



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

**ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ
ΥΠΟΛΟΓΙΣΤΩΝ**

ΕΠΙΣΤΗΜΗ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

**Αξιολόγηση Γεννητικών Ανταγωνιστικών
Δικτύων για την Ανακατασκευή χωρίς Σύννεφα
Δορυφορικών Εικόνων Sentinel-2**

Μεταπτυχιακή Εργασία

Σιούκα Φωτεινή

Επιβλέπων Καθηγητής
Κ. Καραντζαλος, Αν. Καθηγητής, ΕΜΠ

Αθήνα, Οκτώβριος 2020



NATIONAL TECHNICAL UNIVERSITY OF ATHENS
SCHOOL OF ELECTRICAL AND COMPUTER
ENGINEERING
DATA SCIENCE AND MACHINE LEARNING

**Evaluating Generative Adversarial Networks
For Reconstructing without Clouds
Satellite Sentinel-2 Data**

Master Thesis

Siouka Foteini

Supervisor
Karantzalos Konstantinos, Assoc. Professor, NTUA

Athens, October 2020



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

**ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ
ΥΠΟΛΟΓΙΣΤΩΝ**

ΕΠΙΣΤΗΜΗ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

**Αξιολόγηση Γεννητικών Ανταγωνιστικών
Δικτύων για την Ανακατασκευή χωρίς Σύννεφα
Δορυφορικών Εικόνων Sentinel-2**

Μεταπτυχιακή Εργασία

Σιούκα Φωτεινή

Επιβλέπων Καθηγητής
Κ. Καράντζαλος, Αν. Καθηγητής, ΕΜΠ

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή στις 11 Νοεμβρίου 2020

.....
Κ. Καράντζαλος
Αν. Καθηγητής
Ε.Μ.Π.

.....
Γ. Στάμου
Αν. Καθηγητής
Ε.Μ.Π.

.....
Δ. Αργιαλάς
Καθηγητής
Ε.Μ.Π.

.....

Σιούκα Φωτεινή

Copyright © Φωτεινή Σιούκα, 2020

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

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 εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς το συγγραφέα. Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

ΠΡΟΛΟΓΟΣ

Εδώ και πολλά χρόνια οι δορυφορικές εικόνες χρησιμοποιούνται σε τομείς όπως η μετεωρολογία, η δασολογία και η γεωλογία. Τη τελευταία δεκαετία όμως, τόσο η αύξηση της διακριτικής ικανότητας των δορυφόρων, όσο και η μείωση του κόστους απόκτησης των δορυφορικών εικόνων έχει δώσει τη δυνατότητα χρήσης τους σε ένα πολύ μεγάλο εύρος εφαρμογών όπως η γεωργία, η χαρτογραφία, οι τηλεπικοινωνίες, οι επιχειρήσεις κοινής ωφέλειας καθώς και σε καταστάσεις έκτακτης ανάγκης, όπως σε πυρκαγιές και φυσικές καταστροφές. Τροχοπέδη στην χρήση δορυφορικών εικόνων είναι η παρουσία σύννεφων, είτε σε ένα μικρό τμήμα τους είτε και σε ολόκληρη την εικόνα έχοντας ως αποτέλεσμα να «κρύβονται» σημαντικά χαρακτηριστικά της επιφάνειας της γης.

Αντικείμενο της παρούσας διπλωματικής εργασίας είναι να εξεταστούν ήδη υπάρχοντα μοντέλα γεννητικών ανταγωνιστικών δικτύων, τα οποία έχουν χρησιμοποιηθεί με επιτυχία σε προβλήματα «μετάφρασης» εικόνων από ένα πεδίο πηγής σε ένα πεδίο στόχου, έτσι ώστε να αξιολογήσουμε κατά πόσο θα είχαν εφαρμογή στο πρόβλημα αφαίρεσης σύννεφων από τις δορυφορικές εικόνες.

ΕΥΧΑΡΙΣΤΙΕΣ

Με την ολοκλήρωση της παρούσας διπλωματικής εργασίας, θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου, κύριο Καραντζάλο Κωνσταντίνο, Αναπληρωτή Καθηγητή Ε.Μ.Π., για την ανάθεση της παρούσας εργασίας, για το ενδιαφέρον του, καθώς και για τις επιστημονικές του παρατηρήσεις και υποδείξεις.

Θα ήθελα επίσης να ευχαριστήσω τον κύριο Ντούσκο Μάκη για τις πολύτιμες επιστημονικές συμβουλές και την καθοδήγησή του.

ΠΕΡΙΛΗΨΗ

Στις μέρες μας, οι δορυφορικές εικόνες είναι καθοριστικής σημασίας για ποικίλες εφαρμογές, όπως η αναγνώριση των αποτυπωμάτων των κτιρίων ή η ανίχνευση αλλαγών στις χρήσεις κάλυψης γης. Συχνά καλύπτονται από μεμβράνες νεφών που καλύπτουν εν μέρει ή πλήρως την εικόνα. Αυτό δυσκολεύει αρκετά τη χρήση τους, ειδικά για περιοχές που έχουν για μεγάλο μέρος του έτους συννεφοκάλυψη, όπως για παράδειγμα το Παρίσι ή το Λονδίνο. Για την αντιμετώπιση του προβλήματος, μέχρι στιγμής χρησιμοποιούνται συνδυαστικά μάσκες που υποδηλώνουν τη θέση των σύννεφων και στη συνέχεια φίλτρα που προσπαθούν να ανακατασκευάσουν την εικόνα στο τμήμα της μάσκας. Επομένως, απαιτείται να βρεθεί μία αυτόματη τεχνική που θα ανιχνεύει και θα αφαιρεί τις περιοχές με σύννεφα και θα τις αντικαθιστά με καθαρές από σύννεφα ανακατασκευασμένες σκηνές.

Τα Γεννητικά Ανταγωνιστικά Δίκτυα (Generative Adversarial Networks, GANs) είναι μία οικογένεια μοντέλων ικανών να παράγουν ρεαλιστικές, συνθετικές εικόνες. Ένα γεννητικό ανταγωνιστικό δίκτυο αποτελείται στην πραγματικότητα από δύο δίκτυα. Με το ένα δίκτυο προσπαθεί να δημιουργήσει όσο το δυνατόν πιο ρεαλιστικές εικόνες (δημιουργός), και με το άλλο προσπαθεί να ξεχωρίσει όσο το δυνατόν καλύτερα τις τεχνητές από τις αληθινές εικόνες (διευκρινιστής). Μέσω του συνεχούς ανταγωνισμού μεταξύ τους, τα δίκτυα βελτιώνονται μέχρι να επέλθει η ισορροπία, όπου οι αληθινές και οι τεχνητές/δημιουργημένες εικόνες φαίνονται πανομοιότυπες στον διευκρινιστή. Κατά την κατάσταση της ισορροπίας ο δημιουργός παράγει τις πιο αληθοφανείς εικόνες. Τα συγκεκριμένα δίκτυα έχουν χρησιμοποιηθεί με μεγάλη επιτυχία σε προβλήματα «μετάφρασης» μίας εικόνας ενός πεδίου A (π.χ άλογα) σε ένα πεδίο B (π.χ ζέβρες).

Αντικείμενο της παρούσας μελέτης είναι να εξεταστούν ήδη υπάρχοντα μοντέλα γεννητικών ανταγωνιστικών δικτύων ως προς την δυνατότητα τους να χρησιμοποιηθούν στο πρόβλημα αφαίρεσης σύννεφων από τις δορυφορικές εικόνες. Τα δεδομένα που χρησιμοποιήθηκαν για την μελέτη ήταν μία χρονοσειρά εικόνων από τον δορυφόρο sentinel-2 για το έτος 2018.

Στο πρώτο στάδιο της μεθοδολογίας έγινε η προεπεξεργασία των εικόνων ώστε να καταλήξουμε στο τελικό σύνολο δεδομένων που χρησιμοποιήθηκε κατά την εκπαίδευση των μοντέλων. Στη συνέχεια, εκπαιδεύτηκαν τρία μοντέλα το Pix2Pix, το CycleGAN και το Learnable Gated Temporal Shifted Module (LGTSM). Στόχος ήταν να αντιμετωπιστεί το πρόβλημα της αφαίρεσης σύννεφων με δύο τρόπους. Αρχικά ως καθαρό πρόβλημα εικόνας με τη χρήση των δύο πρώτων μοντέλων και στη συνέχεια ως πρόβλημα χρονοσειράς, με τη

χρήση του τρίτου μοντέλου, λαμβάνοντας υπόψη και τα χρονικά χαρακτηριστικά της εικόνας κατά την ανακατασκευή της.

Τα αποτελέσματα του κάθε μοντέλου αξιολογήθηκαν με τα διαγράμματα απώλειας που δημιουργήθηκαν κατά τη διάρκεια της εκπαίδευσης του μοντέλου και με τις εικόνες που δημιουργήθηκαν. Για τα δύο πρώτα μοντέλα που αντιμετωπίζουν το πρόβλημα με τον ίδιο τρόπο χρησιμοποιήθηκε το ίδιο σύνολο ελέγχου για να μπορέσουν να συγκριθούν μεταξύ τους. Τέλος, υπολογίστηκαν κάποιες μετρικές που αφορούν την ποιότητα εικόνας ώστε να αξιολογηθούν τα αποτελέσματα του κάθε μοντέλου συγκριτικά με τα υπόλοιπα.

Λέξεις-κλειδιά

Αφαίρεση σύννεφων, Γεννητικά Ανταγωνιστικά Δίκτυα, Βαθιά Μάθηση, Αποστολή Sentilel-2, Κανονικοποίηση, Τεχνικές συμπλήρωσης εικόνας

ABSTRACT

Nowadays, satellite imagery is crucial for a variety of applications, such as the recognition of building footprints or the detection of changes in land coverage uses. They are often covered by cloud membranes that partially or completely cover the image. This makes them quite difficult to use, especially for areas that are covered by clouds most days of the year, such as Paris or London. So far, in order to face the problem, masks which indicate the position of the clouds are used in combination. Afterwards, filters are used to reconstruct the image in the part of the mask. Therefore, it is necessary to find an automatic technique that will detect and remove areas with clouds and replace them with cloudless reconstructed scenes.

Generative Adversarial Networks (GANs) are a family of models capable of producing realistic synthetic images. A generative adversarial network actually consists of two networks. The first network tries to create as realistic images as possible (generator), while the other tries to discriminate the artificial images from the real ones as accurately as possible (discriminator). Through constant competition between them, the networks are improved until equilibrium is reached, where the real and the artificially created images look identical to the discriminator. During the equilibrium state the generator produces the most realistic images. These networks have been used with great success in "translation" problems of an image of a field A (e.g. horses) to a field B (e.g. zebras).

The object of this study is to examine existing models of generative adversarial networks in terms of their potential to be used in the problem of removing clouds from satellite images. The data used for the study was a time series of images from the Sentinel-2 satellite for the year 2018.

In the first stage of the methodology, the images were pre-processed in order to reach the final dataset which was used during the model training. Then, three models were trained: Pix2Pix, CycleGAN and the Learnable Gated Temporal Shifted Module (LGTSM). The aim was to address the problem of cloud removal in two ways. Initially as a pure image problem, by utilizing the first two models. Afterwards, by using the third model as a time series problem, taking into account the temporal characteristics of the image during its reconstruction as well.

The results of each model were evaluated with the loss diagrams created during the model training and as well as with the generated images. For the first two models facing the problem in the same way, the same testing set was used so that they could be compared with each

other. Finally, some metrics related to image quality were calculated in order to evaluate the results of each model compared to the others.

Key-words

Cloud Removal, Generative Adversarial Networks, Deep Learning, Sentinel-2, Normalization, Inpainting techniques

ΠΕΡΙΕΧΟΜΕΝΑ

ΠΡΟΛΟΓΟΣ.....	I
ΕΥΧΑΡΙΣΤΙΕΣ.....	II
ΠΕΡΙΛΗΨΗ.....	III
ABSTRACT.....	V
1. ΕΙΣΑΓΩΓΗ.....	9
1.1 Γενικά.....	9
1.2 Σκοπός εργασίας.....	9
1.3 Δομή εργασίας.....	10
2. ΑΝΑΣΚΟΠΗΣΗ ΣΧΕΤΙΚΗΣ ΒΙΒΛΙΟΓΡΑΦΙΑΣ.....	12
3. ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ.....	18
3.1 Ιστορική Αναδρομή - Βασικές Έννοιες.....	18
3.2 Συνελκτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks).....	24
3.2.1 Αρχιτεκτονική Συνελκτικών Δικτύων.....	24
3.2.2 Συστατικά Μέρη.....	25
3.2.2.1 Επίπεδο συνέλιξης (Convolution Layer).....	26
3.2.2.2 Συναρτήσεις Ενεργοποίησης (Activation Functions).....	27
3.2.2.3 Επίπεδο Συγκέντρωσης (Pooling).....	29
3.2.2.4 Πλήρως συνδεδεμένο επίπεδο.....	30
3.3 Μεταφορά Μάθησης.....	31
3.4 Γεννητικά Ανταγωνιστικά Δίκτυα (Generative Adversarial Networks).....	32
3.4.1 Διευκρινιστής (Discriminator).....	33
3.4.2 Δημιουργός (Generator).....	34
3.4.3 Γνωστές Αρχιτεκτονικές Γεννητικών Ανταγωνιστικών Δικτύων.....	34
3.4.3.1 U-Net Δημιουργός.....	34
3.4.3.2 Patch-GAN Διευκρινιστής.....	35
3.4.4 Συναρτήσεις Κόστους Γεννητικών Ανταγωνιστικών Δικτύων.....	36
3.4.4.1 Σιγμοειδής Εγκάρσια Εντροπία (Sigmoid Cross Entropy).....	36
3.4.4.2 Μέσο Απόλυτο Σφάλμα (Mean Absolute Error - MAE).....	36
3.4.4.3 Απώλεια Συνέπειας Κύκλου και Απώλεια Ταυτότητας (Cycle consistency loss and Identity loss).....	37
3.4.4.4 Απώλεια Μάσκας (Masked Loss).....	37
3.4.4.5 Αντιληπτική Απώλεια (Perceptual Loss).....	38
3.4.4.6 Απώλεια Στυλ (Style Loss).....	38
3.4.4.7 TSMGAN απώλεια.....	38
3.4.5 Αξιολόγηση Γεννητικών Ανταγωνιστικών Δικτύων – Μετρικές.....	39
3.5 Συμπλήρωση Εικόνας (Image Inpainting).....	41
3.6 Τηλεπισκοπικά Δεδομένα Sentinel-2.....	42

4.	ΜΕΘΟΔΟΛΟΓΙΑ.....	47
4.1	Περιγραφή συνόλου δεδομένων (Dataset)	47
4.2	Επεξεργασία αρχικού συνόλου δεδομένων	48
4.2.1	Επιλογή των Καναλιών του Sentinel-2	49
4.2.2	Κανονικοποίηση τιμών (Normalization)	49
4.2.3	Επεξεργασία μασκών.....	50
4.2.4	Δημιουργία μικρότερων εικόνων (patches) από τις αρχικές	50
4.2.5	Διαχωρισμός εικόνων σε κατηγορίες	51
4.3	Πρώτο Μοντέλο (Pix2Pix)	51
4.3.1	Αρχιτεκτονική	52
4.3.2	Εκπαίδευση.....	55
4.4	Δεύτερο Μοντέλο (CycleGAN)	56
4.4.1	Αρχιτεκτονική	57
4.4.2	Εκπαίδευση.....	58
4.5	Τρίτο Μοντέλο (LGTSM)	59
4.5.1	Αρχιτεκτονική	61
4.5.2	Εκπαίδευση.....	62
5.	ΠΟΙΟΤΙΚΗ ΚΑΙ ΠΟΣΟΤΙΚΗ ΑΞΙΟΛΟΓΗΣΗ	64
5.1	Αποτελέσματα και αξιολόγηση μοντέλου Pix2Pix	64
5.2	Αποτελέσματα και αξιολόγηση μοντέλου CycleGAN	68
5.3	Αποτελέσματα και αξιολόγηση μοντέλου LGTSM	71
5.3.1	1 ^ο πείραμα (χρονοσειρά).....	71
5.3.2	2 ^ο πείραμα (χρονοσειρά).....	74
5.3.3	3 ^ο πείραμα (χρονοσειρά).....	76
5.3.4	4 ^ο πείραμα (χρονοσειρά).....	78
5.3.5	5 ^ο Πείραμα (χρονοσειρά).....	79
6.	ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΠΡΟΟΠΤΙΚΕΣ.....	89
6.1	Συμπεράσματα.....	89
6.2	Προοπτικές	91
7.	ΒΙΒΛΙΟΓΡΑΦΙΑ.....	92

1. ΕΙΣΑΓΩΓΗ

1.1 Γενικά

Οι εικόνες τηλεπισκόπισης βρίσκονται στον πυρήνα πολλών δραστηριοτήτων παρατήρησης της γης. Η τακτική, συνεπής και παγκόσμιας κλίμακας φύση των δορυφορικών δεδομένων αξιοποιείται σε πολλές εφαρμογές, όπως η παρακολούθηση των καλλιεργήσιμων εκτάσεων, οι αλλαγές χρήσης γης και η αξιολόγηση καταστροφών. Ωστόσο, ένα κύριο πρόβλημα που επηρεάζει τη χρονική και χωρική διαθεσιμότητα των επιφανειακών παρατηρήσεων είναι η παρουσία σύννεφων. Η ύπαρξη σύννεφων στις δορυφορικές εικόνες αποτελεί ένα σημαντικό πρόβλημα το οποίο πολλές φορές εμποδίζει την ανάλυση της επιφάνειας της γης, είτε επειδή τα σύννεφα είναι τόσο πυκνά με αποτέλεσμα να χάνονται πληροφορίες, είτε επειδή δημιουργούν θολές εικόνες. Για την αντιμετώπιση του προβλήματος έχουν γίνει προσπάθειες αφαίρεσης των σύννεφων κυρίως με τη χρήση φίλτρων. Τα τελευταία χρόνια όμως με την ανάπτυξη των τεχνικών μηχανικής μάθησης έχουν αναπτυχθεί αρκετές παραλλαγές των γεννητικών ανταγωνιστικών μοντέλων που ασχολούνται με την ανακατασκευή εικόνων από ένα πεδίο πηγής σε ένα πεδίο στόχου με πολύ θετικά αποτελέσματα.

Η παρούσα διπλωματική εργασία έχει ως αντικείμενο την μελέτη γεννητικών ανταγωνιστικών μοντέλων και την δυνατότητα τους να εφαρμοστούν στο πρόβλημα αφαίρεσης σύννεφων, με σκοπό την ανακατασκευή της εικόνας. Τα τρία μοντέλα που εξετάζονται έχουν ήδη χρησιμοποιηθεί με επιτυχία σε άλλες εφαρμογές και εδώ γίνεται προσπάθεια να εκπαιδευτούν με τα δικά μας δεδομένα για να διαπιστωθεί κατά πόσο είναι κατάλληλα για το πρόβλημα αφαίρεσης σύννεφων.

1.2 Σκοπός εργασίας

Στην παρούσα εργασία έγινε μία έρευνα για το ποιο μοντέλο από αυτά που υπάρχουν ήδη στη βιβλιογραφία, θα μπορούσε να χρησιμοποιηθεί για την αφαίρεση των σύννεφων από τις δορυφορικές εικόνες Sentinel-2. Για την υλοποίηση της εργασίας έπρεπε αρχικά να επεξεργαστεί το σύνολο δεδομένων και στη συνέχεια να εκπαιδευτούν τα μοντέλα που επιλέχθηκαν. Τα δεδομένα που χρησιμοποιήθηκαν προέρχονταν από την αποστολή Sentinel-2 και πρόκειται για μία χρονοσειρά εικόνων του έτους 2018. Το πρόβλημα που εξετάζει η διπλωματική αντιμετωπίστηκε με δύο τρόπους. Πρώτον ως πρόβλημα μετάφρασης εικόνας και δεύτερον ως πρόβλημα συμπλήρωσης εικόνων που ανήκουν σε χρονοσειρά, ώστε να

ενταχθεί και η έννοια του χρόνου. Τα μοντέλα που επιλέχθηκαν έχουν στο παρελθόν χρησιμοποιηθεί για την επίλυση τέτοιων προβλημάτων σε διαφορετικές όμως εφαρμογές.

1.3 Δομή εργασίας

Κεφάλαιο 1

Στο πρώτο κεφάλαιο γίνεται μία πρώτη εισαγωγή στο πρόβλημα που καλείται η παρούσα διπλωματική να επιλύσει και γίνεται αναφορά και στους στόχους που έχει η συγκεκριμένη μελέτη.

Κεφάλαιο 2

Στο δεύτερο κεφάλαιο παρουσιάζεται η βιβλιογραφική ανασκόπηση. Γίνεται μία σύντομη αναφορά στις κυριότερες εργασίες που έχουν ασχοληθεί με παρόμοιο θέμα ή έχουν χρησιμοποιήσει παρόμοιες τεχνικές με αυτές που χρησιμοποιούνται και στην παρούσα εργασία.

Κεφάλαιο 3

Στο τρίτο κεφάλαιο αναπτύσσεται όλο το θεωρητικό υπόβαθρο πάνω στο οποίο βασίστηκε η εργασία. Αρχικά, γίνεται αναφορά σε βασικές έννοιες μηχανικής μάθησης και περιγράφεται ο τρόπος λειτουργίας των δικτύων βαθιάς μάθησης. Στη συνέχεια, περιγράφεται η λειτουργία και η αρχιτεκτονική των γεννητικών ανταγωνιστικών δικτύων. Τέλος, γίνεται αναφορά στη μέθοδο συμπλήρωσης εικόνας και στα χαρακτηριστικά των δορυφορικών εικόνων Sentinel-2.

Κεφάλαιο 4

Στο τέταρτο κεφάλαιο περιγράφονται τα δεδομένα που χρησιμοποιήθηκαν στην εργασία και όλη η μεθοδολογία που εφαρμόστηκε. Αρχικά παρουσιάζεται το στάδιο της προεπεξεργασίας, όπου έγινε εξαγωγή των καναλιών που θέλαμε να χρησιμοποιήσουμε και «κόπηκαν» οι εικόνες σε μικρότερα τμήματα ώστε να διευκολύνουν την εκπαίδευση. Στη συνέχεια, έγινε κανονικοποίηση των τιμών των εικονοστοιχείων και χωρίστηκαν οι εικόνες σε δύο κατηγορίες, τις εικόνες με σύννεφα και τις εικόνες χωρίς σύννεφα. Από αυτές τις κατηγορίες έγινε και η τελική επιλογή των εικόνων που χρησιμοποιήθηκαν για την εκπαίδευση των μοντέλων. Τέλος, παρουσιάζεται η επιλογή των μοντέλων, η αρχιτεκτονική τους και η επιλογή των παραμέτρων που χρησιμοποιήθηκαν κατά την εκπαίδευση.

Κεφάλαιο 5

Στο πέμπτο κεφάλαιο παρουσιάζεται η αξιολόγηση των αποτελεσμάτων της εκπαίδευσης των μοντέλων με βάση τις εικόνες που δημιουργήθηκαν από κάθε μοντέλο. Επίσης γίνεται χρήση μετρικών ποιότητας εικόνας ώστε να γίνει σύγκριση των αποτελεσμάτων των τριών μοντέλων.

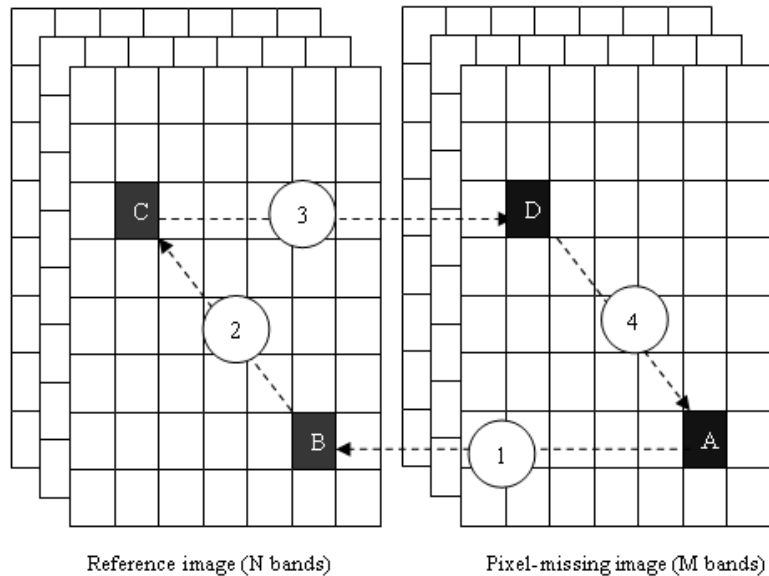
Κεφάλαιο 6

Στο έκτο και τελευταίο κεφάλαιο παρουσιάζονται τα συμπεράσματα της εργασίας και προτείνονται ιδέες και προοπτικές για συνέχιση της συγκεκριμένης έρευνας.

2. ΑΝΑΣΚΟΠΗΣΗ ΣΧΕΤΙΚΗΣ ΒΙΒΛΙΟΓΡΑΦΙΑΣ

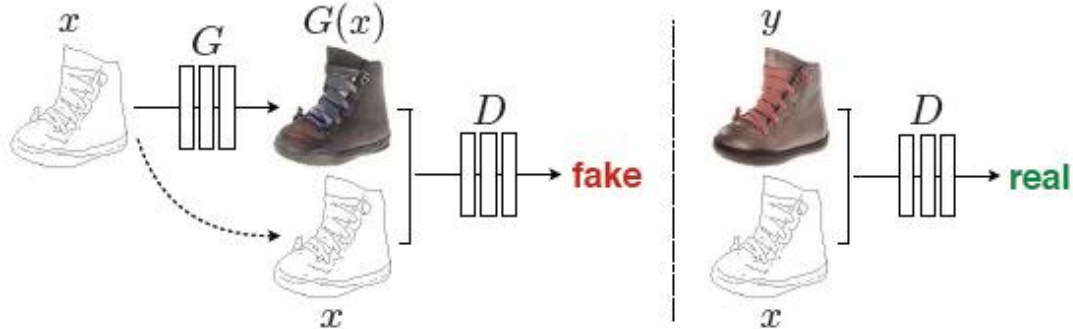
Το πρόβλημα που μελετά η παρούσα εργασία θα μπορούσε να αντιμετωπιστεί είτε ως πρόβλημα συμπλήρωσης εικόνας (Image Inpainting), είτε ως πρόβλημα «μετάφρασης» της εικόνας, από εικόνα με σύννεφα σε εικόνα χωρίς σύννεφα. Μελετώντας τη βιβλιογραφία βρέθηκαν πολλές μελέτες που προσπαθούν να δώσουν λύση στα συγκεκριμένα ζητήματα, καθώς πρόκειται για προβλήματα που έχουν απασχολήσει τους επιστήμονες την τελευταία δεκαετία. Σαφώς και η αύξηση των μελετών έχει επηρεαστεί από την ανάπτυξη των τεχνικών μηχανικής μάθησης και ιδιαίτερα από τα Γεννητικά Ανταγωνιστικά Δίκτυα που προτάθηκαν το 2014 από τον Ian Goodfellow [8]. Επιπλέον, έχουν παρουσιαστεί ενδιαφέρουσες μελέτες που εντάσσουν και το χαρακτηριστικό του χρόνου στην εκπαίδευση των προτεινόμενων μοντέλων, γεγονός που θα μπορούσε να φανεί πολύ χρήσιμο για την αντιμετώπιση του προβλήματος που εξετάζουμε. Παρακάτω, παρουσιάζονται οι σημαντικότερες εργασίες που μελετήθηκαν για την εκπόνηση της διπλωματικής.

Το 2011 οι **Weiliang Fan et al.** [5] παρουσίασαν την εργασία τους με τίτλο «*A Pixel Missing Patch Inpainting Method for Remote Sensing Image*». Η συγκεκριμένη εργασία καταπιάνεται με την αποκατάσταση των εικονοστοιχείων που λείπουν από τις εικόνες τηλεπισκόπησης, λόγω των σύννεφων. Προτείνει μία νέα μέθοδο σε σχέση με τις μέχρι τότε περιορισμένες τεχνικές, στην οποία χρησιμοποιείται μία πρόσφατη εικόνα της σκηνής (εικόνα αναφοράς) ως πίνακας αναζήτησης και γεμίζει τις τιμές των εικονοστοιχείων που λείπουν χρησιμοποιώντας τα εικονοστοιχεία που είναι παρόντα. Η μέθοδος περιλαμβάνει τέσσερα στάδια. Πρώτα προσδιορίζεται η θέση ενός εικονοστοιχείου και στις δύο εικόνες (A: αρχική εικόνα, B: εικόνα αναφοράς). Στη συνέχεια, για το εικονοστοιχείο της εικόνας αναφοράς (B) υπολογίζονται οι φασματικές αποστάσεις του από τα υπόλοιπα εικονοστοιχεία της εικόνας κι επιλέγεται εκείνο που έχει την μικρότερη απόσταση (C). Έπειτα, με βάση τα γεωμετρικά χαρακτηριστικά του (C) εντοπίζεται το αντίστοιχο εικονοστοιχείο στην αρχική εικόνα (D) και τέλος συμπληρώνονται οι τιμές του αρχικού εικονοστοιχείου (A) με βάση τις τιμές του εικονοστοιχείου (D) για τα διάφορα κανάλια. Παρακάτω φαίνεται το σχηματικό διάγραμμα της μεθόδου η οποία μπορεί να χρησιμοποιηθεί μεταξύ διαφορετικών τύπων εικόνων τηλεπισκόπησης που έχουν όμως την ίδια χωρική ανάλυση.



Εικόνα 2-1: Σχηματικό διάγραμμα της μεθόδου που παρουσιάζεται στο «*A pixel missing patch inpainting method for remote sensing image*».

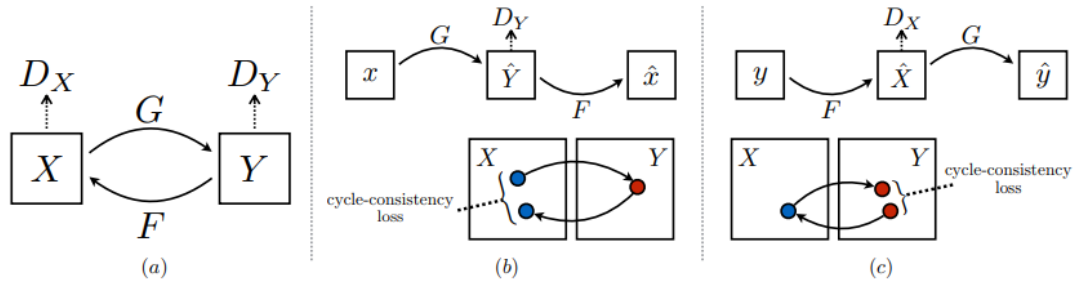
Το 2016 δημοσιεύθηκε η εργασία των **Isola Phillip et al. [1]** με τίτλο «*Image-to-Image Translation with Conditional Adversarial Networks*» οι οποίοι πρότειναν ένα νέο γεννητικό ανταγωνιστικό δίκτυο το Pix2Pix για την «μετάφραση» μία εικόνας από ένα πεδίο πηγής σε εικόνα ενός πεδίου στόχου. Η συγκεκριμένη μέθοδος βασίζεται στα γεννητικά ανταγωνιστικά δίκτυα υπό όρους, όπου δημιουργείται μια εικόνα στόχου, εξαρτώμενη από μια δεδομένη εικόνα εισόδου. Ο δημιουργός δέχεται μια δεδομένη εικόνα ως είσοδο και δημιουργεί μια μεταφρασμένη έκδοση της. Στον διευκρινιστή δίνεται μια εικόνα αναφοράς και μια πραγματική ή δημιουργημένη εικόνα και πρέπει να καθορίσει εάν η εικόνα είναι πραγματική ή ψεύτικη. Τέλος, το μοντέλο του δημιουργού εκπαιδεύεται τόσο για να ξεγελάσει τον διευκρινιστή όσο και για να ελαχιστοποιήσει την απώλεια μεταξύ της δημιουργούμενης εικόνας και της αναμενόμενης εικόνας στόχου. Το μοντέλο εκπαιδεύεται σε σύνολα δεδομένων εικόνων που αποτελούνται από εικόνες εισόδου (πριν από τη μετάφραση) και εικόνες εξόδου ή στόχου (μετά τη μετάφραση). Η αρχιτεκτονική του δημιουργού βασίζεται στη U-Net αρχιτεκτονική, ενώ για τον διευκρινιστή χρησιμοποιείται η PatchGAN. Αυτή η γενική αρχιτεκτονική επιτρέπει στο μοντέλο να εκπαιδευτεί για μια σειρά από εφαρμογές μετάφρασης εικόνας σε εικόνα.



Εικόνα 2-2: Λογική του μοντέλου Pix2Pix που παρουσιάζεται στο «*Image-to-Image Translation with Conditional Adversarial Networks*».

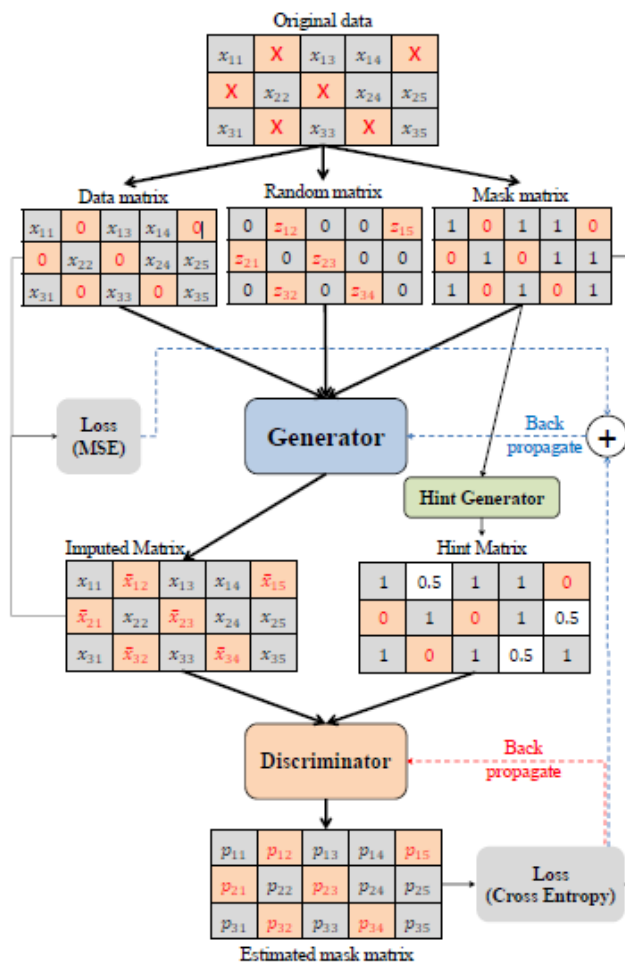
Το 2017 πάλι οι **Isola Phillip et al.** [6] επέκτειναν την μελέτη τους για την «μετάφραση» εικόνας στην εργασία τους με τίτλο «*Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks*». Στη συγκεκριμένη μελέτη προτείνουν ένα μοντέλο, το CycleGAN, το οποίο είναι μία τεχνική που περιλαμβάνει την αυτόματη εκπαίδευση μοντέλων για τη μετάφραση εικόνας σε εικόνα, χωρίς να απαιτείται η είσοδος στο μοντέλο κάποιας εικόνας αναφοράς. Τα μοντέλα εκπαιδεύονται χωρίς επίβλεψη, χρησιμοποιώντας μια συλλογή εικόνων από τον τομέα πηγής και στόχου και δεν χρειάζεται να σχετίζονται με κανέναν τρόπο. Το CycleGAN περιλαμβάνει την ταυτόχρονη εκπαίδευση δύο μοντέλων δημιουργών και δύο μοντέλων διευκρινιστών. Ο ένας δημιουργός λαμβάνει ως είσοδο εικόνες από το πρώτο πεδίο και εξάγει εικόνες για το δεύτερο πεδίο, ενώ ο δεύτερος δημιουργός δέχεται ως είσοδο εικόνες από το δεύτερο πεδίο και δημιουργεί εικόνες για το πρώτο πεδίο. Στη συνέχεια, τα μοντέλα των διευκρινιστών χρησιμοποιούνται για να προσδιορίσουν πόσο αληθοφανείς είναι οι δημιουργημένες εικόνες και να ενημερώσουν αναλόγως τα μοντέλα των δημιουργών. Για να είναι επαρκής αυτή η διαδικασία για τη δημιουργία μεταφράσεων των εικόνων εισαγωγής οι συγγραφείς εισήγαγαν την έννοια της «συνέπειας κύκλου». Η έννοια βασίζεται στην ιδέα ότι μια έξοδος εικόνας από τον πρώτο δημιουργό θα μπορούσε να χρησιμοποιηθεί ως είσοδος στον δεύτερο δημιουργό και η έξοδος του δεύτερου δημιουργού θα πρέπει να ταιριάζει με την αρχική εικόνα. Επίσης, θα πρέπει να ισχύει και το αντίστροφο. Η ενθάρρυνση της συνέπεια του κύκλου γίνεται με την πρόσθεση μιας επιπλέον απώλειας που μετράει τη διαφορά μεταξύ της παραγόμενης εικόνας εξόδου του δεύτερου δημιουργού και της αρχικής εικόνας. Η αρχιτεκτονική του δημιουργού περιέχει πολλαπλά μπλοκ συνελκτικών επιπέδων, ενώ για τον διευκρινιστή χρησιμοποιείται η αρχιτεκτονική PatchGAN. Το 2018 οι **Singh P. και Komodakis N.** [4] χρησιμοποίησαν την παραπάνω τεχνική στην εργασία τους «*CLOUD-GAN: Cloud Removal for Sentinel-2 Imagery Using a Cyclic Consistent Generative Adversarial Networks*» με σκοπό την

αφαίρεση των σύννεφων από τις δορυφορικές εικόνες Sentinel-2. Τα αποτελέσματα της έρευνας τους ήταν ικανοποιητικά για εικόνες με χαμηλή αραιή συννεφοκάλυψη.



Εικόνα 2-3: α) Οι συναρτήσεις «μετάφρασης» G , F και οι διευκρινιστές D_X , D_Y του μοντέλου CycleGAN. β) Η λογική της συνέπειας του κύκλου στο μοντέλο CycleGAN που παρουσιάζεται στο «*Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks*».

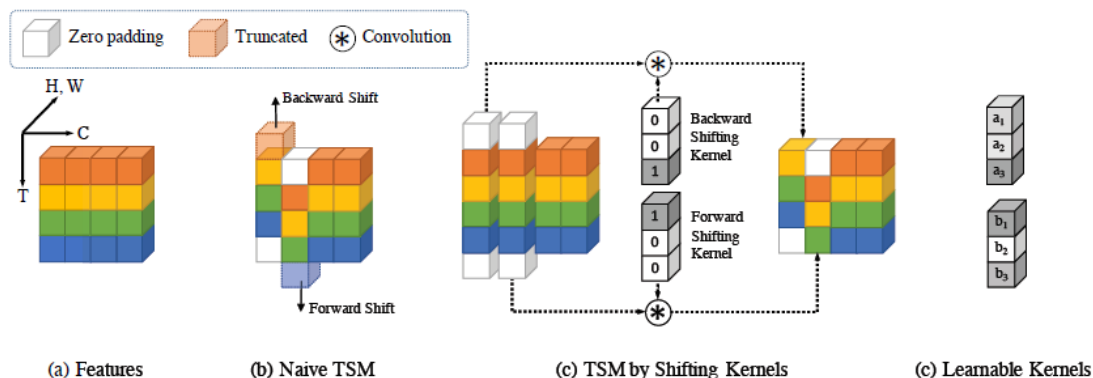
Το 2018 οι **Jinsung Yoon et al. [7]** στην εργασία τους με τίτλο «*GAIN: Missing Data Imputation using Generative Adversarial Nets*» πρότειναν μια νέα μέθοδο για τον υπολογισμό των «χαμένων» δεδομένων προσαρμόζοντας τα γνωστά γεννητικά ανταγωνιστικά δίκτυα. Η μέθοδος τους ονομάστηκε Generative Adversarial Imputation Networks (GAIN). Στη συγκεκριμένη μέθοδο ο δημιουργός (G) παρατηρεί ορισμένα στοιχεία ενός πραγματικού διανύσματος δεδομένων, υπολογίζει τα στοιχεία που λείπουν εξαρτώμενος από το τι πραγματικά παρατηρεί και εξάγει ένα ολοκληρωμένο διάνυσμα. Ο διευκρινιστής (D) στη συνέχεια παίρνει το ολοκληρωμένο διάνυσμα και προσπαθεί να προσδιορίσει ποια στοιχεία πράγματι παρατηρήθηκαν και ποια υπολογίστηκαν από τον δημιουργό. Για να διασφαλιστεί ότι ο διευκρινιστής αναγκάζει τον δημιουργό να μάθει την επιθυμητή κατανομή, παρέχεται στον διευκρινιστή κάποια επιπλέον πληροφορία, σαν υπόδειξη με τη μορφή διανύσματος. Η υπόδειξη αποκαλύπτει στον διευκρινιστή μερικές πληροφορίες σχετικά με την απουσία του αρχικού δείγματος, ώστε να εστιάσει την προσοχή του στον υπολογισμό συγκεκριμένων συστατικών. Αυτή η υπόδειξη διασφαλίζει ότι ο δημιουργός στην πραγματικότητα μαθαίνει να δημιουργεί σύμφωνα με την αληθινή κατανομή δεδομένων.



