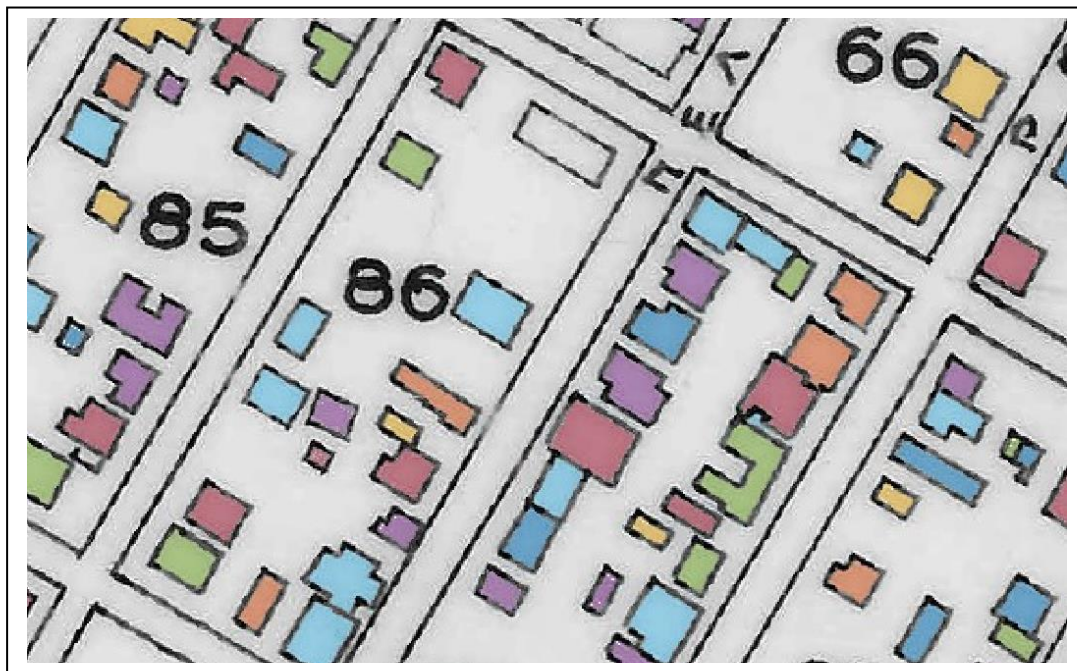




ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ
ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΤΟΠΟΓΡΑΦΙΑΣ ΚΑΙ ΓΕΩΠΛΗΡΟΦΟΡΙΚΗΣ

Διπλωματική Εργασία

ΑΝΑΛΥΣΗ ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑ ΙΣΤΟΡΙΚΩΝ ΧΑΡΤΩΝ ΜΕ ΧΡΗΣΗ ΤΕΧΝΙΚΩΝ ΒΑΘΙΑΣ ΜΑΘΗΣΗΣ



Σταυρούλα-Θεοφανώ Μαλαχία
ΑΜ: 15033

Επιβλέπων Καθηγητής:
Αναστάσιος Α. Κεσίδης
Καθηγητής ΠΑΔΑ

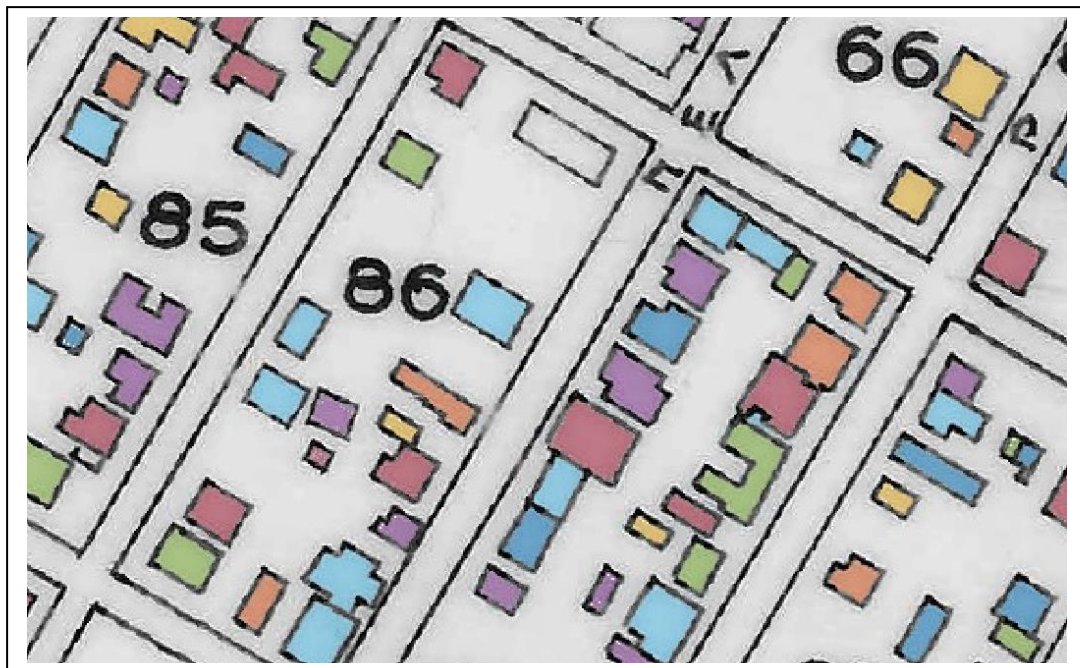
ΑΘΗΝΑ-ΑΙΓΑΛΕΩ, ΙΟΥΛΙΟΣ, 2023



UNIVERSITY OF WEST ATTICA
FACULTY OF ENGINEERING
DEPARTMENT OF ELECTRICAL & ELECTRONICS ENGINEERING

Diploma Thesis

**HISTORICAL MAPS PROCESSING AND ANALYSIS USING
DEEP LEARNING TECHNIQUES**



Stavroula-Theofano Malachia
Registration Number: 15033

Supervisor
Anastasios L. Kesidis
Professor

ATHENS-EGALEO, July, 2023



ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ

ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΤΟΠΟΓΡΑΦΙΑΣ ΚΑΙ ΓΕΩΠΛΗΡΟΦΟΡΙΚΗΣ

Ανάλυση και Επεξεργασία Ιστορικών Χαρτών με Χρήση Τεχνικών Βαθιάς Μάθησης

Μέλη Εξεταστικής Επιτροπής συμπεριλαμβανομένου και του Εισηγητή

Η πτυχιακή/διπλωματική εργασία εξετάστηκε επιτυχώς από την κάτωθι Εξεταστική Επιτροπή:

A/a	ΟΝΟΜΑ ΕΠΩΝΥΜΟ	ΒΑΘΜΙΑΔΑ/ΙΔΙΟΤΗΤΑ	ΨΗΦΙΑΚΗ ΥΠΟΓΡΑΦΗ
1	Αναστάσιος Λ. Κεσίδης (Επιβλέπων)	Καθηγητής ΠΑΔΑ	
2	Νικόλαος Μερλέμης	Αναπληρωτής Καθηγητής ΠΑΔΑ	
3	Βασίλειος Κρασανάκης	Επίκουρος Καθηγητής ΠΑΔΑ	

Copyright © Με επιφύλαξη παντός δικαιώματος. All rights reserved.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ ΚΑΙ ΜΑΛΑΧΙΑ ΣΤΑΥΡΟΥΛΑ-ΘΕΟΦΑΝΩ, ΙΟΥΛΙΟΣ 2023

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τους συγγραφείς.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον/την συγγραφέα του και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις θέσεις του επιβλέποντος, της επιτροπής εξέτασης ή τις επίσημες θέσεις του Τμήματος και του Ιδρύματος.

ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

Η κάτωθι υπογεγραμμένη Μαλαχία Σταυρούλα-Θεοφανώ του Γεωργίου, με αριθμό μητρώου 509150330227 φοιτήτρια του Πανεπιστημίου Δυτικής Αττικής της Σχολής ΜΗΧΑΝΙΚΩΝ του Τμήματος Μηχανικών Τοπογραφίας Και Γεωπληροφορικής,

δηλώνω υπεύθυνα ότι:

«Είμαι συγγραφέας αυτής της διπλωματικής εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, οι όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε ακριβώς είτε παραφρασμένες, αναφέρονται στο σύνολό τους, με πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο. Επίσης, βεβαιώνω ότι αυτή η εργασία έχει συγγραφεί από μένα αποκλειστικά και αποτελεί προϊόν πνευματικής ιδιοκτησίας τόσο δικής μου, όσο και του Ιδρύματος.

Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του διπλώματός μου.

Η Δηλούσα

Σταυρούλα-Θεοφανώ Μαλαχία



Περιεχόμενα

<i>Περίληψη</i>	1
<i>Abstract</i>	2
<i>Ευχαριστίες</i>	3
<i>Πίνακας Εικόνων:</i>	4
1 <i>Εισαγωγή</i>	7
2 <i>Νευρωνικά Δίκτυα</i>	10
2.1 <i>Εισαγωγή</i>	10
2.2 <i>Ιστορική αναδρομή</i>	10
2.3 <i>Τεχνητή Νοημοσύνη</i>	15
2.4 <i>Μηχανική μάθηση</i>	15
2.5 <i>Τεχνητά νευρωνικά δίκτυα</i>	17
2.6 <i>Συνελικτικά νευρωνικά δίκτυα</i>	19
3 <i>Βασικά στάδια προ-επεξεργασίας της εικόνας</i>	30
3.1 <i>Εισαγωγή</i>	30
3.2 <i>Ψηφιακή εικόνα</i>	30
3.3 <i>Μορφολογικές πράξεις</i>	34
3.4 <i>Τμηματοποίηση της εικόνας</i>	37
3.5 <i>Εξαγωγή υπο-εικόνων</i>	38
3.6 <i>Επαύξηση δεδομένων</i>	39
4 <i>Εντοπισμός κτιρίων σε ιστορικούς χάρτες</i>	40
4.1 <i>Εισαγωγή</i>	40
4.2 <i>Προβλήματα</i>	41
4.3 <i>Εργαλείο επισήμανσης - Annotation tool</i>	43
4.4 <i>Εξαγωγή αποτυπωμάτων των κτιρίων</i>	44
4.5 <i>Εξαγωγή περιγραμμάτων των κτιρίων</i>	51
4.6 <i>Κατάτμηση σε υπο-εικόνες</i>	55
4.7 <i>Επαύξηση δεδομένων</i>	60
4.8 <i>Εκπαίδευση νευρωνικού δικτύου U-NET</i>	62
4.9 <i>Πειραματικά Αποτελέσματα</i>	63

4.10	Αποτίμηση	71
5	Συμπεράσματα	78
5.1	Προτάσεις μελλοντικής βελτίωσης	79
	Βιβλιογραφία	80

Περίληψη

Οι ιστορικοί τοπογραφικοί χάρτες είναι πηγή πολύτιμων πληροφοριών. Παρέχουν πληροφορίες σχετικά με το γεωγραφικό περιβάλλον και τις ανθρώπινες δραστηριότητες ενός συγκεκριμένου χώρου και χρόνου. Λόγω της χαμηλής γραφικής ποιότητας που χαρακτηρίζει τους παλιούς χάρτες, η εξαγωγή πληροφορίας από αυτούς και μάλιστα με αυτοματοποιημένο τρόπο είναι μια απαιτητική διαδικασία. Η παρούσα διπλωματική εργασία αφορά την ανάλυση και επεξεργασία ιστορικών χαρτών με χρήση τεχνικών βαθιάς μάθησης. Συγκεκριμένα, σκοπός ήταν ο αυτόματος προσδιορισμός των περιγραμμάτων και των αποτυπωμάτων των κτιρίων σε ιστορικούς πολεοδομικούς χάρτες. Για τον σκοπό αυτό χρησιμοποιήθηκε ένα βαθύ νευρωνικό δίκτυο το οποίο εκπαιδεύτηκε με διάφορα δεδομένα εισόδου προκειμένου να εξετασθεί η αποτελεσματικότητά του στον εντοπισμό των κτιρίων. Η εκπαίδευση του νευρωνικού δικτύου βασίστηκε σε ένα σύνολο δεδομένων στα οποία είχαν προηγουμένως επισημανθεί τα κτίρια πάνω στην τοπογραφική εικόνα (ground truth). Για να διευκολυνθεί η διαδικασία επισήμανσης των κτιρίων αναφοράς, αναπτύχθηκε επιπλέον στα πλαίσια της παρούσας εργασίας ένα εργαλείο επισήμανσης (annotation tool) σε ψηφιακές εικόνες προσαρμοσμένο ειδικά για την επισήμανση κτιρίων σε τοπογραφικές εικόνες. Στη συνέχεια πραγματοποιήθηκε η εκπαίδευση και η αξιολόγηση ενός συνελκτικού νευρωνικού δικτύου αρχιτεκτονικής U-Net. Η αποτίμηση του δικτύου πραγματοποιήθηκε με 12 σετ δεδομένων. Αυτά διαφέρουν μεταξύ τους ως προς τον τρόπο δειγματοληψίας με πλεγματική και τυχαία μορφή και ως προς την επιθυμητή έξοδο του δικτύου που αφορούσε τα περιγράμματα όσο και το συνολικό αποτύπωμα των κτιρίων. Ο προσδιορισμός του περιγράμματος των κτιρίων μέσω του αποτυπώματός τους βασίστηκε σε μορφολογικές πράξεις. Η αποτελεσματικότητα του προτεινόμενου δικτύου βαθιάς μάθησης αποτιμήθηκε μέσω διαφόρων μετρικών και αναδείχθηκαν εκείνες οι παράμετροι του συστήματος που δίνουν τα καλύτερα αποτελέσματα στον προσδιορισμό του περιγράμματος και του αποτυπώματος των κτιρίων στον τοπογραφικό χάρτη.

Abstract

Historical topographic maps are a source of valuable information. They provide information about the geographical environment and human activities for a particular place and time. Due to the low graphic quality, extracting information from old maps in an automated manner, is a challenging process. This thesis is about the analysis and processing of historical maps by using deep learning techniques. In particular, it aims to automatically determine the boundaries and the footprints of buildings on historical urban maps. For this purpose, a deep neural network is involved which is equipped with various input data in order to test its effectiveness in identifying buildings. The training of the neural network is based on a dataset in which the buildings on the topographic image have been previously labeled (ground truth). In order to facilitate the process of labeling the reference buildings, an annotation tool on digital images is also developed in the context of this work, which is specifically adapted for the labeling of buildings on topographic images. Subsequently, the training and evaluation of a convolutional neural network of U-Net architecture is carried out. The evaluation of the network is performed with 12 datasets. These differed from each other in terms of the sampling approach which followed a grid and a random mode as well as in terms of the desired network's output which concerned the boundary and the footprint of the buildings. The determination of the building's boundaries through its footprint is based on morphological operations. The effectiveness of the proposed deep learning network is evaluated through various metrics and those system parameters that give the best results in determining the boundary and footprint of the buildings on the topographic map are highlighted.

Ευχαριστίες

Με την ολοκλήρωση της παρούσας Διπλωματικής εργασίας «κλείνει» ο κύκλος των προπτυχιακών σπουδών μου στο πενταετές πρόγραμμα, στο τμήμα Μηχανικών Τοπογραφίας και Γεωπληροφορικής του Πανεπιστημίου Δυτικής Αττικής. Για τον λόγο αυτό ήθελα να ευχαριστήσω θερμά όλους όσους συνέβαλλαν στην εκπόνησή της.

Θα ήθελα να ευχαριστήσω τον επιβλέπων καθηγητή της εν λόγω εργασίας, κύριο Αναστάσιο Κεσίδη, τόσο για την άμεση ανταπόκριση και τη βοήθεια που μου προσέφερε, όσο και για τις εύστοχες συμβουλές και τη καθοδήγησή του καθ' όλη τη διάρκεια ως και την υλοποίηση της.

Τέλος, δεν θα μπορούσα να παραλείψω την οικογένειά μου και τους φίλους μου. Τους ευχαριστώ θερμά για την στήριξη και την ενθάρρυνση τους καθ' όλη την διάρκεια των σπουδών μου, πόσο μάλλον κατά τη διάρκεια συγγραφής την παρούσας εργασίας.

Πίνακας Εικόνων:

Εικόνα 2.1 Πρώτη παρουσίαση ενός τεχνητού νευρώνα [2] _____	11
Εικόνα 2.2 Ο Rosenblatt με το Mark I Perceptron -Ένα νευρωνικό δίκτυο που πραγματοποιούσε οπτικές ταξινομήσεις (Αριστερά), Μια οπτική αναπαράσταση του (Δεξιά) [3] _____	13
Εικόνα 2.3 Επιβλεπόμενη Μάθηση (Supervised learning) [4] _____	16
Εικόνα 2.4 Μη Επιβλεπόμενη Μάθηση (Unsupervised Learning) [5] _____	16
Εικόνα 2.5 Ενισχυτική Μάθηση (Reinforcement Learning) [6] _____	17
Εικόνα 2.6 Νευρωνικό Κύτταρο [7] _____	18
Εικόνα 2.7 Δομή νευρωνικού κυττάρου _____	18
Εικόνα 2.8 Δομή τεχνητού νευρωνικού δικτύου _____	19
Εικόνα 2.9 Παράδειγμα διαφορετικών επιπέδων όπου το καθένα «κοιτά» σε διαφορετικό παράθυρο της εικόνας. (Τοπική Συνδεσιμότητα) [9] _____	20
Εικόνα 2.10 Layers τα οποία «κοιτούν» στο ίδιο παράθυρο της εικόνας [8] _____	21
Εικόνα 2.11 Συνέλιξη, ως ένας εναλλακτικός τρόπος της πλήρους συνδεσιμότητας [8] _____	22
Εικόνα 2.12 Αριστερή εικόνα: stride=0, μεσαία εικόνα stride=1, δεξιά εικόνα stride=2 [12] _____	23
Εικόνα 2.13 Αριστερή εικόνα: stride=0, μεσαία εικόνα stride=1, δεξιά εικόνα stride=2 [12] _____	23
Εικόνα 2.14 Βήμα [13] _____	23
Εικόνα 2.15 Μηδενικό Γέμισμα [8] _____	24
Εικόνα 2.16 Οι πιο διαδεδομένες συναρτήσεις ενεργοποίησης [8] _____	25
Εικόνα 2.17 Εφαρμογή Μέγιστης Συγκέντρωσης με παράθυρο 2x2 [8] _____	26
Εικόνα 2.18 Χωρική μείωση των διαστάσεων μιας εικόνας με την μέθοδο συγκέντρωσης [14] _____	26
Εικόνα 2.19 (α) Ένα συνελκτικό νευρωνικό δίκτυο. (β) Πλήρως συνδεδεμένο δίκτυο [15] _____	27
Εικόνα 2.20 Οπτικοποίηση βάθους των επιπέδων ενός νευρωνικού δικτύου [8] _____	27
Εικόνα 2.21 Αρχιτεκτονική U-Net [16] _____	28
Εικόνα 3.1 Η κάθε θέση (i,j) αποτελεί ένα εικονοστοιχείο της εικόνας [18] _____	31
Εικόνα 3.2 Δυαδική μορφή εικόνας [19] _____	31
Εικόνα 3.3 Grayscale εικόνα [19] _____	32
Εικόνα 3.4 RGB εικόνα [19] _____	32
Εικόνα 3.5 Λήψη Εικόνας [20] _____	33
Εικόνα 3.6 Συνεχής Εικόνα _____	33
Εικόνα 3.7 Συλλογή Ψηφιακής Εικόνας [18] _____	34
Εικόνα 3.8 Χωρική Αναπαράσταση Εικόνας 2D Pixels [18] _____	34
Εικόνα 3.9 Δομικό Στοιχείο B 3x3 _____	35
Εικόνα 3.10 Εικόνα A _____	35
Εικόνα 3.11 Η κίνηση του δομικού στοιχείου σε τα ενεργά pixel της εικόνας _____	36
Εικόνα 3.12 Διαστολή Εικόνας A με δομικό στοιχείο B _____	36
Εικόνα 3.13 Σε ποιο θέση της εικόνας χωράει ολόκληρο το δομικό στοιχείο _____	36
Εικόνα 3.14 Συστολή Εικόνας A με δομικό στοιχείο B _____	37
Εικόνα 3.15 Συνδεσιμότητα 4 _____	37
Εικόνα 3.16 Συνδεσιμότητα 8 _____	37
Εικόνα 3.17 (α) Πλέγμα 4 Γειτόνων, (β) Πλέγμα 8 Γειτόνων [18] _____	38
Εικόνα 3.18 Η διαδικασία του αλγορίθμου floodfill. Η αρχική εικόνα είναι η πρώτη από τα αριστερά. Προχωρώντας δεξιά, επιλέγεται ο αρχικός κόμβος-pixel, τα γειτονικά pixel γεμίζουν και μετέπειτα οι γείτονες των γειτόνων γεμίζουν, έως ότου γεμίσει όλη η εικόνα [21] _____	38
Εικόνα 4.1 Τοπογραφικοί χάρτες περιοχής Πετρούπολης τεσσάρων δεκαετιών οι οποίοι διατέθηκαν από την ΕΛΣΤΑΤ _____	40
Εικόνα 4.2 Τοπογραφικός χάρτης περιοχής Πετρούπολης 1971 _____	41

Εικόνα 4.3 (α) Νούμερα που επικαλύπτουν τα περιγράμματα (β) Το κείμενο επικαλύπτει τα περιγράμματα	42
Εικόνα 4.4 Στενοί δρόμοι που εφάπτονται στα οικοδομικά τετράγωνα	42
Εικόνα 4.5 Μη κλειστά πολύγωνα	42
Εικόνα 4.6 Κτίρια τα οποία ποικίλλουν σε μέγεθος, περίγραμμα και προσανατολισμό	43
Εικόνα 4.7 Ο τοπογραφικός χάρτης της Πετρούπολης σε κλίμακα του γκρι	44
Εικόνα 4.8 Grayscale αποσπάσματα του τοπογραφικού χάρτη	45
Εικόνα 4.9 Χάρτης με όλα τα περιγράμματα στα οποία έχει γίνει επισήμανση	45
Εικόνα 4.10 Απόσπασμα χάρτη με όλα τα σπίτια που έχουν επισημανθεί	46
Εικόνα 4.11 Αποσπάσματα χάρτη με όλα τα σπίτια που έχουν επισημανθεί	46
Εικόνα 4.12 Περιβάλλον του annotation tool. (α) toolbar, (β) labels tool, (γ) zoom in, zoom out, pan, home button	47
Εικόνα 4.13 Πως μπορεί ο χρήστης να φορτώσει τα ήδη αποθηκευμένα labels	48
Εικόνα 4.14 Παράδειγμα μεγέθυνσης και επισήμανσης κτιρίου με ένα κλικ	48
Εικόνα 4.15 Παράδειγμα, πως γίνεται η ίδια επισήμανση σε ένα «κομμένο» κτίριο με την ροδέλα του ποντικιού	49
Εικόνα 4.16 Πως μπορεί να αποθηκεύσει ο χρήστης τα κτίρια που έχουν επισημανθεί.	50
Εικόνα 4.17 Παράδειγμα προβληματικών οικοδομικών τετραγώνων	50
Εικόνα 4.18 (α) Συστολή ενός περιγράμματος, (β) Διαστολή ενός περιγράμματος, (γ) Μορφολογική κλίση	51
Εικόνα 4.19 Αρχική ασπρόμαυρη εικόνα που περιέχει όλα τα κτίρια που έχουν επισημανθεί	52
Εικόνα 4.20 Αποσπάσματα περιγραμμάτων	53
Εικόνα 4.21 Μορφολογική κλίση των περιγραμμάτων	54
Εικόνα 4.22 Απόσπασμα περιγραμμάτων με μορφολογική κλίση	54
Εικόνα 4.23 Πλεγματική μορφή patches 64x64 για την αρχική εικόνα (αριστερά) και την εικόνα με τα περιγράμματα των κτιρίων (δεξιά)	55
Εικόνα 4.24 Πλεγματική μορφή patches 128x128 για την αρχική εικόνα (αριστερά) και την εικόνα με τα περιγράμματα των κτιρίων (δεξιά)	56
Εικόνα 4.25 Πλεγματική μορφή patches 256x256 για την αρχική εικόνα (αριστερά) και την εικόνα με τα περιγράμματα των κτιρίων (δεξιά)	56
Εικόνα 4.26 Πλεγματική μορφή ground truth επισημασμένων κτιρίων	57
Εικόνα 4.27 Patches 256x256 με τα αντίστοιχα επισημασμένα κτίρια.	58
Εικόνα 4.28 Patches 256x256 με τα αντίστοιχα επισημασμένα κτίρια	58
Εικόνα 4.29 Random patches 64x64 με τα αντίστοιχα επισημασμένα περιγράμματα	58
Εικόνα 4.30 Random patches 128x128 με τα αντίστοιχα επισημασμένα περιγράμματα	59
Εικόνα 4.31 Random patches 256x256 με τα αντίστοιχα επισημασμένα περιγράμματα	59
Εικόνα 4.32 Random patches 64x64 με τα αντίστοιχα επισημασμένα κτίρια	59
Εικόνα 4.33 Random patches 128x128 με τα αντίστοιχα επισημασμένα κτίρια	60
Εικόνα 4.34 Random patches 256x256 με τα αντίστοιχα επισημασμένα κτίρια	60
Εικόνα 4.35 α) αρχικό patch 64x64, β) στροφή 90°, γ) στροφή 180°, δ) στροφή 270°	61
Εικόνα 4.36 α) οριζόντια αναστροφή patch, β) στροφή 90°, γ) στροφή 180°, δ) στροφή 270°	61
Εικόνα 4.37 α) patch με κάθετη αναστροφή, β) στροφή 90°, γ) στροφή 180°, δ) στροφή 270°	61
Εικόνα 4.38 Αρχιτεκτονική U-Net	63
Εικόνα 4.39 Περιγράμματα, grid patches, size 64x64.	64
Εικόνα 4.40 Περιγράμματα, grid patches, size 64x64	64
Εικόνα 4.41 Περιγράμματα, grid patches, size 128x128	64
Εικόνα 4.42 Περιγράμματα, grid patches, size 128x128	65
Εικόνα 4.43 Περιγράμματα, grid patches, size 256x256	65
Εικόνα 4.44 Περιγράμματα, grid patches, size 256x256	65
Εικόνα 4.45 Αποτυπώματα, grid patches, size 64x64	66
Εικόνα 4.46 Αποτυπώματα, grid patches, size 64x64	66
Εικόνα 4.47 Αποτυπώματα, grid patches, size 128x128	66

Εικόνα 4.48 Αποτυπώματα, grid patches, size 128x128	66
Εικόνα 4.49 Αποτυπώματα, grid patches, size 256x256	67
Εικόνα 4.50 Αποτυπώματα, grid patches, size 256x256	67
Εικόνα 4.51 Περιγράμματα, random patches, size 64x64	68
Εικόνα 4.52 Περιγράμματα, random patches, size 64x64	68
Εικόνα 4.53 Περιγράμματα, random patches, size 128x128	68
Εικόνα 4.54 Περιγράμματα, random patches, size 128x128	68
Εικόνα 4.55 Περιγράμματα, random patches, size 256x256	69
Εικόνα 4.56 Περιγράμματα, random patches, size 256x256	69
Εικόνα 4.57 Αποτυπώματα, random patches, size 64x64	69
Εικόνα 4.58 Αποτυπώματα, random patches, size 64x64	70
Εικόνα 4.59 Αποτυπώματα, random patches, size 128x128	70
Εικόνα 4.60 Αποτυπώματα, random patches, size 128x128	70
Εικόνα 4.61 Αποτυπώματα, random patches, size 256x256	70
Εικόνα 4.62 Αποτυπώματα, random patches, size 256x256	71

1 Εισαγωγή

Ο τοπογραφικός ιστορικός χάρτης είναι ένα είδος χάρτη, το οποίο απεικονίζει γεωγραφικές πληροφορίες και λεπτομέρειες για την έκταση μιας περιοχής σε συγκεκριμένο χρονικό διάστημα στο παρελθόν. Στις περισσότερες περιπτώσεις, τέτοιου είδους χάρτες δημιουργούνται και αξιοποιούνται ώστε να καταγράψουν τόσο τα φυσικά όσο και τα ανθρωπογενή χαρακτηριστικά μιας περιοχής. Αναφορικά με τα πρώτα ενδεικτικά αναφέρεται η τοπογραφία, η μορφολογία του εδάφους, τα ποτάμια και οι λίμνες. Σχετικά με τα δεύτερα θα μπορούσαν να αναπαρίστανται οι οδοί, τα κτίρια, οι καλλιέργειες, οι οικισμοί και οι πόλεις. Οι τοπογραφικοί ιστορικοί χάρτες δύνανται να παρέχουν πολύ σημαντικές πληροφορίες για τις αλλαγές που έχουν συμβεί στη γεωγραφία μιας περιοχής κατά το πέρασμα του χρόνου, καθώς και για την ανάπτυξη και εξέλιξή της σε ιστορικό επίπεδο. Οι εν λόγω χάρτες μπορούν να αποτελέσουν πηγή πληροφορίας για διάφορους τομείς, όπως η ανθρωπολογία, η έρευνα, η αρχαιολογία, η ιστορία των γεωγραφικών μελετών, καθώς και για προσωπική και εκπαιδευτική χρήση. Σε πολλές περιπτώσεις οι τοπογραφικοί ιστορικοί χάρτες οπτικοποιούν πληροφορίες που χαρακτηρίζονται από λεπτομέρεια και κάνουν χρήση διαφορετικών συμβολισμών ώστε να αναπαραστήσουν τις διάφορες πτυχές του τοπίου. Στην εποχή μας, εξαιτίας της εκτεταμένης χρήσης τόσο της τεχνολογίας όσο και των αλγορίθμων επεξεργασίας εικόνας, πλέον παρέχεται η δυνατότητα εκμετάλλευσης των πληροφοριών αυτών. Έτσι, έχουν αναπτυχθεί μέθοδοι αυτόματης εξαγωγής και ανάλυσης των τοπογραφικών πληροφοριών από τους ιστορικούς χάρτες, με τη βοήθεια νευρωνικών δικτύων και αλγορίθμων σημασιολογικής κατάταξης εικόνας.

Τα νευρωνικά δίκτυα είναι στην ουσία συστήματα υπολογιστικής μάθησης, τα οποία λειτουργούν με πρότυπο τον τρόπο λειτουργίας του ανθρώπινου εγκεφάλου. Εκείνα αποτελούνται από τους «νευρώνες», οι οποίοι επικοινωνούν μεταξύ τους μέσω συνδέσεων. Τέτοιου είδους δίκτυα εκπαιδεύονται με στόχο την αναγνώριση μοτίβων και την πραγματοποίηση τόσο προβλέψεων όσο και κατηγοριοποιήσεων βάσει των δεδομένων που τους τροφοδοτούνται. Οι σύγχρονες προσεγγίσεις των νευρωνικών δικτύων, οι οποίες είναι ευρέως γνωστές ως «βαθιά μάθηση», έχουν καταφέρει πολύ σημαντικά αποτελέσματα σε αρκετούς κλάδους, όπως αυτόματη ανάλυση ιατρικών εικόνων, τα αυτόνομα συστήματα οδήγησης και αναγνώριση εικόνων. Τα νευρωνικά δίκτυα αποτελούν ένα εργαλείο βαρύνουσας σημασίας στον τομέα της τεχνητής νοημοσύνης και της μηχανικής μάθησης. Η χρήση ενός νευρωνικού δικτύου σε έναν ιστορικό χάρτη μπορεί να έχει ως στόχο την αυτόματη αναγνώριση και επεξεργασία των πληροφοριών που εκείνος περιέχει. Τέτοιου είδους χάρτες περιλαμβάνουν συνήθως πληθώρα πληροφοριών. Τέτοιες μπορεί να είναι, όπως έχει ήδη λεχθεί, κτίρια, τοπογραφικά στοιχεία, ποτάμια, οδοί και άλλα γεωγραφικά στοιχεία, που χαρακτηρίζονται από πολυπλοκότητα και δυσκολία τόσο στην αναγνώριση όσο και την επεξεργασία με το χέρι. Είναι απαραίτητο να επισημανθεί ότι κατά την επεξεργασία ενός ιστορικού χάρτη, ένα νευρωνικό δίκτυο μπορεί να μάθει να αναγνωρίζει και να εξαγει αυτόματα πληροφορίες από τον χάρτη. Το σύστημα παρέχει τη δυνατότητα αναγνώρισης και καταγραφής της παρουσίας και της θέσης των κτιρίων στο χάρτη μέσω των τεχνικών επεξεργασίας εικόνας σε συνδυασμό με κατάλληλα δομημένα νευρωνικά δίκτυα. Αυτό μπορεί να είναι χρήσιμο για διάφορους σκοπούς, όπως η μελέτη των αλλαγών του τοπίου

μέσα στον χρόνο, η ανάλυση της αστικής ανάπτυξης ή ακόμα και η αποκατάσταση ιστορικών κατασκευών.

Λαμβάνοντας υπόψη όλα τα προαναφερθέντα εύκολα ανάγεται το συμπέρασμα ότι ένα νευρωνικό δίκτυο συνδυαζόμενο με την επεξεργασία εικόνας μπορεί να καταστήσει δυνατή την αυτόματη αναγνώριση και επεξεργασία των γεωγραφικών πληροφοριών που περιέχονται σε ιστορικούς χάρτες, καθιστώντας την ανάλυση τους αποτελεσματικότερη, ακριβέστερη και ταχύτερη. Ένα εργαλείο επισημάνσης Ground Truth είναι απαραίτητο για την αυτόματη αναγνώριση πληροφοριών σε έναν χάρτη για αρκετούς λόγους. Χρησιμοποιείται για τη δημιουργία ενός συνόλου δεδομένων εκπαίδευσης, τα οποία περιλαμβάνουν παραδείγματα των επιθυμητών πληροφοριών, τα οποία αφενός φέρουν (labels) και αφετέρου αποτελούν το ζητούμενο αναγνώρισης για το ίδιο το μοντέλο. Η δημιουργία του ground truth πραγματοποιείται από ανθρώπους/ειδικούς, οι οποίοι διαλέγουν και επισημαίνουν τις πληροφορίες του χάρτη που είναι απαραίτητες και ορίζουν ένα σύνολο από ζεύγη δεδομένων εισόδου και επιθυμητής εξόδου του συστήματος.

Ένα εργαλείο επισημάνσης αληθών δεδομένων χρησιμοποιείται στην προ-επεξεργασία των εικόνων πριν το στάδιο της εκπαίδευσης του μοντέλου. Κατά την διάρκεια αξιολόγησης του μοντέλου γίνεται σύγκριση των αποτελεσμάτων του με το ground truth. Έτσι καθίσταται εφικτή η ακριβής μέτρηση και η αποτίμηση της απόδοσης του. Αναφορικά με τη βελτίωση του μοντέλου, αυτή γίνεται βάσει των ετικετών που έχουν επισημανθεί. Με τον τρόπο αυτό αναγνωρίζονται τόσο οι αδυναμίες όσο και τα λάθη του μοντέλου και κατόπιν μπορούμε να προχωρήσουμε στην διόρθωση τους. Επιπρόσθετα, παρέχεται η δυνατότητα προσθήκης περισσότερων παραδειγμάτων στο σύνολο εκπαίδευσης καθώς και η προσαρμογή των παραμέτρων του ίδιου του μοντέλου με στόχο την αύξηση της αποδοτικότητας του. Συνολικά, το εργαλείο επισημάνσης Ground Truth διαδραματίζει σημαντικό ρόλο στην ανάπτυξη και εκπαίδευση ενός αυτόματου συστήματος αναγνώρισης πληροφοριών σε έναν χάρτη. Τέλος, αξιολογείται για τη δημιουργία εκπαιδευτικού συνόλου δεδομένων, την αξιολόγηση της απόδοσης και τη βελτίωση του μοντέλου.

Σκοπός της παρούσας διπλωματικής εργασίας είναι ο αυτόματος προσδιορισμός περιγραμμάτων και αποτυπωμάτων κτιρίων σε ιστορικούς πολεοδομικούς χάρτες. Αυτό επιτεύχθηκε χρησιμοποιώντας ένα νευρωνικό δίκτυο βαθιάς μάθησης, το U-Net, το οποίο εκπαιδεύτηκε με ποικίλα σετ δεδομένων που περιείχαν επισημασμένα κτίρια. Επιπλέον, αναπτύχθηκε ένα εξειδικευμένο εργαλείο επισημάνσης, σχεδιασμένο ειδικά για τη διευκόλυνση της διαδικασίας επισημάνσης των κτιρίων σε τοπογραφικές εικόνες. Η επίδοση του προτεινόμενου νευρωνικού δικτύου αξιολογήθηκε με βάση 12 σετ δεδομένων, τα οποία διαφέρουν σε όρους δειγματοληψίας (πλεγματική ή τυχαία) και στην επιθυμητή έξοδο του δικτύου, που περιλαμβάνει τόσο τα περιγράμματα όσο και το συνολικό αποτύπωμα των κτιρίων. Υπολογίστηκαν διάφορες μετρικές για την αξιολόγηση της αποτελεσματικότητας του δικτύου, και εντοπίστηκαν οι βέλτιστες παράμετροι του συστήματος που παρήγαγαν τα καλύτερα αποτελέσματα για τον προσδιορισμό των περιγραμμάτων και των αποτυπωμάτων των κτιρίων στους τοπογραφικούς χάρτες.

Η δομή της διπλωματικής εργασίας έχει ως εξής: στο δεύτερο κεφάλαιο πραγματοποιείτε μια βιβλιογραφική ανασκόπηση του θεωρητικού υποβάθρου των Τεχνητών Νευρωνικών Δικτύων. Στο τρίτο κεφάλαιο αναφέρονται τα βασικά στάδια προ-επεξεργασίας μιας εικόνας για την εξαγωγή των αποτυπωμάτων των κτιρίων και την εκπαίδευση και αξιολόγηση του συνελκτικού νευρωνικού δικτύου. Στο τέταρτο κεφάλαιο αναπτύσσεται η μεθοδολογία που ακολουθήθηκε για την δημιουργία του Annotation Tool, την εκπαίδευση του νευρωνικού

δικτύου καθώς και τα πειραματικά αποτελέσματα. Τέλος στο πέμπτο κεφάλαιο παρουσιάζονται τα συμπεράσματα που προέκυψαν καθώς και προτάσεις για μελλοντικές βελτιώσεις.

