παροχή, επεξεργασία και ανάλυση δεδομένων από πολλαπλές πηγές υψηλής χωρικής ανάλυσης και χρονικής συχνότητας με σκοπό τη λήψη αποφάσεων και τη διαχείριση λειτουργιών της φυτικής παραγωγής. Η εισαγωγή των συνελικτικών νευρωνικών δικτύων (CNNs) σε εφαρμογές στη γεωργία ήταν αναπόφευκτη λόγω της επιτυχίας των αποτελεσμάτων τους σε διάφορους τομείς. Οι αρχιτεκτονικές βαθιάς εκμάθησης μηχανής (Deep learning) είναι σύγχρονες τεχνικές για την επεξεργασία εικόνας και αποτελούν υποσύνολο του τομέα της τεχνητής νοημοσύνης ή machine learning που επιτρέπει τη δημιουργία μοντέλων ή προτύπων από ένα σύνολο δεδομένων σε ένα υπολογιστικό σύστημα. Τα μοντέλα που αντιπροσωπεύουν τη λογική της εκμάθησης μηχανής (Machine Learning), χρησιμοποιούνται από διάφορους επιστημονικούς κλάδους. Η λειτουργία τους στηρίζεται στη δημιουργία μοντέλων τα οποία σχηματίζονται από το χρήστη. Τα μοντέλα αυτά βελτιώνονται σταδιακά μέσα από επαναλαμβανόμενη εκπαίδευση, και είναι σε θέση να προβλέπουν με κάποιο ποσοστό επιτυχίας την κατηγορία στην οποία ανήκει οποιοδήποτε δείγμα τους δοθεί μετά το πέρας της διαδικασίας εκπαίδευσης. Τα δίκτυα CNN χρησιμοποιούνται στο τομέα της επεξεργασίας εικόνας τόσο για ταξινόμηση όσο και για ανίχνευση αντικειμένου από εικόνες.

Ένα δίκτυο CNN περιλαμβάνει σε σειρά ένα συνελικτικό επίπεδο, ένα επίπεδο που περιέχει μια μη γραμμική συνάρτηση ενεργοποίησης (activation function) και τέλος ένα pooling επίπεδο. Αυτή η δομή μπορεί να επαναλαμβάνεται μια ή περισσότερες φορές μέσα σε ένα μοντέλο ανάλογα με την αρχιτεκτονική του. Το συνελικτικό επίπεδο εκτελεί μια δισδιάστατη συνέλιξη για μια τρισδιάστατη εικόνα εισόδου και ένα τρισδιάστατο φίλτρο. Σε ένα μοντέλο CNN εκτελούνται διάφορες συνελίξεις σε επίπεδα του δικτύου, δημιουργώντας διαφορετικές αναπαραστάσεις του συνόλου δεδομένων εκπαίδευσης, ξεκινώντας από πιο γενικά χαρακτηριστικά στα πρώτα μεγαλύτερα επίπεδα, και προχωρώντας στα πιο λεπτομερή στα μετέπειτα βαθύτερα επίπεδα. Τα συνελικτικά επίπεδα λειτουργούν ως εξαγωγείς χαρακτηριστικών από τις εικόνες εισόδου των οποίων η διάσταση μειώνεται στη συνέχεια από τα pooling επίπεδα. Τα συνελικτικά επίπεδα κωδικοποιούν πολλαπλά μικρά χαρακτηριστικά σε διακριτικά χαρακτηριστικά. Ουσιαστικά εφαρμόζουν συνελικτικά φίλτρα στην εικόνα εισόδου με σκοπό να βρουν και να εξαγάγουν συγκεκριμένα μοτίβα. Τα πλήρως συνδεδεμένα επίπεδα, που είναι τοποθετημένα σε πολλές περιπτώσεις κοντά στην έξοδο του μοντέλου, λειτουργούν ως ταξινομητές που εκμεταλλεύονται τις δυνατότητες υψηλού επιπέδου που εκπαιδεύτηκε να ταξινομεί τις εικόνες εισόδου σε προκαθορισμένες κατηγορίες ή να κάνει αριθμητικές προβλέψεις. Παίρνουν ένα διάνυσμα ως είσοδο και παράγουν ένα άλλο διάνυσμα ως έξοδο.

Οι τρέχοντες κύριοι αλγόριθμοι ανίχνευσης αντικειμένων αποτελούνται κυρίως από δύο κύριες κατηγορίες, αλγόριθμους ανίχνευσης αντικειμένων ενός σταδίου και αλγόριθμους ανίχνευσης αντικειμένων δύο σταδίων, όπου και οι δύο βασίζονται σε μεθόδους βαθιάς μηχανικής μάθησης. Η κύρια διάκριση μεταξύ αυτών των δύο μεθόδων είναι εάν θα δημιουργηθεί μια πρόταση περιοχής. Οι αλγόριθμοι ανίχνευσης αντικειμένων ενός σταδίου δεν χρειάζεται να δημιουργήσουν μια πρόταση περιοχής. Μπορούν να λάβουν άμεσα την ακρίβεια ταξινόμησης του αντικειμένου και τη θέση συντεταγμένων του. Οι αλγόριθμοι ανίχνευσης αντικειμένων δύο σταδίων πρέπει να δημιουργήσουν μια πρόταση περιοχής πριν γίνει η ταξινόμηση και η ανίχνευση. Σε γενικές γραμμές, οι αλγόριθμοι ενός σταδίου έχουν πλεονέκτημα ταχύτητας και οι αλγόριθμοι δύο σταδίων έχουν πλεονεκτήματα στην ακρίβεια.

Transfer learning είναι η διαδικασία που βασίζεται σε ένα προϋπάρχον εκπαιδευμένο δίκτυο βαθιάς εκμάθησης μηχανής, το οποίο προσαρμόζεται για να εκπαιδευτεί σε ένα νέο σετ δεδομένων. Η δημιουργία ενός νέου δικτύου CNN, είναι μια αρκετά χρονοβόρα διαδικασία που απαιτεί τον εκ νέου σχεδιασμό της αρχιτεκτονικής του δικτύου και την εκπαίδευση αυτού από την αρχή για να επιτευχθεί η βέλτιστη διαμόρφωση του δικτύου. Για το λόγο αυτό,

μπορεί κανείς να εκμεταλλευτεί ένα προεκπαιδευμένο δίκτυο και να το επανεκπαιδεύσει σε νέα μοτίβα και δεδομένα. Εξάλλου, είναι χρήσιμο όταν δεν έχει κανείς αρκετά δεδομένα για να εκπαιδεύσει το δίκτυο, με αυτή τη διαδικασία το δίκτυο μπορεί να εκπαιδευτεί σε λιγότερα δεδομένα. Έτσι, χρησιμοποιείται ένα προεκπαιδευμένο μοντέλο σε ένα κατάλληλο σύνολο δεδομένων για την εκάστοτε εργασία. Η βασική ιδέα είναι η διατήρηση ορισμένων επιπέδων ενός προεκπαιδευμένου δικτύου και η προσαρμογή συνήθως των επιπέδων εισόδου και εξόδου.
Με βάση τα παραπάνω αποφασίστηκε στη παρούσα διπλωματική εργασία να εξεταστεί η αποδοτικότητα των αρχιτεκτονικών βαθιάς μηχανικής μάθησης εφαρμόζοντας μοντέλα ταξινόμησης και ανίχνευσης αντικειμένου για την ταξινόμηση εικόνων RGB και την ανίχνευση από εικόνες 3 κατηγοριών φρούτων (μήλο, μπανάνα, πορτοκάλι). Στο μέρος της ταξινόμησης χρησιμοποιήθηκαν τα μοντέλα MobileNet, Resnet50 και VGG16 και αξιοποιήθηκαν οι τεχνικές του transfer learning, fine tuning και ως μέθοδος προ-επεξεργασίας των δεδομένων η επαύξηση δεδομένων. Ενώ για το τμήμα της ανίχνευσης φρούτων επιλέχθηκαν οι αρχιτεκτονικές SSD, Faster R-CNN σε συνδυασμό με το μοντέλο ResNet50 και YOLO v3 δηλαδή τεχνικές τόσο ενός όσο και 2 σταδίων. Τα μοντέλα εκπαιδεύτηκαν και αξιολογήθηκαν σε ένα ελεύθερα διαθέσιμο σετ δεδομένων από τον ιστότοπο Kaggle. Μετά την εκπαίδευση των μοντέλων στο dataset των εικόνων αξιολογήθηκαν τα αποτελέσματα τους.

1.2 Αντικείμενο και Στόχος

Κύριο αντικείμενο της παρούσας διπλωματικής εργασίας ήταν η επιλογή, η εφαρμογή και η αξιολόγηση μοντέλων ταξινόμησης και ανίχνευσης αντικειμένων στηριζόμενα σε αρχιτεκτονικές βαθιάς εκμάθησης μηχανής, για τον εντοπισμό τριών κατηγοριών φρούτων από εικόνες RGB. Ειδικότερα, αξιοποιήθηκαν τρία μοντέλα συνελικτικών νευρωνικών δικτύων με σκοπό την ταξινόμηση εικόνων RGB που περιείχαν τις τρεις κατηγορίες φρούτων. Επιπλέον για το κομμάτι της ανίχνευσης αντικειμένων αξιοποιήθηκαν τρία διαφορετικά μοντέλα με σκοπό να εντοπιστούν τα φρούτα πάνω στις εικόνες τοποθετώντας πλαίσια οριοθέτησης.

Η παρούσα εργασία έχει σκοπό να διερευνήσει τεχνικές βαθιάς εκμάθησης μηχανής (Deep Learning) για την ανίχνευση φρούτων από εικόνες και την ταξινόμηση εικόνων με φρούτα. Στόχος είναι να αποκτηθεί καλύτερη κατανόηση των δυνατοτήτων και των περιορισμών των τεχνικών βαθιάς εκμάθησης μηχανής, να γίνει εμβάθυνση στη λειτουργία τους. Παράλληλα μελετήθηκαν τεχνικές που χρησιμοποιούνται για την αύξηση της ακρίβειας των μοντέλων CNN και διευκολύνουν τη χρονοβόρα διαδικασία εκπαίδευσης τους.

1.3 Κίνητρο

Η αλυσίδα του αγροδιατροφικού τομέα αντιμετωπίζει αυξανόμενες προκλήσεις, οι οποίες απαιτούν την εφαρμογή νέων καινοτόμων τεχνολογιών για τη βελτίωση της παραγωγικότητάς της με την παράλληλη διατήρηση της ποιότητας των παραγόμενων αγαθών. Η Όραση Υπολογιστών και η Μηχανική Μάθηση είναι από τα πιο χρησιμοποιούμενα τεχνολογικά εργαλεία στον αγροβιομηχανικό τομέα. Οι αρχιτεκτονικές βαθιάς μηχανικής μάθησης (Deep learning) είναι σύγχρονες τεχνικές για την επεξεργασία εικόνας και αποτελούν υποσύνολο του τομέα της μηχανικής μάθησης. Τα συνελικτικά νευρωνικά δίκτυα (CNNs) αποτελούν την πιο διαδεδομένη μορφή δικτύων Βαθιάς Μάθησης. Η εισαγωγή των Συνελικτικών Νευρωνικών Δικτύων (CNNs) σε εφαρμογές στη γεωργία ήταν αναπόφευκτη λόγω της μεγάλης αποτελεσματικότητας που σημείωσαν σε διάφορους τομείς. Βάσει των παραπάνω, η παρούσα διπλωματική έχει ως κίνητρο την ανάδειξη των δυνατοτήτων και των περιορισμών των δικτύων CNNs για την ταξινόμηση και εντοπισμό φρούτων σε εικόνες RGB.

1.4 Δομή Εργασίας

Η παρούσα εργασία οργανώνεται σε πέντε κεφάλαια:

Στο κεφάλαιο 2 γίνεται ανασκόπηση της βιβλιογραφίας. Αρχικά αναλύεται το γενικό πλαίσιο στο όποιο εντάσσεται η παρούσα εργασία και το θεωρητικό υπόβαθρο αυτής. Στην συνέχεια παρουσιάζονται τα κίνητρα και οι περιορισμοί στην ανίχνευση φρούτων από εικόνα. Τέλος παρουσιάζονται δημοσιεύσεις από τη διεθνή βιβλιογραφία που βοήθησαν στη διεξαγωγή της.

Στο κεφάλαιο 3 αναπτύσσεται η μεθοδολογία που εφαρμόστηκε στην πειραματική διαδικασία. Παράλληλα, γίνεται ανάλυση του συνόλου των δεδομένων που χρησιμοποιήθηκαν, των μοντέλων που αξιοποιήθηκαν, των επιμέρους βημάτων της διαδικασίας, καθώς και των παραμέτρων που επιλέχθηκαν.

Στο κεφάλαιο 4 παρουσιάζονται τα αποτελέσματα της ταξινόμησης και της ανίχνευσης αντικειμένου, ενώ γίνονται σχετικές παρατηρήσεις και προτάσεις που αποσκοπούν στην βελτίωση της ταξινόμησης.

Στο κεφάλαιο 5 αναφέρονται τα συμπεράσματα που προκύπτουν από την εργασία ενώ προτείνονται μελλοντικές βελτιώσεις και προοπτικές για περαιτέρω εξέλιξη και έρευνα.

2. Ανασκόπηση Βιβλιογραφίας

Στο παρόν κεφάλαιο αναλύεται το θεωρητικό υπόβαθρο και οι βασικές έννοιες πάνω στις οποίες στηρίζεται η εργασία. Πιο συγκεκριμένα τίθεται το γενικό πλαίσιο της διπλωματικής. Στη συνέχεια γίνεται προσέγγιση του θεωρητικού υποβάθρου πάνω στο οποίο βασίζεται το αντικείμενο της εργασίας δηλαδή τα συνελικτικά νευρωνικά δίκτυα. Αναλύεται η λειτουργία τους και οι τεχνικές που χρησιμοποιούνται τόσο για ταξινόμηση όσο και για αναγνώριση αντικειμένων. Παρουσιάζονται κίνητρα και περιορισμοί που υπάρχουν για την αναγνώριση φρούτων από εικόνες. Τέλος για όλα τα παραπάνω παρατίθενται δημοσιεύσεις από την διεθνή βιβλιογραφία που σχετίζονται με το αντικείμενο της εργασίας.

2.1 Γενικό Πλαίσιο

Η γεωργία συνιστά τμήμα του πρωτογενή τομέα, ενώ η εξέλιξη της και οι τεχνικές που χρησιμοποιούνται στο γεωργικό τομέα είναι κρίσιμα στοιχεία για την επισιτιστική ασφάλεια παγκοσμίως. Με την εξέλιξη της τεχνολογίας έχουν εισαχθεί νέες πρακτικές καλλιέργειας, παρακολούθησης και ανάλυσης των γεωργικών οικοσυστημάτων, που έχουν ως σκοπό να εφαρμόσουν διαφορετικά επίπεδα εισροών σε επιμέρους περιοχές του αγρού ανάλογα με το δυναμικό παραγωγής και τις εδαφοκλιματικές συνθήκες (παραλλακτικότητα). Αυτό το νέο σύστημα διαχείρισης των καλλιεργειών ονομάζεται Γεωργία Ακριβείας (Precision Agriculture). Η «έξυπνη» γεωργία ή γεωργία ακριβείας είναι σημαντική για την αντιμετώπιση των προκλήσεων της γεωργικής παραγωγής από άποψη παραγωγικότητας, περιβαλλοντικών επιπτώσεων, επισιτιστικής ασφάλειας και βιωσιμότητας. Καθώς ο παγκόσμιος πληθυσμός αυξάνεται συνεχώς, πρέπει να επιτευχθεί αύξηση της παραγωγής τροφίμων (FAO,2009), διατηρώντας παράλληλα την επάρκεια τους αλλά και την υψηλή θρεπτική τους ποιότητα σε όλο τον κόσμο, προστατεύοντας τα φυσικά οικοσυστήματα χρησιμοποιώντας βιώσιμες γεωργικές διαδικασίες. Για την επίτευξη αυτών των στόχων, που είναι περίπλοκοι και πολυπαραγοντικοί, χρειάζεται τα γεωργικά οικοσυστήματα να κατανοηθούν καλύτερα μέσω παρακολούθησης, μέτρησης και ανάλυσης διαφόρων φυσικών χαρακτηριστικών και φαινομένων. Αυτό προϋποθέτει την ανάλυση μεγάλου όγκου γεωργικών δεδομένων και τη χρήση νέων πληροφοριακών τεχνολογιών. Η συλλογή και επεξεργασία εικόνων αποτελεί σε πολλές περιπτώσεις μέθοδο που εξυπηρετεί τους παραπάνω στόχους για αυτό και είναι ιδιαίτερα δημοφιλής στον τομέα της γεωργίας ακριβείας. Δημοφιλείς εφαρμογές που καταγράφονται στη διεθνή βιβλιογραφία είναι η χαρτογράφηση και η παρακολούθηση καλλιεργειών, η εκτίμηση παραγωγικότητας, η ανίχνευση υδατικού στρες, η ανίχνευση ζιζανίων, η παρακολούθηση θερμοκηπίων κ.α. Ανάλογα με το φαινόμενο που μελετάται, χρησιμοποιούνται εικόνες από διαφορετικούς δέκτες για ευρείας κλίμακας παρατήρηση καλλιεργειών, όπως λήψεις από δορυφόρους, αεροπλάνα και μη επανδρωμένα αεροσκάφη, ενώ για παρατήρηση προβλημάτων μικρότερης κλίμακας χρησιμοποιούνται δέκτες χειρός ή αισθητήρες τοποθετημένοι πάνω σε ειδικά διαμορφωμένες κατασκευές. Η λήψη και επεξεργασία εικόνων έχει πολλά πλεονεκτήματα όταν εφαρμόζεται στη γεωργία, είναι μη καταστροφική μέθοδος συλλογής πληροφοριών όσον αφορά τα χαρακτηριστικά γης, ενώ τα δεδομένα μπορούν να ληφθούν συστηματικά είτε σε μεγάλες είτε σε μικρές γεωγραφικές περιοχές. Ως εκ τούτου, η επεξεργασία εικόνας είναι ένας σημαντικός ερευνητικός τομέας στη γεωργία στον οποίο εφαρμόζονται τεχνικές ανάλυσης δεδομένων για αναγνώριση / ταξινόμηση εικόνας, ανίχνευση ανωμαλιών κ.λπ., σε διάφορες γεωργικές εφαρμογές.Οι πιο συχνά χρησιμοποιούμενες τεχνικές για την ανάλυση εικόνων είναι η εκμάθηση μηχανής (machine learning- ML), ο K-means, η υποστήριξη διανυσματικών μηχανών (SVM), τα τεχνητά νευρωνικά δίκτυα (ANN), τα φίλτρα εικόνας, οι δείκτες βλάστησης και η ανάλυση παλινδρόμησης. Συχνά στη διεθνή βιβλιογραφία διεξάγονται συγκρίσεις ανάμεσα στις διαφορετικές τεχνικές επεξεργασίας εικόνας και όρασης υπολογιστών, ώστε να εκτιμηθεί η ακρίβεια τους και ποια εξυπηρετεί επαρκέστερα τους στόχους κάθε ερευνητικής προσπάθειας. Σύμφωνα με τους *Kamilaris et al., 2018* , τα μοντέλα βαθιάς μηχανικής μάθησης υπερισχύουν έναντι άλλων τεχνικών επεξεργασίας εικόνας σε γεωργικές εφαρμογές. Στη συγκεκριμένη δημοσίευση αναφέρονται μεμονωμένες ερευνητικές προσπάθειες όπου έχει διεξαχθεί σύγκριση ανάμεσα σε DL με άλλες τεχνικές που χρησιμοποιούνται για την επίλυση του υπό μελέτη προβλήματός πάνω στο ίδιο σύνολο δεδομένων. Τα συνελικτικά νευρωνικά δίκτυα παρουσίασαν κατά 5% υψηλότερη ακρίβεια ταξινόμησης

έναντι των SVM, των απλών νευρωνικών δικτύων, των μη επιβλεπόμενων τεχνικών μάθησης χαρακτηριστικών και των μοντέλων παλινδρόμησης βασισμένα στην υφή και διαφόρων ταξινομητών όπως των LMC, Naïve-Bayes.
Μέσα από δημοσιεύσεις ανασκόπησης όπως των Kamilaris et al. (2018), Santos Luís et al. (2020) και Naranjo-Torres et al. (2020), όπου όλες μελετούν τις εφαρμογές των νευρωνικών συνελικτικών δικτύων στη γεωργία με την τελευταία συγκεκριμένα να αναφέρει ερευνητικές προσπάθειες που το αντικείμενο μελέτης είναι η ανίχνευση φρούτων από εικόνες ή ταξινόμηση εικόνων με φρούτα, προκύπτει ότι τα πιο δημοφιλή πεδία χρήσης των CNN είναι:Η αναγνώριση ζιζανίων

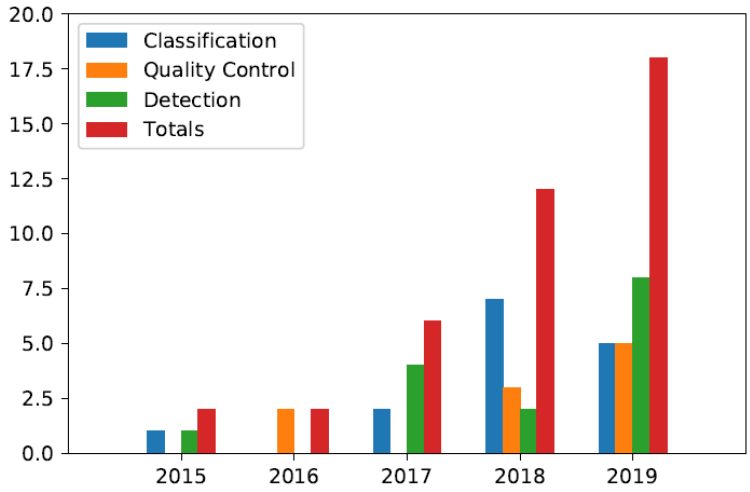
- Η ταξινόμηση κατηγοριών κάλυψης γης
- Η αναγνώριση φυτών
- Η καταμέτρηση φρούτων
- Η ταξινόμηση του είδους των καλλιεργειών

Επίσης μέσα από αυτές τις συγκεντρωτικές δημοσιεύσεις μπορεί κάποιος να συμπεράνει ποιες αρχιτεκτονικές και ποια μοντέλα βαθιάς μηχανικής μάθησης είναι τα πιο συχνά χρησιμοποιούμενα. Αξίζει να σημειωθεί ότι σχεδόν όλα αυτά τα μοντέλα ακολουθούνται από τα βάρη τους είναι δηλαδή προεκπαιδευμένα, πράγμα που σημαίνει ότι το δίκτυό τους είχε ήδη εκπαιδευτεί από κάποιο σύνολο δεδομένων και έτσι έχει τη δυνατότητα να παρέχει ακριβή ταξινόμηση για κάποιο συγκεκριμένο πρόβλημα. Τέλος αναδεικνύονται τα πιο δημοφιλή περιβάλλοντα, βιβλιοθήκες, frameworks όπου είναι κατάλληλα για την ανάπτυξη ή τη χρήση μοντέλων μηχανικής μάθησης αφού τα περισσότερα ενσωματώνουν τέτοιου είδους μοντέλα. Στο παρακάτω πίνακα συγκεντρώνονται τα πιο δημοφιλή μοντέλα & αρχιτεκτονικές DL, σύνολα δεδομένων και Frameworks που συναντήθηκαν στις διεθνείς δημοσιεύσεις:

Αρχιτεκτονικές DL	**Μοντέλα DL**	**Σύνολα Δεδομένων**	**Frameworks/Platforms**
CNN	VGG	ImageNet	Tensorflow/Keras
FCN	ResNet	Pascal VOC	Theano/ Keras
Faster R-CNN	GoogleNet	COCO dataset	Py-Torch
SSD	YoLo	Fruits-360	Darknet
R-FCN	AlexNet	VegFru	Caffe
RNN	MobileNet	Supermarket Data	TF Learn

Πίνακας 1: Δημοφιλή Μοντέλα DL, Αρχιτεκτονικές, Datasets και Frameworks

Η παρούσα διπλωματική εργασία εντάσσεται στο παραπάνω πλαίσιο αλλά επικεντρώνεται στην ανίχνευση και ταξινόμηση φρούτων σε RGB εικόνες με CNN μοντέλα. Μεταξύ των εφαρμογών που αναφέρθηκαν και παραπάνω, έχουν υιοθετηθεί τεχνικές της όρασης υπολογιστών και για εφαρμογές στην αναγνώριση φρούτων και στον αποτελεσματικό εντοπισμό συγκεκριμένων ελαττωμάτων τους, τόσο σε αγορές χονδρικής όσο και λιανικής. Η όραση υπολογιστών είναι ένα από τα πιο χρησιμοποιούμενα τεχνολογικά εργαλεία στον αγροτοβιομηχανικό τομέα, τόσο σε αυτόματη συγκομιδή φρούτων, μηχανήματα διαλογής φρούτων και σάρωση φρούτων. Στη μελέτη των Naranjo-Torres et al. (2020) προσδιορίστηκαν τρεις βασικοί τομείς στις εφαρμογές σχετικά με τα φρούτα. Ο πρώτος είναι η ταξινόμηση των φρούτων, διαδικασία όπου τα φρούτα ταξινομούνται ανάλογα με τον τύπο τους σε εφαρμογές για λιανική και χονδρική πώληση. Ο δεύτερος είναι ο ποιοτικός έλεγχος των φρούτων, ο οποίος χρησιμοποιείται σε εφαρμογές για τον εντοπισμό εσωτερικών και εξωτερικών ανωμαλιών στο σχήμα των φρούτων, του βαθμού ωριμότητάς τους, καθώς και σε ανίχνευση έλλειψης θρεπτικών συστατικών ή ασθενειών. Η τρίτη περιοχή μελέτης που προσδιορίστηκε είναι η ανίχνευση φρούτων, η οποία εφαρμόζεται για τη συγκομιδή των καρπών στα περιβόλια και επίσης στον προσδιορισμό της θέσης τους για αυτοματοποίηση της συγκομιδής συνολικά. Στο Σχήμα 1 επιβεβαιώνεται αυτό που αναφέρεται και παραπάνω δηλαδή οι 3 βασικοί τομείς όπου αναπτύσσονται εφαρμογές με CNN μοντέλα στον αγροβιομηχανικό τομέα σχετικά με τα φρούτα.

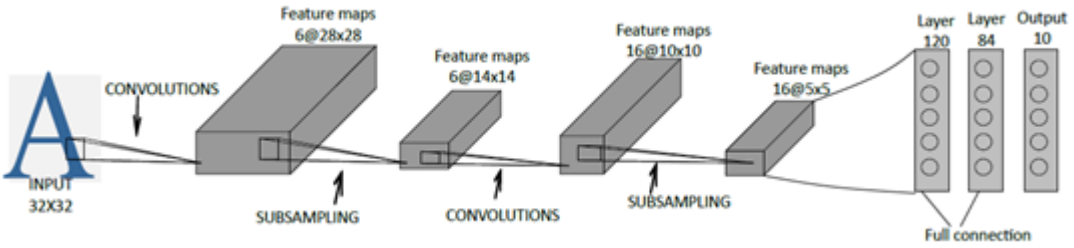


Σχήμα 1: Δημοσιεύσεις με CNN μοντέλα για ταξινόμηση, έλεγχο ποιότητας και εντοπισμό φρούτων

2.2 Υπόβαθρο σχετικά με τα CNNs για ταξινόμηση και ανίχνευση αντικειμένου

Η βαθιά μηχανική μάθηση (deep learning/ DL) είναι μέρος των μεθόδων μηχανικής μάθησης (machine learning) και είναι βασισμένη στα τεχνητά νευρωνικά δίκτυα (ANN, Artificial Neural Networks). Το DL έχει ποικίλες εφαρμογές από τη αναγνώριση της φυσικής ομιλίας μέχρι την επεξεργασία εικόνας. Διαφορετικές αρχιτεκτονικές βαθιάς μηχανικής μάθησης είναι τα: Deep Neural Networks (DNN), Deep Belief Networks (DBN), Recurrent Neural Networks (RNN), recursive neural networks, Fully Convolutional Networks (FCN) και Convolutional Neural Networks (CNN) που έχουν εφαρμοστεί με επιτυχία σε διάφορους ερευνητικούς τομείς, συμπεριλαμβανομένων της γεωργίας ακριβείας.

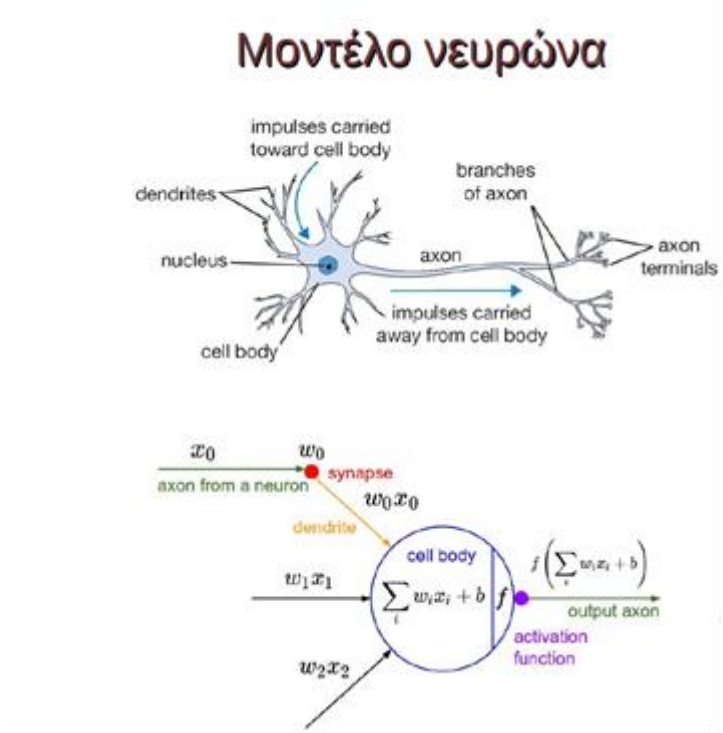
Τα τεχνητά νευρωνικά είναι δίκτυα με πολλαπλά επίπεδα. Τα πολυεπίπεδα δίκτυα μπορούν να εκπαιδευτούν σε πολύπλοκα και υψηλών διαστάσεων μοτίβα από μεγάλα σύνολα δεδομένων, ενώ αυτή τους η ιδιότητα τα θέτει κύριους υποψηφίους για εργασίες στο πεδίο της επεξεργασίας εικόνας όπως η ταξινόμηση και η αναγνώριση αντικειμένου. Ειδικότερα, τα συνελικτικά νευρωνικά δίκτυα είναι ένα είδος πολυστρωματικού νευρωνικού δικτύου, το οποίο προτάθηκε αρχικά από τους LeCun et al. (1998) και βρίσκει μέχρι και σήμερα αρκετές πρακτικές εφαρμογές. Η εικόνα 1 δείχνει την αρχική αρχιτεκτονική του πρώτου Μοντέλο CNN, που ονομάζεται LeNet-5. Τα CNN απέκτησαν μεγάλη δημοτικότητα όταν το μοντέλο AlexNet κέρδισε στο διαγωνισμός ImageNet (ILSVRC) το 2012.



Εικόνα 1: Αναπαράσταση της αρχιτεκτονικής του μοντέλου LetNet-5

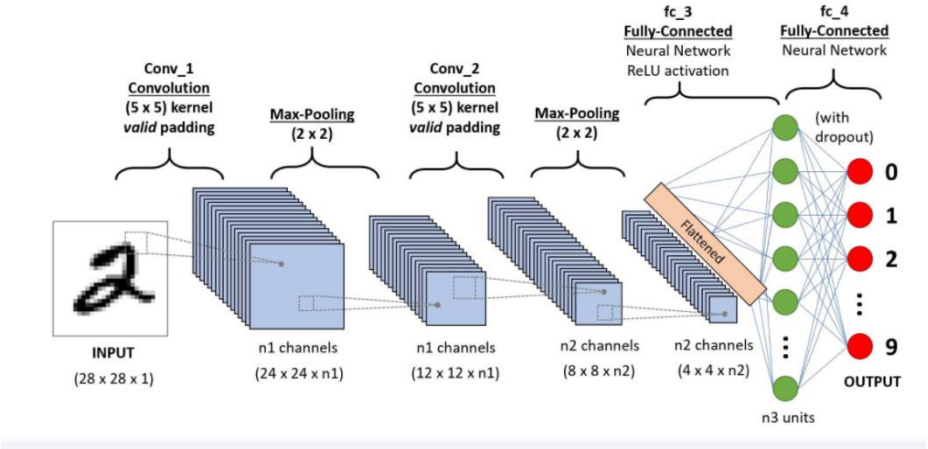
Στη παρούσα εργασία θα εξεταστούν τα συνελικτικά νευρωνικά δίκτυα (CNN) στην ταξινόμηση εικόνας και στην ανίχνευση αντικειμένου σε RGB εικόνες, τα οποία θεωρούνται η τρέχουσα τάση για τη δημιουργία μοντέλων

υπολογιστικής όρασης. Γενικά, τα τεχνητά νευρωνικά δίκτυα έχουν σκοπό να προσομοιάσουν υπολογιστικά το νευρικό σύστημα του ανθρώπινου εγκεφάλου.



Εικόνα 2: Αντιπαραβολή ανθρώπινου νευρώνα με την βασική αρχιτεκτονική ενός τεχνητού νευρωνικού δικτύου

Τα συνελικτικά νευρωνικά δίκτυα έχουν την ικανότητα να αναγνωρίζουν πρότυπα πάνω στις εικόνες. Για να καταλάβει κανείς αυτή τους την ικανότητα πρέπει να μελετήσει τη βασική αρχιτεκτονική που εμφανίζουν τα μοντέλα των CNN και τα δομικά στοιχεία που τα απαρτίζουν. Παρακάτω, θα αναλυθεί μια απλή αρχιτεκτονική CNN, όπως της Εικόνας 3 για να γίνει κατανοητό ποια είναι τα βασικά στοιχεία όλων των CNN.



Εικόνα 3: Αναπαράσταση απλής αρχιτεκτονικής νευρωνικού δικτύου

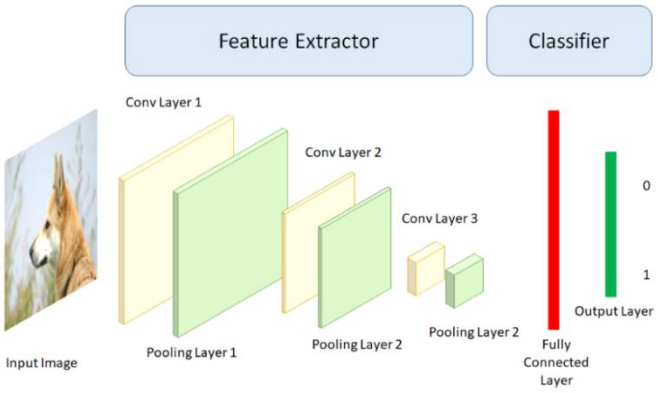
Ένα CNN χρησιμοποιεί ως δεδομένο εισόδου μια εικόνα αποσκοπώντας στην εξαγωγή συγκεκριμένων ανά περίπτωση χαρακτηριστικών από αυτήν. Τα συνήθη επίπεδα που συναντάμε σε ένα CNN είναι τα εξής:

Συνελικτικό επίπεδο: Τα συνελικτικά επίπεδα είναι στην ουσία φίλτρα εικόνας δηλαδή πίνακες βαρών *nxn* διαστάσεων που εφαρμόζονται πάνω στην εικόνα εισόδου με σκοπό να εξαχθούν μορφολογικά χαρακτηριστικά από την εικόνα όπως, ακμές, γωνίες ή βασικά σχήματα όπως ο κύκλος κλπ. Η συνέλιξη γενικά είναι ένα φίλτρο που περνά πάνω από την εικόνα, την επεξεργάζεται και εξάγει κοινά χαρακτηριστικά από την εικόνα π.χ. οριζόντιες ακμές.

Συναθροιστικό επίπεδο / Pooling: Τα συνελικτικά επίπεδα ακολουθούνται, από τα συναθροιστικά επίπεδα τα οποία στόχο έχουν να τονίσουν τα χαρακτηριστικά που ανίχνευσαν προηγουμένως τα συνελικτικά επίπεδα. Στην ουσία αποτελούν συνήθως φίλτρα μέγιστης τιμής (max pooling) ή μέσης τιμής (avg pooling) , που εφαρμόζονται στην εικόνα. Π.χ. στο παραπάνω παράδειγμα της εικόνας τα max-pooling layer 2x2 θα εφαρμόζονται πάνω στην εικόνα που εξάγεται από το συνελικτικό επίπεδο και σε περιοχές της εικόνας 2x2 θα κρατούν την μέγιστη τιμή.

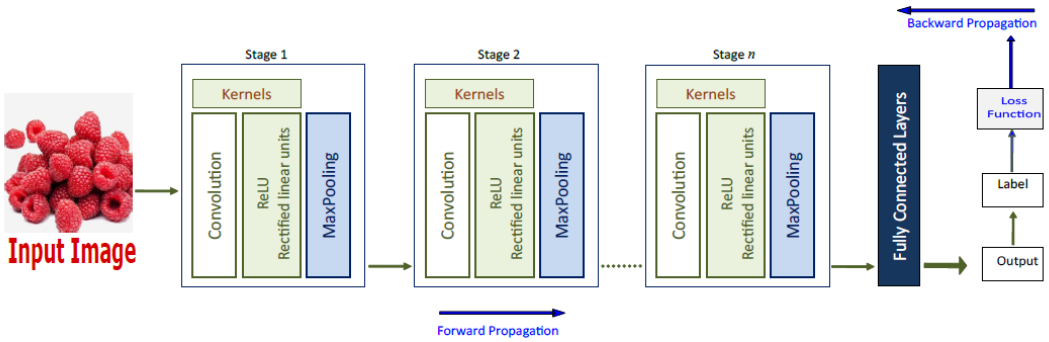
Συνάρτηση Ενεργοποίησης: Η συνάρτηση ενεργοποίησης είναι ένας κόμβος που τοποθετείται στο τέλος ή μεταξύ των επιπέδων των νευρωνικών δικτύων. Βοηθά να αποφασιστεί εάν ο νευρώνας θα συμμετέχει ή όχι. Υπάρχουν διαφορετικοί τύποι συναρτήσεων ενεργοποίησης π.χ. στην εικόνα 3 εφαρμόζεται η (ReLU) στο τέλος του δικτύου. Άλλα είδη συναρτήσεων είναι η σιγμοειδής ή εφαπτομενική, η επιλογή της συνάρτησης γίνεται ανάλογα το πρόβλημα και ποιο είδος το περιγράφει καλύτερα.

Τα 3 παραπάνω στοιχεία είναι τα βασικά στοιχεία κάθε συνελικτικού νευρωνικού δικτύου. Ανάλογα με την αρχιτεκτονική του μοντέλου που επιλέγεται αλλάζει το πλήθος των επιπέδων, τα μεγέθη των επιπέδων, καθώς και τα είδη των συναρτήσεων που εφαρμόζονται σε κάθε επίπεδο. Ανακεφαλαιώνοντας, τα συνελικτικά επίπεδα λειτουργούν ως εξαγωγείς χαρακτηριστικών για τις εικόνες εισόδου, των οποίων οι διαστάσεις μειώνονται από τα επίπεδα συνάθροισης (pooling). Τα χαρακτηριστικά που εξάγονται είναι πιο γενικά στα πρώτα επίπεδα αλλά στα "βαθύτερα" επίπεδα γίνονται πιο ολοκληρωμένα και λεπτομερή. Τα πλήρως συνδεδεμένα επίπεδα δρουν ως ταξινομητές, αξιοποιώντας χαρακτηριστικά υψηλού επιπέδου για την ταξινόμηση των εικόνων εισόδου στην αντίστοιχη κλάση. Άρα η γενική αρχιτεκτονική των CNN αποτελείται από 2 τμήματα, η αλληλουχία επιπέδων conv-pool δημιουργούν το $1^{ο}$ τμήμα του μοντέλου που εξάγει χαρακτηριστικά, ενώ τα πλήρως συνδεδεμένα επίπεδα λειτουργούν ως ταξινομητές αυτών των χαρακτηριστικών και είναι το $2^{ο}$ τμήμα του μοντέλου.



Εικόνα 4: Γενική αρχιτεκτονική των CNN

Στη δημοσίευση των Naranjo-Torres et al. (2020) εξηγείται πιο αναλυτικά η αρχιτεκτονική των CNNs. Σε αντίθεση με τα παραδοσιακά νευρωνικά δίκτυα, τα CNN χρησιμοποιούν τη συνέλιξη σε τουλάχιστον ένα από τα επίπεδα τους. Η αρχιτεκτονική του CNN περιλαμβάνει πολλαπλά στάδια ή μπλοκ που αποτελούνται από τέσσερα κύρια στοιχεία: μια «τράπεζα» φίλτρων που ονομάζονται πυρήνες, ένα επίπεδο συνέλιξης, μια μη γραμμική συνάρτηση ενεργοποίησης ένα αθροιστικό επίπεδο και ένα επίπεδο υποδειγματοληψίας. Κάθε στάδιο στοχεύει να αναγνωρίσει και αναπαραστήσει χαρακτηριστικά μέσω ενός συνόλου πινάκων που ονομάζονται χάρτες χαρακτηριστικών. Στην εικόνα 5 απεικονίζεται μια τυπική αρχιτεκτονική CNN που αποτελείται από μια σειρά από συνελικτικά στάδια και ένα ή περισσότερα πλήρως συνδεδεμένα επίπεδα, τα οποία δίνουν το τελικό αποτέλεσμα της ταξινόμησης.



Εικόνα 5: Τυπική αρχιτεκτονική ενός CNN

Στη συνέχεια παρουσιάζονται τα κύρια στοιχεία μιας τυπικής αρχιτεκτονικής CNN.

Σετ φίλτρων ή πυρήνες: Κάθε φίλτρο ή πυρήνας στοχεύει να ανιχνεύσει ένα συγκεκριμένο χαρακτηριστικό σε κάθε θέση του δεδομένου εισόδου, επομένως, η χωρική μετάφραση του δεδομένου εισόδου (εικόνα) από ένα επίπεδο ανίχνευσης θα μεταφέρεται στο επίπεδο εξόδου χωρίς αλλαγές. Όπως ορίζεται από τον *LeCun et al. (2010)*, υπάρχει μια "τράπεζα" m_1 φίλτρων σε κάθε συνελικτικό επίπεδο και το αποτέλεσμα $Y_i^{(l)}$ του i^{th} επιπέδου το οποίο αποτελείται από $m_1^{(l)}$ χάρτες χαρακτηριστικών μεγέθους $m_2^{(l)}$ x $m_3^{(l)}$. Ο χάρτης χαρακτηριστικών i^{th} υπολογίζεται ως εξής:

$$Y_i^{(l)} = B_i^{(l)} + \sum_{j=1}^{m_1^{(l-1)}} K_{ij}^{(l)} * Y_j^{(l-1)}$$

Εξίσωση 1

όπου $B_i^{(l)}$ δηλώνει τον πίνακα παραμέτρων εκπαίδευσης, $K_{ij}^{(l)}$ είναι το φίλτρο με διαστάσεις ($2h_1^{(l)+1}$ x $2h_2^{(l)+1}$) που συνδέουν τον j^{th} χάρτη χαρακτηριστικών του επιπέδου (l - 1) με τον i^{th} χάρτη χαρακτηριστικών του του επιπέδου (l) και το (*) είναι ο 2D τελεστής διακριτής συνέλιξης.

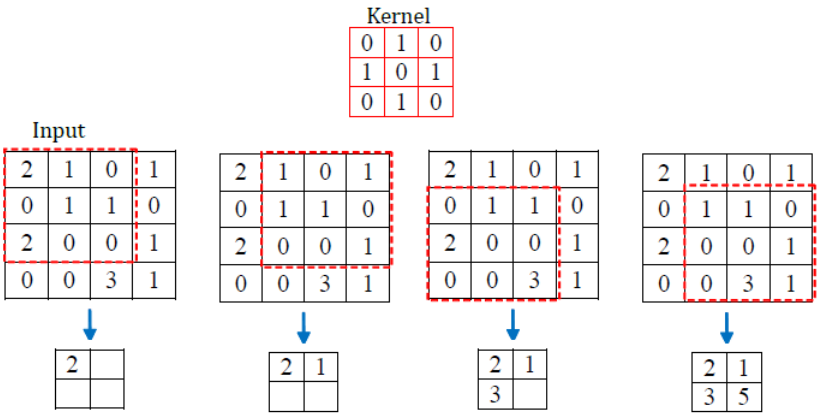
Επίπεδο συνέλιξης: Η διαδικασία της συνέλιξης χρησιμοποιείται ευρέως στην ψηφιακή επεξεργασία εικόνας όπου ένας πίνακας 2 διαστάσεων (2D) που αντιπροσωπεύει την εικόνα (I) συνελίσεται από ένα μικρότερο πίνακα 2D (K),

$$S_{i,j} = (I * K)_{i,j} = \sum_m \sum_n I_{i,j} \cdot K_{i-m,j-n}.$$

στη συνέχεια η μαθηματική διατύπωση με μηδενική συμπλήρωση δίνεται από :

Εξίσωση 2

Στη συνέλιξη, ένα κυλιόμενο φίλτρο μικρών διαστάσεων εφαρμόζεται στην εικόνα από αριστερά προς τα δεξιά και από πάνω μέχρι κάτω. *Η εικόνα 6* παρουσιάζει ένα παράδειγμα της διαδικασίας της συνέλιξης με μια εικόνα εισόδου διαστάσεων (4x4) και ένα φίλτρο συνέλιξης (3x3), λαμβάνοντας ως αποτέλεσμα μια νέα φιλτραρισμένη εικόνα. Σε κάθε θέση του φίλτρου συνέλιξης, υπολογίζεται το άθροισμα των γινομένων μεταξύ κάθε στοιχείου του φίλτρου και του αντίστοιχου στοιχείου της εικόνας εισόδου. Αυτή η διαδικασία επαναλαμβάνεται χρησιμοποιώντας διαφορετικά φίλτρα για να σχηματιστούν όσοι χάρτες χαρακτηριστικών εξόδου είναι επιθυμητό στη κάθε περίπτωση.



Εικόνα 6: Παράδειγμα της διαδικασίας της συνέλιξης με εικόνα εισόδου (4x4) και φίλτρο (3x3)

Οι διαστάσεις του χάρτη χαρακτηριστικών εξόδου είναι μειωμένες σε σχέση με την εικόνα εισόδου. Εναλλακτικά, μπορούμε να εφαρμόσουμε μια τεχνική padding για να διατηρήσουμε την ίδια διάσταση προσθέτοντας μηδενικά γύρω από την εικόνα εισόδου και τοποθετώντας το κέντρο του φίλτρου στα εξωτερικά στοιχεία. Άλλωστε το βήμα δηλώνει το μέγεθος του περάσματος μεταξύ δύο διαδοχικών θέσεων του φίλτρου. Γενικά, το βήμα του φίλτρου επιλέγεται ίσο με 1, αλλά μερικές φορές χρησιμοποιείται ένα βήμα μεγαλύτερο από 1 για τη μείωση της ανάλυσης του χάρτη χαρακτηριστικών κατά την διαδικασία της υποδειγματοληψίας.

