ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ
ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΚΑΙ ΗΛΕΚΤΡΟΝΙΚΩΝ ΜΗΧΑΝΙΚΩΝ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΒΙΟΜΗΧΑΝΙΚΗΣ ΣΧΕΔΙΑΣΗΣ ΚΑΙ ΠΑΡΑΓΩΓΗΣ
http://www.eee.uniwa.gr
http://www.idpe.uniwa.gr
Θηβών 250, Αθήνα-Αιγάλεω 12241
Τηλ: +30 210 538-1614
Διατμηματικό Πρόγραμμα Μεταπτυχιακών Σπουδών
Τεχνητή Νοημοσύνη και Βαθιά Μάθηση
https://aidl.uniwa.gr/

UNIVERSITY OF WEST ATTICA
FACULTY OF ENGINEERING
DEPARTMENT OF ELECTRICAL & ELECTRONICS ENGINEERING
DEPARTMENT OF INDUSTRIAL DESIGN AND PRODUCTION ENGINEERING
http://www.eee.uniwa.gr
http://www.idpe.uniwa.gr
250, Thivon Str., Athens, GR-12241, Greece
Tel: +30 210 538-1614
Master of Science in
Artificial Intelligence and Deep Learning
https://aidl.uniwa.gr/

Master of Science Thesis

Road Signs Detection using A.I.

Space to insert image, picture or diagram related to the thesis subject (optional)

Student: Vyros, Kyriakos
Registration Number: AIDL-0003

MSc Thesis Supervisor

Piromalis, Dimitrios
Assist. Professor

ATHENS-EGALEO, September 2022

ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ
ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΚΑΙ ΗΛΕΚΤΡΟΝΙΚΩΝ ΜΗΧΑΝΙΚΩΝ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΒΙΟΜΗΧΑΝΙΚΗΣ ΣΧΕΔΙΑΣΗΣ ΚΑΙ ΠΑΡΑΓΩΓΗΣ
http://www.eee.uniwa.gr
http://www.idpe.uniwa.gr
Θηβών 250, Αθήνα-Αιγάλεω 12241
Τηλ: +30 210 538-1614
Διατμηματικό Πρόγραμμα Μεταπτυχιακών Σπουδών
Τεχνητή Νοημοσύνη και Βαθιά Μάθηση
https://aidl.uniwa.gr/

UNIVERSITY OF WEST ATTICA
FACULTY OF ENGINEERING
DEPARTMENT OF ELECTRICAL & ELECTRONICS ENGINEERING
DEPARTMENT OF INDUSTRIAL DESIGN AND PRODUCTION ENGINEERING
http://www.eee.uniwa.gr
http://www.idpe.uniwa.gr
250, Thivon Str., Athens, GR-12241, Greece
Tel: +30 210 538-1614
Master of Science in
Artificial Intelligence and Deep Learning
https://aidl.uniwa.gr/

Μεταπτυχιακή Διπλωματική Εργασία

Αναγνώριση Οδικής Σήμανσης με Τ.Ν.

Χώρος για εικόνα/σχήμα/διάγραμμα σχετικό με την εργασία (προαιρετικά)

Φοιτητής: Βύρος, Κυριάκος
ΑΜ: AIDL-0003

Επιβλέπων Καθηγητής
Πυρομάλης, Δημήτριος
Επίκ. Καθηγητής

ΑΘΗΝΑ-ΑΙΓΑΛΕΩ, Σεπτέμβριος 2022

This MSc Thesis has been accepted, evaluated, and graded by the following committee:

Supervisor	Member	Member
Dimitrios Pyromalis Digitally signed by Dimitrios Pyromalis Date: 2022.10.19 10:28:02 +03'00'	Panagiotis Papageorgas Digitally signed by Panagiotis Papageorgas Date: 2022.10.19 17:41:20 +03'00'	Digitally signed by Georgios Priniotakis Date: 2022.10.20 18:44:15 +03'00'
Pyromalis Dimitrios	Papageorgas Panagiotis	Priniotakis Georgios
Assistant Professor	Professor	Professor
Electrical & Electronics Engineering Department	Electrical & Electronics Engineering Department	Industrial Design & Production Engineering Department
University of West Attica	University of West Attica	University of West Attica

ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΜΕΤΑΠΤΥΧΙΑΚΗΣ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

Ο κάτωθι υπογεγραμμένος Κυριάκος Βύρος του Κωνσταντίνου, με αριθμό μητρώου AIDL-0003 μεταπτυχιακός φοιτητής του ΔΠΜΣ «Τεχνητή Νοημοσύνη και Βαθιά Μάθηση» του Τμήματος Ηλεκτρολόγων και Ηλεκτρονικών Μηχανικών και του Τμήματος Μηχανικών Βιομηχανικής Σχεδίασης και Παραγωγής, της Σχολής Μηχανικών του Πανεπιστημίου Δυτικής Αττικής,

δηλώνω υπεύθυνα ότι:

«Είμαι συγγραφέας αυτής της μεταπτυχιακής διπλωματικής εργασίας και κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, οι όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε ακριβώς είτε παραφρασμένες, αναφέρονται στο σύνολό τους, με πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο. Επίσης, βεβαιώνω ότι αυτή η εργασία έχει συγγραφεί από μένα αποκλειστικά και αποτελεί προϊόν πνευματικής ιδιοκτησίας τόσο δικής μου, όσο και του Ιδρύματος. Η εργασία δεν έχει κατατεθεί στο πλαίσιο των απαιτήσεων για τη λήψη άλλου τίτλου σπουδών ή επαγγελματικής πιστοποίησης πλην του παρόντος.
Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του διπλώματός μου.
Επιθυμώ την απαγόρευση πρόσβασης στο πλήρες κείμενο της εργασίας μου μέχρι …… και έπειτα από αίτησή μου στη Βιβλιοθήκη και έγκριση του επιβλέποντος καθηγητή.»

Ο Δηλών
Κυριάκος Βύρος

Declaration of the author of this MSc thesis

I, Kyriakos, Konstantinos, Vyros with the following student registration number: AIDL-0003, postgraduate student of the MSc programme in "Artificial Intelligence and Deep Learning", which is organized by the Department of Electrical and Electronic Engineering and the Department of Industrial Design and Production Engineering of the Faculty of Engineering of the University of West Attica, hereby declare that:

I am the author of this MSc thesis and any help I may have received is clearly mentioned in the thesis. Additionally, all the sources I have used (e.g., to extract data, ideas, words or phrases) are cited with full reference to the corresponding authors, the publishing house or the journal; this also applies to the Internet sources that I have used. I also confirm that I have personally written this thesis and the intellectual property rights belong to myself and to the University of West Attica. This work has not been submitted for any other degree or professional qualification except as specified in it.

Any violations of my academic responsibilities, as stated above, constitutes substantial reason for the cancellation of the conferred MSc degree.

I wish to deny access to the full text of my MSc thesis until ……., following my application to the Library of UNIWA and the approval from my supervisor.

The author
Kyriakos Vyros

Abstract

In a world that is constantly changing and rapidly moving, artificial intelligence is making ever greater leaps forward. The use of deep learning in more and more areas is now a fact. Ever-faster algorithms, graphics cards and response times are emerging due to the growing demand to meet needs in many areas. This unstoppable momentum for volume reduction and the gigantic potential at the same time, has led processor production companies to a frantic struggle of achievements. From this "rally" could not be missing industries that demand continuous technological development of their products, as well as the leading companies for the development of such technologies. On one podium stands the car industry proudly and on the other companies such as Nvidia, respectively. Autonomous vehicles are a realization of a dream of many decades that only recently began to take shape. The road to the complete autonomy of a vehicle may still be far away, but the distance traveled so far is already remarkable and with quite encouraging results. The identification of road signs, pedestrians, obstacles, and other elements of the vehicle's surroundings are key points for the development and achievement of full autonomy, but above all greater road safety for drivers and non-drivers alike. Artificial intelligence, then, plays the role of the redeemer, which rises levels, rapidly. It overcomes every obstacle over time, but all this could not have happened without the parallel development of appropriate firmware, capable of coping with the constantly well-sounding computational requirements. The way and method of identifying objects in a car depends on what kind of object is to be detected. They are available with laser technologies, radar and more. In the case of road signage recognition, the only candidate mode is the cameras. Deep learning undertook the transformation of simple cameras into computer vision and cars into "thinking" machines. New methods, smaller codes, updated and demanding data sets and GPUs ready to respond to this current, arise continuously, endlessly. After some basic concepts and terminologies are understood, in terms of detection and recognition, the procedures for analyzing the data and algorithms are described. The tracking strategy concerns distinction in shapes, colors and their combination. In the part of the methods of approaching the issue of road marking recognition, algorithms and improvement techniques are proposed. Of course, the backbone of these systems, the convolutional neural network as well as the collaborating frameworks are analyzed extensively. Structures, architectures, trends, speeds as well as problems and lags, receive extensive analysis and are interpreted. So here comes Jetson Nano and a beautiful learning challenge. That is the part of the implementation. It is very interesting to program such a small computer system, and to be able to do something that until a few years ago, entire desktop computers with oversized graphic cards could not do, is very interesting. Collecting thousands of photos, commenting on them correctly and the appropriate partition in training, testing and validation sets, is a time-consuming and demanding process. It is also demanding to use a processor with 128 CUDA cores and just 2 GB of RAM, to complete such a large computing load. So, using the familiar environment of Jetson Nano, after creating appropriate data sets, and many trainings of different models, road sign recognition was achieved. This, of course, occurred under conditions and limitations concerning, methods and data used.

Keywords

Road signs detection, object recognition, deep learning, artificial intelligence, autonomous vehicles, convolutional neural networks.

Περίληψη

Σε έναν κόσμο συνεχώς μεταβαλλόμενο και γρηγορά κινούμενο, η τεχνητή νοημοσύνη κάνει άλματα προόδου ολοένα και μεγαλύτερα. Η χρήση της βαθιάς μάθησης (DL) σε συνεχώς περισσοτέρους τομείς, είναι πλέον γεγονός. Ολοένα ταχύτεροι αλγόριθμοι, κάρτες γραφικών και χρόνοι απόκρισης προκύπτουν λόγω της αυξανομένης ζήτησης για κάλυψη αναγκών σε πολλούς τομείς. Αυτή η ασταμάτητη ορμή για μείωση όγκου και γιγάντωση δυνατοτήτων ταυτόχρονα, έχει οδηγήσει τις εταιρίες παραγωγής επεξεργαστών σε ένα ξέφρενο αγώνα επιτευγμάτων. Από αυτό το «ράλι» δεν θα μπορούσαν να λείπουν βιομηχανίες που ζητούν συνεχή τεχνολογική εξέλιξη των προϊόντων τους, καθώς και οι κορυφαίες εταιρίες ανάπτυξης τέτοιων τεχνολογιών. Στο ένα βάθρο στέκεται περίφανη η αυτοκινητοβιομηχανία και στο άλλο εταιρίες όπως η Nvidia, αντίστοιχα. Τα αυτόνομα οχήματα είναι μια πραγματοποίηση ενός ονείρου πολλών δεκαετιών που μόλις πρόσφατα όμως άρχισε να παίρνει σάρκα και οστά. Ο δρόμος μέχρι την πλήρη αυτονόμηση ενός οχήματος, ίσως είναι μακριά ακόμα, αλλά η απόσταση που έχει διανυθεί ως εδώ είναι ήδη αξιοσημείωτη και με αρκετά ενθαρρυντικά αποτελέσματα. Η αναγνώριση των οδικών σημάνσεων, των πεζών, των εμποδίων και των λοιπών στοιχείων του περιβάλλοντος χώρου του οχήματος, είναι σημεία κλειδιά για την εξέλιξη και επίτευξη πλήρους αυτονομίας, αλλά κυρίως μεγαλύτερης οδικής ασφάλειας για οδηγούς και μη. Η τεχνητή νοημοσύνη λοιπόν, διαδραματίζει τον ρολό του λυτρωτή, που ανεβαίνει επίπεδα, τάχιστα. Ξεπερνά κάθε εμπόδιο με τη πάροδο του χρόνου, όμως όλα αυτά δεν θα μπορούσαν να συμβούν χωρίς την παράλληλη ανάπτυξη κατάλληλου υλικολογισμικού, ικανού να ανταπεξέρχεται στις συνεχώς καλπάζουσες υπολογιστικές απαιτήσεις. Ο τρόπος και η μέθοδος αναγνώρισης αντικειμένων σε ένα αυτοκίνητο εξαρτάται από το είδος αντικειμένου που πρόκειται να ανιχνευθεί. Διατίθενται μέσα με τεχνολογίες λέιζερ, ραντάρ και άλλα. Στην περίπτωση αναγνώρισης οδικής σήμανσης, ο μονός υποψήφιος τρόπος, είναι οι κάμερες. Η βαθιά μάθηση ανέλαβε την μετατροπή των απλών καμερών σε υπολογιστική όραση και τα αυτοκίνητα σε «σκεπτόμενες» μηχανές. Νέες μέθοδοι, μικρότεροι κώδικες, ανανεωμένα και απαιτητικά σύνολα δεδομένων και GPUs έτοιμες να ανταποκριθούν στο ρεύμα αυτό, προκύπτουν συνεχώς και ασταμάτητα. Αφού γίνουν κατανοητές κάποιες βασικές έννοιες και ορολογίες, όσον αφορά τον εντοπισμό και την ανίχνευση, περιγράφονται οι διαδικασίες ανάλυσης των δεδομένων και οι αλγόριθμοι. Η στρατηγική εντοπισμού αφορά διάκριση σε σχήματα, χρώματα και συνδυασμό τους. Στο κομμάτι των μεθόδων προσέγγισης του ζητήματος της αναγνώρισης οδικών σημάνσεων, προτείνονται αλγόριθμοι και τεχνικές βελτίωσης. Βεβαίως αναλύεται εκτενώς, η ραχοκοκαλιά των συστημάτων αυτών, το συνελικτικό νευρωνικό δίκτυο καθώς και τα συνεργαζόμενα πλαίσια. Οι δομές, οι αρχιτεκτονικές, οι τάσεις, οι ταχύτητες καθώς και τα προβλήματα και οι υστερήσεις, λαμβάνουν εκτενή ανάλυση και ερμηνεύονται. Εδώ λοιπόν έρχεται το Jetson Nano και μια όμορφη μαθησιακή πρόκληση. Αυτό αποτελεί το κομμάτι της υλοποίησης. Το να προγραμματίσει κάποιος ένα τόσο μικρο υπολογιστικό σύστημα, και να μπορεί να κάνει κάτι που μέχρι πριν μερικά χρονιά δεν μπορούσαν ολόκληροι σταθεροί υπολογιστές με υπερμεγέθεις κάρτες γραφικών, είναι πολύ ενδιαφέρον. Η συλλογή χιλιάδων φωτογραφιών, ο σωστός σχολιασμός τους και το κατάλληλο χώρισμα σε σετ εκπαίδευσης, δοκιμής και επικύρωσης, είναι χρονοβόρα και απαιτητική διαδικασία. Όπως επίσης απαιτητικό είναι και το να χρησιμοποιηθεί ένας επεξεργαστής με 128 πυρήνες CUDA και μόλις 2 GB μνημης RAM, για την περάτωση τόσο μεγάλου υπολογιστικού φόρτου. Χρησιμοποιώντας λοιπόν το οικείο περιβάλλον του Jetson nano, μετα από δημιουργία κατάλληλων σετ δεδομένων, και πολλές εκπαιδεύσεις διαφορετικών μοντέλων επετεύχθη η

αναγνώριση οδικής σήμανσης. Αυτό, βέβαια, συνέβη υπο ορούς και περιορισμούς που αφορούν, μεθόδους και δεδομένα που χρησιμοποιήθηκαν.

Λέξεις – κλειδιά

Αναγνώριση οδικής σήμανσης, αναγνώριση αντικειμένων, βαθιά μάθηση, τεχνητή νοημοσύνη, αυτόνομα οχήματα, συνελικτικά νευρωνικά δίκτυα.

Πίνακας Περιεχομένων

Λίστα σχημάτων

Ακρωνύμια

ΟΣ: Οδική Σήμανση

ΚΟΚ: Κώδικας Οδικής Κυκλοφορίας

ΤΝ: Τεχνίτη Νοημοσύνη

ADAS: Advanced driver-assistance system

ADS: Automated Driving Systems

AI: Artificial Intelligence

API: Application Programming Interface

BGD: Batch Gradient Descent

CIFAR: Canadian Institute for Advanced Research

CPU: Central Processing Unit

CSI: Camera Serial Interface

CNN: Convolutional Neural Network

cuDNN: CUDA Deep Neural Network

CUDA: Compute Unified Device Architecture

CV: Computer Vision

DNN: Deep Neural Network

DL: Deep Learning

FPN: Feature Pyramid Network

FOV: Field Of View

FC: Fully Connected

GIOU: Generalized Intersection over Union

GPU: Graphics Processing Unit

HOG: Histogram of Oriented Gradients

IoU: Intersection Over Union

LBP: Local Binary Patterns

MSE: Mean Square Error

MLP: Multilayer Perceptron

MBGD: Mini Batch Gradient Descent

MCT: Microsoft Computer Terminal

MNIST: Modified National Institute of Standards and Technology

ML: Machine Learning

MS COCO: Microsoft Common Objects in Context

NMS: Non Maximum Supression

NLP: NeuroLinguistic Programming

ONNX: Open Neural Network Exchange

RAM: Random Access Memory

ReLu: Rectified Linear Activation Unit

RGB: Red Green Blue

R-CNN: Region-based Convolutional Neural Network

RPN: Region Proposal Network

ROI: Region of Interest

NN: Neural Network

SAE: Society of Automotive Engineers

SDK: Software Development Kit

SGD: Stochastic Gradient Descent

SIFT: Scale Invariant Feature Transform

SPP: Spatial Pyramid Pooling

SSD: Single Shot Detector

SVHN: Street View House Numbers

SVM: Support Vector Machine

TSR: Traffic Sign Recognition

VGG: Visual Geometry Group

VOC: Visual Object Classes

YOLO: You Only Look Once

YOLOR: You Only Learn One Representation

YOLOX: You Only Look Once X

ΕΙΣΑΓΩΓΗ

Τα συστήματα αναγνώρισης οδικής σήμανσης (ΟΣ) αποτελούν αξιοσημείωτο θέμα για τη μείωση του κυκλοφοριακού φόρτου και την ενίσχυση της βιωσιμότητας των αυτοοδηγούμενων οχημάτων χωρίς ατυχήματα. Η αναγνώριση των σωστών σημάνσεων τη σωστή στιγμή και στο σωστό μέρος είναι σημαντική για τους οδηγούς αυτοκινήτων, για τους επιβάτες τους αλλά και τους πεζούς που διασχίζουν το δρόμο εκείνη τη στιγμή. Μεγάλο ποσοστό οχημάτων οδηγείται χωρίς την τήρηση των κανόνων οδικής κυκλοφορίας (Κ.Ο.Κ), οπότε υπάρχει όλο και περισσότερη κίνηση στο δρόμο. Η έντονη κίνηση χωρίς απόλυτη τήρηση του Κ.Ο.Κ., είναι μία από τις κύριες αιτίες ατυχημάτων κάθε χρόνο. Τελευταία, τα τροχαία ατυχήματα αυξάνονται σε όλο τον κόσμο. Η κυρίαρχη αιτία των μέγιστων τραυματισμών λόγω τροχαίων ατυχημάτων είναι η αγνόηση των οδικών σημάτων. Στόχος του προτεινόμενου συστήματος είναι η ανάπτυξη ενός σύνθετου αυτόνομου συστήματος για την προώθηση της ασφαλέστερης συμπεριφοράς των χρηστών του οδικού δικτύου και την αποφυγή ή ελαχιστοποίηση των ατυχημάτων στο δρόμο.

Στο οδικό δίκτυο, τα σήματα κυκλοφορίας διαδραματίζουν ουσιαστικό ρόλο [1]. Οι πρωταρχικοί σκοποί της ΟΣ είναι να εμφανίζουν το περιεχόμενο που πρέπει να παρατηρείται στα σύγχρονα τμήματα των οδών, να προειδοποιούν τους οδηγούς μπροστά από το δρόμο για την απειλή και τα προβλήματα στο περιβάλλον, να υπενθυμίζουν στους οδηγούς να οδηγούν με την καθορισμένη ταχύτητα και να παρέχουν πολύτιμη διασφάλιση για ασφαλή οδήγηση. Κατά συνέπεια, ο εντοπισμός και η αναγνώριση των πινακίδων είναι μια ουσιαστική και σημαντική διαδρομή μελέτης για τη διάσωση των τραυματισμών όλων και τη διαφύλαξη της ατομικής προστασίας των οδηγών αυτοκινήτων.

Τα αυτόνομα αυτοκίνητα είναι το μέλλον της αυτοκινητοβιομηχανίας [1]–[4] Τα αυτόνομα ευφυή οχήματα, τα τελευταία χρόνια, έχουν προσελκύσει σημαντικό ενδιαφέρον λόγω των υψηλών δυνατοτήτων τους για πρακτικές εφαρμογές. Για να λειτουργούν αποτελεσματικά σε πραγματικούς δρόμους, τα οχήματα χρειάζονται ένα γρήγορο και ακριβές σύστημα επεξεργασίας εικόνας για να αποκτήσουν αντίληψη του περιβάλλοντος. Τα τελευταία χρόνια έχει σημειωθεί τεράστια ανάπτυξη στην επεξεργασία εικόνας εκμεταλλευόμενοι τα βαθιά νευρωνικά δίκτυα και οι αλγόριθμοι μηχανικής μάθησης έχουν αποκτήσει ιδιαίτερη σημασία, ειδικά στις μέρες μας. Τα αυτόνομα οχήματα έχουν ένα ευρύ φάσμα πιθανών εφαρμογών, συμπεριλαμβανομένων των επικίνδυνων περιβαλλόντων και των έξυπνων συστημάτων αυτοκινητοδρόμων.

Σε ζώνες κυκλοφορίας, ο εντοπισμός και η αναγνώριση κυκλοφορίας μπορεί να χρησιμοποιηθεί για την αυτόματη αναγνώριση των πινακίδων κυκλοφορίας. Αυτό γίνεται αυτόματα από το σύστημα καθώς ανιχνεύεται η ΟΣ και εμφανίζεται το όνομα και η ιδιότητα της σήμανσης. Έτσι, ακόμη και αν κάποια σήμανση χαθεί από τη προσοχή του οδηγού ή έχει προκύψει κάποιο κενό στη συγκέντρωση του, το σύστημα θα φροντίσει αυτή να ανιχνευθεί. Αυτό βοηθά στην ανάλογη προειδοποίηση των οδηγών και την απαγόρευση ορισμένων ενεργειών όπως η υπερβολική ταχύτητα.

Η ταξινόμηση των σημάτων κυκλοφορίας είναι πολύ χρήσιμο εργαλείο στα αυτόματα συστήματα υποβοήθησης οδηγού (ADAS). Τα συστήματα ανίχνευσης ΟΣ μπορούν να χωριστούν σε δύο επιμέρους καθήκοντα, ο εντοπισμός και η αναγνώριση. Ο πρωταρχικός ρόλος είναι η ανακάλυψη των στόχων εντός της εικόνας. Ο δεύτερος, επιδιώκει την αναγνώριση των σημάνσεων που εντοπίζονται, σε υποκατηγορίες. Παρέχει επίσης τις απαραίτητες πληροφορίες

σχετικά με τους φωτεινούς σηματοδότες και τον εξοπλισμό ελέγχου της κυκλοφορίας σε κοντινή περίμετρο του οχήματος, για τη διασφάλιση της ομαλής οδήγησης.

Το θέμα αυτής της διπλωματικής εργασίας

Η συγκεκριμένη διπλωματική εργασία αναλύει τον τρόπο με τον οποίο ένα λογισμικό και ένα ενσωματωμένο σύστημα, είναι σε θέση να χρησιμοποιήσουν τα διδόμενα εργαλεία και μεθόδους, ώστε να προβούν σε αναγνώριση αντικειμένων, και πιο συγκεκριμένα, οδικών σημάνσεων. Ένα ή πολλά συνελικτικά νευρωνικά δίκτυα λειτουργούν ταυτόχρονα η μεμονωμένα. Μέσω παραλλάξεων της αρχιτεκτονικής, γίνονται ταχύτερα, «ελαφρύτερα» και αποδοτικότερα. Αυτό είναι που σε μεγάλο βαθμό βοηθά ένα μηχάνημα να επιτυγχάνει περισσότερες ενέργειες με λιγότερο κόστος (χρόνου, ενέργειας κ.α.). Μαζί με συνεργαζόμενα πλαίσια και εφαρμογές, η διαδικασία αυτή φτάνει στο στάδιο της οπτικοποίησης των αποτελεσμάτων. Στην περίπτωση μας χρησιμοποιούμε ένα Jetson Nano και τα εργαλεία του οικοσυστήματος του, ώστε να προβεί στην αναγνώριση οδικής σήμανσης.

Σκοπός και στόχοι

Σκοπός της εργασίας είναι η εμβάθυνση στην αναγνώριση οδικής σήμανσης. Μετα την κατανόηση των αρχιτεκτονικών , μεθόδων, τάσεων και βέλτιστων τεχνικών, σειρά έχει η πλήρης εκμετάλλευση του hardware που διαθέτουμε. Στόχος είναι να προβούμε σε αναγνώριση οδικής σήμανσης με το Jetson Nano.

Μεθοδολογία

Στη διάθεση μας έχουμε, ένα Jetson Nano 2GB Developer Kit της Nvidia. Συνδέεται και χρησιμοποιείται ως αυτόνομο desktop. Με διάθεση πληκτρολογίου, ποντικιού, αφιερωμένη οθόνη και σύνδεση δικτύου. Χρησιμοποιώντας τη κάμερα Raspberry Pi (V2) συλλέγουμε φωτογραφίες με καθορισμό πλαισίου και κατάλληλο σχολιασμό και ετικέτες. Ο σχολιασμός είναι σε μορφή XML και η υλοποίηση γίνεται με PASCAL VOC. Στη συνέχεια προβαίνουμε σε εκπαίδευση του μοντέλου με DNN. Συγκεκριμένα χρησιμοποιούμε cuDNN που μας προμήθευσε η Nvidia μέσω του SDK. Το εξαγόμενο προϊόν πρέπει να μετατραπεί σε ONNX και με τη σειρά του χρησιμοποίει το detectnet. Μετα από πολλά epochs, και πολλές αναπροσαρμογές (και διόρθωση σετ δεδομένων αλλά και ρυθμίσεων), διαπιστώνουμε πως δημιουργήσαμε ένα σύστημα ικανό να αναγνωρίσει την οδική σήμανση όπως ορίστηκε να πράξει. Ο TensorRT είναι ο μεταγλωττιστής που χρησιμοποιείται και όλο το εγχείρημα βασίζεται στο πλαίσιο PyTorch.

Καινοτομία

Το απαιτητικό σε αυτή την αποστολή είναι οι περιορισμένοι πόροι σε σχέση με ένα τόσο δύσκολο υπολογιστικό έργο. Πολλοί χρησιμοποιούν την εξολοκλήρου βοήθεια μέσω Dockers η μέσω Images. Υπάρχουν πολλοί χρήστες του συγκεκριμένου συστήματος, που είτε έτρεξαν έτοιμα σετ δεδομένων, είτε εκπαίδευσαν το μοντέλο τους αλλού και απλά μετέφεραν το .pth/.pt στο Jetson και έτρεξαν απευθείας έτοιμα προεκπαιδευμένα μοντέλα. Στη δική μας περίπτωση όλα γίνονται υπο τη σκέπη του ήδη καταχωρημένου λογισμικού, με δεδομένα και εκπαίδευση εξολοκλήρου στο Jetson Nano και σε αντικείμενο (Οδική Σήμανση) που δεν υπάρχει σε Clones η preprocessed σε τεράστιες προ-τρεγμένες αποθήκες αναγνώρισης αντικειμένων (πχ. εντοπισμός αυτοκίνητου, ανθρώπου, σκύλου, γάτας κοκ.)

Δομή

Στο πρώτο κεφάλαιο γίνεται μια προσπάθεια εξοικείωσης με τις έννοιες και τις βασικές διαδικασίες της αναγνώρισης αντικειμένων. Εξηγείται η σημαντικότητα της, και η εφαρμογή της, μέσω βαθιάς μάθησης, δίνεται σαν γενική δομή, καθώς και οι δημοφιλέστεροι αλγόριθμοι υλοποίησης της, περιγράφονται. Εντός του ιδίου κεφαλαίου γίνεται μια ανασκόπηση στη πρόοδο που έχει πραγματοποιηθεί σε αυτόν τον τομέα και σε σχετικές εργασίες που συνέβαλαν σε αυτή.

Η αναγνώριση οδικής σήμανσης, συγκεκριμένα, μπαίνει στο μικροσκόπιο εντός του δεύτερου κεφαλαίου. Τα καθήκοντα ενός τέτοιου συστήματος όπως και οι δυσκολίες του, αναφέρονται και σχολιάζονται. Οι μέθοδοι ανίχνευσης των οδικών σημάνσεων αναλύονται εκτενώς.

Ο ρόλος της βαθιάς μάθησης και των νευρωνικών δικτυών, αναφέρεται σχολαστικά και σε βάθος, στο τρίτο κεφάλαιο. Αφού αποσαφηνιστούν βασικές έννοιες και οροί, γίνεται διείσδυση στην αρχιτεκτονική των συνελικτικών νευρωνικών δικτυών και στον τρόπο εκπαίδευσης τους. Πριν το κλείσιμο της συγκεκριμένης ενότητας, δίδεται μια περιγραφή των προκλήσεων στην εκπαίδευση των βαθιών δικτύων, που περιλαμβάνει συνήθεις τακτικές για καλύτερα αποτελέσματα και αποφυγή λαθών.

Στο τελευταίο κεφάλαιο, περιλαμβάνεται η υλοποίηση της εργασίας καθώς και η πλήρης αναφορά και περιγραφή των συστημάτων (υλικών και μη). Αφού αναλυθεί κάθε έννοια που συμμετέχει στην διαδικασία παραγωγής δεδομένων, εκπαίδευσης και οπτικοποίησης, φτάνουμε στην εφαρμογή τους και τέλος την εμφάνιση των αποτελεσμάτων της υλοποίησης.

1 ΚΕΦΑΛΑΙΟ 1: Αναγνώριση Αντικειμένων

Η «αναγνώριση αντικειμένων» (Object recognition) είναι ένας γενικός όρος για να περιγράψει μια συλλογή σχετικών εργασιών υπολογιστικής όρασης που περιλαμβάνουν την αναγνώριση αντικειμένων με ψηφιακό αποτύπωμα. Είναι μια σημαντική εργασία υπολογιστικής όρασης (CV) που χρησιμοποιείται για την ανίχνευση περιπτώσεων οπτικών αντικειμένων ορισμένων κατηγοριών (για παράδειγμα, ανθρώπων, ζώων, αυτοκινήτων ή κτιρίων) σε ψηφιακές εικόνες όπως φωτογραφίες ή βίντεο. Ο στόχος της ανίχνευσης αντικειμένων είναι η ανάπτυξη υπολογιστικών μοντέλων που παρέχουν τις πιο θεμελιώδεις πληροφορίες που χρειάζονται οι εφαρμογές υπολογιστικής όρασης: "Ποιά αντικείμενα βρίσκονται και πού;".

1.1 Τι είναι η αναγνώριση αντικειμένων

Όταν ένας χρήστης ή επαγγελματίας αναφέρεται στην «αναγνώριση αντικειμένων» (Object recognition), συχνά εννοεί "ανίχνευση αντικειμένων" (Object detection). Η «ταξινόμηση εικόνας» (Image classification) περιλαμβάνει την πρόβλεψη της κλάσης ενός αντικειμένου σε μια εικόνα. Ο «εντοπισμός αντικειμένων» (Object localization) αναφέρεται στον προσδιορισμό της θέσης ενός ή περισσότερων αντικειμένων σε μια εικόνα και στη σχεδίαση ενός αφθονούντος πλαισίου γύρω από την έκτασή τους. Η «ανίχνευση αντικειμένων» (Object recognition/detection) συνδυάζει αυτές τις δύο εργασίες και μετατοπίζει και ταξινομεί ένα ή περισσότερα αντικείμενα σε μια εικόνα. Ως εκ τούτου, μεταξύ αυτών των τριών εργασιών υπολογιστικής όρασης, μπορούμε να διακρίνουμε:

1.1.1 Ταξινόμηση εικόνας

Οι αλγόριθμοι παράγουν μια λίστα κατηγοριών, των αντικειμένων που υπάρχουν στην εικόνα. Προβλέπει τον τύπο ή την κλάση ενός αντικειμένου σε μια εικόνα. Ως είσοδο δέχεται μια εικόνα με ένα μόνο αντικείμενο, όπως μια φωτογραφία και ως έξοδο μια ετικέτα κλάσης (π.χ. ένας ή περισσότεροι ακέραιοι που αντιστοιχίζονται σε ετικέτες κλάσης).

1.1.2 Εντοπισμός αντικειμένου

Οι αλγόριθμοι παράγουν μια λίστα κατηγοριών αντικειμένων που υπάρχουν στην εικόνα, μαζί με ένα πλαίσιο οριοθέτησης με στοίχιση άξονα που υποδεικνύει τη θέση και την κλίμακα μιας παρουσίας κάθε κατηγορίας αντικειμένων. Δηλαδή, εντοπίζει την παρουσία αντικειμένων σε μια εικόνα και υποδεικνύει τη θέση τους με ένα πλαίσιο οριοθέτησης. Ως είσοδο λαμβάνει μια εικόνα με ένα ή περισσότερα αντικείμενα, όπως μια φωτογραφία και ως έξοδο ένα ή περισσότερα πλαίσια οριοθέτησης (π.χ. που ορίζονται από ένα σημείο, πλάτος και ύψος).

1.1.3 Αναγνώριση αντικειμένων

Οι αλγόριθμοι παράγουν μια λίστα κατηγοριών αντικειμένων που υπάρχουν στην εικόνα μαζί με ένα πλαίσιο οριοθέτησης με στοίχιση άξονα που υποδεικνύει τη θέση και την κλίμακα κάθε παρουσίας και κάθε κατηγορίας αντικειμένων. Πιο απλά, εντοπίζει την παρουσία αντικειμένων με πλαίσιο οριοθέτησης και τύπους ή κλάσεις των αντικειμένων που βρίσκονται σε μια εικόνα. Ως είσοδος λαμβάνεται μια εικόνα με ένα ή περισσότερα αντικείμενα, όπως μια φωτογραφία, και ως έξοδος ένα ή περισσότερα πλαίσια οριοθέτησης (π.χ. που ορίζονται από ένα σημείο, πλάτος και ύψος) και μια ετικέτα κλάσης για κάθε πλαίσιο οριοθέτησης.

1.1.4 Τμηματοποίηση

Μια περαιτέρω επέκταση αυτής της ανάλυσης των εργασιών υπολογιστικής όρασης είναι η τμηματοποίηση αντικειμένων, που ονομάζεται «τμηματοποίηση παρουσίας αντικειμένου» (object instance segmentation) ή «σημασιολογική τμηματοποίηση» (semantic segmentation), όπου οι παρουσίες αναγνωρισμένων αντικειμένων υποδεικνύονται επισημαίνοντας τα συγκεκριμένα pixel του αντικειμένου αντί για ένα πλαίσιο χονδροειδούς οριοθέτησης. Από

αυτή την ανάλυση, μπορούμε να δούμε ότι η αναγνώριση αντικειμένων αναφέρεται σε μια σειρά απαιτητικών εργασιών υπολογιστικής όρασης.

1.1.5 Διαφορές ανίχνευσης αντικειμένων και ταξινόμησης εικόνας

Η ταξινόμηση εικόνας στέλνει μια ολόκληρη εικόνα μέσω ενός ταξινομητή (όπως ένα βαθύ νευρωνικό δίκτυο) για να εξάγει μια ετικέτα. Οι ταξινομητές λαμβάνουν υπόψη ολόκληρη την εικόνα, αλλά δεν αναφέρουν πού εμφανίζεται η ετικέτα στην εικόνα. Ο εντοπισμός αντικειμένων είναι ελαφρώς πιο προηγμένος, καθώς δημιουργεί ένα πλαίσιο οριοθέτησης γύρω από το ταξινομημένο αντικείμενο. Η ταξινόμηση έχει τα πλεονεκτήματά της, καθώς, είναι μια καλύτερη επιλογή για ετικέτες που δεν έχουν πραγματικά φυσικά όρια, όπως «θολωμένη» ή «φωτόλουστη». Ωστόσο, τα συστήματα ανίχνευσης αντικειμένων σχεδόν πάντα θα ξεπεράσουν τα δίκτυα ταξινόμησης στον εντοπισμό αντικειμένων που έχουν υλική παρουσία, όπως πχ. ένα αυτοκίνητο.

1.2 Σημαντικότητα της αναγνώρισης αντικειμένων

Η αναγνώριση αντικειμένων είναι ένα από τα θεμελιώδη προβλήματα της υπολογιστικης ορασης. Αποτελεί τη βάση πολλών άλλων εργασιών υπολογιστικής όρασης, για παράδειγμα, τμηματοποίηση στιγμιότυπων, λεζάντες εικόνων, παρακολούθηση αντικειμένων και πολλά άλλα. Συγκεκριμένες εφαρμογές ανίχνευσης αντικειμένων περιλαμβάνουν ανίχνευση πεζών, καταμέτρηση ατόμων, ανίχνευση προσώπου, ανίχνευση κειμένου, ανίχνευση πόζας ή αναγνώριση ΟΣ.

1.3 Αναγνώριση αντικειμένων και Βαθιά Μάθηση

Τα τελευταία χρόνια, οι ραγδαίες εξελίξεις των τεχνικών βαθιάς μάθησης έχουν επιταχύνει σημαντικά την ορμή της ανίχνευσης αντικειμένων. Με τα δίκτυα βαθιάς μάθησης και την υπολογιστική ισχύ των GPU, η απόδοση των ανιχνευτών αντικειμένων και των ιχνηλατών έχει βελτιωθεί σημαντικά, επιτυγχάνοντας σημαντικές ανακαλύψεις στην ανίχνευση αντικειμένων. Η μηχανική μάθηση (ML) είναι ένας κλάδος της τεχνητής νοημοσύνης (AI) και ουσιαστικά περιλαμβάνει μοτίβα εκμάθησης από παραδείγματα ή δείγματα δεδομένων καθώς η μηχανή έχει πρόσβαση στα δεδομένα και έχει την ικανότητα να μαθαίνει από αυτά (εποπτευόμενη μάθηση σε σχολιασμένες εικόνες). Η Βαθιά Μάθηση (DL) είναι μια εξειδικευμένη μορφή μηχανικής μάθησης που περιλαμβάνει μάθηση σε διαφορετικά στάδια.

1.4 Η λειτουργία της αναγνώρισης αντικειμένων

Η ανίχνευση αντικειμένων μπορεί να πραγματοποιηθεί χρησιμοποιώντας είτε παραδοσιακές τεχνικές επεξεργασίας εικόνας είτε σύγχρονα δίκτυα DL. Οι τεχνικές επεξεργασίας εικόνας γενικά δεν απαιτούν ιστορικά δεδομένα για την εκπαίδευση και πραγματοποιούνται χωρίς επίβλεψη. Το «OpenCV» είναι ένα δημοφιλές εργαλείο για εργασίες επεξεργασίας εικόνας. Ως εκ τούτου, αυτές οι εργασίες δεν απαιτούν σχολιασμένες εικόνες, όπου οι άνθρωποι επισήμαναν τα δεδομένα χειροκίνητα (πχ. για εποπτευόμενη εκπαίδευση). Όμως, αυτές οι τεχνικές περιορίζονται σε πολλούς παράγοντες, όπως πολύπλοκα σενάρια (χωρίς μονόχρωμο φόντο), απόφραξη (μερικώς κρυμμένα αντικείμενα), φωτισμό και σκιές και εφέ. Από την άλλη μεριά, οι μέθοδοι βαθιάς μάθησης εξαρτώνται γενικά από την εποπτευόμενη και μη εποπτευόμενη μάθηση, με τις εποπτευόμενες μεθόδους ομως να είναι το πρότυπο στις εργασίες υπολογιστικής όρασης. Η απόδοση περιορίζεται από την υπολογιστική ισχύ των GPU, η οποία αυξάνεται ραγδαία χρόνο με το χρόνο. Η ανίχνευση αντικειμένων DL είναι σημαντικά πιο ισχυρή σε

απόφραξη, πολύπλοκες σκηνές και προκλητικό φωτισμό. Στα μειονεκτήματα όμως εντάσσεται η απαίτηση τεράστιου όγκου δεδομένων εκπαίδευσης. η διαδικασία σχολιασμού εικόνας απαιτεί έντονη εργασία και καθίσταται δαπανηρή. Για παράδειγμα, η επισήμανση (labeling) 600.000 εικόνων για την εκπαίδευση ενός προσαρμοσμένου αλγορίθμου ανίχνευσης αντικειμένων DL, θεωρείται μικρό σύνολο δεδομένων. Ωστόσο, πολλά σύνολα δεδομένων αναφοράς (MS COCO, Caltech, KITTI, PASCAL VOC, V5) παρέχουν τη διαθεσιμότητα δεδομένων με ετικέτα.

1.5 Ανιχνευτές ενός σταδίου και δύο σταδίων

Η ανίχνευση αντικειμένων γενικά κατηγοριοποιείται σε ανιχνευτές αντικειμένων ενός σταδίου και ανιχνευτές αντικειμένων δύο σταδίων. Οι υπερσύγχρονες αρχιτεκτονικές ανίχνευσης, πολλές από τις οποίες έχουν εκπαιδευτεί εκ των προτέρων στο σύνολο δεδομένων αντικειμένων COCO, αποτελούνται από δύο αρχιτεκτονικές σταδίων. Το COCO είναι ένα σύνολο δεδομένων εικόνας που αποτελείται από 90 διαφορετικές κατηγορίες αντικειμένων (αμάξια, ανθρώπους, αθλητικές μπάλες, ποδήλατα, σκύλους, γάτες, άλογα κ.λπ.). Το σύνολο δεδομένων συγκεντρώθηκε για την επίλυση κοινών προβλημάτων ανίχνευσης αντικειμένων. Σήμερα γίνεται ξεπερασμένο καθώς οι εικόνες του τραβήχτηκαν κυρίως στις αρχές της δεκαετίας του ‘00 καθιστώντας τις πολύ μικρότερες, χαμηλότερης ανάλυσης και με διαφορετικές σιλουέτες αντικειμένων σε σχέση με τις σημερινές παραστάσεις/απαιτήσεις. Νεότερα σύνολα δεδομένων, όπως το OpenImages, παίρνουν τη θέση τους ως τα de-facto σύνολα δεδομένων πριν από κάθε εκπαίδευση.

1.5.1 Ανιχνευτές ενός σταδίου

Οι ανιχνευτές ενός σταδίου προβλέπουν πλαίσια οριοθέτησης πάνω από τις εικόνες χωρίς το βήμα της «πρότασης περιοχής». Αυτή η διαδικασία καταναλώνει λιγότερο χρόνο και επομένως μπορεί να χρησιμοποιηθεί σε εφαρμογές σε πραγματικό χρόνο. Αυτοί οι ανιχνευτές δίνουν προτεραιότητα στην ταχύτητα εξαγωγής συμπερασμάτων και είναι εξαιρετικά γρήγοροι αλλά όχι τόσο καλοί στην αναγνώριση αντικειμένων ακανόνιστου σχήματος ή μιας ομάδας μικρών αντικειμένων. Οι πιο δημοφιλείς ανιχνευτές αυτού του είδους, περιλαμβάνουν τα YOLO, SSD και RetinaNet. Οι πιο πολύ-τρεγμένοι και πρόσφατοι ανιχνευτές σε πραγματικό χρόνο είναι οι YOLOv4-Scaled (2020) και YOLOR (2021). Το κύριο πλεονέκτημα του ενός σταδίου είναι ότι αυτοί οι αλγόριθμοι είναι γενικά ταχύτεροι από τους ανιχνευτές πολλαπλών σταδίων και δομικά απλούστεροι.

1.5.2 Ανιχνευτές δύο σταδίων

Στους ανιχνευτές αντικειμένων δύο σταδίων, προτείνονται οι κατά προσέγγιση περιοχές αντικειμένων, χρησιμοποιώντας βαθιά χαρακτηριστικά πριν χρησιμοποιηθούν για την ταξινόμηση, καθώς και η παλινδρόμηση πλαισίου οριοθέτησης για το υποψήφιο αντικείμενο. Η αρχιτεκτονική δύο σταδίων περιλαμβάνει τη πρόταση περιοχής αντικειμένου με συμβατικές μεθόδους Υπολογιστικής Όρασης ή βαθιά δίκτυα, ακολουθούμενη από την ταξινόμηση αντικειμένων με βάση χαρακτηριστικά που εξάγονται από την προτεινόμενη περιοχή με παλινδρόμηση πλαισίου οριοθέτησης. Οι μέθοδοι που επιτυγχάνουν την υψηλότερη ακρίβεια ανίχνευσης, συνήθως είναι πιο αργές. Λόγω των πολλών βημάτων εξαγωγής συμπερασμάτων ανά εικόνα, η απόδοση (καρέ ανά δευτερόλεπτο) δεν είναι τόσο καλή όσο οι ανιχνευτές ενός σταδίου. Διάφοροι ανιχνευτές δύο σταδίων περιλαμβάνουν συνεκτικό νευρωνικό δίκτυο

περιοχής (RCNN), με εξελίξεις Faster R-CNN ή Mask R-CNN. Η τελευταία εξέλιξη είναι το κοκκοποιημένο RCNN (granulated R-CNN = G-RCNN). Οι ανιχνευτές αντικειμένων βρίσκουν πρώτα μια περιοχή ενδιαφέροντος και χρησιμοποιούν αυτήν την περικομμένη περιοχή για ταξινόμηση. Ωστόσο, τέτοιοι ανιχνευτές πολλαπλών σταδίων συνήθως δεν μπορούν να εκπαιδευτούν από άκρο σε άκρο, επειδή η περικοπή είναι μια μη διαφορίσιμη διαδικασία.

1.6 Οι πιο δημοφιλείς αλγόριθμοι

Οι δημοφιλείς αλγόριθμοι που χρησιμοποιούνται για την εκτέλεση ανίχνευσης αντικειμένων περιλαμβάνουν τα συνελικτικά νευρωνικά δίκτυα (CNN) R-CNN, Fast R-CNN και YOLO (You Only Look Once).

1.6.1 R-CNN

Τα συνελικτικά νευρωνικά δίκτυα που βασίζονται σε περιοχές (ή περιοχές με χαρακτηριστικά CNN) είναι πρωτοποριακές προσεγγίσεις που εφαρμόζουν βαθιά μοντέλα στην ανίχνευση αντικειμένων. Τα μοντέλα R-CNN διαλέγουν αρχικά ικανά προτεινόμενα χωρικά πεδία από μια παράσταση (πχ, τα πλαίσια αγκίστρωσης αποτελούν έναν τύπο συστήματος διαλογής), μετά, επισημαίνουν τα πλαίσια οριοθέτησης και τις κατηγορίες (π.χ. μετατοπίσεις). Αυτές οι ετικέτες δημιουργούνται με βάση προκαθορισμένες πληροφορίες και εντολές που δίνονται στο πρόγραμμα. Στη συνέχεια εκμεταλλεύονται ένα CNN για να εκτελέσουν υπολογισμό με φορά προς τα μπροστά με σκοπό να προβούν στην εξαγωγή χαρακτηριστικών από την εκάστοτε περιοχή που προτείνεται.

Στο R-CNN, η εισαγόμενη εικόνα χωρίζεται πρώτα σε σχεδόν δύο χιλιάδες τμήματα περιοχής και στη συνέχεια εφαρμόζεται ένα συνελικτικό νευρωνικό δίκτυο για κάθε περιοχή, αντίστοιχα. Υπολογίζεται το μέγεθος των περιοχών και η σωστή περιοχή εισάγεται στο νευρωνικό δίκτυο. Μπορούμε να συμπεράνουμε ότι μια λεπτομερής μέθοδος όπως αυτή μπορεί να προκαλέσει χρονικούς περιορισμούς. Ο χρόνος εκπαίδευσης είναι σημαντικά μεγαλύτερος σε σύγκριση με το YOLO επειδή ταξινομεί και δημιουργεί κουτιά οριοθέτησης μεμονωμένα και ένα νευρωνικό δίκτυο εφαρμόζεται σε μία περιοχή, κάθε φορά.

Το 2015, το Fast R-CNN αναπτύχθηκε με σκοπό να μειώσει σημαντικά τον χρόνο. Ενώ το αρχικό R-CNN υπολόγισε τα χαρακτηριστικά του νευρωνικού δικτύου ανεξάρτητα σε κάθε μία, ξεχωριστά, από τις δύο χιλιάδες περιοχές ενδιαφέροντος, το Fast R-CNN τρέχει το νευρωνικό δίκτυο μία φορά σε ολόκληρη την εικόνα. Αυτό είναι πολύ συγκρίσιμο με την αρχιτεκτονική του YOLO, αλλά το YOLO παραμένει μια ταχύτερη εναλλακτική λύση για το Fast R-CNN λόγω της απλότητας του κώδικα.

