



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ  
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

# Υπερανάλυση και Σημασιολογική Κατάτμηση Δορυφορικών Εικόνων για τον Εντοπισμό<sup>1</sup> Δασών

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

**Ελευθέριου Λυμπερόπουλου**



**Επιβλέπων:**

Στέφανος Κόλλιας  
Καθηγητής

**Συνεπιβλέπουσα:**

Παρασκευή Τζούθελη  
ΕΔΙΠ

Αθήνα, Σεπτέμβριος 2022





ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ  
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

**Υπερανάλυση και Σημασιολογική Κατάτμηση  
Δορυφορικών Εικόνων για τον Εντοπισμό Δασών**

**ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ**

του

**Ελευθέριου Λυμπερόπουλου**

**Επιβλέπων:**

Στέφανος Κόλλιας  
Καθηγητής

**Συνεπιβλέπουσα:**

Παρασκευή Τζούβελη  
ΕΔΙΠ

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 27η Σεπτεμβρίου 2022.

(Υπογραφή)

(Υπογραφή)

(Υπογραφή)

.....  
.....  
.....  
Στέφανος Κόλλιας      Αθανάσιος Βουλόδημος      Γεώργιος Στάμου  
Καθηγητής      Επίκουρος Καθηγητής      Καθηγητής

Αθήνα, Σεπτέμβριος 2022





Copyright © – All rights reserved. Με την επιφύλαξη παντός δικαιώματος.  
Ελευθέριος Λυμπερόπουλος, 2022.

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.

Το περιεχόμενο αυτής της εργασίας δεν απηχεί απαραίτητα τις απόψεις του Τμήματος, του Επιβλέποντα, ή της επιτροπής που την ενέκρινε.

#### **ΔΗΛΩΣΗ ΜΗ ΛΟΓΟΚΛΟΠΗΣ ΚΑΙ ΑΝΑΛΗΨΗΣ ΠΡΟΣΩΠΙΚΗΣ ΕΥΘΥΝΗΣ**

Με πλήρη επίγνωση των συνεπειών του νόμου περί πνευματικών δικαιωμάτων, δηλώνω ενυπογράφως ότι είμαι αποκλειστικός συγγραφέας της παρούσας Πτυχιακής Εργασίας, για την ολοκλήρωση της οποίας κάθε βοήθεια είναι πλήρως αναγνωρισμένη και αναφέρεται λεπτομερώς στην εργασία αυτή. Έχω αναφέρει πλήρως και με σαφείς αναφορές, όλες τις πηγές χρήσης δεδομένων, απόψεων, θέσεων και προτάσεων, ιδεών και λεκτικών αναφορών, είτε κατά κυριολεξία είτε βάσει επιστημονικής παράφρασης. Αναλαμβάνω την προσωπική και ατομική ευθύνη ότι σε περίπτωση αποτυχίας στην υλοποίηση των ανωτέρω δηλωθέντων στοιχείων, είμαι υπόλογος έναντι λογοκλοπής, γεγονός που σημαίνει αποτυχία στην Πτυχιακή μου Εργασία και κατά συνέπεια αποτυχία απόκτησης του Τίτλου Σπουδών, πέραν των λοιπών συνεπειών του νόμου περί πνευματικών δικαιωμάτων. Δηλώνω, συνεπώς, ότι αυτή η Πτυχιακή Εργασία προετοιμάστηκε και ολοκληρώθηκε από εμένα προσωπικά και αποκλειστικά και ότι, αναλαμβάνω πλήρως όλες τις συνέπειες του νόμου στην περίπτωση κατά την οποία αποδειχθεί, διαχρονικά, ότι η εργασία αυτή ή τμήμα της δεν μου ανήκει διότι είναι προϊόν λογοκλοπής άλλης πνευματικής ιδιοκτησίας.

(Υπογραφή)

.....  
Ελευθέριος  
Λυμπερόπουλος

27 Σεπτεμβρίου 2022



## Περίληψη

---

Με τις ολοένα και αυξανόμενες δορυφορικές αποστολές, η γεωπαρατήρηση και η τηλεπισκόπηση εμφανίζουν ταχεία πρόοδο τα τελευταία χρόνια και χρησιμοποιούνται σε εφαρμογές όπως η παρακολούθηση του οικοσυστήματος και η πρόληψη φυσικών καταστροφών. Παράλληλα, με την άνοδο της μηχανικής μάθησης στην υπολογιστική όραση και την ανάγκη για αποδοτικότερη παρατήρηση, αναπτύσσονται συνεχώς νέα εργαλεία, βασισμένα στα Συνελικτικά Νευρωνικά Δίκτυα, τα οποία χρησιμοποιούν εικόνες τηλεπισκόπησης με σκοπό τη χαρτογράφηση και τον εντοπισμό αλλαγών στην επιφάνεια της Γης σε επίπεδο εικονοστοιχείου. Μια τέτοια ειδικότερη εφαρμογή είναι η παρακολούθηση των δασικών εκτάσεων, η οποία, λαμβάνοντας υπόψη την απειλή της αποψίλωσης των δασών, κρίνεται στις μέρες μας ιδιαιτέρως αναγκαία. Ωστόσο, οι ανοιχτά διαθέσιμες εικόνες τηλεπισκόπησης δεν παρουσιάζουν υψηλή χωρική ανάλυση. Ένας τρόπος για να αντιμετωπιστεί αυτό είναι η αύξηση της χωρικής ανάλυσης των εικόνων με χρήση ειδικών αρχιτεκτονικών Νευρωνικών Δικτύων, των Δικτύων Υπερανάλυσης Εικόνας.

Στόχος της παρούσας διπλωματικής είναι η υλοποίηση ενός συστήματος Μηχανικής Μάθησης, το οποίο θα επεξεργάζεται δορυφορικές εικόνες ώστε να εντοπίσει δασικές εκτάσεις με τη μέθοδο της σημασιολογικής κατάτμησης, η οποία αναγνωρίζει τα εικονοστοιχεία που αντιστοιχούν σε δάση. Αρχικά, δημιουργείται ένα σύνολο δεδομένων από εικόνες τηλεπισκόπησης της δορυφορικής αποστολής Sentinel-2 και τις αντίστοιχες ετικέτες τους, οι οποίες προέρχονται από το OpenStreetMap. Έπειτα, χρησιμοποιείται ένα Δίκτυο Υπερανάλυσης Εικόνας, το οποίο αυξάνει την ανάλυση των εικόνων και των ετικετών και έτσι προκύπτουν δύο σύνολα δεδομένων, ένα με εικόνες χωρίς υπερανάλυση και ένα με εικόνες με υπερανάλυση. Υστερα, εκπαιδεύονται και εξετάζονται 3 μοντέλα σημασιολογικής κατάτμησης πάνω στα σύνολα αυτά και μελετούνται τόσο οι επιδόσεις των μοντέλων μεταξύ τους, όσο και η επίδραση του συνόλου με υπερανάλυση στις αποδόσεις των μοντέλων. Τέλος, παρατίθενται τα συμπεράσματα της παραπάνω πειραματικής διαδικασίας και συζητούνται πιθανές βελτιώσεις και επεκτάσεις.

## Λέξεις Κλειδιά

Δορυφορική Τηλεπισκόπηση, Εντοπισμός Δασών, Sentinel-2, OpenStreetMap, Βαθιά Μηχανική Μάθηση, Συνελικτικά Νευρωνικά Δίκτυα, Σημασιολογική Κατάτμηση, Υπερανάλυση Εικόνας, Δίκτυα κωδικοποιητή-αποκωδικοποιητή, Μετασχηματιστές, Αναγεννητικά Ανταγωνιστικά Δίκτυα, ESRGAN, ResUNet-a, DeepLabv3+, Swin Transformer



## **Abstract**

---

With the ever-increasing amount of satellite missions in orbit, Earth observation and remote sensing have advanced rapidly and are applied in a multitude of fields, such as ecosystem monitoring and natural disaster prevention. At the same time, due to the remarkable progress of Machine Learning in Computer Vision, combined with the need for more efficient observations, more and more tools are developed, which rely on Convolutional Neural Networks and attempt to map and detect pixel-scale changes on the Earth's surface. Such an application is forest monitoring, which is deemed nowadays highly critical, considering the threat of deforestation on a global scale. However, the spatial resolution of available open-source satellite imagery is not sharp enough. One way to address this issue is to augment the spatial resolution using special Neural Network architectures, the Image Super Resolution Networks.

This diploma thesis aims to the development of a Machine Learning system, which processes satellite images in order to detect forest areas with the Semantic Segmentation method, which identifies the image pixels belonging to a forest area. Initially, a dataset is created which comprises of remote sensing images from the Sentinel-2 satellite mission along with the corresponding ground truth labels, which are collected from OpenStreetMap. After this, an Image Super Resolution architecture is used, which increases the resolution of the images and labels, resulting in having two datasets, one with super-resolved images and one with the original images. Then 3 Semantic Segmentation networks are trained and tested on the two datasets. The performance of the networks compared to one another, as well as the effect of the super-resolved dataset on the performance of the networks, are examined. Finally, conclusions on the aforementioned process are drawn and potential improvements and extensions are discussed.

## **Keywords**

Remote Sensing, Forest Detection, Sentinel-2, OpenstreetMap, Deep Learning, Convolutional Neural Networks, Semantic Segmentation, Image Super Resolution, Encoder-decoder networks, Transformers, Generative Adversarial Networks, ESRGAN, ResUNet-a, DeepLabv3+, Swin Transformer



*στοὺς γονεῖς μου*



## **Ευχαριστίες**

---

Θα ήθελα καταρχάς να ευχαριστήσω τον Καθηγητή κ. Στέφανο Κόλλια που μου έδωσε την ευκαιρία να ασχοληθώ με αυτήν την διπλωματική και για την εμπιστοσύνη που μου έδειξε. Επίσης, ευχαριστώ Θερμά την κ. Παρασκευή Τζούθελη για την άψογη συνεργασία που είχαμε και για την υποστήριξή της, χωρίς την οποία δεν θα ήταν εφικτή η παρούσα διπλωματική.

Τέλος, δεν θα μπορούσα να μην ευχαριστήσω την οικογένειά μου και τους φίλους μου που με στηρίζουν και με ανέχονται σε όλη τη διάρκεια αυτής της περιπέτειας που λέγεται ΣΗΜΜΥ, και με έχουν βοηθήσει στο να διαμορφώσω την προσωπικότητά μου και να γίνω αυτός που είμαι σήμερα.

Αθήνα, Σεπτέμβριος 2022

*Ελευθέριος Λυμπερόπουλος*



# Περιεχόμενα

---

<b>Περίληψη</b>	<b>1</b>
<b>Abstract</b>	<b>3</b>
<b>Ευχαριστίες</b>	<b>7</b>
<b>Πρόλογος</b>	<b>17</b>
<b>1 Εισαγωγή</b>	<b>19</b>
1.1 Αντικείμενο της διπλωματικής . . . . .	20
1.2 Σχετικές Εργασίες . . . . .	21
1.3 Οργάνωση του τόμου . . . . .	21
<b>I Θεωρητικό Μέρος</b>	<b>23</b>
<b>2 Δορυφορική Τηλεπισκόπηση</b>	<b>25</b>
2.1 Εισαγωγή . . . . .	25
2.2 Η ηλεκτρομαγνητική ενέργεια . . . . .	25
2.3 Άλληλεπίδραση με την ατμόσφαιρα και την επιφάνεια της Γης . . . . .	27
2.4 Το πρόγραμμα Copernicus . . . . .	28
2.5 Sentinel-2 . . . . .	29
<b>3 Νευρωνικά Δίκτυα στην 'Οραση Υπολογιστών</b>	<b>33</b>
3.1 Εισαγωγή . . . . .	33
3.2 Συνελικτικά Νευρωνικά Δίκτυα (CNN) . . . . .	34
3.2.1 Συνελικτικό Επίπεδο (Convolutional Layer) . . . . .	35
3.2.2 Επίπεδο Ενεργοποίησης (Activation Layer) . . . . .	35
3.2.3 Επίπεδο Υποδειγματοληψίας (Pooling Layer) . . . . .	36
3.2.4 Επίπεδο Κανονικοποίησης Παρτίδας (Batch Normalization Layer) . . . . .	37
3.2.5 Πλήρως Συνδεδεμένο Επίπεδο (Fully-Connected Layer) . . . . .	37
3.2.6 Residual Blocks . . . . .	38
3.3 Υπερανάλυση Εικόνας (Image Super-Resolution) . . . . .	39
3.4 Σημασιολογική Κατάτμηση Εικόνας (Semantic Segmentation) . . . . .	41

<b>II Προετοιμασία</b>	<b>45</b>
<b>4 Δημιουργία Συνόλου Δεδομένων</b>	<b>47</b>
4.1 Συγκέντρωση δορυφορικών εικόνων . . . . .	47
4.1.1 Google Earth Engine . . . . .	47
4.1.2 Προεπεξεργασία και λήψη των εικόνων . . . . .	48
4.2 Συγκέντρωση Ετικετών . . . . .	49
4.2.1 OpenStreetMap . . . . .	50
4.2.2 Λήψη Ετικετών . . . . .	50
4.3 Οργάνωση Δεδομένων . . . . .	50
<b>III Πειραματικό μέρος</b>	<b>53</b>
<b>5 Υλοποίηση</b>	<b>55</b>
5.1 Υπερανάλυση Εικόνας (Image Super-Resolution) . . . . .	55
5.1.1 SRGAN . . . . .	55
5.1.2 ESRGAN . . . . .	57
5.2 Σημασιολογική Κατάτμηση (Semantic Segmentation) . . . . .	58
5.2.1 ResUNet-a . . . . .	58
5.2.2 DeepLabv3+ . . . . .	60
5.2.3 Swin Transformer . . . . .	62
5.3 Μετρικές Αξιολόγησης . . . . .	65
5.3.1 Συναρτήσεις Απώλειας (Loss Functions) . . . . .	65
5.3.2 Μετρικές Ακρίβειας . . . . .	67
5.4 Λειπομέρειες Υλοποίησης . . . . .	68
5.4.1 Διαδικασία και παράμετροι Μοντέλων . . . . .	68
5.4.2 Υπολογιστικό Σύστημα . . . . .	69
<b>6 Πειραματικά Αποτελέσματα</b>	<b>71</b>
6.1 Ποσοτική Αξιολόγηση Αποτελεσμάτων . . . . .	71
6.1.1 Αποτελέσματα για τις εικόνες χωρίς υπερανάλυση . . . . .	71
6.1.2 Αποτελέσματα για τις εικόνες με υπερανάλυση . . . . .	72
6.1.3 Αποτελέσματα όλων των μοντέλων σε ένα κοινό σύνολο . . . . .	74
6.2 Ποιοτική Αξιολόγηση Αποτελεσμάτων . . . . .	75
6.3 Παρατηρήσεις ως προς τα αποτελέσματα . . . . .	78
6.3.1 Ιδιαιτερότητες Δεδομένων . . . . .	78
6.3.2 Επίδραση της Υπερανάλυσης Εικόνων . . . . .	79
<b>IV Επίλογος</b>	<b>81</b>
<b>7 Συμπεράσματα και Μελλοντικές Επεκτάσεις</b>	<b>83</b>
7.1 Συμπεράσματα . . . . .	83
7.2 Μελλοντικές Επεκτάσεις . . . . .	84





# Κατάλογος Σχημάτων

---

2.1	Οι δύο συνιστώσες της ηλεκτρομαγνητικής ακτινοβολίας: Ηλεκτρικό πεδίο και Μαγνητικό Πεδίο [1] . . . . .	26
2.2	Το ηλεκτρομαγνητικό φάσμα [1] . . . . .	26
2.3	Το ορατό ηλεκτρομαγνητικό φάσμα [2] . . . . .	27
2.4	Το ποσοστό της ανακλώμενης ακτινοβολίας για βλάστηση και νερό αναλόγως του μήκους κύματός της [1] . . . . .	28
2.5	Διαχωρισμός της ηλιακής ακτινοβολίας στο σύστημα του MSI [3] . . . . .	30
3.1	Παράδειγμα ενός Συνελικτικού Δικτύου για ταξινόμηση εικόνας. Με τα διαδοχικά συνελικτικά στρώματα μπορούν να υπολογιστούν τα διάφορα χαρακτηριστικά της εικόνας, ώστε να γίνει η σωστή ταξινόμηση από το πλήρως συνδεδεμένο στρώμα. [4] . . . . .	34
3.2	Παράδειγμα της συνέλιξης μιας RGB εικόνας με έναν πυρήνα. Το αποτέλεσμα είναι ένας χάρτης ενεργοποίησης. [4] . . . . .	36
3.3	Διάφορες συναρτήσεις ενεργοποίησης . . . . .	36
3.4	Υποδειγματοληψία με τη συνάρτηση μεγίστου . . . . .	37
3.5	Τα τελευταία επίπεδα ενός Συνελικτικού Δικτύου ταξινόμησης . . . . .	38
3.6	Σχηματική απεικόνιση ενός residual block . . . . .	38
3.7	Το πρόβλημα της Υπερανάλυσης Εικόνας . . . . .	39
3.8	Το πρόβλημα της Σημασιολογικής Κατάτμησης . . . . .	41
3.9	To FCN σχηματικά [5] . . . . .	42
3.10	Η αρχιτεκτονική κωδικοποιητή-αποκωδικοποιητή που συναντάται στο UNet. [6]	42
3.11	Ο πυρήνας της atrous συνέλιξης, για ρυθμό διαστολής 1, 2 και 3 αντίστοιχα. [7]	43
3.12	Η λειτουργία του Atrous Spatial Pyramid Module. Τα διάφορα field-of-views του εικονοστοιχείου είναι τα χρωματιστά πλαίσια στον κάτω χάρτη ενεργοποίησης. [8] . . . . .	43
4.1	Ζώνες της ίδιας δορυφορικής εικόνας . . . . .	49
4.2	Το αίτημα στο Overpass API. . . . .	50
4.3	Το αποτέλεσμα στο Overpass Turbo. . . . .	51
4.4	Παραδείγματα δορυφορικών εικόνων με τις ετικέτες τους. . . . .	52
5.1	Η δομή του γεννήτορα του SRGAN [9] . . . . .	56
5.2	Η δομή του διαχωριστή του SRGAN [9] . . . . .	56

5.3	Οι δύο βελτιώσεις του γεννήτορα του ESRGAN: Στα αριστερά βλέπουμε την αφαίρεση των επιπέδων της κανονικοποίησης παρτίδας μέσα σε κάθε residual block και στα δεξιά την προσθήκη των skip connections για τη δημιουργία της residual-in-residual μορφής. [10] . . . . .	57
5.4	Η διαφορά μεταξύ του κλασικού διαχωριστή του SRGAN και του οχετικιστικού διαχωριστή του ESRGAN. [10] . . . . .	58
5.5	(a) Η αρχιτεκτονική του ResUNet-a, στα αριστερά ο κωδικοποιητής, στα δεξιά ο αποκωδικοποιητής, (b) Η δομή του βασικού ResUNet-a block του ResUNet-a, με τα $d_i$ να αναπαριστούν τους διάφορους ρυθμούς διαστολής, (c) Το στρώμα Pyramid Scene Parsing Pooling. [11] . . . . .	59
5.6	Η βελτίωση του μοντέλου DeepLabv3, το οποίο ακολουθεί το μοντέλο Spatial Pyramid Pooling (a), με την αρχιτεκτονική encoder-decoder (b). Το DeepLabv3+ περιέχει πλούσιες σημασιολογικές πληροφορίες από τον κωδικοποιητή και παράγει λεπτομερή περιγράμματα αντικειμένων από τον απλό αλλά αποτελεσματικό αποκωδικοποιητή (c). [12] . . . . .	61
5.7	(a) Depthwise συνέλιξη, (b) pointwise συνέλιξη, (c) atrous depthwise συνέλιξη. [12] . . . . .	61
5.8	Η αρχιτεκτονική του DeepLabv3+ με το DeepLabv3 ως κωδικοποιητή. [12] . . . . .	62
5.9	Μια απεικόνιση της μεθόδου των shifting windows. Αριστερά, στο στρώμα 1, εφαρμόζεται ένας απλός διαχωρισμός της εικόνας σε παράθυρα, και το self-attention πραγματοποιείται στο καθένα από αυτά. Δεξιά, στο στρώμα l+1, τα παράθυρα έχουν μετακινηθεί. Ο υπολογισμός του self-attention στα νέα παράθυρα διασχίζει τα όρια των παραθύρων του προηγούμενου στρώματος, δημιουργώντας συνδέσεις μεταξύ τους. [13] . . . . .	64
5.10	(a) Η αρχιτεκτονική του μοντέλου Swin-T του Swin Transformer (b) Δύο συνεχόμενα Swin Transformer blocks. [13] . . . . .	64
5.11	Η δομή του UperNet με το Feature Pyramid Network-FPN ως backbone. Στο κάτω μέρος βλέπουμε τη δομή κάθε head αναλόγως με το επίπεδο στο οποίο ταξινομεί. [14] . . . . .	65
5.12	Η διαδικασία για τις εικόνες με υπερανάλυση. . . . .	68
6.1	Εικόνες του test set χωρίς υπερανάλυση μαζί με τις ετικέτες τους και τις προβλέψεις των μοντέλων. . . . .	76
6.2	Εικόνες του test set με υπερανάλυση μαζί με τις ετικέτες τους και τις προβλέψεις των μοντέλων. . . . .	77
6.3	Ένα παράδειγμα δορυφορικής εικόνας με μια μη ακριβή ετικέτα της. . . . .	78
6.4	Η ίδια δορυφορική εικόνα χωρίς και με Υπερανάλυση . . . . .	79

# Κατάλογος Πινάκων

---

2.1	Οι περιοχές του ηλεκτρομαγνητικού φάσματος . . . . .	27
2.2	Οι ζώνες που καλύπτει ο MSI . . . . .	31
6.1	Τα πειραματικά αποτελέσματα των μοντέλων για τη μετρική Dice Coefficient για το σύνολο εικόνων χωρίς υπερανάλυση. . . . .	71
6.2	Τα πειραματικά αποτελέσματα των μοντέλων για τη μετρική IoU για το σύνολο εικόνων χωρίς υπερανάλυση. . . . .	72
6.3	Τα πειραματικά αποτελέσματα των μοντέλων για τη μετρική Precision για το σύνολο εικόνων χωρίς υπερανάλυση. . . . .	72
6.4	Τα πειραματικά αποτελέσματα των μοντέλων για τη μετρική Recall για το σύνολο εικόνων χωρίς υπερανάλυση. . . . .	72
6.5	Τα πειραματικά αποτελέσματα των μοντέλων για τη μετρική Dice Coefficient για το σύνολο εικόνων με υπερανάλυση. . . . .	73
6.6	Τα πειραματικά αποτελέσματα των μοντέλων για τη μετρική IoU για το σύνολο εικόνων με υπερανάλυση. . . . .	73
6.7	Τα πειραματικά αποτελέσματα των μοντέλων για τη μετρική Precision για το σύνολο εικόνων με υπερανάλυση. . . . .	73
6.8	Τα πειραματικά αποτελέσματα των μοντέλων για τη μετρική Recall για το σύνολο εικόνων με υπερανάλυση. . . . .	73
6.9	Τα πειραματικά αποτελέσματα όλων των μοντέλων για τη μετρική Dice Coefficient στο κοινό σύνολο δοκιμής. ΣΗΜΕΙΩΣΗ: Με χ.υ. σημειώνονται τα μοντέλα που προπονήθηκαν στο σύνολο χωρίς υπερανάλυση και με μ.υ. σημειώνονται τα μοντέλα που προπονήθηκαν στο σύνολο με υπερανάλυση. . . . .	74
6.10	Τα πειραματικά αποτελέσματα όλων των μοντέλων για τη μετρική IoU στο κοινό σύνολο δοκιμής. . . . .	74
6.11	Τα πειραματικά αποτελέσματα όλων των μοντέλων για τη μετρική Precision στο κοινό σύνολο δοκιμής. . . . .	75
6.12	Τα πειραματικά αποτελέσματα όλων των μοντέλων για τη μετρική Recall στο κοινό σύνολο δοκιμής. . . . .	75



## **Πρόλογος**

---

Η παρούσα διπλωματική εκπονήθηκε στην Αθήνα, το έτος 2022, στο Εργαστήριο Συστημάτων Τεχνητής Νοημοσύνης και Μηχανικής Μάθησης που ανήκει στον Τομέα Τεχνολογίας Πληροφορικής και Υπολογιστών του Εθνικού Μετσόβιου Πολυτεχνείου, με επιβλέποντες τον Καθηγητή κ. Στέφανο Κόλλια και την κ. Παρασκευή Τζούθελη.



### Εισαγωγή

---

Τα δάση καλύπτουν περίπου το ένα τρίτο της χερσαίας επιφάνειας της Γης και αποθηκεύουν περίπου το 45% του παγκόσμιου άνθρακα. Περίπου η μισή από την έκτασή τους αποτελείται από τροπικά δάση, των οποίων τα δέντρα συλλέγουν κατά μέσο όρο 50% παραπάνω άνθρακα από τα δέντρα άλλων δασών [15]. Εκτός από κρίσιμο παράγοντα ρύθμισης του παγκόσμιου κλίματος, τα δάση συμβάλλουν κοινωνικά, οικονομικά και αισθητικά στην ζωή των ανθρώπων. [16] Είναι κοινώς αποδεκτό ότι τα δάση προσφέρουν πολύ καλή προστασία από τη διάθρωση εξαιτίας της βροχής, του ανέμου και των παράκτιων κυμάτων, ενώ οι ρίζες των δέντρων βοηθούν και στη σκληρότητα του εδάφους. Επίσης, ένα μεγάλο μέρος του διαθέσιμου πόσιμου νερού προέρχεται από φυσικά ή τεχνητά δάση και τουλάχιστον το ένα τρίτο από τις μεγαλύτερες πόλεις του κόσμου αντλούν νερό από δασικές περιοχές [17]. Επιπλέον, τα δάση αποτελούν σημαντικές πηγές ενέργειας ιδιαίτερα στον αναπτυσσόμενο κόσμο με τη μορφή ξυλείας.

Ωστόσο, ο ανθρώπινος παράγοντας παρεμβαίνει στα δάση, μειώνοντας την έκτασή τους για αγροτικές και κτηνοτροφικές δραστηριότητες και για την επέκταση πόλεων [18]. Η αλλαγή αυτή της επιφάνειας της Γης παίζει καίριο ρόλο στην κλιματική αλλαγή. Η αποψίλωση των δασών είναι μια παγκόσμια απειλή, όχι μόνο γιατί διαταράσσει τα φυσικά περιβάλλοντα και τη βιοποικιλότητα, αλλά και γιατί έχει επιπτώσεις στο κλίμα με την απελευθέρωση αερίων του θερμοκηπίου και συγκεκριμένα του διοξειδίου του άνθρακα. Η εκπομπή μεγάλης ποσότητας διοξειδίου του άνθρακα επιφέρει την υπερθέρμανση του πλανήτη και κατά συνέπεια την διαταραχή των οικοσυστημάτων και την εμφάνιση ακραίων φαινομένων και καταστροφών.

Η παρακολούθηση των δασών επιτρέπει την καταγραφή των αλλαγών. Η συχνή και ακριβής παρακολούθηση συμβάλλει στη λήψη μέτρων για την προστασία και διατήρηση των δασών ώστε να διασφαλιστεί ο ρόλος των δασών στο περιβάλλον και τον άνθρωπο [18]. Η σημασία της παρακολούθησης των δασών είναι καίρια, όπως τονίζεται σε διεθνή πλαίσια, όπως το Πλαίσιο των Ηνωμένων Εθνών για την Κλιματική Αλλαγή (United Nations Framework Convention on Climate Change). Κύριο μέλημα των συστημάτων παρακολούθησης είναι η καταγραφή των δασικών εκτάσεων καθώς και των μεταβολών τους, αλλά και των επιπέδων άνθρακα. Στο έργο αυτό, η γεωπαρατήρηση έχει σημαντική συμβολή, καθώς ολοένα και περισσότερες δορυφορικές αποστολές και τεχνολογίες είναι αφιερωμένες στην παρατήρηση της επιφάνειας της Γης ή και σε τμήματά της όπως οι δασικές περιοχές, και της ανακλώμενης από αυτή ηλιακής ακτινοβολίας, παράγοντας έναν μεγάλο όγκο δεδομένων δουρφορικής τηλεπισκόπησης (remote sensing) [19].

Παράλληλα, η άνοδος της Μηχανικής Μάθησης και συγκεκριμένα του κλάδου της Υπολογιστικής Όρασης έχει συγκεντρώσει το ενδιαφέρον της ερευνητικής κοινότητας για την επεξεργασία αυτού του όγκου δεδομένων, που αποτελείται στο μεγαλύτερο μέρος του από δορυφορικές εικόνες. Τα Συνελικτικά Νευρωνικά Δίκτυα, ο θεμελιώδης αλγόριθμος της Υπολογιστικής Όρασης, χρησιμοποιούνται πλέον σε όλο και περισσότερες σύγχρονες εφαρμογές στο χώρο της γεωπαρατήρησης και του εντοπισμού αντικειμένων και αλλαγών στην επιφάνεια της Γης. Όσον αφορά την παρακολούθηση δασικών εκτάσεων, τα τελευταία χρόνια έχουν αναπτυχθεί αρχιτεκτονικές Νευρωνικών Δικτύων και αλγόριθμοι που σχετίζονται μεταξύ άλλων με την καταγραφή δασικών εκτάσεων [20], της αποψίλωσης [21] και της πρόβλεψης πιθανότητας πυρκαϊάς [22].

Ωστόσο, ένα συνηθισμένο πρόβλημα με τέτοιες εφαρμογές είναι ότι οι διαθέσιμες open-source δορυφορικές εικόνες τηλεπισκόπησης δεν είναι υψηλής ανάλυσης, ενώ η πρόσθαση σε εικόνες υψηλότερης ανάλυσης είναι δύσκολη και δαπανηρή. Για τη διαχείριση τέτοιων εικόνων, στην Όραση Υπολογιστών έχουν προταθεί ορισμένες αρχιτεκτονικές Νευρωνικών Δικτύων, οι λεγόμενες αρχιτεκτονικές Υπερανάλυσης Εικόνας, οι οποίες αυξάνουν την ανάλυση των δορυφορικών εικόνων, οξύνοντας τα χαρακτηριστικά τους. Βασισμένοι σε αυτή την ιδέα, θα χρησιμοποιήσουμε μια τέτοια αρχιτεκτονική στην παρούσα εργασία.

## 1.1 Αντικείμενο της διπλωματικής

Αντικείμενο της διπλωματικής είναι η δημιουργία ενός εργαλείου, το οποίο θα λαμβάνει ως είσοδο δορυφορικές εικόνες και θα εντοπίζει σε αυτές δασικές εκτάσεις με ακρίβεια εικονοστοιχείου. Για να το πετύχουμε αυτό, ασχολούμαστε με τη μελέτη διάφορων αρχιτεκτονικών Νευρωνικών Δικτύων. Χρησιμοποιούνται 3 σύγχρονες αρχιτεκτονικές και πραγματοποιείται σύγκριση των αποτελεσμάτων τους. Επιπλέον, εξετάζεται η επίδραση της μεθόδου της Υπερανάλυσης Εικόνας με τη χρήση μιας κατάλληλης αρχιτεκτονικής, η οποία αυξάνει την ανάλυση των εικόνων και των χαρακτηριστικών τους.