Μη γραμμική συνάρτηση ενεργοποίησης: Αφού η συστοιχία φίλτρων παράγει το αποτέλεσμα, μια μη γραμμική συνάρτηση ενεργοποίησης εφαρμόζεται στην *(Εξίσωση (1))* για την παραγωγή των χαρτών ενεργοποίησης, όπου μόνο τα ενεργοποιημένα χαρακτηριστικά μεταφέρονται στο επόμενο επίπεδο. Αυτή η συνάρτηση καθορίζει τη συμπεριφορά του αποτελέσματος του νευρώνα. Στη συνέχεια, η λειτουργία της συνάρτησης ενεργοποίησης f (.) είναι η εξής:

$$\phi(Y_i^{(l)}) = f\left(B_i^{(l)} + \sum_{j=1}^{m_1^{(l-1)}} K_{ij}^{(l)} * Y_j^{(l-1)}\right)$$

Εξίσωση 3

Υπάρχουν διάφοροι είδη συναρτήσεων ενεργοποίησης. Επί του παρόντος, οι πιο ευρέως χρησιμοποιούμενες στα CNN είναι:

- **Η συνάρτηση Rectified Linear Unit (ReLU)**: Η ReLU είναι η πιο χρησιμοποιούμενη συνάρτηση

$$f(x) = max(0, x) = \begin{cases} 0 & \text{if} \quad x < 0 \\ x & \text{if} \quad x \geq 0 \end{cases}$$

ενεργοποίησης για τα συνελικτικά επίπεδα,(βλ. *Εικόνα 7a*). Ορίζεται μαθηματικά ως:

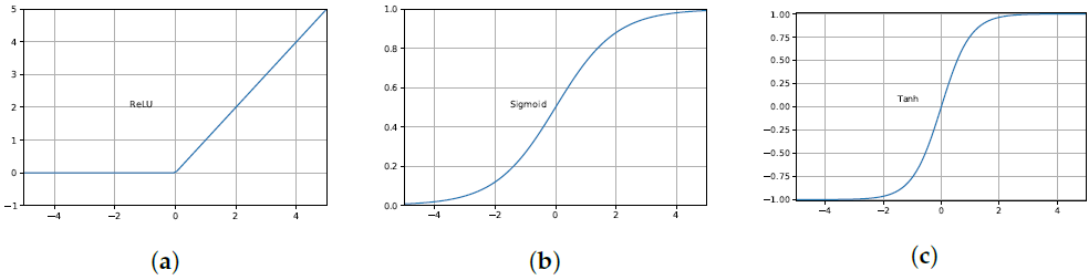
- **Σιγμοειδής συνάρτηση**: Η καμπύλη της μοιάζει με σχήμα S όπως φαίνεται στην *Εικόνα 7b*. Τιμές της συνάρτησης κυμαίνονται μεταξύ [0, 1], επομένως χρησιμοποιείται για να προβλέψει μια πιθανότητα ως αποτέλεσμα. Μαθηματικά αυτό έχει τη μορφή:

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

- **Συνάρτηση Εφαπτομένης Hyperbolic Tangent (tanh)**: Η συνάρτηση tanh έχει παρόμοια μορφή με τη σιγμοειδή, όπως φαίνεται στην *Εικόνα 7c*, αλλά το εύρος των τιμών της είναι [-1, 1]. Το πλεονέκτημα είναι ότι οι μηδενικές τιμές θα να αντιστοιχίζονται κοντά στο μηδέν και οι αρνητικές τιμές θα αντιστοιχιστούν

$$f(x) = \tanh(x) = \frac{2}{1 + e^{-2x}} - 1$$

έντονα αρνητικές. Μαθηματικός της ορισμός είναι:



(a) (b) (c)

Εικόνα 7: Αναπαραστάσεις των πιο χρησιμοποιούμενων συναρτήσεων ενεργοποίησης (a) RELU (b) Sigmoid (c) Hyperbolic Tangent

Συναθροιστικό Επίπεδο/ Pooling Layer: μειώνει τον αριθμό των παραμέτρων του δικτύου μειώνοντας το χωρικό μέγεθος συνελικτικών αποτελεσμάτων. Επιπλέον, τα συνελικτικά επίπεδα συμβάλλουν στην απόκτηση μιας αμετάβλητης αναπαράστασης σε μικρές απεικονίσεις του δεδομένου εισόδου. Τα δύο κύρια συναθροιστικά επίπεδα που συναντώνται είναι:

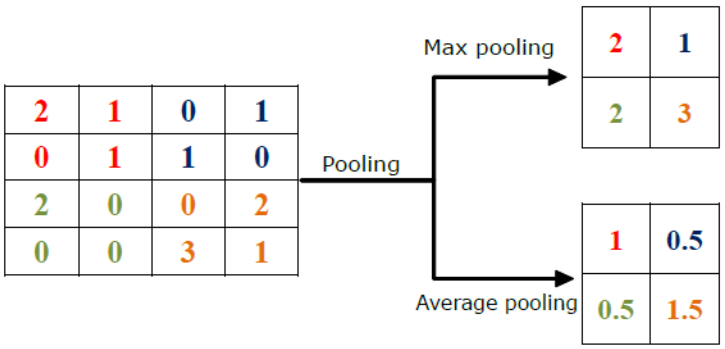
- Max pooling επίπεδο: Υπολογίζει τη μέγιστη τιμή. Το max-pooling είναι το επίπεδο που διατηρεί τη μέγιστη τιμή του τμήματος της εικόνας που εφαρμόζεται(patch), δηλαδή λειτουργεί σαν κυλιόμενο φίλτρο μέγιστης τιμής πάνω από τον χάρτη χαρακτηριστικών. Μαθηματικά έχει τη μορφή:

$$f_{max}(A) = max_{n\times m}(A_{n\times m})$$

  Συνήθως, τα max pooling επίπεδα εφαρμόζονται ως 2x2 φίλτρα με βήμα 2. Έτσι μειώνονται οι διαστάσεις των δεδομένων εισόδου στο φίλτρο κατά 2 κατά και απορρίπτει το 75% των συνελικτικών αποτελεσμάτων.

- Average pooling επίπεδο: Υπολογίζει τη μέση τιμή του τμήματος της εικόνας που εφαρμόζεται(patch). Το επίπεδο της μέσης τιμής εξομαλύνει τη συνελικτική ενεργοποίηση διαιρώντας το δεδομένο εισόδου σε ομαδοποιημένες περιοχές και υπολογίζοντας τις μέσες τιμές τους. Ορίστηκε μαθηματικά ως εξής:

$$f_{ave}(A) = \frac{1}{n+m}\sum_{i=1}^{n}\sum_{k=1}^{m}(A_{i,k})$$



Εικόνα 8: Παραδείγματα συναθροιστικών επιπέδων εφαρμοζόμενα ως 2x2 φίλτρα με βήμα 2

Dropout επίπεδο: Είναι ένα επίπεδο τακτοποίησης που απορρίπτει τυχαία μονάδες νευρώνων του δικτύου, αποτρέποντας την υπερβολική προσαρμογή των μονάδων. Η τεχνική dropout επιτρέπει την αντιμετώπιση του προβλήματος της υπερπροσαρμογής των δεδομένων, και ταυτόχρονα, βελτιώνει την απόδοση του δικτύου. Μπορεί να εφαρμοστεί σε οποιοδήποτε επίπεδο στο δίκτυο.

Πλήρως συνδεδεμένο επίπεδο (Fully Connected): Το τελικό αποτέλεσμα των συνελικτικών σταδίων μειώνεται σε ένα μονοδιάστατο πίνακα και συνδέεται σε ένα πλήρως συνδεδεμένο επίπεδο. Τα επίπεδα FC λαμβάνουν τα αποτελέσματα των επιπέδων της συνέλιξης/pooling τα επεξεργάζονται και τα χρησιμοποιούν για να ταξινομήσουν την εικόνα σε μια ετικέτα (δηλαδή, τάξη κατηγορία), όπως ένα παραδοσιακό νευρωνικό δίκτυο. Έτσι, η συνάρτηση

ενεργοποίησης του τελευταίου επιπέδου (δηλαδή του επιπέδου εξόδου) υπολογίζει τις τελικές πιθανότητες για κάθε τάξη και επιλέγεται ανάλογα με το πρόβλημα. Συνήθως, μια εργασία ταξινόμησης πολλών κλάσεων χρησιμοποιεί τη συνάρτηση Softmax, όπου η τιμή πιθανότητας για κάθε κλάση κυμαίνεται μεταξύ [0, 1] και το συνολικό άθροισμά τους είναι ίσο με 1. Τέλος, κάθε νευρώνας εξόδου αποφασίζει για κάθε μία από τις ετικέτες, όποια κατηγορία/ ετικέτα λάβει τη μεγαλύτερη τιμή εξόδου ορίζεται και ως απόφαση της ταξινόμησης.

2.3 Λειτουργία Συνελικτικών Νευρωνικών Δικτύων

Σκοπός ενός CNN είναι η επιτυχής αναγνώριση συγκεκριμένων κατά περίπτωση προτύπων. Για να επιτευχθεί η αναγνώριση αυτή ακολουθείται μια διαδικασία εκπαίδευσης του μοντέλου πάνω σε ένα σετ δεδομένων εικόνων και ετικετών (κατηγοριοποιημένα δεδομένα), ώστε το μοντέλο να εκπαιδευτεί και να αναγνωρίζει τα επιθυμητά πρότυπα και σε άγνωστες εικόνες.

Η διαδικασία εκπαίδευσης είναι μια επαναληπτική διαδικασία, της οποίας το αποτέλεσμα ορίζεται ως συνάρτηση των δεδομένων εισόδου. Οι παράμετροι της συνάρτησης αυτής λέγονται βάρη (w) και σταθερές πόλωσης (b).

$$X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} \qquad W = \begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \\ W_{31} & W_{32} \\ W_{41} & W_{42} \end{bmatrix} \qquad b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

Input Data | Randomly Initialized Weight Matrix | Randomly Initialized bias Matrix

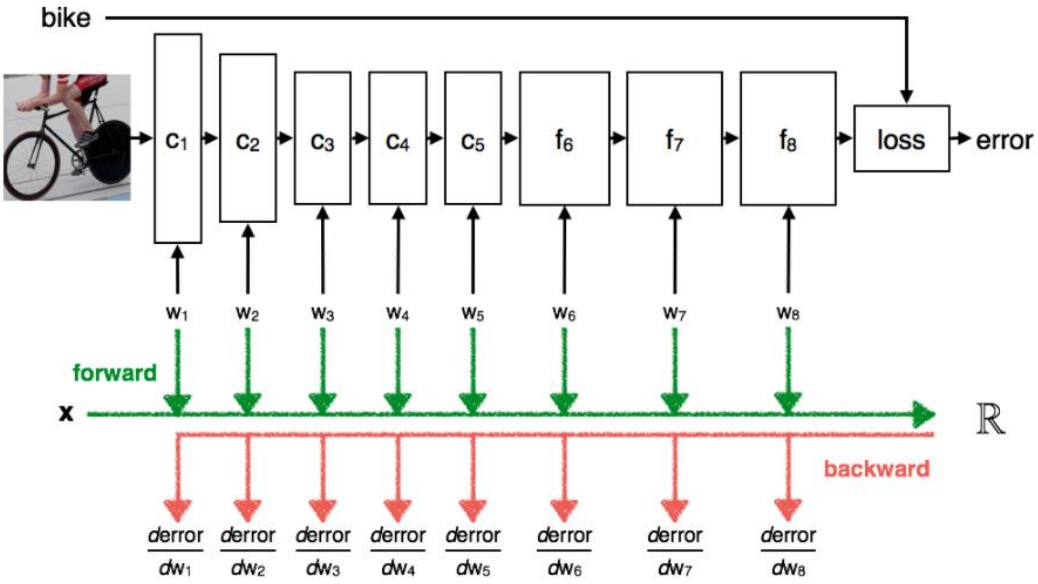
$$Z = W^T . X + b$$

$$Z = \begin{bmatrix} W_{11} & W_{21} & W_{31} & W_{41} \\ W_{12} & W_{22} & W_{32} & W_{42} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

$$Z_{2x2} = \begin{bmatrix} W_{11}X_1 + W_{21}X_2 + W_{31}X_3 + W_{41}X_4 \\ W_{12}X_1 + W_{22}X_2 + W_{32}X_3 + W_{42}X_4 \end{bmatrix}$$

Εικόνα 9: Μαθηματική αναπαράσταση των υπολογισμών των CNN

Κατά τη διαδικασία εκπαίδευσης εκτελούνται οι πρόσθιοι υπολογισμοί (forward pass) του μοντέλου έχοντας δώσει τυχαίες τιμές στα w και b. Στο τέλος κάθε πρόσθιου υπολογισμού των παραμέτρων, ορίζεται μια συνάρτηση απώλειας η οποία υπολογίζει πόσο κοντά είναι η πρόβλεψη του μοντέλου με το πραγματικό αποτέλεσμα. Από τη διαφορά αυτή που προκύπτει, καθορίζονται τα dw και db κατά την οπισθοδιάδοση του σφάλματος. Με τρόπο αυτό θα ανανεωθούν οι τιμές για τα w και b για το επόμενο forward pass. Στο τέλος της εκπαίδευσης η συνάρτηση κόστους υπολογίζει τη μέση τιμή της συνάρτησης απώλειας για όλο το σετ δεδομένων εκπαίδευσης. Στην ουσία η εκπαίδευση θέλει να βρει τις κατάλληλες τιμές των παραμέτρων w και b ώστε να μειώσει τη τιμή της συνάρτησης κόστους.



Εικόνα 10: Η Διαδικασία Εκπαίδευσης περιλαμβάνει τους εμπρόσθιους υπολογισμούς και την οπισθοδιάδοση σφάλματος για τις παραμέτρους w & b

2.4 Διαδικασία Εκπαίδευσης ενός CNN

Η διαδικασία εκπαίδευσης βελτιστοποιεί διαφορετικές παραμέτρους του επιπέδου ενός νευρωνικού δικτύου για να ελαχιστοποιήσει τις διαφορές μεταξύ των δεδομένων εκπαίδευσης και των προβλέψεων που είναι το αποτέλεσμα του δικτύου. Συνήθως, ο αλγόριθμος backpropagation (οπισθοδιάδοση σφάλματος) είναι η πιο χρησιμοποιούμενη μέθοδος για την εκπαίδευση νευρωνικών δικτύων. Η διαδικασία της εκπαίδευσης με το backpropagation είναι η εξής:

1. Επιλέγεται ένα σύνολο δεδομένων εκπαίδευσης στην προκειμένη ένα σύνολο εικόνων, που λαμβάνονται συνήθως ανά τμήμα (batch) με μικρότερες διαστάσεις.
2. Κάθε batch εικόνων διέρχεται μέσω του δικτύου και λάβετε το αποτέλεσμα.
3. Υπολογίζεται το σφάλμα μεταξύ των δεδομένων ετικετών (ground truth) και των προβλέψεων που έκανε το μοντέλο χρησιμοποιώντας μια συνάρτηση απώλειας L.
4. Διαδίδεται το σφάλμα σε όλο το δίκτυο με τον αλγόριθμο backpropagation.
5. Ενημερώνονται τα βάρη W για να ελαχιστοποιηθεί το σφάλμα.

6. Η παραπάνω διαδικασία είναι επαναλαμβανόμενη μέχρι να συγκλίνουν οι δεδομένες ετικέτες (ground truth) και οι προβλέψεις που έκανε το μοντέλο ή να φτάσει το μοντέλο σε ένα όριο επαναλήψεων όπου έχει οριστεί από τον χρήστη.

Για να εκτελεστούν τα προηγούμενα βήματα και να εκπαιδευτεί ένα CNN, πρέπει να εξεταστούν οι ακόλουθες πτυχές:

- Ορισμός της αρχιτεκτονικής του CNN: Αποτελείται από τον καθορισμό του αριθμού των επιπέδων για τον κάθε αντίστοιχο τύπο, καθώς και το μέγεθος και τον αριθμό των φίλτρων για κάθε επίπεδο. Ο σχεδιασμός της αρχιτεκτονικής εξαρτάται πάντα από τον στόχο του CNN.
- Συνάρτηση απώλειας (loss function) : Μετρά τη διαφορά μεταξύ των δεδομένων ετικετών (ground truth) και των αποτελεσμάτων του δικτύου. Συνήθως, εφαρμόζεται η συνάρτηση μέσου τετραγώνου σφάλματος και δίνεται από:

$$L = \sum (target - output)^2$$

Ως εκ τούτου, το L πρέπει να ελαχιστοποιηθεί για να βρεθεί η συμβολή κάθε βάρους και να βελτιστοποιηθούν οι τιμές τους. Ο αλγόριθμος gradient descent υιοθετείται ευρέως για τη διαδικασία αυτής της ελαχιστοποίησης, η οποία εκφράζεται μαθηματικά ως μερική παράγωγος της συνάρτησης απώλειας. Στη συνέχεια, η ενημέρωση των παραμέτρων διατυπώνεται ως εξής:

$$W_k = W_{k-1} - \alpha * \frac{\partial L}{\partial W},$$

Στη παραπάνω εξίσωση, το a υποδηλώνει το ρυθμό εκπαίδευσης (learning rate). Ο ρυθμός εκπαίδευση είναι μια σημαντική παράμετρος και πρέπει να οριστεί πριν από την έναρξη της εκπαίδευσης. Πρέπει να σημειωθεί ότι όσο μικρότερη η τιμή του ρυθμού εκπαίδευσης, μπορεί να δώσει πιο ακριβές αποτέλεσμα στο μοντέλο, αλλά το δίκτυο μπορεί να χρειαστεί περισσότερο χρόνο για να εκπαιδευτεί.

- Δεδομένα εκπαίδευσης: τα διαθέσιμα δεδομένα χωρίζονται γενικά σε τρία υποσύνολα: το 1^{ο} σετ είναι τα δεδομένα εκπαίδευσης που χρησιμοποιούνται για την εκπαίδευση του δικτύου, το 2^{ο} σετ είναι το σύνολο δεδομένων επικύρωσης για την αξιολόγηση του μοντέλου κατά τη διαδικασία της εκπαίδευσης και το 3^{ο} σετ δεδομένων είναι τα δεδομένα δοκιμής πάνω στα οποία θα αξιολογήσει το τελικό εκπαιδευμένο μοντέλο.
  Τα περισσότερα frameworks για CNN απαιτούν όλα τα δεδομένα εκπαίδευσης να έχουν τις ίδιες διαστάσεις. Επομένως, η προεπεξεργασία των δεδομένων είναι το πρώτο βήμα πριν τη διαδικασία εκπαίδευσης για την ομαλοποίηση των δεδομένων.

  Ένα άλλο σημαντικό σημείο είναι ότι το σύνολο δεδομένων πρέπει να είναι ισορροπημένο, που σημαίνει ότι για κάθε κατηγορία εκπαίδευσης θα πρέπει να υπάρχει ο ίδιος αριθμός εικόνων. Σε περίπτωση που το σύνολο δεδομένων δεν έχει επαρκή αριθμό εικόνων, συνιστάται η εφαρμογή της τεχνικής επαύξησης δεδομένων. Η επαύξηση δεδομένων (data augmentation), συντελεί στην αύξηση της ποσότητας των δεδομένων εκπαίδευσης εκτελώντας μια σειρά μετασχηματισμών, όπως περιστροφή, κατοπτρισμός κ.α.

2.5 Transfer Learning & Fine Tuning

Για την αποτελεσματική εκπαίδευση ενός CNN από την αρχή, απαιτείται μεγάλος όγκος δεδομένων εκπαίδευσης. Η διαδικασία συλλογής και προεπεξεργασίας (εισαγωγή ετικετών) των δεδομένων εκπαίδευσης είναι χρονοβόρα

διαδικασία, καθώς δεν είναι πάντα εύκολο να βρεις τόσο μεγάλο όγκο δεδομένων για το πρόβλημα που εξετάζεις που να επαρκεί για την εκπαίδευση του μοντέλου.

Στις μέρες μας υπάρχει στο διαδίκτυο τεράστιος όγκος διαθέσιμων εικόνων που έχουν κατηγοριοποιηθεί και παρέχονται ως σετ δεδομένων με εκατομμύρια εικόνες, όπως το Image Net και το COCO dataset, τα οποία είναι ελεύθερα για να εκπαιδεύσει κανείς το μοντέλο του.

Εικόνα 11 : Περιγραφή του ImageNet

Καθώς η βαθιά μηχανική μάθηση είναι πολύ δημοφιλής τεχνική για την επεξεργασία εικόνας, έχουν δημιουργηθεί μοντέλα deep learning με υψηλά ποσοστά ακρίβειας, τα οποία διατίθενται ελεύθερα στο κοινό για έρευνα όπως τα AlexNet, VGG, CaffeNet, GoogleNet και το ResNet κ.α. Το πλεονέκτημα των μοντέλων αυτών είναι ότι τα περισσότερα από αυτά είναι ήδη προ-εκπαιδευμένα με ανοιχτά σετ δεδομένων, πράγμα που σημαίνει ότι το δίκτυο είναι έτοιμο για εντοπίζει με επιτυχία πολλά χαρακτηριστικά. Τα μοντέλα αυτά έχουν συνήθως εκπαιδευτεί στα σύνολα δεδομένων που αναφέρθηκαν παραπάνω όπως τα Image-Net και COCO. Το κύριο μειονέκτημα του DL μπορεί να είναι ο εκτεταμένος χρόνος εκπαίδευσης καθώς και η αναγκαιότητα κατάλληλου hardware, ενώ οι κλασικές μέθοδοι όπως (SVM) ή το Scale-invariant feature transform (SIFT) έχουν απλούστερες διαδικασίες εκπαίδευσης. Εντούτοις, το μειονέκτημα αυτό μπορεί να ξεπεραστεί σε ένα βαθμό με τα προ-εκπαιδευμένα μοντέλα.

Στη παρούσα εργασία, τα τρία μοντέλα που παρουσιάζονται και συγκρίνονται είναι προ-εκπαιδευμένα στο ImageNet. Δηλαδή τα βάρη τους έχουν εκ των προτέρων υπολογιστεί μέσω μιας χρονοβόρας διαδικασίας εκπαίδευσης με υψηλές απαιτήσεις σε υπολογιστικό κόστος. Μέσω των τεχνικών του transfer learning και του fine tuning που θα περιγραφούν παρακάτω τα προ-εκπαιδευμένα μοντέλα προσαρμόστηκαν στο σετ δεδομένων της παρούσης εργασίας, ώστε με επιτυχία να αναγνωρίζουν τις κατηγόριες των φρούτων του dataset.

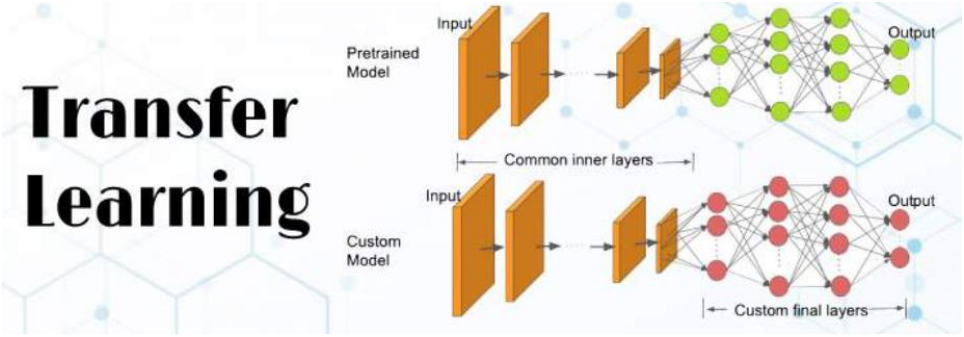
Transfer Learning

Το transfer learning είναι μια μέθοδος μηχανικής μάθησης όπου ένα δεδομένο μοντέλο CNN που αναπτύχθηκε για ένα συγκεκριμένο σκοπό επαναχρησιμοποιείται ώστε να δημιουργηθεί ένα νέο μοντέλο που αντιμετωπίζει κάποιο άλλο πρόβλημα. Είναι μια δημοφιλής προσέγγιση της βαθιάς μηχανική μάθηση όπου τα προ-εκπαιδευμένα μοντέλα χρησιμοποιούνται ως αφετηρία, λόγω των τεράστιων υπολογιστικών και χρονικών πόρων που απαιτούνται για την ανάπτυξη μοντέλων νευρωνικών δικτύων σε αυτά τα προβλήματα, μπορούν να παρέχουν καλά

αποτελέσματα και σε άλλα σχετικά προβλήματα, επειδή εμφανίζουν μεγάλη γενίκευση. Τα βάρη αυτών των προ-εκπαιδευμένων μοντέλων είναι προσαρμοσμένα ώστε να αναγνωρίζουν αποτελεσματικά πληθώρα χαρακτηριστικών.

Βασικά βήματα χρήσης ενός προ-εκπαιδευμένου μοντέλου:

1. Βασικό Μοντέλο: Επιλέγεται ένα προ-εκπαιδευμένο μοντέλο-βάση από τα διαθέσιμα στο διαδίκτυο μοντέλα. Πολλά ερευνητικά ιδρύματα δημοσιεύουν και διαθέτουν μοντέλα που είναι προ-εκπαιδευμένα σε μεγάλα και απαιτητικά σύνολα δεδομένων και έχουν εκπαιδευτεί να αναγνωρίζουν πολλές κατηγορίες (κλάσεις) π.χ. μοντέλα που έχουν εκπαιδευτεί στο ImageNet και μπορούν να αναγνωρίζουν 1000 κατηγορίες.
2. Επαναχρησιμοποίηση βασικού μοντέλου: Το προ-εκπαιδευμένο μοντέλο-βάση μπορεί στη συνέχεια να χρησιμοποιηθεί και να αντικαταστήσει τα αρχικά επίπεδα ενός νέου μοντέλου/δικτύου. Αυτό μπορεί να περιλαμβάνει τη χρήση ολόκληρου ή τμημάτων του μοντέλου, ανάλογα με την τεχνική μοντελοποίησης που χρησιμοποιείται. Η συνήθης πρακτική είναι να αφαιρείται το 2^{o} τμήμα /επίπεδο ταξινόμησης (εικόνα 12) από το μοντέλο που αποτελεί τη βάση, έτσι τα βάρη των επιπέδων που είναι εξαγωγείς χαρακτηριστικών έχουν έτοιμες τιμές από τα μεγάλα σετ δεδομένων που έχουν προ-εκπαιδευτεί.
3. Συντονισμός μοντέλου: Προαιρετικά, το μοντέλο μπορεί να χρειαστεί να προσαρμοστεί ή να βελτιωθεί (fine tuning) στα διαθέσιμα νέα δεδομένα εισόδου-εξόδου για το νέο πρόβλημα που εξετάζεται. Δηλαδή τα πλήρως συνδεδεμένα επίπεδα του νέου μοντέλου στόχου εκπαιδεύονται στο νέο μικρό σετ δεδομένων που μελετάται ώστε να μεταβληθούν τα βάρη τους και να εξειδικευτούν στις κατηγορίες ταξινόμησης του νέου προβλήματος.



Εικόνα 12: Σχηματική Αναπαράσταση του Transfer Learning

Fine Tuning

Το fine-tuning (συντονισμός του μοντέλου) είναι μια πρακτική της βαθιάς μηχανικής μάθησης που συναντάται ως επί το πλείστον σε συνδυασμό με το transfer learning. Με το fine-tuning στην ουσία επανεκπαιδεύονται τα τελευταία επίπεδα (classifier) του νέου μοντέλου ώστε να ταιριάζουν με τις κατηγορίες του σετ δεδομένων που μελετάται, όπως αναφέρθηκε και πριν στο transfer learning. Αλλά η διαφορά έγκειται στο ότι προσαρμόζονται και τα άλλα επίπεδα στο νέο σετ δεδομένων, χωρίς να μένουν τα βάρη τους σταθερά όπως πριν στο transfer learning. Είναι σημαντικό να σημειωθεί ότι σε ένα νευρωνικό δίκτυο, τα πρώτα στρώματα συνελίξεων και συνάθροισης εντοπίζουν απλούστερα και γενικότερα μοτίβα και όσο περισσότερο προχωράει η αρχιτεκτονική, τόσο πιο συγκεκριμένα γίνονται τα χαρακτηριστικά για το σύνολο δεδομένων που γίνεται η εκπαίδευση και πιο περίπλοκα τα πρότυπα που ανιχνεύουν. Επομένως, με το fine tuning επιτρέπεται η επανεκπαίδευση του τελευταίου μπλοκ στρωμάτων συνέλιξης και συνάθροισης (όχι όλων των layer του τμήματος της εξαγωγής χαρακτηριστικών), όπου τα πρότυπα που εντοπίζονται από το μοντέλο είναι σημαντικό να εξειδικευτούν πάνω πάνω στο σετ δεδομένων εικόνων και κατηγοριών που μελετάται.

2.6 Εντοπισμός Αντικειμένου

Στη παρούσα εργασία χρησιμοποιήθηκαν τα δίκτυα CNN τόσο για ταξινόμηση όσο και για εντοπισμό φρούτων σε εικόνες. Σε αυτή την ενότητα αρχικά θα αναλυθούν οι συγκεκριμένες έννοιες. Στη συνέχεια θα παρουσιαστεί η διαδικασία εντοπισμού ενός αντικειμένου. Τέλος αναλύονται τα 2 κύρια είδη αρχιτεκτονικών για εντοπισμό αντικειμένου που καταγράφονται στη βιβλιογραφία, δηλαδή αλγόριθμοι εντοπισμού ενός ή δυο σταδίων.

Η αναγνώριση αντικειμένων στην Όραση Υπολογιστών αποτελεί έναν γενικό όρο που αναφέρεται κατά περίπτωση στις επιμέρους διαδικασίες της ταξινόμησης, της ανίχνευσης και του εντοπισμού αντικειμένων.
Η **ταξινόμηση** εικόνων περιλαμβάνει την πρόβλεψη της κλάσης ενός αντικειμένου σε μια εικόνα.
Η **ανίχνευση αντικειμένων** αναφέρεται στον εντοπισμό της θέσης ενός ή περισσότερων αντικειμένων σε μια εικόνα και απόδοση πλαισίων οριοθέτησης γύρω από αυτά.
Ο **εντοπισμός αντικειμένων** συνδυάζει τις δύο παραπάνω εργασίες δηλαδή εντοπίζει και ταξινομεί ένα ή περισσότερα αντικείμενα σε μια εικόνα.

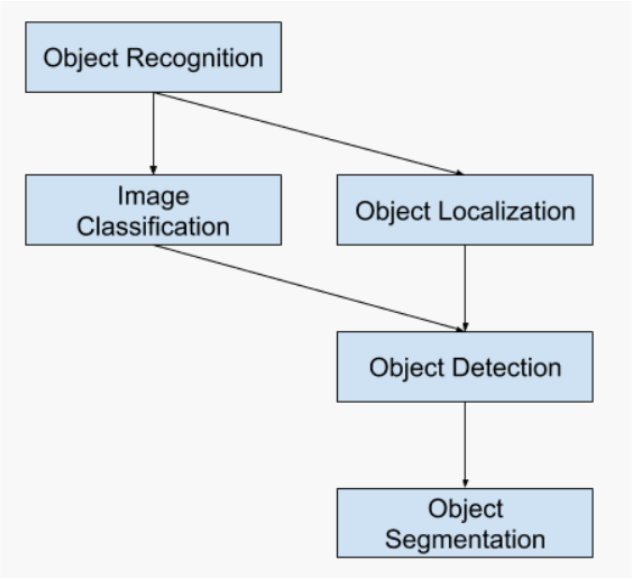
Ως εκ τούτου, μπορεί να γίνει διάκριση μεταξύ αυτών των τριών εργασιών της όρασης υπολογιστών:

Ταξινόμηση εικόνας: Προβλέπει το είδος ή την κατηγορία ενός αντικειμένου σε μια εικόνα.
Δεδομένο: Μια εικόνα με ένα μόνο αντικείμενο, όπως μια φωτογραφία.
Αποτέλεσμα: Μια ετικέτα της κατηγορίας που έχει προβλέψει ότι ανήκει το αντικείμενο της εικόνας (π.χ. ένας ή περισσότεροι ακέραιοι αριθμοί που αντιστοιχίζονται σε ετικέτες της κατηγορίας).

Ανίχνευση αντικειμένου: Ανιχνεύει την παρουσία αντικειμένων σε μια εικόνα και υποδεικνύει τη θέση τους με ένα πλαίσιο οριοθέτησης.
Δεδομένο: Μια εικόνα με ένα ή περισσότερα αντικείμενα, όπως μια φωτογραφία.
Αποτέλεσμα: Ένα ή περισσότερα οριοθετημένα πλαίσια (π.χ. που ορίζονται από σημείο και διαστάσεις όπως πλάτος και ύψος).

Εντοπισμός αντικειμένων: Εντοπίζει σε μια εικόνα την παρουσία αντικειμένων με πλαίσιο οριοθέτησης και προβλέπει τα είδη ή τις κατηγορίες των εντοπισμένων αντικειμένων.
Δεδομένο: Μια εικόνα με ένα ή περισσότερα αντικείμενα, όπως μια φωτογραφία.
Αποτέλεσμα: Ένα ή περισσότερα οριοθετημένα πλαίσια (π.χ. που ορίζονται από ένα σημείο, πλάτος και ύψος) και μια ετικέτα της κατηγορίας που έχει προβλέψει ότι ανήκει το αντικείμενο της εικόνας για κάθε πλαίσιο οριοθέτησης.

Μια περαιτέρω επέκταση αυτής της ανάλυσης των εργασιών της όρασης υπολογιστών είναι η **κατάτμηση** αντικειμένων, που ονομάζεται επίσης "σημασιολογική κατάτμηση", όπου σε αυτή τη περίπτωση τα αναγνωρισμένα αντικείμενα υποδεικνύονται επισημαίνοντας τα συγκεκριμένα εικονοστοιχεία του αντικειμένου αντί για ένα πλαίσιο οριοθέτησης που είναι πιο γενικό.

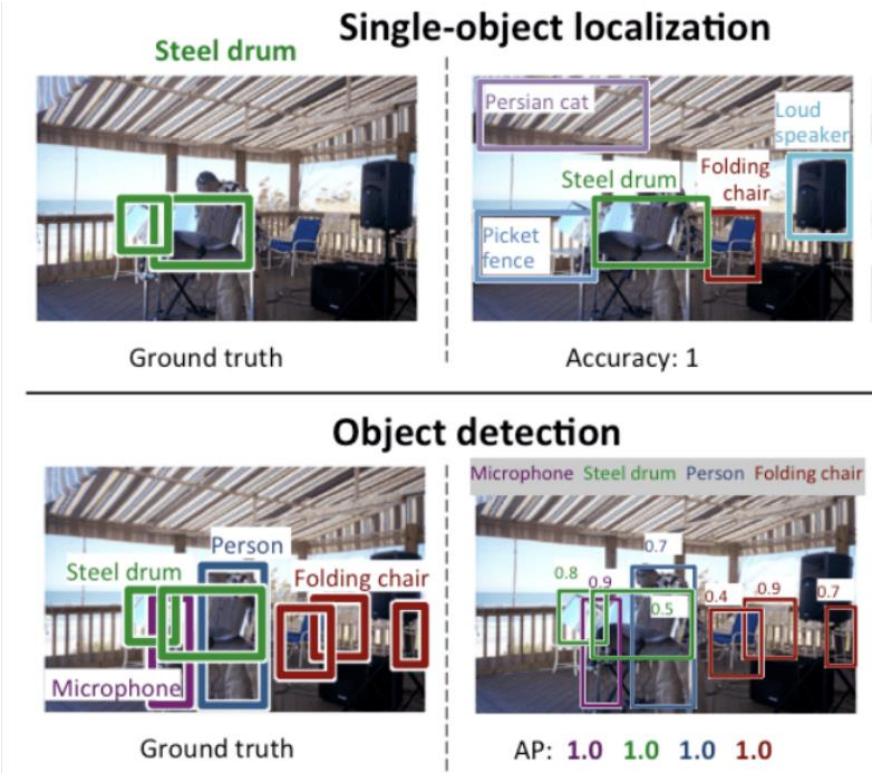


Εικόνα 13: Αναπαράσταση των διαφορετικών εργασιών της όρασης υπολογιστών για αναγνώριση αντικειμένου

Η απόδοση ενός μοντέλου για ταξινόμηση εικόνων αξιολογείται χρησιμοποιώντας το μέσο σφάλμα ταξινόμησης στις προβλεπόμενες ετικέτες κλάσεων.

Η απόδοση ενός μοντέλου για την αναγνώριση ενός αντικειμένου αξιολογείται χρησιμοποιώντας την απόσταση μεταξύ του αναμενόμενου και του προβλεπόμενου πλαισίου οριοθέτησης για την αναμενόμενη κλάση.

Ενώ η απόδοση ενός μοντέλου για τον εντοπισμό αντικειμένων αξιολογείται χρησιμοποιώντας την ακρίβεια και την ανάκληση σε καθένα από τα οριοθετημένα πλαίσια για τα γνωστά αντικείμενα στην εικόνα.



Εικόνα 14: Σύγκριση μεταξύ της αναγνώρισης ενός αντικειμένου σε εικόνα και στον εντοπισμό αντικειμένων σε εικόνα από το διαγωνισμό ImageNet Large Scale Visual Recognition Challenge

2.7 Μεθοδολογία και Τεχνικές στον Εντοπισμό Αντικειμένου

Γενικά, υπάρχουν δύο διαφορετικές προσεγγίσεις/τεχνικές για τον εντοπισμό αντικειμένου. Στην πρώτη προσέγγιση, ο αλγόριθμος υπολογίζει ένα σταθερό αριθμό προβλέψεων στην εικόνα και πραγματοποιείται σε ένα στάδιο. Στη δεύτερη προσέγγιση αξιοποιείται ένα δίκτυο για να βρεθούν αντικείμενα/προτάσεις στην εικόνα (α' στάδιο) και στη συνέχεια χρησιμοποιείται ένα δεύτερο δίκτυο για να τελειοποιήσει αυτές τις προτάσεις και να εξάγει ένα τελικό αποτέλεσμα (β' στάδιο). Οι τεχνικές αυτές θα αναλυθούν παρακάτω, μέσω και των αρχιτεκτονικών που τις εκπροσωπούν όπως το YOLO και SSD (μεθοδολογίες ενός σταδίου) και το Faster R-CNN (μεθοδολογία δυο σταδίων). Ανεξάρτητα από την προσέγγιση, μελετώντας την βιβλιογραφία εξήχθησαν τα παρακάτω βήματα ως βασική μεθοδολογία για τον εντοπισμό αντικειμένου από εικόνα.

Η κοινή μεθοδολογία ανίχνευσης αντικειμένου περιλαμβάνει:

1. προεπεξεργασία της εικόνας εισόδου (αλλαγή μεγέθους, ομαλοποίηση τιμών, επεξεργασία χρώματος κ.λπ.)
2. ανίχνευση αντικειμένων
3. τοποθέτηση πλαισίου οριοθέτησης (bounding box), συνήθως ορθογωνίου γύρω από τα αντικείμενα που ανιχνεύθηκαν για την διαδικασία του εντοπισμού
4. υπολογισμός του επιπέδου εμπιστοσύνης για κάθε κλάση
5. τελικό φιλτράρισμα των ανιχνεύσεων του μοντέλου βάσει του επιπέδου εμπιστοσύνης που θα λειτουργήσει ως κατώφλι επικάλυψης μεταξύ των κλάσεων με σκοπό την συγχώνευση πολλαπλών ανιχνεύσεων ενός αντικειμένου.

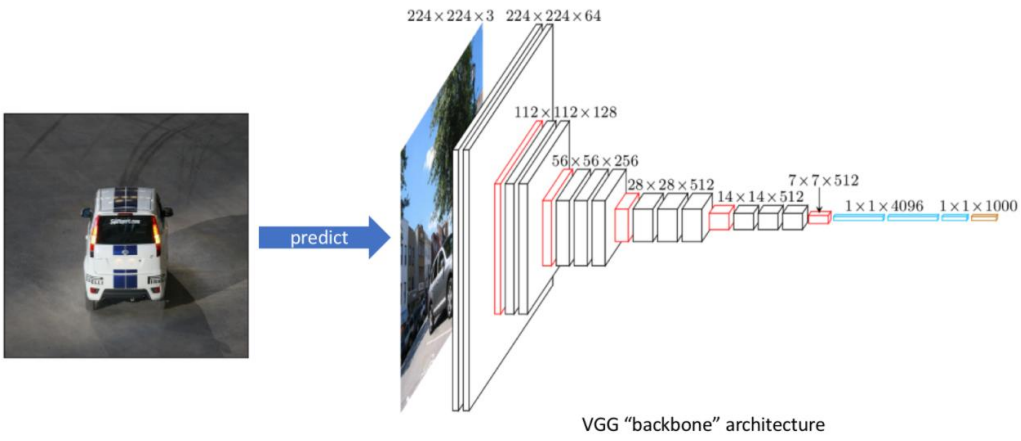
Τεχνικές ενός σταδίου

Ο στόχος της ανίχνευσης αντικειμένων είναι να αναγνωρίσει περιπτώσεις ενός προκαθορισμένου συνόλου κατηγοριών αντικειμένων (π.χ. άνθρωποι, αυτοκίνητα, ποδήλατα, ζώα, φρούτα) και να περιγράψει τις θέσεις κάθε αντικειμένου που ανιχνεύτηκε στην εικόνα χρησιμοποιώντας ένα πλαίσιο οριοθέτησης. Η χρήση ορθογώνιων πλαισίων οριοθέτησης μπορεί να οδηγήσει σε ατελή εντοπισμό λόγω των σχημάτων των αντικειμένων. Μια εναλλακτική προσέγγιση είναι η κατάτμηση εικόνας που παρέχει εντοπισμό αντικειμένου σε επίπεδο pixel.

Εικόνα 15: Αριστερά αρχική εικόνα και δεξιά εικόνα με ανιχνευμένα αντικείμενα

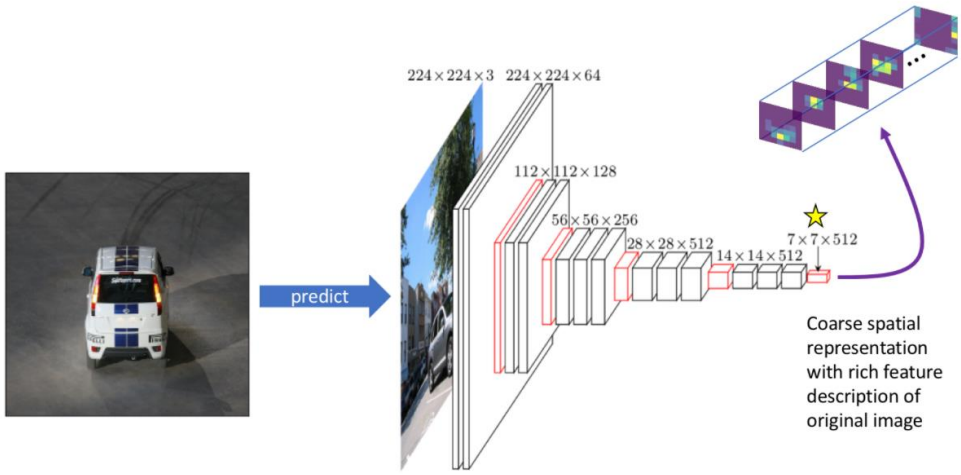
Αρχικά συναντάει κανείς τις μεθόδους, οι οποίες προβλέπουν άμεσα πλαίσια οριοθέτησης αντικειμένων σε μια εικόνα και ονομάζονται μέθοδοι ενός σταδίου. Χαρακτηρίζονται από απλούστερες και ταχύτερες αρχιτεκτονικές μοντέλου, αν και μερικές φορές μπορεί να δυσκολεύονται να προσαρμοστούν σε διαφορετικά εφαρμογές. Για να κατανοηθούν οι αρχιτεκτονικές ενός σταδίου θα αναλυθεί παρακάτω ο τρόπος λειτουργίας τους.

Το 1^{ο} τμήμα του δικτύου ονομάζεται, “backbone” δίκτυο, το οποίο συνήθως είναι προ-εκαπιδευμένο ως ταξινομητής εικόνας, για να είναι ευκολότερη και όχι τόσο χρονοβόρα η εξαγωγή χαρακτηριστικών από μια εικόνα. Έτσι βοηθά να εκπαιδεύεται σε ένα μεγάλο σύνολο δεδομένων (όπως το ImageNet),ώστε να αποκτήσει τη δυνατότητα να ανιχνεύει πληθώρα αντικειμένων.



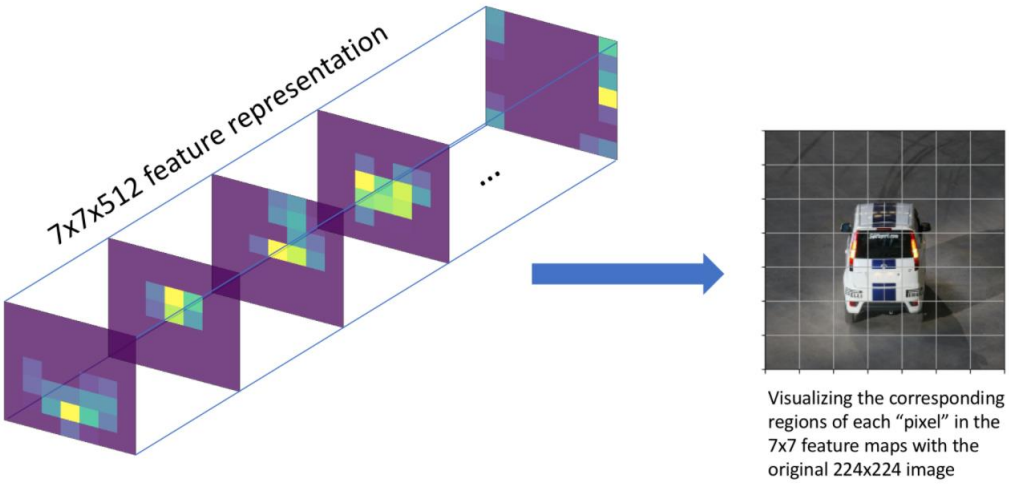
Εικόνα 16: Η αρχιτεκτονική του μοντέλου VGG χρησιμοποιούμενη ως δίκτυο backbone

Έχοντας εκπαιδευτεί εκ των προτέρων η αρχιτεκτονική κορμού (backbone) και αφαιρώντας τα τελευταία επίπεδα του δικτύου αποτελεί τον ταξινομητή εικόνας. Το δίκτυο κορμού είναι μια συλλογή χαρτών χαρακτηριστικών σε στοίβα που περιγράφουν την αρχική εικόνα σε χαμηλή χωρική ανάλυση, όμως έχει υψηλή ανάλυση χαρακτηριστικών. Στο παρακάτω παράδειγμα, υπάρχει μια αναπαράσταση 7x7x512 της αρχικής εικόνας. Καθένας από τους 512 χάρτες χαρακτηριστικών περιγράφει διαφορετικά χαρακτηριστικά της αρχικής εικόνας.



