



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΑΓΡΟΝΟΜΩΝ & ΤΟΠΟΓΡΑΦΩΝ ΜΗΧΑΝΙΚΩΝ –
ΜΗΧΑΝΙΚΩΝ ΓΕΩΠΛΗΡΟΦΟΡΙΚΗΣ

ΕΡΓΑΣΤΗΡΙΟ ΦΩΤΟΓΡΑΜΜΕΤΡΙΑΣ

Αυτοματοποιημένη Διανυσματοποίηση από Εικόνες με Τεχνικές Μηχανικής Μάθησης

Διπλωματική Εργασία

Ταλαβέρος Γιώργος

Επιβλέποντες:

- Ανδρέας Γεωργόπουλος (Καθηγητής ΕΜΠ)
- Dr. Π. Αγραφιώτης (ΑΤΜ-ΜΓ ΕΜΠ, ΔΠΜΣ Γεωπληροφορική)

Αθήνα Οκτώβρης 2022

στην κοινωνική μου ζωή, την
μουσική και όλη τη φοιτητική
ζωή που έχασα γι' αυτή τη
σχολή

Περιεχόμενα

1. ΕΙΣΑΓΩΓΗ	6
2. ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΑΝΑΣΚΟΠΗΣΗ	7
2.1 ΑΛΓΟΡΙΘΜΟΙ ΑΝΙΧΝΕΥΣΗΣ ΑΚΜΩΝ	8
2.2 ΕΙΣΑΓΩΓΗ ΤΗΣ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ ΣΤΗΝ ΑΝΙΧΝΕΥΣΗ ΑΚΜΩΝ	10
3. ΜΕΘΟΔΟΛΟΓΙΑ	12
3.1 ΑΠΟ ΟΡΘΟΦΩΤΟΓΡΑΦΙΑ ΣΕ ΕΙΚΟΝΑ ΑΚΜΩΝ	12
3.1.1 Λειτουργία της AI στην rix2rix	14
3.1.2 Βελτιστοποίηση των παραμέτρων του μοντέλου	17
3.1.3 Συμπεριφορά με διαφορετικά δεδομένα	18
3.2 ΑΠΟ ΕΙΚΟΝΑ ΣΕ ΔΙΑΝΥΣΜΑ	19
3.2.1 Φίλτρα	19
3.2.2 Διανυσματοποίηση	21
4. ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΔΙΚΤΥΟΥ	22
5. ΑΞΙΟΛΟΓΗΣΗ	26
5.1 ΕΡΩΤΗΜΑΤΟΛΟΓΙΟ	26
5.2 ΑΞΙΟΛΟΓΗΣΗ ΑΠΟ ΕΙΔΙΚΟ	28
6. ΣΥΜΠΕΡΑΣΜΑΤΑ	30
7. ΒΙΒΛΙΟΓΡΑΦΙΑ	32
ΠΑΡΑΡΤΗΜΑ Α	35
ΠΙΝΑΚΑΣ ΠΙΝΑΚΩΝ	35
ΠΙΝΑΚΑΣ ΣΧΗΜΑΤΩΝ	35
ΠΙΝΑΚΑΣ ΕΙΚΟΝΩΝ	35
ΟΡΟΛΟΓΙΕΣ	36

Περίληψη

Στην εποχή που ζούμε η επιστήμη της τεχνητής νοημοσύνης έχει αρχίσει και εμφανίζεται παντού. Ήδη βρίσκεται σε αυτοκίνητα, βιομηχανίες, οικιακές συσκευές αλλά και στο κλάδο των μηχανικών, σε συστήματα αποφάσεων, αναγνώρισης αντικειμένων και αυτοματισμού. Ωστόσο, στον εντοπισμό ακμών και στη ψηφιοποίηση ορθοφωτογραφιών, έχει επικρατήσει μια μεθοδολογία, που απαιτεί αρκετό χρόνο και σταθερή επίβλεψη από ένα ειδικευμένο επαγγελματία. Η συγκεκριμένη διπλωματική εργασία, προτείνει την αυτοματοποίηση της ψηφιοποίησης ακμών από ορθοφωτογραφίες λιθόχτιστων κτισμάτων, με την χρήση της τεχνητής νοημοσύνης και ήδη υπαρχόντων εργαλείων σε ανοιχτές βιβλιοθήκες κώδικα προγραμματισμού. Αρχικά, ανασκοπείται η βιβλιογραφία πάνω σε όλες τις μεθόδους εντοπισμών ακμών που έχουν προταθεί μέχρι και σήμερα. Ακολουθεί μια ανάλυση της μεθοδολογίας που εφαρμόζεται και το πως αυτή χωρίζεται στο κομμάτι του εντοπισμού των ακμών, (από ένα νευρωνικό δίκτυο) και στο κομμάτι της ψηφιοποίησης των εντοπισμένων ακμών. Γίνεται αναφορά στο νευρωνικό δίκτυο που χρησιμοποιήθηκε, στα χαρακτηριστικά του και μικρή περιγραφή της λειτουργίας του. Τονίζεται το στάδιο της βελτιστοποίησης του νευρωνικού δικτύου που χρησιμοποιήθηκε και αιτιολογούνται οι αποφάσεις που πάρθηκαν. Παρουσιάζεται μια γενική αξιολόγηση του αποτελέσματος της εργασίας αυτής από το ευρύ κοινό με την χρήση ενός ερωτηματολόγιου και από ειδικευμένο επαγγελματία. Τέλος παρουσιάζονται τα τελικά συμπεράσματα, οι αστοχίες της εργασίας, οι παραδοχές που έγιναν κατά την διάρκεια αυτής και οι βελτιώσεις που μπορούν να γίνουν σε μελλοντικές παρόμοιες εργασίες.

National Technical university of Athens, Greece
School of Rural, Surveying and Geoinformatics Engineering

Automated Vectorization of Images using Machine Learning Techniques

Talaveros George

Diploma Thesis

October 2022

Abstract

In the era we live in, the science of artificial intelligence can be found everywhere. It is being used in car systems, industries, inside our houses and in the engineering industry in decision making, object detection and automation systems. However, in edge detection and digitization of orthophotos, a methodology has prevailed that requires a considerable amount of time and constant supervision by a qualified professional. This diploma thesis proposes the automation of edge digitization from orthophotos of stone-built buildings, with the use of artificial intelligence and already existing tools in open-source libraries. Firstly, the literature on all edge detection methods that have been proposed up to this day is being reviewed. Then, follows an analysis of the methodology that is being followed and how it is divided into the part of locating the edges (by a neural network) and the part of digitizing the already located edges. After that, there is a reference made to the neural network that is being used, its characteristics and a short description of its function. The stage of optimization of the neural network and the reasoning behind the decisions made, is being highlighted. Then, an evaluation of the results of this work by the public (with the use of a questionnaire) and by a qualified professional is presented. Finally, the thesis closes with the final conclusions, the problems of this thesis, the assumptions made during its execution and improvements that can be made in future projects similar to this one.

1. Εισαγωγή

Σε σύγχρονη εποχή, αναπτύσσεται και της δίνεται βάρος ως το επόμενο βήμα της ανθρώπινης προόδου η επιστήμη της Τεχνητής Νοημοσύνης^[01] (AI). Σε αυτή την εργασία παρουσιάζεται πώς με τα ήδη υπάρχοντα και αναπτυσσόμενα εργαλεία αυτής της επιστήμης, γίνεται να διευκολυνθεί η διαδικασία ψηφιοποίησης ορθοφωτογραφιών λιθόχτιστων κατασκευών. Παρά την τεχνολογική ανάπτυξη, αυτή η διαδικασία παραμένει ακόμα χειροκίνητη και απαιτεί την χρήση ενός εξειδικευμένου επαγγελματία.

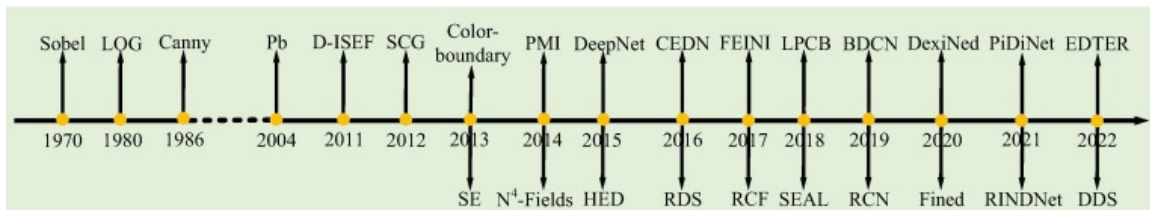
Συγκεκριμένα, μετά την δημιουργία της ορθοφωτογραφίας, ακολουθεί η διαδικασία της ψηφιοποίησης, για να παραχθεί ένα διανυσματικό (αρχιτεκτονικό) σχέδιο. Η ορθοφωτογραφία δίνεται σε έναν ειδικό (συνήθως αρχιτέκτονα) και ζητείται η δημιουργία ενός διανυσματικού αρχείου όπου καταγράφονται όλες οι ακμές των λίθων. Αυτή είναι μια χρονοβόρα διαδικασία μιας και απαιτεί αρκετές ώρες ενασχόλησης του μηχανικού και κατ' επέκταση κοστοβόρα. Με σκοπό την οικονομικότερη και συντομότερη περάτωση αυτής της διαδικασίας, εξετάζεται η χρήση της τεχνολογίας AI και των βιβλιοθηκών ψηφιακής επεξεργασίας εικόνων, ως εργαλεία ψηφιοποίησης που εντοπίζουν, σχεδιάζουν και αποθηκεύουν τις ακμές λίθων, σε επεξεργάσιμη μορφή.

Η εργασία αποτελείται από 5 κεφάλαια. Στο 2^ο κεφάλαιο παρουσιάζεται η τωρινή κατάσταση των μηχανισμών εντοπισμού ακμών μέσα από επιστημονικά άρθρα και δημοσιεύματα. Στο 3^ο κεφάλαιο αναπτύσσεται η μεθοδολογία που ακολουθήθηκε για την επίτευξη του τελικού στόχου. Η μεθοδολογία αποτελείται από δύο στάδια μετατροπών α) από ορθοφωτογραφία σε δυαδική (binary) εικόνα με εντοπισμένες τις ακμές και β) από εικόνα σε διανυσματική μορφή. Αντίστοιχα σε κάθε στάδιο παρουσιάζονται προβλήματα και εξηγούνται τα εργαλεία που χρησιμοποιήθηκαν. Στο 4^ο κεφάλαιο δίνεται έμφαση στο στάδιο πειραματισμού κατά την βελτιστοποίηση του νευρωνικού δικτύου και αιτιολόγηση των βασικών αποφάσεων για τις παραμέτρους που επιλέχθηκαν. Στο 5^ο κεφάλαιο αξιολογούνται τα αποτελέσματα από το κοινό με την χρήση ερωτηματολογίου και από έναν ειδικό με εμπειρία στην ψηφιοποίηση. Τέλος στο 6^ο κεφάλαιο παρουσιάζεται το συμπέρασμα της εργασίας, οι ιδέες και σκέψεις για την χρήση αυτής της μεθόδου, όπως και οι ενδεχόμενες αδυναμίες της.

2. Βιβλιογραφική Ανασκόπηση

Ο εντοπισμός ακμών είναι ένα από τα σημαντικότερα ζητήματα που συναντάται στη επεξεργασία εικόνων και της όρασης υπολογιστών^[92] (computer vision). Αυτό οφείλεται στο ότι η οριοθέτηση της ακμής στα απεικονιζόμενα αντικείμενα μιας εικόνας, είναι κύριος παράγοντας σε συστήματα εντοπισμού, οριοθέτησης, αναγνώρισης προσώπων και αντικειμένων. Συστήματα που συναντιούνται σε προγράμματα αναγνώρισης εκφράσεων (Mohan et al., 2021), στην ιατρική (με ενίσχυση ακμών αντικειμένων σε ακτινογραφίες, μαγνητικές που μαρτυρούν κάποια ασθένεια (Qiu et al., 2021)), σε συστήματα εντοπισμού αντικειμένων (Chi et al., 2013) και σε συστήματα παρακολούθησης καλλιερχειών (Wang et al., 2020).

Οι μηχανισμοί εντοπισμού ακμών χωρίζονται σε δύο βασικές κατηγορίες. Στους αλγορίθμους κατασκευασμένους πλήρως από τον άνθρωπο και στις μεθόδους με συστήματα τεχνητής μάθησης. Σε συνάρτηση με το χρόνο, πρώτα εμφανίστηκαν τα αλγοριθμικά συστήματα (Sobel, LOG, Canny) και μετά το 2010 με την κατασκευή δυνατότερων υπολογιστικών μηχανών, δόθηκε η δυνατότητα χρήσης της τεχνητής νοημοσύνης. Αυτή, ξεκίνησε από απλό υποστηρικτικό εργαλείο σε αλγορίθμους και κατέληξε σε αυτόνομη μέθοδο εντοπισμού ακμών.



Εικόνα 01: Χρονοδιάγραμμα εμφάνισης μεθόδων εντοπισμού ακμών (Junfen et al., 2022)

Όπως προαναφέρθηκε, από το 1970 που εμφανίστηκε ο πρώτος μηχανισμός εντοπισμού ακμών, υπάρχει μια συνεχής προσπάθεια εύρεσης μεθόδων και βελτιστοποίησης των ήδη υφιστάμενων, έτσι ώστε να έχουν γενική εφαρμογή με όσο το δυνατόν καλύτερα αποτελέσματα. Τον στόχο αυτό δεν έχει καταφέρει κάποια μέθοδος να τον ικανοποιήσει πλήρως μέχρι στιγμής. Ο λόγος είναι ότι οι απαιτήσεις ενός ιδανικού συστήματος εντοπισμού ακμών είναι απλές στη κατανόηση, αλλά δύσκολες στην επίτευξη. Οι βασικές απαιτήσεις είναι:

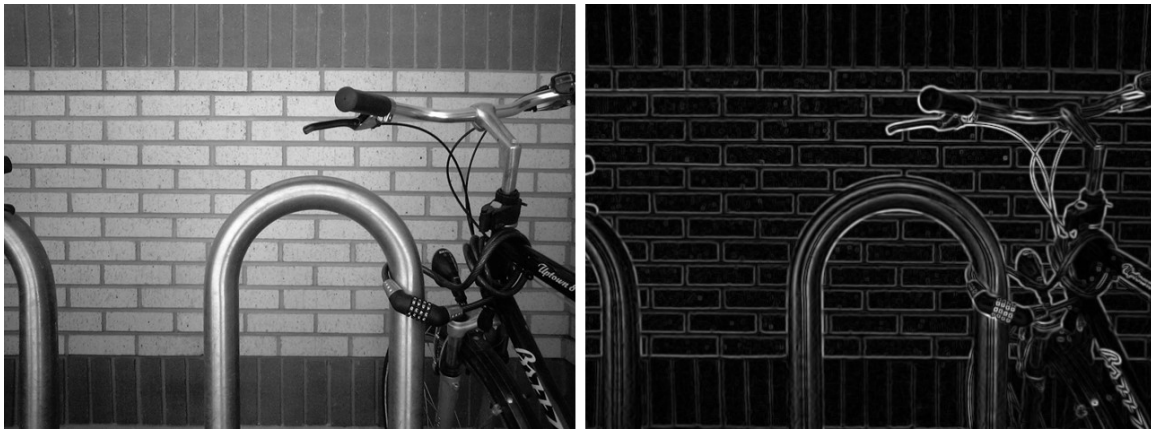
- I. η ορθή τοποθέτηση του ορίου στην εικόνα (π.χ. στη περίπτωση ενός προσώπου, μόλις εντοπιστεί η μύτη και σκιαγραφηθεί να μην τοποθετηθεί στο τμήμα που εμφανίζεται ένα μάτι)
- II. η ακρίβεια στον εντοπισμό της ακμής (περίπτωση ευθείας ακμής να μην εντοπίζεται ως καμπύλη)
- III. η ικανότητα διαφοροποίησης θορύβου από την δημιουργία της εικόνας ή έντονης υφής αντικειμένου (σε περίπτωση εντοπισμού ακμών σε ένα κάδρο να μην μπερδεύεται με την υφή του ξύλου)
- IV. ο χρόνος για την εκτέλεση του συστήματος να είναι όσο το δυνατόν μικρότερος

- V. ο ψηφιακός χώρος που χρειάζεται το σύστημα εντοπισμού να είναι όσο το δυνατόν μικρότερος

2.1 Αλγόριθμοι Ανίχνευσης Ακμών

Οι πρώτες απόπειρες επίλυσης του προβλήματος εντοπισμού ακμών ήταν αλγοριθμικές. Οι μέθοδοι όπου προτάθηκαν ανά χρονικά διαστήματα ήταν βασισμένες στη διαφοροποίηση (Differentiation), στην ανισότροπη διάχυση (Anisotropic Diffusion), στο ενεργό περίγραμμα (Active Contour), στην ασαφή λογική (Fuzzy Logic), στη χωρική συχνότητα (spatial frequency), στη στατιστική (statistics) και στην παρεμβολή υπο-εικονοστοιχείων (sub-pixel interpolation).

