

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

June 27, 2018



# ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

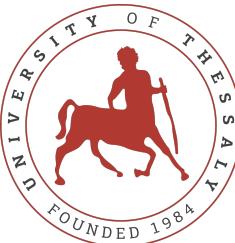
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ  
ΜΗΧΑΝΙΚΩΝ Η/Υ

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 19 κατηγορίες. Στην εργασία χρησιμοποιείται η βάση δεδομένων Cityscapes που αποτελείται από 19 διαφορετικές κατηγορίες αντικειμένων η οποία έχει δημιουργηθεί με χρήση κάμερας τοποθετημένη στο εμπρόσθιο μέρος 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 των τιμών των εικονοστοιχείων αλλά και της δομής που απεικονίζουν. Για την ταξινόμηση χρησιμοποιήσαμε 2 πανομοιότυπες αρχιτεκτονικές πλήρως συνελικτικών νευρωνικών δικτύων (FCNNs) και 2 διαφορετικές μονάδες μετα-επεξεργασίας.

Στόχος της εργασίας ήταν η δημιουργία διαφορετικών ταξινομητών καθώς και η σύγκριση μεταξύ των μεθόδων, αλλά και η δημιουργία λογισμικού για την οπτικοποίηση των αποτελεσμάτων. Για την οπτικοποίηση 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 οπτικοποίησης 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 19 various classes. The Cityscapes database introduced in CVPR 2016, consists of 19 various classes of objects created using a camera mounted on automobiles. Images have been recorded in 50 European cities in different seasons and weather conditions.

Using information from coloured images, a classifier was implemented to recognize the category of objects where each individual pixel belongs to. As part of classification, two different Fully Convolutional Neural Networks models along with another two post processing units were implemented.

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

# *Eυχαριστίες*

Θα ήθελα να ευχαριστήσω τους επιβλέποντες καθηγητές μου, καθηγητές Γεράσιμο Ποταμιάνο και Αντώνιο Αργυρίου 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 συνεπιβλέποντα ερευνητή Θεόδωρο Γιαννακόπουλο από το Ε.Κ.Ε.Φ.Ε Δημόκριτος για την υποστήριξη και τους ανθρώπους από το εργαστήριο Υπολογιστικής Ευφυΐας που μου έδωσαν χώρο και πόρους για να υλοποιηθεί αυτή η διπλωματική.

# Κατάλογος Εικόνων

1.1	Αναγνώριση YOLO	15
1.2	Παράδειγμα Σημασιολογίας	15
1.3	Παράδειγμα Εικόνων Βάσης	16
2.1	Brain Stimulus	20
2.2	Νευρωνικό Δίκτυο	21
2.3	Επίπεδο Συνέλιξης	22
2.4	ΣΝΔ Lenet-5	22
2.5	ΣΝΔ Alex-Net	23
2.6	Fully-CNN	24
3.1	Παράλληλο ΣΝΔ	29
3.2	SELU Function	32
3.3	Συναρτήσεις Ενεργοποίησης	33
3.4	Ενέργεια Συνάρτησης Κόστους	33
3.5	Τμήμα Συνέλιξης	36
3.6	Συνέλιξη με Zero-Padding	37
3.7	Στάδιο Κωδικοποίησης	38
3.8	Παράδειγμα Μέγιστη Συγκέντρωσης	38
3.9	Παράλληλη Μονάδα Επεξεργασίας	39
3.10	Διεσταλμένη Συνέλιξη	40
3.11	Στάδιο Αποκωδικοποίησης με βηματισμό	40
3.12	Διγραμμική Μονάδα Αποκωδικοποίησης	42

3.13 Αρχιτεκτονικές ΣΝΔ	43
4.1 MeanField as CNN	47
4.2 CRF-RNN Network	50
4.3 CNN CRF-RNN Network	51
5.1 Εικόνες από ΠΣΝΔ-ΤΥΣΠΙ-ΕΝΔ	58
5.2 Εικόνες από ΠΣΝΔ	59
5.3 Εικόνες Διγραμμικών ΠΣΝΔ	60
5.4 Πίνακες Σύγχυσης χωρίς μετα-επεξεργασία	62
5.5 Πίνακες Σύγχυσης με ΤΥΣΠΙ-ΕΝΔ	63
6.1 Επιλογή Εικόνας	67
6.2 Επιλογή Αρχείου	67
6.3 Προβολή Ετικέτας	68
6.4 Παράδειγμα Διαφάνειας	68
6.5 Παράδειγμα της Λειτουργίας Μεγέθυνσης	69

# Κατάλογος Πινάκων

5.1	Αποτελέσματα Μέσου Φίλτρου . . . . .	54
5.2	Αποτελέσματα CNN-CRF . . . . .	55
5.3	Σύγχριση με σύγχρονα μοντέλα . . . . .	55
5.4	Αποτέλεσμα μοντέλου ανά κατηγορία . . . . .	56
5.5	Αποτέλεσμα ανά υπερ-κατηγορία . . . . .	56
6.1	Πίνακας Χρωμάτων των Κλάσεων . . . . .	66

# Περιεχόμενα

<b>Κατάλογος Εικόνων</b>	<b>6</b>
<b>Κατάλογος Πινάκων</b>	<b>8</b>
<b>1 Εισαγωγή</b>	<b>14</b>
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 Διπλωματικής . . . . .	17
1.4.2 Συνεισφορά της Διπλωματικής . . . . .	18
1.4.3 Δομή της Διπλωματικής . . . . .	18
1.5 Συναφείς Εργασίες . . . . .	19
<b>2 Νευρωνικά Δίκτυα και Βαθιά Μάθηση</b>	<b>20</b>
2.1 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 . . . . .	26
<b>3 Μεθοδολογία</b>	<b>28</b>
3.1 Εισαγωγή . . . . .	28

3.2	Πρώτη Προσέγγιση . . . . .	28
3.3	Προετοιμασία Δεδομένων . . . . .	29
3.3.1	Τυποδειγματοληψία . . . . .	29
3.3.2	Κανονικοποίηση Χαρακτηριστικών . . . . .	30
3.3.3	Διυσαναλογία των Κλάσεων . . . . .	30
3.3.4	Επισκόπηση Αρχιτεκτονικής . . . . .	31
3.3.5	Στάδιο Κωδικοποίησης . . . . .	36
3.3.6	Μονάδα Παράλληλης Επεξεργασίας Χαρακτηριστικών . . . . .	38
3.3.7	Στάδια Αποκωδικοποίησης . . . . .	40
3.3.8	Ολοκληρωμένες Αρχιτεκτονικές . . . . .	43
<b>4</b>	<b>Μονάδες Μετα-Επεξεργασίας</b>	<b>44</b>
4.1	Επισκόπηση . . . . .	44
4.2	Μεσαίο Φίλτρο . . . . .	44
4.3	Τυχαία υπό Συνθήκη Πεδία (CRF) . . . . .	45
4.3.1	Επισκόπηση Αλγορίθμου . . . . .	45
4.3.2	Αρχικοποίηση . . . . .	47
4.3.3	Πέρασμα Μηνυμάτων . . . . .	47
4.3.4	Στάθμιση Εξόδου Φίλτρου . . . . .	48
4.3.5	Μετασχηματισμός Συμβατότητας . . . . .	48
4.3.6	Κανονικοποίηση . . . . .	49
4.3.7	Τυχαία υπό Συνθήκη Πεδία ως Επαναλαμβανόμενα Νευρωνικά Δίκτυα (CRF as RNN) . . . . .	49
<b>5</b>	<b>Πειράματα και Αποτελέσματα</b>	<b>52</b>
5.1	Εκπαίδευση των Νευρωνικών Δικτύων . . . . .	52
5.1.1	Σημεία Ελέγχου (Checkpoints) . . . . .	52
5.1.2	Πρώιμο Σταμάτημα (Early Stopping) . . . . .	53
5.1.3	Ρυθμός Μάθησης . . . . .	53
5.2	Αποτελέσματα . . . . .	53

<b>6 Συμπεράσματα και Μελλοντική Εργασία</b>	<b>64</b>
6.1 Συμπεράσματα . . . . .	64
6.2 Μελλοντική Εργασία . . . . .	65
<b>Παράρτημα Α</b>	
<b>Βιβλιογραφία</b>	<b>70</b>

*Στην Οικογένειά μου...*

# Ακρωνύμια

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

# Κεφάλαιο 1

## Εισαγωγή

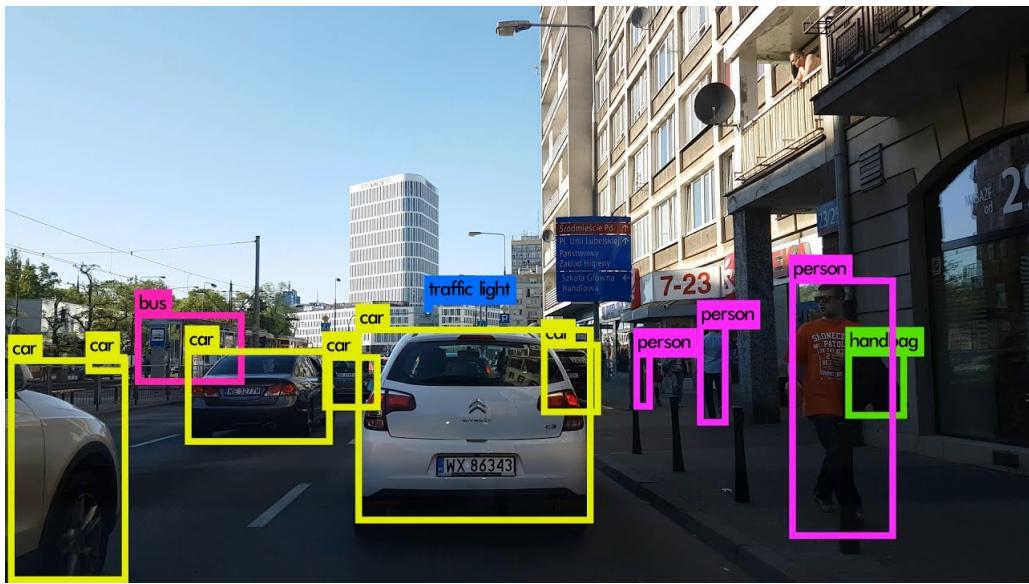
### 1.1 Μηχανική Μάθηση και Σημασιολογική Κατάτμηση

Μηχανική Μάθηση είναι ένας τομέας ο οποίος ανήκει στην Επιστήμη των Υπολογιστών ο οποίος επικεντρώνεται σε εκλεπτυσμένους αλγορίθμους οι οποίοι δεν έχουν δημιουργηθεί 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 η εξέλιξη της υπολογιστικής δύναμης καθώς και ο μεγάλος όγκος δεδομένων που είναι διαθέσιμος επιτρέπει στους επιστήμονες να πειραματιστούν με πιο πολύπλοκους αλγόριθμους. Ο συγκεκριμένος κλάδος καλύπτει ένα μεγάλο εύρος εφαρμογών, από μηχανές αναζήτησης [20] και μετάφραση κειμένου [48] μέχρι εκτιμήσεις για ασθένειες στον κλάδο της Ιατρικής [7].

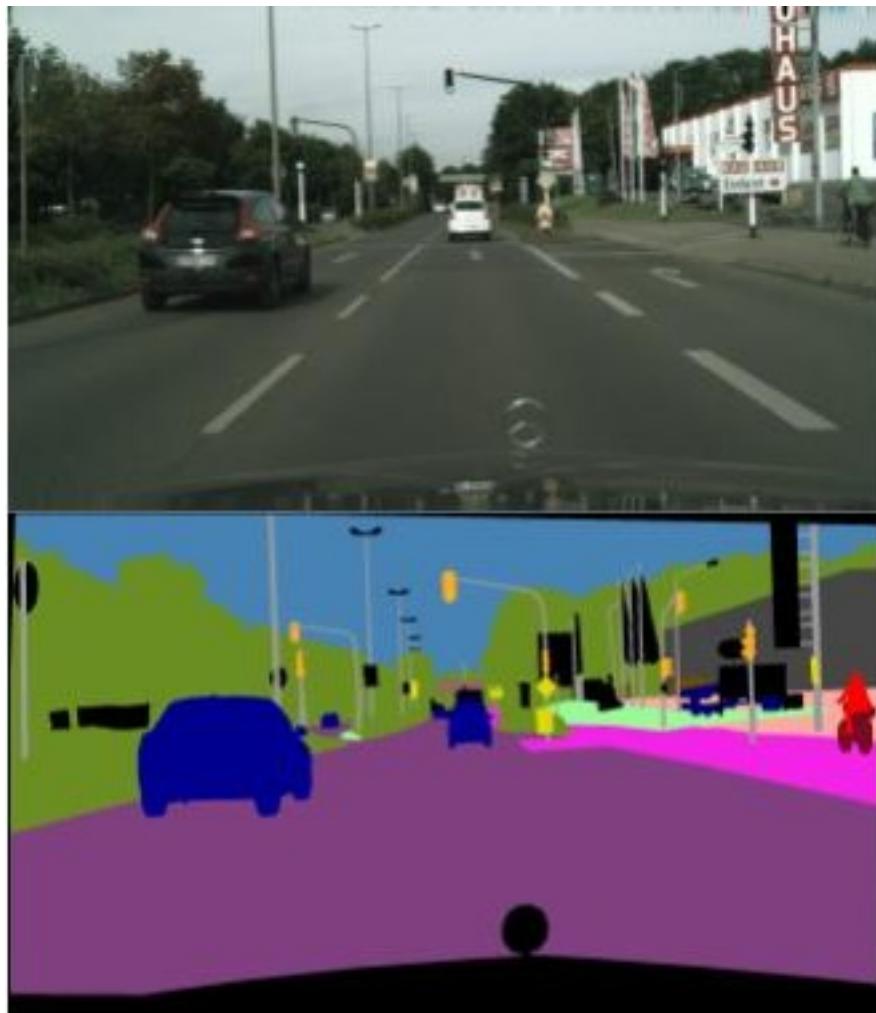
### 1.2 Σημασιολογική Κατάτμηση και Αναγνώριση Αντικειμένων

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 στον ίδιο στόχο, έχουν μια πολύ σημαντική διαφορά. Όταν μιλάμε για αναγνώριση αντικειμένων δεν αναφερόμαστε στην ακριβή εύρεση της τοποθεσίας ενός αντικειμένου στην εικόνα αλλά στη γενική του μορφή, δηλαδή χωρίς την ακριβή εύρεση των ορίων του αντικειμένου (εικόνα 1.1). Εν αντιθέσει, στο θέμα της Σημασιολογικής Κατάτμησης αντικειμένων, μας ενδιαφέρει η τοποθεσία ενός αντικειμένου στην εικόνα αλλά και η εύρεση των ακριβών ορίων του αντικειμένου, καθώς τοποθετούμε κάθε εικονοστοιχείο της εικόνας σε μια κατηγορία αντικειμένου (εικόνα 1.2).

Στις παρακάτω εικόνες φαίνονται ξεκάθαρα οι διαφορές των 2 προβλημάτων.



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



Εικόνα 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 \[11\]](#) η οποία αποτελείται από ένα σύνολο έγχρωμων εικόνων υψηλής ευκρίνειας τραβηγμένες σε αστικές περιοχές. Περιλαμβάνει 19 κατηγορίες αντικειμένων ενώ διαθέτει και μία επιπλέον κατηγορία για την ταξινόμηση των αντικειμένων που δεν ανήκουν σε καμία κατηγορία. Τα αντικείμενα στις εικόνες έχουν επισημειωθεί σε επίπεδο εικονοστοιχείου. Δηλαδή, όλα τα εικονοστοιχεία της εικόνας ανήκουν σε κάποιο αντικείμενο. Η βάση περιέχει περισσότερες από 5000 εικόνες μεγέθους 1024x2048 εικονοστοιχείων από 50 πόλεις της Ευρώπης. Οι εικόνες έχουν τραβηγχτεί από μια κάμερα ακολουθιακά με διάστημα ενός δευτερολέπτου μεταξύ τους.

Το σύνολο δεδομένων διαθέτει επίσης ένα άλλο κομμάτι το οποίο αποτελείται από 20000 εικόνες των οποίων τα εικονοστοιχεία έχουν επισημειωθεί πιο αφηρημένα. Οι παρακάτω εικόνες δείχνουν τη διαφορά μεταξύ των δύο συνόλων, οι μερικώς επισημειωμένες εικόνες δεν έχουν όλα τα εικονοστοιχεία ταξινομημένα. Το συγκεκριμένο σύνολο από εικόνες δεν έχει χρησιμοποιηθεί στα πειράματά μας.



Εικόνα 1.3: Αριστερά: Πλήρως επισημειωμένες εικόνες.

Εικόνα Δεξιά: Μερικώς επισημειωμένες εικόνες.

### 1.3.1 Περιγραφή Αντικειμένων

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

Classes	Κλάσεις	Categories	Κατηγορίες
Road	Δρόμος	Flat	Επίπεδο
Sidewalk	Πεζοδρόμιο	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	Άνθρωπος
Rider	Αναβάτης	Human	Άνθρωπος
Car	Αυτοκίνητο	Vehicle	Όχημα
Truck	Φορτηγό	Vehicle	Όχημα
Bus	Λεωφορείο	Vehicle	Όχημα
Train	Τρένο	Vehicle	Όχημα
Motorcycle	Μοτοσυκλέτα	Vehicle	Όχημα
Bicycle	Ποδήλατο	Vehicle	Όχημα
Unlabeled	Χωρίς Ετικέτα	Void	Κενό

## 1.4 Με μια Ματιά

### 1.4.1 Στόχοι Διπλωματικής

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 σύγχριση 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 Διπλωματικής

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 οπτικοποίηση των αποτελεσμάτων προχωρήσαμε στην υλοποίηση της πλατφόρμας που μας δείχνει διαισθητικά τα αποτελέσματα των μεθόδων.

Εν ολίγοις οι συνεισφορές της 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 σύγχρονων εργαλείων Keras, Tensorflow, OpenCV και PyQt.

### 1.4.3 Δομή της Διπλωματικής

Η Διπλωματική αποτελείται από 5 κεφάλαια, όπου το καθένα επικεντρώνεται σε μία συγκεκριμένη πτυχή του προβλήματος. Πιο συγκεκριμένα, το:

- **ΚΕΦΑΛΑΙΟ 2** Περιέχει μια εισαγωγή στην 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 τα κεφάλαια 3 και 4.
- **ΚΕΦΑΛΑΙΟ 6** Περιέχει θέματα για συζήτηση πάνω στα αποτελέσματα καθώς και μελλοντικές κατευθύνσεις έρευνας.

