



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
Δ.Π.Μ.Σ. ΕΠΙΣΤΗΜΗ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ  
ΕΡΓΑΣΤΗΡΙΟ ΣΥΣΤΗΜΑΤΩΝ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ ΚΑΙ  
ΜΑΘΗΣΗΣ

Μεταπτυχιακή Διπλωματική Εργασία  
ΑΝΙΧΝΕΥΣΗ ΜΕΤΑΒΟΛΩΝ ΣΕ ΔΟΡΥΦΟΡΙΚΕΣ  
ΕΙΚΟΝΕΣ ΜΕ ΤΗΝ ΧΡΗΣΗ ΜΕΘΟΔΩΝ ΒΑΘΙΑΣ  
ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ

Έλλη Κάρκαλου  
Διπλ. Αγρονόμος και Τοπογράφος Μηχανικός Ε.Μ.Π., ΔΠΜΣ Γεωπληροφορική

Αθήνα, Μάρτιος 2022



NATIONAL TECHNICAL UNIVERSITY OF ATHENS  
MSc. DATA SCIENCE AND MACHINE LEARNING  
ARTIFICIAL INTELLIGENCE AND LEARNING SYSTEMS  
LABORATORY

Master Thesis

# CHANGE DETECTION IN SATELLITE IMAGES WITH THE USE OF DEEP LEARNING METHODS

Elli Karkalou

Dipl-Eng, Rural and Surveying Engineering, MSc. Geoinformatics

Athens, March 2022



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
Δ.Π.Μ.Σ. ΕΠΙΣΤΗΜΗ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ  
ΕΡΓΑΣΤΗΡΙΟ ΣΥΣΤΗΜΑΤΩΝ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ ΚΑΙ  
ΜΑΘΗΣΗΣ

Μεταπτυχιακή Διπλωματική Εργασία  
**ΑΝΙΧΝΕΥΣΗ ΜΕΤΑΒΟΛΩΝ ΣΕ ΔΟΡΥΦΟΡΙΚΕΣ  
ΕΙΚΟΝΕΣ ΜΕ ΤΗΝ ΧΡΗΣΗ ΜΕΘΟΔΩΝ ΒΑΘΙΑΣ  
ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ**

Έλλη Κάρκαλου  
Διπλ. Αγρονόμος και Τοπογράφος Μηχανικός Ε.Μ.Π., ΔΠΜΣ Γεωπληροφορική

Τριμελής εξεταστική επιτροπή:

Γ. Στάμου

Κ.Καράντζαλος

Μ.Βακαλοπούλου

.....

.....

.....

Καθηγητής Ε.Μ.Π.  
Επιβλέπων

Αναπλ.Καθηγητής Ε.Μ.Π.

Assistant Professor CentraleSupélec

Αθήνα, Μάρτιος 2022

Έλλη Κάρκαλου

Αγρονόμος και Τοπογράφος Μηχανικός Ε.Μ.Π.

Copyright © All rights reserved. Έλλη Κάρκαλου, 2022

Με επιφύλαξη παντός δικαιώματος.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ' ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς το συγγραφέα. Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν το συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.



# ΠΕΡΙΛΗΨΗ

Σε αυτή την μεταπτυχιακή διπλωματική εργασία αναπτύχθηκαν, εφαρμόστηκαν και αξιολογήθηκαν αρχιτεκτονικές βαθιών νευρωνικών δικτύων για την ανίχνευση μεταβολών σε δορυφορικές εικόνες. Η πρώτη εφαρμογή που αναπτύχθηκε διερευνά και αξιολογεί την απόδοση των μπλοκ LSTM σε μια αρχιτεκτονική με πλήρη συνελκτικά επίπεδα της μορφής U-Net που προτάθηκε από τους Papadomanolaki et al (2019). Ειδικότερα, συγκρίθηκε η πλήρης εκδοχή του δικτύου με κάθε πιθανή περίπτωση στην οποία λείπει από ένα μπλοκ LSTM έως και τα 4 από τα 5. Η δεύτερη εφαρμογή αφορά την χρήση μιας αρχιτεκτονικής με διεσταλμένα (dilated) φίλτρα (Dilated+LSTM) που προτάθηκε από τους Vakaloroulou et al. (2019) για το αντικείμενο της αντιστοίχισης εικόνων. Στην παρούσα εργασία χρησιμοποιήθηκε για το αντικείμενο της ανίχνευσης μεταβολών και διενεργήθηκε σύγκριση της απόδοσής της, σε σχέση με το δίκτυο L-UNet των Papadomanolaki et al (2019). Οι δύο πρώτες εφαρμογές έκαναν χρήση του σετ εικόνων OSCD (Onera Satellite Change Detection) των Daudt et al. (2018). Το εν λόγω σετ εικόνων διαθέτει 24 ζεύγη γεωμετρικά διορθωμένων εικόνων, από τις οποίες χρησιμοποιούνται 14 ζεύγη για εκπαίδευση και επικύρωση και 10 ζεύγη για αξιολόγηση. Η τρίτη υλοποίηση αφορά την εφαρμογή των δικτύων L-UNet και Dilated+LSTM για πρώτη φορά στο απαιτητικό σετ εικόνων LEVIR-CD (Chen & Shi, 2020). Συνολικά χρησιμοποιήθηκαν 115 ζεύγη γεωμετρικά διορθωμένων εικόνων για εκπαίδευση και επικύρωση και 128 ζεύγη εικόνων για αξιολόγηση. Για την συγκριτική αξιολόγηση των μοντέλων χρησιμοποιήθηκαν οι διαθέσιμοι αληθείς χάρτες μεταβολών. Πιο συγκεκριμένα, υπολογίστηκαν πίνακες σύγχυσης και παράχθηκαν εικόνες διαφορών από την σύγκριση των παραγόμενων χαρτών μεταβολών με τους αληθείς. Η εργασία ολοκληρώνεται με τα τελικά συμπεράσματα και ορισμένες προτάσεις για πιθανά θέματα μελλοντικής έρευνας. Για την υλοποίηση των παραπάνω διαδικασιών συντάχθηκε κώδικας στην γλώσσα προγραμματισμού Python με την βοήθεια της βιβλιοθήκης PyTorch.

# ABSTRACT

In this Master Thesis, architectures of deep neural networks for change detection in satellite images have been implemented and evaluated. The scope of the first approach is the investigation and evaluation of the performance of LSTM blocks used in a U-Net architecture with fully convolutional levels, proposed by Papadomanolaki et al. (2019). More specifically, the complete version of the model was compared against every version with fewer LSTM blocks, varying from one up to four LSTM blocks. The second task is about the use of a dilated architecture named Dilated+LSTM, which was first proposed by Vakalopoulou et al. (2019) for the image registration task. In the present work it was used for the purpose of change detection and the model's performance was compared against the L-UNet model of Papadomanolaki et al. (2019). These first two tasks made use of the OSCD (Onera Satellite Change Detection) dataset (Daudt et al., 2018). This dataset provides 24 pairs of co-registered images, 14 of which are used for training and validation and 10 of which are used for testing. As a third task, the two aforementioned models (L-UNet and Dilated+LSTM) were applied for the first time to the demanding dataset LEVIR-CD (Chen & Shi, 2020). In total 115 pairs of co-registered images were used for training and validation, whereas 128 image pairs were used for testing. Quantitative and qualitative comparison between the different models were conducted based on the ground truth change maps provided by the two datasets. In particular, confusion matrices were computed and difference images were produced by comparing the obtained change maps with the ground truth ones. Our main conclusions regarding the overall assessment and performance of the validated frameworks are presented as well as certain suggestions for further possible extensions of this work. The techniques and aforementioned frameworks were implemented in Python exploiting the PyTorch library.

# ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα κατ'αρχάς να ευχαριστήσω τον επιβλέποντά μου καθηγητή Γιώργο Στάμου για την βοήθεια και τον χρόνο που διέθεσε καθ'όλη την διάρκεια εκπόνησης της μεταπτυχιακής διπλωματικής μου εργασίας, καθώς επίσης για την υποστήριξη και την εμπιστοσύνη που μου έδειξε σε όλη την διάρκεια της συνεργασίας μας.

Ακόμη, θα ήθελα να ευχαριστήσω την Δρ.Μαρία Βακαλοπούλου, για την βοήθεια και την καθοδήγηση που μου παρείχε, τόσο για τα πειράματα, όσο και κατά την διάρκεια της συγγραφής. Η ολοκλήρωση της εργασίας θα ήταν αδύνατη χωρίς την συμβολή της. Επιπλέον, ευχαριστώ τον αναπληρωτή καθηγητή Κωνσταντίνο Καραντζαλο για τις χρήσιμες παρατηρήσεις του σχετικά με το κείμενο της εργασίας. Επίσης, ευχαριστώ θερμά την υποψήφια διδάκτορα Μαρία Παπαδομανωλάκη για την βοήθεια που μου παρείχε σε πρακτικά ζητήματα που άπτονταν της εργασίας. Ακόμη, θα ήθελα να ευχαριστήσω τον ομότιμο πλέον καθηγητή κύριο Γεώργιο Καρρά, που μου είχε διδάξει όταν βρισκόμουν σε προπτυχιακό επίπεδο τους κανόνες για την συγγραφή μιας ερευνητικής εργασίας.

Τέλος, θα ήθελα να ευχαριστήσω την οικογένειά μου, τους φίλους και συναδέλφους για την συμπαράσταση που μου έδειξαν.

# ΠΕΡΙΕΧΟΜΕΝΑ

<b>ΠΕΡΙΛΗΨΗ</b>	<b>i</b>
<b>ABSTRACT</b>	<b>ii</b>
<b>ΕΥΧΑΡΙΣΤΙΕΣ</b>	<b>iii</b>
<b>ΠΕΡΙΕΧΟΜΕΝΑ</b>	<b>iv</b>
<b>1. ΕΙΣΑΓΩΓΗ</b>	<b>1</b>
1.1 Γενικότερες Έννοιες	1
1.2 Αντικείμενο και Στόχος της εργασίας	2
1.3 Συνεισφορά	3
1.4 Δομή της εργασίας	3
<b>2. ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΑΝΑΣΚΟΠΗΣΗ</b>	<b>4</b>
2.1 Κατηγοριοποίηση αλγορίθμων ανίχνευσης μεταβολών	4
2.2 Βασικές αρχιτεκτονικές για την σημασιολογική κατάτμηση εικόνας	5
2.2.1 Πλήρη Συνελικτικά Δίκτυα (Fully Convolutional Networks-FCN)	5
2.2.2 U-Net	6
2.2.3 Dilated networks	7
2.3. Μέθοδοι ανίχνευσης μεταβολών (change detection)	11
2.3.1 Μέθοδοι ανίχνευσης μεταβολών χωρίς βαθιά μηχανική μάθηση	11
2.3.2 Πρώτες αρχιτεκτονικές με νευρωνικά δίκτυα	12
2.3.3 Αρχιτεκτονικές με Συνελικτικά LSTM	14
2.3.4 Αρχιτεκτονικές με μηχανισμούς προσοχής και transformers	19
2.3.5 Multi-task αρχιτεκτονικές	21
2.4 Διαθέσιμα σετ δεδομένων για ανίχνευση μεταβολών	22
<b>3. ΜΕΘΟΔΟΛΟΓΙΑ ΚΑΙ ΥΛΟΠΟΙΗΣΗ</b>	<b>25</b>
3.1 Δεδομένα	25
3.1.1 Onera Satellite Change Detection dataset (OSCD, Daudt et al. 2018)	25
3.1.2 LEVIR-CD dataset (Chen & Shi, 2020)	25
3.2 Μοντέλα	27
3.2.1 L-UNet (Papadomanolaki et al, 2019)	27
3.2.2 Αρχιτεκτονική με διεσταλμένα φίλτρα	28
3.3 Λεπτομέρειες υλοποίησης	29
3.4 Τρόποι Αξιολόγησης των Αποτελεσμάτων	30
3.4.1 Διαγράμματα σφάλματος εκπαίδευσης και επικύρωσης	30
3.4.2 Πίνακας σύγκρισης-Μετρικές για τα δεδομένα ελέγχου	31
3.4.3 Ποιοτική αξιολόγηση των εικόνων	32
<b>4. ΠΕΙΡΑΜΑΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΚΑΙ ΑΞΙΟΛΟΓΗΣΗ</b>	<b>34</b>
4.1 Σημαντικότητα LSTM στο δίκτυο	34
4.1.1 Εκπαίδευση	34

4.1.2 Ποσοτικά αποτελέσματα στο σετ ελέγχου	35
4.1.3 Ποιοτικά αποτελέσματα στο σετ ελέγχου	36
4.2 Αρχιτεκτονική με διεσταλμένα (dilated) φίλτρα	40
4.2.1 Εκπαίδευση	40
4.2.2 Ποσοτικά αποτελέσματα στο σετ ελέγχου	40
4.2.3 Ποιοτικά αποτελέσματα στο σετ ελέγχου	41
4.3 Εφαρμογή στο σετ εικόνων LEVIR-CD	43
4.3.1 Εκπαίδευση	43
4.3.2 Ποσοτικά αποτελέσματα στο σετ ελέγχου	44
4.3.3 Ποιοτικά αποτελέσματα στο σετ ελέγχου	44
<b>5. ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΠΡΟΟΠΤΙΚΕΣ</b>	<b>50</b>
5.1 Συμπεράσματα	50
5.2 Προτάσεις	52
<b>ΒΙΒΛΙΟΓΡΑΦΙΑ</b>	<b>53</b>
<b>ΠΑΡΑΡΤΗΜΑ</b>	<b>57</b>

# 1. ΕΙΣΑΓΩΓΗ

Σε αυτό το κεφάλαιο γίνεται αναφορά στις βασικές έννοιες που αφορούν την εργασία. Ακόμη, περιγράφεται το αντικείμενο και ο στόχος της εργασίας και αναλύεται η συνεισφορά της. Τέλος, περιγράφεται η δομή της εργασίας.

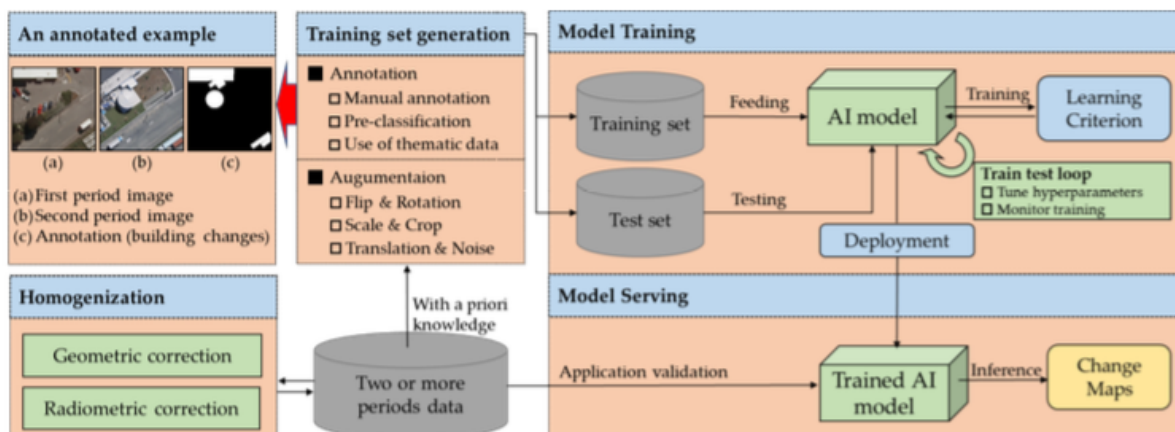
## 1.1 Γενικότερες Έννοιες

Τα τελευταία χρόνια έχει αυξηθεί κατακόρυφα η διαθεσιμότητα τηλεπισκοπικών δεδομένων, για παράδειγμα μέσω των προγραμμάτων Copernicus (<https://www.copernicus.eu/el>) και Landsat (<https://landsat.gsfc.nasa.gov/data/>), άρα και οι δυνατότητες για διάφορες εφαρμογές επεξεργασίας αυτών των δεδομένων, μεταξύ των οποίων και η ανίχνευση μεταβολών (change detection).

Αντικείμενο της ανίχνευσης μεταβολών είναι ο ποσοτικός καθορισμός των αλλαγών της επιφάνειας της γης κατά την διάρκεια του χρόνου μέσω της από κοινού ανάλυσης δύο ή περισσότερων εικόνων που απεικονίζουν την ίδια περιοχή. Οι αλλαγές αυτές μπορεί να οφείλονται για παράδειγμα σε φυσικές καταστροφές, αστική εξάπλωση, την κλιματική αλλαγή. Ακόμη, η ανίχνευση μεταβολών έχει εφαρμογές στην διαχείριση των φυσικών πόρων και στην χαρτογράφηση των χρήσεων και των καλύψεων γης (Daudt, 2020).

Σύμφωνα με τους Song et al. (2021), η αρχή του πεδίου της ανίχνευσης μεταβολών ορίστηκε από τους Weismiller et al. (1977), οι οποίοι δοκίμασαν την μέθοδο της διαφοράς μεταξύ δύο τηλεπισκοπικών εικόνων για να ανιχνεύσουν την αλλαγή μεταξύ τους. Στην βιβλιογραφία υπάρχουν πολλές κατηγοριοποιήσεις των μεθόδων που χρησιμοποιούνται και κάποιες από αυτές περιγράφονται αναλυτικότερα στο Κεφάλαιο 2. Για παράδειγμα, οι μέθοδοι μπορούν να διαχωριστούν στις παραδοσιακές μεθόδους, ειδικότερα σε εκείνες που βασίζονται στις εικονοψηφίδες (pixel-based) και σε εκείνες που βασίζονται στα αντικείμενα (object-based), ανάλογα με την κλίμακα των αντικειμένων και στις μεθόδους που βασίζονται στα νευρωνικά δίκτυα.

Οι εικόνες που μπορεί να χρησιμοποιηθούν για την ανίχνευση μεταβολών είναι μονοχρωματικές, πολυφασματικές, υπερφασματικές, εικόνες SAR, κλπ. Σύμφωνα με τους (Shi et al., 2020), οι εικόνες αρχικά πρέπει να υποστούν προεπεξεργασία, και ειδικότερα να γίνει αντιστοίχιση ("registration"- Vakaloroulou et al., 2015), ραδιομετρικές διορθώσεις και αφαίρεση του θορύβου. Στη συνέχεια, για την περίπτωση όπου ακολουθηθεί μια μέθοδος μηχανικής μάθησης, κατασκευάζεται το σετ με τις εικόνες εκπαίδευσης και αν χρειάζεται γίνεται επαύξηση των δεδομένων. Έπειτα εκπαιδεύεται το μοντέλο με βάση ένα καθορισμένο κριτήριο και στη συνέχεια εφαρμόζεται στις εικόνες του σετ ελέγχου. Το τελικό αποτέλεσμα είναι ο χάρτης αλλαγών (change map).



Εικόνα 1.1. Διάγραμμα ροής για έναν αλγόριθμο μηχανικής μάθησης με αντικείμενο την ανίχνευση μεταβολών (Shi et al., 2020).

Οι προκλήσεις που αντιμετωπίζει η ανίχνευση μεταβολών οφείλονται κυρίως σε λάθη στην αντιστοίχιση (registration) ή σε αλλαγές στη φωτεινότητα μεταξύ των εικόνων που δεν αντιστοιχούν σε ουσιαστικές αλλαγές στα αντικείμενα που απεικονίζονται, αλλαγές λόγω λήψεων σε διαφορετικές εποχές (για την βλάστηση), στην ύπαρξη σύννεφων, κλπ. Επιπροσθέτως, προκλήσεις μπορεί να δημιουργούνται λόγω των σετ εικόνων που χρησιμοποιούνται και τα οποία για παράδειγμα μπορεί να παρουσιάζουν μεγάλη ανισορροπία μεταξύ των δύο κατηγοριών προς ταξινόμηση (αλλαγή/μη αλλαγή), γεγονός που καθιστά λιγότερο εύκολη την εκπαίδευση των αλγορίθμων. Τέλος, οι υπάρχουσες μέθοδοι ανίχνευσης μεταβολών έχουν περιθώριο βελτίωσης, καθώς παράγουν πολλά ψευδώς θετικά παραδείγματα ή παρουσιάζουν χαμηλά νούμερα σε μετρικές, καθιστώντας με αυτό τον τρόπο δύσκολη την πρακτική εφαρμογή τους.

## 1.2 Αντικείμενο και Στόχος της εργασίας

Η παρούσα εργασία εστιάζει στην μελέτη της ανίχνευσης μεταβολών με χρήση αρχιτεκτονικών πλήρως συνελκτικών με μοντέλα LSTM (Long Short-Term Memory Network).

Έτσι, σε πρώτο στάδιο διερευνήθηκε η σημαντικότητα της ύπαρξης των LSTM σε κάθε επίπεδο όσον αφορά την ακρίβεια που επιτυγχάνεται, βάσει του δικτύου L-UNet που προτάθηκε από τους Papadomanolaki et al, 2019. Ειδικότερα, έγινε σύγκριση της πλήρους εκδοχής του δικτύου, δηλαδή με ένα μπλοκ LSTM σε κάθε επίπεδο, με κάθε πιθανή περίπτωση στην οποία λείπει από ένα μπλοκ LSTM, δηλαδή συνολικά προκύπτουν 8 περιπτώσεις.

Η δεύτερη εφαρμογή αφορά την χρήση μιας αρχιτεκτονικής με διεσταλμένα (dilated) φίλτρα όπως περιγράφεται στους Vakaloroulou et al. (2019), για να βοηθήσει την σημασιολογική κατάτμηση και κατ'επέκταση την ανίχνευση μεταβολών. Διενεργήθηκε σύγκριση της απόδοσής της, σε σχέση με το δίκτυο L-UNet (Papadomanolaki et al, 2019) στο σετ εικόνων OSCD (Onera Satellite Change Detection).

Η τρίτη υλοποίηση αφορά την εφαρμογή των δικτύων L-UNet (Papadomanolaki et al, 2019) και Dilated+LSTM για πρώτη φορά στο σετ εικόνων LEVIR-CD (Chen & Shi, 2020).

## 1.3 Συνεισφορά

Η παρούσα εργασία διερευνά και αξιολογεί την σημαντικότητα της ύπαρξης των μπλοκ LSTM που κωδικοποιούν την χρονική πληροφορία, για το αντικείμενο της ανίχνευσης μεταβολών, σε διαφορετικά επίπεδα συνελκτικών δικτύων. Πιο συγκεκριμένα, τα αποτελέσματα υποδεικνύουν ότι η ακρίβεια του μοντέλου μειώνεται με την αφαίρεση μπλοκ LSTM, καθώς αξιοποιείται λιγότερο η χρονική πληροφορία.

Ακόμη, εξερευνά μια διαφορετική αρχιτεκτονική για την αποτελεσματικότερη εξαγωγή της χωρικής πληροφορίας, μέσω της χρήσης διεσταλμένων (“dilated”) φίλτρων, τα οποία βοηθούν στην αύξηση του αντιληπτικού πεδίου. Τα ευρήματα υποδεικνύουν ότι έχει παρόμοια απόδοση όσον αφορά την εξισορροπημένη ακρίβεια, ενώ έχει λιγότερες παραμέτρους από μοντέλο L-UNet (Papadomanolaki et al, 2019) με το οποίο συγκρίνεται.

Τέλος, γίνεται εφαρμογή για πρώτη φορά των δύο αρχιτεκτονικών που δοκιμάστηκαν νωρίτερα στο σετ εικόνων LEVIR-CD, το οποίο είναι ένα πιο μεγάλο σε μέγεθος και πιο απαιτητικό σετ εικόνων πολύ υψηλής ανάλυσης, σε αντίθεση με το σετ εικόνων OSCD που χρησιμοποιήθηκε στις προηγούμενες εφαρμογές. Το βασικό συμπέρασμα που εξάγεται είναι ότι και γ’ αυτό το σετ εικόνων η απόδοση του μοντέλου L-UNet είναι ποσοτικά λίγο καλύτερη από την απόδοση του δικτύου με τα διεσταλμένα φίλτρα, ενώ εξάγονται ιδιαίτερως χρήσιμα συμπεράσματα από την επισκόπηση της εφαρμογής των δύο μεθόδων στις εικόνες.

## 1.4 Δομή της εργασίας

Η εργασία αναπτύσσεται σε συνολικά 5 κεφάλαια. Στο Κεφάλαιο 2 παρουσιάζεται το θεωρητικό πλαίσιο στο οποίο βασίζεται η εργασία, δηλαδή αναλύονται βασικές αρχιτεκτονικές των δικτύων που χρησιμοποιούνται στην εργασία και γενικότερα στο πεδίο της ανίχνευσης μεταβολών. Στο Κεφάλαιο 3 αναλύεται η μεθοδολογία που ακολουθήθηκε για τις τρεις εφαρμογές, δηλαδή τα στοιχεία σχετικά με τα δεδομένα, τα μοντέλα και τους τρόπους αξιολόγησης των αποτελεσμάτων, όπως και λεπτομέρειες για την υλοποίηση των εφαρμογών. Ακολούθως, στο Κεφάλαιο 4 παρουσιάζονται τα ποσοτικά και ποιοτικά αποτελέσματα από τις τρεις εφαρμογές. Τα συμπεράσματα που προέκυψαν, καθώς και οι προτάσεις για πιθανή μελλοντική έρευνα παρουσιάζονται στο Κεφάλαιο 5. Τέλος, παρατίθεται η βιβλιογραφία επί της οποίας βασίστηκε η εργασία και ένα παράρτημα με ορισμένους χάρτες μεταβολών που παράχθηκαν από τις εφαρμογές.



## 2. ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΑΝΑΣΚΟΠΗΣΗ

Σε αυτό το κεφάλαιο παρουσιάζονται αρχικά διάφορες κατηγοριοποιήσεις των αλγορίθμων ανίχνευσης μεταβολών. Έπειτα περιγράφονται οι βασικές αρχιτεκτονικές που χρησιμοποιούνται συνήθως στους αλγορίθμους σημασιολογικής κατάτμησης εικόνων και οι οποίες εφαρμόζονται κατ' επέκταση και στις μεθόδους ανίχνευσης μεταβολών, όπως το U-Net και τα δισταλμένα φίλτρα. Στη συνέχεια, αναλύονται ειδικότερα μέθοδοι ανίχνευσης μεταβολών. Μετά από μια σύντομη ανασκόπηση στους αλγορίθμους χωρίς νευρωνικά δίκτυα, περιγράφονται στη συνέχεια οι πρώτες αρχιτεκτονικές με τη χρήση συνελκτικών νευρωνικών δικτύων (Convolutional Neural Networks) που προτάθηκαν, αρχιτεκτονικές που αξιοποιούν καλύτερα την παράμετρο του χρόνου με την προσθήκη LSTM ή μεθόδων προσοχής, καθώς και αρχιτεκτονικές που χρησιμοποιούνται και για άλλα αντικείμενα, παράλληλα με την ανίχνευση μεταβολών. Τέλος, γίνεται μια ανασκόπηση στα διαθέσιμα σετ δεδομένων για ανίχνευση μεταβολών.

### 2.1 Κατηγοριοποίηση αλγορίθμων ανίχνευσης μεταβολών

Στην βιβλιογραφία υπάρχουν αρκετοί τρόποι διαχείρισης του προβλήματος της ανίχνευσης μεταβολών. Στην συγκεκριμένη εργασία θα εστιάσουμε στην ανάλυση μεθόδων που στηρίζονται σε μεθόδους μηχανικής μάθησης και πιο συγκεκριμένα σε μεθόδους βαθιάς μάθησης. Όσον αφορά τους τύπους της ανίχνευσης μεταβολών, υπάρχει η δυαδική ανίχνευση μεταβολών (binary change detection), η οποία χρησιμοποιείται και στην παρούσα εργασία. Στόχος της είναι η ανάθεση μιας ετικέτας (label) σε κάθε εικονοψηφίδα, που μπορεί να πάρει μια από τις εξής τιμές αλλαγή/όχι αλλαγή (change/no change) και προκύπτει από ένα ζεύγος ή μια σειρά εικόνων (που έχουν υποστεί αντιστοίχιση) μιας συγκεκριμένης περιοχής σε διαφορετικές χρονικές στιγμές (Liu et al., 2019). Ως επέκταση της δυαδικής, υπάρχει η σημασιολογική ανίχνευση μεταβολών (semantic change detection), κατά την οποία ταξινομούνται οι αλλαγές σε σημασιολογικές κατηγορίες.

Ακόμη, η ανίχνευση μεταβολών μπορεί να αντιμετωπίζεται ως ένα πρόβλημα όπου ο στόχος είναι να υπολογιστεί το μέτρο της "απόστασης", της διαφοράς δηλαδή μεταξύ των εικονοψηφίδων των δύο εικόνων, ή ως ένα πρόβλημα ταξινόμησης, όπου οι εικονοψηφίδες ταξινομούνται σε αλλαγμένες ή μη (Daudt, 2020). Επιπλέον, μπορεί να γίνει διαφοροποίηση των υπαρχουσών μεθόδων βάσει του τρόπου συγχώνευσης που χρησιμοποιούν (early/late fusion). Στην πρώτη περίπτωση ("early fusion"), συγχωνεύονται τα τμήματα εικόνων πριν εισαχθούν στο δίκτυο, δηλαδή αντιμετωπίζονται ως διαφορετικά χρωματικά κανάλια. Στην δεύτερη περίπτωση ("late fusion"), γίνεται ξεχωριστή επεξεργασία κάθε εικόνας από το δίκτυο και σε επόμενο στάδιο γίνεται συγχώνευση των χαρτών χαρακτηριστικών που έχουν παραχθεί (Daudt et al., 2018).

Η ύπαρξη δύο ή περισσότερων εικόνων αλλά και ο τρόπος διαχείρισής τους ως διαφορετικές εικόνες της ίδιας περιοχής ("per date") ή ως μιας ακολουθίας εικόνων στον χρόνο ("per sequence") συνιστά ένα ακόμη πεδίο διαφοροποίησης για τους υπάρχοντες αλγορίθμους. Επιπροσθέτως, το πώς θα γίνει η διαχείριση της χωρικής και της χρονικής πληροφορίας είναι ένα ακόμη ζήτημα. Οι αρχιτεκτονικές που χρησιμοποιούν ταυτόχρονα χωρική και χρονική πληροφορία συνήθως χρησιμοποιούν υβριδικές μεθόδους που συνδυάζουν την αρχιτεκτονική U-Net με αναδρομικά δίκτυα, όπως οι Papadomanolaki et al. (2019) ή

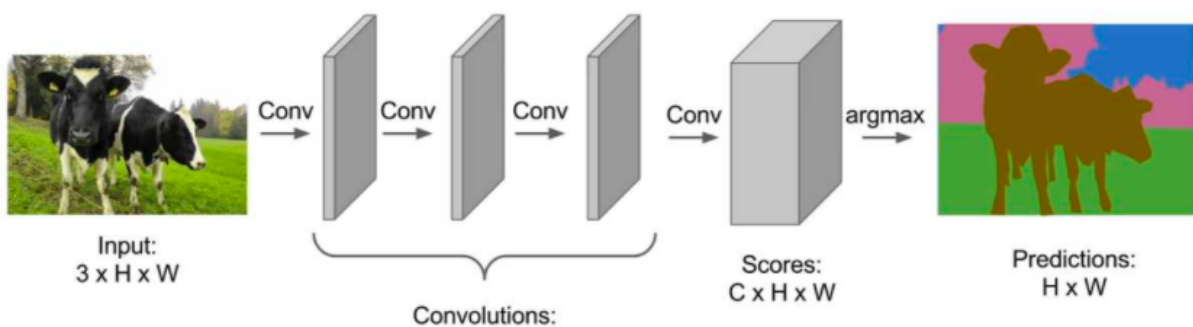
μεθόδους που βασίζονται στην προσοχή (“attention”), όπως οι Sainte Fare Garnot et al (2020).

## 2.2 Βασικές αρχιτεκτονικές για την σημασιολογική κατάτμηση εικόνας

Η ανίχνευση μεταβολών συνιστά συγγενές αντικείμενο της σημασιολογικής κατάτμησης (semantic segmentation), δεδομένου ότι και στις δύο περιπτώσεις χρειάζεται να γίνει αναγνώριση των διαφορετικών κατηγοριών που υπάρχουν στην εικόνα ανά εικονοπηφίδα, με την διαφορά ότι στην ανίχνευση μεταβολών αυτή η διαδικασία πραγματοποιείται για παραπάνω από μία εικόνες, οι οποίες εν συνεχεία συγκρίνονται για να παραχθεί το τελικό αποτέλεσμα, ο χάρτης που απεικονίζει τις μεταβολές που έχουν λάβει χώρα στην περιοχή μεταξύ των δύο λήψεων. Κατά συνέπεια αρκετές από τις αρχιτεκτονικές που χρησιμοποιούνται στην ανίχνευση μεταβολών (ειδικότερα για το χωρικό κομμάτι του προβλήματος, όχι για το χρονικό) έχουν πρώτα εφαρμοστεί στην σημασιολογική κατάτμηση και κρίνεται σκόπιμο να αναλυθούν ξεχωριστά.

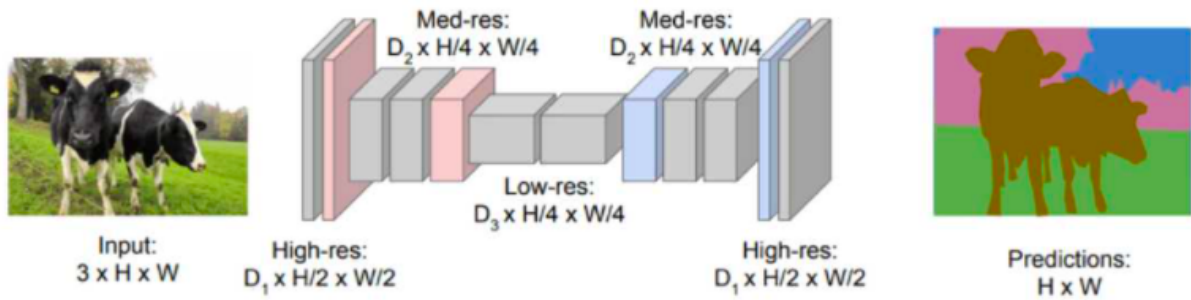
### 2.2.1 Πλήρη Συνελικτικά Δίκτυα (Fully Convolutional Networks-FCN)

Η αρχική ιδέα ήταν να χρησιμοποιηθούν διαδοχικά συνελικτικά επίπεδα με διαστάσεις όσο η εικόνα εισόδου ώστε η εικόνα εξόδου με τις προβλέψεις να έχει επίσης τις ίδιες διαστάσεις (Εικ. 2.1). Όμως αυτό θα ήταν υπολογιστικά ακριβό. Από την μια πλευρά, δεν είναι δυνατό να χρησιμοποιηθούν επίπεδα μείωσης της ανάλυσης (downsampling) ή επίπεδα συγκέντρωσης (pooling), καθώς θα χανόταν η πληροφορία για την θέση των αντικειμένων. Από την άλλη πλευρά, χρειάζονται πολλά επίπεδα για να γίνει μάθηση τόσο των χαρακτηριστικών χαμηλού επιπέδου όσο και υψηλού επιπέδου και να επιτευχθεί καλύτερη απόδοση.



Εικόνα 2.1. Αρχιτεκτονική πλήρους συνελικτικού δικτύου ([http://cs231n.stanford.edu/slides/2017/cs231n\\_2017\\_lecture11.pdf](http://cs231n.stanford.edu/slides/2017/cs231n_2017_lecture11.pdf)).

Για να λυθεί αυτό το πρόβλημα, προτάθηκε από τους Long et al. (2015) μια αρχιτεκτονική κωδικοποιητή-αποκωδικοποιητή (encoder-decoder), όπου ο κωδικοποιητής είναι ένα τυπικό συνελικτικό δίκτυο όπως το AlexNet ή το ResNet ενώ ο αποκωδικοποιητής αποτελείται από αποσυνελικτικά (“deconvolutional”) επίπεδα ή επίπεδα αύξησης της χωρικής ανάλυσης (“upsampling”) [Εικ. 2.2]. Ο στόχος του κωδικοποιητή όπου μειώνεται η ανάλυση είναι να αντιληφθεί την σημασιολογική πληροφορία ενώ ο στόχος του αποκωδικοποιητή όπου αυξάνεται η ανάλυση είναι να ανακτήσει την χωρική πληροφορία.