Με την μέθοδο της διαφοροποίησης εφαρμόζονται φίλτρα στην αρχική εικόνα όπου αναδεικνύουν τις ακμές των αντικειμένων. Αναλυτικότερα, ένα παράθυρο εικονοστοιχείων από την αρχική εικόνα πολλαπλασιάζεται με ένα συγκεκριμένο, σε διαστάσεις και τιμές, πίνακα και το αποτέλεσμα αποθηκεύεται σε ένα νέο εικονοστοιχείο μιας νέας παράγωγης εικόνας. Σε αυτή τη μέθοδο βασίζονται οι 3 βασικοί αλγόριθμοι/προτάσεις Sobel ([Sobel, 1970](#)), Laplacian-of-Gaussian (LOG) ([Marr και Hildreth, 1980](#)) και Canny ([Canny, 1986](#)). Η αδυναμίες αυτής της μεθόδου είναι ο θόρυβος των εικόνων, η έντονη υφές, η απώλεια πληροφορίας στις επιφάνειες και γενική αδυναμία στον προσδιορισμό μεγέθους και τιμών του πίνακα στα φίλτρα.

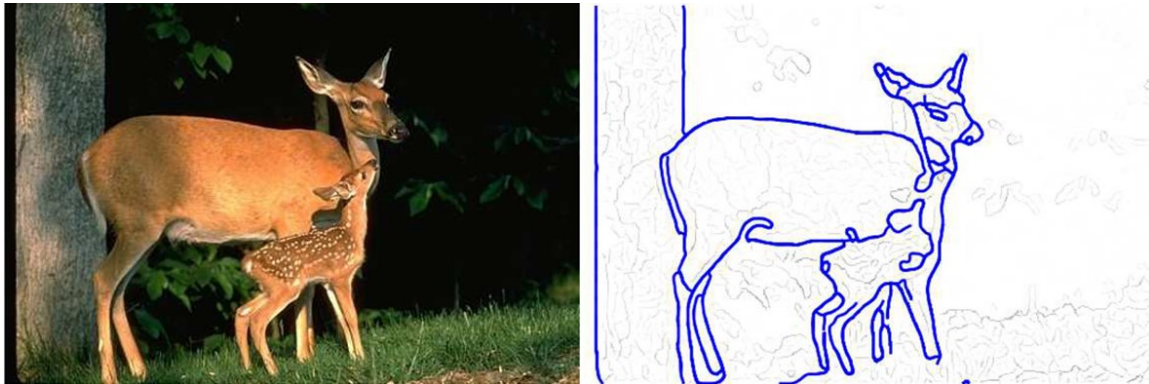


Εικόνα 02: Παράδειγμα εφαρμογής Sobel ([Wikipedia, 2022](#))

Με την μέθοδο διαφοροποίησης εφαρμόζονται φίλτρα με σκοπό να “απαλύνουν” την εικόνα, αυτό έχει ως αποτέλεσμα αυτή και ταυτόχρονα οι ακμές να θολώνουν. Στη μέθοδο ανισότροπης διάχυσης, για να μην χαθεί η πληροφορία των ακμών, εντοπίζονται τα τμήματα της εικόνας όπου έχουν κοινή επιφάνεια και εφαρμόζονται εκεί τα φίλτρα που “απαλύνουν” την εικόνα. Έτσι μένει ένα τελικό προϊόν, όπου οι ακμές παραμένουν όπως ήταν και τα τμήματα που θα μπέρδευαν τον αλγόριθμο εντοπισμού, διορθώνονται. Αυτή η μέθοδος προτείνεται από τους Perona και Malik ([Perona and Malik, 1990](#)).

Η μέθοδος ενεργού περιγράμματος είναι μία μέθοδος που εμφανίζεται στην πρόταση του Bastan ([Bastan et al., 2017](#)). Σε αυτή τη μέθοδο πρώτα γίνεται τοπική

αναγνώριση ακμών και μετά μια γενική. Τέλος εφαρμόζεται μια σύζευξη μεταξύ αυτών και καταλήγει στην εικόνα των ακμών.



Εικόνα 03: Παράδειγμα από εφαρμογή ενεργού περιγράμματος ([Bastan et al., 2017](#))

Η χρήση της ασαφούς λογικής, είναι μια μέθοδος που λόγω της ικανότητας να αντιμετωπίζει ζητήματα χωρίς να υπάρχει προσχεδιασμένο σχέδιο είναι καταλληλότερο για εικόνες που εμφανίζουν πολύ θόρυβο. Αναλυτικότερα, παίρνοντας το ιστόγραμμα από ένα μικρό τμήμα/παράθυρο της εικόνας, παρατηρείται ποιες τιμές εμφανίζονται συχνότερα και θεωρούνται ως επιφάνειες, είτε αντικειμένου, είτε φόντου. Έτσι οι ενδιάμεσες τιμές, άρα και τα εικονοστοιχεία που τις έχουν, θεωρούνται όρια-ακμές ανάμεσα στο αντικείμενο και το φόντο. Οι προτάσεις βασισμένες σε αυτή τη μέθοδο είναι από τους Versaci και Morabito ([Versaci and Morabito, 2021](#)) και η πρόταση από τον Russo ([Russo, 1998](#)). Όλες οι προτάσεις είχαν καλά αποτελέσματα σε εικόνες με έντονη εμφάνιση θορύβου.



Εικόνα 04: Παράδειγμα από εφαρμογή βασισμένη στη ασαφή λογική ([Wang et al., 2004](#))

Η μέθοδος της χωρικής συχνότητας εντοπίζει τις ακμές με τη βοήθεια του μετασχηματισμού των κυμάτων που ενισχύουν την διαφοροποίηση των ακμών. Ένα σύστημα βασισμένο σε αυτό είναι το Shearlet ([Yi et al., 2009](#)).

Προτάσεις που βασίζονται στην στατιστική είναι αυτές των Rakesh, Chaudhuri, Murthy ([Rakesh et al., 2004](#)) και του Konishi ([Konishi, 2003](#)). Για να αποφασιστεί εάν ένα

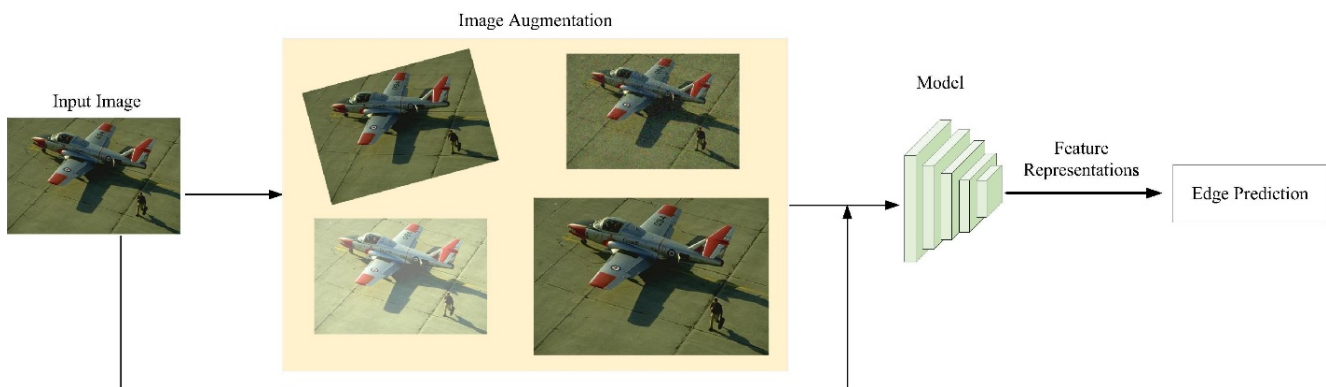
εικονοστοιχείο αποτελεί τμήμα ακμής, χρησιμοποιείται η υφή/μοτίβο των εικονοστοιχείων και η συμπεριφορά τους ανά θέση στη εικόνα. Συγκεκριμένα για κάθε εικονοστοιχείο παράγονται στατιστικά δεδομένα με βάση τα γειτονικά εικονοστοιχεία και με αυτά τα δεδομένα ελέγχεται κάθε εικονοστοιχείο εάν διαταράσσει την ομοιομορφία και αναγνωρίζεται αντίστοιχα ως ακμή ή όχι.

Η μέθοδος με παρεμβολή υπο-εικονοστοιχείων χρησιμοποιήθηκε στην πρόταση των Ghosal, Mehrotra ([Ghosal and Mehrotra, 1993](#)), Trujillo-Pino ([Trujillo et al., 2013](#)) και Seo ([Seo, 2018](#)). Σε αυτή τη μέθοδο γίνεται προσπάθεια εντοπισμού των ακμών με την χρήση των γειτονικών υπο-εικονοστοιχείων, βαρών και την χρήση των Zernike πολυωνύμων.

2.2 Εισαγωγή της Τεχνητής Νοημοσύνης στην Ανίχνευση Ακμών

Με την ανάπτυξη της μηχανικής μάθησης την τελευταία 20ετία, φυσικό ήταν να εμφανιστούν προτάσεις πάνω στο εντοπισμό ακμών με ενσωματωμένα μοντέλα τεχνητής νοημοσύνης. Οι προτάσεις αυτές χωρίζονται σε δύο κατηγορίες, με βάση τον τρόπο λειτουργίας τους. Αφενός στα συστήματα χωρίς επίβλεψη (unsupervised learning based image edge detection) και αφετέρου στα συστήματα με επίβλεψη (supervised learning based image edge detection). Βασική τους διαφορά είναι εάν τα δεδομένα εκπαίδευσης συνοδεύονται ή όχι από τα ζητούμενα αποτελέσματα (label features).

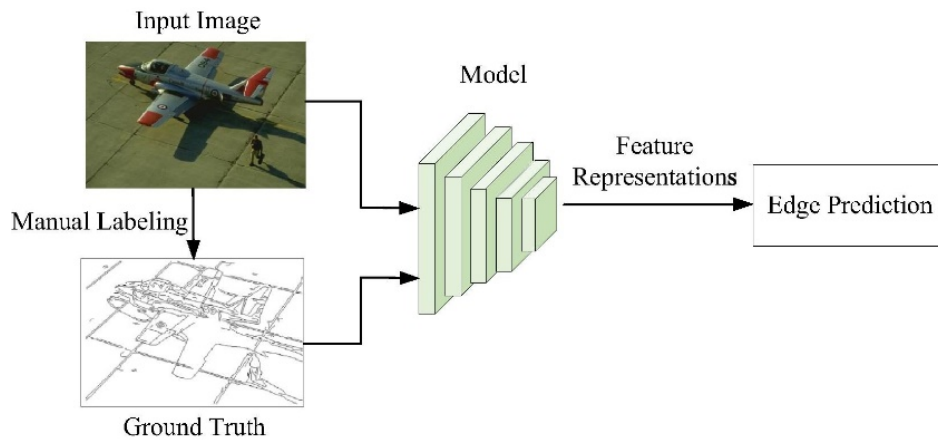
Στα μη επιβλεπόμενα συστήματα εκμάθησης για τον εντοπισμό ακμών, δεν χρειάζονται έτοιμα τα ζητούμενα αποτελέσματα. Διότι τα συστήματα αυτά μπορούν να προσεγγίσουν τα όρια των αντικειμένων, γνωρίζοντας μόνο την αρχική εικόνα. Κάποιες από τις προτάσεις που έχουν γίνει με τέτοια συστήματα είναι από τους Xiaofeng και Bo ([Xiaofeng and Bo, 2012](#)), Yang ([Yang et al., 2016](#)), Li ([Li et al., 2016](#)) και τους Xia και Kulis ([Xia and Kulis et al., 2017](#)).



Εικόνα 05: Διάγραμμα λειτουργίας μη επιβλεπόμενου συστήματος εκμάθησης ([Junfen et al., 2022](#))

Ένας μεγάλος αριθμός από τα τωρινά συστήματα εντοπισμού ακμών με χρήση AI, χρησιμοποιούν επιβλεπόμενα συστήματα εκμάθησης. Τα συστήματα αυτά απαιτούν μέσα στα δεδομένα εκπαίδευσης να υπάρχουν έτοιμες οι εικόνες του επιθυμητού τελικού αποτελέσματος, με σκοπό να μπορεί το σύστημα να παράγει όμοιες εικόνες. Οι προτάσεις που έχουν γίνει με τέτοια συστήματα είναι από Rahebi ([Rahebi and Tajik, 2011](#)), Zitnick

([Dollar and Zitnick, 2014](#)), Ganinand ([Ganinand and Lempitsky, 2014](#)), Xie ([Xie and Ju, 2015](#)), He ([He et al. 2022](#)) και Liu ([Liou, 2019](#)).



Εικόνα 06: Διάγραμμα λειτουργίας επιβλεπόμενου συστήματος εκμάθησης ([Junfen et al., 2022](#))

Περισσότερα πάνω στην πρόοδο του εντοπισμού ακμών βρίσκονται στο δημοσίευμα ([Junfen et al., 2022](#)).

3. Μεθοδολογία

Για την επίτευξη του τελικού στόχου, έγινε η εκπαίδευση ενός δικτύου τεχνητής νοημοσύνης (AI) ώστε να παράγεται μία δυαδική (binary) εικόνα όπου έχουν απεικονιστεί οι ακμές των λίθων. Ακολούθησε μια μικρή επεξεργασία των εικόνων αυτών με γνωστά φίλτρα και τελείωσε η μετατροπή τους από εικονικό δεδομένο, σε διανυσματικό, με την χρήση μιας συνάρτησης μετατροπής ψηφιδωτής (raster) πληροφορίας, σε διανυσματική (vector). Τα εργαλεία επιλέχθηκαν να είναι ένας έτοιμος κώδικας/βιβλιοθήκη για AI (open source), η rix2pix ([Zhu et al., 2017](#)), δύο γνωστές ανοιχτές βιβλιοθήκες η OpenCV (επεξεργασίας εικόνων) και η GDAL (επεξεργασίας γεωχωρικών δεδομένων).