Αναλυτικότερα, η διαδικασία που πραγματοποιήθηκε είναι η εξής: Αρχικά, δημιουργήθηκε το σύνολο δεδομένων με τις δορυφορικές εικόνες και τις ετικέτες (ground truths) που αντιστοιχούν στις δασικές εκτάσεις. Τόσο οι εικόνες όσο και οι ετικέτες αντλήθηκαν με χρήση open-source εργαλείων. Έπειτα, χρησιμοποιούμε ένα Νευρωνικό Δίκτυο Υπερανάλυσης Εικόνας για να μεγεθύνουμε τις δορυφορικές εικόνες με τις ετικέτες τους, δημιουργώντας έτσι δύο σύνολα δεδομένων, ένα σύνολο χωρίς υπερανάλυση και ένα σύνολο με υπερανάλυση. Εκπαιδεύουμε τα 3 μοντέλα που αναφέραμε παραπάνω ξεχωριστά για κάθε σύνολο δεδομένων, αλλά και για κάθε μια από 4 συναρτήσεις σφάλματος που επιλέξαμε για την εκπαίδευση. Για κάθε σύνολο δεδομένων, συγκρίνουμε τις επιδόσεις κάθε αρχιτεκτονικής σε συνδυασμό με τη συνάρτηση σφάλματος με την οποία εκπαιδεύτηκε. Τέλος, για να εξετάσουμε την επίδραση της μεθόδου της Υπερανάλυσης Εικόνας, δοκιμάζουμε όλες τις αρχιτεκτονικές, είτε έχουν εκπαιδευτεί στο σύνολο χαμηλής ανάλυσης είτε στο σύνολο υψηλής ανάλυσης, σε ένα κοινό σύνολο δοκιμής, το οποίο αποτελείται από εικόνες και υψηλής και χαμηλής ανάλυσης, και εξάγουμε τα ανάλογα συμπεράσματα.

## 1.2 Σχετικές Εργασίες

Τα τελευταία χρόνια έχουν αναπτυχθεί αρχιτεκτονικές Νευρωνικών Δικτύων σκοπό την καταγραφή και χαρτογράφηση δασικών εκτάσεων. Στο [20], χρησιμοποιείται μια αρχιτεκτονική Συνελικτικού Νευρωνικού Δικτύου, η οποία βελτιώνει τη χαρτογράφηση δασικών εκτάσεων του Αμαζονίου σε σχέση με προηγούμενες μεθόδους που δεν χρησιμοποιούσαν μηχανική μάθηση, ενώ είναι σε θέση να καταγράφει αλλαγές και να παράγει τη χαρτογράφηση πολύ γρηγορότερα από τις παλαιότερες μεθόδους. Στο [23] συγκρίνονται στο σύνολο δεδομένων TanDEM-X τρεις αρχιτεκτονικές, οι ResNet, DenseNet και UNet για την αναγνώριση των εικονοστοιχείων που ανήκουν σε δασική έκταση, με το UNet να πετυχαίνει τις καλύτερες επιδόσεις. Μια πιο πρόσφατη εφαρμογή στο [24] χρησιμοποιεί πάνω σε δορυφορικές εικόνες από την αποστολή Sentinel-2 μια αρχιτεκτονική τύπου UNet για να χαρτογραφήσει δασικές εκτάσεις στην Αφρική καθώς και να υπολογίσει το ύψος αυτών των εκτάσεων. Η μέθοδος αυτή μάλιστα είχε χαμηλό υπολογιστικό κόστος που επιτρέπει την επεξεργασία των εικόνων του Sentinel-2 σε σχεδόν πραγματικό χρόνο, ακόμα και σε ηπειρωτική κλίμακα.

Γενικότερα, αρχιτεκτονικές Βαθιών Νευρωνικών Δικτύων έχουν υλοποιηθεί και χρησιμοποιηθεί σε διάφορες εφαρμογές από μέλη του Εργαστηρίου Συστημάτων Τεχνητής Νοημοσύνης και Μάθησης του ΕΜΠ. Ειδικότερα τεχνικές CNN και CNN-RNN έχουν εφαρμοστεί για ιατρική διάγνωση νευροεκφυλιστικών ασθενειών, όπως της νόσου του Πάρκινσον [25–29] ή της Covid-19 [30–32], βασισμένες σε 2-Δ ή 3-Δ εικόνες. Εμφαση έχει δοθεί στην διαφάνεια και στην προσαρμογή των μοντέλων [33–35], αλλά και στην ανάπτυξη πλέον σύνθετων αρχιτεκτονικών, μπαϋεσιανών, με κάψουλες και αβεβαιότητα [36–39]. Βαθιές 3-Δ νευρωνικές αρχιτεκτονικές έχουν εφαρμοστεί στην ανίχνευση βλαβών σε πυρηνικούς αντιδραστήρες [40, 41], στην πρόβλεψη της παραγωγής στον αγροτικό τομέα [42, 43] και στην αναγνώριση και σύνθεση συναισθήματος [44–47], ενώ άλλες εφαρμόζονται σε προβλήματα ανάλυσης εικόνων και αλληλεπίδρασης ανθρώπου-υπολογιστή [48–50].

## 1.3 Οργάνωση του τόμου

Η εργασία αυτή είναι οργανωμένη σε επτά κεφάλαια: Στα πρώτα δύο δίνεται το θεωρητικό υπόβαθρο που είναι απαραίτητο για την κατανόηση της εργασίας. Στο κεφάλαιο 2 γίνεται μια εισαγωγή στη Δορυφορική Τηλεπισκόπηση και το ηλεκτρομαγνητικό φάσμα και παρουσιάζονται πληροφορίες σχετικά με τις δορυφορικές εικόνες που χρησιμοποιούνται στη συνέχεια. Στο κεφάλαιο 3 παρουσιάζονται βασικές έννοιες των Νευρωνικών Δικτύων στην υπολογιστική όραση και τα προβλήματα της Σημασιολογικής Κατάτμησης και Υπερανάλυσης Εικόνας που θα μας απασχολήσουν στη συνέχεια. Στο κεφάλαιο 4 αναλύονται τα εργαλεία και η διαδικασία με την οποία αποκτήσαμε και οργανώσαμε το σύνολο δεδομένων μας. Επίτητα, στο κεφάλαιο 5 εξετάζονται αναλυτικά οι αρχιτεκτονικές των Νευρωνικών Δικτύων που χρησιμοποιήσαμε στην πειραματική διαδικασία. Επίσης, εξηγούνται οι μετρικές αξιολόγησης που χρησιμοποιήθηκαν και περιγράφεται λεπτομερώς η πειραματική διαδικασία. Στο κεφάλαιο 6 παρουσιάζονται τα αποτελέσματα της πειραματικής διαδικασίας και σχολιάζονται και, τέλος, στο κεφάλαιο 7 πραγματοποιείται η σύνοψη της εργασίας και η εξαγωγή των τελικών συμπερασμάτων.



## **Μέρος Ι**

### **Θεωρητικό Μέρος**

---



## Κεφάλαιο 2

### Δορυφορική Τηλεπισκόπηση

---

#### 2.1 Εισαγωγή

Η τηλεπισκόπηση (remote sensing), σύμφωνα με έναν από τους ευρύτερους ορισμούς της, είναι η επιστήμη και τεχνολογία της παρατήρησης και εξαγωγής δεδομένων για ένα αντικείμενο ή φαινόμενο από απόσταση, χωρίς να απαιτείται άμεση επαφή με αυτό. Τα δεδομένα αυτά εξάγονται όταν κάποιο όργανο καταγράφει την ενέργεια που εκπέμπει ή ανακλάται από το αντικείμενο ή φαινόμενο ενδιαφέροντος. Αυτή η ενέργεια μπορεί να είναι ηλεκτρομαγνητική, ηχητική, θερμότητα ή άλλη. [51] Αν και η τηλεπισκόπηση έχει ευρείες εφαρμογές, στο πλαίσιο αυτής της εργασίας θα ασχοληθούμε με δορυφορική τηλεπισκόπηση για την παρατήρηση της επιφάνειας της Γης. Αυτή πραγματοποιείται κυρίως με καταγραφή της προσπίπτουσας και ανακλώμενης ηλεκτρομαγνητικής ενέργειας από αισθητήρες αεροσκάφους ή δορυφορικούς. [2]

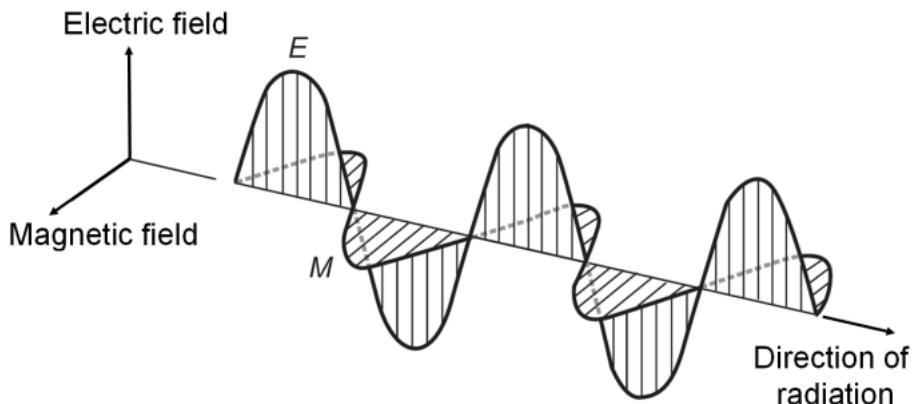
Ιστορικά, η τηλεπισκόπηση, με τη μορφή αεροφωτογραφιών, ξεκίνησε από την δεκαετία του 1910 και χρησιμοποιήθηκε αρχικά για χαρτογράφηση και στρατιωτικούς σκοπούς. Ωστόσο, από τα τέλη της δεκαετίας του 1950, γνώρισε μεγάλη ανάπτυξη και χρησιμοποιήθηκε για σκοπούς όπως οι κλιματολογικές και μετεωρολογικές παρατηρήσεις, η αγρονομία και η δασονομία. Μέχρι τα τέλη του 1990, η δορυφορική τηλεπισκόπηση μπορούσε, μεταξύ άλλων, να εντοπίσει και να προβλέψει φυσικές καταστροφές όπως οι πλημμύρες και να παρακολουθήσει μεταβολές στα οικοσυστήματα, τον αστικό ιστό και την αγροτική δραστηριότητα του ανθρώπου. [52] Από τους πρώτους δορυφόρους που εκτοξεύθηκαν, τον σοβιετικό Sputnik-1 το 1957 και τον αμερικανικό TIROS-1 το 1960, έχουν ακολουθήσει εκατοντάδες [53] δορυφορικές αποστολές, πρωτοπόρες από τις οποίες υπήρξαν οι SPOT του Γαλλικού Εθνικού Κέντρου Διαστημικών Ερευνών CNES (Centre national d'études spatiales) καθώς και η Landsat της Αμερικανικής Εθνικής Υπηρεσίας Αεροναυπηγικής και Διαστήματος NASA (National Aeronautics and Space Administration). [54]

#### 2.2 Η ηλεκτρομαγνητική ενέργεια

Όπως αναφέραμε πιο πάνω, η δορυφορική τηλεπισκόπηση πραγματοποιείται με καταγραφή της προσπίπτουσας και ανακλώμενης ηλεκτρομαγνητικής ακτινοβολίας. Επομένως σίναι απαραίτητη η αναφορά στην ηλεκτρομαγνητική ενέργεια. Η ηλεκτρομαγνητική ενέργεια μεταφέρεται αποτελείται από ηλεκτρομαγνητικά κύματα που οφείλονται σε δύο ταλα-

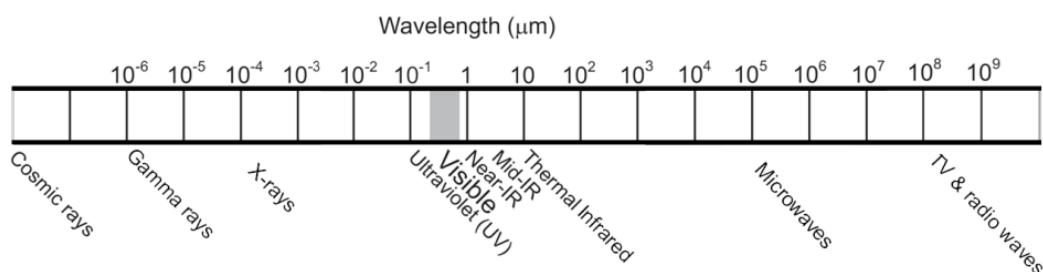
ντούμενες συνιστώσες, το ηλεκτρικό και το μαγνητικό πεδίο. Τα πεδία αυτά ταλαντώνονται σε επίπεδα κάθετα μεταξύ τους και κάθετα προς τη διεύθυνση διάδοσης. Επίσης, τα πεδία αυτά ταλαντώνονται συγχρονισμένα, για παράδειγμα ένα θετικό υψηλό της ηλεκτρικής ενέργειας και ένα αρνητικό υψηλό της μαγνητικής ενέργειας συμβαίνουν ταυτόχρονα. Τα ηλεκτρομαγνητικά κύματα ταξιδεύουν στο κενό με την ταχύτητα του φωτός  $c$  ( $c = 299.792.458 \text{ m/s}$ ) και χαρακτηρίζονται από τη συχνότητα  $f$  και το μήκος κύματος  $\lambda$ , το γινόμενο των οποίων ισούται με την ταχύτητα  $c$ . [2]

$$f\lambda = c$$

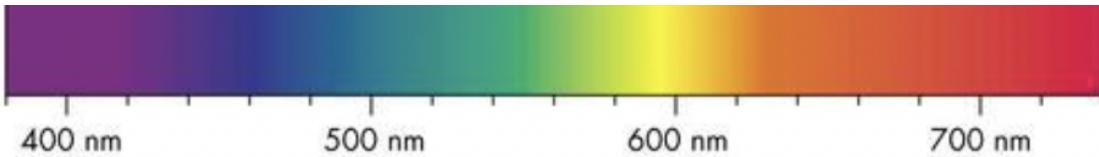


**Σχήμα 2.1:** Οι δύο συνιστώσες της ηλεκτρομαγνητικής ακτινοβολίας: Ηλεκτρικό πεδίο και Μαγνητικό Πεδίο [1]

Καλούμε το σύνολο των μηκών κύματος της ηλεκτρομαγνητικής ακτινοβολίας το **ηλεκτρομαγνητικό φάσμα**. Συμβατικά, χωρίζουμε το ηλεκτρομαγνητικό φάσμα σε περιοχές, ανάλογα με το μήκος κύματος της. Αυτές είναι τα ραδιοκύματα, τα μικροκύματα, η υπέρυθρη, η ορατή και η υπεριώσης ακτινοβολία, οι ακτίνες X και οι ακτίνες γάμμα. Το ηλεκτρομαγνητικό φάσμα είναι συνεχές και δεν υπάρχει κάποιο ξεκάθαρο όριο μεταξύ των περιοχών. [2] Στην παρούσα εργασία μας ενδιαφέρουν οι περιοχές που δημιουργούνται από την ηλιακή ακτινοβολία και συγκεκριμένα οι περιοχές της υπέρυθρης, ορατής και υπεριώδους ακτινοβολίας.



**Σχήμα 2.2:** Το ηλεκτρομαγνητικό φάσμα [1]



Σχήμα 2.3: Το ορατό ηλεκτρομαγνητικό φάσμα [2]

Περιοχή	Μήκος Κύματος
Ραδιοκύματα	> 0.3m
Μικροκύματα	1mm – 0.3m
Υπέρυθρη ακτινοβολία	700nm – 1mm
Ορατή ακτινοβολία	400nm – 700nm
Υπεριώδης ακτινοβολία	10nm – 400nm
Ακτίνες X	0.01nm – 10nm
Ακτίνες γάμμα	< 0.01nm

Πίνακας 2.1: Οι περιοχές του ηλεκτρομαγνητικού φάσματος

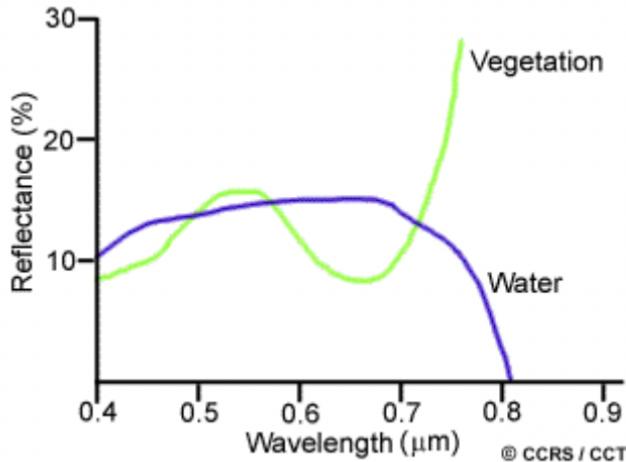
## 2.3 Αλληλεπίδραση με την ατμόσφαιρα και την επιφάνεια της Γης

Προτού η ακτινοβολία φτάσει τη γήινη επιφάνεια πρέπει να περάσει από την ατμόσφαιρα. Εκεί τα σωματίδια και τα αέρια της ατμόσφαιρας επηρεάζουν την ακτινοβολία που θα φτάσει τελικά στην επιφάνεια και λαμβάνουν χώρα κυρίως τα φαινόμενα της σκέδασης και της απορρόφησης. Η σκέδαση (diffusion) συμβαίνει όταν σωματίδια ή μεγάλα μόρια αερίων στην ατμόσφαιρα αλληλεπιδρούν με την ηλεκτρομαγνητική ακτινοβολία και την ανακατευθύνουν από την αρχική της πορεία. Η απορρόφηση συμβαίνει όταν μόρια της ατμόσφαιρας απορροφούν ενέργεια διαφόρων μηκών κύματος. Κατά κύριο λόγο η απορρόφηση οφείλεται στο όζον, το οποίο απορροφά την υπεριώδη ακτινοβολία, και την παρουσία διοξειδίου του άνθρακα και μορίων εξατμισμένου νερού. [1]

Το μέρος της ακτινοβολίας που δεν σκεδάζεται ή απορροφάται από την ατμόσφαιρα αλληλεπιδρά με την επιφάνεια της Γης. Όταν η ακτινοβολία προσπίπτει με την επιφάνεια της Γης, μπορεί να συμβεί ένα ή περισσότερα από τα ακόλουθα τρία φαινόμενα: *απορρόφηση*, *μεταβίβαση* και *η ανάκλαση*. Απορρόφηση (absorption) συμβαίνει όταν η ακτινοβολία απορροφάται από το αντικείμενο στο οποίο προσπίπτει, μεταβίβαση (transmission) συμβαίνει όταν το διαπερνά και ανάκλαση (/ενρεφλεξτιον) συμβαίνει όταν η πρόσπιτωση με το αντικείμενο ανακατευθύνει τις ακτίνες από την αρχική τους κατέυθυνση. Το κατά πόσο συμβαίνουν αυτά τα φαινόμενα εξαρτάται από το αντικείμενο καθώς και την επιφάνειά του. [1]

Για παράδειγμα, τα φύλλα των δέντρων περιέχουν μια χημική ένωση, τη χλωροφύλλη, η οποία απορροφά την κόκκινη και μπλε ορατή ακτινοβολία καθώς και ένα μέρος της υπέρυθρης, ενώ ανακλά την πράσινη και το υπόλοιπο μέρος της υπέρυθρης. Για αυτό το καλοκαίρι, όταν η ποσότητα της χλωροφύλλης είναι η μέγιστη, τα φύλλα φαίνονται πιο πράσινα, ενώ το φθινόπωρο, επειδή η χλωροφύλλη μειώνεται, ανακλάται περισσότερη κόκκινη ορατή ακτινοβολία και τα φύλλα φαίνονται κίτρινα. Επίσης, όσον αφορά τους ωκεανούς, απορροφούνται

περισσότερο τα μεγάλα μήκη κύματος, δηλαδή οι υπέρυθρες και οι κόκκινες ακτινοβολίες, από τα άλλα μήκη κύματος και για αυτό οι ωκεανοί φαίνονται μπλε ή μπλε-πράσινοι.



**Σχήμα 2.4:** Το ποσοστό της ανακλώμενης ακτινοβολίας για βλάστηση και νερό αναλόγως του μήκους κύματός της [1]

Αντίστοιχα, για κάθε αντικείμενο, αν μετρήσουμε το ποσοστό και το μήκος κύματος της ενέργειας που απορροφάται, μεταβιβάζεται ή διαχέεται, μπορούμε να σχηματίσουμε τη φασματική απόκριση (*spectral response*) του αντικειμένου αυτού. Το να γνωρίζουμε τους παράγοντες που καθορίζουν τη συμπεριφορά των αντικειμένων ενδιαφέροντος είναι πολύ σημαντικό ώστε να ερμηνεύσουμε σωστά την αλληλεπίδραση της ηλεκτρομαγνητικής ακτινοβολίας με αυτά. [1]

## 2.4 Το πρόγραμμα Copernicus

Το πρόγραμμα Copernicus αποτελεί το πρόγραμμα γεωπαρατήρησης της Ευρωπαϊκής Ένωσης. Χρηματοδοτείται και οργανώνεται από την Ευρωπαϊκή Επιτροπή (European Commission) σε συνεργασία με τον Ευρωπαϊκό Οργανισμό Διαστήματος (European Space Agency - ESA). [55] Στόχος του προγράμματος είναι η δημιουργία ενός συστήματος για τη συλλογή μεγάλου όγκου δεδομένων γεωπαρατήρησης παγκόσμιας κλίμακας και πραγματικού χρόνου για να χρησιμοποιηθούν σε ευρείες εφαρμογές. Συγκεκριμένα, τα δεδομένα αυτά επεξεργάζονται και χρησιμοποιούνται από τις υπηρεσίες του προγράμματος, οι οποίες αφορούν 6 θεματικές ενότητες: την παρακολούθηση της ατμόσφαιρας, του θαλάσσιου οικοσυστήματος, της χερσαίας επιφάνειας της Γης και της κλιματικής αλλαγής, καθώς και την ασφάλεια και τη διαχείριση έκτακτων καταστάσεων.

Για το πρόγραμμα έχει δημιουργηθεί μια οικογένεια εξειδικευμένων δορυφόρων, οι Sentinel. Οι Sentinel καλύπτουν πολλές από τις ανάγκες του προγράμματος και στόχος είναι να αντικαταστήσουν τις παλαιότερες δορυφορικές αποστολές, όπως η ERS και η Envisat, διασφαλίζοντας τη συνέχεια στα δεδομένα. [56]. Το πρόγραμμα Copernicus περιλαμβάνει 6 δορυφορικές αποστολές Sentinel:

- Η αποστολή **Sentinel-1** χρησιμοποιεί προηγμένα όργανα ραντάρ για να καταγράψει,

ανεξαρτήτως καιρικών συνθηκών, εικόνες της χερσαίας και θαλάσσιας επιφάνειας της Γης. Αποτελείται από δύο δορυφόρους, τον Sentinel-1A, που εκτοξεύθηκε στις 3 Απριλίου 2014, και τον Sentinel-1B, που εκτοξεύθηκε στις 25 Απριλίου 2016. Στο μέλλον θα αποτελείται επίσης από τους δορυφόρους Sentinel-1C και Sentinel-1D, με την εκτόξευση του πρώτου να είναι προγραμματισμένη για το πρώτο μισό του 2023. [57]

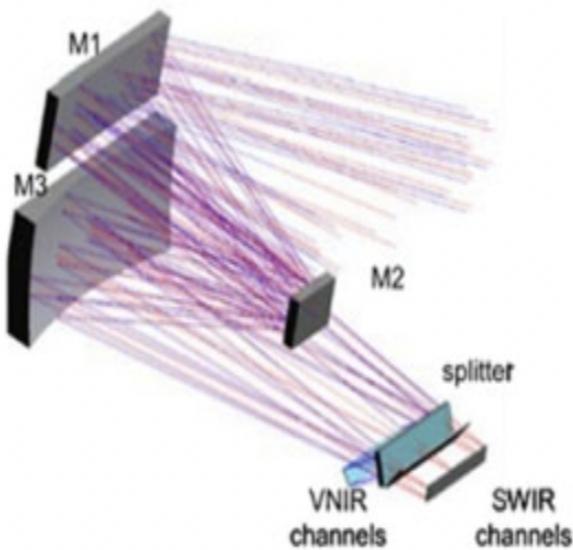
- Η αποστολή **Sentinel-2** δίνει με υψηλή χωρική ανάλυση πολυφασματικές εικόνες και παρακολουθεί τα χερσαία οικοσυστήματα, τα δάση, την ποιότητα του νερού και φυσικές καταστροφές. Αποτελείται από δύο δορυφόρους, τον Sentinel-2A, που εκτοξεύθηκε στις 23 Ιουνίου 2015, και τον Sentinel-2B, που εκτοξεύθηκε στις 7 Μαρτίου 2017. Στο μέλλον θα αποτελείται επίσης από τους δορυφόρους Sentinel-2C και Sentinel-2D, με την εκτόξευση του πρώτου να είναι προγραμματισμένη για τις αρχές του 2024. [58]
- Η αποστολή **Sentinel-3** παρέχει υψηλής ακρίβειας δεδομένα σχετικά με την τοπογραφία της επιφάνειας των θαλασσών και λιμνών, τη θερμοκρασία του νερού και του εδάφους και τον χρωματισμό των ωκεανών και της ξηράς. Αποτελείται από δύο δορυφόρους, τον Sentinel-3A, που εκτοξεύθηκε στις 16 Φεβρουαρίου 2016, και τον Sentinel-3B, που εκτοξεύθηκε στις 25 Απριλίου 2018. [59] Στο μέλλον θα αποτελείται επίσης από τους δορυφόρους Sentinel-3C και Sentinel-3D.
- Οι αποστολές **Sentinel-4**, **Sentinel-5** και **Sentinel-5P** λειτουργούν συμπληρωματικά μεταξύ τους και επικεντρώνονται στη μελέτη της σύνθεσης της ατμόσφαιρας, την καταγραφή της ποιότητας του αέρα και του ποσού της ηλιακής ακτινοβολίας με υψηλή χωρική και χρονική ανάλυση. [60] Ο Sentinel-5P εκτοξεύθηκε στις 13 Οκτωβρίου του 2017 ως πρόδρομος του Sentinel-5, ενώ ο Sentinel-4 είναι προγραμματισμένος να εκτοξευθεί το 2023.
- Η αποστολή **Sentinel-6** επικεντρώνεται στη μελέτη της παγκόσμιας στάθμης της θάλασσας, του ύψους των κυμάτων και άλλων χαρακτηριστικών που άπτονται της ωκεανολογίας, της μετεωρολογίας και της υδρολογίας, παρέχοντας υψηλής ακρίβειας και πραγματικού χρόνου δεδομένα. Αποτελείται από δύο δορυφόρους, ο ένας εκ των οποίων εκτοξεύθηκε το 2020, ενώ ο δεύτερος είναι προγραμματισμένος να εκτοξευθεί το 2025. [61]

## 2.5 Sentinel-2

Καθώς η παρούσα εργασία σχετίζεται με την κατάτμηση των δορυφορικών εικόνων με βάση τη μορφολογία του εδάφους, επιλέχθηκαν ως πλέον κατάλληλα τα δεδομένα της αποστολής Sentinel-2.

Όπως αναφέρθηκε, η αποστολή Sentinel-2 αποτελείται από δύο δορυφόρους οι οποίοι κινούνται στην ίδια τροχιά, με μέσο ύψος από την επιφάνεια της Γης 786 km, και με διαφορά φάσης μεταξύ τους 180°. Με τη χρήση δύο δορυφόρων επιτυγχάνεται ο μισός χρόνος επαναδιέλευσης, δηλαδή από 10 ημέρες σε 5 στον Ισημερινό σε συνθήκες χωρίς νέφη. Τα γεωγραφικά πλάτη που καλύπτουν είναι από 56° Νότια έως 84° Βόρεια. [62]

Κάθε ένας από τους δορυφόρους φέρει έναν υψηλής ανάλυσης Πολυφασματικό Σαρωτή MSI (MultiSpectral Instrument) με εύρος πεδίου 290km. Πρόκειται για ένα παθητικού τύπου σύστημα το οποίο μετράει την ηλιακή ακτινοβολία που έχει ανακλαστεί από τη Γη. Η ακτινοβολία αυτή συλλέγεται από ένα τηλεσκόπιο τριών κατόπτρων (M1, M2, M3) και εστιάζεται σε δύο συγκροτήματα εστιακού επιπέδου (Focal Plane Assemblies - FPA). Το πρώτο είναι για τις ορατές και τα κοντινά υπέρυρθρα μήκη κύματος (Visible and Near-Infrared - VNIR) και το δεύτερο είναι για τα υπέρυθρα βραχέα κύματα (Short Wave Infrared - SWIR). Στη συνέχεια πραγματοποιείται φασματικός διαχωρισμός σε ζώνες με χρήση κατάλληλων φίλτρων. [3]



Σχήμα 2.5: Διαχωρισμός της ηλιακής ακτινοβολίας στο σύστημα του MSI [3]

Ο MSI πραγματοποιεί ανάλυση σε 13 φασματικές ζώνες με χωρική ανάλυση 10, 20 ή 60 μέτρων. Αναλυτικότερα:

- 4 ζώνες έχουν χωρική ανάλυση 10 μέτρων. Αυτές είναι οι τρεις κλασικές ζώνες RGB (μπλε (493nm), πράσινη (560nm) και κόκκινη (665 nm)) και μια κοντινή υπέρυθρη (833 nm).
- 6 ζώνες έχουν χωρική ανάλυση 20 μέτρων. Αυτές είναι 4 VNIR ζώνες (704 nm, 740 nm, 783 nm και 865 nm), καθώς και 2 SWIR ζώνες (1610 nm και 2190nm) για εφαρμογές όπως ο εντοπισμός νεφών, πάγου και χιονιού και η μελέτη της υγρασίας της βλάστησης.
- 3 ζώνες έχουν χωρική ανάλυση 60 μέτρων. Είναι η ζώνη των αερολυμάτων (443 nm), η ζώνη των υδρατμών (945 nm) και η ζώνη των νεφών (1374 nm) και χρησιμοποιούνται κυρίως για εντοπισμό νεφών και ατμοσφαιρικές διορθώσεις. [63]

Τα δεδομένα αυτά περνούν από διάφορα στάδια επεξεργασίας προτού παραχθούν τα τελικά προϊόντα, ώστε να γίνουν διαθέσιμα στους χρήστες. Τα προϊόντα που είναι διαθέσιμα στους χρήστες είναι δύο, το Level-1C και το Level-2A. Το Level-1C παρέχει εικόνες

Αριθμός ζώνης	Μήκος Κύματος (nm)	Εύρος ζώνης(nm)	Χωρική Ανάλυση(m)
1	442	21	60
2	492	66	10
3	559	36	10
4	664	31	10
5	704	15	20
6	740	15	20
7	782	20	20
8	833	106	10
8a	864	21	20
9	945	20	60
10	1374	31	60
11	1613	91	20
12	2202	175	20

Πίνακας 2.2: Οι ζώνες που καλύπτει ο MSI

Top-of-Atmosphere σε χαρτογραφική γεωμετρία. Το Level-2A παρέχει εικόνες Bottom-of-Atmosphere και έχει προκύψει από αλγορίθμους εντοπισμού και ταξινόμησης των εικονοστοιχείων του εδάφους, του νερού και των νεφών, το αποτέλεσμα των οποίων έχει συμπεριληφθεί στα δεδομένα. Τα δεδομένα και των δύο αυτών προϊόντων είναι οργανωμένα σε μωσαϊκό τετράγωνων εικόνων (tiles), η κάθε μια από τις οποίες αντιστοιχεί σε έκταση  $100km^2$ .