---

## 1.5 Συναφείς Εργασίες

Η βάση δεδομένων Cityscapes ολοκληρώθηκε και παρουσιάστηκε το 2016. Οι ομάδες που έχουν δημοσιεύσει μέχρι τώρα τα αποτελέσματά τους στην ιστοσελίδα της βάσης χρησιμοποιούν κυρίως τις λεπτομερώς επισημειωμένες εικόνες σε συνδυασμό με τις μερικώς επισημειωμένες εικόνες. Ωστόσο, υπάρχουν και ομάδες οι οποίες έχουν χρησιμοποιήσει δύο ξεχωριστά μοντέλα τα οποία δέχονται σαν είσοδο έγχρωμες εικόνες και εικόνες με πληροφορία βάθους αντίστοιχα [45]. 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 [21, 30]. Σε άλλες εργασίες όπως [8, 9, 47] χρησιμοποιήθηκε παράλληλη μονάδα επεξεργασίας για λήψη πολλαπλών χαρακτηριστικών από διαφορετικά οπτικά πεδία στην εικόνα. Με αυτόν τον τρόπο εξάγεται χρήσιμη πληροφορία από τις εικόνες, καθώς οι παράλληλες μονάδες καταφέρνουν να πάρουν πολύπλοκα χαρακτηριστικά βοηθώντας στην κατανόηση των δομών των αντικειμένων στην εικόνα. Μία από τις πρωτότυπες εργασίες η οποία πέτυχε πολύ μεγάλη ακρίβεια στην κατηγοριοποίηση των εικονοστοιχείων είναι η ομάδα που δημιούργησε το PSP-Net [51] 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.

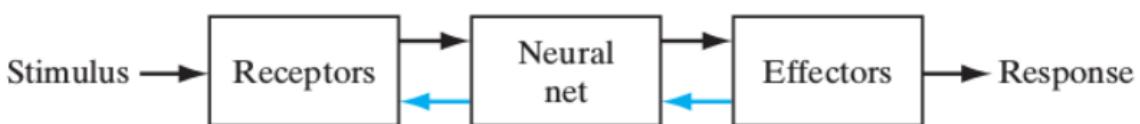
Τέλος, μια λίγο διαφορετική προσέγγιση ήταν το ΠΣΝΔ SegNet [4] το οποίο βασίστηκε σε μια αρχιτεκτονική ΠΣΝΔ κωδικοποιητή-αποκωδικοποιητή χρησιμοποιώντας μόνο επίπεδα συνέλιξης και διγραμμική παρεμβολή για την υπερδειγματοληψία των χαρακτηριστικών, ενώ στο [32] χρησιμοποιήθηκε παρόμοιο μοντέλο χρησιμοποιώντας επίπεδα αποσυνέλιξης στο στάδιο της αποκωδικοποιήσης.

# Κεφάλαιο 2

## Νευρωνικά Δίκτυα και Βαθιά Μάθηση

### 2.1 Νευρωνικά Δίκτυα

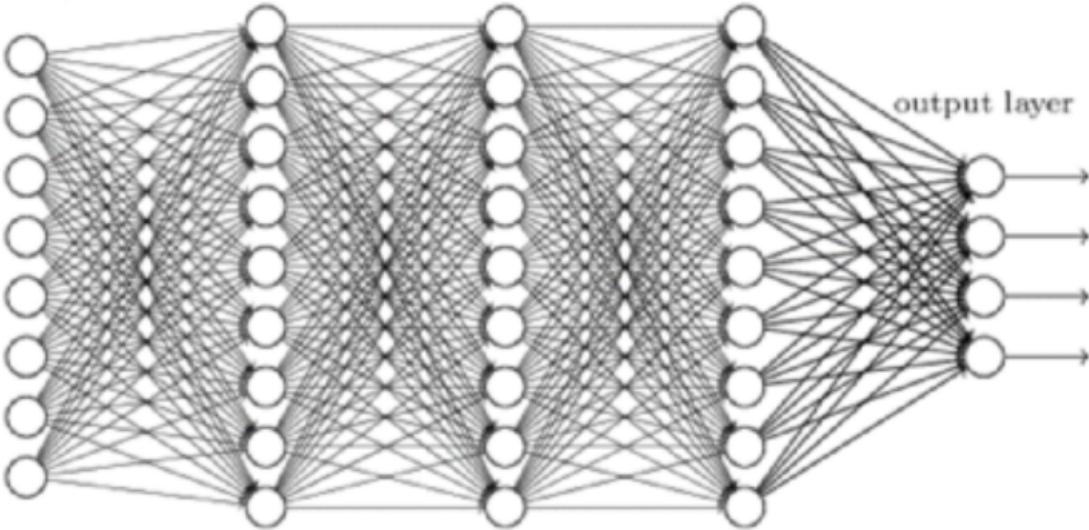
Η βασική αρχή των Νευρωνικών δικτύων ήταν η δημιουργία ενός μοντέλου το οποίο θα μπορεί να προσαρμόζεται σε δεδομένα και να αξιοποιεί την τις πληροφορίες. Η δημιουργία τους εμπνεύστηκε από την βιολογία, συγκεκριμένα, από τον τρόπο που ο εγκέφαλός μας επεξεργάζεται πληροφορίες. Σύμφωνα με το βιολογικό μοντέλο που παρουσίασαν οι H. Hubel and T. Wiesel [16], 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 ανταλλάσσουν 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.



Εικόνα 2.1: Τρόπος επικοινωνίας νευρώνων στον ανθρώπινο εγκέφαλο [13]

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

input layer    hidden layer 1    hidden layer 2    hidden layer 3



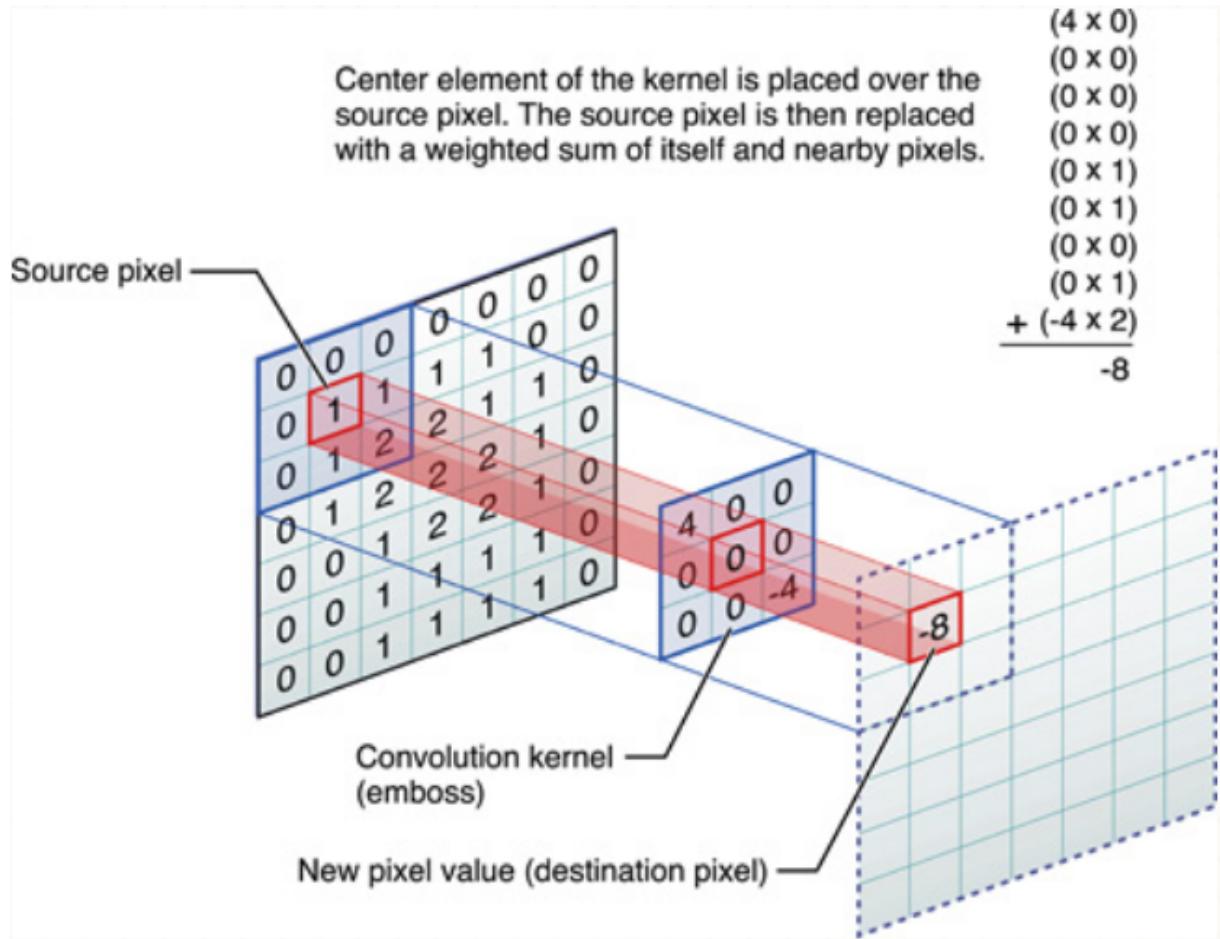
Εικόνα 2.2: **Τεχνητό Νευρωνικό Δίκτυο.** Η είσοδος αποτελείται από κόμβους οι οποίοι δέχονται ένα γεγονός. Η επεξεργασία γίνεται εσωτερικά στους εσωτερικούς κόμβους. Ο αριθμός των επιπέδων ενός Δικτύου δεν είναι προκαθορισμένος και μπορεί να περάσει από πολλά επίπεδα μέχρι να πάρουμε στην έξοδο ένα επιθυμητό αποτέλεσμα. [13]

## 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 ακολουθούμενα από ένα επίπεδο υποδειγματοληψίας ακολουθούμενο από ένα ή περισσότερα πλήρως διασυνδεδεμένα επίπεδα όπως συναντάμε και σε ένα πολυεπίπεδο Νευρωνικό Δίκτυο. Η αρχιτεκτονική του ΣΝΔ σχεδιάζεται έτσι ώστε να εκμεταλλεύεται την δισδιάστατη δομή των εικόνων εισόδου ή άλλα δισδιάστατα σήματα όπως σήματα ήχου (π.χ. Φασματόγραμμα). Αυτό επιτυγχάνεται με τοπικές συνδέσεις και κατάλληλα βάρη προκειμένου να δημιουργηθούν χαρακτηριστικά ανεξαρτήτως μετατοπίσεων (translation-invariant). Άλλο ένα πλεονέκτημα των ΣΝΔ είναι ότι είναι ευκολότερα στην εκπαίδευση και έχουν πολύ λιγότερες παραμέτρους από τα ΝΔ που έχουν πλήρως συνδεδεμένα επίπεδα.

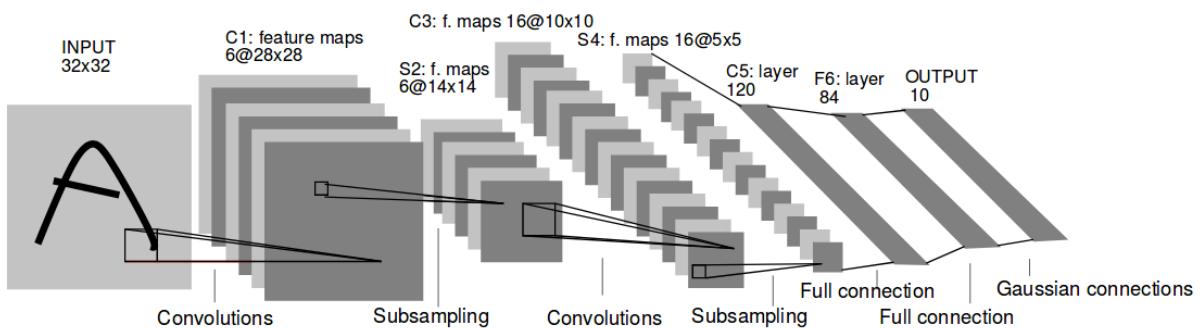
Η είσοδος σε ένα επίπεδο συνέλιξης είναι μια  $m \times n \times c$  εικόνα όπου  $m$  και  $n$  είναι το ύψος και το πλάτος της εικόνας αντίστοιχα, ενώ το  $c$  είναι ο αριθμός των καναλιών, για παράδειγμα για έγχρωμες εικόνες RGB  $c = 3$ . Το επίπεδο συνέλιξης έχει  $k$  φίλτρα (kernels) μεγέθους  $w \times w \times r$  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 φίλτρο. Το μέγεθος των φίλτρων προκαλεί τοπικά συνδεδεμένη δομή όπου το καθένα συνελίσσεται με κάθε εικόνα για να παράγουν χάρτες χαρακτηριστικών (feature maps) μεγέθους  $(m - w + 1) \times (n - w + 1)$ . 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 συνήθως ακολουθεί μια προσθήκη μίας παραμέτρου πόλωσης (bias) και μια συνάρτηση ενεργοποίησης σε κάθε χάρτη χαρακτηριστικών. Η εικόνα 2.3 μας δείχνει ένα παράδειγμα ενός επίπεδου

συνέλιξης.



Εικόνα 2.3: Επίδειξη ενός βήματος εφαρμογής ενός μικρού μεγέθους φίλτρου ( $3 \times 3$ ) σε έναν χάρτη χαρακτηριστικών εισόδου και το αποτέλεσμά της [34].

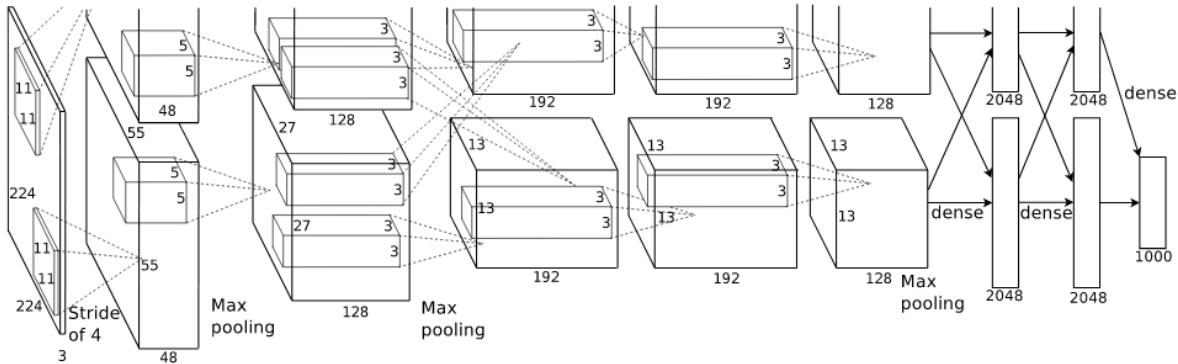
Στην εικόνα 2.4 βλέπουμε την πρώτη αρχιτεκτονική CNN από τον Yann LeCun για εφαρμογή σε προβλήματα αναγνώρισης ψηφίων.



Εικόνα 2.4: **CNN LeNet-5.** Αρχιτεκτονική του πρώτου Συνελικτικού Νευρωνικού Δικτύου για αναγνώριση ψηφίων από εικόνες. Κάθε επίπεδο αποτελεί ένα χαρτογράφημα των χαρακτηριστικών [29].

Το Alex-Net (εικόνα 2.5) ήταν μια δημιουργία των Alex Krizhevsky, Ilya Sutskever, και Geoffrey Hinton, και σηματοδότησε μια νέα εποχή στην υπολογιστική άραση καθώς

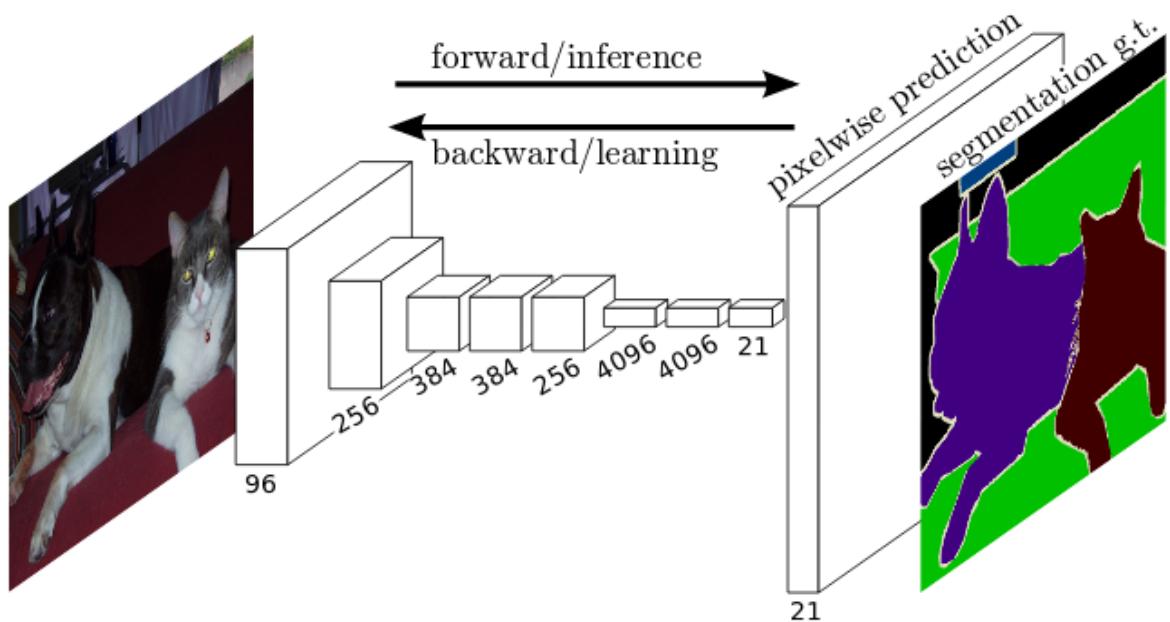
πλέον περάσαμε στα *Bαθιά ΝΔ*. Το εφάρμοσαν σε ένα από τα πιο απαιτητικά προβλήματα, το *Image-Net* [12]. Η συγκεκριμένη αρχιτεκτονική κατάφερε να πετύχει ένα σημαντικό αποτέλεσμα μειώνοντας πάνω από 10% το σφάλμα σε σχέση με τον προηγούμενο νικητή το 2012, πάνω σε ένα πρόβλημα με 15 εκατομμύρια εικόνες 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.



Εικόνα 2.5: **Alex-Net**. Αρχιτεκτονική του Alex-Net ένα από τα πρώτα Βαθιά ΝΔ με 60 εκατομμύρια παραμέτρους και 650.000 νευρώνες. [26].

## 2.3 Πλήρως Συνελικτικά Νευρωνικά Δίκτυα

Μία ακόμα κατηγορία ΝΔ με τα οποία θα ασχοληθούμε στο υπόλοιπο της εργασίας είναι τα Πλήρως ΣΝΔ (Fully-CNN) [40]. Η κύρια διαφορά με τα ΣΝΔ είναι η απώλεια πλήρως συνδεδεμένων επιπέδων στην έξοδο (fully-connected layers), εν αντιθέσει με τα ΣΝΔ που είδαμε προηγουμένως, δηλαδή τα ΠΣΝΔ μαθαίνουν πληροφορία μόνο από φίλτρα. Τα ΠΣΝΔ θεωρούνται κατάλληλα για προβλήματα Σημασιολογικής Κατάτμησης αντικειμένων από εικόνες (εικόνα 2.6).



Εικόνα 2.6: Πλήρως ΣΝΔ (Fully CNN) μπορούν να μάθουν αποδοτικά να κάνουν προβλέψεις σε προβλήματα αναγνώρισης σε επίπεδο εικονοστοιχείων [40].