Η διαδικασία μπορεί να χωριστεί σε δύο στάδια. Το πρώτο στάδιο είναι η μετατροπή ορθοφωτογραφιών που απεικονίζουν λίθινες κατασκευές, σε αντίστοιχες ασπρόμαυρες εικόνες όπου τονίζονται τα όρια των λίθων. Το δεύτερο στάδιο είναι η μετατροπή των παραγόμενων εικόνων σε διανυσματικό αρχείο που δίνει την δυνατότητα για επεξεργασία.

3.1 Από ορθοφωτογραφία σε εικόνα ακμών

Στο στάδιο αυτό, έγιναν δύο επαναλήψεις πειραματισμών. Η πρώτη, έγινε με σκοπό τον εντοπισμό των βέλτιστων παραμέτρων της rix2pix, που να παράγουν τα καλύτερα δυνατά αποτελέσματα. Η δεύτερη, είχε ως στόχο την παρατήρηση συμπεριφοράς του πλέον βελτιστοποιημένου μοντέλου, με την προσθήκη νέων δεδομένων.

Συνολικά έγιναν 76 πειράματα. Τα πρώτα 69, (που αποτελούν και τις δοκιμές του πρώτου γύρου δοκιμών), έγιναν στο περιβάλλον του Google Colaboratory⁽⁰¹⁾, με την χρήση του προσωπικού ψηφιακού χώρου αποθήκευσης Google Drive⁽⁰²⁾ και του Online Documentation Weights & Biases⁽⁰³⁾. Τα υπόλοιπα 4 πειράματα, (δηλ. οι δοκιμές που έγιναν για παρατήρηση συμπεριφοράς των βέλτιστων μοντέλων με νέα δεδομένα), έγιναν στις εγκαταστάσεις του ΣΑΤΜ - ΜΓ ΕΜΠ.

Πίνακας 01: Κάρτες γραφικών που χρησιμοποιήθηκαν

Περιβάλλον	Όνομα	Memory (GB)	Memory Bandwidth (GB/s)	Base Clock (MHz)
Google Colaboratory	Tesla K80	12 x2	240.6 x2	562
	Tesla T4	16	320.0	585
	Tesla P100	16	732.2	1190
ΣΑΤΜ - ΜΓ	RTX 3060	12	360.0	1320

Τα δεδομένα που χρησιμοποιήθηκαν είναι τμήματα από παλαιότερες ορθοφωτογραφίες του Εργαστηρίου Φωτογραμμετρίας, που απεικονίζουν λίθινους τοίχους και τα αντίστοιχα έτοιμα τμήματα διανυσματικής απεικόνισης των ακμών των

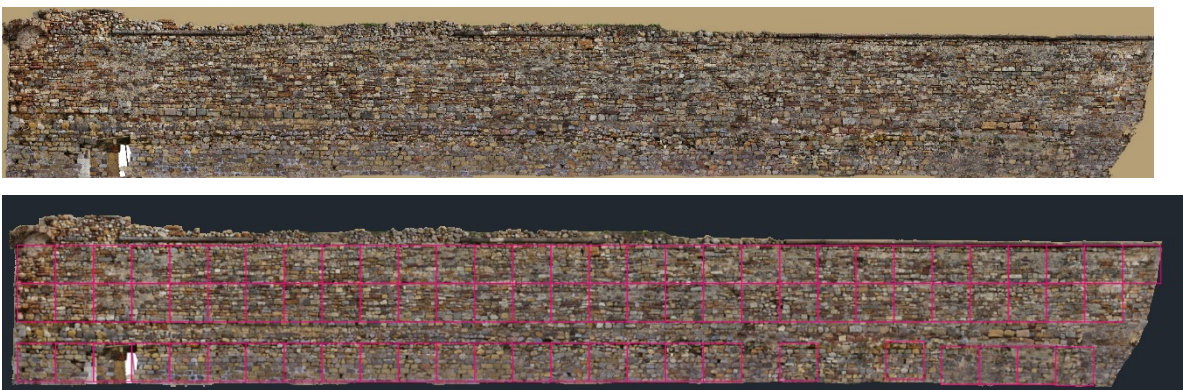
λίθων. Αυτό σημαίνει πως οι εικόνες ήδη έχουν υποστεί μια προ-επεξεργασία πριν τη χρήση τους, είναι δηλαδή με διορθωμένη γεωμετρία (έχουν εξαλειφθεί οι παραμορφώσεις και οι αποκλίσεις) και αποχρώσεις.

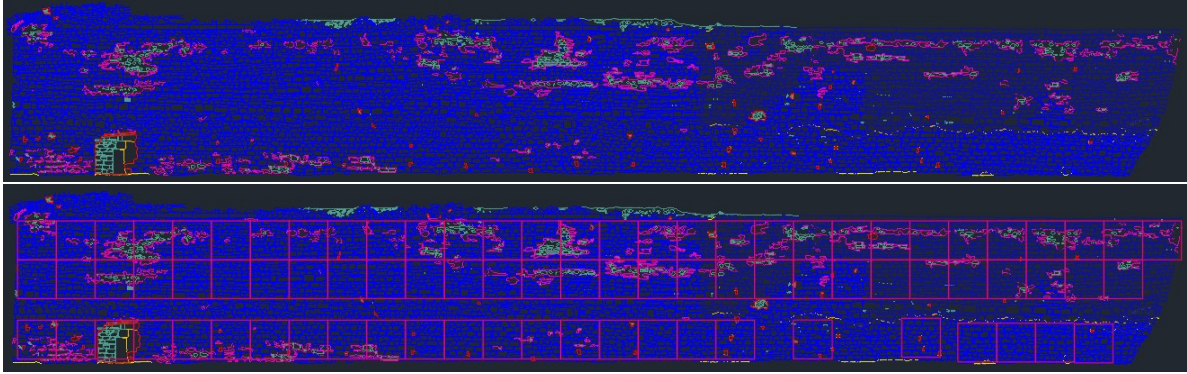
Επίσης στην πληροφορία της εικόνας, πέρα από λίθους υπάρχουν προσημασμένα με στόχους φωτοσταθερά, κενά σημεία λόγω στηρίξεων σκαλωσιάς (σκαλότρυπες) καλυμμένες με πολύγωνα σταθερού χρώματος, ή άλλου εμποδίου εκτός θεματολογίας, εξωτερικά όρια θέματος, βλάστηση και σκαλίσματα πάνω στις πέτρες. Αυτά αποτελούν πληροφορίες, που στην διανυσματική απεικόνιση δεν υπάρχουν και πρέπει το μοντέλο που θα προκύψει μετά την εκπαίδευση να έχει μάθει να τα αγνοεί κατά την σχεδίαση των ακμών.



Εικόνα 07: Πληροφορία που δεν πρέπει να κρατηθεί: α) στόχοι φωτοσταθερών, β) εξωτερικά όρια θέματος, γ) βλάστηση, δ) σκαλίσματα, ε) σκαλότρυπες

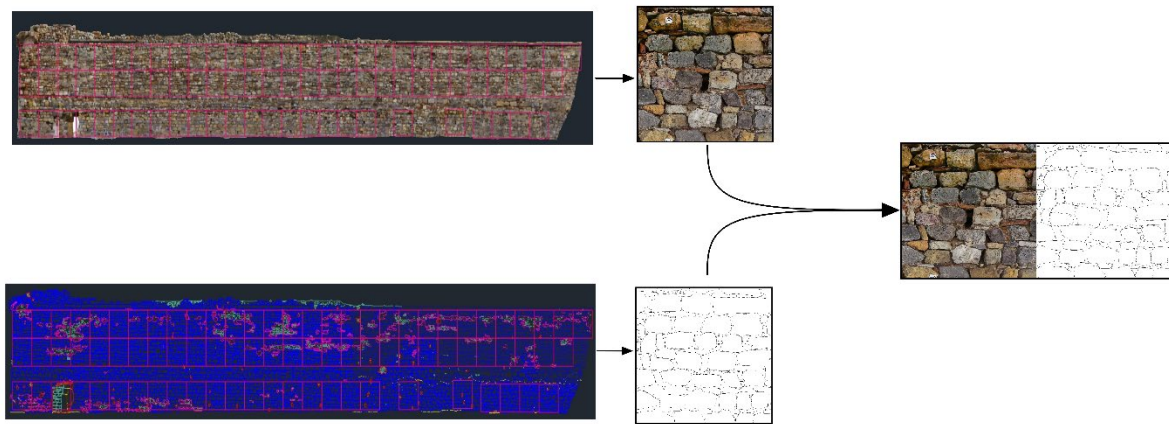
Ο κώδικας που χρησιμοποιήθηκε, απαιτεί ως δεδομένα ένα ζεύγος εικόνων, με την μία να απεικονίζει την πραγματικότητα και την άλλη το επιθυμητό αποτέλεσμα. Στα δικά μας δεδομένα, η μία εικόνα είναι η ορθοφωτογραφία και η άλλη, η διανυσματική απεικόνιση. Λόγω του μεγάλου όγκου των αρχικών εικόνων, για τις οποίες δεν υπάρχει διαθέσιμο υπολογιστικό σύστημα που να μπορεί να τις διαχειριστεί, κρίθηκε αναγκαία η τμηματοποίησή των εικόνων αυτών, όπως και η ενοποίησή τους.





Εικόνα 08: Παράδειγμα α) Αρχικής Πραγματικής Εικόνας β) Τμηματοποίησης Αρχικής Εικόνας γ) Διανυσματική Απεικόνιση δ) Τμηματοποίηση Διανυσματικής Απεικόνισης (με σειρά αντιστοιχίας από πάνω προς τα κάτω)

Η τελική διάσταση των τμημάτων που εμπεριέχουν αρκετή πληροφορία λίθων και λοιπών στοιχείων, κατέληξε 1024x1024, άρα 1024x2048 για τις ενωμένες εικόνες. Ο λόγος είναι ένας λειτουργικός περιορισμός του κώδικα, που δέχεται εικόνες με διαστάσεις πολλαπλάσιες του 4 (ο περιορισμός αυτός παρουσιάζεται παρακάτω στη λειτουργία του ίδιου του κώδικα). Επομένως έγινε αποκοπή τμημάτων των αρχικών ορθοφωτογραφιών με τις αντίστοιχες διανυσματικές απεικονίσεις τους στο περιβάλλον του AutoCAD⁽⁰⁴⁾ και ακολούθησε η ένωση αυτών με το δοσμένο κώδικα (από την βιβλιοθήκη rix2rix).



Εικόνα 09: Διάγραμμα Ενοποίησης Εικόνων

3.1.1 Λειτουργία της AI στην rix2rix

Όπως προαναφέρθηκε χρησιμοποιήθηκε ήδη μία υπάρχουσα βιβλιοθήκη-κώδικας AI, η rix2rix (Zhu et al., 2017). Η βιβλιοθήκη αυτή, έχει δομηθεί να δίνει όλα τα απαραίτητα εργαλεία, έτσι ώστε να μπορεί να εφαρμόσει την επιστήμη της τεχνητής νοημοσύνης, σε αρκετά ζητήματα επεξεργασίας εικόνας, χωρίς να υπάρχει κάποια ειδική προ-απαιτήση. Στην ίδια την δημοσίευση, παρουσιάζονται πληθώρα παραδειγμάτων από καλλιτεχνικής φύσεως εφαρμογές, (π.χ. δημιουργία αναγεννησιακού πίνακα από μία εικόνα) έως και πρακτική χρήση (π.χ. μετατροπή αεροφωτογραφίας σε απλοποιημένο χάρτη).

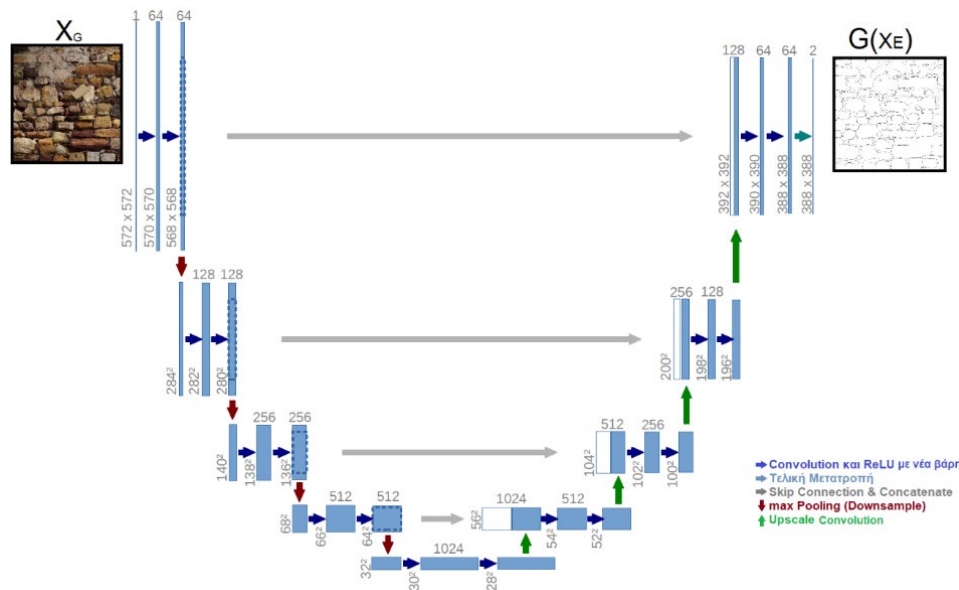
Ο κώδικας χρησιμοποιεί ένα νευρωνικό δίκτυο τύπου, cGAN^[03] και τα βασικά κομμάτια που τον αποτελούν, είναι ο Generator και ο Discriminator, κομμάτια που αναλύονται παρακάτω. Κατά τη χρήση του εφαρμόζονται δύο στάδια: Αυτό της εκπαίδευσης, όπου εκπαιδεύεται το νευρωνικό δίκτυο και αυτό της δοκιμής, όπου το εκπαιδευμένο πλέον δίκτυο επεξεργάζεται νέες εικόνες και αξιολογείται.

Στο στάδιο της εκπαίδευσης, ο Generator και ο Discriminator βρίσκονται σε μια συνεχόμενη προσπάθεια να μπερδέψουν το ένα το άλλο, μέχρι να ολοκληρωθεί ο αριθμός των εποχών^[04]. Ο σκοπός αυτού του σταδίου είναι ο Generator να καταφέρει να ξεγελάσει τον Discriminator και να είναι όσο το δυνατόν κοντά στην πραγματικότητα το τελικό προϊόν του.