Στο τέλος του δικτύου υπάρχει μια νέα μέθοδος γνωστή ως Region of Interest (ROI) Pooling, η οποία αποκόπτει κάθε «περιοχή ενδιαφέροντος» από τον τανυστή εξόδου του δικτύου, την αναδιαμορφώνει και την ταξινομεί. Αυτό καθιστά το Fast R-CNN πιο ακριβές απ' το αρχικό R-CNN. Ωστόσο, λόγω αυτής της τεχνικής αναγνώρισης, απαιτούνται λιγότερες είσοδοι δεδομένων για την εκπαίδευση ανιχνευτών στο Fast R-CNN όπως και στο R-CNN.

1.6.2 Mask R-CNN

Mask R-CNN είναι μια πρόοδος του Fast R-CNN. Η διαφορά αυτών των δύο είναι πως η Mask R-CNN πρόσθεσε έναν κλάδο για την πρόβλεψη μιας μάσκας αντικειμένου παράλληλα με τον

υπάρχοντα κλάδο για την αναγνώριση πλαισίου οριοθέτησης. Η Mask R-CNN είναι απλή στην εκπαίδευση και προσθέτει μόνο ένα μικρό λιθαράκι στο Faster R-CNN.

1.6.3 YOLO – You Only Look Once

Όντας ένα κατασκευασμένο σύστημα για ανίχνευση αντικειμένων σε πραγματικό χρόνο, η ανίχνευση αντικειμένων YOLO χρησιμοποιεί ένα μόνο νευρωνικό δίκτυο. Η έκδοση του ImageAI v2.1.0 υποστηρίζει την εκπαίδευση ενός προσαρμοσμένου μοντέλου YOLO για την ανίχνευση οποιουδήποτε είδους και αριθμού αντικειμένων. Τα συνελικτικά νευρωνικά δίκτυα είναι περιπτώσεις συστημάτων που βασίζονται σε ταξινομητές όπου το σύστημα επανατοποθετεί ταξινομητές (classifiers) ή τοπικοποιητές (localizers) για να εκτελέσει ανίχνευση και εφαρμόζει το μοντέλο ανίχνευσης σε μια εικόνα σε πολλαπλές τοποθεσίες και κλίμακες. Χρησιμοποιώντας αυτή τη διαδικασία, οι περιοχές "υψηλής βαθμολογίας" της εικόνας θεωρούνται ανιχνεύσεις. Με απλά λόγια, οι περιοχές που μοιάζουν περισσότερο με τις εικόνες εκπαίδευσης που δίνονται, προσδιορίζονται θετικά. Ως ανιχνευτής ενός σταδίου (single-stage detector) το YOLO εκτελεί παλινδρόμηση πλαισίου ταξινόμησης και οριοθέτησης σε ένα βήμα, καθιστώντας το πολύ ταχύτερο από τα περισσότερα CNN. Για παράδειγμα, η ανίχνευση αντικειμένων YOLO είναι περισσότερο από 1000x ταχύτερη από το R-CNN και 100x ταχύτερη από το Fast R-CNN.

Το YOLOv3 επιτυγχάνει 57.9% mAP στο σύνολο δεδομένων MS COCO σε σύγκριση με το 53.3% του DSSD513 και το 61.1% του RetinaNet. Το YOLOv3 χρησιμοποιεί ταξινόμηση πολλαπλών ετικετών με επικαλυπτόμενα μοτίβα για εκπαίδευση. Επομένως, μπορεί να χρησιμοποιηθεί σε σύνθετα σενάρια για ανίχνευση αντικειμένων. Λόγω των δυνατοτήτων πρόβλεψης πολλαπλών κλάσεων, το YOLOv3 μπορεί να χρησιμοποιηθεί για ταξινόμηση μικρών αντικειμένων, ενώ παρουσιάζει χειρότερη απόδοση στην ανίχνευση αντικειμένων μεγάλου ή μεσαίου μεγέθους. Το YOLOv4 είναι μια βελτιωμένη έκδοση του YOLOv3. Οι κύριες καινοτομίες είναι η βελτίωση δεδομένων μωσαϊκού, η αυτο-αντιπαραθετική εκπαίδευση και η κανονικοποίηση μεταξύ μίνι δεσμίδων.

1.6.4 YOLOR

Το YOLOR είναι ένας νέος ανιχνευτής αντικειμένων που παρουσιάστηκε το 2021. Μέχρι τώρα, η YOLOR μπορεί να μάθει μια γενική αναπαράσταση και να ολοκληρώσει πολλαπλές εργασίες μέσω αυτής της γενικής αναπαράστασης. Η έμμεση γνώση ενσωματώνεται στη ρητή γνώση μέσω της ευθυγράμμισης του χώρου του πυρήνα, της βελτίωσης της πρόβλεψης και της μάθησης πολλαπλών εργασιών. Μέσω αυτής της μεθόδου, η YOLOR επιτυγχάνει σημαντικά βελτιωμένα αποτελέσματα ανίχνευσης αντικειμένων. Σε σύγκριση με άλλες μεθόδους ανίχνευσης αντικειμένων στο σημείο αναφοράς του συνόλου δεδομένων COCO, το MAP του YOLOR είναι 3,8% υψηλότερο από το PP-YOLOv2 με την ίδια ταχύτητα εξαγωγής συμπερασμάτων. Σε σύγκριση με το Scaled-YOLOv4, η ταχύτητα εξαγωγής συμπερασμάτων έχει αυξηθεί κατά 88%, κάτι που τον καθιστά τον ταχύτερο ανιχνευτή αντικειμένων σε πραγματικό χρόνο που διατίθεται σήμερα. Το YOLOR σημαίνει «Μαθαίνεις Μόνο Μία Αναπαράσταση».

1.6.5 SSD – Single shot detector

Το SSD είναι ένας δημοφιλής ανιχνευτής ενός σταδίου που μπορεί να προβλέψει πολλές κλάσεις. Η μέθοδος ανιχνεύει αντικείμενα σε εικόνες χρησιμοποιώντας ένα μόνο βαθύ

νευρωνικό δίκτυο, διαχωρίζοντας το χώρο εξόδου των πλαισίων οριοθέτησης σε ένα σύνολο προεπιλεγμένων πλαισίων σε διαφορετικές αναλογίες διαστάσεων και κλίμακες, ανά θέση χάρτη χαρακτηριστικών. Ο ανιχνευτής αντικειμένων δημιουργεί βαθμολογίες για την παρουσία κάθε κατηγορίας αντικειμένων σε κάθε προεπιλεγμένο πλαίσιο και προσαρμόζει το πλαίσιο ώστε να ταιριάζει καλύτερα στο σχήμα του αντικειμένου. Επίσης, το δίκτυο συνδυάζει προβλέψεις από πολλούς χάρτες χαρακτηριστικών με διαφορετικές αναλύσεις για το χειρισμό αντικειμένων διαφορετικών μεγεθών. Ο ανιχνευτής SSD είναι εύκολο να εκπαιδευτεί και να ενσωματωθεί σε συστήματα λογισμικού που απαιτούν ένα στοιχείο ανίχνευσης αντικειμένων. Σε σύγκριση με άλλες μεθόδους ενός σταδίου, ο SSD έχει πολύ καλύτερη ακρίβεια, ακόμη και με μικρότερα μεγέθη εικόνας εισόδου.

1.6.6 MobileNet

Το MobileNet είναι ένα δίκτυο ανίχνευσης πολλαπλών κουτιών μίας λήψης που χρησιμοποιείται για την εκτέλεση εργασιών εντοπισμού αντικειμένων. Αυτό το μοντέλο υλοποιείται χρησιμοποιώντας το πλαίσιο Caffe. Η έξοδος του μοντέλου είναι ένα τυπικό διάνυσμα που περιέχει τα δεδομένα αντικειμένων που παρακολουθούνται, όπως περιγράφηκε προηγουμένως.

1.6.7 SqueezeDet

Το SqueezeDet είναι το όνομα ενός βαθιού νευρωνικού δικτύου για υπολογιστική όραση που κυκλοφόρησε το 2016. Το SqueezeDet αναπτύχθηκε ειδικά για αυτόνομη οδήγηση, όπου εκτελεί ανίχνευση αντικειμένων χρησιμοποιώντας τεχνικές υπολογιστικής όρασης. Όπως και το YOLO, είναι ένας αλγόριθμος ανιχνευτή ενός σταδίου. Στο SqueezeDet, τα συνελικτικά επίπεδα χρησιμοποιούνται μόνο για την εξαγωγή χαρτών χαρακτηριστικών, αλλά και ως επίπεδο εξόδου για τον υπολογισμό πλαισίων οριοθέτησης και πιθανοτήτων κλάσης. Ο αγωγός ανίχνευσης των μοντέλων SqueezeDet περιέχει μόνο μεμονωμένα εμπρόσθια περάσματα νευρωνικών δικτύων, επιτρέποντάς τους να είναι εξαιρετικά γρήγορα.

1.7 Ανασκόπηση στη πρόοδο της ανίχνευσης αντικειμένων

Λόγω της αξιοσημείωτης απόδοσης των CNN στο πρόβλημα ταξινόμησης εικόνων, η εφαρμογή του προς την κατεύθυνση της ανίχνευσης αντικειμένων ήταν αναμενόμενη. Σε αντίθεση με το ζήτημα της ταξινόμηση εικόνας, το ζήτημα εντοπισμού αντικειμένων απαιτεί ανίχνευση και εντοπισμό πολλών συγκεκριμένων αντικείμενων από εικόνες. Η πλειονότητα των συμβατικών μεθόδων ανίχνευσης αντικειμένων βασίζονται στην αναγνώριση εικόνας (image recognition). Την τελευταία δεκαετία, το παραδοσιακό πεδίο μηχανικής όρασης συχνά χρησιμοποιεί τις περιγραφές των χαρακτηριστικών για το χειρισμό εργασιών αναγνώρισης αντικειμένων. Το πιο κοινό από αυτούς περιλαμβάνουν το μετασχηματισμό αναλλοίωτης δυνατότητας κλίμακας (SIFT) [5] και το ιστόγραμμα προσανατολισμένης κλίσης (HOG) [6]. Εν τω μεταξύ, η εμφάνιση του AlexNet έχει κάνει τα CNNs την κύρια μέθοδο ανίχνευσης αντικειμένων. Τώρα η βασική ιδέα της ανίχνευσης αντικειμένων που βασίζεται στο CNN είναι να προτείνει υποψήφιες περιοχές από πριν (δηλ. πριν ταξινομηθούν από το CNN). Η τρέχουσα ενότητα περιγράφει κυρίως το σχηματισμό και ανάπτυξη CNNs ανίχνευσης αντικειμένων.

1.7.1 CNN

Εμπνευσμένοι από το AlexNet, οι Girshick et al. [7]προσπάθησαν να γενικεύσουν την ικανότητα του AlexNet στην αναγνώριση αντικειμένων με το ImageNet, στην PASCAL (VOC)

Challenge το 2014, προτείνοντας επίσης ένα μοντέλο R-CNN [7]. Για να επιτευχθεί ο εντοπισμός αντικειμένων, η εικόνα εισόδου περνάει από τρεις ενότητες, αντίστοιχα, προτάσεις περιοχής, εξαγωγή χαρακτηριστικών και ταξινόμηση περιοχής. Παρά την υψηλή ακρίβεια στην ανίχνευση περιοχής αντικειμένων, το R-CNN εξακολουθεί να έχει προφανείς ελλείψεις. Τα περισσότερα CNN, συμπεριλαμβανομένου του R-CNN, περιορίζονται στην αποδοχή εισροών σταθερού μεγέθους. Στην πραγματικότητα, τα επίπεδα συνέλιξης των CNN μπορούν να δημιουργήσουν επιφάνειες χαρακτηριστικών αυθαίρετου μεγέθους. Ο περιορισμός αυτός υπόκειται στην απαίτηση για εισροές σταθερού μήκους από τα επίπεδα FC. Επιπλέον η εκπαίδευση του R-CNN είναι μια περίπλοκη και περιττή διαδικασία. Τόσο τα SVM όσο και η εκμάθηση παλινδρόμησης πλαισίου οριοθέτησης (bounding box regression) απαιτεί εξαγωγή δυνατοτήτων για κάθε πρόταση περιοχής και την αποθήκευσή τους στον σκληρό δίσκο, ο οποίος είναι υπολογιστικά τεράστιος και καταναλώνει το αποθηκευτικό χώρο.

1.7.2 SPP - SPPNet

Το 2015, οι He et al. [8] πρότειναν ένα μοντέλο SPPNet. Το μοντέλο προσθέτει ένα επίπεδο συγκέντρωσης χωρικής πυραμίδας (SPP) μεταξύ του τελευταίου στρώματος συνέλιξης και του πρώτου στρώματος FC ενός CNN. Το SPP αποτελείται από τρία επίπεδα συνολικά. Ανάλογα με την ανάλυση των χωρικών μπλοκ, διαμερισματοποιημένα από το χάρτη χαρακτηριστικών εισόδου, μπορεί να χωριστεί σε ένα επίπεδο χαμηλής, ένα μεσαίας και ένα υψηλής ανάλυσης. Τα τρία επίπεδα χωρίζουν την είσοδο χαρακτηριστικού χάρτη σε 1, σε 4 και σε 16 χωρικά μπλοκ, αντίστοιχα. Για χάρτες χαρακτηριστικών διαφορετικών μεγεθών, ο συνολικός αριθμός των χωρικών τμημάτων εξόδου παραμένει αμετάβλητος. Έτσι, μετά τη σύνδεση στο τελευταίο επίπεδο συνέλιξης, το επίπεδο SPP εξάγει τανυστές σταθερού μήκους και στέλνει στο στρώμα της FC για την ταξινόμηση κάθε πρότασης περιοχής. Σε σύγκριση με το R-CNN, ο χρόνος δοκιμής του SPPNet μπορεί να επιταχυνθεί κατά 10-100 φορές . Προφανώς, το υφιστάμενο στρώμα SPP επιτρέπει σε εισόδους CNNs διαφορετικών μεγεθών να παράγουν εξόδους του ίδιου μεγέθους, κάτι δηλαδή που σπάει τον περιορισμό ότι οι είσοδοι των μοντέλων CNN έχουν όλες σταθερά μεγέθη. Παρ 'όλα αυτά, το SPPNet εξακολουθεί να διαθέτει περίπλοκη εκπαίδευση και διαδικασίες ανάγνωσης και γραφής σκληρού δίσκου. Εκτός αυτού, το SPPNet δεν μπορεί να ενημερώσει τα επίπεδα συνέλιξης πριν από το επίπεδο SPP κατά τη διάρκεια της ανίχνευσης αντικειμένων, περιορίζοντας την ακρίβεια αναγνώρισής του.

1.7.3 Fast και Faster R-CNN

Ως απάντηση σε αυτά τα προβλήματα, οι Girshick et al. [7] ανέπτυξαν το Fast R-CNN το 2015, το οποίο εισάγει ένα ειδικό SPP ενός στρώματος που ονομάζεται «RoI pooling» χωρίς περιορισμό του μεγέθους εισόδου των χαρτών δυνατοτήτων, ενοποιώντας έτσι τα μεγέθη εισόδου στο επίπεδο FC [7]. Στην αρχιτεκτονική του Fast R-CNN, οι εικόνες και οι προτάσεις της περιοχής τους υπόκεινται σε συνέλιξη και μέγιστες λειτουργίες συγκέντρωσης μαζί ως η είσοδος, αποκτώντας έτσι έναν κοινό χάρτη χαρακτηριστικών που περιέχει προτάσεις περιοχής. Το RoI pooling εξάγει τις προτάσεις της περιοχής σε τέτοιο χάρτη χαρακτηριστικών όπως ιδιοδιανύσματα σταθερού μήκους. Μετά την επεξεργασία από το στρώμα FC, αυτά τα ιδιοδιανύσματα αποστέλλονται σε δύο στρώματα εξόδου, τα οποία είναι αντίστοιχα υπεύθυνα για την έξοδο των εκτιμήσεων πιθανότητας softmax (κατηγορίες αντικειμένων K και 1 κατηγορία υποβάθρου) και των θέσεων πλαισίων οριοθέτησης (κατηγορίες αντικειμένων K).

Σε αντίθεση με το R-CNN και το SPPNet, το οποίο απασχολεί μοντέλα παλινδρόμησης όπως το γραμμικό SVM για την οριακή παλινδρόμηση πλαισίου, το Fast R-CNN υιοθετεί softmax

και παλινδρόμηση κουτιού οριοθέτησης. Εκτός από την αντιμετώπιση της αδυναμίας της συνέλιξης, επίπεδα πριν από το στρώμα SPP στο SPPNet για την ενημέρωση των παραμέτρων δικτύου, το Fast R-CNN επίσης παρουσιάζει προφανώς ταχύτερη εκπαίδευση και δοκιμές από το SPPNet. Δεδομένης της κατανάλωσης σημαντικών υπολογιστικών πόρων που προκαλείται από το Fast R-CNN λόγω της εξάρτησης σχετικά με την επιλογή υποψήφιων περιφερειών, οι Ren et al. [9] ανέπτυξαν ένα δίκτυο προτάσεων περιφέρειας (RPN) για ανίχνευση αντικειμένων σε πραγματικό χρόνο, το 2017, για την αντικατάσταση της μεθόδου επιλεκτικής αναζήτησης. Ως ένα πλήρως συνελικτικό δίκτυο (FCN), το RPN μπορεί να λάβει μια σειρά βαθμολογιών αντικειμενικότητας από εικόνες αυθαίρετου μεγέθους. Οι ίδιοι [9]συνδύασαν το Fast R-CNN μαζί με το RPN για να θεμελιώσουν το Faster R-CNN, το οποίο είναι ένα δίκτυο που μοιράζεται τα χαρακτηριστικά των επιπέδων συνέλιξης. Το δίκτυό του αποτελείται από δύο μέρη, ένα RPN που δημιουργεί προτάσεις περιοχής και έναν ανιχνευτή, που υιοθετεί και χρησιμοποιεί τις προτεινόμενες περιοχές. Το Faster R-CNN πέτυχε τα καλύτερα αποτελέσματα ανίχνευσης στα PASCAL VOC2007, 2012 και σύνολα δεδομένων MS COCO, με την υπολογιστική διαδικασία να είναι σχεδόν σε πραγματικό χρόνο.

1.7.4 YOLO

Το 2016, οι Redmon et al. [10] πρότειναν μια νέα στρατηγική ανίχνευσης για το πρόβλημα ανίχνευσης αντικειμένων. Σε αντίθεση με προηγούμενες πρακτικές, τα περισσότερα δίκτυα ανίχνευσης εντοπίζουν αντικείμενα με βάση την ταξινόμηση, και το You Only Look Once (YOLO) θεωρείται ως μια νέα προσέγγιση ανίχνευσης [10]. Μετατρέπει το πρόβλημα ανίχνευσης αντικειμένων σε πρόβλημα πιθανοτικής παλινδρόμησης πολλαπλών πλαισίων οριοθέτησης και σχετικών κατηγοριών. Βασισμένο στο GoogLeNet,, το YOLO χρησιμοποιεί ένα μόνο νευρωνικό δίκτυο και μία μόνο αξιολόγηση για την άμεση πρόβλεψη των πλαισίων οριοθέτησης και κατηγορίες από ολόκληρη την εικόνα εισόδου. Η Εικόνα 18 απεικονίζει την ιδέα ανίχνευσης αντικειμένων του YOLO. Αρχικά, το δίκτυο τμηματοποιεί την εικόνα εισόδου σε πλέγματα SS, με κάθε πλέγμα να είναι υπεύθυνο για την ανίχνευση αντικειμένων-στόχων των οποίων το κεντρικό σημείο εμπίπτει στο πλέγμα, και για την πρόβλεψη των πλαισίων οριοθέτησης «B» και για τη σήμανση των αντίστοιχων βαθμολογιών. Η βαθμολογία υψηλής εμπιστοσύνης υποδεικνύει ότι η πιθανότητα να περιέχεται το αντικείμενο στο πλαίσιο οριοθέτησης είναι ψηλός. Οι τιμές IoU των προβλεπόμενων και πραγματικών πλαισίων οριοθέτησης θεωρούνται ως οι βαθμολογίες εμπιστοσύνης, οι οποίες εξαρτώνται από κοινού από την πιθανότητα P_r (object) του πλαισίου οριοθέτησης που περιέχει αντικείμενο και τη γεωμετρική ακρίβεια IoU (αληθής πρόβλεψη) του κουτιού οριοθέτησης. Ο σχετικός υπολογιστικός τύπος είναι:

$$P_r(Κλάση_i|Αντικείμενο) * P_r(Αντικείμενο) * IoU^{αληθ.}_{προβλ.} = P_r(Κλάση_i) * IoU^{αληθ.}_{προβλ.}$$

όπου P_r (αντικείμενο) υποδηλώνει την πιθανότητα αντικειμένου, $IoU^{αληθ.}_{προβλ.}$είναι η γεωμετρική ακρίβεια και η πιθανότητα της, υπό όρους, κατηγορίας. Κάθε πλαίσιο οριοθέτησης του YOLO περιέχει πέντε προβλεπόμενες τιμές: x, y, w, h και εμπιστοσύνη, εκ των οποίων (x, y) υποδηλώνει τη θέση συντεταγμένων του κέντρου του πλαισίου οριοθέτησης σε σχέση με το

όριο του πλέγματος, και με w και h υποδηλώνει το προβλεπόμενο πλάτος και ύψος σε σχέση με ολόκληρη την εικόνα. Και οι τέσσερις τιμές είναι συντελεστές αναλογικότητας μεταξύ 0 και 1.

1.7.4.1 YOLOv1

Στο σύνολο δοκιμών VOC 2007, η ταχύτητα ανίχνευσης του SSD είναι 59 fps, ενώ αυτή του ταχύτερου R-CNN είναι μόνο 7 fps και του YOLOv1 είναι 45 fps. Παρά την ταχύτερη αναγνώριση από το Faster R-CNN, το YOLOv1 έχει ασθενέστερη ικανότητα εντοπισμού αντικειμένων, λόγω της επιλογής μόνο λίγων πλαισίων οριοθέτησης για πρόβλεψη, καθώς και χαμηλό ποσοστό γενίκευσης σχετικά με την αναλογία διαστάσεων των αντικειμένων.

1.7.4.2 YOLOv2

Το YOLOv2, ως βελτιωμένη μέθοδος, χρησιμοποιεί μια ποικιλία στρατηγικών για την ενίσχυση της ακρίβειας εντοπισμού [11], όπως για παράδειγμα, το YOLOv2 προσθέτει στρώμα BN μετά από κάθε συνέλιξη και διαγράφει τα dropout, ενισχύοντας έτσι το mAP του κατά 2,4%. Επίσης, το YOLOv2 βασίζεται στη στρατηγική χωρίς άγκυρα (anchor) για να διευκολύνει τη σύγκλιση του δικτύου. Τελος, το YOLOv2 χρησιμοποιεί ένα δίκτυο χαρακτηριστικών που ονομάζεται Darknet-19, το οποίο συμβάλλει μειώνοντας την υπολογιστική επιβάρυνση κατά περίπου 33%.

1.7.4.3 YOLOv3

Το YOLOv3, εισάγοντας σύντηξη πολλαπλών κλιμάκων residual networks (υπολειμματικών δικτύων) και χαρακτηριστικών (features), βελτιώνει περαιτέρω το ποσοστό αναγνώρισης δικτύου [11]. Εμβαθύνει το δίκτυο μέσω της εισαγωγής residual κατασκευής δικτύου στο Darknet. Δεδομένου ότι το βελτιωμένο δίκτυο έχει 53 στρωματα συνέλιξης, ονομάζεται προς τούτοις Darknet-53. Στο ImageNet, το Darknet-53 έχει παρόμοιο ποσοστό αναγνώρισης με αυτό των ResNet-101 και ResNet-152, παρόλο που η ταχύτητα αναγνώρισής του είναι πολύ μεγαλύτερη.

1.7.4.4 YOLOv4

Το YOLOv4 εφαρμόζει το πιο ικανό CSPNet ως ραχοκοκαλιά. Μέσω του data augmentation της Mosaic, το YOLOv4 συνδυάζει τέσσερις εικόνες εκπαίδευσης σε μία, για εκπαίδευση, αυξάνοντας έτσι την ευρωστία του μοντέλου. Εκτός αυτού, το YOLOv4 εισάγει αυτο-ανταγωνιστική εκπαίδευση για την ενίσχυση της αντίστασης του δικτύου σε ανταγωνιστικές επιθέσεις. Στο, MS COCO dataset, το YOLOv4 χρησιμοποιεί το Tesla V100 GPU για την επίτευξη ακρίβειας δικτύου AP 43,5% με ταχύτητα εξαγωγής συμπερασμάτων σχεδόν 65 fps.

1.7.4.5 YOLOv5

Μέσα στην ίδια χρονιά, ανέβηκε το YOLOv5 στο Github. Σε σύγκριση με τη προηγούμενη σειρά, το YOLOv5 προσθέτει τη λειτουργία του προσαρμοστικού υπολογισμού αγκύρωσης και η αναλογία διαστάσεων της αγκίστρωσης χρησιμοποιείται επίσης ως παράμετρος για το backpropagation και την επανάληψη ώστε να προσαρμοστεί στη πραγματικότητα. Εκτός από αυτό, το YOLOv5 χρησιμοποιείται για την τροποποίηση της δομής CSPnet και στη συνέχεια εφαρμόζεται στο λαιμό του ανιχνευτή σε σύγκριση με το YOLOv4. Σε συνδυασμό με αυτές τις στρατηγικές βελτιστοποίησης, το YOLOv5 επιτυγχάνει πιο αποτελεσματική εκπαίδευση.

1.7.4.6 YOLOX

Το 2021, μια αρχιτεκτονική δικτύου για το YOLOX προτάθηκε από τους Ge et al., οι οποίοι θεώρησαν οτι οι συζευγμένες κεφαλές ανιχνευτή ενδέχεται να επηρεάσουν τις επιδόσεις του δικτύου και έλαβαν την προσέγγιση της αποσύνδεσης «κεφαλής» για τη βελτίωση του AP κατά

4,2 % σε σύγκριση με την προσέγγιση από άκρο σε άκρο [12]. Επιπλέον, οι Ge et al [12] παρουσίασαν ένα σχήμα αντιστοίχισης δειγμάτων του SimOTA (Simplified Optimal Transport Assignment - Απλοποιημένη βέλτιστη εκχώρηση μεταφοράς) για τον υπολογισμό της σχέσης αντιστοίχισης μεταξύ του πλαίσιου αγκίστρωσης και του groundtruth αποτελέσματος, βασιζόμενοι στην πρόβλεψη του ίδιου του δικτύου, αντί να χρησιμοποιούν την αγκύρωση ως το προηγούμενο πλαίσιο. Το YOLOX επιτυγχάνει ταχύτερη και υψηλότερη ακρίβεια στο σύνολο δεδομένων COCO σε σύγκριση με τα υπάρχοντα μοντέλα.

2 ΚΕΦΑΛΑΙΟ 2: Αναγνώριση Οδικής Σήμανσης

Τα συστήματα αναγνώρισης ΟΣ (TSR) διασφαλίζουν ότι το ισχύον όριο ταχύτητας και άλλες μορφές ΟΣ αναρτώνται στον οδηγό σε συνεχή βάση. Η αυτόματη αναγνώριση λειτουργεί μέσω σύνδεσης, μεταξύ των εικόνων (που τραβήχτηκαν από κάμερα τοποθετημένη σε όχημα) και των ΟΣ που είναι αποθηκευμένες στο σύστημα πλοήγησης. Με αυτόν τον τρόπο, ακόμη και σημάνσεις που δεν είναι ρητά ορατές, θα εμφανίζονται στον οδηγό. Γενικά, αυτά τα συστήματα αναγνωρίζουν την ΟΣ για να ενημερώσουν τον οδηγό εμφανίζοντας ένα σύμβολο που αντιπροσωπεύει το αναγνωρισμένο σήμα. Μια κάμερα εγκατεστημένη στο αυτοκίνητο σαρώνει την πλευρά του δρόμου για αυτές τις σημάνσεις. Οι λειτουργίες υποβοήθησης οδηγού δεν πρέπει να υποκαθιστούν την κρίση του οδηγού. Αρκετοί ερευνητές [13]–[16] χρησιμοποίησαν CNN για να λύσουν το πρόβλημα της αναγνώρισης οδικών σημάτων. Τα CNN είναι το εργαλείο επιλογής για αυτού του είδους τα προβλήματα. Αυτοί είναι οι κορυφαίοι αλγόριθμοι της DL, αντικείμενα έντονης έρευνας, της οποίας ο πλούτος μπορεί να εντυπωσιάσει. Πρόσφατες μελέτες στη DL έδειξαν πόσο καλό είναι ένα CNN για την ταξινόμηση εικόνων και ανέπτυξαν αρκετά προηγμένα μοντέλα με ακρίβεια ταξινόμησης μεγαλύτερη από 99%. Τα RNN είναι περισσότερο γνωστά για την επίλυση προβλέψεων χρονοσειρών, ενώ τα CNN είναι γνωστά για την επεξεργασία εικόνας και είναι ισχυρός υποψήφιος για τη μοντελοποίηση ακολουθιών καθώς και για πρόβλεψη χρονοσειρών [17].

2.1 Η οδική σήμανση

Τα κοινά σήματα κυκλοφορίας αποτελούνται γενικά από σύμβολα, γράμματα, αριθμούς και άλλα στοιχεία, τα οποία παρέχουν σαφώς και ευδιάκριτα οδικές πληροφορίες, όπως απαγόρευση, προειδοποίηση ή καθοδήγηση σε συνδυασμένη μορφή. Με αυτό το είδος σημαντικών πληροφοριών, τα ευφυή οχήματα είναι ικανά να αντιλαμβάνονται αποτελεσματικά τις περιβαλλοντικές συνθήκες των οδικών σκηνών, αποφεύγοντας τα εμπόδια και λαμβάνοντας αποφάσεις έγκαιρα, γεγονός που μπορεί να μειώσει σημαντικά την εμφάνιση ατυχημάτων. Βοηθώντας παράλληλα να βελτιωθεί η ασφάλεια των οχημάτων, διευκολύνει επίσης την ομαλή ροή των αστικών οδικών δικτύων, συμβάλλοντας σημαντικά στη συντονισμένη ανάπτυξη έξυπνων πόλεων και έξυπνων δρόμων. Ο εντοπισμός σημάτων κυκλοφορίας αναγνωρίζεται ως μια εφικτή λύση που μπορεί να επιφέρει βελτίωση της απόδοσης της οδηγικής ασφάλειας. Αξιοποιεί τις τεχνολογίες υπολογιστικής όρασης για να εντοπιστούν αυτόματα αντικείμενα ενδιαφέροντος στα συλλεγόμενα δεδομένα. Ως μία από τις σημαντικές λειτουργίες ενός αυτοματοποιημένου συστήματος οδήγησης (ADS) [18] και ενός προηγμένου συστήματος υποβοήθησης οδηγού (ADAS) [18], [19], η ανίχνευση σημάτων κυκλοφορίας έχει χρησιμοποιηθεί ευρέως στους τομείς της αυτόνομης οδήγησης, την υποβοήθηση οδηγού μέχρι και την παρακολούθηση και συντήρηση οδικού δικτύου. Επομένως, θεωρείται μία από τις αξιοσημείωτες εκδηλώσεις της ανάπτυξης της αυτοκινητιστικής νοημοσύνης.

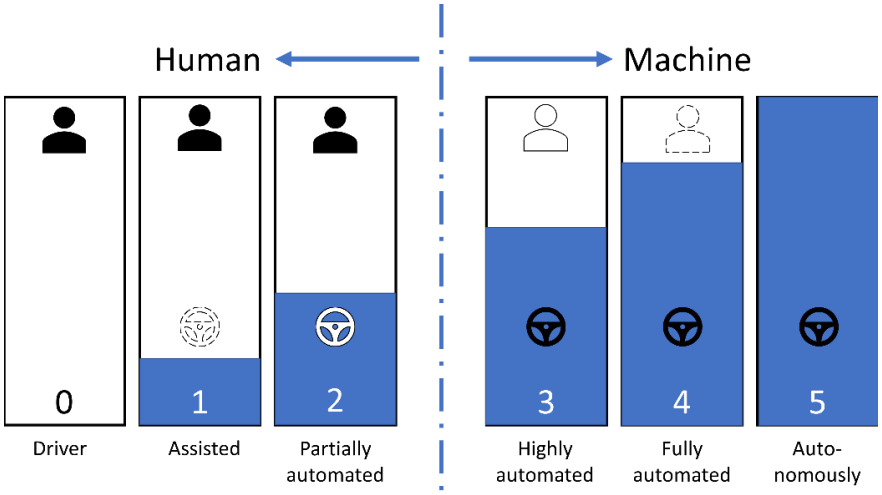
Η ανίχνευση ΟΣ σε πραγματικές σκηνές του δρόμου είναι ευαίσθητη σε αντικειμενικούς παράγοντες, όπως απόφραξη, ρύπανση, αλλαγές φωτισμού, θόλωση κίνησης και καθυστέρηση. Οι συμβατικές μέθοδοι ανίχνευσης βασίζονται συνήθως στο χρώμα ή το συγκεκριμένο σχήμα, η επίδραση της εξαγωγής χαρακτηριστικών δεν είναι αρκετά σταθερή, και το μοντέλο ανίχνευσης έχει κακή προσαρμοστικότητα στο περιβάλλον και ανεπαρκή ευρωστία. Οι μέθοδοι ανίχνευσης ΟΣ βασισμένες στη DL προέρχονται γενικά από τα καθολικά πλαίσια ανίχνευσης αντικειμένων. Αν και έχουν επιτύχει αξιοσημείωτα αποτελέσματα στην ακρίβεια, η έλλειψη απόδοσης σε πραγματικό χρόνο καθιστά τις περισσότερες από τις μεθόδους δύσκολες στην εφαρμογή τους σε πρακτικές εργασίες ανίχνευσης. Επιπλέον, αν και το ελαφρύ πλαίσιο εντοπισμού που αναπτύχθηκε για κινητές συσκευές και ενσωματωμένες συσκευές βελτιώνει σημαντικά την ταχύτητα ανίχνευσης και μειώνει τη ζήτηση για συνθήκες υλικού, η απόδοση όσον αφορά την ακρίβεια ήταν πάντα κατώτερη. Ως εκ τούτου, λόγω του συνθέτου περιβάλλοντος στο οποίο βρίσκονται οι ΟΣ, είναι ζωτικής σημασίας να σχεδιαστεί μια μέθοδος εντοπισμού τους με υψηλή ακρίβεια, εξαιρετική απόδοση σε πραγματικό χρόνο και ισχυρή ευρωστία.

Αν και έχει σημειωθεί εντυπωσιακή πρόοδος στην έρευνα στον εντοπισμό και την αναγνώριση ΟΣ, το πρόβλημα που προκλήθηκε από την ποικιλομορφία των ρεαλιστικών συνθηκών που αναφέρονται παραπάνω, παραμένει. Επιπλέον, η ποιότητα των αποτελεσμάτων κάθε μελέτης σε αυτό το πεδίο διαφέρει από τη μία ερευνητική ομάδα στην άλλη. Είναι εξαιρετικά δύσκολο να αποφασισθεί το ποιά μέθοδος παράγει ανώτερα συνολικά αποτελέσματα, κυρίως λόγω της έλλειψης προτύπου σύνολο δεδομένων εικόνων ΟΣ. Για παράδειγμα, είναι αρκετά αδύνατο να γνωρίζουμε πόσο αποτελεσματικά προσαρμόζονται τα συστήματα στις αλλαγές του φωτισμού της εικόνας, καθώς δεν είναι πάντα σαφές εάν εικόνες με χαμηλό φωτισμό χρησιμοποιήθηκαν στις μελέτες και τα πειράματα. Ένα άλλο μειονέκτημα της έλλειψης τυποποιημένου συνόλου δεδομένων είναι ότι ορισμένα έργα βασίζονται σε μικρά σύνολα εικόνων καθώς η σύνταξη ενός μεγάλου συνόλου εικόνων ΟΣ είναι μια χρονοβόρα διαδικασία. Το μειονέκτημα της χρήσης μικρών συνόλων δεδομένων είναι ότι είναι δύσκολο να αξιολογηθεί η αξιοπιστία των αποτελεσμάτων.

2.2 Κλίμακα αυτονομίας αυτοκίνητων

Σύμφωνα με το SAE [20], υπάρχουν έξι διαφορετικά επίπεδα ενός αυτόνομου συστήματος οχημάτων, όπως φαίνεται στο Σχήμα 1. Σήμερα, η πλειοψηφία των οχημάτων που χρησιμοποιούνται μεταξύ της κοινότητας βρίσκονται στο επίπεδο 0, το οποίο απαιτεί τον πλήρη έλεγχο των ανθρώπων. Στο πρώτο επίπεδο, διάφορα ειδικά συστήματα, για παράδειγμα, το cruise control ή το αυτόματο φρενάρισμα, θα μπορούσαν να ελέγχονται ξεχωριστά από το σύστημα υποβοήθησης οδήγησης. Τα αυτόνομα οχήματα στο δεύτερο επίπεδο προσφέρουν τουλάχιστον δύο ταυτόχρονες αυτόματες λειτουργίες, όπως το τιμόνι και η επιτάχυνση/επιβράδυνση, με την προσδοκία ότι οι οδηγοί θα κάνουν όλες τις υπόλοιπες εργασίες οδήγησης. Σε αυτά τα χαμηλότερα επίπεδα αυτόνομης οδήγησης, οι οδηγοί πρέπει να παρακολουθούν συνεχώς τις οδηγικές καταστάσεις για να εκτελούν παρεμβάσεις. Αντίθετα, από το τρίτο επίπεδο και μετά, τον έλεγχο του οχήματος, υπο ορισμένες συνθήκες, αναλαμβάνει το αυτόνομο σύστημα οδήγησης, αλλά οι οδηγοί πρέπει να παρεμβαίνουν στο σύστημα ελέγχου εάν αυτό δεν είναι σε θέση να χειριστεί πολύπλοκες καταστάσεις κυκλοφορίας. Στο τέταρτο επίπεδο, ένα πλήρως αυτοματοποιημένο σύστημα παρέχει όλες τις απαιτούμενες λειτουργίες των οδηγικών συμπεριφορών για τα οχήματα. Ενώ δεν απαιτούνται ανθρώπινες παρεμβάσεις,

ο οδηγός βρίσκεται σε ετοιμότητα ανάληψης του ελέγχου και να οδηγήσει το όχημα χειροκίνητα. Στο υψηλότερο επίπεδο, τα οχήματα είναι πλήρως ικανά για αυτονομία σε όλες τις περιπτώσεις. Τα οχήματα δεν χρειάζονται πλέον καμία ανθρώπινη αλληλεπίδραση και θα μπορούσαν να λειτουργήσουν ανεξάρτητα.



Σχήμα 1. Επίπεδα συστήματος αυτόνομης οδήγησης [21].

2.3 Βασικά καθήκοντα συστήματος

Δύο βασικά καθήκοντα ενός συστήματος επεξεργασίας εικόνας σε ένα αυτόνομο «νοήμον» όχημα, είναι η ανίχνευση αντικειμένων κυκλοφορίας και η ανίχνευση ΟΣ. Έχουν εκπονηθεί αρκετές μελέτες για την ανίχνευση κυκλοφοριακών αντικειμένων [22]–[24]. Επικεντρώνονται επίσης οι έρευνες και στον εντοπισμό πινακίδων κυκλοφορίας [25]. Κοινά χαρακτηριστικά αυτών των έργων είναι ο συνδυασμός παραδοσιακών τεχνικών υπολογιστικής όρασης με προηγμένα DNN για την επεξεργασία εικόνας.

2.3.1 Βελτιστοποιήσεις

Πολλές βελτιστοποιήσεις έχουν εισαχθεί προκειμένου να βελτιωθεί η ακρίβεια, η υπολογιστική απόδοση καθώς και η ευρωστία σε συνεχώς μεταβαλλόμενα περιβάλλοντα. Οι βελτιστοποιήσεις μπορεί να είναι:

1. Η χρήση της σύντηξης πολλαπλών αισθητήρων για τη λήψη πολυφασματικών εικόνων, η οποία επιτρέπει την ισχυρή ανίχνευση διαφόρων τύπων αντικειμένων όπως αυτοκίνητα, ανθρώπους και ποδήλατα σε διάφορες συνθήκες όπως κατά τη διάρκεια της ημέρας αλλά και της νύχτας [22].
2. Η αξιοποίηση ελαφρών NN μεγάλου βάθους για την επίτευξη υπολογιστικής αποδοτικότητας για εφαρμογές σε πραγματικό χρόνο [23].
3. Ο επανασχεδιασμός της αρχιτεκτονικής δικτύου για την ακριβή ανίχνευση αντικειμένων υπο πολύπλοκα σενάρια, συμπεριλαμβανομένης της διαφοροποιημένης εμφάνισης αντικειμένων και φόντου, της θόλωσης κίνησης, των δυσμενών καιρικών συνθηκών και των πολύπλοκων αλληλεπιδράσεων μεταξύ αντικειμένων [24].

2.3.2 Επισημάνσεις

Είναι ορατό ότι τρία σημαντικά χαρακτηριστικά της ανίχνευσης αντικειμένου κυκλοφορίας / πινακίδας είναι η υψηλή ακρίβεια, η ανθεκτικότητα και ο γρήγορος χρόνος απόκρισης (υπό την προϋπόθεση του περιορισμού υλικών πόρων). Αυτά τα χαρακτηριστικά είναι πραγματικά δύσκολο να αποκτηθούν. Έτσι, είναι ουσιαστικά απαραίτητη η ανάπτυξη μοντέλων ανίχνευσης αντικειμένων κυκλοφορίας/σημάνσεων υπό τέτοιου είδους συνθήκες.

2.4 Μέθοδοι ανίχνευσης ΟΣ

Σκοπός του εντοπισμού ΟΣ είναι κυρίως: να διαχωρίσει τα σήματα κυκλοφορίας από άλλα αντικείμενα της εικόνας και περιοχές φόντου και για να βρεί περιοχές στην εικόνα που μπορεί να είναι ΟΣ. Η ανίχνευση ΟΣ είναι ένα φλέγον θέμα, και υπάρχουν πολλά αποτελεσματικά μοντέλα ανίχνευσης. Με βάση ορισμένα χαρακτηριστικά, οι υπάρχουσες μέθοδοι ανίχνευσης μπορούν να χωριστούν χονδρικά σε μεθόδους με βάση το χρώμα, με βάση το σχήμα και μεθόδους βασισμένους στην μηχανική μάθηση.

2.4.1 Με βάση το χρώμα

Στα πρώτα στάδια ανίχνευσης ΟΣ, η μέθοδος ανίχνευσης με βάση το χρώμα, λαμβάνει το χρώμα ως βασικό παράγοντα, μειώνει το εύρος αναζήτησης σύμφωνα με το χρώμα της εικόνας προορισμού και χρησιμοποιεί τμηματοποίηση χρώματος για τον προσδιορισμό της κατά προσέγγισης θέσης της ΟΣ στην εικόνα. Αυτή η μέθοδος εκτελεί τμηματοποίηση χρωμάτων και επεξεργασία κατωφλίου σε διαφορετικά χρώματα για την εύρεση ΟΣ από το φόντο. Η πιο άμεση μέθοδος είναι η εκτέλεση χρωματικής τμηματοποίησης σε συγκεκριμένα χρώματα ΟΣ στον χρωματικό χώρο κόκκινο-πράσινο-μπλε (RGB) [26]. Το χρώμα των ΟΣ επηρεάζεται εύκολα από τις συνθήκες φωτισμού και σε ορισμένα σύνθετα σενάρια κυκλοφορίας, πολλά αντικείμενα φόντου ενδέχεται επίσης να είναι παρόμοια με το χρώμα των ΟΣ. Επομένως, ορισμένοι μελετητές αρχίζουν να χρησιμοποιούν πληροφορία χρώματος για να φιλτράρουν αρχικά την περιοχή που μπορεί να υπάρχει ΟΣ, στη συνέχεια, χρησιμοποιούν άλλες μεθόδους για να εντοπίσουν με ακρίβεια τις ΟΣ [27].

2.4.2 Με βάση το σχήμα

Επειδή οι ΟΣ έχουν ειδικά σχήματα, πολλοί μελετητές χρησιμοποιούν επίσης πληροφορίες σχήματος για τη διάκριση των ΟΣ από άλλα αντικείμενα, σε εικόνες. Για το σκοπό αυτό, προτείνεται μια σειρά μεθόδων ανίχνευσης σημάτων κυκλοφορίας με βάση τα χαρακτηριστικά σχήματος τους. Η διαδικασία χρησιμοποιεί το μετασχηματισμό Hough για να εντοπίσει τις γραμμές και τους κύκλους στην εικόνα, και να συμπεράνει τη θέση των πινακίδων με βάση το σχήμα κλειστού διαστήματος που σχηματίζεται από τις γραμμές της εικόνας [28], [29]. Ωστόσο, ο υπολογισμός της μεθόδου του μετασχηματισμού Hough είναι περίπλοκος και πολύ χρονοβόρος, και δεν είναι κατάλληλος για την ανίχνευση πινακίδων κυκλοφορίας σε πραγματικό χρόνο. Επομένως, η μέθοδος ανίχνευσης σημάτων κυκλοφορίας που βασίζεται στο χρώμα και το σχήμα, έχει κάποιους περιορισμούς.

Αν και αυτές οι δύο προσεγγίσεις (με βάση το χρώμα και το σχήμα) λειτουργούν καλά ξεχωριστά, τα αποτελέσματα μπορούν να βελτιωθούν εάν συνδυαστούν [30]. Όταν συνδυαστούν, οποιαδήποτε μέθοδος μπορεί να χρησιμοποιηθεί πρώτη ενώ η δεύτερη μέθοδος, συνήθως εφαρμόζεται για το φιλτράρισμα των αποτελεσμάτων. Μετα τον εντοπισμό των

πινακίδων, το ύστερο βήμα αποτελεί η αναγνώριση. Για το βήμα αυτό, οι πιο συνηθισμένες μέθοδοι είναι τα νευρωνικά δίκτυα [31], [32], γενετικοί αλγόριθμοι, ταξινομητές AdaBoost [33] και SVMs.

2.4.3 Με βάση τη ML

Εκτός από τις παραδοσιακές μεθόδους ανίχνευσης που βασίζονται στο χρώμα και στο σχήμα, υπάρχουν επίσης μέθοδοι μηχανικής μάθησης που συνδυάζουν ταξινομητές με δυνατότητες εικόνας για τον εντοπισμό των ΟΣ. Οι ταξινομητές συνήθως χρησιμοποιούν χαρακτηριστικά εικόνας και, στη συνέχεια, αυτά τα χαρακτηριστικά συνδυάζονται με μεθόδους μηχανικής εκμάθησης για τον εντοπισμό ΟΣ στις εικόνες εισόδου [34], [35]. Κάποιοι ερευνητές [36] πρότειναν ένα σύστημα υπολογιστικής όρασης για εντοπισμό σε πραγματικό χρόνο και ισχυρή αναγνώριση των πινακίδων ορίου ταχύτητας. Σε αυτό το σύστημα, μια επικαλυπτόμενη αρχιτεκτονική δύο διανυσματικών μηχανών γραμμικής υποστήριξης χρησιμοποιείται ως ταξινομητής. Άλλοι [35] πρότειναν μια ακριβή και αποτελεσματική τεχνική εντοπισμού ΟΣ διερευνώντας το AdaBoost και την υποστήριξη διανυσματικής παλινδρόμησης (SVR) για την διακριτική εκμάθηση ανιχνευτή. Έχει προταθεί επίσης μέθοδος ανίχνευσης για τα απαγορευτικά σήματα κυκλοφορίας με βάση την προσαρμοστική λειτουργία σύντηξης με ιστογράμματα προσανατολισμένων κλίσεων (HOG) και χαρακτηριστικά τοπικών δυαδικού μοτίβου (LBP). Αν και οι παραδοσιακές μέθοδοι εντοπισμού πινακίδων κυκλοφορίας που βασίζονται στη μηχανική μάθηση έχουν επιτύχει καλύτερα αποτελέσματα από αυτά που βασίζονται σε χρώματα και σχήματα, οι μέθοδοι μηχανικής μάθησης πρέπει να εξάγουν χαρακτηριστικά HOG ή Harr και η υπολογιστική διαδικασία είναι πιο περίπλοκη. Επιπλέον, η εκφραστική ικανότητα αυτών των χαρακτηριστικών είναι περιορισμένη, και δεν είναι αρκετά ισχυρή σε διάφορα πολύπλοκα σενάρια. Έτσι, ορισμένοι μελετητές έχουν προχωρήσει σε μεθόδους DL για ανίχνευση και αναγνώριση σημάτων κυκλοφορίας.