Μερικά όφετα που καταστούν τα ΠΣΝΔ κατάλληλα για Σημασιολογική Κατάτμηση είναι σε αντίθεση με άλλου είδους ΝΔ είναι:

1. Χρησιμοποιούν όλη την πληροφορία της εικόνας.
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. Τα ΣΝΔ έχουν δημιουργηθεί με την προοπτική να επεξεργάζονται δεδομένα από εικόνες. Αποτελούνται από πολλά επίπεδα συνέλιξης τα οποία είναι συνδεδεμένα σειριακά για να επεξεργάζονται οπτική πληροφορία. Τα συνελικτικά επίπεδα αποτελούνται από μια σειρά από φίλτρα  $K$  και τις πολώσεις  $b$  (biases), ενώ δέχονται έναν χάρτη χαρακτηριστικών στην είσοδο  $I$ .

Στην περίπτωση που έχουμε εικόνες για αναγνώριση, η είσοδος αποτελείται από μια εικόνα με ύψος  $H$ , πλάτος  $W$  και αριθμό καναλιών  $C = 3$  (χόκινο, πράσινο και μπλε),  $I \in \mathbb{R}^{H \times W \times C}$ . Επομένως, για μια σειρά από  $D$  φίλτρα έχουμε  $K \in \mathbb{R}^{k_1 \times k_2 \times C \times D}$  και  $b \in \mathbb{R}^D$  πολώσεις, μία για κάθε φίλτρο. Για κάθε στοιχείο  $i, j$  του χάρτη χαρακτηριστικών εισόδου  $I$  εφαρμόζουμε την συνέλιξη με τον πυρήνα  $K$ :

$$(I * K)_{ij} = \sum_{m=0}^{k_1-1} \sum_{n=0}^{k_2-1} K_{m,n,c} \cdot I_{i+m,j+n,c} + b \quad (2.1)$$

Πιο αναλυτικά, η παραπάνω εξίσωση αναλύεται για κάθε επίπεδο συνέλιξης με τις εξής παραμέτρους:

1.  $l$ : Συμβολίζει το επίπεδο συνέλιξης  $l$  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  $w$  διαστάσεων  $k_1 \times k_2$  όπου έχει ως δείκτες τα  $m, n$ .
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  $l$ :

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

6.  $b^l$  είναι το διάνυσμα πόλωσης.
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  $l$ .

## 2.5 Αλγόριθμος Οπισθοδρόμησης

Ο αλγόριθμος οπισθοδρόμησης (backpropagation) [31] αποτελεί την μέθοδο με την οποία ένα NΔ επαναπροσδιορίζει τις παραμέτρους του. Αποτελεί έναν από τους πιο βασικούς αλγορίθμους για την εκπαίδευση των NΔ σε προβλήματα επιβλεπόμενης μάθησης (supervised learning). Η συνάρτηση κόστους που χρησιμοποιούμε σε αυτό το πρόβλημα όπως θα αναλύσουμε παρακάτω (παράγραφος 3.3.4) είναι η συνάρτηση διεντροπίας και είναι παραγωγίσιμη και έχει πολύ απλή παράγωγο.

Ο αλγόριθμος οπισθοδρόμησης χρησιμοποιείται για τον υπολογισμό των παραγώγων σφάλματος. Ο αλγόριθμος αποτελείται από την εφαρμογή του κανόνα της αλυσίδας για τον υπολογισμό των μερικών παραγώγων. Αρχικά, υπολογίζουμε την μερική παράγωγο της συνάρτησης κόστους ως προς την μεταβλητή εξόδου και την μερική παράγωγο της εξόδου από την softmax ( $y_i$ ) ως προς την μεταβλητή της εισόδου της μονάδας softmax ( $s_i$ ). Επομένως, χρησιμοποιώντας τον κανόνα της αλυσίδας για να υπολογίσουμε τις παραγώγους της συνάρτησης κόστους ως προς την είσοδο της μονάδας softmax. Στις εξισώσεις παρακάτω,  $y_i$  είναι η έξοδος του  $i$  στοιχείου ενώ  $s_i$  είναι η η είσοδος του  $i$  στοιχείου [38].

$$\frac{\partial E}{\partial y_i} = -\frac{t_i}{y_i} \quad (2.2)$$

$$\frac{\partial y_i}{\partial s_i} = y_i(1 - y_i) \quad (2.3)$$

$$\frac{\partial E}{\partial s_i} = \frac{\partial E}{\partial y_i} \frac{\partial y_i}{\partial s_i} = y_i - t_i \quad (2.4)$$

Εφαρμόζοντας τον κανόνα της αλυσίδας, προχωράμε προς τα κρυφά επίπεδα (hidden layers) προς τα πίσω. Για να διαδώσουμε τις παραγώγους του  $i$  στοιχείου στο  $j$  στοιχείο το οποίο ανήκει στο προηγούμενο επίπεδο, προκύπτουν οι παρακάτω εξισώσεις (2.5 και 2.6) όπου  $w_{ji}$  είναι το βάρος που ανατίθεται στην σύνδεση μεταξύ της εισόδου του  $j$  στοιχείου με την κρυφή μονάδα  $i$ . Εφαρμόζοντας επαναληπτικά τον κανόνα της αλυσίδας διαδίδουμε τις παραγώγους του σφάλματος προς την είσοδο του NΔ. Με αυτήν την μέθοδο έχουμε καταφέρει να υπολογίσουμε τις παραγώγους σφάλματος για κάθε βάρος.

$$\frac{\partial E}{\partial y_i} = \frac{\partial E}{\partial s_i} \frac{\partial s_i}{\partial y_j} = \sum_i w_{ji} \frac{\partial E}{\partial s_i} \quad (2.5)$$

$$\frac{\partial E}{\partial w_{ji}} = \frac{\partial E}{\partial s_i} \frac{\partial s_i}{\partial w_{ji}} = (y_i - t_i) \frac{\partial s_i}{\partial w_{ji}} \quad (2.6)$$

Στα επίπεδα συγκέντρωσης διαφέρει η διαδικασία της οπισθοδρόμησης ανάλογα με το είδος του επιπέδου συγκέντρωσης. Ένα επίπεδο συγκέντρωσης δεν περιέχει παραμέτρους μάθησης [28]. Τα πιο δημοφιλή επίπεδα συγκέντρωσης είναι τα επίπεδα μέγιστης και μέσης συγκέντρωσης αντίστοιχα. 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 επικρατέστερη τιμή του κάθε τμήματος.

Για να κρατήσουμε την θέση της επικρατέστερης τιμής από το επίπεδο συγκέντρωσης, σημειώνουμε την θέση κατά την εμπρόσθια διάδοση και μετά την χρησιμοποιούμε για να οδηγήσουμε τις παραγώγους του σφάλματος κατά την οπισθοδρόμηση. Η δρομολόγηση των παραγώγων επιτυγχάνεται με κάποια από τις παρακάτω μεθόδους, ανάλογα το επίπεδο συγκέντρωσης:

- **Max-Pooling** Το σφάλμα απλώς ανατίθεται στο στοιχείο το οποίο επικράτησε κατά το εμπρόσθιο πέρασμα.
- **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 κανονικοποιημένη τιμή.

# Κεφάλαιο 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 Κατάτμηση 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).

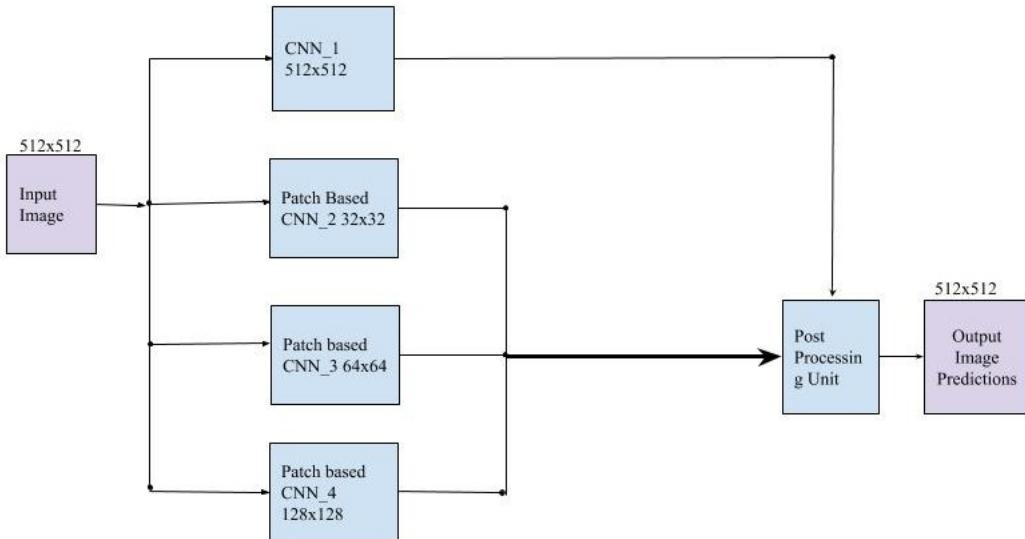
### 3.2 Πρώτη Προσέγγιση

Η πρώτη μας προσέγγιση στο πρόβλημα βασίστηκε σε μια παράλληλη αρχιτεκτονική από πολλαπλά ΣΝΔ (εικόνα 3.1). Η ιδέα στηρίχθηκε στην υλοποίηση τεσσάρων ΣΝΔ όπου τα 3 από αυτά δέχονται σαν είσοδο σειριακά κομμάτια από την εικόνα. Η μέθοδος αυτή αναφέρεται ως ‘Ολίσθηση Παραθύρων’ (*Sliding-Windows*) [39]. Τα τρία από τα τέσσερα ΣΝΔ δέχονται κομμάτια διαφορετικού μεγέθους από την εικόνα, ενώ το τέταρτο ΣΝΔ 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 αρκετές δυσκολίες:

1. Τα μικρά κομμάτια τμηματοποιούν 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.



Εικόνα 3.1: Παράλληλη αρχιτεκτονική βασισμένη σε πολλαπλά ΣΝΔ 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 παράμετροι είναι συνάρτηση της εισόδου του νευρωνικού δικτύου, επομένως η υποδειγματοληψία στις αρχικές εικόνες είναι απαραίτητη για να μπορέσουμε να κάνουμε εφικτά τα πειράματά μας. Αυτή η τεχνική φυσικά έχει κάποιο κόστος, καθώς 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 κατά την διαδικασία της υποδειγματοληψίας προσεγγίζεται από μια σταθμισμένη εκτίμηση από άλλα τέσσερα σημεία ως μια καλύτερη προσέγγιση των εικονοστοιχείων.

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.

### 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 ΝΔ καθώς μπορεί να πάρουν ανεξέλεγκτες τιμές οι νευρώνες στα hidden layers και να μην συγκλίνει το ΣΝΔ. 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 3.2 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 [41].

$$\hat{\mu} = \frac{\sum_{i=1}^N X_i}{N} \quad (3.1)$$

$$X = \frac{X - \hat{\mu}}{\max(X) - \min(X)} \quad (3.2)$$

### 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 μειονότητα 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 τεχνική είναι ανέφικτη σε ένα σύνολο από εικόνες όπου 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) [4]. Με αυτή την συνάρτηση βρίσκουμε τους συντελεστές και τους εφαρμόζουμε στην συνάρτηση κόστους (εξίσωση 3.11).

Η ιδέα είναι να βρεθούν 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 μειονότητα) 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 χλάση που είναι μειονότητα οφείλουμε να εισάγουμε μεγαλύτερη ποινή για να βοηθήσουμε στην εκμάθηση τους.

Με την εξίσωση 3.3 βρίσκουμε την συχνότητα εμφάνισης των εικονοστοιχείων για κάθε χλάση στο σύνολο δεδομένων και αφού ταξινομήσουμε τις τιμές συχνότητας παίρνουμε την μεσαία συχνότητα και την χρησιμοποιούμε σαν επίκεντρο τοποθετώντας την στον αριθμητή (εξίσωση 3.4). Με αυτή την μέθοδο πετυχαίνουμε να έχουμε υψηλούς συντελεστές  $a_i$  στα χαμηλής συχνότητας εμφάνισης εικονοστοιχεία. Με την εξίσωση 3.4 βρίσκουμε τους συντελεστές  $a_i$  για κάθε κατηγορία αντικειμένων. Ένας μεγάλος συντελεστής  $a_i$  προσθέτει μεγαλύτερη ποινή όταν ταξινομηθεί λάθος ένα εικονοστοιχείο που ανήκει σε μια χλάση που δεν υπάρχει πολλές φορές στο σύνολο δεδομένων. Τέλος, στα δεδομένα μας, υπάρχουν εικονοστοιχεία τα οποία δεν ανήκουν σε κάποια κατηγορία, για να μην μάθει το NΔ από αυτά θέσαμε τον συντελεστή  $a_i$  στο μηδέν έτσι ώστε να μην συνεισφέρουν στο σφάλμα κατά την εκπαίδευση.

$$freq(C_i) = \frac{C_i}{\sum_{i=1}^{Classes} C_i} \quad (3.3)$$

$$\alpha_i = \frac{median(freq)}{freq(C_i)} \quad (3.4)$$

### 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 εισόδου 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. Τα βάρη του ΣΝΔ δειγματοληπτήθηκαν από μία Γκαουσιανή κατανομή με μέση τιμή ίση με το μηδέν ( $\mu = 0$ ) και διασπορά ίση με ένα ( $\sigma^2 = 1/N$ ). Η εξίσωση 3.5 μας δείχνει την τυπική απόκλιση που εφαρμόστηκε στην κατανομή Γκάους (εξίσωση 3.6) για την δειγματοληψία των βαρών όπου με  $N$  συμβολίζεται το πλήθος των χαρακτηριστικών της εισόδου σε κάθε επίπεδο συνέλιξης.

$$\sigma = \sqrt{\frac{1}{N}} \quad (3.5)$$

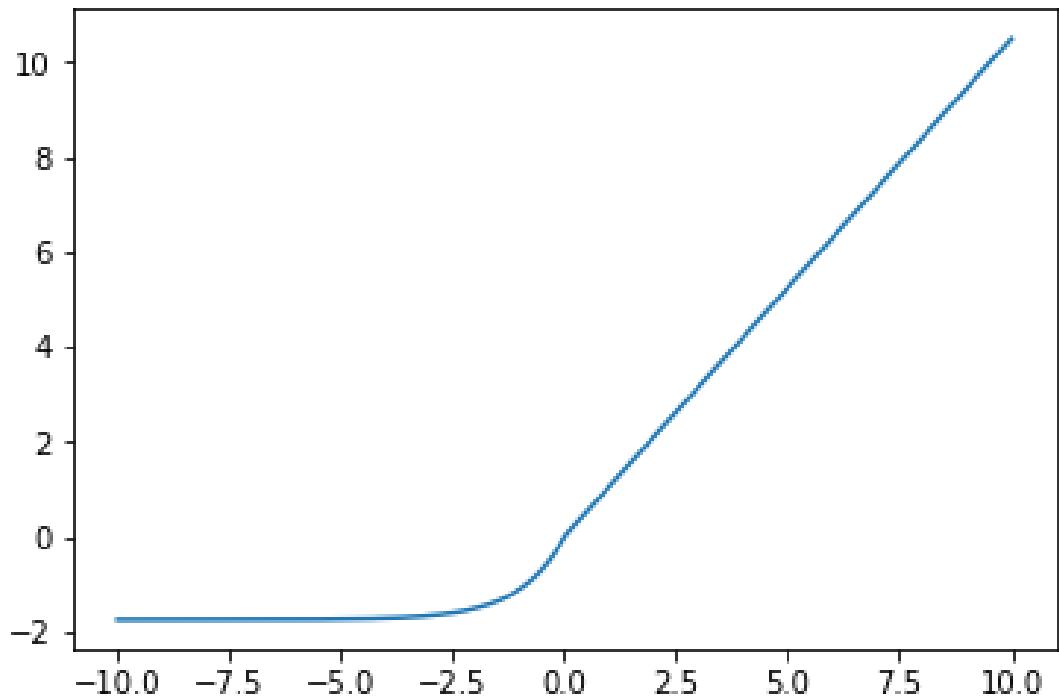
$$g(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \quad (3.6)$$

#### 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 μηνυμάτων μεταξύ νευρώνων στα επίπεδα από νευρώνες. Τα βαθειά νευρωνικά δίκτυα έρχονται αντικέτωπα με το πρόβλημα της εξαφάνισης των αποκλίσεων του σφάλματος κατά την διάδοσή τους προς τα πίσω. Για τον λόγο αυτό επινοήθηκαν συναρτήσεις που

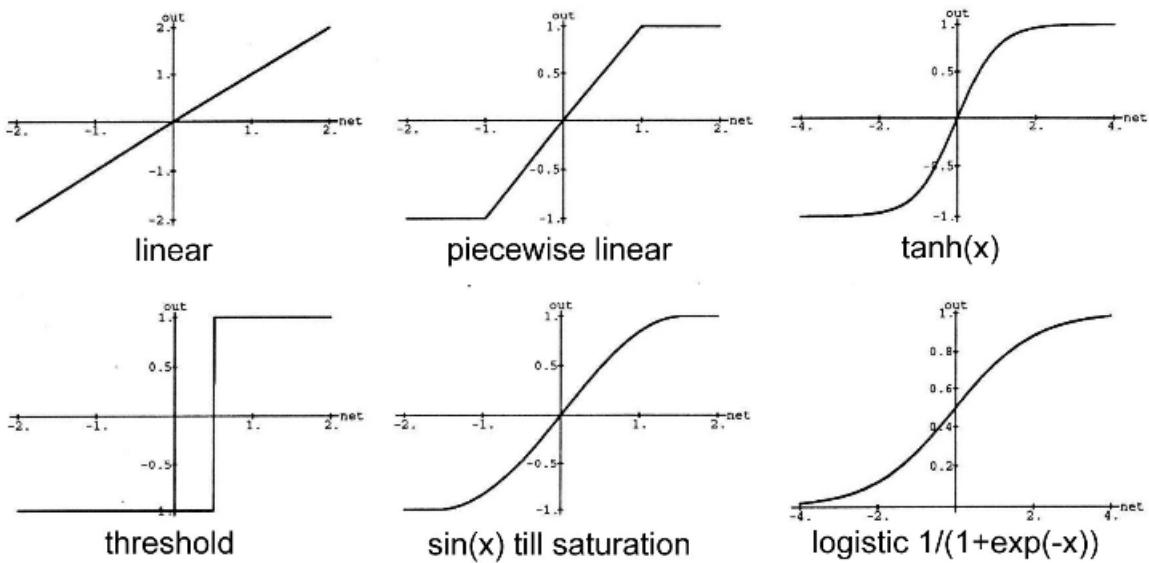
ονομάζονται Γραμμικοί Ανορθωτές (Linear Rectifiers). Οι γραμμικοί ανορθωτές συνήθως κάτω από το μηδέν έχουν μηδενική τιμή. Όταν οι αποκλίσεις πέφτουν κάτω από το μηδέν τα βάρη δεν αλλάζουν, δηλαδή οι νευρώνες μένουν απενεργοποιημένοι σε μια τέτοια περίπτωση. Το ύστερο σε αυτήν την περίπτωση είναι ότι εφόσον 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 εισάγουμε στα νευρωνικά μας την Κλιμακωτή Εκθετική Γραμμική Συνάρτηση (Scaled Exponential Linear Unit-SELU). Στην πραγματικότητα όπως απέδειξαν στο [24] η συγκεκριμένη συνάρτηση ενεργοποίησης σε συνδυασμό με την αρχικοποίηση (εξισώσεις 3.5, 3.6) όχι μόνο καταπολεμά αυτό το πρόβλημα αλλά και ιστά περιττή την εφαρμογή του αλγορίθμου Batch-Normalization [17] καθώς η κανονικοποίηση των εισόδων σε κάθε επίπεδο του νευρωνικού γίνεται μέσα σε αυτή την συνάρτηση, πετυχαίνοντας έτσι μείωση των παραμέτρων.



