



Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης
Πολυτεχνική Σχολή
Τμήμα Ηλεκτρολόγων Μηχανικών &
Μηχανικών Γπολογιστών
Τομέας Ηλεκτρονικής και Γπολογιστών

Διπλωματική Εργασία

Έρευνα ταξινομητών DNN για αναγνώριση
αντικειμένων και βελτιστοποίηση για το
ενσωματωμένο σύστημα Jetson-TK1

Εκπόνηση:
Παναγιώτου Κωνσταντίνος
ΑΕΜ: 7316

Επίβλεψη:
Αν. Καθ. Πέτρου Λουκάς
Υπ. Δρ. Μουσουλιώτης
Παναγιώτης

The robots are going to help us find our crystal...

ΕΥΧΑΡΙΣΤΙΕΣ

Thank you! =]

Περίληψη

Title

Thesis Title here!

Abstract

Konstantinos Panayiotou
Electrical & Computer Engineering Department,
Aristotle University of Thessaloniki, Greece
September 2016

Περιεχόμενα

Ευχαριστίες	iii
Περίληψη	v
Abstract	vii
Ακρωνύμια	xv
1 Εισαγωγή	1
1.1 Περιγραφή του Προβλήματος	2
1.2 Σκοπός - Συνεισφορά της Διπλωματικής Εργασίας	3
1.3 Διάρθρωση της Αναφοράς	3
2 Επισκόπηση της Ερευνητικής Περιοχής	4
3 Τεχνικές Βαθιάς Μηχανικής Μάθησης και Αναγνώρισης Αντικειμένων στον χώρο	6
3.1 Εισαγωγή στην Επιστήμη της Σύγχρονης Μηχανικής Μάθησης	7
3.2 Νευρωνικά Δίκτυα με Βάθος	15
3.2.1 Συναρτήσεις Ενεργοποίησης	20
3.2.2 Συναρτήσεις Σφάλματος/Κόστους	22
3.2.3 Αλγόριθμος Backpropagation	23
3.3 Νευρωνικά Δίκτυα Συνέλιξης	25
3.3.1 Επίπεδο Συνέλιξης	27
3.3.2 Επίπεδο Υπό-δειγματοληφίας - Pooling layer	30
3.3.3 Πλήρως Συνδεδεμένο Επίπεδο - Fully-connected layer	31
4 Εργαλεία Hardware και Software	32
4.1 NVIDIA Jetson TK1 development board	32
4.2 Εργαλεία Λογισμικού	36
5 Υλοποιήσεις	40
5.1 Μοντέλα CNN	40
5.1.1 AlexNet	42
5.1.2 VGG16	44
5.1.3 YOLO Net	46
5.2 Βελτιστοποίηση των εργαλείων λογισμικού στο Jetson TK1	47
6 Πειράματα - Αποτελέσματα	50
6.1 Πρώτη Φάση Πειραμάτων	50
6.2 Δεύτερη Φάση Πειραμάτων	51

ΠΕΡΙΕΧΟΜΕΝΑ

7 Συμπεράσματα	53
8 Μελλοντικές επεκτάσεις	54
Βιβλιογραφία	55

Κατάλογος Σχημάτων

3.1	Κλάδοι και εφαρμογές της επιστήμης της Τεχνητής Νοημοσύνης	7
3.2	Διάγραμμα Venn των διαφόρων κατηγοριών μηχανικής μάθησης	9
3.3	Παράδειγμα διαφορετικών αναπαραστάσεων των δεδομένων	13
3.4	Απλό μοντέλο Autoncoder με ένα χρυφό επίπεδο	13
3.5	Παράδειγμα απεικόνισης των φίλτρων ενός μοντέλου CNN για αναγνώριση προσώπου	14
3.6	Παράδειγμα απεικόνισης των φίλτρων ενός μοντέλου CNN για αναγνώριση προσώπου	15
3.7	Βιολογικός Νευρώνας	16
3.8	Μαθηματικό μοντέλο του νευρώνα	17
3.9	Γλοποίηση πύλης AND με χρήση του μαθηματικού μοντέλου του νευρώνα	18
3.10	Απλό μοντέλο NN με ένα χρυφό επίπεδο - Perceptron	19
3.11	Μορφή ενός πολυεπίπεδου νευρωνικού δικτύου	19
3.12	Συνάρτηση Σιγμοειδούς συνάρτησης	20
3.13	Συνάρτηση Υπερβολικής Εφαπτωμένης	21
3.14	Συνάρτηση Rectified Linear Unit - ReLU	21
3.15	Συνάρτηση Leaky ReLU	22
3.16	Συνάρτηση Maxout	22
3.17	Διαδικασία Backward propagation	24
3.18	Τρισδιάστατη κατανομή των νευρώνων στα CNN	26
3.19	Παράδειγμα διασύνδεσης τρισδιάστατης εισόδου με την τρισδιάστατη δομή των νευρόνων ενός επιπέδου συνέλιξης (CONV)	27
3.20	Συνέλιξη φίλτρου ενός επιπέδου συνέλιξης με τον όγκο εισόδου και παραγωγή ενός χάρτη ενεργοποίησης	28
3.21	Διαδικασία Συνέλιξης	28
3.22	Αντιστοιχία του αριθμού των φίλτρων ενός επιπέδου συνέλιξης με το βάθος του όγκου στην έξοδο	29
3.23	Διαδοχικές εφαρμογές του τελεστή συνέλιξης προκαλούν μείωση των διαστάσεων μήκους και πλάτους του όγκου	29
3.24	Zero Padding	30
3.25	Επίπεδο Υποδειγματοληφίας - Pooling layer	31
3.26	Συνάρτηση υπό-δειγματοληφίας Max - Max Pooling	31
4.1	Tegra K1 SOC	33
4.2	Jetson TK1 development board	33
4.3	ARM Cortex-A15 processor	35

ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ

4.4 Παράδειγμα αναπαράστασης ενός μοντέλου νευρωνικού δικτύου σε γράφο. Το NN αποτλείται από ένα χρυφό επίπεδο και ένα ταξινομητή Softmax στο επίπεδο εξόδου	37
5.1 Μοντέλο δικτύου AlexNet	42
5.2 Πλήρης μορφή του δικτύου AlexNet της υλοποίησης	43
5.3 Μορφή του δικτύου VGG16	44
5.4 Πλήρης μορφή του δικτύου VGG16 της υλοποίησης	45
5.5 Παράδειγμα τρόπου λειτουργίας του δικτύου YOLO	46

Κατάλογος πινάκων

Κατάλογος Αλγορίθμων

3.1 Αλγόριθμος Feedforward για τον υπολογισμό των εξόδων ενός επιπέδου του NN	20
3.2 Backpropagation learning algorithm	26

Ακρωνύμια Εγγράφου

Παρακάτω παρατίθενται ορισμένα από τα πιο συχνά χρησιμοποιούμενα ακρωνύμια της παρούσας διπλωματικής εργασίας:

AI	→ Artificial Intelligence
ML	→ Machine Learning
NN	→ Neural Network
ANN	→ Artificial Neural Network
DL	→ Deep Learning
DNN	→ Deep Neural Network
MLP	→ Multilayer Perceptron
CNN	→ Convolutional Neural Network
ConvNet	→ Convolutional Networks
RNN	→ Recurrent Neural Network
YOLO	→ You Only Look Once
SOC	→ System On Chip
OS	→ Operating System

1

Εισαγωγή

Ο όρος *Τεχνητή Νοημοσύνη* αναφέρεται στην ικανότητα των υπολογιστικών συστημάτων να μιμούνται στοιχεία της ανθρώπινης συμπεριφοράς. Η επιθυμία των ανθρώπων να κατασκευάσουν "έξυπνες" μηχανές, καταγράφεται από την εποχή της αρχαίας Ελλάδας. Μυθικές μορφές όπως οι Πυγμαλίωνας, Δαίδαλος και Ήφαιστος μπορούν να θεωρηθούν ως θρυλικοί εφευρέτες και δημιουργοί νοούντων μηχανών όπως η Γαλάτεια, ο Τάλος και η Πανδώρα.

Ένας πιο ολοκληρωμένος ορισμός της Τεχνητής Νοημοσύνης είναι:

Ο κλάδος/τομέας της επιστήμης της πληροφορικής, που ασχολείται με την σχεδίαση και κατασκευή ευφυών συστημάτων, δηλαδή συστημάτων που διαθέτουν χαρακτηριστικά που σχετίζονται με την ανθρώπινη νοημοσύνη και συμπεριφορά.

Με την εμφάνιση των πρώτων ηλεκτρονικών και (επανα)προγραμματιζόμενων υπολογιστικών συστημάτων, οι άνθρωποι ξεκίνησαν να σκέφτονται τρόπους για να κατασκευάσουν "έξυπνες" μηχανές. Η ραγδαία εξέλιξη στον κλάδο της επιστήμης της πληροφορικής τις τελευταίες δεκαετίες, έφερε και την εξέλιξη στην επιστήμη της Τεχνητής Νοημοσύνης. Το 1997, η IBM κατασκεύασε ένα υπολογιστικό σύστημα το οποίο μπορούσε να παίξει σκάκι (Deep Blue) [1], το οποίο κέρδισε τον παγκόσμιο τότε πρωταθλητή στο σκάκι, Garry Kasparov. Το σκάκι έχει εξήντα τέσσερις θέσεις και τριάντα δύο πιόνια που μπορούν να κινούνται με συγκεκριμένο τρόπο. Η μηχανή Deep Blue, μπορούσε να εκτιμήσει και να αξιολογήσει διακόσια εκατομμύρια πιθανές καταστάσεις της σκακιέρας. Ωστόσο, πρέπει να σημειώσουμε ότι η επίλυση του προβλήματος του σκακιού, παρ' ότι είναι ένα πρόβλημα το οποίο μπορεί να περιγραφεί πλήρως μέσα από μια λίστα με κανόνες, δεν είναι δυνατό να λυθεί αναλυτικά εξαιτίας του τεράστιου αριθμού των δυνατών κινήσεων.

Η "δύναμη" της τεχνητής νοημοσύνης υπερβαίνει τα όρια της φαντασίας. Τα ρομποτικά συστήματα ήδη έχουν περάσει φάση δοκιμής, σε διάφορες χώρες ανά τον κόσμο, όσον αφορά την ένταξή τους στην καθημερινότητα του ανθρώπου.

ΚΕΦΑΛΑΙΟ 1. ΕΙΣΑΓΩΓΗ

Η δυνατότητα ενός "ευφυούς" ρομποτικού συστήματος να αντιλαμβάνεται το περιβάλλον του, είναι απαραίτητη ικανότητα που πρέπει να διαθέτει. Ένας ρομποτικός πράκτορας, σε πληθώρα εφαρμογών, πρέπει να διαθέτει τόσο την ικανότητα να αναγνωρίζει, μέσω εικόνων λήψης από κάμερα, διάφορα αντικείμενα, όσο και την ικανότητα να εντοπίζει επακριβώς την θέση του εκάστοτε αντικειμένου στον χώρο. Δεν θα είχε νόημα για ένα ρομπότ, να έχει άποψη για την παρουσία αντικειμένων γύρω του, αν δεν έχει και την ικανότητα να γνωρίζει και την θέση των αντικειμένων αυτών.

Ένας σημαντικός παράγοντας της ραγδαίας εξέλιξης της επιστήμης της ρομποτικής, αλλά και της τεχνητής νοημοσύνης γενικότερα, είναι η εμφάνιση και εξέλιξη του κλάδου της *Βαθιάς Μηχανικής Μάθησης - Deep learning*.

Η χρήση τεχνικών βαθιάς μάθησης στην επίλυση προβλημάτων *Μηχανικής Όρασης*, έχει φέρει την συγκεκριμένη τεχνολογία σε θέση να μπορεί να αντιμετωπίσει περίπλοκα προβλήματα τα οποία μέχρι και πριν λίγα χρόνια θεωρείτο ακατόρθωτο να λυθούν. Ένα στοχευμένο παράδειγμα είναι τα αυτόνομα αυτοκίνητα, τα οποία σήμερα είναι σε φάση δοκιμής.

Σήμερα, το γενικότερο πρόβλημα της αναγνώρισης και εντοπισμού αντικειμένων σε εικόνες λύνεται με την χρήση *Νευρωνικών Δικτύων Συνέλιξης* (Convolutional Neural Networks - CNN). Το γεγονός ότι τα CNN έχουν την δυνατότητα να κατηγοριοποιήσουν αντικείμενα σε προβλήματα όπου οι κλάσεις αντικειμένων κυμαίνονται από δεκάδες έως και χιλιάδες, τα καθιστά ικανά να χρησιμοποιηθούν σε πραγματικού χρόνου εφαρμογές και ιδιαίτερα στην επιστήμη της ρομποτικής.

1.1 ΠΕΡΙΓΡΑΦΗ ΤΟΥ ΠΡΟΒΛΗΜΑΤΟΣ

Παρόλο που τα CNN είναι ικανά να δώσουν λύσεις με μεγάλη ακρίβεια, απαιτούν μεγάλο όγκο επεξεργαστικής ισχύος, τόσο για την εκπαίδευσή τους, όσο και για την εκτέλεση ενός πειράματος, ιδιαίτερα όταν το πρόβλημα για το οποίο έχουν σχεδιαστεί να δώσουν λύση είναι περίπλοκο. Η απαίτηση αυτή προέρχεται από το βάθος των μοντέλων CNN για αναγνώριση και εντοπισμό αντικειμένων σε εικόνες. Για παράδειγμα, ένα από τα πρώτα μοντέλα CNN το οποίο σχεδιάστηκε για την αναγνώριση αντικειμένων σε εικόνες, αποτελείται από 16 επίπεδα (AlexNet) και έχει εξήντα εκατομμύρια (60,000,000) παραμέτρους και εξακόσιες πενήντα χιλιάδες (650,000) νευρώνες από τους οποίους οι περισσότεροι εκτελούν πράξεις συνέλιξης. Ο Alex Krizhevsky απέδειξε το 2012 ότι η χρήση σύγχρονων GPU για την εκτέλεση πράξεων συνέλιξης, φέρει σαν αποτέλεσμα την εκπαίδευση μοντέλων CNN σε χρόνους εώς και δύο τάξεις μεγέθους πιο κάτω σε σχέση με έναν ισχυρό επεξεργαστή [2]. Ο χρόνος εκτέλεσης ενός πειράματος του Δικτύου AlexNet έχει μετρηθεί στα 7.39 δευτερόλεπτα σε οκταπύρηγο επεξεργαστή Haswell @2.9Ghz και στα 0.71 δευτερόλεπτα σε μονάδα GPU, Nvidia K520 [3].

Είναι ιδιαίτερα σημαντικό, ένα ρομπότ να μπορεί να αντιλαμβάνεται το περιβάλλον του. Αυτό περιλαμβάνει την αναγνώριση ανθρώπων, ζώων, και γενικότερα αντικειμένων. Ωστόσο, θέλουμε τα ρομπότ να είναι και όσο πιο "ελκυστικά" γίνεται στον άνθρωπο ή/και μικρότερα, ανάλογα με το task που επιθυμούμε να εκτελέσουν. Αυτό, φέρει σαν αποτέλεσμα να μην μπορούμε να τοποθετήσουμε ογκώδη, άρα με

1.2. ΣΚΟΠΟΣ - ΣΥΝΕΙΣΦΟΡΑ ΤΗΣ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

μεγάλη υπολογιστική ισχύ, υπολογιστικά συστήματα στο σώμα των ρομποτικών συστημάτων σε όλες τις περιπτώσεις.

Παρόλο που σήμερα έχουν σχεδιαστεί μοντέλα CNN, τα οποία έχουν την δυνατότητα να αναγνωρίσουν και να εντοπίσουν αντικείμενα από χιλιάδες, αν όχι και περισσότερες, κλάσεις, ο χρόνος που απαιτείται για να κατηγοριοποιήσει αντικείμενα σε μία εικόνα είναι αρκετά μεγάλος (της τάξης των μερικών δευτερολέπτων σε σύγχρονους υπολογιστές). Αυτό κάνει την χρήση των CNN σε εφαρμογές πραγματικού χρόνου, όπως για παράδειγμα στην ρομποτική, ακατάλληλη. Ωστόσο, η επιστημονική κοινότητα σήμερα προσπαθεί να δώσει λύσεις στο συγκεκριμένο πρόβλημα εστιάζοντας το ενδιαφέρον στην εξέλιξη των ενσωματωμένων συστημάτων και την σχεδίαση γρήγορου, σε χρόνο εκτέλεσης, λογισμικού για υλοποιήσεις μοντέλων CNN τα οποία εκμεταλλεύονται κυρίως την υπολογιστική ισχύ των μονάδων GPU, αλλά και άλλων πολυπύρηνων επεξεργαστικών μονάδων.

1.2 Σκοπος - Συνεισφορα της Διπλωματικης Εργασιας

Η παρούσα διπλωματική εργασία έχει ως στόχο την υλοποίηση και σύγχριση διαφόρων μοντέλων CNN για αναγνώριση και εντοπισμό αντικειμένων σε εικόνες (object detection), σε ένα ρομποτικό σύστημα που χρησιμοποιεί το ενσωματωμένο σύστημα Jetson TK1 της NVIDIA, το οποίο παρουσιάζεται στο [υποκεφάλαιο 4.1](#).

Επίσης εξετάζεται η χρήση υβριδικών ενσωματωμένων συστημάτων, τα οποία φέρουν μονάδες CPU αλλά και GPU (GPU+CPU SOC), σε εφαρμογές αναγνώρισης και εντοπισμού αντικειμένων σε πραγματικό χρόνο (Real-Time Object Detection).

...

1.3 ΔΙΑΡΘΡΩΣΗ ΤΗΣ ΑΝΑΦΟΡΑΣ

Η διάρθρωση της παρούσας διπλωματικής εργασίας είναι η εξής:

- **Κεφάλαιο 2:** Παρατίθεται η ανασκόπηση της ερευνητικής περιοχής σχετικά με τα αντικείμενα στα οποία επιδιώκει να προτείνει λύσεις η διπλωματική εργασία.
- **Κεφάλαιο 3:** Περιγράφονται τα βασικά θεωρητικά στοιχεία στα οποία βασίστηκαν οι υλοποιήσεις.
- **Κεφάλαιο 4:** Παρουσιάζονται οι διάφορες τεχνικές και τα εργαλεία που χρησιμοποιήθηκαν.
- **Κεφάλαιο 5:** Πλήρης περιγραφή των υλοποιήσεων.
- **Κεφάλαιο 6:** Παρουσιάζεται αναλυτικά η μεθοδολογία των πειραμάτων.
- **Κεφάλαιο 7:** Παρουσιάζονται τα τελικά συμπεράσματα.
- **Κεφάλαιο 8:** Αναφέρονται τα προβλήματα που προέκυψαν και προτείνονται θέματα για μελλοντική μελέτη, αλλαγές και επεκτάσεις.

2

Επισκόπηση της Ερευνητικής Περιοχής

Τόσο η αναγνώριση αντικειμένων (object recognition) όσο και ο εντοπισμός της θέσης των αντικειμένων αυτών (detection/localization) σε εικόνες, είναι μία ερευνητική περιοχή με τεράστιο ενδιαφέρον, η οποία απασχολεί πληθώρα ερευνητών. Η επιστήμη της Μηχανικής Όρασης (Computer Vision), στοχεύει στο να δώσει λύσεις στα συγκεκριμένα προβλήματα, εισάγοντας αναλυτικά ή και πιθανοτικά μαθηματικά μοντέλα.

Ο κλάδος της Βαθιάς Μηχανικής Μάθησης (Deep Learning - DL) [4], ανάγει το πρόβλημα της εύρεσης χαρακτηριστικών σημείων για την αναγνώριση αντικειμένων, στην εκμάθηση αναπαραστάσεων [5], με την χοήση Νευρωνικών Δικτύων Συνέλιξης (Convolutional Neural Networks - CNN). Οι πρώτες εφαρμογές Νευρωνικών Δικτύων Συνέλιξης, για την αναγνώριση αντικειμένων σε εικόνες, αναπτύχθηκαν το 1990 από τον Yann LeCun. Η πιο γνωστή και επιτυχής είναι το δίκτυο LeNet [6], το οποίο χρησιμοποιήθηκε για την αναγνώριση ψηφίων σε εικόνες. Ωστόσο, η εισαγωγή των CNN στον κλάδο της Μηχανικής Όρασης έγινε το 2012 με την ανάπτυξη του δικτύου AlexNet [2], από τους Alex Krizhevsky, Ilya Sutskever και Geoffrey E. Hinton. Το δίκτυο AlexNet χρησιμοποιήθηκε στον διαγωνισμό ImageNet ILSVRC challenge, το 2012, κερδίζοντας με διαφορά 10,9%, στο σφάλμα αναγνώρισης αντικειμένων σε σύνολο 1000 κλάσεων. Με την εμφάνιση του δικτύου AlexNet, η ερευνητική κοινότητα ξεκίνησε να πιστεύει στην αποτελεσματικότητα των Νευρωνικών Δικτύων Συνέλιξης σε εφαρμογές αναγνώρισης αντικειμένων σε εικόνες. Συνέχεια στο έργο του Alex Krizhevsky δόθηκε το 2013, αναπτύσσοντας το ZF-Net [7], το οποίο είναι βασισμένο στην αρχιτεκτονική του δικτύου AlexNet. Μέχρι σήμερα, έχουν σχεδιαστεί και αναπτυχθεί διάφορα μοντέλα Νευρωνικών Δικτύων Συνέλιξης για αναγνώριση αντικειμένων, με πιο πρόσφατο το ResNet, το οποίο αναπτύχθηκε από τον Kaiming He [8]. Το ResNet (Residual Network) έχει την ιδιαιτερότητα απουσίας πλήρως συνδεδεμένων επιπέδων και είναι από τα πιο δημοφιλή μοντέλα που εφαρμόζονται σε πρακτικά προβλήματα αναγνώρισης αντικειμένων σε εικόνες [9].

Τα προαναφερθέντα μοντέλα Νευρωνικών Δικτύων Συνέλιξης δίνουν λύσεις μόνο

στο πρόβλημα της αναγνώρισης και όχι του εντοπισμού της θέσης των αντικειμένων αυτών. Το 2013, ερευνητές εργάζομενοι στην Google Inc., σχεδίασαν και υλοποίησαν ένα μοντέλο Νευρωνικού Δικτύου το οποίο δίνει λύση στο πρόβλημα της ταυτόχρονης αναγνώρισης και εντοπισμού αντικειμένων πάνω σε πλαίσια εικόνων [10]. Το μοντέλο αυτό, που φέρει το όνομα DetectorNet, είναι ομαδική εργασία των Christian Szegedy, Alexander Toshev και Dumitru Erhan. Το μοντέλο αυτό είναι πιθανοτικό αφού ”προβλέπει” τις οριοθετημένες θέσεις για διάφορες κλάσεις αντικειμένων στον πλαίσιο μίας εικόνας. Ωστόσο, ένα βασικό μειονέκτημα του DetectorNet που το κάνει ακατάλληλο για εφαρμογή σε προβλήματα σχεδόν πραγματικού χρόνου (όπως για παράδειγμα σε ένα ρομποτικό σύστημα), είναι οι τεράστιες απαιτήσεις του σε υπολογιστικούς πόρους.

Την ίδια χρονιά (2013) η ερευνητική ομάδα του πανεπιστημίου New York University (Pierre Sermanet, David Eigen, Xiang Zhang, Michael Mathieu, Rob Fergus, Yann LeCun) εμφανίστηκε στον διαγωνισμό ILSVRC challenge 2013 με ένα μοντέλο CNN το οποίο έχει την μορφή ενός απλού δικτύου και το οποίο έχει την ικανότητα εκμάθησης και των πλαισίων (bounding boxes) στα οπία ανήκουν τα αντικείμενα που αναγνωρίζονται. Το μοντέλο φέρει το όνομα OverFeat [?] και ήταν ο νικητής του διαγωνισμού στην κατηγορία εντοπισμού αντικειμένων στον συγκεκριμένο διαγωνισμό.