Εικόνα 2-4: Αρχιτεκτονική του μοντέλου GAIN που παρουσιάζεται στο «*GAIN: Missing Data Imputation using Generative Adversarial Nets*».

Το 2019 οι Chang Ya-Liang et al. [2] στη μελέτη τους με τίτλο «*Learnable Gated Temporal Shift Module for Deep Video Inpainting*» επέκτειναν τη μέθοδο που είχαν προτείνει στην προγενέστερη μελέτη τους που είχε δημοσιευτεί την ίδια χρονολογία «*Free-form Video Inpainting with 3D Gated Convolution and Temporal PatchGAN*» [3]. Στη συγκεκριμένη μελέτη εξετάζεται το πως θα συμπληρωθούν τμήματα των καρέ του βίντεο που καλύπτονται από κάποια μάσκα. Οι συγγραφείς πρότειναν ένα μοντέλο που χρησιμοποιεί μία χρονικά μετατοπισμένη μονάδα (TSM), και έχει την δυνατότητα να αντιμετωπίζει τις χρονικές πληροφορίες σε δύο διαστάσεων συνελιξείς. Αν η είσοδος για κάθε συνελικτικό επίπεδο έχει διαστάσεις (B, C, L, H, W), όπου το B είναι ο αριθμός των δειγμάτων που διαβιβάζονται ταυτόχρονα στο δίκτυο, C είναι ο αριθμός των καναλιών, L είναι το χρονικό μήκος, το H είναι το ύψος και το W είναι το πλάτος, τότε για κάθε καρέ στο χρονικό μήκος L, το TSM μετατοπίζει ένα τμήμα καναλιών στο προηγούμενο και στο επόμενο του καρέ πριν από την εφαρμογή των συνελιξεων. Αυτά τα μετατοπισμένα κανάλια περιέχουν χαρακτηριστικά από άλλα καρέ, με αποτέλεσμα μαζί με τα μη μετατοπισμένα χαρακτηριστικά, να δίνουν τη δυνατότητα στις αρχικές δύο διαστάσεων συνελιξείς να μπορέσουν να «μάθουν» τις χρονικές

δομές. Ωστόσο, το γεγονός ότι πολλές περιοχές της εικόνας είναι καλυμμένες με μάσκα, σημαίνει ότι πολλά χαρακτηριστικά που θα μετατοπιστούν δεν θα είναι ωφέλιμα για το μοντέλο κι επομένως δεν βοηθούν το TSM να μάθει τη διαφορά μεταξύ έγκυρων και μη έγκυρων χαρακτηριστικών. Για την αντιμετώπιση αυτού του ζητήματος, σχεδιάστηκε το Gated Temporal Shift Module (GTSM) κατά το οποίο εκτός από το TSM εφαρμόζεται ένα συνελκτικό φίλτρο που λειτουργεί ως πύλη (gating convolutional filter) στα χαρακτηριστικά της εισόδου και αναγνωρίζει τις περιοχές με μάσκα, τις περιοχές χωρίς μάσκα και τις περιοχές που έχουν «βαφτεί» για τα χαρακτηριστικά εξόδου από το συνδυασμό του TSM με το συνελκτικό φίλτρο. Η αρχιτεκτονική του δημιουργού βασίζεται στη U-Net αρχιτεκτονική, ενώ για τον διευκρινιστή χρησιμοποιείται η PatchGAN. Αξίζει να σημειωθεί ότι η συγκεκριμένη τεχνική έχει μικρότερο κόστος όσον αφορά τον υπολογιστικό χρόνο και τη μνήμη χρόνου εκτέλεσης σε σύγκριση με τις παραδοσιακές 2-διαστάσεων συνελίξεις.



Εικόνα 2-5: Επεξήγηση των μετατοπισμένων πυρήνων στο προτεινόμενο μοντέλο LGTSM που παρουσιάζεται στο «*Learnable Gated Temporal Shift Module for Deep Video Inpainting*». **a)** Χαρακτηριστικά εισόδου. Η μετατόπιση γίνεται στις διαστάσεις channel * time. **b)** Το αρχικό TSM. **c)** Ισοδύναμο TSM με χρονικά μετατοπισμένους πυρήνες. **d)** Οι πυρήνες που χρησιμοποιούνται στο LGTSM, μπορούν να «μάθουν» κι ας έχουν διαφορετικό μέγεθος.

3. ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ

3.1 Ιστορική Αναδρομή - Βασικές Έννοιες

Τα τελευταία χρόνια οι εφαρμογές της μηχανικής μάθησης, και γενικότερα της τεχνητής νοημοσύνης έχουν εισέλθει σε πολλούς τομείς της καθημερινότητας μας, όπως για παράδειγμα τα φίλτρα spam (spam filtering) ή οι μηχανές αναζήτησης. Η Τεχνητή Νοημοσύνη όμως, δεν είναι καινούργια έννοια καθώς πρώτη φορά αναφέρθηκε το 1950, στη μελέτη του άγγλου μαθηματικού Άλαν Τούρινγκ. Ως πεδίο θεμελιώθηκε το 1956 σε συνέδριο στο Dartmouth College που αφορούσε τη μελέτη προσομοίωσης της ανθρώπινης νοημοσύνης με χρήση υπολογιστών. Ένας καλά διατυπωμένος ορισμός, που καλύπτει τους πρακτικούς στόχους της Τεχνητής Νοημοσύνης και, πιο συγκεκριμένα, του κλάδου που καλείται Υπολογιστική Τεχνητή Νοημοσύνη, είναι ο ακόλουθος: *«Τεχνητή Νοημοσύνη είναι εκείνος ο κλάδος της επιστήμης των υπολογιστών που ασχολείται με το σχεδιασμό ευφών υπολογιστικών συστημάτων, δηλαδή συστημάτων με χαρακτηριστικά τα οποία σχετίζονται με την ευφυΐα στην ανθρώπινη συμπεριφορά (μάθηση, αιτίαση, επίλυση προβλημάτων, κατανόηση φυσικής γλώσσας, αναγνώριση αντικειμένων κτλ.)»*. Μερικά παραδείγματα εφαρμογών της τεχνητής νοημοσύνης είναι στην επεξεργασία φυσικής γλώσσας, στην αναγνώριση ομιλίας, καθώς και στο σχεδιασμό, τον προγραμματισμό και τη βελτιστοποίηση.

Η Μηχανική Μάθηση είναι υπό-πεδίο της τεχνητής νοημοσύνης. Αντικείμενο διερεύνησης της είναι η μελέτη και η κατασκευή αλγορίθμων που μπορούν να μαθαίνουν από τα δεδομένα και να κάνουν προβλέψεις σχετικά με αυτά. Τέτοιοι αλγόριθμοι λειτουργούν κατασκευάζοντας μοντέλα από πειραματικά δεδομένα, προκειμένου να κάνουν προβλέψεις βασισζόμενες στα δεδομένα ή να εξάγουν αποφάσεις που εκφράζονται ως το αποτέλεσμα. Αρκετοί ορισμοί έχουν δοθεί για τον προσδιορισμό του όρου «Μηχανική Μάθηση». Ένας από του πιο χαρακτηριστικούς δόθηκε από τους Carbonell & Gil το 1987 *«Μηχανική Μάθηση είναι η μελέτη υπολογιστικών μεθόδων για την απόκτηση νέας γνώσης, νέων δεξιοτήτων και νέων τρόπων οργάνωσης της υπάρχουσας γνώσης»*. Οι τρόποι με τους οποίους αναπτύσσεται η μηχανική μάθηση είναι ανάλογοι με τους τρόπους που μαθαίνει ο άνθρωπος και είναι οι ακόλουθοι:

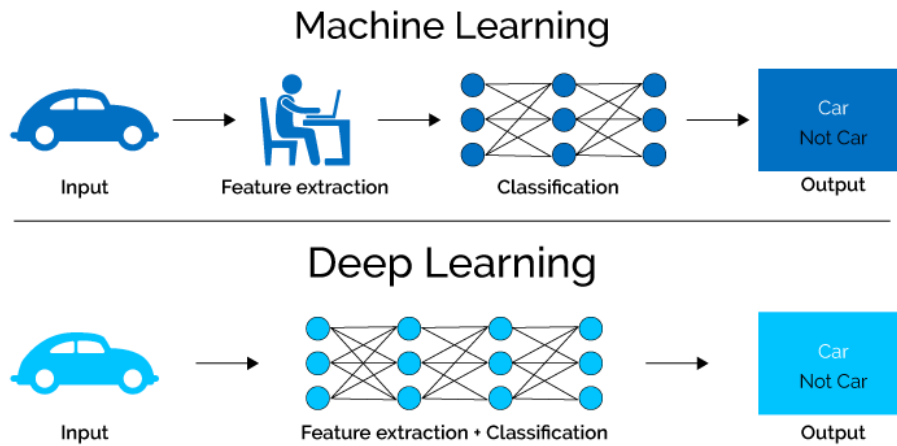
- Επιβλεπόμενη μάθηση (supervised Learning): ο αλγόριθμος κατασκευάζει μία συνάρτηση που απεικονίζει δεδομένες εισόδους (σύνολο εκπαίδευσης) σε γνωστές επιθυμητές εξόδους, με απώτερο στόχο τη γενίκευση της συνάρτησης αυτής για

εισόδους με άγνωστη έξοδο. Αυτό το είδος μάθησης χρησιμοποιείται σε προβλήματα ταξινόμησης (Classification) και πρόβλεψης (Prediction)

- Μη Επιβλεπόμενη Μάθηση (Unsupervised Learning): ο αλγόριθμος κατασκευάζει ένα μοντέλο για κάποιο σύνολο εισόδων υπό μορφή παρατηρήσεων χωρίς να γνωρίζει τις επιθυμητές εξόδους. Χρησιμοποιείται σε προβλήματα ανάλυσης συσχετισμών (Association Analysis) και ομαδοποίησης (Clustering)
- Ενισχυτική Μάθηση (Reinforcement Learning): ο αλγόριθμος μαθαίνει μια στρατηγική ενεργειών μέσα από άμεση αλληλεπίδραση με το περιβάλλον. Χρησιμοποιείται κυρίως σε προβλήματα σχεδιασμού (Planning), όπως για παράδειγμα ο έλεγχος κίνησης ρομπότ και η βελτιστοποίηση εργασιών σε εργοστασιακούς χώρους.

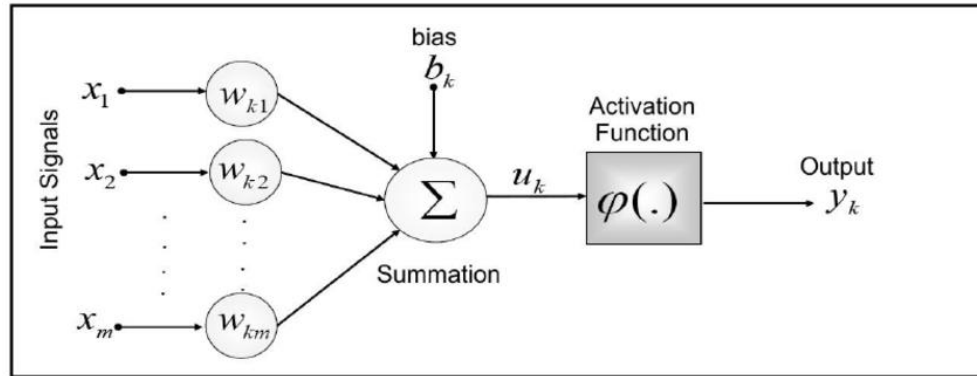
Εκτός από τις παραπάνω γενικές κατηγορίες υπάρχουν και ενδιάμεσες – συνδυαστικές περιπτώσεις αλγορίθμων και εφαρμογών Μηχανικών Μάθησης, όπως για παράδειγμα η ημι-επιβλεπόμενη μάθηση.

Ο κλάδος της Μηχανικής Μάθησης που καταπιάνεται με περίπλοκες μη γραμμικές αρχιτεκτονικές αλγορίθμων, έχοντας ως στόχο την εξαγωγή σύνθετων χαρακτηριστικών από τα δεδομένα ονομάζεται Βαθιά Μάθηση. Η βασική διαφοροποίηση στις εφαρμογές της Βαθιάς Μάθησης είναι ο αριθμός των αναπροσαρμογών που δέχεται ένα σήμα, από τη στιγμή της «εισόδου» του έως τη στιγμή της «εξόδου» του από το υπολογιστικό σύστημα. Ο αριθμός των αναπροσαρμογών, πρακτικά μεταφράζεται στον αριθμό των διάφορων επιπέδων (layers) με τα οποία είναι δομημένος ένας αλγόριθμος Βαθιάς Μάθησης. Τα επίπεδα αυτά δομούνται με ιεραρχικό τρόπο, ώστε οι συντελεστές των υψηλότερων επιπέδων να αναπροσαρμόζονται με βάση αυτούς των χαμηλότερων. Τα θεμέλια της βαθιάς μάθησης είναι τα νευρωνικά δίκτυα, τα οποία συνδυάζονται και δημιουργούν τα βαθιά νευρωνικά δίκτυα



Εικόνα 3-1: Διαφορά Μηχανικής Μάθησης (ML) και Βαθιάς Μάθησης (DL).
 Πηγή: <https://towardsdatascience.com/>

Η περιοχή των νευρωνικών δικτύων ξεκίνησε πριν περίπου 50 χρόνια, αλλά η μεγάλη ώθηση σ' αυτά δόθηκε μετά το 1980, χάρη στην αυξανόμενη εμφάνιση μαζικά παράλληλων ηλεκτρονικών υπολογιστών. Αποτελούν μία απόπειρα προσέγγισης της λειτουργίας του ανθρώπινου εγκεφάλου από μία μηχανή και η αρχιτεκτονική τους στηρίζεται στην αρχιτεκτονική των Βιολογικών Νευρωνικών. Τα τεχνητά νευρωνικά δίκτυα είναι μια συλλογή από νευρώνες που συνδέονται μεταξύ τους. Κάθε νευρώνας έχει πολλές εισόδους αλλά μόνο μια έξοδο, η οποία με την σειρά της μπορεί να αποτελέσει είσοδο για άλλους νευρώνες. Οι συνδέσεις μεταξύ των νευρώνων διαφέρουν ως προς την σημαντικότητά τους, η οποία καθορίζεται από τον συντελεστή βάρους (σύναψη). Η επεξεργασία κάθε νευρώνα καθορίζεται από την συνάρτηση ενεργοποίησης, η οποία καθορίζει την κάθε έξοδο σε σχέση με τις εισόδους και τους συντελεστές βάρους. Για να χρησιμοποιηθεί ένα τεχνητό νευρωνικό δίκτυο πρέπει πρώτα να εκπαιδευτεί για να «μάθει». Η μάθηση συνίσταται στον προσδιορισμό των κατάλληλων συντελεστών βάρους, ώστε το δίκτυο να εκτελεί τους επιθυμητούς υπολογισμούς. Ο ρόλος των συντελεστών βάρους μπορεί να ερμηνευτεί ως αποθήκευση γνώσης, η οποία παρέχεται μέσω παραδειγμάτων. Με αυτό τον τρόπο τα νευρωνικά δίκτυα μαθαίνουν το περιβάλλον τους, δηλαδή το φυσικό μοντέλο που παρέχει τα δεδομένα. Στο παρακάτω σχήμα παρουσιάζεται η δομή ενός τεχνητού νευρώνα που όπως αναφέρθηκε αποτελεί τη βάση για τον σχεδιασμό των τεχνητών νευρωνικών δικτύων.

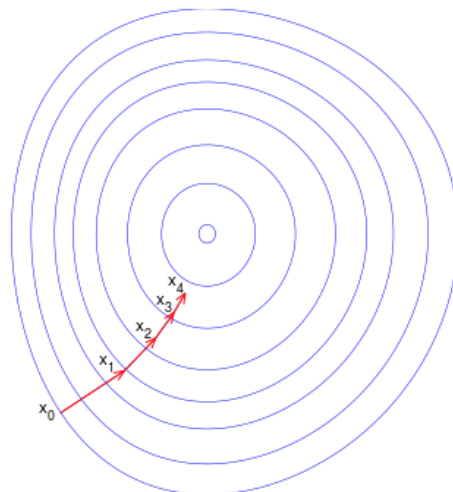


Εικόνα 3-2: Αναπαράσταση/Δομή τεχνητού νευρώνα. Πηγή: www.researchgate.net

Οι νευρώνες ενός τεχνητού νευρωνικού δικτύου οργανώνονται σε παράλληλη διάταξη μέσω διαφόρων επιπέδων. Τα επίπεδα που συναντώνται σε ένα τεχνητό νευρωνικό δίκτυο είναι α) το επίπεδο εισόδου που λαμβάνει τα δεδομένα εισόδου και παρέχει πληροφορίες από τον έξω κόσμο στο δίκτυο χωρίς περαιτέρω υπολογισμούς, β) το κρυφό επίπεδο που συνδέει το προηγούμενο και το επόμενο κατά σειρά επίπεδο νευρώνων και μέσω αυτού η είσοδος υπόκειται σε επεξεργασία και εξάγονται τα χαρακτηριστικά της, γ) το επίπεδο εξόδου που είναι το τελευταίο επίπεδο και σε αυτό λαμβάνεται η τελική απόφαση του δικτύου. Ανάλογα με τον τρόπο σύνδεσης των νευρώνων διακρίνονται διαφορετικές κατηγορίες δικτύων. Οι δύο βασικές κατηγορίες είναι α) τα δίκτυα πρόσθιας τροφοδότησης (feed-forward), στα οποία το σήμα μεταφέρεται μέσω των νευρώνων από την είσοδο έως την έξοδο και β) τα δίκτυα με ανατροφοδότηση (feedback), στα οποία το σήμα εξόδου τροφοδοτείται πίσω στις εισόδους όλων των άλλων νευρώνων από έναν βρόγχο ανατροφοδότησης.

Κατά την εκπαίδευση ενός νευρωνικού δικτύου γίνεται προσπάθεια πρόβλεψης κάθε εξόδου όσο πιο κοντά στην αντίστοιχη πραγματική τιμή. Την ακρίβεια της δυνατότητας αυτής του δικτύου την μετράμε χρησιμοποιώντας τη συνάρτηση κόστους (Loss Function). Με τη χρήση της συνάρτησης κόστους επιβάλλεται κάποια «διόρθωση» στο δίκτυο, στην περίπτωση που γίνονται λάθη. Στόχος είναι να αυξηθεί η ακρίβεια πρόβλεψης και να μειωθεί το σφάλμα ελαχιστοποιώντας τη συνάρτηση κόστους. Αυτό επιτυγχάνεται αναπροσαρμόζοντας τα βάρη και τις πολώσεις (bias) του δικτύου καθ' όλη τη διάρκεια της εκπαίδευσης. Ανάλογα με το είδος και την μορφή των δεδομένων μας, τη συνάρτηση ενεργοποίησης που χρησιμοποιούμε, καθώς και το κατά πόσο τα επιμέρους σφάλματα είναι μεγάλα ή μικρά σε σχέση με το βαθμό εκμάθησης (learning rate) που θέλουμε να πετύχουμε ορίζουμε και τη συνάρτηση κόστους. Μερικές συναρτήσεις κόστους που χρησιμοποιούνται ευρέως είναι: το Μέσο Τετραγωνικό Σφάλμα (MSE), το Μέσο Απόλυτο Σφάλμα (MAE) και η Εγκάρσια Εντροπία (Cross Entropy Loss).

Για την εύρεση ελάχιστων σημείων της συνάρτησης κόστους χρησιμοποιείται ο αλγόριθμος καθοδικής κλίσης. Πρόκειται για έναν επαναληπτικό αλγόριθμο βελτιστοποίησης και λαμβάνει υπόψη μόνο τις πρώτες μερικές παραγώγους της συνάρτησης, εκμεταλλευόμενος το γεγονός ότι το διάνυσμά τους δείχνει την κατεύθυνση στον N-διάστατο χώρο προς την οποία θα αυξηθεί περισσότερο η τιμή της συνάρτησης. Για το λόγο αυτό για την εύρεση των ελάχιστων σημείων ο αλγόριθμος κάνει βήματα προς την αντίθετη κατεύθυνση, αυτή δηλαδή που θα οδηγήσει στη μεγαλύτερη μείωση. Αρχικά οι μερικές παράγωγοι πολλαπλασιάζονται με τον ρυθμό μάθησης (learning rate), βάσει του οποίου καθορίζεται το μέγεθος των βημάτων και είναι αρκετά μικρός για λόγους σταθερότητας στην εκπαίδευση. Στη συνέχεια το αποτέλεσμα αφαιρείται από τις τιμές των παραμέτρων. Οι παράγωγοι μπορούν να ληφθούν με διάφορους τρόπους, όπως για παράδειγμα η Οπισθοδρομική Διάδοση (Backpropagation). Για τον αλγόριθμο καθοδικής κλίσης υπάρχουν αρκετές παραλλαγές. Χαρακτηριστικοί είναι ο Batch Gradient Descent, ο Stochastic Gradient Descent, ο Mini-batch Gradient Descent, ο Momentum, ο Adagrad, ο RMSprop, ο Adadelta και ο Adam. Οι συγκεκριμένοι αλγόριθμοι διαφέρουν στην ποσότητα των δεδομένων που χρειάζονται για να υπολογιστούν οι μερικές παράγωγοι. Όσο πιο πολλά τα δεδομένα, τόσο πιο ακριβείς είναι οι παράγωγοι που προκύπτουν, αλλά παράλληλα απαιτείται και περισσότερος χρόνος για τον υπολογισμό τους. Παρακάτω γίνεται εκτενής αναφορά στον αλγόριθμο βελτιστοποίησης Adam, που είναι και ο αλγόριθμος που χρησιμοποιήθηκε κατά τη διάρκεια της μεθοδολογίας.



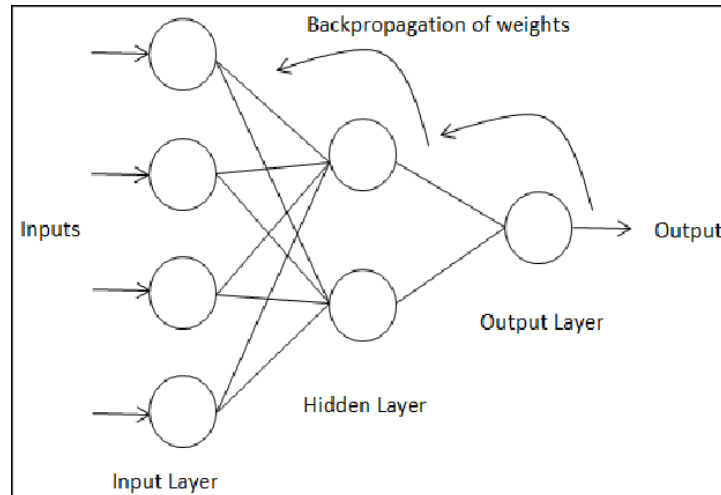
Εικόνα 3-3: Απεικόνιση της κατάβασης κλίσης σε μια σειρά συνόλων επιπέδων.
 Πηγή: <https://en.wikipedia.org/>

Ο αλγόριθμος της Προσαρμοζόμενης Εκτίμησης Ροπών, Adaptive Moment Estimation (Adam) συνδυάζει την προσαρμογή του ρυθμού μάθησης με έναν όρο ορμής (momentum).

Συγκεκριμένα, εκτιμά τις αναμενόμενες τιμές της κλίσης και του τετραγώνου της με δύο κινούμενους μέσους όρους, m_t για την κλίση, g_t και v_t για το - ανά στοιχείο διανύσματος - τετράγωνο της κλίσης g_t^2 . Τα m_t και v_t είναι εκτιμήσεις της πρώτης και δεύτερης ροπής των παραγώγων αντίστοιχα (μέση τιμή και μη προσανατολισμένη διακύμανση). Τα διανύσματα αυτά αρχικοποιούνται με μηδενικά, πράγμα που τα καθιστά προ-διατεθειμένα να λάβουν τιμές κοντά στο μηδέν, ειδικά στις αρχικές στιγμές ή όταν τα β_1 και β_2 είναι πολύ κοντά στο 1, όπως ισχύει συνήθως. Οι πολώσεις αυτές αντιμετωπίζονται με τον υπολογισμό «διορθωμένων» εκτιμήσεων των ροπών ($\widehat{m}_t, \widehat{v}_t$) που χρησιμοποιούνται τελικά για την ανανέωση των παραμέτρων. Οι εξισώσεις που περιγράφουν τη μεταβολή των παραμέτρων στο βήμα t είναι:

$$\begin{aligned} m_t &= \beta_1 * m_{t-1} + (1 - \beta_1) * g_t \\ v_t &= \beta_2 * v_{t-1} + (1 - \beta_2) * g_t^2 \\ \widehat{m}_t &= \frac{m_t}{1 - (\beta_1)^t} \\ \widehat{v}_t &= \frac{v_t}{1 - (\beta_2)^t} \\ \theta_t &= \theta_{t-1} - \eta \frac{\widehat{m}_t}{\sqrt{\widehat{v}_t} + \epsilon} \end{aligned}$$

Για τον υπολογισμό των μερικών παραγώγων της συνάρτησης κόστους ως προς τα βάρη και τις πολώσεις τους, ώστε να πραγματοποιηθεί στη συνέχεια ένα βήμα του αλγορίθμου καθοδικής κλίσης, χρησιμοποιείται στην εκπαίδευση νευρωνικών δικτύων ο αλγόριθμος της οπισθοδρομικής διάδοσης. Για να εφαρμοστεί, πρέπει η συνάρτηση κόστους πρώτον να μπορεί να γραφτεί ως μέσος όρος των επιμέρους απωλειών των δεδομένων εισόδου, και δεύτερον να μπορεί να γραφτεί ως συνάρτηση των εξόδων του δικτύου. Η πρώτη συνθήκη πρέπει να ισχύει διότι η οπισθοδρομική διάδοση υπολογίζει τις μερικές παραγώγους της συνάρτησης κόστους για μία μόνο είσοδο, οπότε οι τελικές τιμές πρέπει να ανακατασκευαστούν με αυτόν τον τρόπο. Η δεύτερη συνθήκη πρέπει να ισχύει για να είναι καλά ορισμένες οι παράγωγοι του τελευταίου επιπέδου, από όπου και ξεκινάει ο υπολογισμός τους. Η διαδικασία υπολογισμού έχει ως εξής: Αρχικά ορίζεται μια ποσότητα ανάλογη με τον όρο σφάλματος των κόμβων εξόδου. Στη συνέχεια γίνεται η οπισθοδρόμηση βασίζόμενοι στην ιδέα ότι κάθε κρυφός κόμβος είναι «υπεύθυνος» για ένα κλάσμα του σφάλματος σε κάθε έναν από τους κόμβους εξόδου στους οποίους συνδέεται. Έτσι οι τιμές του σφάλματος διαιρούνται ανάλογα με την ισχύ της σύνδεσης μεταξύ του κρυφού κόμβου και του κόμβου εξόδου και διαδίδονται προς τα πίσω για να δώσουν τιμές σφάλματος για το κρυφό επίπεδο.



Εικόνα 3-4: Αλγόριθμος Οπισθοδρομικής διάδοσης σφάλματος (Backpropagation).
 Πηγή: <https://deeptai.org/>

3.2 Συνελκτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks)

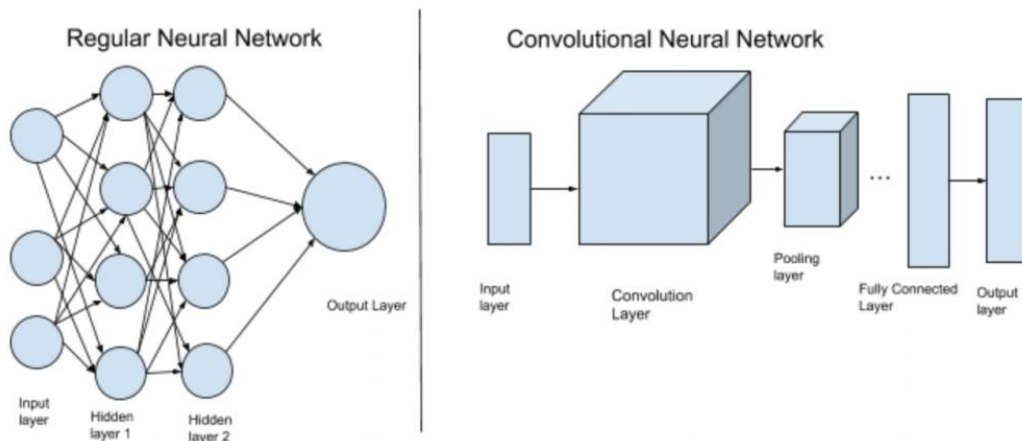
Τα συνελκτικά νευρωνικά δίκτυα χρησιμοποιούνται συνήθως σε εφαρμογές όπου τα δεδομένα είναι εικόνες, όπως προβλήματα ταξινόμησης ή αναγνώρισης. Αυτό οφείλεται στην ικανότητά τους να μειώνουν τον αριθμό των παραμέτρων, διευκολύνοντας έτσι την εκπαίδευση των δεδομένων. Έμπνευση για τη σχεδίασή τους αποτέλεσε η βιολογία και συγκεκριμένα οι έρευνες που έγιναν από τους Hubel και Wiesel. Οι ερευνητές παρατήρησαν ότι ο οπτικός φλοιός της γάτας διαθέτει μία περίπλοκη διάταξη νευρώνων. Η διάταξη αυτή επιβάλλει τη σύνδεση των νευρώνων που είναι ευαίσθητοι σε συγκεκριμένες πολύ μικρές υπό-περιοχές του ορατού πεδίου, γνωστές ως περιοχές ευαισθητοποίησης (receptive fields). Οι συγκεκριμένοι νευρώνες λειτουργούν ως τοπικά «φίλτρα» στις υπό-περιοχές αυτές με αποτέλεσμα να βοηθούν τις γάτες να εκμεταλλεύονται πλήρως τη χωρική συσχέτιση διαφόρων υπό-περιοχών ώστε να έχουν την καλύτερη αντίληψη του χώρου.

3.2.1 Αρχιτεκτονική Συνελκτικών Δικτύων

Τα συνελκτικά νευρωνικά δίκτυα είναι δίκτυα πρόσθιας τροφοδότησης (feed-forward networks) και μοιάζουν με τα κλασικά νευρωνικά δίκτυα. Η διαφορά τους έγκειται στο ότι τα νευρωνικά δίκτυα λόγω της πλήρους συνδεσιμότητας δεν μπορούν να αντιμετωπίσουν με επιτυχία προβλήματα που αφορούν εικόνες μεγάλης διάστασης, εξαιτίας του μεγάλου

αριθμού παραμέτρων. Από την άλλη πλευρά τα συνελκτικά νευρωνικά δίκτυα διαθέτουν νευρώνες που έχουν τρεις διαστάσεις (πλάτος – ύψος - βάθος) και όπως αναφέρθηκε και προηγουμένως εκμεταλλεύονται την ιδιαιτερότητα να συνδέονται σε μία μικρή υπό-περιοχή του προηγούμενου επιπέδου κι όχι με όλους τους νευρώνες.

Στο παρακάτω σχήμα απεικονίζονται στα αριστερά η αρχιτεκτονική ενός κλασσικού νευρωνικού δικτύου, όπου οι νευρώνες κάθε επιπέδου συνδέονται πλήρως μεταξύ τους και στα δεξιά η αρχιτεκτονική ενός συνελκτικού δικτύου.



Εικόνα 3-5: Δομή ενός νευρωνικού δικτύου (αριστερά) με πλήρη διασύνδεση και δομή ενός συνελκτικού (δεξιά). Πηγή: <https://www.topcoder.com/>

Αξίζει να τονιστεί ότι κάθε επίπεδο ενός συνελκτικού νευρωνικού δικτύου μετασχηματίζει τον τρισδιάστατο όγκο εισόδου (input volume), σε έναν τρισδιάστατο όγκο εξόδου (output volume), που παράγεται από τις ενεργοποιήσεις των νευρώνων του.

3.2.2 Συστατικά Μέρη

Οι βασικές λειτουργίες που εκτελούνται στα συνελκτικά νευρωνικά δίκτυα είναι οι παρακάτω:

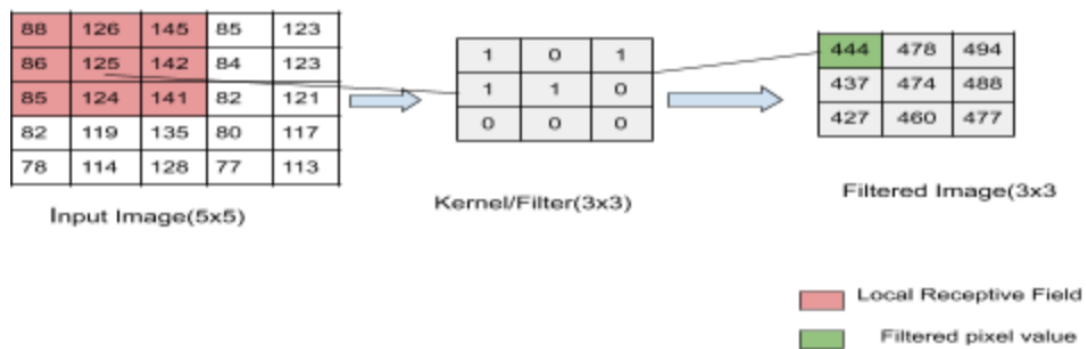
1. *Συνέλιξη*
2. *Μη γραμμικότητα*
3. *Συγκέντρωση ή υπό-δειγματοληψία*
4. *Κατηγοριοποίηση από πλήρως διασυνδεδεμένο επίπεδο*

Παρακάτω γίνεται μία σύντομη αναφορά στα επίπεδα νευρώνων που χρησιμοποιούνται στην αρχιτεκτονική των νευρωνικών δικτύων, με σκοπό την εκτέλεση των λειτουργιών που αναφέρθηκαν.

3.2.2.1 Επίπεδο συνέλιξης (Convolution Layer)

Η λειτουργία του συνελκτικού επιπέδου είναι η εξαγωγή χωρικών χαρακτηριστικών ή μοτίβων από την εικόνα εισόδου. Το συγκεκριμένο επίπεδο δημιουργείται εφαρμόζοντας μια σειρά διαφορετικών φίλτρων, γνωστών ως συνελκτικών πυρήνων, στην εικόνα που εισάγεται στο δίκτυο. Με τον όρο φίλτρο εννοούμε ένα μικρό πλέγμα τιμών με τη μορφή πινάκων, που ολισθαίνει πάνω από την εικόνα εισόδου ανά εικονοστοιχείο για να παράγει μια φιλτραρισμένη εικόνα εξόδου που θα έχει το ίδιο μέγεθος με την αρχική εικόνα. Κάθε πυρήνας θα εξαγάγει ένα διαφορετικό χαρακτηριστικό από μια εικόνα εισόδου, οπότε πολλοί πυρήνες θα παράγουν διαφορετικές φιλτραρισμένες εικόνες εξόδου και τελικά αυτά τα χαρακτηριστικά θα βοηθήσουν στην ταξινόμηση της εικόνας εισόδου.

Τα φίλτρα συνήθως έχουν μικρότερες διαστάσεις από αυτές της αρχικής εικόνας, αλλά διατηρούν τη διάσταση του βάθους ίδια με αυτές. Οι τιμές των φίλτρων/πυρήνων ονομάζονται βάρη και καθορίζουν πόσο σημαντικό είναι το κάθε εικονοστοιχείο για την παραγωγή της τελικής εικόνας.



Εικόνα 3-6: Διαδικασία συνέλιξης. Πηγή: <https://www.topcoder.com/>

Το μέγεθος της εικόνα εξόδου υπολογίζεται από την σχέση:

$$\frac{(I+2*P-F)}{s} + 1 ,$$

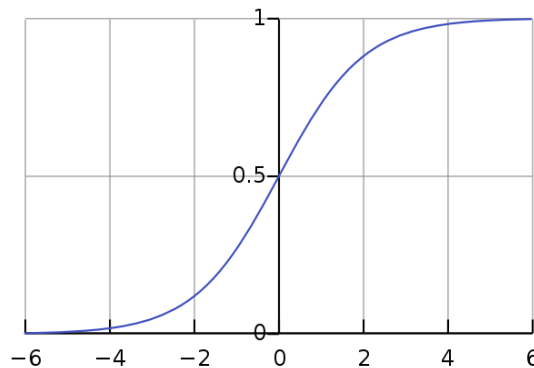
όπου \mathbf{I} = διάσταση της εικόνας εισόδου, \mathbf{P} = γέμισμα περιθωρίου με μηδενικά (Padding), \mathbf{F} = διαστάσεις του φίλτρου και \mathbf{S} = Άλμα (Stride).

3.2.2.2 Συναρτήσεις Ενεργοποίησης (Activation Functions)

Οι συναρτήσεις ενεργοποίησης είναι κόμβοι που χρησιμοποιούν μη-γραμμικές συναρτήσεις, με σκοπό να αποφασίσουν αν ένας νευρώνας πρέπει να ενεργοποιηθεί ή όχι. Ένα νευρωνικό δίκτυο χρειάζεται να προσθέσει μη-γραμμικές συναρτήσεις για να έχει ακριβή αποτελέσματα. Παρακάτω παρουσιάζονται συνοπτικά οι κυριότερες συναρτήσεις ενεργοποίησης που χρησιμοποιούνται.

- Σιγμοειδής (Sigmoid): Η σιγμοειδής είναι η συνάρτηση που χρησιμοποιείται πιο συχνά στα τεχνητά νευρωνικά δίκτυα. Χρησιμοποιείται για να περιγράψει και να μετασχηματίσει μη γραμμικά δεδομένα. Η συγκεκριμένη συνάρτηση παίρνει μία πραγματική τιμή και την περιορίζει στο διάστημα $[0, 1]$. Συνήθως χρησιμοποιείται η λογιστική μορφή της, που έχει την εξής μαθηματική μορφή και γραφική παράσταση:

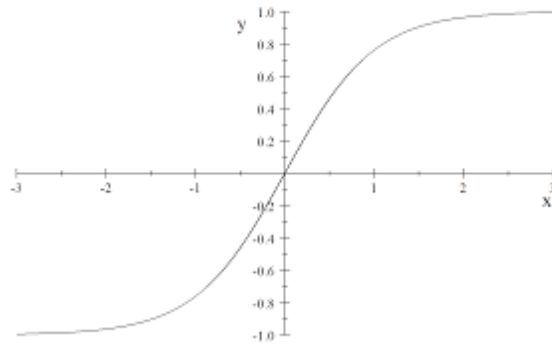
$$\sigma(x) = \frac{1}{1+e^{-x}}$$



Σχήμα 3-1: Γραφική παράσταση σιγμοειδούς συνάρτησης.