2.4.4 Με βάση τη DL

Με τη συνεχή βελτίωση της αναγνώρισης εικόνας και υπολογιστικής ικανότητας, περισσότεροι μελετητές χρησιμοποιούν μεθόδους DL για να αναγνωρίσουν τα σήματα κυκλοφορίας. Σε σύγκριση με τις παραδοσιακές μεθόδους ML, η DL μπορεί να μάθει αυτόματα μια λειτουργία εξαγωγής χαρακτηριστικών σε ένα μεγάλο αριθμό datasets και μπορεί να επιτευχθεί, υψηλή ακρίβεια αναγνώρισης. Οι Ahmed et al. [37] πρότειναν ένα CNN με βάση το προηγούμενο βελτιωμένο πλαίσιο, επικεντρωμένο στην αναγνώριση ΟΣ. Οι Haque et al. [16] πρότειναν ένα νέο ελαφρύ μοντέλο CNN για αναγνώριση σημάτων κυκλοφορίας, και κάθε συνελικτικό στρώμα περιέχει λιγότερα από 50 χαρακτηριστικά επιτρέποντας στο CNN να εκπαιδεύεται γρήγορα ακόμη και χωρίς τη βοήθεια μιας μονάδας επεξεργασίας γραφικών. Οι Xiong et al. [38] πρότειναν μια μέθοδος που βασίζεται σε συμπιεστική ανίχνευση και διασυνδεδεμένο μοντέλο CNN με πλαίσιο 9 στρωμάτων, συμπεριλαμβανομένου ενός στρώματος εισόδου, έξι κρυφών στρωμάτων, ένα πλήρως συνδεδεμένο στρώμα και ένα στρώμα εξόδου. Οι Zhou et al. [39] πρότειναν μια γενετική δυνατότητα δικτύου αναγνώρισης ΟΣ που βασίζεται στην εκμάθηση με καλή ικανότητα γενίκευσης και υψηλή υπολογιστική απόδοση. Οι Yu et al. [40] πρότειναν ένα ισχυρό μοντέλο DL με δύο προγράμματα κατάρτισης ενισχυμένα με γενίκευση για την αναγνώριση ΟΣ. Οι Kamal et al. [41] πρότειναν ένα νέο δίκτυο, το SegU-Net, το οποίο σχηματίζεται συγχωνεύοντας τις state-of-the-art αρχιτεκτονικές τμηματοποίησης SegNet και U-Net για αναγνώριση ΟΣ από βίντεο ακολουθίες.

2.4.4.1 Παραγωγή συνθετικών δεδομένων

Μέθοδοι παραγωγής συνθετικών δεδομένων για βαθιά προπόνηση έχουν μελετηθεί εκτενώς στο παρελθόν. Οι μεθοδολογίες αυτές δύνανται να χωριστούν σε δύο σύνολα: σε μη μαθησιακές μεθόδους (δηλ. δεν υπάρχει μάθηση κατά τη διάρκεια της παραγωγής συνθετικών δεδομένων) και μεθόδους που βασίζονται στη μάθηση (δηλαδή, υπάρχει μάθηση κατά τη διάρκεια της παραγωγής συνθετικών δεδομένων). Για την εργασία ανίχνευσης, οι περισσότερες μη μαθησιακές μέθοδοι αποκόπτουν και επικολλούν αντικείμενα από μία εικόνα πάνω σε μια άλλη. Για παράδειγμα, οι Dwibedi et al.[42] κόβουν αντικείμενα με μια μάσκα τμηματοποίησης και στη συνέχεια τα αναμειγνύουν πάνω σε άλλες εικόνες από τον τομέα προορισμού. Οι Wang et al. [43] χρησιμοποιούν μια παρόμοια προσέγγιση, όπου διαφορετικά αντικείμενα από την ίδια κατηγορία αλλάζουν θέσεις. Τα μοντέλα 3D έχουν επίσης χρησιμοποιηθεί για τη δημιουργία δεδομένων για εκπαίδευση, επικάλυπτοντας 2D αποδόσεις αυτών των μοντέλων σε πραγματικές εικόνες [44]. Στο έργο της ταξινόμησης των σημάτων κυκλοφορίας, ορισμένες εργασίες χρησιμοποίησαν επεξεργασία εικόνας για τη δημιουργία δειγμάτων εκπαίδευσης από πρότυπα ΟΣ [45], [46]. Στο έργο ανίχνευσης, οι Møgelmose et al. [47] προσπάθησαν να χρησιμοποιήσουν συνθετικά δεδομένα για να εκπαιδεύσουν έναν ανιχνευτή πινακίδων Viola-Jones, αλλά τα αποτελέσματα δεν ήταν ικανοποιητικά. Οι προσεγγίσεις που βασίζονται στη μάθηση παρακινούνται από τη προϋπόθεση ότι η εκπαίδευση του μοντέλου με πιο ρεαλιστικά συνθετικά δεδομένα θα οδηγήσει σε καλύτερη απόδοση των «real-world» δεδομένων. Η μαθησιακή διαδικασία μπορεί να χρησιμοποιηθεί, για παράδειγμα, για τη δημιουργία δεδομένων με αντικείμενα σε πιο φυσική θέση, όπως οι Dvornik et al.[48] έχουν δείξει ότι η κοπή και επικόλληση αντικειμένων σε τυχαίες θέσεις, μερικές φορές, μπορεί να μην είναι ιδανική.

Ειδικότερα, οι Georgakis et al. [49] τοποθετούν περικομμένα αντικείμενα ενδιαφέροντος από δημόσια σύνολα δεδομένων σε τοποθεσίες που είναι πιθανότερο να αποτελούν επιφάνεια, με τέτοιες τοποθετήσεις, οι οποίες εκτιμώνται μέσω κατάτμησης σημασιολογίας (semantic segmentation). Η κλίμακα των αντικειμένων που θα τοποθετηθούν, καθορίζεται ανάλογα με το βάθος που σχετίζεται με τη θέση της επιφάνειας. Οι Gupta et al. [50] χρησιμοποιούν μια παρόμοια προσέγγιση για εντοπισμό κειμένου, προβλέποντας έναν χάρτη βάθους για κάθε φόντο εικόνας. Με τον χάρτη βάθους, οι περιοχές φιλτράρονται για να συγκεντρωθούν κατάλληλες περιοχές για την τοποθέτηση κειμένου. Στη συνέχεια, το κείμενο τοποθετείται πάνω σε αυτές τις περιοχές, χρησιμοποιώντας επίσης το «depth map» για τον προσδιορισμό της οπτικής του κειμένου. Στο πλαίσιο της ταξινόμησης με τη χρήση one-shot μάθησης, ο Grigorescu [51] προτείνει τη δημιουργία δεδομένων με προκαθορισμένες συναρτήσεις που κάνουν τα πρότυπα πιο ρεαλιστικά. Για να οριστούν οι παράμετροι αυτών των συναρτήσεων, εκπαιδεύεται ένα δίκτυο χρησιμοποιώντας πρότυπα (αντικείμενα one-shot) και πραγματικά δείγματα ΟΣ. Οι Kim et al. [52] προτείνουν μια προσέγγιση με κωδικοποιητή πρωτοτύπων παραλλαγής. Στην εκπαιδευτική διαδικασία, οι πραγματικές εικόνες εκπαίδευσης κωδικοποιούνται σε λανθάνοντα χώρο και στη συνέχεια αποκωδικοποιούνται σε έναν πρωτότυπο (template). Στη φάση δοκιμής, ο κωδικοποιητής χρησιμοποιείται ως εξολκέας χαρακτηριστικών και ως πλησιέστερος γείτονας και ένας ταξινομητής κοντινότερου γείτονα (nearest neighbor) χρησιμοποιείται στα χαρακτηριστικά που εξάγονται από την εικόνα δοκίμων και τα πρότυπα. Στο [53], οι συγγραφείς χρησιμοποιούν ενισχυμένη μάθηση (reinforcement learning) για να μάθουν τις παραμέτρους μιας συνθετικής γεννήτριας δεδομένων από πραγματικά δεδομένα. Σε όλα τα προαναφερθέντα έργα, η προτεινόμενη μέθοδος είτε

αξιολογείται μόνο στην εργασία ταξινόμησης είτε απαιτεί σχολιασμένα πραγματικά δεδομένα από τον προβληματικό τομέα.

2.5 Δυσκολίες συστήματος

Μία από τις σημαντικότερες δυσκολίες που αντιμετωπίζει η ADAS είναι η αντίληψη του τοπίου και της καθοδήγησης των οχημάτων σε πραγματικούς εξωτερικούς χώρους σκηνές, συμπεριλαμβανομένης της ανίχνευσης πεζών [54], [55], αντίληψη περιβάλλοντος οχημάτων [56]–[58], ανίχνευση ΟΣ [59], και ούτω καθεξής. Η οδήγηση είναι μια δραστηριότητα που σχεδόν αποκλειστικά εξαρτάται από τις οπτικές γνώσεις και μία από τις εργασίες που εμπλέκονται στην καλή οδήγηση είναι να αναγνωρίζουν την ΟΣ. Αλλιώς είναι δυνατόν να συναποτελέσει κίνδυνο για τη ζωή των ανθρώπων λόγω έλλειψης συγκέντρωσης ή άγνοιας.

Σε ορισμένες ρεαλιστικές καταστάσεις, ο εντοπισμός των ΟΣ είναι δύσκολος, αν όχι δυσεπίλυτος. Ορισμένες από αυτές τις καταστάσεις απεικονίζονται στο Σχήμα 2 και που αναφέρονται παρακάτω:

1. Εμπόδια, π.χ. δέντρα, αυτοκίνητα και άνθρωποι μπορεί να επηρεάσουν την αναγνώριση ΟΣ (Σχήμα 2α).
2. Καιρικές συνθήκες όπως χιόνι, βροχή και ομίχλη και ρύπανση, μπορούν να κάνουν τις φάσεις ανίχνευσης και αναγνώρισης πολυσύνθετες (Σχήμα 2β).
3. Ξεθώριασμα χρώματος: Το χρώμα της πινακίδας ξεθωριάζει με το χρόνο ως αποτέλεσμα της μακράς έκθεσης στο ηλιακό φως και της αντίδρασης της βαφής στον αέρα (Σχήμα 2γ).
4. Αλλαγές στις συνθήκες φωτισμού σε διάφορες περιόδους (ημέρα και νύχτα) (Σχήμα 2δ).

Σχήμα 2. Παραδείγματα προκλήσεων της αναγνώρισης των ΟΣ [60].

2.6 Εξωγενείς παράγοντες επηρεασμού της αναγνώρισης ΟΣ

Με την ανάπτυξη της ΤΝ και της DL, το TSR αναπτύσσεται επίσης γρήγορα, πολλά μοντέλα υψηλών προδιαγραφών περιλαμβάνουν πλέον συστήματα υποβοήθησης οδηγού TSR για να βοηθήσουν τους οδηγούς στο δρόμο, με ασφάλεια. Τα τρέχοντα TSR έχουν υψηλή ακρίβεια

μόνο στις εικόνες που λαμβάνονται σε ηλιόλουστες ημέρες, και εφόσον οι πινακίδες κυκλοφορίας είναι ανεμπόδιστες, όμως σε δυσμενείς καιρικές συνθήκες (π.χ. ομίχλη, βροχή, χιόνι κ.λπ.), δύσκολες συνθήκες φωτισμού (π.χ. νύχτα, άμεσο ηλιακό φως, σκιές κ.λπ.) και εάν τα σήματα κυκλοφορίας είναι καλυμμένα, προκύπτουν ψευδείς συναγερμοί ή δεν προκύπτει αποτελέσματα αναγνώρισης.

2.6.1 Επιπτώσεις της βροχής

Το TSR στις εικόνες που λαμβάνονται σε βροχερές μέρες είναι ένα πολύ δύσκολο έργο επειδή η ομίχλη είναι στατική και δεν έχει προφανείς κινήσεις, ενώ η βροχή ή το χιόνι έχει δυναμικές κινήσεις [61]. Όλα αυτά είναι πολύ δύσκολα για το TSR, το οποίο απαιτεί ήδη ανίχνευση αντικειμένων σε πραγματικό χρόνο. Το μέγεθος, η ταχύτητα και το σχήμα των σταγόνων βροχής είναι αβέβαια, οι θέσεις των σταγόνων είναι τυχαίες [62]. Οι πτώση σταγόνων βροχής αλλά και η κάθοδος τους μειώνουν την ορατότητα και εμποδίζουν την κάμερα, γεγονός που μπορεί να προκαλέσει παραμόρφωση στην εικόνα, οπότε οι συνθήκες θα αναγκάσουν τη κάμερα και το σύστημα να προβεί σε εσφαλμένη αναγνώριση. Εάν αυτές οι μικροσκοπικές σταγόνες νερού εμφανιστούν στα παράθυρα των αυτοκινήτων, αφήνουν ένα ίχνος. Εάν η βροχή είναι έντονη, οι σταγόνες βροχής θα είναι αρκετά γρήγορες για να καλύψουν τα παράθυρα σαν κουρτίνα, γεγονός που θα κάνει τις εικόνες ΟΣ που τραβηχτήκαν, θολές. Μετά τη βροχή, το νερό στο έδαφος σαν καθρέφτης που αντανακλά τις ΟΣ μπορεί να κάνει το TSR να δυσκολευτεί πολύ, μέχρι και να μην επιτύχει καν. Ως αποτέλεσμα αυτών των προβλημάτων, οι Chen et al. πρότειναν ένα σύστημα TSR με μια μονάδα βελτίωσης ορατότητας [63], έναν συνελικτικό κωδικοποιητή, αναμεμειγμένο με έναν συνελικτικό αποκωδικοποιητή και ένα σκοτεινό κανάλι [64]. Στα πειράματα, το ποσοστό ακρίβειας βελτιώθηκε κατά 50% χρησιμοποιώντας τις εικόνες από βροχερές μέρες.

2.6.2 Επιπτώσεις του φωτισμού

Ο φωτισμός είναι κρίσιμος παράγοντας για το TSR. Κατά τη διάρκεια της ημέρας, εάν ο ήλιος λάμπει απευθείας στην ΟΣ ή τη κάμερα, θα βρεθούν να είναι υπερβολικά εκτεθειμένα. Τη νύχτα, σε πολλά σημεία είναι συχνά πολύ σκοτεινά, και είναι αναπόφευκτο πως τα φώτα του αυτοκινήτου θα αντανακλούν στις πινακίδες εν μέσω οδήγησης. Αυτό θα οδηγήσει σε εσφαλμένη αναγνώριση των ΟΣ. Σε απάντηση σε αυτά τα προβλήματα, οι Greenhalgh et al. πρότειναν μια μέθοδο ανίχνευσης που προσαρμόζει την ένταση του εισερχόμενου φωτός και επιτρέπει καλύτερη TSR [65]. Μειώνει αποτελεσματικά την επίδραση των σκιών και του φωτισμού στην αναγνώριση που συνδυάζεται με τα γεωμετρικά χαρακτηριστικά των σημάτων κυκλοφορίας για να αυξήσει την ακρίβεια αναγνώρισης υπό διάφορες συνθήκες φωτισμού. Εν τω μεταξύ, κάποιοι ερευνητές πρότειναν ένα byte-MCT και ταξινομητή AdaBoost με βάση την ΟΣ για το TSR, το οποίο πρώτα εφάρμοσε SVM για να επαληθεύσει τις υποψήφιες περιοχές και στη συνέχεια το CNN για να εξαγάγει οπτικά χαρακτηριστικά. Αυτή η μέθοδος πέτυχε πολύ καλά αποτελέσματα με ισχυρή συνέπεια υπό διάφορες συνθήκες φωτισμού.

2.6.3 Επιπτώσεις της ομίχλης

Ενώ τραβάμε μια φωτογραφία χρησιμοποιώντας ψηφιακή φωτογραφική μηχανή στο εξωτερικό περιβάλλον υπό συνθήκες ομίχλης, το φως του περιβάλλοντος θα διασκορπιστεί σοβαρά, με αποτέλεσμα να εμφανιζονται θολά χαρακτηριστικά εικόνας, με συνέπεια τη μη εξαγωγή χαρακτηριστικών και των άλλων σχετικών λειτουργιών [66]. Μέσω της ανάλυσης ενός μεγάλου αριθμού εικόνων ομίχλης και καθαρών εικόνων της ίδιας σκηνής, διαπιστώνουμε ότι οι εικόνες

ομίχλης έχουν συγκεκριμένα χαρακτηριστικά. Η ανάλυση αυτών των χαρακτηριστικών εικόνας προσδιορίζει αυτόματα την θολούρα στην εικόνα με τον ταξινομητή για την ταξινόμηση μοτίβων. Άρα, σύμφωνα με αυτά τα χαρακτηριστικά, η αντίθεση της εικόνας ομίχλης μειώνεται σημαντικά σε σύγκριση με τις καθαρές εικόνες, και ο βαθμός εξασθένησης της αντίθεσης δείχνει εκθετικά χαρακτηριστικά εξασθένησης με την αύξηση των πληροφοριών δομής της σκηνής. Καθώς τα κύρια χαρακτηριστικά της εικόνας είναι θολά, το περίγραμμα και η άκρη ξεθωριάζουν [67]. Τα χρώματα της εικόνας εμφανίζονται σε παραμόρφωση και μετατόπιση, ο κορεσμός των χρωμάτων μειώνεται. Η κατανομή των τιμών των εικονοστοιχείων της εικόνας συγκεντρώνεται, το δυναμικό εύρος συρρικνώνεται και το ιστόγραμμα εικονοστοιχείων παρουσιάζει ομοιόμορφη κατανομή [68]. Τέλος, από την άποψη του πεδίου συχνότητας εικόνας, υπάρχει μεγάλη ποσότητα πληροφοριών χαμηλής συχνότητας στην εικόνα, και οι πληροφορίες υψηλής συχνότητας είναι σχετικά μειωμένες. Όσον αφορά το οπτικό αποτέλεσμα και την εμφάνιση, το επίπεδο κλίμακας του γκρι των εικόνων αλλάζει ομοιόμορφα, οι λεπτομερείς πληροφορίες μειώνονται και η άκρη του περιγράμματος της εικόνας εξασθενεί. Οπότε, σε καιρό ομίχλης, τα αίτια υποβάθμισης και θόλωσης της ποιότητας της εικόνας οφείλεται στην ύπαρξη αιωρούμενων σωματιδίων αερολύματος στον αέρα. Τα σωματίδια αερολύματος συγχέουν τη μετάδοση φωτός, προκαλούν σοβαρή απορρόφηση και σκέδαση φωτός και παράγουν άμεσα την εξασθένηση της ενέργειας που μεταφέρεται ανακλώντας τις ακτίνες φωτός από την επιφάνεια του αντικειμένου στο πλάνο [69]. Σε ολόκληρη την οπτική διαδρομή των ακτινών φωτός προς τη συσκευή απεικόνισης, το ατμοσφαιρικό φως του περιβάλλοντος διασκορπίζεται επίσης στη συσκευή απεικόνισης έτσι ώστε να συμμετέχει στην απεικόνιση των αντικειμένων σκηνής. Οι δύο παράγοντες συνεργάζονται για να μειώσουν την αντίθεση της εικόνας και την ανάλυση της εικόνας.

Η επίδραση των σωματιδίων αερολύματος στη σκέδαση του φωτός ποικίλλει ανάλογα με το μέγεθος και το σχήμα των σωματιδίων. Παράλληλα, με την απεικόνιση του αντικειμένου στο περιβάλλον, τα ίδια τα σωματίδια συμμετέχουν και στην απεικόνιση της σκηνής, η οποία θεωρείται θόρυβος εικόνας, με αποτέλεσμα την απώλεια πληροφοριών της λεπτομέρειας της εικόνας. Το φως που το αντικείμενο, από την επιφάνεια του, αντανακλά στη σκηνή, διασκορπίζεται κατά μήκος της διαδρομής φωτός προς τον φακό της κάμερας. Ταυτόχρονα, το φως που αντανακλάται από άλλες επιφάνειες αντικειμένων στη σκηνή συμμετέχει επίσης στην απεικόνιση μετά από πολυεπίπεδη σκέδαση, με αποτέλεσμα την θολή εικόνα [70]. Σε καιρό ομίχλης, η ένταση της ακτινοβολίας του φωτός θα εξασθενήσει λόγω της νεφοκάλυψης, κάτι που αποφέρει μείωση της φωτεινής εντάσεως στο περιβάλλον, με αποτέλεσμα η φωτεινότητα της εικόνας που συλλέγεται να μην είναι αρκετή [71]. Συνοψίζοντας, ο κύριος λόγος της θολής εικόνας σε περιβάλλον ομίχλης είναι η παρουσία ενός μεγάλου αριθμού αιωρούμενων σωματιδίων στο περιβάλλον, γεγονός που καθιστά σοβαρή τη σκέδαση του φωτός.

2.6.4 Ατμοσφαιρικό Μοντέλο Σκέδασης

Η σκέδαση αναφέρεται στο φαινόμενο κατά το οποίο, ενώ το φως διέρχεται από ένα μη ενιαίο μέσο, η φωτεινή ακτίνα αποκλίνει από την αρχική κατεύθυνση διάδοσης (διάσπαρτο φως). Τα σωματίδια σκέδασης στην ατμόσφαιρα περιλαμβάνουν μικροσκοπικούς κρυστάλλους πάγου, μικροσκοπικά σταγονίδια νερού κ.λπ. Η αντανάκλαση σκέδασης είναι ο κύριος λόγος για την παρεμβολή του πεδίου στην υπολογιστική όραση. Το 1925, προτάθηκε ότι η χαμηλή ορατότητα της ομιχλώδους εικόνας δημιουργείται από την απορρόφηση και τη σκέδαση των αντανακλάσεων των σωματιδίων στην ατμόσφαιρα. Στην πραγματική λήψη της εικόνας, η

απορρόφηση του φωτός από τα ατμοσφαιρικά σωματίδια γενικά αγνοείται και λαμβάνεται υπόψη μόνο η σκέδαση του φωτός.

Σε καιρό ομίχλης, η σκέδαση των ατμοσφαιρικών σωματιδίων όχι μόνο συμβάλει στην απώλεια μέρους του ανακλώμενου φωτός από την επιφάνεια του αντικειμένου, αλλά και στην προσθήκη του ατμοσφαιρικού διάσπαρτου φωτός στο ανακλώμενο από το αντικείμενο φως. Ο συνδυασμός των προαναφερθέντων οδηγεί στο συμπέρασμα ότι τα χαρακτηριστικά αντίθεσης και χρώματος των εικόνων που συλλέγονται σε ομιχλώδεις καιρικές συνθήκες θα εξασθενήσουν σοβαρά. Επιπρόσθετα, με βάση τη θεωρία σκέδασης, το ατμοσφαιρικό μοντέλο σκέδασης έχει εκτεταμένο αντίκτυπο στις εικόνες. Το 1999, οι Nayar et al. [72] εξήγησαν τη διαδικασία απεικόνισης σε ομιχλώδεις καιρικές συνθήκες δημιουργώντας το αντίστοιχο μαθηματικό μοντέλο.

3 ΚΕΦΑΛΑΙΟ 3: Βαθιά Μάθηση και Νευρωνικά Δίκτυα

3.1 Τι είναι η Βαθιά Μάθηση

Η βαθιά μάθηση (DL) είναι τμήμα της ML. Προσφέρει λειτουργία επιδεξιότητας εντός του συστήματος, μιμούμενο τα μοτίβα του εγκέφαλου μας όταν λαμβάνει αποφάσεις. Ένα σύστημα DL διδάσκεται από την εξέταση διαφορετικών προτύπων και ειδών δεδομένων, όπως επίσης και κατά την εξαγωγή των συμπερασμάτων βασιζόμενο σε αυτά. Τα μέρη ενός συστήματος DL είναι ένα μεγάλο PSU (τροφοδοτικό), η κάρτα GPU και οπωσδήποτε αρκετά επαρκής μνήμη RAM. Το γεγονός ότι η γένεση του δικτύου αυτού είναι πολύπλοκη, απαιτείται πολύς χρόνος και μόχθος για την αποτελεσματική εκπαίδευση του. Η αφετηρία της δομής της DL είναι σαφώς τα Συνελικτικά Νευρωνικά Δίκτυα, όπως επίσης τα Επαναλαμβανόμενα Νευρωνικά Δίκτυα, τα Μη-Εποπτευόμενα Προπαιδευμένα Δίκτυα καθώς και το Αναδρομικό Νευρωνικό Δίκτυο.

Μπορούμε να καταλάβουμε πώς λειτουργεί η βαθιά μάθηση χρησιμοποιώντας το παράδειγμα της διάκρισης ενός σκύλου από μια γάτα. Το μοντέλο βαθιάς μάθησης χρησιμοποιεί εικόνες ως είσοδο και τις στέλνει κατευθείαν στους αλγόριθμους. Αυτό σημαίνει ότι δεν υπάρχει ανάγκη για έναν άνθρωπο να κάνει τη δουλειά της εξαγωγής χαρακτηριστικών από εικόνες. Τα διάφορα επίπεδα τεχνητού νευρωνικού δικτύου λαμβάνουν τις εικόνες και προβλέπουν τα αποτελέσματα. Οι εικονικοί βοηθοί των διαδικτυακών παρόχων υπηρεσιών, όπως η Alexa, η Siri και η Cortana, χρησιμοποιούν βαθιά μάθηση για να αναγνωρίσουν την ομιλία και τη γλώσσα που χρησιμοποιούμε για να αλληλοεπιδράσουμε με αυτά.

Η εκπαίδευση βαθιάς μάθησης είναι όταν ένα βαθύ νευρωνικό δίκτυο (DNN) «μαθαίνει» πώς να αναλύει ένα προκαθορισμένο σύνολο δεδομένων και να κάνει προβλέψεις. Περιλαμβάνει πολλές δοκιμές και σφάλματα έως ότου το δίκτυο είναι σε θέση να εξάγει με ακρίβεια συμπεράσματα με βάση το επιθυμητό αποτέλεσμα. Τα DNNs αναφέρονται συχνά ως ένας τύπος τεχνητής νοημοσύνης, καθώς μιμούνται την ανθρώπινη νοημοσύνη μέσω της χρήσης τεχνητών νευρώνων. Ένα DNN έχει τη δυνατότητα να κοσκινίζει αφηρημένους τύπους δεδομένων, όπως εικόνες και ηχογραφήσεις.

Κάθε φορά που καταλήγει σε ένα λανθασμένο συμπέρασμα, αυτό το αποτέλεσμα ανατροφοδοτείται πίσω στο σύστημα έτσι ώστε να «μαθαίνει» από το λάθος του. Αυτή η διαδικασία κάνει τις συνδέσεις μεταξύ των τεχνητών νευρώνων ισχυρότερες με την πάροδο του χρόνου και αυξάνει την πιθανότητα, το σύστημα να κάνει ακριβείς προβλέψεις στο μέλλον.

Καθώς του παρουσιάζονται νέα δεδομένα, το DNN θα οφείλει να βρίσκεται σε θέση να κατηγοριοποιήσει και να αναλύσει νέες και πιθανώς πιο σύνθετες πληροφορίες. Τελικά, θα συνεχίσει να μαθαίνει και να διαμορφώνεται ως πιο διαισθητικό με την παρέλευση του χρόνου.

3.1.1 Γιατί Βαθιά Μάθηση

Η επιλογή των δυνατοτήτων που αντιπροσωπεύει ένα σύνολο δεδομένων έχει σημαντικό αντίκτυπο στην επιτυχία ενός συστήματος μηχανικής μάθησης. Καλύτερα αποτελέσματα δεν μπορούν να επιτευχθούν χωρίς να προσδιοριστούν ποιες πτυχές του ζητήματος που πρέπει να συμπεριληφθούν για την εξαγωγή χαρακτηριστικών , θα ήταν πιο χρήσιμα για τον αλγόριθμο μηχανικής μάθησης. Αυτό απαιτεί τη συνεργασία ενός ειδικού στη μηχανική μάθηση με εμπειρογνώμονα του πεδίου ώστε να αποκτηθεί ένα σύνολο χρήσιμων χαρακτηριστικών. Ένας βιολογικός εγκέφαλος μπορεί εύκολα να προσδιορίσει σε ποιες πτυχές του προβλήματος πρέπει να επικεντρωθεί με συγκριτικά λίγη καθοδήγηση. Αυτό δεν συμβαίνει με τους τεχνητούς παράγοντες, καθιστώντας έτσι δύσκολη τη δημιουργία συστημάτων εκμάθησης υπολογιστών που μπορούν να ανταποκριθούν σε υψηλών διαστάσεων δεδομένα και να εκτελέσουν δύσκολες εργασίες AI (artificial intelligence). Οι επαγγελματίες μηχανικής μάθησης έχουν περάσει τεράστιο χρόνο για να εξαγάγουν πληροφορίες- χαρακτηριστικά από τα δεδομένα. Την εποχή της εισαγωγής στην βαθιά μάθηση οι υπερσύγχρονοι αλγόριθμοι μηχανικής μάθησης είχαν ήδη κοστίσει δεκαετίες ανθρώπινης προσπάθειας συσσώρευσης σχετικού συνόλου χαρακτηριστικών που απαιτούνται για την ταξινόμηση της εισόδου. Η βαθιά μάθηση έχει ξεπεράσει αυτούς τους συμβατικούς αλγόριθμους σε ακρίβεια καθώς τα χαρακτηριστικά διδάχθηκαν από τα δεδομένα χρησιμοποιώντας μια διαδικασία γενικής εκμάθησης αντί να έχουν σχεδιαστεί από ανθρώπους (μηχανικούς). Τα βαθιά δίκτυα έχουν βελτιώσει δραματικά την υπολογιστική όραση και την αυτόματη μετάφραση, και θεωρούνται αποδεδειγμένα μια αποτελεσματική τεχνική AI που έχει την ικανότητα να αναγνωρίζει προφορικά λέξεις σχεδόν τόσο καλά όσο μπορεί ένας άνθρωπος. Έχει επιδείξει εξαιρετική ακρίβεια στη μοντελοποίηση της μηχανικής μάθησης, αλλά επίσης εξαιρετική δύναμη γενίκευσης που έχει προσελκύσει ακόμη και επιστήμονες από άλλους ακαδημαϊκούς κλάδους. Στις μέρες μας, χρησιμοποιείται ως οδηγός για τη λήψη βασικών αποφάσεων σε τομείς όπως η ιατρική, χρηματοδότηση, μεταποίηση και όχι μόνο.

Η βαθιά μάθηση έγινε γνωστή το 2007, με ελπιδοφόρα αποτελέσματα στα προβλήματα αντίληψης όπως της ακοής και της όρασης, στα οποία οι άνθρωποι είναι πολύ καλοί, αλλά για τις μηχανές, τότε, ήταν αρκετά εύστοχο. Έδωσε τη δυνατότητα στους επιστήμονες να αξιοποιήσουν την τεράστια υπολογιστική ισχύ και να χρησιμοποιήσουν μεγάλους όγκους δεδομένων (ήχο, βίντεο), να διδάξουν στους υπολογιστές πώς να κάνουν πράγματα που φαίνονται διαισθητικά φυσικά για τους ανθρώπους, όπως ο εντοπισμός αντικειμένων στις φωτογραφίες, η αναγνώριση λέξεων ή προτάσεων και η μετάφραση ενός κειμένου από μια γλώσσα σε άλλη. Έχει επίσης επιτρέψει στα μηχανήματα να παράγουν την απομαγνητοφώνηση από ένα ηχητικό απόσπασμα (αναγνώριση ομιλίας), για να προσδιορίσει αν ένα μήνυμα ηλεκτρονικού ταχυδρομείου είναι ανεπιθύμητο ή όχι, την πιθανότητα εάν ένας πελάτης θα αποπληρώσει το δάνειό του και ούτω καθεξής. Για όσο θα υπάρχουν αρκετά δεδομένα για την εκπαίδευση μηχανών, οι δυνατότητες θα είναι ατελείωτες. Έχει επιτύχει αποτελέσματα, τεχνολογίας αιχμής, σε πολλές εφαρμογές, όπως η ανάλυση φυσικής γλώσσας, η μοντελοποίηση της γλώσσας, η αναγνώριση εικόνας και χαρακτήρων, παίζοντας απαιτητικά παιχνίδια, βιντεοπαιχνίδια και σε άλλες εφαρμογές. Στις μέρες μας, πολλές εταιρείες

τεχνολογικών κολοσσών - Facebook, Baidu, Amazon, Microsoft και Google έχουν αναπτύξει εμπορικές εφαρμογές βαθιάς εκμάθησης. Αυτές οι εταιρείες έχουν τεράστιο όγκο δεδομένων και η βαθιά μάθηση λειτουργεί καλά όποτε υπάρχουν τεράστιοι όγκοι δεδομένων και σύνθετα προβλήματα προς επίλυση. Πολλές εταιρείες χρησιμοποιούν βαθιά μάθηση για την ανάπτυξη πιο χρήσιμων και ρεαλιστικών εκπροσώπων εξυπηρέτησης πελατών - Ρομπότ συνομιλίας. Ειδικότερα, η βαθιά μάθηση έχει επηρεάσει ικανοποιητικά σε ιστορικά δύσκολους τομείς μηχανικής μάθησης, όπως: η ταξινόμηση εικόνας, η αναγνώριση ομιλίας, μεταγραφή χειρόγραφου, τα βελτιωμένα αυτοοδηγούμενα αυτοκίνητα, οι ψηφιακοί βοηθοί (Amazon, Alexa, Google Now, Siri, Microsoft Cortana), η βελτιωμένη στόχευση διαφημίσεων (Baidu, Google, Bing), τα βελτιωμένα αποτελέσματα αναζήτησης στον ιστό, η ικανότητα απάντησης σε ερωτήσεις φυσικής γλώσσας και παιχνίδια (υπεράνθρωπος Go, Shogi, σκάκι).

Η εξαιρετική απόδοση των βαθιών μοντέλων μπορεί να αποδοθεί κυρίως στην ευελιξία τους στην αναπαράσταση ενός εξαιρετικά πλούσιου συνόλου μη γραμμικών συναρτήσεων καθώς και στην επινόηση μεθόδων για την αποτελεσματική εκπαίδευση αυτών των ισχυρών δικτύων. Επιπλέον, χρησιμοποιώντας διάφορες τεχνικές κανονικοποίησης, εξασφάλισαν ότι τα βαθιά μοντέλα με τεράστιους αριθμούς ελεύθερων παραμέτρων, είναι στατιστικά επιθυμητά με την έννοια ότι θα γενικευθούν καλά ως προς τα αθέατα δεδομένα. Η αυτόματη και γενική προσέγγιση της εκμάθησης χαρακτηριστικών σε βαθιά μοντέλα επιτρέπει σε κάποιον να τα χρησιμοποιήσει σε διαφορετικές εφαρμογές (π.χ. ταξινόμηση εικόνας, αναγνώριση ομιλίας, μοντελοποίηση γλώσσας και ανάκτηση πληροφοριών) με σχετικά μικρές προσαρμογές. Ως εκ τούτου, τα βαθιά μοντέλα φαίνεται να αγνοούν τον τομέα με την έννοια ότι για την χρήση του σε διαφορετικές εφαρμογές, απαιτούνται προσαρμογές μόνο σε ένα μικρό ποσοστό του συγκεκριμένου τομέα. Ιδανικά, η λήθη του τομέα των βαθιών δικτύων συμφέρει, καθώς η πρόσβαση σε ένα καθολικό και γενικό μοντέλο μειώνει την ταλαιπωρία της προσαρμογής για νέες εφαρμογές.

Η βαθιά μάθηση είναι ακόμα στα σπάργανα, αλλά είναι πιθανό ότι θα έχει πολλές επιτυχίες στο εγγύς μέλλον, καθώς απαιτεί λίγη μηχανική ώθηση και έτσι μπορεί επωφεληθεί από τον τεράστιο όγκο δεδομένων και της υπολογιστικής ισχύος. Η βαθιά μάθηση έχει πέτυχει σε προηγουμένως άλυτα προβλήματα που ήταν αρκετά δύσκολο να επιλυθούν χρησιμοποιώντας μηχανική μάθηση καθώς και άλλα δίκτυα. Στο εγγύς μέλλον, η βαθιά μάθηση μπορεί να βοηθήσει τους ανθρώπους στην ανάπτυξη λογισμικού, την επιστήμη και πολλά άλλα. Η ενσωμάτωση της βαθιάς μάθησης με άλλες τεχνικές τεχνητής νοημοσύνης μπορεί να επιφέρει εκπληκτικά αποτελέσματα που θα έχουν μεγάλο αντίκτυπο στον τομέα της τεχνολογίας.

3.1.2 Προκλήσεις

Τα δίκτυα βαθιάς μάθησης έχουν φέρει το δικό τους σύνολο προβλημάτων και προκλήσεων που υπερίσχυσε των πλεονεκτημάτων των βαθιών αρχιτεκτονικών για αρκετές δεκαετίες. Η εκπαίδευση αυτών των αρχιτεκτονικών για γενική χρήση ήταν πρακτικά αργή. Με περιορισμένη υπολογιστική δύναμη, τα δίκτυα βαθιάς μάθησης είχαν ήδη ξεπεραστεί από άλλες προσεγγίσεις όπως μέθοδοι πυρήνα. Με τη σημαντική αύξηση της υπολογιστικής ισχύος (ιδιαίτερα στην GPU) και η πρόσβαση σε σύνολα δεδομένων άνοιξαν το δρόμο για την επιστροφή του. Ωστόσο, παρά τις αξιοσημείωτες προόδους σε αυτόν τον τομέα, η εκπαίδευση βαθιών μοντέλων με έναν τεράστιο αριθμό ελεύθερων παραμέτρων είναι μια περίπλοκη διαδικασία. Πολλές ερευνητικές εργασίες έχουν αφιερωθεί στη δημιουργία μεθόδων αποτελεσματικής εκπαίδευσης για βαθιές αρχιτεκτονικές. Οι στρατηγικές που αναφέρονται στη

βιβλιογραφία που ασχολούνται με τις δυσκολίες της κατάρτισης βαθιών δικτύων, περιλαμβάνουν την ανάπτυξη καλύτερων βελτιστοποιητών, χρησιμοποιώντας καλά σχεδιασμένες στρατηγικές προετοιμασίας, τοπικές λειτουργίες ενεργοποίησης που βασίζονται στον ανταγωνισμό και τη χρήση συνδέσεων skip μεταξύ των επιπέδων με στόχο τη βελτίωση της ροή πληροφοριών. Ωστόσο, η βαθιά εκπαίδευση δικτύου εξακολουθεί να αντιμετωπίζει προβλήματα κάτι που προκαλείται από τη στοίβαξη αρκετών μη γραμμικών μετασχηματισμών και πρέπει να αντιμετωπιστεί. Επιπλέον, η βαθιά μάθηση περιλαμβάνει τη χρήση μεγάλων ποσοτήτων δεδομένων για τη σταδιακή μάθηση. Ενώ μεγάλες ποσότητες δεδομένων είναι διαθέσιμες σε πολλές εφαρμογές, σε ορισμένες περιοχές, είναι σπάνια διαθέσιμες άφθονες ποσότητες δεδομένων. Πιο ευέλικτα μοντέλα απαιτούνται για την επίτευξη βελτιωμένης ικανότητας μάθησης όταν μόνο ένας περιορισμένος όγκος δεδομένων είναι διαθέσιμος.

Τα δίκτυα βαθιάς μάθησης είναι πολύ καλά στην επίλυση ενός προβλήματος. Ωστόσο, χρησιμοποιώντας τα βαθιά δίκτυα για την επίλυση ενός πολύ παρόμοιου προβλήματος απαιτούν επανεκπαίδευση και επαναξιολόγηση. Αν και υπάρχουν πολλές εξελίξεις σε αυτή την πτυχή, απαιτείται περισσότερη δουλειά στην ανάπτυξη μοντέλων βαθιάς μάθησης που μπορούν να εκτελούν πολλαπλές εργασίες χωρίς την ανάγκη ανακατασκευής ολόκληρης της αρχιτεκτονικής.

3.2 Τi είναι η Μηχανική Μάθηση

Η ML είναι μια εφαρμογή AI που χρησιμοποιεί αλγόριθμους που αναλύουν δεδομένα, αποκτούν γνώσεις από αυτές τις πληροφορίες και στη συνέχεια χρησιμοποιούν αυτές τις γνώσεις για να λάβουν αποφάσεις. Με τη χρήση της Μηχανικής Μάθησης, τα υπολογιστικά συστήματα μπορούν να προγραμματιστούν για να κατανοήσουν και να μάθουν από δοθέντα δεδομένα εισόδου χωρίς να χρειάζεται να αναπρογραμματίζονται τακτικά. Αν το θέσουμε διαφορετικά, βελτιώνουν συνεχώς τις επιδόσεις τους σε μια συγκεκριμένη εργασία - όπως παίζοντας ένα συγκεκριμένο παιχνίδι - χωρίς περαιτέρω βοήθεια από έναν άνθρωπο. Η μηχανική μάθηση χρησιμοποιείται σε μεγάλο βαθμό σε διάφορους κλάδους, συμπεριλαμβανομένων των οικονομικών, της υγειονομικής περίθαλψης, της τέχνης και της επιστήμης. Δύο δημοφιλείς μέθοδοι Μηχανικής Μάθησης που χρησιμοποιούνται πολύ συχνά είναι η Μηχανική μάθηση με python και με R.

Το παράδειγμα αναγνώρισης της εικόνας της γάτας και του σκύλου, πάλι, μπορεί να εξηγήσει πώς λειτουργούν τα μοντέλα μηχανικής μάθησης. Το μοντέλο ML χρησιμοποιεί φωτογραφίες εισόδου τόσο των γατών όσο και των σκύλων για προσδιορισμό τους, εξάγοντας χαρακτηριστικά όπως το σχήμα, το ύψος, τη μύτη και τα μάτια πριν εφαρμόσει τη μέθοδο ταξινόμησης και προβλέψει το αποτέλεσμα. Σε ένα ευρύ φάσμα επιχειρήσεων, η μηχανική μάθηση επιτρέπει μια ποικιλία αυτοματοποιημένων διαδικασιών, από εταιρείες ασφάλειας δεδομένων που αναζητούν κακόβουλο λογισμικό έως οικονομικούς εμπειρογνώμονες που αναζητουν ειδοποιήσεις για κερδοφόρες συναλλαγές. Μια υπηρεσία ροής μουσικής κατ' απαίτηση χρησιμεύει ως απλή απεικόνιση ενός αλγόριθμου μηχανικής μάθησης. Οι αλγόριθμοι μηχανικής εκμάθησης βοηθούν στη σύνδεση των μουσικών προτιμήσεων των ανθρώπων με εκείνες άλλων ανθρώπων που τους αρέσει το ίδιο είδος μουσικής. Με αυτόν τον τρόπο, το μηχάνημα μπορεί να καταλάβει ποια νέα τραγούδια ή καλλιτέχνες να προτείνει σε έναν ακροατή. Πολλές υπηρεσίες που παρέχουν αυτοματοποιημένες προτάσεις χρησιμοποιούν αυτήν τη μέθοδο, η οποία συχνά αναφέρεται ως AI.

3.2.1 Περιορισμοί

Το Feature engineering είναι η διαδικασία χειρισμού χαρακτηριστικών με τέτοιο τρόπο ώστε να οδηγούν σε καλά μοντέλα. Με τη μηχανική μάθηση, πρέπει να δοθεί η πληροφορία στον υπολογιστή για το ποια είναι τα μήλα και τα πορτοκάλια δίνοντάς του πληροφορίες μεγέθους ή διαφορές χρώματος μεταξύ τους. Στα νευρωνικά δίκτυα, στη βαθιά μάθηση, επιλέγει αυτόματα και διαφοροποιεί και τους δύο τύπους φρούτων χωρίς να απαιτείται ανθρώπινη βοήθεια. Στην ουσία, η μηχανική χαρακτηριστικών εκτελείται σκόπιμα από ανθρώπους στη μηχανική μάθηση, ενώ πραγματοποιείται αυτόματα από το μοντέλο στη βαθιάς μάθησης. Επιπλέον, σύνθετα προβλήματα AI όπως η αναγνώριση παραστάσεων, η επεξεργασία φυσικής γλώσσας, κ.λπ., δεν μπορούν να επιλυθούν με τεχνικές μηχανικής μάθησης. Όπως επίσης, η μηχανική εκμάθηση δεν αποδίδει με τον καλύτερο δυνατό τρόπο με μεγαλύτερα σύνολα δεδομένων, ενώ η βαθιά μάθηση μπορεί να αποδώσει τόσο με μεγάλα όσο και με μικρά σύνολα δεδομένων.

3.2.2 Διαφορές με τη Βαθιά Μάθηση

Σε αντίθεση με τα συστήματα ML, τα οποία απαιτούν από τον χρήστη να κωδικοποιεί χειροκίνητα τα εφαρμοζόμενα χαρακτηριστικά με βάση τον τύπο δεδομένων, ένα σύστημα βαθιάς μάθησης επιδιώκει να μάθει τέτοια χαρακτηριστικά χωρίς πρόσθετη ανθρώπινη συμβολή (για παράδειγμα, τιμή εικονοστοιχείων, σχήμα και προσανατολισμός). Όσον αφορά τα δεδομένα, τα συστήματα DL χρειάζονται πολύ πιο ισχυρό υλικό από ό, τι τα συστήματα ML λόγω του όγκου των δεδομένων που χειρίζονται και της πολυπλοκότητας των μαθηματικών υπολογισμών που συνεπάγονται οι αλγόριθμοι που χρησιμοποιούνται. Ως προς το χρόνο ανάπτυξης, ένα σύστημα DL είναι δυνατόν να εξακολουθήσει από κάποιες ώρες μέχρι μερικές βδομάδες, καθώς χειρίζεται μεγάλες ποσότητες δεδομένων, έχει πολλές παραμέτρους και χρησιμοποιεί περίπλοκους μαθηματικούς υπολογισμούς, ενώ η ML μπορεί να ολοκληρωθεί σε μόλις λίγα δευτερόλεπτα έως κάποιες ώρες. Επίσης, οι αλγόριθμοι ML συχνά χωρίζουν τα δεδομένα σε μέρη, και στη συνέχεια, ενσωματώνουν αυτά τα μέρη για να πάρουν ένα αποτέλεσμα ή μια λύση. Τα συστήματα DL υιοθετούν μια ολοκληρωμένη προσέγγιση σε ένα πρόβλημα ή περίσταση. Τέλος, ως προς τα πεδία εφαρμογών, οι βασικές εφαρμογές ML περιλαμβάνουν ανιχνευτές ανεπιθύμητων μηνυμάτων ηλεκτρονικού ταχυδρομείου, αλγόριθμους που δημιουργούν τεκμηριωμένα σχέδια θεραπείας για ασθενείς και προγράμματα πρόβλεψης (όπως για την πρόβλεψη των διακυμάνσεων των τιμών του χρηματιστηρίου ή της τοποθεσίας και του χρόνου του επόμενου τυφώνα). Η DL έχει πολλαπλές χρήσεις. Τα αυτοκίνητα που οδηγούν μόνα τους χρησιμοποιούν βαθιά μάθηση για να κάνουν πράγματα όπως να αποφεύγουν εμπόδια και να αναγνωρίζουν τα φανάρια. Μπορεί επίσης να παρατηρηθεί σε υπηρεσίες όπως το Netflix και οι υπηρεσίες ροής μουσικής. Η DL χρησιμοποιείται επίσης για την αναγνώριση προσώπου.

3.3 Τι είναι το Νευρωνικό Δίκτυο

Τα ΝΝ, όπως εξυπονοεί και το όνομα, έχουν ως βάση τη λειτουργία ανθρώπινου νευρικού συστήματος. Αυτή η μέθοδος εργάζεται παραπλήσια με μια αλληλουχία νευρώνων, στο ανθρώπινο σύστημα, που λαμβάνουν τις πληροφορίες και κατόπιν τις επεξεργάζονται. Ένα ΝΝ μεταφράζει αριθμητικά μοτίβα που ενδέχεται να υφίστανται υπό τη μορφή διανύσματος/ων. Αυτοί οι φορείς, με τη αρωγή των ΝΝ, μεταγλωττίζονται. Το κύριο έργο που εκπληρώνει ένα ΝΝ είναι με βάση την ομοιότητα, ταξινόμηση και ομαδοποίηση των στοιχείων. Το

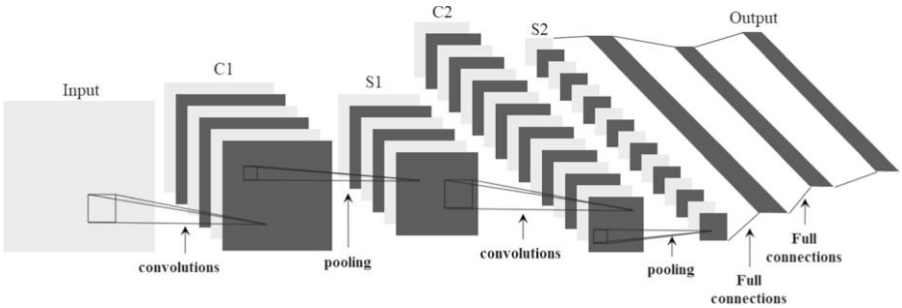
σημαντικότερο πλεονέκτημα για ένα NN είναι πως δεν είναι απαραίτητο να τροποποιείται κάθε φορά με βάση την παρεχόμενη είσοδο, καθώς προσαρμόζεται ευχερώς στο εναλλασσόμενο μοτίβο της εξόδου.