Τα πιο πρόσφατα και πλέον δημοφιλή σήμερα νευρωνικά δίκτυα για ταυτόχρονη αναγνώριση και εντοπισμό αντικειμένων σε εικόνες είναι τα μοντέλα ResNet [8], Fast-RCNN [?] και YOLO (You Only Look Once) [?], με τα τελευταία δύο να επικεντρώνονται στο πρόβλημα της γρήγορης εκτέλεσης. Συγκεκριμένα το δίκτυο YOLO έχει την ικανότητα να επεξεργαστεί πλαίσια με ταχύτητα 45 fps (frames per second) σε μονάδα GPU (NVIDIA Titan).

3

Τεχνικές Βαθιάς Μηχανικής Μάθησης και Αναγνώρισης Αντικειμένων στον χώρο

Τα τελευταία χρόνια, ο κλάδος της Τεχνητής Νοημοσύνης είναι ένας από τους πιο ραγδαία αναπτυσσόμενους κλάδους της επιστήμης της πληροφορικής με τεράστιο ερευνητικό και πρακτικό ενδιαφέρον. Διαιρείται σε δύο υποκατηγορίες; την συμβολική τεχνητή νοημοσύνη και την υποσυμβολική τεχνητή νοημοσύνη. Η πρώτη προσπαθεί να επιλύσει προβλήματα χρησιμοποιώντας αλγορίθμικές διαδικασίες, δηλαδή σύμβολα και λογικούς κανόνες, ενώ η δεύτερη προσπαθεί να αναπαράγει την ανθρώπινη "ευφυΐα" μέσα από την χρήση αριθμητικών μοντέλων που με την σύνθεσή τους προσομοιώνουν την λειτουργία του ανθρώπινου εγκεφάλου (υπολογιστική νοημοσύνη).

Η ικανότητα ενός νοούμενου (AI) συστήματος να αποκτά από μόνο του γνώση, εξάγοντας πρότυπα ή/και χαρακτηριστικά σημεία από τα δεδομένα, είναι γνωστή ως *Μηχανική Μάθηση* (ML).

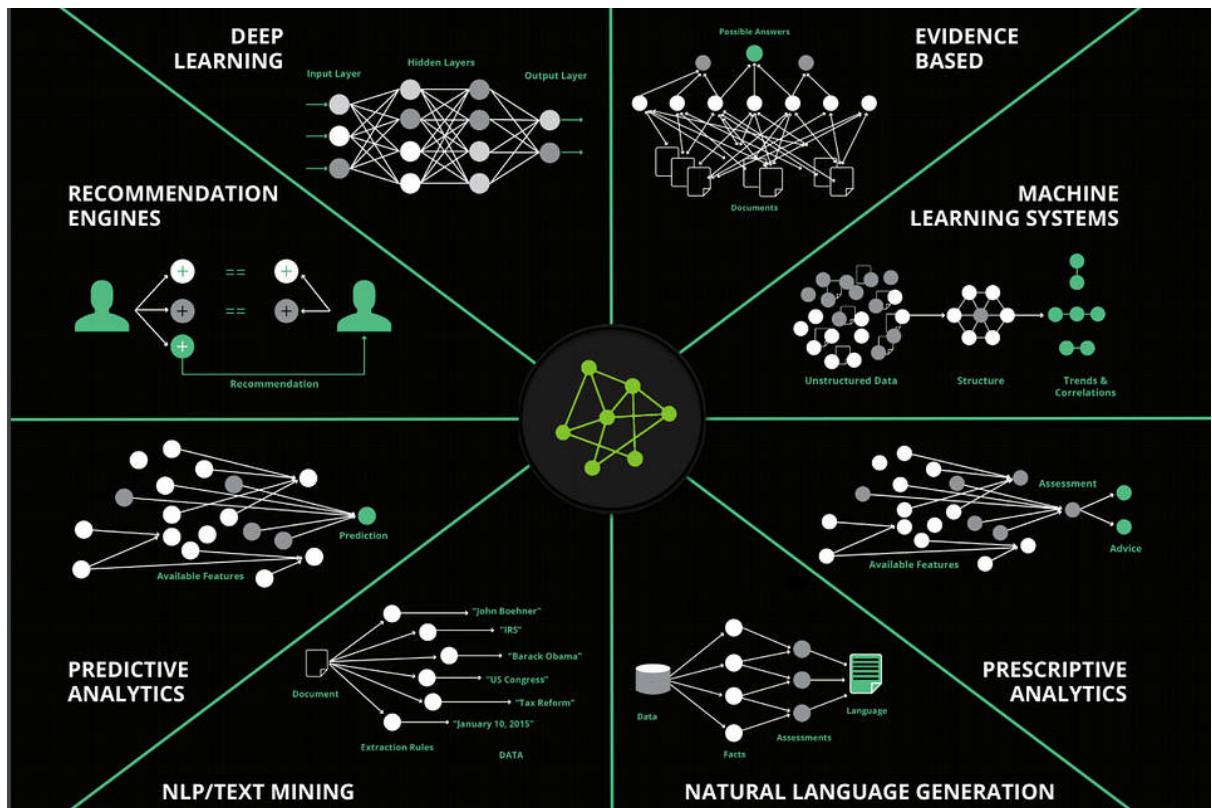
Στο κεφάλαιο αυτό, παρουσιάζονται και αναλύονται τεχνικές και αλγόριθμοι *Μηχανικής Μάθησης*, με επίκεντρο τα νευρωνικά δίκτυα με βάθος (Deep Neural Networks - DNN). Στόχος είναι ο αναγνώστης να αντιληφθεί και να κατανοήσει τις βασικές αρχές και λειτουργίες των νευρωνικών αυτών δικτύων, αφού είναι οι βάσεις για την περαιτέρω μελέτη των νευρωνικών δικτύων συνέλιξης (CNN) και των εφαρμογών αυτών στο πρόβλημα της αναγνώρισης και εντοπισμού αντικειμένων σε εικόνες.

3.1. ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΕΠΙΣΤΗΜΗ ΤΗΣ ΣΥΓΧΡΟΝΗΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ

3.1 ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΕΠΙΣΤΗΜΗ ΤΗΣ ΣΥΓΧΡΟΝΗΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ

Η εισαγωγή του κλάδου της μηχανικής μάθησης στην επιστήμη των υπολογιστών, επέτρεψε στους υπολογιστές να μπορούν να αντιμετωπίσουν προβλήματα που εμπλέκουν την αντίληψη για τον πραγματικό κόσμο, όσο και να παίρνουν υποκειμενικές αποφάσεις.

Οι αλγόριθμοι ML, επιτρέπουν σε συστήματα AI να προσαρμόζονται εύκολα σε καινούργια προβλήματα, απαιτώντας ελάχιστη επέμβαση από τον άνθρωπο. Για παράδειγμα, ένα Νευρωνικό Δίκτυο που έχει εκπαιδευτεί να αναγνωρίζει γάτες σε εικόνες, δεν απαιτεί να σχεδιαστεί και να εκπαιδευτεί από το μηδέν για να έχει την ικανότητα να αναγνωρίζει και σκύλους.



Σχήμα 3.1: Κλάδοι και εφαρμογές της επιστήμης της Τεχνητής Νοημοσύνης

Πολλά προβλήματα, τα οποία μέχρι και πριν μία μερικά χρόνια λύνονταν με "χειρόγραφη", προγραμματισμένη από τον άνθρωπο γνώση, σήμερα επιλύονται με χρήση αλγορίθμων ML (σχήμα 3.1). Κάποια παραδείγματα ακολουθούν:

- Αναγνώριση ομιλίας - Speech Recognition
- Μηχανική όραση - Computer Vision
 - Αναγνώριση αντικειμένων σε εικόνες - Object Recognition

ΚΕΦΑΛΑΙΟ 3. ΤΕΧΝΙΚΕΣ ΒΑΘΙΑΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ ΚΑΙ ΑΝΑΓΝΩΡΙΣΗΣ ΑΝΤΙΚΕΙΜΕΝΩΝ ΣΤΟΝ ΧΩΡΟ

- Αναγνώριση και εντοπισμός της θέσης αντικειμένων σε εικόνες - Object Detection
- Αναγνώριση ηλεκτρονικών επιθέσεων στο διαδίκτυο - Cyberattack detection
- Επεξεργασία φυσικής γλώσσας - Natural Language Processing
 - Κατανόηση της φυσικής γλώσσας του ανθρώπου - Natural Language Understanding
 - Μοντελοποίηση και χρήση της φυσικής γλώσσας του ανθρώπου από μηχανές - Natural Language Generation
- Μηχανές αναζήτησης - Search Engines
- Αναπαράσταση γνώσης - Knowledge Representation
- Ρομποτική

Τα προβλήματα Μηχανικής Μάθησης χωρίζονται σε τρεις μεγάλες κατηγορίες:

- Ύπό επίβλεψη Μάθηση - Supervised Learning: Στο υπολογιστικό σύστημα δίνονται παραδείγματα εισόδου και επιθυμητής εξόδου, δηλαδή στα δεδομένα έχουν προηγουμένως ανατεθεί ετικέτες (labels), και στόχος είναι να εξαχθεί ένας γενικός κανόνας αντιστοίχισης της εισόδου στην επιθυμητή έξοδο. Η αναγνώριση προκαθορισμένων αντικειμένων σε εικόνες είναι ένα πρόβλημα που ανήκει σε αυτή την κατηγορία.
- Χωρίς επίβλεψη Μάθηση - Unsupervised Learning: Τα δεδομένα δεν έχουν ετικέτες (labels), αφήνοντας έτσι τον αλγόριθμο ML να βρει από μόνος του δομές στα δεδομένα εισόδου.
- Εκμάθηση δια ανταμοιβής - Reinforcement Learning: Ο πράκτορας αλληλεπιδρά με ένα δυναμικό περιβάλλον στο οποίο πρέπει να εκτελέσει ένα συγκεκριμένο στόχο, χωρίς την ύπαρξη ενός "δασκάλου" που να ορίζει ρητά αν έχει φθάσει κοντά στον στόχο. Ένα παράδειγμα εφαρμογής είναι η αυτόματη πλοήγηση ενός οχήματος.

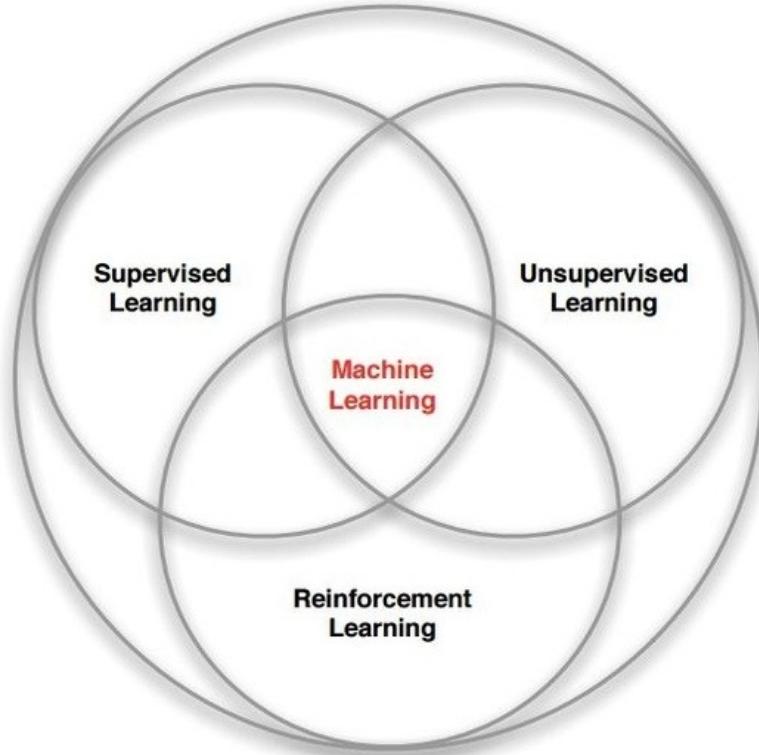
Επιπλέον, κάποια προβλήματα είναι υβριδικά, δηλαδή συνδυασμός των πιο πάνω. Στο [σχήμα 3.2](#) απεικονίζεται το διάγραμμα Venn των διαφόρων αλγορίθμικών κατηγοριών ML.

Περαιτέρω, οι Supervised Learning αλγόριθμοι χωρίζονται σε 2 κατηγορίες, ανάλογα με την επιθυμητή μορφή της εξόδου του αλγόριθμου ML:

- Ταξινόμησης - Classification: Όταν η έξοδος παίρνει διακριτές τιμές (discrete).
- Regression: Όταν η έξοδος παίρνει συνεχείς τιμές.

Γενικότερα, οι αλγόριθμοι ML ομαδοποιούνται και ανάλογα με την ομοιότητα τους σε σχέση με την λειτουργία που εκτελούν. Πιο κάτω αναφέρονται οι πιο δημοφιλείς αλγόριθμοι ML, ομαδοποιημένοι με βάση την λειτουργία τους.

3.1. ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΕΠΙΣΤΗΜΗ ΤΗΣ ΣΥΓΧΡΟΝΗΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ

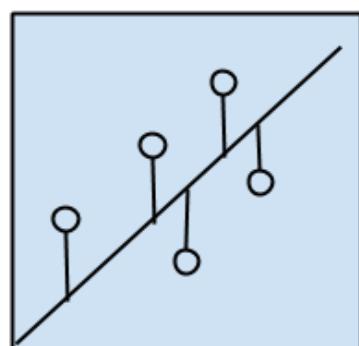


Σχήμα 3.2: Διάγραμμα Venn των διαφόρων κατηγοριών μηχανικής μάθησης

Regression

Ασχολείται με τη μοντελοποίηση της σχέσης μεταξύ των μεταβλητών που ανανεώνονται επαναληπτικά χρησιμοποιώντας ένα μέτρο σφάλματος για τις προβλέψεις που γίνονται από το μοντέλο

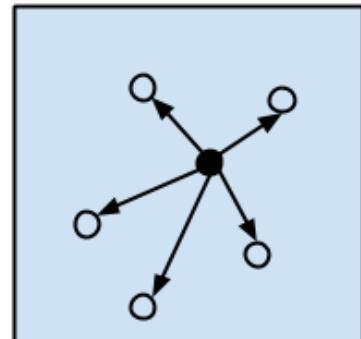
- Ordinary Least Squares Regression (OLSR)
- Linear Regression
- Logistic Regression
- Stepwise Regression
- Multivariate Adaptive Regression Splines (MARS)
- Locally Estimated Scatterplot Smoothing (LOESS)



Instance-based

Αυτές οι μέθοδοι δημιουργούν μία βάση με παραδείγματα δεδομένων και συγκρίνουν τις νέες εισόδους με αυτές που έχουν καταχωρηθεί στην βάση δεδομένων χρησιμοποιώντας ένα μέτρο ομοιότητας, για την πιθανοτική εύρεση της καλύτερης αντιστοιχίας.

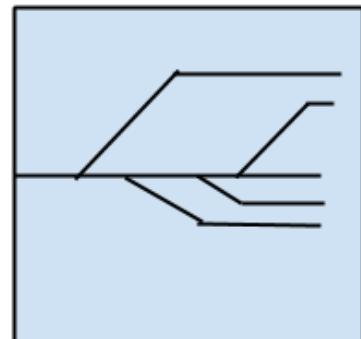
- k-Nearest Neighbour (kNN)
- Learning Vector Quantization (LVQ)
- Self-Organizing Map (SOM)
- Locally Weighted Learning (LWL)



Regularization

Χρησιμοποιούνται σαν επεκτάσεις άλλων μεθόδων και "τιμωρούν" πολύπλοκα μοντέλα, ευνοώντας έτσι απλούστερα μοντέλα τα οποία είναι συνήθως καλύτερα στην γενίκευση της επίλυσης του εκάστοτε προβλήματος.

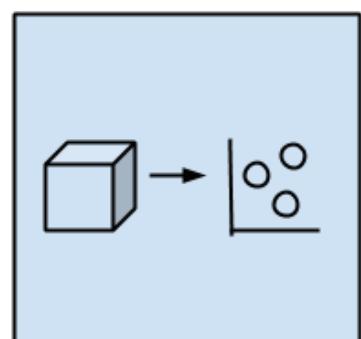
- Ridge Regression
- Least Absolute Shrinkage and Selection Operator (LASSO)
- Least-Angle Regression (LARS)
- Elastic Net



Dimensionality Reduction

Χρησιμοποιούνται για την αφαίρεση ασήμαντης πληροφορίας από τα δεδομένα. Πολλές από τις μεθόδους αυτές χρησιμοποιούνται σαν επεκτάσεις σε μοντέλα επίλυσης προβλημάτων regression και classification

- Principal Component Analysis (PCA)
- Discriminant Analysis: Linear (LDA), Mixture (MDA), Quadratic (QDA), Flexible (FDA)
- Principal Component Regression (PCR)
- Multidimensional Scaling (MDS)

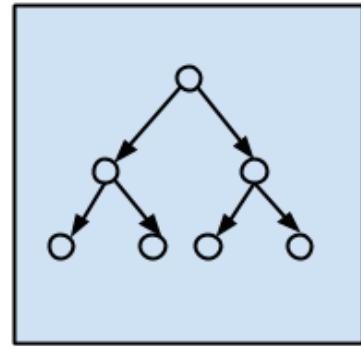


3.1. ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΕΠΙΣΤΗΜΗ ΤΗΣ ΣΥΓΧΡΟΝΗΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ

Decision Trees

Χρησιμοποιούνται για την κατασκευή μοντέλων λήψης αποφάσεων, τα οποία χρησιμοποιούν τις πραγματικές τιμές των χαρακτηριστικών των δεδομένων.

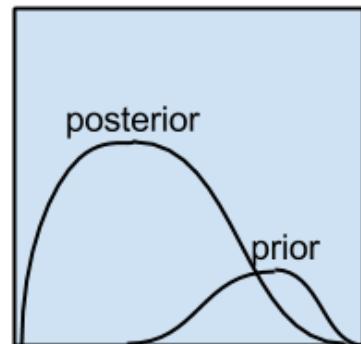
- Classification and Regression Tree (CART)
- Conditional Decision Trees
- M5



Bayesian

Εφαρμόζουν το θεώρημα του Bayes για την επίλυση τόσο προβλημάτων regression, αλλά και classification

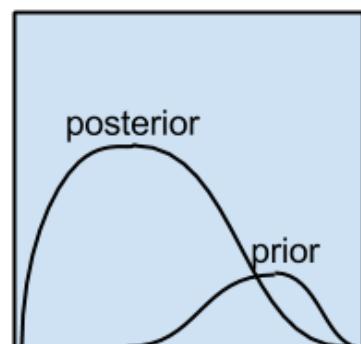
- Naive Bayes
- Gaussian Naive Bayes
- Bayesian Network (BN)
- Bayesian Belief Network (BBN)



Clustering

Περιγράφουν τις κλάσεις του προβλήματος

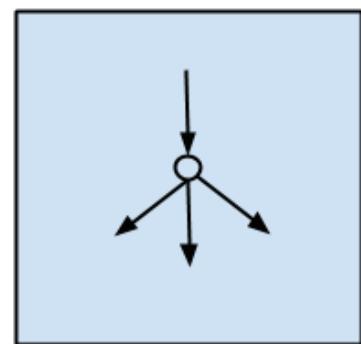
- k-Means
- k-Medians
- Expectation Maximisation (EM)
- Hierarchical Clustering



Artificial Neural Networks (ANN)

Μοντέλα εμπνευσμένα από τη δομή ή/και την λειτουργία των βιολογικών νευρωνικών δικτύων. Χρησιμοποιούνται στην επίλυση προβλημάτων classification ή/και regression.

- Perceptron
- Back-Propagation
- Radial Basis Function Network (RBFN)

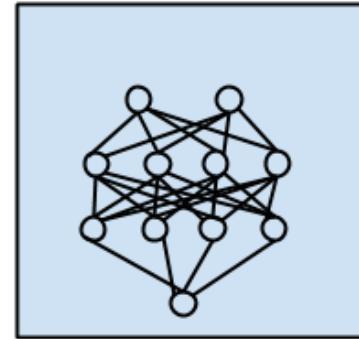


ΚΕΦΑΛΑΙΟ 3. ΤΕΧΝΙΚΕΣ ΒΑΘΙΑΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ ΚΑΙ ΑΝΑΓΝΩΡΙΣΗΣ ΑΝΤΙΚΕΙΜΕΝΩΝ ΣΤΟΝ ΧΩΡΟ

Deep Learning (DL)

Οι αλγόριθμοι DL είναι η σύγχρονη επέκταση των ANN, τα οποία εκμεταλλεύονται την αφθονία υπολογιστικής ισχύς των σύγχρονων υπολογιστικών συστημάτων.

- Autoncoder
- Multilayer Perceptron (MLP)
- Deep Boltzmann Machine (DBM)
- Deep Belief Networks (DBN)
- Convolutional Neural Network (CNN)
- Stacked Auto-Encoders
- Recurrent Neural Networks (RNN)

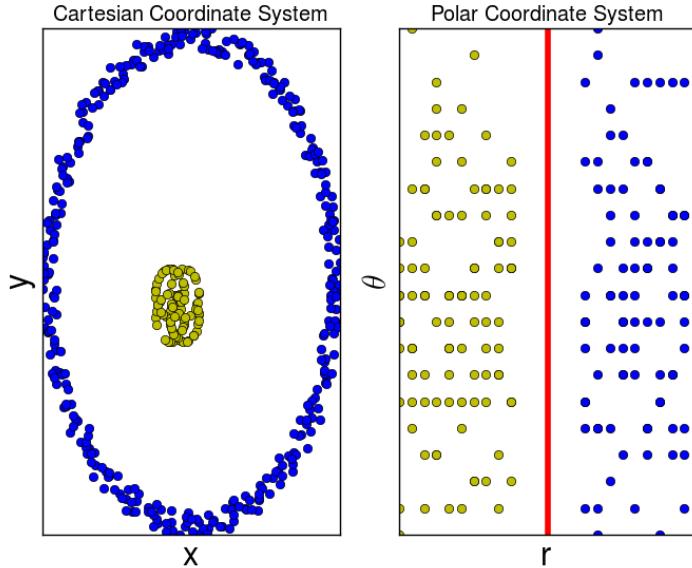


Η μορφή της αναπαράστασης των δεδομένων αποτελεί σημαντικό παράγοντα στην απόδοση των αλγορίθμων ML. Μία αναπαράσταση αποτελείται από χαρακτηριστικά (features). Για παράδειγμα, ένα χρόσιμο χαρακτηριστικό, στην ταυτοποίηση ομιλητή, από δεδομένα ήχου, είναι η εκτίμηση του μεγέθους της φωνητικής έκτασης του ομιλητή. Έτσι, πολλά προβλήματα τεχνητής νοημοσύνης, μπορούν να λυθούν με κατάλληλη σχεδίαση και επιλογή των χαρακτηριστικών για το συγκεκριμένο πρόβλημα. Το σύνολο των χαρακτηριστικών αυτών αποτελεί την αναπαράσταση των δεδομένων σε ένα πιο υψηλό και αφαιρετικό επίπεδο αντίληψης για τους υπολογιστές, η οποία στην συνέχεια δίνεται σαν είσοδος σε έναν απλό ML αλγόριθμο, ο οποίος έχει μάθει να αντιστοιχεί την αναπαράσταση των δεδομένων στην επιθυμητή έξοδο.

