

UNIVERSITY OF THESSALY

DEPARTMENT OF ELECTRICAL AND
COMPUTER ENGINEERING

DIPLOMA THESIS

Semantics Segmentation of Urban Environment Images

Author:

Dimitrios Mallios

Supervisors:

Gerasimos Potamianos

Antonios Argyriou

May 9, 2018



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

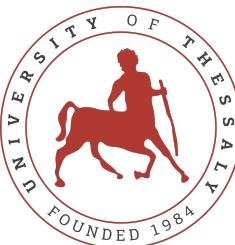
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ
ΜΗΧΑΝΙΚΩΝ Η/Υ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Σημασιολογική Κατάτμηση Εικόνων Αστικού Περιβάλλοντος

Συγγραφέας:
Δημήτριος Μάλλιος

Επιβλέποντες:
Γεράσιμος Ποταμιάνος
Αντώνιος Αργυρίου



Περίληψη

Η παρούσα διπλωματική εξετάζει το πρόβλημα της αναγνώρισης αντικειμένων από εικόνες, των οποίων τα εικονοστοιχεία είναι ταξινομημένα σε μια από τις 19 κατηγορίες. Η εργασία χρησιμοποιεί μια βάση δεδομένων που αποτελείται από 19 διαφορετικές κατηγοριές αντικειμένων η οποία έχει δημιουργηθεί με χρήση κάμερας τοποθετημένη στο εμπρόσθιο μέρος του αυτοκινήτου. Οι εικόνες έχουν απαθανατιστεί από 50 διαφορετικές πόλεις της Ευρώπης σε διάφορες εποχές και καιρικές συνθήκες.

Με την χρήση πληροφορίας από έγχρωμες εικόνες κατασκευάζουμε έναν ταξινομητή ο οποίος μπορεί να αναγνωρίσει την κατηγορία αντικειμένων που ανήκει το κάθε εικονοστοιχείο στην εικόνα ως συνάρτηση του μεγέθους των εικονοστοιχείων αλλά και της δομής που απεικονίζουν. Για την ταξινόμηση χρησιμοποιήσαμε 2 πανομοιότυπα μοντέλα πλήρως συνελικτικών νευρωνικών δικτύων (FCNNs) και 2 διαφορετικά μοντέλα μετά επεξεργασίας για σύγκριση αποτελεσμάτων.

Στόχος της εργασίας ήταν η δημιουργία διαφορετικών ταξινομητών καθώς και η σύγκριση μεταξύ των μεθόδων, αλλά και η δημιουργία λογισμικού για την εικονοποίηση των αποτελεσμάτων. Για την εικονοποίηση των παραπάνω αποτελεσμάτων υλοποιήθηκε λογισμικό που απεικονίζει τα αποτελέσματα των μεθόδων. Για την κατασκευή των παραπάνω μοντέλων γίνεται χρήση των βιβλιοθηκών Keras και Tensorflow, ενώ για την υλοποίηση του λογισμικού εικονοποίησης έγινε η χρήση της βιβλιοθήκης pyQt.

Abstract

This thesis focuses on the problem of recognizing objects from images which are pixel-wise classified in one of the 19 various classes. The database literally introduced in CVPR 2016, consists of 19 various classes of objects created using a camera located on the top of the car. Images have been recorded in 50 individual European cities in different seasons and weather conditions.

Using information from colored images, a classifier implemented due to recognize the category of objects where each individual pixel belongs to. As part of classification, 2 different Fully Convolutional Neural Networks models implemented and another 2 post-processing units.

The aim of this thesis is to create and compare the results from various individual architectures, we also integrate a sophisticated visualizer which presents the results of our model. The tools used in this project are the Keras and Tensorflow and pyQt for the implementation of the visualizer.

Eυχαριστίες

Θα ήθελα να ευχαριστήσω τους επιβλέποντες καθηγητές μου, Δρ. Γεράσιμο Ποταμιάνο και Αντώνιο Αργυριού για την υποστήριξη αλλά και την απαραίτητη γνώση και τα κίνητρα που μου έδωσαν μέσα από τα μαθήματα τους ώστε να πραγματοποιηθεί αυτή η διπλωματική. Επίσης, θα ήθελα να ευχαριστήσω τον συνεπιβλέπων καθηγητή Δρ. Θεόδωρο Γιαννακόπουλο από το Ε.Κ.Ε.Φ.Ε για την υποστήριξη καθώς και τους ανθρώπους από το εργαστήριο Υπολογιστικής Τεχνητής Νοημοσύνης που μου έδωσαν χώρο και πόρους για να υλοποιηθεί αυτή η διπλωματική.

List of Figures

1.1	Αναγνώριση YOLO	15
1.2	Παράδειγμα Σημασιολογίας	15
1.3	Παράδειγμα Εικόνων Βάσης	16
2.1	Brain Stimulus	20
2.2	ANN	21
2.3	Επίπεδο Συνέλιξης	22
2.4	Lenet-5	22
2.5	Alex-Net	23
2.6	Fully-CNN	24
2.7	Διάγραμμα Συνέλιξης	26
2.8	Αναστροφή Φίλτρου	27
2.9	Ολίσθηση Πυρήνα	27
2.10	Πράξη Νευρωνικού	28
2.11	Αναστροφή των δ Πινάκων	29
2.12	Παράδειγμα Οπισθοδρόμησης	30
2.13	Απόκτηση Βαρών	30
2.14	Συμβολή περιοχής στην έξοδο	31
3.1	Παράλληλο Σ.Ν.Δ	35
3.2	Μέθοδοι Παρεμβολής	36
3.3	SELU Function	39
3.4	Συναρτήσεις Ενεργοποίησης	39

3.5	Ενέργεια Συνάρτησης Κόστους	40
3.6	Τμήμα Συνέλιξης	42
3.7	Συνέλιξη με zero-pad	42
3.8	Στάδιο Κωδικοποίησης	43
3.9	Μέγιστη Συγκέντρωση	44
3.10	Παράλληλη Μονάδα Επεξεργασίας	45
3.11	Διεσταλμένη Συνέλιξη	45
3.12	Στάδιο Αποκωδικοποίησης Με βηματισμό	46
3.13	Διγραμμική Μονάδα Αποκωδικοποίησης	47
3.14	Αρχιτεκτονικές	48
4.1	MeanField as CNN	52
4.2	CRF-RNN Network	55
4.3	CNN CRF-RNN Network	55
5.1	Εικόνες από ΠΣΝΔ-ΤΥΣΠΙ-ΕΝΔ	61
5.2	Εικόνες από ΠΣΝΔ	62
5.3	Εικόνες Διγραμμικών ΠΣΝΔ	63
5.4	Πίνακες Σύγχυσης χωρίς CRF Πίνακες σύγχυσης των μοντέλων χωρίς μονάδες μετά-επεξεργασίας και ισοστάθμιση των χλάσεων.	65
5.5	Πίνακες Σύγχυσης με ΤΥΣΠ-ΕΝΔ	66
6.1	Επιλογή Εικόνας	70
6.2	Επιλογή Αρχείου	70
6.3	Προβολή ετικέτας	71
6.4	Παράδειγμα διαφάνειας	71
6.5	Παράδειγμα λογισμικού	72
6.6	Παράδειγμα Συνέλιξης	73

List of Tables

5.1	Αποτελέσματα Μέσου Φίλτρου	58
5.2	Αποτελέσματα CNN-CRF	59
6.1	Color Table	69

Contents

List of Figures	6
List of Tables	8
1 Εισαγωγή	14
1.1 Μηχανική Μάθηση και Σημασιολογική Κατάτμηση	14
1.2 Σημασιολογική Κατάτμηση και Αναγνώριση Αντικειμένων	14
1.3 Η Βάση Δεδομένων Cityscapes	16
1.3.1 Περιγραφή Αντικειμένων	16
1.4 Με Μια Ματιά	17
1.4.1 Στόχοι	17
1.4.2 Συνεισφορά της Εργασίας	18
1.4.3 Δομή της Διατριβής	18
1.5 Συναφείς Εργασίες	19
2 Νευρωνικά Δίκτυα και Βαθειά Μάθηση	20
2.1 Τα Νευρωνικά Δίκτυα	20
2.2 Συνελικτικά Νευρωνικά Δίκτυα CNN	21
2.3 Πλήρως Συνελικτικά Νευρωνικά Δίκτυα	23
2.4 Εμπρόσθια Διάδοση	25
2.5 Αλγόριθμος Οπισθοδρόμησης	27
3 Μεθοδολογία	34
3.1 Εισαγωγή	34

3.2	Πρώτη Προσέγγιση	34
3.3	Προετοιμασία Δεδομένων	35
3.3.1	Υποδειγματοληψία	35
3.3.2	Κανονικοποίηση Χαρακτηριστικών	36
3.3.3	Δυσαναλογία των Κλάσεων	37
3.3.4	Επισκόπηση Αρχιτεκτονικής	37
3.3.5	Στάδιο Κωδικοποίησης	42
3.3.6	Μονάδα Παράλληλης Επεξεργασίας Χαρακτηριστικών	44
3.3.7	Στάδια Αποκωδικοποίησης	46
3.3.8	Ολοκληρωμένες αρχιτεκτονικές	48
4	Μονάδες Μετα-Επεξεργασίας	49
4.1	Επισκόπηση	49
4.2	Μεσαίο Φίλτρο	49
4.3	Conditional Random Fields as Recurrent Neural Network	50
4.3.1	Επισκόπηση Αλγορίθμου	51
4.3.2	Αρχικοποίηση	52
4.3.3	Πέρασμα Μηνυμάτων	52
4.3.4	Στάθμιση Εξόδου Φίλτρου	53
4.3.5	Μετασχηματισμός Συμβατότητας	53
4.3.6	Κανονικοποίηση	54
4.3.7	CRF as RNN	54
5	Experiments and Results	56
5.1	Εκπαίδευση των ΝΔ	56
5.1.1	Σημεία Ελέγχου (Checkpoints)	56
5.1.2	Πρώιμο Σταμάτημα (Early Stopping)	57
5.1.3	Ρυθμός Μάθησης	57
5.2	Αποτελέσματα	57

6 Συμπεράσματα και Μελλοντική Εργασία	67
6.1 Συμπεράσματα	67
6.2 Μελλοντική Εργασία	68
Παράρτημα Α	
Παράρτημα Β	
Bibliography	74

Στην Οικογένειά μου...

Ακρωνύμια

CNN	Convolutional Neural Network
FCNN	Fully Convolutional Neural Network
RNN	Recurrent Neural Network
NN	Neural Network
CRF	Conditional Random Field
CRF-RNN	Conditional Random Field as Recurrent Neural Network
NΔ	Νευρωνικό Δίκτυο
ΤΥΣΠ	Τυχαίο Υπό Συνθήκη Πεδίο
ΤΥΣΠ-ΕΝΔ	Τυχαίο Υπό Συνθήκη Πεδίο ως Επαναλαμβανόμενο Νευρωνικό Δίκτυο
ΣΝΔ	Συνελικτικό Νευρωνικό Δίκτυο
ΠΣΝΔ	Πλήρως Συνελικτικό Νευρωνικό Δίκτυο
ΕΝΔ	Επαναλαμβανόμενο Νευρωνικό Δίκτυο

Chapter 1

Εισαγωγή

1.1 Μηχανική Μάθηση και Σημασιολογική Κατάτμηση

Μηχανική Μάθηση είναι ένας τομέας ο οποίος ανήκει στην Επιστήμη των Υπολογιστών ο οποίος επικεντρώνεται σε εκλεπτυσμένους αλγορίθμους οι οποίοι δεν είναι φτιαχμένοι ρητά από τους επιστήμονες, αλλά μαθαίνουν από τα δεδομένα και προσαρμόζονται σε αυτά για να κάνουν προβλέψεις ή για να πάρουν αποφάσεις. Τα τελευταία χρόνια με την εξέλιξη της υπολογιστικής δύναμης και τον τερράστιο όγκο δεδομένων που είναι διαθέσιμος, επιτρέπει στους επιστήμονες να πειραματιστούν με πιο πολύπλοκους αλγόριθμους. Ο συγκεκριμένος κλάδος καλύπτει ένα τερράστιο εύρος εφαρμογών, από μηχανές αναζήτησης [22] και μετάφραση κειμένου [45] μέχρι εκτιμήσεις για ασθένειες στον κλάδο της Ιατρικής [8].

1.2 Σημασιολογική Κατάτμηση και Αναγνώριση Αντικειμένων

Στην Επιστήμη των Υπολογιστών υπάρχει μια διαφοροποιήση μεταξύ ενός προβλήματος Αναγνώρισης Αντικειμένων και ενός προβλήματος Σημασιολογικής Κατάτμησης αντικειμένων. Αυτά τα δύο προβλήματα ενώ στην ουσία αποσκοπούν στο ίδιο αντικείμενο, έχουν μια πολύ συμαντική διαφορά. Όταν μιλάμε για Αναγνώριση αντικειμένων αναφερόμαστε στην εύρεση της τοποθεσίας ενός αντικειμένου στην εικόνα αλλά στη γενική του μορφή, δηλαδή χωρίς την ακριβή εύρεση των ορίων του αντικειμένου στην εικόνα 1.1. Εν αντιθέσει με το θέμα της Σημασιολογικής Κατάτμησης αντικειμένων, στο οποίο μας ενδιαφέρει η τοποθεσία ενός αντικειμένου στην εικόνα αλλά και η εύρεση των ακριβών ορίων του αντικειμένου στην εικόνα, καθώς τοποθετούμε κάθε εικονοστοιχείο της εικόνας σε μια κατηγορία αντικειμένου (εικόνα 1.2).

Στις παρακάτω εικόνες φαίνονται ξεκάθαρα οι διαφορές των 2 προβλημάτων.

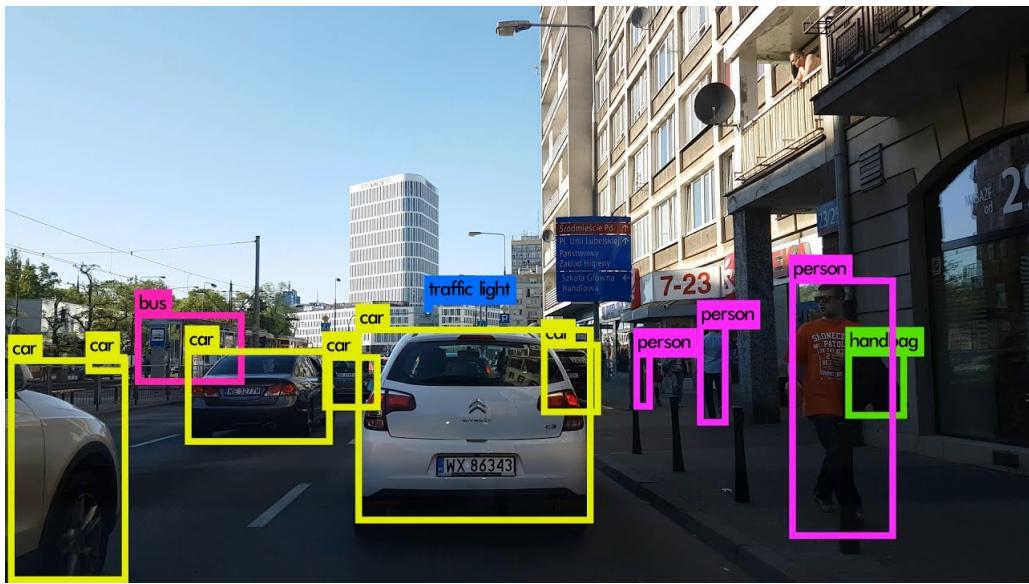


Figure 1.1: Παράδειγμα ενός συστήματος αναγνώρισης αντικειμένων. Αποτέλεσμα του συστήματος YOLO [35]

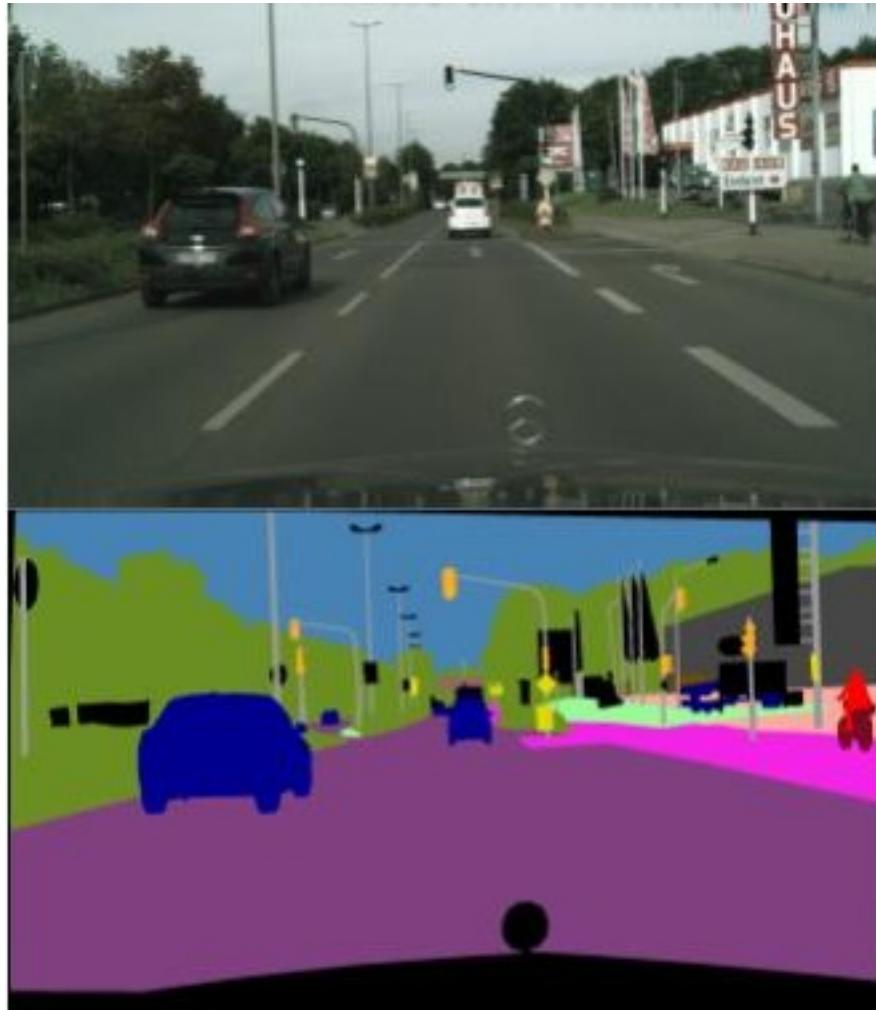


Figure 1.2: Παράδειγμα ενός συστήματος Σημασιολογικής Κατάτμησης αντικειμένων από εικόνες.

1.3 Η Βάση Δεδομένων Cityscapes

Σε αυτή την εργασία χρησιμοποιήθηκε η βάση δεδομένων [Cityscapes \[12\]](#) η οποία αποτελείται από ένα σύνολο έγχρωμων εικόνων υψηλής ευχρίνειας τραβηγμένες σε αστικές περιοχές. Περιλαμβάνει πάνω από 19 κατηγορίες αντικειμένων. Τα αντικείμενα στις εικόνες έχουν σημειωθεί σε βαθμό εικονοστοιχείου. Δηλαδή όλα τα εικονοστοιχεία της εικόνας ανήκουν σε κάποιο αντικειμένο. Η βάση περιέχει περισσότερες από 5000 εικόνες μεγέθους 1024x2048 εικονοστοιχείων από 50 πόλεις της Ευρώπης. Κάθε εικόνα δειγματίζεται από το εικοστό καρέ μιας βιντεοκάμερας, όπου το κάθε καρέ προήλθε από μια ακολουθία βίντεο 30 χρονικών περιόδων μεταξύ τους.

Το σύνολο δεδομένων διαθέτει επίσης ένα άλλο κομμάτι με χονδροειδείς σχολιασμένα εικονοστοιχεία το οποίο αποτελείται από 20000 εικόνες. Οι παρακάτω εικόνες δείχνουν τη διαφορά μεταξύ των δύο συνόλων από εικόνες, οι μερικές σημειωμένες εικόνες δεν έχουν όλα τα εικονοστοιχεία σημειωμένα, όλλα μόνο τα πιο σημαντικά μέρη των αντικειμένων, οι μερικές σημειωμένες εικόνες δεν είναι προς το παρόν στα πειράματά μας.



Figure 1.3: Αριστερά: Πλήρως σημειωμένες εικόνες.
Δεξιά: μερικές σημειωμένες εικόνες.

1.3.1 Περιγραφή Αντικειμένων

Η βάση περιέχει 19 αντικείμενα τα οποία ανήκουν σε 6 υπέρ κατηγορίες. Τα εικονοστοιχεία που ανήκουν σε κατηγορίες αντικειμένων που δεν μας αφορούν στην διαδικασία της αναγνώρισης είναι σημειωμένα ως 'Χωρίς Επικέτα'. Αυτά τα αντικείμενα πιο αναλυτικά είναι:

Classes	Κλάσεις	Categories	Κατηγορίες
Road	Δρόμος	Flat	Επίπεδο
Sidewalk	Πεζόδρομος	Flat	Επίπεδο
Building	Κτίριο	Construction	Κατασκευή
Wall	Τοίχος	Construction	Κατασκευή
Fence	Φράχτης	Construction	Κατασκευή
Pole	Ιστός	Object	Αντικείμενο
Traffic light	Φανάρι κυκλοφορίας	Object	Αντικείμενο
Traffic sign	Πινακίδα κυκλοφορίας	Object	Αντικείμενο
Vegetation	Βλάστηση	Nature	Φύση
Terrain	Έδαφος	Nature	Κατηγορίες
Sky	Ουρανός	Sky	Ουρανός
Person	Άνθρωπος	Human	Άνρυθρωπος
Rider	Αναβάτης	Human	Άνθρωπος
Car	Αυτοκίνητο	Vehicle	Όχημα
Truck	Φορτηγό	Vehicle	Όχημα
Bus	Λειοφωρείο	Vehicle	Όχημα
Train	Τρένο	Vehicle	Όχημα
Motorcycle	Μοτοσυκλέτα	Vehicle	Όχημα
Bicycle	Ποδήλατο	Vehicle	Όχημα
Unlabeld	Χωρίς Ετικέτα	Void	Κενό

1.4 Με Μια Ματιά

1.4.1 Στόχοι

Ο σκοπός αυτής της Διπλωματικής είναι να μελετήσουμε το πρόβλημα της Σημασιολογικής Κατάτμησης Έγχρωμων Εικόνων οι οποίες αναπαριστούν αστικά περιβάλλοντα με την χρήση μεθόδων μηχανικής μάθησης. Η προσπάθειά μας στοχεύει στην πλήρη και ολοκληρωμένη ανασκόπηση ορισμένων από τους αλγορίθμους και τα εργαλεία που θα μπορούσαν να χρησιμοποιηθούν σε αυτόν τον συγκεκριμένο τομέα καθώς και στη σύγκριση των διαφόρων μεθόδων ταξινόμησης. Η δουλειά μας βασίζεται στην έρευνα που δημοσιεύτηκε στον ιστότοπο του Cityscapes-Dataset προκειμένου να αποκτηθούν γνώσεις

στον τομέα και ως εκ τούτου να επεκτείνουμε αυτή την έρευνα με τις δικές μας συνεισφορές.

1.4.2 Συνεισφορά της Εργασίας

Η Σημασιολογική Κατάτμηση πληροφορίας από εικόνες είναι ο τομέας ο οποίος στοχεύει να αλλάξει τον τρόπο με τον οποίο οι μηχανές αντιλαμβάνονται τον κόσμο. Συγκεκριμένα, υπάγεται στον κλάδο της Όρασης Υπολογιστών και αποσκοπεί στο να δώσουμε την ικανότητα στις μηχανές να μπορούν να αναγνωρίζουν τα αντικείμενα με λεπτομερή ακρίβεια, δηλαδή την τμηματοποίηση των αντικειμένων σε σχέση με το υπόβαθρο αλλά και μεταξύ των υπολοίπων αντικειμένων διαγράφοντας με λεπτομέρεια τα όρια των αντικειμένων. Αυτή η πτυχιακή παρουσιάζει μία επισκόπηση του κλάδου της Σημασιολογικής Κατάτμησης πληροφορίας από εικόνες αστικών περιοχών αλλά και στην περαιτέρω έρευνα του προβλήματος. Μέσα από την έρευνα και των μεθόδων και των αλγορίθμων που χρειάζονται, παρέχουμε τις δικές μας λύσεις αλλά και συγκρίσεις μεταξύ των μεθόδων που πειραματιστήκαμε. Για την εικονοποίηση των αποτελεσμάτων προχωρήσαμε στην υλοποίηση της πλατφόρμας που μας δείχνει διαισθητικά τα αποτελέσματα των μεθόδων. Εν ολίγοις οι συνεισφορές της εκάστοτε εργασίας μπορούν να συνοψιστούν ως εξής:

- Στην περαιτέρω έρευνα στον τομέα της Σημασιολογικής Κατάτμησης Αντικειμένων από Εικόνες.
- Στην σύγκριση των αποτελεσμάτων μεταξύ των μεθόδων που πειραματιστήκαμε πάνω σε ένα αληθινό πρόβλημα με την χρήση της βάση δεδομένων Cityscapes-Dataset.
- Στην χρήση των υπερσύγχονων εργαλείων Keras, Tensorflow, OpenCV και PyQt.

1.4.3 Δομή της Διατριβής

Η Διατριβή κατανέμεται σε 5 κεφάλαια, όπου το καθέ ένα επικεντρώνεται σε μία συγκεκριμένη πτυχή του προβλήματος

- **ΚΕΦΑΛΑΙΟ 2** Περιέχει μια εισαγωγή για την ιδέα και την θεωρία των Νευρωνικών Δικτύων.
- **ΚΕΦΑΛΑΙΟ 3** Αναλύει τις αρχιτεκτονικές που χρησιμοποιήθηκαν στην εργασία καθώς και τις απαραίτητες παραμέτρους που επιλέξαμε.
- **ΚΕΦΑΛΑΙΟ 4** Αναλύει τις μονάδες μετα-επεξεργασίας που χρησιμοποιήθηκαν μαζί με τις μεθόδους και αρχιτεκτονικές που συζητήθηκαν στο κεφάλαιο 3.
- **ΚΕΦΑΛΑΙΟ 5** Παρουσιάζει τα αποτελέσματα από τα πειράματα που πραγματοποιήθηκαν κάνοντας χρήση των μεθόδων που εξάγαμε από το κεφάλαιο 3 και 4.
- **ΚΕΦΑΛΑΙΟ 6** Περιέχει θέματα για συζήτηση πάνω στα αποτελέσματα καθώς και μελλοντικές κατευθύνσεις της εργασίας.

1.5 Συναφείς Εργασίες

Η βάση δεδομένων Cityscapes ολοκληρώθηκε και παρουσιάστηκε το 2016. Οι ομάδες που έχουν δημοσιεύσει μέχρι τώρα τα αποτελεσματά τους στην ιστοσελίδα της βάσης χρησιμοποιούν κυρίως τις λεπτομερώς σημειωμένες εικόνες σε συνδυασμό με τις χονδροειδώς σημειωμένες εικόνες. Ωστόσο, υπάρχουν και ομάδες οι οποίες έχουν χρησιμοποιήσει δύο ξεχωριστά μοντέλα δέχοντας σαν είσοδο έγχρωμες εικόνες και εικόνες με πληροφορία βάθους αντίστοιχα[42]. Οι ομάδες με τις καλύτερες επιδόσεις έκαναν χρήση πολύ βαθειών ΣΝΔ σε συνδυασμό με προ-εκπαιδευμένα ΣΝΔ τα οποία είχαν εκπαιδευθεί σε κάποιο άλλο δύσκολο πρόβλημα υπολογιστικής όρασης [23, 31]. Σε άλλες εργασίες όπως [9, 10, 44] χρησιμοποιήθηκε παράλληλη μονάδα επεξεργασίας για λήψη πολλαπλών χαρακτηριστικών από διαφορετικά οπτικά πεδία στην εικόνα. Με αυτόν τον τρόπο εξάγεται χρησιμη πληροφορία από τις εικόνες, καθώς οι παράλληλες μονάδες καταφέρνουν παίρνουν πολύπλοκα χαρακτηριστικά βοηθώντας στην κατανόηση των δομών των αντικειμένων στην εικόνα. Μία από τις πρωτότυπες εργασίες η οποία πέτυχε πολύ μεγάλη ευστοχία στην κατηγοριοποίηση των στοιχείων είναι η ομάδα του που δημιούργησε το PSP-Net [47] το οποίο χρησιμοποιεί ένα πολύ βαθύ ΠΣΝΔ για την εξαγωγή χαρακτηριστικών από ολόκληρη την εικόνα τροφοδοτώντας μια παράλληλη μονάδα επεξεργασίας η οποία εφαρμόζει την τεχνική της μέσης συγκέντρωσης χαρακτηριστικών (Average Pooling) από διαφορετικού μεγέθους περιοχές της εικόνας αξιοποιώντας πληροφορία από πολλές διαφορετικές οπτικές πλευρές.

Τέλος, μια λίγο διαφορετική προσέγγιση ήταν το ΠΣΝΔ SegNet [5] το οποίο βασίστηκε σε μια αρχιτεκτονική ΠΣΝΔ κωδικοποιητής-αποκωδικοποιητής χρησιμοποιώντας μόνο επίπεδα συνέλιξης και διγραμμική παρεμβολή για την υπερδειγματοληψία των χαρακτηριστικών, ενώ στο [33] χρησιμοποιήθηκε παρόμοιο μοντέλο χρησιμοποιώντας επίπεδα αποσυνέλιξης στο στάδιο της αποκωδικοποιήσης.

Chapter 2

Νευρωνικά Δίκτυα και Βαθειά Μάθηση

2.1 Τα Νευρωνικά Δίκτυα

Η βασική αρχή των Νευρωνικών δικτύων ήταν η δημιουργία ενός μοντέλου το οποίο θα μπορεί να προσαρμόζεται σε δεδομένα και να αξιοποιεί την τις πληροφορίες. Η δημιουργία τους εμπνεύστηκε από την βιολογία, συγκεκριμένα, από τον τρόπο που ο εγκεφαλός μας επεξεργάζεται πληροφορίες. Σύμφωνα με το βιολογικό μοντέλο που παρουσίασαν οι H. Hubel and T. Wiesel [18], ο ανθρώπινος εγκέφαλος αποτελείται από κύτταρα τα οποία ονομάζονται νευρώνες. Οι νευρώνες είναι συνδεδεμένοι μεταξύ τους με νευρωνικές γέφυρες, δηλαδή ένα είδος επικοινωνίας που επιτρέπει στους νευρώνες να ανταλλάσουν σήματα μεταξύ τους και να αλληλεπιδρούν. Με αυτό τον τρόπο επιτυγχάνεται η κίνηση, οι αισθήσεις και η δυνατότητα να παίρνουμε αποφάσεις. Ακόμα και η συμπεριφορά μας είναι αποτέλεσμα της διέγερσης των νευρώνων μεταξύ τους αφού επεξεργάζονται πληροφορίες από το περιβάλλον.