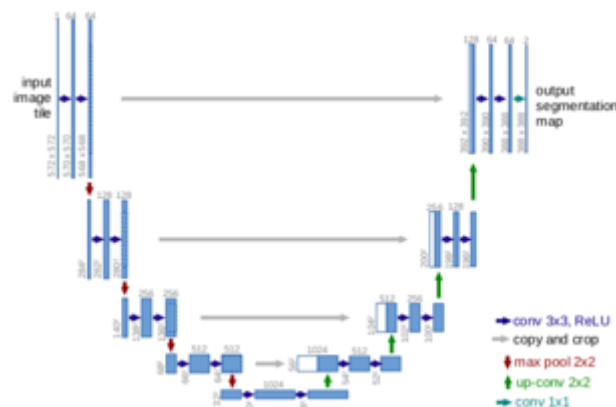


Εικόνα 2.2. Εξέλιξη της αρχιτεκτονικής του πλήρους συνελκτικού δικτύου ([http://cs231n.stanford.edu/slides/2017/cs231n\\_2017\\_lecture11.pdf](http://cs231n.stanford.edu/slides/2017/cs231n_2017_lecture11.pdf)).

Με αυτό τον τρόπο, μειώθηκε η πολυπλοκότητα και ο χρόνος εκτέλεσης αλλά μειώθηκε και η ακρίβεια του αποτελέσματος. Ειδικότερα, επειδή ο κωδικοποιητής μειώνει την χωρική ανάλυση, δεν υπάρχει μεγάλη ακρίβεια στις ακμές και στα όρια μεταξύ των αντικειμένων.

### 2.2.2 U-Net

Το U-Net προτάθηκε από τους Ronneberger et al. (2015) ως επέκταση των Fully Convolutional Networks (FCN), χρησιμοποιώντας και αυτό την ιδέα του κωδικοποιητή-αποκωδικοποιητή και των συνδέσεων παράλειψης (skip connections). Το U-Net δημιουργεί συμμετρικά FCN μέσω της αύξησης του μεγέθους του αποκωδικοποιητή για να ταιριάζει με τον κωδικοποιητή και συνενώνει (“concatenate”) τους χάρτες χαρακτηριστικών κατά την διαδικασία των συνδέσεων παράλειψης αντί απλώς να τους αθροίζει (Εικ. 2.3). Χάρη στη συμμετρία, μπορεί να γίνει μεταφορά πολύ περισσότερης πληροφορίας από τα αντίστοιχα επίπεδα, βελτιώνοντας έτσι την ανάλυση της τελικής εικόνας. Δημιουργήθηκε για το αντικείμενο της κατάτμησης ιατρικών εικόνων, αλλά γρήγορα άρχισε να χρησιμοποιείται και σε άλλες εφαρμογές (<https://theaisummer.com/Semantic Segmentation/>).



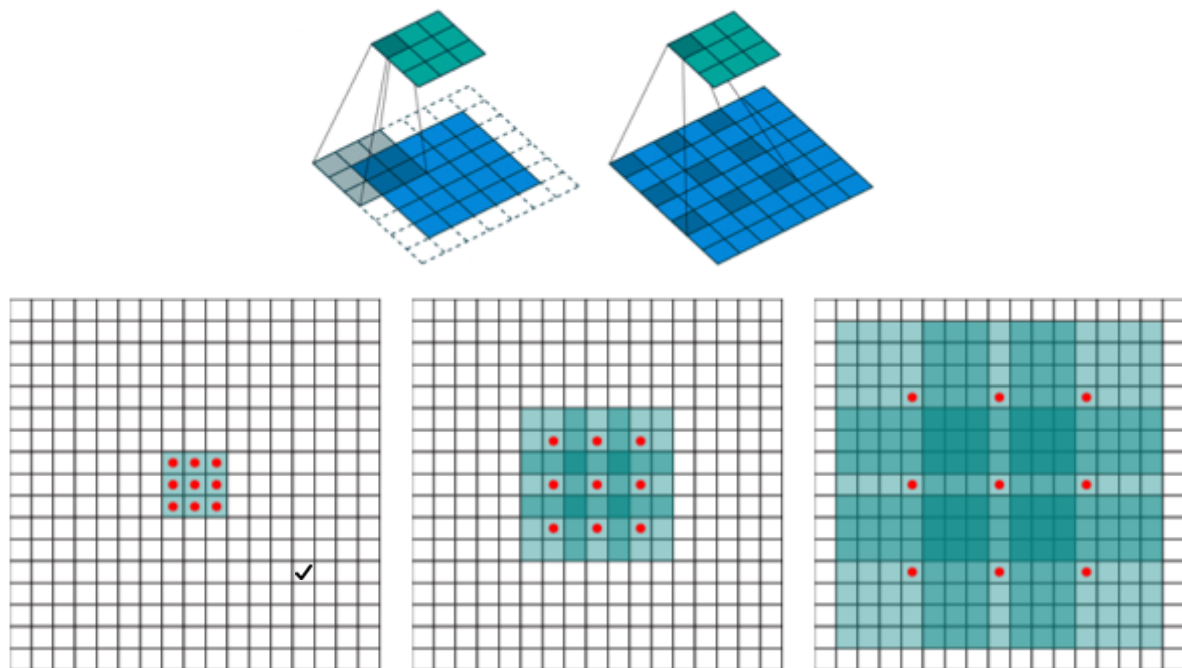
Εικόνα 2.3. Αρχιτεκτονική U-Net (Ronneberger et al., 2015).

Ειδικότερα, το δίκτυο αποτελείται από ένα μονοπάτι στο οποίο μειώνονται οι χωρικές διαστάσεις, και ένα μονοπάτι στο οποίο αυξάνονται. Το πρώτο αποτελείται σε κάθε επίπεδο ανάλυσης από 2 συνελκτικά επίπεδα με φίλτρο μεγέθους 3x3, ένα επίπεδο ReLU και ένα επίπεδο 2x2 μέγιστης συγκέντρωσης με βήμα 2, το οποίο μειώνει την ανάλυση. Σε κάθε

επίπεδο που μειώνεται η ανάλυση, αυξάνεται ο αριθμός των χαρτών χαρακτηριστικών. Στο δεύτερο μονοπάτι, αυξάνεται σε κάθε επίπεδο η ανάλυση με την εφαρμογή συνέλιξης μεγέθους 2x2 (“upconvolution”) και μειώνεται κατά το ήμισυ ο αριθμός των χαρτών χαρακτηριστικών. Έπειτα συνενώνονται οι χάρτες χαρακτηριστικών από το αντίστοιχο επίπεδο του πρώτου μονοπατιού και στη συνέχεια εφαρμόζονται δύο συνέλιξεις μεγέθους 3x3 και ReLU. Στο τελευταίο επίπεδο χρησιμοποιείται συνέλιξη μεγέθους 1x1 ώστε να παραχθεί το τελικό αποτέλεσμα.

### 2.2.3 Dilated networks

Η συνέλιξη atrous (ή dilated) αποτελεί την γενίκευση της γνωστής συνέλιξης, καθώς εισάγει έναν επιπλέον όρο, τον ρυθμό (“rate”). Ειδικότερα, εισάγει «τρύπες» (trous) στο φίλτρο της συνέλιξης, αυξάνοντας με αυτό τον τρόπο το αντιληπτικό πεδίο (“receptive field”), το εύρος δηλαδή της περιοχής που “βλέπει” το φίλτρο (Εικ. 2.4). Αξίζει να σημειωθεί ότι το αντιληπτικό πεδίο αυξάνεται εκθετικά ενώ ο αριθμός των παραμέτρων γραμμικά.



Εικόνα 2.4. Συνέλιξη με ρυθμό=1 και ρυθμό=2 (πάνω-<https://towardsdatascience.com/types-of-convolutions-in-deep-learning-717013397f4d>), απεικόνιση του αντιληπτικού πεδίου με ρυθμό=1, ρυθμό=2, ρυθμό=4 (κάτω-από αριστερά προς τα δεξιά-Yu et al., 2016).

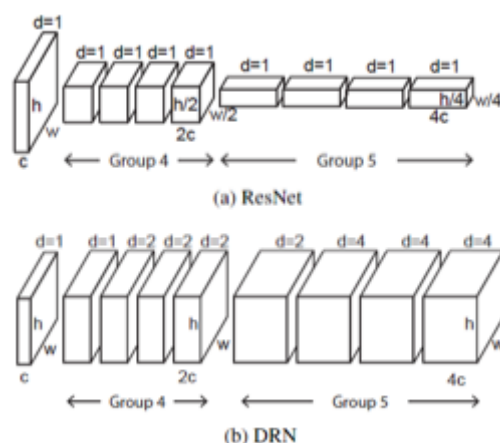
Η χρήση της διεσταλμένης συνέλιξης σε συνδυασμό με νευρωνικά δίκτυα για το πρόβλημα της σημασιολογικής κατάτμησης προτάθηκε από τους Yu et al. (2016) [DilatedNet, ICML 2016]. Στόχος ήταν να συσσωματωθεί πληροφορία από διαφορετικές κλίμακες, χωρίς να μειώνεται η χωρική ανάλυση. Πιο συγκεκριμένα, προτείνουν ένα δομικό στοιχείο “context module”, το οποίο έχει 7 επίπεδα που εφαρμόζουν συνέλιξεις με φίλτρο μεγέθους 3 x 3 και διαφορετικό ρυθμό διαστολής (1,1,2,4,8,16 και 1). Το τελευταίο επίπεδο είναι μια συνέλιξη 1x1, ώστε το πλήθος των εξόδων να είναι ίδιο με τον αριθμό των εισόδων. Οπότε το δομικό στοιχείο μπορεί να εισαχθεί σε διαφορετικών ειδών νευρωνικά δίκτυα. Στην Εικ. 2.5 περιγράφεται το βασικό δομικό στοιχείο, που έχει μόνο ένα κανάλι και το μεγάλο δομικό στοιχείο, που έχει κυμαινόμενο αριθμό καναλιών, από 1 έως 32.

Στην προκειμένη περίπτωση, το δομικό στοιχείο χρησιμοποιείται σε συνδυασμό με το μοντέλο VGG-16.

Layer	1	2	3	4	5	6	7	8
Convolution	3×3	3×3	3×3	3×3	3×3	3×3	3×3	1×1
Dilation	1	1	2	4	8	16	1	1
Truncation	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Receptive field	3×3	5×5	9×9	17×17	33×33	65×65	67×67	67×67
Output channels								
Basic	$C$	$C$	$C$	$C$	$C$	$C$	$C$	$C$
Large	$2C$	$2C$	$4C$	$8C$	$16C$	$32C$	$32C$	$C$

Εικόνα 2.5. Περιγραφή στοιχείων βασικού και μεγάλου δομικού στοιχείου (Yu et al., 2016).

Στη συνέχεια, οι Yu et al. (2017) πρότειναν τα Dilated Residual Networks [DRN, CVPR 2017], όπου αντικαθιστούν τα επίπεδα του γκρουπ 4 και 5 του ResNet με επίπεδα με ρυθμό διαστολής μεγαλύτερο του 1 και αφαιρούν την μέγιστη συγκέντρωση (maxpooling) (Εικ. 2.6). Με αυτό τον τρόπο, οι χάρτες χαρακτηριστικών που προκύπτουν έχουν μεγαλύτερες διαστάσεις και αυξάνεται η ακρίβεια για την σημασιολογική κατάτμηση. Αν είχε αφαιρεθεί απλώς η μέγιστη συγκέντρωση, θα είχε μειωθεί το αντιληπτικό πεδίο, άρα και η πληροφορία που αξιοποιείται για να παραχθεί το τελικό αποτέλεσμα. Με την προσθήκη των επιπέδων με μεγαλύτερο ρυθμό διαστολής, το αντιληπτικό πεδίο μπορεί να διατηρηθεί.



Εικόνα 2.6. Σύγκριση δικτύων ResNet και DRN (Yu et al., 2017).

Ειδικότερα, το DRN έχει τον ίδιο αριθμό επιπέδων και παραμέτρων με το αρχικό ResNet. Όμως η διαφορά τους έγκειται στο γεγονός ότι το αρχικό ResNet μειώνει την αρχική εικόνα κατά 32 σε κάθε διάσταση, ενώ το DRN μόνο κατά 8. Για παράδειγμα, αν η αρχική εικόνα έχει διαστάσεις 224x224, από το ResNet προκύπτει 7x7, που όπως μπορεί να γίνει αντιληπτό δεν επαρκεί για να αντληθεί αρκετή πληροφορία από τον ταξινομητή, ενώ από το DRN προκύπτει 28x28.

Επίσης, επειδή παρατηρήθηκε ότι δημιουργούνται διάφορα σφάλματα τύπου “gridding artifacts”, έγιναν περαιτέρω προσθήκες στο προτεινόμενο δίκτυο DRN των Yu et al. (2017). Αρχικά αντικαταστάθηκε το πρώτο επίπεδο μέγιστης συγκέντρωσης με 2 μπλοκ διαφορών (“residual blocks”) στην αρχή και στο τέλος του δικτύου (DRN-B). Στη συνέχεια, ο ρυθμός διαστολής μειώθηκε σταδιακά στο τέλος του δικτύου, δηλαδή μια συνέλιξη με ρυθμό=2

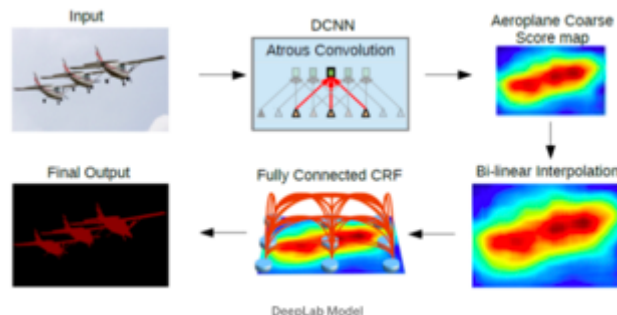
ακολουθείται από μια συνέλιξη με ρυθμό=1 και οι συνδέσεις διαφορών των 2 τελευταίων μπλοκ διαφορών αφαιρέθηκαν (Εικ. 2.7).



Εικόνα 2.7. Σχηματική απεικόνιση των δικτύων DRN-A, DRN-B και DRN-C (Yu et al., 2017).

Εδώ αξίζει να σημειωθεί ότι η μείωση της ανάλυσης είναι ιδιαίτερα σημαντική για το αντικείμενο της σημασιολογικής κατάτμησης, όπου πρέπει να αναγνωριστούν όλα τα αντικείμενα και το σχήμα τους, αντί για παράδειγμα για την απλή ταξινόμηση της εικόνας (χωρίς βέβαια να υπάρχουν περιττές πληροφορίες).

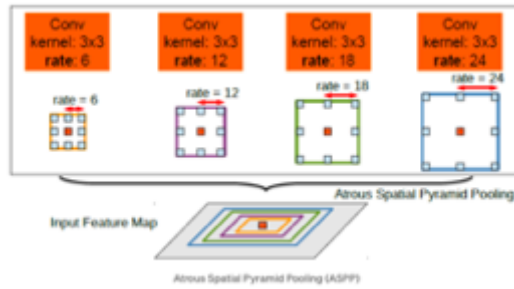
Τα DeepLabv1 (Chen et al, 2014) και DeepLabv2 (Chen et al, 2016) χρησιμοποιούν την συνέλιξη με δισταλμένα φίλτρα, καθώς και για την μετεπεξεργασία κάνουν χρήση πλήρως συνδεδεμένων Υπό Συνθήκη Τυχαίων Πεδίων (“Conditional Random Fields”-CRF) (Εικ. 2.8). Το DeepLabv1 χρησιμοποιεί το VGGNet ως μοντέλο, ενώ το DeepLabv2 χρησιμοποιεί το ResNet και το VGGNet, όπως επίσης και μια πιο εξελιγμένη μορφή της δισταλμένης συνέλιξης (“Atrous Spatial Pyramid Pooling”-ASPP).



Εικόνα 2.8. Αρχιτεκτονική δικτύων DeepLabv1 και DeepLabv2 (Chen et al, 2014).

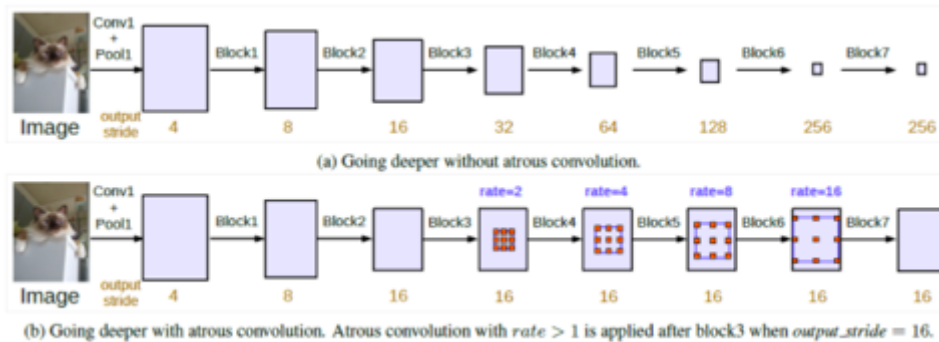
Η εικόνα εισόδου περνάει πρώτα μέσα από το δίκτυο (VGG-16 ή ResNet) όπου χρησιμοποιείται η δισταλμένη συνέλιξη ή και το ASPP. Στο ASPP (Εικ. 2.9) εφαρμόζονται παράλληλα δισταλμένες συνελίξεις με διαφορετικό ρυθμό, που τελικά συγχωνεύονται. Αυτό βοηθάει για παράδειγμα σε περιπτώσεις όπου εμφανίζονται στην εικόνα αντικείμενα της ίδιας κατηγορίας σε διαφορετικές κλίμακες. Στη συνέχεια η έξοδος του δικτύου υπόκειται σε διγραμμική παρεμβολή και εισάγεται στο πλήρες συνδεδεμένο CRF για να βελτιωθεί περαιτέρω (fine-tuning) το αποτέλεσμα.





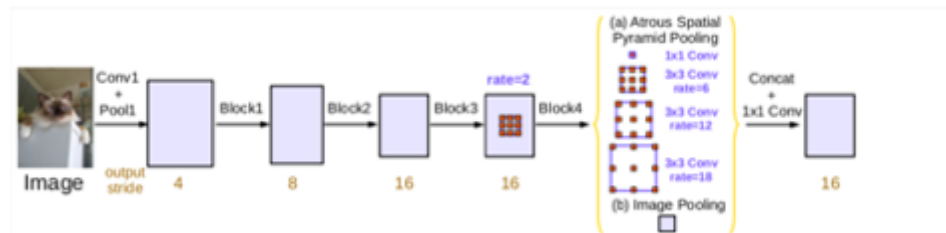
Εικόνα 2.9. Σχηματική απεικόνιση του ASPP (Chen et al, 2016).

Στο DeepLabv3 (Chen et al., 2017) αφαιρείται το CRF, που εμπόδιζε την άμεση (end-to-end) εκπαίδευση και εισάγεται η έννοια του Multi-Grid, όπου διατηρούνται οι διαστάσεις (“stride”) της εικόνας αλλά με μεγαλύτερο αντιληπτικό πεδίο, χωρίς να αυξάνεται ο αριθμός των παραμέτρων ή το υπολογιστικό κόστος (Εικ. 2.10). Επίσης, ο χάρτης χαρακτηριστικών που προκύπτει έχει μεγαλύτερες διαστάσεις.



Εικόνα 2.10. Σχηματική απεικόνιση του Multi-Grid (Chen et al., 2017).

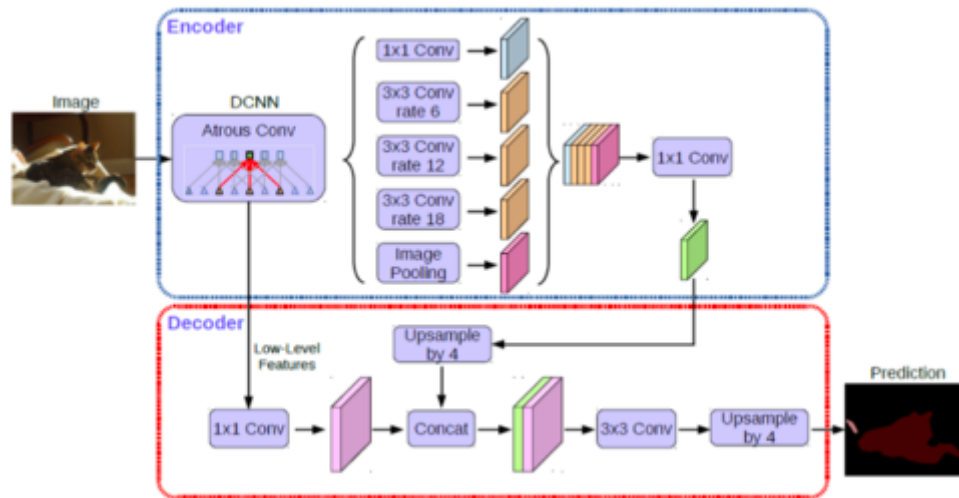
Ακόμη, χρησιμοποιείται ξανά το ASPP με κάποιες βελτιώσεις. Το πλεονέκτημά του έγκειται στο ότι έχει ανακαλυφθεί ότι όσο ο ρυθμός μεγαλώνει, τα βάρη που αντιστοιχούν σε τμήματα της εικόνας και όχι σε τμήματα με μηδενικά λόγω της προσαύξησης (“padding”) γίνονται λιγότερα. Το ASPP εφαρμόζεται ως εξής: όταν το βήμα (“stride”) είναι 16, πραγματοποιείται μια συνέλιξη με φίλτρο μεγέθους 1x1 και τρεις συνέλιξεις με φίλτρο 3x3 και ρυθμούς (6,12,18) και επίσης εφαρμόζεται image pooling, όπως στο ParseNet, για να αξιοποιηθεί πληροφορία από όλη την εικόνα. Όλα τα επίπεδα έχουν 256 χάρτες χαρακτηριστικών και εφαρμόζεται batch normalization. Στην συνέχεια συνενώνονται και εφαρμόζεται άλλη μια συνέλιξη 1x1 (επίσης με 256 χάρτες χαρακτηριστικών και batch normalization) πριν από την τελική συνέλιξη 1x1 που παράγει το τελικό αποτέλεσμα (Εικ. 2.11).



Εικόνα 2.11. Σχηματική απεικόνιση του ASPP (Chen et al., 2017).

Τέλος, οι Chen et al. (2018) πρότειναν το DeepLabv3+, το οποίο ενσωματώνει το DeepLabv3 μέσα σε μια αρχιτεκτονική κωδικοποιητή-αποκωδικοποιητή (Εικ. 2.12). Αυτή η αρχιτεκτονική

θεωρείται ότι διατηρεί καλύτερα τις ακμές των αντικειμένων επειδή η χωρική ανάλυση βελτιώνεται σταδιακά μέσα στο δίκτυο. Από την άλλη, το spatial pyramid pooling module είναι ικανό να αποθηκεύσει πλουσιότερη πληροφορία συνδυάζοντας τις διάφορες χωρικές αναλύσεις.



Εικόνα 2.12. Αρχιτεκτονική του δικτύου DeepLabv3+ (Chen et al., 2018).

Ειδικότερα, χρησιμοποιείται το DeepLabv3 ως κωδικοποιητής και έπειτα σε αυτά τα χαρακτηριστικά αυξάνεται η ανάλυσή τους κατά παράγοντα 4 και έπειτα συγχωνεύονται με χαρακτηριστικά χαμηλότερου επιπέδου. Έπειτα εφαρμόζονται μερικές 3x3 συνελίξεις και τέλος αυξάνεται η ανάλυσή τους κατά παράγοντα 4 ξανά. Ως αρχικό δίκτυο (backbone) χρησιμοποιείται το ResNet-101 (He et al., 2016) ή μια παραλλαγή του δικτύου Aligned Xception (Chollet, 2017, Qi et al., 2017) ως “Modified Aligned Xception”, όπου όλα τα επίπεδα μέγιστης συγκέντρωσης έχουν αντικατασταθεί από atrous depthwise separable convolution (όπως παρουσιάστηκε στη παρούσα δημοσίευση των Chen et al., 2018), που ακολουθείται από batch normalization και ReLU.

## 2.3. Μέθοδοι ανίχνευσης μεταβολών (change detection)

Στο παρόν υποκεφάλαιο παρουσιάζονται οι κυριότερες μέθοδοι ανίχνευσης μεταβολών που συναντώνται στην βιβλιογραφία. Σημειώνεται ότι οι μέθοδοι εφαρμόζονται σε εικόνες που έχουν υποστεί ήδη γεωμετρική διόρθωση (αντιστοίχιση-registration).

### 2.3.1 Μέθοδοι ανίχνευσης μεταβολών χωρίς βαθιά μηχανική μάθηση

Σύμφωνα με τους Mou et al (2019), υπάρχουν πολλές κατηγορίες αλγορίθμων που χρησιμοποιήθηκαν για την ανίχνευση μεταβολών πριν από την έλευση των νευρωνικών δικτύων. Η πρώτη κατηγορία αφορά προσεγγίσεις βασισμένες στην άλγεβρα εικόνων, όπως η Ανάλυση Διανύσματος Μεταβολής (Change Vector Analysis- CVA) που προτάθηκε από τον Malila (1980). Η CVA σχεδιάζεται για να αναλύσει πιθανές αλλαγές στις εικονοψηφίδες από τα ζεύγη πολυφασματικών ζευγών εικόνων και χρησιμοποιείται μαζί με μη επιβλεπόμενες τεχνικές επιλογής κατωφλιού, που βασίζονται σε διαφορετικά πιθανά μοντέλα κατανομής των δεδομένων.



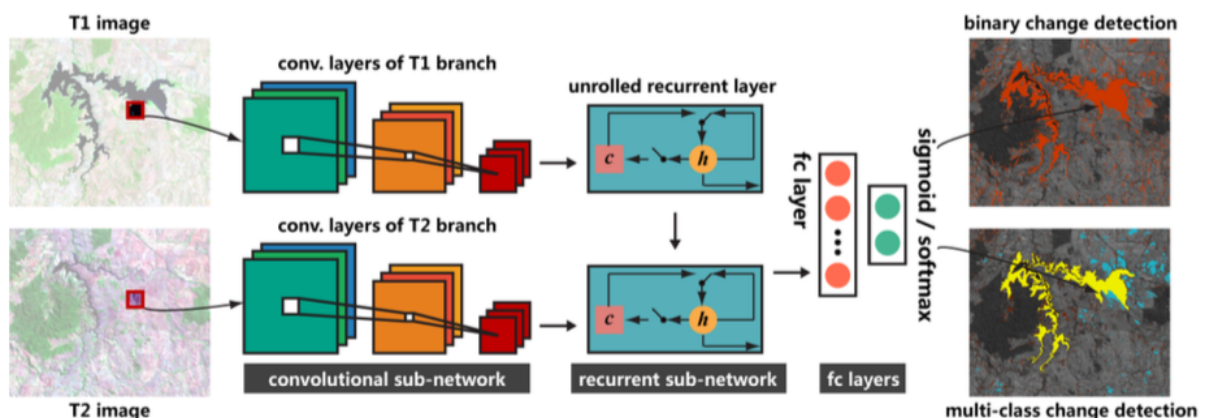
Η δεύτερη κατηγορία αφορά μοντέλα που βασίζονται στον μετασχηματισμό εικόνων και έχουν ως στόχο να μάθουν μια καινούρια, μετασχηματισμένη αναπαράσταση των χαρακτηριστικών του αρχικού φασματικού χώρου, ώστε στον καινούριο χώρο χαρακτηριστικών να επισημανθούν οι περιοχές αλλαγών και να δοθεί μικρότερο βάρος στις περιοχές χωρίς μεταβολή. Χαρακτηριστικό παράδειγμα αλγορίθμου αυτής της κατηγορίας είναι η Ανάλυση Κυρίων Συνιστωσών (Principal Component Analysis-PCA, Deng et al., 2008 και Li & Yeh, 1998).

Η τρίτη κατηγορία μεθόδων ανίχνευσης μεταβολών προσεγγίζει το πρόβλημα ως ταξινόμηση. Για παράδειγμα, οι Bruzzone & Serpico (1997) πρότειναν ένα επιβλεπόμενο μη παραμετρικό μοντέλο για την ανίχνευση αλλαγών στην κάλυψη γης από δύο εικόνες ληφθείσες σε διαφορετικές χρονικές στιγμές. Η Βακαλοπούλου (2017) ανέπτυξε μεθοδολογίες βασισμένες στα Μαρκοβιανά Τυχαία Πεδία (Markov Random Fields-MRF) για την ενιαία επίλυση των προβλημάτων της αντιστοίχισης, της σημασιολογικής κατάτμησης και της ανίχνευσης μεταβολών στον αστικό και περιαστικό χώρο από εικόνες πολύ υψηλής ανάλυσης.

Άλλες κατηγοριοποιήσεις μπορεί να γίνονται βάσει του δομικού στοιχείου που χρησιμοποιούν (pixel-based ή object-based), την χωρική ανάλυση των δεδομένων που χρησιμοποιούν, αν είναι επιβλεπόμενες ή μη επιβλεπόμενες μέθοδοι, κλπ (Βακαλοπούλου, 2017).

### 2.3.2 Πρώτες αρχιτεκτονικές με νευρωνικά δίκτυα

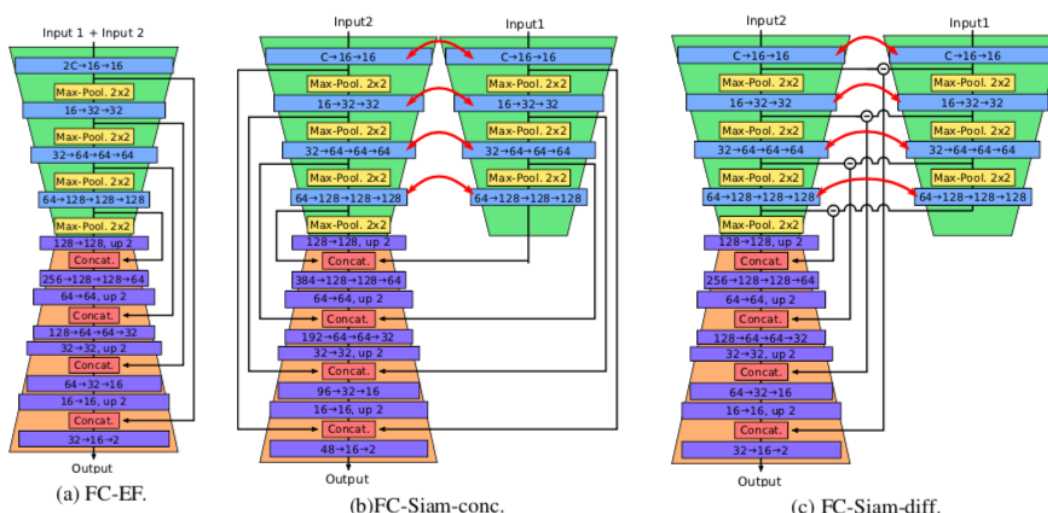
Αρχικά οι Mou et al. (2019) εισήγαγαν ένα αναδρομικό δίκτυο (ReCNN) σε μια αρχιτεκτονική με συνελκτικά δίκτυα, για να αξιοποιήσουν τα χωρικά και χρονικά χαρακτηριστικά που εξαγονται και μπορούν να εκπαιδευτούν end-to-end. Ειδικότερα, τμήματα εικόνων από ζεύγη που απεικονίζουν την ίδια περιοχή σε δύο διαφορετικές ημερομηνίες αρχικά υπόκεινται σε επεξεργασία από μια σειρά παράλληλων αλλά ίδιων διεσταλμένων συνελκτικών επιπέδων. Στη συνέχεια, τα παραγόμενα διανύσματα φασματικών και χωρικών χαρακτηριστικών εισάγονται σε ένα αναδρομικό δίκτυο, το οποίο υπολογίζει την χρονική συσχέτιση μεταξύ τους. Τέλος, πλήρη συνδεδεμένα επίπεδα λαμβάνουν την έξοδο του αναδρομικού δικτύου και αποφασίζουν αν υπάρχει αλλαγή για την κεντρική εικονοψηφίδα του τμήματος της εικόνας (Εικ. 2.13).



Εικόνα 2.13. Αρχιτεκτονική του δικτύου των Mou et al. (2019).

Οι Daudt et al. (2018) πρότειναν δύο αρχιτεκτονικές, την Early Fusion και την Siamese, οι οποίες εφαρμόστηκαν στο σετ εικόνων Opera Satellite Change Detection (“OSCD”). Στην πρώτη αρχιτεκτονική (“Early Fusion”) γινόταν πρώτα συγχώνευση του ζεύγους των τμημάτων των εικόνων (“patches”) (που αφορούσαν την ίδια περιοχή σε δύο διαφορετικές χρονικές στιγμές) κατά την διάσταση των καναλιών και έπειτα χρησιμοποιούνταν ως είσοδος σε μια σειρά 7 συνελικτικών και 2 πλήρως συνδεδεμένων επιπέδων. Η δεύτερη αρχιτεκτονική (“Siamese”) λάμβανε ως είσοδο 2 τμήματα εικόνων, τα οποία επεξεργάζονταν από δύο ξεχωριστούς αλλά με κοινά βάρη κλάδους με 4 συνελικτικά επίπεδα. Έπειτα, τα διανύσματα χαρακτηριστικών που προέκυπταν ως έξοδος συγχωνεύονταν και προωθούνταν σε μια σειρά δύο πλήρως συνδεδεμένων επιπέδων. Η έξοδος κάθε δικτύου είναι ένα ζεύγος τιμών που εκφράζει την πιθανότητα το τμήμα της εικόνας να ανήκει σε κάθε κατηγορία (“μη αλλαγή”, “αλλαγή”). Αυτές ήταν οι πρώτες αρχιτεκτονικές που εκπαιδεύτηκαν end-to-end στο πεδίο της ανίχνευσης μεταβολών, χωρίς την χρήση προεκπαιδευμένου σε άλλες εικόνες μοντέλου.

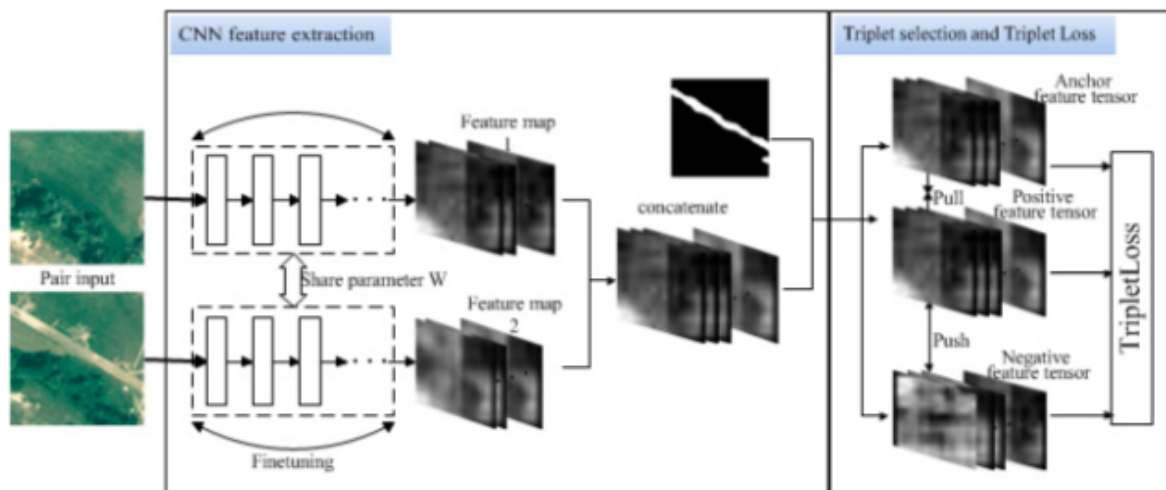
Ακολούθως οι Daudt et al (2018) πρότειναν τρεις αρχιτεκτονικές με U-Net μορφή, ως εξέλιξη των προηγούμενων (Εικ. 2.14). Η αρχιτεκτονική “fully convolutional Early Fusion” (“FC-EF”) λαμβάνει ως είσοδο το συγχωνευμένο ζεύγος τμημάτων εικόνων και σταδιακά μειώνει την ανάλυσή του μέσα στον κωδικοποιητή, ενώ στον αποκωδικοποιητή αυξάνεται ξανά η ανάλυσή τους, προσθέτοντας πληροφορία από τα συμμετρικά επίπεδα του κωδικοποιητή με την χρήση συνδέσεων παράλειψης (skip connections). Η αρχιτεκτονική “fully convolutional siamese concatenation” (“FC-Siam-Conc”), χρησιμοποιεί δύο διαφορετικούς κλάδους με κοινά βάρη ως κωδικοποιητή, οι οποίοι λαμβάνουν ως είσοδο ένα από τα 2 τμήματα εικόνων αντίστοιχα. Σε αυτή την περίπτωση, σε κάθε επίπεδο του αποκωδικοποιητή προστίθεται η πληροφορία του αντίστοιχου επιπέδου και από τους 2 κλάδους του κωδικοποιητή, μέσω των συνδέσεων παράλειψης. Τέλος, η αρχιτεκτονική “FC-Siam-Diff” είναι παρόμοια με την προηγούμενη, με την διαφορά ότι σε κάθε επίπεδο του αποκωδικοποιητή ενσωματώνεται ως πληροφορία από τον κωδικοποιητή η απόλυτη διαφορά των διανυσμάτων των χαρακτηριστικών των 2 κλάδων.