Ένα απλό και κατανοητό παράδειγμα το οποίο δείχνει την εξάρτηση της επίδοσης των αλγορίθμων ML από την μορφή της αναπαράστασης που του δίνεται, φαίνεται στο **σχήμα 3.3**. Έστω ότι θέλουμε να διαχωρίσουμε τα δεδομένα μας σε δύο κλάσεις, χαράζοντας μία ευθεία μεταξύ τους. Αν αναπαραστήσουμε τα δεδομένα στο Καρτεσιανό σύστημα συντεταγμένων (αριστερό διάγραμμα), τότε η επίλυση του προβλήματος είναι αδύνατη αφού δεν υπάρχει καμία ευθεία που να διαχωρίζει τις δύο κλάσεις. Ωστόσο, αν αναπαραστήσουμε τα δεδομένα στο πολικό σύστημα συντεταγμένων (δεξιό διάγραμμα), τότε το πρόβλημα λύνεται εύκολα, χαράζοντας μία κάθετη ευθεία, με $r = a$, $a \in [r_1, r_2]$.

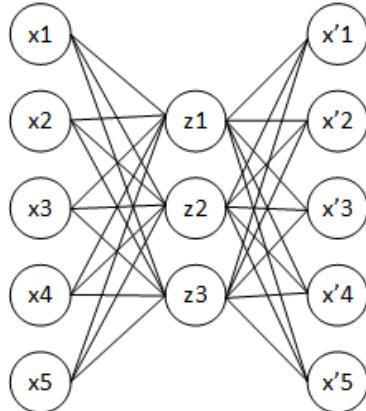
Σε πληθώρα προβλημάτων τεχνητής νοημοσύνης, η επιλογή κατάλληλων χαρακτηριστικών είναι δύσκολο και χρονοβόρο έργο. Έστω ότι θέλουμε να αναγνωρίσουμε πρόσωπα σε εικόνες. Ένα χαρακτηριστικό θα μπορούσε να είναι τα μάτια. Δυστυχώς όμως, η αναγνώριση ματιών είναι και αυτό ένα δύσκολο πρόβλημα, αφού δεν μπορεί να περιγραφεί πάντα επακριβώς έχοντας σαν δεδομένα τις τιμές των pixel της εικόνας. Η γεωμετρική, για παράδειγμα, μορφή των ματιών σε μία εικόνα λήψης εξαρτάται από την γωνία λήψης της εικόνας, τον φωτισμό, τις ανακλάσεις του φωτισμού, την απόσταση από την οποία γίνεται η λήψη, την ανάλυση της κάμερας, κτλ.

3.1. ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΕΠΙΣΤΗΜΗ ΤΗΣ ΣΥΓΧΡΟΝΗΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ



Σχήμα 3.3: Παράδειγμα διαφορετικών αναπαραστάσεων των δεδομένων

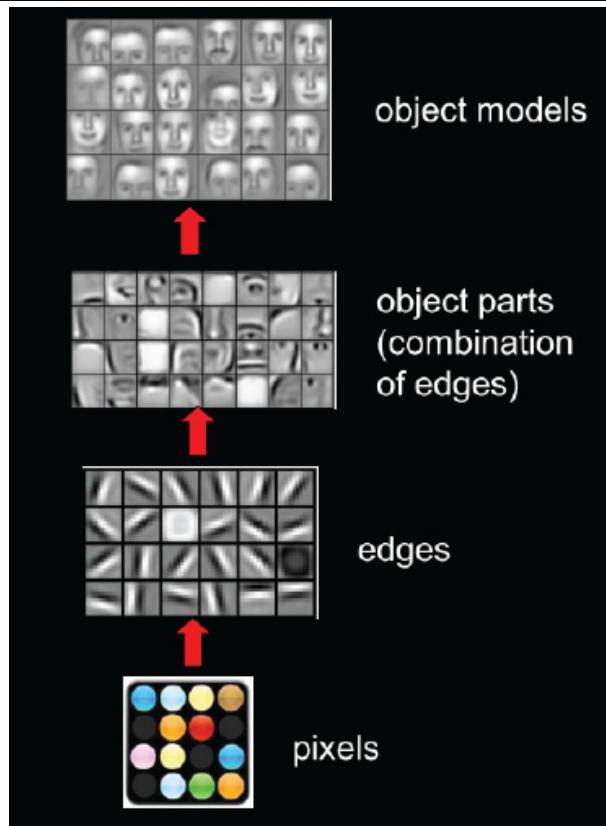
Το πρόβλημα αυτό, της επιλογής κατάλληλης αναπαράστασης, μπορεί να λυθεί χρησιμοποιώντας τεχνικές μηχανικής μάθησης για την εκμάθηση της ίδιας της αναπαράστασης. Αυτή η προσέγγιση είναι γνωστή ως *Εκμάθηση Αναπαραστάσεων* (*Representation Learning*). Οι αλγόριθμοι εκμάθησης αναπαραστάσεων είναι ικανοί να "μάθουν" ένα καλό σετ χαρακτηριστικών (features). Ένα απλό παράδειγμα αλγο-



Σχήμα 3.4: Απλό μοντέλο Autoencoder με ένα κρυφό επίπεδο.

ρίθμου εκμάθησης αναπαραστάσεων είναι αυτό του Autoencoder [11] που φαίνεται στο (σχήμα 3.4). Ο Autoencoder, στην πιο απλή του μορφή, είναι ο συνδυασμός ενός κωδικοποιητή (encoder) ο οποίος μετατρέπει τα δεδομένα εισόδου σε μία διαφορετική αναπαράσταση, και ενός αποκωδικοποιητή ο οποίος επαναφέρει την αναπαράσταση αυτή στην αρχική μορφή της αναπαράστασης των δεδομένων εισόδου. Οι Autoencoders ανήκουν στην κατηγορία των Νευρωνικών Δικτύων και είναι Unsupervised ML αλγόριθμοι. Ένα συχνά εμφανιζόμενο πρόβλημα σε εφαρμογές τεχνητής νοημοσύνης είναι η εύρεση και εξαγωγή χαρακτηριστικών υψηλού επιπε-

ΚΕΦΑΛΑΙΟ 3. ΤΕΧΝΙΚΕΣ ΒΑΘΙΑΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ ΚΑΙ ΑΝΑΓΝΩΡΙΣΗΣ ΑΝΤΙΚΕΙΜΕΝΩΝ ΣΤΟΝ ΧΩΡΟ



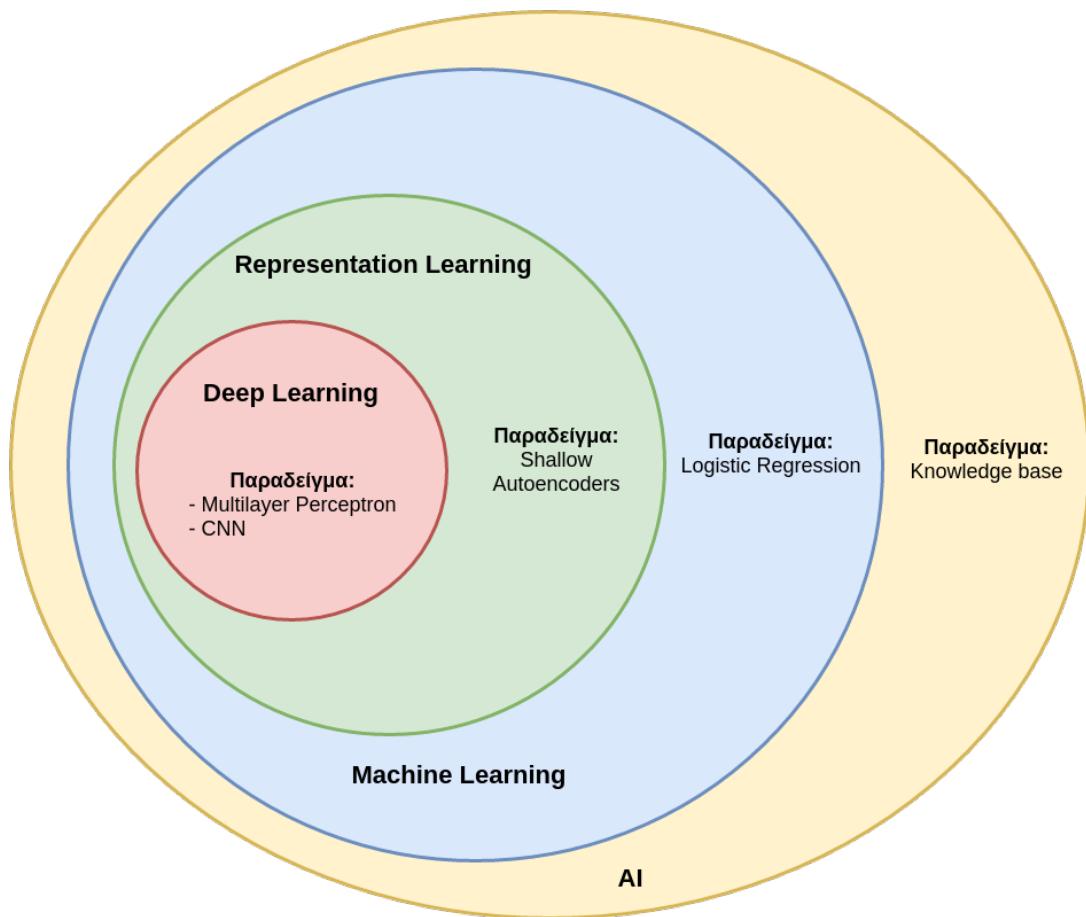
Σχήμα 3.5: Παράδειγμα απεικόνισης των φίλτρων ενός μοντέλου CNN για αναγνώριση προσώπου

δου από τα δεδομένα. Τα μοντέλα *Βαθιάς Μηχανικής Μάθησης* δίνουν λύσεις σε αυτό το πρόβλημα εκμάθησης αναπαραστάσεων με την εισαγωγή χαρακτηριστικών τα οποία εκφράζονται με βάση άλλες, απλούστερες αναπαραστάσεις. Αυτή η προσέγγιση δίνει την δυνατότητα στους υπολογιστές να κατασκευάζουν σύνθετες έννοιες χρησιμοποιώντας απλούστερες. Για παράδειγμα, η αναγνώριση αντικειμένων μπορεί να εκφραστεί με έννοιες όπως το γεωμετρικό σχήμα των αντικειμένων, το οποίο με την σειρά του ορίζεται από γωνίες και περιγράμματα. Επίσης, οι γωνίες και τα περιγράμματα ορίζονται από ακμές. Στο [σχήμα 3.5](#), παρουσιάζονται τα φίλτρα που έμαθε ένα μοντέλο CNN για αναγνώριση προσώπων σε εικόνες. Το συγκεκριμένο μοντέλο CNN έχει 3 κρυφά επίπεδα (hidden layers); το πρώτο κρυφό επίπεδο εξάγει από τα δεδομένα εισόδου (τιμές των πίξελ) πληροφορία σχετικά με τις ακμές, το δεύτερο, έχοντας σαν είσοδο την πληροφορία παρουσίας ακμών, εξάγει πληροφορία σχετικά με τις γωνίες και τα περιγράμματα και το τρίτο παίρνει σαν είσοδο την πληροφορία αυτή και κατασκευάζει μοντέλα αντικειμένων, δηλαδή εξάγει πληροφορία σχετικά με το γεωμετρικό σχήμα των αντικειμένων.

Συνοψίζοντας, η βαθιά μηχανική μάθηση, είναι μία υποκατηγορία του ML και ένας χαρακτηριστικός αντιπρόσωπος της σύγχρονης τεχνητής νοημοσύνης. Πιο συγκεκριμένα, είναι ένα είδος μηχανικής μάθησης, η οποία προσδίδει σε υπολογιστικά συστήματα ευφυΐα, δηλαδή την ικανότητα εκμάθησης με την χρήση εμπειρίας και δεδομένων. Σύμφωνα με τους *Ian Goodfellow, Yoshua Bengio και Aaron Courville*, η

3.2. ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ ΜΕ ΒΑΘΟΣ

μηχανική μάθηση είναι η μόνη βιώσιμη προσέγγιση στην κατασκευή συστημάτων AI τα οποία μπορούν να αντεπεξέλθουν σε πολύπλοκα περιβάλλοντα και προβλήματα [4]. Η βαθιά μηχανική μάθηση καταφέρνει να μαθαίνει να αναπαριστά τον κόσμο ως μία ένθετη ιεραρχία εννοιών όπου η κάθε έννοια ορίζεται σε σχέση με άλλες πιο απλές έννοιες, και πιο αφηρημένες μορφές αναπαραστάσεων σε σχέση με λιγότερο αφηρημένες. Από το διάγραμμα Venn που βλέπουμε στο [σχήμα 3.6](#) παρατηρούμε ότι η βαθιά μηχανική μάθηση ανήκει στην κατηγορία της εκμάθησης αναπαραστάσεων, η οποία με την σειρά της είναι ένα είδος μηχανικής μάθησης που χρησιμοποιείται για την κατασκευή νοούμενων συστημάτων.

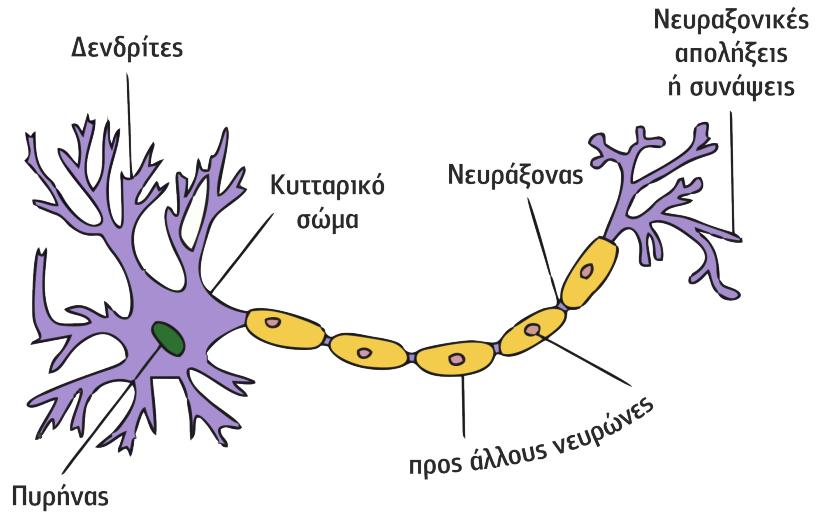


Σχήμα 3.6: Παράδειγμα απεικόνισης των φίλτρων ενός μοντέλου CNN για αναγνώριση προσώπου

3.2 ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ ΜΕ ΒΑΘΟΣ

Τα Νευρωνικά Δίκτυα είναι εμπνευσμένα από το βιολογικό νευρικό σύστημα του ανθρώπου. Η βασική επεξεργαστική μονάδα του εγκεφάλου είναι ο νευρώνας ([σχήμα 3.7](#)), ενώ το ανθρώπινο νευρικό σύστημα αποτελείται από περίπου 86 εκατομμύρια νευρώνες και περίπου $10^{14} - 10^{15}$ διασυνδέσεις. Τα κύρια μέρη ενός νευρώνα είναι τα εξής:

ΚΕΦΑΛΑΙΟ 3. ΤΕΧΝΙΚΕΣ ΒΑΘΙΑΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ ΚΑΙ ΑΝΑΓΝΩΡΙΣΗΣ ΑΝΤΙΚΕΙΜΕΝΩΝ ΣΤΟΝ ΧΩΡΟ



Σχήμα 3.7: Βιολογικός νευρώνας.

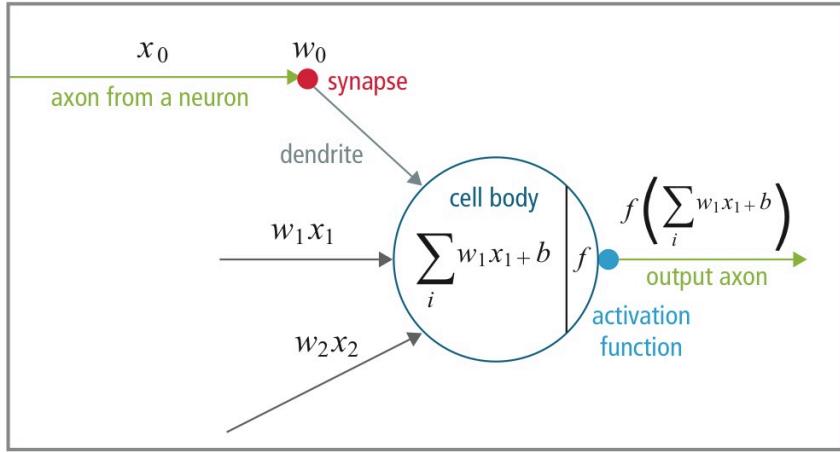
- **Δενδρίτης (Dendrites):** Δέχεται είσοδο από άλλους νευρώνες.
- **Σώμα του κυττάρου (Cell body):** Εξάγει συμπεράσματα, με βάση τις εισόδους.
- **Νευράξονας (Axon):** Συνδέει την έξοδο που λαμβάνεται από το σώμα του κυττάρου με τις νευραξονικές απολήψεις.
- **Νευραξονικές απολήψεις:** Συνδέει τον νευράξονα του εκάστοτε νευρώνα με τους τερματικούς κόμβους, από όπου και μεταφέρεται η πληροφορία στην είσοδο άλλων νευρώνων.

Κάθε νευρώνας δέχεται είσοδο από άλλους νευρώνες μέσω των δενδριτών του και στην συνέχεια επεξεργάζεται το σήμα που λαμβάνει στην είσοδο και στέλνει το αποτέλεσμα στον νευράξονα. Τέλος, άλλοι νευρώνες συνδέονται με αυτόν μέσω των συνάψεων του.

Το αντίστοιχο μαθηματικό μοντέλο του νευρώνα, φαίνεται στο [σχήμα 3.8](#). Η πληροφορία που μεταφέρεται από τις νευραξονικές απολήψεις (x_0), προτού μεταφερθεί στους δενδρίτες των επόμενων νευρώνων, αλληλεπιδρά πολλαπλασιαστικά με τις συνάψεις ($w_0 * x_0$). Οι παράγοντες πολλαπλασιασμού w_n ονομάζονται βάρη και αποτελούν τις μεταβλητές παραμέτρους ενός νευρώνα. Η τιμή των παραμέτρων αυτών ελέγχει την επίδραση μεταξύ των νευρώνων. Η συνάρτηση ενεργοποίησης f ελέγχει την ροή της πληροφορίας στους συνδεδεμένους νευρώνες και προσδίδει ευελιξία και ικανότητα εκτίμησης όσον αφορά πολύπλοκες μη γραμμικές σχέσεις στα δεδομένα εισόδου. Η έξοδος από ένα νευρώνα υπολογίζεται από τη σχέση:

$$a = f\left(\sum_{i=0}^N w_i x_i + b\right)$$

Η πιο απλή μορφή συνάρτησης ενεργοποίησης είναι η σιγμοειδής συνάρτηση $\sigma(x) = 1/(1 + e^{-x})$. Εναλλακτικά, η σιγμοειδής συνάρτηση ενεργοποίησης μπορεί να



Σχήμα 3.8: Μαθηματικό μοντέλο του νευρώνα

εκφραστεί σε διακριτή μορφή ως

$$f(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}$$

Η επιλογή κατάλληλης συνάρτησης ενεργοποίησης των νευρώνων δεν είναι τυχαία, αφού όπως θα διόμε στην συνέχεια επηρεάζει την απόδοση του αλγορίθμου *Backpropagation*, ο οποίος χρησιμοποιείται για την εκπαίδευση των νευρωνικών δικτύων.

Γενικότερα, ο νευρώνας μπορεί να είναι και πολωμένος (biased - b) και έτσι το μαθηματικό μοντέλο που τον περιγράφει πλήρως παίρνει την μορφή:

$$\sigma = f(\sum_i w_i x_i + b) = \frac{1}{1+e^{-\sum_i w_i x_i + b}}$$

Θα μπορούσαμε να ερμηνεύσουμε το αποτέλεσμα της εφαρμογής της σιγμοειδούς συνάρτησης ενεργοποίησης ως την πιθανότητα μίας από τις κλάσεις:

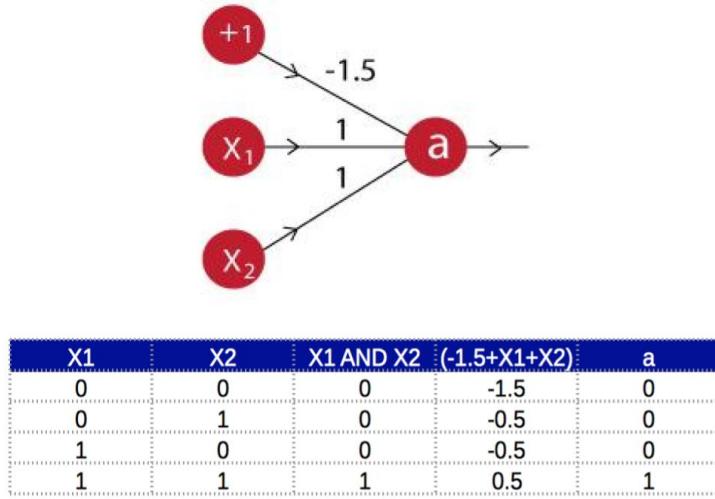
$$P(y_i = 1|x_i; w)$$

$$P(y_i = 0|x_i; w) = 1 - P(y_i = 1|x_i; w)$$

Σαν παρατήρηση, με εφαρμογή κατάλληλης συνάρτησης σφάλματος στην έξοδο, ο νευρώνας έχει την συμπεριφορά ενός γραμμικού ταξινομητή (linear classifier). Πιο συγκεκριμένα, σε περίπτωση που χρησιμοποιήσουμε την cross-entropy συνάρτηση σφάλματος ο νευρώνας μετατρέπεται σε δυαδικό ταξινομητή **Softmax**, τον οποίο και θα συναντήσουμε στην συνέχεια.

Οι τρεις θεμελιώδεις εφαρμογές του μαθηματικού μοντέλου του νευρώνα είναι η μοντελοποίηση των λογικών πυλών *AND*, *OR* και *NOT*. Στο [σχήμα 3.9](#) παρουσιάζεται το αντίστοιχο μοντέλο της λογικής πύλης *AND*. Ο νευρώνας δέχεται σαν είσοδο

ΚΕΦΑΛΑΙΟ 3. ΤΕΧΝΙΚΕΣ ΒΑΘΙΑΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ ΚΑΙ ΑΝΑΓΝΩΡΙΣΗΣ ΑΝΤΙΚΕΙΜΕΝΩΝ ΣΤΟΝ ΧΩΡΟ



Σχήμα 3.9: Υλοποίηση πύλης AND με χρήση του μαθηματικού μοντέλου του νευρώνα

2 σήματα (X_1, X_2) και μία παράμετρο πόλωσης ($b = -1.5$). Η έξοδος a ορίζεται ως:

$$a = f(x) = f(X_1, X_2) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}$$

$$x = X_1 + X_2 - 1.5$$

Η σύνδεση πολλών νευρώνων σε σε ένα γράφο δομεί ένα *Νευρωνικό Δίκτυο - NN*. Το μοντέλο NN που φαίνεται στο [σχήμα 3.10](#) ονομάζεται *Perceptron* ή αλλιώς *Feedforward Artificial Neural Network - ANN*. Feedforward γιατί η πληροφορία ρέει προς μία μόνο κατεύθυνση, δηλαδή η έξοδος νευρώνα στο i επίπεδο δεν συνδέεται με την είσοδο νευρώνα που βρίσκεται το επίπεδο $k \leq i$.