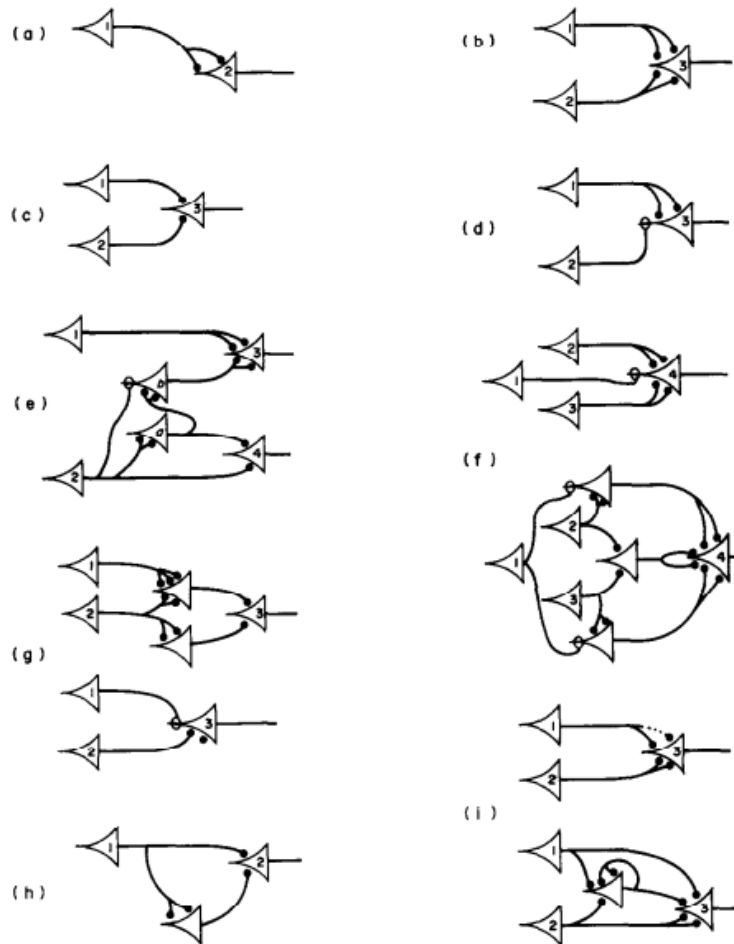
2 Νευρωνικά Δίκτυα

2.1 Εισαγωγή

Το ανθρώπινο είδος ονομάζεται *Homosapiens*, «ο σοφός άνθρωπος», επειδή η σοφία-νοημοσύνη του έχει συμβάλει σημαντικά στην επιβίωση του. Ορμώμενοι από την παραπάνω παραδοχή, τα τελευταία χρόνια καταβάλλεται προσπάθεια για να γίνει κατανοητό πώς ο ανθρώπινος εγκέφαλος, μπορεί να αντιληφθεί, να κατανοήσει, να προβλέψει και να χειριστεί έναν κόσμο πολύ μεγαλύτερο και πιο πολύπλοκο από τον δικό του. Ο τομέας της τεχνητής νοημοσύνης (*Artificial Intelligence*, AI), ή TN, ασχολείται με τη δημιουργία ευφών συστημάτων - μηχανών που έχουν την δυνατότητα όχι μόνο να κατανοούν, αλλά και να βρίσκουν τρόπους να ενεργούν αποτελεσματικά και με ασφάλεια σε μια ποικιλία νέων καταστάσεων. Η τεχνητή νοημοσύνη συγκεντρώνει ένα ευρύ φάσμα από γενικούς τομείς όπως η μάθηση, η συλλογιστική και η αντίληψη και έχει πλήθος παραδοσιακών αλλά και σύγχρονων εφαρμογών όπως τα παίγνια, η απόδειξη μαθηματικών θεωρημάτων, η σύνθεση φυσικής γλώσσας, η οδήγηση αυτοκινήτου και η διάγνωση ασθενειών [1].

2.2 Ιστορική αναδρομή

Ο νευροφυσιολόγος Warren McCulloch και ο μαθηματικός Walter Pitts (1943) διεξήγαγαν την πρώτη έρευνα τεχνητής νοημοσύνης, στηριζόμενοι σε τρεις πηγές. Η πρώτη ήταν η βασική φυσιολογία και η λειτουργία των νευρώνων του εγκεφάλου. Η δεύτερη σχετιζόταν με την τυπική ανάλυση της προτασιακής λογικής των Russell και Whitehead, ενώ η τρίτη ανάγεται στη θεωρία υπολογισμού του Turing. Βασίζόμενοι λοιπόν στον ορισμό του νευρώνα, οι McCulloch και Pitts παρουσίασαν για πρώτη φορά την ιδέα ότι ένα νευρωνικό δίκτυο αποτελείται από ένα μεγάλο αριθμό νευρώνων και πώς οι τελευταίοι μπορούν να λειτουργούν με τις δικές τους συνδέσεις. Αν αντιληφθούμε τα νευρωνικά δίκτυα ως ένα φαινόμενο το οποίο θα μπορούσε να παρουσιαστεί, ως προς την ιστορική του εξέλιξη, σε εικόνες, η παρακάτω θα ήταν η πρώτη.



Εικόνα 2.1 Πρώτη παρουσίαση ενός τεχνητού νευρώνα [2]

Σύμφωνα με αυτό το μοντέλο ο κάθε τεχνητός νευρώνας μπορεί να δεχθεί μόνο δυαδικές τιμές εισόδου δηλαδή 0 (μηδέν) ή 1 (ένα). Στο εσωτερικό του νευρώνα μπορούν να λάβουν χώρα διάφορες διαδικασίες. Αρχικά, υπολογίζονται δύο συναρτήσεις, από τις οποίες η πρώτη υπολογίζει το άθροισμα όλων των τιμών εισόδου και η δεύτερη καταλήγει στο αν το άθροισμα αυτό είναι μεγαλύτερο από μια ορισμένη τιμή κατωφλίου. Εάν είναι μεγαλύτερο, ο νευρώνας εξάγει την τιμή 1. Διαφορετικά, ο νευρώνας εξάγει την τιμή 0. Έτσι, δεδομένης της δομής του δικτύου, κάθε νευρώνας, είτε είναι ενεργός (1) είτε ανενεργός (0), μεταβιβάζει την αντίστοιχη τιμή σε έναν ή πολλούς άλλους νευρώνες. Έτσι, αποδείχθηκε αφενός ότι οποιαδήποτε υπολογίσιμη συνάρτηση μπορεί να υπολογιστεί από ένα τέτοιο δίκτυο νευρώνων και αφετέρου ότι όλες οι λογικές πράξεις (και, ή, όχι, κ.λπ.) μπορούν να μοντελοποιηθούν με την τοποθέτηση μιας κατάλληλης τιμής κατωφλίου.

Το 1949 ο Donald Hebb παρουσίασε έναν κανόνα μάθησης, ο οποίος παραμένει σημαντικός ως σήμερα. Το μοντέλο μάθησης «Hebb» τροποποιεί τις συνδέτικες συνδέσεις μεταξύ των νευρώνων. Στην ουσία πρόκειται για έναν κανόνα βάσει του οποίου αποδυναμώνονται ή ενδυναμώνονται αυτές οι συνδέσεις, δίνοντας λύση σε ένα βασικό μειονέκτημα της πρότασης των Pitts και McCulloch. Έναν χρόνο αργότερα οι Marvin Minsky και Dean Edmonds κατασκεύασαν τον πρώτο υπολογιστή νευρωνικού δικτύου. Εκείνος χρησιμοποιούσε 3000 λυχνίες κενού και μηχανισμό αυτόματου πιλότου από βομβαρδιστικό B-24 για να προσομοιώσει 40 νευρώνες. Οι Newell και Simon παρουσίασαν ένα μαθηματικό σύστημα για την απόδειξη θεωρημάτων, το οποίο έμεινε γνωστό ως «Logic Theorist (LT)». Λίγο μετά, το πρόγραμμα αυτό μπόρεσε να αποδείξει τα περισσότερα από τα θεωρήματα των Principia

Mathematica, των Russell και Whitehead. Οι επιτυχίες των Newell και Simon μετά το LT συνεχίστηκαν με το πρόγραμμα «General Problem Solver» (GPS). Σε αντίθεση με το πρώτο, το δεύτερο πρόγραμμα τους σχεδιάστηκε εξ αρχής για να μιμηθεί τα πρωτόκολλα επίλυσης ανθρώπινων προβλημάτων. Επιπλέον, η σειρά με την οποία το πρόγραμμα εξέταζε τόσο τους επιμέρους στόχους όσο και τις πιθανές ενέργειες σε έναν περιορισμένο τύπο γρίφων ήταν παρόμοια με τον τρόπο με τον οποίο οι άνθρωποι αντιμετωπίζουν τα ίδια προβλήματα. Έτσι, το GPS ήταν πιθανότατα το πρώτο πρόγραμμα που ενσωμάτωσε την προσέγγιση της «ανθρώπινης σκέψης». Οι Newell και Simon προέβαλαν την περίφημη υπόθεση του συστήματος φυσικών συμβόλων και επισήμαναν ότι κάθε σύστημα (ανθρώπινο ή μηχανικό) που εκφράζει νοημοσύνη πρέπει να λειτουργεί χειριζόμενο δομές δεδομένων που αποτελούνται από σύμβολα. Αν κρίνουμε τα ερευνητικά έργα της περιόδου αυτής μακροπρόθεσμα, ίσως να εξαγόταν το συμπέρασμα ότι το σημαντικότερο ήταν το παιχνίδι ντάμας του Arthur Samuel. Το «πείραμα» του κατάφερε να διδαχθεί να παίζει ντάμα σε ερασιτεχνικό επίπεδο χρησιμοποιώντας μια μέθοδο που τώρα ονομάζεται ενισχυτική μάθηση (Reinforcement Learning). Το πρόγραμμά σύντομα έμαθε να παίζει καλύτερα από τον δημιουργό του.

Το 1958 ο John McCarthy στο υπόμνημά του. MIT AI Lab Memo No. 1, όρισε τη γλώσσα υψηλού επιπέδου Lisp. Η τελευταία θα εξελισσόταν σε κυρίαρχη γλώσσα προγραμματισμού για την τεχνητή νοημοσύνη για τα επόμενα 30 χρόνια. Ο ίδιος δημοσίευσε μια εννοιολογική πρόταση σχετικά με τα συστήματα TN, η οποία βασιζόταν στη γνώση και τη συλλογιστική. Το Advice Taker, όπως ονομάστηκε, αποτελεί ένα υποθετικό πρόγραμμα που αφενός θα περιελάμβανε τη γενική γνώση του κόσμου και αφετέρου θα είχε τη δυνατότητα να την αξιοποιήσει ώστε να παράγει σχέδια δράσης. Επιπλέον, το εν λόγω πρόγραμμα μπορούσε να δέχεται νέα αξιώματα, κατά την κανονική του λειτουργία, γεγονός που καθιστούσε δυνατή την απόκτηση δεξιοτήτων σε νέους για εκείνο τομείς χωρίς τον επαναπρογραμματισμό του. Η συγκεκριμένη δημοσίευση επηρέασε την εξέλιξη της TN και παραμένει επίκαιρη ως και σήμερα.

Τόσο οι πρώτες εργασίες που βασίστηκαν στα νευρωνικά δίκτυα όσο και οι μέθοδοι μάθησης συνέχισαν να ευδοκιμούν. Χαρακτηριστικό παράδειγμα αυτής της ενίσχυσης αποτελούν τα δίκτυα adalines του Bernie Widrow και τα perceptron του Frank Rosenblatt. Μάλιστα, σύμφωνα με το θεώρημα της σύγκλισης των τελευταίων, οι αλγόριθμοι μάθησης έχουν την δυνατότητα προσαρμογής της ισχύος των συνδέσεων ενός perceptron για να συμφωνούν με όλα τα δεδομένα εισόδου, με την προϋπόθεση, όμως, ότι υπάρχει αυτή η αντιστοίχιση.

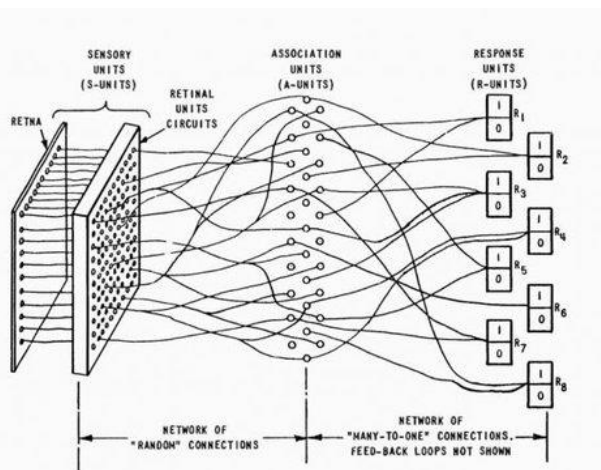
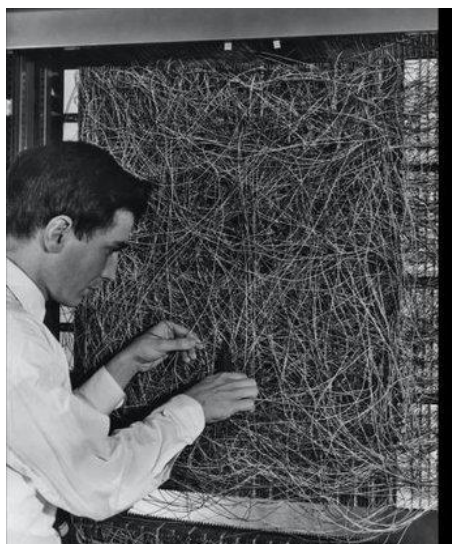


Figure 1 ORGANIZATION OF THE MARK I PERCEPTRON

Εικόνα 2.2 Ο Rosenblatt με το Mark I Perceptron -Ένα νευρωνικό δίκτυο που πραγματοποιούσε οπτικές ταξινομήσεις (Αριστερά), Μια οπτική αναπαράσταση του (Δεξιά) [3]

Ορμώμενοι από την προαναφερθείσα καλπάζουσα εξέλιξη, οι ερευνητές της TN εμφανίζονταν ιδιαίτερα αισιόδοξοι για την εξέλιξη και τα πιθανά επιτεύγματα των συστημάτων. Παρά ταύτα, αυτή η αισιοδοξία τους δεν επιβεβαιώθηκε από την πραγματικότητα της εποχής. Ως επιβεβαίωση για την παραπάνω διαπίστωση θα αξιοποιηθούν τα πρώτα πειράματα στον κλάδο της μηχανικής εξέλιξης, δηλαδή τον σημερινό γενετικό προγραμματισμό. Εκείνα βασίστηκαν περισσότερο στον ενθουσιασμό της εποχής παρά στις πραγματικές δυνατότητες τους. Κεντρική ιδέα αποτελούσε η δοκιμή τυχαίων μεταλλάξεων μέσω μίας διαδικασίας επιλογής, η οποία θα αξιοποιούσε μόνο εκείνες που φαίνονταν χρήσιμες. Η αποτυχία βασίστηκε σε τρία προβλήματα. Αρχικά, δεν σημειώθηκε καμία πρόοδος, παρά τις ατελείωτες ώρες χρόνου CPU. Στη συνέχεια, παρουσιάστηκε η αδυναμία επίλυσης του ζητήματος της «συνδυαστικής έκρηξης», παράγοντας που δέχτηκε πληθώρα κριτικών. Το τρίτο αίτιο αποτυχίας οφειλόταν στους περιορισμούς των βασικών δομών που αξιοποιήθηκαν, όπως για παράδειγμα το ότι ένα perceptron δύο εισόδων αδυνατούσε να μάθει να αναγνωρίζει πότε δύο εισοδοί του ήταν ανόμοιες. Όλα τα παραπάνω είχαν ως απόρροια την διακοπή της χρηματοδότησης των ερευνών για τα νευρωνικά δίκτυα.

Από τα τέλη της δεκαετίας του 1960 ως και τις αρχές του 1980 έγιναν προσπάθειες αναζωπύρωσης του ενδιαφέροντος για την TN. Από τις «ασθενείς μεθόδους» και το πρόγραμμα Dendral ως τα «έμπειρα συστήματα» και τα προγράμματα Mycin, R1, Shrdlu, Planner καθώς και με τη συνεισφορά των Charniac, Schank και άλλων σημαντικών ερευνητών, η TN κατάφερε όχι μόνο να ανακάμψει αλλά και να αποτελέσει πόλο ερευνητικής έλξης. Συνολικά, κατά το διάστημα 1980-1988, ο τομέας της TN γνώρισε ιλιγγιώδη ανάπτυξη, η οποία χρηματικά ισούταν με δισεκατομμύρια. Χαρακτηριστικό της έξαρσης αυτής είναι πως εκατοντάδες εταιρίες άρχισαν να κατασκευάζουν έμπειρα συστήματα, ρομπότ, συστήματα όρασης και εξειδικευμένο υλικό και λογισμικό για τους κλάδους αυτούς. Η εξέλιξη αυτή σύντομα αντικαταστάθηκε από μια περίοδο ύφεσης, γνωστή και ως «Χειμώνας της TN» (AI Winter), οπότε πληθώρα εταιριών άρχισε να γνωρίζει την παρακμή.

Αν και στις αρχές της δεκαετίας του 1960 αναπτύχθηκε για πρώτη φορά ο αλγόριθμος μάθησης με οπισθοδιάδοση (Back-Propagation), φαίνεται πως κατά τη περίοδο αυτή αρχίζει να αξιοποιείται πραγματικά. Πιο συγκεκριμένα, εκείνος εφαρμόστηκε σε αρκετά προβλήματα

μάθησης σε διάφορα πεδία (υπολογιστές, ψυχολογία) και λόγω της ευρείας διάδοσης του προσέλυσε σημαντικό ενδιαφέρον. Ταυτόχρονα, κάνουν την εμφάνισή τους τα «συνδεδεμένα μοντέλα» ευφυών συστημάτων. Τα τελευταία θεωρήθηκαν από ορισμένους ερευνητές ως ανταγωνιστές τόσο των συμβολικών μοντέλων όσο και της λογιστικής προσέγγισης. Εκείνα σχηματίζουν εσωτερικές έννοιες οι οποίες όμως μεταβάλλονται. Αυτή η μεταβλητότητα τους, σχετικά με τον τρόπο δημιουργίας τους, ταιριάζει καλύτερα στον πραγματικό κόσμο και την πολυπλοκότητα του. Η «ελαστικότητα» που επικρατούσε στις προηγούμενες περιόδους, όσον αφορά τη μεθοδολογία και τα παραδείγματα που αξιοποιούνταν, κατέληξε την εν λόγω εποχή σε μια αυστηροποίηση τους. Ενδεικτικό της τάσης αυτής είναι τα «κρυφά μοντέλα Markov», στον τομέα της αναγνώρισης ομιλίας. Εκείνα αφενός βασίζονται σε μια αυστηρή μαθηματική θεωρία και αφετέρου παράγονται με μια διαδικασία εκπαίδευσης, η οποία έχει ως βάση της μια εκτεταμένη συλλογή πραγματικών δεδομένων ομιλίας. Το γεγονός αυτό εξασφάλιζε ότι η απόδοση θα είναι η επιθυμητή και ότι σε αυστηρές τυφλές δοκιμές τα μοντέλα θα βελτιώνουν την απόδοσή τους. Ως αποτέλεσμα, ο συγκεκριμένος τομέας εξελίχθηκε.

Κατά το 1988 η TN συνδέθηκε με πολλούς τομείς, όπως η θεωρία ελέγχου, η επιχειρησιακή έρευνα, η θεωρία αποφάσεων και η στατιστική. Παράλληλα, αναπτύσσονται τα δίκτυα Bayes από τον Judea Pearl, εξέλιξη η οποία οδήγησε σε μια αυστηρή και αποδοτική τυποποίηση της αναπαράστασης βέβαιης γνώσης, καθώς και σε πρακτικούς αλγορίθμους για την πιθανοτική συλλογιστική. Ένα εξίσου σημαντικό γεγονός που συνέβη την ίδια χρονία ήταν η εργασία του Rich Sutton, η οποία συνέδεε την ενισχυτική μάθηση με τη θεωρία των διαδικασιών αποφάσεων Markov (Markov Decision Process-MDP) που γνώρισε ιδιαίτερη ανάπτυξη στον τομέα της επιχειρησιακής έρευνας. Εν συνεχεία, ακολούθησαν πολλές εργασίες που πραγματεύονταν τη σύνδεση της έρευνας σχεδιασμού TN με διαδικασίες MDP. Έτσι, η ενισχυτική μάθηση ήρθε σε επαφή με την ρομποτική με σκοπό τον έλεγχο διαδικασιών, ενώ ταυτόχρονα απέκτησε βαθύτερο θεωρητικό υπόβαθρο. Αυτή η νέα εκτίμηση προς την TN, αναφορικά με τα δεδομένα, τη μοντελοποίηση, τη στατιστική, τη μηχανική μάθηση και την βελτιστοποίηση, είχε ως συνέπεια τη σταδιακή επανένωση υποπεδίων, όπως η ρομποτική, τα πολυπρακτορικά συστήματα, η μηχανική όραση, η αναγνώριση ομιλίας και η επεξεργασία φυσικής γλώσσας που είχαν αποσχισθεί έως ένα βαθμό από τον πυρήνα της TN. Αυτή η επανένταξη έχει ωφελήσει σημαντικά τόσο αναφορικά με τις εφαρμογές τους όσο και προς μια καλύτερη θεωρητική κατανόηση των βασικών προβλημάτων της TN.

Πολλοί και σημαντικοί παράγοντες έχουν συμβάλει στη δημιουργία πολύ μεγάλων συνόλων δεδομένων, τα οποία είναι γνωστά ως «μεγα-δεδομένα» (Big Data) και περιλαμβάνουν τρισεκατομμύρια λέξεις, δισεκατομμύρια εικόνες και δεδομένα τα οποία έχουν συλλεχθεί από διάφορες πηγές όπως π.χ. τα κοινωνικά δίκτυα. Αυτά, με τη σειρά τους, έχουν οδηγήσει στην ανάπτυξη αλγορίθμων μάθησης, οι οποίοι έχουν σχεδιαστεί ειδικά για να εκμεταλλεύονται τεράστια σύνολα δεδομένων. Τόσο η διαθεσιμότητα μεγα-δεδομένων όσο και η μετάβαση προς τη μηχανική μάθηση συνέβαλλαν στο να ανακτήσει η TN την εμπορική της ελκυστικότητα. Τα μεγα-δεδομένα συνέβαλλαν σημαντικά στη νίκη του συστήματος Watson της IBM απέναντι σε πολύ σημαντικά πρόσωπα στο παιχνίδι γνώσεων Jeopardy, το 2011, γεγονός που είχε σημαντικό αντίκτυπο στην κοινή αντίληψη για την TN.

Το 2011 άρχισαν να χρησιμοποιούνται ευρύτατα οι μέθοδοι βαθιάς μάθησης αρχικά στον τομέα της αναγνώρισης ομιλίας και έπειτα στην οπτική αναγνώριση αντικειμένων. Βέβαια, τέτοιου είδους δίκτυα είχαν αξιοποιηθεί σε πειράματα ήδη από το 1970, ενώ τη δεκαετία του 1990, με τη μορφή συνελκτικών νευρωνικών δικτύων (Convolutional Neural Networks), σημείωσαν επιτυχία στην αναγνώριση χειρόγραφων ψηφίων. Ως όρος, η «βαθιά μάθηση» (Deep Learning) αναφέρεται στην χρήση πολλαπλών επιπέδων αποτελούμενων από

συστοιχίες απλών, προσαρμοσμένων υπολογιστικών στοιχείων. Ένα βασικό χαρακτηριστικό των τεχνικών βαθιάς μάθησης είναι ότι οι υπολογιστικές τους απαιτήσεις είναι ιδιαίτερα αυξημένες και απαιτεί συνήθως εξειδικευμένο υλικό (hardware). Για παράδειγμα, ένας αλγόριθμος βαθιάς μάθησης που εκτελείται σε εξειδικευμένο υλικό μπορεί να πραγματοποιεί μεταξύ 10^{14} και 10^{17} πράξεις ανά δευτερόλεπτο, κυρίως υπό τη μορφή εξαιρετικά πολύπλοκων πράξεων μεταξύ πινάκων και διανυσμάτων. Επιπλέον, όπως έχει ήδη αναφερθεί, η βαθιά μάθηση εξαρτάται σε μεγάλο βαθμό και από τη διαθεσιμότητα μεγάλων ποσοτήτων δεδομένων εκπαίδευσης [1].

2.3 Τεχνητή Νοημοσύνη

Σύμφωνα με τον ορισμό των Charniak και McDermott, (1985) Τεχνητή Νοημοσύνη (Artificial Intelligence), είναι η μελέτη των νοητικών ικανοτήτων με τη χρήση υπολογιστικών μοντέλων». Αναλυτικότερα, είναι η προσπάθεια να προσομοιωθούν οι λειτουργίες του ανθρώπινου εγκεφάλου στον υπολογιστή. Έτσι, μια μερίδα ερευνητών ορίζει την Τεχνητή Νοημοσύνη (TN) ως προς την πιστότητα με την ανθρώπινη απόδοση. Η αντιστοιχία αυτή, αναφέρεται στο πόσο καλά ή ακριβώς μπορεί η TN να μιμηθεί τις ανθρώπινες ικανότητες ή να εκτελέσει καθήκοντα που συνήθως απαιτούν ανθρώπινη νοημοσύνη. Στον αντίποδα, μια άλλη μερίδα προτιμά να χρησιμοποιεί την έννοια της ορθολογικότητας (Rationality) η οποία αναφέρεται στο «να δρα ορθά το υποκείμενο». Επιπρόσθετα, ένα υποσύνολο του ερευνητικού δυναμικού αντιλαμβάνεται τη νοημοσύνη ως μια σύμφυτη διαδικασία σκέψης ή ένα χαρακτηριστικό συλλογισμού. Αντίθετα, κάποιοι ερευνητές εστιάζουν στην ευφυή συμπεριφορά ως ένα μη εγγενές χαρακτηριστικό. Η επίτευξη νοημοσύνης που μοιάζει με την ανθρώπινη απαιτεί πειραματική επιστήμη που σχετίζεται με την ψυχολογία, η οποία παρατηρεί και υποθέτει πραγματικές ανθρώπινες συμπεριφορές και διαδικασίες σκέψης, ενώ η ορθολογική προσέγγιση απαιτεί ένα συνδυασμό μαθηματικών και μηχανικής και σχετίζεται με τη στατιστική, τη θεωρία ελέγχου και τα οικονομικά [1].

2.4 Μηχανική μάθηση

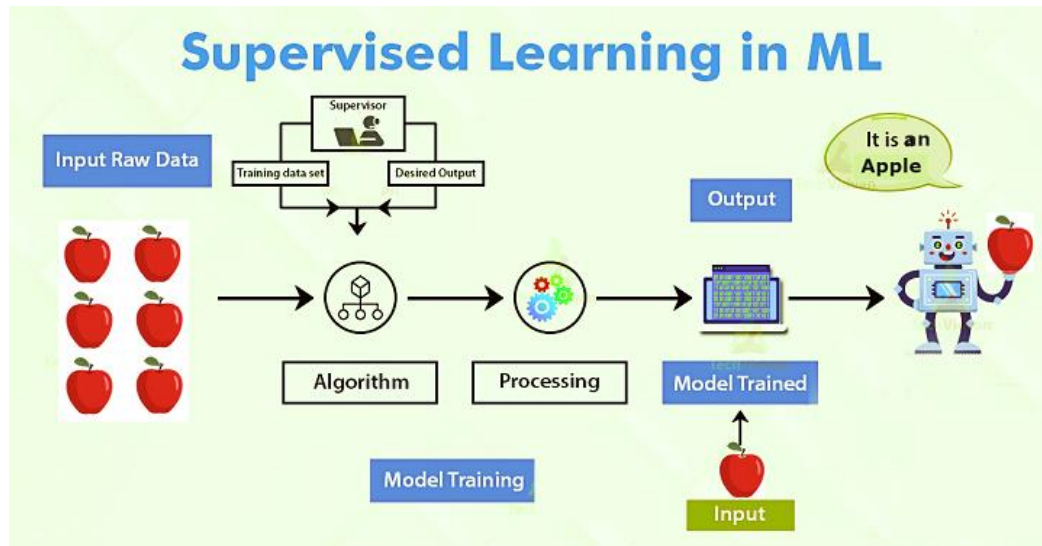
Η μηχανική μάθηση είναι μια υποκατηγορία της TN που εξετάζει τη δυνατότητα βελτίωσης της απόδοσης ενός αλγορίθμου με βάση την εμπειρία. Ειδικότερα, σύμφωνα με τον ορισμό του Arthur Samuel (1959) «Μηχανική Μάθηση (Machine Learning) είναι ο κλάδος που παρέχει στους υπολογιστές την δυνατότητα να μαθαίνουν χωρίς να έχουν προγραμματιστεί λεπτομερώς». Με άλλα λόγια, δίχως να έχουν προγραμματιστεί βάσει αναλυτικών κανόνων ορισμένων και γραμμένων από τους χειριστές. Ένας άλλος χρήσιμος ορισμός, πιο λεπτομερής και περιεκτικός είναι του Tom Mitchell (1997), σύμφωνα με τον οποίο «Ένα πρόγραμμα υπολογιστή λέγεται ότι μαθαίνει από την εμπειρία E αναφορικά με μια εργασία T και ένα μέτρο απόδοσης P , εάν η απόδοση της T , όπως μετράται από το P , βελτιώνεται από την εμπειρία E ».

Λαμβάνοντας υπόψη τα προαναφερθέντα εύκολα καταλήγει κανείς στο ότι η Μηχανική Μάθηση (Machine Learning) αφορά γενικά υπολογιστικά συστήματα ικανά να μαθαίνουν από δεδομένα εκπαίδευσης (training data). Στο σημείο αυτό είναι σκόπιμο να επισημανθεί πως «μάθηση» σημαίνει ότι το σύστημα βελτιώνει την απόδοση του βάσει της γνώσης και της εμπειρίας, η οποία προέρχεται από τα δεδομένα που το τροφοδοτεί ο χρήστης. Ορισμένα συστήματα TN είναι αποδοτικότερα με τη χρήση μεθόδων μηχανικής μάθησης, ενώ άλλα όχι.

Υπάρχουν τρεις κατηγορίες αλγορίθμων μηχανικής μάθησης:

Επιβλεπόμενη Μάθηση (Supervised Learning)

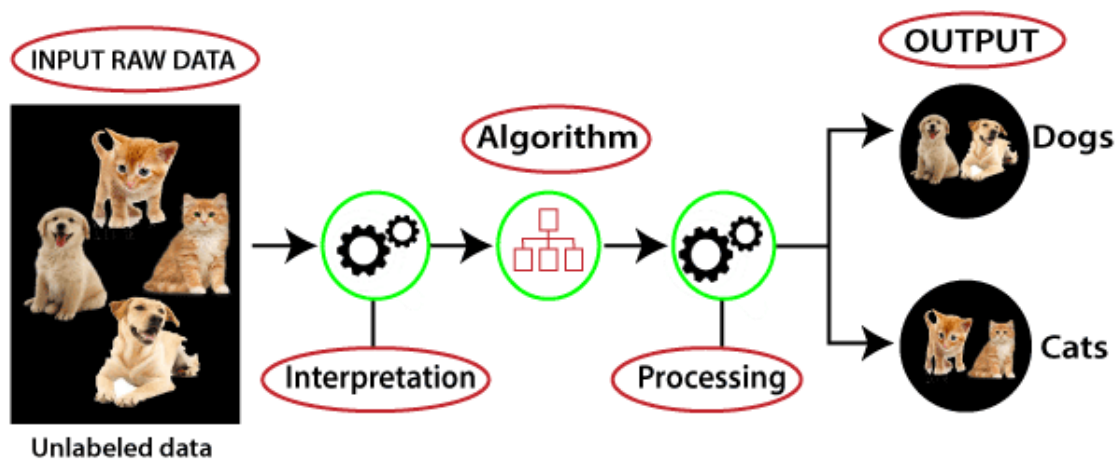
Οι αλγόριθμοι αυτοί εκπαιδεύονται ώστε να δημιουργηθεί μια συνάρτηση απεικόνισης της εισόδου σε κάποια επιθυμητή έξοδο. Η εκπαίδευση πραγματοποιείται με δεδομένα (π.χ. εικόνες) εισόδου, για τα οποία υπάρχουν οι αντίστοιχες επιθυμητές έξοδοι (π.χ. κατηγορίες εικόνων). Ο απώτερος στόχος είναι η γενίκευση της συνάρτησης ώστε να δίνει τα σωστά αποτελέσματα εξόδου και για άγνωστα δεδομένα εισόδου για τα οποία δεν έχει εκπαιδευτεί.



Εικόνα 2.3 Επιβλεπόμενη Μάθηση (Supervised learning) [4]

Μη Επιβλεπόμενη Μάθηση (Unsupervised learning)

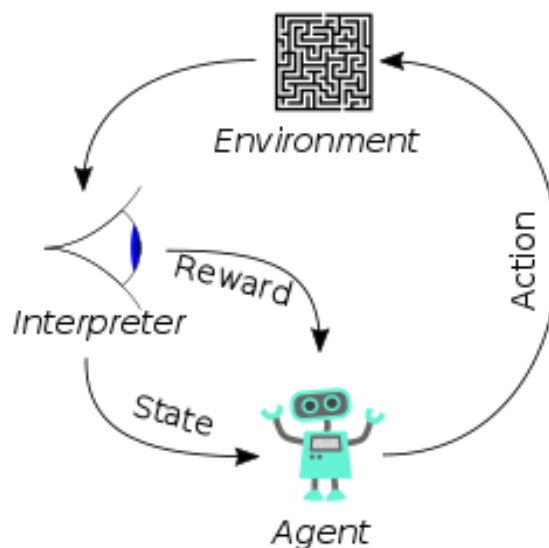
Οι αλγόριθμοι αυτοί εκπαιδεύονται χωρίς είναι γνωστές οι επιθυμητές τιμές εξόδου στις οποίες αντιστοιχούν τα σετ δεδομένων εισόδου. Δημιουργείται έτσι ένα μοντέλο το οποίο βασίζεται σε μοτίβα (π.χ. ομοιότητας).



Εικόνα 2.4 Μη Επιβλεπόμενη Μάθηση (Unsupervised Learning) [5]

Ενισχυτική Μάθηση (Reinforcement Learning)

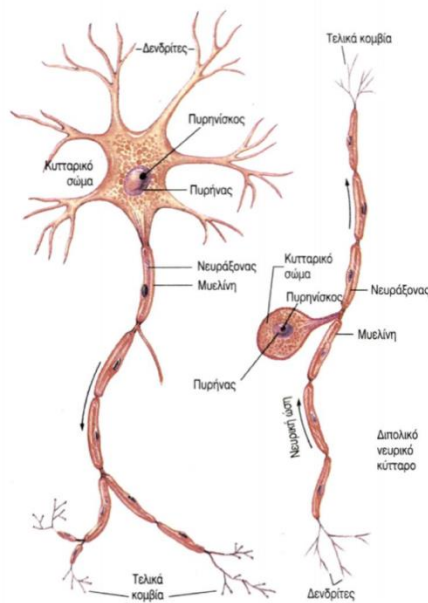
Οι αλγόριθμοι αυτοί μαθαίνουν μια στρατηγική ενεργειών, μέσα από μια άμεση αλληλεπίδραση με το περιβάλλον. Εφόσον η συμπεριφορά είναι επιθυμητή υπάρχει η επιβράβευση, ενώ σε αντίθετη περίπτωση, η τιμωρία.



Εικόνα 2.5 Ενισχυτική Μάθηση (Reinforcement Learning) [6]

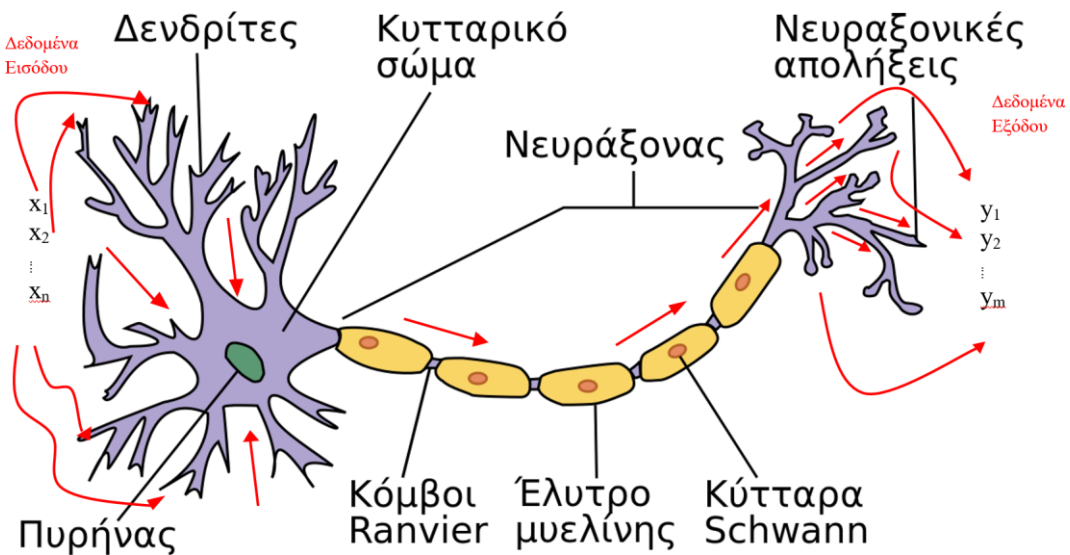
2.5 Τεχνητά νευρωνικά δίκτυα

Η δομή των τεχνητών νευρωνικών δικτύων (Artificial neural networks) έχει εμπνευστεί από την αντίστοιχη του βιολογικού νευρώνα. Ο ανθρώπινος εγκέφαλος αποτελείται από 10^{10} νευρώνες. Νευρώνας ονομάζεται το κύτταρο του εγκεφάλου το οποίο αποτελεί το δομικό και λειτουργικό μέρος του. Τα κύτταρα-νευρώνες- συνδέονται μεταξύ τους μέσω των συνάψεων, δηλαδή τα σημεία που συναντιούνται οι δενδρίτες. Στην απλή περίπτωση, από τις συνάψεις μεταβιβάζεται μια είσοδος, ένα ηλεκτρικό σήμα, το οποίο καταλήγει στον πυρήνα και πραγματοποιείται μια ενέργεια, ένας υπολογισμός. Στην συνέχεια, παράγεται το αποτέλεσμα του υπολογισμού αυτού και μεταβιβάζεται σε μια έξοδο. Η μεταβίβαση της εξόδου στους άλλους νευρώνες πραγματοποιείται μέσω του άξονα. Ένας νευρώνας ενεργοποιείται και μεταδίδει σήματα σε άλλους, ανάλογα με τα σήματα τα οποία δέχεται. Αυτά πρέπει να ξεπεράσουν μια τιμή κατωφλίου ώστε να ενεργοποιηθεί ο νευρώνας.



Εικόνα 2.6 Νευρωνικό Κύτταρο [7]

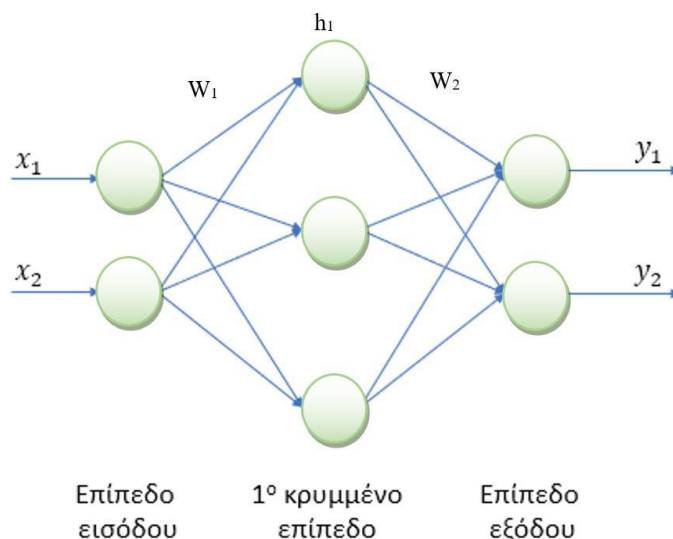
Εφοδιάζοντας με αυτήν την γνώση το νευρωνικό δίκτυο, δημιουργείται η δομή που φαίνεται στην **Εικόνα 2.7**. Οι δενδρίτες αναπαριστούν τα δεδομένα εισόδου (x_i) και τα δεδομένα εξόδου (y_i) του συστήματος. Το σώμα και ο νευροάξονας του κυττάρου πραγματοποιούν τους υπολογισμούς και τις αποφάσεις του συστήματος ώστε να παρθεί ένα αποτέλεσμα. Το κάθε δεδομένο εξόδου (y_i) μεταβιβάζεται μέσω των απολήξεων σε άλλο κύτταρο το οποίο με την σειρά του θα πραγματοποιήσει άλλους υπολογισμούς και θα παρθεί μια άλλη απόφαση. Αυτή η διαδικασία θα συνεχιστεί μέχρις ότου όλα τα κύτταρα έχουν κάνει αυτήν την διαδικασία. Όλα τα κύτταρα μαζί σχηματίζουν ένα δίκτυο κυττάρων, τα οποία βοηθώντας το ένα το άλλο επιλύουν προβλήματα και σχηματίζουν αποφάσεις.