Εικόνα 2.14. Σχηματική απεικόνιση των τριών αρχιτεκτονικών των Daudt et al (2018).

Οι Zhang et al. (2018) χρησιμοποίησαν το DeepLabv2 των Chen et al. (2017) που αναλύθηκε νωρίτερα για το πρόβλημα της ανίχνευσης μεταβολών. Αρχικά το ζεύγος εικόνων

υπέστη προεπεξεργασία και ειδικότερα αντιστοίχιση ιστογράμματος. Στη συνέχεια εισάχθηκε σε ένα βαθύ σιαμαίο δίκτυο και παράχθηκαν δύο χάρτες χαρακτηριστικών, ένας για κάθε εικόνα. Το δίκτυο αυτό έχει χρησιμοποιηθεί για το αντικείμενο της κατάρτησης εικόνας και εδώ χρησιμοποιείται για να εξαχθούν χαρακτηριστικά που θα χρησιμοποιηθούν για την ανίχνευση μεταβολών. Με την χρήση της συνελίξης *atrous* αυξάνεται το αντιληπτικό πεδίο χωρίς να αυξηθεί το πλήθος των παραμέτρων ή το υπολογιστικό κόστος. Επίσης, με το ASPP, μπορούν να ανιχνευτούν καλύτερα μεγάλες και μικρές περιοχές αλλαγών, καθώς εφαρμόζονται παράλληλα συνελίξεις *atrous* στο ίδιο επίπεδο, με 4 διαφορετικά ρυθμούς. Στη συνέχεια τα χαρακτηριστικά υπόκεινται περαιτέρω ξεχωριστή επεξεργασία και ακολούθως συνενώνονται για να δημιουργήσουν τον τελικό χάρτη χαρακτηριστικών. Έπειτα με διγραμμική παρεμβολή κάθε χάρτη χαρακτηριστικών παίρνει τις διαστάσεις των αρχικών εικόνων και υπολογίζεται χάρτης αποστάσεων, με την εφαρμογή Ευκλείδειας απόστασης μεταξύ των χαρτών χαρακτηριστικών. Τέλος, με την εφαρμογή ενός κατωφλιού, παράγεται από τον χάρτη αποστάσεων το τελικό ζητούμενο, ο χάρτης αλλαγών (Εικ. 2.15).



Εικόνα 2.15. Αρχιτεκτονική του δικτύου των Zhang et al. (2018).

Μια ακόμη καινοτομία των Zhang et al. (2018) είναι η δημιουργία μιας καινούριας συνάρτησης σφάλματος, η οποία δεν έχει στόχο μόνο να αυξήσει την διαφορά μεταξύ θετικών και αρνητικών παραδειγμάτων, αλλά και να μειώσει την διαφορά μεταξύ των παραδειγμάτων που ανήκουν στην ίδια κατηγορία.

### 2.3.3 Αρχιτεκτονικές με Συνελικτικά LSTM

Τα Αναδρομικά Νευρωνικά Δίκτυα (Recurrent Neural Networks-RNN) συνιστούν ένα τύπο νευρωνικών δικτύων, τα οποία προωθούν την έξοδό τους ξανά προς την είσοδό τους. Ακόμη, μπορούν να επεξεργάζονται δεδομένα μεταβλητού μήκους και διαθέτουν μνήμη, γεγονός που τους επιτρέπει να ανακαλύπτουν συσχετίσεις μεταξύ των δεδομένων εισόδου. Τα δίκτυα αυτά χρησιμοποιούνται για κάθε είδους δεδομένα με ακολουθίες, όπως η φωνή, η γραφή και η οπτική πληροφορία που προκύπτει από μία κίνηση. Είναι σύνηθες να χρησιμοποιούνται τα Αναδρομικά Νευρωνικά Δίκτυα για προβλήματα που χρησιμοποιούν δεδομένα που εξαρτώνται από τον χρόνο, αφού είναι ικανά να περιγράψουν την χρονική συσχέτιση μεταξύ χαρακτηριστικών σε ακολουθία, όπως για παράδειγμα οι εικόνες που απεικονίζουν μια περιοχή σε διαφορετικές χρονικές στιγμές.

Στην πιο απλή τους μορφή, αυτά τα δίκτυα επεξεργάζονται δεδομένα της μορφής  $X=[X_1, X_2, \dots, X_T]$ , όπου  $X \in \mathbb{R}^N$  είναι μια λίστα που περιλαμβάνει πληροφορία συσχετισμένη με  $t \in [1, \dots, T]$  διαφορετικά χρονικά βήματα. Σε κάθε χρονικό βήμα  $t$ , το αντίστοιχο στοιχείο της λίστας  $X_t$  πολλαπλασιάζεται ανά στοιχείο με έναν αντίστοιχο πίνακα βαρών  $W_x$ . Ταυτόχρονα, μια αναπαράσταση των προηγούμενων στοιχείων της λίστας, γνωστή ως “hidden state”, πολλαπλασιάζεται με έναν πίνακα βαρών  $W_h$ . Το άθροισμα αυτών των ποσοτήτων τροφοδοτείται σε μια συνάρτηση υπερβολικής εφαπτομένης για να παραχθεί η αναπαράσταση “hidden state” του τρέχοντος χρονικού βήματος, ως εξής:

$$H_t = \tanh(W_x \cdot X_t + W_h \cdot X_{t-1})$$

Σημειώνεται ότι οι πίνακες βαρών παραμένουν ίδιοι σε όλα τα χρονικά βήματα, καθώς και ότι οι προκαταλήψεις (biases) έχουν παραλειφθεί για λόγους ευκολίας. Επειδή τα βάρη είναι ίδια σε όλα τα χρονικά βήματα, το δίκτυο μπορεί να παρουσιάζει το πρόβλημα των συρρικνούμενων κλίσεων (vanishing gradient).

Το πρόβλημα αντιμετωπίζεται με έναν ειδικό τύπο Αναδρομικών Νευρωνικών Δικτύων, τα Long Short-Term Memory Network (LSTM), που προτάθηκαν από τους Hochreiter & Schmidhuber (1997). Σε αυτό το δίκτυο προστέθηκε ένα κελί μνήμης, γνωστό ως “cell state”, που χρησιμοποιεί συναρτήσεις πύλης (“gating functions”) για να φιλτράρει πιο αποτελεσματικά την πληροφορία που ρέει στο δίκτυο. Σε αντίθεση με τα συμβατικά Αναδρομικά Νευρωνικά Δίκτυα που χρησιμοποιούν ένα μόνο επίπεδο υπερβολικής εφαπτομένης σε κάθε χρονικό βήμα, τα LSTM χρησιμοποιούν τέσσερα αλληλοσυνδεδεμένα επίπεδα γνωστά και ως “πύλες”. Οι πύλες forget και input χρησιμοποιούν την σιγμοειδή συνάρτηση και τους αντίστοιχους πίνακες  $W_f$  και  $W_i$ . Η πύλη forget προσπαθεί να αφαιρέσει ανεπαρκείς αναπαραστάσεις χαρακτηριστικών, ενώ η πύλη input καθορίζει ποιες πληροφορίες θα χρησιμοποιηθούν για την ενημέρωση του cell state. Η πύλη cell χρησιμοποιεί την συνάρτηση υπερβολικής εφαπτομένης για να ρυθμίσει τα δεδομένα και να υπολογίσει πιθανές τιμές για το cell state, ενώ η πύλη output φιλτράρει περαιτέρω την πληροφορία και υπολογίζει την έξοδο του δικτύου.

$$f_t = \sigma(W_f \cdot (X_t, H_{t-1}))$$

$$i_t = \sigma(W_i \cdot (X_t, H_{t-1}))$$

$$c_t = \tanh(W_c \cdot (X_t, H_{t-1}))$$

$$o_t = \sigma(W_o \cdot (X_t, H_{t-1}))$$

Έπειτα το δίκτυο είναι έτοιμο να υπολογίσει το νέο cell state “ξεχνώντας” τα χαρακτηριστικά που δεν χρειάζονται από το προηγούμενο cell state και κρατώντας μόνο τα χρήσιμα χαρακτηριστικά για το τρέχον cell state.

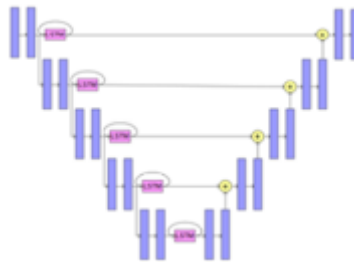
$$C_t = f_t \cdot C_{t-1} + i_t \cdot c_t$$

Αφού υπολογιστεί το cell state, εισάγεται σε μια συνάρτηση υπερβολικής εφαπτομένης και πολλαπλασιάζεται με το αποτέλεσμα της output gate για να παράξει το hidden state για το τρέχον χρονικό βήμα  $t$ .

$$H_t = o_t \cdot \tanh(C_t)$$

Στην περίπτωση όπου τα δεδομένα είναι υψηλών διαστάσεων, όπως οι εικόνες που χρησιμοποιούνται στην τηλεπισκόπηση  $I=[I_1, I_2, \dots, I_T]$ , όπου  $I_t \in \mathbb{R}^{C_h \times H \times W}$ , με  $C_h$  τα φασματικά κανάλια της εικόνας και  $H, W$  οι διαστάσεις της εικόνας, οι προηγούμενες σχέσεις θα περιλάμβαναν μεγάλο αριθμό παραμέτρων και θα ήταν αναγκαίος ο πολλαπλασιασμός τεράστιων πολυδιάστατων πινάκων. Αυτό θα οδηγούσε στο να γίνει η εκπαίδευση υπολογιστικά ακριβή και θα εμπόδιζε την σύγκλιση του μοντέλου.

Για να αντιμετωπιστεί αυτό το πρόβλημα, οι Papadomanolaki et al. (2019) αντικατέστησαν τους πίνακες βαρών  $W_f, W_i, W_c$  και  $W_o$  με συνελκτικά επίπεδα με φίλτρα μεγέθους  $3 \times 3$ ,  $\text{padding}=1$  και  $\text{stride}=1$ . Με αυτό τον τρόπο, μπορεί να γίνει end-to-end εκπαίδευση. Ειδικότερα, χρησιμοποίησαν μία αρχιτεκτονική τύπου U-Net, όπου η είσοδος έχει διαστάσεις  $B_s \times C_h \times H \times W$ , με  $B_s$  το μέγεθος του batch που έχει οριστεί. Καθεμιά από τις εικόνες  $I_t$  επεξεργάζεται ξεχωριστά από τον κωδικοποιητή (Εικ. 2.16).



Εικόνα 2.16. Αρχιτεκτονική του δικτύου των Papadomanolaki et al. (2019).

Στον κωδικοποιητή εισάγεται η εικόνα από κάθε ημερομηνία ξεχωριστά. Κάθε συνελκτικό μπλοκ περιέχει ένα συνελκτικό επίπεδο, ένα επίπεδο batch normalization και ένα επίπεδο ReLU. Όλα τα συνελκτικά επίπεδα χρησιμοποιούν φίλτρα μεγέθους  $3 \times 3$  με βήμα και προσαύξηση (padding) ίσα με 1. Στο πρώτο συνελκτικό μπλοκ κατασκευάζονται 16 χάρτες χαρακτηριστικών (feature maps), ενώ το ύψος και πλάτος παραμένει ίδιο με της εικόνας εισόδου. Σε καθένα από τα επόμενα τέσσερα μπλοκ διπλασιάζεται ο αριθμός των χαρτών χαρακτηριστικών και εφαρμόζεται μέγιστη συγκέντρωση μεγέθους  $2 \times 2$ . Οπότε στο τέλος του κωδικοποιητή οι διαστάσεις σε σχέση με την εικόνα εισόδου έχουν μειωθεί στο  $1/4$  και έχουν δημιουργηθεί 256 χάρτες χαρακτηριστικών.

Τα LSTM που βρίσκονται μετά από κάθε επίπεδο του κωδικοποιητή υπολογίζουν την χρονική σχέση των εξόδων τους. Τα cell state και hidden state αρχικοποιούνται ως μηδενικοί πίνακες μεγέθους  $B_s \times C_h \times H \times W$ . Κάθε μηχανισμός πύλης ορίζεται πλέον από την σχέση:

$$G_k^t = \Phi(W_{G_k} * (X_k^t, H_k^{t-1}))$$

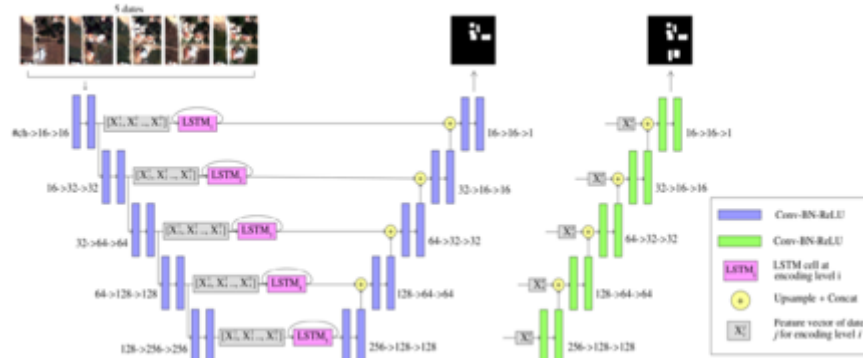
όπου  $G_k^t$  είναι η πύλη forget, input, output ή cell την χρονική στιγμή  $t$  στο επίπεδο  $k$  του αποκωδικοποιητή,  $\Phi$  είναι συνάρτηση ενεργοποίησης και  $W_{G_k}^t$  είναι ένα συνελκτικό επίπεδο που εφαρμόζεται στη συνένωση της τρέχουσας εισόδου  $X_k^t$  και του προηγούμενου hidden state  $H_k^{t-1}$  στη διάσταση των καναλιών.

Ακολούθως, υπολογίζεται η τιμή του τρέχοντος cell state και hidden state με τις σχέσεις που είχαν αναφερθεί προηγουμένως. Το τελικό hidden state σε κάθε επίπεδο του κωδικοποιητή αποθηκεύεται ώστε αργότερα να προστεθεί στο αποτέλεσμα του αντίστοιχου επιπέδου του αποκωδικοποιητή.

Στη συνέχεια, ο αποκωδικοποιητής λαμβάνει την έξοδο από το τελευταίο επίπεδο του κωδικοποιητή συνενωμένη για όλες τις ημερομηνίες και χρησιμοποιώντας πέντε συνελκτικά μπλοκ αντίστοιχα με του κωδικοποιητή αυξάνει την ανάλυση έως τις αρχικές διαστάσεις. Αυτό γίνεται με την εφαρμογή αύξησης της ανάλυσης (“upsampling”) μεγέθους  $2 \times 2$ , αντίστοιχες με εκείνη της μείωσης ανάλυσης μέσω μέγιστης συγκέντρωσης (max pooling). Ακόμη, μετά από κάθε αύξηση της ανάλυσης, προστίθεται στο αποτέλεσμα η έξοδος από το LSTM του αντίστοιχου επιπέδου του κωδικοποιητή. Με αυτό τον τρόπο συνδυάζεται πληροφορία υψηλότερης ανάλυσης με χαμηλότερης ανάλυσης δημιουργώντας πιο σύνθετα χαρακτηριστικά και διατηρώντας έτσι την χωρικά και χρονική γνώση. Το τελικό επίπεδο του δικτύου είναι ένα  $1 \times 1$  συνελκτικό επίπεδο, το οποίο υπολογίζει τον τελικό χάρτη μεταβολών.

Για την εκπαίδευση, παράχθηκαν τμήματα εικόνων μεγέθους  $32 \times 32$  με βήμα 6 σε περίπτωση που περιλαμβάνονται εικονοψηφίδες με ετικέτα “αλλαγή” ή με βήμα 32 στην αντίθετη περίπτωση. Αυτή η τακτική χρησιμοποιείται ως μια μέθοδος επαύξησης δεδομένων, για να δημιουργηθούν περισσότερα δεδομένα εκπαίδευσης που να περιλαμβάνουν αλλαγές. Επιπλέον, έγινε στροφή των τμημάτων εικόνων που είχαν παραπάνω από 5% εικονοψηφίδες με ετικέτα “αλλαγή” σε όλες τις πιθανές γωνίες ανάλογες των 90 μοιρών. Τέλος, κάθε κλάση (αλλαγή/μη αλλαγή) συσχετίζεται με ένα βάρος αντιστρόφως ανάλογο με τον συνολικό αριθμό εικονοψηφίδων που περιλαμβάνει. Για να παραχθούν πιο στιβαρά αποτελέσματα, οι τελικές προβλέψεις προκύπτουν από ένα σύνολο από 5 διαφορετικά εκπαιδευμένα μοντέλα, ακολουθώντας την διαδικασία της διασταυρούμενης επικύρωσης (cross-validation). Ως βελτιστοποιητής χρησιμοποιήθηκε ο Adam, το μέγεθος του batch είναι 64 και ο ρυθμός μάθησης είναι 0.0001.

Οι Papadomanolaki et al. (2020) πρόσθεσαν στην υπάρχουσα αρχιτεκτονική έναν επιπλέον κλάδο αποκωδικοποιητή που πραγματοποιεί σημασιολογική κατάτμηση (semantic segmentation) στα κτήρια και παρέχει στο δίκτυο βοηθητικά δεδομένα για τα χαρακτηριστικά κατά την εκπαίδευση (Εικ. 2.17).



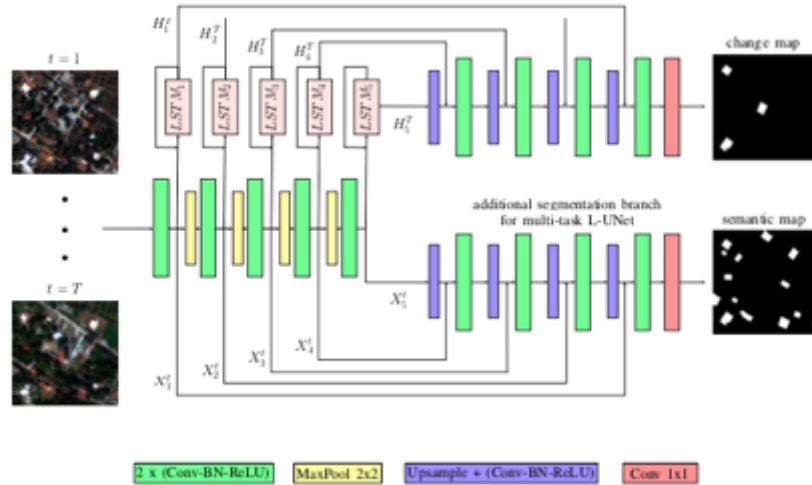
Εικόνα 2.17. Αρχιτεκτονική του δικτύου των Papadomanolaki et al. (2020).

Σε κάθε επίπεδο αυτού του κλάδου προστίθενται τα αντίστοιχα χωρικά χαρακτηριστικά από τον κλάδο του κωδικοποιητή, και όχι η χρονική πληροφορία από τα LSTM όπως στον άλλο κλάδο του αποκωδικοποιητή. Σημειώνεται ότι οι χάρτες που δείχνουν την κατάτμηση των κτηρίων μπορούν να παραχθούν για όλες τις ημερομηνίες ή για μερικές, ανάλογα με την εφαρμογή και την επιθυμητή υπολογιστική πολυπλοκότητα. Οπότε πλέον η αρχιτεκτονική αποτελείται από έναν κωδικοποιητή με 5 επίπεδα και 2 κλάδους αποκωδικοποιητή.

Ακόμη, γίνεται χρήση 5 διαφορετικών σφαλμάτων κατά την διάρκεια της εκπαίδευσης, που χρησιμοποιούνται με διαφορετικά βάρη για να συνεισφέρουν στο συνολικό σφάλμα που χρειάζεται για την βελτιστοποίηση του δικτύου. Όλα τα losses είναι τύπου cross entropy loss. Το πρώτο loss αφορά την δημιουργία του χάρτη ανίχνευσης μεταβολών για τα κτήρια, τα δύο επόμενα losses αφορούν την κατασκευή του σημασιολογικού χάρτη για τα κτήρια για την πρώτη και την τελευταία ημερομηνία. Το τέταρτο loss αφορά την ανίχνευση μεταβολών στα κτήρια και βασίζεται στο αποτέλεσμα της αφαίρεσης της σημασιολογικής κατάτμησης των κτηρίων την τελευταία ημερομηνία και την πρώτη ημερομηνία. Το τελευταίο loss αφορά την σημασιολογική κατάτμηση των κτηρίων για την τελευταία ημερομηνία, και προκύπτει από το άθροισμα της σημασιολογικής κατάτμησης των κτηρίων της πρώτης ημερομηνίας και την ανίχνευσης μεταβολών. Τα δύο τελευταία losses συνδυάζουν τα αποτελέσματα των δύο decoding κλάδων και βοηθούν ώστε να μειωθεί ο αριθμός των ψευδών θετικών παραδειγμάτων (false positives) στον χάρτη μεταβολών των κτηρίων.

Οι Papadomanolaki et al. (2021) πρόσθεσαν στην παραπάνω περιγραφείσα αρχιτεκτονική (“multi-task L-UNet”) επίπεδα dropout στα hidden states των μπλοκ LSTM (Εικ. 2.18).



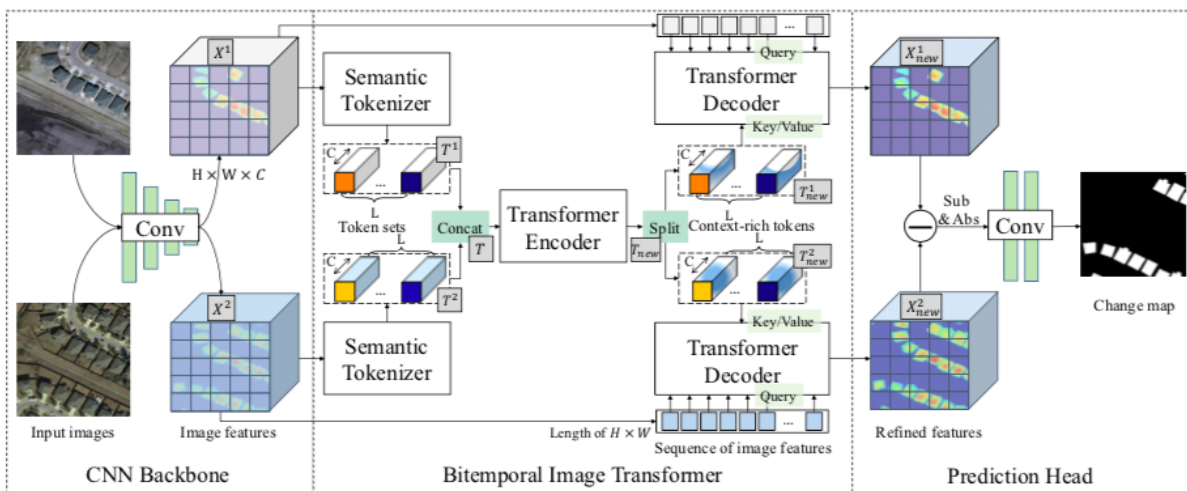


Εικόνα 2.18. Αρχιτεκτονική του δικτύου των Papadomanolaki et al. (2021).

Αξίζει να σημειωθεί ότι συνελκτικά αναδρομικά επίπεδα είχαν επίσης χρησιμοποιηθεί από τους Russwurm & Koerner (2018) για το αντικείμενο της ταξινόμησης της κάλυψης γης.

### 2.3.4 Αρχιτεκτονικές με μηχανισμούς προσοχής και transformers

Οι Chen et al. (2021) χρησιμοποίησαν έναν transformer (BIT model). Αρχικά χρησιμοποιούν το ResNet για να εξάγουν σημασιολογικά χαρακτηριστικά υψηλού επιπέδου από το ζεύγος εικόνων που εισάγεται. Έπειτα εφαρμόζουν χωρική προσοχή (spatial attention) για να μετατρέψουν κάθε χρονικό χάρτη χαρακτηριστικών σε ένα σετ σημασιολογικών token και χρησιμοποιούν έναν κωδικοποιητή transformer για να μοντελοποιήσουν την πληροφορία μεταξύ των δύο σετ από tokens. Στη συνέχεια το αποτέλεσμα επαναπροβάλλεται στον χώρο των εικονοψηφίδων από ένα σιαμαίο αποκωδικοποιητή transformer, για να ενισχύσουν τα αρχικά χαρακτηριστικά. Τέλος, υπολογίζονται οι εικόνες διαφορών των χαρακτηριστικών από τους βελτιωμένους χάρτες χαρακτηριστικών και εισάγονται σε ένα CNN για να παραχθούν προβλέψεις για τις αλλαγές σε επίπεδο εικονοψηφίδας (Εικ. 2.19).

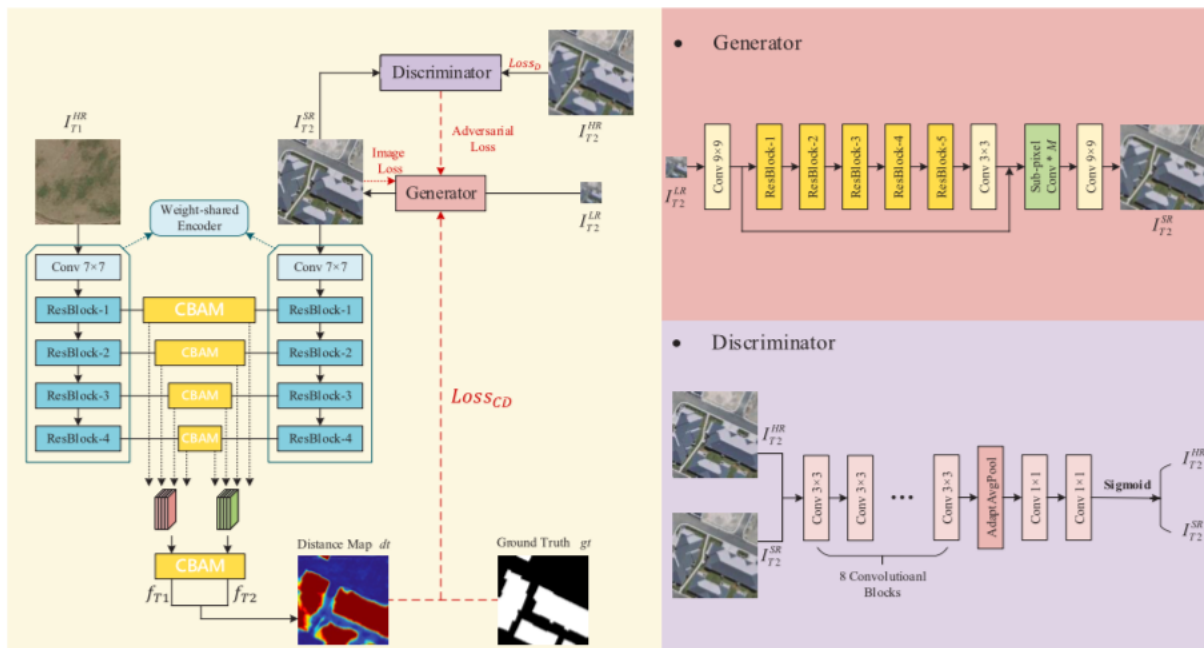


Εικόνα 2.19. Σχηματική απεικόνιση της προτεινόμενης αρχιτεκτονικής από τους Chen et al. (2021).

Οι Liu et al. (2021) εφάρμοσαν ανίχνευση μεταβολών σε ζεύγη εικόνων με διαφορετική ανάλυση, ένα σενάριο το οποίο μπορεί να υπάρξει στην πραγματική ζωή αν παραστεί



ανάγκη για γρήγορα αποτελέσματα σε μεγάλη κλίμακα και χρησιμοποιηθούν εικόνες από διαφορετικές πηγές. Αρχικά το δίκτυο που προτείνουν (SRCDNet) εφαρμόζει το δομικό στοιχείο SRM για να γίνει μάθηση μέσω ενός GAN (Generative Adversarial Network) μιας εικόνας super-resolution από τις χαμηλής ανάλυσης εικόνες ώστε να ανακτηθεί περισσότερη σημασιολογική πληροφορία και να αποφευχθούν τα πλεονάζοντα λάθη. Η εικόνα super-resolution που παράγεται στη συνέχεια εισάγεται σε ένα εξαγωγέα χαρακτηριστικών μαζί με τις εικόνες υψηλής ανάλυσης που αφορούν τις υπόλοιπες ημερομηνίες. Ακόμη, προστίθεται ένα stacked attention module (SAM) που περιλαμβάνει 5 συνελικτικά attention modules (CBAM). Έπειτα παράγεται ο χάρτης απόστασης και συγκρίνεται με τον αληθή, ενώ χρησιμοποιείται ένα contrastive loss, το οποίο χρησιμοποιείται συχνά στο metric learning, ώστε να αυξάνεται η απόσταση για τις αλλαγμένες περιοχές και να μειώνεται η απόσταση όπου δεν υπάρχει αλλαγή. Τέλος, ο χάρτης αλλαγών προκύπτει από τον χάρτη απόστασης με την εφαρμογή ενός κατωφλιού (Εικ. 2.20).

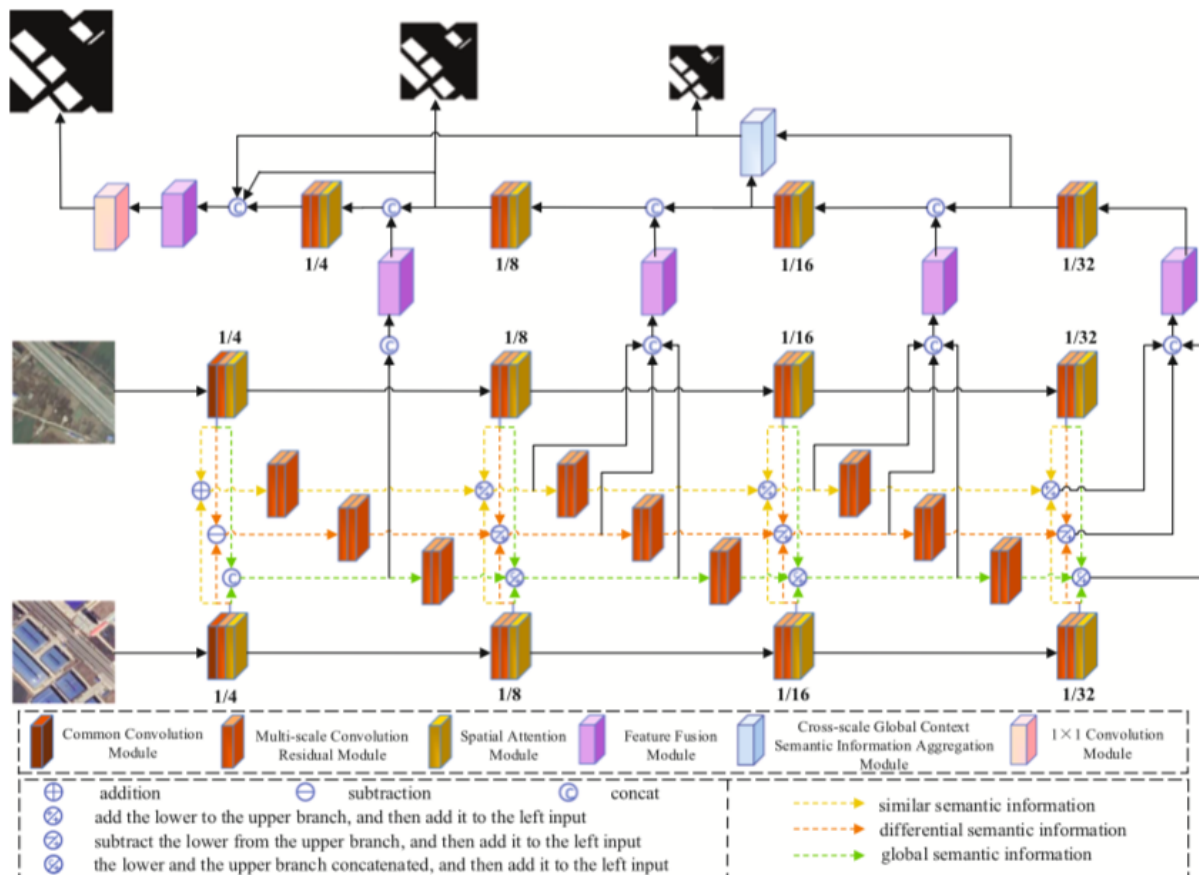


Εικόνα 2.20. Σχηματική απεικόνιση της προτεινόμενης αρχιτεκτονικής από τους Liu et al. (2021).

Οι Song et al. (2021) πρότειναν το SUACDNet, ένα σιαμαίο δίκτυο με U-Net αρχιτεκτονική. Στο στάδιο της κωδικοποίησης χαρακτηριστικών (feature encoding), προστέθηκαν 3 επιπλέον κλάδοι που εστιάζουν στον υπολογισμό της συνολικής (global) πληροφορίας, της διαφοράς και της ομοιότητας των πληροφοριών των 2 εικόνων. Με αυτό τον τρόπο, γίνεται συσχετισμός των πληροφοριών των δύο εικόνων από νωρίς. Στο στάδιο της αποκωδικοποίησης (feature decoding), οι πληροφορίες από τους κλάδους προστίθενται ανά επίπεδο ανάλυσης ώστε να διατηρηθούν κατά το δυνατόν οι λεπτομέρειες των ακμών από τις εικόνες εισόδου.

Ακόμη, χρησιμοποιούνται τρία επιπλέον δομικά στοιχεία (Spatial Attention Module, Feature Fusion Module, Cross-scale Global Context Semantic Information Aggregation Module) που βοηθούν την απόδοση του δικτύου, κάνοντάς το πιο ευαίσθητο στις αλλαγές που ανιχνεύονται ενώ παράλληλα φιλτράρεται και απομακρύνεται ο θόρυβος. Ειδικότερα, το SAM μαθαίνει στο δίκτυο να δίνει περισσότερη προσοχή στις περιοχές με μεταβολές στην δόμηση, το FFM συσσωματώνει χαρακτηριστικά με βάρη από τις διάφορες κατηγορίες σημασιολογικής πληροφορίας και το CGCAM χρησιμοποιείται για να εξαγάγει συνολική

σημασιολογική πληροφορία από τα βαθιά χαρακτηριστικά των εικόνων και για να τα ενσωματώσει στους χάρτες χαρακτηριστικών στις διάφορες αναλύσεις. Τέλος, για την εξαγωγή των χαρακτηριστικών στις διάφορες αναλύσεις, χρησιμοποιείται μια νέα δομή συνελκτικών επιπέδων, το Convolution Residual Module (Εικ. 2.21).



Εικόνα 2.21. Σχηματική απεικόνιση της προτεινόμενης αρχιτεκτονικής από τους Song et al. (2021).

Εκτός του πεδίου της ανίχνευσης μεταβολών, ένα δίκτυο με χρονική προσοχή (temporal attention- Russwurm & Koerner, 2019) έχει χρησιμοποιηθεί από τους Sainte Fare Garnot et al. (2020) για την ταξινόμηση της κάλυψης γης από ακολουθίες δορυφορικών εικόνων. Πιο συγκεκριμένα, γεωτεμάχια από εικόνες του δορυφόρου Sentinel-2 όπως λήφθηκαν σε διάφορες χρονικές στιγμές μετασχηματίστηκαν σε αταξινομήτα σετ εικονοψηφίδων. Κάθε σετ εισάχθηκε σε έναν κωδικοποιητή και παράχθηκε ένας περιγραφέας χαρακτηριστικών, ο οποίος στη συνέχεια εισάχθηκε σε ένα δίκτυο χρονικής προσοχής.