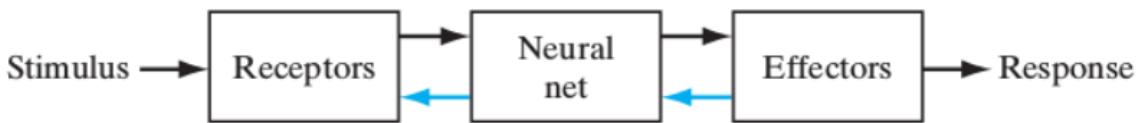


Figure 2.1: Τρόπος επικοινωνίας νευρώνων στον ανθρώπινο εγκέφαλο [14]

Ιδανικά, ένα μαθηματικό μοντέλο ενός νευρωνικού δικτύου προσωμοιώνει την συμπεριφορά του βιολογικού νευρωνικού δικτύου. Για να επιτευχθεί κάτι τέτοιο, οι επιστήμονες έχουν δημιουργήσει ένα μοντέλο το οποίο το οποίο αποτελείται από ένα σύνολο χόμβων οι οποίοι είναι διασυνδεδεμένοι μεταξύ τους και ανταλλάσουν πληροφορία. Ένα παράδειγμα βρίσκεται στην Figure 2.2. Το θέμα των τεχνητών νευρωνικών δικτύων (ANN) είναι πολύ ενδιαφέρον και συνεχίζει να είναι καθώς μας ανοίγει τον δρόμο προς την τεχνητή νοημοσύνη. Επίσης, όπως θα δούμε παρχάτω υπάρχουν πολλά είδη τέτοιων μοντέλων που έχουν πληθώρα εφαρμογών ανάλογα με το πρόβλημα.

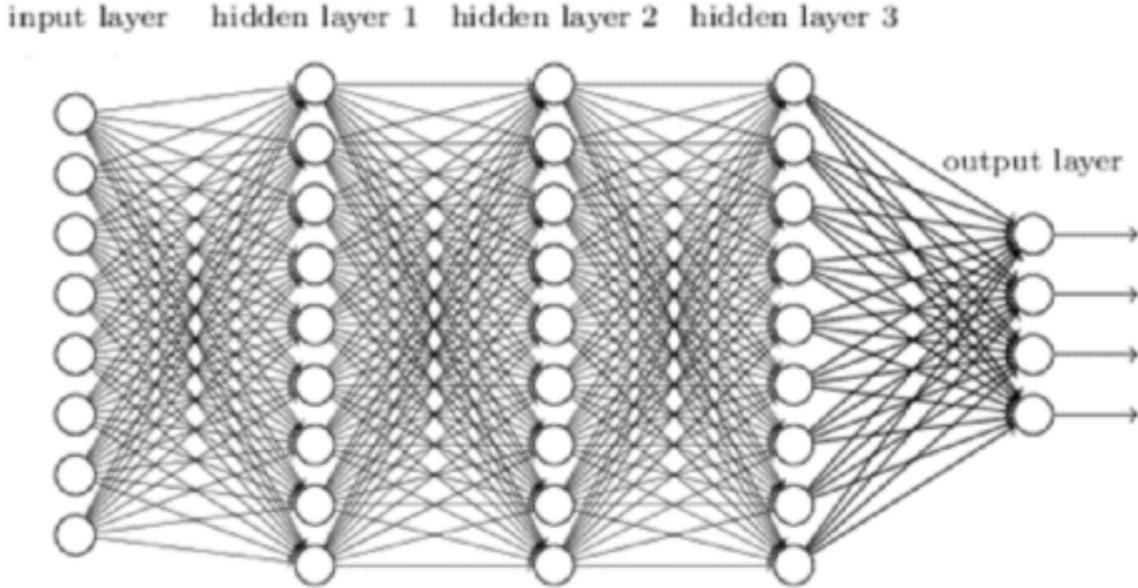


Figure 2.2: **Τεχνητό Νευρωνικό Δίκτυο.** Η είσοδος αποτελείται από κόμβους οι οποίοι δέχονται ένα γεγονός. Η επεξεργασία γίνεται εσωτερικά στους εσωτερικούς κόμβους. Ο αριθμός των επιπέδων ενός Δικτύου δεν είναι προκαθορισμένος και μπορεί να περάσει από πολλά επίπεδα μέχρι να πάρουμε στην έξοδο ένα επιθυμητό αποτέλεσμα. [14]

2.2 Συνελικτικά Νευρωνικά Δίκτυα CNN

Μια κατηγορία Νευρωνικών Δικτύων είναι τα Συνελικτικά Νευρωνικά Δίκτυα, τα οποία έχουν εφαρμογές σε προβλήματα επεξεργασίας εικόνας και υπολογιστικής όρασης. Αποτελούνται από ένα ή περισσότερα επίπεδα συνέλιξης (convolutional layers) συχνά ακολουθέμενα από ένα επίπεδο υποδειγματοληψίας ακολουθόμενο από ένα ή περισσότερα fully connected επίπεδα όπως συναντάμε και σε ένα πολύ-επίπεδο Νευρωνικό Δίκτυο. Η αρχιτεκτονική του NN σχεδιάζεται έτσι ώστε να εκμεταλλεύεται την δυσδιάστατη δομή των εικόνων εισόδου ή άλλα δυσδιάστατα σήματα όπως σήματα ήχου (π.χ. Φασματόγραμμα). Αυτό επιτυγχάνεται με τοπικές συνδέσεις και κατάλληλα βάρον προκειμένου να δημιουργηθούν ανεξαρτήτως μετατοπίσεων χαρακτηριστικά (Translation-Invariant). Άλλο ένα πλεονέκτημα των CNNs είναι ότι είναι ευχολότερα στην εκπαίδευση και έχουν πολύ λιγότερες παραμέτρους από τα N.D. που έχουν fully-connected επίπεδα.

Η είσοδος σε ένα επίπεδο συνέλιξης είναι μια $m \times n \times c$ εικόνα όπου m και n είναι το ύψος και το πλάτος της εικόνας αντίστοιχα, ενώ το c είναι ο αριθμός των καναλιών πχ για RGB $c = 3$. Το επίπεδο συνέλιξης έχει k φίλτρα (kernels) μεγέθους $n \times n \times r$ όπου είναι μικρότερο από τη διάσταση της εικόνας και μπορεί να είναι ίδιου μεγέθους με τα κανάλια ή μικρότερου και μπορεί να ποικίλει για κάθε κερνελ. Το μέγεθος των φίλτρων προκαλεί τοπικά συνδεδεμένη δομή όπου το καθένα συνελίσσεται με κάθε εικόνα για να παράγουν χάρτες χαρακτηριστικών (featuremaps) μεγέθους $m - n + 1$. Κάθε χαρακτηριστικό υποδειγματοληπτείται τυπικά με κάποιο pooling επίπεδο σε $p \times p$ συνεχείς περιοχές όπου το p παίρνει συνήθως τιμές μεταξύ 2 και 5 αλλά για μεγάλες εικόνες εισόδου συναντάμε και μεγαλύτερα. Πριν ή μετά το pooling layer ακολουθεί μια προσθήκη βιας και μια συνάρτηση ενεργοποίησης σε κάθε χάρτη χαρακτηριστικών. Η Figure 2.3 μας δείχνει ένα παράδειγμα ενός επιπέδου συνέλιξης.

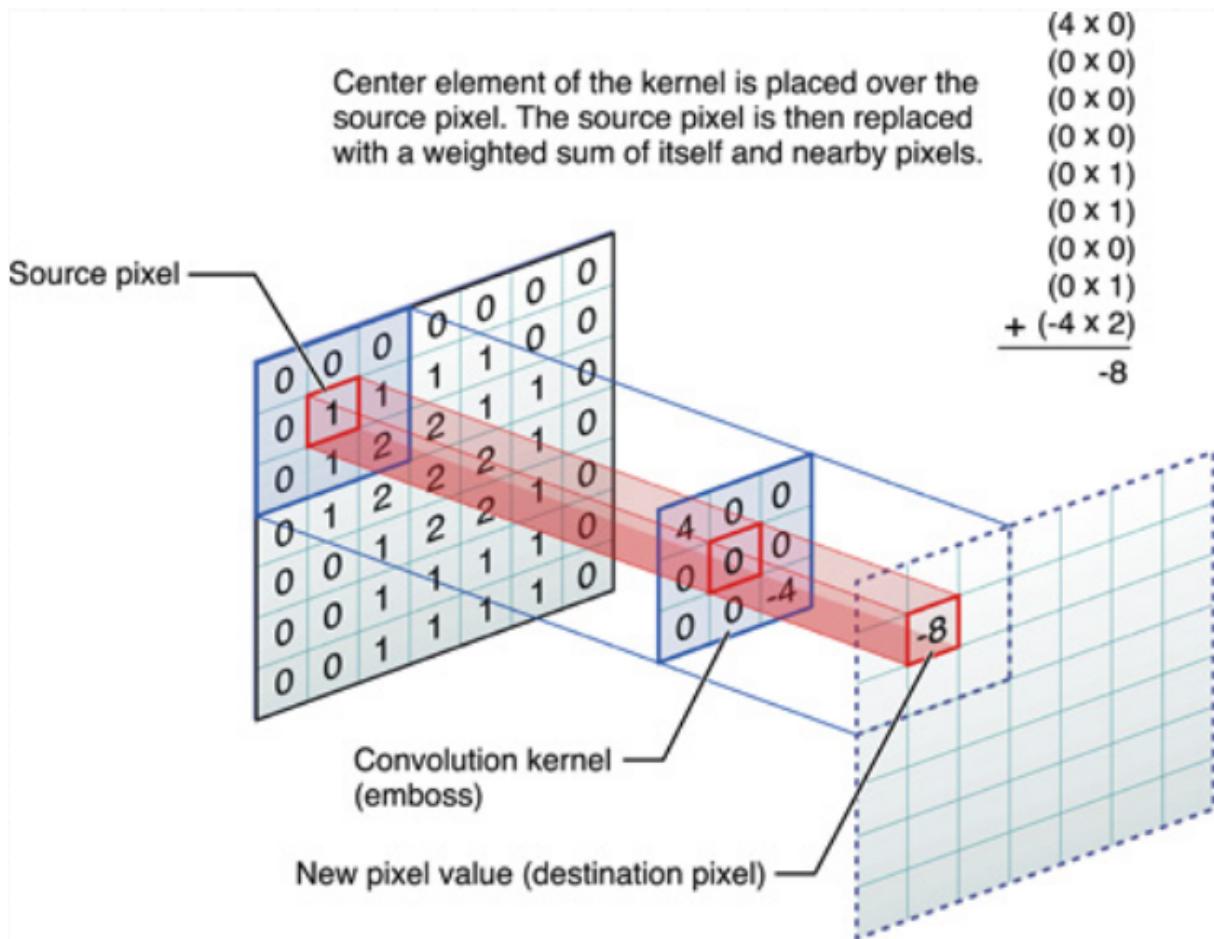


Figure 2.3: Εφαρμογή ενός μικρού μεγέθους φίλτρου σε μία εικόνα και το αποτελεσμά της link.

Στην [Figure 2.4](#) βλέπουμε την πρώτη αρχιτεκτονική CNN από τον Yann Lecun για εφαρμογή σε προβλήματα αναγνώρισης ψηφίων.

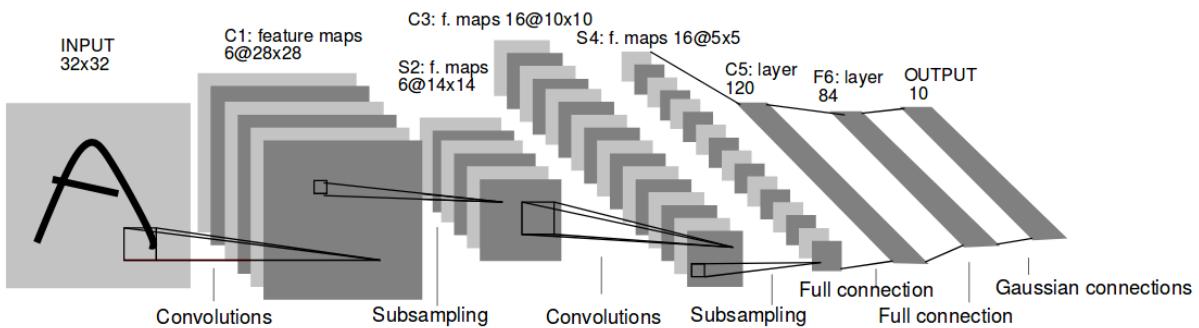


Figure 2.4: **CNN LeNet-5.** Αρχιτεκτονική του πρώτου Συνελικτικού Νευρωνικού Δικτύου για αναγνώριση ψηφίων από εικόνες. Κάθε επίπεδο αποτελεί ένα χαρτογράφημα των χαρακτηριστικών [30].

To Alex-Net [Figure 2.5](#) ήταν μια δημιουργία των Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton σηματοδότησε μια νέα εποχή στην υπολογιστική όραση καθώς πλέον περάσαμε στα *Bαθειά N.Δ.* Το εφάρμοσαν σε ένα από τα πιο προκλητικά προβλήματα, το *Image-Net* [13]. Η συγκεκριμένη αρχιτεκτονική κατάφερε να πετύχει ένα σημαντικό

αποτέλσμα μειώνοντας πάνω από 10% το σφάλμα σε σχέση με τον προηγούμενο νικητή το 2012, πάνω σε ένα πρόβλημα με 15 εκατομμύρια εικόνες και 1000 κατηγορίες για αναγνώριση. Σε αυτό το μοντέλο ήταν και η πρώτη εφαρμογή των γραμμικών ανορθωτών ως συνάρτηση ενεργοποίησης αλλά και η χρήση συνθετικών δεδομένων. Αυτή η συνεισφορά είναι τόσο σημαντική καθώς οι περισσότερες τεχνικές χρησιμοποιούνται μέχρι και σήμερα.

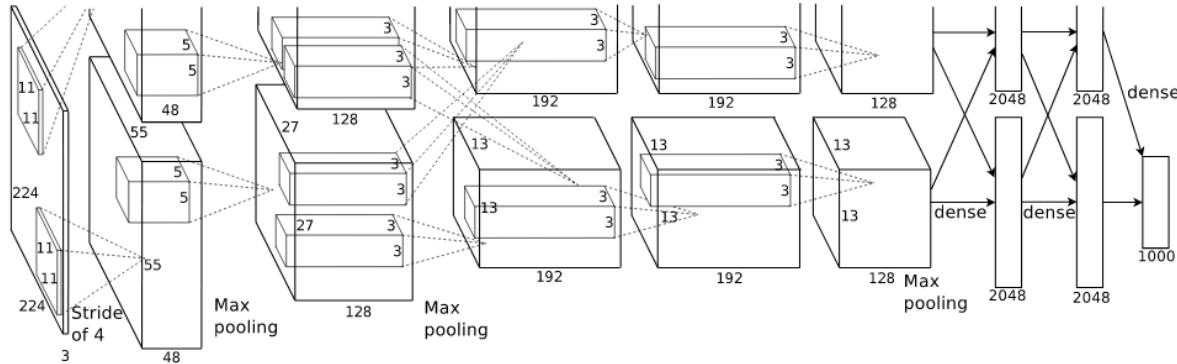


Figure 2.5: **Alex-Net**. Αρχιτεκτονική του Alex-Net ένα από τα πρώτα Βαθειά N.Δ με 60 εκατομμύρια παραμέτρους και 650.000 νευρώνες. [27].

2.3 Πλήρως Συνελικτικά Νευρωνικά Δίκτυα

Μία ακόμα κατηγορία N.Δ. η οποία ανήκει στα Σ.Ν.Δ και θα ασχοληθούμε στο υπόλοιπο της εργασίας είναι τα Πλήρως Σ.Ν.Δ (Fully-CNN) [38]. Η κύρια διαφόρα με τα Σ.Ν.Δ είναι η απώλεια πλήρως συνδεδεμένων επιπέδων στην έξοδο (fully-connected layers), εν αντιθέσει με τα Σ.Ν.Δ που είδαμε προηγουμένως, δηλαδή τα Π.Σ.Ν.Δ. μαθαίνουν πληροφορία μόνο από φίλτρα. Τα Π.Σ.Ν.Δ θεωρούνται κατάλληλα για προβλήματα Σημασιολογικής Κατάτμησης αντικειμένων από εικόνες (εικόνα 2.6).

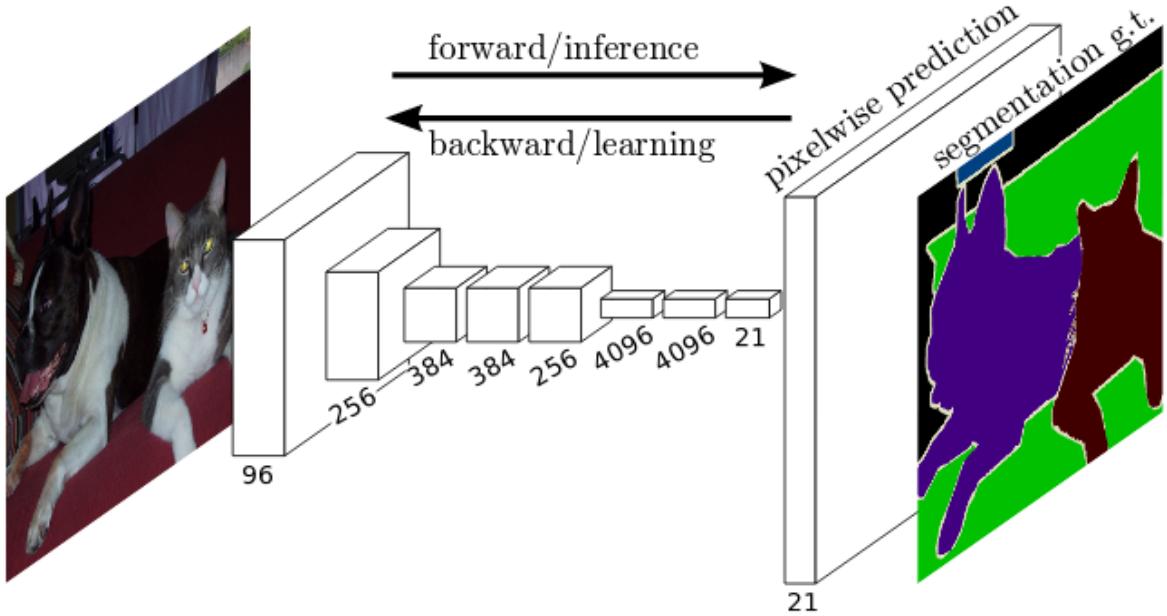


Figure 2.6: Πλήρως Σ.Ν.Δ (Fully CNN) μπορούν να μάθουν αποδοτικά να κάνουν προβλέψεις σε προβλήματα αναγνώρισης σε επίπεδο εικονοστοιχείων [38].

Οι δύο παρακάτω εξισώσεις 2.1, 2.2 πρέπει να ισχύουν για να έχουμε ένα ΠΣΝΔ. Πιο συγκεκριμένα, η μεταβλητή x_{ij} αντιπροσωπεύει το διάνυσμα με τα χαρακτηριστικά στην θέση (i, j) σε κάποιο επίπεδο συνέλιξης και y_{ij} είναι το διάνυσμα εισόδου για το επόμενο επίπεδο ή το διάνυσμα εξόδου του παρόντος επιπέδου που θέλουμε να υπολογίσουμε. Το k συμβολίζει συμβολίζει το μέγεθος του πυρήνα του φίλτρου, ενώ s είναι το άλμα που μπορούμε να θέσουμε στο φίλτρο ή ο παράγοντας της υπο-δειγματοληψίας. Με f_{ks} ορίζεται ο τύπος του επιπέδου στο ΣΝΔ, δηλαδή ένας πολλαπλασιασμός πινάκων σε περίπτωση συνέλιξης ή ένα επίπεδο συγκέντρωσης (pooling) και διάφορα άλλα διαφορετικού τύπου επίπεδα.

$$y_{ij} = f_{ks}(\{x_{si+\delta i, sj+\delta j}\}_{0 \leq \delta i, \delta j < k}) \quad (2.1)$$

Με επιλεγμένο μέγεθος του πυρήνα και του άλματος, θα πρέπει να ισχύει η παρακάτω συνάρτηση μετασχηματισμού. Γενικά ένα NN υπολογίζει μια Μη-γραμμική συνάρτηση, ενώ ένα ΣΝΔ αυτής της μορφής υπολογίζει μη-γραμμικά φίλτρα.

$$f_{ks} \circ g_{k's'} = (f \circ g)_{k'+(k-1)s', ss'} \quad (2.2)$$

Τα κύρια θετικά στοιχεία που καταστούν τα Π.Σ.Ν.Δ κατάλληλα για Σημασιολογική Κατάτυπηση είναι:

1. Χρησιμοποιούν όλη την πληροφορία της εικόνας και μαθαίνουν καθολική πληροφορία.
2. Κρατάνε την χωρική πληροφορία (spatial information) από την εικόνα.
3. Είναι πιο γρήγορα στην εκπαίδευση αλλά και στην συμπερασματολογία.
4. Είναι αμετάβλητα ως προς το μέγεθος εισόδου της εικόνας.

2.4 Εμπρόσθια Διάδοση

Η Εμπρόσθια Διάδοση (Forward Propagation) είναι ο τρόπος με τον οποίο το ΣΝΔ επεξεργάζεται τα δεδομένα, δηλαδή ο τρόπος με τον οποίο τα ΣΝΔ δίνουν πιθανότητες για την κατηγοριοποίηση των δεδομένων. Αρχικά όμως δούμε με λεπτομέρεια την διαδικασία της συνέλιξης η οποία αποτελεί την βασική πράξη για την υλοποίηση των ΣΝΔ.

Τα ΣΝΔ αποτελούνται από συνελικτικά επίπεδα τα οποία χαρακτηρίζονται από έναν χάρτη χαρακτηριστικών εισόδου I , μια σειρά από φίλτρα K και τις πολώσεις b (biases).

Στην περίπτωση που έχουμε εικόνες για αναγνώριση, η είσοδος αποτελείται από μια εικόνα με ύψος H , πλάτος W και αριθμό καναλιών $C = 3$ (χόκκινο, μπλε πράσινο), $I \in \mathbb{R}^{H \times W \times C}$. Επομένως, για μια σειρά από D φίλτρα έχουμε $K \in \mathbb{R}^{k_1 \times k_2 \times C \times D}$ και μεροληφίες $b \in \mathbb{R}^D$, ένα για κάθε φίλτρο. Συνεπώς, το αποτέλεσμα της συνέλιξης είναι όπως βλέπουμε παρακάτω:

$$(I * K)_{ij} = \sum_{m=0}^{k_1-1} \sum_{n=0}^{k_2-1} K_{m,n,c} \cdot I_{i+m,j+n,c} + b \quad (2.3)$$

Για την εξήγηση των μεθόδων της εμπρόσθιας διάδοσης και οπισθοδρόμησης, παραθέτουμε τις παρακάτω συμβολισμούς, για λόγους απλότητας ότι $C = 1$:

1. l : συμβολίζει το l -στο επίπεδο όπου $l = 1$ είναι το πρώτο επίπεδο και $l = L$ το τελευταίο επίπεδο.
2. x είναι η είσοδος με διαστάσεις $H \times W$ και με i, j συμβολίζουμε τους δείκτες του πολυδιάστατου διανύσματος.
3. Φίλτρο ή πυρήνας w διαστάσεων $k_1 \times k_2$ όπου έχει ως δείκτες m, n .
4. $w_{m,n}^l$ είναι ο πίνακας με τα βάρη που συνδέει τους νευρώνες του επιπέδου l με τους νευρώνες του επιπέδου $l - 1$.
5. $x_{i,j}^l$ είναι το διάνυσμα εισόδου του επιπέδου l μαζί με το διάνυσμα της μεροληφίας:

$$x_{i,j}^l = \sum_m \sum_n w_{m,n}^l o_{i+m,j+n}^{l-1} + b^l$$

6. b^l είναι το διάνυσμα της μεροληφίας.
 7. $o_{i,j}^l$ είναι το διάνυσμα εξόδου στο επίπεδο l :
- $$o_{i,j}^l = f(x_{i,j}^l)$$
8. $f(\cdot)$ είναι η συνάρτηση ενεργοποίησης, η οποία εφαρμόζεται στην είσοδο μετά την διαδικασία της συνέλιξης στο επίπεδο l .

Για να εκτελέσουμε την συνέλιξη ο πυρήνας περιστρέφεται κατά 180 μοίρες και ολισθαίνει στον χάρτη χαρακτηριστικών εισόδου με προκαθορισμένο βήμα και άλμα. Σε κάθε θέση, το γινόμενο μεταξύ κάθε στοιχείου του πυρήνα και του στοιχείου εισόδου του χάρτη χαρακτηριστικών, υπολογίζονται και τα αποτελέσματα προσθέτονται για να αναπαραχθεί το αποτέλεσμα στην παρούσα θέση.

Αυτή η διαδικασία επαναλαμβάνεται χρησιμοποιώντας διαφορετικούς πυρήνες για να παραχθούν όσο γίνεται περισσότεροι χάρτες χαρακτηριστικών επιθυμούμε. Η έννοια της μοιρασιάς των βαρών όπως ονομάζεται επιδεικνύεται στο παρακάτω διάγραμμα:

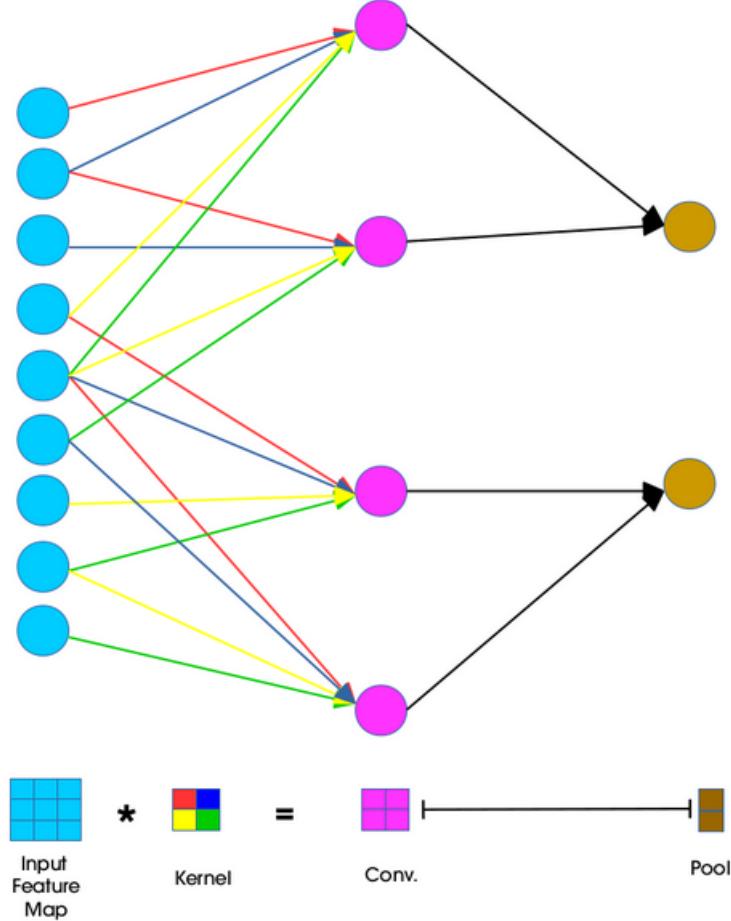


Figure 2.7: Διάγραμμα που επιδεικνύει την διαδικασία της συνέλιξης ενός πυρήνα 2×2 με έναν χάρτη χαρακτηριστικών εισόδου μεγέθους 3×3 . Με καφέ χρώμα βλέπουμε το επίπεδο της συγκέντρωσης των χαρακτηριστικών.

Τα στοιχεία στο συνελικτικό επίπεδο που διαχρίνουμε πιο πάνω στην εικόνα έχουν δεκτικό πεδίο μεγέθους 4 στον χάρτη χαρακτηριστικών και για αυτό είναι συνδεδεμένα μόνο σε 4 γειτονικούς νευρώνες του επιπέδου εισόδου. Αυτή είναι η ιδέα της **αραιής σύνδεσης (sparse connectivity)** στα ΣΝΔ όπου και διατηρείται το πρότυπο της τοπικής σύνδεσης με τους νευρώνες ανάμεσα σε γειτονικά επίπεδα.

Τα χρώματα του πυρήνα υποδεικνύουν τα βάρη τα οποία συνεισφέρουν στην συνέλιξη και μας δείχνουν πως τα βάρη του πυρήνα κατανέμονται στο μεταξύ των νευρώνων στα γειτονικά επιπέδα. Τα βάρη με ίδιο χρώμα είναι ταυτόσημα.

Στην παρακάτω εικόνα βλέπουμε τον πυρήνα ο οποίος έχει αναστραφεί οριζοντίως και καθέτως:

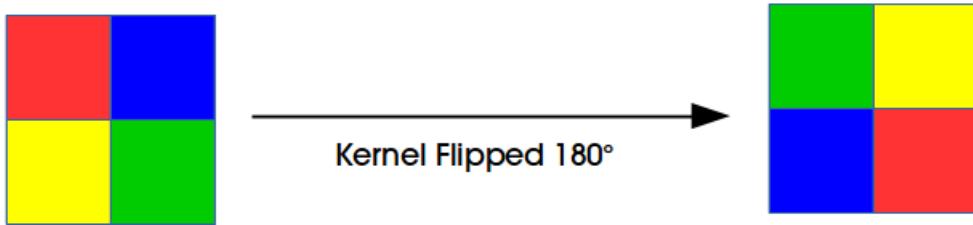


Figure 2.8: Αναστροφή πυρήνα κατά 180 μοίρες.

