ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΑΓΡΟΝΟΜΩΝ ΚΑΙ ΤΟΠΟΓΡΑΦΩΝ ΜΗΧΑΝΙΚΩΝ - ΜΗΧΑΝΙΚΩΝ ΓΕΩΠΛΗΡΟΦΟΡΙΚΗΣ
ΤΟΜΕΑΣ ΤΟΠΟΓΑΦΙΑΣ – ΕΡΓΑΣΤΗΡΙΟ ΦΩΤΟΓΡΑΜΜΕΤΡΙΑΣ

Εφαρμογή Τεχνικών Ανίχνευσης Περιοχών Ενδιαφέροντος σε Ιατρικές Εικόνες

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

ΜΠΟΛΛΑΝΟ ΟΡΕΣΤΗ

Επιβλέπων : Αναστάσιος Δουλάμης, Καθηγητής Ε.Μ.Π.
Σταύρος Συκιώτης, Διδακτορικός

Αθήνα, Φεβρουάριος 2022

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΑΓΡΟΝΟΜΩΝ ΚΑΙ ΤΟΠΟΓΡΑΦΩΝ ΜΗΧΑΝΙΚΩΝ - ΜΗΧΑΝΙΚΩΝ ΓΕΩΠΛΗΡΟΦΟΡΙΚΗΣ
ΤΟΜΕΑΣ ΤΟΠΟΓΑΦΙΑΣ – ΕΡΓΑΣΤΗΡΙΟ ΦΩΤΟΓΡΑΜΜΕΤΡΙΑΣ

Εφαρμογή Τεχνικών Ανίχνευσης Περιοχών Ενδιαφέροντος σε Ιατρικές Εικόνες

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

ΜΠΟΛΛΑΝΟ ΟΡΕΣΤΗ

Επιβλέπων : Αναστάσιος Δουλάμης, Καθηγητής Ε.Μ.Π.
Σταύρος Συκιώτης, Διδακτορικός

Αθήνα, Φεβρουάριος 2022

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΑΓΡΟΝΟΜΩΝ ΚΑΙ ΤΟΠΟΓΡΑΦΩΝ ΜΗΧΑΝΙΚΩΝ - ΜΗΧΑΝΙΚΩΝ ΓΕΩΠΛΗΡΟΦΟΡΙΚΗΣ
ΤΟΜΕΑΣ ΤΟΠΟΓΡΑΦΙΑΣ – ΕΡΓΑΣΤΗΡΙΟ ΦΩΤΟΓΡΑΜΜΕΤΡΙΑΣ

Εφαρμογή Τεχνικών Ανίχνευσης Περιοχών Ενδιαφέροντος σε Ιατρικές Εικόνες

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

ΜΠΟΛΛΑΝΟ ΟΡΕΣΤΗ

Επιβλέπων : Αναστάσιος Δουλάμης, Καθηγητής ΕΜΠ
Σταύρος Συκιώτης, Διδακτορικός

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή:

(Υπογραφή)	*(Υπογραφή)*	*(Υπογραφή)*
.....	.....	.....
Αναστάσιος Δουλάμης Καθηγητής Ε.Μ.Π.	Νικόλαος Δουλάμης Καθηγητής Ε.Μ.Π.	Βασίλειος Βεσκούκης Καθηγητής Ε.Μ.Π.

Αθήνα Φεβρουάριος 2022

(Υπογραφή)

.....
ΟΡΕΣΤΗΣ ΜΠΟΛΛΑΝΟ

Διπλωματούχος Αγρονόμος και Τοπογράφος Μηχανικός – Μηχανικός Γεωπληροφορικής Ε.Μ.Π.

ΕΥΧΑΡΙΣΤΙΕΣ

Με την αφορμή που μου δίνεται, θα ήθελα να ευχαριστήσω: Τον Καθηγητή κ. Αναστάσιο Δουλάμη για τη διδασκαλία των μαθημάτων (Προγραμματιστικές Τεχνικές, Φωτογραμμετρία ΙΙΙ, Στοιχεία Επεξεργασίας Σημάτων) που έπαιξαν καθοριστικό ρόλο στην ανάπτυξη των ενδιαφερόντων μου και στην απόφαση να ασχοληθώ με το συγκεκριμένο αντικείμενο, καθώς και για την ευκαιρία που μου έδωσε να εκπονήσω τη διπλωματική μου εργασία στο εργαστήριό του.

Τον διδακτορικό φοιτητή Σταύρο Συκιώτη, για την πολύ καλή συνεργασία που είχαμε όλη αυτήν την περίοδο. Η καθοδήγηση, οι συμβουλές, οι προτάσεις και οι διορθώσεις του ήταν πολύτιμες τόσο στην διεκπεραίωση της έρευνας και στη διαμόρφωση της εργασίας όσο και στην αντιμετώπιση των όποιων προβλημάτων, αποριών ή εμποδίων δημιουργούνταν.

Τους γονείς μου και τον αδερφό μου, για την αμέριστη στήριξη που μου προσφέρουν, όλα αυτά τα χρόνια, την φίλους μου που είναι σταθερά δίπλα μου, στις ευχάριστες και δυσάρεστες στιγμές μου, και ειδικά στην Ελίζα Κοντού και την Κυριακή Κεχαγιά που αυτά τα 5 χρόνια της σχολής χωρίς την βοήθεια τους και την ανιδιοτελή υποστήριξη τους όλα αυτά θα ήταν ένα μακρινό όνειρο.

Ορέστης Μπολλάνο
Φεβρουάριος 2022

ΠΕΡΙΕΧΟΜΕΝΑ

ΕΥΧΑΡΙΣΤΙΕΣ 1
ΠΕΡΙΕΧΟΜΕΝΑ 3
ΚΑΤΑΛΟΓΟΣ ΣΥΝΤΟΜΟΓΡΑΦΙΩΝ 5
PREFACE 7
ΑΝΤΙ ΠΡΟΛΟΓΟΥ 9
ABSTRACT 11
ΠΕΡΙΛΗΨΗ 13
1. Ανίχνευση Αντικειμένων 15
1.1. Εισαγωγή 16
1.2. Επεξήγηση τεχνικής Ανίχνευσης Αντικειμένων 17
1.2.1. Η Ανίχνευσης Αντικειμένων ως το βασικό βήμα για αναγνώριση εικόνων 17
1.2.2. Ανίχνευση αντικειμένων με χρήση Συνελεκτικών Νευρωνικών Δικτύων 17
1.2.3. Πλαίσια και υπηρεσίες για Ανίχνευση Αντικειμένων 19
1.3. Περιοχές Εφαρμογής 20
1.4. Υπερσύγχρονες προσεγγίσεις βαθιάς μάθησης για Ανίχνευση Αντικειμένων 21

2. Αλγόριθμοι Ανίχνευσης Αντικειμένων 25
2.1. Εισαγωγή 26
2.2. Αλγόριθμοι 26
2.2.1. Histogram of Oriented Gradients (HOG) 26
2.2.2. RetinaNET 28
2.2.3. Single Shot Detector (SSD) 29
2.2.4. YOLO (You Only Look Once) 31
2.2.5. Region-based Convolutional Neural Networks (R-CNN) 32

3. Αλγόριθμος Faster R-CNN 35
3.1. Εισαγωγή 36
3.2. Faster R-CNN 36
3.3. Δίκτυο Προτάσεων Περιοχής 37
3.3.1. Άγκυρες 38
3.3.2. Συνάρτηση Απώλειας 40
3.3.3. Εκπαίδευση Δικτύου Προτάσεων Περιοχής 42
3.4. Κοινά χαρακτηριστικά RPN & Fast R-CNN 42
3.5. Πείραμα Faster R-CNN 44

4. Εφαρμογές Ανίχνευσης Αντικειμένων σε Ιατρικές Εικόνες 49
4.1. Εισαγωγή 50
4.2. Ανίχνευση Αντικειμένων σε Ιατρικές Εικόνες 52
4.3. Ανίχνευση Αντικειμένων σε Αξονικές Τομογραφίες (CT) 55

5. Πειραματική διαδικασία 57
5.1. Περιγραφή δεδομένων 58
5.2. Παράμετροι Εκπαίδευσης 60
5.3. Αξιολόγηση αποτελεσμάτων 63
5.4. Αποτελέσματα και συγκρίσεις 67

6. Συμπεράσματα και Μελλοντικά Σχέδια 69
6.1. Συμπεράσματα 70

Κατάλογος Σχημάτων 72
Κατάλογος Πινάκων 73
Βιβλιογραφία 75

ΚΑΤΑΛΟΓΟΣ ΣΥΝΤΟΜΟΓΡΑΦΙΩΝ

CNN – Convolutional Neural Network – Συνελεκτικά Νευρωνικά Δίκτυα
FCN – Fully Convolutional Network – Ολοκληρωμένα Συνελεκτικά Δίκτυα
CT – Computed Tomography – Αξονική Τομογραφία
HOG – Histogram of Oriented Gradients – Ιστόγραμμα Προσανατολισμένων Διαβαθμίσεων
SSD – Single Shot Detector – Ανιχνευτής μιας Λήψης
YOLO – You Only Look Once
RPN – Region Proposal Network – Δίκτυο Προτάσεων Περιοχής
R-CNN – Region-based Convolutional Neural Network – Συνελεκτικό Νευρωνικό Δίκτυο με βάση την περιοχή
VRS – Visual Recognition System – Σύστημα οπτικής αναγνώρισης
MRI – Magnetic Resonance Imaging – Μαγνητική Τομογραφία
DBN – Deep Belief Network – Δίκτυο Βαθιάς Πίστης
AUC – Area Under the Curve – Περιοχή κάτω απτήν καμπύλη
mAP – mean Average Precision – Μέση Μέσης Ακρίβειας
ROC – Receiver Operator Characteristic – Χαρακτηριστικά Χειριστη Δέκτη
RNN – Recurrent Neural Network – Επαναλαμβανόμενο Νευρωνικό Δίκτυο
RBM – Restricted Boltzmann Machine – Περιορισμένη Μηχανή Boltzmann
SVM – Support-Vector Machine – Μηχανή υποστήριξης διανύσματος
SIFT – Scale-Invariant Feature Transform – Μετασχηματισμός αναλλοίωτων χαρακτηριστικών κλίμακας
FPN – Feature Pyramid Network – Δίκτυο χαρακτηριστικών πυραμίδας
IOU – Intersection over Union – Διατομή πάνω απ'την Ένωση
AR – Average Recall – Μέση Ανάκληση
CPU – Central Processing Unit – Κεντρική Μονάδα Επεξεργασίας
GPU – Graphics Processing Unit – Μονάδα Επεξεργασίας Γραφικών
TP – True Positive – Αληθές Θετικό
TN – True Negative – Αληθές Αρνητικό
FP – False Positive – Ψευδές Θετικό
FN – False Negative – Ψευδές Αρνητικό
VGG – Visual Geometry Group – Ομάδα Οπτικής Γεωμετρίας
CLS – Classification – Ταξινόμηση
REG – Regression – Παλινδρόμηση
ROI – Region of Interest – Περιοχή Ενδιαφέροντος
SGD – Stochastic Gradient Descent – Στοχαστική Μέθοδος Κλίσης
EB – Edge Boxes – Πλαίσια Ακμών
NMS – Network Management System – Σύστημα Διαχείρισης Δικτύου
PET – Positron Emission Tomography – Τομογραφίας εκπομπής ποζιτρονίων
WLE – White Light Endoscopy – Ενδοσκόπηση Λευκού Φωτός
CE – Chromo Endoscopy – Χρωμοενδοσκόπηση
ML – Machine Learning – Μηχανική Μάθηση
DL – Deep Learning – Βαθιά Μάθηση
VSM – Vector Space Model – Μοντέλο Διανυσματικού Χώρου
DICOM - Digital Imaging and Communications in Medicine – Ψηφιακή απεικόνιση και επικοινωνίες στην Ιατρική

PREFACE

The subject of this thesis is the application of region of interest detection techniques in medical digital images. The thesis was implemented with the aim of analyzing and developing the technology of analyzing Medical Digital Images for faster and more reliable results.

The thesis aims to establish the general use of the system, the applications it can address, with its respective performance and its comparison with other methods. The paper is summarized by the following chapters:

- **The first chapter** introduces Object Detection, explains its use and how it is implemented and in which areas it is established. At the end of the chapter an approach is taken to the influence of deep learning in the field of Object Detection.
- **The second chapter** separates the different Object Detection implementation algorithms, gives a brief explanation of the most widely used models and a quick review of the advantages and disadvantages of each model.
- **Chapter three** provides an extensive analysis of the Faster R-CNN algorithm and explains why it was chosen for this thesis.
- **Chapter four** discusses the use of Object Detection in Medical images. An analysis of how the industry has been affected by this technique and examples of the method are given for different parts of the human body. Finally, the data used for this thesis is presented and more specifically the processing on CT scans for the purpose of tumor detection.
- **Chapter five** presents in detail the experimental procedure followed to obtain the results. The parameters that were implemented for the proper conduct of the procedure are analyzed. Finally, a comparison is made between the different versions used for the analysis of the model

Although the process of analyzing and using the Faster R-CNN model is relatively complex and requires programming skills, many tutorials are presented on the Internet that enable even a simple user to train such models and then use their own images for personal use.

System training is constantly monitored as there is the possibility of overfitting in which a system is completely applied to the training data and does not perform correctly on unknown implementation sources. To monitor and correct such random errors, loss functions are used which are functions that allocate an event or values of one or more variables to a real number that intuitively represents an "error" associated with the event, which in this case is overfitting.

The selection of the appropriate number of iterations in model training, the number of images entered into the train, test, and validation sets, and the training time are all variables that contribute to the quality of the result and the efficiency of the model analysis.

ΑΝΤΙ ΠΡΟΛΟΓΟΥ

Αντικείμενο της παρούσας διπλωματικής εργασίας είναι η εφαρμογή τεχνικών ανίχνευσης περιοχών ενδιαφέροντος σε Ιατρικές ψηφιακές εικόνες. Η διπλωματική εργασία υλοποιήθηκε με σκόπό την ανάλυση και την εξέλιξη της τεχνολογίας ανάλυσης των Ιατρικών ψηφιακών εικόνων για πιο γρήγορα και αξιόπιστα αποτελέσματα.

Η εργασία αποσκοπεί στη διαπίστωση της γενικότερης χρήσης του συστήματος, τις εφαρμογές που δύναται να αντιμετωπίσει, με τις αντίστοιχες επιδόσεις του και τη σύγκρισή του με άλλες μεθόδους. Η εργασία συνοψίζεται από τα εξής κεφάλαια:

- **Στο πρώτο κεφάλαιο** γίνεται εισαγωγή στην Ανίχνευση Αντικειμένων, επεξηγείται η χρήση του και με ποιούς τρόπους υλοποιείται και σε ποιους τομείς έχει εδραιωθεί. Στο τέλος του κεφαλαίου γίνεται μια προσέγγιση για την επιρροή της βαθιάς μάθησης στον τομέα της Ανίχνευση Αντικειμένων.
- **Στο δεύτερο κεφάλαιο** διαχωρίζονται οι διαφορετικοί αλγόριθοι υλοποίησης Ανίχνευση Αντικειμένων, γίνεται μια σύντομη επεξήγηση για τα πιο διαδεδομένα μοντέλα και μια γρήγορη ανασκόπηση στα πλεονεκτήματα και μειονεκτήματα του κάθε μοντέλου.
- **Στο τρίτο κεφάλαιο** γίνεται εκτενής ανάλυση του αλγορίθμου Faster R-CNN και επεξηγείται ο λόγος για τον οποίο επιλέχθηκε για την εκπόνηση αυτής της διπλωματικής εργασίας.
- **Στο τέταρτο κεφάλαιο** γίνεται αναφορά στην χρήση της Ανίχνευση Αντικειμένων στις Ιατρικές εικόνες. Γίνεται ανάλυση του πως έχει επηρεάστει ο κλάδος από την συγκεκριμένη τεχνική και παρουσιάζονται παραδείγματα της μεθόδου και σε διαφορετικά πλαίσια του ανθρώπινου σώματος. Τέλος, παρουσιάζονται τα δεδομένα που χρησιμοποιήθηκαν για την παρούσα εργασία και πιο συγκεκριμένα η επεξεργασία σε αξονικές τομογραφίες με σκοπό την αναγνώριση όγκου.
- **Στο πέμπτο κεφάλαιο** παρουσιάζεται αναλυτικά η πειραματική διαδικασία που ακολουθήθηκε για την εξαγωγή των αποτελεσμάτων. Αναλύονται οι παράμετροι που υλοποιήθηκαν για την σωστή διεξαγωγή της διαδικασίας. Τέλος, γίνεται σύγκριση μεταξύ των διαφορετικών εκδοχών που χρησιμοποιήθηκαν για την ανάλυση του μοντέλου

Παρόλο που η διαδικασία ανάλυσης και χρήσης του μοντέλου Faster R-CNN είναι σχετικά πολύπλοκη και απαιτούνται προγραμματιστικές δεξιότητες, στο διαδίκτυο παρουσιάζονται πολλά μαθήματα που δίνουν την δυνατότητα και σε έναν απλό χρήστη να εκπαιδεύσει τέτοιου είδους μοντέλα και στην συνέχεια να χρησιμοποιήσει δικές του εικόνες για προσωπική χρήση.

Η εκπαίδευση του συστήματος παρακολουθείται διαρκώς καθώς υπάρχει η πιθανότητα της υπερπροσαρμογής (Overfitting) κατά την οποία ένα σύστημα εφαρμόζεται ολοκληρωτικά στα δεδομένα εκπαίδευσης και δεν αποδίδει σωστά σε άγνωστες πηγές υλοποίησης. Για την παρακολούθηση και την διόρθωση τέτοιων τυχών λαθών χρησιμοποιούνται συναρτήσεις απώλειας ή συνάρτηση σφάλματος (Loss Functions) οι οποίες είναι συναρτήσεις που κατανέμουν ένα γεγονός ή τιμές μιας ή περισσότερων μεταβλητών σε έναν πραγματικό αριθμό που αντιπροσωπεύει διαισθητικά κάποιο "λάθος" που σχετίζεται με το συμβάν, που στην συγκεκριμένη περίπτωση είναι η υπερπροσαρμογή.

Η επιλογή του κατάλληλου αριθμού επαναλήψεων στην εκπαίδευση του μοντέλου, ο αριθμός των εικόνων που εισήχθησαν στην κατηγορία εκπαίδευση, δοκιμή και επιβεβαίωση (train, test, validation sets) και ο χρόνος εκπαίδευσης είναι όλες μεταβλητές που συμβάλουν στην ποιότητα του αποτελέσματος και την αποτελεσματικότητα της ανάλυσης του μοντέλου.

ABSTRACT

The current era is characterized by both the rapid development of technology and the spread of information and the shift of the economy towards the digitalization of the structures of society. In this context, the use of human resources for tasks that used to be almost compulsory has now become optional.

In response to the challenges and needs imposed by reality, the supply of services and goods must adapt to keep pace with current trends. Robotics, autonomous driving, video surveillance, digital analysis of medical images are some of the areas where technology has made leaps and bounds and the automation of operations is now a given.

The use of Visual Recognition Systems (VRS) is at the core of all these applications. Due to the significant development in Neural Networks these systems have achieved remarkable performance. More specifically, Object Detection is one of these areas. The different algorithms that have been developed over the years (YOLO, Fast R-CNN, SSD etc.) offer a plethora of options for its exploitation methods. Each different algorithm offers different advantages and disadvantages based on the data to be used.

The medical sector benefits significantly from the development of this field as fast and reliable processing of medical images by medical staff is now available. Accurate analysis of CT scans, MRI scans and simple tomographs is becoming a simple process that will probably in the future be able to be performed by the patient himself.

The ease of the Object Detection system and the possibility of its development in the medical field, one of the most important areas of society, was the impetus for the formulation of the topic of this thesis. The evaluation of the methodology of the Faster R-CNN algorithm is done at the experimental level as real CT scans are used. Conclusions are drawn from the use of loss functions in artificial intelligence models trained for the specific function.

Keywords: Visual Recognition Systems (VRS), Object Detection, Neural Networks, Medical Images, CT scans, Faster R-CNN

ΠΕΡΙΛΗΨΗ

Η σημερινή εποχή χαρακτηρίζεται τόσο από τη ραγδαία ανάπτυξη της τεχνολογίας και τη διάδοση της πληροφορίας όσο και από τη στροφή της οικονομίας προς την ψηφιοποίηση των δομών της κοινωνίας. Στο πλαίσιο αυτό, η χρήση ανθρώπινου δυναμικού για εργασίες που παλαιότερα ήταν σχεδόν υποχρεωτικές έχει γίνει πλέον προαιρετική.

Ανταποκρινόμενη στις προκλήσεις και τις ανάγκες που επιβάλλει η πραγματικότητα, η προσφορά υπηρεσιών και αγαθών πρέπει να προσαρμοστεί ώστε να συμβαδίζει με τις τρέχουσες τάσεις. Η ρομποτική, η αυτόνομη οδήγηση, η βιντεοεπιτήρηση, η ψηφιακή ανάλυση ιατρικών εικόνων είναι μερικοί από τους τομείς στους οποίους η τεχνολογία έχει κάνει άλματα και η αυτοματοποίηση των εργασιών είναι πλέον δεδομένη.

Η χρήση συστημάτων οπτικής αναγνώρισης (VRS) βρίσκεται στον πυρήνα όλων αυτών των εφαρμογών. Λόγω της σημαντικής ανάπτυξης των νευρωνικών δικτύων, τα συστήματα αυτά έχουν επιτύχει αξιοσημείωτες επιδόσεις. Πιο συγκεκριμένα, η ανίχνευση αντικειμένων είναι ένας από αυτούς τους τομείς. Οι διάφοροι αλγόριθμοι που έχουν αναπτυχθεί με την πάροδο των ετών (YOLO, Fast R-CNN, SSD κ.λπ.) προσφέρουν μια πληθώρα επιλογών για τις μεθόδους αξιοποίησής του. Κάθε διαφορετικός αλγόριθμος προσφέρει διαφορετικά πλεονεκτήματα και μειονεκτήματα με βάση τα δεδομένα που θα χρησιμοποιηθούν.

Ο ιατρικός τομέας επωφελείται σημαντικά από την ανάπτυξη αυτού του τομέα, καθώς είναι πλέον διαθέσιμη η γρήγορη και αξιόπιστη επεξεργασία ιατρικών εικόνων από το ιατρικό προσωπικό. Η ακριβής ανάλυση των αξονικών τομογραφιών, των μαγνητικών τομογραφιών και των απλών τομογραφιών γίνεται μια απλή διαδικασία που πιθανόν στο μέλλον θα μπορεί να πραγματοποιείται από τον ίδιο τον ασθενή.

Η ευκολία του συστήματος Ανίχνευσης Αντικειμένων και η δυνατότητα ανάπτυξής του στον ιατρικό τομέα, έναν από τους σημαντικότερους τομείς της κοινωνίας, αποτέλεσε το έναυσμα για τη διατύπωση του θέματος της παρούσας διπλωματικής εργασίας. Η αξιολόγηση της μεθοδολογίας του αλγορίθμου Faster R-CNN γίνεται σε πειραματικό επίπεδο καθώς χρησιμοποιούνται πραγματικές αξονικές τομογραφίες. Εξάγονται συμπεράσματα από τη χρήση των

συναρτήσεων απωλειών σε μοντέλα τεχνητής νοημοσύνης που εκπαιδεύονται για τη συγκεκριμένη λειτουργία.

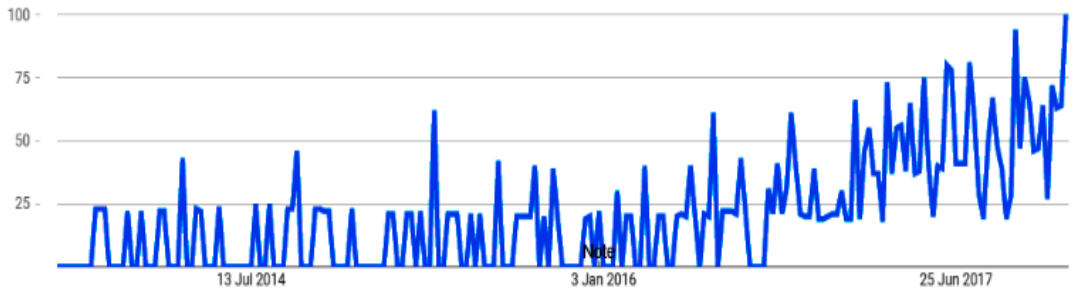
Keywords: Visual Recognition Systems (VRS), Object Detection, Neural Networks, Medical Images, CT scans, Faster R-CNN

Κεφάλαιο 1: Ανίχνευση Αντικειμένων

Στο κεφάλαιο αυτό παρουσιάζεται η έννοια του Object Detection. Πιο συγκεκριμένα γίνεται μια μικρή εισαγωγή στην διαδικασία αυτή καθαυτή, ακολουθούν οι τομείς οι οποίοι εκμεταλλεύονται τα χαρακτηριστικά αυτής της δομής και τέλος παρουσιάζεται ο τρόπος εφαρμογής στις διαδικασίες αυτές.

1.1 Εισαγωγή

Η τεχνολογία της βαθιάς μάθησης έχει γίνει στις μέρες μας σήμα κατατεθέν λόγω των κορυφαίων αποτελεσμάτων που έχουν επιτευχθεί στην τομέα της ταξινόμησης εικόνων (image classification), της ανίχνευσης αντικειμένων (object detection), της επεξεργασίας φυσικής γλώσσας (natural language processing). Οι λόγοι πίσω από τη δημοτικότητα της βαθιάς μάθησης είναι δύο, αρχικά η μεγάλη διαθεσιμότητα συνόλων δεδομένων και οι ισχυροί επεξεργαστές. Καθώς η βαθιά μάθηση απαιτεί μεγάλα σύνολα δεδομένων και ισχυρούς πόρους για την εκτέλεση της εκπαίδευσης, και οι δύο απαιτήσεις έχουν ήδη ικανοποιηθεί σε σημερινής εποχής. Η *Εικόνα 1.1* δείχνει την άνοδο της βαθιάς μάθησης όσον αφορά την όραση υπολογιστών (Computer Vision) την πενταετία από το 2013 έως το 2018.



Εικόνα 1.1: *Άνοδος βαθιάς μάθησης στην όραση υπολογιστών (Computer Vision) από τον Μάρτιο 2013 έως το Ιανουάριο 2018 [1]*

Η ταξινόμηση εικόνων, που είναι ο ευρύτερα ερευνημένος τομέας στον τομέα της όρασης υπολογιστών, έχει επιτύχει αξιοσημείωτα αποτελέσματα σε παγκόσμιους διαγωνισμούς όπως οι ILSVRC, PASCAL VOC και Microsoft COCO με τη βοήθεια της βαθιάς μάθησης. Με κίνητρο τα αποτελέσματα της ταξινόμησης εικόνων, έχουν αναπτυχθεί μοντέλα βαθιάς μάθησης για αντικείμενα ανίχνευσης αντικειμένων και ανίχνευση αντικειμένων με βάση τη βαθιά μάθηση που έχει επίσης επιτύχει κορυφαία αποτελέσματα.