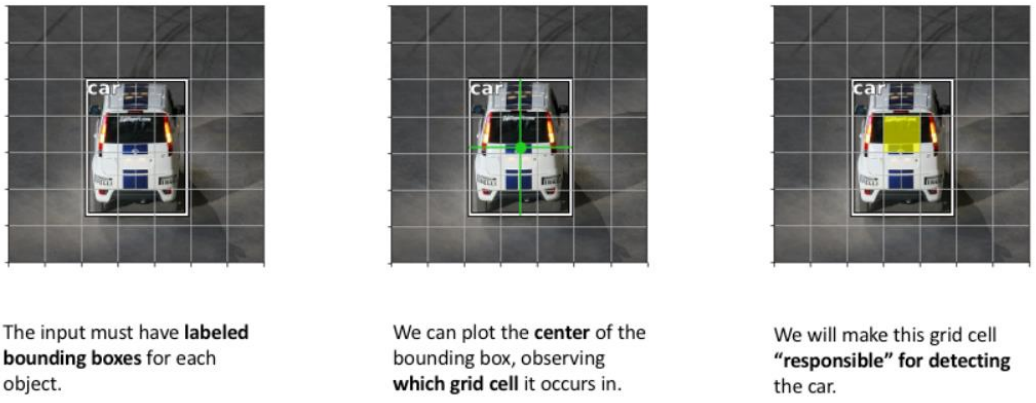
Εικόνα 17: Το backbone δίκτυο αποτελείται από ένα σύνολο χαρτών χαρακτηριστικών με χαμηλή χωρική ανάλυση αλλά με πλούσια περιγραφή των χαρακτηριστικών της αρχικής εικόνας

Συσχετίζοντας αυτό το πλέγμα 7x7 με το αρχικό δεδομένο μπορεί να κατανοηθεί τι αντιπροσωπεύει κάθε κελί του καννάβου σε σχέση με την αρχική εικόνα.



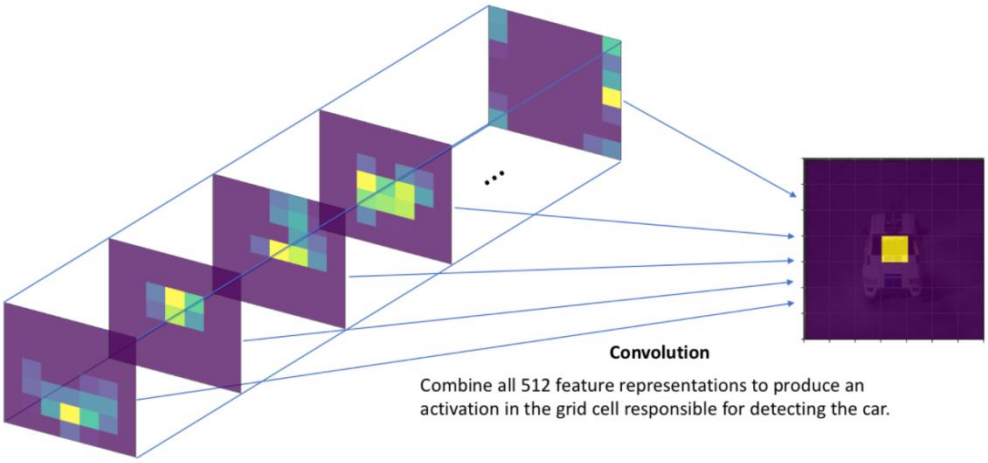
Εικόνα 18: Αντιστοίχιση των 7x7 χαρτών χαρακτηριστικών με την αρχική εικόνα

Μπορεί επίσης να προσδιοριστεί χονδρικά πού βρίσκονται τα αντικείμενα στους χάρτες χαρακτηριστικών (7x7), παρατηρώντας ποιο κελί του καννάβου περιέχει το κέντρο του πλαισίου οριοθέτησης. Θα οριστεί αυτό το κελί του καννάβου ως "υπεύθυνο" για τον εντοπισμό του συγκεκριμένου αντικειμένου.



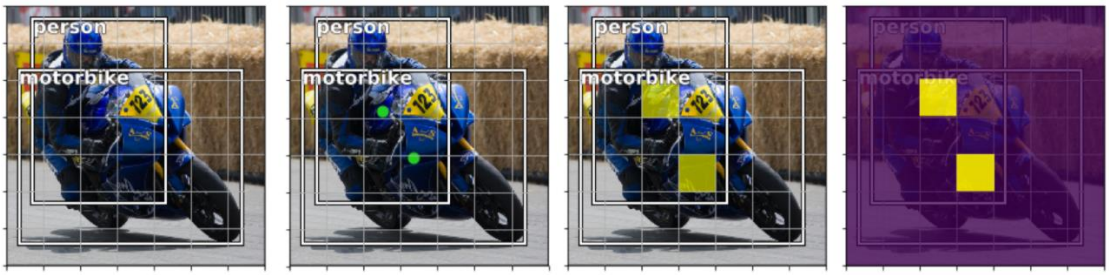
Εικόνα 19: Ορισμός του «υπεύθυνου» κελιού για τον εντοπισμό αντικειμένου

Για να ανιχνευτεί αυτό το αντικείμενο, θα προστεθεί ένα άλλο συνελικτικό επίπεδο και θα εκπαιδευτεί στις παραμέτρους του πυρήνα που συνδυάζουν το περιεχόμενο και των 512 χαρτών χαρακτηριστικών για να παράγει μια ενεργοποίηση που αντιστοιχεί στο κελί του καννάβου που περιέχει το εξεταζόμενο αντικείμενο.



Εικόνα 20: Συνδυάζοντας όλα τα χαρακτηριστικά που έχουν εξαχθεί, παράγεται η ενεργοποίηση του «υπεύθυνου» κελιού για τον εντοπισμό του αντικειμένου

Αν η εικόνα εισόδου περιέχει πολλαπλά αντικείμενα, πρέπει να υπάρχουν πολλαπλές ενεργοποιήσεις στον κάνναβο, δηλώνοντας ότι ένα αντικείμενο βρίσκεται σε καθεμιά από τις ενεργοποιημένες περιοχές.

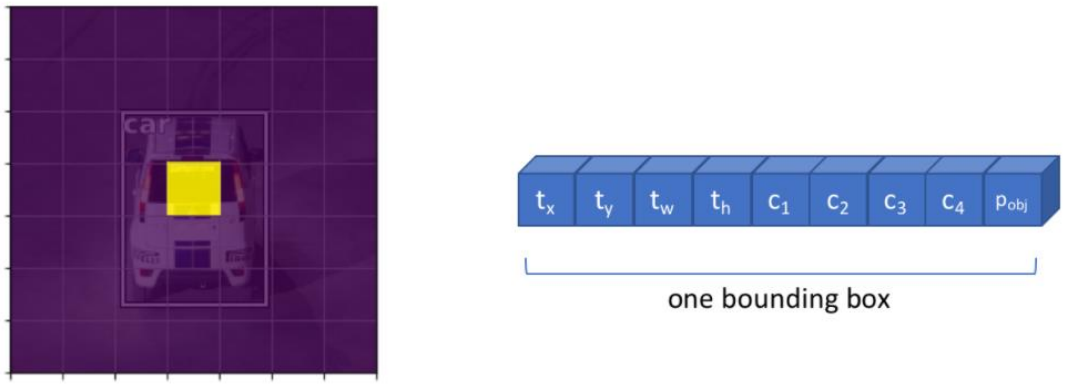


Εικόνα 21: Ενεργοποιήσεις 2 περιοχών σε εικόνα με 2 αντικείμενα

Ωστόσο, δεν μπορεί να περιγραφεί επαρκώς κάθε αντικείμενο με μία μόνο ενεργοποίηση. Για να περιγραφεί πλήρως ένα αντικείμενο που έχει εντοπιστεί, θα πρέπει να οριστεί:

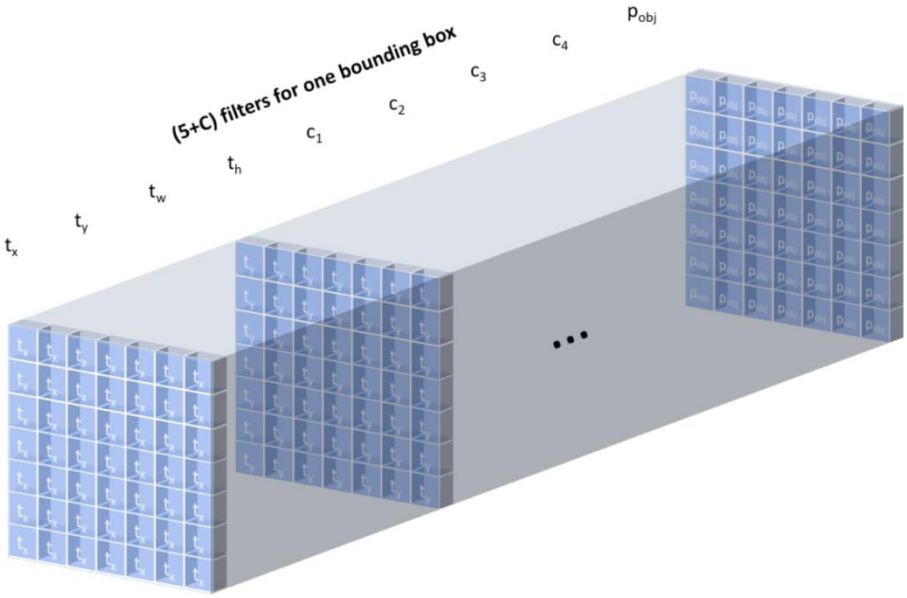
- Η πιθανότητα ότι ένα κελί του καννάβου περιέχει ένα αντικείμενο (p_{obj})
- Σε ποια κλάση ανήκει το αντικείμενο (c_1, c_2, ..., c_C)
- Τέσσερις περιγραφείς για το πλαίσιο οριοθέτησης για να περιγράψουν τις συντεταγμένες x και y, το πλάτος και το ύψος ενός πλαισίου οριοθέτησης με ετικέτα κλάσης (t_x, t_y, t_w, t_h)

Επομένως, θα χρειαστεί να εκπαιδευτεί ένα φίλτρο συνέλιξης για κάθε μια από τις παραπάνω παραμέτρους, έτσι ώστε να παραχθούν 5+C αποτελέσματα για να περιγράψουν ένα μόνο πλαίσιο οριοθέτησης σε κάθε του κελιού του καννάβου. Άρα θα εκπαιδευτεί ένα σύνολο βαρών για να εξεταστούν οι 512 χάρτες χαρακτηριστικών και να προσδιοριστεί ποια κελιά του καννάβου είναι πιθανό να περιέχουν ένα αντικείμενο, ποιες κλάσεις είναι πιθανό να υπάρχουν σε κάθε κελί του καννάβου και πώς να περιγραφεί το πλαίσιο οριοθέτησης για πιθανά αντικείμενα σε κάθε κελί του καννάβου.



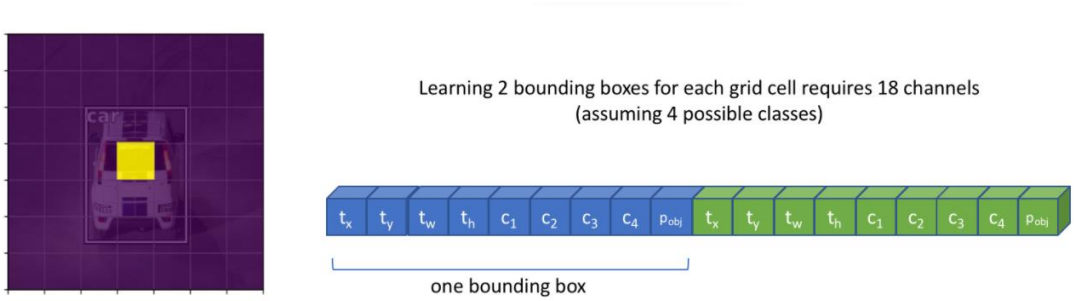
Εικόνα 22: Οι παράμετροι που περιγράφουν ένα bounding box

Το πλήρες αποτέλεσμα της εφαρμογής 5+C συνελικτικών φίλτρων παρουσιάζεται στην παρακάτω εικόνα, στην ουσία παράγεται ένα πλαίσιο οριοθέτησης για να περιγράψει κάθε κελί του καννάβου.



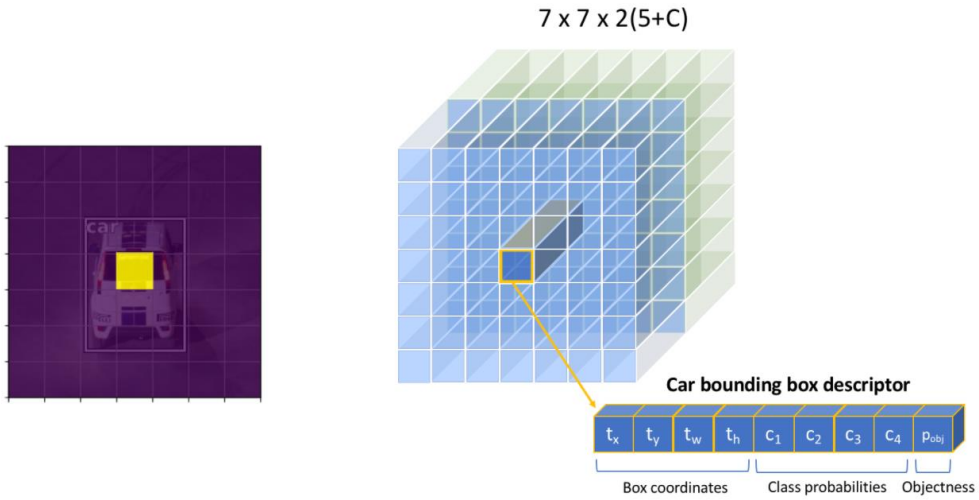
Εικόνα 23: Πλήρες αποτέλεσμα 5+C συνελικτικών φίλτρων σε μια εικόνα

Ωστόσο, ορισμένες εικόνες μπορεί να έχουν πολλά αντικείμενα που "ανήκουν" στο ίδιο κελί του καννάβου. Χρειάζεται να αλλάξουν τα επίπεδα για να παραχθούν B*(5+C) φίλτρα έτσι ώστε να μπορούν να προβλεφθούν τα B πλαίσια οριοθέτησης για κάθε θέση κελιού του καννάβου.



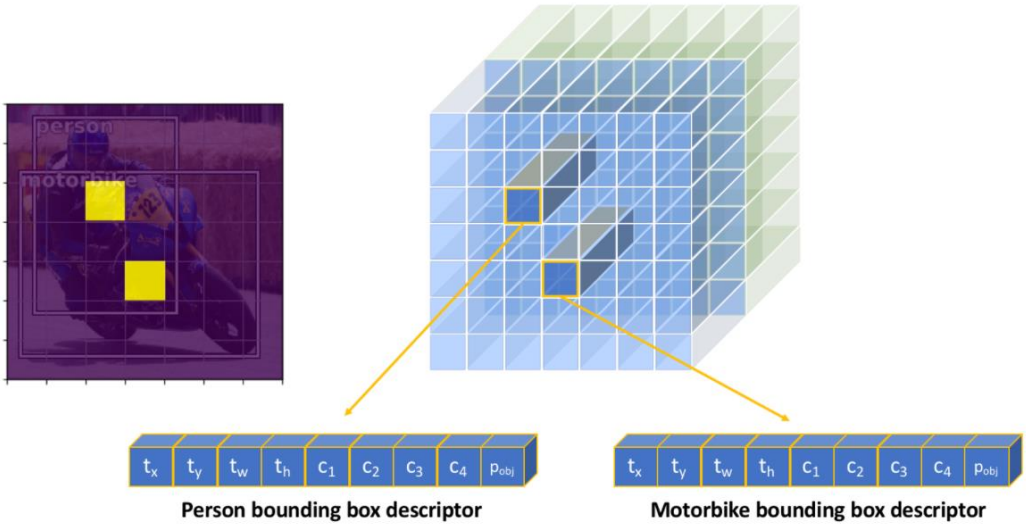
Εικόνα 24: Παράδειγμα εκπαίδευσης 2 πλαισίων οριοθέτησης για ένα κελί του καννάβου

Οπτικοποιώντας το πλήρες συνελικτικό αποτέλεσμα των φίλτρων B(5+C), παρατηρείται ότι το μοντέλο θα παράγει πάντα έναν σταθερό αριθμό προβλέψεων N×N×B για μια δεδομένη εικόνα. Στη συνέχεια φιλτράρονται οι προβλέψεις αυτές, για να ληφθούν υπόψη μόνο τα οριοθετημένα πλαίσια που έχουν πιθανότητα (p_{obj})πάνω από κάποιο καθορισμένο όριο.



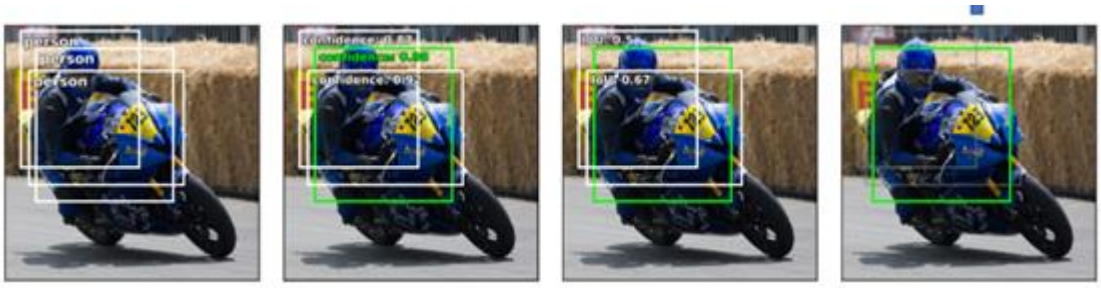
Εικόνα 25: Πλήρες αποτέλεσμα B(5+C) συνελικτικών φίλτρων σε μια εικόνα και φιλτράρισμα των προβλέψεων που παράγονται με όριο p_{obj}

Λόγω της συνελικτικής φύσης της διαδικασίας ανίχνευσης, πολλά αντικείμενα μπορούν να ανιχνευθούν παράλληλα. Ωστόσο, το μοντέλο καταλήγει να προβλέπει ένα μεγάλο αριθμό κελιών καννάβου όπου δεν υπάρχει αντικείμενο. Αν και μπορεί να γίνει φιλτράρισμα για αυτά τα οριοθετημένα πλαίσια με βάση την τιμή p_{obj} τους, αυτό εισάγει μια αρκετά μεγάλη ανισορροπία μεταξύ των οριοθετημένων πλαισίων που περιέχουν ένα αντικείμενο και εκείνων που δεν περιέχουν ένα αντικείμενο και έχουν προβλεφθεί από το μοντέλο.



Εικόνα 26: Παράλληλη πρόβλεψη 2 αντικειμένων

Τα 2 μοντέλα (YOLO και SSD) που θα αναλυθούν παρακάτω χρησιμοποιούν τη λογική «προβλέψεων σε κάνναβο» για να ανιχνεύσουν έναν σταθερό αριθμό πιθανών αντικειμένων μέσα σε μια εικόνα. Η προσέγγιση "προβλέψεων σε κάνναβο" παράγει έναν σταθερό αριθμό προβλέψεων οριοθέτησης για κάθε εικόνα. Ωστόσο, ο στόχος είναι να φιλτραριστούν αυτές οι προβλέψεις προκειμένου να εξαχθούν μόνο πλαίσια οριοθέτησης για αντικείμενα που είναι στην πραγματικότητα πιθανό να βρίσκονται στην εικόνα. Επιπλέον, το μοντέλο πρέπει να καταλήγει σε μια πρόβλεψη ενός πλαισίου οριοθέτησης για κάθε αντικείμενο που ανιχνεύεται. Μπορούν να φιλτραριστούν τα περισσότερα πλαίσια οριοθέτησης που έχουν προβλεφθεί λαμβάνοντας υπόψη μόνο προβλέψεις με p_{obj} πάνω από κάποιο καθορισμένο όριο εμπιστοσύνης. Ωστόσο, μπορεί να υπάρχουν ακόμα πολλές προβλέψεις υψηλής εμπιστοσύνης που περιγράφουν το ίδιο αντικείμενο. Έτσι, χρειάζεται μια μέθοδος για την αφαίρεση των προβλέψεων περιττών αντικειμένων έτσι ώστε κάθε αντικείμενο να περιγράφεται από ένα μόνο πλαίσιο οριοθέτησης. Για να επιτευχθεί αυτό, θα χρησιμοποιηθεί μια τεχνική γνωστή ως μη μέγιστης συμπίεσης. Στη πράξη, αυτή η τεχνική θα εξετάσει τα οριοθετημένα πλαίσια που επικαλύπτονται σε μεγάλο βαθμό και θα απορρίψει όλες τις προβλέψεις εκτός από την πρόβλεψη με το υψηλότερο επίπεδο εμπιστοσύνης.



Εικόνα 27: Εφαρμογή non-maximum suppression

Παρατηρώντας και τη παραπάνω εικόνα, γίνεται αντιληπτό ότι μετά από ένα πρώτο φιλτράρισμα απορρίπτονται οι προβλέψεις με χαμηλή πιθανότητα. Παρόλα αυτά μπορεί να υπάρχουν ακόμα περιττές ανιχνεύσεις. Μετέπειτα επαναληπτικά γίνονται οι παρακάτω διαδικασίες:

- Επιλέγεται το πλαίσιο οριοθέτησης με το μεγαλύτερο ποσοστό εμπιστοσύνης.
- Υπολογίζεται η τομή πάνω στην ένωση (IoU) των εναπομέντων πλαισίων με το επιλεγμένο πλαίσιο.
- Απομακρύνονται όποια πλαίσια έχουν τιμή IoU πάνω από ένα ορισμένο όριο.

Εκτελείται για κάθε κλάση ξεχωριστά η τεχνική της μη μέγιστης συμπίεσης. Ο στόχος και σε αυτή τη περίπτωση είναι να αφαιρεθούν οι περιττές προβλέψεις, επομένως μπορεί να υπάρχουν δύο προβλέψεις που αλληλεπικαλύπτονται εάν το ένα πλαίσιο περιγράφει μια κλάση και το άλλο πλαίσιο περιγράφει μια άλλη κλάση.

YOLO: You Look Only Once

Το μοντέλο YOLO δημοσιεύθηκε για πρώτη φορά από τους Joseph Redmon et al. το 2015 και στη συνέχεια αναθεωρήθηκε σε δύο επόμενες δημοσιεύσεις.

Backbone δίκτυο

Το αρχικό δίκτυο YOLO χρησιμοποιεί τροποποιημένο το δίκτυο GoogLeNet ως βασικό (backbone) δίκτυο. Ο Redmond δημιούργησε αργότερα ένα νέο μοντέλο με το όνομα DarkNet-19 το οποίο ακολουθεί τη γενική σχεδίαση των φίλτρων 3×3, διπλασιάζοντας τον αριθμό των καναλιών σε κάθε συναθροιστικό επίπεδο. Τα φίλτρα 1×1 χρησιμοποιούνται επίσης για την περιοδική συμπίεση της αναπαράστασης χαρακτηριστικών σε όλο το δίκτυο. Η τελευταία του δημοσίευση παρουσιάζει ένα νέο, μεγαλύτερο μοντέλο που ονομάζεται DarkNet-53 που προσφέρει βελτιωμένη απόδοση σε σχέση με την προηγούμενη έκδοση του μοντέλου. Όλα τα προαναφερόμενα μοντέλα εκπαιδεύτηκαν πρώτα ως ταξινομητές εικόνων πριν προσαρμοστούν για τον εντοπισμό αντικειμένων. Στη δεύτερη έκδοση του μοντέλου YOLO, ο Redmond ανακάλυψε ότι η χρήση εικόνων υψηλότερης ανάλυσης στο τέλος της προεκπαίδευσης της ταξινόμησης βελτίωσε την απόδοση ανίχνευσης και έτσι υιοθετήθηκε αυτή η πρακτική. Η προσαρμογή ενός δικτύου ταξινόμησης σε δίκτυο ανίχνευσης αντικειμένων συνίσταται απλώς στην αφαίρεση των τελευταίων επιπέδων του δικτύου και στην προσθήκη ενός συνελικτικού επιπέδου με φίλτρα B(5+C) για την παραγωγή των προβλέψεων των πλαισίων οριοθέτησης N×N×B.

Πλαίσια οριοθέτησης και Πλαισίων Αγκύρωσης

Η πρώτη επανάληψη του μοντέλου YOLO προβλέπει άμεσα και τις τέσσερις τιμές που περιγράφουν ένα πλαίσιο οριοθέτησης. Οι συντεταγμένες x και y κάθε πλαισίου οριοθέτησης ορίζονται σε σχέση με την επάνω αριστερή γωνία κάθε κελιού του καννάβου και κανονικοποιούνται από τις διαστάσεις του κελιού έτσι ώστε οι τιμές των συντεταγμένων να οριοθετούνται μεταξύ 0 και 1. Ο χρήστης ορίζει το πλάτος και το ύψος των πλαισίων έτσι ώστε το μοντέλο να προβλέπει το πλάτος και το ύψος της τετραγωνικής ρίζας. Ορίζοντας το πλάτος και το ύψος των πλαισίων ως η τετραγωνική ρίζα της τιμής , οι διαφορές μεταξύ μεγάλων αριθμών είναι λιγότερο σημαντικές από τις διαφορές μεταξύ μικρών αριθμών. Ο Redmond επέλεξε αυτή τη διατύπωση επειδή "οι μικρές αποκλίσεις στα μεγάλα κουτιά έχουν μικρότερη σημασία από ό,τι σε μικρά κουτιά" και επομένως κατά τον υπολογισμό της συνάρτησης απώλειας ήθελε να δοθεί έμφαση στο να γίνουν πιο ακριβή τα μικρά κουτιά. Το πλάτος και το ύψος του πλαισίου κανονικοποιούνται από το πλάτος και ύψος εικόνας και επομένως οριοθετούνται μεταξύ 0 και 1. Μια απώλεια L2 εφαρμόζεται κατά τη διάρκεια της εκπαίδευσης.

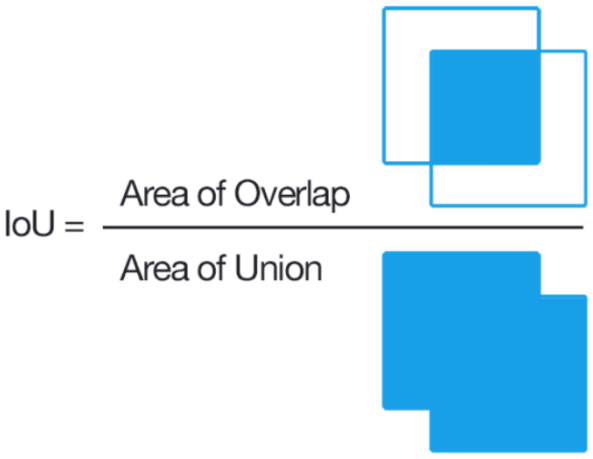
Η παραπάνω διατύπωση αναθεωρήθηκε αργότερα για να εισαχθεί η έννοια του πλαισίου αγκύρωσης. Αντί το μοντέλο να παράγει άμεσα μοναδικούς περιγραφείς οριοθέτησης για κάθε νέα εικόνα, ορίζεται από το χρήστη μια συλλογή οριοθετημένων πλαισίων με ποικίλους λόγους διαστάσεων που ενσωματώνουν κάποιες προηγούμενες πληροφορίες σχετικά με το σχήμα των αντικειμένων που περιμένει να ανιχνεύσει το μοντέλο. Ο Redmond πρόσφερε μια προσέγγιση μέσω της οποίας επιτεύχθηκαν καλύτερες αναλογίες διαστάσεων κάνοντας ομαδοποίηση k-means (προσαρμοσμένη μέτρηση απόστασης) σε όλα τα οριοθετημένα πλαίσια στο σύνολο των δεδομένων εκπαίδευσης. Στην παρακάτω εικόνα, εμφανίζονται μια συλλογή από 5 πλαίσια αγκύρωσης για το κελί του καννάβου που είναι τονισμένο με κίτρινο χρώμα. Με αυτή τη διατύπωση, καθένα από τα B πλαίσια οριοθέτησης ειδικεύεται ρητά στην ανίχνευση αντικειμένων συγκεκριμένου μεγέθους και αναλογίας διαστάσεων.

Αν και δεν οπτικοποιούνται τα πλαίσια αγκύρωσης υπάρχουν για κάθε κελί στο κάνναβο πρόβλεψής. Αντί να προβλέπονται απευθείας οι διαστάσεις του πλαισίου οριοθέτησης, προβλέπεται απλώς η μετατόπιση από τις προηγούμενες διαστάσεις του πλαισίου οριοθέτησης, έτσι ώστε να μπορούν να προσαρμοστούν οι προβλεπόμενες διαστάσεις του πλαισίου οριοθέτησης. Αυτή η αναδιατύπωση καθιστά την εργασία πρόβλεψης πιο εύκολη στην εκπαίδευση.

Εικόνα 28: Πλαίσια αγκύρωσης

«Αντικειμενικότητα»

Στην πρώτη έκδοση του μοντέλου, η τιμή "αντικειμενικότητας" p_{obj} εκπαιδεύτηκε για να προσεγγίζει την τομή πάνω από την ένωση (IoU) μεταξύ του πλαισίου που εξήγαγε το μοντέλο και του ground truth. Όταν υπολογίζεται η απώλειά κατά τη διάρκεια της εκπαίδευσης, θα αντιστοιχίσει αντικείμενα με όποια πρόβλεψη πλαισίου οριοθέτησης (στο ίδιο κελί καννάβου) έχει την υψηλότερη τιμή IoU. Για μη αντιστοιχισμένα πλαίσια, ο μόνος περιγραφέας που θα συμπεριληφθεί στη συνάρτηση απώλειας είναι το p_{obj}. Μετά την προσθήκη των πλαισίων αγκύρωσης στο YOLOv2, αντιστοιχίζονται τα αντικείμενα με ετικέτα σε όποιο πλαίσιο αγκύρωσης έχει την υψηλότερη τιμή IoU με το αντικείμενο με ετικέτα. Στην τρίτη έκδοση, ο Redmond επαναπροσδιορίστηκε η τιμή στόχος του p_{obj} σε 1 για τα οριοθετημένα πλαίσια με την υψηλότερη τιμή του IoU για κάθε δεδομένο στόχο και 0 για όλα τα υπόλοιπα πλαίσια. Ωστόσο, δεν θα συμπεριληφθούν τα οριοθετημένα κουτιά που έχουν υψηλή τιμή IoU (πάνω από κάποιο όριο) αλλά όχι την υψηλότερη τιμή κατά τον υπολογισμό της απώλειας. Στην ουσία αν μια πρόβλεψη είναι καλή συμπεριλαμβάνεται παρότι μπορεί να μην είναι η καλύτερη.



Εικόνα 29: Σχηματικά ο υπολογισμός της έννοιας IoU η τομή πάνω από την ένωση

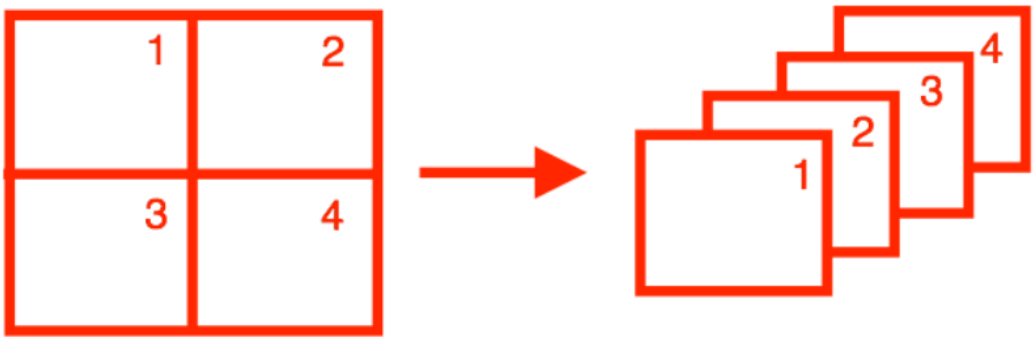
Ετικέτες των κλάσεων

Αρχικά, η πρόβλεψη κλάσης πραγματοποιούνταν σε επίπεδο κελιού καννάβου. Αυτό σήμαινε ότι ένα μεμονωμένο κελί καννάβου δεν μπορούσε να προβλέψει πολλαπλά πλαίσια οριοθέτησης διαφορετικών κλάσεων. Αυτό αναθεωρήθηκε αργότερα για να προβλέπει την κλάση για κάθε πλαίσιο οριοθέτησης χρησιμοποιώντας μια ενεργοποίηση softmax μεταξύ των κλάσεων και μια απώλεια διασταυρούμενης εντροπίας. Ο Redmond άλλαξε αργότερα την πρόβλεψη της κλάσης για να χρησιμοποιήσει σιγμοειδείς ενεργοποιήσεις για ταξινόμηση πολλαπλών ετικετών, καθώς διαπίστωσε ότι η softmax δεν απαραίτητα η μόνη επιλογή για την καλή απόδοση του μοντέλου.

Αυτή η επιλογή θα εξαρτηθεί από το σύνολο δεδομένων που χρησιμοποιούνται και από το εάν οι ετικέτες τους επικαλύπτονται ή όχι.

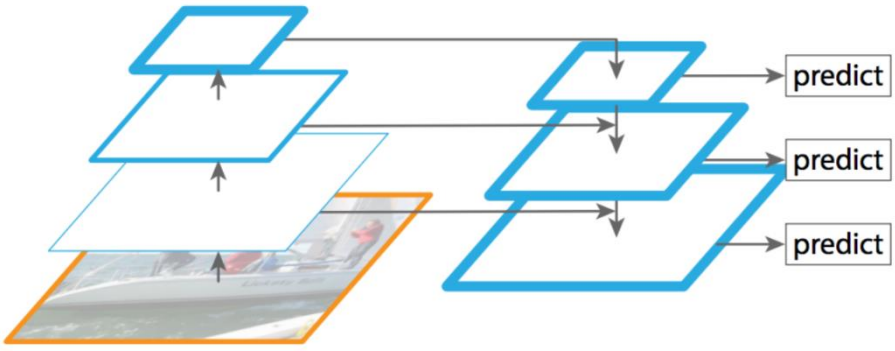
Επίπεδο εξόδου

Το πρώτο μοντέλο YOLO απλώς προβλέπει τα N×N×B πλαίσια χρησιμοποιώντας το αποτέλεσμα του δικτύου του κορμού (backbone). Στο YOLOv2, ο Redmond προσθέτει μια σύνδεση παράβλεψης χωρίζοντας έναν χάρτη χαρακτηριστικών υψηλότερης ανάλυσης σε πολλά κανάλια όπως απεικονίζεται παρακάτω.



Εικόνα 30: Η σύνδεση παράβλεψης στο YOLOv2

Η σύνδεση παράβλεψης άλλαξε στην τρίτη έκδοση του μοντέλου αντικαταστάθηκε από μια δομή εξόδου δικτύου πυραμίδας. Με αυτήν τη μέθοδο, γίνεται εναλλαγή μεταξύ της παραγωγής μιας πρόβλεψης και της δειγματοληψίας στους χάρτες χαρακτηριστικών (με συνδέσεις παράλειψης). Αυτό επιτρέπει οι προβλέψεις που μπορούν να επωφεληθούν από πιο λεπτομερείς πληροφορίες από νωρίτερα στο δίκτυο, βοηθώντας στον εντοπισμό μικρών αντικειμένων στην εικόνα.



Εικόνα 31: Δομή πυραμίδας στο επίπεδο εξόδου του YOLOv3

SSD: Single Shot Detection

Το μοντέλο SSD δημοσιεύθηκε από τους Wei Liu et al. το 2015, λίγο μετά το μοντέλο YOLO, και βελτιώθηκε αργότερα σε μια επόμενη δημοσίευση.

Backbone Δίκτυο

Ένα μοντέλο VGG-16, προεκπαιδευμένο στο ImageNet για ταξινόμηση εικόνων, χρησιμοποιείται ως βασικό δίκτυο. Τροποποιώντας το VGG-16 κατά την προσαρμογή του μοντέλου στην διαδικασία ανίχνευσης, με ενέργειες όπως: αντικατάσταση πλήρως συνδεδεμένων επιπέδων με συνελικτικές υλοποιήσεις, αφαίρεση επιπέδων dropout και αντικατάσταση του τελευταίου συναθροιστικού στρώματος (max pooling) με διεσταλμένη συνέλιξη.

Πλαίσια οριοθέτησης και Πλαίσια Αγκύρωσης

Το μοντέλο SSD ορίζει χειροκίνητα μια συλλογή αναλογιών διαστάσεων (π.χ. {1, 2, 3, 1/2, 1/3}) που θα χρησιμοποιηθούν για τα B πλαίσια οριοθέτησης σε κάθε θέση κελιού του καννάβου. Για κάθε πλαίσιο οριοθέτησης, θα προβλεφθούν οι μετατοπίσεις από το πλαίσιο αγκύρωσης τόσο για τις συντεταγμένες του πλαισίου οριοθέτησης (x και y) όσο και για τις διαστάσεις (πλάτος και ύψος). Χρησιμοποιούνται ενεργοποιήσεις ReLU που έχουν εκπαιδευτεί με απώλεια L1.

«Αντικειμενικότητα»

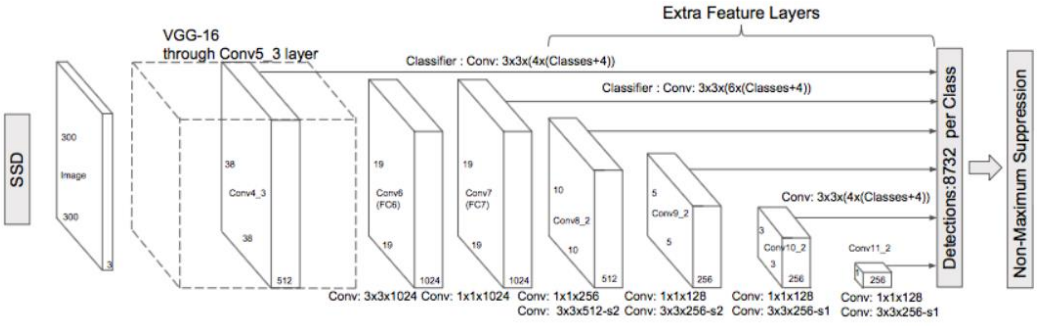
Μια σημαντική διαφορά μεταξύ YOLO και SSD είναι ότι το SSD δεν προσπαθεί να προβλέψει μια τιμή για το p_{obj}. Ενώ το μοντέλο YOLO προέβλεψε την πιθανότητα ενός αντικειμένου και στη συνέχεια προέβλεψε την πιθανότητα κάθε κλάσης δεδομένου ότι υπήρχε ένα αντικείμενο, το μοντέλο SSD προσπαθεί να προβλέψει άμεσα την πιθανότητα να υπάρχει μια κλάση σε ένα δεδομένο πλαίσιο οριοθέτησης. Κατά τον υπολογισμό της απώλειας, θα αντιστοιχίσει κάθε πλαίσιο ground truth με το πλαίσιο αγκύρωσης με το υψηλότερο IoU — ορίζοντας αυτό το πλαίσιο ως "υπεύθυνο" για την πραγματοποίηση της πρόβλεψης. Ωστόσο, θα ταιριάξει επίσης τα πλαίσια του ground truth με οποιαδήποτε άλλα πλαίσια αγκύρωσης με IoU πάνω από κάποιο καθορισμένο όριο (0,5) υπό το ίδιο πρίσμα που συμβαίνει και στο YOLO. Τέλος η μη μέγιστη συμπίεση φιλτράρει τις περιττές προβλέψεις.

Ετικέτες των κλάσεων

Στο SSD οι προβλέψεις των κλάσεων για τα κουτιά οριοθέτησης δεν εξαρτώνται από το γεγονός ότι υπάρχει ένα αντικείμενο. Έτσι, προβλέπεται άμεσα η πιθανότητα κάθε κλάσης χρησιμοποιώντας μια ενεργοποίηση softmax και απώλεια διασταυρούμενης εντροπίας. Επειδή δεν προβλέπεται ρητά το p_{obj}, είναι σημαντικό να υπάρχει μια κλάση για το "background" ώστε να μπορεί να προβλέψει όταν δεν υπάρχει αντικείμενο. Λόγω του γεγονότος ότι τα περισσότερα από τα πλαίσια θα ανήκουν στην κατηγορία "background", χρησιμοποιείται μια τεχνική γνωστή ως "hard negative mining" για να δοκιμαστούν αρνητικές (χωρίς αντικείμενο) προβλέψεις έτσι ώστε να υπάρχει το πολύ μια αναλογία 3:1 μεταξύ αρνητικών και θετικών προβλέψεων κατά τον υπολογισμό της απώλειας.

Επίπεδο εξόδου

Το SSD επιτρέπει προβλέψεις σε πολλαπλές κλίμακες, το επίπεδο εξόδου του SSD μειώνει σταδιακά τους χάρτες συνελικτικών χαρακτηριστικών, παράγοντας κατά διαστήματα προβλέψεις πλαισίων οριοθέτησης (όπως φαίνεται με τα βέλη στην παρακάτω εικόνα).



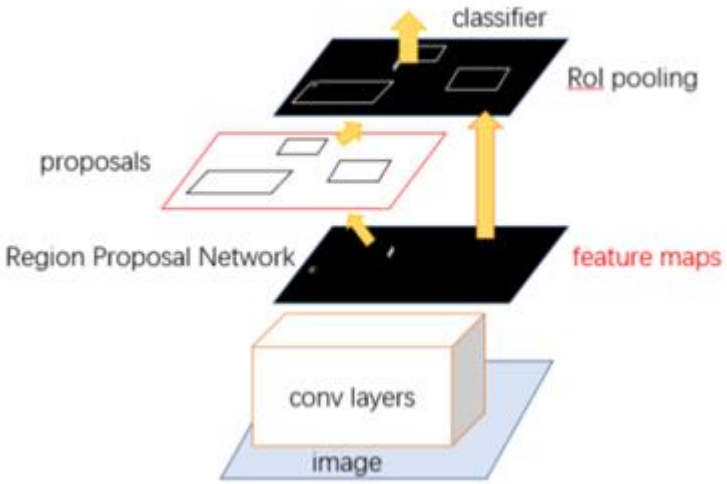
Εικόνα 32: Σχηματική αναπαράσταση του μοντέλου SSD

Τεχνικές δυο σταδίων

Οι αλγόριθμοι δύο σταδίων περιλαμβάνουν δύο στάδια, όπως αναφέρει και το όνομα τους. Στο πρώτο στάδιο, πραγματοποιούνται πρώτα προκαταρκτικές δοκιμές, όλα τα θετικά δείγματα ελέγχονται και δημιουργούνται περιοχές ενδιαφέροντος (RoIs). Στο δεύτερο στάδιο, ο αλγόριθμος εκτελεί περιφερειακή ταξινόμηση και βελτίωση της θέσης στις RoI που δημιουργήθηκαν στο προηγούμενο στάδιο. Κοινοί αλγόριθμοι δύο σταδίων είναι οι R-CNN, Fast R-CNN, Faster R-CNN, R-FCN, FPN κ.λπ. Ο Faster R-CNN αλγόριθμος εισάγει ένα Δίκτυο Προτάσεων Περιοχής (RPN) επίσης βασίζεται στο Fast R-CNN, και ενσωματώνει τα δύο αυτά στοιχεία σε ένα πλήρες δίκτυο που μπορεί να εκπαιδευτεί από άκρο σε άκρο, το οποίο όχι μόνο εγγυάται την ακρίβεια αλλά και αυξημένη ταχύτητα. Ο αλγόριθμος FPN είναι μια μέθοδος που χρησιμοποιεί συμβατικά μοντέλα CNN για την αποτελεσματική εξαγωγή χαρακτηριστικών διαφόρων διαστάσεων από μια εικόνα, και είναι βελτίωση του παραδοσιακού δικτύου CNN για να εκφράσει και να εξάγει πληροφορίες από μια εικόνα.

Αρχιτεκτονική του Faster R-CNN

Το Faster R-CNN είναι η αρχιτεκτονική που βελτιώνει περαιτέρω τις αρχιτεκτονικές R-CNN και Fast R-CNN τόσο από πλευρά ταχύτητας εκπαίδευσης όσο και όσο και από πλευράς ακρίβειας στον εντοπισμό. Το παρουσίασαν οι Shaoqing Ren, et al. στο Microsoft Research το 2016 με τίτλο «Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks». Είναι ένα πλήρες πλαίσιο ανίχνευσης αντικειμένων δύο σταδίων που αντικατέστησε τη μέθοδο επιλεκτικής αναζήτησης του R-CNN με ένα δίκτυο πρόταση περιοχής (RPN), με το πλεονέκτημα της υψηλότερης ταχύτητας ανίχνευσης. Το RPN χρησιμοποιεί ένα παράθυρο 3 × 3 το οποίο μετακινείται πάνω στον τελικό χάρτη χαρακτηριστικών και σε κάθε θέση του παραθύρου θεωρεί k διαφορετικά πλαίσια αγκύρωσης με επίκεντρο την τοποθεσία για τη δημιουργία πιθανών προτάσεων περιοχής. Κάθε πρόταση περιοχής αποτελείται από μια τιμή αντικειμενικότητας για αυτήν την περιοχή καθώς και από τις συντεταγμένες των πλαισίων. Οι προτάσεις φιλτράρονται με βάση το όριο αντικειμενικότητας και περνούν στο στάδιο της ανίχνευσης αντικειμένων. Γενικά, το Faster R-CNN μπορεί να θεωρηθεί ως Fast R-CNN που στην κορυφή του τοποθετείται ένα RPN. Το Faster R-CNN δίνει πρόβλεψη με πλαίσια οριοθέτησης, που αποτελούνται από τρεις κλίμακες και τρεις λόγους διαστάσεων και χρησιμοποιεί ένα επίπεδο ανίχνευσης. Τα CNN που χρησιμοποιούνται ως βάση μπορεί να είναι τα ZFNet και VGGNet (VGG-16) που έχουν 7 και 16 επίπεδα αντίστοιχα.



Εικόνα 33: Διάγραμμα Δομής του μοντέλου Faster R-CNN

Το Faster R-CNN χωρίζεται κυρίως στα ακόλουθα μέρη:

1. Επίπεδο συνέλιξης: Εξάγει τον χάρτη χαρακτηριστικών ολόκληρης της εικόνας.
2. Δίκτυο πρότασης περιοχών (RPN): Ως το βασικό μέρος του Faster R-CNN, το RPN αντικαθιστά τη μέθοδο επιλεκτικής αναζήτησης στον αλγόριθμο Fast R-CNN για τη δημιουργία μιας πρότασης περιοχής. Η χρήση του RPN για τη λήψη μιας πρότασης περιοχής μπορεί να χρησιμοποιήσει πιο γρήγορα και αποτελεσματικά

το δίκτυο CNN. Το RPN δημιουργεί αγκυρώσεις ενώ δημιουργεί μια πρόταση περιοχής. Η συνάρτηση κρίσης καθορίζει εάν οι αγκυρώσεις είναι στο προσκήνιο ή στο παρασκήνιο και, στη συνέχεια, προσαρμόζει τις αγκυρώσεις μέσω παλινδρόμησης των ορίων για να δώσει μια πρόταση περιοχής με ακρίβεια.

3. Συναθροιστικό επίπεδο RoI: Αντιμετωπίζει το πρόβλημα ότι διαφορετικά μεγέθη χαρτών χαρακτηριστικών εισέρχονται στο δίκτυο με πλήρως συνδεδεμένα επίπεδα. Το σταθερό μέγεθος λαμβάνεται με αναδειγματοληψία.
4. Επίπεδο ταξινόμησης και επίπεδο παλινδρόμησης: Το επίπεδο ταξινόμησης είναι υπεύθυνο για να κρίνει σε ποια κατηγορία ανήκει ένα αντικείμενο. Το στρώμα παλινδρόμησης προσαρμόζει τις θέσεις των περιοχών ενδιαφέροντος (RoIs) για να εξαχθεί το τελικό αποτέλεσμα ανίχνευσης αντικειμένου.