Η συνάρτηση της συνέλιξης της εισόδου στο επίπεδο l δίνεται από:

$$x_{i,j}^l = \text{rot}_{180} \{w_{m,n}^l\} * o_{i,j}^{l-1} + b_{i,j}^l \quad (2.4)$$

$$x_{i,j}^l = \sum_m \sum_n w_{m,n}^l o_{i+m, j+n}^{l-1} + b_{i,j}^l \quad (2.5)$$

$$o_{i,j}^l = f(x_{i,j}^l) \quad (2.6)$$

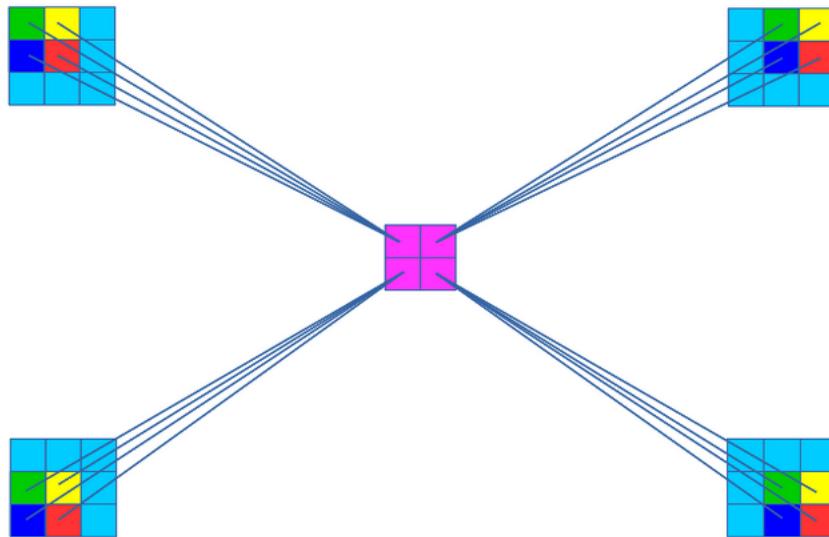


Figure 2.9: Διάγραμμα που επιδεικνύει την διαδικασία της συνέλιξης ενός πυρήνα 2×2 με έναν χάρτη χαρακτηριστικών εισόδου μεγέθους 3×3 σε όλα τα στάδια ολίσθησης δείχνοντας την συνεισφορά του κάθε στοιχείου εισόδου με τα γειτονικά στοιχεία στο στοιχείο εξόδου.

2.5 Αλγόριθμος Οπισθοδρόμησης

Ο τρόπος με τον οποίο εκπαιδεύονται τα N.Δ. όπως και τα Σ.Ν.Δ. είναι ο αλγόριθμος Οπισθοδρόμησης (Backpropagation). Η Οπισθοδρόμηση στα Σ.Ν.Δ. διαφέρει σε σχέση με τα κοινά N.Δ. καθώς ένα βήμα οπισθοδρόμησης μοντελοποιείται σαν μια πράξη συνέλιξης αλλά εφαρμόζοντας το φίλτρο ανάποδα κατά 180 μοίρες (εξίσωση 2.17). Ο αλγόριθμος οπισθοδρόμησης αποτελεί το επόμενο βήμα μετά τον αλγόριθμο εμπρόσθιας διάδοσης, όπου γίνονται οι πρώτες εκτιμήσεις για την κατηγοριοποίηση των δεδομένων αφού έχουμε

υπολογίζει το σφάλμα με κάποια συνάρτηση κόστους και επαναζυγίζουμε τα βάρη των πυρήνων στο $\Sigma \Delta$ για την καλύτερη κατηγοριοποίηση των δεδομένων.

Κατά την Οπισθοδρόμηση πραγματοποιούνται δύο ενημερώσεις πινάκων, μία για τα βάρη και μία των συντελεστών δέλτα (δ). Στην ενημέρωση των βαρών υπολογίζουμε την $\frac{\partial E}{\partial w_{m',n'}}$ η οποία συμβολίζει την επίδραση στην συνάρτηση κόστους E που μπορεί να επιφέρει μια αλλαγή των βαρών $w_{m',n'}$ ενός εικονοστοιχείου.

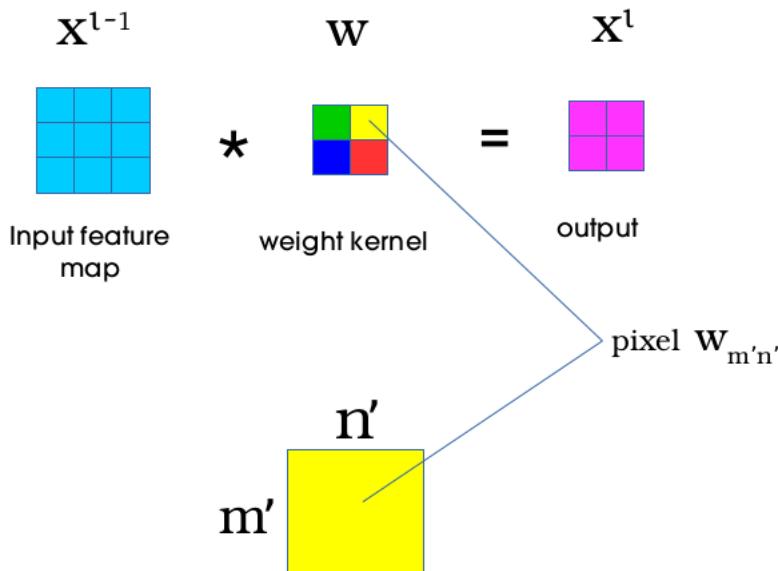


Figure 2.10: Διάγραμμα που δείχνει την διαδικασία της συνέλιξης ενός πυρήνα 2×2 με έναν χάρτη χαρακτηριστικών εισόδου μεγέθους 3×3 επιδεικνύοντας την συνεισφορά στοιχειού του πυρήνα μεταξύ των στοιχείων εισόδου αλλά και των γειτόνων του φέρει επικάλυψη.

Κατά την εμπρόσθια διάδοση, η διαδικασία της συνέλιξης μας βεβαιώνει πως το κίτρινο εικονοστοιχείο $w_{m,n}$ του πυρήνα βαρών συνεισφέρει σε όλα τα γινόμενα μεταξύ κάθε στοιχείου του πυρήνα και της περιοχής που έχει επικάλυψη με τον χάρτη χαρακτηριστικών εισόδου. Κατά συνέπεια, αυτό σημαίνει ότι το εικονοστοιχείο $w_{m,n}$ θα επηρεάσει όλα τα στοιχεία της εξόδου του χάρτη χαρακτηριστικών.

Η συνέλιξη ανάμεσα στον χάρτη χαρακτηριστικών διαστάσεων $H \times W$ και στον πυρήνα βαρών διαστάσεων $k_1 \times k_2$, παράγει ως έξοδο έναν χάρτη χαρακτηριστικών διαστάσεων $(H - K_1 + 1) \times (W - K_2 + 1)$. Ο υπολογισμός της κλίσης για κάθε ένα από τα στοιχεία του πυρήνα βαρών μπορεί να παρθεί εφαρμόζοντας τον κανόνα της αλυσίδας:

$$\begin{aligned} \frac{\partial E}{\partial w_{m,n}^l} &= \sum_{i=0}^{H-k_1} \sum_{j=0}^{W-k_2} \frac{\partial E}{\partial x_{i,j}^l} \frac{\partial x_{i,j}^l}{\partial w_{m',n'}^l} \\ &= \sum_{i=0}^{H-k_1} \sum_{j=0}^{W-k_2} \delta_{i,j}^l \frac{\partial x_{i,j}^l}{\partial w_{m',n'}^l} \end{aligned} \quad (2.7)$$

Στην εξίσωση 2.7, η μεταβλητή $x_{i,j}^l$ ισούται με $\sum_m \sum_n w_{m,n}^l o_{i+m,j+n}^{l-1} + b^l$ και επεκτείνοντας αυτή την εξίσωση παίρνουμε:

$$\frac{\partial x_{i,j}^l}{\partial w_{m',n'}^l} = \frac{\partial}{\partial w_{m,n}^l} \left(\sum_m \sum_n w_{m,n}^l o_{i+m,j+n}^{l-1} + b^l \right) \quad (2.8)$$

Επεκτείνοντας περισσότερο την παραπάνω εξίσωση και υπολογιζόντας τις μερικές διαφορικές από όλα τα στοιχεία παίρνουμε μηδενικό αποτέλεσμα εκτός από τα στοιχεία όπου $m = m'$ και $n = n'$ στο $w_{m,n}^l o_{i+m,j+n}^{l-1}$:

$$\begin{aligned} \frac{\partial x_{i,j}^l}{\partial w_{m',n'}^l} &= \frac{\partial}{\partial w_{m',n'}^l} \left(w_{0,0}^l o_{i+0,j+0}^{l-1} + \cdots + w_{m',n'}^l o_{i+m',j+n'}^{l-1} + \cdots + b^l \right) \\ &= \frac{\partial}{\partial w_{m',n'}^l} \left(w_{m,n}^l o_{i+m,j+n}^{l-1} \right) \\ &= o_{i+m,j+n}^{l-1} \end{aligned} \quad (2.9)$$

Αντικαθιστώντας την εξίσωση 2.9 στην 2.7 μας δίνει το εξής αποτέλεσμα:

$$\begin{aligned} \frac{\partial E}{\partial w_{m,n}^l} &= \sum_{i=0}^{H-k_1} \sum_{j=0}^{W-k_2} \delta_{i,j}^l o_{i+m',j+n'}^{l-1} \\ &= rot_{180} \left\{ \delta_{i,j}^l \right\} * o_{m',n'}^{l-1} \end{aligned} \quad (2.10)$$

Η διπλή σειρά στην εξίσωση 2.10 είναι ως αποτέλεσμα του μοιράσματος βαρών του πυρήνα στο Σ.Ν.Δ, δηλαδή ο ίδιος πυρήνας βαρών ολισθαίνει κατα μηκός και πλάτος σε ολόκληρο τον χάρτη χαρακτηριστικών εισόδου. Οι σειρές αναπαριστούν μια συλλογή όλων των κλίσεων $\delta_{i,j}^l$ που προέρχονται από όλα τα στοιχεία εξόδου του επιπέδου l .

Λαμβάνοντας τις κλίσεις ως προς τους χάρτες των φίλτρων, έχουμε μια πράξη αυτό-συσχέτισης η οποία μετατρέπεται σε μια πράξη συνέλιξης αναστρέφοντας τους πίνακες $\delta_{i,j}^l$ οριζοντίως και καθέτως, με τον ίδιο τρόπο που αναστρέφουμε τα φίλτρα κατά την Οπισθοδρόμηση.



Figure 2.11: Παράδειγμα αναστροφής ενός πίνακα με συντελεστές δέλτα κατά 180 μοίρες.

Το παρακάτω διάγραμμα μας δείχνει ένα βήμα της Οπισθοδρόμησης σε ένα επίπεδο επιδεικνύοντας τις κλίσεις ($\delta_{11}, \delta_{12}, \delta_{21}, \delta_{22}$) που αναπαράγονται κατά την διαδικασία της Οπισθοδρόμησης:

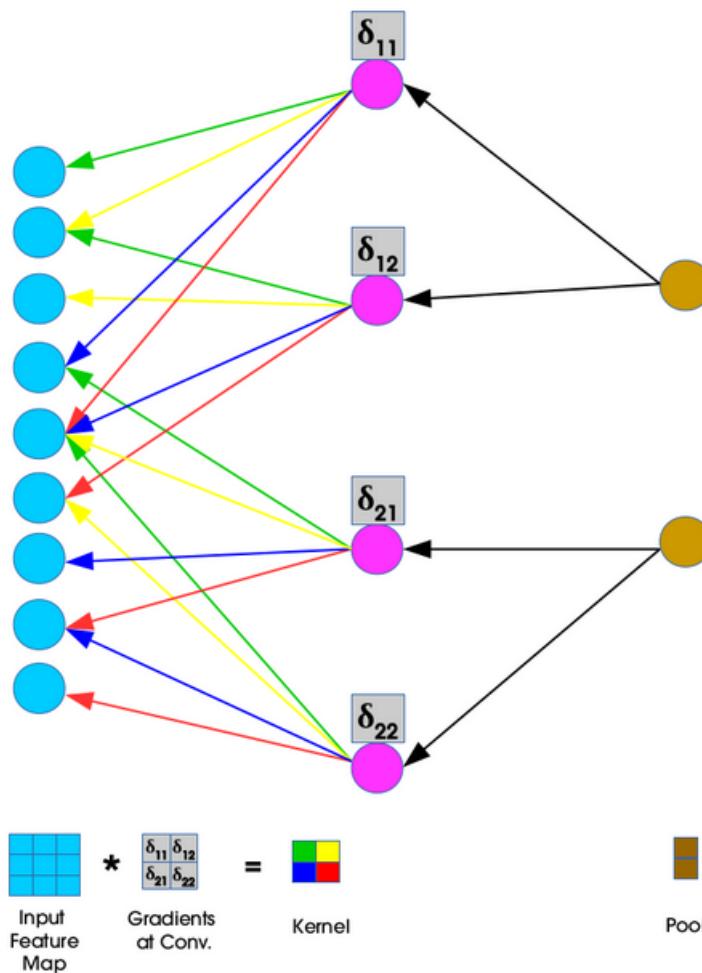


Figure 2.12: Επισκόπηση Οπισθοδρόμησης σε ένα επίπεδο συνέλιξης.

Η συνέλιξη χρησιμοποιήθηκε για την απόκτηση των καινούριων βαρών όπως φαίνεται στην παρακάτω εικόνα:

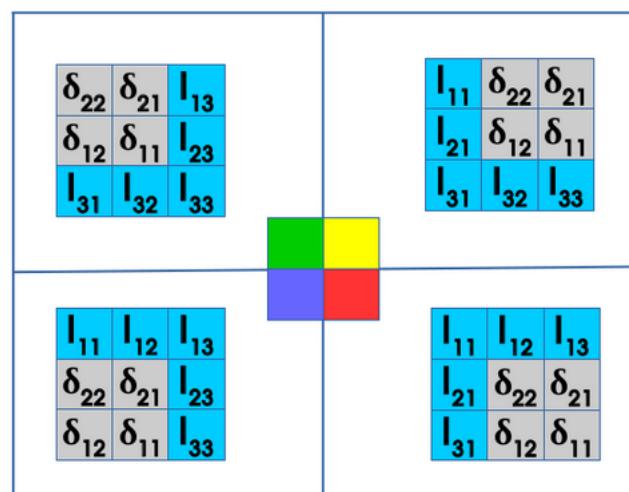


Figure 2.13: Συνέλιξη για την απόκτηση των νέων βαρών.

Κατά την διαδικασία της ανακατασκευής, χρησιμοποιούνται οι μεταβλητές δ ($\delta_{11}, \delta_{12}, \delta_{21}, \delta_{22}$) οι οποίες προκύπτουν από την παρακάτω εξίσωση:

$$\delta_{i,j}^l = \frac{\partial E}{\partial x_{i,j}^l} \quad (2.11)$$

Ο υπολογισμός των μεταβλητών δ από την παραπάνω εξίσωση επιδεικνύει των υπολογισμό των αποχλίσεων. Δηλαδή, η εξίσωση μεταφράζεται ως μια μέτρηση, κατά πόσο μια αλλαγή ενός εικονοστοιχείου $x_{i,j}^l$ στον χάρτη χαρακτηριστικών εισόδου επηρεάζει την συνάρτηση κόστους E .

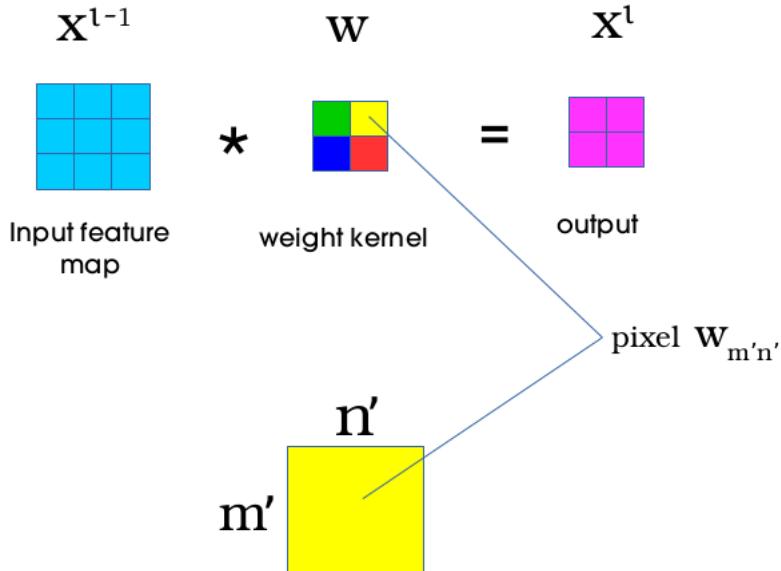


Figure 2.14: Συμβολή της περιοχής εισόδου στο εικονοστοιχείο της εξόδου κατά την συνέλιξη.

Από το παραπάνω διάγραμμα, είναι φανερό ότι η περιοχή η οποία επηρεάζει το εικονοστοιχείο $x_{i',j'}$ από την είσοδο είναι η περιοχή η οποία στην έξοδο περιορίζεται από τις διακεκομένες γραμμές. Το εικονοστοιχείο της επάνω αριστερής γωνίας δίνεται από $(i' - k_1 + 1, j' - k_2 + 1)$ και το κάτω δεξιά εικονοστοιχείο δίνεται από (i', j') .

Χρησιμοποιώντας τον κανόνα της αλυσίδας και των σειρών προκύπτει η παρακάτω εξίσωση:

$$\begin{aligned} \frac{\partial E}{\partial x_{i',j'}^l} &= \sum_{i,j \in Q} \frac{\partial E}{\partial x_Q^{l+1}} \frac{\partial x_Q^{l+1}}{\partial x_{i',j'}^l} \\ &= \sum_{i,j \in Q} \delta_Q^{l+1} \frac{\partial x_Q^{l+1}}{\partial x_{i',j'}^l} \end{aligned} \quad (2.12)$$

Ως Q στο άθροισμα παραπάνω ορίζεται περιοχή εξόδου η οποία περιβάλλεται από τις διακεκομένες γραμμές και η συνθεσή του αποτελείται από τα εικονοστοιχεία στην έξοδο τα οποία επηρεάζονται από το εικονοστοιχείο $x_{i',j'}$ του χάρτη χαρακτηριστικών εισόδου. Μία πιο προσιτή αναπάρασταση της παραπάνω εξίσωσης επιδεικνύεται παρακάτω:

$$\begin{aligned} \frac{\partial E}{\partial x_{i',j'}^l} &= \sum_{m=0}^{k_1-1} \sum_{n=0}^{k_2-1} \frac{\partial E}{\partial x_{i'-m,j'-n}^{l+1}} \frac{\partial x_{i'-m,j'-n}^{l+1}}{\partial x_{i',j'}^l} \\ &= \sum_{m=0}^{k_1-1} \sum_{n=0}^{k_2-1} \delta_{i'-m,j'-n}^{l+1} \frac{\partial x_{i'-m,j'-n}^{l+1}}{\partial x_{i',j'}^l} \end{aligned} \quad (2.13)$$

Στην περιοχή Q , το ύψος ορίζεται από $i' - 0$ έως $i' - (k_1 - 1)$ και το πλάτος από $j' - 0$ έως $j' - k_2 - 1$. Για λόγους ευχέρειας, οι δύο ορισμοί μπορούν να αναπαραστηθούν από $i' - m$ και $j' - n$ στην σειρά αθροίσματος από την στιγμή που οι μεταβλητές m και n ορίζονται στις ίδιες περιοχές, από $0 \leq m \leq k_1 - 1$ και $0 \leq n \leq k_2 - 1$.

Στην εξίσωση 2.13, $x_{i'-m,j'-n}^{l+1}$ ισοδυναμεί με $w_{m',n'}^{l+1} o_{i'-m+m',j'-n+n'}^l + b^{l+1}$ και επεκτείνοντας την αυτό το χομμάτι της εξίσωσης μας δίνει:

$$\begin{aligned} \frac{\partial x_{i'-m,j'-n}^{l+1}}{\partial x_{i',j'}^l} &= \frac{\partial}{\partial x_{i',j'}^l} \left(\sum_{m'} \sum_n w_{m',n'}^{l+1} o_{i'-m+m',j'-n+n'}^l + b^{l+1} \right) \\ &= \frac{\partial}{\partial x_{i',j'}^l} \left(\sum_{m'} \sum_n w_{m',n'}^{l+1} f(x_{i'-m+m',j'-n+n'}) + b^{l+1} \right) \end{aligned} \quad (2.14)$$

Επεκτείνοντας λίγο περισσότερο την εξίσωση 2.13 και υπολογίζοντας τις μερικές παραγώγους για όλα τα στοιχεία, καταλήγουμε σε μηδενικές τιμές εκτός των στοιχείων που $m' = m$ και $n' = n$ έτσι ώστε η εξίσωση $f(x_{i'-m+m',j'-n+n'}^l)$ γίνεται $f(x_{i',j'})$ και η $w_{m',n'}^{l+1}$ μετατρέπεται σε $w_{m,n}^{l+1}$. Επομένως αντικαθιστώντας στην παραπάνω εξίσωση προκύπτει:

$$\begin{aligned} \frac{\partial x_{i'-m,j'-n}^{l+1}}{\partial x_{i',j'}^l} &= \frac{\partial}{\partial x_{i',j'}^l} \left(w_{m',n'}^{l+1} f(x_{0-m+m',0-n+n'}^l) + \dots + w_{m,n}^{l+1} f(x_{i',j'}^l) + \dots + b^{l+1} \right) \\ &= \frac{\partial}{\partial x_{i',j'}^l} \left(w_{m,n}^{l+1} f(x_{i',j'}^l) \right) \\ &= w_{m,n}^{l+1} \frac{\partial}{\partial x_{i',j'}^l} f(x_{i',j'}^l) \\ &= w_{m,n}^{l+1} f'(x_{i',j'}^l) \end{aligned} \quad (2.15)$$

Αντικαθιστώντας την εξίσωση 2.15 στην εξίσωση 2.13 μας δίνει το ακόλουθο αποτέλεσμα:

$$\frac{\partial E}{\partial x_{i',j'}^l} = \sum_{m=0}^{k_1-1} \sum_{n=0}^{k_2-1} \delta_{i'-m,j'-n}^{l+1} w_{m,n}^{l+1} f'(x_{i',j'}^l) \quad (2.16)$$

Για την Οπισθοδρόμηση, χρησιμοποιούμε τον ανάστροφο πυρήνα και ως αποτέλεσμα έχουμε μια συνέλιξη η οποία εκφράζεται ως μια πράξη αυτό-συσχέτισης με ανάστροφο πυρήνα:

$$\begin{aligned}
 \frac{\partial E}{\partial x_{i',j'}^l} &= \sum_{m=0}^{k_1-1} \sum_{n=0}^{k_2-1} \delta_{i'-m,j'-n}^{l+1} w_{m,n}^{l+1} f'(x_{i',j'}^l) \\
 &= rot_{180} \left\{ \sum_{m=0}^{k_1-1} \sum_{n=0}^{k_2-1} \delta_{i'+m,j'+n}^{l+1} w_{m,n}^{l+1} \right\} f'(x_{i',j'}^l) \\
 &= \delta_{i',j'}^{l+1} * rot_{180} \left\{ w_{m,n}^{l+1} \right\} f'(x_{i',j'}^l)
 \end{aligned} \tag{2.17}$$

Κάτι ακόμα που αξίζει να συζητηθεί είναι η διαδικασία της οπισθοδρόμησης σε ένα επίπεδο συγκέντρωσης. Η βασική λειτουργία ενός επιπέδου συγκέντρωσης είναι να μειώνει σταδιακά τα χωρικά μεγέθη των επιπέδων ενός Σ.Ν.Δ. πετυχαίνοντας την μείωση των παραμέτρων και του υπολογιστικού χρόνου αλλά και στην μείωση της υπερμάθησης. Σε ένα επίπεδο συγκέντρωσης δεν λαμβάνει χώρα κάποια πράξη μάθησης [29].

Τα επίπεδα συγκέντρωσης συνήθως λειτουργούν με μεθόδους όπως μέγιστης συγκέντρωση, μέση συγκέντρωση και ακόμα και *L2 – norm* συγκέντρωση. Στο επίπεδο συγκέντρωσης, η εμπρόσθια διάδοση έχει ως αποτέλεσμα στην έξοδο, έναν μειωμένο χάρτη χαρακτηριστικών όπου έχει εφαρμοστεί ένα $N \times N$ τμήμα συγκέντρωσης σε κάθε περιοχή και στην έξοδο εξέρχεται μόνο ένα στοιχείο από το τμήμα. Η Οπισθοδρόμηση στο επίπεδο συγκέντρωσης, οπισθοδρομεί το σφάλμα το οποίο έχει προέλθει από την μοναδική επιχρατέστερη τιμή του εκάστοτε τμήματος.

Για να κρατήσουμε την θέση της επιχρατέστερης τιμής από το επίπεδο συγκέντρωσης, σημειώνουμε την θέση κατά την εμπρόσθια διάδοση και μετά την χρησιμοποιούμε για να οδηγήσουμε τις αποκλίσεις κατά την οπισθοδρόμηση. Η δρυπολόγηση των αποκλίσεων επιτυγχάνεται με κάποια από τις παρακάτω μεθόδους:

- **Max-Pooling** Το σφάλμα απλώς ανατίθεται στο στοιχείο το οποίο επιχράτησε κατά το εμπρόσθιο πέρασμα. Επειδή, τα υπόλοιπα στοιχεία δεν έιχαν συνεισφέρει δεν τους ανατίθεται κάποια τιμή παρά μόνο το μηδέν.
- **Average-Pooling** Το σφάλμα πολλαπλασιάζεται με τον παράγοντα $\frac{1}{N \times N}$ και η προκύπτουσα τιμή ανατίθεται σε ολόκληρο το τμήμα συγκέντρωσης, δηλαδή όλα τα στοιχεία παίρνουν την ίδια τιμή.

Chapter 3

Μεθοδολογία

3.1 Εισαγωγή

Στο κεφάλαιο αυτό θα συζητήσουμε για τις μεθόδους και τις τεχνικές που χρησιμοποιήθηκαν στην εργασία μας αλλά και την ανάλυση με λεπτομέρειες των αλγορίθμων που εφαρμόστηκαν. Συγκεκριμένα, θα δούμε τις αρχιτεκτονικές βαθειάς μάθησης που χρησιμοποιήσαμε στα πειράματα, καθώς και την θεωρία αυτών. Η Μεθοδολογία μας βασίστηκε στους αλγορίθμους των Νευρωνικών Δικτύων και πιο συγκεκριμένα στα Πλήρως Συνελικτικά Νευρωνικά Δίκτυα FCNN που έχουν εφαρμογές σε προβλήματα της όρασης υπολογιστών και πιο συγκεκριμένα στην Σημασιολογική Κατάτυπη πληροφορίες από εικόνες. Η προσέγγιση μας περιλαμβάνει δύο μοντέλα τα οποία αποτελούνται από τρία στάδια: Κωδικοποίηση Χαρακτηριστικών, Παράλληλη Επεξεργασία και Αποκωδικοποίηση (Encoder-Parallel Processing-Decoder).

3.2 Πρώτη Προσέγγιση

Η πρώτη μας προσέγγιση στο πρόβλημα ήταν βασισμένη σε μια παράλληλη αρχιτεκτονική από πολλαπλά Σ.Ν.Δ (εξίσωση 3.1). Η ιδέα στηρίχθηκε στην υλοποίηση τεσσάρων Σ.Ν.Δ. όπου τα 3 από αυτά δέχονται σαν είσοδο σειριακά κομμάτια από την εικόνα, η μέθοδος αυτή αναφέρεται ως ‘Ολίσθηση Παραθύρων’ (*Sliding-Windows*) [37]. Τα τρία από τα τέσσερα Σ.Ν.Δ. δέχονται κομμάτια διαφορετικού μεγέθους από την εικόνα, ενώ το τέταρτο Σ.Ν.Δ. δέχεται σαν είσοδο ολόκληρη την εικόνα. Έτσι παίρνουμε 4 διαφορετικές προβλέψεις για κάθε εικονοστοιχείο και αποφασίζουμε κατά πλειψηφία το επικρατέστερο.

Η ιδέα αυτή αν και πολύ απλή είχε αρχετές δυσκολίες:

1. Τα μικρά κομμάτια τημηματοποιούν κάποια αντικείμενα κατά την εκπαίδευση και καθίσταται δύσκολη η αναγνώριση τους καθώς δεν μαθαίνουν κάποια ολοκληρωμένη δομή από αυτά.
2. Η διαδικασία της δοκιμής ήταν σχεδόν ανέφικτη καθώς ένα τέτοιο μοντέλο είναι πολύ δαπανηρό σε πόρους.
3. Η μέθοδος Ολίσθησης Παραθύρων είναι πολύ αργή.

Σύμφωνα με τα παραπάνω, δεν θα ασχοληθούμε περαιτέρω με αυτή την αρχιτεκτονική, αλλά προχωρήσαμε σε διαφορετική προσέγγιση του προβλήματος όπως θα δούμε στην συνέχεια.

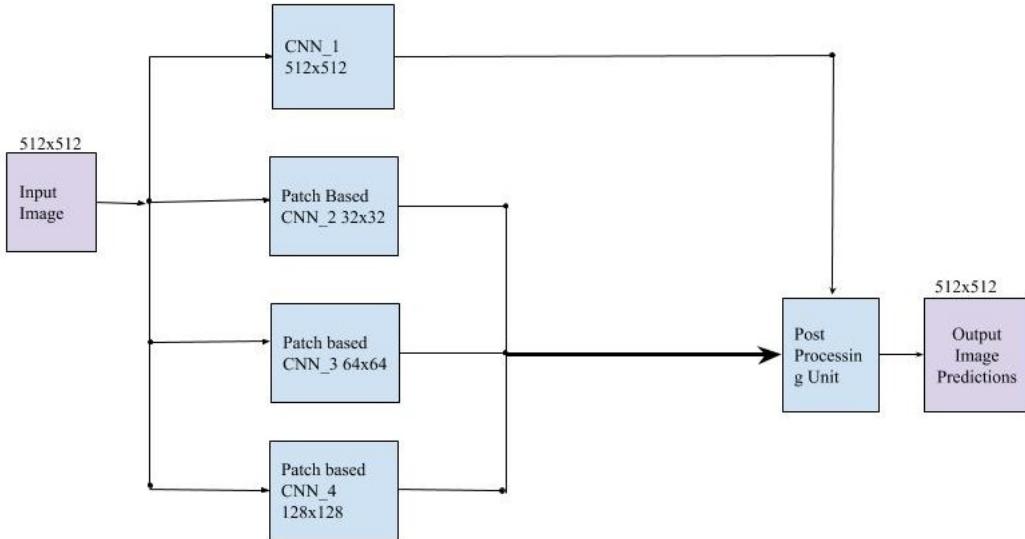


Figure 3.1: Παράλληλη αρχιτεκτονική βασισμένη σε πολλαπλά Σ.Ν.Δ. με διαφορετικά μεγέθη ειδόδου το καθένα.

3.3 Προετοιμασία Δεδομένων

Η προετοιμασία των δεδομένων μας αποτελεί το πρώτο στάδιο, το οποίο χρίνεται αναγκαίο ώστε να γίνει εφικτή η εφαρμογή των αλγορίθμων βαθειάς μάθησης καθώς χωρίς αυτό το στάδιο δεν θα μπορέσουμε να έχουμε τα επιθυμητά αποτελέσματα.