### 2.3.5 Multi-task αρχιτεκτονικές

Σε ορισμένες περιπτώσεις έχουν δοκιμαστεί αρχιτεκτονικές που έχουν παραπάνω από ένα αντικείμενα, δηλαδή εκτός από το πρόβλημα της ανίχνευσης μεταβολών υπολογίζουν για παράδειγμα την κατάτμηση της εικόνας. Θεωρείται ότι με αυτό τον τρόπο παρέχονται παραπάνω πληροφορίες στα μοντέλα κατά την διάρκεια της εκπαίδευσης, βελτιώνοντας με αυτό τον τρόπο την απόδοσή τους.

Οι Liu et al. (2019) συνδύασαν την ανίχνευση μεταβολών σε αστικό περιβάλλον με την σημασιολογική κατάτμηση κτηρίων με την χρήση ενός πλήρους συνελκτικού σιαμαίου δικτύου. Αντίστοιχα, οι Daudt et al. (2018) συνδύασαν την ανίχνευση μεταβολών με τον

υπολογισμό της κάλυψης γης. Ειδικότερα, σε πρώτη φάση η εκπαίδευση εστιάζει στην αναγνώριση των διάφορων σημασιολογικών κατηγοριών της κάλυψης γης και σε δεύτερη φάση το δίκτυο επανεκπαιδεύεται για την ανίχνευση μεταβολών χρησιμοποιώντας ως αρχικοποίηση τα βάρη που είχαν υπολογιστεί προηγουμένως.

Τέλος, οι Yang et al. (2019) χρησιμοποίησαν transfer learning για να μειώσουν την ανισορροπία μεταξύ των δεδομένων για τα οποία ήταν διαθέσιμα ετικέτες και για τα δεδομένα που δεν είχαν. Ειδικότερα, ένα δίκτυο με αρχιτεκτονική κωδικοποιητή-αποκωδικοποιητή (encoder-decoder) πραγματοποιεί ανίχνευση μεταβολών σε ζεύγη εικόνων για τα οποία είναι διαθέσιμοι οι αληθείς χάρτες μεταβολών. Παράλληλα, το δίκτυο υπολογίζει τις διαφορές των δεδομένων χωρίς ετικέτες, ενισχύοντας τα εξαγόμενα χαρακτηριστικά κατά την διάρκεια της εκπαίδευσης. Έπειτα από την προεκπαίδευση, γίνεται εφαρμογή μεθόδων για να βελτιωθεί περαιτέρω το αποτέλεσμα (fine-tuning) ώστε να γίνει μη επιβλεπόμενη εκπαίδευση στα δεδομένα χωρίς ετικέτες.

## 2.4 Διαθέσιμα σετ δεδομένων για ανίχνευση μεταβολών

Οι Benedeck & Szirnyi (2009) δημιούργησαν το Air Change Dataset, το οποίο περιλαμβάνει περίπου 8 χιλιάδες εικονοψηφίδες με ετικέτες, διαιρεμένες σε 3 περιοχές. Οι εικόνες προέρχονται από το Ουγγρικό Ινστιτούτο Γεωδαισίας, Χαρτογραφίας και Τηλεπισκόπησης και το Google Earth και υπάρχει μεγάλη διαφορά μεταξύ της ανάλυσης των εικόνων, όπως και της έκθεσης που έχουν. Πιο συγκεκριμένα, το πρώτο τμήμα του σετ δεδομένων περιλαμβάνει 7 ζεύγη εικόνων ανάλυσης 1.5 μέτρου ληφθέντα από το 2000 και το 2005, το δεύτερο τμήμα αποτελείται από 5 ζεύγη εικόνων από το 2000 και το 2007, ενώ το τρίτο σετ εικόνων αποτελείται από εικόνες του Google Earth από το 1984 και το 2007.

Στη συνέχεια οι Daudt et al. (2018) δημιούργησαν το OSCD σετ δεδομένων από εικόνες του δορυφόρου Sentinel 2, συνολικού μεγέθους 24 ζευγών εικόνων και 9 χιλιάδων εικονοψηφίδων με ετικέτες. Οι Papadomanolaki et al. (2019) επέκτειναν αυτό το σετ δεδομένων από 2 ημερομηνιών για κάθε εικόνα σε 5 διαθέσιμων ημερομηνιών, έχοντας ως στόχο να μειώσουν τα προβλήματα που αφορούν αλλαγές λόγω διαφοράς στην εποχή και όχι διαφορά στην δόμηση, που είναι το ζητούμενο.

Ακολούθως το High Resolution Semantic Change Detection σετ δεδομένων (Daudt et al., 2018) αποτελεί ένα από τα μεγαλύτερα διαθέσιμα σετ δεδομένων. Περιλαμβάνει 291 ζεύγη RGB εικόνων ανάλυσης 0.5 μέτρου από την βάση δεδομένων BD ORTHO του IGN (Institut national de l'information géographique et forestiere από την Γαλλία). Είναι διαθέσιμες ετικέτες αλλαγής και μη-αλλαγής, όπως και ετικέτες για 6 κατηγορίες κάλυψης γης, οι οποίες έχουν δημιουργηθεί από τους Urban Atlas 2006, Urban Atlas 2012 και Urban Atlas 2006-2012.

Οι Lebedev et al. (2018) κατασκεύασαν ένα σετ δεδομένων που περιέχει 16000 ζεύγη εικόνων από το Google Earth μεγέθους 256x256 εικονοψηφίδων, που απεικονίζουν την ίδια περιοχή αλλά σε διαφορετικές εποχές. Οι εικόνες έχουν πολύ υψηλή χωρική ανάλυση, της τάξης των 3-100 εκατοστών, οπότε δεν παρέχει πληροφορία μόνο για τις αλλαγές σε κτήρια, δρόμους και δάση αλλά και αντικειμένων όπως αυτοκίνητα και δεξαμενές.

Οι Ji et al. (2018) κατασκεύασαν ένα σετ δεδομένων με ζεύγη εικόνων ανάλυσης 0.2 μέτρων και χρησιμοποιήθηκε από τους Liu et al. (2021) και Song et al. (2021). Οι εικόνες αφορούν μια περιοχή όπου συνέβη ένας σεισμός μεγέθους 6.3 της κλίμακας Ρίχτερ τον Φεβρουάριο του 2011 και λήφθηκαν το 2012 και το 2016, οπότε έχουν συμβεί πολλές μεταβολές στα κτήρια λόγω της ανακατασκευής που ακολούθησε μετά από τον σεισμό. Επίσης, το xBD (Gurta et al., 2019) συνιστά ένα σετ δεδομένων που εστιάζει στις καταστροφές των κτηρίων μετά από φυσικές καταστροφές.

Το LEVIR-CD (Chen & Shi, 2020) είναι ένα σετ εικόνων μεγάλης κλίμακας που ανιχνεύει την μεταβολή των κτηρίων και συνολικά περιλαμβάνει 31333 διαφορές κτηρίων. Ακόμη, το SpaceNet7 (Van Etten et al., 2021) αποτελείται από δορυφορικές εικόνες του Planet με 4 μέτρα ανάλυση. Παρέχει περίπου 24 εικόνες (μια εικόνα ανά μήνα για 2 χρόνια) σε 100 θέσεις. Συνολικά περιλαμβάνει πάνω από 10 εκατομμύρια ετικέτες.

Οι Yang et al. (2020) κατασκεύασαν το σετ εικόνων SECOND (SEmantic Change detectiON Dataset), το οποίο περιέχει 4662 ζεύγη εναέριων εικόνων και 30 κατηγορίες μεταβολών. Οι εικόνες έχουν ληφθεί από διάφορες πλατφόρμες και σένσορες και έχουν μέγεθος 512 x 512 εικονοψηφίδες. Σημειώνεται ότι γίνεται ξεχωριστά επισήμανση της μεταβολής των εικονοψηφίδων και του είδους της μεταβολής.

Επίσης, οι Verma et al. (2021) δημιούργησαν το σετ δεδομένων QFabric, που περιλαμβάνει 450000 πολύγωνα αλλαγών που αντιστοιχούν σε 504 θέσεις εκατό πόλεων. Οι εικόνες προέρχονται από τον δορυφόρο WorldView-2 και λήφθηκαν μεταξύ του Ιανουαρίου 2014 και Ιουλίου 2020. Για καθεμιά από τις 504 θέσεις, επιλέχθηκαν εικόνες για 5 διαφορετικές ημερομηνίες έτσι ώστε η διάρκεια μεταξύ δύο συνεχόμενων σετ εικόνων να είναι παρόμοια. Αυτό το σετ δεδομένων μπορεί να χρησιμοποιηθεί σε τέσσερα διαφορετικά αντικείμενα: (δυναμική) ανίχνευση μεταβολών, ταξινόμηση των μεταβολών, παρακολούθηση του τύπου της αλλαγής (“change status tracking”) και ταξινόμηση γειτονιάς (“neighborhood classification”-αναγνώριση αστικών και γεωγραφικών χαρακτηριστικών γύρω από το πολύγωνο που έχει αλλάξει). Συνολικά περιέχει 6 τύπους αλλαγών (για το δεύτερο αντικείμενο) και 9 τύπους αλλαγών κατηγορίας (για το 3ο αντικείμενο), ενώ τα πολύγωνα αλλαγών των γειτονιών μπορούν να ταξινομηθούν σε 5 αστικές κατηγορίες και 11 γεωγραφικές κατηγορίες.

Τέλος, οι Song et al. (2021) κατασκεύασαν το σετ δεδομένων Bitemporal Remote Sensing Image Change Detection Dataset (BTCDD), το οποίο περιέχει 5281 ζεύγη εικόνων υψηλής ανάλυσης από το Google Earth με μέγεθος 256x256 εικονοψηφίδες. Οι εικόνες έχουν ληφθεί σε διαφορετικές ημερομηνίες μέσα στο διάστημα 2010-2020 σε διαφορετικές περιοχές στην Κίνα και απεικονίζουν εργοστάσια, αγροτικές περιοχές, δρόμους, κτήρια και περιοχές εξόρυξης.

Συγκεντρωτικά τα διαθέσιμα σετ εικόνων παρουσιάζονται στον Πίν. 2.1.

Δημοσίευση	Μέγεθος σετ εικόνων	Είδος εικόνων (χωρική ανάλυση)	Τύπος αλλαγής
Air Change Dataset (Benedeck & Szirnyi, 2009)	13 ζεύγη εικόνων	Διαφορετικές αναλύσεις: 1.5 μ./ Google Earth	Δυαδική
OSCD (Daudt et al., 2018)	24 ζεύγη εικόνων, 2 ημερομηνίες	Sentinel-2 (10 μ.)	Δυαδική
MultiDate OSCD (Papadomanolaki et al., 2019)	120 εικόνες, 5 ημερομηνίες	Sentinel-2 (10 μ.)	Δυαδική
High Resolution Semantic Change Detection (Daudt et al., 2018)	291 ζεύγη εικόνων	0.5 μ.	Δυαδική/ 6 κατηγορίες κάλυψης γης
Lebedev et al. (2018)	16000 ζεύγη εικόνων 256x256 εικ.	Google Earth 0.03-1 μ.	Δυαδική
Ji et al. (2018)	13000 ζεύγη εικόνων 256x256 εικ.	0.03-1 μ.	Δυαδική
xBD (Gupta et al., 2019)	22068 εικόνες	Από διάφορους σένσορες (ακρίβεια φωτοσταθερών 0.8 μ)	Ταξινόμηση (4 κατηγορίες)
LEVIR-CD (Chen & Shi, 2020)	637 ζεύγη εικόνων 1024 x 1024 εικ.	Google Earth (πολύ υψηλή ανάλυση)	Δυαδική
SpaceNet7 (Van Etten et al., 2021)	2389 εικόνες, 1024x1024 εικ., πολλές ημερομηνίες	4μ.	Δυαδική (για τα κτήρια)
SECOND (Yang et al., 2020)	4662 ζεύγη εικόνων 512 x 512 εικ.	Από διάφορες πλατφόρμες και σένσορες	Σημασιολογική
QFabric (Verma et al., 2021)	2520 εικόνες, 5 ημερομηνίες	0.31-0.7 μ. (WorldView-2)	Δυαδική, σημασιολογική, παρακολούθηση αλλαγής, ταξινόμηση γειτονιάς
Bitemporal Remote Sensing Image Change Detection Dataset (Song et al., 2021)	5281 ζεύγη εικόνων 256 x 256 εικ.	Google Earth (υψηλή ανάλυση)	Δυαδική

Πίνακας 2.1. Διαθέσιμα σετ εικόνων για ανίχνευση μεταβολών.

## 3. ΜΕΘΟΔΟΛΟΓΙΑ ΚΑΙ ΥΛΟΠΟΙΗΣΗ

Στο παρόν κεφάλαιο παρουσιάζονται τα δεδομένα που χρησιμοποιήθηκαν στο πλαίσιο της εργασίας, τα μοντέλα που υλοποιήθηκαν και εφαρμόστηκαν, οι τρόποι αξιολόγησης των αποτελεσμάτων, καθώς και οι λεπτομέρειες υλοποίησης.

### 3.1 Δεδομένα

Στην παρούσα εργασία χρησιμοποιήθηκαν δύο σετ δεδομένων, που αναλύονται στη συνέχεια.

#### 3.1.1 Onera Satellite Change Detection dataset (OSCD, Daudt et al. 2018)

Το σετ δεδομένων αποτελείται από ζεύγη εικόνων του δορυφόρου Sentinel-2 που απεικονίζει 24 διαφορετικές πόλεις και περιλαμβάνει 13 φασματικά κανάλια. Οι εικόνες έχουν μέγεθος περίπου 500x500 εικονοψηφίδες και ανάλυση 10-60 μ. Αφορούν χρονικό διάστημα περίπου δύο ετών, από τον Ιούνιο του 2015 έως το τέλος του 2017. Οι Papadomanolaki et al. (2019) αύξησαν κατά 3 τις εικόνες του σετ εικόνων και προσπάθησαν να χρησιμοποιήσουν εικόνες παρόμοιων ημερομηνιών και εποχών ανάμεσα στις υπάρχουσες ημερομηνίες, ώστε να προσομοιάζουν στις υπάρχουσες εικόνες του dataset. Είναι εστιασμένο στις αστικές περιοχές και δεν επισημαίνει αλλαγές στην βλάστηση.

Στην παρούσα εργασία χρησιμοποιούνται 14 από τις εικόνες για την εκπαίδευση και την επικύρωση (aguasclaras, bercy, bordeaux, nantes, paris, rennes, saclay\_e, abudhabi, cupertino, pisa, beihai, hongkong, beirut, mumbai) και 10 από τις εικόνες για την αξιολόγηση της απόδοσης (brasilia, montpellier, norcia, rio, saclay\_w, valencia, dubai, lasvegas, milano, chongqing).



Εικόνα 3.1. Ζεύγος εικόνων “abudhabi” και η αληθής εικόνα μεταβολών.

#### 3.1.2 LEVIR-CD dataset (Chen & Shi, 2020)

Το LEVIR-CD είναι ένα σετ δεδομένων μεγάλης κλίμακας που αφορά το αντικείμενο της δυαδικής ανίχνευσης μεταβολών (binary change detection). Αποτελείται από 637 πολύ υψηλής ανάλυσης ζεύγη RGB εικόνων από το Google Earth μεγέθους 1024 x 1024 εικονοψηφίδων, με μέγεθος εδαφοψηφίδας 0.5μ, που έχουν ληφθεί με διαφορά 5 έως 14

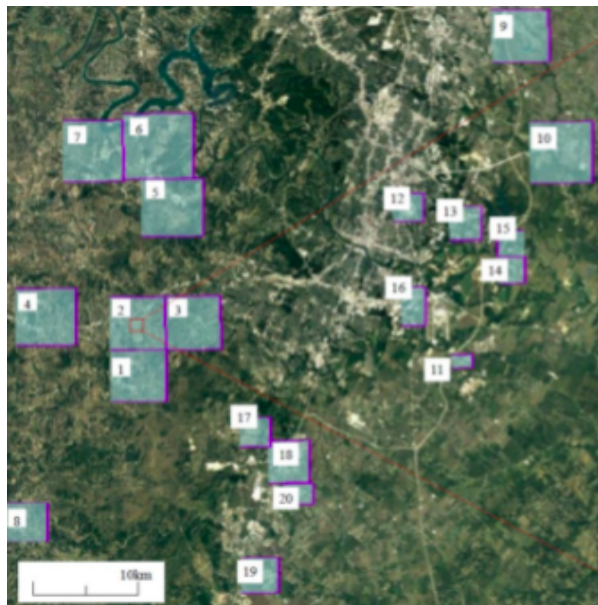


ετών και συνοδεύονται από τον αληθή (ground truth) χάρτη μεταβολών (change map) (Εικ. 3.2).



Εικόνα 3.2. Αντιπροσωπευτικό ζεύγος εικόνων και αληθής (ground truth) χάρτης μεταβολών του dataset LEVIR-CD.

Οι εικόνες απεικονίζουν 20 διαφορετικές περιοχές στο Τέξας των ΗΠΑ από το 2002 έως το 2018 (Εικ. 3.3). Περιέχουν αλλαγές λόγω των διαφορετικών εποχών και λόγω της διαφορετικής φωτεινότητας. Εστιάζει στην μεταβολή διάφορων ειδών κτηρίων, όπως βίλες και αποθήκες, ψηλά κτήρια ή μικρά γκαράζ.



Εικόνα 3.3. Περιοχές λήψης των εικόνων του dataset LEVIR-CD (Chen & Shi, 2020).

Οι μεταβολές έχουν επισημανθεί από ειδικούς φωτερμηνευτές, χρησιμοποιώντας την τιμή 1 ως ετικέτα για την αλλαγή και την τιμή 0 για την μη μεταβολή (Εικ. 3.4). Συνολικά το σετ δεδομένων αποτελείται από 31333 περιπτώσεις μεταβολών (<https://justchenhao.github.io/LEVIR/>).



Εικόνα 3.4. Ζεύγη εικόνων από το dataset LEVIR-CD και χαρακτηριστικά παραδείγματα μεταβολών (Chen & Shi, 2020).

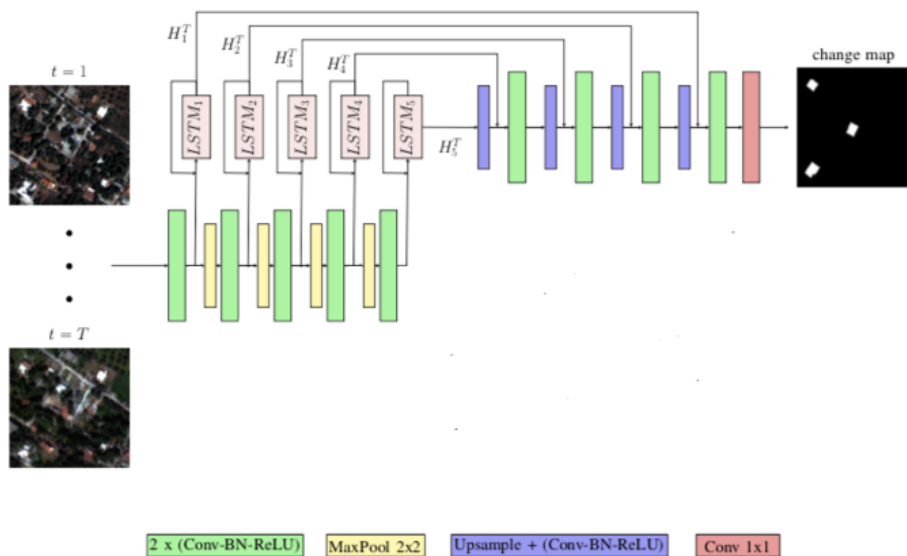
Στην παρούσα εργασία χρησιμοποιήθηκαν 100 από τις 445 για εκπαίδευση (train), μετά από αρκετές δοκιμές για το πόσες εικόνες είναι αρκετές ώστε να εκπαιδευτεί το δίκτυο ικανοποιητικά, χωρίς να χρειαστεί να χρησιμοποιηθούν όλες λόγω περιορισμένης υπολογιστικής ισχύος. Αντίστοιχα, χρησιμοποιήθηκαν 15 από τις 64 εικόνες επικύρωσης (validation), ενώ στον έλεγχο (test) χρησιμοποιήθηκαν όλες οι εικόνες (128).

## 3.2 Μοντέλα

Ακολούθως γίνεται περιγραφή των δύο μοντέλων που υλοποιήθηκαν και εφαρμόστηκαν για τις ανάγκες της εργασίας.

### 3.2.1 L-UNet (Papadomanolaki et al, 2019)

Ένα από τα μοντέλα που χρησιμοποιήθηκαν είναι το L-UNet (Papadomanolaki et al, 2019) (Εικ. 3.5), το οποίο αναλύθηκε στο Κεφάλαιο 2, ενότητα 2.3.3. Ειδικότερα, έγιναν πειράματα για το κατά πόσον χρειάζονται τα μπλοκ LSTM, αν είναι απαραίτητα σε όλα τα επίπεδα του κωδικοποιητή και μελετήθηκε πόσο συμβάλλει το κάθε μπλοκ LSTM στο τελικό αποτέλεσμα. Επίσης, έγινε σύγκριση με ένα δίκτυο με διεσταλμένα φίλτρα (δες ενότητα 3.2.2) πάνω σε δύο dataset, το OSCD και το LEVIR-CD.



Εικόνα 3.5. Αρχιτεκτονική του L-UNet (Papadomanolaki et al, 2019).



Στην πλήρη εκδοχή του δικτύου, στον κωδικοποιητή εισάγεται ξεχωριστά η εικόνα από κάθε ημερομηνία. Σε όλες τις εικόνες χρησιμοποιούνται τα ίδια βάρη στα επίπεδα του κωδικοποιητή. Μετά από κάθε επίπεδο του κωδικοποιητή υπάρχει ένα μπλοκ LSTM, όπου εκεί αποτυπώνεται η χρονική συσχέτιση των δύο εικόνων βάσει των χωρικών χαρακτηριστικών τους, που εισάγονται όπως αναφέρθηκε διαδοχικά. Το αποτέλεσμα του μπλοκ LSTM σε κάθε επίπεδο του κωδικοποιητή αποθηκεύεται ώστε αργότερα να προστεθεί στο αντίστοιχο επίπεδο του αποκωδικοποιητή.

Στις υπόλοιπες εκδοχές, αφαιρέθηκαν από ένα έως τέσσερα μπλοκ LSTM. Ειδικότερα, τα ονόματα των μοντέλων έχουν αντιστοιχηθεί με τους αριθμούς των επιπέδων του κωδικοποιητή, δηλαδή η περίπτωση hybrid\_1 αφορά την αφαίρεση του μπλοκ LSTM του 1ου επιπέδου του κωδικοποιητή, η περίπτωση hybrid\_345 αφορά την αφαίρεση των μπλοκ του 3ου, 4ου και 5ου επιπέδου του κωδικοποιητή και ούτω καθεξής. Συνολικά εξετάζονται 8 περιπτώσεις.

Τα συνελκτικά LSTM που χρησιμοποιούνται, αντί των συμβατικών, προτάθηκαν πρώτα από τους Shi et al. (2015) για άλλο αντικείμενο (precipitation nowcasting) και εν συνεχεία εφαρμόστηκαν από τους Russwurm & Koerner (2018) για την κατάτμηση εικόνων που είναι ανθεκτική στην ύπαρξη σύννεφων. Τα πλεονεκτήματά τους έγκεινται στο ότι μπορεί να γίνει end-to-end εκπαίδευση και χρησιμοποιούν λιγότερες παραμέτρους, άρα είναι εφικτό να χρησιμοποιηθούν σε όλα τα επίπεδα του κωδικοποιητή.

Τα μπλοκ LSTM χρησιμοποιούν τα χωρικά χαρακτηριστικά του αντίστοιχου επιπέδου του κωδικοποιητή. Οπότε με την αφαίρεση των μπλοκ LSTM που βρίσκονται στα πρώτα επίπεδα του κωδικοποιητή, χάνεται η ακριβέστερη χωρική ανάλυση, αφού σε κάθε επίπεδο μειώνεται μέσω των επιπέδων μέγιστης συγκέντρωσης (maxpooling). Από την άλλη πλευρά, όταν αφαιρούνται μπλοκ LSTM από τα τελευταία επίπεδα, δεν αξιοποιούνται τα πιο σύνθετα χωρικά χαρακτηριστικά που εξάγονται από αυτά τα επίπεδα.

Όσον αφορά την εκπαίδευση, χρησιμοποιήθηκε το cross entropy loss, όπου  $n$  είναι ο αριθμός των κατηγοριών (στην προκειμένη περίπτωση, είναι 2: αλλαγή/μη αλλαγή),  $y_{s,l}$  είναι ένας δυαδικός δείκτης που δείχνει αν η κατηγορία  $l$  είναι η ορθή για την παρατήρηση  $s$  και το  $P_{ch(s,l)}$  αντιπροσωπεύει την πιθανότητα η παρατήρηση  $s$  να ανήκει στην κατηγορία  $l$ .

$$Loss_{ch} = - \sum_{l=0}^n y_{s,l} \log(p_{ch(s,l)})$$

### 3.2.2 Αρχιτεκτονική με διεσταλμένα φίλτρα

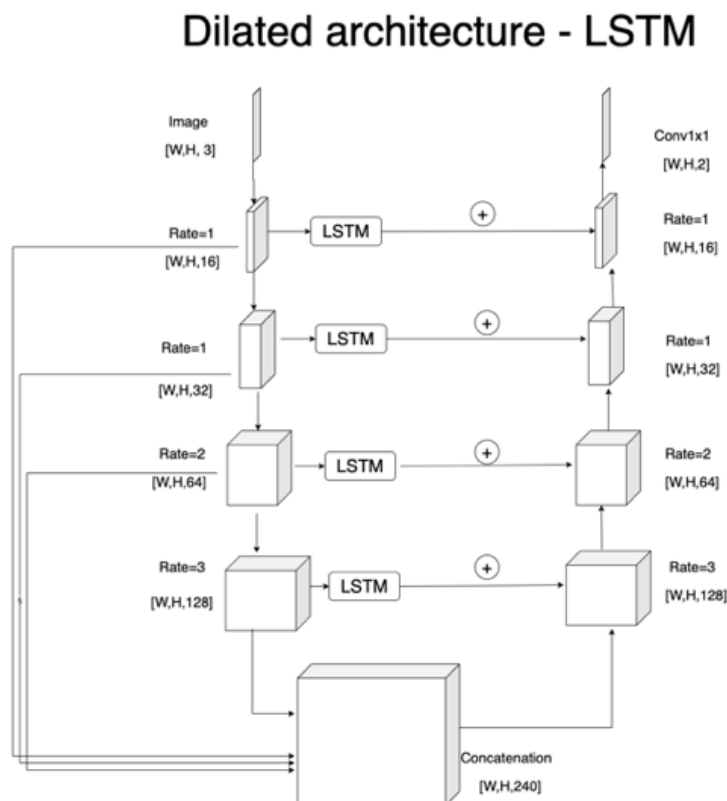
Η αρχιτεκτονική που υλοποιήθηκε βασίστηκε στην αρχιτεκτονική που παρουσιάστηκε από τους Vakaloroulou et al. (2019), όπου είχε εφαρμοστεί για το αντικείμενο της αντιστοίχισης εικόνων (registration).

Ειδικότερα, βασίζεται στην αρχιτεκτονική κωδικοποιητή-αποκωδικοποιητή. Ο κωδικοποιητής έχει 4 συνελκτικά επίπεδα με φίλτρο μεγέθους 3x3, τα 2 πρώτα με ρυθμό διαστολής=1 και τα επόμενα με ρυθμό 2 και 3 αντίστοιχα και καθένα ακολουθείται από ένα επίπεδο

ενεργοποίησης LeakyReLU. Οι διαστάσεις παραμένουν ίδιες κατά πλάτος και ύψος, ενώ ο αριθμός των χαρτών χαρακτηριστικών διπλασιάζεται σε κάθε επίπεδο. Πιο συγκεκριμένα, στο πρώτο επίπεδο παράγονται 16 χάρτες χαρακτηριστικών, στο δεύτερο 32, στο τρίτο επίπεδο 64, ενώ στο τέταρτο επίπεδο καταλήγουν να παράγονται 128 χάρτες χαρακτηριστικών. Έπειτα οι χάρτες χαρακτηριστικών από κάθε επίπεδο του κωδικοποιητή συνενώνονται πριν από τον αποκωδικοποιητή ώστε να δημιουργηθεί ένας χώρος χαρακτηριστικών με περισσότερη πληροφορία και πολλές αναλύσεις.

Ο αποκωδικοποιητής έχει 4 συνελκτικά επίπεδα με φίλτρο μεγέθους 3x3, το πρώτο με ρυθμό διαστολής=3, το δεύτερο με ρυθμό 2 και τα επόμενα με ρυθμό 1 και καθένα ακολουθείται από ένα επίπεδο ενεργοποίησης LeakyReLU. Οι διαστάσεις παραμένουν ίδιες κατά πλάτος και ύψος, ενώ ο αριθμός των χαρτών χαρακτηριστικών υποδιπλασιάζεται σε κάθε επίπεδο. Σε κάθε επίπεδο του αποκωδικοποιητή προστίθεται η έξοδος από το LSTM που είχε εφαρμοστεί στο αντίστοιχο επίπεδο του κωδικοποιητή. Στο τέλος εφαρμόζεται ένα συνελκτικό επίπεδο με μέγεθος φίλτρου 1x1 για να υπολογιστούν οι τελικές πιθανότητες.

Όσον αφορά τα μπλοκ LSTM, χρησιμοποιήθηκαν και εδώ τα συνελκτικά LSTM, όπως στο προηγούμενο δίκτυο L-UNet. Επίσης, και στην εκπαίδευση χρησιμοποιήθηκε και πάλι το cross entropy loss.



Εικόνα 3.6. Προτεινόμενη αρχιτεκτονική με διεσταλμένα συνελκτικά φίλτρα.

### 3.3 Λεπτομέρειες υλοποίησης

Για τις εικόνες του OSCD, ως προεπεξεργασία έγινε κανονικοποίηση του ιστογράμματος των εικόνων ανά κανάλι και έπειτα αντιστοίχιση του ιστογράμματος κάθε καναλιού στο

ιστόγραμμα της εικόνας του πρώτου καναλιού. Ακόμη, έγινε προσαρμογή του μεγέθους της εικόνας κάθε καναλιού στο μέγεθος του αληθούς χάρτη μεταβολών. Στις εικόνες του LEVIR-CD δεν εφαρμόστηκε κάποια τεχνική προεπεξεργασίας.

Για την διαδικασία της εκπαίδευσης, χρησιμοποιήθηκαν τμήματα εικόνων (“patches”) μεγέθους 32x32 από τις εικόνες με την χρήση βήματος μεγέθους 19 εικονοψηφίδων. Μια τακτική επαύξησης δεδομένων που χρησιμοποιήθηκε βάσει των Papadomanolaki et al. (2019) είναι η περιστροφή των τμημάτων εικόνων κατά 90,180 και 270 μοίρες εάν το ποσοστό των εικονοψηφίδων τους που ανήκουν στην κατηγορία «αλλαγή» είναι μεγαλύτερο του 5%. Τέλος, κάθε κατηγορία συσχετίζεται με ένα βάρος που είναι αντιστρόφως ανάλογο των εικονοψηφίδων που περιλαμβάνει σε σχέση με το συνολικό αριθμό εικονοψηφίδων. Αυτή η διαδικασία ακολουθήθηκε για την εκπαίδευση και των δύο σετ εικόνων, OSCD και LEVIR-CD.

Συνολικά προέκυψαν 12775 τμήματα εικόνων (“patches”) από τις 14 εικόνες εκπαίδευσης του OSCD, από τα οποία το  $\frac{1}{5}$  χρησιμοποιήθηκαν για επικύρωση. Από τις 100 εικόνες εκπαίδευσης του LEVIR-CD δημιουργήθηκαν 189180 τμήματα εικόνων και από τις 15 εικόνες επικύρωσης δημιουργήθηκαν 20047 τμήματα εικόνων. Από αυτά τα δεδομένα μπορεί να διαπιστωθεί η διαφορά μεγέθους μεταξύ των δύο σετ εικόνων.

Για το σετ εικόνων OSCD, χρειάστηκαν 15-30 εποχές εκπαίδευσης για τα υδριβικά μοντέλα και το L-UNet και 40 εποχές για το μοντέλο Dilated+LSTM. Για το σετ εικόνων LEVIR-CD, χρειάστηκαν 15-30 εποχές εκπαίδευσης για τα δύο μοντέλα.

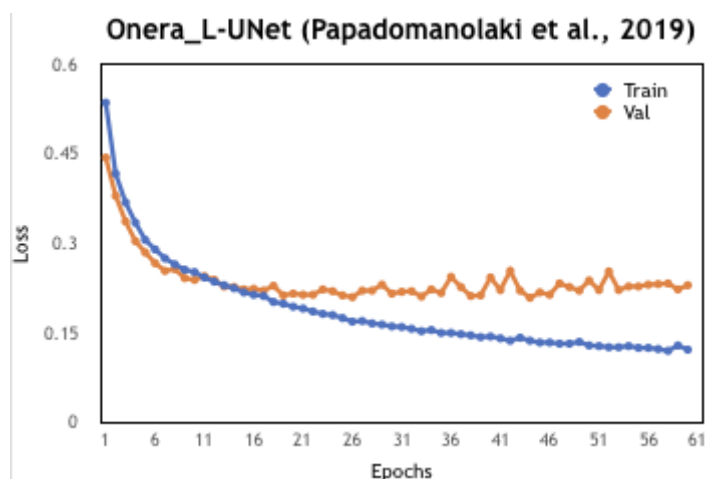
Όσον αφορά τις υπερπαραμέτρους της εκπαίδευσης, ως βελτιστοποιητής επιλέχθηκε ο Adam, το μέγεθος του batch ήταν 64 και ο ρυθμός εκπαίδευσης 0.0001. Χρησιμοποιήθηκε η βιβλιοθήκη PyTorch (<https://pytorch.org/>) και μια κάρτα γραφικών GeForce GTX 1080 Ti.

## 3.4 Τρόποι Αξιολόγησης των Αποτελεσμάτων

Είναι απαραίτητο τα μοντέλα να αξιολογούνται σε κάθε στάδιο της εφαρμογής τους, δηλαδή κατά την εκπαίδευσή τους μέσω του σφάλματος εκπαίδευσης και επικύρωσης, όπως και κατά την εφαρμογή τους στο σετ ελέγχου μέσω διάφορων μετρικών που προκύπτουν από τον πίνακα σύγχυσης, αλλά και από το τελικό αποτέλεσμα που παράγουν, τον χάρτη μεταβολών.

### 3.4.1 Διαγράμματα σφάλματος εκπαίδευσης και επικύρωσης

Για την εύρεση των βέλτιστων παραμέτρων του δικτύου κατά την διάρκεια της εκπαίδευσης (training) και της επικύρωσης (validation), κατασκευάζονται διαγράμματα του σφάλματος σε κάθε εποχή (Εικ. 3.7). Το διάγραμμα του σφάλματος εκπαίδευσης δείχνει πόσο καλά εκπαιδεύεται το μοντέλο, ενώ από την επικύρωση παρουσιάζεται η ικανότητα του μοντέλου για γενίκευση. Επιλέγεται να χρησιμοποιηθούν οι παράμετροι από εκείνη την εποχή που εμφανίζεται το μικρότερο σφάλμα επικύρωσης.



Εικόνα 3.7. Τυπικό διάγραμμα σφάλματος εκπαίδευσης και επικύρωσης συναρτήσει του αριθμού των εποχών, για το μοντέλο L-UNet (Papadomanolaki et al., 2019) στο σετ εικόνων OSCD.

### 3.4.2 Πίνακας σύγχυσης-Μετρικές για τα δεδομένα ελέγχου

Για την αξιολόγηση της απόδοσης των αλγορίθμων, είναι σημαντικό να οριστούν ποσοτικά μέτρα αξιολόγησης (Daudt, 2020). Για την δυαδική περίπτωση ταξινόμησης, έστω ότι  $Y_n \sim Y$  είναι η n-ιοστή ετικέτα στο σετ εικόνων (είτε στο υποσύνολο εκπαίδευσης είτε στο υποσύνολο ελέγχου) και η  $\hat{Y}_n \sim \hat{Y}$  η n-ιοστή ετικέτα που προβλέπει ο αλγόριθμος υπό εξέταση, όπου  $Y_n, \hat{Y}_n \in \{0,1\}$ . Οι τέσσερις πιθανοί συνδυασμοί δίνονται στον πίνακα που ακολουθεί (Πίν. 3.1), που αποτελεί τον πίνακα σύγχυσης (confusion matrix).