Για την παρούσα εργασία επιλέχθηκε η χρήση δεδομένων από το επίπεδο Level-2A, καθώς τα Bottom-Of-Atmosphere δεδομένα είναι σαφώς πιο κατάλληλα για τη συγκεκριμένη εφαρμογή.



# Νευρωνικά Δίκτυα στην 'Οραση Υπολογιστών

### 3.1 Εισαγωγή

Η όραση υπολογιστών είναι ο κλάδος της τεχνητής νοημοσύνης που ασχολείται με την ικανότητα των υπολογιστών να "βλέπουν", δηλαδή να εξάγουν πληροφορία από ψηφιακές εικόνες, βίντεο και άλλες οπτικές εισόδους και να κάνουν προτάσεις ή να λαμβάνουν αποφάσεις βάσει αυτής της πληροφορίας. [64] Η όραση υπολογιστών έχει ευρείες εφαρμογές σήμερα όπως η αναγνώριση παθήσεων μέσα από ιατρικές εικόνες, η σχεδίαση αυτοοδηγούμενων οχημάτων, η παρακολούθηση της υγείας των φυτών και των ζώων σε αγροτικές και κτηνοτροφικές μονάδες και ο ποιοτικός έλεγχος των παραγόμενων προϊόντων σε εργοστάσια. [65]

Τις τελευταίες δεκαετίες το ενδιαφέρον αυξήθηκε ιδιαίτερα για τις εφαρμογές της 'Ορασης Υπολογιστών, με αποτέλεσμα να δημιουργηθούν και να γίνουν διαθέσιμα πολλά σύνολα δεδομένων από εικόνες, με το πρώτο μεγάλης κλίμακας σύνολο που δημιουργήθηκε να είναι το ImageNet το 2009, που επικεντρώνεται στην αναγνώριση καθημερινών αντικειμένων. [66] Η μεγάλη επανάσταση στο χώρο της υπολογιστικής όρασης έγινε ωστόσο το 2012, όταν μια ομάδα από το Πανεπιστήμιο του Τορόντο δημιούργησε ένα Συνελικτικό Νευρωνικό Δίκτυο (Convolutional Neural Network - CNN), το AlexNet και το υπέβαλε στον διαγωνισμό αναγνώρισης αντικειμένων ImageNet Large-Scale Visual Recognition Challenge (ILSVRC). [67] Έγινε φανερό ότι τα Συνελικτικά Νευρωνικά Δίκτυα είχαν πολύ καλύτερες επιδόσεις από τα παραδοσιακά και έκτοτε έγιναν η βάση μιας πληθώρας αρχιτεκτονικών, που ασχολούνται με πολλά προβλήματα της 'Ορασης Υπολογιστών.

Σήμερα, τα κυριότερα από αυτά τα προβλήματα είναι:

- Ταξινόμηση Εικόνας (Image Classification): Σκοπός του προβλήματος είναι να αντιστοιχίσει μια εικόνα εξ' ολοκλήρου σε μία κλάση, προβλέποντας τι απεικονίζει. Δεν πραγματοποιείται σε επίπεδο εικονοστοιχείου
- Ανίχνευση Αντικειμένου (Object Detection): Σκοπός είναι, δεδομένης μιας εικόνας, να εντοπίσει τις κλάσεις των διαφόρων αντικειμένων πάνω στην εικόνα, καθώς και τις θέσεις τους. Όπως και η ταξινόμηση εικόνας, δεν πραγματοποιείται σε επίπεδο εικονοστοιχείου.
- Σημασιολογική κατάτμηση (Semantic Segmentation): Σκοπός είναι να αντιστοιχίσει όλα τα εικονοστοιχεία μιας εικόνας σε κλάσεις.

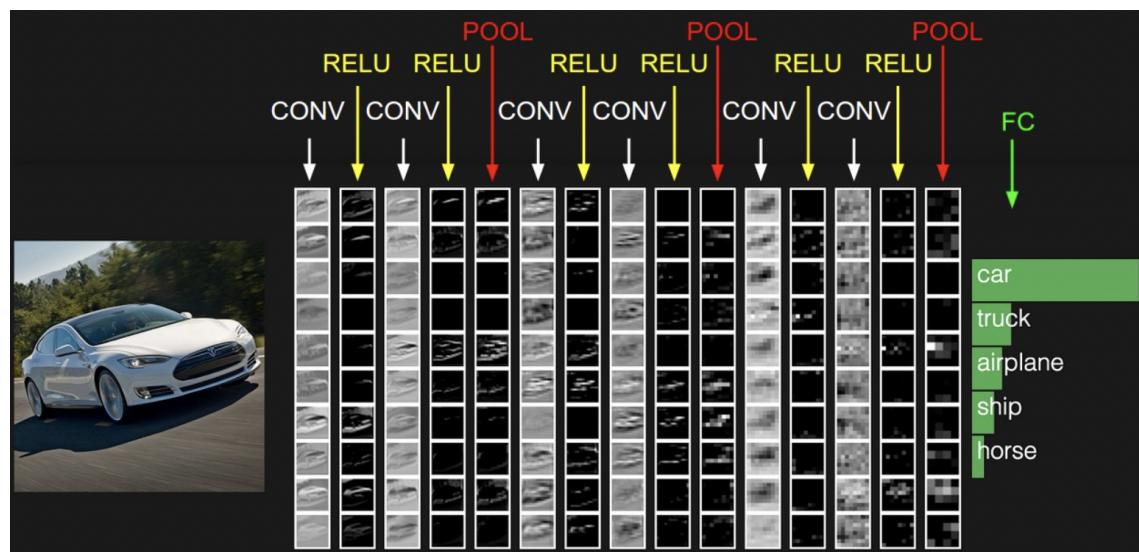
- Κατάτμηση Στιγμιοτύπου (Instance Segmentation): Σκοπός είναι, δεδομένης μιας εικόνας, να εντοπίσει τα διάφορα αντικείμενα που υπάρχουν σε μια εικόνα και να τα αντιστοιχίσει σε κλάσεις, σε επίπεδο εικονοστοιχείου.

Το πρόβλημα με το οποία ασχολείται η παρούσα εργασία ανήκει στην τρίτη κατηγορία. Συγκεκριμένα, μας ενδιαφέρει η σημασιολογική κατάτμηση δορυφορικών εικόνων με σκοπό την ταξινόμηση των εικονοστοιχείων τους σε δασική και μη δασική έκταση.

Παρακάτω θα αναλυθούν αρχικά κάποιες εισαγωγικές έννοιες σχετικά με τη δομή των Συνελικτικών Νευρωνικών Δικτύων και το πρόβλημα της ανίχνευσης αντικειμένων. Στη συνέχεια θα γίνει μία συνοπτική περιγραφή των αρχιτεκτονικών που χρησιμοποιούνται για κάθε ένα από τα προβλήματα της υπολογιστικής όρασης.

## 3.2 Συνελικτικά Νευρωνικά Δίκτυα (CNN)

Τα Συνελικτικά Νευρωνικά Δίκτυα είναι μια κατηγορία Νευρωνικών Δικτύων που δέχονται ως είσοδο εικόνες και εξάγει χαρακτηριστικά από αυτές. Χρησιμοποιούνται ευρέως στην επεξεργασία εικόνας πάνω σε προβλήματα ταξινόμησης (classification) και αναγνώρισης αντικειμένων (object detection), καθώς απαιτούν λιγότερη υπολογιστική ισχύ σε σχέση με άλλες μεθόδους ταξινόμησης και εξαγωγής χαρακτηριστικών, όπως οι ακμές. Η βασική δομή τους αποτελείται από τρία στρώματα, το συνελικτικό στρώμα (convolutional layer), το στρώμα ομαδοποίησης (pooling layer) και το πλήρως συνδεδεμένο στρώμα (fully-connected layer), αν και όπως θα δούμε παρακάτω, υπάρχουν και πιο περίπλοκες αρχιτεκτονικές.



**Σχήμα 3.1:** Παράδειγμα ενός Συνελικτικού Δικτύου για ταξινόμηση εικόνας. Με τα διαδοχικά συνελικτικά στρώματα μπορούν να υπολογιστούν τα διάφορα χαρακτηριστικά της εικόνας, ώστε να γίνει η σωστή ταξινόμηση από το πλήρως συνδεδεμένο στρώμα. [4]

Η προτίμηση των Συνελικτικών Δικτύων από τα παραδοσιακά Πλήρως Συνδεδεμένα Νευρωνικά Δίκτυα οφείλεται σε δύο κύριους λόγους. Ο πρώτος είναι ότι στα παραδοσιακά Νευρωνικά Δίκτυα όλοι οι νευρώνες ενός στρώματος είναι συνδεδεμένοι με όλους τους νευρώνες του επόμενου στρώματος. Αυτό σημαίνει ότι, όταν η είσοδος είναι μια εικόνα και μάλιστα

έχει πολλά κανάλια (RGB και πολυφασματικές εικόνες), ο αριθμός των συνδέσεων και κατ' επέπταση των παραμέτρων που πρέπει να εκπαιδευτούν αυξάνεται ραγδαία. Στα Συνελικτικά Δίκτυα, κάθε νευρώνας συνδέεται μόνο με συγκεκριμένους άλλους μέσω συνέλιξης με κατάλληλα φίλτρα. Τα φίλτρα αυτά έχουν τα ίδια βάρη για όλους τους νευρώνες του ίδιου επιπέδου και έχουν μικρότερες διαστάσεις από αυτές των εικόνων, με αποτέλεσμα ο αριθμός των παραμέτρων προς εκπαίδευση να είναι σημαντικά μικρότερος. Ο δεύτερος λόγος είναι ότι στα παραδοσιακά Νευρωνικά Δίκτυα η είσοδος δίνεται ως μονοδιάστατο διάνυσμα. Αυτό σημαίνει ότι στην περίπτωση των εικόνων δεν θα μπορεί να εκμεταλλευτεί η χωρική εξάρτηση των γειτονικών εικονοστοιχείων pixel. Αντιθέτως, ένα Συνελικτικό Νευρωνικό Δίκτυο, αφού δέχεται την εικόνα όπως είναι, είναι ικανό να “αντιληφθεί” την χρονική και χωρική εξάρτηση μεταξύ των εικονοστοιχείων με τη χρήση της συνέλιξης ανάμεσά τους. Με άλλα λόγια, μπορεί να εκπαιδευτεί για να κατανοήσει ολιστικά την εικόνα. Παρακάτω θα δούμε διάφορα επίπεδα των Συνελικτικών Δικτύων που θα μας απασχολήσουν στη συνέχεια.

### 3.2.1 Συνελικτικό Επίπεδο (Convolutional Layer)

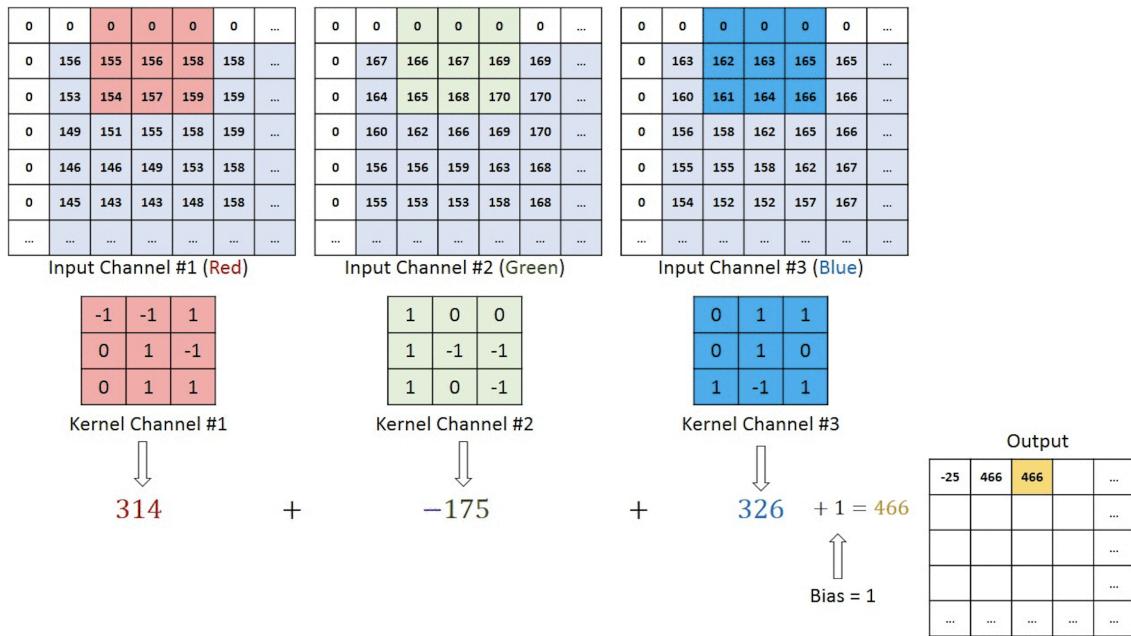
Το κάθε συνελικτικό επίπεδο αποτελείται από εκπαιδεύσιμους πυρήνες (kernels) ή αλιώς φίλτρα. Οι πυρήνες είναι πίνακες με μικρές διαστάσεις αλλά ίδιου βάθους (πλήθους καναλιών) με την εικόνα, για παράδειγμα, ένα τυπικό μέγεθος πυρήνα για μια εικόνα 3 καναλιών (RGB εικόνα) θα μπορούσε να είναι  $5 \times 5 \times 3$ . Καθώς η εικόνα εισέρχεται στο συνελικτικό επίπεδο, πραγματοποιείται η πράξη της συνέλιξης μεταξύ των εικονοστοιχείων της και των βαρών των πυρήνων. Η συνέλιξη μεταξύ δύο διακριτών δισδιάστατων σημάτων, στην περίπτωσή μας μεταξύ μιας εικόνας  $f$  και ενός πυρήνα  $g$  εκφράζεται μαθηματικά ως εξής:

$$(f * g)[x, y] = f[x, y] * g[x, y] = \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} f[n_1, n_2]g[x - n_1, y - n_2]$$

Δηλαδή, όπως μπορούμε να δούμε στην εικόνα 3.2, κάθε εικονοστοιχείο της εικόνας πολλαπλασιάζεται με τα αντίστοιχα των πυρήνων και το αποτέλεσμα κάθε συνέλιξης με έναν από τους πυρήνες αποτελεί έναν ξεχωριστό χάρτη ενεργοποίησης (activation map). Τα βάρη των πυρήνων είναι εκαπιδεύσιμες παράμετροι του δικτύου, οι οποίες εκπαιδεύονται για να υπολογίζουν χρήσιμα χαρακτηριστικά των εικόνων. Μερικές φορές, ενάντια στον ορισμό της συνέλιξης, κατά τον οποίο ο πυρήνας ολισθαίνει κατά ένα εικονοστοιχείο κατά την συνέλιξη, το βήμα ολισθησης (stride) μπορεί να οριστεί μεγαλύτερο. Επίσης, σε κάθε συνελικτικό επίπεδο έχουμε πολλαπλούς πυρήνες, καθένας από τους οποίους μπορεί να αντιστοιχεί σε ένα διαφορετικό χαρακτηριστικό.

### 3.2.2 Επίπεδο Ενεργοποίησης (Activation Layer)

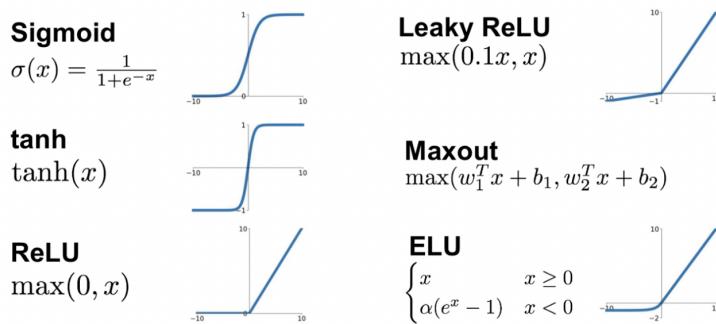
Αφού το συνελικτικό επίπεδο παράξει χάρτες ενεργοποίησης, το επίπεδο ενεργοποίησης καλείται να εισάγει μη γραμμικότητα στο δίκτυο, καθώς οι εικόνες που δέχεται και τα προβλήματα προς επίλυση πολλές φορές δεν ακολουθούν γραμμικότητα. Για αυτό εφαρμόζει στους χάρτες ενεργοποίησης μια μη γραμμική συνάρτηση. Μια δημοφιλής συνάρτηση ενεργοποίησης είναι η Rectified Linear Unit (ReLU), η οποία εκφράζεται ως εξής:



**Σχήμα 3.2:** Παράδειγμα της συνέλιξης μιας RGB εικόνας με έναν πυρήνα. Το αποτέλεσμα είναι ένας χάρτης ενεργοποίησης. [4]

$$ReLU(x) = \max(0, x)$$

Η RELU έχει αποδειχτεί ότι είναι λιγότερο απαιτητική υπολογιστικά από συναρτήσεις όπως η σιγμοειδής (sigmoid) και η υπερβολική εφαπτομένη (tanh) και συγκλίνει πιο γρήγορα.



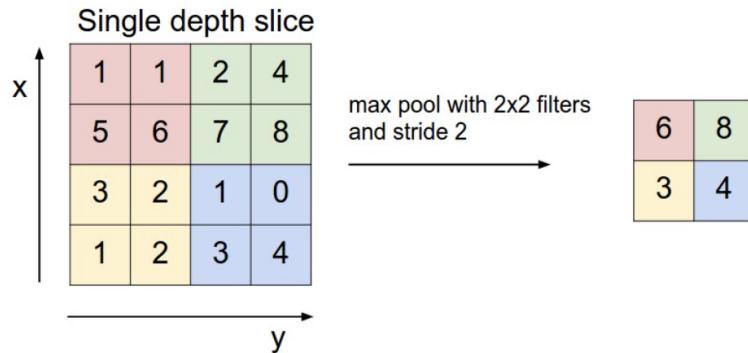
**Σχήμα 3.3:** Διάφορες συναρτήσεις ενεργοποίησης<sup>1</sup>

### 3.2.3 Επίπεδο Υποδειγματοληψίας (Pooling Layer)

Ο σκοπός των επιπέδων υποδειγματοληψίας είναι να μειώσουν τις διαστάσεις των παραγόμενων χαρτών ενεργοποίησης, αφού περάσουν από το επίπεδο ενεργοποίησης, με αποτέλεσμα να μειώσουν τον αριθμό των παραμέτρων προς εκπαίδευση και το υπολογιστικό κόστος και κατ' επέκταση να αποφευχθεί το φαινόμενο της υπερεκπαίδευσης (overfitting). Το επίπεδο αυτό λειτουργεί μεμονωμένα σε κάθε κανάλι του χάρτη ενεργοποίησης, αλλάζοντας

<sup>1</sup> <https://medium.com/dataseries/basic-overview-of-convolutional-neural-network-cnn-4fcc7dbb4f17>

τις διαστάσεις χρησιμοποιώντας τη συνάρτηση μεγίστου σε μη επικαλυπτόμενα τμήματα, οπότε και έχουμε το λεγόμενο Max-Pooling. Σε ορισμένες περιπτώσεις χρησιμοποιείται και η συνάρτηση του αριθμητικού μέσου όρου, ώστόσο η συνάρτηση μεγίστου έχει δείξει καλύτερα αποτελέσματα.



Σχήμα 3.4: Υποδειγματοληψία με τη συνάρτηση μεγίστου<sup>2</sup>

### 3.2.4 Επίπεδο Κανονικοποίησης Παρτίδας (Batch Normalization Layer)

Η εκπαίδευση βαθιών Νευρωνικών Δικτύων με πολλά επίπεδα μπορεί να γίνεται ασταθής, καθώς ανά παρτίδα (batch) η προσαρμογή των βαρών των αρχικών επιπέδων μπορεί να αλλάξει την κατανομή των βαρών των μετέπειτα επιπέδων. Αυτό το φαινόμενο ονομάζεται στην βιβλιογραφία internal covariate shift [68] και η αντιμετώπισή του επιταχύνει την εκπαίδευση και αυξάνει τις πιθανότητες σύγκλισης του δικτύου. Για να επιτευχθεί αυτό συνηθίζεται να προστίθενται επίπεδα κανονικοποίησης παρτίδας, τα οποία κανονικοποιούν την είσοδο, ώστε να ακολουθεί μια πιο σταθερή κατανομή. Η κανονικοποίηση γίνεται έτσι ώστε ο μέσος όρος της παρτίδας να είναι 0 και η διακύμανση μοναδιαία.

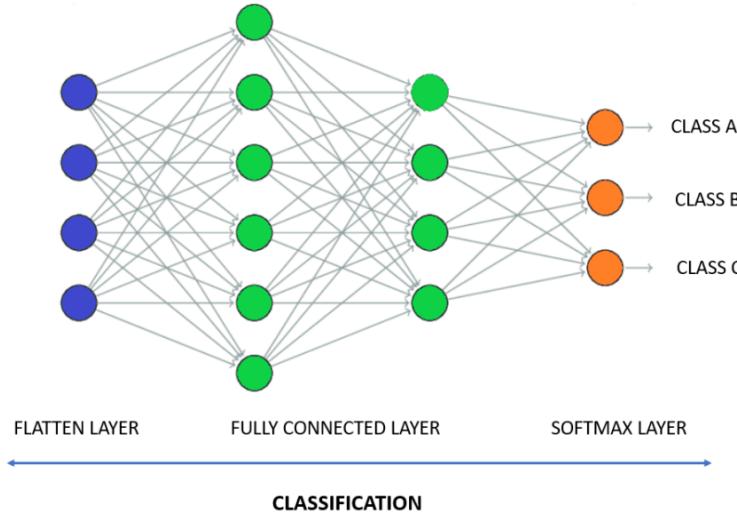
### 3.2.5 Πλήρως Συνδεδεμένο Επίπεδο (Fully-Connected Layer)

Τα Πλήρως Συνδεδεμένα Επίπεδα είναι τα τελευταία επίπεδα ενός Συνελικτικού Νευρωνικού Δικτύου. Αφού η εικόνα έχει περάσει από όλες τις επαναλήψεις των παραπάνω επιπέδων, περνάει από το επίπεδο Flatten, το οποίο μετατρέπει τους χάρτες ενεργοποίησης σε μονοδιάστατο διάνυσμα. Υστερα περνάει από μια σειρά Πλήρως Συνδεδεμένων Επιπέδων, τα οποία εφαρμόζουν πάνω στο διάνυσμα αφινικούς μετασχηματισμούς (affine transformations) και μη γραμμικές συναρτήσεις. Στα δίκτυα ταξινόμησης, το τελευταίο Επίπεδο χρησιμοποιεί συνήθως την συνάρτηση Softmax, η οποία υπολογίζει την πιθανότητα η εικόνα της εισόδου να αντιστοιχεί σε κάθε κλάση:

$$\sigma(x_i) = \frac{e^{x_i}}{\sum_j x_j}$$

<sup>2</sup><https://medium.com/dataseries/basic-overview-of-convolutional-neural-network-cnn-4fcc7dbb4f17>

<sup>3</sup><https://indiantechwarrior.com/fully-connected-layers-in-convolutional-neural-networks>

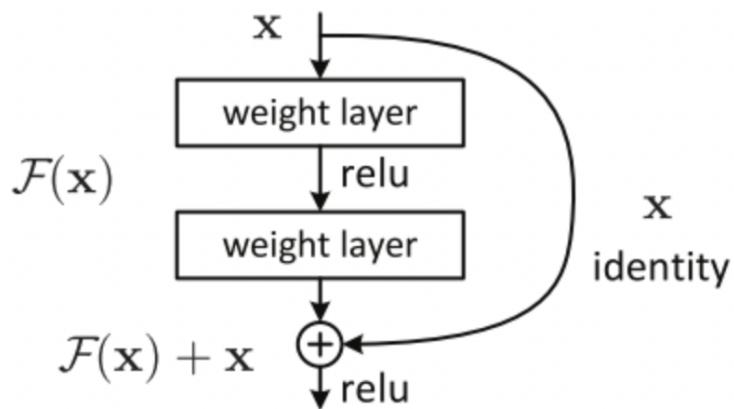


Σχήμα 3.5: Τα τελευταία επίπεδα ενός Συνελικτικού Δικτύου ταξινόμησης<sup>3</sup>

### 3.2.6 Residual Blocks

Τα residual blocks είναι μια αρχιτεκτονική που θα μας απασχολήσει με τα μοντέλα που υλοποιούνται σε αυτή την διπλωματική. Εισήχθησαν το 2015 με τα μοντέλα ResNet και από τότε χρησιμοποιούνται κατά κόρον σε νευρωνικά δίκτυα αναγνώρισης αντικειμένων και ταξινόμησης εικόνας.

Η κύρια ιδέα του residual block είναι η προσθήκη μιας σύνδεσης, η λεγόμενη skip connection, η οποία παρακάμπτει μια σειρά από συνελικτικά επίπεδα. Έτσι, ένα αντίγραφο της εισόδου περνάει μέσα από τα συνελικτικά επίπεδα και ένα αντίγραφο της εισόδου περνάει μέσα από το skip connection και έρχεται να προστεθεί στην έξοδο των συνελικτικών δικτύων πριν προχωρήσει στα επόμενα επίπεδα.



Σχήμα 3.6: Σχηματική απεικόνιση ενός residual block<sup>4</sup>

Το μεγάλο πλεονέκτημα με τα residual blocks είναι η αντιμετώπιση του προβλήματος των vanishing gradients. Το πρόβλημα αυτό εμφανίζεται κυρίως σε βαθιά Νευρωνικά Δίκτυα, στα οποία κατά το backpropagation, όσο προχωράει η ενημέρωση των βαρών, οι μερικές

<sup>4</sup> <https://towardsdatascience.com/residual-blocks-building-blocks-of-resnet-fd90ca15d6ec>

παράγωγοι της συνάρτησης κόστους όταν μειώνονται συνεχώς μέχρι να φτάσουν στο μηδέν. Αυτό έχει ως αποτέλεσμα κάποια επίπεδα του νευρωνικού να μην εκπαιδευτούν ποτέ. Τα residual blocks περιορίζουν αυτό το πρόβλημα με τα skip connections, καθώς πλέον η πληροφορία ακολουθεί συντομότερες διαδρομές και η μερική παράγωγος του κάθε επιπέδου δεν θα έχει μειωθεί από όλα τα προηγούμενα επίπεδα. [69] Έτσι μπορούν να δημιουργηθούν πιο βαθιές και πολύπλοκες αρχιτεκτονικές Νευρωνικών Δικτύων χωρίς αυτό το πρόβλημα.

### 3.3 Υπερανάλυση Εικόνας (Image Super-Resolution)



Σχήμα 3.7: Το πρόβλημα της Υπερανάλυσης Εικόνας<sup>5</sup>

Η υπερανάλυση εικόνων αποτελεί ένα πρόβλημα της Όρασης Υπολογιστών με ευρείες εφαρμογές, καθώς οι υψηλής ανάλυσης εικόνες είναι απαραίτητες για παράδειγμα στις ιατρικές απεικονίσεις και στην τηλεπισκόπηση. Η ανάγκη για να αποκτηθούν υψηλότερης ανάλυσης εικόνες χωρίς να γίνουν σημαντικές αναβαθμίσεις στο υλικό (hardware), καθώς αυτό θα αύξανε πολύ το κόστος, έχει οδηγήσει τους ερευνητές στο να αναπτύξουν διάφορες αρχιτεκτονικές Νευρωνικών Δικτύων για την επίλυση αυτού του προβλήματος.

Το πρόβλημα της Υπερανάλυσης Εικόνας αναφέρεται στο να ανακατασκευαστεί μια εικόνα υψηλής ανάλυσης από μια εικόνα χαμηλής ανάλυσης. Γενικά, θεωρείται ότι αν έχουμε μια εικόνα χαμηλής ανάλυσης  $Y$ , αυτή έχει αλλοιωθεί από μια υψηλής ανάλυσης εικόνα  $X$  μέσω της ακόλουθης σχέσης:

$$Y = D(X, \theta_D)$$

όπου  $D$  είναι η συνάρτηση αλλοίωσης, ορισμένη από τις παραμέτρους της  $\theta_D$ . [70] Σε ένα πραγματικό σενάριο, οι παραμέτροι  $\theta_D$  είναι άγνωστοι και το μόνο που έχουμε είναι η εικόνα χαμηλής ανάλυσης  $Y$ . Για αυτό η Υπερανάλυση Εικόνας στοχεύει στο να δώσει μια καλή εκτίμηση μιας πιθανής εικόνας υψηλής ανάλυσης με το να εκτιμήσει την αντίστροφη συνάρτηση  $R$  της συνάρτησης αλλοίωσης  $D$ , δηλαδή να κατασκευάσει μια εικόνα  $\hat{X}$  με βάση τον τύπο:

<sup>5</sup><https://blog.paperspace.com/image-super-resolution/>

$$\hat{X} = R(Y, \partial_R)$$

Όπου  $\partial_R$  οι παράμετροι της συνάρτησης  $R$ .

Οι συναρτήσεις  $D$  και  $R$  είναι αντίστροφες μεταξύ τους. Στη βιβλιογραφία ορισμένοι ερευνητές προσεγγίζουν την αλλοίωση ως αποτέλεσμα θόλωσης, υποδειγματοληψίας και εισαγωγής θορύβου στην εικόνα. Αυτό εκφράζεται μαθηματικά ως εξής:

$$Y = SBX + n$$

όπου  $B$  και  $S$  υποδηλώνουν τις συναρτήσεις θόλωσης και υποδειγματοληψίας, αντίστοιχα. Γενικά, η θόλωση περιγράφεται ως αποτέλεσμα της συνέλιξης της εικόνας υψηλής ανάλυσης με έναν Γκαουσιανό πυρήνα. Το  $n$  αντιπροσωπεύει τον θόρυβο, ο οποίος συχνά θεωρείται πως είναι ο λευκός Γκαουσιανός θόρυβος. Αρκετοί ερευνητές υιοθετούν μια απλούστερη συνάρτηση αλλοίωσης, όπως τη δικυβική παρεμβολή (bicubic interpolation), ώστε να παράξουν μια εικόνα χαμηλής ανάλυσης από μια εικόνα υψηλής ανάλυσης.