3.3.1 Υποδειγματοληψία

Η βάση δεδομένων μας αποτελείται από εικόνες υψηλής ευχρίνειας. Τα νευρωνικά δίκτυα έχουν πολλά εκατομμύρια παραμέτρους, οι παραμέτροι είναι συναρτήσει της εισόδου του νευρωνικού δικτύου, επομένως η υποδειγματοληψία στις αρχικές εικόνες είναι απαραίτητη για να μπορέσουμε να κάνουμε εφικτά τα πειραματά μας. Αυτή η τεχνική φυσικά έχει κάποιο αντάλλαγμα, η μείωση των διαστάσεων των εικόνων σημαίνει απώλεια σε πληροφορία.

Για την υποδειγματοληψία στις εικόνες χρησιμοποιήθηκαν δύο διαφορετικοί αλγόριθμοι. Ο πρώτος είναι ο αλγόριθμος της Διγραμμικής Παρεμβολής (Bilinear Interpolation) και ο δεύτερος είναι αυτός των Πλησιέστερων Γειτόνων.

Τον αλγόριθμο της Διγραμμικής Παρεμβολής τον χρησιμοποιήσαμε για την υποδειγματοληψία της εικόνας καθώς το εικονοστοιχείο που δημιουργείται χατά την διαδικασία της υποδειγματοληψίας προσεγγίζεται από μια ζυγισμένη εκτίμηση από άλλα τέσσερα σημεία ως μια καλύτερη προσέγγιση των εικονοστοιχείων (εξίσωση 3.1). Ο λόγος που ονομάζεται διγραμμικός είναι επειδή προκύπτει από το γινόμενο δύο γραμμικών συναρτήσεων, επομένως είναι Μη-Γραμμικός αλγόριθμος.

$$f(x, y) = \sum_{i=0}^1 \sum_{j=0}^1 \alpha_{ij} x^i y^j \quad (3.1)$$

Ο αλγόριθμος του Πλησιέστερου Γείτονα, γνωστός και ως αλγόριθμος Παρεμβολής μηδενικής τάξης εφαρμόστηκε στις εικόνες με τις ετικέτες των εικονοστοιχείων (ground truth). Ο λόγος που χρησιμοποιήσαμε αυτήν την απλή προσέγγιση είναι η εξασφάλιση των επιθυμητών ετικετών κατά τη διάρκεια της δειγματοληψίας. Συγκεκριμένα, το καινούριο εικονοστοχείο προέρχεται από το πλησιέστερο ως προς το μέγεθος εικονοστοιχείο, επομένως κάποιος άλλος αλγόριθμος θα μας παραποιούσε τις ετικέτες των εικονοστοιχείων. Η εικόνα 3.2 μας δείχνει μια ποικιλία αλγορίθμων Παρεμβολής.

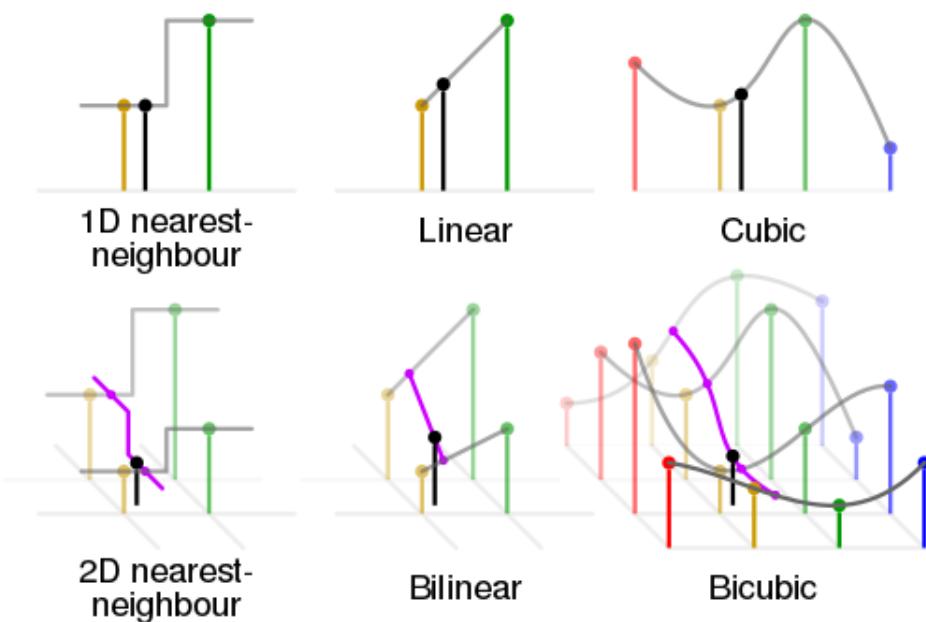


Figure 3.2: Μέθοδοι Παρεμβολής [2].

3.3.2 Κανονικοποίηση Χαρακτηριστικών

Η κανονικοποίηση των χαρακτηριστικών (feature normalization), εφαρμόζεται στα χαρακτηριστικά των δεδομένων, στην δική μας περίπτωση τις εικόνες και έχει ως αποτέλεσμα να φέρει τα δεδομένα στην ίδια κλίμακα με μικρές διακυμάνσεις μεταξύ τους. Ο χώρος των χρωμάτων των εικόνων έχει μεγάλο εύρος [0,255] αυτό δημιουργεί πρόβλημα στην εκπαίδευση των N.Δ. καθώς μπορεί να πάρουν ανεξέλεγκτες τιμές οι νευρώνες στα hidden layers και να μην συγκλίνει το Σ.Ν.Δ. Ο λόγος που βελτιώνει την σύγκλιση είναι επειδή οι τιμές στην είσοδο έχουν μέση τιμή μηδέν και διασπορά ένα, ως αποτέλεσμα οι νευρώνες στα ενδιάμεσα επίπεδα δεν μπαίνουν σε κορεσμό τόσο εύκολα και τόσο γρήγορα. Η εξίσωση 3.3 μας εξασφαλίζει τα χαρακτηριστικά να βρίσκονται στον χώρο [-1,1], έχοντας μέση τιμή μηδέν και διασπορά κοντά στο ένα. Για τον υπολογισμό της μέσης τιμής του συνόλου δεδομένων χρησιμοποιήσαμε ένα δείγμα από αυτό, από 500 δείγματα [39].

$$\hat{\mu} = \frac{\sum_{i=1}^N X_i}{N} \quad (3.2)$$

$$X = \frac{X - \hat{\mu}}{\max(X) - \min(X)} \quad (3.3)$$

3.3.3 Δυσαναλογία των Κλάσεων

Ένα πολύ συχνό πρόβλημα που υπάρχει στα περισσότερα σύνολα δεδομένων, είναι η δυσαναλογία των κλάσεων ή κατηγοριών. Η δυσαναλογία προκύπτει όταν σε ένα σύνολο δεδομένων υπάρχουν μεγάλες διαφορές μεταξύ του πλήθους των στοιχείων που ανήκουν σε ορισμένες κατηγορίες. Το πρόβλημα αυτό δεν το λύνουν τα νευρωνικά δίκτυα από μόνα τους, καθώς τείνουν να μάθουν καλύτερα πληροφορίες για τα στοιχεία που αποτελούν πλειοψηφία στο σύνολο δεδομένων μας, ενώ τα στοιχεία που αποτελούν μεινότητα φτάνουν σε σημείο μέχρι και να αγνοούνται. Μία λύση σε αυτό το πρόβλημα θα μπορούσε να είναι η υπερδειγματοληψία των κλάσεων που είναι μειονότητα, έτσι ώστε να δημιουργήσουμε ισόποσα σύνολα κλάσεων για να υπάρξει ισοστάθμιση. Όμως αυτή η επιλογή είναι ανέφτικη σε ένα σύνολο από εικόνες οπού θα πρέπει να δημιουργήσουμε καινούρια εικονοστοιχεία που να ανήκουν σε κάποια συγκεκριμένη κατηγορία.

Για την λύση αυτού του προβλήματος εφαρμόστηκε η Συνάρτηση Μέσης Συχνότητας Ισορροπίας (Median Frequency Balance) [5]. Με αυτή την συνάρτηση βρίσκουμε τους συντελεστές και τους εφαρμόζουμε στην συνάρτηση κόστους (εξίσωση 3.12).

Η ιδέα είναι να βρέθουν οι συντελεστές συχνότητας οι οποίοι προέρχονται από την συχνότητα εμφάνισης ενός εικονοστοιχείου που ανήκει σε μια κατηγορία, όταν ένα εικονοστοιχείο i ανήκει στην κατηγορία j (όπου είναι μεινότητα) και βρίσκεται κατά την διαδικασία της μάθησης να είναι στην κατηγορία k τότε επιβάλλεται μεγαλύτερη ποινή και διαδίδεται μεγαλύτερο σφάλμα προς τα πίσω. Αυτό συμβαίνει επειδή το νευρωνικό δίκτυο δεν θα δει πολλές φορές μια κλάση που είναι μεινότητα οφείλουμε να εισάγουμε μεγαλύτερη ποινή για να βοηθήσουμε στην εχμάθηση τους, διότι θα διαδοθεί μεγαλύτερο σφάλμα για διάδοση με την προς τα πίσω διάδοση.

Με την εξίσωση 3.4 βρίσκουμε την συχνότητα εμφάνισης των εικονοστοιχείων στο σύνολο δεδομένων και αφού ταξινομήσουμε τις τιμές συχνοτήτων παίρνουμε την μεσαία συχνότητα και την χρησιμοποιύμε σαν επίκεντρο (εξίσωση 3.5) τοποθετώντας την στον αριθμητή και στον παρονομαστή έχουμε τις συχνότητες των κλάσεων. Με αυτή την μέθοδο πετυχαίνουμε να έχουμε υψηλούς συντελεστές στης χαμηλής συχνότητας εμφάνισης εικονοστοιχείων. Τέλος, στα δεδομένα μας, υπάρχουν εικονοστοιχεία τα οποία δεν ανήκουν σε κάποια κατηγορία, για να μην μάθει το N.D από αυτά θέσαμε τον συντελεστή στο μηδέν έτσι ώστε να μην συνεισφέρουν στο σφάλμα κατά την διάρκεια της εκπαίδευσης.

$$freq(C_i) = \frac{C_i}{\sum_{i=1}^{Classes} C_i} \quad (3.4) \qquad \alpha_i = \frac{median(freq)}{freq(C_i)} \quad (3.5)$$

3.3.4 Επισκόπηση Αρχιτεκτονικής

Αρχικοποίηση Παραμέτρων Πυρήνα

Η αρχικοποίηση των παραμέτρων των φίλτρων αποτελεί ένα σημαντικό στάδιο στην εκπαίδευση των Νευρωνικών Δικτύων. Στόχος της αρχικοποίησης είναι η μέση τιμή της

είσοδος και εξόδου ενός επιπέδου να είναι κοντά στο μηδέν αλλά και η διασπορά τους να είναι κοντά στο ένα, καθώς αποτρέπει τους Νευρώνες να μπουν σε κορεσμό. Η εξίσωση 3.6 μας δείχνει την συνάρτηση αρχικοποίησης που χρησιμοποιήσαμε για την αρχικοποίηση των παραμέτρων. Στην ουσία πρόκειται για μία γκαουσιανή κατανομή 3.7 την οποία την ορίζουμε στον χώρο $[-L, L]$ από τον οποίο παίρνουμε δείγματα για να αρχικοποιήσουμε τα βάρη.

$$L = \sqrt{\frac{1}{Fan\ In}} \quad (3.6)$$

$$g(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \quad (3.7)$$

Συνάρτηση Ενεργοποίησης

Άλλη μια απαραίτητη συνάρτηση για τα Νευρωνικά Δίκτυα είναι η συνάρτηση ενεργοποίησης. Εφαρμόζεται στην έξοδο των επιπέδων των νευρωνικών δικτύων και είναι υπεύθυνη για την ανταλλαγή μυημάτων μεταξύ νευρώνων στα επίπεδα από νευρώνες. Τα βαθειά νευρωνικά δίκτυα έρχονται αντιμέτωπα με το πρόβλημα της εξαφάνισης των αποκλίσεων του σφάλματος κατά την διαδοσή τους προς τα πίσω. Για τον λόγο αυτό επινοήθηκαν συναρτήσεις που ονομάζονται Γραμμικοί Ανορθωτές (Linear Rectifiers). Οι γραμμικοί ανορθωτές συνήθως κάτω από το μηδέν έχουν μηδενική τιμή, όταν οι αποκλίσεις πέφτουν κάτω από το μηδέν τα βάρη δεν αλλάζουν, δηλαδή οι νευρώνες μένουν απενεργοποιημένοι σε μια τέτοια περίπτωση. Το θετικό σε αυτήν την περίπτωση είναι ότι εφ' όσον κάποιοι νευρώνες τείνουν σε αδράνεια, το νευρωνικό γίνεται ελαφρύτερο από την άποψη των υπολογισμών. Από την άλλη, το μεγάλο μειονέκτημα είναι ότι αν βρεθούν σε αυτή την κατάσταση μπορεί να μην ξανά ενεργοποιηθούν οι νευρώνες και δεν θα ανταποκριθούν σε αλλαγές από μικρά σφάλματα. Αυτό ονομάζεται Φαινόμενο Νεκρών Νευρώνων.

Μία λύση σε αυτό το πρόβλημα είναι η εισαγωγή μιας παραμετρικής συνάρτησης κάτω από το μηδέν, η οποία θα δίνει ένα μικρό ερέθισμα στους νευρώνες ώστε να αποφευχθεί αυτό το πρόβλημα. Για τον λόγο αυτό εισάγαμε στα νευρωνικά μας την Εκθετική Γραμμική Συνάρτηση (Scaled Exponential Linear Unit-SELU)εικόνα 3.3. Στην πραγματικότητα όπως απέδειξαν στο [25]η συγκεκριμένη συνάρτηση ενεργοποίησης σε συνδυασμό με την αρχικοποίηση (εξίσωση 3.6) όχι μόνο καταπολεμά αυτό το πρόβλημα αλλά εξαλείφει την εφαρμογή του αλγορίθμου Batch-Normalization [19] καθώς η κανονικοποίηση των εισόδων σε κάθε επίπεδο του νευρωνικού γίνεται μέσα σε αυτή την συνάρτηση, πετυχαίνοντας έτσι μείωση των παραμέτρων.

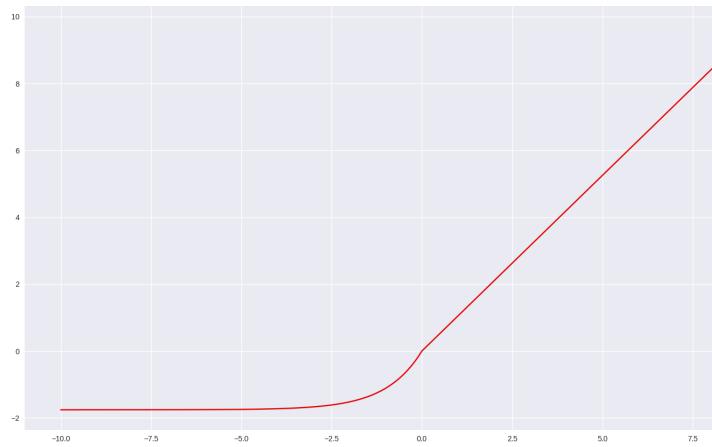


Figure 3.3: Μεταβαλόμενη Εκθετική Συνάρτηση Ενεργοποίησης (SELU) με τις προεπιλεγμένες παραμέτρους $\alpha = 1.6732$, $\lambda = 1.0507$ [1].

$$f(x) = \lambda \begin{cases} x & \text{if } x > 0 \\ \alpha e^x - \alpha & \text{if } x \leq 0 \end{cases} \quad (3.8)$$

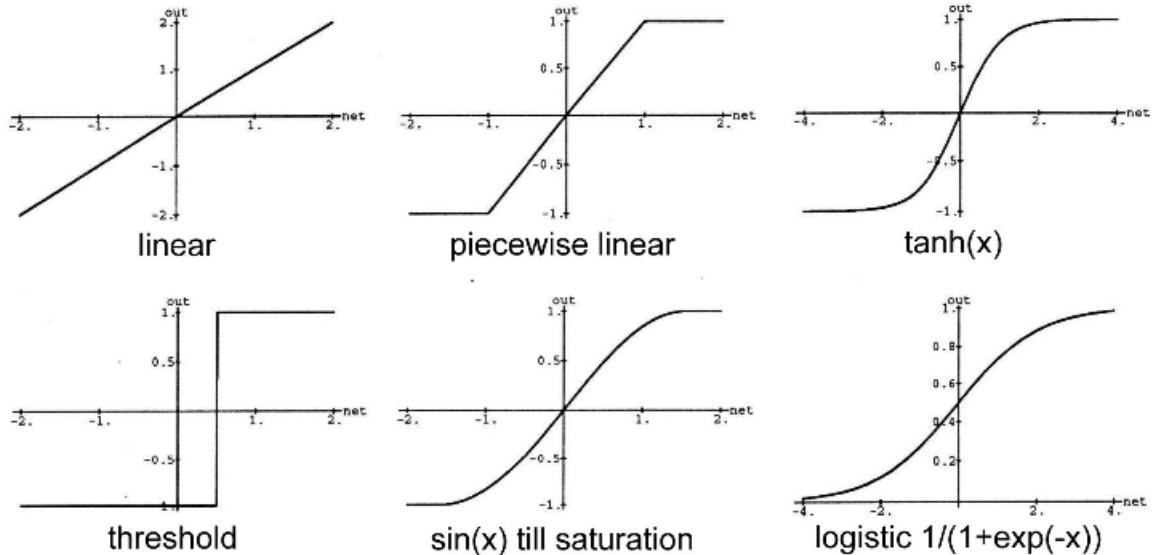


Figure 3.4: Διάφορες συναρτήσεις Ενεργοποίησης [17]

Αλγόριθμοι Βελτιστοποίησης

Ο αλγόριθμος βελτιστοποίησης αποτελεί έναν πολύ σημαντικό παράγοντα για την εκπαίδευση ενός Νευρωνικού Δικτύου, όπως είδαμε πριν τα Νευρωνικά Δίκτυα χρησιμοποιούν τον αλγόριθμο της Οπισθοδρόμησης για να υπολογίσουν τα παράγωγα του σφάλματος ώστε να αναβαθμίσουν τις κρυφές παραμέτρους με στόχο να μειώσουν το σφάλμα της σύναρτησης κόστους. Η ανάγκη για αναζήτηση αλγορίθμων βελτιστοποίησης προήλθε από δύο σημαντικούς παράγοντες. Πρώτον, λόγω των πολλών δεδομένων για επεξεργασία και των βαθειών νευρωνικών δικτύων που έχανε την διαδικασία της μάθησης αργή. Αυτοί οι λόγοι μας ώθησαν σε τεχνικές μείωσης του σφάλματος από κομμάτια του συνόλου δεδομένων και δεύτερον για την επιτάχυνση της σύγκλισης του Νευρωνικού

δικτύου προφανώς. Ο πιο συνηθισμένος και βασικός αλγόριθμος βελτιστοποίησης είναι ο Στοχαστικός Αλγόριθμος Απότομης καθόδου (SGD) [46], ο οποίος υπολογίζει απλά την αποκλίση των παραμέτρων ως προς της συνάρτηση κόστους πάνω σε ένα μικρό σύνολο δειγμάτων από τα δεδομένα. Πλέον υπάρχουν πιο προχωρημένοι αλγόριθμοι βελτιστοποίησης. Η επιλόγη του αλγόριθμου βελτιστοποίησης γίνεται ανάλογα με την αρχιτεκτονική του Νευρωνικου Δικτύου. Η εξίσωση 3.9 μας δείχνει την εξίσωση όπου α είναι ο ρυθμός μάθησης και ο υπολογισμός της απόκλισης γίνεται πάνω σε ένα σύνολο ζευγών (x^i, y^j) .

$$\vartheta = \vartheta - \alpha \nabla_{\vartheta} \mathcal{J}(\vartheta; x^i, y^j) \quad (3.9)$$

Αν η συνάρτηση κόστους έχει την μορφή μίας χαράδρας που οδηγεί προς το βέλτιστο κόστος και έχει στα πλάγια υψηλά τοιχώματα (εικόνα ??) τότε ο κλασικός αλγόριθμος SGD τείνει να ταλαντεύεται στο φαράγγι επειδή η αρνητική απόκλιση τείνει προς τις απότομες πλευρές κάθε φορά αντί να πηγαίνει προς το βέλτιστο. Αυτό το φαινόμενο συμβαίνει σχεδόν πάντα και δημιουργεί πρόβλημα διότι μας κάνει την σύγκλιση πολύ αργή.

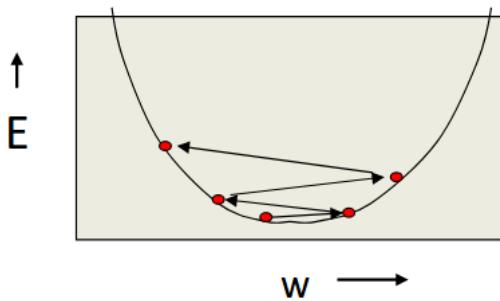


Figure 3.5: Ταλάντωση του SGD κατά την εκτιμήση των βαρών προσπαθώντας να φτάσει στο βέλτιστο σημείο [16].

Μία λύση σε αυτό το πρόβλημα είναι η χρήση της ορμής, η οποία βοηθάει να φτάσουμε στο βέλτιστο πιο γρήγορα. Στην εξίσωση 3.10 υ είναι το διάνυσμα ταχύτητας το οποίο είναι φυσικά ίδιων διαστάσεων με το διάνυσμα των παραμέτρων θ . Πέρα από την παράμετρο α που είδαμε και προηγουμένως η οποία είναι ο ρυθμός μάθησης, παρατηρούμε και την παράμετρο $\gamma \in [0, 1]$ η οποία ορίζει το ποσοστό συνεισφόρας των προηγούμενων αποκλίσεων στην παρούσα ανανέωση των παραμέτρων. Συνήθως αυτή η ποσότητα ορίζεται στο 0.9 .

$$\begin{aligned} v &= \gamma v + \alpha \nabla_{\vartheta} \mathcal{J}(\vartheta; x^i, y^j) \\ \vartheta &= \vartheta - v \end{aligned} \quad (3.10)$$

Στην εργασία μας έγινε χρήση τόσο του αλγορίθμου της Στοχαστικής Απότομης Καθόδου, αλλά και του αλγόριθμου Adam [24] ενός πιο αποδοτικού αλγορίθμου σε θέματα στοχαστικής βελτιστοποίησης καθώς χρησιμοποιεί πρώτης τάξης παράγωγα. Ο αλγόριθμος υπολογίζει τις παραμέτρους του ρυθμού μάθησης για διάφορες παραμέτρους από εκτιμήσεις των ορμών πρώτης και δεύτερης τάξης των κλίσεων, όπως φαίνεται αναλυτικά στο . Συγκεκριμένα ο αλγόριθμος Adam είναι μια εξέλιξη του RMSProp [41].

Algorithm Αλγόριθμος Adam, Αναλυτική περιγραφή των βημάτων, όλες οι πράξεις των διανυσμάτων είναι ανά στοιχείο. Η g_t^2 δείχνει τον ανά στοιχείο πολλαπλασιασμό $g_t \odot g_t$. Οι προτεινόμενες τιμές των παραμέτρων είναι: $\alpha = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$ και $\epsilon = 10^{-8}$.

Require: : α : Ρυθμός Βήματος
Require: $\beta_1, \beta_2 \in [0, 1]$: Εκθετικοί ρυθμοί καθόδου για τις εκτιμησεις των ορμών
Require: $f(\theta)$: Στοχαστική συνάρτηση κόστους
Require: θ_0 : Αρχικοποίηση διανύσματος παραμέτρων

- 1: $m_0 \leftarrow 0$ (Αρχικοποίηση 1ης τάξης διανύσματος)
- 2: $u_0 \leftarrow 0$ Αρχικοποίηση 2ης τάξης διανύσματος
- 3: $t \leftarrow 0$ (Αρχικοποίηση βήματος χρόνου)
- 4: **while** θ_t not converged **do**
- 5: $t = t + 1$
- 6:
- 7: $g_t \leftarrow \nabla_{\theta} f_t(\theta_{t-1})$ {Αποκλίσεις ως προς την συναρτηση f την στιγμή t }
- 8:
- 9: $m_t \leftarrow \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t$ {Ενημέρωση της μεροληπτικής εκτίμησης 1ης τάξης}
- 10:
- 11: $v_t \leftarrow \beta_2 \cdot m_{t-1} + (1 - \beta_2) \cdot g_t^2$ {Ενημέρωση της μεροληπτικής εκτίμησης 2ης τάξης}
- 12:
- 13: $\hat{m}_t \leftarrow m_t / (1 - \beta_1^t)$ {Διόρθωση της μεροληπτικής εκτίμησης 1ης τάξης}
- 14:
- 15: $\hat{v}_t \leftarrow v_t / (1 - \beta_2^t)$ {Διόρθωση της μεροληπτικής εκτίμησης 2ης τάξης}
- 16:
- 17: $\theta_t \leftarrow \theta_{t-1} - \alpha \cdot \hat{m}_t / (\sqrt{\hat{v}_t} + \epsilon)$ {Ενημέρωση παραμέτρων}
- 18: **end while**

return θ_t {Αποτελέσματα παραμέτρων}

Συνάρτηση Κόστους

Σύνηθως, σε προβλήματα πολλαπλής ταξινόμησης στοιχειών, όπως στην Σημασιολογική Κατάτμηση, θέλουμε τα Νευρωνικά Δίκτυα να δέχονται στην είσοδο ένα διάνυσμα και να μας δίνουν στην έξοδο ένα διάνυσμα με την πιθανότητα των εικονοστοιχείων να ανήκουν σε μια από τις L κατηγορίες. Για να το επιτύχουμε αυτό τοποθετούμε ένα επίπεδο *Softmax* L εξόδων, στην έξοδο του Νευρωνικού Δικτύου. Η $softmax(z)_i$ περιγράφει την i^{th} πιθανότητα ενός εικονοστοιχείου να ανήκει σε μια από τις L κατηγορίες. Η $softmax$ μετατρέπει το διάνυσμα L διαστάσεων σε μια πιθανοτική κατανομή όπου όλες οι τιμές αθροίζονται στο ένα (εξίσωση 3.11).

$$softmax(z)_i = \frac{e^{z_i}}{\sum_{l=1}^L e^{z_l}} \quad (3.11)$$

$$Loss(P, Q) = -\frac{1}{N} \sum_{x \in N} \sum_{k \in L} P(x, k) \times log(Q(x, k)) \times \alpha_{coefficients} \quad (3.12)$$

3.3.5 Στάδιο Κωδικοποίησης

Σκοπός της μονάδας Κωδικοποίησης είναι η εξαγωγή χαρακτηριστικών από την έγχρωμη εικόνα, δηλαδή η δημιουργία μιας αναπαράστασης πολυδιάστατων χαρακτηριστικών από τα εικονοστοιχεία της εικόνας σε μια συμπιεσμένη μορφή ώστέ να γίνεται εφικτή η εκπαίδευση του συστήματος. Η μονάδα κωδικοποίησης αποτελείται από 4 τμήματα και συγκεκριμένα από ομάδες συνελικτικών επιπέδων και μονάδων συγκέντρωσης. Η εικόνα 3.6 μας δίνει μια διαίσθηση του κάθε τμήματος το οποίο αποτελείται από 2 επίπεδα συνέλιξης ακολουθύμενα από ένα επίπεδο συγκέντρωσης μέγιστων τιμών ανά περιοχή (Max-Pooling).

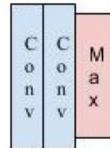


Figure 3.6: Τμήμα συνέλιξης: 2 επίπεδα συνέλιξης και ένα Max-pooling επίπεδο.

Πιο συγκεκριμένα, στα επίπεδα συνέλιξης γεμίζουμε περιφερειακά την χαρτογράφηση των χαρακτηριστικών με μηδενικά ανάλογα με το μέγεθος του πυρήνα για να μπορέσουμε να κρατήσουμε το μέγεθος τους αναλοίωτο κρατώντας την θέση των χαρακτηριστικών αλλά και επειδή χρειαζόμαστε την πληροφορία από τις γωνίες των χαρακτηριστικών. Η εικόνα 3.7 μας δείχνει ένα παράδειγμα της διαδικασίας, ενώ η εξίσωση 3.13 μας δείχνει ότι για να πετύχουμε μέγεθος εισόδου (input) ίδιο με το μέγεθος εξόδου, πρέπει να ισχύει η παρακάτω εξίσωση, για οποιοδήποτε μέγεθος εισόδου $input$ και για μονό αριθμό στοιχείων πυρήνα k όπου ($k = 2n + 1, n \in \mathbb{N}$), s είναι το βήμα ολίσθησης το οποίο είναι 1 και p είναι το γέμισμα των μηδενικών περιφερειακά της εισόδου όπου $p = \lfloor k/2 \rfloor = n$.

$$\begin{aligned} output &= input + 2\lfloor k/2 \rfloor - (k - 1) \\ &= input + 2n - 2n \\ &= input \end{aligned} \tag{3.13}$$

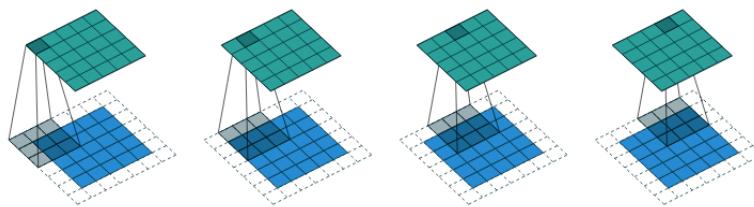


Figure 3.7: Εφαρμογή ενός πυρήνα 3×3 με ολίσθηση μονού βηματισμού σε επίπεδο εισόδου μεγέθους 5×5 με γέμισμα μηδενικών περιφερειακά της εισόδου.