Εικόνα 2.7 Δομή νευρωνικού κυττάρου

Η αντιστοίχιση αυτή με το τεχνητό νευρωνικό δίκτυο έχει την μορφή που φαίνεται στην **Εικόνα 2.8**. Συγκεκριμένα, τα δεδομένα εισόδου (x_i) τροφοδοτούνται στο δίκτυο από το επίπεδο εισόδου (Input Layer), και αφού περάσουν από πολλαπλά "κρυφά επίπεδα" (Hidden

layers) καταλήγουν στην έξοδο (Output Layer) απεικονίζοντας την επιθυμητή πληροφορία (y_i). Ο κάθε νευρώνας μαθαίνει και εκπαιδεύεται από τα βάρη (w_i) τα οποία ορίζονται από τον χρήστη, και παράγουν την τιμή h_i η οποία είναι αυτή η πληροφορία η οποία περνάει στα επόμενα επίπεδα.



Εικόνα 2.8 Δομή τεχνητού νευρωνικού δικτύου

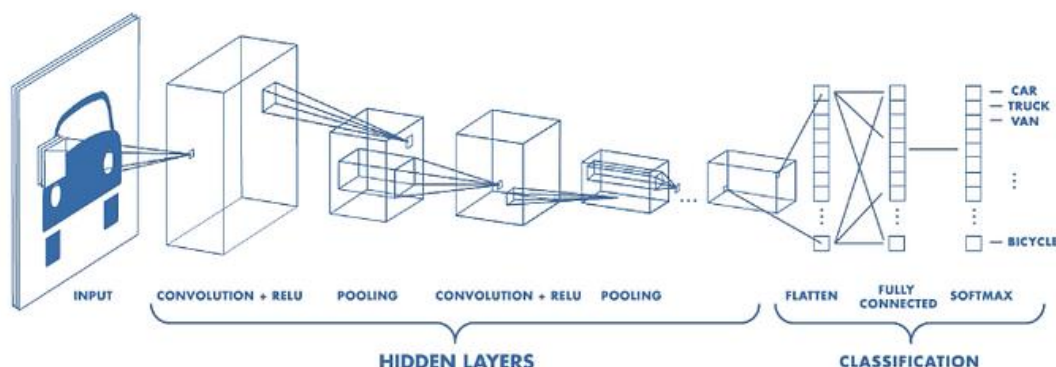
2.6 Συνελκτικά νευρωνικά δίκτυα

Ο όρος βαθιά μάθηση ή βαθιά νευρωνικά δίκτυα αναφέρεται στα νευρωνικά δίκτυα με πολλαπλά επίπεδα (layers). Τις τελευταίες δεκαετίες έχει αναφερθεί ως ένα από τα πιο ισχυρά εργαλεία και έχει αναγνωριστεί ευρέως καθώς μπορεί να διαχειριστεί μεγάλους όγκους δεδομένων. Το πιο διαδεδομένο νευρωνικό δίκτυο είναι το συνελκτικό νευρωνικό δίκτυο (Convolutional Neural Network - CNN). Το δίκτυο αυτό παίρνει το όνομα του από την γραμμική μαθηματική πράξη της συνέλιξης που χρησιμοποιείται μεταξύ των πινάκων (που υπάρχουν τα layers). Ένα CNN αποτελείται από πολλαπλά layers [8] στα οποία περιλαμβάνονται τα εξής: Επίπεδο Εισόδου (Input Layer), Επίπεδο Συνέλιξης (Convolutional Layer), Επίπεδο Συνάρτησης Ενεργοποίησης ReLU (ReLU Layer), Επίπεδο Συγκέντρωσης (Pooling layer), Επίπεδο Ανεστραμμένης Συνέλιξης (Transposed Convolution Layer) και Πλήρως συνδεδεμένο επίπεδο (Fully- Connected layer).

Συνέλιξη (Convolution)

Στην πιο απλή μορφή ενός νευρωνικού δικτύου με ένα νευρώνα δημιουργείται το πρόβλημα της συνδεσιμότητας του κάθε νευρώνα με τον επόμενο. Για παράδειγμα σε μια εφαρμογή όπου η είσοδος είναι μια εικόνα, οι νευρώνες συνδέονται κατευθείαν με την εικόνα εισόδου και δημιουργείται ένας τεράστιος όγκος από παραμέτρους βαρών συνδεσιμότητας (Weight Connections) που χρειάζονται επίλυση και υπολογιστική ισχύ. Η πολυπλοκότητα του προβλήματος και οι παράμετροι διπλασιάζονται όταν προστεθεί ένας δεύτερος νευρώνας. Προφανώς δεν είναι αποτελεσματικό και ούτε χρήσιμο ένα νευρωνικό δίκτυο με δυο νευρώνες καθώς δεν μπορεί να πραγματοποιηθεί καμία εύλογη διαδικασία επεξεργασία. Το ζήτημα αυτό επιλύεται χρησιμοποιώντας αντί για πλήρη συνδεσιμότητα μεταξύ των νευρώνων και της εικόνας εισόδου, η διασύνδεση τοπικών μόνο περιοχών (Local Regions) των εικόνων το οποίο ονομάζεται τοπική συνδεσιμότητα (Local Connectivity).

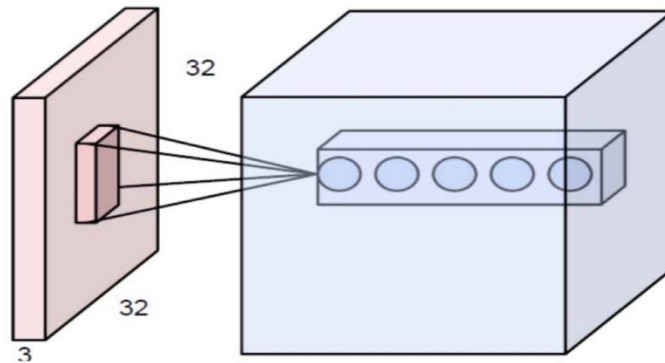
Συγκεκριμένα, οι κρυφοί νευρώνες στα κρυφά layers λαμβάνουν ως εικόνα εισόδου μόνο την αντίστοιχη περιοχή η οποία τροφοδοτεί το προηγούμενο layer, όπως φαίνεται στην **Εικόνα 2.9** [8].



Εικόνα 2.9 Παράδειγμα διαφορετικών επιπέδων όπου το καθένα «κοιτά» σε διαφορετικό παράθυρο της εικόνας. (Τοπική Συνδεσιμότητα) [9]

Τα κρυφά layer είναι εκείνα τα οποία πραγματοποιούν μαθηματικούς υπολογισμούς [10]. Παρόλο που με αυτό τον τρόπο οι παράμετροι συνδεσιμότητας ελαττώθηκαν υπάρχουν και άλλες παράμετροι που χρήζουν επίλυσης. Μια βελτιστοποίηση είναι, τα βάρη της τοπικής συνδεσιμότητας να παραμείνουν σταθερά σε όλα τα επόμενα layer. Με αυτό τον τρόπο συνδέονται οι γειτονικοί νευρώνες του επόμενου layer με ακριβώς τα ίδια βάρη του προηγούμενου layer. Δυο ενδιαφέροντα πλεονεκτήματα αυτών των βελτιστοποιήσεων είναι, πρώτον ότι μειώνεται δραστικά ο αριθμός συνδεσιμότητας των νευρώνων και δεύτερον η πιο ενδιαφέρουσα ιδέα είναι, πως κρατώντας σταθερά τα βάρη στις τοπικές συνδεσιμότητες, αυτό παρομοιάζεται με ένα κυλιόμενο παράθυρο των νευρώνων το οποίο περνάει από όλη την εικόνα και «χαρτογραφεί» την αντίστοιχη περιοχή στο αντίστοιχο μέρος. Αυτό παρέχει μια ευκαιρία για την εύρεση ακμών και των χαρακτηριστικών στην εικόνα ασχέτως με το που βρίσκονται. Για αυτούς τους λόγους ονομάζεται και συνέλιξη. Οι πίνακες που χρησιμοποιούνται ονομάζονται αλλιώς και φίλτρα, καθώς λειτουργούν όπως τα κλασσικά φίλτρα στην επεξεργασία εικόνας.

Για να γίνει αυτή η μέθοδος ακόμα πιο ωφέλιμη, είναι εύλογο να προστεθούν περισσότερα layers μετά το layer εισόδου. Κάθε layer έχει το δικό του φίλτρο και εξάγει διαφορετικά χαρακτηριστικά από το κάθε δεδομένο εισόδου. Η **Εικόνα 2.10** δείχνει νευρώνες οι οποίοι να μην χρησιμοποιούν διαφορετικά φίλτρα «κοιτούν» στο ίδιο παράθυρο της δοσμένης εικόνας [8].



Εικόνα 2.10 Layers τα οποία «κοιτούν» στο ίδιο παράθυρο της εικόνας [8]

Η συνέλιξη για ένα εικονοστοιχείο για το επόμενο layer υπολογίζεται με βάση την ακόλουθη εξίσωση:

$$net(i, j) = (x * w)[i, j] = \sum_m \sum_n x[m, n]w[i - m, j - n] \quad 2.6.1$$

Όπου $net(i, j)$ είναι η έξοδος στο επόμενο layer, x είναι η εικόνα εισόδου, w το φίλτρο και $*$ είναι ο μαθηματικός τελεστής της συνέλιξης.

Η συνέλιξη εκφράζεται ως πολλαπλασιασμός πινάκων. Για παράδειγμα, ας υποθέσουμε πως υπάρχει μια είσοδος 4x4 και ένα φίλτρο 3x3.

$$K = \begin{bmatrix} W_{0,0} & W_{0,1} & W_{0,2} \\ W_{1,0} & W_{1,1} & W_{1,2} \\ W_{2,0} & W_{2,1} & W_{2,2} \end{bmatrix} \quad I = \begin{bmatrix} X_{0,0} & X_{0,1} & X_{0,2} & X_{0,3} \\ X_{1,0} & X_{1,1} & X_{1,2} & X_{1,3} \\ X_{2,0} & X_{2,1} & X_{2,2} & X_{2,3} \\ X_{3,0} & X_{3,1} & X_{3,2} & X_{3,3} \end{bmatrix}$$

Η συνέλιξη ανάμεσα στο του φίλτρο K και τον πίνακα εισόδου I , με μοναδιαία βήματα και χωρίς προθήκη μηδενικών (θα περιγραφούν στην συνέχεια τα στάδια αυτά) δύναται να πραγματοποιηθεί ως πολλαπλασιασμός. Ο πίνακας I θα αναδιαταχθεί σε ένα διάνυσμα δεκαέξι γραμμών και θα δημιουργηθεί ο πίνακας συνέλιξης C , από το φίλτρο K .

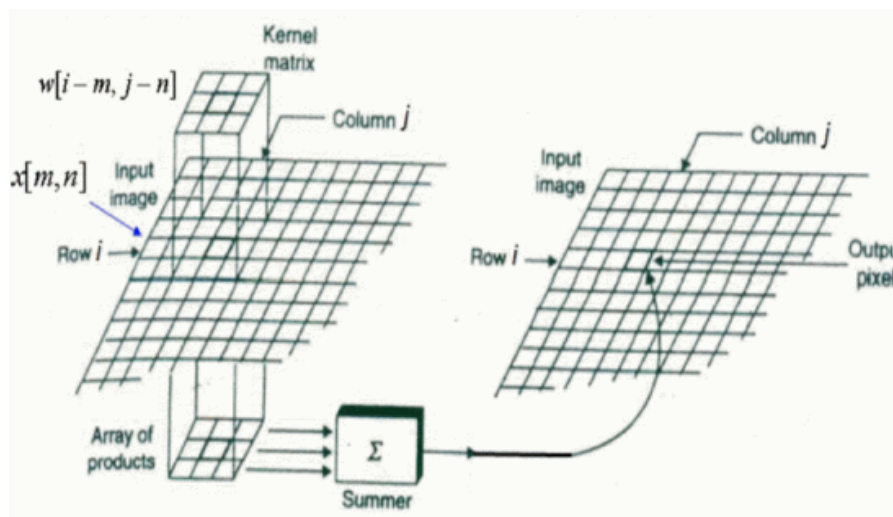
$$I = \begin{bmatrix} x_{0,0} \\ x_{0,1} \\ x_{0,2} \\ x_{0,3} \\ \vdots \\ x_{3,0} \\ x_{3,1} \\ x_{3,2} \\ x_{3,3} \end{bmatrix}$$

$$C = \begin{bmatrix} w_{0,0} & w_{0,1} & w_{0,2} & 0 & w_{1,0} & w_{1,1} & w_{1,2} & 0 & w_{2,0} & w_{2,1} & w_{2,2} & 0 & 0 & 0 & 0 & 0 \\ 0 & w_{0,0} & w_{0,1} & w_{0,2} & 0 & w_{1,0} & w_{1,1} & w_{1,2} & 0 & w_{2,0} & w_{2,1} & w_{2,2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & w_{0,0} & w_{0,1} & w_{0,2} & 0 & w_{1,0} & w_{1,1} & w_{1,2} & 0 & w_{2,0} & w_{2,1} & w_{2,2} & 0 \\ 0 & 0 & 0 & 0 & 0 & w_{0,0} & w_{0,1} & w_{0,2} & 0 & w_{1,0} & w_{1,1} & w_{1,2} & 0 & w_{2,0} & w_{2,1} & w_{2,2} \end{bmatrix} \quad 2.6.2$$

Ο πολλαπλασιασμός $C \times I$, καταλήγει σε ένα διάνυσμα 4 τιμών, που μπορεί έχει τη δυνατότητα αναδιάταξης σε πίνακα 2×2 , τον πίνακα τιμών εξόδου, δηλαδή, από τη διαδικασία της συνέλιξης [11]:

$$G = C \times I = \begin{bmatrix} y_{0,0} \\ y_{1,0} \\ y_{2,0} \\ y_{3,0} \end{bmatrix} = \begin{bmatrix} y_{0,0} & y_{1,0} \\ y_{2,0} & y_{3,0} \end{bmatrix} \quad 2.6.3$$

Η **Εικόνα 2.11** δείχνει πως λειτουργεί η συνέλιξη.



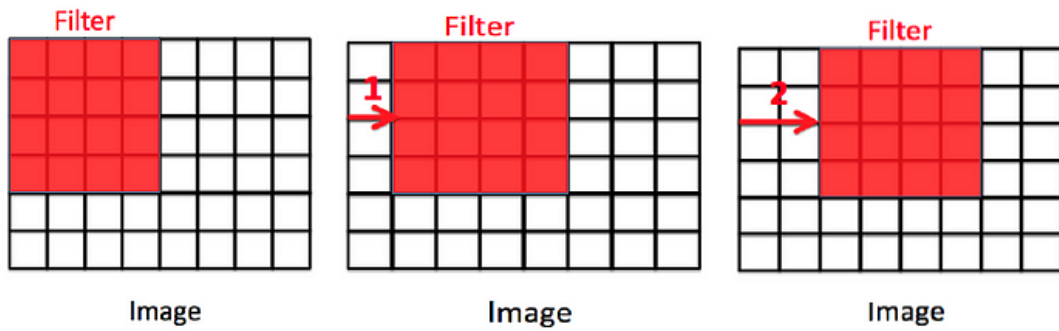
Εικόνα 2.11 Συνέλιξη, ως ένας εναλλακτικός τρόπος της πλήρους συνδεσιμότητας [8]

Βήμα (Stride)

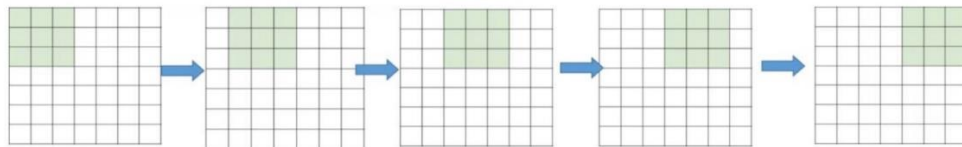
Τα CNN παρέχουν πολλές εναλλακτικές οι οποίες παρέχουν πολλές ευκαιρίες για να μειωθούν ακόμα περισσότερο οι παράμετροι του δικτύου. Μια από αυτές τις ιδιότητες είναι το stride. Με την προαναφερθέν προσέγγιση υπάρχουν πολλές επικαλύψεις μεταξύ των νευρώνων που «κοιτούν» σε μια περιοχή. Έτσι μπορεί αυτή η επικάλυψη να ελεγχθεί με την επιλογή του βήματος του παραθύρου. Αναλυτικά είναι το βήμα με το οποίο το παράθυρο του κάθε νευρώνα μετακινείται στην εικόνα ώστε να μην υπάρχει επικάλυψη.

$$O = 1 + \frac{N - F}{S} \quad 2.6.4$$

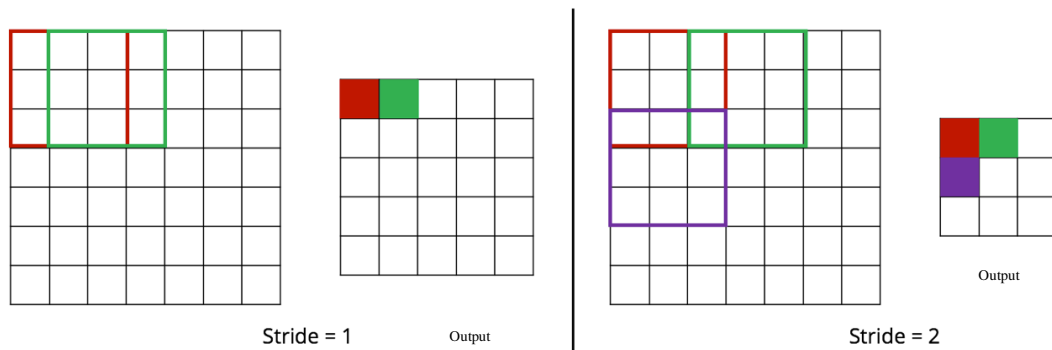
Όπου N είναι το μέγεθος της εικόνας, F το μέγεθος του φίλτρου και S το μέγεθος του βήματος stride και O το τελικό μέγεθος της εικόνας εξόδου [8].



Εικόνα 2.12 Αριστερή εικόνα: stride=0, μεσαία εικόνα stride=1, δεξιά εικόνα stride=2 [12]



Εικόνα 2.13 Αριστερή εικόνα: stride=0, μεσαία εικόνα stride=1, δεξιά εικόνα stride=2 [12]



Εικόνα 2.14 Βήμα [13]

Γέμισμα (Padding)

Ένα μειονέκτημα της συνέλιξης είναι πως κάθε φορά που πραγματοποιείται η πράξη αυτή οι διαστάσεις της εικόνας μειώνονται (Down-Sampling). Έτσι χάνεται ένας όγκος σημαντικής πληροφορίας της εικόνας, η οποία μπορεί να βρίσκεται στο περίγραμμα. Ένας απλός αλλά αποτελεσματικός τρόπος επίλυσης είναι να χρησιμοποιηθεί το λεγόμενο γέμισμα (Padding) στο περίγραμμα της εικόνας. Το σύνηθες είναι το μηδενικό γέμισμα (Zero Padding). Με αυτόν τον τρόπο διαχειρίζονται και οι διαστάσεις της εικόνας. Με το padding αποτρέπεται να συρρικνωθεί το τελικό αποτέλεσμα, όσο μεγαλώνει το βάθος του νευρωνικού.

$$O = 1 + \frac{N + 2P - F}{S} \quad 2.6.5$$

Όπου N είναι το μέγεθος της εικόνας, F το μέγεθος του φίλτρου, S το μέγεθος του βήματος, P ο αριθμός των στρωμάτων του μηδενικού padding (για παράδειγμα, στην Εικόνα 2.15 το $P=1$) [8].

0	0	0	0	0	0	0	0	0
0								0
0								0
0								0
0								0
0								0
0								0
0								0
0								0
0	0	0	0	0	0	0	0	0

Εικόνα 2.15 Μηδενικό Γέμισμα [8]

Ανεστραμμένη Συνέλιξη (Transposed Convolution Layer)

Όπως προαναφέρθηκε η συνέλιξη μειώνει τις διαστάσεις της κάθε εικόνας εισόδου. Το αποτέλεσμα αυτού είναι πως ένας μεγάλος όγκος σημαντικής πληροφορίας χάνεται. Εν αντιθέσει, η ανεστραμμένη συνέλιξη αυξάνει τις διαστάσεις της εικόνας. Αυτό πραγματοποιείται για να αυξηθεί η ανάλυση της και τα σημαντικά χαρακτηριστικά των εικόνων να φαίνονται με περισσότερη ευκρίνεια στα επόμενα layer. Συγκεκριμένα, η ανεστραμμένη συνέλιξη αντιστρέφει τη διαδικασία της κανονικής συνέλιξης. Στη δεύτερη, το φίλτρο ολισθαίνει πάνω στον πίνακα εισόδου, πολλαπλασιάζεται με το αντίστοιχο κομμάτι του πίνακα και τα αποτελέσματα προστίθενται για να δημιουργηθεί ο πίνακας εξόδου. Αυτό επαναλαμβάνεται για όλες τις θέσεις του φίλτρου. Αντίθετα, στην ανεστραμμένη συνέλιξη, υπάρχει ο πίνακας εξόδου ο οποίος ολισθαίνει πάνω στο φίλτρο και πραγματοποιείται η πράξη της συνέλιξης. Για κάθε θέση του φίλτρου το αντίστοιχο τμήμα του πίνακα εξόδου πολλαπλασιάζεται με τις αντίστοιχες τιμές του φίλτρου. Έτσι δημιουργείται η ανακατασκευασμένη έξοδος.

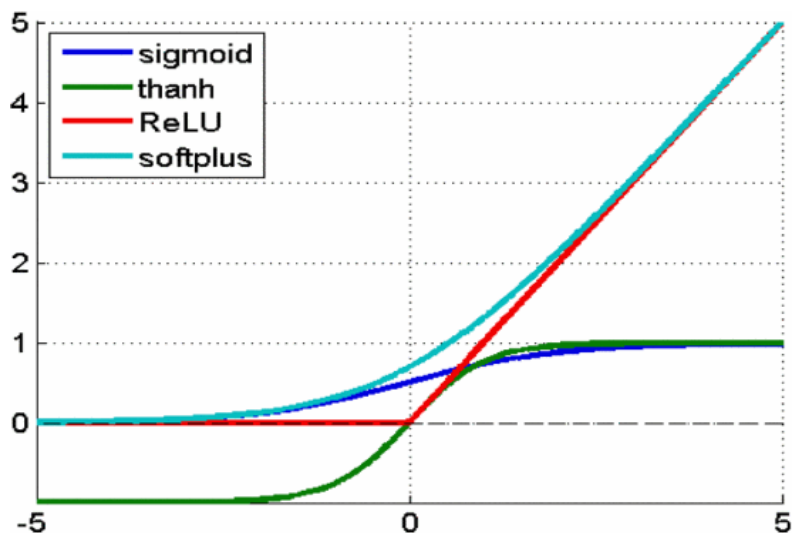
Συνάρτηση Ενεργοποίησης ReLU (ReLU Layer)

Το επόμενο στάδιο μετά την συνέλιξη είναι το στάδιο της ενεργοποίησης. Οι μη γραμμικές συναρτήσεις μπορεί να χρησιμοποιηθούν για να γίνει προσαρμογή ή να αποκοπεί το παραγόμενο αποτέλεσμα. Αυτές οι συναρτήσεις εφαρμόζονται για να γίνει κορεσμός του αποτελέσματος ή να περιοριστεί το ολικό παραγόμενο αποτέλεσμα. Για πολλά χρόνια η σιγμοειδής (sigmoid) και η υπερβολική εφαπτόμενη (tanh) ήταν οι πιο συνηθισμένες μη γραμμικές συναρτήσεις ενεργοποίησης. Παρόλα αυτά, η συνάρτηση Rectified Linear Unit (ReLU) χρησιμοποιείται ευρέως τα τελευταία χρόνια. Έχει ένα πολύ πιο απλό ορισμό όσο αναφορά τη συνάρτηση και την πρώτη παράγωγο της.

$$RelU(x) = \max(0, x) \quad 2.6.6$$

$$\frac{d}{dx}(x) = \{1 \text{ if } x > 0; \text{ otherwise } 0\} \quad 2.6.7$$

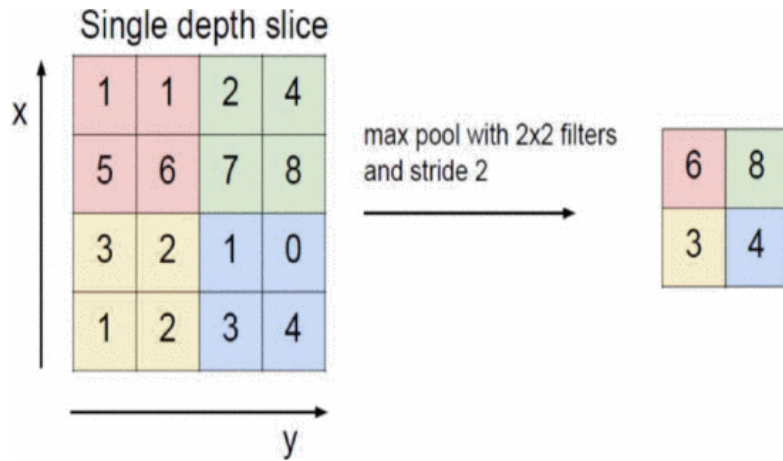
Ακόμα, η κορεσμένη σιγμοειδής και η υπερβολική εφαπτόμενη συνάρτηση δημιουργούν προβλήματα στην οπισθοδρόμηση (Back Propagation). Όσο το νευρωνικό δίκτυο μεγαλώνει σε βάθος, η παραγωγός τους τείνει να εξαφανίζεται. Αυτό συμβαίνει γιατί η παραγωγός αυτών τείνει παντού -εκτός του κέντρου- στο μηδέν 0. Σε αντίθεση, η συνάρτηση ReLU έχει μια συνεχή παραγωγό για τα θετικά δεδομένα. Παρόλο που η παράγωγος της δεν είναι διαχωρίσιμη, μπορεί όμως να παραλειφθεί στην εκτέλεση. Τέλος η ReLU είναι μια συνεχής συνάρτηση που τείνει στο μηδέν. Το μηδέν στην παράγωγο οδηγεί στην διατήρηση του συνολικού μηδέν. Εν αντιθέσει, η σιγμοειδή και η εφαπτομένη πάντα έχουν μη μηδενικά αποτελέσματα στην παράγωγο τα οποία δεν ευνοούν την εκπαίδευση. Η Εικόνα 2.16 δείχνει τις πιο διαδεδομένες συναρτήσεις ενεργοποίησης [8].



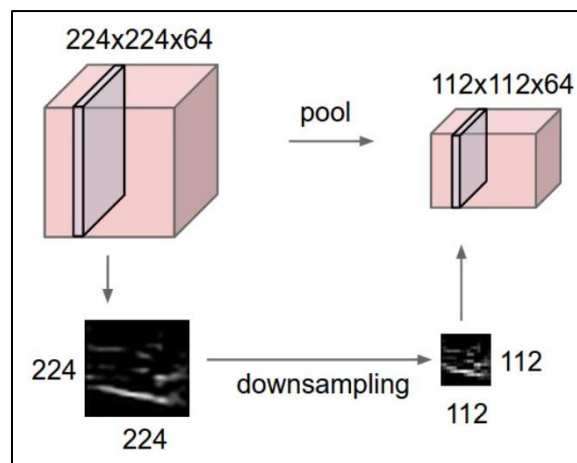
Εικόνα 2.16 Οι πιο διαδεδομένες συναρτήσεις ενεργοποίησης [8]

Επίπεδο Συγκέντρωσης (Pooling Layer)

Ο κύριος σκοπός του επιπέδου συγκέντρωσης (Pooling Layer) είναι η μείωση των διαστάσεων της εικόνας (Down-Sampling), ώστε να μειωθεί η περιπλοκότητα των επόμενων layers. Μπορεί να θεωρηθεί παρόμοιο με την μείωση της ανάλυσης στην επεξεργασία εικόνας. Η συγκέντρωση δεν επηρεάζει τον αριθμό των φίλτρων. Η μέγιστη συγκέντρωση (Max-Pooling) είναι ένας από τους πιο συνηθισμένους τύπους συγκέντρωσης. Διαχωρίζει την εικόνα σε ορθογώνια υπο-περιοχών και επιστρέφει μόνο την μέγιστη τιμή που υπάρχει σε αυτήν την υπό-περιοχή. Το πιο διαδεδομένο μέγεθος για την μέγιστη συγκέντρωση είναι το 2x2. Όπως φαίνεται στην **Εικόνα 2.17**, όταν εκτελείτε η μέγιστη συγκέντρωση στα πάνω αριστερά τετράγωνα (ροζ περιοχή), λαμβάνεται η μέγιστη τιμή και στην συνέχεια μετακινείται δυο τετράγωνα. Επικεντρώνεται λοιπόν μετά στο επάνω δεξί μέρος (πράσινη περιοχή) και υπολογίζεται η μέγιστη τιμή. Αυτό σημαίνει ότι χρησιμοποιείται βήμα 2. Για να αποφευχθεί το down sampling, μπορεί να χρησιμοποιηθεί βήμα ίσο με 1, κάτι το οποίο βεβαία δεν είναι συνηθισμένο. Θα πρέπει να ληφθεί υπόψη ότι η χωρική μείωση των διαστάσεων της εικόνας, δεν διατηρεί την θέση των πληροφοριών. Τέλος, η συγκέντρωση μπορεί να χρησιμοποιηθεί με άνισα φίλτρα και βήματα. Για παράδειγμα, ένα 3x3 max-pooling με βήμα 2 διατηρεί ορισμένες επικαλύψεις μεταξύ των περιοχών [8].



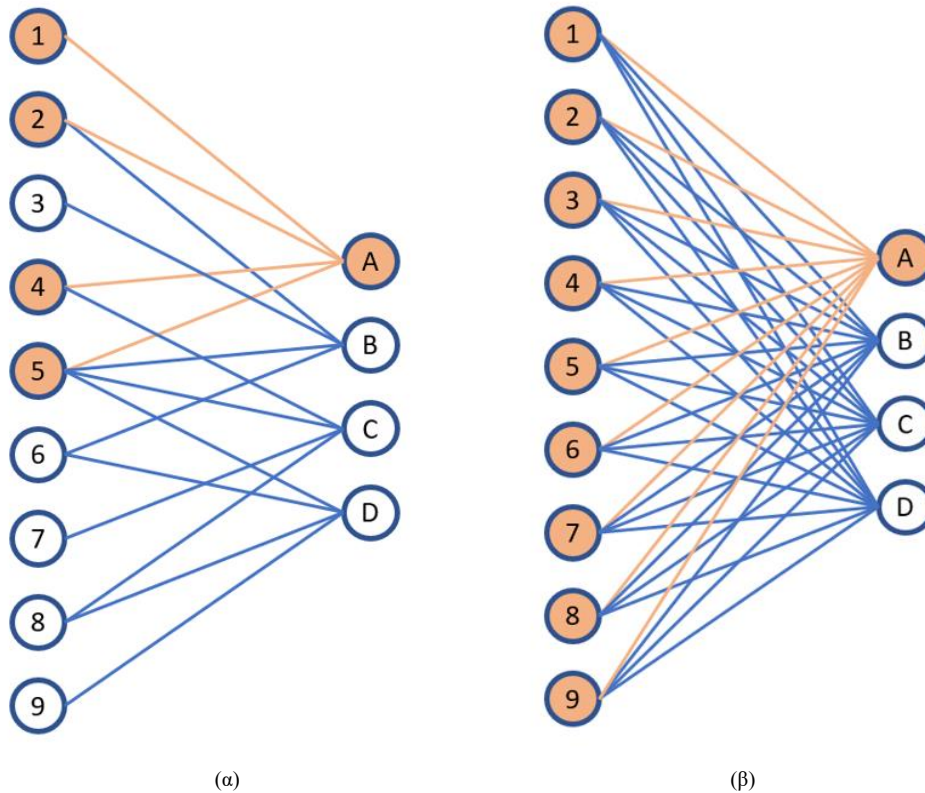
Εικόνα 2.17 Εφαρμογή Μέγιστης Συγκέντρωσης με παράθυρο 2x2 [8]



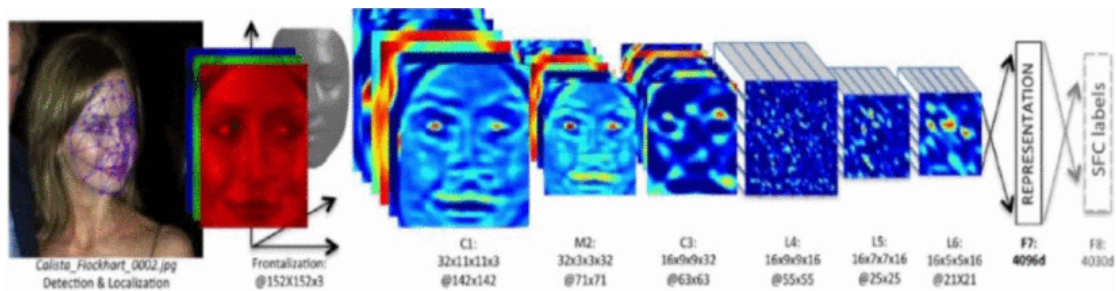
Εικόνα 2.18 Χωρική μείωση των διαστάσεων μιας εικόνας με την μέθοδο συγκέντρωσης [14]

Πλήρως Συνδεδεμένο Επίπεδο (Fully Connected Layer)

Το πλήρως συνδεδεμένο επίπεδο (Fully-Connected Layer) είναι το τελευταίο σε ένα νευρωνικό δίκτυο. Η αρχιτεκτονική αυτού είναι παρόμοια με τον τρόπο που οι νευρώνες είναι διατεταγμένοι σε ένα παραδοσιακό νευρωνικό δίκτυο. Διαφοροποιείται στο ότι κάθε κόμβος σε ένα πλήρως συνδεδεμένο επίπεδο συνδέεται απευθείας με κάθε κόμβο τόσο στο προηγούμενο όσο και στο επόμενο επίπεδο, όπως φαίνεται στην **Εικόνα 2.19β**. Από αυτό το σχήμα παρατηρείται ότι κάθε ένας από τους κόμβους, στο μέγιστο επίπεδο συγκέντρωσης συνδέεται ως διάνυσμα στο πρώτο layer του πλήρως συνδεδεμένου επιπέδου. Αυτές είναι και οι πιο απαιτητικές και χρονοβόρες παράμετροι των συναρτήσεων εκπαίδευσης, μέσα στις συνδέσεις. Το κύριο μειονέκτημα ενός πλήρως συνδεδεμένου επιπέδου είναι ότι περιλαμβάνει πολλές παραμέτρους που χρειάζονται πολύπλοκους υπολογισμούς στην εκπαίδευση. Επομένως, πρέπει να εξαιρεθεί ο αριθμός των κόμβων και των συνδέσεων. Οι κόμβοι και οι συνδέσεις αφαιρούνται χρησιμοποιώντας την τεχνική dropout [8].



Εικόνα 2.19 (α) Ένα συνελκτικό νευρωνικό δίκτυο. (β) Πλήρως συνδεδεμένο δίκτυο [15]



Εικόνα 2.20 Οπτικοποίηση βάθους των επιπέδων ενός νευρωνικού δικτύου [8]

U-NET

Το μοντέλο U-Net έχει δομηθεί με βάση τη λογική των Πλήρως Συνδεδεμένων δικτύων. Η βασική ιδέα του δικτύου αυτού δεν έχει υποστεί σημαντικές αλλαγές. Η εικόνα εισόδου ακολουθεί δύο μονοπάτια. Το πρώτο είναι εκείνο της συστολής (Contracting Path) ή κωδικοποιητής, στο οποίο η εικόνα συμπιέζεται σε μικρότερες διαστάσεις. Το δεύτερο είναι εκείνο της διαστολής (Expansive Path) ή αλλιώς αποκωδικοποιητής, στο οποίο η εικόνα αναδομείται στις διαστάσεις που είχε στην αρχή, δηλαδή πριν υποστεί οποιαδήποτε αλλαγή.

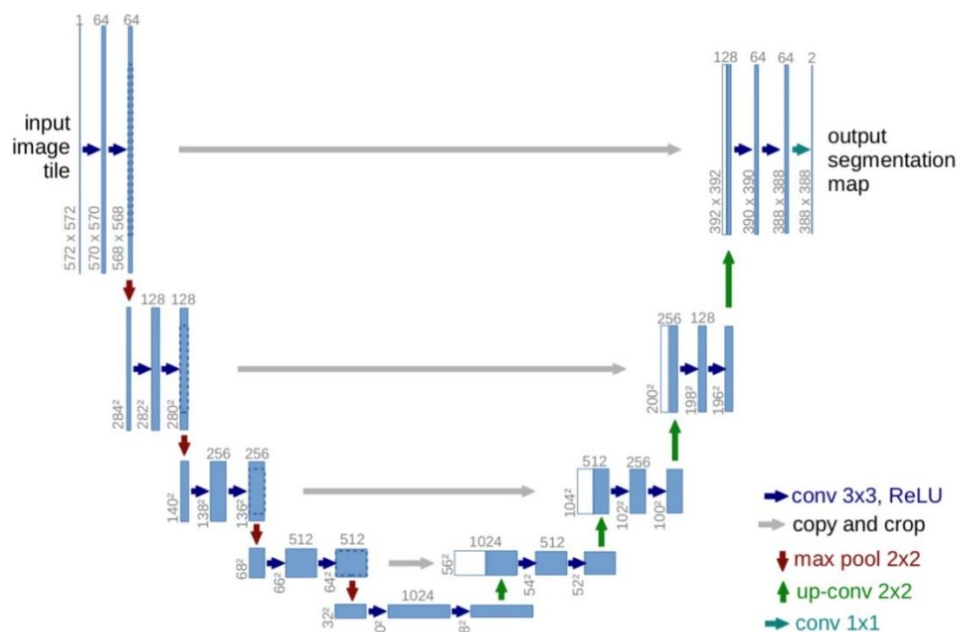
Η διαφορά που εισάγεται στο δίκτυο αυτό εντοπίζεται στις συνδέσεις παράκαμψης (Skip Connections). Οι τελευταίες εφαρμόζονται στα σημεία όπου παρουσιάζουν τις ίδιες διαστάσεις και δίνουν στον τομέα της υπερδειγματοληψίας (αύξηση των διαστάσεων της εικόνας) την δυνατότητα διατήρησης ενός σημαντικού αριθμού πληροφοριών περιβάλλοντος.

Με τον τρόπο αυτό, το δίκτυο δύναται να μεταδώσει τις πληροφορίες περιεχομένου σε επίπεδα υψηλότερης ανάλυσης. Όλα τα παραπάνω, έχουν ως αποτέλεσμα τη δημιουργία μιας συμμετρίας ανάμεσα στα μονοπάτια της διαστολής και της συστολής, η οποία εμφανίζεται σε ένα σχήμα μορφής που παραπέμπει στο γράμμα «U» της αγγλικής γλώσσας, εξ ου κι η ονομασία «U-Net».

Το δίκτυο δεν έχει πλήρως συνδεδεμένα επίπεδα και αξιοποιεί μόνο το έγκυρο κομμάτι της Συνέλιξης. Δηλαδή, η μάσκα κατάτμησης σχετίζεται μόνο με τα εικονοστοιχεία που αντιστοιχούν στην εικόνα εισόδου. Η διαδικασία αυτή, σε συνδυασμό με την πρόβλεψη εικονοστοιχείων στα όρια της εικόνας μέσω τόσο της εξαγωγής πληροφοριών περιεχομένου όσο και του κατοπτρισμού της εικόνας εισόδου, καθιστούν εφικτό τον ατέρμονο διαμοιρασμό μεγάλων εικόνων, ο οποίος όμως γίνεται αυθαίρετα.