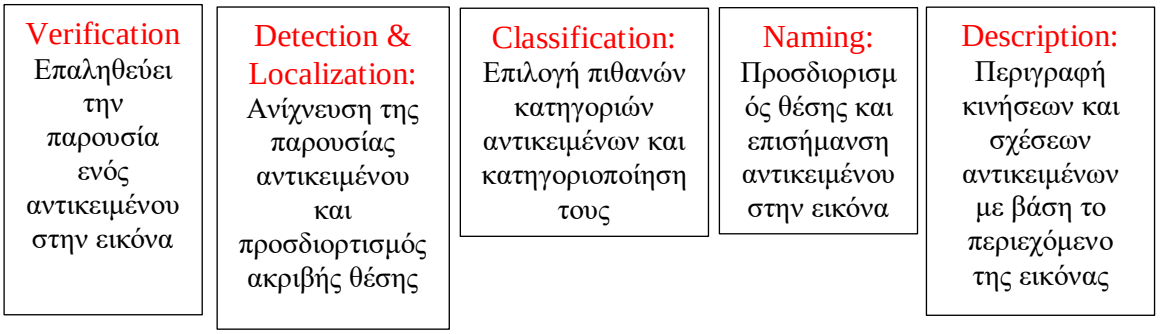
Οι αρχιτεκτονικές βαθιών νευρώνων (Deep Neural Architectures) χειρίζονται πολύπλοκα μοντέλα αποτελεσματικότερα από τα ρηχά δίκτυα (shallow networks). Τα Συνελεκτικά Νευρωνικά Δίκτυα είναι λιγότερο ακριβή για μικρότερα δεδομένα, αλλά παρουσιάζουν σημαντική ακρίβεια που σπάει ρεκόρ στα μεγάλα σύνολα δεδομένων εικόνων. Όμως, τα Συνελεκτικά Νευρωνικά Δίκτυα απαιτούν μεγάλη ποσότητα επισημασμένων συνόλων δεδομένων για την εκτέλεση

εργασιών που σχετίζονται με την όραση υπολογιστών (αναγνώριση, ταξινόμηση και ανίχνευση).

1.2 Επεξήγηση Ανίχνευσης Αντικειμένων

1.2.1 Η Ανίχνευση Αντικειμένων ως το βασικό βήμα για Αναγνώριση Εικόνων

Η Ανίχνευση Αντικειμένων είναι η διαδικασία προσδιορισμού της περίπτωσης της κλάσης στην οποία ανήκει το αντικείμενο και την εκτίμηση της θέσης του αντικειμένου με την εξαγωγή του πλαισίου οριοθέτησης (bounding box) γύρω από το αντικείμενο. Η ανίχνευση μιας μεμονωμένης περίπτωσης της κλάσης από την εικόνα ονομάζεται ανίχνευση αντικειμένου μίας κλάσης, ενώ η ανίχνευση των κλάσεων όλων των αντικειμένων που υπάρχουν στην είναι γνωστή ως ανίχνευση αντικειμένων πολλαπλών κλάσεων. Διαφορετικές προκλήσεις, όπως η μερική/πλήρης απόκρυψη, οι ποικίλες συνθήκες φωτισμού, πόζες, κλίμακα κ.λπ. πρέπει να αντιμετωπιστούν κατά την εκτέλεση της ανίχνευσης αντικειμένων. Όπως φαίνεται στο σχήμα 3, η Ανίχνευση Αντικειμένων είναι το πρώτο βήμα σε κάθε δραστηριότητα οπτικής αναγνώρισης.

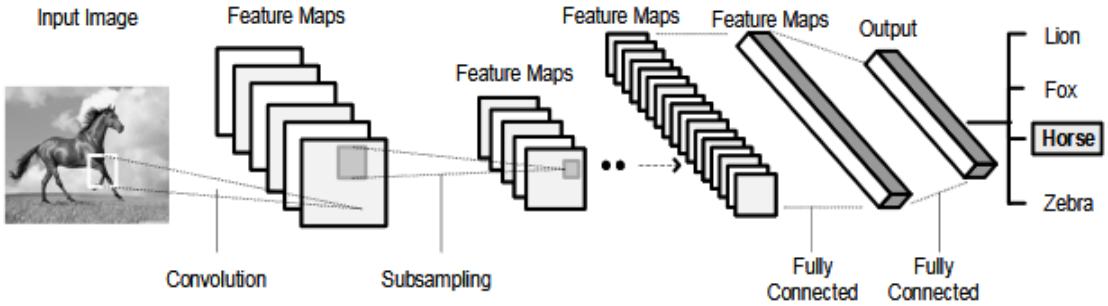


Σχήμα 1.1: *Η Ανίχνευση Αντικεικμένων ως το βασικό βήμα για Αναγνώριση Εικόνων*

1.2.2 Ανίχνευση Αντικειμένων με χρήση Συνελεκτικών Νευρωνικών Δικτύων

Τα Συνελεκτικά Νευρωνικά Δίκτυα (Convolutional Neural Network -CNN) έχουν χρησιμοποιηθεί εκτενώς για την ανίχνευση αντικειμένων. Το CNN είναι ένα νευρωνικό δίκτυο τροφοδότησης τύπου feed-forward και λειτουργεί με βάση την αρχή της κατανομής βαρών. Η συνέλιξη είναι μια ενοποίηση που δείχνει πώς μια συνάρτηση επικαλύπτεται με μια άλλη συνάρτηση και είναι ένα μείγμα δύο συναρτήσεων που πολλαπλασιάζονται. Στην *Εικόνα 1.2* παρουσιάζεται η πολυεπίπεδη αρχιτεκτονική του CNN για το αντικείμενο της Ανίχνευσης Αντικειμένων. Η εικόνα συνελίσσεται με τη συνάρτηση ενεργοποίησης για να προκύψουν χάρτες χαρακτηριστικών. Για να

μειωθεί η χωρική πολυπλοκότητα του δικτύου, οι χάρτες χαρακτηριστικών επεξεργάζονται με στρώματα συγκέντρωσης για να προκύψουν αφηρημένοι χάρτες χαρακτηριστικών. Η διαδικασία αυτή επαναλαμβάνεται για τα επιθυμητό αριθμό φίλτρων και αντίστοιχα δημιουργούνται χάρτες χαρακτηριστικών. Τελικά, αυτοί οι χάρτες χαρακτηριστικών υποβάλλονται σε επεξεργασία με πλήρως συνδεδεμένα στρώματα για να προκύψει η έξοδος της αναγνώρισης εικόνας που δείχνει το σκορ εμπιστοσύνης για την προβλεπόμενη ετικέτα της κλάσης. Για τη βελτίωση της πολυπλοκότητας του δικτύου και τη μείωση του αριθμού των παραμέτρων, το CNN χρησιμοποιεί διαφορετικά είδη στρωμάτων συγκέντρωσης, όπως φαίνεται στον *Πίνακα 1.1*. Τα στρώματα συγκέντρωσης είναι μεταφραστικά αμετάβλητα. Χάρτες ενεργοποίησης τροφοδοτούνται ως είσοδος στα στρώματα συγκέντρωσης. Λειτουργούν σε κάθε τμήμα του επιλεγμένου χάρτη.



Εικόνα 1.2: *Χρήση Συνελεκτικών Νευρωνικών Δικτύων για Ανίχνευση Αντικειμένων [1]*

Pooling Layer	Περιγραφή
Max Pooling	Χρησιμοποιείται ευρέως για την συγκέντρωση των CNN. Παίρνει την μέγιστη τιμή από ένα τμήμα μιας εικόνας και τοποθετεί τον πίνακα για την αποθήκευσεις και άλλων μέγιστων τιμών από άλλες εικόνες.
Average Pooling	Η συγκέντρωση αυτή χρησιμοποιεί τον μέσο όρο των γειτονικών pixel
Deformation pooling	Η παραμορφώσιμη συγκέντρωση έχει την δυνατότητα να εξάγει δεδομένα, γεωμετρικούς περιορισμούς από τα αντικείμενα
Spatial pyramid pooling	Αυτή η συγκέντρωση εκτελεί μείωση της δειγματοληψίας της εικόνας και παράγει διάνυσμα χαρακτηριστικών με σταθερό μήκος. Αυτό το διάνυσμα χαρακτηριστικών μπορεί να χρησιμοποιηθεί για την ανίχνευση αντικειμένων χωρίς να γίνουν παραμορφώσεις στηναρχική εικόνα. Αυτή η συγκέντρωση είναι ανθεκτική στις παραμορφώσεις αντικειμένων.
Scale dependent pooling	Αυτή η ομαδοποίηση αντιμετωπίζει τις διακυμάνσεις κλίμακας στο Object Detection και συμβάλλει στη βελτίωση της ακρίβειας της ανίχνευσης.

Πίνακας 1.1: *Στρώματα συγκέντρωσης που χρησιμοποιούνται για την Ανίχνευση Αντικειμένων*

1.2.3 Πλαίσια και Υπηρεσίες της Ανίχνευσης Αντικειμένων

Ο κατάλογος των πλαισίων βαθιάς μάθησης που είναι διαθέσιμα μέχρι σήμερα είναι εξαντλητικός. Αναφέρονται τα μελετημένα σε αυτή την διπλωματική εργασία και δύο από τα πιο σημαντικά πλαίσια βαθιάς μάθησης στον *Πίνακα 1.2*. Τα πλαίσια μελετώνται από την άποψη των παρουσιαζόμενων χαρακτηριστικών, της διεπαφής, υποστήριξη για το μοντέλο βαθιάς μάθησης, δηλαδή το νευρωνικό δίκτυο συνελίξεων, το επαναλαμβανόμενο νευρωνικό δίκτυο (RNN), το περιορισμένο νευρωνικό δίκτυο (Restricted Botltzmann Machine (RBM) και το δίκτυο βαθιάς πίστης (Deep Belief Network (DBN)) και υποστήριξη για παράλληλη εκτέλεση πολλαπλών κόμβων. Google Cloud Visio API & Microsoft Cognitive Service είναι μερικές από τις υπηρεσίες που μπορούν να χρησιμοποιηθούν για την ανίχνευση αντικειμένων. Αυτές οι υπηρεσίες μπορούν να χρησιμοποιηθούν μέσω των Integrated REST API & REST API αντίστοιχα.

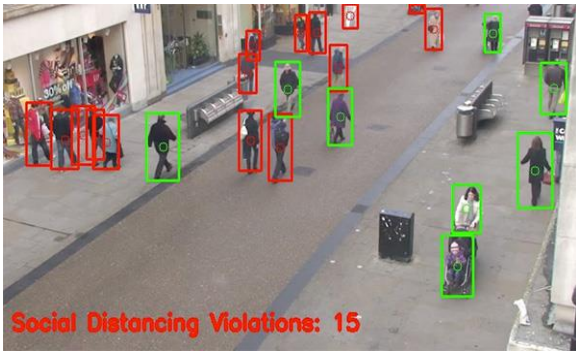
Name	**Features**	**Interface**	**Deep Learning Model**	**Multi-node parallel execution**	**Developer**	**License**
Tensor-Flow	Μαθηματικοί υπολογισμοί με χρήση γραφημάτων ροής δεδομένων, εισαγωγή, ταξινόμηση εικόνων, αυτόματη διαφοροποίηση, φορητότητα	C++, Python, Java, Go	CNN, RNN, DBN/RBM	Ναι	Google, Brain team	Apache 2.0
Keras	Γρήγορη δημιουργία πρωτοτύπων, αρθρωτή, μινιμαλιστικές ενότητες, επεκτάσιμη, αυθαίρετη σύνδεση σχημάτων	Python	CNN, RNN, DBN/RBM	Ναι	F. Chollet	MIT License
PyTorch	N-διάστατος πίνακας υποστήριξη, αυτόματη διαφοροποίηση κλίσης, υποστήριξη νευρωνικών μοντέλων και ενεργειακά μοντέλα	C, C++, Lua, Python	CNN, RNN, DBN/RBM	Ναι	R. Collobert, K. Kavukcuoglu, C. Farabet	BSD License

Πίνακας 1.2: Πλαίσια βαθιάς μάθησης

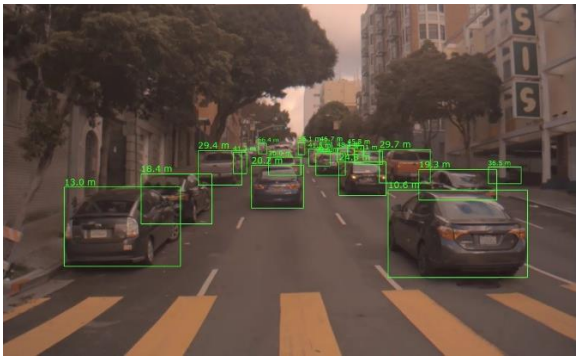
1.3 Περιοχές Εφαρμογής

Η ανίχνευση αντικειμένων εφαρμόζεται σε πολλούς τομείς, από την άμυνα (επιτήρηση), την αλληλεπίδραση ανθρώπου-υπολογιστή, ρομποτική, μεταφορές, ανάκτηση κ.λπ. Οι αισθητήρες που χρησιμοποιούνται για τη διαρκή επιτήρηση παράγουν petabyte από δεδομένα εικόνας μέσα σε λίγες ώρες. Τα δεδομένα αυτά ανάγονται σε γεωχωρικά δεδομένα και ενσωματώνονται με άλλα δεδομένα για να αποκτήσουν σαφή εικόνα του τρέχοντως σεναρίου. Η διαδικασία αυτή περιλαμβάνει την ανίχνευση αντικειμένων για τον εντοπισμό οντοτήτων όπως ανθρώπους, οχήματα και ύποπτα αντικείμενα από από τα ακατέργαστα δεδομένα εικόνας. Ο εντοπισμός και η ανίχνευση των άγριων ζώων στην επικράτεια αποστειρωμένων ζωνών όπως οι βιομηχανικές ζώνες, η ανίχνευση των οχημάτων που σταθμεύουν σε απαγορευμένες περιοχές είναι επίσης ορισμένες εφαρμογές ανίχνευσης αντικειμένων.

Η ανίχνευση των αφύλακτων αποσκευών είναι πολύ σημαντική εφαρμογή της ανίχνευσης αντικειμένων. Για την αυτόνομη οδήγηση, η ανίχνευση αντικειμένων στο δρόμο θα διαδραματίσει σημαντικό ρόλο. Ανίχνευση ελαττωματικών ηλεκτρικών καλωδίων όταν η εικόνα είναι από κάμερες μη επανδρωμένων αεροσκαφών είναι επίσης εφαρμογή ανίχνευσης αντικειμένων. Η ανίχνευση της υπνηλίας των οδηγών στο αυτοκινητόδρομο για την αποφυγή ατυχήματος μπορεί να επιτευχθεί με την ανίχνευση αντικειμένων.



Εικόνα 1.3: *Ανίχνευση Αντικειμένων για βιντεοπαρακολούθηση [2]*



Εικόνα 1.4: *Ανίχνευση Αντικειμένων για αυτόνομη οδήγηση [3]*

Οι απαιτήσεις των προαναφερόμενων εφαρμογών ποικίλλουν ανάλογα με την περίπτωση χρήσης. Οι αναλύσεις ανίχνευσης αντικειμένων μπορούν και εκτελούνται εκτός σύνδεσης, σε απευθείας σύνδεση ή σχεδόν σε πραγματικό χρόνο. Άλλοι παράγοντες, όπως οι αποκρύψεις, η αναλλοίωτη περιστροφή, η ενδο ταξινόμηση και η

παραλλαγή της, και η ανίχνευση αντικειμένων πολλαπλών θέσεων πρέπει να λαμβάνονται υπόψη για την ανίχνευση αντικειμένων.

1.4 Υπερσύγχρονες προσεγγίσεις βαθιάς μάθησης για Ανίχνευση Αντικειμένων

Ο *Πίνακας 1.3* συγκρίνει τις μεθόδους βαθιάς μάθησης για την ανίχνευση αντικειμένων, οι οποίες είναι χρήσιμες για την ερευνητική κοινότητα να εργαστεί περαιτέρω στον τομέα της ανίχνευσης αντικειμένων με βάση τη βαθιά μάθηση. Ο Szegedy κ.α. πρωτοστάτησαν στη χρήση των βαθιών Συνελεκτικών Νευρωνικών Δικτύων για την ανίχνευση αντικειμένων μοντελοποιώντας την ανίχνευση αντικειμένων ως πρόβλημα παλινδρόμησης. Αντικατέστησαν το τελευταίο στρώμα στο AlexNet με στρώμα παλινδρόμησης για την ανίχνευση αντικειμένων. Και τα δύο καθήκοντα της ανίχνευσης και του εντοπισμού πραγματοποιήθηκαν με τη χρήση μάσκας παλινδρόμησης αντικειμένου. Το DeepMultiBox επέκτεινε την προσέγγιση για την ανίχνευση πολλαπλών αντικειμένων σε μια εικόνα.

Ο τρόπος με τον οποίο το Συνελεκτικό Νευρωνικό Δίκτυο μαθαίνει το χαρακτηριστικό γνώρισμα είναι ένα σημαντικό ζήτημα. Το έργο της οπτικοποίησης των χαρακτηριστικών του Συνελεκτικού Νευρωνικού Δικτυού γίνεται από τον Zeiler και λοιπούς. Εφάρμοσαν τόσο τα Συνελεκτικά Νευρωνικά Δίκτυα όσο και τη διαδικασία αποσυνέλιξης για την οπτικοποίηση των χαρακτηριστικών. Αυτή η προσέγγιση υπερτερεί έναντι στο AlexNet. Αιτιολόγησαν επίσης ότι η απόδοση του βαθιού μοντέλου επηρεάζεται από το βάθος του δικτύου. Μοντέλο υπερχείλισης εφαρμόζει την προσέγγιση του ολισθαίνοντος παραθύρου με βάση την πολυκλιμάκωση για την από κοινού εκτέλεση της ταξινόμησης, της ανίχνευσης και του εντοπισμού. Ο Girshick κ.α. πρότειναν βαθύ μοντέλο που βασίζεται σε προτάσεις περιοχής. Σε αυτή την προσέγγιση, η εικόνα διαιρείται σε μικρές περιοχές και στη συνέχεια χρησιμοποιείται το βαθύ Συνελεκτικό Νευρωνικό Δίκτυο για τη λήψη διανυσμάτων χαρακτηριστικών. Τα διανύσματα χαρακτηριστικών χρησιμοποιούνται για την ταξινόμηση με γραμμικό SVM. Ο εντοπισμός του αντικειμένου γίνεται με τη χρήση παλινδρόμησης bounding-box. Σε παρόμοιες γραμμές, χρησιμοποιήθηκαν regionlets για τη γενική ανίχνευση αντικειμένων ανεξάρτητα από τις πληροφορίες περιβάλλοντος. Σχεδίασαν τη μετρική Support Pixel Integral Image για την εξαγωγή χαρακτηριστικών με βάση το ιστόγραμμα των κλίσεων, τα χαρακτηριστικά συνδιακύμανσης και το αραιό Συνελεκτικό Νευρωνικό Δίκτυο.

Πριν από το ξέσπασμα της βαθιάς μάθησης, η ανίχνευση αντικειμένων γινόταν κατά προτίμηση με τη χρήση της τεχνικής του παραμορφώσιμου μοντέλου. Η τεχνική του παραμορφώσιμου μοντέλου μέρους εκτελεί ανίχνευση και εντοπισμό αντικειμένων βάσει πολλαπλών κλιμάκων. Με βάση τις αρχές αυτού του μοντέλου, ο Ouyang κ.ά. πρότειναν στρώμα συγκέντρωσης για το χειρισμό της παραμόρφωσης των ιδιοτήτων των αντικειμένων για λόγους ανίχνευσης.

Αναμένεται ότι τα συστήματα ανίχνευσης αντικειμένων θα πρέπει να εκτελούν σταθερά την ανίχνευση αντικειμένων αναλλοίωτη από τον φωτισμό, αποκρύψεις, παραμορφώσεις και διακυμάνσεις εντός της κλάσης. Καθώς οι συγκαλύψεις και οι παραμορφώσεις ακολουθούν στατιστική κατανομή μακράς ουράς (long-tail statistical distribution), υπάρχει πιθανότητα τα σύνολα δεδομένων να χάσουν τις σπάνιες αποκρύψεις και παραμορφώσεις των αντικειμένων. Αυτό εμποδίζει την απόδοση των συστημάτων ανίχνευσης αντικειμένων. Ως εκ τούτου, ο Wang κ.ά. πρότειναν την προσέγγιση που βασίζεται σε αντίπαλο δίκτυο στο οποίο το δίκτυο παράγει επιλεκτικά τα χαρακτηριστικά των αποκρύψεων και των παραμορφώσεων που είναι δύσκολο να αναγνωριστούν από τον ανιχνευτή αντικειμένων. Χρησιμοποίησαν το δίκτυο χωρικής απόρριψης και το δίκτυο χωρικού μετασχηματιστή με βάση αντίπαλο δίκτυο για τη δημιουργία χαρακτηριστικών απόκρυψης και παραμόρφωσης αντίστοιχα.

Μέθοδος	Τρόπος λειτουργίας	Χαρακτηριστικά
Deep saliency network	Τα CNN χρησιμοποιούνται για την εξαγωγή χαρακτηριστικών υψηλού επιπέδου και πολλαπλής κλίμακας	Είναι δύσκολο να εντοπιστούν τα όρια των σημαντικών περιοχών λόγω του γεγονότος ότι τα εικονοστοιχεία που βρίσκονται στην οριακή περιοχή έχουν παρόμοια δεκτικά πεδία. Λόγω αυτού, το δίκτυο μπορεί να καταλήξει σε ανακριβή χάρτη και σχήμα ως προς την ανίχνευση του αντικειμένου.
Generating image (or pixels)	Η μέθοδος αυτή χρησιμοποιείται όταν η εμφάνιση αποφράξεων και παραμορφώσεων είναι σπάνια σε στη βάση δεδομένων	Η μέθοδος αυτή δημιουργεί νέες εικόνες με αποκρύψεις και παραμορφώσεις μόνο όταν τα δεδομένα εκπαίδευσης περιέχουν εμφανίσεις αποκρύψεων και παραμορφώσεων.
Generating all possible occlusions and deformations	σε αυτή τη μέθοδο, όλα τα σύνολα των πιθανών αποκρύψεων και παραμορφώσεων παράγονται για την εκπαίδευση των ανιχνευτών αντικειμένων	Η μέθοδος αυτή δεν είναι επεκτάσιμη, δεδομένου ότι οι παραμορφώσεις και οι αποκρύψεις απαιτούν μεγάλο χώρο.
Adversarial learning	Αντί να δημιουργούνται όλες οι παραμορφώσεις και οι αποκρύψεις, η μέθοδος αυτή χρησιμοποιεί αντιφατικό δίκτυο το οποίο παράγει επιλεκτικά χαρακτηριστικά που μιμούνται τις αποκρύψεις και παραμορφώσεις που είναι δύσκολο να αναγνωρίζονται από τον ανιχνευτή αντικειμένων.	Καθώς αυτή η μέθοδος παράγει τα παραδείγματα κατα την διαρκεια, είναι η καλύτερη για να εφαρμοστεί σε πραγματικό χρόνο για ανίχνευση αντικειμένων. Καθώς, παράγει επιλεκτικά τα χαρακτηριστικά, είναι και επεκτάσιμη.
Part-based method	Αυτή η μέθοδος αναπαριστά το αντικείμενο ως συλλογή τοπικών τμημάτων και χωρικής δομής. Αυτή η μέθοδος αναζητά εξαντλητικά πολλαπλά μέρη για την ανίχνευση αντικειμένων	Η μέθοδος αυτή αντιμετωπίζει το ζήτημα της ενδο-ταξικής διαφοροποίησης στις κατηγορίες αντικειμένων. Τέτοιες διαφοροποιήσεις εμφανίζονται εξαιτίας της διαφοροποίησης στις πόζες, στο ακατάστατο φόντο και στην μερική απόκρυψη
CNN with part-based method	Σε αυτή τη μέθοδο, το μοντέλο παραμορφώσιμου μέρους χρησιμοποιείται για τη μοντελοποίηση της χωρικής δομής του των τοπικών τμημάτων, ενώ το CNN χρησιμοποιείται για την εκμάθηση των διακριτικών χαρακτηριστικών	Αυτή η μέθοδος αντιμετωπίζει το ζήτημα των μερικών αποκρύψεων. Απαιτεί όμως πολλαπλά μοντέλα CNN για μερική ανίχνευση αντικειμένων. Η εύρεση του βέλτιστου αριθμού τμημάτων ανά αντικείμενο είναι επίσης πρόκληση.
Fine-grained object detection method	Η μέθοδος αυτή λειτουργεί σε annotated αντικείμενα κατα την διάρκεια της εκπαίδευσης. Ο μερικός εντοπισμός είναι το θεμελιώδες μέρος στο δοκιμαστικό κομμάτι	Αυτή η μέθοδος έχει τη δυνατότητα να υπολογίσει τις διαφορές στα αντικείμενα μεταξύ των κατηγοριών σε πιο λεπτό επίπεδο. Και εργάζονται περισσότερο σε διακριτικά μέρη σε σύγκριση με τις γενικές μεθόδους ανίχνευσης αντικειμένων.

Πίνακας 1.3: *Σύγκριση μεθόδων ανίχνευσης αντικειμένων με βάση τη βαθιά μάθηση*

Κεφάλαιο 2: Αλγόριθμοι Ανίχνευσης Αντικειμένων

Στο κεφάλαιο αυτό παρουσιάζεται οι διαφορετικοί αλγόριθμοι που χρησιμοποιούνται ανα καιρούς για την εκπόνηση της Ανίχνευσης Αντικειμένων. Για κάθε αλγόριθμο θα γίνεται μια σύντομη παρουσίαση του και θα επεξηγούνται τα βασικά χαρακτηριστικά.

2.1 Εισαγωγή

Τα μοντέλα ανίχνευσης αντικειμένων συνήθως εκπαιδεύονται για την ανίχνευση της παρουσίας συγκεκριμένων αντικειμένων. Τα κατασκευασμένα μοντέλα μπορούν να χρησιμοποιηθούν σε εικόνες, βίντεο ή λειτουργίες πραγματικού χρόνου. Ακόμη και πριν από τις μεθοδολογίες βαθιάς μάθησης και τις σύγχρονες τεχνολογίες επεξεργασίας εικόνας, η ανίχνευση αντικειμένων ήταν ένα μεγάλο πεδίο ενδιαφέροντος. Ορισμένες μέθοδοι (όπως οι SIFT και HOG με τις τεχνικές εξαγωγής χαρακτηριστικών και ακμών τους) είχαν επιτυχία στην ανίχνευση αντικειμένων και υπήρχαν σχετικά λίγοι άλλοι ανταγωνιστές σε αυτόν τον τομέα.