Η βασική διαφορά της DL και των NN βρίσκεται στο ότι, η πρώτη, ορίζεται ως DNN αποτελούμενο από αρκετά διαφορετικά στρωματά και κάθε στρώμα εμπερικλείει πολλούς κόμβους, μεταξύ τους διαφορετικούς. Ένα NN ενισχύει την εκτέλεση εργασίας με λιγότερη ακρίβεια, ενώ στη DL, εξαιτίας της εμφάνισης πολλαπλών στρωμάτων, η αποστολή περατώνεται με τελεσφόρηση. Ένα NN προϋποθέτει πιο λίγο χρόνο για να εκπαιδευτεί, μιας και είναι λιγότερο πολύπλοκο, ενώ για να εκπαιδευτεί ένα δίκτυο DL ενδέχεται να χρειαστεί πολύς χρόνος

Ευρίσκεται σημαντική ομοιότητα ανάμεσα στη DL και τα NN, επομένως γίνεται δύσκολο να δημιουργηθεί διάκριση ανάμεσα τους. Από τη μία, τα NN αποπερατώνουν τα καθήκοντά τους με τη επικουρία των νευρώνων, ενώ η DL υπολογίζει στην εξέταση ενός συνόλου δεδομένων και, με βάση το ίδιο, την εξαγωγή συμπερασμάτων.

3.4 Συνελικτικά Νευρωνικά Δίκτυα

Η έμπνευση των NN συνέλιξης προέρχεται από τη φυσιολογία του οπτικού φλοιού της γάτας. Υπάρχουν κύτταρα νευρώνων που είναι εξαιρετικά ευαίσθητα στην εξωτερική είσοδο σε μια συγκεκριμένη περιοχή του οπτικού φλοιού της γάτας και η περιοχή αυτή ονομάζεται δεκτικό πεδίο [73]. Από τότε, ο οπτικός φλοιός άρχισε να εισέρχεται στο πεδίο όρασης (FOV) και προσέλκυσε την προσοχή. Το συνελικτικό νευρωνικό δίκτυο CNN εφαρμόστηκε αρχικά στη χειρόγραφη αναγνώριση ψηφίων. Το μοντέλο δεδομένων [74] εκπαιδεύτηκε με CNN και πέτυχε καλά αποτελέσματα στο πείραμα, το οποίο δημιούργησε προηγούμενο για το συνελικτικό νευρωνικό δίκτυο ώστε αυτό να χρησιμοποιηθεί ευρέως στην αναγνώριση εικόνας, ομιλίας, προσώπου και σε άλλους τομείς. Το νευρωνικό δίκτυο συνέλιξης βελτιώνεται και προωθείται βασισμένο στο τεχνητό νευρωνικό δίκτυο. Το συνελικτικό νευρωνικό δίκτυο αποτελείται από πέντε μέρη: Επίπεδο πλήρους σύνδεσης, επίπεδο συνέλιξης, επίπεδο εισόδου, επίπεδο εξόδου και επίπεδο συγκέντρωσης. Στην τεχνητή νοημοσύνη όπως στην επεξεργασία εικόνας και βίντεο, η δομή της αποτελείται από πολλά μοντέλα. Το Le-Net 5 είναι το πιο δημοφιλές και πρακτικό μοντέλο και εισήχθη από τον Yann LeCun το 1998. Στην περίπτωση της βασικής αρχιτεκτονικής του LeNet-5 που φαίνεται στο Σχήμα 3, οι αρχιτεκτονικές των CNNs αποτελούνται από τέσσερα μέρη: το στρώμα εισόδου, το στρώμα συνέλιξης, το στρώμα συγκέντρωσης, το επίπεδο FC και το επίπεδο εξόδου. Τα δίκτυα των CNNs διαδραματίζουν σημαντικό ρόλο στο σχεδιασμό αρχιτεκτονικών νευρωνικών δικτύων, καθώς μια πιο λογική αρχιτεκτονική δικτύου μπορεί να ενισχύσει την προσαρμογή μεταξύ των επιπέδων ή μείωση των περιττών υπολογισμών στο δίκτυο, οι οποίοι συνήθως μπορεί να αποφέρουν ανώτερη απόδοση.



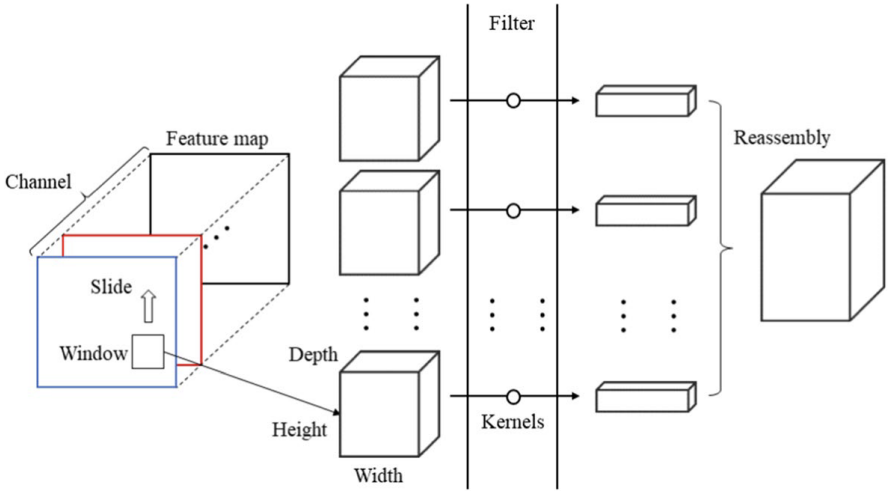
Σχήμα 3. Βασική αρχιτεκτονική LeNet-5 [75]

Τα CNN είναι δίκτυα τροφοδοσίας και η χωρική σχέση μεταξύ των pixel διατηρείται. Εδώ οι πληροφορίες δρομολογούνται από την είσοδο x, μέσω των ενδιάμεσων στρωμάτων στην έξοδο y το οποίο προσεγγίζεται από την συνάρτηση f*. Η συνάρτηση χαρτογράφησης δίνεται ως y = f*(x; θ), όπου η παράμετρος θ οδηγεί σε ακριβέστερη προσέγγιση της συνάρτησης. Έτσι, η είσοδος στο δίκτυο ρέει μέσω διαφόρων επιπέδων συνέλιξης που μαθαίνουν τα χαρακτηριστικά της εικόνας και συνοψίζουν τα χαρακτηριστικά αυτά όπως φαίνεται στο Σχήμα 4.

Σχήμα 4. Συνελικτικά Νευρωνικά Δίκτυα [76]

3.4.1 Στρώμα συνέλιξης

Το επίπεδο συνέλιξης περιλαμβάνει ένα σύνολο πυρήνων συνέλιξης, των οποίων το διάγραμμα εργασίας εμφανίζεται στο Σχήμα 5. Η διαδικασία στοχεύει να σύρει ένα προκαθορισμένο, σταθερού μεγέθους παράθυρο, στο χάρτη χαρακτηριστικών, να εξαγάγει παρακείμενα πλακίδια χαρακτηριστικών σε διάφορες θέσεις, (βήμα προς βήμα) και να εκτελέσει tensor προϊόντα κάθε πλακιδίου χαρακτηριστικών με τον εκπαιδευμένο πυρήνα βάρους συνέλιξης της μήτρας, ακολουθούμενο από αναδιοργάνωση του ληφθέντος διανυσματικού χώρου για την απόκτηση ενός νέου τανυστή (tensor).



Σχήμα 5 Διάγραμμα εργασίας συνελικτικών νευρωνικών δικτύων [75].

Ο μαθηματικός τύπος μεταξύ της εισόδου και της εξόδου της πράξης συνέλιξης είναι:

$$x^l{}_j = f^l(u^l{}_j) \qquad (1)$$

$$u^l{}_j = \sum_{i \in M_j} x_j{}^{l-1} \cdot k^l{}_{ij} + b^l{}_j \qquad (2)$$

Όπου το $x^l{}_j$ αντιπροσωπεύει την έξοδο του jth καναλιού στο επίπεδο συνέλιξης l, το fl συμβολίζει την λειτουργία ενεργοποίησης το στρώματος συνέλιξης, το $u^l{}_j$ αντιπροσωπεύει την ενεργοποίηση του δικτυου του jth καναλιού στο επίπεδο συνέλιξης l, το Mj αντιπροσωπεύει το υποσύνολο του χάρτη χαρακτηριστικών που λαμβάνεται ως δείγμα από το συρόμενο παράθυρο (sliding window) στο χάρτη χαρακτηριστικών, και $k^l{}_{ij}$, $b^l{}_j$ αντίστοιχα υποδηλώνουν τον πυρήνα συνέλιξης του επίπεδου συνέλιξης l και το σφάλμα του χάρτη χαρακτηριστικών. Πιο περιγραφικά, η λειτουργία του στρώματος συνέλιξης είναι η εξαγωγή των χαρακτηριστικών των δεδομένων εισόδου. Η εισαγωγή του στρώματος συνέλιξης μειώνει τον αριθμό των βαρών που χρειάζεται το παραδοσιακό δίκτυο για να μάθει και να ανακουφίσει το βάρος της μνήμης και του υπολογισμού.

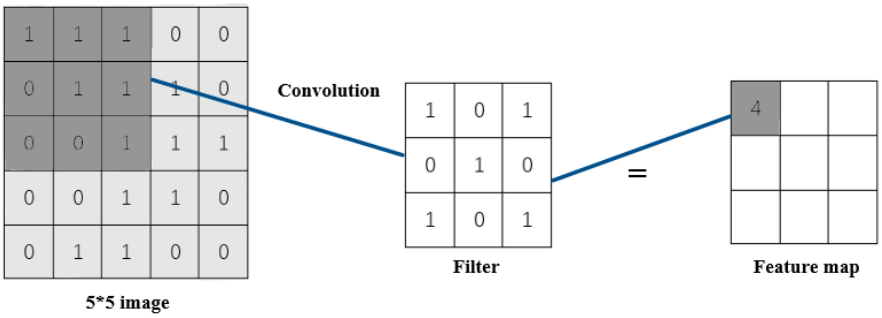
Προκειμένου να εμπλουτιστούν τα χαρακτηριστικά της εξαγόμενης εικόνας, κάθε επίπεδο συνέλιξης περιέχει πολλαπλούς πυρήνες συνέλιξης. Η διαδικασία εκπαίδευσης παραμέτρων συνέλιξης είναι παρόμοια με αυτή του παραδοσιακού νευρωνικού δικτύου. Πρώτον, η τιμή εξόδου του δικτύου λαμβάνεται με υπολογισμό forward propagation, τότε η τιμή της διαφοράς λαμβάνεται συγκρίνοντας την με την πραγματική τιμή. Το σφάλμα κάθε κόμβου σε κάθε στρώμα υπολογίζεται με backpropagation, και το βάρος της μάθησης βελτιστοποιείται σύμφωνα με το σφάλμα.

Τα συνελικτικά στρώματα μειώνουν τον αριθμό των παραμέτρων κυρίως με δύο τρόπους: Ο ένας είναι να εισαχθεί η ιδέα του πεδίου τοπικής αντίληψης, όπου υπάρχει συσχέτιση μεταξύ tsr από την τοπική εικόνα προς την καθολική. Ωστόσο, η συσχέτιση μεταξύ των pixels στο χωρικό παρακείμενο πεδίο είναι αρκετά στενή, ενώ η συσχέτιση μεταξύ των pixels που βρίσκονται μακριά είναι σχετικά αδύναμη. Από αυτή την άποψη, κάθε νευρώνας σε κάθε στρώμα πρέπει να αντιλαμβάνεται μόνο τα χαρακτηριστικά του δικού του στρώματος, αντιλαμβανόμενος μόνο την τοπική εικόνα. Τέλος, οι καθολικές πληροφορίες της εικόνας λαμβάνονται με την ενσωμάτωση των τοπικών πληροφοριών που λαμβάνονται στο επόμενο στρώμα. Μετά από αυτή τη λειτουργία, οι παράμετροι εκπαίδευσης μπορούν να μειωθούν. Ωστόσο, εάν ο όγκος των δεδομένων εκπαίδευσης είναι πολύ μεγάλος, θα υπάρξουν πολλά προβλήματα σε αυτήν τη λειτουργία. Εάν κάθε νευρώνας αντιστοιχεί μόνο σε ένα μέρος της εικόνας, το εξαγόμενο περιεχόμενο αυτού του νευρώνα δεν μπορεί να εφαρμοστεί σε άλλους νευρώνες. Στην περίπτωση αυτή, λαμβάνεται η μέθοδος επιμερισμού του βάρους. Διατηρώντας τις παραμέτρους κάθε νευρώνα σταθερούς, η ακρίβεια του μοντέλου δεν θα επηρεαστεί από την αλλαγή της θέσης του αντικειμένου. Η ανεπιφύλακτη αρχή είναι ότι τα τοπικά στατιστικά χαρακτηριστικά της εικόνας είναι συνεπή με χαρακτηριστικά από άλλα μέρη, και έτσι εφαρμόζουμε τα τοπικά χαρακτηριστικά μάθησης σε άλλα μέρη.

3.4.1.1 Σημαντικές παράμετροι

Οι πιο σημαντικές παράμετροι των συνελικτικών στρωμάτων είναι το μέγεθος της εικόνας εισόδου, το μέγεθος βήματος συνέλιξης, το μέγεθος του πυρήνα συνέλιξης, το μέγεθος πλήρωσης και ούτω καθεξής. Η συγκεκριμένη διαδικασία λειτουργίας της συνέλιξης παρουσιάζεται στο Σχήμα 6. Η εικόνα 5×5 στο σχήμα αντιπροσωπεύει τον χάρτη χαρακτηριστικών εισόδου, ο οποίος συμβολίζεται με το γράμμα Β. Το γκρι τμήμα (3×3) αντιπροσωπεύει τον πυρήνα συνέλιξης, ο οποίος σημειώνεται με το γράμμα Κ. Το μήκος κάθε κίνησης του πυρήνα συνέλιξης καθορίζεται ως το μήκος βήματος, το οποίο εκφράζεται με το γράμμα s. Το μέγεθος βήματος του πυρήνα συνέλιξης που κινείται οριζόντια είναι το ίδιο με

αυτό του πυρήνα συνέλιξης που κινείται κάθετα, όπου το μήκος βήματος είναι 1. Η πλήρωση των άκρων δεν ενδείκνυται. Η προεπιλεγμένη τιμή πλήρωσης είναι γενικά 0, το γράμμα q υποδεικνύει το μέγεθος της πλήρωσης των άκρων. Σε γενικές γραμμές, η έννοια του στρώματος συνέλιξης προτείνεται επειδή το συνελικτικό στρώμα έχει τα χαρακτηριστικά της αυτόματης εξαγωγής εικόνας και ο αριθμός των παραμέτρων στο δίκτυο μειώνεται, και αυτά μας βοηθούν να λύσουμε το δύσκολο πρόβλημα της εξαγωγής χαρακτηριστικών και των τεράστιων παραμέτρων εκπαίδευσης.



Σχήμα 6. Λειτουργία CNN [77]

3.4.2 Στρώμα συγκέντρωσης

Το στρώμα υπο-δειγματοληψίας (subsampling layer), γνωστό και ως στρώμα συγκέντρωσης (pooling layer), βρίσκεται συνήθως πίσω από το συνελικτικό στρώμα. Η κύρια λειτουργία του επιπέδου συγκέντρωσης είναι να μειώσει τις άχρηστες πληροφορίες χαρακτηριστικών, να διατηρήσει τις αποτελεσματικές πληροφορίες, να μειώσει τη διάσταση των χαρακτηριστικών και να επιτύχει το σκοπό της συμπίεσης του αριθμού των δεδομένων και των παραμέτρων. Από τη μία πλευρά, η μείωση του μεγέθους του χάρτη χαρακτηριστικών είναι σε θέση να ανακουφίσει την υπολογιστική πολυπλοκότητα. Αν και η ποσότητα του υπολογισμού μειώνεται μετά την προηγούμενη λειτουργία συνέλιξης, εξακολουθεί να αντιμετωπίζει την πρόκληση της υπολογιστικής πολυπλοκότητας λόγω της πολυπλοκότητας των δεδομένων εκπαίδευσης.

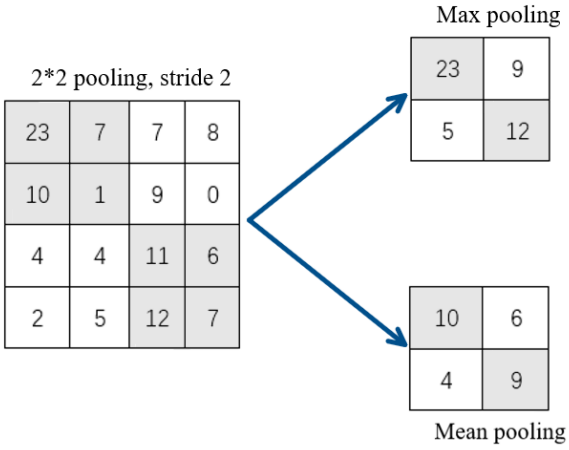
Από την άλλη, η περαιτέρω μείωση του αριθμού των παραμέτρων με τη συγκέντρωση μπορεί να αποτρέψει την υπερβολική προσαρμογή (over fitting). Το στρώμα συγκέντρωσης περιλαμβάνει γενικά το maxpooling, mean-pooling και stochastic pooling. Η μέγιστη συγκέντρωση (maxpooling) λαμβάνει τη μέγιστη τιμή των σημείων των χαρακτηριστικών στη γειτονιά, η μέση συγκέντρωση (mean-pooling) λαμβάνει τη μέση τιμή, ενώ η στοχαστική συγκέντρωση (stochastic pooling) παίρνει την τιμή τυχαίων σημείων στην γειτονιά. Ο μαθηματικός τύπος μεταξύ της εισόδου και της εξόδου της λειτουργίας συγκέντρωσης είναι:

$$x^l{}_j = f^l\left(u^l{}_j\right)$$

$$u_j{}^l = a_j{}^l \cdot f^l{}_{pooling}\left(x_j{}^{l-1}\right) + b_j{}^l$$

όπου $x^l{}_j$ι και $u_j{}^l$αντίστοιχα αντιπροσωπεύουν την έξοδο και την ενεργοποίηση δικτύου του καναλιού jth στο στρώμα συγκέντρωσης l, το f^l αντιπροσωπεύει τη λειτουργία ενεργοποίησης που συνδέεται με το επίπεδο συγκέντρωσης l, $a_j{}^l$ δηλώνει το συντελεστή βάρους του στρώματος συγκέντρωσης l και το $f^l{}_{pooling}$ αντιπροσωπεύει τη λειτουργία συγκέντρωσης του στρώματος συγκέντρωσης l. Το στρώμα συγκέντρωσης μειώνει τον αριθμό των στοιχείων που

υποβάλλονται σε επεξεργασία στο χάρτη χαρακτηριστικών μέσω υποδειγματοληψίας, και εισάγει μια ιεραρχική χωρική δομή φίλτρου αυξάνοντας σταδιακά το παράθυρο συνεχούς στρώματος συνέλιξης. Το επίπεδο συνέλιξης μπορεί να μειώσει σημαντικά τις παραμέτρους του δικτύου και να αποτρέψει την υπερπροσαρμογή του. Η συγκεκριμένη πράξη παρουσιάζεται στο Σχήμα 7. Η επάνω πλευρά δείχνει τη μέγιστη συγκέντρωση, και μόνο ένα από τα βάρη στον πυρήνα συνέλιξης είναι 1, τα άλλα βάρη έχουν οριστεί σε 0. Το αποτέλεσμα της μέγιστης συγκέντρωσης είναι η μείωση του μεγέθους των εικόνων στο ένα τέταρτο του αρχικού τους μεγέθους, στο οποίο διατηρείται η μέγιστη τιμή κάθε περιοχής 2×2. Η κάτω πλευρά αντιπροσωπεύει τη μέση συγκέντρωση, τα βάρη που ορίζονται στον πυρήνα συνέλιξης είναι όλα 0,25. Η μέση συγκέντρωση του πυρήνα συνέλιξης σκοπό έχει να να αποδυναμώσει το θόλωμα της αρχικής εικόνας στο 1/4 του αρχικού.



Σχήμα 7. Λειτουργία Pooling [77]

3.4.3 Συνάρτηση ενεργοποίησης

Είναι ένας μη-γραμμικός παράγοντας για την επίλυση του προβλήματος της ανεπαρκούς ικανότητας έκφρασης του γραμμικού μοντέλου [78]. Στη δομή του νευρωνικού δικτύου, πρέπει να προστεθεί μια συνάρτηση ενεργοποίησης μετά την υπέρθεση κάθε στρώματος. Οι συναρτήσεις ενεργοποίησης διαδραματίζουν ρόλο λήψης αποφάσεων στα CNN, τα οποία ευνοούν την εκμάθηση μη γραμμικών σύνθετων μοτίβων. Εφαρμόζονται κυρίως για την εισαγωγή μη γραμμικών παραγόντων στα νευρωνικά δίκτυα, για την ενίσχυση της ικανότητας προσαρμογής τους. Ο Πίνακας 1 περιγράφει τους τυπικούς τύπους συναρτήσεων ενεργοποίησης σε CNN, συμπεριλαμβανομένων των σιγμοειδών, tanh, διορθωμένης γραμμικής μονάδας (ReLU) και οι παραλλαγές της Exponential Linear Unit [79], Leaky ReLU[80], Randomized ReLU[81], Parametric ReLU [82], Switch και Maxout.

Function	Definition	Parameter of α
Sigmoid	$f(x) = \frac{1}{1+e^{-x}}$	–
Tanh	$f(x) = \frac{2}{1+e^{-2x}} - 1$	–
ReLU	$f(x) = \begin{cases} x & x \geqslant 0 \\ 0 & x < 0 \end{cases}$	–
LeakyReLU	$f(x) = \begin{cases} x & x \geqslant 0 \\ \alpha x & x < 0 \end{cases}$	$\alpha \in (0, 1)$
PReLU	$f(x) = \begin{cases} x & x \geqslant 0 \\ \alpha x & x < 0 \end{cases}$	Learned parameter
RReLU	$f(x) = \begin{cases} x & x \geqslant 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$	Uniform(a, b)
ELU	$f(x) = \begin{cases} x & x \geqslant 0 \\ x & x < 0 \end{cases}$	Predefined parameter
Switch	$f(x) = x.sigmoid(\alpha x)$	Learned parameter
Maxout	$f(x) = \max_{j \in [1,k]} x^T W + b_{ij}\, W \in R^{d \times m \times k}$	–

Πίνακας 1. Σύνοψη συνάρτησης ενεργοποίησης [75]

Προφανώς, σε σύγκριση με τις συναρτήσεις όπως το σιγμοειδές και το tanh, οι υπερβολικά μεγάλες και μικρές είσοδοι δεν κάνουν το ReLU να τείνει να κορεστεί. Έτσι, το ReLU και οι παραλλαγές του είναι ανώτερες από τη συμβατική ενεργοποίηση συναρτήσεων όπως το σιγμοειδές και το tanh για την αντιμετώπιση του προβλήματος κλίσης που εξαφανίζεται[83]. Όσον αφορά τη συνάρτηση ενεργοποίησης Maxout, αυτό αποφεύγει την ακύρωση του νευρώνα και άλλα προβλήματα με βάση τη διατήρηση της γραμμικότητας και των πλεονεκτημάτων ακορεστότητας της συνάρτησης ReLU. Η έκφραση της σιγμοειδούς συνάρτησης παρουσιάζεται στην εξίσωση:

$$h(a) = \frac{1}{1 - e^{-a}}$$

Από την εικόνα της σιγμοειδούς συνάρτησης, βλέπουμε ότι η συνάρτηση χρησιμοποιείται για να μετατρέψει την τιμή εισόδου στο εύρος από 0 έως 1,0. Η λειτουργία είναι συνεχής, μονοτονική και λειτουργεί εύκολα. Η σιγμοειδής συνάρτηση εφαρμοζόταν ευρέως στο παρελθόν, αλλά έχει δύο ελαττώματα. Πρώτον, η τιμή εξόδου της σιγμοειδούς συνάρτησης δεν είναι συμμετρική με το 0, γεγονός που καθιστά τον υπολογισμό πολύ περίπλοκο. Δεύτερον, η λειτουργία έχει μαλακό κορεσμό. Εάν η τιμή εισόδου είναι πολύ μεγάλη ή πολύ μικρή, η κλίση βάρους θα αλλάξει αργά στο μηδέν, δηλαδή η κλίση θα εξαφανιστεί. Η υπερβολική εφαπτομένη συνάρτηση εξελίσσεται από τη σιγμοειδή συνάρτηση, η οποία εκφράζεται στην παρακάτω εξίσωση.

$$h(a) = \frac{e^{a} - e^{-a}}{e^{a} + e^{-1}}$$

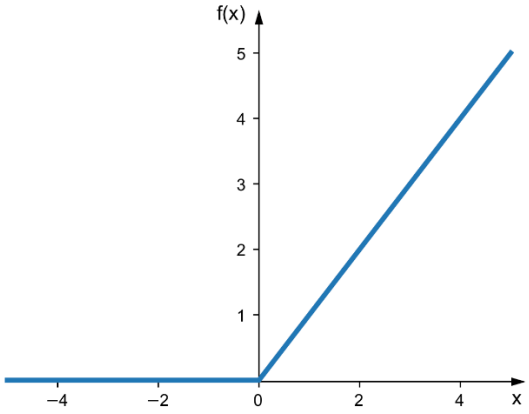
Η συνάρτηση Tanh είναι πιο δημοφιλής από τη σιγμοειδή συνάρτηση στην πράξη. Σε σύγκριση με τη σιγμοειδή συνάρτηση, η έξοδος της έχει συμμετρία με μηδενικό κέντρο, η παράγωγος της σιγμοειδούς συνάρτησης είναι πολύ μικρότερη από την παράγωγο της συνάρτησης tanh, κάτι που δείχνει ότι η συνάρτηση tanh συγκλίνει γρηγορότερα στο στρώμα συνέλιξης. Η προφανής

αλλαγή κλίσης συμβαίνει κατά τη διάρκεια της εκπαιδευτικής διαδικασίας, αλλά το πρόβλημα εξαφάνισης της κλίσης λόγω της συνάρτησης ενεργοποίησης όπως η σιγμοειδής συνάρτηση και η συνάρτηση tanh εξακολουθούν να υπάρχουν.

Οι γραμμικές διορθωμένες μονάδες (ReLU) είναι οι πιο σημαντικές συναρτήσεις ενεργοποίησης τα τελευταία χρόνια, η έκφρασή τους εμφανίζεται στην παρακάτω εξίσωση:

$$f(x) = \begin{cases} 0, x & \leq 0 \\ x, x & > 0 \end{cases} = max(0, x)$$

Η εικόνα της συνάρτησης ReLU φαίνεται στο Σχήμα 8, βλέπουμε ότι η διαδικασία υπολογισμού της συνάρτησης ReLU είναι απλούστερη από αυτή της συνάρτησης tanh και της σιγμοειδούς συνάρτησης. Το ReLU είναι ιδιαίτερα γρήγορο επειδή χρειάζεται μόνο να καθορίσει εάν το αποτέλεσμα είναι μεγαλύτερο από 0. Επιπλέον, σε σύγκριση με τις δύο πρώτες συναρτήσεις ενεργοποίησης, η ταχύτητα σύγκλισης είναι προφανώς ταχύτερη. Επιπλέον, το πρόβλημα της εξαφάνισης της κλίσης μετριάζεται αποτελεσματικά ενώ μειώνεται το μέγεθος του υπολογισμού.



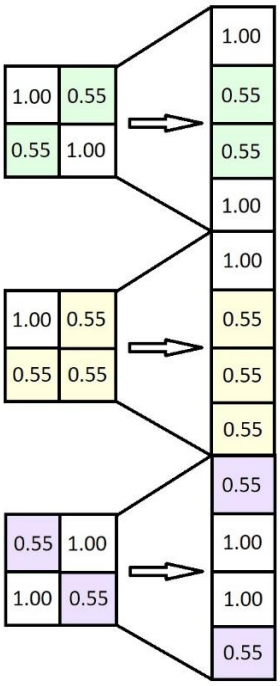
Σχήμα 8. Συνάρτηση ReLu

Συγκρίνοντας τις παραπάνω τρεις συναρτήσεις ενεργοποίησης, βλέπουμε ότι η έξοδος της σιγμοειδούς συνάρτησης είναι ένας θετικός πραγματικός αριθμός. Η έξοδος της συνάρτησης tanh είναι ένας θετικός αριθμός ή ένας αρνητικός αριθμός, ενώ η είσοδος της συνάρτησης ReLU είναι μόνο ένας αριθμός μεγαλύτερος από 0.

3.4.4 Πλήρως συνδεδεμένο στρώμα

Το πλήρως συνδεδεμένο στρώμα (FC) βρίσκεται γενικά στο τελευταίο μέρος ενός CNN, το οποίο υιοθετήθηκε για τη μετατροπή πληροφοριών των δισδιάστατων χαρακτηριστικών εξόδου (2D) που εξάγονται από το προηγούμενο στρώμα στη μονοδιάστατη (1D) πληροφορία ταξινόμησης. Μοιάζει με το κρυφό στρώμα πολυστρωματικών perceptrons (MLPs), όπου η έξοδος λαμβάνεται με συνδυασμό βαρών από νευρώνες FC στο προηγούμενο στρώμα.

Δηλαδή, όλοι οι νευρώνες στο CNN, είναι απαραίτητο να συνδεθούν με όλους τους νευρώνες στο ανώτερο στρώμα[14], μετατρέποντας τελικά τα εξαγόμενα χαρακτηριστικά σε n × 1. Το σχηματικό διάγραμμα του στρώματος σύνδεσης φαίνεται στην Σχήμα 9.



Σχήμα 9. Διάγραμμα στρώματος FC

Ο μαθηματικός τύπος μεταξύ της λειτουργίας εισόδου και εξόδου κάθε νευρώνα είναι:

$$x_j{}^l = f^l(u_j{}^l)$$

$$u_j{}^l = max\left(0, \sum_i y_i{}^l \cdot w_{ij}{}^l + b_j{}^l\right)$$

όπου τα $x_j{}^l$ και $u_j{}^l$ αντιπροσωπεύουν αντίστοιχα την έξοδο και την ενεργοποίηση δικτύου του καναλιού jth στο στρώμα FC l, ενώ το f^l αντιπροσωπεύει τη συνάρτηση ενεργοποίησης που συνδέεται με το στρώμα συγκέντρωσης l, $w_{ij}{}^l$ και $b_j{}^l$ είναι αντίστοιχα ο συντελεστής βάρους και το σφάλμα του στρώματος FC l, και το max υποδηλώνει το μεγαλύτερο αριθμό στην επιλεγμένη περιοχή του πίνακα.

3.4.5 Εργαλείο βελτιστοποίησης

Στη διαδικασία backpropagation στη βαθιά μάθηση, ο βελτιστοποιητής προσαρμόζει συνεχώς τις παραμέτρους της συνάρτησης απώλειας για να κάνει την είσοδο καλύτερα προσαρμόσιμη για την έξοδο και να διασφαλίσει ότι η συνάρτηση απώλειας προσεγγίζει το συνολικό ελάχιστο. Στην ουσία, είναι το πρόβλημα της εύρεσης της βέλτιστης λύσης των συναρτήσεων. Η βασική ιδέα του τρέχοντος βελτιστοποιητή είναι να χρησιμοποιήσει τη μέθοδο καθόδου κλίσης για τη βελτιστοποίηση των παραμέτρων. Η ταχύτερη συγκλίνουσα κατεύθυνση της τιμής της συνάρτησης είναι η κατεύθυνση κλίσης. Όταν το μοντέλο βαθιάς μάθησης έχει ρυθμιστεί να επιλύει την ελάχιστη τιμή της αντικειμενικής συνάρτησης, η κατεύθυνση κινείται κατά μήκος της φθίνουσας κατεύθυνσης κλίσης, και τα αποτελέσματα προσεγγίζουν συνεχώς τη βέλτιστη τιμή. Οι δύο πιο σημαντικές παράμετροι του βελτιστοποιητή είναι η κατεύθυνση βελτιστοποίησης και το μέγεθος βήματος. Η κατεύθυνση βελτιστοποίησης αντικατοπτρίζεται ως η κλίση ή το μομέντουμ στον βελτιστοποιητή, το μέγεθος του βήματος αντικατοπτρίζεται

ως ο ρυθμός εκμάθησης στον βελτιστοποιητή. Οι δημοφιλείς βελτιστοποιητές είναι η στοχαστική κάθοδος κλίσης (SGD), ο αλγόριθμος AdaDelta και η προσαρμοστική στιγμιαία εκτίμηση (Adam), η Mini Batch Gradient Descent (MBGD), ο προσαρμοστικός αλγόριθμος κλίσης (AdaGrad), η Batch Gradient Descent (BGD) κ.λπ.

Η διαδικασία βελτιστοποίησης των παραμέτρων χωρίζεται σε τέσσερα βήματα: Πρώτον, η αντικειμενική συνάρτηση, αντιστοιχεί στην κλίση των τρεχουσών παραμέτρων. Στη συνέχεια το διαφορετικό μομέντουμ υπολογίζεται σύμφωνα με την κλίση της παραμέτρου. Τρίτον, υπολογίζεται η κλίση καθόδου την τρέχουσα χρονική στιγμή. Τελικά οι παράμετροι ενημερώνονται σύμφωνα με την κλίση καθόδου. Στα δύο τελευταία βήματα, ο αλγόριθμος είναι συναφής, διαφορές υπάρχουν μόνο στα δύο πρώτα βήματα. Ο αλγόριθμος καθόδου κλίσης επιλέγει ένα δείγμα τυχαία από το σύνολο προπόνησης και δεν περιλαμβάνει μομέντουμ. Το πλεονέκτημα του αλγορίθμου είναι ότι μαθαίνει μόνο ένα δείγμα κάθε φορά και έτσι η ταχύτητα προπόνησης είναι γρήγορη. Το εργαλείο βελτιστοποίησης ικανοποιεί το τοπικό βέλτιστο σημείο ή το σημείο ράχης-saddle εάν η κλίση είναι 0, γεγονός που καθιστά αδύνατη την ενημέρωση των παραμέτρων. Μια πολύ κλασσική μέθοδος σε πειράματα καθόδου-κλίσης δεσμίδας, είναι αυτή που χρησιμοποιεί όλα τα δείγματα σε κάθε ενημέρωση. Παρέχει καλύτερη αναπαράσταση όλων των δεδομένων δειγμάτων και υποδεικνύει με μεγαλύτερη ακρίβεια την θέση των ακραίων σημείων. Ωστόσο, αυτή η μέθοδος θα υπολογίζει ολόκληρο το σύνολο δεδομένων κάθε φορά ενώ ενημερώνεται, γεγονός που οδηγεί στην αύξηση του υπολογισμού και στην αδυναμία να προσθέσει νέα δείγματα. Η βελτιωμένη μέθοδος καθόδου κλίσης, μικρής δεσμίδας, βασίζεται στις μεθόδους BGD και SGD. Σε αυτή την περίπτωση επιλέγεται ένας μικρός αριθμός δειγμάτων από το σύνολο δεδομένων για τον υπολογισμό των απωλειών και την ενημέρωση των παραμέτρων δικτύου. Από τη μία πλευρά, μειώνει τη διακύμανση της ενημέρωσης παραμέτρων και καθιστά τη σύγκλιση σταθερή. Ταυτόχρονα, αξιοποιεί πλήρως τα δεδομένα για τον πιο αποτελεσματικό υπολογισμό κλίσης. Το μειονέκτημα είναι ότι δεν μπορεί να εγγυηθεί καλή σύγκλιση, η οποία είναι απαραίτητη για τον καθορισμό ενός κατάλληλου ποσοστού μάθησης. Εάν η ρύθμιση είναι πολύ μικρή, η ταχύτητα σύγκλισης είναι πολύ αργή, ενώ εάν η ρύθμιση είναι πολύ μεγάλη, είναι εύκολο να ταλαντωθεί γύρω από το ελάχιστο. Ο προσαρμοστικός αλγόριθμος μάθησης προσθέτει την έννοια του μομέντουμ δεύτερης τάξης. Ο αλγόριθμος SGD και ο αλγόριθμος παραλλαγής του, προσαρμόζουν κάθε παράμετρο για τον ίδιο ρυθμό μάθησης.

Το βαθύ νευρωνικό δίκτυο περιέχει συχνά μεγάλο αριθμό παραμέτρων οι οποίες δεν χρησιμοποιούνται όλες πάντοτε. Επομένως, μέσω του αλγορίθμου προσαρμοστικού ρυθμού εκμάθησης, ορίζουμε διαφορετικούς ρυθμούς μάθησης ανάλογα με τη συγκεκριμένη κατάσταση. Όσον αφορά τις παραμέτρους που ενημερώνονται συχνά, ο ρυθμός μάθησης είναι πιο αργός. Όσον αφορά τις περιστασιακά ενημερωμένες παραμέτρους, ο ρυθμός μάθησης θα πρέπει να είναι μεγαλύτερος. Ωστόσο, εάν επαναληφθεί πολλές φορές, ο ρυθμός μάθησης θα μειωθεί, το μομέντουμ δεύτερης τάξης θα συσσωρεύεται συνεχώς, γεγονός που θα οδηγήσει στην ολοκλήρωση της εκπαιδευτικής διαδικασίας νωρίτερα.

Η προσαρμογή του ρυθμού μάθησης στον αλγόριθμο AdaGrad είναι πολύ ριζοσπαστική, ο αλγόριθμος AdaDelta προτείνει την ιδέα όχι μόνο της συσσώρευσης ιστορικών διαβαθμίσεων αλλά και της εστίασης στην φθίνουσα κλίση του χρονικού παραθύρου στο παρελθόν. Ο αλγόριθμος Adam έχει ένα σωρό οφέλη που ελέγχει το μήκος του ρυθμού εκμάθησης και την

κατεύθυνση κλίσης στον αλγόριθμο, γεγονός που αποτρέπει τα προβλήματα όπως οι ταλαντώσεις κλίσης και η στασιμότητα του διάσελου.

3.5 R-CNN

Το R-CNN περιεγράφηκε το 2014 από τον Ross Girshick, et al.[7] στο UC Berkeley. Μπορεί να ήταν μια από τις πρώτες μεγάλες και επιτυχημένες εφαρμογές συνελικτικών νευρωνικών δικτύων στο πρόβλημα του εντοπισμού αντικειμένων, της ανίχνευσης και της τμηματοποίησης. Η προσέγγιση επιδείχθηκε σε Datasets αναφοράς, κατορθώνοντας εν συνεχεία σκορ κορυφαίας-τεχνολογίας στο «VOC-2012» και στο 200ων κλάσεων - ανίχνευσης αντικειμένων - «ILSVRC-2013». Το προτεινόμενο μοντέλο R-CNN αποτελείται από τρεις ενότητες:

- Ενότητα 1: Πρόταση Περιφέρειας (Region Proposal). Δημιουργεί και εξαγάγει προτάσεις κατηγοριών, ανεξάρτητων περιοχών, π.χ. υποψηφία πλαισίων οριοθέτησης.
- Ενότητα 2: Εξολκέας χαρακτηριστικών (Feature Extractor). Εξαγάγει χαρακτηριστικά από κάθε υποψήφια περιοχή, π.χ. χρησιμοποιώντας ένα βαθύ συνελικτικό νευρωνικό δίκτυο.
- Ενότητα 3: Ταξινομητής (Classifier). Ταξινομεί τα χαρακτηριστικά ως μία κλάση από τις γνωστές, π.χ. γραμμικό μοντέλο ταξινόμησης SVM.

Μια τεχνική υπολογιστικής όρασης (computer vision), που ονομάζεται "επιλεκτική αναζήτηση" («selective search»), χρησιμοποιείται για να προτείνει υποψήφιες περιοχές ή πλαίσια οριοθέτησης πιθανών αντικειμένων στην εικόνα, αν και η ευελιξία του σχεδιασμού επιτρέπει τη χρήση άλλων αλγορίθμων προτάσεων περιοχής. Ο εξολκέας χαρακτηριστικών που χρησιμοποιήθηκε από το μοντέλο ήταν το «AlexNet deep CNN» που κέρδισε τον διαγωνισμό ταξινόμησης εικόνας «ILSVRC-2012». Η έξοδος του CNN ήταν ένα διάνυσμα 4.096 στοιχείων που περιγράφει τα περιεχόμενα της εικόνας που τροφοδοτούνται σε ένα γραμμικό SVM για ταξινόμηση. Συγκεκριμένα ένα SVM εκπαιδεύεται για κάθε γνωστή κλάση. Είναι μια σχετικά απλή εφαρμογή των CNNs στο πρόβλημα του εντοπισμού και της αναγνώρισης αντικειμένων. Ένα μειονέκτημα της προσέγγισης είναι ότι είναι αργή, απαιτώντας την εξαγωγή χαρακτηριστικών που βασίζεται στο CNN να περνά στην κάθε μία από τις υποψήφιες περιοχές που δημιουργούνται από τον αλγόριθμο πρότασης περιοχής.

3.5.1 Fast R-CNN

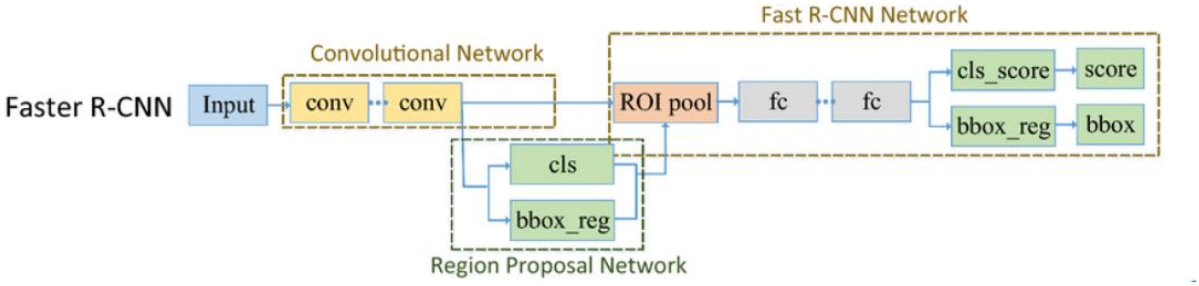
Δεδομένης της μεγάλης επιτυχίας του R-CNN, το 2015, ο Ross Girshick, τότε στη Microsoft Research, πρότεινε μια επέκταση για την αντιμετώπιση των ζητημάτων ταχύτητας του R-CNN. Ανασκοπικά, για τα R-CNN ισχύει ότι η εκπαίδευση είναι ένας αγωγός πολλαπλών σταδίων. Περιλαμβάνει την προετοιμασία και τη λειτουργία τριών ξεχωριστών μοντέλων. Η εκπαίδευση είναι δαπανηρή σε χώρο και χρόνο. Η εκπαίδευση ενός βαθιού CNN σε τόσες πολλές προτάσεις περιοχής ανά εικόνα είναι πολύ αργή. Ο εντοπισμός αντικειμένων είναι αργός. Το να κάνεις προβλέψεις χρησιμοποιώντας ένα βαθύ CNN για τόσες πολλές προτάσεις περιοχών είναι πολύ αργό.

Μια εργασία πρότεινε την επιτάχυνση της τεχνικής με την χρήση δικτυων SPPnets (χωρικά δίκτυα συγκέντρωσης πυραμίδων) το 2014. Αυτό επιτάχυνε την εξαγωγή χαρακτηριστικών, αλλά ουσιαστικά χρησιμοποίησε έναν τύπο αλγόριθμου «forward pass caching». Το Fast R-CNN προτείνεται ως ένα ενιαίο μοντέλο αντί για έναν αγωγό (pipeline) για την άμεση εκμάθηση και εξαγωγή περιοχών και ταξινομήσεων. Η αρχιτεκτονική του μοντέλου εισαγάγει μια

φωτογραφία ενός συνόλου προτάσεων περιοχής και την περνα μέσα από ένα βαθύ συνελικτικό νευρωνικό δίκτυο. Ένα προ-εκπαιδευμένο CNN, όπως ένα «VGG-16», χρησιμοποιείται για την εξαγωγή χαρακτηριστικών. Το τέλος του βαθιού CNN είναι ένα προσαρμοσμένο επίπεδο που ονομάζεται «Region of Interest Pooling Layer» («Επίπεδο συγκέντρωσης περιοχής ενδιαφέροντος») ή RoI Pooling, το οποίο εξάγει ειδικά χαρακτηριστικά για μια δεδομένη υποψήφια περιοχή εισόδου. Η έξοδος του CNN στη συνέχεια ερμηνεύεται από ένα πλήρως συνδεδεμένο στρώμα (fully connected layer) και στη συνέχεια το μοντέλο διακλαδίζεται σε δύο εξόδους, μία για την πρόβλεψη κλάσης μέσω ενός στρώματος softmax και μια άλλη με γραμμική έξοδο για το πλαίσιο οριοθέτησης (bounding box). Αυτό το μοτίβο ξανασυμβαίνει στη συνέχεια πολλές φορές για κάθε περιοχή ενδιαφέροντος σε μια δεδομένη εικόνα. Το μοντέλο είναι σημαντικά πιο γρήγορο στην εκπαίδευση και στην πραγματοποίηση προβλέψεων, αλλά εξακολουθεί να απαιτεί να προταθεί ένα σύνολο υποψήφιων περιοχών μαζί με κάθε εικόνα εισόδου.

3.5.2 Faster R-CNN

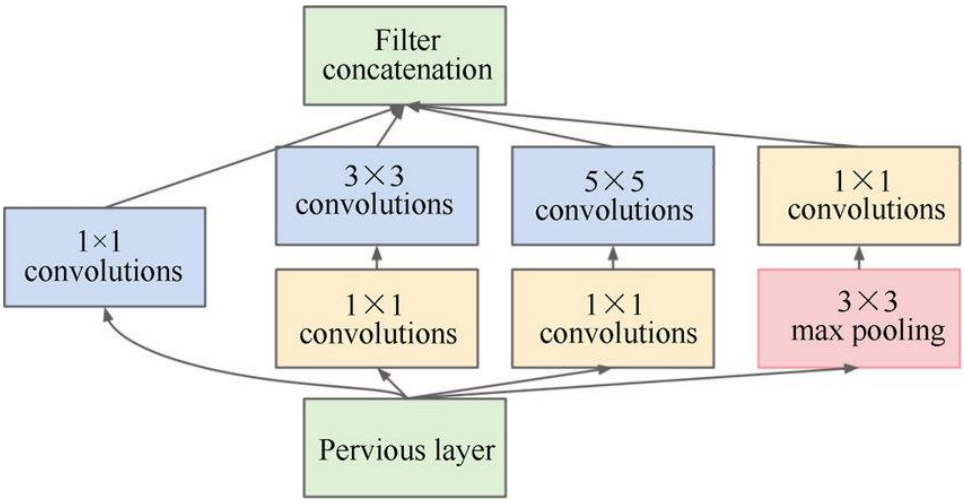
Το Faster R-CNN είναι μια βελτίωση του Fast R-CNN. Στο Faster R-CNN, αντί να χρησιμοποιηθεί αλγόριθμος επιλεκτικής αναζήτησης[84] για την εξαγωγή υποψηφίων πλαισίων, χρησιμοποιήθηκε το RPN[82] (συνιστώμενο δίκτυο περιοχής) για την τμηματοποίηση υποψήφιων περιοχών. Το δίκτυο RPN όχι μόνο βελτιώνει το ποσοστό ανάκλησης του δικτύου, αλλά και αυξάνει το ποσοστό αναγνώρισης και την ακρίβεια του δικτύου. Το Σχήμα 10 δείχνει τη δομή του Faster δικτύου R-CNN.