Ειδικότερα, η δομή του δικτύου αποτελείται από ένα κομμάτι συστολής στο αριστερό μέρος και ένα μονοπάτι διαστολής στο δεξί. Το πρώτο υπάγεται στην πεπατημένη διάταξη που εμφανίζουν τα συνελκτικά δίκτυα. Επιπρόσθετα, αποτελείται από επαναλαμβανόμενες εφαρμογές διπλών Συνελίξεων 3×3 (χωρίς γέμισμα ακριανών τιμών), τις οποίες ακολουθεί η συνάρτηση ενεργοποίησης ReLU κι η εφαρμογή Μέγιστης Συγκέντρωσης (Max-Pooling) με βήμα 2 για την επίτευξη της μείωσης των διαστάσεων της εικόνας (Down-Sampling).

Σε όλα τα στάδια της μείωσης των διαστάσεων της εικόνας διπλασιάζονται τα συμμετέχοντα κανάλια. Κάθε επίπεδο στον τομέα της διαστολής απαρτίζεται από την υπερδειγματοληψία των χαρτογραφημένων χαρακτηριστικών, η οποία ακολουθείται από μια εφαρμογή Ανεστραμμένης Συνέλιξη 2×2 , δύο Συνελίξεις 3×3 , που τις ακολουθεί μία συνάρτηση ενεργοποίησης ReLU. Στο τελευταίο στάδιο, μία Συνέλιξη 1×1 αξιοποιείται ώστε να αντιστοιχήσει κάθε διάνυσμα στο επιθυμητό αριθμό κλάσεων. Το δίκτυο αποτελείται από 23 συνελκτικά επίπεδα στο σύνολό του. Τα ελαττώματα του δικτύου σχετίζονται με το ότι αφενός χάνεται σημαντικός όγκος χωρικής πληροφορίας εξαιτίας των επιπέδων μέγιστης συγκέντρωσης και αφετέρου στο ότι δεν μπορεί να συσχετίσει το περιεχόμενο με τις συντεταγμένες του [16].



Εικόνα 2.21 Αρχιτεκτονική U-Net [16]

Ταξινόμηση Εικόνων, Ανίχνευση αντικειμένων και Σημασιολογική Κατάτμηση

Η αναγνώριση αντικειμένων είναι μια εντυπωσιακή ικανότητα των ανθρώπων. Με μια και μόνο ματιά σε ένα αντικείμενο, οι άνθρωποι μπορούν να αναγνωρίσουν την ταυτότητα και την ευρύτερη κατηγορία στην οποία ανήκει. Για τον λόγο αυτό, χρειάστηκε να αναπτυχθούν υπολογιστικά συστήματα που να είναι εφοδιασμένα με την ικανότητα αυτή. Η απόδοση των συνελκτικών νευρωνικών δικτύων έχει οδηγήσει στη ραγδαία εξέλιξη της έρευνας που πραγματοποιείται στο συγκεκριμένο πεδίο και στην εφαρμογή τους σε πολλά άλλα θέματα υπολογιστικής όρασης. Παρακάτω αναφέρονται οι βασικές τεχνικές που χρησιμοποιούνται για την ανάλυση και την κατανόηση του περιεχομένου των εικόνων.

Ταξινόμηση Εικόνων (Image Classification)

Η ταξινόμηση εικόνων περιλαμβάνει την αντιστοίχιση μιας εικόνας ή ενός βίντεο σε μια τάξη ή κατηγορία. Αυτό πραγματοποιείται με την εκπαίδευση ενός μοντέλου μηχανικής εκμάθησης με ένα σύνολο δεδομένων εικόνων που έχουν επισημανθεί με τις αντίστοιχες κατηγορίες τους. Μόλις το μοντέλο εκπαιδευτεί, μπορεί να χρησιμοποιηθεί για την πρόβλεψη της κατηγορίας ή της τάξης νέων, άγνωστων, εικόνων.

Ανίχνευση αντικειμένων (Object Detection)

Η ανίχνευση αντικειμένων πραγματοποιεί μια διάκριση μεταξύ διάφορων αντικειμένων σε μια εικόνα ή ένα βίντεο, χρησιμοποιώντας οριοθετημένα πλαίσια για να υποδειχθεί η θέση τους εντός του χώρου των εικονοστοιχείων. Η ανίχνευση αντικειμένων είναι χρήσιμη κατά τον εντοπισμό συγκεκριμένων αντικειμένων σε μια σκηνή, όπως αυτοκίνητα σταθμευμένα στο δρόμο ή άτομα εντός μιας οριοθετημένης γραμμής.

Σημασιολογική Κατάτμηση (Semantic Segmentation)

Η τμηματοποίηση είναι ένας άλλος τύπος επισημάνσης όπου κάθε εικονοστοιχείο σε μια εικόνα επισημαίνεται με βάση κάποιες προκαθορισμένες τάξεις ή κατηγορίες, παρέχοντας λεπτομερή επισημάνση σε pixel. Η κατάτμηση της εικόνας θεωρείται πιο ακριβής από άλλες μεθόδους ανίχνευσης αντικειμένων, επειδή επισημαίνονται μεμονωμένα pixel.

3 Βασικά στάδια προ-επεξεργασίας της εικόνας

3.1 Εισαγωγή

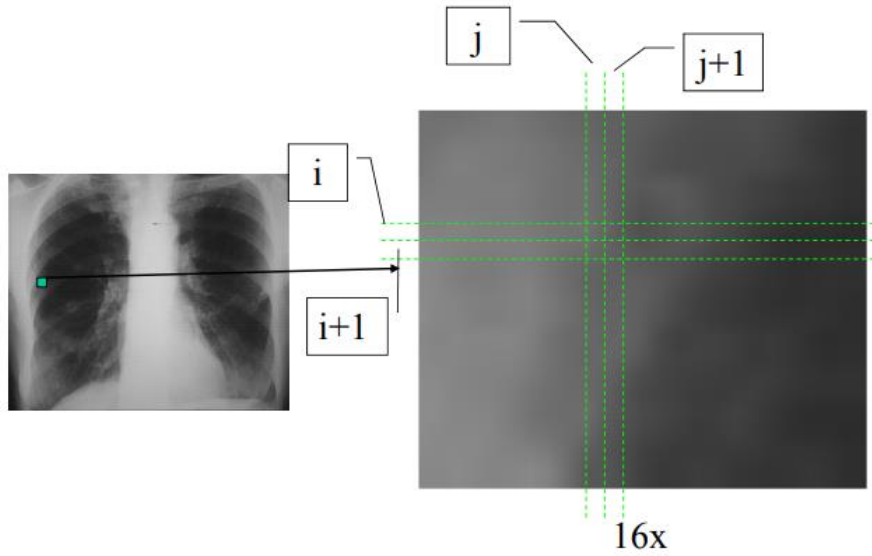
Annotation Tool, ή αλλιώς «εργαλεία επισήμανσης», χρησιμοποιούνται κατά κόρον σε σημασιολογικές κατατμήσεις εικόνων (Image Segmentation), στις οποίες είναι αναγκαίο να υπάρχει η αρχική εικόνα με την αντίστοιχη «αληθή» της κατάτμηση (Ground Truth). Για να επιτευχθεί αυτό χρειάζεται ένα εργαλείο, το οποίο να μπορεί αφενός να «μαρκάρει» τις επιθυμητές περιοχές και να τους αποδίδει την αντίστοιχη ετικέτα (label) «χτίζοντας» με αυτό τον τρόπο την αληθή κατάτμηση της εικόνας. Ειδικότερα, στην περίπτωση της επισήμανσης χαρτών, η δημιουργία του ground truth γίνεται είτε μέσω της επιτόπιας εξαγωγής από ειδικούς του τομέα, είτε μέσω ημι-αυτοματοποιημένων μεθόδων, οι οποίες χρησιμοποιούν λογισμικά που συνήθως παρέχονται από γεωγραφικά συστήματα πληροφοριών. Τα λογισμικά αυτά μπορούν να διανυσματοποιήσουν (vectorize) την εικόνα, αλλά με σημαντικό υπολογιστικό κόστος. Ωστόσο, η μέθοδος αυτή δεν καταφέρνει να διαχωρίσει αποτελεσματικά γραμμικά στοιχεία από πολύγωνα (όπως γράμματα ή αριθμούς) και σημειακά δεδομένα [17]. Επομένως, δημιουργείται η ανάγκη για ένα εργαλείο το οποίο, με εύχρηστο τρόπο, να μπορεί να ορίζει πολλαπλές περιοχές της εικόνας, να γεμίζει το περίγραμμα περιοχών, να τις απεικονίζει πάνω σε μια εικόνα και να τις αποθηκεύει για περαιτέρω χρήση. Τόσο η όραση υπολογιστών όσο και οι αλγόριθμοι τεχνητής νοημοσύνης χρησιμοποιούν τέτοια εργαλεία σε πολλές πρακτικές εφαρμογές. Ενδεικτικά θα μπορούσε να αναφέρει κανείς την ανάλυση ιατρικών εικόνων, την όραση υπολογιστών για αυτόνομα οχήματα, την αναγνώριση και ανίχνευση προσώπων ή ακόμα και την ανάλυση δορυφορικών εικόνων. Στην παρούσα εργασία υλοποιήθηκε ένα τέτοιο εργαλείο επισήμανσης ειδικά για εικόνες από σκαναρισμένους ιστορικούς χάρτες με στόχο την αποτελεσματική επισήμανση τους

3.2 Ψηφιακή εικόνα

Οι εικόνες βρίσκονται παντού γύρω μας. Παρέχονται από την επικοινωνία του ανθρώπου με τον εξωτερικό κόσμο μέσω της πιο βασικής αίσθησης του, της όρασης. Η τελευταία είναι μια πολύ σύνθετη διεργασία μέσω της οποίας τα ανθρώπινα μάτια συλλέγουν πληροφορία. Οι εικόνες μετά την ψηφιοποίηση τους αναπαρίστανται στον υπολογιστή ως ορθογώνιες περιοχές, όπου κάθε σημείο σε αυτές αποτελεί μια σημειακή πληροφορία, γνωστή ως pixel. Κάθε pixel μπορεί να εκφράζει την φωτεινότητα του σημείου σε μια ασπρόμαυρη εικόνα ή την τιμή του χρώματος σε μια πολύχρωμη εικόνα. Συλλέγοντας όλα τα pixel και ταξινομώντας τα χωρικά στο επίπεδο, προκύπτει το τελικό αποτέλεσμα της εικόνας, το οποίο περιλαμβάνει πληροφορίες όπως το χρώμα, η φωτεινότητα και οι περιοχές με σκίαση. Συνεπώς η ψηφιακή εικόνα $I(i,j)$ είναι το αποτέλεσμα της δειγματοληψίας μια δισδιάστατης συνάρτησης $f(x,y)$, όπου η συνάρτηση f αποτελεί έναν μετασχηματισμό που αντιστοιχεί στο ζεύγος ακέραιων συντεταγμένων στο σύνολο των πραγματικών αριθμών.

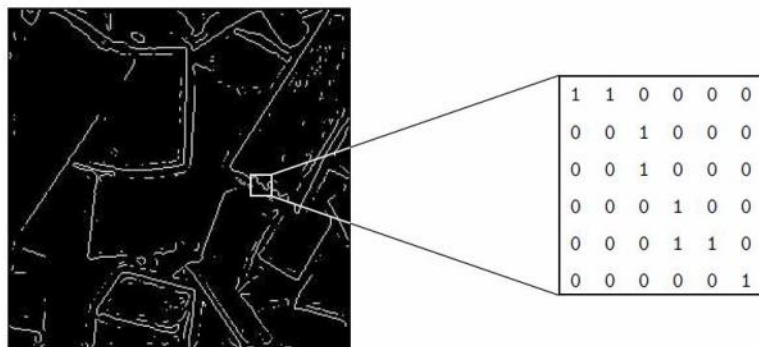
$$f(x, y): N^2 \rightarrow R$$

3.2.1



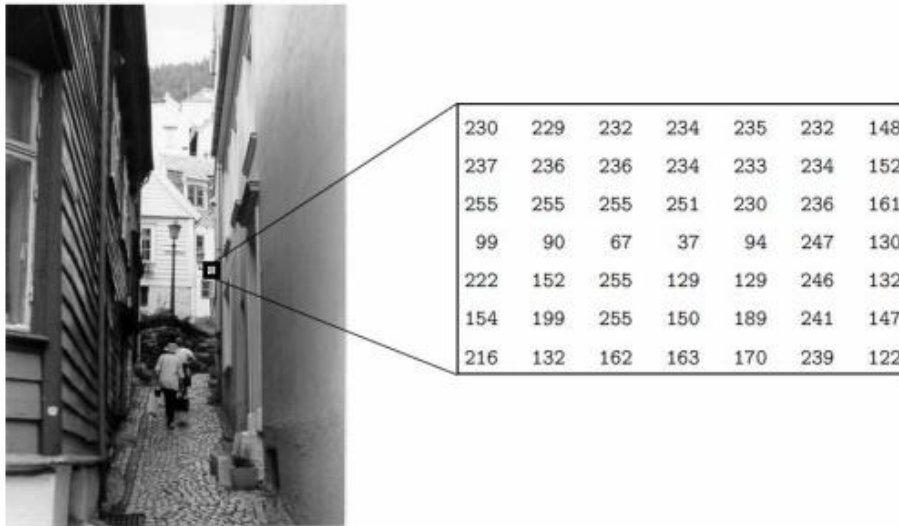
Εικόνα 3.1 Η κάθε θέση (i, j) αποτελεί ένα εικονοστοιχείο της εικόνας [18]

Οι τιμές των pixel σε μια δυαδική μορφή εικόνας μπορούν να πάρουν μόνο μια τιμή, μαύρο (0) ή λευκό (1).



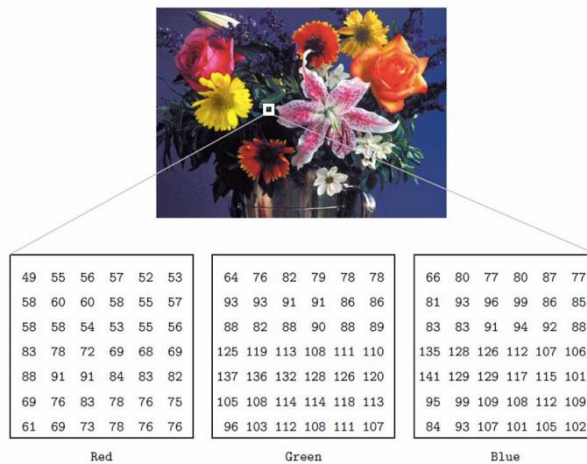
Εικόνα 3.2 Δυαδική μορφή εικόνας [19]

Σε μια grayscale εικόνα, μια εικόνα δηλαδή με διαβαθμίσεις του γκρι, οι τιμές των pixel κυμαίνονται από το 0 έως το 255. Το 0 είναι το μαύρο, το 255 το λευκό και οι ενδιάμεσες τιμές είναι οι διαβαθμίσεις του γκρι.



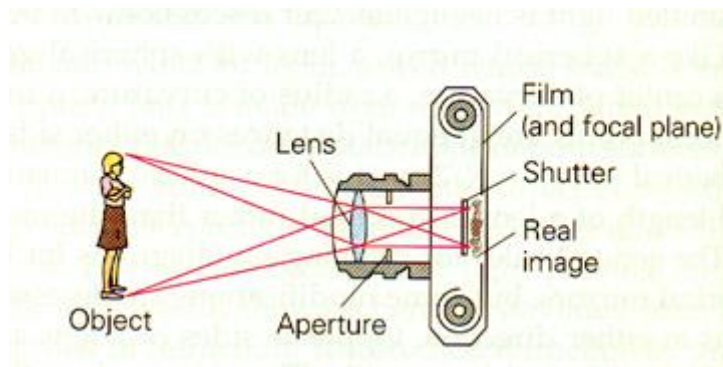
Εικόνα 3.3 Grayscale εικόνα [19]

Οι true colour ή αλλιώς RGB, είναι οι εικόνες στις οποίες το κάθε pixel έχει την τιμή ενός διανύσματος, το οποίο περιγράφει το χρώμα του. Αναλυτικά, το χρώμα αναπαρίσταται με βάση το ποσοστό που περιέχει σε κόκκινο (Red), πράσινο (Green) και μπλε (Blue). Κάθε ένα από τα RGB μπορεί να λάβει μια τιμή από 0 έως 255. Αυτό σημαίνει $255^3=16.777.216$ διαφορετικά χρώματα [19].



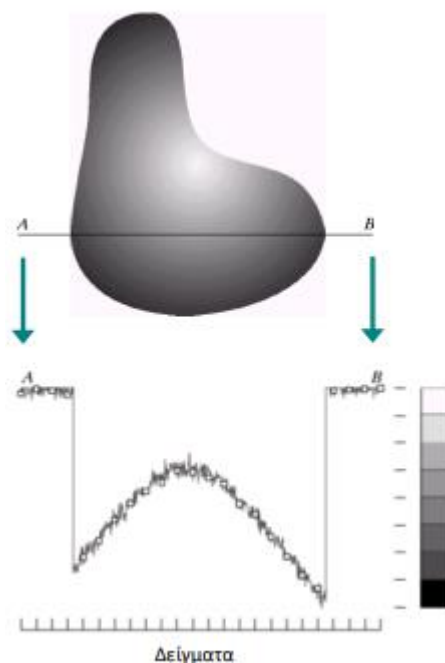
Εικόνα 3.4 RGB εικόνα [19]

Για να γίνει λήψη μιας εικόνας μέσω φωτογραφικής κάμερας με φακό, ο τελευταίος συγκλίνει τις ακτίνες φωτός στο εστιακό σημείο, σε μια συγκεκριμένη απόσταση που ονομάζεται εστιακό μήκος. Το εισερχόμενο φως, ή αλλιώς ενέργεια, προσπίπτει στους οπτικούς αισθητήρες που βρίσκονται σε μια επιφάνεια ευαίσθητη σε αυτόν τον τύπο ενέργειας, η οποία μετατρέπει το φως σε ηλεκτρική τάση. Για να πραγματοποιηθεί η λήψη μιας εικόνας, χρησιμοποιούνται συστοιχίες φωτοαισθητήρων. Οι ψηφιακοί οπτικοί αισθητήρες είναι σε θέση να μετρήσουν έναν πεπερασμένο αριθμό δειγμάτων, σε ένα διακριτό σύνολο επιπέδων ενέργειας (φωτεινότητας).



Εικόνα 3.5 Λήψη Εικόνας [20]

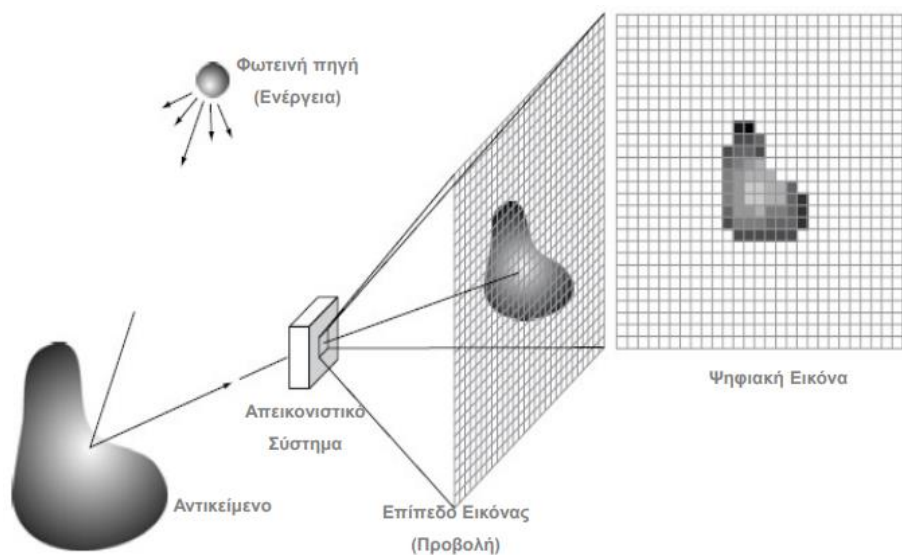
Για να δημιουργηθεί μια ψηφιακή εικόνα, χρειάζεται να μετατραπούν τα συνεχή δεδομένα σε ψηφιακή μορφή. Αναλυτικότερα, μετατρέπονται οι έξοδοι του αισθητήρα, μια κυματομορφή συνεχούς τάσης, το πλάτος και η χωρική συμπεριφορά. Τα προαναφερθέντα δεδομένα είναι πλέον ψηφιακά. Αυτή η εναλλαγή πραγματοποιείται μέσω δύο διαδικασιών, της δειγματοληψίας και του κβαντισμού.



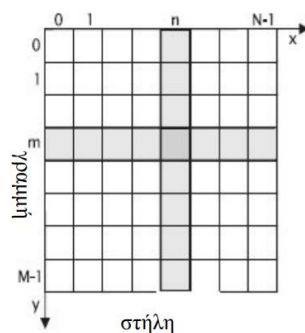
Εικόνα 3.6 Συνεχής Εικόνα

Συγκεκριμένα, η Εικόνα 3.6 απεικονίζει μια δισδιάστατη εικόνα f , η οποία πρόκειται να μετατραπεί σε ψηφιακή μορφή. Η εικόνα μπορεί να είναι συνεχής τόσο σε σχέση με τις συντεταγμένες x και y , όσο και σε σχέση με το πλάτος της. Προκειμένου να πραγματοποιηθεί η μετατροπή σε ψηφιακή μορφή, θα πρέπει να γίνει καταγραφή δειγμάτων της συνάρτησης σε σχέση με τις δύο συντεταγμένες και το πλάτος της. Η διαδικασία της δειγματοληψίας αναφέρεται στη ψηφιοποίηση των τιμών των συντεταγμένων, ενώ ο κβαντισμός αναφέρεται στη ψηφιοποίηση των τιμών του πλάτους. Η μονοδιάστατη συνάρτηση της Εικόνα 3.6 αναπαρίσταται ως ένα γράφημα των τιμών του πλάτους, δηλαδή του επιπέδου έντασης, στο ευθύγραμμο τμήμα AB της Εικόνα 3.6. Οι τυχαίες διακυμάνσεις που παρατηρούνται οφείλονται στο θόρυβο της εικόνας. Για να πραγματοποιηθεί η δειγματοληψία της συνάρτησης της εικόνας, θεωρούνται ισόποσα δείγματα πάνω στην καμπύλη AB , όπως

φαίνονται στην Εικόνα 3.6. Η χωρική θέση κάθε δείγματος επισημαίνεται από μια μικρή κάθετη γραμμή στο κάτω μέρος της εικόνας. Τα δείγματα απεικονίζονται ως μικρά λευκά τετράγωνα που υπερτίθενται επάνω στο γράφημα. Για να σχηματιστεί μια ψηφιακή συνάρτηση, αυτές οι τιμές έντασης πρέπει να κβαντισθούν σε διακριτές ποσότητες.



Εικόνα 3.7 Συλλογή Ψηφιακής Εικόνας [18]



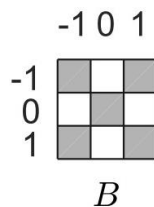
Εικόνα 3.8 Χωρική Αναπαράσταση Εικόνας 2D Pixels [18]

3.3 Μορφολογικές πράξεις

Η μορφολογική επεξεργασία εικόνας χρησιμοποιεί τον όρο «μορφολογία» για να περιγράψει έναν κλάδο της βιολογίας που ασχολείται με τη μορφή και τη δομή τόσο των ζώων όσο και των φυτών. Ωστόσο, στο πεδίο της μαθηματικής μορφολογίας, ο ίδιος όρος χρησιμοποιείται ως εργαλείο για την εξαγωγή συνιστωσών της εικόνας, οι οποίες είναι χρήσιμες για την αναπαράσταση και την περιγραφή της μορφής περιοχών, όπως τα όρια, οι σκελετοί και τα κυρτά περιγράμματα. Αυτές οι μέθοδοι μορφολογικής επεξεργασίας επιτρέπουν την ανίχνευση και την αναγνώριση χαρακτηριστικών της εικόνας που σχετίζονται με το σχήμα και την δομή των αντικειμένων που εμφανίζονται σε αυτήν. Η μορφολογική επεξεργασία

εικόνας αναφέρεται σε μια σειρά μαθηματικών τεχνικών που χρησιμοποιούνται για την ανάλυση και την επεξεργασία της γεωμετρικής δομής και της μορφής των αντικειμένων που εμφανίζονται σε μια εικόνα. Η τεχνική αυτή συνήθως χρησιμοποιείται για να βελτιώσει την ποιότητα της εικόνας, να αφαιρέσει θόρυβο, να ενισχύσει τα χαρακτηριστικά των αντικειμένων και να εξάγει σημαντικές πληροφορίες.

Οι βασικές μορφολογικές επεξεργασίες περιλαμβάνουν τη συστολή (Erosion) και τη διαστολή (Dilation). Αυτές οι δύο διαδικασίες συνδυάζονται με διάφορους τρόπους με διττό στόχο. Ο πρώτος σκοπός σχετίζεται με τη βελτίωση της ποιότητας της εικόνας, ενώ ο δεύτερος με την ανάδειξη συγκεκριμένων χαρακτηριστικών. Η συστολή αφαιρεί μικρές περιοχές ή ακμές από τα αντικείμενα σε μια εικόνα, ενώ η διαστολή προσθέτει περιοχές ή ακμές στα αντικείμενα. Οι δύο αυτές διαδικασίες μπορούν να εφαρμοστούν μόνες τους ή σε συνδυασμό μεταξύ τους για να επιτευχθούν πιο σύνθετες μορφολογικές πράξεις. Μια βασική έννοια της μορφολογικής επεξεργασίας είναι το δομικό στοιχείο, το οποίο αποτελεί γεωμετρικές συσχετίσεις μεταξύ των pixel. Εκείνο μπορεί να έχει τη μορφή είτε μιας εικόνας είτε ενός πίνακα. Τα ενεργά Pixel συμβολίζονται ως μαύρα-γκρι ενώ τα ανενεργά ως λευκά. Είναι μια σύμβαση που εξυπηρετεί την οπτικοποίηση των αποτελεσμάτων, καθώς στον υπολογιστή αυτά αναπαριστούνται αντίθετα.



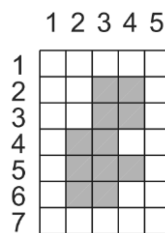
Εικόνα 3.9 Δομικό Στοιχείο B 3x3

Διαστολή (Dilation)

Ορίζεται ως εξής:

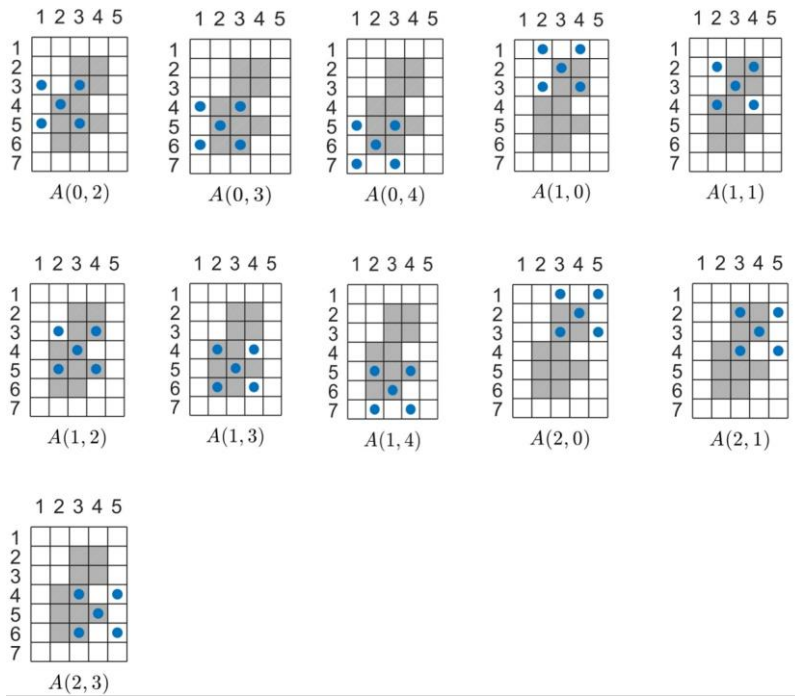
$$A \oplus B = \bigcup_{b \in B} A_b \tag{3.3.1}$$

Όπου A η εικόνα, B το δομικό στοιχείο και b διάνυσμα το οποίο παρέχεται από το δομικό στοιχείο.



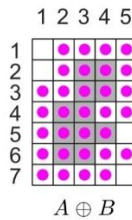
Εικόνα 3.10 Εικόνα A

Μετατοπίζεται το B δομικό στοιχείο πάνω σε κάθε ένα ενεργό pixel της εικόνας A και εφαρμόζεται. Αρχικά το δομικό στοιχείο B μεταθέτεται στην θέση (4,2) και ενεργοποιούνται τα μπλε pixel όπως φαίνονται στην **Εικόνα 3.11**. Κατόπιν μεταφέρεται στη θέση (5,2) και ούτω κάθε εξής.



Εικόνα 3.11 Η κίνηση του δομικού στοιχείου σε τα ενεργά pixel της εικόνας

Η συνένωση όλων αυτών των ενεργών Pixel αποτελούν την διαστολή της εικόνας η οποία οδηγεί σε αύξηση των διαστάσεων της εικόνας



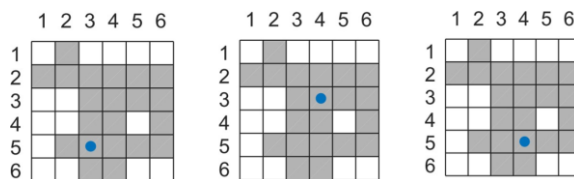
Εικόνα 3.12 Διαστολή Εικόνας A με δομικό στοιχείο B

Συστολή (Erosion)

$$A \ominus B = \bigcap_{b \in B} A_{-b} \quad 3.3.2$$

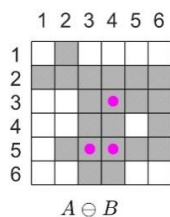
Όπου A η εικόνα, B το δομικό στοιχείο και b διάνυσμα το οποίο παρέχεται από το δομικό στοιχείο.

Η συστολή μεταφράζεται ως το πόσες φορές «χωράει» το δομικό στοιχείο στην αρχική εικόνα. Όλα τα pixel της εικόνας τέτοια ώστε, η μετατόπιση του b σε κάθε ενεργοποιημένο pixel να είναι ένα υποσύνολο της εικόνας A.



Εικόνα 3.13 Σε ποιο θέση της εικόνας χωράει ολόκληρο το δομικό στοιχείο

Η συνένωση όλων αυτών των ενεργών Pixel αποτελούν την συστολή της εικόνας.



Εικόνα 3.14 Συστολή Εικόνας A με δομικό στοιχείο B

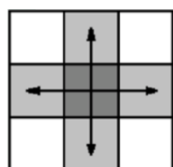
3.4 Τμηματοποίηση της εικόνας

Η τμηματοποίηση μιας εικόνας αναφέρεται στον διαμερισμό της σε κάποιες διακριτές περιοχές ή αντικείμενα με βάση κάποιο χαρακτηριστικό όπως το χρώμα, το σχήμα ή ακόμα και την υφή τους. Υπάρχουν δυο βασικές κατηγορίες ως προς τη μεθοδολογία που μπορεί να ακολουθηθεί. Στην πρώτη ανήκουν οι μέθοδοι που βασίζονται στην ασυνέχεια, δηλαδή στην απότομη αλλαγή της πληροφορίας της εικόνας. Στην δεύτερη ανήκουν οι μέθοδοι που βασίζονται στην ομοιότητα κάποιων περιοχών όπου και βασίζεται και η έννοια του τμήματος κάποιων περιοχών.

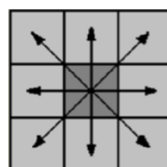
Αλγόριθμος FloodFill

Ένας βασικός αλγόριθμος στο πεδίο των γραφικών υπολογιστών οποίος χρησιμοποιεί τεχνικές τμηματοποίησης, είναι ο αλγόριθμος floodfill, ο οποίος χρησιμοποιείται για να ενώσει συσχετισμένα και γειτονικά στοιχεία σε έναν πίνακα. Ο εν λόγω αλγόριθμος μπορεί να χρησιμοποιηθεί με ποικίλους τρόπους, κάποιιο περισσότερο ή λιγότερο περίπλοκοι, αποτελεσματικοί ή/και αποδοτικοί από άλλους. Εκτός από την επιφάνεια, δίνονται σε εκείνον άλλοι δύο παράμετροι. Ο πρώτος είναι το λεγόμενο «κατώφλι», ενώ ο δεύτερος είναι ο αρχικός κόμβος, όπου η τιμή του πρέπει να ξεπερνά το δοσμένο κατώφλι. Ο αλγόριθμος εντοπίζει όλα τα σημεία που συνδέονται με τον αρχικό κόμβο-pixel μέσω μιας διαδρομής κατά μήκος της οποίας όλες οι τιμές υπερβαίνουν το κατώφλι. Έτσι, δημιουργείται μια ομάδα από συνδεδεμένα pixel, η οποία έχει τη δυνατότητα να τα συσσωρεύει και να τα ταξινομεί σε στοίβα.

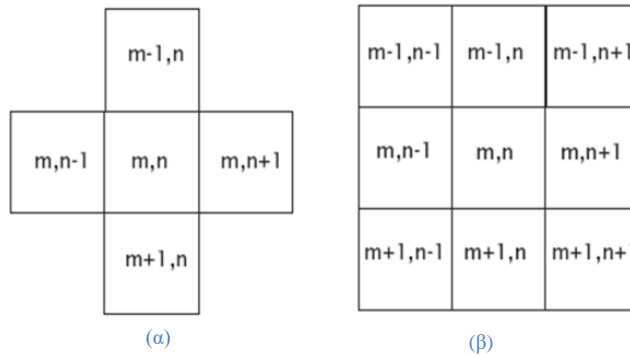
Ο αρχικός κόμβος τροφοδοτεί αυτή τη στοίβα, καθώς είναι ο πρώτος που προστίθεται. Όλα τα pixel που γειτνιάζουν με το αρχικό ελέγχονται. Η «γειτονιά», βάσει της οποίας ο αλγόριθμος ελέγχει, περιλαμβάνει 8 κατευθύνσεις: αυτή πάνω από το pixel, αυτήν κάτω από αυτό, στα αριστερά και δεξιά του καθώς και τις τέσσερις γωνίες του (συνδεσιμότητα 8). Η συνδεσιμότητα 4 δεν περιλαμβάνει τις γωνίες. Κάθε γειτονικό pixel, το οποίο έχει τιμή μεγαλύτερη του κατωφλίου, προστίθεται στην στοίβα και εφαρμόζεται ο αλγόριθμος σε αυτόν σαν αρχικό κόμβο. Η διαδικασία συνεχίζεται μέχρις ότου όλα τα συνδεδεμένα στοιχεία να έχουν προστεθεί στην λίστα [21].



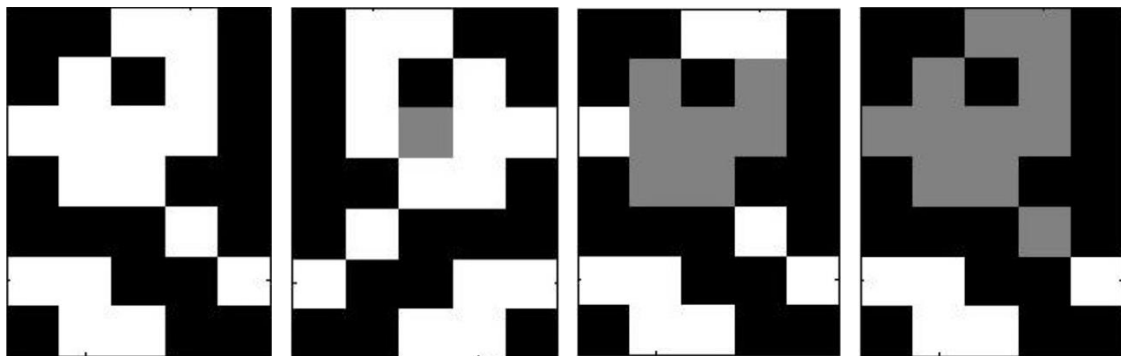
Εικόνα 3.15 Συνδεσιμότητα 4



Εικόνα 3.16 Συνδεσιμότητα 8



Εικόνα 3.17 (α) Πλέγμα 4 Γειτόνων, (β) Πλέγμα 8 Γειτόνων [18]



Εικόνα 3.18 Η διαδικασία του αλγορίθμου floodfill. Η αρχική εικόνα είναι η πρώτη από τα αριστερά. Προχωρώντας δεξιά, επιλέγεται ο αρχικός κόμβος-pixel, τα γειτονικά pixel γεμίζουν και μετέπειτα οι γείτονες των γειτόνων γεμίζουν, έως ότου γεμίσει όλη η εικόνα [21]

3.5 Εξαγωγή υπο-εικόνων

Οι προσεγγίσεις που γίνονται με βάση τα patches πολλές φορές προτιμώνται ώστε να γίνει μια εργασία σε ολόκληρη την εικόνα. Αυτό δικαιολογείται αν αναλογιστεί κανείς πως εκείνα προσφέρουν μια εναλλακτική παρουσίαση της εικόνας ως ένα σετ από υπό-εικόνες (patches). Σε πολλές εφαρμογές επεξεργασίας εικόνας, το στάδιο αυτό αποτελεί ένα ζωτικό κομμάτι της προ-επεξεργασίας. Για τον λόγο αυτό, χρησιμοποιήθηκε στην παρούσα διπλωματική το εργαλείο patchIT το οποίο προσφέρει μια ενιαία δομή, κατάλληλη για συστηματική και αυτοματοποιημένη εξαγωγή των patches (patch extraction) μιας εικόνας [22]. Ο προσδιορισμός των patches που θα εξαχθούν βασίζεται σε γεωμετρικά και χωρικά κριτήρια, τα οποία καθορίζονται από τον χρήστη. Το patchIT μπορεί να εξάγει υπό-εικόνες (patches) τόσο με πλεγματική μορφή (Grid ή Sliding mode) όσο και σε τυχαίες θέσεις πάνω στην εικόνα (Random Mode). Ο χρήστης καθορίζει ορισμένες παραμέτρους όπως το μέγεθος του patch, το βήμα της μετατόπισης καθώς και την κατανομή των patches στην εικόνα. Ο αριθμός των patches που δημιουργούνται είναι ένας συνδυασμός από διάφορες παραμέτρους, οι οποίες επηρεάζουν το φόρτο εργασίας του υπολογιστή τόσο σε επίπεδο χρόνου όσο και μνήμης.

Ειδικότερα, στην Grid λειτουργία δημιουργούνται patches σε πλεγματική μορφή, στην οποία η απόσταση μεταξύ τους μπορεί να καθοριστεί και για τον οριζόντιο και για τον κατακόρυφο άξονα. Στην προσέγγιση αυτή δημιουργούνται patches τα οποία είτε μπορούν να επικαλύπτονται και να είναι πυκνά κατανομημένα είτε να έχουν μια απόσταση μεταξύ τους.

Μια εναλλακτική τεχνική δημιουργίας των patches είναι η λεγόμενη «random» (τυχαία). Σε αυτή τα patches δημιουργούνται τυχαία σε όλη την εικόνα με βάση τα κριτήρια που έχει ορίσει ο χρήστης, όπως είναι ο αριθμός των patches, η εγγύτητα τους και η χωρική τους κατανομή. Επιπρόσθετα, ένα ιδιαίτερα σημαντικό πλεονέκτημα του patchIT είναι πως έχει τη δυνατότητα, μέσω της χρήσης κατάλληλης δεικτοδότησης, να δημιουργήσει patches διαφόρων μορφών. Για παράδειγμα, περιστρεμμένα, ανεστραμμένα, συμμετρικά, ανεστραμμένα κατά τον οριζόντιο και κατακόρυφο άξονα, κλπ.