Εικόνα 3.2: Κλιμακωτή Εκθετική Γραμμική Συνάρτηση Ενεργοποίησης (SELU) με τις προεπιλεγμένες παραμέτρους  $\alpha = 1.6732$  και  $\lambda = 1.0507$ .

$$f(x) = \lambda \begin{cases} x & \text{if } x > 0 \\ \alpha e^x - \alpha & \text{if } x \leq 0 \end{cases} \quad (3.7)$$

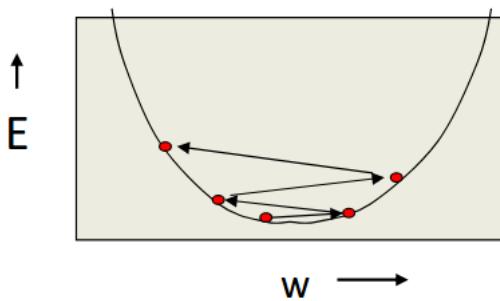


Εικόνα 3.3: Διάφορες συναρτήσεις Ενεργοποίησης [15]

### Αλγόριθμοι Βελτιστοποίησης

Ο αλγόριθμος βελτιστοποίησης αποτελεί έναν πολύ σημαντικό παράγοντα για την εκπαίδευση ενός Νευρωνικού Δικτύου. Η ανάγκη για αναζήτηση αλγορίθμων βελτιστοποίησης προήλθε από δύο σημαντικούς παράγοντες. Πρώτον, λόγω των πολλών δεδομένων για επεξεργασία και των βαθιών νευρωνικών δικτύων που έκανε την διαδικασία της μάθησης αργή. Αυτοί οι λόγοι μας ώθησαν σε τεχνικές μείωσης του σφάλματος από κοινάτια του συνόλου δεδομένων και δεύτερων για την επιτάχυνση της σύγκλισης του Νευρωνικού δικτύου προφανώς.

Παρακάτω στην εικόνα 3.4 βλέπουμε πως αν η συνάρτηση κόστους έχει την μορφή μίας χαράδρας που οδηγεί προς το βέλτιστο κόστος και έχει στα πλάγια υψηλά τοιχώματα, τότε με ένα μεγάλο ρυθμό μάθησης τα βάρη τείνουν να ταλαντεύονται μπρος και πίσω 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.



Εικόνα 3.4: Ταλάντωση κατά την εκτίμηση των βαρών όπου ο μεγάλος ρυθμός μάθησης οδηγεί σε αντίθετο αποτέλεσμα [43].

Ο πιο συνηθισμένος και βασικός αλγόριθμος βελτιστοποίησης είναι ο Στοχαστικός Αλγόριθμος Απότομης καθόδου (SGD) [50], ο οποίος υπολογίζει απλά την απόκλιση των παραμέτρων ως προς της συνάρτηση κόστους πάνω σε ένα μικρό σύνολο δειγμάτων

---

από τα δεδομένα. Πλέον υπάρχουν πιο προχωρημένοι αλγόριθμοι βελτιστοποίησης που χρησιμοποιούν περισσότερες παραμέτρους. Η επιλογή του αλγόριθμου βελτιστοποίησης γίνεται ανάλογα με την αρχιτεκτονική του Νευρωνικού Δικτύου. Η εξίσωση 3.8 μας δείχνει την εξίσωση όπου α είναι ο ρυθμός μάθησης και ο υπολογισμός της απόκλισης γίνεται πάνω σε ένα σύνολο ζευγών  $(x^i, y^j)$ .

$$\vartheta = \vartheta - \alpha \nabla_{\vartheta} \mathcal{J}(\vartheta; x^i, y^j) \quad (3.8)$$

Με την χρήση της παραμέτρους της ορμής, ο αλγόριθμος τείνει να φτάσει στο βέλτιστο σημείο πιο γρήγορα. Στην εξίσωση 3.9 ν είναι το διάνυσμα ταχύτητας το οποίο είναι φυσικά ίδιων διαστάσεων με το διάνυσμα των παραμέτρων  $\vartheta$ . Πέρα από την παράμετρο  $\alpha$  που είδαμε και προηγουμένως η οποία είναι ο ρυθμός μάθησης, παρατηρούμε και την παράμετρο  $\gamma \in [0, 1)$  η οποία ορίζει το ποσοστό συνεισφοράς των προηγούμενων αποκλίσεων στην παρούσα ανανέωση των παραμέτρων. Συνήθως αυτή η ποσότητα ορίζεται στο 0.9.

$$\begin{aligned} v &= \gamma v + \alpha \nabla_{\vartheta} \mathcal{J}(\vartheta; x^i, y^j) \\ \vartheta &= \vartheta - v \end{aligned} \quad (3.9)$$

Στην εργασία μας έγινε χρήση τόσο του αλγορίθμου της Στοχαστικής Απότομης Καθόδου, αλλά και του αλγόριθμου Adam [23] ενός πιο αποδοτικού αλγορίθμου σε θέματα στοχαστικής βελτιστοποίησης καθώς χρησιμοποιεί πρώτης τάξης παραγώγους. Ο αλγόριθμος υπολογίζει τις παραμέτρους του ρυθμού μάθησης για διάφορες παραμέτρους από εκτιμήσεις των ορμών πρώτης και δεύτερης τάξης των κλίσεων, όπως φαίνεται αναλυτικά παρακάτω. Συγκεκριμένα ο αλγόριθμος Adam είναι μια εξέλιξη του RMSProp [36, 44].

---

**Algorithm** Αλγόριθμος Adam [22]. 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:** :  $\alpha$  : Ρυθμός Μάθησης  
**Require:**  $\beta_1, \beta_2 \in [0, 1]$  : Εκθετικοί ρυθμοί καθόδου για τις εκτιμήσεις των ορμών  
**Require:**  $f(\theta)$  : Στοχαστική συνάρτηση κόστους  
**Require:**  $\theta_0$  : 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**  $\theta_t$  not converged **do**
- 5:      $t = t + 1$
- 6:
- 7:      $g_t \leftarrow \nabla_{\theta} f_t(\theta_{t-1})$  {Αποκλίσεις ως προς την συνάρτηση  $f$  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:      $v_t \leftarrow \beta_1 \cdot m_{t-1} + (1 - \beta_2) \cdot g_t^2$  {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**  $\theta_t$  {Αποτελέσματα παραμέτρων}

---

## Συνάρτηση Κόστους

Συνήθως, σε προβλήματα πολλαπλής ταξινόμησης στοιχείων, 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 αυθορίζονται στο ένα (εξίσωση 3.10).

$$softmax(z)_i = \frac{e^{z_i}}{\sum_{l=1}^L e^{z_l}} \quad (3.10)$$

Επειδή η έξοδος του ΝΔ είναι μια μονάδα softmax, πρέπει να χρησιμοποιηθεί και η κατάλληλη συνάρτηση κόστους. Η συνάρτηση κόστους μετράει την διαφορά μεταξύ της εξόδου του ΝΔ και της επιθυμητής εξόδου. Όταν η έξοδος του ΝΔ είναι μια πιθανοτική κατανομή, η πιο κατάλληλη συνάρτηση κόστους είναι η αρνητική λογαριθμική πιθανότητα της επιθυμητής εξόδου. Η συνάρτηση αυτή ονομάζεται Συνάρτηση Διεντροπίας (Cross-Entropy). Βέβαια λόγω της εισαγωγής των συντελεστών α στην συνάρτηση για την ισοστάθμιση των κλάσεων έχουμε παραμετροποιήσει την συνάρτηση καταλλήλως (εξίσωση 3.11). Με  $p_{ij}$  συμβολίζεται

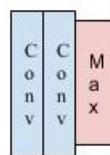
η κατανομή των πραγματικών τιμών του εικονοστοιχείου  $i$  που ανήκει σε μια κατηγορία  $j$  ενώ με  $q_{ij}$  συμβολίζεται η έξοδος που δίνει το μοντέλο για το εικονοστοιχείο  $i$  με μια πιθανοτική κατανομή πάνω στις  $L$  κατηγορίες του μοντέλου όπου  $j \in L$ . Στην πραγματικότητα, οι πραγματικές τιμές  $p_i$  ενός εικονοστοιχείου είναι ένα διάνυσμα  $L$  διαστάσεων όπου η σωστή κατηγορία που ανήκει το εικονοστοιχείο ι διαθέτει ένα στην θέση  $j$  της σωστής κατηγορίας ενώ στις υπόλοιπες θέσεις των λάθος κατηγοριών διαθέτει μηδέν. Επίσης, έχουμε προσθέσει τους συντελεστές  $\alpha_j$  οι οποίοι ρυθμίζουν την ποινή για κάθε λάθος πρόβλεψη κατηγορίας. Στο τέλος γίνεται μια κανονικοποίηση ως προς το πλήθος των εικονοστοιχείων της εικόνας  $N$ .

$$Loss = -\frac{1}{N} \sum_{i \in N} \sum_{j \in L} p_{ij} \log(q_{ij}) \alpha_j \quad (3.11)$$

Κατά την εκπαίδευση ενός  $N\Delta$ , αυτό που γίνεται είναι η βελτιστοποίηση της συνάρτησης κόστους (cross-entropy). Με αυτό τον τρόπο χρησιμοποιούμε το σφάλμα από την συνάρτηση κόστους και επαναληπτικά επαναπροσδιορίζουμε τις παραμέτρους του  $N\Delta$  με σκοπό να ελαχιστοποιήσουμε το κόστος. Η μείωση της συνάρτησης κόστους είναι ισοδύναμο με το γεγονός της αύξησης της πιθανότητας της σωστής απάντησης.

### 3.3.5 Στάδιο Κωδικοποίησης

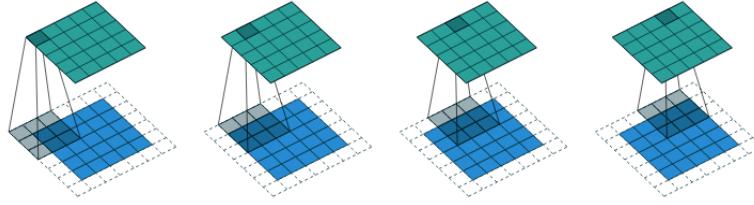
Σκοπός της μονάδας Κωδικοποίησης είναι η εξαγωγή χαρακτηριστικών από την έγχρωμη εικόνα, δηλαδή η δημιουργία μιας αναπαράστασης πολυδιάστατων χαρακτηριστικών από τα εικονοστοιχεία της εικόνας σε μια συμπιεσμένη μορφή ώστε να γίνεται εφικτή η εκπαίδευση του συστήματος. Η μονάδα κωδικοποίησης αποτελείται από 4 τμήματα και συγκεκριμένα από ομάδες συνέλιξης επιπέδων και μονάδων συγκέντρωσης. Η εικόνα 3.5 μας δίνει μια διαίσθηση του κάθε τμήματος το οποίο αποτελείται από 2 επίπεδα συνέλιξης ακολουθούμενα από ένα επίπεδο συγκέντρωσης μέγιστων τιμών ανά περιοχή (Max-Pooling).



Εικόνα 3.5: 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 αναλλοίωτο κρατώντας την θέση των χαρακτηριστικών αλλά και επειδή χρειαζόμαστε την πληροφορία από τις γωνίες των χαρακτηριστικών. Η εικόνα 3.6 μας δείχνει ένα παράδειγμα της διαδικασίας, ενώ η εξίσωση 3.12 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 δεν λαμβάνεται υπόψη στην εξίσωση. Με το  $p$  συμβολίζεται το γέμισμα των μηδενικών περιφερειακά της εισόδου όπου  $p = \lfloor k/2 \rfloor = n$ .

$$\begin{aligned}
output &= input + 2\lfloor k/2 \rfloor - (k-1) \\
&= input + 2n - 2n \\
&= input
\end{aligned} \tag{3.12}$$

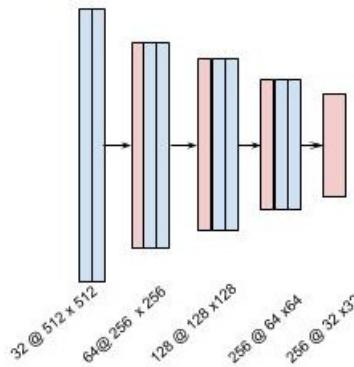


Εικόνα 3.6: Εφαρμογή ενός πυρήνα  $3 \times 3$  με ολίσθηση μονού βηματισμού σε επίπεδο εισόδου μεγέθους  $5 \times 5$  με γέμισμα μηδενικών περιφερειακά της εισόδου [14].

Το στάδιο κωδικοποίησης αποτελείται από 4 τμήματα (εικόνα 3.5) 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. Το επίπεδο μέγιστης συγκέντρωσης είναι μια ανορθόδοξη τεχνική στην οποία επιδιώκουμε να μειώσουμε τον αριθμό των παραμέτρων σταδιακά, κανθάρισμα προχωράμε στα επόμενα τμήματα αυξάνεται ο αριθμός του βάθους των επιπέδων (δηλαδή των φίλτρων) και επομένως και ο αριθμός των παραμέτρων. Ένας άλλος λόγος είναι η προσπάθεια της εξάλειψης της υπερμάθησης ως αποτέλεσμα της μείωσης των παραμέτρων. Στην δική μας περίπτωση συγκεντρώσαμε από κάθε περιοχή  $2 \times 2$  την μέγιστη τιμή των χαρακτηριστικών. Συγκεκριμένα ολισθαίνουμε ένα παράθυρο μεγέθους  $2 \times 2$  στα χαρακτηριστικά και παίρνουμε την μέγιστη τιμή. Διαισθητικά, αυτό σημαίνει ότι χρατήσαμε την τιμή που υπάρχει μεγαλύτερο ερέθισμα στον εκάστοτε νευρώνα (εικόνα 3.8). Επομένως ο χάρτης των χαρακτηριστικών χώρου μετά από κάθε τμήμα μειώνεται κατά το ήμισυ, οι εξισώσεις παρακάτω 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 συγκέντρωσης το οποίο είναι ίσο με 2.

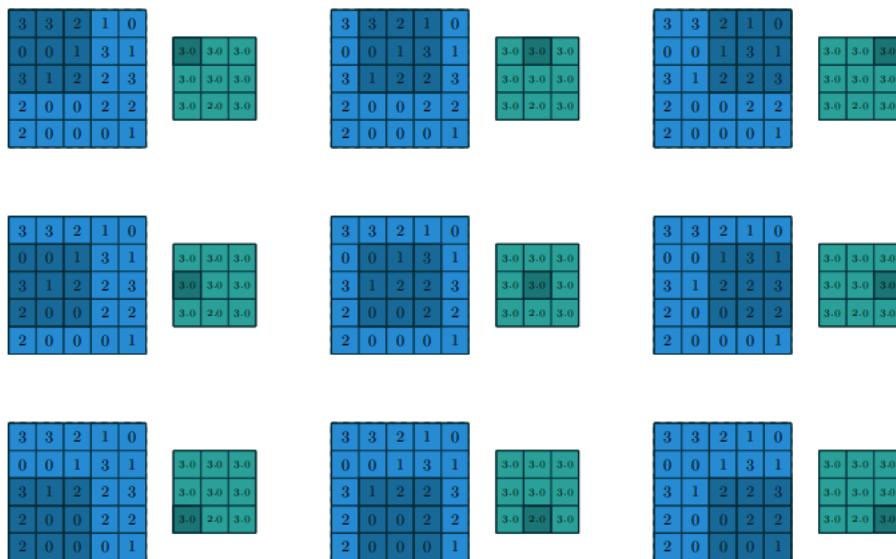
$$\begin{aligned}
Image &= H \times W \times D \\
Height &= (Height - Poolsize + 2 \times P)/S + 1 \\
Width &= (Width - Poolsize + 2 \times P)/S + 1 \\
Output &= Height \times Width \times D
\end{aligned} \tag{3.13}$$

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



Εικόνα 3.7: Στάδιο κωδικοποίησης των ΣΝΔ επίπεδο.

Η εικόνα 3.8 μας δείχνει ένα παράδειγμα εφαρμογής ενός πυρήνα  $3 \times 3$  πάνω σε ένα επίπεδο χαρακτηριστικών  $5 \times 5$  εφαρμόζοντας την μέθοδο της μέγιστης συγκέντρωσης. Όπως βλέπουμε συγκεντρώνουμε το μέγιστο στοιχείο από τον  $3 \times 3$  πυρήνα που ολισθαίνει σε όλο το επίπεδο ενός χάρτη χαρακτηριστικών, από αριστερά προς τα δεξιά.



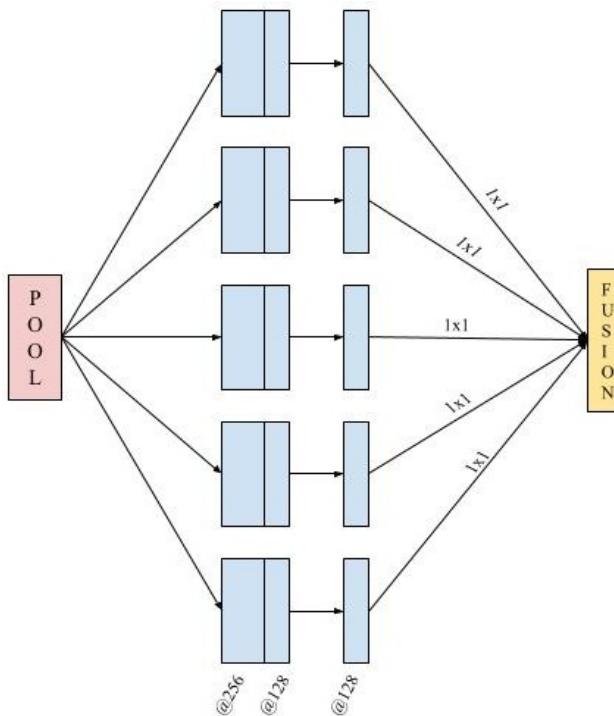
Εικόνα 3.8: Παράδειγμα της μεθόδου της μέγιστης συγκέντρωσης, εφαρμόζοντας ένα παράθυρο  $3 \times 3$  σε ένα επίπεδο χαρακτηριστικών εισόδου  $5 \times 5$  με μονό βήμα ολίσθησης. Τα βήματα είναι από αριστερά προς τα δεξιά [14].