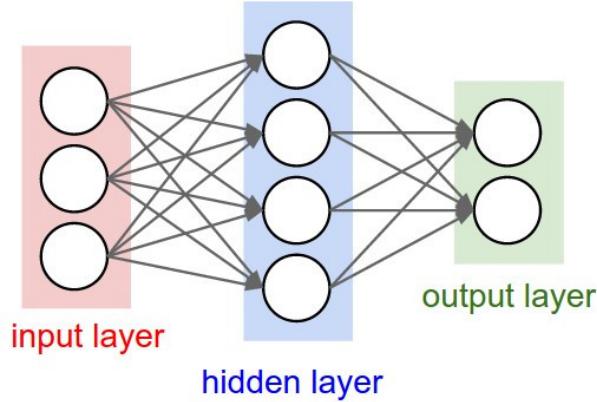
Ένα ANN έχει τα εξής χαρακτηριστικά:

- Οι διασυνδέσεις μεταξύ των νευρώνων δεν σχηματίζουν σε καμία περίπτωση κύκλο.
- Αποτελείται από 2 ή περισσότερα κρυφά επίπεδα; ένα κρυφό και το επίπεδο εξόδου
- Χρησιμοποιείται η σιγμοειδής συνάρτηση ενεργοποίησης

Το Perceptron δεν είναι το μόνο μοντέλο NN που ανήκει στην κατηγορία των Feedforward ANN. Όπως θα δούμε στο [υποκεφάλαιο 3.3](#), τα *Νευρωνικά Δίκτυα Συνέλιξης (Convolutional Neural Networks - CNN)* ανήκουν και αυτά στην κατηγορία αυτή.

Η γενική δομή ενός νευρωνικού δικτύου φαίνεται στο [σχήμα 3.11](#). Τα γνωρίσματα ενός τέτοιου, πολυ-επίπεδου NN, είναι τα εξής:

- Αριθμός κρυφών επιπέδων



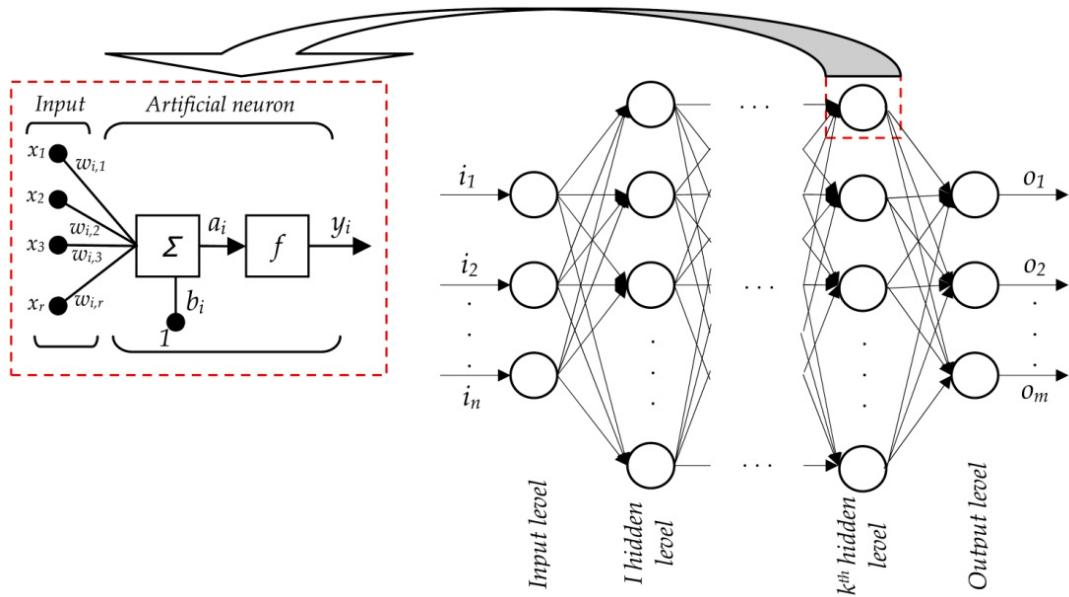
Σχήμα 3.10: Απλό μοντέλο NN με ένα κρυφό επίπεδο - Perceptron

- Αριθμός των νευρώνων στο κάθε επίπεδο

Η έξοδος από το εκάστοτε επίπεδο μπορεί να εκφραστεί ως

$$A_{i+1} = f_i(A_i \bullet W_i + B_i)$$

όπου ο πίνακας A_i αναφέρεται στην είσοδο και έχει διαστάσεις $M \times N$, W είναι ο πίνακας με τα βάρη των νευρώνων του εκάστοτε επιπέδου με διαστάσεις $K \times M$, και τέλος ο πίνακας B διαστάσεων $K \times N$ αναφέρεται στις τιμές πόλωσης. Η τιμή i αναφέρεται στον αριθμό του εκάστοτε επιπέδου του NN. Ο αριθμός των επιπέδων, ή καλύτερα το μέγιστο μήκος του μονοπατιού που ακολουθεί η πληροφορία από την είσοδο μέχρι την έξοδο του NN, ορίζει το βάθος ενός NN.



Σχήμα 3.11: Μορφή ενός πολυεπίπεδου νευρωνικού δικτύου

Ο [αλγόριθμος 3.1](#) υλοποιεί την διαδικασία υπολογισμού της εξόδου ενός νευρωνικού δικτύου (forward pass), έχοντας σαν δεδομένα τα βάρη και τις τιμές πόλωσης των νευρώνων του κάθε επίπεδου, καθώς και τα δεδομένα εισόδου.

ΚΕΦΑΛΑΙΟ 3. ΤΕΧΝΙΚΕΣ ΒΑΘΙΑΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ ΚΑΙ ΑΝΑΓΝΩΡΙΣΗΣ ΑΝΤΙΚΕΙΜΕΝΩΝ ΣΤΟΝ ΧΩΡΟ

Αλγόριθμος 3.1 Αλγόριθμος Feedforward για τον υπολογισμό των εξόδων ενός επιπέδου του NN

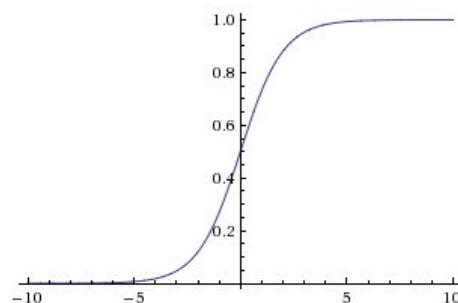
```
1: function ACTIVATION(a)
2:     return  $1.0/(1.0 + e^{-a})$ 
3: end function
4:
5: procedure NN_FORWARD(X, W, B, num_layers)
6:     Starting from the input layer, use  $\sigma$  to do a forward pass through the network,
       computing the activities of the neurons at each layer.
7:      $k \leftarrow 0$ 
8:     while  $k < num\_layers$  do
9:          $X^k \leftarrow ACTIVATION(X \bullet W_k + B_k)$            # If we want to keep output from
           intermediate layers, we must add up one dimension on X.
10:         $k \leftarrow k + 1$ 
11:    end while
12: end procedure
```

3.2.1 Συναρτήσεις Ενεργοποίησης

Σιγμοειδές - Sigmoid

Η σιγμοειδής μη γραμμική συνάρτηση έχει την μορφή που είδαμε στην αρχή του κεφαλαίου. Παίρνει σαν είσοδο έναν προγματικό αριθμό και τον κανονικοποιεί στο διάστημα $[0, 1]$

$$\sigma(x) = 1/(1 + e^{-x})$$



Σχήμα 3.12: Συνάρτηση Σιγμοειδούς συνάρτησης

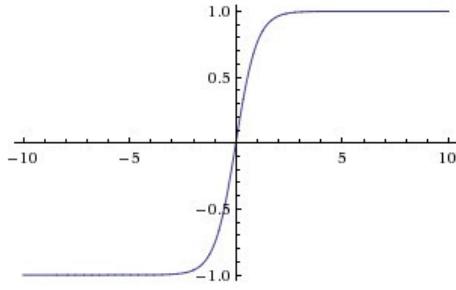
Πλέον δεν χρησιμοποιείται σε πρακτικές εφαρμογές. Προτιμούνται οι συναρτήσεις Tanh, ReLU και Maxout.

Υπερβολική Εφαπτομένη - Tanh

Παίρνει σαν είσοδο έναν θετικό αριθμό και τον κανονικοποιεί στο διάστημα $[-1, 1]$ χρησιμοποιώντας την πιο κάτω σχέση:

$$\tanh x = 2\sigma(2x) - 1$$

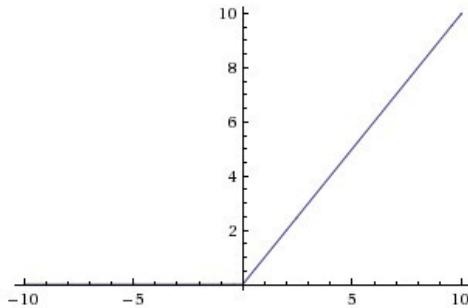
ReLU



Σχήμα 3.13: Συνάρτηση Υπερβολικής Εφαπτομένης

Μία από τις πιο δημοφιλές συναρτήσεις ενεργοποίησης τα τελευταία χρόνια. Πρακτικά κρατά την ενεργοποίηση οριοθετημένη στο μηδέν και είναι γρήγορη στον υπολογισμό.

$$f(x) = \max(0, x) \equiv f(x) = \begin{cases} x, & \text{Αν } x > 0 \\ 0, & \text{Διαφορετικά} \end{cases}$$



Σχήμα 3.14: Συνάρτηση Rectified Linear Unit - ReLU

Το μειονέκτημά της είναι ότι κατά την διάρκεια της εκπαίδευσης του νευρωνικού δικτύου τα βάρη να ανανεώνονται με τέτοιο τρόπο που ο νευρώνας να μην ενεργοποιηθεί ποτέ. Αυτό έχει σαν αποτέλεσμα να "σκοτώσει" τον εκάστοτε νευρώνα.

Leaky ReLU

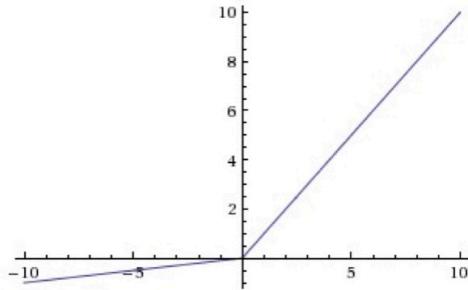
Η συνάρτηση ενεργοποίησης Leaky ReLU προσπαθεί να λύσει το προαναφερθέν πρόβλημα που εμφανίζεται με την χοήση της συνάρτησης ReLU. Αντί να μηδενίζεται για $x < 0$, η συνάρτηση Leaky ReLU έχει μικρή κλίση:

$$f(x) = 1(x < 0)(ax) + 1(x \geq 0)(x) = \begin{cases} x, & \text{Αν } x > 0 \\ ax, & \text{Διαφορετικά} \end{cases}$$

Η τιμή της σταθεράς a ορίζει την κλίση της συνάρτησης για $x < 0$ και μπορεί να δοθεί σαν παράμετρος του εκάστοτε νευρώνα.

Η πρώτη εφαρμογή της συνάρτησης Leaky ReLU σαν συνάρτηση ενεργοποίησης νευρώνων έγινε το 2015 από τον Kaiming He. Οι νευρώνες οι οποίοι χρησιμοποιούν την συνάρτηση Leaky ReLU ονομάζονται νευρώνες *PReLU* [12].

ΚΕΦΑΛΑΙΟ 3. ΤΕΧΝΙΚΕΣ ΒΑΘΙΑΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ ΚΑΙ ΑΝΑΓΝΩΡΙΣΗΣ ΑΝΤΙΚΕΙΜΕΝΩΝ ΣΤΟΝ ΧΩΡΟ



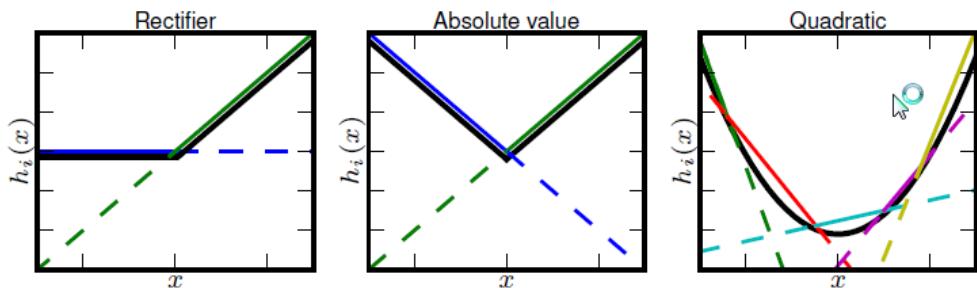
Σχήμα 3.15: Συνάρτηση Leaky ReLU

Maxout

Η συνάρτηση ενεργοποίησης Maxout [13] είναι η γενίκευση της συνάρτησης Leaky ReLU. Ένας νευρώνας Maxout, υπολογίζει την συνάρτηση

$$f(x) = \max(w_1x + b_1, w_2x + b_2)$$

Παρατηρούμε ότι η πιο πάνω συνάρτηση έχει τέσσερις παραμέτρους (w_1, w_2, b_1 και b_2). Επίσης, παρατηρούμε οι συναρτήσεις ReLU και Leaky ReLU είναι ειδικές περιπτώσεις της συνάρτησης Maxout. Για παράδειγμα, για $w_1, b_1 = 0$ παίρνει την μορφή της συνάρτησης ReLU. Το μειονέκτημα αυτής της συνάρτησης ενεργοποίησης, σε σχέση με την συνάρτηση ReLU, είναι ότι διπλασιάζει τις παραμέτρους κάθε νευρώνα.



Σχήμα 3.16: Συνάρτηση Maxout

Πρακτικά, τα μοντέλα NN που χρησιμοποιούν την συνάρτηση Maxout έχουν την ικανότητα να μάθουν, πέρα από την συσχέτιση μεταξύ των κρυμμένων επιπέδων, και την συνάρτηση ενεργοποίησης όλων των νευρώνων ενός ή περισσότερων επιπέδων. Στο [σχήμα 3.16](#) φαίνονται διάφορες μορφές της συνάρτησης Maxout, μετά από την εκπαίδευση δικτύων Maxout.

3.2.2 Συναρτήσεις Σφάλματος/Κόστους

Είναι κατανοητό ότι τα νευρωνικά δίκτυα χρησιμοποιούνται για την επίλυση προβλημάτων classification και regression. Γενικότερος στόχος είναι να παρθεί μία απόφαση, η οποία φαίνεται στην έξοδο του μοντέλου, έχοντας σαν είσοδο τα δεδομένα του εκάστοτε προβλήματος. Άρα καταλαβαίνουμε ότι στα μοντέλα NN τα οποία χρησιμοποιούνται για την επίλυση τέτοιων προβλημάτων, ενσωματώνεται

ένας ταξινομητής (classifier). Ο ταξινομητής αυτός συνήθως αποτελεί το τελευταίο επίπεδο ενός NN.

Οι συναρτήσεις σφάλματος έχουν στόχο να τιμωρήσουν τις λανθασμένες αποφάσεις που λαμβάνονται από τον ταξινομητή, και άρα στην έξοδο ενός NN, κατά την διάρκεια της εκπαίδευσης του.

Μία συνάρτηση σφάλματος έχει την γενική μορφή:

$$E_{total} = f(target - output)$$

Σε προβλήματα κατηγοριοποίησης (classification), όπως για παράδειγμα η αναγνώριση αντικειμένων σε εικόνες, η πιο γνωστή και συχνά χρησιμοποιούμενη συνάρτηση ταξινόμησης είναι η συνάρτηση *Softmax*, η οποία έχει την μορφή:

$$f_j(z) = \frac{e^{z_j}}{\sum_{k \neq j} e^{z_k}} = P(y = j|x)$$

όπου ο δείκτης j αναφέρεται σε ένα στοιχείο του διαγύσματος των πιθανών κλάσεων. Αμέσως καταλαβαίνουμε ότι η πιο πάνω μαθηματική εξίσωση μας δίνει την πιθανότητα y να ανήκει σε μία εκ του συνόλου των πιθανών κλάσεων, έχοντας υπό συνθήκη τα δεδομένα εισόδου x . Πιο συγκεκριμένα, παίρνει σαν είσοδο ένα διάνυσμα πραγματικών τιμών (z) και ”κατασκευάζει” ένα καινούργιο διάνυσμα του οποίου τα στοιχεία είναι κανονικοποιημένα στο διάστημα $[0, 1]$ και το άθροισμα τους ισούται με την μονάδα, δηλαδή $\sum_{k=1}^N f_j(z) = 1$.

Η συνάρτηση σφάλματος που χρησιμοποιείται στην περίπτωση του ταξινομητή Softmax είναι η συνάρτηση *cross-entropy*:

$$L_i = -\log \left(\frac{e^{f_{y_i}}}{\sum_j e^{f_j}} \right)$$

Σε προβλήματα regression οι πιο συχνά χρησιμοποιούμενες συναρτήσεις σφάλματος είναι οι *L1 Norm* και *L2 Norm*.

$$\begin{aligned} L1_i &= |f - y_i|_1 \\ L2_i &= |f - y_i|^2 \end{aligned}$$

Όπως καταλαβαίνουμε, η πρώτη είναι η συνάρτηση μέσης τιμής ενώ η δεύτερη είναι η συνάρτηση του τετραγώνου της μέσης τιμής του σφάλματος.

3.2.3 Αλγόριθμος Backpropagation

Ο αλγόριθμος *backpropagation* (αλγόριθμος 3.2) εμφανίστηκε το 1970 και υποτιμήθηκε μέχρι το 1986, όταν και οι David Rumelhart, Geoffrey Hinton, και Ronald Williams απέδειξαν την αποδοτικότητα του στην εκπαίδευση των νευρωνικών δικτύων, κυρίως όσον αφορά στην ταχύτητα [14]. Σήμερα ο αλγόριθμος backpropagation χρησιμοποιείται κατά κόρον για την εκπαίδευση μεγάλων νευρωνικών δικτύων με εκατομμύρια παραμέτρους.

Αυτό που προσπαθεί να καταφέρει ο αλγόριθμος *backpropagation* είναι να ελαχιστοποιήσει το σφάλμα, διοσμένης μίας συνάρτησης σφάλματος, ορισμένη στον χώρο

ΚΕΦΑΛΑΙΟ 3. ΤΕΧΝΙΚΕΣ ΒΑΘΙΑΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ ΚΑΙ ΑΝΑΓΝΩΡΙΣΗΣ ΑΝΤΙΚΕΙΜΕΝΩΝ ΣΤΟΝ ΧΩΡΟ

των βαρών w , χρησιμοποιώντας τον αλγόριθμο *Gradient Descent*. Υπολογίζει δε το σφάλμα και ανανεώνει ανάλογα τις τιμές του πολυεπίπεδου νευρωνικού δικτύου, ακολουθώντας μία προς τα πίσω διαδικασία.

Πρακτικά, ο αλγόριθμος backpropagation εφαρμόζει τον κανόνα της αλυσιδωτής παραγώγισης (gradient chain rule) για τον υπολογισμό των μερικών παραγώγων, δοσμένης μίας συνάρτησης σφάλματος.

$$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial y} \bullet \frac{\partial y}{\partial x}$$

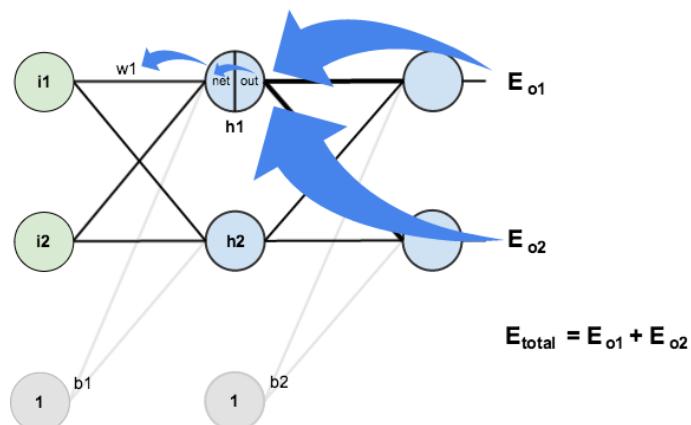
Το πρώτο βήμα για την ελαχιστοποίηση του σφάλματος είναι ο υπολογισμός της παραγώγου της συνάρτησης σφάλματος ως προς τις παραμέτρους w κάθε επιπέδου, ή καλύτερα κάθε νευρώνα, του νευρωνικού δικτύου.

$$\frac{\partial E_{total}}{\partial w_{ij}}$$

Στο [σχήμα 3.17](#) φαίνεται η διαδικασία υπολογισμού της παραγώγου της συνάρτησης σφάλματος ως προς τις παραμέτρους w ενός νευρωνικού δικτύου που αποτελείται από 1 κρυφό επίπεδο και 1 άκρη επίπεδο αποτελείται από 2 ουδέτερους νευρώνες.

$$\frac{\partial E_{total}}{\partial w_1} = \frac{\partial E_{total}}{\partial out_{h1}} * \frac{\partial out_{h1}}{\partial net_{h1}} * \frac{\partial net_{h1}}{\partial w_1}$$

$$\frac{\partial E_{total}}{\partial out_{h1}} = \frac{\partial E_{o1}}{\partial out_{h1}} + \frac{\partial E_{o2}}{\partial out_{h1}}$$



Σχήμα 3.17: Διαδικασία Backward propagation

Η εξίσωση μερικής παραγώγισης της συνάρτησης σφάλματος ως προς την πα-

3.3. ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ ΣΥΝΕΛΙΞΗΣ

ράμετρο w_1 μπορεί να γραφεί σε πιο αναλυτική μορφή:

$$\begin{aligned}\frac{\partial E_{total}}{\partial w_1} &= \left(\sum_{k=1}^O \frac{\partial E_{total}}{\partial out_k} * \frac{\partial out_k}{\partial net_k} * \frac{\partial net_k}{\partial out_{h1}} \right) * \frac{\partial out_{h1}}{\partial net_{h1}} * \frac{\partial net_{h1}}{\partial w_1} \\ \frac{\partial E_{total}}{\partial w_1} &= \left(\sum_{k=1}^O \delta_{ho} * w_{ho} \right) * out_{h1}(1 - out_{h1}) * i_1 \\ \frac{\partial E_{total}}{\partial w_1} &= \delta_{h1} i_1\end{aligned}$$

Η αλυσιδωτή παραγώγιση εμπλέκει και την μερική παραγώγιση της συνάρτησης ενεργοποίησης κάθε νευρώνα ως προς τις παραμέτρους w του. Άρα καταλαβαίνουμε ότι επιλογή της συνάρτησης ενεργοποίησης επηρεάζει και την απόδοση, τόσο σε χρόνο εκτέλεσης, όσο και σε σφάλματα λόγω παραγώγισης, του αλγορίθμου backpropagation.

Η δε ανανέωση των βαρών w γίνεται με βάση την σχέση

$$w_i^+ = w_i - step * \frac{\partial E_{total}}{\partial w_i}$$

δηλαδή αφαιρώντας από την αρχική τιμή την μεταβολή του σφάλματος πολλαπλασιασμένη με ένα βαθμωτό μέγεθος. Το μέγεθος αυτό οποίο ονομάζεται ρυθμός ανανέωσης των βαρών και είναι σταθερά της διαδικασίας εκπαίδευσης.

Η πλήρης ανάλυση του αλγόριθμου backpropagation ξεφεύγει από τα πλαίσια της παρούσας διπλωματικής εργασίας. Στην περίπτωση πολυεπίπεδων NN, η ανάλυση του γίνεται αρκετά περίπλοκη. Ο αναγνώστης παραπέμπεται στο κεφάλαιο 6.5 από το βιβλίο Deep Learning [4], όπου και ο αλγόριθμος backpropagation παρουσιάζεται και αναλύεται πλήρως. Πιο συγκεκριμένα, ο αλγόριθμος 6.1 από το συγκεκριμένο βιβλίο αναφέρεται στην απλή μορφή της διαδικασίας forward propagation, ο 6.2 είναι μία απλουστευμένη μορφή του αλγορίθμου backpropagation, ενώ οι αλγόριθμοι 6.3 και 6.4 παρουσιάζουν πλήρως τις διαδικασίες forward και backward propagation οι οποίοι χρησιμοποιούνται για την εκπαίδευση πολυεπίπεδων νευρωνικών δικτύων, καταλήγοντας έτσι σε μία ολοκληρωμένη μορφή του αλγορίθμου backpropagation.