Το στάδιο κωδικοποίησης αποτελείται από 4 τμήματα (3.6) όπου δέχεται σαν είσοδο την εικόνα μεγέθους $512 \times 512 \times 3$ και παράγει στην έξοδο μια χαρτογράφηση χαρακτηριστικών $32 \times 32 \times 256$, όπου η τρίτη διάσταση είναι ο αριθμός των φίλτρων. Πιο αναλυτικά στο πρώτο τμήμα έχουμε την συνέλιξη της εικόνας με μια σειρά από 32 φίλτρα και ακόμα ένα ίδιο επίπεδο πριν καταλήξουμε να εφαρμόσουμε το επίπεδο μέγιστης συγκέντρωσης ή (Max-Pooling). Το επίπεδο μέγιστης συγκέντρωσης είναι μια ανορθόδοξη

τεχνική στην οποία επιδιώκουμε να μειώσουμε τον αριθμό των παραμέτρων σταδιακά, καθώς όσο προχωράμε στα επόμενα τμήματα αυξάνεται ο αριθμός του βάθους των επιπέδων (δηλαδή των φίλτρων) και επομένως και ο αριθμός των παραμέτρων. Ένας άλλος ο λόγος είναι η προσπάθεια της εξάλειψης της υπερμάθησης ως αποτέλεσμα της μείωσης των παραμέτρων. Στην δική μας περίπτωση συγκεντρώσαμε από κάθε περιοχή 2×2 την μέγιστη τιμή των χαρακτηριστικών. Συγκεκριμένα ολισθαίνουμε ένα παράθυρο μεγέθους 2×2 στα χαρακτηριστικά και πάρουμε την μέγιστη τιμή. Διαισθητικά, αυτό σημαίνει ότι κρατήσαμε την τιμή που υπάρχει μεγαλύτερο ερέθισμα στον εκάστοτε νευρώνα (εικόνα 3.9). Επομένως ο χάρτης των χαρακτηριστικών χώρου μετά από κάθε τμήμα μειώνεται κατά το ήμισυ, οι εξισώσεις στη 3.14 μας δείχνουν τον υπολογισμό των διαστάσεων εξόδου μετά την εφαρμογή του επιπέδου μέγιστης συγκέντρωσης. Η μεταβλητή P ορίζει τυχόν γέμισμα στο στάδιο της συγκέντρωσης το οποίο στην δική μας περίπτωση είναι μηδέν και με S ορίζεται το άλμα ολίσθησης του πυρήνα συγκέτρωσης.

$$\begin{aligned} Image &= H \times W \times D \\ Height &= (Height - Poolsize + 2 \times P)/S + 1 \\ Width &= (Width - Poolsize + 2 \times P)/S + 1 \\ Output &= Height \times Width \times D \end{aligned} \tag{3.14}$$

Ο αλγόριθμος μέγιστης συγκέντρωσης εφαρμόζεται ανεξάρτητα σε κάθε φίλτρο εισόδου των χαρακτηριστικών. Δηλαδή δεν επηρεάζει το μέγεθος των φίλτρων καθώς εφαρμόζεται μόνο στις χωρικές διαστάσεις των χαρακτηριστικών.

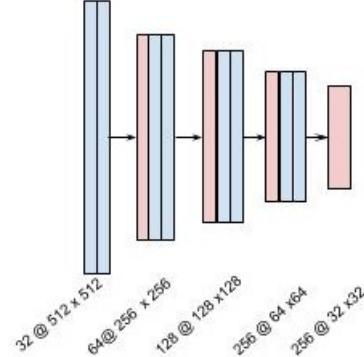


Figure 3.8: Στάδιο κωδικοποίησης των ΣΝΔ επίπεδο.

Η εικόνα 3.9 μας δείχνει ένα παράδειγμα εφαρμογής ενός πυρήνα 3×3 πάνω σε ένα επιπέδο χαρακτηριστικών 5×5 εφαρμόζοντας την μέθοδο της μέγιστης συγκέντρωσης. Όπως βλέπουμε συγκεντρώνουμε το μέγιστο στοιχείο από τον 3×3 πυρήνα που ολισθαίνει σε όλο το επίπεδο, ορίζοντιως και καθέτως αντίστοιχα.

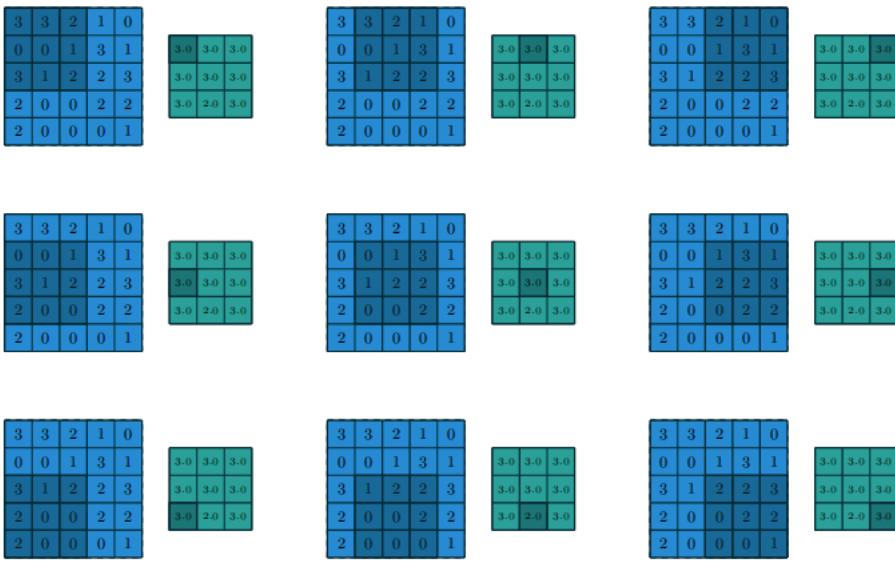


Figure 3.9: Παράδειγμα της μεθόδου της μέγιστης συγκέντρωσης, εφαρμόζοντας ένα παράθυρο 3×3 σε μια είσοδο 5×5 .

3.3.6 Μονάδα Παράλληλης Επεξεργασίας Χαρακτηριστικών

Η παράλληλη μονάδα επεξεργασίας χαρακτηριστικών αποτελείται από 5 διαφορετικά τμήματα τα οποία δέχονται ως είσοδο τον χάρτη με τα κωδικοποιημένα χαρακτηριστικά από το στάδιο της κωδικοποίησης. Η εξίσωση 3.15 μας δείχνει την συνέλιξη σε ένα επίπεδο σήμα είσαγωντας την διαστολή που υποδεικνύεται με r . Η συγκεκριμένη συνάρτηση υπάγεται στην θεωρία ως Διεσταλμένη Συνέλιξη (Dilated Convolution) ενώ η εικόνα 3.11 μας δίνει μια διαίσθηση γύρω από αυτή την τεχνική.

$$g[i, j] = \sum_k \sum_k f[i + r \cdot k, j + r \cdot k] h[k, k] \quad (3.15)$$

Πιο συγκεκριμένα, κάθε παρακλάδι της μονάδας επεξεργασίας διαφέρει στην διαστολή των στοιχείων του πυρήνα που αλληλεπιδρούν με την είσοδο (εικόνα 3.10). Σκοπός αυτού του τμήματος είναι η μάθηση χαρακτηριστικών από διαφορετικά πεδία όρασης. Κάθε παρακλάδι διαιθέτει έναν πυρήνα με διαφορετική διαστολή στο πρώτο επίπεδο. Ο πυρήνας σε όλους τους κλάδους έχει μέγεθος 3×3 , εκτός από το τελευταίο επίπεδο πριν την ένωση των χαρακτηριστικών όπου η πυρήνας έχει μέγεθος 1×1 . Στην εικόνα 3.10 βλέπουμε σε κάθε επίπεδο το μέγεθος κάθε επιπέδου και τον αριθμό του βάθους των φίλτρων τα οποία είναι 256, 128 και 128 αντίστοιχα. Ο λόγος που μειώνουμε τις διαστάσεις είναι για την μείωση των παραμέτρων κατά την εκπαίδευση. Επίσης, για να κρατήσουμε σταθερό το μέγεθος των χαρακτηριστικών και για να μην υπάρχει περαιτέρω αλλοίωση της πληροφορίας γεμίζουμε περιφερειακά με μηδενικά την είσοδο πριν την διαδικασία της συνέλιξης. Στο στάδιο της ένωσης πραγματοποιείται η πράξη της πρόσθεσης όλως των χαρακτηριστικών που καταλήγουν από κάθε κλάδο αντίστοιχα.

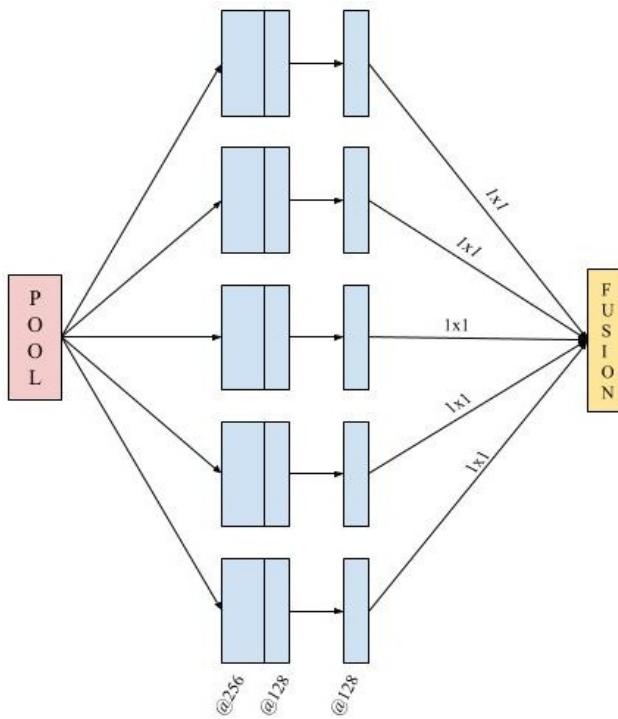


Figure 3.10: Η παράλληλη μονάδα επεξεργασίας με τα 5 ξεχωριστά μονοπάτια. Το κάθε μονοπάτι έχει στο πρώτο επίπεδο συνέλιξης μια διαστολή: 3×3 , 6×6 , 9×9 , 12×12 και 1×1 αντίστοιχα. Επίσης, βλέπουμε και τον αριθμό των φίλτρων του κάθε επιπέδου συνέλιξης.

Παρακάτω βλέπουμε ένα παράδειγμα για την διεσταλμένη συνέλιξη όπου εφαρμόζουμε έναν πυρήνα 3×3 πάνω σε ένα επίπεδο εισόδου 7×7 με ολίσθηση του πυρήνα ίσο με 1 και χωρίς γέμισμα περιφερειακά της εισόδου με μηδενικά. Στην δική μας περίπτωση υπάρχει γέμισμα του χάρτη χαρακτηριστικών περιφερειακά με μηδενικά καθώς θέλουμε να κρατήσουμε το μέγεθος αναλοίωτο αλλά και να προσπαθήσουμε να κρατήσουμε την θέση της πληροφορίας όσο περισσότερο γίνεται.

Η διεσταλμένη συνέλιξη γεμίζει τον πυρήνα του φίλτρου με μηδενικά ανάμεσα στα στοιχεία του πυρήνα. Για την ακρίβεια, για έναν ρυθμό διαστολής d εισάγουμε $d - 1$ μηδενικά ανάμεσα στα στοιχεία του πυρήνα και προφανώς για $d = 1$ αναφερόμαστε σε μια τυπική συνέλιξη. Η συνέλιξη με διαστολή συνήθως χρησιμοποείται για την αύξηση του δεκτικού πεδίου ενός νευρώνα χωρίς να χρειαστεί να αυξηθεί το μέγεθος του πυρήνα. Μία σημαντική ιδιότητα αυτής της τεχνικής είναι ο αριθμός των παραμέτρων ο οποίος αυξάνεται γραμμικά ενώ ο αριθμός του δεκτικού πεδίου αυξάνεται εκθετικά.

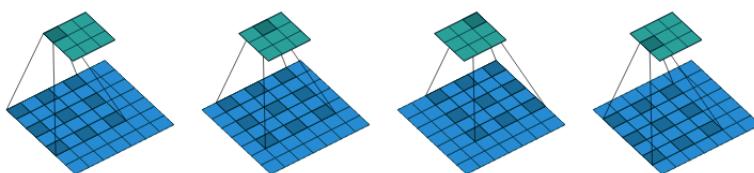


Figure 3.11: Συνέλιξη ενός πυρήνα μεγέθους 3×3 πάνω σε ένα επίπεδο εισόδου μεγέθους 7×7 και με διαστολή μεγέθους 2. Τα μπλέ σκούρα στοιχεία δείχνουν την συμμετοχή για τον υπολογισμό της τιμής του στοιχείου (πράσινο σκούρο) [15].

3.3.7 Στάδια Αποκωδικοποίησης

Επισκόπηση

Παρακάτω θα εξηγήσουμε τις 2 παραλαγές των μονάδων αποκωδικοποίησης που υλοποιήσαμε για να πειραματίστουμε με αυτές και να συγχρίνουμε τα αποτελέσματα τους στο πρόβλημα της Σημασιολογικής Κατάτμησης. Η κύρια διαφορά μεταξύ των 2 μονάδων είναι ο τρόπος που γίνεται η υπερδειγματοληψία. Η πρώτη μονάδα χρησιμοποιεί την μέθοδο της αποσυνέλιξης με εισαγωγή ενός βήματος για την επίτευξη της υπερδειγματοληψίας, ενώ στην δεύτερη μονάδα υλοποιήθηκε ένα επίπεδο διγραμμικής παρεμβολής που λειτουργεί με τον τρόπο που εξηγήσαμε στο τμήμα 3.3.1.

Μονάδα Αποσυνέλιξης Με Άλμα Ολίσθησης

Η συγκεκριμένη μονάδα αποκωδικοποίησης αποσκοπεί στην ανακατασκεύη των χαρακτηριστικών από τον χάρτη χαρακτηριστικών πίσω στην μορφή της εικόνας. Η εικόνα 3.12 μας δείχνει την αρχιτεκτονική της μονάδας αποκωδικοποίησης.

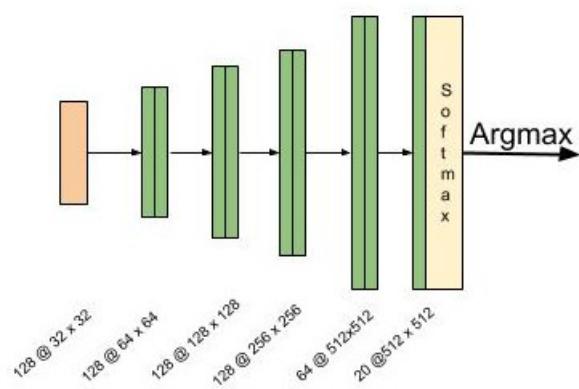


Figure 3.12: Στάδιο αποκωδικοποίησης του ΣΝΔ με χρήση επιπέδων αποσυνέλιξης.

Για να μπορέσουμε να εξηγήσουμε καλύτερα την μονάδα αποκωδικοποίησης θα πρέπει πρώτα να μιλήσουμε για την διαδικασία της αποσυνέλιξης ή ανάστροφης συνέλιξης όπως την βρίσκουμε στην βιβλιογραφία. Η ιδέα και η ανάγκη της ανάστροφης συνέλιξης προκύπτει από την επιθυμία να χρησιμοποιηθεί ένας μετασχηματισμός που να μας οδηγεί από τον χάρτη των χαρακτηριστικών, δηλαδή στον μετασχηματισμό από ένα σχήμα κάποιου αντικειμένου πίσω στον ανασχηματισμό του σε σχέση με την εικόνα εισόδου. Με λίγα λόγια γίνεται μια ανακατεσκευή της εικόνας από τα μεγάλων διαστάσεων χαρακτηριστικά.

Η τεχνική της ανάστροφης συνέλιξης μας οδηγεί από έναν χάρτη χαρακτηριστικών μικρού μεγέθους σε έναν χάρτη μεγαλύτερου μεγέθους ενώ συγχρατεί τα μοτίβα διασύνδεσης μεταξύ των νευρώνων. Η ανάστροφη συνέλιξη δουλεύει εναλλάσσοντας το μπροστινό πέρασμα με το πέρασμα της οπισθοδόμησης της συνέλιξης. Με λίγα λόγια, η κανονική συνέλιξη με την ανάστροφη συνέλιξη είναι ο τρόπος με τον οποίο υπολογίζονται τα προς τα εμπρός και προς τα πίσω περάσματα (feed-forward and backward passes).

Για παράδειγμα μπορεί ένας πυρήνας w να ορίζει μια συνέλιξη όπου τα περάσματα (εμπρός-πίσω) να υπολογίζονται από έναν πίνακα C και C^T αντίστοιχα, αλλά επίσης αν αναστρέψουμε τους πίνακες ορίζουμε την ανάστροφη συνέλιξη ορίζοντας τους πίνακες ως C^T και $(C^T)^T = C$ για τα εμπρός και πίσω περάσματα αντίστοιχα. Στο παράτημα B 6.2 περιγράφεται αναλυτικά η διαδικασία με τον πίνακα C . Τέλος, το τελευταίο επίπεδο πριν την εφαρμογή του επιπέδου softmax έχουμε ένα επίπεδο με αριθμό βάθους χαρτών ίσο με 20, όσο είναι και οι κατηγορίες αντικειμένων. Κάθε επίπεδο του βάθους από τα 20 επίπεδα αποτελεί ένα heatmap της κάθε χλάσης ως προς τις υπόλοιπες. Εφαρμόζοντας την softmax μετασχηματίζεται η έξοδος σε μια κατανομή πιθανοτήτων.

Διγραμμική Μονάδα Αποκωδικοποίησης

Η διγραμμική μονάδα αποκωδικοποίησης όπως βλέπουμε στην εικόνα 3.13 έχει πανομοιότυπη αρχιτεκτονική με την προηγούμενη μονάδα 3.3.7, με την διαφορά στην μέθοδο που επιτυγχάνεται η υπερδειγματοληψία. Για την υπερδειγματοληψία του χάρτη χαρακτηριστικών χρησιμοποείται η μέθοδος της διγραμμικής παρεμβολής. Κατά την διγραμμική παρεμβολή, ένα στοιχείο δημιουργείται από μια μέση τιμή βαρύτητας από τέσσερα γειτονικά στοιχεία από τον χάρτη χαρακτηριστικών εισόδου. Το επίπεδο αυτό δεν διαιθέτει παραμέτρους μάθησης.

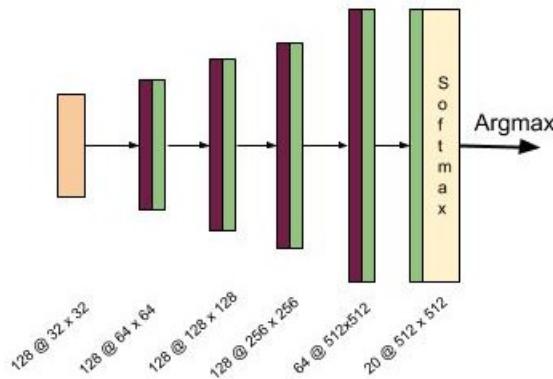
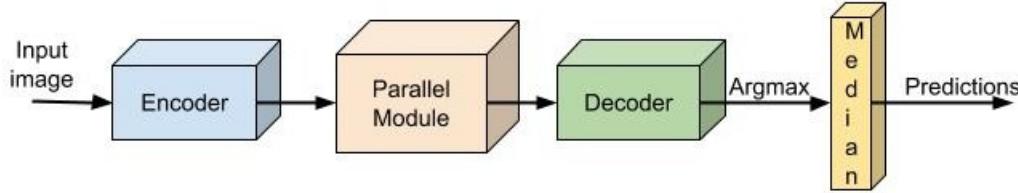


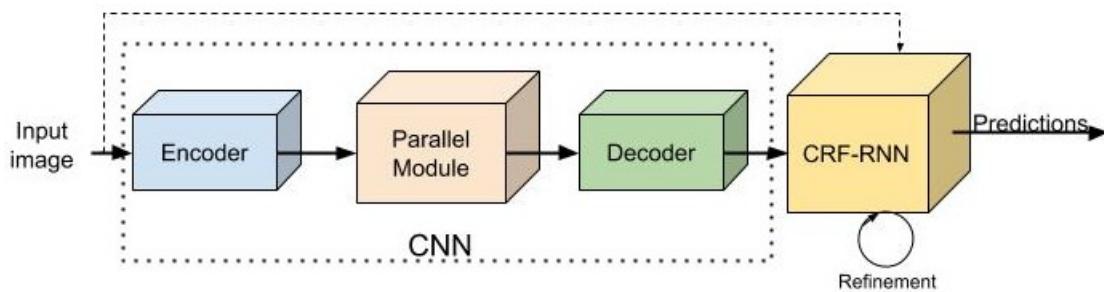
Figure 3.13: Στάδιο αποκωδικοποίησης του ΣΝΔ με χρήση επιπέδων διγραμμικής παρεμβολής για την υπερδειγματοληψία των χαρακτηριστικών. Τα μωβ επίπεδα υποδεικνύουν το επίπεδο της διγραμμικής παρεμβολής.

3.3.8 Ολοκληρωμένες αρχιτεκτονικές

Η εικόνα 3.14 μας δείχνει τις μονάδες των αρχιτεκτονικών που περιγράψαμε προηγουμένως ολοκληρωμένες, μαζί με τις μονάδες μετά-επεξεργασίας που θα αναλύσουμε στο επόμενο κεφάλαιο.



(a) Ολοκληρωμένη αρχιτεκτονική με την μονάδα μετα-επεξεργασίας Μεσαίου Φίλτρου.



(b) Ολοκληρωμένη αρχιτεκτονική με την μονάδα μετα-επεξεργασίας ΤΥΣΠ-ΕΝΔ.

Figure 3.14: Ολοκληρωμένες Αρχιτεκτονικές.

Chapter 4

Μονάδες Μετα-Επεξεργασίας

4.1 Επισκόπηση

Το πρόβλημα με τα Συνελικτικά Δίκτυα (CNNs) είναι η προσαρμογή σε συνελικτικά φίλτρα με μεγάλα οπτικά πεδία όπως έχουμε και στο δίκη μας εργασία με τα διαφορετικά οπτικά πεδία που εφαρμόζονται στην παράλληλη μονάδα επεξεργασίας. Συνεπώς παράγουν χονδροειδείς εξόδους όταν αναδιαρθρώνονται για να παράγουν προβλέψεις σε επίπεδο εικονοστοιχείων και καταλήγουμε να έχουμε πιο γενικά όρια από ότι θα περιμέναμε. Επίσης, τα (CNNs) δεν έχουν περιορισμούς ομαλότητας. Για να μπορέσουμε να διυθετήσουμε αυτό το πρόβλημα υιοθετήσθηκαν δύο δαφορετικές προσεγγίσεις που θα εξηγήσουμε λεπτομερώς στα επόμενα τμήματα. Πρώτον ο αλγόριθμος Μέσου Φίλτρου που προσαρμόστηκε ως μονάδα μετα-επεξεργασίας και ο αλγόριθμος των Υποθετικών Τυχαίων Πεδίων ως Επαναλαμβανόμενα Νευρωνικά Δίκτυα (CRFs as RNN).

4.2 Μεσαίο Φίλτρο

Ο αλγόριθμος του Μεσαίου Φίλτρου είναι μια μη γραμμική τεχνική ψηφιακού φιλτραρίσματος, που συχνά χρησιμοποιείται για την απομάκρυνση του θορύβου από μια εικόνα ή ένα σήμα. Μια τέτοια μείωση θορύβου είναι ένα τυπικό στάδιο προεπεξεργασίας για τη βελτίωση των αποτελεσμάτων της μεταγενέστερης επεξεργασίας (για παράδειγμα, ανίχνευση ακμής σε μια εικόνα). Το μεσαίο φιλτράρισμα χρησιμοποιείται ευρέως στη ψηφιακή επεξεργασία εικόνων επειδή, υπό ορισμένες συνθήκες, διατηρεί τις άκρες ενώ απομακρύνει τον θόρυβο, έχοντας επίσης εφαρμογές στην επεξεργασία σήματος.

Ο λόγος που χρησιμοποιήθηκε στα πειράματα μας είναι για να επιτύχουμε μια εξομάλυνση στις στην έξοδο του συστήματος, δηλαδή στις προβλέψεις του συστήματος για τα εικονοστοιχεία. Για παράδειγμα, αν μια περιοχή της εικόνας απεικονίζει εναν δρόμο, μπορεί να υπάρχουν ορισμένα εικονοστροιχεία που να έχουν προβλεφθεί ως πεζόδρομος, τότε με αυτό το φίλτρο θα πετύχουμε την μείωση των λανθασμένων εικονοστοιχείων της περιοχής της εικόνας. Η εξίσωση 4.1 μας δείχνει την γενική εξίσωση της εφαρμογής ενός φίλτρου στην εικόνα, όπου η τιμή του εικονοστοιχείου ($g(i, j)$) εξαρτάται από ένα σταθμισμένο άθροισμα των εικονοστοιχείων εισόδου ($f(i + k, j + l)$) και $h(k, l)$ ονομάζεται ο πυρήνας που περιέχει τους συντελεστές του φίλτρου [40].

$$g(i, j) = \sum_{k,l} f(i+k, j+l)h(k, l) \quad (4.1)$$

Ο αλγόριθμος παρακάτω μας δείχνει βήμα-βήμα τον αλγόριθμο του Μεσαίου Φίλτρου:

Algorithm Αλγόριθμος Μεσαίου Φίλτρου Median Filter

```

Require: Output image[ $W \times H$ ]
Require: Input image[ $W \times H$ ]
Require: Window[ $K \times K$ ]
Require: edgeX  $\leftarrow \text{round}(K/2)$ 
Require: edgeY  $\leftarrow \text{round}(K/2)$ 
    for  $x$  from edgeX to  $W - \text{edgeX}$  do
        2:   for  $y$  from edgeY to  $H - \text{edgeY}$  do
             $i = 0$ 
            4:     for  $Fx$  from 0 to  $K$  do
                5:         for  $Fy$  from 0 to  $K$  do
                    6:              $Window[i] = \text{Input image}[x + Fx - \text{edgeX}][y + Fy - \text{edgeY}]$ 
                     $i \leftarrow i + 1$ 
                7:         end for
                8:     end for
                9:     sort values in Window
                10:     $\text{Output Image}[x][y] \leftarrow Window[K * K / 2]$ 
            11:     end for
            12:     end for
        return {Output Image}
    
```

Ένα μειονέκτημα του αλγορίθμου είναι ότι για κάθε υπολογισμό ενός εικονοστοιχείου πρέπει να ταξινομήσουμε τα στοιχεία για να πάρουμε την ενδιάμεση τιμή. Επομένως, προσθέτει υπολογιστικό κόστος καθώς προσθέτει επιπλέον $O(N^2)$ πράξεις.

4.3 Conditional Random Fields as Recurrent Neural Network

Τα Τυχαία υπό Συνθήκη Πεδία (CRF) παρουσιάστηκαν ως μονάδα μετά-επεξεργασίας για την βελτίωση των αποτελεσμάτων. Χρησιμοποιούνται συνήθως σε προβλήματα σημασιολογικής κατάτμησης, ενώ ανήκουν στη κατηγορία των στατιστικών μοντέλων γράφων. Στην πραγματικότητα, πριν από την έλευση των νευρωνικών δικτύων και συγκεκριμένα των Συνελικτικών (CNN), τα CRF αποτελούσαν την καλύτερη δυνατή προσέγγιση σε θέματα σημασιολογικής κατάτμησης, ενώ πλέον χρησιμοποιούνται για βελτίωση αποτελεσμάτων καθώς τείνουν να βελτιώνουν την διαγράμμιση των ορίων των αντικειμένων στις εικόνες. Ένα άλλο γεγονός είναι ότι τα CRF είναι ένα Τυχαίο Πεδίο Markov (MRF) όπου οι συντελεστές του καθορίζονται από κάποιες συνθήκες στα δεδομένα.

4.3.1 Επισκόπηση Αλγορίθμου

Στην πραγματικότητα υπάρχουν πολλές παραλλαγές τέτοιων μοντέλων. Έμεις θα ασχοληθούμε με τα πυκνά μοντέλα CRF και στην προκειμένη περίπτωση μια υλοποίηση που είναι βασισμένη σε επαναλαμβανόμενα νευρωνικά δίκτυα (CRF as RNN). Θα δώσουμε μία συνοπτική περιγραφή του αλγορίθμου πριν προχωρήσουμε στην ανάλυση του. Τα CRF όπως αναφέραμε, χρησιμοποιούνται για πρόβλεψη των εικονοστοιχείων, μοντελοποιούν τα εικονοστοιχεία σαν τυχαίες κατανομές που δημιουργούν ένα MRF όταν υπόκεινται σε μια μέγαλη κλίμακα παρατηρήσεων. Στην προκειμένη περίπτωση η μεγάλη κλίμακα παρατηρήσεων είναι η εικόνα.

Ας υποθέσουμε ότι X_i είναι μια τυχαία μεταβλητή που σχετίζεται με το εικονοστοιχείο i το οποίο μπορεί να πάρει οποιαδήποτε τιμή από ένα σύνολο τιμών που ανήκει στο \mathcal{L} . Αν υποθέσουμε ότι \mathbf{X} είναι το διάνυσμα των τυχαίων μεταβλητών X_1, X_2, \dots, X_N όπου N ο αριθμός των εικονοστοιχείων της εικόνας.

Παίρνοντας σαν δεδομένο τον γράφο $G = (V, E)$ όπου $V = X_1, X_2, \dots, X_N$, μία παρατήρηση της εικόνας \mathbf{I} , το ζευγάρι (\mathbf{I}, \mathbf{X}) μπορεί να μοντελοποιηθεί ως μια κατανομή Gibbs της μορφής $P(\mathbf{X} = \mathbf{x}|\mathbf{I}) = \frac{1}{Z(\mathbf{I})} \exp(-E((\mathbf{x}|\mathbf{I}))$. Η συνάρτηση $E(x)$ είναι η ενέργεια των παρατηρήσεων $x \in \mathcal{L}^N$, και η $Z(\mathbf{I})$ είναι η συνάρτηση διαμέρισης (partition function) [28]. Στα πλήρως συνδεδεμένα CRF ζεύγους [26] η ενέργεια της ανάθεσης ενός εικονοστοιχείου σε μια κατηγορία \mathbf{x} δίνεται από την εξίσωση 4.2.

$$E(\mathbf{x}) = \sum_i \psi_u(x_i) + \sum_{i < j} \psi_p(x_i, x_j) \quad (4.2)$$

Όπου $\psi_u(x_i)$ είναι οι ενιαίοι συντελεστές ενέργειας οι οποίοι μετράνε την αντίστροφη πιθανότητα του εικονοστοιχείου i να παίρνει την ετικέτα x_i κα η οι συντελεστές ενέργειας ζεύγους $\psi_p(x_i, x_j)$ μετράνε το κόστος της ανάθεσης της τιμής x_i, x_j στα εικονοστοιχεία i, j ταυτόχρονα.

$$\psi_p(x_i, x_j) = \mu(x_i, x_j) \sum_{m=1}^M w^{(m)} k_G^{(m)}(\mathbf{f}_i, \mathbf{f}_j) \quad (4.3)$$

όπου $K_G^{(m)}$ είναι ένας Γκαουσιανός Πυρήνας (Gaussian Kernel) ο οποίος εφαρμόζεται στα διανύσματα των στοιχείων $\mathbf{f}_i, \mathbf{f}_j$ τα οποία προέρχονται από τα χαρακτηριστικά της εικόνας, όπως πληροφορία θέσης των εικονοστοιχείων και τις τιμές των εικονοστοιχείων (RGB values). Η συνάρτηση $\mu(\cdot, \cdot)$ ονομάζεται συνάρτηση συμβατότητας, η οποία βρίσκει την συμβατότητα μεταξύ ενός ζεύγους εικονοστοιχείων ανάλογα με την ετικέτα που έχει ανατεθεί. Μειώνοντας την συνάρτηση ενέργειας παίρνουμε την πιο πιθανή τιμή (ετικέτα) στο x δεδομένου μιας εικόνας.