2.9 Κίνητρα και Περιορισμοί στην Επεξεργασία Εικόνας για την Ανίχνευση Φρούτων

Η βαθιά μηχανική μάθηση (Deep Learning) αποτελεί μια πρόσφατη, σύγχρονη τεχνική επεξεργασίας εικόνας και ανάλυσης δεδομένων, με πολλά υποσχόμενα αποτελέσματα και μεγάλες δυνατότητες. Καθώς η βαθιά μηχανική μάθηση έχει εφαρμοστεί ήδη με επιτυχία σε διάφορους τομείς, πρόσφατα έχει αρχίσει να χρησιμοποιείται επίσης στον τομέα της γεωργίας. Οι ερευνητικές προσπάθειες που μελετήθηκαν αναδεικνύουν τις δυνατότητες των CNNs τόσο στην ταξινόμηση όσο και στην ανίχνευση και τη σημασία να επεκταθεί η έρευνα πάνω στα CNNs για τα πολλαπλά οφέλη που θα επιφέρει στο γεωργικό τομέα. Ταυτόχρονα όπως σε κάθε ερευνητικό πεδίο προκύπτουν και περιορισμοί προσπαθώντας να εφαρμοστούν τα αποτελέσματα της έρευνας πάνω στα CNNs σε εφαρμογές της πραγματικής ζωής. Παρακάτω παρουσιάζονται τόσο κίνητρα που ωθούν τις ερευνητικές προσπάθειες να εξελιχθούν, όσο και περιορισμοί που είναι αντικείμενο για περαιτέρω εργασία ώστε να ξεπεραστούν.

Κίνητρα

Η ανίχνευση φρούτων είναι ένα κρίσιμο συστατικό για τον αυτοματισμό της γεωργίας σε συνθήκες χωραφιού σύμφωνα με τους *Bargoti and Underwood (2017)*. Με την ακριβή γνώση των μεμονωμένων θέσεων των φρούτων στο χωράφι, είναι δυνατή η εκτίμηση της σοδειάς και η χαρτογράφηση του, οι οποίες είναι σημαντικές διαδικασίες για τους καλλιεργητές καθώς διευκολύνουν την αποτελεσματική χρήση των πόρων (λίπανση, πότισμα)και με αυτό το τρόπο μπορεί να γίνει ζωνοποίηση της καλλιέργειας και να γίνονται στοχευμένες επεμβάσεις σε αυτή βελτιώνοντας την παραγωγικότητα ανά περιοχή και εξοικονομώντας πόρους. Επίσης ο ακριβής εντοπισμός των φρούτων είναι απαραίτητο συστατικό για την ανάπτυξη ενός αυτοματοποιημένου ρομποτικού συστήματος συγκομιδής, το οποίο μπορεί να βοηθήσει στην άμβλυνση των κοπιαστικών και εντατικών εργασιών σε έναν οπωρώνα.

Στην δημοσίευση των *Xu Liu et al.(2018)* που στόχο είχε την καταμέτρηση καρπών φρούτων καταλήγουν ότι η μεθοδολογία τους με CNNs θα μπορούσε να επεκταθεί για την ανίχνευση, την παρακολούθηση και την καταμέτρηση ενός πλήθους άλλων «ακίνητων» χαρακτηριστικών ενδιαφέροντος όπως οι κηλίδες των φύλλων, ο μαρασμός και η άνθηση. Επιπρόσθετα τονίζουν ότι η δυνατότητα λήψης αριθμών καρπών από βίντεο, όπως ήταν και το αντικείμενο τους, επιτρέπει στους καλλιεργητές να βελτιστοποιούν καλύτερα τις αποφάσεις διαχείρισης και συγκομιδής όπως τη κατανομή εργασίας στο πεδίο, την αποθήκευση, τη συσκευασία και τον προγραμματισμός της συγκομιδής.

Επίσης πέρα από τα παραπάνω που αναδεικνύουν τα οφέλη που θα φέρει η αυτοματοποίηση της γεωργίας μέσω της τεχνικής των CNNs, ως κίνητρο μπορεί να λειτουργήσει και υπεροχή που φαίνεται να παρουσιάζουν οι αποδόσεις των CNNs έναντι πιο παραδοσιακών μεθόδων. Σύμφωνα με τους *Bargoti and Underwood (2017),* συνήθως, οι προηγούμενες τεχνικές χρησιμοποιούσαν χαρακτηριστικά σχεδιασμένα με το χέρι για να κωδικοποιήσουν οπτικά χαρακτηριστικά που διακρίνουν τα φρούτα από τις περιοχές των εικόνων που δεν περιέχουν φρούτα. Αν και αυτές οι προσεγγίσεις είναι κατάλληλες για το σύνολο δεδομένων για το οποίο έχουν σχεδιαστεί, η κωδικοποίηση χαρακτηριστικών είναι γενικά μοναδική για ένα συγκεκριμένο φρούτο και τις συνθήκες κάτω από τις οποίες καταγράφηκαν τα δεδομένα και δεν μπορεί να γενικευτεί. Πρόσφατα, η πρόοδος στην κοινότητα της όρασης υπολογιστών που έχει μεταφραστεί στην γεωργία σε agrovision (όραση υπολογιστών στη γεωργία), επιτυγχάνει υψηλά αποτελέσματα με τη χρήση Deep Neural Networks (DNN) για την ανίχνευση αντικειμένων και τη σημασιολογική κατάτμηση εικόνων. Τα DNN αποφεύγουν την ανάγκη για χαρακτηριστικά που έχουν σχεδιαστεί με το χέρι αλλά εκπαιδεύονται αυτόματα σε αναπαραστάσεις χαρακτηριστικών που καταγράφονται διακριτά μέσα στη κατανομή των δεδομένων εκπαίδευσης.

Προκλήσεις

Οι προσεγγίσεις που βασίζονται στα CNNs θα πρέπει να αντιμετωπίσουν σημαντικές προκλήσεις προκειμένου να τις εφαρμόσουν σε σενάρια πραγματικού κόσμου. Συνήθεις προκλήσεις που συναντά το αντικείμενο ανίχνευσης των φρούτων είναι ότι η ανίχνευση σε πραγματικές συνθήκες καλλιέργειας περιλαμβάνει μεταβλητότητα στο φωτισμό, αποκρύψεις από τα φυλλώματα των δέντρων καθώς και αποκρύψεις μεταξύ των φρούτων με αποτέλεσμα καρποί είτε να μην ανιχνεύονται είτε να διπλό-μετριούνται. Επιπρόσθετα τα δεδομένα που αποτελούνται από εικόνες υπαίθριου οπωρώνα παρουσιάζουν πρόσθετες προκλήσεις για την ανίχνευση φρούτων. Για αποτελεσματική λειτουργία σε μεγάλης κλίμακας δεδομένα όπως ένα οπωρώνας, το οπτικό πεδίο του αισθητήρα πρέπει να καλύπτει ολόκληρα δέντρα, με αποτέλεσμα να απαιτούνται εικόνες με υψηλή ανάλυση για το σχετικά μικρό μέγεθος του καρπού των φρούτων. Πιο τεχνικοί περιορισμοί παρουσιάζονται παρακάτω:

1. Μέγεθος του σετ δεδομένων - Το σετ δεδομένων πρέπει να είναι επαρκές σε μέγεθος και καλά επισημασμένο για να εκπαιδεύσει το CNN. Ώστε το μοντέλο να αποφύγει το πρόβλημα της υπερπροσαρμογής στα δεδομένα και να εκτελέσει της αποτελεσματικά το στόχο του. Ως εκ τούτου, η προετοιμασία του σετ των δεδομένων είναι μια χρονοβόρα και απαιτητική διαδικασία για την εκπαίδευση ενός CNN. Παρόλο που υπάρχει μια μεγάλη πληθώρα βάσεων δεδομένων που προτείνονται από τη διεθνή βιβλιογραφία, δεν είναι διαθέσιμες όλες, για το λόγο αυτό, η αναπαραγωγιμότητα όλων των μελετών δεν είναι πάντα εφικτή. Επιπλέον, σε πολλές περιπτώσεις, οι βάσεις δεδομένων συλλέγονται ανάλογα με την με το σκοπό χειροκίνητα.
2. Ορισμός παραμέτρων CNN - Ο αριθμός των επιπέδων και των φίλτρων όταν προτείνεται μια αρχιτεκτονική CNN για ένα συγκεκριμένο πρόβλημα, καθώς και ο προσδιορισμός των παραμέτρων και των υπερπαραμέτρων του μοντέλου, παραμένει ένα σχετικό πρόβλημα που συνήθως λύνεται δοκιμάζοντας και αποτυγχάνοντας (trial and error) μέχρι το μοντέλο να ρυθμιστεί κατάλληλα. Η διαδικασία αυτή είναι πολύ χρονοβόρα για μοντέλα με βαθιά αρχιτεκτονική (πολλά επίπεδα). Σε αυτό το πρόβλημα, τα προ-εκπαιδευμένα CNN μοντέλα παρέχουν μεγάλη βοήθεια, καθώς μπορούν να ληφθούν ως βασικός άξονας για την κατασκευή άλλων CNNS.

2.10 Πρόσφατη Βιβλιογραφία

Στη παρούσα διπλωματική εργασία μελετήθηκαν τα συνελικτικά νευρωνικά δίκτυα, καθώς και η πρόσφατη βιβλιογραφία όσον αφορά μεθόδους ταξινόμησης και ανίχνευσης φρούτων μέσω δικτύων CNN για εφαρμογές στην γεωργία. Σκοπός ήταν να μελετηθούν, κατανοηθούν και αναλυθούν οι πιο διαδεδομένες αρχιτεκτονικές των σχετικών δικτύων, καθώς και να διερευνηθεί η επάρκεια ελεύθερων σετ δεδομένων για πειραματισμό. Παρακάτω παρατίθενται στοιχεία από δημοσιεύσεις όπου συνέβαλαν στην ανάπτυξη της μεθοδολογίας της παρούσας εργασίας.

Το πρόβλημα του υπολογισμού του αριθμού των φρούτων από εικόνες βίντεο μπορεί να αντιμετωπιστεί χρησιμοποιώντας παρακολούθηση του αντικειμένου μέσω ανίχνευσης, που αποτελείται από δύο φάσεις: ανίχνευση φρούτων, ακολουθούμενη από παρακολούθηση και μέτρηση τους. Στη δημοσίευση των *Xu Liu et al. (2018)* περιγράφεται μια διαδικασία βημάτων για την καταμέτρηση φρούτων, που συνδυάζει την κατάτμηση με βαθιά μηχανική μάθηση, την παρακολούθηση από καρέ σε καρέ και τον τρισδιάστατο εντοπισμό για την ακριβή μέτρηση φρούτων που είναι ορατά σε μια ακολουθία εικόνων. Η μεθοδολογία εφαρμόστηκε σε εικόνες σε σειρά από κάμερα με ένα φακό, τραβηγμένες τόσο σε φυσικό φως όσο και σε τεχνητό κατά τη διάρκεια της νύχτας. Πρώτα εκπαιδεύτηκε ένα πλήρως συνδεδεμένο συνελικτικό δίκτυο, ώστε να εκτελεί κατάτμηση σε κάθε εικονοστοιχείο / pixel της εικόνας –frame του βίντεο σε pixel με ή χωρίς φρούτα. Στη συνέχεια η παρακολούθηση των φρούτων στις εικόνες γίνεται χρησιμοποιώντας τον αλγόριθμο Hungarian, όπου το αντικειμενικό κόστος υπολογίζεται μέσω του φίλτρου Kalman και διορθώνεται μέσω του Kanade-Lucas-Tomasi (KLT) Tracker. Προκειμένου να διορθωθεί το εκτιμώμενο πλήθος των φρούτων από τη διαδικασία παρακολούθησης, συνδυάζεται η παρακολούθηση με τα αποτελέσματα του αλγόριθμου Structure from Motion (SfM) που υπολογίζει τις σχετικές θέσεις στον χώρο των φρούτων και το μέγεθος τους με σκοπό να απορριφθούν ακραίες τιμές και διπλά μετρημένα φρούτα. Τα αποτελέσματα της έρευνας των *Xu Liu et al.(2018)* έδειξαν ότι η μεθοδολογία που αναπτύχθηκε είναι ικανή να μετράει με ακρίβεια και αξιοπιστία τα φρούτα σε αλληλουχίες εικόνων, και το στάδιο τη διόρθωσης που

εφαρμόστηκε μπορεί να βελτιώσει σημαντικά την καταμέτρηση με ακρίβεια και σταθερότητα. Η βαθιά μηχανική μάθηση εφαρμόζεται στο πρώτο βήμα της παραπάνω μεθοδολογίας όπου χρησιμοποιεί ένα σύνολο εικόνων ως δεδομένα εκπαίδευσης για την εκπαίδευση ενός Πλήρως Συνδεδεμένου Συνελικτικού Δικτύου (FCN) που εκτελεί κατάτμηση σε μεμονωμένες εικόνες από το βίντεο σε εικονοστοιχεία με φρούτα ή χωρίς. Το FCN διαφέρει από τα κλασικά συνελικτικά δίκτυα επειδή δεν αξιοποιεί τα πλήρως συνδεδεμένα επίπεδα του δικτύου. Αξιοποιεί μόνο τα συνελικτικά και αποσυνελικτικά στρώματα επιτρέποντας στο δίκτυο να εκτελεί κατάτμηση αντί για ταξινόμηση ή παλινδρόμηση. Το FCN λαμβάνει ως δεδομένο εισόδου μια εικόνα και βγάζει ως αποτέλεσμα έναν χάρτη τιμών διαστάσεων h × w × n, όπου h και w είναι το ύψος και το πλάτος της εικόνας και n είναι ο αριθμός των κλάσεων. Κάθε pixel έχει τη δική του τιμή συμμετοχής στην κάθε κλάση. Για παράδειγμα, η τιμή για το pixel (i, j) που ανήκει στην κλάση k αντιπροσωπεύεται από το στοιχείο x_{ijk} στον χάρτη τιμών. Το FCN που χρησιμοποιείται στη δημοσίευση βασίζεται στη δομή του δικτύου VGG που χρησιμοποιείται στην ταξινόμηση εικόνων . Η κύρια διαφορά είναι ότι το δίκτυό της δημοσίευσης έχει 2 κατηγορίες (φρούτα και όχι-φρούτα) αντί για τις αρχικές 21 κατηγορίες του VGG. Χρησιμοποιώντας την ίδια αρχιτεκτονική η εκπαίδευση του μοντέλου είναι γρήγορη καθώς μπορούν να αξιοποιηθούν τα βάρη του δικτύου VGG για να αρχικοποιηθούν τα βάρη του FCN (transfer learning & fine tuning).

Η δημοσίευση των *Bargoti and Underwood (2017)* παρουσιάζει τη χρήση, της πρόσφατης και δημοφιλούς αρχιτεκτονικής για την ανίχνευση αντικειμένων, Faster R-CNN, στο πλαίσιο της ανίχνευσης φρούτων σε οπωρώνες, για τις κατηγορίες μάνγκο, αμύγδαλα και μήλα. Στη δημοσίευση τους εμβαθύνουν στην αρχιτεκτονική που χρησιμοποιούν, στο πόσα δεδομένα εκπαίδευσης απαιτούνται για να αντιπροσωπεύεται η μεταβλητότητα του συνόλου των δεδομένων, στις τεχνικές επαύξησης δεδομένων που απέδωσαν σημαντικά, με αποτέλεσμα μεγαλύτερη από δύο φορές μείωση του αριθμού των απαιτούμενων εικόνων εκπαίδευσης. Σε αντίθεση, αναφέρουν ότι στη συγκεκριμένη προσπάθεια το transfer learning μεταξύ οπωρώνων συνέβαλε αμελητέα στην απόδοση του δικτύου σε σχέση με την αρχικοποίηση των βαρών απευθείας από τις δυνατότητες του ImageNet. Επίσης προτείνουν τεχνικές τροποποίησης των αρχικών εικόνων για την εκτέλεση ανίχνευσης σε δεδομένα υψηλής ανάλυσης που περιέχουν περισσότερα από 1000 αντικείμενα το καθένα, με τη λογική κάθε αρχική εικόνα να χωριστεί σε μικρότερες και να τροφοδοτήσει το μοντέλο του Faster R-CNN. Η μελέτη τους είχε απόδοση στην ανίχνευση F1 > 0,9 για τα μήλα και τα μάνγκο σε εικόνες από πραγματικές συνθήκες χωραφιού, που χαρακτηρίζεται πολύ υψηλή σε σχέση και με άλλες σχετικές προσπάθειες. Κατά τη διάρκεια της εκπαίδευσης, τα δεδομένα εισόδου για το μοντέλο είναι μια έγχρωμη εικόνα (rgb) 3 καναλιών μαζί με τα πλαίσια οριοθέτησης /bounding boxes τριγύρω από κάθε φρούτο επισημασμένα με το είδος του φρούτου. Σε αυτή την εργασία επιλέχθηκαν να χρησιμοποιηθούν τα εξής CNN μοντέλα το μοντέλο ZF, το οποίο περιέχει 5 συνελικτικά επίπεδα και το βαθύτερο μοντέλο VGG16, που περιέχει 13 συνελικτικά επίπεδα. Και στη συγκεκριμένη προσπάθεια χρησιμοποιήθηκε η τεχνική του transfer learning/ fine tuning, δηλαδή τα μοντέλα ήταν προεκπαιδευμενα και δεν χρειάστηκε η αρχικοποίηση των βαρών τους εξ αρχής. Συγκεκριμένα είχαν εκπαιδευτεί στο ImageNet που είναι σύνηθες να χρησιμοποιείται σαν βάση καθώς περιλαμβάνει 1000 κατηγορίες και 1.2 εκατομμύρια εικόνες. Η τεχνική επαύξησης δεδομένων που χρησιμοποιήθηκε στη συγκεκριμένη δημοσίευση είναι η PCA που είχε χρησιμοποιηθεί και στο AlexNet, που στην ουσία γίνεται ένα χρωματικός μετασχηματισμός στις εικόνες. Τέλος να αναφερθεί ότι το σετ δεδομένων τους υπάρχει διαθέσιμο στο internet και μπορεί κανείς να πειραματιστεί πάνω σε αυτό.

Στο άρθρο των *A.Koirala et al.(2019)* συγκρίνονται οι αποδόσεις έξι υπαρχουσών αρχιτεκτονικών βαθιάς μηχανικής μάθησης με σκοπό να εντοπιστούν μάνγκο σε εικόνες από πραγματικές συνθήκες καλλιέργειας. Συλλέχθηκαν 1515 εικόνες δέντρων από 5 διαφορετικές καλλιέργειες μάνγκο. Η απεικόνιση ενός οπωρώνα πραγματοποιείται με την συλλογή δύο εικόνων ανά δέντρο από αντίθετες πλευρές. Η κάμερα που χρησιμοποιήθηκε ήταν ψηφιακή 5 Mega-pixel RGB, τοποθετημένη πάνω σε αγροτικό όχημα που κινούνταν σταθερά με 6 km/h. Οι λήψεις έγιναν νύχτα με τεχνητό φωτισμό. Η χρήση νυχτερινής απεικόνισης επιτρέπει σταθερές συνθήκες φωτισμού σε σύγκριση με τη διακύμανση που παρατηρείται στο φως της ημέρας λόγω της γωνίας του ήλιου και των καιρικών συνθηκών. Στη συγκεκριμένη δημοσίευση εκπαιδεύτηκαν οι μέθοδοι δύο σταδίων του Faster R-CNN(VGG) και του Faster R-CNN(ZF) και οι τεχνικές ενός σταδίου YOLOv3, YOLOv2, YOLOv2(tiny) και SSD. Οι παραπάνω αρχιτεκτονικές εκπαιδεύτηκαν τόσο στις εικόνες του σετ δεδομένων με αρχική ανάλυση όσο και σε εκδόσεις των εικόνων με χαμηλότερη ανάλυση με αποτέλεσμα να υλοποιηθούν συνολικά έντεκα μοντέλα. Αναπτύχθηκε επίσης μια νέα αρχιτεκτονική, βασισμένη στα χαρακτηριστικά των YOLOv3 και YOLOv2 (tiny) και προσαρμοσμένη στα κριτήρια σχεδιασμού, της ακρίβειας και

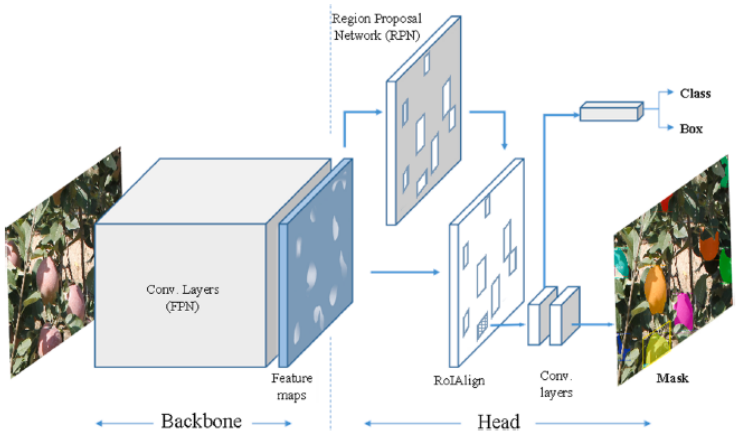
της ταχύτητας για την τρέχουσα εφαρμογή. Αυτή η αρχιτεκτονική, ονομάστηκε «MangoYOLO», εκπαιδεύτηκε χρησιμοποιώντας: (i) το σετ εκπαίδευσης 1300 πλακιδίων, (ii) το σύνολο δεδομένων COCO πριν από την εκπαίδευση στο σετ εκπαίδευσης με μάνγκο και (iii) ένα σετ εκπαίδευσης με εικόνες κατά τη διάρκεια της ημέρας μιας προηγούμενης δημοσίευσης.
Με το τρόπο αυτό δημιουργήθηκαν 3 εκδόσεις/μοντέλα του MangoYOLO 's', 'pt' και 'bu', αντίστοιχα. Το Mango YOLO ‘s’ εκπαιδεύτηκε εξ αρχής στα δεδομένα εκπαίδευσης εικόνων με Mango, δηλαδή η αρχικοποίηση των βαρών στα συνελικτικά φίλτρα ήταν τυχαία. Στο Mango YOLO ‘pt’ χρησιμοποιήθηκε η τεχνική του transfer learning από το coco dataset. Τέλος το Mango YOLO ‘bu’ εκπαιδεύτηκε και αυτό εξ αρχής στα δεδομένα εκπαίδευσης εικόνων ημέρας μιας προηγούμενης δημοσίευσης με μάνγκο. Το βασικό σκεπτικό πίσω από τις τροποποιήσεις των αρχιτεκτονικών ήταν να επιτευχθεί η ανίχνευση σε πολλαπλούς χάρτες χαρακτηριστικών από διαφορετικά επίπεδα του δικτύου με την προϋπόθεση ότι αυτό θα επέτρεπε την ακριβή ανίχνευση φρούτου ακόμη και με μειωμένο αριθμό επιπέδων στο δίκτυο. Περνώντας στα αποτελέσματα το MangoYOLO(pt) κατέληξε με μέση ακρίβεια 0,983 στο σετ δεδομένων αξιολόγησης, που ήταν ανεξάρτητο από το σετ εκπαίδευσης, ξεπερνώντας τους άλλους αλγόριθμους, με ταχύτητα ανίχνευσης 8 ms ανά εικόνα 512 × 512 pixel. Το μοντέλο MangoYOLO ξεπέρασε επίσης τα άλλα μοντέλα στην επεξεργασία πλήρους εικόνων, απαιτώντας μόλις 70 ms ανά εικόνα (2048 × 2048 pixel). Το μοντέλο αποδείχτηκε αρκετά γενικευμένο στη χρήση εικόνων από άλλους οπωρώνες, άλλων ποικιλιών και συνθηκών φωτισμού. Το MangoYOLO (bu) έδωσε ακρίβεια F1=0,89 στις εικόνες με μάνγκο, που η λήψη τους είχε γίνει την ημέρα. Τέλος το MangoYOLO (pt) υπολόγισε την παραγωγή του οπωρώνα με ποσοστό μεταξύ 4,6 και 15,2% των καρπών που συσκευάστηκαν για τους πέντε εξεταζόμενους οπωρώνες, συγκρινόμενο με τη χρήση ενός συντελεστή διόρθωσης. Ο συντελεστής διόρθωσης υπολογίζεται από την αναλογία, του αριθμού των φρούτων που έχει εκτιμηθεί από άνθρωπο μέσω εικόνων που απεικονίζουν και τις δύο πλευρές των δέντρων του δείγματος ανά οπωρώνα και έναν αριθμό συγκομιδής υπολογισμένο χειροκίνητα σε όλα τα φρούτα αυτών των δέντρων.
Στη δημοσίευση των *Yu-Dong Zhang et al.(2019)*, η ομάδα τους σχεδίασε ένα συνελικτικό νευρωνικό δίκτυο (CNN) 13 επιπέδων, για τον προσδιορισμό της κατηγορίας φρούτων από εικόνες. Χρησιμοποιήθηκαν οι μέθοδοι επαύξησης δεδομένων, περιστροφή εικόνας, διόρθωση γάμμα και έγχυση θορύβου. Οι συγγραφείς σύγκριναν επίσης την απόδοση που προσφέρουν στο δίκτυο τα επίπεδα μέγιστης και μέσης συνάθροισης. Η συνολική ακρίβεια της μεθόδου τους είναι 94,94%. Τα κύρια σημεία αυτής της μελέτης είναι:

- Προτείνεται ένα συνελικτικό νευρωνικό δίκτυο 13 επιπέδων, στο οποίο αξιολόγησαν το βέλτιστο αριθμό επιπέδων συνέλιξης και επιπέδων συνάθροισης.
- Οι συγγραφείς κατέληξαν ότι το επίπεδο μέγιστης συνάθροισης παρέχει καλύτερη ελαφρά απόδοση από το επίπεδο μέσης συνάθροισης.
- Η μέθοδος τους απέδωσε συνολική ακρίβεια 94,94% και συγκρινόμενη με πέντε άλλες μεθοδολογίες αιχμής (π.χ. PCA, SVM κ.α.) έδωσε υψηλότερα αποτελέσματα ακρίβειας.
- Αξιολόγησαν τη μέθοδό τους σε εικόνες με διαφορετικά προβλήματα, όπως η λανθασμένη εστίαση, η υπερέκθεση στο φως και η απουσία των φρούτων από το πρώτο πλάνο ενώ η υπολογιζόμενη ακρίβεια ήταν 89.6%.
- Σύγκριναν υπολογιστικούς πόρους, δηλαδή υπολογισμούς με CPU και με GPU και διαπίστωσαν ότι η GPU μπορεί να επιτύχει επιτάχυνση ×177 στα δεδομένα εκπαίδευσης, και επιτάχυνση ×175 στα δεδομένα της δοκιμής σε σχέση με τη CPU.
- Χρησιμοποιήσαν πέντε διαφορετικούς τύπους μεθόδων επαύξησης δεδομένων και σύγκριναν την απόδοση ταξινόμησης χρησιμοποιώντας την επαύξηση δεδομένων και χωρίς αυτήν. Καταλήγοντας ότι η επαύξηση δεδομένων μπορεί να αυξήσει τη συνολική ακρίβεια της ταξινόμησης.

Η μέθοδος και το μοντέλο που αναπτύχθηκε, κρίθηκε αποτελεσματική στην ταξινόμηση φρούτων από εικόνες.

Το άρθρο των *Jordi Gene-Mola et al. (2020)*, παρουσιάζει μια νέα μεθοδολογία για την ανίχνευση φρούτων και τον προσδιορισμό της θέσης τους στο χώρο που αποτελείται από: (1) Δισδιάστατη ανίχνευση φρούτων και κατάτμηση με χρήση νευρωνικού δικτύου κατάτμησης Mask R-CNN. (2) Δημιουργία τρισδιάστατου νέφους σημείων των ανιχνευμένων μήλων χρησιμοποιώντας τη φωτογραμμετρική τεχνική δομής από κίνηση/ Structure from Motion (SfM). (3) Προβολή των δισδιάστατων ανιχνεύσεων από την εικόνα στον τρισδιάστατο χώρο (4) Αφαίρεση των ψευδών θετικών προβλέψεων με χρήση SVM. Αυτή η μεθοδολογία δοκιμάστηκε σε 11 Μηλιές Fuji που περιέχουν συνολικά 1455 μήλα. Τα αποτελέσματα έδειξαν ότι, συνδυάζοντας την κατάτμηση στιγμιότυπων με SfM η απόδοση

του συστήματος αυξήθηκε από F1 = 0,816 (2D ανίχνευση φρούτων) σε F1=0,881 (3D ανίχνευση φρούτων) σε σχέση με τη συνολική ποσότητα των φρούτων. Τα κύρια πλεονεκτήματα αυτής της μεθοδολογίας είναι ο μειωμένος αριθμός ψευδώς θετικών προβλέψεων και το υψηλότερο ποσοστό ανίχνευσης, ενώ το κύριο μειονέκτημα είναι ο μεγάλος χρόνος επεξεργασίας που απαιτείται για τον SfM, γεγονός που την καθιστά επί του παρόντος ακατάλληλη για εφαρμογές σε πραγματικό χρόνο. Σχετικά με τα δεδομένα που χρησιμοποιήθηκαν, να αναφερθεί ότι λόγω της μεγάλης ποσότητας μήλων ανά εικόνα και του γεγονότος ότι η απόδοση των συνελικτικών νευρωνικών δικτύων μειώνεται κατά την ανίχνευση μικρών αντικειμένων, πριν από την εφαρμογή του βήματος της κατάτμησης οι εικόνες χωρίστηκαν σε 24 υπό-εικόνες των 1024 × 1024 pixel. Μετά το συνελικτικό νευρωνικό δίκτυο Mask R-CNN (He et al., 2017) χρησιμοποιήθηκε για τον εντοπισμό και την κατάτμηση των μήλων. Το βαθύ νευρωνικό δίκτυο Mask R-CNN που χρησιμοποιήθηκε στη συγκεκριμένη εφαρμογή παρέχει πλαίσια οριοθέτησης και σημασιολογικές μάσκες για τα αντικείμενα της εικόνας εισόδου. Στην ουσία αποτελεί μια επέκταση του δικτύου Faster R-CNN (Ren et al., 2017) που προσθέτει ένα τμήμα για πρόβλεψη μάσκας κατάτμησης σε κάθε περιοχή ενδιαφέροντος (RoI). Στην αρχιτεκτονική του δικτύου μπορούν να διαφοροποιηθούν δύο μέρη: το δίκτυο κορμού (backbone), που χρησιμοποιείται για την εξαγωγή χαρακτηριστικών στη προκειμένη χρησιμοποιήθηκε ένα προεκπαιδευμένο μοντέλο ResNet-101 FPN, και η κεφαλή του δικτύου που αναγνωρίζει το πλαίσιο οριοθέτησης (ταξινόμηση και παλινδρόμηση)και προβλέπει τη μάσκα του αντικείμενου, που εφαρμόζεται χωριστά σε κάθε RoI.



Εικόνα 34: Αρχιτεκτονική Mask R-CNN

3. Μεθοδολογία

Η παρούσα εργασία έχει ως αντικείμενο την ταξινόμηση και την αναγνώριση αντικειμένου από ένα σετ δεδομένων έγχρωμων εικόνων που απεικονίζουν φρούτα, με μοντέλα CNN. Στο παρόν κεφάλαιο περιγράφεται πως οργανώθηκαν τα πειράματα τόσο στην ταξινόμηση όσο και στην αναγνώριση φρούτων από εικόνες. Στην αρχή του κεφαλαίου παρατίθενται πληροφορίες σχετικά με το σύνολο δεδομένων και τις τεχνολογίες που χρησιμοποιήθηκαν τόσο κατά την ταξινόμηση όσο και κατά την αναγνώριση αντικειμένου. Μετέπειτα το κεφάλαιο οργανώνεται σε 2 μέρη όπου στο καθένα αναλύεται η μεθοδολογία και τα πειράματα που πραγματοποιήθηκαν τόσο στο τμήμα της ταξινόμησης όσο και στο τμήμα της αναγνώρισης αντικειμένου. Για το $1^{ο}$ μέρος της ταξινόμησης στόχος ήταν να υπάρξει εξοικείωση με τα μοντέλα CNN, τη θεωρία γύρω από αυτά καθώς και με τις τεχνολογίες που μπορεί κανείς να τα διαχειριστεί. Επιλέχθηκαν 3 δημοφιλή προ-εκπαιδευμένα μοντέλα το MobileNet V2, το Resnet 50 και το VGG-16, όπου χρησιμοποιούνται στη διεθνή βιβλιογραφία για αντίστοιχα και πιο σύνθετα προβλήματα ταξινόμησης, κατάτμησης και αναγνώρισης εικόνας. Τα 3 επιλεγμένα μοντέλα κυκλοφορούν ελεύθερα για χρήση σε ερευνητικούς σκοπούς. Αποφασίστηκε να υπάρξει σύγκριση της ακρίβειας των 3 μοντέλων πάνω στο κοινό dataset εικόνων ακολουθώντας κοινές πρακτικές στην προεπεξεργασία των εικόνων καθώς και των τεχνικών και των παραμέτρων που εφαρμόστηκαν κατά τη διαδικασία εκπαίδευσης των $3^{ων}$ μοντέλων. Πέρα από την $1^{η}$ σύγκριση των $3^{ων}$ μοντέλων αποφασίστηκε να υπάρξουν άλλες 2 παραλλαγές συγκρίσεων όπου αλλάζει η προεπεξεργασία των εικόνων ώστε να μελετηθεί εάν και πόσο επηρεάζουν την ακρίβεια κάθε μοντέλο στην διαδικασία της ταξινόμησης. Για το $2^{ο}$ μέρος της αναγνώρισης αντικειμένου από εικόνα, επιλέχθηκαν 3 διαφορετικές αρχιτεκτονικές μοντέλων SSD , YOLOv3 και Faster R-CNN, οι δυο πρώτες ανήκουν στην κατηγορία των αρχιτεκτονικών ενός σταδίου ενώ η $3^{η}$ στην κατηγορία των 2 σταδίων. Στα πειράματα που εκτελέστηκαν καθένα από τα μοντέλα εκπαιδεύτηκε πάνω στο ίδιο σετ δεδομένων , το οποίο χρησιμοποιήθηκε και στην ταξινόμηση. Με σκοπό στην συνέχεια να αξιολογηθεί η ακρίβεια τους στην αναγνώριση αντικειμένου με πλαίσια οριοθέτησης σε εικόνες που δεν είχαν χρησιμοποιηθεί στην εκπαίδευση των μοντέλων.

3.1 Σετ Δεδομένων

Το σετ δεδομένων που χρησιμοποιήθηκε διατίθεται ελεύθερο στον ιστότοπο www.kaggle.com , ο οποίος αποτελεί ένα τεράστιο αποθετήριο δημοσιευμένων δεδομένων. Το σετ δεδομένων αποτελείται από έγχρωμες εικόνες RGB, οι οποίες περιέχουν 4 κατηγορίες με φρούτα για την ταξινόμηση [μήλο, μπανάνα, πορτοκάλι, μικτή κατηγορία φρούτων]. Το σετ δεδομένων αποτελείται συνολικά από 300 εικόνες, και ήταν δομημένο ώστε οι 240 εικόνες να χρησιμοποιούνται για την εκπαίδευση του κάθε μοντέλου και 60 εικόνες για τη δοκιμή του. Στη παρούσα εργασία σε όλες τις δοκιμές το τμήμα του dataset που χρησιμοποιήθηκε για την εκπαίδευση των μοντέλων (240 εικόνες) χωρίστηκε σε ποσοστό 80% /20%, για εκπαίδευση (192 εικόνες) και αξιολόγηση (48 εικόνες) του μοντέλου, πριν την τελική δοκιμή του κάθε μοντέλου στις 60 εικόνες του testing.

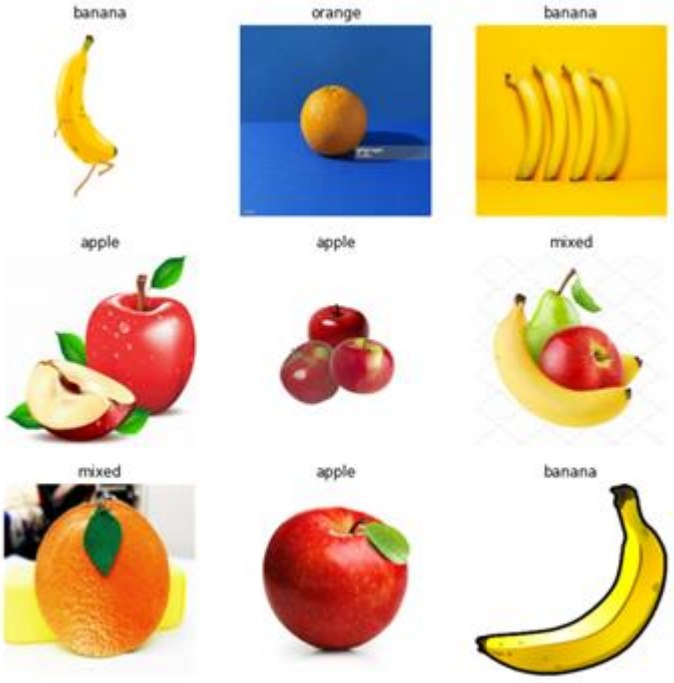
Άρα το σετ δεδομένων οργανώθηκε ως εξής:

Training dataset: Οι 192 εικόνες αποτελούν το σύνολο δεδομένων εκπαίδευσης δηλαδή το σύνολο δεδομένων το οποίο είναι ένα σύνολο εικόνων που χρησιμοποιούνται για την προσαρμογή των παραμέτρων (π.χ. βάρη) του μοντέλου. Το μοντέλο εκπαιδεύεται στο σύνολο των δεδομένων χρησιμοποιώντας μια επιβλεπόμενη μέθοδο. Στην πράξη, το σύνολο δεδομένων εκπαίδευσης αποτελείται συχνά από ζεύγη ενός διανύσματος εισόδου και του αντίστοιχου διανύσματος εξόδου, όπου η σύνδεση του υποδηλώνεται συνήθως από την ετικέτα που περιέχουν. Το

τρέχον μοντέλο προσαρμόζεται στο σύνολο δεδομένων εκπαίδευσης και παράγει ένα αποτέλεσμα, το οποίο στη συνέχεια συγκρίνεται με τη ετικέτα που περιείχε, για κάθε διάνυσμα εισόδου στο σύνολο δεδομένων εκπαίδευσης. Με βάση το αποτέλεσμα της σύγκρισης και τον συγκεκριμένο αλγόριθμο μάθησης που χρησιμοποιείται, οι παράμετροι του μοντέλου προσαρμόζονται.

Validation dataset: Οι 48 εικόνες θα αποτελέσουν το σύνολο δεδομένων αξιολόγησης, κατά τη διαδικασία της εκπαίδευσης. Ουσιαστικά πρόκειται για ένα τμήμα εικόνων που έχει οριστεί ώστε κατά την εκπαίδευση του μοντέλου να ελέγχεται πάνω σε αυτές πως προχωράει η εκπαίδευση και αν παρουσιάζονται προβλήματα όπως υπερπροσαρμογής των δεδομένων δηλαδή μεγάλη απόκλιση ανάμεσα στα ποσοστά ακρίβειας του συνόλου δεδομένων εκπαίδευσης και αξιολόγησης.

Testing dataset: Οι 60 εικόνες θα αποτελέσουν το σύνολο δεδομένων δοκιμής, δηλαδή είναι ένα σύνολο δεδομένων που χρησιμοποιείται για την παροχή αμερόληπτης αξιολόγησης ενός τελικού μοντέλου που έχει ολοκληρώσει τη διαδικασία εκπαίδευσης του. Το σύνολο των δεδομένων δοκιμής δεν έχουν χρησιμοποιηθεί ποτέ στην εκπαίδευση του μοντέλου.



Εικόνα 35: Εικόνες από το σετ δεδομένων που μελετάται με τις 4 κατηγορίες που εξετάζονται

3.2 Τεχνολογίες

Για την ανάπτυξη των πειραμάτων της ταξινόμησης χρησιμοποιήθηκαν η βιβλιοθήκη Tensorflow και το περιβάλλον του Google-Colaboratory, παρακάτω αναφέρονται λίγα λόγια για το καθένα.

Tensorflow: Είναι βιβλιοθήκη ανοιχτού κώδικα που βοηθά στην ανάπτυξη και την εκπαίδευση μοντέλων Μηχανικής Μάθησης. Χρησιμοποιεί ως γλώσσα προγραμματισμού την Python για να παρέχει στο χρήστη μια βολική διεπαφή για τη δημιουργία εφαρμογών, ενώ εκτελεί αυτές τις εφαρμογές σε C ++.

Google Colaboratory: Είναι προϊόν της Google. Το Colab επιτρέπει σε οποιονδήποτε να γράψει και να εκτελέσει κώδικα σε γλώσσα python μέσω του λογισμικού περιήγησης και είναι κατάλληλο για μηχανική μάθηση και ανάλυση δεδομένων. Τεχνικά, το Colab ενσωματώσει περιβάλλον Jupyter και δεν απαιτεί εγκατάσταση για να το χρησιμοποιήσει κανείς, ενώ παρέχει δωρεάν πρόσβαση σε υπολογιστικούς πόρους, συμπεριλαμβανομένων της κάρτας γραφικών.

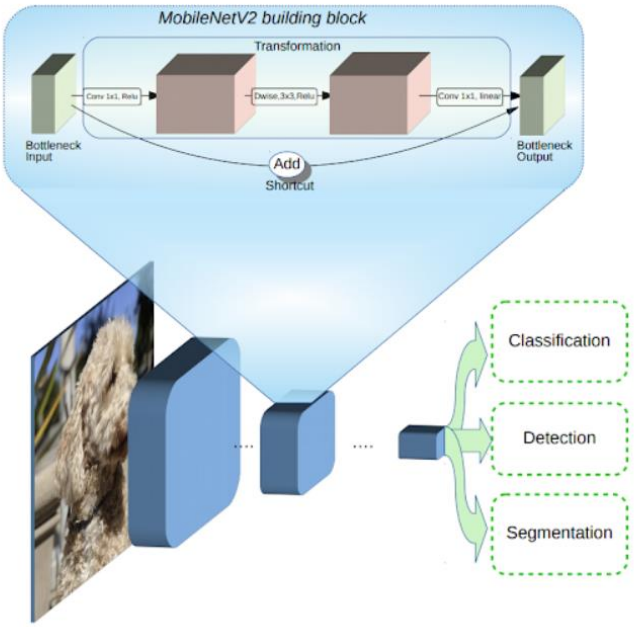
3.3 Μεθοδολογία Ταξινόμησης

Παρακάτω περιγράφονται τα μοντέλα που χρησιμοποιήθηκαν στο κομμάτι της ταξινόμησης καθώς και τα πειράματα που υλοποιήθηκαν στο τμήμα αυτό της εργασίας.

Επιλεγμένα CNN Μοντέλα

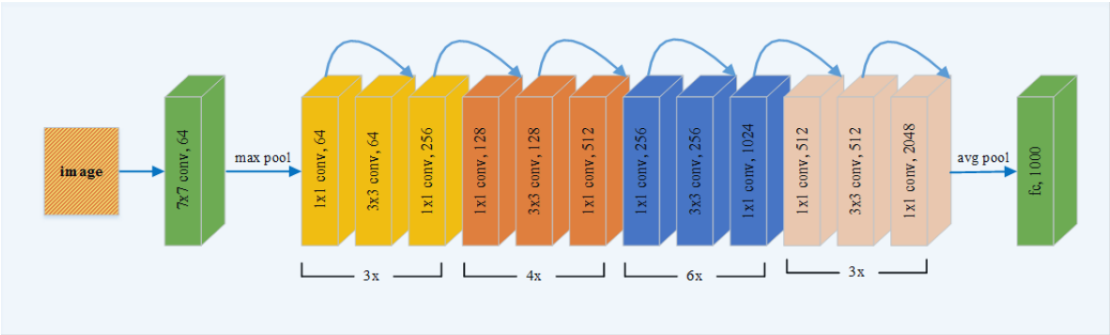
Τα 3 μοντέλα που επιλέχθηκαν για μελέτη είναι το MobileNetV2, το Resnet50 και το VGG-16. Η επιλογή τους έγινε με κριτήριο τη διεθνή βιβλιογραφία δηλαδή να έχουν χρησιμοποιηθεί σε αντίστοιχα προβλήματα καθώς και να παρουσιάζουν καλά ποσοστά ακρίβειας πέρα από το κομμάτι της ταξινόμησης τόσο στο κομμάτι της ανίχνευσης αντικειμένων όπου είναι ο στόχος να γίνει περαιτέρω έρευνα μελλοντικά. Ακολουθεί μια μικρή περιγραφή για καθένα από αυτά.

Το **MobileNet-v2** είναι ένα συνελικτικό νευρωνικό δίκτυο με 53 επίπεδα βάθους. Στη παρούσα εργασία χρησιμοποιείται προ-εκπαιδευμένο στο ImageNet. Το MobileNetV2 είναι μια βελτιωμένη εκδοχή σε σχέση με το MobileNetV1 και χρησιμοποιεί προηγμένη τεχνολογία για την οπτική αναγνώριση των κινητών τηλεφώνων, συμπεριλαμβανομένης της ταξινόμησης, της ανίχνευσης αντικειμένων και της σημασιολογικής κατάτμησης. Το MobileNetV2 κυκλοφορεί ως μέρος της βιβλιοθήκης TensorFlow-Slim Image. Το MobileNetV2 βασίζεται στο MobileNetV1, χρησιμοποιώντας διαχωρίσιμα συνελικτικά επίπεδα με βάθος ως δομικά στοιχεία. Ωστόσο, το V2 εισάγει δύο νέα χαρακτηριστικά στην αρχιτεκτονική: 1) bottleneck επίπεδα μεταξύ των συνελικτικών δομικών στοιχείων προστίθενται για να μειώσουν τον αριθμό των χαρτών με χαρακτηριστικά που εξάγονται 2) συνδέσεις συντόμευσης μεταξύ των bottleneck επιπέδων. Η βασική δομή φαίνεται παρακάτω στην παρακάτω εικόνα.



Εικόνα 39: Η αρχιτεκτονική του MobileNet V2 / Τα μπλε τμήματα του μοντέλου αντιπροσωπεύουν σύνθετα συνελικτικά δομικά στοιχεία όπως αυτό που αναλύεται στην εικόνα

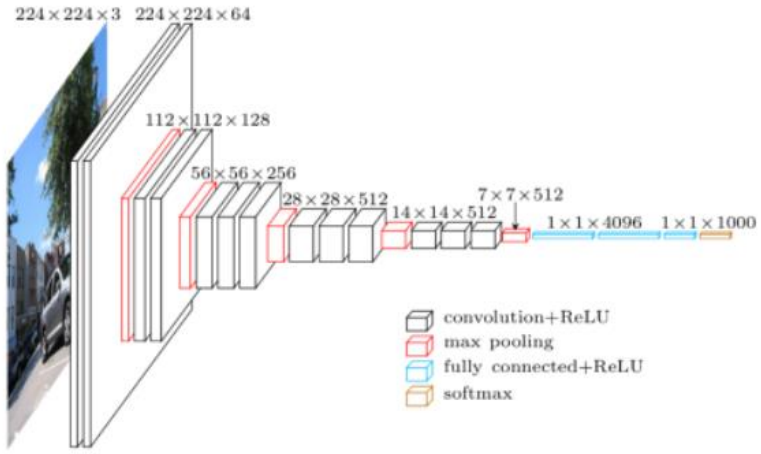
Το **ResNet50** είναι ένα συνελικτικό νευρωνικό δίκτυο με 50 επίπεδα βάθους. Στη παρούσα εργασία χρησιμοποιείται προ-εκπαιδευμένο στο ImageNet. Το ResNet50 είναι μια παραλλαγή του μοντέλου ResNet. Η αρχιτεκτονική ResNet εισήγαγε συνδέσεις παράκαμψης, γνωστές και ως υπολειπόμενες συνδέσεις για την αποφυγή απώλειας πληροφοριών κατά την εκπαίδευση σε βάθος δικτύου. Η τεχνική παράλειψης σύνδεσης επιτρέπει την εκπαίδευση πολύ βαθιών δικτύων και μπορεί να αυξήσει την απόδοση του μοντέλου. Το ResNet50 έχει 48 επίπεδα συνελίξεων μαζί με 1 επίπεδο MaxPool (συνάθροισης μέγιστης τιμής) και 1 επίπεδο Average Pool (συνάθροισης μέσης τιμής).