Για τη σύγκριση και αξιολόγησή τους, οι μέθοδοι Υπερανάλυσης Εικόνας προπονούνται και ελέγχονται σε "συνθετικές" εικόνες χαμηλής ανάλυσης, δηλαδή εικόνες που έχουν προκύψει από εικόνες υψηλής ανάλυσης με κάποια συνάρτηση αλλοίωσης. Ωστόσο, παρότι οι υπάρχουσες μέθοδοι και ιδιαίτερα τα νευρωνικά δίκτυα βαθιάς μάθησης πετυχαίνουν ικανοποιητικές ικανότητες ανακατασκευής των συνθετικών εικόνων χαμηλής ανάλυσης, η ανακατασκευή των πραγματικών εικόνων είναι ένα πιο περίπλοκο πρόβλημα διότι η ποικιλομορφία τους δεν καθιστά επαρκείς κάποιες συγκεκριμένες συναρτήσεις αλλοίωσης, ώστε οι μέθοδοι Υπερανάλυσης Εικόνας να προσεγγίσουν τις αντίστροφές τους. [70]

Έχουν προταθεί διάφορα είδη μεθόδων Υπερανάλυσης Εικόνας. Οι πρώτες που δημιουργήθηκαν, οι λεγόμενες μέθοδοι μοντελοποίησης της αλλοίωσης (Degradation Modeling-based methods) προσπαθούν με Νευρωνικά Δίκτυα Βαθίας Μάθησης να εκτιμήσουν τη συνάρτηση θόλωσης και κατά συνέπεια να προσεγγίσουν την αντίστροφή της για να κατασκευάσουν μια εικόνα υψηλής ανάλυσης. Παραδείγμα αυτής της μεθόδου είναι το Deep Alternating Network (DAN) [71], το οποίο με μια σειρά εκτιμητών και restorers προσπαθεί να εκτιμήσει τον πυρήνα θόλωσης με βάση το πόσο καλά αποκαθιστά την εικόνα υψηλής ανάλυσης.

Άλλες μέθοδοι, οι μέθοδοι βασισμένες σε ζεύγη εικόνων (Image Pairs-based Methods) συλλέγουν πολλαπλές εκδοχές μιας εικόνας με διαφορετικές αναλύσεις, δημιουργώντας ζεύγη εικόνας χαμηλότερης ανάλυσης με εικόνα υψηλότερης ανάλυσης, ώστε να προσεγγίσουν την αντιστοιχία (mapping) μεταξύ τους. Ένα παράδειγμα αυτής της μεθόδου είναι το μοντέλο Component Divide-and-Conquer (CDC). [72] Το μοντέλο αυτό εκπαιδεύει τρία ξεχωριστά blocks, το καθένα από τα οποία προβλέπει ξεχωριστά χαρακτηριστικά, τα οποία είναι τα επίπεδα χαρακτηριστικά, οι αικμές και οι γωνίες της εικόνας υψηλής ανάλυσης. Τα χαρακτηριστικά αυτά συνδυάζονται για να προκύψει η τελική εικόνα.

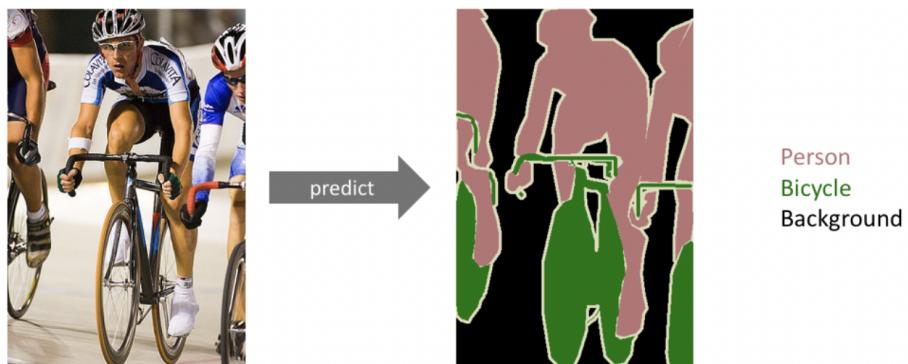
Άλλες μέθοδοι, οι μέθοδοι μετάφρασης χώρου (Domain Translation-based Methods) είναι σχεδιασμένες για σύνολα δεδομένων στα οποία λείπουν ζεύγη εικόνων υψηλής και χαμηλής ανάλυσης. Αυτές "μεταφράζουν" μια πραγματική εικόνα χαμηλής ανάλυσης σε μια συνθετική εικόνα χαμηλής ανάλυσης και ύστερα χρησιμοποιούν κάποιο προεκπαιδευμένο

Νευρωνικό Δίκτυο, τον γεννήτορα, για αναβάθμιση της ανάλυσης της συνθετικής εικόνας. Αυτή η κατασκευασμένη εικόνα υψηλής ανάλυσης, στις περισσότερες τέτοιες μεθόδους, τροφοδοτείται σε ένα άλλο Νευρωνικό Δίκτυο, τον διαχωριστή (discriminator), ο οποίος, υπολογίζει την πιθανότητα αυτή η εικόνα να είναι μια πραγματική εικόνα υψηλής ανάλυσης. Η πλειονότητα αυτών των Νευρωνικών Δικτύων ακολουθούν την αρχιτεκτονική των Αναγεννητικών Ανταγωνιστικών Δικτύων (Generative Adversarial Networks-GAN). [73] Στόχος αυτής της αρχιτεκτονικής είναι ο γεννήτορας να “ξεγελάσει” τον διαχωριστή, δηλαδή ο διαχωριστής να κατατάξει τις εικόνες που δημιουργήθηκαν ως πραγματικές, έτσι ώστε οι παραγόμενες εικόνες υψηλής ανάλυσης να ανταποκρίνονται όσο πιο πιστά γίνεται στην πραγματικότητα.

Το μοντέλο που θα χρησιμοποιήσουμε, το ESRGAN [10], ακολουθεί αυτήν την αρχιτεκτονική. Αποτελεί μια βελτίωση του μοντέλου SRGAN [9] και έχει πετύχει πολύ καλές επιδόσεις σε γνωστά benchmarks.

### 3.4 Σημασιολογική Κατάτμηση Εικόνας (Semantic Segmentation)

Το πρόβλημα της Σημασιολογικής Κατάτμησης Εικόνας περιγράφεται ως το πρόβλημα της ταξινόμησης κάθε εικονοστοιχείου μιας εικόνας σε μια κλάση. Δηλαδή η έξοδος ενός συστήματος Σημασιολογικής Κατάτμησης είναι μια εικόνα ίδιων διαστάσεων με την αρχική, στην οποία το κάθε εικονοστοιχείο ανήκει σε μια κλάση.



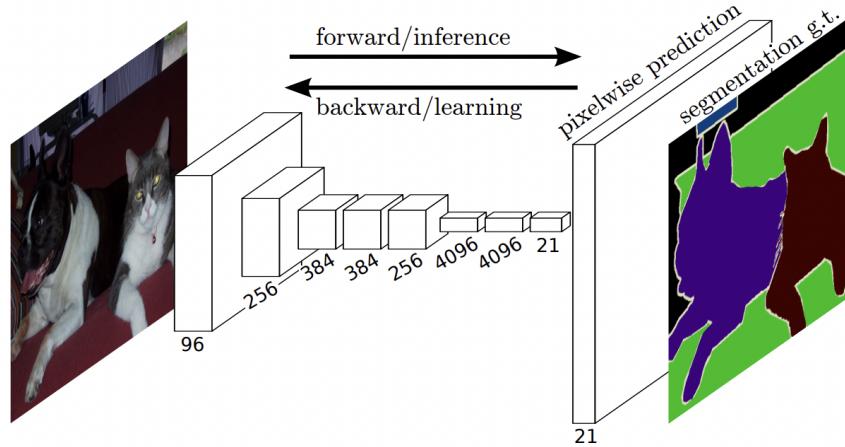
Σχήμα 3.8: Το πρόβλημα της Σημασιολογικής Κατάτμησης<sup>6</sup>

Μια αφελής προσέγγιση για την επίλυση αυτού του προβλήματος θα ήταν ένα δίκτυο, στο οποίο θα είχαμε πολλά διαδοχικά συνελικτικά επίπεδα, διατηρώντας τις διαστάσεις της εικόνας και στο τέλος θα παίρναμε έναν τελικό χάρτη ενεργοποίησης. Ωστόσο, το υπολογιστικό κόστος για τη διατήρηση της αρχικής ανάλυσης της εικόνας είναι δυσθεώρητα μεγάλο.

Για αυτό, από τις πρώτες αρχιτεκτονικές Νευρωνικών Δικτύων Βαθιάς Μάθησης που χρησιμοποιήθηκε για τη Σημασιολογική Κατάτμηση είναι το Πλήρως Συνελικτικό Δίκτυο - Fully Convolutional Network (FCN) [5]. Το FCN αποτελείται από πολλαπλά κρυφά συνελικτικά επίπεδα και επίπεδα υποδειγματοληψίας, των οποίων οι διαστάσεις μειώνονται όσο προχωράμε στο δίκτυο. Ο τελικός χάρτης τμηματοποίησης προκύπτει από ένα τελικό στρώμα

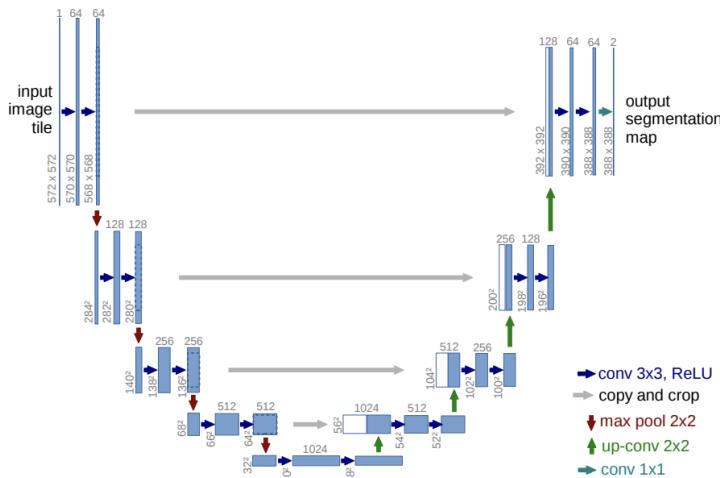
<sup>6</sup><https://www.jeremyjordan.me/semantic-segmentation/>

ταξινόμησης. Μετέπειτα αρχιτεκτονικές εφαρμόζουν πιο περίπλοκα στοιχεία, όπως τα skip connections που είδαμε στα residual blocks.



Σχήμα 3.9: To FCN σχηματικά [5]

Υστερα, αναπτύχθηκε η αρχιτεκτονική του κωδικοποιητή-αποκωδικοποιητή (encoder-decoder) με το μοντέλο UNet [6]. Σε αυτή την αρχιτεκτονική, η εικόνα πρώτα περνάει από τον κωδικοποιητή, ο οποίος αποτελεί ένα πλήρως συνελικτικό δίκτυο. Παράγει χαμηλών διαστάσεων χάρτες ενεργοποίησης και εκπαιδεύεται ώστε να εξάγει κύρια χαρακτηριστικά των διάφορων κλάσεων, διαφοροποιώντας αυτές τις κλάσεις μεταξύ τους. Υστερα, αυτοί οι χάρτες περνούν στον αποκωδικοποιητή, ένα πλήρως συνελικτικό δίκτυο το οποίο όμως υπερδειγματοληπτεί την είσοδό του, αυξάνοντας τις διαστάσεις των συνελικτικών του επιπέδων και προωθώντας "συμφραζόμενη" πληροφορία στα επόμενα επίπεδα. Τέλος, παράγει έναν τελικό χάρτη σημασιολογικής κατάτμησης, ίδιας ανάλυσης με την αρχική εικόνα. Η αρχιτεκτονική αυτή είναι πολύ δημοφιλής στη σημασιολογική κατάτμηση και αυτήν ακολουθούν τα μοντέλα ResUNet-a και DeepLabv3+ που θα χρησιμοποιήσουμε σε αυτή την εργασία.

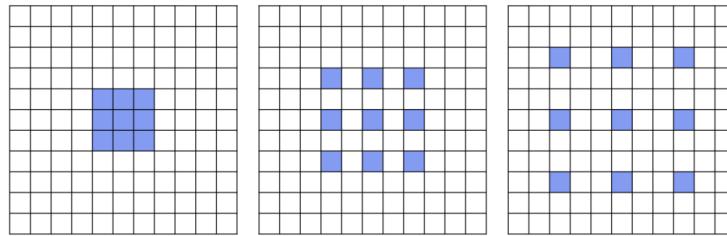


Σχήμα 3.10: Η αρχιτεκτονική κωδικοποιητή-αποκωδικοποιητή που συναντάται στο UNet. [6]

Μια επόμενη τεχνική που εφαρμόστηκε είναι η τεχνική της atrous συνέλιξης και των

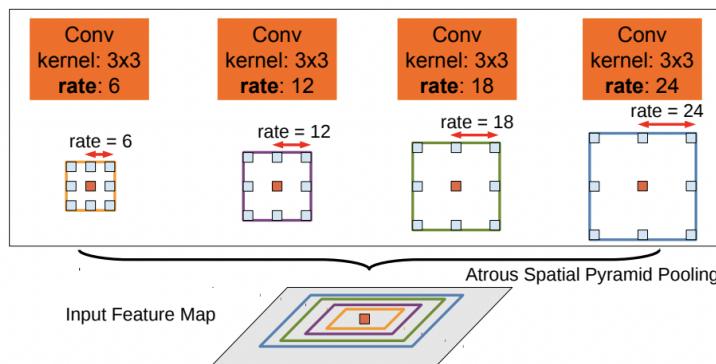
Atrous Spatial Pyramid Pooling [74]. Η atrous συνέλιξη είναι η συνέλιξη με έναν πυρήνα, οι μη μηδενικές τιμές του οποίου έχουν μεταξύ τους απόσταση έναν αριθμό εικονοστοιχείων ίσο με έναν ρυθμό διαστολής μειωμένο κατά 1. Μαθηματικά, η atrous συνέλιξη για μονοδιάστατο σήμα  $x$  και πυρήνα  $w$ , καθώς και ρυθμό διαστολής  $r$ , ορίζεται ως:

$$y[i] = \sum_k x[i + rk] \cdot w[k]$$



Σχήμα 3.11: Ο πυρήνας της atrous συνέλιξης, για ρυθμό διαστολής 1, 2 και 3 αντίστοιχα. [7]

Με το να αντικατασταθεί η κλασική συνέλιξη με την atrous συνέλιξη, κάθε εικονοστοιχείο στον χάρτη χαρακτηριστικών αποκτά μεγαλύτερο πεδίο θέασης field-of-view, δηλαδή συσχετίζεται με εικονοστοιχεία της εισόδου από τα οποία έχει μεγαλύτερη απόσταση. Αυτό αίρει τον περιορισμό της τοπικότητας που επιβάλλει η κλασική συνέλιξη και εξάγει χαρακτηριστικά σε μεγαλύτερη κλίμακα, χωρίς να εισάγει παραπάνω παραμέτρους εκπαίδευσης. Πάνω σε αυτή τη συνέλιξη δημιουργήθηκαν τα Atrous Spatial Pyramid modules. Τα modules αυτά πραγματοποιούν παράλληλες atrous συνελίξεις με διαφορετικούς ρυθμούς διαστολής και συνδυάζει τα αποτελέσματά τους, με σκοπό να εξάγουν χαρακτηριστικά σε πολλαπλές κλίμακες. Χρησιμοποιούνται σε μοντέλα όπως το DeepLab [8] και το DeepLabv3 [75].



Σχήμα 3.12: Η λειτουργία του Atrous Spatial Pyramid Module. Τα διάφορα field-of-views του εικονοστοιχείου είναι τα χρωματιστά πλαίσια στον κάτω χάρτη ενεργοποίησης. [8]

Τα τελευταία χρόνια, έχουν εισαχθεί στον χώρο της υπολογιστικής όρασης και της σημασιολογικής κατάτμησης οι μετασχηματιστές (Transformers) [76]. Οι μετασχηματιστές σχεδιάστηκαν αρχικά στα πλαίσια της Επεξεργασίας Κειμένου και Φυσικής Γλώσσας για το έργο της μετάφρασης γλωσσών. Ωστόσο, λόγω της ικανότητάς τους να εντοπίζουν μακριές εξαρτήσεις της εισόδου, έχουν αναπτυχθεί αρχιτεκτονικές μετασχηματιστών και για την επεξεργασία εικόνων, με σκοπό την αποτελεσματικότερη εξαγωγή χαρακτηριστικών τους.

Πράγματι, πολλά σύγχρονα μοντέλα πετυχαίνουν state-of-the-art αποτελέσματα χρησιμοποιώντας ως βάση μετασχηματιστές [77] [78] [79] και μια τέτοια δημοφιλής αρχεκτονική είναι ο Swin Transformer [13], τον οποίο θα χρησιμοποιήσουμε στην παρούσα εργασία.

## Μέρος **III**

### Προετοιμασία



## Κεφάλαιο 4

### Δημιουργία Συνόλου Δεδομένων

Στο κεφάλαιο αυτό παρουσιάζονται τα εργαλεία και οι μέθοδοι με τις οποίες δημιουργήθηκε το σύνολο δεδομένων που χρησιμοποιήθηκε στην παρούσα εργασία. Ήταν απαραίτητη η δημιουργία ενός συνόλου δεδομένων με επαρκή αριθμό δορυφορικών εικόνων, η κάθε μία εκ των οποίων θα συνοδεύεται από την επιθυμητή έξοδο ή ετικέτα (ground truth) των σημείων που αντιστοιχούν σε κάποια δασική έκταση. Στις επόμενες ενότητες συζητούνται θέματα που σχετίζονται με τη συγκέντρωση των δορυφορικών εικόνων, την προεπεξεργασία τους και την συγκέντρωση των ετικετών τους.

#### 4.1 Συγκέντρωση δορυφορικών εικόνων

Όπως αναφέρθηκε στο κεφάλαιο 2, σε αυτή την εργασία συγκεντρώθηκαν δεδομένα από την αποστολή Sentinel-2 του προγράμματος Copernicus. Τα δεδομένα αυτά, εκτός από την απευθείας λήψη τους από τη σελίδα του προγράμματος Copernicus (<https://scihub.copernicus.eu/>), φιλοξενούνται από εταιρείες που παρέχουν υποδομές νέφους (cloud infrastructure) όπως η Amazon (Amazon Web Services - AWS) και η Google (Google Cloud Platform - GCP). Οι εταιρείες αυτές επιτρέπουν την ανάλυση και την επεξεργασία του μεγάλου όγκου αυτών των δεδομένων, στα πλαίσια των υπηρεσιών τους, χωρίς να απαιτείται λήψη από το χρήστη. Αυτό διευκολύνει την ανάπτυξη γεωχωρικών εφαρμογών, καθώς η επεξεργασία πολύ μεγάλων όγκων δεδομένων, που απαιτεί μεγάλη υπολογιστική ισχύ, πραγματοποιείται στις υποδομές των εταιρειών αυτών. Το Google Earth Engine, το οποίο χρησιμοποιείται σε αυτή την εργασία, είναι ένα εργαλείο ανάλυσης και λήψης δορυφορικών εικόνων, που διαθέτει μια Διασύνδεση Προγραμματιστικών Εφαρμογών (Application Programming Interface - API) για την επεξεργασία τους.

##### 4.1.1 Google Earth Engine

Το Google Earth Engine είναι μια πλατφόρμα επιστημονικής ανάλυσης και οπτικοποίησης γεωχωρικών συνόλων δεδομένων για ακαδημαϊκούς, μη κερδοσκοπικούς ή και κυβερνητικούς σκοπούς. Φιλοξενεί δορυφορικές εικόνες και τις διατηρεί σε ένα δημόσιο αρχείο που περιέχει εικόνες της Γης από περισσότερα από 40 έτη πίσω. Οι εικόνες ανανεώνονται καθημερινά και γίνονται άμεσα διαθέσιμες για ανάλυση και επεξεργασία. Παράλληλα, διαθέτει και ένα εύχρηστο API, το οποίο επιτρέπει την ανάλυση και άντληση δεδομένων από σύνολα

μεγάλης κλίμακας. Για την πρόσθαση στο Google Earth Engine χρειάζεται να γίνει αίτημα πρόσθασης στο <https://signup.earthengine.google.com/>, το οποίο περιέχει τα στοιχεία του χρήστη και τον σκοπό του project. Η επεξεργασία του αιτήματος γίνεται άμεσα και η χρήση της υπηρεσίας είναι ελεύθερη.

#### 4.1.2 Προεπεξεργασία και λήψη των εικόνων

Για τις ανάγκες του συνόλου δεδομένων που θέλουμε να δημιουργήσουμε, είναι απαραίτητη μία δορυφορική εικόνα ανά περιοχή. Ωστόσο, αν απλά πάρουμε μια εικόνα από μια περιοχή για μια συγκεκριμένη ημερομηνία, ενδέχεται αυτή η εικόνα να μην είναι κατάλληλη για την εφαρμογή μας για τους εξής λόγους. Πρώτον, είναι πολύ πιθανό τη στιγμή της λήψης της δορυφορικής εικόνας να παρεμβάλλονται σύννεφα, τα οποία προφανώς συνεπάγονται απώλεια πληροφορίας. Δεύτερον, για λόγους που σχετίζονται με τη λειτουργία του αισθητήρα MSI και την τροχιά του δορυφόρου, κάποιες εικόνες δεν είναι ολοκληρωμένες και υπάρχουν περιοχές που λείπουν.

Για αυτό, συγκεντρώνουμε για κάθε περιοχή ένα σύνολο εικόνων που λήφθηκαν σε ένα επαρκές χρονικό διάστημα, και στη συνέχεια κατεβάζουμε ως τελική εικόνα τη διάμεσο (median) όλων των εικόνων του συνόλου αυτού. Η επιλογή αρκούντως μεγάλου χρονικού διαστήματος διασφαλίζει ότι η τελική εικόνα θα είναι αντιπροσωπευτική της περιοχής. Συγκεκριμένα, επιλέχθηκε το χρονικό διάστημα από την 01-06-2018 έως την 01-09-2018. Επιλέχθηκαν θερινοί μήνες, διότι είναι λιγότερο πιθανό έντονα καιρικά φαινόμενα να επηρεάσουν την ποιότητα των εικόνων.

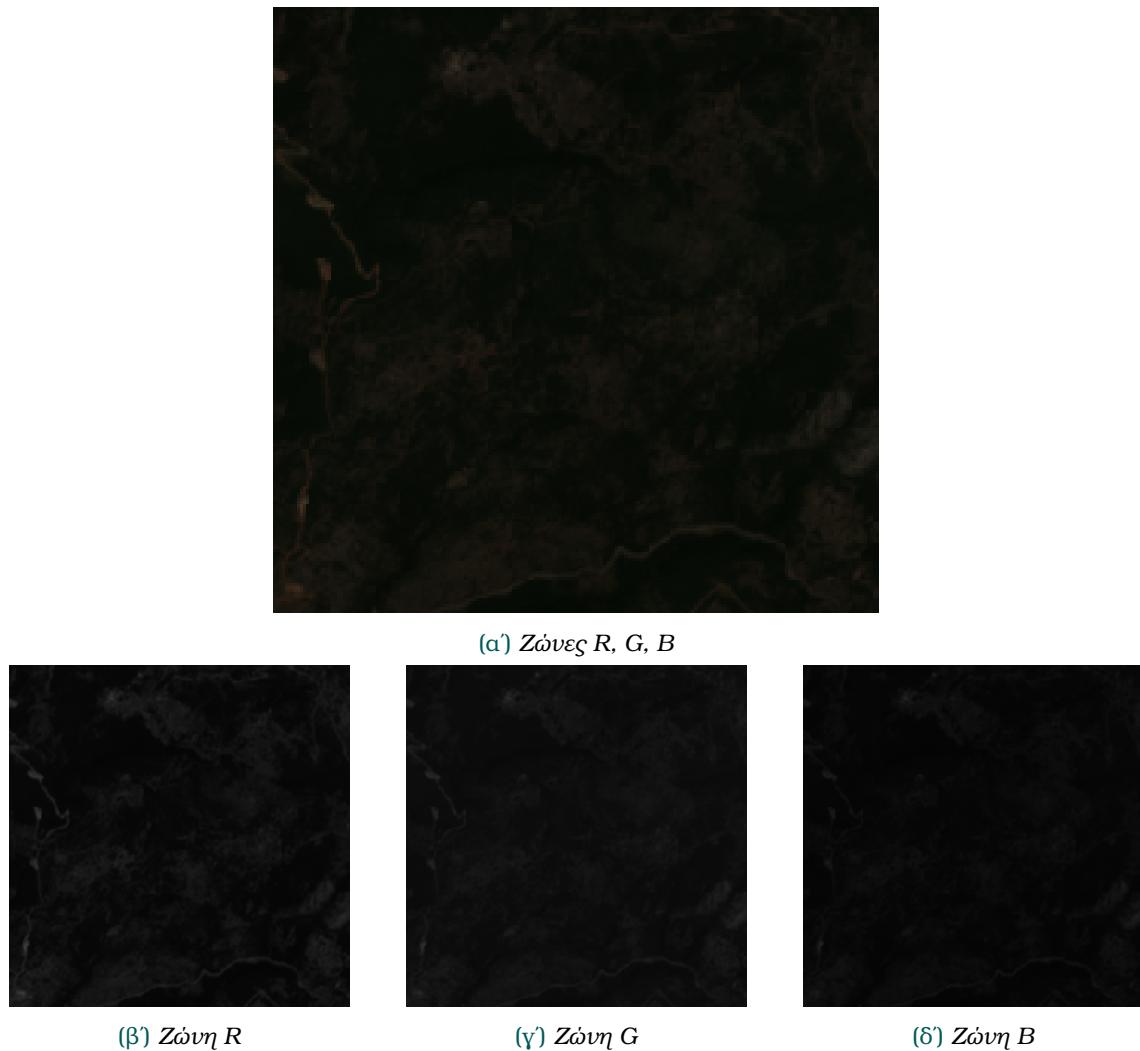
Η λήψη των εικόνων πραγματοποιήθηκε με χρήση του API του Google Earth Engine, το οποίο παρέχεται στη γλώσσα Python από το Python Package Index (PIP). Το API αυτό επιτρέπει την επιλογή όλων των εικόνων για μια συγκεκριμένη περιοχή ανά χρονική περίοδο, καθώς και την επεξεργασία αυτών των εικόνων πριν την τελική λήψη, όπως φιλτράρισμα αναλόγως των ιδιοτήτων των εικόνων και επιλογή συγκεκριμένων φασματικών ζωνών.

Με κατάλληλο script γραμμένο σε γλώσσα Python, μπορέσαμε να συγκεντρώσουμε τις εικόνες από όλες τις περιοχές δασικών εκτάσεων της Πελοποννήσου. Το εργαλείο αυτό δέχεται ως ορίσματα:

- Το πλαίσιο οριοθέτησης (bounding box), δηλαδή τα ελάχιστα και μέγιστα γεωγραφικά μήκη και πλάτη που θέλουμε να καλύπτει η εικόνα
- Το χρονικό διάστημα, δηλαδή την αρχική και τελική ημερομηνία λήψης της εικόνας
- Το όνομα που θα δοθεί στην εικόνα.

To script αυτό εκτελεί την παρακάτω διαδικασία για να παράξει μια τελική εικόνα ανά περιοχή:

- Με βάση τα ορίσματα, επιλέγεται από το Google Earth Engine API μια συλλογή εικόνων με όλες τις διαθέσιμες εικόνες για την περιοχή, για την συγκεκριμένη χρονική περίοδο.
- Οι εικόνες του Sentinel-1C περιέχουν, μεταξύ άλλων, στα μεταδεδομένα τους, την ιδιότητα cloudy\_pixel\_percentage, η οποία περιγράφει το ποσοστό των εικονοστοιχείων



Σχήμα 4.1: Ζώνες της ίδιας δορυφορικής εικόνας

της εικόνας που αντιστοιχούν σε σύννεφα. Με την ιδιότητα αυτή φιλτράρονται οι εικόνες ώστε να παραμέινουν μόνο αυτές με ποσοστό άνω του 20%.

- Στη συνέχεια, επιλέγονται οι φασματικές ζώνες της δορυφορικής εικόνας. Στην συγκεκριμένη εφαρμογή θα χρησιμοποιήσουμε μόνο τις 3 ζώνες του ορατού φάσματος (R, G, B).
  - Τέλος, υπολογίζεται η διάμεσος της συλλογής εικόνων.

## 4.2 Συγκέντρωση Ετικετών

Μετά την λήψη των δορυφορικών εικόνων πρέπει να συλλέξουμε τις απαραίτητες ετικέτες, δηλαδή αυτές που προσδιορίζουν τη θέση των δασικών εκτάσεων σε κάθε εικόνα. Για την εφαρμογή αυτή αντλήσαμε δεδομένα από το OpenStreetMap, τα οποία αν και δεν είναι πλήρη και περιέχουν ανακρίβειες σε κάποιο βαθμό, είναι επαρκή για να εκπαιδευτούν τα Νευρωνικά Δίκτυα, οδηγώντας αργότερα σε μια πληρέστερη βάση δεδομένων.

### 4.2.1 OpenStreetMap

Το OpenStreetMap ξεκίνησε το 2004 από το Ηνωμένο Βασίλειο και στόχος του είναι η δημιουργία ενός δωρεάν προσθάσιμου και επεξεργάσιμου χάρτη όλου του κόσμου. Τα δεδομένα σε αυτόν τον χάρτη προέρχονται ως επί το πλείστον από εθελοντές, επεκτείνονται και ενημερώνονται καθημερινά, καλύπτοντας όλο τον πλανήτη. Επίσης, περιλαμβάνουν όχι μόνο φυσικά χαρακτηριστικά, όπως δάση και λίμνες, αλλά και χαρακτηριστικά που σχετίζονται με την ανθρώπινη δραστηριότητα, όπως δρόμοι και κατοικημένες περιοχές.

Για την άντληση δεδομένων στην βάση του OpenStreetMap έχει αναπτυχθεί η διεπαφή Overpass API, στην οποία ο χρήστης στέλνει αίτημα (request) με τα δεδομένα που θέλει να αποκτήσει, και εκείνη του τα επιστρέφει σε μορφή λίστας.

### 4.2.2 Λήψη Ετικετών

Αποφασίστηκε να ληφθούν όλες οι διαθέσιμες δασικές εκτάσεις στην Πελοπόννησο, καθώς εξασφαλίζουμε επαρκή όγκο δεδομένων για την εκπαίδευση των Νευρωνικών Δικτύων. Στην περίπτωση που διαλέγαμε μεγαλύτερη έκταση, για παράδειγμα όλη την Ελλάδα, ο όγκος των δεδομένων θα ήταν απαγορευτικά μεγάλος. Για αυτό, με ένα script που χρησιμοποιεί το overpy, τον python wrapper του Overpass API, στάλθηκε το παρακάτω αίτημα για τη συλλογή των δασικών εκτάσεων της Πελοποννήσου. Στο σχήμα 4.3 φαίνεται η έξοδος της διαδικτυακής επαφής του Overpass API, Overpass Turbo, για το ίδιο αίτημα.

```

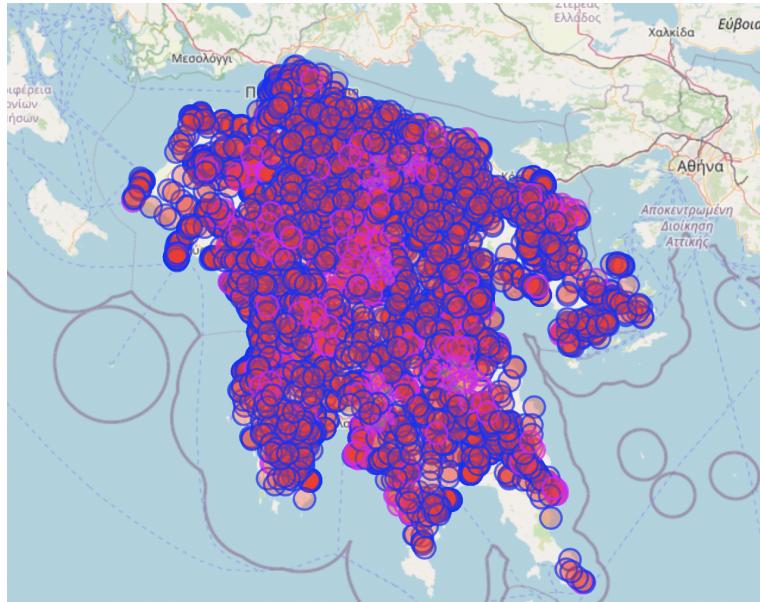
1 [out:json][timeout:1000];
2 area["name:en"]="Peloponnese"]->.searchArea;
3 (
4     way["natural"]="wood"](.area.searchArea);
5     relation["natural"]="wood"](.area.searchArea);
6 );
7 out body;
8 >;
9 out body;
```

Σχήμα 4.2: Το αίτημα στο Overpass API.