	$Y_n=0$	$Y_n=1$
$\hat{Y}_n=0$	Αληθώς αρνητικό (True Negative-TN)	Ψευδώς Αρνητικό (False Negative-FN)
$\hat{Y}_n=1$	Ψευδώς θετικό (False Positive-FP)	Αληθώς θετικό (True Positive-TP)

Πίνακας 3.1. Πίνακας σύγχυσης για 2 κατηγορίες.

Ως αληθώς θετική ορίζεται η εικονοψηφίδα που απεικονίζει αλλαγή και έχει προβλεφθεί ως αλλαγή, ενώ ψευδώς θετική θεωρείται όταν έχει προβλεφθεί ως αλλαγή ενώ στην πραγματικότητα δεν είναι. Ως αληθώς αρνητική ορίζεται μια εικονοψηφίδα που δεν αντιστοιχεί σε αλλαγή και ούτε έχει προβλεφθεί από το μοντέλο ως αλλαγή. Τέλος, ψευδώς αρνητική θεωρείται όταν έχει προβλεφθεί από το μοντέλο ως μη αλλαγή, ενώ στην πραγματικότητα είναι αλλαγή.

Χρησιμοποιώντας τα στοιχεία του πίνακα προκύπτουν διάφορες ενδιαφέρουσες μετρικές. Η ακρίβεια (accuracy) ή συνολική ακρίβεια (overall accuracy) αφορά το ποσοστό των σωστά ταξινομημένων εικονοψηφίσεων στο σύνολο των εικονοψηφίδων. Μπορεί να πάρει τιμές από 0 έως και 1, με το 1 να είναι η καλύτερη πιθανή τιμή.

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

Το Precision μετράει πόσο συχνά οι θετικές ταξινομήσεις είναι σωστές. Υψηλό precision σημαίνει ότι μπορούμε να έχουμε εμπιστοσύνη στο ότι όσα προβλέφθηκαν από τον

αλγόριθμο ως “θετικά” είναι πραγματικά, αλλά δεν παρέχει καμία πληροφορία σχετικά με το πόσα θετικά δεν προβλέφθηκαν από τον αλγόριθμο. Επίσης μπορεί να πάρει τιμές από 0 έως και 1, με το 1 να είναι η καλύτερη πιθανή τιμή.

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

Το Recall μετράει πόσα από τα θετικά παραδείγματα (δηλαδή με τιμή 1) ανιχνεύτηκαν σωστά από τον αλγόριθμο. Υψηλές τιμές αυτής της μετρικής σημαίνουν ότι μπορούμε να είμαστε ασφαλείς ότι έχουν ανιχνευτεί τα περισσότερα θετικά παραδείγματα, αλλά δεν παρέχει πληροφορία αναφορικά με το αν έχουν ανιχνευτεί πολλά ψευδώς θετικά. Και αυτή η μετρική μπορεί να πάρει τιμές από 0 έως και 1, με το 1 να είναι η καλύτερη πιθανή τιμή.

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

Το F1 score (ή Dice score) εξάγει ένα ισορροπημένο αποτέλεσμα μεταξύ precision και recall, ως ο αρμονικός μέσος των δύο μετρικών. Αν μία από τις δύο μετρικές έχει χαμηλή τιμή, θα μειωθεί το F1 score, γεγονός που την καθιστά μια εύρωστη μετρική για την ποσοτικοποίηση των αλγορίθμων ταξινόμησης. Επίσης είναι καλύτερη από την μετρική της ακρίβειας (accuracy) όταν υπάρχει ανισορροπία μεταξύ των παραδειγμάτων κάθε κατηγορίας (όπως ισχύει συνήθως στο αντικείμενο της ανίχνευσης μεταβολών), καθώς η μετρική accuracy θα είναι ισχυρά προκατειλημμένη προς την κατηγορία με τα περισσότερα παραδείγματα. Το F1 score μπορεί να πάρει τιμές από 0 έως και 1, με το 1 να είναι η καλύτερη πιθανή τιμή.

$$F1score = \frac{2TP}{2TP + FP + FN}$$

Η μετρική specificity είναι το αντίστοιχο της recall, αλλά για τα αρνητικά παραδείγματα, δηλαδή μετράει πόσα από τα αρνητικά παραδείγματα (δηλαδή με τιμή 0) ανιχνεύτηκαν σωστά από τον αλγόριθμο.

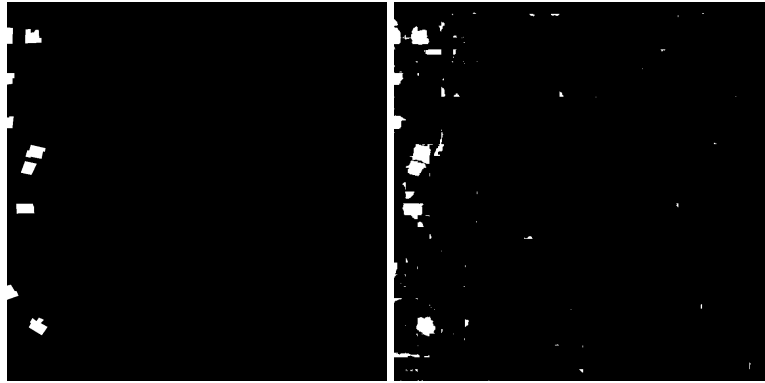
$$Specificity = \frac{TN}{FN + TP}$$

Τέλος, η εξισορροπημένη ακρίβεια είναι ο μέσος όρος των μετρικών recall και specificity, που είναι ιδιαίτερα χρήσιμη στην περίπτωση που οι κατηγορίες είναι ανομοιόμορφες ως προς το μέγεθός τους.

$$BalancedAccuracy = \frac{Recall + Specificity}{2}$$

### 3.4.3 Ποιοτική αξιολόγηση των εικόνων

Η ποιοτική αξιολόγηση των εικόνων εφαρμόζεται με οπτική σύγκριση με τους διαθέσιμους αληθείς (ground truth) χάρτες μεταβολών (Εικ. 3.8).



Εικόνα 3.8. Τυπικός αληθής χάρτης μεταβολών (αριστερά) και παραγόμενος χάρτης μεταβολών (δεξιά).

Ακόμη, μπορεί να γίνει καλύτερη επισκόπηση της απόδοσης των μοντέλων με την κατασκευή εικόνων διαφορών. Υπολογίζονται από την διαφορά του χάρτη μεταβολών κάθε μοντέλου με τον αληθή και απεικονίζονται με διαφορετικά χρώματα οι εικονοψηφίδες που είναι ορθώς θετικές (πράσινο), ψευδώς θετικές (κόκκινο), ορθώς αρνητικές (μαύρο), ψευδώς αρνητικές (κίτρινο) (Εικ. 3.9).



Εικόνα 3.9. Αληθής χάρτης μεταβολών, παραγόμενος χάρτης μεταβολών και εικόνα διαφορών, για το μοντέλο L-UNet, ζεύγος test111 από το σετ εικόνων LEVIR-CD (από αριστερά προς τα δεξιά). Με πράσινο χρώμα απεικονίζονται οι εικονοψηφίδες που είναι ορθώς θετικές, με μαύρο οι ορθώς αρνητικές, με κόκκινο οι ψευδώς θετικές και με κίτρινο οι ψευδώς αρνητικές.

Με αυτό τον τρόπο μπορούν να παρατηρηθούν πιο εύκολα τα σφάλματα που δημιουργούν τα μοντέλα, όπως πρόβλεψη πολλών ψευδών θετικών παραδειγμάτων ή μη ακριβής πρόβλεψη των ακμών των κτηρίων και να γίνουν συγκρίσεις για το πού υπερτερεί το καθένα.

## 4. ΠΕΙΡΑΜΑΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΚΑΙ ΑΞΙΟΛΟΓΗΣΗ

Στο παρόν κεφάλαιο περιγράφονται οι εφαρμογές που πραγματοποιήθηκαν στο πλαίσιο της εργασίας και αναλύονται τα αποτελέσματά τους.

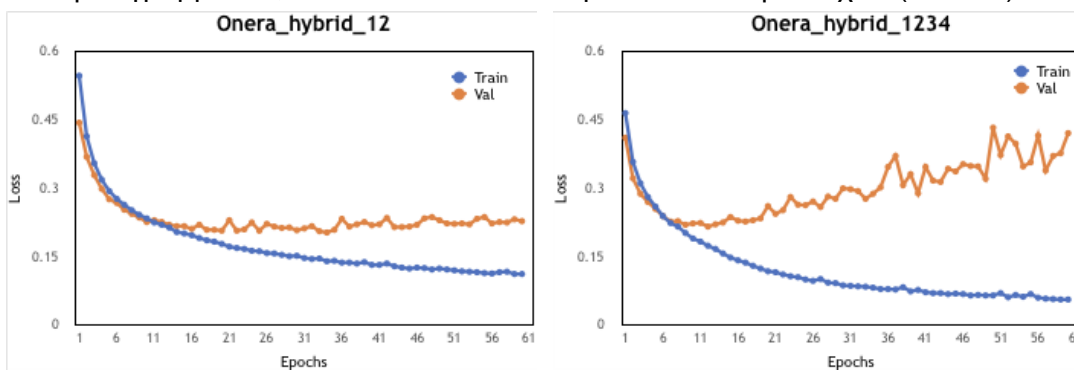
### 4.1 Σημαντικότητα LSTM στο δίκτυο

Η πρώτη εφαρμογή που διενεργήθηκε στο πλαίσιο της παρούσας διπλωματικής εργασίας αφορά το κατά πόσον χρειάζονται οι λειτουργίες LSTM του δικτύου L-UNet (Paradomanolaki et al, 2019). Οπότε συγκρίθηκε η πλήρης εκδοχή του δικτύου με κάθε πιθανή περίπτωση στην οποία λείπει από ένα μπλοκ LSTM έως και τα 4 από τα 5. Οι εικόνες που χρησιμοποιήθηκαν είναι από το OSCD (Daudt et al. 2018).

Τα μπλοκ LSTM έχουν αριθμηθεί κατ' αντιστοιχία των επιπέδων του κωδικοποιητή (encoder), δηλαδή η περίπτωση hybrid\_1 αφορά την αφαίρεση του μπλοκ LSTM του 1ου επιπέδου του κωδικοποιητή, η περίπτωση hybrid\_345 αφορά την αφαίρεση των μπλοκ του 3ου, 4ου και 5ου επιπέδου του κωδικοποιητή και ούτω καθεξής.

#### 4.1.1 Εκπαίδευση

Αρχικά επιλέχθηκε να τρέξουν τα μοντέλα για 60 εποχές. Η απόδοση της εκπαίδευσης αξιολογήθηκε βάσει του σφάλματος εκπαίδευσης και του σφάλματος αξιολόγησης με την κατασκευή διαγραμμάτων, δύο εκ των οποίων παρατίθενται στη συνέχεια (Εικ. 4.1).



Εικόνα 4.1. Διαγράμματα σφάλματος εκπαίδευσης και αξιολόγησης σε κάθε εποχή, ενδεικτικά για δύο μοντέλα (hybrid\_12 και hybrid\_1234).

Από τα διαγράμματα παρατηρήθηκε ότι το σφάλμα της εκπαίδευσης βαίνει πάντα μειούμενο για όλες τις περιπτώσεις, ενώ της αξιολόγησης αυξανόταν μετά από κάποιες εποχές. Αξίζει να τονιστεί ότι το σφάλμα αξιολόγησης αυξάνεται περισσότερο στα μοντέλα από τα οποία έχουν αφαιρεθεί LSTM από τα υψηλότερα επίπεδα ή μεγαλύτερο πλήθος μπλοκ LSTM, γεγονός που υποδεικνύει την ύπαρξη υπερπροσαρμογής (overfitting). Αντίθετα, το μοντέλο με όλα τα μπλοκ LSTM έχει την πιο ομαλή καμπύλη σφάλματος αξιολόγησης, δηλαδή είναι πιο σταθερό στην εκπαίδευσή του.

Τελικά για κάθε μοντέλο επιλέχθηκε η εποχή όπου παρουσιάζεται το μικρότερο σφάλμα επικύρωσης (Πίνακας 4.1).

Μοντέλο	hybrid_1	hybrid_12	hybrid_123	hybrid_1234	hybrid_5	hybrid_45	hybrid_345	hybrid_2345
Βέλτιστη εποχή	25	30	20	25	25	15	15	15

Πίνακας 4.1. Εποχή με το μικρότερο σφάλμα αξιολόγησης, για κάθε μοντέλο.

#### 4.1.2 Ποσοτικά αποτελέσματα στο σετ ελέγχου

Στη συνέχεια έγινε εφαρμογή του μοντέλου που αντιστοιχεί στην βέλτιστη εποχή στις εικόνες του σετ ελέγχου, παράχθηκε ο χάρτης μεταβολών και υπολογίστηκε ο πίνακας σύγχυσης για κάθε χάρτη σε σχέση με τον αληθή, αλλά και για το σύνολο των χαρτών. Τα αποτελέσματα για το σύνολο των εικόνων του σετ ελέγχου παρουσιάζονται στον Πίνακα 4.2.

Network	precision	recall	f1	specificity	Balanced accuracy
L-UNet (Papadomanolaki et al., 2019)	44.964	<b>65.443</b>	<b>53.304</b>	95.634	<b>80.539</b>
hybrid_1	39.168	<b>66.426</b>	49.279	94.378	<b>80.402</b>
hybrid_12	39.954	63.807	49.139	94.774	79.291
hybrid_123	43.049	60.815	50.412	95.615	78.215
hybrid_1234	41.75	58.379	48.684	95.561	76.97
hybrid_5	42.267	64.956	51.211	95.165	80.06
hybrid_45	<b>45.06</b>	59.638	<b>51.334</b>	<b>96.037</b>	77.837
hybrid_345	41.339	57.517	48.104	95.552	76.535
hybrid_2345	<b>45.565</b>	55.21	49.926	<b>96.405</b>	75.808

Πίνακας 4.2. Αποτελέσματα μετρικών πάνω στις εικόνες ελέγχου του OSCD για το δίκτυο L-UNet (Papadomanolaki et al, 2019) και όλες τις παραλλαγές του χωρίς LSTM.

Παρατηρείται ότι η εξισοροπημένη ακρίβεια μειώνεται όσο αφαιρούνται περισσότερα μπλοκ LSTM, επειδή μειώνεται η αξιοποίηση της χρονικής πληροφορίας από το δίκτυο. Ειδικότερα, μετά από το δίκτυο με όλα τα μπλοκ LSTM (L-UNet), καλύτερη απόδοση έχουν τα δίκτυα hybrid\_1 και hybrid\_5, όπου έχει αφαιρεθεί μόνο ένα μπλοκ LSTM. Επίσης, παρατηρείται ότι η ακρίβεια μειώνεται περισσότερο όταν αφαιρούνται μπλοκ LSTM από τα υψηλότερα επίπεδα (79.291 για το hybrid\_12 έναντι 77.837 για το hybrid\_45, και το ίδιο συνεχίζεται και με την αφαίρεση περισσότερων μπλοκ).

Σχετικά με την μετρική recall, που δείχνει πόσα θετικά παραδείγματα προβλέπονται, το μεγαλύτερο ποσοστό εμφανίζεται για το δίκτυο hybrid\_1 κι έπειτα ακολουθεί το L-UNet. Όσον αφορά την μετρική specificity, που υποδεικνύει πόσα αρνητικά παραδείγματα ανιχνεύει το κάθε δίκτυο, τα μεγαλύτερα ποσοστά επιτυγχάνουν τα δίκτυα hybrid\_2345 και hybrid\_45.

Για την μετρική precision, καλύτερη απόδοση έχουν δύο δίκτυα που έχουν αφαιρεθεί τα μπλοκ LSTM των υψηλότερων επιπέδων (hybrid\_2345 και hybrid\_45) και ακολουθεί το δίκτυο με όλα τα LSTM. Αντίθετα, χειρότερη απόδοση παρουσιάζουν τα δίκτυα από τα οποία έχουν αφαιρεθεί τα LSTM των χαμηλότερων επιπέδων (hybrid\_12 και hybrid\_1).

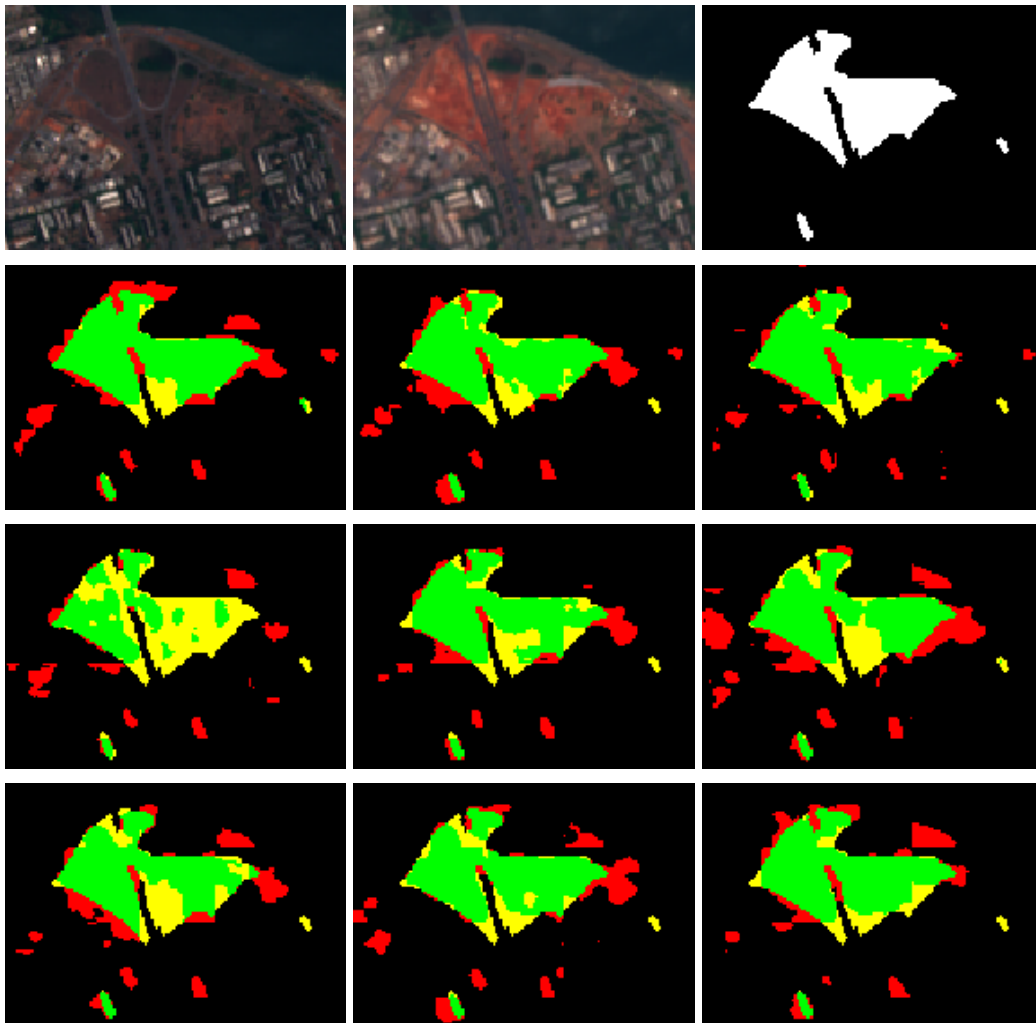


Όσον αφορά την μετρική F1, που αποτελεί τον αρμονικό μέσο των precision και recall, παρατηρείται ότι το δίκτυο L-UNet έχει την καλύτερη απόδοση με αρκετή διαφορά (2%) από τα επόμενα δίκτυα, που είναι τα hybrid\_45 και hybrid\_5.

Σε απόλυτους αριθμούς, η μετρική specificity παρουσιάζει τα μεγαλύτερα ποσοστά, δηλαδή όλα τα μοντέλα ανιχνεύουν σε μεγάλο βαθμό τα αρνητικά παραδείγματα, ενώ η μετρική precision εμφανίζει τα μικρότερα ποσοστά, που σημαίνει ότι δεν ανιχνεύονται ορθά αρκετά από τα θετικά παραδείγματα, δηλαδή οι εικονοψηφίδες που υποδεικνύουν τις αλλαγές μεταξύ των εικόνων.

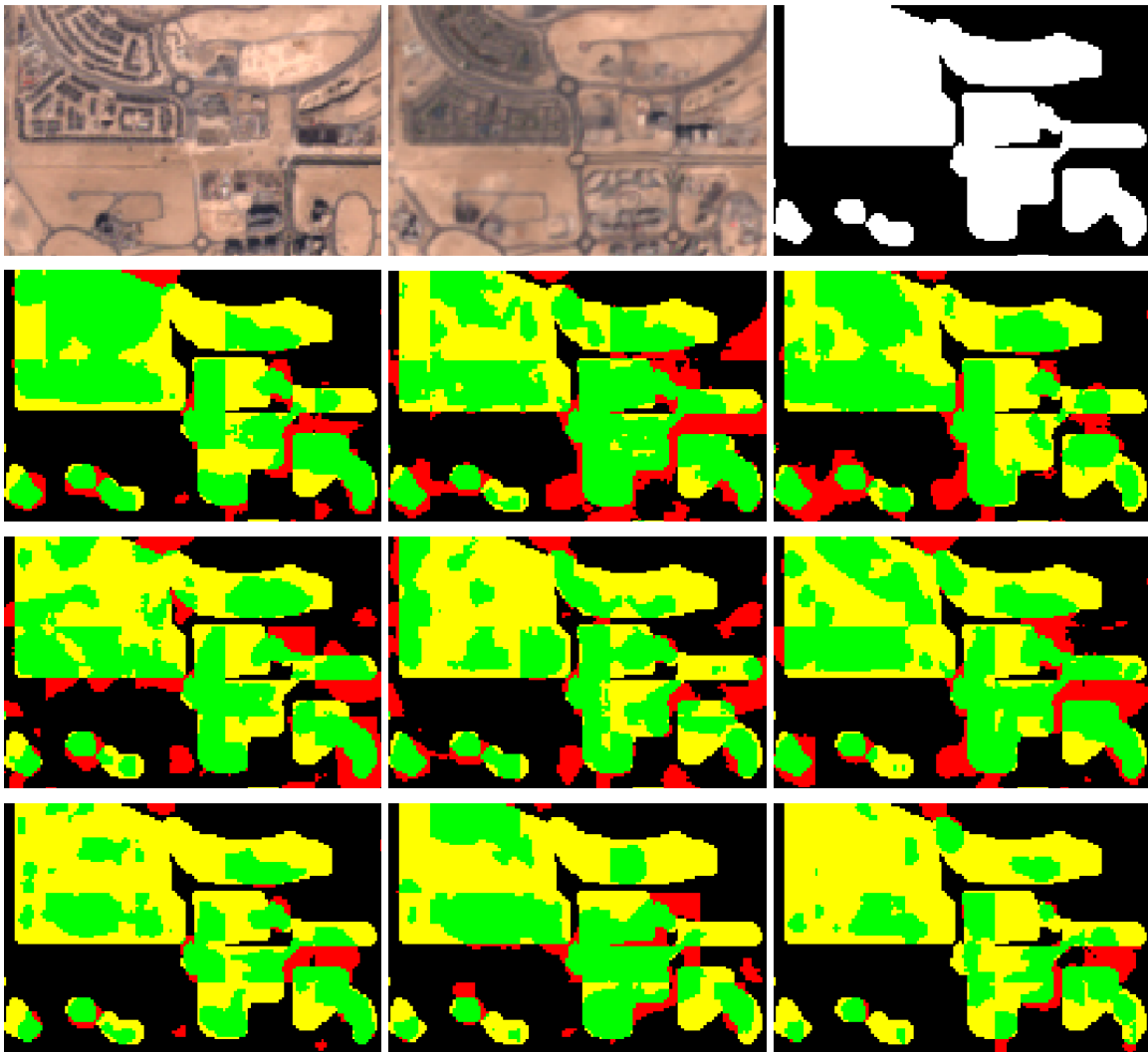
#### 4.1.3 Ποιοτικά αποτελέσματα στο σετ ελέγχου

Για την περαιτέρω μελέτη της απόδοσης των μεθόδων, κατασκευάστηκαν εικόνες διαφορών, όπου απεικονίζονται με διαφορετικά χρώματα οι εικονοψηφίδες που είναι ορθώς θετικές (πράσινο), ψευδώς θετικές (κόκκινο), ορθώς αρνητικές (μαύρο), ψευδώς αρνητικές (κίτρινο). Όπως ήταν αναμενόμενο βάσει των ποσοτικών αποτελεσμάτων, παρατηρήθηκαν πολλές περιπτώσεις όπου το μοντέλο L-UNet με όλα τα LSTM προέβλεπε τα περισσότερα αληθώς θετικά και στη συνέχεια παρουσιάζονται ενδεικτικά δύο παραδείγματα.



Εικόνα 4.2. Τμήμα από τις εικόνες του ζεύγους Brasilia, αληθής χάρτης μεταβολών, χάρτες μεταβολών από L-UNet, hybrid\_1, hybrid\_12, hybrid\_123, hybrid\_1234, hybrid\_5, hybrid\_45, hybrid\_345, hybrid\_2345 (από αριστερά προς τα δεξιά).

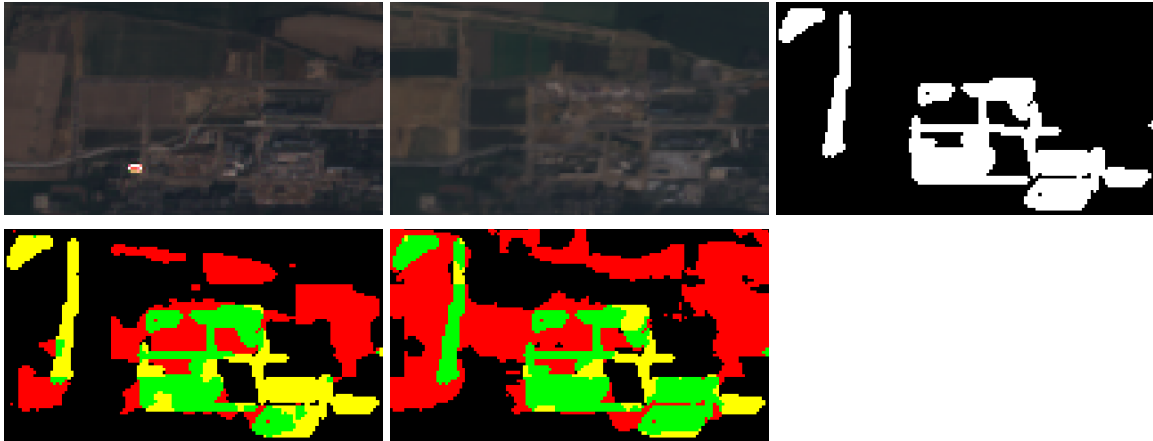
Όπως φαίνεται στην Εικ.4.2, το L-UNet ανιχνεύει πληρέστερα την αλλαγή που υπάρχει μεταξύ των εικόνων απ'ό,τι όλα τα μοντέλα με λιγότερα μπλοκ LSTM. Στη συνέχεια, καλύτερη απόδοση εμφανίζουν τα δίκτυα όπου έχει αφαιρεθεί ένα μπλοκ LSTM (hybrid\_1 και hybrid\_5), ενώ την χειρότερη απόδοση εμφανίζει το δίκτυο hybrid\_123.



Εικόνα 4.3. Τμήμα από τις εικόνες του ζεύγους Dubai, αληθής χάρτης μεταβολών, χάρτες μεταβολών από L-UNet, hybrid\_1, hybrid\_12, hybrid\_123, hybrid\_1234, hybrid\_5, hybrid\_45, hybrid\_345, hybrid\_2345 (από αριστερά προς τα δεξιά).

Από την Εικ.4.3, μπορεί να παρατηρηθεί ότι το L-UNet ανιχνεύει σε μεγαλύτερο βαθμό την αλλαγή αριστερά, ενώ το hybrid\_1 ανιχνεύει πληρέστερα την αλλαγή δεξιά. Ακόμη, το L-UNet αλλά και τα hybrid\_45 και hybrid\_2345 παρουσιάζουν τα λιγότερα ψευδώς θετικά παραδείγματα.

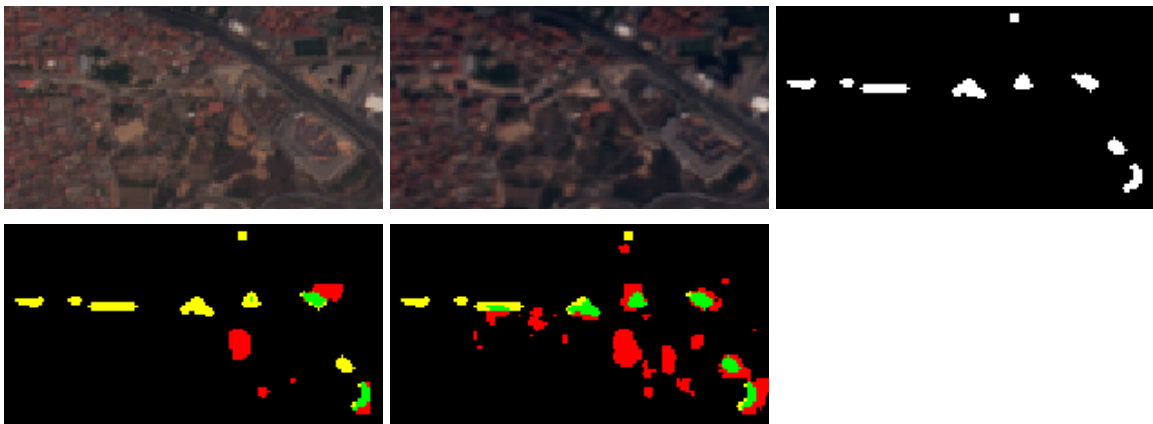
Παρατηρήθηκαν όμως και περιπτώσεις όπου τα μοντέλα χωρίς τα μπλοκ LSTM του δεύτερου, ή και του τρίτου και τέταρτου επιπέδου μπόρεσαν και ανίχνευσαν αλλαγές που δεν έγιναν αντιληπτές από το μοντέλο με όλα τα μπλοκ LSTM. Αυτό ίσως συνέβη επειδή τα μοντέλα με τα λιγότερα LSTM σύγκριναν μόνο την υψηλότερου επιπέδου πληροφορία και έτσι ανίχνευσαν την μεταβολή. Παρακάτω παρουσιάζονται δύο χαρακτηριστικές περιπτώσεις.



Εικόνα 4.4. Τμήμα από τις εικόνες του ζεύγους Saclay\_w, αληθής χάρτης μεταβολών, εικόνες διαφορών από L-UNet και hybrid\_12 (από αριστερά προς τα δεξιά)

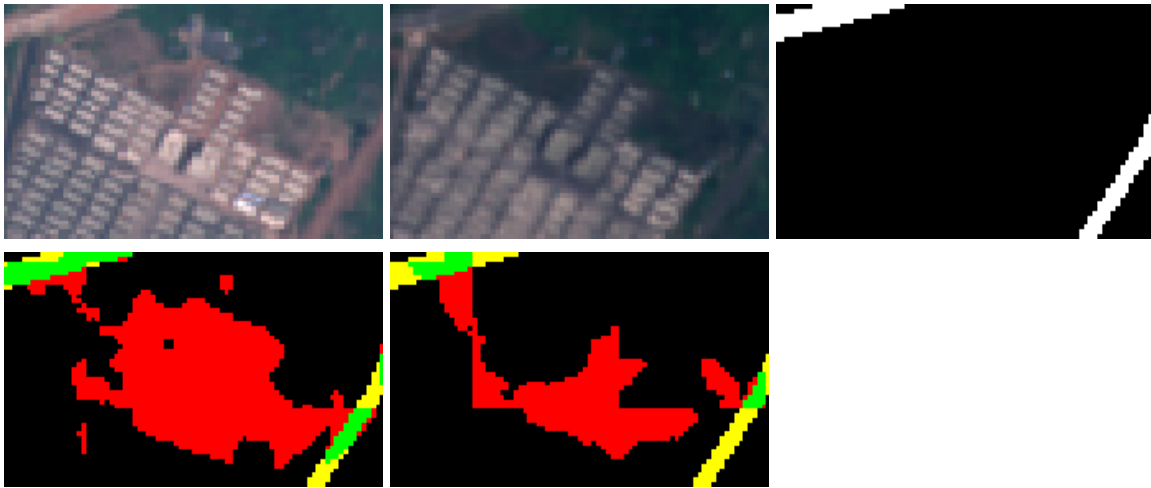
Στην Εικ. 4.4 παρατηρείται ότι το μοντέλο hybrid\_12 ανίχνευσε περισσότερα αληθώς θετικά παραδείγματα σε σχέση με το L-UNet, όμως ανιχνεύει και αρκετές ψευδώς θετικές εικονοψηφίδες, δηλαδή διαπιστώνει ότι υπάρχει αλλαγή στην δόμηση, ενώ στην πραγματικότητα υπάρχει αλλαγή στην βλάστηση ή μόνο σε φωτεινότητα μεταξύ των δύο εικόνων.

Στην Εικ. 4.5 φαίνεται ότι οι μικρές περιοχές αλλαγών ανιχνεύονται καλύτερα από ένα μοντέλο hybrid (συγκεκριμένα, το hybrid\_12), απ'ό,τι από το L-UNet. Όμως και σε αυτή την περίπτωση, το μοντέλο hybrid προβλέπει περισσότερα ψευδώς θετικά παραδείγματα.



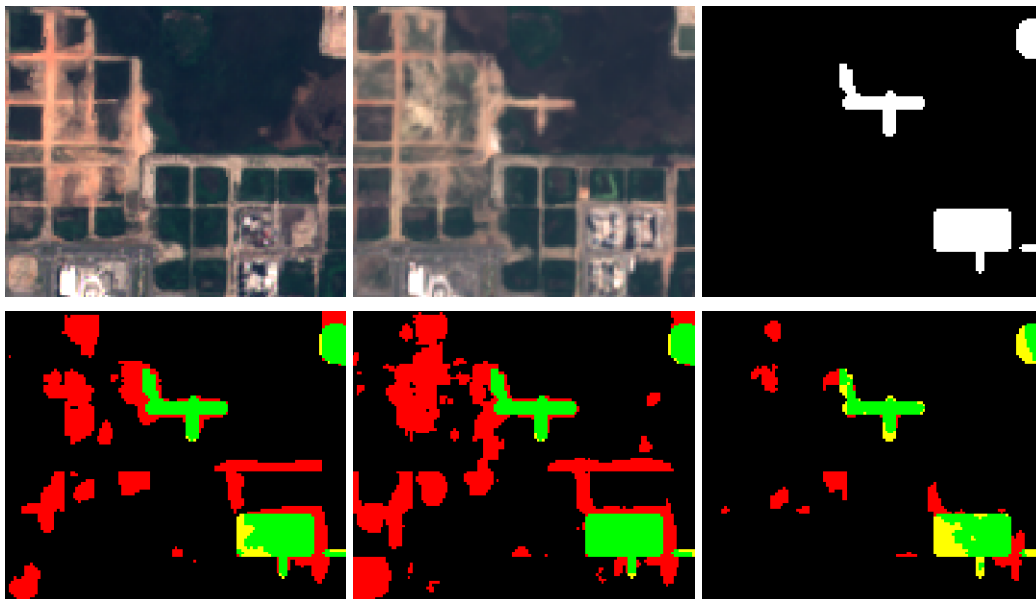
Εικόνα 4.5. Τμήμα από τις εικόνες του ζεύγους Valencia, αληθής χάρτης μεταβολών, εικόνες διαφορών από L-UNet και hybrid\_12 (από αριστερά προς τα δεξιά)

Στην Εικ. 4.6 παρουσιάζεται μια περίπτωση, όπου όλα τα μοντέλα έκαναν λανθασμένες προβλέψεις, παράγοντας ψευδώς θετικά, λόγω αλλαγής της φωτεινότητας μεταξύ των δύο εικόνων. Σημειώνεται ότι ένα μοντέλο hybrid έχει καλύτερη απόδοση απ'ό,τι το πλήρες μοντέλο L-UNet.