3.6 Επαύξηση δεδομένων

Τα μοντέλα βαθιάς μάθησης, στην πλειονότητα τους, χρειάζονται ένα σημαντικό όγκο δεδομένων ώστε να φθάσουν στο επιθυμητό αποτέλεσμα. Είναι γνωστό πως για να πραγματοποιηθεί η εκμάθηση ενός νευρωνικού δικτύου χρειάζονται ένα ικανό σετ δεδομένων, τα οποία δυστυχώς δεν είναι πάντα διαθέσιμα ή επαρκή. Επιπλέον, ο ανεπαρκής αριθμός εικόνων συχνά έχει ως αποτέλεσμα την υπέρ εκτίμηση των αποτελεσμάτων (Overfitting). Βέβαια, το πρόβλημα αυτό μπορεί να προκληθεί και από την πολυπλοκότητα του μοντέλου που χρησιμοποιείται. Ένας τρόπος για να εμπλουτισθούν τα δεδομένα εκπαίδευσης είναι η επαύξηση των δεδομένων (data augmentation). Προκειμένου να επιτευχθεί ένα ακριβέστερο αποτέλεσμα πραγματοποιούνται βασικοί γεωμετρικοί μετασχηματισμοί στα patches που έχουν εξαχθεί από την εικόνα. Τέτοιοι μπορεί να είναι η αναστροφή (flip) και η στροφή (rotation), οι οποίοι δεν ενέχουν ιδιαίτερο βαθμό δυσκολίας. Η πρώτη αναφέρεται στην διαδικασία αναστροφής των εικόνων κατά τον οριζόντιο και κατακόρυφο άξονα. Η δεύτερη αποτελεί έναν απλό γεωμετρικό μετασχηματισμό, που ορίζεται από την γωνία που υποδεικνύει ο χρήστης. Έτσι, δημιουργούνται νέα patches και με αυτό τον τρόπο εμπλουτίζεται το σετ δεδομένων και όγκος τους [23].

4 Εντοπισμός κτιρίων σε ιστορικούς χάρτες

4.1 Εισαγωγή

Οι ιστορικοί τοπογραφικοί χάρτες είναι πηγή πολύτιμων πληροφοριών. Παρέχουν πληροφορίες σχετικά με το γεωγραφικό περιβάλλον και τις ανθρώπινες δραστηριότητες ενός συγκεκριμένου χώρου και χρόνου. Μέσα από αυτούς εξάγονται πληροφορίες ανά τα έτη από πολλά επιστημονικά πεδία. Δυο από τις βασικότερες επιστήμες, όπως η κοινωνιολογία και η αρχαιολογία, οι οποίες χρησιμοποιούν αυτούς τους χάρτες, εξάγουν συμπεράσματα σχετικά με την ανάπτυξη των περιοχών, τις κοινωνικοπολιτικές καθώς και τις πολεοδομικές αλλαγές στον χρόνο. Οι ιστορικοί χάρτες αποτυπώνουν τις οικιστικές περιοχές, τα χωριά και τις πόλεις, καθώς και τις υποδομές, όπως δρόμους, γέφυρες, σιδηροδρομικές γραμμές, λιμάνια και αεροδρόμια. Αυτές οι πληροφορίες είναι χρήσιμες για τη μελέτη της ανάπτυξης των πόλεων, την ανάλυση των υποδομών και την κατανόηση των μεταβολών στον αστικό χώρο. Οι πληροφορίες από τους τοπογραφικούς ιστορικούς χάρτες συμβάλλουν στην ανακάλυψη, την έρευνα και την κατανόηση της ανθρώπινης ιστορίας και της γεωγραφίας μιας περιοχής.



(a) 1961



(b) 1971



(c) 1981



(d) 1991

Εικόνα 4.1 Τοπογραφικοί χάρτες περιοχής Πετρούπολης τεσσάρων δεκαετιών οι οποίοι διατέθηκαν από την ΕΛΣΤΑΤ



Εικόνα 4.2 Τοπογραφικός χάρτης περιοχής Πετρούπολης 1971

Οι πηγές των ιστορικών χαρτών της Ελλάδας μπορούν να ποικίλουν ανάλογα τον σκοπό για τον οποίο δημιουργήθηκαν. Ορισμένες από τις κύριες πηγές αυτών περιλαμβάνουν μουσεία, οργανισμούς, Εθνικές και περιφερειακές βιβλιοθήκες. Συχνά διαθέτουν συλλογές ιστορικών χαρτών που καλύπτουν διάφορες περιόδους και περιοχές της Ελλάδας.

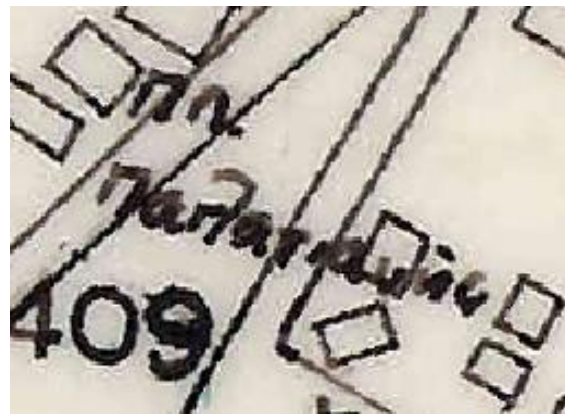
Ένα από τα προβλήματα που προκύπτουν στους αναλογικούς χάρτες είναι η περιορισμένη ακρίβεια που παρέχουν και η περιορισμένη αναπαράσταση γεωγραφικών πληροφοριών. Λόγω της χαμηλής γραφικής ποιότητας που χαρακτηρίζει τους παλιούς χάρτες, η εξαγωγή πληροφορίας από αυτούς είναι μια δύσκολη διαδικασία. Μια σημαντική εργασία είναι η εξαγωγή συγκεκριμένου περιεχομένου, όπως ένας είδος πληροφορίας, που αναφέρεται σε μια παρελθούσα χρονική εποχή για την ανάλυση και την μέτρηση του υποβάθρου του παρελθόντος. Μια τέτοια πληροφορία είναι εξαιρετικά χρήσιμη καθώς παρέχει πληροφορία για το αστικό περιβάλλον του παρελθόντος. Για να πραγματοποιηθεί η εξαγωγή πληροφορίας γρήγορα και με ακρίβεια χρειάζεται ένας τρόπος που βασίζεται τόσο σε συγκεκριμένη μεθοδολογία όσο και συγκεκριμένα χαρακτηριστικά. Η συγκεκριμένη πληροφορία μπορεί να εξαχθεί για πάρα πολλούς σκοπούς. Παρά ταύτα η διαδικασία αυτή μπορεί αποδειχθεί τόσο χρονοβόρα όσο και απαιτητική διότι μόνο ένας ειδικός μπορεί να την πραγματοποιήσει [17].

4.2 Προβλήματα

Σε όλη την έκταση του τοπογραφικού χάρτη, υπάρχουν περιπτώσεις στις οποίες είτε το κείμενο είτε κάποιου άλλου είδους σχέδιο επικαλύπτει τα πολύγωνα. Τα τελευταία αναπαριστούν οικοδομικά τετράγωνα ή σπίτια, όπως φαίνεται στην Εικόνα 4.3.



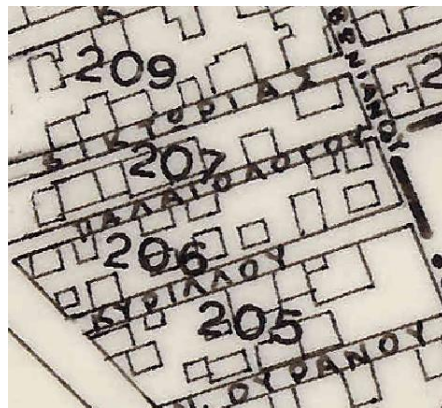
(α)



(β)

Εικόνα 4.3 (α) Νούμερα που επικαλύπτουν τα περιγράμματα (β) Το κείμενο επικαλύπτει τα περιγράμματα

Συχνά, η αναγνώριση είτε των σπιτιών είτε των οικοδομικών τετραγώνων γίνεται ακόμη δυσκολότερη εξαιτίας των ονομάτων των δρόμων, οι οποίοι, κατά περιπτώσεις, όταν οι δρόμοι είναι στενοί εφάπτονται στις άκρες, οι οποίες ορίζουν τα όρια του οικοδομικού τετραγώνου. Ένα ενδεικτικό παράδειγμα φαίνεται στην Εικόνα 4.4.



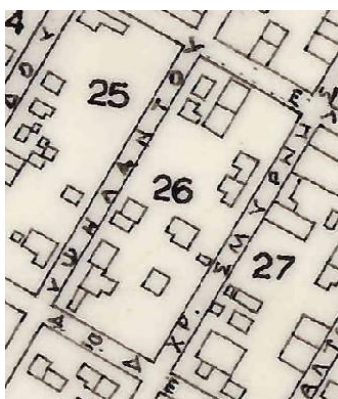
Εικόνα 4.4 Στενοί δρόμοι που εφάπτονται στα οικοδομικά τετράγωνα

Ένα εξίσου σημαντικό πρόβλημα δημιουργείται λόγω των ασαφών ορίων των οικοδομικών τετραγώνων, των οποίων η περιγραφή μέσω πολυγώνου είναι συνήθως προβληματική ή ακόμα και αδύνατη, όπως μπορεί κανείς να διακρίνει στην Εικόνα 4.5.



Εικόνα 4.5 Μη κλειστά πολύγωνα

Επιπρόσθετα, τα κτίρια, που βρίσκονται μέσα στον ιστό των οικοδομικών τετραγώνων, ποικίλλουν σε μέγεθος, περίγραμμα και προσανατολισμό. Χαρακτηριστικό παράδειγμα αποτελεί η Εικόνα 4.6.



Εικόνα 4.6 Κτίρια τα οποία ποικίλλουν σε μέγεθος, περίγραμμα και προσανατολισμό

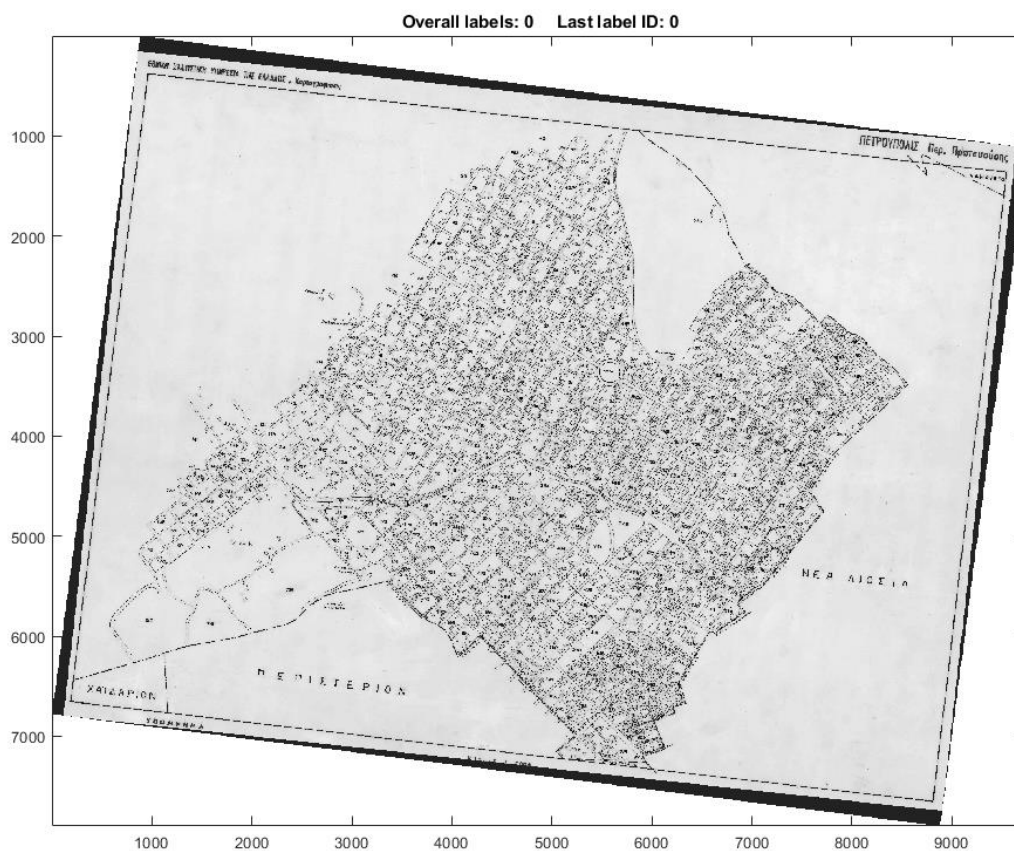
Επιπλέον, υπάρχουν απομονωμένα σπίτια καθώς επίσης και κτίρια τα οποία βρίσκονται το ένα δίπλα στο άλλο [17].

4.3 Εργαλείο επισήμανσης - Annotation tool

Μια σημαντική απαίτηση για την μελέτη των τοπογραφικών χαρτών και την περαιτέρω εξαγωγή πληροφοριών από αυτούς είναι η επισήμανση των κτιρίων στον χάρτη και η απόδοση μιας ξεχωριστής σήμανσης (label) σε κάθε ένα από αυτά. Στο διαδίκτυο μπορεί κανείς να βρει διάφορα εργαλεία [24-34] τα οποία μπορούν να χρησιμοποιηθούν για την δημιουργία επισημάνσεων (ground truth) σε εικόνες. Η πολυπλοκότητα και η ευχρηστία των εργαλείων αυτών ποικίλει και στα περισσότερα για να προσδιορίσει ο χρήστης κάποια περιοχή / πολύγωνο πάνω στην εικόνα θα πρέπει να ορίσει τόσα σημεία όσες και οι κορυφές του εκάστοτε πολυγώνου [24]. Η παρούσα διπλωματική έρχεται να επιλύσει αυτό το πρόβλημα προσφέροντας ένα εργαλείο επισήμανσης (annotation tool) στο οποίο ο προσδιορισμός ενός οικήματος στον εκάστοτε τοπογραφικό χάρτη γίνεται με ένα απλό «κλικ» στο εσωτερικό του περιγράμματος του. Αυτό έχει σαν αποτέλεσμα να ορίζεται άμεσα το περίγραμμα του κτίσματος και να του αποδίδεται αυτόματα μια ξεχωριστή σήμανση (label). Με αυτό τον τρόπο επιτυγχάνεται ο βασικός στόχος του εργαλείου που είναι να διευκολυνθεί η διαδικασία επισήμανσης περιοχών όπως τα κτίρια σε εικόνες που αντιστοιχούν σε ιστορικούς χάρτες. Για την υλοποίηση του εργαλείου χρησιμοποιήθηκε ως εικόνα εισόδου ο τοπογραφικός χάρτης της περιοχής της Πετρούπολης που φαίνεται στην Εικόνα 4.2. Στην παρούσα διπλωματική, ο σχετικός χάρτης διατέθηκε από την Ελληνική Στατιστική Αρχή (ΕΛΣΤΑΤ) και είχε χρησιμοποιηθεί για την απογραφή του πληθυσμού κατά το έτος 1971. Η κλίμακα του είναι 1:5000 και οι διαστάσεις του 50x70 cm. Το εργαλείο υλοποιήθηκε σε προγραμματιστικό περιβάλλον MATLAB. Ένα σημαντικό πλεονέκτημα του περιβάλλοντος αυτού είναι η πληθώρα των πληροφοριών που είναι διαθέσιμες σε διαδικτυακούς τόπους δημόσιας συζήτησης, κάτι που διευκολύνει την αντιμετώπιση προβλημάτων που ενδέχεται να προκύψουν κατά τη διάρκεια της διαδικασίας υλοποίησης.

4.4 Εξαγωγή αποτυπωμάτων των κτιρίων

Αρχικά, το εργαλείο επισήμανσης «διαβάζει» τις διαστάσεις της εικόνας (7897x9757 pixels) και στη συνέχεια δημιουργεί έναν πίνακα Z στον οποίο θα αποθηκευτούν τα αποτελέσματα της επισήμανσης. Ο πίνακας αυτός έχει τις ίδιες διαστάσεις με την αρχική εικόνα και αρχικοποιείται με τιμή 0. Ύστερα, η εικόνα εισόδου μετατρέπεται σε grayscale μορφή, δηλαδή σε εικόνα διαβαθμίσεων του γκρι. Στην συνέχεια, εφαρμόζεται ο αλγόριθμος FloodFill, ο οποίος «γεμίζει» μια περιοχή με εικονοστοιχεία διαφορετικού χρώματος στα πλαίσια μιας ανοχής (κατώφλι). Η τελευταία έχει οριστεί στα 20 graylevels. Ο αριθμός αυτός προέκυψε μέσα από δοκιμές και αποφασίστηκε να εφαρμοστεί μια πιο αυστηρή τιμή, με σκοπό να περιοριστούν ορισμένα προβλήματα. Εφόσον κάνει κλικ ο χρήστης στο κέντρο του κτιρίου, ο αλγόριθμος διατηρεί την τιμή του εικονοστοιχείου και ορίζεται ως ο αρχικός κόμβος.

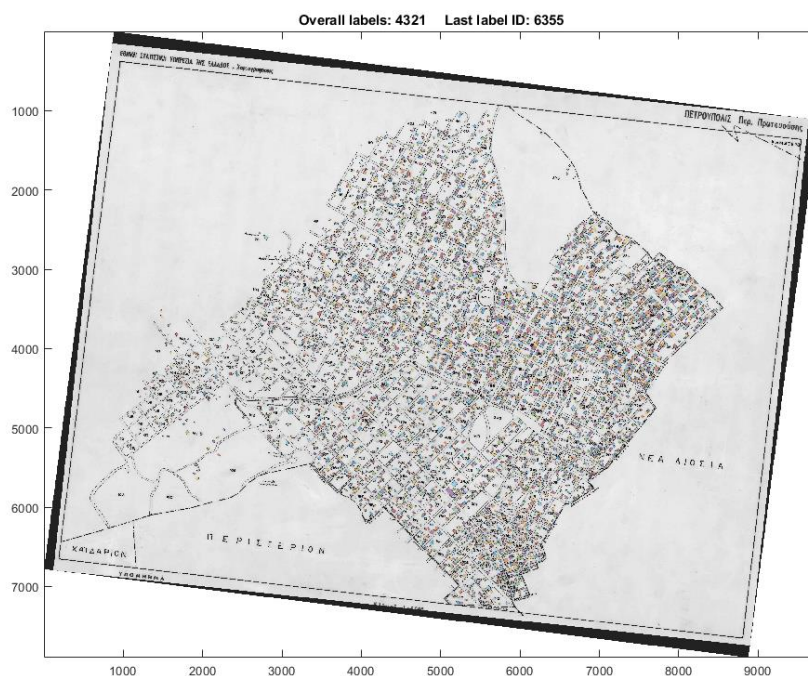


Εικόνα 4.7 Ο τοπογραφικός χάρτης της Πετρούπολης σε κλίμακα του γκρι



Εικόνα 4.8 Grayscale αποσπάσματα του τοπογραφικού χάρτη

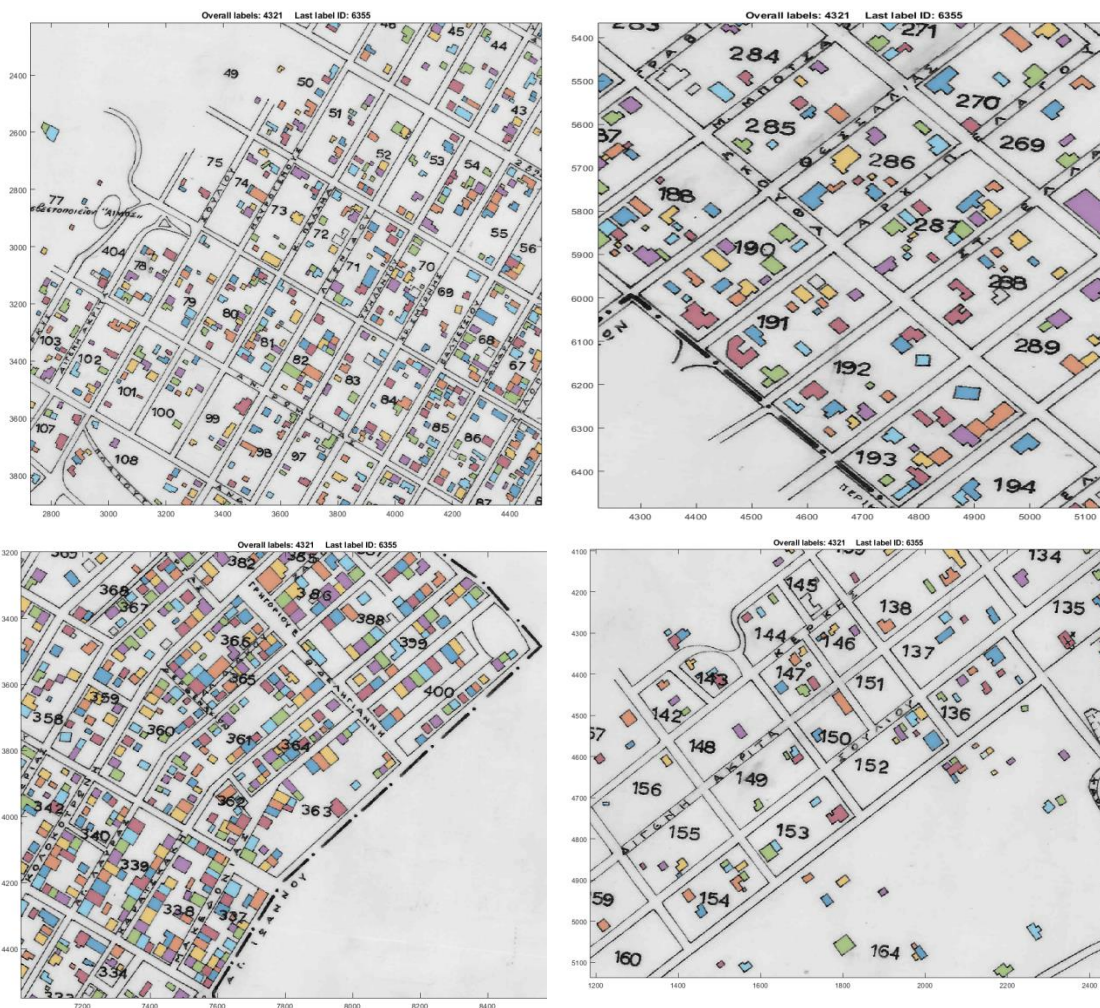
Η συνάρτηση αυτή επιστρέφει μια δυαδική εικόνα που περιέχει τις θέσεις των ρixel όπου οι εντάσεις τους γεινιάζουν με την ένταση του αρχικού κόμβου. Η "γειτονιά" του αλγορίθμου αποτελείται από 8 κατευθύνσεις: πάνω, κάτω, αριστερά, δεξιά και τις τέσσερις γωνίες (συνδεσιμότητα 8). Έτσι, ελέγχονται όλες οι γειτονικές εντάσεις των γειτονικών ρixel. Με βάση την προκαθορισμένη ανοχή των 20 graylevels, υπολογίζεται ένα εύρος εντάσεων που περιλαμβάνεται στη δυαδική εικόνα. Το εύρος αυτό καθορίζεται ως η τιμή του εικονοστοιχείου μείον την ανοχή, έως την τιμή του εικονοστοιχείου συν την ανοχή. Έπειτα, για να χρωματιστεί η περιοχή των παρόμοιων εντάσεων που προέκυψαν από την προηγούμενη διαδικασία, χρησιμοποιείται η συνάρτηση "labeloverlay". Αυτή η συνάρτηση χρωματίζει μόνο τα ρixel της δυαδικής εικόνας που έχουν την τιμή 1 και εμφανίζουν παρόμοια ένταση. Κάθε φορά που πραγματοποιείται ένα κλικ στην εικόνα, γεμίζει το αντίστοιχο κομμάτι του πίνακα Z με το εκάστοτε αριθμημένο label. Αναλυτικότερα, στο πρώτο κτίριο που θα γίνει επισήμανση, η αντίστοιχη θέση του στον πίνακα Z θα λάβει τιμή 1. Όταν επιλεγεί το δεύτερο κτίριο, τότε η αντίστοιχη θέση του στον πίνακα Z θα λάβει την τιμή 2, και ούτω κάθε εξής.



Εικόνα 4.9 Χάρτης με όλα τα περιγράμματα στα οποία έχει γίνει επισήμανση

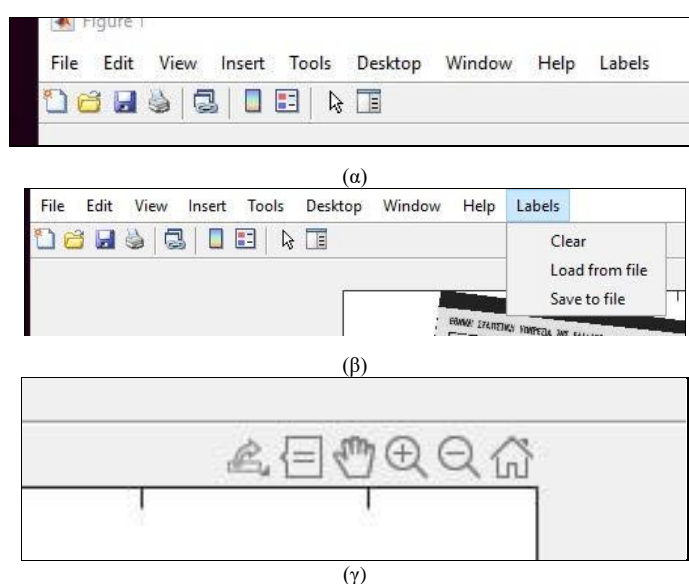


Εικόνα 4.10 Απόσπασμα χάρτη με όλα τα σπίτια που έχουν επισημανθεί



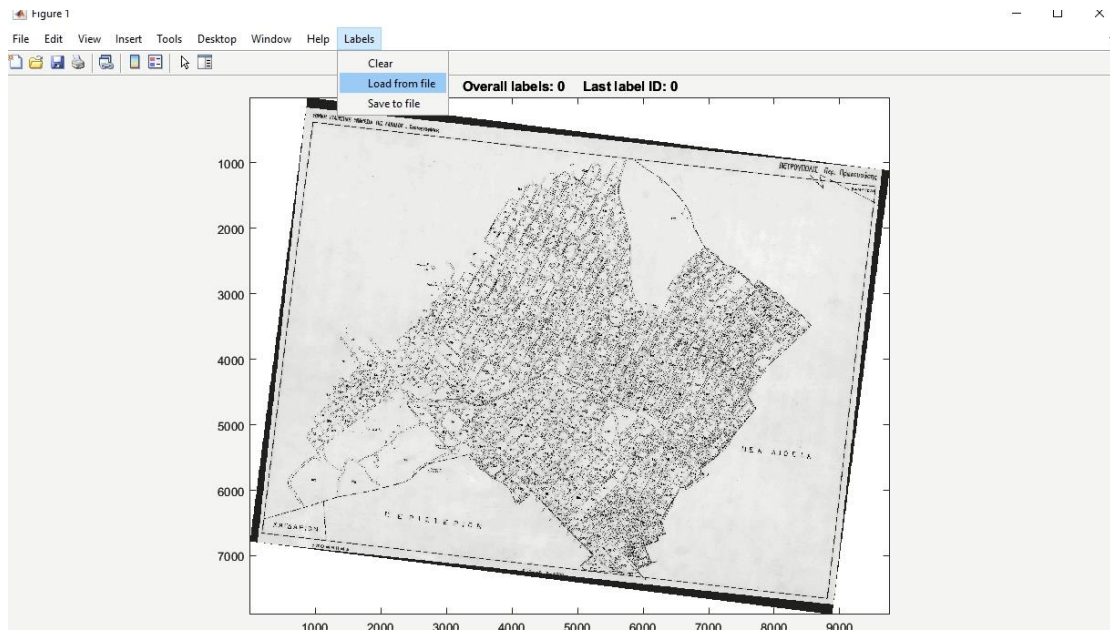
Εικόνα 4.11 Αποσπάσματα χάρτη με όλα τα σπίτια που έχουν επισημανθεί

Το annotation tool προσφέρει και μια μπάρα εργαλείων (toolbar) στο άνω δεξί μέρος της εικόνας για να μπορεί ο χρήστης να κάνει μεγέθυνση ή σμίκρυνση της εικόνας και να μετακινείται πάνω σε αυτήν. Αυτά προστέθηκαν ώστε να διακρίνονται καλύτερα ορισμένες λεπτομέρειες και να γίνεται με μεγαλύτερη ακρίβεια η επισήμανση. Ταυτόχρονα, παρέχεται η δυνατότητα της αποθήκευσης των ήδη επισημασμένων labels σε αρχείο, καθώς και το διάβασμα των labels από αρχείο. Η επιλογή ή από-επιλογή μιας ετικέτας γίνεται με δεξιά κλικ πάνω στο αντίστοιχο κτίριο. Στην περίπτωση που ένα κτίριο είναι κομμένο στην μέση, ή για κάποιο λόγο χρειάζεται να μπει η ίδια επισήμανση σε άλλο μέρος του χάρτη, αυτό γίνεται με τη ροδέλα του ποντικιού. Με αυτό τον τρόπο μπορούν δύο ή περισσότερες περιοχές του χάρτη να έχουν τον ίδιο αριθμό επισήμανσης. Επιπλέον, ο χρήστης μπορεί να διαγράψει όλα τα υπάρχοντα labels και να ξαναρχίσει τη διαδικασία από την αρχή. Τέλος, στο συγκεκριμένο γράφημα αναφέρεται ο συνολικός αριθμός των ετικετών που έχουν δημιουργηθεί καθώς και ο αριθμός της τελευταίας ετικέτας.

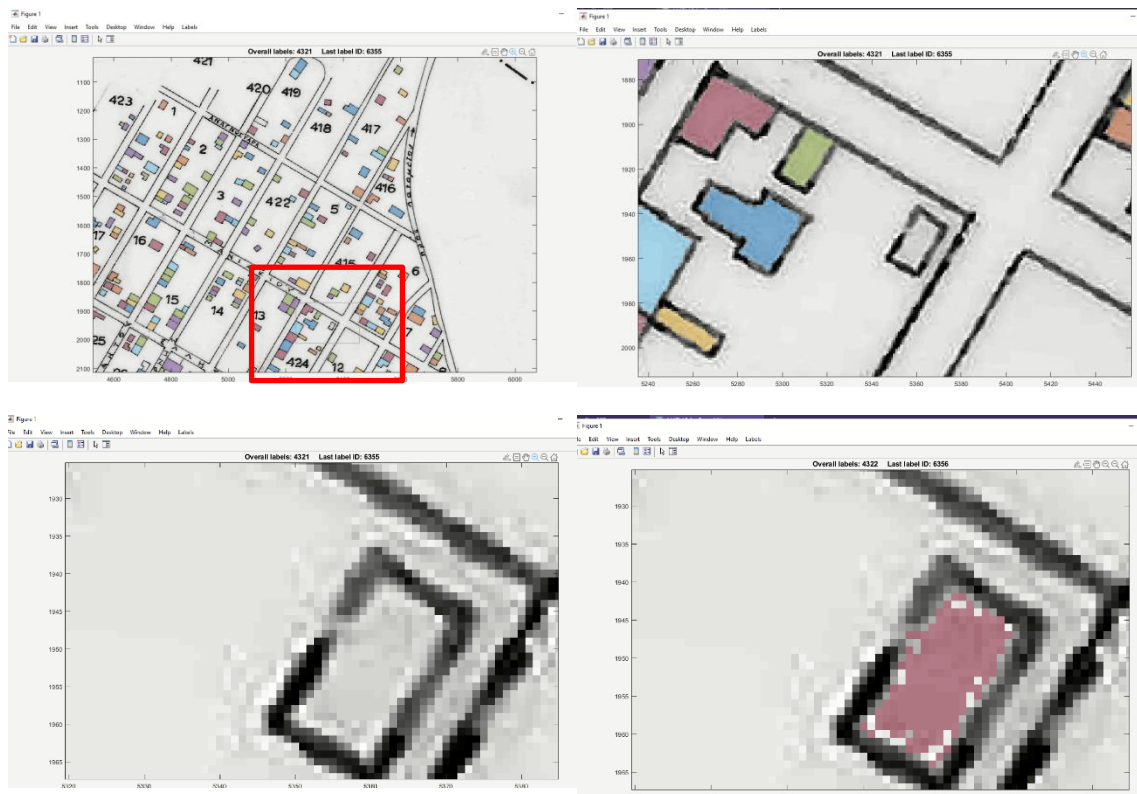


Εικόνα 4.12 Περιβάλλον του annotation tool. (α) toolbar, (β) labels tool, (γ) zoom in, zoom out, pan, home button

Το παρόν εργαλείο μπορεί να διαχειριστεί μεγάλων διαστάσεων εικόνες όπως για παράδειγμα τον συγκεκριμένο χάρτη.



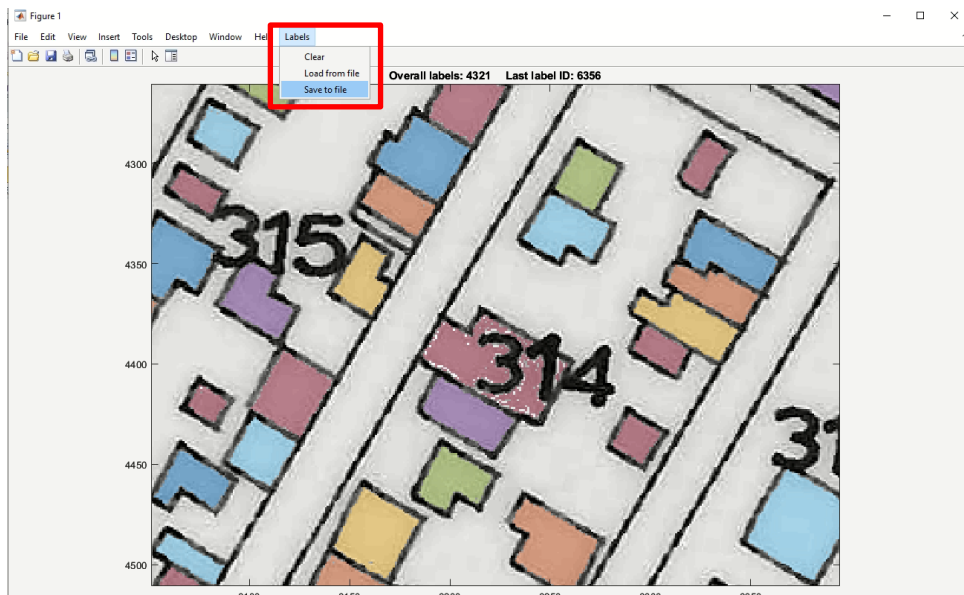
Εικόνα 4.13 Πως μπορεί ο χρήστης να φορτώσει τα ήδη αποθηκευμένα labels



Εικόνα 4.14 Παράδειγμα μεγέθυνσης και επισήμανσης κτιρίου με ένα κλικ

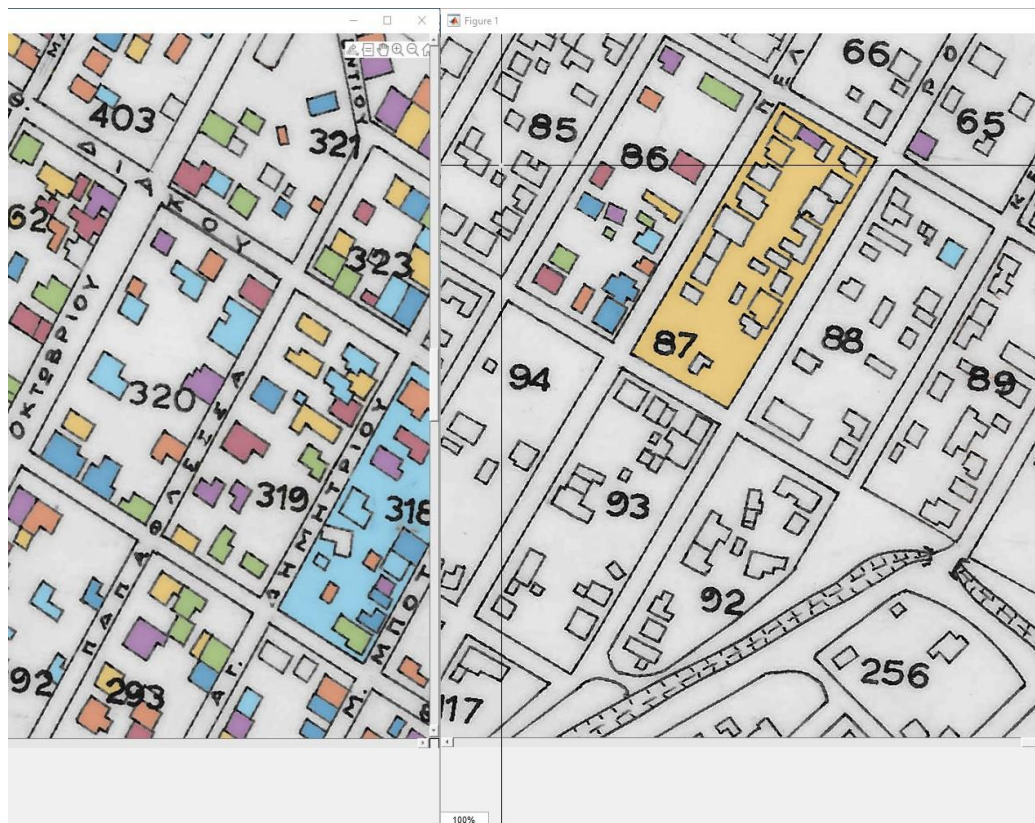


Εικόνα 4.15 Παράδειγμα, πως γίνεται η ίδια επισήμανση σε ένα «κομμένο» κτίριο με την ροδέλα του ποντικιού



Εικόνα 4.16 Πως μπορεί να αποθηκεύσει ο χρήστης τα κτίρια που έχουν επισημανθεί.

Μερικά προβλήματα που αντιμετωπίστηκαν ήταν πως το χρώμα ξεχείλιζε σε μερικά ανοιχτά πολύγωνα. Αυτό αντιμετωπίστηκε «καθαρίζοντας» τον πίνακα Z διαγράφοντας τα διπλόπερασμένα labels.



Εικόνα 4.17 Παράδειγμα προβληματικών οικοδομικών τετραγώνων

4.5 Εξαγωγή περιγραμμάτων των κτιρίων

Όταν ολοκληρωθεί η διαδικασία επισήμανσης των κτιρίων, ο πίνακας Z περιέχει τις επισημάνσεις όλων των κτιρίων, δηλαδή, περιέχει μηδενικά στις θέσεις που δεν υπάρχει κάποιο κτίριο, ενώ έχει την αντίστοιχη τιμή επισήμανσης στα σημεία που υπάρχουν κτίρια. Ο πίνακας μετατρέπεται σε δυαδική μορφή οπότε κάποιο pixel είτε θα είναι μέρος του παρασκηνίου και θα πάρει την τιμή 0 (μαύρο), είτε θα είναι pixel του προσκηνίου (άρα θα ανήκει σε κάποιο κτίριο) και θα πάρει την τιμή 1 (λευκό). Στη συνέχεια, αφού μετατραπεί ο πίνακας Z σε δυαδική εικόνα, μέσω μορφολογικών πράξεων γίνεται ο προσδιορισμός του περιγράμματος των κτιρίων. Για το σκοπό αυτό πραγματοποιήθηκαν κάποιες δοκιμές για το τι δομικό στοιχείο θα χρησιμοποιηθεί. Αρχικά, χρησιμοποιήθηκε ένα οριζόντιο δομικό στοιχείο, το οποίο λόγω της δομής του δημιουργούσε μια παραμόρφωση στον κάθετο άξονα και όλα τα κτίρια ήταν «τραβηγμένα». Έτσι επιλέχθηκε ένα κυκλικό δομικό στοιχείο με ακτίνα 1 το οποίο πραγματοποιούσε μια πιο ομαλή εξομάλυνση.