- Υπερβολική Εφαπτομένη (Hyperbolic tanh): Η συνάρτηση υπερβολικής εφαπτομένης περιορίζει την τιμή ενός πραγματικού αριθμού στο διάστημα $[-1, 1]$. Όπως φαίνεται και στο διάγραμμα της οι τιμές της είναι κεντραρισμένες στο μηδέν, γεγονός που βοηθάει τον επόμενο νευρώνα στη διαδικασία διάδοσης.

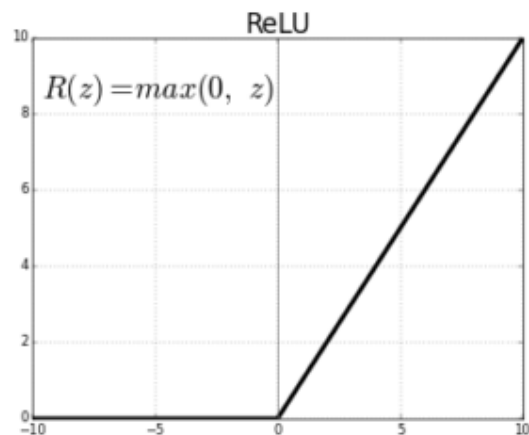
$$f(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$



Σχήμα 3-2: Γραφική παράσταση υπερβολικής εφαπτομένης.

- Rectified Linear Unit (ReLU): Η ReLU χρησιμοποιείται ευρέως ως συνάρτηση ενεργοποίησης κι έχει όριο το μηδέν. Για κάθε αρνητική είσοδο η συνάρτηση δεν ενεργοποιείται, γεγονός που σημαίνει ότι μόνο νευρώνες που πιθανόν επεξεργάζονται σημαντικά χαρακτηριστικά του προβλήματος ενεργοποιούνται.

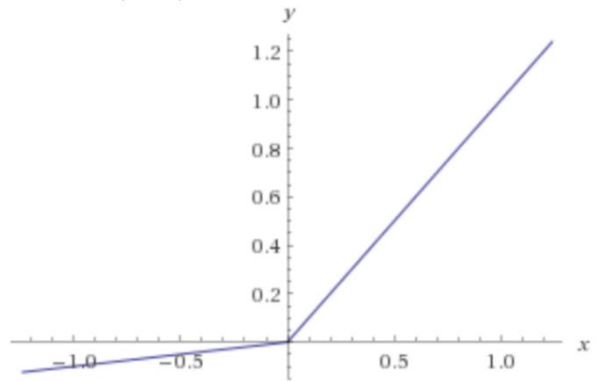
$$f(x) = x^+ = \max(0, x)$$



Σχήμα 3-3: Γραφική παράσταση ReLU.

- Leaky ReLU: Σε αντίθεση με την ReLU η συνάρτηση Leaky ReLU επιτρέπει κατά τη διάρκεια της οπίσθιας διάδοσης μία μικρή αρνητική τιμή. Σύμφωνα με τη σχέση της, το a είναι μία μικρή σταθερά κι έχει ως στόχο να ξεπεράσει το πρόβλημα της ReLU να είναι απενεργοποιημένη για αρνητικές εισόδους.

$$f(x) = \begin{cases} x, & \text{αν } x \geq 0 \\ a * x, & \text{αν } x < 0 \end{cases}$$

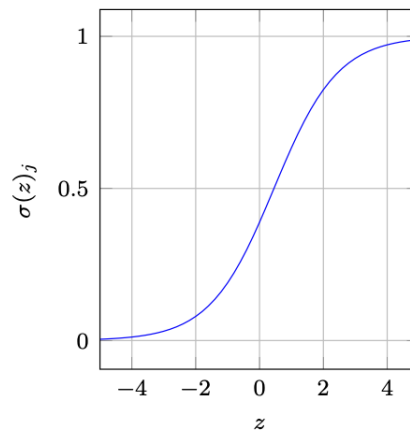


Σχήμα 3-4: Γραφική παράσταση Leaky ReLU.

- **Softmax:** Η συνάρτηση softmax χρησιμοποιείται ευρέως στα βαθιά νευρωνικά δίκτυα κι ο λόγος είναι διότι οι προηγούμενες συναρτήσεις δεν μπορούν να χρησιμοποιηθούν σε προβλήματα κατηγοριοποίησης με παραπάνω από δύο κλάσεις. Γι' αυτές τις περιπτώσεις κατάλληλη είναι η συνάρτηση softmax η οποία είναι μία γενίκευση της λογιστικής συνάρτησης. Χρησιμοποιείται στο επίπεδο εξόδου ενός δικτύου και η λειτουργία της είναι να συμπυκνώνει τις τιμές ώστε να είναι μεταξύ του εύρους $[0, 1]$ και το άθροισμά τους να είναι ίσο με τη μονάδα. Η τιμή εξόδου που προκύπτει από τη συνάρτηση softmax είναι ισοδύναμη με μία κατηγορική συνάρτηση πιθανότητας.

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}},$$

for $i=1, \dots, K$ and $z=(z_1, \dots, z_k) \in R^k$



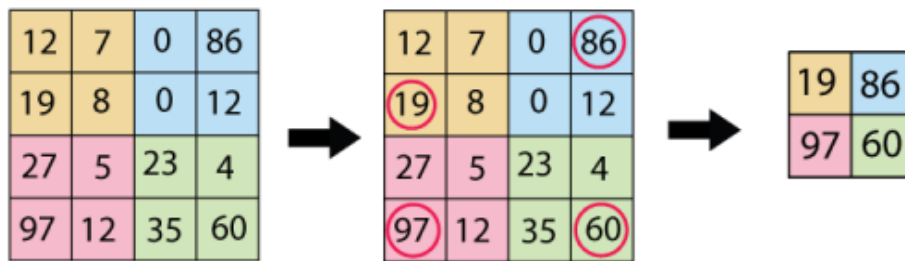
Σχήμα 3-5: Γραφική παράσταση συνάρτησης softmax.

3.2.2.3 Επίπεδο Συγκέντρωσης (Pooling)

Τα επίπεδα συγκέντρωσης τοποθετούνται συνήθως μετά από συνελκτικά επίπεδα. Η λειτουργία τους είναι να μειώσουν το πλήθος των παραμέτρων, μειώνοντας τη διάσταση του

δικτύου και άρα και την υπολογιστική του πολυπλοκότητα. Δέχονται ως είσοδο ένα πλέγμα από το προηγούμενο συνελκτικό επίπεδο και έχουν ως έξοδο μία μοναδική τιμή που μπορεί να είναι συνήθως είτε η μέγιστη τιμή του πλέγματος, είτε η μέση τιμή. Τα επίπεδα που επιστρέφουν την μέγιστη τιμή ονομάζονται επίπεδα μέγιστης συγκέντρωσης (Max Pooling) και χρησιμοποιούνται πιο συχνά.

Στα επίπεδα μέγιστης συγκέντρωσης διαλέγεται μόνο η μέγιστη τιμή από το πλέγμα με τη χρήση ενός φίλτρου που «γλιστρά» στην εικόνα εισόδου και κάθε φορά κρατείται μόνο η μέγιστη παράμετρος. Η διαδικασία φαίνεται στην παρακάτω εικόνα.

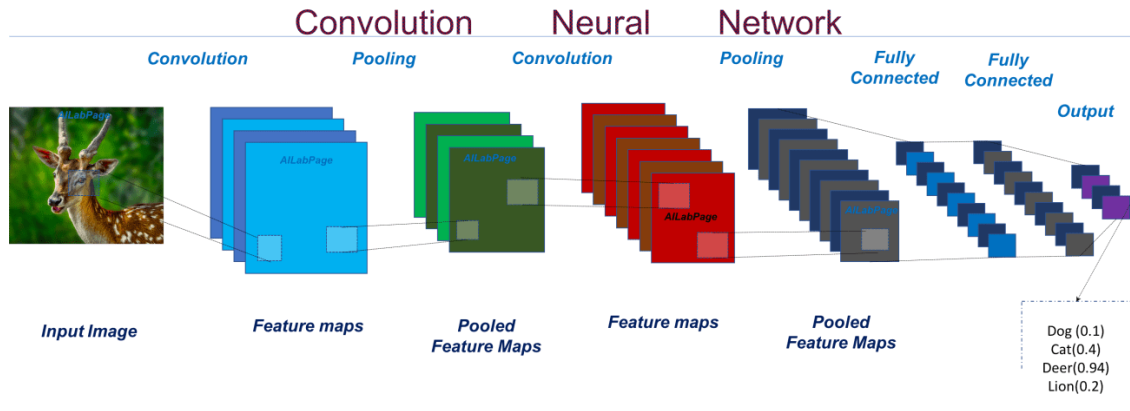


Εικόνα 3-7: Διαδικασία Max Pooling, με χρήση 2*2 φίλτρου.

Πηγή: <https://principlesofdeeplearning.com/>

3.2.2.4 Πλήρως συνδεδεμένο επίπεδο

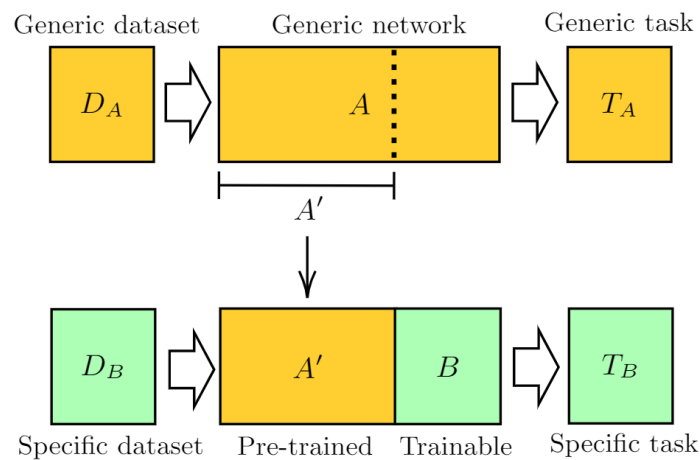
Το τελευταίο επίπεδο σε ένα συνελκτικό νευρωνικό δίκτυο είναι το πλήρως συνδεδεμένο. Σε αυτό το επίπεδο οι νευρώνες έχουν πλήρεις συνδέσεις με όλες τις εξόδους των νευρώνων του προηγούμενου επιπέδου. Σκοπός του συγκεκριμένου επιπέδου είναι να εκτελέσει την ταξινόμηση βάσει των χαρακτηριστικών που εξάχθηκαν από τα προηγούμενα επίπεδα. Στην παρακάτω εικόνα φαίνεται όλη η αρχιτεκτονική ενός συνελκτικού νευρωνικού δικτύου, με βάση τα στοιχεία που αναφέρθηκαν παραπάνω.



Εικόνα 3-8: Αρχιτεκτονική συνελκτικού νευρωνικού δικτύου. Πηγή: <https://vinodsblog.com>

3.3 Μεταφορά Μάθησης

Στη μηχανική μάθηση (και ειδικά στη βαθιά μάθηση - deep learning), αντιμετωπίζουμε ένα πολύ σημαντικό πρόβλημα. Αυτό είναι το γεγονός ότι τα δίκτυα που επιλύουν περίπλοκα προβλήματα απαιτούν τεράστιες ποσότητες δεδομένων. Ωστόσο, η απόκτηση αυτών των δεδομένων για τα επιβλεπόμενα μοντέλα είναι συχνά ανέφικτη λόγω χρονικών ή υπολογιστικών περιορισμών. Επιπλέον, τα μοντέλα που έχουν εκπαιδευτεί σε μικρά σύνολα δεδομένων έχουν χειρότερη απόδοση όταν χρησιμοποιούνται για να αντιμετωπίσουν ένα διαφορετικό πρόβλημα, το οποίο μπορεί να είναι σχετικά παρεμφερές με το πρόβλημα στο οποίο έχουν εκπαιδευτεί. Ο στόχος της μεταφοράς μάθησης είναι να βελτιώσει την εκμάθηση του προβλήματος-στόχου (specific task) αξιοποιώντας γνώση από το πρόβλημα-πηγή (generic task).

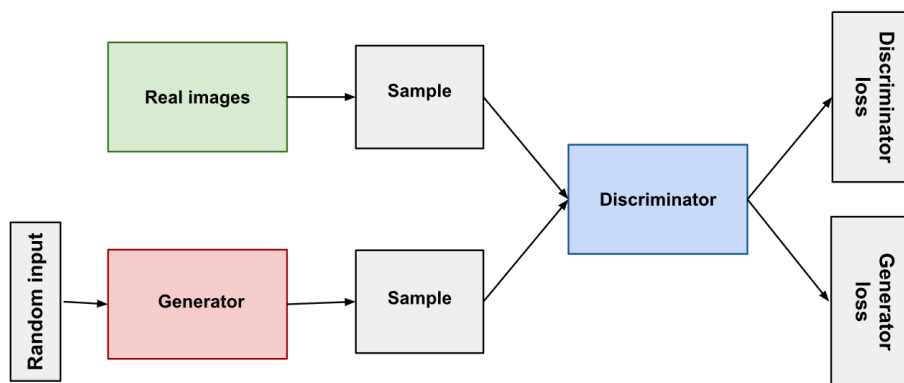


Εικόνα 3-9: Αρχή της Μεταφοράς Μάθησης. Πηγή: <https://pennylane.ai/>

3.4 Γεννητικά Ανταγωνιστικά Δίκτυα (Generative Adversarial Networks)

Εκτός από τα μοντέλα που δέχονται ως είσοδο δεδομένα και παράγουν ως έξοδο πληροφορίες γι' αυτά τα δεδομένα, όπως για παράδειγμα σε ποια κλάση ανήκουν, υπάρχουν και τα γεννητικά μοντέλα, τα οποία δέχονται ως είσοδο κάποια πληροφορία κι έχουν ως στόχο να δημιουργήσουν νέα, «ψεύτικα» δεδομένα που μοντελοποιούν όσο το δυνατόν καλύτερα την πληροφορία αυτή. Στην κατηγορία των γεννητικών μοντέλων ανήκουν και τα μοντέλα τα οποία έχουν επιλεγθεί για την υλοποίηση των πειραμάτων που θα παρουσιαστούν στο κεφάλαιο 4.

Τα Γεννητικά Ανταγωνιστικά Δίκτυα προτάθηκαν το 2014 από τον Ian Goodfellow και τους συναδέλφους του, αρχικά στα πλαίσια της μη επιβλεπόμενης μάθησης, με σκοπό τη δημιουργία τεχνητών δεδομένων. Ως προς την αρχιτεκτονική τους αποτελούνται από δύο δίκτυα, τον δημιουργό (generator) και τον διευκρινιστή (discriminator). Η είσοδος του δημιουργού είναι συνήθως ένα διάνυσμα τυχαίων μεταβλητών που ακολουθούν μία συγκεκριμένη κατανομή, ενώ η έξοδός του, τα τεχνητά δεδομένα, προωθείται μαζί με αληθινά δεδομένα για αξιολόγηση στον διευκρινιστή. Η έξοδος του διευκρινιστή, τουλάχιστον στην αρχική υλοποίηση της αρχιτεκτονικής, ήταν η πιθανότητα ένα δείγμα να προέρχεται από τα δεδομένα εκπαίδευσης. Σκοπός του διευκρινιστή είναι να μάθει να ξεχωρίζει με όσο το δυνατόν περισσότερη ακρίβεια τα αληθινά δεδομένα από αυτά που παράγει ο δημιουργός, ενώ ο σκοπός του δημιουργού είναι με τη σειρά του να μάθει την κατανομή των δεδομένων και να «ξεγελάσει» τον διευκρινιστή, να παράγει δηλαδή δεδομένα τα οποία ο διευκρινιστής δεν μπορεί να ξεχωρίσει από τα αληθινά. Η τυπική δομή μίας αρχιτεκτονικής GAN φαίνεται στην παρακάτω εικόνα.



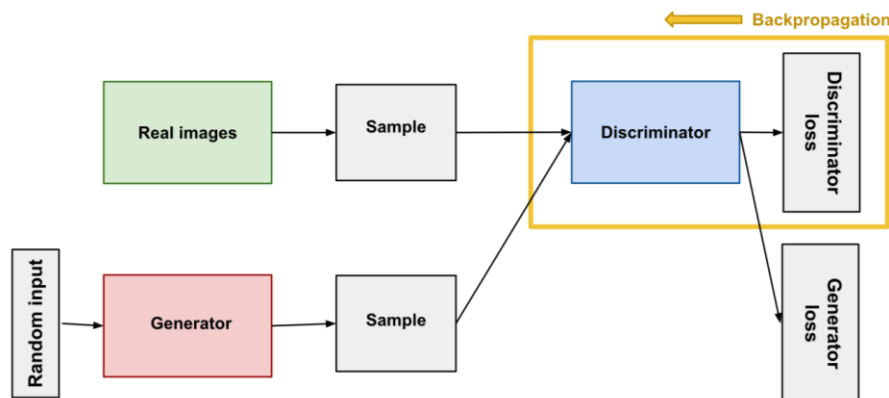
Εικόνα 3-10: Τυπική αρχιτεκτονική Γεννητικών Ανταγωνιστικών Δικτύων.

Πηγή: <https://developers.google.com/>

Ο δημιουργός και ο διευκρινιστής είναι και οι δύο νευρωνικά δίκτυα. Η έξοδος του δημιουργού συνδέεται απευθείας με την είσοδο του διευκρινιστή και μέσω της διαδικασίας οπισθοδρομικής διάδοσης σφάλματος, ο δημιουργός χρησιμοποιεί την απόφαση του διευκρινιστή για να ενημερώσει τα βάρη του.

3.4.1 Διευκρινιστής (Discriminator)

Ο διευκρινιστής σε ένα GAN είναι απλά ένας ταξινομητής. Προσπαθεί να διακρίνει τα πραγματικά δεδομένα, από τα δεδομένα που δημιουργούνται από τον δημιουργό. Ως προς την αρχιτεκτονική του μπορεί να χρησιμοποιεί οποιαδήποτε αρχιτεκτονική δικτύου κατάλληλη για τον τύπο δεδομένων που ταξινομεί.



Εικόνα 3-11: Διαδικασία οπισθοδρομικής διάδοσης σφάλματος κατά την εκπαίδευση του διευκρινιστή. Πηγή: <https://developers.google.com/>

Όπως παρατηρείται στην παραπάνω εικόνα τα δεδομένα εκπαίδευσης του διευκρινιστή προέρχονται από δύο πηγές. Πρόκειται για τα αληθινά δεδομένα και για τα «ψεύτικα/πλαστά» δεδομένα τα οποία δημιουργούνται από τον δημιουργό. Ο διευκρινιστής χρησιμοποιεί αυτά τα δεδομένα ως αρνητικά παραδείγματα κατά την εκπαίδευση. Κατά τη διάρκεια εκπαίδευσης του διευκρινιστή, ο δημιουργός δεν εκπαιδεύεται. Τα βάρη του παραμένουν σταθερά, ενώ παράγει παραδείγματα για τον διευκρινιστή. Παρά του ότι, ο διευκρινιστής συνδέεται με δύο συναρτήσεις κόστους, κατά την εκπαίδευση του αγνοεί εκείνη που αφορά τον δημιουργό και χρησιμοποιεί μόνο το κόστος του διευκρινιστή. Κατά τη διάρκεια εκπαίδευσης του διευκρινιστή συμβαίνουν τα ακόλουθα:

- Ο διευκρινιστής ταξινομεί τα αληθινά και τα «ψεύτικα» δεδομένα από τον δημιουργό.

- Η απώλεια του διευκρινιστή «τιμωρεί» των διευκρινιστή για την λάθος ταξινόμηση ενός αληθινού δεδομένου ως «ψεύτικου» και αντίστοιχα ενός «ψεύτικου» δεδομένου ως αληθινού.
- Ο διευκρινιστής ενημερώνει τα βάρη του μέσω της οπισθοδρομικής διάδοσης σφάλματος.

3.4.2 Δημιουργός (Generator)

Ο δημιουργός μαθαίνει να δημιουργεί ψεύτικα δεδομένα από την ανατροφοδότηση του διευκρινιστή, με απλά λόγια ο στόχος του είναι να μάθει τον διευκρινιστή να ταξινομεί την έξοδο του ως αληθινό δεδομένο. Η εκπαίδευση του δημιουργού αποτελείται από τα παρακάτω συστατικά:

- Τυχαίος θόρυβος. Τα γεννητικά ανταγωνιστικά δίκτυα δέχονται τυχαίο θόρυβο ως είσοδο. Στη συνέχεια, ο δημιουργός μετατρέπει αυτό τον θόρυβο σε μία έξοδο με σημασία. Με το να εισάγουμε θόρυβο στα γεννητικά ανταγωνιστικά δίκτυα τα αναγκάζουμε στο να παράγουν μεγάλη ποικιλία δεδομένων, κάνοντας δειγματοληψία από διαφορετικά μέρη στην κατανομή στόχου.
- Δίκτυο δημιουργού, το οποίο μετατρέπει τον τυχαίο θόρυβο σε δεδομένα.
- Δίκτυο διευκρινιστή, ο οποίος ταξινομεί τα δεδομένα που παράγονται.
- Έξοδος διευκρινιστή.
- Απώλεια δημιουργού, η οποία «τιμωρεί» τον δημιουργό για την αποτυχία του να «κοροϊδέψει» τον διευκρινιστή.

3.4.3 Γνωστές Αρχιτεκτονικές Γεννητικών Ανταγωνιστικών Δικτύων

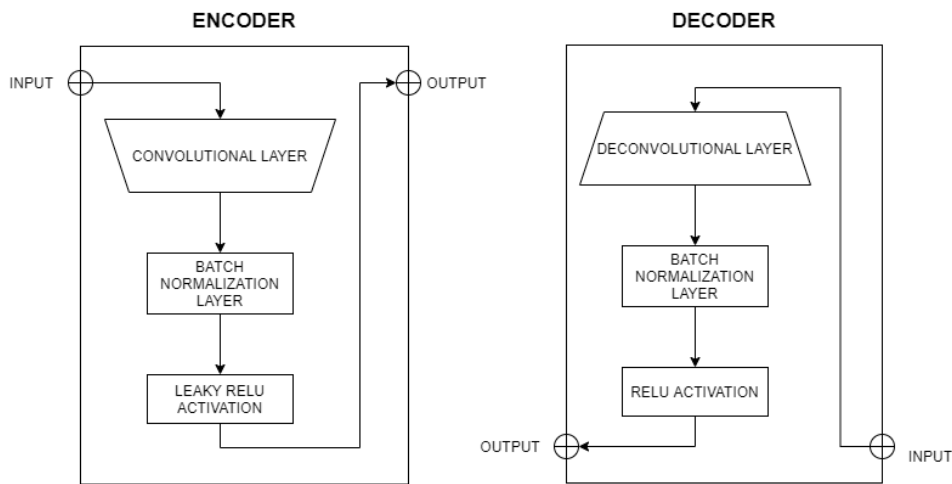
Στις παρακάτω υποενότητες παρουσιάζονται γνωστές αρχιτεκτονικές για τον δημιουργό και τον διευκρινιστή, στις οποίες έχουν βασιστεί τα μοντέλα τα οποία χρησιμοποιήθηκαν στο κεφάλαιο της μεθοδολογίας.

3.4.3.1 U-Net Δημιουργός

Η αρχιτεκτονική U-Net αποτελείται από τρία τμήματα: α) τα επίπεδα κωδικοποίησης, β) το σημείο συμφόρησης (bottleneck) και γ) τα επίπεδα αποκωδικοποίησης. Είναι χαρακτηριστική για τις συνδέσεις παράλειψης, οι οποίες συνενώνουν την έξοδο των επιπέδων κωδικοποίησης, με τις αντίστοιχες εξόδους των επιπέδων αποκωδικοποίησης. Πιο συγκεκριμένα, η σύνδεση

παράλειψης από ένα επίπεδο κωδικοποίησης που έχει διαστάσεις εξόδου $128 * 128 * 256$ θα συνενωθεί με ένα επίπεδο του αποκωδικοποιητή που έχει αντίστοιχα τις ίδιες διαστάσεις. Δεδομένου ότι το U-Net είναι συμμετρικό δίκτυο, κάθε επίπεδο αποκωδικοποίησης θα έχει ένα αντίστοιχο επίπεδο κωδικοποίησης για τη διευκόλυνση των συνδέσεων παράλειψης.

Η εικόνα εισόδου περνάει από μια σειρά κωδικοποιητών που προσπαθούν να εξάγουν χαρακτηριστικά από αυτήν, συμπιέζοντάς την (downsampling) σε μια πολύ μικρότερη αναπαράσταση, μέχρι να φτάσει σε ένα επίπεδο συμφόρησης (bottleneck layer). Έπειτα, το αποτέλεσμα των κωδικοποιητών περνάει από ένα δίκτυο αποκωδικοποιητών, με σκοπό να το αποσυμπιέσουν μέχρι να φτάσει στο μέγεθος της αρχικής αναπαράστασης. Στο παρακάτω σχήμα φαίνεται πως είναι ένα μπλοκ του κωδικοποιητή και του αποκωδικοποιητή.



Σχήμα 3-6: Μοντέλο κωδικοποιητή (αριστερά), και αποκωδικοποιητή (δεξιά).

3.4.3.2 Patch-GAN Διευκρινιστής

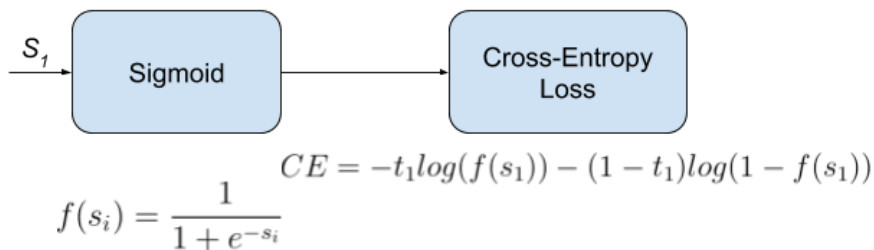
Η αρχιτεκτονική PatchGAN βασίζεται στην λογική ότι ο διευκρινιστής δεν χρειάζεται να προβλέψει αν ολόκληρη η εικόνα είναι ψεύτικη ή πραγματική, αλλά παίρνει ένα τετράγωνο τμήμα (patch) της εικόνας διαστάσεων $N * N$ και προβλέπει κάθε εικονοστοιχείο σε αυτό το τμήμα εάν είναι πραγματικό ή ψεύτικο. Κάθε τιμή του πίνακα εξόδου αντιπροσωπεύει την πιθανότητα εάν το αντίστοιχο patch της εικόνας είναι πραγματικό ή έχει δημιουργηθεί από τον δημιουργό. Με αυτό τον τρόπο ο διευκρινιστής μπορεί να λάβει υπόψη του τα χαρακτηριστικά υψηλών συχνοτήτων της εικόνας.

3.4.4 Συναρτήσεις Κόστους Γεννητικών Ανταγωνιστικών Δικτύων

Για τις συναρτήσεις κόστους των γεννητικών ανταγωνιστικών δικτύων έχουν δημιουργηθεί αρκετές παραλλαγές ανάλογα με το πρόβλημα που εξετάζει η κάθε εφαρμογή. Στις παρακάτω υποενότητες παρουσιάζονται με συνοπτικό τρόπο οι συναρτήσεις κόστους που χρησιμοποιούνται στα μοντέλα που περιγράφονται στο κεφάλαιο της μεθοδολογίας.

3.4.4.1 Σιγμοειδής Εγκάρσια Εντροπία (Sigmoid Cross Entropy)

Πρόκειται για μία σιγμοειδή συνάρτηση ενεργοποίησης σε συνδυασμό με μία συνάρτηση κόστους εγκάρσιας εντροπίας. Είναι ανεξάρτητη για κάθε κλάση δεδομένων, που σημαίνει ότι η απώλεια που υπολογίζεται για κάθε διάνυσμα από ένα δίκτυο δεν επηρεάζεται από άλλες τιμές συνιστωσών.



Εικόνα 3-12: Υπολογισμός σιγμοειδούς εγκάρσιας εντροπίας για δυαδικό πρόβλημα ταξινόμησης. Πηγή: <https://gombu.github.io/>

3.4.4.2 Μέσο Απόλυτο Σφάλμα (Mean Absolute Error - MAE)

Σε αυτή τη συνάρτηση κόστους η απώλεια είναι η μέση τιμή των απόλυτων διαφορών μεταξύ αληθινών και προβλεπόμενων τιμών. Δεν είναι ευαίσθητη έναντι των ακραίων τιμών. Ένα μειονέκτημα του μέσου απόλυτου σφάλματος είναι ότι το μέγεθος της κλίσης δεν εξαρτάται από το μέγεθος του σφάλματος, αλλά μόνο από το πρόσημο του $y - \hat{y}$. Αυτό οδηγεί στο ότι το μέγεθος της κλίσης θα είναι μεγάλο ακόμη και όταν το σφάλμα είναι μικρό, το οποίο με τη σειρά του μπορεί να οδηγήσει σε προβλήματα σύγκλισης. Σε προβλήματα ανακατασκευής εικόνας ενθαρρύνει τον δημιουργό να κατασκευάζει λιγότερο θολές εικόνες.

$$L(y, \hat{y}) = \frac{1}{N} \sum_{i=0}^N |y - \hat{y}_i|$$

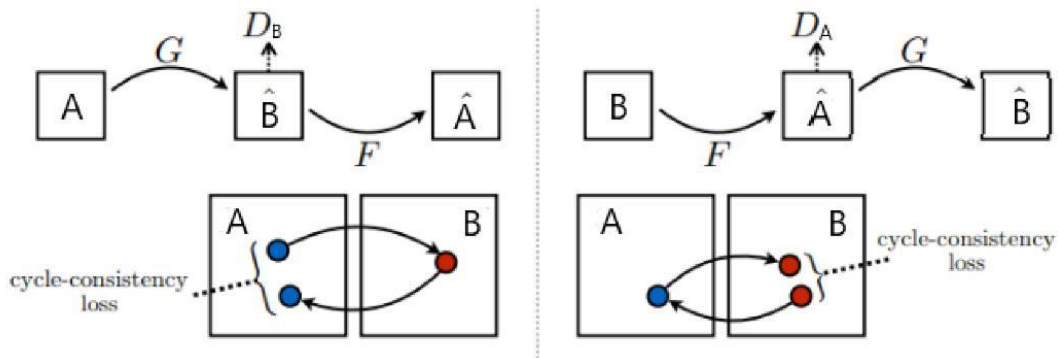
Όπου \hat{y} οι τιμές πρόβλεψης.

3.4.4.3 Απώλεια Συνέπειας Κύκλου και Απώλεια Ταυτότητας (Cycle consistency loss and Identity loss)

Στην απώλεια συνέπειας στον κύκλο υπολογίζεται η απώλεια μέσου απόλυτου σφάλματος (mean absolute error - MAE) μεταξύ της δημιουργούμενης εικόνας από τον δημιουργό G και της δημιουργημένης εικόνας από τον F.

$$\text{forward cycle consistency loss} : X \rightarrow G(X) \rightarrow F(G(X)) \sim \hat{X}$$

$$\text{backward cycle consistency loss} : Y \rightarrow F(Y) \rightarrow G(F(Y)) \sim \hat{Y}$$



Σχήμα 3-7: Απώλεια Cycle-consistency. Πηγή: <https://www.tensorflow.org/>

Η identity loss (απώλεια ταυτότητας) λέει ότι εφόσον ο δημιουργός G είναι υπεύθυνος για την μετατροπή της εικόνας X στην εικόνα Y, τότε αν τροφοδοτήσουμε την εικόνα Y στον δημιουργό G, αυτός θα πρέπει να παράγει την πραγματική εικόνα Y ή μία εικόνα πολύ κοντά στην Y.

3.4.4.4 Απώλεια Μάσκας (Masked Loss)

Η απώλεια μάσκας έχει χρησιμοποιηθεί σε εφαρμογές ανακατασκευής βίντεο, όπου τμήματα του καλύπτονται με μάσκα. Εστιάζει στα χαρακτηριστικά των εικονοστοιχείων, δεδομένου ότι η μη καλυμμένη περιοχή θα επικολληθεί στο τελικό αποτέλεσμα εξόδου. Η εξίσωση για την απώλεια της καλυμμένης περιοχής είναι:

$$L_{l1_{mask}} = \mathbb{E}_{t,x,y} [M_{t,x,y} |O_{t,x,y} - V_{t,x,y}|]$$

3.4.4.5 Αντιληπτική Απώλεια (Perceptual Loss)

Η αντιληπτική απώλεια χρησιμοποιείται για την διατήρηση του περιεχομένου της εικόνας και για τον μετριάσμό της θολότητας που προκαλείται από την απώλεια μάσκας που αναφέρθηκε προηγουμένως. Για την συγκεκριμένη απώλεια χρησιμοποιείται το προεκπαιδευμένο δίκτυο VGG.

$$L_{perc} = \sum_{t=1}^n \sum_{p=0}^{P-1} \frac{|\Psi_p^{O_t} - \Psi_p^{V_t}|}{N_{\Psi_p^{V_t}}}$$

Όπου $\Psi_p^{V_t}$ είναι η ενεργοποίηση από το p -οστό επίπεδο του προεκπαιδευμένου δικτύου για την είσοδο V_t και $N_{\Psi_p^{V_t}}$ ο αριθμός των στοιχείων του p -οστού επιπέδου.

3.4.4.6 Απώλεια Στυλ (Style Loss)

Χρησιμοποιείται για να διατηρηθεί το περιεχόμενο της εικόνας και είναι παρόμοια με την αντιληπτική απώλεια. Η μόνη διαφορά είναι ότι εφαρμόζει πρώτα στα χαρακτηριστικά την αυτόματη συσχέτιση (Gram matrix).

$$L_{style} = \sum_{t=1}^n \sum_{p=0}^{P-1} \frac{1}{C_p C_p} \frac{|(\Psi^{O_{t_p}})^T (\Psi^{O_{t_p}}) - (\Psi^{V_{t_p}})^T (\Psi^{V_{t_p}})|}{C_p H_p W_p}$$

Όπου $\Psi^{O_{t_p}}$ και $\Psi^{V_{t_p}}$ είναι χαρακτηριστικά του δικτύου VGG διαστάσεων (H_p, W_p, C_p) .

3.4.4.7 TSMGAN απώλεια

Χρησιμοποιείται για την διαχείριση υψηλού-επιπέδου χαρακτηριστικών της εικόνας και για να ενισχυθεί η αληθοφάνεια της. Η απώλεια TSMGAN L_D βοηθάει τον διευκρινιστή να αποφανθεί εάν το βίντεο εισόδου z είναι πραγματικό ή ψεύτικο και τον δημιουργό L_G να ξεγελάσει τον διευκρινιστή.

$$\begin{aligned} L_D &= \mathbb{E}_{x \sim P_{data}(x)} [ReLU(1 + D(x))] \\ &\quad + \mathbb{E}_{z \sim P_z(z)} [ReLU(1 - D(G(z)))] \\ L_G &= -\mathbb{E}_{z \sim P_z(z)} [D(G(z))] \end{aligned}$$

3.4.5 Αξιολόγηση Γεννητικών Ανταγωνιστικών Δικτύων – Μετρικές

Στα γεννητικά ανταγωνιστικά δίκτυα τόσο ο δημιουργός όσο και ο διευκρινιστής εκπαιδεύονται μαζί για να διατηρήσουν μια ισορροπία. Ως εκ τούτου, δεν υπάρχει αντικειμενική λειτουργία απώλειας που χρησιμοποιείται για την εκπαίδευση δημιουργού και κανέναν τρόπο αντικειμενικής αξιολόγησης της προόδου της εκπαίδευσης και της σχετικής ή απόλυτης ποιότητας του μοντέλου από την απώλεια μόνο. Αντ' αυτού, αναπτύχθηκε μια σειρά ποιοτικών και ποσοτικών τεχνικών για την αξιολόγηση της απόδοσης ενός γεννητικού ανταγωνιστικού μοντέλου με βάση την ποιότητα και την ποικιλομορφία των παραγόμενων εικόνων.

Στη συνέχεια γίνεται αναφορά σε μετρικές που χρησιμοποιούνται για να αξιολογήσουν την ποιότητα των εικόνων που παράγονται από τον δημιουργό.

1. Μέσο Τετραγωνικό Σφάλμα (Mean Squared Error - MSE): Πρόκειται για την απλούστερη και πιο ευρέως χρησιμοποιημένη μέτρηση για την ποιότητα εικόνας. Υπολογίζει το σφάλμα αφαιρώντας κάθε εικονοστοιχείο της εικόνας εξόδου από το αντίστοιχο εικονοστοιχείο μίας εικόνας αναφοράς. Στη συνέχεια, υπολογίζει το μέσο όρο των σφαλμάτων. Το MSE αντιπροσωπεύει το αθροιστικό τετράγωνο σφάλμα μεταξύ της ανακατασκευασμένης και της αρχικής εικόνας. Ορίζεται ως εξής:

$$MSE = \frac{1}{M * N} \sum_{i=1}^M \sum_{j=1}^N (x(i,j) - y(i,j))^2$$

Όπου $x(i,j)$ η εικόνα αναφοράς, $y(i,j)$ η μετασχηματισμένη εικόνα, i και j οι θέσεις των εικονοστοιχείων στην $M * N$ εικόνα. Το μέσο τετραγωνικό σφάλμα είναι μηδέν όταν ισχύει $x(i,j) = y(i,j)$.

2. Μέγιστη αναλογία σήματος προς θόρυβο (Peak Signal to Noise Ratio - PSNR): Υπολογίζει την μέγιστη αναλογία σήματος προς θόρυβο, σε ντεσιμπέλ, μεταξύ δύο εικόνων. Αυτή η αναλογία χρησιμοποιείται ως μέτρηση ποιότητας μεταξύ της αρχικής και της ανακατασκευασμένης εικόνας. Όσο υψηλότερο είναι το PSNR, τόσο καλύτερη είναι η ποιότητα της ανακατασκευασμένης εικόνας. Το PSNR αντιπροσωπεύει ένα μέτρο του μέγιστου σφάλματος. Για τον υπολογισμό του, υπολογίζεται πρώτα το μέσο τετραγωνικό σφάλμα όπως αναφέρθηκε προηγουμένως και στη συνέχεια το PSNR με την ακόλουθη εξίσωση:

$$PSNR = 10 * \log_{10} \left(\frac{R^2}{MSE} \right)$$

Όπου το R είναι η μέγιστη διακύμανση στον τύπο δεδομένων της εικόνας εισόδου. Για παράδειγμα, εάν η εικόνα εισαγωγής έχει έναν ακέραιο τύπο δεδομένων 8-bit τότε το R είναι 255.

3. Δείκτης Δομικής Ομοιότητας (Structural Similarity Index – SSIM): Πρόκειται για μια μέθοδο μέτρησης της ομοιότητας μεταξύ δύο εικόνων, που βασίζεται σε μια αρχική χωρίς παραμόρφωση εικόνα αναφοράς. Το SSIM είναι ένα μοντέλο βασισμένο στην αντίληψη που θεωρεί την υποβάθμιση της εικόνας ως αντιληπτή αλλαγή στις δομικές πληροφορίες, ενώ ενσωματώνει επίσης σημαντικά χαρακτηριστικά όπως η φωτεινότητα και η αντίθεση των εικόνων. Η διαφορά με άλλες τεχνικές όπως το MSE ή το PSNR είναι ότι υπολογίζει απόλυτα σφάλματα. Οι δομικές πληροφορίες είναι η ιδέα ότι τα εικονοστοιχεία έχουν ισχυρές αλληλεξαρτήσεις, ειδικά όταν είναι χωρικά κοντά. Αυτές οι εξαρτήσεις οπτικά δίνουν σημαντικές πληροφορίες σχετικά με τη δομή των αντικειμένων. Οι τιμές του SSIM κυμαίνονται μεταξύ 0 και 1, όπου το 1 σημαίνει ότι η ανακατασκευασμένη εικόνα ταιριάζει τέλεια με την αρχική εικόνα. Ο Συγκεκριμένος δείκτης υπολογίζεται ως εξής:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1) * (2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1) * (\sigma_x^2 + \sigma_y^2 + c_2)}$$

όπου μ_x ο μέσος όρος του x , μ_y ο μέσος όρος του y , σ_x^2 η διακύμανση του x , σ_y^2 η διακύμανση του y , σ_{xy} η συνδιακύμανση των x και y , c_1/c_2 δύο μεταβλητές για την σταθεροποίηση της διαίρεσης με ασθενή παρανομαστή.