Εικόνα 40: Βασική αρχιτεκτονική του ResNet50

Το **VGG16** είναι ένα συνελικτικό νευρωνικό δίκτυο με 16 επίπεδα βάθους. Στη παρούσα εργασία χρησιμοποιείται προ-εκπαιδευμένο στο ImageNet. Το VGG16 είναι μια αρχιτεκτονική νευρωνικού δικτύου (CNN), η οποία χρησιμοποιήθηκε για να κερδίσει τον διαγωνισμό ILSVR (Imagenet) το 2014. Θεωρείται ως μία από τις εξαιρετικές αρχιτεκτονικές μοντέλων όρασης υπολογιστών μέχρι σήμερα. Το πιο μοναδικό με το VGG16 είναι ότι αντί να έχει μεγάλο αριθμό παραμέτρων, η ανάπτυξη του έχει επικεντρωθεί στο να έχει επίπεδα συνέλιξης 3x3 και

συναθροιστικά επίπεδα μεγίστου 2x2. Ακολουθεί αυτή τη διάταξη σε επίπεδα συνέλιξης και συναθροιστικά επίπεδα μεγίστου με συνέπεια σε όλη την αρχιτεκτονική του. Στο τέλος έχει 2 πλήρως συνδεδεμένα επίπεδα ακολουθούμενα από τη συνάρτηση softmax για έξοδο, η οποία μετατρέπει ένα διάνυσμα x πραγματικών τιμών σε ένα διάνυσμα x πραγματικών τιμών με άθροισμα 1. Το 16 στο VGG16 αναφέρεται ότι έχει 16 επίπεδα που έχουν βάρη. Αυτό το δίκτυο είναι ένα αρκετά μεγάλο αποτελούμενο περίπου από 138 εκατομμύρια παραμέτρους.



Εικόνα 41: Η αρχιτεκτονική του VGG16

Περιγραφή Πειραμάτων

Στη παρούσα εργασία μελετάται το πρόβλημα της ταξινόμησης εικόνων που περιέχουν διαφορετικά είδη φρούτων χρησιμοποιώντας μοντέλα Βαθιάς Μηχανικής Μάθησης. Στόχος είναι να αποκτηθεί καλύτερη κατανόηση των δυνατοτήτων και των περιορισμών των τεχνικών βαθιάς μηχανικής μάθησης. Για αυτό πραγματοποιήθηκαν τα ακόλουθα πειράματα τα οποία χωρίζονται σε 3 μέρη.

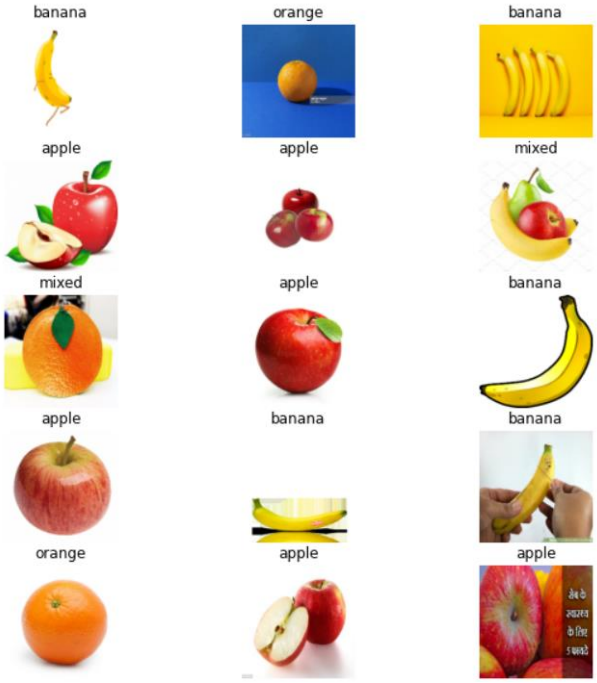
• Επιλέχθηκαν 3 μοντέλα (MobileNet, Resnet50, VGG16) που είναι προεκπαιδευμένα στο ImageNet, με σκοπό την ταξινόμηση του επιλεγμένου σετ δεδομένων με εικόνες φρούτων. Δημιουργώντας μια ακολουθία βημάτων για την προεπεξεργασία των δεδομένων, την εκπαίδευση του μοντέλου και την τελική δοκιμή του, η οποία περιλαμβάνει τη χρήση των τεχνικών του transfer-learning και fine-tuning καθώς και τη μελέτη των κλασσικών μεθόδων προεπεξεργασίας των δεδομένων, που είναι τεχνικές επαύξησης του συνόλου των δεδομένων (αναστροφή εικόνας, περιστροφή κ.λπ.), για το πως επηρεάζουν την απόδοση ενός μοντέλου. Τέλος συγκρίθηκαν οι αποδόσεις του κάθε μοντέλου.

• Αλλαγή της ανάλυσης των εικόνων στα δεδομένα εκπαίδευσης, μελετάται η απόδοση του μοντέλου σε δεδομένα χαμηλότερης ανάλυσης.

• Αλλαγή της αντίθεσης και μορφολογικών χαρακτηριστικών στις εικόνες εκπαίδευσης, μελετάται η απόδοση των μοντέλων και αν επηρεάζονται από αυτές τις αλλαγές.

Παρά τις παραλλαγές των πειραμάτων που παρουσιάζονται παραπάνω ακολουθούνται κάποια κοινά βήματα για τη δημιουργία των μοντέλων και την εκπαίδευση τους καθώς είναι κοινός ο ορισμός βασικών παραμέτρων, ώστε να μπορούν να συγκριθούν τα μοντέλα στη συνέχεια. Αυτή η κοινή διαδικασία παρουσιάζεται παρακάτω:

- Εισαγωγή απαραίτητων βιβλιοθηκών για την επεξεργασία των εικόνων και για την εκτέλεση των μοντέλων και των τεχνικών της βαθιάς μηχανικής μάθησης.
- Εξέταση και κατανόηση του σετ εικόνων που θα χρησιμοποιηθεί.
- Δήλωση των δεδομένων και χωρισμός τους σε δεδομένα εκπαίδευσης , αξιολόγησης και δοκιμής και των 4 κατηγοριών που περιλαμβάνουν. Στο σημείο αυτό δηλώνονται το μέγεθος των εικόνων καθώς και το μέγεθος του batch η τιμή που πήρε στη παρούσα εργασία είναι 32 που σημαίνει ότι 32 εικόνες θα εκπαιδευτούν πριν ο αλγόριθμος βελτιστοποίησης του μοντέλου ανανεώσει τις εσωτερικές του παραμέτρους με σκοπό να βελτιώσει τις παραμέτρους του μοντέλου ώστε κάθε πρόβλεψη του να πλησιάζει τα πραγματικά δεδομένα. Επίσης δηλώνεται ότι ετικέτες για τις εικόνες δημιουργούνται από τη δομή του φακέλου που βρίσκονται οι εικόνες. Δηλαδή το σετ δεδομένων για εκπαίδευση είναι χωρισμένο σε 4 υποφακέλους που καθένας περιέχει τις εικόνες της κάθε κατηγορίας και πήρε και το όνομα της.



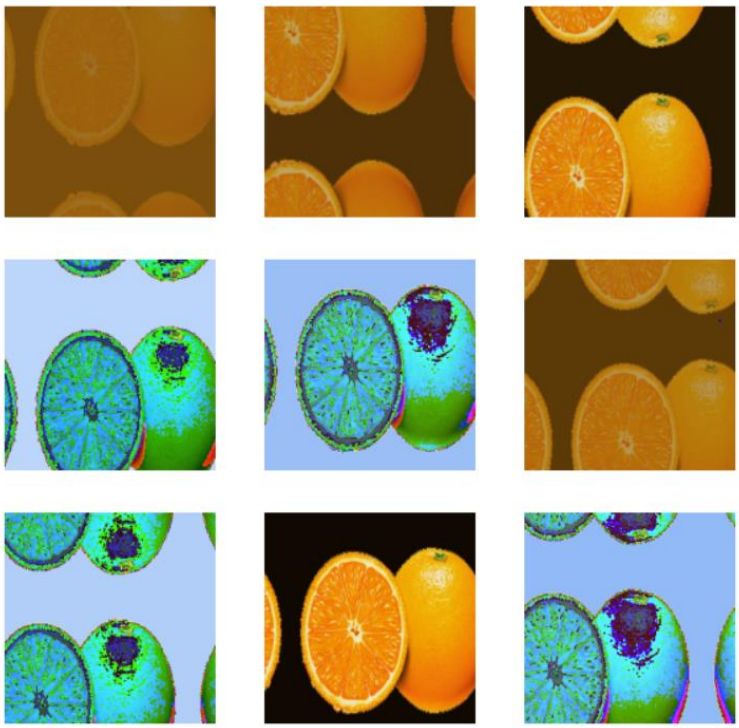
Εικόνα 42: Οπτικοποίηση από τα δεδομένα εκπαίδευσης

- Στη συνέχεια διαμορφώνονται τόσο τα δεδομένα εκπαίδευσης και αξιολόγησης όσο και δοκιμής ώστε όταν επεξεργάζονται από την κάρτα γραφικών του υπολογιστή να είναι αποδοτικά και να μην δημιουργούνται καθυστερήσεις ανάμεσα στην επεξεργασία κάθε batch.
- Επόμενο βήμα αποτελεί η δοκιμή τεχνικών επαύξησης δεδομένων (data augmentation), είναι συνηθής τεχνική που εφαρμόζεται στα μοντέλα νευρωνικών δικτύων, ώστε να αποκτήσουν πρόσθετα συνθετικά τροποποιημένα δεδομένα, ειδικά όταν τα σετ εικόνων είναι μικρά όπως στην παρούσα εργασία. Στην ουσία πρόκειται για αλγορίθμους που μεταβάλουν γεωμετρικά και μορφολογικά χαρακτηριστικά των εικόνων όπως κλίμακα , φωτεινότητα και προσανατολισμό. Στην $1^{η}$ και $3^{η}$ εκδοχή των πειραμάτων αποφασίστηκε να

εφαρμοστούν οι εξής μετασχηματισμοί για την επαύξηση δεδομένων 1) περιστροφή της εικόνας οριζόντια και κάθετα, 2)περιστροφή του αντικειμένου και 3) τυχαίο zoom στην εικόνα.

Εικόνα 43: Επαύξηση Δεδομένων

Στην 2η εκδοχή των πειραμάτων και συγκρίσεων μεταξύ των cnn μοντέλων αποφασίστηκε να αλλάξουν οι μετασχηματισμοί επαύξησης δεδομένων και να χρησιμοποιηθούν 1) διαφοροποίηση στην αντίθεση/ contrast (μείωση) των εικόνων 2) χωρικός και χρωματικός μετασχηματισμός. Με σκοπό να εξεταστεί αν επηρεάζεται η ικανότητα εκπαίδευσης των μοντέλων.



Εικόνα 44: Επαύξηση Δεδομένων στην 2η εκδοχή των πειραμάτων

- Εγκατάσταση του προεκπαιδευμένου μοντέλου και των έτοιμων βαρών που περιέχει.
- Ξεκινώντας το transfer learning, έγινε εξαγωγή του τμήματος των προεκπαιδευμένων επιπέδων που θα λειτουργήσει ως εξαγωγέας χαρακτηριστικών, δηλαδή το μοντέλο που θα αποτελέσει τη βάση. Το μοντέλο της βάσης εξάγει χαρακτηριστικά που είναι γενικώς χρήσιμα για την ταξινόμηση εικόνων καθώς αποτελεί τμήμα του προεκπαιδευμένου μοντέλου. Σε αυτό σημείο ορίζεται ότι τα βάρη του κυρίως μοντέλου θα είναι διατηρηθούν σταθερά κατά την εκπαίδευση του μοντέλου σε αυτή τη φάση του πειράματος.

```
Total params: 2,257,984
Trainable params: 0
Non-trainable params: 2,257,984
```

Εικόνα 45: Οι παράμετροι του βασικού μοντέλου του MobileNet, όπως φαίνεται στην εικόνα το σύνολο τους δεν θα εκπαιδευτεί

- Για να ολοκληρωθεί το μοντέλο, τοποθετούνται στο μοντέλο επίπεδα που θα λειτουργήσουν ώς ταξινομητές και τους δίνεται ο αριθμός των κατηγοριών που εξετάζεται. Αυτά είναι και τα επίπεδα που θα προσαρμοστούν στα δεδομένα μας σε αυτή τη φάση της εκπαίδευσης. Απλώς προστέθηκε ένας νέος ταξινομητής, ο οποίος θα εκπαιδευτεί από την αρχή, πάνω από το προ-εκπαιδευμένο μοντέλο, ώστε να μπορεί να επαναπροσδιορίσει τους χάρτες χαρακτηριστικών που έχουν εκπαιδευτεί προηγουμένως για άλλο σύνολο δεδομένων.

```
Total params: 2,263,108
Trainable params: 5,124
Non-trainable params: 2,257,984
```

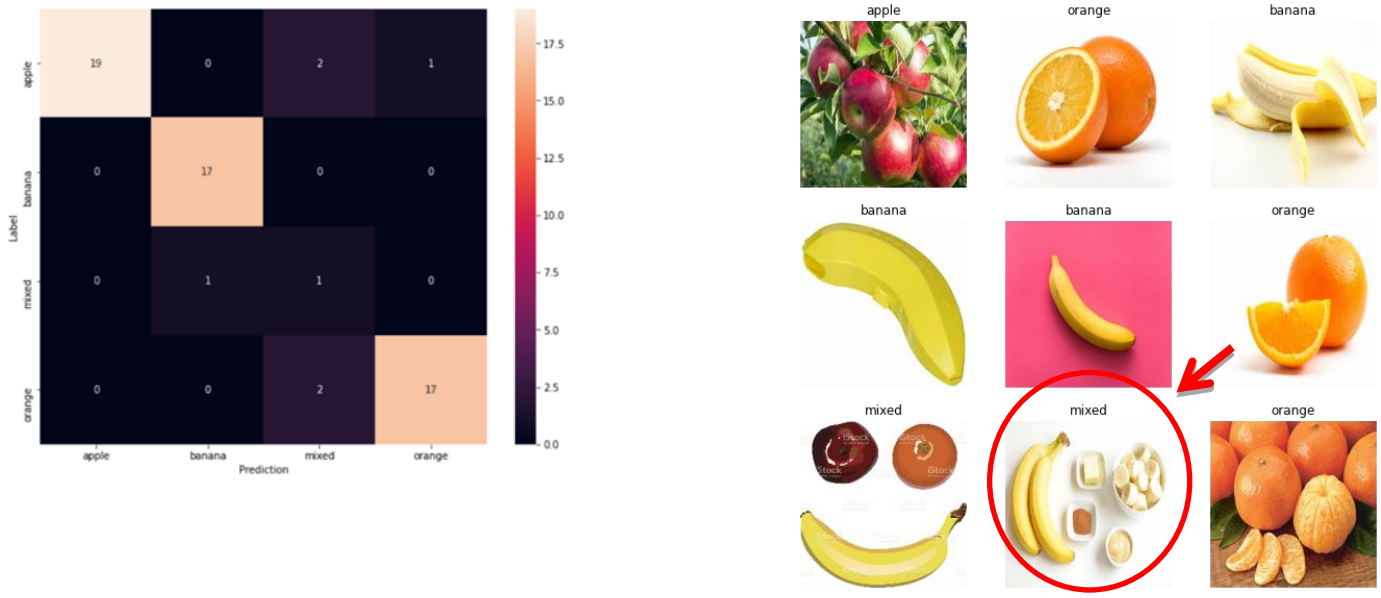
Εικόνα 46: Οι παράμετροι του μοντέλου του που έχει προκύψει από την συνένωση του feature extractor του MobileNet και του ταξινομητή που προστέθηκε, όπως φαίνεται στην εικόνα θα εκπαιδευτούν μόνο οι παράμετροι των επιπέδων του ταξινομητή

- Όλα τα παραπάνω μέρη συνθέτουν το μοντέλο, δηλαδή το μέγεθος του διανύσματος εισόδου (εικόνα), η συνάρτηση επαύξησης δεδομένων, το βασικό μοντέλο, τα επίπεδα που προστέθηκαν ως ταξινομητής, δηλώνονται σε αυτό το βήμα ως ενιαίο μοντέλο.
- Η εκπαίδευση ξεκινάει δηλώνοντας το ρυθμό εκπαίδευσης και τις εποχές. Εδώ δηλώθηκαν 0.001 και 10 epochs αντίστοιχα για αυτό το κομμάτι εκπαίδευσης που έχει εκτελεστεί μόνο transfer learning. Να αναφερθεί ότι η παράμετρος epoch καθορίζει τον αριθμό των φορών που ο αλγόριθμος εκπαίδευσης θα λειτουργήσει σε ολόκληρο το σύνολο δεδομένων εκπαίδευσης. Ενώ ο ρυθμός εκπαίδευσης καθορίζει το ποσό που θα ενημερώνονται τα βάρη κατά τη διάρκεια της εκπαίδευσης.
- Στη συνέχεια υπολογίζονται οι δείκτες απόδοσης του μοντέλου για τις πρώτες 10 εποχές εκπαίδευσης, όπου έχει εφαρμοστεί μόνο το transfer learning.
- Ακολουθεί η διαδικασία του fine tuning δηλαδή μερικά από τα τελευταία επίπεδα του βασικού μοντέλου θα εκπαιδευτούν από κοινού με τα επίπεδα της ταξινόμησης. Αυτό επιτρέπει να προσαρμοστούν, πέρα από τα βάρη των επιπέδων της ταξινόμησης, και τα βάρη από τα επίπεδα της εξαγωγής χαρακτηριστικών. Τα επίπεδα αυτά επειδή είναι από τα τελευταία του μοντέλου εξάγουν αναπαραστάσεις χαρακτηριστικών υψηλότερης τάξης στο βασικό μοντέλο, και με αυτό το τρόπο γίνονται πιο συναφείς για τα δεδομένα που εξετάζονται. Δηλώνεται από ποιο επίπεδο του βασικού μοντέλου και μετά αυτά θα συμμετέχουν στην εκπαίδευση του συνολικού μοντέλου. Επομένως δηλώνονται για πόσα epochs θα γίνει η εκπαίδευση του μοντέλου συμμετέχοντας πλέον περισσότερα επίπεδα αυτά της εξαγωγής χαρακτηριστικών.

```
Total params: 2,263,108
Trainable params: 1,866,564
Non-trainable params: 396,544
```

Εικόνα 47: Οι παράμετροι του μοντέλου του που έχει προκύψει από την συνένωση του feature extractor του MobileNet και του ταξινομητή που προστέθηκε, φαίνονται οι παράμετροι που θα εκπαιδευτούν εφαρμόζοντας fine tuning

- Τυπώνονται οι δείκτες απόδοσης του μοντέλου για τις επόμενες 10 εποχές εκπαίδευσης, όπου έχει εφαρμοστεί transfer learning σε συνδυασμό με το fine tuning.
- Στη συνέχεια γίνεται αξιολόγηση του μοντέλου στο σετ εικόνων δοκιμής. Δίνοντας το μοντέλο μια συνολική ακρίβεια.
- Θέλοντας να εξεταστούν περαιτέρω οι προβλέψεις του μοντέλου, μετατράπηκαν σε πιθανότητες και τυπώθηκαν.
- Τέλος αναπτυχθηκε συνάρτηση ώστε να τυπώνεται ο πίνακας σύγχυσης για κάθε μοντέλο, ώστε να είναι εύληπτο ποιες κατηγορίες ταξινομήθηκαν καλά από το μοντέλο και σε ποιες παρουσιάζει αδυναμία.



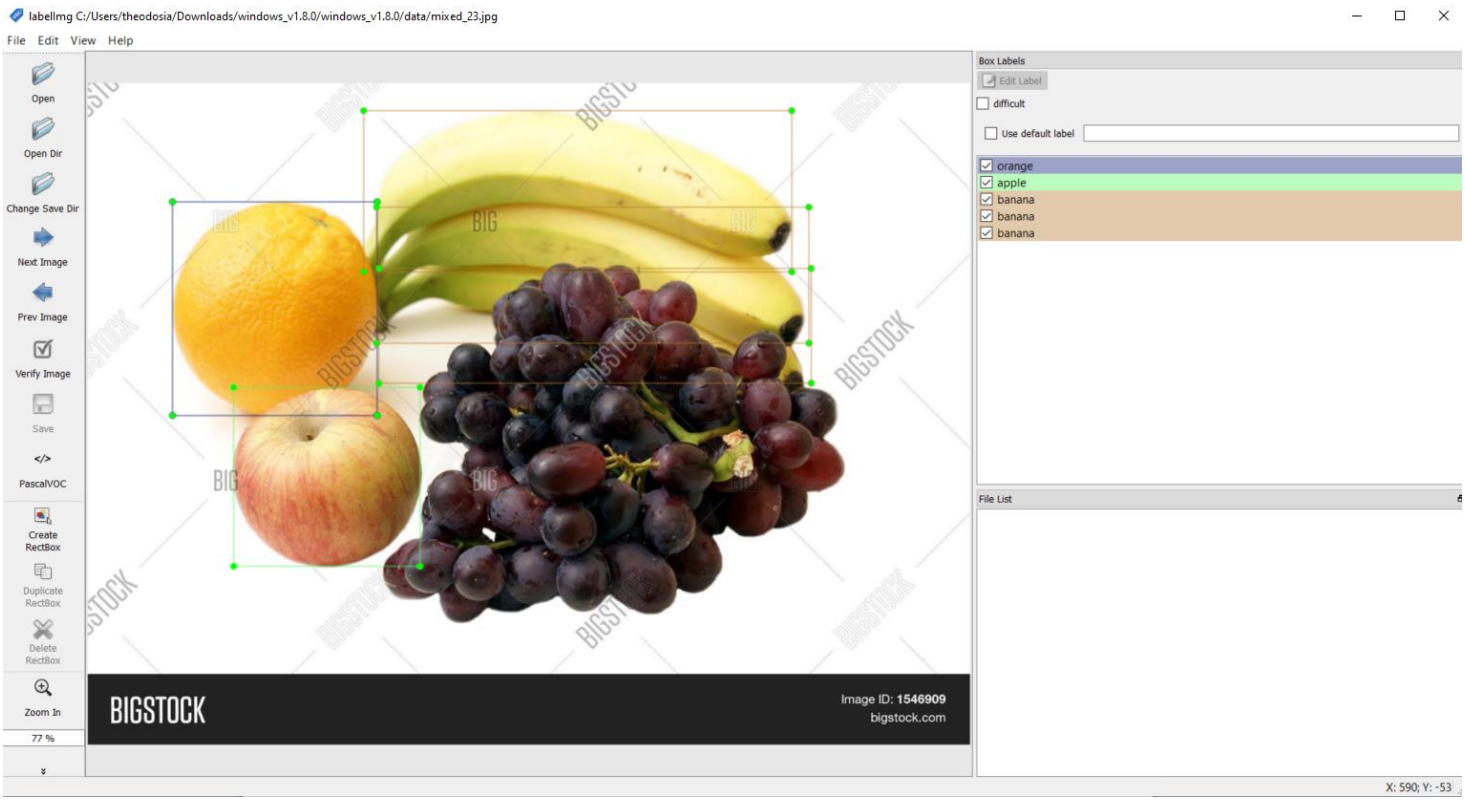
Εικόνα 48: Παράδειγμα του πίνακα σύγχυσης που δημιουργήθηκε (δεξιά) καθώς και προβλέψεων του MobileNet (αριστερά)

3.4 Μεθοδολογία Αναγνώρισης Αντικειμένου

Στο μέρος αυτό της εργασίας έγινε προσπάθεια να φτιαχτεί μια ροή εργασιών για τρεία μοντέλα CNN (SSD, Faster R-CNN, YOLOv3) με στόχο την αναγνώριση αντικειμένων με πλαίσια οριοθέτησης από εικόνα. Όπως και παραπάνω το σύνολο των δεδομένων που χρησιμοποιήθηκε είναι το προαναφερόμενο σετ με εικόνες φρούτων. Στην προκειμένη περίπτωση, στόχος είναι σε κάθε εικόνα να αναγνωρισθούν με επιτυχία τα φρούτα μήλο, πορτοκάλι, μπανάνα με τα αντίστοιχα πλαίσια οριοθέτησης.

Προετοιμασία των δεδομένων

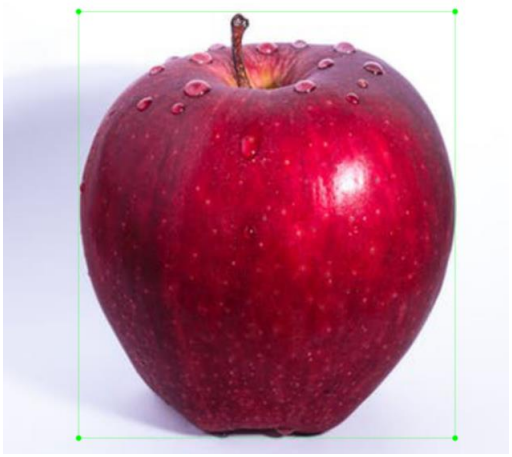
Για τη διαδικασία εκπαίδευσης ενός μοντέλου για την αναγνώριση αντικειμένου πρέπει κάθε εικόνα του σετ δεδομένων εκπαίδευσης να συνοδεύεται από τα πλαίσια οριοθέτησης με την αντίστοιχη ετικέτα του αντικειμένου που περιέχει. Για το λόγο αυτό χρησιμοποιήθηκε το πρόγραμμα labelImg, με το οποίο κάθε εικόνα του dataset επισημάνθηκε με bounding box και δόθηκε και στο καθένα η ετικέτα της κατηγορίας του εικονιζόμενου φρούτου.



Εικόνα 49: Το περιβάλλον του λογισμικού labelImg με το οποίο δόθηκαν ετικέτες στις εικόνες του dataset.

Στην ουσία μέσω του labelImg δημιουργήθηκαν αρχεία για κάθε εικόνα του dataset που περιγράφουν τα εικονιζόμενα αντικείμενα της εικόνας. Κάθε αρχείο επισήμανσης περιέχει τις συντεταγμένες του κάθε bounding box καθώς και την κλάση του αντικειμένου (ετικέτα). Τα αρχεία αυτά συνοδεύουν τις εικόνες που περιγράφουν και τροφοδοτούν το δίκτυο κατά τη διαδικασία της εκπαίδευσης.

Ανάλογα με το δίκτυο που θα χρησιμοποιηθούν τα αρχεία επισήμανσης μπορεί να αλλάζουν μορφότυπο αλλά και δομή. Για τα δίκτυα SSD και Faster R-CNN, τα αρχεία επισήμανσης είναι μορφότυπου .xml και έχουν την δομή που περιγράφεται στη παρακάτω εικόνα.



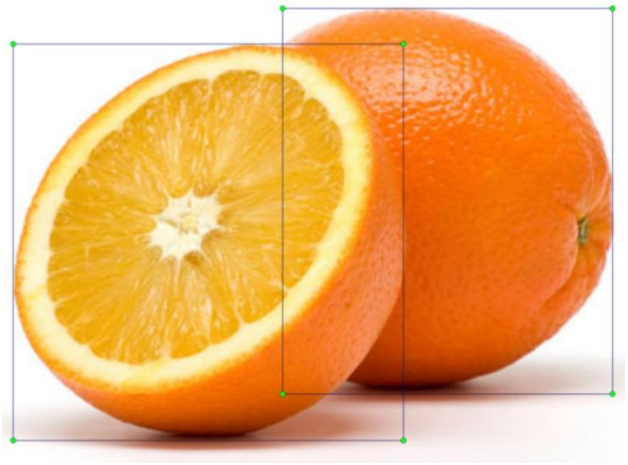
apple_93 - Σημειωματάριο

Αρχείο Επεξεργασία Μορφή Προβολή Βοήθεια

```
<annotation>
        <folder>data</folder>
        <filename>apple_93.jpg</filename>
        <path>C:\Users\theodosia\Downloads\windows_v1.8.0\windows_v1.8.0\data\apple_93.jpg</path>
        <source>
                <database>Unknown</database>
        </source>
        <size>
                <width>528</width>
                <height>350</height>
                <depth>3</depth>
        </size>
        <segmented>0</segmented>
        <object>
                <name>apple</name>
                <pose>Unspecified</pose>
                <truncated>0</truncated>
                <difficult>0</difficult>
                <bndbox>
                        <xmin>145</xmin>
                        <ymin>22</ymin>
                        <xmax>413</xmax>
                        <ymax>326</ymax>
                </bndbox>
        </object>
</annotation>
```

Εικόνα 50: Αρχείο xml που παράχθηκε μέσω του labelImg και περιγράφει το bounding box και την κλάση του αντικειμένου για τα δίκτυα SSD και Faster R-CNN

Ενώ για το δίκτυο YOLOv3 τα αρχεία επισήμανσης είναι μορφότυπου .txt και έχουν την δομή που εμφανίζεται στη παρακάτω εικόνα.



orange_84 - Σημειωματάριο

Αρχείο Επεξεργασία Μορφή Προβολή Βοήθεια

```
2 0.343478 0.506944 0.547826 0.702778
2 0.679348 0.431944 0.454348 0.680556
```

Εικόνα 51: Αρχείο txt που παράχθηκε μέσω του labelImg και περιγράφει το bounding box και την κλάση του αντικειμένου για το δίκτυο YOLOv3

Διαδικασία εκπαίδευσης και Αξιολόγησης των Μοντέλων

Η διαδικασία εκπαίδευσης για τις αρχιτεκτονικές SSD και Faster R-CNN ήταν παρόμοια καθώς έγινε μέσω της βιβλιοθήκης Tensorflow Object Detection API και επιλέχθηκε κοινό δίκτυο κορμού το ResNet-50. Αντίθετα το YOLO v3 είχε κάποιες διαφοροποιήσεις γιατί υλοποιήθηκε μέσω του Darknet-19. Όλες οι διαδικασίες αναπτύχθηκαν στο google Colab. Τα βήματα και οι αποφάσεις κατά τη διαδικασία εκπαίδευσης των μοντέλων αναλύονται παρακάτω.

Εκπαίδευση Μοντέλων SSD και Faster R-CNN με το Tensorflow Object Detection API 2

Το TensorFlow Object Detection API είναι ένα πλαίσιο ανοιχτού κώδικα χτισμένο πάνω στο TensorFlow που διευκολύνει την κατασκευή, την εκπαίδευση και την ανάπτυξη μοντέλων ανίχνευσης αντικειμένων.

- Υπάρχουν ήδη προεκπαιδευμένα μοντέλα στο Tensorflow τα οποία αναφέρονται ως Model Zoo. Τα μοντέλα, που παρέχονται, έχουν διαφορετική αρχιτεκτονική και επομένως παρέχουν διαφορετικές ακρίβειες, αλλά υπάρχει μια αντιστάθμιση μεταξύ της ταχύτητας εκτέλεσης και της ακρίβειας στην τοποθέτηση πλαισίων οριοθέτησης.

- Περιλαμβάνει μια συλλογή προεκπαιδευμένων μοντέλων που έχουν εκπαιδευτεί σε διάφορα σύνολα δεδομένων όπως το σύνολο δεδομένων COCO, KITTI και το Open Images Dataset.

Το TensorFlow Object Detection API 2 που χρησιμοποιήθηκε στην παρούσα εργασία, υποστηρίζει το TensorFlow 2, επιτρέπει τη χρήση αρχιτεκτονικών από μοντέλα αιχμής για την ανίχνευση αντικειμένων και επιτρέπει την διαμόρφωση των διαθέσιμων μοντέλων με απλό τρόπο.

Βήματα Εκπαίδευσης

Εγκατάσταση και ρύθμιση βιβλιοθηκών και αποθετηρίων που είναι απαραίτητες για τη διαδικασία ανίχνευσης αντικειμένων.

- Εγκατάσταση της βασικής βιβλιοθήκης Tensorflow.
- Εγκατάσταση του TensorFlow Model Garden.
- Εγκατάσταση του Protobuf. Το TensorFlow Object Detection API χρησιμοποιεί τη βιβλιοθήκη Protobuf για να διαμορφώσει τις παραμέτρους του μοντέλου και της εκπαίδευσης.
- Εγκατάσταση του COCO API
- Εγκατάσταση του Object Detection API

- Προετοιμασία εικόνων που θα χρησιμοποιηθούν στην εκπαίδευση και την αξιολόγηση του μοντέλου με αντίστοιχα πλαίσια οριοθέτησης και ετικέτες. Διαχωρισμός των εικόνων που θα χρησιμοποιηθούν για κάθε εργασία.

- Συγκεκριμένη δόμηση φακέλων τόσο για τα δεδομένα όσο και για τα βοηθητικά αρχεία που χρησιμοποιούνται για την εκπαίδευση ώστε να μπορεί να τα προσπελάσει το μοντέλο.

Δημιουργία χάρτη ετικετών

- Το TensorFlow απαιτεί έναν χάρτη ετικετών ο οποίος αντιστοιχίζει κάθε μία από τις χρησιμοποιούμενες ετικέτες σε ακέραιες τιμές. Αυτός ο χάρτης ετικετών χρησιμοποιείται τόσο κατά την διαδικασία εκπαίδευσης όσο και κατά τον εντοπισμό αντικειμένων.

label_map.pbtxt

```
item {
    id: 1
    name: 'apple'
}

item {
    id: 2
    name: 'banana'
}

item {
    id: 3
    name: 'orange'
}
```

Εικόνα 52: Χάρτης ετικετών για τις 3 κατηγορίες φρούτων που είναι τα αντικείμενα ανίχνευσης

Έχοντας δημιουργηθεί οι επισημάνσεις για τα δεδομένα και έχοντας χωρίσει το σύνολο των δεδομένων στα επιθυμητά υποσύνολα εκπαίδευσης και αξιολόγησης, το επόμενο βήμα είναι ο μετασχηματισμός των αρχείων επισήμανσης στη λεγόμενη μορφή TFRecord.

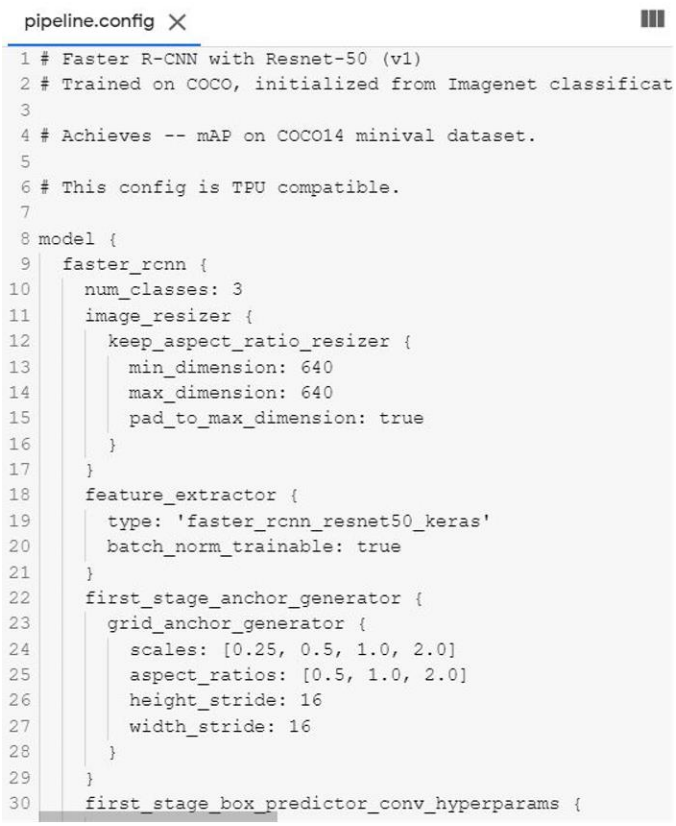
Επιλογή και διαμόρφωση μοντέλου

- Αρχικά, επιλέγεται μια αρχιτεκτονική CNN για αξιοποιηθεί στη συγκεκριμένη εργασία. Το μοντέλα που θα χρησιμοποιηθούν όπως έχει αναφερθεί είναι το SSD ResNet50 V1 FPN 640x640 και το Faster R-CNN ResNet50 V1 FPN 640x640 Ωστόσο, υπάρχει μια σειρά άλλων μοντέλων που μπορούν να χρησιμοποιηθούν, τα οποία παρατίθενται στο TensorFlow 2 Detection Model Zoo.

Model name	Speed (ms)	COCO mAP	Outputs
CenterNet HourGlass104 512x512	70	41.9	Boxes
CenterNet HourGlass104 Keypoints 512x512	76	40.0/61.4	Boxes/Keypoints
CenterNet HourGlass104 1024x1024	197	44.5	Boxes
CenterNet HourGlass104 Keypoints 1024x1024	211	42.8/64.5	Boxes/Keypoints
CenterNet Resnet50 V1 FPN 512x512	27	31.2	Boxes
CenterNet Resnet50 V1 FPN Keypoints 512x512	30	29.3/50.7	Boxes/Keypoints
CenterNet Resnet101 V1 FPN 512x512	34	34.2	Boxes
CenterNet Resnet50 V2 512x512	27	29.5	Boxes
CenterNet Resnet50 V2 Keypoints 512x512	30	27.6/48.2	Boxes/Keypoints
CenterNet MobileNetV2 FPN 512x512	6	23.4	Boxes
CenterNet MobileNetV2 FPN Keypoints 512x512	6	41.7	Keypoints
EfficientDet D0 512x512	39	33.6	Boxes
EfficientDet D1 640x640	54	38.4	Boxes
EfficientDet D2 768x768	67	41.8	Boxes
EfficientDet D3 896x896	95	45.4	Boxes
EfficientDet D4 1024x1024	133	48.5	Boxes
EfficientDet D5 1280x1280	222	49.7	Boxes
EfficientDet D6 1280x1280	268	50.5	Boxes

Εικόνα 53: Ορισμένα από τα διαθέσιμα μοντέλα στο TF2 Model Zoo

- Στη συνέχεια πραγματοποιείται η διαμόρφωση του μοντέλου που το καθιστά αποτελεσματικό και του επιτρέπει να γενικεύει αρκετά καλά ώστε να μπορεί να χρησιμοποιείται και σε άλλα δεδομένα. Για τη διαμόρφωση του μοντέλου στην ουσία τροποποιείται ένα αρχείο pipeline.config που μέσα δίδεται σε κώδικα η αρχιτεκτονική του μοντέλου (μέγεθος επιπέδων, είδη συναρτήσεων κλπ.) καθώς και οι παράμετροι αυτού.



pipeline.config

```
# Faster R-CNN with Resnet-50 (v1)
# Trained on COCO, initialized from Imagenet classificat

# Achieves -- mAP on COCO14 minival dataset.

# This config is TPU compatible.

model {
  faster_rcnn {
    num_classes: 3
    image_resizer {
      keep_aspect_ratio_resizer {
        min_dimension: 640
        max_dimension: 640
        pad_to_max_dimension: true
      }
    }
    feature_extractor {
      type: 'faster_rcnn_resnet50_keras'
      batch_norm_trainable: true
    }
    first_stage_anchor_generator {
      grid_anchor_generator {
        scales: [0.25, 0.5, 1.0, 2.0]
        aspect_ratios: [0.5, 1.0, 2.0]
        height_stride: 16
        width_stride: 16
      }
    }
    first_stage_box_predictor_conv_hyperparams {
```

Εικόνα 54: Τμήμα του αρχείου pipeline.config για το μοντέλο Faster R-CNN

Για τα δύο μοντέλα που χρησιμοποιήθηκαν, οι παράμετροι που χρησιμοποιήθηκαν μέσα στα αντίστοιχα αρχεία διαμόρφωσης τους είναι ο αριθμός το κλάσεων (3), το μέγεθος του batch (8) και ο αριθμός των βημάτων (1000 ή 2000) που θα υλοποιήσει το μοντέλο κατά την εκπαίδευση του. επίσης, ορίζονται τα paths μέσα στο αρχείο διαμόρφωσης του μοντέλου, όπου το μοντέλο θα μπορεί να διαβάσει είτε άλλα αρχεία διαμόρφωσης είτε τον χάρτη ετικετών και tfrecord για τα δεδομένα εκπαίδευσης και αξιολόγησης.

Αξίζει να αναφερθεί ότι το βήμα / step είναι μια παράμετρος που αφορά την ενημέρωση των βαρών του μοντέλου. Επομένως, ο αριθμός των βημάτων καθορίζει πόσες φορές θα ενημερωθούν τα βάρη από το εργαλείο βελτιστοποίησης (π.χ. Gradient Descent). Κατά το στάδιο της ενημέρωσης των βαρών, οι εικόνες εισόδου εισάγονται ομαδοποιημένες. Για κάθε ομαδοποιημένη είσοδο, τα βάρη ενημερώνονται μία φορά. Όταν η είσοδος καλύπτει το συνολικό αριθμό των δεδομένων εικόνων, έχει ολοκληρωθεί μία εποχή.

Κατά τη διάρκεια της εκπαίδευσης, κάθε 100 βήματα το μοντέλο τυπώνει κάποιες μετρήσεις όπως το loss (κόστος) του μοντέλου και το ρυθμό εκπαίδευσης του. Με τις μετρήσεις αυτές μπορεί να αξιολογηθεί η διαδικασία εκπαίδευσης καθώς μπορεί να βγει και ένα συμπέρασμα για την ακρίβεια που θα έχει το μοντέλο στη συνέχεια για την ανίχνευση αντικειμένων.

Επόμενο βήμα είναι να εξαχθεί το μοντέλο και τα βάρη που έχει διαμορφώσει κατά τη διαδικασία της εκπαίδευσης.

Τέλος γίνεται αξιολόγηση του μοντέλου πάνω σε εικόνες του dataset όπου δεν έχουν χρησιμοποιηθεί κατά την εκπαίδευση καθώς και σε εικόνες άγνωστες εκτός dataset, ώστε να κριθεί η ικανότητα του μοντέλου να αναγνωρίζει τις τρεις κατηγορίες των φρούτων στις οποίες έχει εκπαιδευτεί (μήλα, μπανάνες και πορτοκάλια).

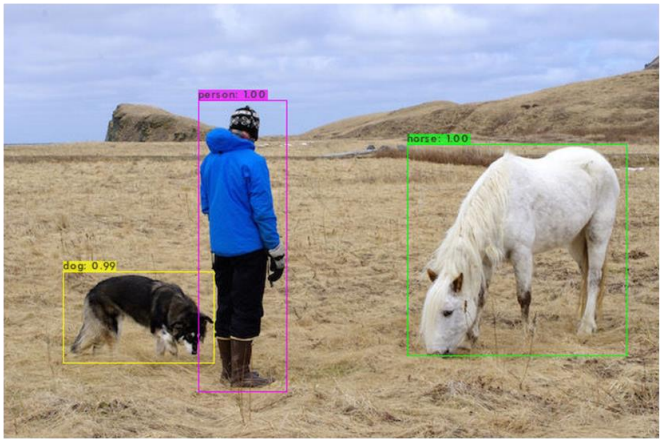
Εκπαίδευση του Μοντέλου YOLOv3

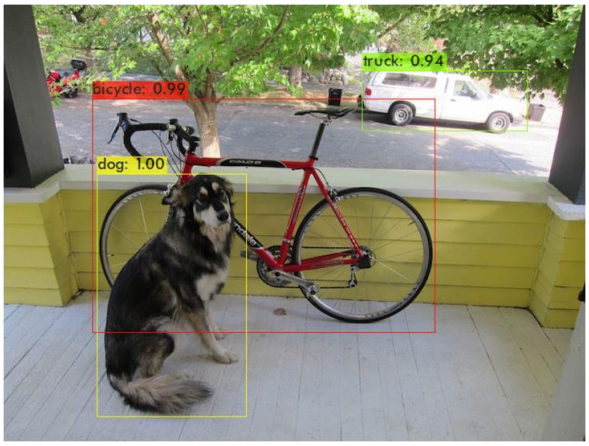
Η εκπαίδευση του δικτύου YOLOv3 υλοποιείται μέσω της βιβλιοθήκης του Darknet. Το Darknet είναι ένα πλαίσιο ανοιχτού κώδικα για νευρωνικά δίκτυα γραμμένο σε C και CUDA. Είναι γρήγορο, εύκολο στην εγκατάσταση και πραγματοποιεί υπολογισμούς σε CPU και GPU. Μέσα από το Darknet μπορεί κανείς να έχει πρόσβαση στην αρχιτεκτονική YOLOv3 που χρησιμοποιήθηκε και στη παρούσα εργασία. Η διαδικασία εκπαίδευσης του έχει ως εξής:

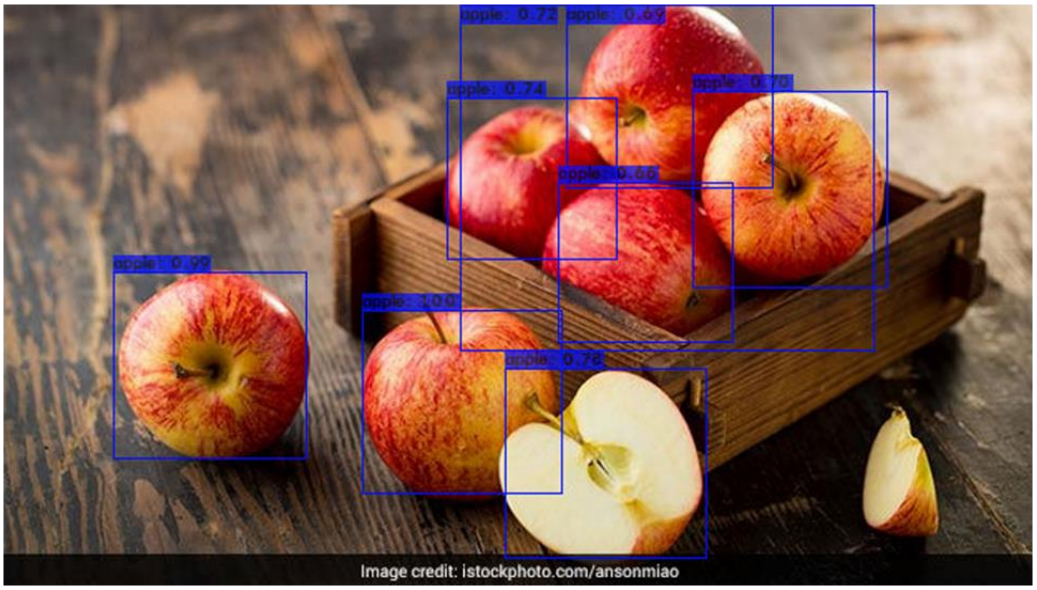
Εγκατάσταση και ρύθμιση του πλαισίου Darknet.
Εγκατάσταση του προεκπαιδευμένου YOLOv3 και των βαρών του.