### 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 είσισωση 3.14 μας δείχνει την συνέλιξη σε ένα επίπεδο σήμα εισάγοντας την διαστολή που υποδεικνύεται με  $r$ . Η συγκεκριμένη συνάρτηση υπάγεται στην θεωρία ως Διεσταλμένη Συνέλιξη (Dilated Convolution) ενώ η εικόνα 3.10 μας δίνει μια διαίσθηση γύρω από αυτή την τεχνική.

$$g[i, j] = \sum_k \sum_k f[i + r \cdot k, j + r \cdot k] h[k, k] \quad (3.14)$$

Πιο συγκεκριμένα, κάθε παρακλάδι της μονάδας επεξεργασίας διαφέρει στην διαστολή των στοιχείων του πυρήνα που αλληλεπιδρούν με την είσοδο (εικόνα 3.9). Σκοπός αυτού του τυμήματος είναι η μάθηση χαρακτηριστικών από διαφορετικά πεδία όρασης. Κάθε παρακλάδι διαθέτει έναν πυρήνα με διαφορετική διαστολή στο πρώτο επίπεδο. Ο πυρήνας σε όλους τους κλάδους έχει μέγεθος  $3 \times 3$ , εκτός από το τελευταίο επίπεδο πριν την ένωση των χαρακτηριστικών όπου η πυρήνας έχει μέγεθος  $1 \times 1$ . Στην εικόνα 3.9 βλέπουμε σε κάθε επίπεδο το μέγεθος κάθε επιπέδου και τον αριθμό 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.

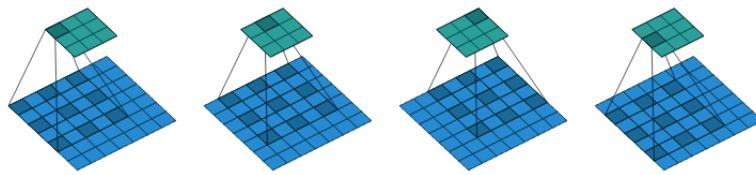


Εικόνα 3.9: Η παράλληλη μονάδα επεξεργασίας με τα 5 ζεχωριστά μονοπάτια. Το κάθε μονοπάτι έχει στο πρώτο επίπεδο συνέλιξης μια διαστολή:  $3 \times 3$ ,  $6 \times 6$ ,  $9 \times 9$ ,  $12 \times 12$  και  $1 \times 1$  αντίστοιχα. Επίσης, βλέπουμε και τον αριθμό των φίλτρων του κάθε επιπέδου συνέλιξης.

Παρακάτω βλέπουμε ένα παράδειγμα για την διεσταλμένη συνέλιξη όπου εφαρμόζουμε έναν πυρήνα  $3 \times 3$  πάνω σε ένα επίπεδο εισόδου  $7 \times 7$  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 αναλογίωτο 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 χρησιμοποιείται 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.



Εικόνα 3.10: Συνέλιξη ενός πυρήνα μεγέθους  $3 \times 3$  πάνω σε ένα επίπεδο εισόδου μεγέθους  $7 \times 7$  και με διαστολή μεγέθους 2. Τα μπλε σκούρα στοιχεία δείχνουν την συμμετοχή για τον υπολογισμό της τιμής του στοιχείου (πράσινο σκούρο) [14].

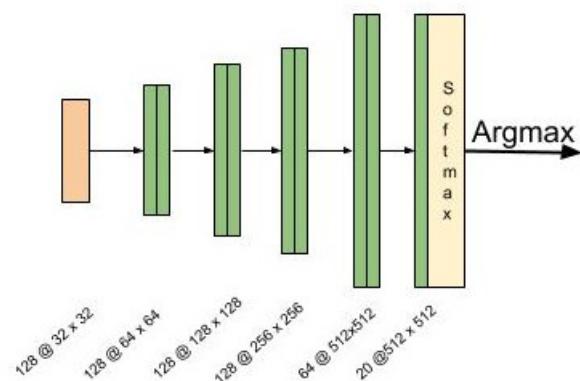
### 3.3.7 Στάδια Αποκωδικοποίησης

#### Επισκόπηση

Παρακάτω θα εξηγήσουμε τις 2 παραλλαγές των μονάδων αποκωδικοποίησης που υλοποιήσαμε για να πειραματιστούμε με αυτές και να συγκρίνουμε 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 ανακατασκευή των χαρακτηριστικών από τον χάρτη χαρακτηριστικών πίσω στην μορφή της εικόνας. Η εικόνα 3.11 μας δείχνει την αρχιτεκτονική της μονάδας αποκωδικοποίησης.



Εικόνα 3.11: Στάδιο αποκωδικοποίησης του ΣΝΔ με χρήση επιπέδων αποσυνέλιξης.

---

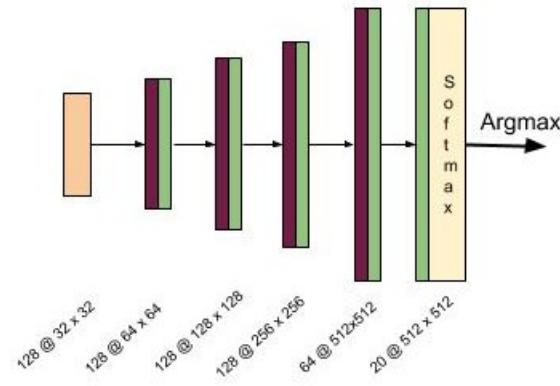
Για να μπορέσουμε να εξηγήσουμε καλύτερα την μονάδα αποκωδικοποίησης ωστε πρώτα να μιλήσουμε για την διαδικασία της αποσυνέλιξης ή ανάστροφης συνέλιξης όπως την βρίσκουμε στην βιβλιογραφία. Η ιδέα και η ανάγκη της ανάστροφης συνέλιξης προκύπτει από την επιθυμία να χρησιμοποιηθεί ένας μετασχηματισμός 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 ανακατασκευή 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 οπισθοδρόμησης 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  $\mathbf{w}$  να ορίζει μια συνέλιξη όπου τα περάσματα (εμπρός-πίσω) να υπολογίζονται από έναν πίνακα  $\mathbf{C}$  και  $\mathbf{C}^T$  αντίστοιχα, αλλά επίσης αν αναστρέψουμε τους πίνακες ορίζουμε την ανάστροφη συνέλιξη ορίζοντας τους πίνακες ως  $\mathbf{C}^T$  και  $(\mathbf{C}^T)^T = \mathbf{C}$  για τα εμπρός και πίσω περάσματα αντίστοιχα. 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 κλάσης ως προς τις υπόλοιπες. Εφαρμόζοντας την softmax μετασχηματίζεται η έξοδος σε μια κατανομή πιθανοτήτων.

## Διγραμμική Μονάδα Αποκωδικοποίησης

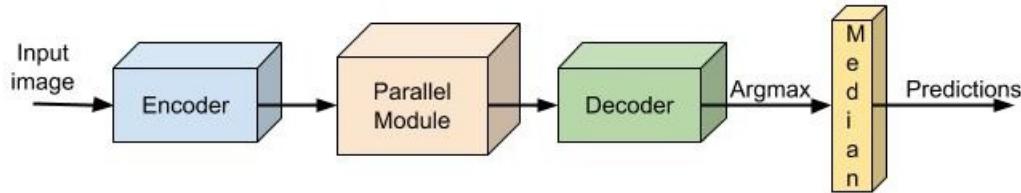
Η διγραμμική μονάδα αποκωδικοποίησης όπως βλέπουμε στην εικόνα 3.12 έχει πανομοιότυπη αρχιτεκτονική με την προηγούμενη μονάδα 3.3.7, με την διαφορά στην μέθοδο που επιτυγχάνεται η υπερδειγματοληψία. Για την υπερδειγματοληψία του χάρτη χαρακτηριστικών χρησιμοποιείται η μέθοδος της διγραμμικής παρεμβολής. Κατά την διγραμμική παρεμβολή, ένα στοιχείο δημιουργείται από μια σταθμισμένη μέση τιμή από τέσσερα γειτονικά στοιχεία από τον χάρτη χαρακτηριστικών εισόδου. Το επίπεδο αυτό δεν διαθέτει παραμέτρους μάθησης.



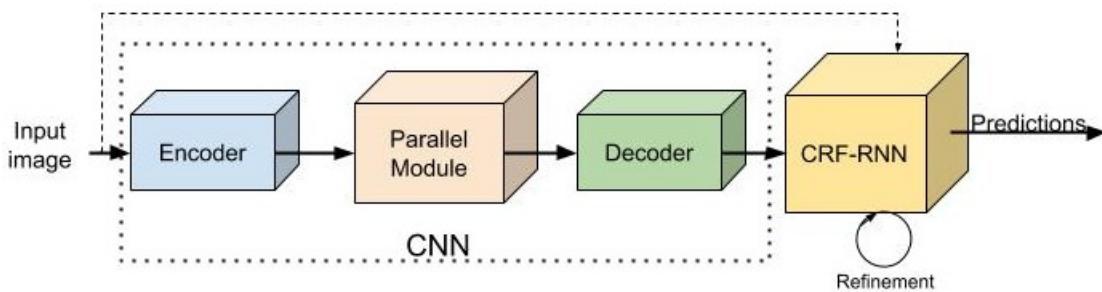
Εικόνα 3.12: Στάδιο αποκωδικοποίησης του ΣΝΔ με χρήση επιπέδων διγραμμικής παρεμβολής για την υπερδειγματοληψία των χαρακτηριστικών. Τα μωβ επίπεδα υποδεικνύουν το επίπεδο της διγραμμικής παρεμβολής.

### 3.3.8 Ολοκληρωμένες Αρχιτεκτονικές

Η εικόνα 3.13 μας δείχνει τις μονάδες των αρχιτεκτονικών που περιγράψαμε προηγουμένως 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.

Εικόνα 3.13: Ολοκληρωμένες Αρχιτεκτονικές.

# Κεφάλαιο 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 στην δική 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 διευθετήσουμε αυτό το πρόβλημα υιοθετήθηκαν δύο διαφορετικές 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 φηφιακού φιλτραρίσματος, που συχνά χρησιμοποιείται για την απομάκρυνση του θορύβου από μια εικόνα ή ένα σήμα. Μια τέτοια μείωση θορύβου είναι ένα τυπικό στάδιο προ-επεξεργασίας για τη βελτίωση των αποτελεσμάτων της μεταγενέστερης επεξεργασίας (για παράδειγμα, ανίχνευση ακμής σε μια εικόνα). Το μεσαίο φιλτράρισμα χρησιμοποιείται ευρέως στη φηφιακή 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 έναν δρόμο, μπορεί να υπάρχουν ορισμένα εικονοστοιχεία 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 [42].