4. Απόσταση Έναρξης Frechet (Frechet Inception Distance - FID): Η FID χρησιμοποιεί το μοντέλο Inception v3. Συγκεκριμένα, το τελευταίο επίπεδο ομαδοποίησης πριν από την ταξινόμηση εξόδου των εικόνων χρησιμοποιείται για τη σύλληψη χαρακτηριστικών όρασης υπολογιστή από μία εικόνα εισόδου. Αυτές οι ενεργοποιήσεις υπολογίζονται για μια συλλογή πραγματικών και δημιουργημένων εικόνων και συνοψίζονται με τον υπολογισμό του μέσου όρου και της συνδιακύμανσης των εικόνων. Η απόσταση μεταξύ των δύο κατανομών υπολογίζεται στη συνέχεια χρησιμοποιώντας την απόσταση Frechet, που ονομάζεται επίσης απόσταση Wasserstein-2. Η βαθμολογία συνοψίζει πόσο παρόμοιες είναι οι δύο ομάδες εικόνων. Οι χαμηλότερες βαθμολογίες δείχνουν ότι οι δύο ομάδες εικόνων είναι πιο παρόμοιες, ή έχουν περισσότερα παρόμοια στατιστικά

στοιχεία, με το τέλειο αποτέλεσμα να είναι 0.0 που δείχνει ότι οι δύο ομάδες εικόνων είναι πανομοιότυπες.

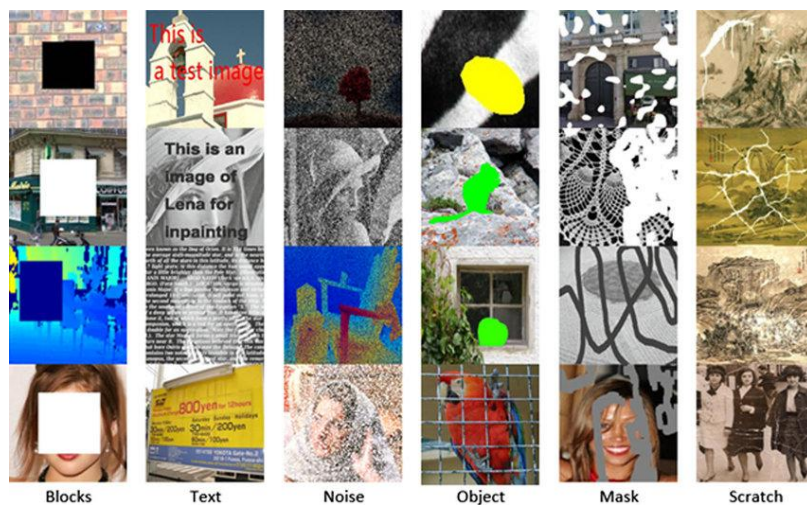
$$FID = \|\mu_r - \mu_g\|^2 + Tr(\Sigma_r + \Sigma_g - 2(\Sigma_r \Sigma_g)^2)$$

Όπου $X_r \sim N(\mu_r, \Sigma_r)$ και $X_g \sim N(\mu_g, \Sigma_g)$ είναι οι 2048-διαστάσεων ενεργοποιήσεις του Inception v3 pool3 επιπέδου για τις πραγματικές και τις δημιουργημένες εικόνες.

3.5 Συμπλήρωση Εικόνας (Image Inpainting)

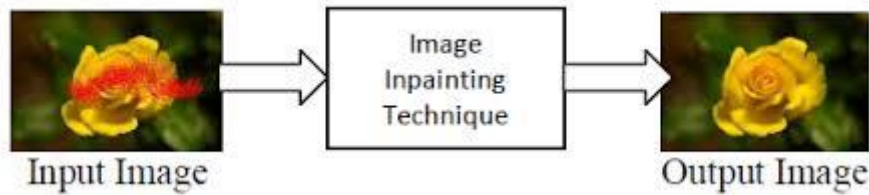
Αν υποθέσουμε ότι έχουμε μία εικόνα και μία κατεστραμμένη περιοχή στο εσωτερικό της τότε το πρόβλημα της συμπλήρωσης εικόνας έγκειται στην τροποποίηση των τιμών των εικονοστοιχείων της εικόνας στην κατεστραμμένη περιοχή με τέτοιο τρόπο έτσι ώστε να μην ξεχωρίζει σε σχέση με τις γειτονικές περιοχές. Σκοπός της συγκεκριμένης τεχνικής είναι η αποκατάσταση κατεστραμμένων τμημάτων μιας εικόνας ή η αφαίρεση ανεπιθύμητων στοιχείων που υπάρχουν στην εικόνα.

Στην παρακάτω εικόνα παρουσιάζονται διάφορες περιπτώσεις που μπορούν να παραμορφώσουν μία εικόνα. Αυτές είναι α) η έλλειψη ολόκληρης περιοχής από την εικόνα, β) η απόκρυψη τμήματος της εικόνας από την εμφάνιση κειμένου πάνω σε αυτήν, γ) η ύπαρξη θορύβου, δ) η εμφάνιση κάποιου αντικειμένου στην εικόνα, ε) οι μάσκες και ζ) οι γραμμές που προκαλούνται από τον χρόνο.



Εικόνα 3-13: Πιθανές παραμορφώσεις εικόνας. Πηγή: <https://www.researchgate.net/>

Στην επόμενη εικόνα εμφανίζεται το αποτέλεσμα μια τεχνικής συμπλήρωσης εικόνας.



Εικόνα 3-14: Αποτέλεσμα τεχνικής συμπλήρωσης εικόνας. Πηγή: Survey on Different Techniques for Image Inpainting - IRJET

3.6 Τηλεπισκοπικά Δεδομένα Sentinel-2

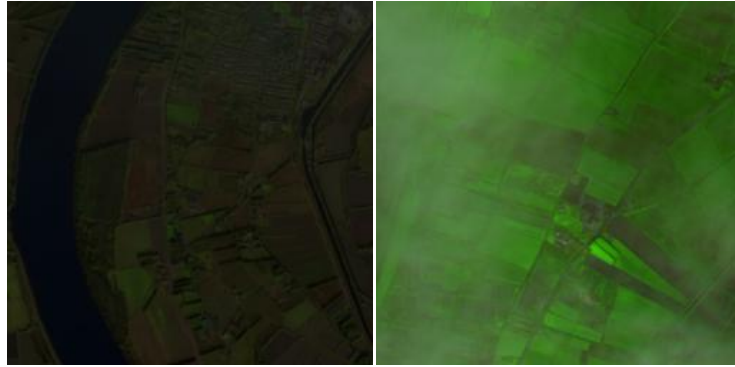
Ο Sentinel-2, είναι μία αποστολή παρακολούθησης της Γης που αναπτύχθηκε από τον Ευρωπαϊκό Οργανισμό Διαστήματος (ESA), ως μέρος του Προγράμματος Copernicus, για την εκτέλεση επίγειων παρατηρήσεων. Αποτελείται από δύο δίδυμους δορυφόρους, τους Sentinel-2A και Sentinel-2B. Οι δύο αυτοί δορυφόροι περιστρέφονται σε αντίθετες πλευρές γύρω από τη Γη. Ο καθένας παρέχει εύρος κάλυψης 290 χιλιομέτρων, επομένως σε ταυτόχρονη λειτουργία μπορούν να απεικονίσουν την υδρόγειο κάθε πέντε ημέρες και την Ευρώπη κάθε δύο με τρεις ημέρες. Κάθε ένας δορυφόρος μεταφέρει έναν οπτικό πολυφασματικό δέκτη MSI (multispectral instrument). Οι επίγειες παρατηρήσεις των δορυφόρων είναι αναγκαίες για την υποστήριξη υπηρεσιών όπως η παρακολούθηση των δασών, η ανίχνευση αλλαγών κάλυψης γης και η διαχείριση φυσικών καταστροφών. Παρέχει φασματικά δεδομένα με 13 ζώνες στο ορατό, στο κοντινό υπέρυθρο και στο μικροκυματικό υπέρυθρο τμήμα του φάσματος και καλύπτει συστηματικά την επιφάνεια της γης από τις 56° N έως τις 84° B, τα παράκτια ύδατα και όλη τη Μεσόγειο θάλασσα. Τα προϊόντα που παράγονται από τον Sentinel-2 έχουν χωρική ανάλυση 10μ, 20μ και 60μ. Στην παρακάτω εικόνα παρουσιάζονται τα 13 κανάλια τα οποία παρέχει ο Sentinel-2.

Sentinel-2 Bands	Central Wavelength (μm)	Resolution (m)
Band 1 - Coastal aerosol	0.443	60
Band 2 - Blue	0.490	10
Band 3 - Green	0.560	10
Band 4 - Red	0.665	10
Band 5 - Vegetation Red Edge	0.705	20
Band 6 - Vegetation Red Edge	0.740	20
Band 7 - Vegetation Red Edge	0.783	20
Band 8 - NIR	0.842	10
Band 8A - Vegetation Red Edge	0.865	20
Band 9 - Water vapour	0.945	60
Band 10 - SWIR - Cirrus	1.375	60
Band 11 - SWIR	1.610	20
Band 12 - SWIR	2.190	20

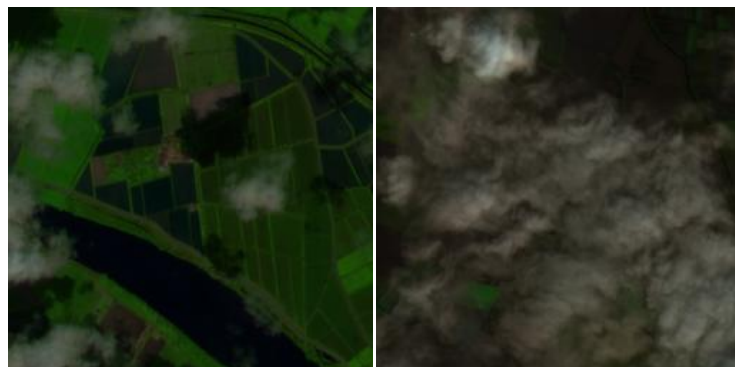
Εικόνα 3-15: Τα φασματικά κανάλια του sentinel-2. Πηγή: <https://www.satimagingcorp.com/>

Πολλές φορές η δυνατότητα χρήσης των τηλεπισκοπικών δεδομένων είναι περιορισμένη. Οι σκιές και τα σύννεφα που υπάρχουν στα δεδομένα καθιστούν την παρατήρηση της επιφάνειας της γης αδύνατη, με αποτέλεσμα ένας μεγάλος αριθμός διαθέσιμων δεδομένων να παραμένει ανεκμετάλλευτος. Η παρουσία σύννεφων και σκιών μπορεί να υπάρχει είτε σε ένα μικρό τμήμα των εικόνων είτε και σε ολόκληρη την εικόνα. Το πρόβλημα των σύννεφων προκαλεί σημαντικά κενά δεδομένων τόσο στους χωρικούς όσο και στους χρονικούς τομείς. Αποτελεί ένα σοβαρό εμπόδιο είτε για εφαρμογές όπου απαιτούνται συνεπείς χρονοσειρές, όπως για παράδειγμα η γεωργική παρακολούθηση, ή όπου μια συγκεκριμένη περιοχή πρέπει να παρατηρηθεί σε ένα συγκεκριμένο χρόνο, π.χ. η παρακολούθηση καταστροφών.

Παρακάτω παρουσιάζονται ενδεικτικά κάποια παραδείγματα τμημάτων τηλεπισκοπικών εικόνων από διαφορετικές περιοχές.



Εικόνα 3-16: Ενδεικτικά παραδείγματα τμήματος εικόνας χωρίς σύννεφα (αριστερά) και τμήματος εικόνας με ιστούς νεφών (δεξιά)

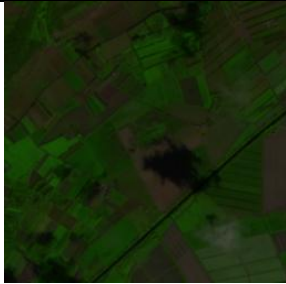
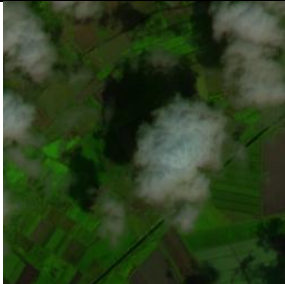
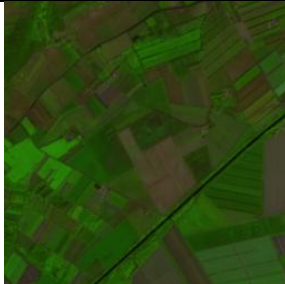
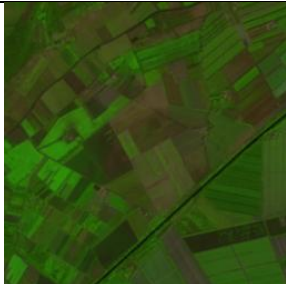
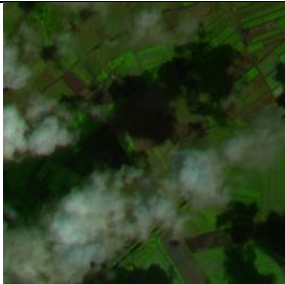
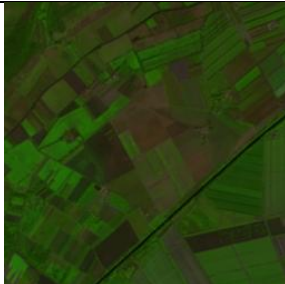
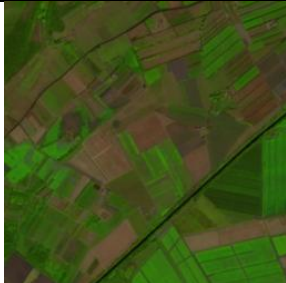


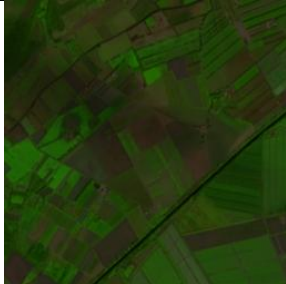
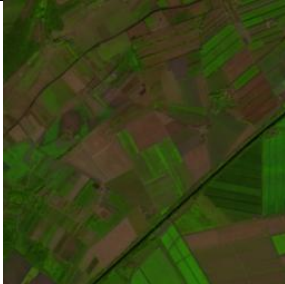
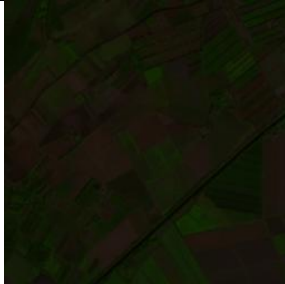





Εικόνα 3-17: Ενδεικτικά παραδείγματα τμήματος εικόνας με λίγα σύννεφα – σκιές (αριστερά) και τμήματος εικόνας με μεγάλο ποσοστό συννεφοκάλυψης (δεξιά)

Παρακάτω παρουσιάζονται ενδεικτικά παραδείγματα μίας χρονοσειράς και απεικονίζεται ακριβώς η ίδια περιοχή στις διαφορετικές χρονικές στιγμές.

14-03-2018	05-04-2018	15-04-2018
28-04-2018	03-05-2018	05-05-2018

ΚΕΦΑΛΑΙΟ 3: ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ

18-05-2018	23-05-2018	25-05-2018
		
09-06-2018	12-06-2018	14-06-2018
		
22-06-2018	27-06-2018	02-07-2018
		
19-07-2018	29-07-2018	21-08-2018
		
17-09-2018	05-10-2018	20-10-2018
		

Είναι εμφανές ότι ακόμα και σε λήψεις που έχουν γίνει τον ίδιο μήνα το ποσοστό συννεφοκάλυψης ποικίλλει με αποτέλεσμα σε πολλές εικόνες να μην είναι δυνατή η

παρατήρηση της επιφάνειας της γης. Λαμβάνοντας υπόψη ότι κάθε μέρα η επιφάνεια τη γης καλύπτεται από σημαντικά ποσοστά σύννεφων, με αποτέλεσμα ένας τεράστιος όγκος δεδομένων να μην μπορεί να χρησιμοποιηθεί, καθίσταται σαφές ότι είναι απαραίτητη η χρήση ενός αλγορίθμου που θα αφαιρεί τα σύννεφα και θα ανακατασκευάζει την εικόνα. Στο επόμενο κεφάλαιο παρουσιάζεται η μεθοδολογία με την οποία έγινε προσπάθεια να αντιμετωπιστεί το συγκεκριμένο πρόβλημα.

4. ΜΕΘΟΔΟΛΟΓΙΑ

Για την επίλυση του προβλήματος που εξετάζει η διπλωματική εφαρμόστηκαν δύο μεθοδολογίες. Η πρώτη αντιμετωπίζει το πρόβλημα ως ένα πρόβλημα εικόνας και η δεύτερη ως χρονοσειρά, παίρνοντας υπόψη και τα χρονικά χαρακτηριστικά. Για την εφαρμογή αυτών των δύο μεθοδολογιών χρησιμοποιήθηκαν και εξετάστηκαν τρία ήδη υπάρχοντα μοντέλα γεννητικών ανταγωνιστικών δικτύων. Παρακάτω ακολουθεί η περιγραφή των δεδομένων, των μοντέλων καθώς και η επιλογή των υπερπαραμέτρων τους. Για την εκτέλεση των πειραμάτων χρησιμοποιήθηκε το Google Colaboratory, το οποίο όμως έχει περιορισμένους υπολογιστικούς πόρους (RAM, GPU) με αποτέλεσμα να αντιμετωπίσουμε αρκετά προβλήματα και να επιβάλουμε αναγκαστικούς περιορισμούς ως προς τον αριθμό των δεδομένων που χρησιμοποιήσαμε για την εκπαίδευση.

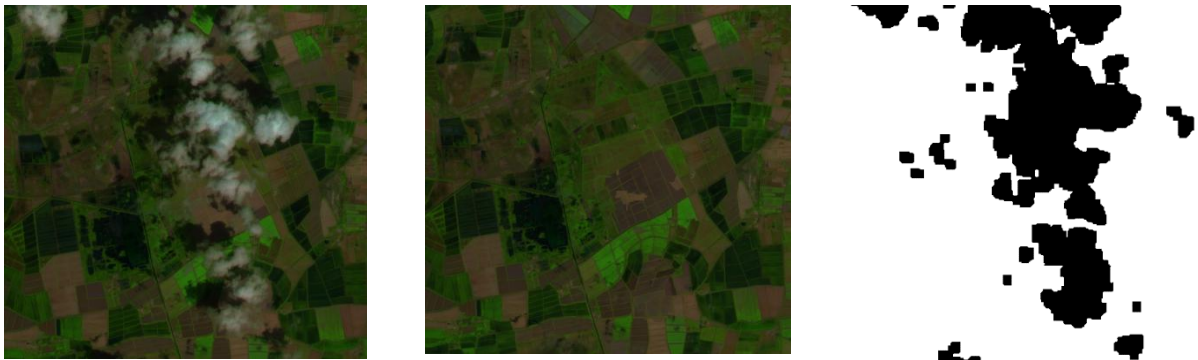
4.1 Περιγραφή συνόλου δεδομένων (Dataset)

Για τη υλοποίηση της παρούσας διπλωματικής εργασίας κρίθηκε απαραίτητο να βρεθεί μία χρονοσειρά δορυφορικών εικόνων, η οποία θα απεικονίζει μία συγκεκριμένη περιοχή για ένα χρονικό διάστημα και θα παρουσιάζει ποικιλία ως προς το ποσοστό συννεφοκάλυψης. Το σύνολο δεδομένων που χρησιμοποιήθηκε τελικώς δημιουργήθηκε από το εργαστήριο Τηλεπισκόπησης της Σχολής Αγρονόμων και Τοπογράφων Μηχανικών. Τα δεδομένα που αποτέλεσαν το σύνολο δεδομένων προέρχονταν από το δορυφόρο Sentinel-2, για το έτος 2018. Όπως αναφέρθηκε και στο κεφάλαιο 3, ο δορυφόρος Sentinel-2 διαθέτει υψηλή χωρική ανάλυση (10 μέτρα), παρέχοντας συστηματική κάλυψη της ελληνικής επικράτειας (και γενικότερα της γης) και πολλαπλή φασματική πληροφορία. Οι εικόνες που μας δόθηκαν είχαν «κοπεί» σε μικρότερο μέγεθος από αυτό που παρέχει ο δορυφόρος, απεικονίζαν όλες την ίδια ακριβώς περιοχή στις διάφορες ημερομηνίες κι εκτός από την δορυφορική εικόνα, μας παρέιχαν μία εικόνα που παρουσίαζε το «ground truth» και μία εικόνα που περιείχε τις μάσκες των σύννεφων. Επομένως, το αρχικό σύνολο δεδομένων, όπως μας δόθηκε, περιελάμβανε αναλυτικά:

- 79 εικόνες (raw) με πλάτος 1802 εικονοστοιχεία, ύψος 2126 εικονοστοιχεία και πληροφορία για 7 φασματικά κανάλια.

- 73 ασπρόμαυρες εικόνες (fmask) με πλάτος 901 εικονοστοιχεία και ύψος 1063 εικονοστοιχεία. Οι συγκεκριμένες εικόνες απεικόνιζαν τη θέση των σύννεφων με μάσκες.
- 79 εικόνες που είχαν επεξεργαστεί με το φίλτρο διάμεσης τιμής (median filter) με ίδιες διαστάσεις όσες και οι raw εικόνες. Οι συγκεκριμένες εικόνες είχαν δημιουργηθεί με βάση τη θέση των σύννεφων στις μάσκες. Στην πραγματικότητα, όπου υπήρχαν σύννεφα χρησιμοποιούταν το φίλτρο διάμεσης τιμής, ώστε από την μέση τιμή των τιμών των χρονικά γειτονικών εικονοστοιχείων να «γεμίσει» η περιοχή της μάσκας. Αυτές οι εικόνες χρησιμοποιήθηκαν ως «ground truth».

Επειδή οι μάσκες ήταν λιγότερες από τις εικόνες, επιλέξαμε μόνο όσες εικόνες (raw) είχαν μάσκα κι επίσης χρειάστηκε να αφαιρέσουμε και δύο εικόνες που το περισσότερο τμήμα τους ήταν μαύρο. Επομένως, το σύνολο δεδομένων αποτελούταν από εικόνες, μάσκες και εικόνες με το φίλτρο μέσης τιμής για 71 ημερομηνίες συνολικά.



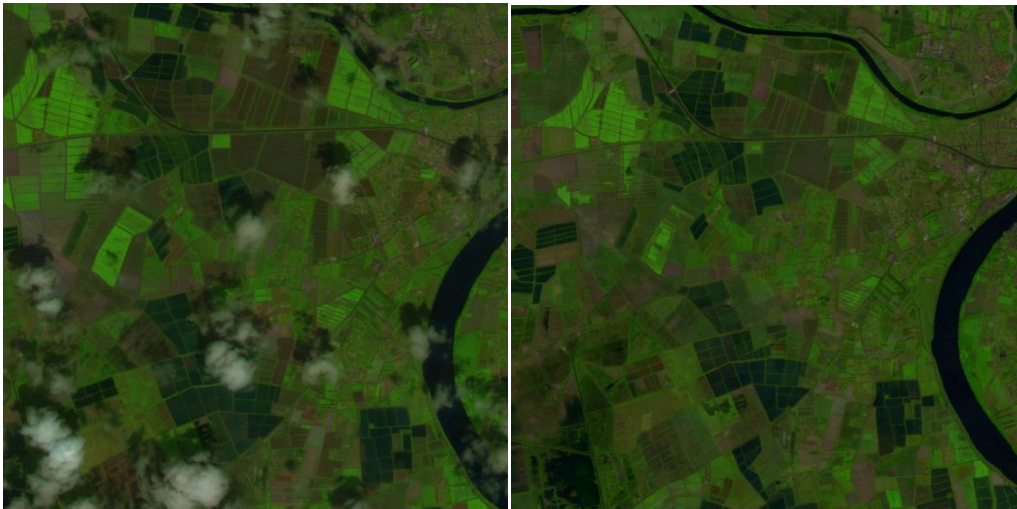
Εικόνα 4-1: Ενδεικτικά παραδείγματα εικόνας RGB με σύννεφα (αριστερά), εικόνας χωρίς σύννεφα με χρήση φίλτρου διάμεσης τιμής (κέντρο), μάσκας σύννεφων (δεξιά) – ημερομηνία: 02-07-2018

4.2 Επεξεργασία αρχικού συνόλου δεδομένων

Πριν την εφαρμογή των μεθοδολογιών που θα παρουσιαστούν στις επόμενες ενότητες του κεφαλαίου έπρεπε να γίνει επεξεργασία του αρχικού συνόλου δεδομένων. Οι συγκεκριμένες επεξεργασίες περιγράφονται παρακάτω.

4.2.1 Επιλογή των Καναλιών του Sentinel-2

Όπως αναφέρθηκε στην προηγούμενη ενότητα οι εικόνες του αρχικού συνόλου δεδομένων καθώς και οι εικόνες που είχαν επεξεργαστεί με το φίλτρο διάμεσης τιμής περιείχαν πληροφορία για 7 κανάλια. Δεδομένου του ότι η εκτέλεση των πειραμάτων έγινε στο Google Colaboratory, το οποίο παρέχει περιορισμένους υπολογιστικούς πόρους (GPU, RAM), κρίθηκε απαραίτητο να μειωθεί η πολυπλοκότητα των μοντέλων σε παραμέτρους επιλέγοντας να δουλέψουμε μόνο με τα τρία από τα επτά κανάλια των εικόνων. Συγκεκριμένα επιλέχθηκαν τα κανάλια 2 (Blue), 3 (Green) και 4 (Red) που προσφέρουν πληροφορία για το ορατό τμήμα του φάσματος. Στην παρακάτω εικόνα παρουσιάζεται στα αριστερά το τμήμα μίας εικόνας από το σύνολο δεδομένων μετά την επιλογή των 3 καναλιών και στα δεξιά το αντίστοιχο τμήμα της εικόνας με το φίλτρο διάμεσης τιμής.



Εικόνα 4-2: Ενδεικτικό τμήμα δορυφορικής εικόνας με 3 κανάλια (RGB). Ανεπεξέργαστο με σύννεφα (αριστερά), επεξεργασμένο με το φίλτρο διάμεσης τιμής (δεξιά).
Ημερομηνία: 09-06-2018.

4.2.2 Κανονικοποίηση τιμών (Normalization)

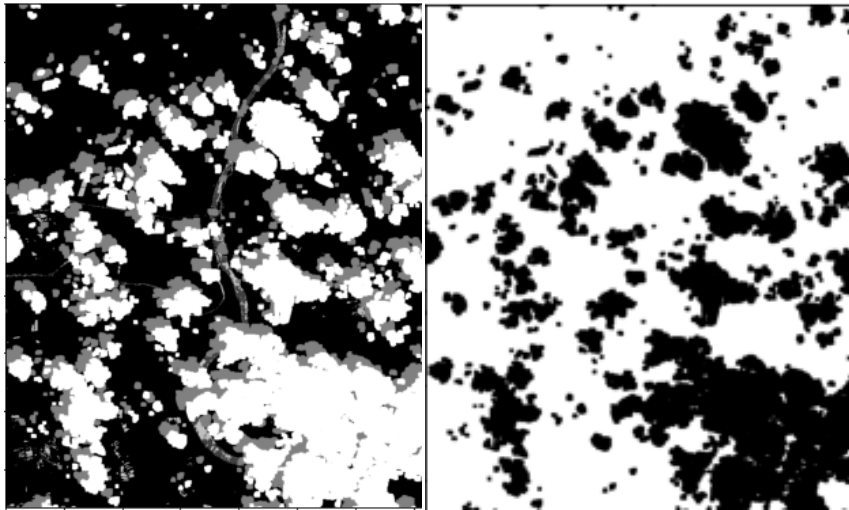
Οι raw εικόνες και οι εικόνες με το φίλτρο διάμεσης τιμής αποτελούνταν από αρνητικές και θετικές ακέραιες τιμές στο διάστημα $[-1161, 18987]$. Τα μοντέλα που εξετάζονται όμως, δέχονται ως είσοδο εικόνες που οι τιμές τους κυμαίνονται στο διάστημα $[0, 255]$. Για το λόγο αυτό, έγινε πρώτα μετασχηματισμός των τιμών με χρήση του αλγορίθμου ελάχιστης-μέγιστης τιμής (4.1). Στο συγκεκριμένο αλγόριθμο για κάθε εικόνα η ελάχιστη τιμή εικονοστοιχείου μετατρέπεται σε 0, η μέγιστη τιμή μετατρέπεται σε 1 και κάθε άλλη τιμή μετατρέπεται σε δεκαδικό μεταξύ του 0 και του 1, ώστε όλα τα εικονοστοιχεία να πάρουν τιμές στο διάστημα

$[0, 1]$ και στη συνέχεια πολλαπλασιάσαμε τις τιμές με το 255, «απλώνοντάς» τις στο διάστημα $[0, 255]$.

$$\text{νέα τιμή} = \frac{\text{τιμή} - \text{ελάχιστο}}{\text{μέγιστο} - \text{ελάχιστο}}, \quad (4.1)$$

4.2.3 Επεξεργασία масκών

Οι μάσκες είχαν το μισό μέγεθος από τις υπόλοιπες εικόνες, γι' αυτό κι έγινε η επιλογή να διπλασιαστεί το μέγεθος τους (resize), ώστε όλες οι εικόνες να έχουν τις ίδιες διαστάσεις $1802 * 2162$ εικονοστοιχεία. Επίσης, οι τιμές των εικονοστοιχείων των масκών ήταν $[0, 1, 2, 3, 4]$ και μετατράπηκε σε δυαδική εικόνα με τιμές 0 και 1, όπου το 0 αντιπροσώπευε τα σύννεφα (απεικονιζόταν με μαύρο) και 1 τα υπόλοιπα στοιχεία της εικόνας (λευκό).



Εικόνα 4-3: Ενδεικτικό παράδειγμα αρχικής μάσκας (αριστερά) και μάσκας μετά την επεξεργασία (δεξιά).

4.2.4 Δημιουργία μικρότερων εικόνων (patches) από τις αρχικές

Όπως αναφέρθηκε και παραπάνω, το σύνολο δεδομένων αποτελείται από 71 εικόνες, ο αριθμός των οποίων είναι αρκετά μικρός για αλγορίθμους βαθιάς μάθησης. Επίσης, το μέγεθος τους είναι αρκετά μεγάλο, γεγονός που θα δυσκόλευε πολύ την εκπαίδευση ενός μοντέλου. Για αυτούς τους λόγους, αποφασίστηκε να κοπούν οι αρχικές εικόνες σε μικρότερα τμήματα (patches) διαστάσεων $256 * 256$ το καθένα. Επομένως, από κάθε εικόνα σύμφωνα με το μέγεθος της δημιουργήθηκαν περίπου 56 μικρότερες εικόνες διάστασης $256 * 256$.

4.2.5 Διαχωρισμός εικόνων σε κατηγορίες

Οι μικρότερες εικόνες που δημιουργήθηκαν χωρίστηκαν σε δύο κατηγορίες ανάλογα με το αν περιέχουν σύννεφα ή όχι. Για το διαχωρισμό τους χρησιμοποιήθηκαν οι μάσκες που δίνονταν με το σύνολο δεδομένων. Επομένως, οι εικόνες που περιείχαν σύννεφα ήταν 2226 ενώ εκείνες χωρίς σύννεφα 1750. Παρόλο τον διαχωρισμό που έγινε σύμφωνα με τις μάσκες έγινε και εποπτικός έλεγχος δεδομένου του ότι οι μάσκες δεν ήταν ακριβείας και διαπιστώθηκε ότι τελικώς οι εικόνες με σύννεφα ήταν λιγότερες από 2000.

4.3 Πρώτο Μοντέλο (Pix2Pix)

Το πρώτο μοντέλο που δοκιμάστηκε για την αφαίρεση των σύννεφων από τις δορυφορικές εικόνες ήταν το Pix2Pix. Το συγκεκριμένο μοντέλο αντιμετωπίζει το πρόβλημα ως πρόβλημα εικόνας και όπως αναφέρθηκε και στο κεφάλαιο 2, έχει σχεδιαστεί για να «μεταφράζει» με συγκεκριμένο τρόπο τις εικόνες που εισάγονται σε αυτό από ένα πεδίο πηγής σε ένα πεδίο στόχου. Επομένως, στην δική μας περίπτωση θα εξετάσουμε αν οι εικόνες με τα σύννεφα (πεδίο πηγής) μπορούν να μεταφραστούν σε εικόνες χωρίς σύννεφα (πεδίο στόχου).

Για την υλοποίηση του πειράματος χρησιμοποιήθηκαν ως είσοδος οι εικόνες με τα σύννεφα (source image) και ως ground truth οι αντίστοιχες εικόνες που είχαν δημιουργηθεί με το φίλτρο μέσης τιμής (target image). Λόγω έλλειψης πόρων (GPU, RAM) δεν ήταν δυνατό να χρησιμοποιηθούν όλες οι εικόνες που δημιουργήθηκαν με τον τρόπο που περιγράφηκε προηγουμένως. Επομένως, έγινε επιλογή 240 ζευγών εικόνων με σύννεφα και χωρίς σύννεφα (φίλτρο μέσης τιμής). Οι εικόνες αυτές αποτέλεσαν τα δεδομένα εκπαίδευσης. Επιπλέον, επιλέχθηκαν 50 εικόνες για να λειτουργήσουν ως δεδομένα ελέγχου. Η επιλογή των εικόνων έγινε με τέτοιο τρόπο ώστε να συμπεριλαμβάνονται εικόνες από όλα τα ποσοστά συννεφοκάλυψης. Οι διαστάσεις των δεδομένων εισόδου ήταν $256*256*3$. Παρακάτω παρουσιάζεται ένα ζεύγος εισόδου.



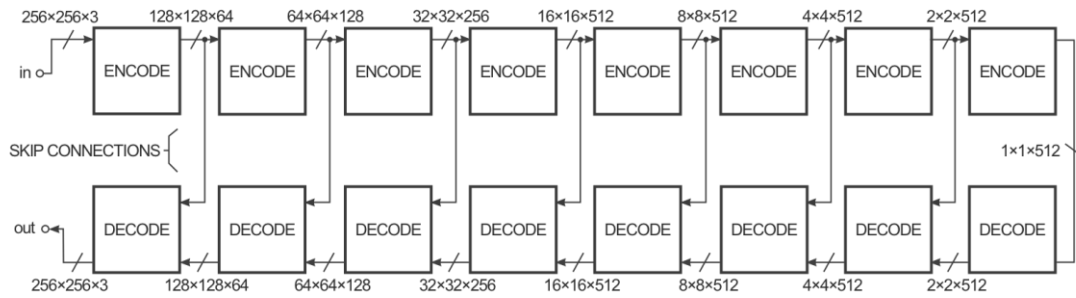
Εικόνα 4-4: Ενδεικτικό παράδειγμα ζεύγους εισόδου, ground truth (αριστερά), εικόνα με σύννεφα (δεξιά)

4.3.1 Αρχιτεκτονική

Το pix2pix αποτελείται από δύο επιμέρους μοντέλα, τον δημιουργό (generator) και τον διευκρινιστή (discriminator). Στο μοντέλο του δημιουργού παρέχεται μια δεδομένη εικόνα ως είσοδος και εκείνος δημιουργεί μια μεταφρασμένη έκδοση της εικόνας. Το μοντέλο του διευκρινιστή συγκρίνει την εικόνα εισόδου (εικόνα με σύννεφα) με μία άγνωστη εικόνα (είτε μια εικόνα στόχου από το σύνολο δεδομένων είτε μια εικόνα εξόδου από τον δημιουργό) και προσπαθεί να μαντέψει εάν αυτή κατασκευάστηκε από το δημιουργό. Στόχος του δημιουργού είναι να εκπαιδευτεί ώστε να ξεγελάσει το μοντέλο του διευκρινιστή και να ελαχιστοποιήσει την απώλεια μεταξύ της δημιουργούμενης εικόνας και της αναμενόμενης εικόνας στόχου.

Ο δημιουργός είναι ένα μοντέλο κωδικοποιητή-αποκωδικοποιητή που χρησιμοποιεί αρχιτεκτονική U-Net. Το δίκτυο του κωδικοποιητή αποτελείται από 8 συνελκτικά μπλοκ. Κάθε συνελκτικό μπλοκ αποτελείται από ένα επίπεδο συνέλιξης (convolutional layer) μεγέθους $4 * 4$, ακολουθούμενο από τη συνάρτηση ενεργοποίησης Leaky Relu. Σε κάθε επίπεδο εκτός από το πρώτο επίπεδο συνέλιξης, εφαρμόζεται ομαλοποίηση του batch (batch normalization) δεδομένων. Το δίκτυο του αποκωδικοποιητή έχει κι αυτό 8 από-συνελκτικά μπλοκ. Καθένα μπλοκ αποτελείται από ένα από-συνελκτικό επίπεδο (deconvolutional layer) μεγέθους $4*4$, ακολουθούμενο από ένα επίπεδο ομαλοποίησης του batch και τη συνάρτηση ενεργοποίησης Relu. Στα πρώτα τρία μπλοκ αποκωδικοποίησης χρησιμοποιήθηκε η τεχνική ομαλοποίησης dropout με τιμή ίση με 0.5. Στο επίπεδο εξόδου του δημιουργού χρησιμοποιήθηκε η συνάρτηση ενεργοποίησης υπερβολικής εφαπτομένης (tanh), που σημαίνει ότι οι τιμές των εικονοστοιχείων στην εικόνα που δημιουργείται θα είναι στην περιοχή $[-1, 1]$.

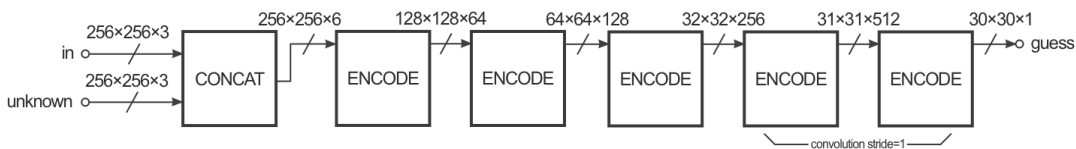
Στο παρακάτω σχήμα φαίνεται το πως κωδικοποιείται και αποκωδικοποιείται η εικόνα από κάθε μπλοκ του δημιουργού. Μπορούμε να παρατηρήσουμε και τις συνενώσεις μεταξύ των εξόδων των μπλοκ των κωδικοποιητών και των αντίστοιχων μπλοκ των αποκωδικοποιητών, η οποία γίνεται στον άξονα των καναλιών.



Σχήμα 4-1: Αρχιτεκτονική Δημιουργού. Πηγή: <https://towardsdatascience.com/cyclegans-and-pix2pix-5e6a5f0159c4>

Ο διακρινιστής είναι ένα βαθύ συνελκτικό νευρωνικό δίκτυο που εκτελεί ταξινόμηση εικόνας. Δέχεται δύο εικόνες, είτε την εικόνα εισόδου και την εικόνα στόχου, που πρέπει να ταξινομήσει ως αληθινή, είτε την εικόνα εισόδου και την παραγόμενη από τον δημιουργό εικόνα που πρέπει να ταξινομήσει ως «ψεύτικη». Αυτές οι δύο εικόνες συνενώνονται για την είσοδο στο μοντέλο, στον κώδικα υλοποίησης.

Η αρχιτεκτονική του βασίζεται στο μοντέλο PatchGAN. Αποτελείται από 5 συνελκτικά μπλοκ κωδικοποιητών που συμπιέζουν την συνενωμένη εικόνα. Η έξοδος είναι ένα πλέγμα εικονοστοιχείων 30 * 30 (patch), όπου προσπαθεί να ταξινομήσει ένα 70 * 70 τμήμα της εικόνας εισόδου. Το 70 επιλέχθηκε να είναι αρκετά μικρότερο από την αρχική διάσταση της εικόνας εισόδου (256) ώστε να έχει λιγότερες παραμέτρους το μοντέλο. Στο παρακάτω σχήμα παρουσιάζεται το πως κωδικοποιείται η συνενωμένη εικόνα από τα επίπεδα του διακρινιστή.