Η εύρεση της ακριβής ελάχιστης τιμής είναι ανέφικτη καθώς δεν μπορούμε να υπολογίσουμε εύκολα την συνάρτηση διαμέρισης. Για αυτό τον σκόπο εφαρμόζεται η προσέγγιση Μέσου-Πεδιου (Mean-Field Approximation) στην κατανομή του CRF, η παραπάνω διαδικασία γίνεται με την προσέγγιση της κατανομής $P(\mathbf{X})$ από μια απλούστερη κατανομή $Q(\mathbf{X})$ η οποία μπορεί να γραφτεί σαν ένα γινόμενο ανεξάρτητων περιθωριακών κατανομών $Q(\mathbf{X}) = \prod_i Q_i(X_i)$. Παρακάτω θα δείξουμε αναλυτικά τα βήματα του

αλγορίθμου και πως ο αλγόριθμος Μέσου Πεδίου μπορεί να αναδιαμορφωθεί σαν μια σειρά από πράξεις ενος Σ.Ν.Δ (εικόνα 4.1) και πως μοντελοποιείται σαν ένα ENΔ (RNN) [48].

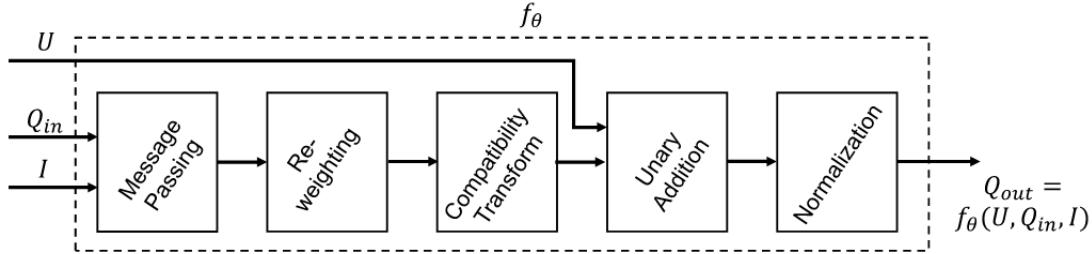


Figure 4.1: Μία επανάληψη Μέσου Πεδίου ως CNN [48]

4.3.2 Αρχικοποίηση

Στο πρώτο βήμα της αρχικοποίησης παρατηρούμε ότι στην ουσία έχουμε την εφαρμογή μιας συνάρτησης softmax όπου $Z_i = \sum_l \exp(U_i(l))$ πάνω στις ενιαίες πιθανότητες (Unary potentials).

$$Q_i(l) \leftarrow \frac{1}{Z_i} \exp(U_i(l)) \quad \triangleright \text{Initialization} \quad (4.4)$$

4.3.3 Πέρασμα Μηνυμάτων

Στα πυκνά μοντέλα CRF το πέρασμα μηνυμάτων πραγματοποιείται εφαρμόζωντας M Γκαουσιανά Φίλτρα στις Q κατανομές. Οι συντελεστές φίλτρων προέρχονται από τα χαρακτηριστικά της εικόνας, όπως οι θέσεις και οι τιμές των εικονοστοιχείων, αλλά και πόσο έντονα ένα εικονοστοιχείο συσχετίζεται με ένα άλλο εικονοστοιχείο της εικόνας καθώς είναι όλα συνδεδεμένα μεταξύ τους. Επειδή ο υπολογισμός μεγάλων διαστάσεων Γκαουσιανών Φίλτρων είναι υπερβολικά μεγάλος, γίνεται η χρήση ενός μεταθετικού πλέγματος (Permutohedral) το οποίο κάνει τον υπολογισμό των φίλτρων σε $O(N)$ [4].

Κατά τη διάρκεια της προς τα πίσω διάδοσης σφάλματος, οι έισοδοι των φίλτρων υπολογίζονται με την αποστολή των παραγώγων σφάλματος ως προς την έξοδο του φίλτρου μέσα από το ίδιο M Γκαουσιανό Φίλτρο με αντίστροφη κατεύθυνση. Όσον αφορά τις πράξεις του πλέγματος, μπορούν να επιτευχθούν μόνο αντιστρέφοντας την σειρά των φίλτρων διαχωρισμού στο στάδιο θολώματος (Blurring stage), κατά την δημιουργία του πλέγματος, τεμαχίζουμε στο πίσω όπως και στο μπροστινό πέρασμα. Ως εκ τούτου, η προς τα πίσω διάδοση μέσω αυτού του σταδίου φιλτραρίσματος μπορεί επίσης να εκτελεστεί σε $O(N)$ χρόνο. Μετά χρησιμοποιούμε δύο Γκαουσιανούς πυρήνες, ένα Πυρήνα Χώρου (Spatial Kernel) και έναν Διμερή Πυρήνα (Bilateral Filter). Επίσης είναι δυνατή η χρήση πολλών χωρικών και διμερών πυρήνων με διαφορετικές τιμές εύρους ζώνης για να μάθουν τη βέλτιστη γραμμική σχέση τους, αλλά χάριν απλότητας χρησιμοποιείται μόνο ένα από το καθένα.

$$\tilde{Q}_i^{(m)}(l) \leftarrow \sum_{j \neq i} k^{(m)}(\mathbf{f}_i, \mathbf{f}_j) Q_j(l) \quad \triangleright \text{Message Passing} \quad (4.5)$$

4.3.4 Στάθμιση Εξόδου Φίλτρου

Το επόμενο βήμα της στον αλγόριθμο μέσου πεδίου λαμβάνει ένα σταθμισμένο όμροισμα των εξόδων φίλτρου M από το προηγούμενο βήμα, για κάθε ετικέτα κλάσης λ . Όταν λαμβάνεται υπόψη κάθε ετικέτα κλάσης μεμονωμένα, αυτό μπορεί να θεωρηθεί ως μια συνήθης συνέλιξη με μέγεθος φίλτρου 1×1 με M κανάλια εισόδου και ένα κανάλι εξόδου. Επειδή και οι είσοδοι και οι έξοδοι σε αυτό το βήμα είναι γνωστές κατά τη διάρκεια της προς τα πίσω διάδοσης, η διαφορά σφάλματος ως προς τα βάρη του φίλτρου μπορεί να υπολογιστεί, καθιστώντας δυνατή την αυτόματη εκμάθηση των βαρών του φίλτρου (σχετικές συνεισφορές από κάθε έξοδο φίλτρου Γκάους από το προηγούμενο στάδιο).

Η παράγωγος του σφάλματος ως προς τις εισόδους μπορούν επίσης να υπολογιστούν με τον ίδιο τρόπο, να περάσουν τα παράγωγα σφάλματος προς τα πίσω, στο πρώτο στάδιο. Για να αποκτήσουμε μεγαλύτερο αριθμό ρυθμιζόμενων παραμέτρων, χρησιμοποιούμε ανεξάρτητα βάρη πυρήνα για κάθε ετικέτα κλάσης. Η διαίσθηση εδώ είναι η σημασία του χωρικού πυρήνα εναντίον του διμερούς πυρήνα η οποία εξαρτάται από την κλάση και την τιμή του εικονοστοιχείου. Για παράδειγμα, οι διμερείς πυρήνες μπορεί από τη μία πλευρά να δίνουν έμφαση στην ανίχνευση ποδηλάτων, η ομοιότητα των χρωμάτων είναι καθοριστική. Αφ' ετέρου δε, μπορεί να έχουν μικρή σημασία για την ανίχνευση της τηλεόρασης, δεδομένου ότι οτιδήποτε βρίσκεται μέσα στην οθόνη της τηλεόρασης μπορεί να έχει πολλούς διαφορετικούς τύπους χρωμάτων.

$$\check{Q}_i(l) \leftarrow \sum_m w^{(m)} \tilde{Q}_i^{(m)}(l) \quad \triangleright \text{Weighting Filter Outputs} \quad (4.6)$$

4.3.5 Μετασχηματισμός Συμβατότητας

Στο βήμα Μετασχηματισμού Συμβατότητας, οι έξοδοι από το προηγούμενο βήμα (εξίσωση 4.6) μοιράζονται μεταξύ των ετικετών, ανάλογα φυσικά με τον βαθμό της συμβατότητας ανάμεσα στις ετικέτες. Η συμβατότητα μεταξύ των ετικετών των εικονοστοιχείων ορίζεται από την συνάρτηση $\mu(l, l')$. Η οποία μαθαίνει την συμβατότητα μεταξύ δύο εικονοστοιχείων, διαισθητικά, η ανάθεση των ετικετών Άνθρωπος και ποδήλατο έχουν μικρότερη ποινή από την ανάθεση των ετικετών ουρανός και ποδήλατο. Επίσης δεν ισχύει η μεταθετικότητα των ετικετών $\mu(l, l') \neq \mu(l', l)$.

Η Συνάρτηση συμβατότητας μπορεί να θεωρηθεί ως ένα επιπλέον Συνελικτικό Επίπεδο όπου το χωρικό πεδίο του φίλτρου είναι 1×1 και ο αριθμός των καναλιών εισόδου και εξόδου είναι Λ . Μαθαίνοντας τα βάρη του φίλτρου είναι ισοδύναμο με την εκπαίδευση της συνάρτησης μ για τις ετικέτες των εικονοστοιχείων.

$$\hat{Q}_i(l) \leftarrow \sum_{l' \in \mathcal{L}} \mu(l, l') \check{Q}_i(l') \quad \triangleright \text{Compatibility Transform} \quad (4.7)$$

Πρόσθεση Πιθανοτήτων

Σε αυτό το βήμα, η έξοδος από τον Μετασχηματισμό Συμβατότητας αφαιρείται από τις ενιαίες πιθανότητες U . Εδώ δεν υπάρχουν παραμετροί, οπότε η διάδοση των διαφορών σφάλματος γίνεται απλά περνώντας τα από την έξοδο προς τις εισόδους.

$$\check{Q}_i(l) \leftarrow U_i(l) - \hat{Q}_i(l') \quad \triangleright \text{Adding Unary Potentials} \quad (4.8)$$

4.3.6 Κανονικοποίηση

Τέλος όπως βλέπουμε στην εξίσωση 4.9 έχουμε την κανονικοποίηση στο τέλος της επανάληψης όπου μπορεί να θεωρηθεί ως μια συνάρτηση softmax χωρίς κάποιες παραμέτρους. Οι διαφορικοί παράγοντες από αυτό το βήμα περνάνε κανονικά προς την είσοδο μέσω της προς τα πίσω διάδοσης.

$$Q_i \leftarrow \frac{1}{Z_i} \exp(\check{Q}_i(l)) \quad \triangleright \text{Normalize} \quad (4.9)$$

4.3.7 CRF as RNN

Εδώ θα εξηγήσουμε πως η επαναληπτική διαδικασία του αλγορίθμου Μέσου Πεδίου μπορεί να μοντελοποιηθεί ως ένα Επαναλαμβανόμενο Νευρωνικό Δίκτυο.

Χρησιμοποιούμε την f_θ για να υποδείξουμε την συνάρτηση μεταφοράς που προκύπτει από μια επανάληψη μέσου πεδίου. Δούσέντος μιας εικόνας I , καθώς και τις ενιαίες πιθανότητες U και την εκτίμηση των περιιωριακών πιθανοτήτων Q_{in} από την προηγούμενη επανάληψη, η επόμενη εκτίμηση των πιθανοτήτων δίνεται από τον εξής τύπο: $f_\theta(U, Q_{in}, I)$. Το διάνυσμα $\theta = \{w^m, \mu(l, l')\}$, $\mu \in \{1, \dots, M\}$, $l, l' \in \{l_1, \dots, l_L\}$ αναπαριστούν τις παραμέτρους του CRF που περιγράψαμε προηγουμένως.

Οι επαναλήψεις του αλγορίθμου του Μέσου Πεδίου υλοποιούνται ως μια στοίβα από επίπεδα με τέτοιο τρόπο ώστε σε κάθε επανάληψη να παίρνει τις εκτιμήσεις Q της προηγούμενης επανάληψης και τις ενιαίες πιθανότητες U από το CNN. Αυτή η διαδικασία που ακολουθείται είναι ίδια με την διαδικασία που ακολουθούν τα RNN για εκπαίδευση. Οι εξισώσεις 4.10, 4.11 και 4.12 μας δείχνουν την διαδικασία της επανεκτίμησης των πιθανοτήτων, όπου T είναι ο αριθμός των επαναλήψεων του Μέσου Πεδίου (Mean-Field Iterations):

$$H_1(t) = \begin{cases} \text{softmax}(U), & t = 0 \\ H_2(t-1), & 0 < t \leq T \end{cases} \quad (4.10)$$

$$H_2(t) = f_\theta(U, H_1(t), I), \quad 0 \leq t \leq T \quad (4.11)$$

$$Y(t) = \begin{cases} 0, & 0 \leq t < T \\ H_2(t), & t = T \end{cases} \quad (4.12)$$

Οι παράμετροι του μοντέλου (CRF-RNN) είναι ίδιες με τις παραμέτρους του αλγορίθμου Μέσου Πεδίου και αναφέρονται ως θ . Επομένως, ο υπολογισμός των διαφορών του σφάλματος ως προς τις παραμέτρους είναι μια επανάληψη Μέσου Πεδίου, μπορούν να εκπαιδευτούν σαν Επαναλαμβανόμενο Νευρωνικό Δίκτυο με τον αλγόριθμο της Προς τα Πίσω Διάδοσης μέσω Χρόνου (Back Propagation Through Time-BPTT) [32, 36].

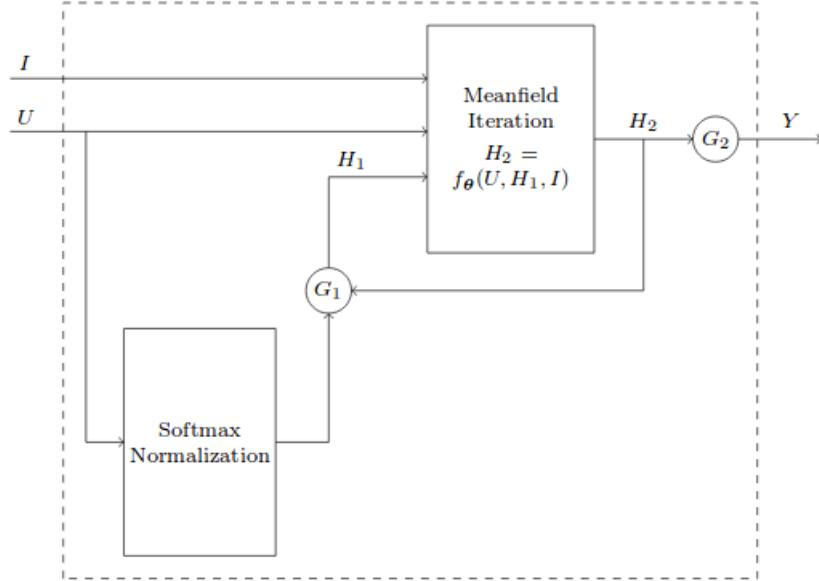


Figure 4.2: **CRF-RNN Network.** Ο επαναληπτικός αλγόριθμος Μέσου Πεδίου ως ένα επαναλαμβανόμενο νευρωνικό δίκτυο. Οι συναρτήσεις G_1, G_2 είναι απλά οι συναρτήσεις εξόδου [48].

Στην 4.3 βλέπουμε την ολοκληρωμένη αρχιτεκτονική, στο δικό μας μοντέλο οι ενιαίες πιθανότητες (Unary potentials U) είναι η έξοδος από το τελευταίο επίπεδο του Π.Σ.Ν.Δ. όπως φαίνεται στην εικόνα.

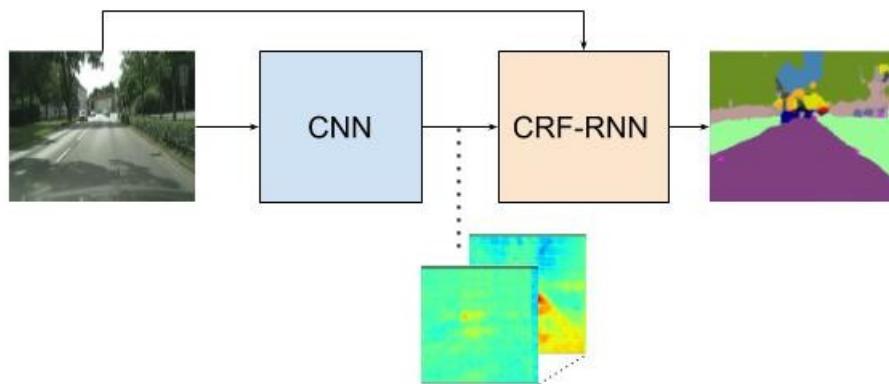


Figure 4.3: **CRF-RNN Network.** Ολοκληρωμένη αρχιτεκτονική του ΠΣΝΔ μαζί με το ΤΥΣΠ-ΕΝΔ. Το ΤΥΣΠ-ΕΝΔ δέχεται ως είσοδο την κανονική εικόνα μαζί με τις ενιαίες πιθανότητες του ΠΣΝΔ.

Chapter 5

Experiments and Results

5.1 Εκπαίδευση των ΝΔ

Σε αυτό το κομμάτι θα παρουσιάσουμε μερικές τεχνικές τις οποίες εφαρμόσαμε για την εκπαίδευση των ΠΣΝΔ καθώς και κάποιες υπερ-παραμέτρους που θέσαμε κατά την διαδικασία της εκπαίδευσης. Η εκπαίδευση των ΠΣΝΔ αποτελεί μια χρονοβόρα διαδικασία και επειδή μπορεί να πάρει μέρες για να συκλίνει, χρίναμε απαραίτητη την τοποθέτηση κάποιων σημείων ελέγχου κατά την διαδικασία της εκπαίδευσης.

5.1.1 Σημεία Ελέγχου (Checkpoints)

Τα σημεία ελέγχου αποτελούν ένα απαραίτητο κομμάτι για την διαδικασία της εκπαίδευσης, ειδικά όταν έχουμε ΠΣΝΔ βαθειάς μάθησης. Η διαδικασία της μάθησης μπορεί να πάρει πολύ χρόνο, όπως στην δική μας περίπτωση που ήταν μερικές μέρες μέχρι να φτάσουμε σε σύγκλιση. Επομένως, πρέπει να αποθηκεύουμε τις παραμέτρους που μαθαίνει το ΠΣΝΔ κατά την μάθηση για να μην συμβεί κάποια αστοχία και χάσουμε χρειαστεί να επανάληψη της διαδικασίας από την αρχή.

Για το δικό μας μοντέλο θέσαμε ως σημείο ελέγχου το τέλος της κάθε εποχής (epoch), όπου αποθηκεύουμε τις παραμέτρους μας σε περίπτωση που χρειαστεί να συνεχίσουμε την εκπαίδευση του ΠΣΝΔ από εκείνο το σημείο. Ο όρος 'Έποχή' αντιπροσωπεύει την τροφοδοσία ενός ΝΔ με το σύνολο δεδομένων εκπαίδευσης. Η διαδικασία η οποία ολόκληρο το σύνολο δεδομένων εκπαίδευσης περνά μία φορά από το στάδιο της εμπρόσθιας διάδοσης και της οπισθοδρόμησης αντίστοιχα ορίζεται ως 'Έποχή'. Ιδανικά, στο τέλος κάθε εποχής ελέγχουμε τα αποτελέσματα της μάθησης, επομένως αν υπάρχει κάποια βελτίωση στην διαδικασία της μάθησης, ελέγχοντας την ακρίβεια του μοντέλου στο σύνολο δεδομένων επαλήθευσης που χρησιμοποιούμε (validation set) στο τέλος κάθε εποχής τότε αποθηκεύουμε τα βάρη του.

5.1.2 Πρώιμο Σταμάτημα (Early Stopping)

Το πρώιμο σταμάτημα είναι ένας μηχανισμός ο οποίος αποσκοπεί στην αποδοτικότητα της εκπαίδευσης του ΠΣΝΔ. Αποσκοπεί στην αποτροπή του μοντέλου από την κατάσταση της υπερ-μάθησης (over-fitting). Ελέγχουμε το σφάλμα από το σύνολο επαλήθευσης σε κάθε εποχή, αν δεν υπάρχει κάποια μείωση του σφάλματος για 12 συνεχώμενα εποχές, τότε σταματάει η διαδικασία της μάθησης. Με αυτόν τον τρόπο σταματάει η διαδικασία της μάθησης πριν το μοντέλο αρίστει να μαθαίνει υπερβολικά το σύνολο δεδομένων της εκπαίδευσης το οποίο αποτελεί πρόβλημα.

5.1.3 Ρυθμός Μάθησης

Ο ρυθμός μάθησης είναι από τις πιο σημαντικές παραμέτρους για την εκπαίδευση των NN. Χρειάζεται να είναι μικρό το μέγεθος για να συγκλίνει, αλλά όχι πολύ μικρό ώστε να πάρει πάρα πολύ χρόνο να βρεθεί σε σύγκλιση. Για την εκπαίδευση των ΣΝΔ βρήκαμε την βέλτιστη τιμή του ρυθμού μάθησης να είναι 10^{-3} χρησιμοποιώντας τον αλγόριθμο Adam ως αλγόριθμο βελτιστοποίησης. Ενώ για την εκπαίδευση του ΣΝΔ μαζί με το ΤΥΣΠ-ΕΝΔ βρήκαμε σαν βέλτιστη επιλογή την χρήση ενός πολύ μικρότερου ρυθμού μάθησης το οποίο ήταν 10^{-13} σε συνδυασμό με τον αλγόριθμο βελτιστοποίησης SGD και με χρήση της παραμέτρου της ορμής επιλεγμένη στο 0.9. Για την ακρίβεια, ξεκινήσαμε με ρυθμό μάθησης 10^{-6} και σταδιακά δοκιμάστηκαν και μικρότεροι ρυθμοί μάθησης μέχρι να καταλήξουμε στο 10^{-13} .

5.2 Αποτελέσματα

Τα μοντέλα εκπαίδευτηκαν σε 2975 εικόνες μεγέθους 512×512 η κάθε μία, ενώ το σύνολο των εικόνων επαλήθευσης το οποίο χρησιμοποιούμε για την επαλήθευση του μοντέλου στο τέλος της κάθε εποχής είναι 500 εικόνες. Επίσης οι εικόνες έχουν επαληθευτεί στο κανονικό τους μέγεθος (1024×2048). Τα μοντέλα εκπαίδευτηκαν σε παρτίδες (batches) όπου το μέγεθος ήταν 2 και 4 εκτός από την ολοκληρωμένη εκπαίδευση του (end-to-end) μοντέλου (ΠΣΝΔ-ΤΥΣΠ-ΕΝΔ) που χρησιμοποιήθηκε μέγεθος ίσο με ένα λόγω των διαθέσιμων πόρων. Τα αποτελέσματα στον πίνακα 5.1 μας δείχνουν τις επιδόσεις των ΠΣΝΔ σε συνδυασμό με την μονάδα επεξεργασίας Μέσου Φίλτρου καθώς δοκιμάζουμε τις επιδόσεις με διαφορετικό μέγεθος παραθύρου, ενώ ο πίνακας 5.2 επιδεικνύει τις επιδόσεις του μοντέλου με μονάδα μετά-επεξεργασίας ΤΥΣΠ-ΕΝΔ καθώς και την σύγκριση με τις αναδρομικές επαναλήψεις κατά την δοκιμή. Η δοκιμή του μοντέλου έγινε στα δεδομένα επαλήθευσης, δηλαδή στις 500 εικόνες. Η μετρική που χρησιμοποιούμε στα αποτελέσματα είναι ο μέσος όρος των σημείων τομής ως προς την ένωση δύο συνόλων (mean Intersection over Union ή Jaccard Similarity) και ορίζεται με την παρακάτω συνάρτηση.

Η εξίσωση 5.1 μας δίνει το σημείο τομής δύο συνόλων ως προς την ένωσή τους. Στην δική μας περίπτωση αυτα τα δύο σύνολα ορίζουν τις εικόνες με τις προβλέψεις και τις εικόνες με τις πραγματικές ταξινομήσεις των εικονοστοιχείων.

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (5.1)$$

Η εξίσωση 5.1 γράφεται όπως βλέπουμε παρακάτω:

$$J = \frac{TP}{TP + FP + FN} \quad (5.2)$$

$TP = \text{True Positives}$
 $FP = \text{False Positives}$
 $FN = \text{False Negatives}$

True Positives συμβολίζονται τα εικονοστοιχεία τα οποία έχουν προβλεφθεί σωστά από τον ταξινομητή. False Positives συμβολίζονται τα εικονοστοιχεία τα οποία έχουν προβλεφθεί σε μια θετική κλάση ενώ ανήκουν σε μία αρνητική και False Negatives είναι οι προβλέψεις οι οποίες έχουν προβλεφθεί αρνητικά ενώ η κανονική τους κατάσταση είναι θετική. Η παραπάνω εξήγηση αφορά δύο μόνο κατηγορίες (Θετική, Αρνητική) για λόγους απλότητας, στην πραγματικότητα για περισσότερες κατηγορίες, αν i είναι η πρόβλεψη του ταξινομητή και j η πραγματική κατηγορία που ανήκει το εικονοστοιχείο, τότε για $j > i$ θεωρούνται ως False Positives ενώ για $i < j$ θεωρούνται ως False Negatives.

Για την μέτρηση των αποτελεσμάτων μετράμε συνολικά από όλη την εικόνα τις παραμέτρους TP, FP, FN και υπολογίζουμε την συνάρτηση J. Επομένως μετράμε την συνάρτηση Jaccard Similarity (J) από κάθε εικόνα και παίρνουμε έναν μέσο όρο από τις 500 εικόνες που χρησιμοποιήσαμε για την δοκιμή του μοντέλου. Επίσης, κατά την μέτρηση του μοντέλου δεν λάβαμε υπόψη μας τα εικονοστοιχεία τα οποία είναι ταξινομημένα ως 'Κενά'. Δηλαδή, αν ένα εικονοστοιχείο έχει ταξινομηθεί σε μία οποιαδήποτε κλάση i και ανήκει στην κλάση 'Κενό' τότε δεν συνεισφέρει στο αποτέλεσμα.

Model	-	9x9	19x19	31x31	45x45	59x59	65x65	71x71	77x77	81x81
SD-CNN-MFB	0.75031	0.75093	0.75134	0.75178	0.75219	0.75245	0.75245	0.75238	0.75226	0.75212
SD-CNN	0.86425	0.86444	0.86446	0.86425	0.86335	0.86231	0.86157	0.86074	0.85985	0.85921
SD-CNN-CRF[5]	0.76086	0.76143	0.76183	0.76231	0.76280	0.76305	0.76306	0.76298	0.76280	0.76264
BD-CNN	0.88850	0.88861	0.88869	0.88566	0.88784	0.88641	0.88552	0.88335	0.88335	0.88251

Table 5.1: Αποτελέσματα των ΠΣΝΔ με τον αλγόριθμο Μέσου Φίτρου ως μονάδα μετά-επεξεργασίας χρησιμοποιώντας διαφορετικά μεγέθη παραθύρων. SD (Strided Deconvolution) είναι η μονάδα με αποκωδικοποίησης με βήμα ολίσθησης ενώ BD Bilinear Deconvolution είναι η διγραμμική μονάδα αποκωδικοποίησης. Με MFB (Median Frequency Balance) συμβολίζουμε την συνάρτηση ισοστάθμισης που χρησιμοποιήσαμε.

Όπως φαίνεται στον πίνακα 5.1 το μεσαίο φίλτρο δεν προσδίδει ιδιαίτερη βελτίωση στα αποτελέσματα παρά μόνο μια μικρή εξομάλυνση. Για την ακρίβεια η βελτίωση είναι της τάξης του 0.2% για έναν συγκεκριμένο αριθμό μεγέθους παραθύρου (19×19).

Τα μοντέλο SD-CNN το οποίο εκπαιδεύτηκε με την συνάρτηση ισοστάθμισης μέσης συχνότητας πήρε 70 εποχές μέχρι να επιτευχθεί σύγκλιση καθώς επειδή προσπαθεί το ΣΝΔ να μάθει πληροφορία από όλες τις κλάσεις ανεξάρτητα της δυσαναλογίας χρειάζεται περισσότερο χρόνο για την σύγκλιση εφόσον η λανθασμένη ταξινόμηση ενός

εικονοστοιχείου που βρίσκεται σε μια σπάνια κατηγορία διαδίδει μεγαλύτερο σφάλμα προς τα πίσω στο ΣΝΔ. Τα ΣΝΔ που εκπαιδεύτηκαν χωρίς ισοστάθμιση, χρειάστηκαν μόνο 40 εποχές για να συγκλίνουν καθώς βρέθηκαν πολύ γρήγορα σε κατάσταση υπερμάθησης.

Ο πίνακας 5.2 δείχνει την επίδοση του μοντέλου SD-CNN μαζί με την μονάδα μετά-επεξεργασίας ΤΥΣΠ-ΕΝΔ (CRF-RNN). Αρχικά επιχειρήσαμε να παγώσουμε την μάθηση στο ΠΣΝΔ και να γίνει η εκπαίδευση μόνο στο ΤΥΣΠ-ΕΝΔ όμως δεν υπήρχε κάποιο θετικό αποτέλεσμα. Εν τέλει, ξεκινήσαμε να εκπαίδεύσουμε το ΠΣΝΔ σε συνδυασμό με το ΤΥΣΠ-ΕΝΔ αρχικά με 10 επαναλήψεις και με ρυθμό μάθησης 10^{-6} για να δούμε την ανταπόκριση του μοντέλου. Σταδιακά μειώσαμε τον ρυθμό μάθησης σε 10^{-13} όμως ακόμα και μετα από 20 εποχές το μοντέλο άρχισε να αποκλίνει. Πιθανότητα λόγω των πολλών επανάληψεων στο ΤΥΣΠ-ΕΝΔ παρουσιάστηκε το φαινόμενο της εξαφάνισης των αποκλίσεων. Επομένως μειώσαμε τον αριθμό των επαναλήψεων σε 5 κατά την μάθηση και πετύχαμε σύγκλιση μετά από 30 εποχές.