Σχήμα 10. Δομή του Faster R-CNN [85]

Η συγκεκριμένη διαδικασία ταυτοποίησης έχει ως εξής: Μετά την είσοδο της εικόνας στο δίκτυο CNN, λαμβάνεται το συνελικτικό στρώμα. Αυτό το συνελικτικό στρώμα έχει δύο σκοπούς: 1) Εισαγωγή στο δίκτυο RPN ως input, 2) εισαγωγή εικόνων σε ένα συγκεκριμένο σταθερό συνελικτικό στρώμα για να αποκτήσει έναν, πιο περίπλοκων διαστάσεων, χάρτη χαρακτηριστικών. Μέσω του δικτύου RPN, οι βαθμολογίες κάθε πρότασης περιοχής και η αντίστοιχη πρόταση περιοχής αποτελούν output, υπολογίζονται οι βαθμολογίες κάθε περιφερειακής σύστασης, στη βαθμολογία χρησιμοποιείται μη μέγιστη καταστολή (NMS) [86] (το όριο είναι 0,7) και η τελική έξοδος είναι μικρότερη από 300 υποψήφια πλαίσια με καλή ποιότητα. Τα υποψήφια πλαίσια συγκεντρώνονται στο ίδιο μέγεθος με το στρώμα συγκέντρωσης ROI. Η απόδοση επένδυσης (ROI) στο δεύτερο βήμα και ο χάρτης χαρακτηριστικών υψηλών διαστάσεων στο πρώτο βήμα, εισάγονται στο στρώμα συγκέντρωσης ROI. Στη συνέχεια, λαμβάνονται τα αντίστοιχα χαρακτηριστικά κάθε περιοχικής πρότασης. Τέλος, τα υποψήφια χαρακτηριστικά στο τρίτο βήμα εισάγονται στο πλήρως συνδεδεμένο στρώμα για να ληφθεί η βαθμολογία και η παλινδρόμηση των συνόρων κάθε υποψήφιας περιοχής. Η ταξινόμηση και η παλινδρόμηση των ορίων, εκπαιδεύονται από κοινού

χρησιμοποιώντας απώλεια softmax και ομαλή απώλεια L1. Το RPN είναι η χρήση παλινδρομήσεων πλαισίου οριοθέτησης και αγκυρών (δηλαδή, υποψήφιου κουτιού - box). Χρειάζεται μόνο να γλιστρήσει στο τελευταίο επίπεδο συνέλιξης για να πάρει υποψήφιες περιοχές πολλαπλών κλιμάκων. Το Σχήμα 11 δείχνει τη δομή δικτύου του RPN. Το βασικό δίκτυο είναι το ZF [87], η διάσταση εξόδου του στρώματος conv5 είναι 256, που αντιστοιχεί σε 256 χαρακτηριστικά γραφήματα. Κάθε συρόμενο παράθυρο προβλέπει δύο υποψήφιες περιοχές, το επίπεδο ταξινόμησης εξάγει δύο βαθμολογίες, δηλαδή πιθανότητα αντικειμένου. Το επίπεδο εξάγει 4 συντεταγμένες, δηλαδή τις συντεταγμένες του ορίου του αντικειμένου. Κάθε άγκυρα χρησιμοποιεί την τρέχουσα θέση ως προέλευση και επιλέγει συρόμενα παράθυρα διαφορετικών μεγεθών και κλιμάκων. Στο Σχήμα 11 χρησιμοποιούνται τρία μεγέθη και τρεις αναλογίες πλάτους μήκους, έτσι ώστε κάθε αντίστοιχο σημείο ολίσθησης να έχει εννέα υποψήφιες περιοχές. Με αυτόν τον τρόπο, οι υποψήφιες περιοχές μεταφράζονται αναλλοίωτες. Το Faster R-CNN συνδυάζει τον αρχικό διαχωρισμό της παλινδρόμησης και της ταξινόμησης των ορίων και χρησιμοποιεί τη μέθοδο από άκρο σε άκρο για την ανίχνευση του αντικειμένου, η οποία όχι μόνο βελτιώνει την ακρίβεια της ανίχνευσης και της αναγνώρισης, αλλά και ενισχύει την ταχύτητα αναγνώρισης. Το πρώτο μοντέλο Faster R-CNN συνδέεται με το μοντέλο VGG [88] και παίρνει αυτό το μοντέλο ως σημαντικό χαρακτηριστικό εξολκέα. Ωστόσο, το δίκτυο VGG ως η ραχοκοκαλιά του Faster R-CNN έχει μεγαλύτερη κλίμακα και χαμηλότερη ταχύτητα αναγνώρισης. Αυτές οι ελλείψεις δεν είναι κατάλληλες για TSR σε πραγματικό χρόνο, το οποίο απαιτεί ταχύτητα και ακρίβεια. Ως ταξινομητής, η απόδοση του GoogLeNet [80] δεν υστερεί από το μοντέλο VGG, του οποίου η κλίμακα δικτύου του είναι μικρή.



Σχήμα 11. Inception Model [88]

3.6 Yolo

Μια άλλη δημοφιλής οικογένεια μοντέλων αναγνώρισης αντικειμένων αναφέρεται συλλογικά ως YOLO ή "You Only Look Once", που αναπτύχθηκε από τον Joseph Redmon, et al. [89] Τα μοντέλα R-CNN μπορεί να είναι γενικά πιο ακριβή, ωστόσο η οικογένεια μοντέλων YOLO είναι γρήγορη, πολύ ταχύτερη από το R-CNN, επιτυγχάνοντας ανίχνευση αντικειμένων σε πραγματικό χρόνο. Η προσέγγιση περιλαμβάνει ένα ενιαίο νευρωνικό δίκτυο εκπαιδευμένο από άκρο σε άκρο που παίρνει μια φωτογραφία ως είσοδο και προβλέπει κουτιά οριοθέτησης και ετικέτες κλάσης για κάθε πλαίσιο οριοθέτησης απευθείας. Η τεχνική προσφέρει χαμηλότερη προγνωστική ακρίβεια (π.χ. περισσότερα σφάλματα εντοπισμού), αν και λειτουργεί με 45 καρέ ανά δευτερόλεπτο και έως 155 καρέ ανά δευτερόλεπτο σε μια βελτιστοποιημένης ταχύτητας

έκδοση του μοντέλου. Το μοντέλο λειτουργεί διαιρώντας πρώτα την εικόνα εισόδου σε ένα πλέγμα κελιών, όπου κάθε κελί είναι υπεύθυνο για την πρόβλεψη ενός πλαισίου οριοθέτησης εάν το κέντρο ενός πλαισίου οριοθέτησης εμπίπτει μέσα στο κελί. Κάθε κελί πλέγματος προβλέπει ένα πλαίσιο οριοθέτησης που περιλαμβάνει τη συντεταγμένη x, y και το πλάτος και το ύψος και την εμπιστοσύνη. Μια πρόβλεψη κλάσης βασίζεται επίσης σε κάθε κελί. Για παράδειγμα, μια εικόνα μπορεί να διαιρεθεί σε ένα πλέγμα 7×7 και κάθε κελί στο πλέγμα μπορεί να προβλέψει 2 πλαίσια οριοθέτησης, με αποτέλεσμα 94 προτεινόμενες προβλέψεις πλαισίων οριοθέτησης. Ο χάρτης πιθανοτήτων κλάσης και τα πλαίσια οριοθέτησης με σιγουριά συνδυάζονται στη συνέχεια σε ένα τελικό σύνολο πλαισίων οριοθέτησης και ετικετών κλάσης.

3.6.1 Αλγόριθμος

Ο πυρήνας του αλγορίθμου ανίχνευσης στόχου YOLO έγκειται στο μικρό μέγεθος και τη γρήγορη ταχύτητα υπολογισμού του μοντέλου. Η δομή του YOLO είναι απλή. Μπορεί να εξάγει απευθείας τη θέση και την κατηγορία του πλαισίου οριοθέτησης μέσω του νευρωνικού δικτύου. Η ταχύτητα του YOLO είναι γρήγορη επειδή το YOLO χρειάζεται να βάλει μόνο την εικόνα στο δίκτυο για να πάρει το τελικό αποτέλεσμα ανίχνευσης, οπότε το YOLO μπορεί επίσης να πραγματοποιήσει την ανίχνευση χρόνου βίντεο. Το YOLO χρησιμοποιεί απευθείας την καθολική εικόνα για ανίχνευση, η οποία μπορεί να κωδικοποιήσει τις καθολικές πληροφορίες και να μειώσει το σφάλμα ανίχνευσης του φόντου ως αντικειμένου. Το YOLO έχει μια ισχυρή ικανότητα γενίκευσης επειδή το YOLO μπορεί να μάθει πολύ γενικευμένα χαρακτηριστικά για να μεταφερθεί σε άλλα πεδία. Μετατρέπει το πρόβλημα της ανίχνευσης στόχου σε πρόβλημα παλινδρόμησης, αλλά η ακρίβεια ανίχνευσης πρέπει να βελτιωθεί. Τα αποτελέσματα των δοκιμών του YOLO είναι φτωχά για αντικείμενα που βρίσκονται πολύ κοντά το ένα στο άλλο και σε ομάδες. Αυτή η κακή απόδοση οφείλεται στο γεγονός ότι προβλέπονται μόνο δύο πλαίσια στο πλέγμα και ανήκουν μόνο σε μια νέα κλάση αντικειμένων της ίδιας κατηγορίας, επομένως εμφανίζεται ένας μη φυσιολογικός λόγος διαστάσεων και άλλες συνθήκες, όπως η αδύναμη ικανότητα γενίκευσης.
Λόγω της λειτουργίας απώλειας, το σφάλμα τοποθέτησης είναι ο κύριος λόγος για τη βελτίωση της αποτελεσματικότητας ανίχνευσης. Ειδικά ο χειρισμός μεγάλων και μικρών αντικειμένων πρέπει να ενισχυθεί. Κατά την υλοποίηση, το πιο σημαντικό είναι ο τρόπος σχεδιάσμου της λειτουργίας απώλειας έτσι ώστε αυτές οι τρεις πτυχές να μπορούν να εξισορροπηθούν καλά. Το YOLO χρησιμοποιεί πολλά χαμηλότερα στρώματα δειγματοληψίας και οι δυνατότητες προορισμού που αποκτήθηκαν από το δίκτυο δεν είναι εξαντλητικές, έτσι ώστε να βελτιωθεί το αποτέλεσμα ανίχνευσης.

Η αρχική αρχιτεκτονική YOLO αποτελείται από 24 στρώματα συνέλιξης, ακολουθούμενα από δύο πλήρως συνδεδεμένα στρώματα. Το YOLO προβλέπει πολλαπλά πλαίσια οριοθέτησης ανά κελί πλέγματος, αλλά επιλέγονται εκείνα τα πλαίσια οριοθέτησης που έχουν την υψηλότερη IOU (Intersection Over Union), η οποία είναι γνωστή ως καταστολή μη μέγιστων. Το YOLO έχει δύο ελαττώματα: το ένα είναι η ανακριβής τοποθέτηση και το άλλο είναι ο χαμηλότερος ρυθμός ανάκλησης σε σύγκριση με τη μέθοδο που βασίζεται στις συστάσεις περιοχής. Ως εκ τούτου, το YOLO V2 βελτιώνεται κυρίως σε αυτές τις δύο πτυχές. Εκτός αυτού, το YOLO V2 δεν εμβαθύνει ή διευρύνει το δίκτυο αλλά απλοποιεί το δίκτυο. Συμπερασματικά, δύο είναι οι βελτιώσεις του YOLO V2: Καλύτερο και ταχύτερο.

3.6.2 Εκδόσεις

Επειδή το YOLO και το YOLOv2 δεν είναι αποτελεσματικά στην ανίχνευση μικρών στόχων, η ανίχνευση πολλαπλών κλιμάκων προστίθεται στο YOLOv3. Το YOLOv4 ταξινόμησε και δοκίμασε όλες τις πιθανές βελτιστοποιήσεις σε ολόκληρη τη διαδικασία και βρήκε το καλύτερο αποτέλεσμα σε κάθε μετάθεση και συνδυασμό. Το YOLOv4 τρέχει δύο φορές πιο γρήγορα από το EfficientDet με συγκρίσιμη απόδοση, βελτιώνει το AP και το FPS του YOLOv3 κατά 10% και 12%, αντίστοιχα. Το YOLOv5 μπορεί να ελέγξει με ευελιξία μοντέλα και η μικρή του έκδοση είναι πολύ εντυπωσιακή. Τα συνολικά διαγράμματα δικτύου των YOLOv3 έως YOLO v5 είναι παρόμοια, αλλά εστιάζουν επίσης στην ανίχνευση αντικειμένων διαφορετικών μεγεθών από τρεις διαφορετικές κλίμακες. Τα κύρια μέτρα βελτίωσης του δικτύου YOLO από v1 σε v5 είναι τα εξής:

1. YOLO: Η διαίρεση του πλέγματος είναι υπεύθυνη για την ανίχνευση και την απώλεια εμπιστοσύνης.
2. YOLOv2: Άγκυρα με πρόσθετα K-means, εκπαίδευση δύο σταδίων, πλήρες συνελικτικό δίκτυο.
3. YOLOv3: Ανίχνευση πολλαπλών κλιμάκων με χρήση FPN.
4. YOLOv4: SPP, συνάρτηση ενεργοποίησης MISH, βελτίωση δεδομένων Mosaic/Mixup, συνάρτηση απώλειας GIOU (Generalized Intersection over Union).
5. YOLOv5: Ευέλικτος έλεγχος του μεγέθους του μοντέλου, εφαρμογή της συνάρτησης ενεργοποίησης Hardswish και βελτίωση δεδομένων.

3.7 Εκπαίδευση Βαθιών Νευρωνικών Δικτύων

Τα νευρωνικά δίκτυα είναι αλγόριθμοι εποπτευόμενης μάθησης, που σημαίνει ότι μαθαίνουν να προβλέπουν στόχους από συνδυασμό εισόδου-ετικέτας (input-label). Στην πράξη εκπαιδεύονται μέσω διαφορετικών παραλλαγών βελτιστοποίησης καθόδου κλίσης. Εδώ όλα τα μερικά παράγωγα μιας συνάρτησης απώλειας (loss function), του μέτρου για την απόκλιση του προβλεπόμενου αποτελέσματος y' από το βασικά αληθές y, έως τα βάρη του δικτύου, προέρχονται μέσω ανάστροφης διάδοσης (backpropagation). Με βάση αυτά τα παράγωγα, ο αλγόριθμος βελτιστοποίησης καθορίζει τον τρόπο με τον οποίο κάθε μια από τις παραμέτρους των δικτύων πρέπει να ενημερώνεται σε κάθε επανάληψη (iteration). Αυτή η διαδικασία εκτελείται επαναληπτικά σε τεράστια σύνολα δεδομένων εκπαίδευσης. Εάν το σύνολο δεδομένων είναι πολύ μικρό ή όχι αρκετά ποικιλόμορφο, το δίκτυο ενδέχεται να μην γενικευτεί για να μοντελοποιήσει μια αντιπροσωπευτική δοκιμή σετ και θα υπερφορτωθεί. Υπάρχουν διάφορες τεχνικές συστηματοποίησης (regularization techniques) για να αποφευχθεί αυτό.

3.7.1 Συνάρτηση απώλειας

Προκειμένου να βελτιστοποιηθεί μια λειτουργία μέσω μάθησης βάσει κλίσης, απαιτείται μια διαφοροποιήσιμη συνάρτηση απώλειας. Βέλτιστα, η κατανομή που ορίζεται από το μοντέλο νευρωνικού δικτύου θα πρέπει να είναι όσο το δυνατόν πλησιέστερα στη διανομή των δεδομένων που πρόκειται να αναπαρασταθούν. Με άλλα λόγια, το μοντέλο πρέπει να βελτιστοποιηθεί για τη μέγιστη πιθανότητα. Μαθηματικά, αυτό μπορεί να επιβληθεί μέσω της ελαχιστοποίησης της απώλειας διασταυρούμενης εντροπίας μεταξύ των ετικετών των δεδομένων εκπαίδευσης και των προβλέψεων του μοντέλου. Αντίστοιχα, η διασταυρούμενη εντροπία μπορεί επίσης να ελαχιστοποιηθεί μέσω της ελαχιστοποίησης του μέσου

τετραγωνικού σφάλματος (MSE) των προβλέψεων του μοντέλου [90]. Η κατηγορική διασταυρούμενη εντροπία έχει αποδειχθεί ότι ταιριάζει καλύτερα σε γενικά βαθιά νευρωνικά δίκτυα. Η συνάρτηση απώλειας και η παραγώγιση της σε σχέση με την έξοδο δίνεται ως:

$$L(t,o) = -\sum_{j}^{C} t_j \log o_j$$

$$\frac{\partial L}{\partial o_j} = \frac{-t_j}{o_j}$$

όπου C είναι ο αριθμός των νευρώνων εξόδου, t_j είναι 1 για την κατηγορία-στόχο, διαφορετικά είναι 0, και o_j είναι η προβλεπόμενη τιμή εξόδου του συστήματος. Το μέσο τετραγωνικό σφάλμα χρησιμοποιείται επίσης συχνά ως συνάρτηση απώλειας. Αυτή η λειτουργία και η παραγώγιση του σε σχέση με το αποτέλεσμα δίνεται ως εξής:

$$L(t,o) = \frac{1}{C}\sum_{j}^{C} \frac{1}{2}\left(t_j - o_j\right)^2$$

$$\frac{\partial L}{\partial o_j} = \frac{1}{C}\left(t_j - o_j\right)$$

Υπάρχουν και άλλες λειτουργίες απώλειας. Hinge-loss, που χρησιμοποιείται συνήθως σε κβαντικά νευρωνικά δίκτυα, είναι μια συνάρτηση μέγιστης απώλειας περιθωρίου [91].

3.7.2 Backpropagation

Μόλις επιλεγεί μια καλή συνάρτηση απώλειας, οι μερικές παράγωγοι αυτής της συνάρτησης σε οποιοδήποτε βάρος ή μεροληψία (bias) που συμβάλλει στο μοντέλο μπορεί να υπολογιστεί μέσω του αλυσιδωτού κανόνα της παραγώγισης. Το Backpropagation είναι ένας αποτελεσματικός αλγόριθμος που υπολογίζει αυτόν τον αλυσιδωτό κανόνα με συγκεκριμένη σειρά λειτουργιών. Αυτός ο κανόνας δίνεται εδώ:

$$\frac{\partial L}{\partial w_{ij}} = \frac{\partial L}{\partial o_j}\frac{\partial o_j}{\partial z_j}\frac{\partial z_j}{\partial w_{ij}}$$

Ή, γενικότερα:

$$\frac{\partial L}{\partial w_{ij}} = \Sigma\left[\frac{\partial L}{\partial o_p}\left(\sum_{k}^{p}\frac{\partial o_p}{\partial z_k}\frac{\partial z_k}{\partial w_{ij}}\right)\right]$$

3.7.3 Βελτιστοποίηση

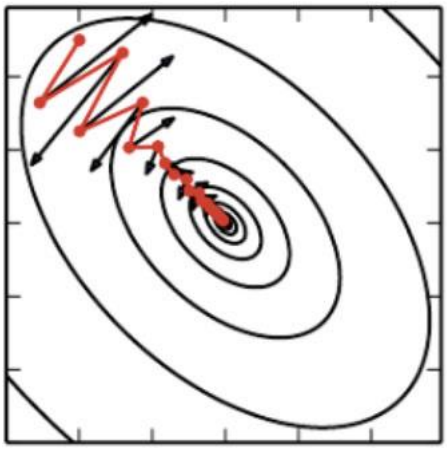
Καθώς η ελαχιστοποίηση της λειτουργίας απώλειας στη βαθιά μάθηση είναι μια ένα άκαμπτο πρόβλημα βελτιστοποίησης, είναι απαραίτητοι πολύ ισχυροί αλγόριθμοι βελτιστοποίησης. Εάν όχι, η βελτιστοποίηση μπορεί να κολλήσει σε ένα από τα πολλά τοπικά ελάχιστα, αντί να συγκλίνει με το καθολικό ελάχιστο στη συνάρτηση απώλειας. Επομένως, τα νευρωνικά δίκτυα είναι συνήθως εκπαιδευμένα μέσω παραλλαγών στον αλγόριθμο στοχαστικής κλίσης καθόδου (stochastic gradient descent = SGD).

3.7.4 SGD

Είναι μια στοχαστική προσέγγιση της βασικής βελτιστοποίησης κλίσης καθόδου, στην οποία οι κλίσεις εκτιμώνται μέσω backpropagation με βάση τον μέσο όρο απωλειών σε m δειγμάτα του συνόλου δεδομένων που περιέχει N >>m δείγματα. Όλες οι παράμετροι θ ενημερώνονται στη συνέχεια μέσω θ = θ − _g, όπου g είναι η εκτίμηση κλίσης για κάθε παράμετρο και _ είναι ένας μικρός ρυθμός μάθησης (learning rate). Αυτή η επαναληπτική μέθοδος έχει αρκετά πλεονεκτήματα σε σύγκριση με την "απλή" κάθοδο κλίσης, όπου η παράγωγος υπολογίζεται στο πλήρες σετ προπόνησης. Πρώτον, χρησιμοποιώντας στοχαστική κάθοδο κλίσης, ο χρόνος υπολογισμού ανά ενημέρωση δεν αυξάνεται με τον αριθμό των παραδειγμάτων εκπαίδευσης, αλλά μόνο με το μέγεθος m, minibatch. Αυτό είναι συνήθως μικρό και κυμαίνεται από 16- 256. Το σημαντικότερο είναι ότι η SGD εισάγει θόρυβο, ακόμη και όταν πλησιάζει το ελάχιστο. Αυτό εμποδίζει τη σύγκλιση του αλγορίθμου σε ένα τοπικό ελάχιστο, αλλά μπορεί επίσης να αποτρέψει οποιαδήποτε σύγκλιση. Στην πράξη, ο ρυθμός μάθησης συνήθως μειώνεται ενόσω η εκπαίδευση προχωρεί ώστε να αποφευχθεί αυτό.

3.7.5 Momentum

Είναι μια επέκταση της SGD για την επιτάχυνση της μαθησιακής διαδικασίας. Η ορμή (Momentum) συσσωρεύει ένα εκθετικά αποσυντιθέμενο κινητό μέσο όρο προηγούμενων διαβαθμίσεων και συνεχίζει να κινείται προς την κατεύθυνσή τους. Η ενημέρωση βάρους είναι τώρα θ = θ+v, όπου v είναι v = αv− _g. Εδώ α ∈ [0, 1] είναι μια υπερπαραμέτρος που καθορίζει την ταχύτητα της εκθετικής διάσπασης, g είναι και πάλι η εκτίμηση κλίσης που υπολογίστηκε σε ένα minibatch. Εάν το α είναι μεγαλύτερο, οι συνεισφορές από τις προηγούμενες κλίσεις θα είναι μεγαλύτερες. Αυτό γενικά οδηγεί σε ταχύτερους χρόνους σύγκλισης. Η έννοια της ορμής έναντι SGD απεικονίζεται στην Σχήμα 12.



Σχήμα 12. Απόδοση του momentum (κόκκινο) σε σύγκριση με απλό SGD (μαύρο)[92]

Υπάρχουν διάφοροι άλλοι αλγόριθμοι βελτιστοποίησης [93], [94]. Ένας είναι το παραδείγματα με τους προσαρμοστικούς ρυθμούς μάθησης. Μια καλή επισκόπηση της απόδοσης διαφορετικών μεθόδων δίνονται στο [95]. Εκτός από τη στρατηγική βελτιστοποίησης, επίσης ο ορισμός αρχικής κατάστασης (initialization) είναι το κλειδί. Υπάρχουν αρκετές στρατηγικές αρχικοποίησης που δεν εμπίπτουν στο πεδίο εφαρμογής αυτού του κειμένου.

Μια άλλη τεχνική στην εκπαίδευση σύγχρονων νευρωνικών δικτύων είναι η ομαλοποίηση δεσμίδων (batch normalization) [96], το οποίο είναι το κλειδί για την εκπαίδευση πολύ βαθιών μοντέλων. Στη κάθοδο κλίσης, όλες οι παράμετροι ενημερώνονται με την υπόθεση ότι όλες οι άλλες παράμετροι παραμένουν σταθερές. Αυτό φυσικά δεν συμβαίνει και μπορεί να οδηγήσει σε απροσδόκητα, ασταθή αποτελέσματα. Τα αποτελέσματα μιας ενημέρωσης στις παραμέτρους ενός επιπέδου θα είναι έντονα εξαρτώμενη από όλα τα άλλα στρώματα. Στην κανονικοποίηση δεσμίδων, το δίκτυο αναπαραμετροποιείται για να μετριάσει αυτό το πρόβλημα. Στην κανονικοποίηση δεσμίδων ένα minibatch των χαρτών χαρακτηριστικών F είναι αναπαραμετρημένο ως F'

$$F' = \frac{F-\mu}{\sigma}$$

όπου μ, σ είναι ο μέσος όρος και η τυπική απόκλιση του χάρτη χαρακτηριστικών minibatch. Στη συνέχεια, η αναπαραστατική ισχύς αποκαθίσταται μέσω της αντικατάστασης των χαρακτηριστικών με:

$$F'' = \gamma F' + \beta$$

όπου γ και β είναι βαθμωτές παράμετροι που μπορούν να εκπαιδευτούν. Αυτή η νέα παραμετροποίηση μπορεί να αντιπροσωπεύει τις ίδιες λειτουργίες όπως πριν, αλλά έχει πολύ διαφορετική, ευκολότερη μαθησιακή δυναμική.

3.7.6 Σετ δεδομένων

Προκειμένου να επιτευχθεί υψηλή ακρίβεια σε μια δεδομένη εργασία, τα νευρωνικά δίκτυα πρέπει να εκπαιδευτούν σε μεγάλο αριθμό ποικίλων δεδομένων. Εάν δεν δοθούν επαρκή δεδομένα στο δίκτυο, ενδέχεται να μην γενικεύει και να αρχίσει να υπερπροσαρμόζεται (overfitting). Η συλλογή τεράστιου όγκου επισημασμένων δεδομένων προπόνησης είναι ζωτικής σημασίας για να λειτουργήσει η βαθιά μάθηση. Τα σύνολα δεδομένων ανοιχτού κώδικα είναι άφθονα στην αναγνώριση ομιλίας, μετάφραση φυσικής γλώσσας και υπολογιστικής όρασης. Ωστόσο, σε πολλά άλλα πεδία, δεν είναι. Η αδυναμία συλλογής αρκετών επισημασμένων δεδομένων είναι σημαντικό εμπόδιο για την εφαρμογή τεχνικών βαθιάς μάθησης εκτός των πιο συνηθισμένων πεδίων. Όπως και τώρα, οι περισσότερες τεχνολογίες βαθιάς μάθησης έχουν εδραιωθεί, ένα επόμενο βήμα θα πρέπει να είναι η εξεύρεση αυτοματοποιημένων τρόπων συλλογής σύνθετων δεδομένων εκτός της ομιλίας, της επεξεργασία φυσικής γλώσσας (NLP) και της υπολογιστικής όρασης.

Ένα τυπικό σύνολο εκπαίδευσης χωρίζεται σε ένα υποσύνολο εκπαίδευσης (train) και εγκυρότητας (validation). Το υποσύνολο εκπαίδευσης χωρίζεται σε δεσμίδες και

χρησιμοποιείται στη διαδικασία SGD για την ελαχιστοποίηση της συνάρτησης απώλειας αυτών των δεδομένων μέσω ενημερώσεων των βαρών (weights). Σε συγκεκριμένα χρονικά διαστήματα, συνήθως μια εποχή (epoch), η απώλεια και η ακρίβεια του μοντέλου ελέγχεται στο υποσύνολο επικύρωσης. Το σύνολο εγκυρότητας (validation set) χρησιμοποιείται για να παρακολουθήσει την απόδοση του αλγορίθμου εκπαίδευσης. Μπορεί να χρησιμοποιηθεί για να καθορίσει πότε να σταματήσει την εκπαίδευση ή να αποτρέψει το δίκτυο από την υπερπροσαρμογή του.

Ως εκ τούτου, η εγκυρότητα μπορεί να θεωρείται μέρος του συνόλου κατάρτισης, καθώς αποτελεί αναπόσπαστο μέρος της εκπαιδευτικής διαδικασίας. Στην ιδανική περίπτωση, το σετ δοκιμών (test) είναι εντελώς ανεξάρτητο από το σετ προπόνησης. Το σετ δοκιμής χρησιμοποιείται για να ελέγξει τον τρόπο με τον οποίο το μοντέλο γενικεύεται σε άγνωστα και αόρατα δείγματα. Ως εκ τούτου, ο αλγόριθμος εκπαίδευσης δεν θα πρέπει να έχει δει δείγματα του σετ δοκιμών κατά τη διάρκεια της προπόνησης. Παραδείγματα συνόλων δεδομένων που χρησιμοποιούνται συχνά στην υπολογιστική όραση γενικά, είναι τα εξής:

3.7.6.1 MNIST

Το MNIST [97] είναι ένα μικρό σύνολο δεδομένων χειρόγραφων ψηφίων. Περιέχει 70.000 εικόνες, 28 × 28, κλίμακας του γκρι. Αυτό το σύνολο δεδομένων δεν θεωρείται ότι σχετίζεται με το να κρίνει την απόδοση των νέων αρχιτεκτονικών δικτύου, αλλά χρησιμοποιείται μάλλον ως γρήγορη περίπτωση δοκιμής στην ανάπτυξη κώδικα. Η ακρίβεια SotA σε αυτό το σημείο αναφοράς υπερβαίνει το 99%

3.7.6.2 IMAGENET

Το IMAGENET [98] είναι ένα μεγάλης κλίμακας σύνολο δεδομένων οπτικής αναγνώρισης που περιέχουν 1000 κλάσεις, μεταξύ των οποίων εικόνες σκυλιών, χρυσόψαρων, βουνών, αυτοκίνητων και ούτω καθεξής. Το πλήρες σύνολο δεδομένων αποτελείται από 150 GB 256 × 256 εικόνες RGB. Εξαιτίας αυτού, τα δίκτυα εκπαίδευσης στο IMAGENET είναι ένα μακρύ και κουραστικό έργο. Ωστόσο, οι καλές επιδόσεις σε αυτό το σύνολο δεδομένων είναι γενικά αξιόπιστες, και είναι ένα καλό μέτρο σύγκρισης απόδοσης ενός δικτύου σε σχέση με παρόμοια ιδιωτικά δίκτυα. Άλλα σύνολα δεδομένων που χρησιμοποιούνται συνήθως είναι το CIFAR-100, Pascal VOC [99], και COCO [100].

3.7.6.3 SVHN

Το SVHN [101] είναι ένα πραγματικό σύνολο δεδομένων εικόνας που περιέχει ψηφία και αριθμούς που κυμαίνονται από 0 έως 9 σε εικόνες φυσικής σκηνής. το SVHN έχει ληφθεί από πραγματικούς αριθμούς σπιτιών, από εικόνες του Google Street View. Το σύνολο δεδομένων αποτελείται από ένα σετ προπόνησης 73,000 φωτογραφιών, ένα σετ δοκιμών 26.000 και ένα επιπλέον σετ προπόνησης 531.000 έγχρωμων εικόνων, 32 × 32.

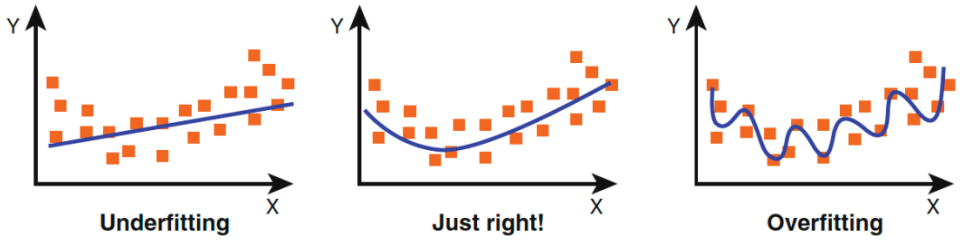
3.7.6.4 CIFAR-10

Το CIFAR-10 [102] είναι ένα μικρό σύνολο δεδομένων που περιέχει 10 κλάσεις, παραδείγματα των οποίων είναι το άλογο και το αυτοκίνητο, μεταξύ άλλων. Το σετ περιέχει 60.000 έγχρωμες εικόνες, 32 × 32, RGB. Αν και σημαντικά πιο περίπλοκο από το MNIST, αυτό το σύνολο δεδομένων χρησιμοποιείται κυρίως για τη γρήγορη αξιολόγηση νέων πλαισίων κατάρτισης ή αρχιτεκτονικές δικτύου. Μια καλή απόδοση ενός αλγορίθμου στο CIFAR-10 δεν αντιστοιχεί

απαραίτητα σε καλό αποδοτικό «custom» σύστημα. Η ακρίβεια σε αυτό το σημείο αναφοράς υπερβαίνει το 95%.

3.7.7 Κανονικοποίηση

Ένα δίκτυο υπερπροσαρμόζεται, όταν αποδίδει καλά στα δεδομένα της εκπαίδευσης του, αλλά όχι σε νέες εισόδους που δεν συνάντησε ξανά. Η έννοια της υποπροσαρμογής και της υπερπροσαρμογής απεικονίζεται στην Σχήμα 13. Ένα δίκτυο είναι υπερπροσαρμοσμένο όταν το σφάλμα εκπαίδευσης ενός μοντέλου (υπολογίζεται σε ένα σύνολο εκπαίδευσης) είναι πολύ χαμηλότερο από το σφάλμα γενίκευσης (υπολογίζεται σε ένα σύνολο δοκιμής ή εγκυρότητας) δηλ. training error < generalization error. Αυτό είναι αντίθετο με την υποπροσαρμογή, όταν ένα μοντέλο δεν δύναται να αποκτήσει, στο σετ εκπαίδευσης του, επαρκώς, χαμηλή τιμή σφάλματος. Η υπερπροσαρμογη συμβαίνει συνήθως όταν το πλήθος των χρησιμοποιούμενων παραμέτρων είναι πολύ μεγαλύτερος από την ποσότητα δειγμάτων κατάρτισης. Η υποπροσαρμογή μπορεί να επιλυθεί με την αύξηση της χωρητικότητας του μοντέλου (πηγαίνοντας σε βαθύτερα και ευρύτερα δίκτυα), τα οποία με τη σειρά τους μπορεί να οδηγήσουν σε υπερπροσαρμογή. Υπάρχουν πολλές αποτελεσματικές τεχνικές κανονικοποίησης, αποτρέποντας ένα μοντέλο από την υπερπροσαρμογή: η Dropout, η Adversarial training, η Data augmentation καθώς και η Parameter norm penalties κ.α.



Σχήμα 13. Απεικόνιση under- και overfitting [103]

3.7.7.1 Dropout

Το Dropout [104] είναι μια μέθοδος κανονικοποίησης στην οποία μερικές συνδέσεις μηδενίζονται τυχαία σε κάθε επανάληψη.

3.7.7.2 Adversarial training

Ο Adversarial training [90] είναι ένας τρόπος για να αποφευχθεί η υπερπροσαρμογή που χρησιμοποιεί αντίθετες εικόνες ως δείγματα εκπαίδευσης για την αύξηση της ακρίβειας του δικτύου.

3.7.7.3 Data augmentation

Το Data augmentation (επαύξηση δεδομένων) είναι ένα αποτελεσματικό μέσο για να γίνει ένα μοντέλο καλύτερο. Μέσω του Data augmentation, ψεύτικα δεδομένα δημιουργούνται και προστίθενται στο σύνολο εκπαίδευσης. Στην αναγνώριση εικόνας αυτό είναι απλό: ανεστραμμένες, περιστρεφόμενες, έγχρωμες εικόνες των προσώπων, εξακολουθούν να είναι μια αναπαράσταση της ίδιας τάξης. Στον πραγματικό κόσμο, το σύστημα θα συναντήσει πολλές τέτοιες παραλλαγές φυσικά. Ως εκ τούτου, ένα επαυξημένο σύνολο δεδομένων θα είναι μια καλύτερη αναπαράσταση του πανταχού παρόντος θορύβου. Η επαύξηση δεδομένων μπορεί να εφαρμόζεται και στην ομιλία, αλλά όχι σε πολλές άλλες εργασίες όπως η επεξεργασία φυσικής γλώσσας. Εάν η αύξηση δεδομένων δεν λειτουργεί, θα πρέπει να εφαρμοστούν άλλες τεχνικές.

3.7.7.4 Parameter norm penalties

Οι Parameter norm penalties (κυρώσεις κανόνα παραμέτρων) ισχύουν γενικά. Εδώ, προστίθεται μια ποινή στη συνάρτηση απώλειας για τον βέλτιστο περιορισμό της χωρητικότητας ενός μοντέλου. Η ολική συνάρτηση απώλειας τότε γίνεται L = L(θ, y) + αρ(θ), όπου L(θ, y) είναι η συνάρτηση απώλειας όπως και πριν, η οποία βελτιστοποιεί την ακρίβεια και ρ(θ) είναι ένας παράγοντας που εξαρτάται μόνο από την τιμή του συνόλου των παραμέτρων θ. Στην τακτοποίηση του L2 ή στην αποσύνθεση βαρών (weight decay), τα βάρη οδηγούνται κοντά στο μηδέν από ρ(θ) = 1 2 |w|22. Αυτό θα αναγκάσει τις ενημερωμένες παραμετρικές, να ενημερώσουν σε κατευθύνσεις που συμβάλλουν σημαντικά στη μείωση των συναρτήσεων που αποτελούν το αντικείμενο της αριστοποίησης. Οι παράμετροι που αντιστοιχούν σε ασήμαντες κατευθύνσεις θα βελτιστοποιηθούν χωριστά. Στην κανονικοποίηση L1, ρ(θ) =|w|1. Αυτή η κανονικοποίηση θα οδηγήσει σε μια λύση που είναι πιο αραιή, καθώς ένα υποσύνολο βαρών θα γίνει μηδέν, το οποίο μπορεί να θεωρηθεί ως μηχανισμός επιλογής χαρακτηριστικών [105].

3.7.8 Πλαίσια εκπαίδευσης

Καμία από τις τεχνικές που συζητήθηκαν παραπάνω δεν πρέπει να εφαρμοστεί από το μηδέν. Υπάρχει ένας αριθμός πλαισίων βαθιάς μάθησης ανοιχτού κώδικα. Αυτά διευκολύνουν την εκπαίδευση νευρωνικών δικτύων, με διαφορετικές τοπολογίες, με backpropagation σε υπολογιστικά συστήματα υψηλής απόδοσης όπως CPU και GPU. Σημαντικοί παράγοντες για να αποφασιστεί το ποιο πλαίσιο θα χρησιμοποιηθεί είναι (α) το μέγεθος της βάσης του χρήστη, (β) η ποσότητα του διαθέσιμου κώδικα με παραδείγματα και προεκπαιδευμένα μοντέλα, (γ) την προσφερόμενη υποστήριξη και (δ) τις δυνατότητες εντοπισμού σφαλμάτων. Μερικά από τα πιο δημοφιλή είναι το Pytorch, το Caffe και το Tensorflow.

3.7.8.1 Pytorch

Το Pytorch [106] και το Torch είναι πλαίσια ανοιχτού κώδικα που βασίζονται στη Lua, με εκτεταμένες βάσεις χρηστών. Καθώς το Pytorch ωθείται από το Facebook, έχει μια αυξανόμενη βάση χρηστών. Στο Pytorch, τα γραφήματα είναι δυναμικά και καταρτίζονται εν κινήσει, γεγονός που τα καθιστά πιο αργά, αλλά ευκολότερα στον εντοπισμό σφαλμάτων. Το συγκεκριμένο πλαίσιο θα αναφερθεί εκτενέστερα στη συνέχεια της εργασίας καθώς αποτελεί κομμάτι της εκπαίδευσης του δικού μας μοντέλου.

3.7.8.2 Caffe

Το Caffe [88] είναι ένα ακαδημαϊκό έργο με στατικά γραφήματα που εξελίχθηκε σε Caffe2. Καθώς το Caffe ήταν ένα από τα πρώτα πλαίσια εκεί έξω, δημιούργησε γρήγορα μια μεγάλη βάση χρηστών. Γενικά, τα χρησιμοποιούμενα πλαίσια βαθιάς μάθησης ποικίλλουν γρήγορα, αν και το Tensorflow τουλάχιστον φαίνεται ότι ήρθε για να μείνει.

3.7.8.3 Tensorflow

Το Tensorflow [105] αναπτύσσεται και υποστηρίζεται από την Google. Αυτό χρησιμοποιείται από μια τεράστια κοινότητα εμπειρογνωμόνων, οι οποίοι παρέχουν υλοποιήσεις ανοιχτού κώδικα και προεκπαιδευμένα μοντέλα δικτύων SotA. Το Keras είναι ένα περιτύλιγμα Tensorflow που παρέχει υποστήριξη για γρήγορη φόρτωση δεδομένων, data augmentation και ούτω καθεξής. Η βάση χρηστών του Tensorflow είναι τεράστια και πολλοί κώδικες είναι άμεσα διαθέσιμοι. Το πλαίσιο είναι καλά τεκμηριωμένο και υποστηρίζεται μέσω της Google, αλλά είναι δύσκολο να εντοπιστούν σφάλματα, καθώς όλα τα γραφήματα δικτύου είναι στατικά και προ-μεταγλωττισμένα, όχι εν κινήσει (όπως στην προεπιλεγμένη python). Αυτό σημαίνει ότι οι

υπολογισμοί θα είναι γρήγοροι, αλλά ο εντοπισμός σφαλμάτων μπορεί να είναι δύσκολος, αν και υπάρχει κάποια υποστήριξη.

3.8 Προκλήσεις στην εκπαίδευση Βαθιών Δικτύων

Η εκπαίδευση ενός βαθιού νευρωνικού δικτύου είναι ένα δύσκολο έργο, και μερικές από τις εξέχουσες προκλήσεις στην εκπαίδευση βαθιών μοντέλων συζητούνται παρακάτω.

3.8.1 Vanishing Gradient

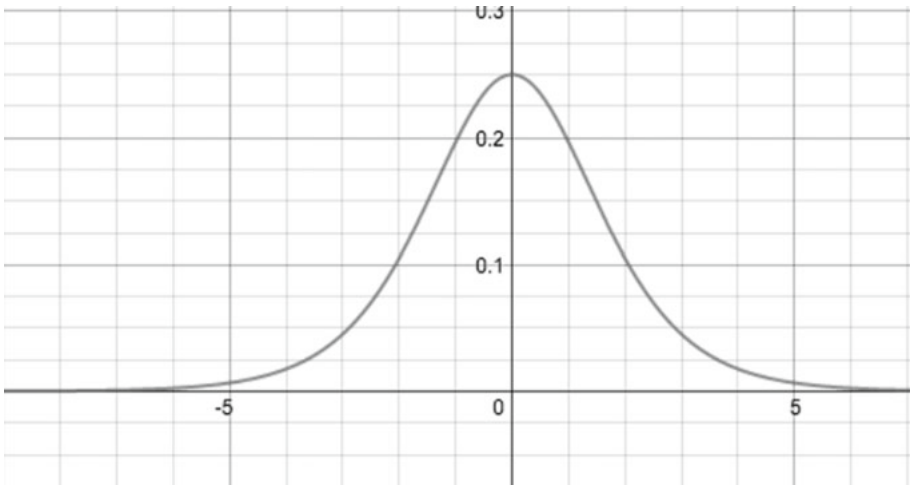
Οποιοδήποτε βαθύ νευρωνικό δίκτυο με λειτουργία ενεργοποίησης όπως σιγμοειδές, tanh κ.λπ. μέσω του backpropagation πάσχει από πρόβλημα εξαφάνισης κλίσης. Το Gradient καθιστά πολύ δύσκολη την εκπαίδευση και την ενημέρωση των παραμέτρων των αρχικών επιπέδων στο δίκτυο. Αυτό το πρόβλημα επιδεινώνεται καθώς αυξάνεται ο αριθμός των επιπέδων στο δίκτυο. Στόχος της οπισθοπορείας στα νευρωνικά δίκτυα είναι η επικαιροποίηση των παραμέτρων έτσι ώστε το σφάλμα του δικτύου να ελαχιστοποιείται και η πραγματική έξοδος να πλησιάζει στην έξοδο-στόχο. Κατά τη διάρκεια του backpropagation, τα βάρη ενημερώνονται χρησιμοποιώντας την κάθοδο κλίσης (ποσοστό μεταβολής του συνολικού σφάλματος E σε σχέση με οποιοδήποτε βάρος w). Στα βαθιά δίκτυα, αυτές οι διαβαθμίσεις καθορίζουν κατά πόσο πρέπει να αλλάξει κάθε βάρος. Οι κλίσεις γίνονται μικρότερες καθώς διαδίδονται μέσω πολλών στρωμάτων. Η σιγμοειδής λειτουργία δίνεται από τον τύπο

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

Το παράγωγο αυτής της σιγμοειδούς λειτουργίας δίνεται ως

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

Το γράφημα της παραπάνω εξίσωσης δίνεται στο Σχήμα 14. Από το γράφημα προκύπτει ότι το μέγιστο σημείο της συνάρτησης είναι 0,25, πράγμα που σημαίνει ότι η έξοδος της παραγώγου της συνάρτησης κόστους θα βρίσκεται πάντα μεταξύ 0 και 0,25. Με άλλα λόγια, τα σφάλματα θα συμπιέζονται στην περιοχή 0 και 0,25 σε κάθε στρώση. Επομένως, οι κλίσεις γίνονται όλο και μικρότερες μετά από κάθε στρώση και τελικά εξαφανίζονται αφήνοντας τα ανώτερα στρώματα ανεκπαίδευτα. Η κλίση εξαφάνισης είναι ο πρωταρχικός λόγος που κάνει τις ενεργοποιήσεις σιγμοειδούς ή tanh ακατάλληλες για βαθιά δίκτυα, και εδώ είναι που οι γραμμικές μονάδες (ReLUs) έρχονται να διασώσουν την κατάσταση. Η λειτουργία ενεργοποίησης ReLU δεν υποφέρει από εξαφανισμένη διαβάθμιση επειδή δεν υπάρχει συμπίεση των εισόδων, καθώς η παράγωγος είναι πάντα 1 για θετική εισροή. Μια διορθωμένη γραμμική μονάδα (ReLU) εξάγει 0 για είσοδο μικρότερη από 0 και ακατέργαστη έξοδο αλλιώς. Δηλαδή, εάν η είσοδος x είναι μικρότερη από 0, τότε η έξοδος είναι 0 και εάν το x είναι μεγαλύτερο από 0, η έξοδος είναι ίση με την είσοδο x και η παράγωγός της είναι 1. Δηλαδή f (x) = x και f_(x) _ 1 για x > 0.



Σχήμα 14. Διάγραμμα παραγώγου σιγμοειδούς συνάρτησης [107]

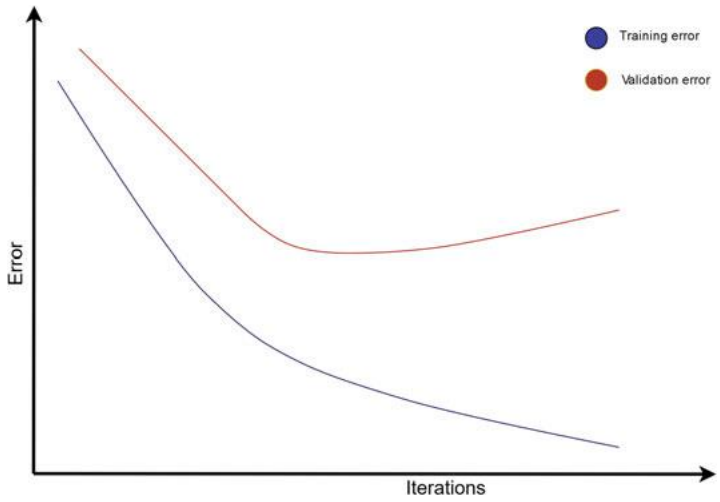
3.8.2 Μέγεθος εκπαιδευτικών δεδομένων

Τα βαθιά δίκτυα χρησιμοποιούν δεδομένα κατάρτισης για μάθηση και είναι ικανά να μαθαίνουν σύνθετες μη γραμμικές σχέσεις μεταξύ δεδομένων εισόδου και ετικέτας εξόδου. Τα βαθιά δίκτυα απαιτούν ένα μεγάλο αριθμό παραμέτρων που πρέπει να μάθουν πριν μπορέσουν να χρησιμοποιηθούν για την παράδοση του επιθυμητού αποτελέσματος. Ο αριθμός των παραμέτρων σε βαθιά μοντέλα είναι μεγάλος. Πιο σύνθετα μοντέλα σημαίνουν πιο ισχυρή αφαίρεση και περισσότερες παράμετροι που απαιτούν περισσότερα δεδομένα. Έτσι, το μέγεθος των δεδομένων κατάρτισης είναι ένας σημαντικός παράγοντας που μπορεί να επηρεάσει την επιτυχία των βαθιών μοντέλων. Στην πραγματικότητα, όλα τα επιτυχημένα βαθιά μοντέλα έχουν εκπαιδευτεί σε κάποιο πολύ μεγάλο σύνολο δεδομένων. Για παράδειγμα, το διαδίκτυο, το GoogleNet, το VGG, το ResNet κ.λπ., έχουν εκπαιδευτεί σε ένα τεράστιο σύνολο δεδομένων εικόνων που ονομάζεται ImageNet. Το ImageNet είναι ένα σύνολο δεδομένων εικόνας που περιέχει περίπου 1,2 εκατομμύρια εικόνες με ετικέτα που διανέμονται σε 1000 κατηγορίες. Ωστόσο, μπορεί κανείς να υποστηρίξει ότι τα βαθιά μοντέλα για αναγνώριση και ανίχνευση αντικειμένων απαιτούν μεγάλο αριθμό παραμέτρων για την αντιμετώπιση διαφορετικών παραλλαγών, διαφορετικών θέσεων, χρωμάτων κ.λπ., και έτσι απαιτούν τεράστιου μεγέθους σύνολο δεδομένων για εκπαίδευση. Από την άλλη πλευρά, λιγότερο σύνθετα προβλήματα (όπως η ταξινόμηση των ιατρικών εικόνων) όπου οι παραλλαγές είναι πολύ μικρές σε σύγκριση με τις παραλλαγές που αναφέρονται παραπάνω μπορούν να επιλυθούν χρησιμοποιώντας λιγότερο περίπλοκα μοντέλα που δεν απαιτούν τεράστια σύνολα δεδομένων κατάρτισης. Ο ισχυρισμός ισχύει σε κάποιο βαθμό, αλλά η πολυπλοκότητα του μοντέλου από μόνη της δεν μπορεί να καθορίσει το μέγεθος των δεδομένων που απαιτούνται για την κατάρτιση. Η ποιότητα των δεδομένων κατάρτισης διαδραματίζει επίσης σημαντικό ρόλο σε αυτό. Θορυβώδη δεδομένα σημαίνει χαμηλός λόγος σήματος προς θόρυβο (SNR) στα δεδομένα και χαμηλότερος SNR σημαίνει ότι απαιτούνται περισσότερα στοιχεία για τη σύγκλιση. Επομένως, το μέγεθος του συνόλου δεδομένων εξαρτάται πραγματικά από την πολυπλοκότητα του προβλήματος που μελετάται και την ποιότητα των δεδομένων. Ανεξάρτητα από την πολυπλοκότητα της εργασίας, το μεγάλο μέγεθος δεδομένων εκπαίδευσης μπορεί να οδηγήσει σε σημαντική βελτίωση της απόδοσης των βαθιών μοντέλων. Δηλαδή, όσο μεγαλύτερο είναι το μέγεθος των δεδομένων προπόνησης, τόσο καλύτερη είναι η ακρίβεια. Αλλά το ερώτημα "Πόσα δεδομένα είναι αρκετά;" παραμένει αναπάντητο και δεν υπάρχει εμπειρικός κανόνας που να μπορεί να

καθορίσει τον ακριβή αριθμό παραδειγμάτων που απαιτείται για την εκπαίδευση ενός συγκεκριμένου μοντέλου βάθους.