Σχήμα 4-2: Αρχιτεκτονική Διακρινιστή. Πηγή: <https://towardsdatascience.com/cyclegans-and-pix2pix-5e6a5f0159c4>

Στους πίνακες που ακολουθούν παρουσιάζονται αναλυτικά οι αρχιτεκτονικές του δημιουργού και του διευκρινιστή, οι έξοδοι που προκύπτουν από κάθε επίπεδο και οι επιλογές των υπερπαραμέτρων.

Αρχιτεκτονική Δημιουργού							
Layers	Kernel	Stride	Channels	W*H	Activation	BN ¹	Dropout
Είσοδος: Εικόνα			3	256*256	LeakyReLU	✗	✗
Conv.Layer 1	4	2	64	128*128	LeakyReLU	✓	✗
Conv.Layer 2	4	2	128	64*64	LeakyReLU	✓	✗
Conv.Layer 3	4	2	256	32*32	LeakyReLU	✓	✗
Conv.Layer 4	4	2	512	16*16	LeakyReLU	✓	✗
Conv.Layer 5	4	2	512	8*8	LeakyReLU	✓	✗
Conv.Layer 6	4	2	512	4*4	LeakyReLU	✓	✗
Conv.Layer 7	4	2	512	2*2	LeakyReLU	✓	✗
Conv.Layer 8	4	2	512	1*1	LeakyReLU	✗	✗
Deconv.Layer 9	4	2	1024	2*2	ReLU	✓	✓
Συνένωση εξόδων (Layer 9, Layer 6)							
Deconv.Layer 10	4	2	1024	4*4	ReLU	✓	✓
Συνένωση εξόδων (Layer 10, Layer 5)							
Deconv.Layer 11	4	2	1024	8*8	ReLU	✓	✓
Συνένωση εξόδων (Layer 11, Layer 4)							
Deconv.Layer 12	4	2	1024	16*16	ReLU	✓	✗
Συνένωση εξόδων (Layer 12, Layer 3)							
Deconv.Layer 13	4	2	512	32*32	ReLU	✓	✗
Συνένωση εξόδων (Layer 13, Layer 2)							
Deconv.Layer 14	4	2	256	64*64	ReLU	✓	✗
Συνένωση εξόδων (Layer 14, Layer 1)							
Deconv.Layer 15	4	2	128	128*128	Tanh	✗	✗
Έξοδος: Μετασηματισμένη Εικόνα			3	256*256	✗	✗	✗

Πίνακας 4-1: Πίνακας που παρουσιάζει την αρχιτεκτονική για τον δημιουργό του μοντέλου pix2pix. (¹:batch normalization)

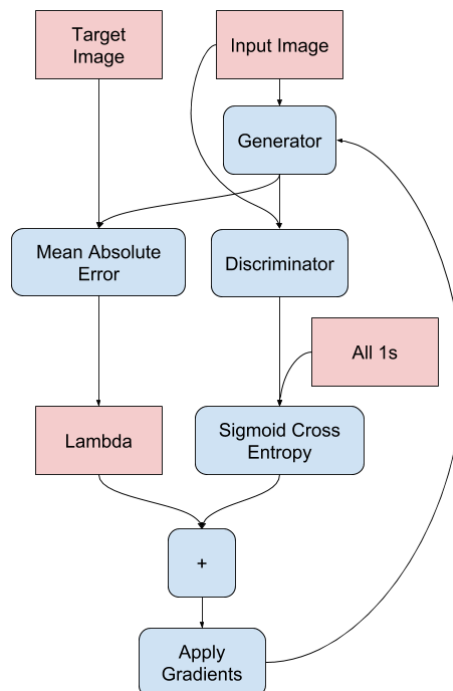
Αρχιτεκτονική Διευκρινιστή						
Layers	Kernel	Stride	Channels	W*H	Activation	BN
Είσοδος: Συνένωση(Εικόνα 1, Εικόνα 2)			6	256*256	✗	✗
Conv.Layer 1	4	2	64	128*128	LeakyReLU	✗
Conv.Layer 2	4	2	128	64*64	LeakyReLU	✓
Conv.Layer 3	4	2	256	32*32	LeakyReLU	✓
Zero Padding			256	34*34	✗	✗
Conv.Layer 4	4	1	512	31*31	LeakyReLU	✓
Zero Padding			512	33*33	✗	✗
Έξοδος: guess (real or fake)			512	30*30	LeakyReLU	✓

Πίνακας 4-2: Πίνακας που παρουσιάζει την αρχιτεκτονική για τον διευκρινιστή του μοντέλου pix2pix.

4.3.2 Εκπαίδευση

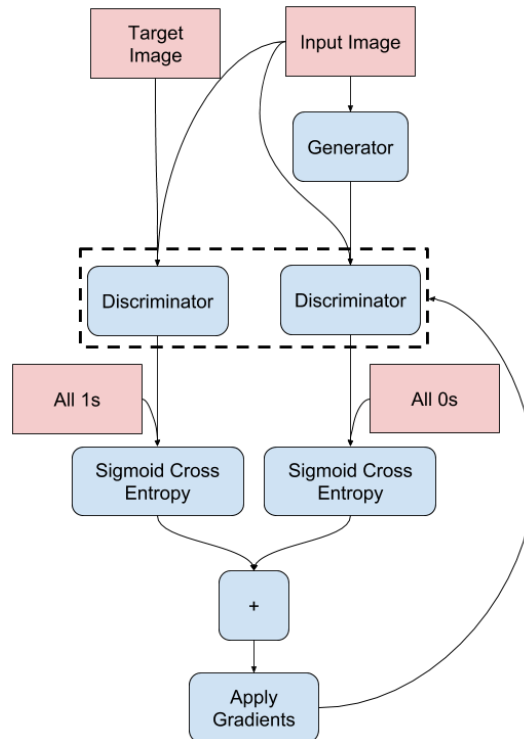
Στο pix2pix μοντέλο η εκπαίδευση του δημιουργού γίνεται μέσω του διευκρινιστή. Πρώτα ο δημιουργός κατασκευάζει μια εικόνα εξόδου και στη συνέχεια ο διευκρινιστής κοιτάζει το ζεύγος εισόδου-στόχου και το ζεύγος εισόδου-εξόδου και δίνει την πρόβλεψη του για το πόσο ρεαλιστικά φαίνονται. Τα βάρη του διευκρινιστή προσαρμόζονται στη συνέχεια με βάση το σφάλμα ταξινόμησης του ζεύγους εισόδου-εξόδου και του ζεύγους εισόδου-στόχου. Τα βάρη του δημιουργού ρυθμίζονται με βάση την έξοδο του διευκρινιστή, καθώς και τη διαφορά μεταξύ της εικόνας εξόδου και του στόχου. Καθώς ο διευκρινιστής γίνεται καλύτερος, το ίδιο συμβαίνει και στον δημιουργό.

Ο δημιουργός ενημερώνεται μέσω του σταθμισμένου αθροίσματος τόσο της σιγμοειδούς εγκάρσιας απώλειας εντροπίας (sigmoid cross entropy) των παραγόμενων εικόνων όσο και της απώλειας μέσου απόλυτου σφάλματος (mean absolute error - MAE) ή L1 μεταξύ της δημιουργούμενης εικόνας και της εικόνας στόχου. Η στάθμιση που προτείνεται είναι 100 έως 1 υπέρ της απώλειας L1. Αυτό συμβαίνει για να ενθαρρυνθεί ο δημιουργός να κατασκευάσει ρεαλιστικές μεταφράσεις της εικόνας εισόδου. Η σιγμοειδής εγκάρσια απώλεια εντροπίας επηρεάζει εάν το μοντέλο του δημιουργού μπορεί να εξάγει εικόνες που αντιστοιχούν στο πεδίο στόχου, ενώ η απώλεια L1 κανονικοποιεί το μοντέλο του δημιουργού για να εξάγει εικόνες που είναι εύλογη μετάφραση της εικόνας προέλευσης. Στα παρακάτω σχήματα φαίνεται η διαδικασία εκπαίδευσης του δημιουργού και του διευκρινιστή.



Σχήμα 4-3: Εκπαίδευση του Δημιουργού.

Πηγή: <https://www.tensorflow.org/tutorials/generative/pix2pix>



Σχήμα 4-4: Εκπαίδευση του Διευκρινιστή (δεξιά).

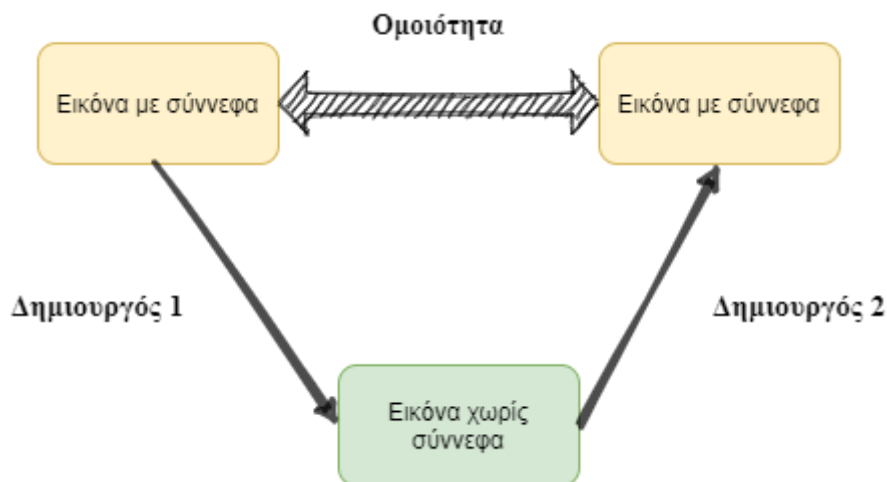
Πηγή: <https://www.tensorflow.org/tutorials/generative/pix2pix>

Όλα τα δίκτυα εκπαιδεύτηκαν από το μηδέν. Τα βάρη ξεκίνησαν από μια κανονική κατανομή με μέση τιμή 0 και τυπική απόκλιση 0.02. Ο βελτιστοποιητής που χρησιμοποιήθηκε για την εκπαίδευση ήταν ο Adam με ρυθμό μάθησης 0.0002 και το batch size επιλέχθηκε ως 1. Η εκπαίδευση του μοντέλου έγινε για 300 εποχές και η τιμή του dropout τέθηκε ίση με 0.5. Ο χρόνος εκπαίδευσης στο Google Colaboratory για 300 εποχές ήταν 3 ώρες. Τα αποτελέσματα παρουσιάζονται στο κεφάλαιο 5.

4.4 Δεύτερο Μοντέλο (CycleGAN)

Το δεύτερο μοντέλο που δοκιμάστηκε για την αφαίρεση των σύννεφων από τις δορυφορικές εικόνες είναι το CycleGAN. Με το συγκεκριμένο μοντέλο έγινε προσπάθεια να αντιμετωπιστεί το πρόβλημα που εξετάζει η παρούσα διπλωματική πάλι ως πρόβλημα μετάφρασης εικόνας. Όπως αναφέρθηκε και στο κεφάλαιο 2, το CycleGAN αποτελείται από δύο γεννητικά ανταγωνιστικά δίκτυα, έχοντας συνολικά δύο δημιουργούς (generators) και δύο διευκρινιστές (discriminators). Λαμβάνοντας υπόψη δύο σετ διαφορετικών εικόνων, στην περίπτωσή μας, εικόνες με σύννεφα και εικόνες χωρίς σύννεφα, ο ένας δημιουργός «μεταφράζει» τις εικόνες με σύννεφα σε εικόνες χωρίς σύννεφα και ο άλλος δημιουργός λειτουργεί αντίστροφα «μεταφράζοντας» τις εικόνες χωρίς σύννεφα σε εικόνες με σύννεφα. Κατά τη διάρκεια της εκπαίδευσης, οι διευκρινιστές ελέγχουν εάν οι εικόνες που

κατασκευάζονται από τους δημιουργούς φαίνεται να είναι αληθινές ή όχι. Μέσω αυτής της διαδικασίας, οι δημιουργοί μπορούν να γίνουν καλύτεροι με την ανατροφοδότηση από τους αντίστοιχους διευκρινιστές τους. Επιπλέον, ο κάθε δημιουργός λαμβάνει και μία επιπλέον ανατροφοδότηση από τον άλλο δημιουργό. Αυτή η ανατροφοδότηση διασφαλίζει ότι η εικόνα που κατασκευάζεται από τον δημιουργό είναι «συνεπής στον κύκλο» (cycle consistent), που σημαίνει ότι εφαρμόζοντας διαδοχικά και τους δύο δημιουργούς στην εικόνα το αποτέλεσμα πρέπει να είναι μία παρόμοια εικόνα.



Σχήμα 4-5: Διαδικασία Cycle-consistent.

Ένα πλεονέκτημα του CycleGAN είναι ότι δεν χρειάζεται τα δεδομένα να είναι ζευγάρια, όπως στο μοντέλο pix2pix που εξετάστηκε προηγουμένως. Για την υλοποίηση του πειράματος χρησιμοποιήθηκαν ως είσοδος 240 εικόνες με σύννεφα (οι ίδιες που είχαν χρησιμοποιηθεί και στο pix2pix) και 270 τυχαία επιλεγμένες εικόνες χωρίς σύννεφα. Οι διαστάσεις των δεδομένων ήταν $256 \times 256 \times 3$. Επιπλέον, ως δεδομένα ελέγχου χρησιμοποιήθηκαν οι ίδιες 50 εικόνες με σύννεφα που χρησιμοποιήθηκαν και στο πείραμα του pix2pix μοντέλου, ώστε να συγκρίνουμε τα τελικά αποτελέσματα σε σχέση με τα δύο μοντέλα.

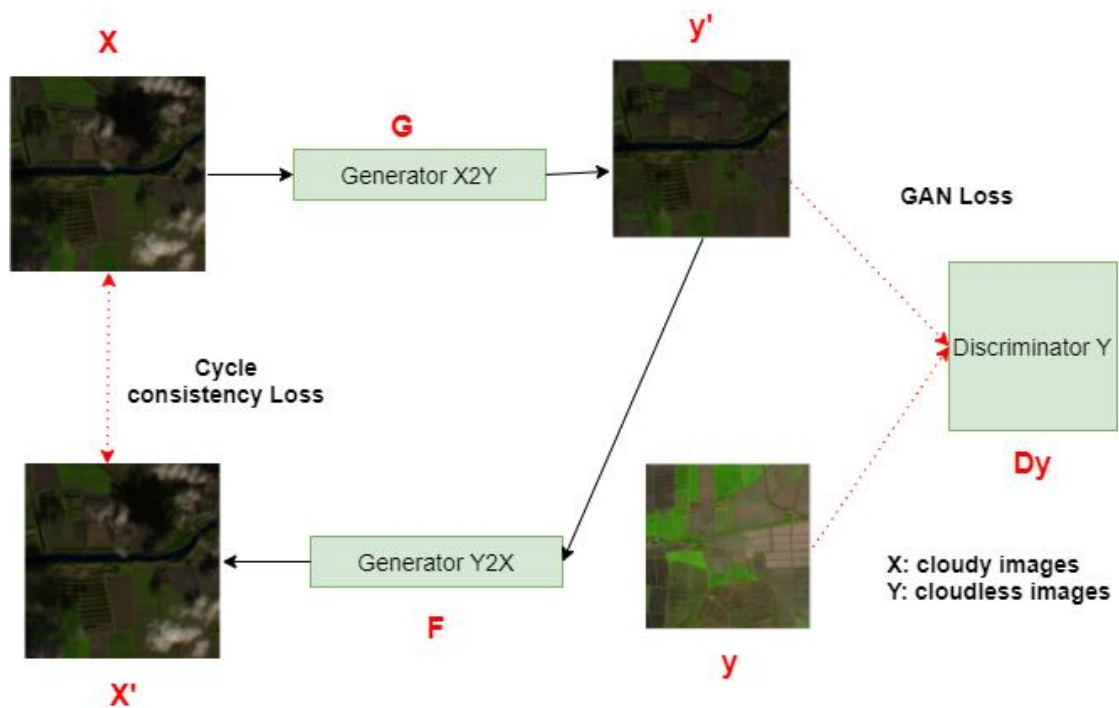
4.4.1 Αρχιτεκτονική

Η αρχιτεκτονική που περιγράφεται στην αντίστοιχη δημοσίευση θέλει τους δημιουργούς να ακολουθούν αρχιτεκτονική Resnet. Στην δική μας υλοποίηση όμως ακολουθήσαμε την αρχιτεκτονική του pix2pix η οποία αναφέρθηκε αναλυτικά στην ενότητα **4.3.1** και για τον δημιουργό χρησιμοποιεί U-Net αρχιτεκτονική. Η μόνη διαφορά είναι ότι στην αρχιτεκτονική

των δημιουργών χρησιμοποιήθηκε η instance normalization αντί της batch. Οι διευκρινιστές ακολουθούν αρχιτεκτονική PatchGAN (όπως στο 4.3.1).

4.4.2 Εκπαίδευση

Αν θεωρήσουμε ότι το μοντέλο αποτελείται από δύο ζεύγη δημιουργού-διευκρινιστή G-D1 και F-D2 η εκπαίδευση ακολουθεί την παρακάτω διαδικασία. Ο G δέχεται ως είσοδο την εικόνα X με σύννεφα και τη «μεταφράζει» σε μία εικόνα χωρίς σύννεφα Y . Το καθήκον του D1 είναι να εξετάσει εάν η εικόνα από τον G είναι πραγματική εικόνα χωρίς σύννεφα ή εικόνα που δημιουργήθηκε από το δίκτυο δημιουργών. Η παραγόμενη εικόνα από τον G μεταφέρεται στον δημιουργό F. Ο στόχος του F είναι να «μεταφράσει» την παραγόμενη εικόνα χωρίς σύννεφα σε μια εικόνα με σύννεφα X' . Το καθήκον του D2 τότε είναι να διακρίνει αν η εικόνα είναι πραγματική ή παραγόμενη.



Σχήμα 4-6: Εκπαίδευση CycleGAN.

Τα μοντέλα δημιουργού είναι κανονικοποιημένα ώστε να μην δημιουργούν μόνο νέες εικόνες στο πεδίο στόχου, αλλά να «μεταφράζουν» περισσότερες εκδόσεις των εικόνων εισαγωγής από το πεδίο πηγής. Αυτό επιτυγχάνεται χρησιμοποιώντας τις παραγόμενες εικόνες ως είσοδο στο αντίστοιχο μοντέλο δημιουργού και συγκρίνοντας την εικόνα εξόδου με τις αρχικές εικόνες. Το πέρασμα μιας εικόνας και από τους δύο δημιουργούς ονομάζεται κύκλος. Κάθε

ζεύγος μοντέλων δημιουργού εκπαιδεύεται για την καλύτερη αναπαραγωγή της αρχικής εικόνας πηγής (συνέπεια κύκλου).

Για την εκπαίδευση του μοντέλου χρησιμοποιούνται οι απώλειες του δημιουργού και του διευκρινιστή, οι οποίες είναι το σταθμισμένο άθροισμα της σιγμοειδούς εγκάρσιας απώλειας εντροπίας (sigmoid cross entropy) των παραγόμενων εικόνων και επίσης εισάγεται η απώλεια cycle consistency (συνέπεια στον κύκλο) και η απώλεια ταυτότητας (identity loss). Λαμβάνοντας τις προβλέψεις από τους διευκρινιστές, υπολογίζονται οι απώλειες και μέσω της οπίσθιας διάδοσης σφάλματος υπολογίζονται οι κλίσεις και εφαρμόζονται στους βελτιστοποιητές.

Και σε αυτή την περίπτωση τα δίκτυα εκπαιδεύτηκαν από το μηδέν. Τα βάρη ξεκίνησαν από μια κανονική κατανομή με μέση τιμή 0 και τυπική απόκλιση 0.02. Ο βελτιστοποιητής που χρησιμοποιήθηκε για την εκπαίδευση ήταν ο Adam με ρυθμό μάθησης 0.0002 και το batch size επιλέχθηκε ως 1. Σκοπός ήταν η εκπαίδευση του μοντέλου να γίνει για 300 εποχές, όμως λόγω των απαιτήσεων σε μνήμη RAM, κάτι τέτοιο αποδείχθηκε αδύνατον στο Google Colaboratory, όπου έγιναν και τα πειράματα. Τελικώς, το Google Colaboratory μας άφησε να εκπαιδεύσουμε το μοντέλο μόνο για 45 συνεχόμενες εποχές. Επομένως η διαδικασία εκπαίδευσης επαναλήφθηκε αρκετές φορές μέχρι να φτάσουμε τις 200 εποχές. Ο χρόνος εκτέλεσης της εκπαίδευσης για τις 45 εποχές στο Google Colaboratory ήταν 2.5 ώρες. Τα αποτελέσματα παρουσιάζονται στο κεφάλαιο 5.

4.5 Τρίτο Μοντέλο (LGTSM)

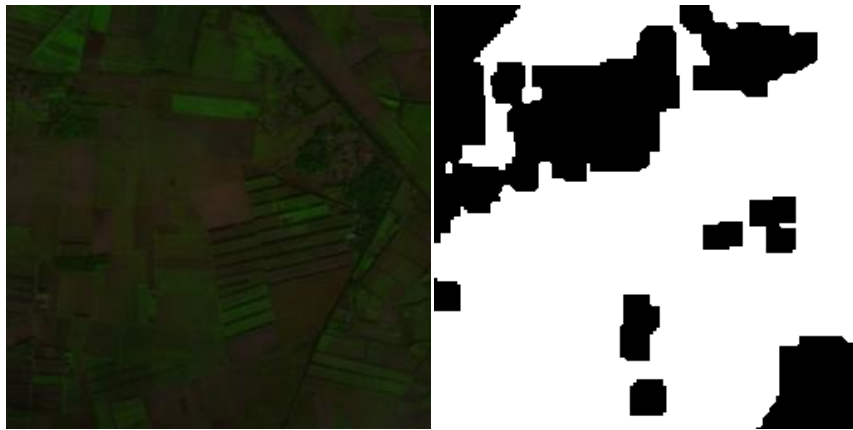
Το τρίτο μοντέλο που δοκιμάστηκε για την αφαίρεση των σύννεφων από τις δορυφορικές εικόνες είναι το Learnable Gated Temporal Shift Module for Deep Video Inpainting (LGTSM). Με το συγκεκριμένο μοντέλο έγινε προσπάθεια να αντιμετωπιστεί το πρόβλημα της αφαίρεσης σύννεφων ως χρονοσειρά. Όπως αναφέρθηκε και στο κεφάλαιο 2, το LGTSM είναι ένα γεννητικό ανταγωνιστικό δίκτυο, που δέχεται ως είσοδο ένα παραμορφωμένο βίντεο (με μάσκες) και δημιουργεί την πλήρη εικόνα του, χάρη στην ενσωμάτωση μίας μονάδας περιφραγμένης χρονικής μετατόπισης (gated temporal shift module) στον δημιουργό και στον διευκρινιστή του δικτύου. Στην δική μας περίπτωση, η συγκεκριμένη μεθοδολογία θα μπορούσε να υιοθετηθεί υπό τη σκέψη ότι όλες μαζί οι δορυφορικές εικόνες που έχουμε από το ίδιο μέρος για διάφορες χρονικές στιγμές αποτελούν μία χρονοσειρά. Αν υποθέσουμε ότι κάθε εικόνα αποτελεί ένα καρέ σε μία συγκεκριμένη χρονική σειρά, μπορούμε να

παραδεχτούμε ότι η χρονοσειρά στην πραγματικότητα είναι ένα βίντεο. Επομένως, με αυτή τη λογική θα εξετάσουμε τα αποτελέσματα του συγκεκριμένου μοντέλου στο πρόβλημα μας.

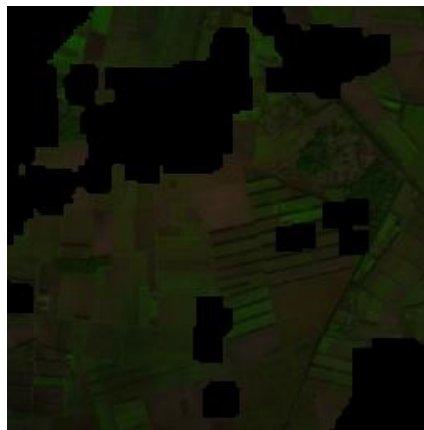
Για την υλοποίηση του πειράματος χρησιμοποιήθηκαν οι εικόνες RGB που δεν περιείχαν σύννεφα, οι οποίες είχαν δημιουργηθεί στην υπό-ενότητα 4.2.5. Αυτές οι εικόνες ήταν μεγέθους 256×256 και απεικόνιζαν κάποιο τμήμα της αρχικής εικόνας σε κάποια χρονική στιγμή. Επομένως, υπήρχαν πολλές τέτοιες εικόνες που τύχαινε να απεικονίζουν το ίδιο τμήμα για διαφορετικές στιγμές. Γι' αυτό το λόγο χωρίστηκαν με βάση το τμήμα που απεικόνιζαν, έτσι ώστε να δημιουργηθούν διαφορετικές χρονοσειρές. Συνολικά, δημιουργήθηκαν 56 χρονοσειρές, όπου κάθε μία αποτελούνταν από 12 έως και 40 εικόνες. Για λόγους περιορισμένων υπολογιστικών πόρων, όπως αναφέρθηκε και στις προηγούμενες υλοποιήσεις αποφασίστηκε να επιλεγθούν 12 εικόνες από κάθε χρονοσειρά, ώστε να μπορέσουν να χρησιμοποιηθούν όλες οι χρονοσειρές για την εκπαίδευση του μοντέλου. Από τις 56 χρονοσειρές, οι 50 χρησιμοποιήθηκαν για την εκπαίδευση και οι 6 ως δεδομένα ελέγχου.

Εκτός από τις εικόνες χωρίς σύννεφα επιλέχθηκαν για κάθε χρονοσειρά με τυχαίο τρόπο μάσκες που θα αντιστοιχούν σε κάθε καρέ της. Για την επιλογή των масκών τέθηκε ο περιορισμός, το ποσοστό συννεφοκάλυψης να είναι μεταξύ 10% έως 90%. Το ποσοστό αυτό επιλέχθηκε επειδή ύστερα από οπτικό έλεγχο διαπιστώθηκε ότι εικόνες με ποσοστό σύννεφων μικρότερο του 10% αναφέρονται κυρίως σε χαρακτηριστικά στα όρια της εικόνας που δεν είναι σύννεφα κι επίσης ποσοστό μεγαλύτερο του 90% σημαίνει ότι όλη η εικόνα καλύπτεται από σύννεφα, το οποίο θεωρείται ακραία περίπτωση. Επίσης η κάθε μάσκα που χρησιμοποιήθηκε είναι μοναδική, δηλαδή δεν υπάρχει η ίδια σε κάποια άλλη χρονοσειρά.

Το μοντέλο «διαβάζει» την χρονοσειρά των εικόνων χωρίς σύννεφα και την αντίστοιχη χρονοσειρά με τις μάσκες. Κανονικοποιεί τις τιμές των εικονοστοιχείων της RGB εικόνας στο διάστημα $[0, 1]$, συνενώνει κάθε εικόνα με την αντίστοιχη μάσκα της και δημιουργεί μία επιπλέον διάσταση σε κάθε εικόνα που αντιπροσωπεύει τη χρονική σειρά. Στο τελευταίο στάδιο προ-επεξεργασίας οι εικόνες κάθε χρονοσειράς συνενώνονται στη διάσταση L του χρόνου με αποτέλεσμα η τελική είσοδος του μοντέλου να είναι της μορφής (Batch, Channels, Length, Height, Width) \rightarrow (1, 3, 12, 256, 256), το 12 προκύπτει από το γεγονός ότι για κάθε χρονοσειρά έχουμε 12 χρονικές στιγμές.



Εικόνα 4-5: Ενδεικτικό παράδειγμα εικόνας RGB χωρίς σύννεφα (αριστερά) και τυχαίας μάσκας (δεξιά).



Εικόνα 4-6: Ενδεικτικό παράδειγμα αποτελέσματος συνένωσης των εικόνων 4.5.

4.5.1 Αρχιτεκτονική

Το LGTSM αποτελείται από έναν δημιουργό του οποίου η αρχιτεκτονική βασίζεται σε U-Net δίκτυο σε συνδυασμό με τη μονάδα περιφραγμένης χρονικής μετατόπισης (gated temporal shift module) και από έναν χρονικά μετατοπισμένο διευκρινιστή αρχιτεκτονικής Patch-GAN (TSMGAN).

Το μοντέλο του δημιουργού αποτελείται από 11 συνελκτικά επίπεδα σε συνδυασμό με τη μονάδα περιφραγμένης χρονικής μετατόπισης και συμπεριλαμβάνει επίσης επίπεδα συμπίεσης (downsampling), διαστολής (dilated) και αποσυμπίεσης (upsampling). Σε αντίθεση με την αρχιτεκτονική U-Net που χρησιμοποιεί συνδέσεις παράλειψης, στο συγκεκριμένο μοντέλο δεν γίνεται αυτό, λόγω του ότι πολλές περιοχές είναι καλυμμένες με μάσκες στα επίπεδα συμπίεσης. Για τα επίπεδα συμπίεσης και αποσυμπίεσης εφαρμόζεται διγραμμική παρεμβολή πριν τις συνελίξεις. Για το πρώτο συνελκτικό επίπεδο χρησιμοποιήθηκε φίλτρο μεγέθους 5×5 . Για τα επίπεδα αποσυμπίεσης χρησιμοποιήθηκαν φίλτρα μεγέθους (4×4) με

stride ίσο με 2. Για τα διεσταλμένα επίπεδα χρησιμοποιήθηκε φίλτρο μεγέθους (3*3) με διαστολή 2, 4, 8 και 16, ενώ για οποιοδήποτε άλλο συνελκτικό επίπεδο στον δημιουργό χρησιμοποιήθηκε φίλτρο μεγέθους 3*3. Για μη γραμμικά επίπεδα χρησιμοποιήθηκε η συνάρτηση ενεργοποίησης LeakyReLU, ενώ σε όλα τα συνελκτικά επίπεδα εκτός από το πρώτο προστέθηκε το Learnable Gated Temporal Shift Module.

Ο διευκρινιστής του μοντέλου αποτελείται από 6 συνελκτικά επίπεδα. Όλα τα επίπεδα χρησιμοποιούν το ίδιο φίλτρο μεγέθους 5*5 και stride ίσο με 2. Για τα μη γραμμικά επίπεδα χρησιμοποιήθηκε η συνάρτηση ενεργοποίησης LeakyReLU, ενώ σε όλα τα συνελκτικά επίπεδα εκτός από το πρώτο προστέθηκε το Learnable Gated Temporal Shift Module.

4.5.2 Εκπαίδευση

Όσον αφορά την εκπαίδευση του μοντέλου παρατηρούμε ότι είναι διαφορετική από αυτή που ακολουθείται στα υπόλοιπα γεννητικά ανταγωνιστικά δίκτυα. Οι συγγραφείς προτείνουν πρώτα την εκπαίδευση του δημιουργού μέχρι να επέλθει σύγκλιση και στην συνέχεια την εισαγωγή του διευκρινιστή στο μοντέλο, ώστε να γίνει fine-tuning. Η είσοδος του μοντέλου είναι η χρονοσειρά των εικόνων χωρίς σύννεφα (λειτουργεί και ως ground truth) και η χρονοσειρά των αντίστοιχων τυχαίων μασκών. Η είσοδος «περνάει» από τον δημιουργό, όπου δημιουργεί τη χρονοσειρά χωρίς τις μάσκες. Η έξοδος του δημιουργού συγκρίνεται με το ground truth και υπολογίζονται οι απώλειες. Στη συνέχεια κι εφόσον έχει επέλθει σύγκλιση προστίθεται ο διευκρινιστής και η έξοδος του δημιουργού μαζί με το ground truth «περνάνε» σε αυτόν, όπου και αποφασίζει αν πρόκειται για αληθινό ή ψεύτικο χαρακτηριστικό και υπολογίζει την απώλεια ανταγωνισμού. Μέσω αυτής της διαδικασίας ενημερώνεται ο δημιουργός ώστε να παράγει χρονοσειρές πιο κοντά στις πραγματικές.

Ενδιαφέρον παρουσιάζουν οι απώλειες που χρησιμοποιούνται. Χρησιμοποιείται η masked loss l_1 για τα χαμηλού-επιπέδου χαρακτηριστικά της εικόνας, η perceptual loss και η style loss για το περιεχόμενο της εικόνας. Επίσης, οι συγγραφείς προτείνουν την απώλεια TSMGAN για να διαχειριστούν τα υψηλού-επιπέδου χαρακτηριστικά της εικόνας και για να ενισχύσουν την αληθοφάνεια της. Στη φάση της προ-εκπαίδευσης όπως προτείνεται από τους συγγραφείς η συγκεκριμένη απώλεια δεν χρησιμοποιείται και εισάγεται στο στάδιο του fine-tuning.

Το συγκεκριμένο μοντέλο αρχικά έγινε προσπάθεια να το εκπαιδεύσουμε από το μηδέν, όπως και τα προηγούμενα μοντέλα. Παρ' όλες τις προσπάθειες όμως και τις αλλαγές που έγιναν, το

γεγονός ότι είχαμε λίγα δεδομένα για την εκπαίδευση δεν μπόρεσε να αφήσει το μοντέλο να φτάσει σε σύγκλιση κατά την προ-εκπαίδευση, που παίρνει μέρος μόνο ο δημιουργός. Για το λόγο αυτό, αποφασίστηκε να χρησιμοποιηθεί το ήδη προ-εκπαιδευμένο μοντέλο που παρείχαν οι συγγραφείς. Το προ-εκπαιδευμένο μοντέλο είχε εκπαιδευτεί για 135 εποχές στο σύνολο δεδομένων Free-form Video Inpainting (FVI) που αποτελείται από 15000 υψηλής ευκρίνειας βίντεο με ανθρώπινες δραστηριότητες. Συνεχίζοντας την εκπαίδευση διαπιστώθηκε ότι δεν χρειάζονταν παραπάνω από 15-20 επιπλέον εποχές για να φτάσουμε σε ένα αξιόλογο αποτέλεσμα. Ως προς το στάδιο του fine-tuning έγιναν κάποιες δοκιμές στα βάρη των απωλειών ώστε να παρατηρήσουμε πώς επηρεάζει κάθε απώλεια το τελικό αποτέλεσμα. Ο χρόνος εκτέλεσης κάθε πειράματος στο Google Colaboratory ήταν περίπου 40 λεπτά για τις 15 εποχές. Τα πειράματα που έγιναν, των οποίων τα αποτελέσματα θα αναφερθούν στο κεφάλαιο 5 ήταν τα εξής:

1. Διατήρηση των ήδη επιλεγμένων τιμών βαρών από τους συγγραφείς και πρόσθεση του TSMGAN
2. Μείωση του βάρους του Style loss σε 1.0
3. Μείωση του TSMGAN σε 0.1
4. Πρόσθεση του masked loss ως 1.0
5. Αύξηση του Style loss σε 10.0 και χρήση του 30% των δεδομένων ως validation set

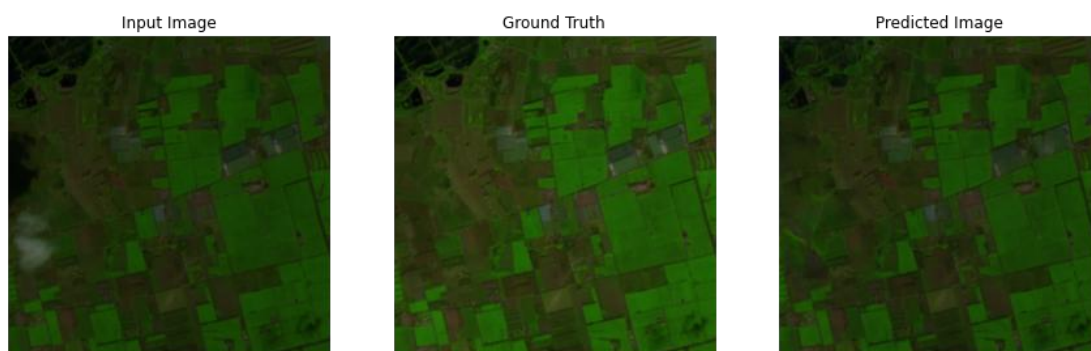
5. ΠΟΙΟΤΙΚΗ ΚΑΙ ΠΟΣΟΤΙΚΗ ΑΞΙΟΛΟΓΗΣΗ

Όπως αναφέρθηκε και στο κεφάλαιο 3, η αξιολόγηση των γεννητικών ανταγωνιστικών δικτύων δεν μπορεί να γίνει με την παρατήρηση της απώλειας καθώς τόσο ο δημιουργός όσο και ο διευκρινιστής εκπαιδεύονται μαζί για να διατηρήσουν μια ισορροπία. Ως εκ τούτου, δεν υπάρχει αντικειμενική λειτουργία απώλειας που χρησιμοποιείται για την εκπαίδευση του δημιουργού και κανένας τρόπος αντικειμενικής αξιολόγησης της προόδου της εκπαίδευσης και της σχετικής ή απόλυτης ποιότητας του μοντέλου από την απώλεια μόνο. Επομένως, για την αξιολόγηση των μεθοδολογιών που παρουσιάστηκαν στο προηγούμενο κεφάλαιο θα παρουσιαστούν οι παραγόμενες εικόνες των δημιουργών και τα αποτελέσματα των μετρικών που υπολογίστηκαν για την αξιολόγηση της ποιότητας των παραγόμενων εικόνων σε σχέση με την εικόνα αναφοράς (ground truth). Οι μετρικές που υπολογίστηκαν ήταν οι MSE, PSRN, SSIM και FID. Οι συγκεκριμένες μετρικές παρουσιάστηκαν αναλυτικά στο κεφάλαιο 3.

5.1 Αποτελέσματα και αξιολόγηση μοντέλου Pix2Pix

Το πρώτο μοντέλο ήταν το pix2pix το οποίο εκπαιδεύτηκε για 300 εποχές. Αρχικά θα παρουσιαστούν οι εικόνες που παράχθηκαν από το μοντέλο για το σύνολο ελέγχου. Επειδή δεν είναι δυνατό να παρουσιαστούν όλες οι εικόνες έγινε μία επιλογή εικόνων ώστε να παρουσιαστούν παραδείγματα επιτυχημένων αποτελεσμάτων και παραδείγματα αποτελεσμάτων που δεν ήταν τα αναμενόμενα.

Επιτυχία Εκπαίδευσης:



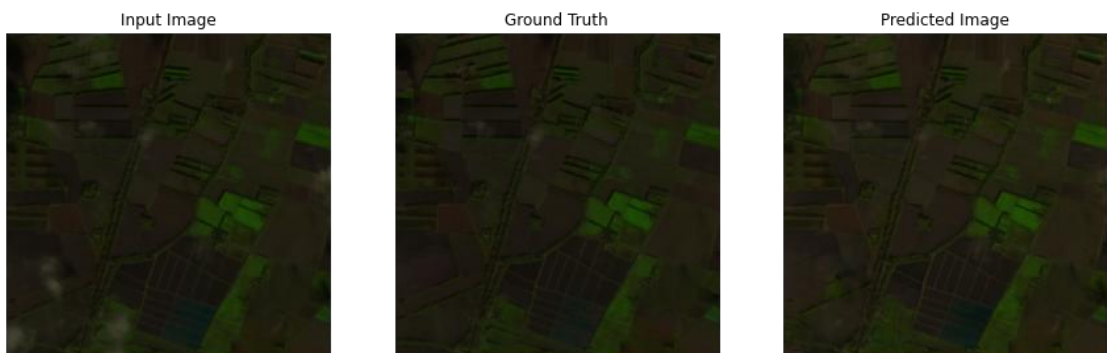
Εικόνα 5-1: 1^ο παράδειγμα επιτυχημένου αποτελέσματος μοντέλου pix2pix.