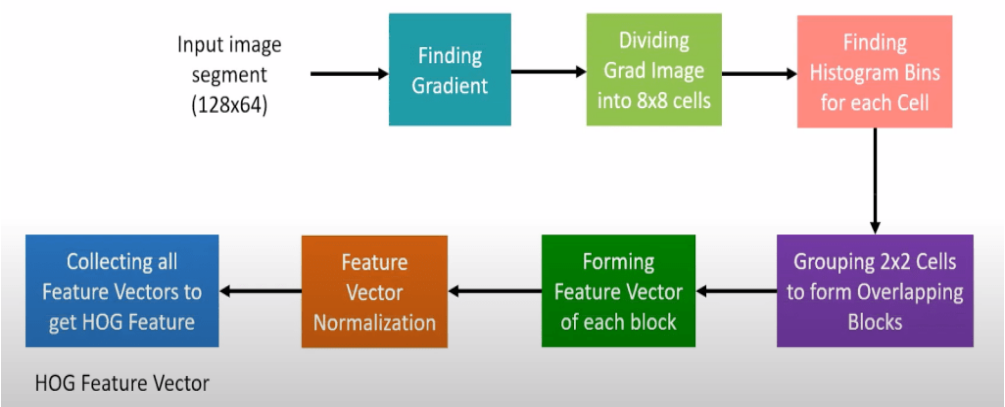
Με την εισαγωγή των συνελικτικών νευρωνικών δικτύων (CNN) και την προσαρμογή των τεχνολογιών υπολογιστικής όρασης, η ανίχνευση αντικειμένων έγινε πολύ πιο διαδεδομένη στη σημερινή γενιά. Το νέο κύμα ανίχνευσης αντικειμένων με προσεγγίσεις βαθιάς μάθησης ανοίγει φαινομενικά ατελείωτες δυνατότητες.

2.2 Αλγόριθμοι

2.2.1 Histogram of Oriented Gradients (HOG)

Το Histogram of Oriented Gradients είναι μια από τις παλαιότερες μεθόδους ανίχνευσης αντικειμένων. Παρουσιάστηκε για πρώτη φορά το 1986. Παρά κάποιες εξελίξεις την επόμενη δεκαετία, η προσέγγιση δεν απέκτησε μεγάλη δημοτικότητα μέχρι το 2005, όταν άρχισε να χρησιμοποιείται σε πολλές εργασίες που σχετίζονται με την όραση υπολογιστών. Η HOG χρησιμοποιεί έναν εκχυλιστή χαρακτηριστικών για τον εντοπισμό αντικειμένων σε μια εικόνα.

Η περιγραφή χαρακτηριστικών που χρησιμοποιείται στο HOG είναι μια αναπαράσταση ενός τμήματος μιας εικόνας όπου εξάγονται μόνο οι πιο απαραίτητες πληροφορίες, ενώ οτιδήποτε άλλο αγνοείται. Η λειτουργία του περιγραφέα χαρακτηριστικών είναι η μετατροπή του συνολικού μεγέθους της εικόνας σε μορφή πίνακα ή διανύσματος χαρακτηριστικών. Στο HOG, χρησιμοποιείται η διαδικασία προσανατολισμού κλίσης για να εντοπιστούν τα πιο κρίσιμα μέρη μιας εικόνας.



Εικόνα 2.1: *Αρχιτεκτονική HOG* ***[44]***

Για ένα συγκεκριμένο εικονοστοιχείο σε μια εικόνα, το ιστόγραμμα της κλίσης υπολογίζεται λαμβάνοντας υπόψη τις κάθετες και οριζόντιες τιμές για να ληφθούν τα διανύσματα χαρακτηριστικών. Με τη βοήθεια του μεγέθους της κλίσης και των γωνιών κλίσης, μπορεί να ληφεθί μια σαφή τιμή για το τρέχον εικονοστοιχείο εξερευνώντας τις άλλες οντότητες στο οριζόντιο και κάθετο περιβάλλον τους.

Όπως φαίνεται στην *Εικόνα 2.1* , θα θεωρηθεί ένα τμήμα εικόνας συγκεκριμένου μεγέθους. Το πρώτο βήμα είναι να βρεθεί η κλίση χωρίζοντας ολόκληρο τον υπολογισμό της εικόνας σε αναπαραστάσεις κλίσης 8×8 κελιών. Με τη βοήθεια των 64 διανυσμάτων κλίσης που επιτυγχάνονται, χωρίζεται κάθε κελί σε γωνιακά bins και νυπολογίζεται το ιστόγραμμα για τη συγκεκριμένη περιοχή. Αυτή η διαδικασία μειώνει το μέγεθος των 64 διανυσμάτων σε ένα μικρότερο μέγεθος 9 τιμών.

Μόλις βρεθεί το μέγεθος των 9 σημειακών τιμών ιστογράμματος (bins) για κάθε κελί, ξεκινά η δημιουργία των επικαλύψεων για τα μπλοκ των κελιών. Τα τελευταία βήματα είναι ο σχηματισμός των μπλοκ χαρακτηριστικών, η κανονικοποίηση των χαρακτηριστικών των λαμβανόμενων διανυσμάτων και η συλλογή όλων των διανυσμάτων χαρακτηριστικών για να ληφθεί ένα συνολικό χαρακτηριστικό HOG.

- **Περιορισμοί**: Ενώ το Histogram of Oriented Gradients (HOG) ήταν αρκετά επαναστατικό στα αρχικά στάδια της ανίχνευσης αντικειμένων, υπήρχαν πολλά προβλήματα σε αυτή τη μέθοδο. Είναι αρκετά χρονοβόρα για πολύπλοκους υπολογισμούς εικονοστοιχείων σε εικόνες και

αναποτελεσματική σε ορισμένα σενάρια ανίχνευσης αντικειμένων με στενότερους χώρους.

- **Αποτελεσματική Χρήση**: Ο HOG συχνά χρησιμοποιείται ως η πρώτη μέθοδος ανίχνευσης αντικειμένων για τη δοκιμή άλλων αλγορίθμων και των αντίστοιχων επιδόσεών τους. Ανεξάρτητα από αυτό, ο HOG βρίσκει σημαντική χρήση στις περισσότερες περιπτώσεις ανίχνευσης αντικειμένων και αναγνώρισης αξιοθέατων με αξιοπρεπή ακρίβεια.

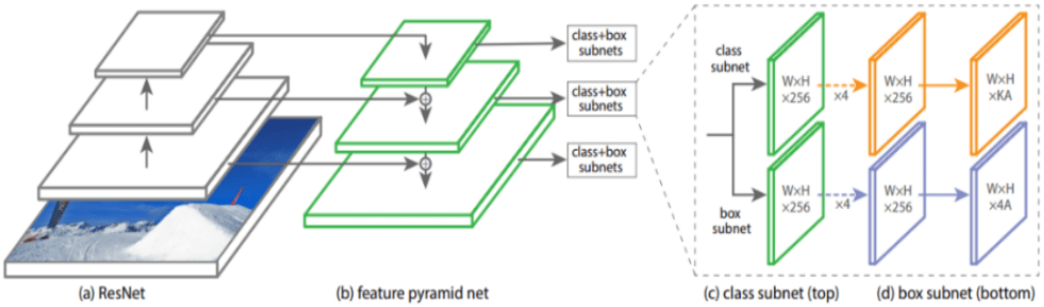
2.2.2 RetinaNet

Το μοντέλο RetinaNet εισήχθη το 2017 και έγινε ένα από τα καλύτερα μοντέλα με δυνατότητες ανίχνευσης αντικειμένων με μία λήψη που μπορούσαν να ξεπεράσουν άλλους δημοφιλείς αλγορίθμους ανίχνευσης αντικειμένων κατά τη διάρκεια αυτής της περιόδου. Όταν κυκλοφόρησε η αρχιτεκτονική RetinaNet, οι δυνατότητες ανίχνευσης αντικειμένων ξεπέρασαν αυτές των μοντέλων Yolo v2 και SSD. Διατηρώντας την ίδια ταχύτητα με αυτά τα μοντέλα, ήταν επίσης σε θέση να ανταγωνιστεί την οικογένεια R-CNN όσον αφορά την ακρίβεια. Λόγω αυτών, το μοντέλο RetinaNet βρίσκει μεγάλη χρήση στην ανίχνευση αντικειμένων μέσω δορυφορικών εικόνων.

Η αρχιτεκτονική του RetinaNet είναι κατασκευασμένη με τέτοιο τρόπο ώστε τα προηγούμενα προβλήματα των ανιχνευτών μιας λήψης να εξισορροπούνται κάπως, ώστε να παράγονται πιο αποτελεσματικά και αποδοτικά αποτελέσματα. Σε αυτή την αρχιτεκτονική μοντέλου, η απώλεια διασταυρούμενης εντροπίας στα προηγούμενα μοντέλα αντικαθίσταται από την απώλεια εστίασης. Η εστιακή απώλεια χειρίζεται τα προβλήματα ανισορροπίας κλάσεων που υπάρχουν σε αρχιτεκτονικές όπως η YOLO και η SSD. Το μοντέλο RetinaNet είναι ένας συνδυασμός τριών κύριων οντοτήτων.

Το RetinaNet κατασκευάζεται χρησιμοποιώντας τρεις παράγοντες, δηλαδή το μοντέλο ResNet (συγκεκριμένα το ResNet-101), το δίκτυο πυραμίδας χαρακτηριστικών (Feature Pyramid Network - FPN) και την εστιακή απώλεια. Το δίκτυο πυραμίδας χαρακτηριστικών είναι μία από τις καλύτερες μεθόδους για την αντιμετώπιση της πλειονότητας των ελλείψεων της προηγούμενης αρχιτεκτονικής. Βοηθά στο συνδυασμό των σημασιολογικά πλούσιων χαρακτηριστικών των εικόνων χαμηλότερης ανάλυσης με εκείνα των σημασιολογικά αδύναμων χαρακτηριστικών των εικόνων υψηλότερης ανάλυσης.

Στην τελική έξοδο, δημιουργούνται τόσο τα μοντέλα ταξινόμησης όσο και παλινδρόμησης παρόμοια με τις άλλες μεθόδους ανίχνευσης αντικειμένων που συζητήθηκαν προηγουμένως. Το δίκτυο ταξινόμησης χρησιμοποιείται για τις κατάλληλες προβλέψεις πολλαπλών κατηγοριών, ενώ το δίκτυο παλινδρόμησης δημιουργείται για την πρόβλεψη των κατάλληλων πλαισίων οριοθέτησης για τις ταξινομημένες οντότητες.



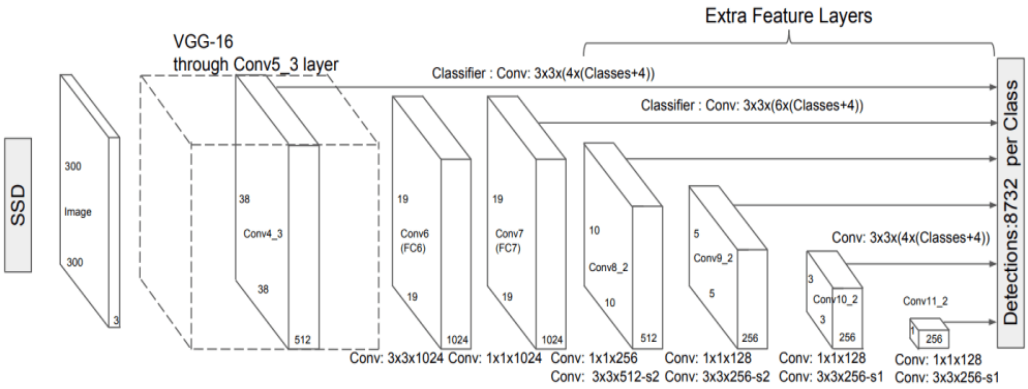
Εικόνα 2.5: *Αρχιτεκτονική ResinaNet [111]*

- **Αποτελεσματική Χρήση:** Το μοντέλο RetinaNet είναι σήμερα μία από τις καλύτερες μεθόδους για την ανίχνευση αντικειμένων σε διάφορες εργασίες. Μπορεί να χρησιμοποιηθεί ως αντικαταστάτης ενός ανιχνευτή μονής λήψης για ένα πλήθος εργασιών για την επίτευξη γρήγορων και ακριβών αποτελεσμάτων για εικόνες.

2.2.3 Single Shot Detector (SSD)

Ο ανιχνευτής μίας λήψης για προβλέψεις πολλαπλών κουτιών είναι ένας από τους ταχύτερους τρόπους για την επίτευξη του υπολογισμού σε πραγματικό χρόνο των εργασιών ανίχνευσης αντικειμένων. Ενώ οι μεθοδολογίες Faster R-CNN μπορούν να επιτύχουν υψηλές ακρίβειες πρόβλεψης, η συνολική διαδικασία είναι αρκετά χρονοβόρα και απαιτεί η εργασία πραγματικού χρόνου να εκτελείται με περίπου 7 καρέ ανά δευτερόλεπτο, κάτι που απέχει πολύ από το επιθυμητό.

Ο ανιχνευτής μίας λήψης (SSD) λύνει αυτό το ζήτημα βελτιώνοντας τα καρέ ανά δευτερόλεπτο σχεδόν πέντε φορές περισσότερο από το μοντέλο Faster R-CNN. Αφαιρεί τη χρήση του δικτύου πρότασης περιοχής και αντ' αυτού κάνει χρήση χαρακτηριστικών πολλαπλής κλίμακας και προεπιλεγμένων πλαισίων.



Εικόνα 2.3: *Αρχιτεκτονική SSD [112]*

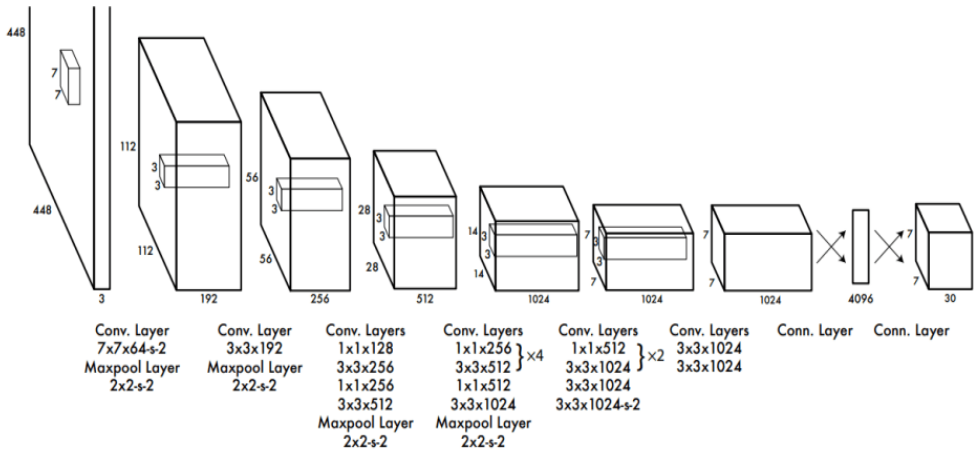
Ο SSD μπορεί να χωριστεί κυρίως σε τρία στοιχεία. Το πρώτο στάδιο του ανιχνευτή single-shot είναι το βήμα εξαγωγής χαρακτηριστικών, όπου επιλέγονται όλοι οι κρίσιμοι χάρτες χαρακτηριστικών. Αυτή η αρχιτεκτονική περιοχή αποτελείται μόνο από πλήρως συνελικτικά στρώματα και από κανένα άλλο στρώμα. Μετά την εξαγωγή όλων των βασικών χαρτών χαρακτηριστικών, το επόμενο βήμα είναι η διαδικασία ανίχνευσης κεφαλών. Αυτό το βήμα αποτελείται επίσης από πλήρως συνελικτικά νευρωνικά δίκτυα.

Ωστόσο, στο δεύτερο στάδιο της ανίχνευσης κεφαλών, ο στόχος δεν είναι να βρεθεί το σημασιολογικό νόημα για τις εικόνες. Αντίθετα, ο πρωταρχικός στόχος είναι η δημιουργία των καταλληλότερων χαρτών οριοθέτησης για όλους τους χάρτες χαρακτηριστικών. Αφού υπολογιστούν τα δύο βασικά στάδια, το τελικό στάδιο είναι να περαστούν από τα στρώματα μη μέγιστης καταστολής για τη μείωση του ποσοστού σφάλματος που προκαλείται από τα επαναλαμβανόμενα πλαίσια οριοθέτησης.

- **Αποτελεσματική Χρήση**: Ο SSD είναι συχνά η προτιμώμενη μέθοδος. Ο κύριος λόγος για τη χρήση του ανιχνευτή μονής βολής είναι ότι προτιμούνται κυρίως ταχύτερες προβλέψεις σε μια εικόνα για την ανίχνευση μεγαλύτερων αντικειμένων, όπου η ακρίβεια δεν αποτελεί εξαιρετικά σημαντικό μέλημα. Ωστόσο, για πιο ακριβείς προβλέψεις για μικρότερα και ακριβή αντικείμενα, εξετάζονται άλλες μέθοδοι.

2.2.4 YOLO (You Only Look Once)

Το YOLO (You only look once) είναι μία από τις πιο δημοφιλείς αρχιτεκτονικές μοντέλων και αλγορίθμων για την ανίχνευση αντικειμένων. Συνήθως, η πρώτη έννοια που συναντάται σε μια αναζήτηση στο Google για αλγόριθμους ανίχνευσης αντικειμένων είναι η αρχιτεκτονική YOLO. Υπάρχουν διάφορες εκδόσεις της YOLO. Το μοντέλο YOLO χρησιμοποιεί ένα από τα καλύτερα αρχέτυπα νευρωνικών δικτύων για την παραγωγή υψηλής ακρίβειας και συνολικής ταχύτητας επεξεργασίας. Αυτή η ταχύτητα και η ακρίβεια είναι ο κύριος λόγος για τη δημοτικότητά του.



Εικόνα 2.4: *Αρχιτεκτονική YOLO* ***[87]***

Η αρχιτεκτονική YOLO χρησιμοποιεί τρεις βασικές ορολογίες για να επιτύχει τον στόχο της ανίχνευσης αντικειμένων. Η κατανόηση αυτών των τριών τεχνικών είναι αρκετά σημαντική για την αντίληψη του, το μοντέλο αποδίδει γρήγορα και με ακρίβεια σε σύγκριση με άλλους αλγορίθμους ανίχνευσης αντικειμένων. Η πρώτη έννοια στο μοντέλο YOLO είναι τα υπολειμματικά μπλοκ. Στον πρώτο αρχιτεκτονικό σχεδιασμό, έχουν χρησιμοποιηθεί μπλοκ υπολειμμάτων 7×7 για να δημιουργήσουν πλέγματα στη συγκεκριμένη εικόνα.

Κάθε ένα από αυτά τα πλέγματα λειτουργεί ως κεντρικό σημείο και μια συγκεκριμένη πρόβλεψη για κάθε ένα από αυτά τα πλέγματα γίνεται αναλόγως. Στη δεύτερη τεχνική, κάθε ένα από τα κεντρικά σημεία για μια συγκεκριμένη πρόβλεψη λαμβάνεται υπόψη για τη δημιουργία των bounding boxes. Ενώ οι εργασίες ταξινόμησης λειτουργούν καλά για κάθε πλέγμα, είναι πιο περίπλοκο να διαχωρίζονται τα bounding boxes για κάθε μία από τις προβλέψεις που γίνονται. Η τρίτη και τελευταία τεχνική είναι η χρήση της τομής της

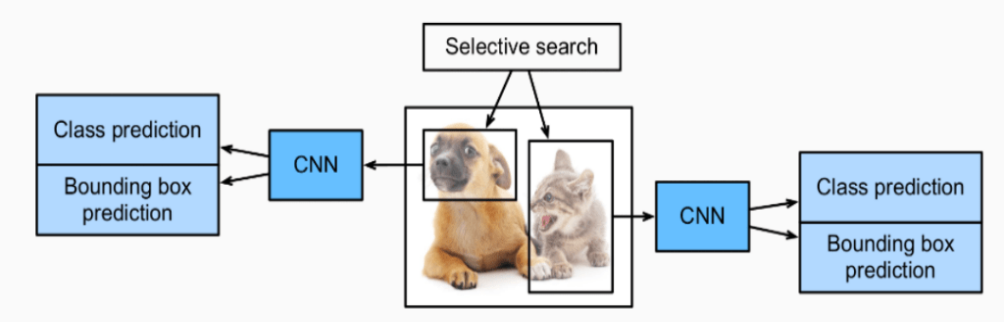
ένωσης (IOU) για τον υπολογισμό των καλύτερων πλαισίων οριοθέτησης για τη συγκεκριμένη εργασία ανίχνευσης αντικειμένων.

- **Περιορισμοί**: α)Αποτυχία εντοπισμού μικρότερων αντικειμένων σε μια εικόνα ή ένα βίντεο λόγω του χαμηλότερου ποσοστού ανάκλησης. β)Δεν μπορεί να ανιχνεύσει δύο αντικείμενα που βρίσκονται εξαιρετικά κοντά το ένα στο άλλο λόγω των περιορισμών των bounding boxes.

- **Αποτελεσματική Χρήση**: Ενώ όλες οι μέθοδοι που συζητήθηκαν προηγουμένως αποδίδουν αρκετά καλά στις εικόνες και μερικές φορές στην ανάλυση βίντεο για την ανίχνευση αντικειμένων, η αρχιτεκτονική YOLO είναι μία από τις πλέον προτιμώμενες μεθόδους για την εκτέλεση της ανίχνευσης αντικειμένων σε πραγματικό χρόνο. Επιτυγχάνει υψηλή ακρίβεια στις περισσότερες εργασίες επεξεργασίας σε πραγματικό χρόνο με αξιοπρεπή ταχύτητα και καρέ ανά δευτερόλεπτο ανάλογα με τη συσκευή στην οποία εκτελείτε το πρόγραμμα.

2.2.5 Region-based Convolutional Neural Networks (R-CNN)

Το R-CNN αποτελεί βελτίωση της διαδικασίας ανίχνευσης αντικειμένων σε σχέση με τις προηγούμενες μεθόδους HOG και SIFT. Στα μοντέλα R-CNN, γίνεται η προσπάθεια να εξαχθούν τα πιο ουσιώδη χαρακτηριστικά (συνήθως περίπου 2000 χαρακτηριστικά) κάνοντας χρήση επιλεκτικών χαρακτηριστικών. Η διαδικασία επιλογής των πιο σημαντικών εξαγωγών μπορεί να υπολογιστεί με τη βοήθεια ενός αλγορίθμου επιλεκτικής αναζήτησης που μπορεί να επιτύχει αυτές τις σημαντικές περιφερειακές προτάσεις.



Εικόνα 2.2: *Αρχιτεκτονική R-CNN [86]*

Η διαδικασία εργασίας του αλγορίθμου επιλεκτικής αναζήτησης για την επιλογή των πιο σημαντικών περιφερειακών προτάσεων είναι να διασφαλίσει την δημιουργία πολλαπλών υποτμηματοποιήσεων σε μια συγκεκριμένη εικόνα και επιλέγει τις υποψήφιες καταχωρήσεις για την εργασία. Ο αλγόριθμος απληστίας μπορεί στη συνέχεια να χρησιμοποιηθεί για τον συνδυασμό των αποτελεσματικών καταχωρίσεων αναλόγως για μια επαναλαμβανόμενη διαδικασία συνδυασμού των μικρότερων τμημάτων σε κατάλληλα μεγαλύτερα τμήματα.

Μόλις ολοκληρωθεί επιτυχώς ο αλγόριθμος επιλεκτικής αναζήτησης, τα επόμενα βήματα είναι η εξαγωγή των χαρακτηριστικών και η πραγματοποίηση των κατάλληλων προβλέψεων. Στη συνέχεια, εξάγονται τελικές υποψήφιες προτάσεις και τα νευρωνικά δίκτυα συνελικτικού τύπου μπορούν να χρησιμοποιηθούν για τη δημιουργία ενός διανύσματος χαρακτηριστικών n-διαστάσεων (είτε 2048 είτε 4096) ως έξοδος. Με τη βοήθεια ενός προεκπαιδευμένου συνελικτικού νευρωνικού δικτύου, επιτυγχάνεται η εξαγωγή χαρακτηριστικών με ευκολία.

Το τελικό βήμα του R-CNN είναι να κάνει τις κατάλληλες προβλέψεις για την εικόνα και να επισημάνει ανάλογα το αντίστοιχο πλαίσιο οριοθέτησης. Προκειμένου να επιτευχθούν τα καλύτερα αποτελέσματα για κάθε εργασία, οι προβλέψεις γίνονται με τον υπολογισμό ενός μοντέλου ταξινόμησης για κάθε εργασία, ενώ ένα μοντέλο παλινδρόμησης χρησιμοποιείται για τη διόρθωση της ταξινόμησης του bounding box για τις προτεινόμενες περιοχές.

- **Αποτελεσματική Χρήση**: Το R-CNN, παρόμοια με τη μέθοδο HOG, χρησιμοποιείται ως μια πρώτη βάση για τη δοκιμή της απόδοσης των μοντέλων ανίχνευσης αντικειμένων. Ο χρόνος που απαιτείται για την πρόβλεψη εικόνων και αντικειμένων μπορεί να διαρκέσει λίγο περισσότερο από το αναμενόμενο, οπότε συνήθως προτιμώνται οι πιο σύγχρονες εκδόσεις του R-CNN.

Κεφάλαιο 3: Αλγόριθμος Faster R-CNN

Στο κεφάλαιο αυτό παρουσιάζεται ο αλγόριθμος που χρησιμοποιήθηκε για την υλοποίηση της διπλωματικής εργασίας. Δίνεται μια επισκόπηση στην δομή του αλγόριθμου, τον τρόπο λειτουργείας και μια εκτενής ανάλυση των βασικών χαρακτηριστικών για τις οποίες επιλέχθηκε.