Το YOLOv3 έχει ήδη εκπαιδευτεί στο σύνολο δεδομένων Coco που έχει 80 κατηγορίες άρα μπορεί να δώσει προβλέψεις σε αυτές τις 80 κατηγορίες. Όπως και τα παραπάνω μοντέλα.

Αρχικά, γίνεται μια πρώτη αξιολόγηση του προεκπαιδευμένου μοντέλου ότι έχει εγκατασταθεί σωστά και ότι λειτουργεί, εξετάζοντας αν μπορεί να ανιχνεύσει αντικείμενα πάνω σε κάποιες από τις 80 κατηγορίες που έχει εκπαιδευτεί.







Εικόνα 55: Ανιχνεύσεις του YOLOv3 ως προεκπαιδευμένο μοντέλο στο COCO dataset

Προετοιμάζονται τα δεδομένα εκπαίδευσης και αξιολόγησης όπως έχει περιγραφεί παραπάνω, δημιουργούνται πλαίσια οριοθέτησης / bounding boxes και γίνεται προσθήκη ετικετών.

Εισάγονται οι κατάλληλα διαμορφωμένες εικόνες και αρχεία επισήμανσης (δεδομένα εκπαίδευσης) στο εικονικό περιβάλλον που έχει διαμορφωθεί.

Διαμορφώνονται τα αρχεία (obj.names, obj.data, custom.cfg, train.txt) του μοντέλου ώστε να προσαρμόζονται πάνω στα χρησιμοποιούμενα δεδομένα εκπαίδευσης.

Το αρχείο obj.names είναι ένα αρχείο κειμένου που περιλαμβάνει τις κατηγορίες αντικειμένων και την αντίστοιχη αριθμητική ετικέτα που τους έχει δοθεί. Δηλαδή:

0 apple

1 banana

2 orange

Κατηγορίες ανίχνευσης αντικειμένων και αντίστοιχες αριθμητικές ετικέτες, περιεχόμενο αρχείου obj.names

Το αρχείο obj.data είναι ένα αρχείο κειμένου που περιλαμβάνει των αριθμό των κλάσεων, τις «τοποθεσίες» των δεδομένων εκπαίδευσης, αξιολόγησης, του αρχείου obj.names και ενός φακέλου backup όπου το μοντέλο κατά την διάρκεια της εκπαίδευσης μπορεί να σώζει ανά 1000 βήματα τα βάρη όπως έχουν διαμορφωθεί μέχρι στιγμής ώστε αν για κάποιο αιφνίδιο λόγο διακοπεί η εκπαίδευση να μην χρειάζεται η διαδικασία να ξεκινήσει από το μηδέν, γιατί είναι και μια χρονοβόρα διαδικασία.

classes = 3

train = data/train.txt

valid = data/test.txt

names = data/obj.names

backup = /mydrive/yolov3/backup/

Περιεχόμενο αρχείου obj.data

Διαμορφώθηκε το αρχείο custom.cfg, όπου στην ουσία περιέχει την αρχιτεκτονική του μοντέλου. Παράμετροι του μοντέλου που διαμορφώθηκαν βάση του dataset των φρούτων είναι οι εξής:

- Batch=64
- Subdivision=16
- Αριθμός κλάσεων=3
- Max_batches=2000*(αριθμός των κλάσεων) =2000*3=6000
- Αριθμός βημάτων= 80%*Max_batches=2400-2700
- Μέγεθος συνελικτικών φίλτρων πριν τα επίπεδα yolo= (αρ.κλάσεων+5)* αρ. κλάσεων=24

Το τελευταίο αρχείο διαμόρφωσης που απαιτείται για να μπορέσει να αρχίσει η εκπαίδευση είναι το αρχείο train.txt που περιέχει τους φακέλους σε όλες τις εικόνες εκπαίδευσης του dataset των φρούτων.

Στο επόμενο βήμα πραγματοποιείται λήψη των βαρών για τα συνελικτικά επίπεδα του δικτύου YOLOv3. Με τη χρήση αυτών των βαρών το μοντέλο γίνεται πιο ακριβές και εύκολα προσαρμοζόμενο στα δεδομένα εκπαίδευσης, ενώ ταυτόχρονα η διαδικασία εκπαίδευσης δεν είναι τόσο χρονοβόρα.

Ακολουθεί η διαδικασία εκπαίδευσης ώστε το μοντέλο YOLO v3 να προσαρμοστεί στα δεδομένα των φρούτων. Σε κάθε βήμα της εκπαίδευσης το μοντέλο εμφανίζει σχετικές πληροφορίες όπως το ρυθμό εκπαίδευσης και το κόστος μέσω των οποίων ελέγχεται η ομαλή λειτουργία της εκπαίδευσης.

Τέλος εξάγεται το εκπαιδευμένο μοντέλο και τα βάρη που έχουν διαμορφωθεί κατά την εκπαίδευση. Αξιολογείται η ακρίβεια του μοντέλου στην αναγνώριση αντικειμένων των τριών κατηγοριών φρούτων σε εικόνες που δεν είχαν χρησιμοποιηθεί στην εκπαίδευση του μοντέλου.

4. Παρουσίαση Αποτελεσμάτων

Σε αυτό το κεφάλαιο θα παρουσιαστούν και θα σχολιαστούν τα αποτελέσματα των πειραμάτων τόσο για την εφαρμογή της ταξινόμησης όσο και για την ανίχνευση αντικειμένων.

4.1 Αποτελέσματα Ταξινόμησης

Παρακάτω δίδονται τα αποτελέσματα των πειραμάτων της ταξινόμησης που υλοποιήθηκε για κάθε μια από τις εκδοχές που αναφέρθηκαν παραπάνω στη παράγραφο 3.3, ενώ γίνεται και σύγκριση των μοντέλων. Αρχικά θα αναφερθούν τα μεγέθη που βοηθούν στην αξιολόγηση της αποτελεσματικότητας των μοντέλων.

Ακρίβεια: Υπολογίζει το ποσοστό των προβλέψεων του μοντέλου που ταυτίζονται με τις πραγματικές τιμές του αντίστοιχου σετ δεδομένων.

Συνάρτηση Απώλειας: Υπολογίζει πόσο κοντά είναι η πρόβλεψη του μοντέλου με το πραγματικό αποτέλεσμα.

Πίνακας Σύγχυσης: Οι στήλες του πίνακα αντιπροσωπεύουν τις προβλέψεις και οι σειρές τις πραγματικές τιμές. Ο πίνακας σύγχυσης είναι πάντα ένας πίνακας δύο διαστάσεων σχήματος [n, n], όπου n είναι ο αριθμός των κλάσεων της ταξινόμησης που εξετάζεται.

Cross Entropy: μετρά την απόδοση ενός μοντέλου ταξινόμησης με βάση την πιθανότητα και το σφάλμα, όπου όσο πιο πιθανό (ή μεγαλύτερη είναι η πιθανότητα) για κάτι, τόσο μικρότερη είναι η διασταυρούμενη εντροπία.

1^{ο} Σετ Πειραμάτων

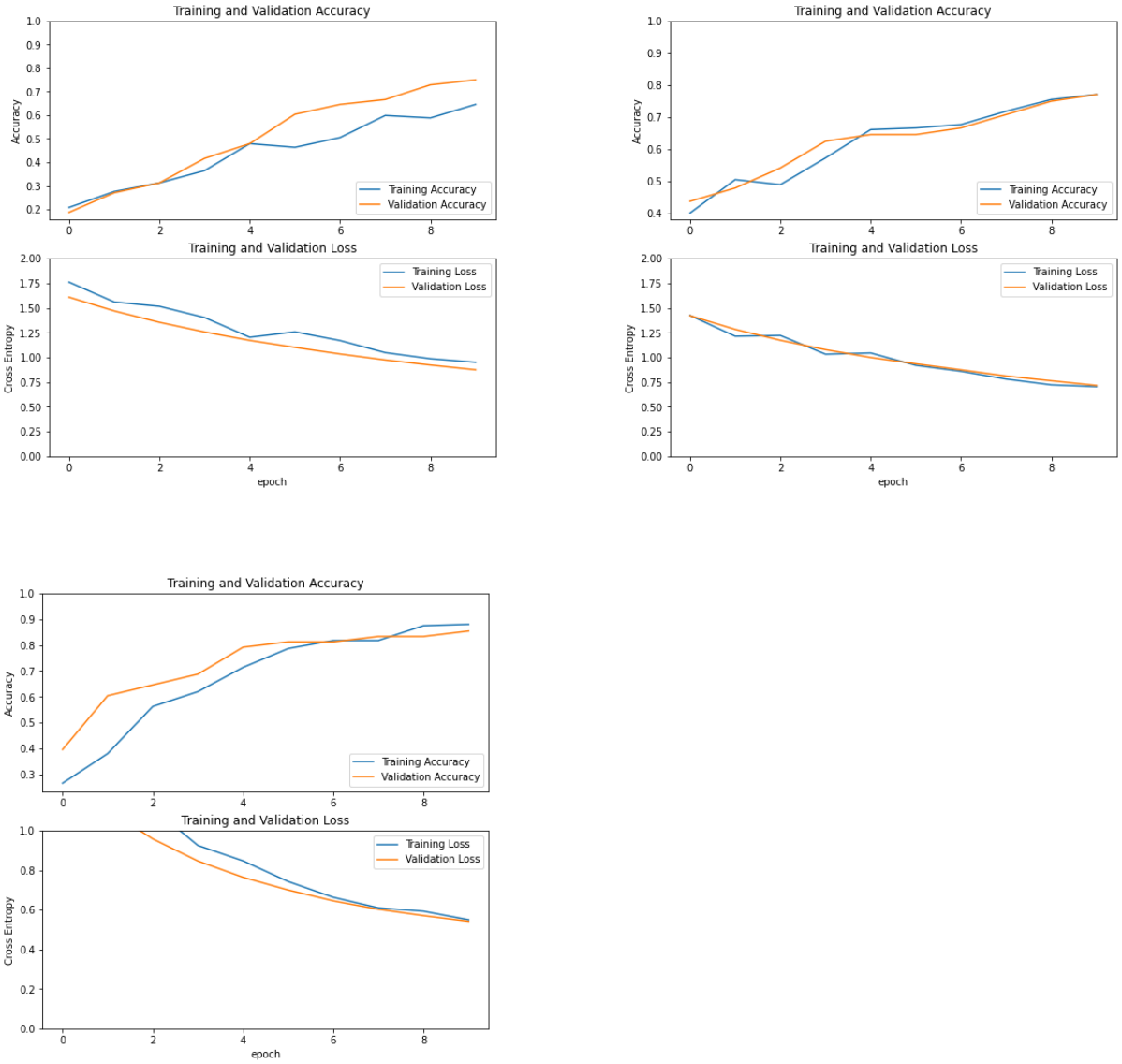
Η διαδικασία που ακολουθήθηκε για την εκπαίδευση των τριών μοντέλων περιγράφεται παραπάνω στη μεθοδολογία καθώς και η επιλογή των παραμέτρων που έγιναν. Οι συναρτήσεις που χρησιμοποιήθηκαν για την επαύξηση των δεδομένων ήταν 1) η περιστροφή της εικόνας οριζόντια και κάθετα, 2) η περιστροφή του αντικειμένου και 3) ένα τυχαίο zoom στην εικόνα.

Στο παρακάτω πίνακα παρουσιάζονται οι ακρίβειες που έδωσαν τα μοντέλα για τις 10 πρώτες εποχές εκπαίδευσης στο σετ εικόνων αξιολόγησης έχοντας εφαρμοστεί μόνο το transfer learning.

Metrics	MobileNet	Resnet50	VGG16
Accuracy	0.66	0.77	0.85
Loss	0.974	0.716	0.542

Πίνακας 2: Αποδόσεις μοντέλων

Για κάθε μοντέλο τυπώθηκαν καμπύλες που δείχνουν πως κυμαίνεται η ακρίβεια και η διασταυρούμενη εντροπία μέσα στις 10 πρώτες εποχές εκπαίδευσης.



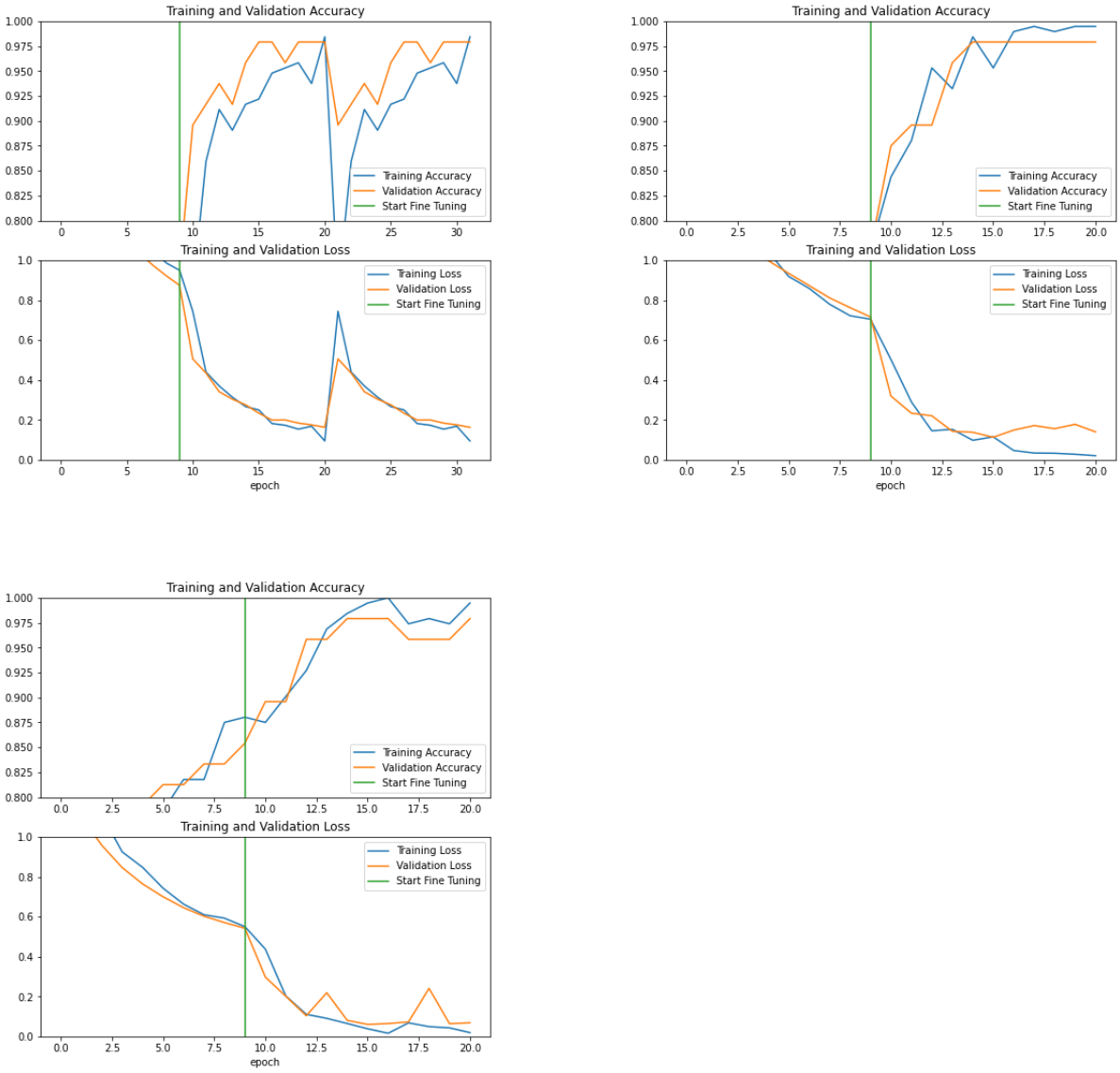
Εικόνα 56: Οι καμπύλες για τα μοντέλα MobileNet, Resnet50,VGG16

Στο παρακάτω πίνακα φαίνονται οι τελικές αποδόσεις του κάθε μοντέλου στο σετ εικόνων δοκιμής εφόσον το κάθε μοντέλο έχει ολοκληρώσει 20 εποχές εκπαίδευσης και έχει εφαρμοστεί πέρα από το transfer learning και το fine tuning.

Metrics	MobileNet	Resnet50	VGG16
Accuracy	0.9	0.93	0.93
Loss	0.168	0.195	0.268

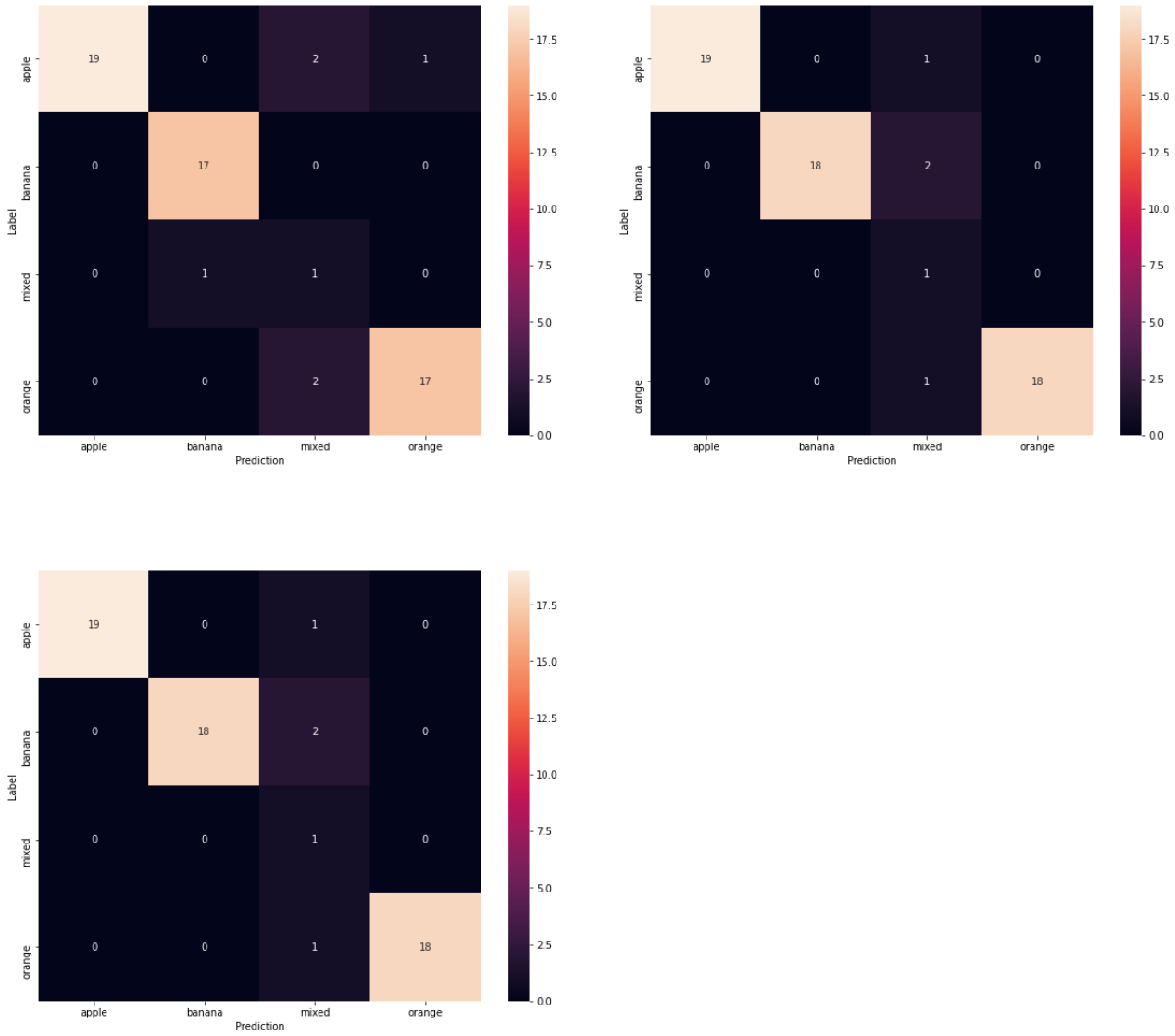
Πίνακας 3: Αποδόσεις μοντέλων

Για κάθε μοντέλο δίδονται οι καμπύλες που δείχνουν πως κυμαίνεται η ακρίβεια και η διασταυρούμενη εντροπία εφόσον το κάθε μοντέλο έχει ολοκληρώσει 20 εποχές εκπαίδευσης και έχοντας εφαρμοστεί πέρα από το transfer learning και το fine tuning.



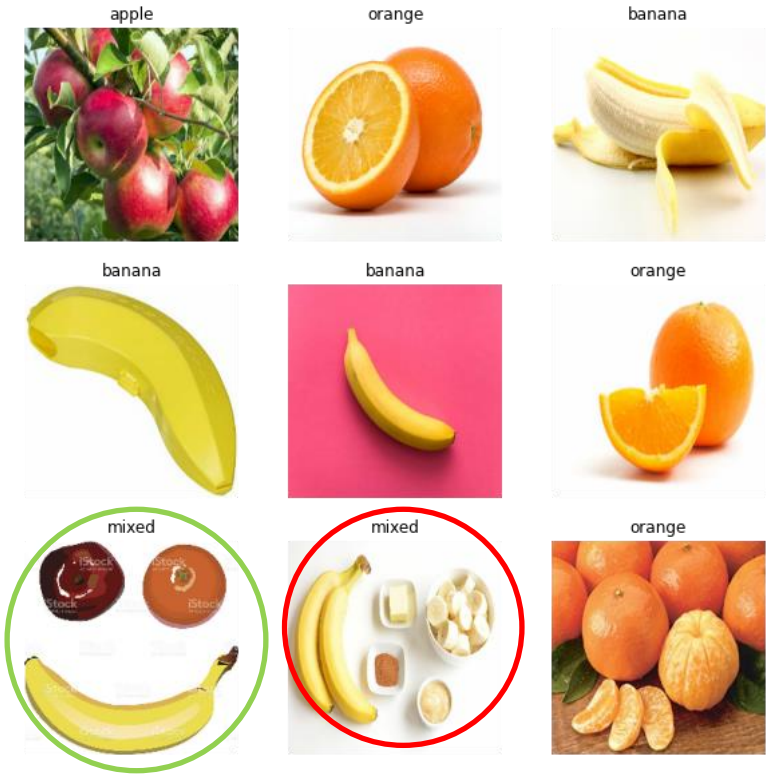
Εικόνα57: Οι καμπύλες για τα μοντέλα MobileNet, Resnet50,VGG16

Παρατίθενται οι πίνακες σύγχυσης για κάθε μοντέλο:



Εικόνα 58: Οι πίνακες σύγχυσης για τα μοντέλα MobileNet, Resnet50,VGG16

Στο παρακάτω σετ εικόνων από το test dataset το MobileNet κατά τα πρώτα πειράματα παρατηρείται άλλοτε να αναγνωρίζει και άλλοτε να συγχέει την κατηγορία mixed.



Εικόνα 59: Παραδείγματα προβλέψεων του MobileNet σωστών και λανθασμένων

2^{ο} Σετ Πειραμάτων

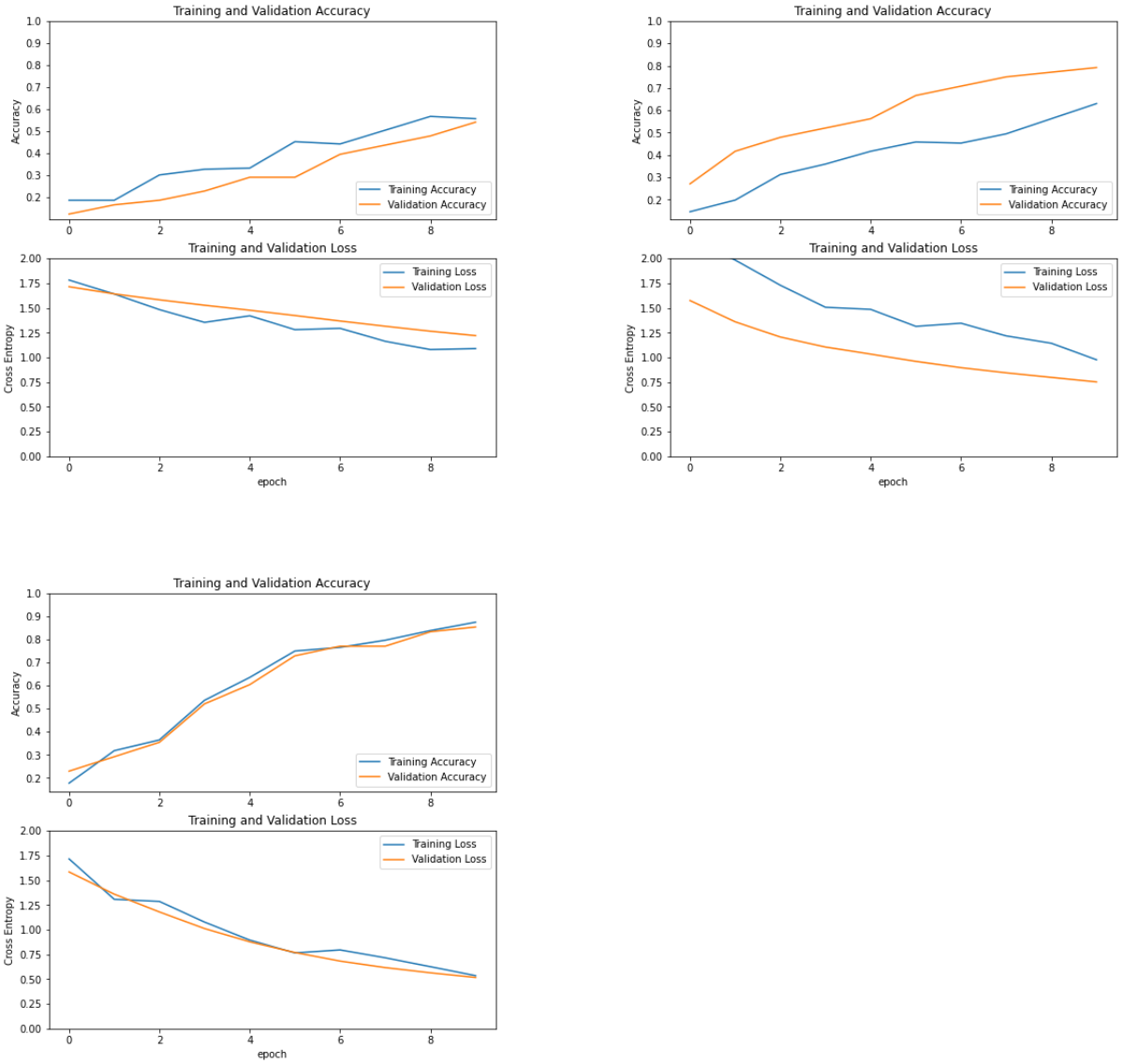
Η διαδικασία που ακολουθήθηκε για την εκπαίδευση των τριών μοντέλων περιγράφεται παραπάνω στη μεθοδολογία καθώς και η επιλογή των παραμέτρων που έγιναν. Οι συναρτήσεις που χρησιμοποιήθηκαν για την επαύξηση δεδομένων ήταν 1) η διαφοροποίηση στην αντίθεση/ contrast (μείωση) των εικόνων και 2) ο χωρικός και χρωματικός μετασχηματισμός.

Στο παρακάτω πίνακα παρουσιάζονται οι ακρίβειες που έδωσαν τα μοντέλα για τις δέκα (10) πρώτες εποχές εκπαίδευσης στο σετ εικόνων αξιολόγησης έχοντας εφαρμοστεί μόνο το transfer learning.

Metrics	MobileNet	Resnet50	VGG16
Accuracy	0.54	0.79	0.85
Loss	1.219	0.751	0.516

Πίνακας 4: Αποδόσεις μοντέλων

Για κάθε μοντέλο τυπώθηκαν καμπύλες που δείχνουν πως κυμαίνεται η ακρίβεια και η διασταυρούμενη εντροπία μέσα στις 10 πρώτες εποχές εκπαίδευσης.



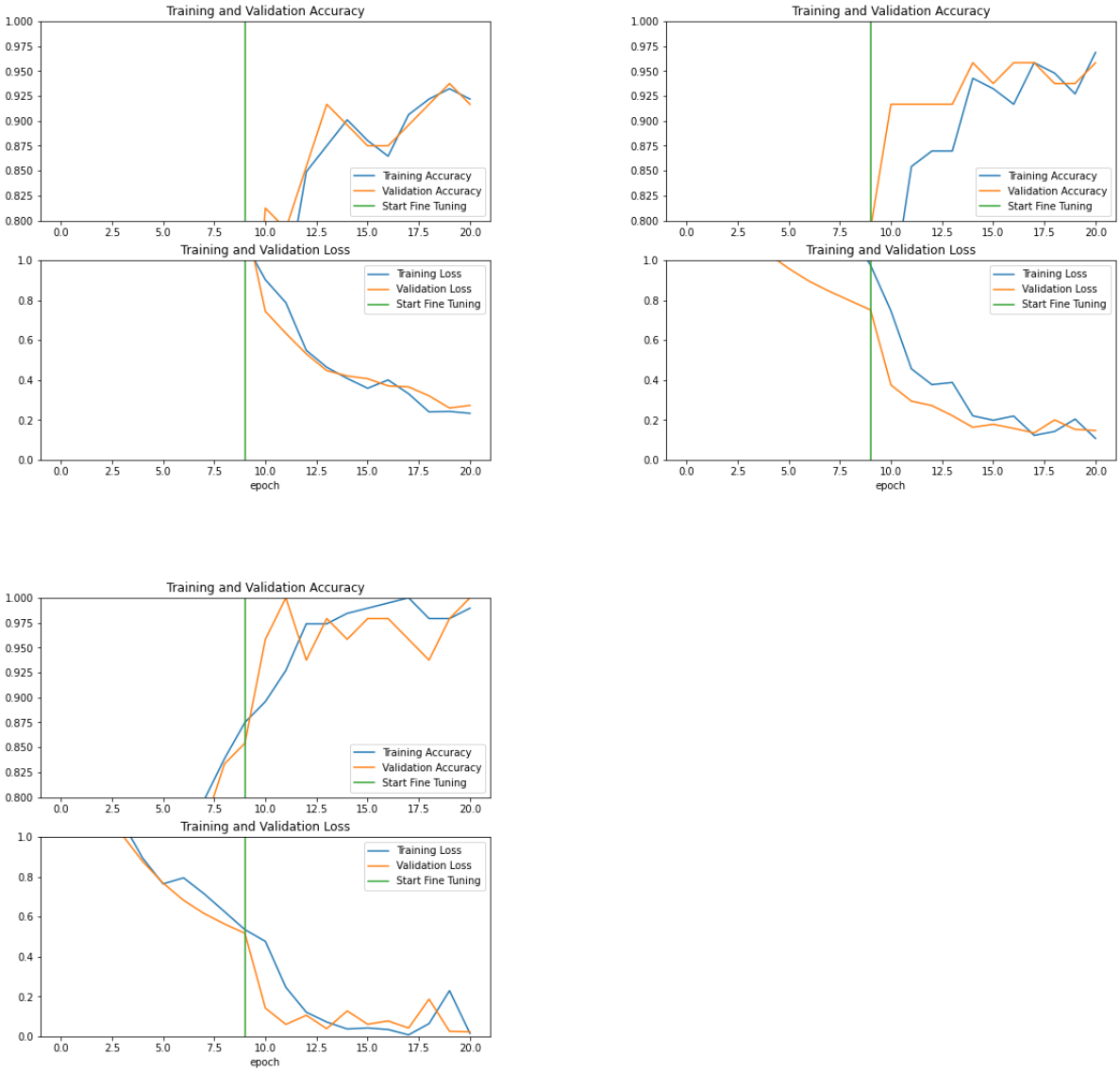
Εικόνα 60: Οι καμπύλες για τα μοντέλα MobileNet, Resnet50,VGG16

Στο παρακάτω πίνακα φαίνονται οι τελικές αποδόσεις του κάθε μοντέλου στο σετ εικόνων δοκιμής εφόσον το κάθε μοντέλο έχει ολοκληρώσει 20 εποχές εκπαίδευσης και έχοντας εφαρμοστεί πέρα από το transfer learning και το fine tuning.

Metrics	MobileNet	Resnet50	VGG16
Accuracy	0.93	0.93	0.93
Loss	0.248	0.219	0.380

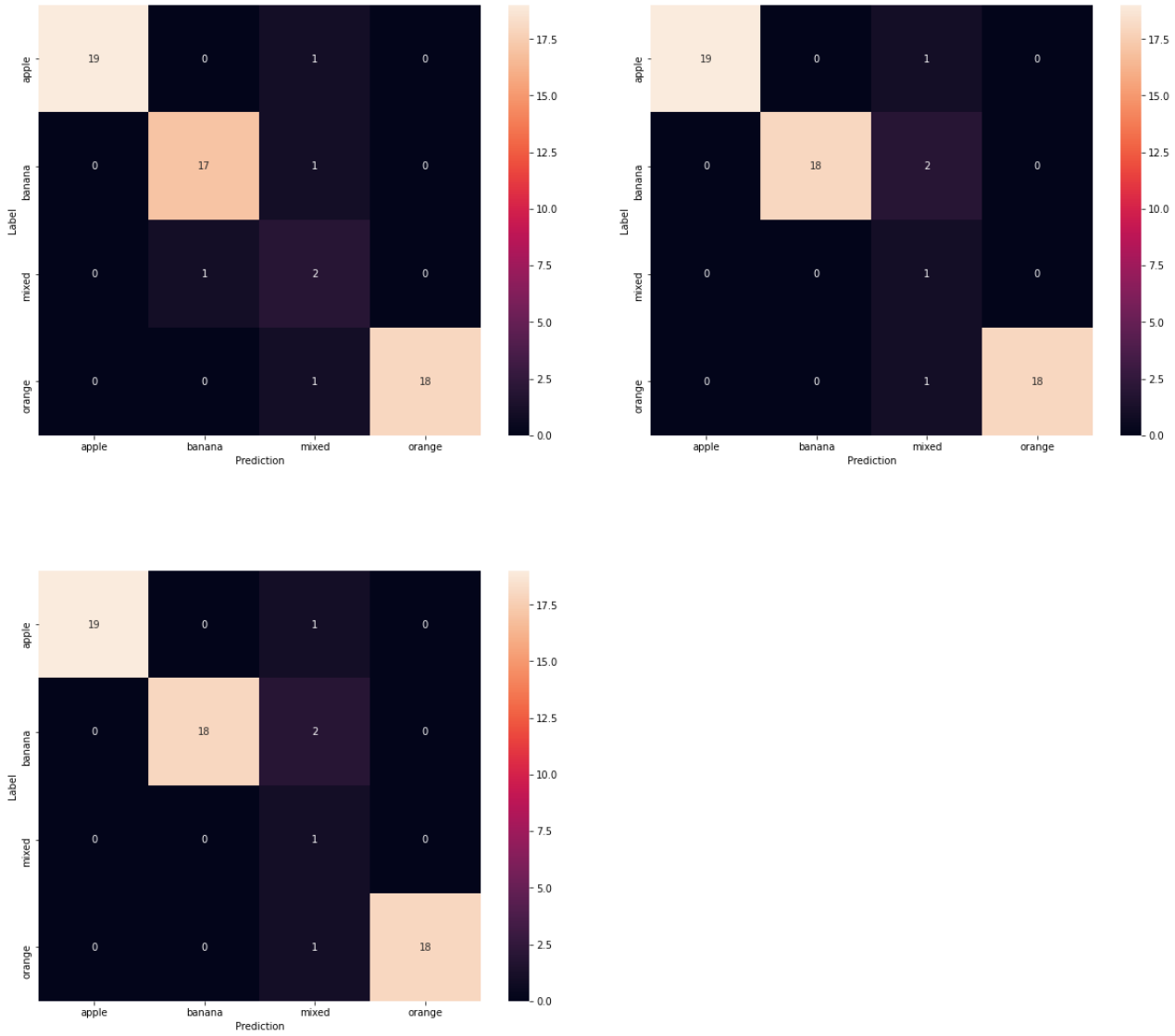
Πίνακας 5: Αποδόσεις μοντέλων

Για κάθε μοντέλο τυπώθηκαν οι καμπύλες που δείχνουν πως κυμαίνεται η ακρίβεια και η διασταυρούμενη εντροπία εφόσον το κάθε μοντέλο έχει ολοκληρώσει 20 εποχές εκπαίδευσης και έχοντας εφαρμοστεί πέρα από το transfer learning και το fine tuning.



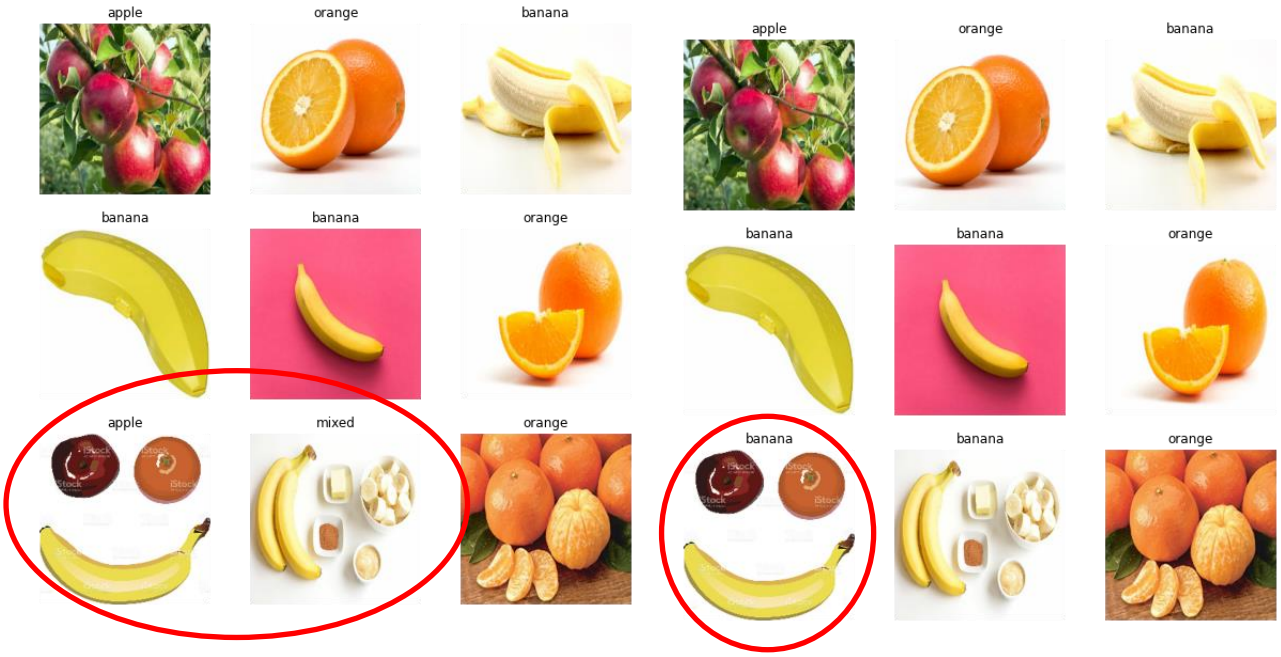
Εικόνα 61: Οι καμπύλες για τα μοντέλα MobileNet, Resnet50,VGG16

Παρατίθενται οι πίνακες σύγχυσης για κάθε μοντέλο:



Εικόνα 62: Οι πίνακες σύγχυσης για τα μοντέλα MobileNet, Resnet50,VGG16

Στο παρακάτω σετ εικόνων από το test dataset το VGG-16, δεν είχαν καμία λάθος πρόβλεψη, ενώ το MobileNet και το ResNet 50 έδειξαν να συγχέουν τις κατηγορίες μπανάνα, μήλο με mixed.



Εικόνα 63: Παραδείγματα προβλέψεων σωστών και λανθασμένων του MobileNet (αριστερά) και ResNet50 (δεξιά)

3^{ο} Σετ Πειραμάτων

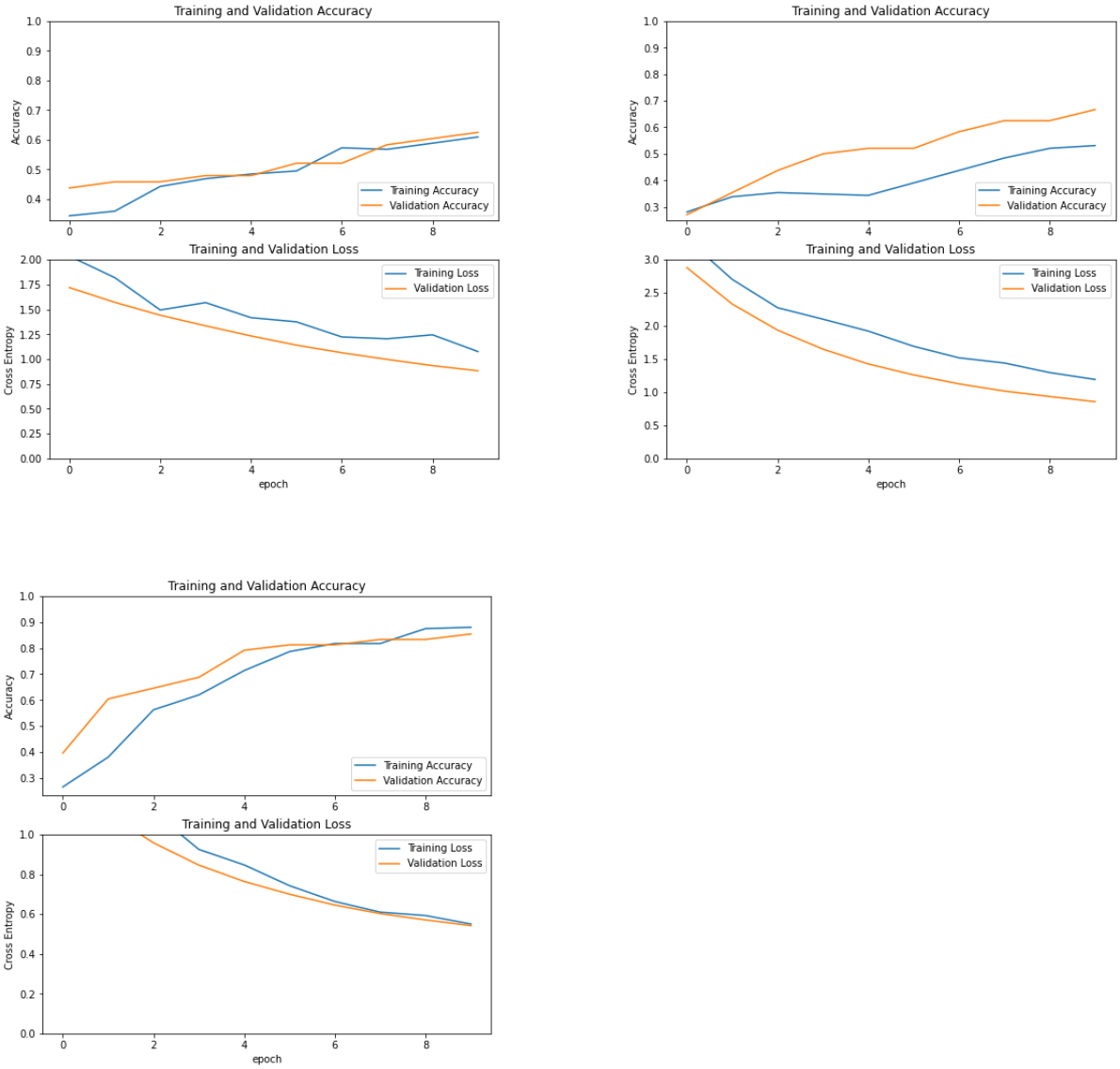
Η διαδικασία που ακολουθήθηκε για την εκπαίδευση των τριών μοντέλων περιγράφεται παραπάνω στη μεθοδολογία καθώς και η επιλογή των παραμέτρων που έγιναν. Οι συναρτήσεις που χρησιμοποιήθηκαν για επαύξηση δεδομένων ήταν 1) περιστροφή της εικόνας οριζόντια και κάθετα, 2)περιστροφή του αντικειμένου και 3) τυχαίο zoom στην εικόνα. Η διαφοροποίηση σε αυτό το σετ είναι ότι μειώθηκε η ανάλυση των εικόνων στο μισό της αρχικής ανάλυσης και επιλέχθηκε η επανασύσταση των εικόνων να γίνει με τη μέθοδο του εγγύτερου γείτονα. Σκοπός ήταν να μελετηθεί αν μικρότερης ανάλυσης εικόνων θα επηρεάζει το μοντέλο.

Στο παρακάτω πίνακα παρουσιάζονται οι ακρίβειες που έδωσαν τα μοντέλα για τις 10 πρώτες εποχές εκπαίδευσης στο σετ εικόνων αξιολόγησης έχοντας εφαρμοστεί μόνο το transfer learning.

Metrics	MobileNet	Resnet50	VGG16
Accuracy	0.62	0.66	0.70
Loss	0.882	0.857	0.716

Πίνακας 6: Αποδόσεις μοντέλων

Για κάθε μοντέλο τυπώθηκαν καμπύλες που δείχνουν πως κυμαίνεται η ακρίβεια και η διασταυρούμενη εντροπία μέσα στις 10 πρώτες εποχές εκπαίδευσης.



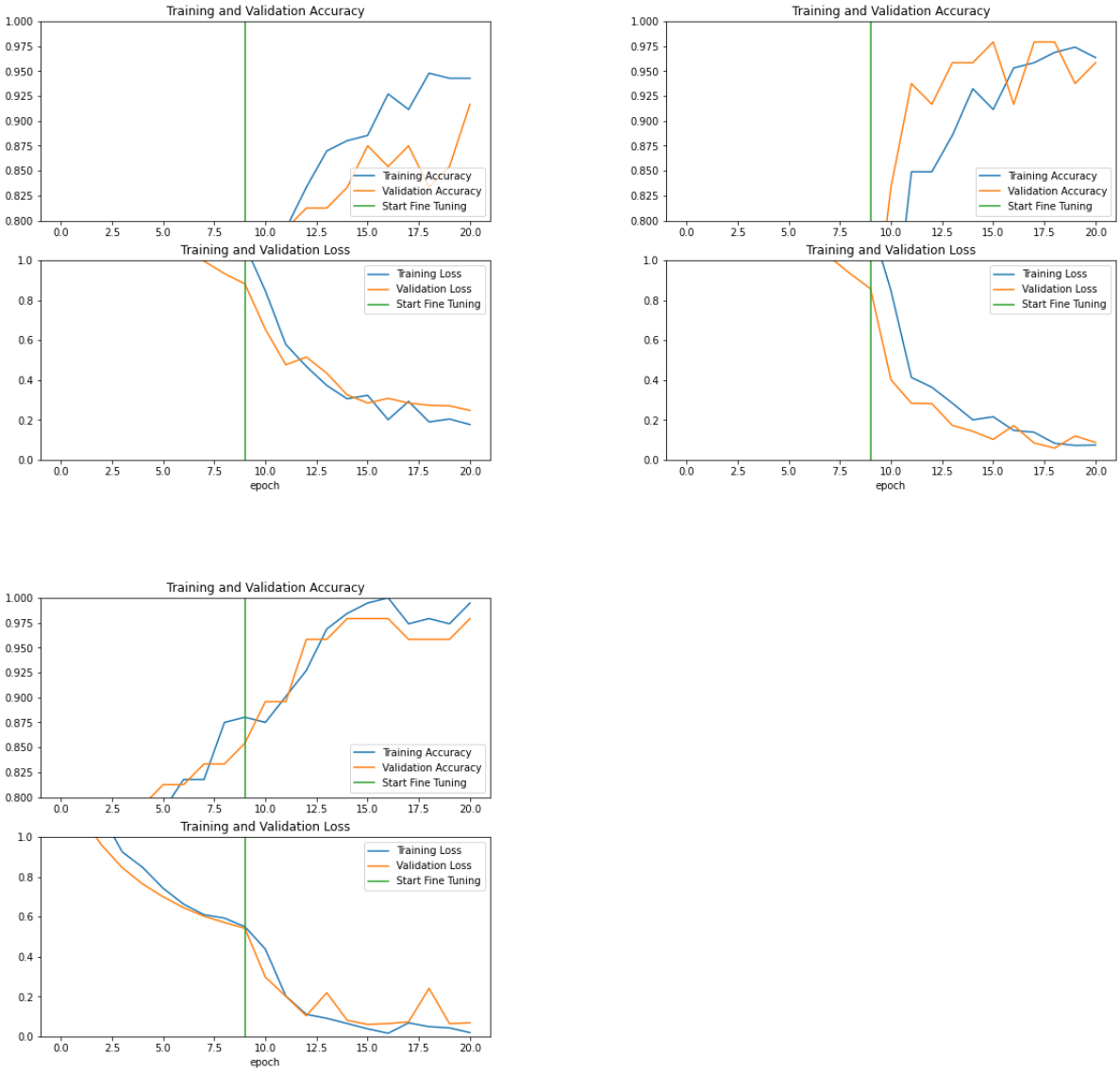
Εικόνα 64: Οι καμπύλες για τα μοντέλα MobileNet, Resnet50,VGG16

Στο παρακάτω πίνακα φαίνονται οι τελικές αποδόσεις του κάθε μοντέλου στο σετ εικόνων δοκιμής εφόσον το κάθε μοντέλο έχει ολοκληρώσει 20 εποχές εκπαίδευσης και έχοντας εφαρμοστεί πέρα από το transfer learning και το fine tuning.