Συμβολισμοί:

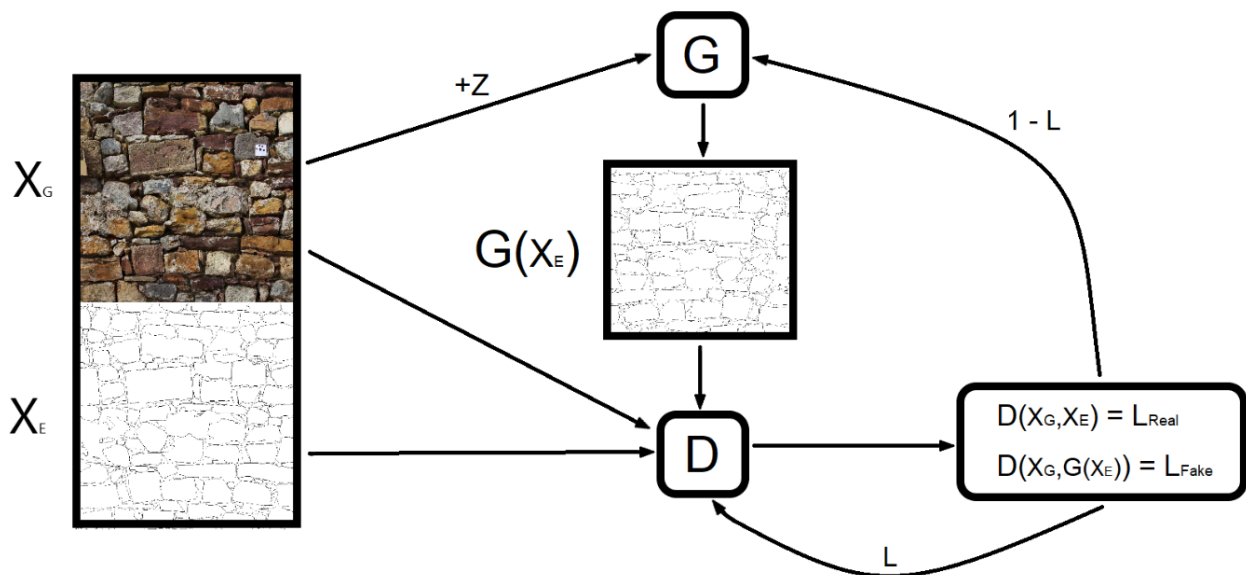
- G = Generator
- D = Discriminator
- X_G = Πραγματική Εικόνα
- X_E = Επιθυμητό Αποτέλεσμα που Βασίζεται στη X_G
- $G(X_E)$ = Εικόνα που μοιάζει με την X_E και δημιουργήθηκε από τον G
- Z = Τυχαίο Διάνυσμα
- L_{Fake} = τιμή που προκύπτει στον έλεγχο μεταξύ συνθετικής με επιθυμητής εικόνας
- L_{Real} = τιμή που προκύπτει στον έλεγχο μεταξύ πραγματικής με επιθυμητής εικόνα

Ο Generator, είναι το νευρωνικό δίκτυο που μετά την εκπαίδευσή του, σε μια ιδανική περίπτωση, θα μπορεί να βλέπει εικόνες ίδιας φύσεως με την πραγματική (X_G) και θα παράγει την αντίστοιχη συνθετική $G(X_E)$, που θα ξεγελά τον Discriminator. Σε κάθε εποχή λαμβάνει την αρχική εικόνα (X_G), το επιθυμητό αποτέλεσμα (X_E), ένα τυχαίο διάνυσμα (Z), ως ενεργοποιητικό μηχανισμό δικτύου (Network Initialization^[05]), την τιμή διόρθωσης $1 - L$ (που επηρεάζει τα βάρη εντός του δικτύου) και καλείται να μάθει να κατασκευάζει εικόνες $G(X_E)$, ίδιες με την επιθυμητή (X_E). Στόχος που ισοδυναμεί με την προσπάθεια αύξησης της τιμής L . Οι αρχιτεκτονικές δομές που μπορεί να έχει είναι είτε U-Net^[06] με 256×256 , ή 128×128 διαστάσεις παραθύρων, είτε ResNET^[07] με βήμα 6 ή 9 block.



Σχήμα 01: Διάγραμμα Generator με αρχιτεκτονική U-Net με εικόνα input και output

Ο Discriminator, είναι κι αυτό νευρωνικό δίκτυο που μετά την εκπαίδευσή του, σε μια ιδανική περίπτωση, μπορεί να διαχωρίσει την συνθετική εικόνα ($G(X_E)$) του Generator από την πραγματική (X_E). Σε κάθε εποχή λαμβάνει την παραγόμενη εικόνα $G(X_E)$, το επιθυμητό αποτέλεσμα (X_E), την αρχική εικόνα (X_G), την τιμή διόρθωσης L και αξιολογεί το πόσο αληθοφάνες είναι το προϊόν του Generator, όπως και την ίδια του την ικανότητα στη διάκριση ψεύτικου από αληθινό. Η αξιολόγηση αυτή μεταφράζεται στις τιμές L_{Fake} και L_{Real} , που συνυπολογίζονται μαζί με τη norm $L1$ ^[08] (που εφαρμόζεται μεταξύ της παραγόμενης εικόνας $G(X_E)$ και της επιθυμητής (X_E)) σε μια τελική τιμή L την οποία προσπαθεί να μειώσει.



Σχήμα 02: Γενικό Σχεδιάγραμμα Εκπαίδευσης

Σε μορφή εξισώσεων ισχύει:

$$L_{cGAN}(G, D) = E_{(X_G, X_E)}(\log[D(X_G, X_E)]) + E_{(X_G, G(X_E))}(\log[1 - D(X_G, G(X_E))])$$

$$L_{L1}(G) = E_{(X_G, X_E)}(\|D(X_G, X_E) - G(X_E)\|_1)$$

Με Σκοπό:

$$G_{Final} = \operatorname{argmin}_G \max_D L_{cGAN}(G, D) + \lambda * L_{L1}(G)$$

Όπου:

- $D(X_G, X_E) = L_{Real}$
- $D(X_G, G(X_E)) = L_{Fake}$
- $E_{(X_G, X_E)}(\log[D(X_G, X_E)]) = \text{λογαριθμική τιμή της } L_{Real}$
- $E_{(X_G, G(X_E))}(\log[1 - D(X_G, G(X_E))]) = \text{λογαριθμική τιμή της } L_{Fake}$
- $L_{cGAN}(G, D) = \text{συνολική τιμή απόκλισης βασισμένη στον } D$
- $L_{L1}(G) = \text{γραμμική απόκλιση της ψεύτικης εικόνας από την πραγματική}$
- $G_{Final} = \text{τελική τιμή απόκλισης βασισμένη στην } L_{cGAN}(G, D) \text{ και } L_{L1}(G)$

Στο στάδιο της δοκιμής χρησιμοποιείται ο εκπαιδευμένος πλέον Generator. Του δίνονται πραγματικές εικόνες (X_G) και παράγει σε δευτερόλεπτα την αντίστοιχη συνθετική (X_E), οι οποίες αξιολογούνται από τον χρήστη.

3.1.2 Βελτιστοποίηση των παραμέτρων του μοντέλου

Η βιβλιοθήκη, μας δίνει πληθώρα επιλογών για την εκπαίδευση και δοκιμή των νευρωνικών δικτύων. Μετά από μελέτη των δυνατοτήτων που παρέχονται, κρίθηκε απαραίτητο ένα στάδιο πειραμάτων, όπου θα βρίσκονταν οι βέλτιστες παράμετροι. Κάποιες από αυτές, λόγω της φύσης του ζητήματός θα είναι συγκεκριμένες, από εκεί και πέρα υπάρχει η ανάγκη για πειραματισμό συνδυασμών στην υπόλοιπη παραμετροποίηση για την επίτευξη καλύτερου αποτελέσματος.

Συνολικά έγιναν 69 δοκιμές, οι οποίες χωρίστηκαν σε 10 φάσεις. Κάθε φάση είχε μία παράμετρο που άλλαζε και οι υπόλοιπες ήταν σταθερές. Με την μέθοδο αυτή, βρέθηκε η καλύτερη παραμετροποίηση.

Πίνακας 02: Παρουσίαση φάσεων με αντίστοιχες μεταβλητές

Φάση	Μεταβαλλόμενη Παράμετρος	Κλειδίωμα
A	Κανάλια Εξόδου + Αρχιτεκτονική Νευρωνικού Δικτύου	1 κανάλι (Μαύρο – Άσπρο) + Unet – 256, ResNET 6blocks

B	Network Initialization ^[09]	Normal ^[10] + Xavier ^[11]
Γ	Λειτουργία Discriminator	PatchGAN ^[12] + Pixel
Δ	Normalization	Batch
E	Τύπος GAN	Vanilla ^[13] + wgangp ^[14]
Z	Μεγέθη Τμημάτων	1024
H	Εποχές – Epoch	400 (100 - 300)
Θ	Learning Rate ^[15]	Linear ^[16]
I	Batch Size	1
K	Discriminator – Generator Filters	32 – 128

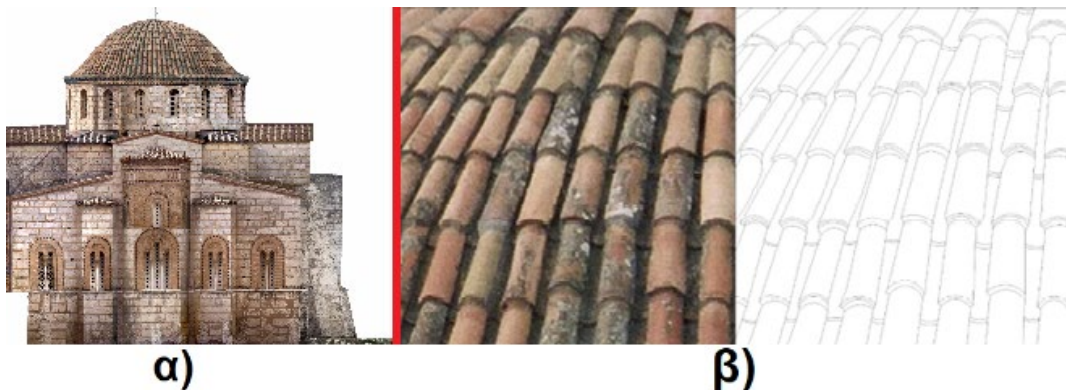
Η αξιολόγηση των αποτελεσμάτων μετά από κάθε εκπαίδευση, έγινε με την παρατήρηση του ρυθμού μεταβολής των δεικτών Loss^[17] που μας δίνονταν και με ανθρώπινη παρατήρηση/σύγκριση κάθε παραγόμενης εικόνας, με την αναμενόμενη.

Περισσότερη έμφαση στη αιτιολόγηση επιλογής σε κύριες παραμέτρους, δίνεται στο [4^ο κεφάλαιο](#) και τα αποτελέσματα από τα πειράματα παρουσιάζονται στο συνημμένο αρχείο [Πειράματα](#).

3.1.3 Συμπεριφορά με διαφορετικά δεδομένα

Μετά την εύρεση των βέλτιστων παραμέτρων, ακολούθησε ένας νέος μικρότερος κύκλος πειραματισμού. Ο σκοπός είναι ο εντοπισμός της καλύτερης κατανομής ώστε να καλύπτονται όσες το δυνατόν περιπτώσεις γίνονται, όπως και παρατήρηση του πώς αντιδρά το βελτιστοποιημένο μοντέλο. Πρώτα έγινε η δοκιμή των παραμέτρων αυτών με διαφορετική κατανομή δεδομένων και ύστερα με την προσθήκη άλλων 33 ζευγών εικόνων, από μια λιθόχτιστη κατασκευή. Αναλυτικότερα, έγιναν μια σειρά από δοκιμές όπου η κατανομή των εικόνων άλλαζε, αλλά με ίδια θεματολογία δεδομένων και μια σειρά από δοκιμές με ανάμεικτα δεδομένα.

Τα δεδομένα που δόθηκαν ήταν από την γεωμετρική τεκμηρίωση του I.N. της Μονής Δαφνίου (Georgoroulos et al. 2004, Georgoroulos 2006). Αντίστοιχα έγινε η διαδικασία κοπής τμημάτων της εικόνας και δημιουργίας ζευγών εικόνων (X_G, X_E) όπως στο προηγούμενο σύνολο δεδομένων.



Εικόνα 10: Παράδειγμα νέας εικόνας α) και νέου ζεύγος τμημάτων αυτής β)

Πίνακας 03: Κατανομή των δεδομένων στο 2^ο κύκλο πειραμάτων

Δοκιμή	Train (Αριθμός Εικόνων)	Test (Αριθμός Εικόνων)
A	Data Χίου (67)	Data Χίου (17)
B	Data Χίου (42)	Data Χίου (42)
C	Data Χίου (17)	Data Χίου (67)
D	Data Χίου (84)	Data Δαφνί (33)
E	Data Δαφνί (33)	Data Χίου (84)
F	Data Δαφνί (26)	Data Δαφνί (7)
G	Data Χίου + Δαφνί (69 + 24)	Data Χίου + Δαφνί (15 + 9)

3.2 Από εικόνα σε διάνυσμα

Μετά την ολοκλήρωση και του 1^{ου} σταδίου απομένουν δυαδικές (binary) εικόνες όπου απεικονίζουν με μαύρο τα όρια των λίθων και άσπρο την επιφάνεια αυτών. Σε αυτό το στάδιο γίνεται η μετατροπή των εικόνων αυτών σε διανυσματικά αρχεία, που δίνει την δυνατότητα επεξεργασίας των ακμών.

Για να επιτευχθεί αυτό χρησιμοποιήθηκαν δύο ανοιχτές βιβλιοθήκες στο περιβάλλον της python, η OpenCV⁽⁰⁵⁾ και η GDAL⁽⁰⁶⁾. Με την OpenCV έγινε η εφαρμογή μιας σειράς από φίλτρα στις εικόνες με σκοπό να τονιστούν τα όρια των λίθων και με την GDAL εφαρμόστηκε η συνάρτηση όπου μετατρέπει μια πληροφορία εικόνας σε διανυσματική.

3.2.1 Φίλτρα

Κατά την διάρκεια των πειραματισμών στο 1^ο στάδιο, είχαν παρατηρηθεί ότι σε πολλές γραμμές ακμών υπήρχαν ασυνέχειες, κενά και μικρά σύννεφα εικονοστοιχείων, παράλληλα των ακμών. Αυτά τα φαινόμενα σε ένα μεγάλο βαθμό μειώθηκαν, με την

χρήση των καλύτερων παραμέτρων που εντοπίστηκαν, κατά τη βελτιστοποίηση του δικτύου ([Κεφάλαιο 4](#)). Για την περαιτέρω βελτίωση των απομενόντων ελαττωμάτων σε παράθυρα μεγέθους 10x10 pixel, αποφασίστηκε η εφαρμογή φίλτρων στις εικόνες. Συγκεκριμένα εφαρμόστηκαν τα εξής φίλτρα: Gaussian Blur, Dilation και του κατωφλίου (threshold).

Αρχικά, έγινε η εφαρμογή του φίλτρου Gaussian Blur. Στόχος χρήσης αυτού του φίλτρου είναι η ένωση των κοντινών ακμών και γέμισμα πιθανών κενών κατά μήκος της μαύρης ακμής/γραμμής. Με το Gaussian Blur, δημιουργείται ένα νέφος γύρω από τις ακμές που απεικονίζεται με διαφορετικές τιμές της κλίμακας του γκρι. Αποτέλεσμα αυτού τα παλιά άσπρα εικονοστοιχεία με τιμή 255, να αποκτήσουν μια τιμή του γκρι από 0 έως 255.

Έχοντας πλέον αποδώσει τόνους του γκρι στα προβληματικά τμήματα των εικόνων εφαρμόζεται κατωφλίωση. Συγκεκριμένα όλα τα εικονοστοιχεία πάνω από την τιμή 234 βάφονται άσπρα και τα κάτω από αυτήν μαύρα. Η τιμή αυτή βρέθηκε με μικρές δοκιμές σε ένα πλήθος από εικόνες, με σκοπό την καλύτερη διόρθωση των εικόνων, χωρίς όμως την δημιουργία μεγαλύτερου προβλήματος.