3.3 ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ ΣΥΝΕΛΙΞΗΣ

Έως τώρα μιλήσαμε για τα πολυεπίπεδα νευρωνικά δίκτυα και την γενικότερη λειτουργία τους. Σε αυτό το υποκεφάλαιο θα μιλήσουμε για συγκεκριμένα μοντέλα πολυεπίπεδων νευρωνικών δικτύων και πιο συγκεκριμένα για τα *Νευρωνικά Δίκτυα Συνέλιξης*. Τα συγκεκριμένα μοντέλα χρησιμοποιούνται σήμερα κυρίως στα προβλήματα της αναγνώρισης και εντοπισμού αντικειμένων σε εικόνες.

Ο τρόπος λειτουργίας τους είναι όμοιος με αυτόν που παρουσιάστηκε στο [υποκεφάλαιο 3.2](#); αποτελούνται από πολλά επίπεδα, όπου το κάθε επίπεδο αποτελείται από έναν αριθμό νευρώνων οι οποίοι έχουν σαν παραμέτρους εκμάθησης τα βάρη τους (w_j^i) και την τιμή πόλωσης (b^i). Κάθε νευρώνας δέχεται ένα σήμα εισόδου, εφαρμόζει μία πράξη εσωτερικού γινομένου σε αυτό, και προαιρετικά εφαρμόζει

ΚΕΦΑΛΑΙΟ 3. ΤΕΧΝΙΚΕΣ ΒΑΘΙΑΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ ΚΑΙ ΑΝΑΓΝΩΡΙΣΗΣ ΑΝΤΙΚΕΙΜΕΝΩΝ ΣΤΟΝ ΧΩΡΟ

Αλγόριθμος 3.2 Backpropagation learning algorithm

```
for d in data do
```

FORWARDS PASS

Starting from the input layer, do a forward pass through the network, computing the activities of the neurons at each layer.

BACKWARDS PASS

Compute the derivatives of the error function with respect to the output layer activities

```
for layer in layers do
```

 Compute the derivatives of the error function with respect to the inputs of the upper layer neurons

 Compute the derivatives of the error function with respect to the weights between the outer layer and the layer below

 Compute the derivatives of the error function with respect to the activities of the layer below

```
end for
```

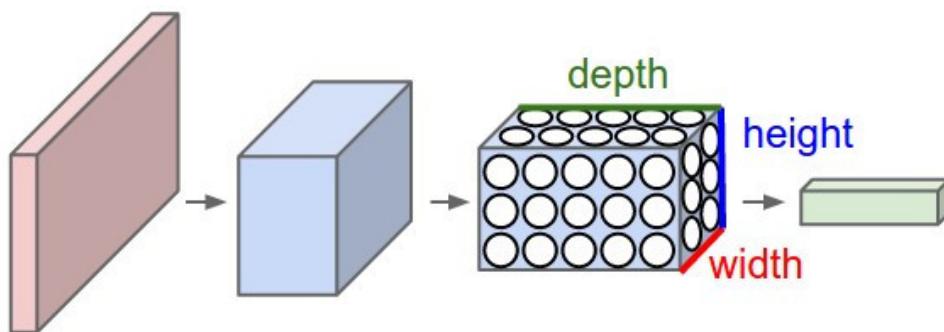
Updates the weights.

```
end for
```

στο αποτέλεσμα μία μη γραμμική συνάρτηση. Το τελευταίο επίπεδο των CNN είναι πλήρες συνδεδεμένο και έχει μία συνάρτηση σφάλματος. Η διαφορά των μοντέλων CNN από τα κλασσικά ANN είναι ότι θεωρούν για δεδομένα εισόδου εικόνες.

Αυτό που καταφέρουν να κάνουν τα CNN είναι να μοντελοποιήσουν μικρά τμήματα πληροφορίας, τα οποία στην συνέχεια ενώνονται για να δημιουργήσουν υψηλότερου επιπέδου πληροφορία. Αν για παράδειγμα παρατηρήσουμε την λειτουργία ενός μοντέλου CNN, το πρώτο επίπεδο προσπαθεί να εντοπίσει ακμές, το δεύτερο επίπεδο παίρνει την πληροφορία των ακμών και προσπαθεί να εντοπίσει περιγράμματα, κτλ.

Σε κάθε εικονοστοιχείο της εικόνας αντιστοιχούν 3 τιμές (RGB) και άρα η είσοδος σε ένα CNN έχει τρεις διαστάσεις όπως φαίνεται και στο [σχήμα 3.18](#). Για παράδειγμα ένα CNN το οποίο έχει σχεδιαστεί να δέχεται σαν είσοδο εικόνες ανάλυσης 80×60 έχει επίπεδο εισόδου διαστάσεων $80 \times 60 \times 3$.



Σχήμα 3.18: Τρισδιάστατη κατανομή των νευρώνων στα CNN

Κάθε επίπεδο ενός CNN παίρνει σαν είσοδο μία μορφή όγκου την οποία και

μετασχηματίζει σε μία άλλη μορφή όγκου.

Οι τρεις βασικοί τύποι επιπέδων που χρησιμοποιούνται σε αρχιτεκτονικές CNN είναι:

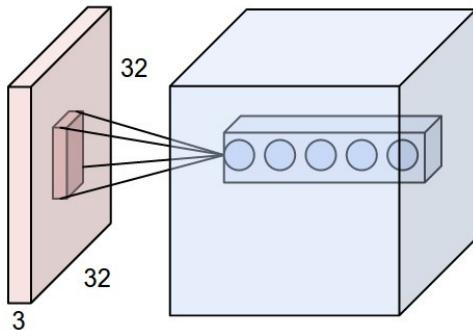
- Επίπεδο Συνέλιξης - Convolutional Layer (CONV)
- Επίπεδο Υπόδειγματοληψίας- Pooling Layer (POOL)
- Πλήρως Συνδεδεμένο Επίπεδο - Fully-Connected Layer (FC)

Μία σημαντική παρατήρηση είναι ότι τα επίπεδα CONV και FC έχουν παραμέτρους, δηλαδή βάρη και τιμή πόλωσης των νευρώνων, ενώ τα επίπεδα POOL εκτελούν λειτουργία δειγματοληψίας στα δεδομένα εισόδου τους.

3.3.1 Επίπεδο Συνέλιξης

Τα επίπεδα συνέλιξης είναι ο πυρήνας των μοντέλων CNN. Οι παράμετροι ενός επιπέδου CONV είναι μία σειρά από δισδιάστατα φίλτρα τα οποία όμως εκτείνονται σε όλο το σε όλο το βάθος του όγκου εισόδου. Το βάθος των φίλτρων αυτών ισούται με το βάθος του όγκου στην είσοδο.

Όπως αναφέραμε και προηγουμένως, τα επίπεδα CONV εφαρμόζουν πράξη συνέλιξης πάνω στα δεδομένα εισόδου. Αυτό επηρεάζει την δομή των "τοπικών" διασυνδέσεων. Στο παράδειγμα του σχήματος 3.19 βλέπουμε πως ο κάθε νευρώνας του επιπέδου συνέλιξης συνδέεται με μία περιοχή του όγκου στην είσοδο του.



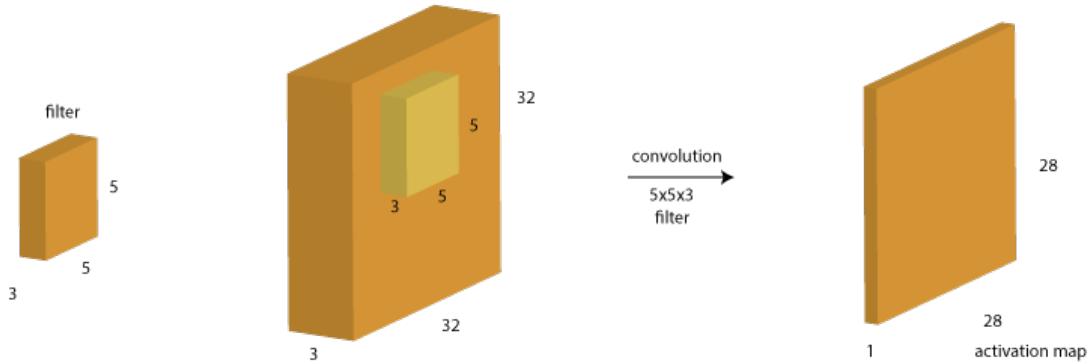
Σχήμα 3.19: Παράδειγμα διασύνδεσης τρισδιάστατης εισόδου με την τρισδιάστατη δομή των νευρώνων ενός επιπέδου συνέλιξης (CONV)

Η συνέλιξη ενός φίλτρου με τον τον όγκο εισόδου παράγει έναν χάρτη ενεργοποίησης (*activation map*), με τον τρόπο που φαίνεται στο σχήμα 3.20. Στο παράδειγμα αυτό εφαρμόζεται φίλτρο διαστάσεων $5 \times 5 \times 3$ σε έναν όγκο $32 \times 32 \times 3$ και παράγεται ένας χάρτης ενεργοποίησης διαστάσεων $28 \times 28 \times 1$.

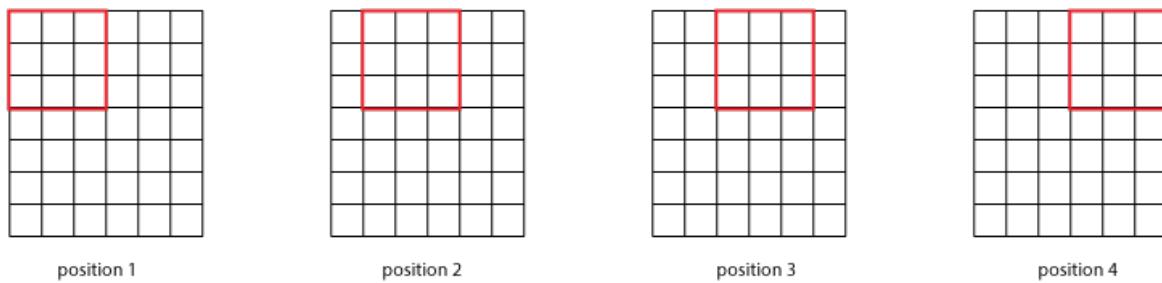
Η μείωση των διαστάσεων μήκους και πλάτους από 32×32 σε 28×28 οφείλεται στον τρόπο με τον οποίο εκτελείται η πράξη της συνέλιξης των φίλτρων με τον όγκο εισόδου (σχήμα 3.21). Οι διαστάσεις του όγκου εξόδου, έχοντας σαν είσοδο όγκο διαστάσεων $N \times N \times d$ και φίλτρων $F \times F \times d$ υπολογίζονται, στην απλούστερη περίπτωση με βάση την σχέση

$$\text{outsize} = (N - F) + 1$$

ΚΕΦΑΛΑΙΟ 3. ΤΕΧΝΙΚΕΣ ΒΑΘΙΑΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ ΚΑΙ ΑΝΑΓΝΩΡΙΣΗΣ ΑΝΤΙΚΕΙΜΕΝΩΝ ΣΤΟΝ ΧΩΡΟ



Σχήμα 3.20: Συνέλιξη φίλτρου ενός επιπέδου συνέλιξης με τον όγκο εισόδου και παραγωγή ενός χάρτη ενεργοποίησης



Σχήμα 3.21: Διαδικασία Συνέλιξης

Η τιμή του βάθους του όγκου στην έξοδο ενός επιπέδου CONV αντιστοιχεί στον αριθμό των φίλτρων που εφαρμόζονται στον όγκο εισόδου. Δηλαδή ο αριθμός των χαρτών ενεργοποίησης αντιστοιχεί στον αριθμό των φίλτρων. Αν για παράδειγμα ο όγκος εισόδου είναι διαστάσεων $32 \times 32 \times 3$ και εφαρμόσουμε δέκα φίλτρα συνέλιξης διαστάσεων $5 \times 5 \times 3$, ο όγκος εξόδου θα είναι διαστάσεων $28 \times 28 \times 3$ [σχήμα 3.22](#). Ο αριθμός των φίλτρων είναι μία παράμετρος, ή καλύτερα υπερ-παράμετρος των επιπέδων συνέλιξης.

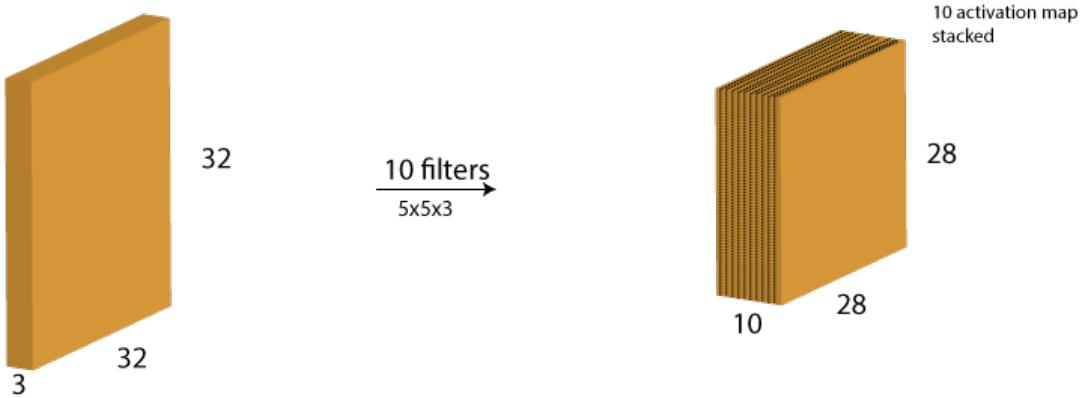
Ωστόσο, το βήμα μετατόπισης (stride) του φίλτρου πάνω στην είσοδο είναι και αυτό μία υπέρ-παράμετρος (hyperparameter) των επιπέδων συνέλιξης. Χρησιμοποιώντας βήμα μετατόπισης (S) διάφορο της μονάδας καταλήγουμε στην πιο κάτω εξίσωση για τον υπολογισμό του όγκου εξόδου:

$$\text{outsize} = (N - F)/S + 1$$

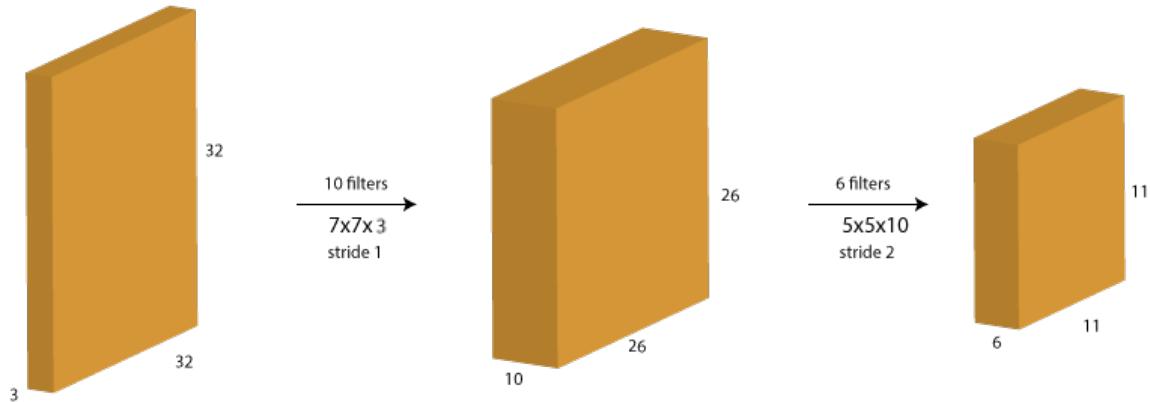
Ένα πρόβλημα που εμφανίζεται στην περίπτωση των μοντέλων CNN με μεγάλο αριθμό χρυφών επιπέδων είναι η γρήγορη μείωση των διαστάσεων μήκους και πλάτους του όγκου, το οποίο είναι αποτέλεσμα της διαδοχικής εφαρμογής πράξεων συνέλιξης ([σχήμα 3.23](#)).

Αυτή η συμπεριφορά είναι ανεπιθύμητη αφού περιορίζει και τις διαστάσεις των φίλτρων που μπορούμε να χρησιμοποιήσουμε σε κάθε επίπεδο CONV. Η χρήση φίλτρων μεγάλων διαστάσεων φέρει σαν αποτέλεσμα την γρήγορη μείωση των διαστάσεων του όγκου.

3.3. ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ ΣΥΝΕΛΙΞΗΣ



Σχήμα 3.22: Αντιστοιχία του αριθμού των φίλτρων ενός επιπέδου συνέλιξης με το βάθος του όγκου στην έξοδο



Σχήμα 3.23: Διαδοχικές εφαρμογές του τελεστή συνέλιξης προκαλούν μείωση των διαστάσεων μήκους και πλάτους του όγκου

Για να αποτρέψουμε αυτή την συμπεριφορά μπορούμε να επεκτείνουμε τις διαστάσεις μήκους και πλάτους, προσθέτοντας μηδενικά στα σύνορα του όγκου εισόδου του εκάστοτε επιπέδου CONV. Η διαδικασία αυτή ονομάζεται *zero-padding* και φαίνεται στο [σχήμα 3.24](#). Το μέγεθος του συνόρου που προστίθεται είναι η τρίτη υπέρ-παράμετρος ενός επιπέδου συνέλιξης.

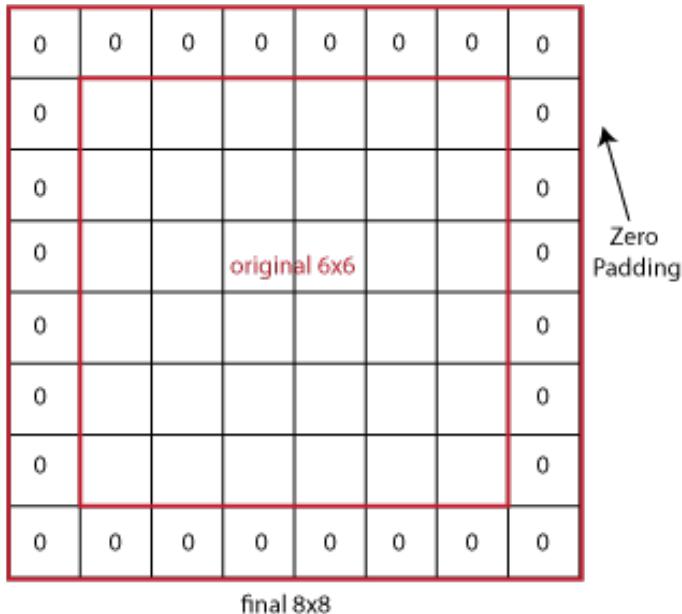
Με την εισαγωγή της υπέρ-παραμέτρου zero-padding η εξίσωση υπολογισμού του όγκου εξόδου έχει την μορφή:

$$\text{outsize} = (N - F + 2P)/S + 1$$

Συνοψίζοντας, ένα επίπεδο συνέλιξης έχει τα εξής χαρακτηριστικά:

- Διαστάσεις όγκου εισόδου: $W_1 \times H_1 \times D_1$
- Hyperparameters:
 - K: Αριθμός φίλτρων
 - F: Μέγεθος του φίλτρου ($F \times F$)

ΚΕΦΑΛΑΙΟ 3. ΤΕΧΝΙΚΕΣ ΒΑΘΙΑΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ ΚΑΙ ΑΝΑΓΝΩΡΙΣΗΣ ΑΝΤΙΚΕΙΜΕΝΩΝ ΣΤΟΝ ΧΩΡΟ



Σχήμα 3.24: Zero Padding

- S: Βήμα μετατόπισης
- P: Ποσότητα zero-padding
- Διαστάσεις όγκου εξόδου: $W_2 \times H_2 \times D_2$, $D_2 = K$ όπου:
 - $W_2 = (W_1 - F)/S + 1$
 - $H_2 = (H_1 - F)/S + 1$
 - $D_2 = D_1$

3.3.2 Επίπεδο Υπό-δειγματοληψίας - Pooling layer

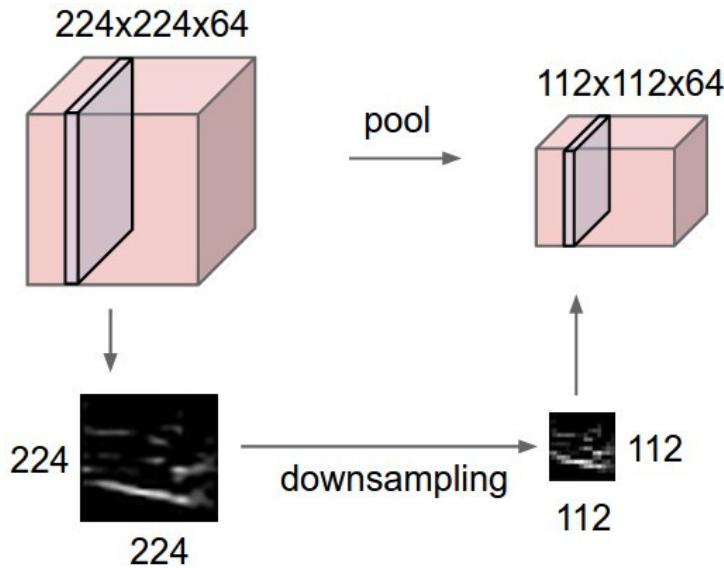
Συνήθως τα επίπεδα υπό-δειγματοληψίας προστίθενται στο δίκτυο, μεταξύ διαδοχικών επιπέδων συνέλιξης. Η λειτουργία τους είναι να μειώσουν τις χωρικές διαστάσεις των αναπαραστάσεων, μειώνοντας έτσι τον αριθμό των παραμέτρων και άρα τους υπολογισμούς που γίνονται στο νευρωνικό δίκτυο, ενεργώντας σαν μία συνάρτηση υποδειγματοληψίας (σχήμα 3.25).

Πιθανές συναρτήσεις υπό-δειγματοληψίας είναι οι συναρτήσεις *max*, *average* και *L2-Norm*.

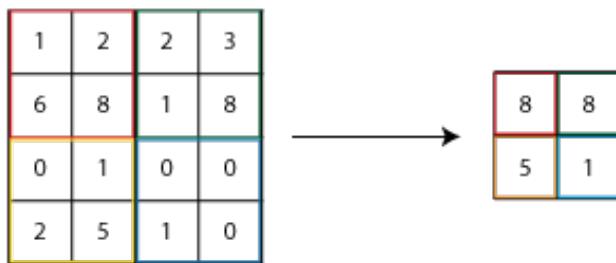
Στο σχήμα 3.26 βλέπουμε το αποτέλεσμα της εφαρμογής της συνάρτησης δειγματοληψίας $\text{max}(\vec{v})$ πάνω σε ένα πλέγμα διαστάσεων 4×4 .

Τα χαρακτηριστικά των συναρτήσεων υπό-δειγματοληψίας είναι:

- Διαστάσεις όγκου εισόδου: $W_1 \times H_1 \times D_1$
- Hyperparameters:
 - F: Χωρική τους έκταση ($F \times F$)



Σχήμα 3.25: Επίπεδο Υπό-δειγματοληψίας - Pooling layer



Σχήμα 3.26: Συνάρτησης υπό-δειγματοληψίας Max - Max Pooling

- S: Βήμα μετατόπισης
- Διαστάσεις όγκου εξόδου: $W_2 \times H_2 \times D_2$, $D_2 = K$ όπου:
 - $W_2 = (W_1 - F)/S + 1$
 - $H_2 = (H_1 - F)/S + 1$
 - $D_2 = D_1$

3.3.3 Πλήρως Συνδεδεμένο Επίπεδο - Fully-connected layer

Ένα πλήρως συνδεδεμένο επίπεδο συνδέεται με όλους τους νευρώνες στο προηγούμενο επίπεδο, όπως γίνεται στα απλά μοντέλα NN (Πολυεπίπεδος Perceptron).