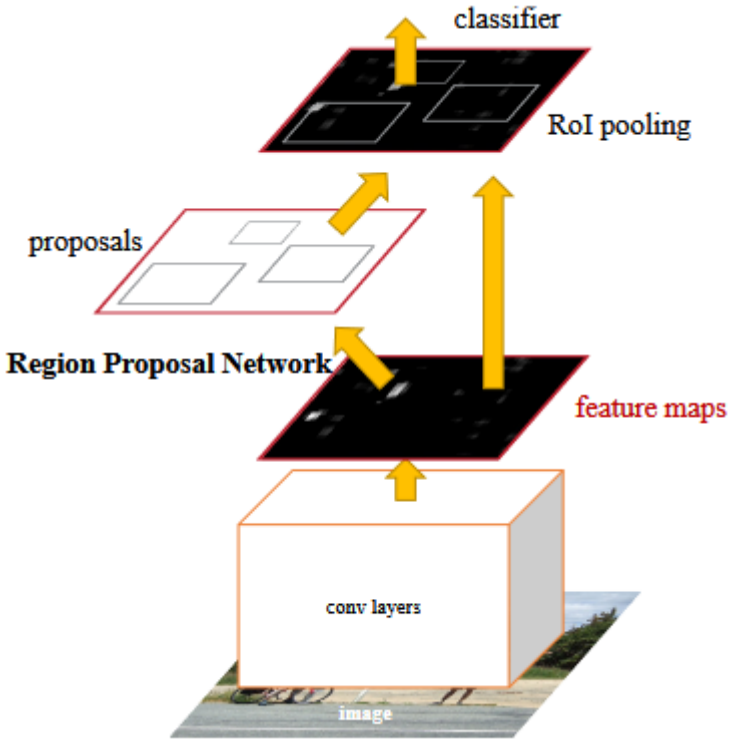
3.1 Εισαγωγή

Οι πρόσφατες εξελίξεις στον τομέα της Ανίχνευσης Αντικειμένων οφείλονται στην επιτυχία του Δίκτυο Προτάσεων Περιοχής (Region Proposal Network (RPN)) και Συνελεκτικό Νευρωνικό Δίκτυο με βάση την περιοχή (region-based convolutional networks (R-CNN)). Παρόλο που τα CNNs με βάση την περιοχή ήταν ακριβά υπολογιστικά, το κόστος τους έχει μειωθεί δραστικά χάρη στην κοινή χρήση των νευρώνων σε όλα τα Region Proposals. Η πιο πρόσφατη ενσάρκωση, το Fast R-CNN, επιτυγχάνει ρυθμούς επεξεργασίας σχεδόν σε πραγματικό χρόνο χρησιμοποιώντας πολύ βαθιά δίκτυα (deep networks), όταν δεν υπολογίζεται ο χρόνος που δαπανάται για τα Region Proposals. Τώρα, τα RPN είναι η υπολογιστική συμφόρηση κατά το χρόνο δοκιμής στα σύγχρονα συστήματα ανίχνευσης. Οι μέθοδοι των Region Proposal βασίζονται συνήθως σε οικονομικά χαρακτηριστικά και σχήματα εξαγωγής συμπερασμάτων. Η επιλεκτική αναζήτηση, μία από τις πιο δημοφιλείς μεθόδους, συγχωνεύει superpixels με βάση μηχανικά χαρακτηριστικά χαμηλού επιπέδου. Ωστόσο, σε σύγκριση με αποδοτικά δίκτυα ανίχνευσης, η επιλεκτική αναζήτηση είναι μια τάξη μεγέθους πιο αργή, σε 2 δευτερόλεπτα ανά εικόνα σε μια υλοποίηση Κεντρικής Μονάδας Επεξεργασίας (CPU). Τα EdgeBoxes παρέχει επί του παρόντος την καλύτερο συμβιβασμό μεταξύ ποιότητας και ταχύτητας των Proposals, σε 0,2 δευτερόλεπτα ανά εικόνα. Παρόλ' αυτα, τα Region Proposals εξακολουθούν να καταναλώνουν τόσο πολύ χρόνο εκτέλεσης όσο το δίκτυο ανίχνευσης. Σημειώνεται πως τα γρήγορα region-based CNNs εκμετταλεύονται την κάρτα γραφικών (GPU), ενώ οι μέθοδοι των region proposal οι οποίοι χρησιμοποιούνται για έρευνα αξιοποιούνται σε επεξεργαστές (CPU), κάνοντας έτσι τον χρόνο λειτουργίας μη συγκρίσιμο. Ένας εμφανής τρόπος για την επιτάχυνση των υπολογισμών των proposals είναι η επανεφαρμογή τους σε GPU. Αυτή είναι μια αποτελεσματική λύση, άλλα η επανεφαρμογή αγνοεί το down-stream του δικτύου ανίχνευσης και επομένως χάνει σημαντικές ευκαιρίες για το μοίρασμα των υπολογισμών.

3.2 Faster R-CNN

Το σύστημα για Ανίχνευση Αντικειμένων που χρησιμοποιήθηκε ονομάζεται Faster R-CNN και αποτελείται από 2 ενότητες. Η 1^{η} ενότητα είναι ένα deep fully convolutional network που προτείνει περιοχές και η 2^{η} είανι το Fast R-CNN που χρησιμοποιεί τις προτεινόμενες περιοχές. Ολόκληρο το σύστημα είναι ένα ενοποιημένο δίκτυο για Ανίχνευση Αντικειμένων. Χρησιμοποιόντας της πρόσφατα

δημοφιλή ορολογία νευρωνικά δίκτυα με μηχανισμούς “προσοχής”, η ενότητα του RPN καθοδηγεί την ενότητα του Fast R-CNN προς τα που να “κοιτάξει”.



***Εικόνα 3.1**: Faster R-CNN ένα ενιαίο, ενοποιημένο δίκτυο για Ανίχνευση Αντικειμένων. Η ενότητα του RPN χρησιμοποιείται ώς η “προσοχή” του ενοποιημένου δικτύου. [10]*

3.3 Region Proposal Networks

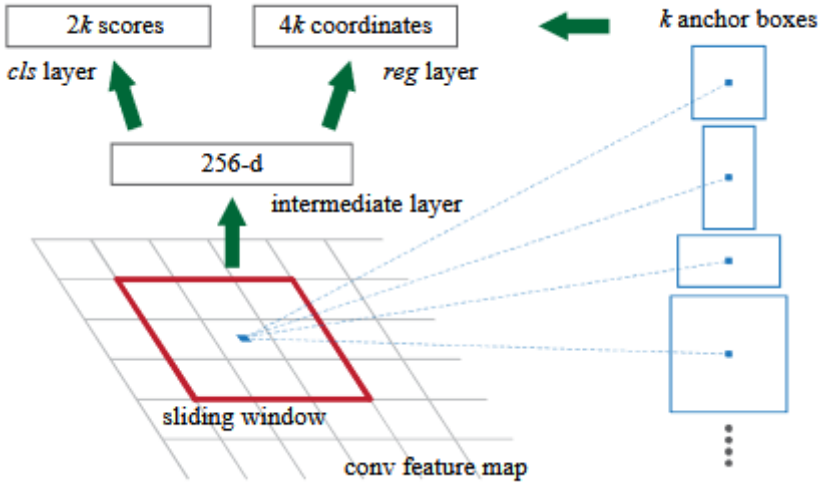
Ένα Δίκτυο Προτάσεων Περιοχής (Region Proposal Network (RPN)) υποδέχεται μια εικόνα (οποιασδήποτε διάστασης) και δημιουργεί ένα σετ παραλληλογράμμων προτάσεων αντικειμένων, όπου το καθένα έχει μια τιμή, η οποία μετρά τη συμμετοχή σε ένα σύνολο κλάσεων αντικειμένων έναντι του φόντου. Η διαδικασία αυτή μοντελοποιείται με βάση ένα ένα ολοκληρομένο convolution network. Καθώς ο τελικός στόχος είναι να μοιραστούν οι υπολογισμοί με ένα δίκτυο Fast R-CNN, γίνεται η υπόθεση πως και τα δύο δίκτυα μοιράζονται ένα κοινό σετ από συνελεκτικών επιπέδων.

Για την παραγωγή των προτάσεων περιοχής, χρησιμοποιείται ένα μικρό δίκτυο πάνω από την έξοδο του συνελικτικού χάρτη χαρακτηριστικών από το τελευταίο κοινόχρηστο επίπεδο συνελίξεων. Αυτό το μικρό δίκτυο δέχεται ως είσοδο ένα χωρικό παράθυρο n × n

του συνελικτικού χάρτη χαρακτηριστικών εισόδου. Κάθε παράθυρο αντιστοιχίζεται σε ένα χαμηλότερης διάστασης χαρακτηριστικό (256-d για το ZF και 512-d για το VGG). Το χαρακτηριστικό αυτό τροφοδοτείται σε δύο αδελφικά πλήρως συνδεδεμένα στρώματα, ένα box-regression (reg) και ένα box-classification (cls) στρώμα. Το μίνι-δίκτυο λειτουργεί με συρόμενο παράθυρο τρόπο, τα πλήρως συνδεδεμένα στρώματα μοιράζονται σε όλες τις θέσεις του χώρου. Αυτή η αρχιτεκτονική εφαρμόζεται φυσικά με ένα n × n συνελικτικό στρώμα που ακολουθείται από δύο αδελφικά στρώματα 1 × 1 συνελικτικού τύπου (για reg και cls, αντίστοιχα).

3.3.1 Άγκυρες

Σε κάθε θέση παραθύρου, προβλέπονται ταυτόχρονα πολλαπλές προτάσεις περιοχής, όπου ο αριθμός των μέγιστων πιθανών προτάσεων για κάθε περιοχή συμβολίζεται με k. Επομένως, το reg επίπεδο έχει 4k outputs κωδικοποιώντας τις συντεταγμένες k πλαισίων, και το cls επίπεδο έχει 2k outputs τιμές που εκτιμούν την πιθανότητα να υπάρχει ή να μην υπάρχει αντικείμενο. Οι k προτάσεις παραμετροποιούνται σχετικά με το k των πλαισίων αναφοράς τα οποία ονομάζονται Άγκυρες (Anchors). Ένα anchor έχει ως κέντρο το κέντρο του παραθύρου στο οποίο ανήκει και έχει μια συγκεκριμένη κλίμακα και έναν λόγο διαστάσεων. Η προεπιλογή χρησιμοποιεί 3 κλίμακες και 3 λόγους διαστάσεων, αποδίδοντας έτσι k=9 anchors σε κάθε θέση, για ένα χάρτη χαρακτηριστικών με μεγέθη W x H, υπάρχουν W x H x k anchors συνολικά.



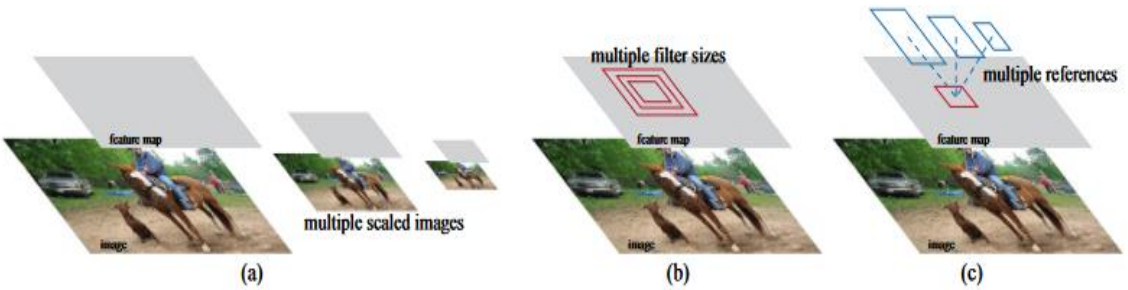
Εικόνα 3.2: *Δίκτυο Προτάσεων Περιοχής - Region Proposal Network (RPN) [10]*

Μεταφραστικά Αναλλοίωτες Άγκυρες

Η σημαντική αυτή ιδιότητα είναι αναγκαία τόσο για τις Άγκυρες (anchors) όσο και για τις συναρτήσεις που υπολογίζουν τις προτάσεις σχετικά με τα anchors. Εάν ένα anchor μεταφράσει ένα αντικείμενο στην εικόνα, η πρόταση πρέπει να μεταφραστεί και η ίδια συνάρτηση πρέπει να είναι σε θέση να προβλέψει την πρόταση σε οποιδήποτε σημείο. Η λειτουργία της αναλλοίωτης μεταφραστικότητας είναι σίγουρη στην περίπτωση σου Faster R-CNN.

Άγκυρες πολλαπλών κλιμάκων ώς αναφορές παλινδρόμησης

Ο σχεδιασμός των anchors όπως παρουσιάστηκε παραπάνω παρουσιάζει ένα νέο σύστημα για την προσέγγιση των πολλαπλών κλιμάκων και των λόγων διαστάσεων. Όπως παρουσιάζεται στην *Εικόνα 3.3* υπάρχουν 2 δημοφιλείς τρόποι για πολυ-κλιμακωτές προβλέψεις. Ενας τρόπος βασίζεται στο χαρακτηριστικό της πυραμίδας και στις μεθόδους Συνελεκτικών Νευρωνικών Δικτύων. Οι εικόνες αλλάζουν μέγεθος μέσα από πολλαπλές κλίμακες και οι πίνακες χαρακτηριστικών (ή/και βαθιά χαρακτηριστικά περίπλεξης) υπολογίζονται για κάθε διαφορετική κλίμακα (*Εικόνα 3.3(α)*). Αυτός ο τρόπος είναι πολύ χρήσιμος αλλά και πολύ χρονοβόρος. Ο δεύτερος τρόπος είναι να χρησιμοποιήθουν παράθυρα διαφορετικών κλιμάκων πάνω σε χάρτες χαρακτηριστικών. Για παράδειγμα, μοντέλα διαφορετικών λόγων διαστάσεων εκπαιδεύονται ξεχωριστά χρησιμοποιώντας διαφορετικά μεγέθη φίλτρων (5 x 7 και 7 x 5). Εαν αυτός ο τρόπος χρησιμοποιηθεί για να προσεγγίσει διαφορετικές κλίμακες μπορεί να θεωρηθεί και ώς μια πυραμίδα από φίλτρα (*Εικόνα 3.3(β)*). Συνήθως και οι 2 μέθοδοι χρησιμοποιούνται ταυτόχρονα.



Εικόνα 3.3: *Διαφορετικά σχέδια που αφορούν διαφορετικές κλίμακες και μεγέθη. (α) Πυραμίδα εικόνων και χαρτών χαρακτηριστικών και ο ταξινομητής τρέχει για όλες τις κλίμακες. (β) Πυραμίδα φίλτρων με πολλαπλές κλίμακες/μεγέθη τρέχουν στο feature map. (γ) Χρήση πυραμίδων πλαισίων αναφοράς σε συναρτήσεις παλινδρόμησης [10]*

Για σύγκριση η μέθοδος που βασίζεται στα anchors χτίζεται πάνω σε μια πυραμίδα από άγκυρες, που είναι οικονομικό. Η μέθοδος αυτή κατηγοριοποιεί και επιστρέφει πλαίσιο δέσμευσης, με σημείο αναφοράς την άγκυρα του κάθε πλαισίου για κάθε διαφορετική κλίμακα και λόγο διαστάσεων. Βασίζεται μόνο σε εικόνες και χάρτες χαρακτηριστικών μιας κλίμακας και χρησιμοποιεί φίλτρα μιας κλίμακας. Λόγω αυτού του πολυκλιμακωτού σχεδιασμού, μπορεί να γίνει χρήση των συνελεκτικών χαρακτηριστικών που έχουν υπολογιστεί από την μονοκλιμακωτή εικόνα όπως και στο Fast R-CNN. Ο σχεδιασμός των πολυκλιμακωτών αγκυρών είναι σημαντικό κόμματι για την μοιρασιά χαρακτηριστικών χωρίς έξτρα κόστος επειδή χρησιμοποιήθηκαν κλίμακες.

3.3.2 Συνάρτηση Απώλειας

Για την εκπαίδευση των Δίκτυο Προτάσεων Περιοχής (RPN), αναθέτεται μια ετικέτα δυαδικής κλάσης σε κάθε anchor για το αν είναι αντικείμενο ή όχι. Αναθέτεται μια θετική ετικέτα σε δύο είδη αγκυρών (anchors): α) στα anchors με την μεγαλύτερη τιμή Διατομής πάνω απ'την Ένωση (Intersection-over-Union (IoU)) που αλληλοκαλύπτονται με ένα πλαίσιο "αλήθειας" (ground-truth box), ή β) μια άγκυρα που έχει IoU μεγαλύτερο από 0.7 με κάποιο ground-truth box. Σημειώνεται πως ένα ground-truth box μπορεί να αναθέσει θετική ετικέτα σε πολλαπλά anchors. Συνήθως η β) κατάσταση είναι αρκετή για να προσδιορίσει τα θετικά δείγματα, αλλά και πάλι χρησιμοποιείται η α) κατάσταση για τον λόγο ότι σε πολύ σπάνιες περιπτώσεις η κατάσταση β) μπορεί να μην βρεί θετικά δείγματα. Αρνητική ετικέτα εισάγεται στα anchor που έχουν αναλογία IoU μικρότερη από 0.3 με όλα τα ground-truth boxes. Τα anchors που δεν είναι ούτε θετικά ούτε αρνητικά δεν συμβάλουν στην εκπαίδευση . Επομένως η συνάρτηση απώλειας (Loss Function) για μια εικόνα υπολογίζεται ώς εξής:

$$L(\{p_i\},\{t_i\}) = \frac{1}{N_{cls}}\sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}}\sum_i p_i^* L_{reg}(t_i, t_i^*) \quad (1)$$

Εδώ i είναι ο δείκτης ενός anchor σε ένα mini-batch και p_i είναι η προβλεπόμενη πιθανότητα ενός anchor να είναι αντικείμενο. Η ετικέτου ground-truth p_i^* είναι 1 αν το anchor είναι αρνητικο και 0 αν είναι θετικό. t_i είναι το διάνυσμα που εκπροσωπεί τις 4 παραμετροποιημένες συντεταγμένες των προβλεπόμενων bounding box και t_i^* είναι αυτές των ground-truth box που αφορούν τα θετικά anchors. Η κατηγοριοποίηση της απώλειας L_{cls} είναι λογαριθμική απώλεια σε δύο κατηγορίες (αντικείμενο και όχι αντικείμενο). Για την

επιστρεφόμενη απώλεια χρησιμοποιείται $L_{reg}(t_i, t_i^*) = R(t_i - t_i^*)$ που το R είναι η συνάρτηση απώλειας που παρουσιάζεται στο paper [2]. Ο όρος $p_i^* L_{reg}$ σημαίνει πως η επιστρεφόμενη απώλεια ενεργοποιείται μόνο σε θετικά anchors αλλιώς απενργοποιείται. Τα outputs του cls και reg περιέχουν $\{p_i\}$ και $\{t_i\}$ αντίστοιχα. Οι 2 όροι κανονικοποιούνται από N_{cls} και N_{reg} και παίρνουν βάρη από το ισορροπιστικό παράγοντα λ. Στους συγκεκριμένους υπολογισμούν το N_{cls} = 256 και το reg κανονικοποιείται από τον αριθμό των περιοχών των anchors. Προεπιλεγμένα το λ = 10 επομένως οι όροι cls και το reg έχουν σχεδόν τα ίδια βάρη. Στον *Πίνακα X* παρουσιάζονται παραδείγματα με διαφορετικές τιμές λ.

Για το regression των bounding box χρησιμοποιούνται οι παρακάτω παραμετρικές τιμές των συντεταγμένων:

$$
\begin{aligned}
t_x &= \frac{x - x_a}{w_a}, \qquad t_y = \frac{y - y_a}{h_a} \\
t_w &= \log\left(\frac{w}{w_a}\right), \qquad t_h = \log(\frac{h}{h_a}) \\
t_x^* &= \frac{x^* - x_a}{w_a}, \qquad t_y^* = \frac{x^* - y_a}{h_a} \\
t_x^* &= \log(\frac{w^*}{w_a}), \qquad t_y^* = \log(\frac{h^*}{h_a})
\end{aligned}
\tag{2}
$$

Όπου x, y, w και h οι συντεταγμένες του κέντρου και το μήκος και το πλάτος. Οι μεταβλητές x, x_a και x^* είναι για τα προβλέψιμα πλαίσια, anchor box και ground-truth box αντίστοιχα (παρομοίως για y, w, h). Αυτό μπορεί να θεωρηθεί σαν μια επιστροφή ενός bounding-box σε anchor-box και στην συνέχεια σε ένα διπλανό ground-truth box.

Ωστόσο, η μέθοδος αυτή πετυχαίνει διαφορετική επιστροφή πλαισίων οριοθέτησης (bounding-box) από αυτή που χρησιμοποιεί η μέθοδος Περιοχών Ενδιαφέροντος (RoI-Regions of Interest) όπου η επιστροφή των πλαισίων οριοθέτησης πραγματοποιείται σε χαρακτηριστικά που εξάγωνται από RoI αυθαίρετων μεγεθών και τα βάρη της επιστροφής μοιράζονται σε όλες τις περιοχές ανεξαρτήτως μεγέθους.

3.3.3 Εκπαίδευση Δικτύου Προτάσεων Περιοχής

Τα Δίκτυα Προτάσεων Περιοχής (Region Proposal Networks – RPN) μπορούν να εκπαιδευτούν από την αρχή ως το τέλος από αντίθετη αναπαραγωγή και στοχαστική κάθοδο κλίσης (Stochastic Gradient Descent SGD). Ακολουθείται η εικονοκεντρική στρατηγική για την εκπαίδευση του δικτύου. Κάθε mini-batch που εξάγεται από μια εικόνα που περιέχει αρκετά θετικά και αρνητικά παραδείγματα anchors. Είναι δυνατόν να γίνει βελτιστοποίηση για την συνάρτηση απώλειας για όλα τα anchors αλλά θα κλίνει προς τα αρνητικά δείγματα καθώς είναι πιο κυρίαρχα. Αντιθέτως, χρησιμοποιούνται 256 τυχαία anchors από την εικόνα για να υπολογιστεί η συνάρτηση απώλειας του mini-batch καθώς τα θετικά και τα αρνητικά anchors έχουν αναλογία 1:1. Αν υπάρχουν λιγότερα από 128 θετικά δείγματα σε μια εικόνα τότε προσθέτονται στο mini-batch αρνητικά.

Τυχαία ενεργοποιούνται όλα τα νέα επίπεδα μετά από την σχεδίαση των βαρών με μια Γκαουσιανή κατανομή zero-mean με τυπική απόκλιση 0.01. Όλα τα υπόλοιπα επίπεδα ενεργοποιούνται με προεκπαιδευμένα μοντέλα για κατηγοριοποίηση ImageNet καθώς είναι βασική εκπαίδευση. Όλα τα επίπεδα του ZF net για να εξοικονωμηθεί μνήμη στο VGG net. Χρησιμοποιείται δείκτης μάθησης 0.001 για 60k mini-batches και 0.0001 για τα επόμενα 20k.

3.4 Κοινά χαρακτηριστικά των RPN & Fast R-CNN

Μέχρι τώρα έχει περιγραφεί η εκπαίδευση ενός δικτύου για RPN, χωρίς να υπολογίζεται το region-based object detection CNN που θα χρησιμοποιήσει αυτές τις προτάσεις. Για το επόμενο δίκτυο αναγνώρισης θα χρησιμοποιηθεί το Fast R-CNN, καθώς χρειάζεται να περιγραφεί ο αλγόριθμος που μαθαίνει ένα ενοποιημένο δίκτυο απο RPN και Fast R-CNN με μοιρασμένα convolution layers *Εικόνα 3.1*.

Τόσο τα RPN και το Fast R-CNN, αν εκπαιδευτούν ξεχωριστά, θα τροποποιήσουν τα convolutional layers τους με διαφορετικούς τρόπους. Επομένως, χρειάζεται μια τεχνική για να επιτρέψει και στα 2 δίκτυα να μοιραστούν τα convolutional layers. Υπάρχουν 4 τρόποι για να επιτευχθεί αυτό:

- **Εναλλασόμενη εκπαίδευση**. Με αυτή την λύση εκπαιδεύεται για αρχή το RPN και με τις προτάσεις εκπαιδεύεται στην συνέχεια το Fast R-CNN. Το δίκτυο που συντονίζεται από το Fast R-CNN χρησιμοποιείται για την εκκίνηση του RPN,

- **Προσεγγιστική κοινή εκπαίδευση**. Με αυτή την λύση το RPN και το Fast R-CNN ενοποιούνται κατα την διάρκεια της εκπαίδευσης όπως στην *Εικόνα 3.1*. Σε κάθε χρήση SGD δημιουργεί region proposals και στην συνέχεια ακολουθεί η οπισθοδρομική διάδοση όπου όλα τα κοινά επίπεδα από την απώλεια του RPN και του Fast R-CNN συνδυάζονται. Είναι η πιο εύκολη λύση για να εφαρμοστεί αλλά αγνοεί τις παραγώγους που δίνουν τις συντεταγμένες των proposal boxes. Η εμπειρική μέθοδος έχει δείξει πως τα αποτελέσματα είναι πολύ κοντινά και μειώνει και τον χρόνο εκπαίδευσης κατα 25-50% συγκριτικά με την Εναλλασόμενη Εκπαίδευση.
- **Μη προσεγγιστική κοινή εκπαίδευση**. Όπως αναφέρθηκε παραπάνω, τα προβλεπόμενα Bounding boxes του RPN είναι και συναρτήσεις των εισόδων. Το RoI επίπεδο στο Fast R-CNN δέχεται τα convolutional χαρακτηριστικά και τα προβλεπόμενα bounding boxes ως είσοδο, επομένως μια θεωριτικά έγκυρη backpropagation λύση είναι αναγκαίο να περιλαμβάνει τις συντεταγμένες του πλαισίου. Αυτές οι κλίσεις αγνοούνται στην παραπάνω λύση. Σε μια μη προσεγγιστική κοινή εκπαίδευση είναι αναγκαία η ύπαρξη RoI επιπέδου που είναι διαφοροποιήσιμο από τις συντεταγμένες του πλαισίου. Υπάρχει λύση σε αυτό το πρόβλημα αλλά δεν παρουσιάζεται σε αυτή την διπλωματική εργασία.
- **Εναλλασόμενη εκπαίδευση 4 – βημάτων.** Είναι η χρησιμοποιημένη μέθοδος. Το 1ο βήμα είναι η εκπαίδευση ενός RPN. Αυτό το δίκτυο ενεργοποιείται με ενα προεκπαιδευμένο ImageNet μοντέλο και ειναι λεπτομερώς ρυθμισμένο από την αρχή εώς το τέλος. Στην συνέχεια εκπαιδεύεται ξεχωριστά ένα δίκτυο αναγνώρισης από το Fast R-CNN χρησιμοποιώντας τις προτάσεις του βήματος 1. Σε αυτό το σημείο τα 2 δίκτυο δεν μοιράζονται convolutional επίπεδα Στο 3ο βήμα χρησιμοποιείται το δίκτυο εντοπισμού για να εκκινήσει την εκπαίδευση του RPN αλλά διορθώνονται μόνο τα μοιρασμένα convolutional επίπεδα και ρυθμίζονται λεπτομερώς μόνο τα επιπέδα που είναι αποκλειστικά στο RPN. Τώρα τα 2 δίκτυα έχουν κοινά convolutional επίπεδα τα οποία διατηρούνται και σταθερά για να ρυθμιστούν και τα επίπεδα του Fast R-CNN. Τώρα πλέον τα 2 δίκτυα λειτουργούν ως ένα ενοποιημένο δίκτυο.