Εικόνα 4.6. Τμήμα από τις εικόνες του ζεύγους Chongqing, αληθής χάρτης μεταβολών, εικόνες διαφορών από L-UNet και hybrid\_1234 (από αριστερά προς τα δεξιά)

Στην Εικ. 4.7 παρουσιάζεται μια παρόμοια περίπτωση, όπου όλα τα μοντέλα έκαναν λάθος προβλέψεις, παράγοντας ψευδώς θετικά. Η απόδοση των μοντέλων hybrid ποικίλλει, για παράδειγμα το μοντέλο hybrid\_1 έχει χειρότερη απόδοση, ενώ το μοντέλο hybrid\_1234 προβλέπει λιγότερα ψευδώς θετικά απ'ό,τι το L-UNet.



Εικόνα 4.7. Τμήμα από τις εικόνες του ζεύγους Rio, αληθής χάρτης μεταβολών, εικόνες διαφορών από L-UNet, hybrid\_1 και hybrid\_1234 (από αριστερά προς τα δεξιά)

Τέλος, στην Εικ. 4.8 φαίνεται μια περίπτωση όπου όλα τα μοντέλα αδυνατούν να προβλέψουν την αλλαγή, εκτός από το μοντέλο hybrid\_1 μπορεί να προβλέψει μερικώς την αλλαγή.



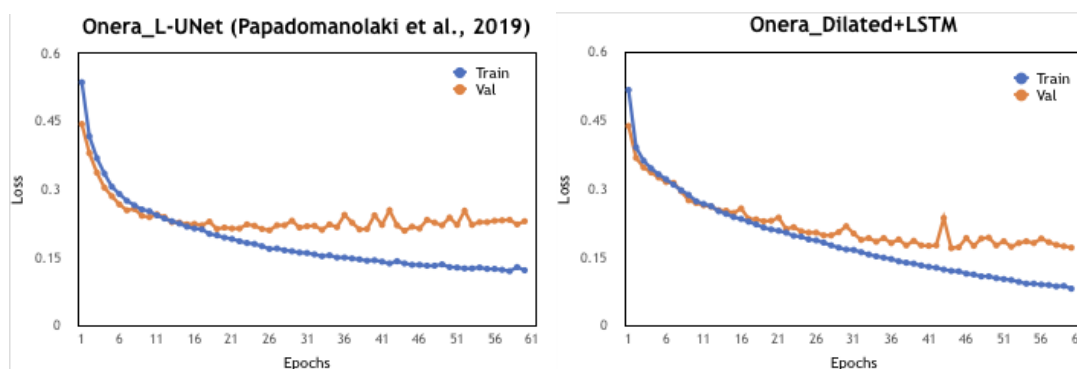
Εικόνα 4.8. Τμήμα από τις εικόνες του ζεύγους Brazilia, αληθής χάρτης μεταβολών, εικόνες διαφορών από L-UNet και hybrid\_1 (από αριστερά προς τα δεξιά)

## 4.2 Αρχιτεκτονική με διεσταλμένα (dilated) φίλτρα

Η δεύτερη εφαρμογή αφορά την υλοποίηση του δικτύου με τα διεσταλμένα φίλτρα από τους Vakaforoulou et al. (2019), του οποίου η αρχιτεκτονική αναλύθηκε στο προηγούμενο κεφάλαιο. Διενεργήθηκε σύγκριση της απόδοσής του, σε σχέση με το δίκτυο L-UNet (Papadomanolaki et al, 2019) στο σετ εικόνων OSCD.

### 4.2.1 Εκπαίδευση

Αρχικά επιλέχθηκε να τρέξουν τα μοντέλα για 60 εποχές. Η απόδοση της εκπαίδευσης αξιολογήθηκε βάσει του σφάλματος εκπαίδευσης και του σφάλματος αξιολόγησης (Εικ. 4.9).



Εικόνα 4.9. Διαγράμματα σφάλματος εκπαίδευσης και αξιολόγησης σε κάθε εποχή, για το δίκτυο L-UNet (αριστερά) και για το δίκτυο Dilated+LSTM (δεξιά) για τις εικόνες OSCD.

Όπως φαίνεται από τα διαγράμματα, το σφάλμα της εκπαίδευσης βαίνει πάντα μειούμενο, ενώ της αξιολόγησης μετά από κάποιες εποχές ήταν σχετικά σταθερό, με λίγες αυξομειώσεις. Παρατηρείται ότι το μοντέλο Dilated+LSTM φτάνει σε μικρότερο σφάλμα, τόσο στο σετ εκπαίδευσης όσο και στο σετ επικύρωσης.

Τελικά για κάθε μοντέλο επιλέχθηκε η εποχή όπου παρουσιάζεται το μικρότερο σφάλμα επικύρωσης, δηλαδή η εποχή 25 για το L-UNet και η εποχή 40 για το Dilated+LSTM.

### 4.2.2 Ποσοτικά αποτελέσματα στο σετ ελέγχου

Στη συνέχεια έγινε εφαρμογή του μοντέλου που αντιστοιχεί στην βέλτιστη εποχή στις εικόνες του σετ ελέγχου, παράχθηκε ο χάρτης μεταβολών και υπολογίστηκε ο πίνακας σύγχυσης για κάθε χάρτη σε σχέση με τον αληθή, αλλά και για το σύνολο των χαρτών, όπως και στην προηγούμενη εφαρμογή. Τα αποτελέσματα για το σύνολο των εικόνων του σετ ελέγχου παρουσιάζονται στον Πίνακα 4.3.

Network	precision	recall	f1	specificity	Balanced accuracy
L-UNet (Papadomanolaki et al., 2019)	44.964	65.443	53.304	95.634	80.539
Dilated+LSTM	38.925	64.66	48.596	94.471	79.565

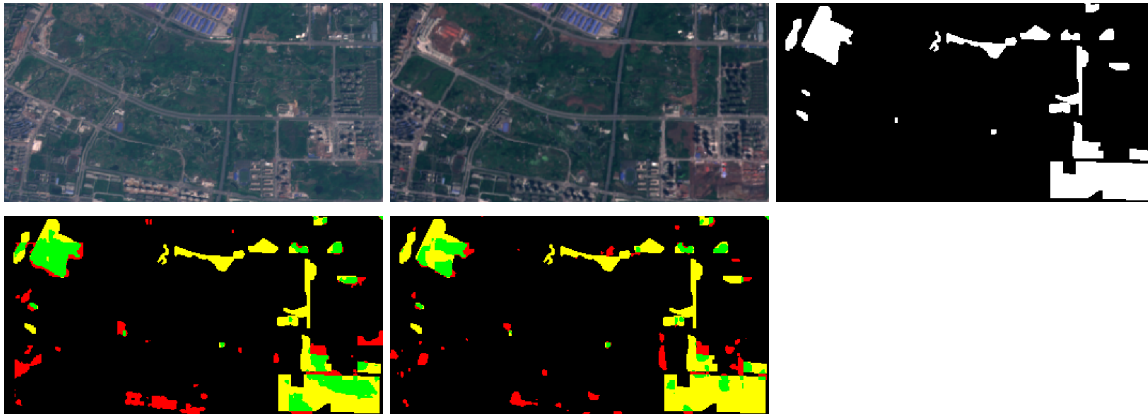
Πίνακας 4.3. Αποτελέσματα μετρικών πάνω στις εικόνες ελέγχου του OSCD για το δίκτυο L-UNet (Papadomanolaki et al, 2019) και το Dilated+LSTM.

Παρατηρείται ότι το δίκτυο L-UNet είναι ελαφρώς καλύτερο από το Dilated+LSTM στις μετρικές recall, specificity και balanced accuracy. Ακόμη, το L-UNet είναι αρκετά καλύτερο στις μετρικές precision και f1-score (που είναι παράγωγο μέγεθος του precision). Κατά συνέπεια συμπεραίνεται ότι το Dilated+LSTM προβλέπει λιγότερα ορθώς θετικά παραδείγματα, δηλαδή αλλαγές στην δόμηση, αν και πρέπει να σημειωθεί ότι έχει το  $\frac{1}{3}$  των παραμέτρων απ'όσες έχει το L-UNet.

#### 4.2.3 Ποιοτικά αποτελέσματα στο σετ ελέγχου

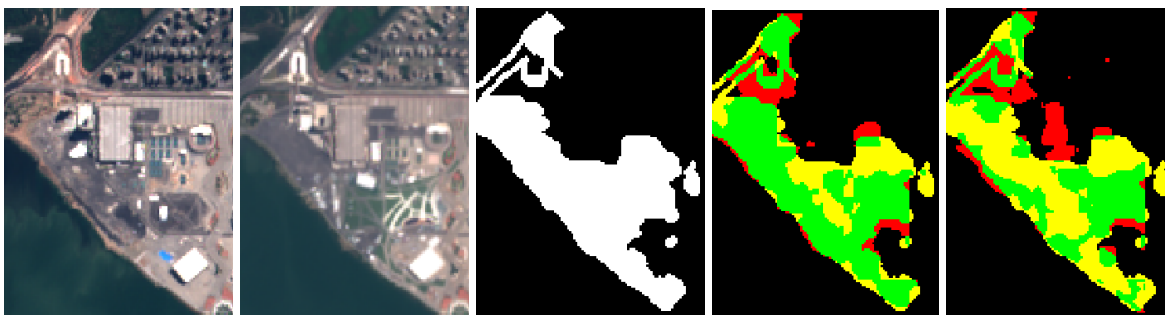
Για να γίνει μια πιο διεξοδική σύγκριση των δύο μοντέλων, κατασκευάστηκαν και σε αυτή την περίπτωση εικόνες διαφορών. Γενικώς παρατηρείται ότι οι χάρτες αλλαγών από την μέθοδο Dilated+LSTM έχουν περισσότερα ψευδώς θετικά παραδείγματα. Στη συνέχεια παρουσιάζονται δύο χαρακτηριστικές περιπτώσεις όπου υπερτερεί το μοντέλο L-UNet και δύο περιπτώσεις όπου το Dilated+LSTM έχει καλύτερη απόδοση.

Στην Εικ.4.10 παρατηρείται ότι το μοντέλο L-UNet ανιχνεύει περισσότερο τα κτήρια που έχουν υποστεί αλλαγή, αν και όχι πλήρως.



Εικόνα 4.10. Τμήμα από τις εικόνες του ζεύγους Chongqing, αληθής χάρτης μεταβολών, χάρτης μεταβολών από L-UNet και από Dilated+LSTM (από αριστερά προς τα δεξιά).

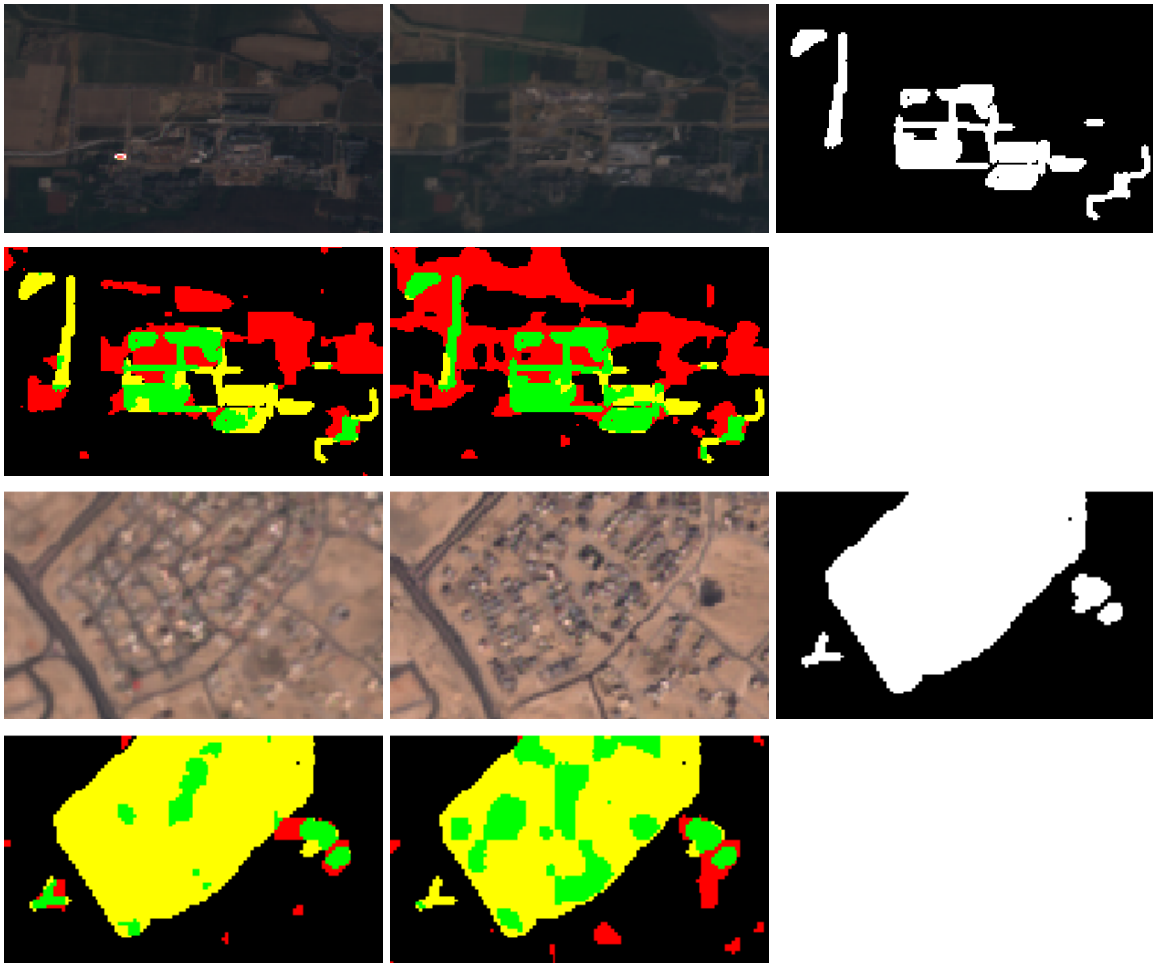
Στην Εικ. 4.11 παρατηρείται ότι το μοντέλο L-UNet ανίχνευσε καλύτερα, αλλά όχι πλήρως, τις αλλαγές που πραγματοποιήθηκαν στο λιμάνι. Επίσης, το μοντέλο Dilated+LSTM εμφανίζει περισσότερα ψευδώς θετικά παραδείγματα.



Εικόνα 4.11. Τμήμα από τις εικόνες του ζεύγους Rio, αληθής χάρτης μεταβολών, χάρτης μεταβολών από L-UNet και από Dilated+LSTM (από αριστερά προς τα δεξιά).

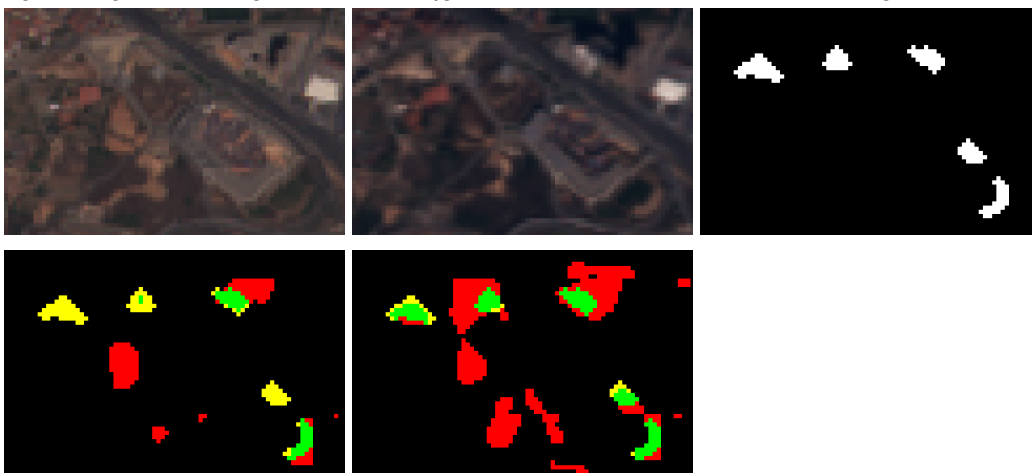
Στην Εικ. 4.12 μπορεί να παρατηρηθεί ότι το μοντέλο Dilated+LSTM έχει ανιχνεύσει επιτυχώς περισσότερες αλλαγές από το μοντέλο L-UNet, αν και προβλέπει και αρκετά ψευδώς θετικά.





Εικόνα 4.12. Τμήμα από τις εικόνες του ζεύγους Saclay\_w (πάνω) και Dubai (κάτω), αληθής χάρτης μεταβολών, χάρτης μεταβολών από L-UNet και από Dilated+LSTM (από αριστερά προς τα δεξιά).

Στην Εικ.4.13 μπορεί να παρατηρηθεί ότι το μοντέλο Dilated+LSTM ανίχνευσε επιτυχώς όλες τις μικρές μεταβολές, αν και επίσης προέβλεψε περισσότερα ψευδώς θετικά.



Εικόνα 4.13. Τμήμα από τις εικόνες του ζεύγους Valencia, αληθής χάρτης μεταβολών, χάρτης μεταβολών από L-UNet και από Dilated+LSTM (από αριστερά προς τα δεξιά).

Τέλος, στην Εικ. 4.14 παρουσιάζεται μια περίπτωση όπου το δίκτυο L-UNet προβλέπει πλήρως τις αλλαγές σε σχέση με το Dilated+LSTM, όμως επίσης προβλέπει περισσότερα ψευδώς θετικά.



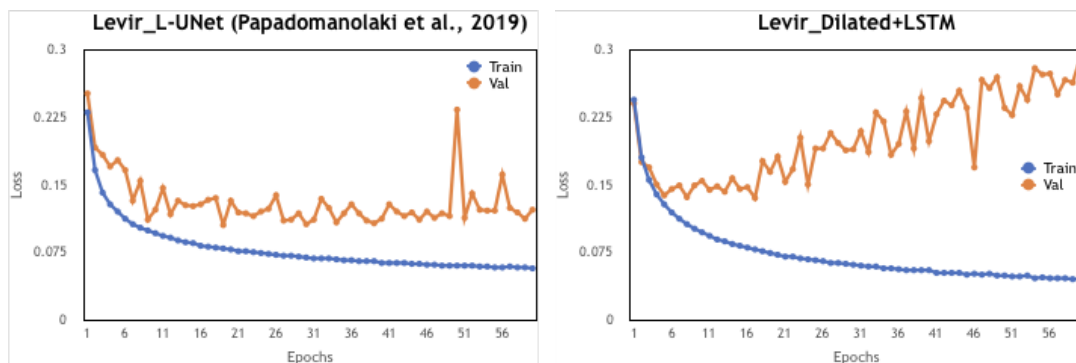
Εικόνα 4.14. Τμήμα από τις εικόνες του ζεύγους Norcia, αληθής χάρτης μεταβολών, χάρτης μεταβολών από L-UNet και από Dilated+LSTM (από αριστερά προς τα δεξιά).

### 4.3 Εφαρμογή στο σετ εικόνων LEVIR-CD

Η τρίτη υλοποίηση αφορά την εφαρμογή των δικτύων L-UNet (Papadomanolaki et al, 2019) και Dilated+LSTM στο σετ εικόνων LEVIR-CD (Chen & Shi, 2020).

#### 4.3.1 Εκπαίδευση

Για την εκπαίδευση χρησιμοποιήθηκε ένα υποσύνολο των διαθέσιμων εικόνων, λόγω περιορισμού στη μνήμη. Ειδικότερα, έγινε χρήση 100 εικόνων από το σετ εκπαίδευσης και 15 εικόνων από το σετ επικύρωσης και επιλέχθηκε να τρέξουν τα μοντέλα για 60 εποχές. Η απόδοση της εκπαίδευσης αξιολογήθηκε όπως και προηγουμένως βάσει του σφάλματος εκπαίδευσης και του σφάλματος αξιολόγησης. Στη συνέχεια παρατίθενται τα αντίστοιχα διαγράμματα (Εικ. 4.15).



Εικόνα 4.15. Διαγράμματα σφάλματος εκπαίδευσης και αξιολόγησης σε κάθε εποχή, για το δίκτυο L-UNet (αριστερά) και για το δίκτυο Dilated+LSTM (δεξιά) για τις εικόνες LEVIR-CD.

Για το δίκτυο L-UNet, το σφάλμα της εκπαίδευσης βαίνει πάντα μειούμενο, ενώ της αξιολόγησης έχει κάποιες αυξομειώσεις. Για το δίκτυο Dilated+LSTM το σφάλμα εκπαίδευσης είναι μικρότερο από το σφάλμα του L-UNet, όπως και στο προηγούμενο σετ εικόνων που εξετάστηκε. Ακόμη, το σφάλμα επικύρωσης του Dilated+LSTM αυξάνεται μετά από κάποιες εποχές, οπότε υπάρχει υπερπροσαρμογή στα δεδομένα (overfitting). Αυτό μπορεί να οφείλεται στο ότι χρειάζεται παραπάνω δεδομένα ή λόγω του ότι χρησιμοποιείται υποσύνολο του σετ εικόνων επικύρωσης.



Τελικά για κάθε μοντέλο επιλέχθηκε η εποχή όπου παρουσιάζεται το μικρότερο σφάλμα επικύρωσης, δηλαδή η εποχή 30 για το L-UNet και η εποχή 15 για το Dilated+LSTM.

### 4.3.2 Ποσοτικά αποτελέσματα στο σετ ελέγχου

Ακολούθως εφαρμόστηκαν τα μοντέλα για τις εικόνες του σετ ελέγχου, με τις παραμέτρους που αντιστοιχούν στην βέλτιστη εποχή. Παράχθηκαν χάρτες μεταβολών και υπολογίστηκε ο πίνακας σύγχυσης για κάθε χάρτη σε σχέση με τον αληθή, αλλά και για το σύνολο των χαρτών. Τα αποτελέσματα για το σύνολο των εικόνων του σετ ελέγχου παρουσιάζονται στον Πίνακα 4.4.

Network	precision	recall	f1	specificity	Balanced accuracy
L-UNet (Papadomanolaki et al., 2019)	49.662	94.862	65.194	94.839	94.851
Dilated+LSTM	45.601	92.483	61.083	94.078	93.280

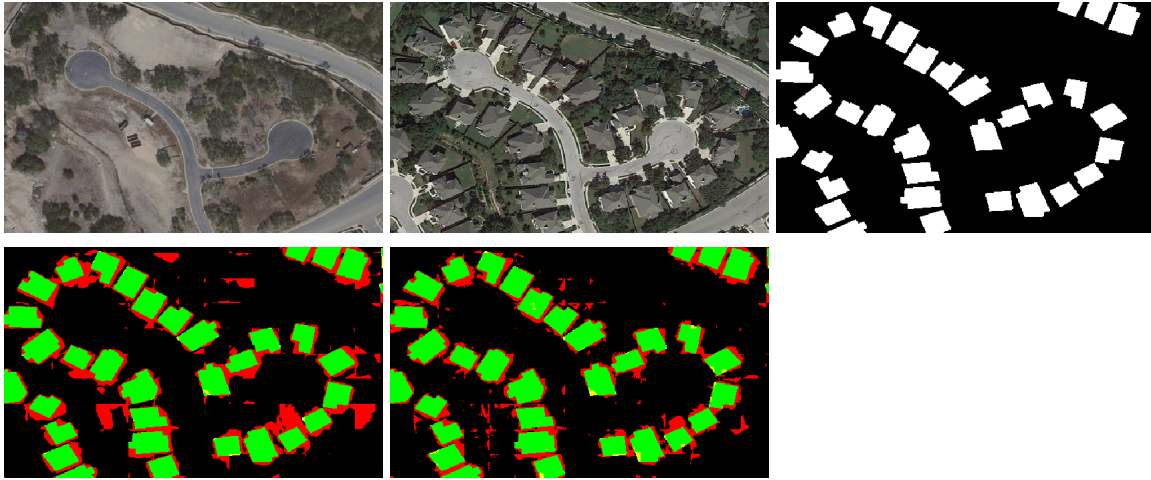
Πίνακας 4.4. Αποτελέσματα μετρικών πάνω στις εικόνες ελέγχου του LEVIR-CD για το δίκτυο L-UNet (Papadomanolaki et al, 2019) και το Dilated+LSTM.

Παρατηρείται ότι το δίκτυο L-UNet είναι ελαφρώς καλύτερο από το Dilated+LSTM στις μετρικές recall, specificity και balanced accuracy. Ακόμη, είναι αρκετά καλύτερο στις μετρικές precision και f1-score (που είναι παράγωγο μέγεθος του precision), προκύπτουν δηλαδή τα ίδια συμπεράσματα με την εφαρμογή των μοντέλων στο σετ εικόνων OSCD. Επίσης, διαπιστώνεται ότι και τα δύο μοντέλα έχουν αυξημένο recall σε σχέση με την εφαρμογή τους στο σετ OSCD, γεγονός που υποδεικνύει ότι προβλέπουν περισσότερες εικονοψηφίδες ως θετικές σε αυτό το σετ εικόνων.

### 4.3.3 Ποιοτικά αποτελέσματα στο σετ ελέγχου

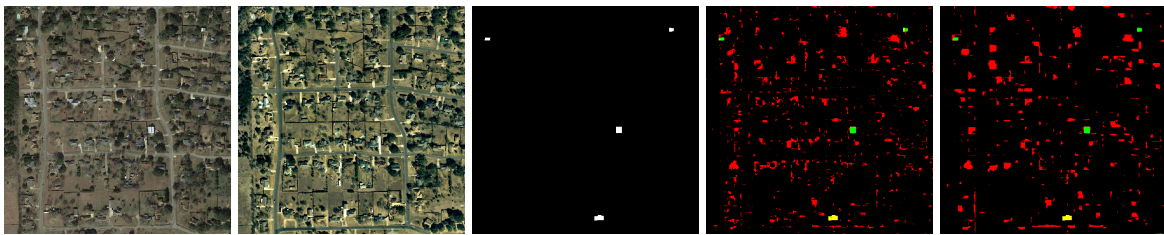
Για να γίνει εμβάθυνση στην απόδοση των μεθόδων, κατασκευάστηκαν και για την τρίτη εφαρμογή εικόνες διαφορών για τις εικόνες του σετ ελέγχου. Από τις 128 εικόνες, ως γενικότερο συμπέρασμα μπορεί να ειπωθεί ότι οι παραγόμενοι χάρτες μεταβολών από τα δύο μοντέλα έχουν αρκετό θόρυβο, δηλαδή παράγουν πολλά ψευδώς θετικά παραδείγματα ή με άλλα λόγια προβλέπουν αλλαγές στη δόμηση που δεν υφίστανται.

Ένα φαινόμενο που παρατηρήθηκε συχνά είναι ότι ανιχνεύονται ψευδώς θετικές εικονοψηφίδες γύρω από τις ακμές των κτηρίων που αποτελούν αλλαγές, δηλαδή δεν ανιχνεύεται σωστά το σχήμα των κτηρίων και από τα δύο μοντέλα (Εικ.4.16).



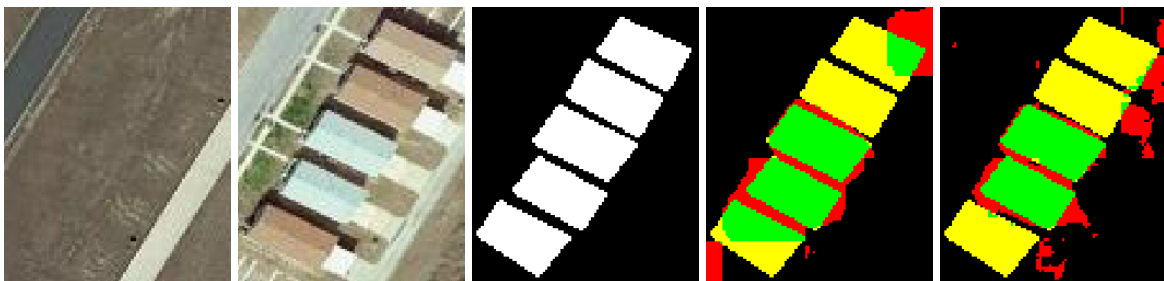
Εικόνα 4.16. Ζεύγος εικόνων test10, αληθής χάρτης μεταβολών, χάρτες μεταβολών από τα μοντέλα L-UNet και Dilated+LSTM (από αριστερά προς τα δεξιά)

Ένας λόγος για την ύπαρξη ψευδώς θετικών είναι οι μεγάλες αλλαγές που παρατηρούνται στην ραδιομετρία του ζεύγους των εικόνων και δεν μπόρεσαν να μοντελοποιηθούν από τα δίκτυα (Εικ.4.17).



Εικόνα 4.17. Ζεύγος εικόνων test107, αληθής χάρτης μεταβολών, χάρτες μεταβολών από τα μοντέλα L-UNet και Dilated+LSTM (από αριστερά προς τα δεξιά).

Στην αντίθετη περίπτωση, όταν η φωτεινότητα των εικονοψηφίδων που απεικονίζουν την αλλαγή είναι παρόμοια με την φωτεινότητα των εικονοψηφίδων της περιοχής πριν από την αλλαγή, τα μοντέλα δυσκολεύονται να αναγνωρίσουν την αλλαγή και παράγουν ψευδώς αρνητικές προβλέψεις. Όπως φαίνεται και στην Εικ. 4.18, τα μοντέλα ανιχνεύουν καλύτερα την αλλαγή των κτηρίων με γκριζα σκεπή, που απέχει πολύ χρωματικά από το χρώμα που υπήρχε στην πρώτη εικόνα, ενώ αντίθετα αποδίδουν χειρότερα όταν το χρώμα της σκεπής είναι παρόμοιο με το χρώμα.



Εικόνα 4.18. Ζεύγος εικόνων test83, αληθής χάρτης μεταβολών, χάρτες μεταβολών από τα μοντέλα L-UNet και Dilated+LSTM (από αριστερά προς τα δεξιά).

Επιπλέον, οι ψευδώς θετικές εικονοψηφίδες μπορεί να προκύπτουν λόγω διαφορετικής γωνίας λήψης των εικόνων, ιδιαίτερα σε αντικείμενα με μεγάλο ύψος, όπως δεξαμενές (Εικ. 4.19).



Εικόνα 4.19. Ζεύγος εικόνων test10, αληθής χάρτης μεταβολών, χάρτες μεταβολών από τα μοντέλα L-UNet και Dilated+LSTM (από αριστερά προς τα δεξιά).

Επιπροσθέτως, παρατηρήθηκε ότι έχει γίνει επισήμανση μόνο των κτηρίων που έχουν υποστεί αλλαγή, ενώ τα μοντέλα μπορεί να προβλέπουν αλλαγές σε δρόμους, κάτι που κάνει σε μεγαλύτερο βαθμό το μοντέλο Dilated+LSTM (Εικ. 4.20).



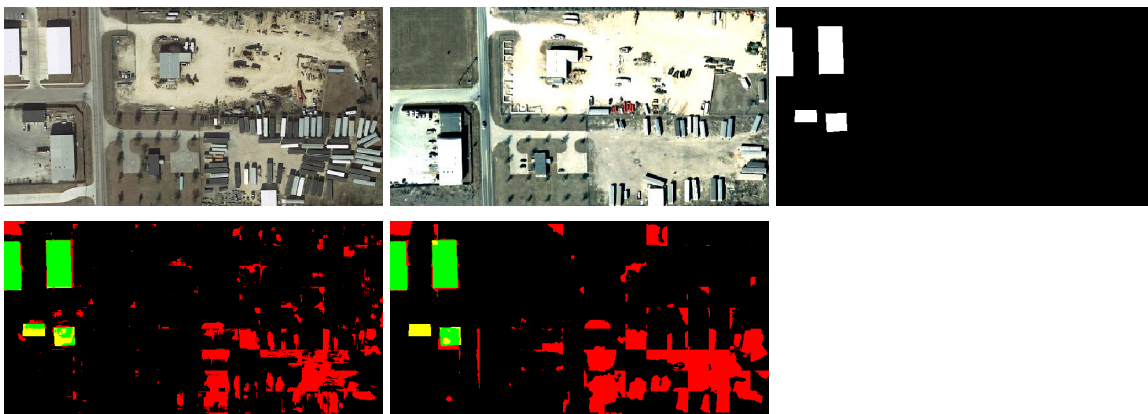
Εικόνα 4.20. Ζεύγος εικόνων test104, αληθής χάρτης μεταβολών, χάρτες μεταβολών από τα μοντέλα L-UNet και Dilated+LSTM (από αριστερά προς τα δεξιά).

Ακολούθως, διαπιστώθηκε ότι τα μοντέλα ανιχνεύουν ψευδώς ως αλλαγές την κατασκευή χώρων στάθμευσης (Εικ. 4.21).



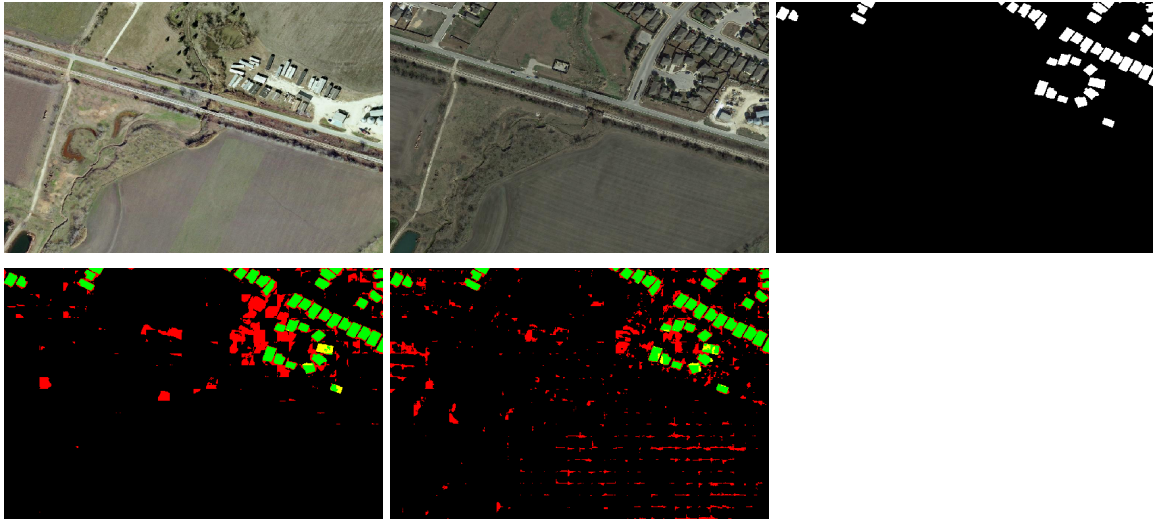
Εικόνα 4.21. Ζεύγος εικόνων test101, αληθής χάρτης μεταβολών, χάρτες μεταβολών από τα μοντέλα L-UNet και Dilated+LSTM (από αριστερά προς τα δεξιά).

Επίσης, παρατηρήθηκε ότι τα μοντέλα ανιχνεύουν ψευδώς ως αλλαγές την αλλαγή θέσης των κοντίνερς (Εικ. 4.22).