---


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

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

---

**Algorithm** Αλγόριθμος Μεσαίου Φίλτρου (Median Filter) [3].

---

```

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
                for Fy from 0 to K do
                    6:     Window[i] = Input image[ $x + Fx - \text{edgeX}$ ][ $y + Fy - \text{edgeY}$ ]
                    i  $\leftarrow i + 1$ 
                    8:   end for
                end for
                10:  sort values in Window
                    Output Image[x][y]  $\leftarrow \text{Window}[K * K / 2]$ 
            12: end for
        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 Τυχαία υπό Συνθήκη Πεδία (CRF)

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. Στην πραγματικότητα 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. Εμείς 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 ως τυχαίες κατανομές που δημιουργούν ένα MRF όταν υπόκεινται σε μια μεγάλη 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  $P(\mathbf{X} = \mathbf{x} | \mathbf{I}) = \frac{1}{Z(\mathbf{I})} \exp(-E(x|I))$ . Η συνάρτηση  $E(x)$  είναι η ενέργεια των παρατηρήσεων  $x \in \mathcal{L}^N$ , ενώ η  $Z(I)$  είναι η συνάρτηση διαμέρισης (partition function) [27]. Στα πλήρως συνδεδεμένα CRF ζεύγους [25] 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  $x_i$  και 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.

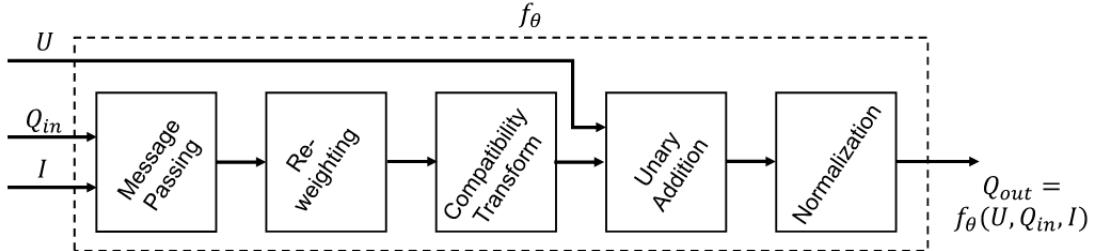
$$\psi_p(x_i, x_j) = \mu(x_i, x_j) \sum_{m=1}^M w^{(m)} k_G^{(m)}(\mathbf{f}_i, \mathbf{f}_j) \quad (4.3)$$

Όπου  $K_G^{(m)}$  (εξίσωση 4.4) συνάρτηση η οποία αποτελείται από δύο Γκαουσιανύς Πυρήνες (Gaussian Kernels) οι οποίοι εφαρμόζονται στα διανύσματα των στοιχείων  $\mathbf{f}_i, \mathbf{f}_j$  τα οποία προέρχονται από τα χαρακτηριστικά της εικόνας, όπως πληροφορία θέσης των εικονοστοιχείων ( $p$ ) και τις τιμές των εικονοστοιχείων ( $I$ : RGB values) και  $w^{(m)}$  είναι τα βάρη των πυρήνων. Ο πρώτος πυρήνας εξαρτάται από την θέση και την τιμή των εικονοστοιχείων ενώ ο δεύτερος εξαρτάται αποκλειστικά από την θέση των εικονοστοιχείων. Οι παράμετροι  $\theta_a, \theta_\beta$  και  $\theta_\gamma$  χρησιμοποιούνται για την κανονικοποίηση των τιμών των πυρήνων. 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.

$$k(\mathbf{f}_i, \mathbf{f}_j) = w^{(1)} \exp\left(-\frac{|p_i - p_j|^2}{2\theta_a^2} - \frac{|I_i - I_j|^2}{2\theta_\beta^2}\right) + w^{(2)} \exp\left(-\frac{|p_i - p_j|^2}{2\theta_\gamma^2}\right) \quad (4.4)$$

Η εύρεση της ακριβής ελάχιστης τιμής δεν είναι εύκολο να υπολογιστεί καθώς δεν μπορούμε να υπολογίσουμε εύκολα την συνάρτηση διαμέρισης. Για αυτό τον σκοπό εφαρμόζεται

η προσέγγιση Μέσου Πεδίου (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 ενός ΣΝΔ (εικόνα 4.1) και πως μοντελοποιείται σαν ένα ENΔ (RNN) [52].



Εικόνα 4.1: Μία επανάληψη Μέσου Πεδίου ως CNN [52]

### 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).

$$Q_i(l) \leftarrow \frac{1}{Z_i} \exp(U_i(l)) \quad \triangleright \text{Initialization} \quad (4.5)$$

### 4.3.3 Πέρασμα Μηνυμάτων

Στα πυκνά μοντέλα CRF το πέρασμα μηνυμάτων πραγματοποιείται εφαρμόζοντας  $M$  Γκαουσιανά Φίλτρα στις  $Q$  κατανομές. Οι συντελεστές των 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 αντιμεταθετικού πλέγματος (Permutohedral lattice) [2] το οποίο κάνει τον υπολογισμό των φίλτρων σε  $O(N)$  χρόνο, όπου  $N$  είναι το πλήθος των εικονοστοιχείων της εικόνας.

Κατά τη διάρκεια της προς τα πίσω διάδοσης σφάλματος, οι είσοδοι των φίλτρων υπολογίζονται με την αποστολή των παραγώγων σφάλματος ως προς την έξοδο του φίλτρου μέσα από τα  $M$  Γκαουσιανά Φίλτρα που θέσαμε αλλά με αντίστροφη κατεύθυνση. Επίσης, στο πλέγμα μία πολυδιάστατη συνέλιξη μπορεί να υλοποιηθεί ως μια ακολουθία από μονοδιάστατες συνέλιξεις κατά μήκος των αξόνων του πλέγματος. Με αυτόν τον τρόπο το αντιμεταθετικό πλέγμα πετυχαίνει έναν πολύ αποδοτικό τρόπο περάσματος μηνυμάτων ανάμεσα στα εικονοστοιχεία.

---


$$\tilde{Q}_i^{(m)}(l) \leftarrow \sum_{j \neq i} k^{(m)}(\mathbf{f}_i, \mathbf{f}_j) Q_j(l) \quad \triangleright \text{Message Passing} \quad (4.6)$$

#### 4.3.4 Στάθμιση Εξόδου Φίλτρου

Το επόμενο βήμα στον αλγόριθμο μέσου πεδίου λαμβάνει ένα σταθμισμένο όμβροισμα των  $M$  εξόδων φίλτρου από το προηγούμενο βήμα για κάθε ετικέτα κλάσης  $l$ . Όταν λαμβάνεται υπόψη κάθε ετικέτα κλάσης μεμονωμένα, μπορεί να θεωρηθεί ως μια συνέλιξη με μέγεθος φίλτρου  $1 \times 1$  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.

Η παράγωγος του σφάλματος ως προς τις εισόδους μπορούν επίσης να υπολογιστούν με τον ίδιο τρόπο, να περάσουν οι παράγωγοι του σφάλματος προς τα πίσω, μέχρι το πρώτο στάδιο. Για να αποκτήσουμε μεγαλύτερο αριθμό ρυθμιζόμενων παραμέτρων, χρησιμοποιούμε ανεξάρτητα βάρη πυρήνα για κάθε ετικέτα κλάσης. Ο χωρικός πυρήνας και ο διμερής πυρήνας έχουν αντίθετες ιδιότητες και η συνεισφορά τους είναι σημαντική. Για παράδειγμα, οι διμερείς πυρήνες μπορεί από τη μία πλευρά να δίνουν έμφαση στην ανίχνευση ποδηλάτων κανώς η ομοιότητα των χρωμάτων είναι καθοριστική. Όμως, μπορεί να μην δίνουν σημασία για την ανίχνευση της τηλεόρασης, δεδομένου ότι οτιδήποτε βρίσκεται μέσα στην ουλόν της τηλεόρασης μπορεί να έχει πολλούς διαφορετικούς τύπους χρωμάτων.

$$\check{Q}_i(l) \leftarrow \sum_m w^{(m)} \tilde{Q}_i^{(m)}(l) \quad \triangleright \text{Weighting Filter Outputs} \quad (4.7)$$

#### 4.3.5 Μετασχηματισμός Συμβατότητας

Στο βήμα Μετασχηματισμού Συμβατότητας, οι έξοδοι από το προηγούμενο βήμα (εξίσωση 4.7) 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. Για παράδειγμα, 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 μέγεθος του φίλτρου είναι  $1 \times 1$  και ο αριθμός των καναλιών εισόδου και εξόδου είναι  $L$ .

---

Μαθαίνοντας τα βάρη του φίλτρου είναι ισοδύναμο με την εκπαίδευση της συνάρτησης μ για τις ετικέτες των εικονοστοιχείων.

$$\hat{Q}_i(l) \leftarrow \sum_{l' \in \mathcal{L}} \mu(l, l') \check{Q}_i(l') \quad \triangleright \text{Compatibility Transform} \quad (4.8)$$

### Πρόσθεση Πιθανοτήτων

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

$$\check{Q}_i(l) \leftarrow U_i(l) - \hat{Q}_i(l') \quad \triangleright \text{Adding Unary Potentials} \quad (4.9)$$

### 4.3.6 Κανονικοποίηση

Τέλος όπως βλέπουμε στην εξίσωση 4.10 έχουμε την κανονικοποίηση στο τέλος της επανάληψης όπου μπορεί να θεωρηθεί ως μια συνάρτηση softmax χωρίς κάποιες παραμέτρους. Οι παράγωγοι από αυτό το βήμα περνάνε κανονικά προς την είσοδο μέσω της προς τα πίσω διάδοσης.

$$Q_i \leftarrow \frac{1}{Z_i} \exp(\check{Q}_i(l)) \quad \triangleright \text{Normalize} \quad (4.10)$$

### 4.3.7 Τυχαία υπό Συνθήκη Πεδία ως Επαναλαμβανόμενα Νευρωνικά Δίκτυα (CRF as RNN)

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

Χρησιμοποιούμε την  $f_\theta$  για να υποδείξουμε την συνάρτηση μεταφοράς που προκύπτει από μια επανάληψη μέσου πεδίου. Δοθέντος 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  $\theta = \{w^m, \mu(l, l')\}$ ,  $\mu \in \{1, \dots, M\}$ , και  $l, l' \in \{l_1, \dots, l_L\}$  αναπαριστούν τις παραμέτρους του CRF που περιγράψαμε προηγουμένως.

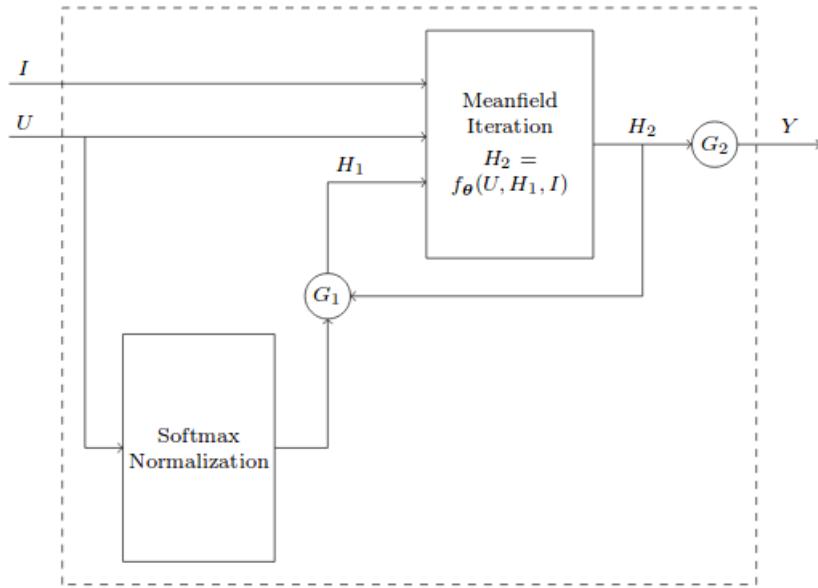
Οι επαναλήψεις του αλγορίθμου του Μέσου Πεδίου υλοποιούνται ως μια στοίβα από επίπεδα με τέτοιο τρόπο ώστε σε κάθε επανάληψη να παίρνει τις εκτιμήσεις  $Q$  της προηγούμενης επανάληψης και τις ενιαίες πιθανότητες  $U$  από το CNN. Αυτή η διαδικασία που ακολουθείται είναι ίδια με την διαδικασία που ακολουθούν τα RNN για εκπαίδευση. Οι εξισώσεις 4.11, 4.12 και 4.13 μας δείχνουν την διαδικασία της επανεκτίμησης των πιθανοτήτων, όπου  $T$  είναι ο αριθμός των επαναλήψεων του Μέσου Πεδίου (Mean-Field Iterations):

$$H_1(t) = \begin{cases} \text{softmax}(U), & t = 0 \\ H_2(t-1), & 0 < t \leq T \end{cases} \quad (4.11)$$

$$H_2(t) = f_\theta(U, H_1(t), I), \quad 0 \leq t \leq T \quad (4.12)$$

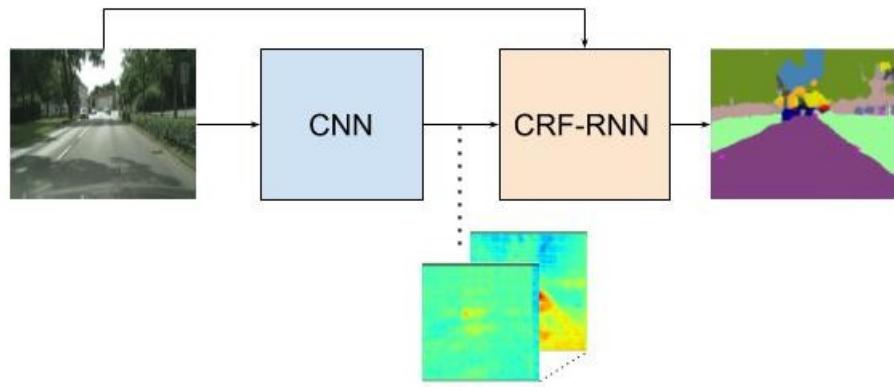
$$Y(t) = \begin{cases} 0, & 0 \leq t < T \\ H_2(t), & t = T \end{cases} \quad (4.13)$$

Οι παράμετροι του μοντέλου (CRF-RNN) είναι ίδιες με τις παραμέτρους του αλγορίθμου Μέσου Πεδίου και αναφέρονται ως  $\Theta$ . 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) [31, 37].



Εικόνα 4.2: Ο επαναληπτικός αλγόριθμος Μέσου Πεδίου ως ένα επαναλαμβανόμενο νευρωνικό δίκτυο. Οι συναρτήσεις  $G_1, G_2$  είναι απλά οι συναρτήσεις εξόδου [52].

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



Εικόνα 4.3: 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.