3.5 Πείραμα Faster R-CNN

Αξιολογείται διεξοδικά η μέθοδό στο PASCAL VOC 2007. Το σύνολο δεδομένων αποτελείται από περίπου 5000 εικόνες και 5000 εικόνες δοκιμής σε 20 κατηγορίες αντικειμένων. Για το προ-εκπαιδευμένο δίκτυο ImageNet, χρησιμοποιείται η "γρήγορη" έκδοση του δικτύου ZF που έχει 5 συνελικτικά στρώματα και 3 πλήρως συνδεδεμένα στρώματα, και το δημόσιο μοντέλο VGG-167 που έχει 13 συνελικτικά συγκεραστικά στρώματα και 3 πλήρως συνδεδεμένα στρώματα. Αξιολογείται πρωτίστως ο μέσος όρος ανίχνευσης Μέσης ακρίβειας (mAP), επειδή αυτή είναι η πραγματική μετρική για την ανίχνευση αντικειμένων ανίχνευση αντικειμένων. Ο μέσος όρος ανίχνευσης Μέσης Ακρίβειας υπολογίζεται από τον τύπο:

$$Precision = \frac{TP}{TP + FP}$$

Η *Εικόνα 3.5.2* δείχνει τα αποτελέσματα του Fast R-CNN όταν εκπαιδεύτηκαν και δοκιμάστηκαν χρησιμοποιώντας διάφορες προτάσεις περιοχής μεθόδους. Αυτά τα αποτελέσματα χρησιμοποιούν το δίκτυο ZF. Για την επιλεκτική αναζήτηση (SS), δημιουργήθηκαν περίπου 2000 προτάσεις με τη "γρήγορη" λειτουργία. Για τα EdgeBoxes (EB), παράγονται οι προτάσεις με την προεπιλεγμένη ρύθμιση EB που συντονίζεται για 0,7 7 IoU. Η SS έχει mAP 58,7% και η EB έχει mAP του 58,6% στο πλαίσιο του Fast R-CNN. Το RPN με Fast R-CNN επιτυγχάνει ανταγωνιστικά αποτελέσματα, με mAP 59,9%, ενώ χρησιμοποιεί έως και 300 προτάσεις. Η χρήση του RPN αποδίδει ένα πολύ ταχύτερο σύστημα ανίχνευσης χρησιμοποιώντας είτε SS ή EB λόγω των κοινών convolutional υπολογισμών. Οι λιγότερες προτάσεις μειώνουν επίσης το κόστος των πλήρως συνδεδεμένων στρωμάτων ανά περιοχή όπως φαίνεται στην *Εικόνα 3.5.1*.

model	system	conv	proposal	region-wise	total	rate
VGG	SS + Fast R-CNN	146	1510	174	1830	0.5 fps
VGG	RPN + Fast R-CNN	141	**10**	47	**198**	**5 fps**
ZF	RPN + Fast R-CNN	31	**3**	25	**59**	**17 fps**

Εικόνα 3.5.1: *Χρόνος λειτουργίας και αποτελέσματα PASCAL VOC 2007 με Fast R-CNN [10]*

train-time region proposals		test-time region proposals		
method	# boxes	method	# proposals	mAP (%)
SS	2000	SS	2000	58.7
EB	2000	EB	2000	58.6
RPN+ZF, shared	2000	RPN+ZF, shared	300	**59.9**
ablation experiments follow below				
RPN+ZF, unshared	2000	RPN+ZF, unshared	300	58.7
SS	2000	RPN+ZF	100	55.1
SS	2000	RPN+ZF	300	56.8
SS	2000	RPN+ZF	1000	56.3
SS	2000	RPN+ZF (no NMS)	6000	55.2
SS	2000	RPN+ZF (no *cls*)	100	44.6
SS	2000	RPN+ZF (no *cls*)	300	51.4
SS	2000	RPN+ZF (no *cls*)	1000	55.8
SS	2000	RPN+ZF (no *reg*)	300	52.1
SS	2000	RPN+ZF (no *reg*)	1000	51.3
SS	2000	RPN+VGG	300	59.2

Εικόνα 3.5.2: *Αποτελέσματα Ανίχνευσης PASCAL VOC 2007. Οι ανιχνευτές είναι Fast R-CNN με ZF αλλα χρησιμοποιώντας διαφορετικές μεθόδους για εκπαίδευσης και εξέταση* ***[10]***

Πειράματα απόσβεσης RPN. Για την διερεύνηση της συμπεριφοράς των RPN ως μέθοδο πρότασης, πραγματοποιήθηκαν διάφορες μελέτες εκτομής. Πρώτον, παρουσιάζεται η επίδραση του διαμοιρασμού των στρωμάτων συνελικτικού συστήματος μεταξύ των RPN και του δικτύου ανίχνευσης Fast R-CNN. Για να γίνει αυτό, σταμάτησε η διαδικασία μετά το δεύτερο βήμα στη διαδικασία εκπαίδευσης 4 βημάτων. Η χρήση ξεχωριστών δικτύων μειώνει ελαφρώς το αποτέλεσμα σε 58,7% (RPN+ZF, χωρίς κοινή χρήση (*Εικόνα 3.5.2)*. Παρατηρείται ότι αυτό οφείλεται στο γεγονός ότι στο τρίτο βήμα, όταν τα χαρακτηριστικα εντοπισμού χρησιμοποιούνται για τη λεπτομερή ρύθμιση του RPN, η ποιότητα της πρότασης βελτιώνεται.

Στη συνέχεια, διαχωρίζεται η επιρροή του RPN στην εκπαίδευση του δικτύου ανίχνευσης Fast R-CNN. Για το σκοπό αυτό, εκπαιδεύεται ένα μοντέλο Fast R-CNN χρησιμοποιώντας 2000 προτάσεις SS και το δίκτυο ZF. Καθορίζεται ο ανιχνευτής και αξιολογείται το mAP ανίχνευσης αλλάζοντας τις περιοχές προτάσεων που χρησιμοποιούνται κατά τη διάρκεια της δοκιμής. Σε αυτά τα καταλυτικά πειράματα, το RPN δεν μοιράζεται χαρακτηριστικά με τον με τον ανιχνευτή. Αντικατάσταση της SS με 300 προτάσεις RPN κατά το χρόνο δοκιμής οδηγεί σε mAP 56,8%. Η απώλεια στο mAP οφείλεται στο γεγονός της ασυνέπειας μεταξύ των προτάσεων εκπαίδευσης/ελέγχου. Αυτό το αποτέλεσμα χρησιμεύει ως βάση για τις ακόλουθες συγκρίσεις.

Παραδόξως, το RPN εξακολουθεί να οδηγεί σε ένα ανταγωνιστικό αποτέλεσμα (55,1%) όταν χρησιμοποιεί τις 100 προτάσεις με την κορυφαία κατάταξη κατά τη διάρκεια της δοκιμής, υποδεικνύοντας ότι η κορυφαία καταταγμένες προτάσεις RPN είναι ακριβείς. Στο άλλο άκρο, η χρήση των 6000 προτάσεων RPN με την υψηλότερη κατάταξη (χωρίς NMS) έχει συγκρίσιμο mAP (55,2%), υποθέτοντας ότι το NMS δεν βλάπτει το mAP ανίχνευσης και μπορεί να μειώσει τους ψευδείς συναγερμούς.

Στη συνέχεια, διερευνούνται ξεχωριστά οι ρόλοι των εξόδων των RPN's cls και reg απενεργοποιώντας οποιαδήποτε από αυτούς κατά τη διάρκεια της δοκιμής. Όταν το στρώμα cls αφαιρείται κατά τη δοκιμής (επομένως δεν χρησιμοποιείται κανένα NMS/ranking), δημιουργούνται τυχαίες δειγματοληπτικές N προτάσεις από τις περιοχές χωρίς βαθμολογία. Το mAP είναι σχεδόν αμετάβλητο με N = 1000 (55,8%), αλλά υποβαθμίζεται σημαντικά στο 44,6% όταν N = 100. Αυτό δείχνει ότι οι βαθμολογίες cls ευθύνονται για την ακρίβεια των των προτάσεων με την υψηλότερη κατάταξη. Από την άλλη πλευρά, όταν αφαιρείται το στρώμα reg κατά τη διάρκεια της δοκιμής (οπότε οι προτάσεις γίνονται πλαίσια αγκύρωσης), το mAP πέφτει στο 52,1%. Αυτό υποδηλώνει ότι η υψηλή ποιότητας προτάσεις οφείλονται κυρίως στο παλινδρομημένο bounding box. Τα anchor boxes, αν και έχουν πολλαπλές κλίμακες και αναλογίες διαστάσεων, δεν επαρκούν για ακριβή ανίχνευση.

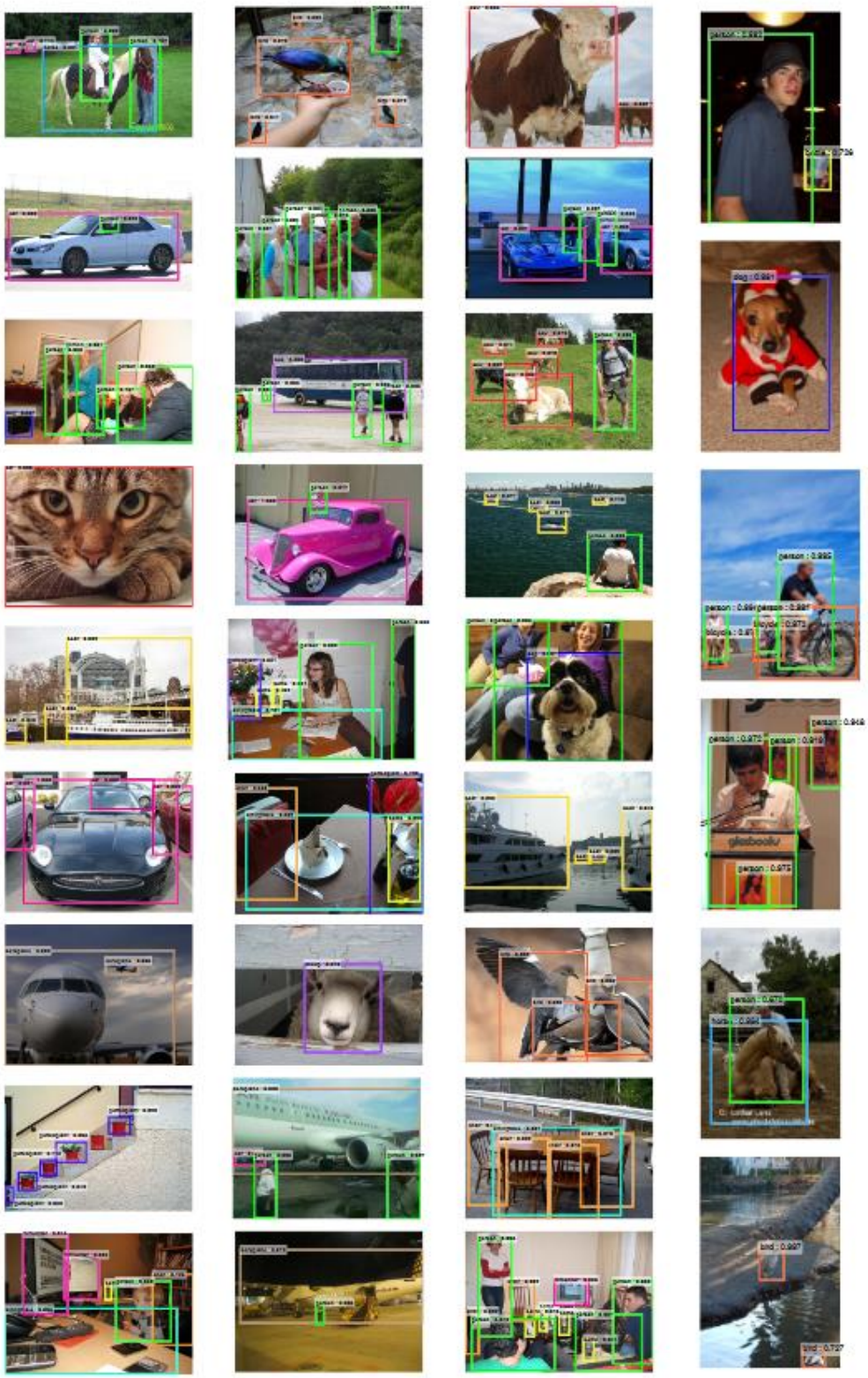
Αξιολογούνται επίσης οι επιδράσεις ισχυρότερων δικτύων στην ποιότητα των προτάσεων μόνο του RPN. Χρησιμοποιείται VGG-16 για να εκπαιδευτεί το RPN, και εξακολουθούν να χρησιμοποιούνται οι παραπάνω ανιχνευτές του SS+ZF. Η mAP βελτιώνεται από 56,8% (χρησιμοποιώντας RPN+ZF) σε 59,2% (χρησιμοποιώντας RPN+VGG). Αυτό είναι ένα ελπιδοφόρο αποτέλεσμα, διότι υποδηλώνει ότι η πρόταση ποιότητα της RPN+VGG είναι καλύτερη από εκείνη της RPN+ZF. Επειδή οι προτάσεις της RPN+ZF είναι ανταγωνιστικές με SS (και οι δύο είναι 58,7% όταν χρησιμοποιούνται σταθερά για εκπαίδευση και δοκιμές), αναμένεται ότι η RPN+VGG θα είναι καλύτερη από την SS. Τα ακόλουθα πειράματα δικαιολογούν αυτή την υπόθεση.

Απόδοση του VGG-16. Η *Εικόνα 3.5.3* παρουσιάζει τα αποτελέσματα του VGG-16 τόσο για την πρόταση όσο και για την ανίχνευση. Στην χρήση του RPN+VGG, το αποτέλεσμα είναι 68,5% για μη κοινά χαρακτηριστικά, ελαφρώς υψηλότερο από τη βασική γραμμή SS. Όπως φαίνεται παρακάτω, αυτό οφείλεται στο γεγονός ότι οι προτάσεις που παράγονται από την RPN+VGG είναι πιο ακριβείς

από τις SS. Σε αντίθεση με το SS που είναι προκαθορισμένο, το RPN εκπαιδεύεται ενεργά και επωφελείται από καλύτερα δίκτυα. Για την παραλλαγή με κοινόχρηστα χαρακτηριστικά, το αποτέλεσμα είναι 69,9% καλύτερο από την ισχυρή βασική γραμμή SS, ωστόσο με προτάσεις σχεδόν χωρίς κόστος. Εκπαιδεύεται περαιτέρω το RPN και το δίκτυο ανίχνευσης στο σύνολο της ένωσης των PASCAL VOC 2007 και 2012. Το mAP είναι 73,2%. Στην *Εικόνα 3.5.6* παρουσιάζονται ορισμένα αποτελέσματα στο σύνολο PASCAL VOC 2007. Στο σύνολο δοκιμών PASCAL VOC 2012 (*Εικόνα 3.5.3*), η μέθοδός έχει εκπαιδευμένο mAP 70,4%. στο σύνολο ένωσης των συνόλων VOC 2007 και VOC 2012.

method	# proposals	data	mAP (%)
SS	2000	12	65.7
SS	2000	07++12	68.4
RPN+VGG, shared†	300	12	67.0
RPN+VGG, shared‡	300	07++12	**70.4**
RPN+VGG, shared§	300	COCO+07++12	**75.9**

Εικόνα 3.5.3: *Αποτελέσματα ανίχνευσης στο σύνολο δοκιμών PASCAL VOC 2012. Ο ανιχνευτής είναι Fast R-CNN και VGG-16* ***[10]***



Εικόνα 3.5.6: *Επιλεγμένα παραδείγματα αποτελεσμάτων ανίχνευσης αντικειμένων στο σύνολο δοκιμών PASCAL VOC 2007 με τη χρήση του Faster R-CNN. Το μοντέλο είναι το VGG-16 και τα δεδομένα εκπαίδευσης είναι τα PASCAL VOC 2007 + 2012 (73,2% mAP στη δοκιμή 2007 σύνολο).* ***[10]***

Κεφάλαιο 4: Εφαρμογές Ανίχνευσης Αντικειμένων σε Ιατρικές Εικόνες

Στο κεφάλαιο αυτό αναφέρεται η χρήση της μεθόδου της Ανίχνευσης Αντικειμένων στις Ιατρικές εικόνες. Παρουσιάζονται ορισμένες εφαρμογές στον Ιατρικό τομέα και τέλος αναλύεται το κίνητρο και η επιλογή της ανάλυσης των αξονικών τομογραφιών που περιέχουν όγκο.

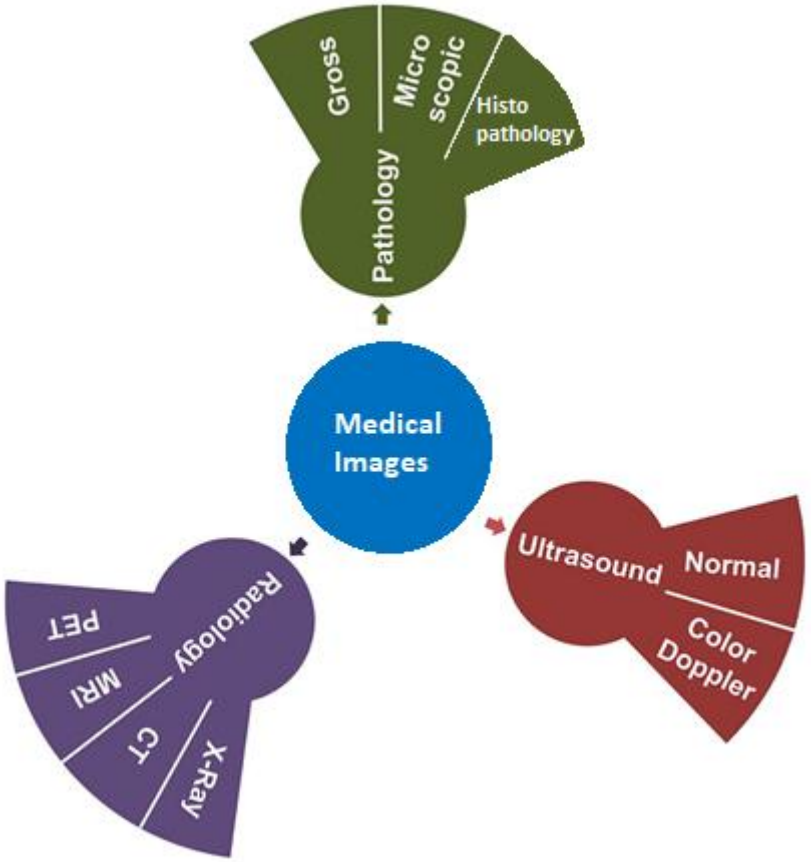
4.1 Εισαγωγή

Στην εποχή της ψηφιακής ιατρικής, παράγεται καθημερινά ένας τεράστιος αριθμός ιατρικών εικόνων. Υπάρχει μεγάλη ζήτηση για έξυπνο εξοπλισμό για συμπληρωματική διάγνωση που θα βοηθάει τους ιατρούς διαφόρων ειδικοτήτων. Με την ανάπτυξη της τεχνητής νοημοσύνης, οι αλγόριθμοι του συνελικτικού νευρωνικού δικτύου (CNN) προχώρησαν με ταχείς ρυθμούς. Το CNN και οι αλγόριθμοι επέκτασής του διαδραματίζουν σημαντικό ρόλο στην ταξινόμηση ιατρικών εικόνων, στην ανίχνευση αντικειμένων και στη σημασιολογική κατάτμηση. Ενώ η ταξινόμηση της ιατρικής απεικόνισης έχει αναφερθεί ευρέως, η ανίχνευση αντικειμένων και η σημασιολογική τμηματοποίηση της απεικόνισης σπάνια περιγράφονται. Στην συνέχεια παρουσιάζεται η εξέλιξη της ανίχνευσης αντικειμένων και της σημασιολογικής τμηματοποίησης στη μελέτη της ιατρικής απεικόνισης. Αναφέρεται επίσης ο ακριβς προσδιορισμός της θέσης και των ορίων των ασθενειών.

Στην ιατρική πρακτική ρουτίνας, παράγεται μεγάλος αριθμός ιατρικών εικόνων κατά τη διαδικασία διαφόρων εξετάσεων, όπως η ακτινολογία, οι υπέρηχοι, η ενδοσκόπηση, η οφθαλμολογία και η παθολογία. Οι εικόνες ακτινοβολίας περιλαμβάνουν τις ακτίνες X, την υπολογιστική τομογραφία (CT), τη μαγνητική τομογραφία (MRI) και την τομογραφία εκπομπής ποζιτρονίων-υπολογιστική τομογραφία (PET-CT). Οι εικόνες υπερήχων περιλαμβάνουν φυσιολογικές εικόνες υπερήχων και έγχρωμες εικόνες υπερήχων Doppler. Οι ενδοσκοπικές εικόνες περιλαμβάνουν ενδοσκόπηση λευκού φωτός (WLE), χρωμοενδοσκόπηση (CE) και μεγεθυντική ενδοσκόπηση - απεικόνιση στενής ζώνης (ME-NBI). Οι εικόνες της οφθαλμολογίας αφορούν εικόνες οπτικής τομογραφίας συνοχής (OCT), ενώ οι παθολογικές εικόνες καλύπτουν ακατέργαστες εικόνες και μικροσκοπικές εικόνες (*Εικόνα 4.1*). Οι κλινικοί γιατροί πρέπει να αφιερώσουν πολύ χρόνο για να εξετάσουν και να αξιολογήσουν αυτές τις εικόνες.

Με την ανάπτυξη της τεχνητής νοημοσύνης (ΤΝ), οι βιομηχανίες ΤΝ εισέρχονται σταδιακά στους ιατρικούς τομείς και εμπλέκονται στην ανάλυση της ιατρικής απεικόνισης, η οποία βοηθά τους γιατρούς να επιλύουν διαγνωστικά προβλήματα και να βελτιώνουν την αποτελεσματικότητα. Η τεχνητή νοημοσύνη είναι ένας κλάδος της επιστήμης των υπολογιστών για το σχεδιασμό και την εκτέλεση εργασιών που αρχικά εκτελούνταν από την ανθρώπινη νοημοσύνη. Η μηχανική μάθηση (ML) είναι ένα είδος τεχνολογίας που χρησιμοποιεί καθορισμένων εργασιών. Η μηχανική μάθηση περιλαμβάνει τη μάθηση

με επίβλεψη, τη μάθηση χωρίς επίβλεψη, τη μάθηση με ημιεπίβλεψη και την ενισχυτική μάθηση. Η επιβλεπόμενη μάθηση σημαίνει ότι το σύνολο δεδομένων εκπαίδευσης επισημαίνεται από ειδικούς ιατρούς. Η μάθηση χωρίς επίβλεψη σημαίνει ότι το σύνολο δεδομένων εκπαίδευσης δεν έχει επισημανθεί. Η μάθηση με ημιεπίβλεψη σημαίνει ότι ένα μέρος των δεδομένων εκπαίδευσης είναι επισημασμένο και άλλα είναι μη επισημασμένα. Η ενισχυτική μάθηση λαμβάνει ανατροφοδότηση για να λάβει τις πληροφορίες μάθησης και να ενημερώσει την παράμετρο του μοντέλου. Η βαθιά μάθηση (DL) είναι μια νέα κατεύθυνση στην μηχανική μάθηση (ML), η οποία βασίζεται στην προσομοίωση της δομής του νευρωνικού δικτύου του ανθρώπινου εγκεφάλου για τη δημιουργία ενός υπολογιστικού μοντέλου. Η βαθιά μάθηση (DL) χρησιμοποιείται συχνά στην ανάλυση δεδομένων υψηλής διάστασης, συμπεριλαμβανομένης της ταξινόμησης εικόνων, της ανίχνευσης αντικειμένων και της σημασιολογικής κατάτμησης. Το συνελικτικό νευρωνικό δίκτυο (CNN) είναι ο αντιπροσωπευτικός αλγόριθμος της βαθιάς μάθησης (DL).



Εικόνα 4.1: *Τομείς εξαγωγής Ιατρικών εικόνων*

4.2 Ανίχνευση Αντικειμένων σε Ιατρικές Εικόνες

Διαφορετικοί τύποι αλγορίθμων μπορούν να εφαρμοστούν σε διάφορες αναλύσεις ιατρικών εικόνων. Η ενδοσκόπηση αποτελεί βασικό εργαλείο για τη διάγνωση των ασθενειών του πεπτικού συστήματος. Η ενδοσκόπηση καθιστά ορατές τις βλάβες του πεπτικού συστήματος και μπορούν να ληφθούν βιοψίες για ιστολογική εξέταση. Χρησιμοποιείται συχνά για την έγκαιρη διάγνωση ή την παρακολούθηση καρκίνων μετεγχειρητικά. Ωστόσο, οι άπειροι γιατροί μπορεί να παραβλέψουν ορισμένες άτυπες αλλοιώσεις, επειδή οι περισσότερες από αυτές τις αλλοιώσεις προέρχονται από ατροφικό βλεννογόνο που οδηγεί σε ψευδώς αρνητικά αποτελέσματα. Ο αλγόριθμος ανίχνευσης αντικειμένων θα μπορούσε να ανιχνεύσει αυτόματα τις βλάβες και να βοηθήσει τη διάγνωση κατά τη διαδικασία της ενδοσκοπικής εξέτασης. Ο Hirasawa κ.α. χρησιμοποίησαν SSD για τη διάγνωση του γαστρικού καρκίνου σε χρωμοενδοσκοπικές εικόνες. Το σύνολο δεδομένων εκπαίδευσης αποτελούνταν από 13.584 εικόνες και το σύνολο δεδομένων δοκιμής περιελάμβανε 2.296 εικόνες από 77 γαστρικές βλάβες σε 69 ασθενείς. Το SSD είχε καλή απόδοση για την εξαγωγή ύποπτων βλαβών και την αξιολόγηση του πρώιμου γαστρικού καρκίνου. Το αποτέλεσμα έδειξε ότι ο χρόνος που δαπανάται για την ανάλυση 2.296 εικόνων είναι 47 δευτερόλεπτα και η συνολική ευαισθησία ήταν 92,2%. Αυτό σήμαινε ότι το μοντέλο SSD μπορούσε να αναλύσει μεγάλο αριθμό ενδοσκοπικών εικόνων σε σύντομο χρονικό διάστημα και μείωσε σημαντικά το φορτίο των ενδοσκοπικών ιατρών. Ο Wu κ.α. πρότειναν ένα μοντέλο ανίχνευσης αντικειμένων-ENDOANGEL για γαστρεντερική ενδοσκοπική εξέταση σε πραγματικό χρόνο. Το ENDOANGEL μπορεί να εξάγει αποτελεσματικά ύποπτες βλάβες και να αξιολογεί τη σοβαρότητα των βλαβών. Το ENDOANGEL έχει χρησιμοποιηθεί σε πολλά νοσοκομεία στην Κίνα για την υποβοήθηση της κλινικής διάγνωσης. Ο Gao κ.α. ανέλυσαν τους περι-γαστρικούς μεταστατικούς λεμφαδένες των εικόνων αξονικής τομογραφίας χρησιμοποιώντας το Faster R-CNN. Για την ανάλυση χρησιμοποιήθηκε η μετριτική τιμή AUC. Η καμπύλη ROC (Receiver Operator Characteristic) είναι μια μετρική αξιολόγησης για προβλήματα δυαδικής ταξινόμησης. Πρόκειται για μια καμπύλη πιθανότητας που απεικονίζει το TPR έναντι του FPR σε διάφορες τιμές κατωφλίου και ουσιαστικά διαχωρίζει το "σήμα" από το "θόρυβο". Η περιοχή κάτω από την καμπύλη (AUC) είναι το μέτρο της ικανότητας ενός ταξινομητή να διακρίνει μεταξύ κλάσεων και χρησιμοποιείται ως σύνοψη της καμπύλης ROC. Η ανάλυση χωρίστηκε σε δύο στάδια, το αρχικό στάδιο εκμάθησης για την εκπαίδευση και το στάδιο ακριβούς εκμάθησης για τη λεπτομερή ρύθμιση και τη δοκιμή. Το αποτέλεσμα

έδειξε ότι, στο αρχικό στάδιο μάθησης, τα ποσοστά ανάκλησης των κλάσεων οζιδίων για το σύνολο εκπαίδευσης και το σύνολο επικύρωσης, το mAP ήταν 0,5019 και η AUC ήταν 0,8995. Στο ακριβές στάδιο εκμάθησης, το mAP και η AUC ήταν 0,7801 και 0,9541, τα οποία ήταν προφανώς βελτιωμένα, σε σύγκριση με το αρχικό στάδιο εκμάθησης. Έτσι, το μοντέλο Faster R-CNN είχε υψηλή αποτελεσματικότητα κρίσης και ακρίβεια αναγνώρισης για την CT διάγνωση των περι-γαστρικών μεταστατικών λεμφαδένων.