Η κάθε δασική έκταση αναπαρίσταται ως ένα πολύγωνο (polygon), οι κορυφές του οποίου αποτελούν τις συντεταγμένες του περιγράμματός της. Τα πολύγωνα αυτά καταγράφονται σε ένα αρχείο .GeoJSON. Στο σχήμα 4.4 φαίνονται κάποια παραδείγματα δασικών εκτάσεων με την αντίστοιχη ετικέτα τους.

## 4.3 Οργάνωση Δεδομένων

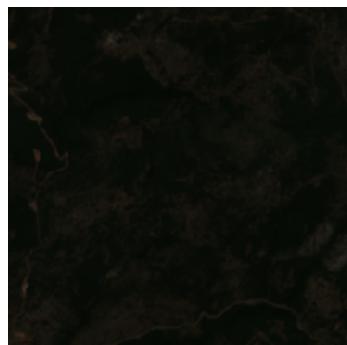
Αποφασίστηκε οι διαστάσεις των εικόνων με τις οποίες θα εκπαιδεύσουμε τα Νευρωνικά Δίκτυα να είναι  $200 \times 200$ , λόγω των μεγάλων απαιτήσεων σε μνήμη που θα είχαμε σε μεγαλύτερες διαστάσεις. Μιας και η χωρική ανάλυση των ζωνών που επιλέξαμε είναι 10m,



*Σχήμα 4.3: Το αποτέλεσμα στο Overpass Turbo.*

η κάθε εικόνα αντιστοιχεί σε μια έκταση  $2km \times 2km$ . Για να κατατμήσουμε αυτές τις εικόνες, χρειαζόμαστε ένα πλέγμα (grid), το οποίο χωρίζει την επιφάνεια της Πελοποννήσου σε περιοχές  $2km \times 2km$  και στη συνέχεια να κρατήσουμε μόνο αυτές που περιέχουν κάποια δασική έκταση. Με ένα script σε Python κατασκευάζουμε αυτό το πλέγμα και στη συνέχεια, εκμεταλλευόμενοι το αρχείο .GeoJSON, το οποίο περιλαμβάνει τις περιοχές όλων των δασικών εκτάσεων της Πελοποννήσου, κρατάμε μόνο τα πλαίσια που περιέχουν τις δασικές αυτές εκτάσεις.

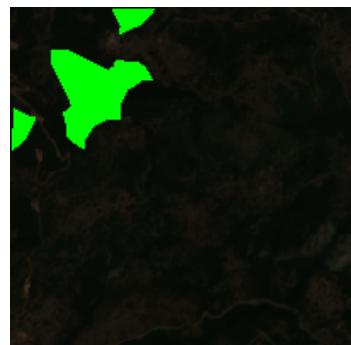
Εντέλει το σύνολο δεδομένων μας αποτελείται από 3204 εικόνες δασικών εκτάσεων με τις επικέτες τους.



(α) Δορυφορική εικόνα



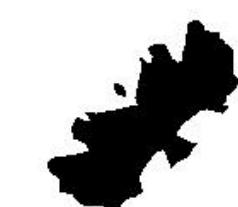
(β) Ετικέτες OSM



(γ) Τελική εικόνα με ετικέτες



(δ) Δορυφορική εικόνα



(ε) Ετικέτες OSM



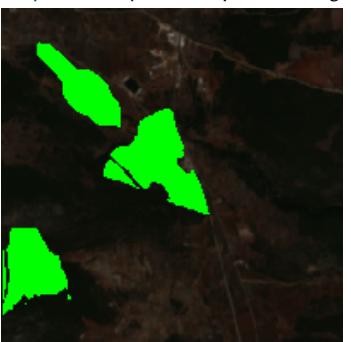
(ζ) Τελική εικόνα με ετικέτες



(η) Δορυφορική εικόνα



(η) Ετικέτες OSM



(θ) Τελική εικόνα με ετικέτες

Σχήμα 4.4: Παραδείγματα δορυφορικών εικόνων με τις ετικέτες τους.

## Μέρος

### Πειραματικό μέρος



## Κεφάλαιο 5

### Υλοποίηση

---

Σε αυτό το κεφάλαιο παρουσιάζονται οι αρχιτεκτονικές των νευρωνικών δικτύων που χρησιμοποιήθηκαν σε αυτή την εργασία, οι συναρτήσεις σφάλματος και οι μετρικές αξιολόγησης, καθώς και η μεθοδολογία που εφαρμόστηκε στην πειραματική διαδικασία.

#### 5.1 Υπερανάλυση Εικόνας (Image Super-Resolution)

Στην παρούσα διπλωματική εξετάζουμε την επίδραση της Υπερανάλυσης Εικόνας στην απόδοση των μοντέλων και για αυτό χρησιμοποιούμε το μοντέλο ESRGAN για την επαύξηση των δορυφορικών εικόνων.

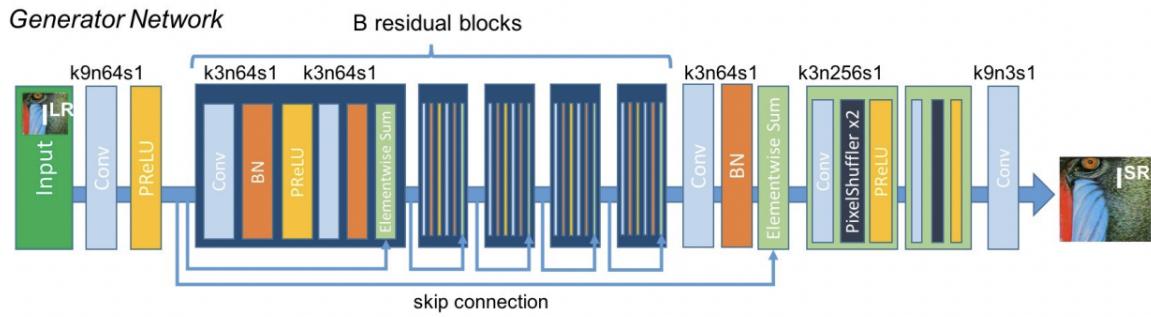
##### 5.1.1 SRGAN

Καθώς το ESRGAN αποτελεί βελτίωση του SRGAN, είναι απαραίτητο να αναφερθούμε στο SRGAN και στην ιδέα πίσω από αυτό.

Όπως αναφέρθηκε, το SRGAN (Super Resolution Generative Adversarial Network) ακολουθεί την αρχιτεκτονική των Αναγεννητικών Ανταγωνιστικών Δικτύων (Generative Adversarial Networks-GAN). Στόχος είναι να εκπαιδευτεί μια συνάρτηση γεννήτορα (generating function)  $G$ , η οποία με είσοδο μια εικόνα χαμηλής ανάλυσης θα προσπαθήσει να κατασκευάσει ένα αντίγραφο της εικόνας σε υψηλότερη ανάλυση. Αυτή η συνάρτηση αποτελεί ένα βαθύ Συνελικτικό Νευρωνικό Δίκτυο. Το αντίγραφο αυτό τροφοδοτείται σε ένα άλλο Νευρωνικό Δίκτυο, τον διαχωριστή (discriminator), ο οποίος προσπαθεί να διαχωρίσει τις πραγματικές εικόνες υψηλής ανάλυσης από τις εικόνες που δημιούργησε ο γεννήτορας. Στόχος του γεννήτορα λοιπόν είναι να “ζεγελάσει” τον διαχωριστή, δηλαδή ο διαχωριστής να κατατάξει τις εικόνες που δημιουργήθηκαν ως πραγματικές, έτοι ώστε οι παραγόμενες εικόνες υψηλής ανάλυσης να ανταποκρίνονται όσο πιο πιστά γίνεται στην πραγματικότητα. [9]

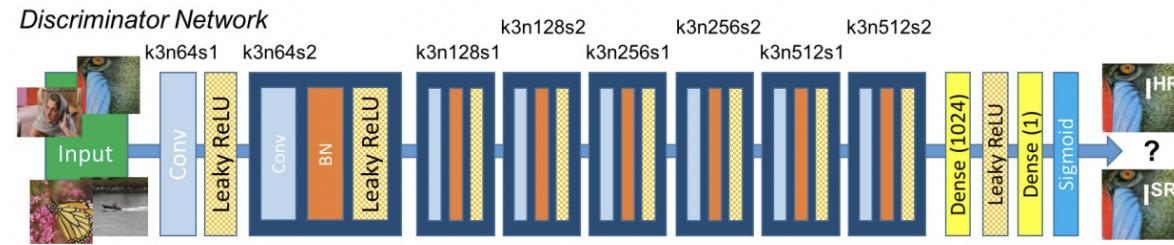
Αναλυτικότερα, το κύριο κομμάτι που απαρτίζει το Νευρωνικό Δίκτυο του γεννήτορα είναι τα residual blocks, καθένα από τα οποία αποτελείται από δύο συνελικτικά στρώματα, στρώματα για κανονικοποίηση παρτίδας (batch normalization) και συνάρτηση ενεργοποίησης την Parametric RELU.

Το Νευρωνικό Δίκτυο του διαχωριστή αποτελείται από οκτώ συνελικτικά στρώματα με συνάρτηση ενεργοποίησης τη LeakyRELU, η δομή των οποίων ακολουθεί αυτή που συναντάται στο Νευρωνικό Δίκτυο VGG για αναγνώριση αντικειμένων. Έπειτα ακολουθούν δύο πυκνά



Σχήμα 5.1: Η δομή του γεννήτορα του SRGAN [9]

στρώματα και στο τέλος η σιγμοειδής συνάρτηση υπολογίζει την πιθανότητα για το κατά πόσο η εικόνα που δόθηκε στην είσοδο είναι πραγματική ή όχι.



Σχήμα 5.2: Η δομή του διαχωριστή του SRGAN [9]

Για την εκπαίδευση του δικτύου του γεννήτορα χρησιμοποιείται μια συνάρτηση κόστους ονομαζόμενη ως αντιληπτικό κόστος (perceptual loss)  $l^{SR}$ , το οποίο αποτελεί το σταθμισμένο άθροισμα (weighted sum) δύο επιμέρους συναρτήσεων κόστους, του κόστους περιεχομένου (content loss)  $l_X^{SR}$  και του ανταγωνιστικού κόστους  $l_{Gen}^{SR}$  ως εξής:

$$l^{SR} = l_X^{SR} + 10^{-3} l_{Gen}^{SR}$$

Για το κόστος περιεχομένου προτιμάται στη βιβλιογραφία [9] το κόστος VGG. Αν θεωρήσουμε ως  $\phi_{ij}$  τον χάρτη χαρακτηριστικών (feature map) που προκύπτει μετά την  $j$ -οστή συνέλιξη (μετά το επίπεδο ενεργοποίησης) και πριν το  $i$ -οστό επίπεδο υποδειγματοληψίας, τότε ορίζουμε το κόστος VGG ως την ευκλείδια απόσταση μεταξύ του χάρτη χαρακτηριστικών μιας ανακατασκευασμένης εικόνας χαμηλής ανάλυσης  $G_{\partial G}(I^{LR})$  και της πραγματικής εικόνας υψηλής ανάλυσης  $I^{HR}$ :

$$l_{VGG/ij}^{SR} = \frac{1}{W_{ij}H_{ij}} \sum_{x=1}^{W_{ij}} \sum_{y=1}^{H_{ij}} (\phi_{ij}(I^{HR})_{x,y} - \phi_{ij}(G_{\partial G}(I^{LR}))_{x,y})^2$$

όπου  $W_{ij}$  και  $H_{ij}$  είναι οι διαστάσεις του χάρτη χαρακτηριστικών για κάποια  $i, j$ . Στη βιβλιογραφία [9] εξετάζονται συγκεκριμένα τα VGG loss για  $i = 2, j = 2$  και για  $i = 5, j = 4$ .

Το ανταγωνιστικό κόστος έχει τη λογική που αναπτύξαμε πιο πάνω σχετικά με τα Αναγεννητικά Ανταγωνιστικά Δίκτυα, δηλαδή ενθαρρύνει τον γεννήτορα να παράγει εικόνες υψηλής ανάλυσης πιστές στην πραγματικότητα, προσπαθώντας να ξεγελάσει τον διαχωριστή.

Το ανταγωνιστικό κόστος  $l_{Gen}^{SR}$  ορίζεται από τις πιθανότητες  $D_{\partial D}(G_{\partial G}(I^{LR}))$  που δίνει ο διαχωριστής σε κάθε ανακατασκευασμένη εικόνα για το κατά πόσο είναι πραγματική. Για όλες τις ανακατασκευασμένη εικόνα, το ανταγωνιστικό κόστος είναι:

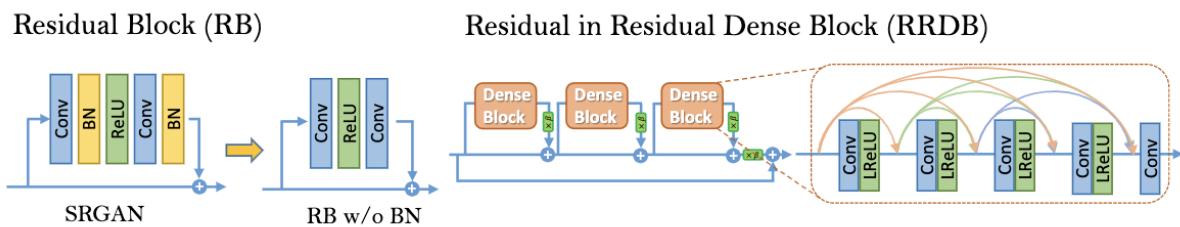
$$l_{Gen}^{SR} = \sum_{n=1}^N -\log D_{\partial D}(G_{\partial G}(I^{LR}))$$

Όπως καταλαβαίνουμε, όσο μικρότερες πιθανότητες δίνει ο διαχωριστής στις ανακατασκευασμένες εικόνες, τόσο μεγαλύτερο γίνεται το κόστος.

### 5.1.2 ESRGAN

Όπως αναφέρθηκε, το ESRGAN (Enhanced Super Resolution Generative Adversarial Network) αποτελεί βελτίωση του SRGAN. Όπως μαρτυρά το όνομά του, και αυτό ακολουθεί την αρχιτεκτονική των Αναγεννητικών Ανταγωνιστικών Δικτύων και οι βελτιώσεις σε σχέση με το SRGAN αφορούν τις δομές του γεννήτορα και του διαχωριστή, καθώς και τη συνάρτηση κόστους. [10]

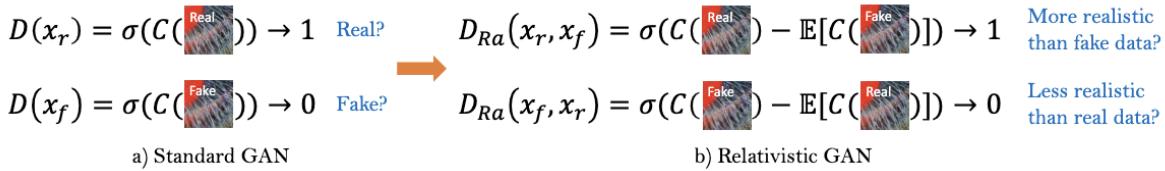
Όσον αφορά το νευρωνικό δίκτυο του γεννήτορα, έχουν υλοποιηθεί δύο κύριες βελτιώσεις. Η πρώτη βελτίωση είναι η αφαίρεση όλων των επιπέδων κανονικοποίησης παρτίδας από τα residual blocks. Έχει αποδειχτεί ότι η αφαίρεσή τους αυξάνει την απόδοση και μειώνει το υπολογιστικό κόστος, καθώς τα επίπεδα κανονικοποίησης παρτίδας καθιστούν ασταθή την απόδοση κατά την εκπαίδευση του νευρωνικού, ιδίως όταν το νευρωνικό είναι πολύ βαθύ και ακολουθεί την δομή ενός Αναγεννητικού Ανταγωνιστικού Δικτύου. Η δεύτερη βελτίωση είναι η προσθήκη ενός skip connection σε κάθε residual block που θα παραλείπει όλο το residual block, δημιουργώντας έτσι την λεγόμενη residual-in-residual μορφή. Έτσι, ο γεννήτορας του ESRGAN έχει μια πιο βαθιά και πολύπλοκη δομή από αυτή του γεννήτορα του SRGAN και έχει παρατηρηθεί ότι, ιδίως στο πρόβλημα της Υπερανάλυσης Εικόνας, οι περίπλοκες δομές βελτιώνουν την απόδοση.



**Σχήμα 5.3:** Οι δύο βελτιώσεις του γεννήτορα του ESRGAN: Στα αριστερά βλέπουμε την αφαίρεση των επιπέδων της κανονικοποίησης παρτίδας μέσα σε κάθε residual block και στα δεξιά την προσθήκη των skip connections για τη δημιουργία της residual-in-residual μορφής. [10]

Όσον αφορά τον διαχωριστή, πλέον δεν χρησιμοποιείται ο κλασικός διαχωριστής που χρησιμοποιείται στο SRGAN, αλλά ένας σχετικιστικός διαχωριστής μέσου όρου (Relativistic average Discriminator - RaD). Αυτός ο διαχωριστής, αντί να προβλέπει την πιθανότητα η εικόνα που δέχεται στην είσοδο να είναι αληθινή, προβλέπει την πιθανότητα η εικόνα της εισόδου να είναι πιο "αληθινή" από μια συλλογή εικόνων που γνωρίζει ότι είναι ψευδείς. [80]

Αυτός ο διαχωριστής αποδεικνύεται πιο αποτελεσματικός στην εκπαίδευση και βοηθάει τον γεννήτορα να κατασκευάσει εικόνες με πιο λεπτομερή χαρακτηριστικά.



**Σχήμα 5.4:** Η διαφορά μεταξύ του κλασικού διαχωριστή του SRGAN και του σχετικιστικού διαχωριστή του ESRGAN. [10]

Τέλος, υπάρχουν διαφορές όσον αφορά και τη συνάρτηση κόστους. Συγκεκριμένα, η συνάρτηση κόστους  $L_G$  του γεννήτορα είναι:

$$L_G = L_{percep} + \lambda L_G^{Ra} + \eta L_1$$

Εδώ το  $L_{percep}$  είναι το κόστος περιεχομένου που έχει οριστεί στο SRGAN με τη διαφορά ότι το κόστος αυτό υπολογίζεται από τους χάρτες χαρακτηριστικών πριν εφαρμοστεί σε αυτούς η συνάρτηση ενεργοποίησης και όχι μετά. Το  $L_G^{Ra}$  είναι το ανταγωνιστικό κόστος από τον σχετικιστικό διαχωριστή και το  $L_1$  είναι ίσο με  $\mathbb{E}_{x_i} \|G(x_i) - y\|$ , δηλαδή η νόρμα L1 μεταξύ της ανακατασκευασμένης εικόνας  $G(x_i)$  και της πραγματικής εικόνας  $y$  που της αντιστοιχεί. Τα  $\lambda, \eta$  είναι παράμετροι για την εξισορρόπηση των όρων της παραπάνω εξίσωσης.

Στην παρούσα διπλωματική θα χρησιμοποιήσουμε έναν προεκπαιδευμένο γεννήτορα του ESRGAN για να αυξήσουμε την ανάλυση των εικόνων από το Sentinel-2 με παράγοντα 4. Ο κώδικας που χρησιμοποιήθηκε είναι ανοιχτός στο [81].

## 5.2 Σημασιολογική Κατάτμηση (Semantic Segmentation)

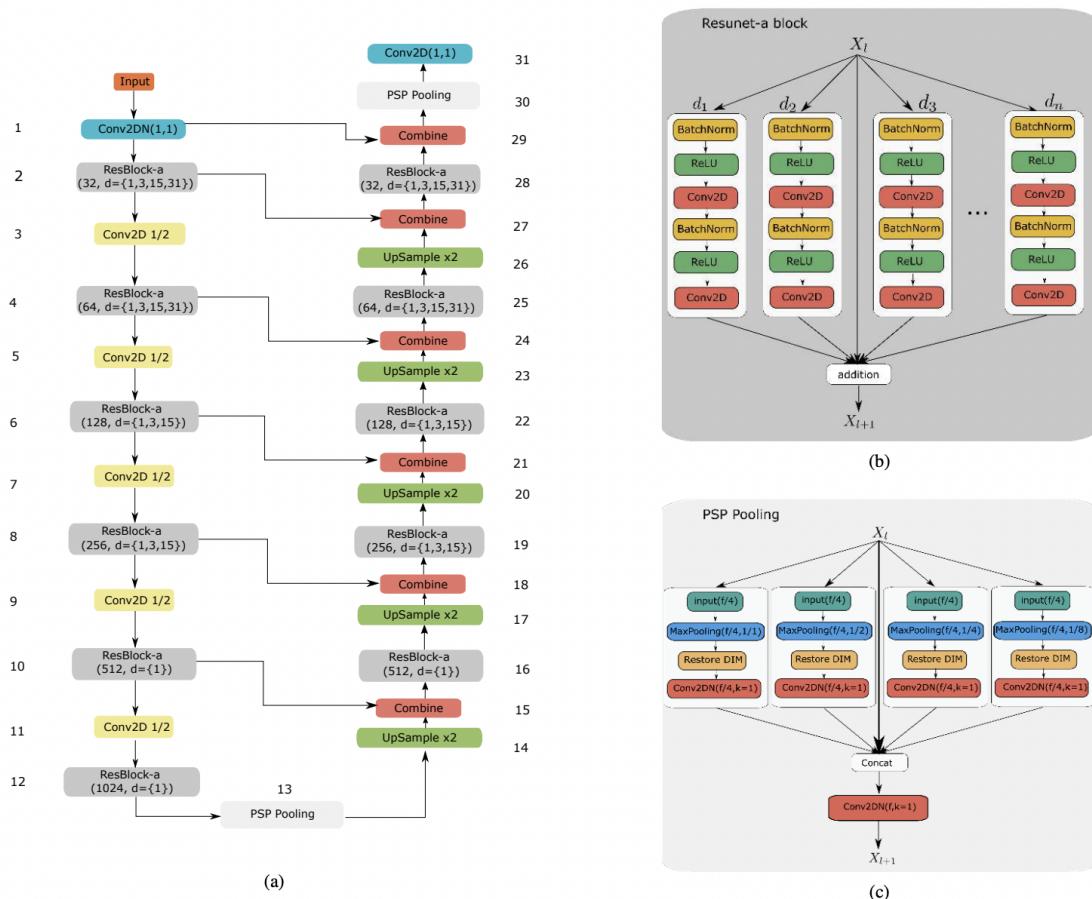
Στην παρούσα διπλωματική θα χρησιμοποιήσουμε δύο μοντέλα για σημασιολογική κατάτμηση, το ResUNet-a και το DeepLabv3+.

### 5.2.1 ResUNet-a

Το ResUNet-a [11] αποτελεί ένα Πλήρως Συνδεδεμένο Συνελικτικό Νευρωνικό Δίκτυο (FCN), το οποίο είναι σχεδιασμένο για την σημασιολογική κατάτμηση εικόνων τηλεπισκόπησης. Ακολουθεί την αρχιτεκτονική κωδικοποιητή-αποκωδικοποιητή (encoder-decoder) που χρησιμοποιεί το UNet, το πιο βασικό Νευρωνικό Δίκτυο αυτής της αρχιτεκτονικής που ασχολήθηκε με τη σημασιολογική κατάτμηση [6]. Ωστόσο, οι αλλαγές σε σχέση με το UNet είναι οι εξής:

- Για να μην είναι ασταθής η εκπαίδευση ειδικά όταν το βάθος του Νευρωνικού Δικτύου μεγαλώνει, τα μπλοκ του UNet αντικαθιστώνται από residual blocks, τα ResUNet-a blocks, που περιέχουν συνελικτικά στρώματα. Αυτό αντιμετωπίζει το πρόβλημα των vanishing και exploding gradients που συναντάται ιδιαίτερα σε βαθιές αρχιτεκτονικές.

- Για καλύτερη κατανόηση σε πολλές κλίμακες, μέσα σε κάθε residual block γίνονται πολλαπλές παράλληλες τραχείς συνελίξεις (atrous convolutions) με διαφορετικούς ρυθμούς διαστολής. Η ιδέα πίσω από αυτό είναι ότι καθιστάται ικανή η εισαγωγή χαρακτηριστικών από διαφορετικά πεδία πρόσληψης (receptive fields).
- Για καλύτερη απόδοση των δικτύων με την εισαγωγή πληροφορίας από το παρασκήνιο, εισάγεται το στρώμα Pyramid Scene Parsing Pooling - PSPP [82]. Το στρώμα αυτό χρησιμοποιείται σε δύο σημεία. Το πρώτο είναι αιμέσως μετά τον κωδικοποιητή και το δεύτερο είναι το προτελευταίο στρώμα πριν τη δημιουργία της μάσκας κατάτμησης (segmentation mask).



**Σχήμα 5.5:** (a) Η αρχιτεκτονική του ResUNet-a, στα αριστερά ο κωδικοποιητής, στα δεξιά ο αποκωδικοποιητής, (b) Η δομή του βασικού ResUNet-a block του ResUNet-a, με τα  $d_i$  να αναπαριστούν τους διάφορους ρυθμούς διαστολής, (c) Το στρώμα Pyramid Scene Parsing Pooling. [11]

Στο [11] παρατίθενται δύο μοντέλα ResUNet-a, το ResUNet-a d6 και το ResUNet-a d7, τα οποία διαφέρουν ως προς το βάθος τους, καθώς στο ResUNet-a d6 ο κωδικοποιητής (encoder) αποτελείται από 6 ResUNet-a blocks και ένα στρώμα PSPP, ενώ στο ResUNet-a d7 ο κωδικοποιητής αποτελείται από 7 ResUNet-a blocks. Η είσοδος περνάει αρχικά από ένα συνελικτικό στρώμα με πυρήνα μεγέθους  $1 \times 1$  για να αυξηθεί ο αριθμός των χαρακτηριστικών

στο επιθυμητό μέγεθος του αρχικού φίλτρου, χωρίς να υπάρξει απώλεια πληροφορίας. Έπειτα ακολουθούν τα ResUNet-a blocks. Σε κάθε ResUNet-a block χρησιμοποιούνται μέχρι 3 σετ από παράλληλες atrous συνελίξεις εκτός από το κλασσικό σετ των 2 συνελίξεων της αρχιτεκτονικής των residual block, δηλαδή έχουμε μέχρι 4 παράλληλα σετ που περιέχουν 2 συνελικτικά στρώματα. Στη συνέχεια η έξοδος προστίθεται στην είσοδο. Η έξοδος από κάθε μπλοκ υποδειγματοληπτείται με μια συνέλιξη με πυρήνα μεγέθους  $1 \times 1$  και βήμα ολίσθησης 2. Στο τέλος του κωδικοποιητή και του αποκωδικοποιητή υπάρχει ένα στρώμα PSP Pooling, το οποίο χωρίζει την είσοδο σε 4 ίσα κομμάτια στον χώρο χαρακτηριστικών (feature space) και στη συνέχεια εκτελεί max pooling σε διαδοχικές διχοτομήσεις της εισόδου, σε 1, 4, 16 και 64 κομμάτια. Στον αποκωδικοποιητή, η υπερδειγματοληψία εκτελείται με τον αλγόριθμο του κοντινότερου γείτονα ακολουθούμενου από ένα συνελικτικό στρώμα  $1 \times 1$  και ένα στρώμα Batch Normalization. Τα στρώματα του κωδικοποιητή και του αποκωδικοποιητή συνενώνονται μεταξύ τους μέσω του στρώματος Combine, το οποίο ενώνει τις δύο εισόδους και τις περνάει από μια συνέλιξη που φέρνει τον αριθμό των χαρακτηριστικών στην επιθυμητή τιμή.

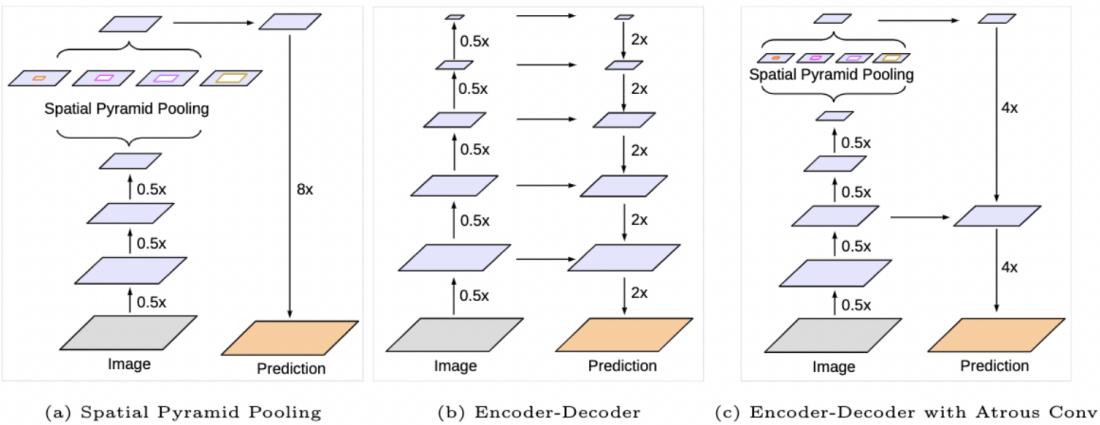
### 5.2.2 DeepLabv3+

Το μοντέλο DeepLabv3+ [12] συνδυάζει δύο επιτυχημένες αρχιτεκτονικές στη σημασιολογική κατάτμηση εικόνας, το Atrous Spatial Pyramid Pooling και την αρχιτεκτονική κωδικοποιητή-αποκωδικοποιητή. Χρησιμοποιεί το μοντέλο DeepLabv3 ως κωδικοποιητή και διαθέτει έναν απλό αλλά αποτελεσματικό αποκωδικοποιητή που βελτιώνει τα αποτελέσματα του σχηματισμού του περιγράμματος των αντικειμένων.