Το πρόβλημα με τους γράφους είναι ότι στηρίζονται πάρα πολύ στον υπάρχων ταξινομητή για να συνεισφέρουν μια καλύτερη επίδοση. Στην προκειμένη περίπτωση πετύχαμε μια βελτίωση της τάξης του 1%. Από τον πίνακα 5.2 είναι ολοφάνερο πως όσο αυξάνονται οι επανάληψεις επιβαρύνεται με επιπλέον χρόνο το μοντέλο καθώς σε κάθε επανάληψη γίνεται επανεκτίμηση της κατανομής. Για επαναλήψεις μεγαλύτερες από 5 το μοντέλο δεν παρουσιάζει κάποια βελτίωση, επίσης στον πίνακα βλέπουμε και τον χρόνο της συμπερασματολογίας καθώς και μια τυπική απόκλιση του χρόνου σε δευτερόλεπτα που μετρήσαμε από την συμπερασματολογία 500 εικόνων.

Model[Iterations]	mean IoU	Χρόνος Διεκπεραίωσης[s]	Απόκλιση Χρόνου
SD-CNN	0.7503	0.1111	0.06
SD-CNN-CRF[3]	0.7601	0.3988	0.09
SD-CNN-CRF[5]	0.7608	0.6994	0.17
SD-CNN-CRF[10]	0.7608	1.1868	0.09
SD-CNN-CRF[20]	0.7608	2.5681	0.48

Table 5.2: Σύγκριση των μοντέλων με την μονάδα μετά-επεξεργασίας ΤΥΣΠ-ΕΝΔ με διφορετικό αριθμό επαναλήψεων.

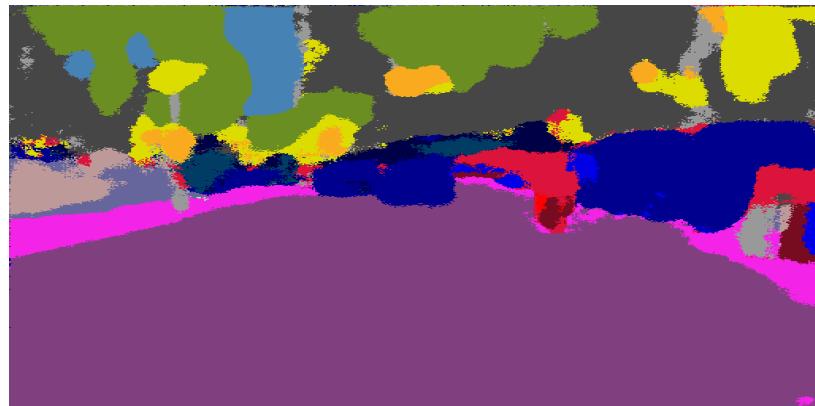
Παρακάτω βλέπουμε μερικές εκτιμήσεις πάνω σε εικόνες από τα μοντέλα που παρουσιάσαμε. Στην εικόνα 5.1 βλέπουμε τα αποτελέσματα από μια εικόνα εισόδου στο μοντέλο SD-CNN-CRF το οποίο έχει εκπαιδευτεί με ισοστάθμιση κλάσεων. Σε γενικές γραμμές έχει καταφέρει να τυμηματοποιήσει αρκετά από τα αντικείμενα αν και όχι στη πιο λεπτομερή μορφή. Ο λόγος πιθανότητα που υπάρχει μια αστοχία και μια υπερκατάτμηση σε αντικείμενα όπως οι πινακίδες κυκλοφορίας και τα φανάρια κυκλοφορίας είναι η συμπίεση που υπέστη το μοντέλο από τα τμήματα συγκέντρωσης τα οποία χάνουν αρκετή πληροφορία ενώ τα επίπεδα υπερδειγματοληψίας φαίνεται να μην μπορούν να ανταπεξέλισουν σε αυτό το πρόβλημα. Επίσης, το μεσαίο φίλτρο ομαλοποιεί κάπως κάποια απομακρυσμένα εικονοστοιχεία από την κατηγορία τους τα οποία έχουν ταξινομηθεί σε λάθος κατηγορία.

Στην εικόνα 5.2 βλέπουμε το ίδιο μοντέλο ΠΣΝΔ με προηγουμένως χωρίς να έχει εκπαιδευτεί με κάποια συνάρτηση ισοστάθμισης. Αυτό είναι προφανές καθώς το ΠΣΝΔ έχει μάθει πολύ λιγότερες κατηγορίες, δηλαδή τις επικρατέστερες κατά πλειοψηφία. Αν και το ΠΣΝΔ έχει μάθει αρκετά καλά τις επικρατέστερες κλάσεις, γενικά αδυνατεί να αναγνωρίσει οποιαδήποτε άλλη κλάση.

Τέλος, στην εικόνα 5.3 βλέπουμε ένα δείγμα από το μοντέλο με την διγραμμική αποκωδικοποίηση το οποίο δεν έχει εκπαιδευση με την συνάρτηση για την δυσαναλογία των κλάσεων. Κάτι που παρατηρούμε είναι πως αν και χωρίς ισοστάθμιση κλάσεων καταφέρνει να ανγνωρίσει περισσότερα εικονοστοιχεία στην εικόνα εισόδου που ανήκουν και σε κλάσεις που αποτελούν μειονότητα. Μια εξήγηση για αυτό είναι ότι το προηγούμενο μοντέλο διαθέτει περισσότερες παραμέτρους λόγω του τμήματος υπερδειγματοληψίας που διαθέτει με εκπαιδευόμενες παραμέτρους τείνει να πάσχει από μεγαλύτερο πρόβλημα υπερμάθησης.



(a) Εικόνα εισόδου



(b) Πρόβλεψη μοντέλου CNN-CRF-RNN.



(c) Πρόβλεψη μοντέλου CNN-CRF-RNN με μεσαίο φίλτρο.



(d) Ground Truth

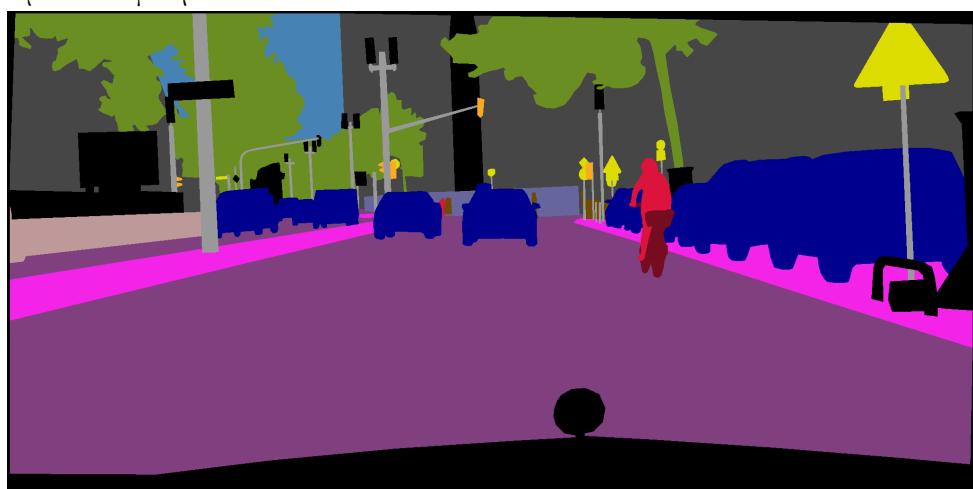
Figure 5.1: Εικόνες αποτελεσμάτων του ΠΣΝΔ με ΤΥΣΠ-ΕΝΔ μαζί με την εικόνα αλήθειας για σύγκριση.



(a) Πρόβλεψη μοντέλου SD-CNN χωρίς ισοστάθμιση κλάσεων.



(b) Πρόβλεψη μοντέλου SD-CNN χωρίς ισοστάθμιση κλάσεων με την εφαρμογή του βέλτιστου παραθύρου μεσαίου φίλτρου.

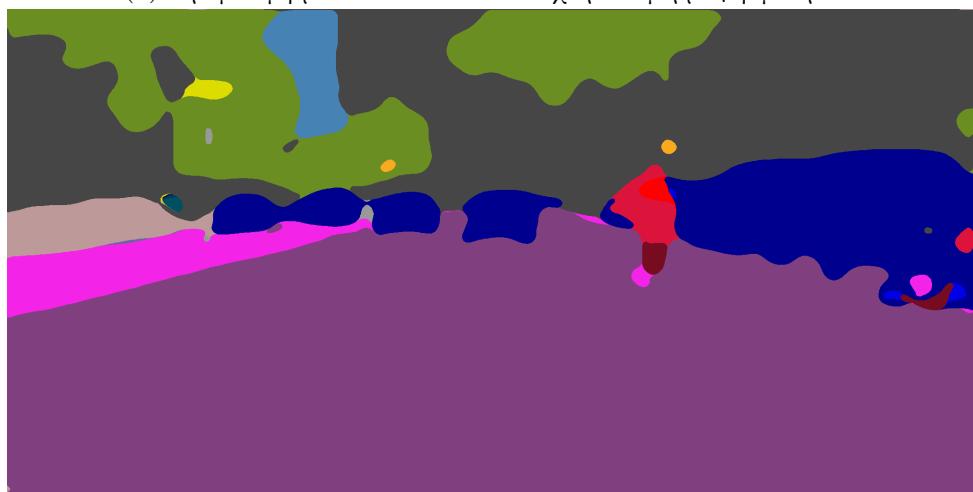


(c) Ground Truth

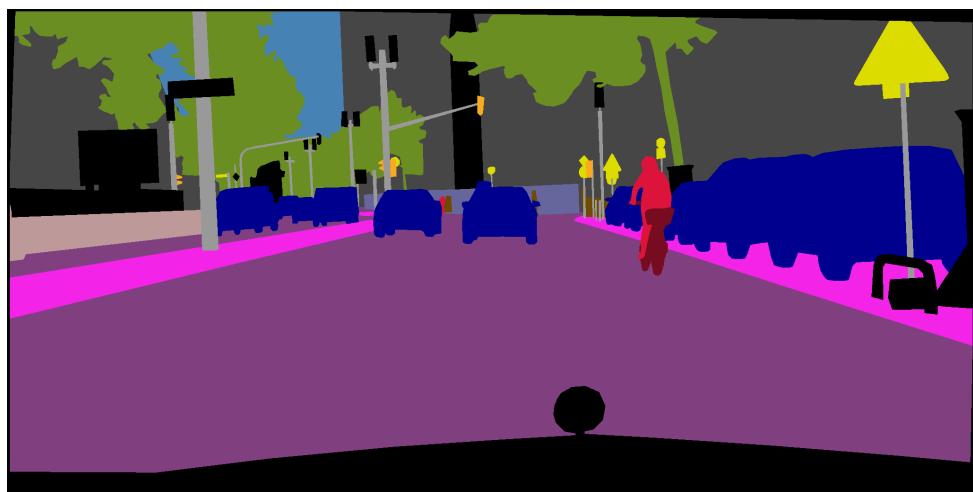
Figure 5.2: Εικόνες αποτελεσμάτων του ΠΣΝΔ χωρίς ισοστάθμιση κλάσεων με εφαρμογή μεσαίου φίλτρου και χωρίς.



(a) Πρόβλεψη μοντέλου BD-CNN χωρίς εφαρμογή φίλτρου.



(b) Πρόβλεψη μοντέλου BD-CNN με την εφαρμογή του βέλτιστου παραθύρου (19×19) μεσαίου φίλτρου.



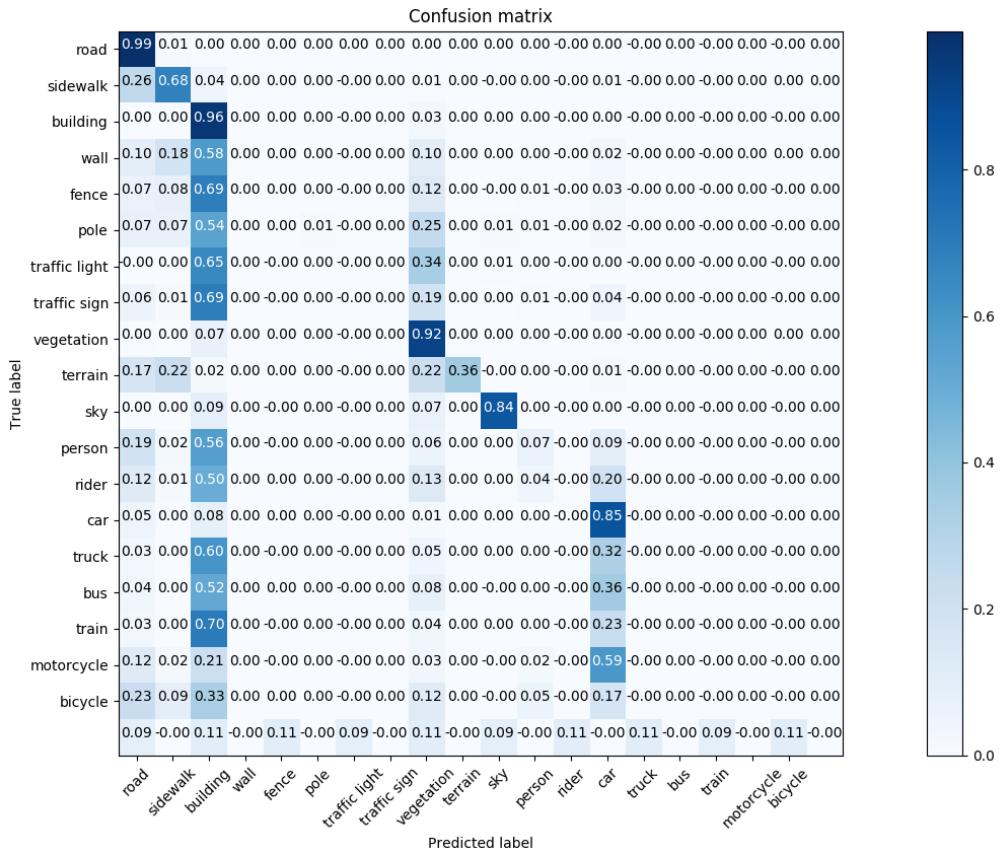
(c) Ground Truth

Figure 5.3: Εικόνες αποτελεσμάτων του ΠΣΝΔ χωρίς ισοστάθμιση κλάσεων με εφαρμογή μεσαίου φίλτρου και χωρίς.

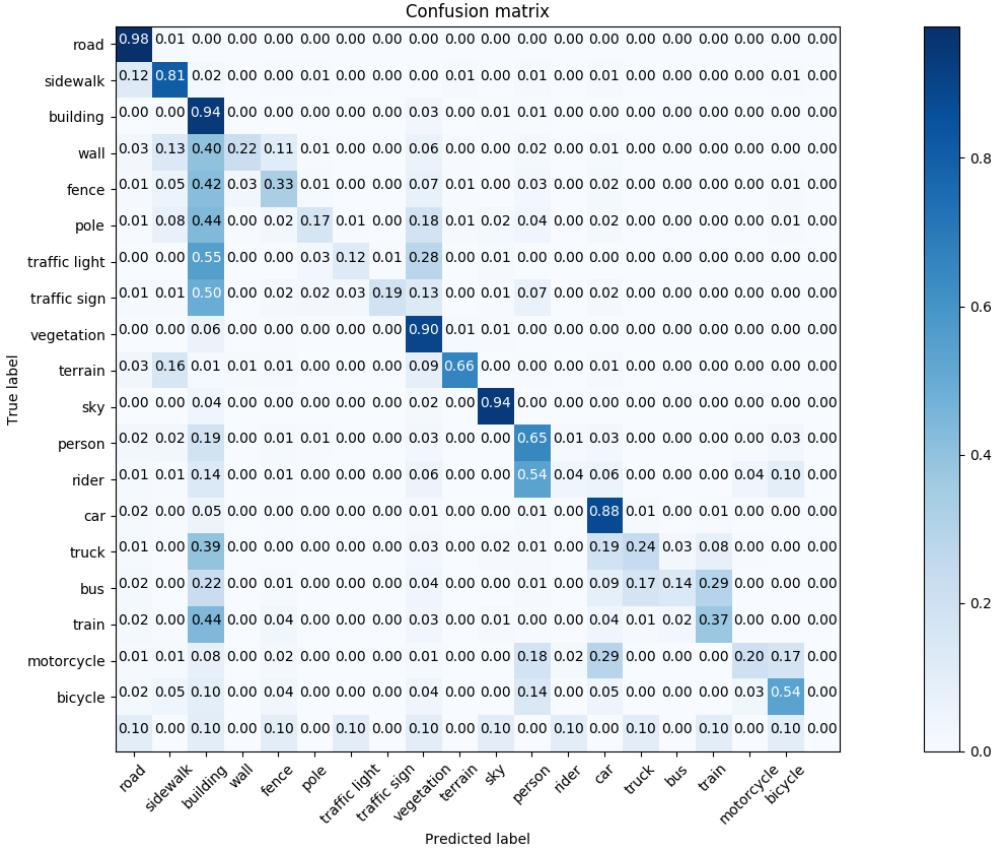
Οι πίνακες σύγχυσης που βλέπουμε παρακάτω στις εικόνες αποτελούν μια πιο αξιόπιστη μέθοδο εικονοποίησης των αποτελεσμάτων για να δούμε πόσο ισχυρά είναι τα μοντέλα μας και ποιες κατηγορίες μπερδεύονται περισσότερο μεταξύ τους. Οι πίνακες παρουσιάζονται με μια κανονικοποιημένη μορφή, ενώ όσο πιο έντονο είναι το χρώμα στην διαγώνιο του πίνακα τόσο πιο δυνατό είναι το μοντέλο. Η μετρική Jaccard Similarity που είδαμε προηγουμένως επειδή υπολογίζει καθολικά σε όλη την εικόνα τις παραμέτρους της βγάζει μια μέση τιμή από αυτές. Στην αριστερή πλευρά έχουμε τις πραγματικές κατηγοροποιήσεις των εικονοστοιχείων ενώ κάθετα έχουμε τις προβλέψεις των μοντέλων μας. Η τελευταία στήλη και σειρά ανήκουν στην κατηγορία 'χωρίς ετικέτα' για αυτό και το αφήσαμε κενό.

Η εικόνα 5.4 μας συγχρίνει τα μοντέλα που δεν έχουν εκπαιδευτεί με την μέθοδο της ισοστάθμισης. Για τους λόγους που εξηγήσαμε προηγουμένως, το μοντέλο με την διγραμμική μονάδα (b) τα πηγαίνει αρκετά καλύτερα.

Τέλος, στην εικόνα 5.5 επιδεικνύεται η σύγκριση μεταξύ του μοντέλου SD-CNN με ισοστάθμιση των κλάσεων (a) καθώς και του end-to-end μοντέλου SD-CNN-CRF-RNN με αριθμό επαναλήψεων επανεκτίμησης ίσο με πέντε. Η διαφορά μεταξύ των δύο μοντέλων δεν είναι πολυ μεγάλη άλλωστε όπως είδαμε και πριν ήταν μόλις 1%. Όμως μπορούμε να δούμε πως πολλές λανθασμένες προβλέψεις έχουν μειωθεί αρκετά και αυτό δείχνει την ισχύ του μοντέλου μετά-επεξεργασίας. Αξίζει να δούμε την κλάση μοτοσυκλέτα η οποία έχει μειωθεί στο ΤΥΣΠ-ΕΝΔ. Πιθανόν τα εικονοστοιχεία που ταιριάζουν σε αυτή την κλάση, μοιάζουν αρκετά με εικονοστοιχεία κάποιας άλλης κλάσης και για αυτό τυχαίνει να υπάρχει μείωση. Οι πίνακες σύγχησης υπολογίστηκαν πάνω στις 500 εικόνες του συνόλου δεδομένων επαλήθευσης στο αρχικό μέγεθος της εικόνας.

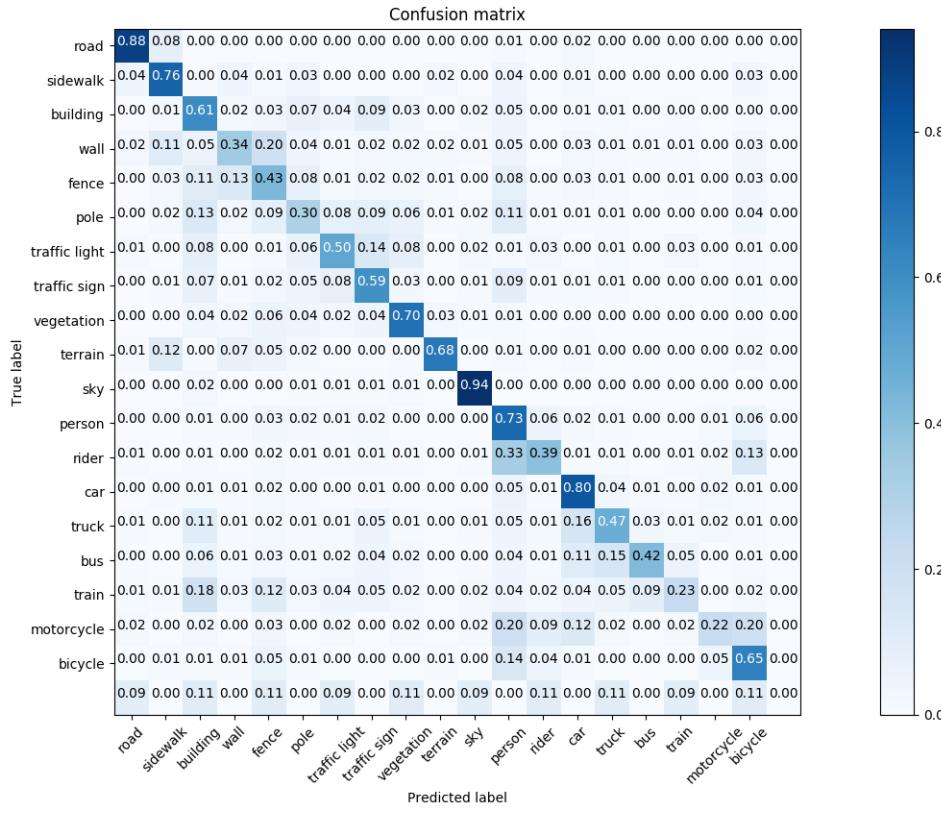


(a) Confusion Matrix του ΠΣΝΔ με μονάδα αποκωδικοπίησης (SD-CNN) χωρίς εφαρμογή της συνάρτησης ισοστάθμισης.

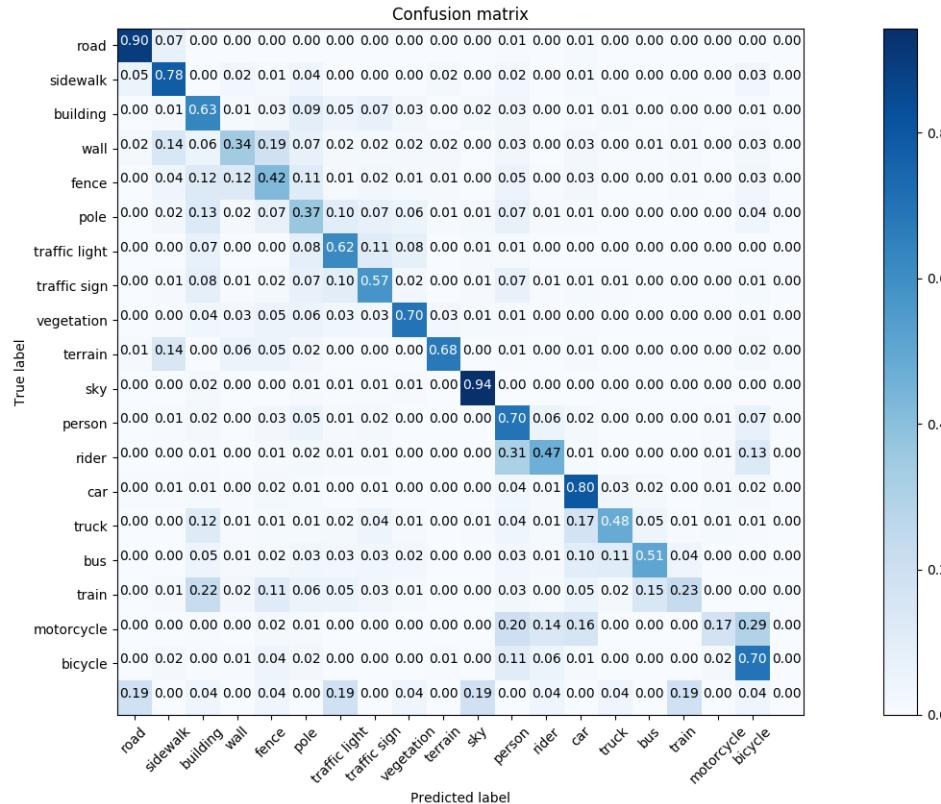


(b) Confusion Matrix του ΠΣΝΔ με διγραμμική μονάδα αποκωδικοπίησης (BD-CNN) χωρίς εφαρμογή της συνάρτησης ισοστάθμισης.

Figure 5.4: Πίνακες Σύγχυσης χωρίς CRF Πίνακες σύγχυσης των μοντέλων χωρίς μονάδες μετά-επεξεργασίας και ισοστάθμιση των κλάσεων.



(a) Confusion Matrix του ΠΣΝΔ με μονάδα αποκωδικοποίησης (SD-CNN) με ισοστάθμιση των κλάσεων.



(b) Confusion Matrix του ΠΣΝΔ με μονάδα αποκωδικοποίησης με άλμα ολίσθησης και με μονάδα μετά-επεξεργασίας ΤΥΣΠ-ΕΝΔ (SD-CNN-CRF) με ισοστάθμιση των κλάσεων.

Figure 5.5: Πίνακες σύγχυσης για την σύγκριση των μοντέλων με και χωρίς μονάδα μετά-επεξεργασίας και ισοστάθμιση των κλάσεων.

Chapter 6

Συμπεράσματα και Μελλοντική Εργασία

6.1 Συμπεράσματα

Ο σκοπός της συγκεκριμένης διπλωματικής ήταν η εξερεύνηση και η σύγκριση τεχνικών τελευταίας γενιάς στο πεδίο της μηχανικής μάθησης για το πρόβλημα της σημασιολογικής κατάτμησης αντικειμένων από εικόνες αλλά και η δημιουργία κατάλληλου εργαλείου για την προβολή των προβλέψεων από τα μοντέλα. Πιο συγκεκριμένα, επικεντρωθήκαμε στην εφαρμογή μεθόδων βαθειάς μάθησης με χρήση αρχιτεκτονικών ΠΣΝΔ με το μοντέλο γράφων ΤΥΣΠ-ΕΝΔ για την εκτίμηση της κατηγορίας που ανήκει κάθε εικονοστοιχείο της εικόνας. Τα δίκτυα ΠΣΝΔ αποτελούν μια από τις τεχνικές τελευταίας γενιάς στα προβλήματα σημασιολογικής κατάτμησης ειδικά σε συνδυασμό με τους γράφους ΤΥΣΠ καιώς έχουν επιδείξει πολύ καλά αποτελέσματα.

Μια ενδιαφέρουσα συνεισφορά της εργασίας είναι η χρήση της εκθετικής συνάρτησης ενεργοποίησης η οποία σε συνδυασμό με την σωστή συνάρτηση αρχικοποίησης των βαρών αντιμετωπίζουν το πρόβλημα των νεκρών νευρώνων το οποίο τείνουν να πάσχουν τα βαθειά ΝΔ. Επίσης, αυτή η προέγγιση είναι πιο αποδοτική κανώς δεν προσθέτει υπολογιστικό κόστος στην εκπαίδευση του ΝΔ σε σχέση με άλλες προσεγγίσεις.

Τέλος, παρουσιάσαμε και συγχρίναμε δύο πανομοιότυπες αρχιτεκτονικές βασισμένες σε ΣΝΔ κωδικοποίησης και αποκωδικοποίησης και πως αυτές ανταποκρίνονται. Είναι φανερό πως το ΣΝΔ με την μονάδα αποκωδικοποίησης με άλμα ολίσθησης αν και διαθέτει μεγαλύτερο αριθμό παραμέτρων, δίνει καλύτερα αποτελέσματα λόγω της μη γραμμικής υπερδειγματοληψίας η οποία διαθέτει παραμέτρους που μαθαίνουν την χαρτογράφηση της υπερδειγματοληψίας κατά την εκπαίδευση. Επίσης, η χρησιμότητα της συνάρτησης μέσης συχνότητας ισορροπίας, η οποία παίζει σημαντικό ρόλο σε τέτοιου είδος προβλήματα στο στάδιο της εκπαίδευσης, κανώς επιτυγχάνεται μια ισσοροπία ως ένα βαθμό μετξύ της δυσαναλογίας των κλάσεων που υπάρχει στα δεδομένα.

6.2 Μελλοντική Εργασία

Το υέμα της σημασιολογικής κατάτμησης κεντρίζει όλο και περίσσοτερο το ενδιαφέρον των επιστημόνων καθώς αποτελεί πρόκληση στον κλάδο, ενώ η συνεχής ανάπτυξη της υπολογιστικής δύναμης η οποία είναι απαραίτητη σε συνδυασμό με την δημιουργία καινούριων αυτόνομων μηχανών που παίρνουν αποφάσεις σύμφωνα με τον ακριβή διαχωρισμό των αντικειμένων στο περιβάλλον [6, 43], έχουν σαν αποτέλεσμα την μεταστροφή από προβλήματα ανίχνευσης αντικειμένων στην σημασιολογική κατάτμηση.

Στο μέλλον θα θέλαμε να χρησιμοποιήσουμε πιο βαθειά μοντέλα χρησιμοποιώντας λιγότερη υποδειγματοληψία στις εικόνες εισόδου για περισσότερη πληροφορία. Επίσης θα θέλαμε να χρησιμοποιήσουμε περισσότερα δεδομένα, όμως υπάρχει δυσκολία σε αυτό το κομμάτι καθώς θα πρέπει να δημιουργηθούν καινούριες εικόνες με κατηγοροποιημένα όλα τα εικονοστοιχεία. Μία καλή προσέγγιση θα ήταν η δημιουργία συνθετικών δεδομένων από τις ήδη υπάρχουσες εικόνες. Οι συνθετικές εικόνες δημιουργούνται με εφαρμογή από μια πληθώρα κατάλληλων φίλτρων πάνω στις εικόνες ώστε να δημιουργήσουμε παραλλαγές των εικόνων και ως αποτέλεσμα περισσότερα δεδομένα για την αντιμετώπιση του προβλήματος της υπερμάθησης. Επίσης, θα θέλαμε να δοκιμάσουμε την αρχιτεκτονική με την διγραμμική υπερδιεγματοληψία με περισσότερα φίλτρα σε συνδυασμό με την μονάδα ΤΥΣΠ-ΕΝΔ καθώς είναι πιθανόν να υπάρχουν προοπτικές.