Πριν τις δοκιμές στο σετ δεδομένων για τις αξονικές τομογραφίες δοκιμάστικε η τεχνική του Faster R-CNN σε ένα σετ για την αναγνώριση των λευκών και των ερυθρών κυττάρων του αίματος. Το dataset περιλάμβανε 364 εικόνες με πληθυσμούς κυττάρων και 4888 ετικέτες που αναγνώριζαν τα λευκά και τα ερυθρά κύτταρα του αίματος άλλα και τα αιμοπετάλια. Το dataset ήταν pre-labeled οπότε απλά έγινε προετοιμασία των εικόνων και των σημειώσεων του μοντέλου

Η γνώση της παρουσίας και της αναλογίας των ερυθρών αιμοσφαιρίων, των λευκών αιμοσφαιρίων και των αιμοπεταλίων των ασθενών είναι το κλειδί για τον εντοπισμό πιθανών ασθενειών. Η παροχή δυνατότητας στους γιατρούς να αυξήσουν την ακρίβεια και την απόδοση του εντοπισμού των εν λόγω μετρήσεων αίματος μπορεί να βελτιώσει μαζικά την υγειονομική περίθαλψη για εκατομμύρια ανθρώπους.

Για την απόδοση σωστών αποτελεσμάτων είναι επιθυμητό να γίνει μια προετοιμασία των εικόνων:

- Επιβεβαίωση πως οι σημειώσεις (annotations) είναι σωστές. (π.χ. δεν υπάρχουν annotations που βρίσκονται εκτός εικόνας)
- Βεβαίωση πως ο προσανατολισμός των εικόνων είναι σωστός (π.χ. οι εικόνες αποθηκεύονται με διαφορετικό τρόπο από τον που παρουσιάζονται στις εφαρμογές)
- Αλλαγή μεγέθους των εικόνων και ενημέρωση των annotations για να ταιριάζει με το νέο μέγεθος των εικόνων.
- Έλεγχος της υγείας του συνόλου δεδομένων, όπως η ισορροπία των τάξεων, τα μεγέθη των εικόνων και οι αναλογίες διαστάσεων - και προσδιορισμός του τρόπου με τον οποίο αυτά μπορεί να επηρεάσουν την προεπεξεργασία και τις επαυξήσεις που θα εκτελεστούν.

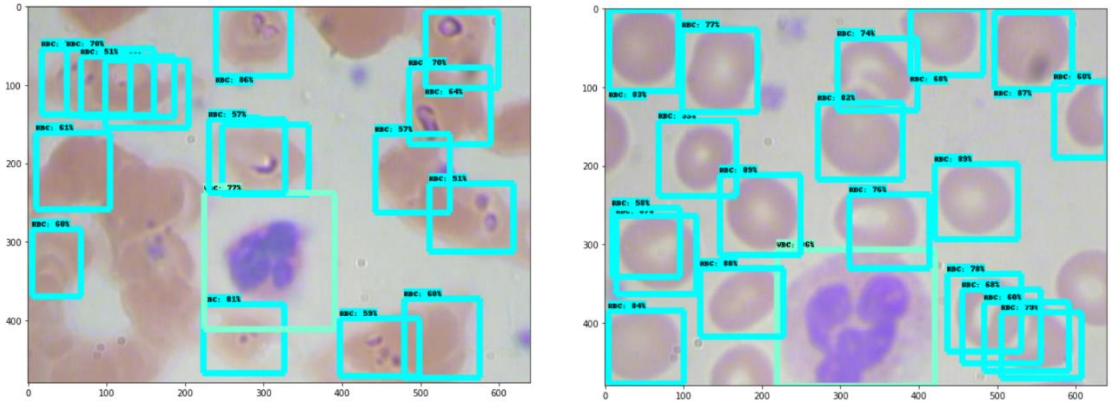
- Διάφορες χρωματικές διορθώσεις που μπορούν να βελτιώσουν την απόδοση του μοντέλου, όπως προσαρμογές κλίμακας του γκρι και αντίθεσης.

Δεν θα αναλυθούν εκτενώς οι λεπτομέρειες για την λειτουργία του συγκεκριμένου μοντέλου καθώς δεν είναι αυτό για το οποίο γίνεται η διπλωματική εργασία αλλά θα παρουσιαστούν ο τρόπος εκπαίδευσης του μοντέλου και τα αποτελέσματα.

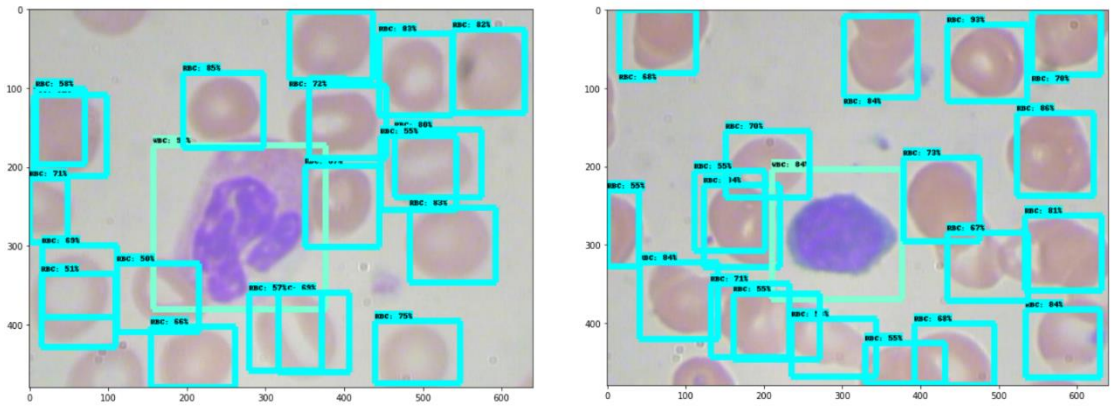
Θα εκπαιδευτεί ένα νευρωνικό δίκτυο Faster R-CNN. Το Faster R-CNN είναι ένας ανιχνευτής αντικειμένων σε δύο στάδια: πρώτα προσδιορίζει τις περιοχές ενδιαφέροντος και στη συνέχεια περνά αυτές τις περιοχές σε ένα νευρωνικό δίκτυο συνελικτικής ανάλυσης. Οι χάρτες χαρακτηριστικών που εξάγονται περνούν σε μια μηχανή διανυσμάτων υποστήριξης (VSM) για ταξινόμηση. Υπολογίζεται η παλινδρόμηση μεταξύ των προβλεπόμενων οριοθετημένων πλαισίων και των οριοθετημένων πλαισίων της βασικής αλήθειας. Το Faster R-CNN, παρά το όνομά του, είναι γνωστό ότι είναι πιο αργό μοντέλο από ορισμένες άλλες επιλογές (όπως το YOLOv3 ή το MobileNet) για συμπερασματολογία, αλλά ελαφρώς πιο ακριβές.

Το Faster R-CNN είναι μία από τις πολλές αρχιτεκτονικές μοντέλων που παρέχει εξ ορισμού το TensorFlow Object Detection API, μεταξύ άλλων και με προ-εκπαιδευμένα βάρη. Αυτό σημαίνει ότι θα ξεκινήσει η διαδικασία με ένα μοντέλο εκπαιδευμένο σε COCO (common objects in context) και θα προσαρμοστεί στα δεδομένα που χρειάζονται στην συνέχεια.

Μετά την ολοκλήρωση της διαδικασίας τα αποτελέσματα που εξήγχθησαν από το μοντέλο παρουσιάζονται παρακάτω:



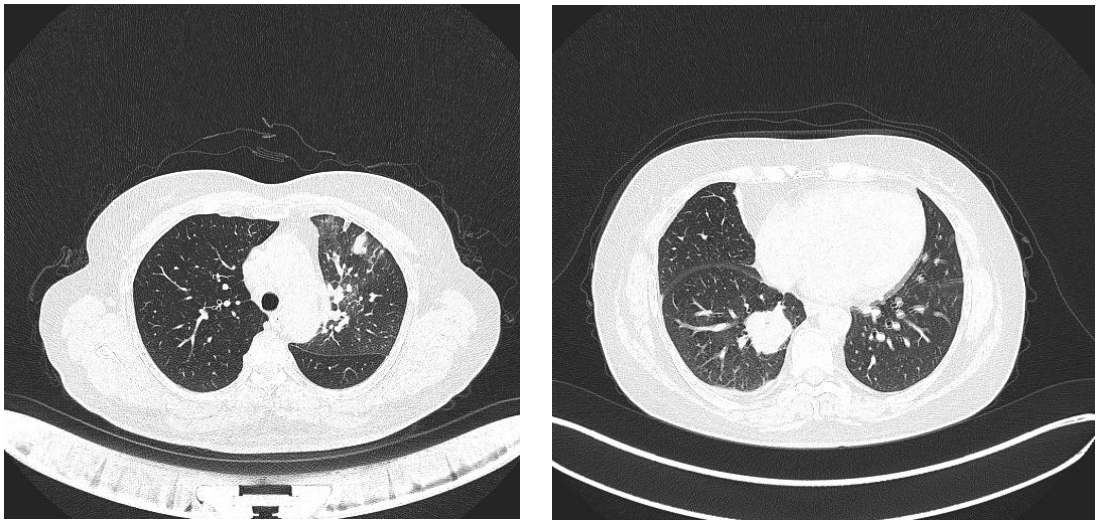
Εικόνα 4.2: *Αποτελέσματα ερυθρών και λευκών κυττάρων του αίματος Με μπλε χρώμα απεικονίζονται τα ερυθρά κύτταρα και με πράσινο τα λευκά*



Εικόνα 4.3: *Αποτελέσματα ερυθρών και λευκών κυττάρων του αίματος Με μπλε χρώμα απεικονίζονται τα ερυθρά κύτταρα και με πράσινο τα λευκά*

4.3 Ανίχνευση Αντικειμένων σε Αξονικές Τομογραφίες (CT)

Κίνητρο για την συγκεκριμένη διπλωματική εργασία ήταν η ανάλυση αξονικών τομογραφιών για την ανίχνευση του καρκίνου στους πνεύμονες. Ο λόγος που επιλέχθηκε να γίνει Object Detection σε CT είναι καθώς θα χρησιμοποιούνταν ο αλγόριθμος Faster R-CNN που όπως έχει αναλυθεί επανειλημμένα στα προηγούμενα κεφάλαια της διπλωματικής εργασίας είναι από τους πιο αξιόπιστους αλγορίθμους όσον αφορά την ακρίβεια των αποτελεσμάτων.



Εικόνα 4.4: *Αξονικές Τομογραφίες Πνεύμονα*

Όπως γίνεται αντιπληπτό από τις παραπάνω εικόνες η μοναδικός τρόπος επεξεργασίας των εικόνων είναι η ανάλυση από έναν έμπειρο ιατρό. Με μια γρήγορη ματιά όμως, μπορεί κάποιος να καταλάβει πως ακόμα και η ανάλυση από ιατρούς στις συγκεκριμένες εικόνες είναι δύσκολη. Η διαδικασία του Deep Learning και η εκπαίδευση ένος

μοντέλου Faster R-CNN έρχεται να επισπεύσει την διαδικασία της ανάλυσης και να βοηθήσει το ιατρικό προσωπικό να αντιληφθεί πληροφορίες οι οποίες θα ήταν δύσκολο να εξαχθούν ακόμα και μετά από εκτενή επεξεργασία. Η εκτενής ανάλυση της διαδικασίας και της επεξήγησης όλων των επεξεργασιών που γίνανε θα ακολουθήσει στο επόμενο κεφάλαιο.

Κεφάλαιο 5: Πειραματική Διαδικασία

Στο κεφάλαιο αυτό ακολουθεί η αναλυτική περιγραφή των δεδομένων, των παραμέτρων εκπαίδευσης, των μετρικών για την αξιολόγηση των επιδόσεων του δικτύου και τέλος η σύγκριση μεταξύ των διαφορετικών διαχωρισμών του dataset σε train, test και validation σετ.

5.1 Περιγραφή δεδομένων

Για την αξιολόγηση απόδοσης της αρχιτεκτονικής Faster R-CNN για τον εντοπισμό όγκου σε αξονικές τομογραφίες, χρησιμοποιήθηκε ένα σύνολο δεδομένων που δώθηκε από το Εργαστήριο Φωτογραμμετρίας του Τομέα Τοπογραφίας της Σχολής Αγρονόμων και Τοπογράφων Μηχανικών – Μηχανικών Γεωπληροφορικής (ΣΑΤΜ-ΜΓ).

Το αποθετήριο των δεδομένων περιέχει 36738 εικόνες εκ των οποίων οι 4751 περιέχουν annotations οπότε είναι και αυτές που χρησιμοποιήθηκαν για την εκπαίδευση του μοντέλου.

Όλες οι εικόνες είναι ασπρόμαυρες καθώς προέρχονται κατευθείαν από την αξονική τομογραφία και εξάγωνται σε μορφή .dcm (Digital Imaging and Communications in Medicine (DICOM)). Η επεξεργασία σε αυτό το είδος αρχείου είναι αρκετά δύσκολη επομένως ακολούθησε μετατροπή των εικόνων σε αρχεία τύπου .jpg και καθώς είναι μια πιο διαδεδομένη μορφή εικόνας αλλά και επειδή μειώνει το μέγεθος του dataset σημαντικά δημιουργώντας περισσότερο ελεύθερο χώρο στην μνήμη της κάρτας γραφικών που θα τις επεξεργαστεί.

Το μειονέκτημα που υπήρχε στο dataset που χρησιμοποιήθηκε ήταν πως δεν είχαν υποστεί καμία επεξεργασία και χρειαζόταν μια προεπεξεργασία για την παραγωγή βέλτιστων αποτελεσμάτων. Για την επεξεργασία αυτή χρησιμοποιήθηκε το Roboflow, μια ηλεκτρονική πλατφόρμα η οποία εξειδικεύεται στην επεξεργασία των dataset πριν αυτά χρησιμοποιηθούν για εκπαίδευση μοντέλων για Deep Learning.

Έγιναν αρκετοί διαχωρισμοί για να δοκιμαστεί ποιός είναι ο κατάλληλος για τα πιο ακριβή αποτελέσματα. Επομένως δημιουργήθηκαν 3 εκδόχες:

Εκδοχή	Training Set (Images - %)	Validation Set (Images - %)	Testing Set(Images - %)
$1^{η}$	3285 – 70%	972 – 20%	494 – 10%
$2^{η}$	2849 – 60%	951 – 20%	951 – 20%
$3^{η}$	9855 – 70%	2916 – 20%	1482 – 10%

Πίνακας 5.1: *Διαφορετικές εκδοχές διαχωρισμού dataset*

Γίνεται αντιληπτό πως η $3^{η}$ εκδοχή περιέχει πολλές περισσότερες εικόνες από ότι αναφέρθηκε πως θα συμμετέχουν στην διαδικασία της εκπαίδευσης του μοντέλου. Η εκδοχή αυτή ουσιαστικά είναι η $1^{η}$ εκδοχή απλώς τριπλασιάζοντας κάθε σετ. Αυτό γίνεται για

προσδιοριθεί εάν ένας μεγαλύτερος αριθμός εικόνων θα βελτιώσει την ακρίβεια του μοντέλου.

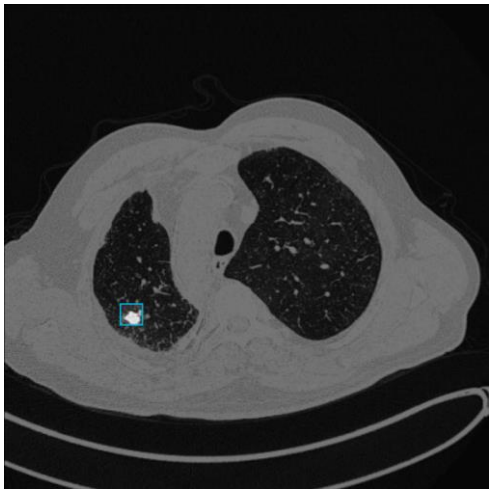
Στην συνέχεια ακολούθησε η επεξεργασία του dataset. Η πρώτη εφαρμογή πάνω στο dataset είναι η αλλαγή του μεγέθους σε 416x416. Αυτό γίνεται καθώς δεν είναι γνωστό το μέγεθος όλων των εικόνων και είναι δύσκολο να ελεγχθεί οπότε για να γίνει η διαδικασία της επεξεργασίας των εικόνων πιο ευκόλη για το πρόγραμμα αλλάζει το μέγεθος. Επιπλέον μπορεί το μέγεθος να είναι μικρότερο και από τις πραγματικές εικόνες οπότε μειώνεται και το μέγεθος της εικόνας σε mb οπότε καταλώνεται και λιγότερος χώρος στην μνήμη επεξεργασίας της κάρτας γραφικών (GPU). Μια άλλη διαδικασία που λαμβάνει χώρα είναι και ο προσανατολισμός των εικόνων. Είναι πιθανό όπως έγινε η λήψη των εικόνων να έχει αλλάξει ο προσανατολισμός τους, επομένως αυτό θα δημιουργούσε μεγάλο πρόβλημα στην εκπαίδευση του προγράμματος καθώς αλλάζουν οι συντεταγμένες στις οποίες οδηγείται το μοντέλο για την αναγνώριση των όγκων.

Μετά την αρχική επεξεργασία ακολουθεί η ενίσχυση των εικόνων (augmentation). Η διαδικασία προκαλεί μικρές αλλαγές στο dataset οι οποίες ενισχύουν την εκπαιδευτική ικανότητα του μοντέλου κατα την διάρκεια της εκπαίδευσης. Τα δείγματα ενίσχυσης που εφαρμόστηκαν παρουσιάζονται στον παρακάτω πίνακα:

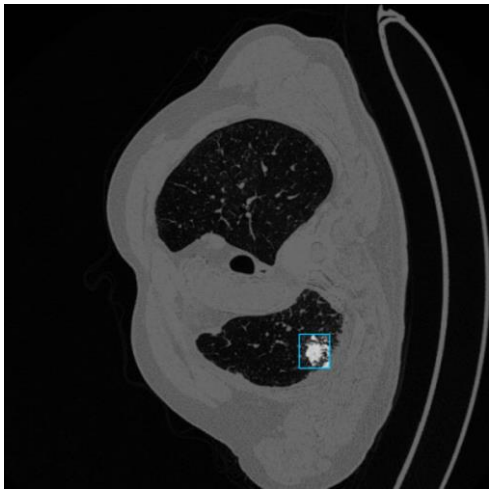
Augmentation	**Application**
Flip	Horizontal, Vertical
90° Rotate	Clockwise, Counter-Clockwise, Upside Down
Crop	0% Minimum Zoom, 15% Maximum Zoom
Hue	Between -25° and +25°
Saturation	Between -25% and +25%
Brightness	Between -15% and +15%
Exposure	Between -20% and +20%

Πίνακας 5.2: *Δείγματα ενίσχυσης*

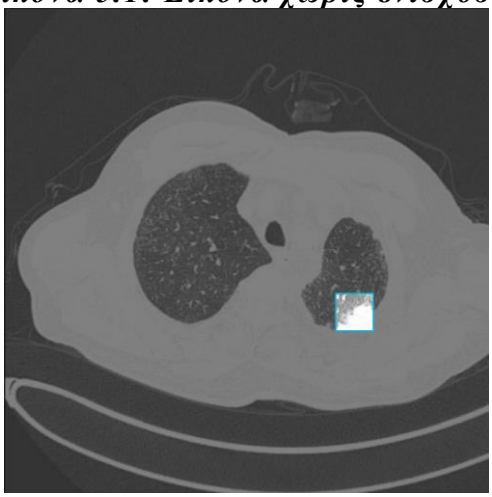
Τέλος παρουσιάζονται και μερικές εικόνες για την παρουσίαση των ενισχύσεων:



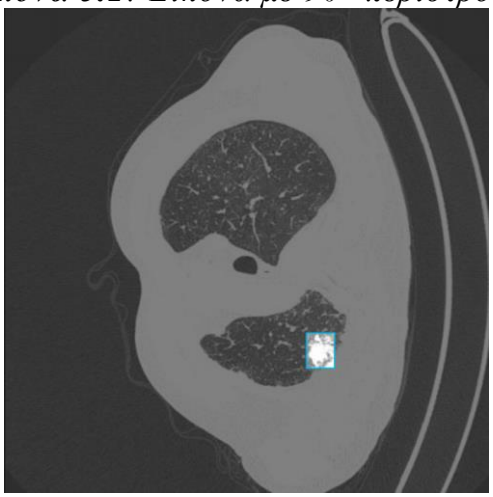
Εικόνα 5.1: Εικόνα χωρίς ενισχύση



Εικόνα 5.2*: Εικόνα με 90° περιστροφή*



Εικόνα 5.3: *Εικόνα με Horizontal flip και Saturation*



Εικόνα 5.4: *Εικόνα με Saturation*

5.2 Παράμετροι Εκπαίδευσης

Σε αυτή την ενότητα γίνεται ανάλυση των παραμέτρων που υλοποιήθηκαν για την βέλτιστη εκπαίδευση του δικτύου Faster R-CNN. Για την παρούσα διπλωματική εργασία δεν χρησιμοποιήθηκε τοπική μονάδα για την εκπαίδευση του μοντέλου, αλλά χρησιμοποιήθηκε το free version του Google Colab. Το Colab επιτρέπει σε οποιονδήποτε να γράφει και να εκτελεί αυθαίρετο κώδικα python μέσω του προγράμματος περιήγησης και είναι ιδιαίτερα κατάλληλο για μηχανική μάθηση, ανάλυση δεδομένων και εκπαίδευση. Πιο τεχνικά, το Colab είναι μια φιλοξενούμενη υπηρεσία σημειωματάριου Jupyter που δεν απαιτεί καμία εγκατάσταση για να χρησιμοποιηθεί, ενώ παρέχει δωρεάν πρόσβαση σε υπολογιστικούς πόρους,

συμπεριλαμβανομένων των GPUs. Βέβαια για την δωρεάν έκδοση υπάρχει ο περιορισμός ώρας χρήσης των υπολογιστικών πόρων και αυτό το μειονέκτημα θα εξηγηθεί στην συνέχεια.

Η δυνατότητα χρήσης μιας τέτοιας ιστοσελίδας και μάλιστα δωρεάν για μηχανική μάθηση, ανάλυση δεδομένων και εκπαίδευση μοντέλος αποτελεί την πιο οικονομική επιλογή αποδεικνύοντας τις άπειρες δυνατότητες που παρέχει η γλώσσα προγραμματισμού python.

Λόγω του μεγάλου όγκου εικόνων και το μέγεθος τους σε mb, μετατράπηκαν οι εικόνες μέσω του Roboflow σε ένα διαφορετικό είδος αρχείου το TFRecord η οποία είναι μια απλή μορφή για την αποθήκευση σε δυαδική μορφή μειώνοντας το μέγεθος των αρχείων στο μέγιστο. Στην συνέχεια έγιναν upload σε ένα Google Drive για να υπάρχει άμεση πρόσβαση και έτσι κλήθηκαν στον κώδικα που χρησιμοποιήθηκε στο Colab.

Η πιο σημαντική παράμετρος της εκπαίδευσης αποτελεέι το batch size που εκφράζει πόσες εικόνες του τετρασδιάστατου τανιστή θα χρησιμοποιούνται σε κάθε εποχή μέχρι να εκπαιδευτεί πλήρως το δίκτυο. Η τιμή αυτή ορίστηκε στο 12, προκειμένου οι παράμετροι της απώλειας να υπολογίζονται για ένα ικανοποιητικό σύνολο και ταυτόχρονα να μην επιβαρύνεται το σύστημα με υπερβολικούς υπολογισμούς.

Το δίκτυο εκπαιδεύτηκε με μέγεθος batch size 12 και για 1 epoch αλλά 10000 steps. Μια εποχή ολοκληρώνεται όταν όλο το σετ δεδομένων έχει ανακυκλωθεί στο δίκτυο. Δεν υπάρχει συγκεκριμένη διαφορά με τα steps απλώς χρησιμοποιούνται για να επιταχυνθεί η διαδικασία και να αποφευχθεί το Overfitting (υπερπροσαρμογή).

Το Overfitting είναι μια έννοια στην επιστήμη των δεδομένων, η οποία συμβαίνει όταν ένα στατιστικό μοντέλο ταιριάζει ακριβώς με τα δεδομένα εκπαίδευσης. Όταν συμβαίνει αυτό, ο αλγόριθμος δυστυχώς δεν μπορεί να αποδώσει με ακρίβεια έναντι αθέατων δεδομένων, με αποτέλεσμα να εξουδετερώνεται ο σκοπός του. Η γενίκευση ενός μοντέλου σε νέα δεδομένα είναι τελικά αυτό που επιτρέπει την καθημερινή χρήση αλγόριθμων μηχανικής μάθησης για να πραγματοποιούνται προβλέψεις και ταξινομήσεις δεδομένων.

Όταν κατασκευάζονται αλγόριθμοι μηχανικής μάθησης, αξιοποιούν ένα σύνολο δειγματικών δεδομένων για την εκπαίδευση του μοντέλου. Ωστόσο, όταν το μοντέλο εκπαιδεύεται για πολύ καιρό σε δειγματικά

δεδομένα ή όταν το μοντέλο είναι πολύ περίπλοκο, μπορεί να αρχίσει να μαθαίνει τον "θόρυβο" ή τις άσχετες πληροφορίες μέσα στο σύνολο δεδομένων. Όταν το μοντέλο απομνημονεύει τον θόρυβο και προσαρμόζεται πολύ στενά στο σύνολο εκπαίδευσης, το μοντέλο γίνεται "υπερβολικά προσαρμοσμένο" και δεν μπορεί να γενικεύσει καλά σε νέα δεδομένα. Εάν ένα μοντέλο δεν μπορεί να γενικεύσει καλά σε νέα δεδομένα, τότε δεν θα είναι σε θέση να εκτελέσει τις εργασίες ταξινόμησης ή πρόβλεψης για τις οποίες προοριζόταν.

Τα χαμηλά ποσοστά σφάλματος και η υψηλή διακύμανση είναι καλοί δείκτες υπερπροσαρμογής. Προκειμένου να αποφευχθεί αυτού του είδους η συμπεριφορά, μέρος του συνόλου δεδομένων εκπαίδευσης συνήθως τίθεται στην άκρη ως "σύνολο δοκιμής" για τον έλεγχο της υπερπροσαρμογής. Εάν τα δεδομένα εκπαίδευσης έχουν χαμηλό ποσοστό σφάλματος και τα δεδομένα δοκιμής έχουν υψηλό ποσοστό σφάλματος, αυτό σηματοδοτεί υπερπροσαρμογή.