Για να συγκεντρώσει τις "συμφραζόμενες" πληροφορίες σε διαφορετικές κλίμακες, το DeepLabv3 εκτελεί πολλαπλές παράλληλες atrous συνελίξεις με διαφορετικούς ρυθμούς (Atrous Spatial Pyramid Pooling) [74]. Παρότι όμως στην έξοδο του DeepLabv3 είναι κωδικοποιημένες πλούσιες σημασιολογικές πληροφορίες, λείπουν λεπτομερείς πληροφορίες σχετικά με το περιγραμμα των αντικειμένων εξαιτίας των πράξεων pooling και των συνέλιξεων με μεγάλα βήματα ολίσθησης. Η αρχιτεκτονική κωδικοποιητή-αποκωδικοποιητή είναι ένας τρόπος να αντιμετωπιστεί το πρόβλημα αυτό, καθώς ο αποκωδικοποιητής δημιουργεί σταδιακά ξεκάθαρα περιγράμματα αντικειμένων. Το DeepLabv3+ λοιπόν συνδυάζει τα πλεονεκτήματα των δύο αυτών προσεγγίσεων με, εμπλουτίζοντας την αρχιτεκτονική κωδικοποιητή-αποκωδικοποιητή με το να συμπεριλαμβάνει τις συμφραζόμενες πληροφορίες του αποκωδικοποιητή.

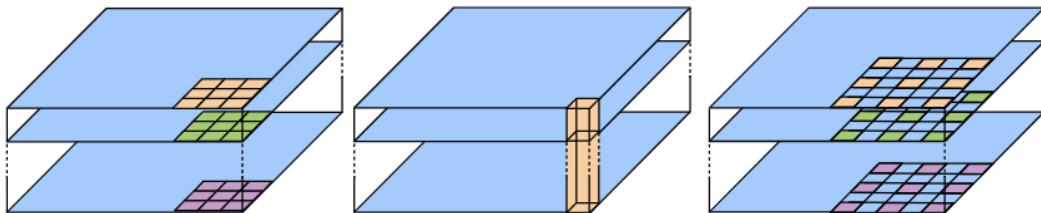
Ειδικότερα, το DeepLabv3+ χρησιμοποιεί τις εξής τεχνικές. Αρχικά, χρησιμοποιεί την atrous συνέλιξη, η οποία επιτρέπει τον έλεγχο της ανάλυσης των υπολογιζόμενων χαρακτηριστικών και ρυθμίζει το πεδίο θέασης (field of view) του φίλτρου ώστε να εξάγει πληροφορία σε πολλαπλές κλίμακες. Στην περίπτωση δισδιάστατων σημάτων, για κάθε θέση  $i$  στον χάρτη χαρακτηριστικών  $y$  και με ένα συνελικτικό φίλτρο  $w$ , η atrous συνέλιξη υπολογίζεται στην είσοδο  $x$  ως εξής:

$$y[i] = \sum_k [x[i] + r \cdot k] \cdot w[k]$$



**Σχήμα 5.6:** Η βελτίωση του μοντέλου DeepLabv3, το οποίο ακολουθεί το μοντέλο Spatial Pyramid Pooling (a), με την αρχιτεκτονική encoder-decoder (b). Το DeepLabv3+ περιέχει πλούσιες σημασιολογικές πληροφορίες από τον κωδικοποιητή και παράγει λεπτομερή περιγράμματα αντικειμένων από τον απλό αλλά αποτελεσματικό αποκωδικοποιητή (c). [12]

όπου ο atrous ρυθμός  $r$  καθορίζει το βήμα ολίσθησης με το οποίο δειγματοληπτούμε την είσοδο.



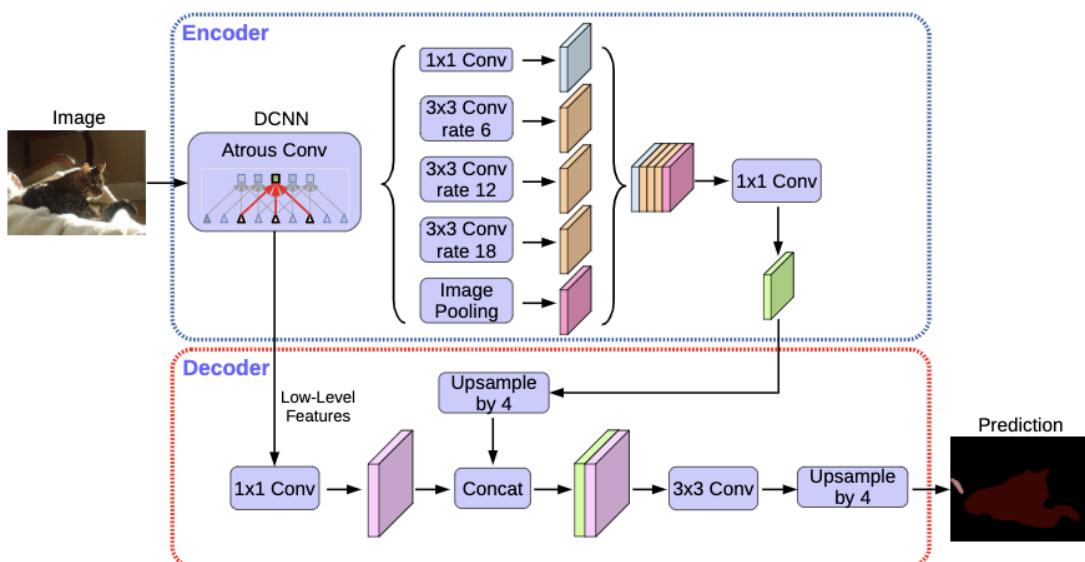
**Σχήμα 5.7:** (a) Depthwise συνέλιξη, (b) pointwise συνέλιξη, (c) atrous depthwise συνέλιξη. [12]

Εκτός από την atrous συνέλιξη, το DeepLabv3+ χρησιμοποιεί και τις *depthwise* συνέλιξεις ακολουθούμενες από τις *pointwise* συνέλιξεις, δηλαδή συνελίξεις με πυρήνα  $1 \times 1$ . Η *depthwise* συνέλιξη είναι μια χωρική συνέλιξη και εκτελείται χωριστά για κάθε κανάλι εισόδου και τα αποτελέσματα αυτών των συνελίξεων συνδυάζονται μέσω των *pointwise* συνελίξεων. Η *atrous depthwise* συνέλιξη που φαίνεται στο σχήμα 5.7 προκύπτει από την *depthwise* συνέλιξη με rate 2 και αποδεικνύεται ότι μειώνει σημαντικά την υπολογιστική πολυπλοκότητα χωρίς να μειώνει την ποιότητα.

Όπως αναφέρθηκε, το DeepLabv3+ χρησιμοποιεί το μοντέλο DeepLabv3 ως κωδικοποιητή. Χρησιμοποιεί την atrous συνέλιξη για να εξάγει χαρακτηριστικά που έχουν υπολογιστεί από βαθιά νευρωνικά δίκτυα. Για την σημασιολογική κατάτμηση εικόνας, το output stride, που είναι ο λόγος της χωρικής ανάλυσης της εισόδου προς της εξόδου, μπορεί να έχει τιμή από 16 έως 8, αφαιρώντας το striding στα τελευταία ένα ή δύο μπλοκ και χρησιμοποιώντας atrous συνέλιξη. Επίσης, με το να εφαρμόζει συνέλιξη σε διαφορετικούς ρυθμούς, επεκτείνει το Atrous Spatial Pyramid Pooling module. Η εξόδος από το DeepLabv3 που δίνεται στον αποκωδικοποιητή είναι ο τελευταίος χάρτης χαρακτηριστικών πριν τα logits, ο οποίος περιέχει 256 κανάλια και πλούσια σημασιολογική πληροφορία.

Από την πλευρά του αποκωδικοποιητή, τα χαρακτηριστικά αυτά αρχικά υπερδειγματολειπτούνται διγραμμικά (bilinear upsampling) με παράγοντα 4 και συνενώνονται με αντίστοιχα χαρακτηριστικά από το backbone του δικτύου που έχουν την ίδια χωρική ανάλυση. Έπειτα, εφαρμόζεται μια συνέλιξη  $1 \times 1$  σε αυτά τα χαρακτηριστικά ώστε να μειωθεί ο αριθμός των καναλίων, ο οποίος σε εκείνο το σημείο είναι μεγάλος (συνήθως 256 ή 512) και δυσχεραίνει την εκπαίδευση. Έπειτα εφαρμόζονται και άλλες συνέλιξεις  $3 \times 3$  οι οποίες αυξάνουν την ποιότητα των χαρακτηριστικών και ακολουθεί τελικά μια ακόμα διγραμμική υπερδειγματοληψία με παράγοντα 4. Παρατηρείται ότι η απόδοση του μοντέλου βελτιώνται όταν χρησιμοποιείται output stride ίσο με 8 για τον κωδικοποιητή, θυσιάζοντας υπολογιστική ισχύ, ενώ η τιμή 16 για το output stride δίνει ένα καλό tradeoff μεταξύ ταχύτητας και ακρίβειας.

Στην παρούσα εργασία το DeepLabv3+ που θα χρησιμοποιήσουμε έχει ως backbone ένα δίκτυο ResNet50 [83] προεκπαιδευμένο στο σύνολο δεδομένων ImageNet.



**Σχήμα 5.8:** Η αρχιτεκτονική του DeepLabv3+ με το DeepLabv3 ως κωδικοποιητή. [12]

### 5.2.3 Swin Transformer

Ο Swin Transformer [13] είναι μια μορφή Νευρωνικού Δικτύου που ανήκει στην κατηγορία των Μετασχηματιστών (Transformers). Οι μετασχηματιστές αρχικά κατασκευάστηκαν στα πλαίσια της Επεξεργασίας Φυσικής Γλώσσας (Natural Language Processing) και έχουν την ικανότητα να μοντελοποιούν μακριές εξαρτήσεις μεταξύ ακολουθιών εισόδου. Επίσης, ο σχεδιασμός τους επιτρέπει την επεξεργασία διαφορετικών τύπων δεδομένων όπως εικόνες, βίντεο και κείμενο. Λόγω της μεγάλης τους επιτυχίας, έχουν εξαπλωθεί και στην υπολογιστική όραση και έχουν οδηγήσει σε υποσχόμενα αποτελέσματα.

Η αρχιτεκτονική ενός μετασχηματιστή αποτελείται από μια σειρά κωδικοποιητών και αποκωδικοποιητών. Ο κωδικοποιητής αρχικά δημιουργεί μια διανυσματική αναπαράσταση εισόδου embedding της ακολουθίας και την περνάει στα επόμενα στρώματά του. Περιέχει μεταξύ άλλων το θεμελιώδες στρώμα για την λειτουργία των μετασχηματιστών Self-Attention, το οποίο υπολογίζει μακριές σχέσεις μεταξύ των στοιχείων μιας ακολουθίας και δημιουργεί

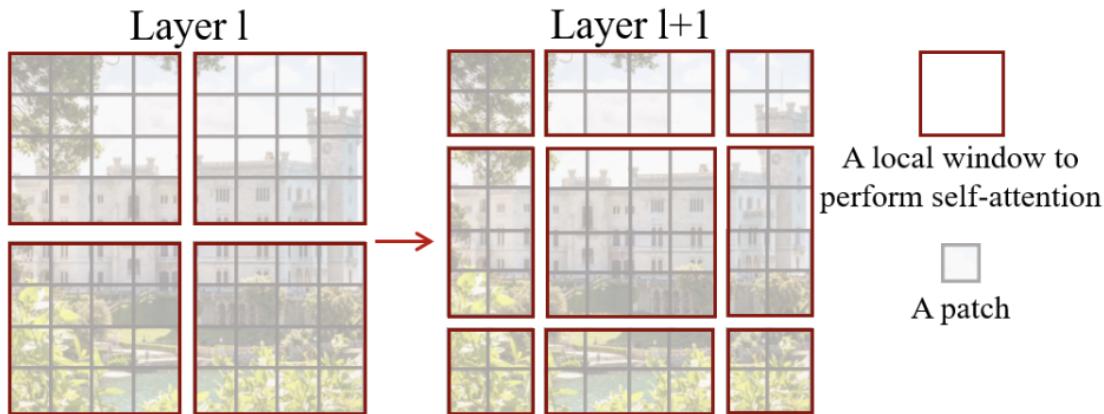
μια κωδικοποίηση για κάθε token της ακολουθίας εισόδου που περιγράφει τη συσχέτισή του με τα υπόλοιπα tokens της ακολουθίας. Ο αποκωδικοποιητής παίρνει ως είσοδο τόσο τα embeddings της εισόδου όσο και τα embeddings της προηγούμενης εξόδου μετατοπισμένα κατά ένα token και προβλέπει το επόμενο token στην ακολουθία. [76]

Ωστόσο, η μεταφορά των μετασχηματιστών από το πεδίο της Επεξεργασίας Γλώσσας στην Υπολογιστική Όραση έχει να αντιμετωπίσει ορισμένα προβλήματα. Ένα από αυτά είναι η κλίμακα, καθώς σε αντίθεση με τα tokens των λέξεων που είναι τα βασικά στοιχεία επεξεργασίας στους μετασχηματιστές γλώσσας, τα οπικά στοιχεία ποικίλουν σημαντικά σε κλίμακα. Ένα άλλο πρόβλημα είναι ότι οι διαστάσεις των εικόνων είναι πολύ μεγαλύτερες από ότι ο αριθμός των λέξεων σε κάποιο κείμενο που θα δινόταν σε μετασχηματιστή γλώσσας. Η σημασιολογική κατάτμηση, για παράδειγμα, που απαιτεί υπολογισμούς σε επίπεδο εικονοστοιχείου, θα είχε τεράστια υπολογιστική πολυπλοκότητα σε έναν κλασικό μετασχηματιστή, αφού σε αυτούς η πολυπλοκότητα κάθε στρώματος Self-Attention είναι τετραγωνική ως προς το μέγεθος της εισόδου. [13]

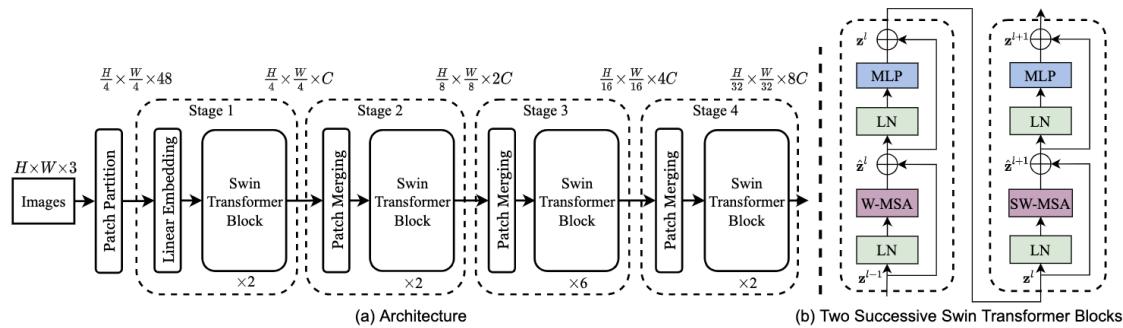
Για αυτό προτάθηκε ένα γενικής χρήσης backbone που ακολουθεί την αρχιτεκτονική των μετασχηματιστών, ο Swin Transformer. Ο Swin Transformer κατασκευάζει ιεραρχικούς χάρτες χαρακτηριστικών ξεκινώντας από μικρά patches της εικόνας, τα οποία έχουν τον ρόλο των tokens, και στα βαθύτερα στρώματα του συνενώνοντας γειτονικά patches, απαιτώντας μάλιστα γραμμική υπολογιστική πολυπλοκότητα ως προς το μέγεθος της εικόνας. Η γραμμική πολυπλοκότητα πετυχαίνεται με τον υπολογισμό του self-attention σε μη επικαλυπτόμενα παράθυρα που χωρίζουν την εικόνα. Ο αριθμός των patches σε κάθε παράθυρο είναι προκαθορισμένος, κάτι που καθιστά την πολυπλοκότητα γραμμική. Επίσης, ανάμεσα στα διαδοχικά στρώματα self-attention, ακολουθείται η μέθοδος των μετακινούμενων παραθύρων (shifting windows), δηλαδή μετακινούνται τα παράθυρα που χωρίζουν την εικόνα. Με αυτόν τον τρόπο μπορεί ένα παράθυρο να γεφυρώνει δύο παράθυρα του προηγούμενου στρώματος και έτσι ο μηχανισμός self-attention αποκτά μεγαλύτερη ικανότητα μοντελοποίησης. Όλα αυτά τα χαρακτηριστικά καθιστούν τον Swin Transformer ένα backbone για ποικίλες εφαρμογές υπολογιστικής όρασης.

Ο αλγόριθμος του Swin Transformer είναι ο εξής. Όπως αναφέρθηκε πιο πάνω, κάθε patch έχει το ρόλο του token και τα χαρακτηριστικά του (features) είναι η συνένωση όλων των τιμών των RGB εικονοστοιχείων του. Επιλέγεται μέγεθος patch τα  $4 \times 4$  εικονοστοιχεία και άρα η διάσταση των χαρακτηριστικών του είναι  $4 \times 4 \times 3 = 48$ . Σε αυτά τα χαρακτηριστικά εφαρμόζεται ένα linear embedding στρώμα για να το μετατρέψει σε κάποια άλλη διάσταση, που ονομάζεται C. Έπειτα ακολουθεί μια σειρά από Swin Transformer blocks με μηχανισμούς self-attention, τα οποία διατηρούν τον αριθμό των patches ( $\frac{H}{4}, \frac{W}{4}$ ) και μαζί με τα linear embedding στρώματα αναφέρονται ως "Επίπεδο 1".

Για να παραχθεί μια ιεραρχική απεικόνιση, ο αριθμός των patches μειώνεται από patch merging στρώματα. Αυτά τα στρώματα συγχωνεύουν τα γειτονικά patches για να δημιουργήσουν νέα, τα οποία έχουν διάσταση  $2 \times 2$  των αρχικών patches. Έπειτα εφαρμόζεται ένα γραμμικό στρώμα (Linear Layer) στα 4C-διάστατα χαρακτηριστικά, μειώνοντας τον αριθμό των διαστάσεων σε 2C. Έπειτα, ακολουθεί μια σειρά από Swin Transformer blocks, τα οποία διατηρούν τον αριθμό των patches σε  $(\frac{H}{8}, \frac{W}{8})$ , ολοκληρώνοντας έτσι το "Επίπεδο 2". Με την ίδια διαδικασία έχουμε και τα επίπεδα 3 και 4.



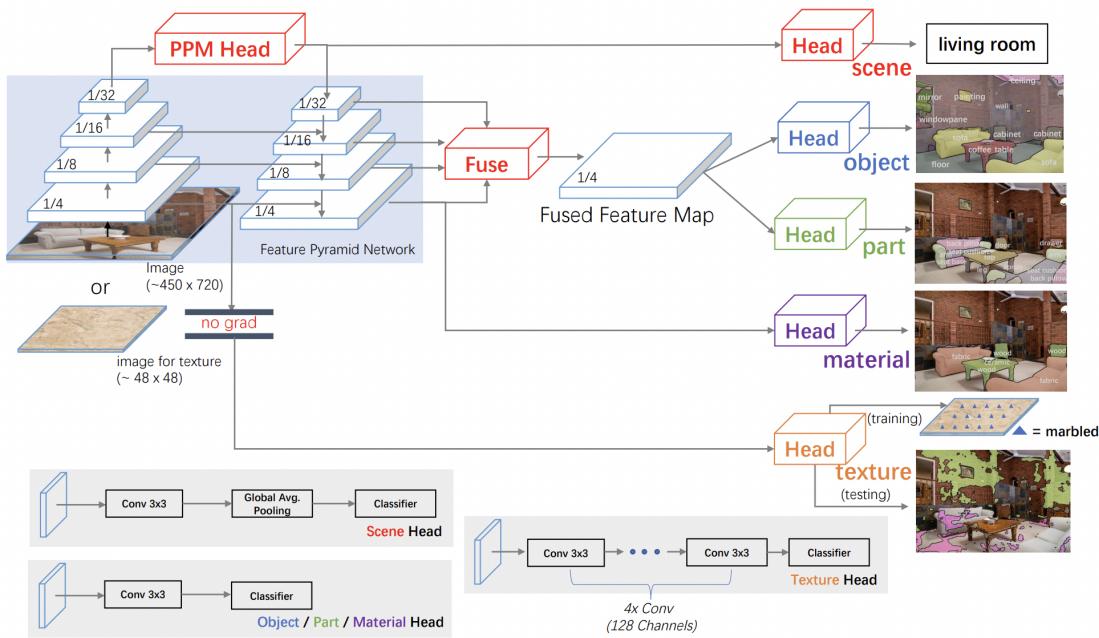
**Σχήμα 5.9:** Μια απεικόνιση της μεθόδου των *shifting windows*. Αριστερά, στο στρώμα  $l$ , εφαρμόζεται ένας απλός διαχωρισμός της εικόνας σε παράθυρα, και το self-attention πραγματοποιείται στο καθένα από αυτά. Δεξιά, στο στρώμα  $l+1$ , τα παράθυρα έχουν μετακινηθεί. Ο υπολογισμός του self-attention στα νέα παράθυρα διασχίζει τα όρια των παραθύρων του προηγούμενου στρώματος, δημιουργώντας συνδέσεις μεταξύ τους. [13]



**Σχήμα 5.10:** (a) Η αρχιτεκτονική του μοντέλου Swin-T του Swin Transformer (b) Δύο συνεχόμενα Swin Transformer blocks. [13]

Όπως φαίνεται και στο σχήμα, ένα Swin Transformer Block αποτελείται από ένα στρώμα self-attention (SA), το οποίο μπορεί να έχει απλό διαχωρισμό σε παράθυρα (W-MSA) ή διαχωρισμό σε μετακινούμενα παράθυρα (SW-MSA), ένα Multi-Layer Perceptron βάθους 2 και στρώματα κανονικοποίησης (Layer-Norm) πριν από τα προηγούμενα δύο στρώματα. Επίσης έχουμε residual connections σε κάθε module LN-MSA και LN-MLP.

Στην παρούσα διπλωματική, καθώς ασχολούμαστε με το πρόβλημα της σημασιολογικής κατάτμησης, ακολουθούμε τη βιβλιογραφία και θα προσθέσουμε τον Swin Transformer ως backbone στο μοντέλο UPerNet. [14] Το UPerNet παίρνει τους χάρτες χαρακτηριστικών από διάφορα επίπεδα του backbone δικτύου, τα συνενώνει ή τα περνάει από ένα στρώμα Pyramid Pooling [84] και ύστερα περνάει το αποτέλεσμα από διάφορα heads τα οποία εκτελέσουν σημασιολογική κατάτμηση σε πολλαπλά επίπεδα. Αυτά τα heads αποτελούνται κυρίως από συνελικτικά στρώματα και ταξινομητές και πραγματοποιούν ταξινόμηση σε επίπεδο σκηνής, αντικειμένου, μερών του αντικειμένου, υλικού και υφής.



**Σχήμα 5.11:** Η δομή του UperNet με το Feature Pyramid Network-FPN ως backbone. Στο κάτω μέρος βλέπουμε τη δομή κάθε head αναλόγως με το επίπεδο στο οποίο ταξινομεί. [14]

## 5.3 Μετρικές Αξιολόγησης

### 5.3.1 Συναρτήσεις Απώλειας (Loss Functions)

Σε αυτή την διπλωματική εξετάζουμε την επίδοση των μοντέλων με 4 διαφορετικές συναρτήσεις απώλειας.

#### Binary Cross-Entropy

Μία από τις μετρικές σφάλματος για σημασιολογική κατάτμηση 2 κλάσεων που έχουν επικρατήσει είναι η Binary Cross Entropy, η οποία υπολογίζεται από τον τύπο :

$$L_{BCE} = -y \log p - (1 - y) \log (1 - p)$$

όπου η  $y$  είναι η πραγματική τιμή της κλάσης στην οποία ανήκει το εικονοστοιχείο (0 ή 1) και  $p$  είναι η πρόβλεψη του μοντέλου για την κλάση του εικονοστοιχείου. Η συνάρτηση αυτή δίνει μεγαλύτερο σφάλμα όσο η πιθανότητα που υπολογίζεται διαφέρει από την πραγματική κλάση.

#### Dice Loss

Η συνάρτηση σφάλματος Dice ορίζεται μαθηματικά ως εξής :

$$L_{DICE} = 1 - \frac{2|y \cap p| + 1}{|y| + |p| + 1}$$

όπου  $|y \cap p|$  είναι ο αριθμός των εικονοστοιχείων της πραγματικής εικόνας  $y$  και της προβλεπόμενης εικόνας  $p$  που είναι ίσα και  $|y|, |p|$  ο αριθμός των εικονοστοιχείων στις δύο εικόνες ξεχωριστά. Το γεγονός ότι το κλάσμα υπολογίζει ουσιαστικά την επικάλυψη των εικόνων σε σχέση με το σύνολο των εικονοστοιχείων τους καθιστά τη συνάρτηση κόστους πολύ χρήσιμη για να επιτευχθούν υψηλές ακρίβεις και πολλές φορές προτιμητέα έναντι της Binary Cross-Entropy.

## Hybrid Loss

Στο [85] γίνεται αναφορά σε μια „ϋβριδική“ συνάρτηση απώλειας, η οποία συνδυάζει τρεις προϋπάρχουσες συναρτήσεις απώλειας, με σκοπό να προκύψουν πιο ακριβείς σημασιολογικούς χάρτες αντικειμένων σε διάφορες κλίμακες. Συγκεκριμένα, η Hybrid Loss εκφράζεται ως εξής:

$$L_{\text{Hybrid}} = L_{\text{FL}} + L_{\text{MS-SSIM}} + L_{\text{IOU}}$$

όπου  $L_{\text{FL}}$  είναι η συνάρτηση Focal Loss,  $L_{\text{MS-SSIM}}$  είναι η συνάρτηση Multi-Scale Structural Similarity και  $L_{\text{IOU}}$  είναι η συνάρτηση Intersection over Union Loss. Στην παρούσα διπλωματική, το hybrid loss που θα χρησιμοποιήσουμε δεν περιλαμβάνει το Multi-Scale Structural Similarity, καθώς σε ορισμένες αρχιτεκτονικές, όπως του ResUNet-a είναι ασταθές στην αρχή της εκπαίδευσης και μπορεί να επιστρέψει μη αριθμητικές NaN τιμές. Εντέλει, η συνάρτηση που θα χρησιμοποιήσουμε είναι το áθροισμα του Focal Tversky Loss, που είναι μια γενίκευση του Focal Loss, και του IoU Loss.

Η Focal Tversky Loss [86] εκφράζεται ως:

$$L_{\text{FTL}} = (1 - \frac{TP}{TP + \alpha FN + \beta FP})^\gamma$$

όπου TP τα true positive προβλεπόμενα εικονοστοιχεία της κάθε κλάσης, FN τα false negative και FP τα false positive. Η συνάρτηση αυτή έχει παραμέτρους τα  $\alpha, \beta, \gamma$ , με τα οποία μπορεί κανείς να προσαρμόσει την εκπαίδευση αναλόγως με το πόσο ανισόρροπο (imbalanced) είναι το σύνολο δεδομένων του. Για παράδειγμα, συνίσταται  $\gamma > 1$  όταν το σύνολο είναι ανισόρροπο. Επίσης, όταν το  $\alpha$  είναι μεγαλύτερο του  $\beta$  η συνάρτηση σφάλματος “τιμωρεί” περισσότερο τα false negatives. Στην παρούσα διπλωματική, επειδή έχουμε μια ανισορροπία ως προς τον αριθμό των εικονοστοιχείων που δεν ανήκουν σε δάση και αυτών που ανήκουν, εφαρμόζουμε  $\alpha = 0.5, \beta = 0.4, \gamma = 4/3$ .

Η IoU Loss υπολογίζεται ως:

$$L_{\text{IOU}} = 1 - \frac{TP}{TP + FN + FP}$$

δηλαδή είναι ειδική περίπτωση του Focal Tversky Loss όταν όλες οι παράμετροι είναι ίσες με 1.

### Tanimoto loss

Στη δημοσίευση του μοντέλου ResUNet-a [11], εισάγεται μια νέα συνάρτηση κόστους, το Tanimoto loss. Αποτελεί ουσιαστικά μια παραλλαγή του Dice Loss και προκύπτει ότι οι μερικές παράγωγοι της βοηθούν στο να συγκλίνει γρηγορότερα το μοντέλο στην αληθινή τιμή, ανεξάρτητα από τα τυχαία αρχικά βάρη του. Εκφράζεται ως:

$$L_{Tanimoto} = 1 - \frac{|y \cap p| + 1}{|y|^2 + |p|^2 - |y \cap p| + 1}$$

όπου χρησιμοποιείται ο ίδιος συμβολισμός με τον ορισμό του Dice Loss.

### 5.3.2 Μετρικές Ακρίβειας

#### Dice Coefficient

Η μετρική Dice Coefficient αναπτύχθηκε στην προηγούμενη παράγραφο και ο τύπος της είναι  $1 - DiceLoss$ . Η τελική τιμή της προκύπτει από τη μέση τιμή της Dice Coefficient για όλα τα ζεύγη πραγματικών και υπολογισμένων εικόνων.

#### Intersection over Union - IoU

Η μετρική Intersection over Union αποτελεί το λόγο της επικάλυψης μεταξύ της πραγματικής και της υπολογισμένης εικόνας, δηλαδή των αριθμών ισων εικονοστοιχείων τους, προς την ένωση των εικονοστοιχείων τους. Ο τύπος της είναι:

$$IOU = \frac{X \cap Y}{X \cup Y}$$

Η τελική τιμή της είναι η μέση τιμή των μετρικών για όλα τα ζεύγη πραγματικών και υπολογισμένων εικόνων.

#### Precision

Η μετρική Precision είναι ο λόγος των σωστών προβλέψεων προς όλες τις προβλέψεις για κάποια κλάση και ο τύπος της είναι:

$$Precision = \frac{TP}{TP + FP}$$

όπου TP, FP τα true και false positives που αναπτύχθηκαν στην προηγούμενη παράγραφο.

#### Recall

Η μετρική Recall είναι ο λόγος των true positives προς το άθροισμα των true positives με τα false negatives, δηλαδή εξετάζει την ικανότητα του μοντέλου να βρίσκει τα θετικά από όλα τα πραγματικά θετικά αποτελέσματα.

$$Recall = \frac{TP}{TP + FN}$$

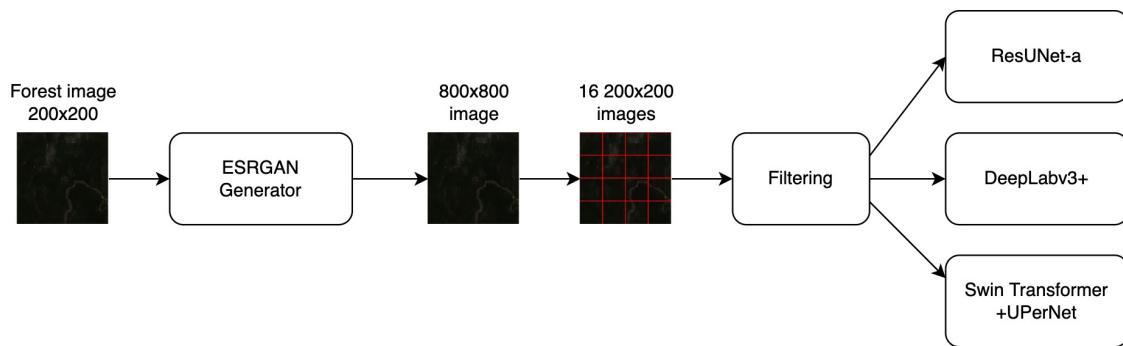
Η μετρική που χρησιμοποιήθηκε για την αξιολογόγηση των μοντέλων κατά τη διάρκεια της εκπαίδευσης είναι η μετρική Dice Coefficient.