Metrics	MobileNet	Resnet50	VGG16
Accuracy	0.93	0.90	0.93
Loss	0.213	0.414	0.268

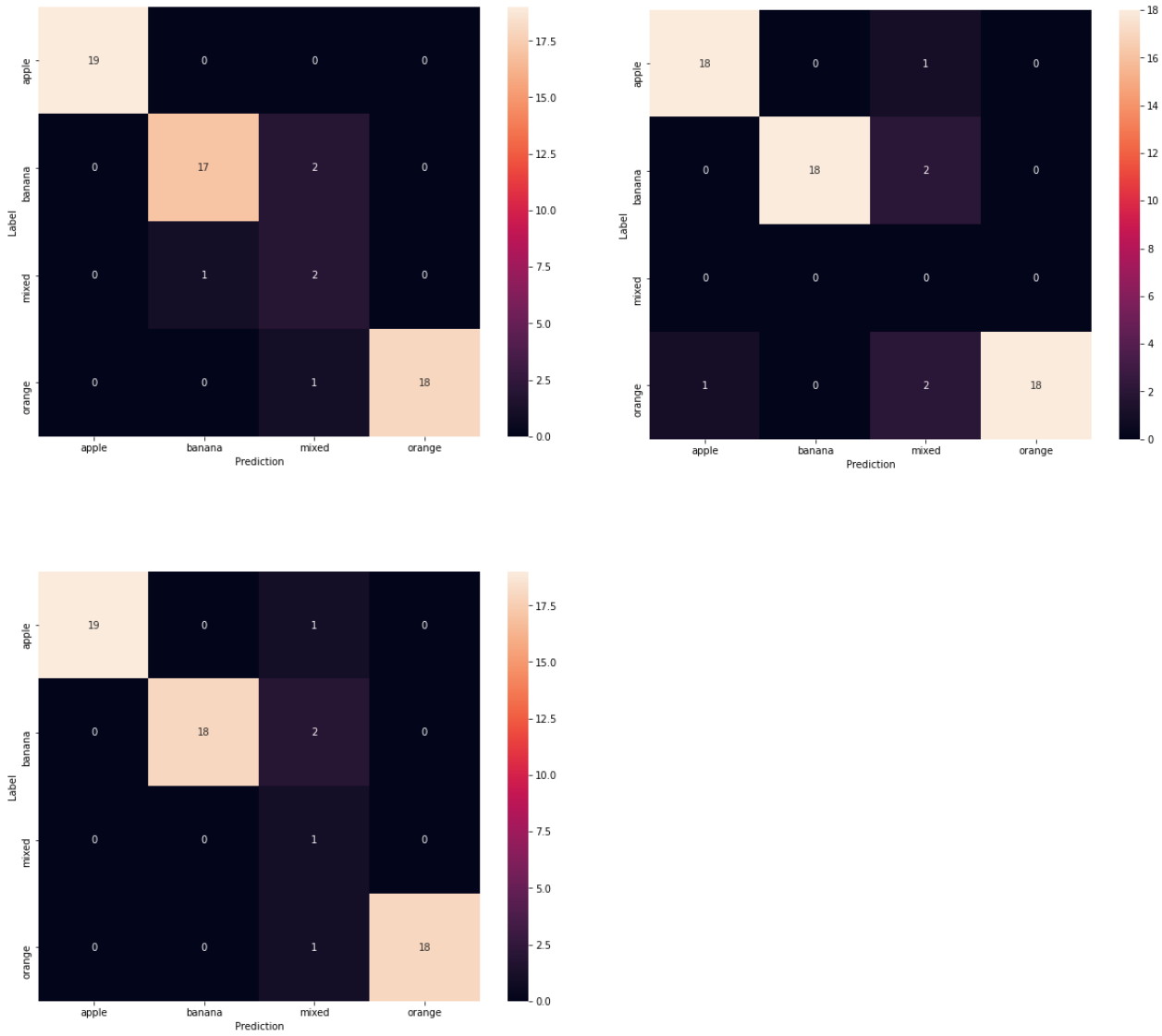
Πίνακας 7: Αποδόσεις μοντέλων

Για κάθε μοντέλο τυπώθηκαν οι καμπύλες που δείχνουν πως κυμαίνεται η ακρίβεια και η διασταυρούμενη εντροπία εφόσον το κάθε μοντέλο έχει ολοκληρώσει 20 εποχές εκπαίδευσης και έχοντας εφαρμοστεί πέρα από το transfer learning και το fine tuning.



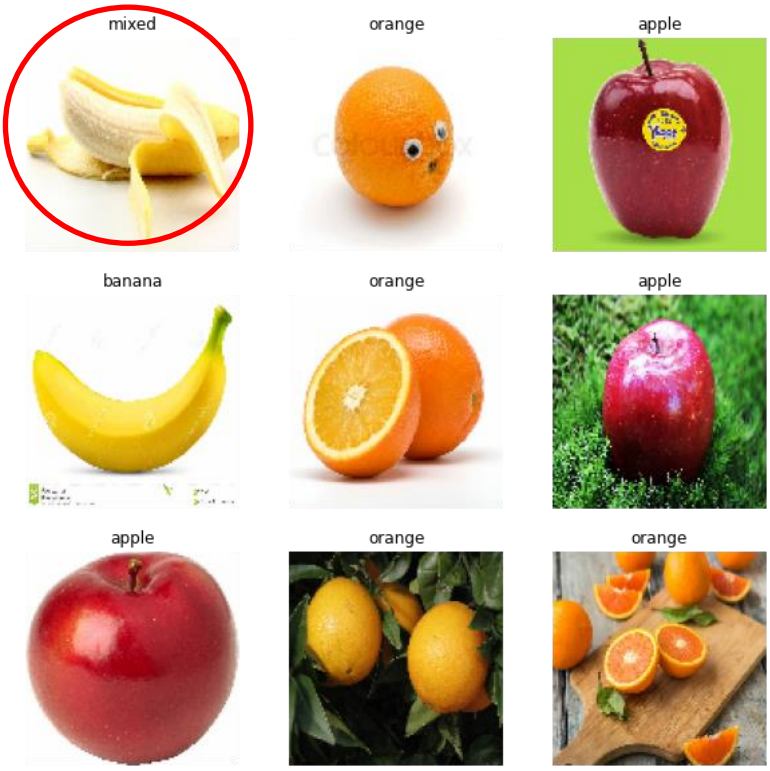
Εικόνα 65: Οι καμπύλες για τα μοντέλα MobileNet, Resnet50,VGG16

Παρατίθενται οι πίνακες σύγχυσης για κάθε μοντέλο:



Εικόνα 66: Οι πίνακες σύγχυσης για τα μοντέλα MobileNet, Resnet50,VGG16

Στο παρακάτω σετ εικόνων από το test dataset το Resnet50 και το VGG-16, δεν είχαν καμία λάθος πρόβλεψη, ενώ το MobileNet κατέταξε λανθασμένα μια εικόνα μπανάνας στην κατηγορία mixed.



Εικόνα 67: Παραδείγματα προβλέψεων σωστών και λανθασμένων του MobileNet

Παρατηρήσεις

1ο Σετ Πειραμάτων

Από τα πρώτα πειράματα επιβεβαιώνεται ότι τα μοντέλα έχουν μεγάλα ποσοστά επιτυχίας παρά το μικρό σύνολο δεδομένων. Φαίνεται ότι το Resnet έχει την καλύτερη ακρίβεια σε συνδυασμό με τη μικρότερη απώλεια. επίσης, όπως ήταν λογικό, η κατηγορία mixed εμφανίζει τη μεγαλύτερη δυσκολία να ταξινομηθεί σε σχέση με υπόλοιπα τα μοντέλα.

2ο Σετ Πειραμάτων

Ως συμπέρασμα από τα δεύτερα πειράματα, παρατηρείται ότι τα μοντέλα συνεχίζουν να έχουν μεγάλα ποσοστά ακρίβειας χρησιμοποιώντας διαφορετικές τεχνικές στο data augmentation (επαύξηση δεδομένων). Φαίνεται ότι το Resnet συνεχίζει να εμφανίζει την καλύτερη ακρίβεια σε συνδυασμό με τη μικρότερη απώλεια. Παρατηρείται ότι το MobileNet αύξησε το ποσοστό ακρίβεια του που θα μπορούσε να αποδίδεται στις διαφορετικές τεχνικές data augmentation, αλλά δεν είναι κάτι στο οποίο μπορεί να καταλήξει κανείς με σιγουριά καθώς υπάρχουν και ποσοστά τυχαιότητας στις προβλέψεις και επίσης παρατηρείται η απώλεια του να αυξάνεται.

3ο Σετ Πειραμάτων

Σαν συμπέρασμα από τα τρίτα πειράματα, παρατηρείται ότι τα μοντέλα συνεχίζουν να έχουν μεγάλα ποσοστά ακρίβειας χρησιμοποιώντας τα δεδομένα με χαμηλότερη ανάλυση. Αλλά παρατηρείται ότι έχει αυξηθεί την απώλεια σε όλα τα μοντέλα.

Ως γενικό συμπέρασμα προκύπτει ότι η ακρίβεια των μοντέλων δεν επηρεάστηκε αισθητά από τις διαφοροποιήσεις στις εικόνες. Τα μοντέλα συνεχίζουν να εμφανίζουν μεγάλα ποσοστά επιτυχίας στην ταξινόμηση. Ταυτόχρονα, παρουσιάζεται μια μικρή αύξηση της συνάρτησης απώλειας και μικρή πτώση των ποσοστών της ακρίβειας χωρίς αυτό να είναι ενιαίο σε όλα τα μοντέλα. Στη κατηγορία mixed, όπως είναι λογικό, όλα τα μοντέλα παρουσιάζουν μεγαλύτερη δυσκολία για ορθή ταξινόμηση. Επίσης είναι εμφανές ότι το transfer learning σε συνδυασμό με το fine tuning, αυξάνουν αισθητά την απόδοση όλων των μοντέλων. Για αυτό αποτελούν και τεχνικές - όπως και το data augmentation - που χρησιμοποιούνται ευρύτατα σε ερευνητικές προσπάθειες.

4.2 Αποτελέσματα Ανίχνευσης Αντικειμένου

Στην ενότητα αυτή θα παρουσιαστούν τα αποτελέσματα των πειραμάτων της ανίχνευση αντικειμένου που υλοποιήθηκαν, για καθένα από τα μοντέλα που αναφέρθηκαν παραπάνω. Αρχικά αναφέρονται τα μεγέθη που θα βοηθήσουν να αξιολογηθεί η αποτελεσματικότητα των μοντέλων.

Ακρίβεια: Δηλώνει πόσο ακριβείς είναι οι προβλέψεις, δηλαδή το ποσοστό των προβλέψεών που είναι σωστές.

Συνάρτηση Απώλειας: Υπολογίζει πόσο κοντά είναι η πρόβλεψη του μοντέλου με το πραγματικό αποτέλεσμα.

Loss localization /Απώλεια Εντοπισμού: είναι η απώλεια (L1) μεταξύ των παραμέτρων του πλαισίου οριοθέτησης που προέβλεψε το μοντέλο και των παραμέτρων του πλαισίου οριοθέτησης που έχει οριστεί ως ground truth. Αυτές οι παράμετροι περιλαμβάνουν τις μετατοπίσεις για το κεντρικό σημείο (cx, cy), το πλάτος (w) και το ύψος (h) του πλαισίου οριοθέτησης.

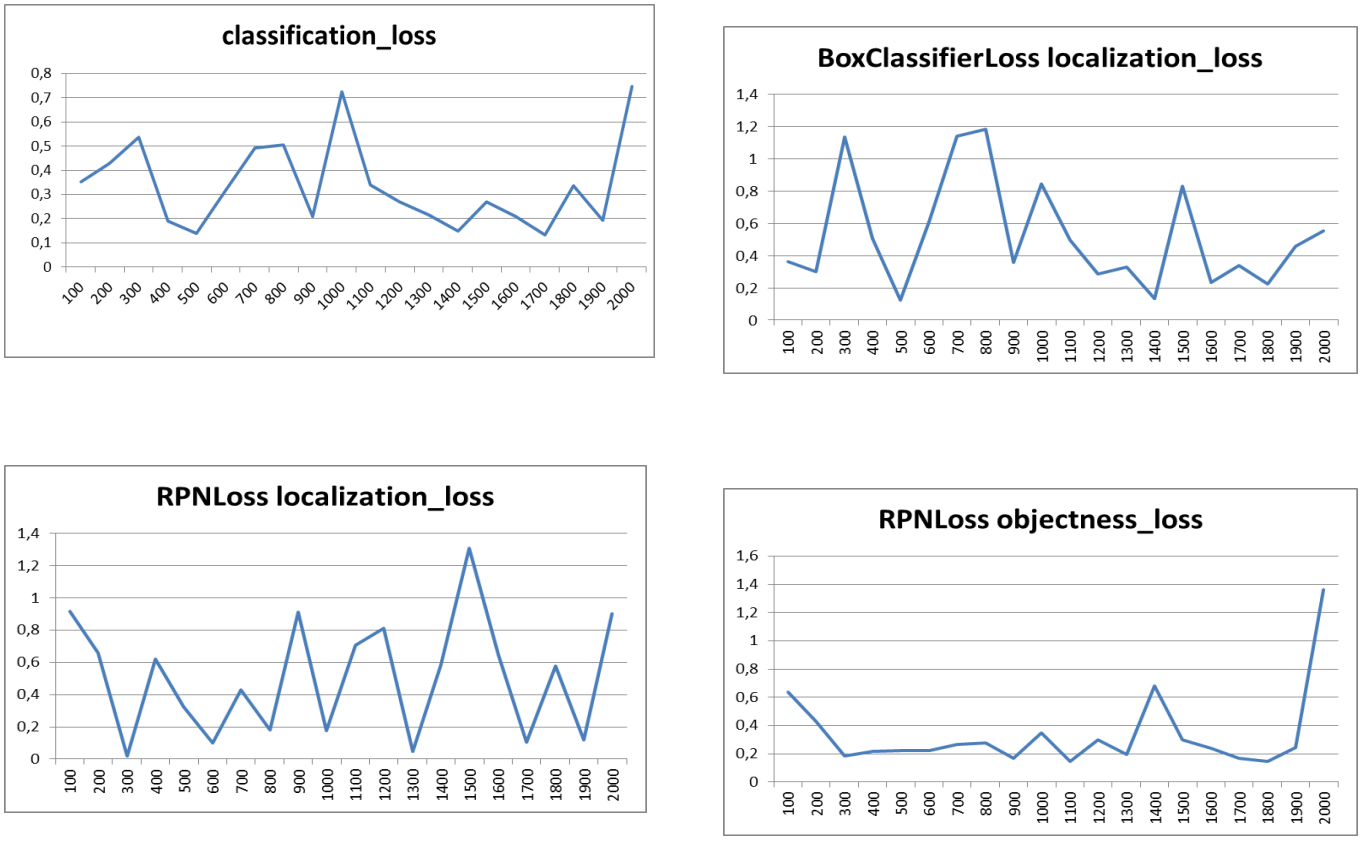
RPN localizationκαι objectness losses, BoxClassifierLoss classification και localization losses: Οι τιμές των 4 αυτών απωλειών αποτελούν αξιολόγηση για μοντέλα δύο σταδίων. Επιπρόσθετα προκύπτουν και άλλες τιμές απωλειών τόσο για το τμήμα του δικτύου που προτείνει περιοχές όσο και διαφορετικές τιμές για το τμήμα του δικτύου που αποτελεί τον ταξινομητή.

Total loss/ συνολική απώλεια: είναι το σταθμισμένο άθροισμα των τεσσάρων απωλειών (RPN localization και objectness losses, BoxClassifier Loss classification και localization losses).

Παρακάτω παρατίθενται τα αποτελέσματα των τριών μοντέλων Faster R-CNN, SSD και YOLO v3 που εκπαιδεύτηκαν πάνω στο dataset των φρούτων που χρησιμοποιείται στην παρούσα διπλωματική. Τα αποτελέσματα αφορούν τις δοκιμές των μοντέλων να αναγνωρίσουν τις τρεις κατηγορίες φρούτων στο σετ δεδομένων αξιολόγησης που δεν χρησιμοποιήθηκε κατά τη διάρκεια της εκπαίδευσης.

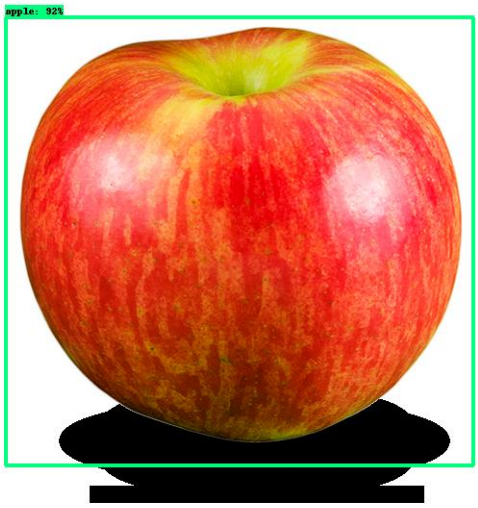
Faster R-CNN

Η διαδικασία εκπαίδευσης για το Faster R–CNN φαίνεται με βάση τα γραφήματα αξιολόγησης που παρουσιάζονται παρακάτω και τα παραδείγματα ανίχνευσης σε εικόνες αξιολόγησης, ότι δεν κύλησε ομαλά. Αυτό συμπεραίνεται καθώς σε όλες τις συναρτήσεις απώλειας φαίνεται να αυξάνονται οι τιμές τους, ενώ θα έπρεπε σταδιακά να μειώνονται κατά την διαδικασία της εκπαίδευσης.



Εικόνα 68: Γραφήματα συναρτήσεων απώλειας για το μοντέλο Faster R-CNN

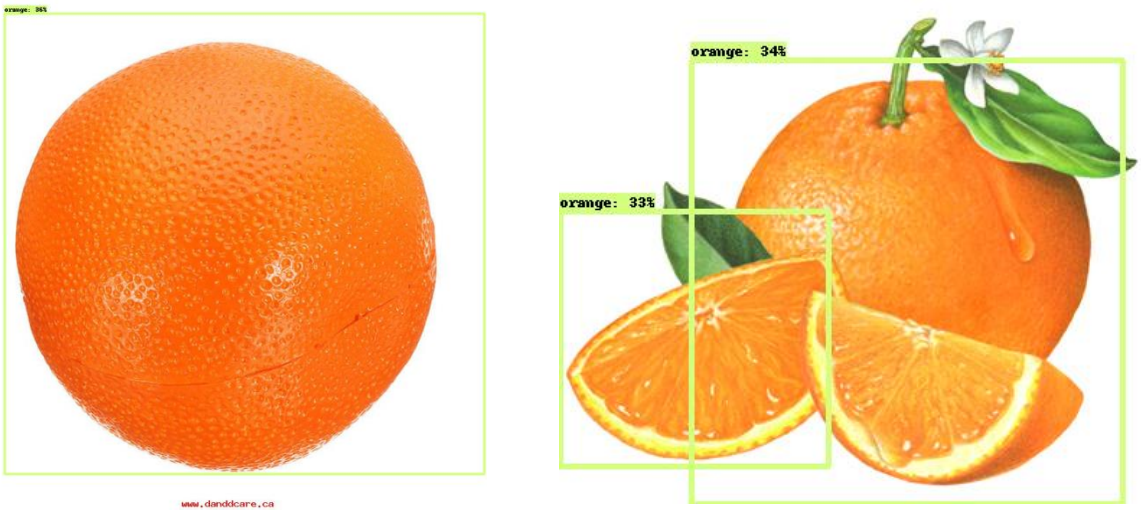
Εντούτοις, με επιτυχία αναγνώρισε το μοντέλο την κατηγορία μήλο αν και σε εικόνες με πάνω από ένα αντικείμενα της κατηγορίας δεν κατάφερε πάντα να τα αναγνωρίσει όλα.



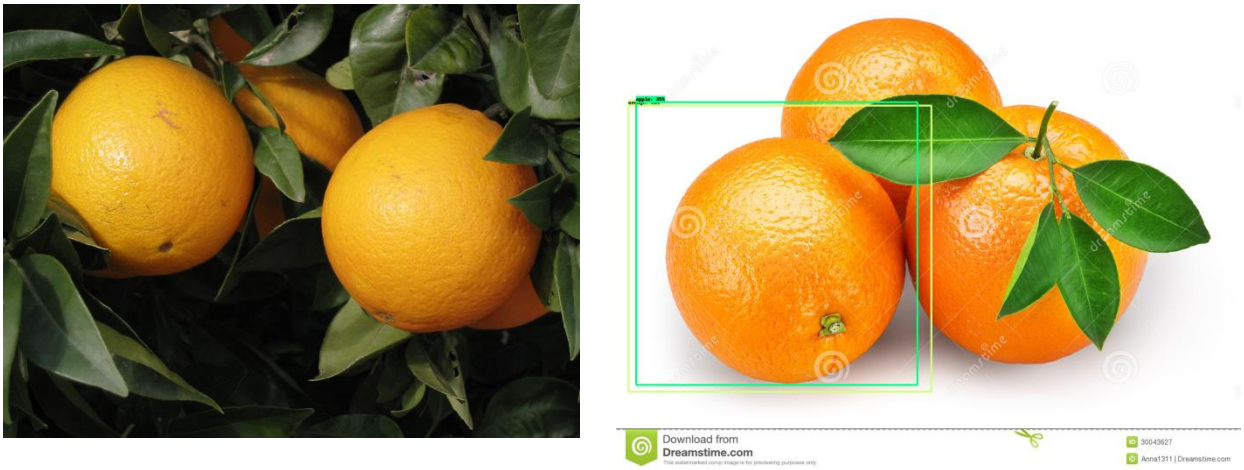
Εικόνα 69: Παραδείγματα σωστής ανίχνευσης του δικτύου Faster R-CNN για την κλάση μήλο με ακρίβεια 94% (αριστερά) και 88% (δεξιά) και 92%(αριστερά κάτω)

Εικόνα 70: Παραδείγματα ανίχνευσης του δικτύου Faster R-CNN για την κλάση μήλο με ακρίβεια 30% (αριστερά) και 60% και 47%(δεξιά) δεν αναγνωρίζει όλα τα αντικείμενα της κατηγορίας που απεικονίζονται

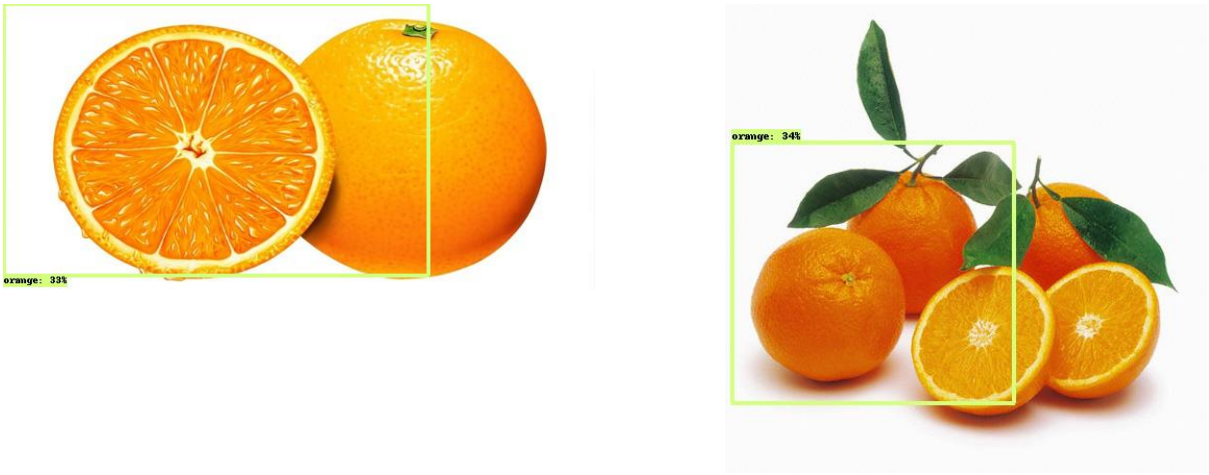
Το μοντέλο αναγνώρισε την κατηγορία πορτοκάλι αν και σε εικόνες με περισσότερα από ένα αντικείμενα της κατηγορίας δεν κατάφερε πάντα να τα αναγνωρίσει όλα. Επίσης σε κάποιες εικόνες το μοντέλο δεν κατάφερε να αναγνωρίσει καθόλου την κατηγορία πορτοκάλι παρότι οι εικόνες περιείχαν μόνο τέτοια αντικείμενα. Τέλος το ποσοστό ακρίβειας ανίχνευσης σε αυτή τη κατηγορία ήταν σχετικά χαμηλό, γύρω στο 34%.



Εικόνα 71: Παραδείγματα σωστής ανίχνευσης του δικτύου Faster R-CNN για την κλάση πορτοκάλι με ακρίβεια 36% (αριστερά) και33% και 34%(δεξιά)

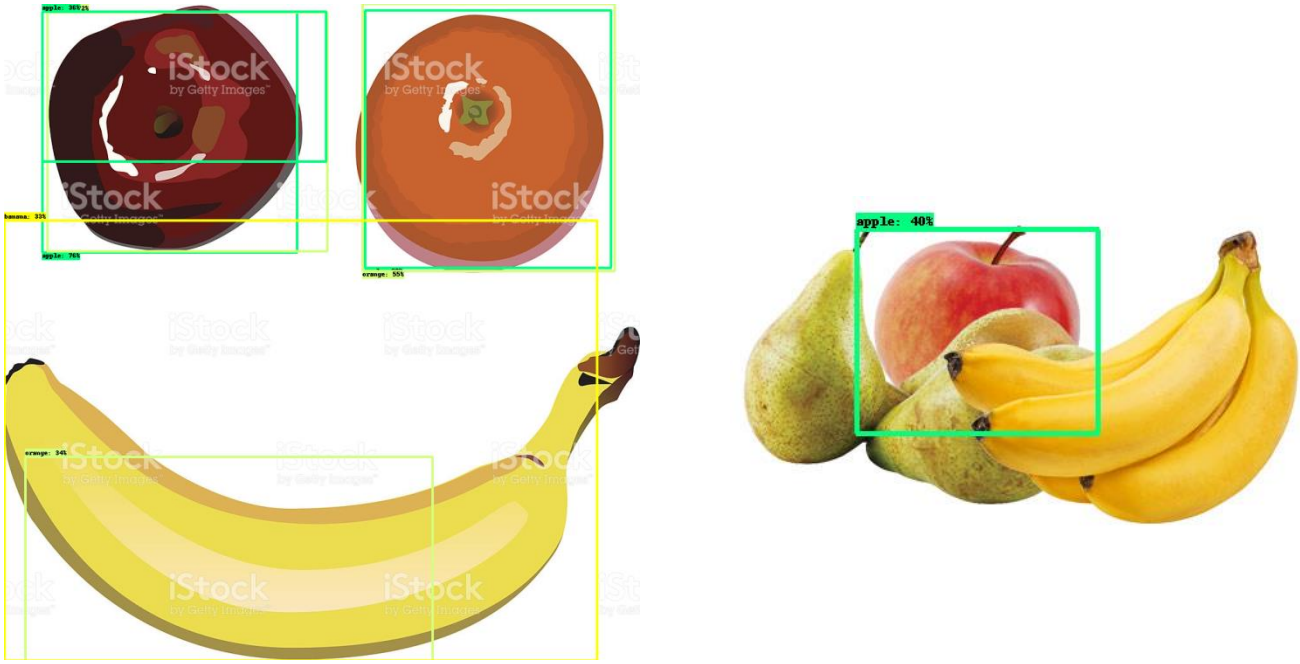


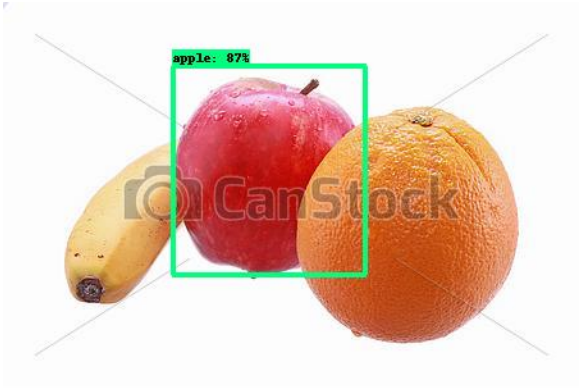
Εικόνα 72: Παραδείγματα λάθος ανίχνευσης του δικτύου Faster R-CNN για την κλάση πορτοκάλι (αριστερά) το μοντέλο δεν κατάφερε να αναγνωρίσει τα αντικείμενα και (δεξιά) δεν έχει καταφέρει να καταλήξει σε μοναδικό bounding box



Εικόνα 73: Παραδείγματα ανίχνευσης του δικτύου Faster R-CNN για την κλάση πορτοκάλι με ακρίβεια 33% (αριστερά) και 34% όμως δεν αναγνωρίζει όλα τα αντικείμενα της κατηγορίας που απεικονίζονται

Στη κατηγορία του φρούτου μπανάνα το μοντέλο δεν κατάφερε να ανιχνεύσει το αντικείμενο. Παρακάτω παρουσιάζονται εικόνες αξιολόγησης όπου το μοντέλο έχει λειτουργήσει ατελώς ή και εσφαλμένα.

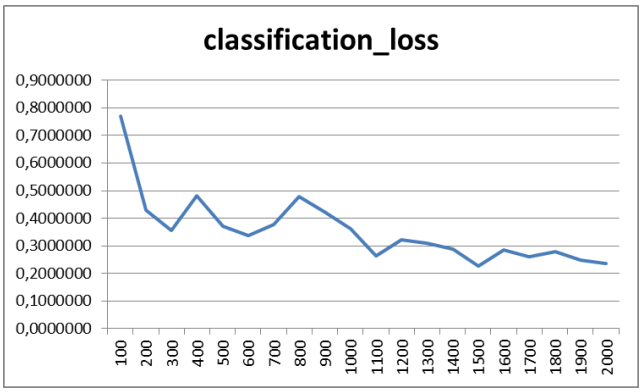


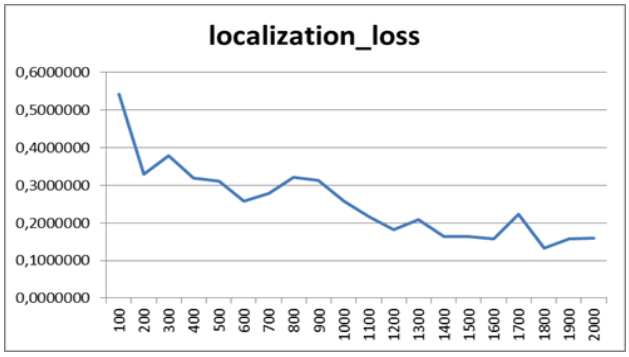


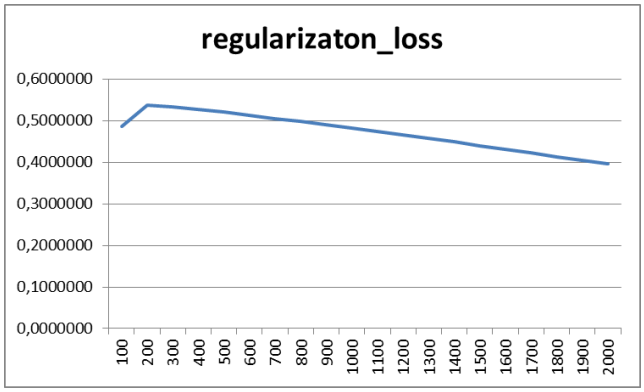
Εικόνα 74: Στην εικόνα αριστερά πάνω το μοντέλο Faster R-CNN έχει τοποθετήσει λανθασμένα πλαίσια οριοθέτησης από όλες τις κατηγορίες και δεν έχει καταλήξει σε μοναδικά πλαίσια ανά κατηγορία, ενώ στις άλλες 2 εικόνες αναγνωρίζει μόνο την κατηγορία μήλο ενώ τις άλλες όχι

SSD

Η διαδικασία εκπαίδευσης για το μοντέλο SSD, φαίνεται με βάση τα γραφήματα αξιολόγησης που παρουσιάζονται παρακάτω και τα παραδείγματα ανίχνευσης σε εικόνες αξιολόγησης, ότι παρουσιάζει παρόμοια προβλήματα με το μοντέλο Faster R-CNN. Αυτό συμπεραίνεται καθώς σε όλες τις συναρτήσεις απώλειας φαίνεται οι τιμές τους, ενώ θα έπρεπε σταδιακά να μειώνονται μέσα στη διαδικασία της εκπαίδευσης, παρουσιάζουν αρκετές εξάρσεις και όχι ομαλή σταδιακή μείωση.

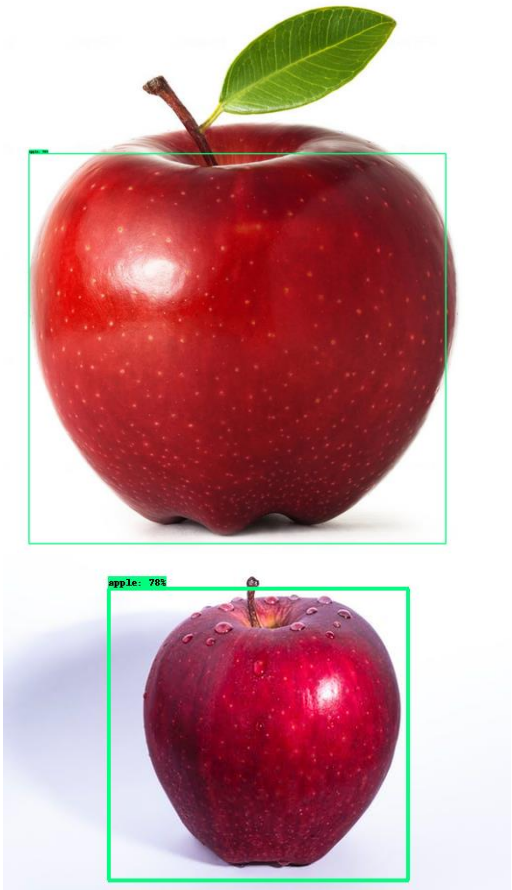


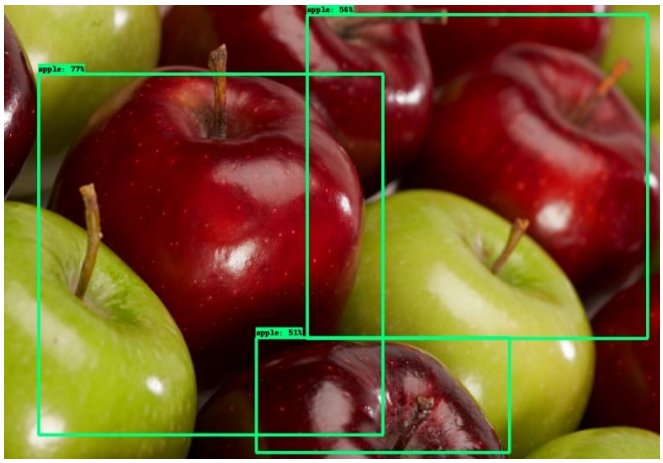




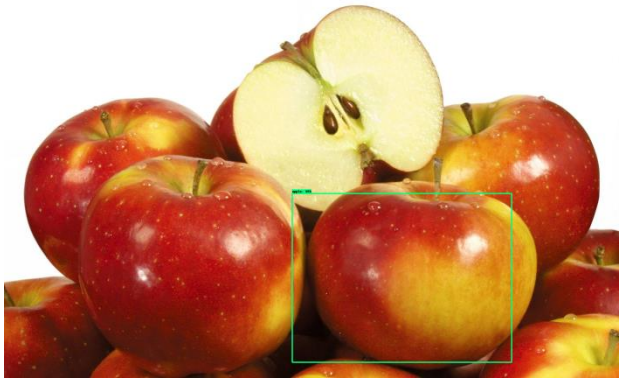
Εικόνα 75: Γραφήματα συναρτήσεων απώλειας για το μοντέλο SSD

Με επιτυχία αναγνώρισε το μοντέλο την κατηγορία μήλο σε κάποιες εικόνες αξιολόγησης αν και σε εικόνες με πάνω από ένα αντικείμενα της κατηγορίας δεν κατάφερε πάντα να τα αναγνωρίσει όλα. Τέλος σε κάποιες εικόνες δεν κατάφερε καν αναγνωρίσει την κατηγορία.



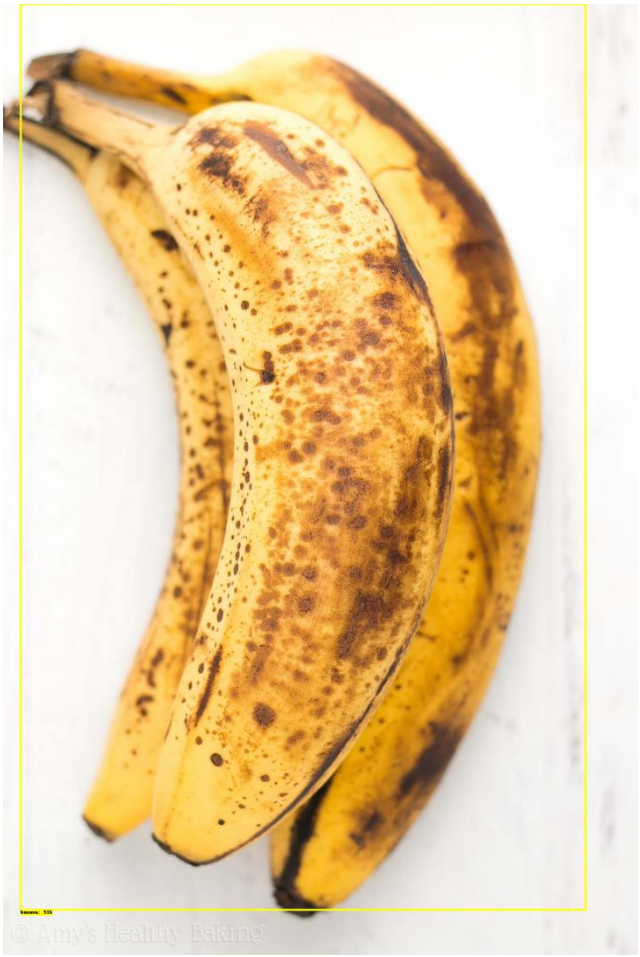


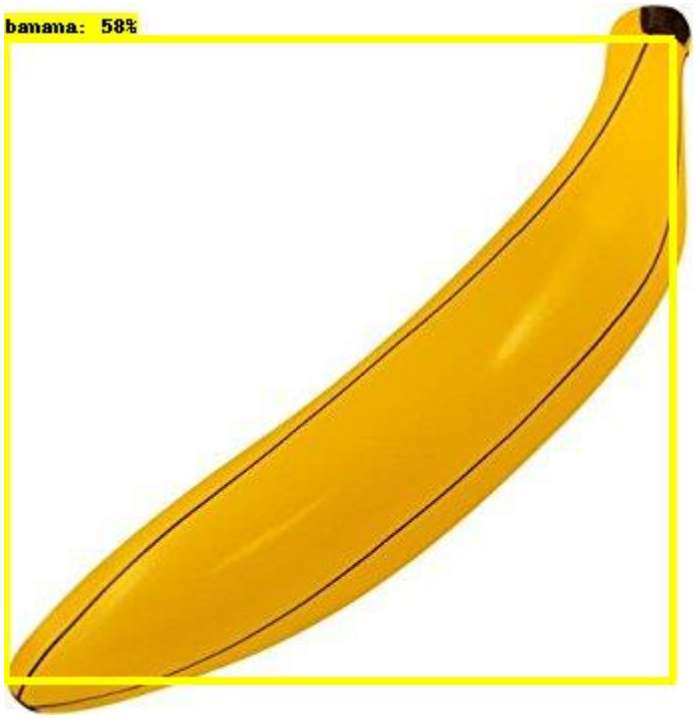
Εικόνα 76: Παραδείγματα σωστής ανίχνευσης του δικτύου SSD για την κλάση μήλο με ακρίβεια 76% (αριστερά) και 77%-51%(δεξιά) και 78%(αριστερά κάτω)

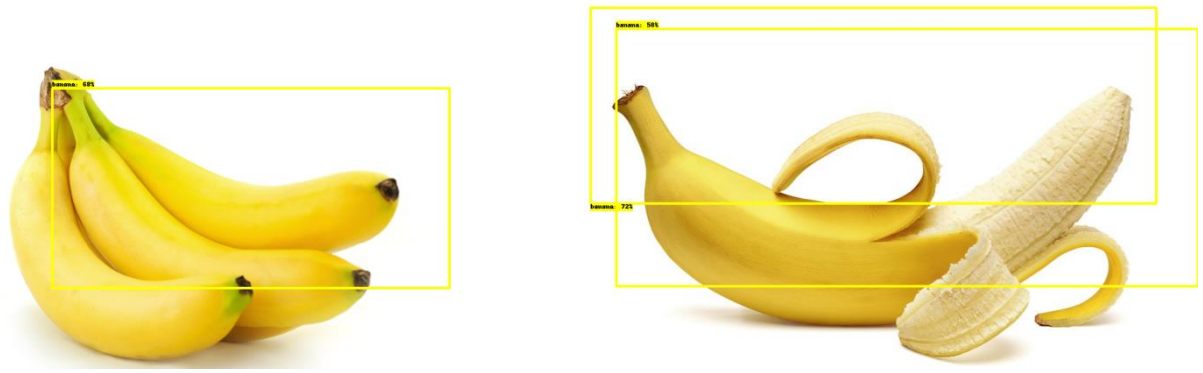


Εικόνα 77: Παράδειγμα ανίχνευσης του δικτύου SSD για την κλάση μήλο με ακρίβεια 65% αλλά δεν αναγνωρίζει όλα τα αντικείμενα της κατηγορίας που απεικονίζονται

Την κατηγορία μπανάνα την αναγνώρισε το μοντέλο SSD σχετικά επιτυχώς με ποσοστό ακρίβειας γύρω στο 60%. Παρακάτω παρατίθενται παραδείγματα από τις εικόνες αξιολόγησης.

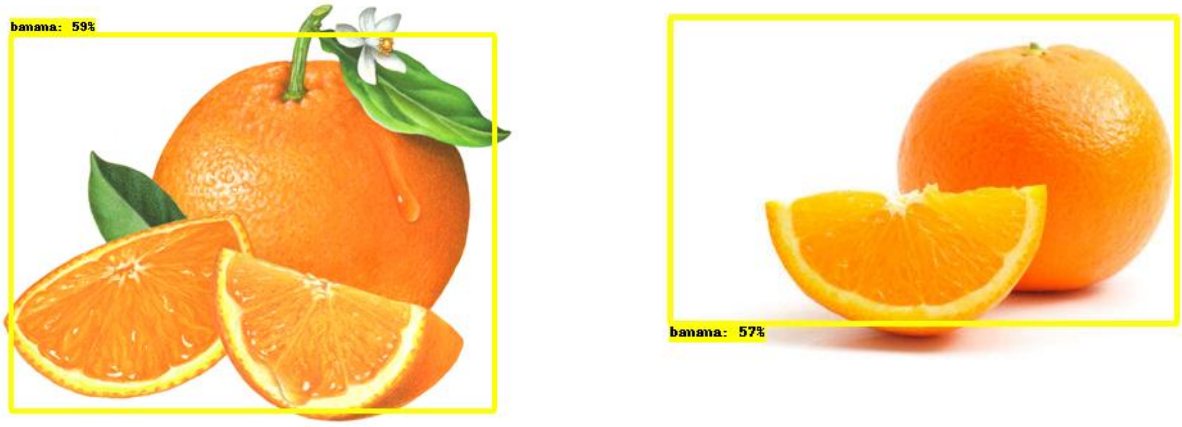


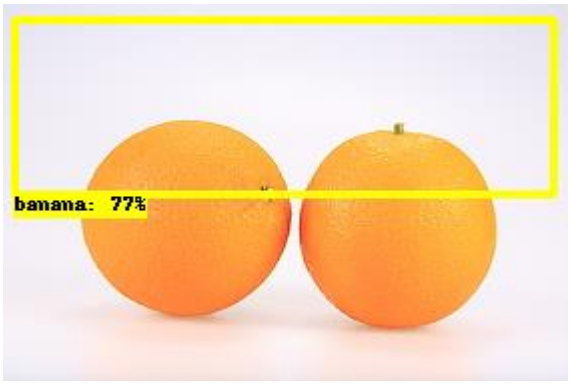


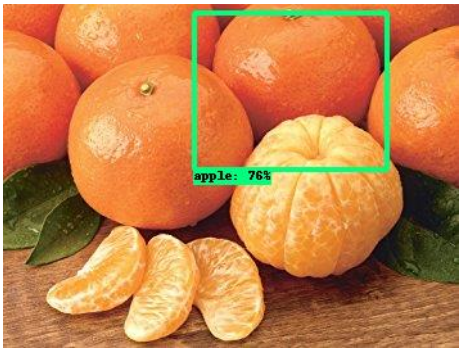


Εικόνα 78: Παραδείγματα σωστής ανίχνευσης του δικτύου Faster R-CNN για την κλάση μπανάνα με ακρίβεια 51% (αριστερά πάνω), 58% (δεξιά πάνω)και 68%(αριστερά κάτω). Στην κάτω δεξιά εικόνα το μοντέλο αναγνωρίζει την κατηγορία αλλά δεν καταλήγει σε ένα πλαίσιο οριοθέτησης

Στη κατηγορία του φρούτου πορτοκάλι το μοντέλο SSD δεν κατάφερε να την ανιχνεύσει σωστά, την συγχέει με την κατηγορία μπανάνα αλλά και με τη κατηγορία μήλο. Παρακάτω παρουσιάζονται εικόνες αξιολόγησης όπου το μοντέλο έχει λειτουργήσει εσφαλμένα.