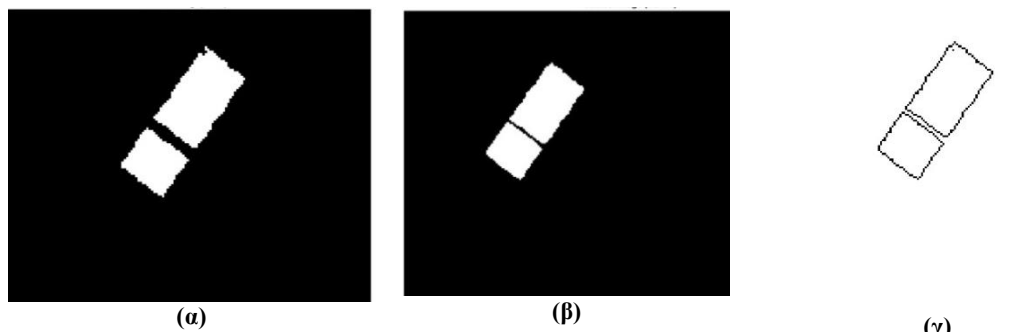
Οι μορφολογικές πράξεις που επιλέχθηκαν βασίστηκαν στον υπολογισμό της μορφολογικής κλίσης και είναι i) μια διαστολή η οποία οδήγησε σε αύξηση των διαστάσεων του κάθε κτιρίου και ii) μια συστολή η οποία οδήγησε σε μείωση των διαστάσεων τους. Από την διαφορά των εικόνων που προέκυψαν μέσω αυτών των δυο πράξεων, εξήχθη το επιθυμητό αποτέλεσμα δηλαδή το περίγραμμα των σπιτιών που έχουν επισημανθεί.

Δηλαδή:

$$(A \oplus B) - (A \ominus B)$$

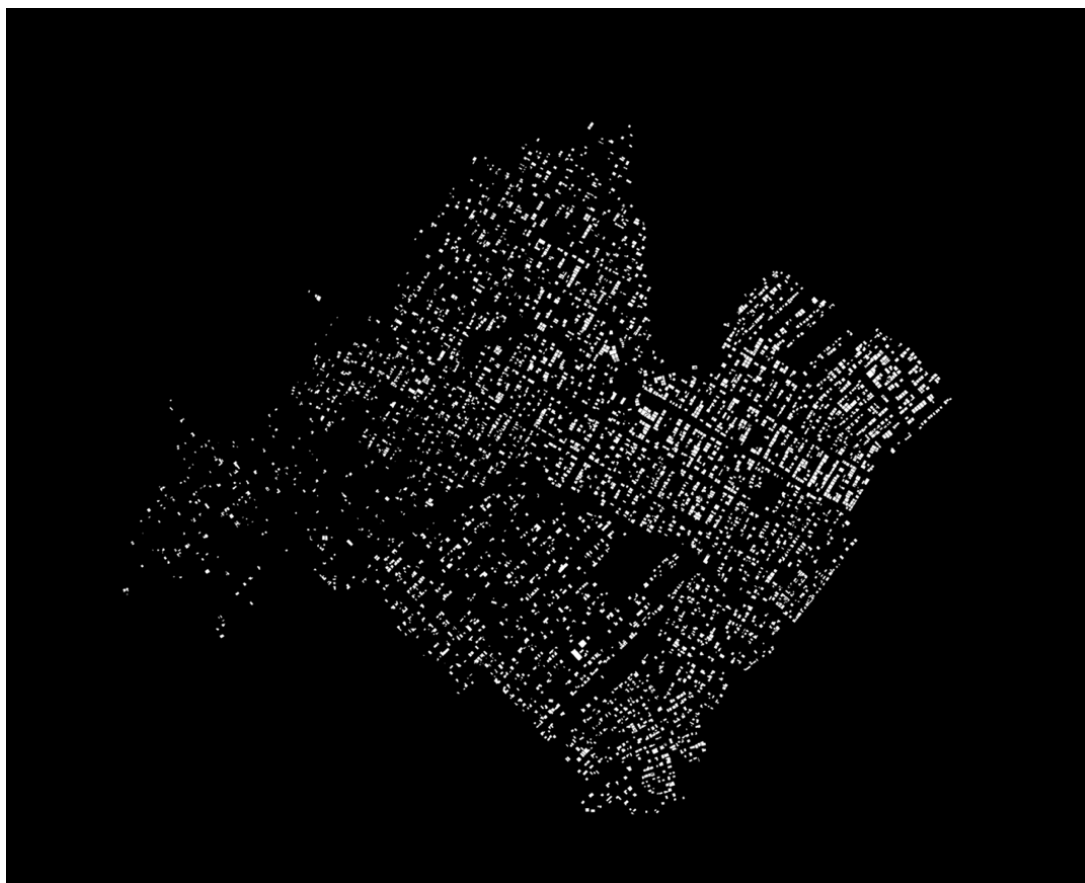
4.5.1

όπου A είναι η αρχική εικόνα και B το κυκλικό δομικό στοιχείο.

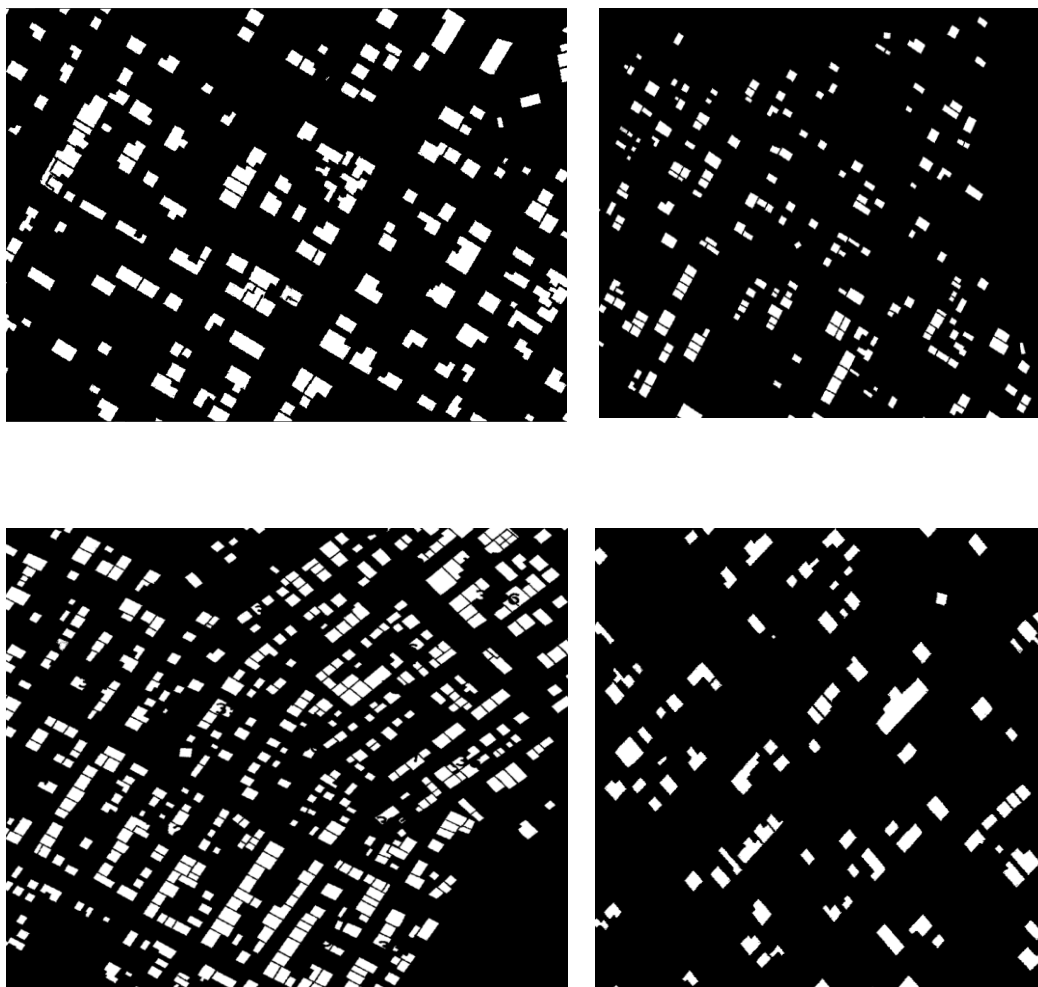


Εικόνα 4.18 (α) Συστολή ενός περιγράμματος, (β) Διαστολή ενός περιγράμματος, (γ) Μορφολογική κλίση

Στην Εικόνα 4.19 και Εικόνα 4.20 φαίνονται όλα τα αποτυπώματα των κτιρίων τα οποία έχουν επισημανθεί και αποτυπώνονται με δυαδική μορφή (0,1) στην παρακάτω εικόνα.

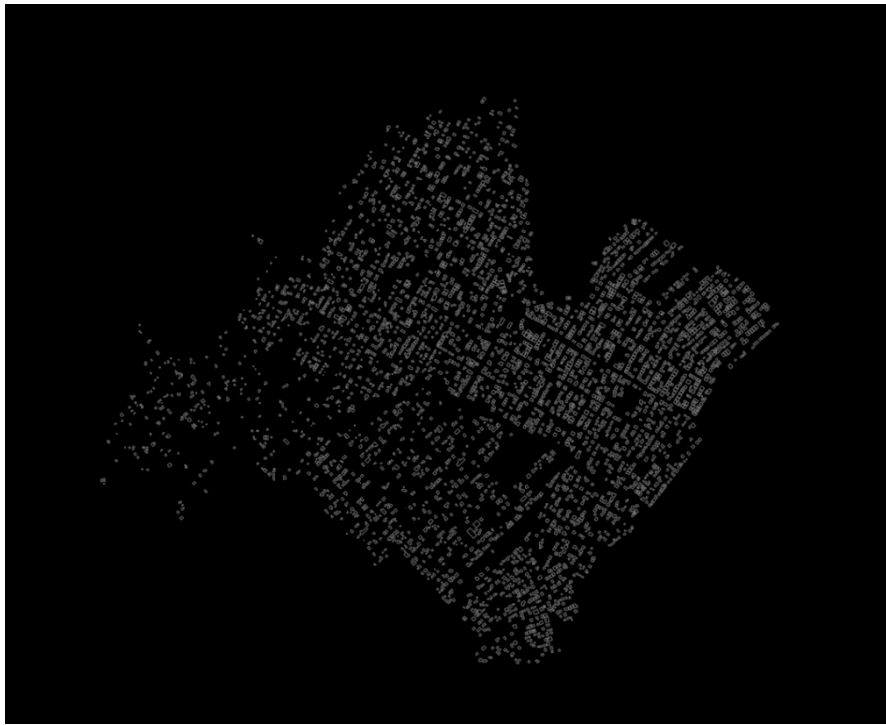


Εικόνα 4.19 Αρχική ασπρόμαυρη εικόνα που περιέχει όλα τα κτίρια που έχουν επισημανθεί

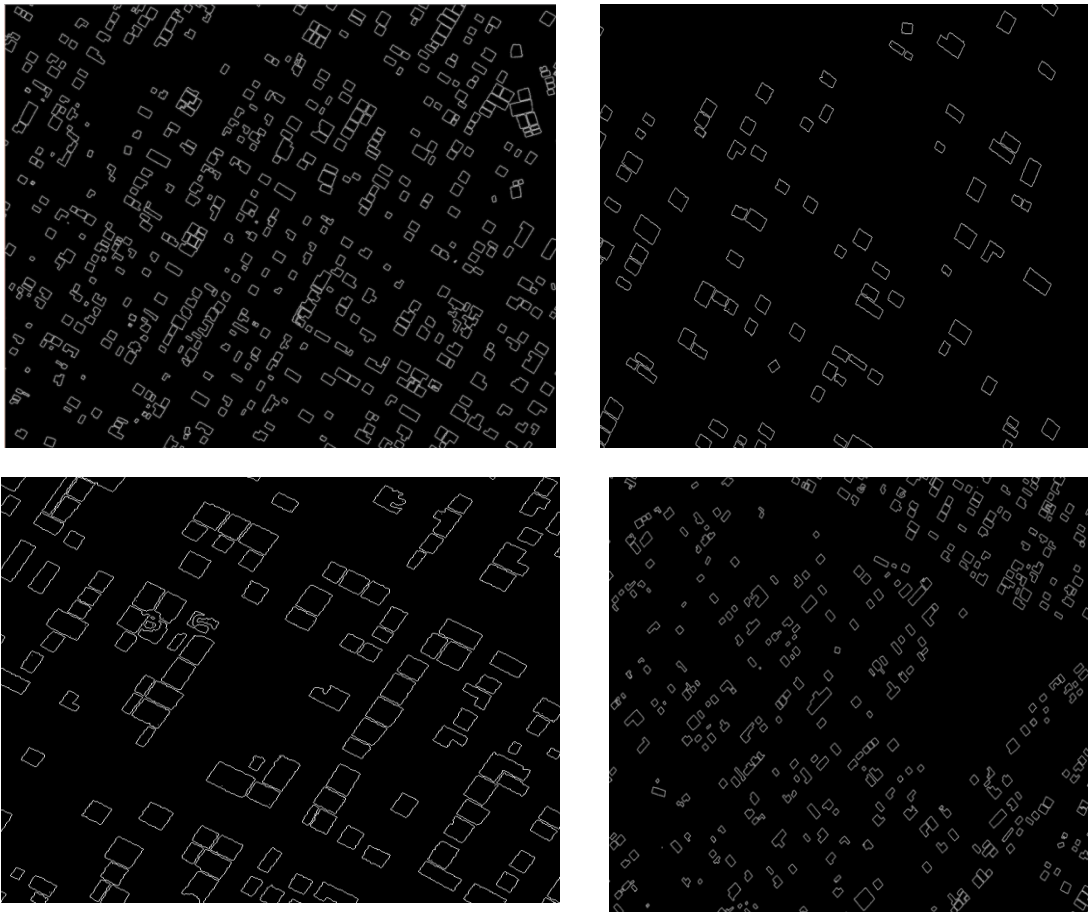


Εικόνα 4.20 Αποσπάσματα περιγραμμάτων

Κατόπιν, μέσω μορφολογικών πράξεων (μορφολογικής κλίσης) γίνεται ο προσδιορισμός του περιγράμματος των κτιρίων και αποτυπώνονται με δυαδική μορφή, όπως φαίνεται στην **Εικόνα 4.21** και **Εικόνα 4.22**.



Εικόνα 4.21 Μορφολογική κλίση των περιγραμμάτων



Εικόνα 4.22 Απόσπασμα περιγραμμάτων με μορφολογική κλίση

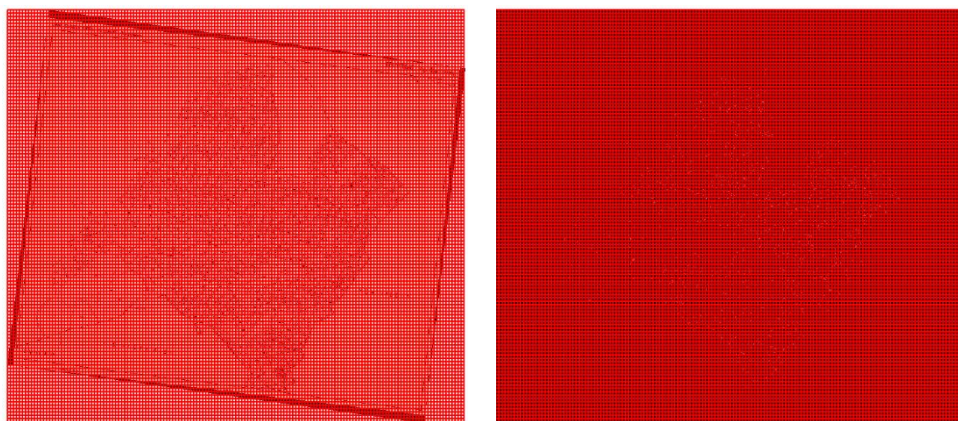
Στην παρούσα διπλωματική μέσω του προτεινόμενου annotation tool επισημάνθηκαν 4.321 περιγράμματα σπιτιών στον τοπογραφικό χάρτη της Πετρούπολης. Από τα συνολικά περιγράμματα που υπάρχουν στον χάρτη, τα «προβληματικά» στα οποία δεν μπόρεσε να εξαχθεί το περίγραμμα δεν ξεπερνούν το 10%. Οι λόγοι αδυναμίας οφείλονται είτε στη ύπαρξη επικαλύψεων γραμμμάτων πάνω στα κτίρια είτε στο γεγονός ότι δεν ήταν κλειστό το πολύγωνο του κτιρίου οπότε αποτύγχανε η flood fill μέθοδος.

4.6 Κατάτμηση σε υπο-εικόνες

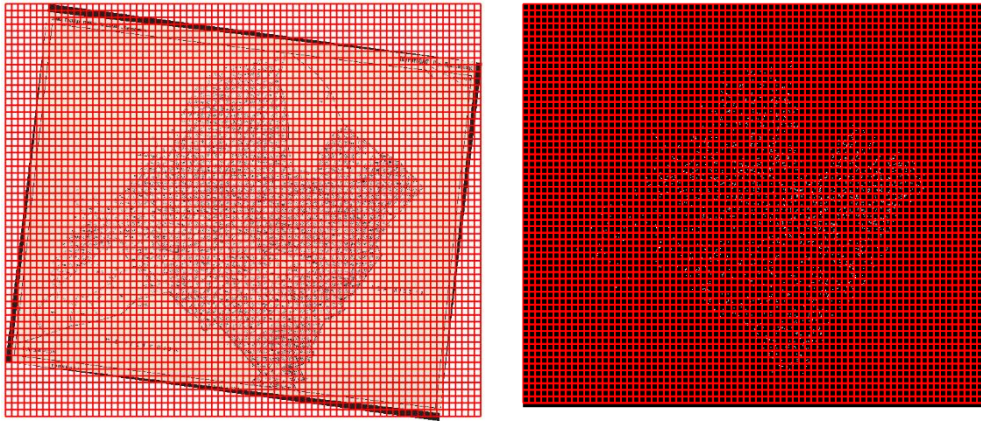
Έχοντας προσδιορίσει τα περιγράμματα των κτιρίων υπάρχει πλέον διαθέσιμη η κατάλληλη πληροφορία για την εκπαίδευση ενός νευρωνικού δικτύου σκοπός του οποίου είναι να λαμβάνει στην είσοδο έναν τοπογραφικό χάρτη και να παράγει στην έξοδο μια εικόνα, ιδίων διαστάσεων με την αρχική, που να περιέχει μόνο τα περιγράμματα των κτιρίων. Για την εκπαίδευση του νευρωνικού δικτύου απαιτείται να δημιουργηθεί ένας μεγάλος αριθμός από ζεύγη υπο-εικόνων εισόδου – εξόδου. Στην παρούσα εργασία χρησιμοποιήθηκε το εργαλείο PatchIt για να παραχθούν οι υπο-εικόνες αυτές τα λεγόμενα patches. Χρησιμοποιήθηκαν δυο προσεγγίσεις, η Πλεγματική (Grid) και η τυχαία (Random). Σε κάθε προσέγγιση δημιουργήθηκαν patches με τρία διαφορετικά μεγέθη. Παίρνοντας σαν δεδομένο εισόδου την αρχική έγχρωμη εικόνα του χάρτη της Πετρούπολης, δημιουργήθηκαν αρχικά με την μέθοδο Grid

- 18.696 patches μεγέθους 64x64
- 4.636 patches μεγέθους 128x128 και
- 1.140 patches μεγέθους 256x256

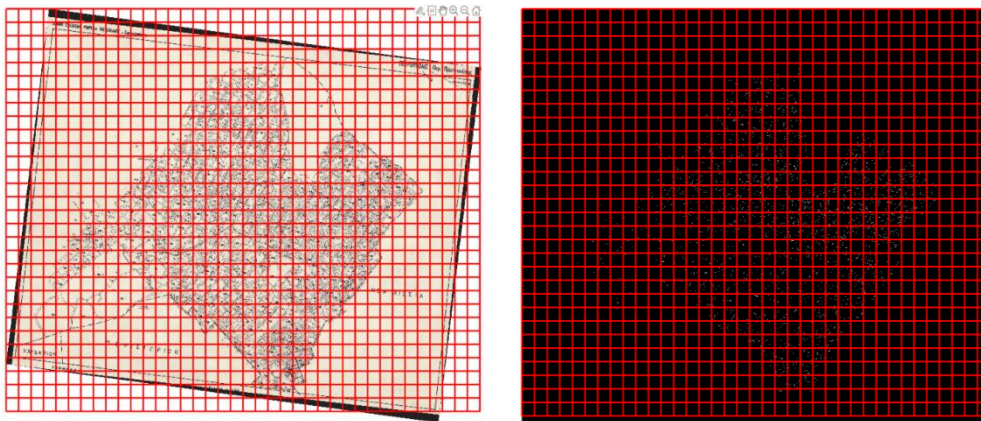
Διατηρώντας το ίδιο βήμα πλέγματος, δημιουργήθηκαν τα αντίστοιχα patches για την εικόνα με τα αληθή στοιχεία, όπως φαίνονται στις εικόνες Εικόνα 4.23, Εικόνα 4.24 και Εικόνα 4.25 αντίστοιχα. Να σημειωθεί πως το πλέγμα διαστάσεων 128x128 και 256x256 δεν καλύπτει όλο το περίγραμμα της εικόνας καθώς το μέγεθος της υπο-εικόνας δεν «χωράει» ώστε να δημιουργηθεί άλλη μια γραμμή πλέγματος.



Εικόνα 4.23 Πλεγματική μορφή patches 64x64 για την αρχική εικόνα (αριστερά) και την εικόνα με τα περιγράμματα των κτιρίων (δεξιά)

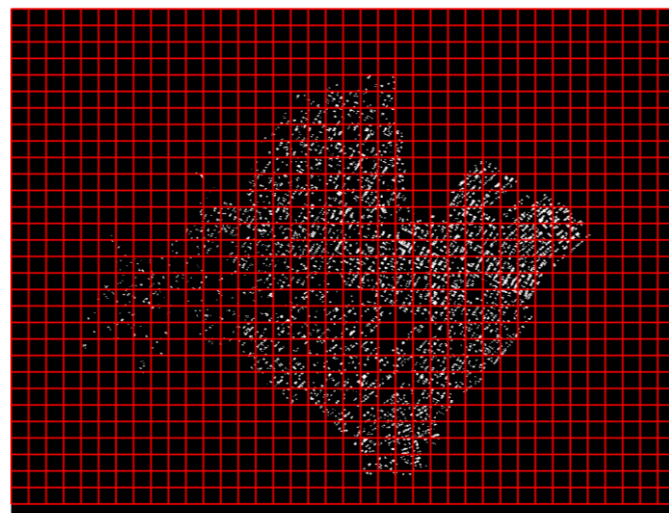
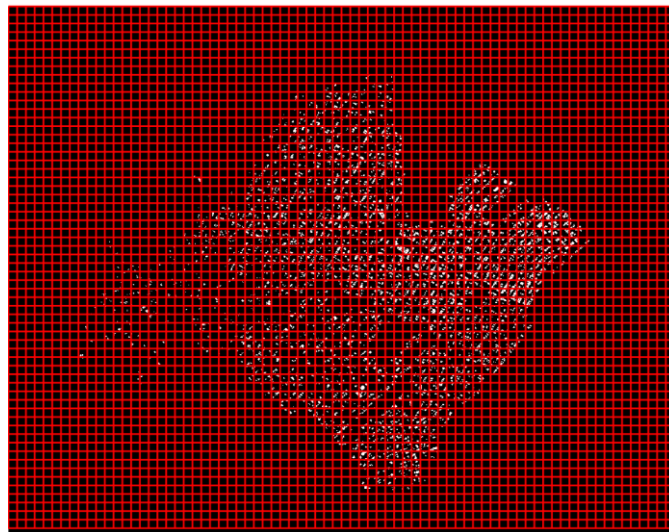
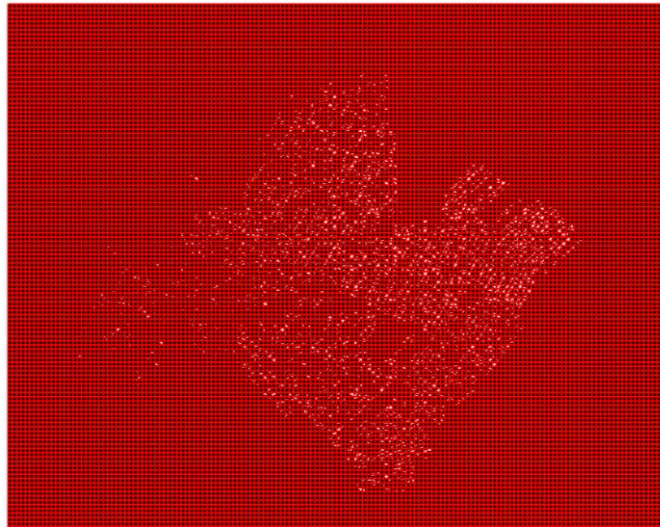


Εικόνα 4.24 Πλεγματική μορφή patches 128x128 για την αρχική εικόνα (αριστερά) και την εικόνα με τα περιγράμματα των κτιρίων (δεξιά)



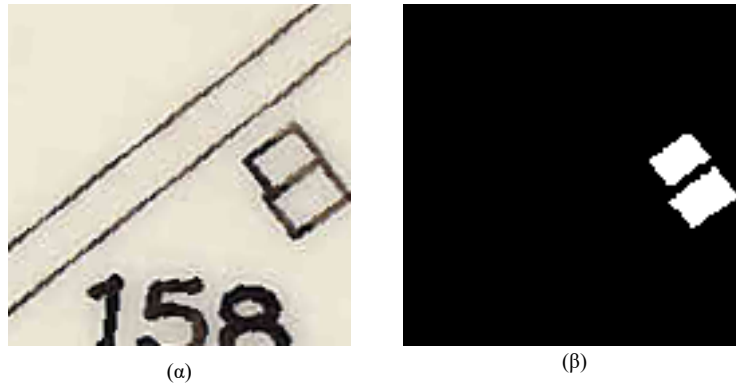
Εικόνα 4.25 Πλεγματική μορφή patches 256x256 για την αρχική εικόνα (αριστερά) και την εικόνα με τα περιγράμματα των κτιρίων (δεξιά)

Ακόμα, δημιουργήθηκε ο ίδιο αριθμός patches και για την εικόνα του πίνακα Z ο οποίος δεν έχει υποστεί καμία μορφολογική επεξεργασία και περιέχει την περιγραφή των κτιρίων με διαφορετική αρίθμηση το καθένα.

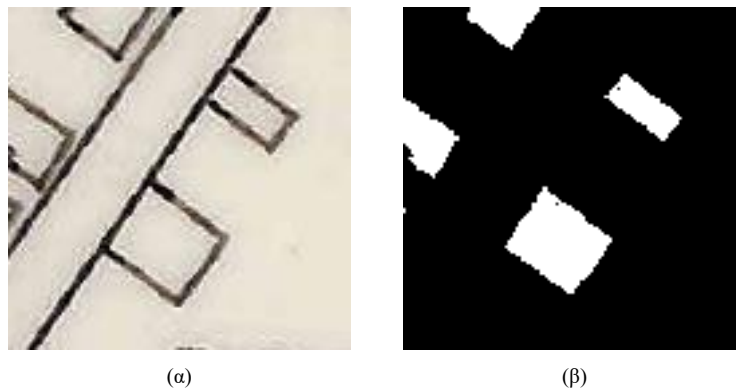


Εικόνα 4.26 Πλεγματική μορφή ground truth επισημασμένων κτιρίων

Παρακάτω φαίνονται δύο παραδείγματα patches τόσο στην αρχική τοπογραφική εικόνα όσο τα αντίστοιχα patches στην εικόνα με τις επισημάνσεις.



Εικόνα 4.27 Patches 256x256 με τα αντίστοιχα επισημασμένα κτίρια.

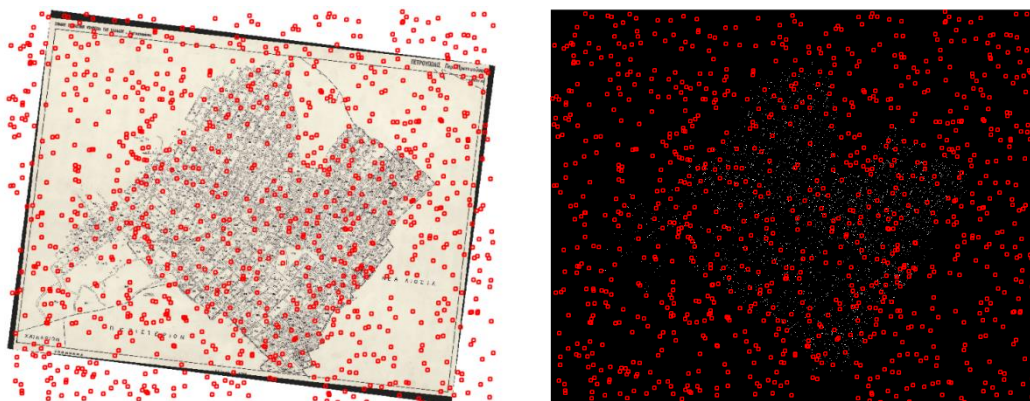


Εικόνα 4.28 Patches 256x256 με τα αντίστοιχα επισημασμένα κτίρια

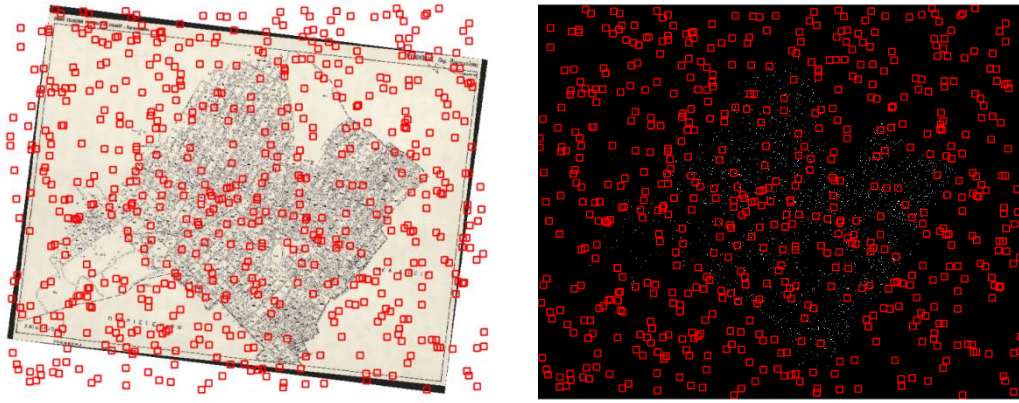
Στην μέθοδο Random δημιουργήθηκαν

- 1000 patches μεγέθους 64x64
- 600 patches μεγέθους 128x128
- 200 patches μεγέθους 256x256

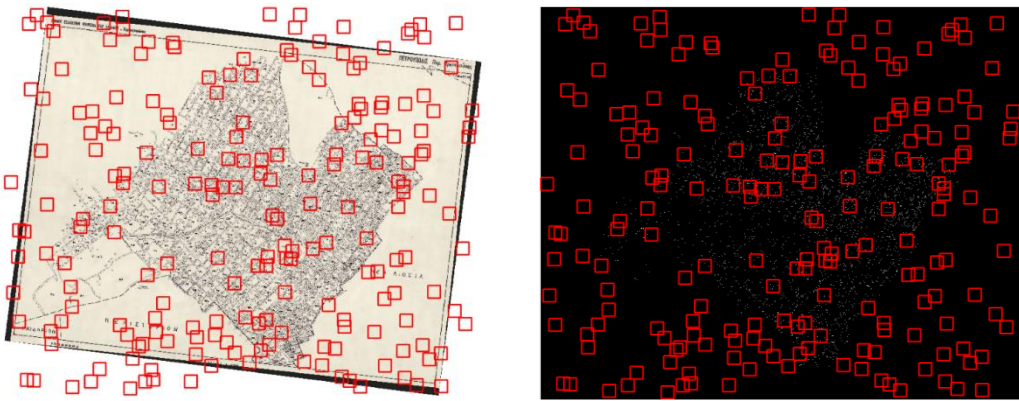
Σε κάθε περίπτωση η μέγιστη μεταξύ τους κάλυψη ορίστηκε στο 70%. Στις παρακάτω εικόνες απεικονίζονται παραδείγματα από random patches διαφόρων μεγεθών τόσο με περιγράμματα όσο και με τις αρχικές επισημάνσεις ολόκληρων των κτιρίων.



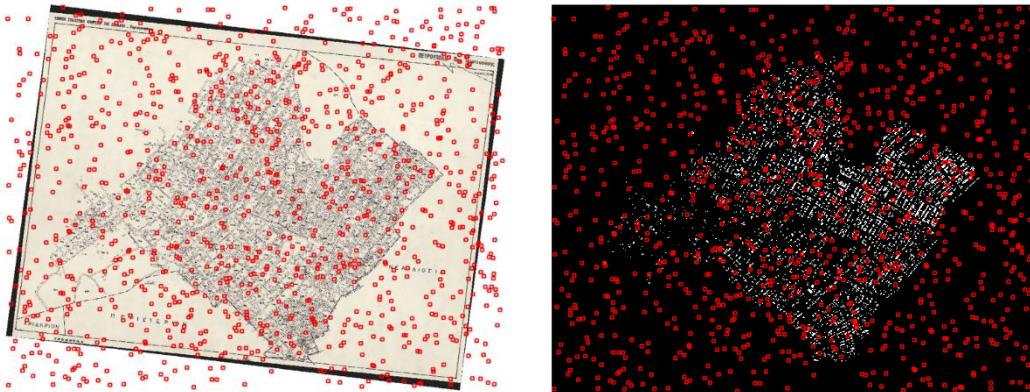
Εικόνα 4.29 Random patches 64x64 με τα αντίστοιχα επισημασμένα περιγράμματα



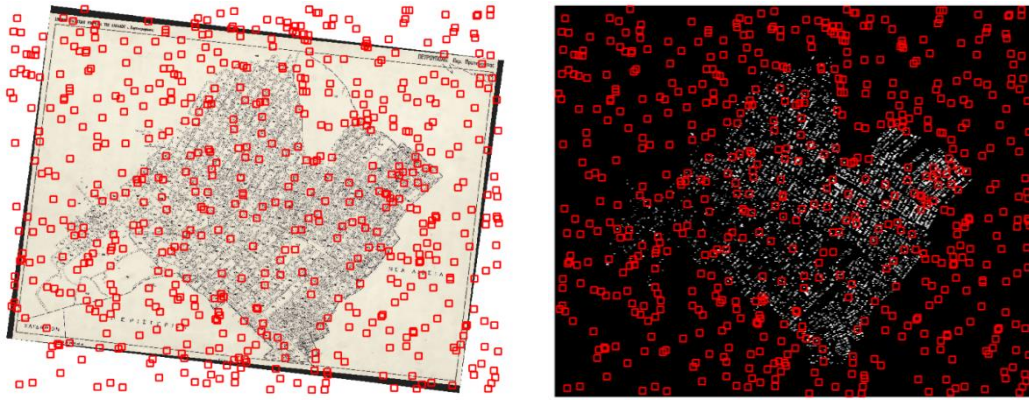
Εικόνα 4.30 Random patches 128x128 με τα αντίστοιχα επισημασμένα περιγράμματα



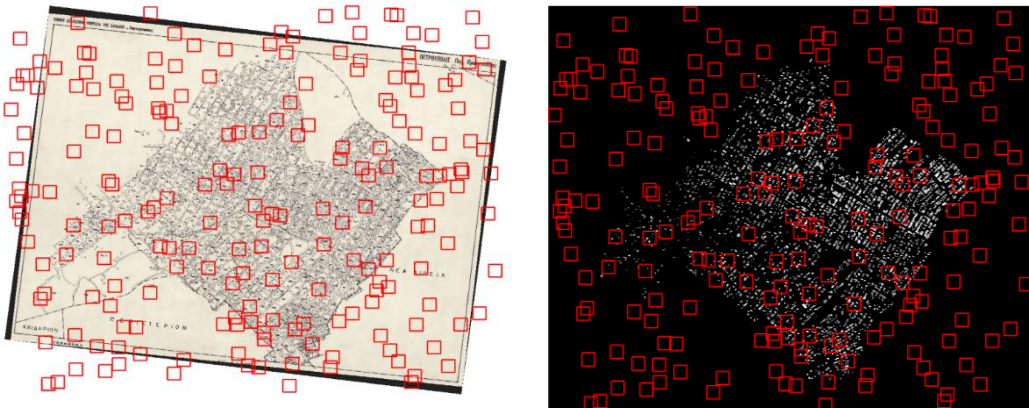
Εικόνα 4.31 Random patches 256x256 με τα αντίστοιχα επισημασμένα περιγράμματα



Εικόνα 4.32 Random patches 64x64 με τα αντίστοιχα επισημασμένα κτίρια



Εικόνα 4.33 Random patches 128x128 με τα αντίστοιχα επισημασμένα κτίρια

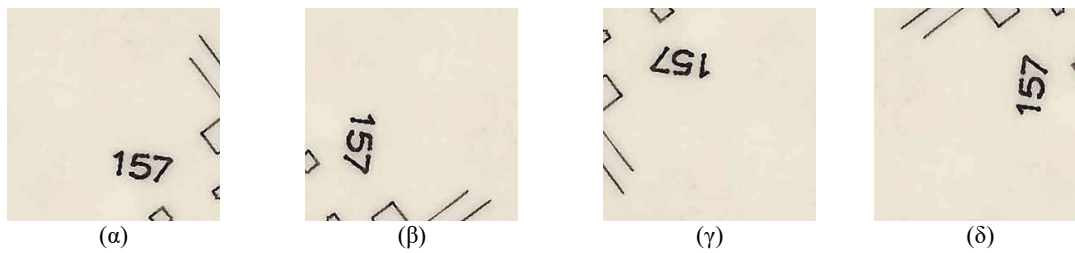


Εικόνα 4.34 Random patches 256x256 με τα αντίστοιχα επισημασμένα κτίρια

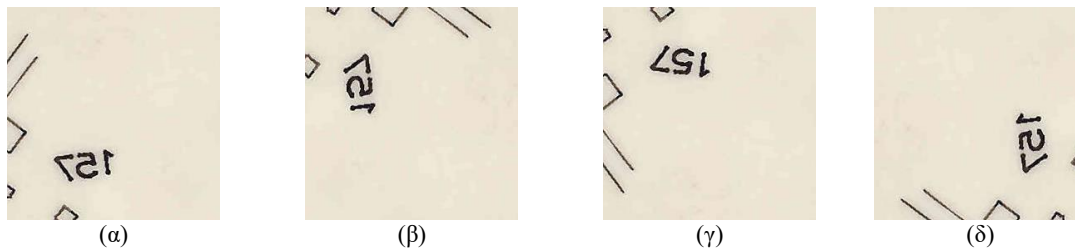
4.7 Επαύξηση δεδομένων

Προκειμένου να εμπλουτισθούν ακόμη περισσότερο τα δεδομένα εκπαίδευσης του νευρωνικού δικτύου χρησιμοποιήθηκε η λειτουργία περιστροφής και αναστροφής του εργαλείου PatchIt προκειμένου να γίνει επαύξηση των διαθέσιμων patches. Η επαύξηση των δεδομένων είναι σημαντική διότι αφενός εμπλουτίζεται το σετ μάθησης των δεδομένων εισόδου του νευρωνικού δικτύου και αφετέρου αυξάνεται η ακρίβεια του. Με τον τρόπο αυτό, το νευρωνικό δίκτυο δημιουργεί καλύτερα και πιο αξιόπιστα μοντέλα κατά την εκπαίδευση του.