Συνήθως το τελευταίο επίπεδο σε ένα CNN είναι πλήρως συνδεδεμένο και πιο συγκεκριμένα έχει τόσους νευρώνες όσες και οι ακλάσεις της πρόβλεψης. Για παράδειγμα, ένα CNN που χρησιμοποιείται για αναγνώριση αντικειμένων σε εικόνες CIFAR-10 έχει το τελευταίο επίπεδο του πλήρως συνδεδεμένο και αποτελείται από 10 νευρώνες.

4

Εργαλεία Hardware και Software

Στο κεφάλαιο αυτό παρουσιάζονται τα βασικά εργαλεία που χρησιμοποιήθηκαν τόσο για τις υλοποιήσεις όσο και για τα πειράματα. Στο υποκεφάλαιο 4.1 παρουσιάζεται το ενσωματωμένο σύστημα της NVIDIA, Jetson TK1, ενώ τα εργαλεία λογισμικού που χρησιμοποιήθηκαν παρουσιάζονται στο υποκεφάλαιο 4.2.

4.1 NVIDIA JETSON TK1 DEVELOPMENT BOARD

Στο κεφάλαιο αυτό παρουσιάζεται η ενσωματωμένη πλατφόρμα Jetson TK1 της NVIDIA, η οποία χρησιμοποιήθηκε για εφαρμογή των υλοποιήσεων για Αναγνώριση και Εντοπισμό Αντικειμένων σε συστήματα πραγματικού χρόνου, χρησιμοποιώντας Νευρωνικά Δίκτυα Συνέλιξης.

Ο Tegra K1 είναι το πρώτο SOC της NVIDIA για φορητές συσκευές, με προηγμένη αρχιτεκτονική και χαρακτηριστικά, καθώς και χαμηλή κατανάλωση ισχύος. Η τυπική κατανάλωση ισχύος αναφέρεται στα 0.6 – 3 Watt ενώ η μέγιστη μπορεί να φθάσει μέχρι και τα 15 Watt σε κατάσταση μέγιστης λειτουργίας των μονάδων CPU, GPU και ISP.

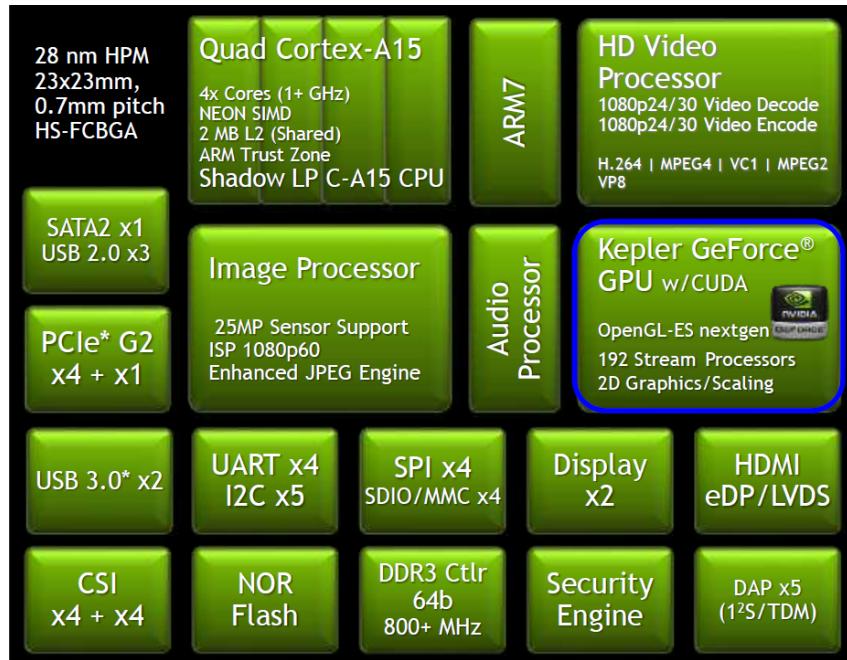
Ένα επίσης ενδιαφέρον χαρακτηριστικό του συγκεκριμένου SOC είναι η τιμή του λόγου της απόδοσης ανά μονάδα ισχύος κατανάλωσης (performance per watt), η οποία αναφέρεται στα $26SPGiga-flops/Watt$ ¹

Όπως φαίνεται στο σχήμα 4.1, τα βασικά τεχνικά χαρακτηριστικά του Tegra K1 SOC είναι:

- CPU: Quad-core ARM Cortex-A15 CPU, 2.3Ghz
- GPU: GK20A Kepler-based GPU with 192 CUDA cores
- RAM: DDR3L/LPDDR3, up to 8GB

¹Λόγος απόδοσης ανά μονάδα ισχύος: <http://wccftech.com/nvidia-tegra-k1-performance-power-consumption-revealed-xia>

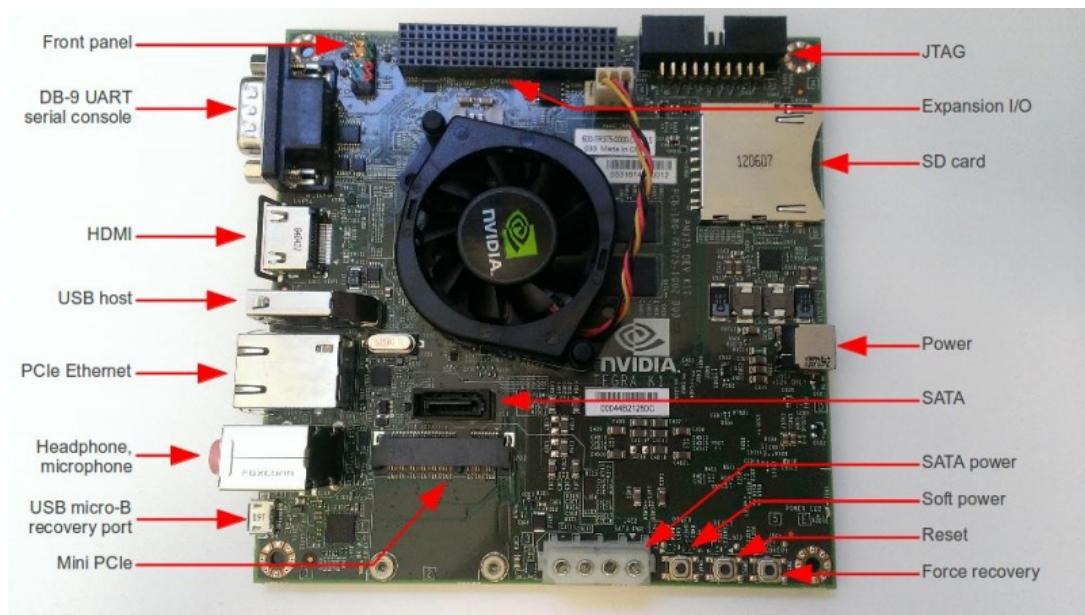
4.1. NVIDIA JETSON TK1 DEVELOPMENT BOARD



Σχήμα 4.1: Tegra K1 SOC

- Peripherals I/O: USB, eMMC/SD-card, LVDS, HDMI, SPI, UART, I2C, SATA, PCIe
- ISP: Image processor

Στα πλαίσια της παρούσας διπλωματικής εργασίας, επιλέχθηκε να χρησιμοποιήσουμε το Jetson TK1 development board της NVIDIA, που φαίνεται στο [σχήμα 4.2](#). Το Jetson TK1 ενσωματώνει το Tegra K1 SOC (CPU+GPU+ISP) και είναι πλήρως



Σχήμα 4.2: Jetson TK1 development board

ΚΕΦΑΛΑΙΟ 4. ΕΡΓΑΛΕΙΑ HARDWARE KAI SOFTWARE

συμβατό με διάφορες διανομές λειτουργικών συστημάτων Linux (Ubuntu, Debian, Arch, Fedora, openSUSE, Gentoo). Η πλήρης συμβατότητα και υποστήριξη του λειτουργικού συστήματος Linux ήταν βασικό κριτήριο στην επιλογή του συγκεκριμένου ενσωματωμένου συστήματος, αφού επιτρέπει την εγκατάσταση εργαλείων με τον ίδιο τρόπο όπως σε ένα σταθερό υπολογιστικό σύστημα (Desktop PC) το οποίο τρέχει Linux OS. Πέρα από την συμβατότητα με κλασσικές διανομές Linux OS, η NVIDIA έχει αναπτύξει δικό της λειτουργικό σύστημα, *Linux4Tegra*, το οποίο έχει σαν βάση τα Ubuntu-14.04, με κάποιες επεκτάσεις για πλήρη υποστήριξη του hardware του Jetson TK1. Επιπλέον, παράγοντας στην επιλογή της συγκεκριμένης πλατφόρμας είναι η πληθώρα των περιφερειακών διεπαφών που διαθέτει, κάνοντάς την χρήσιμη για εφαρμογή σε ρομποτικά συστήματα όπου η σύνδεση διαφόρων περιφερειακών συσκευών, όπως για παράδειγμα αισθητήρες, κάμερες, κινητήρες, σερβο-κινητήρες, είναι απαίτηση. Πιο κάτω δίνονται οι βασικές διεπαφές που προσφέρει η πλατφόρμα Jetson TK1:

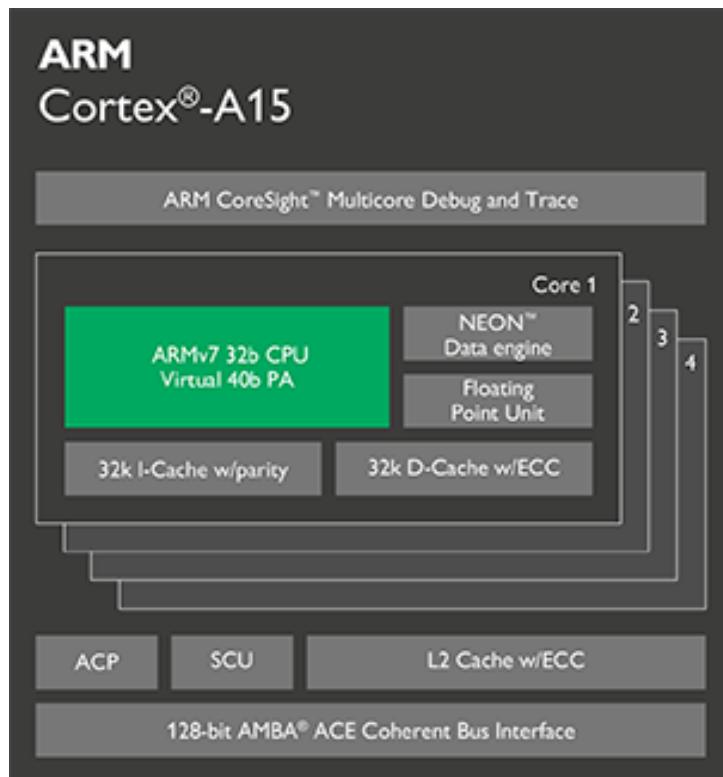
- mini-PCIe: Σύνδεση πρόσθετων συσκευών στον δίαυλο PCI-Express όπως, Wifi cards, SSD δίσκων, κτλ.
- USB 2.0 port: Σύνδεση συσκευών ή/και αισθητήρων με διεπαφή eHCI (Extended Host Controller Interface)
- USB 3.0 port: Σύνδεση συσκευών ή/και αισθητήρων με διεπαφή xHCI (eXtensible Host Controller Interface)
- HDMI: Δυνατότητα σύνδεσης οθόνης
- RS232 port: Παρόλο που το RS232 είναι αρκετά παλιό πρωτόκολλο, εξακολουθεί να ενσωματώνεται σε διάφορες συσκευές που δεν απαιτούν μεγάλο όγκο μεταφοράς δεδομένων, όπως οι οδηγοί κινητήρων
- Audio IO
- Gigabit Ethernet LAN: Δικτύωση της πλατφόρμας με τον "έξω" κόσμο
- SATA: Επιτρέπει την σύνδεση σκληρού δίσκου SATA
- JTAG port: Το JTAG προσφέρει την δυνατότητα σύνδεσης συσκευής/προγράμματος εντοπισμού σφαλμάτων (debugger), για HW-oriented αποσφαλμάτωση
- UART port
- I2C ports: Διαθέτει τρεις θύρες I2C για σύνδεση αισθητήρων/συσκευών που οδηγούνται με το συγκεκριμένο πρωτόκολλο
- GPIO: Προσφέρει δύο θύρες επέκτασης (expansion ports), με 50 και 75 ακροδέκτες αντίστοιχα. Χρήσιμο κυρίως για επικοινωνία συσκευών με SPI, οδήγηση σερβο-κινητήρων, διακλάδωση τροφοδοσίας σε τρίτες συσκευές, κτλ.

Η κατανάλωση ισχύος της πλατφόρμας Jetson TK1 εξαρτάται τόσο από την χρήση του SOC, αλλά και από την χρήση των περιφερειακών διεπαφών που διαθέτει. Η ζήτηση (κατα απόλυτο μέγεθος) σε ηλεκτρική ισχύ για το παρόν ενσωματωμένο σύστημα, σε περίπτωση πλήρους λειτουργίας του (περιλαμβανόμενης της χρήσης του διάυλου SATA & PCIe) αναφέρεται περίπου στα 58 Watt ².

²http://elinux.org/Jetson/Jetson_TK1_Power#Typical_power_draw_of_Jetson_TK1

4.1. NVIDIA JETSON TK1 DEVELOPMENT BOARD

Όσον αφορά την υλοποίηση και ανάπτυξη Νευρωνικών Δικτύων σε ενσωματωμένα συστήματα, η πλατφόρμα Jetson TK1 θεωρείται ιδανική αφού υποστηρίζει CUDA και cuDNN. Η cuDNN (CUDA Deep Neural Network library) είναι GPU-accelerated βιβλιοθήκη για Νευρωνικά Δίκτυα, η οποία αναπτύχθηκε από την NVIDIA και προσφέρει υψηλού επιπέδου υλοποιήσεις για ρουτίνες όπως συνέλιξη, κανονικοποίηση δεδομένων, pooling, επίπεδα ενεργοποίησης, κτλ. Η βιβλιοθήκη cuDNN χρησιμοποιείται και ενσωματώνεται στα πιο δημοφιλή εργαλεία σχεδίασης και υλοποίησης μοντέλων DNN, τα οποία και παρουσιάζονται στο [υποκεφάλαιο 4.2](#).



Σχήμα 4.3: ARM Cortex-A15 processor

Ένα από τα μειονεκτήματα της πλατφόρμας Jetson TK1 είναι η αρχιτεκτονική 32 bit του επεξεργαστή ARM Cortex-A15 MPCore που έχει ενσωματωμένο ([σχήμα 4.3](#)). Με την εμφάνιση των επεξεργαστών ARM Cortex-A57, οι οποίοι είναι αρχιτεκτονικής 64 bit, η υποστήριξη σε λογισμικά, τόσο από την πλευρά της NVIDIA όσο και από την ανοικτού-κώδικα (open-source) κοινότητα, μειώθηκε. Η NVIDIA σταμάτησε να υποστηρίζει μηχανήματα αρχιτεκτονικής 32 bit, τις βιβλιοθήκες CUDA και cuDNN, από την έκτη και δεύτερη έκδοση αντίστοιχα. Η CUDA είναι σήμερα στην έκδοση 8 και η cuDNN στην έκδοση 5, οι οποίες φέρουν τρομερές αναβαθμίσεις στην επίδοση των αλγορίθμων βαθιάς εκμάθησης³. Όπως θα δούμε στο [κεφάλαιο 5](#) το γεγονός αυτό δημιούργησε πολλά προβλήματα καθ όλη την διάρκεια των υλοποιήσεων.

³<https://developer.nvidia/cudnn>

4.2 ΕΡΓΑΛΕΙΑ ΛΟΓΙΣΜΙΚΟΥ

Εξ' αιτίας της ραγδαίας εξέλιξης της επιστήμης της βαθιάς μηχανικής μάθησης, τα τελευταία χρόνια έχουν αναπτυχθεί πολλά εργαλεία λογισμικού (βιβλιοθήκες, SDKs, frameworks) για γρήγορη ή/και αποτελεσματική σχεδίαση και υλοποίηση πολυεπίπεδων νευρωνικών δικτύων.

Θωρούμε σημαντικό να αναφέρουμε μερικά από τα πιο γνωστά εργαλεία σχεδίασης και ανάπτυξης αλγορίθμων και μοντέλων πολυεπίπεδων νευρωνικών δικτύων, τα οποία χρησιμοποιούνται από την επιστημονική κοινότητα:

- Caffe ⁴ [15]: Framework για σχεδίαση, υλοποίηση και δοκιμή μοντέλων βαθιάς μηχανικής μάθησης. Είναι ένα από τα πρώτα ανοικτού κώδικα εργαλεία βαθιάς μηχανικής μάθησης που αναπτύχθηκαν.
- Tensorflow ⁵ [16]: Ανοικτού κώδικα βιβλιοθήκη για αριθμητικούς υπολογισμούς χρησιμοποιώντας γράφους ροής δεδομένων ανεπτυγμένο από την ερευνητική ομάδα της Google. Η ευέλικτη αρχιτεκτονική της επιτρέπει την ανάπτυξη λογισμικού σε μία ή και περισσότερες κεντρικές μονάδες επεξεργασίας CPU ή μονάδες GPU.
- Theano ⁶ [17][18][19]: Βιβλιοθήκη για Python η οποία επιτρέπει τον ορισμό, βελτιστοποίηση και αξιολόγηση μαθηματικών συναρτήσεων που αφορούν πράξεις με πολυδιάστατους πίνακες.
- Torch ⁷ [20][21][22]: Framework για γρήγορη ανάπτυξη και δοκιμή αλγορίθμων μηχανικής μάθησης.
- Keras ⁸ [23]: Ανοικτού κώδικα, υψηλού επιπέδου βιβλιοθήκη σε Python για σχεδίαση και ανάπτυξη πολυεπίπεδων νευρωνικών δικτύων. Ενσωματώνει τις βιβλιοθήκες Theano και Tensorflow και προσφέρει.
- CNTK ⁹ [24]: Σετ από εργαλεία για εφαρμογές βαθιάς μηχανικής μάθησης, ανεπτυγμένο από την ερευνητική ομάδα της Microsoft.

Μία σημαντική παρατήρηση είναι ότι όλα τα προαναφερθέντα εργαλεία χρησιμοποιούν μοντέλα γράφων για την συμβολική αναπαράσταση των μαθηματικών υπολογισμών. Η αναπαράσταση σε μορφή γράφων έχει αποδειχθεί αποτελεσματική όσον αφορά την ταχύτητα υπολογισμών μαθηματικών εκφράσεων, αλλά και την απλότητα ανάπτυξης σχήμα 4.4.

Οι κόμβοι στον γράφο αναπαριστούν μαθηματικές πράξεις ή εκφράσεις, ενώ τα φύλλα αναπαριστούν πολυδιάστατους πίνακες δεδομένων, οι οποίοι ονομάζονται Tensors.

⁴<http://caffe.berkeleyvision.org/>

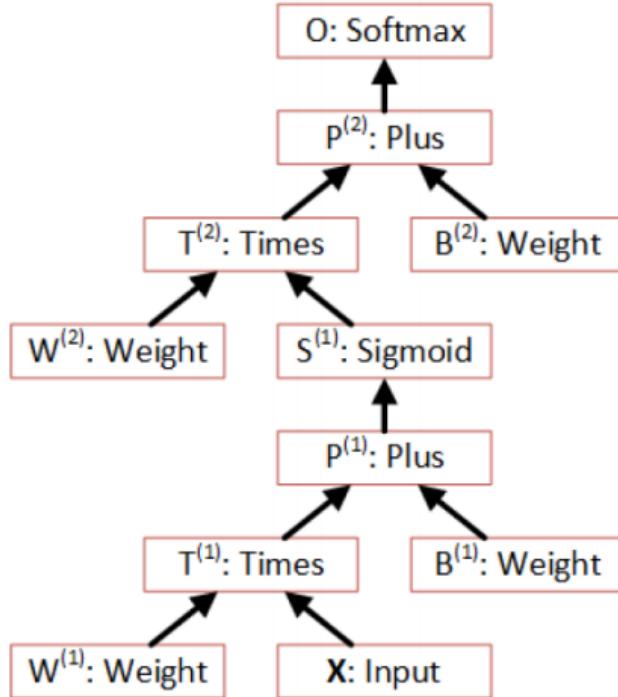
⁵<https://www.tensorflow.org/>

⁶<http://deeplearning.net/software/theano/>

⁷<http://torch.ch/>

⁸<https://keras.io/>

⁹<https://www.cntk.ai/>



Σχήμα 4.4: Παράδειγμα αναπαράστασης ενός μοντέλου νευρωνικού δικτύου σε γράφο. Το NN απειλείται από ένα κρυφό επίπεδο με συνάρτηση ενεργοποίησης την σιγμοειδή συνάρτηση και έναν ταξινομητή Softmax στο επίπεδο εξόδου

Καταλαβαίνουμε ότι η αναπαράσταση των νευρωνικών δικτύων σε μορφή γράφου ροής δεδομένων προσθέτει ένα επίπεδο ευελιξίας στην ανάπτυξη τους. Μία σημαντική παρατήρηση είναι ότι ο γράφος μπορεί χωριστεί σε μικρότερους υπογράφους και άρα να κατανεμηθεί σε περισσότερες από μία υπολογιστικές μονάδες (CPUs / GPUs) ή ακόμα και σε ετερογενή κατανεμημένα συστήματα (CPUs + GPUs) [25].

Ένα ακόμη αξιοσημείωτο εργαλείο είναι η βιβλιοθήκη ConvNetJS [26], η οποία αναπτύχθηκε από τον Andrej Karpathy και υποστηρίζεται απευθείας από σύγχρονους Web Browsers¹⁰ ή/και σε υπολογιστικά συστήματα¹¹.

Στα πλαίσια της παρούσας διπλωματικής εργασίας επιλέξαμε να χρησιμοποιήσουμε την βιβλιοθήκη Keras για τους εξής λόγους:

- Προσφέρει υψηλού επιπέδου ρουτίνες για ανάπτυξη νευρωνικών δικτύων
- Εύκολη και γρήγορη σχεδίαση και ανάπτυξη
- Υποστηρίζει Νευρωνικά Δίκτυα Συνέλιξης

¹⁰Στον ιστοχώρο <http://cs.stanford.edu/people/karpathy/convnetjs/> υπάρχουν παραδείγματα μοντέλων νευρωνικών δικτύων τα οποία μπορεί ο αναγνώστης να δοκιμάσει κατευθείαν μέσα από τον browser του

¹¹Η βιβλιοθήκη ConvNetJS μπορεί να τρέξει σε συστήματα τα οποία υποστηρίζουν Node.js (server-side javascript interpreter)

- Παρέχει πολλά παραδείγματα σχεδίασης και ανάπτυξης
- Τρέχει σε CPU και GPU
- Επιτρέπει την επιλογή μεταξύ των βιβλιοθηκών Theano και Tensorflow για την εκτέλεση των μαθηματικών εκφράσεων
- Είναι από τα πιο ενεργή προγράμματα ανάπτυξης λογισμικού για DL

Το τελευταίο είναι σημαντικό, αφού μας επιτρέπει να αναπτύξουμε μοντέλα και στην συνέχεια να αξιολογήσουμε την απόδοσή τους, χωρίως σε χρόνο εκτέλεσης, με χρήση τόσο του Theano αλλά και του Tensorflow, χωρίς να χρειαστεί επαναπρογραμματισμός.