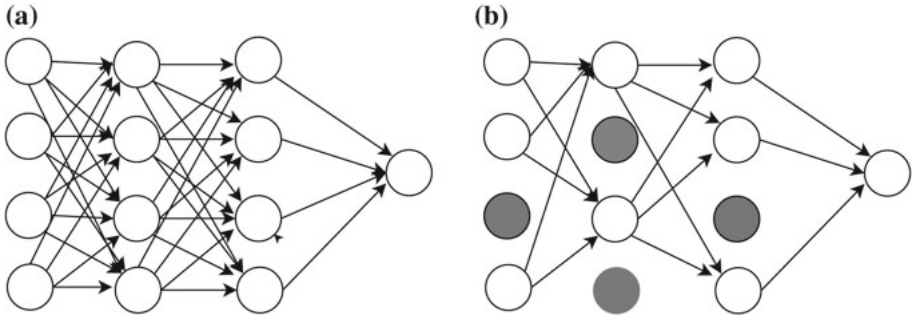
3.8.3 Υπερπροσαρμογή και υποπροσαρμογή

Μόλις ένα μοντέλο εκπαιδευτεί σε ένα σύνολο δεδομένων προπόνησης, αναμένεται να έχει καλή απόδοση σε νέα, προηγουμένως αθέατα, δεδομένα που δεν υπήρχαν κατά τη διάρκεια της μάθησης. Η ικανότητα ενός μοντέλου μηχανικής εκμάθησης για καλή απόδοση σε νέα και μη ορατά δεδομένα καλείται γενίκευση. Η γενίκευση είναι ένας από τους στόχους μιας καλής βαθιάς μάθησης. Για να εκτιμηθεί η ικανότητα γενίκευσης ενός μοντέλου βαθιάς μάθησης, δοκιμάζεται σε δεδομένα που συλλέγονται χωριστά από το σύνολο εκπαίδευσης. Τα μοντέλα βαθιάς μάθησης μπορεί να υποφέρουν από δύο προβλήματα, την υπερπροσαρμογή και την υποπροσαρμογή, και τα δύο μπορούν να οδηγήσουν σε κακή απόδοση του μοντέλου. Η υπερπροσαρμογή συμβαίνει όταν ένα μοντέλο εκπαιδεύεται και αποδίδει τόσο καλά στα προπονημένα δεδομένα που δεν είναι σε θέση να γενικεύσει σε νέα δεδομένα. Δηλαδή, το μοντέλο έχει χαμηλό σφάλμα εκπαίδευσης, αλλά δεν μπορεί να επιτύχει χαμηλό σφάλμα δοκιμής. Σε αυτή την περίπτωση, το μοντέλο απομνημονεύει τα δεδομένα αντί να τα μαθαίνει (Σχήμα 15). Η υποπροσαρμογή συμβαίνει όταν το μοντέλο δεν είναι σε θέση να μάθει σωστά και η απόδοσή του στο προπονητικό σετ είναι φτωχό. Το πιο κοινό πρόβλημα στη βαθιά μάθηση είναι η υπερβολική προσαρμογή. Βαθιά μοντέλα όπως τα συνελικτικά νευρωνικά δίκτυα (ConvNet) έχουν μεγάλο αριθμό παραμέτρων που μπορούν να εκπαιδευτούν, που πρέπει να μάθουν πριν μπορέσουν να χρησιμοποιηθούν για την εκτέλεση του καθήκοντος ενδιαφέροντος. Προκειμένου να εκπαιδευτούν εκτενώς αυτά τα μοντέλα, μεγάλα δεδομένα προπόνησης με συγκεκριμένο τελικό στόχο απαιτούνται για την επίτευξη της επιθυμητής απόδοσης. Εάν τα δεδομένα προπόνησης είναι πολύ μικρά σε σύγκριση με τον αριθμό των βαρών που πρέπει να μάθουν, τότε το δίκτυο πάσχει από υπερπροσπάθεια. Η υπερπροσαρμογή είναι ένα κοινό πρόβλημα στα βαθιά δίκτυα, ωστόσο, υπάρχουν λίγες τεχνικές διαθέσιμες που μπορούν να χρησιμοποιηθούν σε μοντέλα βαθιάς μάθησης για τον περιορισμό της υπερπροσαρμογής: (α) Αύξηση του συνόλου δεδομένων εκπαίδευσης. β) Μείωση του μεγέθους του δικτύου. (γ) Αύξηση δεδομένων: Τροποποίηση των τρεχόντων δεδομένων εκπαίδευσης με τυχαίο τρόπο (Scaling, ζουμ, μετάφραση κ.λπ.) για τη δημιουργία περισσότερων δεδομένων εκπαίδευσης. δ) Παρεμβολή κυρώσεων βάρους όπως η τακτοποίηση L1 και L2 και κοινή χρήση μαλακού βάρους. (ε) Εγκατάλειψη: Η πιο δημοφιλής τεχνική για τη μείωση της, είναι η εγκατάλειψη. Με την έννοια εγκατάλειψη ονομάζεται η εγκατάλειψη νευρώνων / μονάδων σε ένα νευρωνικό δίκτυο κατά τη διάρκεια της εκπαίδευσης.



Σχήμα 15. Overfitting: training error (μπλε), validation error (κόκκινο) ως συνάρτηση του αριθμού των iterations [107]

Η απόρριψη μιας μονάδας σημαίνει την προσωρινή αποκόλλησή της από το δίκτυο, συμπεριλαμβανομένων όλων των εσωτερικών και εξερχόμενων συνδέσεών της (Σχήμα 16). Οι νευρώνες που εγκαταλείφθηκαν ούτε συμβάλλουν στην προς τα εμπρός ούτε συμβάλλουν στην προς τα πίσω λειτουργεία. Χρησιμοποιώντας την εγκατάλειψη, το δίκτυο αναγκάζεται να μάθει πιο ισχυρές δυνατότητες και η αρχιτεκτονική αλλάζει με κάθε είσοδο.



Σχήμα 16. Ένα απλό νευρωνικό (a) και νευρωνικό δίκτυο μετα την απόρριψη (b) [107]

3.8.4 Υψηλής απόδοσης Hardware

Η εκπαίδευση μοντέλων σε βάθος σε τεράστια σύνολα δεδομένων απαιτεί μηχανήματα με επαρκή επεξεργαστική δύναμη και μνήμη. Για να έχετε υψηλή απόδοση και γρήγορο χρόνο προπόνησης, συνιστάται ιδιαίτερα η χρήση πολυπύρηνων μονάδων επεξεργασίας γραφικών υψηλής απόδοσης (GPU). Αυτά τα μηχανήματα υψηλής απόδοσης, οι GPU και η μνήμη, είναι πολύ δαπανηρά και καταναλώνουν πολλή ενέργεια. Ως εκ τούτου, η DL στον πραγματικό κόσμο γίνεται μια δαπανηρή και ενεργοβόρα εργασία.

4 ΚΕΦΑΛΑΙΟ 4: Υλοποίηση στο Jetson Nano

Το Jetson Nano Developer Kit είναι ένας μικρός, ισχυρός υπολογιστής που επιτρέπει να εκτελεστούν πολλά παράλληλα νευρωνικά δίκτυα για εφαρμογές όπως ταξινόμηση εικόνας, ανίχνευση αντικειμένων, τμηματοποίηση και επεξεργασία ομιλίας. Είναι ιδανικό για τη διδασκαλία, τη μάθηση και την ανάπτυξη τεχνητής νοημοσύνης και ρομποτικής. Είναι το μέσο που χρησιμοποιήθηκε για την εργασία αυτή. Σε αυτή την ενότητα θα αναφερθούν τα υλικολογισμικά χαρακτηριστικά του συγκεκριμένου συστήματος, καθώς και των παρελκομένων που χρίζονται ως απαραίτητα για την εκτέλεση μοντέλων υπο αυτή τη διαδικασία.

4.1 Τεχνικά χαρακτηριστικά

Πιο συγκεκριμένα, αποτελείται από μια 128-πύρηνη GPU NVIDIA αρχιτεκτονικής Maxwell, 4-πυρηνο CPU ARM A57 στα 1.43 GHz και μνήμη ram 2 GB των 64-bit τεχνολογίας LPDDR4. Στη κωδικοποίηση βίντεο παρέχεται η δυνατότητα έως 4K στα 30fps καθώς και μικρότερες αναλύσεις όπως 1080p30 (4x) και 720p30 (9x) σε H.264/H.265. Στην αποκωδικοποίηση αγγίζει την 4K στα 60 και 30 fps (2x, στην τελευταία περίπτωση) με τις μικρότερες αναλύσεις επίσης 1080p30 (8x) και 720p30 (18x) σε H.264/H.265. Η συνδεσιμότητα διατίθεται σε Gigabit

Ethernet και 802.11ac ασύρματα. Στην υποδοχή της κάμερας συναντάμε έναν σύνδεσμο τύπου MIPI CSI-2 και για την απεικόνιση έχει θύρα HDMI. Υπάρχουν θύρες USB. Μια USB 3.0 Type A, δυο USB 2.0 Type A και ένα USB 2.0 Micro-B. Άλλοι σύνδεσμοι που διατίθενται για διεπαφές είναι 40-pin, 12-pin και 4-pin υποδοχές. Οι διαστάσεις του board είναι: 100 mm x 80 mm x 29 mm.

4.2 JetPack SDK

Το JetPack παρέχει ένα πλήρες περιβάλλον ανάπτυξης για ανάπτυξη AI με επιτάχυνση υλικού. Όλες οι μονάδες Jetson και τα κιτ προγραμματιστών υποστηρίζονται από το JetPack. Περιλαμβάνει Jetson Linux με πρόγραμμα φόρτωσης εκκίνησης, πυρήνα Linux, περιβάλλον επιφάνειας εργασίας Ubuntu και ένα πλήρες σύνολο βιβλιοθηκών για επιτάχυνση υπολογιστών GPU, πολυμέσων, γραφικών και υπολογιστικής όρασης [108]. Περιλαμβάνει επίσης κιτ προγραμματιστών και υποστηρίζει SDK υψηλότερου επιπέδου, όπως το DeepStream για ροή αναλυτικών στοιχείων βίντεο, για τη ρομποτική το Isaac και για συνομιλία AI το Riva. Ειδικά χαρακτηριστικά του JetPack είναι τα κάτωθι [108]:

4.2.1 cuDNN

Το CUDA Toolkit προμηθεύει τους προγραμματιστές C και C++ με ένα άρτιο περιβάλλον ανάπτυξης την επινόηση εφαρμογών που απαιτούν GPU με επιτάχυνση. Το κιτ εργαλείων εμπεριέχει ένα πρόγραμμα μεταγλώττισης για κάρτες γραφικών της NVIDIA, βιβλιοθήκες μαθηματικών και τα κατάλληλα εργαλεία για να εντοπίζονται σφάλματα καθώς και να βελτιστοποιείται η απόδοση των χρησιμοποιούμενων εφαρμογών [108]. Η έκδοση που χρησιμοποιούμε, περιλαμβάνει CUDA 10.2 [108].

4.2.2 TensorRT

Το TensorRT είναι ένας χρόνος εκτέλεσης συμπερασμάτων βαθιάς μάθησης υψηλής απόδοσης για την ταξινόμηση εικόνων, την τμηματοποίηση και την ανίχνευση αντικειμένων νευρωνικών δικτύων. Το TensorRT βασίζεται στο CUDA, το μοντέλο παράλληλου προγραμματισμού της NVIDIA, και επιτρέπει την βελτιστοποίηση της εξαγωγής συμπερασμάτων για όλα τα πλαίσια βαθιάς μάθησης [108]. Περιλαμβάνει έναν βελτιστοποιητή συμπερασμάτων βαθιάς μάθησης και χρόνο εκτέλεσης που παρέχει χαμηλό λανθάνοντα χρόνο και υψηλή απόδοση για εφαρμογές συμπερασμάτων βαθιάς μάθησης. Η έκδοση που χρησιμοποιούμε εμπεριέχει το TensorRT 8.2.1 [108].

4.2.3 API Πολυμέσων

Το πακέτο API πολυμέσων Jetson παρέχει API χαμηλού επιπέδου για ευέλικτη ανάπτυξη εφαρμογών. Το libargus προσφέρει ένα API χαμηλού επιπέδου συγχρονισμού καρέ για εφαρμογές κάμερας, με έλεγχο παραμέτρων κάμερας ανά καρέ, πολλαπλής (συμπεριλαμβανομένης της συγχρονισμένης) υποστήριξης κάμερας και εξόδους ροής EGL. Οι κάμερες CSI εξόδου RAW που απαιτούν ISP δύνανται να χρησιμοποιηθούν είτε με το πρόσθετο GStreamer είτε με το libargus. Και στις δύο περιπτώσεις, γίνεται χρήση του API του προγράμματος οδήγησης αισθητήρα ελεγκτή μέσων V4L2 [108]. Το API V4L2 κάνει δυνατή την αποκωδικοποίηση βίντεο, την κωδικοποίηση, τη μετατροπή μορφής και τη λειτουργία κλιμάκωσης. Όσον αφορά την κωδικοποίηση, με τη χρήση του V4L2, ανοίγει πολλές δυνατότητες όπως έλεγχο ρυθμού bit, προεπιλογές ποιότητας, κωδικοποίηση χαμηλού λανθάνοντος χρόνου, χρονική αντιστάθμιση, διανυσματικούς χάρτες κίνησης και πολλά άλλα.

Το JetPack 4.6.1 περιλαμβάνει τα ακόλουθα κύρια σημεία στα πολυμέσα: Υποστήριξη για κλιμακούμενη κωδικοποίηση βίντεο (SVC). Η υποστήριξη για YUV444 8, 10 bit κωδικοποίηση και αποκωδικοποίηση [108]

4.2.4 Υπολογιστική Όραση

Το VPI (Vision Programing Interface) είναι μια βιβλιοθήκη λογισμικού που παρέχει αλγόριθμους υπολογιστικής όρασης / επεξεργασίας εικόνας που υλοποιούνται σε PVA1 (προγραμματιζόμενος επιταχυντής όρασης), GPU και CPU. Το OpenCV είναι μια κορυφαία βιβλιοθήκη ανοιχτού κώδικα για υπολογιστική όραση, επεξεργασία εικόνας και μηχανική μάθηση. Το VisionWorks2 είναι ένα πακέτο ανάπτυξης λογισμικού για υπολογιστική όραση (CV) και επεξεργασία εικόνας [108]. Το JetPack 4.6.1 περιλαμβάνει το VPI 1.2 στο οποίο παρέχονται: υποστήριξη ποιότητας παραγωγής για συνδέσεις Python, υποστήριξη πολλαπλών ροών σε συνδέσεις Python για να επιτρέπεται η δημιουργία πολλαπλών ροών για τον παραλληλισμό των λειτουργιών, υποστήριξη για την κλήση δεσμών ενεργειών Python σε ροή VPI και νέοι αλγόριθμοι. Η έκδοση που χρησιμοποιούμε, περιλαμβάνει το OpenCV 4.1.1 και το VisionWorks 1.6 [108]

4.2.5 Εργαλεία προγραμματιστών

Το CUDA Toolkit παρέχει ένα ολοκληρωμένο περιβάλλον ανάπτυξης για προγραμματιστές C και C++ που δημιουργούν εφαρμογές υψηλής απόδοσης με επιτάχυνση GPU με βιβλιοθήκες CUDA. Η εργαλειοθήκη περιλαμβάνει το Nsight Eclipse Edition, εργαλεία εντοπισμού σφαλμάτων και δημιουργίας προφίλ, συμπεριλαμβανομένου του Nsight Compute, και μια αλυσίδα εργαλείων για τη διασταυρούμενη μεταγλώττιση εφαρμογών [108].

Το NVIDIA Nsight Systems είναι ένα εργαλείο δημιουργίας χαμηλής επιβάρυνσης προφίλ σε όλο το σύστημα, παρέχοντας τις πληροφορίες που χρειάζονται οι προγραμματιστές για να αναλύσουν και να βελτιστοποιήσουν την απόδοση του λογισμικού. Είναι μια αυτόνομη εφαρμογή για εφαρμογές εντοπισμού σφαλμάτων και δημιουργίας προφίλ γραφικών. Η έκδοση που χρησιμοποιούμε περιλαμβάνει NVIDIA Nsight Graphics 2021.2 [108].

4.2.6 Υποστηριζόμενα SDK και εργαλεία

Το NVIDIA DeepStream SDK είναι ένα πλήρες κιτ εργαλείων ανάλυσης για επεξεργασία πολλαπλών αισθητήρων που βασίζεται σε AI και κατανόηση βίντεο και ήχου. Το JetPack 4.6.1 υποστηρίζει το DeepStream SDK 6.0. Ο διακομιστής συμπερασμάτων NVIDIA Triton™ απλοποιεί την ανάπτυξη μοντέλων AI σε κλίμακα. Είναι ανοιχτού κώδικα και υποστηρίζει την ανάπτυξη εκπαιδευμένων μοντέλων AI από την NVIDIA TensorRT, tensorflow και χρόνο εκτέλεσης ONNX στο Jetson [108]. Στο Jetson, ο διακομιστής συμπερασμάτων Triton παρέχεται ως κοινόχρηστη βιβλιοθήκη για άμεση ενοποίηση με το C API. Το PowerEstimator είναι μια εφαρμογή web που απλοποιεί τη δημιουργία προσαρμοσμένων προφίλ λειτουργίας ισχύος και εκτιμά την κατανάλωση ισχύος της μονάδας Jetson. Ο εκτιμητής ισχύος v1.1 υποστηρίζει το JetPack 4.6 [108].

4.2.7 Cloud Native

Η Jetson φέρνει το Cloud-Native και επιτρέπει τεχνολογίες όπως τα κοντέινερ και την ενορχήστρωση κοντέινερ. Το JetPack περιλαμβάνει χρόνο εκτέλεσης κοντέινερ NVIDIA με

ενοποίηση Docker, επιτρέποντας εφαρμογές με επιτάχυνση GPU σε κοντέινερ στην πλατφόρμα Jetson [108]. Η NVIDIA φιλοξενεί πολλές εικόνες κοντέινερ για το Jetson στο NVIDIA NGC. Ορισμένα είναι κατάλληλα για ανάπτυξη λογισμικού με δείγματα και τεκμηρίωση και άλλα είναι κατάλληλα για ανάπτυξη λογισμικού παραγωγής, το οποίο περιέχει μόνο στοιχεία χρόνου εκτέλεσης [108].

4.2.8 Ασφάλεια

Οι μονάδες NVIDIA Jetson περιλαμβάνουν διάφορα χαρακτηριστικά ασφαλείας όπως Hardware Root of Trust, ασφαλή εκκίνηση, επιτάχυνση κρυπτογράφησης υλικού, περιβάλλον αξιόπιστης εκτέλεσης, κρυπτογράφηση δίσκου και μνήμης, προστασία από φυσικές επιθέσεις και πολλά άλλα [108].

4.2.9 Λειτουργική Ασφάλεια

Η προσέγγιση της NVIDIA Jetson στη λειτουργική ασφάλεια είναι να παρέχει πρόσβαση στη βάση διαγνωστικών σφαλμάτων υλικού που μπορεί να χρησιμοποιηθεί στο πλαίσιο του σχεδιασμού συστημάτων που σχετίζονται με την ασφάλεια. Το πακέτο επέκτασης ασφάλειας Jetson (JSEP) παρέχει διαγνωστικό σφάλμα και πλαίσιο αναφοράς σφαλμάτων για την εφαρμογή λειτουργιών ασφαλείας και την επίτευξη συμμόρφωσης με τα πρότυπα λειτουργικής ασφάλειας. Όταν χρησιμοποιείται μια GPU NVIDIA και το SDK βαθιάς εκμάθησης, υπάρχουν μερικές βέλτιστες πρακτικές που μπορούν να βοηθήσουν στην εξασφάλιση της καλύτερης απόδοσης [108].

4.3 Το μοντέλο CUDA

Το CUDA είναι ένα μοντέλο προγραμματισμού και μια πλατφόρμα για παράλληλο υπολογισμό που δημιουργήθηκε από την NVIDIA. Ο προγραμματισμός CUDA σχεδιάστηκε για υπολογιστές με τις μονάδες επεξεργασίας γραφικών (GPU) της NVIDIA. Το CUDA επιτρέπει στους προγραμματιστές να μειώσουν το χρόνο που απαιτείται για την εκτέλεση εργασιών έντονου υπολογισμού, επιτρέποντας στους φόρτους εργασίας να εκτελούνται σε GPU και να διανέμονται σε παράλληλες GPU.

Κατά την εκτέλεση υπολογιστικών λειτουργιών χρησιμοποιώντας GPU χρησιμοποιούνται τόσο κεντρικές μονάδες επεξεργασίας (CPU) όσο και GPU. Οι CPU εκτελούν τα διαδοχικά τμήματα του φόρτου εργασίας, καθώς είναι βελτιστοποιημένα για απόδοση ενός νήματος. Εν τω μεταξύ, οι μη διαδοχικές εργασίες πολλαπλών νημάτων εκτελούνται σε GPU οι οποίες στη συνέχεια επιστρέφουν το αποτέλεσμά τους στην CPU.

4.3.1 Εργαλειοθήκη CUDA

Το CUDA Toolkit περιλαμβάνει βιβλιοθήκες, εργαλεία εντοπισμού σφαλμάτων και βελτιστοποίησης, ένα πρόγραμμα μεταγλώττισης, τεκμηρίωσης και μια βιβλιοθήκη χρόνου εκτέλεσης για την ανάπτυξη των εφαρμογών. Έχει στοιχεία που υποστηρίζουν βαθιά μάθηση, γραμμική άλγεβρα, επεξεργασία σήματος και παράλληλους αλγόριθμους. Σε γενικές γραμμές, οι βιβλιοθήκες CUDA υποστηρίζουν όλες τις οικογένειες GPU NVIDIA, αλλά αποδίδουν καλύτερα στην τελευταία γενιά, όπως η V100, το οποίο μπορεί να είναι 3 φορές ταχύτερο από την P100 για φόρτους εργασίας εκπαίδευσης βαθιάς μάθησης. Η A100 μπορεί να προσθέσει μια επιπλέον επιτάχυνση 2x. Η χρήση μίας ή περισσότερων βιβλιοθηκών είναι ο ευκολότερος τρόπος για την καλύτερη εκμετάλλευση της GPU, αρκεί οι αλγόριθμοι που απαιτούνται, να έχουν υλοποιηθεί στην κατάλληλη βιβλιοθήκη.

4.3.2 OpenCL vs. CUDA

Ο ανταγωνιστής της CUDA, OpenCL κυκλοφόρησε το 2009, σε μια προσπάθεια να παρέχει ένα πρότυπο για ετερογενείς υπολογιστές που δεν περιοριζόταν σε επεξεργαστές Intel / AMD με GPU NVIDIA. Ενώ το OpenCL ακούγεται ελκυστικό λόγω της γενικότητάς του, δεν έχει αποδώσει τόσο καλά όσο το CUDA σε GPU NVIDIA και πολλά πλαίσια βαθιάς μάθησης είτε δεν υποστηρίζουν το OpenCL είτε το υποστηρίζουν μόνο ως δεύτερη σκέψη, μεχρι να κυκλοφορήσει η υποστήριξη για CUDA.

4.4 Τι είναι το cuDNN;

Το NVIDIA CUDA Deep Neural Network (cuDNN) είναι μια βιβλιοθήκη με επιτάχυνση GPU για βαθιά νευρωνικά δίκτυα, παρέχοντας εξαιρετικά συντονισμένες τυπικές υλοποιήσεις ρουτίνας, όπως κανονικοποίηση, ομαδοποίηση, συνέλιξη εμπρός και πίσω και στρώματα ενεργοποίησης. Η βιβλιοθήκη cuDNN επιτρέπει στους προγραμματιστές και τους ερευνητές πλαισίων βαθιάς μάθησης παντού να αξιοποιήσουν την επιτάχυνση GPU για υψηλή απόδοση. Μειώνει την ανάγκη βελτίωσης της απόδοσης της GPU σε χαμηλό επίπεδο, εξοικονομώντας χρόνο, ώστε να καθίσταται δυνατή η επικέντρωση στην ανάπτυξη του λογισμικού και στην εκπαίδευση των νευρωνικών δικτύων. Η επιτάχυνση cuDNN υποστηρίζει δημοφιλή πλαίσια βαθιάς μάθησης όπως Κεράς, Caffe2, Chainer, MxNet, MATLAB, TensorFlow και PyTorch.

4.4.1 Χαρακτηριστικά του cuDNN

Τα βασικά χαρακτηριστικά της cuDNN της NVIDIA περιλαμβάνουν την επιτάχυνση των συγχωνευμένων λειτουργιών για όλες τις αρχιτεκτονικές CNN. Υποστηρίζει μορφές ακεραίων UINT8 και INT8 και μορφές κινητής υποδιαστολής BF16, FP16, FP32 και TF32. Χρησιμοποιείται η επιτάχυνση απόδοσης πυρήνα tensor για ευρέως χρησιμοποιούμενες συνελίξεις όπως 2D, 3D, ομαδοποιημένες, διαχωρίσιμες σε βάθος και διασταλμένες (με εισόδους και εξόδους NCHW και NHWC). Επιπλεον βελτιστοποιεί τον πυρήνα για μοντέλα ομιλίας και υπολογιστικής όρασης όπως ResNet, ResNext, EfficientDet, EfficientNet, SSD, MaskRCNN, Tacotron2, Unet και VNet. Τέλος, επιτυγχάνει ενσωμάτωση με όλες τις υλοποιήσεις νευρωνικών δικτύων με χρήση αυθαίρετης ταξινόμησης διαστάσεων, υποπεριοχών τανυστή 4D και διασκελισμού (striding).

Το cuDNN απολαμβάνει υποστήριξη από Linux και Windows με μια ποικιλία αρχιτεκτονικών κινητής GPU και κέντρων δεδομένων, συμπεριλαμβανομένων των Ampere, Volta, Turing, Pascal, Kepler και Maxwell. Η τελευταία έκδοση του cuDNN είναι η 8.3, η οποία παρέχει βελτιωμένη απόδοση με GPU A100 (έως και πέντε φορές υψηλότερη από τις έτοιμες GPU V100). Προσφέρει επίσης νέα API και βελτιστοποιήσεις για εφαρμογές υπολογιστικής όρασης και συνομιλίας ΤΝ.

Ο επανασχεδιασμός της έκδοσης 8.3 είναι φιλικός προς το χρήστη και προσφέρει βελτιωμένη ευελιξία και εύκολη ενσωμάτωση εφαρμογών. Περιλαμβάνει βελτιστοποιήσεις για την επιτάχυνση μοντέλων βαθιάς μάθησης που βασίζονται σε μετασχηματισμούς, συγχώνευση χρόνου εκτέλεσης για τη μεταγλώττιση πυρήνων με νέους τελεστές και μικρότερο πακέτο λήψης, ιδιαίτερα επωφελές για μικρά συστήματα που χρειάζονται ελαφριά προγράμματα εγκατάστασης.

4.5 Περιβάλλον ONNX

Η κοινότητα των συστημάτων βαθιάς μάθησης, των πλατφορμών και των gadgets αυξάνεται με εκπληκτικά τεράστια ταχύτητα. Πριν από μερικά χρόνια, υπήρχαν λίγα πλαίσια βαθιάς μάθησης και μηχανικής μάθησης διαθέσιμα στον κόσμο, δεν περνάει ούτε ένας μήνας αυτές τις μέρες χωρίς να ακούσουμε για ένα νέο εξαιρετικό πλαίσιο βαθιάς μάθησης που ειδικεύεται σε έναν συγκεκριμένο τομέα διαλειτουργικότητας μεταξύ των διαφορετικών στοιβών βαθιάς μάθησης στην αγορά που σκοτώνει κάθε ελπίδα επαναχρησιμοποίησης μοντέλων και δικτύων σε διαφορετικούς χρόνους εκτέλεσης. Η βασική λειτουργία της βαθιάς μάθησης ολοκληρώνεται μέσω υπολογισμού σε γραφήματα ροής δεδομένων. Κάτω από το υπόστεγο της βαθιάς γνώσης, τα γραφήματα χωρίζονται σε Δυναμικά Γραφήματα και Στατικά Γραφήματα.

Διαφορετικά πλαίσια βαθιάς μάθησης χρησιμοποιούν διαφορετικά είδη γραφημάτων. Πλαίσια όπως το CNTK, το Caffe2, το Theano και το Tensorflow προτιμούν τη χρήση στατικών γραφημάτων. Από την άλλη πλευρά, πλαίσια όπως το PyTorch και το Chainer χρησιμοποιούν δυναμικά γραφήματα. Αυτά τα γραφήματα παρέχουν μια ενδιάμεση αναπαράσταση που συλλαμβάνει τη συγκεκριμένη πρόθεση οποιουδήποτε πηγαίου κώδικα. Μπορεί να τρέξει στον αριθμό των συσκευών (FPGA, GPU, CPU κ.λπ.).

Η εμφάνιση και ανάδυση διαφορετικών πλαισίων απαιτεί επίσης τη διαλειτουργικότητα τους. Η ζήτηση για αυτά τα πλαίσια ευελιξίας και φορητότητας γίνεται πιο κρίσιμη από ποτέ. Το πρώτο βήμα που πλησιάζει για να αγγίξει την προαναφερθείσα ζήτηση είναι η άνοδος των ανοιχτών περιβαλλόντων γνωστών ως Open Neural Network Exchange. Πρόκειται για μια μορφή ανοιχτού κώδικα που υπόκειται σε διαμόρφωση ανοιχτού κώδικα για μοντέλα ΤΝ. Χαρακτηρίζει ένα επεκτάσιμο υπολογιστικό μοντέλο γραφήματος και τη μετάφραση ενσωματωμένων τελεστών και έγκυρων τύπων δεδομένων.

4.5.1 Λόγοι χρησιμοποίησης ONNX

Δύο βασικά πράγματα για ένα Μοντέλο ΤΝ είναι η παραγωγή του μοντέλου και φυσικά η ανάπτυξή του (deployment). Η τεχνική ικανότητα που παρέχει το ONNX επιτρέπει σε έναν επιστήμονα δεδομένων να εισάγει γρήγορα εξαιρετικές ιδέες στην παραγωγή. Δίνει ένα επιπλέον πλεονέκτημα για την επιλογή ενός συγκεκριμένου πλαισίου για μια συγκεκριμένη εργασία από διαφορετικά διαθέσιμα πλαίσια, με αποτέλεσμα να δαπανάται λιγότερος χρόνος για να καταστεί ένα μοντέλο έτοιμο για παραγωγή και ανάπτυξή του. Ένα οικοσύστημα εργαλείων για την οπτικοποίηση και την επιτάχυνση των μοντέλων παρέχεται επίσης υπό τις λειτουργίες των μοντέλων ONNX. Για την υποστήριξη της έννοιας της μεταγραφικής μάθησης διατίθενται επίσης προ-εκπαιδευμένα μοντέλα ONNX για κοινά σενάρια.

4.6 Πλαίσιο PyTorch

Το PyTorch είναι ένα πλαίσιο μηχανικής μάθησης ανοιχτού κώδικα που βασίζεται στην Python. Δίνει τη δυνατότητα να εκτελεστούν επιστημονικοί και τανυστικοί υπολογισμοί με τη βοήθεια γραφικών μονάδων επεξεργασίας (GPU). Μπορεί να το χρησιμοποιηθεί για να αναπτυχθούν και να εκπαιδευθούν νευρωνικά δίκτυα βαθιάς μάθησης χρησιμοποιώντας αυτόματη διαφοροποίηση (μια διαδικασία υπολογισμού που δίνει ακριβείς τιμές σε σταθερό χρόνο). Τα βασικά χαρακτηριστικά του PyTorch περιλαμβάνουν ένα εύχρηστο API που μπορεί να χρησιμοποιηθεί με Python, C++ ή Java, όπως και το Pythonic το οποίο ενσωματώνεται ομαλά με τη στοίβα επιστήμης δεδομένων της Python και επιτρέπει την αξιοποίηση των υπηρεσιών

και των λειτουργιών της Python. Τέλος διαθέτει δυνατότητες για δυναμικά υπολογιστικά γραφήματα που μπορείτε να προσαρμόσετε κατά τη διάρκεια του χρόνου εκτέλεσης.

4.6.1 Υποστηριξη CUDA για PyTorch

Το CUDA είναι το κυρίαρχο API που χρησιμοποιείται για βαθιά μάθηση, αν και υπάρχουν και άλλες επιλογές, όπως το OpenCL. Το PyTorch παρέχει υποστήριξη για το CUDA στη βιβλιοθήκη torch.cuda. Η βιβλιοθήκη CUDA του PyTorch επιτρέπει την παρακολούθηση της GPU που χρησιμοποιείται και προκαλεί την αυτόματη αντιστοίχιση των τανυστών που δημιουργούνται σε αυτήν τη συσκευή. Αφού εκχωρηθεί ένας τανυστής, μπορεί να εκτελέσει λειτουργίες και τα αποτελέσματα εκχωρούνται επίσης στην ίδια συσκευή. Από προεπιλογή, μέσα στο PyTorch, δεν μπορούν να χρησιμοποιηθούν λειτουργίες μεταξύ GPU. Εξαίρεση αποτελεί η χρήση μεθόδων copy_() ή αντιγραφής, όπως to() και cuda(). Για την εκκίνηση λειτουργιών σε κατανεμημένους τανυστές, πρέπει πρώτα να ενεργοποιηθεί η ομότιμη πρόσβαση στη μνήμη.

Οι λειτουργίες GPU είναι ασύγχρονες από προεπιλογή για να επιτρέψουν την παράλληλη εκτέλεση μεγαλύτερου αριθμού υπολογισμών. Οι ασύγχρονες λειτουργίες είναι γενικά αόρατες στο χρήστη, επειδή το PyTorch συγχρονίζει αυτόματα τα δεδομένα που αντιγράφονται μεταξύ CPU και GPU ή GPU και GPU. Επιπλέον, οι λειτουργίες εκτελούνται με τη σειρά αναμονής. Αυτό εξασφαλίζει ότι οι λειτουργίες εκτελούνται με τον ίδιο τρόπο όπως εάν οι υπολογισμοί ήταν σύγχρονοι. Εάν πρέπει να χρησιμοποιηθούν σύγχρονες λειτουργίες, μπορεί να επιβληθεί αυτή η ρύθμιση με τη μεταβλητή περιβάλλοντος CUDA_LAUNCH_BLOCKING=1. Για παράδειγμα, μπορεί να χρειαστεί να γίνει αυτό εάν παρατηρούνται σφάλματα στις GPU. Η σύγχρονη εκτέλεση διασφαλίζει ότι τα σφάλματα αναφέρονται όταν προκύπτουν και διευκολύνει τον εντοπισμό του αιτήματος από το οποίο προήλθε το σφάλμα. Μια άλλη περίπτωση που πρέπει να προσεχθεί αν θα χρησιμοποιηθούν λειτουργίες μη συγχρονισμού ή συγχρονισμού είναι οι μετρήσεις χρόνου. Με τις ασύγχρονες λειτουργίες, οι μετρήσεις δεν θα είναι ακριβείς. Για να επιλυθεί αυτό το πρόβλημα αφήνοντας ενεργοποιημένο το async, μπορεί να κληθεί το torch.cuda.synchronize() πριν από τη μέτρηση ή μπορεί να χρησιμοποιηθεί το torch.cuda.event για να καταγράψει τους χρόνους.

Οι ροές CUDA είναι γραμμικές ακολουθίες εκτέλεσης σε συγκεκριμένες GPU. Αυτές οι ροές δημιουργούνται από προεπιλογή κατά τη λειτουργία. Μέσα σε κάθε ροή, οι λειτουργίες σειριοποιούνται κατά σειρά δημιουργίας. Ωστόσο, οι λειτουργίες από διαφορετικές ροές μπορούν να εκτελεστούν ταυτόχρονα με οποιαδήποτε σχετική σειρά. Η εξαίρεση αφορά την χρησιμοποίηση μεθόδων synchronize() ή wait_stream(). Υπόψη ότι, αν έχει ορίστει η προεπιλεγμένη ροή σε " current stream ", το PyTorch συγχρονίζει αυτόματα τα δεδομένα. Ωστόσο, εάν χρησιμοποιηθούν μη προεπιλεγμένες ροές, απαιτείται ιδιαίτερη προσοχή στην εκτέλεση αυτού του συγχρονισμού.

4.7 Βέλτιστες πρακτικές για DL

4.7.1 Ενεργοποίηση πυρήνων τανυστή (Tensor)

Οι πυρήνες tensor επεξεργάζονται πυρήνες που έχουν σχεδιαστεί ειδικά για την επιτάχυνση των διαδικασιών DL. Μπορούν να χρησιμοποιηθούν αποτελεσματικά πυρήνες tensor με δεδομένα INT8, FP 16 ή FP 32. Για το τελευταίο, μπορούν να χρησιμοποιηθούν μικτές μέθοδοι ακρίβειας. Εν τω μεταξύ, θα πρέπει να επιλεγεί ένας αριθμός βασικών παραμέτρων που διαιρείται με οκτώ με FP16 και 16 εάν χρησιμοποιείται INT8. Το αν μπορούν να χρησιμοποιηθούν αυτοί οι

πυρήνες καθορίζεται από τις παραμέτρους. Αυτοί οι προσδιορισμοί ποικίλλουν ανάλογα με τον τύπο αρχιτεκτονικής.

1. Πλήρως συνδεδεμένα στρώματα (Fully-connected layers) — καθορίζεται από τον αριθμό των εισόδων και εξόδων και τα μεγέθη των παρτίδων.
2. Συνελικτικά επίπεδα (Convolutional layers) — καθορίζονται μόνο από τον αριθμό των καναλιών εισόδου και εξόδου.
3. Επαναλαμβανόμενα επίπεδα (Recurrent layers) — που καθορίζονται από τα κρυφά μεγέθη και τα μεγέθη minibatch.

4.7.2 Λειτουργία με μαθηματική νοοτροπία

Οι GPU έχουν σχεδιαστεί για να αυξάνουν παράλληλα την απόδοση των υπολογισμών. Ωστόσο, αυτό απαιτεί τη φόρτωση και αποθήκευση δεδομένων, πράγμα που σημαίνει ότι η απόδοση μπορεί να περιορίζεται από το εύρος ζώνης ή τη μνήμη. Αυτό συμβαίνει συχνά όταν οι λειτουργίες δεν μπορούν να αναπαρασταθούν από πολλαπλάσια πίνακα, όπως με λειτουργίες ομαδοποίησης, κανονικοποίησης παρτίδας ή ενεργοποίησης. Σε αυτές τις περιπτώσεις, μπορούμε είτε να αυξήσουμε το εύρος ζώνης είτε τη χωρητικότητα της μνήμης. Εναλλακτικά, μπορούμε να δώσουμε προτεραιότητα σε υλοποιήσεις που είναι μαθηματικά συνδεδεμένες. Αυτό σημαίνει ότι η απόδοση περιορίζεται από τον αριθμό των υπολογισμών που μπορούν να εκτελέσουν οι GPU (οι οποίοι μπορούν να αυξηθούν ενεργοποιώντας πυρήνες Tensor).

4.7.3 Επιλογή παραμέτρων για τη μεγιστοποίηση της απόδοσης εκτέλεσης

Λόγω της παράλληλης μορφής επεξεργασίας των GPU, πρέπει να προσεχθεί το πόσο ομοιόμορφα μπορούν να χωριστούν οι παράμετροί. Όσο πιο ομοιόμορφη είναι η διαίρεση, τόσο καλύτερα θα είναι τα κέρδη απόδοσης. Γενικά, θα πρέπει να στοχευθεί κατάλληλα ώστε οι παράμετροι να διαιρούνται με έναν ζυγό αριθμό κάπου μεταξύ 64 και 256. Μπορούν να χρησιμοποιηθούν τιμές μεγαλύτερες από 256, αλλά μόνο με μειωμένα κέρδη. Επιπλέον, η εύρεση του σωστού συνδυασμού περιορισμένων παραμέτρων και υψηλής διαιρετότητας θα δώσει την καλύτερη απόδοση. Αυτό προϋποθέτει ότι οι πράξεις είναι μαθηματικά συνδεδεμένες με υψηλή αριθμητική ένταση (δηλαδή περισσότερες πράξεις κινητής υποδιαστολής από τις προσβάσεις μνήμης).

4.8 Το MobileNet SSD

Τα συνελικτικά νευρωνικά δίκτυα χρησιμοποιούνται για την ανάπτυξη ενός μοντέλου το οποίο αποτελείται από πολλά επίπεδα για την ταξινόμηση των δεδομένων αντικειμένων σε οποιαδήποτε από τις καθορισμένες. Αυτά τα αντικείμενα ανιχνεύονται από χρήση χαρτών δυνατοτήτων υψηλότερης ανάλυσης και είναι δυνατή λόγω της πρόσφατης προόδου στη βαθιά μάθηση με επεξεργασία εικόνας. Ο MobileNet SSD είναι ένα μοντέλο εντοπισμού αντικειμένων που υπολογίζει το πλαίσιο οριοθέτησης εξόδου και την κλάση ενός αντικειμένου από μια εικόνα εισόδου. Αυτό το μοντέλο εντοπισμού αντικειμένων Single Shot Detector (SSD) χρησιμοποιεί το MobileNet ως ραχοκοκαλιά και μπορεί να επιτύχει γρήγορο εντοπισμό αντικειμένων βελτιστοποιημένο για κινητές συσκευές.

4.8.1 Το μοντέλο MobileNet

Το MobileNet είναι μια κατηγορία αποδοτικών μοντέλων που καλούνται για κινητά και ενσωματωμένες εφαρμογές όρασης. Αυτή η κατηγορία μοντέλων βασίζεται σε μια απλοποιημένη αρχιτεκτονική που χρησιμοποιεί διαχωρίσιμες σε βάθος συνελίξεις για τη

δημιουργία ελαφρών βαθιών νευρωνικών δικτύων. Το μοντέλο έχει ως βάση συνελίξεις διαχωρίσιμες που στην ουσία αποτελούν ένα είδος συνελίξεων που έχουν παραγοντοποιηθεί. Αυτό παραγοντοποιεί την συνέλιξη από τυπική σε συνέλιξη βάθους και συνέλιξη 1×1 γνωστή ως Σημειακή Συνέλιξη (pointwise Convolution). Η, συνέλιξη βάθους, στην περίπτωση των MobileNets, για κάθε ένα κανάλι εισόδου, εφαρμόζει μονό φίλτρο. Στη συνέχεια, η κατά σημείο συνέλιξη δημιουργεί μια 1×1 (συνέλιξη) ώστε οι έξοδοι της κατά βάθος συνέλιξης να συνδυαστούν,. Μια κλασσική συνέλιξη έχει ένα μόνο βήμα τόσο για το φιλτράρισμα όσο και για τον συνδυασμό εισόδων σε ένα νέο σύνολο εξόδων. Αλλά η κατά βάθος διαχωρίσιμη συνέλιξη το χωρίζει σε δύο στρώματα, ένα διαφορετικό στρώμα για το φιλτράρισμα, και ένα άλλο επίσης διαφορετικό στρώμα για τον συνδυασμό. Με τη συγκεκριμένη παραγοντοποίηση μειώνεται δραστικά ο υπολογισμός και το μέγεθος του μοντέλου [109]. Τα μοντέλα MobileNet μπορούν να εφαρμοστούν σε διάφορες εργασίες αναγνώρισης για αποτελεσματική ευφυΐα στη συσκευή.

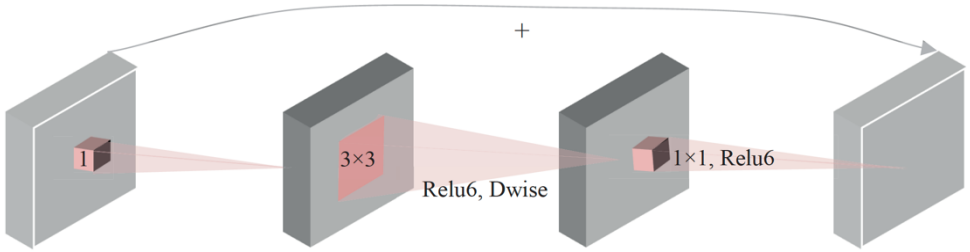
4.8.2 Η τεχνική SSD

Η τεχνική βασίζεται σε ένα συνελικτικό δίκτυο τροφοδοσίας (feed-forward convolutional network) που δημιουργεί μια συλλογή πλαισίων οριοθέτησης σταθερού μεγέθους και βαθμολογίες για την παρουσία εμφανίσεων κλάσης αντικειμένου σε αυτές που ακολουθείται από ένα μη μέγιστο βήμα καταστολής για να παράγουν τις τελικές ανιχνεύσεις[9]. Τα πλαίσια περιέχουν τιμές μετατόπισης (cx,cy,w,h) από το προεπιλεγμένο πλαίσιο. Οι βαθμολογίες περιέχουν τιμές βεβαιότητας για την παρουσία καθεμιάς από τις κατηγορίες αντικειμένων, η τιμή 0 προορίζεται για το υπόβαθρο. Ο SSD εισάγει πολυ-αναφορικές και πολυ-αναλυτικές τεχνικές ανίχνευσης. Οι τεχνικές πολλαπλής αναφοράς ορίζουν ένα σύνολο κουτιών αγκύρωσης διαφορετικών μεγεθών και αναλογιών διαστάσεων σε διαφορετικές θέσεις μιας εικόνας και, στη συνέχεια, προβλέπει τον εντοπισμό με βάση αυτές τις αναφορές. Οι τεχνικές πολλαπλής ανάλυσης επιτρέπουν την ανίχνευση αντικειμένων σε διάφορες κλίμακες και σε διαφορετικά στρώματα του δικτύου. Ένα δίκτυο SSD υλοποιεί έναν αλγόριθμο για τον εντοπισμό πολλών κλάσεων αντικειμένων στις εικόνες μέσω δημιουργίας βαθμολογιών εμπιστοσύνης / βεβαιότητας που σχετίζονται με την παρουσία οποιασδήποτε κατηγορίας αντικειμένων στο κάθε προεπιλεγμένο πλαίσιο.

Παράγει επίσης προσαρμογές σε πλαίσια για καλύτερη αντιστοίχιση των σχημάτων των αντικειμένων. Αυτό το δίκτυο είναι κατάλληλο για εφαρμογές σε πραγματικό χρόνο, καθώς δεν πραγματοποιεί αναδειγματοληψίες των χαρακτηριστικών για τις υποθέσεις οριοθέτησης πλαισίου (bounding box). Ο SSD βασίζεται στο CNN και για την ανίχνευση των κλάσεων-στόχων των αντικειμένων και ακολουθεί δύο στάδια: (1) εξαγωγή των “feature maps”, και (2) εφαρμογή φίλτρων συνελίξεων για την ανίχνευση των αντικειμένων. Ο SSD χρησιμοποιεί το VGG16 για την εξαγωγή “feature maps”. Στη συνέχεια, ανιχνεύει αντικείμενα χρησιμοποιώντας το στρώμα Conv4_3 του VGG16. Κάθε πρόβλεψη αποτελείται από ένα πλαίσιο οριοθέτησης και 21 βαθμολογίες για κάθε τάξη (μία επιπλέον κατηγορία χωρίς αντικείμενο) και η κατηγορία με την υψηλότερη βαθμολογία επιλέγεται ως η κατάλληλη για να παραστεί ως το οριοθετημένο αντικείμενο [110]. Ο μεγάλος στόχος κατά τη διάρκεια της εκπαίδευσης είναι η απόκτηση υψηλής θέσης βαθμολογίας εμπιστοσύνης και αυτό μπορεί να επιτευχθεί με την αντιστοίχιση των προεπιλεγμένων πλαισίων με τα αληθινά πλαίσια.