# Κεφάλαιο 5

## Πειράματα και Αποτελέσματα

### 5.1 Εκπαίδευση των Νευρωνικών Δικτύων

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 συγκλίνει, 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 χρειαστεί να κάνουμε την διαδικασία της μάθησης 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 'Εποχή' 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 συνεχόμενες εποχές, τότε σταματάει η διαδικασία της μάθησης. Με αυτόν τον τρόπο σταματάει η διαδικασία της μάθησης πριν το μοντέλο αρχίσει να μαυρίζει 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 εκπαιδεύτηκαν σε 2975 εικόνες μεγέθους  $512 \times 512$  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 ( $1024 \times 2048$ ). Τα μοντέλα εκπαιδεύτηκαν 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 περιορισμένων 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 μετα-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 ένας μέσος όρος του πλήθους των επιτυχών προβλέψεων του μοντέλου ως προς το άνθρωπισμα των λανθασμένων προβλέψεων του μοντέλου μαζί με τα  $TP$  για την κάθε εικόνα. Πιο συγκεκριμένα μετράμε συνολικά από όλη την εικόνα τις παραμέτρους  $TP, FP, FN$  και υπολογίζουμε την συνάρτηση  $J$ . Επομένως μετράμε την συνάρτηση Jaccard Similarity ( $J$ ) (εξίσωση 5.1) ή Intersection over Union από κάθε εικόνα και παίρνουμε έναν μέσο όρο από τις 500 εικόνες που χρησιμοποιήσαμε για την δοκιμή του μοντέλου (εξίσωση 5.2). 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.

$$J = \frac{TP}{TP + FP + FN}$$

$TP$  = True Positives

(5.1)

$FP$  = False Positives

$$mIoU = \frac{1}{N} \sum_{i=1}^N J(i) \quad (5.2)$$

$FN$  = False Negatives

True Positives συμβολίζονται τα εικονοστοιχεία τα οποία έχουν προβλεφθεί σωστά από τον ταξινομητή. Από την σκοπιά της στατιστικής, False Positives είναι όταν το μοντέλο που έχει προβλέψει ένα αποτέλεσμα απορρίπτει την αναγνώριση του σωστού αποτελέσματος λανθασμένα. Ενώ False Negatives είναι όταν το μοντέλο λανθασμένα απέτυχε να απορρίψει το αποτέλεσμα που πρόβλεψε.

Model	-	9x9	19x19	31x31	45x45	59x59	65x65	71x71	77x77	81x81
SD-CNN-MFB	0.607106	0.60792	0.60847	0.60905	0.60960	0.60997	<b>0.60998</b>	0.60989	0.60976	0.60957
SD-CNN	0.76681	0.76711	<b>0.76716</b>	0.76684	0.76574	0.76380	0.76265	0.76136	0.75996	0.75897
SD-CNN-CRF[3]	0.62051	0.62132	0.62186	0.62251	0.62316	0.62352	<b>0.62355</b>	0.62345	0.62322	0.62303
BD-CNN	0.80342	0.80360	<b>0.80373</b>	0.80351	0.80234	0.80001	0.79857	0.79695	0.79509	0.79375

Πίνακας 5.1: Αποτελέσματα των ΠΣΝΔ με τον αλγόριθμο Μέσου Φίλτρου ως μονάδα μετα-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%.

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 μετα-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 ανταπόχριση του μοντέλου. Σταδιακά μειώσαμε τον ρυθμό μάθησης σε  $10^{-13}$  όμως ακόμα και μετα από 20 εποχές το μοντέλο άρχισε να αποκλίνει. Πιλινότητα λόγω των πολλών επαναλήψεων στο ΤΥΣΠ-ΕΝΔ παρουσιάστηκε το φαινόμενο της εξαφάνισης των αποκλίσεων. Επομένως μειώσαμε τον αριθμό των επαναλήψεων σε 5 κατά την μάθηση και πετύχαμε σύγκλιση μετα από 30 εποχές.

Το πρόβλημα με τους γράφους είναι ότι στηρίζονται πάρα πολύ στον ταξινομητή (ΣΝΔ) που τα τροφοδοτεί για να συνεισφέρουν περισσότερο στην επίδοση. Στην προκειμένη περίπτωση πετύχαμε μια βελτίωση της τάξης του 1%. Από τον πίνακα 5.2 είναι ολοφάνερο πως όσο αυξάνονται οι επαναλήψεις επιβαρύνεται με επιπλέον χρόνο το μοντέλο καθώς σε κάθε επανάληψη γίνεται επανεκτίμηση της κατανομής. Για περισσότερες από 3 επαναλήψεις 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	0.60710	0.1111	0.06
SD-CNN-CRF[3]	0.62058	0.3988	0.09
SD-CNN-CRF[5]	0.62051	0.6994	0.17
SD-CNN-CRF[10]	0.62051	1.1868	0.09
SD-CNN-CRF[20]	0.62051	2.5681	0.48

Πίνακας 5.2: Σύγκριση των μοντέλων με την μονάδα μετα-επεξεργασίας ΤΥΣΠ-ΕΝΔ με διαφορετικό αριθμό επαναλήψεων.

Στον πίνακα 5.3 βλέπουμε την ακρίβεια του μοντέλου μας σε σχέση με τα σύγχρονα μοντέλα στο test set του συνόλου δεδομένων. Για τα αποτελέσματα χρησιμοποιήθηκε η μετρική IoU της εξίσωσης 5.1 με την διαφορά ότι υπολογίστηκε για κάθε κλάση και ο μέσος όρος βγήκε από τον αριθμό των κλάσεων που υπάρχουν σε κάθε εικόνα. Στην δεξιά στήλη βλέπουμε την ακρίβεια που είχαν τα μοντέλα στην σωστή ταξινόμηση των εικονοστοιχείων σε μια από της εφτά υπερ-κατηγορίες. Το σύνολο δεδομένων αποτελείται από 1525 εικόνες, ενώ η αξιολόγηση έγινε στον server της βάσης Cityscapes [11]. Ο πίνακας 5.4 μας δείχνει την ακρίβεια των μοντέλων για την αναγνώριση της κάθε κατηγορίας.

Model	mean IoU Class	mean IoU Category
PSPNet [51]	80.2	90.2
ResNet-DUC-HDC [47]	80.1	-
GRN-LRN-ResNet [49]	77.27	-
PEARL-ResNet101 [21]	74.9	-
RefineNet-ResNet101 [30]	73.06	-
AdapNet [45]	72.91	-
SD-CNN-CRF[3] (Ours)	34.08	60.15

Πίνακας 5.3: Σύγκριση του μοντέλου μας με σύγχρονα μοντέλα στο test set.

Τα περισσότερα μοντέλα που κατέχουν τις πρώτες θέσεις διαθέτουν πολύ βαθιά ΣΝΔ με

εκατοντάδες εκατομμύρια παραμέτρους εν αντιθέσει με το δικό μας μοντέλο το οποίο διαθέτει μόλις 7 εκατομμύρια. Το PSPNet για παράδειγμα διαθέτει ένα προ-εκπαιδευμένο ΣΝΔ (ResNet) με 101 επίπεδα συνέλιξης το οποίο είχε εκπαιδευτεί πρώτα σε ένα άλλο σύνολο δεδομένων (ImageNet) και αποσκοπούσε στην εξαγωγή πολύπλοκων χαρακτηριστικών για την τροφοδοσία του ΣΝΔ (PSPNet) το οποίο πέτυχε την πρώτη θέση στην κατάταξη.

Model	road	sidewalk	building	wall	fence	pole	traffic light	traffic sign	vegetation	terrain	sky	person	rider	car	truck	bus	train	motorcycle	bicycle
SD-CNN-CRF[3]	86.99	44.03	60.00	14.35	9.20	9.60	10.86	10.54	70.72	49.31	81.04	27.55	18.26	71.99	10.51	18.33	21.49	7.13	25.64

Πίνακας 5.4: Αποτελέσματα για κάθε κατηγορία από τον server της βάσης με την μετρική IoU (%).

Όπως μπορούμε να δούμε από τον πίνακα 5.5 βλέπουμε ότι τα αποτελέσματα σε επίπεδο υπερ-κατηγορίας είναι αρκετά ανεβασμένα. Αυτό μας δείχνει ότι το μοντέλο μπερδεύει περισσότερο τα εικονοστοιχεία τα οποία βρίσκονται στην ίδια υπερ-κατηγορία. Αντιθέτως, τα αποτελέσματα για τις υπερ-κατηγορίες 'Άντικείμενο' (object) και 'Ανθρωπος' (human) είναι αρκετά χαμηλά. Μια εξήγηση σε αυτό είναι η αναλογία των εικονοστοιχείων που ανήκουν σε αυτές τις υπερ-κατηγορίες σε σχέση με τα υπόλοιπα εικονοστοιχεία. Τα εικονοστοιχεία που ανήκουν στις παραπάνω υπερ-κατηγορίες είναι πολύ λιγότερα σε σχέση με τα υπόλοιπα και αυτή η μεγάλη δυσαναλογία δεν μπορεί να αντιμετωπιστεί εύκολα από την μέθοδο ισοστάθμισης κλάσεων. Επίσης τα αντικείμενα που ανήκουν σε αυτές τις κατηγορίες είναι πιο δύσκολο να αναγνωριστούν στην εικόνα.

Model	flat	nature	object	sky	construction	human	vehicle
SD-CNN-CRF[3]	93.42	71.01	14.67	81.04	60.19	30.60	70.15

Πίνακας 5.5: Αποτελέσματα μοντέλου ανά υπερ-κατηγορία στο test set.

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 τιμηματοποιήσει 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 χυκλοφορίας και τα φανάρια χυκλοφορίας 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 ανταπεξέλθουν σε αυτό το πρόβλημα. Επίσης, το μεσαίο φίλτρο ομαλοποιεί κάπως κάποια απομακρυσμένα εικονοστοιχεία από την κατηγορία τους τα οποία έχουν ταξινομηθεί σε λάθος κατηγορία.

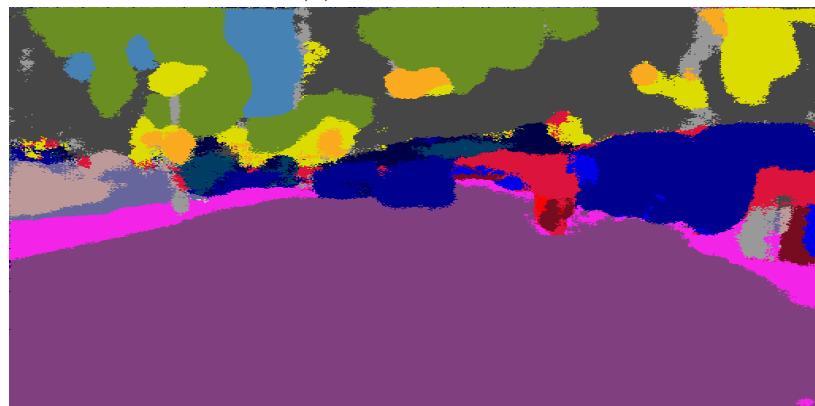
---

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.

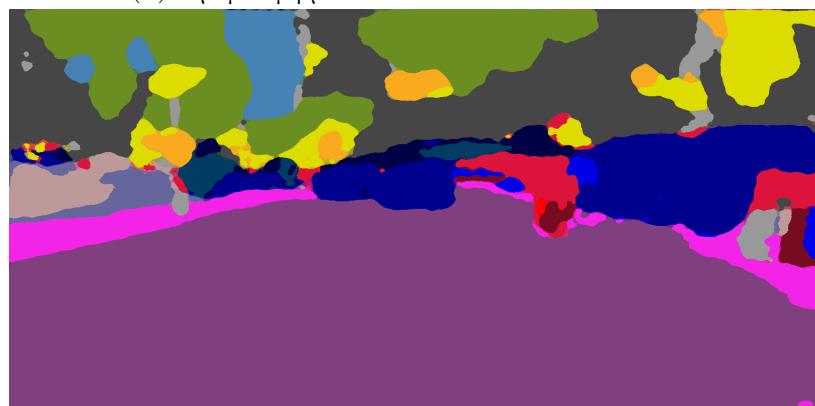
Τέλος, στην εικόνα 5.3 βλέπουμε ένα δείγμα από το μοντέλο με την διγραμμική αποκωδικοποίηση το οποίο δεν έχει εκπαιδευση με την συνάρτηση για την δυσαναλογία των κλάσεων. Κάτι που παρατηρούμε είναι πως αν και χωρίς ισοστάθμιση κλάσεων καταφέρνει να αναγνωρίσει περισσότερα εικονοστοιχεία στην εικόνα εισόδου 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.



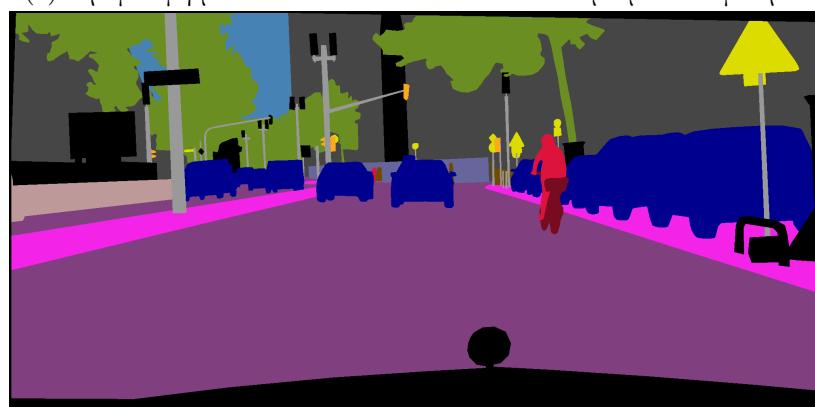
(a) Εικόνα εισόδου



(b) Πρόβλεψη μοντέλου SD-CNN-CRF-RNN.



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



(d) Ground Truth

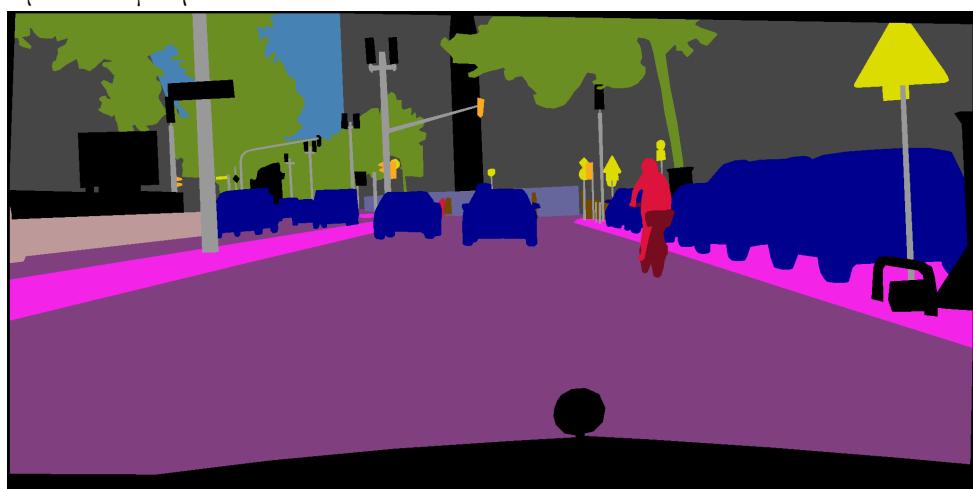
Εικόνα 5.1: Εικόνες αποτελεσμάτων του ΠΣΝΔ με ΤΥΣΠ-ΕΝΔ με ισοστάθμιση χλάσεων.



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



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

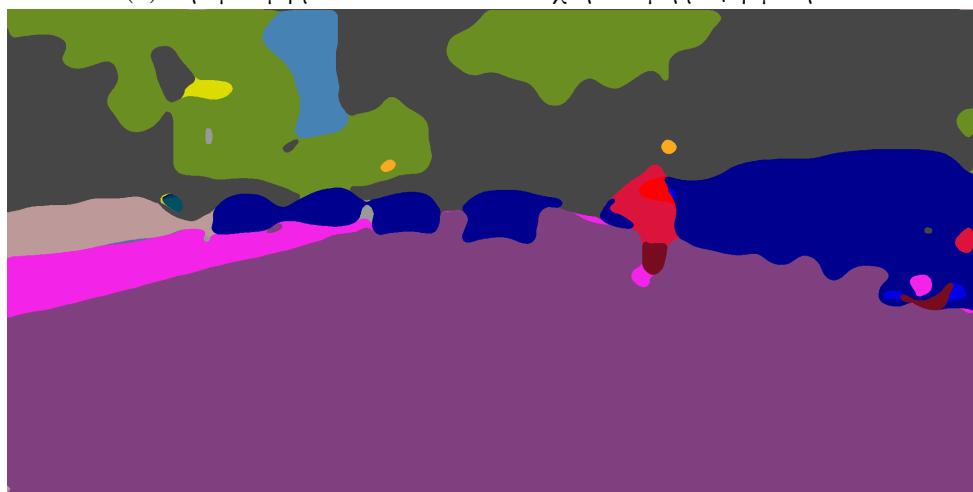


(c) Ground Truth

Εικόνα 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 ( $19 \times 19$ ) του μεσαίου φίλτρου.



(c) Ground Truth

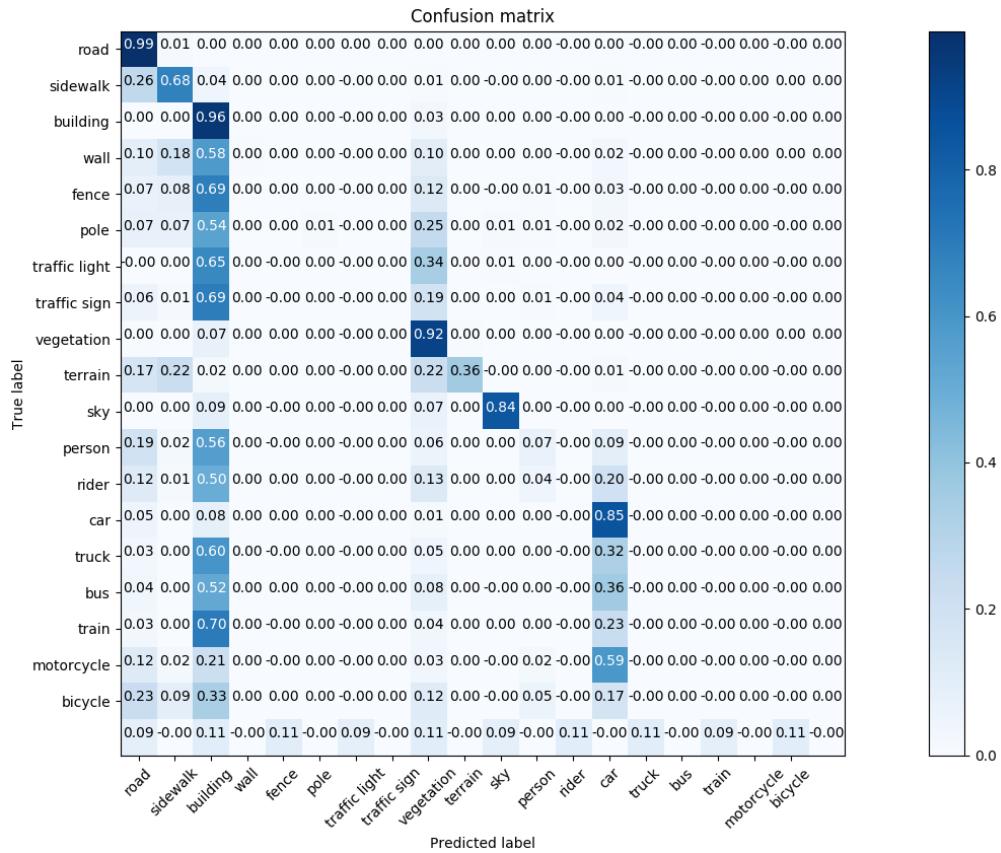
Εικόνα 5.3: Εικόνες αποτελεσμάτων του ΠΣΝΔ χωρίς ισοστάθμιση κλάσεων με εφαρμογή μεσαίου φίλτρου και χωρίς.

---

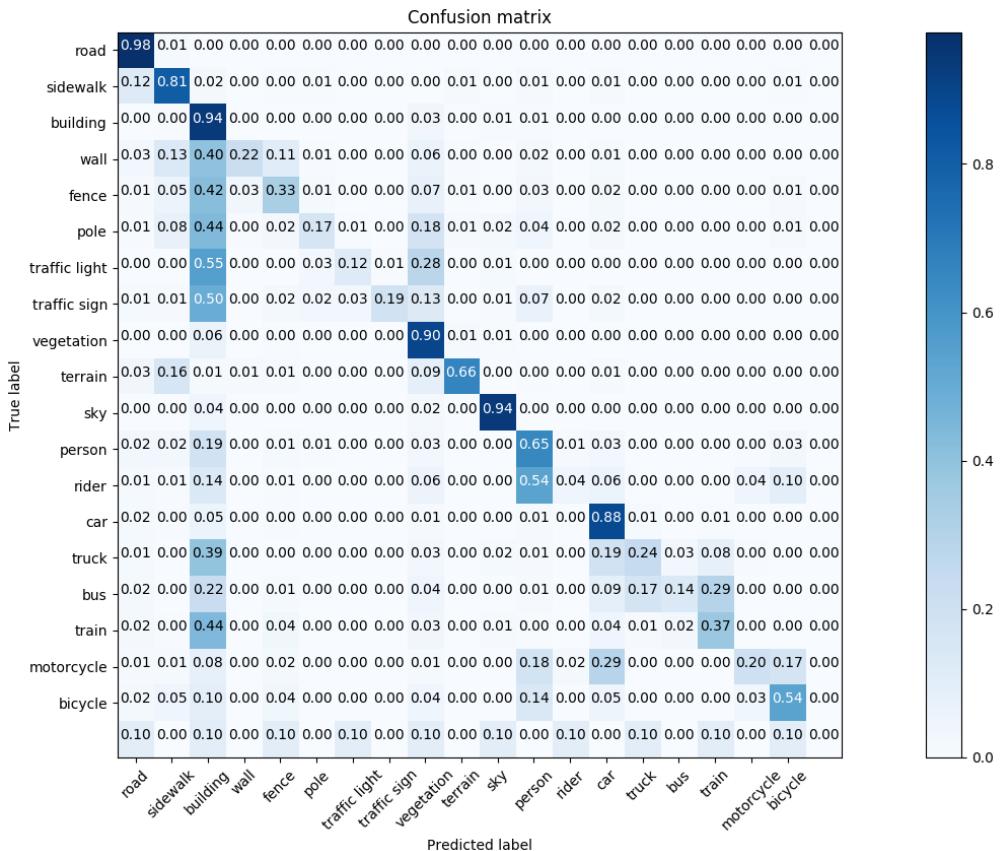
Οι πίνακες σύγχυσης που βλέπουμε παρακάτω στις εικόνες αποτελούν μια πιο αξιόπιστη μέθοδο οπτικοποίησης των αποτελεσμάτων για να δούμε πόσο ισχυρά είναι τα μοντέλα μας και ποιες κατηγορίες μπερδεύονται περισσότερο μεταξύ τους. Οι πίνακες παρουσιάζονται με μια κανονικοποιημένη μορφή, ενώ όσο πιο έντονο 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 και βγάζει μια μέση τιμή από αυτές δεν δύναται να μας δώσει σωστά την ακρίβεια για κάθε κατηγορία. Στην αριστερή πλευρά των πινάκων σύγχυσης έχουμε τις πραγματικές κατηγοριοποιήσεις 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 (εικόνα 5.3) τα πηγαίνει αρκετά καλύτερα.

Τέλος, στην εικόνα 5.5 επιδεικνύεται η σύγκριση μεταξύ του μοντέλου SD-CNN με ισοστάθμιση των κλάσεων (εικόνα 5.2) 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 πολύ 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 μετα-επεξεργασίας. Αξίζει να δούμε την κλάση μοτοσυκλέτα η οποία έχει μειωθεί στο ΤΥΣΠΙ-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 σύγχυσης υπολογίστηκαν πάνω στις 500 εικόνες του συνόλου δεδομένων επαλήθευσης στο αρχικό μέγεθος της εικόνας.