Εικόνα 5-2: 2^ο παράδειγμα επιτυχημένου αποτελέσματος μοντέλου rix2pix.



Εικόνα 5-3: 3^ο παράδειγμα επιτυχημένου αποτελέσματος μοντέλου rix2pix.



Εικόνα 5-4: 4^ο παράδειγμα επιτυχημένου αποτελέσματος μοντέλου rix2pix.



Εικόνα 5-5: 5^ο παράδειγμα επιτυχημένου αποτελέσματος μοντέλου rix2pix.

Αποτυχία Εκπαίδευσης:



Εικόνα 5-6: 1^ο παράδειγμα μη επιτυχημένου αποτελέσματος μοντέλου rix2rix.



Εικόνα 5-7: 2^ο παράδειγμα μη επιτυχημένου αποτελέσματος μοντέλου rix2rix.



Εικόνα 5-8: 3^ο παράδειγμα μη επιτυχημένου αποτελέσματος μοντέλου rix2rix.



Εικόνα 5-9: 4^ο παράδειγμα μη επιτυχημένου αποτελέσματος μοντέλου rix2rix.



Εικόνα 5-10: 5^ο παράδειγμα μη επιτυχημένου αποτελέσματος μοντέλου rix2rix.

Από τα αποτελέσματα μπορούμε να παρατηρήσουμε ότι το rix2rix έχει καταφέρει να αφαιρέσει με επιτυχία τα σύννεφα σε εικόνες όπου η συννεφοκάλυψη ήταν σε μικρό ποσοστό και να ανακατασκευάσει την εικόνα σε σημαντικό βαθμό. Αντίθετα σε εικόνες όπου το ποσοστό συννεφοκάλυψης είναι μεγαλύτερο δεν έχει καταφέρει να ανακατασκευάσει με ακρίβεια τα χαρακτηριστικά της εικόνας. Κάτι τέτοιο φαίνεται λογικό, αν λάβουμε υπόψη ότι το συγκεκριμένο μοντέλο εκπαιδεύτηκε από το μηδέν με πολύ λιγότερα δεδομένα από αυτά που απαιτούσε. Παρ' όλα αυτά και στις εικόνες που έχει αποτύχει η εκπαίδευση μπορούμε να διακρίνουμε ότι το μοντέλο έχει προσπαθήσει προς την σωστή κατεύθυνση να ανακατασκευάσει την εικόνα κι επίσης και στις δύο περιπτώσεις (επιτυχημένα – αποτυχημένα αποτελέσματα) έχει καταφέρει να εξαφανίσει τις σκιές των σύννεφων. Από τα αποτελέσματα καταλαβαίνουμε ότι η εκπαίδευση είναι ελλιπής, λόγω έλλειψης δεδομένων.

Παρακάτω παρουσιάζονται τα αποτελέσματα των μετρικών που υπολογίστηκαν για τις 50 εικόνες του συνόλου ελέγχου που δημιούργησε ο δημιουργός του μοντέλου rix2rix. Ως εικόνα αναφοράς χρησιμοποιήθηκε η εικόνα που δινόταν με το σύνολο δεδομένων και είχε δημιουργηθεί με το φίλτρο μεσαίας τιμής. Οι συγκεκριμένες εικόνες χρησιμοποιήθηκαν και κατά την εκπαίδευση του μοντέλου.

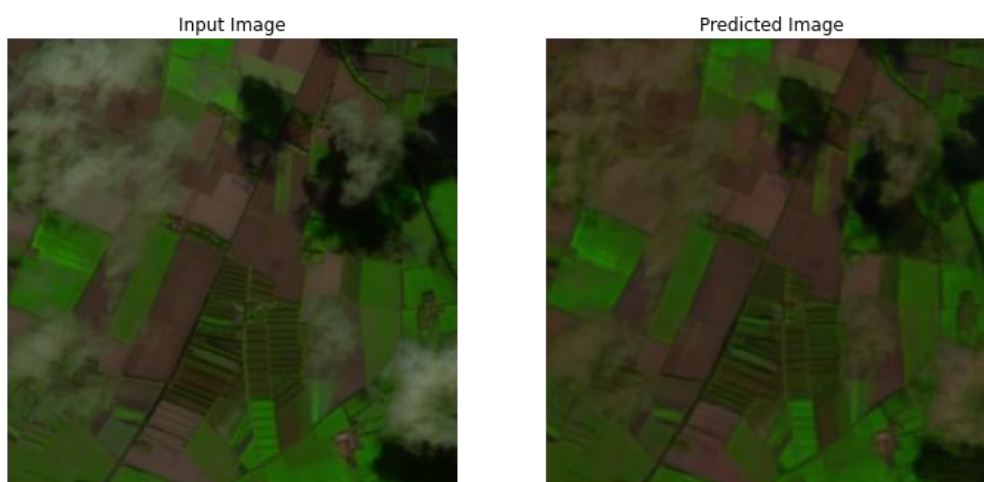
Metric	MSE	PSNR	SSIM	FID
Value	0.010	29.194	0.849	6.092

Το μέσο τετραγωνικό σφάλμα (MSE) είναι σχετικά κοντά στο μηδέν που σημαίνει ότι οι τιμές των εικόνων που δημιούργησε ο δημιουργός είναι κοντά στις τιμές της εικόνας αναφοράς. Το ίδιο ισχύει και για την τιμή της μέγιστης αναλογίας σήματος προς θόρυβο

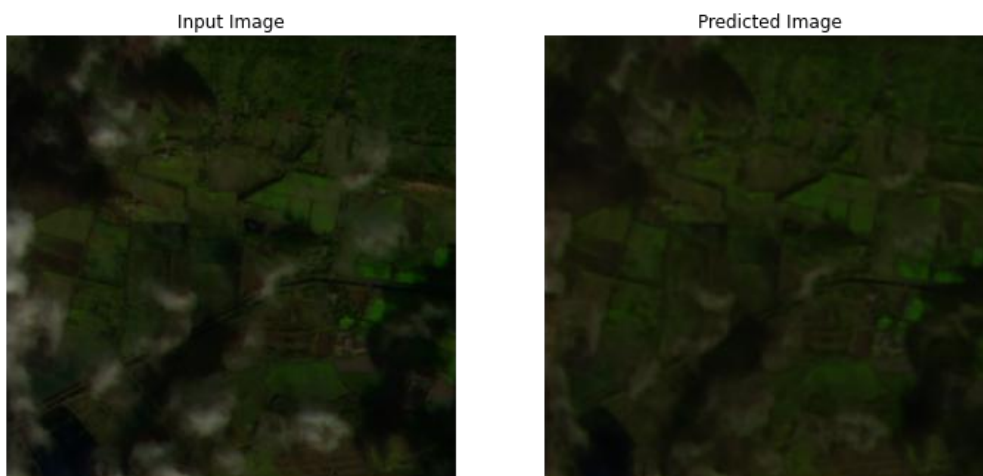
(PSNR) όπου είναι κοντά στα 30 db. Συνήθως τιμές μεταξύ του 30-50 db μας υποδεικνύουν ότι η εικόνα είναι σε καλή ποιότητα. Αυτές οι δύο μετρικές απλώς ποσοτικοποιούν την ποιότητα της ανακατασκευασμένης εικόνας με την εικόνα αναφοράς. Μεγαλύτερη βάση δίνεται στις τιμές των επόμενων δύο μετρικών που θεωρούνται πιο αξιόπιστες. Ο δείκτης δομικής ομοιότητας (SSIM) βασίζεται στις δομικές πληροφορίες των εικόνων. Οι τιμές που μπορεί να πάρει ο συγκεκριμένος δείκτης είναι μεταξύ του 0 και του 1, όπου το 1 σημαίνει ότι η ανακατασκευασμένη εικόνα ταιριάζει τέλεια με την αρχική. Επομένως η τιμή 0.849 δείχνει ότι η ανακατασκευασμένη εικόνα ταιριάζει δομικά σε ένα καλό ποσοστό με την αρχική. Τέλος για την απόσταση έναρξης Frechet (FID) οι ιδανικές τιμές είναι κοντά στο μηδέν, άρα η τιμή 6.092 δεν θεωρείται καλή. Θα πρέπει όμως να λάβουμε υπόψη μας ότι η συγκεκριμένη μετρική επηρεάζεται από τον αριθμό δεδομένων κι επομένως ο μικρός αριθμός των δεδομένων με τον οποίο εκπαιδεύτηκε το μοντέλο επηρέασε αρνητικά τη τιμή της μετρικής.

5.2 Αποτελέσματα και αξιολόγηση μοντέλου CycleGAN

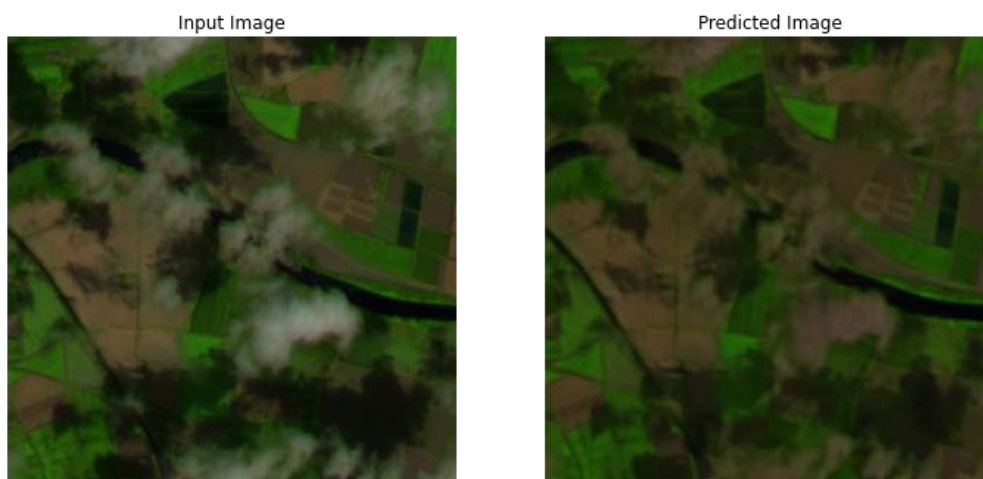
Το δεύτερο μοντέλο ήταν το CycleGAN το οποίο εκπαιδεύτηκε για 200 εποχές. Το συγκεκριμένο μοντέλο δεν εκπαιδεύτηκε συνεχόμενα λόγω του σφάλματος που προέκυψε από την έλλειψη RAM στο Google Colaboratory, όπου κάθε 45 εποχές μας σταματούσε την εκπαίδευση. Παρακάτω ακολουθούν κάποια ενδεικτικά αποτελέσματα του μοντέλου.



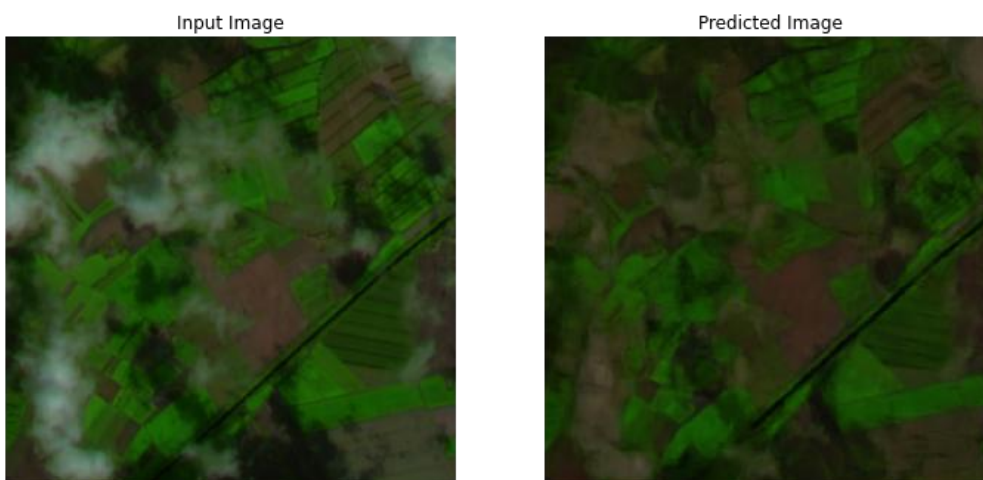
Εικόνα 5-11: 1^ο ενδεικτικό αποτέλεσμα μοντέλου CycleGAN.



Εικόνα 5-12: 2^ο ενδεικτικό αποτέλεσμα μοντέλου CycleGAN.



Εικόνα 5-13: 3^ο ενδεικτικό αποτέλεσμα μοντέλου CycleGAN.



Εικόνα 5-14: 4^ο ενδεικτικό αποτέλεσμα μοντέλου CycleGAN.

Όπως παρατηρούμε στην εποχή 200 το μοντέλο έχει εντοπίσει τα σύννεφα και τα έχει αφαιρέσει. Παρ' όλα αυτά δεν έχει ανακατασκευάσει με ακρίβεια την εικόνα. Σίγουρα χρειάζεται αρκετή εκπαίδευση και πολύ μεγαλύτερο αριθμό δεδομένων για να μπορέσει να δημιουργήσει αληθοφανείς εικόνες, καθώς το συγκεκριμένο μοντέλο δεν χρησιμοποιεί κάποια εικόνα αναφοράς.

Παρακάτω παρουσιάζονται τα αποτελέσματα των μετρικών που υπολογίστηκαν για τις 50 εικόνες του συνόλου ελέγχου που δημιούργησε ο δημιουργός του μοντέλου CycleGAN. Ως εικόνα αναφοράς χρησιμοποιήθηκε η εικόνα που δινόταν με το σύνολο δεδομένων και είχε δημιουργηθεί με το φίλτρο μεσαίας τιμής. Υπενθυμίζεται ότι πάνω στο ίδιο σύνολο ελέγχου υπολογίστηκαν και οι μετρικές του προηγούμενου μοντέλου ώστε να γίνει η σύγκριση τους.

Metric	MSE	PSNR	SSIM	FID
Value	0.016	28.934	0.821	7.488

Παρατηρούμε ότι οι τιμές των μετρικών συμφωνούν με τα οπτικά αποτελέσματα. Οι τιμές και των τεσσάρων μετρικών υποδεικνύουν ότι οι παραγόμενες εικόνες δεν είναι ποιοτικές. Παρ' όλα αυτά οι τιμές δεν απέχουν πάρα πολύ από τις τιμές που θα θεωρούσαμε ότι τα αποτελέσματα είναι ποιοτικά, γεγονός που μας δείχνει ότι αν υπήρχε η δυνατότητα να εκπαιδευτεί το μοντέλο με περισσότερα δεδομένα και για περισσότερες εποχές θα μπορούσε να δώσει ικανοποιητικά αποτελέσματα. Σε σύγκριση με τις τιμές των μετρικών του μοντέλου rix2rix, διαπιστώνουμε ότι το rix2rix υπερτερεί ως προς τα αποτελέσματα. Ωστόσο, τα αποτελέσματα τους δεν απέχουν και τόσο πολύ, αν αναλογιστούμε ότι το rix2rix εκπαιδεύεται με εικόνα αναφοράς ενώ το CycleGAN εκπαιδεύεται στα «τυφλά» γνωρίζοντας μόνο την κατανομή των εικόνων στο πεδίο στόχου. Το γεγονός ότι οι τιμές των μετρικών των δύο μοντέλων είναι τόσο κοντά, έρχεται σε αντίθεση με την παρατήρηση των οπτικών αποτελεσμάτων κατά τα οποία φάνηκε ότι το CycleGAN δεν έχει ανακατασκευάσει σωστά καμία εικόνα. Αυτό οφείλεται στο γεγονός ότι οι μετρικές εξετάζουν χαρακτηριστικά της εικόνας, όπως είναι η απόχρωση. Αυτό σημαίνει ότι κατά την εκπαίδευση του CycleGAN ο δημιουργός δεν κατάφερε να δημιουργήσει αληθοφανείς εικόνες αλλά στην προσπάθεια του να ξεγελάσει τον διευκρινιστή δημιούργησε εικόνες που έχουν την ίδια απόχρωση και αντίθεση με την αρχική εικόνα. Γι' αυτό το λόγο οι τιμές των μετρικών εμφανίζονται να είναι αρκετά καλές.

Επιπρόσθετα αξίζει να γίνει αναφορά στα αποτελέσματα του μοντέλου Cloud-GAN. Το μοντέλο Cloud-GAN [4], όπως αναφέρθηκε και στο κεφάλαιο 2 χρησιμοποιεί το μοντέλο CycleGAN για την αντιμετώπιση του ίδιου προβλήματος που εξετάζει η παρούσα διπλωματική, δηλαδή την ανακατασκευή χωρίς σύννεφα δορυφορικών εικόνων Sentinel-2. Οι διαφορές είναι ότι στο Cloud-GAN ο δημιουργός ακολουθεί αρχιτεκτονική ResNet, ενώ στη δική μας μεθοδολογία ακολουθεί αρχιτεκτονική U-Net. Επίσης, η εκπαίδευση του Cloud-GAN έγινε με έναν αρκετά μεγαλύτερο αριθμό δεδομένων. Συγκεκριμένα χρησιμοποιήθηκαν 1677 εικόνες από κάθε κατηγορία (α. με σύννεφα, β. χωρίς σύννεφα) ως δεδομένα εκπαίδευσης και 837 εικόνες ως δεδομένα ελέγχου, σε αντίθεση με τα δικά μας δεδομένα που ήταν 240 και 50 εικόνες αντίστοιχα. Οι εικόνες που χρησιμοποιήθηκαν προέρχονταν από τον Sentinel-2 και για την εκπαίδευση επιλέχθηκαν τα κανάλια του ορατού φάσματος (RGB), όπως και στην δική μας μεθοδολογία. Η εκπαίδευση του μοντέλου έγινε για 200 εποχές. Τα αποτελέσματα όσον αφορά τις παραγόμενες εικόνες από τον δημιουργό έδειξαν ότι μπορεί να ανακατασκευάσει με επιτυχία περιοχές όπου η συννεφοκάλυψη είναι αραιή, ενώ για περιοχές όπου τα σύννεφα είναι πυκνά και καταλαμβάνουν το μεγαλύτερο τμήμα της εικόνας το Cloud-GAN απέτυχε εντελώς να παράγει ανακατασκευασμένες εικόνες. Επομένως από την αξιολόγηση των παραγόμενων εικόνων συμπεραίνουμε ότι και τα δύο μοντέλα είχαν τα ίδια αποτελέσματα όσον αφορά την ποικιλία της συννεφοκάλυψης. Όσον αφορά τις μετρικές, στο Cloud-GAN δίνεται ο δείκτης PSRN μόνο για 5 εικόνες. Η τιμή του είναι περίπου 28-28.5 db που είναι πολύ κοντά στο δικό μας αποτέλεσμα (28.934).

5.3 Αποτελέσματα και αξιολόγηση μοντέλου LGTSM

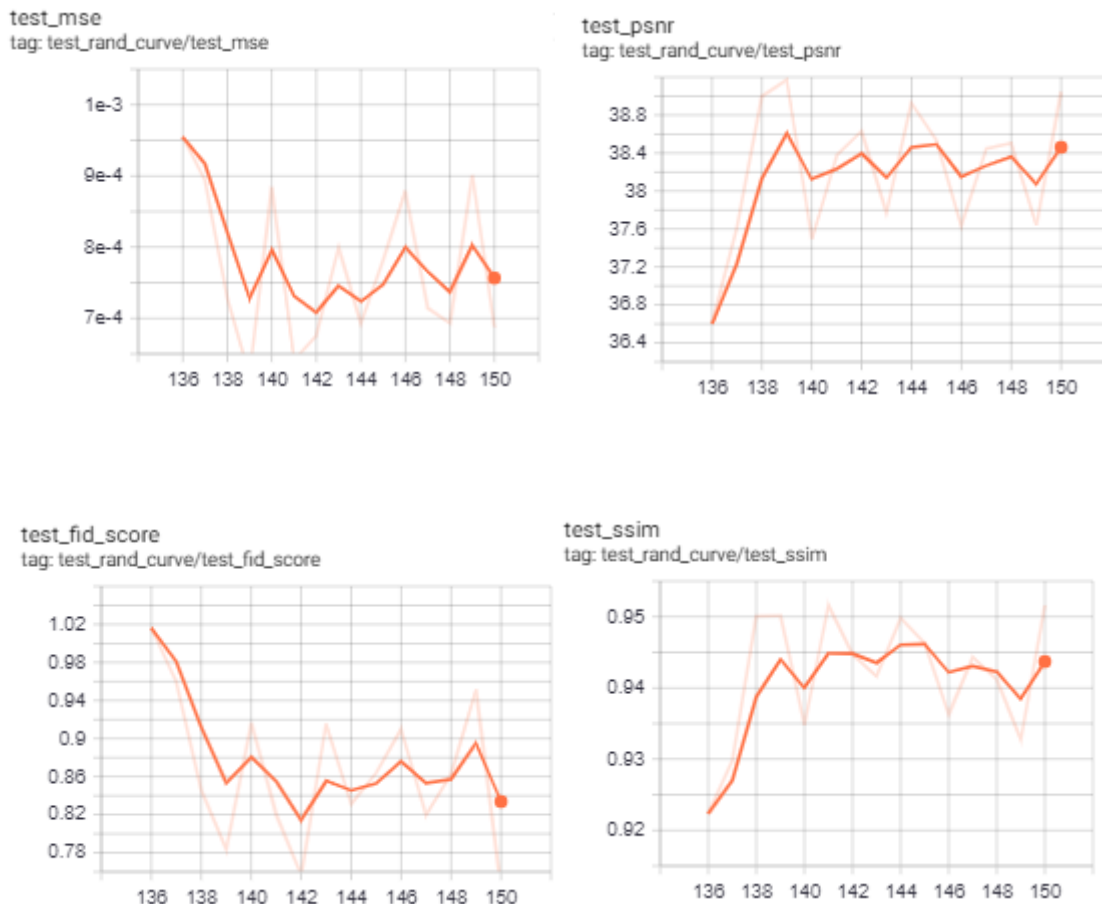
Το τρίτο μοντέλο ήταν το Learnable Gated Temporal Shift Module (LGTSM). Όπως αναφέρθηκε και στο κεφάλαιο της μεθοδολογίας το συγκεκριμένο μοντέλο δεν ήταν δυνατό να εκπαιδευτεί από την αρχή λόγω έλλειψης υπολογιστικών πόρων. Η λύση που βρέθηκε ήταν να χρησιμοποιηθεί το προ-εκπαιδευμένο μοντέλο που παρείχαν οι συγγραφείς και να συνεχιστεί η εκπαίδευση για 15 εποχές ακόμα, ώστε να γίνει fine-tuning. Για το σκοπό αυτό έγιναν κάποια πειράματα ως προς το βάρος που έχουν οι απώλειες. Παρακάτω, εξετάζονται τα αποτελέσματα ανά πείραμα.

5.3.1 1^ο πείραμα (χρονοσειρά)

Για το πρώτο πείραμα αφήσαμε τα βάρη όπως ήταν κατά τη διάρκεια της προ-εκπαίδευσης, προσθέτοντας μόνο το gan-temporal loss με βάρος ίσο με τη μονάδα.

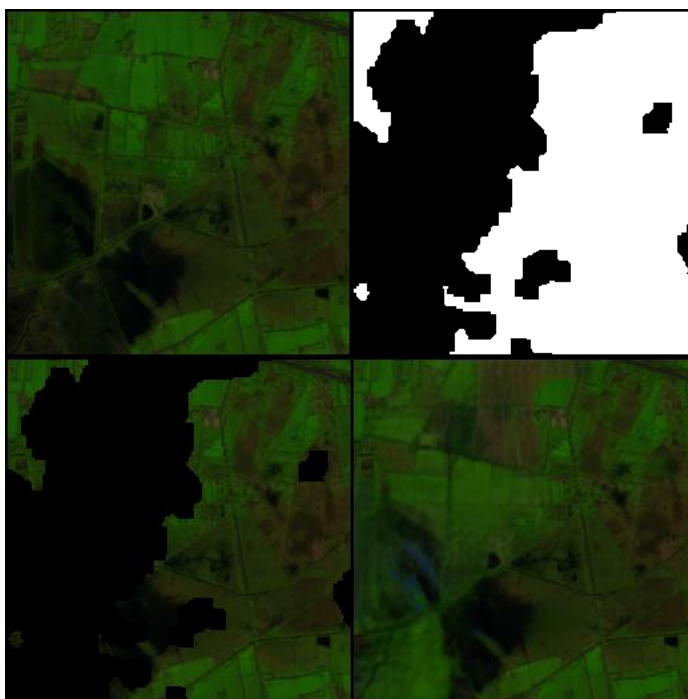
Loss reconstruction = 1, Perceptual loss = 1, Style loss =10, Gan-temporal loss =1

Παρακάτω φαίνονται πως μεταβάλλονται οι τιμές των μετρικών κατά την εκπαίδευση:

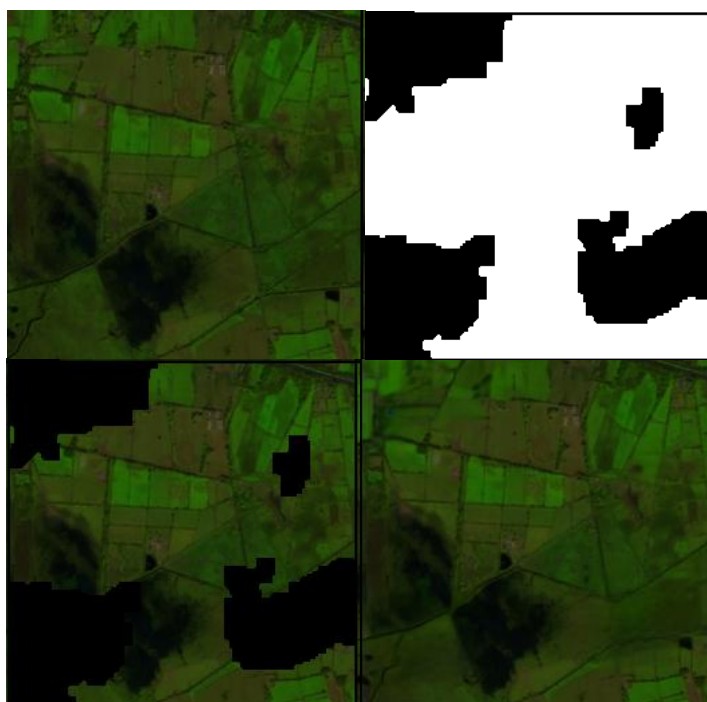


Εικόνα 5-15: Διαγράμματα μετρικών για το 1^ο πείραμα (MSE: πάνω αριστερά, PSNR: πάνω δεξιά, FID: κάτω αριστερά και SSIM: κάτω δεξιά).

Όπως παρατηρούμε το MSE στην εποχή 150 πλησιάζει το μηδέν ($6.8 \cdot 10^{-4}$) που σημαίνει ότι η ανακατασκευασμένη εικόνα είναι κοντά στις τιμές με την εικόνα αναφοράς (χωρίς τις μάσκες). Η τιμή του SSIM στην εποχή 150 είναι περίπου 0.95, πολύ κοντά στο 1, που είναι το τέλειο. Η τιμή του PSNR βρίσκεται μεταξύ του 30-50db και η τιμή της FID απόστασης είναι 0.74 και γενικά σε όλες τις εποχές είναι κάτω του 1, που σημαίνει ότι οι ανακατασκευασμένες εικόνες είναι πολύ κοντά με τις εικόνες αναφοράς. Παρακάτω ακολουθούν δύο ενδεικτικά παραδείγματα εικόνων που προέκυψαν κατά την εκπαίδευση για το σύνολο ελέγχου.



Εικόνα 5-16: 1^ο ενδεικτικό αποτέλεσμα 1^{ου} πειράματος μοντέλου LGTSM (ground truth: πάνω αριστερά, μάσκα: πάνω δεξιά, είσοδος μοντέλου: κάτω αριστερά και ανακατασκευασμένη εικόνα: κάτω δεξιά).



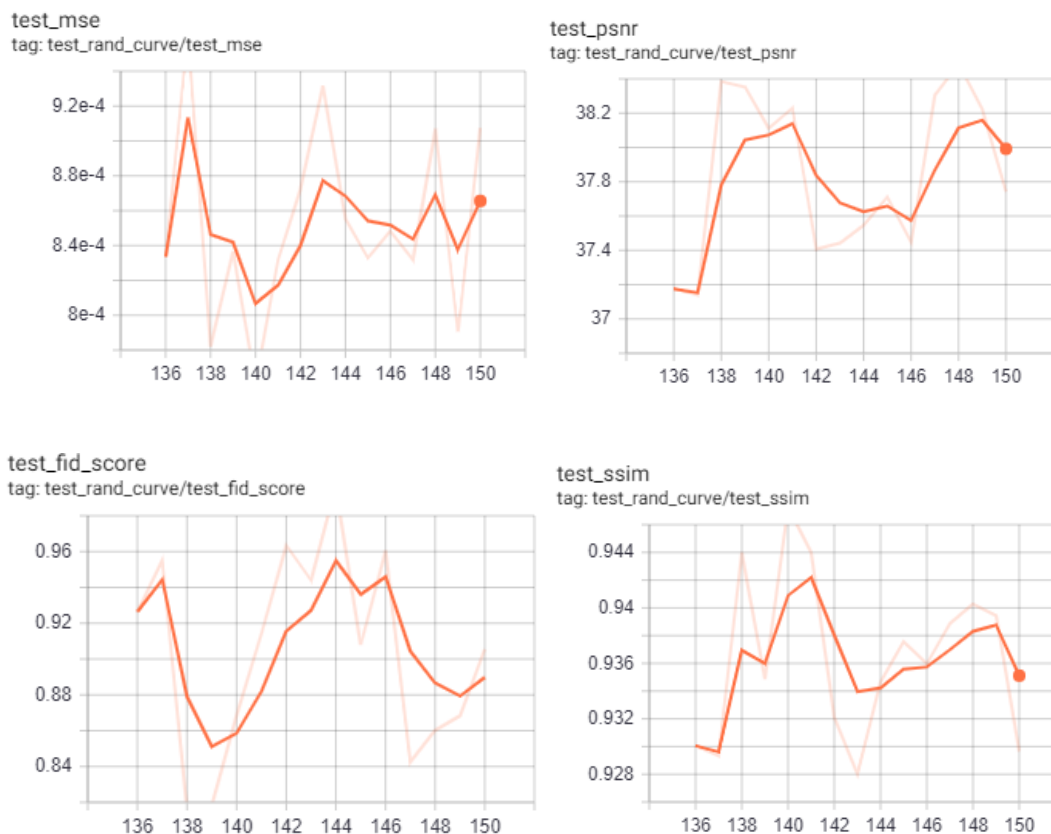
Εικόνα 5-17: 2^ο ενδεικτικό αποτέλεσμα 1^{ου} πειράματος μοντέλου LGTSM (ground truth: πάνω αριστερά, μάσκα: πάνω δεξιά, είσοδος μοντέλου: κάτω αριστερά και ανακατασκευασμένη εικόνα: κάτω δεξιά).

Παρατηρούμε ότι στην πρώτη περίπτωση που η μάσκα καλύπτει μεγαλύτερο τμήμα της εικόνας έχει αλλοιώσει λίγο ένα τμήμα της. Στην δεύτερη περίπτωση φαίνεται το μοντέλο να έχει λειτουργήσει τέλεια.

5.3.2 2^ο πείραμα (χρονοσειρά)

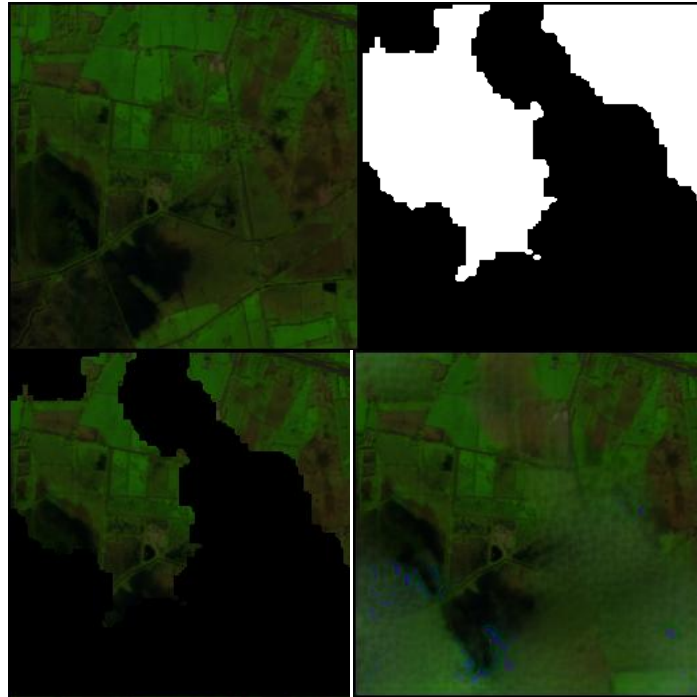
Για το δεύτερο πείραμα μειώσαμε το βάρος του loss style σε 1, για να δούμε πως θα επηρεαστεί το τελικό αποτέλεσμα.

Loss reconstruction = 1, Perceptual loss = 1, Style loss = 1, Gan-temporal loss = 1

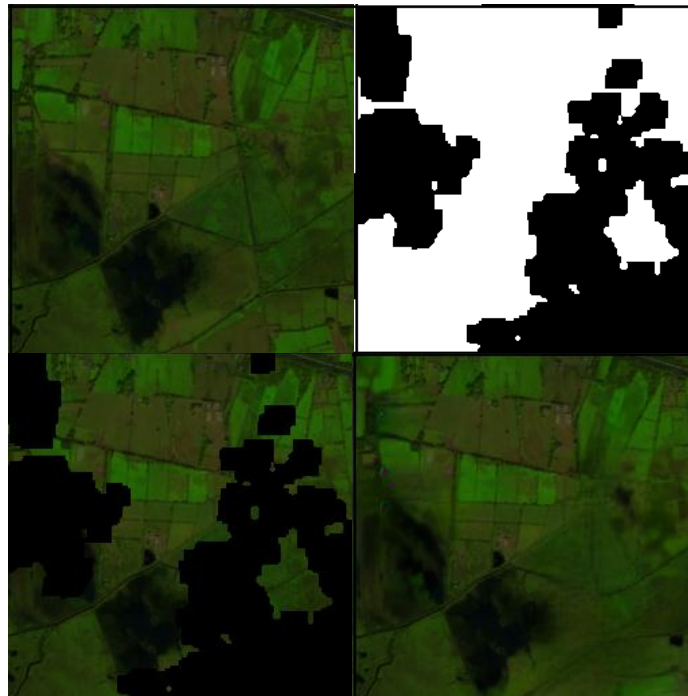


Εικόνα 5-18: Διαγράμματα μετρικών για το 2^ο πείραμα (MSE: πάνω αριστερά, PSNR: πάνω δεξιά, FID: κάτω αριστερά και SSIM: κάτω δεξιά).

Από τη μεταβολή των τιμών των μετρικών κατά την εκπαίδευση παρατηρούμε ότι η εκπαίδευση ίσως θα έπρεπε στη συγκεκριμένη περίπτωση να σταματήσει νωρίτερα. Στην εποχή 140 περίπου οι τιμές είναι καλύτερες ως προς την αξιολόγηση της ποιότητας της ανακατασκευασμένης εικόνας από την εικόνα αναφοράς, σε σχέση με την εποχή 150. Παρακάτω ακολουθούν κάποια αποτελέσματα της εκπαίδευσης του 2^{ου} πειράματος για το σύνολο ελέγχου.

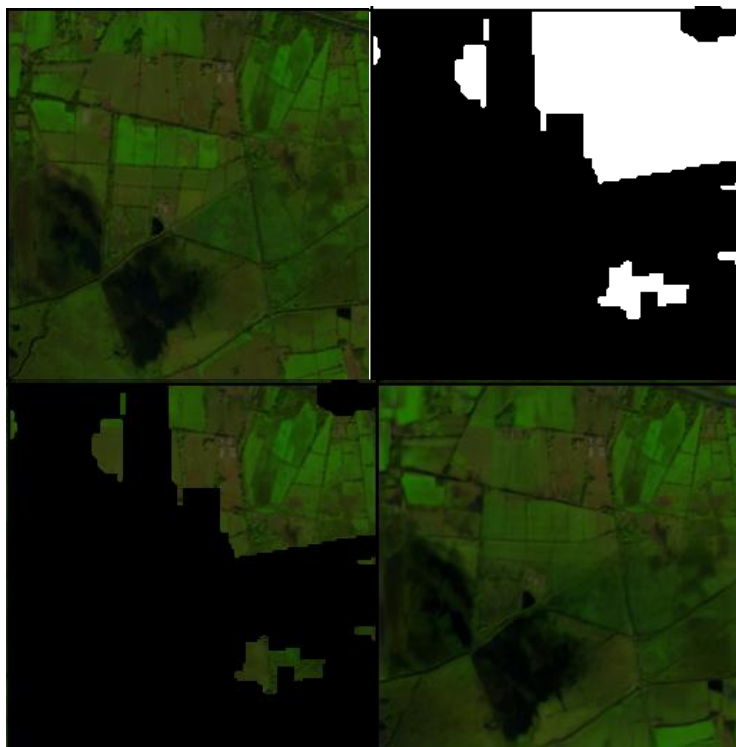


Εικόνα 5-19: 1^ο ενδεικτικό αποτέλεσμα 2^{ου} πειράματος μοντέλου LGTSM για την εποχή 150 (ground truth: πάνω αριστερά, μάσκα: πάνω δεξιά, είσοδος μοντέλου: κάτω αριστερά και ανακατασκευασμένη εικόνα: κάτω δεξιά).



Εικόνα 5-20: 2^ο ενδεικτικό αποτέλεσμα 2^{ου} πειράματος μοντέλου LGTSM για την εποχή 150 (ground truth: πάνω αριστερά, μάσκα: πάνω δεξιά, είσοδος μοντέλου: κάτω αριστερά και ανακατασκευασμένη εικόνα: κάτω δεξιά).

Όντως σε σύγκριση με το προηγούμενο πείραμα βλέπουμε ότι υπάρχει αρκετή αλλοίωση στην τελική εικόνα του 1^{ου} αποτελέσματος κυρίως. Ακολουθούν τα ενδεικτικά αποτελέσματα για την εποχή 140.



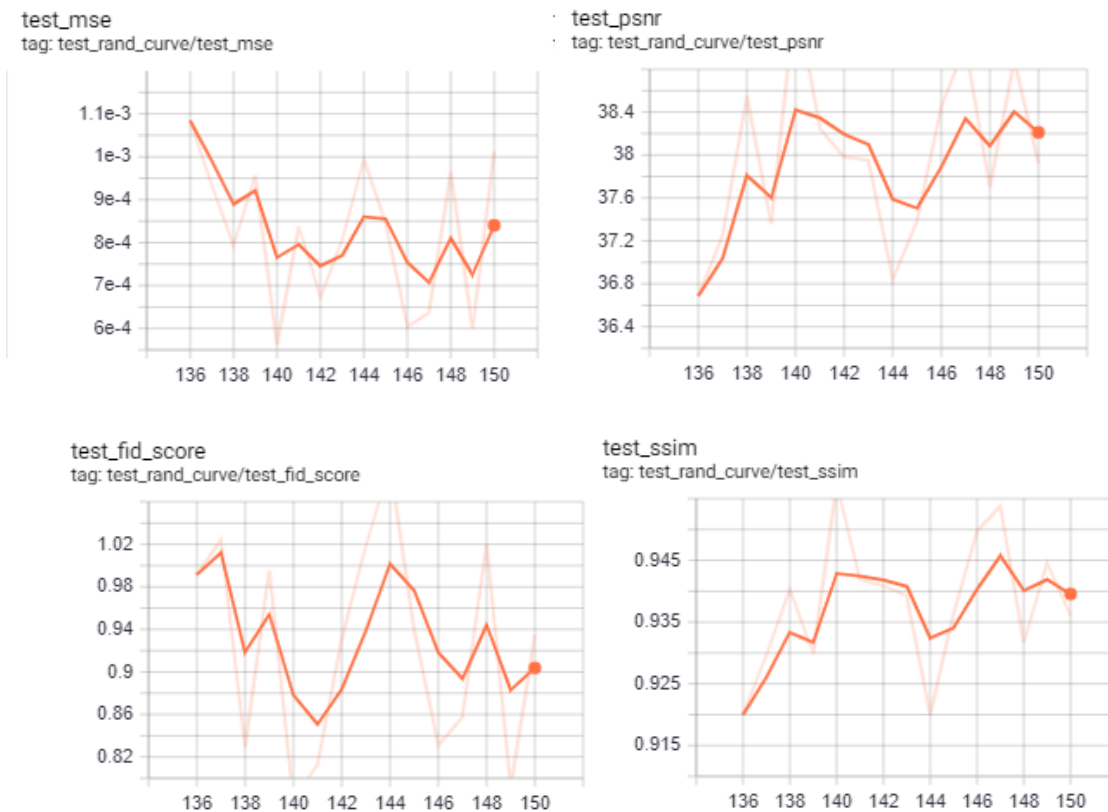
Εικόνα 5-21: 2^ο ενδεικτικό αποτέλεσμα 2ου πειράματος μοντέλου LGTSM για την εποχή 140 (ground truth: πάνω αριστερά, μάσκα: πάνω δεξιά, είσοδος μοντέλου: κάτω αριστερά και ανακατασκευασμένη εικόνα: κάτω δεξιά).