Εικόνα 11: Παράδειγμα εφαρμογής των φίλτρων α) αρχική εικόνα β) εφαρμογή φίλτρο Gaussian γ) κατωφλιωμένη εικόνα

Τέλος, έγινε εφαρμογή του φίλτρου Dilation. Με την εφαρμογή αυτού του φίλτρου επιτυγχάνεται η αύξηση του πάχους των ακμών που απεικονίζονται κατά 2 pixel. Έτσι συμπληρώνεται οποιοδήποτε μικρό κενό δεν κατάφεραν τα προηγούμενα φίλτρα να καλύψουν.



Εικόνα 12: Παράδειγμα εικόνας πριν και μετά την εφαρμογή του Dilation φίλτρου

Αξίζει να σημειωθεί, πως λόγω των μικρών τιμών που δόθηκαν στα φίλτρα για εφαρμογή, σε καμία από τις εικόνες που έγιναν δοκιμές δεν εμφανίστηκαν νέες ανύπαρκτες αρχικά γραμμές.

3.2.2 Διανυσματοποίηση

Σε αυτό το στάδιο πλέον, υπάρχει μία εικόνα όπου εμφανίζονται όλες οι ακμές των λίθων και επόμενο πρόβλημα είναι η μετατροπή της εικόνας αυτής σε διανυσματική μορφή. Την λύση την δίνει η GDAL σε μορφή συνάρτησης, την Polygonize. Η συνάρτηση αυτή εντοπίζει τα εικονοστοιχεία που έχουν ίδιες τιμές, αναγνωρίζει ποια από αυτά έχουν κοινή πλευρά, κρατά μόνο την εξωτερική και ζωγραφίζει το περίγραμμά των εξωτερικών πλευρών.



Εικόνα 13: Παράδειγμα εικόνας πριν και μετά την εφαρμογή της συνάρτησης Polygonize

4. Βελτιστοποίηση Δικτύου

Όπως προαναφέρθηκε, η rix2pix παρέχει αρκετές επιλογές παραμετροποίησης δικτύου. Μετά από την πρώτη δοκιμή, με την προκαθορισμένη παραμετροποίηση, βγήκαν υποσχόμενα αποτελέσματα. Όμως όπως και σε κάθε διεργασία με AI, απαιτείται μια σειρά από αλλαγές στις παραμέτρους, με σκοπό την βελτιστοποίηση των αποτελεσμάτων.

Για την βελτιστοποίηση αυτή χρησιμοποιήθηκαν τα δεδομένα της Χίου, σε ποσοστά 80% (67 εικόνες), για εκπαίδευση (training) και 20% (17 εικόνες), για έλεγχο (test). Με αυτή τη κατανομή έγιναν επαναλαμβανόμενες εκπαιδεύσεις και έλεγχοι με διαφορετικούς συνδυασμούς παραμέτρων, σε 10 φάσεις. Οι παράμετροι που δίνονται και μεταβλήθηκαν κατά την διάρκεια των πειραμάτων με τις αντίστοιχες φάσεις παρουσιάζονται στον [Πίνακα 02](#). Κατά την διάρκεια των πειραμάτων αρκετοί συνδυασμοί με μη ικανοποιητικά αποτελέσματα κρατήθηκαν και σε περαιτέρω φάσεις με την προοπτική ότι με την μεταβολή κάποιας άλλης παραμέτρου τα αποτελέσματα θα άλλαζαν.

Μετά την Α φάση, αποφασίστηκε η χρήση του μονού καναλιού (grayscale) ως τελική έξοδο εικόνας. Ο λόγος αυτής της απόφασης είναι η εμφάνιση μιας σταθερής σύγχυσης που “λερώνει/ζαρώνει” την περιοχή των ορίων κάθε εικόνας όταν χρησιμοποιούνται τρία κανάλια εξόδου (RGB). Ως αρχιτεκτονική δικτύου αποφασίστηκε η χρήση του ResNET 6blocks αφού βγάζει καθαρότερα και κοντινότερα στην πραγματικότητα αποτελέσματα, σε αντίθεση με την αρχιτεκτονική Unet που εμφάνιζαν μαύρες κουκίδες. Όμως η χρήση της αρχιτεκτονικής τύπου Unet 256 δεν απορρίφθηκε πλήρως. Παρέμεινε με την υπόθεση ό,τι με την αλλαγή κάποιας άλλης παραμέτρου, το αποτέλεσμα θα βελτιωθεί.



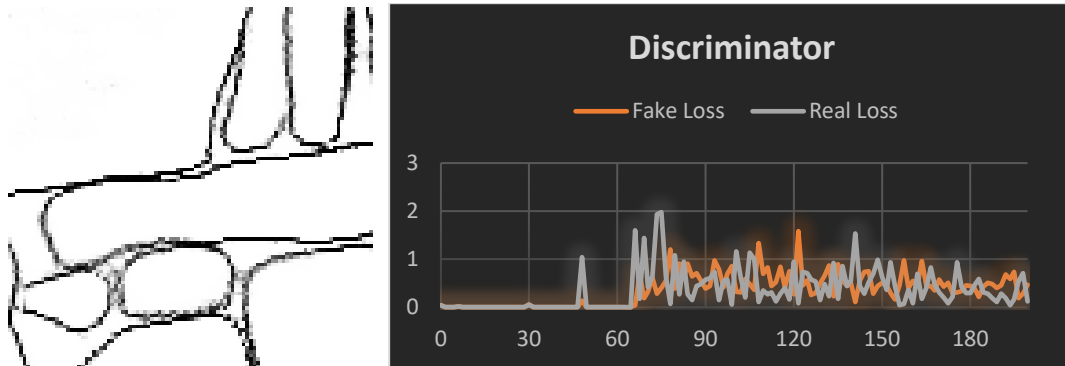
Εικόνα 14: Παράδειγμα ζαρωμένων ακμών σε RGB (1^ο Πείραμα)



Εικόνα 15: Παράδειγμα μαύρων κουκίδων με χρήση Unet αρχιτεκτονικής (2^ο Πείραμα)

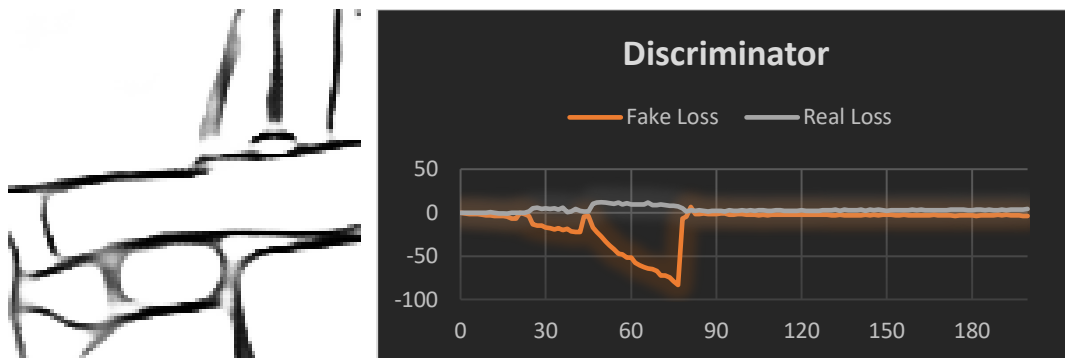
Στη φάση Β κατάληξε η χρήση των Normal και Xavier ως επιλογή για network initialization. Στη απόφαση βοήθησαν τα εικονικά αποτελέσματα, αφού αυτά υπέδειξαν ποια επιλογή ήταν πιο κοντά στη πραγματικότητα. Σε αυτή τη φάση, εμφανίστηκε το πρώτο στοιχείο, ότι τα διαγράμματα των Loss ενώ προϋδέαζαν ένα σωστό αποτέλεσμα, στη πραγματικότητα υπήρχε μεγάλη απόκλιση. Για παράδειγμα, στην εικόνα 16 το

διάγραμμα στις πρώτες 60 εποχές έδειξε μηδενική μεταβολή, που προϋπέθετε ότι η εκπαίδευση ήταν ανεπιτυχής.



Εικόνα 16: Παράδειγμα σχετικά καλού αποτελέσματος, αλλά κακού διαγράμματος Loss με χρήση Xavier ως initialization network (9^ο Πείραμα)

Ο τύπος GAN επιλέχθηκε μετά την ολοκλήρωση της φάσης E. Συγκεκριμένα, κατάληξε η χρήση του Vanilla, λόγω των ρεαλιστικών αποτελεσμάτων του και του wgangp, λόγω της ιδιαιτερότητας των αποτελεσμάτων.




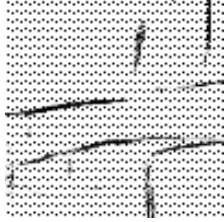





Εικόνα 17: Παράδειγμα εικόνας αποτελέσματος με wgangp με αντίστοιχο Loss διάγραμμα (31^ο Πείραμα)

Τέλος, το πλήθος το εποχών αποφασίστηκε στη φάση H, όπου το πλήθος που ελέγχθηκε ήταν έως και 600 εποχές, χωρισμένες στις δύο διαδικασίες του αλγορίθμου. Οι διαδικασίες είναι αυτές όπου το νευρωνικό δίκτυο λαμβάνει το network initialization και αυτές όπου το δίκτυο μαθαίνει, μέσω του learning rate, adam κλπ. Ο καλύτερος συνδυασμός που προέκυψε, ήταν αυτός των 400 εποχών, 100 στη 1^η διαδικασία και 300 στη 2^η. Η απόφαση αυτή πάρθηκε, διότι με την αύξηση στη 1^η διαδικασία δεν άλλαζαν πολλά και με αύξηση στη 2^η το δίκτυο άρχιζε να μπερδεύεται από μόνο του και να χειροτερεύει τα αποτελέσματα.

Κατά τη διάρκεια των πειραμάτων προέκυψαν ενδιαφέροντα αποτελέσματα, τα οποία και κρατήθηκαν κατά την διάρκεια των πειραμάτων. Ο σκοπός ήταν να παρατηρηθεί

η πιθανή μεταβολή των αποτελεσμάτων αυτών προς το καλύτερο, όσο άλλαζαν οι λοιπές παράμετροι.

Πίνακας 04: Αποτελέσματα στο 1^ο Κύκλο Πειραμάτων

Πείραμα	Πραγματική Εικόνα	Παραγόμενη Εικόνα	Ζητούμενη Εικόνα
13°			
17°			
22°			
33°			
35°			

Για αναλυτικότερη θέαση των πειραμάτων, δίνεται συνημμένο ένα [αρχείο](#), στο οποίο παρουσιάζονται όλων των πειραμάτων τα διαγράμματα τιμών Loss, τα εικονικά τμήματα ενδιαφέροντος, η αντίστοιχη παραμετροποίηση και σχόλια πάνω σε κάθε ολοκλήρωση φάσης. Παρακάτω παρατίθεται μια γενική εικόνα των δοκιμών από πλευράς απαιτούμενου χρόνου εκτέλεσης, απαιτήσεις σε μνήμη, η κάρτα γραφικών που χρησιμοποιήθηκε και οι τελικές τιμές Loss. Αξίζει να σημειωθεί, πως το μόνο απαιτητικό στάδιο είναι αυτό της εκπαίδευσης, ανεξάρτητα της κάρτας γραφικών, ή των τιμών Loss.

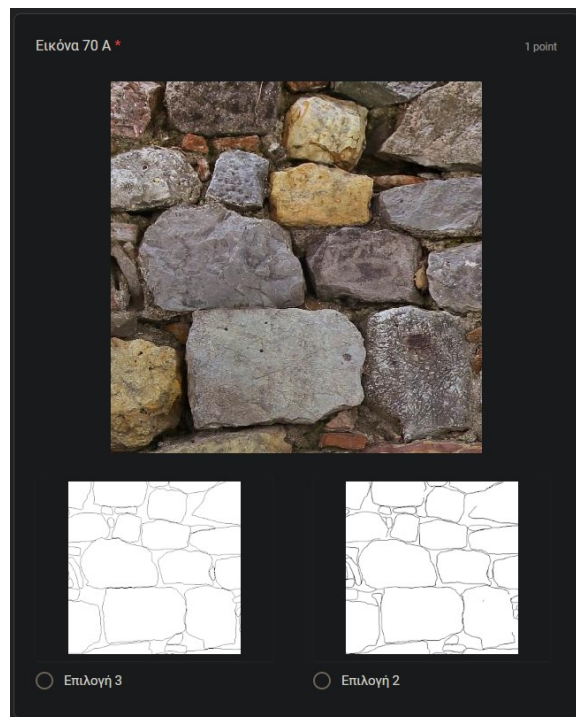
5. Αξιολόγηση

Για την καλύτερη εξαγωγή συμπερασμάτων, θεωρήθηκε απαραίτητη η στατιστική επεξεργασία ερωτηματολογίου αξιολόγησης, από μεγάλο αριθμό ατόμων. Γι' αυτό δημιουργήθηκε ερωτηματολόγιο, με σκοπό την αξιολόγηση από άτομα μη γνώστες. Παράλληλα, αποφασίστηκε να επιδειχθούν τα αποτελέσματα σε ειδικευμένους επιστήμονες, ώστε να εξαχθούν και πρόσθετα συμπεράσματα.

5.1 Ερωτηματολόγιο

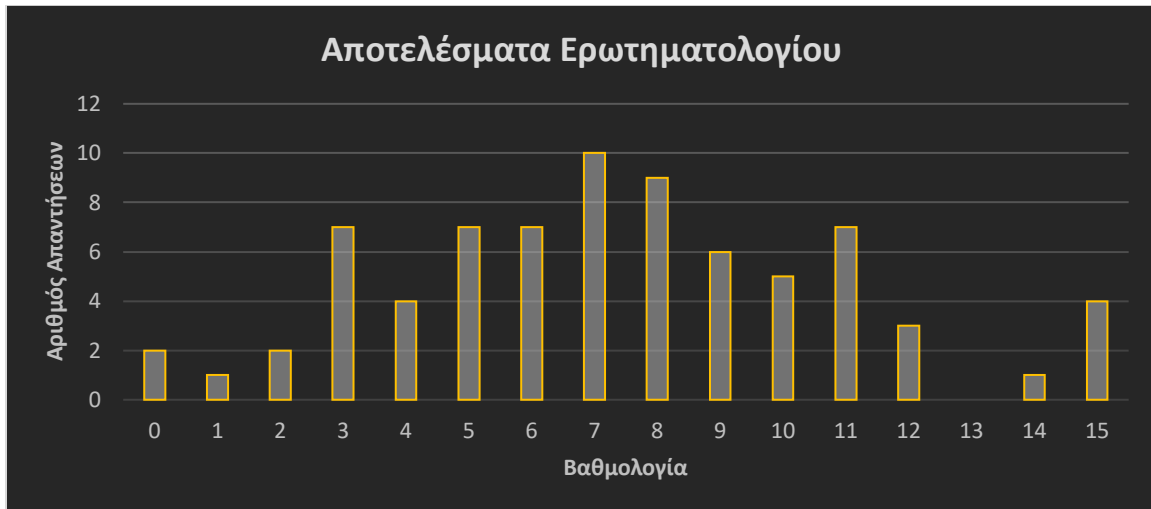
Για την δημιουργία του ερωτηματολογίου χρησιμοποιήθηκε η πλατφόρμα Google Forms, που προσφέρει εύκολη και γρήγορη δημιουργία ερωτηματολογίων. Η δομή του ερωτηματολογίου είναι απλή, αποτελείται από 15 ερωτήσεις όπου κάθε μία δίνει το πραγματικό στιγμιότυπο ενός λίθινου τείχους και από κάτω ζητείται να γίνει η επιλογή ενός από τα 2 στιγμιότυπα του ίδιου τμήματος όπου εμφανίζονται μόνο οι ακμές τους. Σε αυτές τις 2 επιλογές είναι σε τυχαία κατανομή το απόσπασμα που δημιουργήθηκε από τον επαγγελματία και το αντίστοιχο συνθετικό απόσπασμα που παρήχθη από τη ΑΙ.