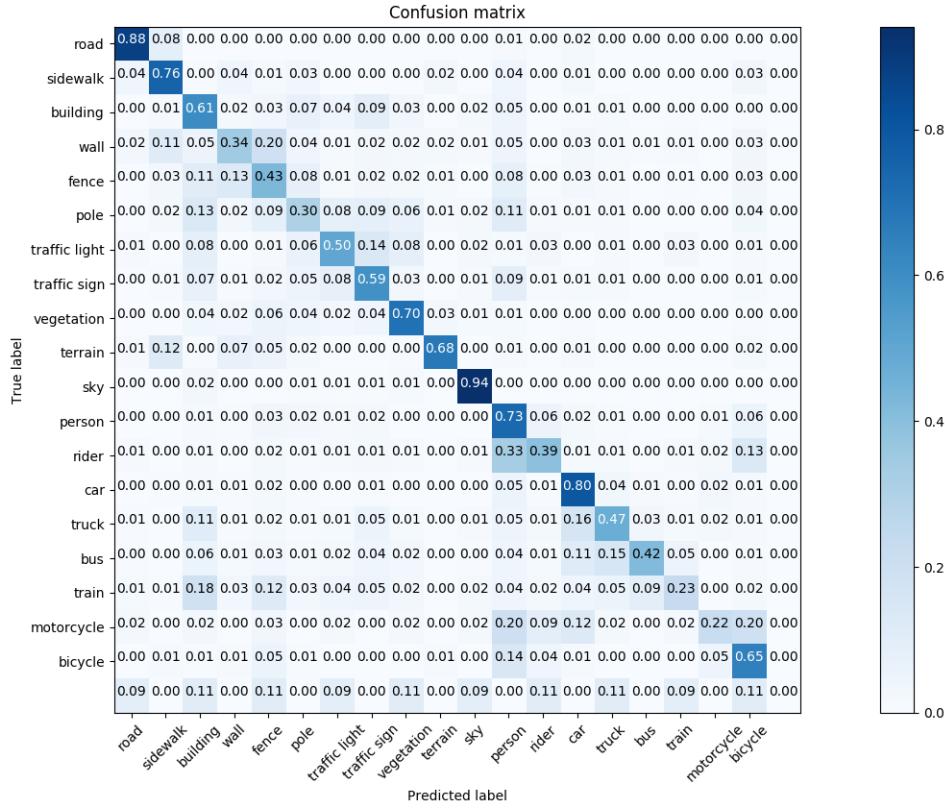


(a) Πίνακας Σύγχυσης του ΠΣΝΔ με μονάδα αποκωδικοποίησης (SD-CNN) χωρίς εφαρμογή της συνάρτησης ισοστάθμισης.

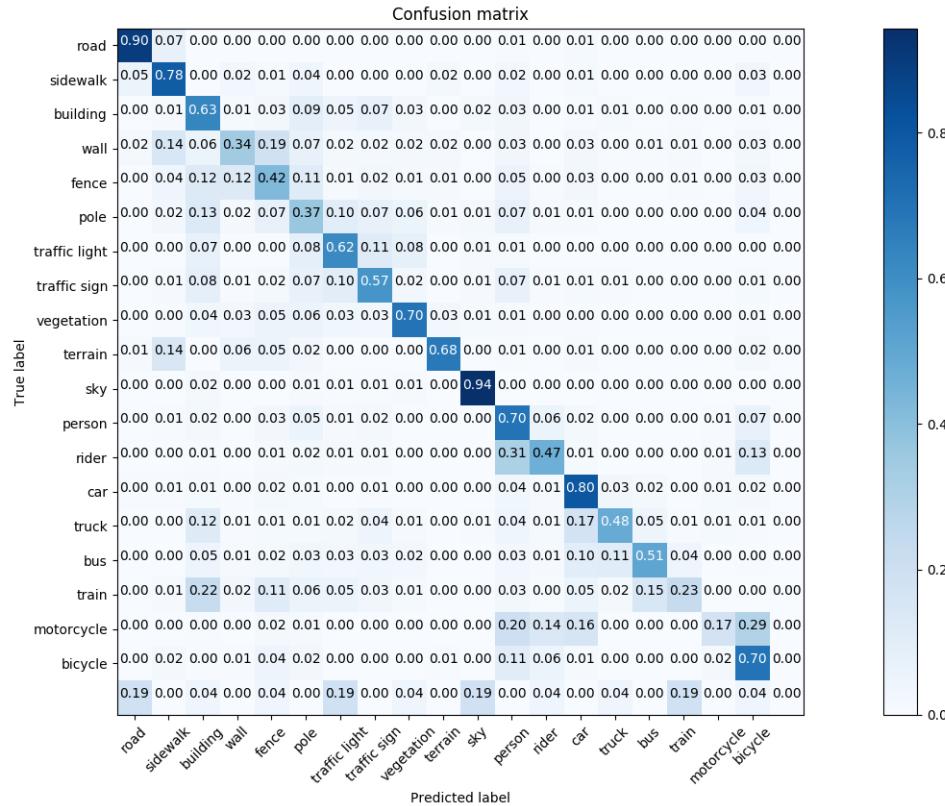


(b) Πίνακας Σύγχυσης του ΠΣΝΔ με διγραμμική μονάδα αποκωδικοποίησης (BD-CNN) χωρίς εφαρμογή της συνάρτησης ισοστάθμισης.

Εικόνα 5.4: Πίνακες σύγχυσης των μοντέλων χωρίς μονάδες μετα-επεξεργασίας και ισοστάθμιση των ακλάσεων.



(a) Πίνακας Σύγχυσης του ΠΣΝΔ με μονάδα αποκωδικοποίησης (SD-CNN) με ισοστάθμιση των κλάσεων.



(b) Πίνακας Σύγχυσης του ΠΣΝΔ με μονάδα αποκωδικοποίησης με άλμα ολίσθησης και με μονάδα μετα-επεξεργασίας ΤΥΣΠ-ΕΝΔ (SD-CNN-CRF) με ισοστάθμιση των κλάσεων.

Εικόνα 5.5: Πίνακες σύγχυσης για την σύγχριση των μοντέλων με και χωρίς μονάδα μετα-επεξεργασίας και ισοστάθμιση των κλάσεων.

# Κεφάλαιο 6

## Συμπεράσματα και Μελλοντική Εργασία

### 6.1 Συμπεράσματα

Ο σκοπός της συγκεκριμένης διπλωματικής ήταν η εξερεύνηση και η σύγχριση 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 βαθιάς 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 καθώς 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 προσέγγιση είναι πιο αποδοτική καθώς δεν προσθέτει υπολογιστικό κόστος στην εκπαίδευση του ΝΔ σε σχέση με άλλες προσεγγίσεις.

Τέλος, παρουσιάσαμε και συγκρίναμε 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, καθώς επιτυγχάνεται μια ισορροπία ως ένα βαθμό μεταξύ της δυσαναλογίας των κλάσεων που υπάρχει στα δεδομένα.

---

## 6.2 Μελλοντική Εργασία

Το θέμα της σημασιολογικής κατάτμησης κεντρίζει όλο και περισσότερο 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 πάρουν αποφάσεις σύμφωνα με τον ακριβή διαχωρισμό των αντικειμένων στο περιβάλλον [5, 46], 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 κατηγοριοποιημένα 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 υπερδειγματοληψία με περισσότερα φίλτρα σε συνδυασμό με την μονάδα ΤΥΣΠ-ΕΝΔ καθώς είναι πιθανόν να υπάρχουν προοπτικές.

Μία διαφορετική κατεύθυνση είναι η χρήση ενός προ-εκπαιδευμένου 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

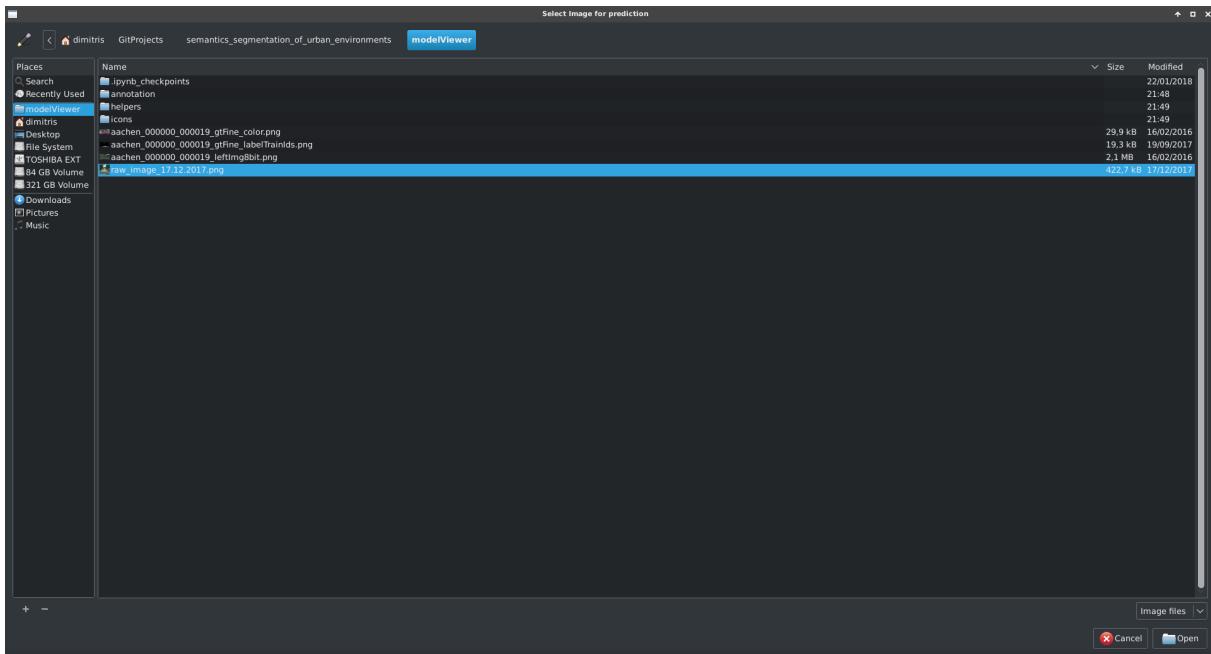
Οι αλγόριθμοι και τα μοντέλα που χρησιμοποιήθηκαν σε αυτή την διπλωματική βρίσκονται στο προφίλ του συγγραφέα στο Github στον παρακάτω σύνδεσμο [link](#). Σε αυτό το τμήμα θα δείξουμε το λογισμικό το οποίο υλοποιήθηκε στα πλαίσια της εργασίας για την οπτικοποίηση των αποτελεσμάτων. Το λογισμικό υλοποιήθηκε με την χρήση των βιβλιοθηκών PyQt4 [33] και OpenCV [18, 19], ενώ για την υλοποίηση των μοντέλων χρησιμοποιήθηκαν οι βιβλιοθήκες Keras [10], Tensorflow [1] και scikit-learn [6].

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

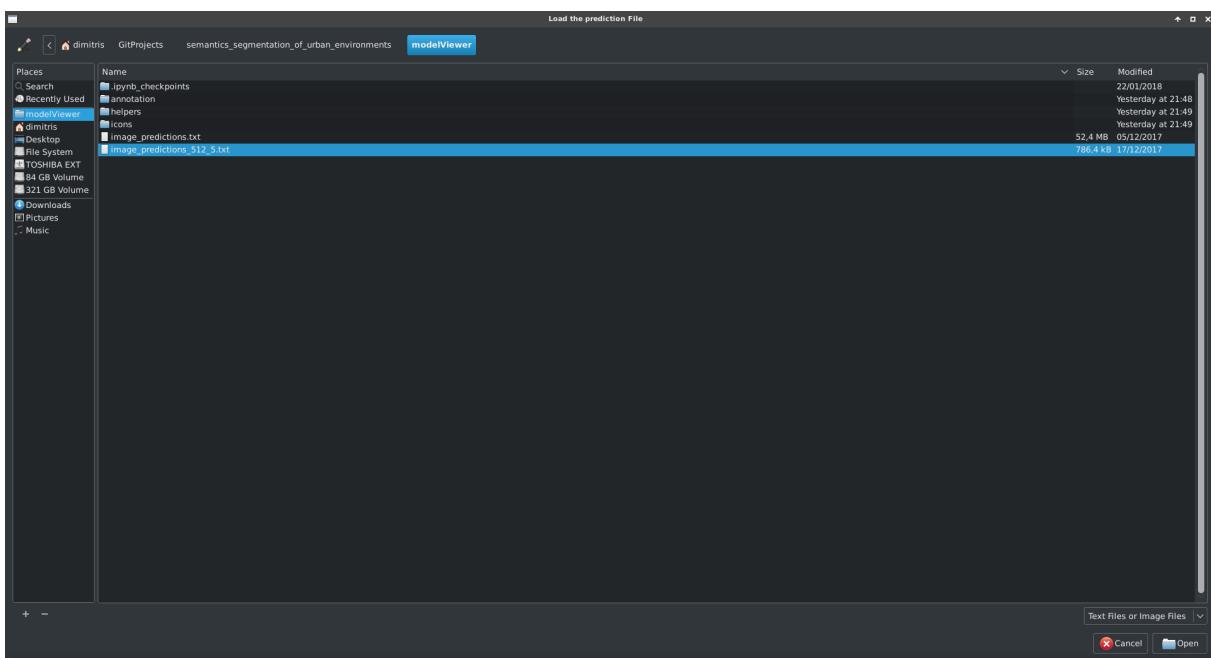
Δρόμος	Πεζοδρόμιο	Κτίριο	Τοίχος	Φράχτης
Ιστός	Φανάρι Κυκλοφορίας	Πινακίδα Κυκλοφορίας	Βλάστηση	Έδαφος
Ουρανός	Άνθρωπος	Αναβάτης	Αυτοκίνητο	Φορτηγό
Λεωφορείο	Τρένο	Μοτοσυκλέτα	Ποδήλατο	Κενό

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

Το λογισμικό που θα δούμε παρακάτω βασίστηκε στο λογισμικό που έχει δημιουργηθεί από την ομάδα της βάσης δεδομένων Cityscapes [11]. Όπως βλέπουμε στις εικόνες 6.1, 6.2 κατά την εκτέλεση του προγράμματος εμφανίζεται παράθυρο επιλογής μίας εικόνας για αναγνώριση και ακολουθεί δεύτερο παράθυρο αντίστοιχα για την επιλογή ενός αρχείου/εικόνας το οποίο περιέχει τις προβλέψεις που έχουν παραχθεί από το μοντέλο μας.

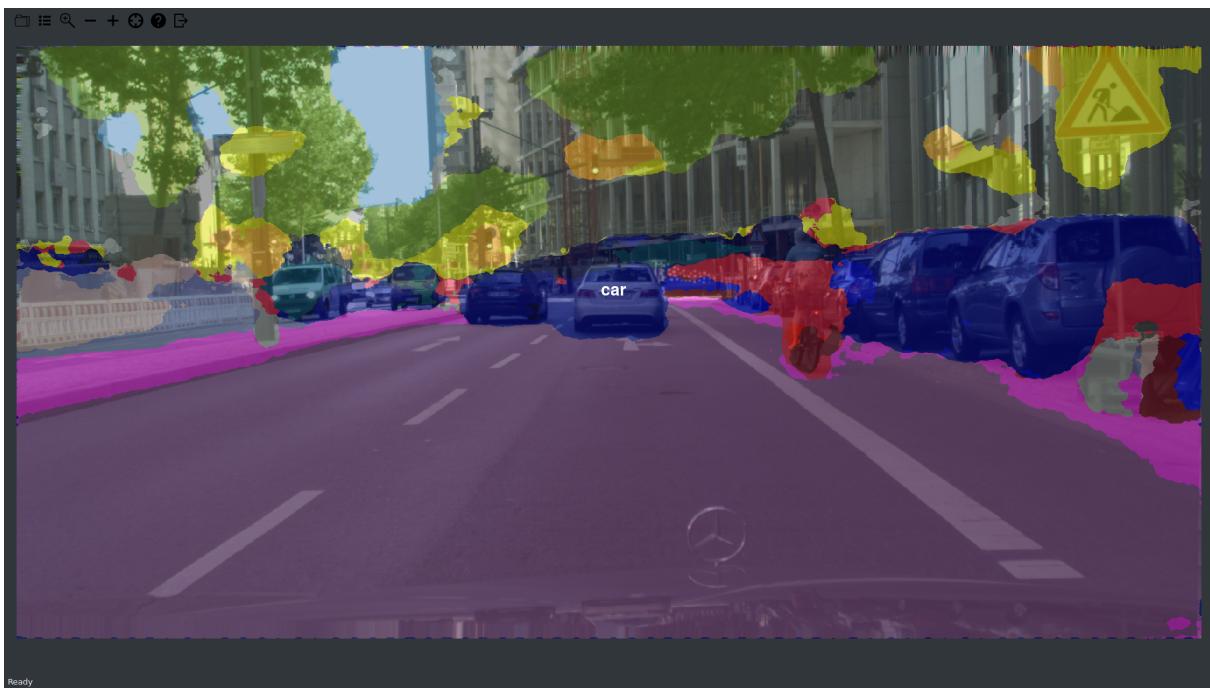


Εικόνα 6.1: Επιλογή αρχικής εικόνας για αναγνώριση.

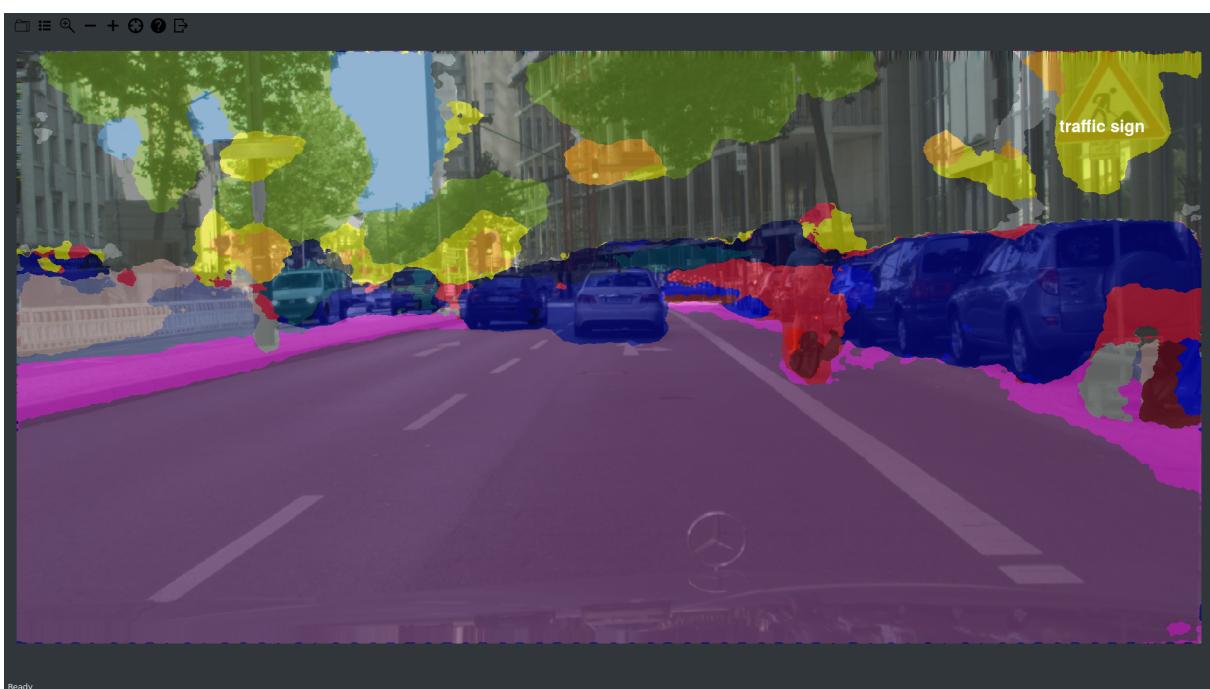


Εικόνα 6.2: Επιλογή αρχείου/εικόνας αποτελεσμάτων

Η εικόνα με τις προβλέψεις ζωγραφίζεται πάνω από την εικόνα εισόδου, όπου κάθε κατηγορία αντικειμένων έχει ένα χρώμα που την αντιπροσωπεύει (εικόνες 6.3 και 6.4). Ο δείκτης ανάλογα με την θέση του δείχνει με άσπρα έντονα γράμματα στα δεξιά του δείκτη την κατηγορία που ανήκει 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 επιθυμητή διαφάνεια ώστε να μπορεί να διαχρίνει ευκολότερα τα αντικείμενα με τις αντίστοιχες κατηγορίες που ανήκουν.

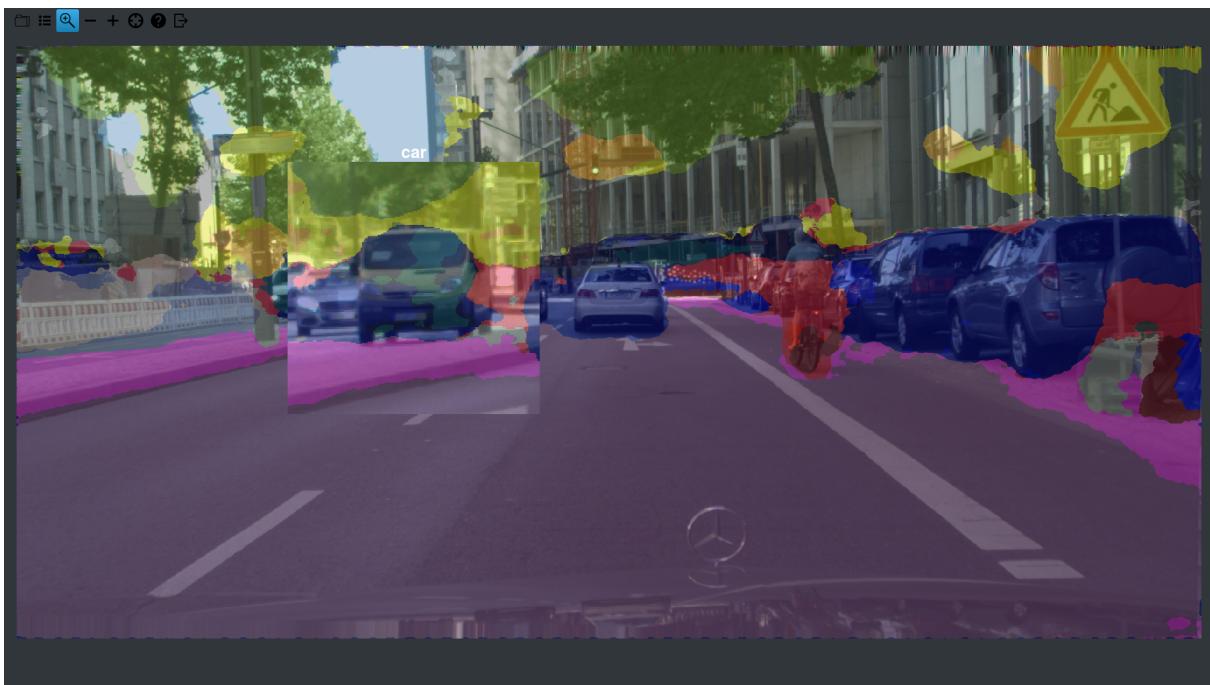


Εικόνα 6.3: Προβολή εικόνας και πρόβλεψης του μοντέλου .



Εικόνα 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).



Εικόνα 6.5: Λειτουργία μεγέθυνσης επιφάνειας.

# Βιβλιογραφία

- [1] 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.
- [2] 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.
- [3] Anonymous. Median filter. wikipedia, 2007. URL [https://en.wikipedia.org/wiki/Median\\_filter](https://en.wikipedia.org/wiki/Median_filter). Accessed 22-May-2018.
- [4] V. Badrinarayanan, A. Kendall, and R. Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *CoRR*, abs/1511.00561, 2015.
- [5] 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>.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] L. Chen, G. Papandreou, F. Schroff, and H. Adam. Rethinking atrous convolution for semantic image segmentation. *CoRR*, abs/1706.05587, 2017.
- [10] F. Chollet et al. Keras, 2015.
- [11] 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.

- 
- [12] 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.
  - [13] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification (2nd Ed)*. Wiley, 2001.
  - [14] 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>.
  - [15] A. Graves. *Supervised Sequence Labelling with Recurrent Neural Networks*, volume 385 of *Studies in Computational Intelligence*. Springer, 2012. ISBN 978-3-642-24796-5. doi: 10.1007/978-3-642-24797-2. URL <https://doi.org/10.1007/978-3-642-24797-2>.
  - [16] M. Hubel and T. N. Wiesel. *Brain and Visual Perception*. Oxford University Press, 2005.
  - [17] 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.
  - [18] *The OpenCV Reference Manual*. Itseez, 2.4.9.0 edition, April 2014.
  - [19] Itseez. Open source computer vision library. <https://github.com/itseez/opencv>, 2015.
  - [20] 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.
  - [21] 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.
  - [22] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014.
  - [23] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014.
  - [24] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter. Self-normalizing neural networks. In *NIPS*, pages 972–981, 2017.
  - [25] P. Krähenbühl and V. Koltun. Efficient inference in fully connected crfs with gaussian edge potentials. *CoRR*, abs/1210.5644, 2012.
  - [26] 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>.

- 
- [27] J. Lafferty. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. pages 282–289. Morgan Kaufmann, 2001.
  - [28] 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.
  - [29] 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.
  - [30] 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.
  - [31] M. C. Mozer. A focused backpropagation algorithm for temporal pattern recognition. *Complex Systems*, 3:349–381, 1989.
  - [32] 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.
  - [33] PyQt. Pyqt reference guide. 2012.
  - [34] J. Redmon. cnn-primer. <https://github.com/pjreddie/cnn-primer/tree/master/1>, 2017.
  - [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] S. Ruder. An overview of gradient descent optimization algorithms. *CoRR*, abs/1609.04747, 2016.
  - [37] 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>.
  - [38] P. Sadowski. Notes on backpropagation, 2016. URL <https://www.ics.uci.edu/~pjsadows/notes.pdf>.
  - [39] 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.
  - [40] E. Shelhamer, J. Long, and T. Darrell. Fully convolutional networks for semantic segmentation. *CoRR*, abs/1605.06211, 2016.
  - [41] 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.
  - [42] R. Szeliski. Computer vision algorithms and applications, 2011.
-

- 
- [43] T. Tieleman and G. Hinton. Lecture 6.1-Overview of mini-batch gradient descent. COURSERA: Neural Networks for Machine Learning, 2012.
  - [44] 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*, 4(2):26–31, 2012.
  - [45] A. Valada, J. Vertens, A. Dhall, and W. Burgard. Adapnet: Adaptive semantic segmentation in adverse environmental conditions. In *ICRA*, pages 4644–4651. IEEE, 2017.
  - [46] 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>.
  - [47] 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.
  - [48] 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.
  - [49] R. Zhang, S. Tang, M. Lin, J. Li, and S. Yan. Global-residual and local-boundary refinement networks for rectifying scene parsing predictions. In *IJCAI*, pages 3427–3433. ijcai.org, 2017.
  - [50] 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>.
  - [51] 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.
  - [52] 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.