Μια επιπλέον εφαρμογή που μπορεί να εφαρμοστεί είναι τα EarlyStoppings και τα model checkpoint. Στη μηχανική μάθηση, το EarlyStopping είναι μια μορφή κανονικοποίησης που χρησιμοποιείται για την αποφυγή υπερβολικής προσαρμογής κατά την εκπαίδευση ενός μοντέλου με μια επαναληπτική μέθοδο, όπως η κάθοδος κλίσης. Τέτοιες μέθοδοι ενημερώνουν τον εκπαιδευτή έτσι ώστε να προσαρμόζεται καλύτερα στα δεδομένα εκπαίδευσης με κάθε επανάληψη. Μέχρι ενός σημείου, αυτό βελτιώνει την απόδοση του μοντέλου σε δεδομένα εκτός του συνόλου εκπαίδευσης. Μετά από αυτό το σημείο, ωστόσο, η βελτίωση της προσαρμογής του μοντέλου στα δεδομένα εκπαίδευσης γίνεται εις βάρος του αυξημένου σφάλματος γενίκευσης. Οι κανόνες του EarlyStopping παρέχουν καθοδήγηση σχετικά με το πόσες επαναλήψεις μπορούν να εκτελεστούν πριν το μοντέλο αρχίσει να προσαρμόζεται υπερβολικά. Όσο για το Model Checkpoint είναι μια διαδικασία που αποθηκεύει το καλύτερο μοντέλο ανάμεσα στις εποχές που έχουν οριστεί και αποθηκεύει τα βάρη, τα Average Precision και τα Average Recall του μοντέλου.

Η τελευταία σημαντική παράμετρος που αφορά το δίκτυο είναι η συνάρτηση απώλειας (Loss Function) πού όπως έχει αναφερθεί αποτελεί καίριο παράγοντα της εκπαίδευσης του δικτύου. Μια συνάρτηση απώλειας μετατρέπει μια θεωρητική πρόταση σε πρακτική. Η δημιουργία ενός εξαιρετικά ακριβούς προβλεπτικού μηχανισμού απαιτεί συνεχή επανάληψη του προβλήματος μέσω ερωτήσεων, μοντελοποίηση του προβλήματος με την επιλεγμένη προσέγγιση και δοκιμές.

Το μόνο κριτήριο με το οποίο εξετάζεται ένα στατιστικό μοντέλο είναι η απόδοσή του - πόσο ακριβείς είναι οι αποφάσεις του μοντέλου. Αυτό απαιτεί έναν τρόπο μέτρησης του πόσο απέχει μια συγκεκριμένη επανάληψη του μοντέλου από τις πραγματικές τιμές. Σε αυτό το σημείο μπαίνουν στο παιχνίδι οι συναρτήσεις απωλειών.

Οι συναρτήσεις απωλειών μετρούν πόσο μακριά βρίσκεται μια εκτιμώμενη τιμή από την πραγματική της τιμή. Μια συνάρτηση απωλειών αντιστοιχίζει τις αποφάσεις στο σχετικό τους κόστος. Οι συναρτήσεις απωλειών δεν είναι σταθερές, αλλά αλλάζουν ανάλογα με την εκάστοτε εργασία και τον επιδιωκόμενο στόχο.

Η συνάρτηση απώλειας πολλαπλών εργασιών της μάσκας R-CNN συνδυάζει τις απώλειες της ταξινόμησης, του εντοπισμού και της μάσκας τμηματοποίησης: $L = L_{cls} + L_{box} + L_{mask}$, όπου το L_{cls} και το L_{box} είναι τα ίδια στο Faster R-CNN. Επομένως, προκύπτει πως η συνάρτηση απώλειας του Faster R-CNN είναι η εξής:

$$L(\{p_i\},\{t_i\}) = \frac{1}{N_{cls}}\sum_i L_{cls}(p_i, p_i^*) + \lambda\frac{1}{N_{reg}}\sum_i p_i^* L_1^{smooth}(t_i, t_i^*)$$

όπου L_{cls} είναι η λογαριθμική συνάρτηση απωλειών για δύο κλάσεις, καθώς μπορεί εύκολα να μεταφραστεί μια ταξινόμηση πολλαπλών κλάσεων σε δυαδική ταξινόμηση προβλέποντας ότι ένα δείγμα είναι αντικείμενο-στόχος ή όχι. L_1^{smooth} είναι η ομαλή απώλεια L1.

$$L_{cls}(p_i, p_i^*) = -p_i^* log p_i - (1 - p_i^*)\log(1 - p_i)$$

5.3 Αξιολόγηση αποτελεσμάτων

Σε αυτή την ενότητα παρουσιάζονται οι μετρικοί δείκτες που εφαρμόστηκαν σε κάθε εικόνα για τα δεδομένα ελέγχου. Πιο αναλυτικά χρησιμοποιήθηκαν οι δείκτες precision & recall.

- Precision: Μετρά πόσο ακριβείς είναι οι προβλέψεις, δηλαδή το ποσοστό των προβλέψεων που είναι σωστές και προκύπτει από τον τύπο:

$$Precision = \frac{TP}{TP + FP}$$

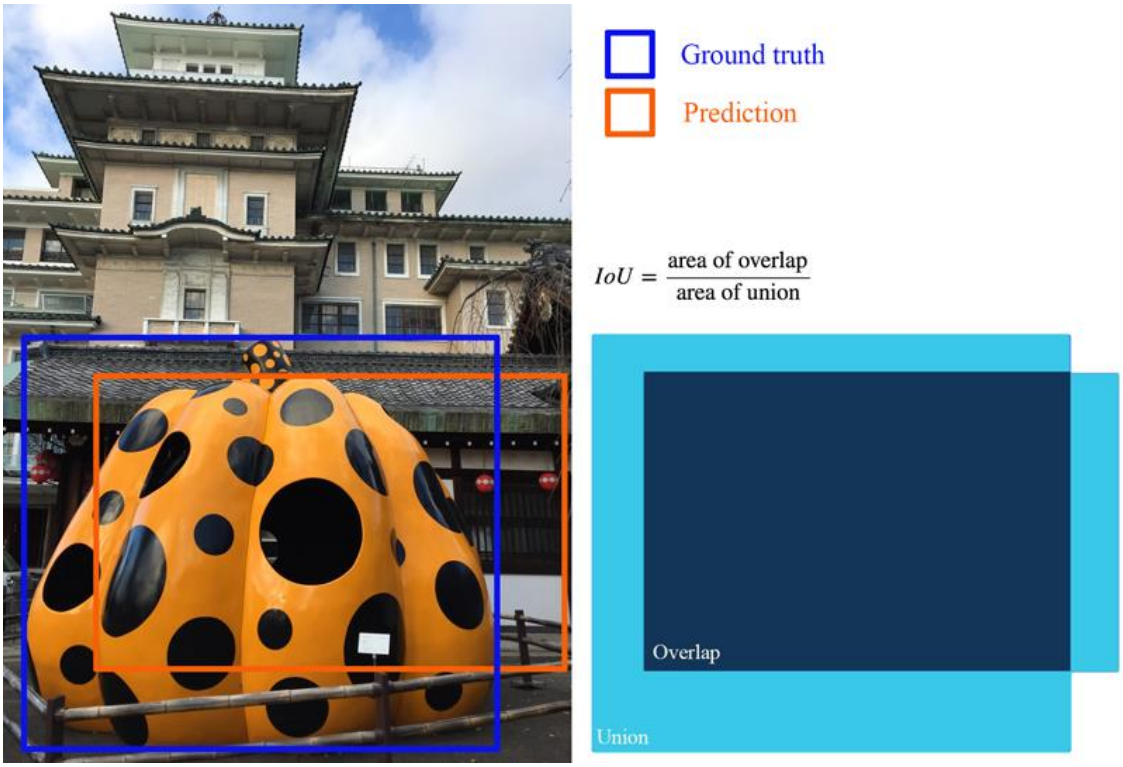
- Recall: Μετράει πόσο καλά βρίσκονται τα θετικά αποτελέσματα. Για παράδειγμα μπορεί να βρεθεί το 80% των θετικών περιπτώσεων στις κορυφαίες κ προβλέψεις

$$Recall = \frac{TP}{TP + FN}$$

Όπου TP = True Positive, TN = True Negative, FP = False Positive & FN = False Negative.

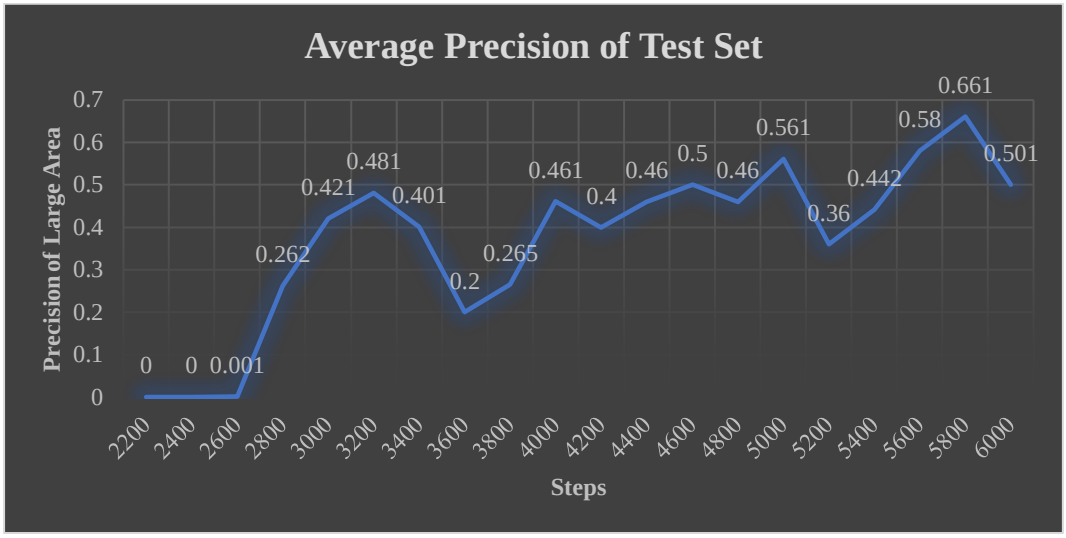
Για την καλύτερη επεξήγηση των παραπάνω ορισμών είναι αναγκαία η επεξήγηση του Intersection over Union (IoU). Η IoU μετρά την επικάλυψη μεταξύ 2 ορίων. Χρησιμοποιείται για την μέτρηση του πόσο επικαλύπτεται το προβλεπόμενο όριο με την αλήθεια του εδάφους (το πραγματικό όριο του αντικειμένου). Σε ορισμένα σύνολα δεδομένων, προκαθορίζεται ένα κατώφλι IoU (π.χ. 0,5) για να ταξινομηθεί αν η πρόβλεψη είναι αληθώς θετική ή ψευδώς θετική. Στην συγκεκριμένη διπλωματική εργασία στο 1ο στάδιο επεξεργασίας χρησιμοποιείται κατώφλι 0.7 και στην συνέχεια 0.6.

Στην *Εικόνα 5.5* παρουσιάζεται ένα παράδειγμα για την περιγραφή του IoU.



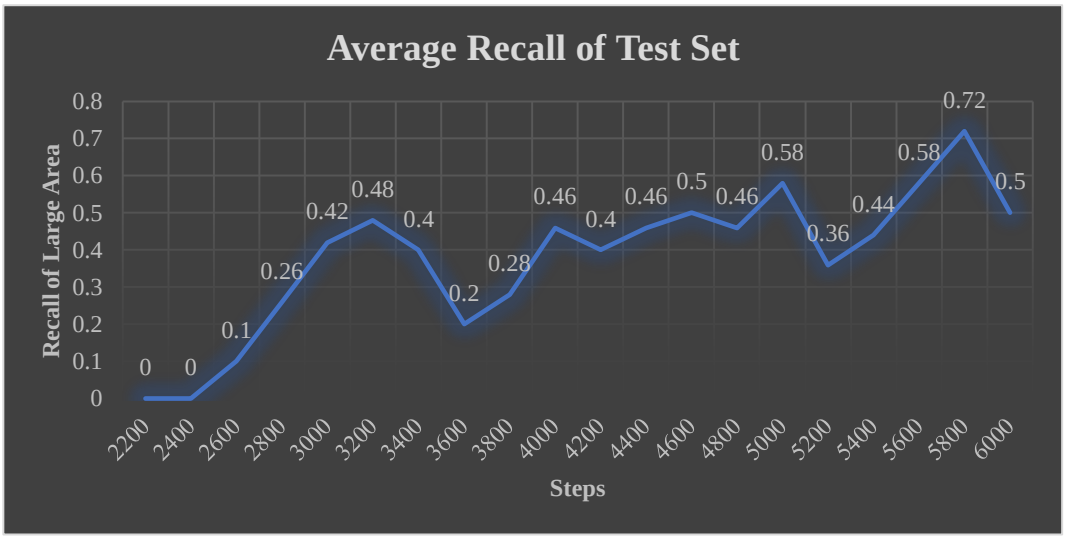
Εικόνα 5.5: *Ερμηνεία Intersection over Union (IoU)* ***[108]***

Για την περίπτωση της καλύτερης μεθόδου για την εκπαίδευση του μοντέλου θα παρουσιαστεί όλο το γράφημα του Precision και του Recall. Για τις άλλες 2 εκδοχές θα χρησιμοποιηθούν μόνο οι καλύτερες τιμές για σύγκριση καθώς βρέθηκαν να είναι χειρότερες εκδοχές για αξιόπιστα αποτελέσματα.

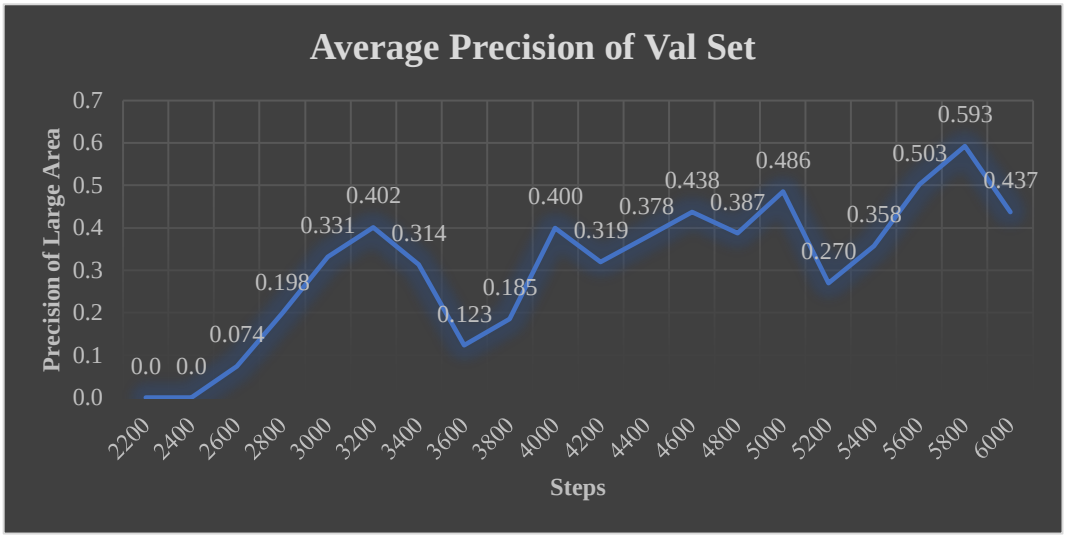


Γράφημα 5.1: *Average Precision 2ης Μεθόδου για το Test Set*

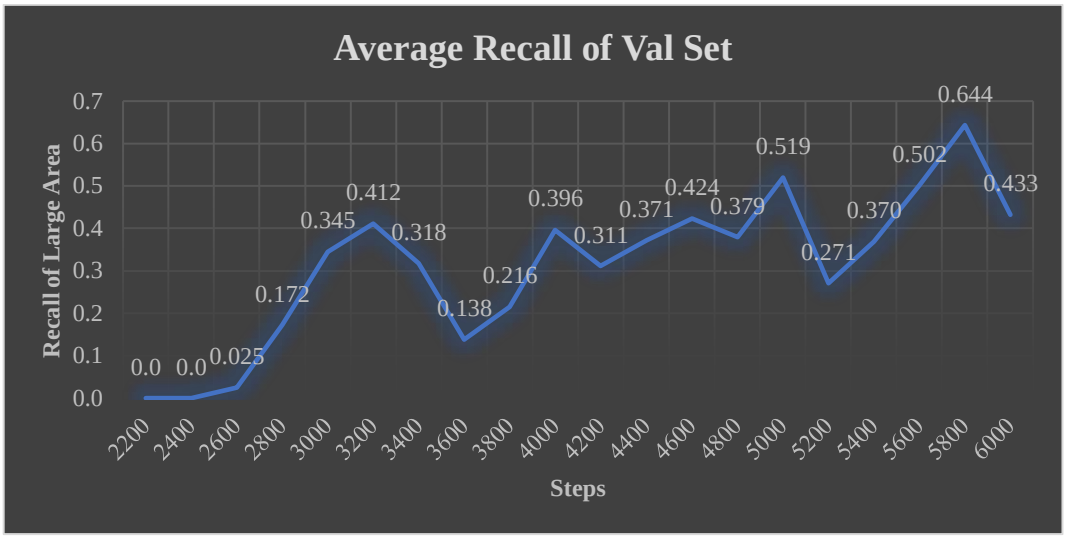
Όπως γίνεται διακριτό και από το γράφημα της 2ης εκδοχής η μέγιστη τιμή είναι 66.1% κατα το 5800 step. Το δίκτυο πέτυχε την πιο υψηλή απόδοση με τον διαχωρισμό του dataset σε 60% training set, 20% validation set & 20% test set. Στο διάγραμμα του Recall που παρουσιάζεται παρακάτω γίνεται αντιληπτή και η επιτυχία αυτής της μεθόδου καθώς η μέγιστη τιμή που λαμβάνει είναι το 72%.



Γράφημα 5.2: *Average Recall 2ης μεθόδου για το Test Set*



Γράφημα 5.3: *Average Precision 2ης Μεθόδου για το Val Set*



Γράφημα 5.4: *Average Recall 2ης μεθόδου για το Val Set*

Διακρίνεται πως τα steps τελειώνουν στα 6000. Πιθανότατα αν συνεχιζόταν η διαδικασία να αυξανόταν και άλλο το ποσοστό και να δημιουργόντουσαν ακόμα καλύτερα αποτελέσματα. Όμως η δωρεάν έκδοση του Google Colab παρέχει περιορισμένη δωρεάν υπολογιστική δύναμη και δεν ήταν δυνατόν να ξεπεραστούν τα 6000 βήματα καθώς η ιστοσελίδα μετά από κάποιες ώρες εργασίας σταματούσε την διαδικασία. Είναι σημαντικό να συμπλωρωθεί πως παρόλο που ο αλγόριθμος του Faster R-CNN είναι πιο αργός από άλλους αλγόριθμους Deep Learning για Object Detection παρέχει τα καλύτερα αποτελέσματα.

	1η Εκδοχή	2η Εκδοχή	3η Εκδοχή
Χρόνος Λειτουργίας	6:23:43	6:56:03	6:19:55
Steps	5800	6000	5600
Loss	0.1672	0.1414	0.1842
Average Precision Val Set	54.4%	66.5%	48.8%
Average Recall Val Set	55.2%	71.4%	48.1%
Average Precision Test Set	55.1%	66.1%	49.5%
Average Recall Test Set	56%	72%	48.8%

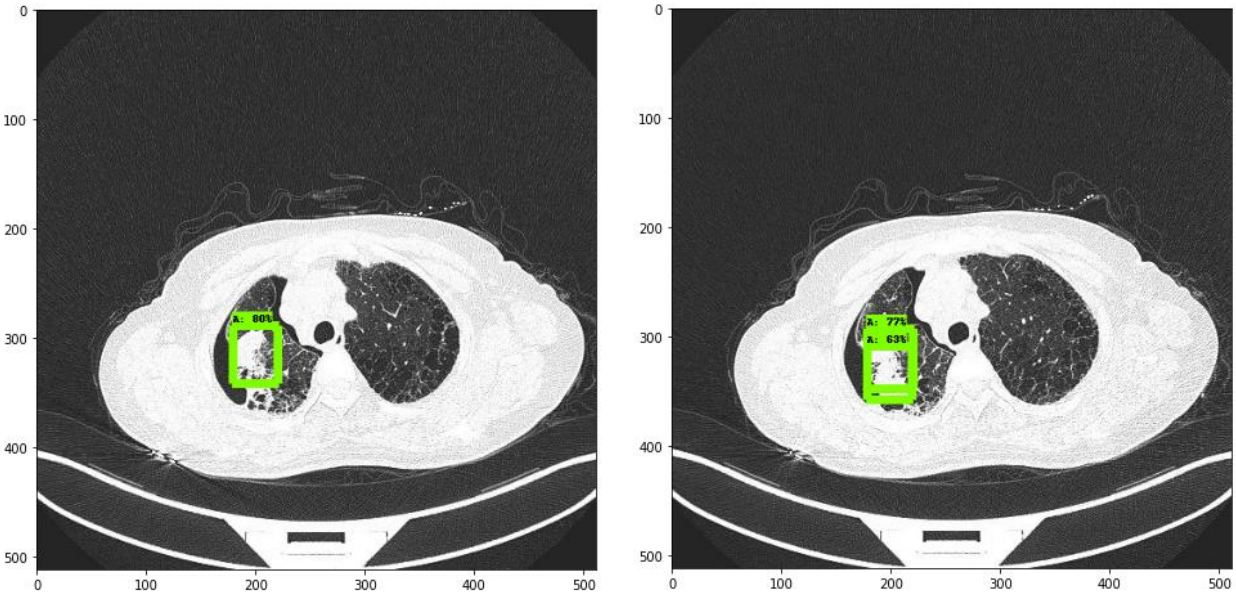
Πίνακας 5.3: *Αξιολόγηση εκπαίδευσης για 3 εκδοχές*

Σύμφωνα με τον παραπάνω πίνακα γίνεται αντιληπτό πως ίσως και οι άλλες 2 εκδοχές να είχαν καταφέρει να φτάσουν στο επίπεδο της 2ης εκδοχής με περισσότερες επαναλήψεις.

Σημαντική πληροφορία είναι και ο χρόνος λειτουργίας καθώς διακρίνεται πόσο χρονοβόρα διαδικασία είναι.

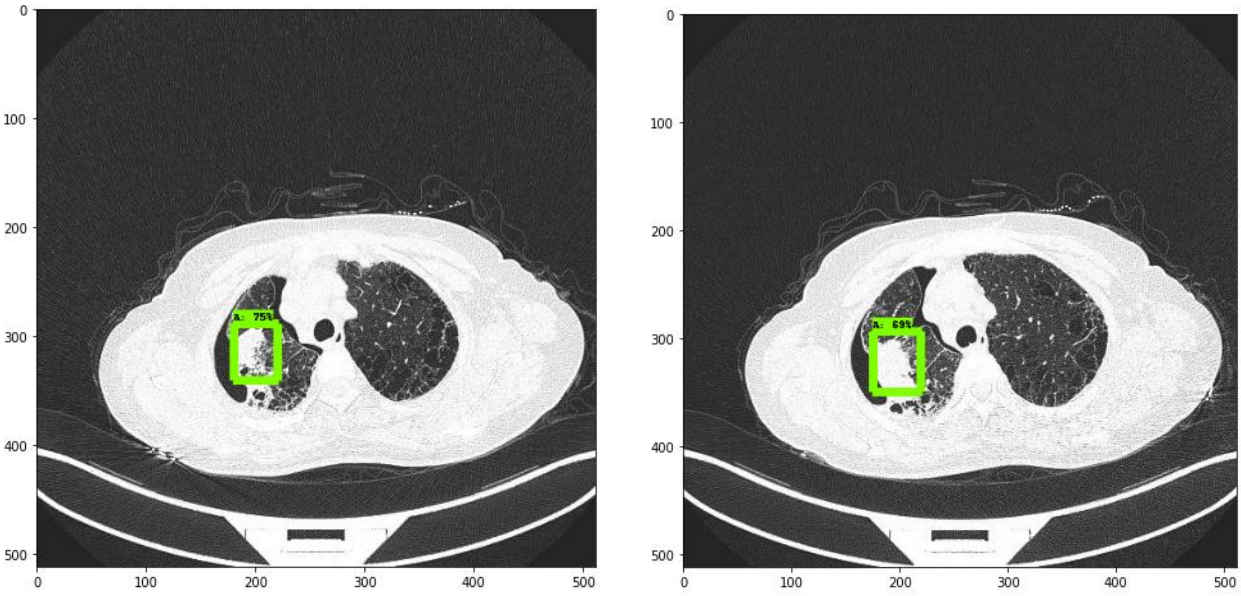
5.4 Αποτελέσματα και συγκρίσεις

Σε αυτή την ενότητα παρουσιάζονται τα καλύτερα αποτελέσματα από κάθε εκδοχή και γίνεται μια σχετική σύγκριση ανάμεσα στις 3 διαφορετικές εκδοχές.



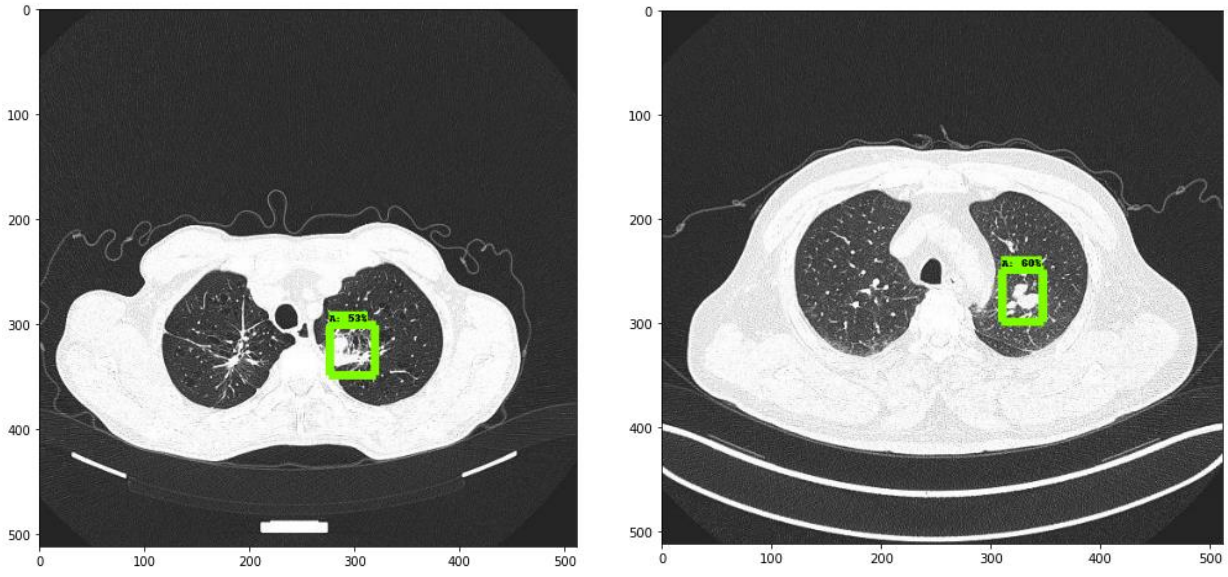
Εικόνα 5.6: *Αποτελέσματα 2ης εκδοχής*

Τα αποτελέσματα της 2ης εκδοχής παρουσιάζουν τιμές Intersection over Union πάνω από 75%, που στις εικόνες απεικονίζεται με το γράμμα Α.