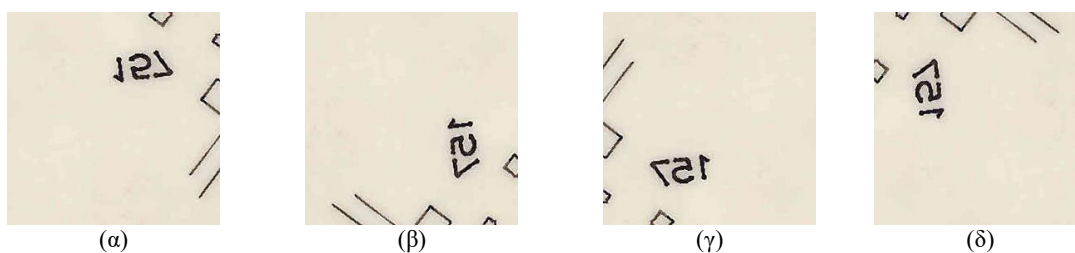
Αρχικά, σε κάθε σετ δεδομένων πραγματοποιήθηκε στροφή του κάθε patch κατά 90,180 και 270 μοίρες. Στην συνέχεια, πραγματοποιήθηκε αναστροφή κατά τον οριζόντιο και τον κατακόρυφο άξονα κατά τις ίδιες μοίρες 90,180 και 270, αντίστοιχα. Επιλέχθηκαν αυτές οι στροφές επειδή οι τυχαίες στροφές δημιουργούν ένα μαύρο πλαίσιο δηλαδή μαύρα pixel περιμετρικά της εικόνας τα οποία επηρεάζουν την κατάτμηση της. Συνολικά έγιναν δέκα παραλλαγές μαζί με την αρχική εικόνα και δημιουργήθηκαν με αυτό τον τρόπο 12 σετ δεδομένων.



Εικόνα 4.35 α) αρχικό patch 64x64, β) στροφή 90°, γ) στροφή 180°, δ) στροφή 270°



Εικόνα 4.36 α) οριζόντια αναστροφή patch, β) στροφή 90°, γ) στροφή 180°, δ) στροφή 270°



Εικόνα 4.37 α) patch με κάθετη αναστροφή, β) στροφή 90°, γ) στροφή 180°, δ) στροφή 270°

Να σημειωθεί πως επειδή το περίγραμμα της εικόνας καταλαμβάνει το μεγαλύτερο μέρος της, είναι φυσικό ένας πολύ μεγάλος αριθμός από patches, να μην περιέχουν πληροφορία. Έτσι στην εκπαίδευση του νευρωνικού δικτύου δεν συμπεριελήφθησαν αυτά τα patches. Τα patches τα οποία όμως συμπεριελήφθησαν περιείχαν τουλάχιστον ένα ενεργό (λευκό) pixel. Ο Πίνακας 1 και ο Πίνακας 2 παρουσιάζουν τα στατιστικά των υπο-εικόνων. Ενδεικτικά, στα patches 64x64 στην πλεγματική προσέγγιση δημιουργήθηκαν συνολικά, συμπεριλαμβανομένων και των επαυξημένων, 186.960 patches εκ των οποίων τα 168.264 ήταν επαυξημένα. Από αυτά το 71% δεν περιείχε πληροφορία.

GRID	OUTLINES & ANNOTATED		
	64X64	128X128	256X256
ΜΕΓΕΘΟΣ PATCH			
PATCHES	18.696	4.636	1.140
ΕΠΑΥΞΗΜΕΝΑ PATCHES	168.264	41.724	10.260
ΣΥΝΟΛΙΚΑ	186.960	46.360	11.400
ΠΟΣΟΣΤΟ ΕΠΑΥΞΗΜΕΝΩΝ PATCHES	90%	90%	90%
ΑΡΙΘΜΟΣ PATCHES ΜΕ ΠΛΗΡΟΦΟΡΙΑ	41.820	13.430	3.870
ΑΡΙΘΜΟΣ PATCHES ΧΩΡΙΣ ΠΛΗΡΟΦΟΡΙΑ	145.140	32.930	7.530
ΠΟΣΟΣΤΟ PATCHES ΧΩΡΙΣ ΠΛΗΡΟΦΟΡΙΑ	71%	59%	49%

Πίνακας 1 Στατιστικά των υπο-εικόνων για την πλεγματική προσέγγιση

RANDOM	OUTLINES & ANNOTATED		
ΜΕΓΕΘΟΣ PATCH	64X64	128X128	256X256
NORMAL PATCHES	1.000	600	200
ΕΠΑΥΞΗΜΕΝΑ	9.000	5.400	1.800
ΣΥΝΟΛΙΚΑ	10.000	6.000	2.000
ΠΟΣΟΣΤΟ ΕΠΑΥΞΗΜΕΝΩΝ PATCHES	90%	90%	90%
ΑΡΙΘΜΟΣ PATCHES ΜΕ ΠΛΗΡΟΦΟΡΙΑ	2.411	1.660	640
ΑΡΙΘΜΟΣ PATCHES ΧΩΡΙΣ ΠΛΗΡΟΦΟΡΙΑ	7.589	4.340	1.360
ΠΟΣΟΣΤΟ PATCHES ΧΩΡΙΣ ΠΛΗΡΟΦΟΡΙΑ	68%	62%	53%

Πίνακας 2 Στατιστικά των υπο-εικόνων για την τυχαία προσέγγιση

4.8 Εκπαίδευση νευρωνικού δικτύου U-NET

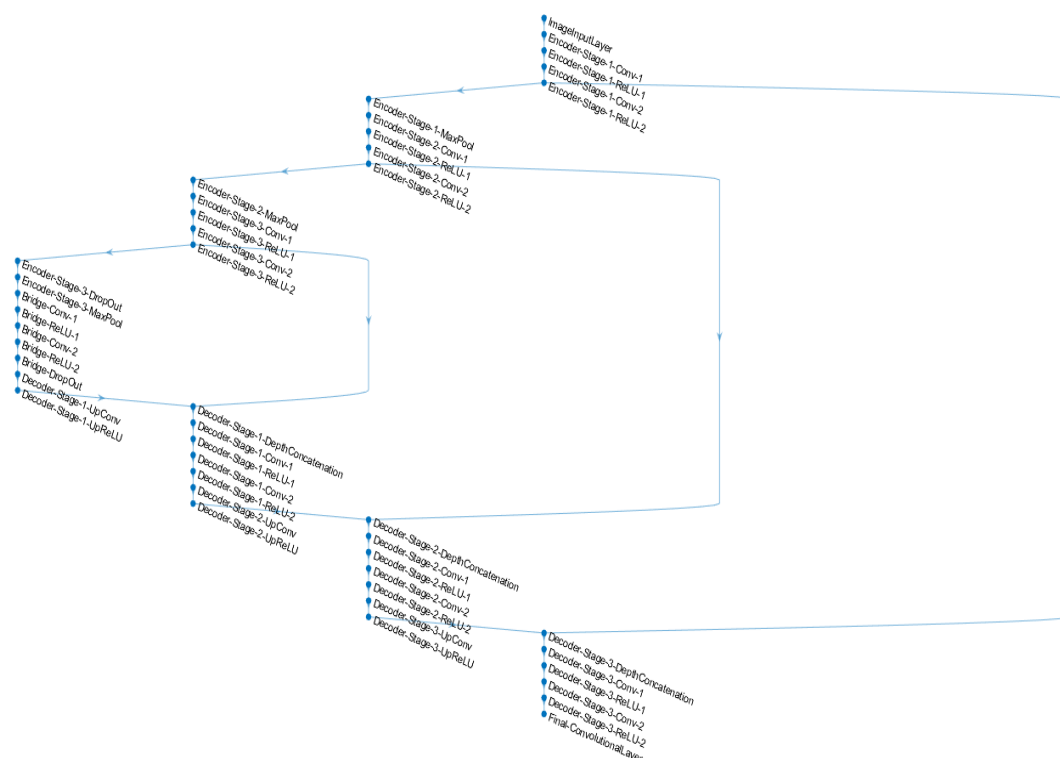
Για την επίλυση του προβλήματος της αναγνώρισης και της εξαγωγής κτιρίων από ψηφιακές εικόνες χρησιμοποιήθηκε μια προσέγγιση βαθιάς συνελκτικής νευρωνικής δικτύωσης (CNN) που βασίζεται στην αρχιτεκτονική U-Net [16]. Το CNN εκπαιδεύεται χρησιμοποιώντας έναν μεγάλο αριθμό patches εικόνας ως ένα βαθύ μοντέλο παλινδρόμησης μεταξύ εικόνων για τον επιτυχή διαχωρισμό των κτιρίων από τον χάρτη και την αγνόηση του υπόλοιπου περιεχομένου. Αυτό το σύνολο δεδομένων χρησιμοποιείται τόσο για την εκπαίδευση του συστήματος όσο και για τον έλεγχο της αποτελεσματικότητάς του, καθώς η επιθυμητή απόκριση του δικτύου παρέχεται για κάθε σύνολο δεδομένων εισόδου. Η αποτελεσματικότητα της μεθόδου αξιολογείται μέσω μιας σειράς πειραμάτων με διαφορετικές μεθόδους συλλογής δειγμάτων εικόνων και διαφορετικά μεγέθη δειγμάτων. Η αποτελεσματικότητα της μεθόδου αξιολογείται ποσοτικά μέσω κατάλληλων μετρικών που συγκρίνουν τις τιμές εξόδου του CNN με τις αντίστοιχες τιμές του συνόλου δεδομένων επικύρωσης έναντι του συνόλου δεδομένων ελέγχου.

Η αρχιτεκτονική του U-Net δικτύου αποτελείται από δύο κύρια μέρη. Το πρώτο μέρος είναι ο κωδικοποιητής, ο οποίος επεξεργάζεται την εικόνα εισόδου "περνώντας" την από μια συστολική διαδρομή, όπου η χωρική διάσταση μειώνεται και η διάσταση βάθους επεκτείνεται μέσω συνεχούς συνέλιξης, ReLU και μέγιστες συγκεντρώσεις. Το δεύτερο μέρος είναι ο αποκωδικοποιητής, ο οποίος επεξεργάζεται το προϊόν του κωδικοποιητή. Αυτό περνάει μέσα από μια συστολική διαδρομή όπου αυξάνονται οι χωρικές διαστάσεις του. Στη συνέχεια, μειώνεται σταδιακά η διάσταση του βάθους μέσω διαδοχικών συνλίξεων, ReLU και ανεστραμμένες συνέλιξεις.

Ένα ιδιαίτερο πλεονέκτημα της αρχιτεκτονικής U-Net είναι η ικανότητά της να βρίσκει τη βέλτιστη ισορροπία μεταξύ θέσης και περιεχομένου. Καθώς οι διαστάσεις πλάτους και μήκους της εικόνας εισόδου μειώνονται, τα φίλτρα στα βαθύτερα στρώματα εντός της διαδρομής συνέλιξης εστιάζουν σε μεγαλύτερα πεδία υποδοχής (μειώνοντας την πληροφορία θέσης στην εικόνα), αλλά αντίστοιχα, κάθε φορά που ο όγκος του χάρτη χαρακτηριστικών περνάει από το στρώμα συνέλιξης, το κανάλι ή η διάσταση βάθους σταδιακά αυξάνεται (αυξάνοντας την πληροφορία περιεχομένου στην εικόνα), με αποτέλεσμα την εκμάθηση σύνθετων χαρακτηριστικών μέσα σε αυτό το στρώμα συνέλιξης (αυξάνοντας την πληροφορία περιεχομένου στην εικόνα).

Η χωρική διάσταση της εικόνας αυξάνεται σταδιακά και η διάσταση βάθους μειώνεται σταδιακά (μειώνοντας την πληροφορία περιεχομένου στην εικόνα). Για να αυξηθεί η ακρίβεια της ανίχνευσης θέσης, χρησιμοποιούνται συνδέσεις παράλειψης σε κάθε βήμα του

αποκωδικοποιητή, όπου η έξοδος του στρώματος αντίστροφης συνέλιξης και οι χάρτες χαρακτηριστικών από τον κωδικοποιητή συνδυάζονται στο ίδιο στρώμα. Τα δύο υπο-δίκτυα του κωδικοποιητή και του αποκωδικοποιητή αποτελούνται από διάφορα στάδια. Για την αρχιτεκτονική της Εικόνα 4.38, ο αριθμός των σταδίων είναι ίσος με τρία. Κάθε στάδιο κωδικοποιητή αποτελείται από δύο σύνολα στρωμάτων συνέλιξης και ReLU, ακολουθούμενα από ένα στρώμα μέγιστης συνάθροισης 2x2. Κάθε στάδιο αποκωδικοποιητή αποτελείται από ένα ζεύγος στρωμάτων αντίστροφης συνέλιξης, ακολουθούμενο από δύο στρώματα συνέλιξης και ένα στρώμα ReLU. Ο στόχος του προτεινόμενου CNN είναι να προβλέψει, για κάθε εικονοστοιχείο στην εικόνα εισόδου, μια τιμή όσο το δυνατόν πιο κοντά στην τιμή του αντίστοιχου εικονοστοιχείου στο τμήμα εικόνας των δεδομένων επικύρωσης [11].



Εικόνα 4.38 Αρχιτεκτονική U-Net

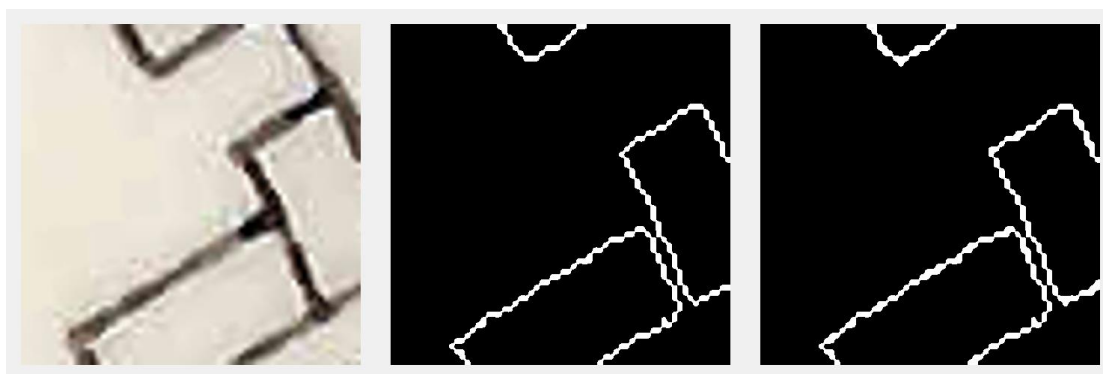
4.9 Πειραματικά Αποτελέσματα

Για την εκπαίδευση του U-Net δικτύου χρησιμοποιήθηκαν κάποιες παράμετροι οι οποίες βελτιστοποιούν τόσο την διαδικασία μάθησης όσο και τον απαιτούμενο χρόνο εκπαίδευσης. Συγκεκριμένα, χρησιμοποιήθηκαν οι παράμετροι που περιγράφονται στην διπλωματική εργασία [11]. Με τα δώδεκα διαφορετικά σετ δεδομένων, πραγματοποιήθηκαν δώδεκα διαφορετικές διαδικασίες εκπαίδευσης. Σε όλες αυτές, για τον υπολογισμό των τιμών του δικτύου χρησιμοποιήθηκε ο αλγόριθμος Stochastic Gradient Descent (SGD), σε συνδυασμό με τον αλγόριθμο βελτιστοποίησης του Adam. Ο αριθμός εποχών ορίστηκε σε 100, χρησιμοποιώντας ένα mini-batch, ίσο με 8, με ένα βαθμό μάθησης ίσο με 0.001, ο οποίος παρέμεινε σταθερός σε όλη τη διάρκεια των 12 διαφορετικών εκπαιδύσεων (δε χρησιμοποιήθηκε προσαρμοστικός βαθμός μάθησης). Σχετικά με τις υπέρ-παραμέτρους του αλγόριθμου Adam, χρησιμοποιήθηκαν οι τιμές τις οποίες προτείνουν ως βέλτιστες στην

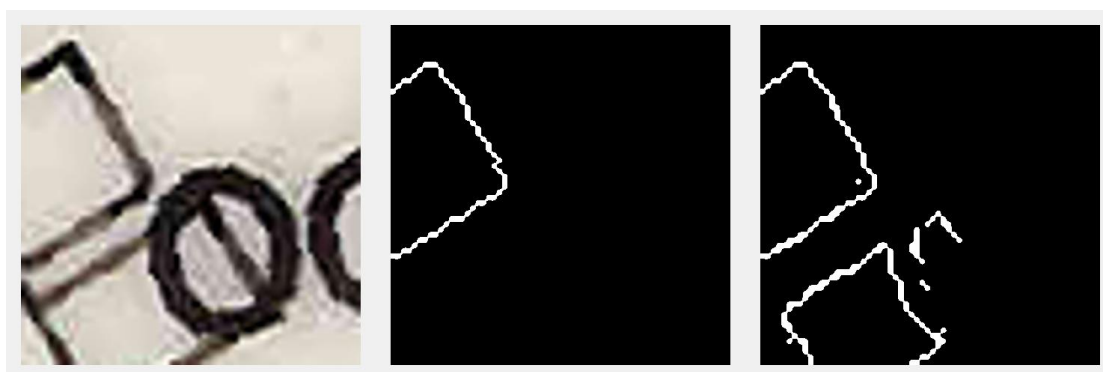
εργασία [35] εκτός της τιμής του συντελεστή μείωσης των παλαιότερων τετραγωνισμένων κλίσεων (β_2), όπου αντί για 0,999, εφαρμόστηκε τιμή ίση με 0,99 [11]. Η πραγματοποίηση των προαναφερθέντων υλοποιήθηκε σε υπολογιστικό σύστημα με τα εξής χαρακτηριστικά: Core i9-2TB HDD + 1TB SSD - 128GB RAM - RTX 3090 | 8 Cores @ 5.3 GHz - 11th Gen CPU - 24GB GDDR6X.

Grid patches – Περιγράμματα

Τα παρακάτω παραδείγματα αφορούν το δίκτυο το οποίο έχει εκπαιδευτεί με τα περιγράμματα των κτιρίων χρησιμοποιώντας grid patches. Σε κάθε παράδειγμα, η πρώτη εικόνα είναι το τμήμα εικόνας από το σετ δεδομένων ελέγχου, η δεύτερη εικόνα είναι το αντίστοιχο τμήμα εικόνας των επαληθευμένων δεδομένων (ground truth) και η τρίτη εικόνα είναι η πρόβλεψη του δικτύου.



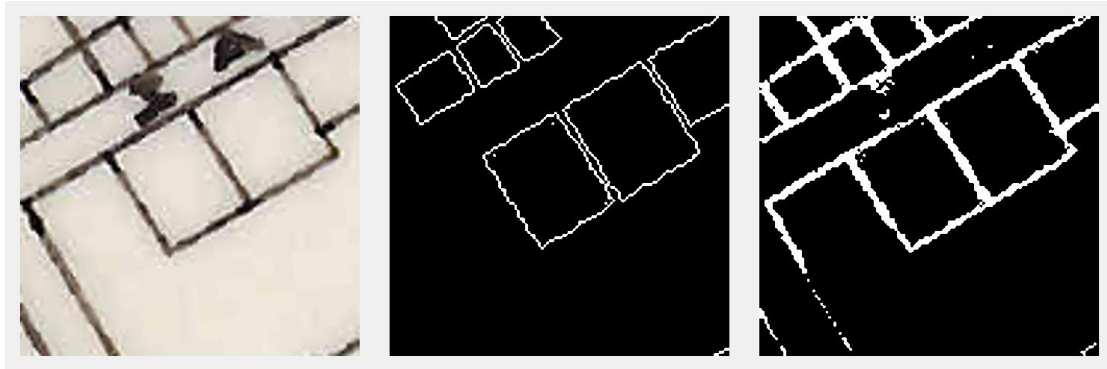
Εικόνα 4.39 Περιγράμματα, grid patches, size 64x64.



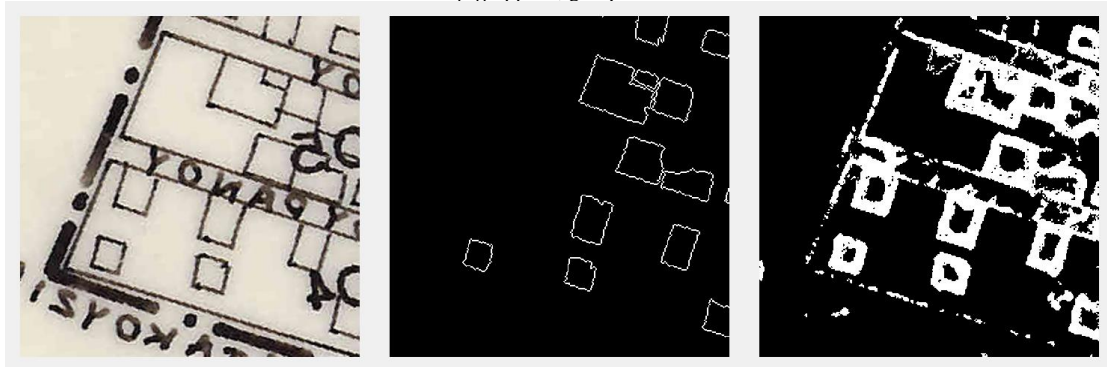
Εικόνα 4.40 Περιγράμματα, grid patches, size 64x64



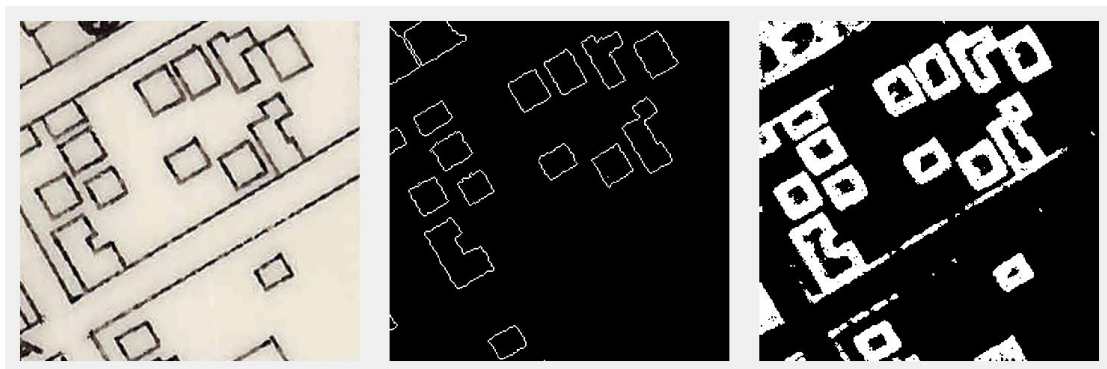
Εικόνα 4.41 Περιγράμματα, grid patches, size 128x128



Εικόνα 4.42 Περιγράμματα, grid patches, size 128x128



Εικόνα 4.43 Περιγράμματα, grid patches, size 256x256



Εικόνα 4.44 Περιγράμματα, grid patches, size 256x256

Από αυτά τα πειράματα παρατηρείται πως εκείνο που απέδωσε καλύτερα είναι εκείνο με patch size 64x64. Ενδιαφέρον έχει η Εικόνα 4.41 όπου το νευρωνικό δίκτυο που εκπαιδεύτηκε με 128x128 patch size, δεν έχει μεγάλη ακρίβεια στα περιγράμματα που βρήκε. Παρόλα αυτά, ο αριθμός 3 που υπάρχει στην εικόνα εισόδου και επικαλύπτει το κτίριο έχει απαλειφθεί επιτυχώς και το νευρωνικό δίκτυο μπόρεσε να εξάγει το αποτύπωμα του κτιρίου, παρά τα λάθη του.

Grid patches - Αποτυπώματα

Τα επόμενα παραδείγματα αφορούν την περίπτωση που το δίκτυο εκπαιδεύεται στο να βγάζει ολόκληρο το αποτύπωμα (footprint) του κτιρίου χρησιμοποιώντας grid patches.



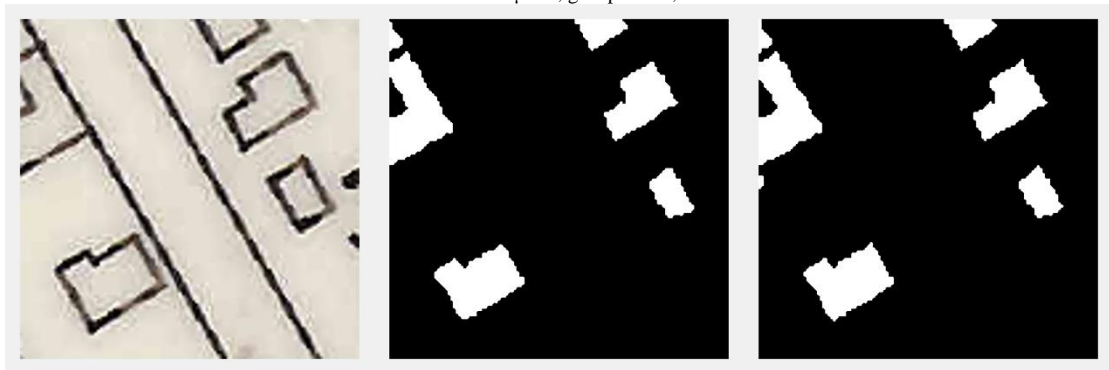
Εικόνα 4.45 Αποτοπώματα, grid patches, size 64x64



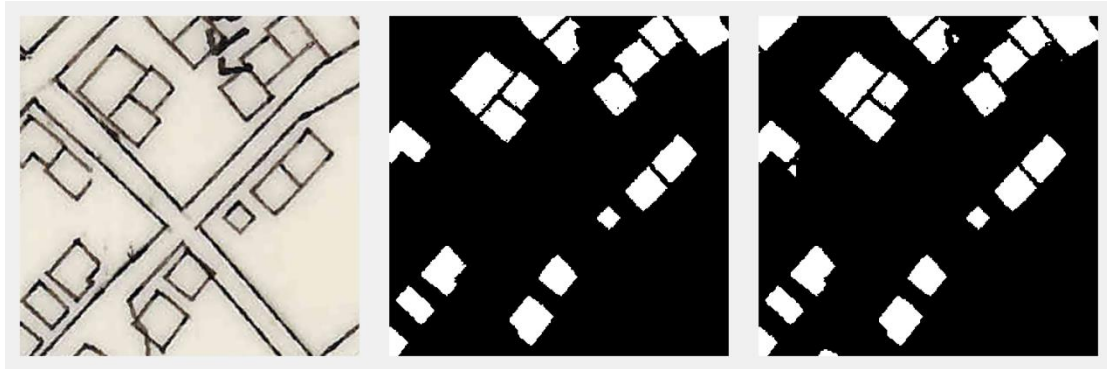
Εικόνα 4.46 Αποτοπώματα, grid patches, size 64x64



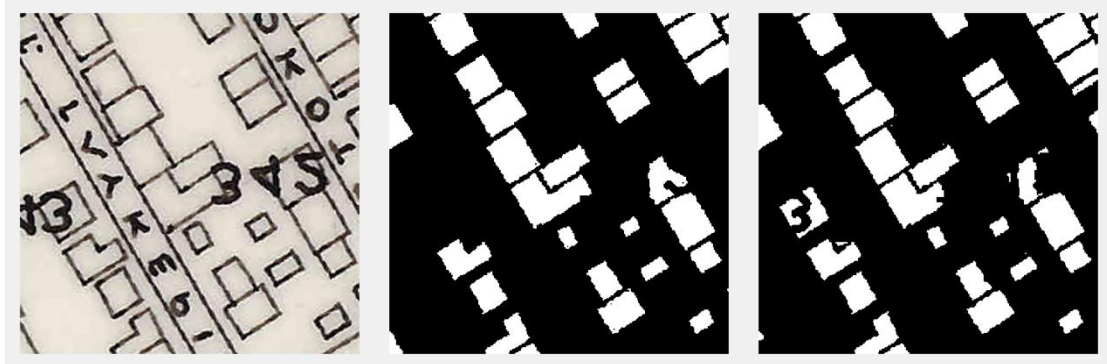
Εικόνα 4.47 Αποτοπώματα, grid patches, size 128x128



Εικόνα 4.48 Αποτοπώματα, grid patches, size 128x128



Εικόνα 4.49 Αποτυπώματα, grid patches, size 256x256

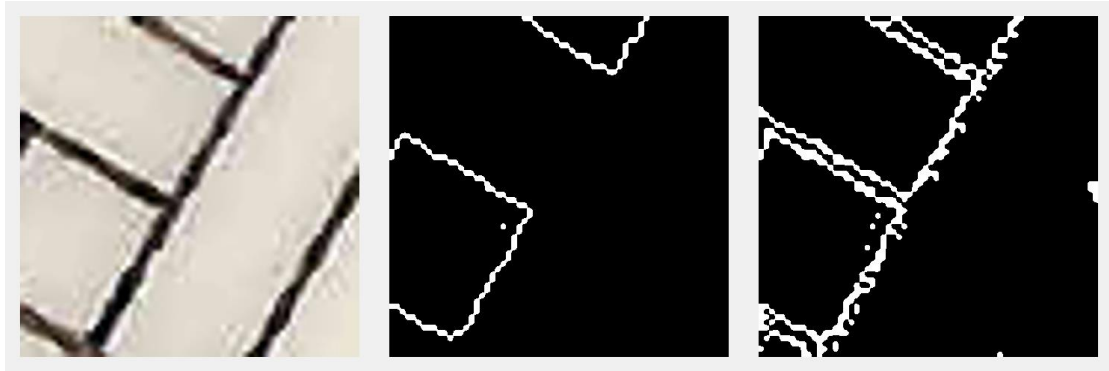


Εικόνα 4.50 Αποτυπώματα, grid patches, size 256x256

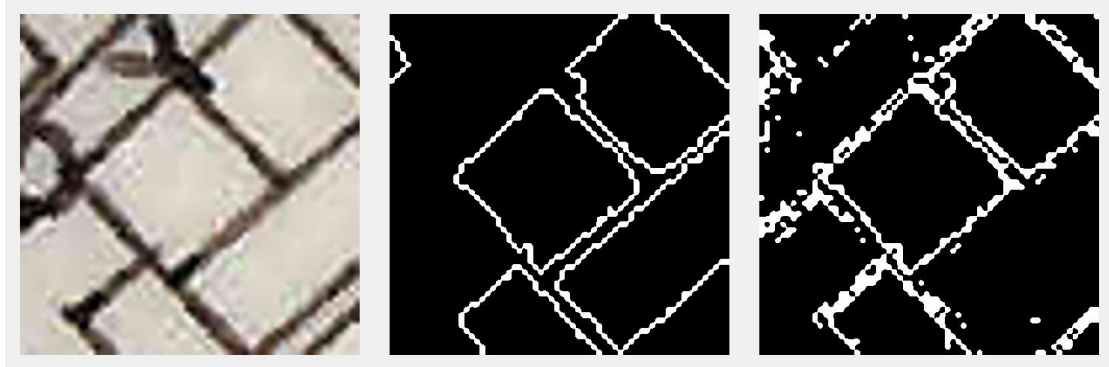
Το δίκτυο στα δεδομένα εισόδου με patch size 128x128 και 256x256 έχει πολύ καλύτερες επιδόσεις σε σύγκριση με τις υπό-εικόνες με patch size 64x64. Αυτό σημαίνει ότι βελτιώνεται η απόδοση τους όσο μεγαλώνει το μέγεθος της υπο-εικόνας. Ενδιαφέρουσα θα μπορούσε να χαρακτηριστεί η Εικόνα 4.46 όπου το δίκτυο έχει εξαγάγει ένα κτίριο πάνω δεξιά, παρόλο που δεν υπάρχει. Ακόμα, στην Εικόνα 4.50 το U-Net δίκτυο πρόβλεψε πολύ αποτελεσματικά, ένα κτίριο (κέντρο και αριστερά της δεξιάς εικόνας) το οποίο δεν υπάρχει καν στο ground truth (μεσαία εικόνα) και εκτός αυτού μπόρεσε να ξεχωρίσει και τον αριθμό ο οποίος επικαλύπτει το κτίριο. Συγκριτικά, ανάμεσα στην προσέγγιση με τα περιγράμματα των κτιρίων σε σύγκριση με τα αποτυπώματά τους, η πρώτη έχει καταλήξει σε ακριβέστερα αποτελέσματα στο μέγεθος 64x64 ενώ η δεύτερη στο 256x256.

Random patches – Περιγράμματα

Τα επόμενα παραδείγματα αφορούν την περίπτωση που το δίκτυο εκπαιδεύεται στο να εξαγάγει τα περιγράμματα των κτιρίων χρησιμοποιώντας random patches .



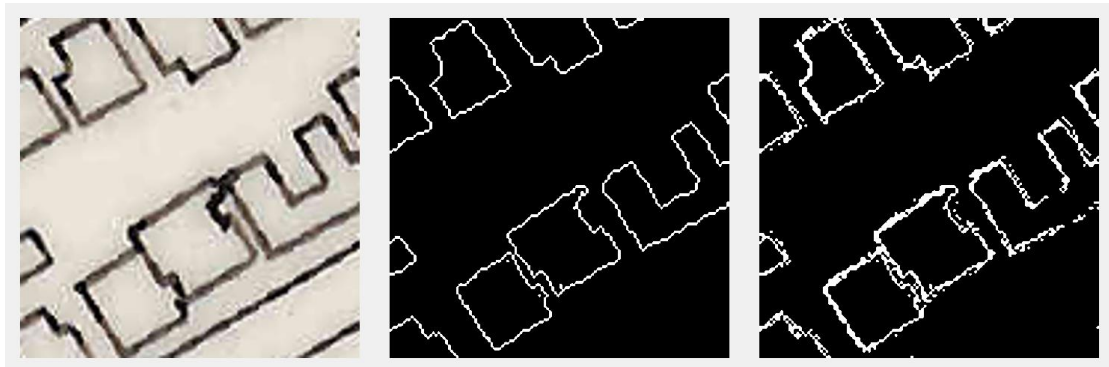
Εικόνα 4.51 Περιγράμματα, random patches, size 64x64



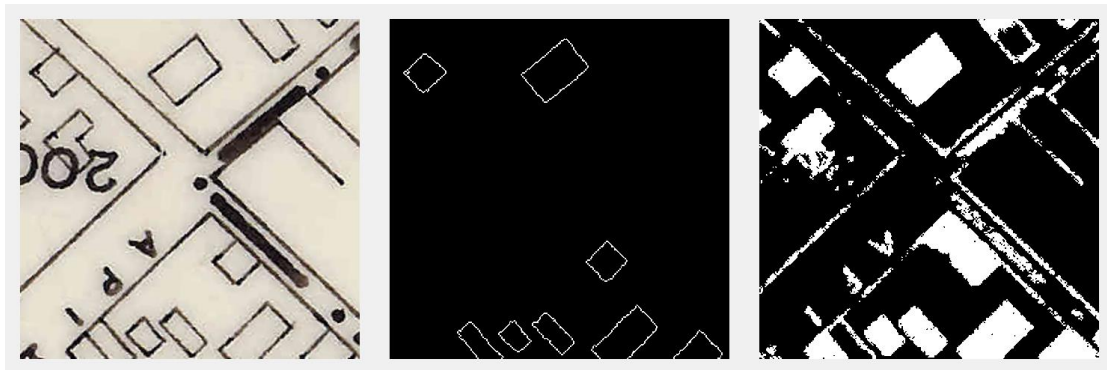
Εικόνα 4.52 Περιγράμματα, random patches, size 64x64



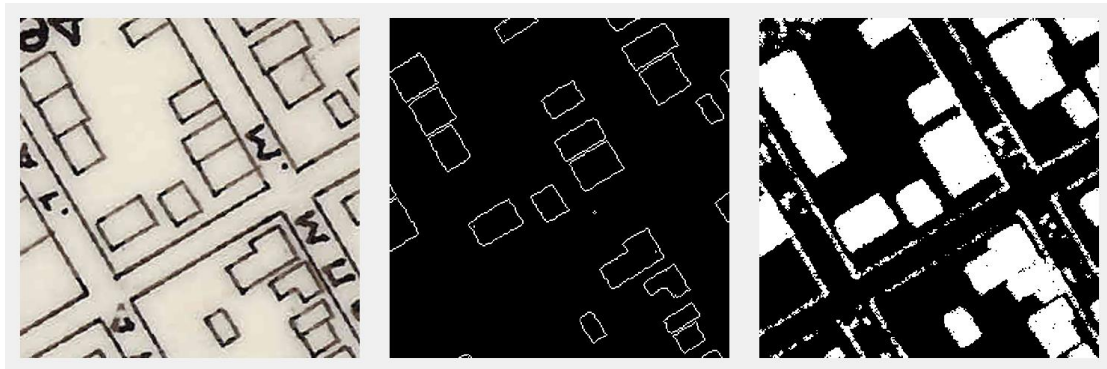
Εικόνα 4.53 Περιγράμματα, random patches, size 128x128



Εικόνα 4.54 Περιγράμματα, random patches, size 128x128



Εικόνα 4.55 Περιγράμματα, random patches, size 256x256

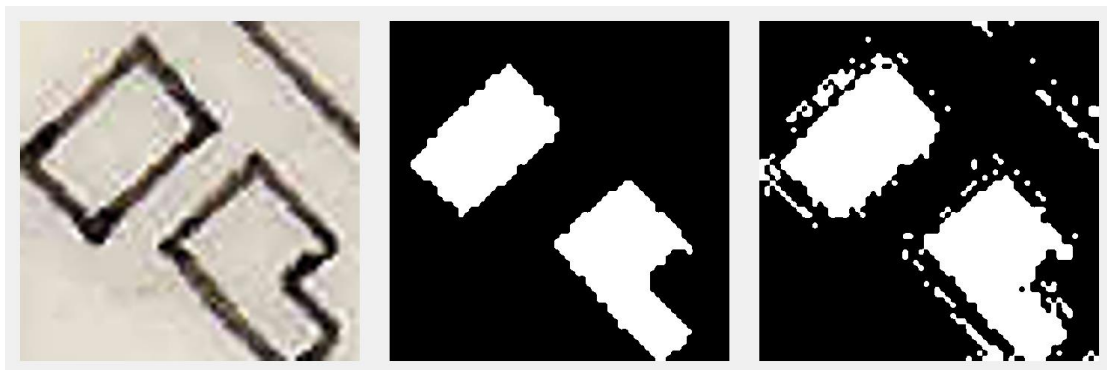


Εικόνα 4.56 Περιγράμματα, random patches, size 256x256

Παρατηρείται ότι με εκπαίδευση με τυχαία patches στα περιγράμματα των κτιρίων, υψηλότερη ακρίβεια επιτυγχάνεται για μέγεθος patch 128x128. Το νευρωνικό δίκτυο που εκπαιδεύτηκε με τα 256x256 patch size, δεν μπορεί να ξεχωρίσει πολύ ορθά τα οικοδομικά τετράγωνα.