## 5.4 Λεπτομέρειες Υλοποίησης

### 5.4.1 Διαδικασία και παράμετροι Μοντέλων

Σε αυτή τη διπλωματική συγκρίνουμε την επίδοση των παραπάνω μοντέλων τόσο μεταξύ τους όσο και με το αν έχουμε ή όχι υπερανάλυση εικόνας. Όσον αφορά το μέγεθος των εικόνων, έχουμε ως αρχικό μέγεθος τα  $200 \times 200$  εικονοστοιχεία. Εφαρμόζουμε σε αυτές υπερανάλυση εικόνας με τον γεννήτορα του ESRGAN με παράμετρο 4 και έτσι προκύπτουν εικόνες με ανάλυση  $800 \times 800$ , τις οποίες χωρίζουμε σε τμήματα με ανάλυση  $200 \times 200$ . Επειδή όμως στα μοντέλα ResUNet-a και DeepLabv3+ οι διαστάσεις των εικόνων πρέπει να είναι δυνάμεις του 2, έγινε resize σε αυτές τις εικόνες σε διαστάσεις  $256 \times 256$ . Επίσης έγινε resize στις εικόνες σε διάσταση  $224 \times 224$  για την εκπαίδευση του Swin Transformer.

Επιπλέον, παρατηρούμε ότι το σύνολο δεδομένων είναι αρκετά ανισορροπημένο, δηλαδή ο αριθμός των εικονοστοιχείων που αντιστοιχούν σε δάση αντιστοιχούσε περίπου στο 3.3% όλων των εικονοστοιχείων των εικόνων χωρίς υπερανάλυση, πράγμα που δυσκολεύει την εκπαίδευση και οδηγεί σε άσχημα αποτελέσματα των μοντέλων. Για αυτό, για τις εικόνες χωρίς υπερανάλυση, για την εκπαίδευση κρατήσαμε μόνο τις εικόνες στις οποίες τα εικονοστοιχεία που αντιστοιχούσαν σε δασική έκταση ξεπερνούσαν το 5% των εικονοστοιχείων της εικόνας, ενώ για τις εικόνες με υπερανάλυση κρατήσαμε τις εικόνες με ποσοστό των εικονοστοιχείων που αντιστοιχούσαν σε δασικές εκτάσεις πάνω από 10%. Τα ποσοστά των εικονοστοιχείων που ανήκουν σε δασικές εκτάσεις μετά από αυτό το φίλτραρισμα ανέρχεται στο 16% για τις εικόνες χωρίς υπερανάλυση και 39% για τις εικόνες με υπερανάλυση.



Σχήμα 5.12: Η διαδικασία για τις εικόνες με υπερανάλυση.

Για την εκπαίδευση των μοντέλων, λόγω απαγορευτικών απαιτήσεων μνήμης, επιλέχθηκε batch size=2. Ο αριθμός των εποχών τέθηκε στις 100, αν και τα μοντέλα είχαν τη μέγιστη απόδοσή τους πολύ πριν φτάσουν στην εκατοστή εποχή. Ως optimizer επιλέχθηκε ο Adam με learning rate=0.001. Τα μοντέλα ResUNet-a και DeepLabv3+ αναπτύχθηκαν στη βιβλιοθήκη Keras της Python, ενώ για τον Swin Transformer χρησιμοποιήθηκε το framework MMSegmentation [87], που είναι γραμμένο στη βιβλιοθήκη Pytorch.

Στο μοντέλο ResUNet-a οι παράμετροι που επιλέχθηκαν είναι 5 επίπεδα downsampling ακολουθούμενα από ένα στρώμα Atrous Spatial Pyramid Pooling με 256 φίλτρα, 5 επίπεδα

upsampling και ένα στρώμα ASPP με 128 φίλτρα. Οι ρυθμοί διαστολής είναι 1, 3, 15, 31, η συνάρτηση ενεργοποίησης είναι η ReLU και η συνάρτηση ενεργοποίησης της εξόδου είναι η σιγμοειδής.

Το μοντέλο DeepLabv3+ που χρησιμοποιήθηκε έχει ως backbone ένα ResNet50 το οποίο είναι προεκπαίδευμένο στο σύνολο δεδομένων ImageNet.

Ο Swin Transformer που χρησιμοποιήθηκε έχει διάσταση embedding  $C = 96$ , αριθμό Swin Transformer blocks [2, 2, 6, 2] ανά επίπεδο, δηλαδή αποτελεί την "tiny" εκδοχή του Swin Transformer (Swin-T). Επίσης έχει πλήθος παραθύρων 7, μέγεθος patch  $4 \times 4$ , αριθμό attention heads [3, 6, 12, 24] για κάθε επίπεδο και την GELU ως συνάρτηση ενεργοποίησης.

Για κάθε συνάρτηση σφάλματος προπονούμε δύο εκδοχές του κάθε μοντέλου: η πρώτη εκδοχή είναι στο σύνολο των εικόνων χωρίς υπερανάλυση και η δεύτερη είναι στο σύνολο με υπερανάλυση. Όσον αφορά το σύνολο δοκιμής (test set), δοκιμάζουμε τις εκδοχές αυτές τόσο στο σύνολο δεδομένων στο οποίο προπονήθηκαν, δηλαδή το μοντέλο που προπονήθηκε στο σύνολο χωρίς υπερανάλυση δοκιμάζεται στο σύνολο χωρίς υπερανάλυση, όσο και σε ένα κοινό σύνολο που περιέχει εικόνες χωρίς υπερανάλυση και εικόνες με υπερανάλυση. Περισσότερα για τα σύνολα δοκιμής περιγράφονται στο επόμενο κεφάλαιο.

## 5.4.2 Υπολογιστικό Σύστημα

Η εκτέλεση των πειραμάτων έγινε στους servers του Εργαστηρίου Τεχνητής Νοημοσύνης και Μάθησης του Εθνικού Μετσόβιου Πολυτεχνείου. Συγκεκριμένα, ο server που χρησιμοποιήθηκε διαθέτει 2 GPU Nvidia GeForce GTX 1080 με 8GB μνήμη η καθεμία.



## Κεφάλαιο 6

### Πειραματικά Αποτελέσματα

Σε αυτό το κεφάλαιο παρουσιάζονται τα αποτελέσματα από τα πειράματα που εκτελέστηκαν με τα μοντέλα και τις συναρτήσεις απώλειας που αναλύθηκαν στο προηγούμενο κεφάλαιο.

#### 6.1 Ποσοτική Αξιολόγηση Αποτελεσμάτων

##### 6.1.1 Αποτελέσματα για τις εικόνες χωρίς υπερανάλυση

Στους παρακάτω πίνακες βλέπουμε τα αποτελέσματα για τα μοντέλα ResUNet-a, DeepLabv3+, Swin Transformer και τις διάφορες συναρτήσεις σφάλματος που είναι εκπαιδευμένα στις εικόνες χωρίς υπερανάλυση. Συγκεκριμένα, βλέπουμε στους πίνακες 6.1, 6.2, 6.3, 6.4 τα αποτελέσματα για τις μετρικές Dice Coefficient, IoU, Precision και Recall αντίστοιχα. Υστερα από τη διαδικασία που περιγράψαμε στην παράγραφο 5.4.1, το πλήθος των εικόνων που έχουμε είναι 861 για το σύνολο εκπαίδευσης (training set), 311 για το σύνολο επικύρωσης (validation set) και 294 για το σύνολο δοκιμής (test set).

Όπως μπορούμε να δούμε, τα αποτελέσματα δεν είναι ικανοποιητικά, καθώς πολλές φορές είτε οι ετικέτες είναι μικρές ως προς το μέγεθος της εικόνας, είτε τα μοντέλα αναγνωρίζουν και άλλες περιοχές ως δάση που δεν αναγράφονται στις ετικέτες. Περισσότερες λεπτομέρειες σχετικά με την ποιότητα των αποτελεσμάτων και των ετικετών αναλύονται στην παράγραφο 6.3. Το μοντέλο που γενικεύει καλύτερα ως προς τις μετρικές Dice Coefficient, IoU και Precision είναι το Swin Transformer, ενώ το DeepLabv3+ πετυχαίνει το μεγαλύτερο Recall. Επίσης, παρατηρούμε ότι η συνάρτηση σφάλματος Cross Entropy δεν μας δίνει καλά αποτελέσματα σε σχέση με τις άλλες, με την υβριδική συνάρτηση καθώς και την Dice Loss να μας δίνουν τις καλύτερες μετρικές σε γενικές γραμμές.

Συνάρτηση Σφάλματος	ResUNet-a	DeepLabv3+	Swin Transformer
Cross Entropy	0.185	0.188	<b>0.282</b>
Dice	0.167	0.153	<b>0.346</b>
Hybrid	0.196	0.126	<b>0.359</b>
Tanimoto	0.193	0.188	<b>0.327</b>

Πίνακας 6.1: Τα πειραματικά αποτελέσματα των μοντέλων για τη μετρική Dice Coefficient για το σύνολο εικόνων χωρίς υπερανάλυση.

Συνάρτηση Σφάλματος	ResUNet-a	DeepLabv3+	Swin Transformer
Cross Entropy	0.151	<b>0.19</b>	0.164
Dice	0.207	0.127	<b>0.209</b>
Hybrid	0.184	0.088	<b>0.219</b>
Tanimoto	0.172	0.174	<b>0.196</b>

Πίνακας 6.2: Τα πειραματικά αποτελέσματα των μοντέλων για τη μετρική IoU για το σύνολο εικόνων χωρίς υπερανάλυση.

Συνάρτηση Σφάλματος	ResUNet-a	DeepLabv3+	Swin Transformer
Cross Entropy	0.187	<b>0.267</b>	<b>0.267</b>
Dice	<b>0.32</b>	0.135	0.254
Hybrid	0.224	0.09	<b>0.275</b>
Tanimoto	0.222	0.213	<b>0.312</b>

Πίνακας 6.3: Τα πειραματικά αποτελέσματα των μοντέλων για τη μετρική Precision για το σύνολο εικόνων χωρίς υπερανάλυση.

Συνάρτηση Σφάλματος	ResUNet-a	DeepLabv3+	Swin Transformer
Cross Entropy	<b>0.434</b>	0.398	0.295
Dice	0.37	<b>0.679</b>	0.543
Hybrid	0.506	<b>0.739</b>	0.517
Tanimoto	0.432	<b>0.484</b>	0.344

Πίνακας 6.4: Τα πειραματικά αποτελέσματα των μοντέλων για τη μετρική Recall για το σύνολο εικόνων χωρίς υπερανάλυση.

### 6.1.2 Αποτελέσματα για τις εικόνες με υπερανάλυση

Στους παρακάτω πίνακες βλέπουμε τα αποτελέσματα για τα μοντέλα ResUNet-a, DeepLabv3+, Swin Transformer και τις διάφορες συναρτήσεις σφάλματος που είναι εκπαιδευμένα στις εικόνες με υπερανάλυση. Συγκεκριμένα, βλέπουμε στους πίνακες 6.5, 6.6, 6.7, 6.8 τα αποτελέσματα για τις μετρικές Dice Coefficient, IoU, Precision και Recall αντίστοιχα. Ύστερα από τη διαδικασία της παραγράφου 5.4.1, έχουμε 4296 εικόνες στο σύνολο εκπαίδευσης (training set), 1281 εικόνες στο σύνολο επικύρωσης (validation set) και 1264 εικόνες στο σύνολο δοκιμής (test set).

Όπως μπορούμε να δούμε, τα μοντέλα έχουν σαφώς καλύτερα αποτελέσματα σε όλες τις μετρικές. Ο Swin Transformer γενικεύει καλύτερα ως προς τις μετρικές Dice Coefficient και Precision, ενώ και η απόδοση του ResUNet-a είναι αρκετά ανταγωνιστική σε σχέση με τις εικόνες χωρίς υπερανάλυση. Όσον αφορά τις συναρτήσεις σφάλματος, πάλι η Dice Loss και η υθριδική συνάρτηση δείχνουν ως οι καλύτερες για το συγκεκριμένο πρόβλημα, ενώ η Cross Entropy φαίνεται ως η λιγότερο κατάλληλη.

Συνάρτηση Σφάλματος	ResUNet-a	DeepLabv3+	Swin Transformer
Cross Entropy	<b>0.633</b>	0.609	0.609
Dice	0.627	0.534	<b>0.687</b>
Hybrid	0.634	0.518	<b>0.685</b>
Tanimoto	0.645	0.628	<b>0.685</b>

Πίνακας 6.5: Τα πειραματικά αποτελέσματα των μοντέλων για τη μετρική Dice Coefficient για το σύνολο εικόνων με υπερανάλυση.

Συνάρτηση Σφάλματος	ResUNet-a	DeepLabv3+	Swin Transformer
Cross Entropy	<b>0.521</b>	0.503	0.513
Dice	0.513	0.422	<b>0.524</b>
Hybrid	<b>0.523</b>	0.393	0.521
Tanimoto	<b>0.532</b>	0.524	0.521

Πίνακας 6.6: Τα πειραματικά αποτελέσματα των μοντέλων για τη μετρική IoU για το σύνολο εικόνων με υπερανάλυση.

Συνάρτηση Σφάλματος	ResUNet-a	DeepLabv3+	Swin Transformer
Cross Entropy	0.622	0.586	<b>0.651</b>
Dice	0.545	0.467	<b>0.644</b>
Hybrid	0.578	0.417	<b>0.649</b>
Tanimoto	0.604	0.619	<b>0.633</b>

Πίνακας 6.7: Τα πειραματικά αποτελέσματα των μοντέλων για τη μετρική Precision για το σύνολο εικόνων με υπερανάλυση.

Συνάρτηση Σφάλματος	ResUNet-a	DeepLabv3+	Swin Transformer
Cross Entropy	0.764	<b>0.785</b>	0.709
Dice	<b>0.895</b>	0.815	0.738
Hybrid	0.845	<b>0.873</b>	0.725
Tanimoto	<b>0.818</b>	0.771	0.747

Πίνακας 6.8: Τα πειραματικά αποτελέσματα των μοντέλων για τη μετρική Recall για το σύνολο εικόνων με υπερανάλυση.

### 6.1.3 Αποτελέσματα όλων των μοντέλων σε ένα κοινό σύνολο

Για να μπορέσουμε να συγκρίνουμε την απόδοση των μοντέλων που εκπαιδεύτηκαν σε εικόνες χωρίς υπερανάλυση με την απόδοση των μοντέλων που εκπαιδεύτηκαν σε εικόνες με υπερανάλυση, χρειάζεται να τα δοκιμάσουμε σε ένα κοινό test set. Για το σκοπό αυτό, επιλέχθηκαν τυχαία 250 εικόνες από το test set του συνόλου χωρίς υπερανάλυση και 250 εικόνες του συνόλου με υπερανάλυση ώστε να δημιουργηθεί ένα σύνολο 500 εικόνων, το οποίο θα είναι το κοινό test set μας. Στους πίνακες 6.9, 6.10, 6.11, 6.12 έχουμε τα αποτελέσματα από όλα τα μοντέλα για τις μετρικές Dice Coefficient, IoU, Precision και Recall αντίστοιχα.

Από τα αποτελέσματα μπορούμε να δούμε ότι σε γενικές γραμμές τα μοντέλα που έχουν εκπαιδευτεί στο σύνολο με υπερανάλυση έχουν καλύτερες μετρικές απόδοσης από τα μοντέλα που έχουν εκπαιδευτεί στο σύνολο χωρίς υπερανάλυση. Αξιοσημείωτες είναι οι υψηλές τιμές της μετρικής Recall, πράγμα που σημαίνει ότι τα μοντέλα αυτά μπορούν να εντοπίσουν αποτελεσματικά ένα μεγάλο μέρος των εικονοστοιχείων που έχουν σημειωθεί από τις ετικέτες ως δάση.

Συνάρτηση Σφάλματος	ResUNet-a χ.υ.	ResUNet-a μ.υ.	DLv3+ χ.υ.	DLv3+ μ.υ.	Swin T. χ.υ.	Swin T. μ.υ.
Cross Entropy	0.301	<b>0.405</b>	0.281	<b>0.388</b>	0.387	<b>0.462</b>
Dice	0.244	<b>0.391</b>	0.321	<b>0.332</b>	0.377	<b>0.526</b>
Hybrid	0.279	<b>0.397</b>	0.312	0.312	0.378	<b>0.54</b>
Tanimoto	0.265	<b>0.406</b>	0.219	<b>0.386</b>	0.25	<b>0.514</b>

**Πίνακας 6.9:** Τα πειραματικά αποτελέσματα όλων των μοντέλων για τη μετρική Dice Coefficient στο κοινό σύνολο δοκιμής. **ΣΗΜΕΙΩΣΗ:** Με χ.υ. σημειώνονται τα μοντέλα που προπονήθηκαν στο σύνολο χωρίς υπερανάλυση και με μ.υ. σημειώνονται τα μοντέλα που προπονήθηκαν στο σύνολο με υπερανάλυση.

Συνάρτηση Σφάλματος	ResUNet-a χ.υ.	ResUNet-a μ.υ.	DLv3+ χ.υ.	DLv3+ μ.υ.	Swin T. χ.υ.	Swin T. μ.υ.
Cross Entropy	0.258	<b>0.38</b>	0.242	<b>0.352</b>	0.24	<b>0.393</b>
Dice	0.255	<b>0.329</b>	<b>0.295</b>	0.27	0.232	<b>0.357</b>
Hybrid	0.227	<b>0.34</b>	<b>0.26</b>	0.232	0.233	<b>0.369</b>
Tanimoto	0.203	<b>0.353</b>	0.164	<b>0.326</b>	0.143	<b>0.346</b>

**Πίνακας 6.10:** Τα πειραματικά αποτελέσματα όλων των μοντέλων για τη μετρική IoU στο κοινό σύνολο δοκιμής.

Συνάρτηση Σφάλματος	ResUNet-a χ.υ.	ResUNet-a μ.υ.	DLv3+ χ.υ.	DLv3+ μ.υ.	Swin T. χ.υ.	Swin T. μ.υ.
Cross Entropy	<b>0.435</b>	0.426	<b>0.515</b>	0.387	<b>0.506</b>	0.475
Dice	<b>0.58</b>	0.342	<b>0.346</b>	0.286	<b>0.471</b>	0.401
Hybrid	<b>0.442</b>	0.363	<b>0.276</b>	0.24	<b>0.495</b>	0.426
Tanimoto	<b>0.447</b>	0.383	0.363	0.363	<b>0.515</b>	0.389

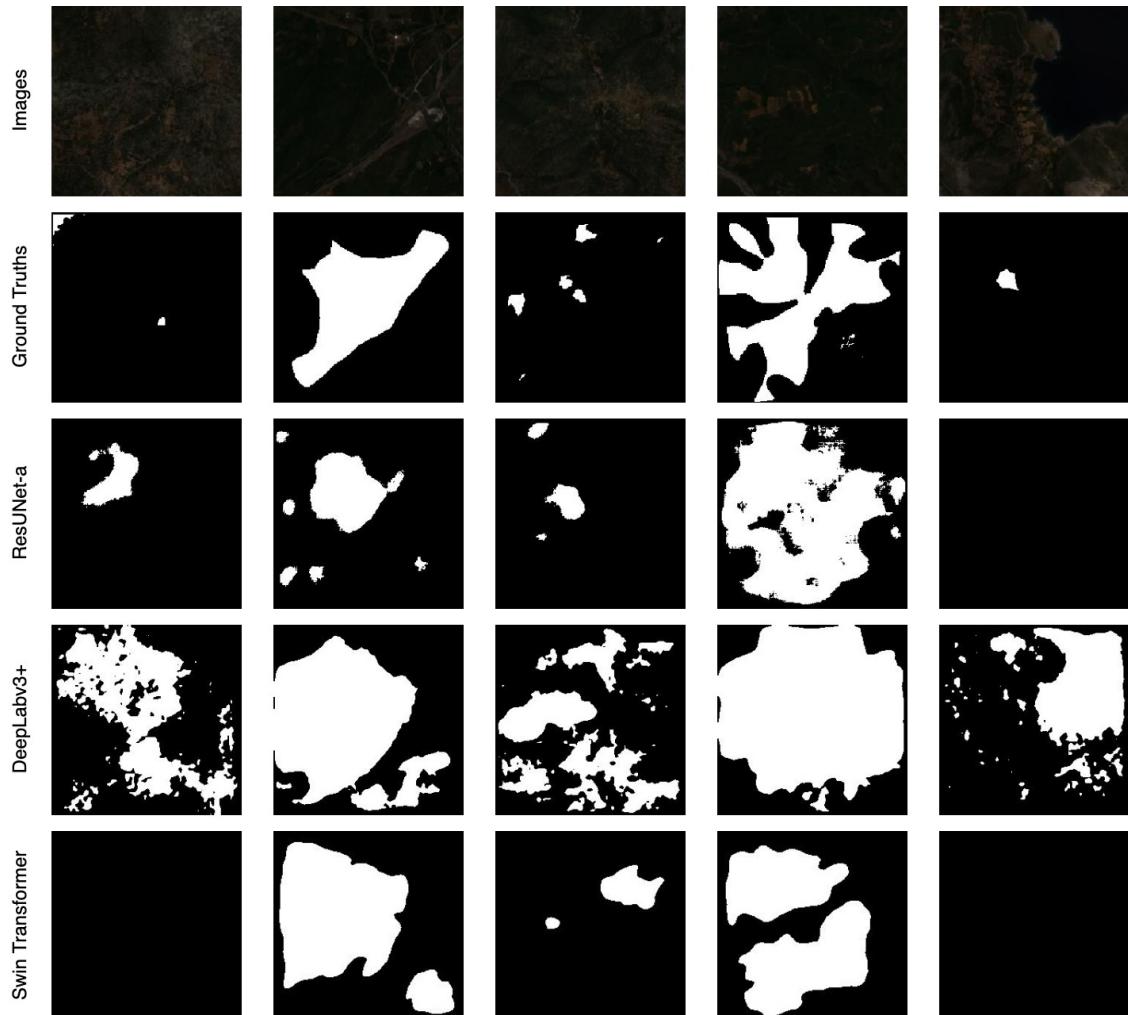
Πίνακας 6.11: Τα πειραματικά αποτελέσματα όμων των μοντέλων για τη μετρική Precision στο κοινό σύνολο δοκιμής.

Συνάρτηση Σφάλματος	ResUNet-a χ.υ.	ResUNet-a μ.υ.	DLv3+ χ.υ.	DLv3+ μ.υ.	Swin T. χ.υ.	Swin T. μ.υ.
Cross Entropy	0.388	<b>0.776</b>	0.314	<b>0.793</b>	0.314	<b>0.782</b>
Dice	0.313	<b>0.899</b>	0.668	<b>0.828</b>	0.314	<b>0.766</b>
Hybrid	0.319	<b>0.847</b>	0.82	<b>0.873</b>	0.306	<b>0.735</b>
Tanimoto	0.27	<b>0.822</b>	0.229	<b>0.76</b>	0.165	<b>0.758</b>

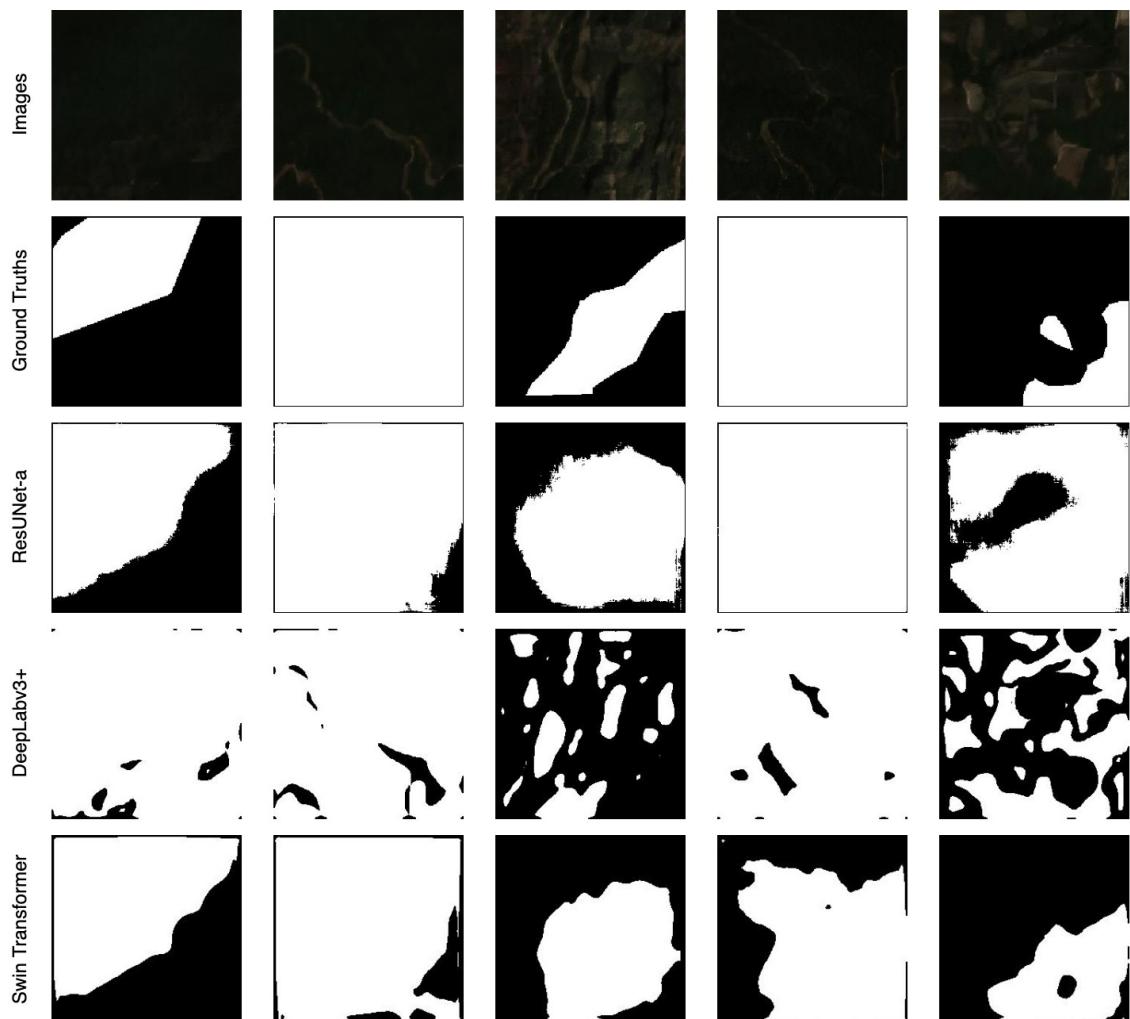
Πίνακας 6.12: Τα πειραματικά αποτελέσματα όμων των μοντέλων για τη μετρική Recall στο κοινό σύνολο δοκιμής.

## 6.2 Ποιοτική Αξιολόγηση Αποτελεσμάτων

Παρακάτω βλέπουμε τα αποτελέσματα της σημασιολογικής κατάτμησης που έκαναν τα μοντέλα στις παρακάτω τυχαία επιλεγμένες εικόνες. Στο πρώτο σχήμα έχουμε τα αποτελέσματα από τις εικόνες χωρίς υπερανάλυση και στο δεύτερο έχουμε τα αποτελέσματα από τις εικόνες με υπερανάλυση. Όλα τα εικονιζόμενα αποτελέσματα είναι από μοντέλα που έχουν εκπαιδευτεί με συνάρτηση σφάλματος Dice Loss.



Σχήμα 6.1: Εικόνες του *test set* χωρίς υπερανάλυση μαζί με τις ετικέτες τους και τις προβλέψεις των μοντέλων.



**Σχήμα 6.2:** Εικόνες του *test set* με υπερανάλυση μαζί με τις ετικέτες τους και τις προβλέψεις των μοντέλων.

### 6.3 Παρατηρήσεις ως προς τα αποτελέσματα

Σε αυτή την παράγραφο σχολιάζονται τα αποτελέσματα σε σχέση με το σύνολο δεδομένων που είχαμε και την υπερανάλυση εικόνας που εφαρμόστηκε.

#### 6.3.1 Ιδιαιτερότητες Δεδομένων

Οι δορυφορικές εικόνες του Sentinel που χρησιμοποιήθηκαν έχουν χωρική ανάλυση 10 m, πράγμα που σημαίνει ότι χαρακτηριστικά των δασών όπως τα δέντρα δεν απεικονίζονται με ευκρίνεια. Αυτό έχει ως συνέπεια να γίνεται δυσκολότερη η εξαγωγή των χαρακτηριστικών των δασών από τα μοντέλα και τα αποτελέσματα να είναι λιγότερο ακριβή. Επίσης, οι ετικέτες που χρησιμοποιήθηκαν προέρχονται από τη βάση δεδομένων του OpenStreetMap και κατά συνέπεια είναι στην πλειοψηφία τους συμπληρωμένες από τους χρήστες του. Κατά συνέπεια, οι ετικέτες δεν είναι πλήρεις και δεν παρουσιάζουν μεγάλη ακριβεία, καθώς κοιτώντας απλά τις εικόνες μπορεί κάποιος να ισχυριστεί ότι η ετικέτα σε πολλές περιπτώσεις καλύπτει μόνο ένα μικρό μέρος της εικονιζόμενης δασικής έκτασης.



(α') Δορυφορική εικόνα



(β') Ετικέτες OSM

**Σχήμα 6.3:** 'Ενα παράδειγμα δορυφορικής εικόνας με μια μη ακριβή ετικέτα της.

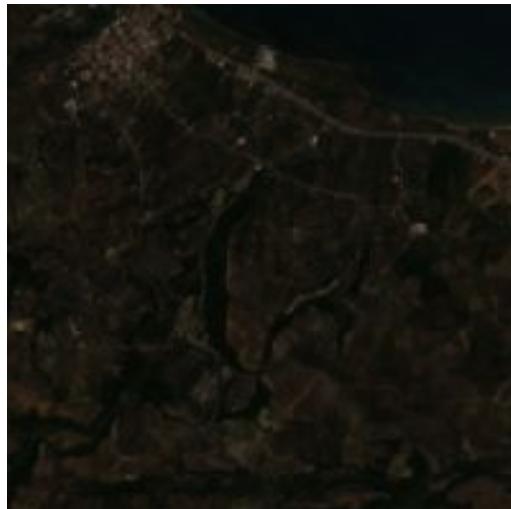
Αυτό έχει δύο αποτελέσματα. Το πρώτο, που είναι και το προφανές, είναι ότι επηρεάζονται αρνητικά οι μετρικές των αποτελεσμάτων των μοντέλων, καθώς σε πολλές περιπτώσεις τα μοντέλα προβλέπουν ως δασική έκταση μια μεγαλύτερη και οπτικώς πιο ακριβή έκταση από αυτή που προσδιορίζουν οι ετικέτες. Με βάση αυτό μπορούμε να ισχυριστούμε ότι η απόδοση των μοντέλων είναι μεγαλύτερη από αυτή που δείχνουν οι μετρικές και τα αποτελέσματα των μοντέλων μπορούν δυνητικά να αποτελέσουν μια βελτίωση στις ετικέτες της βάσης δεδομένων του OpenStreetMap. Το δεύτερο αποτέλεσμα είναι ότι με αυτές τις ετικέτες το σύνολο δεδομένων καθίσταται μη ισορροπημένο (imbalanced), καθώς όπως αναφέρθηκε, στις εικόνες χωρίς υπερανάλυση, σχεδόν μόλις το 16% των εικονοστοιχείων των εικόνων εμφανίζεται να ανήκει σε δασική έκταση, ενώ το ποσοστό αυτό θα μπορούσε να είναι μεγαλύτερο. Αυτό, όπως είναι αντιληπτό, εμποδίζει την σωστή εξαγωγή των χαρακτηριστικών των δασικών εκτάσεων και κατά συνέπεια έχει επιπτώσεις στην απόδοση των μοντέλων. Σε αυτό το σημείο

έχει σημαντική συμβολή η υπερανάλυση που εφαρμόστηκε στις εικόνες.