Επομένως, ο ερωτώμενος καλείται να διαλέξει από 15 διαφορετικές περιπτώσεις, ποια απεικόνιση είναι πιο ρεαλιστική, όπου κάθε επιλογή που αντιστοιχεί με το απόσπασμα του επαγγελματία, δίνει από ένα πόντο. Άρα όσο κοντινότερη η βαθμολογία στη τιμή 0, τόσο πιο ρεαλιστικές φάνηκαν στον ερωτώμενο οι δημιουργημένες εικόνες της ΑΙ. Αντίστοιχα, όσο κοντινότερη η βαθμολογία στη τιμή 15, τόσο πιο ρεαλιστικές φάνηκαν στον ερωτώμενο τα τμήματα από τον επαγγελματία.



Εικόνα 18: Παράδειγμα μίας ερώτησης από το ερωτηματολόγιο

Μετά την ολοκλήρωση του χρονικού διαστήματος των 2 εβδομάδων έγινε η ολοκλήρωση και συλλογή 75 ερωτηματολογίων. Το εύρος βαθμολογιών που συλλέχθηκε καλύπτει όλες τις τιμές από 0/15 έως 15/15, με μέσο όρο το 7.36.



Σχήμα 03: Κατανομή βαθμολογίας

Με βάση τα αποτελέσματα που συλλέχθηκαν παρατηρείται μια γενική σύγχυση μεταξύ διάκρισης πραγματικής από συνθετικής εικόνας. Φαινόμενο που μαρτυράει την ομοιότητα των συνθετικών εικόνων με τις πραγματικές. Παράλληλα παρατηρούνται και μερικές περιπτώσεις πλήρως επιτυχούς διάκρισης αυτών. Αυτό δείχνει πως οι συνθετικές εικόνες δεν είναι ακόμα τέλειες και χρειάζονται βελτίωση. Αντίστοιχα όμως εμφανίζονται και περιπτώσεις πλήρως ανεπιτυχούς διάκρισης αυτών. Όλα αυτά μαρτυρούν πως τα συνθετικά αποτελέσματα της AI είναι ικανοποιητικά σε έναν βαθμό για την περαιτέρω χρήση τους.

Αξίζει να σημειωθεί ότι σε 2 ερωτήματα επικράτησε η επιλογή της συνθετικής εικόνας κατά 70.7% (53 άτομα). Στο 1^ο ερώτημα σε μεγάλο βαθμό δικαιολογείται μιας και η συνθετική εικόνα διαχώριζε ήδη μια ραγισμένη πέτρα και εντοπίζει μικρότερες πέτρες που στο απόσπασμα του αρχιτέκτονα δεν εμφανίζονται. Στο 2^ο ερώτημα ευθύνεται η κεντρική πέτρα όπου καλύπτεται από σοβά και ενώ ο αρχιτέκτονας μπόρεσε εμπειρικά και οριοθέτησε την πέτρα. Ο ανειδίκευτος άνθρωπος μπερδεύεται και το θεωρεί όλο μία πέτρα.



Εικόνα 19: Πρώτο ερωτήματα που υπερίσχυσε η συνθετική εικόνα



Εικόνα 20: Δεύτερο ερωτήματα που υπερίσχυσε η συνθετική εικόνα

5.2 Αξιολόγηση από Ειδικό

Παράλληλα με την συμπλήρωση του ερωτηματολογίου από το ευρύ κοινό, δόθηκε για συμπλήρωση και σε δύο ειδικευμένους επιστήμονες. Μία από αυτούς, ήταν η αρχιτέκτονας που δημιούργησε τα δεδομένα εκπαίδευσης που χρησιμοποιήθηκαν σε αυτή την εργασία. Μετά την συμπλήρωση του ερωτηματολογίου ακολούθησε μια συζήτηση με την αρχιτέκτονα με στόχο τη καταγραφή των σχολίων της πάνω στα αποτελέσματα και το όλο εγχείρημα αυτής της εργασίας.

Τα αποτελέσματα από τη συμπλήρωση του ερωτηματολογίου ήταν προβλέψιμα. Η αρχιτεκτόνισσα, που ήταν υπεύθυνη και για τα δεδομένα που χρησιμοποιήθηκαν, είχε πλήρη επιτυχία με βαθμολογία 15/15. Αντίστοιχα ο δεύτερος ειδικευμένος επιστήμονας είχε βαθμολογία 11/15, βαθμολογία που είναι πάνω από το μέσο όρο (7). Επομένως μέσα από τις βαθμολογίες, συμπεραίνεται ότι οι συνθετικές εικόνες δεν κατάφεραν να μπερδέψουν ειδικευμένα άτομα του τομέα.

Μέσα από την συζήτηση που ακολούθησε με την αρχιτέκτονα καταγράφηκαν θετικά σχόλια και συμπεράσματα. Αναφέρθηκε ότι οι συνθετικές εικόνες είναι αρκετά καλές/πειστικές, με αποτέλεσμα να κάνουν δύσκολη την διαχώριση μεταξύ αληθινής από

συνθετικής εικόνας. Επίσης κατά την διάρκεια συμπλήρωσης του ερωτηματολογίου, παρατηρήθηκε πως το βασικό στοιχείο που έδειχνε τη διαφοροποίηση των δύο εικόνων δεν ήταν το κομμάτι της απεικόνισης των ακμών, αλλά η οριοθέτηση “κρυμμένων ακμών πίσω από εμπόδια (σοβάς). Τέλος για παρόμοια μελλοντικά έργα, η χρήση αυτής της διαδικασίας εντοπισμού ακμών θεωρήθηκε αρκετά βοηθητική. Οι λόγοι είναι ότι προσφέρει αποδεκτά αποτελέσματα σε επαναλαμβανόμενα τμήματα που δεν απαιτούν ιδιαίτερη προσοχή και ότι προσφέρει παραπάνω πληροφορίες (ραγίσματα, φθορές ή σκαλίσματα), που πολλές φορές είναι και το ζητούμενο των έργων.

6. Συμπεράσματα

Ο σκοπός της έρευνας ήταν ο εντοπισμός ακμών σε ορθοφωτογραφίες λιθοδομών και η αυτόματη δημιουργία διανυσματικών αρχείων αυτών. Για να επιτευχθεί αυτό χρησιμοποιήθηκε το μοντέλο rix2rix (χρησιμοποιείτε στην τεχνολογία τεχνητής νοημοσύνης), η γνωστή βιβλιοθήκη γεωχωρικής επεξεργασίας GDAL και η βιβλιοθήκη OpenCV επεξεργασίας εικόνων.

Το κομμάτι μετατροπής από ορθοφωτογραφία λιθόχτιστης κατασκευής σε εικόνα ακμών το ανέλαβε ένα νευρωνικό δίκτυο. Συγκεκριμένα το νευρωνικό δίκτυο καλείται να λαμβάνει μία εικόνα από λιθόχτιστο τμήμα κτίσματος και να κατασκευάζει μια δική του εικόνα με εμφανισμένες τις ακμές των λίθων. Για την εύρεση του βέλτιστου νευρωνικού δικτύου έγιναν 76 διαφορετικά πειράματα. Ύστερα, στις συνθετικές εικόνες της AI, εφαρμόστηκε μια σειρά φίλτρων με σκοπό τη συμπλήρωση μικρών κενών κατά το μήκος των γραμμών. Τέλος το κομμάτι μετατροπής από εικόνα ακμών σε διανυσματικό αρχείο έγινε με την χρήση των εργαλείων της βιβλιοθήκης GDAL.

Για την πλήρη αξιολόγηση των συνθετικών εικόνων της AI, συλλέχθηκε η γνώμη του ευρέος κοινού και η εκτίμηση από έναν ειδικό επαγγελματία πάνω στο τομέα. Στο ευρύ κοινό δόθηκε ένα ερωτηματολόγιο 15 ερωτήσεων, που οι ερωτώμενοι καλούνταν να διακρίνουν ανάμεσα στην πραγματική εικόνα ακμών από την συνθετική. Με τον ειδικό επαγγελματία ακολούθησε μια επικοινωνία και συζητήθηκε το εγχείρημα αυτής της εργασίας όπως και η εκτίμησή του για τα αποτελέσματά της. Από τα αποτελέσματα του ερωτηματολογίου, το ευρύ κοινό δεν κατάφερε την πλήρη διάκριση μεταξύ συνθετικής με πραγματικής εικόνας, κάτι που μαρτυρά μια σχετική επιτυχία στο στόχο αυτής της διπλωματικής. Ο επαγγελματίας αντίθετα κατάφερε και διέκρινε τις συνθετικές από τις πραγματικές εικόνες, όμως έκρινε ότι τα συνθετικά αποτελέσματα ήταν πειστικά και εμφάνιζαν μια σταθερή δυσκολία στον εντοπισμό τους. Επίσης σχολίασε ότι μια διαδικασία, όπως αυτή που πραγματεύεται αυτή η διπλωματική, θα ήταν αρκετά χρήσιμη μιας και μπορεί να γλυτώσει χρόνο εντοπίζοντας τις ακμές σε επαναλαμβανόμενα απλά τμήματα που δεν απαιτούν ιδιαίτερη προσοχή.

Ο στόχος της εργασίας αυτής έχει επιτευχθεί σε ένα καλό βαθμό. Όντως έχει δημιουργηθεί ένα νευρωνικό δίκτυο που δίνοντάς του μία εικόνα λιθοδομής, μπορεί και κατασκευάζει σε δευτερόλεπτα, μια συνθετική εικόνα των ακμών που εμφανίζονται. Επίσης με την εφαρμογή της συγκεκριμένης διαδικασίας των φίλτρων, οι σύνθετες εικόνες πλησιάζουν ακόμα περισσότερο το επιθυμητό αποτέλεσμα. Τέλος με την χρήση της GDAL, η μετατροπή από εικονικό δεδομένο σε διανυσματικό γίνεται εξίσου γρήγορα και με άμεσα αποτελέσματα. Άρα όντως από μία ορθοφωτογραφία λιθοδομής, αυτόματα και ταχύτατα φτάνουμε στο επιθυμητό προϊόν των ακμών σε διανυσματική μορφή.

Παρόλα αυτά όμως, υπάρχουν κάποιες μικρές αστοχίες. Πρώτη είναι η αδυναμία της δημιουργίας ενός μοναδικού νευρωνικού δικτύου και την εφαρμογή του σε όλα τα είδη λίθινων κατασκευών. Αυτό οφείλεται στη περιορισμένη διαφοροποίηση κατηγοριών λίθινων κατασκευών στα δεδομένα που προσφέρθηκαν στο πλαίσιο αυτής της διπλωματικής. Δεύτερη αστοχία είναι ότι ο εντοπισμός μόνο των εξωτερικών ακμών λίθων και αγνόηση περαιτέρω στοιχείων όπως φωτοσταθερά και κενά δεν έγινε πλήρως, αντίθετα σε αρκετές περιπτώσεις εντοπίστηκαν και αναδείχθηκαν οι ακμές αυτών όπως και ρωγμές της ίδιας της επιφάνειας των λίθων. Αστοχία που μπορεί να θεωρηθεί και θετική, μιας και παράγεται περισσότερη επεξεργάσιμη πληροφορία που ίσως ένας άνθρωπος να αγνοούσε. Τέλος παρά την εφαρμογή φίλτρων και της βελτιστοποίησης του νευρωνικού δικτύου, σε αρκετές περιπτώσεις κάποιες ακμές δεν εντοπίστηκαν, αρκετά

μικρά κενά ασυνέχειας παρέμειναν και σε πολύ μικρά τμήματα δεν βρέθηκαν τα σωστά όρια.

Λόγω της φύσης μιας διπλωματικής έγιναν κάποιες παραδοχές. Αρχικά λόγω των λίγων δεδομένων, υπάρχει η πιθανότητα κατά τη διάρκεια βελτιστοποίησης του νευρωνικού δικτύου να πάρθηκε λάθος απόφαση επιλογής παραμέτρου, που με την ύπαρξη περισσότερων δεδομένων ίσως να είχε φανεί και να είχε αποφευχθεί. Επίσης το πλήθος των δοκιμών θα μπορούσε να ήταν μεγαλύτερος και να κάλυπτε μεγαλύτερη γκάμα παραμέτρων, όμως κάτι τέτοιο θα ξέφευγε χρονικά από το πλαίσιο μιας διπλωματικής. Τέλος η αξιολόγηση με ερωτηματολόγιο προς ευρύ κοινό περιορίζεται στο κύκλο του ίδιου του φοιτητή και δεν εμπεριέχει πληθώρα διαφοροποίηση μεταξύ ερωτώμενων.

Η διαδικασία που πραγματεύεται η διπλωματική αυτή δέχεται και παραπάνω βελτίωση. Ήδη μετά από την ολοκλήρωση της ίδιας της εργασίας, προέκυψαν αρκετές ιδέες που ίσως μπορέσουν να βελτιώσουν το αποτέλεσμα. Για παράδειγμα μετά την δημιουργία των διανυσματικών αρχείων, λόγω της διαδικασίας που ακολουθεί η συνάρτηση Polygonize της GDAL, δημιουργούνται σκαλοπάτια γραμμών μεγέθους pixel. Με σκοπό την εξομάλυνση αυτών, η χρήση μιας αντίστοιχης συνάρτησης που βρίσκει την βέλτιστη ευθεία που διέρχεται από τις κορυφές των “σκαλιών” ίσως βοηθήσει. Επιπλέον μπορεί να βελτιστοποιηθεί κι άλλο η παραμετροποίηση ή η ίδια η δομή του νευρωνικού δικτύου με την παρέμβαση σε επίπεδο κώδικα στην rix2rix και όχι με απλή χρήση έτοιμων επιλογών. Αντίστοιχα, μετά την παραγωγή των σύνθετων εικόνων από το νευρωνικό δίκτυο, μπορεί να δίνεται η δυνατότητα επέμβασης σε αυτές, μέσω κάποιου προγράμματος επεξεργασίας εικόνων, ώστε όσες ακμές δεν είναι επιθυμητές να γίνει μια απλή αφαίρεση αυτών σε επίπεδο pixel. Επίσης και η εφαρμογή εντελώς διαφορετικής σειράς φίλτρων ή συναρτήσεων στο αποτέλεσμα του νευρωνικού να επιφέρει καλύτερα τελικά προϊόντα. Τέλος, ακόμα και η αλλαγή του βασικής ιδέας από «Εντοπισμός Ακμών Λίθινων Κτισμάτων» σε «Εντοπισμό Αρμών Λίθινων Κτισμάτων» ίσως επιφέρει μεθόδους καλύτερης προσέγγισης του ζητήματος.