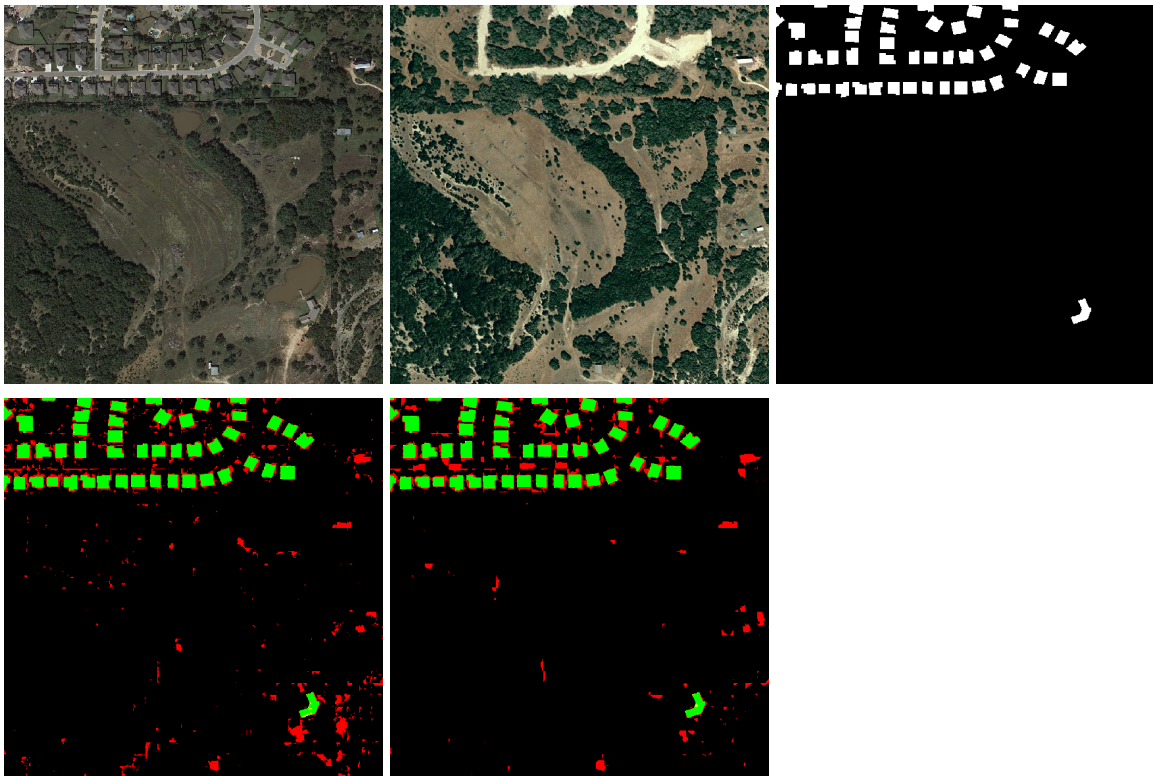
Εικόνα 4.22. Ζεύγος εικόνων test116, αληθής χάρτης μεταβολών, χάρτες μεταβολών από τα μοντέλα L-UNet και Dilated+LSTM (από αριστερά προς τα δεξιά).

Ακόμη, παρατηρήθηκε ότι το Dilated+LSTM σε γενικές γραμμές είχε περισσότερα ψευδώς θετικά παραδείγματα, για παράδειγμα σε περιοχές με αλλαγές σε βλάστηση (Εικ.4.23).



Εικόνα 4.23. Ζεύγος εικόνων test72, αληθής χάρτης μεταβολών, χάρτες μεταβολών από τα μοντέλα L-UNet και Dilated+LSTM (από αριστερά προς τα δεξιά).

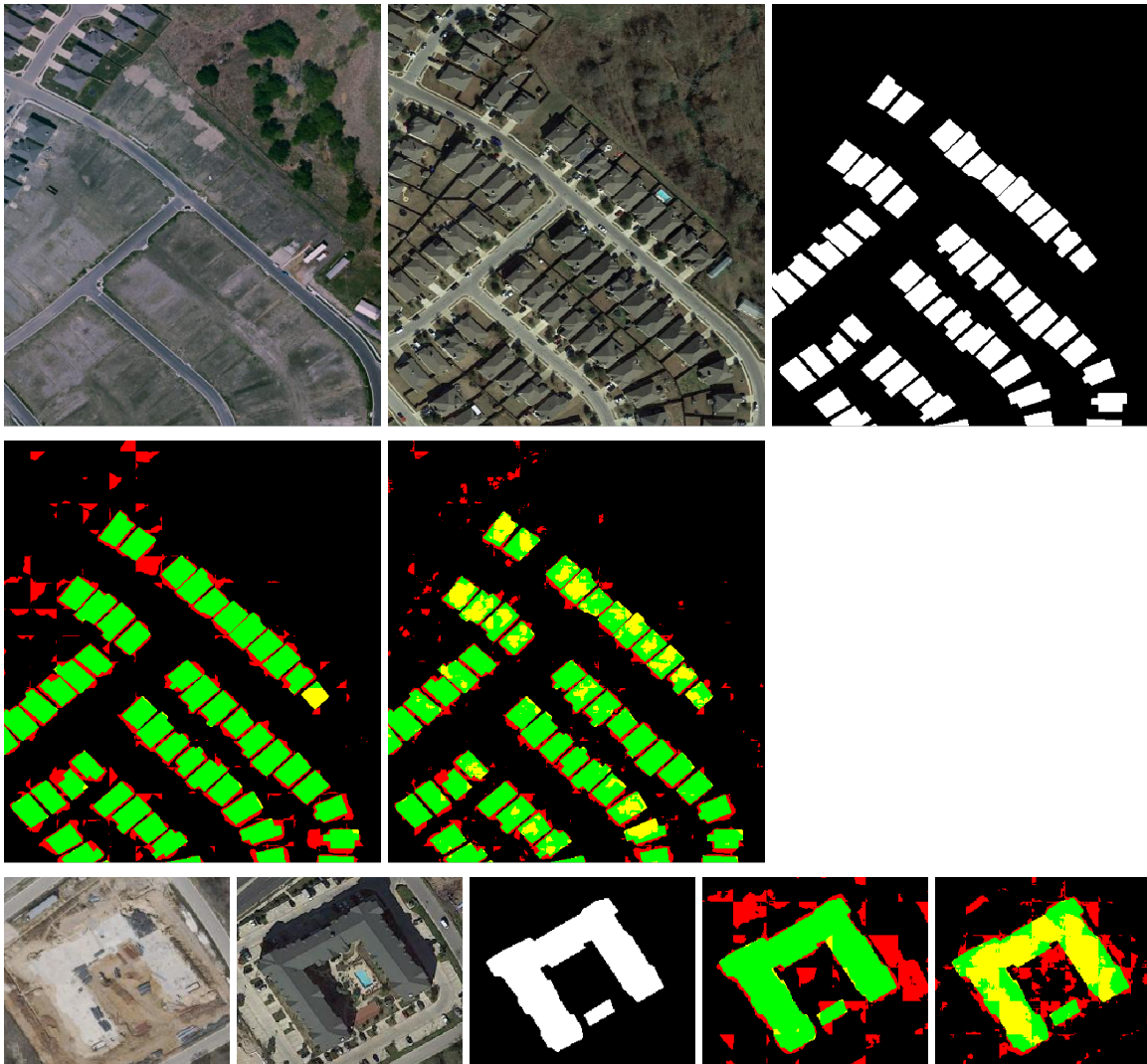
Θα πρέπει όμως να σημειωθεί ότι υπήρχαν περιπτώσεις με μεγάλη ραδιομετρική αλλαγή σε περιοχές με βλάστηση, οι οποίες δεν ανιχνεύτηκαν, ορθώς, από τα μοντέλα ως μεταβολές (Εικ. 4.24).



Εικόνα 4.24. Ζεύγος εικόνων test2, αληθής χάρτης μεταβολών, χάρτες μεταβολών από τα μοντέλα L-UNet και Dilated+LSTM (από αριστερά προς τα δεξιά).

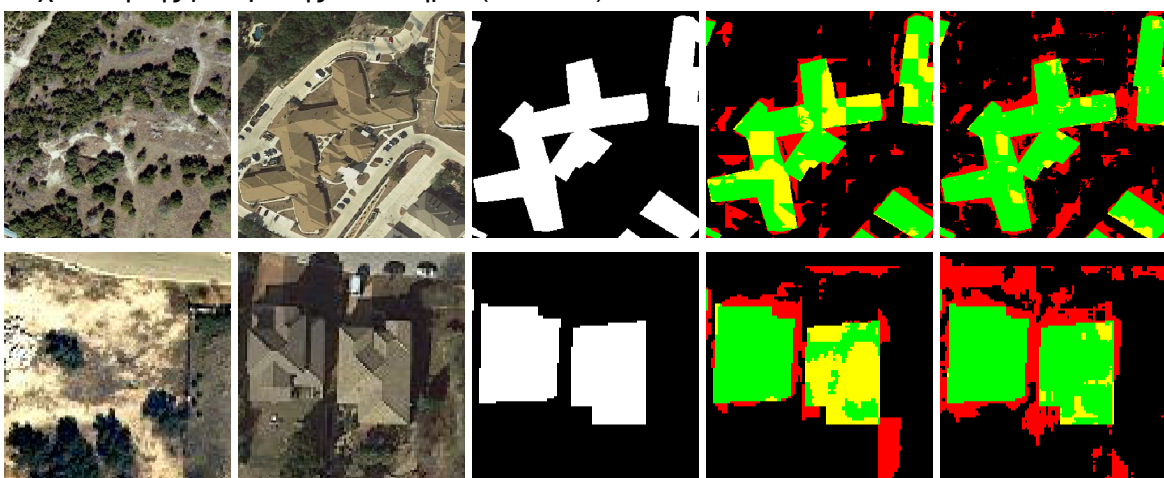
Σχετικά με την απόδοση των μοντέλων βάσει των εικόνων διαφορών, διαπιστώθηκε ότι το L-UNet μπορούσε να προβλέπει περισσότερα και πληρέστερα κτήρια που μεταβλήθηκαν (Εικ.4.25).





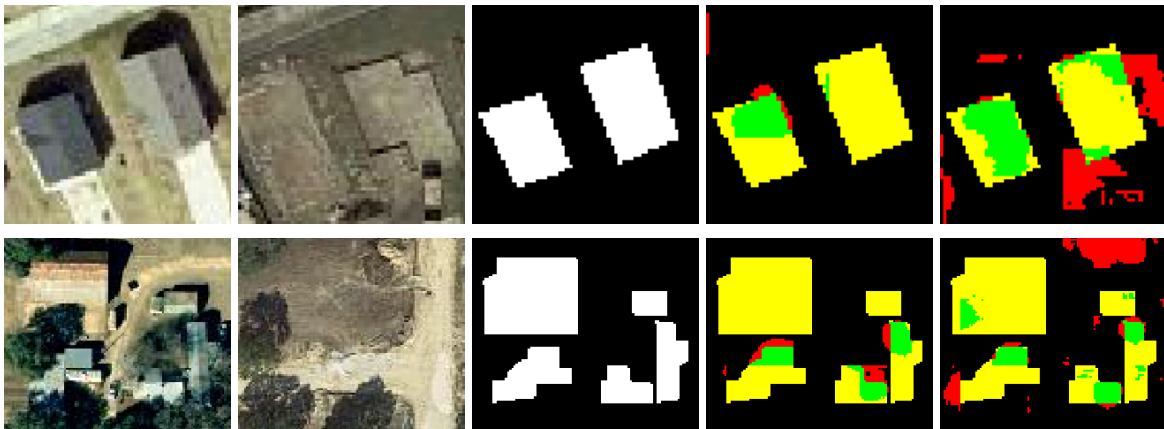
Εικόνα 4.25. Ζεύγος εικόνων (test55 πάνω, test102 κάτω), αληθής χάρτης μεταβολών, χάρτες μεταβολών από τα μοντέλα L-UNet και Dilated+LSTM (από αριστερά προς τα δεξιά).

Στη συνέχεια παρουσιάζονται δύο περιπτώσεις όπου το Dilated+LSTM ήταν καλύτερο στην ανίχνευση της μεταβολής στα κτήρια (Εικ.4.26).



Εικόνα 4.26. Ζεύγος εικόνων (test21 πάνω, test34 κάτω), αληθής χάρτης μεταβολών, χάρτες μεταβολών από τα μοντέλα L-UNet και Dilated+LSTM (από αριστερά προς τα δεξιά).

Επιπλέον, παρατηρείται ότι σε περιπτώσεις που η αλλαγή ήταν η κατεδάφιση και όχι η κατασκευή ενός νέου κτηρίου, τα μοντέλα δυσκολεύονταν περισσότερο να καταλάβουν την αλλαγή, με το Dilated+LSTM να έχει καλύτερη απόδοση (Εικ. 4.27).



Εικόνα 4.27. Ζεύγος εικόνων (test84 πάνω, test111 κάτω), αληθής χάρτης μεταβολών, χάρτες μεταβολών από τα μοντέλα L-UNet και Dilated+LSTM (από αριστερά προς τα δεξιά).

Ακόμη, παρατηρήθηκε μια περίπτωση όπου το μοντέλο L-UNet ανιχνεύει περισσότερα ψευδώς θετικά παραδείγματα απ'ό,τι το Dilated+LSTM (Εικ. 4.28).



Εικόνα 4.28. Ζεύγος εικόνων test45, αληθής χάρτης μεταβολών, χάρτες μεταβολών από τα μοντέλα L-UNet και Dilated+LSTM (από αριστερά προς τα δεξιά).

## 5. ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΠΡΟΟΠΤΙΚΕΣ

Στο κεφάλαιο αυτό περιγράφονται τα συμπεράσματα που προκύπτουν από τα αποτελέσματα των εφαρμογών που παρουσιάστηκαν στο Κεφάλαιο 4 και παρουσιάζονται προτάσεις για μελλοντική επέκταση της εργασίας.

### 5.1 Συμπεράσματα

Σχετικά με την πρώτη εφαρμογή, όπου αφαιρούνται μπλοκ LSTM από τα επίπεδα του κωδικοποιητή (encoder) πριν από την ενσωμάτωση της πληροφορίας στο αντίστοιχο επίπεδο του αποκωδικοποιητή (decoder), εξάγεται το συμπέρασμα ότι η εξισορροπημένη ακρίβεια του μοντέλου μειώνεται όσο αφαιρούνται περισσότερα μπλοκ LSTM. Αυτό είναι αναμενόμενο, επειδή με την αφαίρεση των LSTM μπλοκ, αξιοποιείται λιγότερο η χρονική πληροφορία και καθίσταται δυσκολότερη η ανίχνευση των μεταβολών μεταξύ των δύο εικόνων. Ακόμη, παρατηρείται ότι επιτυγχάνεται μικρότερη ακρίβεια όταν αφαιρούνται μπλοκ LSTM που είναι σε πιο υψηλό επίπεδο, εκεί δηλαδή όπου γίνεται εξαγωγή χαρακτηριστικών σε χαμηλότερη μεν ανάλυση αλλά με υψηλότερη εννοιολογική σημασία.

Ακόμη, παρατηρείται ότι στην μετρική precision τα μοντέλα hybrid\_2345 και hybrid\_45 υπερτερούν του L-UNet και έπονται τα υπόλοιπα μοντέλα. Στην μετρική F1 το L-UNet παρουσιάζει το μεγαλύτερο ποσοστό, με αρκετή διαφορά (2%) από το αμέσως επόμενο. Σε απόλυτους αριθμούς, τα μοντέλα εμφανίζουν τα μικρότερα ποσοστά για την μετρική precision.

Από τους χάρτες μεταβολών και τις εικόνες διαφορών που παράχθηκαν με την βοήθεια των αληθών χαρτών, μπορεί να παρατηρηθεί ότι γενικά το L-UNet προβλέπει πληρέστερα τις αλλαγές στις περισσότερες περιπτώσεις. Φυσικά υπάρχουν και μεμονωμένες περιπτώσεις όπου τα υδριβικά μοντέλα προέβλεψαν πληρέστερα τις αλλαγές στις εικόνες (για παράδειγμα, το hybrid\_12), αν και με περισσότερα ψευδώς θετικά παραδείγματα. Τέλος, διαπιστώθηκε ότι ορισμένα υδριβικά μοντέλα προβλέπουν λιγότερα ψευδώς θετικά απ'ό,τι η πλήρης εκδοχή του L-UNet.

Όσον αφορά την εκπαίδευση, παρατηρείται ότι τα δίκτυα στα οποία έχουν αφαιρεθεί μπλοκ LSTM από τα υψηλότερα επίπεδα ή μεγαλύτερο πλήθος μπλοκ LSTM, είναι πιο ρηχά και φτάνουν πιο γρήγορα σε κατάσταση υπερεκπαίδευσης, δηλαδή το σφάλμα της επικύρωσης αυξάνεται αντί να μειώνεται από μια εποχή και μετά. Σημειώνεται ότι τα αποτελέσματα των μετρικών του δικτύου L-UNet είναι σε αντιστοιχία με τα αντίστοιχα αποτελέσματα του δικτύου στο ίδιο σετ εικόνων, όπως περιγράφονται στη δημοσίευση των Papadomanolaki et al (2019). Καταλήγοντας, συμπεραίνεται ότι η ενσωμάτωση των μπλοκ LSTM στην U-Net αρχιτεκτονική με τα συνελκτικά επίπεδα, βοηθάει πολύ στην καλύτερη σύγκριση μεταξύ των δύο εικόνων που απεικονίζουν την ίδια περιοχή σε διαφορετική χρονική στιγμή και άρα στην ανίχνευση των μεταβολών τους.

Αναφορικά με την δεύτερη εφαρμογή, εφαρμόστηκε μια αρχιτεκτονική με διεσταλμένα συνελκτικά φίλτρα (Vakaloroulou et al., 2019), η οποία μέχρι στιγμής είχε εφαρμοστεί μόνο στο πεδίο της αντιστοίχισης (registration) εικόνων και δεν είχε εφαρμοστεί στο πεδίο της ανίχνευσης μεταβολών. Η χρήση διεσταλμένων φίλτρων βοηθά στην αύξηση του

αντιληπτικού πεδίου και άρα αυξάνεται η πληροφορία που λαμβάνεται υπόψη κατά την εξαγωγή των χαρακτηριστικών. Διαπιστώθηκε ότι το μοντέλο Dilated+LSTM παρουσιάζει μικρότερο σφάλμα στο σετ εκπαίδευσης και επικύρωσης απ'ό,τι το L-UNet. Η αρχιτεκτονική με διεσταλμένα φίλτρα είναι πιο ρηχή, πιο συγκεκριμένα έχει 30% λιγότερες παραμέτρους από το L-UNet ενώ έχουν παρόμοια απόδοση όσον αφορά την εξισορροπημένη ακρίβεια στο σετ ελέγχου. Στις μετρικές precision και f1-score υπερτερεί σε απόδοση το δίκτυο L-UNet, γεγονός που υποδεικνύει ότι το δίκτυο με τα διεσταλμένα φίλτρα δεν προβλέπει ορθά όσα θετικά παραδείγματα, δηλαδή αλλαγές στη δόμηση, μπορεί να προβλέψει το L-UNet.

Ακόμη, από τους χάρτες μεταβολών που παράχθηκαν από τις δύο μεθόδους, διαπιστώθηκε ότι οι χάρτες που παράχθηκαν από το δίκτυο L-UNet έχουν λιγότερα ψευδώς θετικά παραδείγματα. Επίσης, παρατηρήθηκε ότι το μοντέλο Dilated+LSTM προβλέπει πληρέστερα αλλαγές μικρές σε εμβαδό.

Η τρίτη εφαρμογή αφορά την χρήση ενός νέου σετ εικόνων, που απαντάται συχνά στην πρόσφατη βιβλιογραφία, και την εφαρμογή των δύο δικτύων, του L-UNet και του δικτύου με τα διεσταλμένα φίλτρα σε αυτό. Επειδή ήταν πολύ μεγαλύτερο σε όγκο από το προηγούμενο σετ εικόνων, και λόγω περιορισμών τόσο στην διαθέσιμη μνήμη αλλά και στον χρόνο, χρησιμοποιήθηκε ένα υποσύνολο των εικόνων του για εκπαίδευση. Μετά από αρκετές δοκιμές βρέθηκε το ιδανικό πλήθος εικόνων, ώστε να γίνεται επαρκής εκπαίδευση με τους διαθέσιμους υπολογιστικούς πόρους. Δυστυχώς όμως εξαιτίας της εκπαίδευσης σε υποσύνολο και όχι στο σύνολο των εικόνων, δεν μπορεί να συγκριθεί η απόδοση των μοντέλων με τα υπόλοιπα μοντέλα από την βιβλιογραφία που έχουν εφαρμοστεί σε αυτό το σετ εικόνων. Ενδεικτικά στους Chen & Shi (2020) αναφέρεται η απόδοση άλλων μεθόδων που εφαρμόστηκαν στο LEVIR-CD και παρατηρείται ότι τα μοντέλα που παρουσιάστηκαν στην παρούσα εργασία έχουν παρόμοια απόδοση όσον αφορά την μετρική recall, αλλά πολύ χαμηλότερη απόδοση όσον αφορά το precision και κατ'επέκταση το F1 score.

Στην εκπαίδευση επιτυγχάνεται μικρότερο σφάλμα για το μοντέλο Dilated+LSTM. Όσον αφορά το σφάλμα επικύρωσης, παρουσιάζεται αύξησή του μετά από κάποια εποχή, που σημαίνει ότι γίνεται υπερπροσαρμογή, δηλαδή χρειάζεται παραπάνω δεδομένα για να εκπαιδευτεί σωστά. Μια ακόμα πιθανή εκδοχή για τις αυξομειώσεις που παρουσιάζουν τα δύο μοντέλα στο σφάλμα επικύρωσης είναι ότι οφείλεται στην χρήση υποσυνόλου των εικόνων επικύρωσης.

Όπως και με το σετ εικόνων OSCD, η απόδοση του μοντέλου L-UNet στο σετ ελέγχου είναι λίγο καλύτερη από την απόδοση του δικτύου με τα διεσταλμένα φίλτρα, τόσο ποσοτικά όσο και όπως αποτυπώνεται στις εικόνες μεταβολών. Ακόμη, παρατηρείται ότι η recall σε αυτό το σετ εικόνων είναι πολύ υψηλότερη απ'ό,τι η αντίστοιχη στο σετ εικόνων OSCD, που σημαίνει ότι εδώ προβλέφθηκαν ως θετικά παραδείγματα ένα μεγάλο ποσοστό των εικονοψηφίδων, που όμως δεν σημαίνει ότι είναι ορθώς θετικά.

Από τις εικόνες διαφορών που παράχθηκαν βάσει των χαρτών μεταβολών των δύο μοντέλων, παρατηρήθηκε ότι το δίκτυο L-UNet προβλέπει πληρέστερα τα κτήρια και παράγει λιγότερα ψευδώς θετικά παραδείγματα. Σημειώθηκαν βέβαια και περιπτώσεις όπου το Dilated+LSTM έχει καλύτερη απόδοση όσον αφορά την ανίχνευση μεταβολών στη δόμηση.



Ακόμη, διαπιστώθηκε ότι το μοντέλο Dilated+LSTM ανιχνεύει περισσότερο την αλλαγή, όσον αφορά την κατεδάφιση και όχι την κατασκευή ενός κτηρίου.

## 5.2 Προτάσεις

Στο πλαίσιο της παρούσας εργασίας χρησιμοποιήθηκε ένα δίκτυο που χρησιμοποιεί διεσταλμένα συνελκτικά φίλτρα. Θα είχε ιδιαίτερο ενδιαφέρον να γίνει μια περαιτέρω διερεύνηση της αρχιτεκτονικής, για παράδειγμα να εφαρμοστούν άλλοι συνδυασμοί ρυθμού διαστολής στα διάφορα επίπεδα. Επιπλέον, θα μπορούσε να δοκιμαστεί να αυξηθεί το μέγεθος του τμήματος εικόνας (patch) που χρησιμοποιείται, το οποίο θα βοηθούσε στην αύξηση του αντιληπτικού πεδίου. Ακόμη, θα μπορούσαν να εξεταστούν πιο σύνθετες μέθοδοι υπολογισμού της loss function ή να συνδυαστεί με κάποιο άλλο αντικείμενο, για παράδειγμα με την σημασιολογική κατάτμηση (αρχιτεκτονική multi-task), όπως από τους Paradomanolaki et al (2021).

Επιπροσθέτως, από την βιβλιογραφία φαίνεται ότι τον τελευταίο καιρό έχουν προταθεί αρκετές αρχιτεκτονικές για το πρόβλημα της ανίχνευσης μεταβολών που βασίζονται σε μηχανισμούς προσοχής (attention mechanisms). Κατά συνέπεια, θα ήταν σκόπιμο να εξεταστούν τέτοιου είδους μοντέλα στο μέλλον, τόσο όσον αφορά τα πλεονεκτήματα που προσφέρουν όσο και σχετικά με τις πιθανές αδυναμίες τους.

Επιπλέον, θα ήταν πολύ σημαντική η εφαρμογή των δύο δικτύων (L-UNet και Dilated+LSTM) σε επιπλέον σετ εικόνων, τα οποία για παράδειγμα να διαθέτουν παραπάνω από 2 ημερομηνίες ανά περιοχή, να απεικονίζουν διαφορετικών ειδών περιοχές από αυτές που έχουν εξεταστεί ήδη ή να έχουν μικρότερη ανισορροπία μεταξύ των δύο κατηγοριών (αλλαγή/μη αλλαγή), για να εκτιμηθεί περαιτέρω η απόδοσή τους. Με την χρήση περισσότερων δεδομένων ή την εφαρμογή μεθόδων επαύξησης δεδομένων, θα μπορούσε να αντιμετωπιστεί και το πολύ συχνό φαινόμενο της πρόβλεψης ψευδώς θετικών παραδειγμάτων.

Τέλος, αποτελεί πρόκληση η δημιουργία μιας νέας αρχιτεκτονικής για ένα νευρωνικό δίκτυο που θα αφορά το πρόβλημα της δυαδικής ανίχνευσης μεταβολών, ώστε να μειωθούν τα προβλήματα που υπάρχουν με τα σημερινά μοντέλα όπως η δημιουργία πολλών ψευδώς θετικών παραδειγμάτων, η καλύτερη διαχείριση των αλλαγών φωτεινότητας και η καλύτερη διατήρηση του σχήματος του αντικειμένου που ανιχνεύεται ως μεταβολή. Θα μπορούσε επίσης να διερευνηθεί το αντικείμενο, να γίνει για παράδειγμα σημασιολογική ανίχνευση μεταβολών.

# ΒΙΒΛΙΟΓΡΑΦΙΑ

- Βακαλοπούλου Μ., 2017. Εντοπισμός Μεταβολών σε Δορυφορικά Δεδομένα Πολύ Υψηλής Ανάλυσης: Συνδυάζοντας Τεχνικές Αντιστοίχισης, Κατάτμησης και Ανίχνευσης Μεταβολών. Διδακτορική Διατριβή, Εθνικό Μετσόβιο Πολυτεχνείο.
- Benedek C., Sziranyi T., 2009. Change detection in optical aerial images by a multilayer conditional mixed markov model. *IEEE Transactions on Geoscience and Remote Sensing*, pages 3416–3430.
- Bruzzone L., Serpico S.B., 1997. An iterative technique for the detection of land-cover transitions in multitemporal remote-sensing images. *IEEE Trans. Geosci. Remote. Sens.*, 35, 858-867.
- Chen L.-C., Papandreou G., Kokkinos I., Murphy K., Yuille A., 2014. Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs. *CoRR. arXiv*.
- Chen L.-C., Papandreou G., Kokkinos I., Murphy K., Yuille A., 2016. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. PP. 10.1109/TPAMI.2017.2699184.
- Chen L.-C., Papandreou G., Schroff F., Adam H., 2017. Rethinking Atrous Convolution for Semantic Image Segmentation. *arXiv:1706.05587*.
- Chen L.-C., Zhu Y., Papandreou G., Schroff F., Adam H., 2018. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation.
- Chen H., Shi Z., 2020. A Spatial-Temporal Attention-Based Method and a New Dataset for Remote Sensing Image Change Detection. *Remote Sensing*. 12. 1662. 10.3390/rs12101662.
- Chen H., Qi Z., Shi Z., 2021. Remote Sensing Image Change Detection With Transformers. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1-14.
- Chollet F., 2017. Xception: Deep learning with depthwise separable convolutions. *CVPR*.
- Daudt R.C., Saux B.L., Boulch A., Gousseau Y., 2018. Urban change detection for multispectral earth observation using convolutional neural networks. *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*.
- Daudt R.C., Saux B.L., Boulch A., 2018. Fully convolutional siamese networks for change detection. *IEEE International Conference on Image Processing (ICIP)*.
- Daudt R.C., Saux B.L., Boulch A., Gousseau Y., 2018. Multitask learning for large-scale semantic change detection. *Computer Vision and Image Understanding*, vol. 187.

Daudt R.C., 2020. Convolutional neural networks for change analysis in earth observation images with noisy labels and domain shifts. Diss. Institut polytechnique de Paris.

Deng J.S., Wang K., Deng Y.H., Qi G.J., 2008. PCA-based land-use change detection and analysis using multitemporal and multisensor satellite data. *International Journal of Remote Sensing*, vol. 29, no. 16, pp. 4823–4838.

He K., Zhang X., Ren S., Sun J., 2016. Deep residual learning for image recognition. *CVPR*.

Hochreiter S., Schmidhuber J., 1997. Long Short-term Memory. *Neural computation*. 9. 1735-80. 10.1162/neco.1997.9.8.1735.

Gupta R., Goodman B., Patel N., Hosfelt R., Sajeev S., Heim E., Doshi J., Lucas K., Choset H., Gaston M.E., 2019. Creating xbd: A dataset for assessing building damage from satellite imagery. *CVPR Workshops*, pages 10–17.

Ji S., Wei S., Lu M., 2018. Fully Convolutional Networks for Multisource Building Extraction From an Open Aerial and Satellite Imagery Data Set. *IEEE Transactions on Geoscience and Remote Sensing*, 57, 574-586.

Lebedev M., Vizilter Y., Vygolov O., Knyaz V., Rubis A., 2018. Change Detection in Remote Sensing Images Using Conditional Adversarial Networks. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. XLII-2. 565-571. 10.5194/isprs-archives-XLII-2-565-2018.

Li X., Yeh A.G.O., 1998. Principal component analysis of stacked multitemporal images for the monitoring of rapid urban expansion in the Pearl River Delta. *International Journal of Remote Sensing*, vol. 19, no. 8, pp. 1501– 1518.

Liu M., Shi Q., Marinoni A., He D., Liu X., Zhang L., 2021. Super-Resolution-Based Change Detection Network With Stacked Attention Module for Images With Different Resolutions. *IEEE Transactions on Geoscience and Remote Sensing*. PP. 1-18. 10.1109/TGRS.2021.3091758.

Liu Y. , Pang C., Zhan Z., Zhang X., Yang X., 2019. Building change detection for remote sensing images using a dual task constrained deep siamese convolutional network model, *ArXiv*, vol. abs/1909.07726.

Long J., Shelhamer E., Darrell T., 2015. Fully convolutional networks for semantic segmentation. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3431-3440, doi: 10.1109/CVPR.2015.7298965.

Malila W. A., 1980. Change vector analysis: An approach for detecting forest changes with Landsat. *Machine Processing of Remotely Sensed Data Symposium*.

Mou L., Bruzzone L., Zhu X, 2019. Learning Spectral-Spatial-Temporal Features via a Recurrent Convolutional Neural Network for Change Detection in Multispectral Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 57, 924-935.

Papadomanolaki M., Verma S., Vakalopoulou M., Gupta S., Karantzalos K., 2019. Detecting Urban Changes with Recurrent Neural Networks from Multitemporal Sentinel-2 Data.

IGARSS - IEEE International Geoscience and Remote Sensing Symposium, Jul 2019, Yokohama, Japan.

Papadomanolaki M., Vakalopoulou M., Karantzas K., 2020. Urban Change Detection Based On Semantic Segmentation and Fully Convolutional LSTM Networks. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences. V-2-2020. 541-547. 10.5194/isprs-annals-V-2-2020-541-2020.

Papadomanolaki M., Vakalopoulou M., Karantzas K., 2021. A Deep Multi-Task Learning Framework Coupling Semantic Segmentation and Fully Convolutional LSTM Networks for Urban Change Detection. IEEE Transactions on Geoscience and Remote Sensing. PP. 1-18. 10.1109/TGRS.2021.3055584.

Qi H., Zhang Z., Xiao B., Hu H., Cheng B., Wei Y., Dai J., 2017. Deformable convolutional networks – coco detection and segmentation challenge 2017 entry. ICCV COCO Challenge Workshop.

Ronneberger O., Fischer P., Brox T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. LNCS. 9351. 234-241. 10.1007/978-3-319-24574-4\_28.

Sainte Fare Garnot V., Landrieu L., Giordano S., Chehata N., 2020. Satellite image time series classification with pixel-set encoders and temporal self-attention. In CVPR.

Rußwurm M., Koerner M., 2018. Multi-temporal land cover classification with sequential recurrent encoders. ISPRS International Journal of Geo-Information, vol. 7, 02.

Rußwurm M., Koerner M., 2018. Convolutional LSTMs for cloud-robust segmentation of remote sensing imagery. NeurIPS Workshop.

Rußwurm M., Koerner M., 2019. Self-attention for raw optical satellite time series classification. arXiv preprint arXiv:1910.10536.

Shi X., Chen Z., Wang H., Yeung D.-Y., Wong W.-K., Woo W.-C., 2015. Convolutional lstm network: A machine learning approach for precipitation nowcasting. In Advances in neural information processing systems, pages 802–810.

Shi W., Zhang M., Zhang R., Chen S., Zhan Z., 2020. Change Detection Based on Artificial Intelligence: State-of-the-Art and Challenges. Remote Sensing. 12. 1688. 10.3390/rs12101688.

Song L., Xia M., Jin J., Qian M., Zhang Y., 2021. SUACDNet: Attentional change detection network based on siamese U-shaped structure, International Journal of Applied Earth Observation and Geoinformation, Volume 105, 102597, ISSN 0303-2434, <https://doi.org/10.1016/j.jag.2021.102597>.

Vakalopoulou M., Karantzas K., Komodakis N., Paragios N., 2015. Simultaneous registration and change detection in multitemporal, very high resolution remote sensing data. 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 61–69.

Vakalopoulou M., Christodoulidis S., Sahasrabudhe M., Mougiakakou S., Paragios N., 2019. Image Registration of Satellite Imagery with Deep Convolutional Neural Networks. 4939-4942. 10.1109/IGARSS.2019.8898220.

Van Etten A., Lindenbaum D., Bacastow T., 2018. Spacenet: A remote sensing dataset and challenge series. IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Van Etten A., Hogan D., Manso J. M., Shermeyer J., Weir N., Lewis R., 2021. The multi-temporal urban development spacenet dataset. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6398-6407.

Verma S., Panigrahi A., Gupta S., 2021. QFabric: Multi-Task Change Detection Dataset. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.

Weismiller R., Kristof S., Scholz D., Anuta P., Momin S., 1977. Change detection in coastal zone environments. Photogrammetric Engineering and Remote Sensing 43, 1533–1539.

Yang M., Jiao L., Liu F., Hou B., Yang S., 2019. Transferred deep learning-based change detection in remote sensing images. IEEE Transactions on Geoscience and Remote Sensing, vol. 57, pp. 6960–6973.

Yang K., Xia G.-S., Liu Z., Du B., Yang W., Pelillo M., Zhang L., 2020. Semantic Change Detection with Asymmetric Siamese Networks. ArXiv, abs/2010.05687.

Yu F., Koltun V., 2016. Multi-scale context aggregation by dilated convolutions. In International Conference on Learning Representations (ICLR).

Yu F., Koltun V., Funkhouser T., 2017. Dilated residual networks. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Zhang M., Xu G., Chen K., Yan M., Sun X., 2018. Triplet-Based Semantic Relation Learning for Aerial Remote Sensing Image Change Detection. IEEE Geoscience and Remote Sensing Letters. PP. 1-5. 10.1109/LGRS.2018.2869608.