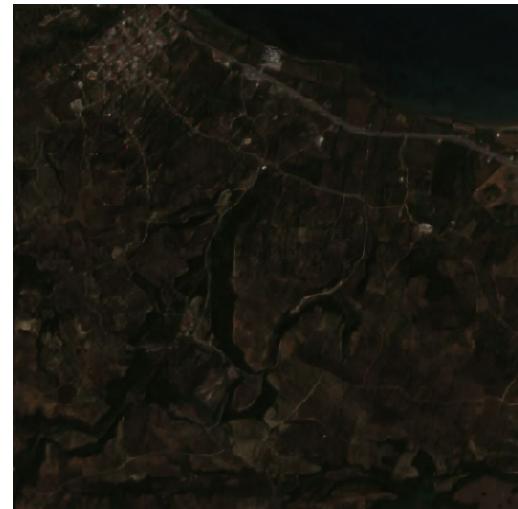
### 6.3.2 Επίδραση της Υπερανάλυσης Εικόνων

Όπως αναφέρθηκε στην προηγούμενη ενότητα, το σύνολο δεδομένων που είχαμε ήταν αρκετά μη ισορροπημένο και οι εικόνες του δεν είχαν ικανοποιητική χωρική ανάλυση. Τα δύο αυτά προβλήματα μετριάστηκαν με την υπερανάλυση εικόνας. Συγκεκριμένα, η ποιότητα των χαρακτηριστικών αυξήθηκε, το οποίο έκανε ευκολότερη την εξαγωγή τους. Εδώ πρέπει να επισημανθούν οι μεγάλες τιμές στις μετρικές του Recall στο κοινό σύνολο δοκιμής, που δείχνουν ότι τα μοντέλα που εκπαιδεύτηκαν στο σύνολο υψηλής υπερανάλυσης έμαθαν αποτελεσματικά τα χαρακτηριστικά των δασικών περιοχών του συνόλου εκπαίδευσης και μπορούν να εντοπίσουν ένα μεγάλο μέρος των ετικετών τόσο σε εικόνες χαμηλής όσο και σε εικόνες υψηλής ανάλυσης. Επιπλέον, καθώς οι εικόνες μετά την υπερανάλυση χωρίστηκαν ξανά σε εικόνες  $200 \times 200$  και φιλτραρίστηκαν με τον τρόπο που περιγράφηκε στην παράγραφο 5.4, το ποσοστό των εικονοστοιχείων που αντιστοιχούσαν σε δάση έφτασε το 39%, το οποίο είναι μια σημαντική βελτίωση στην εξισορρόπηση του συνόλου δεδομένων και κατά συνέπεια στα αποτελέσματα των μοντέλων.

Μπορούμε από αυτό να οδηγούμε στο συμπέρασμα, ότι η τεχνική της Υπερανάλυσης Εικόνας μπορεί να αποτελέσει ένα σημαντικό στάδιο προεπεξεργασίας, ιδιαίτερα σε σύνολα δεδομένων που έχουν εικόνες χαμηλής ανάλυσης και μη ισορροπημένες ετικέτες, για να βελτιωθεί η απόδοση των Νευρωνικών Δικτύων Σημασιολογικής Κατάτμησης.



(a') Χωρίς Υπερανάλυση



(β) Με Υπερανάλυση

Σχήμα 6.4: Η ίδια δορυφορική εικόνα χωρίς και με Υπερανάλυση



## **Μέρος IV**

### **Επίλογος**

---



# Συμπεράσματα και Μελλοντικές Επεκτάσεις

## 7.1 Συμπεράσματα

Σύμφωνα με την ανάλυση των πειραματικών αποτελεσμάτων μπορούμε να καταλήξουμε στα εξής. Όσον αφορά τα μοντέλα, παρατηρούμε ότι το DeepLabv3+, το οποίο δεν είναι σχεδιασμένο για σημασιολογική κατάτμηση σε εικόνες τηλεπισκόπησης, είχε τα πιο άσχημα αποτελέσματα. Ωστόσο, ειδικά στις εικόνες χωρίς τηλεπισκόπηση, μπορεί κάποιος να ισχυριστεί πως εντόπιζε περιοχές που δεν εντόπιζαν τα άλλα δύο μοντέλα, οι οποίες οπικώς θα μπορούσαν να χαρακτηριστούν ως δασικές, αλλά δεν ήταν σημειωμένες στις ετικέτες. Τα δύο καλύτερα μοντέλα για την σημασιολογική κατάτμηση ήταν από το Swin Transformer και το ResUNet-a. Ειδικότερα, το ResUNet-a που χρησιμοποιήθηκε ήταν μια απλοποιημένη εκδοχή των μοντέλων που προτάθηκαν στη σχετική βιβλιογραφία, λόγω των περιορισμών μνήμης του συστήματος. Πιθανότατα μια πιο περίπλοκη δομή, με περισσότερα blocks στον κωδικοποιητή και αποκωδικοποιητή και με περισσότερα φίλτρα, θα έχει ακόμα καλύτερη απόδοση. Όσο για τον Swin Transformer, αποδείχτηκε ότι η δομή του μετασχηματιστή βοήθησε πολύ στην εξαγωγή των χαρακτηριστικών των δασικών περιοχών και φαίνεται ότι η μέθοδος του κυλιόμενου παραθύρου προτιμάται και σε σύγχρονες εφαρμογές τμηματοποίησης σε δορυφορικές εικόνες. [88] [89] [90] Πιθανότατα και εδώ μια από τις πιο πολύπλοκες εκδοχές του, όπως τα Swin-S, Swin-B, Swin-L [13], να είναι ακόμα πιο αποδοτικά.

Όσον αφορά τις συναρτήσεις σφάλματος, σε αυτό το πρόβλημα τα καλύτερα αποτελέσματα τα έδωσαν η Dice Loss καθώς και η Hybrid Loss, ενώ και η Tanimoto Loss αποδείχτηκε πιο αποδοτική στις εικόνες με υπερανάλυση. Η Dice Loss προσπαθεί να αντιμετωπίσει το πρόβλημα της ανισορροπίας του συνόλου δεδομένων, ενώ η υβριδική συνάρτηση φιλοδοξεί να βοηθήσει τα μοντέλα να εντοπίσουν τόσο υψηλού όσο και χαμηλού επιπέδου χαρακτηριστικά. Η Tanimoto Loss, αν και παρουσιάζεται ως βελτίωση της Dice Loss, δεν βγάζει τόσο καλά αποτελέσματα στα σύνολα δεδομένων, αν και στις εικόνες με υπερανάλυση προσεγγίζει περισσότερο τις επιδόσεις που έχουν οι Dice Loss και Hybrid Loss.

Επίσης, από την δοκιμή όλων των μοντέλων σε ένα κοινό test set, παρατηρούμε ότι τα μοντέλα που εκπαιδεύτηκαν στο σύνολο υψηλής υπερανάλυσης απέδωσαν καλύτερα σε γενικές γραμμές από τα μοντέλα που εκπαιδεύτηκαν στο σύνολο χαμηλής ανάλυσης. Η υψηλή τιμή της μετρικής Recall που πέτυχαν δείχνει ότι η υπερανάλυση εικόνας τα βοήθησε να εξάγουν τα χαρακτηριστικά των δασικών περιοχών και συνεπώς τα μοντέλα αυτά είναι σε θέση να εντοπίσουν ένα μεγάλο μέρος των εικονοστοιχείων των ετικετών του OpenStreetMap.

Τέλος, αναφορικά με το σύνολο δεδομένων, όπως αναλύθηκε και στην παράγραφο 6.3.1, τόσο η χωρική ανάλυση όσο και η ανακρίβεια στις ετικέτες δυσχεραίνουν την εκπαίδευση των μοντέλων. Σε αυτό συμβάλλει και το γεγονός ότι σε πολλές περιπτώσεις τα όρια των δασικών περιοχών δεν είναι σαφώς καθορισμένα. Δεδομένων αυτών των περιορισμάν, μπορούμε να καταλήξουμε στο ότι η μετρική Recall αποτελεί πιο πιστή για την αξιολόγηση των μοντέλων, καθώς λαμβάνει υπόψη πόσα εικονοστοιχεία δασικών εκτάσεων από τις ετικέτες προβλέφθηκαν σωστά σε σχέση με τα εικονοστοιχεία που δεν προβλέφθηκαν και δεν λαμβάνει υπόψη τα false positives, που με πιο ακριβείς ετικέτες θα μπορούσαν να είναι true positives. Με βάση αυτό η απόδοση των μοντέλων που εκπαιδεύτηκαν στις εικόνες με υπερανάλυση κρίνεται ικανοποιητική και όπως έχει αναφερθεί, τα αποτελέσματά τους μπορούν να αποτελέσουν βελτιώσεις στις ετικέτες της βάσης δεδομένων του OpenStreetMap.

## 7.2 Μελλοντικές Επεκτάσεις

Μια επέκταση της εργασίας αυτής προτείνεται η χρήση αυτού του εργαλείου στο να εντοπίζει όχι μόνο δασικές εκτάσεις, αλλά και εκτάσεις άλλων κατηγοριών που βρίσκονται στη βάση δεδομένων του OpenStreetMap, όπως αγροτικές ή κατοικίσιμες εκτάσεις. Ο τρόπος που δημιουργήθηκε το σύνολο δεδομένων είναι ανεξάρτητος από το αντικείμενο προς εντοπισμό και άρα με ελάχιστες τροποποιήσεις, η διαδικασία θα μπορούσε να επαναληφθεί για χιλιάδες από τα χαρακτηριστικά που περιλαμβάνονται στο OpenStreetMap. Θα μπορούσε λοιπόν, αυτή η εργασία να μετατραπεί σε ένα παρματεροποιήσιμο εργαλείο, στο οποίο ο χρήστης θα δίνει ως είσοδο το αντικείμενο που θέλει να εντοπίσει και εκείνο θα δημιουργεί το σύνολο δεδομένων και θα εκπαιδεύει τα μοντέλα σημασιολογικής κατάτμησης. Τέλος, τη διαδικασία αυτή θα μπορούσε να την κάνει όχι μόνο για ένα αντικείμενο, αλλά για πολλά διαφορετικά ταυτόχρονα.

## Βιβλιογραφία

---

- [1] *Tutorial: Fundamentals of Remote Sensing*, 2016.
- [2] Nicholas C. Coops και Thoreau Rory Tooke. *Introduction to Remote Sensing*. Springer New York, New York, NY, 2017.
- [3] *MSI Instrument - Sentinel-2 MSI Technical Guide - Sentinel Online*. <https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-2-msi/msi-instrument>. Ημερομηνία πρόσθασης: 25-5-2022.
- [4] *CS231n Convolutional Neural Networks for Visual Recognition*. <https://cs231n.github.io/understanding-cnn/>. Ημερομηνία πρόσθασης 9-8-2022.
- [5] Jonathan Long, Evan Shelhamer και Trevor Darrell. *Fully Convolutional Networks for Semantic Segmentation*, 2014.
- [6] Olaf Ronneberger, Philipp Fischer και Thomas Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. CoRR, abs/1505.04597, 2015.
- [7] Bruno Artacho και Andreas Savakis. *Waterfall Atrous Spatial Pooling Architecture for Efficient Semantic Segmentation*. Sensors, 19(24):5361, 2019.
- [8] Liang Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy και Alan L. Yuille. *DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs*, 2016.
- [9] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang και Wenzhe Shi. *Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network*, 2016.
- [10] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Chen Change Loy, Yu Qiao και Xiaou Tang. *ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks*, 2018.
- [11] Foivos I. Diakogiannis, François Waldner, Peter Caccetta και Chen Wu. *ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data*. ISPRS Journal of Photogrammetry and Remote Sensing, 162:94–114, 2020.
- [12] Liang Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff και Hartwig Adam. *Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation*, 2018.

- [13] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin και Baining Guo. *Swin Transformer: Hierarchical Vision Transformer using Shifted Windows*, 2021.
- [14] Tete Xiao, Yingcheng Liu, Bolei Zhou, Yuning Jiang και Jian Sun. *Unified Perceptual Parsing for Scene Understanding*, 2018.
- [15] Richard A Houghton. *Tropical deforestation as a source of greenhouse gas emissions. Tropical deforestation and climate change*, 13, 2005.
- [16] Satoru Miura, Michael Amacher, Thomas Hofer, Jesús San-Miguel-Ayanz, Ernawati και Richard Thackway. *Protective functions and ecosystem services of global forests in the past quarter-century*. *Forest Ecology and Management*, 352:35–46, 2015. Changes in Global Forest Resources from 1990 to 2015.
- [17] Daisy Núñez, Laura Nahuelhual και Carlos Oyarzún. *Forests and water: The value of native temperate forests in supplying water for human consumption*. *Ecological Economics*, 58(3):606–616, 2006.
- [18] Erika Romijn, Celso B. Lantican, Martin Herold, Erik Lindquist, Robert Ochieng, Arief Wijaya, Daniel Murdiyarsa και Louis Verchot. *Assessing change in national forest monitoring capacities of 99 tropical countries*. *Forest Ecology and Management*, 352:109–123, 2015. Changes in Global Forest Resources from 1990 to 2015.
- [19] Veronique De Sy, Martin Herold, Frédéric Achard, Gregory P Asner, Alex Held, Josef Kellndorfer και Jan Verbesselt. *Synergies of multiple remote sensing data sources for REDD+ monitoring*. *Current Opinion in Environmental Sustainability*, 4(6):696–706, 2012. 4/6 Climate systems.
- [20] Ricardo Dal Molin και Paola Rizzoli. *Potential of Convolutional Neural Networks for Forest Mapping Using Sentinel-1 Interferometric Short Time Series*. *Remote Sensing*, 14(6), 2022.
- [21] Ahmad Alzu’bi και Lujain Alsmadi. *Monitoring deforestation in Jordan using deep semantic segmentation with satellite imagery*. *Ecological Informatics*, 70:101745, 2022.
- [22] Guoli Zhang, Ming Wang και Kai Liu. *Forest Fire Susceptibility Modeling Using a Convolutional Neural Network for Yunnan Province of China*. *International Journal of Disaster Risk Science*, 10(3):386–403, 2019.
- [23] Antonio Mazza, Francescopaolo Sica, Paola Rizzoli και Giuseppe Scarpa. *TanDEM-X Forest Mapping using Convolutional Neural Networks*. *Remote Sensing*, 11, 2019.
- [24] Anders U. Waldeland, Øivind Due Trier και Arnt Børre Salberg. *Forest mapping and monitoring in Africa using Sentinel-2 data and deep learning*. *International Journal of Applied Earth Observation and Geoinformation*, 111:102840, 2022.

- [25] Dimitrios Kollias, Athanasios Tagaris, Andreas Stafylopatis, Stefanos Kollias και Georgios Tagaris. *Deep neural architectures for prediction in healthcare*. *Complex & Intelligent Systems*, 4(2):119–131, 2018.
- [26] Athanasios Tagaris, Dimitrios Kollias και Andreas Stafylopatis. *Assessment of Parkinson’s disease based on deep neural networks*. *International Conference on Engineering Applications of Neural Networks*, σελίδες 391–403. Springer, 2017.
- [27] Athanasios Tagaris, Dimitrios Kollias, Andreas Stafylopatis, Georgios Tagaris και Stefanos Kollias. *Machine learning for neurodegenerative disorder diagnosis—survey of practices and launch of benchmark dataset*. *International Journal on Artificial Intelligence Tools*, 27(03):1850011, 2018.
- [28] Ilianna Kollia, Andreas Georgios Stafylopatis και Stefanos Kollias. *Predicting Parkinson’s disease using latent information extracted from deep neural networks*. *2019 International Joint Conference on Neural Networks (IJCNN)*, σελίδες 1–8. IEEE, 2019.
- [29] James Wingate, Ilianna Kollia, Luc Bidaut και Stefanos Kollias. *Unified deep learning approach for prediction of Parkinson’s disease*. *IET Image Processing*, 14(10):1980–1989, 2020.
- [30] Dimitrios Kollias, Anastasios Arsenos, Levon Soukissian και Stefanos Kollias. *Mia-cov19d: Covid-19 detection through 3-d chest ct image analysis*. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, σελίδες 537–544, 2021.
- [31] Dimitrios Kollias, Anastasios Arsenos και Stefanos Kollias. *Ai-mia: Covid-19 detection & severity analysis through medical imaging*. *arXiv preprint arXiv:2206.04732*, 2022.
- [32] Anastasios Arsenos, Dimitrios Kollias και Stefanos Kollias. *A Large Imaging Database and Novel Deep Neural Architecture for Covid-19 Diagnosis*. *2022 IEEE 14th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)*, σελίδες 1–5. IEEE, 2022.
- [33] Dimitrios Kollias, Miao Yu, Athanasios Tagaris, Georgios Leontidis, Andreas Stafylopatis και Stefanos Kollias. *Adaptation and contextualization of deep neural network models*. *2017 IEEE symposium series on computational intelligence (SSCI)*, σελίδες 1–8. IEEE.
- [34] D Kollias, N Bouas, Y Vlaxos, V Brillakis, M Seferis, I Kollia, L Sukissian, J Wingate και S Kollias. *Deep Transparent Prediction through Latent Representation Analysis*. *arXiv preprint arXiv:2009.07044*, 2020.
- [35] Dimitris Kollias, Y Vlaxos, M Seferis, Ilianna Kollia, Levon Sukissian, James Wingate και S Kollias. *Transparent adaptation in deep medical image diagnosis*. *International Workshop on the Foundations of Trustworthy AI Integrating Learning, Optimization and Reasoning*, σελίδες 251–267. Springer, 2020.

- [36] Fabio De Sousa Ribeiro, Francesco Caliva, Mark Swainson, Kjartan Gudmundsson, Georgios Leontidis και Stefanos Kollias. *Deep bayesian self-training*. *Neural Computing and Applications*, 32(9):4275–4291, 2020.
- [37] Fabio De Sousa Ribeiro, Georgios Leontidis και Stefanos Kollias. *Capsule routing via variational bayes*. *Proceedings of the AAAI Conference on Artificial Intelligence*, τόμος 34, σελίδες 3749–3756, 2020.
- [38] Fabio De Sousa Ribeiro, Georgios Leontidis και Stefanos Kollias. *Introducing routing uncertainty in capsule networks*. *Advances in Neural Information Processing Systems*, 33:6490–6502, 2020.
- [39] Nikolaos Simou και Stefanos Kollias. *Fire: A fuzzy reasoning engine for imprecise knowledge*. Citeseer.
- [40] Francesco Caliva, Fabio Sousa De Ribeiro, Antonios Mylonakis, Christophe Demaziere, Paolo Vinai, Georgios Leontidis και Stefanos Kollias. *A deep learning approach to anomaly detection in nuclear reactors*. *2018 International joint conference on neural networks (IJCNN)*, σελίδες 1–8. IEEE, 2018.
- [41] Bashar Alhnaity, Stefanos Kollias, Georgios Leontidis, Shouyong Jiang, Bert Schamp και Simon Pearson. *An autoencoder wavelet based deep neural network with attention mechanism for multi-step prediction of plant growth*. *Information Sciences*, 560:35–50, 2021.
- [42] Bashar Alhnaity, Simon Pearson, Georgios Leontidis και Stefanos Kollias. *Using deep learning to predict plant growth and yield in greenhouse environments*. *International Symposium on Advanced Technologies and Management for Innovative Greenhouses: GreenSys2019* 1296, σελίδες 425–432, 2019.
- [43] Stefanos Kollias, Miao Yu, James Wingate, Aiden Durrant, Georgios Leontidis, Georgios Alexandridis, Andreas Stafylopatis, Antonios Mylonakis, Paolo Vinai και Christophe Demaziere. *Machine learning for analysis of real nuclear plant data in the frequency domain*. *Annals of Nuclear Energy*, 177:109293, 2022.
- [44] Andreas Psaroudakis και Dimitrios Kollias. *MixAugment & Mixup: Augmentation Methods for Facial Expression Recognition*. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, σελίδες 2367–2375, 2022.
- [45] Dimitrios Kollias και Stefanos Zafeiriou. *Training deep neural networks with different datasets in-the-wild: The emotion recognition paradigm*. *2018 International Joint Conference on Neural Networks (IJCNN)*, σελίδες 1–8. IEEE, 2018.
- [46] Dimitrios Kollias και Stefanos Zafeiriou. *Va-stargan: Continuous affect generation*. *International Conference on Advanced Concepts for Intelligent Vision Systems*, σελίδες 227–238. Springer, 2020.

- [47] G Caridakis, A Raouzaiou, K Karpouzis και S Kollias. *Synthesizing Gesture Expressivity Based on Real Sequences*. Workshop Programme, τόμος 10, σελίδα 19.
- [48] Phivos Mylonas, Evangelos Spyrou, Yannis Avirthis και Stefanos Kollias. *Using visual context and region semantics for high-level concept detection*. *IEEE Transactions on Multimedia*, 11(2):229–243, 2009.
- [49] Stefanos Kollias και Dimitris Anastassiou. *A unified neural network approach to digital image halftoning*. *IEEE Transactions on signal processing*, 39(4):980–984, 1991.
- [50] Paraskevi Tzouveli, Andreas Schmidt, Michael Schneider, Antonis Symvonis και Stefanos Kollias. *Adaptive reading assistance for the inclusion of students with dyslexia: The AGENT-DYSL approach*. *2008 Eighth IEEE International Conference on Advanced Learning Technologies*, σελίδες 167–171. IEEE, 2008.
- [51] K. Tempfli, G.C. Huurneman, W.H. Bakker, L.L.F. Janssen, W.F. Feringa, A.S.M. Gieske, K.A. Grabmaier, C.A. Hecker, J.A. Horn, N. Kerle, F.D. van der Meer, G.N. Parodi, C. Pohl, C.V. Reeves, F.J.A. van Ruitenbeek, E.M. Schetselaar, M.J.C. Weir, E. Westinga και T. Woldai. *Principles of remote sensing : an introductory textbook*. ITC Educational Textbook Series. International Institute for Geo-Information Science and Earth Observation, Netherlands, 2009.
- [52] Shu Kun Lin. *Introduction to Remote Sensing*. Fifth Edition. Edited by James B. Campbell and Randolph H. Wynne, The Guilford Press, 2011; 662 pages, ISBN 978-1-60918-176-5. *Remote Sensing*, vol. 5, issue 1, pp. 282-283, 5:282–283, 2013.
- [53] All satellites - ITC Satellites and Sensors database. <https://webapps.itc.utwente.nl/sensor/default.aspx?view=allsatellites>. Ημερομηνία πρόσθασης: 22-5-2022.
- [54] Scott Madry. *Introduction and History of Space Remote Sensing*, σελίδες 657-666. 2013.
- [55] Copernicus in Detail. <https://www.copernicus.eu/en/about-copernicus/copernicus-detail>. Ημερομηνία πρόσθασης: 24-5-2022.
- [56] Missions - Sentinel Online - Sentinel Online. <https://sentinels.copernicus.eu/web/sentinel/missions>. Ημερομηνία πρόσθασης: 24-5-2022.
- [57] Arianespace wins new contract to launch Sentinel-1C observation satellite on board Vega-C - Arianespace. <https://www.arianespace.com/press-release/arianespace-wins-new-contract-to-launch-sentinel-1c-observation-satellite-on-board-vega-c/>. Ημερομηνία πρόσθασης: 24-5-2022.
- [58] Gearing up for third Sentinel-2 satellite. [https://www.esa.int/Applications/Observing\\_the\\_Earth/Copernicus/Sentinel-2/Gearing\\_up\\_for\\_third\\_Sentinel-2\\_satellite](https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-2/Gearing_up_for_third_Sentinel-2_satellite). Ημερομηνία πρόσθασης: 24-5-2022.
- [59] Sentinel-3 Infographic. [https://sentinels.copernicus.eu/documents/247904/4757086/ESA\\_Sentinel-3+infographic\\_1.pdf](https://sentinels.copernicus.eu/documents/247904/4757086/ESA_Sentinel-3+infographic_1.pdf). Ημερομηνία πρόσθασης: 24-5-2022.

- [60] *Sentinel-4 - Missions - Sentinel Online - Sentinel Online*. <https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-4>. Ημερομηνία πρόσθασης: 24-5-2022.
- [61] *Sentinel-6 Overview - Missions - Sentinel Online - Sentinel Online*. <https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-6/overview>. Ημερομηνία πρόσθασης: 24-5-2022.
- [62] SHUET. *Sentinel-2 User Handbook*. [https://sentinels.copernicus.eu/documents/247904/685211/Sentinel-2\\_User\\_Handbook](https://sentinels.copernicus.eu/documents/247904/685211/Sentinel-2_User_Handbook), 2015.
- [63] *Radiometric - Resolutions - Sentinel-2 MSI - User Guides - Sentinel Online*. <https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi/resolutions/radiometric>. Ημερομηνία πρόσθασης: 25-5-2022.
- [64] *What is Computer Vision? | IBM*. <https://www.ibm.com/topics/computer-vision>. Ημερομηνία πρόσθασης 8-6-2022.
- [65] *27+ Most Popular Computer Vision Applications in 2022*. <https://viso.ai/applications/computer-vision-applications/>. Ημερομηνία πρόσθασης 8-6-2022.
- [66] Jia Deng, Wei Dong, Richard Socher, Li Jia Li, Kai Li και Li Fei-Fei. *ImageNet: A large-scale hierarchical image database*. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, σελίδες 248-255, 2009.
- [67] Alex Krizhevsky, Ilya Sutskever και Geoffrey E Hinton. *ImageNet Classification with Deep Convolutional Neural Networks*. *Advances in Neural Information Processing Systems*F. Pereira, C.J. Burges, L. Bottou και K.Q. Weinberger, επιμελητές, τόμος 25. Curran Associates, Inc., 2012.
- [68] Sergey Ioffe και Christian Szegedy. *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*. CoRR, abs/1502.03167, 2015.
- [69] Andreas Veit, Michael Wilber και Serge Belongie. *Residual Networks Behave Like Ensembles of Relatively Shallow Networks*, 2016.
- [70] Honggang Chen, Xiaohai He, Linbo Qing, Yuanyuan Wu, Chao Ren και Ce Zhu. *Real-World Single Image Super-Resolution: A Brief Review*, 2021.
- [71] Zhengxiong Luo, Yan Huang, Shang Li, Liang Wang και Tieniu Tan. *Unfolding the Alternating Optimization for Blind Super Resolution*, 2020.
- [72] Pengxu Wei, Ziwei Xie, Hannan Lu, Zongyuan Zhan, Qixiang Ye, Wangmeng Zuo και Liang Lin. *Component Divide-and-Conquer for Real-World Image Super-Resolution*. 2020.
- [73] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville και Yoshua Bengio. *Generative Adversarial Networks*, 2014.

- [74] Liang Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy και Alan L. Yuille. *DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs*, 2016.
- [75] Liang Chieh Chen, George Papandreou, Florian Schroff και Hartwig Adam. *Rethinking Atrous Convolution for Semantic Image Segmentation*, 2017.
- [76] Salman H. Khan, Muzammal Naseer, Munawar Hayat, Syed Waqas Zamir, Fahad Shahbaz Khan και Mubarak Shah. *Transformers in Vision: A Survey*. CoRR, abs/2101.01169, 2021.
- [77] Yixuan Wei, Han Hu, Zhenda Xie, Zheng Zhang, Yue Cao, Jianmin Bao, Dong Chen και Baining Guo. *Contrastive Learning Rivals Masked Image Modeling in Fine-tuning via Feature Distillation*, 2022.
- [78] Ze Liu, Han Hu, Yutong Lin, Zhuliang Yao, Zhenda Xie, Yixuan Wei, Jia Ning, Yue Cao, Zheng Zhang, Li Dong, Furu Wei και Baining Guo. *Swin Transformer V2: Scaling Up Capacity and Resolution*. 2021.
- [79] Jitesh Jain, Anukriti Singh, Nikita Orlov, Zilong Huang, Jiachen Li, Steven Walton και Humphrey Shi. *SeMask: Semantically Masked Transformers for Semantic Segmentation*, 2021.
- [80] Alexia Jolicoeur-Martineau. *The relativistic discriminator: a key element missing from standard GAN*, 2018.
- [81] Francesco Cardinaleet al. *ISR*. <https://github.com/idealo/image-super-resolution>, 2018.
- [82] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang και Jiaya Jia. *Pyramid Scene Parsing Network*, 2016.
- [83] Kaiming He, Xiangyu Zhang, Shaoqing Ren και Jian Sun. *Deep Residual Learning for Image Recognition*, 2015.
- [84] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang και Jiaya Jia. *Pyramid Scene Parsing Network*, 2016.
- [85] Huimin Huang, Lanfen Lin, Ruofeng Tong, Hongjie Hu, Qiaowei Zhang, Yutaro Iwamoto, Xianhua Han, Yen Wei Chen και Jian Wu. *UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation*. ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), σελίδες 1055–1059, 2020.
- [86] Nabila Abraham και Naimul Mefraz Khan. *A Novel Focal Tversky loss function with improved Attention U-Net for lesion segmentation*. CoRR, abs/1810.07842, 2018.
- [87] MM Segmentation Contributors. *MM Segmentation: OpenMMLab Semantic Segmentation Toolbox and Benchmark*. <https://github.com/open-mmlab/mmsegmentation>, 2020.

- [88] Hang Gong, Tingkui Mu, Qiuxia Li, Haishan Dai, Chunlai Li, Zhiping He, Wenjing Wang, Feng Han, Abudusalamu Tuniyazi, Haoyang Li, Xuechan Lang, Zhiyuan Li και Bin Wang. *Swin-Transformer-Enabled YOLOv5 with Attention Mechanism for Small Object Detection on Satellite Images*. *Remote Sensing*, 14(12), 2022.
- [89] Xingjian Gu, Sizhe Li, Shougang Ren, Hengbiao Zheng, Chengcheng Fan και Huan-liang Xu. *Adaptive enhanced swin transformer with U-net for remote sensing image segmentation*. *Computers and Electrical Engineering*, 102:108223, 2022.
- [90] Linus Scheibenreif, Joëlle Hanna, Michael Mommert και Damian Borth. *Self-supervised Vision Transformers for Land-cover Segmentation and Classification*. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, σελίδες 1421–1430, 2022.