Από το αποτέλεσμα της εποχής 140 μπορούμε να δούμε ότι η ανακατασκευασμένη εικόνα φαίνεται να μοιάζει αρκετά με την εικόνα αναφοράς, χωρίς να υπάρχουν αλλοιώσεις όπως στην εποχή 150. Δεν έχουν αποδοθεί πλήρως κάποιες ακμές της εικόνας αλλά σε γενικές γραμμές το αποτέλεσμα είναι πολύ καλό σε σύγκριση με την εποχή 150 του ίδιου πειράματος. Επομένως, με τη μείωση του βάρους του Style loss καταλαβαίνουμε ότι το μοντέλο μπορεί να δώσει αξιολογικά αποτελέσματα σε σύντομο χρόνο.

5.3.3 3^ο πείραμα (χρονοσειρά)

Για το τρίτο πείραμα μειώθηκε το βάρος του gan temporal loss σε 0.1.

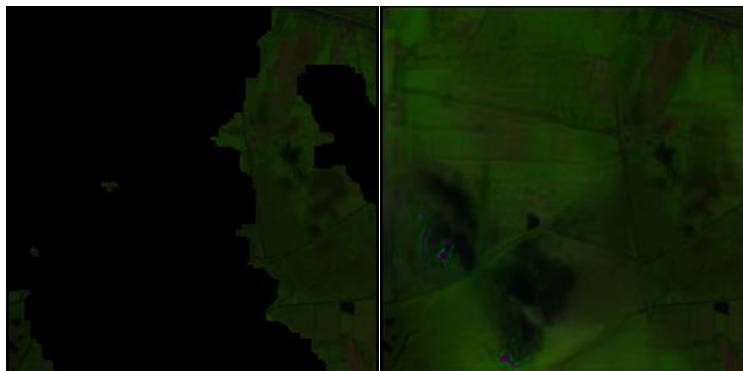
Loss reconstruction = 1, Perceptual loss = 1, Style loss = 1, Gan-temporal loss = 0.1



Εικόνα 5-22: Διαγράμματα μετρικών για το 3^ο πείραμα (MSE: πάνω αριστερά, PSNR: πάνω δεξιά, FID: κάτω αριστερά και SSIM: κάτω δεξιά).

Παρατηρούμε ότι η μείωση του βάρους του gan temporal loss δεν έδωσε καλύτερα αποτελέσματα. Οι τιμές των μετρικών είναι μέσα στα όρια αλλά έχουν χειροτερέψει. Το ίδιο προκύπτει και από την παρατήρηση του οπτικού αποτελέσματος, όπου φαίνεται ότι το τελικό αποτέλεσμα είναι αλλοιωμένο και θολό.



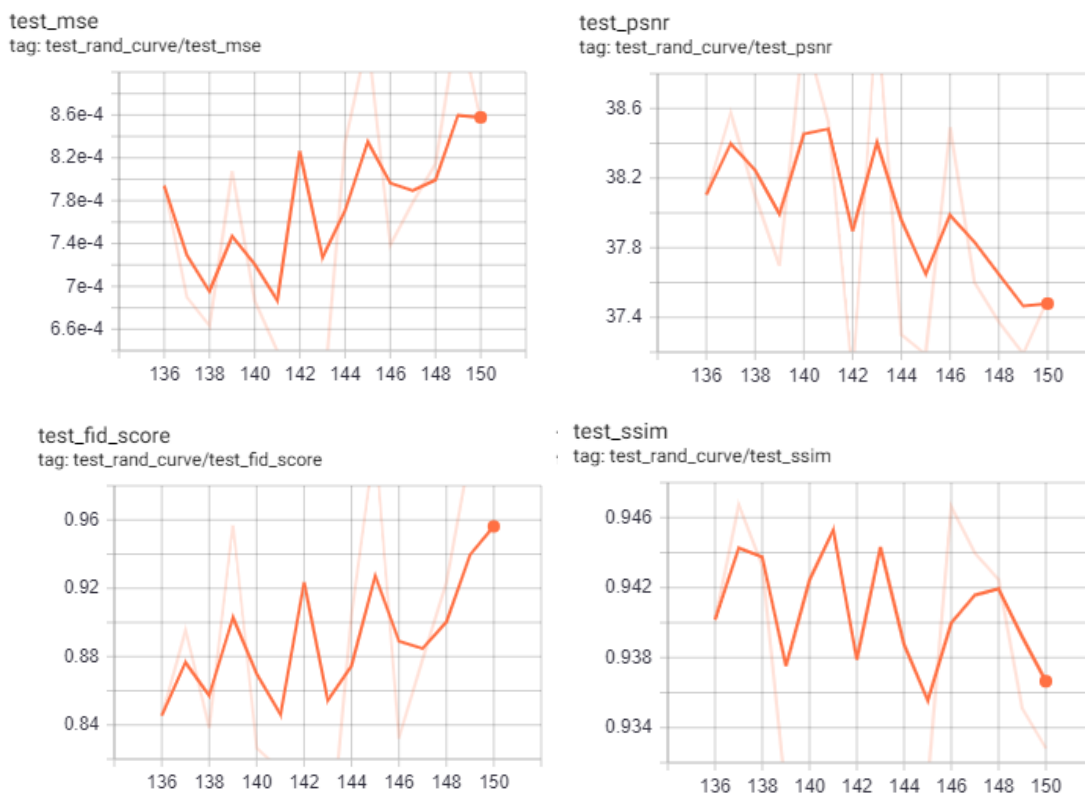


Εικόνα 5-23: Ενδεικτικό αποτέλεσμα 3^{ου} πειράματος μοντέλου LGTSM (ground truth: πάνω αριστερά, μάσκα: πάνω δεξιά, είσοδος μοντέλου: κάτω αριστερά και ανακατασκευασμένη εικόνα: κάτω δεξιά).

5.3.4 4^ο πείραμα (χρονοσειρά)

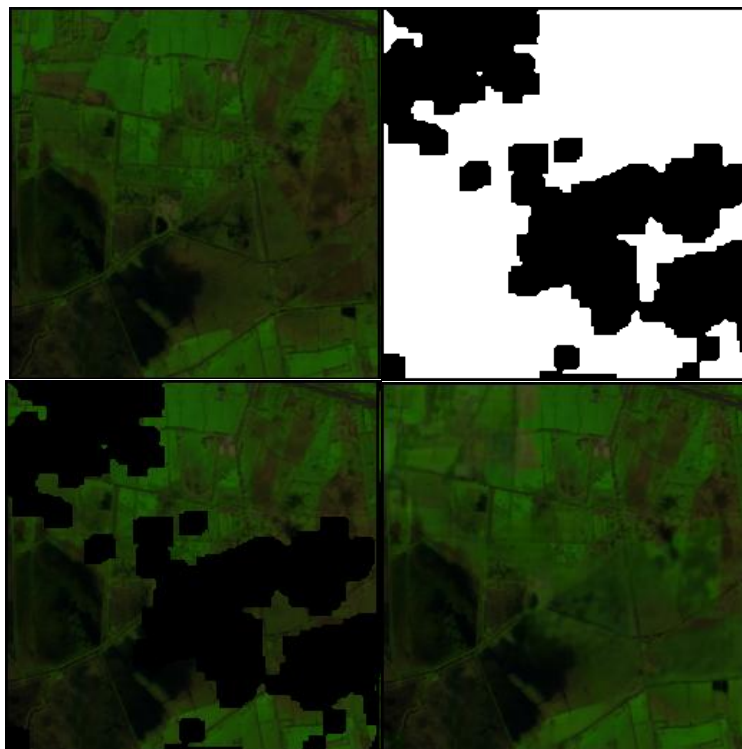
Για το τέταρτο πείραμα προσθέσαμε το βάρος του masked loss ίσο με τη μονάδα.

Loss reconstruction = 1, Masked loss reconstruction = 1, Perceptual loss = 1, Style loss =1, Gan-temporal loss =1



Εικόνα 5-24: Διαγράμματα μετρικών για το 4^ο πείραμα (MSE: πάνω αριστερά, PSNR: πάνω δεξιά, FID: κάτω αριστερά και SSIM: κάτω δεξιά).

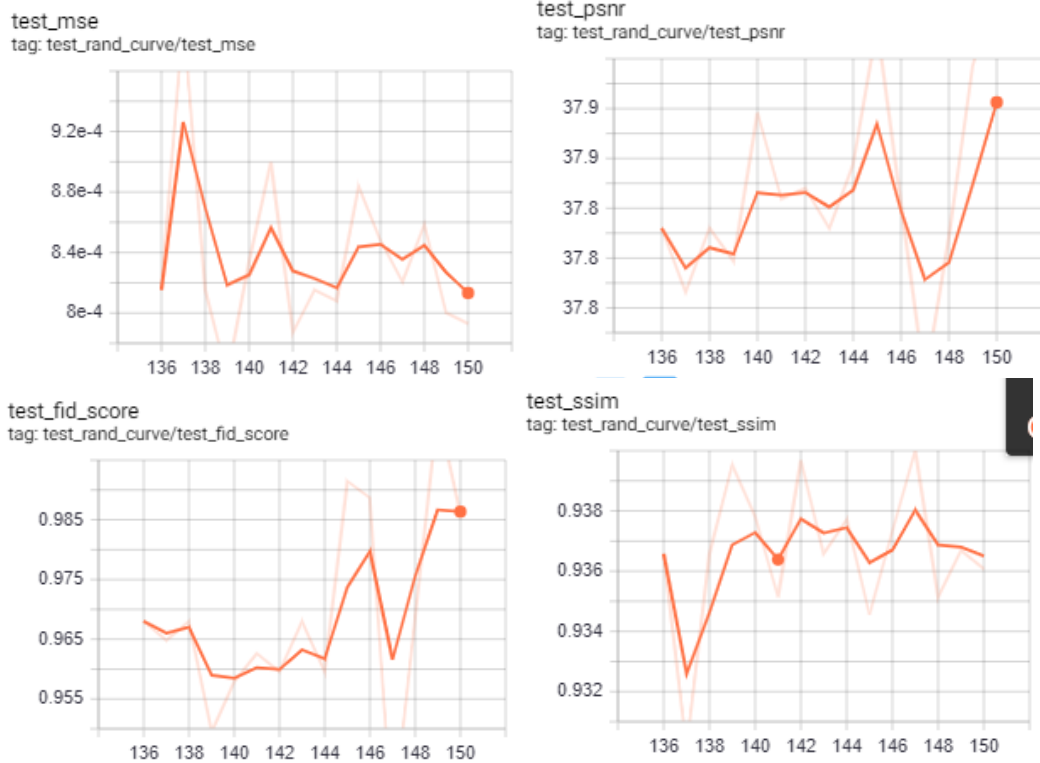
Και σε αυτό το πείραμα παρατηρούμε ότι οι τιμές των μετρικών βρίσκονται μέσα στα επιθυμητά όρια. Όμως οι καλύτερες τιμές έχουν εμφανιστεί στην εποχή 141. Επομένως παρά την εισαγωγή του masked loss στην εκπαίδευση βλέπουμε ότι το μοντέλο σύμφωνα με τα διαγράμματα συνεχίζει να έχει ποιοτικά αποτελέσματα σε πιο λίγες εποχές. Από την παρατήρηση των εικόνων που προέκυψαν παρατηρούμε ότι το αποτέλεσμα μοιάζει αρκετά με την εικόνα αναφοράς αλλά είναι πιο θολό.



Εικόνα 5-25: Ενδεικτικό αποτέλεσμα 4^{ου} πειράματος μοντέλου LGTSM (ground truth: πάνω αριστερά, μάσκα: πάνω δεξιά, είσοδος μοντέλου: κάτω αριστερά και ανακατασκευασμένη εικόνα: κάτω δεξιά).

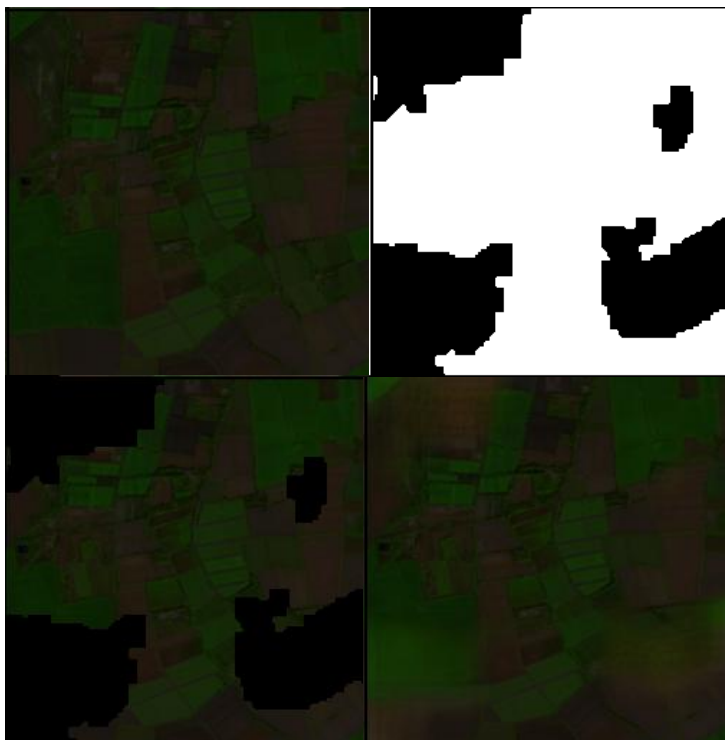
5.3.5 5^ο Πείραμα (χρονοσειρά)

Από το προηγούμενα πειράματα είδαμε ότι οι τιμές των μετρικών ήταν καλύτερες στο πρώτο πείραμα, όπου αφήσαμε τα βάρη ίδια με αυτά που είχε το μοντέλο κατά την προ-εκπαίδευση του, προσθέτοντας μόνο το βάρος του gan temporal loss ίσο με τη μονάδα. Γι' αυτό το λόγο εκπαιδεύσαμε ξανά το μοντέλο με τα ίδια βάρη, όπως στο πείραμα 1 αλλά με ένα ποσοστό 30% του συνόλου δεδομένων να λειτουργεί ως validation set, θέλοντας να παρατηρήσουμε αν θα επηρεάσει τα αποτελέσματα της εκπαίδευσης.



Εικόνα 5-26: Διαγράμματα μετρικών για το 5^ο πείραμα (MSE: πάνω αριστερά, PSNR: πάνω δεξιά, FID: κάτω αριστερά και SSIM: κάτω δεξιά).

Από τα διαγράμματα παρατηρούμε ότι οι τιμές των μετρικών σε σχέση με το πρώτο πείραμα δεν βελτιώθηκαν. Επομένως, το validation set δεν βοήθησε στο fine-tuning του μοντέλου. Και από τις εικόνες που προέκυψαν συμπεραίνουμε το ίδιο καθώς υπάρχει σημαντική αλλοίωση στην τελική εικόνα.



Εικόνα 5-27: Ενδεικτικό αποτέλεσμα 5^{ου} πειράματος μοντέλου LGTSM (ground truth: πάνω αριστερά, μάσκα: πάνω δεξιά, είσοδος μοντέλου: κάτω αριστερά και ανακατασκευασμένη εικόνα: κάτω δεξιά).

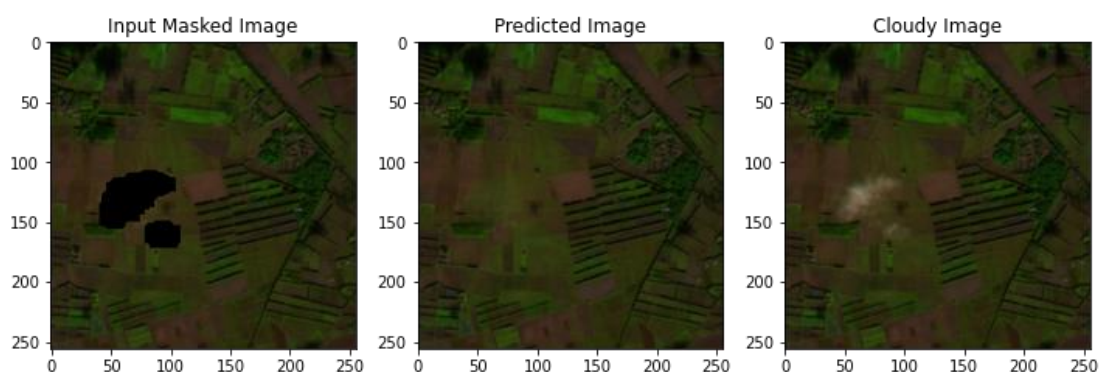
Παρακάτω παρουσιάζονται συγκριτικά οι μέσοι όροι για τις 15 εποχές των τιμών των μετρικών που προέκυψαν για τα πειράματα που τρέξαμε.

	MSE	PSRN	SSIM	FID
Πείραμα 1	0.0007	38.226	0.941	0.866
Πείραμα 2	0.013	37.846	0.936	0.903
Πείραμα 3	0.012	38.014	0.938	0.923
Πείραμα 4	0.011	37.926	0.940	0.899
Πείραμα 5	0.0008	35.328	0.936	1.099

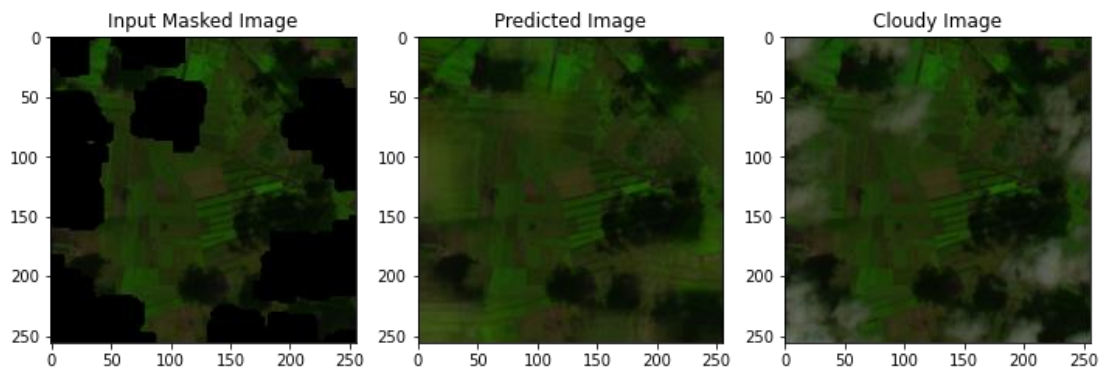
Από τις τιμές παρατηρούμε ότι διαπιστώσαμε και από την προηγούμενη ανάλυση, ότι το πείραμα 1 υπερτερεί με μικρή διαφορά ως προς τις τιμές των μετρικών για την ποιότητα των εικόνων που παράγει. Επίσης σε σύγκριση με τις αντίστοιχες τιμές που προέκυψαν από το μοντέλο Pix2Pix και CycleGAN διαπιστώνουμε ότι αν και τα σύνολο δεδομένων που χρησιμοποιήθηκε ήταν διαφορετικό το μοντέλο LGTSM έχει δώσει σε όλα τα πειράματα του πολύ καλύτερες τιμές μετρικών.

Γι' αυτό το λόγο το συγκεκριμένο μοντέλο (1^ο πείραμα) εφαρμόστηκε σε πραγματικές εικόνες με σύννεφα, των οποίων τη μάσκα την είχαμε από το αρχικό σύνολο δεδομένων. Λόγω του ότι δεν γίνεται να παρουσιαστούν όλα τα αποτελέσματα, έγινε επιλογή κάποιων ενδεικτικών αποτελεσμάτων με τέτοιο τρόπο, ώστε να παρουσιαστούν κι επιτυχημένα αποτελέσματα αλλά και αποτελέσματα στα οποία το μοντέλο δεν λειτούργησε όπως περιμέναμε.

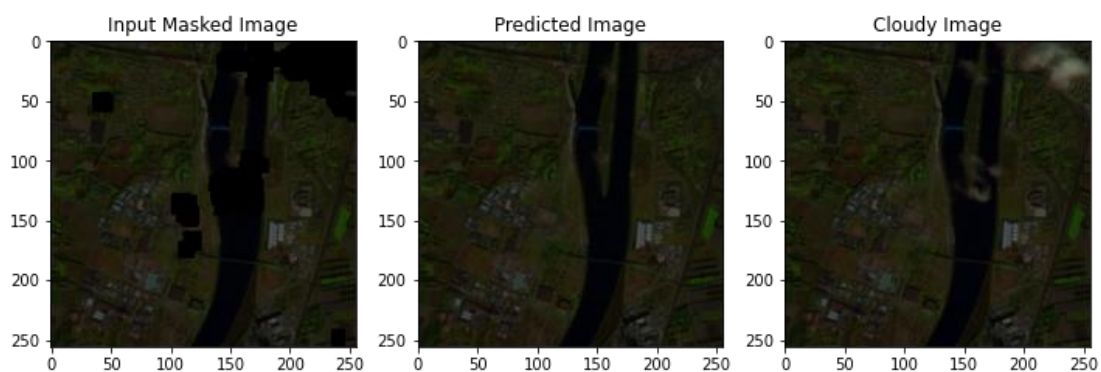
Επιτυχία Εκπαίδευσης:



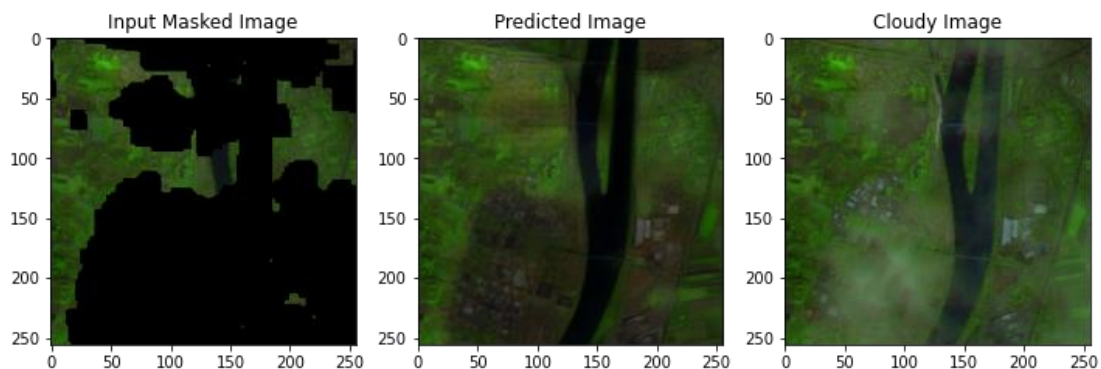
Εικόνα 5-28: 1^ο παράδειγμα επιτυχημένης εφαρμογής μοντέλου LGTSM σε εικόνες με σύννεφα (αριστερά: εικόνα εισόδου, κέντρο: παραγόμενη εικόνα, δεξιά: εικόνα με σύννεφα).



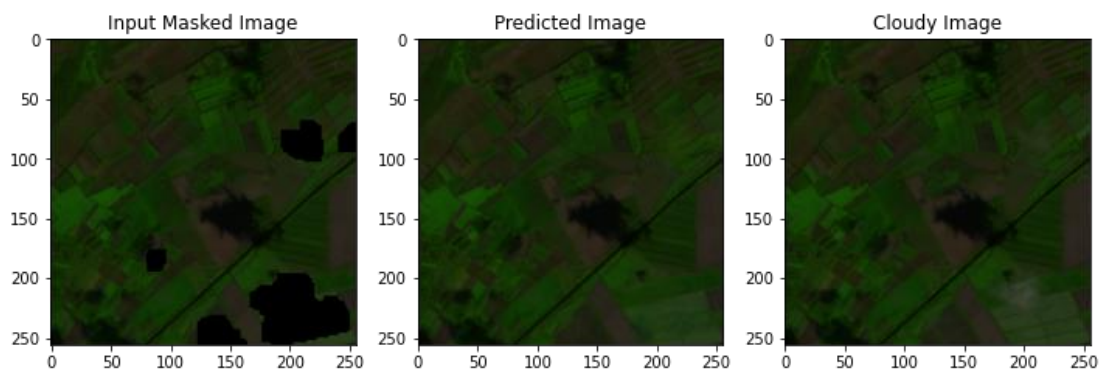
Εικόνα 5-29: 2^ο παράδειγμα επιτυχημένης εφαρμογής μοντέλου LGTSM σε εικόνες με σύννεφα (αριστερά: εικόνα εισόδου, κέντρο: παραγόμενη εικόνα, δεξιά: εικόνα με σύννεφα).



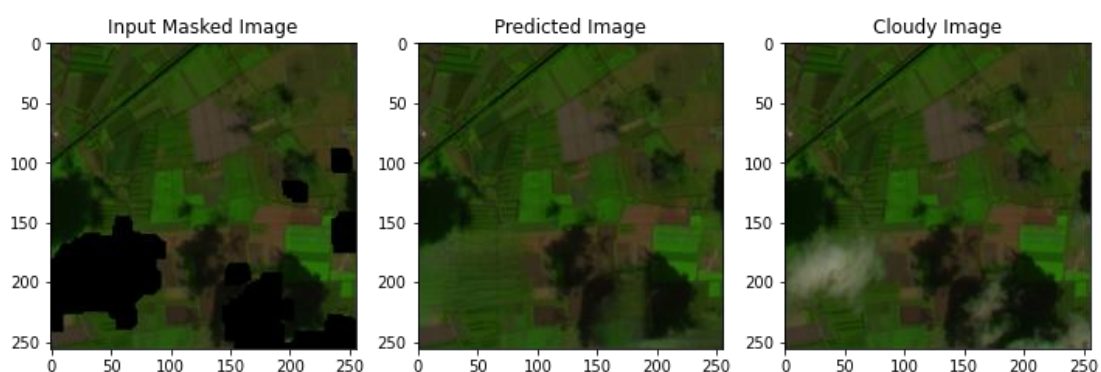
Εικόνα 5-30: 3^ο παράδειγμα επιτυχημένης εφαρμογής μοντέλου LGTSM σε εικόνες με σύννεφα (αριστερά: εικόνα εισόδου, κέντρο: παραγόμενη εικόνα, δεξιά: εικόνα με σύννεφα).



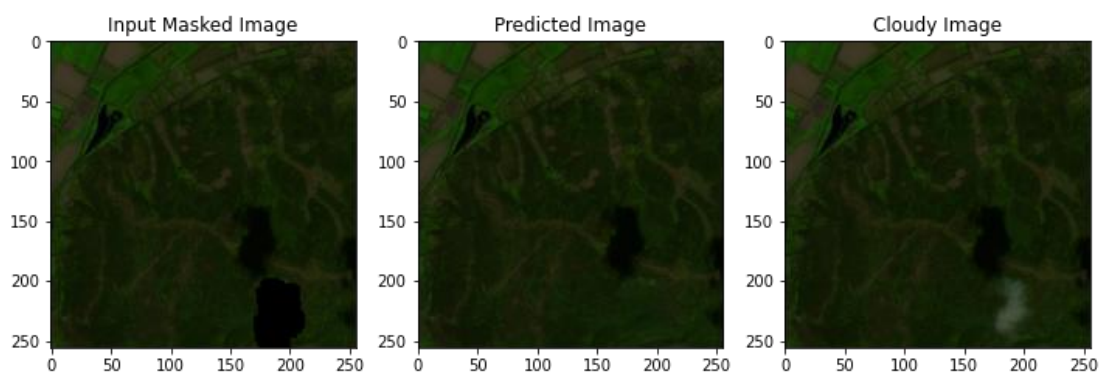
Εικόνα 5-31: 4^ο παράδειγμα επιτυχημένης εφαρμογής μοντέλου LGTSM σε εικόνες με σύννεφα (αριστερά: εικόνα εισόδου, κέντρο: παραγόμενη εικόνα, δεξιά: εικόνα με σύννεφα).



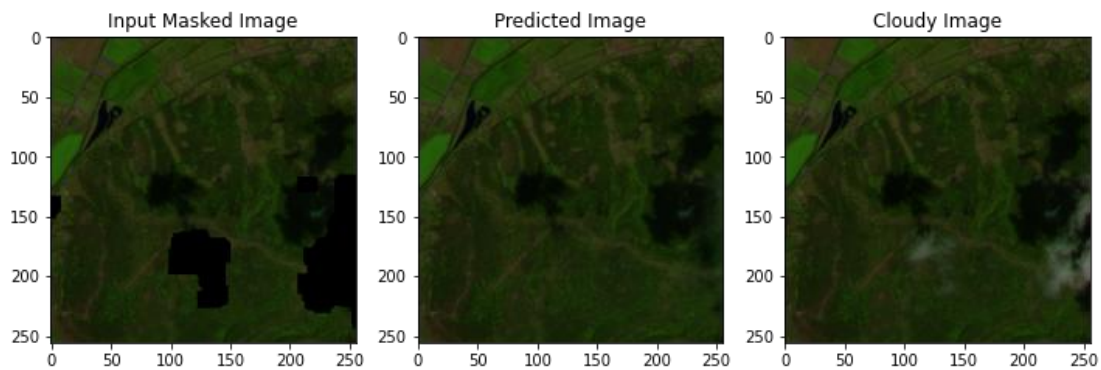
Εικόνα 5-32: 5^ο παράδειγμα επιτυχημένης εφαρμογής μοντέλου LGTSM σε εικόνες με σύννεφα (αριστερά: εικόνα εισόδου, κέντρο: παραγόμενη εικόνα, δεξιά: εικόνα με σύννεφα).



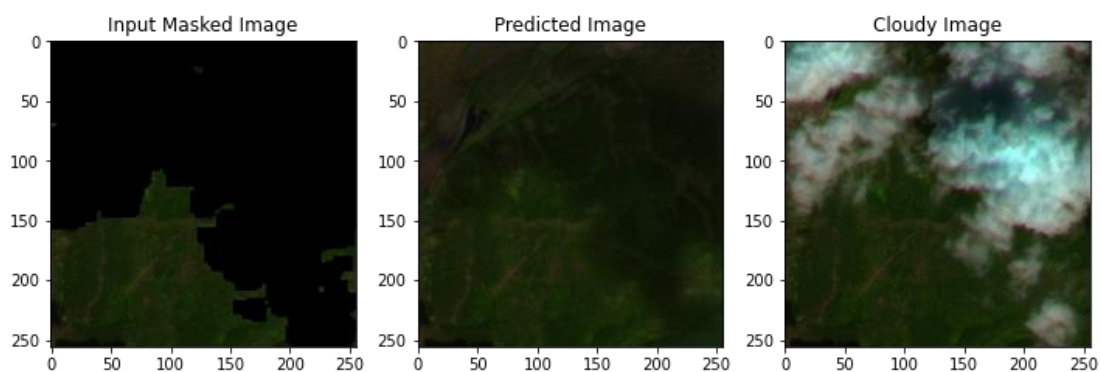
Εικόνα 5-33: 6^ο παράδειγμα επιτυχημένης εφαρμογής μοντέλου LGTSM σε εικόνες με σύννεφα (αριστερά: εικόνα εισόδου, κέντρο: παραγόμενη εικόνα, δεξιά: εικόνα με σύννεφα).



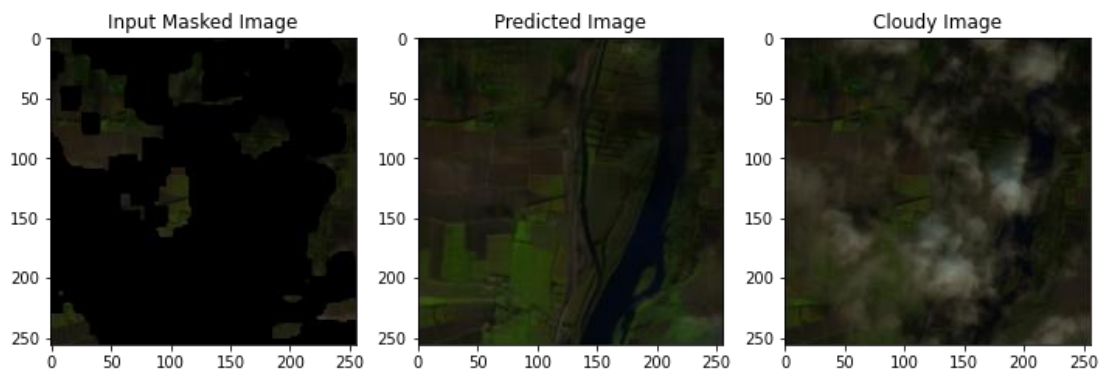
Εικόνα 5-34: 7^ο παράδειγμα επιτυχημένης εφαρμογής μοντέλου LGTSM σε εικόνες με σύννεφα (αριστερά: εικόνα εισόδου, κέντρο: παραγόμενη εικόνα, δεξιά: εικόνα με σύννεφα).



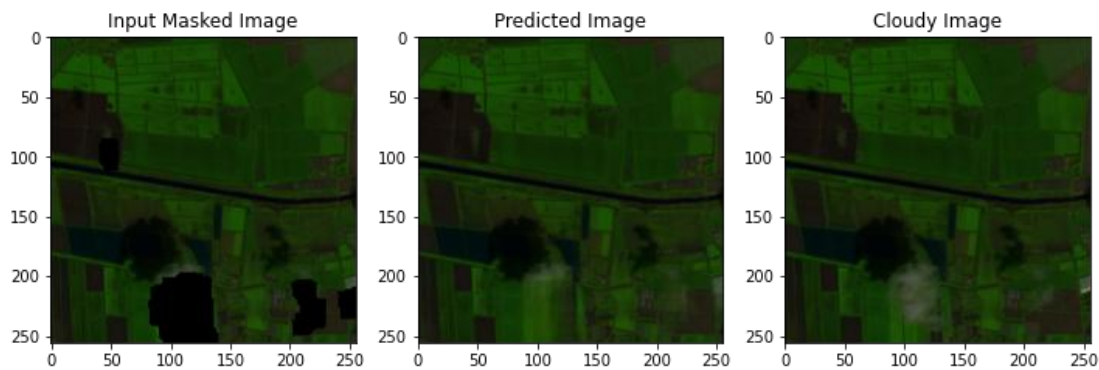
Εικόνα 5-35: 8^ο παράδειγμα επιτυχημένης εφαρμογής μοντέλου LGTSM σε εικόνες με σύννεφα (αριστερά: εικόνα εισόδου, κέντρο: παραγόμενη εικόνα, δεξιά: εικόνα με σύννεφα).



Εικόνα 5-36: 9^ο παράδειγμα επιτυχημένης εφαρμογής μοντέλου LGTSM σε εικόνες με σύννεφα (αριστερά: εικόνα εισόδου, κέντρο: παραγόμενη εικόνα, δεξιά: εικόνα με σύννεφα).



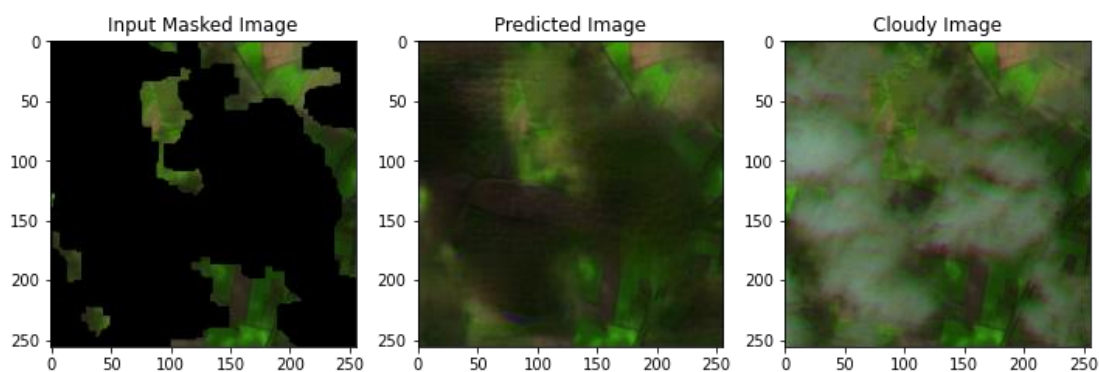
Εικόνα 5-37: 10^ο παράδειγμα επιτυχημένης εφαρμογής μοντέλου LGTSM σε εικόνες με σύννεφα (αριστερά: εικόνα εισόδου, κέντρο: παραγόμενη εικόνα, δεξιά: εικόνα με σύννεφα).



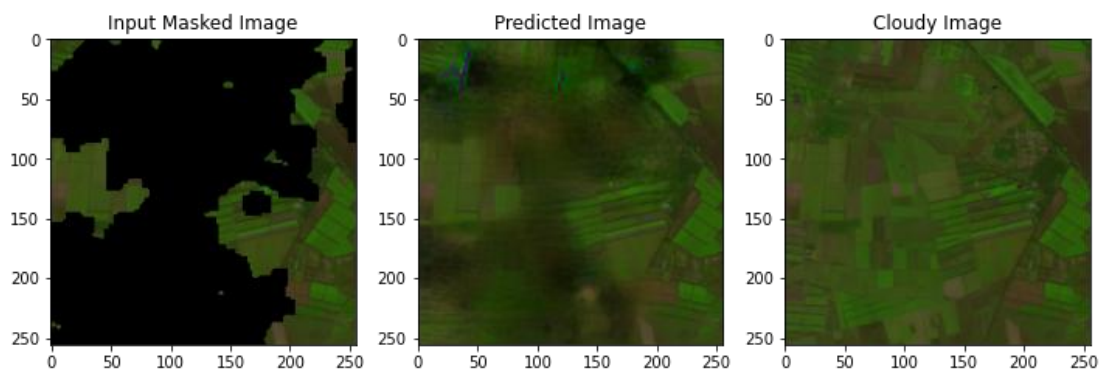
Εικόνα 5-38: 1^ο παράδειγμα επιτυχημένης εφαρμογής μοντέλου LGTSM σε εικόνες με σύννεφα (αριστερά: εικόνα εισόδου, κέντρο: παραγόμενη εικόνα, δεξιά: εικόνα με σύννεφα).

Παρατηρούμε ότι το μοντέλο ανταποκρίνεται πολύ καλά σε περιοχές με μικρή συννεφοκάλυψη αλλά και σε μερικές περιπτώσεις πιο μεγάλης συννεφοκάλυψης όπως για παράδειγμα στις εικόνες 5.36 και 5.37. Σημαντικό είναι να τονιστεί ότι το μοντέλο απ' ότι φαίνεται και στα αποτελέσματα ανακατασκευάζει μόνο την περιοχή με τη μάσκα, πράγμα που σημαίνει ότι αν οι μάσκες δεν έχουν σχεδιαστεί σωστά και δεν καλύπτουν όλο το σύννεφο, το μοντέλο δεν θα το ανακατασκευάσει όπως για παράδειγμα στο αποτέλεσμα της εικόνας 5.38. Επίσης παρατηρήθηκε ότι αν η εικόνα ανήκε σε χρονοσειρά όπου όλες οι εικόνες που την αποτελούσαν περιείχαν μεγάλη συννεφοκάλυψη τότε το αποτέλεσμα δεν ήταν καλό όπως στα παραδείγματα που ακολουθούν.