Random patches – Αποτυπώματα

Τα επόμενα παραδείγματα αφορούν την περίπτωση που το δίκτυο εκπαιδεύεται στο να βγάξει ολόκληρο το αποτύπωμα του κτιρίου χρησιμοποιώντας random patches



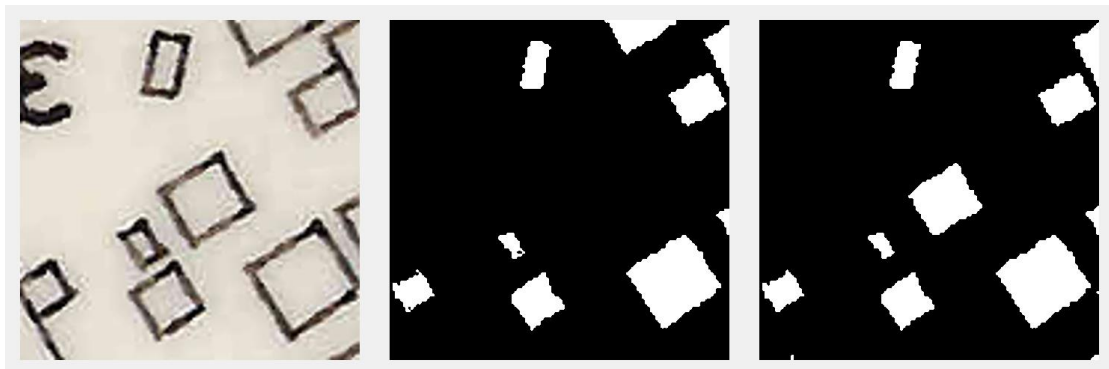
Εικόνα 4.57 Αποτυπώματα, random patches, size 64x64



Εικόνα 4.58 Αποτοπώματα, random patches, size 64x64



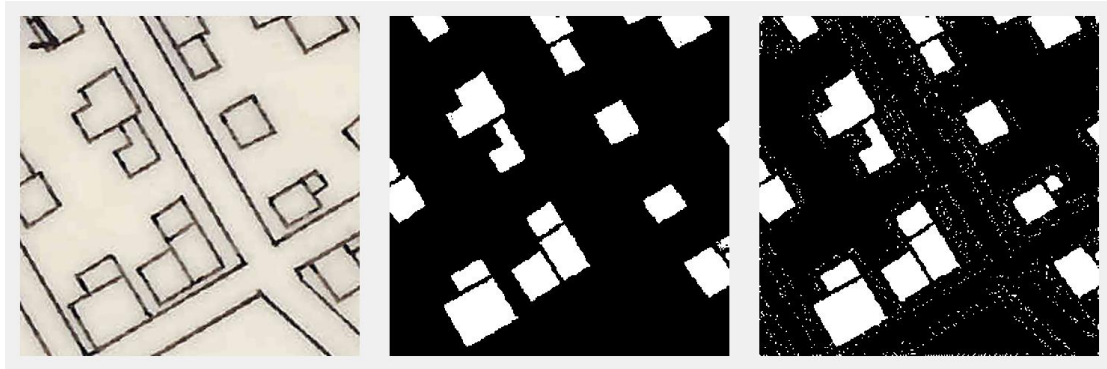
Εικόνα 4.59 Αποτοπώματα, random patches, size 128x128



Εικόνα 4.60 Αποτοπώματα, random patches, size 128x128



Εικόνα 4.61 Αποτοπώματα, random patches, size 256x256



Εικόνα 4.62 Αποτυπώματα, random patches, size 256x256

Στα δίκτυα που εκπαιδευτήκαν με τα αποτυπώματα, υψηλότερη ακρίβεια παρέχουν τα patch size 128x128 και 256x256. Αυτό συμβαίνει γιατί το νευρωνικό δίκτυο με patch size 64 δεν έχει εκπαιδευτεί με δείγματα που να καλύπτουν όλη την εικόνα χωρικά και το μέγεθος είναι πολύ μικρό ώστε να «χωρέσει» αρκετές πληροφορίες. Οι Πίνακες 3 και 5 παρέχουν τα αντίστοιχα αναλυτικά αποτελέσματα.

4.10 Αποτίμηση

Η έξοδος του νευρωνικού δικτύου σε όλα τα παραπάνω πειράματα είναι δυαδική μορφής. Αυτό σημαίνει ότι τα εικονοστοιχεία των προβλέψεων ανήκουν σε κάποια από τις δύο κλάσεις. Συγκεκριμένα, τα λευκά pixel θα λαμβάνονται υπόψιν ως το προσκήνιο δηλαδή τα περιγράμματα ή τα αποτυπώματα των σπιτιών, και τα μαύρα pixel ως το παρασκήνιο. Ο τρόπος που αξιολογήθηκε το κάθε δίκτυο των δώδεκα διαφορετικών σετ δεδομένων είναι βασιζόμενο σε συγκρίσεις επιπέδου pixel. Αναλυτικά γίνεται καταμέτρηση των pixel στην εκάστοτε εικόνα εισόδου και στην αντίστοιχη εικόνα που αποτελεί την έξοδο του δικτύου. Έτσι λοιπόν γίνεται καταμέτρηση των true positives (TP) ή αλλιώς αληθώς θετικών εικονοστοιχείων, τα οποία είναι όλα εκείνα στα οποία η πρόβλεψη και η αντίστοιχη αληθής τιμή (ground truth) είναι και οι δύο θετικές. Στην συνέχεια καταμετρώνται τα εσφαλμένα θετικά, false positives (FP), δηλαδή τα εικονοστοιχεία όπου η πρόβλεψη είναι θετική, αλλά η αντίστοιχη αληθής τιμή είναι αρνητική. Τέλος γίνονται άλλες δυο καταμετρήσεις στα εσφαλμένα αρνητικά, false negatives (FN), εικονοστοιχεία στα οποία πρόβλεψη είναι αρνητική αλλά η αντίστοιχη αληθής τιμή είναι θετική και στα true negatives (TN) στα οποία η πρόβλεψη και η αντίστοιχη αληθής τιμή είναι και οι δύο αρνητικές. Οι μετρικές που χρησιμοποιήθηκαν είναι:

Καθολική ακρίβεια (Global Accuracy): Είναι ο λόγος όλων των εικονοστοιχείων που προβλέφθηκαν ορθά, προς το συνολικό αριθμό εικονοστοιχείων, ανεξαρτήτως σε ποια κλάση ανήκουν.

$$Global Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad 4.10.1$$

Μέση ακρίβεια (Mean Accuracy): Ο λόγος των ορθών θετικών εικονοστοιχείων προς τον συνολικό αριθμό εικονοστοιχείων σε κάθε κλάση.

Μέση Τομή προς Ένωση (Mean Intersection over Union-Mean IoU): Το μέσο σκορ Τομή προς Ένωση όλων των κλάσεων σε όλες τις εικόνες. Το μέσο σκορ IoU (Mean IoU) υπολογίζεται ως

$$\text{Mean IoU} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad 4.10.2$$

Είναι δηλαδή ο λόγος των ορθών ταξινομημένων pixel προς το συνολικό αριθμό των Ground truth pixel προσθέτοντας αυτά που έχουν προβλεφθεί για την κάθε κλάση.

Σταθμισμένη Τομή προς Ένωση (Weighted Intersection over Union-Weighted IoU): Ο σταθμισμένος μέσος του σκορ Τομή προς Ένωση, όλων των τάξεων σε όλες τις εικόνες. Για κάθε τάξη ο σταθμισμένος μέσος Τομής προς Ένωση, είναι το μέσο σκορ Τομής προς Ένωση, σταθμισμένο βάσει του αριθμού pixel που ανήκουν στην αντίστοιχη κλάση.

Μέσο F1 Σκορ Περιγράμματος (Mean Boundary F1 Score): Το σκορ το οποίο υποδεικνύει κατά πόσο το περίγραμμα το οποίο προβλέφθηκε για κάθε κλάση συμφωνεί με το αντίστοιχο αληθές περίγραμμα [11].

Ακολουθούν οι μετρικές αξιολόγησης, για τα grid και random patches μεταξύ των περιγραμμάτων και των αποτυπωμάτων για όλα τα μεγέθη.

EVALUATION METRICS		Grid Patches					
		Περιγράμματα			Αποτυπώματα		
		64X64	128X128	256X256	64X64	128X128	256X256
Global Accuracy		99.01%	94.76%	81.50%	93.94%	96.66%	98.09%
Mean Accuracy		95.35%	93.80%	89.76%	91.69%	94.84%	97.63%
Mean IoU		84.66%	61.49%	45.40%	82.29%	88.36%	92.59%
Weighted IoU		98.25%	93.18%	79.69%	89.16%	93.79%	96.38%
F1 Score		86.36%	68.54%	33.85%	74.36%	85.85%	91.86%
Accuracy	Background	99.21%	94.81%	81.16%	95.14%	97.45%	98.26%
	Houses	91.48%	92.79%	98.36%	88.23%	92.23%	97.00%
IoU	Background	98.99%	94.65%	81.13%	92.84%	96.13%	97.80%
	Houses	70.34%	28.33%	9.67%	71.74%	80.60%	87.37%
F1 Score	Background	92.21%	71.76%	42.27%	83.29%	90.08%	94.48%
	Houses	80.42%	65.33%	25.44%	65.39%	81.59%	89.24%

Πίνακας 3 Μετρικές αξιολόγησης για grid patches

EVALUATION METRICS		Grid Patches					
		Περι- γράμματα	Αποτυ- πώματα	Περι- γράμματα	Αποτυ- πώματα	Περι- γράμματα	Αποτυ- πώματα
		64X64	64X64	128X128	128X128	256X256	256X256
Global Accuracy		99.01%	93.94%	94.76%	96.66%	81.50%	98.09%
Mean Accuracy		95.35%	91.69%	93.80%	94.84%	89.76%	97.63%
Mean IoU		84.66%	82.29%	61.49%	88.36%	45.40%	92.59%
Weighted IoU		98.25%	89.16%	93.18%	93.79%	79.69%	96.38%
F1 Score		86.36%	74.36%	68.54%	85.85%	33.85%	91.86%
Accuracy	Back-ground	99.21%	95.14%	94.81%	97.45%	81.16%	98.26%
	Houses	91.48%	88.23%	92.79%	92.23%	98.36%	97.00%
IoU	Back-ground	98.99%	92.84%	94.65%	96.13%	81.13%	97.80%
	Houses	70.34%	71.74%	28.33%	80.60%	9.67%	87.37%
F1 Score	Back-ground	92.21%	83.29%	71.76%	90.08%	42.27%	94.48%
	Houses	80.42%	65.39%	65.33%	81.59%	25.44%	89.24%

Πίνακας 4 Μετρικές αξιολόγησης για grid patches μεταξύ περιγραμμάτων και αποτυπωμάτων.

EVALUATION METRICS		Random Patches					
		Περιγράμματα			Αποτυπώματα		
		64X64	128X128	256X256	64X64	128X128	256X256
Global Accuracy		96.45%	98.27%	83.64%	91.31%	98.21%	97.05%
Mean Accuracy		93.01%	95.24%	91.35%	92.22%	95.83%	96.00%
Mean IoU		65.73%	72.04%	44.92%	73.67%	91.52%	85.71%
Weighted IoU		95.07%	97.42%	82.59%	86.07%	96.58%	94.73%
F1 Score		63.07%	85.16%	30.06%	42.35%	92.11%	53.78%
Accuracy	Background	96.60%	98.37%	83.46%	91.01%	98.88%	97.28%
	Houses	89.42%	92.10%	99.24%	93.42%	92.79%	94.72%
IoU	Background	96.38%	98.25%	83.45%	90.17%	98.01%	96.76%
	Houses	35.08%	45.82%	6.39%	57.18%	85.03%	74.66%
F1 Score	Background	77.02%	90.88%	40.99%	53.14%	96.12%	54.92%
	Houses	49.12%	79.44%	19.13%	31.56%	88.10%	52.63%

Πίνακας 5 Μετρικές αξιολόγησης για random patches

EVALUATION METRICS		Random Patches					
		Περιγράμματα		Foot-prints		Foot-prints	
		64X64	64X64	128X128	128X128	256X256	256X256
Global Accuracy		96.45%	91.31%	98.27%	98.21%	83.64%	97.05%
Mean Accuracy		93.01%	92.22%	95.24%	95.83%	91.35%	96.00%
Mean IoU		65.73%	73.67%	72.04%	91.52%	44.92%	85.71%
Weighted IoU		95.07%	86.07%	97.42%	96.58%	82.59%	94.73%
F1 Score		63.07%	42.35%	85.16%	92.11%	30.06%	53.78%
Accuracy	Back-ground	96.60%	91.01%	98.37%	98.88%	83.46%	97.28%
	Houses	89.42%	93.42%	92.10%	92.79%	99.24%	94.72%
IoU	Back-ground	96.38%	90.17%	98.25%	98.01%	83.45%	96.76%
	Houses	35.08%	57.18%	45.82%	85.03%	6.39%	74.66%
F1 Score	Back-ground	77.02%	53.14%	90.88%	96.12%	40.99%	54.92%
	Houses	49.12%	31.56%	79.44%	88.10%	19.13%	52.63%

Πίνακας 6 Μετρικές αξιολόγησης για random patches μεταξύ Περιγραμμάτων και Αποτυπώματα

EVALUATION METRICS		Περιγράμματα					
		Grid Patches			Random Patches		
		64X64	128X128	256X256	64X64	128X128	256X256
Global Accuracy		99.01%	94.76%	81.50%	96.45%	98.27%	83.64%
Mean Accuracy		95.35%	93.80%	89.76%	93.01%	95.24%	91.35%
Mean IoU		84.66%	61.49%	45.40%	65.73%	72.04%	44.92%
Weighted IoU		98.25%	93.18%	79.69%	95.07%	97.42%	82.59%
F1 Score		86.36%	68.54%	33.85%	63.07%	85.16%	30.06%
Accuracy	Background	99.21%	94.81%	81.16%	96.60%	98.37%	83.46%
	Houses	91.48%	92.79%	98.36%	89.42%	92.10%	99.24%
IoU	Background	98.99%	94.65%	81.13%	96.38%	98.25%	83.45%
	Houses	70.34%	28.33%	9.67%	35.08%	45.82%	6.39%
F1 Score	Background	92.21%	71.76%	42.27%	77.02%	90.88%	40.99%
	Houses	80.42%	65.33%	25.44%	49.12%	79.44%	19.13%

Πίνακας 7 Μετρικές αξιολόγησης για τα Περιγράμματα μεταξύ των grid και random patches

EVALUATION METRICS		Περιγράμματα					
		Grid Patches	Random Patches	Grid Patches	Random Patches	Grid Patches	Random Patches
		64X64	64X64	128X128	128X128	256X256	256X256
Global Accuracy		99.01%	96.45%	94.76%	98.27%	81.50%	83.64%
Mean Accuracy		95.35%	93.01%	93.80%	95.24%	89.76%	91.35%
Mean IoU		84.66%	65.73%	61.49%	72.04%	45.40%	44.92%
Weighted IoU		98.25%	95.07%	93.18%	97.42%	79.69%	82.59%
F1 Score		86.36%	63.07%	68.54%	85.16%	33.85%	30.06%
Accuracy	Back-ground	99.21%	96.60%	94.81%	98.37%	81.16%	83.46%
	Houses	91.48%	89.42%	92.79%	92.10%	98.36%	99.24%
IoU	Back-ground	98.99%	96.38%	94.65%	98.25%	81.13%	83.45%
	Houses	70.34%	35.08%	28.33%	45.82%	9.67%	6.39%
F1 Score	Back-ground	92.21%	77.02%	71.76%	90.88%	42.27%	40.99%
	Houses	80.42%	49.12%	65.33%	79.44%	25.44%	19.13%

Πίνακας 8 Μετρικές αξιολόγησης για τα Περιγράμματα μεταξύ των grid και random patches

EVALUATION METRICS		Grid Patches			Random Patches		
		Αποτυπώματα			Αποτυπώματα		
		64X64	128X128	256X256	64X64	128X128	256X256
Global Accuracy		93.94%	96.66%	98.09%	91.31%	98.21%	97.05%
Mean Accuracy		91.69%	94.84%	97.63%	92.22%	95.83%	96.00%
Mean IoU		82.29%	88.36%	92.59%	73.67%	91.52%	85.71%
Weighted IoU		89.16%	93.79%	96.38%	86.07%	96.58%	94.73%
F1 Score		74.36%	85.85%	91.86%	42.35%	92.11%	53.78%
Accuracy	Background	95.14%	97.45%	98.26%	91.01%	98.88%	97.28%
	Houses	88.23%	92.23%	97.00%	93.42%	92.79%	94.72%
IoU	Background	92.84%	96.13%	97.80%	90.17%	98.01%	96.76%
	Houses	71.74%	80.60%	87.37%	57.18%	85.03%	74.66%
F1 Score	Background	83.29%	90.08%	94.48%	53.14%	96.12%	54.92%
	Houses	65.39%	81.59%	89.24%	31.56%	88.10%	52.63%

Πίνακας 9 Μετρικές αξιολόγησης για τα αποτυπώματα μεταξύ των grid και random patches

EVALUATION METRICS		Αποτυπώματα					
		Grid Patches	Random Patches	Grid Patches	Random Patches	Grid Patches	Random Patches
		64X64	64X64	128X128	128X128	256X256	256X256
Global Accuracy		93.94%	91.31%	96.66%	98.21%	98.09%	97.05%
Mean Accuracy		91.69%	92.22%	94.84%	95.83%	97.63%	96.00%
Mean IoU		82.29%	73.67%	88.36%	91.52%	92.59%	85.71%
Weighted IoU		89.16%	86.07%	93.79%	96.58%	96.38%	94.73%
F1 Score		74.36%	42.35%	85.85%	92.11%	91.86%	53.78%
Accuracy	Back-ground	95.14%	91.01%	97.45%	98.88%	98.26%	97.28%
	Houses	88.23%	93.42%	92.23%	92.79%	97.00%	94.72%
IoU	Back-ground	92.84%	90.17%	96.13%	98.01%	97.80%	96.76%
	Houses	71.74%	57.18%	80.60%	85.03%	87.37%	74.66%
F1 Score	Back-ground	83.29%	53.14%	90.08%	96.12%	94.48%	54.92%
	Houses	65.39%	31.56%	81.59%	88.10%	89.24%	52.63%

Πίνακας 10 Μετρικές αξιολόγησης για τα αποτυπώματα μεταξύ των grid και random patches

Από τον Πίνακα 3 προκύπτουν οι εξής παρατηρήσεις: Το νευρωνικό δίκτυο που έχει εκπαιδευτεί με patch size 64x64 έχει πολύ καλύτερες επιδόσεις στα περιγράμματα με την μόνη απόκλιση στην μέση ακρίβεια των κτιρίων, να έχει καλύτερη επίδοση το νευρωνικό δίκτυο που εκπαιδεύτηκε με patch size 256x256. Εν αντιθέσει το δίκτυο που εκπαιδεύτηκε με τα αποτυπώματα με patch size 256x256 έχει την καλύτερη απόδοση από τα υπόλοιπα patch size. Από τον Πίνακα 4 προκύπτει ότι, το νευρωνικό δίκτυο που έχει εκπαιδευτεί με patch size 64x64 έχει πολύ καλύτερες επιδόσεις στα περιγράμματα. Τα αποτυπώματα έχουν καλύτερο δείκτη IoU των κτιρίων στα patch size 128x128. Αυτό σημαίνει ότι οι προβλέψεις αυτού του δικτύου έχουν την μεγαλύτερη επικάλυψη με τα αληθή δεδομένα. Αυτό και μόνο όμως δεν είναι αντιπροσωπευτικό διότι όλοι οι υπόλοιποι δείκτες παρουσιάζουν μεγαλύτερες τιμές στα περιγράμματα. Όσον αφορά τα μεγέθη 128x128 το δίκτυο που έχει πάει καλύτερα είναι εκείνο που έχει εκπαιδευτεί με τα αποτυπώματα. Η τιμή της ακρίβειας των σπιτιών είναι μεγαλύτερη όμως στα περιγράμματα με patch size 128x128.

Στον Πίνακα 5 φαίνονται οι μετρικές αξιολόγησης του δικτύου για τα περιγράμματα και τα αποτυπώματα στα random patches. Το δίκτυο που έχει καλύτερη απόδοση είναι εκείνο που έχει εκπαιδευτεί με τα περιγράμματα και patch size 128x128. Αντίστοιχα, και το δίκτυο που εκπαιδεύτηκε στα αποτυπώματα με patch size 128x128 έχει την καλύτερη απόδοση από τα υπόλοιπα patch sizes. Υπάρχει ξανά η τιμή της ακρίβειας των σπιτιών η οποία έχει την μεγαλύτερη τιμή στο patch size 256x256, αλλά όπως προαναφέρθηκε δεν είναι αντιπροσωπευτική μόνο μια τιμή. Σύμφωνα με τον Πίνακα 6 προκύπτει πως το νευρωνικό δίκτυο το οποίο έχει την καλύτερη απόδοση στα random patches, με patch size 64x64, είναι αυτό που εκπαιδεύτηκε με τα περιγράμματα. Αντίθετα, τα αποτυπώματα έχουν μεγαλύτερη απόδοση στους δείκτες του μέσου IoU, την ακρίβεια και το IoU των κτιρίων. Ακόμα, στα

patches με μέγεθος 128x128, τα αποτυπώματα έχουν γενικά καλύτερη επίδοση, αλλά τα περιγράμματα έχουν την υψηλότερη συνολική ακρίβεια.

Σε μεγέθη patch 256x256, τα αποτυπώματα επιδεικνύουν καλύτερη απόδοση. Ακόμα οι μόνες τιμές οι οποίες είναι μεγαλύτερες είναι το σταθμισμένο IoU, και το, IoU του παρασκηνίου. Στα patches με patch size 256x256 το δίκτυο που έχει πάει καλύτερα είναι αυτό με τα αποτυπώματα. Όσον αφορά ποιο patch size είναι πιο αποδοτικό, με κριτήριο την ποια patches χρησιμοποιήθηκαν random ή grid, αναλύονται στον Πίνακα 7 και τον Πίνακα 8. Στον Πίνακα 7 παρατηρείται πως το δίκτυο που είναι το πιο αποδοτικό είναι εκείνο που εκπαιδεύτηκε με τα περιγράμματα, με grid patches και μέγεθος 64x64. Αντίθετα, στα random patches αυτό με τις υψηλότερες αποδόσεις είναι εκείνο με patch size 128x128. Αναλυτικά, στον Πίνακα 8 φαίνεται πως το δίκτυο με patches 64x64 είναι πιο αποδοτικό στα grid patches, ενώ τα patch size 128x128 και 256x256 αποδίδουν καλύτερα στα random patches.

Όσον αφορά ποιο patch size είναι πιο αποδοτικό, και σε ποια προσέγγιση στα αποτυπώματα, αυτό φαίνεται στους Πίνακες 9 και 10. Στον Πίνακα 9 παρατηρείται πως το δίκτυο που είναι το πιο αποδοτικό είναι εκείνο που εκπαιδεύτηκε με grid patches και μέγεθος 256x256. Αντίθετα, στα random patches αυτό με τις υψηλότερες αποδόσεις είναι εκείνο με patch size 128x128. Στον Πίνακα 10 φαίνεται πως το δίκτυο με patches 64x64 είναι πιο αποδοτικό στα grid patches, ενώ τα patch size 128x128 και 256x256 αποδίδουν καλύτερα στα random patches. Τέλος, να επισημανθεί ότι, όπως φαίνεται στους παραπάνω πίνακες, στην περίπτωση των περιγραμμάτων των κτιρίων, το patch size 256x256 δίνει χαμηλά ποσοστά στις μετρικές IOU και F1 Score τόσο στην Grid όσο και στην Random προσέγγιση. Αυτό φαίνεται και στα παραδείγματα στις Εικόνες 77 και 78 (για την Grid προσέγγιση) και τις Εικόνες 89 και 90 (για την Random προσέγγιση). Στις περιπτώσεις αυτές διαπιστώνεται ότι το δίκτυο παράγει αποτελέσματα με μεγάλο θόρυβο ο οποίος έχει την μορφή είτε ενεργών περιφερειακών pixel που δεν ανήκουν σε κάποιο κτίριο είτε με την μορφή ενεργών pixel στο εσωτερικό των κτιρίων πράγμα που δεν είναι επιθυμητό όταν στόχος είναι ο προσδιορισμός των περιγραμμάτων μόνο.

5 Συμπεράσματα

Η παρούσα διπλωματική εργασία αφορά την ανάλυση και επεξεργασία ιστορικών χαρτών με χρήση τεχνικών βαθιάς μάθησης. Συγκεκριμένα, σκοπός ήταν ο αυτόματος προσδιορισμός των περιγραμμάτων ή των αποτυπωμάτων των κτιρίων σε ιστορικούς πολεοδομικούς χάρτες. Για τον σκοπό αυτό προτάθηκε ένα βαθύ νευρωνικό δίκτυο το οποίο εκπαιδεύτηκε με διάφορα δεδομένα εισόδου προκειμένου να εξετασθεί η αποτελεσματικότητά του στον εντοπισμό των κτιρίων. Η εκπαίδευση του νευρωνικού δικτύου βασίστηκε σε ένα σύνολο δεδομένων στα οποία είχαν προηγουμένως επισημανθεί τα κτίρια πάνω στην τοπογραφική εικόνα (ground truth). Η διαδικασία επισημάνσης σε εικόνες είναι μια ιδιαίτερα απαιτητική και χρονοβόρα διαδικασία και πραγματοποιείται συνήθως από κάποιον ειδικό. Για να διευκολυνθεί η διαδικασία επισημάνσης του ground truth των εικόνων αναφοράς, αναπτύχθηκε επιπλέον στα πλαίσια της παρούσας εργασίας ένα εργαλείο επισημάνσης (annotation tool) σε ψηφιακές εικόνες προσαρμοσμένο ειδικά για την επισημάνση κτιρίων σε τοπογραφικές εικόνες. Το annotation tool που υλοποιήθηκε διευκολύνει και επιταχύνει σημαντικά την διαδικασία επισημάνσης. Στη συνέχεια πραγματοποιήθηκε η εκπαίδευση και η αξιολόγηση ενός συνελκτικού νευρωνικού δικτύου αρχιτεκτονικής U-Net. Για τον σκοπό αυτό απαιτήθηκε να δημιουργηθεί ένα μεγάλο σύνολο δεδομένων εκπαίδευσης με την μορφή ζευγών από υπο-εικόνες (patches) της αρχικής εικόνας και της αντίστοιχης επιθυμητής εικόνας που περιέχει τα περιγράμματα ή τα αποτυπώματα των κτιρίων. Η αποτίμηση του δικτύου πραγματοποιήθηκε με 12 σετ δεδομένων. Αυτά διαφέρουν μεταξύ τους ως προς τον τρόπο δειγματοληψίας των υπο-εικόνων με πλεγματική και τυχαία μορφή και ως προς την επιθυμητή έξοδο του δικτύου που αφορούσε τα περιγράμματα όσο και το συνολικό αποτύπωμα των κτιρίων. Ο προσδιορισμός του περιγράμματος των κτιρίων μέσω του αποτυπώματός τους βασίστηκε σε μορφολογικές πράξεις.

Μέσω του annotation tool επισημάνθηκαν 4.231 περιγράμματα. Σε αυτά δεν συμπεριελήφθησαν τα «προβληματικά» περιγράμματα είτε λόγω του ότι το ίδιο το πολύγωνο δεν ήταν κλειστό, είτε γιατί ένα νούμερο ή γράμμα επικάλυπτε μεγάλο μέρος του. Αυτά τα πολύγωνα ήταν ελάχιστα σε σχέση με τον συνολικό αριθμό τους, γύρω στο 10%. Αναφορικά με την εκπαίδευση του νευρωνικού δικτύου με τα 12 σετ εκπαίδευσης η διαδικασία ήταν αρκετά χρονοβόρα και απαιτητική. Τα πειράματα αφορούσαν το μέγεθος του patch, την προσέγγιση που χρησιμοποιήθηκε (Grid ή Random) καθώς και ποια εικόνα εισόδου θα απέδιδε καλύτερα όσον αφορά την εκπαίδευση και την αξιολόγηση. Οι εικόνες εξόδου περιείχαν τα περιγράμματα ή/και το αποτύπωμα των κτιρίων.

Όσον αφορά τη πλεγματική προσέγγιση, το δίκτυο που τα πήγε καλύτερα ήταν εκείνο που εκπαιδεύτηκε με το αποτύπωμα των κτιρίων και patch size 256x256. Είναι αξιοσημείωτο πως οι προβλέψεις σε αυτό το CNN «μεγαλώνουν» το μέγεθος του περιγράμματος και το απεικονίζουν πιο έντονο από ότι είναι. Στην τυχαία προσέγγιση το δίκτυο που τα πήγε καλύτερα στην μετρική F1 Score είναι εκείνο με patch size 128x128 και αφορούσε τα αποτυπώματα των κτιρίων. Όμως, και το δίκτυο του ίδιο μεγέθους που εκπαιδεύτηκε με τα περιγράμματα είχε παρόμοια τιμή F1 Score του παρόλο που η καθολική του ακρίβεια (global accuracy) ήταν μεγαλύτερη.

Συγκριτικά , ανάμεσα στις μεθόδους Grid και Random, στα δίκτυα που εκπαιδευτήκαν με τα περιγράμματα, η πλεγματική μέθοδος απέδωσε καλύτερα στα Patch size 64x64 και 256x256, ενώ η random μέθοδος έδωσε τα καλύτερα αποτελέσματα για μέγεθος patch 128x128. Κατ' αντιστοιχία, στα δίκτυα που εκπαιδευτήκαν με τα αποτυπώματα των κτιρίων, και πάλι η πλεγματική μέθοδος είχε καλύτερα αποτελέσματα στα Patch size 64x64 και 256x256, ενώ η random για 128x128. Τέλος, να επισημανθεί ότι στην περίπτωση των περιγραμμάτων και για patch size 256x256 απεικονίζονται εσφαλμένα πολλά Pixel του παρασκήνιου ως προσκήνιο με αποτέλεσμα να δημιουργείται θόρυβος πράγμα που αποτυπώνεται και στις χαμηλές τιμές των μετρικών αποτίμησης.

5.1 Προτάσεις μελλοντικής βελτίωσης

Ένας περιορισμός στην πλήρη επισήμανση του τοπογραφικού χάρτη που χρησιμοποιήθηκε στην παρούσα εργασία αποτέλεσε το γεγονός πως υπήρχαν κτίρια στα οποία δεν μπορούσε να γίνει η επισήμανση με αποτελεσματικό τρόπο. Αυτό το ζήτημα δεν αντιμετωπίστηκε στην παρούσα εργασία και θα ήταν ωφέλιμο να εξερευνηθεί. Ενδιαφέρον θα είχε να κατασκευαζόταν ένα εργαλείο με το οποίο θα μπορούσε να επισημάνει χειρωνακτικά ο χρήστης τα προβληματικά σημεία όπως τα ανοιχτά πολύγωνα και τα επικαλυπτόμενα γράμματα και νούμερα. Έτσι, θα εμπλουτιζόταν το σετ δεδομένων και με αυτές τις ιδιαίτερες περιπτώσεις κτιρίων. Επιπρόσθετα θα μπορούσε να περιλαμβάνει και την δυνατότητα διαφορετικών κατηγοριών επισήμανσης που θα ορίζει ο χρήστης. Με άλλα λόγια να υπάρχει παραπάνω από μια ετικέτα και να ορίζει ο χρήστης το όνομα της και σε ποια θα γίνει εκείνη την στιγμή η επισήμανση. Με αυτό τον τρόπο θα μπορούσε να επισημανθούν όχι μόνο κτίρια αλλά κι άλλες πληροφορίες πάνω στην εικόνα, π.χ. λεκτικές περιγραφές, ειδικά γεωμετρικά χαρακτηριστικά του χάρτη κλπ. Τέλος, θα μπορούσε αυτό το εργαλείο να δοκιμαστεί και σε άλλου είδους προβλήματα και εικόνες. Αναλυτικότερα, μπορεί να δοκιμαστεί και σε άλλες εικόνες τοπογραφικών και μη χαρτών ή ακόμα και σε άλλους επιστημονικούς κλάδους όπως στη επεξεργασία ιατρικών εικόνων.

Αναφορικά με την εκπαίδευση και την αξιολόγηση του νευρωνικού δικτύου, το U-Net είναι ένα πολύ αποτελεσματικό CNN για την σημασιολογική κατάτμηση. Πάρα ταύτα υπάρχουν περιθώρια βελτίωσης. Αρχικά, θα μπορούσαν να γίνουν περαιτέρω εκπαιδεύσεις με διαφορετικές παραμέτρους ώστε να γίνει μια αξιολόγηση σχετικά με το κατά πόσο επηρεάζει η κάθε παράμετρος την διαδικασία εκπαίδευσης. Τέλος, θα μπορούσε να δοκιμαστεί το πρόβλημα και σε άλλα δίκτυα βαθιάς μάθησης τα οποία να παρέχουν είτε βελτίωση στην ταχύτητα εκπαίδευσης ή/και βελτίωση στην αποτελεσματικότητά τους.

Βιβλιογραφία

- [1] Russel, S. και Norvig, P. 2021. Τεχνητή Νοημοσύνη Μια σύγχρονη προσέγγιση. Αθήνα. Κλειδάριθμος
- [2] McCulloch, Warren S., and Walter Pitts. “A Logical Calculus of the Ideas Immanent in Nervous Activity.” *The Bulletin of Mathematical Biophysics* 5, no. 4 (1943): 115–33. <https://doi.org/10.1007/bf02478259>.
- [3] ResearchGate. “Figure 3.4: Frank Rosenblatt with His Mark I Perceptron(Left), and A...” ResearchGate, 2020. https://www.researchgate.net/figure/Frank-Rosenblatt-with-his-Mark-I-perceptronleft-and-a-graphical-representation-of_fig2_345813508.
- [4] Sharma, Anmol. “An Introduction to Supervised Learning.” InsideAIML. InsideAIML. August 9, 2021. <https://insideaiml.com/blog/an-introduction-to-supervised-learning-1188>.
- [5] “Unsupervised Machine Learning - Javatpoint.” www.javatpoint.com. 2021. <https://www.javatpoint.com/unsupervised-machine-learning>.
- [6] Wikipedia Contributors. “Reinforcement Learning.” Wikipedia. Wikimedia Foundation. June 23, 2023. https://en.wikipedia.org/wiki/Reinforcement_learning.
- [7] “9. ΝΕΥΡΙΚΟ ΣΥΣΤΗΜΑ.” Ebooks.edu.gr. 2023. http://ebooks.edu.gr/ebooks/v/html/8547/2666/Biologia_A-Lykeiou_html-empl/index9.html.
- [8] Albawi, Saad, Tareq Abed Mohammed, and Saad Al-Zawi. “Understanding of a Convolutional Neural Network.” *International Conference on Engineering and Technology (ICET)*, 2017. <https://doi.org/10.1109/icengtechnol.2017.8308186>.
- [9] Mishra, Mayank. “Convolutional Neural Networks, Explained - towards Data Science.” Medium. Towards Data Science. August 26, 2020. <https://towardsdatascience.com/convolutional-neural-networks-explained-9cc5188c4939>.
- [10] Malhotra, Priyanka, Sheifali Gupta, Deepika Koundal, Atef Zaguia, and Wegayehu Enbeyle. “Deep Neural Networks for Medical Image Segmentation.” *Journal of Healthcare Engineering* 2022 (2022): 1–15. <https://doi.org/10.1155/2022/9580991>.
- [11] Ξύδας, Χ. “Χρήση Τεχνικών Βαθιάς Μάθησης για την Αναγνώριση Γεωγραφικών χαρακτηριστικών σε Ιστορικούς Χάρτες”, Διπλωματική Εργασία, Πανεπιστήμιο Δυτικής Αττικής (2021).
- [12] Abhishek Kumar Pandey, “Convolution, Padding, Stride, and Pooling in CNN - Analytics Vidhya - Medium.” Medium. Analytics Vidhya. June 25, 2020. <https://medium.com/analytics-vidhya/convolution-padding-stride-and-pooling-in-cnn-13dc1f3ada26>.
- [13] “Convolutional Neural Networks (CNN).” Github.io. 2018. https://humboldt-wi.github.io/blog/research/information_systems_1718/02convolutionalneuralnetworks/
- [14] “Down Sampling Using Pooling .” ResearchGate. ResearchGate. 2020. https://www.researchgate.net/figure/Down-sampling-using-pooling-5_fig1_334208543.
- [15] Unzueta, Diego. 2021. “Convolutional Layers vs Fully Connected Layers - towards Data Science.” Medium. Towards Data Science., November 13, 2021. <https://towardsdatascience.com/convolutional-layers-vs-fully-connected-layers-364f05ab460b>.

- [16] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. “U-Net: Convolutional Networks for Biomedical Image Segmentation.” *Lecture Notes in Computer Science*, 2015, 234–41, (2015) https://doi.org/10.1007/978-3-319-24574-4_28.
- [17] Kesidis, Anastasios, Kleomenis Kalogeropoulos, and Andreas Tsatsaris. Challenges of object recognition in historical topographic maps., 4th BAB International Conference, 2019.
- [18] “Επεξεργασία Και Ανάλυση Εικόνας: Βασικές Αρχές.” http://archive.eclass.uth.gr/eclass/modules/document/file.php/DIB229/image_proc_intro.pdf.
- [19] Εισαγωγή στην Επεξεργασία Εικόνας, Παραδείγματα με χρήση Matlab, Αθανασία Κολοβού, Αθήνα 2012
- [20] “How It Works.” 2023. The Camera. 2023. <http://opticalphenomenalcamera.weebly.com/how-it-works.html>.
- [21] Nosal, Eva-Marie. “Flood-Fill Algorithms Used for Passive Acoustic Detection and Tracking.” *New Trends for Environmental Monitoring Using Passive Systems*, 2008. <https://doi.org/10.1109/passive.2008.4786975>.
- [22] Kesidis, Anastasios L., Vassilios Krassanakis, Loukas-Moysis Misthos, and Nikolaos Merlemis. “Patchit: A Multipurpose Patch Creation Tool for Image Processing Applications.” *Multimodal Technologies and Interaction* 6, no. 12 (2022): 111. <https://doi.org/10.3390/mti6120111>.
- [23] Alomar, Khaled, Halil Ibrahim Aysel, and Xiaohao Cai. “Data Augmentation in Classification and Segmentation: A Survey and New Strategies.” *Journal of Imaging* 9, no. 2 (2023): 46. <https://doi.org/10.3390/jimaging9020046>.
- [24] Zhang, Chuanhai, Kurt Loken, Zhiyu Chen, Zhiyong Xiao and Gary Kunkel. “Mask Editor : an Image Annotation Tool for Image Segmentation Tasks.” (2018) ArXiv abs/1809.06461 (2018)
- [25] Daturks API. Accessed May 24, 2023. <https://docs.daturks.com/>.
- [26] “Object Annotation Platform for Computer Vision: Oclavi.” Object Annotation Platform for Computer Vision | AI Data Annotation Tool. Accessed May 24, 2023. <https://oclavi.com/>.
- [27] “The Leading AI Platform for Building Intelligent Applications.” Labelbox. Accessed July 24, 2023. <https://www.labelbox.com/>.
- [28] Tzotalin. “LabelImg”. Git code (2015). Available: <https://github.com/tzotalin/labelImg>
- [29] “Memorize Anything.” Pixorize. Accessed July 24, 2023. <https://pixorize.com/>.
- [30] Russell, Bryan C., Antonio Torralba, Kevin P. Murphy, and William T. Freeman. “LabelMe: A Database and Web-Based Tool for Image Annotation.” *International Journal of Computer Vision* 77, no. 1–3 (2007): 157–73. <https://doi.org/10.1007/s11263-007-0090-8>.
- [31] “JavaScript Image Annotation Library.” Annotorious. Accessed July 24, 2023. <https://annotorious.github.io/>.
- [32] Iakovidis, D. K., T. Goudas, C. Smailis, and I. Maglogiannis. “Ratsnake: A Versatile Image Annotation Tool with Application to Computer-Aided Diagnosis.” *The Scientific World Journal* 2014 (2014): 1–12. <https://doi.org/10.1155/2014/286856>.
- [33] Fiedler, Niklas, Marc Bestmann, and Norman Hendrich. “ImageTagger: An Open Source Online Platform for Collaborative Image Labeling.” *RoboCup 2018: Robot World Cup XXII*, 2019, 162–69. https://doi.org/10.1007/978-3-030-27544-0_13.
- [34] “VGG Image Annotator (Via).” Visual Geometry Group - University of Oxford. Accessed July 24, 2023. <http://www.robots.ox.ac.uk/~vgg/software/via>.
- [35] Kingma, Diederik P, and Jimmy Ba. “Adam: A Method for Stochastic Optimization.” arXiv.org, 2014. <https://arxiv.org/abs/1412.6980>.