4.8.3 Το Δίκτυο MobileNet-V2

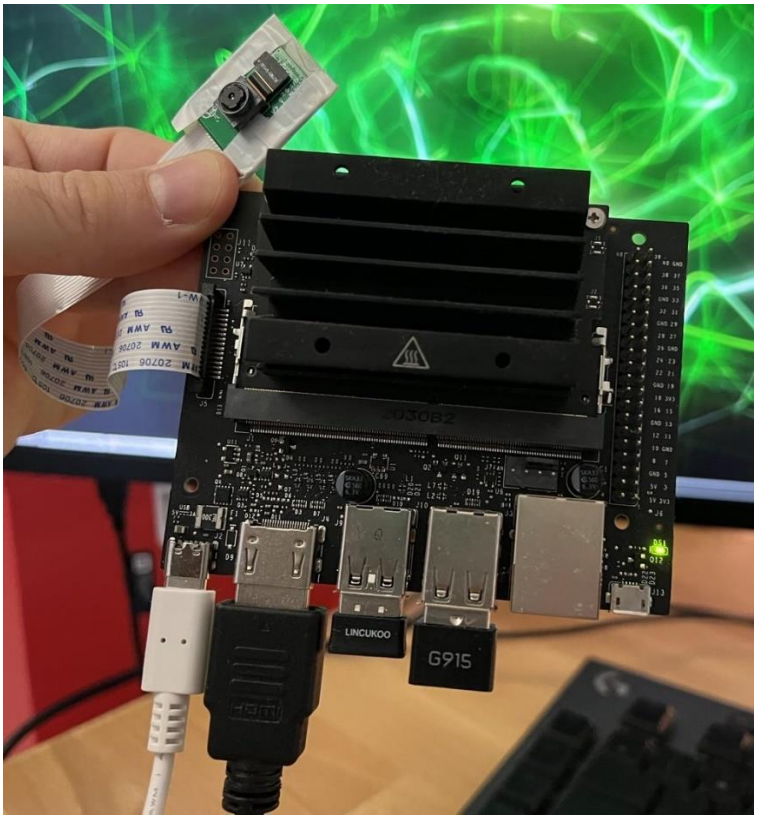
Το παραδοσιακό συνελικτικό νευρωνικό δίκτυο έχει μεγάλες απαιτήσεις μνήμης και μεγάλους υπολογισμός, ώστε να μπορεί να χρησιμοποιείται σε κινητές συσκευές ή ενσωματωμένες συσκευές. Η Google πρότεινε MobileNet-V1 [111], MobileNet - V2 [112] , MobileNet-V3 [113] και άλλες δομές δικτύου. Σε σύγκριση με το συνηθισμένο συνελικτικό νευρωνικό δίκτυο, οι παράμετροι του μοντέλου και του υπολογισμού στο δίκτυο του MobileNet μειώνονται σημαντικά, αλλά η ακρίβεια του δεν μειώνεται πολύ. Το δίκτυο MobileNet-V2 εισάγει μια ανεστραμμένη υπολειμματική δομή (όπως φαίνεται στο Σχήμα 17) και γραμμική συμφόρηση. Η προς τα πίσω υπολειμματική δομή χωρίζεται σε τρία βήματα: πρώτα, 1×1 συνέλιξη βελτίωσης διάστασης, στη συνέχεια 3×3 βαθιά συνέλιξη και τέλος 1×1 συνέλιξη μείωσης διαστάσεων. Προκειμένου να αποφευχθεί η απώλεια πληροφοριών που προκαλείται από μη γραμμικό μετασχηματισμό, ο μη γραμμικός μετασχηματισμός είναι προσεγγισμένος ως γραμμικός μετασχηματισμός με στρώση γραμμικής συμφόρησης.



Σχήμα 17. Ανάστροφη υπολειμματική δομή MobileNet-V2 [114]

4.9 Στήνοντας το Jetson Nano

Πρέπει πρώτα να βεβαιωθούμε ότι φέρουμε τα περιφερειακά που απαιτούνται για τη ρύθμιση του Nvidia Jetson Nano (Εικόνα 1). Οι επιλογές είναι είτε με οθόνη δική του, δλδ. σαν desktop, είτε «headless», που σημαίνει χωρίς περιφερικά εκτός από σύνδεση με το δίκτυο (σε αυτή τη περίπτωση οι εντολές και οι διαδικασίες εκτελούνται από απομακρυσμένο υπολογιστή). Εμείς επιλέξαμε τη πρώτη επιλογή η οποία είναι καλύτερη και πιο διαδραστική με το ίδιο το Jetson. Τα περιφερειακά, λοιπόν, που απαιτούνται για να λειτουργήσει είναι το ποντίκι, το πληκτρολόγιο, το καλώδιο HDMI, κάμερα CSI (στη δική μας περίπτωση), οθόνη, σύνδεση στο διαδίκτυο με είτε WiFi dongle είτε με ethernet, καλώδιο τροφοδοσίας, και τον σκληρό μας δίσκο που είναι σε μορφή κάρτας microSD. Προτιμάται κάρτα τουλάχιστον 64GB microSD με ταχύτητα διασύνδεσης διαύλου τουλάχιστον 100MB/s. Αυτή η σύσταση οφείλεται σε περίπτωση εκμάθησης μεγάλου συνόλου δεδομένων.



Εικόνα 1.

4.9.1 Βασικές συστάσεις έναρξης

Είναι πολύ σημαντικό να προσέξουμε μερικά πράγματα πριν ξεκινήσουμε. Το ένα είναι η τροφοδοσία η οποία μπορεί να είναι είτε 5v είτε 15v σε power mode. Αυτό παίζει πολύ σημαντικό ρολό στο τι μπορεί να σηκώσει το σύστημα. Το άλλο είναι η δημιουργία χώρου SWAP που επιτρέπει στη μνήμη ram (που είναι μόνο 2GB) να δανείζεται μερικούς πόρους (ανάλογα με το πόσο θα του ζητήσεις) από την microSD. Αυτό βέβαια είναι μερικώς ρίσκο, καθότι εκθέτει τον δίσκο σε μεγαλύτερες φθορές, όπως πράγματι μου συνέβη μια φορά και καταστράφηκε από υπερφόρτωση. Κάτι που έσβησε όλα τα μέχρι τότε δεδομένα, την εκπαίδευση και τα αποτελέσματα της. Οπότε το συνετό είναι να κάνει κάποιος Swap 4 GB, σε κάρτα από 64 GB και πάνω. Όσον αφορά το terminal, πρέπει να βεβαιωθούμε ότι έχουμε κάνει όλα τα updates και upgrades πριν συνεχίσουμε με οποιαδήποτε άλλα πακέτα, dockers, βιβλιοθήκες κ.α. Πρέπει να δοθεί μεγάλη προσοχή σε συμβατότητες. Κάθε βιβλιοθήκη, εργαλείο, πλαίσιο, γλώσσα κοκ, πρέπει να έχει συμβατότητα με το αντίστοιχο περιβάλλον που θέλουμε να συνεργαστούμε.

4.10 Προετοιμασία της κάρτας SD

Το πρώτο πράγμα που πρέπει να κάνουμε είναι να προετοιμάσουμε το λειτουργικό σύστημα. Το Jetson Nano χρησιμοποιεί μια κάρτα microSD για την αποθήκευση του λειτουργικού συστήματος.. Δεδομένου, λοιπόν, ότι το λειτουργικό σύστημα θα λειτουργεί στην κάρτα microSD, η επιλογή της κάρτας είναι σημαντική. Πρέπει να επιλεγεί κάρτα υψηλής ποιότητας και χαρακτηριστικών, για να υπάρξει μια αξιοπρεπής ταχύτητα γραφής/ανάγνωσης χωρίς προβλήματα. Το host και κύριο σύστημα μας είναι macOS, οπότε από εδώ θα γράψουμε το Image της NVIDIA στη microSD. Σε αυτό το βήμα απαιτείται από τον κεντρικό μας υπολογιστή, σύνδεση στο διαδίκτυο. Μπορούμε να κατεβάσουμε ένα γραφικό πρόγραμμα για

να γίνει η εγγραφή ή μπορεί να γίνει ακόμα και από το τερματικό. Στη προκείμενη περίπτωση κατεβάσαμε το Etcher που κάνει το flash πολύ εύκολο και γρήγορο. Όταν ολοκληρώθηκε η αναδιαμόρφωση του δίσκου, τότε είμασταν έτοιμοι να την εισάγουμε στο Jetson. Μετα από πολύ γρήγορες ρυθμίσεις, ανοίγει στο λειτουργικό Linux (Ubuntu).

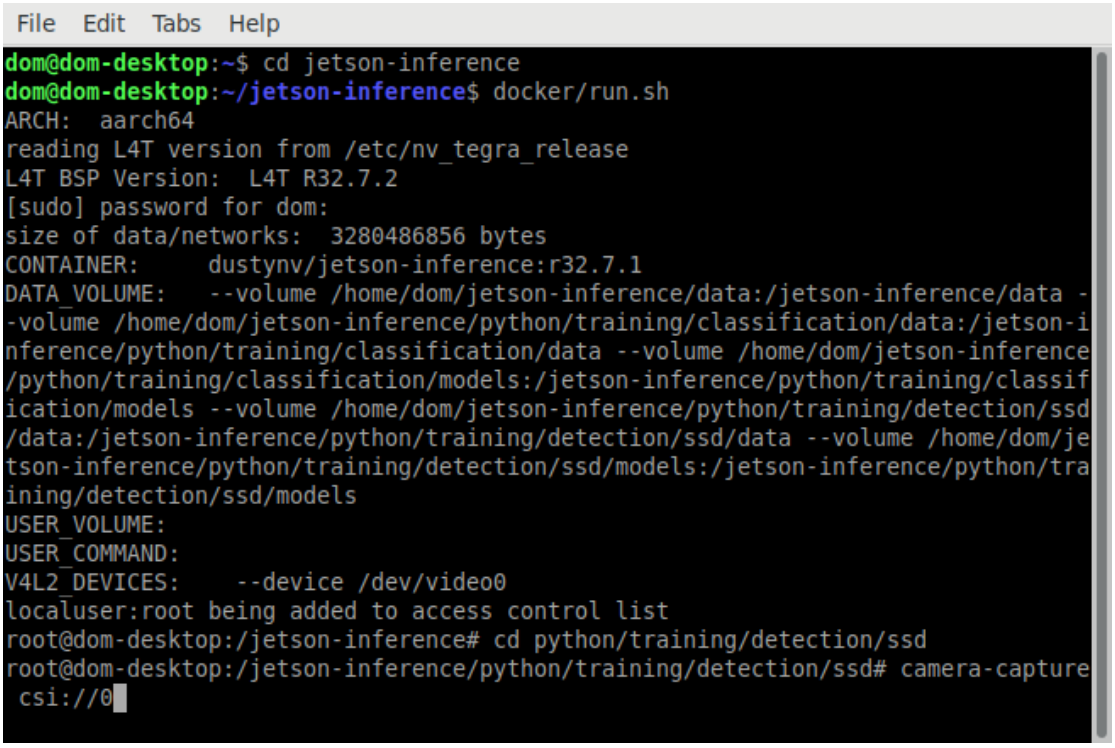
4.11 Σετ δεδομένων

Για τη διπλωματική αυτή εργασία δημιουργήθηκαν 2 σετ δεδομένων. Το ένα αφορά 1000 φωτογραφίες που ανήκουν σε 10 κολάσεις. Κάθε κλάση έχει 100 φωτογραφίες. Κατά τη λήψη τους οροθετήθηκαν οι σημάνσεις με πλαίσιο εικόνας (Bounding Box). Οι σχολιασμοί δημιουργήθηκαν υπο τη μορφή XML. Το δεύτερο σετ δεδομένων δημιουργήθηκε λόγω μη ικανοποιητικών αποτελεσμάτων του πρώτου. Περιέχει 10000 φωτογραφίες με 30 κλάσεις. Γίνεται χρήση των 1000 ήδη υπαρχουσών φωτογραφιών με τη προσθήκη επιπλέον 4000 ακόμα (λήφθηκες μέσω της κάμερας και οριοθετημένες μέσω του εργαλείου της *camera-capture* λειτουργίας του SDK) και άλλες 5000 που ανασυρθήκαν από dataset στο διαδίκτυο. Το dataset αυτό περιείχε 10000 εικόνες και 29 κολάσεις, όμως οι εικόνες του ήταν πολύ μικρές, πολλές ήταν σχεδόν μη αξιοποιήσιμες και δεν είχαν φόντο. Οπότε έτσι προέκυψε ο συγκερασμός, 5000 από αυτές και 5000 δικών μας, με φόντο πραγματικών οδικών συνθήκων. Οι αποκτήθηκες έξωθεν φωτογραφίες ήταν σχολιασμένες αλλά σε άλλο φόρμα, πραγματοποιήθηκε μετατροπή από TXT σε XML και έγινε η ετικετοποίηση. Ο χωρισμός του Dataset έγινε στη, πρώτη περίπτωση, σε 80% για το σετ εκπαίδευσης, 10% για το σετ δοκιμής και 10% για το σετ επικύρωσης. Στη δεύτερη περίπτωσης το χώρισμα έγινε σε 75-15-10 αντίστοιχα.

4.12 Διαδικασία συλλογής δεδομένων – (Δημιουργία Dataset)

Αφού μπούμε στον φάκελο jetson-inference από οπού μπορούμε συνδεθούμε στο docker της NVidia, μετα μπορούμε με μια εντολή (Εικόνα 2) να ενεργοποιήσουμε το εργαλείο “data capture control”

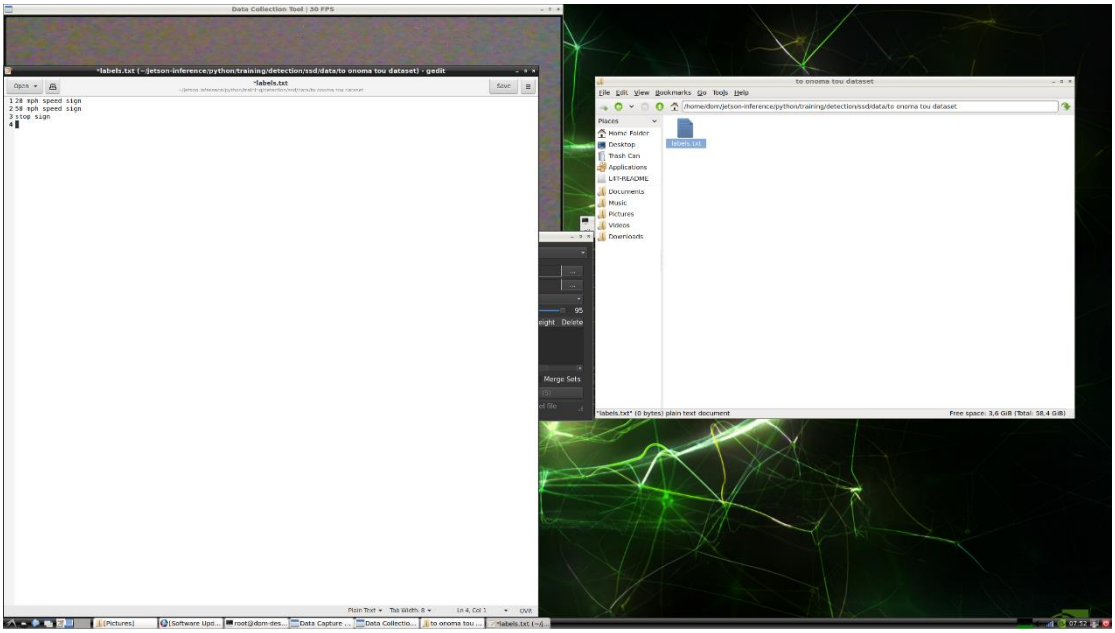
Πριν όμως κάνουμε αυτό, εντός του jetson-inference, επιλέγουμε φάκελο «data» (ο οποίος βρίσκεται στο /jetson-inference/python/training/detection/ssd/data), και εντός του δημιουργούμε έναν φάκελο με το όνομα που θέλουμε να χρησιμοποιήσουμε για το dataset. Εντός αυτού του φακέλου, έπειτα, δημιουργούμε ένα αρχείο με την ονομασία labels και του προσδίδουμε τον χαρακτήρα .txt δηλαδή «labels.txt». Εντός του αρχείου labels, καταχωρούμε τις κολάσεις μας, όπως φαίνεται και στο παράδειγμα της Εικόνας 3.



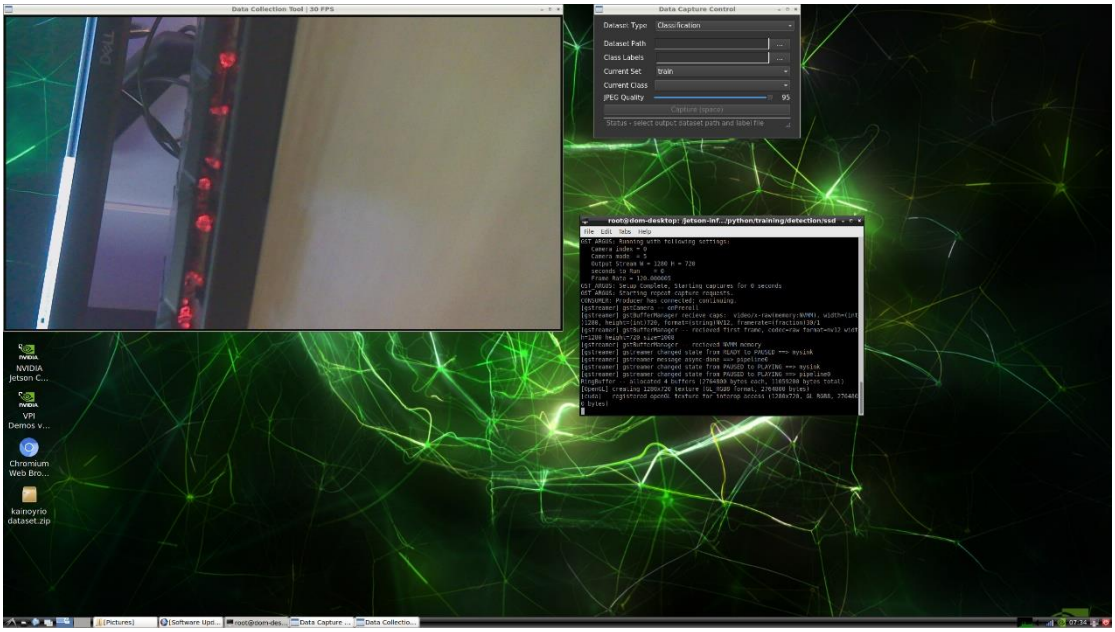
Εικόνα 2.

Η εφαρμογή αυτή είναι απαραίτητη για τη συλλογή σετ δεδομένων, βέβαια δίνεται η δυνατότητα εισαγωγής dataset είτε δικού μας από άλλον υπολογιστή, είτε δημοσίου dataset, η διαδικασία όμως και οι μετατροπές που χρειάζονται για να λειτουργήσουν σωστά σε αυτό το σύστημα θέλουν εξοικείωση και γνώση χειρισμού. Εμείς αφιερώσαμε τον χρόνο να δημιουργήσουμε έτσι το dataset ώστε να αποκτήσουμε μια συνολική εμπειρία του περιβάλλοντος του jetson καθώς και να προσέξουμε πιο πολύ τα δεδομένα με τα οποία θα το τροφοδοτήσουμε. Ήδη οι 10000 εικόνες είναι πολύ μεγάλος φόρτος για ένα Nano και είναι γνωστό πως τόσο μεγάλο σετ δεν μπορεί να τρέξει ευκολά πάνω από 50-100 epochs.

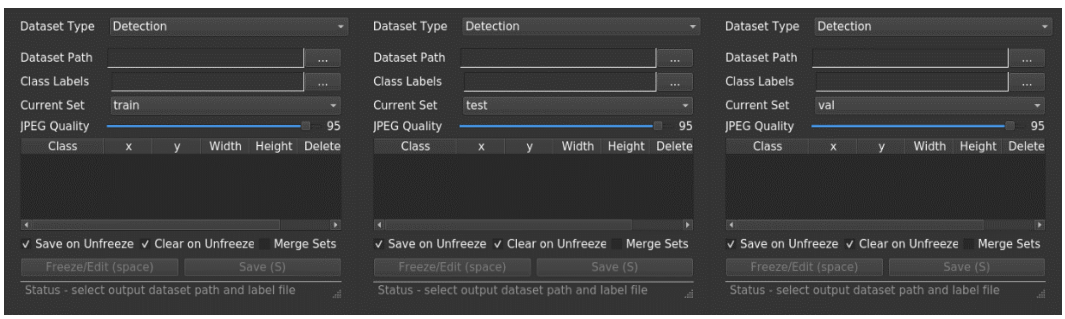
Εισερχόμενοι στην εφαρμογή, μας επιτρέπει να διαλέξουμε τον τύπο του dataset (Classification / Detection). Όπως βλέπουμε βρίσκεται στην ταξινόμηση (Εικόνα 4). Εμείς επιλέγουμε το Detection.



Εικόνα 3.

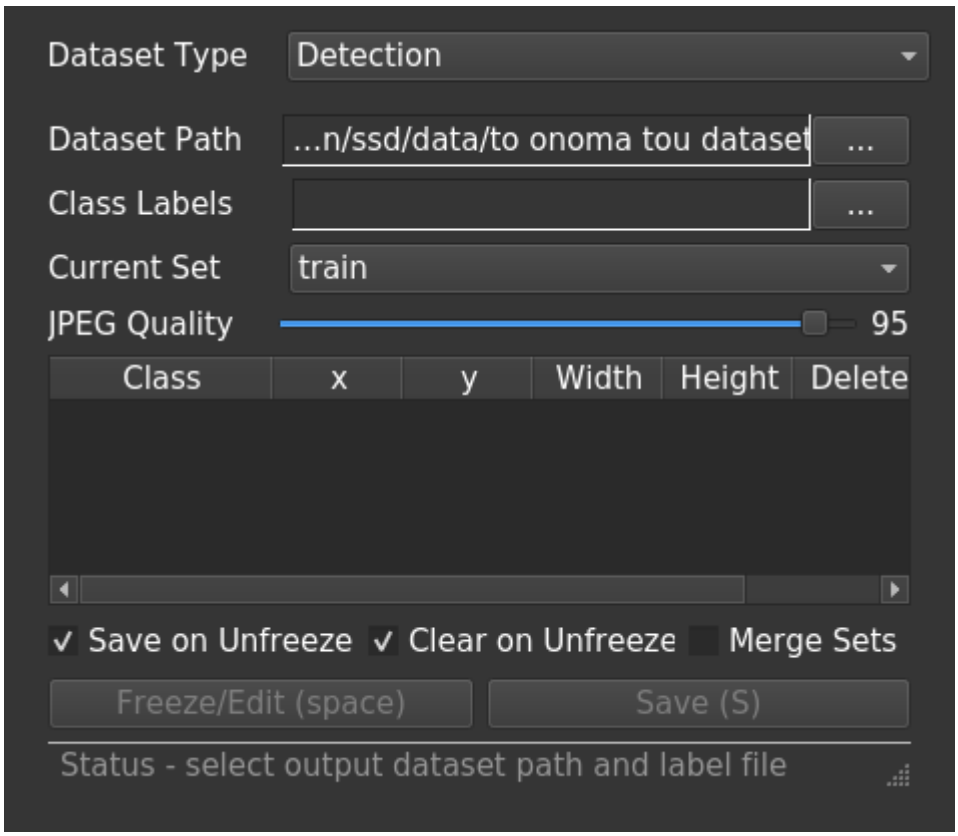


Εικόνα 4.



Εικόνα 5.

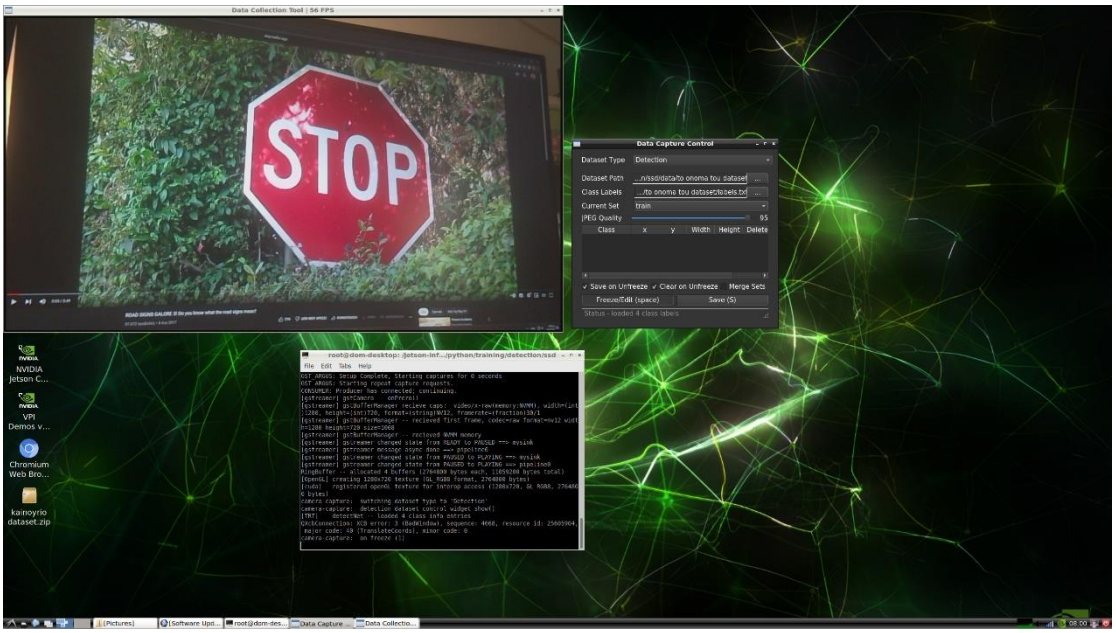
Αφού λοιπόν κάνουμε την επιλογή μας. Μας δίνει τη δυνατότητα στο σημείο «Current set» να επιλέξουμε σε ποιο κατάλογο θα αποθηκευτούν οι εικόνες που θα συλλέξουμε. στις εικόνες (Εικόνα 5) φαίνεται να έχουμε επιλέξει και τις 3 επιλογές (train, test, val). Επιπλέον δίνεται η επιλογή του παγώματος εικόνας «Freeze/Edit» και του «Save», οπού στην πρώτη περίπτωση παγώνει την κάμερα στο σημείο, ώστε να μπορεί να γίνει επεξεργασία της εικόνας (οριοθέτηση πλαισίου, χειροκίνητα με τον κέρσορα), ενώ στη δεύτερη αποθηκεύει την εικόνα μας. Ως επιλογές επίσης διατίθενται και οι «Save on Unfreeze», «Clear on Unfreeze» που κάνουν αυτό ακριβώς που λένε (για διευκόλυνση ίσως μερικών που κάνουν τη διαδικασία πιο ευκολά και γρηγορά), και η «Merge Sets» η οποία αποτελεί σημαντική επιλογή. Η «Merge Sets» σου δίνει τη δυνατότητα, λαμβάνοντας και αφού επεξεργαστείς την εικόνα, σώζοντας τη, να αποθηκευτεί στους καταλόγους και των τριών μας κατηγοριών (test, train, val) κάτι που δεν συνίσταται σε μεγάλα σετ δεδομένων. Εμείς ακολουθήσαμε τη διαδικασία εισαγωγής εικόνων ξεχωριστά σε κάθε κατηγορία. (7500 train, 1500 test, 1000 val). Τέλος, η επιλογή «JPEG Quality» είναι μια πολύ ιδιαίτερη λεπτομέρεια , διότι σου επιτρέπει σε περίπτωση που έχουμε ξεκάθαρα πλανά, ή κοντινά, του αντικειμένου που μας ενδιαφέρει, τότε μειώνει τον όγκο του σετ δεδομένων ρίχνοντας τη ποιότητα και συνεπώς ελαφρύνει την υπολογιστική ισχύ που θα απαιτηθεί. Στο «Dataset Path» (Εικόνα 6) πρέπει να εισαχθεί το μονοπάτι του φακέλου που δημιουργήσαμε για το συγκεκριμένο dataset και στο «Class Labels» εισάγουμε το μονοπάτι για την τοποθεσια εντός του φακέλου dataset, στο αρχείο «labels.txt».



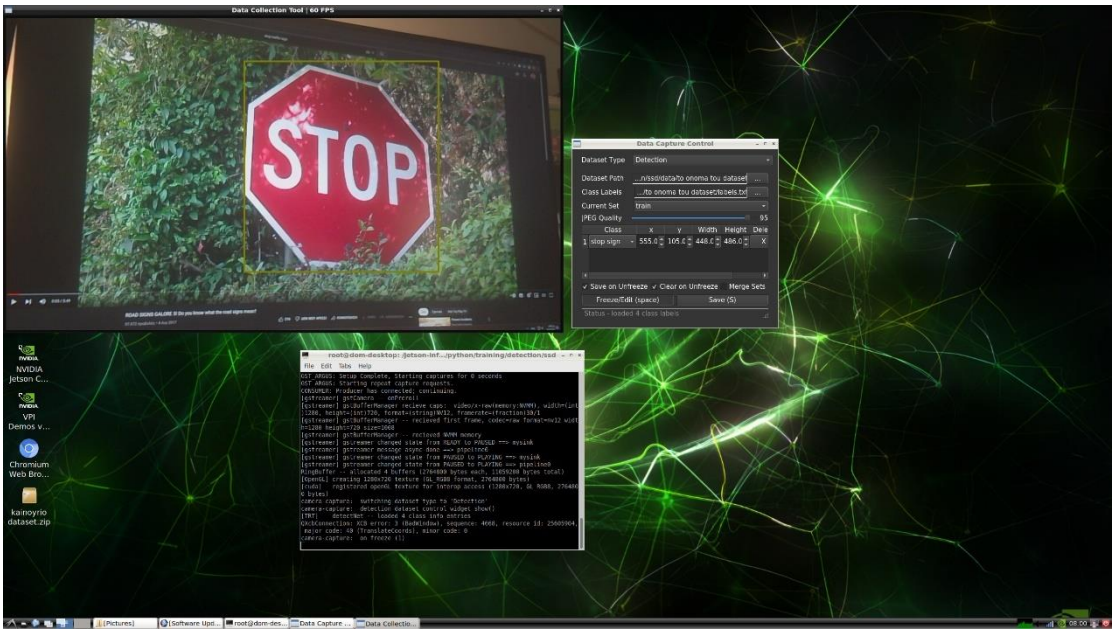
Εικόνα 6.

Τώρα είμαστε έτοιμοι λοιπόν να διαλέξουμε υποψήφιες εικόνες μέσω της κάμερας μας. Σε περίπτωση που χρειαζόμαστε στιγμιότυπο από βίντεο, όπως στην Εικόνα 7, πατάμε το «freeze» αφού εντοπίσουμε στο κάδρο το στοιχείο που μας ενδιαφέρει και στη συνέχεια αφού

το παγώσει, οριοθετούμε με ακρίβεια το στοιχείο εντός του πλαισίου που εμείς οι ίδιοι δημιουργούμε με τον κέρσορα μας, όπως φαίνεται στην Εικόνα 8.



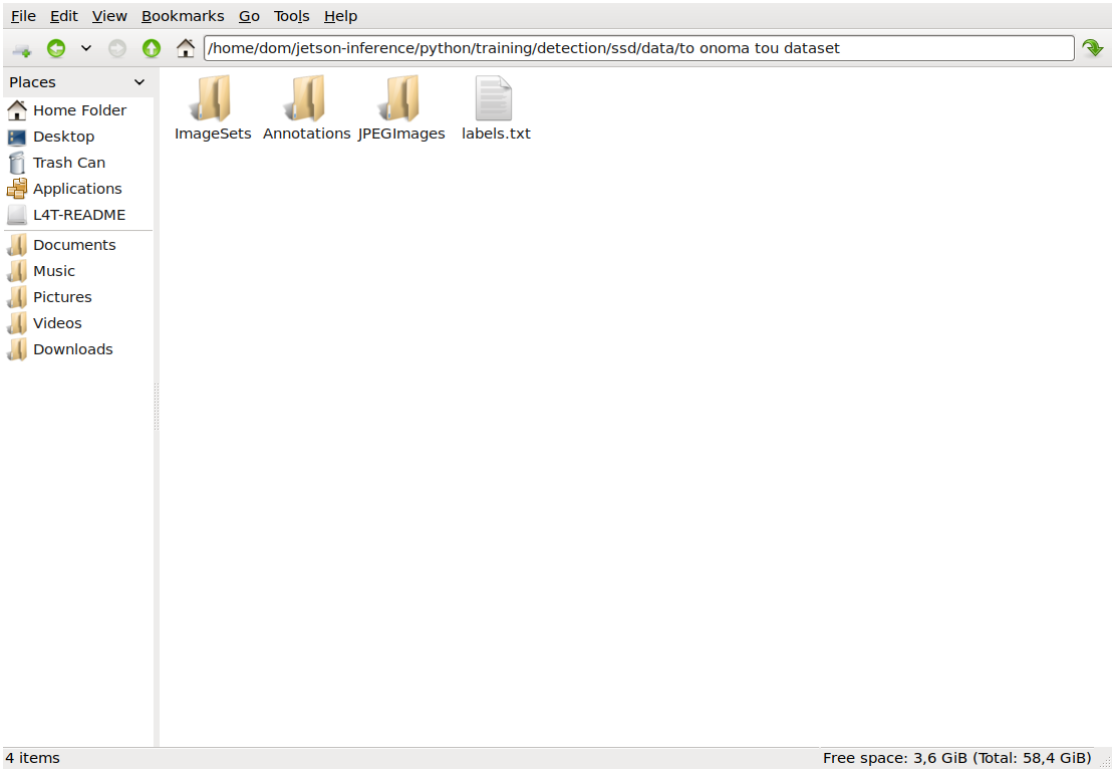
Εικόνα 7.



Εικόνα 8.

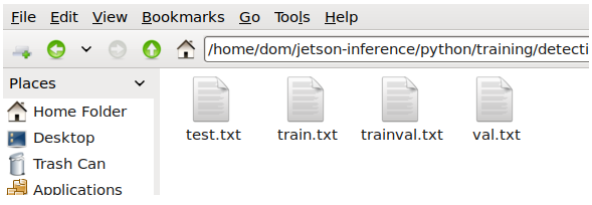
Όπως βλέπουμε, με το που δημιουργήσαμε το πλαίσιο, εμφανίστηκε η επιλογή της κλάσης. Στη συγκεκριμένη περίπτωση, βρίσκεται στην σωστή («stop sign»), σε άλλη περίπτωση μπορούμε εμείς να επιλέξουμε πατώντας την σωστη κλάση, καθώς υπάρχει λίστα εντός αντίστοιχη με τις κολάσεις που εισήγαμε στο labels.txt μιας και από εκεί του έχουμε ορίσει το αντίστοιχο path.

Με όλη αυτή την διαδικασία, από την πρώτη κιόλας εικόνα που εισάγουμε στο dataset μας με τον τρόπο αυτό, αυτόματα δημιουργούνται τα απαραίτητα αρχεία εντός του φακέλου με το dataset μας, όπως φαίνεται στην Εικόνα 9.

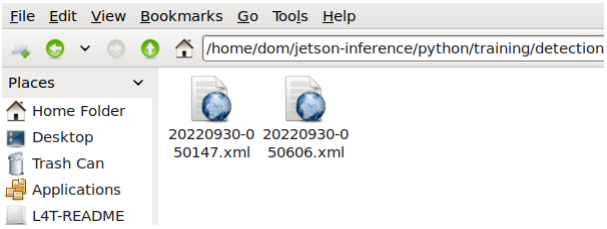


Εικόνα 9.

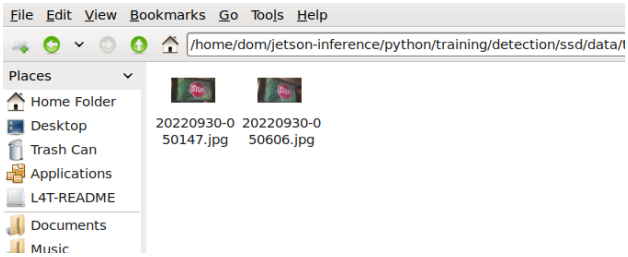
Εντός του φακέλου έχουν δημιουργηθεί τρεις φάκελοι «ImageSets», «Annotations» και «JPEGImages». Στον πρώτο φάκελο περιέχονται τα αρχεία «test.txt» «train.txt» «trainval.txt» «val.txt» (Εικόνα 10). Στον δεύτερο φάκελο αποθηκεύονται οι σχολιασμοί σε μορφή .xml (Εικόνα 11). Στον τελευταίο αποθηκεύονται οι εικόνες που επιλέξαμε (Εικόνα 12). Τα δεδομένα των σχολιασμών φαίνονται όπως στην Εικόνα 13.



Εικόνα 10.



Εικόνα 11.



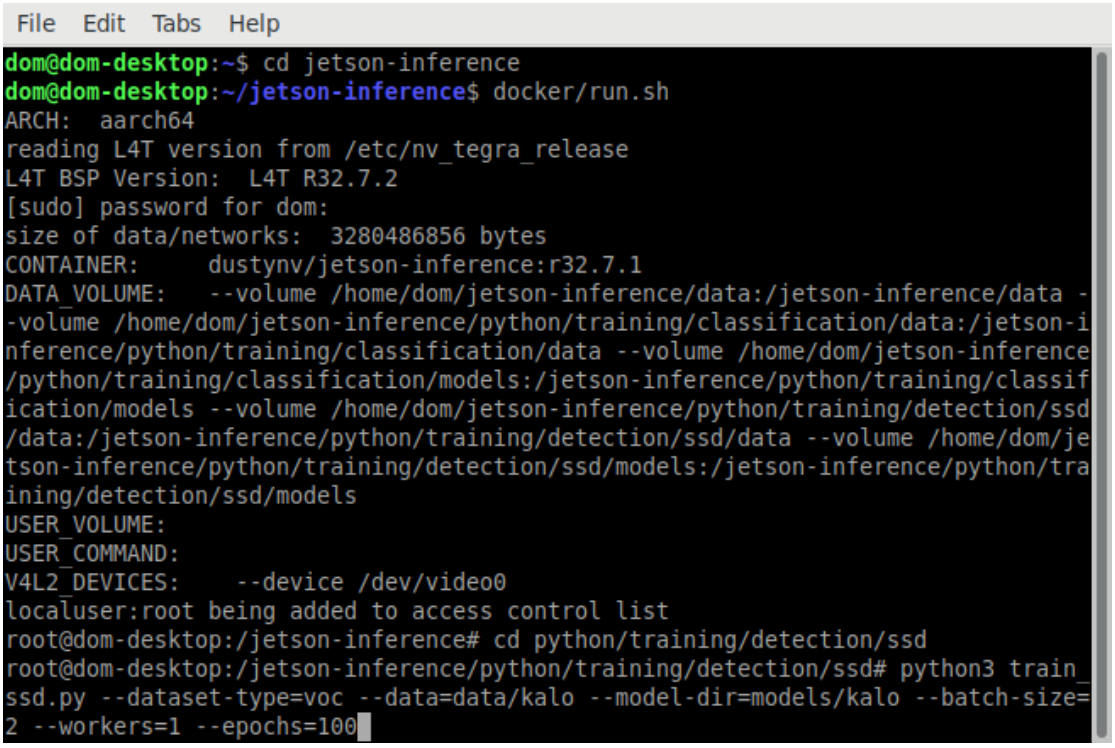
Εικόνα 12.

Εικόνα 13.

(Σημείωση: οι εικόνες 8, 9 και 10 αφορούν αναπαράσταση της διαδικασίας αρχειοθέτησης του dataset προς ευκολία επεξήγησης και δεν αποτελεί τα στοιχεία του πλήρους dataset μας.)

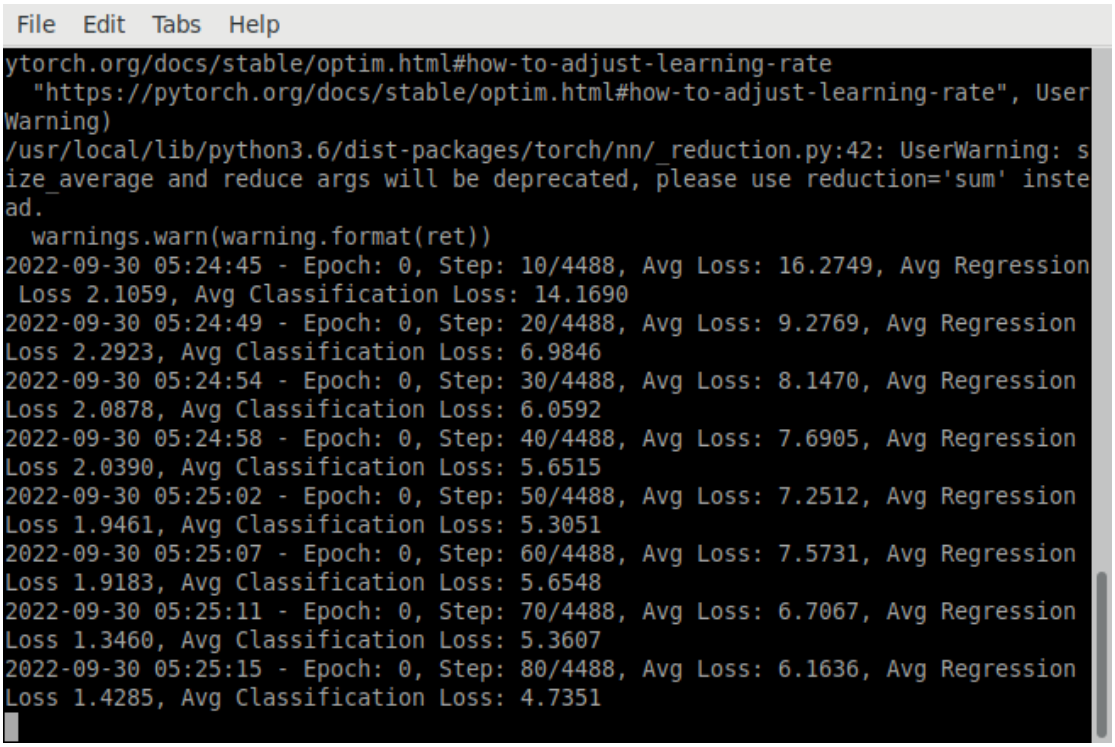
4.13 Διαδικασία εκπαίδευσης του μοντέλου

Αφού συλλέξαμε το dataset μας, το εκπαιδεύσαμε πολλές φορές μέχρι να βρούμε την χρυσή τομή αντοχών/αποδοτικότητας του jetson και της σωστής διαχείρισης/ επεξεργασίας του σετ δεδομένων μας. Με τον κώδικα στην εικόνα 14, εκπαιδεύουμε χρησιμοποιώντας το train_ssd.py και λόγω του μεγάλου φόρτου, ορίζουμε το --batch-size σε 2 και –workers σε 1 (--batch-size (default 4) και --workers (default 2)). Δοκιμάσαμε και σε μικρότερο dataset και σε μεγαλύτερο να ορίσουμε διαφορετικές τιμές ή έστω και τις default τους , αλλά κατέστη μάταιο και το σύστημα δεν ανταποκρινόταν. Σε αυτή την περίπτωση αξίζει να αναφερθεί η σημαντικότητα του Mounting SWAP και την απενεργοποίηση του Desktop GUI, για την αύξηση των πιθανοτήτων λογιότερων προβλημάτων και καλύτερης και σταθερότερης υπολογιστικής ταχύτητας.



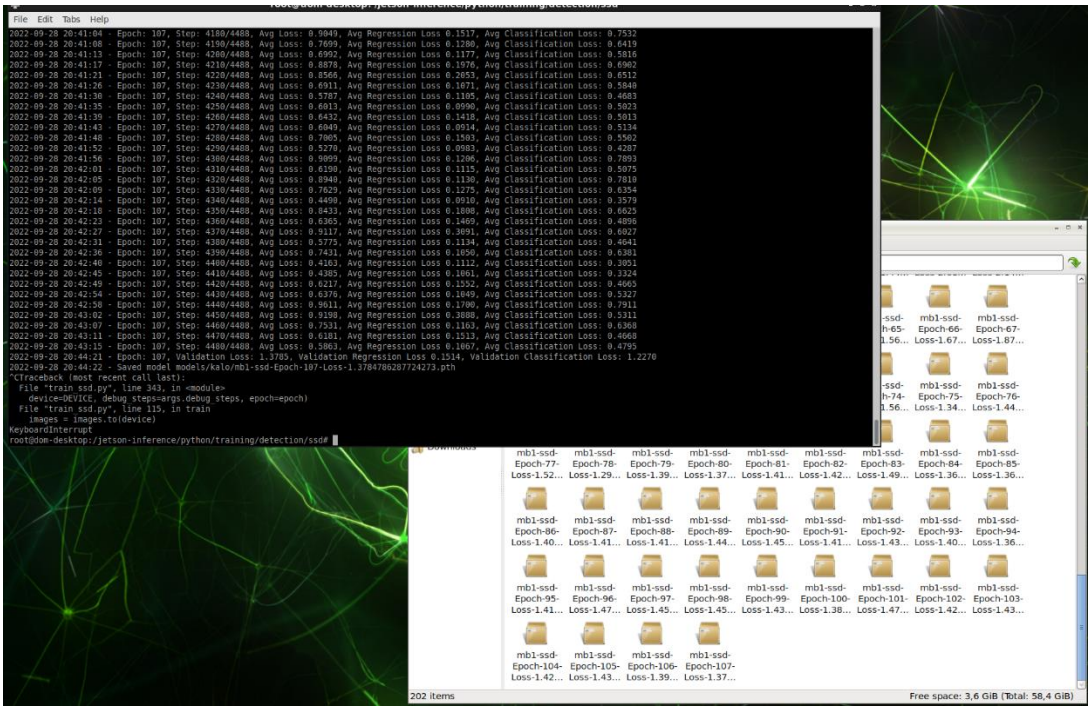
Εικόνα 14.

Κάπως έτσι λοιπόν ξεκινά μια προσπάθεια εκπαίδευσης με τιμές απωλειών αρκετά υψηλές (Εικόνα 15)



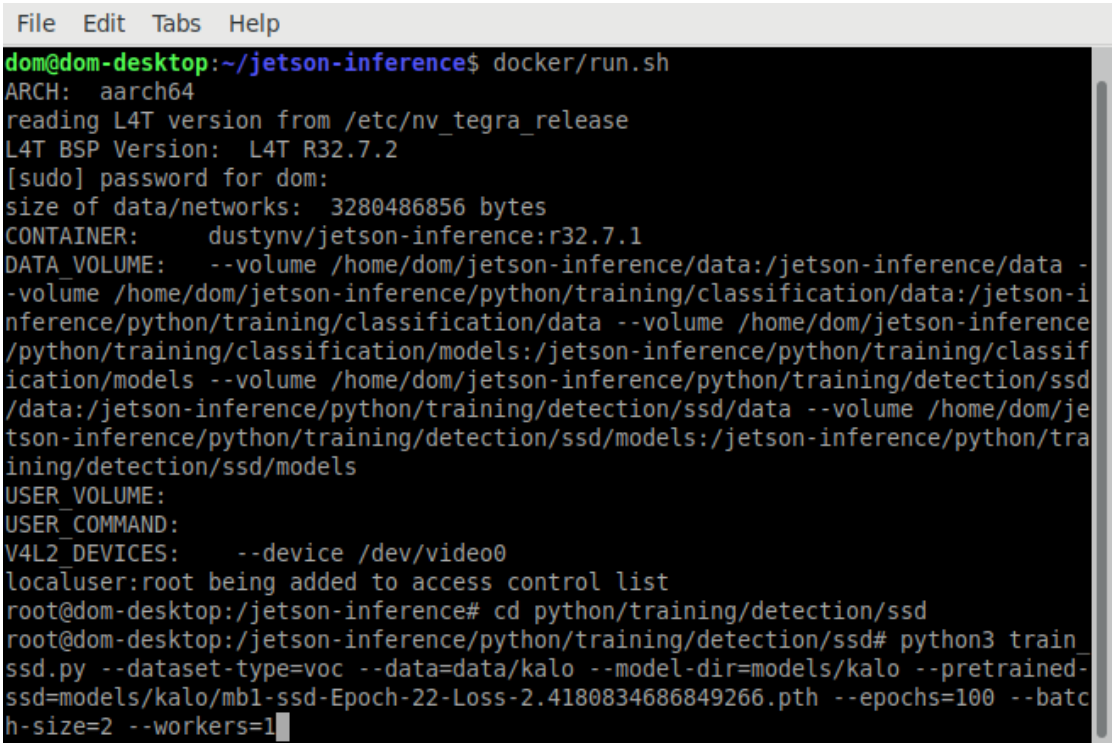
Εικόνα 15.

Και στην περίπτωση μας, μετα από πολλές προσπάθειες βελτιστοποίησης και αναπροσαρμογής καταλήξαμε στο καλύτερο μοντέλο μας το οποίο έφτασε μέχρι την 107 epoch (από την 67η και μετα σταμάτησε να βελτιώνεται και παρέμεινε στα ιδιά μεγέθη απωλειών)(Εικόνα 16).



Εικόνα 16.