Στο κομμάτι της μετάβασης από το θεωρητικό/ερευνητικό τμήμα στη εφαρμογή στο πεδίο υπάρχουν εμπόδια. Η διαδικασία αυτή πρακτικά δεν μπορεί να χρησιμοποιηθεί από οποιονδήποτε. Απαιτείται ελαφριά γνώση python από τον χρήστη και έναν καλό υπολογιστικό εξοπλισμό. Επίσης το συγκεκριμένο νευρωνικό δίκτυο εμφανίζει ένα περιορισμό μεγέθους εικόνας (1024x1024 pixel) όπως και ένα περιορισμό ανάλυσης της πληροφορίας (ο μικρότερος αρμός να απεικονίζεται το ελάχιστο με 3 εικονοστοιχεία). Μελλοντικά μπορεί η διαδικασία αυτή να γίνει ένα κουμπί μιας εφαρμογής, όπου δίνονται εικόνες λίθων κι αυτό εντοπίζει ακμές, όμως απαιτεί αρκετή δουλειά που ξεπερνάει για άλλη μια φορά το όριο μιας διπλωματικής. Τέλος με την βοήθεια του ήδη υπάρχοντος νευρωνικού δικτύου μπορεί να γίνει μια διαφορετική χρήση του. Αντί να του απαιτείται η πλήρως αυτοματοποιημένη αναγνώριση ακμών ολόκληρου του έργου, να χρησιμοποιείτε ως βοηθητικό εργαλείο της ήδη υπάρχουσας διαδικασίας. Συγκεκριμένα, σε κάθε έργο, να γίνεται η ήδη γνωστή διαδικασία ψηφιοποίησης των ακμών στο 1/5 της έκτασής του. Αυτό το 1/5 να δίνεται στο νευρωνικό δίκτυο για μετεκπαίδευση και αυτό με τη σειρά του να παράγει τα υπόλοιπα 4/5 του έργου. Ουσιαστικά, σε κάθε έργο να γίνεται μια εξειδίκευση του νευρωνικού δικτύου σε αυτό.

Φτάνοντας στο τέλος, με σιγουριά γίνεται η επαλήθευση ότι η χρήση της AI μπορεί και έχει δυνατότητες να βοηθήσει έως και να απαλλάξει στο μέλλον τον μηχανικό από χρονοβόρες και δαπανηρές διαδικασίες. Απόδειξη αυτού, η παρούσα διπλωματική εργασία που αν και δεν δημιούργησε ένα νέο εργαλείο για το μηχανικό, έδειξε ότι η AI μπορεί να χρησιμοποιηθεί και να βγάλει αξιόλογα αποτελέσματα.

7. Βιβλιογραφία

[Baştan M., Bukhari S.S., and Breuel T., \(2017\), Active Canny: edge detection and recovery with open active contour models. IET Image Processing, 11: 1325-1332. <https://doi.org/10.1049/iet-ipr.2017.0336>](#)

[Canny J., A computational approach to edge detection, IEEE Trans. Pattern Anal. Mach. Intell. 8 \(1986\) 679–698](#)

[Chaubey A. "Downsampling and Upsampling of Images — Demystifying the Theory", 2020, Analytics Vidhya \[Τελευταία Πρόσβαση 07/09/2022\]](#)

[Chi Z., Li H., Lu H. and Yang H., "Dual Deep Network for Visual Tracking," in IEEE Transactions on Image Processing, vol. 26, no. 4, pp. 2005-2015, April 2017, doi: \[10.1109/TIP.2017.2669880\]\(https://doi.org/10.1109/TIP.2017.2669880\)](#)

[Dollár P., Zitnick C.L., Fast edge detection using structured forests, IEEE Trans. Pattern Anal. Mach. Intell. 37 \(2014\) 1558–1570](#)

[Dong-Su K., Wang-Heon L., In-So K., "Automatic edge detection using 3×3 ideal binary pixel patterns and fuzzy-based edge thresholding", Pattern Recognition Letters, Vol 25, Issue 1, 2004, p. 101-106, ISSN 0167-8655](#)

[Ganin G., Lempitsky V. "N4-Fields: Neural Network Nearest Neighbor Fields for Image Transforms", in Computer Vision and Pattern Recognition, 2014, arXiv:1406.6558](#)

[Ghosal S., Mehrotra R., "Orthogonal moment operators for subpixel edge detection", Pattern Recognition, vol 26, Issue 2, 1993, p. 295-306, ISSN 0031-3203](#)

[He J., Zhang S., Yang M., Shan Y. and Huang T., "BDCN: Bi-Directional Cascade Network for Perceptual Edge Detection," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 1, pp. 100-113, 1 Jan. 2022, doi: \[10.1109/TPAMI.2020.3007074\]\(https://doi.org/10.1109/TPAMI.2020.3007074\)](#)

[Junfeng J., Shenjuan L., Gang W., Weichuan Z., Changming S., "Recent advances on image edge detection: A comprehensive review", in Neurocomputing, vol 503, 2022, Pages 259-271, <https://doi.org/10.1016/j.neucom.2022.06.083>](#)

[Konishi S., Yuille A.L., Coughlan J.M. and Song Chun Zhu, "Statistical edge detection: learning and evaluating edge cues," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, no. 1, pp. 57-74, Jan. 2003, doi: \[10.1109/TPAMI.2003.1159946\]\(https://doi.org/10.1109/TPAMI.2003.1159946\)](#)

[Lamba H., "Understanding Semantic Segmentation with UNET", 2019, Towards Data Science \[Τελευταία Πρόσβαση 07/09/2022\]](#)

[Li Y., Paluri M., Rehg J.M. and Dollár P., "Unsupervised Learning of Edges," 2016 IEEE Conference on Computer Vision and Pattern Recognition \(CVPR\), 2016, pp. 1619-1627, doi: \[10.1109/CVPR.2016.179\]\(https://doi.org/10.1109/CVPR.2016.179\)](#)

[Liu Y. et al., "Richer Convolutional Features for Edge Detection," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 8, pp. 1939-1946, 1 Aug. 2019, doi: \[10.1109/TPAMI.2018.2878849\]\(https://doi.org/10.1109/TPAMI.2018.2878849\)](#)

[Marr D., Hildreth E., Theory of edge detection, Proc. R. Soc. London. Series B. Biol. Sci. 207 \(1980\) 187–217](#)

[Mohan K., Seal A., Krejcar O. and Yazidi A., "Facial Expression Recognition Using Local Gravitational Force Descriptor-Based Deep Convolution Neural Networks," in IEEE](#)

[Transactions on Instrumentation and Measurement, vol. 70, pp. 1-12, 2021, Art no. 5003512, doi: 10.1109/TIM.2020.3031835](#)

[Perona P. and Malik J., "Scale-space and edge detection using anisotropic diffusion," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12, no. 7, pp. 629-639, July 1990, doi: 10.1109/34.56205](#)

[Qiu B. , Guo J., Kraeima J., Glas HH, Zhang W., Borra RJH, MJH Witjes , van Ooijen PMA. Recurrent Convolutional Neural Networks for 3D Mandible Segmentation in Computed Tomography. J Pers Med. 2021 May 31;11\(6\):492. doi: 10.3390/jpm11060492](#)

[Rahebi J. and Tajik, H.R. \(2011\). Biomedical Image Edge Detection using an Ant Colony Optimization Based on Artificial Neural Networks](#)

[Rakesh R.R., Chaudhuri P. and Murthy C.A., "Thresholding in edge detection: a statistical approach," in IEEE Transactions on Image Processing, vol. 13, no. 7, pp. 927-936, July 2004, doi: 10.1109/TIP.2004.828404](#)

[Russo F., "Edge detection in noisy images using fuzzy reasoning," in IEEE Transactions on Instrumentation and Measurement, vol. 47, no. 5, pp. 1102-1105, Oct. 1998, doi: 10.1109/19.746564](#)

[Seo S., "Subpixel Edge Localization Based on Adaptive Weighting of Gradients," in IEEE Transactions on Image Processing, vol. 27, no. 11, pp. 5501-5513, Nov. 2018, doi: 10.1109/TIP.2018.2860241](#)

[Sobel I.E., Camera models and machine perception, Stanford University, 1970 J. Jing, S. Liu, G. Wang et al. Neurocomputing 503 \(2022\) 259–271](#)

[Sobel Operator, Wikipedia \[Τελευταία Πρόσβαση 07/09/2022\]](#)

[Trujillo-Pino A., Krissian K., Alemán-Flores M., Santana-Cedrés D., "Accurate subpixel edge location based on partial area effect", Image and Vision Computing, Vol 31, Issue 1, 2013, p. 72-90, ISSN 0262-8856](#)

[Versaci M., Morabito, F.C. Image Edge Detection: A New Approach Based on Fuzzy Entropy and Fuzzy Divergence. Int. J. Fuzzy Syst. 23, 918–936 \(2021\). <https://doi.org/10.1007/s40815-020-01030-5>](#)

[Wang D., Li C., Song H., Xiong H., Liu C. and He D., "Deep Learning Approach for Apple Edge Detection to Remotely Monitor Apple Growth in Orchards," in IEEE Access, vol. 8, pp. 26911-26925, 2020, doi: 10.1109/ACCESS.2020.2971524](#)

[Xia X., Kulis B., 2017. W-Net: A deep model for fully unsupervised image segmentation. arXiv preprint arXiv:171.85.](#)

[Xiaofeng R., Bo L., Discriminatively trained sparse code gradients for contour detection, in: Advances in Neural Information Processing Systems 25, 2012](#)

[Xie S. and Tu Z., "Holistically-Nested Edge Detection," 2015 IEEE International Conference on Computer Vision \(ICCV\), 2015, pp. 1395-1403, doi: 10.1109/ICCV.2015.164](#)

[Yang J., Price B., Cohen S., Lee H., Yang M.H., Object contour detection with a fully convolutional encoder-decoder network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 193–202](#)

[Yi S., Labate D., Easley G.R. and Krim H., "A Shearlet Approach to Edge Analysis and Detection," in IEEE Transactions on Image Processing, vol. 18, no. 5, pp. 929-941, May 2009, doi: 10.1109/TIP.2009.2013082](#)

[Zhu J-Y, Isola P., Zhou T., Efros A., "Image-to-Image Translation with Conditional Adversarial Networks", in Computer Vision and Pattern Recognition \(CVPR\), 2017 IEEE Conference, 2017](#)

Ιστότοποι

[AutoCAD Software](#)

[Colaboratory](#)

[GDAL](#)

[OpenCV](#)

[Personal Cloud Storage & File Sharing Platform](#)

[Weights & Biases](#)

Παράρτημα Α

Πίνακας Πινάκων

[Πίνακας 01](#): Κάρτες γραφικών που χρησιμοποιήθηκαν

[Πίνακας 02](#): Παρουσίαση φάσεων με αντίστοιχες μεταβλητές

[Πίνακας 03](#): Κατανομή των δεδομένων στο 2^ο κύκλο πειραμάτων

[Πίνακας 04](#): Ενδιαφέροντα αποτελέσματα στο 1^ο Κύκλο Πειραμάτων

[Πίνακας 05](#): Καταγραφών σε απαιτήσεις και τελικών τιμών για τα πειράματα του 1^{ου} κύκλου

Πίνακας Σχημάτων

[Σχήμα 01](#): Διάγραμμα Generator με αρχιτεκτονική U-Net με εικόνα input και output

[Σχήμα 02](#): Γενικό Σχεδιάγραμμα Εκπαίδευσης

[Σχήμα 03](#): Κατανομή Βαθμολογίας

[Σχήμα A.01](#): Απλό Σχεδιάγραμμα από Νευρωνικό Δίκτυο με 2 κρυμμένα layer

[Σχήμα A.02](#): Σχεδιάγραμμα Λειτουργίας U-Net 64 Φίλτρων

Πίνακας Εικόνων

[Εικόνα 01](#): Χρονοδιάγραμμα εμφάνισης μεθόδων εντοπισμού ακμών ([Junfen et al., 2022](#))

[Εικόνα 02](#): Παράδειγμα εφαρμογής Sobel ([Wikipedia, 2022](#))

[Εικόνα 03](#): Παράδειγμα από εφαρμογή ενεργού περιγράμματος ([Bastan et al., 2017](#))

[Εικόνα 04](#): Παράδειγμα από εφαρμογή βασισμένη στη ασαφή λογική ([Wang et al., 2004](#))

[Εικόνα 05](#): Διάγραμμα λειτουργίας μη επιβλεπόμενου συστήματος εκμάθησης ([Junfen et al., 2022](#))

[Εικόνα 06](#): Διάγραμμα λειτουργίας επιβλεπόμενου συστήματος εκμάθησης ([Junfen et al., 2022](#))

[Εικόνα 07](#): Πληροφορία που δεν πρέπει να κρατηθεί: α) στόχοι φωτοσταθερών, β) εξωτερικά όρια θέματος, γ) βλάστηση, δ) σκαλίσματα, ε) σκαλότρυπες

[Εικόνα 08](#): Παράδειγμα α) Αρχικής Πραγματικής Εικόνας β) Τμηματοποίησης Αρχικής Εικόνας γ) Διανυσματική Απεικόνιση δ) Τμηματοποίηση Διανυσματικής Απεικόνισης (με σειρά αντιστοιχίας από πάνω προς τα κάτω)

[Εικόνα 09](#): Διάγραμμα Ενοποίησης Εικόνων

[Εικόνα 10](#): Παράδειγμα νέας εικόνας α) και νέου ζεύγος τμημάτων αυτής β)

[Εικόνα 11](#): Παράδειγμα εφαρμογής των φίλτρων α) αρχική εικόνα β) εφαρμογή φίλτρο Gaussian γ) εφαρμογή φίλτρο Threshold

[Εικόνα 12](#): Παράδειγμα εικόνας πριν και μετά την εφαρμογή του Dilation φίλτρου

[Εικόνα 13](#): Παράδειγμα εικόνας πριν και μετά την εφαρμογή της συνάρτησης Polygonize

[Εικόνα 14](#): Παράδειγμα ζαρωμένων ακμών σε RGB (1^ο Πείραμα)

[Εικόνα 15](#): Παράδειγμα μαύρων κουκίδων με χρήση Unet αρχιτεκτονικής (2^ο Πείραμα)

[Εικόνα 16](#): Παράδειγμα καλού αποτελέσματος – κακού διαγράμματος Loss με χρήση xavier ως initialization network (9^ο Πείραμα)

[Εικόνα 17](#): Παράδειγμα εικόνας αποτελέσματος με wganr με αντίστοιχο Loss διάγραμμα (31^ο Πείραμα)

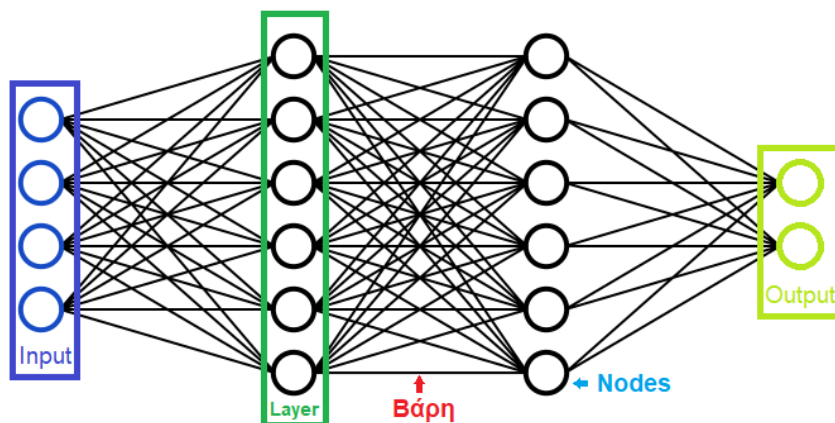
[Εικόνα 18](#): Παράδειγμα μίας ερώτησης από το ερωτηματολόγιο