Εικόνα 5.7: *Αποτελέσματα 1ης εκδοχής*

Παρόλο που τα metrics της συγκεκριμένης εκδοχής είναι κάπως χειρότερα από ότι της 2ης παρατηρείται πως τα ποσοστά των RoI είναι επίσης υψηλά. Αυτό οφείλεται στο μέγεθος του dataset και την αντιστοιχία των επαναλήψεων των steps. Καθώς όπως φαίνεται στα αποτελέσματα της 3ης εκδοχής επειδή το dataset είναι πάρα πολύ μεγάλο αλλά οι επαναλήψεις σχετικά με το μέγεθος είναι λίγες διακρίνονται ποσοστά μεγέθους 50-60%.



Εικόνα 5.8: *Αποτελέσματα 3ης εκδοχής*

Κεφάλαιο 6: Συμπεράσματα και Μελλοντικά Σχέδια

Στο κεφάλαιο αυτό παρουσιάζονται τα συμπεράσματα και οι παρατηρήσεις που αφορούν τον αλγόριθμο Faster R-CNN και πιθανές μελλοντικές χρήσεις.

6.1 Συμπεράσματα

Στην συγκεκριμένη διπλωματική εργασία πραγματοποιήθηκε έρευνα που αφορούσε τον εντοπισμό όγκου σε αξονικές τομογραφίες με την μέθοδο του Object Detection. Αποδείχθηκε πως είναι δυνατή η ανάλυση ιατρικών εικόνων με σκοπό την πρόληψη σοβαρότερων καταστάσεων. Η διαδικασία εκπονήθηκε με την βοήθεια μοντέλου βαθιάς μάθησης και πιο συγκεκριμένα με την χρήση συνελικτικών νευρωνικών δικτύων ή αλλιώς CNN.

Ο αλγόριθμος που έλαβε χώρα ήταν ο Faster R-CNN ο οποίος είναι η τελευταία έκδοση της οικογένειας συνελικτικών νευρωνικών δικτύων συνδιάζοντας τα προηγούμενα μοντέλα της οικογένειας CNN & Fast R-CNN. Στόχος του αλγορίθμου είναι ο εντοπισμός αντικειμένων σε εικόνες μέσα από την εκπαίδευση ενός μοντέλου. Τα πιο βασικά πλεονεκτήματα του αλγορίθμου είναι η μεγάλη ακρίβεια που παραθέτει αλλά και οι πολλαπλή ανάλυση καθώς έχει την δυνατότητα επεξεργασίας πολλών εικόνων ταυτόχρονα. Αυτό διακρίνεται και στην συγκεκριμένη διπλωματική εργασία καθώς εκπαιδεύεται μοντέλο με 4751 εικόνες και παραπάνω και παρόλο τον τεράστιο όγκο δεδομένων τα αποτελέσματα είναι εξαιρετικά, πετυχαίνοντας 66% Average Precision στην καλύτερη εκδοχή με 72% Average Recall. Παρόλο που τα νούμερα μπορεί να φαίνονται μικρά παγκόσμια papers αποδεικνύουν πως δεν είναι και τόσο με μέγιστο Average Precision για τον Faster R-CNN να φτάνει το 78%. Όσον αφορά τον χρόνο εκπαίδευσης του ο μέσος όρος των 6 ωρών είναι αρκετά μεγάλος αλλά ανταμείβει τον χρήστη με πολύ καλά αποτελέσματα.

Στην διπλωματική εργασία παρουσιάζονται και ελαφρώς άλλοι αλγόριθμοι που θα μπορούσαν να χρησιμοποιηθούν για την εκάστωτε εργασία αλλά τα μειονεκτήματα του καθενός οδήγησαν στην τελική επιλογή του Faster R-CNN.

Επιπλέον, η διαδικασία την διπλωματικής εργασίας απέδειξε πως μια εξεζητημένη διαδικασία όπως η ανάγνωση αξονικών τομογραφιών γίνεται μια απλή διαδικασία με ελάχιστη προεπεξεργασία, με χαμηλό έως και μηδαμινό κόστος και μπορούν να προσφέρουν χρήσιμες πληροφορίες και στους ειδικούς για μια πιο εύκολη διάγνωση.

Σε μελλοντικά σχέδια, το Faster R-CNN χρειάζεται βελτιώσεις σχετικά με τον χρόνο επεξεργασίας των δεδομένων καθώς είναι υπερβολικά μεγάλος. Η βελτίωση ακόμα περισσότερο των

αποτελεσμάτων θα το εκτοξεύσει στην λίστα των αλγορίθμων για Object Detection καθώς είναι το μεγάλο ατού του αλγορίθμου. Η χρήση του αλγορίθμου από λιγότερο απαιτητικές υπολογιστικές μονάδες θα το κάνει πιο προσητό στο κοινό με αποτέλεσμα να γίνει πιο ευρεία η χρήση του.

Τέλος, μελλοντική δουλειά αποτελεί η χρήση του αλγορίθμου για την ανίχνευση πολλαπλών παθογενιών, όπως η πνευμονία, διαφορετικοί τύποι καρκίνου κ.α. Επιπλέον, σημαντικό ορόσημο θα ήταν και η εφαρμογή της διαδικασία που υλοποιήθηκε στην παρούσα διπλωματική εργασία σε άλλες Ιατρικές εικόνες αλλά και σε άλλους τύπους καρκίνου.

Κατάλογος Σχημάτων

Άνοδος βαθιάς μάθησης στην όραση υπολογιστών (Computer Vision) από τον Μάρτιο 2013 έως το Ιανουάριο 2018 [1] 14
Το Object Detection ως το βασικό βήμα για visual recognition 15
Χρήση Συνελεκτικών Νευρωνικών Δικτύων για Ανίχνευση Αντικειμένων [1] 16
Ανίχνευση Αντικειμένων για βιντεοπαρακολούθηση [2] 18
Ανίχνευση Αντικειμένων για αυτόνομη οδήγηση [3] 18
Αρχιτεκτονική HOG [4] 24
Αρχιτεκτονική ResinaNet [5] 26
Αρχιτεκτονική SSD [6] 27
Αρχιτεκτονική YOLO [7] 28
Αρχιτεκτονική R-CNN [8] 29
Faster R-CNN ένα ενιαίο, ενοποιημένο δίκτυο για Object Detection. Η ενότητα του RPN χρησιμοποιείται ώς η "προσοχή" του ενοποιημένου δικτύου. [9] *33*
Δίκτυο Προτάσεων Περιοχής - Region Proposal Network (RPN) [9] 34
Διαφορετικά σχέδια που αφορούν διαφορετικές κλίμακες και μεγέθη. (α) Πυραμίδα εικόνων και χαρτών χαρακτηριστικών και ο ταξινομητής τρέχει για όλες τις κλίμακες. (β) Πυραμίδα φίλτρων με πολλαπλές κλίμακες/μεγέθη τρέχουν στο feature map. (γ) Χρήση πυραμίδων πλαισίων αναφοράς σε συναρτήσεις παλινδρόμησης [9] 35
Χρόνος λειτουργίας και αποτελέσματα PASCAL VOC 2007 με Fast R-CNN [9] 40
Αποτελέσματα Ανίχνευσης PASCAL VOC 2007. Οι ανιχνευτές είναι Fast R-CNN με ZF αλλα χρησιμοποιώντας διαφορετικές μεθόδους για εκπαίδευσης και εξέταση [9] 41
Αποτελέσματα ανίχνευσης στο σύνολο δοκιμών PASCAL VOC 2012. Ο ανιχνευτής είναι Fast R-CNN και VGG-16 [9] 43
Επιλεγμένα παραδείγματα αποτελεσμάτων ανίχνευσης αντικειμένων στο σύνολο δοκιμών PASCAL VOC 2007 με τη χρήση του Faster R-CNN. Το μοντέλο είναι το VGG-16 και τα δεδομένα εκπαίδευσης είναι τα PASCAL VOC 2007 + 2012 (73,2% mAP στη δοκιμή 2007 σύνολο). [9] 44
Τομείς εξαγωγής Ιατρικών εικόνων 47
Αποτελέσματα ερυθρών και λευκών κυττάρων του αίματος Με μπλε χρώμα απεικονίζονται τα ερυθρά κύτταρα και με πράσινο τα λευκά 50
Αποτελέσματα ερυθρών και λευκών κυττάρων του αίματος Με μπλε χρώμα απεικονίζονται τα ερυθρά κύτταρα και με πράσινο τα λευκά 51
Αξονικές Τομογραφίες Πνεύμονα 51
Εικόνα χωρίς ενισχύση 56

Εικόνα με 90° περιστροφή 56
Εικόνα με Horizontal flip και Saturation 56
Εικόνα με Saturation 56
Ερμηνεία Intersection over Union (IoU) [10] 60
Average Precision 2ης Μεθόδου για το Test Set 61
Average Recall 2ης Μεθόδου για το Test Set 61
Average Precision 2ης Μεθόδου για το Val Set 62
Average Recall 2ης Μεθόδου για το Val Set 62
Αποτελέσματα 2ης εκδοχής 63
Αποτελέσματα 1ης εκδοχής 64
Αποτελέσματα 3ης εκδοχής 64

Κατάλογος Πινάκων

Στρώματα συγκέντρωσης που χρησιμοποιούνται για την Ανίχνευση Εικόνων 16
Πλαίσια βαθιάς μάθησης 17
Σύγκριση μεθόδων ανίχνευσης αντικειμένων με βάση τη βαθιά μάθηση 21
Διαφορετικές εκδοχές διαχωρισμού dataset 54
Δείγματα ενίσχυσης 55
Αξιολόγηση εκπαίδευσης για 3 εκδοχές 63

ΒΙΒΛΙΟΓΡΑΦΙΑ

[1] https://reader.elsevier.com/reader/sd/pii/S1877050918308767?token=9DED6EC0FAF623F3E4253B77098FAB660790A5DC0D70F50A9797A55A12C084A4254BC7E8ED24945F42DFBCD9407A7FD8&originRegion=eu-west-1&originCreation=20220218114254

[2] https://groundup.ai/wp-content/uploads/2021/03/social_distance_detector_people_detections.jpg [Accessed 16 2 2022]

[3] https://i.ytimg.com/vi/ftsUg5VlzIE/maxresdefault.jpg [Accessed 16 2 2022]

[4] https://neptune.ai/blog/object-detection-algorithms-and-libraries [Accessed 16 2 2022]

[5] https://arxiv.org/pdf/1512.02325.pdf [Accessed 10 2 2022]

[6] https://arxiv.org/pdf/1708.02002.pdf [Accessed 8 2 2022]

[7] Redmon, Joseph, Santosh Divvala, Ross Girshick, and Ali Farhadi. 2016. "You Only Look Once: Unified, Real-Time Object Detection." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 779-88.

[8] https://d2l.ai/chapter_computer-vision/rcnn.html

[9] S. Ren, K.He, R.Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in Neural Information Processing Systems (NIPS), 2015.

[10] https://jonathan-hui.medium.com/map-mean-average-precision-for-object-detection-45c121a31173 [Accessed 23 1 2022]

[11] K.He, X.Zhang, S.Ren, and J.Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," in European Conference on Computer Vision (ECCV), 2014.

[12] R.Girshick, "Fast R-CNN," in IEEE International Conference on Computer Vision (ICCV), 2015.

[13] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in International Conference on Learning Representations (ICLR), 2015.

[14] Voulodimos, A., Protopapadakis, E., Katsamenis, I., Doulamis, A., & Doulamis, N. (2021, June). Deep learning models for COVID-19 infected area segmentation in CT images. In The 14th PErvasive Technologies Related to Assistive Environments Conference (pp. 404-411).

[15] Voulodimos, A., Protopapadakis, E., Katsamenis, I., Doulamis, A., & Doulamis, N. (2021). A few-shot U-net deep learning model for COVID-19 infected area segmentation in CT images. Sensors, 21(6), 2215.

[16] Katsamenis, I., Protopapadakis, E., Voulodimos, A., Doulamis, A., & Doulamis, N. (2020, November). Transfer learning for COVID-19 pneumonia detection and classification in chest X-ray images. In 24th Pan-Hellenic Conference on Informatics (pp. 170-174).

[17] Protonotarios, Nicholas E., et al. "A few-shot U-Net deep learning model for lung cancer lesion segmentation via PET/CT imaging." Biomedical Physics & Engineering Express (2022).

[18] J.R.Uijlings, K.E.van de Sande, T. Gevers, and A. W. Smeulders, "Selective search for object recognition," International Journal of Computer Vision (IJCV), 2013.

[19] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic seg mentation," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014.

[20] C.L.Zitnick and P.Doll´ar, "Edge boxes: Locating object proposals from edges," in European Conference on Computer Vision (ECCV), 2014.

[21] J.Long, E.Shelhamer, and T.Darrell, "Fully convolutional networks for semantic segmentation," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

[22] P.F.Felzenszwalb, R.B.Girshick, D.McAllester, and D.Ramanan, "Object detection ith discriminatively trained part based models," IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2010.

[23] P.Sermanet, D.Eigen, X.Zhang, M.Mathieu, R.Fergus, and Y. LeCun, "Overfeat: Integrated recognition, localization and detection using convolutional networks," in International Conference on Learning Representations (ICLR), 2014.

[24] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results," 2007.

[25] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Doll´ar, and C. L. Zitnick, "Microsoft COCO: Common Objects in Context," in European Conference on Computer Vision (ECCV), 2014.

[26] S. Song and J. Xiao, "Deep sliding shapes for amodal 3d object detection in rgb-d images," arXiv:1511.02300, 2015.

[27] J. Zhu, X. Chen, and A. L. Yuille, "DeePM: A deep part-based model for object detection and semantic part localization," arXiv:1511.07131, 2015.

[28] J. Dai, K. He, and J. Sun, "Instance-aware semantic segmentation via multi-task network cascades," arXiv:1512.04412, 2015.

[29] J. Johnson, A. Karpathy, and L. Fei-Fei, "Densecap: Fully convolutional localization networks for dense captioning," arXiv:1511.07571, 2015.

[30] D. Kislyuk, Y. Liu, D. Liu, E. Tzeng, and Y. Jing, "Human curation and convnets: Powering item-to-item recommendations on pinterest," arXiv:1511.04003, 2015.

[31] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," arXiv:1512.03385, 2015.

[32] J. Hosang, R. Benenson, and B. Schiele, "How good are de tection proposals, really?" in British Machine Vision Conference (BMVC), 2014.

[33] J. Hosang, R. Benenson, P. Doll´ar, and B. Schiele, "What makes for effective detection proposals?" IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2015.

[34] N. Chavali, H. Agrawal, A. Mahendru, and D. Batra, "Object-Proposal Evaluation Protocol is 'Gameable'," arXiv 1505.05836, 2015.

[35] J. Carreira and C. Sminchisescu, "CPMC: Automatic object segmentation using constrained parametric min-cuts," IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2012.

[36] P. Arbel´aez, J. Pont-Tuset, J. T. Barron, F. Marques, and J. Malik, "Multiscale combinatorial grouping," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014.

[37] B. Alexe, T. Deselaers, and V. Ferrari, "Measuring the objectness of image windows," IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2012.

[38] C. Szegedy, A. Toshev, and D. Erhan, "Deep neural networks for object detection," in Neural Information Processing Systems (NIPS), 2013.

[39] D. Erhan, C. Szegedy, A. Toshev, and D. Anguelov, "Scalable object detection using deep neural networks," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014.

[40] C. Szegedy, S. Reed, D. Erhan, and D. Anguelov, "Scalable, high-quality object detection," arXiv:1412.1441 (v1), 2015.

[41] P. O. Pinheiro, R. Collobert, and P. Dollar, "Learning to segment object candidates," in Neural Information Processing Systems (NIPS), 2015.

[42] J. Dai, K. He, and J. Sun, "Convolutional feature masking for joint object and stuff segmentation," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

[43] S. Ren, K. He, R. Girshick, X. Zhang, and J. Sun, "Object detection networks on convolutional feature maps," arXiv:1504.06066, 2015.

[44] J. K. Chorowski, D. Bahdanau, D. Serdyuk, K. Cho, and Y. Bengio, "Attention-based models for speech recognition," in Neural Information Processing Systems (NIPS), 2015.

[45] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional neural networks," in European Conference on Computer Vision (ECCV), 2014.

[46] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in International Conference on Machine Learning (ICML), 2010.

[47] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, and A. Rabinovich, "Going deeper with convolutions," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

[48] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," Neural computation, 1989.

[49] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," in International Journal of Computer Vision (IJCV),
2015.

[50] Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classi fication with deep convolutional neural networks," in Neural Information Processing Systems (NIPS), 2012.

[51] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," arXiv:1408.5093, 2014.

[52] K. Lenc and A. Vedaldi, "R-CNN minus R," in British Machine Vision Conference (BMVC), 2015

[53] Technology Trends, https://www.gartner.com/smarterwithgartner/gartner-top-10- trategic-technology-trends-for-2018.

[54] Krizhevsky Alex, Ilya Sutskever, GeoffreyE. Hinton Imagenet Classification with Deep Convolutional Neural Networks. Advances in Neural Information Processing Systems (2012), pp. 1097-1105

[55] Ren Shaoqing, Kaiming He, Ross Girshick, Jian Sun Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks.

[56] IEEE transactions on pattern analysis and machine intelligence, 39 (6) (2017), pp. 1137-1149

[57] Caffe, http://caffe.berkeleyvision.org/.

[58] Microsoft Cognitive Toolkit CNTK, https://www.microsoft.com/en-us/research/product/cognitive-toolkit/.

[59] TensorFlow, https://www.tensorflow.org/.

[60] Theano, http://deeplearning.net/software/theano/.

[61] Torch, http://torch.ch/.

[62] Chainer, http://chainer.org/.

[63] Keras, https://keras.io/.

[64] Deeplearning4j, https://deeplearning4j.org.

[65] Apache Singa, http://singa.incubator.apache.org/.
[66] MXnet, http://mxnet.io/.
[67] Neon, http://neon.nerva-nasys.com/docs/latest.
[68] Clarifai, https://clarifai.com/.
[69] Google Cloud Vision API, https://cloud.google.com/vision/.
[70] Microsoft Cognitive Service, https://www.microsoft.com/cognitive-services/en-us/computer-vision-api.
[71] IBM Watson Vision Recognition Service,
[72] Amazon Rekognition, https://aws.amazon.com/rekognition/.
[73] CloudSight, https://cloudsight.readme.io/v1.0/docs.
[74] Lin, Tsung-Yi et al. (2014). "Microsoft Coco: Common Objects in Context." In European Conference on Computer Vision, 740-55.
[75] Deng Jia Imagenet: A Large-Scale Hierarchical Image Database. Computer Vision and Pattern Recognition (2009), pp. 248-255
[76] Torralba Antonio, Rob Fergus, WilliamT. Freeman 80 Million Tiny Images: A Large Data Set for Nonparametric Object and Scene Recognition. IEEE transactions on pattern analysis and machine intelligence, 30 (11) (2008), pp. 1958-1970
[77] Krizhevsky, Alex, and Geoffrey Hinton. (2009). "Learning Multiple Layers of Features from Tiny Images." Thesis ch.3.
[78] Wah, Catherine et al. (2011). "The Caltech-Ucsd Birds-200-2011 Dataset.".
[79] Welinder P., Branson S., Mita T., Wah C., Schroff F., Belongie S., Perona P. Caltech-UCSD Birds 200, California Institute of Technology (2010) CNS-TR-2010-001.
[80] Griffin, Gregory, Alex Holub, and Pietro Perona. (2007). "Caltech-256 Object Category Dataset.".
[81] Russakovsky Olga ImageNet Large Scale Visual Recognition Challenge. Int. Journal of CV, 115 (3) (2015), pp. 211-252
[82] Everingham Mark The Pascal Visual Object Classes Challenge: A Retrospective. Int. journal of CV, 111 (1) (2015), pp. 98-136
[83] Russell Bryan C, Antonio Torralba, KevinP Murphy, WilliamT Freeman LabelMe: A Database and Web-Based Tool for Image Annotation. International journal of computer vision, 77 (1-3) (2008), pp. 157-173
[84] Xiao Jianxiong Sun Database: Large-Scale Scene Recognition from Abbey to Zoo., CVPR (2010), pp. 3485-3492
[85] Chang, Wo L. (2015). NIST Big Data Interoperability Framework: Volume 3, Use Cases and General Requirements.
[86] Szegedy Christian, Alexander Toshev, Dumitru Erhan Deep Neural Networks for Object Detection. Advances in Neural Information Processing Systems (2013), pp. 2553-2561
[87] Erhan, Dumitru, Christian Szegedy, Alexander Toshev, and Dragomir Anguelov. (2014). "Scalable Object Detection Using Deep Neural Networks." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2147-54.

[88] Zeiler, Matthew D, and Rob Fergus. (2014). "Visualizing and Understanding Convolutional Networks." In European Conference on Computer Vision, 818-33.

[89] Sermanet, Pierre et al. (2013). "Overfeat: Integrated Recognition, Localization and Detection Using Convolutional Networks." arXiv preprint arXiv:1312.6229.

[90] Girshick, Ross, Jeff Donahue, Trevor Darrell, and Jitendra Malik. (2014). "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 580-87.

[91] Wang Xiaoyu, Ming Yang, Shenghuo Zhu, Yuanqing Lin Regionlets for Generic Object Detection. IEEE transactions on pattern analysis and machine intelligence, 37 (10) (2015), pp. 2071-2084

[92] Felzenszwalb Pedro F, RossB Girshick, David McAllester, Deva Ramanan Object Detection with Discriminatively Trained Part-Based Models. IEEE transactions on pattern analysis and machine intelligence, 32 (9) (2010), pp. 1627-1645

[93] Ouyang, W et al. (2015). "DeepID-Net: Deformable Deep Convolutional Neural Networks for Object Detection." In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2403-2412.

[94] Huang Chen, Zhihai He, Guitao Cao, Wenming Cao Task-Driven Progressive Part Localization for Fine-Grained Object Recognition. IEEE Transactions on Multimedia, 18 (12) (2016), pp. 2372-2383

[95] Huang Chen, Zhihai He, Guitao Cao, Wenming Cao Task-Driven Progressive Part Localization for Fine-Grained Object Recognition. IEEE Transactions on Multimedia, 18 (12) (2016), pp. 2372-2383

[96] Ohn-Bar Eshed, Mohan Manubhai Trivedi Multi-Scale Volumes for Deep Object Detection and Localization. Pattern Recognition, 61 (2017), pp. 557-572

[97] Girshick, Ross. (2015). "Fast R-Cnn." arXiv preprint arXiv:1504.08083.

[98] Ren Shaoqing, Kaiming He, Ross Girshick, Jian Sun Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE transactions on pattern analysis and machine intelligence, 39 (6) (2017), pp. 1137-1149

[99] Kim, Kye-Hyeon et al. (2016). "PVANET: Deep but Lightweight Neural Networks for Real-Time Object Detection." arXiv preprint arXiv:1608.08021.

[100] Liu, Nian, and Junwei Han. (2016). "Dhsnet: Deep Hierarchical Saliency Network for Salient Object Detection." In Computer Vision and Pattern Recognition.

[101] Li Xi Deepsaliency: Multi-Task Deep Neural Network Model for Salient Object Detection. IEEE Transactions on Image Processing, 25 (8) (2016), pp. 3919-3930

[102] Wang Lijun, Huchuan Lu, Xiang Ruan, Yang Ming-Hsuan Deep Networks for Saliency Detection via Local Estimation and Global Search. Computer Vision and Pattern Recognition (CVPR) (2015), pp. 183-192

[103] Li, Guanbin, and Yizhou Yu. (2016). "Deep Contrast Learning for Salient Object Detection." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 478-87.

[104] Bojarski, Mariusz et al. (2016). "End to End Learning for Self-Driving Cars." arXiv preprint arXiv:1604.07316.

[105] He Kaiming, Xiangyu Zhang, Shaoqing Ren, Jian Sun Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. IEEE transactions on pattern analysis and machine intelligence, 37 (9) (2015), pp. 1904-1916

[106] Yang, Fan, Wongun Choi, and Yuanqing Lin. (2016). "Exploit All the Layers: Fast and Accurate Cnn Object Detector with Scale Dependent Pooling and Cascaded Rejection Classifiers." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2129-37.

[107] Denton Emily L, Soumith Chintala, Rob Fergus Deep Generative Image Models Using A Laplacian Pyramid of Adversarial Networks. Advances in Neural Information Processing Systems (2015), pp. 1486-1494

[108] Shrivastava, Abhinav, Abhinav Gupta, and Ross Girshick. (2016). "Training Region-Based Object Detectors with Online Hard Example Mining." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 761-69.

[109] Takác Martin, Avleen Singh Bijral, Peter Richtárik, Nati Srebro Mini-Batch Primal and Dual Methods for SVMs. ICML (3) (2013), pp. 1022-1030

[110] Wang, Xiaolong, Abhinav Shrivastava, and Abhinav Gupta. (2017). "A-Fast-Rcnn: Hard Positive Generation via Adversary for Object Detection." arXiv preprint arXiv:1704.03414 2.

[111] Girshick, Ross, Forrest Iandola, Trevor Darrell, and Jitendra Malik. (2015). "Deformable Part Models Are Convolutional Neural Networks." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 437-46.

[112] Wan, Li, David Eigen, and Rob Fergus. (2015). "End-to-End Integration of a Convolution Network, Deformable Parts Model and Non-Maximum Suppression." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 851-59.

[113] Girshick, Ross, Forrest Iandola, Trevor Darrell, and Tian, Yonglong, Ping Luo, Xiaogang Wang, and Xiaoou Tang. (2015). "Deep Learning Strong Parts for Pedestrian Detection." In Proceedings of the IEEE International Conference on Computer Vision, 1904-12.

[114] Chai, Yuning, Victor Lempitsky, and Andrew Zisserman. (2013). "Symbiotic Segmentation and Part Localization for Fine-Grained Categorization." In Computer Vision (ICCV), 2013 IEEE International Conference on, 321-28.

[115] Göring Christoph, Erik Rodner, Alexander Freytag, Joachim Denzler Nonparametric Part Transfer for Fine-Grained Recognition., CVPR (2014), p. 7

[116] Lin, Di, Xiaoyong Shen, Cewu Lu, and Jiaya Jia. (2015). "Deep Lac: Deep Localization, Alignment and Classification for Fine-Grained Recognition." In Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on, 1666-74.

[117] Zhang, Ning, Jeff Donahue, Ross Girshick, and Trevor Darrell. (2014). "Part-Based R-CNNs for Fine-Grained Category Detection." In European Conference on Computer Vision, 834-49.

[118] Redmon, J, and A Farhadi. (2017). “YOLO9000: Better, Faster, Stronger.” In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 6517-25.

[119] Shih, Ya-Fang et al. (2017). “Deep Co-Occurrence Feature Learning for Visual Object Recognition.” In Proc. Conf. Computer Vision and Pattern Recognition.