Η πλήρης λίστα των εργαλείων λογισμικού που χρησιμοποιήθηκαν παρουσιάζεται πιο κάτω:

- Γλωσσα προγραμματισμού Python
- Keras: Σχεδίαση και ανάπτυξη των CNN
- Theano: Keras backend που χρησιμοποιείται για τους υπολογισμούς των μαθηματικών εκφράσεων
- Matplotlib¹²: Βιβλιοθήκη Python για κατασκευή διαγραμμάτων
- PyDot¹³: Βιβλιοθήκη Python για κατασκευή γράφων

¹²<http://matplotlib.org/>

¹³<https://pypi.python.org/pypi/pydot>

5

Υλοποιήσεις

Όπως ήδη έχει κατανοηθεί, η παρούσα διπλωματική εργασία καταπιάνεται με το πρόβλημα της ανάπτυξης νευρωνικών δικτύων στην πλατφόρμα Jetson TK1. Συγκεκριμένα στοχεύει στην ανάπτυξη ενός νευρωνικού δικτύου για ταυτόχρονη αναγνώριση και εντοπισμό αντικειμένων (object detection) σε εικόνες.

Ένα CNN αποτελείται από πολλά επίπεδα συνέλιξης, τα οποία με την σειρά τους αποτελούνται από εκατομμύρια νευρώνες το κάθε ένα. Αυτό σημαίνει και εκατομμύρια παραμέτρους όπου η καθεμία είναι ένας αριθμός κινητής υποδιαστολής 32 bit. Υπενθυμίζουμε ότι το Jetson TK1 έχει περιορισμένη χωρητικότητα μνήμης (2GB), γεγονός το οποίο λήφθηκε υπόψη για την επιλογή του μοντέλου CNN που υλοποιήθηκε.

Οι υλοποιήσεις χωρίζονται σε 2 μέρη:

- Ανάπτυξη διαφόρων μοντέλων CNN σε έναν σταθερό υπολογιστή με επεξεργαστή Intel-i7-6700
- Ρυθμίσεις και βελτιστοποιήσεις στο Jetson TK1

Στο πρώτο μέρος παρουσιάζονται και αναλύονται τα μοντέλα CNN που αναπτύχθηκαν σε έναν υπολογιστή με επεξεργαστή Intel-i7-6700 και μνήμη 16GB, ενώ στο δεύτερο δίνεται μία πλήρης περιγραφή των διαδικασιών ρύθμισης και βελτιστοποίησης που έγιναν στο ενσωματωμένο σύστημα Jetson TK1 καθώς και τα εργαλεία που χρειάστηκε να εγκαταστήσουμε.

5.1 ΜΟΝΤΕΛΑ CNN

Όπως αναφέραμε [υποκεφάλαιο 4.2](#), το βασικό εργαλείο λογισμικού που χρησιμοποιήσαμε για την ανάπτυξη των μοντέλων CNN είναι το Keras.

Η βιβλιοθήκη Keras προσφέρει τις υλοποιήσεις όλων των επιπέδων που απαιτούνται για για την ανάπτυξη ενός CNN¹⁴. Πιο συγκεκριμένα, τα επίπεδα που χρησιμοποιούνται, καθώς και οι βασικές παράμετροι τους περιγράφονται πιο κάτω:

- **InputLayer:** Επίπεδο Εισόδου του νευρωνικού δικτύου
 - Διαστάσεις του όγκου εισόδου (tensor shape)
- **Convolution2D:** Επίπεδο Συνέλιξης
 - Μορφολογία της εισόδου
 - Αριθμός των φίλτρων συνέλιξης
 - Διαστάσεις των φίλτρων συνέλιξης
 - Συνάρτηση ενεργοποίησης
- **MaxPooling2D:** Επίπεδο Ύποδειγματοληψίας
 - Διαστάσεις πλαισίου
 - Βήμα μετατόπισης
- **ZeroPadding2D:** Προσθέτει πλαίσιο με μηδενικά στον όγκο εισόδου
 - Διαστάσεις του πλαισίου
- **Activation:** Επίπεδο ενεργοποίησης. Εφαρμόζει συνάρτηση ενεργοποίησης στον όγκο εξόδου του προηγούμενου επιπέδου
- **Dropout:** Επίπεδο πρόληψης υπέρ-προσαρμογής [27]
- **Dense:** Πλήρες συνδεδεμένο επίπεδο
 - Διαστάσεις του όγκου εισόδου (Προαιρετικό)
 - Διαστάσεις του όγκου εξόδου
 - Συνάρτηση ενεργοποίησης
- **Flatten:** Μετασχηματίζει τον όγκο εισόδου σε επίπεδη αναπαράσταση (π.χ. για όγκο εισόδου $64 \times 32 \times 32$ η έξοδος θα είναι επίπεδη με 65536 νευρώνες)
- **BatchNormalization:** Εφαρμόζει μετασχηματισμό για να διατηρήσει την μέση τιμή και την τυπική απόκλιση των ενεργοποιήσεων του προηγούμενου επιπέδου στις τιμές 0 και 1 αντίστοιχα

Σημαντική παρατήρηση είναι το γεγονός ότι η επιστήμη της βαθιάς μηχανικής μάθησης βρίσκεται σε πρώιμο στάδιο, με αποτέλεσμα να μην υπάρχουν συγκεντρωμένες οι υλοποιήσεις των διαφόρων επιπέδων και γενικότερα των μοντέλων σύγχρονων ANN.

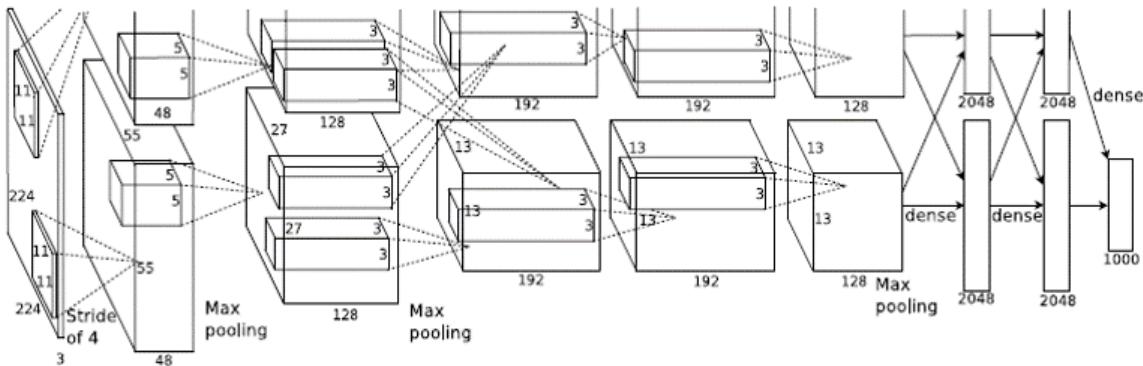
Περαιτέρω, η επιλογή των μοντέλων CNN στηρίχθηκε στην ύπαρξη προ-εκπαιδευμένων βαρών για τα αντίστοιχα CNN στο διαδίκτυο για 2 λόγους:

¹⁴Πλήρες περιγραφή της λίστας των διαθέσιμων επιπέδων: <https://keras.io/layers/core/>

- Η διαδικασία εκπαίδευσης είναι χρονοβόρα διαδικασία και προϋποθέτει τη χοήση μίας ή περισσοτέρων ισχυρών μονάδων GPU (NVIDIA Titan X GPU)
- Η εκπαίδευση νευρωνικών δικτύων ξεφεύγει από τα πλαίσια της παρούσας διπλωματικής εργασίας

5.1.1 AlexNet

Το δίκτυο AlexNet ήταν η αρχή της εισαγωγής της βαθιάς μηχανικής μάθησης στην επιστήμη της μηχανικής όρασης. Εμφανίστηκε και χρησιμοποιήθηκε στον διαγωνισμό ImageNet ILSVRC challenge, το 2012, κερδίζοντας με διαφορά 10,9%, στο σφάλμα αναγνώρισης αντικειμένων σε σύνολο 1000 κλάσεων.



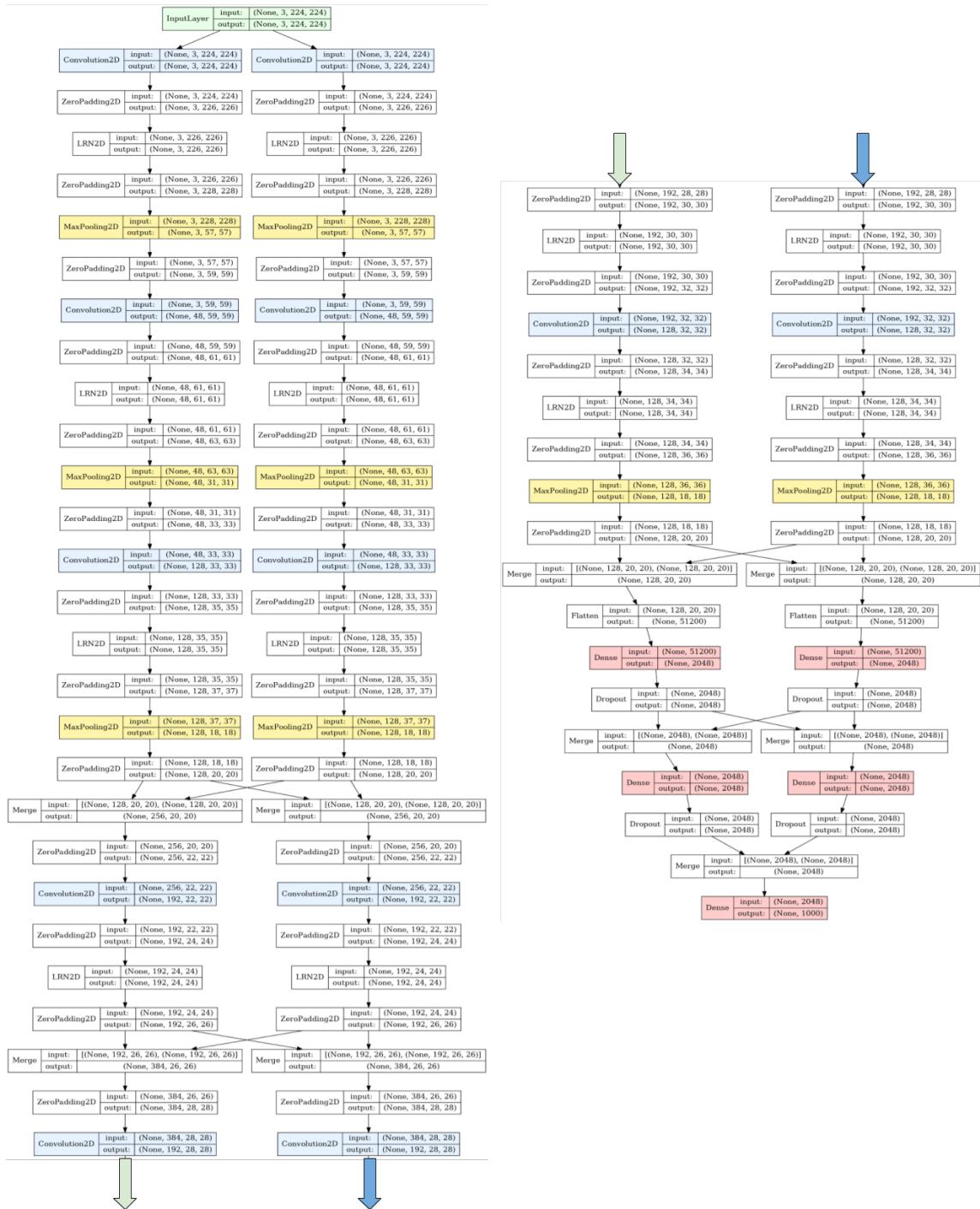
Σχήμα 5.1: Μοντέλο δικτύου AlexNet [2]

Το συγκεκριμένο νευρωνικό δίκτυο συνέλιξης αποτελείται από σύνολο οκτώ επίπεδα, ή καλύτερα ομάδες επιπέδων; πέντε (5) επίπεδα συνέλιξης και τρία (3) πλήρως συνδεδεμένα.

Επίπεδο	Τύπος	Αριθμός καναλιών	Διαστάσεις φίλτρων
1	Conv+Pool+Norm	96	11 × 11
2	Conv+Pool+Norm	256	5 × 5
3	Conv	384	3 × 3
4	Conv	384	3 × 3
5	Conv+Pool	256	3 × 3
6	Full	4096	N/A
7	Full	4096	N/A
8	Full	1000	N/A

Η συνάρτηση υποδειγματοληψίας που χρησιμοποιεί το μοντέλο είναι η συνάρτηση MaxPooling2D με βήμα μετατόπισης ίσο με την μονάδα και στους 2 άξονες (1×1). Σε όλα τα επίπεδα εκτός του τελευταίου χρησιμοποιήθηκε η συνάρτηση ενεργοποίησης ReLU. Επίσης χρησιμοποιεί και πολλά επίπεδα ZeroPadding με πλαίσιο διαστάσεων 1×1 . Το τελευταίο επίπεδο παίζει τον ρόλο του ταξινομητή και συγκεκριμένα είναι ένας ταξινομητής Softmax.

Η υλοποίηση στηρίχτηκε στην αντίστοιχη που υπάρχει με το εργαλείο Caffe¹⁵. Πρακτικά μεταφράστηκε ο πηγαίος κώδικας της υλοποίησης από το Caffe στο Keras. Ο αντίστοιχος γράφος της υλοποίησης του μοντέλου φαίνεται φαίνεται στο σχήμα 5.2.



Σχήμα 5.2: Πλήρης μορφή του δικτύου AlexNet της υλοποίησης

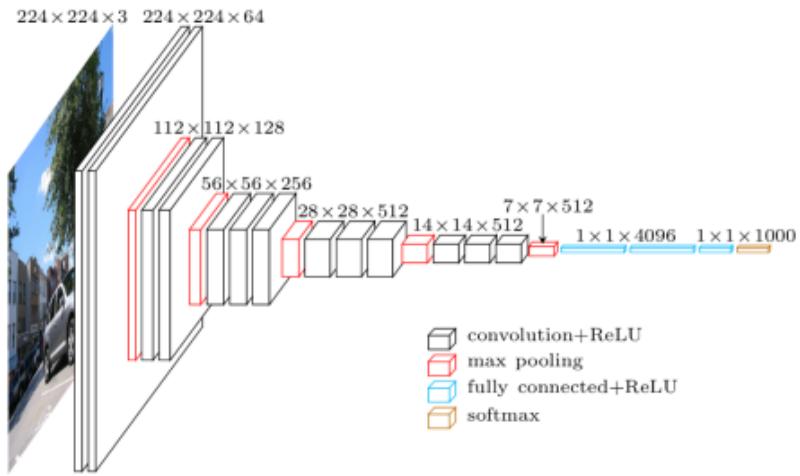
¹⁵Υλοποίηση του δικτύου AlexNet στο Caffe: https://github.com/BVLC/caffe/tree/master/models/bvlc_alexnet

ΚΕΦΑΛΑΙΟ 5. ΥΛΟΠΟΙΗΣΕΙΣ

Το AlexNet είναι ένα μοντέλο που υποστηρίζει μόνο αναγνώριση και όχι εντοπισμό των αντικειμένων, δηλαδή η μόνη πληροφορία που μας δίνει είναι η ύπαρξη των κλάσεων των αντικειμένων.

5.1.2 VGG16

Το δίκτυο VGG16 ήταν ο νικητής του διαγωνισμού ImageNet ILSVRC-2014 με σφάλμα κατηγοριοποίησης 7.5% σε σύνολο 1000 κλάσεων αντικειμένων [?].



Σχήμα 5.3: Μορφή του δικτύου VGG16

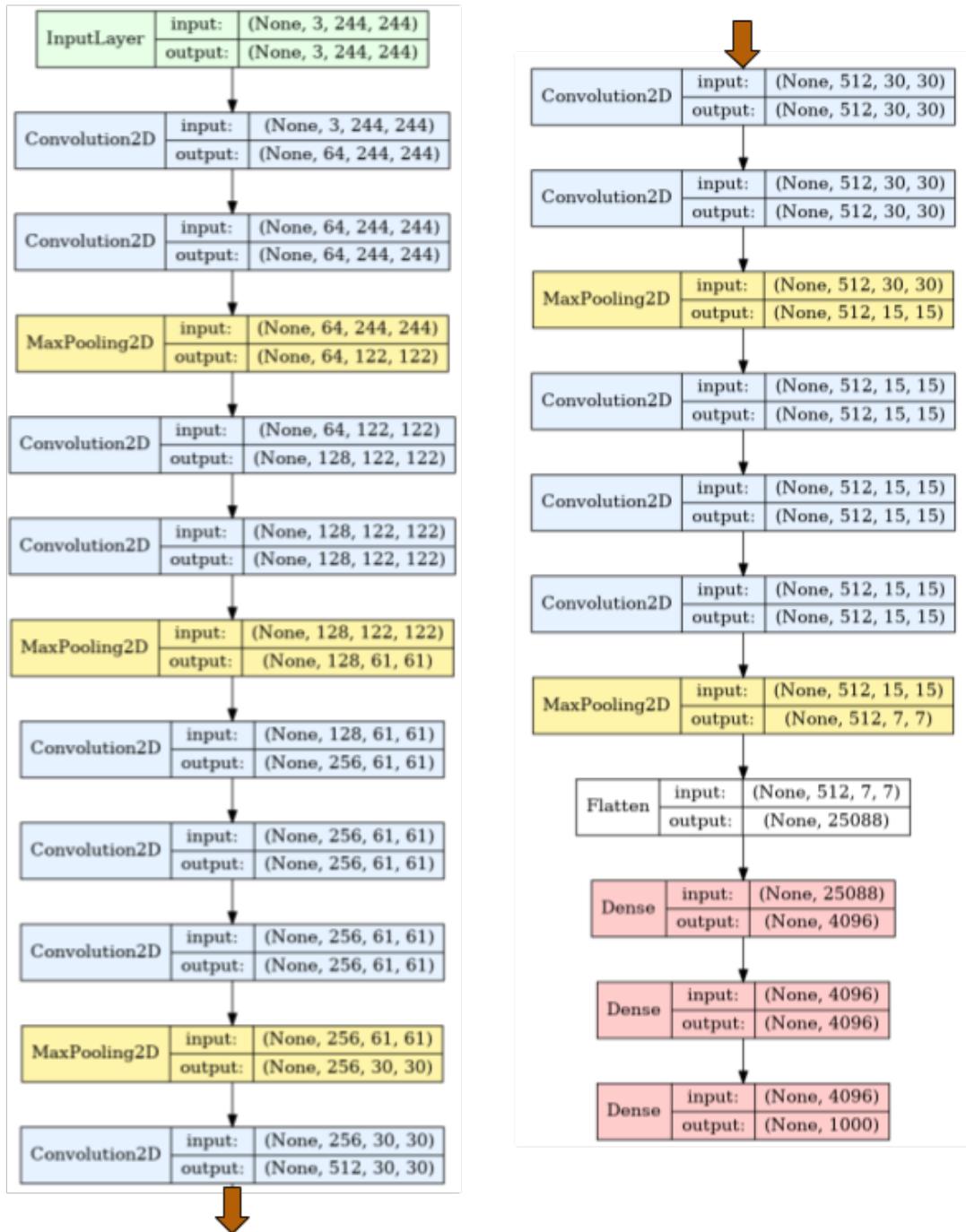
Αποτελείται από 16 επίπεδα; 13 επίπεδα συνέλιξης και 3 πλήρως συνδεδεμένα (σχήμα 5.3), των οποίων τα χαρακτηριστικά δίνονται στον πιο κάτω πίνακα:

Επίπεδο	Τύπος	Αριθμός καναλιών	Διαστάσεις φίλτρων
1	Conv	64	3×3
2	Conv+Pool	64	3×3
3	Conv	128	3×3
4	Conv+Pool	128	3×3
5	Conv	256	3×3
6	Conv	256	3×3
7	Conv+Pool	256	3×3
8	Conv	512	3×3
9	Conv	512	3×3
10	Conv+Pool	512	3×3
11	Conv	512	3×3
12	Conv	512	3×3
13	Conv+Pool	512	3×3
14	FullyConnected	4096	N/A
15	FullyConnected	4096	N/A
16	FullyConnected	1000	N/A

Παρομοίως με το δίκτυο AlexNet, σε όλα τα επίπεδα εκτός του τελευταίου χρησιμοποιήθηκε η συνάρτηση ενεργοποίησης ReLU, ενώ το τελευταίο επίπεδο παίζει

τον ρόλο του ταξινομητή και συγκεκριμένα είναι ένας ταξινομητής Softmax. Το βήμα μετατόπισης των συναρτήσεων υποδειγματοληψίας είναι διαστάσεων 1×1 .

Το μοντέλο του δικτύου VGG16 υπάρχει υλοποιημένο στην λίστα με τα παραδείγματα του εργαλείου Keras¹⁶. Ο αντίστοιχος γράφος της υλοποίησης του μοντέλου φαίνεται στο σχήμα 5.4.



Σχήμα 5.4: Πλήρης μορφή του δικτύου VGG16 της υλοποίησης

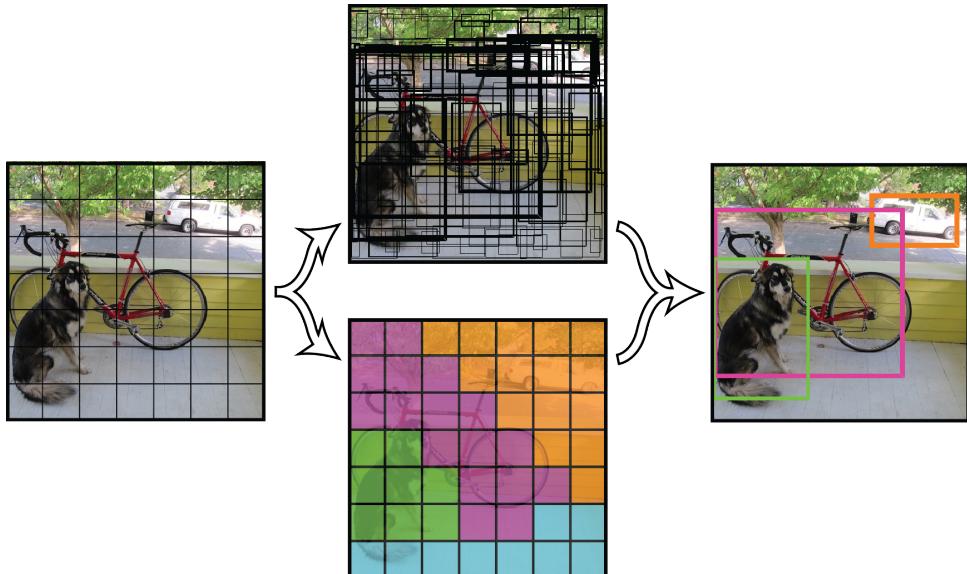
¹⁶<https://github.com/fchollet/keras/tree/master/keras/applications>

5.1.3 YOLO Net

Το δίκτυο YOLO (You Only Look Once), είναι το πρώτο μοντέλο νευρωνικού δικτύου συνέλιξης το οποίο προσπαθεί να επιλύσει το πρόβλημα της ταυτόχρονης αναγνώρισης και εντοπισμού αντικειμένων σε εικόνες, με ένα προς-τα-εμπρός πέρασμα (forward pass). Η ιδιαιτερότητά του είναι ότι αντιμετωπίζει το πρόβλημα σαν ένα πρόβλημα regression και όχι classification.

Μία ακόμη ιδιαιτερότητα του συγκεκριμένου δικτύου είναι ότι θέτει σαν απαίτηση την εφαρμογή του σε προβλήματα σχεδόν πραγματικού χρόνου και άρα στοχεύει κυρίως στην ταχύτητα της αναγνώρισης. Φυσικά αυτό έχει σαν αποτέλεσμα την μείωση της ακρίβειας αναγνώρισης, η οποία είναι χαμηλότερη σε σχέση με άλλα μοντέλα, όπως για παράδειγμα τα δίκτυα Fast-RCNN [?], Overfeat και DetectorNet.