[Εικόνα 19](#): Πρώτο ερωτήματα που υπερίσχυσε η συνθετική εικόνα

[Εικόνα 20](#): Δεύτερο ερωτήματα που υπερίσχυσε η συνθετική εικόνα

Ορολογίες

- **Τεχνητή Νοημοσύνη**: η προσπάθεια του ανθρώπου να φτιάξει ένα υπολογιστικό σύστημα που να μην ακολουθεί έναν αλγόριθμο, αλλά να μπορεί με την κατάλληλη μάθηση να προσαρμόζεται και να επιλύει ζητήματα όπως ένας άνθρωπος με μεγαλύτερη ακρίβεια και ταχύτητα
- **Όραση Υπολογιστή**: Το επιστημονικό πεδίο όπου ασχολείται με το πώς μπορούν οι υπολογιστές να αποκτήσουν υψηλού επιπέδου κατανόηση από ψηφιακά δεδομένα.
- **Νευρωνικά δίκτυα**: είναι μια μέθοδο μηχανικής μάθησης της τεχνητής νοημοσύνης. Έχει πολύ καλά αποτελέσματα στον εντοπισμό μοτίβων και η λειτουργία τους θυμίζει τους νευρώνες των έμβιων όντων. Συγκεκριμένα, αποτελείται από το layer εισόδου (nodes με την αρχική πληροφορία), τα εσωτερικά κρυμμένα layer, το layer εξόδου (nodes με την τελική πληροφορία) και τις συνδέσεις μεταξύ αυτών (με το κάθε ένα να έχει τη δική του τιμή και το δικό του βάρος)

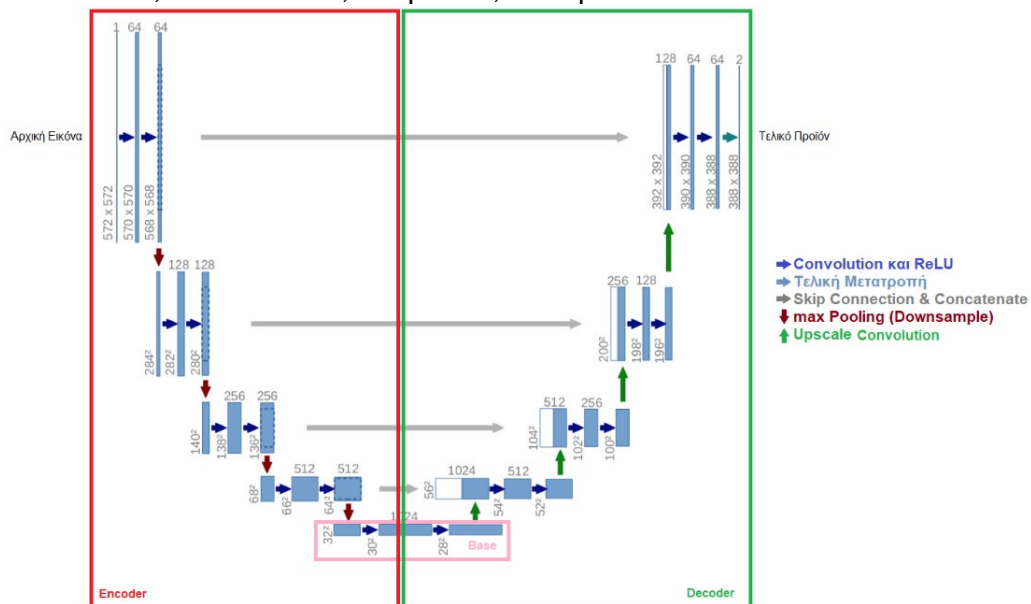


Σχήμα A.01: Απλό Σχεδιάγραμμα από Νευρωνικό Δίκτυο με 2 κρυμμένα layer

- **Βάρη**: είναι τα γνωστά βάρη της στατιστικής. Κάθε τιμή που εμπεριέχετε στις συνδέσεις μεταξύ των nodes έχει ένα βάρος. Η τελικές τιμές των βαρών προκύπτουν μετά από την τελική επανάληψη του κύκλου εκπαίδευσης, όπου σε

κάθε προηγούμενη επανάληψη αυτής, οι τιμές τους διαμορφώνονται με βάση την τιμή του Loss Function

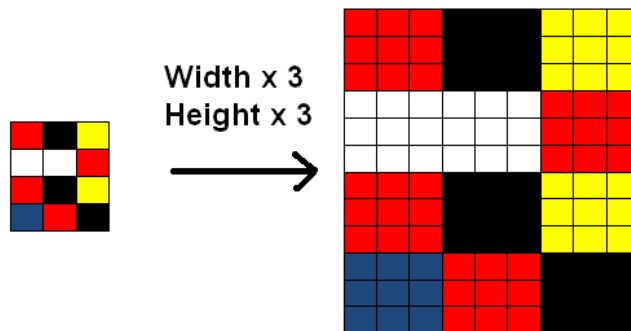
- Nodes (ή units ή neurons): η αρχική ή τελική ή ενδιάμεση πληροφορία που προκύπτει από την εικόνα όταν δέχεται ένα φίλτρο με συγκεκριμένο βάρος
- Επίπεδο (Layer): ονομάζεται ένα σύνολο από nodes, συναντιέται στο αρχικό, στο τελικό layer και στα εσωτερικά κρυφά layer του νευρωνικού δικτύου
- Conditional Adversarial Networks (cGAN): είναι μία κατηγορία νευρωνικών δικτύων που χρησιμοποιεί η βιβλιοθήκη pix2pix, αποτελείται από τον Generator και τον Discriminator. Ο Generator είναι το νευρωνικό δίκτυο που εκπαιδεύεται να προβλέπει και να παράγει ένα επιθυμητό αποτέλεσμα με βάση μία αρχική εικόνα. Ο Discriminator είναι το νευρωνικό δίκτυο που εκπαιδεύεται να διαχωρίζει μεταξύ αληθινής αρχικής και επιθυμητής εικόνας, όπως και αληθινής με παραγόμενης εικόνας από το Generator. Οι έτοιμες επιλογές που υπήρχαν (ουσιαστικά διαφορετικές παραμέτρους δομής των Generator και Discriminator):
 - vanilla = χρησιμοποιεί ως loss function την cross-entropy
 - wgangp = χρησιμοποιεί ως loss function την Wasserstein φόρμουλα και ένα βαθμιαίο σύστημα πινών για να πετύχει την συνάρτηση Lipschitz συνέχειας
- Residual Block (ResNET): ήταν η πρώτη αρχιτεκτονική εκπαίδευσης που εισήγαγε την έννοια του skip connections. Δίνεται η δυνατότητα στο πρόγραμμα να αγνοήσει μια σειρά από layers και να επηρεάσει κατευθείαν ένα επόμενο output
- U-Net: αρχιτεκτονική εκπαίδευσης, όπου αποτελείται από το κομμάτι του encoder και του decoder με ίδιο αριθμό επαναλήψεων. Ως εργαλεία έχει τα ReLU, pooling, convolution, downscale, upscale, skip connection και concatenate



Σχήμα Α.02: Σχεδιάγραμμα Λειτουργίας U-Net 64 Φίλτρων

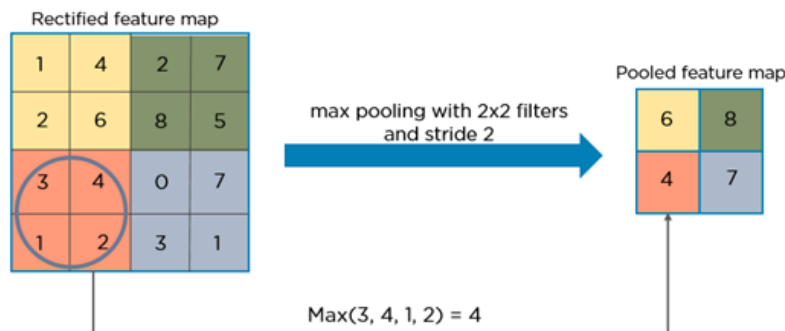
- Encoder: ονομασία της διαδικασίας όπου επαναλαμβάνονται οι διαδικασίες Convolution, BatchNorm, ReLU και downsampling (max pooling), με σκοπό να φτάσει μια εικόνα από την αρχική της ανάλυση στη βάση της

- Βάση (Base): το κατώτατο στάδιο του down-sampling, όπου μια εικόνα έχει τέτοιες διαστάσεις που δεν μπορεί να ξαναδεχτεί down-sample
- Decoder: ονομασία της διαδικασίας όπου επαναλαμβάνονται οι διαδικασίες upscale, convolution, BatchNorm και ReLU, με σκοπό να φτάσει μια εικόνα από την βάση της σε μια κατασκευασμένη από το δίκτυο με τις διαστάσεις της αρχικής
- Down-sample: η διαδικασία στην οποία μία εικόνα, με την χρήση διάφορων εργαλείων (pooling) μικραίνει σε διαστάσεις και σε πληροφορία
- Upscale: η διαδικασία όπου μία εικόνα, με την χρήση διάφορων εργαλείων (nearest neighbor/bilinear interpolation) αυξάνει σε διαστάσεις και πληροφορία (αλλά συνθετική)



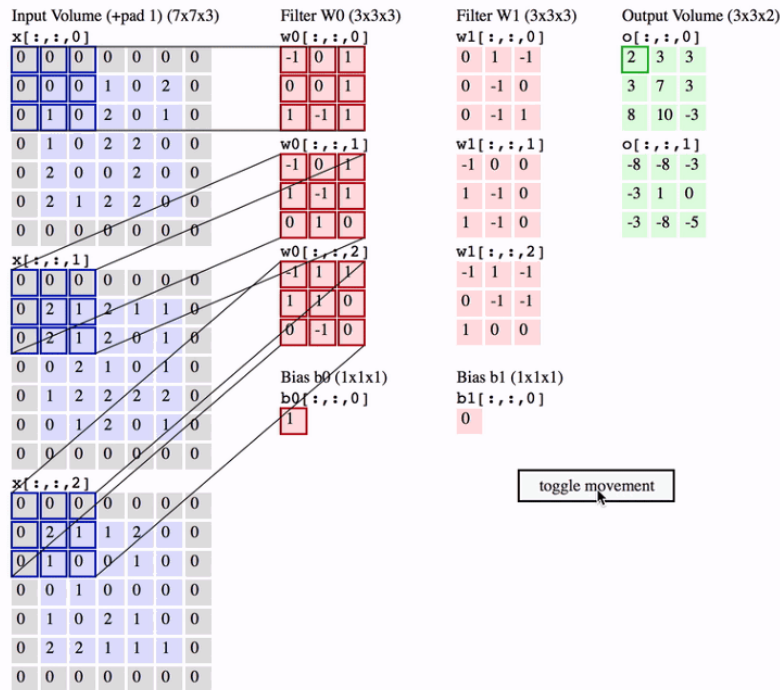
Εικόνα A.01: Upscale εικόνας από 4x3 σε 12x9 ([Chaubey, 2020](#))

- Pooling: δημιουργία νέων εικόνων μικρότερης ανάλυσης, όπου κάθε τιμή του pixel της νέας εικόνας, προκύπτει από μαθηματική επιλογή τιμών (max, min, average) ενός παραθύρου/πίνακα από τα pixel της αρχικής εικόνας



Εικόνα A.02: Παράδειγμα max pooling σε 2x2 παράθυρο ([Lamba, 2019](#))

- Συνέλιξη (Convolution): η εφαρμογή φίλτρων σε εικόνες με συγκεκριμένα βάρη



Εικόνα Α.03: Παράδειγμα Συνέλιξης (Lamba, 2019)

- Rectified Linear Unit (ReLU): συνάρτηση ενεργοποίησης νευρώνα, όπου εμφανίζεται σε αυτή τη περίπτωση ως ένα φίλτρο απογραμμικοποίησης των εικόνων. Συγκεκριμένα μηδενίζει όλες τις αρνητικές τιμές και μετατρέπει τις θετικές σε τιμές (0,1) που υπακούουν στη σιγμοειδή συνάρτηση
- Skip Connection και Concatenate: διαδικασία που συναντάμε στη αρχιτεκτονική εκπαίδευσης U-net και ResNET, τα οποία ουσιαστικά αγνοούν κάποια βήματα της διαδικασίας και ενώνουν την αρχική εικόνα με αυτή μετά τη επεξεργασία. Ο σκοπός αυτών είναι να βοηθήσουν στη χωρική οριοθέτηση των θεματικών κατηγοριών
- Network Initialization: η αρχικοποίηση του νευρωνικού δικτύου για να ξεκινήσει η διαδικασία. Επιλογές:
 - Xavier: τυχαίος αριθμός μεταξύ των τιμών $\pm \frac{1}{\sqrt{n}}$, όπου n είναι ο αριθμός των αρχείων εισόδων σε κάθε node
 - normal: τυχαίος αριθμός
 - kaing: αρχικοποίηση που παίρνει υπόψιν τη μη-γραμμικότητα των συναρτήσεων έναρξης με σκοπό να μην αυξάνει ή μειώνει τα σήματα εισόδου στα layer
 - orthogonal: αρχικές τιμές πινάκων που ανήκουν στο διάστημα [-1,1]. Χρήσιμη μέθοδος που αντιμετωπίζει το χάσιμο ή σπάσιμο των εναλλαγών ακμών στις εικόνες (αλλαγές ή τάσεις αλλαγής χρωμάτων)
- Dropout: μέθοδος όπου σε κάποια στάδια εκπαίδευσης τυχαία επιλέγονται και αγνοούνται κάποια nodes. Αυτό γίνεται ώστε το μοντέλο που εκπαιδεύεται να μην περιοριστεί και παγιδευτεί στην επίλυση μόνο των αρχικών δεδομένων αλλά να μπορεί να διαχειριστεί και άλλα δεδομένα

- Εποχή (Epoch): ο αριθμός που το δίκτυο είδε/πέρασε από μία φορά τα δεδομένα
- Adam: είναι αλγόριθμος βελτιστοποίησης, μετά από την ολοκλήρωση κάθε εποχής ενημερώνει τα βάρη των νευρωνικών δικτύων με σκοπό την ταχύτερη εύρεση τελικών παραμέτρων για την επίτευξη του στόχου
 - Adam momentum: η τάση του δείκτη Adam (χρησιμοποιεί τις τιμές από προηγούμενα και επόμενα αποτελέσματα για να δει πού να κινηθεί καλύτερα)
 - Initial Learning Rate for Adam: αρχική τιμή για τον ρυθμό εκπαίδευσης του Adam, καλύτερη από 0,002 – 0,003
- Learning rate policy: ο ρυθμός εκπαίδευσης, οι επιλογές που υπάρχουν είναι
 - Linear = αλλάζει γραμμικά το ρυθμό εκπαίδευσης μέχρι ο αριθμός των epoch να φτάσει σε μία προκαθορισμένη τιμή
 - Step = μειώνει το ρυθμό εκπαίδευσης κατά ένα βήμα (τιμή του gamma)
 - Plateau = αλλάζει το ρυθμό εκπαίδευσης με βάση τα στοιχεία απόδοσης
 - Cosine = κλειδώνεται ένα σταθερό ρυθμό εκπαίδευσης
- Loss Functions: ονομάζεται η συνάρτηση όπου με διάφορα μαθηματικά μοντέλα μπορεί και υπολογίζει επιβαρύνσεις τιμών ώστε να βελτιστοποιείται το νευρωνικό δίκτυο
 - PatchGAN: Είναι η αρχιτεκτονική σύγκρισης 2 εικόνων. Η λειτουργία της είναι η σύγκριση μικρών κομματιών κάθε εικόνας που βγάζει μια τιμή και μετά καταλήγει σε ένα Loss
 - Norm L1: κριτήριο που υπολογίζει την μέση θετική απόκλιση μεταξύ 2 πραγματικής και την παραγόμενη εικόνα