<https://www.copernicus.eu/el> (τελευταία πρόσβαση Μάρτιος 2022)

<https://landsat.gsfc.nasa.gov/data/> (τελευταία πρόσβαση Μάρτιος 2022)

[http://cs231n.stanford.edu/slides/2017/cs231n\\_2017\\_lecture11.pdf](http://cs231n.stanford.edu/slides/2017/cs231n_2017_lecture11.pdf) (τελευταία πρόσβαση Μάρτιος 2022)

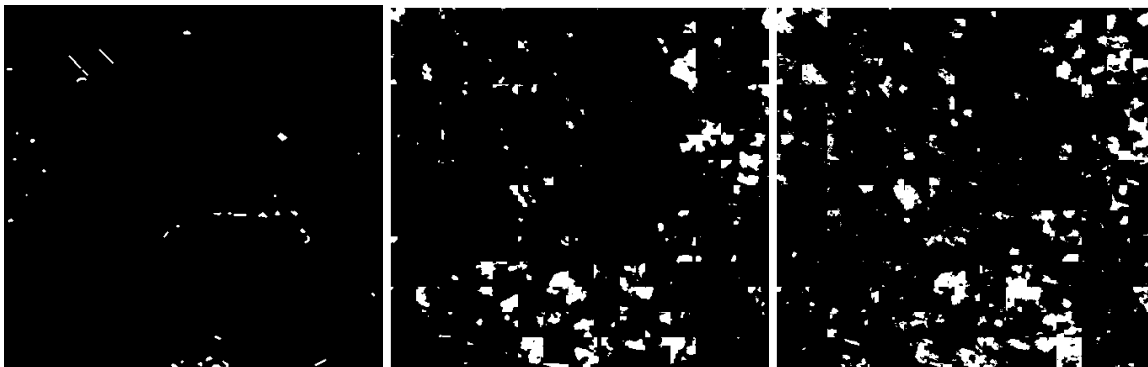
[https://theaisummer.com/Semantic\\_Segmentation/](https://theaisummer.com/Semantic_Segmentation/) (τελευταία πρόσβαση Μάρτιος 2022)

<https://towardsdatascience.com/types-of-convolutions-in-deep-learning-717013397f4d> (τελευταία πρόσβαση Μάρτιος 2022)

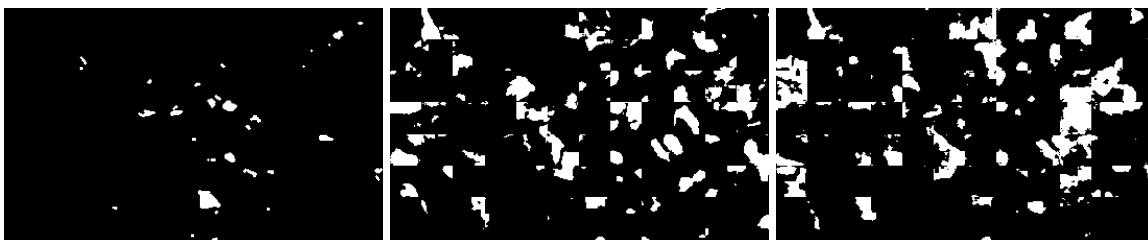
<https://justchenhao.github.io/LEVIR/> (τελευταία πρόσβαση Μάρτιος 2022)

<https://pytorch.org/> (τελευταία πρόσβαση Μάρτιος 2022)

## ΠΑΡΑΡΤΗΜΑ



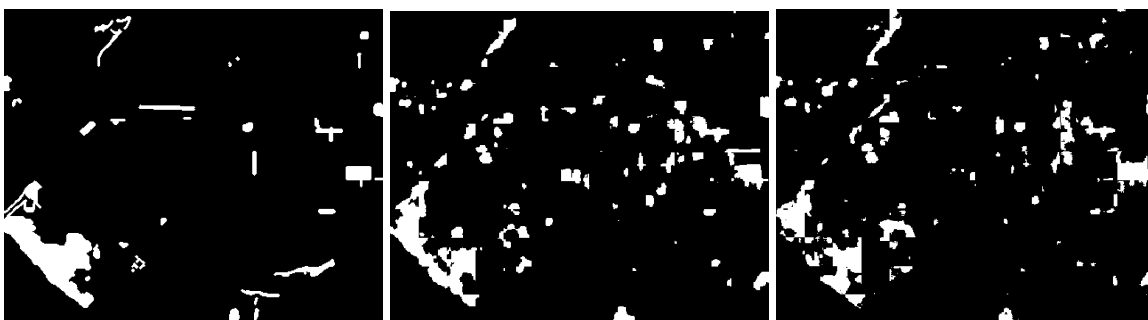
Εικόνα 1. Αληθής χάρτης μεταβολών, χάρτης μεταβολών από L-UNet και από “hybrid\_1” (από αριστερά προς τα δεξιά) [με μεγαλύτερη θετική απόκλιση για το hybrid\_1](Εικόνες Valencia).



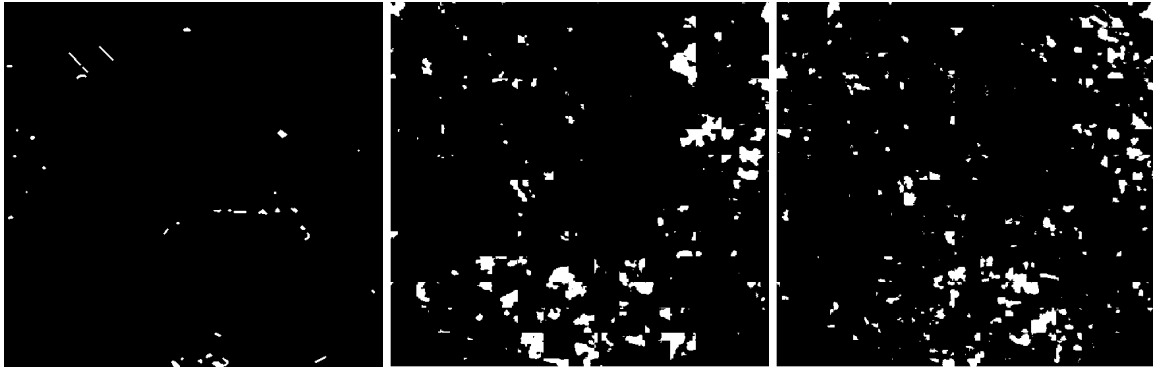
Εικόνα 2. Αληθής χάρτης μεταβολών, χάρτης μεταβολών από L-UNet και από “hybrid\_1” (από αριστερά προς τα δεξιά) [με μεγαλύτερη αρνητική απόκλιση για το hybrid\_1](Εικόνες Norcia).



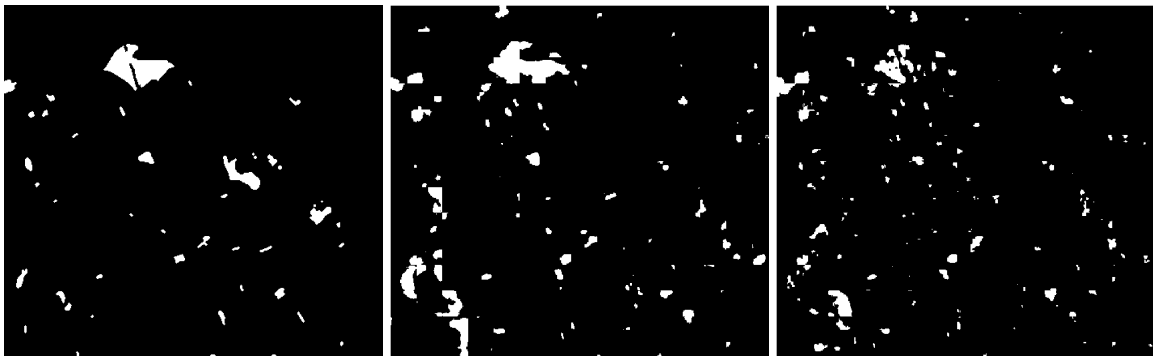
Εικόνα 3. Αληθής χάρτης μεταβολών, χάρτης μεταβολών από L-UNet και από “hybrid\_12” (από αριστερά προς τα δεξιά) [με μεγαλύτερη θετική απόκλιση για το hybrid\_12](Εικόνες Saclay\_w).



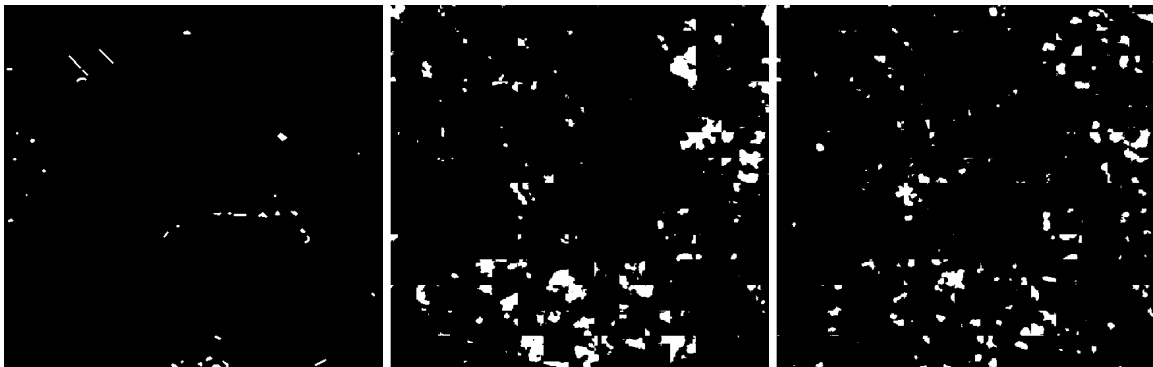
Εικόνα 4. Αληθής χάρτης μεταβολών, χάρτης μεταβολών από L-UNet και από “hybrid\_12” (από αριστερά προς τα δεξιά) [με μεγαλύτερη αρνητική απόκλιση για το hybrid\_12](Εικόνες Rio).



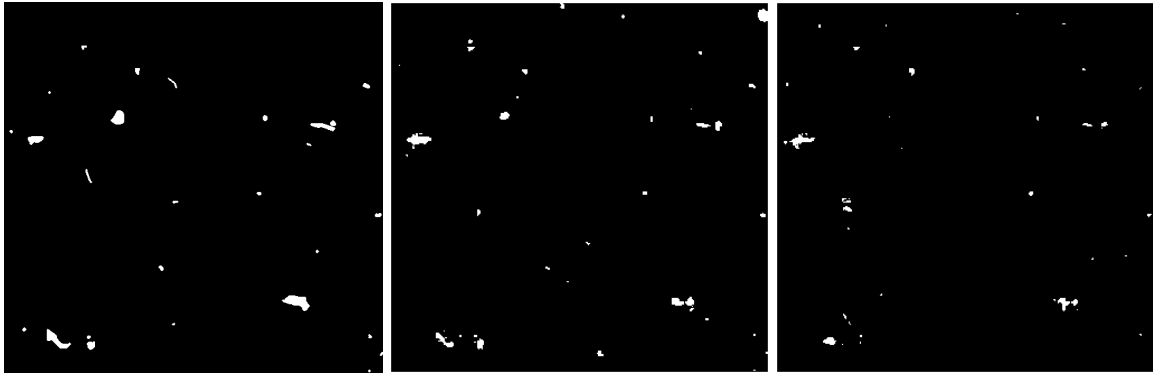
Εικόνα 5. Αληθής χάρτης μεταβολών, χάρτης μεταβολών από L-UNet και από “hybrid\_123” (από αριστερά προς τα δεξιά) [με μεγαλύτερη θετική απόκλιση για το hybrid\_123] (Εικόνες Valencia).



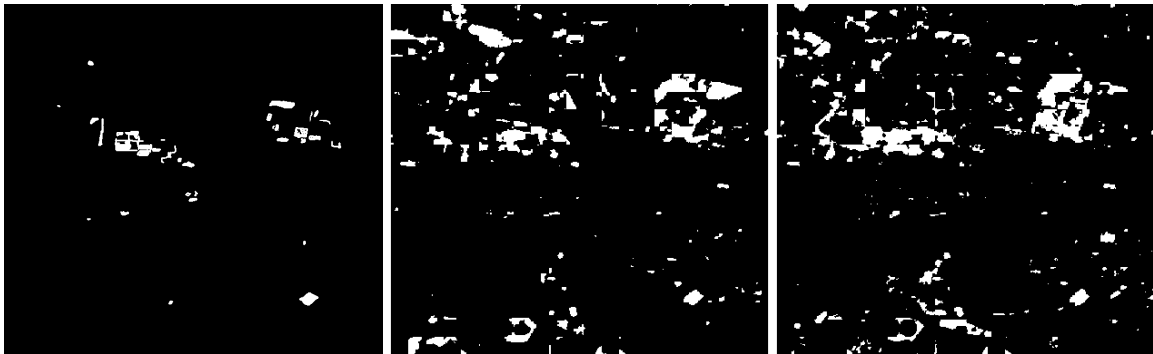
Εικόνα 6. Αληθής χάρτης μεταβολών, χάρτης μεταβολών από L-UNet και από “hybrid\_123” (από αριστερά προς τα δεξιά) [με μεγαλύτερη αρνητική απόκλιση για το hybrid\_123] (Εικόνες Brasilia).



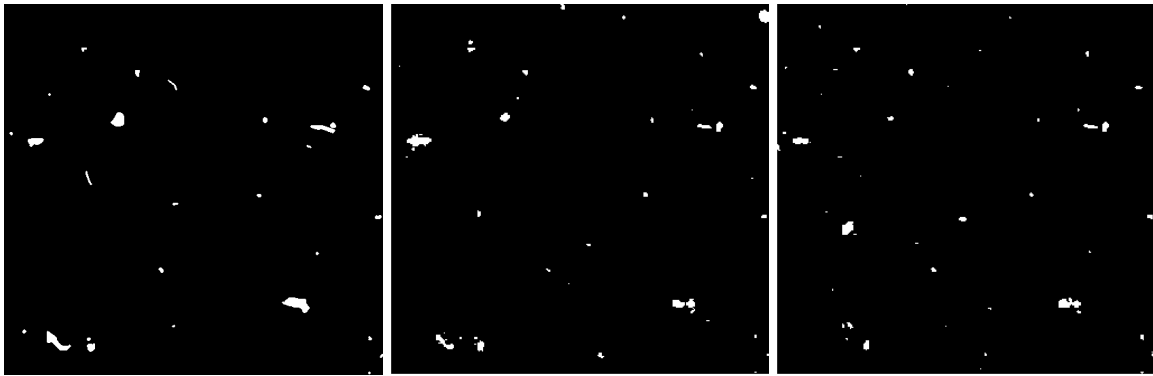
Εικόνα 7. Αληθής χάρτης μεταβολών, χάρτης μεταβολών από L-UNet και από “hybrid\_1234” (από αριστερά προς τα δεξιά) [με μεγαλύτερη θετική απόκλιση για το hybrid\_1234 (αλλά αρκετά μικρή)] (Εικόνες Valencia).



Εικόνα 8. Αληθής χάρτης μεταβολών, χάρτης μεταβολών από L-UNet και από “hybrid\_1234” (από αριστερά προς τα δεξιά) [με μεγαλύτερη αρνητική απόκλιση για το hybrid\_1234] (Εικόνες Milano).

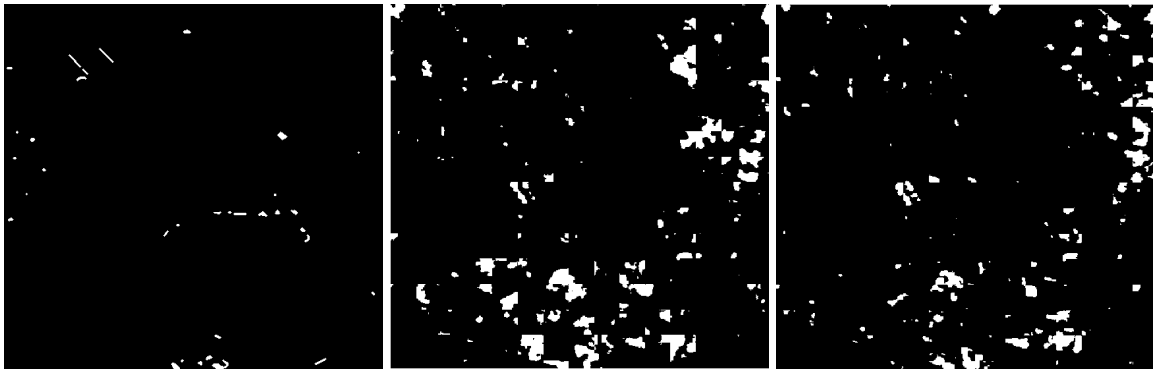


Εικόνα 9. Αληθής χάρτης μεταβολών, χάρτης μεταβολών από L-UNet και από “hybrid\_5” (από αριστερά προς τα δεξιά) [με μεγαλύτερη θετική απόκλιση για το hybrid\_5] (Εικόνες Saclay\_w).



Εικόνα 10. Αληθής χάρτης μεταβολών, χάρτης μεταβολών από L-UNet και από “hybrid\_5” (από αριστερά προς τα δεξιά) [με μεγαλύτερη αρνητική απόκλιση για το hybrid\_5] (Εικόνες Milano).

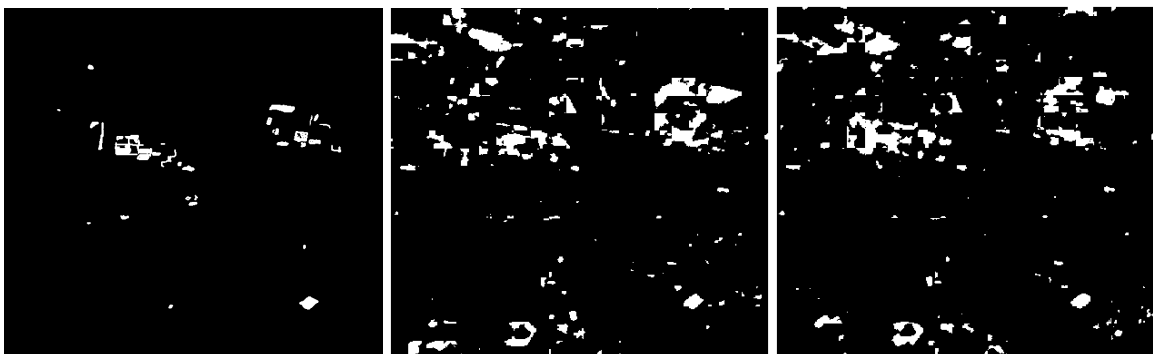




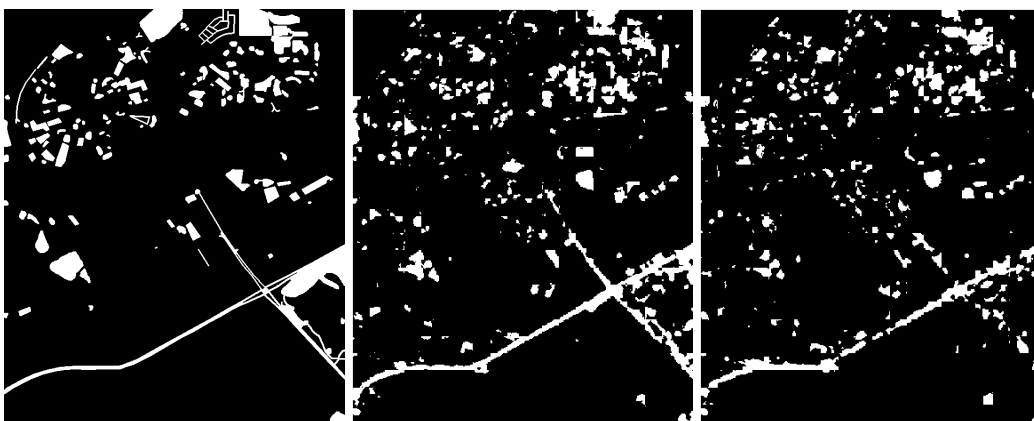
Εικόνα 11. Αληθής χάρτης μεταβολών, χάρτης μεταβολών από L-UNet και από “hybrid\_45” (από αριστερά προς τα δεξιά) [με μεγαλύτερη θετική απόκλιση για το hybrid\_45] (Εικόνες Valencia).



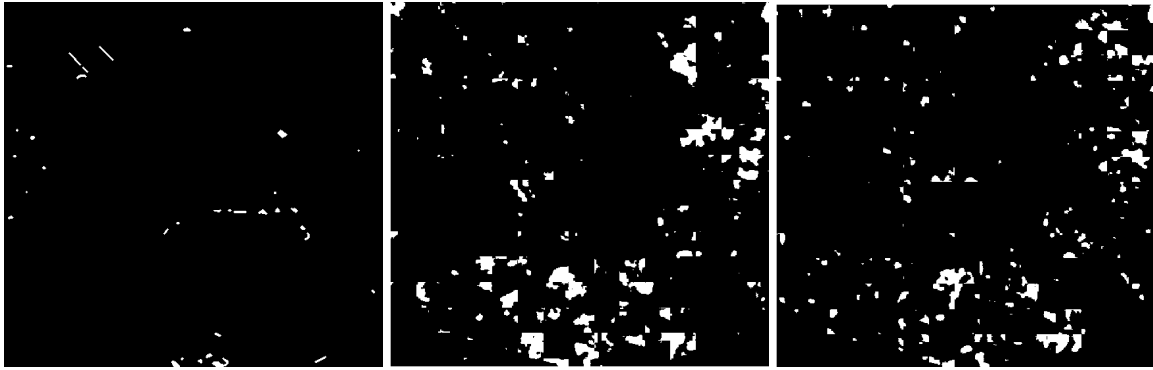
Εικόνα 12. Αληθής χάρτης μεταβολών, χάρτης μεταβολών από L-UNet και από “hybrid\_45” (από αριστερά προς τα δεξιά) [με μεγαλύτερη αρνητική απόκλιση για το hybrid\_45] (Εικόνες Norcia).



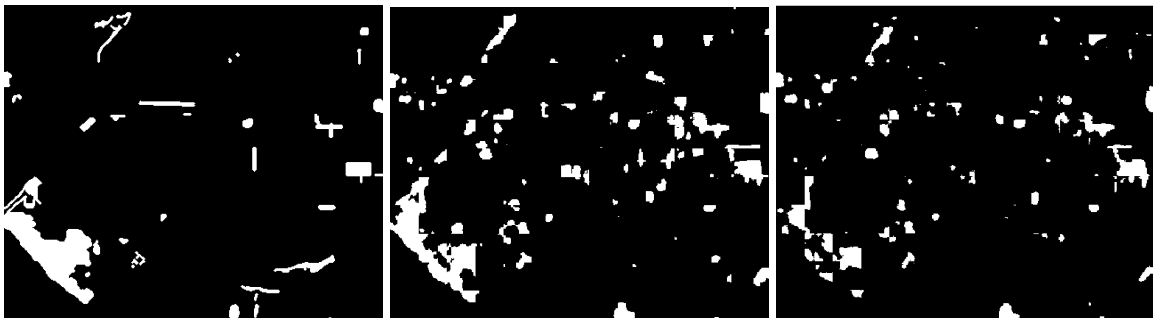
Εικόνα 13. Αληθής χάρτης μεταβολών, χάρτης μεταβολών από L-UNet και από “hybrid\_345” (από αριστερά προς τα δεξιά) [με μεγαλύτερη θετική απόκλιση για το hybrid\_345] (Εικόνες Saclay\_w).



Εικόνα 14. Αληθής χάρτης μεταβολών, χάρτης μεταβολών από L-UNet και από “hybrid\_345” (από αριστερά προς τα δεξιά) [με μεγαλύτερη αρνητική απόκλιση για το hybrid\_345] (Εικόνες Dubai).



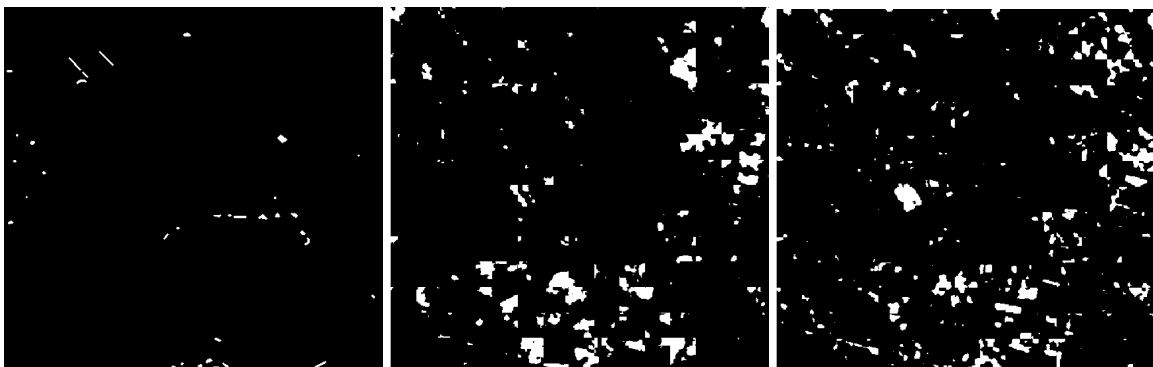
Εικόνα 15. Αληθής χάρτης μεταβολών, χάρτης μεταβολών από L-UNet και από “hybrid\_2345” (από αριστερά προς τα δεξιά) [με μεγαλύτερη θετική απόκλιση για το hybrid\_2345] (Εικόνες Valencia).



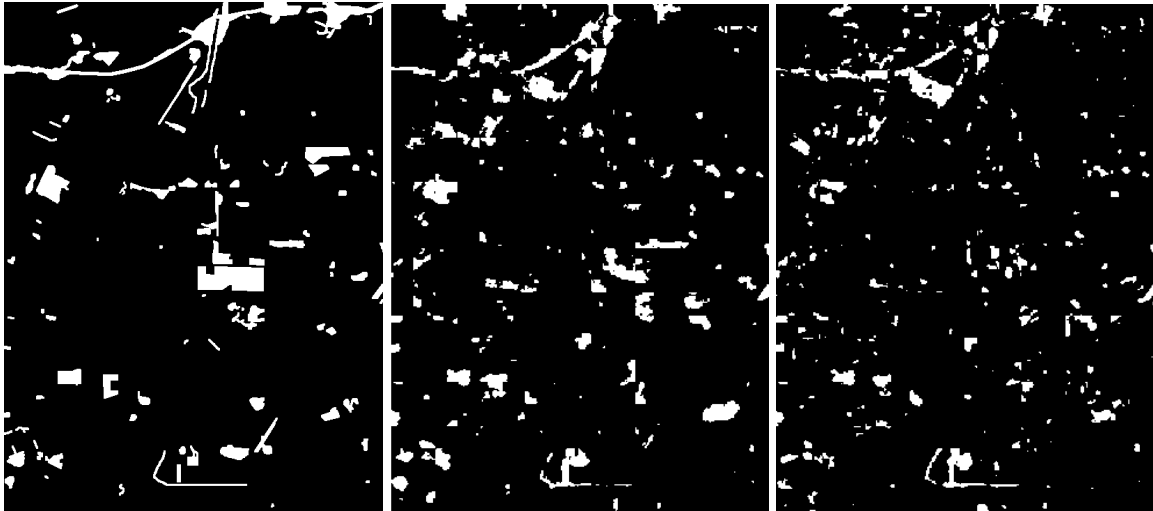
Εικόνα 16. Αληθής χάρτης μεταβολών, χάρτης μεταβολών από L-UNet και από “hybrid\_2345” (από αριστερά προς τα δεξιά) [με μεγαλύτερη αρνητική απόκλιση για το hybrid\_2345] (Εικόνες Rio).



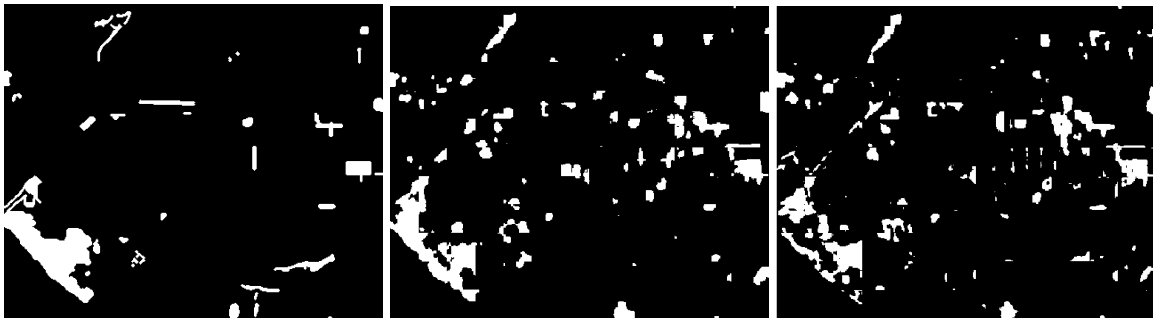
Εικόνα 17. Αληθής χάρτης μεταβολών, χάρτης μεταβολών από L-UNet και από Dilated+LSTM (από αριστερά προς τα δεξιά) [με μεγαλύτερη θετική απόκλιση για το Dilated+LSTM] (Εικόνες Saclay\_w).



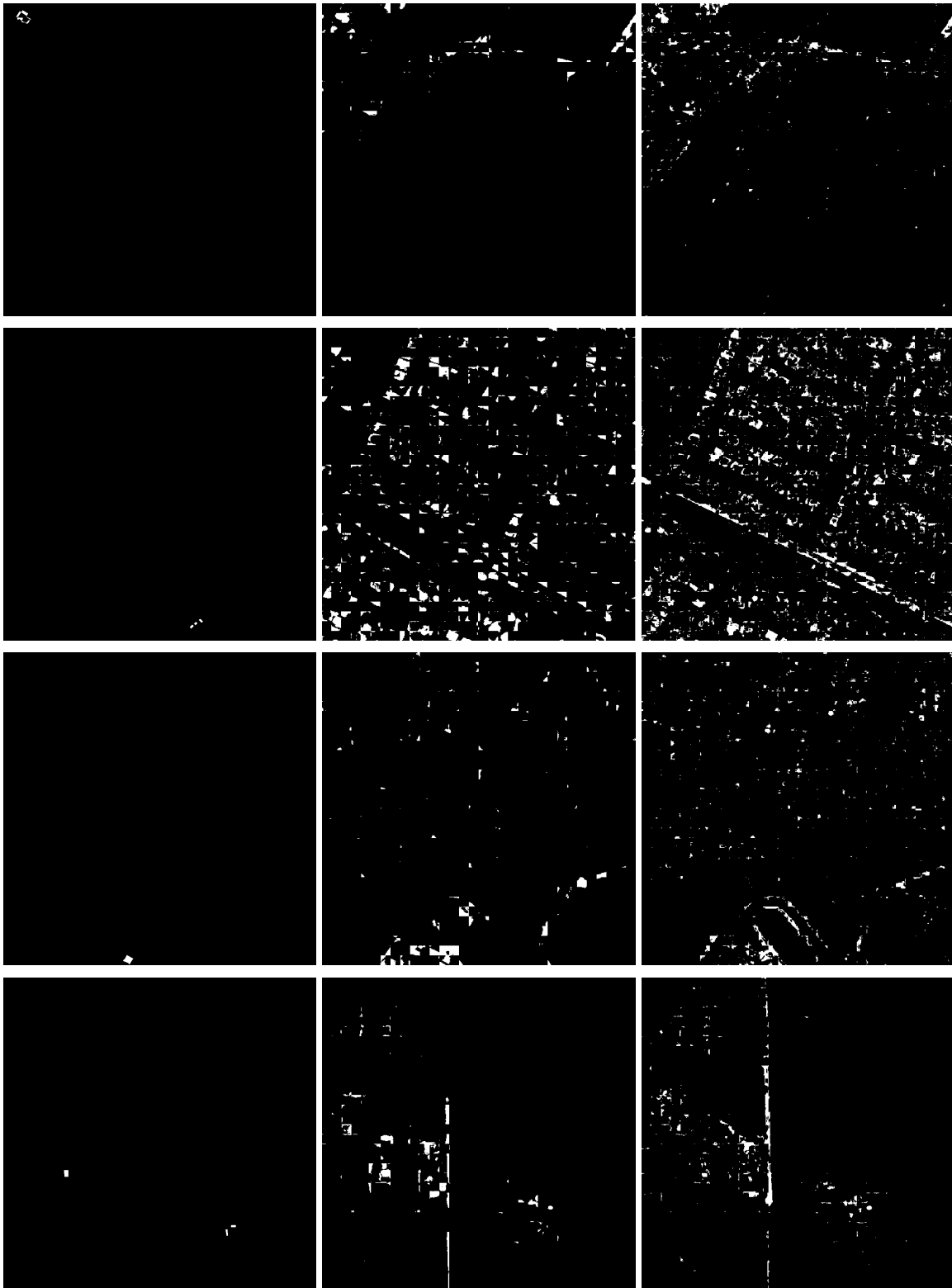
Εικόνα 18. Αληθής χάρτης μεταβολών, χάρτης μεταβολών από L-UNet και από Dilated+LSTM (από αριστερά προς τα δεξιά) [με μεγαλύτερη θετική απόκλιση για το Dilated+LSTM] (Εικόνες Valencia).



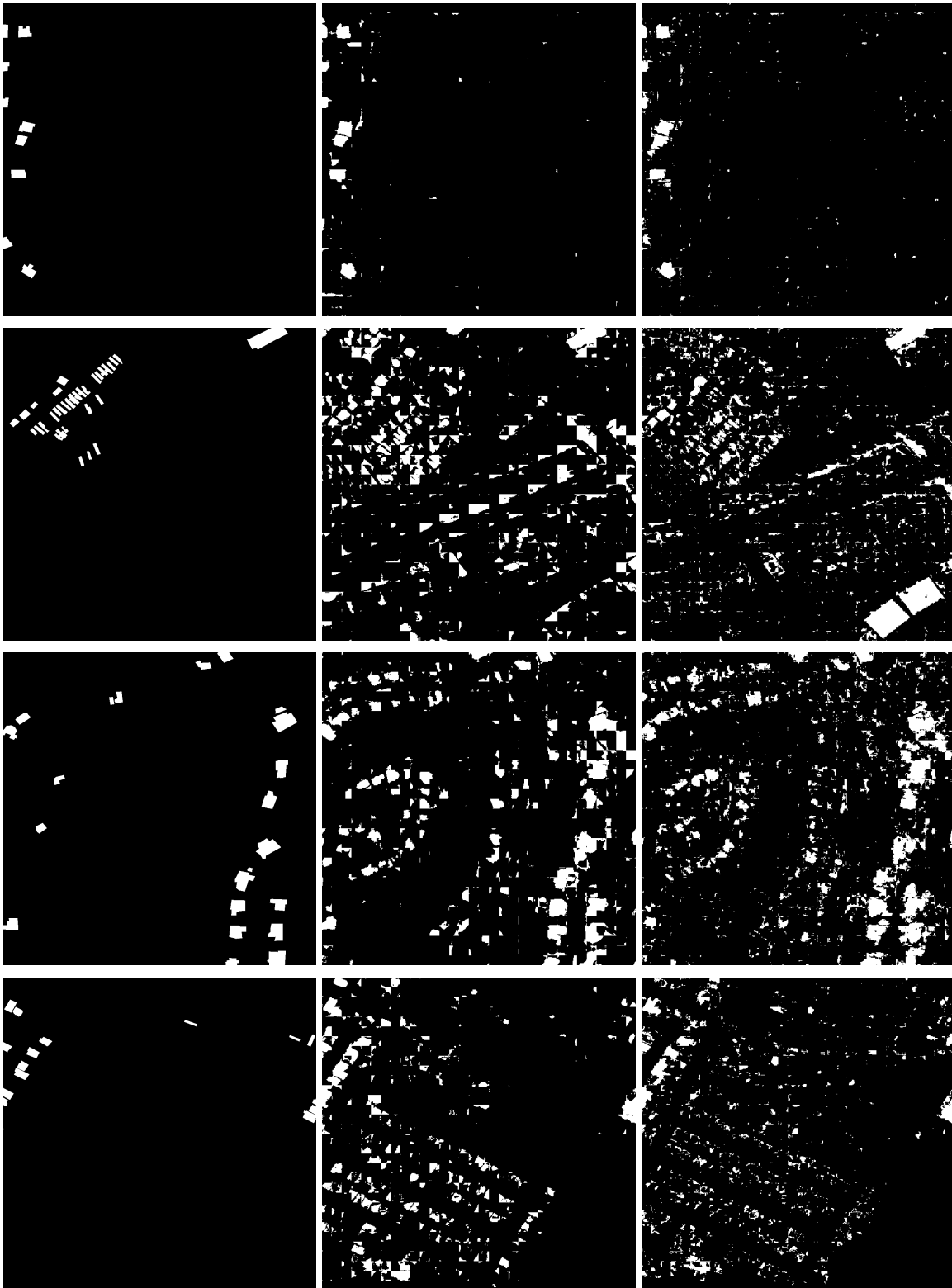
Εικόνα 19. Αληθής χάρτης μεταβολών, χάρτης μεταβολών από L-UNet και από Dilated+LSTM (από αριστερά προς τα δεξιά) [με μεγαλύτερη αρνητική απόκλιση για το Dilated+LSTM] (Εικόνες Chongqing).



Εικόνα 20. Αληθής χάρτης μεταβολών, χάρτης μεταβολών από L-UNet και από Dilated+LSTM (από αριστερά προς τα δεξιά) [με μεγαλύτερη αρνητική απόκλιση για το Dilated+LSTM] (Εικόνες Rio).



Εικόνα 21. Αληθής χάρτης μεταβολών και χάρτες μεταβολών από το L-UNet και από το Dilated+LSTM (από αριστερά προς τα δεξιά)-περιπτώσεις όπου το L-UNet έχει την καλύτερη απόδοση σε σχέση με το Dilated+LSTM.



Εικόνα 22. Αληθής χάρτης μεταβολών και χάρτες μεταβολών από το L-UNet και από το Dilated+LSTM (από αριστερά προς τα δεξιά)-περιπτώσεις όπου το Dilated+LSTM έχει την καλύτερη απόδοση σε σχέση με το L-UNet.