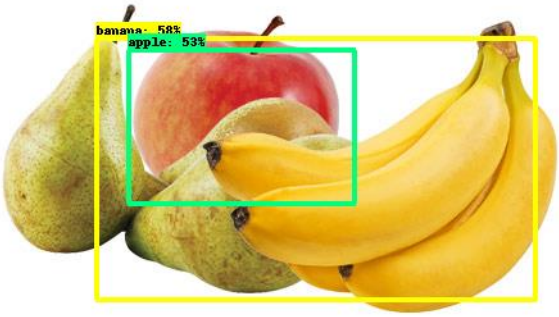


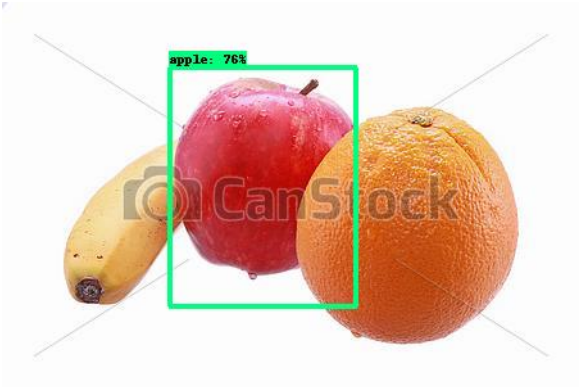




Εικόνα 79: Παραδείγματα ανίχνευσης του δικτύου SSD για την κλάση πορτοκάλι όπου εσφαλμένα εντόπισε τα ατικείμενα και τα κατέταξε στην κλάση μπανάνα και μήλο

Στις παρακάτω εικόνες παρουσιάζονται επιτυχείς και ελλιπείς ανιχνεύσεις του μοντέλου SSD.

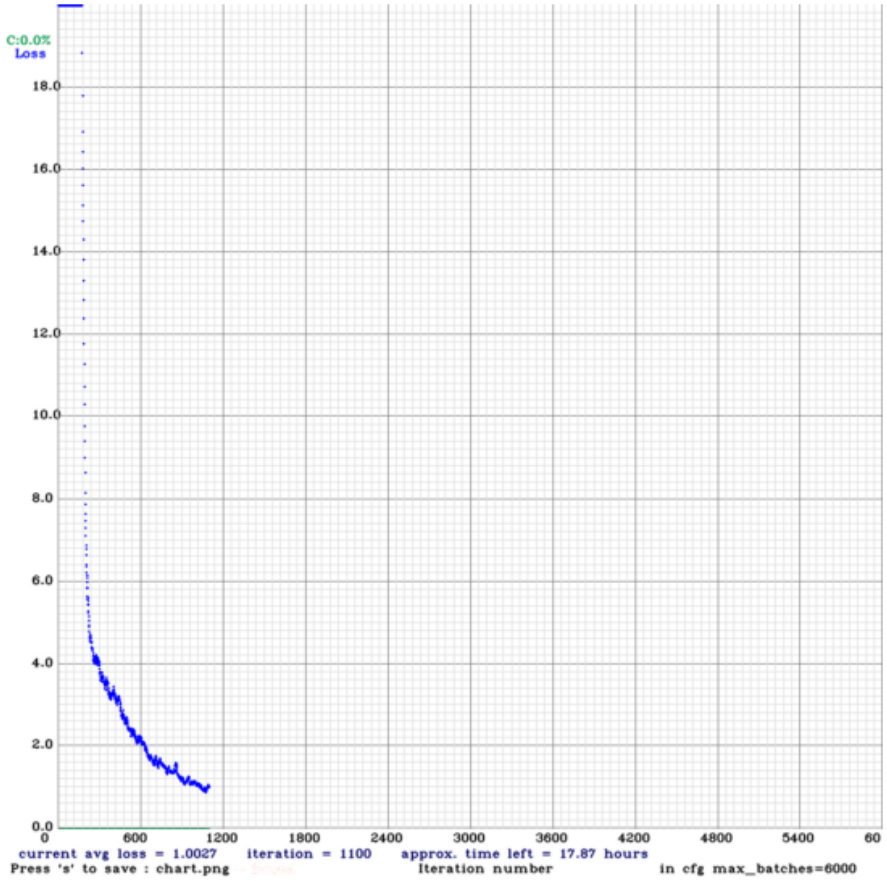




Εικόνα 80 : Στην εικόνα αριστερά πάνω το μοντέλο SSD έχει τοποθετήσει σωστά τα πλαίσια οριοθέτησης για τις κατηγορίες μήλο και μπανάνα αλλά λανθασμένα για το πορτοκάλι, στη δεξιά πάνω αναγνωρίζει επιτυχώς τα αντικείμενα και κάτω αριστερά αναγνωρίζει μόνο το μήλο

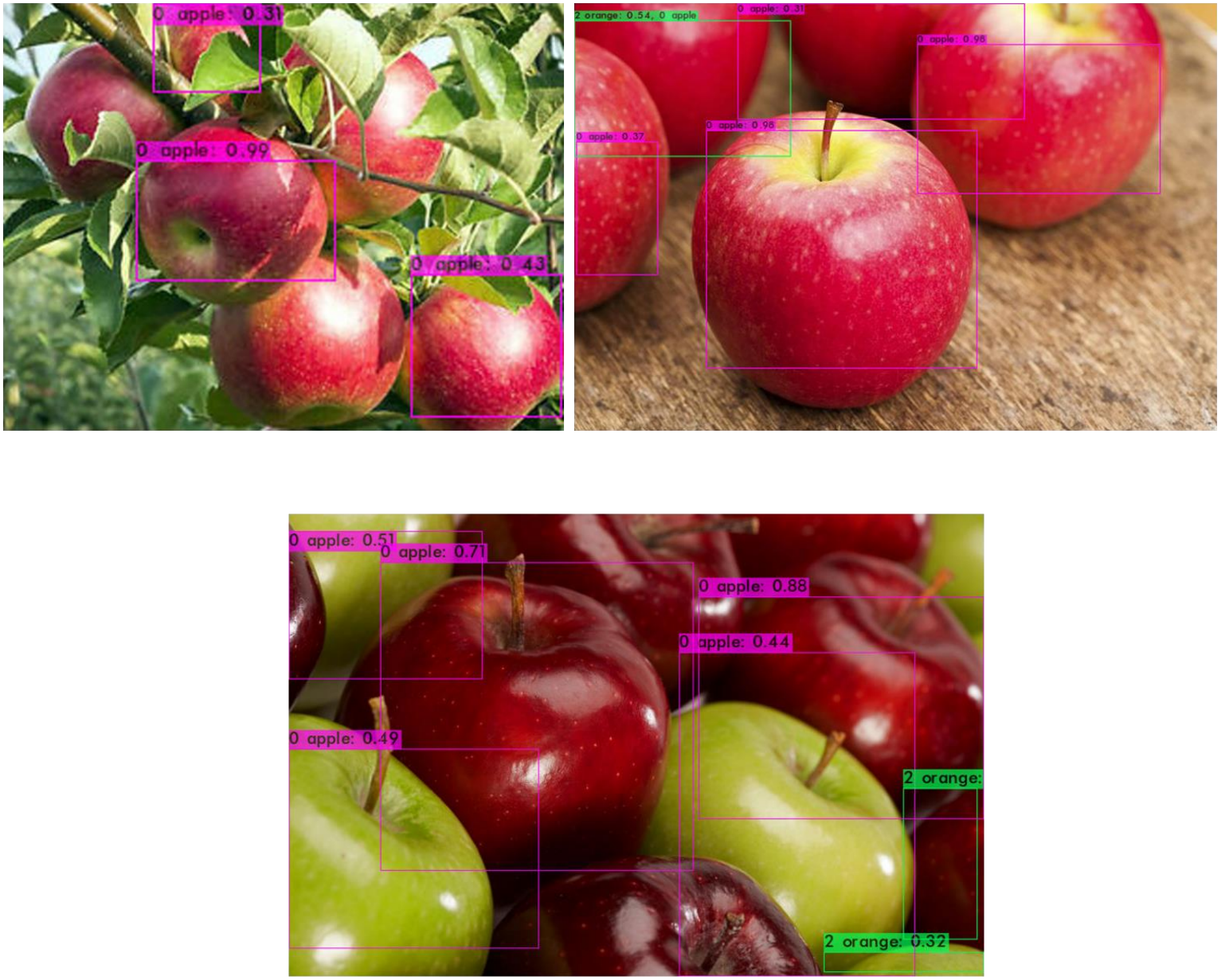
YOLOv3

Το τελευταίο μοντέλο που εκπαιδεύτηκε ήταν το YOLO v3 που εμφάνισε υψηλά ποσοστά επιτυχίας στον εντοπισμών αντικειμένων όλων των κατηγοριών καθώς και στο πλήθος των αντικειμένων για τις περισσότερες των περιπτώσεων. Παρακάτω παρουσιάζεται η συνάρτηση απώλειας κατά την εκπαίδευση του μοντέλου.

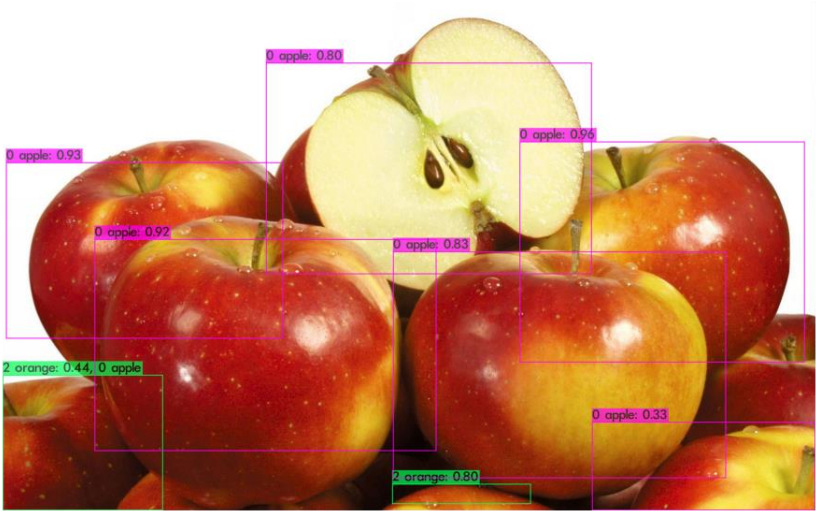


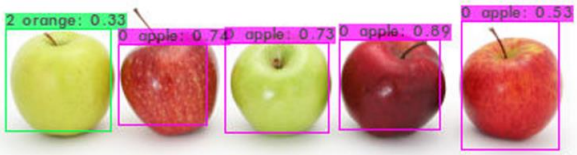
Εικόνα 81: Συνάρτηση απώλειας για το μοντέλο YOLOv3

Επίσης, παρουσιάζονται τα αποτελέσματα εντοπισμού του μοντέλου YOLOv3 στις εικόνες αξιολόγησης. Για την κατηγορία μήλο, το μοντέλο αναγνώρισε επιτυχώς σχεδόν όλα τα εικονιζόμενα αντικείμενα. Μια αστοχία που παρουσίασε το μοντέλο ήταν να μπερδεύει κάποια φορές τα μήλα με τη κατηγορία πορτοκάλι ειδικά αν αυτά δεν απεικονίζονταν ολόκληρα ή αν υπήρχαν πράσινα μήλα..



Εικόνα 82: Ανιχνεύσεις του μοντέλου YOLOv3 στην κατηγορία μήλο με μικρές αστοχίες σε μικρή περιοχή της εικόνας (κάτω)



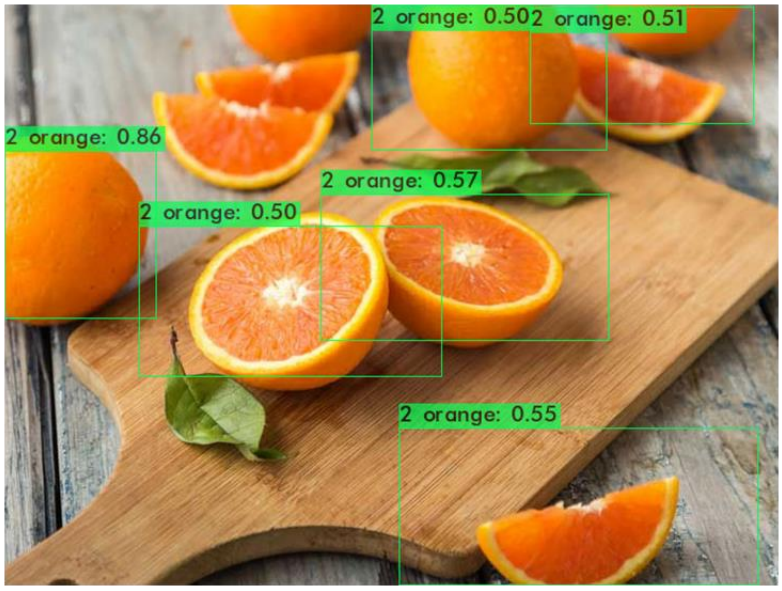


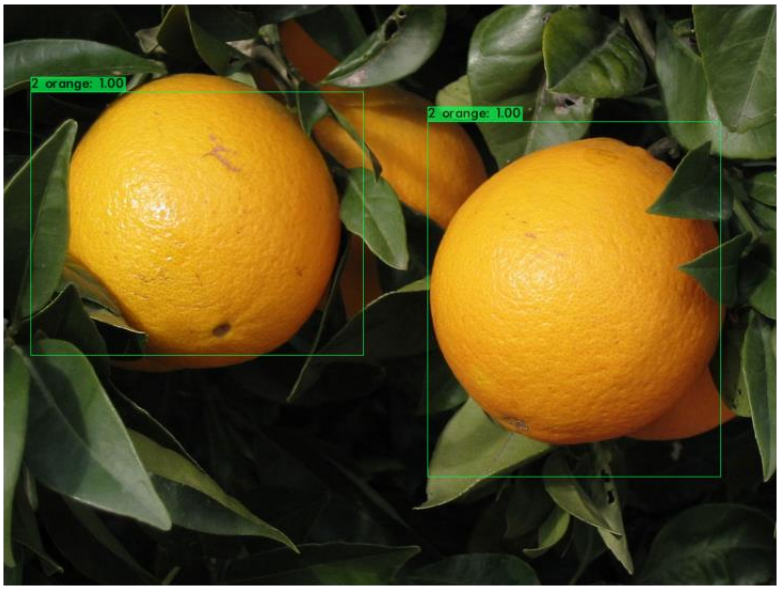
Εικόνα 83: Ανιχνεύσεις του μοντέλου YOLOv3 στην κατηγορία μήλο εμφανίζεται η αστοχία του μοντέλου σε κάποια αντικείμενα που τα κατατάσσει ως πορτοκάλια

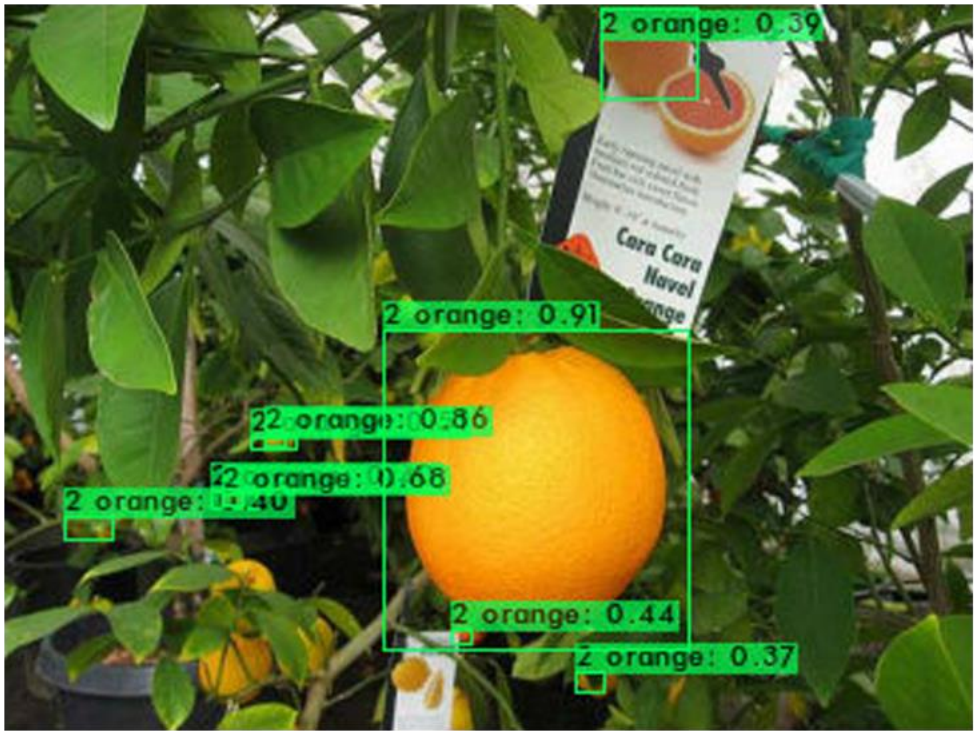
Για την κατηγορία μπανάνα υπήρχαν επιτυχείς ανιχνεύσεις στις εικόνες αξιολόγησης.

Εικόνα 84: Ανιχνεύσεις του μοντέλου YOLOv3 στην κατηγορία μπανάνα

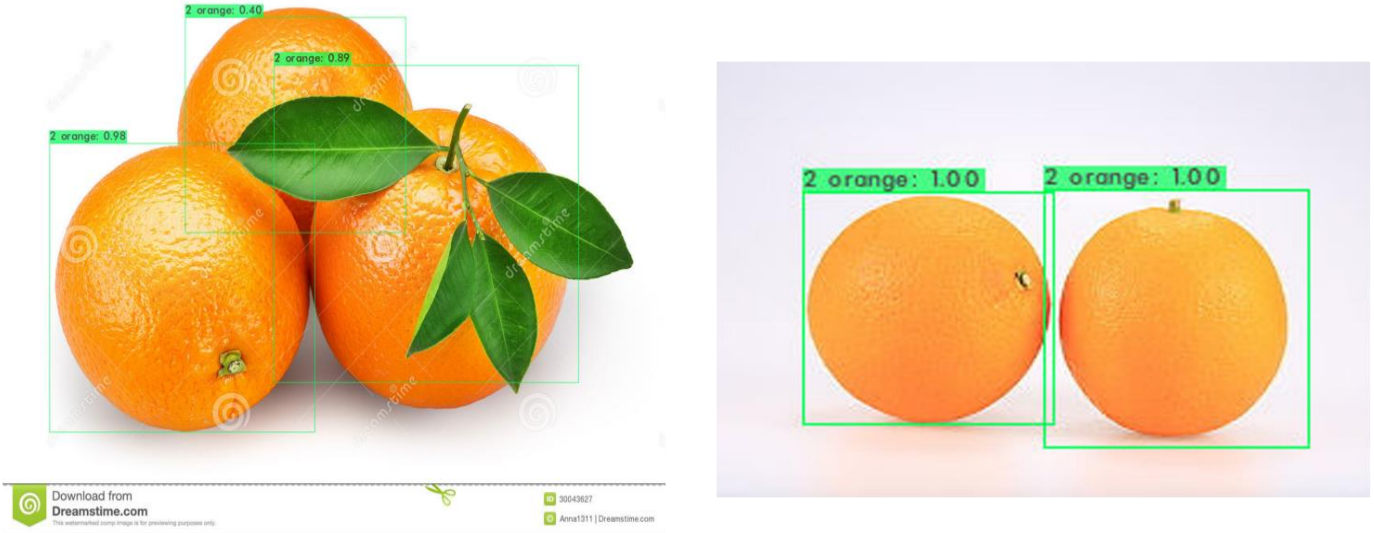
Τέλος στη κατηγορία πορτοκάλι, ανίχνευσε αρκετά αντικείμενα ακόμα και μικρού μεγέθους χωρίς απαραίτητα να είναι τοποθετημένα σε πρώτο πλάνο μέσα στην εικόνα.





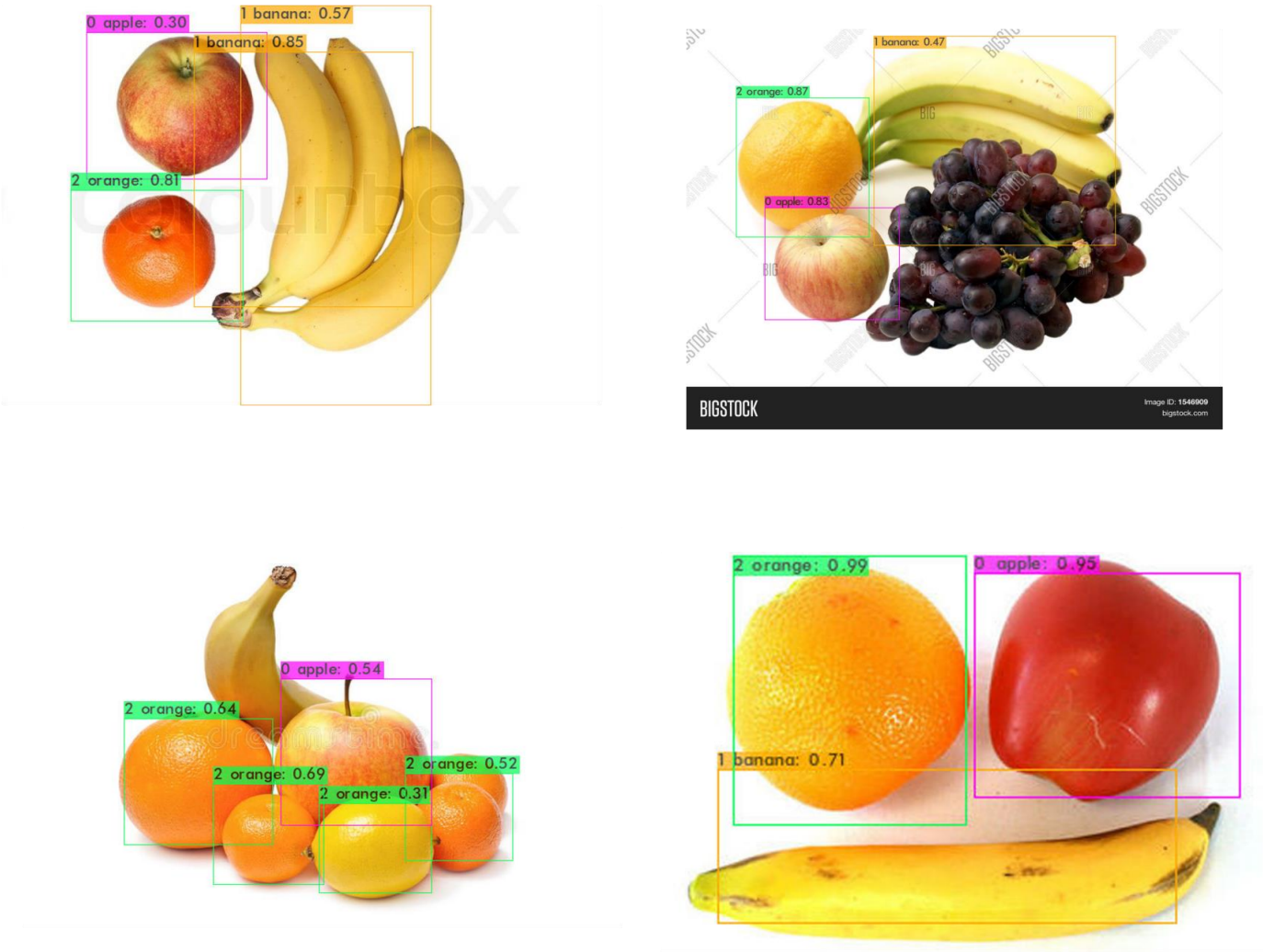


Εικόνα 85: Ανιχνεύσεις του μοντέλου YOLOv3 στην κατηγορία πορτοκάλι



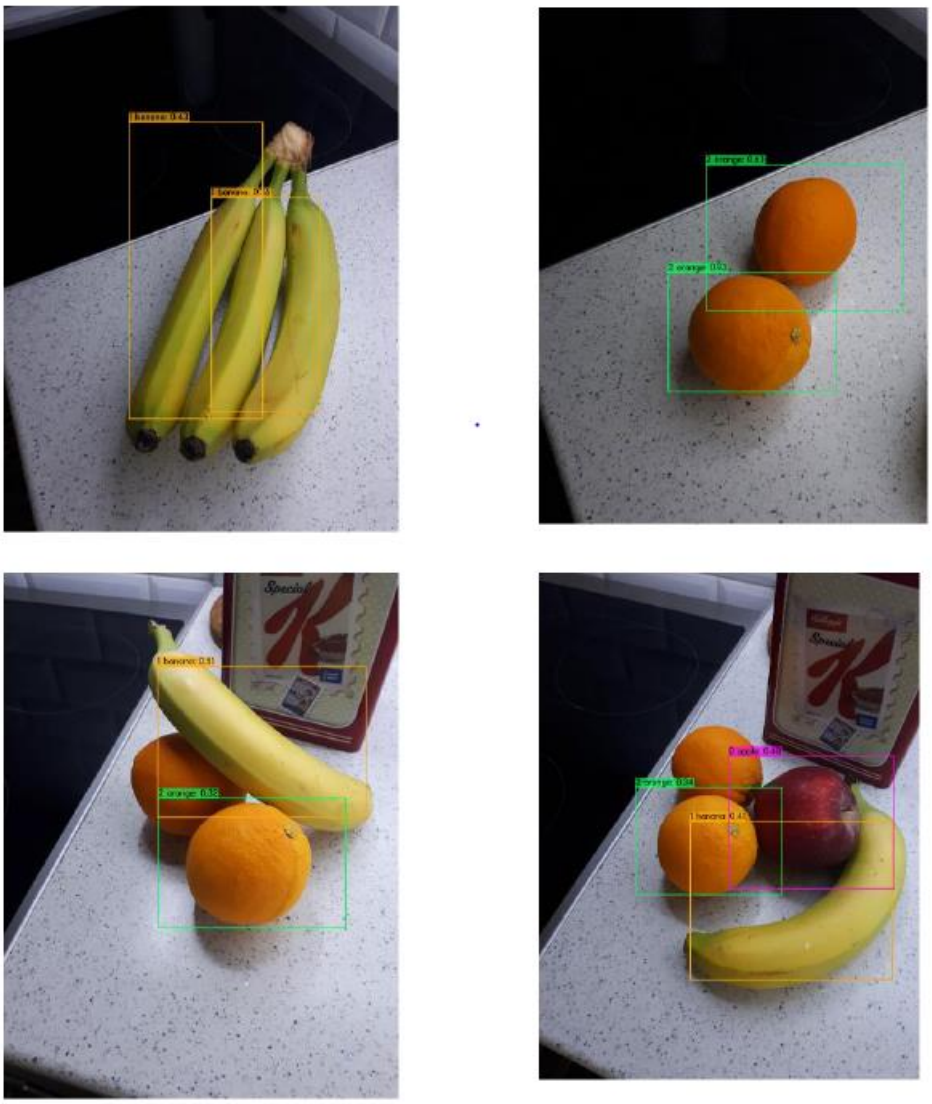
Εικόνα 86: Ανιχνεύσεις του μοντέλου YOLOv3 στην κατηγορία πορτοκάλι

Το YOLOv3 ήταν το μόνο από τα τρία μοντέλα που κατάφερε σε εικόνες με διαφορετικές κατηγορίες φρούτων να τις ανιχνεύσει όλες επιτυχώς.

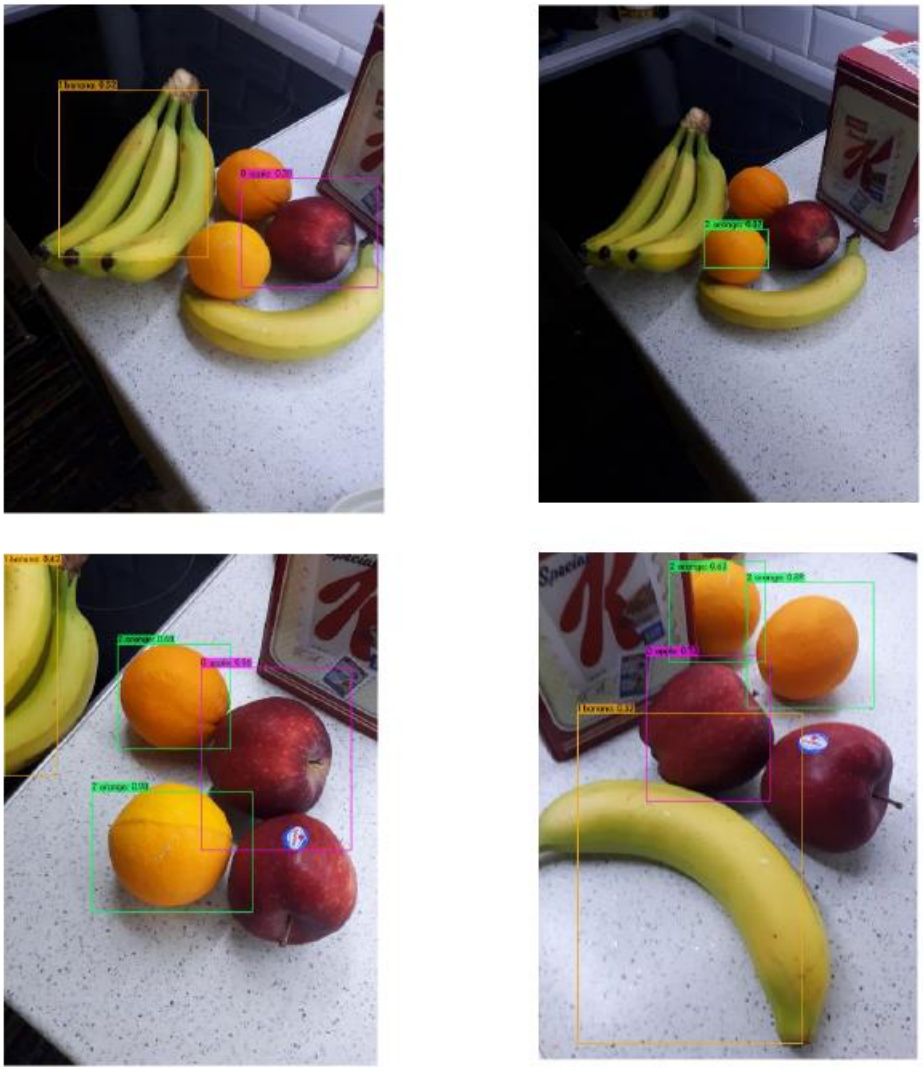


Εικόνα 87: Ανιχνεύσεις του μοντέλου YOLOv3 σε εικόνες με όλες τις κατηγορίες εκπαίδευσης

Λόγω των ικανοποιητικών αποτελεσμάτων του μοντέλου YOLO v3, αποφασίστηκε να δοκιμαστεί η γενίκευση του και σε εικόνες εκτός dataset. Αρχικά λήφθησαν εικόνες με φρούτα για πειραματισμό στο μοντέλο. επίσης πραγματοποιήθηκε δοκιμή και με εικόνες από το άρθρο των *Bargoti et al. (2017)* όπου τα δεδομένα τους κυκλοφορούν ελεύθερα για ερευνητικούς σκοπούς. Παρακάτω παρουσιάζονται τα αποτελέσματα των δοκιμών αυτών.



Εικόνα 88: Ανιχνεύσεις του μοντέλου YOLOv3 σε εικόνες εκτός dataset



Εικόνα 89: Ανιχνεύσεις του μοντέλου YOLOv3 σε εικόνες εκτός dataset

Εικόνα90: Ανιχνεύσεις του μοντέλου YOLOv3 σε εικόνες εκτός dataset (από το άρθρο Deep Fruit Detection in Orchards)

Παρατηρήσεις

Στο προηγούμενο κεφάλαιο παρουσιάστηκαν τα αποτελέσματα αυτόματου εντοπισμού φρούτων σε εικόνες μέσω εφαρμογής τριών μοντέλων (Faster R-CNN, SSD, YOLOv3). Αρχικά για τα μοντέλα Faster R-CNN και SSD έγινε προσπάθεια να βελτιωθεί η ακρίβεια εντοπισμού τους αλλάζοντας παραμέτρους όπως την παράμετρο steps και τον ρυθμό εκπαίδευσης ώστε τα βάρη να τροποποιούνται πιο αργά, χωρίς όμως κάποια ουσιαστική βελτίωση. Συμπεραίνοντας από τις καμπύλες απώλειας κατά τη διάρκεια εκπαίδευσης των μοντέλων Faster R-CNN και SSD ότι λόγω του μικρού συνόλου δεδομένων που είχαν στη διάθεση τους ως δεδομένα εκπαίδευσης, υπερπροσαρμόστηκαν (overfitting). Με επακόλουθο την μη επιτυχή εκπαίδευσης τους και τα ελλιπή αποτελέσματα τους στην ανίχνευση αντικειμένων. Οι αντικειμενικές δυνατότητες των 2 αυτών μοντέλων είναι αντίστοιχες του YOLO v3, θα μπορούσαν να βελτιωθούν είτε με μεγαλύτερο σετ δεδομένων είτε με διαφορετική παραμετροποίηση αλλά τέτοιες παρεμβάσεις δεν ήταν στα πλαίσια της παρούσας διπλωματικής εργασίας. Συμπερασματικά, το YOLO v3 έδωσε τα πιο ικανοποιητικά αποτελέσματα στην αναγνώριση αντικειμένου στα πλαίσια των πειραμάτων της εργασίας. Εδώ πρέπει να επισημανθεί ότι όλα τα μοντέλα ήταν προεκπαιδευμένα στο coco dataset αλλά τα βάρη των Faster R-CCN και SSD δεν χρησιμοποιήθηκαν κατά την εκπαίδευση τους, έγινε δηλαδή εκ νέου εκπαίδευση. Αντίθετα, όπως αναφέρθηκε στη μεθοδολογία για το YOLO v3, φορτώθηκαν κάποια έτοιμα βάρη για τα συνελικτικά του επίπεδα, πριν ξεκινήσει η εκπαίδευση στο παρόν σετ δεδομένων. Συνεπώς αυτός μπορεί να είναι και ο λόγος που εμφανίζει το YOLOv3 καλύτερα αποτελέσματα από ότι τα άλλα δύο μοντέλα

5. Συμπεράσματα και Προοπτικές

Στο κεφάλαιο αυτό περιγράφονται συμπεράσματα που προέκυψαν τόσο κατά την εκπόνηση της εργασίας όσο και από τα αποτελέσματα της ταξινόμησης και των συγκρίσεων που παρουσιάστηκαν στο Κεφάλαιο 4. Επίσης προτείνονται βελτιώσεις της εφαρμογής που αποσκοπούν στην εξέλιξη της έρευνας στο συγκεκριμένο αντικείμενο.

5.1 Συμπεράσματα

Στη παρούσα διπλωματική μελετάται το πρόβλημα της ταξινόμησης εικόνων που περιέχουν διαφορετικά είδη φρούτων χρησιμοποιώντας μοντέλα βαθιάς μηχανικής μάθησης, αλλά και ο εντοπισμός φρούτων από εικόνες. Στόχος ήταν η εμβάθυνση στο θεωρητικό πλαίσιο αυτών των τεχνικών, καθώς και η καλύτερη κατανόηση των δυνατοτήτων και των περιορισμών τους μέσω της εφαρμογής τους σε σύνολα δεδομένων. Από την εκτέλεση πειραμάτων στο τμήμα της ταξινόμησης, αναδείχθηκε ο σημαντικός ρόλος των τεχνικών transfer learning και fine tuning. Οι τεχνικές αυτές διευκολύνουν τη διαδικασία της εκπαίδευσης που είναι ιδιαίτερα χρονοβόρα, ενώ ταυτόχρονα αυξάνουν την ακρίβεια των μοντέλων ακόμα και με μικρό χρόνο εκπαίδευσης ή με περιορισμένα δεδομένα εκπαίδευσης. Η τεχνική της επαύξησης δεδομένων (data augmentation) φάνηκε ιδιαίτερα χρήσιμη για την δημιουργία ενός ολοκληρωμένου dataset. Από τα μοντέλα που χρησιμοποιήθηκαν, όλα εμφάνισαν ιδιαίτερα υψηλά ποσοστά ακρίβειας (>90%) στην ταξινόμηση εικόνας. Τα μοντέλα Deep Learning ήταν ιδιαίτερα αποδοτικά στην συγκεκριμένη εφαρμογή τους, παρά του περιορισμένου πλήθους των εικόνων.
Το συγκεκριμένο συμπέρασμα συνδέεται και με τη μέχρι τώρα εφαρμογή των αρχιτεκτονικών deep learning στη διεθνή βιβλιογραφία όπου έχουν αναπτυχθεί και εφαρμοστεί με επιτυχία σε πολλές πρακτικές εφαρμογές (π.χ. εντοπισμός οχημάτων, ανθρώπων κ.λπ.). Το Resnet-50 ξεχώρισε λόγω του σταθερά υψηλού ποσοστού σε συνδυασμό με τη μικρή τιμή απώλειας που εμφάνισε. Η κλάση που συγχέεται πιο συχνά σε όλα τα τα μοντέλα ήταν η κλάση mixed.
Στο τμήμα της αναγνώρισης αντικειμένου παρατηρήθηκε ότι τα μοντέλα Faster R-CCN και SSD υπερπροσαρμόστηκαν στα δεδομένα εκπαίδευσης, γεγονός που κάνει εμφανή την ανάγκη για μεγαλύτερο dataset. Το μοντέλο YOLO v3 εμφάνισε τα μεγαλύτερα ποσοστά ακρίβειας στον εντοπισμό των φρούτων από τις εικόνες.
Στο τμήμα της αναγνώρισης αντικειμένου παρατηρήθηκε ότι τα μοντέλα Faster R-CCN και SSD υπερπροσαρμόστηκαν στα δεδομένα εκπαίδευσης άρα θα ήταν σημαντικό το dataset να ήταν μεγαλύτερο. Το μοντέλο YOLO v3 εμφάνισε τα μεγαλύτερα ποσοστά ακρίβειας στον εντοπισμό των φρούτων από τις εικόνες.

5.2 Προοπτικές

Οι τεχνικές deep learning για ταξινόμηση και ανίχνευση αντικείμενων από εικόνες είναι ιδιαίτερα αποδοτικές και τα αποτελέσματα τους βελτιώνονται κάθε χρόνο όλο και περισσότερο. Για το λόγο αυτό, η περαιτέρω ενασχόληση με το συγκεκριμένο πεδίο έρευνας θα ήταν ιδιαίτερα ελπιδοφόρα, αφού εξακολουθούν να υπάρχουν αρκετά περιθώρια βελτίωσης των μοντέλων. Βάσει των συγκεκριμένων πειραμάτων που αναπτύχθηκαν και της βιβλιογραφικής μελέτης που διεξήχθει σαν μελλοντική έρευνα προτείνονται τα παρακάτω :

- Το σετ δεδομένων που χρησιμοποιείται ως δεδομένα εκπαίδευσης και αξιολόγησης του μοντέλου πρέπει να περιέχει μεγάλο αριθμό εικόνων ώστε το μοντέλο να αναγνωρίζει με επιτυχία τα ζητούμενα αντικείμενα και να καταφέρει να γενικευτεί.
- Οι κατηγορίες αντικειμένων που ορίζονται στο μοντέλο προς αναγνώριση, πρέπει να είναι διακριτές και να αντιπροσωπεύονται με επάρκεια στα δεδομένα εκπαίδευσης.
- Οι εικόνες του σετ δεδομένων πρέπει να παρουσιάζουν τα φρούτα σε περιβάλλον χωραφιού ώστε να προσεγγίζει η εφαρμογή μια πραγματική ανάγκη που υπάρχει για τον αυτοματισμό της γεωργίας σε συνθήκες χωραφιού.

- Ο συνδυασμός των CNNs με φωτογραμμετρικές τεχνικές όπως sfm συνίσταται για ακριβέστερο εντοπισμό των φρούτων και μοντελοποίησης του χωραφιού.
- Στη παρούσα εργασία η ταξινόμηση αφορά την ύπαρξη ή όχι των εξεταζόμενων κατηγοριών σε επίπεδο εικόνας. Στην αναγνώριση αντικειμένου η κλίμακα μικραίνει και εξετάζονται περιοχές με pixels. Ως επόμενο βήμα θα είχει ενδιαφέρον να εξεταστεί η κατάτμηση σε επίπεδο pixel με τεχνικές deep learning.
- Ερευνητικό ενδιαφέρον θα εμφάνιζε η ανάπτυξη cnn μοντέλου για πολλαπλές ανιχνεύσεις και ταξινόμηση διαφόρων ειδών φρούτων ταυτόχρονα.
- Σχεδόν όλα τα έργα ελέγχου ποιότητας πραγματοποιούνται υπό εργαστηριακές συνθήκες χρησιμοποιώντας αισθητήρες που δεν είναι έτοιμοι για πραγματικές συνθήκες χωραφιού. Άρα προτείνεται η δοκιμή εικόνων με φρούτα ως δεδομένα αναφοράς για έλεγχο ποιότητας.

Βιβλιογραφία

1) Andreas Kamilaris, Francesc X. Prenafeta-Boldú, (2018) Deep learning in agriculture: A survey, Computers and Electronics in Agriculture (Volume 147,2018, Pages 70-90)
Διαθέσιμο σε: (https://www.sciencedirect.com/science/article/pii/S0168169917308803)

2) Naranjo-Torres J., Mora M., Hernández-García R., Barrientos R.J., Fredes C., Valenzuela A. ,(2020) A Review of Convolutional Neural Network Applied to Fruit Image Processing, Appl. Sci. 2020, 10, 3443.
Διαθέσιμο σε: (https://doi.org/10.3390/app10103443)

3) X. Liu and Steven W. Chen and Shreyas Aditya and Nivedha Sivakumar and Sandeep Dcunha and Chao Qu and Camillo Jose Taylor and Jnaneshwar Das and Vijay R. Kumar, (2018) Robust Fruit Counting: Combining Deep Learning, Tracking, and Structure from Motion, journal article: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
Διαθέσιμο σε: (https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8594239&tag=1)

4) Chen, Steven W. and Shivakumar, Shreyas S. and Dcunha, Sandeep and Das, Jnaneshwar and Okon, Edidiong and Qu, Chao and Taylor, Camillo J. and Kumar, Vijay,(2017) "Counting Apples and Oranges With Deep Learning: A Data-Driven Approach," in IEEE Robotics and Automation Letters, vol. 2, no. 2, pp. 781-788
Διαθέσιμο σε: (https://ieeexplore.ieee.org/document/7814145)

5) Suchet Bargoti and James Patrick Underwood, (2017) Deep fruit detection in orchards, journal in IEEE International Conference on Robotics and Automation (ICRA),(Pages 3623-3633)
Διαθέσιμο σε: (https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7989417)

6) Koirala, A., Walsh, K.B., Wang, Z. et al. Deep learning for real-time fruit detection and orchard fruit load estimation: benchmarking of 'MangoYOLO'. Precision Agric 20, 1107–1135 (2019)
Διαθέσιμο σε: (https://doi.org/10.1007/s11119-019-09642-0)

7) Santos Luís, Santos Filipe N., Oliveira Paulo Moura, Shinde Pranjali, (2020) Deep Learning Applications in Agriculture: A Short Review, Robot 2019: Fourth Iberian Robotics Conference, (Pages 139- 151)
Διαθέσιμο σε: (https://link.springer.com/chapter/10.1007/978-3-030-35990-4_12)

8) LeCun Y., Bottou L., Bengio Y., Haffner P., Gradient-based learning applied to document recognition. (1998) Proc. IEEE 1998, 86, 2278–2324.
Διαθέσιμο σε: (https://ieeexplore.ieee.org/document/72679112)

9) Krizhevsky A., Sutskever I., Hinton G.E., (2012) ImageNet Classification with Deep Convolutional Neural Networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems—Volume 1, NIPS'12, Curran Associates Inc.: Red Hook, NY, USA, 2012; pp. 1097–1105.
Διαθέσιμο σε: (https://papers.nips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf)

10) LeCun Y., Kavukcuoglu K., Farabet C., Convolutional networks and applications in vision. In Proceedings of the IEEE 2010 IEEE international symposium on circuits and systems, Paris, France, pp. 253–256.
Διαθέσιμο σε: (https://koray.kavukcuoglu.org/publis/lecun-iscas-10.pdf)

11) Yudong Zhang , Zhengchao Dong ,Xianqing Chen ,Wenjuan Jia , Sidan Du , Khan Muhammad , Shuihua Wang, (2017), Image based fruit category classification by 13-layer deep convolutional neural network and data augmentation, Multimedia Tools and Applications vol. 78,pp. 3613-3632
Διαθέσιμο σε: (https://link.springer.com/article/10.1007/s11042-017-5243-3)

12) Jordi Gené-Mola, Ricardo Sanz-Cortiella, Joan R. Rosell-Polo, Josep-Ramon Morros, Javier Ruiz-Hidalgo, Verónica Vilaplana, Eduard Gregorio, (2020), Fruit detection and 3D location using instance segmentation neural networks and structure-from-motion photogrammetry, article: Computers and Electronics in Agriculture, Volume 169
Διαθέσιμο σε: (https://doi.org/10.1016/j.compag.2019.105165)

13) Lixuan Du, Rongyu Zhang, Xiaotian Wang (2020), Overview of two-stage object detection algorithms, J. Phys.: Conf. Ser. 1544 012033
Διαθέσιμο σε: (https://iopscience.iop.org/article/10.1088/1742-6596/1544/1/012033/pdf)

14) Jordi Gené-Mola, Ricardo Sanz-Cortiella, Joan R. Rosell-Polo, Josep-Ramon Morros, Javier Ruiz-Hidalgo, Verónica Vilaplana, Eduard Gregorio,(2020),Fruit detection and 3D location using instance segmentation neural networks and structure-from-motion photogrammetry,Computers and Electronics in Agriculture, Volume 169
Διαθέσιμο σε: (https://www.sciencedirect.com/science/article/pii/S0168169919321507)

15) Jordi Gené-Mola, Ricardo Sanz-Cortiella, Joan R. Rosell-Polo, Josep-Ramon Morros, Javier Ruiz-Hidalgo, Verónica Vilaplana, Eduard Gregorio,(2020), Fuji-SfM dataset: A collection of annotated images and point clouds for Fuji apple detection and location using structure-from-motion photogrammetry, Data in Brief, Volume 30,
Διαθέσιμο σε: (https://www.sciencedirect.com/science/article/pii/S2352340920304856)

Ηλεκτρονικές Πηγές

16) https://machinelearningmastery.com

17) https://www.jeremyjordan.me/object-detection-one-stage

18) https://www.pyimagesearch.com/2016/11/07/intersection-over-union-iou-for-object-detection/

19) https://hamxam1.medium.com/custom-data-tensorflow-object-detection-api-603baccc416

20) https://www.kaggle.com/mbkinaci/fruit-images-for-object-detection

21) https://github.com/tensorflow/models/blob/master/research/object_detection/g3doc/tf2_detection_zoo.md

22) https://github.com/tensorflow/models

23) https://tensorflow-object-detection-api-tutorial.readthedocs.io/en/latest/install.html

24) https://www.tensorflow.org/hub/tutorials/object_detection

25) https://www.tensorflow.org/hub/tutorials/tf2_object_detection

26) https://blog.roboflow.com/train-a-tensorflow2-object-detection-model

27) https://github.com/aieml/tensorflow-object-detection-api-configuration/blob/master/1.0%20Object%20Detection%20Tutorial.ipynb