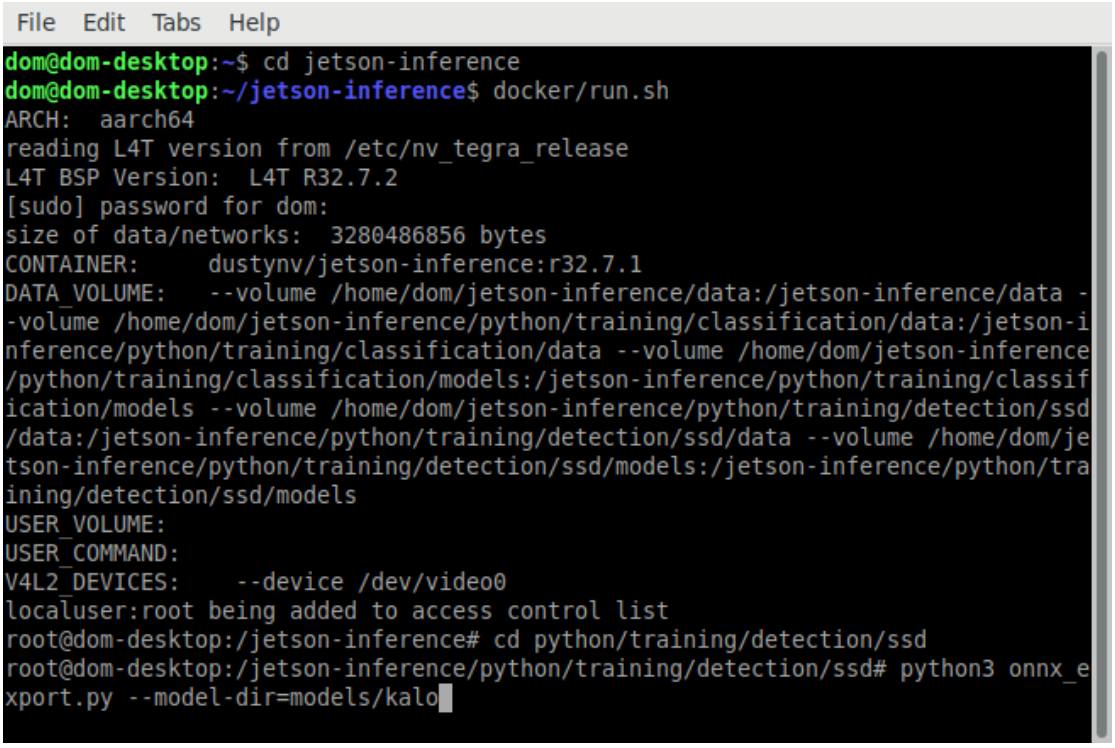
Όταν ολοκληρωθεί η εκπαίδευση, υπάρχει η δυνατότητα επανεκπαίδευσης, σε περίπτωση που το αποτέλεσμα οπτικοποίηση δεν είναι ικανοποιητικό. Όμως σε αυτή την περίπτωση παρόλο που η διαδικασία ξεκινά από το καλύτερο βρεγμένο validation loss, η διαδικασία της εκπαίδευσης ως -pretrained, δεν είναι απαραίτητο ούτε σίγουρο πως θα υπάρξουν γρηγορότερα και καλυτέρα αποτελέσματα, καθώς όλο αυτό εξαρτάται καθαρά από το μοντέλο και την δομή του. Στην περίπτωση μας, χρειάστηκε αρκετές φορές να επανεκτιμήσουμε την εκπαίδευση (λόγω διακοπής ρεύματος -κολλήματος του προγράμματος, ή πολύ αργής ανάπτυξης epochs από κάποιο σημείο και μετα, κ.α.). Η διαδικασία που ακολουθείται εμφανίζεται στην Εικόνα 17.



Εικόνα 17.

4.14 Μετατροπή από PyTorch σε ONNX

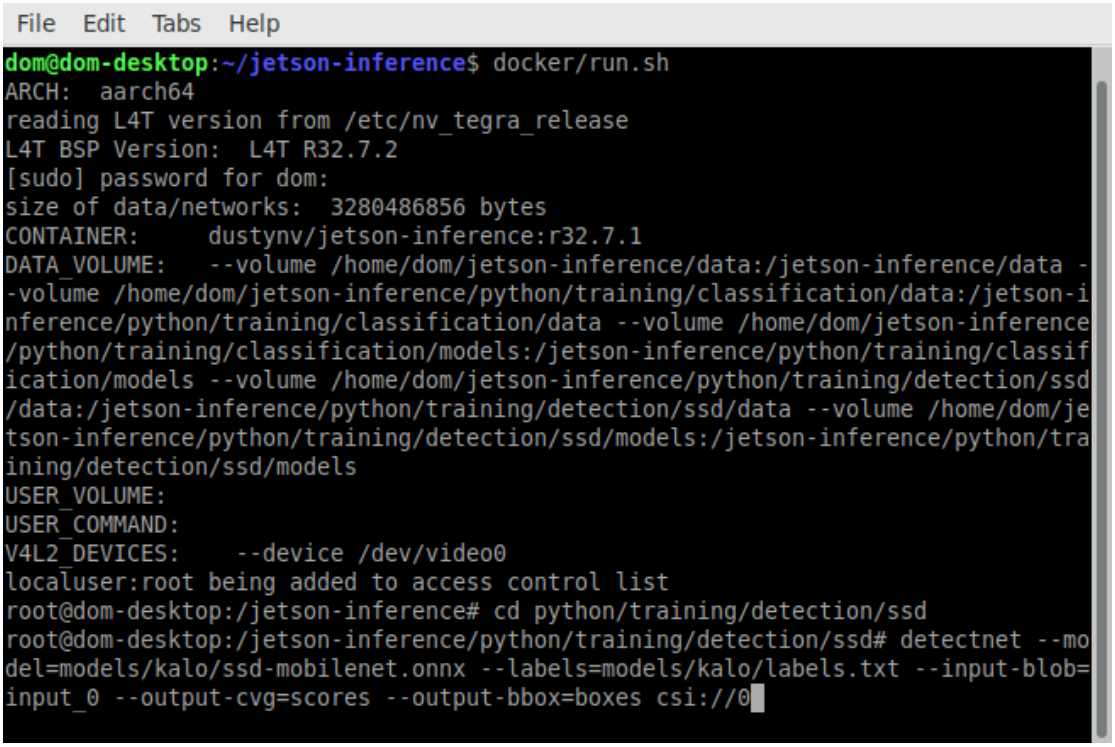
Όταν λοιπόν εκπαιδευτεί το μοντέλο και έχουμε στα χεριά μας ένα ικανοποιητικό αποτέλεσμα απωλειών, περνάμε στο επόμενο βήμα το οποίο είναι η μετατροπή του μοντέλου μας από porch σε ONNX (Εικόνα 18). Το ONNX πλέον μοντέλο αποθηκεύεται ως /ssd-mobilenet.onnx το οποίο μετα μπορούμε να φορτώσουμε με το detectnet.



Εικόνα 18.

4.15 Εκτέλεση στο detectnet

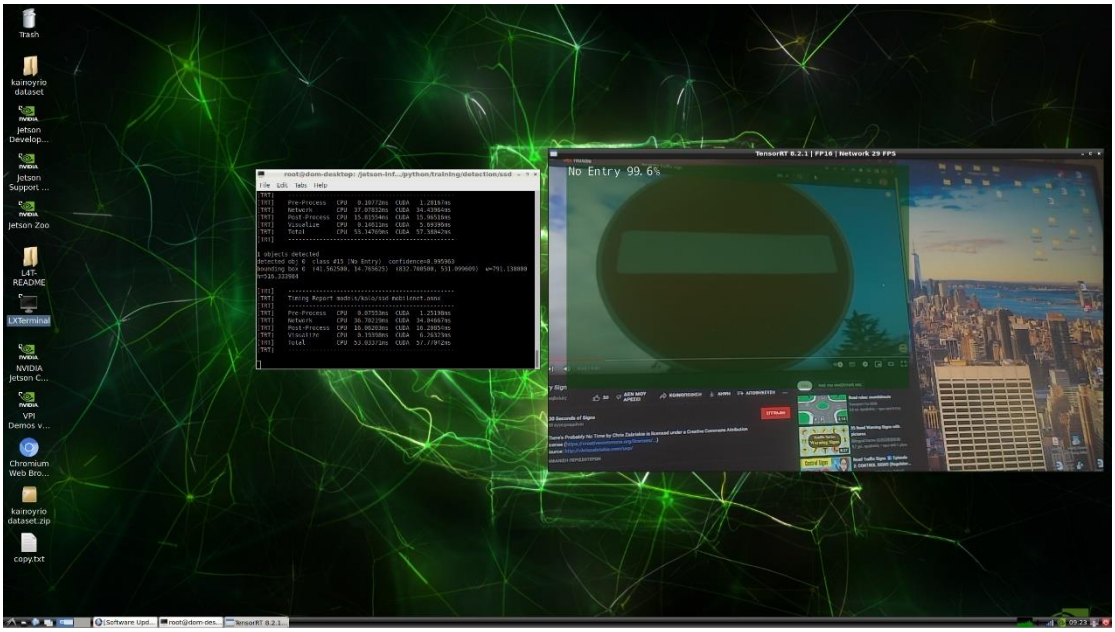
Η διαδικασία εισαγωγής στο detectnet και των εργαλείων του φαίνεται στην εικόνα 19.



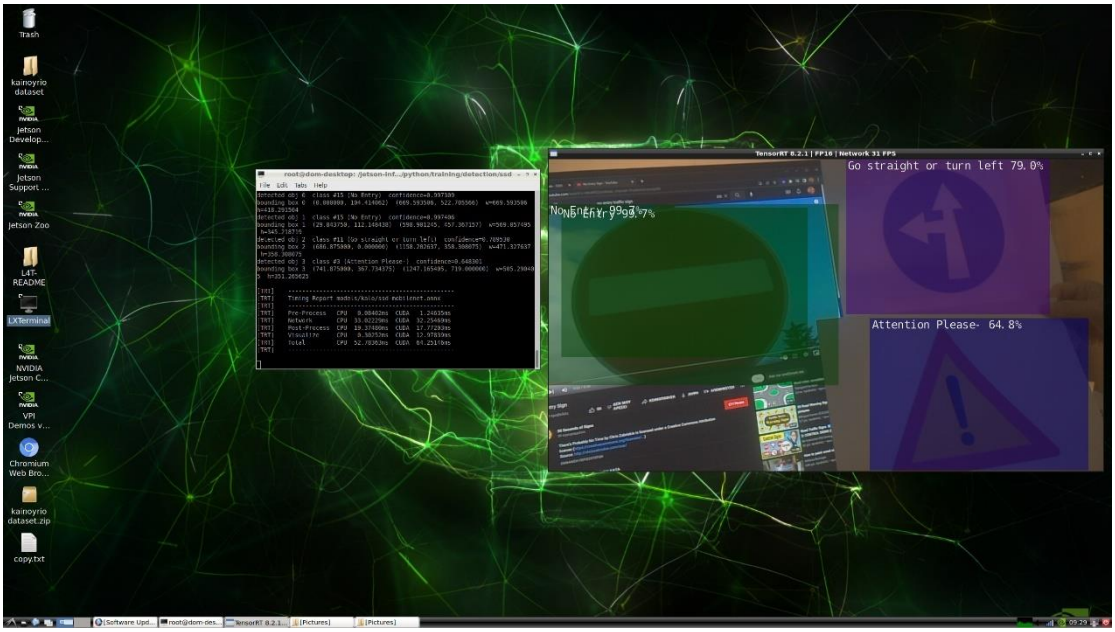
Εικόνα 19.

4.16 TensorRT και συμπεράσματα

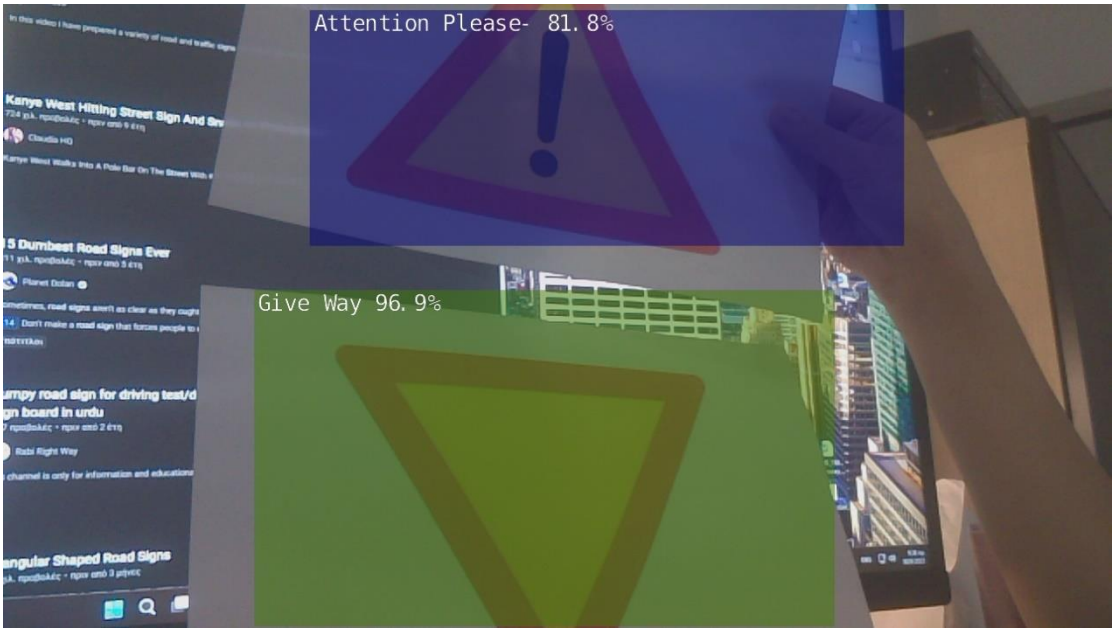
Αφού τρέξει την εντολή αρχίζουμε να απολαμβάνουμε την οπτικοποίηση των δεδομένων και την αποδοτικότητα και ταχύτητα του TensorRT. Χρησιμοποιούμε λοιπόν την κάμερα σαν δεκτή του σήματος και τρέχουμε βίντεο που αναπαράγει εικόνες οδικής σήμανσης. Η αναγνώριση της σήμανσης κατά την αναπαραγωγή βίντεο (real time) είναι πολύ πιο απαιτητική από την απλή εισαγωγή εικόνας προς αναγνώριση. Παρόλα αυτά βλέπουμε υψηλό ποσοστό αναγνώρισης και όσον αφορά αναγνώριση ενός μεμονωμένου στόχου (Εικόνα 20) αλλά και πολλαπλών ταυτόχρονα (Εικόνα 21 και 22 και 23).



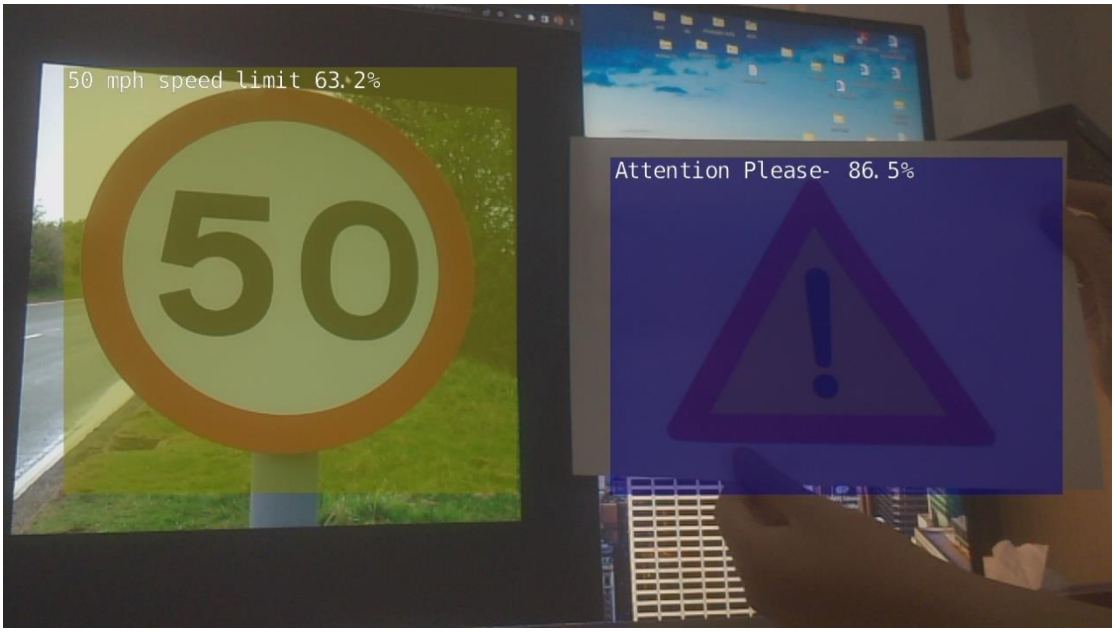
Εικόνα 20.



Εικόνα 21.



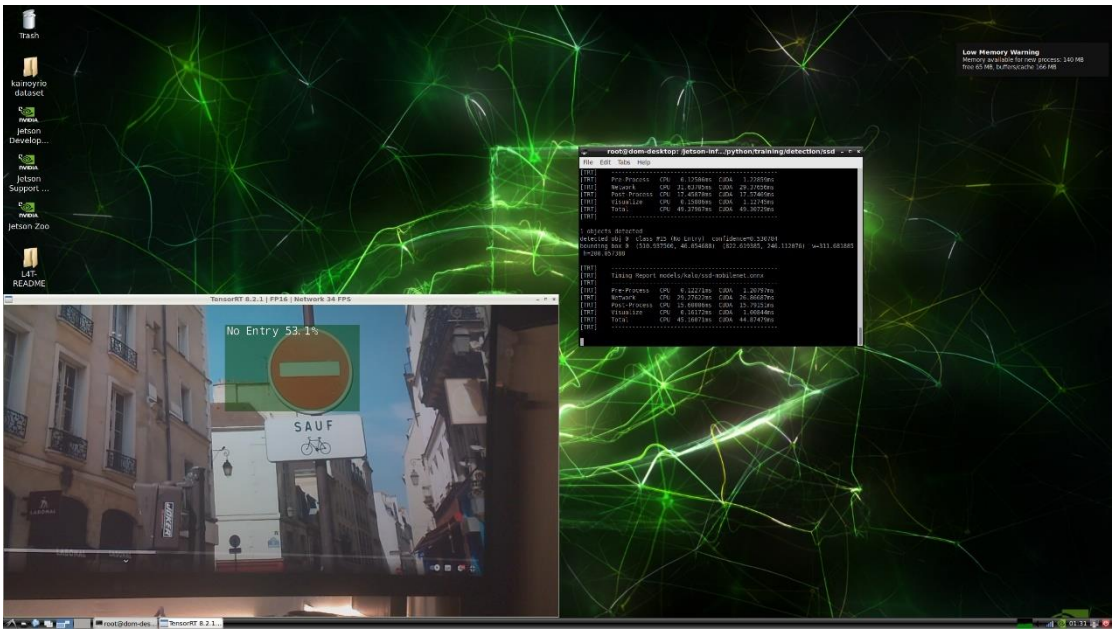
Εικόνα 22.



Εικόνα 23.

Παρότι έγιναν προσπάθειες ανίχνευσης ΟΣ σε πιο μακρινούς στόχους, μερικές απέτυχαν, μερικές όμως πέτυχαν. Ο λόγος αποτυχίας σίγουρα έχει να κάνει με πολλούς παράγοντες όπως για παράδειγμα η κάμερα δεν διαθέτει autofocus (οπότε ήταν δύσκολο να ληφθούν κοντινές και μάκρυνες εικόνες για εκπαίδευση καθότι θα απαιτούσε συνεχή επίδραση στον φακό, κάτι που ενδεχομένως να έφθειρε την κάμερα – σε τέτοιες περιπτώσεις συνίσταται αναπροσαρμοσμένο και κατάλληλα επιλεγμένο dataset και σίγουρα απαιτεί μια κάμερα καλύτερης ανάλυσης και με καλυτέρα γενικότερα χαρακτηριστικά), κ.α.. Παρόλα’ αυτά λόγω της μείξης του σετ δεδομένων με εικόνες που ελήφθησαν από video 4K pov μέσω πλατφόρμας αναπαραγωγής βίντεο (δηλαδή σαν να είναι τραβηγμένες από πραγματική κίνηση αυτοκίνητου σε δρόμο με ρεαλιστικές συνθήκες), η συσκευή μας μπόρεσε να εντοπίσει μερικά ακόμα και υπο αυτές τις απαιτητικές συνθήκες (Εικόνα 24). Τέλος να σημειώσουμε ότι όλη η διαδικασία της αναγνώρισης συμβαίνει

με ένα μέσο ορό 30fps που είναι αρκετά ικανοποιητικό αν σκεφτούμε πως οι υψηλοί ρυθμοί ανανέωσης συναντιούνται σε ελαφρύτερα μοντέλα, δυνατότερους επεξεργαστές ή πολύ χαμηλότερο ποσοστό αυτοπεποίθησης.



Εικόνα 24.

5 ΣΥΜΠΕΡΑΣΜΑΤΑ

Η τεχνολογία κάνει άλματα με ρυθμούς ασύλητους. Έννοιες που ηχούσαν ως επιστημονική φαντασία, πλέον είναι διαθέσιμες στη κυριολεξία και υλοποιήσιμες σε μεγάλο βαθμό. Πριν λίγα χρονιά, δεν φανταζόταν κάνεις πως ένα αυτοκίνητο θα μπορούσε να προσέχει τη σήμανση στο δρόμο ίσως καλυτέρα κι από αρκετούς οδηγούς τριγύρω μας. Ήδη εγκαθίστανται τέτοια και πιο πολυσύνθετα συστήματα σε αυτοκίνητα, κυρίως όμως ακριβά και πιο «premium». Σύμφωνα με τα αποτελέσματα του δικού μας εγχειρήματος, δεν θα ήταν δύσκολο να υποθέσει κάνεις πως είναι θέμα λίγου χρόνου να εγκατασταθούν σχεδόν σε όλα τα αυτοκίνητα παραγωγής, τέτοια συστήματα, μιας και το κόστος είναι εξαιρετικά μικρο και σίγουρα υπάρχουν και καλύτεροι τρόποι εκπαίδευσης και εξαγωγής συμπερασμάτων, από τον δικό μας. Η τεχνητή νοημοσύνη είναι κάτι που σίγουρα ήρθε για να μείνει και αν η βαθιά μάθηση συνεχίζει να εξαπλουστεύται με τον ίδιο ρυθμό, δεν είναι μακριά η μέρα που ο καθένας θα μπορεί να προγραμματίζει με ιδιαίτερη ευκολία.

Σε αυτή την εργασία, αφού αναλύσαμε βασικές ορολογίες και ξεδιαλύναμε το τοπίο, μπήκαμε πιο βαθιά στον κόσμο των νευρωνικών δικτυών και του τρόπου που χρησιμοποιούνται για την αναγνώριση αντικειμένων. Εστιάσαμε ακόμα βαθύτερα στην αναγνώριση συγκεκριμένων αντικειμένων, των οδικών σημάνσεων. Συνεχίσαμε με την ανάλυση εργασιών που αφορούσαν την συγκεκριμένη διαδικασία και διεργασία. Αναφέρθηκαν τεχνικές, μέθοδοι, στρατηγικές και τρόποι βελτιστοποίησης και αποφυγής σφαλμάτων. Τέλος αναλύθηκε όλο το περιβάλλον του Jetson Nano, που εκτός από μια πολύ ισχυρή κινητή συσκευή μεγέθους τσέπης, συνοδεύεται από μια αρίστη παρέα βιβλιοθηκών, εργαλείων και υποστήριξης. Θα τολμούσα να πω πως όλο το προγραμματιστικό πακέτο που περιλαμβάνεται με την συσκευή αυτή, είναι ανεκτίμητης επιστημονικής και ερευνητικής σημασίας καθώς είναι προσιτότατο και ευκολόχρηστο. Μόλις αποδείξαμε, πως μια συσκευή με ιδιαίτερα χαμηλό κόστος για τις δυνατότητες της, μαζί με το παρελκόμενο λογισμικό της, μπορεί να αξιοποιηθεί ακόμα και από κάποιον ερασιτέχνη, και να τη μετατρέψει σε εργαλείο καινοτομίας ή ερευνάς. Τα αποτελέσματα μας είναι ικανοποιητικά.

Βιβλιογραφία

[1] J. Ondruš, E. Kolla, P. Vertaľ, and Ž. Šarić, "How Do Autonomous Cars Work?," in *Transportation Research Procedia*, 2020, vol. 44, pp. 226–233. doi: 10.1016/j.trpro.2020.02.049.

[2] A. Gupta, A. Anpalagan, L. Guan, and A. S. Khwaja, "Deep learning for object detection and scene perception in self-driving cars: Survey, challenges, and open issues," *Array*, vol. 10, p. 100057, Jul. 2021, doi: 10.1016/j.array.2021.100057.

[3] N. G. S. Sai Srinath, A. Z. Joseph, S. Umamaheswaran, C. L. Priyanka, M. Malavika Nair, and P. Sankaran, "NITCAD - Developing an object detection, classification and stereo vision dataset for autonomous navigation in Indian roads," in *Procedia Computer Science*, 2020, vol. 171, pp. 207–216. doi: 10.1016/j.procs.2020.04.022.

[4] E. Khatab, A. Onsy, M. Varley, and A. Abouelfarag, "Vulnerable objects detection for autonomous driving: A review," *Integration*, vol. 78, pp. 36–48, May 2021, doi: 10.1016/j.vlsi.2021.01.002.

[5] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," 2004.

[6] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*, 2005, vol. I, pp. 886–893. doi: 10.1109/CVPR.2005.177.

[7] R. Girshick, "Fast R-CNN," in *2015 IEEE International Conference on Computer Vision (ICCV)*, Dec. 2015, pp. 1440–1448. doi: 10.1109/ICCV.2015.169.

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition," Jun. 2014, doi: 10.1007/978-3-319-10578-9_23.

[9] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," Jun. 2015, [Online]. Available: http://arxiv.org/abs/1506.01497

[10] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," Dec. 2016, [Online]. Available: http://arxiv.org/abs/1612.08242

[11] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," Apr. 2018, [Online]. Available: http://arxiv.org/abs/1804.02767

[12] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: Exceeding YOLO Series in 2021," Jul. 2021, [Online]. Available: http://arxiv.org/abs/2107.08430

[13] A. Vennelakanti, S. Shreya, R. Rajendran, D. Sarkar, D. Muddegowda, and P. Hanagal, "Traffic Sign Detection and Recognition using a CNN Ensemble; Traffic Sign Detection and Recognition using a CNN Ensemble," 2019.

[14] Z. Liu, J. Du, F. Tian, and J. Wen, "MR-CNN: A Multi-Scale Region-Based Convolutional Neural Network for Small Traffic Sign Recognition," *IEEE Access*, vol. 7, pp. 57120–57128, 2019, doi: 10.1109/ACCESS.2019.2913882.

[15] A. Bouti, M. A. Mahraz, J. Riffi, and H. Tairi, “Traffic Sign Detection: A Comparative Study Between CNN and RNN,” in *EAI/Springer Innovations in Communication and Computing*, Springer Science and Business Media Deutschland GmbH, 2022, pp. 53–67. doi: 10.1007/978-3-030-77185-0_4.

[16] W. A. Haque, S. Arefin, A. S. M. Shihavuddin, and M. A. Hasan, “DeepThin: A novel lightweight CNN architecture for traffic sign recognition without GPU requirements,” *Expert Syst Appl*, vol. 168, Apr. 2021, doi: 10.1016/j.eswa.2020.114481.

[17] T. Nguyen, G. Nguyen, and B. M. Nguyen, “EO-CNN: An enhanced CNN model trained by equilibrium optimization for traffic transportation prediction,” in *Procedia Computer Science*, 2020, vol. 176, pp. 800–809. doi: 10.1016/j.procs.2020.09.075.

[18] Q. Tang, G. Cao, and K. H. Jo, “Integrated Feature Pyramid Network with Feature Aggregation for Traffic Sign Detection,” *IEEE Access*, 2021, doi: 10.1109/ACCESS.2021.3106350.

[19] R. Ayachi, M. Afif, Y. Said, and M. Atri, “Traffic Signs Detection for Real-World Application of an Advanced Driving Assisting System Using Deep Learning,” *Neural Process Lett*, vol. 51, no. 1, pp. 837–851, Feb. 2020, doi: 10.1007/s11063-019-10115-8.

[20] “Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles,” *SAE*, 2016.

[21] X.-H. Nguyen, T.-T. Ngo, and D.-A. Nguyen, “Development of Real-Time Traffic-Object and Traffic-Sign Detection Models Applied for Autonomous Intelligent Vehicles,” *JST: Smart Systems and Devices*, vol. 32, no. 1, pp. 17–24, Jan. 2022, doi: 10.51316/jst.155.ssad.2022.32.1.3.

[22] T. Karasawa, K. Watanabe, Q. Ha, A. Tejero-De-Pablos, Y. Ushiku, and T. Harada, “Multispectral object detection for autonomous vehicles,” in *Thematic Workshops 2017 - Proceedings of the Thematic Workshops of ACM Multimedia 2017, co-located with MM 2017*, Oct. 2017, pp. 35–43. doi: 10.1145/3126686.3126727.

[23] J. Varagula, P. A. N. Kulproma, and T. Itob, “Object Detection Method in Traffic by On-Board Computer Vision with Time Delay Neural Network,” in *Procedia Computer Science*, 2017, vol. 112, pp. 127–136. doi: 10.1016/j.procs.2017.08.185.

[24] H. Zhang, K. Wang, Y. Tian, C. Gou, and F. Y. Wang, “MFR-CNN: Incorporating Multi-Scale Features and Global Information for Traffic Object Detection,” *IEEE Trans Veh Technol*, vol. 67, no. 9, pp. 8019–8030, Sep. 2018, doi: 10.1109/TVT.2018.2843394.

[25] Á. Arcos-García, J. A. Álvarez-García, and L. M. Soria-Morillo, “Evaluation of deep neural networks for traffic sign detection systems,” *Neurocomputing*, vol. 316, pp. 332–344, Nov. 2018, doi: 10.1016/j.neucom.2018.08.009.

[26] S. Wang, H. Pan, C. Zhang, and Y. Tian, “RGB-D image-based detection of stairs, pedestrian crosswalks and traffic signs,” *J Vis Commun Image Represent*, vol. 25, no. 2, pp. 263–272, Feb. 2014, doi: 10.1016/j.jvcir.2013.11.005.

[27] H. Gomez-Moreno, S. Maldonado-Bascon, P. Gil-Jimenez, and S. Lafuente-Arroyo, “Goal evaluation of segmentation algorithms for traffic sign recognition,” *IEEE Transactions on*

Intelligent Transportation Systems, vol. 11, no. 4, pp. 917–930, Dec. 2010, doi: 10.1109/TITS.2010.2054084.

[28] N. Barnes, A. Zelinsky, and L. S. Fletcher, "Real-time speed sign detection using the radial symmetry detector," *IEEE Transactions on Intelligent Transportation Systems*, vol. 9, no. 2, pp. 322–332, Jun. 2008, doi: 10.1109/TITS.2008.922935.

[29] X. Yuan, X. Hao, H. Chen, and X. Wei, "Robust traffic sign recognition based on color global and local oriented edge magnitude patterns," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 4, pp. 1466–1474, 2014, doi: 10.1109/TITS.2014.2298912.

[30] G. Loy, "Fast Shape-based Road Sign Detection for a Driver Assistance System," 2004.

[31] J. P. Carrasco, A. de la Escalera, and J. M. Armingol, "Recognition stage for a speed supervisor based on road sign detection," *Sensors (Switzerland)*, vol. 12, no. 9, pp. 12153–12168, Sep. 2012, doi: 10.3390/s120912153.

[32] H. K. Kim, J. H. Park, and H. Y. Jung, "An Efficient Color Space for Deep-Learning Based Traffic Light Recognition," *J Adv Transp*, vol. 2018, 2018, doi: 10.1155/2018/2365414.

[33] C. C. Lin and M. S. Wang, "Road sign recognition with fuzzy adaptive pre-processing models," *Sensors (Switzerland)*, vol. 12, no. 5, pp. 6415–6433, May 2012, doi: 10.3390/s120506415.

[34] Z. Liu, M. Qi, C. Shen, Y. Fang, and X. Zhao, "Cascade saccade machine learning network with hierarchical classes for traffic sign detection," *Sustain Cities Soc*, vol. 67, Apr. 2021, doi: 10.1016/j.scs.2020.102700.

[35] T. Chen and S. Lu, "Accurate and Efficient Traffic Sign Detection Using Discriminative AdaBoost and Support Vector Regression," *IEEE Trans Veh Technol*, vol. 65, no. 6, pp. 4006–4015, Jun. 2016, doi: 10.1109/TVT.2015.2500275.

[36] Y. Saadna, A. Behloul, and S. Mezzoudj, "Speed limit sign detection and recognition system using SVM and MNIST datasets," *Neural Comput Appl*, vol. 31, no. 9, pp. 5005–5015, Sep. 2019, doi: 10.1007/s00521-018-03994-w.

[37] S. Ahmed, U. Kamal, and M. K. Hasan, "DFR-TSD: A Deep Learning Based Framework for Robust Traffic Sign Detection Under Challenging Weather Conditions," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 6, pp. 5150–5162, Jun. 2022, doi: 10.1109/TITS.2020.3048878.

[38] J. Xiong, L. Ye, D. Jiang, T. Ye, F. Wang, and L. Y. Zhu, "Efficient Traffic Sign Recognition Using Cross-Connected Convolution Neural Networks Under Compressive Sensing Domain," *Mobile Networks and Applications*, vol. 26, no. 2, pp. 629–637, Apr. 2021, doi: 10.1007/s11036-019-01409-1.

[39] S. Zhou, C. Deng, Z. Piao, and B. Zhao, "Few-shot traffic sign recognition with clustering inductive bias and random neural network," *Pattern Recognit*, vol. 100, Apr. 2020, doi: 10.1016/j.patcog.2019.107160.

[40] F. Yu, Z. Qin, C. Liu, D. Wang, and X. Chen, "REIN the RobuTS: Robust DNN-Based Image Recognition in Autonomous Driving Systems," *IEEE Transactions on Computer-Aided Design of*

Integrated Circuits and Systems, vol. 40, no. 6, pp. 1258–1271, Jun. 2021, doi: 10.1109/TCAD.2020.3033498.

[41] U. Kamal, T. I. Tonmoy, S. Das, and M. K. Hasan, "Automatic Traffic Sign Detection and Recognition Using SegU-Net and a Modified Tversky Loss Function with L1-Constraint," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 4, pp. 1467–1479, Apr. 2020, doi: 10.1109/TITS.2019.2911727.

[42] D. Dwibedi, I. Misra, and M. Hebert, "Cut, Paste and Learn: Surprisingly Easy Synthesis for Instance Detection," in *Proceedings of the IEEE International Conference on Computer Vision*, Dec. 2017, vol. 2017-October, pp. 1310–1319. doi: 10.1109/ICCV.2017.146.

[43] H. Wang, Q. Wang, F. Yang, W. Zhang, and W. Zuo, "Data Augmentation for Object Detection via Progressive and Selective Instance-Switching," Jun. 2019, [Online]. Available: http://arxiv.org/abs/1906.00358

[44] X. Peng, B. Sun, K. Ali, and K. Saenko, "Learning Deep Object Detectors from 3D Models," in *2015 IEEE International Conference on Computer Vision (ICCV)*, Dec. 2015, pp. 1278–1286. doi: 10.1109/ICCV.2015.151.

[45] B. Moiseev, A. Konev, A. Chigorin, and A. Konushin, "LNCS 8192 - Evaluation of Traffic Sign Recognition Methods Trained on Synthetically Generated Data," 2013. [Online]. Available: http://graphics.cs.msu.ru/files/research/

[46] A. Stergiou, G. Kalliatakis, and C. Chrysoulas, "Traffic sign recognition based on synthesised training data," *Big Data and Cognitive Computing*, vol. 2, no. 3, pp. 1–16, Sep. 2018, doi: 10.3390/bdcc2030019.

[47] A. Møgelmose, M. M. Trivedi, and T. B. Moeslund, *Learning to Detect Traffic Signs: Comparative Evaluation of Synthetic and Real-world Datasets*. 2012. doi: 10.0/Linux-x86_64.

[48] N. Dvornik, J. Mairal, and C. Schmid, "Modeling Visual Context is Key to Augmenting Object Detection Datasets," Jul. 2018, [Online]. Available: http://arxiv.org/abs/1807.07428

[49] G. Georgakis, A. Mousavian, A. C. Berg, and J. Kosecka, "Synthesizing Training Data for Object Detection in Indoor Scenes," Feb. 2017, [Online]. Available: http://arxiv.org/abs/1702.07836

[50] A. Gupta, A. Vedaldi, and A. Zisserman, "Synthetic Data for Text Localisation in Natural Images," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Dec. 2016, vol. 2016-December, pp. 2315–2324. doi: 10.1109/CVPR.2016.254.

[51] S. M. Grigorescu, "Generative one-shot learning (GOL): A semi-parametric approach to one-shot learning in autonomous vision," in *Proceedings - IEEE International Conference on Robotics and Automation*, Sep. 2018, pp. 7127–7134. doi: 10.1109/ICRA.2018.8461174.

[52] J. Kim, T. H. Oh, S. Lee, F. Pan, and I. S. Kweon, "Variational prototyping-encoder: One-shot learning with prototypical images," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun. 2019, vol. 2019-June, pp. 9454–9462. doi: 10.1109/CVPR.2019.00969.

[53] J. Devaranjan, A. Kar, and S. Fidler, “Meta-Sim2: Unsupervised Learning of Scene Structure for Synthetic Data Generation,” Aug. 2020, [Online]. Available: http://arxiv.org/abs/2008.09092

[54] K. R. Sri Preethaa and A. Sabari, “Intelligent video analysis for enhanced pedestrian detection by hybrid metaheuristic approach,” *Soft comput*, vol. 24, no. 16, pp. 12303–12311, Aug. 2020, doi: 10.1007/s00500-020-04674-5.

[55] R. Lahmyed, M. el Ansari, and A. Ellahyani, “A new thermal infrared and visible spectrum images-based pedestrian detection system,” *Multimed Tools Appl*, vol. 78, no. 12, pp. 15861–15885, Jun. 2019, doi: 10.1007/s11042-018-6974-5.

[56] Z. Kerkaou and M. el Ansari, “Support vector machines based stereo matching method for advanced driver assistance systems,” *Multimed Tools Appl*, vol. 79, no. 37–38, pp. 27039–27055, Oct. 2020, doi: 10.1007/s11042-020-09260-3.

[57] D. Sudha and J. Priyadarshini, “An intelligent multiple vehicle detection and tracking using modified vibe algorithm and deep learning algorithm,” *Soft comput*, vol. 24, no. 22, pp. 17417–17429, Nov. 2020, doi: 10.1007/s00500-020-05042-z.

[58] M. el Ansari, S. Mousset, and A. Bensrhair, “Temporal consistent real-time stereo for intelligent vehicles,” *Pattern Recognit Lett*, vol. 31, no. 11, pp. 1226–1238, Aug. 2010, doi: 10.1016/j.patrec.2010.03.023.

[59] Y. Liu, J. Ling, Q. Wu, and B. Qin, “Scalable privacy-enhanced traffic monitoring in vehicular ad hoc networks,” *Soft comput*, vol. 20, no. 8, pp. 3335–3346, Aug. 2016, doi: 10.1007/s00500-015-1737-y.

[60] R. Lahmyed, M. el Ansari, and Z. Kerkaou, “Automatic road sign detection and recognition based on neural network,” *Soft comput*, vol. 26, no. 4, pp. 1743–1764, Feb. 2022, doi: 10.1007/s00500-021-06726-w.

[61] K. Garg and S. K. Nayar, “Detection and Removal of Rain from Videos *,” 2004.

[62] M. Hnewa and H. Radha, “Object Detection Under Rainy Conditions for Autonomous Vehicles: A Review of State-of-the-Art and Emerging Techniques,” Jun. 2020, doi: 10.1109/MSP.2020.2984801.

[63] X. Z. Chen, C. M. Chang, C. W. Yu, and Y. L. Chen, “A real-time vehicle detection system under various bad weather conditions based on a deep learning model without retraining,” *Sensors (Switzerland)*, vol. 20, no. 20, pp. 1–22, Oct. 2020, doi: 10.3390/s20205731.

[64] N. Nan, R. Gang, and R. Song, “Image Defogging Algorithm Based on Fisher Criterion Function and Dark Channel Prior,” in *Proceedings - 2020 13th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics, CISP-BMEI 2020*, Oct. 2020, pp. 98–103. doi: 10.1109/CISP-BMEI51763.2020.9263582.

[65] J. Greenhalgh and M. Mirmehdi, “TRAFFIC SIGN RECOGNITION USING MSER AND RANDOM FORESTS,” 2012.

[66] B. Li, S. Wang, J. Zheng, and L. Zheng, “Single image haze removal using content-adaptive dark channel and post enhancement,” *IET Computer Vision*, vol. 8, no. 2, pp. 131–140, 2014, doi: 10.1049/iet-cvi.2013.0011.

[67] E. Provenzi, M. Fierro, A. Rizzi, L. de Carli, D. Gadia, and D. Marini, “Random spray retinex: A new retinex implementation to investigate the local properties of the model,” *IEEE Transactions on Image Processing*, vol. 16, no. 1, pp. 162–171, Feb. 2007, doi: 10.1109/TIP.2006.884946.

[68] A. Rizzi, C. Gatta, and D. Marini, “A new algorithm for unsupervised global and local color correction,” *Pattern Recognit Lett*, vol. 24, no. 11, pp. 1663–1677, 2003, doi: 10.1016/S0167-8655(02)00323-9.

[69] A. Deshmukh and S. Singh, “DESIGN AND DEVELOPMENT OF IMAGE DEFOGGING SYSTEM.”

[70] G. Hines, Z.-U. Rahman, D. Jobson, G. Woodell, and N. Langley, “Single-Scale Retinex Using Digital Signal Processors.”

[71] H. Xu, J. Guo, Q. Liu, and L. Ye, *Fast Image Dehazing Using Improved Dark Channel Prior*. IEEE, 2012.

[72] S. K. Nayar and S. G. Narasimhan, “Vision in Bad Weather *,” 1999.

[73] D. H. Hubel and T. N. Wiesel, “RECEPTIVE FIELDS AND FUNCTIONAL ARCHITECTURE OF MONKEY STRIATE CORTEX,” 1968.

[74] B. D. Ripley, *Pattern recognition and neural networks*. Cambridge University Press, 1996.

[75] S. Cong and Y. Zhou, “A review of convolutional neural network architectures and their optimizations,” *Artif Intell Rev*, 2022, doi: 10.1007/s10462-022-10213-5.

[76] R. S. Sandhya Devi, V. R. Vijay Kumar, and P. Sivakumar, “A Review of image Classification and Object Detection on Machine learning and Deep Learning Techniques,” 2021. doi: 10.1109/ICECA52323.2021.9676141.

[77] J. Xing, “Traffic Sign Recognition From Digital Images Using Deep Learning,” 2021.

[78] C. Ajmi, J. Zapata, S. Elferchichi, A. Zaafouri, and K. Laabidi, “Deep Learning Technology for Weld Defects Classification Based on Transfer Learning and Activation Features,” *Advances in Materials Science and Engineering*, vol. 2020, 2020, doi: 10.1155/2020/1574350.

[79] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, “Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs),” Nov. 2015, [Online]. Available: http://arxiv.org/abs/1511.07289

[80] A. L. Maas, A. Y. Hannun, and A. Y. Ng, “Rectifier Nonlinearities Improve Neural Network Acoustic Models,” 2013.

[81] B. Xu, N. Wang, T. Chen, and M. Li, “Empirical Evaluation of Rectified Activations in Convolutional Network,” May 2015, [Online]. Available: http://arxiv.org/abs/1505.00853

[82] K. He, X. Zhang, S. Ren, and J. Sun, “Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition,” *IEEE Trans Pattern Anal Mach Intell*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015, doi: 10.1109/TPAMI.2015.2389824.

[83] C. Gulcehre, K. Cho, R. Pascanu, and Y. Bengio, “LNAI 8724 - Learned-Norm Pooling for Deep Feedforward and Recurrent Neural Networks.”

[84] J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers, and A. W. M. Smeulders, “Selective search for object recognition,” *Int J Comput Vis*, vol. 104, no. 2, pp. 154–171, Sep. 2013, doi: 10.1007/s11263-013-0620-5.

[85] C. Li, B. Zhang, H. Hu, and J. Dai, “Enhanced Bird Detection from Low-Resolution Aerial Image Using Deep Neural Networks,” *Neural Process Lett*, vol. 49, no. 3, pp. 1021–1039, Jun. 2019, doi: 10.1007/s11063-018-9871-z.

[86] J. Hosang, R. Benenson, and B. Schiele, “Learning non-maximum suppression,” in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, Nov. 2017, vol. 2017-January, pp. 6469–6477. doi: 10.1109/CVPR.2017.685.

[87] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” Sep. 2014, [Online]. Available: http://arxiv.org/abs/1409.1556

[88] C. Szegedy *et al.*, “Going Deeper with Convolutions,” Sep. 2014, [Online]. Available: http://arxiv.org/abs/1409.4842

[89] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” Jun. 2015, [Online]. Available: http://arxiv.org/abs/1506.02640

[90] I. J. Goodfellow *et al.*, “Generative Adversarial Networks,” Jun. 2014, [Online]. Available: http://arxiv.org/abs/1406.2661

[91] K. Janocha and W. M. Czarnecki, “On Loss Functions for Deep Neural Networks in Classification,” Feb. 2017, [Online]. Available: http://arxiv.org/abs/1702.05659

[92] J. Heaton, “Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep learning,” *Genet Program Evolvable Mach*, vol. 19, no. 1–2, pp. 305–307, Jun. 2018, doi: 10.1007/s10710-017-9314-z.

[93] J. Duchi JDUCHI and Y. Singer, “Adaptive Subgradient Methods for Online Learning and Stochastic Optimization * Elad Hazan,” 2011.

[94] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” Dec. 2014, [Online]. Available: http://arxiv.org/abs/1412.6980

[95] S. Ruder, “An overview of gradient descent optimization algorithms,” Sep. 2016, [Online]. Available: http://arxiv.org/abs/1609.04747

[96] S. Ioffe and C. Szegedy, “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,” Feb. 2015, [Online]. Available: http://arxiv.org/abs/1502.03167

[97] L. Cun *et al.*, “Handwritten Digit Recognition with a Back-Propagation Network,” 1990.

[98] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, and Li Fei-Fei, “ImageNet: A large-scale hierarchical image database,” Mar. 2010, pp. 248–255. doi: 10.1109/cvpr.2009.5206848.

[99] M. Everingham, L. van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, “The pascal visual object classes (VOC) challenge,” *Int J Comput Vis*, vol. 88, no. 2, pp. 303–338, Jun. 2010, doi: 10.1007/s11263-009-0275-4.

[100] T.-Y. Lin *et al.*, "Microsoft COCO: Common Objects in Context," May 2014, [Online]. Available: http://arxiv.org/abs/1405.0312

[101] Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, and A. Y. Ng, "Reading Digits in Natural Images with Unsupervised Feature Learning." [Online]. Available: http://ufldl.stanford.edu/housenumbers/

[102] A. Krizhevsky, "Learning Multiple Layers of Features from Tiny Images," 2009.

[103] A. Gondaliya, "REGULARIZATION IMPLEMENTATION IN R : BIAS AND VARIANCE DIAGNOSIS," May 22, 2014. https://pingax.com/regularization-implementation-r/ (accessed Sep. 25, 2022).

[104] N. Srivastava, G. Hinton, A. Krizhevsky, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," 2014.

[105] M. Abadi *et al.*, "TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems," Mar. 2016, [Online]. Available: http://arxiv.org/abs/1603.04467

[106] A. Paszke *et al.*, "Automatic differentiation in PyTorch," Long Beach, CA, 2017.

[107] M. Arif, W. Farooq, A. B. Saduf, A. Asif, and I. Khan, "Studies in Big Data 57 Advances in Deep Learning." [Online]. Available: http://www.springer.com/series/11970

[108] NVIDIA, "JetPack SDK 4.6.1." https://developer.nvidia.com/embedded/jetpack-sdk-461 (accessed Sep. 23, 2022).

[109] A. G. Howard *et al.*, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," Apr. 2017, [Online]. Available: http://arxiv.org/abs/1704.04861

[110] Á. Morera, Á. Sánchez, A. B. Moreno, Á. D. Sappa, and J. F. Vélez, "Ssd vs. Yolo for detection of outdoor urban advertising panels under multiple variabilities," *Sensors (Switzerland)*, vol. 20, no. 16, pp. 1–23, Aug. 2020, doi: 10.3390/s20164587.

[111] A. G. Howard *et al.*, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," Apr. 2017, [Online]. Available: http://arxiv.org/abs/1704.04861

[112] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," Jan. 2018, [Online]. Available: http://arxiv.org/abs/1801.04381

[113] A. Howard *et al.*, "Searching for mobileNetV3," in *Proceedings of the IEEE International Conference on Computer Vision*, Oct. 2019, vol. 2019-October, pp. 1314–1324. doi: 10.1109/ICCV.2019.00140.

[114] J. Liu, Y. Fan, S. Wu, Z. Liang, W. Xie, and L. Fu, "An Identification Method for Mechanical Parts Using on Deep Learning with Single Shot-multibox Detector MobileNet."