Η έξοδος του δικτύου αντιστοιχεί τόσο στις κλάσεις των αντικειμένων που αναγνωρίστηκαν, χαθώς και στις συντεταγμένες όπου αυτά εντοπίστηκαν.



Σχήμα 5.5: Παράδειγμα τρόπου λειτουργίας του δικτύου YOLO

Το βασικό μοντέλο του δικτύου YOLO έχει την δυνατότητα να επεξεργάζεται εικόνες με ταχύτητα 45 fps χρησιμοποιώντας μονάδα GPU (Nvidia Titan X).

Λόγω της ιδιαιτερότητας του συγκεκριμένου δικτύου, όσον αφορά τον τρόπο προσέγγισης του προβλήματος της ταυτόχρονης αναγνώρισης και εντοπισμού των αντικειμένων, θεωρούμε σημαντικό να αναφέρουμε και να εξηγήσουμε τον τρόπο λειτουργίας του. Το πρώτο βήμα είναι να χωρίσει την εικόνα εισόδου σε ένα πλέγμα διαστάσεων $S \times S$ (αριστερή εικόνα στο [σχήμα 5.5](#)). Κάθε κελί του πλέγματος προβλέπει B οριοθετημένα πλαίσια (bounding boxes) μαζί με ένα σκορ "εμπιστοσύνης" για το κάθε πλαίσιο (πάνω μεσαία εικόνα στο [σχήμα 5.5](#)). Το σκορ εμπιστοσύνης ερμηνεύεται ως η βεβαιότητα να ανήκει ένα αντικείμενο στο συγκεκριμένο πλαίσιο μαζί με την ακρίβεια ότι το συγκεκριμένο αντικείμενο ανήκει σε αυτό το πλαίσιο. Η μαθηματική έκφραση ορίζεται ως:

$$confidence = P(object|box = i) * IOU_{pred}^{truth}$$

5.2. ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΤΩΝ ΕΡΓΑΛΕΙΩΝ ΛΟΓΙΣΜΙΚΟΥ ΣΤΟ JETSON TK1

Από την παραπάνω μαθηματική έκφραση καταλαβαίνουμε ότι σε περίπτωση που σε ένα κελί δεν υπάρχει αντικείμενο, το σκορ εμπιστοσύνης μηδενίζεται. Επίσης, κάθε κελί του πλέγματος προβλέπει και τις πιθανότητες της κλάσης του αντικειμένου, $P(class_i|object)$ (κάτω μεσαία εικόνα στο [σχήμα 5.5](#)).

Χρησιμοποιώντας τις 2 αυτές μαθηματικές εκφράσεις των υπό-συνθήκη πιθανοτήτων, καταλήγουμε στην σχέση

$$\text{class - confidence} = P(class_i|object) * P(object|box = i) * IOU_{pred}^{truth} = \\ P(class_i) * IOU_{pred}^{truth}$$

η οποία μας δίνει το σκορ εμπιστοσύνης για μία συγκεκριμένη κλάση.

Κάθε οριοθετημένο πλαίσιο αποτελείται από πέντε προβλέψεις; x, y, w, h και το σκορ εμπιστοσύνης. Τα x και y αντιστοιχούν στις συντεταγμένες του κέντρου του πλαισίου ενώ οι τιμές w και h αναφέρονται στο μήκος και το πλάτος του.

Τέλος, οι προβλέψεις στην έξοδο του δικτύου φαίνονται σαν ένας όγκος διαστάσεων

$$S \times S \times (B * 5 + C)$$

Το πλήρες μοντέλο YOLO αποτελείται από 24 επίπεδα συνέλιξης και 2 πλήρως συνδεδεμένα.

Πέρα από το πλήρες μοντέλο, έχει σχεδιαστεί και εκπαιδευτεί, από την ίδια ομάδα ερευνητών, ένα μοντέλο με παρόμοιο τρόπο λειτουργίας αλλά με πολύ λιγότερα επίπεδα συνέλιξης και άρα πιο γρήγορο στην εκτέλεση (Tiny/Fast YOLO). Το TinyYOLO αποτελείται από 9 επίπεδα συνέλιξης και 2 πλήρως συνδεδεμένα. Σε όλα τα επίπεδα εκτός του τελευταίου χρησιμοποιείται η συνάρτηση ενεργοποίησης LeakyReLU που παρουσιάστηκε στο [υποκεφάλαιο 3.3](#).

Η υλοποίηση μας στηρίχτηκε στην αντίστοιχη του μοντέλου Tiny-YOLO που έγινε από την αντίστοιχη ερευνητική ομάδα ¹⁷ [?].

Θεωρούμε σημαντικό να αναφέρουμε ότι αρχικά προσπαθήσαμε να τρέξουμε την βασική υλοποίηση, η οποία είναι γραμμένη σε γλώσσα προγραμματισμού C , ανεπιτυχώς. Η εκτέλεση του προγράμματος υλοποίησης του μοντέλου Tiny-YOLO εμφανίζει σφάλματα τύπου *segmentation fault*.

5.2 ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΤΩΝ ΕΡΓΑΛΕΙΩΝ ΛΟΓΙΣΜΙΚΟΥ ΣΤΟ JETSON TK1

Στο [υποκεφάλαιο 4.1](#) είχαμε αναφέρει ότι η πλατφόρμα διαθέτει λειτουργικό Linux4Tegra το οποίο είναι ουσιαστικά διανομή Ubuntu 14.04 με εγκατεστημένους τους απαραίτητους drivers για το συγκεκριμένο σύστημα.

Ένα από τα πρώτα βήματα ήταν να απ-εγκαταστήσουμε από το λειτουργικό το γραφικό περιβάλλον, για να περιορίσουμε την κατανάλωση μνήμης αφού η μνήμη του ενσωματωμένου συστήματος Jetson TK1 είναι περιορισμένη στα 2 Gigabytes και μάλιστα είναι κοινή μεταξύ των μονάδων GPU και CPU (shared memory). Το γραφικό περιβάλλον που είναι προ-εγκατεστημένο στην διανομή Linux4Tegra απαιτεί γύρω στα 114 Megabytes μνήμη RAM. Το Jetson TK1 συνδέθηκε μέσω θύρας ethernet

¹⁷Βασική υλοποίηση του δικτύου YOLO: <http://pjreddie.com/darknet/>

ΚΕΦΑΛΑΙΟ 5. ΥΛΟΠΟΙΗΣΕΙΣ

στο δίκτυο ενός σταθερού υπολογιστή και έτσι ο χειρισμός του έγινε μέσω καναλιού ssh¹⁸, κερδίζοντας έτσι 114 Megabytes σε μνήμη. Αυτό το κέρδος, όπως θα φανεί αργότερα, έπαιξε σημαντικό ρόλο, καθώς τα νευρωνικά δίκτυο που υλοποιήθηκαν απαιτούν αρκετή μνήμη κατά την διάρκεια εκτέλεσής τους (της τάξης των Gigabytes).

Πιο κάτω παρατίθενται οι βασικές τροποποιήσεις και ρυθμίσεις που έγιναν στο λειτουργικό σύστημα:

- Απενεργοποίηση της θύρας HDMI (κέρδος 0.6 Watt σε κατανάλωση ισχύος - αναφέρεται στο τεχνικό εγχειρίδιο)
- Ρύθμιση λειτουργίας των πυρήνων του επεξεργαστή σε κατάσταση μέγιστης απόδοσης
- Ρύθμιση λειτουργίας της μονάδας GPU σε κατάσταση μέγιστη απόδοση

Στην συνέχεια εγκαταστάθηκαν οι βιβλιοθήκες CUDA και cuDNN για να εκμεταλλευτούμε την υπολογιστική ισχύ της μονάδας GPU. (ο οδηγός εγκατάστασης λήφθηκε από τον ιστοχώρο: http://elinux.org/Jetson_TK1).

Ένα μεγάλο μέρος της συνεισφοράς της παρούσας διπλωματικής εργασίας ήταν να γίνουν βελτιστοποιήσεις σε επίπεδο λογισμικού των εργαλείων που χρησιμοποιήθηκαν για το συγκεκριμένο μηχάνημα.

Αρχικά, ακολουθήθηκαν οι οδηγίες εγκατάστασης των εργαλείων *numpy*, *scipy*, *Theano* και *Keras*. Όλα τα προαναφερθέντα πακέτα εγκαταστάθηκαν μέσα από τον επίσημο διαχειριστή πακέτων της Python, *pip*. Τα πρώτα αποτελέσματα σε χρόνους εκτέλεσης πράξεων γραμμικής άλγεβρας ήταν απαγορευτικά (σχεδόν μία τάξη μεγέθους κάτω από το αναμενόμενο). Τόσο η βιβλιοθήκη Theano, όσο και η *numpy*, χρησιμοποιούν εξωτερικές βιβλιοθήκες για πράξεις γραμμικής άλγεβρας, οι οποίες ονομάζονται BLAS (Basic Linear Algebra Subprograms).

Καταλάβαμε ότι το πρόβλημα ήταν στους υπολογισμούς πράξεων γραμμικής άλγεβρας, αφού σχεδόν όλες οι μαθηματικές εκφράσεις που εκτελούνται σε ένα νευρωνικό δίκτυο είναι πράξεις πινάκων και εκτελέστηκε profiling των ρουτινών BLAS χρησιμοποιώντας και τις τρεις βιβλιοθήκες που είναι διαθέσιμες για επεξεργαστές ARM οι οποίες είναι οι εξής:

- libblas-dev + liblapack3 + libumfpack5.6.2: Προ-εγκατεστημένες στην διανομή Linux4Tegra
- ATLAS: Ανοικτού κώδικα
- OpenBLAS: Ανοικτού κώδικα

Τόσο η βιβλιοθήκη ATLAS, όσο και η OpenBLAS, μεταγλωτίστηκαν από τον πηγαίο κώδικα τους, λαμβάνοντας έτσι υπόψη τις απαραίτητες βελτιστοποιήσεις κατά την διαδικασία της μεταγλωτίσης. Συγκεκριμένα ορίστηκε στον μεταγλωτιστή να χρησιμοποιήσει την αρχιτεκτονική του συγκεκριμένου επεξεργαστή (armv7 - Cortex-A15) και τις μονάδες NEON¹⁹ που διαθέτει για τον υπολογισμό πράξεων

¹⁸SSH: Secure Shell <https://tools.ietf.org/html/rfc4254>

¹⁹url`https://www.arm.com/products/processors/technologies/neon.php`

5.2. ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΤΩΝ ΕΡΓΑΛΕΙΩΝ ΛΟΓΙΣΜΙΚΟΥ ΣΤΟ JETSON TK1

με αριθμούς κινητής υποδιαστολής. Επίσης ορίσαμε την χρησιμοποίηση τεσσάρων νημάτων (threads) για την εκτέλεση των ρουτινών BLAS.

Στην συνέχεια μετρήθηκε η απόδοση σε χρόνο εκτέλεσης των εξής ρουτινών BLAS:

- Εσωτερικό γινόμενο δύο διανυσμάτων με 1000 στοιχεία το καθένα
- Εσωτερικό γινόμενο δύο πινάκων διαστάσεων 1000×1000
- Αντιστροφή πίνακα διαστάσεων 1000×1000
- Υπολογισμός διακρίνουσας πίνακα διαστάσεων 1000×1000
- Υπολογισμός ιδιοτιμών πίνακα διαστάσεων 2000×1000
- Υπολογισμός ιδιοδιανυσμάτων πίνακα διαστάσεων 1500×1500

Κάθε μία από τις πιο πάνω ρουτίνες εκτελέστηκε 1000 φορές και λήφθηκε η μέση τιμή του χρόνου εκτέλεσης. Οι καλύτεροι χρόνοι λήφθηκαν χρησιμοποιώντας την βιβλιοθήκη OpenBLAS. Τα συγκριτικά αποτελέσματα των τριών βιβλιοθηκών BLAS παρουσιάζονται στον πιο κάτω πίνακα:

	libblas+liblapack+libumfpack	ATLAS	OpenBLAS
$\vec{x} \odot \vec{x}$	11.21 us	10.99 us	11.49 us
$X \odot X$	2602.7 ms	209.2 ms	148.8 ms
X^{-1}	3420.279 ms	570.630 ms	296.545 ms
$ X $	972.815 ms	283.391 ms	101.563 ms
$SVD(X)$	94.985 s	51.635 s	11.200 s
$Eigen(X)$	98.585 s	38.157 s	28.708 s

Αφού καταλήξαμε στην χρησιμοποίηση της βιβλιοθήκης OpenBLAS, συνεχίσαμε με την μεταγλώττιση από πηγαίο κώδικα των βιβλιοθηκών numpy, scipy και Theano, οι οποίες εκτελούν τις πράξεις BLAS, ορίζοντας αυτή την φορά στον μεταγλωττιστή να χρησιμοποιήσει την συγκεκριμένη βιβλιοθήκη αντί της προ-εγκατεστημένης (link against shared library).

Στο Jetson TK1 εγκαταστάθηκε η βιβλιοθήκη *CNMeM*²⁰, η οποία έχει αναπτυχθεί από την Nvidia και χρησιμοποιείται για την εκ των προτέρων δέσμευση μνήμης σε μονάδες GPU που υποστηρίζουν CUDA. Όπως θα φανεί στο επόμενο κεφάλαιο, η εκ των προτέρων δέσμευση μνήμης επιταχύνει την διαδικασία εκτέλεσης των μαθηματικών υπολογισμών στην μονάδα GPU. Ωστόσο απαιτεί σωστή ρύθμιση για να αποφευχθούν περιπτώσεις δέσμευσης περισσότερης μνήμης από αυτή που το πρόγραμμα εκτέλεσης απαιτεί.

Τέλος, είναι σημαντικό να αναφέρουμε ότι όλες οι προαναφερθέντες διαδικασίες εγκατάστασης των εργαλείων λογισμικού καθώς και τα πειράματα που εκτελέσθηκαν περιγράφονται πλήρως στον ιστοχώρο:

<https://github.com/klpanagi/Thesis/tree/master/jetson-tk1>

²⁰<https://github.com/NVIDIA/cnmem>

6

Πειράματα - Αποτελέσματα

Τα πειράματα χωρίζονται σε δύο φάσεις. Στη πρώτη φάση τα CNN εκτελέστηκαν σε έναν υπολογιστή με επεξεργαστή Intel i7 6700 και χωρητικότητα μνήμης 16GB. Στην συνέχεια, με βάση τα αποτελέσματα των πειραμάτων που λήφθηκαν στην πρώτη φάση των πειραμάτων, επιλέχθηκαν τα αντίστοιχα μοντέλα CNN για εκτέλεση πάνω στο ενσωματωμένο σύστημα Jetson TK1. Οι παράμετροι αξιολόγησης ήταν οι απαιτήσεις σε μνήμη και ο χρόνος εκτέλεσης της διαδικασίας προς-ταξιμπρός περάσματος (forward propagation).

Για λόγους πληρότητας, πιο κάτω παρουσιάζονται τα βασικά χαρακτηριστικά των δύο συστημάτων:

Ο αριθμός νημάτων (8) που υποστηρίζει ο επεξεργαστής i7-6500 οφείλεται στην υποστήριξη της τεχνολογίας *Hyper-Threading*²¹. Σημαντικό επίσης να αναφέρουμε ότι η τιμή της κατανάλωσης ισχύος για την πλατφόρμα Jetson TK1, αντιστοιχεί στην κατάσταση μέγιστης λειτουργίας των μονάδων CPU και GPU.

6.1 ΠΡΩΤΗ ΦΑΣΗ ΠΕΙΡΑΜΑΤΩΝ

TODO!!

²¹<http://www.intel.com/content/www/us/en/architecture-and-technology/hyper-threading/hyper-threading-technology.html>

Μηχάνημα	# πυρήνων	# νημάτων	clock	Κατ. Ισχύος	RAM	Cache
Host PC	4	8	3.4 Ghz	65Watts	16GB	8MB L3
TK1	4	4	2.32	11Watts	2GB (shared)	2MB L2

6.2 ΔΕΥΤΕΡΗ ΦΑΣΗ ΠΕΙΡΑΜΑΤΩΝ

TODO!!

7

Συμπεράσματα

TODO!!

8

Μελλοντικές επεκτάσεις

TODO!!

Βιβλιογραφία

- [1] Murray Campbell, A Joseph Hoane, and Feng-hsiung Hsu. “Deep blue“. Artificial intelligence, 134(1):57–83, 2002.
- [2] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. “ImageNet Classification with Deep Convolutional Neural Networks“. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, “Advances in Neural Information Processing Systems 25“, pages 1097–1105. Curran Associates, Inc., 2012.
- [3] Firas Abuzaid. “Optimizing CPU Performance for Convolutional Neural Networks“.
- [4] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. “Deep Learning“. Book in preparation for MIT Press, 2016. URL <http://www.deeplearningbook.org>.
- [5] Yoshua Bengio, Aaron Courville, and Pascal Vincent. “Representation learning: A review and new perspectives“. IEEE transactions on pattern analysis and machine intelligence, 35(8):1798–1828, 2013.
- [6] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. “Gradient-based learning applied to document recognition“. Proceedings of the IEEE, 86(11):2278–2324, 1998.
- [7] Matthew D. Zeiler and Rob Fergus. “Visualizing and Understanding Convolutional Networks“. CoRR, abs/1311.2901, 2013.
- [8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. “Deep Residual Learning for Image Recognition“. CoRR, abs/1512.03385, 2015.
- [9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. “Identity Mappings in Deep Residual Networks“. CoRR, abs/1603.05027, 2016.
- [10] Christian Szegedy, Alexander Toshev, and Dumitru Erhan. “Deep neural networks for object detection“. In “Advances in Neural Information Processing Systems“, pages 2553–2561, 2013.
- [11] Pierre Baldi. “Autoencoders, unsupervised learning, and deep architectures.“. ICML unsupervised and transfer learning, 27(37-50):1, 2012.
- [12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. “Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification“. CoRR, abs/1502.01852, 2015.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- [13] Ian J Goodfellow, David Warde-Farley, Mehdi Mirza, Aaron C Courville, and Yoshua Bengio. “*Maxout networks*“. ICML (3), 28:1319–1327, 2013.
- [14] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. “*Learning representations by back-propagating errors*“. Cognitive modeling, 5(3):1, 1988.
- [15] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. “*Caffe: Convolutional Architecture for Fast Feature Embedding*“. arXiv preprint arXiv:1408.5093, 2014.
- [16] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, Manjunath Kudlur, Josh Levenberg, Rajat Monga, Sherry Moore, Derek Gordon Murray, Benoit Steiner, Paul A. Tucker, Vijay Vasudevan, Pete Warden, Martin Wicke, Yuan Yu, and Xiaoqiang Zhang. “*TensorFlow: A system for large-scale machine learning*“. CoRR, abs/1605.08695, 2016.
- [17] Rami Al-Rfou, Guillaume Alain, Amjad Almahairi, Christof Angermueller, Dzmitry Bahdanau, Nicolas Ballas, Frédéric Bastien, Justin Bayer, Anatoly Belikov, Alexander Belopolsky, Yoshua Bengio, Arnaud Bergeron, James Bergstra, Valentin Bisson, Josh Bleecher Snyder, Nicolas Bouchard, Nicolas Boulanger-Lewandowski, Xavier Bouthillier, Alexandre de Brébisson, Olivier Breuleux, Pierre-Luc Carrier, Kyunghyun Cho, Jan Chorowski, Paul Christiano, Tim Cooijmans, Marc-Alexandre Côté, Myriam Côté, Aaron Courville, Yann N. Dauphin, Olivier Delalleau, Julien Demouth, Guillaume Desjardins, Sander Dieleman, Laurent Dinh, Mélanie Ducoffe, Vincent Dumoulin, Samira Ebrahimi Kahou, Dumitru Erhan, Ziye Fan, Orhan Firat, Mathieu Germain, Xavier Glorot, Ian Goodfellow, Matt Graham, Caglar Gulcehre, Philippe Hamel, Iban Harlouchet, Jean-Philippe Heng, Balázs Hidasi, Sina Honari, Arjun Jain, Sébastien Jean, Kai Jia, Mikhail Korobov, Vivek Kulkarni, Alex Lamb, Pascal Lamblin, Eric Larsen, César Laurent, Sean Lee, Simon Lefrancois, Simon Lemieux, Nicholas Léonard, Zhouhan Lin, Jesse A. Livezey, Cory Lorenz, Jeremiah Lowin, Qianli Ma, Pierre-Antoine Manzagol, Olivier Mastropietro, Robert T. McGibbon, Roland Memisevic, Bart van Merriënboer, Vincent Michalski, Mehdi Mirza, Alberto Orlandi, Christopher Pal, Razvan Pascanu, Mohammad Pezeshki, Colin Raffel, Daniel Renshaw, Matthew Rocklin, Adriana Romero, Markus Roth, Peter Sadowski, John Salvatier, François Savard, Jan Schlüter, John Schulman, Gabriel Schwartz, Iulian Vlad Serban, Dmitriy Serdyuk, Samira Shabanian, Étienne Simon, Sigurd Spieckermann, S. Ramana Subramanyam, Jakub Sygnowski, Jérémie Tanguay, Gijs van Tulder, Joseph Turian, Sebastian Urban, Pascal Vincent, Francesco Visin, Harm de Vries, David Warde-Farley, Dustin J. Webb, Matthew Willson, Kelvin Xu, Lijun Xue, Li Yao, Saizheng Zhang, and Ying Zhang. “*Theano: A Python framework for fast computation of mathematical expressions*“. arXiv e-prints, abs/1605.02688, May 2016.
- [18] James Bergstra, Olivier Breuleux, Frédéric Bastien, Pascal Lamblin, Razvan Pascanu, Guillaume Desjardins, Joseph Turian, David Warde-Farley, and Yoshua

- Bengio. “*Theano: a CPU and GPU Math Expression Compiler*“. In “*Proceedings of the Python for Scientific Computing Conference (SciPy)*“, June 2010. Oral Presentation.
- [19] Frédéric Bastien, Pascal Lamblin, Razvan Pascanu, James Bergstra, Ian J. Goodfellow, Arnaud Bergeron, Nicolas Bouchard, and Yoshua Bengio. “*Theano: new features and speed improvements*“. Deep Learning and Unsupervised Feature Learning NIPS 2012 Workshop, 2012.
- [20] Ronan Collobert, Samy Bengio, and Johnny Mariéthoz. “*Torch: a modular machine learning software library*“. Technical report, Idiap, 2002.
- [21] Ronan Collobert, Koray Kavukcuoglu, and Clément Farabet. “*Torch7: A matlab-like environment for machine learning*“. In “*BigLearn, NIPS Workshop*“, number EPFL-CONF-192376, 2011.
- [22] Ronan Collobert, Koray Kavukcuoglu, and Clément Farabet. “*Implementing neural networks efficiently*“. In “*Neural Networks: Tricks of the Trade*“, pages 537–557. Springer, 2012.
- [23] François Chollet. “*Keras*“. <https://github.com/fchollet/keras>, 2015.
- [24] Frank Seide and Amit Agarwal. “*CNTK: Microsoft’s Open-Source Deep-Learning Toolkit*“. In “*Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*“, KDD ’16, pages 2135–2135, New York, NY, USA, 2016. ACM. ISBN 978-1-4503-4232-2.
- [25] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Gregory S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian J. Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Józefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Rajat Monga, Sherry Moore, Derek Gordon Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul A. Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda B. Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. “*TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems*“. CoRR, abs/1603.04467, 2016.
- [26] Andrej Karpathy. “*Convnetjs-deep learning in your browser, 2015*“. <https://github.com/karpathy/convnetjs>, 2015.
- [27] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. “*Deep learning*“. Nature, 521 (7553):436–444, 2015.