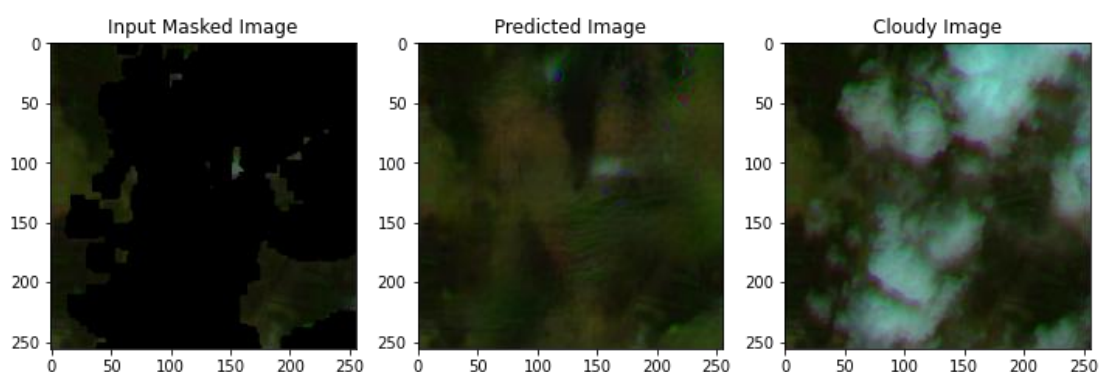
Αποτυχία Εκπαίδευσης:



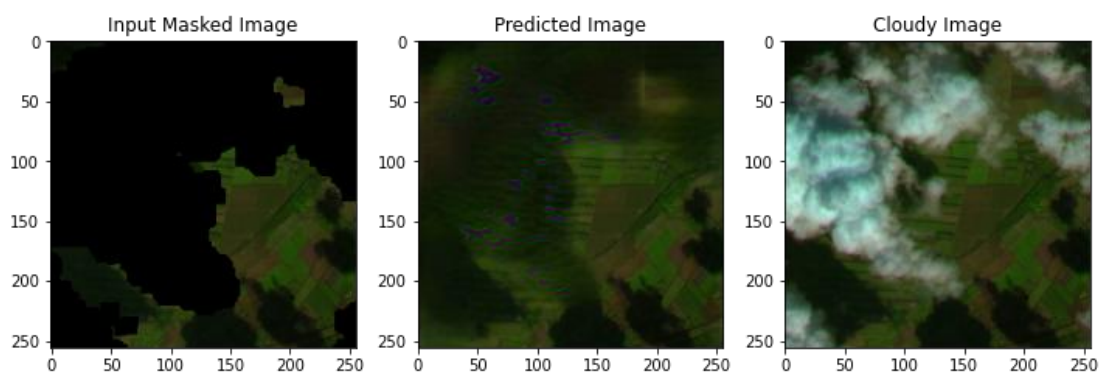
Εικόνα 5-39: 1^ο παράδειγμα μη επιτυχημένης εφαρμογής μοντέλου LGTSM σε εικόνες με σύννεφα (αριστερά: εικόνα εισόδου, κέντρο: παραγόμενη εικόνα, δεξιά: εικόνα με σύννεφα).



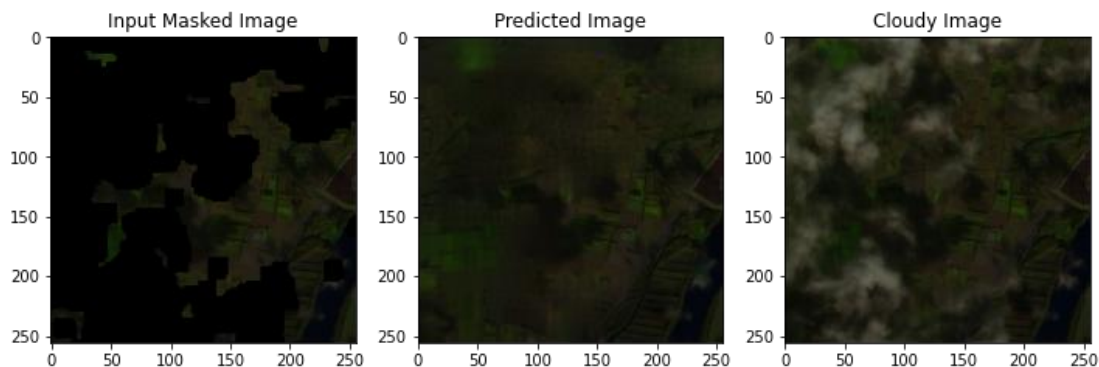
Εικόνα 5-40: 2^ο παράδειγμα μη επιτυχημένης εφαρμογής μοντέλου LGTSM σε εικόνες με σύννεφα (αριστερά: εικόνα εισόδου, κέντρο: παραγόμενη εικόνα, δεξιά: εικόνα με σύννεφα)



Εικόνα 5-41: 3^ο παράδειγμα μη επιτυχημένης εφαρμογής μοντέλου LGTSM σε εικόνες με σύννεφα (αριστερά: εικόνα εισόδου, κέντρο: παραγόμενη εικόνα, δεξιά: εικόνα με σύννεφα)

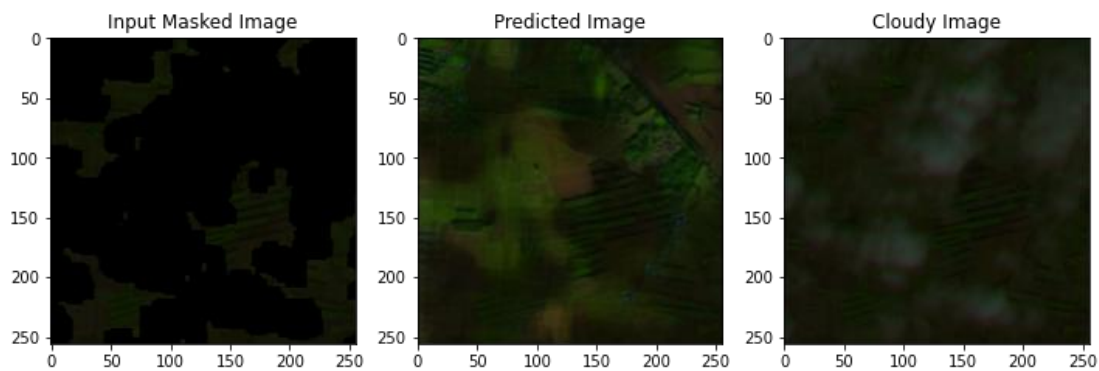


Εικόνα 5-42: 4^ο παράδειγμα μη επιτυχημένης εφαρμογής μοντέλου LGTSM σε εικόνες με σύννεφα (αριστερά: εικόνα εισόδου, κέντρο: παραγόμενη εικόνα, δεξιά: εικόνα με σύννεφα)

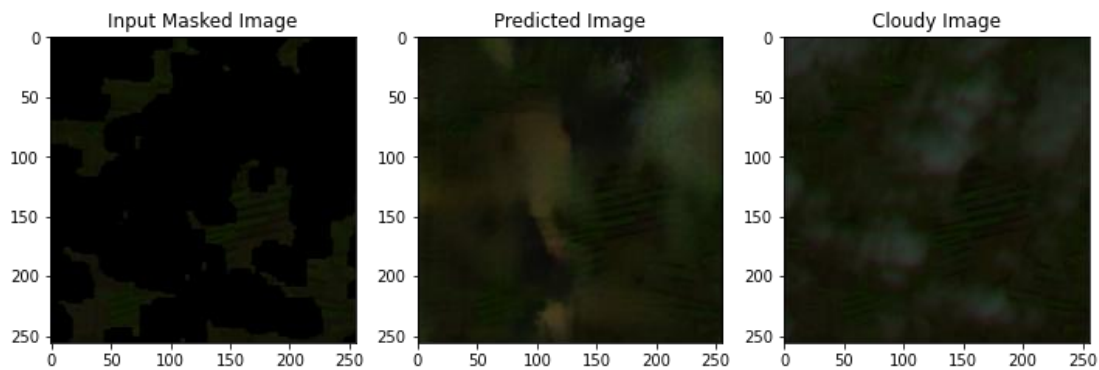


Εικόνα 5-43: 5^ο παράδειγμα μη επιτυχημένης εφαρμογής μοντέλου LGTSM σε εικόνες με σύννεφα (αριστερά: εικόνα εισόδου, κέντρο: παραγόμενη εικόνα, δεξιά: εικόνα με σύννεφα)

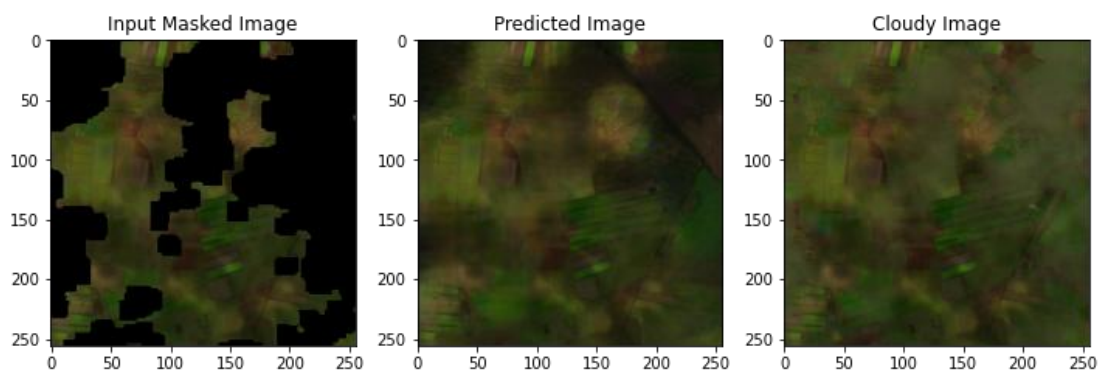
Από τις μη επιτυχημένες εφαρμογές που παρουσιάστηκαν παρατηρούμε ότι το μοντέλο έχει αφαιρέσει τα σύννεφα αλλά δεν έχει ανακατασκευάσει σωστά τις εικόνες. Οι εικόνες αυτές ανήκουν σε χρονοσειρές που οι μάσκες καλύπτουν μεγάλο μέρος της εικόνας, επομένως θα έπρεπε να είχε γίνει μία πιο ισότιμη κατανομή των εικόνων των χρονοσειρών με βάση και το ποσοστό συννεφοκάλυψης. Για το λόγο αυτό σε μία ήδη υπάρχουσα χρονοσειρά με 10 εικόνες προσθέσαμε άλλες 10 εικόνες με μικρότερο ποσοστό συννεφοκάλυψης και παρατηρήσαμε ότι όντως τα αποτελέσματα ήταν καλύτερα. Παρακάτω παρουσιάζονται δύο ενδεικτικά παραδείγματα.



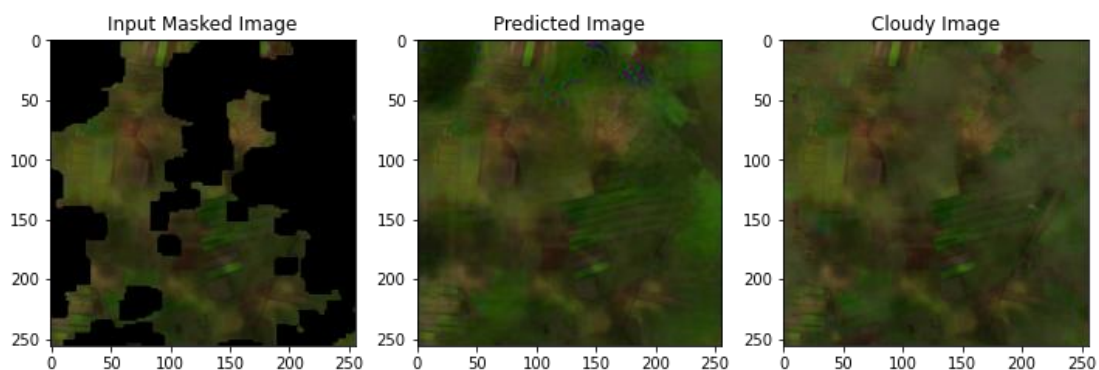
Εικόνα 5-44: Ενδεικτικό παράδειγμα αποτελέσματος εικόνας που ανήκει σε χρονοσειρά (20 εικόνες) με μεγαλύτερη ποικιλία σε συννεφοκάλυψη (παράδειγμα 1 - α).



Εικόνα 5-45: Ενδεικτικό παράδειγμα αποτελέσματος εικόνας που ανήκει στην αρχική χρονοσειρά με τις 10 εικόνες, όπου όλες είχαν μεγάλη συννεφοκάλυψη (παράδειγμα 1 – β).



Εικόνα 5-46: Ενδεικτικό παράδειγμα αποτελέσματος εικόνα που ανήκει σε χρονοσειρά (20 εικόνες) με μεγαλύτερη ποικιλία σε συννεφοκάλυψη (παράδειγμα 2 – α).



Εικόνα 5-47: Ενδεικτικό παράδειγμα αποτελέσματος εικόνας που ανήκει στην αρχική χρονοσειρά με τις 10 εικόνες, όπου όλες είχαν μεγάλη συννεφοκάλυψη (παράδειγμα 2 – β).

Παρατηρούμε ότι και πάλι το μοντέλο δεν έχει ανακατασκευάσει πλήρως την εικόνα αλλά έχει μπορέσει να αποδώσει καλύτερα κάποια χαρακτηριστικά στην περίπτωση όπου η χρονοσειρά αποτελείται από περισσότερες εικόνες που έχουν ποικιλία σε συννεφοκάλυψη.

6. ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΠΡΟΟΠΤΙΚΕΣ

6.1 Συμπεράσματα

Ο στόχος της συγκεκριμένης διπλωματικής εργασίας ήταν η μελέτη και αξιολόγηση γεννητικών ανταγωνιστικών δικτύων που υπάρχουν ήδη στη βιβλιογραφία, ως προς το αν μπορούν να εφαρμοστούν στο πρόβλημα αφαίρεσης σύννεφων από τις δορυφορικές εικόνες Sentinel-2. Για το σκοπό αυτό επιλέχθηκε να μελετηθούν τα μοντέλα Pix2Pix, CycleGAN και Learnable Gated Temporal Shift Module (LGTSM). Τα δύο πρώτα μοντέλα χρησιμοποιούνται για τη μετάφραση μίας εικόνας από ένα πεδίο πηγής σε ένα πεδίο στόχου. Το Pix2Pix δέχεται ως είσοδο ζεύγη εικόνων (εικόνα με σύννεφα – ground truth), ενώ το CycleGAN εκπαιδεύεται στα «τυφλά» χωρίς να γνωρίζει το ground truth αλλά γνωρίζει την κατανομή των εικόνων που ανήκουν στο πεδίο στόχου. Το LGTSM από την άλλη εισάγει την έννοια του χρόνου ανακατασκευάζοντας βίντεο, τα καρέ του οποίου έχουν κάποιο είδος μάσκας. Η λογική αυτή φάνηκε εξ' αρχής ιδιαίτερα ενδιαφέρουσα, λόγω του ότι τα δεδομένα μας αποτελούν μία χρονοσειρά. Αρχικά έγινε προσπάθεια να εκπαιδευτούν και τα τρία μοντέλα από το μηδέν. Κάτι τέτοιο δεν ήταν δυνατό λόγω του ότι το Google Colaboratory στο οποίο εκτελέστηκαν τα πειράματα δεν μας επέτρεπε να συνεχίσουμε την εκπαίδευση μετά από κάποιο σημείο λόγω κατανάλωσης της διαθέσιμης RAM. Τελικώς, για να μπορέσουμε να αντιμετωπίσουμε αυτό το πρόβλημα χρησιμοποιήθηκε ένας μικρός αριθμός δεδομένων για την εκπαίδευση. Το Pix2Pix εκπαιδεύτηκε για 300 εποχές, το CycleGAN εκπαιδεύτηκε για 200 εποχές και για το τρίτο μοντέλο χρησιμοποιήθηκε ένα προ-εκπαιδευμένο μοντέλο που έδιναν οι συγγραφείς ώστε να γίνει fine-tuning στο μοντέλο για λίγες εποχές ακόμα. Επομένως, συμπεραίνουμε ότι και τα τρία μοντέλα για να εκπαιδευτούν χρειάζονται αρκετή μνήμη καθώς κι έναν μεγάλο όγκο δεδομένων.

Όσον αφορά τα αποτελέσματα που λάβαμε από κάθε μοντέλο διαπιστώσαμε ότι το Pix2Pix με τη βοήθεια του ground truth μπόρεσε για εικόνες με μικρή συννεφοκάλυψη να δώσει καλές ποιοτικά ανακατασκευασμένες εικόνες, ενώ για εικόνες με μεγάλη συννεφοκάλυψη φάνηκε ότι ήθελε να εκπαιδευτεί κι άλλο με περισσότερα δεδομένα. Παρ' όλα αυτά και σε αυτές τις εικόνες εντόπιζε τα σύννεφα και τα αφαιρούσε απλώς η υπόλοιπη εικόνα δεν ήταν καλή ποιοτικά. Σημαντική διαφορά με τα υπόλοιπα μοντέλα ήταν ότι το Pix2Pix εντόπιζε και αφαιρούσε και τις σκιές εκτός από τα σύννεφα. Για το CycleGAN παρατηρήθηκε από τα αποτελέσματα ότι παρ' ότι εκπαιδεύεται χωρίς να έχει καμία εικόνα αναφοράς για εικόνες με χαμηλή συννεφοκάλυψη μπόρεσε να βρει τα σύννεφα και να τα αφαιρέσει από την εικόνα. Το

συγκεκριμένο μοντέλο αν εκπαιδευτεί με περισσότερα δεδομένα και σε ένα άλλο σύστημα περιμένουμε ότι θα μπορέσει να δώσει καλά αποτελέσματα. Το LGTSM για την εκπαίδευση του χρησιμοποίησε εικόνες χωρίς σύννεφα στις οποίες είχαν περαστεί μάσκες, κι επομένως το μοντέλο προσπαθούσε να αφαιρέσει τις μάσκες έχοντας ως εικόνα αναφοράς την εικόνα χωρίς τα σύννεφα. Για το συγκεκριμένο μοντέλο έγιναν κάποια πειράματα στο στάδιο του fine-tuning ως προς τα βάρη των απωλειών. Αυτό που διαπιστώθηκε είναι ότι παρά την διαφορετική φύση και κατανομή των δεδομένων εισόδου τα καλύτερα αποτελέσματα τα λάβαμε με τις τιμές που είχαν επιλέξει οι συγγραφείς να χρησιμοποιήσουν κατά το στάδιο της προ-εκπαίδευσης. Τα αποτελέσματα που λάβαμε κατά το στάδιο αυτό ήταν ικανοποιητικά ως προς ένα βαθμό αλλά παρατηρήσαμε ότι σε αρκετές δημιουργημένες εικόνες υπήρχαν κάποιες αλλοιώσεις. Σίγουρα αν είχαμε τη δυνατότητα να εκπαιδεύσουμε το συγκεκριμένο μοντέλο σε ένα άλλο σύστημα και με πολύ περισσότερα δεδομένα θα μπορούσαμε να βγάλουμε πιο ασφαλή συμπεράσματα. Παρ' όλα αυτά οι τιμές των μετρικών που υπολογίστηκαν ως προς την ποιότητα των παραγόμενων εικόνων κατά το στάδιο εκπαίδευσης έδειξαν ότι το συγκεκριμένο μοντέλο υπερτερεί έναντι του pix2pix και του cycleGAN. Το συγκεκριμένο μοντέλο εφαρμόστηκε και σε πραγματικές εικόνες με σύννεφα των οποίων τη μάσκα την είχαμε από το αρχικό σύνολο δεδομένων για να δούμε αν μπορεί να ανακατασκευάσει την εικόνα. Αυτό που παρατηρήσαμε είναι ότι το αποτέλεσμα εξαρτιόταν από τη μάσκα και αν εκείνη είχε παραχθεί σωστά. Για παράδειγμα σε εικόνα που δεν είχαν όλα τα σύννεφα καλυφθεί με μάσκα, το μοντέλο αφαίρεσε τα σύννεφα που ήταν κάτω από τη μάσκα αλλά άφησε τα υπόλοιπα. Επίσης, το αποτέλεσμα εξαρτιόταν κατά πολύ και από τις υπόλοιπες εικόνες της χρονοσειράς. Αν για παράδειγμα όλες οι εικόνες της χρονοσειράς είχαν μεγάλο ποσοστό συννεφοκάλυψης τότε δεν μπορούσαν να ανακατασκευαστούν σωστά οι εικόνες. Αυτό είναι λογικό εφόσον το συγκεκριμένο μοντέλο όπως αναφέρθηκε και στα προηγούμενα κεφάλαια χρησιμοποιεί μία χρονικά μετατοπισμένη μονάδα (TSM), η οποία μετατοπίζει ένα τμήμα καναλιών στο προηγούμενο και στο επόμενο του καρέ πριν από την εφαρμογή των συνελίξεων με στόχο μαζί με τα μετατοπισμένα και μη χαρακτηριστικά να «μάθει» το μοντέλο τις χρονικές δομές.

Από τα αποτελέσματα μπορούμε να συμπεράνουμε ότι σίγουρα και τα τρία μοντέλα αξίζει να μελετηθούν περισσότερο για το συγκεκριμένο πρόβλημα καθώς κρίνεται ότι μπορούν με σωστή εκπαίδευση να δώσουν σημαντικά αποτελέσματά. Παρ' όλα αυτά το LGTSM φαίνεται να υπερτερεί και να αντιμετωπίζει το πρόβλημα με πιο σωστό τρόπο καθώς λαμβάνει υπόψη και τις χρονικές δομές.

6.2 Προοπτικές

Λαμβάνοντας υπόψη όσα έχουν περιγραφεί στις προηγούμενες παραγράφους και με δεδομένα τα αποτελέσματα της συγκεκριμένης διπλωματικής εργασίας, αναφέρονται παρακάτω ορισμένες προτάσεις για μελλοντική έρευνα :

- Εκπαίδευση και των τριών μοντέλων σε σύστημα με επαρκή μνήμη και χρησιμοποιώντας έναν αρκετά μεγάλο όγκο δεδομένων.
- Δημιουργία καλύτερου συνόλου δεδομένων. Δημιουργία масκών με μεγαλύτερη ακρίβεια καθώς κατά τη διάρκεια των πειραμάτων παρατηρήθηκε ότι οι μάσκες που μας παρείχαν δεν ήταν ακριβείας ή περιλάμβαναν και άλλα στοιχεία της εικόνας εκτός από σύννεφα.
- Εκπαίδευση του μοντέλου LGTSM από την αρχή, κάνοντας πειράματα για το μέγεθος της TSM μονάδας, δηλαδή το πόσα καρέ πίσω και πόσα μπροστά θα κοιτάζει το μοντέλο για να εξάγει τις χρονικές δομές.
- Επέκταση του LGTSM λαμβάνοντας υπόψη την πραγματική χρονική διαφορά μεταξύ των εικόνων (λόγω μη ομοιόμορφης δειγματοληψίας στο χρόνο).

7. ΒΙΒΛΙΟΓΡΑΦΙΑ

Δημοσιεύσεις:

- [1] P. Isola, J.-Y. Zhu, T. Zhou and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," CVPR, Berkeley AI Research (BAIR) Laboratory, UC Berkeley, 2017.
- [2] Y.-L. Chang, Z. Y. Liu, K.-Y. Lee and W. Hsu, "Learnable Gated Temporal Shift Module for Deep Video Inpainting," British Machine Vision Conference (BMVC), 2019.
- [3] Y.-L. Chang, Z. Y. Liu, K.-Y. Lee and W. Hsu, "Free-form Video Inpainting with 3D Gated Convolution and Temporal," International Conference on Computer Vision, 2019.
- [4] P. Singh and N. Komodakis, "Cloud-GAN: Cloud Removal for Sentinel-2 Imagery Using a Cyclic Consistent Generative Adversarial Network," IGARSS, Valencia, Spain., 2018.
- [5] F. Weiliang, J. Weimin and C. Jing, "A pixel missing patch inpainting method for remote sensing image," International Conference on Geoinformatics, Shanghai, 2011.
- [6] J.-Y. Zhu, T. Park, P. Isola and A. A. Efros, "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks," ICCV, Berkeley AI Research (BAIR) laboratory, UC Berkeley, 2017.
- [7] J. Yoon, J. Jordon and M. v. d. Schaar, "GAIN: Missing Data Imputation using Generative Adversarial Nets," International Conference on Machine, Stockholm Sweden, 2018.
- [8] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio, "Generative Adversarial Networks," *Proceedings of the International Conference on Neural Information Processing Systems - NIPS*, p. 2672–2680, 2014.
- [9] Α. Χρονοπούλου, «Τεχνικές Μεταφοράς Μάθησης σε Βαθιά Νευρωνικά Δίκτυα για Ανάλυση Συναισθήματος και Σημασιολογική Μοντελοποίηση,» ΕΜΠ, Αθήνα, 2019.
- [10] Χ. Κοντόπουλος, «Τεχνικές βαθιάς μηχανικής μάθησης και Συνελκτικά Νευρωνικά Δίκτυα για την ταξινόμηση υπερφασματικών δεδομένων,» ΕΜΠ, Αθήνα, 2015.
- [11] Φ. Π. Κονιδάρης, «Χρήση Γεννητικών Ανταγωνιστικών Δικτύων ως Τεχνική Εμπλουτισμού Δεδομένων σε Ιατρικές Εφαρμογές,» ΕΜΠ, Αθήνα, 2018.

Ηλεκτρονικές Πηγές:

<https://machinelearningmastery.com/>
<https://medium.com/>
<https://kharshit.github.io/blog/2019/02/15/autoencoder-downsampling-and-upsampling>
<https://towardsdatascience.com/>
<http://deeplearning.net/tutorial/unet.html>
<https://affinelayer.com/pix2pix/>
<https://paperswithcode.com/method/pix2pix>
<https://paperswithcode.com/method/u-net>
<https://www.researchgate.net/>
<https://developers.google.com/machine-learning/gan>
<https://www.tensorflow.org/tutorials/generative/pix2pix>
<https://cs231n.github.io/neural-networks-1/>

ΕΥΡΕΤΗΡΙΟ ΕΙΚΟΝΩΝ

Εικόνα 2-1: Σχηματικό διάγραμμα της μεθόδου που παρουσιάζεται στο « <i>A pixel missing patch inpainting method for remote sensing image</i> ».....	13
Εικόνα 2-2: Λογική του μοντέλου Pix2Pix που παρουσιάζεται στο « <i>Image-to-Image Translation with Conditional Adversarial Networks</i> ».....	14
Εικόνα 2-3: α) Οι συναρτήσεις «μετάφρασης» G, F και οι διευκρινιστές D_X , D_Y του μοντέλου CycleGAN. β) Η λογική της συνέπειας του κύκλου στο μοντέλο CycleGAN που παρουσιάζεται στο « <i>Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks</i> ».....	15
Εικόνα 2-4: Αρχιτεκτονική του μοντέλου GAIN που παρουσιάζεται στο « <i>GAIN: Missing Data Imputation using Generative Adversarial Nets</i> ».....	16
Εικόνα 2-5: Επεξήγηση των μετατοπισμένων πυρήνων στο προτεινόμενο μοντέλο LGTSM που παρουσιάζεται στο « <i>Learnable Gated Temporal Shift Module for Deep Video Inpainting</i> ». α) Χαρακτηριστικά εισόδου. Η μετατόπιση γίνεται στις διαστάσεις channel * time. β) Το αρχικό TSM. γ) Ισοδύναμο TSM με χρονικά μετατοπισμένους πυρήνες. δ) Οι πυρήνες που χρησιμοποιούνται στο LGTSM, μπορούν να «μάθουν» κι ας έχουν διαφορετικό μέγεθος.....	17
Εικόνα 3-1: Διαφορά Μηχανικής Μάθησης (ML) και Βαθιάς Μάθησης (DL).....	20
Εικόνα 3-2: Αναπαράσταση/Δομή τεχνητού νευρώνα. Πηγή: www.researchgate.net	21
Εικόνα 3-3: Απεικόνιση της κατάβασης κλίσης σε μια σειρά συνόλων επιπέδων.....	22
Εικόνα 3-4: Αλγόριθμος Οπισθοδρομικής διάδοσης σφάλματος (Backpropagation).....	24
Εικόνα 3-5: Δομή ενός νευρωνικού δικτύου (αριστερά) με πλήρη διασύνδεση και δομή ενός συνελκτικού (δεξιά). Πηγή: https://www.topcoder.com/	25
Εικόνα 3-6: Διαδικασία συνέλιξης. Πηγή: https://www.topcoder.com/	26
Εικόνα 3-7: Διαδικασία Max Pooling, με χρήση 2*2 φίλτρου.....	30
Εικόνα 3-8: Αρχιτεκτονική συνελκτικού νευρωνικού δικτύου. Πηγή: https://vinodsblog.com	31
Εικόνα 3-9: Αρχή της Μεταφοράς Μάθησης. Πηγή: https://pennylane.ai/	31
Εικόνα 3-10: Τυπική αρχιτεκτονική Γεννητικών Ανταγωνιστικών Δικτύων.....	32
Εικόνα 3-11: Διαδικασία οπισθοδρομικής διάδοσης σφάλματος κατά την εκπαίδευση του διευκρινιστή. Πηγή: https://developers.google.com/	33
Εικόνα 3-12: Υπολογισμός σιγμοειδούς εγκάρσιας εντροπίας για δυαδικό πρόβλημα ταξινόμησης. Πηγή: https://gombbru.github.io/	36
Εικόνα 3-13: Πιθανές παραμορφώσεις εικόνας. Πηγή: https://www.researchgate.net/	41
Εικόνα 3-14: Αποτέλεσμα τεχνικής συμπλήρωσης εικόνας. Πηγή: Survey on Different Techniques for Image Inpainting - IRJET.....	42
Εικόνα 3-15: Τα φασματικά κανάλια του sentinel-2. Πηγή: https://www.satimagingcorp.com/	43
Εικόνα 3-16: Ενδεικτικά παραδείγματα τμήματος εικόνας χωρίς σύννεφα (αριστερά) και τμήματος εικόνας με ιστούς νεφών (δεξιά).....	44
Εικόνα 3-17: Ενδεικτικά παραδείγματα τμήματος εικόνας με λίγα σύννεφα – σκιές (αριστερά) και τμήματος εικόνας με μεγάλο ποσοστό συννεφοκάλυψης (δεξιά).....	44
Εικόνα 4-1: Ενδεικτικά παραδείγματα εικόνας RGB με σύννεφα (αριστερά), εικόνας χωρίς σύννεφα με χρήση φίλτρου διάμεσης τιμής (κέντρο), μάσκας σύννεφων (δεξιά) – ημερομηνία: 02-07-2018.....	48
Εικόνα 4-2: Ενδεικτικό τμήμα δορυφορικής εικόνας με 3 κανάλια (RGB). Ανεπεξέργαστο με σύννεφα (αριστερά), επεξεργασμένο με το φίλτρο διάμεσης τιμής (δεξιά).	49
Εικόνα 4-3: Ενδεικτικό παράδειγμα αρχικής μάσκας (αριστερά) και μάσκας μετά την επεξεργασία (δεξιά).	50
Εικόνα 4-4: Ενδεικτικό παράδειγμα ζεύγους εισόδου, ground truth (αριστερά), εικόνα με σύννεφα (δεξιά).....	52

Εικόνα 4-5: Ενδεικτικό παράδειγμα εικόνας RGB χωρίς σύννεφα (αριστερά) και τυχαίας μάσκας (δεξιά).	61
Εικόνα 4-6: Ενδεικτικό παράδειγμα αποτελέσματος συνένωσης των εικόνων 4.5	61
Εικόνα 5-1: 1 ^ο παράδειγμα επιτυχημένου αποτελέσματος μοντέλου rix2rix	64
Εικόνα 5-2: 2 ^ο παράδειγμα επιτυχημένου αποτελέσματος μοντέλου rix2rix	65
Εικόνα 5-3: 3 ^ο παράδειγμα επιτυχημένου αποτελέσματος μοντέλου rix2rix	65
Εικόνα 5-4: 4 ^ο παράδειγμα επιτυχημένου αποτελέσματος μοντέλου rix2rix	65
Εικόνα 5-5: 5 ^ο παράδειγμα επιτυχημένου αποτελέσματος μοντέλου rix2rix	65
Εικόνα 5-6: 1 ^ο παράδειγμα μη επιτυχημένου αποτελέσματος μοντέλου rix2rix	66
Εικόνα 5-7: 2 ^ο παράδειγμα μη επιτυχημένου αποτελέσματος μοντέλου rix2rix	66
Εικόνα 5-8: 3 ^ο παράδειγμα μη επιτυχημένου αποτελέσματος μοντέλου rix2rix	66
Εικόνα 5-9: 4 ^ο παράδειγμα μη επιτυχημένου αποτελέσματος μοντέλου rix2rix	66
Εικόνα 5-10: 5 ^ο παράδειγμα μη επιτυχημένου αποτελέσματος μοντέλου rix2rix	67
Εικόνα 5-11: 1 ^ο ενδεικτικό αποτέλεσμα μοντέλου CycleGAN	68
Εικόνα 5-12: 2 ^ο ενδεικτικό αποτέλεσμα μοντέλου CycleGAN	69
Εικόνα 5-13: 3 ^ο ενδεικτικό αποτέλεσμα μοντέλου CycleGAN	69
Εικόνα 5-14: 4 ^ο ενδεικτικό αποτέλεσμα μοντέλου CycleGAN	69
Εικόνα 5-15: Διαγράμματα μετρικών για το 1 ^ο πείραμα (MSE: πάνω αριστερά, PSNR: πάνω δεξιά, FID: κάτω αριστερά και SSIM: κάτω δεξιά)	72
Εικόνα 5-16: 1 ^ο ενδεικτικό αποτέλεσμα 1 ^{ου} πειράματος μοντέλου LGTSM (ground truth: πάνω αριστερά, μάσκα: πάνω δεξιά, είσοδος μοντέλου: κάτω αριστερά και ανακατασκευασμένη εικόνα: κάτω δεξιά)	73
Εικόνα 5-17: 2 ^ο ενδεικτικό αποτέλεσμα 1 ^{ου} πειράματος μοντέλου LGTSM (ground truth: πάνω αριστερά, μάσκα: πάνω δεξιά, είσοδος μοντέλου: κάτω αριστερά και ανακατασκευασμένη εικόνα: κάτω δεξιά)	73
Εικόνα 5-18: Διαγράμματα μετρικών για το 2 ^ο πείραμα (MSE: πάνω αριστερά, PSNR: πάνω δεξιά, FID: κάτω αριστερά και SSIM: κάτω δεξιά)	74
Εικόνα 5-19: 1 ^ο ενδεικτικό αποτέλεσμα 2 ^{ου} πειράματος μοντέλου LGTSM για την εποχή 150 (ground truth: πάνω αριστερά, μάσκα: πάνω δεξιά, είσοδος μοντέλου: κάτω αριστερά και ανακατασκευασμένη εικόνα: κάτω δεξιά)	75
Εικόνα 5-20: 2 ^ο ενδεικτικό αποτέλεσμα 2 ^{ου} πειράματος μοντέλου LGTSM για την εποχή 150 (ground truth: πάνω αριστερά, μάσκα: πάνω δεξιά, είσοδος μοντέλου: κάτω αριστερά και ανακατασκευασμένη εικόνα: κάτω δεξιά)	75
Εικόνα 5-21: 2 ^ο ενδεικτικό αποτέλεσμα 2ου πειράματος μοντέλου LGTSM για την εποχή 140 (ground truth: πάνω αριστερά, μάσκα: πάνω δεξιά, είσοδος μοντέλου: κάτω αριστερά και ανακατασκευασμένη εικόνα: κάτω δεξιά)	76
Εικόνα 5-22: Διαγράμματα μετρικών για το 3 ^ο πείραμα (MSE: πάνω αριστερά, PSNR: πάνω δεξιά, FID: κάτω αριστερά και SSIM: κάτω δεξιά)	77
Εικόνα 5-23: Ενδεικτικό αποτέλεσμα 3 ^{ου} πειράματος μοντέλου LGTSM (ground truth: πάνω αριστερά, μάσκα: πάνω δεξιά, είσοδος μοντέλου: κάτω αριστερά και ανακατασκευασμένη εικόνα: κάτω δεξιά)	78
Εικόνα 5-24: Διαγράμματα μετρικών για το 4 ^ο πείραμα (MSE: πάνω αριστερά, PSNR: πάνω δεξιά, FID: κάτω αριστερά και SSIM: κάτω δεξιά)	78
Εικόνα 5-25: Ενδεικτικό αποτέλεσμα 4 ^{ου} πειράματος μοντέλου LGTSM (ground truth: πάνω αριστερά, μάσκα: πάνω δεξιά, είσοδος μοντέλου: κάτω αριστερά και ανακατασκευασμένη εικόνα: κάτω δεξιά)	79
Εικόνα 5-26: Διαγράμματα μετρικών για το 5 ^ο πείραμα (MSE: πάνω αριστερά, PSNR: πάνω δεξιά, FID: κάτω αριστερά και SSIM: κάτω δεξιά)	80
Εικόνα 5-27: Ενδεικτικό αποτέλεσμα 5 ^{ου} πειράματος μοντέλου LGTSM (ground truth: πάνω αριστερά, μάσκα: πάνω δεξιά, είσοδος μοντέλου: κάτω αριστερά και ανακατασκευασμένη εικόνα: κάτω δεξιά)	80
Εικόνα 5-28: 1 ^ο παράδειγμα επιτυχημένης εφαρμογής μοντέλου LGTSM σε εικόνες με σύννεφα (αριστερά: εικόνα εισόδου, κέντρο: παραγόμενη εικόνα, δεξιά: εικόνα με σύννεφα)	81

Εικόνα 5-47: Ενδεικτικό παράδειγμα αποτελέσματος εικόνας που ανήκει στην αρχική χρονοσειρά με τις 10 εικόνες, όπου όλες είχαν μεγάλη συννεφοκάλυψη (παράδειγμα 2 – β).....	88
---	----

ΕΥΡΕΤΗΡΙΟ ΠΙΝΑΚΩΝ

Πίνακας 4-1: Πίνακας που παρουσιάζει την αρχιτεκτονική για τον δημιουργό του μοντέλου pix2pix. (':batch normalization).....	54
Πίνακας 4-2: Πίνακας που παρουσιάζει την αρχιτεκτονική για τον διευκρινιστή του μοντέλου pix2pix.	54

ΕΥΡΕΤΗΡΙΟ ΣΧΗΜΑΤΩΝ

Σχήμα 3-1: Γραφική παράσταση σιγμοειδούς συνάρτησης.	27
Σχήμα 3-2: Γραφική παράσταση υπερβολικής εφαπτομένης.....	28
Σχήμα 3-3: Γραφική παράσταση ReLU.	28
Σχήμα 3-4: Γραφική παράσταση Leaky ReLU.	29
Σχήμα 3-5: Γραφική παράσταση συνάρτησης softmax.....	29
Σχήμα 3-6: Μοντέλο κωδικοποιητή (αριστερά), και αποκωδικοποιητή (δεξιά).	35
Σχήμα 3-7: Απώλεια Cycle-consistency. Πηγή: https://www.tensorflow.org/	37
Σχήμα 4-1: Αρχιτεκτονική Δημιουργού. Πηγή: https://towardsdatascience.com/cyclelegans-and-pix2pix-5e6a5f0159c4	53
Σχήμα 4-2: Αρχιτεκτονική Διευκρινιστή. Πηγή: https://towardsdatascience.com/cyclelegans-and-pix2pix-5e6a5f0159c4	53
Σχήμα 4-3: Εκπαίδευση του Δημιουργού.....	55
Σχήμα 4-4: Εκπαίδευση του Διευκρινιστή (δεξιά).	56
Σχήμα 4-5: Διαδικασία Cycle-consistent.	57
Σχήμα 4-6: Εκπαίδευση CycleGAN.....	58