Μία διαφορετική κατεύθυνση είναι η χρήση ενός προεκπαιδευμένου ΠΣΝΔ το οποίο έχει εκπαιδευτεί σε κάποιο διαφορετικό πρόβλημα. Η χρήση ενός τέτοιου μοντέλου βοηθάει στην εξαγωγή πολύπλοκων χαρακτηριστικών τα οποία μπορούν να τροφοδοτήσουν ένα ΠΣΝΔ όπως το δικό μας. Η εκπαίδευση ενός τέτοιου μοντέλου σε συνδυασμό με το δικό μας μοντέλο θα μπορούσε να προσφέρει καλύτερα αποτελέσματα. Ο μεγαλύτερος αριθμός παρτίδας επίσης, θα μπορούσε να επιφέρει καλύτερα αποτελέσματα, καθώς χρησιμοποιήσαμε μία παρτίδα της τάξης του 4 στην καλύτερη περίπτωση, ο υπολογισμός σε περισσότερα δεδομένα σε κάθε επανάληψη και ανανέωση των κρυμμένων στοιχείων ανά περισσότερα κομμάτια θα μπορούσε να βελτιώσει τα αποτελέσματα. Δυστυχώς, η αύξηση της παρτίδας και ειδικά σε δεδομένα πολύ μεγάλων διαστάσεων απαιτούν και αρκετούς πόρους.

Τέλος, η παράλληλη μονάδα επεξεργασίας θα μπορούσε να βελτιώσει τα αποτελέσματα αν μπορούσαμε να την χρησιμοποιήσουμε με μεγαλύτερα μεγέθη χαρτών χαρακτηριστικών στην είσοδο, καθώς θα μπορούσαν να αποδώσουν καλύτερα στην αξιοποίηση της πληροφορίας λόγω της μικρότερης συμπίεσης. Επίσης η αύξηση των αριθμών των φίλτρων σε κάθε κλάδο θα μπορούσαν να αποδώσουν θετικά καθώς θα υπήρχαν περισσότεροι χάρτες χαρακτηριστικών, όμως αυτή η επιλογή έρχεται με αντάλλαγμα την αύξηση των παραμέτρων.

Παράρτημα Α

Οι αλγόριθμοι και τα μοντέλα που χρησιμοποιήθηκαν σε αυτή την διπλωματική βρίσκονται στο προφίλ του συγγραφέα στο Github στον παρακάτω σύνδεσμο [link](#). Σε αυτό το τμήμα θα δείξουμε το λογισμικό το οποίο υλοποιήθηκε στα πλαίσια της εργασίας για την εικονοποίηση των αποτελεσμάτων. Το λογισμικό υλοποιήθηκε με την χρήση των βιβλιοθηκών PyQt4 [34] και OpenCV [20, 21], ενώ για την υλοποίηση των μοντέλων χρησιμοποιήθηκαν οι βιβλιοθήκες Keras [11], Tensorflow [3] και scikit-learn [7].

Ο πίνακας 6.1 επιδεικνύει τα χρώματα που αντιστοιχούν στην κάθε κλάση τα οποία χρησιμοποιούνται στο λογισμικό για την εικονικοποίηση των κλάσεων. Κάθε εικονοστοιχείο το οποίο ανήκει σε κάποια συγκεκριμένη κλάση αντιστοιχεί και το αντίστοιχο χρώμα.

Δρόμος	Πεζόδρομος	Κτίριο	Τοίχος	Φράχτης
Ιστός	Φανάρι Κυκλοφορίας	Πινακίδα Κυκλοφορίας	Βλάστηση	Έδαφος
Ουρανός	Άνθρωπος	Αναβάτης	Αυτοκίνητο	Φορτηγό
Λεωφορείο	Τρένο	Μοτοσυκλέτα	Ποδήλατο	Κενό

Table 6.1: Χρώματα των διαφορετικών αντικειμένων τα οποία φαίνονται στο παρακάτω λογισμικό.

Όπως βλέπουμε στις εικόνες 6.1, 6.2 κατά την εκτέλεση του προγράμματος εμφανίζεται παράνυρο επιλογής μίας εικόνας για αναγνώριση και ακολουθεί δεύτερο παράνυρο αντίστοιχα για την επιλογή ενός αρχείου/εικόνας το οποίο περιέχει τις προβλέψεις που έχουν παραχθεί από το μοντέλο μας.

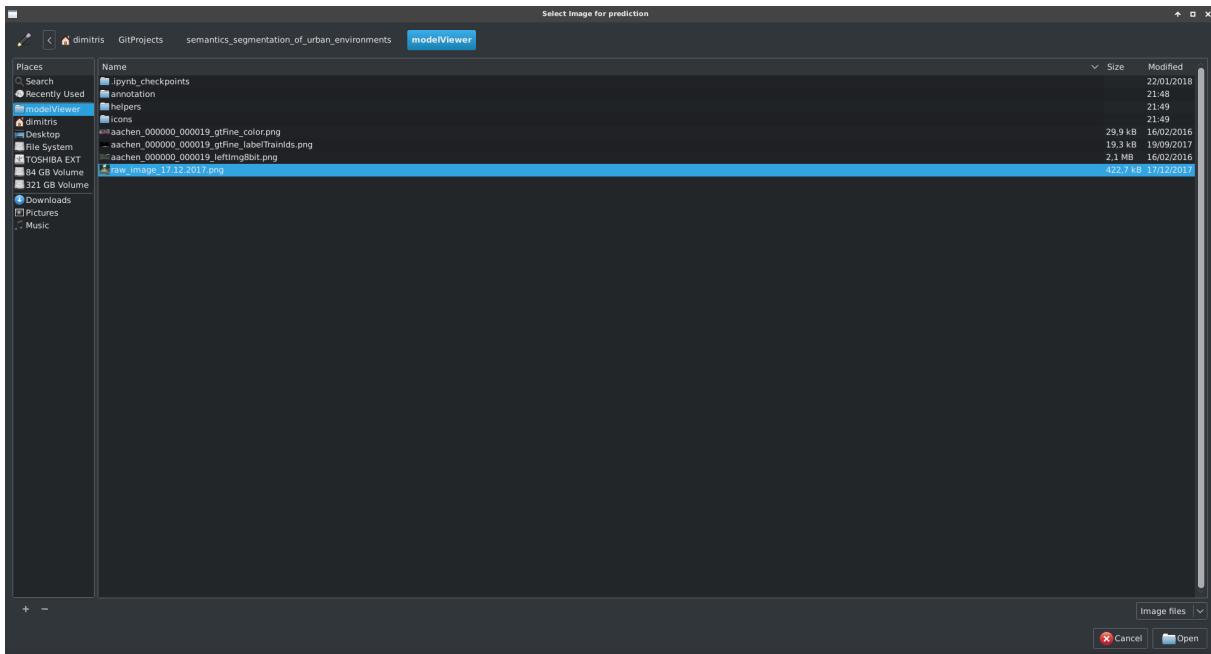


Figure 6.1: Επιλογή αρχικής εικόνας για αναγνώριση.

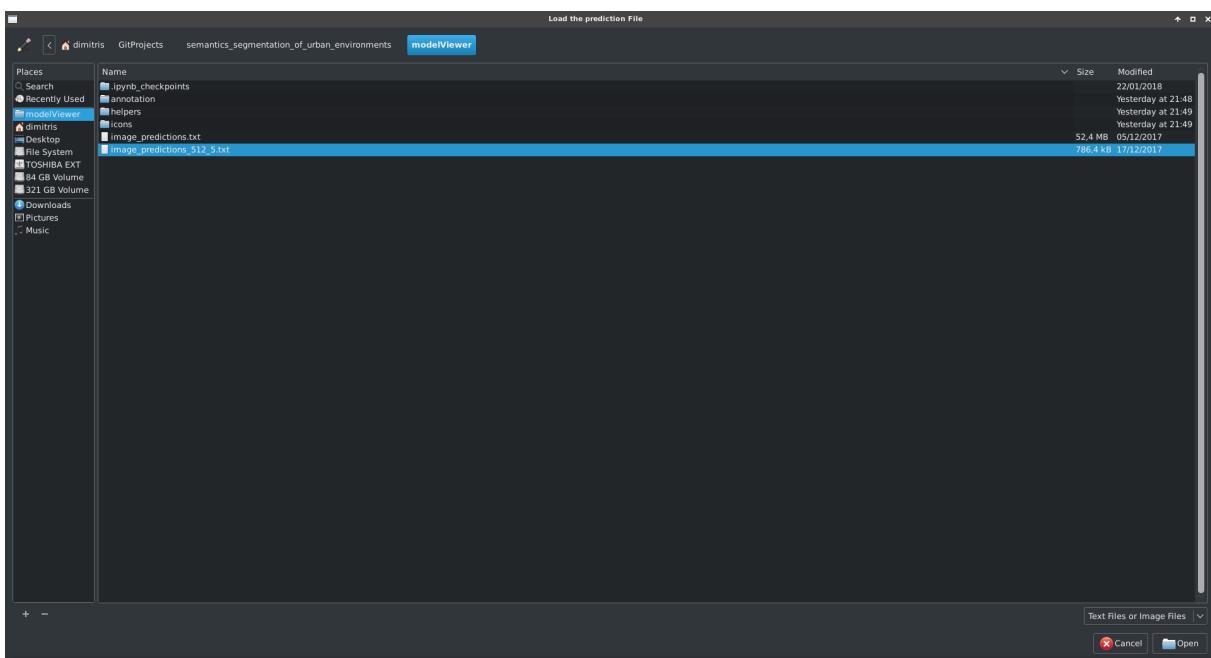


Figure 6.2: Επιλογή αρχείου/εικόνας αποτελεσμάτων

Η εικόνα με τις προβλέψεις ζωγραφίζεται πάνω από την εικόνα εισόδου, όπου κάθε κατηγορία αντικειμένων έχει ένα χρώμα που την αντιπροσωπεύει [6.3, 6.4]. Ο δείκτης ανάλογα με την θέση του δείχνει με άσπρα έντονα γράμματα στα δεξιά του δείκτη την κατηγορία που ανηκεί το εικονοστοιχείο. Επίσης, έχουμε δώσει την εξής λειτουργία, την ρυθμιζόμενη διαφάνεια του στρώματος με τις προβλέψεις για να δώσουμε την ευχέρεια στον χρήστη να επιλέξει την κατάλληλη επιμυητή διαφάνεια ώστε να μπορεί να διαχρίνει ευκολότερα τα αντικείμενα με τις αντίστοιχες κατηγορίες που ανήκουν.

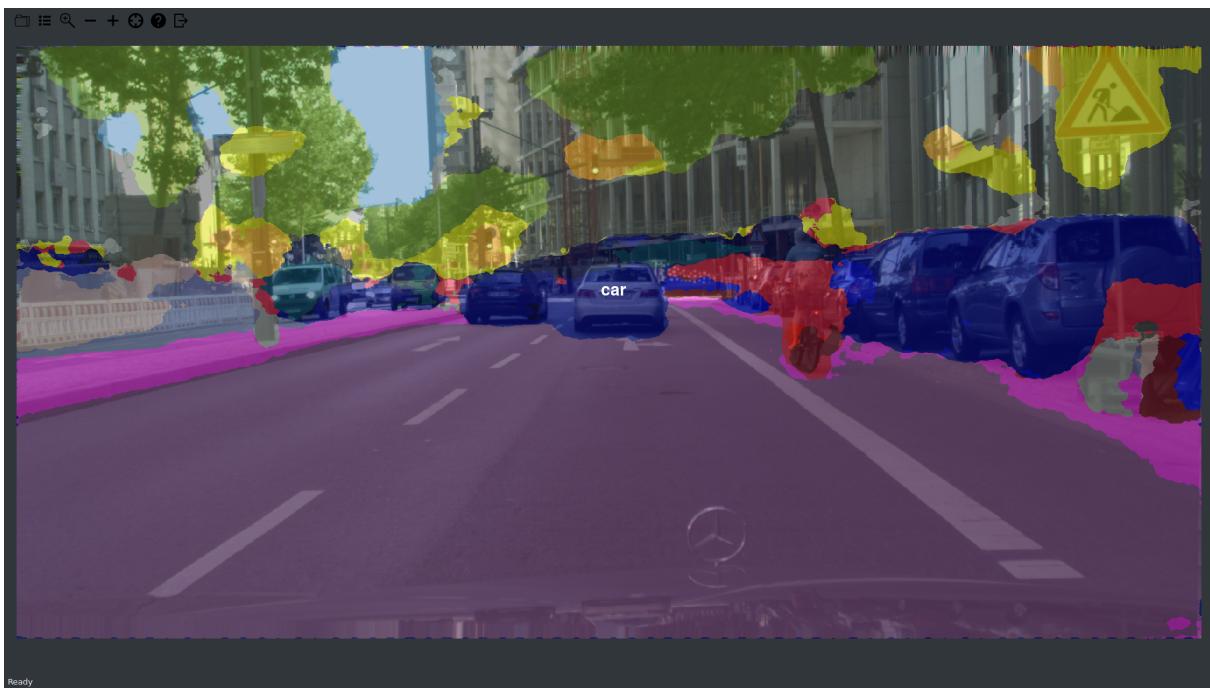


Figure 6.3: Προβολή εικόνας και πρόβλεψης του μοντέλου .

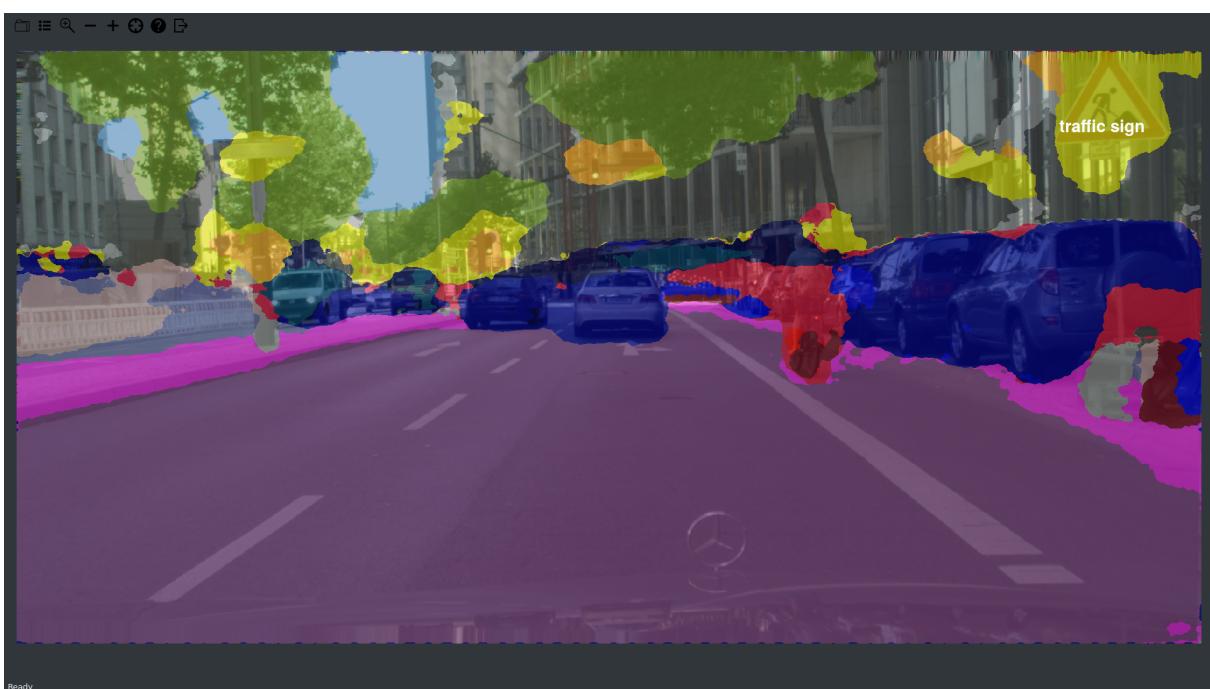


Figure 6.4: Παράδειγμα απεικόνισης ετικέτας σε επιλεγμένο βαθμό διαφάνειας.

Τέλος, υπάρχει η δυνατότητα της μεγέθυνσης συγκεκριμένης επιφάνειας της εικόνας στατικού μεγέθους ανάλογα με την θέση που βρίσκεται ο δείκτης από το ποντίκι, δείχνοντας την ετικέτα του κεντρικού εικονοστοιχείου (εικόνα 6.5).

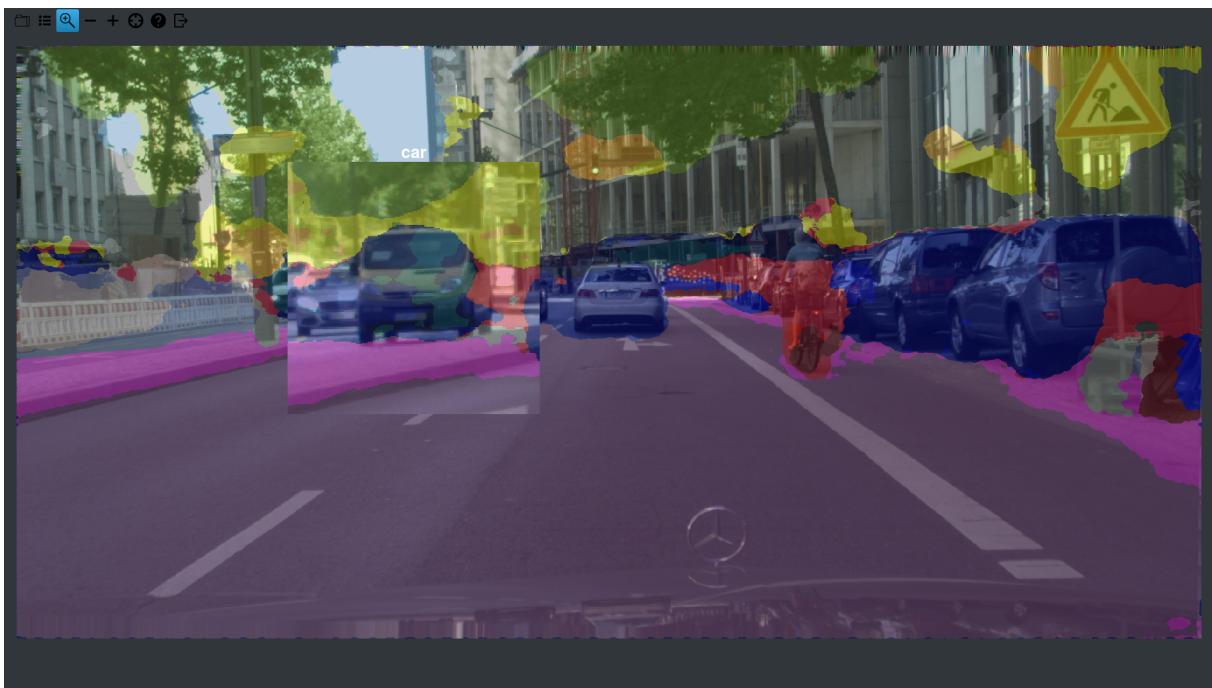


Figure 6.5: Λειτουργία μεγέθυνσης επιφάνειας.

Παράρτημα Β

Θα δείξουμε με ένα παράδειγμα την διαδικασία της συνέλιξης ως μια πράξη πίνακα, παίρνοντας σαν παράδειγμα την συνέλιξη της εικόνας 6.6. Αν ξετυλίξουμε την είσοδο και την έξοδο σε ένα διάνυσμα από αριστερά προς τα δεξιά και από πάνω προς τα κάτω, η συνέλιξη μπορεί να αναπαρασταθεί ως έναν αραιό πίνακα \mathbf{C} (6.1), όπου τα μη-μηδενικά στοιχεία είναι τα στοιχεία $w_{i,j}$ του πυρήνα, όπου i και j είναι είναι η γραμμή και η στήλη αντίστοιχα.

$$\begin{pmatrix} w_{0,0} & w_{0,1} & w_{0,2} & 0 & w_{1,0} & w_{1,1} & w_{1,2} & 0 & w_{2,0} & w_{2,1} & w_{2,2} & 0 & 0 & 0 & 0 \\ 0 & w_{0,0} & w_{0,1} & w_{0,2} & 0 & w_{1,0} & w_{1,1} & w_{1,2} & 0 & w_{2,0} & w_{2,1} & w_{2,2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & w_{0,0} & w_{0,1} & w_{0,2} & 0 & w_{1,0} & w_{1,1} & w_{1,2} & 0 & w_{2,0} & w_{2,1} & w_{2,2} & 0 \\ 0 & 0 & 0 & 0 & 0 & w_{0,0} & w_{0,1} & w_{0,2} & 0 & w_{1,0} & w_{1,1} & w_{1,2} & 0 & w_{2,0} & w_{2,1} & w_{2,2} \end{pmatrix} \quad (6.1)$$

Αυτή η γραμμική πράξη δέχεται σαν είσοδο έναν πίνακα ο οποίος έχει μετασχηματιστεί σε διάνυσμα 16-Διαστάσεων και παράγει έναν 4 διαστάσεων διάνυσμα το οποίο αργότερα μετασχηματίζεται σε έναν 2×2 πίνακα εξόδου. Χρησιμοποιώντας αυτήν την αναπαράσταση, η διαδικασία της οπισθοδόμησης επιτυγχάνεται εύκολα αναστρέφοντας τον πίνακα \mathbf{C} . Με άλλα λόγια, η διαφορά του σφάλματος διαδίδεται προς τα πίσω πολλαπλασιάζοντας το με τον πίνακα \mathbf{C}^T .

Αυτή η πράξη παίρνει σαν είσοδο ένα διάνυσμα τεσσάρων διαστάσεων και παράγει ένα διάνυσμα 16 διαστάσεων, ενώ η διασύνδεση των νευρώνων παραμένει συμβατή με τον πίνακα \mathbf{C} . Προφανώς, ο πυρήνας w παραμένει ίδιος και εξ' ορισμού και για τους δυό πίνακες \mathbf{C} και \mathbf{C}^T .

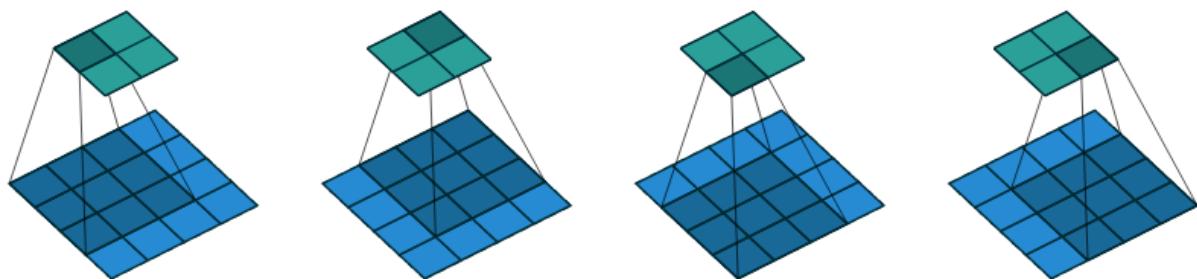


Figure 6.6: Συνέλιξη ενός πυρήνα μεγέθους 3×3 πάνω από μία είσοδο ενός χάρτη χαρακτηριστικών μεγέθους 4×4 χωρίς γέμισμα μηδενικών περιφερειακά της εισόδου και με μοναδιαίο βήμα ολίσθησης.

Bibliography

- [1] Selu make fnns great again (snn). URL <https://towardsdatascience.com/selu-make-fnns-great-again-snn-8d61526802a9>.
- [2] Biliner interpolation. URL https://en.wikipedia.org/wiki/Bilinear_interpolation.
- [3] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.
- [4] A. Adams, J. Baek, and M. A. Davis. Fast High-Dimensional Filtering Using the Permutohedral Lattice. *Computer Graphics Forum*, 2010. ISSN 1467-8659. doi: 10.1111/j.1467-8659.2009.01645.x.
- [5] V. Badrinarayanan, A. Kendall, and R. Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *CoRR*, abs/1511.00561, 2015.
- [6] S. Brodeur, E. Perez, A. Anand, F. Golemo, L. Celotti, F. Strub, J. Rouat, H. Larochelle, and A. C. Courville. Home: a household multimodal environment. *CoRR*, abs/1711.11017, 2017. URL <http://arxiv.org/abs/1711.11017>.
- [7] L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, R. Layton, J. VanderPlas, A. Joly, B. Holt, and G. Varoquaux. API design for machine learning software: experiences from the scikit-learn project. *CoRR*, abs/1309.0238, 2013.
- [8] Z. Che, Y. Cheng, S. Zhai, Z. Sun, and Y. Liu. Boosting deep learning risk prediction with generative adversarial networks for electronic health records. *CoRR*, abs/1709.01648, 2017.
- [9] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *CoRR*, abs/1606.00915, 2016.
- [10] L. Chen, G. Papandreou, F. Schroff, and H. Adam. Rethinking atrous convolution for semantic image segmentation. *CoRR*, abs/1706.05587, 2017.
- [11] F. Chollet et al. Keras, 2015.

-
- [12] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele. The cityscapes dataset for semantic urban scene understanding. *CoRR*, abs/1604.01685, 2016.
 - [13] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*, 2009.
 - [14] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification (2nd Ed)*. Wiley, 2001.
 - [15] V. Dumoulin and F. Visin. A guide to convolution arithmetic for deep learning. *CoRR*, abs/1603.07285, 2016. URL <http://dblp.uni-trier.de/db/journals/corr/corr1603.html#DumoulinV16>.
 - [16] N. S. Geoffrey Hinton and K. Swersky. Neural networks for machine learning.
 - [17] A. Graves. *Supervised Sequence Labelling with Recurrent Neural Networks*. 2011.
 - [18] M. Hubel and T. N. Wiesel. *Brain and Visual Perception*. Oxford Univeristy Press, 2005.
 - [19] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *ICML*, volume 37 of *JMLR Workshop and Conference Proceedings*, pages 448–456. JMLR.org, 2015.
 - [20] *The OpenCV Reference Manual*. Itseez, 2.4.9.0 edition, April 2014.
 - [21] Itseez. Open source computer vision library.
<https://github.com/itseez/opencv>, 2015.
 - [22] P. A. Jadhav, P. N. Chatur, and K. P. Wagh. Integrating performance of web search engine with machine learning approach. In *2016 2nd International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB)*, pages 519–524, Feb 2016. doi: 10.1109/AEEICB.2016.7538344.
 - [23] X. Jin, X. Li, H. Xiao, X. Shen, Z. Lin, J. Yang, Y. Chen, J. Dong, L. Liu, Z. Jie, J. Feng, and S. Yan. Video scene parsing with predictive feature learning. *CoRR*, abs/1612.00119, 2016.
 - [24] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014.
 - [25] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter. Self-normalizing neural networks. In *NIPS*, pages 972–981, 2017.
 - [26] P. Krähenbühl and V. Koltun. Efficient inference in fully connected crfs with gaussian edge potentials. *CoRR*, abs/1210.5644, 2012.
 - [27] A. Krizhevsky, I. Sutskever, and G. 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. URL <http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>.

-
- [28] J. Lafferty. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. pages 282–289. Morgan Kaufmann, 2001.
 - [29] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1:541–551, 1989.
 - [30] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, Nov 1998. ISSN 0018-9219. doi: 10.1109/5.726791.
 - [31] G. Lin, A. Milan, C. Shen, and I. Reid. Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5168–5177, July 2017. doi: 10.1109/CVPR.2017.549.
 - [32] M. C. Mozer. A focused backpropagation algorithm for temporal pattern recognition. *Complex Systems*, 3:349–381, 1989.
 - [33] H. Noh, S. Hong, and B. Han. Learning deconvolution network for semantic segmentation. In *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, ICCV ’15, pages 1520–1528, Washington, DC, USA, 2015. IEEE Computer Society. ISBN 978-1-4673-8391-2. doi: 10.1109/ICCV.2015.178.
 - [34] PyQt. Pyqt reference guide. 2012.
 - [35] J. Redmon, S. K. Divvala, R. B. Girshick, and A. Farhadi. You only look once: Unified, real-time object detection. *CoRR*, abs/1506.02640, 2015.
 - [36] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Parallel distributed processing: Explorations in the microstructure of cognition, vol. 1. chapter Learning Internal Representations by Error Propagation, pages 318–362. MIT Press, Cambridge, MA, USA, 1986. ISBN 0-262-68053-X. URL <http://dl.acm.org/citation.cfm?id=104279.104293>.
 - [37] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun. Overfeat: Integrated recognition, localization and detection using convolutional networks. *CoRR*, abs/1312.6229, 2013.
 - [38] E. Shelhamer, J. Long, and T. Darrell. Fully convolutional networks for semantic segmentation. *CoRR*, abs/1605.06211, 2016.
 - [39] M. Suarez-Alvarez, D. Pham, M. Prostov, and Y. I. Prostov. Statistical approach to normalization of feature vectors and clustering of mixed datasets. 468: 2630–2651, 09 2012.
 - [40] R. Szeliski. Computer vision algorithms and applications, 2011.
 - [41] T. Tieleman and G. Hinton. Lecture 6.5—RmsProp: Divide the gradient by a running average of its recent magnitude. COURSERA: Neural Networks for Machine Learning, 2012.
 - [42] A. Valada, J. Vertens, A. Dhall, and W. Burgard. Adapnet: Adaptive semantic segmentation in adverse environmental conditions. In *ICRA*, pages 4644–4651. IEEE, 2017.
-

-
- [43] C. Wachinger, M. Reuter, and T. Klein. Deepnat: Deep convolutional neural network for segmenting neuroanatomy. *CoRR*, abs/1702.08192, 2017. URL <http://arxiv.org/abs/1702.08192>.
 - [44] P. Wang, P. Chen, Y. Yuan, D. Liu, Z. Huang, X. Hou, and G. W. Cottrell. Understanding convolution for semantic segmentation. *CoRR*, abs/1702.08502, 2017.
 - [45] Y. Wu, M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, J. Klingner, A. Shah, M. Johnson, X. Liu, L. Kaiser, S. Gouws, Y. Kato, T. Kudo, H. Kazawa, K. Stevens, G. Kurian, N. Patil, W. Wang, C. Young, J. Smith, J. Riesa, A. Rudnick, O. Vinyals, G. Corrado, M. Hughes, and J. Dean. Google’s neural machine translation system: Bridging the gap between human and machine translation. *CoRR*, abs/1609.08144, 2016.
 - [46] T. Zhang. Solving large scale linear prediction problems using stochastic gradient descent algorithms. In *Proceedings of the Twenty-first International Conference on Machine Learning*, ICML ’04, pages 116–, New York, NY, USA, 2004. ACM. ISBN 1-58113-838-5. doi: 10.1145/1015330.1015332. URL <http://doi.acm.org/10.1145/1015330.1015332>.
 - [47] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia. Pyramid scene parsing network. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6230–6239, July 2017. doi: 10.1109/CVPR.2017.660.
 - [48] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. H. S. Torr. Conditional random fields as recurrent neural networks. *CoRR*, abs/1502.03240, 2015.